

**UNIVERSIDAD NACIONAL DEL SANTA**  
**FACULTAD DE INGENIERÍA**  
**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E**  
**INFORMÁTICA**



**UNS**  
UNIVERSIDAD  
NACIONAL DEL SANTA

**IMPLEMENTACIÓN DE UN ALGORITMO RECOMENDADOR  
MACHINE LEARNING USANDO DATAMART PARA LA TOMA DE  
DECISIONES EN EL ÁREA DE VENTAS DE LA EMPRESA SPORT  
ANTHONY EN CHIMBOTE**

**Tesis para optar el título de Ingeniero de Sistemas e Informática**

**TESISTA:**

Bach. MARCELO GÓMEZ LUIS RICARDO

**ASESORA:**

DRA. BRIONES PEREYRA, LIZBETH DORA

**Nuevo Chimbote – Perú**

**2022**

**UNIVERSIDAD NACIONAL DEL SANTA**  
**FACULTAD DE INGENIERÍA**  
**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E**  
**INFORMÁTICA**



**IMPLEMENTACIÓN DE UN ALGORITMO RECOMENDADOR**  
**MACHINE LEARNING USANDO DATAMART PARA LA TOMA DE**  
**DECISIONES EN EL ÁREA DE VENTAS DE LA EMPRESA SPORT**  
**ANTHONY EN CHIMBOTE**

**Tesis para optar el título de Ingeniero de Sistemas e Informática**

**Revisado y Aprobado por:**

---

**Dra. Briones Pereyra, Lizbeth Dora**

**Nuevo Chimbote – Perú**

**2022**

**UNIVERSIDAD NACIONAL DEL SANTA**

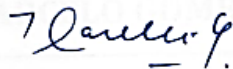
**FACULTAD DE INGENIERÍA**

**ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E  
INFORMÁTICA**

**IMPLEMENTACIÓN DE UN ALGORITMO RECOMENDADOR  
MACHINE LEARNING USANDO DATAMART PARA LA TOMA DE  
DECISIONES EN EL ÁREA DE VENTAS DE LA EMPRESA SPORT  
ANTHONY EN CHIMBOTE**

**Tesis para optar el título de Ingeniero de Sistemas e Informática**

**Revisado y aprobado ante el siguiente jurado:**



**DR. HUGO ESTEBAN CASELLI GISMONDI  
PRESIDENTE**



**MS. MIRKO MARTÍN MANRIQUE RONCEROS  
SECRETARIO**



**DRA. LIZBETH DORA BRIONES PEREYRA  
INTEGRANTE**

**Nuevo Chimbote – Perú**

**2022**



**ACTA DE EVALUACIÓN PARA SUSTENTACIÓN DE TESIS**

En el Campus Universitario de la Universidad Nacional del Santa, siendo las 10:00 am. del día viernes 27 de mayo de 2022, en el Aula S3 del Pabellón nuevo de la EPISI, en atención a la Resolución Decanal N° 264-2022-UNS-FI de Declaración de Expedito de fecha 23.05.2022; se llevó a cabo la instalación del jurado Evaluador, designado mediante Resolución N° 017 - 2022 -UNS- CFI de fecha 21.01.2022, integrado por el **DR. HUGO ESTEBAN CASELLI GISMONDI (Presidente)**, **MS. MIRKO MARTIN MANRIQUE RONCEROS (Secretario)**, **DRA. LIZBETH DORA BRIONES PEREYRA (Integrante)**, para dar inicio a la sustentación del Informe Final de Tesis, cuyo título es: **"IMPLEMENTACIÓN DE UN ALGORITMO RECOMENDADOR MACHINE LEARNING USANDO DATAMART PARA LA TOMA DE DECISIONES EN EL AREA DE VENTAS DE SPORT ANTHONY"** perteneciente al bachiller: **MARCELO GOMEZ LUIS RICARDO** con código de matrícula N° **0201014036**, tienen como **ASESORA** a la **Dra. Lizbeth Dora Briones Pereyra**, según T/R.D. N° 746 -2019-UNS-FI de fecha 04.12.2019.

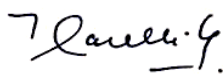
Terminada la sustentación, el tesista respondió a las preguntas formuladas por los miembros del Jurado Evaluador y el público presente.

El Jurado después de deliberar sobre aspectos relacionados con el trabajo, contenido y sustentación del mismo y con las sugerencias pertinentes y en concordancia con el artículo 73º y 103º del Reglamento General de Grados y Títulos, vigente de la Universidad Nacional del Santa; considera la siguiente nota final de Evaluación:

BACHILLER	CALIFICACIÓN	CONDICIÓN
<b>MARCELO GOMEZ LUIS RICARDO</b>	<i>DIECISEIS (16)</i>	<i>BUENO</i>

Siendo la 11: 00 am. se dio por terminado el Acto de Sustentación y en señal de conformidad, firma el Jurado la presente Acta.

Nuevo Chimbote, 27 de mayo de 2022

  
\_\_\_\_\_  
**DR. HUGO ESTEBAN CASELLI GISMONDI**  
PRESIDENTE

  
\_\_\_\_\_  
**MS. MIRKO MARTIN MANRIQUE RONCEROS**  
SECRETARIO

  
\_\_\_\_\_  
**DRA. LIZBETH DORA BRIONES PEREYRA**  
INTEGRANTE

## **DEDICATORIA**

Con todo mi corazón:

A mi familia, en especial a mi madre Nelly, a mi padre Luis, a mis adoradas hermanas Fiorella y Brigitt, a mi abuelita María y a mi abuelito Walter, en quienes estuve pensando en todo mi proceso de aprendizaje y elaboración de mi tesis.

*Luis Ricardo Marcelo Gómez*

## **AGRADECIMIENTOS**

*Terminar la presente investigación no se hubiera dado sin el apoyo incondicional de las personas que siempre están, para quienes dirijo mi especial agradecimiento.*

*A la Dra. Lizbeth Dora, Briones, quien, como Asesora de Tesis, con su paciencia, conocimientos y experiencia, colaboró con el avance, la realización, retroalimentación y depuración de la presente investigación.*

*Al personal Administrativo y técnico de la empresa Sport Anthony, por el soporte y apoyo prestado para llevar a cabo la presente investigación.*

*A mi familia, quienes con amor, orientación y paciencia transmitida a lo largo de toda mi carrera universitaria y luego de egresar, hicieron posible llevar a cabo la presente investigación.*

*Y en su totalidad, a todos y todas, quienes hayan aportado un granito de arena en llevar a cabo el presente informe.*

*Luis Ricardo Marcelo Gómez  
Bach. Ing. Sistemas e Informática*

## PRESENTACIÓN

Señores miembros del jurado:

En cumplimiento a lo dispuesto en el Reglamento General de Grados y Títulos de la Universidad Nacional del Santa, pongo a vuestra disposición el presente informe de tesis titulado: **“IMPLEMENTACIÓN DE UN ALGORITMO RECOMENDADOR MACHINE LEARNING USANDO DATAMART PARA LA TOMA DE DECISIONES EN EL ÁREA DE VENTAS DE LA EMPRESA SPORT ANTHONY EN CHIMBOTE”**, siendo requisito para lograr el título Profesional de Ingeniero de Sistemas e Informática.

La presente investigación tiene como lugar de aplicación a la Empresa Sport Anthony ubicada en Chimbote, de la cual, su propósito es la implementación de un Algoritmo Recomendador Machine Learning para la toma de decisiones en el área de ventas de la empresa Sport Anthony en Chimbote.

Esperando se cubra con las exigencias que se solicitan en las leyes vigentes y requisitos de la universidad para la respectiva aprobación de la presente investigación, quedo de Ustedes.

Atentamente;

*Luis Ricardo Marcelo Gómez*  
*Bach. Ing. Sistemas e Informática*

## RESUMEN

La presente investigación está orientada a la implementación de un Algoritmo Machine Learning, usando DATAMART para la toma de decisiones en el área de ventas de la empresa Sport Anthony. Se utilizó la Metodología de Ciencia de Datos (CRISP-DM). El tipo de investigación es cuasi experimental, útil para hacer la comparación entre un escenario previo con la experimentación de los resultantes que se obtiene luego de esta. De esta manera tendremos una medición previa de la variable independiente, algoritmo recomendador (Pre - test), su aplicación y una nueva medida para la variable dependiente luego de aplicar la independiente (Pos - test). También usamos herramientas como Google Analytics para exportar datos, Visual Studio 2017 para la extracción y transformación de datos, SQL server 2012 para almacenar los datos, Anaconda Navigator y Jupyter Notebook para implementar el algoritmo en Python, entre otras que hacen posible la implementación del algoritmo usando DATAMART para la toma de decisiones en el área de ventas de la empresa Sport Anthony, con el apoyo también del personal representante de la empresa, el personal de ventas y marketing, 6 personas, las cuales forman parte de la muestra para los indicadores cualitativos, mientras que, con los indicadores cuantitativos, 11 con respecto a la cantidad de otros Recomendadores y 61 que corresponden a los tiempos, días en que se hicieron las pruebas. Los resultados obtenidos, como el grado de satisfacción de los usuarios, que se ha incrementado en un 28,74%, la calidad de la Información con un incremento del 13,2%, y con respecto a los tiempos de búsqueda de productos a ofertar, una reducción del 93,88%, se tiene también una reducción de costos por la utilización del Algoritmo Recomendador Propuesto en un 78,78%, y un incremento de ventas del 55,82%, los cuales corroboran la hipótesis planteada, siendo útiles para la toma de decisiones de la empresa del estudio.

**PALABRAS CLAVES:** Algoritmo Recomendador, SVD, Machine Learning, DataMart, Python.



## ABSTRACT

The present investigation is oriented to the implementation of a Machine Learning Algorithm, using DATAMART for decision making in the sales area of the company Sport Anthony. In the development of this research, the Data Science Methodology (CRISP-DM) was used, the type of research is quasi experimental, which helps us to make the comparison between a previous situation with the experimentation of the results obtained. after this. In this way we will have a previous measurement of the independent variable, recommender algorithm (Pretest), the application and a new measurement of the dependent variable after application of the independent variable (Posttest). We also use tools such as Google Analytics to export data, Visual Studio 2017 for data extraction and transformation, SQL server 2012 to store the data, Anaconda Navigator and Jupyter Notebook to implement the algorithm in Python, among others that make the implementation of the algorithm possible. using DATAMART for decision making in the sales area of the company Sport Anthony, also with the support of the representative staff of the company, the sales and marketing staff, 6 people, who are part of the sample for the qualitative indicators, while with the quantitative indicators, 11 with respect to the number of other Recommenders and 61 that correspond to the times, days in which the tests were done.

The results obtained such as the degree of user satisfaction, which has increased by 28.74%, the quality of the Information with an increase of 13.2%, and with respect to the search times for products to be offered, a reduction of 93.88%, there is also a reduction in costs due to the use of the Proposed Recommender Algorithm by 78.78%, and an increase in sales of 55.82%, which corroborating the hypothesis raised, being useful for decision-making of the company of the present investigation.

**KEY WORDS:** Recommender Algorithm, SVD, Machine Learning, DataMart, Python.

## INTRODUCCIÓN

Hoy en día si tu negocio no está en la red, estarás perdiendo mucha rentabilidad, hoy en día si no analizas los datos de tu empresa, estarás perdiendo mucho valor, el comercio electrónico está muy ligado a la ciencia de datos, hay mucha data por analizar en un sitio web y su histórico, comportamiento de los usuarios, historial de visitas, historial de compras, porcentajes de rebotes, etc; si bien es cierto, Machine Learning, como una de las ramas de la IA, y que se puede desplegar en distintas materias y escenarios, el machine learning como aliado en el ecommerce nos puede brindar protección de datos, resultados de búsquedas optimizados, precios optimizados, en la presente investigación su uso está enfocado en el ecommerce y en la segmentación de clientes y recomendaciones de productos.

Los volúmenes de datos y el saber darles valor para la toma de decisiones se ha ido incrementando en los años, es muy común oír que los datos son el nuevo petróleo, pero creo que al igual que el petróleo, para que deje de ser crudo pasa por procesos de refinamientos, lo mismo pasa con los volúmenes de datos, para obtener valor e involucrarlos a la toma de decisiones, hay que extraerlos, transformarlos, cargarlos, procesarlos, analizarlos, etc; y en ese sentido podríamos definir al Machine Learning, como la capacidad para aprender de la data y para poder utilizarlo en la toma de decisiones. (Rouhiainen, 2018).

La presente investigación tiene como objetivo hacer uso de un algoritmo recomendador Machine Learning para la toma de decisiones en el área de ventas de la empresa Sport Anthony, utilizando Python y siguiendo la metodología CRISP-DM.

Para el Capítulo:

- **I:** Donde muestra el problema, se justifica el estudio y posterior a ello definiendo las limitaciones, se procede a detallar los objetivos.

- **II:** Se presenta la teoría, términos y definiciones que se usaron para la comprensión en el proceso de elaboración de la investigación.
- **III:** Se describe el proceso de la metodología utilizada para esta investigación, CRISP – DM, y las fases que le pertenecen, la cual funciona como un espiral por sus ciclos de retroalimentación.
- **IV:** Se presenta la hipótesis, las variables, los indicadores cuantitativos y cualitativos y los métodos que comprueban la veracidad de la hipótesis planteada en la presente investigación.
- **V:** Finalmente lo que se concluye y se recomienda del presente estudio y las encuestas que se usaron como anexos para probar la hipótesis, de igual manera la factibilidad de la investigación.

## ÍNDICE

DEDICATORIA.....	- 5 -
AGRADECIMIENTOS .....	- 6 -
PRESENTACIÓN.....	- 7 -
RESUMEN .....	- 8
-	
INTRODUCCIÓN .....	- 10 -
ÍNDICE.....	- 12 -
ÍNDICE DE TABLAS .....	- 16 -
ÍNDICE DE FIGURAS.....	- 17 -
CAPÍTULO I	
PROBLEMA DE LA INVESTIGACIÓN .....	21
1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA .....	22
1.3. ANTECEDENTES DE LA INVESTIGACIÓN.....	- 25 -
1.3.1.A Nivel Internacional.....	- 25 -
1.3.2. A Nivel Nacional.....	- 27 -
1.3.3. A Nivel Local .....	- 29 -
1.3 FORMULACIÓN DEL PROBLEMA DE INVESTIGACIÓN .....	30
1.4 DELIMITACIÓN DEL ESTUDIO .....	30
1.5 JUSTIFICACIÓN DE LA INVESTIGACIÓN .....	31
1.6. LIMITACIONES .....	- 32 -
1.7 OBJETIVOS DE LA INVESTIGACIÓN: .....	- 33 -
1.7.1 Objetivo general.....	33

1.7.2 Objetivos específicos. ....	- 33 -
-----------------------------------	--------

## CAPÍTULO II

MARCO TEÓRICO.....	- 34 -
--------------------	--------

2.1 FUNDAMENTOS .....	- 35 -
-----------------------	--------

2.1.1. Toma de Decisiones. ....	- 35 -
---------------------------------	--------

2.1.2. Machine Learning.....	- 35 -
------------------------------	--------

2.1.3. SVD.....	- 35 -
-----------------	--------

2.1.4. Anaconda Navigator.....	- 36 -
--------------------------------	--------

2.1.5. Jupyter Nothebook.....	- 37 -
-------------------------------	--------

2.1.6. Python.....	- 37 -
--------------------	--------

2.1.7. Inteligencia de Negocios (BI).....	- 37 -
---	--------

2.1.8. Ciencia de Datos.....	- 38 -
------------------------------	--------

2.1.9. Metodología CRISP-DM .....	- 38 -
-----------------------------------	--------

2.1.10. ETL .....	- 38 -
-------------------	--------

2.1.11. DATAMART .....	- 38 -
------------------------	--------

2.1.11. Proceso Analítico en Línea (OLAP).....	- 39 -
--	--------

2.1.12. SQL SERVER.....	- 39 -
-------------------------	--------

## CAPÍTULO III

DESARROLLO DE LA METODOLOGÍA.....	- 40 -
-----------------------------------	--------

3.1. Fase de comprensión del negocio .....	41
--	----

3.1.1. Determinar el objetivo del negocio .....	41
---	----

3.1.2. Valorar la situación.....	42
----------------------------------	----

3.1.3. Realizar un plan de proyecto .....	45
---	----

3.2. Fase de comprensión de datos .....	45
---	----

3.2.1. Recolección de datos iniciales.....	46
--	----

3.2.2. Descripción de los datos .....	46
3.2.3. Verificación de la calidad de los datos .....	46
3.3. Fase de preparación de los datos.....	47
3.3.1. Selección de datos - Implementación: Proceso ETL .....	47
3.3.2. Limpieza de datos - Implementación: Proceso ETL .....	48
3.3.3. Limpieza de datos - tablas .....	50
3.3.4. Estructuración, Integración y formateo de los datos.....	51
3.4. Fase de modelado.....	73
3.4.1. Selección de la técnica del modelo .....	73
3.4.2. Generación de un plan de prueba.....	74
3.4.3. Construcción del modelo .....	75
3.4.4. Evaluación del modelo.....	75

## CAPÍTULO IV

PRUEBA DE HIPÓTESIS .....	- 89 -
4.1. Indicadores Cualitativos .....	90
4.1.1. Grado de satisfacción De Los Usuarios - Cálculo.....	- 91 -
4.1.2. Calidad de las recomendaciones de los productos a ofertar con el Alg. ML – Cálculo .....	- 99 -
4.2. Indicadores Cuantitativos .....	- 106 -
4.2.1. Indicador cuantitativo – Prueba de Hipótesis: .....	- 106 -
Tiempos de búsqueda de productos a ofertar.....	- 106 -
4.2.2. Indicador cuantitativo – Prueba de Hipótesis: .....	- 112 -
Costos por el uso del algoritmo recomendador Machine Learning .....	- 112 -
4.2.3. Prueba de la Hipótesis para el indicador cuantitativo:.....	- 116 -
Ventas al usar el Algoritmo de Machine Learning .....	- 116 -

4.3. Resultados .....	- 121 -
4.3.1. Indicador Cualitativo: Grado de satisfacción de los usuarios.....	- 121 -
4.3.2. Indicador Cualitativo: Calidad de la información. ....	- 122 -
4.3.3. Indicador Cuantitativo: Tiempos de búsqueda de productos a ofertar. ....	- 123 -
4.3.4. Costos por obtención de recomendaciones con el algoritmo de ML.....	- 124 -
4.3.5. Ventas usando el algoritmo de Machine Learning.....	- 125 -
CAPÍTULO V CONCLUSIÓN Y RECOMENDACIONES .....	- 126 -
5.1 Conclusiones.....	- 127 -
5.2 Recomendaciones .....	- 128 -
REFERENCIAS.....	- 130 -
ANEXO A.....	- 134 -
ANEXO B.....	- 135 -
CUESTIONARIOS.....	- 135 -
6.1. Cuestionario de Nivel de Satisfacción - Sin Algoritmo Recomendador.....	- 135 -
6.2. Cuestionario de Nivel de Satisfacción - Con Algoritmo Recomendador .....	- 136 -
6.3. Cuestionario de Calidad de la Información - Sin Algoritmo Rec. ....	- 137 -
6.4. Cuestionario de Calidad de la Información - Con Algoritmo Rec. ....	- 138 -
ANEXO C .....	- 139 -
ALFA DE CRONBACH.....	- 139 -
7.1. Indicador: Nivel de satisfacción de los usuarios - ENCUESTA.....	- 139 -
7.2. Indicador: Calidad de la Información - ENCUESTA.....	- 140 -

## ÍNDICE DE TABLAS

Tabla N° 1: SVD frente a otros Algoritmos Recomendadores .....	36 -
Tabla N° 2: Tabla de Terminología. ....	43
Tabla N° 3: Tabla de Costes - Hardware.....	43
Tabla N° 4: Tabla de Costes – Software .....	44
Tabla N° 5: Tabla de Costes – Internet .....	44
Tabla N° 6: Tabla de Costes – Recursos Humanos.....	44
Tabla N° 7: Tabla de Costes – Materiales.....	44
Tabla N° 8: Conformación de equipo y usuarios .....	45
Tabla N° 9: Fases .....	45
Tabla N° 10: Calidad de datos, Código – Término .....	47
Tabla N° 11: Prueba estadística de los indicadores.....	90
Tabla N° 12: Rango de Valores.....	91
Tabla N° 13: Ponderación de Indicador “Grado de satisfacción de los usuarios” .....	92 -
Tabla N° 14: Tabulación - Grado de satisfacción de los usuarios. (Pre Test). ....	94 -
Tabla N° 15: Tabulación – Grado de satisfacción de los usuarios. (Post Test) .....	95 -
Tabla N° 16: Contrastación - Resultados Pre & Post Test.....	96 -
Tabla N° 17: Ponderación de Indicador: “Calidad de la información obtenida con las recomendaciones del algoritmo Machine Learning” .....	100
Tabla N° 18: Tabulación para la Calidad de la información (Pre Test) Obtenida con las recomendaciones del algoritmo Machine Learning. ....	101
Tabla N° 19: Tabulación para la Calidad de la información (Post Test) Obtenida con las recomendaciones del algoritmo Machine Learning .....	102 -
Tabla N° 20: Contrastación - Resultados Pre & Post Test.....	103 -
Tabla N° 21: Toma de Tiempos de búsquedas de productos a Ofertar:.....	108 -



Tabla N° 22: Contrastación de los <i>Costos por el uso del algoritmo recomendador Machine Learning</i> .....	- 112 -
Tabla N° 23: Ventas al usar el algoritmo de Machine Learning.....	- 117 -
Tabla N° 24: Presentación de Resultados del indicador Grado de satisfacción de los usuarios. .....	- 121 -
Tabla N° 25: Resultados - <i>Calidad de la Información</i> .....	- 122 -
Tabla N° 26: Resultados - <i>Tiempos de búsqueda de productos a ofertar</i> .....	- 123 -
Tabla N° 27: Resultados - Costos por obtención de recomendaciones con el algoritmo Machine Learning. ....	- 124 -
Tabla N° 28: Resultados - <i>Ventas usando el algoritmo de Machine Learning</i> .....	- 125 -
Tabla N° 29 Rangos y magnitud de confiabilidad por Alfa de Cronbach.....	- 139 -
Tabla N° 30 Indicador Nivel de Satisfacción - Estadísticos .....	- 139 -
Tabla N° 31 Indicador Calidad de la Información - Estadísticos.....	- 140 -

## ÍNDICE DE FIGURAS

<i>Figura N° 1</i> : Nuevo proyecto - Business Intelligence. ....	47
<i>Figura N° 2</i> : Origen de datos – Prueba de conexión .....	48
<i>Figura N° 3</i> : Destino de datos – Prueba de conexión.....	49
<i>Figura N° 4</i> : Data Mart de ventas - ETL .....	50
<i>Figura N° 5</i> : Script Limpiar tablas. ....	50
<i>Figura N° 6</i> : ETL Cargar Productos.....	51
<i>Figura N° 7</i> : ETL Cargar Productos – Leer Productos.....	51
<i>Figura N° 8</i> : ETL Cargar Productos – Obtener Talla - Script.....	52
<i>Figura N° 9</i> : ETL Cargar Productos – Obtener Talla – Entradas y Salidas.....	53
<i>Figura N° 10</i> : ETL Cargar Productos – Destino – Conexión y Asignaciones .....	54

<i>Figura N° 11:</i> ETL Cargar Productos – Ejecución .....	55
<i>Figura N° 12:</i> ETL Cargar Regiones .....	55
<i>Figura N° 13:</i> ETL Cargar Regiones – Leer Regiones.....	55
<i>Figura N° 14:</i> ETL Cargar Regiones – Conversión de datos .....	56
<i>Figura N° 15:</i> ETL Cargar Regiones – Ordenar datos .....	56
<i>Figura N° 16:</i> ETL Cargar Regiones – Obtener ID - Script .....	57
<i>Figura N° 17:</i> ETL Cargar Regiones – Obtener ID – Columnas, entradas y salidas.....	58
<i>Figura N° 18:</i> ETL Cargar Regiones – Destino – Conexión y Asignaciones .....	59
<i>Figura N° 19:</i> ETL Cargar Regiones – Ejecución .....	60
<i>Figura N° 20:</i> ETL Cargar Páginas .....	60
<i>Figura N° 21:</i> ETL Cargar Regiones – Leer Regiones.....	60
<i>Figura N° 22:</i> ETL Cargar Páginas – Obtener ID .....	61
<i>Figura N° 23:</i> ETL Cargar Páginas – Destino .....	62
<i>Figura N° 24:</i> ETL Leer Fechas .....	63
<i>Figura N° 25:</i> ETL Dividir Fechas – Script.....	64
<i>Figura N° 26:</i> ETL Destino Fechas Conexión BD-BIUNS.....	65
<i>Figura N° 27:</i> ETL Destino Fechas especificaciones Anio, mes, dia, id_tiempo.....	65
<i>Figura N° 28:</i> Dimensión Visitas.....	66
<i>Figura N° 29:</i> ETL Dimensión Visitas – Leer Visitas.....	66
<i>Figura N° 30:</i> ETL Dimensión Visitas – Ordenar Visitas.....	66
<i>Figura N° 31:</i> ETL Dimensión Visitas – Leer Región .....	67
<i>Figura N° 32:</i> ETL Dimensión Visitas – Ordenar 1 (Región).....	67
<i>Figura N° 33:</i> ETL Dimensión Visitas – Combinación de mezcla.....	68
<i>Figura N° 34:</i> ETL Dimensión Visitas – Ordenar 3 .....	68
<i>Figura N° 35:</i> ETL Dimensión Visitas – Leer producto:.....	69

<i>Figura N° 36:</i> ETL Dimensión Visitas – Ordenar 2 .....	70
<i>Figura N° 37:</i> ETL Dimensión Visitas – Combinación de mezcla 1 (Visitass, Regiones, Productos) .....	70
<i>Figura N° 38:</i> ETL Destino Visitas Conexión BD-BIUNS .....	71
<i>Figura N° 39:</i> ETL Destino Malla Ejecución .....	72
<i>Figura N° 40:</i> Cubo .....	72
<i>Figura N° 41:</i> Arquitectura .....	74
<i>Figura N° 42:</i> Dataset de Productos/Codigo/Región/Visitass/Link .....	75
<i>Figura N° 43:</i> Importación Numpy, Pandas .....	76
<i>Figura N° 44:</i> Creación de los Diccionarios .....	76
<i>Figura N° 45:</i> Creación Matriz – Pivot Table.....	77
<i>Figura N° 46:</i> Matriz_visitas .....	77
<i>Figura N° 47:</i> Identificadores de Regiones y de Productos.....	78
<i>Figura N° 48:</i> Convertir el formato .....	78
<i>Figura N° 49:</i> Mostrar Matriz recreada.....	79
<i>Figura N° 50:</i> Mostrar Matriz_svd .....	80
<i>Figura N° 51:</i> Mostrar Matriz index_sort .....	80
<i>Figura N° 52:</i> Matriz True/False productos_no_visitados .....	81
<i>Figura N° 53:</i> Mostrar 5 Id de los productos recomendados .....	82
<i>Figura N° 54:</i> Función Recomendar .....	83
<i>Figura N° 55:</i> Recomendaciones Ancash .....	83
<i>Figura N° 56:</i> Recomendaciones Arequipa .....	84
<i>Figura N° 57:</i> Recomendaciones Cusco .....	84
<i>Figura N° 58:</i> Recomendaciones Ica .....	85
<i>Figura N° 59:</i> Recomendaciones La Libertad .....	85

<i>Figura N° 60: Recomendaciones Lambayeque</i> .....	86
<i>Figura N° 61: Recomendaciones Lima Región</i> .....	86
<i>Figura N° 62: Recomendaciones Tacna</i> .....	- 87 -
<i>Figura N° 63: Recomendaciones Loreto</i> .....	- 87 -
<i>Figura N° 64: Recomendaciones Piura</i> .....	- 88 -
<i>Figura N° 65: Resumen Recomendaciones</i> .....	- 88 -
<i>Figura N° 66: Aceptación y Rechazo “Región” – Satisfacción de los usuarios</i> .....	- 99 -
<i>Figura N° 67: Aceptación y Rechazo “Región” – Calidad de la información obtenida con las recomendaciones del algoritmo Machine Learning</i> .....	- 106 -
<i>Figura N° 68: Aceptación y Rechazo “Región” – Tiempos de búsqueda</i> .....	- 112 -
<i>Figura N° 69: Aceptación y Rechazo “Región” – Costos por el uso del Algoritmo Recomendador Machine Learning</i> .....	- 115 -
<i>Figura N° 70: Aceptación y Rechazo “Región” – Ventas al usar el algoritmo de Machine Learning</i> .....	- 120 -
<i>Figura N° 71: Resultados - Grado de satisfacción de los Usuarios</i> .....	- 121 -
<i>Figura N° 72: Resultados - Calidad de la Información</i> .....	- 122 -
<i>Figura N° 73: Resultados del Indicador Tiempos de búsqueda</i> .....	- 123 -
<i>Figura N° 74: Resultados del Indicador Costos por obtención de recomendaciones con el algoritmo de Machine Learning</i> .....	- 124 -
<i>Figura N° 75: Gráficos de Resultados del Indicador Costos por obtención de recomendaciones con el algoritmo de Machine Learning</i> .....	- 125 -
<i>Figura N° 76 Cuestionario de Nivel de Satisfacción - Sin Algoritmo Recomendador</i> .-	135 -
<i>Figura N° 77 Cuestionario de Nivel de Satisfacción - Con Algoritmo Recomendador</i> -	136 -
<i>Figura N° 78 Cuestionario de Calidad de la Información - Sin Algoritmo Rec</i> .....	- 137 -
<i>Figura N° 79 Cuestionario de Calidad de la Información - Con Algoritmo Rec</i> .....	- 138 -

## **CAPÍTULO I**

### **PROBLEMA DE LA INVESTIGACIÓN**

## **1. Problema de la investigación**

### **1.1. Descripción de la realidad problemática**

“Sport Anthony” tiene como principal actividad la comercialización de calzados casuales, deportivos, ropa y accesorios, tras sus 5 años en el mercado, está en pleno crecimiento, de la misma forma su información, esta información la tienen guardadas o almacenadas en hojas de cálculo, en un sistema de escritorio, es decir; un sistema transaccional, y además de vender online en una web ecommerce que tiene deficiencias en la parte de ventas; con estas herramientas actuales, a la empresa le toma tiempo demás, tener información valiosa y así mejores resultados, las cuales se transforman en una excelente alternativa para saber qué acciones se deberán tomar. Las bases de datos son distintas y de distintas plataformas, lo que genera que la obtención y análisis de la información demore más tiempo, de igual forma una generación de cálculos en tablas dinámicas engorrosos y poco ordenados.

Tomaría horas de horas ponerse a clasificar, ordenar, escoger, analizar qué productos elegir para ofertar y se tiene el riesgo de lo que escojamos, no de un impacto esperado, no se tiene una base de la cuál afianzarse y decir, estos productos son ganadores, voy a ofertarlos.

También obtener un sistema de recomendación que no afecte en las utilidades del negocio, pues no es muy conveniente pagar una membresía mensual o anual por una herramienta que no sabemos si nos dará los resultados que incrementen las utilidades del negocio, es importante saber que las probabilidades de ventas nunca son exactas, pero pagar membresías incrementan los gastos que no necesariamente se van a recuperar con su uso.

Por otro lado, el encargado de administrar el sitio web, con respecto a recomendar modelos y preferencias en calzado, artículos, etc.; los determinan en base a tendencias de

revistas internacionales, pero no se han enfocado en un estudio de los productos que hayan comprado, o visitas que hayan recibido en su web de los clientes en una temporada anterior, lo que le podría generar un ahorro de costes, de tiempos, y obtener mejores decisiones al saber la preferencia de los clientes.

El no saber qué ofertar, tener dudas qué productos escoger, trae consigo pérdidas tanto en tiempo como en ganancias, las recomendaciones de productos, ropa, calzado, etc.; existieron desde que se vendieron las primeras revistas como técnica para ventas personalizadas. Desde sugerencias de ingredientes para usar en una receta hasta los famosos artículos “Completa el look”, que se encuentran en las revistas de moda, pero clasificar y tenerlas toma mucho tiempo y es ambiguo, va de acuerdo al criterio del encargado de ventas.

Estos métodos antiguos, proporcionaban sugerencias subjetivas por la razón de no ser inteligentes. Para el comercio electrónico se necesita de una recomendación inteligente que esté impulsada por las matemáticas y ciencia de datos.

Los encargados de ventas, tiendas y sitios de comercio electrónico, utilizan el comportamiento anterior de un comprador o visitante para ofrecer recomendaciones de productos que podrían gustarle o del que en alguna anterior visita captó su atención de alguna forma. Algo así como si ya estuviera utilizando las preferencias de un usuario para enviar anuncios o correos electrónicos direccionados.

A pesar de saber lo fuerte que es el poder de la personalización, no todos los especialistas en ventas, marketing, webs master de comercio electrónico, etc.; aún no lo están haciendo correctamente y envían todas las recomendaciones de productos incorrectas, basándose únicamente en la intuición o moda.

De continuarse presentando esta problemática en Sport Anthony, se ve comprometida la rentabilidad económica que toda empresa debe tener para continuar su funcionamiento y cubrir ciertos gastos, también la ambigüedad en la decisión del encargado de ventas en qué productos ofertar, de cierta forma también se verán afectados los clientes que eligen comprar en esta tienda, y claro, la popularidad y la imagen de la misma.

## **1.2. Análisis del problema**

### ***1.2.1. Selección ambigua de los productos a ofertar***

Causas y consecuencias: Todos los productos que fabrica un proveedor, la información y la cantidad de productos que existen en la red y en el mercado es tanta, que la persona encargada de publicitar los productos los selecciona y clasifica a su propio criterio. La selección, búsqueda, clasificación de los productos a publicitar se ven afectados, también la utilidad y la eficiencia del trabajador.

### ***1.2.2. Remates de Productos***

Causas y consecuencias: Los productos que se ofertan y que no tienen mayor relevancia, con pocas o nulas salidas, pasan a rematarse ofertándose, buscando la recuperación de esos gastos, no siempre se logra recuperar lo invertido en los lotes invertidos en las campañas, de esta manera se ven afectadas las ganancias en la empresa.

### ***1.2.3. Tiempos extensos de productos a ofertar***

Causas y consecuencias: A parte de ser ambigua la selección, búsqueda, clasificación de los productos a publicitar, los tiempos para realizar estos procesos también se ven afectados, al realizarse al criterio del encargado del área, este proceso puede llevarse a cabo con extensos periodos de tiempo.



#### **1.2.4. Vulnerabilidad de la imagen y prestigio de la empresa**

Causas y consecuencias: Al no ofertarse productos aceptados por el público, puede verse afectada la imagen y el prestigio de la empresa, las especulaciones, cambio de gustos, hábitos o preferencias del público, traen consigo caídas en la demanda.

### **1.3. Antecedentes de la investigación**

#### **1.3.1. A nivel internacional**

##### **1.3.1.1. Tesis 01**

**1.3.1.1.1. Autores.** Coronel Romero Edison Leonardo y Victor Francisco Jumbo Sinchire

**1.3.1.1.2. Título.** SISTEMA RECOMENDADOR DE CUESTIONARIOS PARA APOYO AL APRENDIZAJE MEDIANTE BLOQUEO DE APLICACIONES EN LA PLATAFORMA ANDROID.

**1.3.1.1.3. Institución.** Universidad Nacional de Loja

**1.3.1.1.4. Año.** 2019

**1.3.1.1.5. Resumen.** Los recomendadores se abordan en la implementación de los algoritmos de descomposición en valores singulares (SVD) junto con técnicas de reducción de dimensionalidad y de cálculo de mínimos, para la implementación en los sistemas de recomendación en el comercio. (Coronel y Jumbo, 2019).

**1.3.1.1.6. Relación.** El utilizar SVD (Descomposición de valores singulares) hace posible lograr recomendaciones, basándonos en el REM y RMSE, nos lleva a obtener información más relevante y precisa, y en la presente investigación se está usando SVD para recomendación en el ámbito comercial, afirmando el supuesto del uso de la técnica para estos fines.

### **1.3.1.2. Tesis 02**

**1.3.1.2.1. Autor.** Jinneth Tique Ortiz

**1.3.1.2.2. Título.** Diseño de Datamart en las Temáticas de Producción e Inversión de Ciencia Tecnología e Innovación en Colombia.

**1.3.1.2.3. Institución.** Escuela Colombiana de Ingeniería Julio Garavito

**1.3.1.2.4. Grado.** Maestro en Ingeniería de Computación y Sistemas

**1.3.1.2.5. Año.** 2016

**1.3.1.2.6. Resumen.** Si la información estuviera disponible desde las entidades territoriales, las decisiones se tomarían informadas y con datos reales; y estas mejorarían. También menciona que la información no solo debe ser integrada, también debe ser íntegra y de calidad. Para ello es necesario que se ajuste, se depure y verifique la información que ha sido capturada por los sistemas (Tique, 2016).

**1.3.1.2.7. Relación.** En la presente investigación, al hacer uso del ETL, la información integrada obtenida ya es valiosa, y al aplicarle SVD y obtener recomendaciones logra ser de gran calidad también.

### **1.3.1.3. Tesis 03**

**1.3.1.3.1. Autor.** Camilo Antonio Ramírez Morales

**1.3.1.3.2. Título.** Algoritmo SVD Aplicado a los Sistemas de Recomendación en el comercio.

**1.3.1.3.3. Institución .** Universidad distrital Francisco José de Caldas – Colombia

**1.3.1.3.4. Grado.** Ingeniero en telemática y tecnólogo en sistematización de datos.

**1.3.1.3.5. Año.** 2018

**1.3.1.3.6. Resumen.** Que hacer uso de SVD en los aplicativos de recomendación del filtrado colaborativo, logra excelentes recomendaciones y de calidad, pero en cuanto a los tiempos de los cálculos de las predicciones es una técnica poco atractiva en los entornos de comercio, es

gracias a eso que se tiene una tarea pendiente para optimizar, los tiempos de ejecución en la etapa de predicciones. (Ramírez, 2018).

**1.3.1.3.7. Relación.** Si bien es cierto en la actualidad existen apps o plugins que se pueden añadir a sitios webs ecommerce, estos pueden ser muy costosos, con SVD podemos lograr recomendaciones con una cierta cantidad de código y sin hacer pagos mensuales de membresías.

## **1.3.2. A Nivel Nacional**

### **1.3.2.1. Tesis 04**

**1.3.2.1.1. Autores.** Sara Landa Tinco y Fernando Jaime Ticona Pacheco

**1.3.2.1.2. Título.** Un Modelo de Recomendación de Productos utilizando Redes Sociales implementado en Tecnologías de la Web Semántica.

**1.3.2.1.3. Institución.** Universidad Nacional Mayor De San Marcos

**1.3.2.1.4. Grado.** Ingeniero de Sistemas

**1.3.2.1.5. Año.** 2015

**1.3.2.1.6. Resumen.** Se puede aplicar el modelo personalizando y recomendando servicios a organizaciones o instituciones que ofrezcan dichos servicios, como por ejemplo dependiendo de las necesidades del cliente, en organizaciones privadas recomendando sus servicios; o en hospitales, programas sociales u otras instituciones públicas se calificaría a los ciudadanos, todo de manera automática con sus historiales, o también en lugares con mayor o menor demanda de servicios sociales para mejorar la toma de decisiones y los procesos distribuyendo los servicios correctamente. (Landa y Ticona, 2015).

**1.3.2.1.7. Relación.** Las recomendaciones en esta investigación son para el personal que se encarga de los productos, los procesos se mejoran en cuánto a rapidez, en disminución de costos, la toma de decisiones se agiliza y se toman en base a los datos.

### **1.3.2.2. Tesis 05**

**1.3.2.2.1. Autores.** Arenas Da Cruz Elfer Joann y Tapia Shuña Paulo Cesar

**1.3.2.2.2. Título.** Un Datamart Con Tecnología de Inteligencia de Negocios en la Optimización de la Toma de Decisiones en el Área de Logística Comercial de un Grupo Comercial.

**1.3.2.2.3. Institución.** Universidad Nacional de Ucayali

**1.3.2.2.4. Grado.** Ingeniero de Sistemas

**1.3.2.2.5. Año.** 2018

**1.3.2.2.6. Resumen.** Que en los procesos del ETL y también la purificación de los datos y en su análisis, se logran automatizar, mejorando notablemente así el tiempo para crear un conocimiento de excelente utilidad en la toma de decisiones y logrando reportes con información valiosa (Arenas y Tapia, 2018).

**1.3.2.2.7. Relación.** Es muy valiosa la información que se obtiene en un ETL, pero lograr predicciones recomendando, es información aún más valiosa, de eso se encarga SVD en Machine Learning, la facilidad en la toma de decisiones con esta tecnología se hace aún mejor.

### **1.3.2.3. Tesis 06**

**1.3.2.3.1. Autor.** Alain Yucra Halanoca

**1.3.2.3.2. Título.** Datamart de la Gerencia de Supermercado Mia Market de Juliaca 2018

**1.3.2.3.3. Institución.** Universidad Nacional del Altiplano

**1.3.2.3.4. Grado.** Ingeniero Estadístico e Informático

**1.3.2.3.5. Año.** 2018

**1.3.2.3.6. Resumen.** Que es una gran ventaja el uso del Datamart en datos multidimensionales de Analysis Services conjuntamente con Excel, Reporting Services y

Performance Point, que son herramientas utilizadas habitualmente en informes BI, logrando respuestas rápidas, desplegando la información con mayor facilidad en el menor tiempo posible y logrando contribuir a la toma de decisiones (Yucra, 2018).

### **1.3.3. A Nivel Local**

#### **1.3.1.1. Tesis 07**

**1.3.1.1.1. Autores.** Gianira Xiomara Espinoza Airac y Eduar Fabián León Muñoz.

**1.3.1.1.2. Título.** Modelo de Machine Learning para la Clasificación de Estudiantes de acuerdo a su Rendimiento Académico en el Centro de Idiomas de la Universidad Nacional Del Santa.

**1.3.1.1.3. Institución.** UNS

**1.3.1.1.4. Grado.** Ingeniero de Sistemas e Informática

**1.3.1.1.5. Año.** 2020

**1.3.1.1.6. Resumen.** Usando modelos Machine Learning se logra obtener predicciones, basándonos en registros de datos históricos, de esta manera desarrollar soluciones que se involucran en la toma de decisiones, en este caso para CEIDUNS, les sirve como herramienta en el proceso de clasificación de sus estudiantes, esta herramienta de conocimiento es según al rendimiento académico. (Espinoza y León, 2020).

**1.3.1.1.7. Relación.** Con los modelos y algoritmos Machine Learning se logran tener predicciones, recomendaciones, etc. En la presente investigación de igual manera que (León, 2020), con datos y registros históricos, en el caso de nuestra investigación los datos históricos se basan en los registros de las visitas a la web, mientras que en la investigación de referencia usan el rendimiento académico.

#### **1.3.1.2. Tesis 08**

**1.3.1.2.1. Autor.** Zegarra Mezarina Alexander Junior

**1.3.1.2.2. Título.** Implementación de un Data Mart usando la Metodología Ralph Kimball para el área de Logística de la Empresa Comercializadora Easy Work E.I.R.L - Casma; 2017

**1.3.1.2.3. Institución.** ULADECH

**1.3.1.2.4. Grado.** Ingeniero de Sistemas e Informática

**1.3.1.2.5. Año.** 2017

**1.3.1.2.6. Resumen.** La empresa Easy Work tuvo la necesidad de implementar un DataMart que ayudó en lo que respecta a gestionar mejor la información, de esta forma mejorar en la toma de decisiones, lográndose la disminución considerable de muchos procesos innecesarios. (Zegarra, 2017).

**1.3.1.2.7. Relación.** En la presente investigación aplicamos Datamart y su proceso ETL para posteriormente aplicar Machine Learning con el Dataset obtenido, y al igual que Zegarra (2017), se busca mejorar la manera de gestionar los datos y progresar en las decisiones que se puedan tomar.

### **1.3 Formulación del problema de investigación**

¿De qué manera la implementación de un Algoritmo Recomendador Machine Learning usando Datamart mejorará la toma de decisiones con recomendaciones a ofertar en el área de ventas de la empresa Sport Anthony en Chimbote?

### **1.4 Delimitación del estudio**

La investigación está enfocada a la implementación de un algoritmo recomendador Machine Learning usando DATAMART para la toma de decisiones en el área de ventas de la empresa Sport Anthony.

## **1.5 Justificación de la investigación**

### ***1.5.1. Justificación teórica.***

Se pretende que aplicando la teoría y práctica de los términos Business Intelligence y Machine Learning (DATAMART, ALGORITMO RECOMENDADOR), respectivamente, explotar la información, con la intención tomar mejores decisiones, mediante la manipulación de forma sencilla, planteando escenarios futuros.

De esa forma luego de la implementación, la información va a estar mejor fundamentada y disponible a mejores tomas de decisiones, habrá menos posibilidades de equivocación, y una idea más clara a futuro, lo cual es beneficioso para la organización, y es motivo para justificar totalmente a la elaboración de este proyecto.

### ***1.5.2. Justificación metodológica.***

En la construcción del Datamart de Sport Anthony, se usará la metodología de Ralph Kimball, mientras que para Machine Learning se usará la Metodología Crisp DM, que es más enfocada a la ciencia de datos. Se justifica porque la investigación se especifica en un departamento de la compañía, en comparación a otras metodologías, con Kimball el coste inicial es menor, se requiere menos tiempo de desarrollo, y siendo específicamente de un área, logra de forma creciente un enfoque más versátil, dejando pase libre a una ampliación más adelante, en cualquier momento del tiempo, asegurando así la usabilidad del sistema por parte de los usuarios de manera sencilla, fácil de entender y rápido de desarrollar.

### ***1.5.3. Justificación tecnológica.***

Tecnológicamente es justificable al estar a la vanguardia de las nuevas tendencias tecnológicas, como aplicar soluciones de Business Intelligence o Machine Learning, un

DATAMART y un algoritmo recomendador respectivamente, al manejo y control de la información para bien de la empresa.

#### **1.5.4. *Justificación social***

Mejorar la imagen de la organización y su popularidad al promocionar productos de mejor alcance, que vayan más acorde a lo que requiere y gusta la sociedad.

#### **1.5.5. *Justificación operativa***

Tratar y manipular la data e información eficazmente y reducir los posibles problemas, bugs o errores al ejecutar los task o tareas para las recomendaciones.

#### **1.5.6. *Justificación técnica.***

Se justifica técnicamente al reducir los tiempos de búsqueda de productos a ofertar, también evitando la pérdida de información valiosa utilizable y teniendo la Disponibilidad inmediata de datos informativos.

### **1.6. Limitaciones**

- No se cuenta con partida presupuestal para la elaboración de una solución BI, por lo que no se puede contratar a terceros, tampoco comprar una solución a medida, ni compra de licencias ni hardware.
- Limitaciones en disponibilidad de tiempo por parte del trabajador de la empresa para obtener información, entrevistas, consultas.



## **1.7. Objetivos de la investigación**

### **1.7.1. *Objetivo general.***

Mejorar la toma de decisiones en el área de ventas de la empresa Sport Anthony en Chimbote con recomendaciones a ofertar a través de un Algoritmo Recomendador Machine Learning y Datamart

### **1.7.2 *Objetivos específicos.***

1.7.2.1. Extender el grado de satisfacción de los usuarios; personal administrativo, de ventas y marketing.

1.7.2.2. Obtener recomendaciones de calidad de productos a ofertar con el algoritmo Machine Learning.

1.7.2.3. Disminuir los tiempos de búsqueda de productos a ofertar.

1.7.2.4. Costos por el uso del algoritmo recomendador Machine Learning

1.7.2.5. Disminuir los tiempos para el análisis de la información y obtener alternativas gracias al uso del Algoritmo en Python con el Dataset resultante obtenido.

## **CAPÍTULO II**

### **MARCO TEÓRICO**

## 2. Marco Teórico

### 2.1. Fundamentos

#### 2.1.1. Toma de Decisiones.

Gunther (2008) nos dice que es importante conocer la cantidad de data que se necesita y cuánto tiempo podemos esperar hasta tomar acción. Se trata de saber lo suficiente para poder dar el paso, más no de ser perfecto, asea, tratemos siempre tener todos los factores necesarios para el éxito, en consideración, tratar de ejecutar control sobre aquellos que se escapen y saber cuáles tenemos bajo.

#### 2.1.2. Machine Learning.

Mitchell (2018) define al aprendizaje automático o en inglés Machine Learning, como el estudio de procedimientos, métodos o algoritmos computacionales que mejoran y evolucionan automáticamente su rendimiento debido a la experiencia. Si el desempeño de las tareas de un sistema o aplicativo informático mejora en base a un conjunto de tareas debido a la experiencia y haciendo uso de una medida de rendimiento, quiere decir que se está haciendo uso de Machine Learning.

#### 2.1.3. SVD.

Moya (2013) nos dice que la Descomposición en Valores Singulares (SVD). Usa una técnica de factorización de matrices que descompone tres matrices U, S, y V a una Matriz A, de la siguiente manera:

$$\text{SVD}(A) = U \times S \times V^t$$

Es decir, que del producto matricial de la matriz  $U \times S \times V^t$ , se obtiene la matriz inicial A.

SVD permite realizar unas recomendaciones muy buenas a los usuarios del sistema de recomendación, en este caso a los encargados de Ventas y/o Marketing de la empresa Sport Anthony.

La prueba de que el método del SVD es uno de los mejores métodos para recomendar, está en que la aplicación de esta técnica logró ganar el premio Netflix.

**Tabla N° 1. SVD frente a otros Algoritmos Recomendadores**

N°	Movielens 100k	RMSE	MAE	Tiempo
1	SVD	0.934	0.737	0:00:11
2	SVD++	0.92	0.722	0:09:03
3	NMF	0.963	0.758	0:00:15
4	Slope One	0.946	0.743	0:00:08
5	K-NN	0.98	0.774	0:00:10
6	Slope One K-NN	0.926	0.7130	0:00:08
7	Centered K-NN	0.951	0.749	0:00:10
8	K-NN Baseline	0.931	0.733	0:00:12
9	Co-Clustering	0.963	0.753	0:00:03
10	Baseline	0.944	0.748	0:00:01
11	Random	1.514	1.215	0:00:01

#### **2.1.4. Anaconda Navigator.**

Toro (2017) define que es una GUI de escritorio utilizable en Windows, que, sin hacer uso de comandos en Linux, logra inicializar aplicaciones y administrar fácilmente paquetes, entornos y canales de Conda. Forma parte de los 4 sectores que agrupa la Distribución de Anaconda, la que engloba distintas aplicaciones, librerías y conceptos diseñados para el desarrollo de la Ciencia de datos con Python.

### ***2.1.5. Jupyter Notebook.***

Martin (2018) dice que es un entorno de trabajo interactivo que de dinámicamente hace posible introducir código en Python, integra también bloques de texto, gráficas, imágenes. Es una herramienta bastante poderosa que se utiliza mucho en la estadística, análisis numérico, Machine Learning, entre otros campos de la informática y las matemáticas, que combina código ejecutable, representaciones gráficas de su ejecución y textos planos muy sencillos de utilizar.

### ***2.1.6. Python.***

Abellán (2019) nos dice que es un lenguaje de programación multiparadigma que se encuentra disponible en diversas plataformas, también nos dice que Python contiene diversas funciones, tipos de datos y una gran cantidad de librerías incorporadas en el propio lenguaje, que, sin la necesidad de programarlas desde cero, nos ayudan a realizar muchas tareas comunes.

### ***2.1.7. Inteligencia de Negocios (BI)***

Muñiz (2018) comenta que con la Inteligencia de negocios se podrían obtener pronósticos, informes o escenarios que contribuyan a la toma de decisiones, que a lo que nos lleva, es a una ventaja super competitiva. La clave de la Inteligencia de negocios es la cantidad y calidad de data que obtiene y uno de sus mayores sino principal beneficio, es de ser utilizada para la toma de decisiones. Hace más sencilla la distribución de la información a los distintos niveles de la organización y le facilita al usuario la información que necesita según el momento. Para su diseño e implantación, un sistema de Inteligencia de negocios requiere llevar a cabo con anticipación una evaluación de las necesidades de información y de las data bases que existen.

### **2.1.8. Ciencia de Datos**

La ciencia de datos impacta enormemente en la manera en que las organizaciones hacen negocio, también señala que la ciencia de datos ayuda a tomar mejores decisiones, conocer qué productos y servicios lanzar y cómo brindar un mejor servicio a los clientes, todo esto se da recopilando y clasificando grandes cantidades de información.

### **2.1.9. Metodología CRISP-DM**

(Cross industry standard process for data mining).

La metodología CRISP-DM consiste en un conjunto de tareas descritas en cada una de sus 6 fases; comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación, e implementación, y que estas están interrelacionadas y directa o bidireccionalmente.

### **2.1.10. ETL**

ETL, cuyas siglas hacen alusión a sus tres pasos principales (extraer, transformar y cargar), es un método para la integración de datos, se utiliza para transformar en un formato homogéneo y consolidado, los datos provenientes de distintas fuentes: CRM, ERP, tablas de Excel y otras fuentes tradicionales de archivos. Lo utilizamos para construir un Data warehouse o almacén de datos; en combinación con otros softwares que nos mejoran el rendimiento de la gestión de datos y nos facilitarán su procesamiento.

### **2.1.11. DATAMART**

Simon (1998) nos dice en su libro que es una base de datos de un departamento del negocio que se especializa en el almacenamiento de la data de igual forma específicamente en un área de negocio, en nuestro caso tiene que ver con el área de ventas. Nos menciona que un

datamart se alimenta de varias fuentes de información, en nuestro caso desde archivos exportados desde Google Analytics y el sitio web ecommerce desarrollado.

#### ***2.1.12. Proceso Analítico en Línea (OLAP).***

Rouse (2015) nos dice que con OLAP se extrae de manera fácil y selectiva los datos y se logra verlos de distintos puntos de vista, También que una base de datos OLAP no necesariamente debe ser inmensa como un data warehouse, puesto que no todos los datos transaccionales son necesarios para analizar tendencias, en nuestro caso necesitaremos ciertas dimensiones como región, productos, tiempo, visitas.

#### ***2.1.13. SQL SERVER***

Parada (2019) nos menciona que es un sistema para gestionar las bases de datos relacionales, el cuál ha sido desarrollado como un servidor que da servicio a otras aplicaciones de software para que funcionen, bien en el mismo ordenador o en otro mediante una red, incluyendo Internet.

**CAPÍTULO III**  
**DESARROLLO DE LA METODOLOGÍA**



### **3. Desarrollo de la metodología**

#### **3.1. Fase de comprensión del negocio**

##### ***3.1.1. Determinar el objetivo del negocio***

El objetivo del estudio y la aplicación de CRISP-DM en el negocio es lograr obtener recomendaciones para el encargado del área de ventas al determinar con mayor rapidez y tener una idea más clara de qué productos ofertar, logrando de esta manera un mejor servicio y captar más posibles clientes que realicen una compra en Sport Anthony.

##### **3.1.1.1. Contexto**

Con respecto a la situación de Sport Anthony, la tienda cuenta con data de clientes que han comprado en la tienda, los productos que se han vendido, los productos visitados en su página web, etc. Pero no existe algún estudio en cuanto al comportamiento de estos datos, de los cuáles se pueden sacar conclusiones o comportamientos para obtener predicciones que podrían ser valiosas para el encargado de las ventas.

##### **3.1.1.2. Objetivos del negocio**

Los objetivos de la tienda, como toda empresa que se dedica a las ventas, es incrementarlas; para esto hay muchos factores que influyen y la toma de decisiones son parte fundamental para lograr este objetivo, lo que se plantea es influenciar positivamente en esta toma de decisiones, el tener una idea clara de qué productos ofertar basándonos en datos históricos, hace más fiable un plan de ventas, el encargado de las ventas tendrá un respaldo de qué productos ofertar gracias a la aplicación del algoritmo machine learning. Por otro lado, el realizar pruebas de testeo para determinar

qué productos son “Ganadores” tiene consigo gastos, los mismos que se pueden evitar conociendo a qué productos hacerles publicidad.

### **3.1.1.3. Criterios de éxito del negocio**

Desde un punto de vista comercial, se tiene como criterio de éxito la obtención de información relevante para el encargado de ventas a partir del algoritmo recomendador, identificando productos a ofertar en el menor tiempo y disminuyendo gastos en pruebas de testeo.

### **3.1.2. Valorar la situación**

Se cuenta con una base de datos con información de los productos, los clientes, etc.; también los datos que se exportan de Google Analytics integrado a su web ecommerce.

#### **3.1.2.1. Inventario de Recursos**

- Sobre los recursos de software; se tiene a disposición programas como Visual Studio 2019 con SSDT (SQL Server Data Tools), SQL Server 2014, para python, Jupyter Notebook en Anaconda Navigator, office 2016 y Windows 10.
- En cuanto a recursos hardware:
  - Procesador: Intel(R) Core (TM) i7-7700 CPU @3.60 GHz
  - Memoria RAM: 16 GB
  - Sistema Operativo: Windows 10
  - Tarjeta gráfica: Intel(R) HD Graphics 630
  - Tarjeta gráfica2: NVIDIA GeForce GT 730
  - Capacidad de almacenamiento: 2TB

### 3.1.2.2. Requisitos, supuestos y restricciones

El requisito indispensable es contar con la data y la información o tener acceso a las fuentes de datos que se necesitan en el procesamiento de los mismos.

Entre los supuestos, se cuenta con la data e información suficiente de la empresa Sport Anthony y la disponibilidad del encargado de ventas para remitirla en cuanto se necesite.

Entre las restricciones son el tiempo de disponibilidad del encargado de ventas para acceder a los datos o para hacer pruebas.

### 3.1.2.3. Riesgos

La disponibilidad el encargado de ventas para atención es muy importante ya que es posible que se necesite más datos, mientras más cantidad de datos se dispongan, mayor cantidad de mediciones pueden realizarse.

**Tabla N° 2: Tabla de Terminología.**

<b>CÓDIGO</b>	<b>TÉRMINO</b>
ML	Machine Learning
CRISP	Cross Industry-Standard Process For Data Mining
SSDT	SQL Server Data Tools

**Tabla N° 3: Tabla de Costes - Hardware**

<b>HARDWARE</b>			
<b>Descripción</b>	<b>Cantidad</b>	<b>Precio</b>	<b>Total</b>
Laptop	1	2500.00	2500.00
Impresora	1	300.00	300.00
<b>TOTAL</b>			<b>2800.00</b>

**Tabla N° 4: Tabla de Costes – Software**

<b>SOFTWARE</b>			
<b>Descripción</b>	<b>Cantidad</b>	<b>Precio</b>	<b>Total</b>
Windows 10 Home	1	658.00	658.00
Office	1	219.00	219.00
<b>TOTAL</b>			<b>877.00</b>

**Tabla N° 5: Tabla de Costes – Internet**

<b>INTERNET</b>			
<b>Descripción</b>	<b>Precio (Mes)</b>	<b>Meses</b>	<b>Total</b>
Windows 10 Home	147.90	4	591,6
<b>TOTAL</b>			<b>591,6</b>

**Tabla N° 6: Tabla de Costes – Personas**

<b>Personas</b>			
<b>Personal</b>	<b>Cantidad meses</b>	<b>Precio</b>	<b>Total</b>
Desarrollador	4	1200.00	4.00
<b>TOTAL</b>			<b>4800.00</b>

**Tabla N° 7: Tabla de Costes – Materiales**

<b>MATERIALES</b>			
<b>Descripción</b>	<b>Cantidad</b>	<b>Precio</b>	<b>Total</b>
Lapiceros	2	0.50	1.00
Folder Manila	2	1.00	2.00
<b>TOTAL</b>			<b>3.00</b>

### 3.1.3. Realizar un plan de proyecto

**Tabla N° 8: Conformación de equipo y usuarios**

N°	MIEMBRO
01	Luis Ricardo Marcelo Gómez • Analista / Programador • Tester
02	Representante de Sport Anthony • Erika Mechán Chavesta
03	• Erika Mechán Chavesta (Adm. / Marketing) • Joselyn Vásquez (Ventas) • Juan Vega (Ventas) • Shirley Castillo (Ventas) • Grace Padilla (Ventas)

**Tabla N° 9: Fases**

Fase	Periodo estimado	Riesgo
Comprensión del negocio	2 semanas	Bajo
Comprensión de los datos	4 semanas	Medio
Preparación de los datos	5 semanas	Alto
Modelado	4 semanas	Alto
Evaluación	2 semanas	Medio
Distribución	1 semana	Medio

### 3.2. Etapa de comprensión de datos

Esta es la etapa número dos de la metodología CRISP-DM, en la que comenzamos a familiarizarnos con los datos y con el problema descrito en la primera fase, se realiza la primera recolección de datos, averiguamos la calidad y sus relaciones más evidentes para ir teniendo la idea clara de la hipótesis.

### ***3.2.1. Recolección de datos iniciales***

Con respecto al origen de los datos, los cuales son de Sport AnthonyDB y de archivos en Excel o de formato csv, provenientes de Google Analytics, tecnologías que se usan en la web de ventas, y de su sistema de escritorio, estos datos son extraídos, transformados y cargados con Visual Studio y sus herramientas de Business intelligence y luego desde allí almacenados a SQL Server, proceso que se denomina ETL.

El proceso se ha realizado extrayendo datos de unas fuentes origen, en este caso de las tecnologías que usa la empresa para con su sitio web, Google Analytics, SportAnthony.com, (WooCommerce), sportanthony.tk, y de los sistemas de escritorio que exportan archivos Excel, de igual manera, luego se procede a transformar esta data y almacenarlos en un destino. Al transformar estos datos, cumplen ciertas reglas, medidas que aplican al destino para realizar un análisis de información de data histórica obtenida en el proceso de migración, que serán indispensables para implementar el algoritmo de recomendación.

### ***3.2.2. Descripción de los datos***

En el presente estudio se utiliza y procesa data dentro del periodo de enero hasta fines de Julio del 2021, contando con datos de la tienda registrados precisos para este, también con data como visitas a un sitio web, las páginas, los productos que se ofrecen, el tiempo que se permanece en las páginas, las regiones del Perú etc., datos que son muy valiosos a la hora de querer sacar conclusiones.

### ***3.2.3. Verificación de la calidad de los datos***

Los datos no contienen errores, valores nulos, ni duplicados, y en el peor de los casos con Python convertiremos los nulos en ceros.

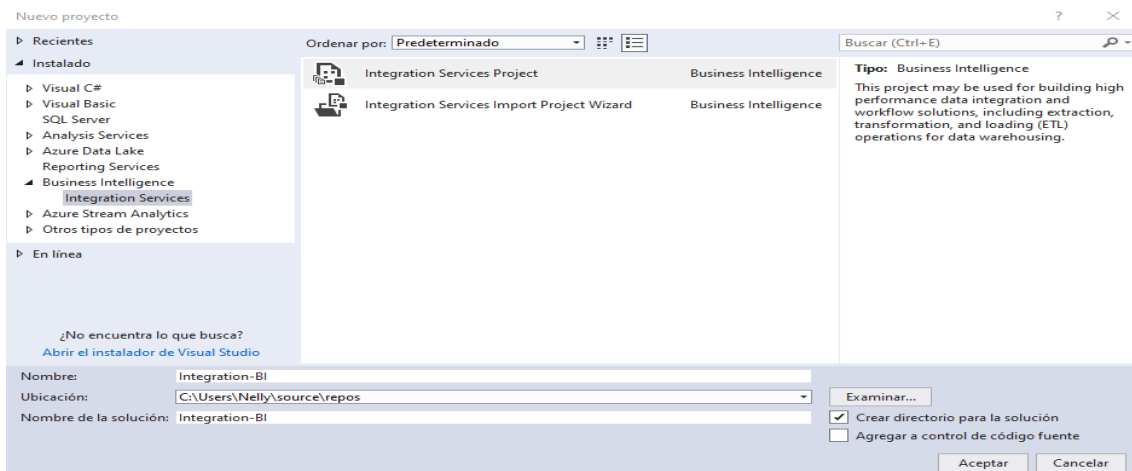
**Tabla N° 10: Calidad de datos, Código – Término**

CÓDIGO	TÉRMINO
CLD – 01	Los campos de la tabla no contienen nulos.
CLD – 02	Los campos no deben tener caracteres extraños o información no válida.
CLD - 03	La información o data no debe estar duplicada.

### 3.3. Fase de preparación de los datos

#### 3.3.1. Selección de datos - Implementación: Proceso ETL

El proceso se ha realizado extrayendo data de unas fuentes origen, en este caso de las tecnologías que usa la empresa para con su sitio web, Google Analytics, SportAnthony.com. (WooCommerce), sportanthony.tk, y de los sistemas de escritorio que exportan archivos Excel, de igual manera luego se procede a transformar esta data y almacenarla en un destino. Al transformar estos datos, cumplen ciertas reglas, medidas que aplican al destino para realizar un análisis de información de datos históricos obtenidos en el proceso de migración, que serán indispensables para implementar el algoritmo de recomendación. Para el ETL aplicado, se realizó usando SQL Server 2017 Integration Services (SSIS) y Visual Studio 2017 (SSDT). Abrimos (SSDT), de Visual Studio 2017, para este proyecto, y creamos uno nuevo de Inteligencia de Negocios tipo Servicios de Integración. El que nombraremos como “Integration-BI”



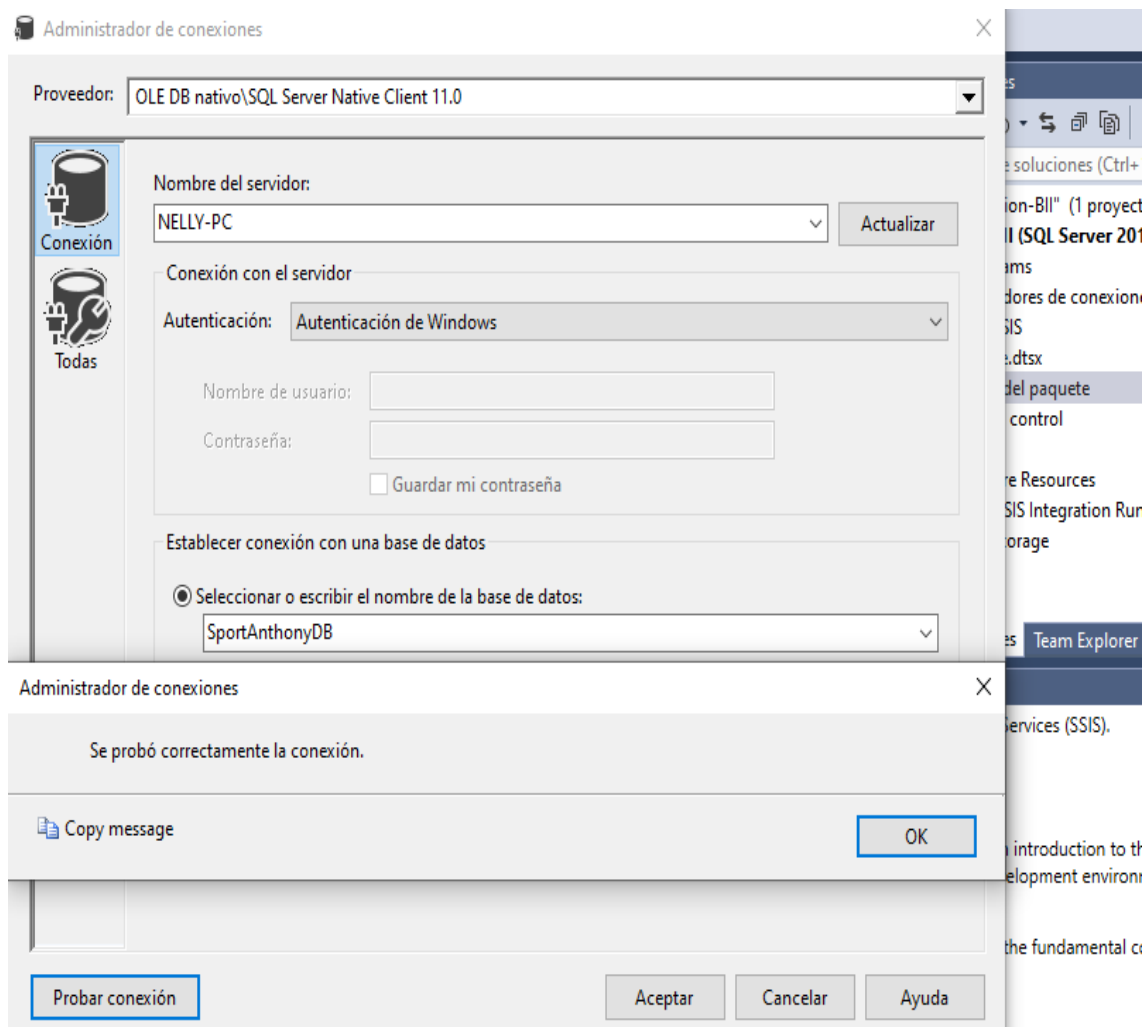
**Figura N° 1 : Nuevo proyecto - Business Intelligence.**

### 3.3.2. Limpieza de datos - Implementación: Proceso ETL

A continuación, configuramos las conexiones que tendremos como fuentes de Origen y Destino. En este caso, la fuente de Origen será la base de datos y la fuente Destino será el Data Mart. Se debe colocar el servidor, en este caso, uno local.

A continuación, preparar la conexión Origen indicando el nombre de la base de datos y probar su conexión haciendo clic en el botón “Probar Conexión”

Nos muestra un mensaje afirmando la correcta conexión. Por lo tanto, ya está configurada la fuente de base de datos Origen.

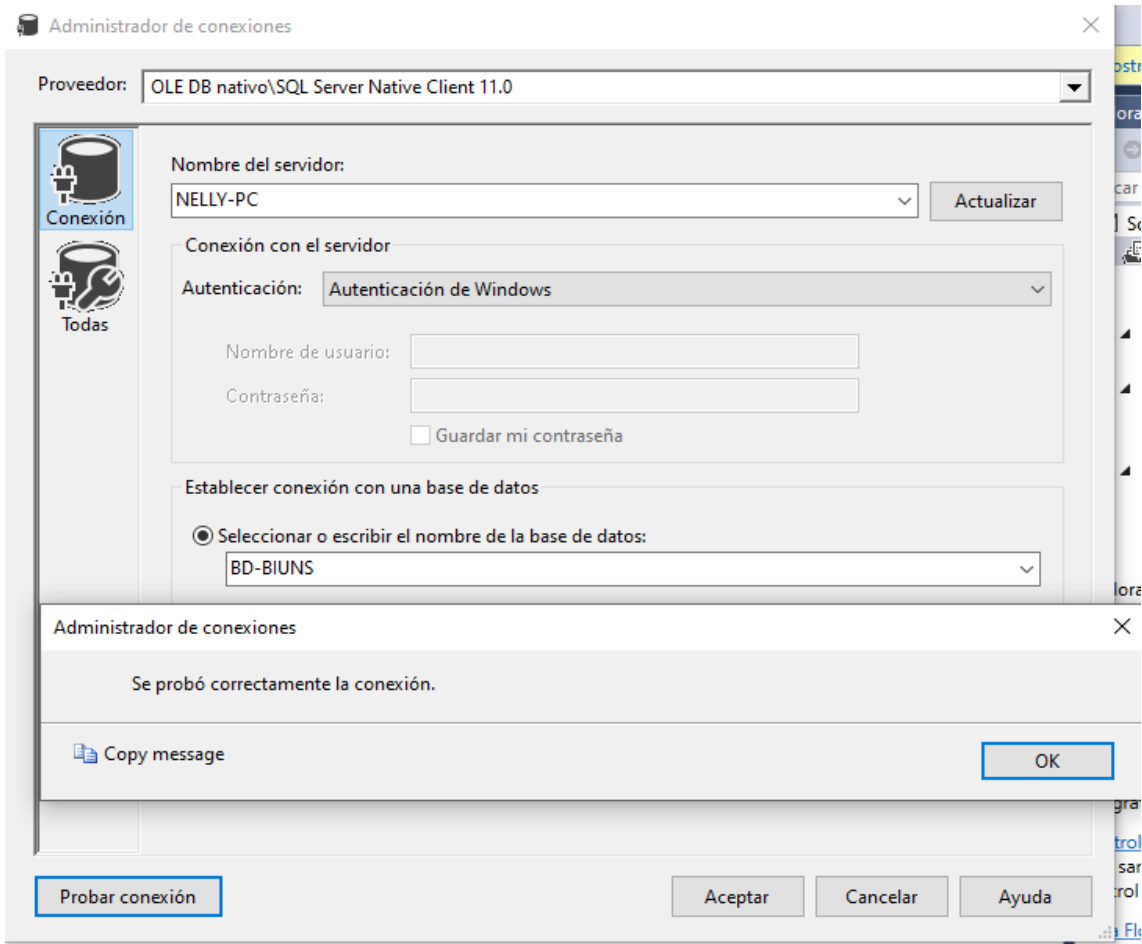


**Figura N° 2: Origen de datos – Prueba de conexión**



Lo siguiente es preparar la conexión de Destino, lo que vendría a ser nuestro Data Mart, y de la misma manera, probamos la conexión, clicando en “Probar conexión” y obtenemos que la conexión es correcta.

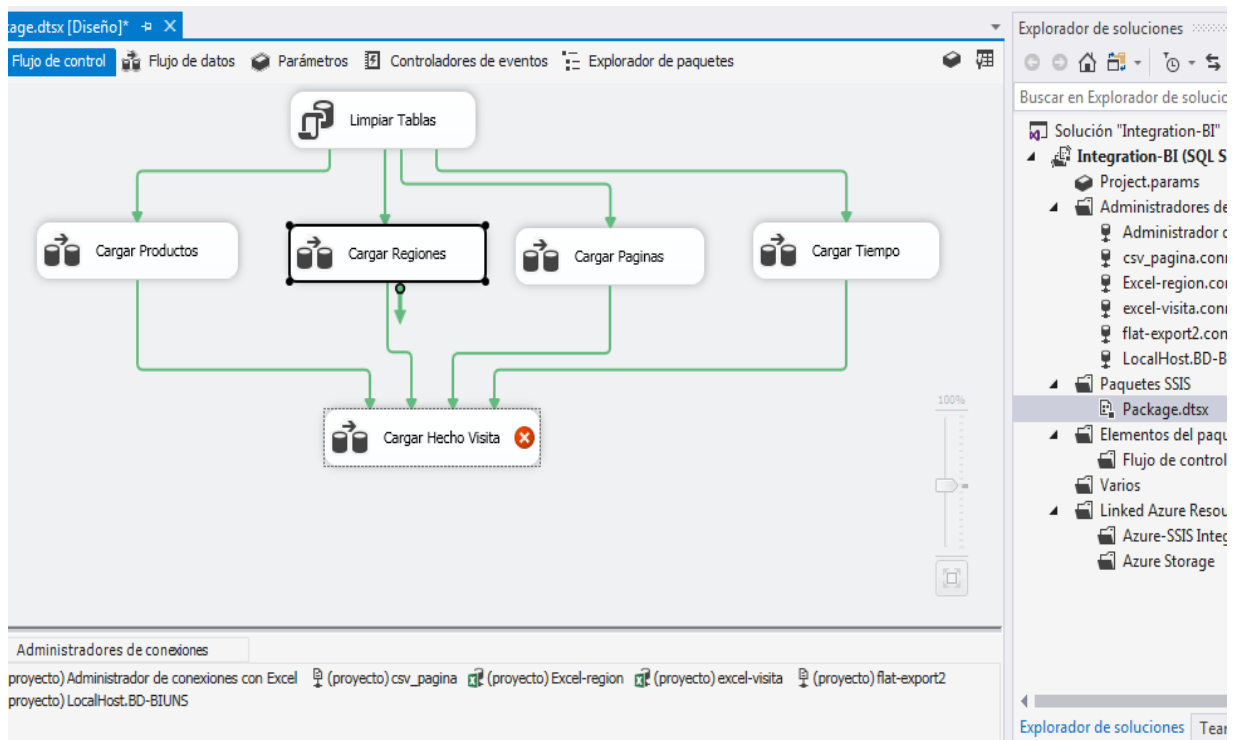
Al culminar este paso, ya están las conexiones configuradas para poder realizar la integración y mediante un proceso ETL cargar datos al Data Mart.



**Figura N° 3: Destino de datos – Prueba de conexión**

Lo que continúa es poner Tareas SQL o Data Flow Task (con tareas internas).

Se tuvo en cuenta la estructura para el ETL de la siguiente manera:



**Figura N° 4: Data Mart de ventas - ETL**

Una vez conectadas las tasks a ejecutar, dependiendo del orden indicado, se cargará en cada proceso las siguientes consultas de base de datos, partiendo de la fuente de Origen.

### 3.3.3. Limpieza de datos - tablas

#### Limpiar Tablas

```
Conjunto de resultados
Escribir consulta SQL
delete from Hech_Visita
delete from DimProducto
delete from DimRegion
delete from DimPagina
delete from DimTiempo
```

**Figura N° 5: Script Limpiar tablas.**

### 3.3.4. Estructuración, Integración y formateo de los datos

#### 3.3.4.1. Dimensión Producto - Cargar Productos

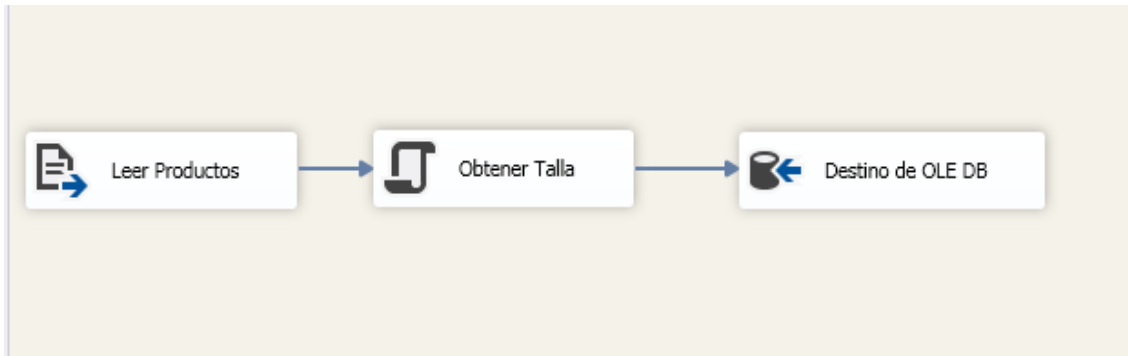


Figura N° 6: ETL Cargar Productos

##### 3.3.4.1.1. Leer Productos

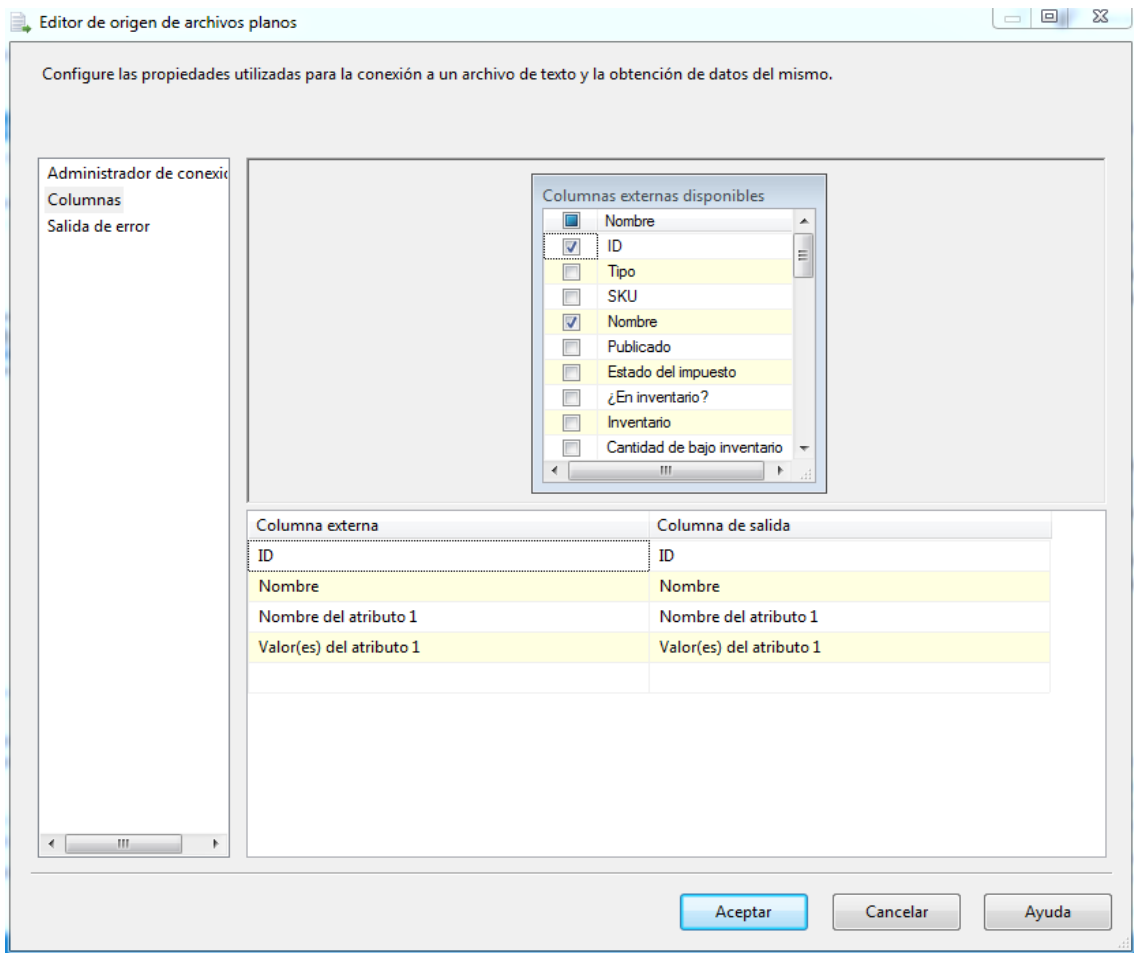


Figura N° 7: ETL Cargar Productos – Leer Productos

### 3.3.4.1.2. Obtener Talla

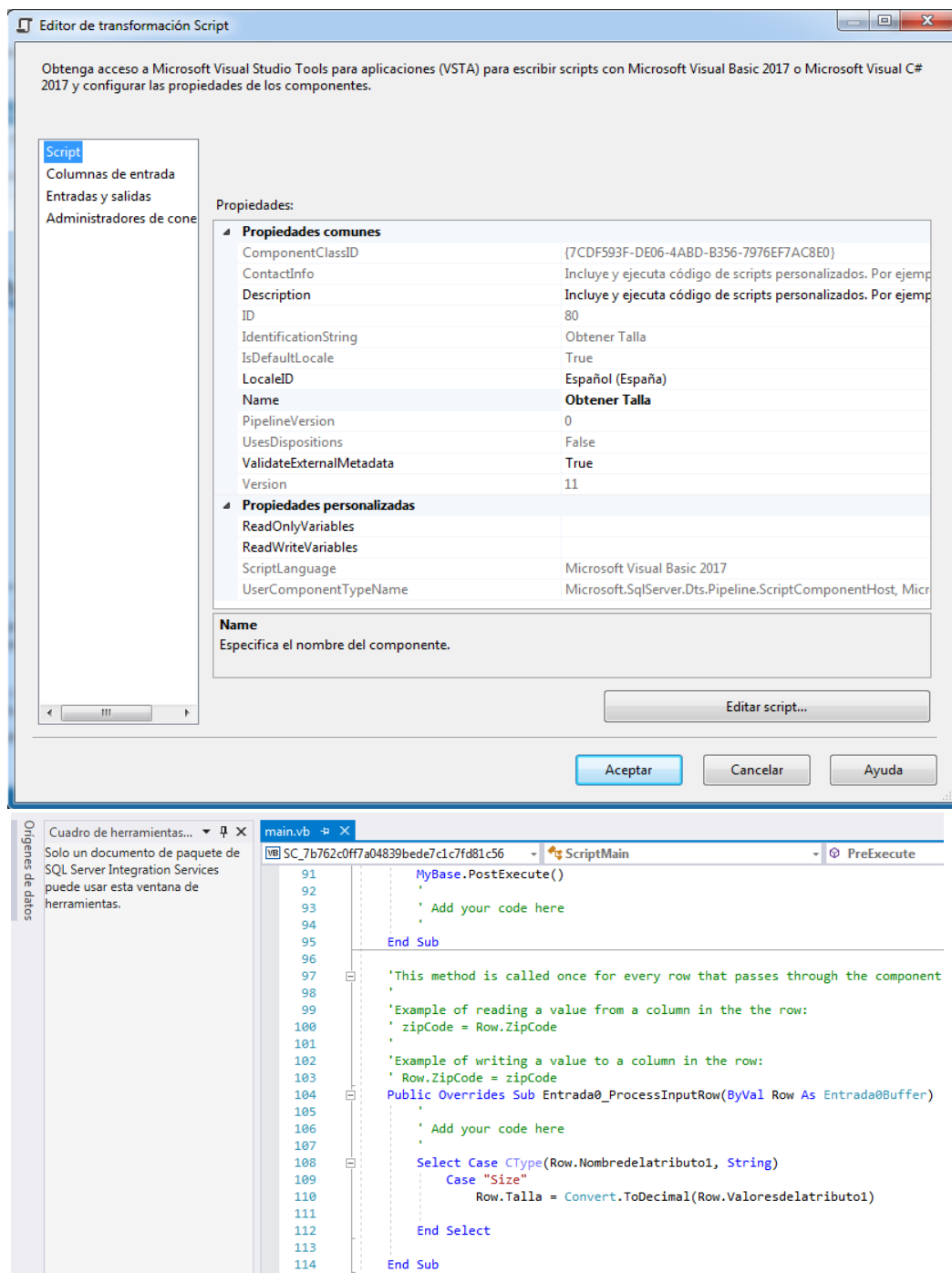
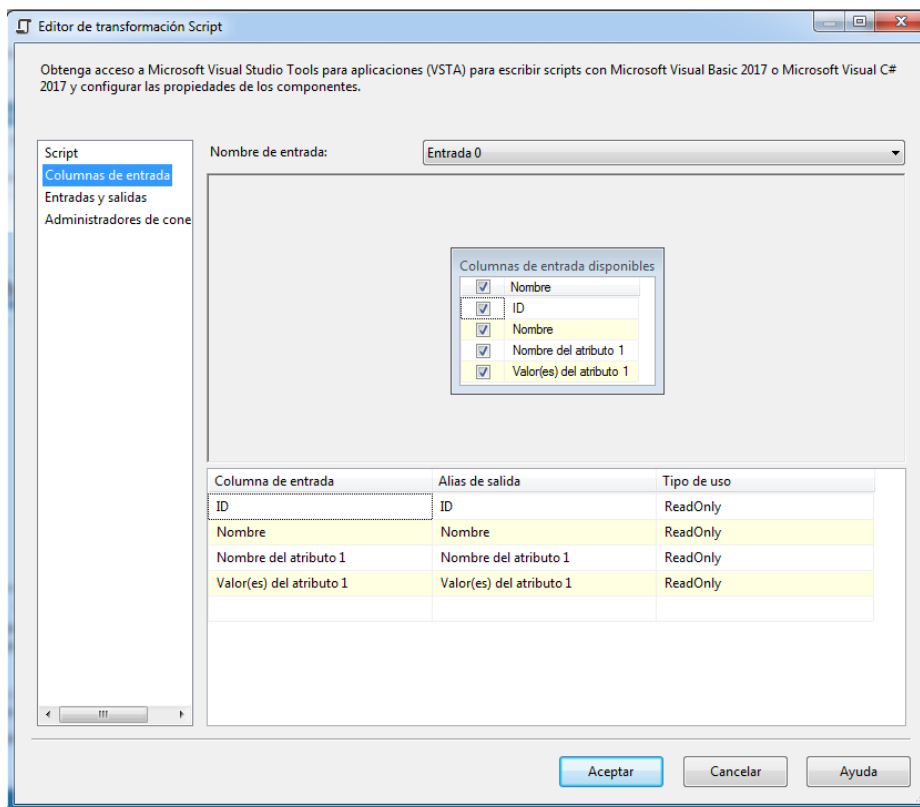
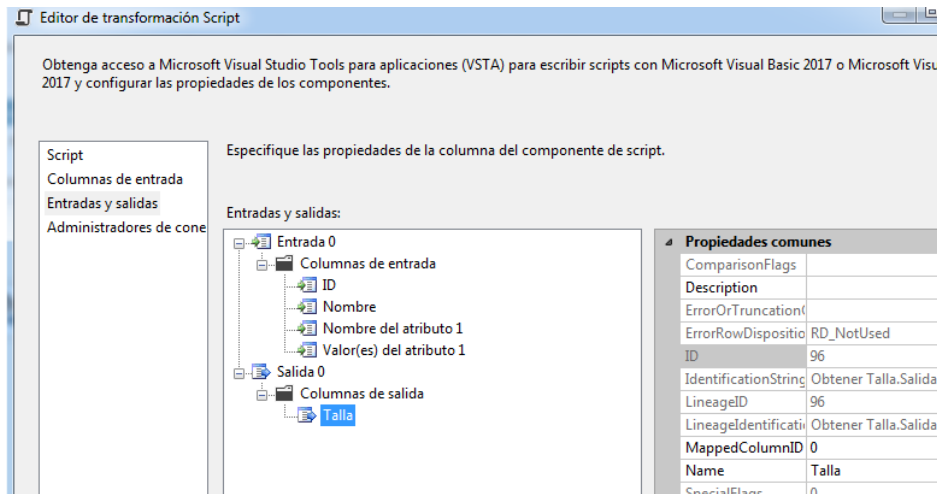


Figura N° 8: ETL Cargar Productos – Obtener Talla - Script



**Figura N° 9: ETL Cargar Productos – Obtener Talla – Entradas y Salidas**

### 3.3.4.1.3. Destino Productos BD-BIUNS

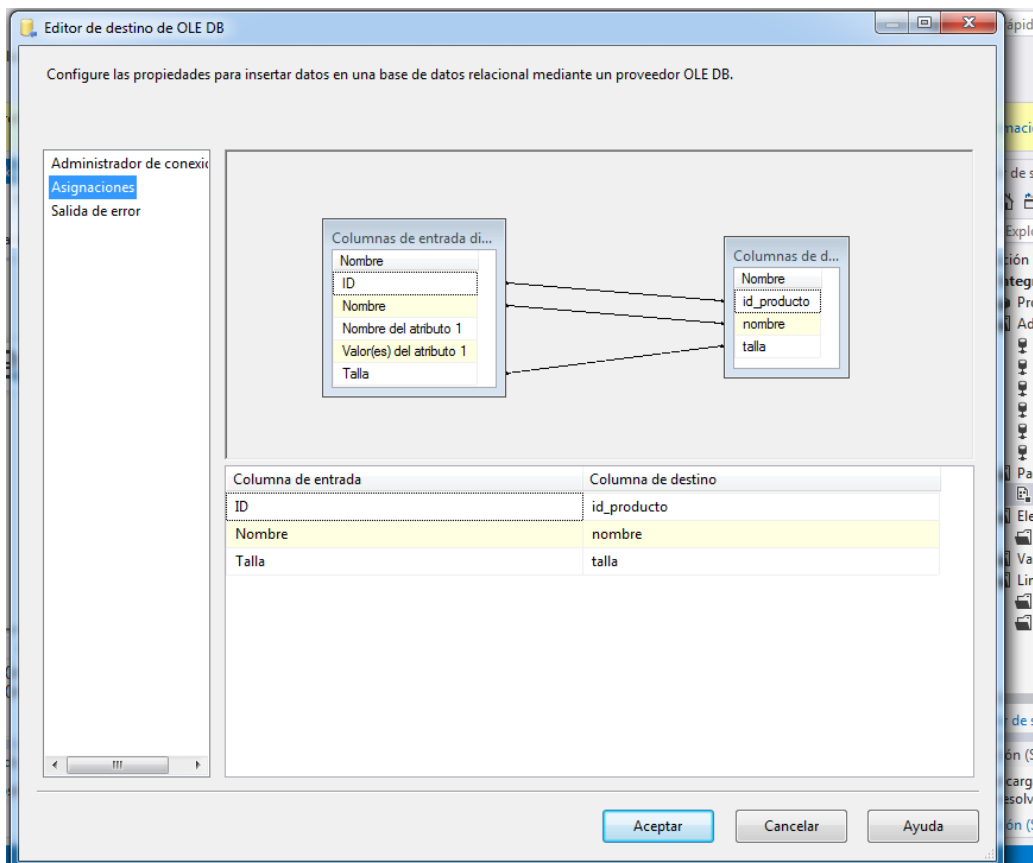
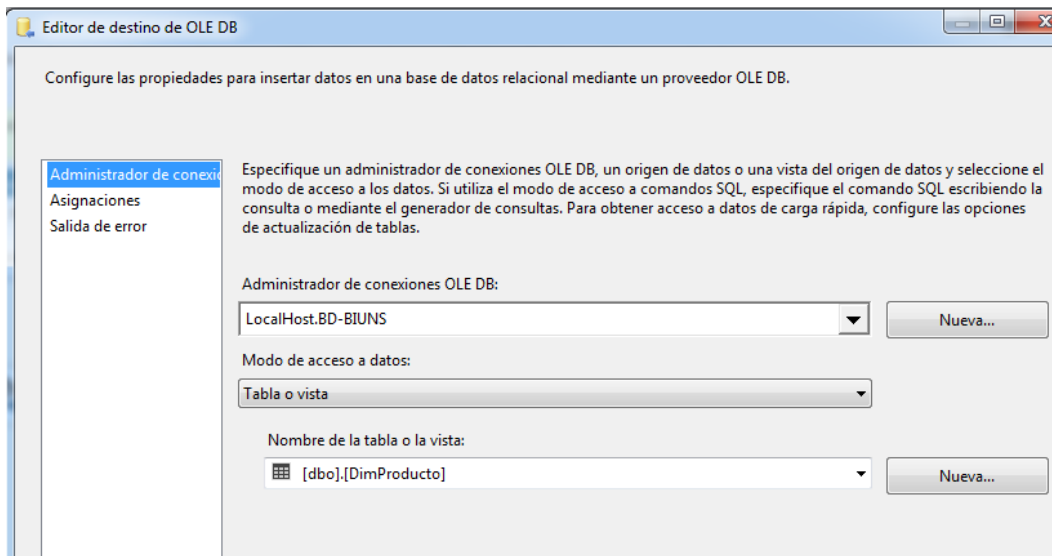


Figura N° 10: ETL Cargar Productos – Destino – Conexión y Asignaciones

### 3.3.4.1.4. Ejecución Cargar Productos – Integración de los Datos

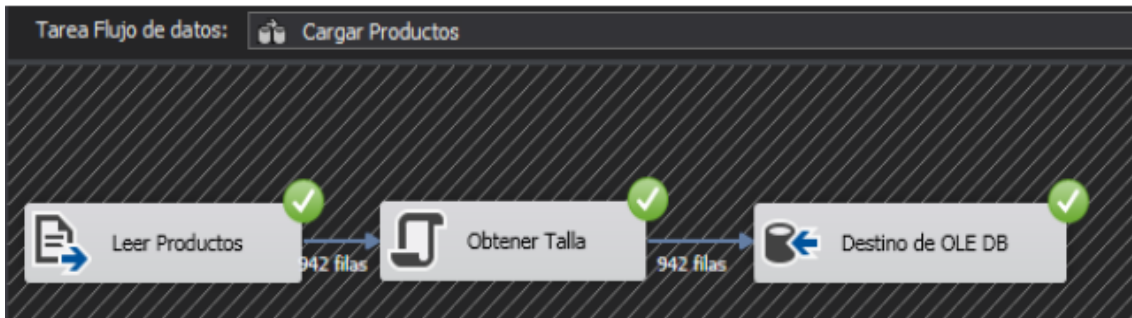


Figura N° 11: ETL Cargar Productos – Ejecución

### 3.3.4.2. Dimensión Regiones – Cargar Regiones

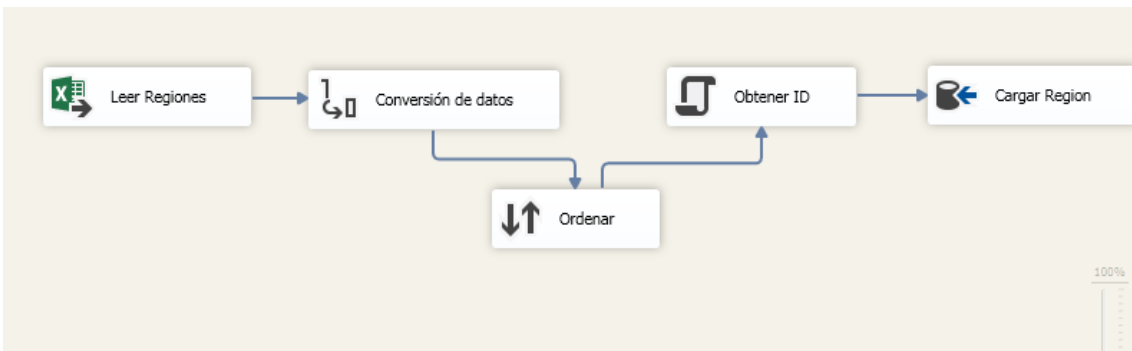


Figura N° 12: ETL Cargar Regiones

#### 3.3.4.2.1. Leer Regiones

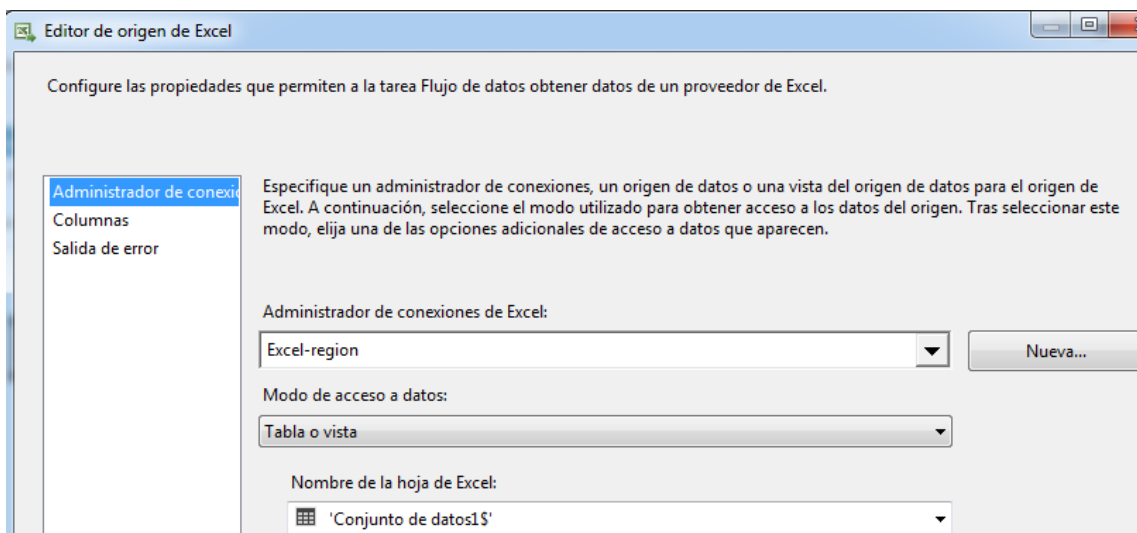
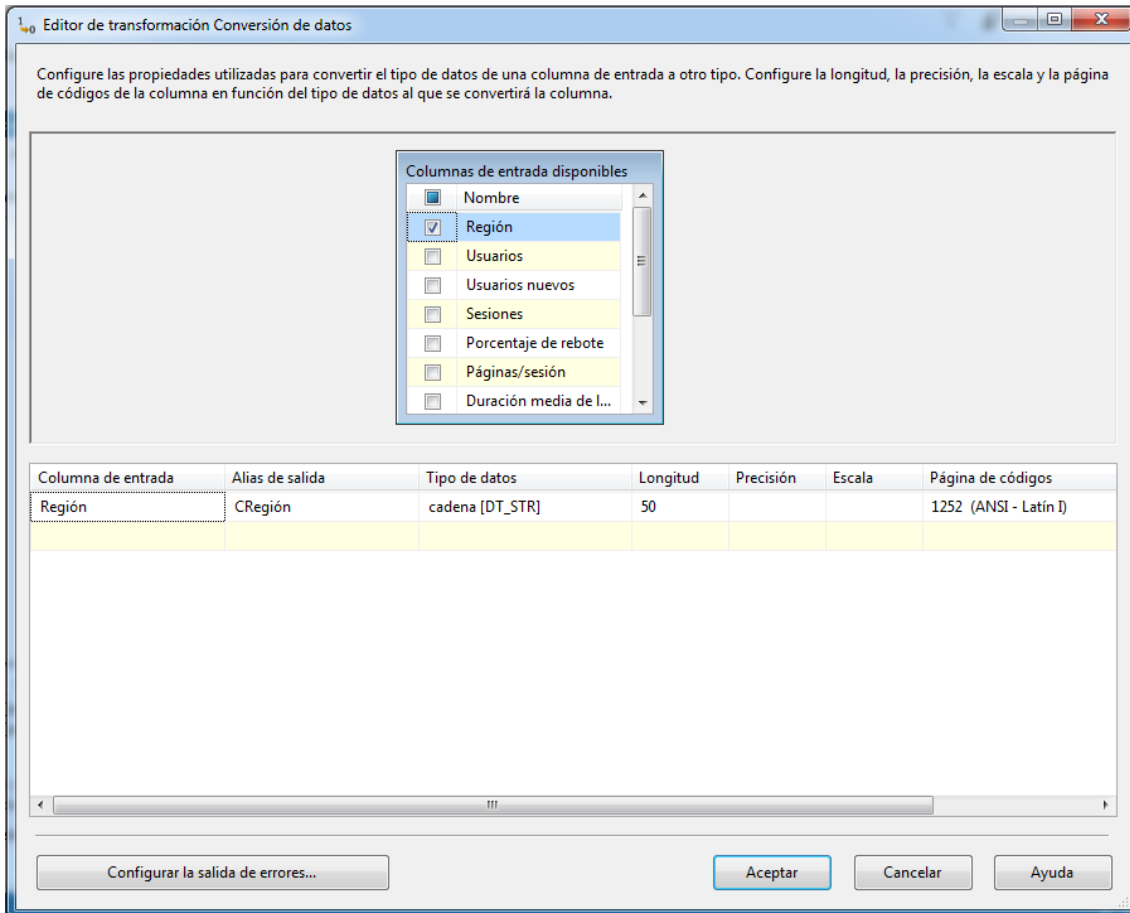


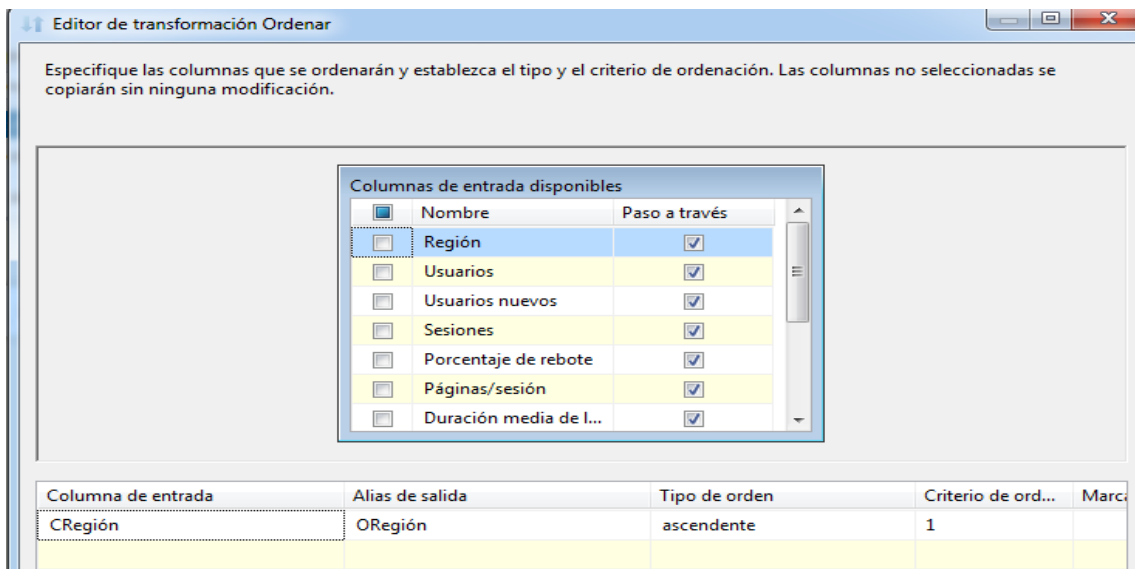
Figura N° 13: ETL Cargar Regiones – Leer Regiones

### 3.3.4.2. Conversión de Datos



**Figura N° 14: ETL Cargar Regiones – Conversión de datos**

### 3.3.4.2.3. Ordenar Datos



**Figura N° 15: ETL Cargar Regiones – Ordenar datos**



### 3.3.4.2.4. Obtener ID

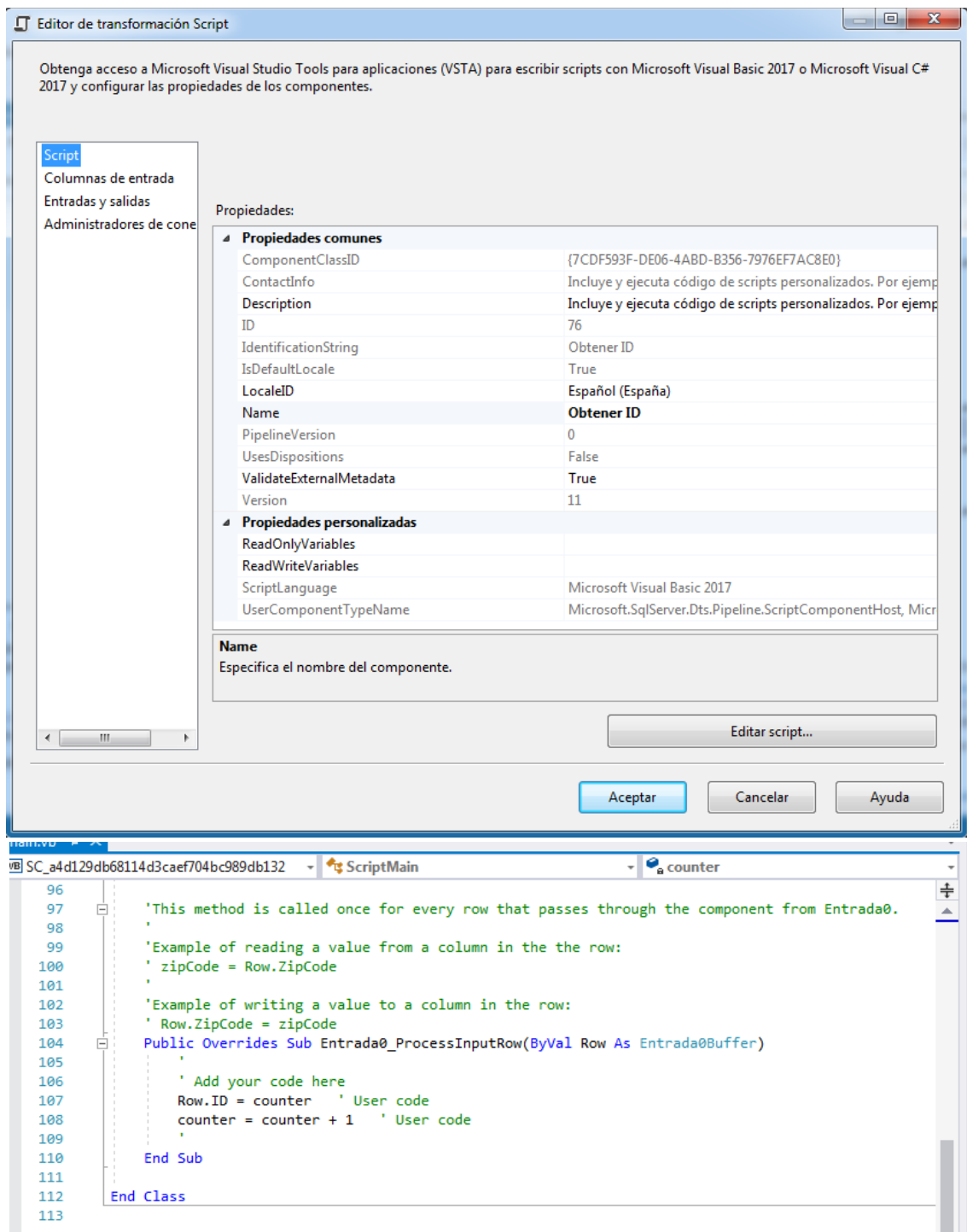
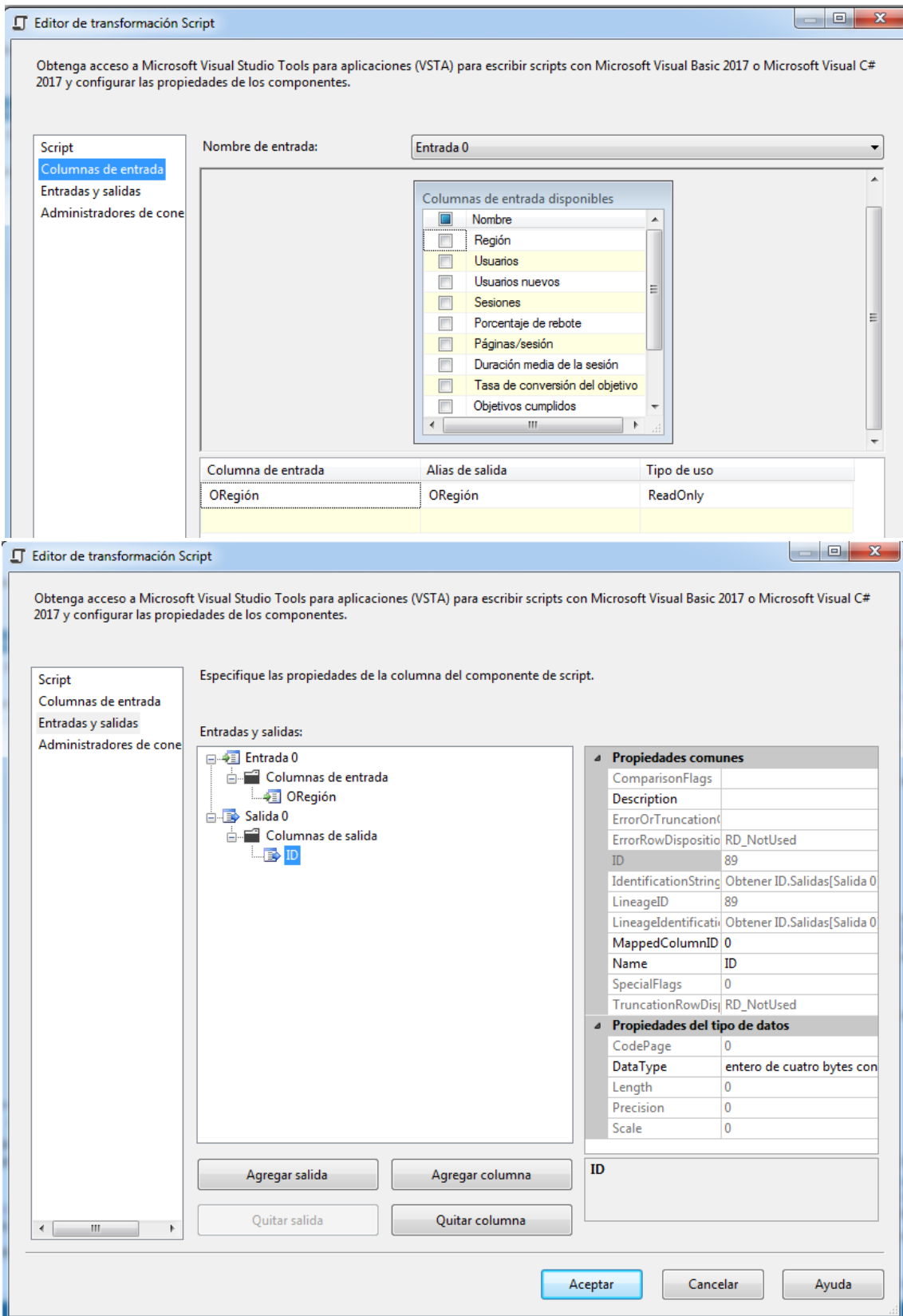


Figura N° 16: ETL Cargar Regiones – Obtener ID - Script



**Figura N° 17: ETL Cargar Regiones – Obtener ID – Columnas, entradas y salidas**

### 3.3.4.2.5. Destino Regiones BD-BIUNS

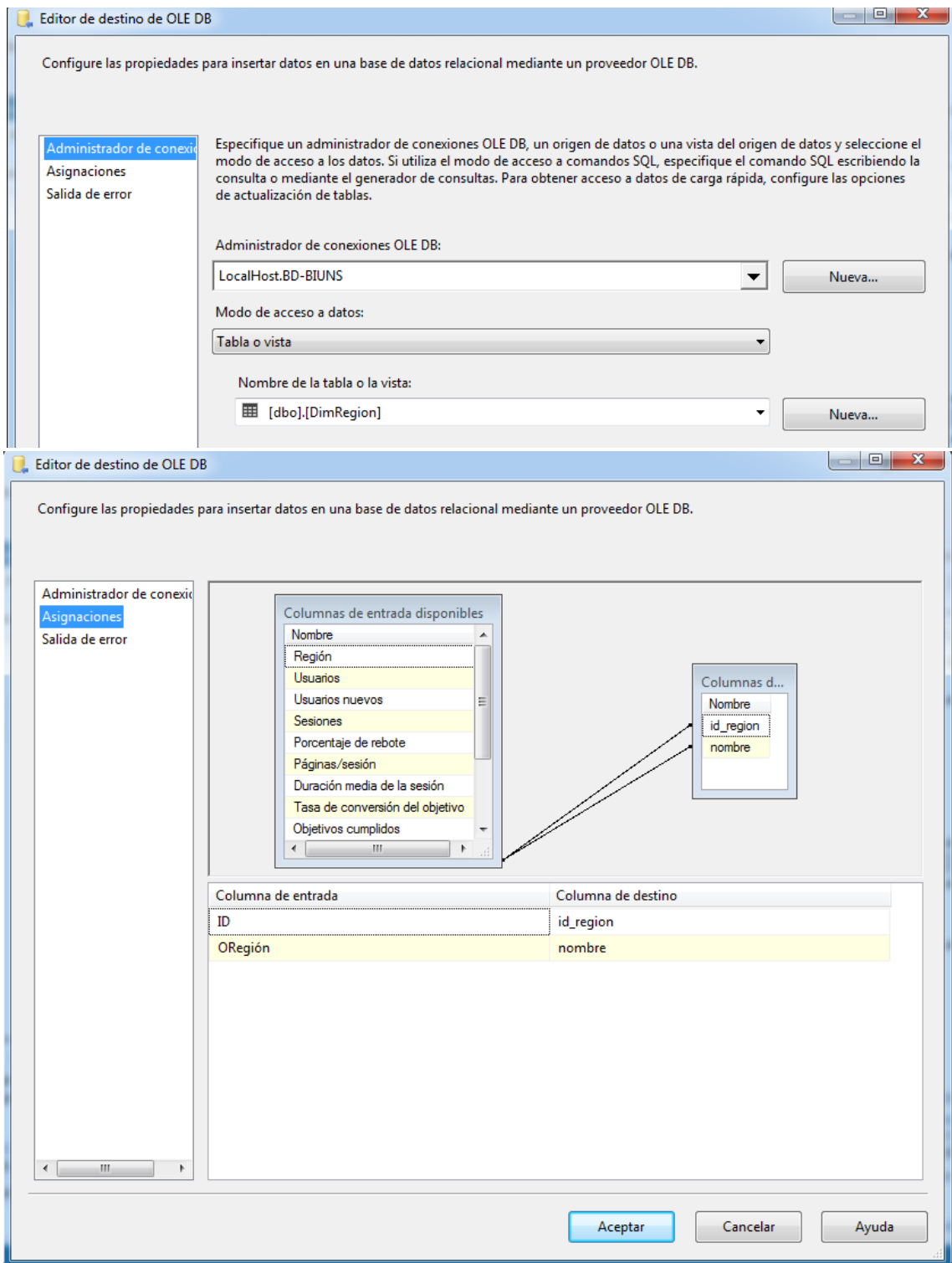


Figura N° 18: ETL Cargar Regiones – Destino – Conexión y Asignaciones

### 3.3.4.2.6. Ejecución Cargar Regiones - Integración de los datos

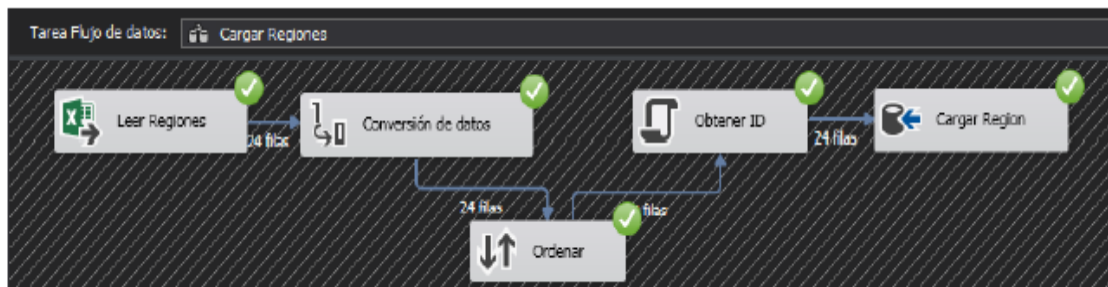


Figura N° 19: ETL Cargar Regiones – Ejecución

### 3.3.4.3 Dimensión Páginas – Cargar Páginas

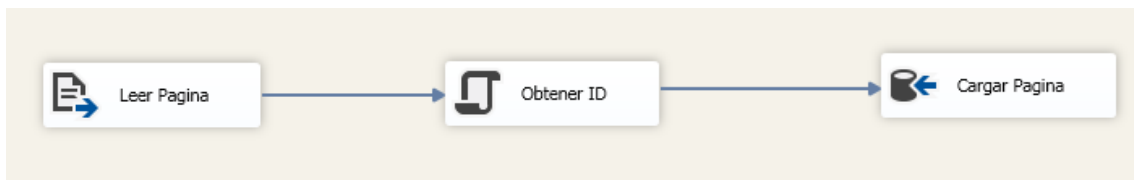


Figura N° 20: ETL Cargar Páginas

#### 3.3.4.3.1. ETL Cargar Páginas - Leer Páginas

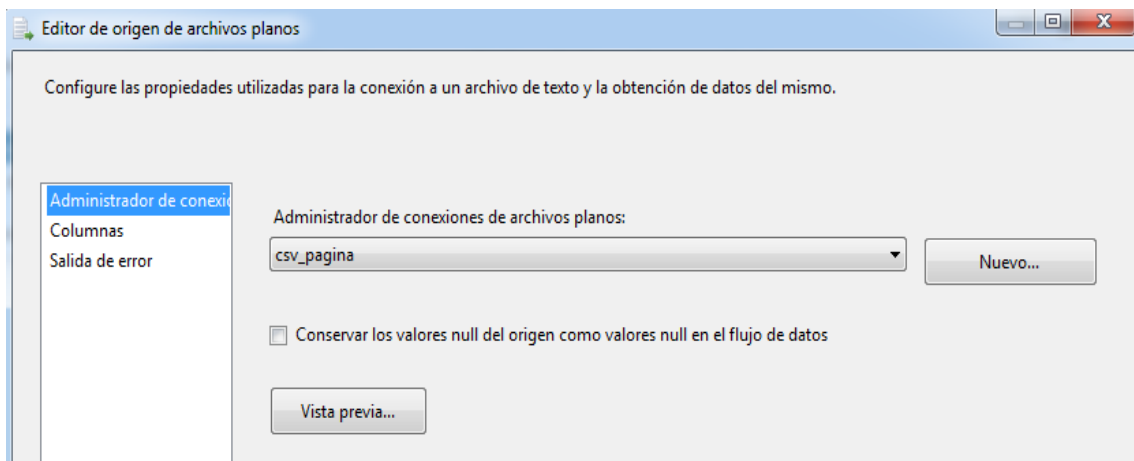


Figura N° 21: ETL Cargar Regiones – Leer Regiones

### 3.3.4.3.2. ETL Cargar Páginas - Obtener ID

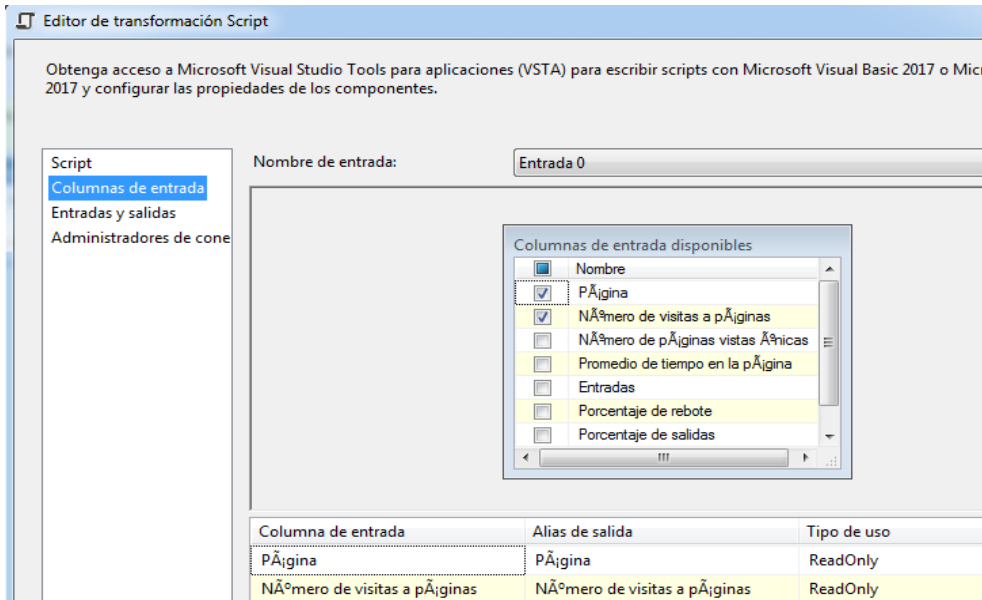
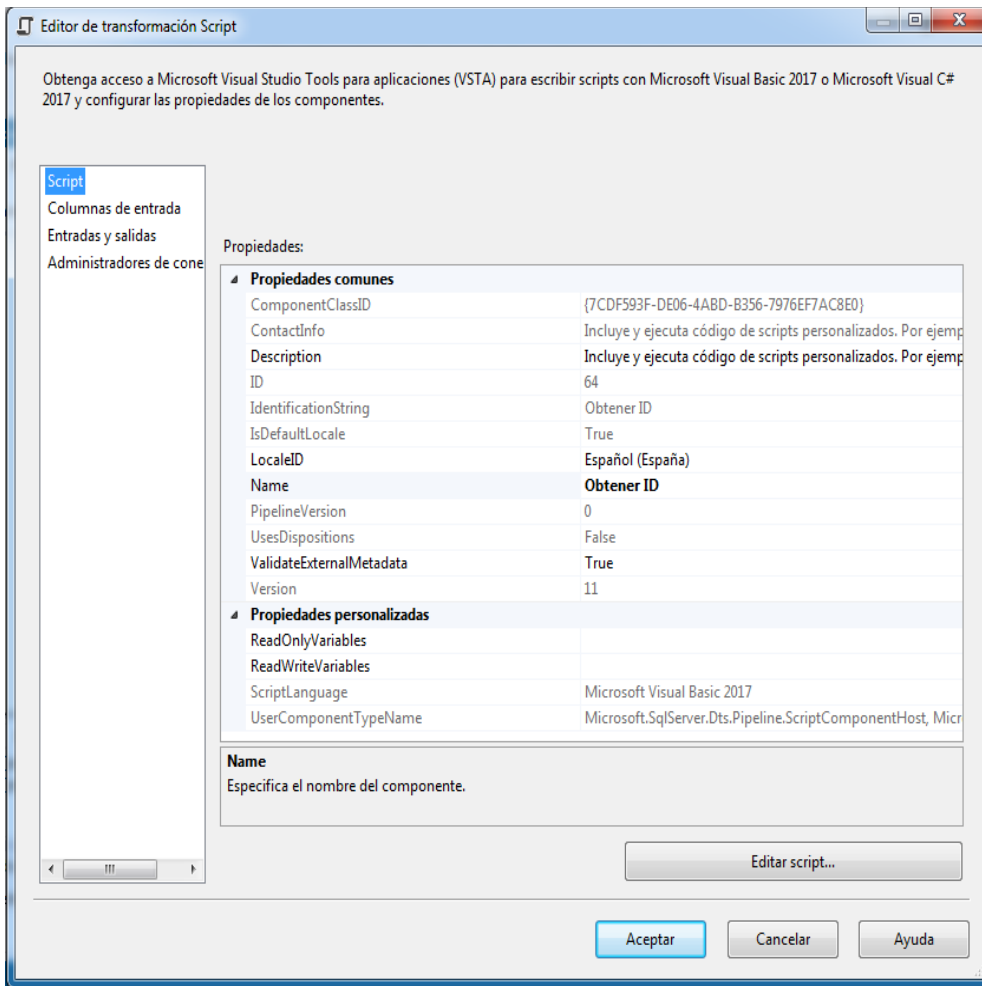


Figura N° 22: ETL Cargar Páginas – Obtener ID

### 3.3.4.3.3. ETL Cargar Páginas – Destino

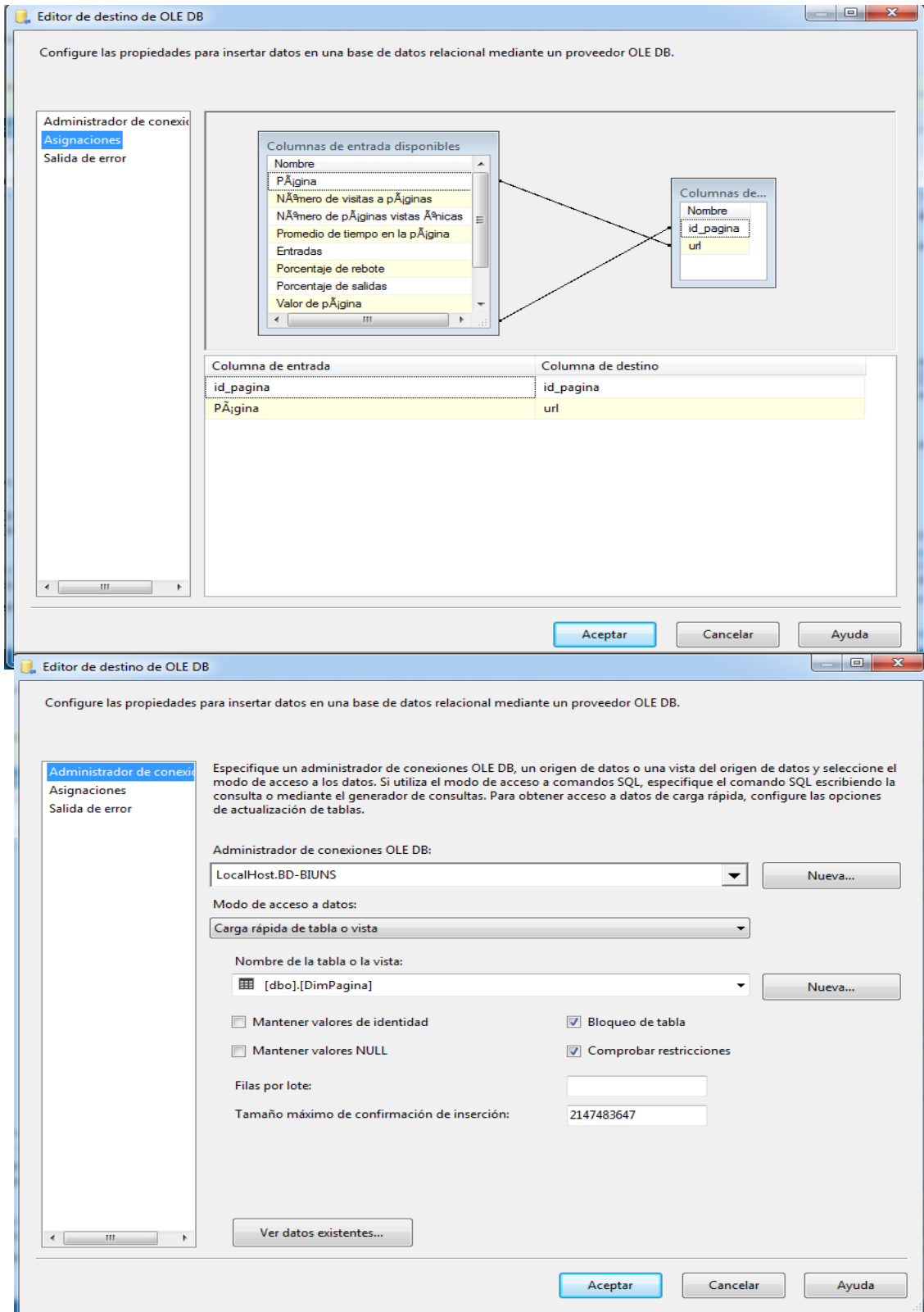


Figura N° 23: ETL Cargar Páginas – Destino

### 3.3.4.4 Dimensión Tiempo

#### 3.3.4.4.1. Leer Fechas

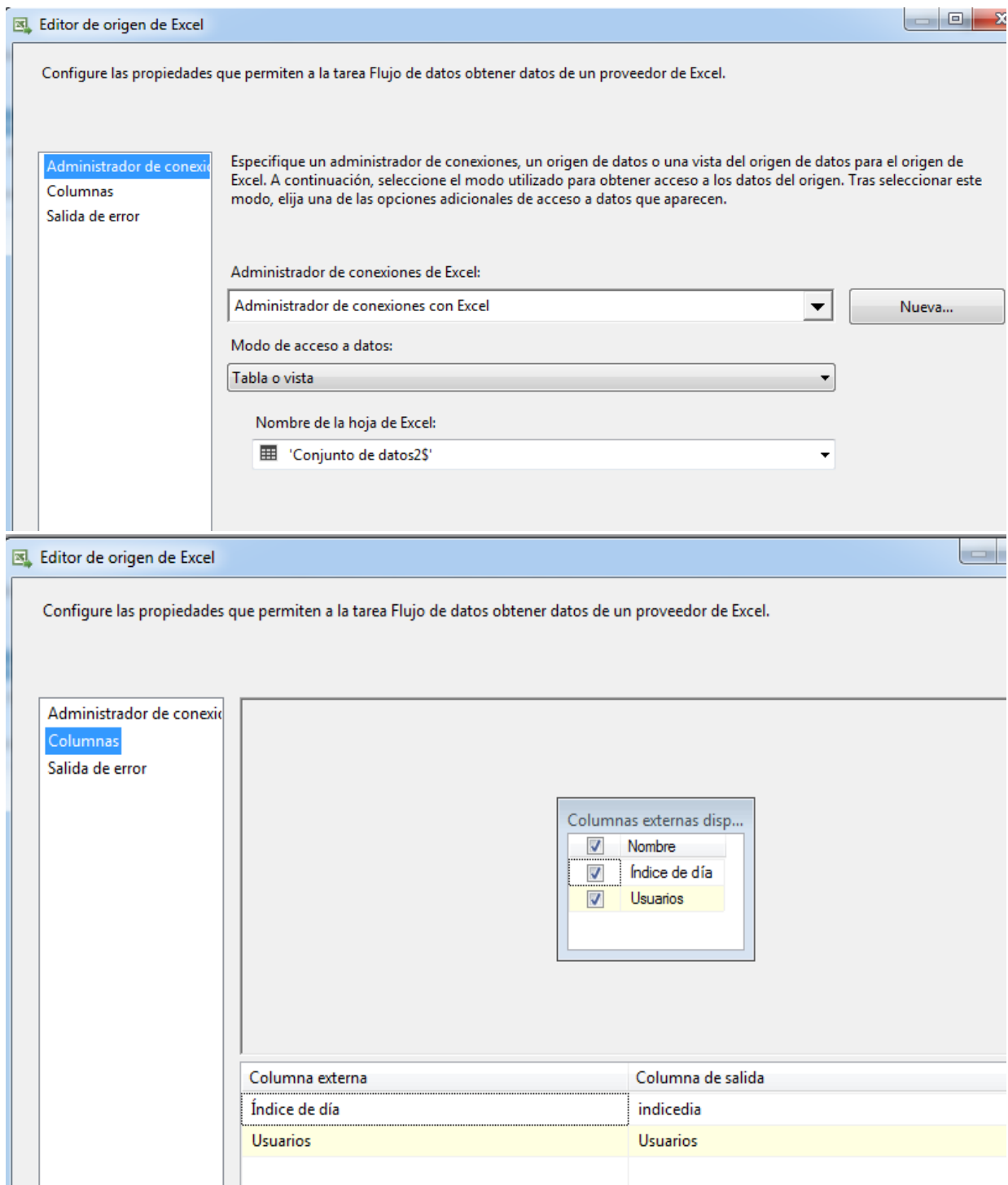


Figura N° 24: ETL Leer Fechas

### 3.3.4.4.2. Dividir fecha - Script

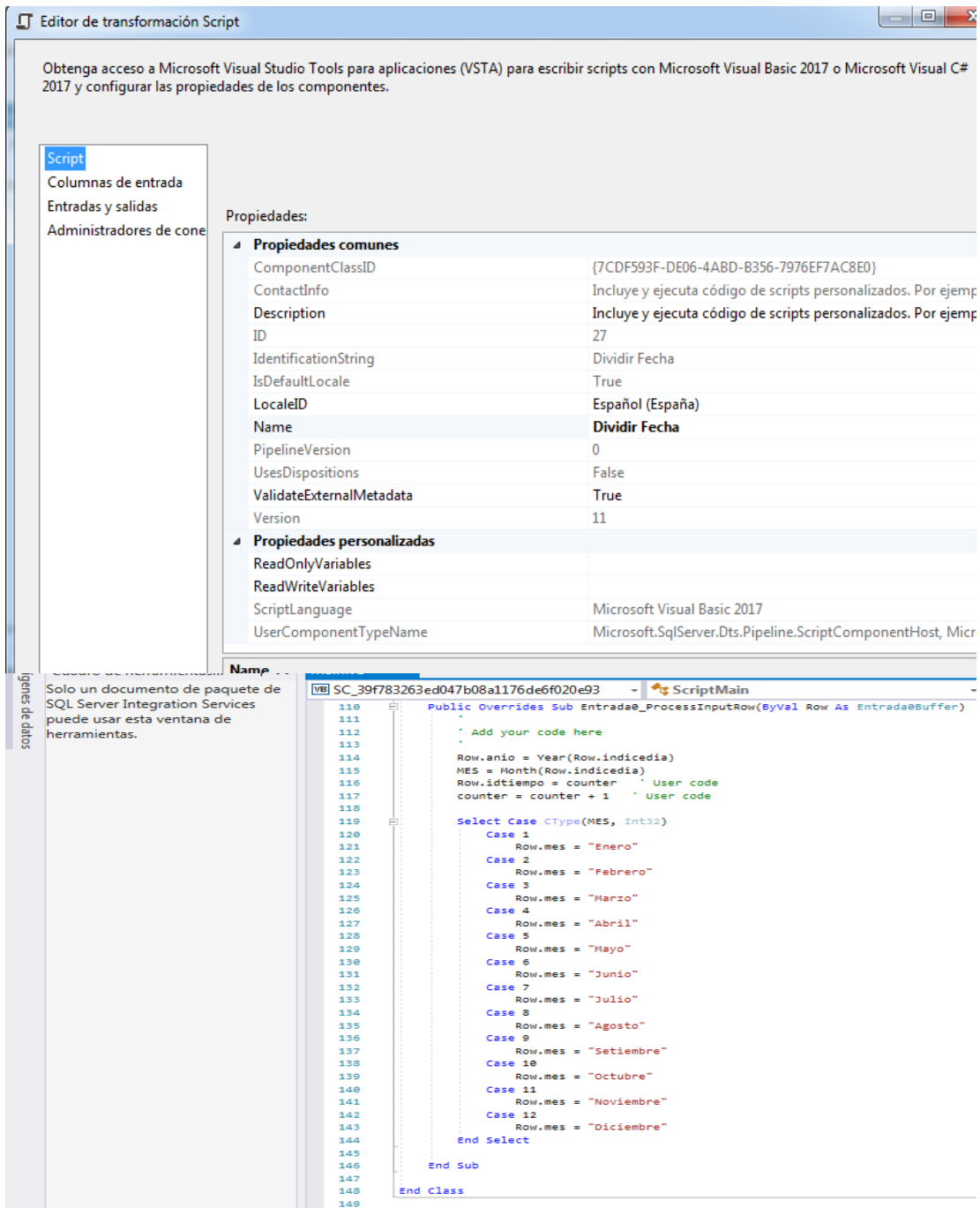


Figura N° 25: ETL Dividir Fechas – Script



### 3.3.4.4.3. Destino BD-BIUNS

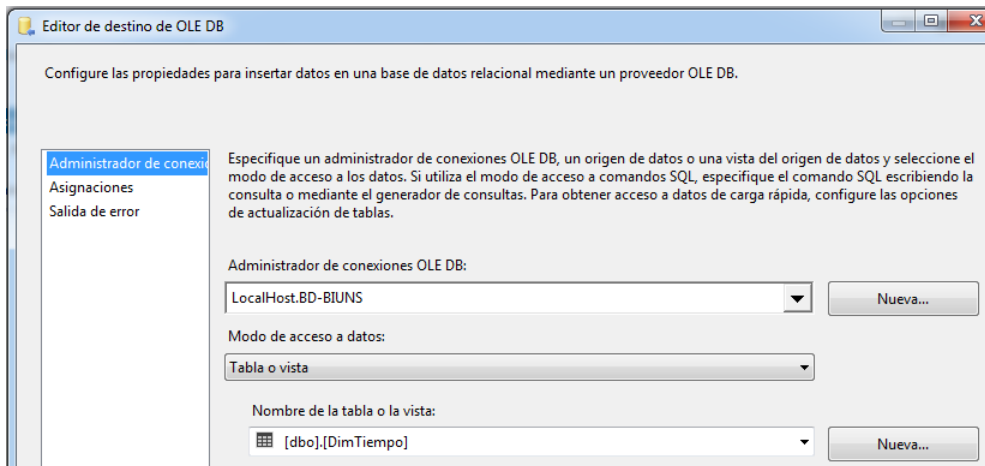


Figura N° 26: ETL Destino Fechas Conexión BD-BIUNS

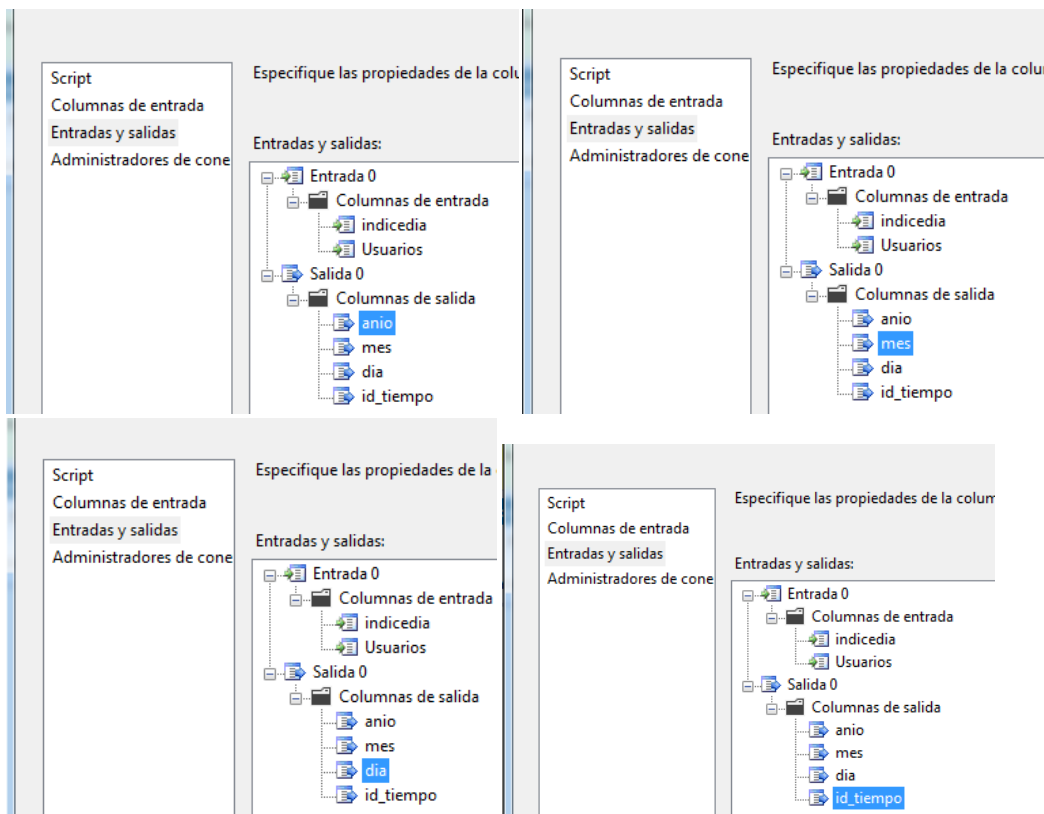
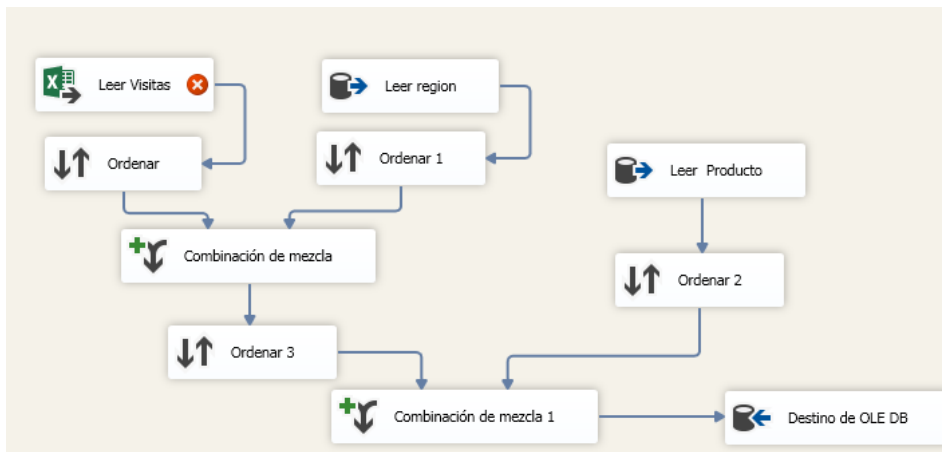


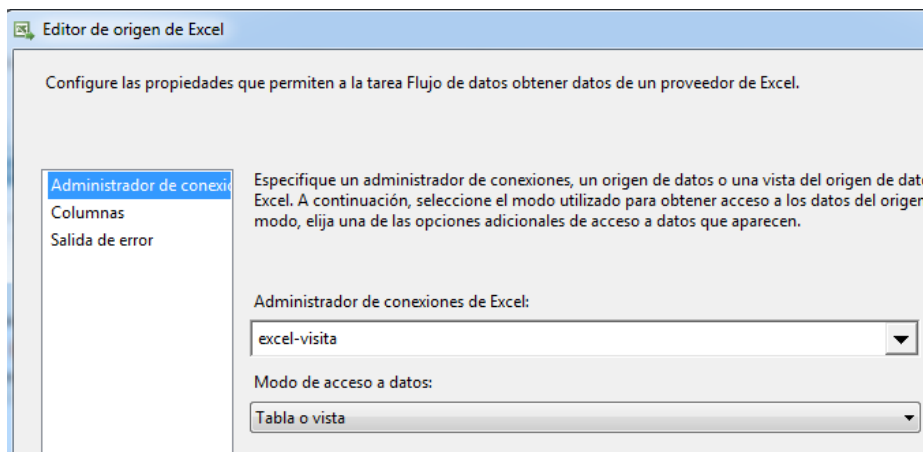
Figura N° 27: ETL Destino Fechas especificaciones Anio, mes, dia, id\_tiempo

### 3.3.4.5. Dimensión Visitas



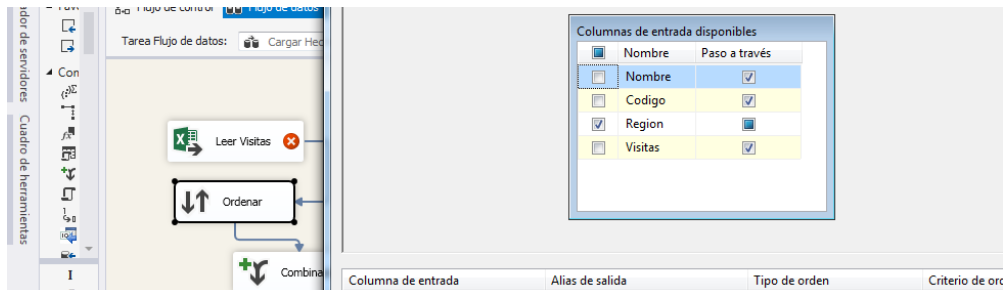
**Figura N° 28: Dimensión Visitas**

#### 3.3.4.5.1. Dimensión Visitas – Leer Visitas Páginas



**Figura N° 29: ETL Dimensión Visitas – Leer Visitas**

#### 3.3.4.5.2. Dimensión Visitas – Ordenar Visitas



**Figura N° 30: ETL Dimensión Visitas – Ordenar Visitas**

### 3.3.4.5.3. Dimensión Visitas – Leer Región

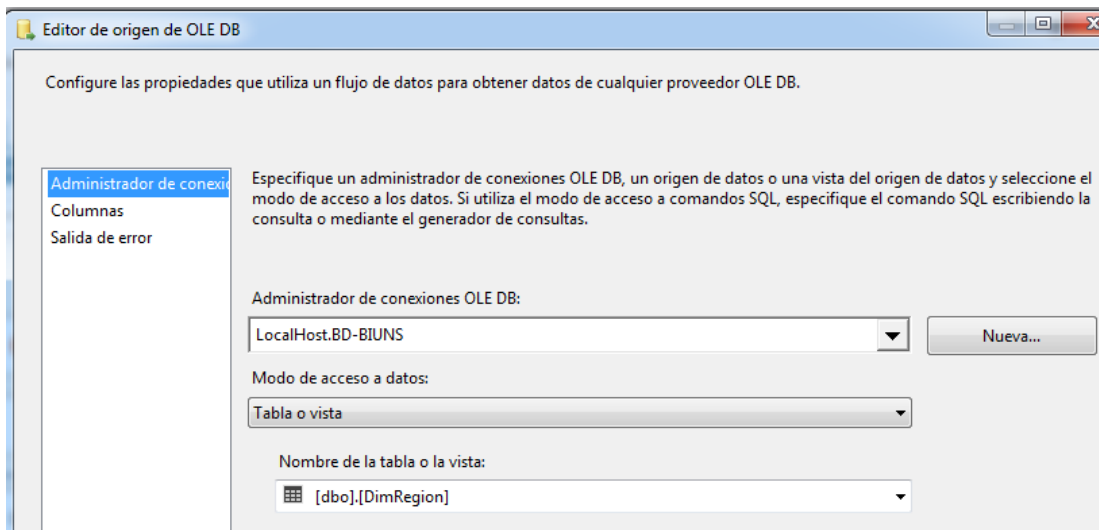


Figura N° 31: ETL Dimensión Visitas – Leer Región

### 3.3.4.5.4. Dimensión Visitas – Ordenar 1 (Región)

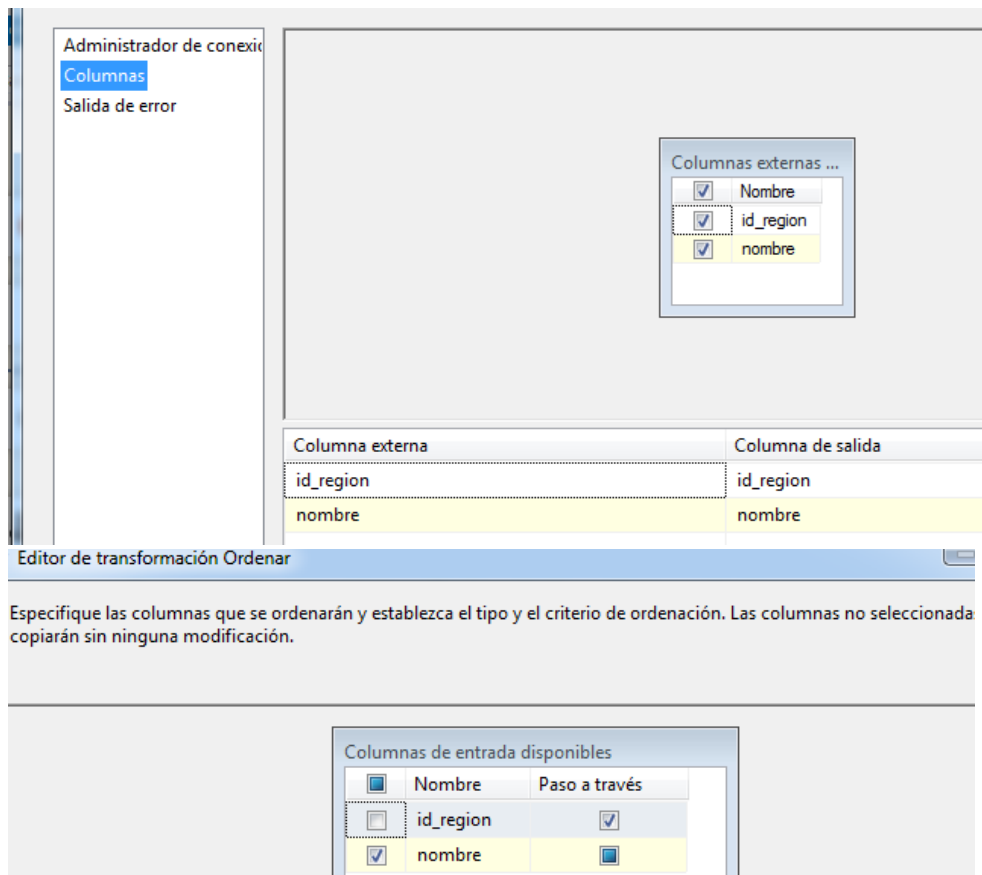


Figura N° 32: ETL Dimensión Visitas – Ordenar 1 (Región)

### 3.3.4.5.5. Dimensión Visitas – Combinación de mezcla

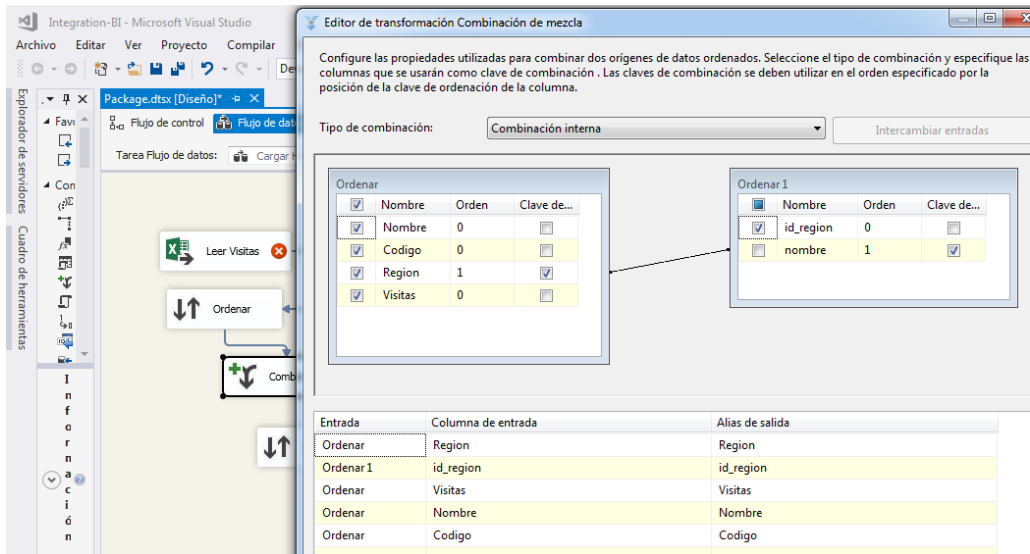


Figura N° 33: ETL Dimensión Visitas – Combinación de mezcla

### 3.3.4.5.6. Dimensión Visitas – Ordenar 3

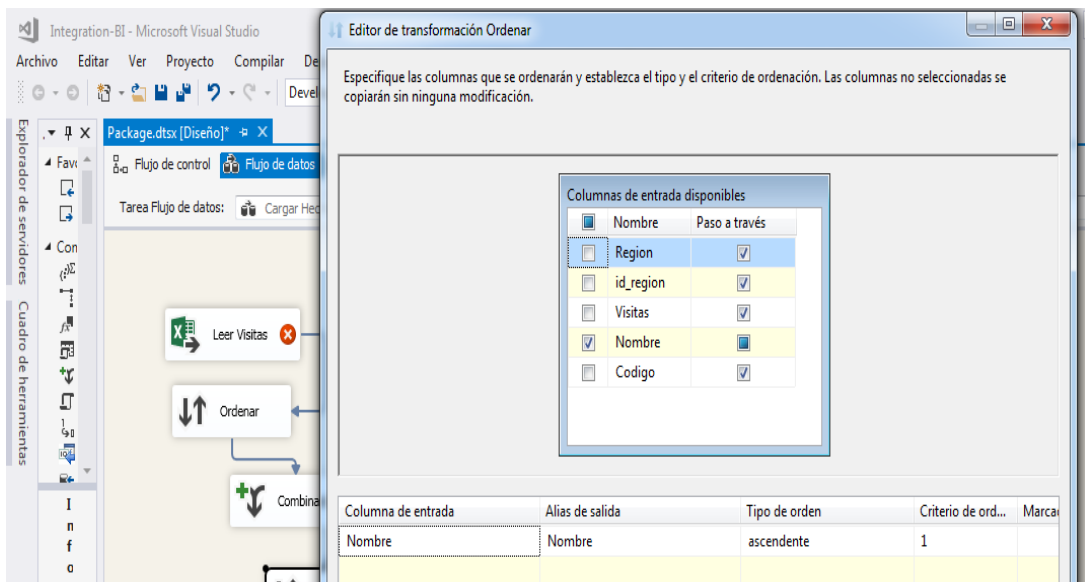
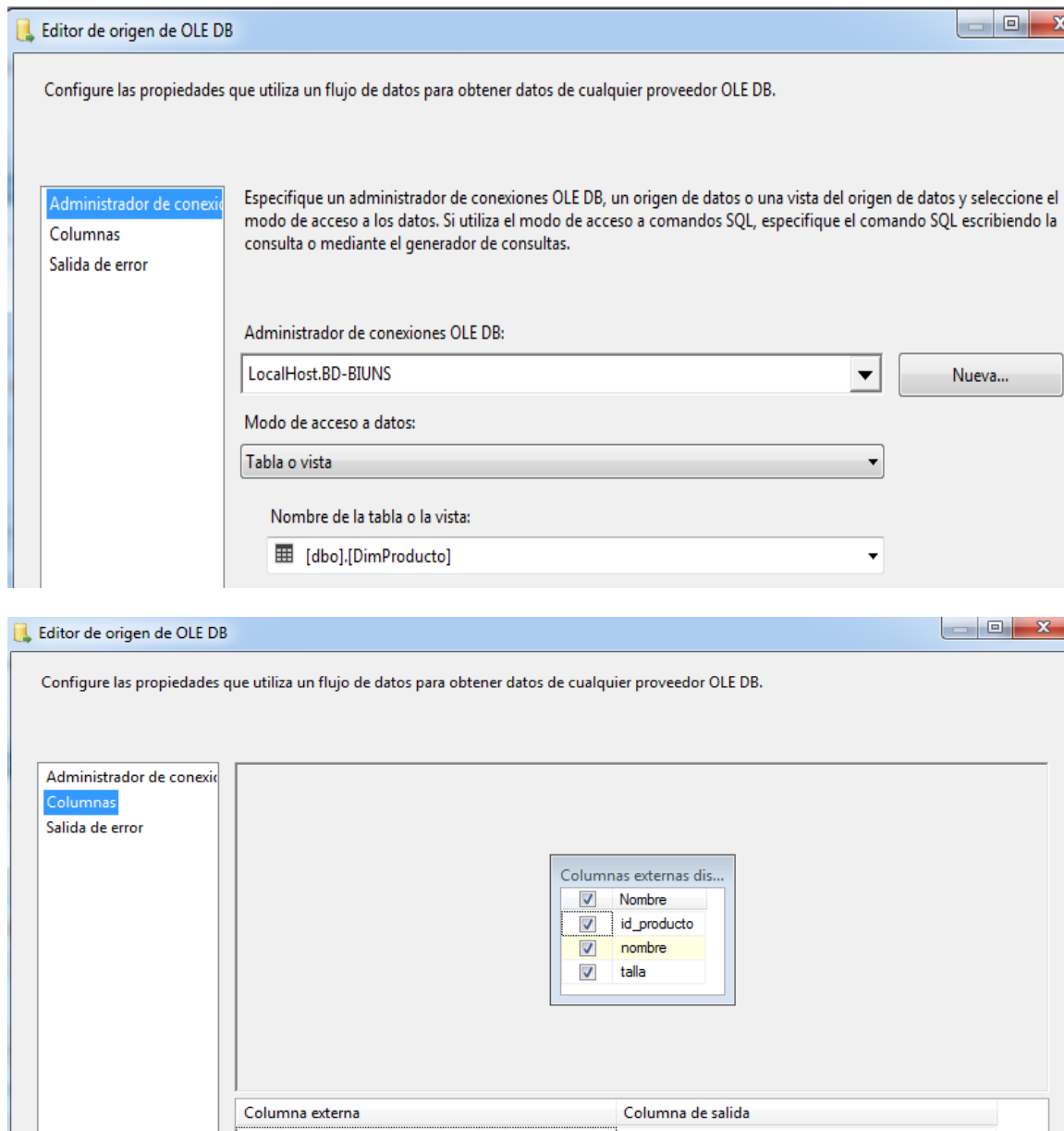


Figura N° 34: ETL Dimensión Visitas – Ordenar 3

### 3.3.4.5.7. Dimensión Visitas – Leer producto



**Figura N° 35: ETL Dimensión Visitas – Leer producto:**

### 3.3.4.5.8. Dimensión Visitas – Ordenar 2

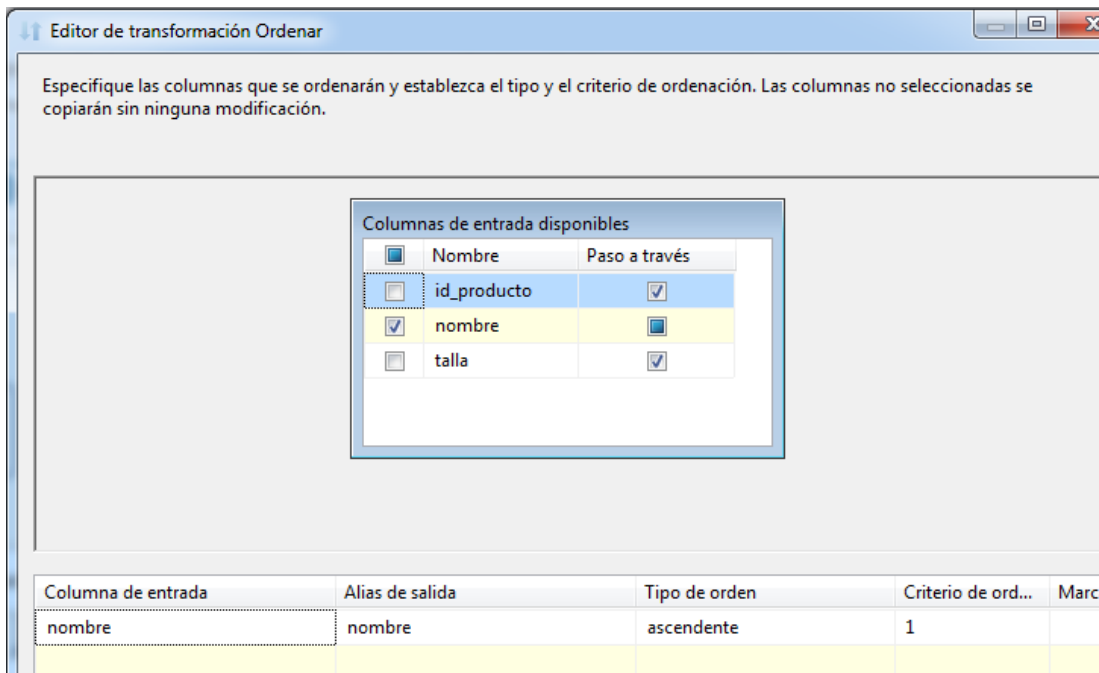


Figura N° 36: ETL Dimensión Visitas – Ordenar 2

### 3.3.4.5.9. Dimensión Visitas – Combinación de mezcla 1 (Visitas, Regiones, Productos)

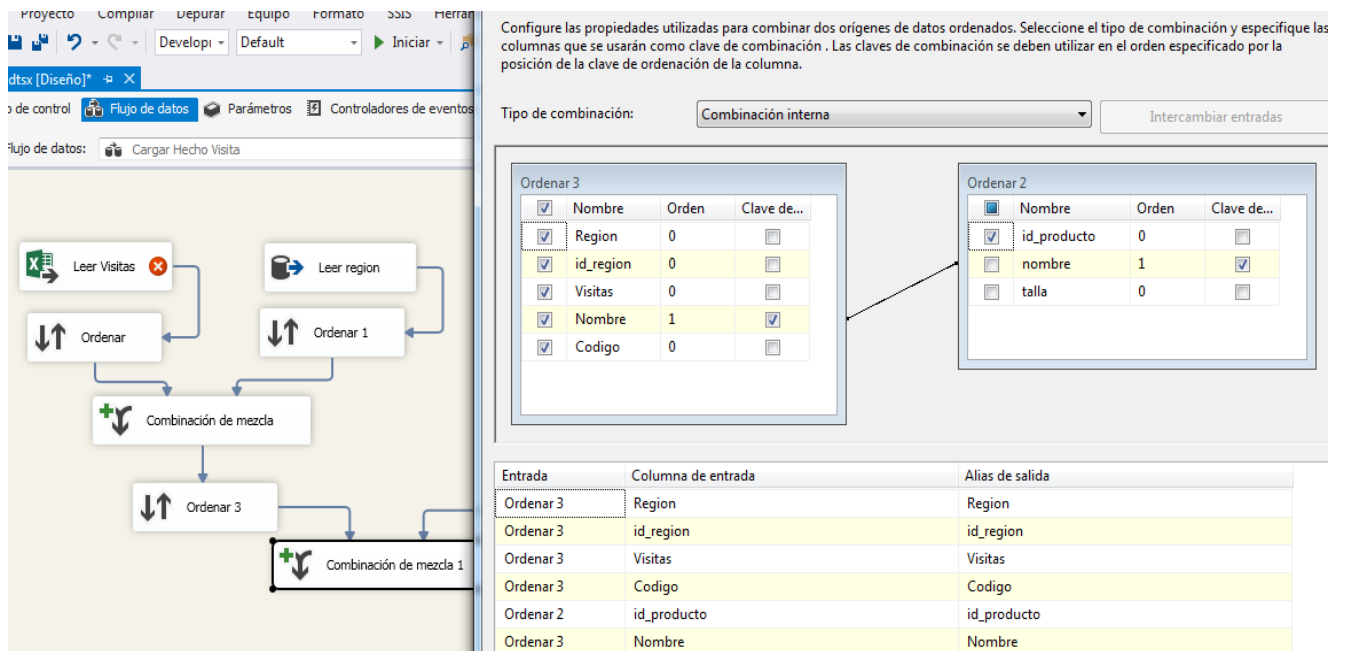


Figura N° 37: ETL Dimensión Visitas – Combinación de mezcla 1 (Visitas, Regiones, Productos)

### 3.3.4.5.10. Destino Visitas BD-BIUNS

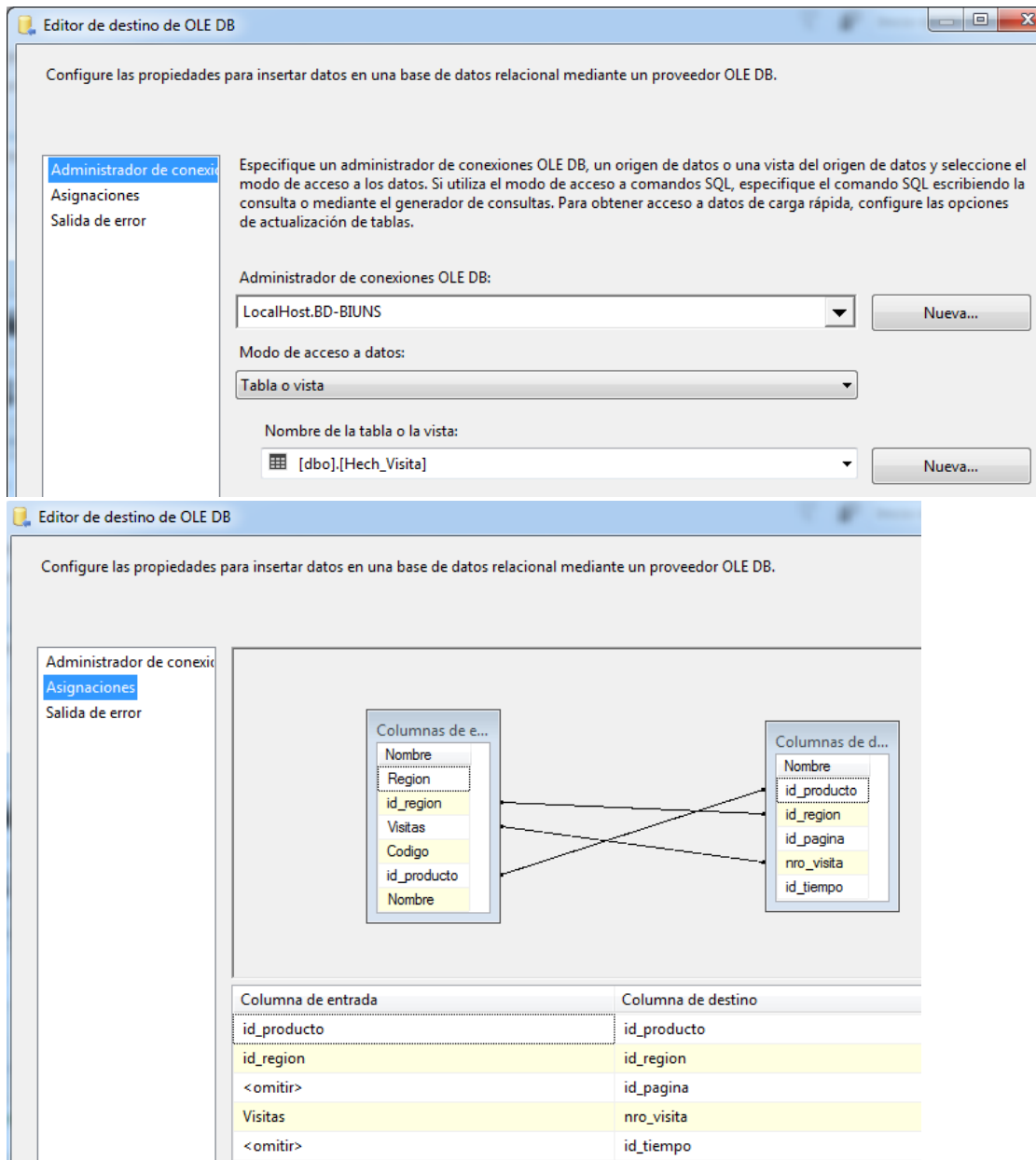
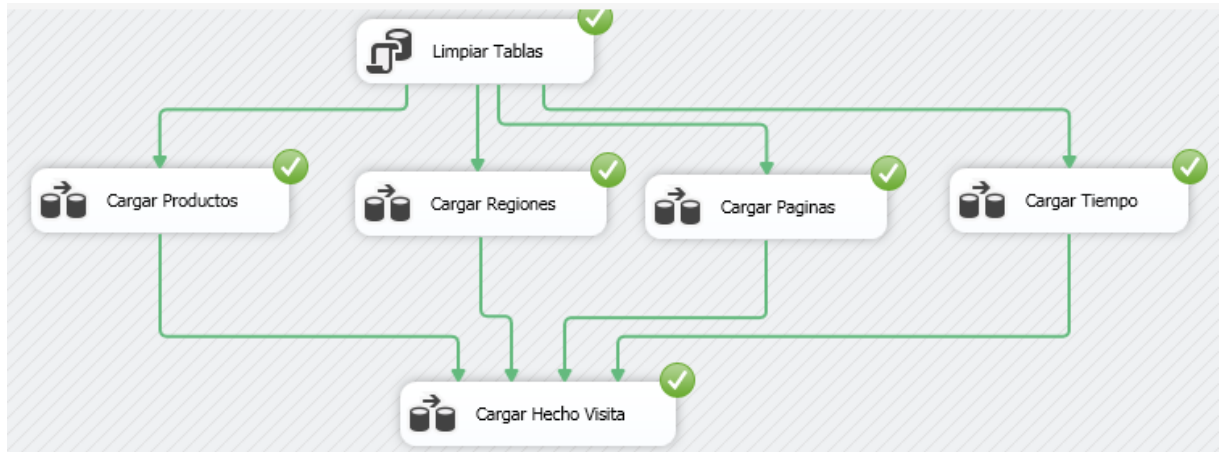


Figura N° 38: ETL Destino Visitas Conexión BD-BIUNS

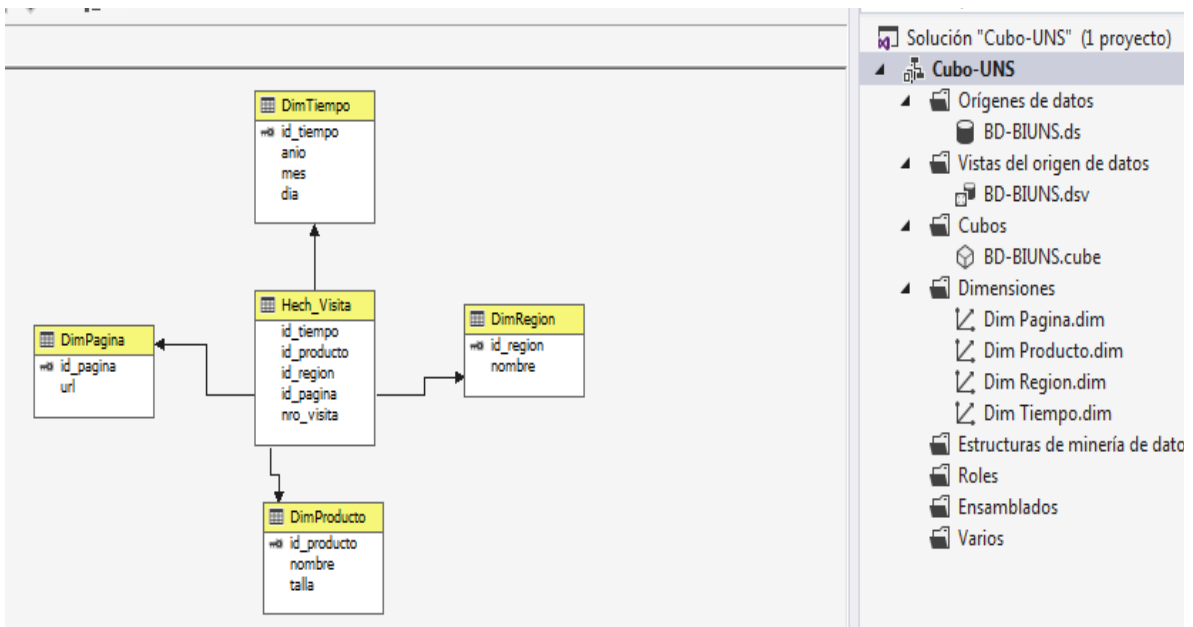
### 3.3.6 ETL Malla Ejecución del proceso

Se continúa con la ejecución del proceso ETL y la prueba que ha ejecutado correctamente.



**Figura N° 39: ETL Destino Malla Ejecución**

### 3.3.7 Cubo



**Figura N° 40: Cubo**



### **3.4. Fase de modelado**

En esta etapa de la metodología se elige la técnica o técnicas más apropiadas para el cumplimiento de los objetivos o desarrollo del modelo, con esto se procede a realizar un plan de prueba para la técnica elegida, construir y evaluar la misma.

#### ***3.4.1. Selección de la técnica del modelo***

En el presente proyecto se está considerando el uso del algoritmo recomendador machine learning denominado Descomposición en Valores Singulares o simplemente SVD, por ajustarse al negocio, debido a que es una técnica aplicada a recomendaciones en entornos comerciales.

### 3.4.2. Generación de un plan de prueba

Se pretende utilizar el dataset generado en el ETL de la fase anterior a partir de las fuentes de datos para luego procesarlas con Python y librerías de machine learning y lograr recomendaciones útiles en la toma de decisiones, el plan de pruebas va de la forma del siguiente esquema:

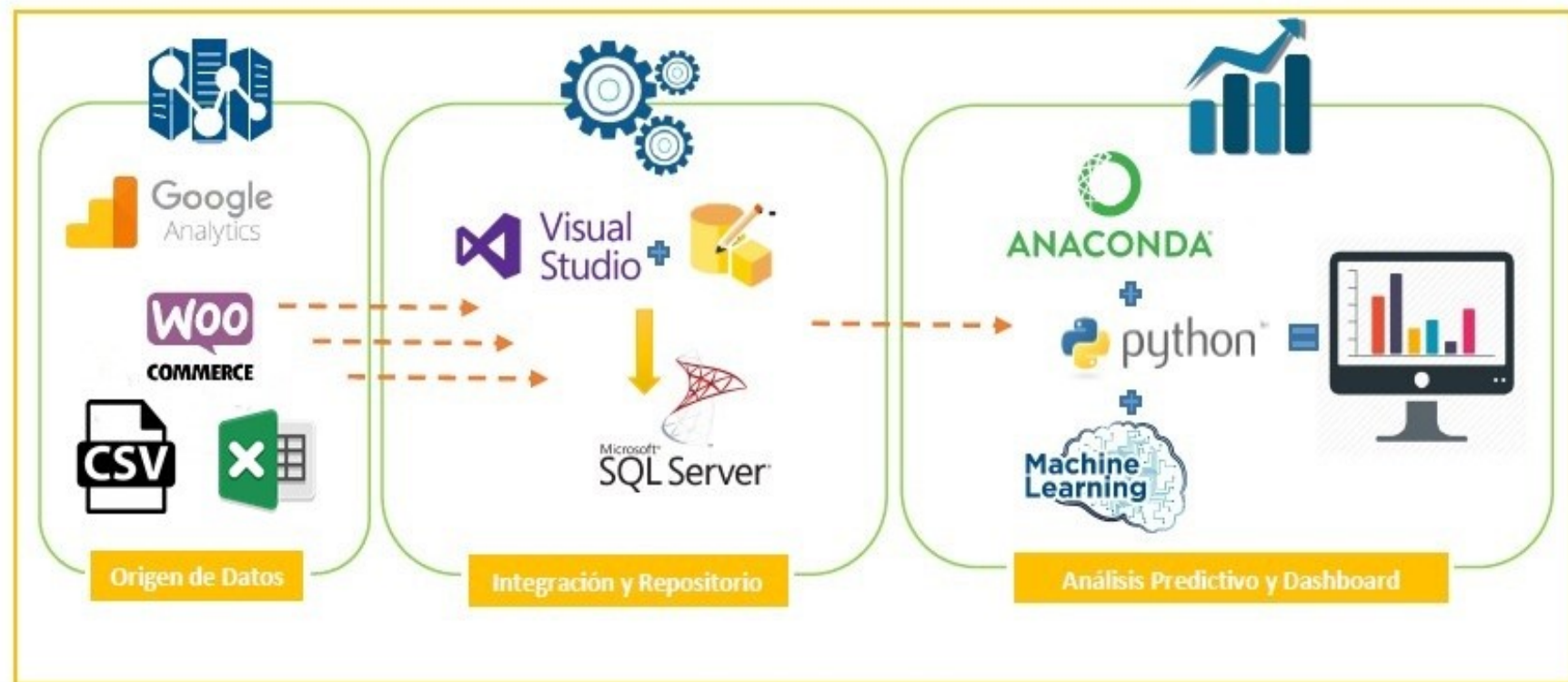
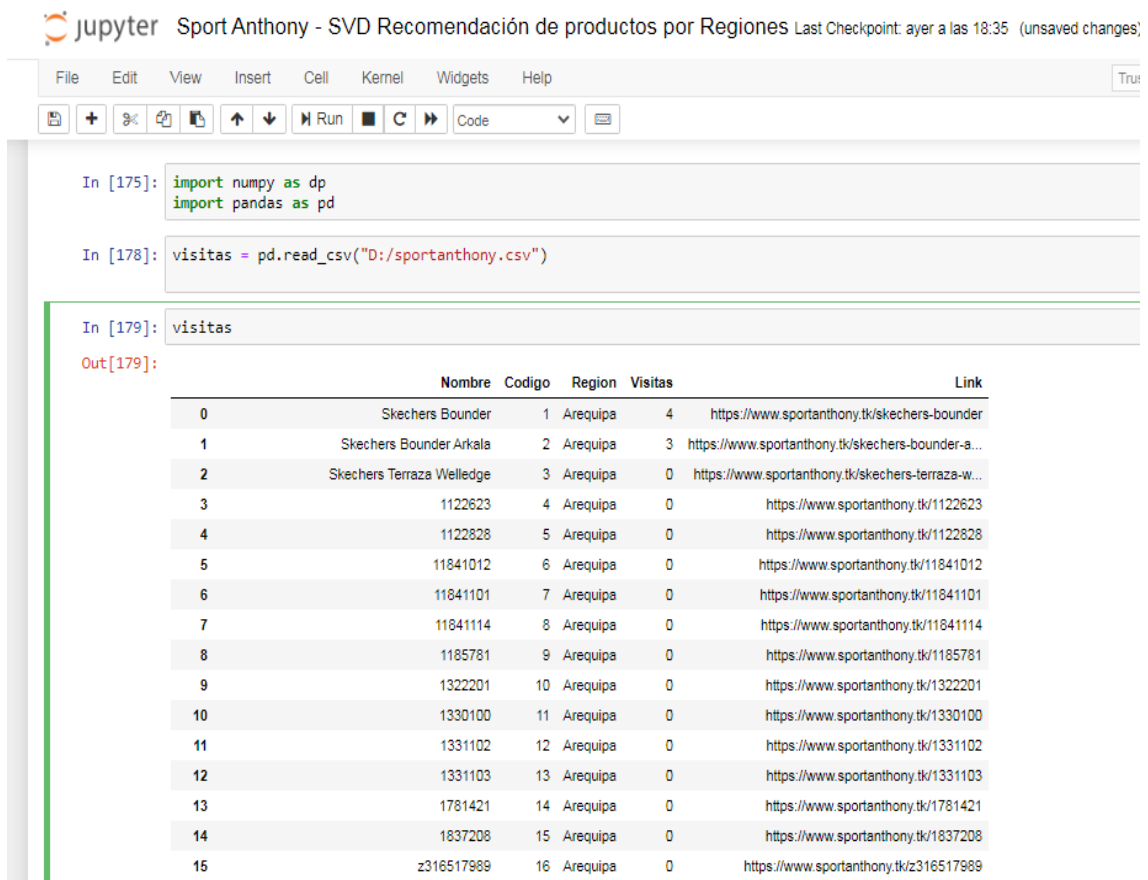


Figura N° 41: Arquitectura

### 3.4.3. Construcción del modelo

Luego de finalizar con el Datamart, del cual haremos uso de las dimensiones, producto, región, visita.

Con Jupyter Notebook se siguieron los siguientes pasos para implementar el algoritmo recomendador mediante SVD en Python usando el Dataset obtenido del Datamart.



```
In [175]: import numpy as np
import pandas as pd

In [178]: visitas = pd.read_csv("D:/sportanthony.csv")

In [179]: visitas
```

	Nombre	Codigo	Region	Visitas	Link
0	Skechers Bounder	1	Arequipa	4	<a href="https://www.sportanthony.tk/skechers-bounder">https://www.sportanthony.tk/skechers-bounder</a>
1	Skechers Bounder Arkala	2	Arequipa	3	<a href="https://www.sportanthony.tk/skechers-bounder-a...">https://www.sportanthony.tk/skechers-bounder-a...</a>
2	Skechers Terraza Welledge	3	Arequipa	0	<a href="https://www.sportanthony.tk/skechers-terraz-a-w...">https://www.sportanthony.tk/skechers-terraz-a-w...</a>
3	1122623	4	Arequipa	0	<a href="https://www.sportanthony.tk/1122623">https://www.sportanthony.tk/1122623</a>
4	1122828	5	Arequipa	0	<a href="https://www.sportanthony.tk/1122828">https://www.sportanthony.tk/1122828</a>
5	11841012	6	Arequipa	0	<a href="https://www.sportanthony.tk/11841012">https://www.sportanthony.tk/11841012</a>
6	11841101	7	Arequipa	0	<a href="https://www.sportanthony.tk/11841101">https://www.sportanthony.tk/11841101</a>
7	11841114	8	Arequipa	0	<a href="https://www.sportanthony.tk/11841114">https://www.sportanthony.tk/11841114</a>
8	1185781	9	Arequipa	0	<a href="https://www.sportanthony.tk/1185781">https://www.sportanthony.tk/1185781</a>
9	1322201	10	Arequipa	0	<a href="https://www.sportanthony.tk/1322201">https://www.sportanthony.tk/1322201</a>
10	1330100	11	Arequipa	0	<a href="https://www.sportanthony.tk/1330100">https://www.sportanthony.tk/1330100</a>
11	1331102	12	Arequipa	0	<a href="https://www.sportanthony.tk/1331102">https://www.sportanthony.tk/1331102</a>
12	1331103	13	Arequipa	0	<a href="https://www.sportanthony.tk/1331103">https://www.sportanthony.tk/1331103</a>
13	1781421	14	Arequipa	0	<a href="https://www.sportanthony.tk/1781421">https://www.sportanthony.tk/1781421</a>
14	1837208	15	Arequipa	0	<a href="https://www.sportanthony.tk/1837208">https://www.sportanthony.tk/1837208</a>
15	z316517989	16	Arequipa	0	<a href="https://www.sportanthony.tk/z316517989">https://www.sportanthony.tk/z316517989</a>

**Figura N° 42: Dataset de Productos/Código/Región/Visitas/Link**

Se hace la importación de Pandas como as y Numpy np, luego se carga el Dataset en formato .csv y en la siguiente línea lo mostramos.

```
In [175]: import numpy as np
import pandas as pd

In [176]: visitas = pd.read_csv("D:/sportanthony.csv")
visitas.head()

Out[176]:
```

	Nombre	Codigo	Region	Visitas	Link
0	Skechers Bounder	1	Arequipa	4	https://www.sportanthony.tk/skechers-bounder
1	Skechers Bounder Arkala	2	Arequipa	3	https://www.sportanthony.tk/skechers-bounder-a...
2	Skechers Terraza Welledge	3	Arequipa	0	https://www.sportanthony.tk/skechers-terraz-a-w...
3	1122623	4	Arequipa	0	https://www.sportanthony.tk/1122623
4	1122828	5	Arequipa	0	https://www.sportanthony.tk/1122828

```
In [ ]: |
```

**Figura N° 43: Importación Numpy, Pandas**

Item\_dict nos crea un diccionario con las columnas Código y Nombre y en la siguiente línea mostramos el producto con el id 250, en ese caso “Skechers Wa vy Lites”, también se ha creado un diccionario 2 con las columnas Código y Link el que también mostramos a modo de prueba, en este caso: “https://www.sportanthony.tk/skechers-wavy-lite”

```
In [172]: visitas.head()

Out[172]:
```

	Nombre	Codigo	Region	Visitas	Link
0	Skechers Bounder	1	Arequipa	4	https://www.sportanthony.tk/skechers-bounder
1	Skechers Bounder Arkala	2	Arequipa	3	https://www.sportanthony.tk/skechers-bounder-a...
2	Skechers Terraza Welledge	3	Arequipa	0	https://www.sportanthony.tk/skechers-terraz-a-w...
3	1122623	4	Arequipa	0	https://www.sportanthony.tk/1122623
4	1122828	5	Arequipa	0	https://www.sportanthony.tk/1122828

```
In [173]: item_dict = dict(zip(visitas['Codigo'], visitas['Nombre']))
item_dict[250]

Out[173]: 'Skechers Wavy Lites'

In [174]: item_dict = dict(zip(visitas['Codigo'], visitas['Link']))
item_dict[250]

Out[174]: 'https://www.sportanthony.tk/skechers-wavy-lites'
```

**Figura N° 44: Creación de los Diccionarios**

En la siguiente línea hemos creado una matriz con el método de pandas pivot table para darle forma a nuestra DataSet de acuerdo al algoritmo con SVD, en este caso

nuestras filas serán las Regiones, nuestras columnas los códigos de los productos y las intersecciones las sumas (Función sum) de las visitas que cada producto ha tenido en determinada Región en el DataSet.

```
In [28]: matriz_visitas_df = visitas.pivot_table(
values= 'Visitas',
index= 'Region',
columns= 'Codigo',
aggfunc="sum"
)
```

**Figura N° 45: Creación Matriz – Pivot Table**

In [29]: matriz\_visitas\_df

Out[29]:

	Codigo	1	2	3	4	5	6	7	8	9	10	...	270	271	272	273	274	275	276	277	278	279	
Region																							
Ancash		0	1	1	0	1	2	1	1	1	2	...	0	1	1	2	2	0	0	0	1	0	
Arequipa		4	3	0	0	0	0	0	0	0	0	...	5	3	6	1	4	0	6	7	4	4	
Cusco		2	0	3	2	2	2	2	2	0	0	...	1	0	1	2	1	0	1	2	2	2	
Ica		0	0	0	0	0	1	1	2	0	0	...	1	2	0	1	0	2	2	1	0	1	
La Libertad		0	1	5	0	4	1	5	5	6	0	...	1	6	5	7	2	4	0	6	3	1	
Lambayeque		0	7	5	0	0	6	4	4	7	10	...	4	3	5	2	4	4	3	0	1	4	
Lima Region		0	1	0	2	1	0	1	2	2	1	...	0	1	2	1	2	2	1	2	0	2	
Loreto		1	2	2	1	0	2	0	1	0	0	...	2	0	2	1	2	2	2	0	0	2	
Piura		0	2	2	0	2	1	0	0	2	0	...	0	1	0	2	0	1	0	2	1	2	
Tacna		1	0	0	0	2	0	0	1	1	0	...	1	2	0	2	0	2	2	1	1	2	

Luego se procede a mostrar la Matriz Visitas Pivot Table que creamos.

**Figura N° 46: Matriz\_visitas**

Con Shape mostramos la cantidad de Regiones y de productos, y luego creamos y mostramos en forma de Arrays, región\_id\_lista y productos\_id\_lista, luego importamos la Matriz escasa coo\_Matriz de Scipy.sparse para SVD

```

In [36]: region_id_lista = np.array(matriz_visitas_df.index.tolist())
region_id_lista
Out[36]: array(['Ancash', 'Arequipa', 'Cusco', 'Ica', 'La Libertad', 'Lambayeque',
'Lima Region', 'Loreto', 'Piura', 'Tacna'], dtype='<U11')

In [37]: productos_id_lista = np.array(matriz_visitas_df.columns.tolist())
productos_id_lista
Out[37]: array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13,
14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26,
27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39,
40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52,
53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65,
66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78,
79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91,
92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104,
105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117,
118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130,
131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143,
144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156,
157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169,
170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182,
183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195,
196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208,
209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221,
222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234,
235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247,
248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260,
261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273,
274, 275, 276, 277, 278, 279])

In [38]: from scipy.sparse import coo_matrix

```

**Figura N° 47: Identificadores de Regiones y de Productos**

Rellenamos con 0 (ceros) los datos nulos de la Matriz visitas\_mtz y luego la convertimos en una matriz escasa (Sparse), al final en este caso Numpy admite valores de tipo float, y nosotros tenemos enteros, entonces los convertimos a float.

```

In [39]: visitas_mtz = matriz_visitas_df.fillna(0).values.copy()
visitas_mtz_sparse = coo_matrix(visitas_mtz)

In [40]: visitas_mtz_sparse
Out[40]: <10x279 sparse matrix of type '<class 'numpy.int64''
with 2039 stored elements in COOrdinate format>

In [41]: visitas_mtz
Out[41]: array([[0, 1, 1, ..., 0, 1, 0],
[4, 3, 0, ..., 7, 4, 4],
[2, 0, 3, ..., 2, 2, 2],
...,
[1, 2, 2, ..., 0, 0, 2],
[0, 2, 2, ..., 2, 1, 2],
[1, 0, 0, ..., 1, 1, 2]], dtype=int64)

In [42]: visitas_mtz_sparse = visitas_mtz_sparse.astype(float)

```

**Figura N° 48: Convertir el formato**

Importamos de Scipy con la función de algebra lineal linalg SVDS, y hacemos la descomposición SVD de Sparse de la matriz original con una diagonal que es un vector S= 5, Filas U=10 y Columnas V=279, luego convertimos ese array diagonal en una Matriz diagonal, luego mostramos con s\_diag la matriz recreada pero rellena con valores latentes producidos por la descomposición SVD.

```
In [181]: from scipy.sparse.linalg import svds

In [182]: U, s, V = svds(visitas_mtz_sparse, k=5)

In [183]: print('Tamaños')
print(f'U: {U.shape}')
print(f's: {s.shape}')
print(f'V: {V.shape}')

Tamaños
U: (10, 5)
s: (5,)
V: (5, 279)

In [187]: # Convertimos el array diagonal a una Matriz Diagonal
s_diag = np.diag(s)

In [188]: s_diag
Out[188]: array([[ 18.37039088,  0.          ,  0.          ,  0.          ,
  0.          ],
 [ 0.          , 24.18086633,  0.          ,  0.          ,
  0.          ],
 [ 0.          ,  0.          , 37.5891361 ,  0.          ,
  0.          ],
 [ 0.          ,  0.          ,  0.          , 45.53388552,
  0.          ],
 [ 0.          ,  0.          ,  0.          ,  0.          ,
 121.18754096]])
```

**Figura N° 49: Mostrar Matriz recreada.**

Lo que nos recomienda para cada Región son los productos donde los valores de las visitas sean las más altas, así obtenemos una matriz con las “puntuaciones” que cada región daría a cada producto y guarda las predicciones por cada región en orden descendente. (Más visitados a menos)

```

In [146]: visitas_svd = U @ s_diag @ V

In [147]: visitas_svd.shape
Out[147]: (10, 279)

In [148]: visitas_svd
Out[148]: array([[ 0.49161813,  0.71552894,  0.94627076, ...,  1.03417181,
                  0.62722741,  1.45096701],
                 [ 3.97697852,  3.00556281, -0.01114274, ...,  7.02707532,
                  3.97927684,  4.00493886],
                 [ 1.84973031,  0.19131071,  2.98657962, ...,  1.96047369,
                  1.90806831,  1.88452381],
                 ...,
                 [ 0.34726207,  1.1958635 ,  0.97489418, ...,  0.73876342,
                  0.48115939,  1.55002186],
                 [ 0.2816565 ,  0.85744766,  0.80004769, ...,  1.02396373,
                  0.52424941,  1.31459595],
                 [ 0.20792072,  0.8302679 ,  0.52921125, ...,  1.02769987,
                  0.40358832,  1.5238797 ]])

```

**Figura N° 50: Mostrar Matriz\_svd**

Tomamos como ejemplo la región Tacna, creamos una matriz donde las regiones coincidan con sus índices, ubicamos que la posición de Tacna en su array es 9, teniendo esto vamos a crear la matriz index\_sort, que serán los valores de los índices de las regiones en la matriz visitas\_svd, y luego mostramos la matriz index\_sort

```

In [149]: id_region = "Tacna"

In [150]: region_index = np.where(region_id_lista == id_region)[0][0]

In [151]: region_index
Out[151]: 9

In [152]: index_sort = visitas_svd[region_index, :].argsort()[::-1]

In [153]: index_sort
Out[153]: array([ 96, 228,  46, 243, 126,  36, 146, 168, 235, 274,  68, 205,  64,
                  54, 157, 110,  45, 132, 163, 145, 259,  48,  74,  38, 272, 250,
                  109, 263, 221, 240, 249, 220, 200, 184, 207, 192,  85, 173, 217,
                  187, 270, 152,  91,  89, 179,  75, 278,  35, 162, 151, 226, 140,
                  149,  42, 238, 193,  33,  24, 239, 246,  49, 160, 237, 219, 188,
                  203, 194, 245,  43,  11, 275,  95, 123, 115, 261, 177, 127,  84,
                  76, 155, 107, 234, 119, 241,  7, 252,  19, 257, 153, 236, 143,
                  27, 141,  71, 258, 268,  53, 198, 195, 175, 233, 244, 206,  79,
                  165, 114, 208,  32,  34, 260, 105,  4, 169,  61, 154,  82, 124,
                  262, 185,  78, 135, 121,  86,  69, 170, 103, 148, 211, 117, 183,
                  90,  28, 215, 265,  17,  60, 171,  8, 199, 191, 276, 248, 167,
                  80, 104,  92, 232, 182, 230, 166, 161, 202, 186, 172,  15,  72,
                  147, 181, 150, 247, 134, 128,  87, 267, 229, 231, 111, 197,  1,
                  12,  99,  37, 209,  5, 255, 144,  21, 227,  63,  57, 273, 213,
                  20, 130,  18,  93, 164, 158,  30, 266,  51, 210, 101, 222,  50,
                  97,  26,  47, 225, 120, 242, 138,  73, 178,  10,  39, 13, 133,
                  100,  58, 269, 113, 174, 216,  88,  77, 136, 271,  70, 176,  65,
                  142, 159, 204,  2, 264, 129, 189,  59,  41,  31, 125, 116,  23,
                  83, 223,  62, 224, 214,  52,  55,  94, 112,  16, 256, 254, 251,
                  102, 156, 277,  25, 137, 190,  6, 131,  67,  22, 106, 196,  3,
                  253, 122,  66, 108,  29,  9,  44,  98,  0,  40, 218, 201,  56,
                  139,  81, 212, 118,  14, 180], dtype=int64)

```

**Figura N° 51: Mostrar Matriz index\_sort**



Se crea una máscara booleana (True/False) con los productos que no fueron visitados, de la matriz `visitas_mtz`, con `región_index` y `index_sort` y los mostramos.

```
In [154]: productos_no_visitados = visitas_mtz[region_index, :] [index_sort] == 0
```

```
In [155]: productos_no_visitados
```

```
Out[155]: array([False, False, False, False, False, False, False, False, False,
                False, False, False, False, False, False, False, False, False,
                False, False, False, False, False, False, False, True, False,
                False, False, False, False, False, False, False, False, False,
                False, False, False, False, False, False, True, False, False,
                False, True, False, False, False, False, False, False, False,
                False, False, False, False, False, False, True, False, False,
                False, False, False, True, False, True, False, False, True,
                False, True, False, False, False, False, False, True, False,
                False, False, False, False, False, True, False, False, False,
                False, False, False, True, False, True, False, False, False,
                False, True, False, False, False, False, True, False, False,
                False, False, False, True, False, False, False, True, False,
                False, True, False, False, False, True, True, True, False,
                True, False, True, True, True, False, False, True, False,
                True, False, False, True, True, False, True, False, False,
                False, True, True, True, False, True, False, True, True,
                True, False, False, True, True, False, False, False, False,
                True, False, False, False, False, True, True, False, False,
                False, False, True, False, True, True, False, True, True,
                False, False, True, False, True, True, False, True, True,
                True, True, True, True, False, True, False, True, True,
                False, True, True, True, True, True, True, True, False, True,
                False, True, True, True, False, True, True, True, True])
```

**Figura N° 52: Matriz True/False productos\_no\_visitados**

Mostramos 5 recomendaciones en forma de array de los productos (productos\_id\_lista), de acuerdo a los productos debidamente ordenados por las recomendaciones por sus índices de la matriz anterior (index\_sort), en este caso nos muestra 5 recomendaciones de los productos con los id: 251, 227, 194, 12, 178.

```

In [156]: rec_index = index_sort[productos_no_visitados]

In [157]: rec_index
Out[157]: array([250, 226, 193, 11, 177, 84, 107, 119, 153, 268, 206, 165, 260,
154, 135, 170, 60, 191, 248, 167, 232, 182, 230, 172, 72, 147,
247, 87, 231, 111, 1, 209, 5, 255, 21, 63, 57, 273, 130,
18, 266, 50, 97, 73, 178, 39, 58, 113, 174, 88, 77, 136,
271, 70, 176, 159, 204, 2, 264, 129, 59, 41, 23, 62, 224,
214, 94, 254, 251, 156, 25, 137, 6, 131, 67, 106, 3, 253,
66, 108, 29, 9, 44, 98, 40, 201, 56, 139, 212, 118, 14,
180], dtype=int64)

In [158]: productos_id_lista
Out[158]: array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13,
14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26,
27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39,
40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52,
53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65,
66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78,
79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91,
92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104,
105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117,
118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130,
131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143,
144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156,
157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169,
170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182,
183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195,
196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208,
209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221,
222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234,
235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247,
248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260,
261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273,
274, 275, 276, 277, 278, 279])

In [159]: productos_id_lista[rec_index][:5]
Out[159]: array([251, 227, 194, 12, 178])

```

**Figura N° 53: Mostrar 5 Id de los productos recomendados**

Finalmente se ha creado la función “recomendar” con parámetros `id_region` y el número de recomendaciones, `region_index` coge la fila de la matriz que corresponde al id de la Región, `index_sort` nos ordena las visitas predichas por las regiones en valor

```
In [160]: def recomendar(id_region, num_recomendaciones=5):
#Cogemos la fila de La Matriz que corresponde a La id de Región
region_index = np.where(region_id_lista == id_region)[0][0]
#Ordenamos Las visitas predichas por Las Regiones en valor descendente
index_sort = visitas_svd[region_index, :].argsort()[::-1]
#Creamos una máscara booleana (True/False) de Los productos que no se han visitados
productos_no_visitas = visitas_mtz[region_index, :][index_sort] == 0
rec_index = index_sort[productos_no_visitados]
rec_ids = productos_id_lista[rec_index]
recomendaciones = rec_ids[:num_recomendaciones]
return recomendaciones
```

**Figura N° 54: Función Recomendar**

#### 3.4.4. Evaluación del modelo

Se hará uso de la función `recomendar` creada y utilizando los diccionarios creados al inicio, `item_dic1` y `item_dict2`, mostrando los nombres de los productos y sus links correspondientes.

```
In [161]: region_id = "Ancash"

In [162]: recomendaciones = recomendar(region_id)
print(recomendaciones)

[221 264 275 234 262]

In [168]: [item_dict[x] for x in recomendaciones]

Out[168]: ['5287203',
'Skechers Skech-Air Ultra Flex',
'Skechers Dynamight 2.0 Quick Concept',
'1837212',
'Skechers Arya']

In [180]: [item_dict2[x] for x in recomendaciones]

Out[180]: ['https://www.sportanthony.tk/5287203',
'https://www.sportanthony.tk/skechers-sketch-air-ultra-flex',
'https://www.sportanthony.tk/skechers-dynamight-quick-concept',
'https://www.sportanthony.tk/1837212',
'https://www.sportanthony.tk/skechers-arya']
```

**Figura N° 55: Recomendaciones Ancash**

```

In [189]: region_id = "Arequipa"

In [191]: recomendaciones = recomendar(region_id)
          print(recomendaciones)

          [190 103 172 140  57]

In [198]: [item_dict[x] for x in recomendaciones]
Out[198]: ['8406200', 'bc1041', '30608312', 'a04485400', 'f35374']

In [199]: [item_dict2[x] for x in recomendaciones]
Out[199]: ['https://www.sportanthony.tk/8406200',
          'https://www.sportanthony.tk/bc1041',
          'https://www.sportanthony.tk/30608312',
          'https://www.sportanthony.tk/a04485400',
          'https://www.sportanthony.tk/f35374']

```

**Figura N° 56: Recomendaciones Arequipa**

```

In [200]: region_id = "Cusco"

In [201]: recomendaciones = recomendar(region_id)
          print(recomendaciones)

          [ 85 126 133 249 205]

In [202]: [item_dict[x] for x in recomendaciones]
Out[202]: ['ee8265', 'aq1701', 'aa7403001', 'Skechers Meridian Best Intent', '62834044']

In [203]: [item_dict2[x] for x in recomendaciones]
Out[203]: ['https://www.sportanthony.tk/ee8265',
          'https://www.sportanthony.tk/aq1701',
          'https://www.sportanthony.tk/aa7403001',
          'https://www.sportanthony.tk/skechers-meridian-best-intent',
          'https://www.sportanthony.tk/62834044']

```

**Figura N° 57: Recomendaciones Cusco**

```

In [204]: region_id = "Ica"

In [205]: recomendaciones = recomendar(region_id)
          print(recomendaciones)

[250 150 251 178  50]

In [206]: [item_dict[x] for x in recomendaciones]
Out[206]: ['Skechers Wavy Lites',
          '631303022',
          'Skechers Dynamight Break Through',
          '19230605',
          'f35627']

In [207]: [item_dict2[x] for x in recomendaciones]
Out[207]: ['https://www.sportanthony.tk/skechers-wavy-lites',
          'https://www.sportanthony.tk/631303022',
          'https://www.sportanthony.tk/skechers-dynamight-break-through',
          'https://www.sportanthony.tk/19230605',
          'https://www.sportanthony.tk/f35627']

```

**Figura N° 58: Recomendaciones Ica**

```

In [208]: region_id = "La Libertad"

In [209]: recomendaciones = recomendar(region_id)
          print(recomendaciones)

[204 141 211 261  26]

In [210]: [item_dict[x] for x in recomendaciones]
Out[210]: ['6292100',
          'a01912006',
          '6281142',
          'Skechers Bobs Sport Squad Glam League',
          'm18555']

In [211]: [item_dict2[x] for x in recomendaciones]
Out[211]: ['https://www.sportanthony.tk/6292100',
          'https://www.sportanthony.tk/a01912006',
          'https://www.sportanthony.tk/6281142',
          'https://www.sportanthony.tk/skechers-bobs-sport-squad-glam-league',
          'https://www.sportanthony.tk/m18555']

```

**Figura N° 59: Recomendaciones La Libertad**

```

In [212]: region_id = "Lambayeque"

In [213]: recomendaciones = recomendar(region_id)
print(recomendaciones)

[202 162 137 279 85]

In [214]: [item_dict[x] for x in recomendaciones]
Out[214]: ['6410102', '36442214', 'aa1636008', "Skechers D'Lites Fresh Start", 'ee8265']

In [215]: [item_dict2[x] for x in recomendaciones]
Out[215]: ['https://www.sportanthony.tk/6410102',
'https://www.sportanthony.tk/36442214',
'https://www.sportanthony.tk/aa1636008',
'https://www.sportanthony.tk/skechers-d-lites-fresh-start',
'https://www.sportanthony.tk/ee8265']

```

**Figura N° 60: Recomendaciones Lambayeque**

```

In [216]: region_id = "Lima Region"

In [217]: recomendaciones = recomendar(region_id)
print(recomendaciones)

[221 188 102 178 276]

In [218]: [item_dict[x] for x in recomendaciones]
Out[218]: ['5287203',
'8407100',
'bd7961',
'19230605',
'Skechers Skech-Air Element Prelude']

In [219]: [item_dict2[x] for x in recomendaciones]
Out[219]: ['https://www.sportanthony.tk/5287203',
'https://www.sportanthony.tk/8407100',
'https://www.sportanthony.tk/bd7961',
'https://www.sportanthony.tk/19230605',
'https://www.sportanthony.tk/skechers-sketch-air-element-prelude']

```

**Figura N° 61: Recomendaciones Lima Región**

```

In [229]: region_id = "Tacna"

In [230]: recomendaciones = recomendar(region_id)
          print(recomendaciones)

[251 227 194 12 178]

In [231]: [item_dict[x] for x in recomendaciones]
Out[231]: ['Skechers Dynamight Break Through',
          '4804102',
          '8387311',
          '1331102',
          '19230605']

In [232]: [item_dict2[x] for x in recomendaciones]
Out[232]: ['https://www.sportanthony.tk/skechers-dynamight-break-through',
          'https://www.sportanthony.tk/4804102',
          'https://www.sportanthony.tk/8387311',
          'https://www.sportanthony.tk/1331102',
          'https://www.sportanthony.tk/19230605']

```

**Figura N° 62: Recomendaciones Tacna**

```

In [220]: region_id = "Loreto"

In [221]: recomendaciones = recomendar(region_id)
          print(recomendaciones)

[169 124 36 227 75]

In [222]: [item_dict[x] for x in recomendaciones]
Out[222]: ['35357204', 'at2506002', 'g26594', '4804102', 'ef0417']

In [223]: [item_dict2[x] for x in recomendaciones]
Out[223]: ['https://www.sportanthony.tk/35357204',
          'https://www.sportanthony.tk/at2506002',
          'https://www.sportanthony.tk/g26594',
          'https://www.sportanthony.tk/4804102',
          'https://www.sportanthony.tk/ef0417']

```

**Figura N° 63: Recomendaciones Loreto**

```

In [225]: region_id = "Piura"

In [226]: recomendaciones = recomendar(region_id)
          print(recomendaciones)
          [180  39 124 220   8]

In [227]: [item_dict[x] for x in recomendaciones]
Out[227]: ['19140401', 'f99256', 'at2506002', '5287210', '11841114']

In [228]: [item_dict2[x] for x in recomendaciones]
Out[228]: ['https://www.sportanthony.tk/19140401',
          'https://www.sportanthony.tk/f99256',
          'https://www.sportanthony.tk/at2506002',
          'https://www.sportanthony.tk/5287210',
          'https://www.sportanthony.tk/11841114']

```

Figura N° 64: Recomendaciones Piura

Ancash	Ica	Lima Región	Tacna	Loreto
5287203	skechers-wavy-lites	5287203	skechers-dynamight-break-through	35357204
skechers-sketch-air-ultra-flex	631303022	8407100	4804102	at2506002
skechers-dinamight-quick-concept	skechers-dynamight-break-through	bd7961	8387311	g26594
1837212	19230605	19230605	1331102	4804102
skechers-arya	f35627	skechers-sketch-air-element-prelude	19230605	ef0417
<b>Modelos más recomendados</b>				
5287203	 5287203-01_2.jpg   5287203-02_2.jpg   5287203-03_2.jpg   5287203-04_2.jpg			
skechers-dynamight-break-through	 skechers-dynamight-break-through-01.jpg   skechers-dynamight-break-through-02.jpg   skechers-dynamight-break-through-03.jpg   skechers-dynamight-break-through-04.jpg   skechers-dynamight-break-through-05.jpg			
19230605	 19230605-01.jpg   19230605-02.jpg   19230605-3.jpg   19230605-4.jpg			
4804102	 4804102-01_2.jpg   4804102-02_2.jpg   4804102-03_2.jpg   4804102-04_2.jpg			

Figura N° 65: Resumen Recomendaciones



## **CAPÍTULO IV**

### **PRUEBA DE HIPÓTESIS**

La contrastación de hipótesis se realizó conforme al Método de Pre Test -Pos Test propuesto, para poder admitir o rechazar la hipótesis. Y en cuanto a la ejecución de este diseño se utilizaron los siguientes indicadores:

**Tabla N° 11: Prueba estadística de los indicadores**

Ítems	Indicadores	n	Criterios	Prueba estadística
1	Grado de satisfacción de los usuarios	6	N menor o igual a 30	Prueba T de Student, diferencia de medias.
2	Calidad de información	6	N menor o igual a 30	Prueba T de Student, diferencia de medias.
3	Tiempos de búsqueda de productos a ofertar	61	N mayor a 30	Prueba estadística Distribución Normal Z, diferencias de medias.
4	Costos por obtención de recomendaciones con el algoritmo de Machine Learning	11	N menor a 30	Prueba T de Student, diferencia de medias.
5	Ventas usando el algoritmo de Machine Learning	61	N mayor a 30	Prueba estadística Distribución Normal Z, diferencias de medias.

Con respecto a lo consistente de las encuestas, se utilizó la confiabilidad de Alfa de Cronbach, obteniendo los resultados entre el rango de fiabilidad (Ver Anexos). Los valores obtenidos están cerca de 1, por lo que se acepta el grado de confiabilidad.

#### **4.1. Indicadores Cualitativos**

Para realizar la contrastación de hipótesis se aplicaron encuestas a los trabajadores(usuarios) internos de la empresa. Tabulados para obtener los resultados que se calculan de acuerdo a los siguientes rangos:

**Tabla N° 12: Valores (Rangos)**

Rango	Niveles	Peso
A-P	Aprobación - Plena	5
A-S	Aprobación - Simple	4
I-N	Indecisión / Indiferencia	3
D-S	Desaprobación - Simple	2
D-P	Desaprobación - Plena	1

Los valores fueron calculados de acuerdo a lo que respondieron de los usuarios, para la ponderación se usó la escala de Likert (1-5). Para cada pregunta y cada una de las que fueran sus respuestas (5) y por cada usuario que se entrevistó, se cuenta la frecuencia con la que ocurren, se calculó luego el total del puntaje y el promedio, en tal sentido se tiene:

$$PT_i = \sum_{j=1}^5 (F_{ij} \times P_j) \dots\dots\dots (1)$$

Donde:

$PT_j$  = Puntaje total de la pregunta  $i$  - esima

$F_{ij}$  = Frecuencia  $j$  - esima de la Pregunta  $i$  - esima

$P_j$  = Peso  $j$  - esima

Y para calcular el promedio ponderado de cada pregunta:

$$\overline{PP}_i = \frac{PT_i}{n} \dots\dots\dots (2)$$

$\overline{PP}_i$  = Promedio de Puntaje Total de la pregunta  $i$ -esima.

$n$  = 6 número de usuarios del sistema.

**4.1.1. Grado de satisfacción De Los Usuarios - Cálculo.**

En la tabla N° 12 vemos la ponderación en cuanto a los criterios de evaluación del indicador cualitativo: “Grado de satisfacción de los usuarios”

**Tabla N° 13: Ponderación de Indicador “Grado de satisfacción de los usuarios”**

N°	Pregunta	PESO					Puntaje Total	Puntaje Promedio
		AP	AS	IN	DS	DP	$PT_i$	$\overline{PP}_j$
		5	4	3	2	1		
1	¿Es fácil el procedimiento de búsqueda de productos sin un algoritmo?	P(1-5)	P(1-4)	P(1-3)	P(1-2)	P(1-1)	$\sum_{j=1}^5 (F_{1j} \times P_j)$	$\frac{PT_1}{n}$
2	¿Se encuentra satisfecho con la variedad de productos a ofertar sin un algoritmo?	P(2-5)	P(2-4)	P(2-3)	P(2-2)	P(2-1)	$\sum_{j=1}^5 (F_{2j} \times P_j)$	$\frac{PT_2}{n}$
3	¿Es amigable el procedimiento de búsqueda de productos a ofertar sin un algoritmo?	P(3-5)	P(3-4)	P(3-3)	P(3-2)	P(3-1)	$\sum_{j=1}^5 (F_{3j} \times P_j)$	$\frac{PT_3}{n}$
4	¿La información sin un recomendador le ayuda a tener una buena visión de los productos a ofertar?	P(4-5)	P(4-4)	P(4-3)	P(4-2)	P(4-1)	$\sum_{j=1}^5 (F_{4j} \times P_j)$	$\frac{PT_4}{n}$
5	¿Cómo calificaría la flexibilidad del procedimiento sin un algoritmo al momento de busca productos?	P(5-5)	P(5-4)	P(5-3)	P(5-2)	P(5-1)	$\sum_{j=1}^5 (F_{5j} \times P_j)$	$\frac{PT_5}{n}$
6	¿Las recomendaciones se entregan en tiempos favorables al área de ventas?	P(6-5)	P(6-4)	P(6-3)	P(6-2)	P(6-1)	$\sum_{j=1}^5 (F_{6j} \times P_j)$	$\frac{PT_6}{n}$

7	¿Cómo califica la manera de obtener las recomendaciones de productos sin un algoritmo?	P(7-5)	P(7-4)	P(7-3)	P(7-2)	P(7-1)	$\sum_{j=1}^5 (F_{7j} \times P_j)$	$\frac{PT_7}{n}$
8	¿Cómo califica usted la calidad de las recomendaciones obtenidas sin un Algoritmo?	P(8-5)	P(8-4)	P(8-3)	P(8-2)	P(8-1)	$\sum_{j=1}^5 (F_{8j} \times P_j)$	$\frac{PT_8}{n}$
9	¿Las recomendaciones sin un algoritmo son fáciles de analizar?	P(9-5)	P9-4	P9-3	P9-2	P9-1	$\sum_{j=1}^5 (F_{9j} \times P_j)$	$\frac{PT_9}{n}$
10	¿Considera que las recomendaciones sin un algoritmo son útiles para el área de ventas?	(10-5)	(10-4)	(10-3)	(10-2)	(10-1)	$\sum_{j=1}^5 (F_{10j} \times P_j)$	$\frac{PT_{10}}{n}$

En la tabla N° 12 observamos las ponderaciones de los criterios de evaluación del indicador cualitativo “Grado de satisfacción del usuario” con los resultados de las encuestas realizadas.

	N°	Pregunta	AP	AS	IN	DS	DP	Puntaje Total	Puntaje Promedio
			5	4	3	2	1		
FUNCIONALIDAD	1	¿Es fácil el procedimiento de búsqueda de productos sin un algoritmo?	1	2	2	1		21	3,5
	2	¿Se encuentra satisfecho con la variedad de productos a ofertar sin un algoritmo?		2	3	1		16	2,7
	3	¿Es amigable el procedimiento de búsqueda de productos a ofertar sin un algoritmo?		2	3	1		16	2,7

	4	¿La información sin un recomendador le ayuda a tener una buena visión de los productos a ofertar?			4	1	1	15	2,5
	5	¿Cómo calificaría la flexibilidad del procedimiento sin un algoritmo al momento de busca productos?			3	2	1	14	2,3
SERVICIO Y CALIDAD	1	¿Las recomendaciones se entregan en tiempos favorables al área de ventas?	1	1	4			21	3,5
	7	¿Cómo califica la manera de obtener las recomendaciones de productos sin un algoritmo?		1	4	1		18	3
	8	¿Cómo califica usted la calidad de las recomendaciones obtenidas sin un Algoritmo?		1	4	1		18	3
	9	¿Las recomendaciones sin un algoritmo son fáciles de analizar?			4	2		16	2,7
	10	¿Considera que las recomendaciones sin un algoritmo son útiles para el área de ventas?		3	3			21	3,5

**Tabla N° 14: Tabulación - Grado de satisfacción de los usuarios. (Pre Test).**

**Grado de satisfacción de los usuarios con el Recomendador propuesto - Cálculo:**

A los usuarios del sistema se le aplicaron las encuestas mostradas y se tabularon de para que se calculen los resultados logrados de acuerdo a la tabla N° 14. Todas las respuestas que se aplicaron tienen un peso. Luego se halla el puntaje promedio de cada criterio usado por cada indicador y al final se hallar el puntaje total de todos los indicadores con las fórmulas que se emplearon.

En la tabla N°14 se muestra lo obtenido en la encuesta que se aplicó para determinar el grado de satisfacción de los usuarios con el recomendador que se propone.

**Tabla N° 15: Tabulación – Grado de satisfacción de los usuarios. (Post Test)**

N°	Pregunta	AP	AS	I N	DS	DP	Puntaje Total	Puntaje Promedio	
		5	4	3	2	1			
FUNCIONALIDAD	1	¿Es de fácil de uso el funcionamiento del algoritmo?	3	3				24	4,5
	2	¿Se encuentra satisfecho con la variedad de recomendaciones que brinda el algoritmo?	2	2	2			24	4
	3	¿La interfaz donde se muestran las recomendaciones es amigable para usted?	2	2	1	1		23	3,8
	4	¿La información del recomendador ayuda a tener una mejor visión de los productos a ofertar?	4	1	1			27	4,5
	5	¿Cómo califica la flexibilidad del recomendador al momento de realizar la emisión de las mismas?	2	2	2			24	4
SERVICIO Y CALIDAD	6	¿Las recomendaciones se entregan en tiempos favorables al área de ventas?	2	2	1	1		23	3,8
	7	¿Cómo considera la manera de obtener las recomendaciones de productos?	3	3				24	4,5
	8	¿Cómo califica usted la calidad de las recomendaciones obtenidas?	4	1	1			27	4,5
	9	¿Las recomendaciones de los productos son fáciles de analizar?	4	2				28	4,67
	10	¿Considera que las recomendaciones son útiles para el área de ventas?	5	1				29	4,8

**Tabla N° 16: Contratación - Resultados Pre & Post Test**

	Pregunta	PRE TEST	POST TEST	$D_i$	$D_i^2$
		$GSUA_i$	$GSUP_i$		
<b>FUNCIONALIDAD</b>	1	3,5	4,5	-1	1
	2	2,7	4	-1,3	1,69
	3	2,7	3,8	-1,1	1,21
	4	2,5	4,5	-2	4
	5	2,3	4	-1,7	2,89
<b>SERVICIO Y CALIDAD</b>	1	3,5	3,8	-0,3	0,09
	2	3	4,5	-1,5	2,25
	3	3	4,5	-1,5	2,25
	4	2,7	4,67	-1,97	3,8809
	5	3,5	4,8	-1,3	1,69
			$\Sigma$	-13,67	20,95

**Donde:**

$GSUA$ : Grado de satisfacción del usuario sin el Algoritmo Recomendador

$GSUP$ : Grado de satisfacción del usuario con el Algoritmo Recomendador

Prop.

**Prueba de la Hipótesis para:**

**Grado de satisfacción de los usuarios**

**A. Definición de Variables**

$GSUA$ : Grado de satisfacción de los usuarios sin el Algoritmo Recomendador

$GSUP$ : Grado de satisfacción de los usuarios con el Algoritmo Recomendador

Propuesto.



## B. Hipótesis Estadísticas

- **Hipótesis  $H_0$ :** El grado de satisfacción de los usuarios sin el Algoritmo Recomendador es mayor o igual que el grado de satisfacción de los usuarios *con el Algoritmo Recomendador Propuesto*.

$$H_0: GSU_A - GSU_P \geq 0$$

- **Hipótesis  $H_a$ :** El grado de satisfacción de los usuarios sin el Algoritmo Recomendador es menor que el grado de satisfacción de los usuarios *con el Algoritmo Recomendador Propuesto*.

$$H_a: GSU_A - GSU_P < 0$$

## C. Nivel de Significancia

La significancia ( $\alpha$ ) elegida para la prueba de hipótesis es de 5%.

Con ( $\alpha$ ) = 0.05 (nivel de significancia) y n-1 (10-1) = 9 grados de libertad, tenemos el valor crítico de T de Student:

**Valor Crítico:**

$$t_{\alpha = 0.05} = -1.833$$

Como  $\alpha = 0.05$  y n-1 = 10-1 = 9 grados de libertad, la zona de rechazo son los valores de **t** que son menores a:

$$-t_{0.05} = -1.833$$

## D. Resultados de la Hipótesis Estadística

### a) Diferencia Promedio

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n} \dots\dots\dots (3)$$

$$\bar{D} = \frac{-13,67}{10}$$

$$\bar{D} = -1,367$$

### b) Desviación Estándar

$$S_{D^2} = \frac{n \sum_{i=1}^n D_i^2 - [\sum_{i=1}^n D_i]^2}{n(n-1)} \dots\dots\dots (4)$$

$$S_{D^2} = \frac{10(20,95) - (-13,67)^2}{10(10-1)}$$

$$S_{D^2} = 0,25$$

### c) Calculando T:

$$t = \frac{\bar{D}\sqrt{n}}{\sqrt{S_D}}$$

$$t = \frac{(-1,367)\sqrt{10}}{\sqrt{0,5}}$$

$$t = -2,56$$

### d) Conclusión

Dado que:  $t_c = -2,56$  ( $t_{\text{calculado}}$ )  $< t_{\alpha} = -1,833$  ( $t_{\text{tabular}}$ ), y ubicándose este valor en la región de rechazo, concluyendo que:  $GSU_A - GSU_P < 0$ , se rechaza  $H_0$  y se acepta  $H_a$ , entonces se prueba la validez de la hipótesis con un nivel de error de 5% ( $\alpha = 0.05$ ), siendo el Recomendador una posible solución para el problema de la presente investigación.

En la figura se observa la zona de aceptación y rechazo para la prueba de la hipótesis Grado de satisfacción de los usuarios.

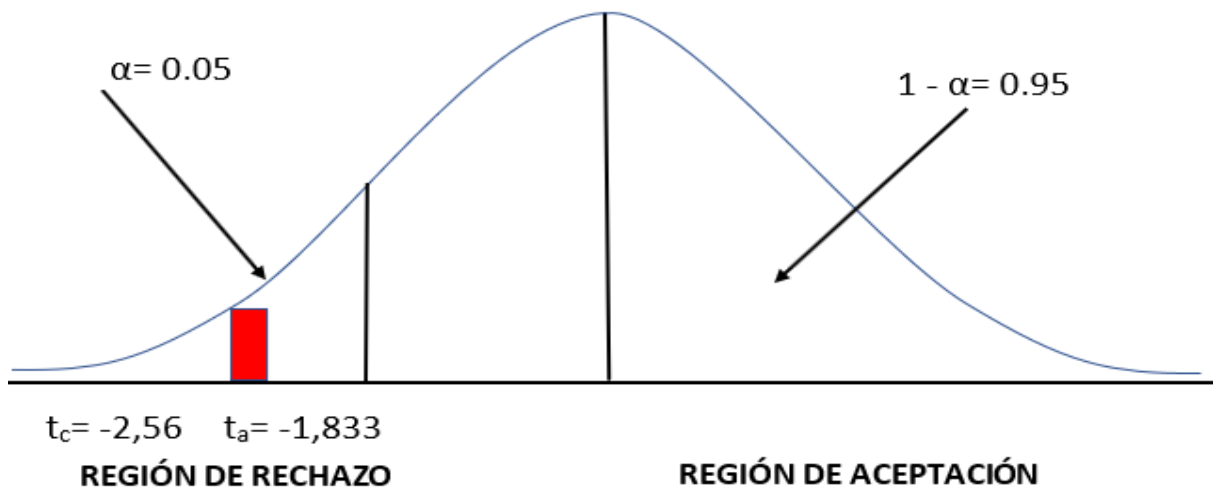


Figura N° 66: Aceptación y Rechazo “Región” – Grado de satisfacción de los usuarios.

#### *4.1.2. Calidad de las recomendaciones de los productos a ofertar con el algoritmo*

##### *Machine Learning – Cálculo.*

En la tabla N° 16 se pueden apreciar los criterios de ponderación que se evalúan para: “Calidad de la información obtenida con las recomendaciones del algoritmo Machine Learning”

**Tabla N° 17: Ponderación de Indicador: “Calidad de la información obtenida con las recomendaciones del algoritmo Machine Learning”**

N°	Pregunta	PESO					Puntaje Total	Puntaje Promedio
		AP	AS	IN	DS	DP		
		5	4	3	2	1	$PT_i$	$\overline{PP}_j$
1	¿Obtiene recomendaciones relevantes para la toma de decisiones?	(1-5)	(1-4)	(1-3)	(1-2)	(1-1)	$\sum_{j=1}^5 (F_{1j} \times P_j)$	$\frac{PT_1}{n}$
2	¿Considera que las recomendaciones son considerables detalles para la toma de decisiones?	(2-5)	(2-4)	(2-3)	(2-2)	(2-1)	$\sum_{j=1}^5 (F_{2j} \times P_j)$	$\frac{PT_2}{n}$
3	¿Considera que hay exactitud en las recomendaciones que brinda el Algoritmo recomendador?	(3-5)	(3-4)	(3-3)	(3-2)	(3-1)	$\sum_{j=1}^5 (F_{3j} \times P_j)$	$\frac{PT_3}{n}$
4	¿Considera que las recomendaciones que muestra el Recomendador son datos actualizados?	(4-5)	(4-4)	(4-3)	(4-2)	(4-1)	$\sum_{j=1}^5 (F_{4j} \times P_j)$	$\frac{PT_4}{n}$
5	¿Considera que las recomendaciones ofrecidas son fáciles de comprender?	(5-5)	(5-4)	(5-3)	(5-2)	(5-1)	$\sum_{j=1}^5 (F_{5j} \times P_j)$	$\frac{PT_5}{n}$
6	¿Considera que las recomendaciones que se le ofrece influyen en la toma de decisiones?	(6-5)	(6-4)	(6-3)	(6-2)	(6-1)	$\sum_{j=1}^5 (F_{6j} \times P_j)$	$\frac{PT_6}{n}$
7	¿Cómo considera la manera de obtener las recomendaciones de productos?	(7-5)	(7-4)	(7-3)	(7-2)	(7-1)	$\sum_{j=1}^5 (F_{7j} \times P_j)$	$\frac{PT_7}{n}$
8	¿Cómo califica usted la calidad de las recomendaciones obtenidas con el Algoritmo?	(8-5)	(8-4)	(8-3)	(8-2)	(8-1)	$\sum_{j=1}^5 (F_{8j} \times P_j)$	$\frac{PT_8}{n}$

En la tabla N° 17 observamos las ponderaciones de los criterios de evaluación para el indicador cualitativo “Calidad de la información obtenida con las recomendaciones del algoritmo Machine Learning” con lo obtenido en las encuestas que se realizaron.

**Tabla N° 18: Tabulación para la Calidad de la información (Pre Test) Obtenida con las recomendaciones del algoritmo Machine Learning.**

N°	Pregunta	AP	AS	IN	DS	DP	Puntaje Total	Puntaje Promedio
		5	4	3	2	1		
1	¿Se obtiene recomendaciones relevantes para la toma de decisiones?		3	2	1		12+6+2=20	3,3
2	¿Considera que las recomendaciones ofrecen considerables detalles para la toma de decisiones?		1	4	1		4+12+2=18	3
3	¿Considera que hay exactitud en las recomendaciones que brinda el Algoritmo recomendador?			5	1		15+2=17	2,8
4	¿Considera que las recomendaciones que muestra el Recomendador son datos actualizados?			4	1	1	12+2+1=15	2,5
5	¿Considera que las recomendaciones ofrecidas son fáciles de comprender?		3	3			12+9=21	3,5
6	¿Considera que las recomendaciones que se le ofrece influyen en la toma de decisiones?		2	3	1		8+9+2=19	3,2
7	¿Cómo considera la manera de obtener las recomendaciones de productos?		1	4	1		4+12+2=18	3
8	¿Cómo califica usted la calidad de las recomendaciones obtenidas con el Algoritmo?		2	4			8+12+=20	3,3

**Calidad de la información obtenida con las recomendaciones del algoritmo Machine  
propuesto - Cálculo:**

Las encuestas que se muestran fueron aplicadas a los usuarios del sistema, estas fueron tabuladas de forma que se calculen los resultados logrados de acuerdo a la tabla N° 17.

A toda respuesta de las encuestas que se aplicaron le pertenece un peso. De esta forma se pasa a calcular el puntaje promedio de cada criterio usado por cada indicador, por último, se halla el puntaje total para cada indicador con las fórmulas empleadas.

En la tabla N° 18 se muestra lo obtenido de la encuesta que se aplicó para conocer el grado de satisfacción de los usuarios con el recomendador propuesto.

**Tabla N° 19: Tabulación para la Calidad de la información (Post Test) Obtenida con las recomendaciones del algoritmo Machine Learning.**

N°	Pregunta	AP	AS	IN	DS	DP	Puntaje Total	Puntaje Promedio
		5	4	3	2	1		
1	¿Obtiene recomendaciones relevantes para la toma de decisiones?	2	3	1			10+12+3=25	4.2
2	¿Considera que las recomendaciones ofrecen considerables detalles para la toma de decisiones?	3	2	1			15+8+3=26	4,3
3	¿Considera que hay exactitud en las recomendaciones que brinda el Algoritmo recomendador?		1	5			4+15=19	3,2
4	¿Considera que las recomendaciones que muestra el Recomendador son datos actualizados?	1	2	3			4+6+9=19	3,2

5	¿Considera que las recomendaciones ofrecidas son fáciles de comprender?	5	1				25+4=29	4,8
6	¿Considera que las recomendaciones que se le ofrece influyen en la toma de decisiones?	5	1				25+4=29	3,2
7	¿Cómo considera la manera de obtener las recomendaciones de productos?		4	1	1		16+3+2=21	3,5
8	¿Cómo califica usted la calidad de las recomendaciones obtenidas con el Algoritmo?	1	4	1			5+16+3=24	4

**Tabla N° 20: Contratación - Resultados Pre & Post Test**

Pregunta	PRE TEST	POST TEST	$D_i$	$D_I^2$
	$CIA_i$	$CIP_i$		
1	3,3	4,2	-0,9	0,81
2	3	4,3	-1,3	1,69
3	2,8	3,2	-0,4	0,16
4	2,5	3,2	-0,7	0,49
5	3,5	4,8	-1,3	1,69
6	3,2	3,2	0	0
7	3	3,5	-0,5	0,25
8	3,3	4	-0,7	0,49
		$\Sigma$	-5,8	5,58

**Donde:**

*CIA:* Calidad de la Información sin el Algoritmo Recomendador

*CIP:* Calidad de la Información sin el Algoritmo Recomendador Propuesto.

## Prueba de la Hipótesis para el Indicador Cualitativo:

### Calidad de la información obtenida con las recomendaciones del algoritmo Machine Learning

#### E. Definición de Variables

*CI<sub>A</sub>*: Calidad de la Información sin el Algoritmo Recomendador

*CI<sub>P</sub>*: Calidad de la Información con el Algoritmo Recomendador Propuesto.

#### F. Hipótesis Estadísticas

- **Hipótesis  $H_0$** : La calidad de Información obtenida sin el Algoritmo Recomendador es mayor o igual a la calidad de Información obtenida con el Algoritmo *Recomendador Propuesto*.

$$H_0: CI_A - CI_P \geq 0$$

- **Hipótesis  $H_a$** : La calidad de Información obtenida sin el Algoritmo Recomendador es menor a la calidad de Información obtenida con el Algoritmo *Recomendador Propuesto*.

$$H_a: CI_A - CI_P < 0$$

#### G. Nivel de Significancia

La significancia ( $\alpha$ ) elegida para la prueba de hipótesis es del 5%.

Siendo ( $\alpha$ ) = 0.05 (nivel de significancia) y n-1 (8-1) = 7 grados de libertad, se obtiene el valor crítico de T de Student:

#### Valor Crítico:

$$t_{\alpha = 0.05} = - 1.8946$$



Como  $\alpha = 0.05$  y  $n-1 = 8-1 = 7$  grados de libertad, la zona de rechazo consiste en aquellos valores de  $t$  menores que:

$$-t_{0.05} = -1.8946$$

## H. Resultados de la Hipótesis Estadística

### a) Diferencia Promedio

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n} \dots\dots\dots (3)$$

$$\bar{D} = \frac{-5,8}{8}$$

$$\bar{D} = -0,725$$

### b) Desviación Estándar

$$S_{D^2} = \frac{n \sum_{i=1}^n D_i^2 - [\sum_{i=1}^n D_i]^2}{n(n-1)} \dots\dots\dots (4)$$

$$S_{D^2} = \frac{8(5,58) - (-5,8)^2}{8(8-1)}$$

$$S_{D^2} = 0,2$$

### c) Calculando T:

$$t = \frac{\bar{D}\sqrt{n}}{\sqrt{S_D}}$$

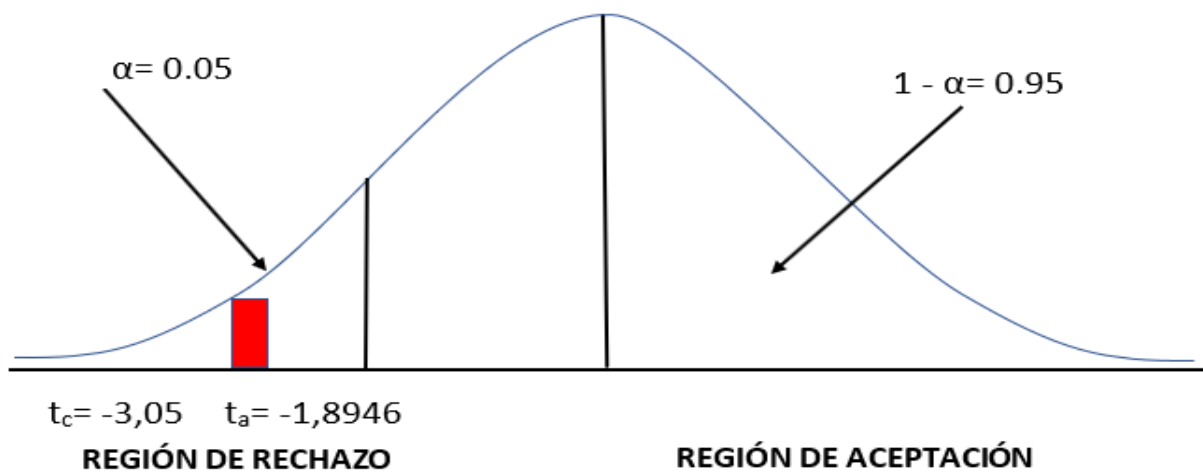
$$t = \frac{(-0,725)\sqrt{8}}{\sqrt{0,48}}$$

$$t = -3,05$$

#### d) Conclusión

Puesto que:  $t_c = -3,05$  ( $t_{\text{calculado}}$ )  $< t_{\alpha} = -1,8946$  ( $t_{\text{tabular}}$ ), y estando este valor dentro de la región de rechazo, se concluye que  $CI_A - CI_P < 0$ , se rechaza  $H_0$  y  $H_a$  es aceptada, por lo tanto, se prueba la validez de la hipótesis con un nivel de error de 5% ( $\alpha = 0,05$ ), siendo el Recomendador una alternativa de solución para el problema de la presente investigación.

En la siguiente imagen se aprecia la zona de aceptación y rechazo para la prueba de la hipótesis Calidad de la información obtenida con las recomendaciones del algoritmo Machine Learning.



**Figura N° 67:** Aceptación y Rechazo “Región” – Calidad de la información obtenida con las recomendaciones del algoritmo Machine Learning.

## 4.2. Indicadores Cuantitativos

### 4.2.1. Indicador cuantitativo – Prueba de Hipótesis:

**Tiempos de búsqueda de productos a ofertar**

## A. Definición de Variables

TBP<sub>A</sub>: Tiempos de búsqueda de productos a ofertar sin el Algoritmo Recomendador. (Minutos).

TBP<sub>P</sub>: Tiempos de búsqueda de productos a ofertar con el Algoritmo Recomendador propuesto (Minutos).

## B. Hipótesis Estadísticas

- **Hipótesis  $H_0$ :** Los tiempos de búsquedas de productos a ofertar sin hacer uso del Algoritmo Recomendador son menor o igual que los tiempos de búsqueda de productos a ofertar haciendo uso del Algoritmo Recomendador *Propuesto*.

$$H_0: TBP_A - TBP_P \leq 0$$

- **Hipótesis  $H_a$ :** Los tiempos de búsquedas de productos a ofertar sin hacer uso del Algoritmo Recomendador son mayor que los tiempos de búsqueda de productos a ofertar haciendo uso del Algoritmo Recomendador *Propuesto*.

$$H_a: TPB_A - TPB_P > 0$$

## C. Nivel de Significancia

Al usar un nivel de significancia del 5% ( $\alpha = 0.05$ ) entonces el nivel de confianza será del 95% ( $1 - \alpha = 0.95$ ).

## D. Estadígrafo de contraste

Como  $n = 61$  es un valor grande, se usa la distribución normal (Z):

**a. Promedio:**

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \dots\dots\dots(5)$$

**b. Calculando la Varianza:**

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \dots\dots\dots(6)$$

**c. Calculando Z:**

$$Z_c = \frac{(\bar{X}_A - \bar{X}_P)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_P^2}{n_P}}} \dots\dots\dots(7)$$

**E. Resultados de la Estadística:**

Para hacer el cálculo del tiempo Promedio para las búsquedas de productos a ofertar se estimó una muestra de 61 tiempos.

**Tabla N° 21: Toma de Tiempos de búsquedas de productos a Ofertar:**

N°	PRE-TEST (Min.)	POST-TEST (Min.)	PRE-TEST (Min.)	POST-TEST (Min.)	PRE-TEST (Min.)	POST-TEST (Min.)
	$TGR_A$	$TGR_P$	$TGR_A - \overline{TGR_A}$	$TGR_P - \overline{TGR_P}$	$(TGR_A - \overline{TGR_A})^2$	$(TGR_P - \overline{TGR_P})^2$
1	129	6	2,77	-1,72	7,67	2,96
2	121	5	-5,23	-2,72	27,35	7,40
3	124	11	-2,23	3,28	4,97	10,76
4	116	10	-10,23	2,28	104,65	5,20
5	112	10	-14,23	2,28	202,49	5,20
6	147	4	20,77	-3,72	431,39	13,84

7	144	4	17,77	-3,72	315,77	13,84
8	120	7	-6,23	-0,72	38,81	0,52
9	117	10	-9,23	2,28	85,19	5,20
10	133	9	6,77	1,28	45,83	1,64
11	106	10	-20,23	2,28	409,25	5,20
12	132	11	5,77	3,28	33,29	10,76
13	129	4	2,77	-3,72	7,67	13,84
14	104	10	-22,23	2,28	494,17	5,20
15	121	8	-5,23	0,28	27,35	0,08
16	110	9	-16,23	1,28	263,41	1,64
17	128	8	1,77	0,28	3,13	0,08
18	100	6	-26,23	-1,72	688,01	2,96
19	120	5	-6,23	-2,72	38,81	7,40
20	150	10	23,77	2,28	565,01	5,20
21	101	7	-25,23	-0,72	636,55	0,52
22	139	5	12,77	-2,72	163,07	7,40
23	105	8	-21,23	0,28	450,71	0,08
24	143	6	16,77	-1,72	281,23	2,96
25	141	8	14,77	0,28	218,15	0,08
26	136	5	9,77	-2,72	95,45	7,40
27	139	6	12,77	-1,72	163,07	2,96
28	146	6	19,77	-1,72	390,85	2,96
29	142	6	15,77	-1,72	248,69	2,96
30	118	7	-8,23	-0,72	67,73	0,52
31	107	8	-19,23	0,28	369,79	0,08
32	114	8	-12,23	0,28	149,57	0,08
33	145	6	18,77	-1,72	352,31	2,96
34	130	8	3,77	0,28	14,21	0,08
35	116	8	-10,23	0,28	104,65	0,08
36	134	10	7,77	2,28	60,37	5,20
37	136	9	9,77	1,28	95,45	1,64
38	109	10	-17,23	2,28	296,87	5,20
39	101	8	-25,23	0,28	636,55	0,08
40	103	6	-23,23	-1,72	539,63	2,96
41	115	9	-11,23	1,28	126,11	1,64

42	128	10	1,77	2,28	3,13	5,20
43	148	5	21,77	-2,72	473,93	7,40
44	144	6	17,77	-1,72	315,77	2,96
45	116	9	-10,23	1,28	104,65	1,64
46	110	8	-16,23	0,28	263,41	0,08
47	131	10	4,77	2,28	22,75	5,20
48	146	8	19,77	0,28	390,85	0,08
49	140	5	13,77	-2,72	189,61	7,40
50	132	6	5,77	-1,72	33,29	2,96
51	132	11	5,77	3,28	33,29	10,76
52	147	9	20,77	1,28	431,39	1,64
53	137	10	10,77	2,28	115,99	5,20
54	112	6	-14,23	-1,72	202,49	2,96
55	133	4	6,77	-3,72	45,83	13,84
56	107	8	-19,23	0,28	369,79	0,08
57	131	10	4,77	2,28	22,75	5,20
58	107	12	-19,23	4,28	369,79	18,32
59	131	8	4,77	0,28	22,75	0,08
60	136	10	9,77	2,28	95,45	5,20
61	146	5	22,77	-2,72	518,47	7,40
Total	7697	471			13280,79	270,26

**a. Promedio:**

$$\overline{TGR_A} = \frac{7697}{61}$$

$$\overline{TGR_A} = 126,18$$

$$\overline{TGR_P} = \frac{471}{61}$$

$$\overline{TGR_P} = 7,72$$

**b. Varianza:**

$$\sigma^2_A = \frac{13280,79}{61}$$

$$\sigma^2_A = 217,71$$

$$\sigma^2_P = \frac{270,26}{61}$$

$$\sigma^2_P = 4,43$$

**c. Cálculo de Z:**

$$Z_c = \frac{(126,18 - 7,72)}{\sqrt{\frac{217,71}{61} + \frac{4,43}{61}}}$$

$$Z_c = \frac{118,46}{\sqrt{(3,57 + 0,07)}}$$

$$Z_c = \frac{118,46}{\sqrt{(3,64)}}$$

$$Z_c = \frac{118,46}{1,91}$$

$$Z_c = 62,02$$

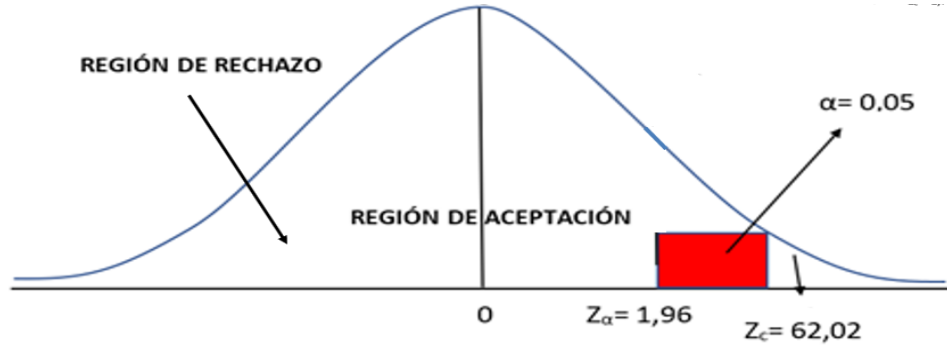
**d. Zona Crítica:**

Para  $\alpha = 0.05$ , en la tabla de Distribución Normal se encuentra a  $Z_{\alpha} = 1.96$ . Por lo tanto, la zona crítica de la prueba es:

$$Z_c = \langle 1.96, \alpha \rangle.$$

**e. Conclusiones:**

Apreciamos en la imagen la zona de aceptación y rechazo para la prueba de hipótesis Tiempos de búsqueda de productos a ofertar.



**Figura N° 68: Aceptación y Rechazo “Región” – Tiempos de búsqueda de productos a ofertar.**

Teniendo que  $Z_c = 62,02$  calculado es mayor que  $Z_{\alpha} = 1,96$  y encontrándose ese valor en la zona de rechazo  $Z_c = <1,96, \alpha >.$ , rechazamos  $H_0$  y  $H_a$  es aceptada. Por lo que podemos concluir que los tiempos de búsqueda de productos a ofertar son menores al hacer uso del Algoritmo Recomendador Machine Learning.

**4.2.2. Indicador cuantitativo – Prueba de Hipótesis:**

**Costos por el uso del algoritmo recomendador Machine Learning**

**Tabla N° 22: Contrastación de los Costos por el uso del algoritmo recomendador Machine Learning.**

N°	Otros Recomendadores	USD/ Mes	Otro Recomendador (SOLES)	Algoritmo Recomendador ML (SOLES)	$D_i$	$D_I^2$
			$OR_i$	$ARP_i$		
1	ActiveCampaign	9	196,2	54.4	141,8	20107,24
2	Drip Pricing	19	234,2	54.4	179,8	32328,04
3	Yusp	49	348,2	54.4	293,8	86318,44



4	AI Product Recommendations	24	253,2	54.4	198,8	39521,44	
5	Product recommendation	12	207,6	54.4	153,2	23470,24	
6	Prestashop Recommended-Products	32	283,6	54.4	229,2	52532,64	
7	Opencart Recommended Product	20	238	54.4	183,6	33708,96	
8	Magento Recommended Products	27	264,6	54.4	210,2	44184,04	
9	Evolve Commerce	29	272,2	54.4	217,8	47436,84	
10	Amazon Native Shopping	27	264,6	54.4	210,2	44184,04	
11	Role Based Up	25	257	54.4	202,6	41046,76	
					Σ	2221	464838,68

**Donde:**

*OR<sub>i</sub>: Otro Recomendador*

*ARP<sub>i</sub>: Algoritmo Recomendador Propuesto*

**Prueba de la Hipótesis para el Indicador Cuantitativo:**

*Costos por el uso del Algoritmo Recomendador Machine Learning*

**A. Definición de Variables:**

*COR<sub>i</sub>: Costos con otro Recomendador*

*CARP<sub>i</sub>: Costos con algoritmo Recomendador Propuesto*

**B. Hipótesis Estadísticas:**

- **Hipótesis  $H_0$ :** El costo por el uso de Otro Recomendador es mayor o igual que el costo por el uso del *Algoritmo Recomendador Propuesto*

$$H_0: COR_i - CARP_i \geq 0$$

- **Hipótesis  $H_a$ :** El costo por el uso de Otro Recomendador es menor que el costo por el uso del *Algoritmo Recomendador Propuesto*.

$$H_a: COR_i - CARP_i < 0$$

### C. Nivel de Significancia

La significancia ( $\alpha$ ) elegida en la prueba de hipótesis es del 5%.

Siendo ( $\alpha$ ) = 0.05 (nivel de significancia) y n-1 (11-1) = 10 grados de libertad, teniendo el valor crítico de T de Student:

**Valor Crítico:**

$$t_{\alpha = 0.05} = 1.812$$

Como  $\alpha = 0.05$  y n-1 = 11-1 = 10 grados de libertad, la zona de rechazo son los valores de **t** menores que:

$$-t_{0.05} = -1.812$$

### D. Resultados de la Hipótesis Estadística

#### a) Diferencia Promedio

$$\bar{D} = \frac{\sum_{i=1}^n D_i}{n} \dots\dots\dots (8)$$

$$\bar{D} = \frac{2221}{11}$$

$$\bar{D} = 201,9$$

**b) Desviación Estándar**

$$S_{D^2} = \frac{n \sum_{i=1}^n D_i^2 - [\sum_{i=1}^n D_i]^2}{n(n-1)} \dots\dots\dots (9)$$

$$S_{D^2} = \frac{11(464838,68) - (2221)^2}{11(11-1)}$$

$$S_{D^2} = 180.38$$

**c) Cálculo de T:**

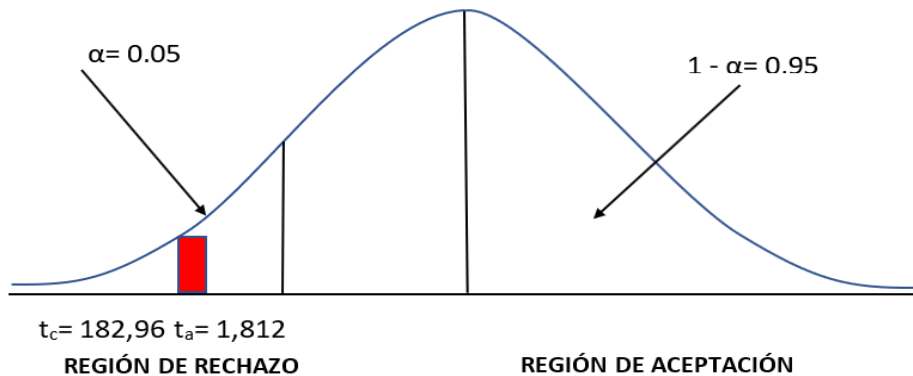
$$t = \frac{\bar{D}\sqrt{n}}{\sqrt{S_D}}$$

$$t = \frac{(201,9)\sqrt{11}}{\sqrt{13,43}}$$

$$t = 182,96$$

**d) Conclusión**

Dado que:  $t_c = 182,96$  ( $t_{\text{calculado}}$ )  $\geq t_{\alpha} = 1.812$  ( $t_{\text{tabular}}$ ), y encontrándose ese valor en de la zona de Rechazo, se concluye que  $COR - CARP \geq 0$ , se rechaza  $H_0$  y  $H_a$  se acepta, entonces se prueba la validez de la hipótesis con un nivel de error de 5% ( $\alpha = 0.05$ ), se concluye entonces que los costos haciendo uso del Algoritmo Recomendador es menor que al utilizar otros Recomendadores del mercado. En la imagen se aprecia la zona de aceptación y rechazo para la prueba de la hipótesis *Costos por el uso del Algoritmo Recomendador Machine Learning*.



**Figura N° 69: Aceptación y Rechazo “Región” – Costos por el uso del Algoritmo Recomendador Machine Learning.**

### 4.2.3. Prueba de la Hipótesis para el indicador cuantitativo:

#### Ventas al usar el Algoritmo de Machine Learning

##### A. Definición de Variables

VSR<sub>A</sub>: Ventas sin el uso del Algoritmo Recomendador Propuesto

VCR<sub>P</sub>: Ventas con el uso del Algoritmo Recomendador Propuesto

##### B. Hipótesis Estadísticas

- **Hipótesis  $H_0$ :** La cantidad de ventas logradas sin hacer uso del Algoritmo Recomendador Propuesto son menor o igual que La cantidad de ventas logradas con hacer uso del Algoritmo Recomendador *Propuesto*.

$$H_0: VSR_A - VCR_P \leq 0$$

- **Hipótesis  $H_a$ :** La cantidad de ventas logradas sin hacer uso del Algoritmo Recomendador Propuesto son mayor que la cantidad de ventas logradas con hacer uso del Algoritmo Recomendador *Propuesto*.

$$H_a: VSR_A - VCR_P > 0$$

##### C. Nivel de Significancia

Al usar un nivel de significancia del 5% ( $\alpha = 0.05$ ) entonces el nivel de confianza será del 95% ( $1 - \alpha = 0.95$ ).

##### D. Estadígrafo de contraste

Como  $n = 61$  es un valor grande, se usa la distribución normal (Z):

###### a. Promedio:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \dots\dots\dots(10)$$

**b. Varianza:**

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \dots\dots\dots(11)$$

**c. Calculando Z:**

$$Z_c = \frac{(\bar{X}_A - \bar{X}_P)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_P^2}{n_P}}} \dots\dots\dots(12)$$

**E. Resultados de la Estadística:**

Para el cálculo del tiempo Promedio se ha considerado una muestra de 61 tiempos para las búsquedas de productos a ofertar.

**Tabla N° 23: Ventas al usar el algoritmo de Machine Learning.**

Días	PRE-TEST	POST-TEST	PRE-TEST	POST-TEST	PRE-TEST	POST-TEST
	VSR <sub>A</sub>	VCR <sub>P</sub>	VSR <sub>A</sub> - $\overline{VSR}_A$	VCR <sub>P</sub> - $\overline{VCR}_P$	(VSR <sub>A</sub> - $\overline{VSR}_A$ ) <sup>2</sup>	(VCR <sub>P</sub> - $\overline{VCR}_P$ ) <sup>2</sup>
1	0	0	-0,25	-0,56	0,0625	0,3136
2	0	0	-0,25	-0,56	0,0625	0,3136
3	0	0	-0,25	-0,56	0,0625	0,3136
4	0	1	-0,25	0,44	0,0625	0,1936
5	0	1	-0,25	0,44	0,0625	0,1936
6	1	2	0,75	1,44	0,5625	2,0736
7	1	3	0,75	2,44	0,5625	5,9536
8	0	0	-0,25	-0,56	0,0625	0,3136
9	0	0	-0,25	-0,56	0,0625	0,3136
10	0	0	-0,25	-0,56	0,0625	0,3136
11	0	0	-0,25	-0,56	0,0625	0,3136
12	0	0	-0,25	-0,56	0,0625	0,3136
13	0	0	-0,25	-0,56	0,0625	0,3136
14	0	2	-0,25	1,44	0,0625	2,0736
15	1	2	0,75	1,44	0,5625	2,0736

16	0	1	-0,25	0,44	0,0625	0,1936
17	1	0	0,75	-0,56	0,5625	0,3136
18	0	0	-0,25	-0,56	0,0625	0,3136
19	0	2	-0,25	1,44	0,0625	2,0736
20	0	0	-0,25	-0,56	0,0625	0,3136
21	1	0	0,75	-0,56	0,5625	0,3136
22	1	1	0,75	0,44	0,5625	0,1936
23	1	1	0,75	0,44	0,5625	0,1936
24	0	2	-0,25	1,44	0,0625	2,0736
25	0	1	-0,25	0,44	0,0625	0,1936
26	0	0	-0,25	-0,56	0,0625	0,3136
27	0	0	-0,25	-0,56	0,0625	0,3136
28	1	2	0,75	1,44	0,5625	2,0736
29	0	0	-0,25	-0,56	0,0625	0,3136
30	0	0	-0,25	-0,56	0,0625	0,3136
31	2	0	1,75	-0,56	3,0625	0,3136
32	0	0	-0,25	-0,56	0,0625	0,3136
33	0	2	-0,25	1,44	0,0625	2,0736
34	0	1	-0,25	0,44	0,0625	0,1936
35	1	0	0,75	-0,56	0,5625	0,3136
36	1	0	0,75	-0,56	0,5625	0,3136
37	0	0	-0,25	-0,56	0,0625	0,3136
38	0	0	-0,25	-0,56	0,0625	0,3136
39	0	0	-0,25	-0,56	0,0625	0,3136
40	1	1	0,75	0,44	0,5625	0,1936
41	2	1	1,75	0,44	3,0625	0,1936
42	0	0	-0,25	-0,56	0,0625	0,3136
43	0	1	-0,25	0,44	0,0625	0,1936
44	0	0	-0,25	-0,56	0,0625	0,3136
45	0	0	-0,25	-0,56	0,0625	0,3136
46	0	0	-0,25	-0,56	0,0625	0,3136
47	0	0	-0,25	-0,56	0,0625	0,3136
48	0	0	-0,25	-0,56	0,0625	0,3136
49	0	0	-0,25	-0,56	0,0625	0,3136
50	0	2	-0,25	1,44	0,0625	2,0736

51	0	0	-0,25	-0,56	0,0625	0,3136
52	0	1	-0,25	0,44	0,0625	0,1936
53	0	0	-0,25	-0,56	0,0625	0,3136
54	0	0	-0,25	-0,56	0,0625	0,3136
55	0	1	-0,25	0,44	0,0625	0,1936
56	0	1	-0,25	0,44	0,0625	0,1936
57	0	1	-0,25	0,44	0,0625	0,1936
58	0	1	-0,25	0,44	0,0625	0,1936
59	0	0	-0,25	-0,56	0,0625	0,3136
60	0	0	-0,25	-0,56	0,0625	0,3136
61	0	0	-0,25	-0,56	0,0625	0,3136
Total	15	34			15,31	37,05

**a. Promedio:**

$$\overline{VSR_A} = \frac{15}{61}$$

$$\overline{VSR_A} = 0,25$$

$$\overline{VCR_P} = \frac{34}{61}$$

$$\overline{VCR_P} = 0,56$$

**b. Varianza:**

$$\sigma^2_A = \frac{15,31}{61}$$

$$\sigma^2_A = 0,25$$

$$\sigma^2_P = \frac{37,05}{61}$$

$$\sigma^2_P = 0,61$$

**c. Cálculo de Z:**

$$Z_c = \frac{(0,25 - 0,56)}{\sqrt{\frac{0,25}{61} + \frac{0,61}{61}}}$$

$$Z_c = \frac{-0,36}{\sqrt{(0,004 + 0,01)}}$$

$$Z_c = \frac{-0,36}{\sqrt{(0,017)}}$$

$$Z_c = \frac{-0,36}{0,13}$$

$$Z_c = -2,77$$

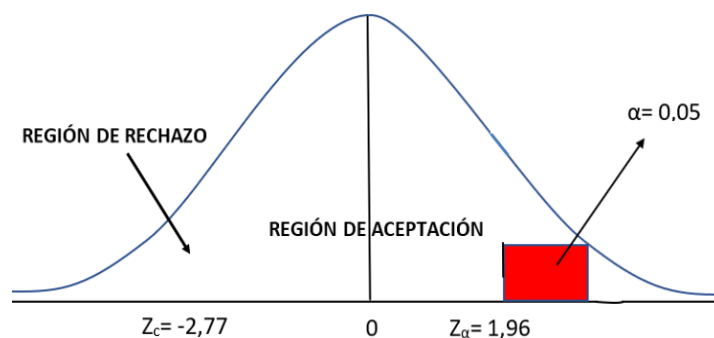
**d. Región Crítica:**

Para  $\alpha = 0.05$ , en la tabla de Distribución Normal se encuentra que  $Z_{\alpha} = 1.96$ . Por lo tanto, la zona crítica de la prueba es:

$$Z_c = \langle 1.96, \alpha \rangle.$$

**e. Conclusiones:**

En la imagen se aprecia la zona de aceptación y rechazo para la prueba de hipótesis Ventas al usar el algoritmo de Machine Learning. Teniendo que  $Z_c = -2,77$  calculado es menor que  $Z_{\alpha} = 1.96$  y encontrándose ese valor en de la zona de rechazo  $Z_c = \langle 1.96, \alpha \rangle$ ., rechazamos  $H_a$  y  $H_0$  es aceptada. Por lo que podemos concluir que haciendo uso del Algoritmo Recomendador Machine Learning, las ventas se han incrementado con un nivel de error de 5% y un nivel de confianza del 95%.



**Figura N° 70: Aceptación y Rechazo “Región” – Ventas al usar el algoritmo de Machine Learning.**



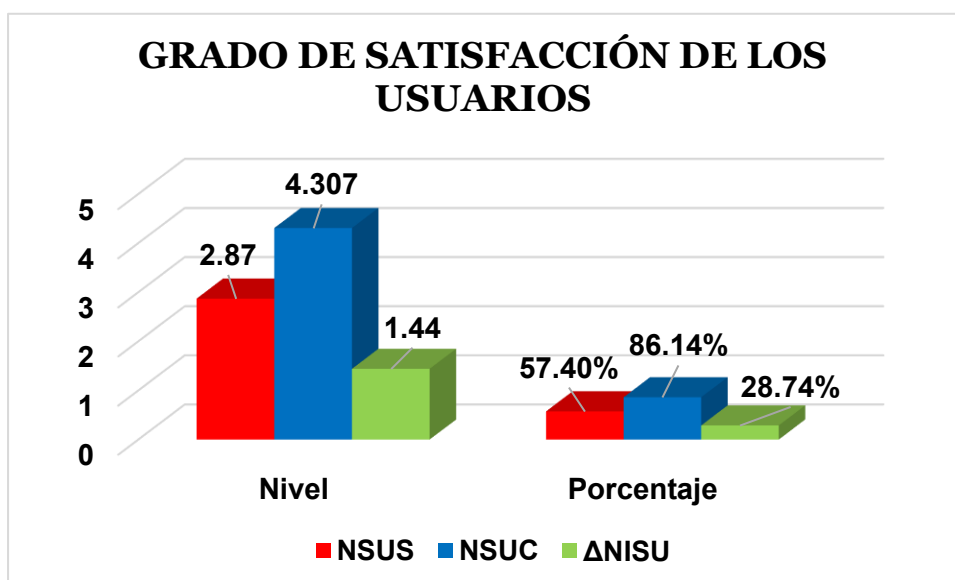
### 4.3. Resultados

#### 4.3.1. Indicador Cualitativo: Grado de satisfacción de los usuarios.

Se utilizó la escala de Likert (1-5), donde 5 es el 100%. Para el Grado de satisfacción de los usuarios sin el Algoritmo Recomendador (GSUA) y el Grado de satisfacción de los usuarios con el Algoritmo Recomendador Propuesto. (GSUP), obteniendo los siguientes resultados:

**Tabla N° 24: Presentación de Resultados del indicador Grado de satisfacción de los usuarios.**

Grado de satisfacción de los usuarios sin el Algoritmo Recomendador		Grado de satisfacción de los usuarios con el Algoritmo Recomendador Propuesto		Nivel de Impacto del indicador Grado de satisfacción de los usuarios (Incremento).	
GSUS (1-5)	GSUS %	GSUC (1-5)	GSUC %	ΔNISU (1-5)	ΔNISU %
2,87	57,4 %	4,307	86,14%	1,44	28,74%



**Figura N° 71: Resultados - Grado de satisfacción de los Usuarios**

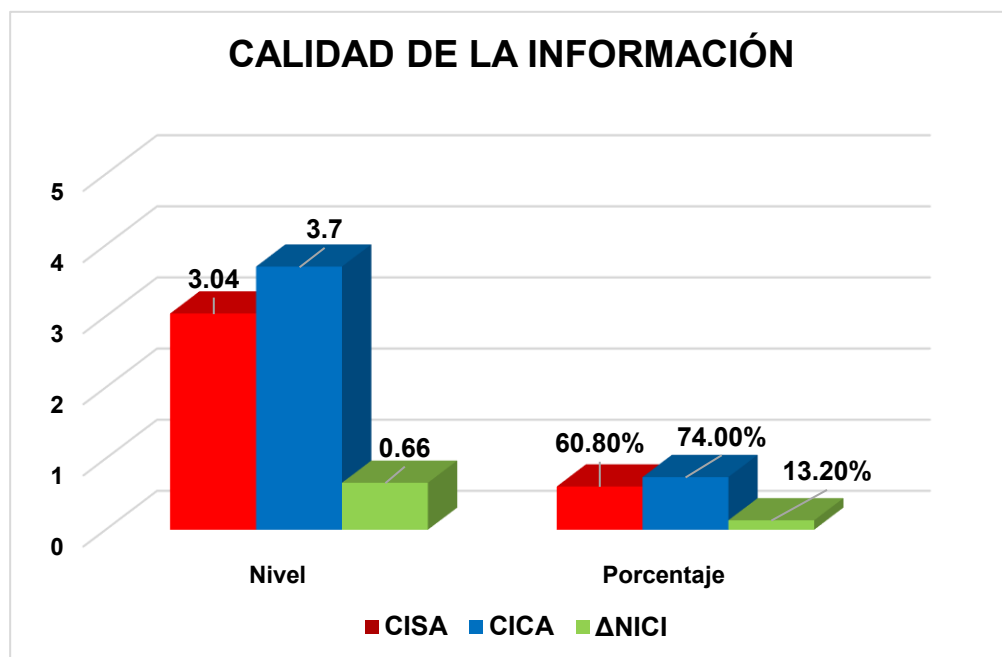
En la Tabla N° 24 se aprecia que el Grado de satisfacción de los usuarios se ha incrementado en un 28,74% y el objetivo cumple con los propuesto en esta investigación.

#### 4.3.2. Calidad de la información.

Se utilizó la escala de Likert (1-5), donde 5 es el 100%. Para la *Calidad de la Información sin el Algoritmo Recomendador (CI<sub>A</sub>)*, y *Calidad de la Información con el Algoritmo Recomendador Propuesto (CI<sub>p</sub>)*, obteniendo los siguientes resultados:

**Tabla N° 25: Resultados - Calidad de la Información.**

Calidad de la Información sin el Algoritmo Recomendador		Calidad de la Información con el Algoritmo Recomendador Propuesto		Nivel de Impacto del indicador Calidad de la Información (Incremento).	
<b>CISA</b> <b>(1-5)</b>	<b>CICA</b> <b>%</b>	<b>CISA</b> <b>(1-5)</b>	<b>GSUC</b> <b>%</b>	<b>ΔNICI</b> <b>(1-5)</b>	<b>ΔNICI</b> <b>%</b>
3,04	60,80%	3,7	74%	0,66	13,2%



**Figura N° 72: Resultados - Calidad de la Información**

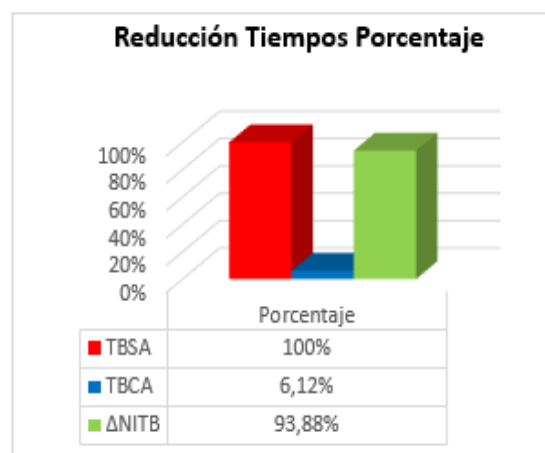
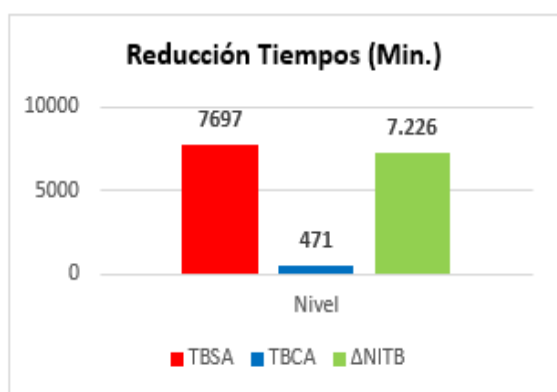
En la Tabla N° 25 se aprecia que la Calidad de la Información a aumentado en un 13,2% y el objetivo cumple con los propuesto en esta investigación.

#### 4.3.3. Indicador Cuantitativo: *Tiempos de búsqueda de productos a ofertar.*

Se utilizó los datos del cuantitativo Tiempo de búsqueda de productos a ofertar sin el Algoritmo Recomendador (TBP<sub>A</sub>), para compararlos con el cuantitativo Tiempo de búsqueda de productos a ofertar con el Algoritmo Recomendador propuesto (TBP<sub>P</sub>), lográndose los resultados:

**Tabla N° 26: Resultados -*Tiempos de búsqueda de productos a ofertar***

<i>Tiempos de búsqueda de productos a ofertar sin el Algoritmo Recomendador</i>		<i>Tiempos de búsqueda de productos a ofertar con el Algoritmo Recomendador Propuesto</i>		Nivel de Impacto del indicador <i>Tiempos de búsqueda de productos a ofertar</i> (Reducción).	
<b>TBSA (Minutos)</b>	<b>TBSA %</b>	<b>TBCA (Minutos)</b>	<b>TBCA %</b>	<b>ΔNITB (Minutos)</b>	<b>ΔNITB %</b>
7697	100%	471	6,12%	7.226	93,88%



**Figura N° 73: Resultados del Indicador *Tiempos de búsqueda de productos a ofertar.***

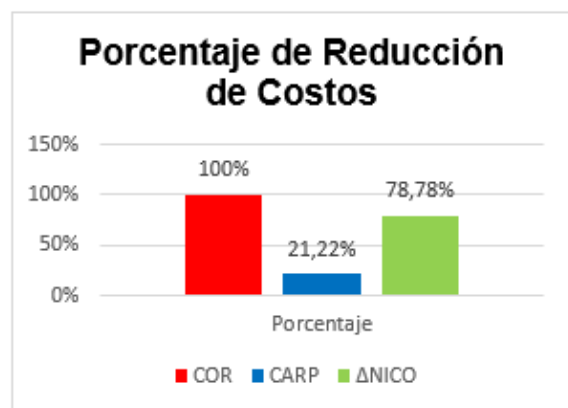
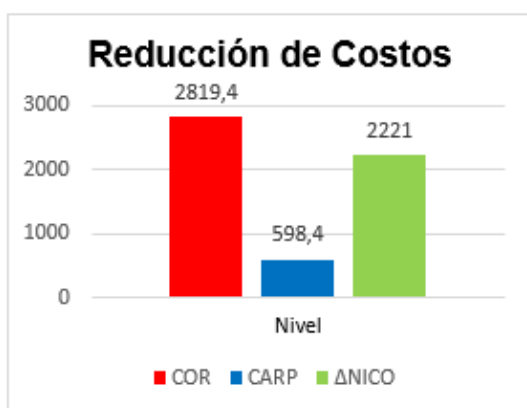
En la Tabla N° 26 se observa que los tiempos de búsqueda de productos a ofertar se han reducido en un 93,88%, lo que indica que se cumple con los objetivos de la presente investigación.

#### 4.3.4. Costos por obtención de recomendaciones con el algoritmo de Machine Learning.

Se utilizó los datos en campo del cuantitativo Costos con otro Recomendador ( $COR_i$ ), para compararlos con el cuantitativo *Costos con algoritmo Recomendador Propuesto* ( $CARP_i$ ), lográndose resultados:

**Tabla N° 27: Resultados - Costos por obtención de recomendaciones con el algoritmo Machine Learning.**

<i>Costos por obtención de recomendaciones con otros Recomendadores.</i>		<i>Costos por obtención de recomendaciones con el Algoritmo Recomendador Propuesto</i>		<i>Nivel de Impacto del indicador Costos por obtención de recomendaciones. (Reducción).</i>	
<b>COR (Soles)</b>	<b>COR %</b>	<b>CARP (Soles)</b>	<b>CARP %</b>	<b><math>\Delta</math>NICO (Soles)</b>	<b><math>\Delta</math>NICO %</b>
2819,4	100%	598,4	21,22%	2221	78,78%



**Figura N° 74: Resultados del Indicador Costos por obtención de recomendaciones con el algoritmo de Machine Learning.**

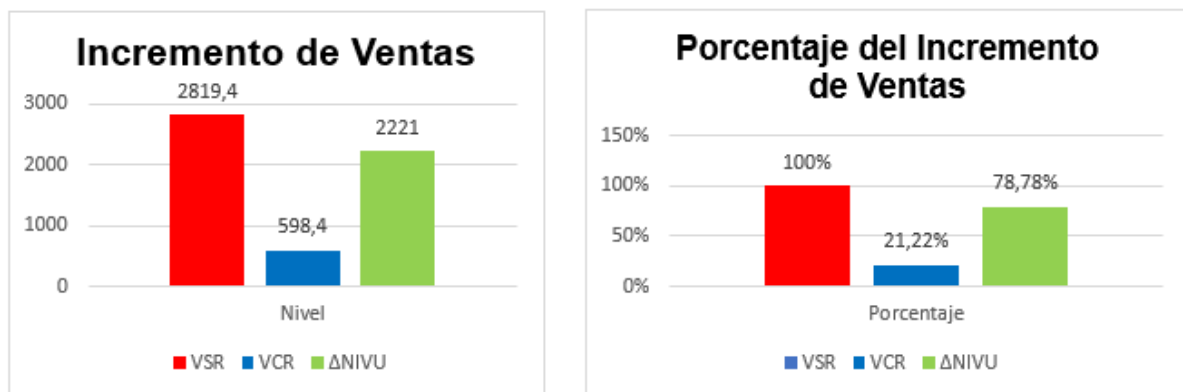
En la Tabla N° 27 y en la Figura N° 73 se observa una reducción de costos por la utilización del Algoritmo Recomendador Propuesto en un 78,78%, lo que indica que se cumple con los objetivos de la presente investigación.

<i>Ventas sin la utilización del Recomendador</i>		<i>Ventas con la utilización del Recomendador Propuesto</i>		<i>Nivel de Impacto del indicador Ventas usando el algoritmo de Machine Learning (Incremento).</i>	
<b>VSR</b>	<b>VSR %</b>	<b>VCR</b>	<b>VCR %</b>	<b>ΔNIVU</b>	<b>ΔNIVU %</b>
15	44,18%	34	100%	22	55,82%

#### 4.3.5. Ventas usando el algoritmo de Machine Learning

Se utilizó los datos en campo del cuantitativo Ventas sin la utilización del Recomendador ( $VSR_A$ ), para compararlos con el cuantitativo Ventas usando el Recomendador Propuesto ( $VCR_P$ ), logrando los resultados:

**Tabla N° 28: Resultados - Ventas usando el algoritmo de Machine Learning**



**Figura N° 75: Gráficos de Resultados del Indicador Costos por obtención de recomendaciones con el algoritmo de Machine Learning.**

En la Tabla N° 28 y en la Figura N° 74 se aprecia un incremento de ventas por la utilización del Algoritmo Recomendador Propuesto en el periodo de 2 meses en un 55,82%, lo que indica que se cumple con los objetivos de la presente investigación.

**CAPÍTULO V**  
**CONCLUSIÓN Y RECOMENDACIONES**

## 5.Conclusiones y recomendaciones

### 5.1. Conclusiones

Realizadas las pruebas mediante el uso del Algoritmo Recomendador:

- Se concluye que, el grado de satisfacción del usuario mejoró conforme a los estadísticos del Pre – Test que se mostraron por la encuesta realizada y con rangos de ponderación [1-5]. Sin el Algoritmo Recomendador el valor de 2.87 (57,4%) y con los estadísticos del Post – Test con el Algoritmo Recomendador Propuesto 4,3 (86,14%), lo que indica que el grado de satisfacción del usuario aumentó en un 1,44 (28,74%).
- Se concluye que, la calidad de la información mejoró conforme a los estadísticos del Pre – Test que se mostraron por la encuesta realizada y con rangos de ponderación [1-5]. Sin el Algoritmo Recomendador el valor de 3,04 (60,8%) y con los estadísticos del Post – Test con el Algoritmo Recomendador Propuesto 3,7 (74%), lo que significa que el nivel de calidad que recibe el usuario se incrementó en 0,66 (13,2%).
- Se concluye que los tiempos de búsqueda de productos a ofertar se disminuyeron conforme a los estadísticos del Pre test que mostraron sin Algoritmo Recomendador el valor de 7697 mins. (100.00%) y los estadísticos del Post-test con el Algoritmo Recomendador el valor de 471 minutos (6,12%), esto quiere decir que los tiempos de búsqueda de productos a ofertar se redujeron en 7226 min. (93,88%).
- Se concluye que los costos por la obtención de recomendaciones se reducen conforme a los estadísticos del Pre test que mostraron con otros Recomendadores el valor de 2819,4

Soles (100.00%) y los estadísticos con el Algoritmo Recomendador el valor de 598,4 Soles (21,22%), lo que significa que los costos por la obtención de recomendaciones se reducen en 2221 Soles (78,78%).

- Se concluye que las ventas se incrementaron conforme a los estadísticos del Pre test que mostraron con el Algoritmo Recomendador la cantidad de 34 Unidades (100.00%) y los estadísticos sin el Algoritmo Recomendador la cantidad de 15 Unidades (44,18%), lo que significa que las ventas utilizando el Algoritmo Recomendador se incrementaron en 22 Unidades (55,82%).
- Finalmente queda demostrado que el uso del Algoritmo Recomendador mejora la toma de decisiones en el área de ventas de la empresa Sport Anthony en base a la mejora del grado de satisfacción del usuario en 28,74%, la calidad de la información en base a las recomendaciones con un 13,2%, el tiempo de búsqueda de productos a ofertar en 93,88%, los costos por la obtención de recomendaciones se reducen en 78,78% y que las ventas se incrementaron en periodo establecido se incrementaron en un 55,82%. En consecuencia, la hipótesis del problema queda demostrada.

## **5.2. Recomendaciones**

Para la continuidad de la mejora del algoritmo, su uso, su operacionalidad, agilidad, etc., se presentan las siguientes recomendaciones:

Planificar horarios de inicio y fin en el proceso del ETL, ya que, al contar con grandes cantidades de datos, pueden verse afectados los usuarios por la ralentización del servidor por la carga en el proceso.



Se recomienda establecer una política en la empresa que indique el uso continuo del algoritmo recomendador para que se conozca muy bien el proceso.

Establecer capacitaciones a los usuarios finales o a toda el área de ventas, a fin que de cualquier colaborador pueda hacer uso del algoritmo, obtener así el máximo rendimiento y su aprovechamiento, para poder tomar las decisiones adecuadas teniendo siempre una buena data.

Programar encuestas a los usuarios para tener propuestas y evaluar mejoras.

## REFERENCIAS

Abellán, M. (2019). *Introducción a Python en Raspberry Pi*. PROGRAMA ERGO SUM.

<https://www.programoergosum.es/tutoriales/introduccion-a-python-en-raspberry-pi/>

Arenas Da Cruz, E. J. y Tapia Shuña, P.C. (2018). *Un Datamart con Tecnología de Inteligencia en la Optimización de la Toma de Decisiones en el Área de Logística Comercial de un Grupo Comercial* [Tesis de pregrado, Universidad Nacional de Ucayali]. Repositorio Institucional.

<http://repositorio.unu.edu.pe/bitstream/handle/UNU/3599/SISTEMAS-2018-ELFERARENAS-RESUMEN.pdf?sequence=1&isAllowed=y>

Coronel Romero, E. L. y Jumbo Sinchire, V. F. (2019). *Sistema Recomendador de Cuestionarios para Apoyo al Aprendizaje Mediante Bloqueo de Aplicaciones en la Plataforma Android*. [Tesis de pregrado, Universidad Nacional de Loja]. Repositorio institucional. <https://dspace.unl.edu.ec/jspui/handle/123456789/22231>

Espinoza Airac, G. X. y León Muñoz, E.F. (2020). *Modelo de Machine Learning para la Clasificación de Estudiantes de acuerdo a su Rendimiento Académico en el Centro de Idiomas de la Universidad Nacional del Santa* [Tesis de pregrado, Universidad Nacional del Santa]. Repositorio Institucional.

<http://repositorio.uns.edu.pe/bitstream/handle/UNS/3588/50119.pdf?sequence=1&isAllowed=y>

- Fernández Iglesias, D. (2014). *Uso de técnicas de recomendación en sistemas dispersos*. [Tesis de pregrado, Universidad La Coruña]. Archivo digital. [https://ruc.udc.es/dspace/bitstream/handle/2183/11819/FernandezIglesias\\_Diego\\_TD\\_2014.pdf;sequence=5](https://ruc.udc.es/dspace/bitstream/handle/2183/11819/FernandezIglesias_Diego_TD_2014.pdf;sequence=5)
- Gunther, R. E. (2008). *La verdad sobre la toma de decisiones* (1 ed.). Editorial Financial Times Press. [https://www.academia.edu/40653958/La\\_Verdad\\_sobre\\_la\\_Toma\\_de\\_Decisiones](https://www.academia.edu/40653958/La_Verdad_sobre_la_Toma_de_Decisiones)
- Kimball, R. y Caserta, J. (2004). *The data warehouse ETL toolkit : practical techniques for extracting, cleaning, conforming, and delivering data*. Wiley Publishing, Inc. <http://users.itk.ppke.hu/~szoer/DW/Kimball%20&%20Caserta%20-The%20Data%20Warehouse%20ETL%20Toolkit%20%5BWiley%202004%5D.pdf>
- Landa Tinco, S. y Ticona Pacheco, F.J. (2015). *Un Modelo de Recomendación de Productos utilizando Redes Sociales implementado en Tecnologías de la Web Semántica*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos]. Repositorio institucional. [https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/4661/Landa\\_ts.pdf?sequence=1&isAllowed=y](https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/4661/Landa_ts.pdf?sequence=1&isAllowed=y)
- Martin, G. (2018). *Primeros pasos con Jupyter Notebook*. <https://www.adictosaltrabajo.com/2018/01/18/primeros-pasos-con-jupyter-notebook/>

Mitchell, T. M. (2018). *Machine Learning*. (1 ed.). Editorial Indian.

<https://www.amazon.com/Machine-Learning-Tom-M-Mitchell/dp/0070428077?asin=0070428077&revisionId=&format=4&depth=1>

Moya García, R. (2013). *SVD Aplicado a Sistemas de Recomendación Basados en Filtrado Colaborativo* [Tesis de maestría, Universidad Politécnica de Madrid]. Archivo digital.

[http://informatico.ricardomoya.es/docs/ProyectoFinDeMaster\\_V1.pdf](http://informatico.ricardomoya.es/docs/ProyectoFinDeMaster_V1.pdf)

Muñiz, L. (2018). *El Business Intelligence herramienta clave para mejorar la gestión*

*empresarial*. <https://docplayer.es/92111400-El-business-intelligence-herramienta-clave-para-mejorar-la-gestion-empresarial.html>

Parada, M. (2019). *Qué es el SQL Server*. Open Webinars. <https://openwebinars.net/blog/que-es-sql-server/>

Ramírez, C. A. (2018). *Algoritmo SVD aplicado a los sistemas de recomendación en el comercio*. TIA, 6(1), pp. 18-27.

<https://revistas.udistrital.edu.co/index.php/tia/article/view/11827/pdf>

Rouhiainen, L. (2018). *Inteligencia Artificial*. Alienta Editorial.

[https://books.google.com.pe/books?id=\\_T9xDwAAQBAJ&source=gbs\\_navlinks\\_s](https://books.google.com.pe/books?id=_T9xDwAAQBAJ&source=gbs_navlinks_s)

Rouse, M. (2015). *OLAP o Procesamiento Analítico en Línea*

<https://www.computerweekly.com/es/definicion/OLAP-o-procesamiento-analitico-en-linea>

Simon, A. (1998). *90 Days to the Data Mart*. Wiley. <https://www.worldcat.org/title/90-days-to-the-data-mart/oclc/45729805?referer=di&ht=edition>

Tique Ortiz, J. (2016). *Diseño de Datamart en las temáticas de producción e inversión de Ciencia, Tecnología e Innovación en Colombia*. [Tesis de maestría, Escuela Colombiana de Ingeniería Julio Garavito]. Repositorio institucional.

<https://repositorio.escuelaing.edu.co/bitstream/handle/001/827/Tique%20Ortiz,%20Jinneth%20-%202018.pdf;jsessionid=FA22079F8E4918B382721BE68FCFAACA?sequence=1>

Toro, L. (2017). *Anaconda Distribution: La Suite más completa para la Ciencia de datos con Python*. <https://blog.desdelinux.net/ciencia-de-datos-con-python/>

Yucra Halanoca, A. (2018). *Datamart de la gerencia de supermercado Mía Market de Juliaca – 2018* [Tesis de pregrado, Universidad Nacional del Altiplano]. Archivo digital.

<http://repositorioslatinoamericanos.uchile.cl/handle/2250/3278920>

## ANEXO A

**CARTA DE AUTORIZACIÓN DE USO DE INFORMACIÓN DE EMPRESA PARA  
OBTENCIÓN DE TÍTULO PROFESIONAL**



Yo, Erika Claudia Mechan Chavesta, identificada con DNI 45219822, en calidad de Representante de la empresa/SportAnthony con RUC 20445795292, ubicada en la ciudad de Chimbote, Perú.

**OTORGO LA AUTORIZACIÓN,**


Al señor Luis Ricardo Marcelo Gómez identificado con DNI N° 70932343, egresado/bachiller de la carrera de Ingeniería de Sistemas e Informática de la Universidad Nacional del Santa para que utilice información SIN FINES DE LUCRO de la empresa con la finalidad de que pueda desarrollar su Trabajo de Investigación para optar el Título Profesional.



---

**Firma del Representante**  
**DNI: 45219822**

El Egresado o Bachiller declara que los datos emitidos en esta carta y en el Trabajo de Investigación, en la Tesis o Trabajo de Suficiencia Profesional son **CON FINES ACADÉMICOS**. En caso de comprobarse lo contrario, el Egresado será sometido al inicio del procedimiento disciplinario correspondiente; y, asimismo, asumirá toda la responsabilidad ante posibles acciones legales que la empresa, otorgante de información, pueda ejecutar.



---

**Firma del Egresado o Bachiller**  
**DNI: 70932343**

## ANEXO B

### CUESTIONARIOS

#### 6.1. Cuestionario de Nivel de Satisfacción - Sin Algoritmo Recomendador

##### CUESTIONARIO DE NIVEL DE SATISFACCIÓN DE LOS USUARIOS - SIN ALGORITMO RECOMENDADOR

En el siguiente cuestionario encontrarás diversas preguntas sobre aspectos referidos a la funcionalidad, calidad y servicio con respecto al procedimiento actual de búsqueda de productos a ofertar. (Sin Algoritmo Recomendador). El objetivo de estas preguntas es conocer el nivel de satisfacción de los usuarios con respecto al procedimiento actual.

Indicaciones: Para contestar marque en cada frase la opción que mejor indique su grado de aprobación mediante la siguiente escala:

AP: Aprobación Plena = 5

AS: Aprobación Simple = 4

IN: Indecisión o Indiferencia = 3

DS: Desaprobación Simple = 2

DP: Desaprobación Plena = 1

	Nº	ITEM	AP	AS	IN	DS	DP
			PUNTUACIÓN				
FUNCIONALIDAD	1	¿Es fácil el procedimiento de búsqueda de productos sin un algoritmo?	5	4	3	2	1
	2	¿Se encuentra satisfecho con la variedad de productos a ofertar sin un algoritmo?	5	4	3	2	1
	3	¿Es amigable el procedimiento de búsqueda de productos a ofertar sin un algoritmo?	5	4	3	2	1
	4	¿La información sin un recomendador le ayuda a tener una buena visión de los productos a ofertar?	5	4	3	2	1
	5	¿Cómo calificaría la flexibilidad del procedimiento sin un algoritmo al momento de busca productos?	5	4	3	2	1
SERVICIO Y CALIDAD	6	¿Las recomendaciones se entregan en tiempos favorables al área de ventas?	5	4	3	2	1
	7	¿Cómo califica la manera de obtener las recomendaciones de productos sin un algoritmo?	5	4	3	2	1
	8	¿Cómo califica usted la calidad de las recomendaciones obtenidas sin un Algoritmo?	5	4	3	2	1
	9	¿Las recomendaciones sin un algoritmo son fáciles de analizar?	5	4	3	2	1
	10	¿Considera que las recomendaciones sin un algoritmo son útiles para el área de ventas?	5	4	3	2	1


  
 Dr. Luis Vladimir Urrelo Human

Figura N° 76 Cuestionario de Nivel de Satisfacción - Sin Algoritmo Recomendador

## 6.2. Cuestionario de Nivel de Satisfacción - Con Algoritmo Recomendador

### CUESTIONARIO DE NIVEL DE SATISFACCIÓN DE LOS USUARIOS - CON ALGORITMO RECOMENDADOR

En el siguiente cuestionario encontrarás diversas preguntas sobre aspectos referidos a la funcionalidad, calidad y servicio con respecto al algoritmo recomendado. El objetivo de estas preguntas es conocer el nivel de satisfacción de los usuarios con respecto al procedimiento actual.

Indicaciones: Para contestar marque en cada frase la opción que mejor indique su grado de aprobación mediante la siguiente escala:

AP: Aprobación Plena = 5

AS: Aprobación Simple = 4

IN: Indecisión o Indiferencia = 3

DS: Desaprobación Simple = 2

DP: Desaprobación Plena = 1

	N°	ITEM	AP	AS	IN	DS	DP
			PUNTUACIÓN				
FUNCIONALIDAD	1	¿Es comprensible el funcionamiento y uso del algoritmo?	5	4	3	2	1
	2	¿Se encuentra satisfecho con la variedad de recomendaciones que brinda el algoritmo?	5	4	3	2	1
	3	¿La interfaz donde se muestran las recomendaciones es amigable para usted?	5	4	3	2	1
	4	¿La información del recomendador ayuda a tener una mejor visión de los productos a ofertar?	5	4	3	2	1
	5	¿Cómo califica la flexibilidad del recomendador al momento de realizar la emisión de las mismas?	5	4	3	2	1
SERVICIO Y CALIDAD	6	¿Las recomendaciones se entregan en tiempos favorables al área de ventas?	5	4	3	2	1
	7	¿Cómo considera la manera de obtener las recomendaciones de productos?	5	4	3	2	1
	8	¿Cómo califica usted la calidad de las recomendaciones obtenidas?	5	4	3	2	1
	9	¿Las recomendaciones de los productos son fáciles de analizar?	5	4	3	2	1
	10	¿Considera que las recomendaciones son útiles para el área de ventas?	5	4	3	2	1



Dr. Luis Vladimir Urrelo Huiman

Figura N° 77 Cuestionario de Nivel de Satisfacción - Con Algoritmo Recomendador



### 6.3. Cuestionario de Calidad de la Información - Sin Algoritmo Recomendador

#### CUESTIONARIO DE CALIDAD DE LA INFORMACIÓN - SIN ALGORITMO RECOMENDADOR

En el siguiente cuestionario encontrarás diversas preguntas sobre aspectos referidos a la calidad de información que se obtiene con respecto al procedimiento actual de búsqueda de productos a ofertas. (Sin Algoritmo Recomendador). El objetivo de estas preguntas es conocer el nivel de calidad de la información que obtienen con respecto al procedimiento actual.

Indicaciones: Para contestar marque en cada frase la opción que mejor indique su grado de aprobación mediante la siguiente escala:

AP: Aprobación Plena = 5

AS: Aprobación Simple = 4

IN: Indecisión o Indiferencia = 3

DS: Desaprobación Simple = 2

DP: Desaprobación Plena = 1

Nº	ITEM	AP	AS	IN	DS	DP
		PUNTUACION				
1	¿Se obtiene recomendaciones relevantes para la toma de decisiones?	5	4	3	2	1
2	¿Considera que las recomendaciones ofrecen considerables detalles para la toma de decisiones?	5	4	3	2	1
3	¿Considera suficientes las recomendaciones que se brindan sin un Algoritmo recomendador?	5	4	3	2	1
4	¿Considera que las recomendaciones que se muestran sin un algoritmo Recomendador son datos actualizados?	5	4	3	2	1
5	¿Considera que las recomendaciones sin un algoritmo son fáciles de comprender?	5	4	3	2	1
6	¿Considera que las recomendaciones que se le ofrece influyen en la toma de decisiones?	5	4	3	2	1
7	¿Cómo considera la manera de obtener las recomendaciones de productos?	5	4	3	2	1
8	¿Cómo califica usted la calidad de las recomendaciones obtenidas sin un Algoritmo?	5	4	3	2	1



Dr. Luis Vladimir Urreio Huiman

Figura N° 78 Cuestionario de Calidad de la Información - Sin Algoritmo Recomendador

## 6.4. Cuestionario de Calidad de la Información - Con Algoritmo Recomendador

### CUESTIONARIO DE CALIDAD DE LA INFORMACIÓN - CON ALGORITMO RECOMENDADOR

En el siguiente cuestionario encontrarás diversas preguntas sobre aspectos referidos a la calidad de información que se obtiene con respecto al algoritmo recomendador. El objetivo de estas preguntas es conocer el nivel de calidad de la información, específicamente con las recomendaciones que obtienen con el algoritmo recomendador.

Indicaciones: Para contestar marque en cada frase la opción que mejor indique su grado de aprobación mediante la siguiente escala:

AP: Aprobación Plena = 5

AS: Aprobación Simple = 4

IN: Indecisión o Indiferencia = 3

DS: Desaprobación Simple = 2

DP: Desaprobación Plena = 1

Nº	ITEM	AP	AS	IN	DS	DP
		PUNTUACIÓN				
1	¿Obtiene recomendaciones relevantes para la toma de decisiones?	5	4	3	2	1
2	¿Considera que las recomendaciones ofrecen considerables detalles para la toma de decisiones?	5	4	3	2	1
3	¿Considera que conjuntamente a su procedimiento actual, el Algoritmo recomendador complementa las recomendaciones de productos a ofertar?	5	4	3	2	1
4	¿Considera que las recomendaciones que muestra el Recomendador son datos actualizados?	5	4	3	2	1
5	¿Considera que las recomendaciones ofrecidas son fáciles de comprender?	5	4	3	2	1
6	¿Considera que las recomendaciones que se le ofrece influyen en la toma de decisiones?	5	4	3	2	1
7	¿Cómo considera la manera de obtener las recomendaciones de productos?	5	4	3	2	1
8	¿Cómo califica usted la calidad de las recomendaciones obtenidas con el Algoritmo?	5	4	3	2	1



Dr. Luis Vladimir Urrelo Huiman

Figura N° 79 Cuestionario de Calidad de la Información - Con Algoritmo Recomendador

## ANEXO C

### ALFA DE CRONBACH

**Tabla N° 29 Rangos y magnitud de confiabilidad por Alfa de Cronbach**

Rangos	Magnitud
0,81 a 1,00	Muy buena
0,61 a 0,80	Buena
0,41 a 0,60	Moderada
0,21 a 0,40	Baja
0,01 a 0,20	Muy baja

#### 7.1. Indicador: Nivel de satisfacción de los usuarios - ENCUESTA

Para asegurar la consistencia de los datos de las encuestas realizadas a los usuarios (personal administrativo, de ventas y marketing), se calculó el coeficiente de Alfa de Cronbach:

**Tabla N° 30 Indicador Nivel de Satisfacción - Estadísticos**

Usuarios	Preguntas										Suma de Ítems
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
U1	4	4	1	2	1	3	2	2	2	3	24
U2	4	1	4	3	2	3	3	3	2	3	28
U3	5	4	1	1	3	5	4	4	3	4	34
U4	2	2	2	3	3	3	3	3	3	3	27
U5	3	4	4	3	2	4	3	3	3	4	33
U6	3	1	4	3	3	3	3	3	3	4	30
<b>Varianza Típica</b>	<b>0,92</b>	<b>1,89</b>	<b>1,89</b>	<b>0,58</b>	<b>0,56</b>	<b>0,58</b>	<b>0,33</b>	<b>0,33</b>	<b>0,22</b>	<b>0,25</b>	<b>29</b>

**Calculamos el Alfa de Cronbach a partir de las varianzas:**

$$\alpha = \left[ \frac{k}{k-1} \right] \left[ 1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right],$$

Donde;

$K$  = Número de Ítems = 10

$\sum_{i=1}^k S_i^2$  = Sumatoria de las Varianzas de los Ítems = 7,56

$S_t^2$  = La varianza de la suma de los Ítems = 29

$\alpha = 0,82$

Teniendo en cuenta el valor hallado de  $\alpha=0,82$  y de acuerdo a la magnitud en los rangos de Alfa de Cronbach (Coeficiente alfa  $> 0.8$  es muy bueno). Se concluye que el instrumento y por consiguiente los datos, son fiables.

## 7.2. Indicador: Calidad de la Información - ENCUESTA

Para asegurar la consistencia de los datos de las encuestas realizadas a los usuarios (personal administrativo, de ventas y marketing), se calculó el coeficiente de Alfa de Cronbach:

**Tabla N° 31 Indicador Calidad de la Información - Estadísticos**

Usuarios	Preguntas								Suma de Ítems
	P1	P2	P3	P4	P5	P6	P7	P8	
U1	4	4	3	2	4	4	3	3	27
U2	4	1	3	3	3	3	3	3	23
U3	3	3	3	1	4	4	4	4	26
U4	2	3	3	3	4	3	3	3	24
U5	4	3	3	3	3	3	2	4	25
U6	3	3	2	3	3	2	3	3	22
<b>Varianza Típica</b>	0,56	0,81	0,14	0,58	0,25	0,47	0,33	0,22	25

**Calculamos el Alfa de Cronbach a partir de las varianzas:**

$$\alpha = \left[ \frac{k}{k-1} \right] \left[ 1 - \frac{\sum_{i=1}^k S_i^2}{S_t^2} \right],$$

Donde;

$K$  = Número de Ítems = 8

$\sum_{i=1}^k S_i^2$  = Sumatoria de las Varianzas de los Ítems = 3,37

$S_t^2$  = La varianza de la suma de los Ítems = 25

$\alpha = 0,96$

Teniendo en cuenta el valor hallado de  $\alpha=0,96$  y de acuerdo a la magnitud en los rangos de Alfa de Cronbach (Coeficiente alfa  $> 0.8$  es muy bueno). Se concluye que el instrumento y por consiguiente los datos, son fiables.



## ACTA DE APROBACIÓN DE ORIGINALIDAD

Yo, **Briones Pereyra, Lizbeth Dora** asesora de la

Facultad:	Ciencias		Educación		Ingeniería	X
Departamento Académico	CIVIL Y SISTEMAS					
Escuela de Posgrado	Doctorado			Maestría		

Programa:

De la Universidad Nacional del Santa. Asesor / Unidad de Investigación revisora del trabajo de Investigación intitulado:

IMPLEMENTACIÓN DE UN ALGORITMO RECOMENDADOR MACHINE LEARNING USANDO DATAMART PARA LA TOMA DE DECISIONES EN EL ÁREA DE VENTAS DE LA EMPRESA SPORT ANTHONY EN CHIMBOTE

Del estudiante / docente: Luis Ricardo Marcelo Gómez

De la escuela / departamento académico: INGENIERÍA DE SISTEMAS E INFORMÁTICA

Constato que la investigación presentada tiene un porcentaje de similitud del 19% el cual se verifica con el reporte de originalidad de la aplicación Turnitin adjunto.

Quién suscribe la presente, declaro el haber analizado dicho reporte y concluyo que las coincidencias detectadas no se conforman como plagio. A mi claro saber y entender, la investigación cumple con las normas de citas y referencias establecidas por la Universidad Nacional del Santa.

Nuevo Chimbote, 24 de junio de 2022

Firma

Nombres y Apellidos del Asesor/Presidente UI: Briones Pereyra, Lizbeth Dora

DNI: 32960646

# IMPLEMENTACIÓN DE UN ALGORITMO RECOMENDADOR MACHINE LEARNING USANDO DATAMART PARA LA TOMA DE DECISIONES EN EL ÁREA DE VENTAS DE LA EMPRESA SPORT ANTHONY EN CHIMBOTE

## INFORME DE ORIGINALIDAD



## FUENTES PRIMARIAS

<b>1</b>	<b>repositorio.uns.edu.pe</b> Fuente de Internet	<b>6%</b>
<b>2</b>	<b>Submitted to Universidad Cesar Vallejo</b> Trabajo del estudiante	<b>2%</b>
<b>3</b>	<b>hdl.handle.net</b> Fuente de Internet	<b>1%</b>
<b>4</b>	<b>1library.co</b> Fuente de Internet	<b>1%</b>
<b>5</b>	<b>dspace.unitru.edu.pe</b> Fuente de Internet	<b>1%</b>
<b>6</b>	<b>alicia.concytec.gob.pe</b> Fuente de Internet	<b>1%</b>
<b>7</b>	<b>repositorio.ucv.edu.pe</b> Fuente de Internet	<b>1%</b>
<b>8</b>	<b>live-with-wisdom.tistory.com</b> Fuente de Internet	<b>1%</b>

9	<a href="http://repositorio.usm.cl">repositorio.usm.cl</a> Fuente de Internet	<1 %
10	<a href="http://docplayer.es">docplayer.es</a> Fuente de Internet	<1 %
11	<a href="http://repositorio.uladech.edu.pe">repositorio.uladech.edu.pe</a> Fuente de Internet	<1 %
12	Submitted to Universidad Wiener Trabajo del estudiante	<1 %
13	<a href="http://repositorio.unc.edu.pe">repositorio.unc.edu.pe</a> Fuente de Internet	<1 %
14	<a href="http://repositorio.unsa.edu.pe">repositorio.unsa.edu.pe</a> Fuente de Internet	<1 %
15	Submitted to Unviersidad de Granada Trabajo del estudiante	<1 %
16	<a href="http://revistas.udistrital.edu.co">revistas.udistrital.edu.co</a> Fuente de Internet	<1 %
17	<a href="http://es.slideshare.net">es.slideshare.net</a> Fuente de Internet	<1 %
18	<a href="http://www.amazon.com">www.amazon.com</a> Fuente de Internet	<1 %
19	<a href="http://repositorio.autonoma.edu.pe">repositorio.autonoma.edu.pe</a> Fuente de Internet	<1 %
20	<a href="http://mineriadatosmiva.home.blog">mineriadatosmiva.home.blog</a> Fuente de Internet	<1 %

21	<a href="https://repository.udistrital.edu.co">repository.udistrital.edu.co</a> Fuente de Internet	<1 %
22	Submitted to Escuela Politecnica Nacional Trabajo del estudiante	<1 %
23	<a href="https://centrodeconocimiento.ccb.org.co">centrodeconocimiento.ccb.org.co</a> Fuente de Internet	<1 %
24	<a href="https://dspace.unl.edu.ec">dspace.unl.edu.ec</a> Fuente de Internet	<1 %
25	Submitted to Universidad Peruana de Las Americas Trabajo del estudiante	<1 %
26	<a href="https://dspace.utb.edu.ec">dspace.utb.edu.ec</a> Fuente de Internet	<1 %
27	<a href="https://repositorio.uancv.edu.pe">repositorio.uancv.edu.pe</a> Fuente de Internet	<1 %
28	<a href="https://sites.google.com">sites.google.com</a> Fuente de Internet	<1 %
29	Submitted to ulatina Trabajo del estudiante	<1 %
30	<a href="https://ludwingalvarado.wordpress.com">ludwingalvarado.wordpress.com</a> Fuente de Internet	<1 %
31	<a href="https://repositorio.utp.edu.pe">repositorio.utp.edu.pe</a> Fuente de Internet	<1 %
32	Submitted to Universidad Carlos III de Madrid Trabajo del estudiante	<1 %



		<1 %
33	<a href="http://digibug.ugr.es">digibug.ugr.es</a> Fuente de Internet	<1 %
34	<a href="http://repositorio.unap.edu.pe">repositorio.unap.edu.pe</a> Fuente de Internet	<1 %
35	<a href="http://riti.es">riti.es</a> Fuente de Internet	<1 %
36	<a href="http://cybertesis.unmsm.edu.pe">cybertesis.unmsm.edu.pe</a> Fuente de Internet	<1 %
37	<a href="http://www.scribd.com">www.scribd.com</a> Fuente de Internet	<1 %
38	<a href="http://www.timetoast.com">www.timetoast.com</a> Fuente de Internet	<1 %
39	<a href="http://documents.mx">documents.mx</a> Fuente de Internet	<1 %
40	<a href="http://repositorio.uta.edu.ec">repositorio.uta.edu.ec</a> Fuente de Internet	<1 %
41	<a href="http://tesis.usat.edu.pe">tesis.usat.edu.pe</a> Fuente de Internet	<1 %
42	<a href="http://www.coursehero.com">www.coursehero.com</a> Fuente de Internet	<1 %
43	<a href="http://www.sigen.gov.ar">www.sigen.gov.ar</a> Fuente de Internet	<1 %

44	Carolina Santillán Torres Torija. "Adherencia terapéutica en personas que viven con VIH/SIDA", Universidad Nacional Autónoma de México, 2014 Publicación	<1 %
45	paolavaldivieso.blogspot.com Fuente de Internet	<1 %
46	repositorio.unp.edu.pe Fuente de Internet	<1 %
47	repositorio.uss.edu.pe Fuente de Internet	<1 %
48	www.ilustrados.com Fuente de Internet	<1 %
49	www.mindmeister.com Fuente de Internet	<1 %
50	pt.scribd.com Fuente de Internet	<1 %
51	repositorio.unfv.edu.pe Fuente de Internet	<1 %
52	repositorio.unu.edu.pe Fuente de Internet	<1 %
53	www.azurian.com Fuente de Internet	<1 %
54	www.clubensayos.com Fuente de Internet	<1 %

55	<a href="http://www.dspace.unitru.edu.pe">www.dspace.unitru.edu.pe</a> Fuente de Internet	<1 %
56	<a href="http://www.graphispack-asociacion.com">www.graphispack-asociacion.com</a> Fuente de Internet	<1 %
57	<a href="http://www.santiagooponiente.cl">www.santiagooponiente.cl</a> Fuente de Internet	<1 %
58	<a href="http://www.seguridadidl.org.pe">www.seguridadidl.org.pe</a> Fuente de Internet	<1 %
59	<a href="http://www.sqlinfo.in">www.sqlinfo.in</a> Fuente de Internet	<1 %
60	<a href="http://moam.info">moam.info</a> Fuente de Internet	<1 %
61	<a href="http://patents.google.com">patents.google.com</a> Fuente de Internet	<1 %

Excluir citas

Activo

Excluir coincidencias

Apagado

Excluir bibliografía

Activo