

**UNIVERSIDAD PERUANA UNIÓN**

**ESCUELA DE POSGRADO**

Unidad de Posgrado de Ingeniería y Arquitectura



**Modelamiento de datos sobre el comportamiento del  
consumidor y hábitos de compra on-line**

Tesis para obtener el Título de Segunda Especialidad Profesional de  
Ingeniería: Estadística Aplicada para Investigación

**Autor:**

Alison Samantha Hernandez Zevallos  
Rosario Violeta Grijalva Salazar

**Asesor:**

Dr. Paulo Jorge Canas Rodríguez

Lima, mayo de 2024

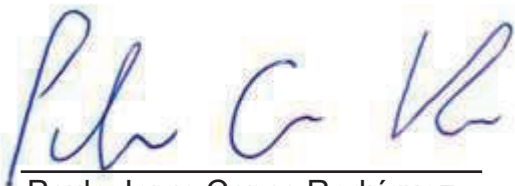
## DECLARACIÓN JURADA DE ORIGINALIDAD DE LA TESIS

Yo Paulo Jorge Canas Rodríguez, docente de la Unidad de Posgrado de la Facultad de Ingeniería y Arquitectura, Escuela de Posgrado de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: **“Modelamiento de datos sobre el comportamiento del consumidor y hábitos de compra on-line”** de los autores Alison Samantha Hernandez Zevallos y Rosario Violeta Grijalva Salazar, tiene un índice de similitud de 11 % verificable en el informe del programa Turnitin, y fue realizada en la Universidad Peruana Unión bajo mi dirección.

En tal sentido asumo la responsabilidad que corresponde ante cualquier falsedad u omisión de los documentos como de la información aportada, firmo la presente declaración en la ciudad de Lima, a los 2 días del mes de julio del año 2024



Paulo Jorge Canas Rodríguez

## ACTA DE SUSTENTACIÓN DE TESIS

En Lima, Ñaña, Villa Unión a 14 días del mes de mayo del año 2024, siendo las 9:30 horas, se reunieron de forma online sincrónica, bajo la dirección del presidente del jurado Dra. Ethel Altez Ortiz, secretaria Mg. Lizeth Geanina Huanca Lopez; los demás miembros: PhD. Javier Linkolk López Gonzales, Dr. Josué Edison Turpo Chaparro y el asesor Dr. Paulo Jorge Canas Rodrigues, con el propósito de administrar el acto académico de sustentación de Tesis de la Segunda Especialidad titulada "Modelamiento de datos sobre el comportamiento del consumidor y hábitos de compra on-line", conducente a la obtención del Título de Segunda Especialidad Profesional de Ingeniería: Estadística Aplicada para Investigación.

El presidente inició el acto académico de sustentación invitando a los candidatos a hacer uso del tiempo determinado para su exposición. Concluida la exposición, el presidente invitó a los demás miembros del jurado a efectuar las preguntas, cuestionamientos y aclaraciones pertinentes, los cuales fueron absueltos por los candidatos. Luego, se produjo un receso para las deliberaciones y la emisión del dictamen del jurado. Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

Candidatos: Rosario Violeta Grijalva Salazar y Alison Samantha Hernandez Zevallos

| Calificación |           |         |                             | Mérito     |
|--------------|-----------|---------|-----------------------------|------------|
|              | Vigesimal | Literal | Cualitativa                 |            |
| Aprobado     | 19        | A       | Con nominación de Excelente | Excelencia |

Finalmente, el presidente del jurado invitó a los candidatos a ponerse de pie, para recibir la evaluación final. Además, el presidente concluyó el acto académico de sustentación, procediéndose a registrar a registrar las firmas respectivas.



Presidente



Secretario



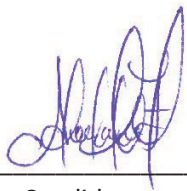
Asesor



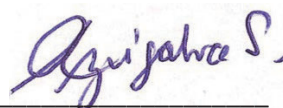
Miembro



Miembro



Candidato



Candidato

# Modelamiento de datos sobre el comportamiento del consumidor y hábitos de compra on-line

Modeling data on consumer behavior and online purchasing habits

## Resumo

Con el crecimiento del comercio electrónico y la necesidad de desarrollar estudios sobre este tema, este artículo tiene como objetivo realizar un análisis exploratorio de datos integral en términos de sector y tipo de producto, temporalidad y ubicación geográfica, con el fin de estudiar patrones y formular hipótesis para investigaciones posteriores. Utilizando un estudio observacional transversal de datos de comercio electrónico en Brasil, realizamos un preprocesamiento en bases de datos transaccionales proporcionadas por la tienda de comercio electrónico Olist en el sitio Kaggle a través de una muestra aleatoria. El análisis se centró en analizar las métricas de precio de productos y la demanda de ventas, donde los principales resultados fueron la identificación de sectores más atractivos, patrones de consumo de estacionalidad y hora/día en relación con el precio y la demanda de ventas, y patrones espaciales en relación con los estados de Brasil. Nuestros resultados mostraron que la distribución del precio medio por sector de producto es homogénea, destacándose los sectores de cama, mesa y baño, deporte y ocio, y belleza y salud con una alta demanda, en comparación con los sectores de relojes y cosas interesantes que tienen una mediana más alta en relación con los otros. En cuanto a la estacionalidad y temporalidad, observamos una tendencia de crecimiento a lo largo del tiempo con patrones de ventas/hora y día más bajos los fines de semana y, finalmente, en el análisis espacial, se muestra la discrepancia en ventas y valor pagado por productos entre las regiones del sureste y las demás regiones.

## Abstract

"With the growth of e-commerce and the need for the development of studies on this subject, this article aims to conduct a comprehensive exploratory data analysis in terms of sector and product type, temporality, and geographical location with the intention of studying patterns and formulating hypotheses for further research. Using a transversal observational study of e-commerce data in Brazil, we conducted preprocessing on transversal databases made available by the Olist e-commerce store on the Kaggle website through random sampling. The analysis focused on examining metrics such as product prices and sales demand, with the main results identifying more attractive sectors, consumption patterns of seasonality and time/day in relation to price and sales demand, and spatial patterns in relation to states in Brazil. Our findings reveal that the distribution of average prices per product sector is homogeneous, with sectors such as Bed Table Bath, Sport Leisure, and Beauty Health standing out with high demand. However, sectors like Watches and Cool Things have a higher median compared to others. Regarding seasonality and temporality, there is a growing trend over time with sales/hour and day patterns significantly lower on weekends. Finally, the spatial analysis shows a discrepancy in sales and the amount paid for products between the Southeast region and other regions."

## Palavras Chaves

Comercio Electrónico, Análisis Exploratorio de Datos, Comportamiento del Consumidor, Hábitos de Compra, Brasil

## Keywords

E-commerce, Exploratory Data Analysis, Consumer Behaviour, Shopping Habits, Brazil.