A blue and red text on a black background

AI-generated content may be incorrect.

**CIENCIA**

**DE**

**DATOS**

REPUBLICA DOMINICANA

INSTITUTO TECNOLOGICO DE LAS AMERICAS



**PROCESADORES DE TEXTO Y TECNICAS DE INFORME**

SECCION:

**2025-C-1-2802-3094-TCD-201**

PROYECTO FINAL PROCESADORES DE TEXTO Y TECNICAS DE INFORME:

Informe de Análisis Exploratorio de Datos

Rossy Elania Arvelo Pérez

20240861

**Jose Arcadio Garcia Garcia.**

**Indice**

[Introducción 4](#_Toc196254417)

[Metodología 5](#_Toc196254418)

[Análisis General 6](#_Toc196254419)

[Conclusiones Generales 14](#_Toc196254420)

[Recomendaciones Generales 15](#_Toc196254421)

[Grafica 1: Venta Promedio Por Rango de Edad 6](#_Toc196253419)

[Grafica 2: Distribución de Edad de los Clientes 7](#_Toc196253420)

[Grafica 3: Edad de los Clientes por categoría de Producto 8](#_Toc196253421)

[Grafica 4: Ventas por categoría de Producto 9](#_Toc196253422)

[Grafica 5: Clientes por país 10](#_Toc196253423)

[Grafica 6: Promedio de Venta por ocupación 11](#_Toc196253424)

[Grafica 7: Distribución por Genero 12](#_Toc196253425)

# **Introducción**

Este informe presenta un análisis exploratorio de un conjunto de datos con 2,216 órdenes de venta, extraído de un archivo Excel titulado "Data Proyecto Final Proce Texto.xlsx". Los datos incluyen información demográfica y comercial de los clientes, como edad, género, ocupación, país, categoría del producto comprado, monto de la venta y fecha de la orden.

El análisis revela que los clientes tienen una edad promedio de 53 años, con una distribución equitativa entre hombres y mujeres, y ocupaciones variadas. Las ventas se concentran principalmente en bicicletas de montaña y carretera, con un precio promedio de $3,193. Los clientes de entre 36 y 55 años son quienes más compran, y la mayoría de las ventas se realizaron en pocos países. También se identificaron diferencias de compra según la edad, ocupación y país de residencia.

Este análisis proporciona una visión general del perfil de los clientes y sus patrones de compra, sirviendo como base para futuras estrategias de segmentación y marketing.

# **Metodología**

Para realizar este análisis exploratorio de datos (EDA), se siguieron varias etapas utilizando el lenguaje de programación **R**, lo que permitió manipular, explorar y visualizar los datos de manera efectiva.

1. **Importación de datos:**

Se comenzó leyendo el archivo en formato Excel mediante la función read\_excel() del paquete readxl.

1. **Inspección inicial:**

Una vez importados los datos, se realizó una primera revisión utilizando funciones básicas como head() para observar las primeras filas, str() para conocer la estructura de las variables, y summary() para obtener estadísticas descriptivas generales. Estas funciones ayudaron a identificar el tipo de variables (categóricas, numéricas, fechas) y posibles inconsistencias.

1. **Revisión de valores únicos y faltantes:**

Para conocer los diferentes valores dentro de cada columna y detectar si había datos nulos/faltantes, se utilizaron funciones como sapply() y colSums(is.na()). Esto facilitó la identificación de variables con poca variabilidad o posibles problemas de calidad en los datos.

1. **Resumen general de los datos:**

Se aplicó la función skim() para obtener un resumen completo y estructurado de las variables según su tipo (por ejemplo, numéricas, de texto o de fecha). Esta herramienta ayudó a visualizar patrones generales y preparar los datos para el análisis más detallado.

# **Análisis General**

Se ha analizado una base de datos con **2,216 ventas realizadas durante el año 2011**. Cada fila representa una compra hecha por un cliente, y contamos con información detallada sobre quién compró, qué compró, cuándo y cuánto gastó.

**Información sobre los clientes**

* **Edad promedio** de los clientes: **53 años**, siendo el más joven de 37 años y el mayor de 92. Esto nos dice que la mayoría de nuestros clientes son **adultos de mediana o mayor edad**.
* En cuanto al **género**, hay una **distribución muy pareja** entre hombres y mujeres. Esto muestra que nuestros productos atraen **por igual a ambos géneros**.
* Los clientes tienen distintos **niveles de educación**, desde educación secundaria hasta estudios universitarios. Esto indica que nuestro producto llega a personas con diferentes niveles académicos.
* Las **ocupaciones** también son variadas: hay obreros, técnicos, profesionales, entre otros. Nuestro mercado es diverso y no se limita a un solo tipo de trabajo o estilo de vida.
* Nuestros clientes provienen de **6 países distintos**. Esto sugiere que la empresa tiene un **alcance internacional**, aunque centrado en pocos países.

**Información sobre los productos y ventas**

* Todos los productos vendidos son **bicicletas**, en dos tipos: **de montaña y de carretera**. La empresa se enfoca en un solo rubro, lo que muestra **especialización**.
* El **precio promedio** por venta fue de aproximadamente **$3,193**, y el máximo alcanzó **$3,578**. Esto nos dice que se trata de **bicicletas de precio medio a alto**, lo cual sugiere buena calidad o tecnología avanzada.

**Periodo de ventas**

* Todas las compras se realizaron en el año **2011**, entre enero y diciembre. Tenemos una visión **completa de ese año**, lo cual permite analizar tendencias y patrones.

**Calidad de los datos**

* No hay datos vacíos o perdidos. Todo está bien organizado y completo. Esto es importante porque nos da **confianza** para seguir analizando sin errores.Gráfico, Gráfico de barras

  El contenido generado por IA puede ser incorrecto.

**Grafica 1: Venta Promedio Por Rango de Edad**

**Interpretación del gráfico:**  
El gráfico muestra cuánto gastan en promedio los clientes según su edad. Las personas entre 36 y 55 años son las que tienen el promedio de ventas más alto. En cambio, los más jóvenes (menores de 25) y los mayores de 65 gastan menos.

**Conclusión:**  
Los adultos entre 36 y 55 años son el grupo de clientes más valioso en términos de ventas. Ellos representan el mayor poder de compra.

**Recomendaciones:**

* Mantener campañas dirigidas a personas entre 36 y 55 años, ya que son quienes más compran.
* Diseñar promociones específicas para atraer a jóvenes y adultos mayores y así aumentar su ticket promedio.
* Evaluar si los productos actuales son atractivos para los extremos de edad (menores de 25 y mayores de 65).

Gráfico, Histograma

El contenido generado por IA puede ser incorrecto.

**Grafica 2: Distribución de Edad de los Clientes**

**Interpretación del gráfico:**  
Este gráfico muestra cuántos clientes hay por grupo de edad. La mayoría está entre 26 y 45 años, especialmente en el grupo de 31 a 35. Hay menos clientes muy jóvenes o mayores.

**Conclusión:**  
La base de clientes está compuesta principalmente por adultos jóvenes y de mediana edad. Hay poco alcance entre los extremos de edad.

**Recomendaciones:**

* Seguir fortaleciendo la conexión con clientes de 26 a 45 años, ya que son la mayoría.
* Investigar por qué no se están atrayendo muchos jóvenes o personas mayores.
* Explorar nuevos canales o productos que interesen a esos grupos menos representados.
* Gráfico, Gráfico de cajas y bigotes

  El contenido generado por IA puede ser incorrecto.

**Grafica 3: Edad de los Clientes por categoría de Producto**

**Interpretación del gráfico:**  
El gráfico compara las edades de los clientes según el tipo de producto que compran. Algunas categorías son más populares entre los adultos, mientras que otras tienen clientes más variados en edad.

**Conclusión:**  
Cada categoría de producto tiene un perfil de edad diferente. Algunas se centran en personas de más edad, otras en jóvenes o en un rango más amplio.

**Recomendaciones:**

* Personalizar las estrategias de marketing por categoría, basándose en el grupo de edad dominante.
* Si se quiere expandir una categoría, considerar qué edad no está siendo bien cubierta y adaptar el producto o la promoción.
* Usar esta información para mejorar la segmentación de campañas y aumentar la efectividad de las ventas.

Gráfico

El contenido generado por IA puede ser incorrecto.

**Grafica 4: Ventas por categoría de Producto**

**Interpretación del gráfico:**

El gráfico muestra cuánto se vendió según la categoría del producto (por ejemplo: tecnología, muebles, ropa, etc.). Cada categoría tiene una barra o porción que indica su nivel de ventas. Algunas categorías tienen muchas ventas, otras menos.

**Conclusión:**

Las categorías con más ventas son las más importantes para el negocio. También se puede notar si hay algunas que no están funcionando tan bien en comparación.

**Recomendaciones:**

* **Seguir potenciando** las categorías con más ventas, ya que están funcionando bien.
* **Revisar** por qué las categorías con menos ventas no están rindiendo tanto (¿hay poco stock? ¿no se promocionan?).
* **Hacer promociones o descuentos** en las categorías más bajas para atraer más atención hacia ellas.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

**Grafica 5: Clientes por país**

**Interpretación del gráfico:**

Este gráfico muestra cuántos clientes hay en cada país. Algunos países tienen muchos clientes, otros tienen muy pocos o casi ninguno.

**Conclusión:**

La mayor parte de nuestros clientes está concentrada en uno o dos países. Hay otros países donde casi no tenemos presencia.

**Recomendaciones:**

* **Fortalecer relaciones** con los países donde tenemos más clientes (por ejemplo, mejor atención, ofertas especiales).
* **Explorar oportunidades** en países con pocos clientes. Tal vez se puede hacer publicidad o alianzas para atraer más gente.
* **Adaptar el marketing** según el país (idioma, cultura, intereses) para lograr más impacto.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

**Grafica 6: Promedio de Venta por ocupación**

**Interpretación del gráfico:**

Aquí vemos cuánto gasta en promedio cada cliente, según su ocupación (por ejemplo: estudiante, profesional, jubilado, etc.). Algunas ocupaciones compran más que otras.

**Conclusión:**

Hay ocupaciones que generan más ingresos por persona. Es decir, algunos grupos de clientes gastan más dinero que otros.

**Recomendaciones:**

* **Enfocar campañas publicitarias** en las ocupaciones que más compran, ya que son más rentables.
* **Entender por qué otras ocupaciones compran menos**: ¿no les interesa el producto?, ¿les parece caro?, ¿no se están alcanzando con la publicidad?
* **Diseñar promociones especiales** para los grupos con bajo promedio de compra, para intentar aumentar sus compras.

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

**Grafica 7: Distribución por Genero**

**Interpretación del gráfico:**

Este gráfico muestra cuántos clientes son mujeres (F) y cuántos son hombres (M).  
Las dos barras son casi del mismo tamaño, lo que significa que hay **casi la misma cantidad de hombres y mujeres** en la base de datos.  
Cada barra representa un poco más de **1,100 personas**.

**Conclusión**

La empresa tiene una **distribución de género muy equilibrada**.  
Esto quiere decir que **tanto hombres como mujeres están bien representados** entre los clientes.

**Recomendaciones**

* **Seguir manteniendo esta igualdad** en campañas de marketing y atención al cliente.
* Se pueden hacer **ofertas o promociones neutrales** que funcionen para ambos géneros.
* En análisis futuros, se puede revisar si **existen diferencias de comportamiento o de compras** entre hombres y mujeres para personalizar mejor las estrategias.

# **Conclusiones Generales**

El análisis de los datos de ventas nos revela patrones claros en cuanto al perfil de los clientes y su comportamiento de compra. La mayoría de los clientes son adultos entre 36 y 55 años, con una distribución de género equilibrada y diversas ocupaciones y niveles educativos. Las bicicletas, tanto de montaña como de carretera, son los únicos productos vendidos, y se consideran de precio medio a alto.

Se identificó que el mayor volumen de ventas y el gasto promedio más alto se concentran en clientes de mediana edad, mientras que los extremos de edad presentan menor participación. Además, algunas categorías de productos tienen un perfil de edad específico, lo cual permite una segmentación más precisa.

En cuanto a la ubicación geográfica, la mayor parte de los clientes se concentra en pocos países, lo que sugiere una oportunidad de expansión. También se observa que ciertas ocupaciones generan mayores ingresos por cliente, lo que puede guiar futuras estrategias comerciales.

Finalmente, la calidad de los datos es alta, sin valores faltantes ni inconsistencias, lo que brinda una base sólida para análisis más complejos en futuras etapas del proyecto.

# **Recomendaciones Generales**

* Segmentar campañas publicitarias enfocándose en clientes de entre 36 y 55 años, ya que representan el grupo con mayor poder adquisitivo.
* Diseñar estrategias específicas para atraer a clientes jóvenes (menores de 25) y mayores de 65 años, con productos o promociones adaptadas a sus intereses y necesidades.
* Aprovechar la información por categoría de producto para personalizar las campañas según el perfil de edad más frecuente en cada línea.
* Fortalecer la presencia en países con menor participación mediante campañas de marketing localizadas y análisis de mercado.
* Fomentar las categorías de productos menos vendidas, evaluando posibles barreras como falta de promoción o atractivo del producto.
* Analizar el comportamiento de compra por ocupación y crear promociones dirigidas a los grupos con mayor gasto promedio.
* Mantener una estrategia de marketing inclusiva en cuanto a género, ya que la base de clientes es equilibrada entre hombres y mujeres
* Dar seguimiento a la limpieza y organización de los datos para garantizar análisis confiables en el futuro.