

Tarea 3: Análisis de datos de ventas de una tienda de electrónica (enunciado 1)

Integrantes:

Gonzalo Muñoz H.

Docente:

Patricio Galdames

Fecha:

07/07/2024

1. Introducción

En este proyecto, se aplicaron conceptos y habilidades de manipulación y análisis de datos utilizando el lenguaje de programación R. El objetivo era procesar, limpiar y analizar datos de ventas de una tienda de electrónica para generar informes y visualizaciones que ayuden en la toma de decisiones estratégicas. A lo largo del proyecto, se generaron datos de ventas de manera aleatoria, se realizaron limpiezas de datos, se manipularon data frames con dplyr y se crearon visualizaciones básicas con ggplot2.

2. Desarrollo

Primero, se generaron datos aleatorios para simular las ventas de una tienda de electrónica. Los datos incluyeron información sobre el ID de venta, ID de producto, nombre del producto, categoría, precio unitario y cantidad vendida. Estos datos se almacenaron en un data frame y se guardaron en un archivo CSV llamado 'ventas.csv'.

```
# Cargar librerías necesarias
library(tidyverse)
set.seed(123)

# Generar datos aleatorios
num_ventas <- 1000
id_venta <- 1:num_ventas
id_producto <- sample(1:100, num_ventas, replace = TRUE)
nombre_producto <- paste("Producto", id_producto)
categorias <- c("Electrónica", "Accesorios", "Hogar", "Juguetes")
categoria <- sample(categorias, num_ventas, replace = TRUE)
precio_unitario <- round(runif(num_ventas, 5, 500), 2)
cantidad_vendida <- sample(1:20, num_ventas, replace = TRUE)

# Crear data frame
ventas <- data.frame(
  id_venta,
  id_producto,
  nombre_producto,
  categoria,
  precio_unitario,
  cantidad_vendida
)

# Guardar en archivo CSV
write.csv(ventas, "ventas.csv", row.names = FALSE)
# Cargar el archivo CSV
ventas <- read.csv("ventas.csv")
```

Se genero un análisis de datos con los datos generados del csv 'ventas.csv'.

Total de Ventas por Categoría: Se calculó el total de ventas para cada categoría de producto.

Precio Promedio por Categoría: Se calculó el precio promedio de los productos en cada categoría.

Productos Más Vendidos: Se identificaron los cinco productos más vendidos.

Productos Menos Vendidos: Se identificaron los cinco productos menos vendidos.

```
#total de ventas por categoría
total_ventas_categoria <- ventas %>%
  group_by(categoría) %>%
  summarise(total_ventas = sum(precio_unitario * cantidad_vendida))

print(total_ventas_categoria)

#precio promedio por categoría
precio_promedio_categoria <- ventas %>%
  group_by(categoría) %>%
  summarise(precio_promedio = mean(precio_unitario))

print(precio_promedio_categoria)

#productos más vendidos
productos_mas_vendidos <- ventas %>%
  group_by(nombre_producto) %>%
  summarise(total_cantidad = sum(cantidad_vendida)) %>%
  arrange(desc(total_cantidad)) %>%
  head(5)

print(productos_mas_vendidos)

#productos menos vendidos
productos_menos_vendidos <- ventas %>%
  group_by(nombre_producto) %>%
  summarise(total_cantidad = sum(cantidad_vendida)) %>%
  arrange(total_cantidad) %>%
  head(5)

print(productos_menos_vendidos)
```

Se creo un grafico que muestra la distribución de precios de los productos por su categoría.

```
#visualización de la distribución de precios por categoría
ggplot(ventas, aes(x = categoria, y = precio_unitario)) +
  geom_boxplot() +
  labs(title = "Distribución de Precios por Categoría",
       x = "Categoría",
       y = "Precio Unitario") +
  theme_minimal()
```

Por último, se guardaron los resultados en archivos csv.

```
#Guardar los resultados en un archivo csv
write.csv(total_ventas_categoria, "total_ventas_categoria.csv", row.names = FALSE)
write.csv(precio_promedio_categoria, "precio_promedio_categoria.csv", row.names = FALSE)
write.csv(productos_mas_vendidos, "productos_mas_vendidos.csv", row.names = FALSE)
write.csv(productos_menos_vendidos, "productos_menos_vendidos.csv", row.names = FALSE)
```

3. Conclusión

Mediante la elaboración de este proyecto de Data Science pude aprender y profundizar lo visto en clases por el profesor.

Este proyecto reforzó la comprensión de las herramientas básicas de R, así como las técnicas de análisis de datos esenciales para la toma de decisiones informadas. La práctica de guardar y cargar archivos CSV resultó fundamental para asegurar la correcta documentación del análisis.