

Proyecto 5: Optimización de Envíos y Logística

Requisitos:

1. Calcular el tiempo promedio de envío para cada modalidad de envío comparando la fecha de pedido y la fecha de envío.
2. Identificar las regiones con los tiempos promedio de envío más largos.
3. Proporcionar información sobre qué modalidad de envío genera el mayor margen de ganancia.
4. Analizar posibles correlaciones entre los tiempos de envío y las ventas totales o la satisfacción del cliente (si los datos están disponibles).
5. Recomendar las mejores modalidades de envío para clientes de alto valor basándose en datos históricos.

1. Calcular el tiempo promedio de envío para cada modalidad de envío comparando la fecha de pedido y la fecha de envío.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Cargar el archivo Excel
file_path = './furniture (1).xlsx' # Cambiar por la ruta de tu archivo
data = pd.ExcelFile(file_path)

# Cargar la hoja con los datos
df = data.parse('sales')

# Convertir las columnas de fechas al formato datetime
df['order_date'] = pd.to_datetime(df['order_date'])
df['ship_date'] = pd.to_datetime(df['ship_date'])
```

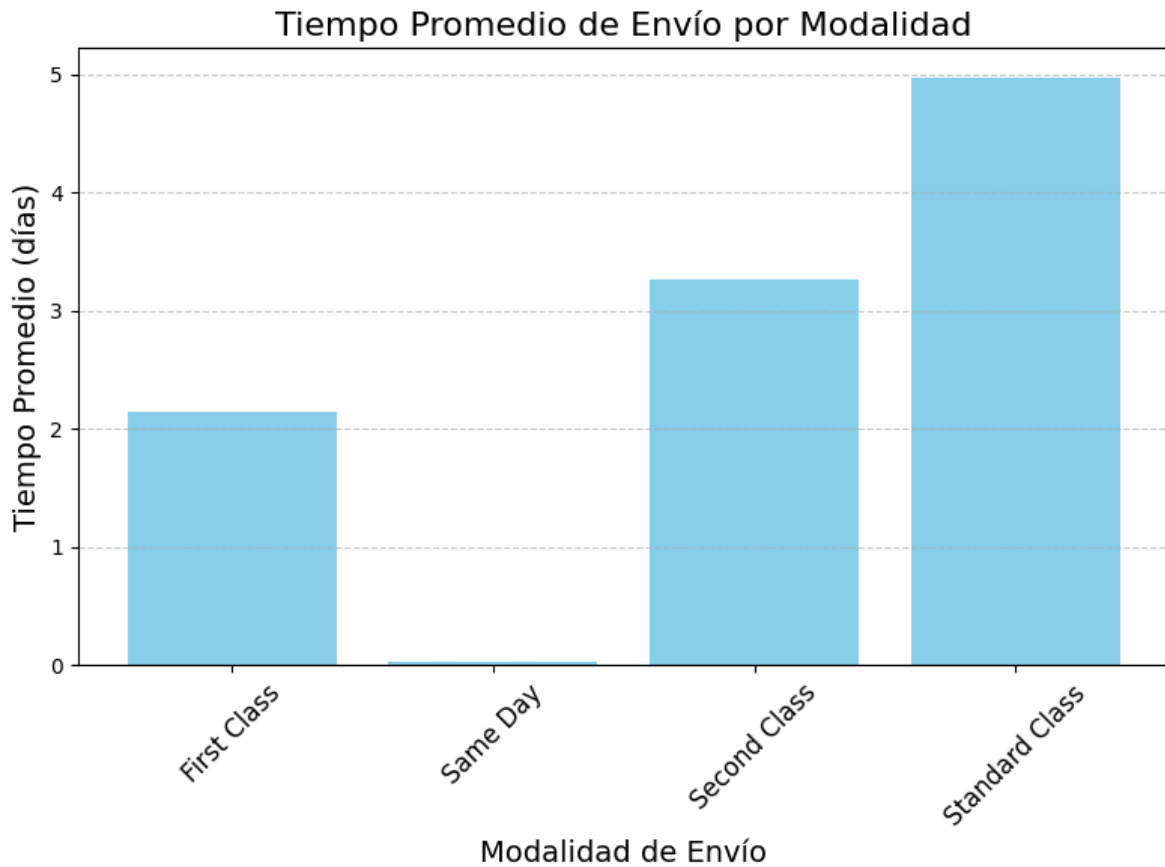
```

# Calcular la diferencia en días entre la fecha de envío y la fecha de pedido
df['shipping_time_days'] = (df['ship_date'] - df['order_date']).dt.days

# Calcular el tiempo promedio de envío para cada modalidad de envío
average_shipping_time = df.groupby('ship_mode')['shipping_time_days'].mean().reset_index()
average_shipping_time.rename(columns={'shipping_time_days': 'average_shipping_time_days'}, inplace=True)

# gráfico
plt.figure(figsize=(8, 6))
plt.bar(average_shipping_time['ship_mode'], average_shipping_time['average_shipping_time_days'])
plt.title('Tiempo Promedio de Envío por Modalidad', fontsize=16)
plt.xlabel('Modalidad de Envío', fontsize=14)
plt.ylabel('Tiempo Promedio (días)', fontsize=14)
plt.xticks(rotation=45, fontsize=12)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

```



2. Identificar las regiones con los tiempos promedio de envío más largos.

```
# Group by region and calculate the average shipping time
average_shipping_time = df.groupby('region')['shipping_time_days'].mean().sort_values(ascending=True)

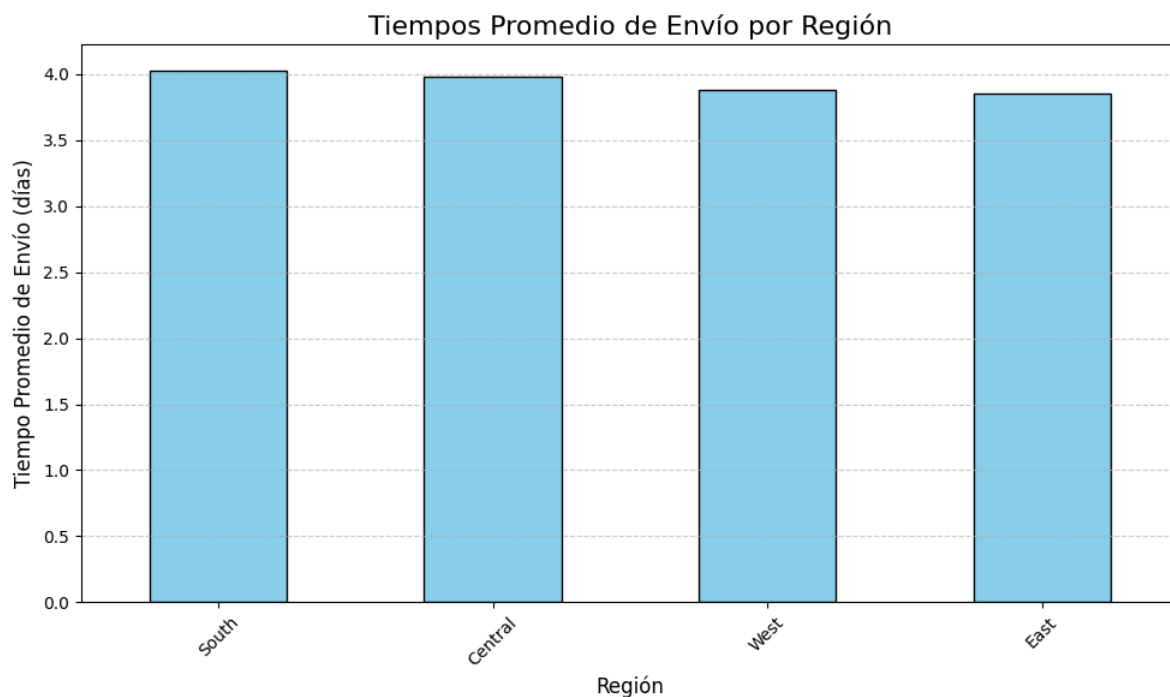
# Display the results
print(average_shipping_time)

# gráfico
plt.figure(figsize=(10, 6))
average_shipping_time.plot(kind='bar', color='skyblue', edgecolor='black')
plt.title('Tiempos Promedio de Envío por Región', fontsize=16)
plt.xlabel('Región', fontsize=12)
plt.ylabel('Tiempo Promedio de Envío (días)', fontsize=12)
plt.xticks(rotation=45)
```

```
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()

plt.savefig('average_shipping_time_by_region.png')
plt.show()
```

```
region
South      4.024096
Central    3.977131
West       3.881188
East       3.850250
Name: shipping_time_days, dtype: float64
```



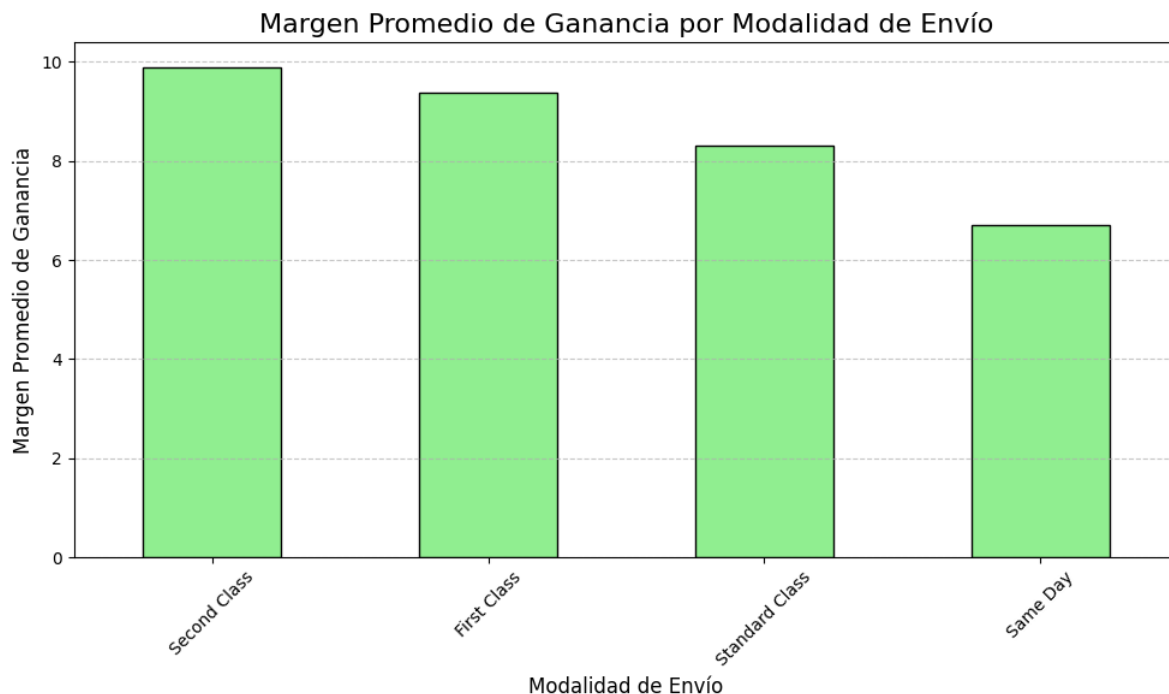
3. Proporcionar información sobre qué modalidad de envío genera el mayor margen de ganancia.

```
# Calcular el margen promedio de ganancia por modalidad de envío
average_profit_by_ship_mode = df.groupby('ship_mode')['profit'].mean().sort_values(ascending=
```

```
# Mostrar los resultados
print(average_profit_by_ship_mode)

# Crear un gráfico de barras para visualizar los resultados
plt.figure(figsize=(10, 6))
average_profit_by_ship_mode.plot(kind='bar', color='lightgreen', edgecolor='black')
plt.title('Margen Promedio de Ganancia por Modalidad de Envío', fontsize=16)
plt.xlabel('Modalidad de Envío', fontsize=12)
plt.ylabel('Margen Promedio de Ganancia', fontsize=12)
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

```
ship_mode
Second Class    9.897568
First Class     9.379044
Standard Class  8.301855
Same Day        6.700407
Name: profit, dtype: float64
```

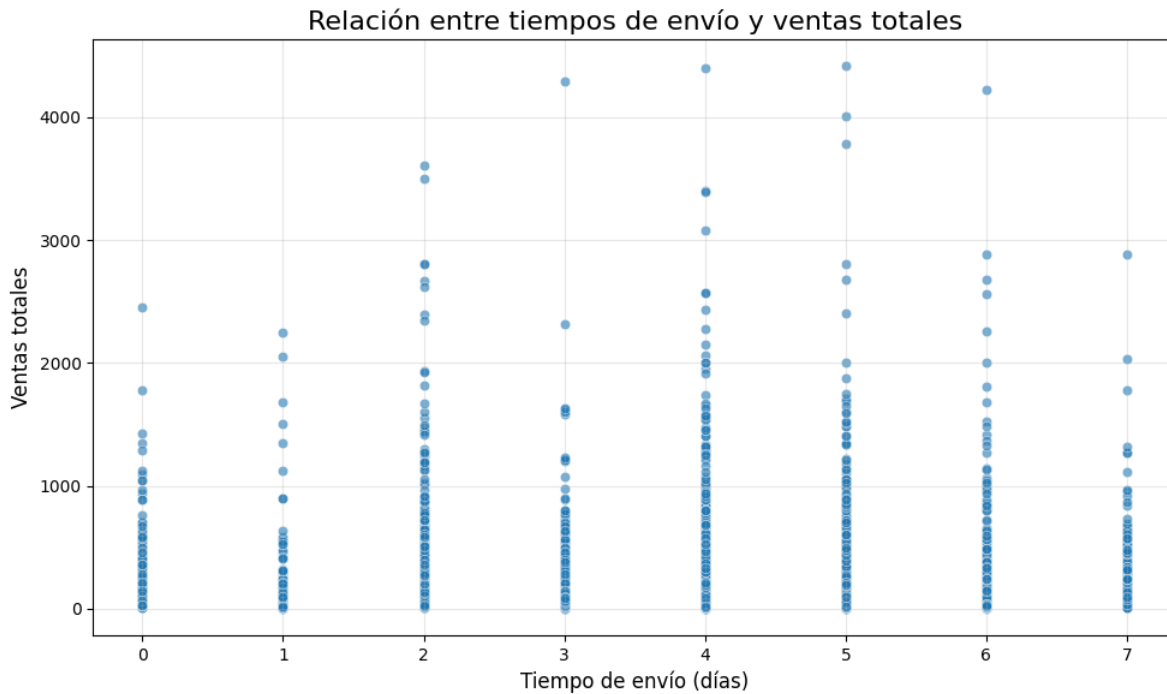


4. Analizar posibles correlaciones entre los tiempos de envío y las ventas totales o la satisfacción del cliente (si los datos están disponibles).

```
# Calcular la correlación entre tiempos de envío y ventas totales
correlation_sales = df['shipping_time_days'].corr(df['total_sales'])
print(f"Correlación entre tiempos de envío y ventas totales: {correlation_sales}")

# Visualizar la relación con un gráfico de dispersión
plt.figure(figsize=(10, 6))
sns.scatterplot(x='shipping_time_days', y='total_sales', data=df, alpha=0.6)
plt.title('Relación entre tiempos de envío y ventas totales', fontsize=16)
plt.xlabel('Tiempo de envío (días)', fontsize=12)
plt.ylabel('Ventas totales', fontsize=12)
plt.grid(alpha=0.3)
plt.tight_layout()
plt.savefig('shipping_time_vs_sales.png')
plt.show()
```

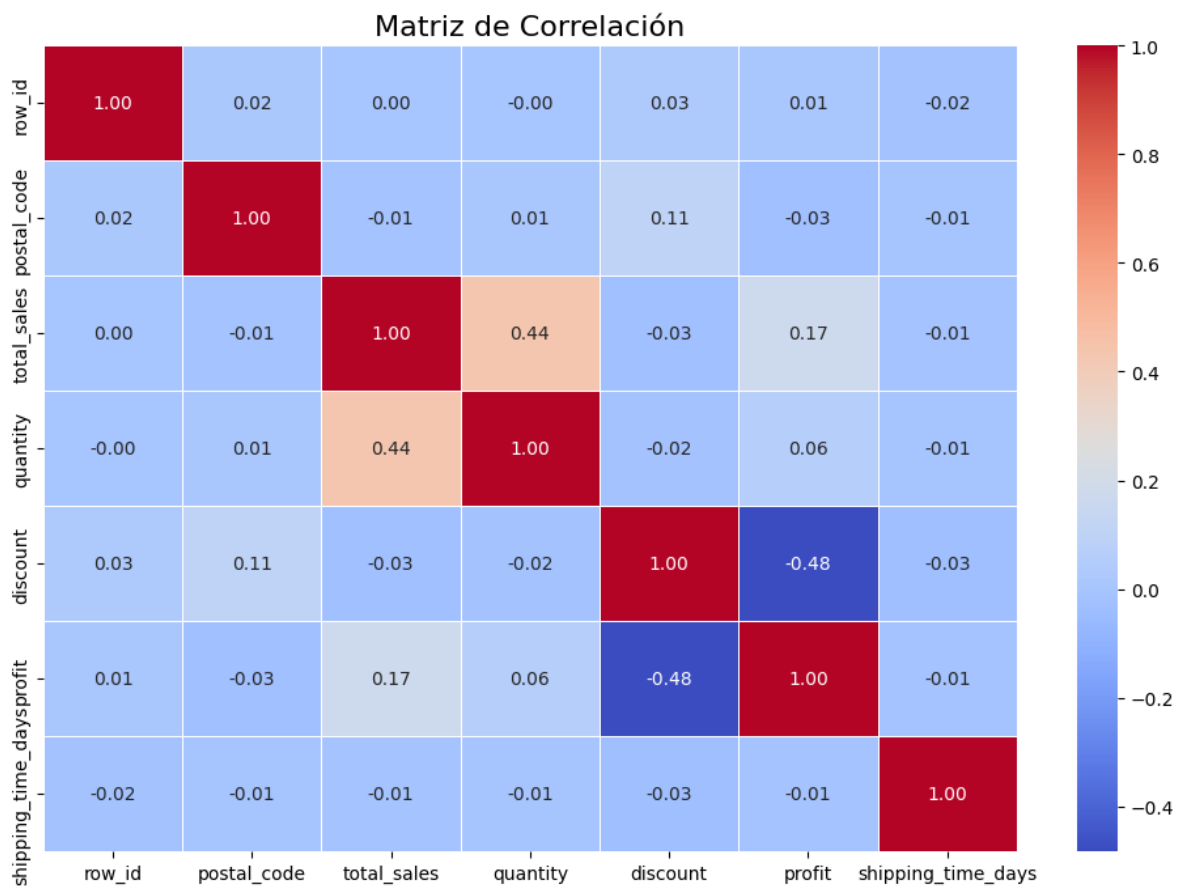
Correlación entre tiempos de envío y ventas totales: -0.009941838996710924



```
# Seleccionar solo las columnas numéricas
numeric_data = df.select_dtypes(include=['float64', 'int64'])

# Calcular la matriz de correlación
correlation_matrix = numeric_data.corr()

# Visualizar la matriz de correlación con un mapa de calor
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)
plt.title('Matriz de Correlación', fontsize=16)
plt.show()
```



5. Recomendar las mejores modalidades de envío para clientes de alto valor basándose en datos históricos.

```
# Definir clientes de alto valor (por ejemplo, aquellos con ventas totales mayores al percentil 75)
high_value_threshold = df['total_sales'].quantile(0.75)
high_value_customers = df[df['total_sales'] > high_value_threshold]

# Agrupar por modalidad de envío y calcular métricas
ship_mode_analysis = high_value_customers.groupby('ship_mode').agg(
    avg_profit=('profit', 'mean'),
    avg_sales=('total_sales', 'mean'),
    count=('ship_mode', 'size')
).sort_values(by='avg_profit', ascending=False)

# Mostrar los resultados
print(ship_mode_analysis)

# Visualizar los resultados
plt.figure(figsize=(12, 6))

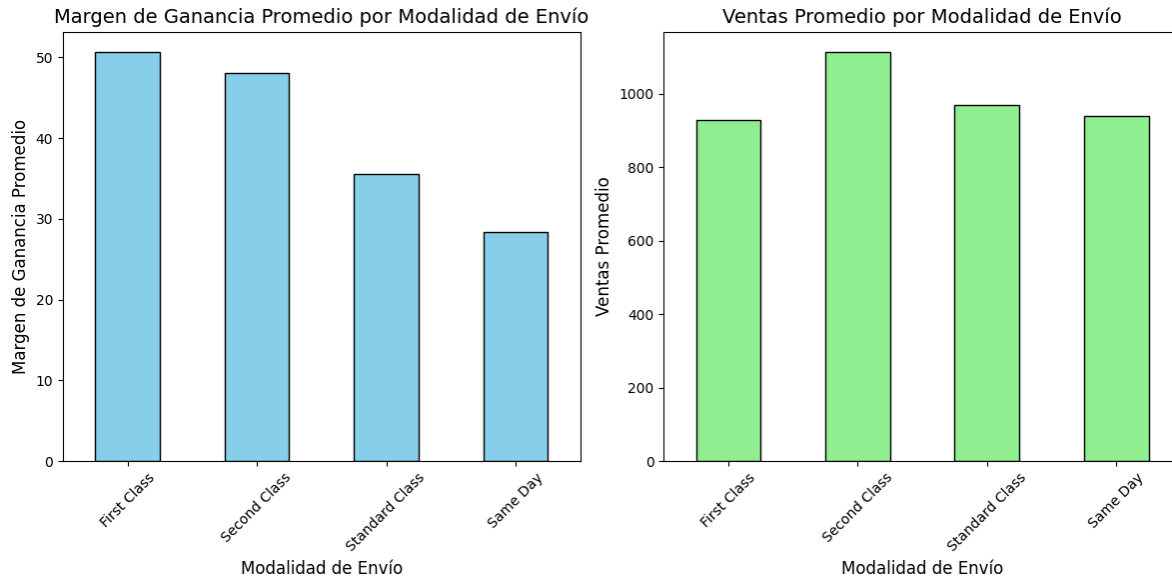
# Gráfico de barras para el margen de ganancia promedio
plt.subplot(1, 2, 1)
ship_mode_analysis['avg_profit'].plot(kind='bar', color='skyblue', edgecolor='black')
plt.title('Margen de Ganancia Promedio por Modalidad de Envío', fontsize=14)
plt.xlabel('Modalidad de Envío', fontsize=12)
plt.ylabel('Margen de Ganancia Promedio', fontsize=12)
plt.xticks(rotation=45)

# Gráfico de barras para las ventas promedio
plt.subplot(1, 2, 2)
ship_mode_analysis['avg_sales'].plot(kind='bar', color='lightgreen', edgecolor='black')
plt.title('Ventas Promedio por Modalidad de Envío', fontsize=14)
plt.xlabel('Modalidad de Envío', fontsize=12)
plt.ylabel('Ventas Promedio', fontsize=12)
plt.xticks(rotation=45)

plt.tight_layout()
plt.show()
```

	avg_profit	avg_sales	count
ship_mode			

First Class	50.637080	929.441188	82
Second Class	48.038505	1113.401527	101
Standard Class	35.510464	968.557922	317
Same Day	28.425138	939.001034	29



Conclusiones e Insights Basados en el Análisis Proporcionado:

Insights sobre el Tiempo de Envío:

La modalidad de envío Standard Class tiene el tiempo promedio de envío más largo, con 5 días, seguida por Second Class, que tiene un promedio de 3 días. Esto indica que la opción Standard Class es menos eficiente en términos de rapidez de entrega.

La región South presenta la mayor demora en los tiempos de envío, aunque no hay diferencias significativas en comparación con otras regiones. Esto sugiere que los factores regionales no influyen de manera notable en la eficiencia de las entregas.

Insights sobre Rentabilidad:

La modalidad de envío Second Class genera el mayor promedio de ganancias. Esto la convierte en la opción más rentable en general. Para los clientes de alto valor, la modalidad de envío First Class es la que más ganancias genera, lo que indica que debería ser priorizada para maximizar la rentabilidad de este segmento.

Correlaciones:

No se encontró correlación entre las ventas totales y el tiempo de envío. Esto implica que el comportamiento de compra de los clientes (en términos de valor de pedido) no se ve afectado por el tiempo que tarda el envío.

La matriz de correlación mostró relaciones débiles entre las diferentes columnas numéricas, lo que sugiere que hay pocas dependencias directas entre las métricas clave en el conjunto de datos.

Insights sobre Clientes de Alto Valor:

Entre los clientes de alto valor, la modalidad de envío Standard Class es la más utilizada (con 317 instancias). Esto podría deberse a su costo más bajo o a las preferencias de los clientes por entregas menos urgentes.

A pesar de ser la más utilizada, Second Class tiene el promedio de ventas más alto, lo que la convierte en una opción importante para la generación de ingresos en este grupo de clientes.

La modalidad First Class es la que más rentabilidad genera para los clientes de alto valor, lo que sugiere que debería ser priorizada al dirigirse a este segmento.

Recomendaciones:

1. Optimizar Métodos de Envío:

Enfocarse en promover la modalidad Second Class, ya que combina rentabilidad y tiempos de entrega relativamente rápidos.

2. Desarrollar Estrategias Regionales:

Aunque las diferencias regionales son mínimas en términos de tiempos de envío, se podrían ofrecer incentivos o promociones en regiones con mayores retrasos (como South).

3. Dirigir Estrategias a Clientes de Alto Valor:

Promover la modalidad First Class a clientes de alto valor para maximizar los márgenes de ganancia. Continuar ofreciendo Standard Class a aquellos clientes que valoran la eficiencia de costos sobre la rapidez de entrega, mientras se ofrecen opciones de mejora a Second Class o First Class para aquellos que valoran la velocidad.