Company Sales Evaluation

Guided By,
Madhavi Nimkar , EDS
Presented By,
Pranay Karkal -628
Onkar Katkamwar – 630
Janardhan Lavhale – 637
Viraj Mandlik - 639

Introduction

In the world of e-commerce, data plays a crucial role in understanding business performance, customer behavior, and market trends. As a data analyst for an e-commerce company, your task is to analyze a dataset containing sales data for various products. By leveraging the power of Python's NumPy and Pandas libraries, you will extract key insights to gain a deeper understanding of the company's performance, customer spending patterns, product popularity, and more.

Motivation

The motivation behind this project is to empower e-commerce companies with the tools and techniques to extract meaningful insights from their sales data. By performing a comprehensive analysis, companies can gain a competitive edge, enhance their decision-making capabilities, and better understand their customers and market dynamics. Through data-driven insights, companies can adapt their strategies, identify growth opportunities, and address challenges effectively.

Furthermore, data analysis provides a means to track key performance indicators (KPIs) and evaluate the success of marketing campaigns, product launches, and other business initiatives. By measuring and monitoring KPIs, companies can identify areas of improvement, optimize operations, and align their efforts with their strategic goals.

Details of Dataset

Transactio	Date	Customer	Product	Category	Quantity	Price
1	01-01-2023	Jay	Smartphor	Electronics	2	5000
2	02-01-2023	Viraj	Laptop	Electronics	1	10000
3	03-01-2023	Onkar	Headphon	Electronics	3	500
4	03-01-2023	Janardhan	T-Shirt	Clothing	5	200
5	04-01-2023	Pranay	Smartphor	Electronics	1	500
6	05-01-2023	Paras	Laptop	Electronics	1	10000
7	06-01-2023	Yash	Headphon	Electronics	3	500
8	07-01-2023	Punnet	Headphon	Electronics	1	150
9	08-01-2023	Angel	Headphon	Electronics	5	750
10	09-01-2023	Priya	T-Shirt	Clothing	6	1200

Data Manipulation

Data manipulation is a fundamental process in data analysis that involves transforming and preparing raw data to make it suitable for further exploration and analysis. It encompasses a range of operations aimed at ensuring data quality, consistency, and usability. Missing values can be imputed or removed, while outliers can be addressed through various methods such as transformation

```
import pandas as pd

# Load the dataset from a CSV file
df = pd.read_csv('sales_data.csv')
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Load the dataset

```
df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/sales_data.csv")
```

1) Calculate the total revenue generated by the company

```
df['Revenue'] = df['Quantity'] * df['Price']
total_revenue = df['Revenue'].sum()
print("Total revenue generated by the company: $", total_revenue)
```

```
Total revenue generated by the company: $ 45600
```

2) Determine the top 3 best-selling products based on the quantity sold

```
top_3_products = df.groupby('Product')['Quantity'].sum().nlargest(3)
print("Top 3 best-selling products:")
print(top_3_products)
```

```
Top 3 best-selling products:
Product
Headphones    12
T-Shirt       11
Smartphone     3
Name: Quantity, dtype: int64
```

3) Identify the customers who made the highest total spending and calculate their average spending per transaction

```
customer_spending = df.groupby('Customer')['Revenue'].sum()
highest_spending_customers = customer_spending.nlargest(3)
average_spending = highest_spending_customers / df.groupby('Customer')['Revenue'].count()
print("Customers with highest total spending:")
print(highest_spending_customers)
print("\nAverage spending per transaction:")
print(average_spending)
```

```
Customers with highest total spending:
Customer
Jay      10000
Paras    10000
Viraj    10000
Name: Revenue, dtype: int64

Average spending per transaction:
Customer
Angel            NaN
Janardhan        NaN
Jay          10000.0
Onkar            NaN
Paras        10000.0
Pranay           NaN
Priya            NaN
Punnet           NaN
Viraj        10000.0
Yash             NaN
Name: Revenue, dtype: float64
```

4) Calculate the monthly sales growth rate for the company

```
df['Date'] = pd.to_datetime(df['Date'])
monthly_sales = df.groupby(df['Date'].dt.to_period('M'))['Revenue'].sum()
monthly_growth_rate = monthly_sales.pct_change()
print("Monthly sales growth rate:")
print(monthly_growth_rate)
```

```
Monthly sales growth rate:
Date
2023-01      NaN
2023-02     0.00
2023-03    -0.75
2023-04    -0.80
2023-05    19.00
2023-06    -0.85
2023-07    -0.90
2023-08    24.00
2023-09     0.92
Freq: M, Name: Revenue, dtype: float64
```

5) Analyze the product sales distribution across different product categories

```
product_category_sales = df.groupby('Category')['Quantity'].sum()
print("Product sales distribution across different categories:")
print(product_category_sales)
```

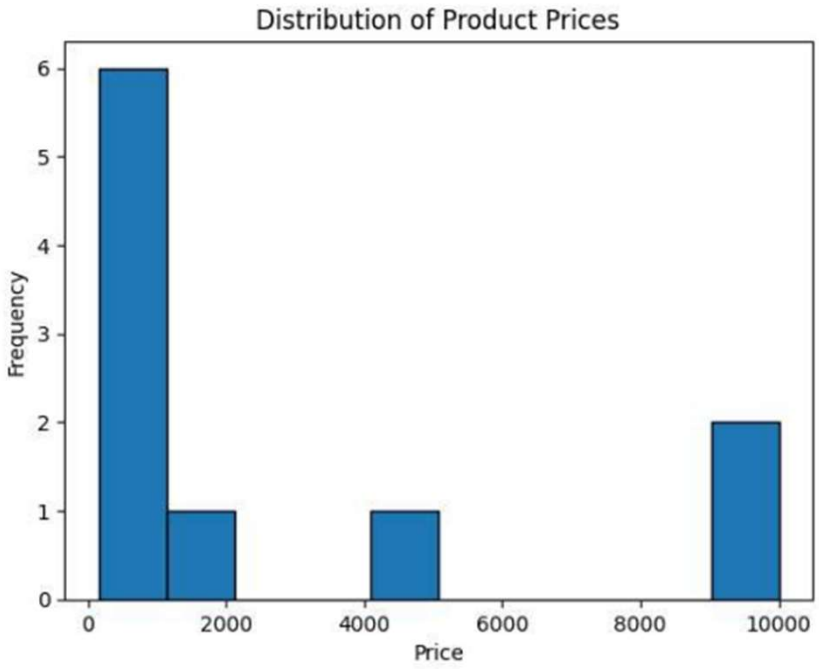
```
Product sales distribution across different categories:
Category
Clothing       11
Electronics    17
Name: Quantity, dtype: int64
```

Data Visualization

Data visualization is the process of representing data and information visually through charts, graphs, maps, and other graphical elements. It is a powerful technique that allows us to effectively communicate complex concepts, patterns, and trends in a visual format. Data visualization transforms complex data into visual representations that enhance understanding, reveal patterns, and support decision-making.

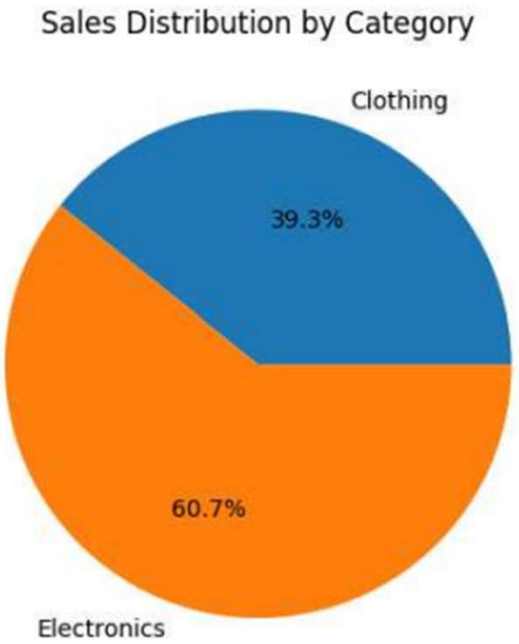
6) Generate a histogram to visualize the distribution of product prices

```
plt.hist(df['Price'], bins=10, edgecolor='black')
plt.xlabel('Price')
plt.ylabel('Frequency')
plt.title('Distribution of Product Prices')
plt.show()
```



7) Create a pie chart to represent the sales distribution across different product categories

```
plt.pie(product_category_sales, labels=product_category_sales.index, autopct='%1.1f%%')
plt.title('Sales Distribution by Category')
plt.show()
```



Predictive Technique (KMeans)

```
from sklearn.cluster import KMeans
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset (replace with your own data)
df = pd.read_csv("sales_data.csv")

# Select the relevant columns for clustering
data = df[["Quantity", "Price"]]

# Define the number of clusters
num_clusters = 4

# Create a K-means clustering model
kmeans = KMeans(n_clusters=num_clusters)

# Fit the data to the model
kmeans.fit(data)

# Get the cluster labels for each data point
cluster_labels = kmeans.labels_

# Add the cluster labels to the DataFrame
df["Cluster"] = cluster_labels

# Visualize the clusters
plt.scatter(df["Quantity"], df["Price"], c=cluster_labels)
plt.xlabel("Quantity")
plt.ylabel("Price")
plt.title("Product Clustering")
plt.show()

# Print the cluster assignments for each product
print("Cluster Assignments:\n", df[["Product", "Cluster"]])
```

EDS_MINI.PROJECT | mini_project.ipynb - Colaborator

colab.research.google.com/drive/1ypyoJzqFhy0QUH662k49dRGFZViW2yJg#scrollTo=knHSt8nHXUIR

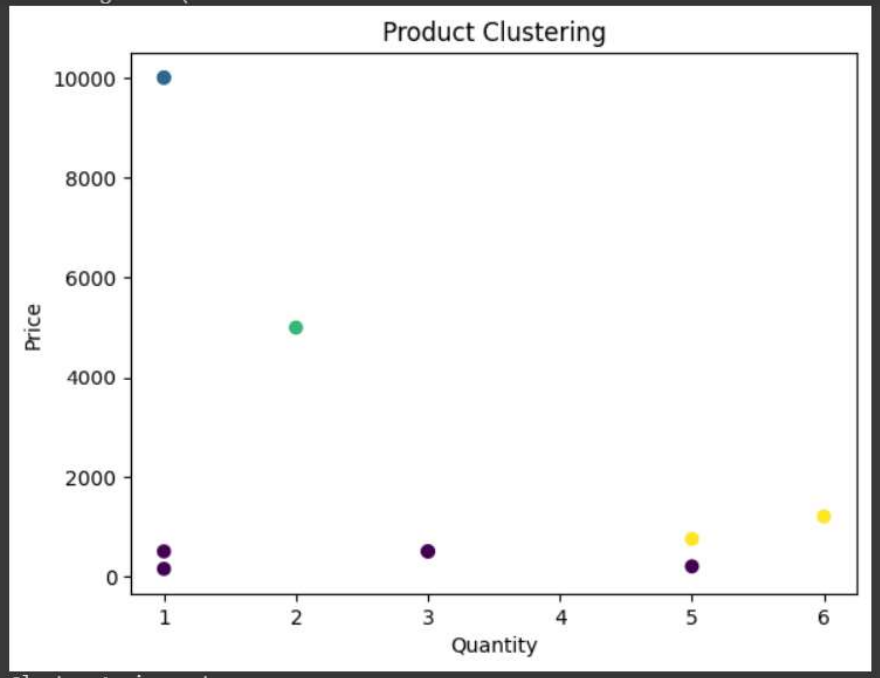
mini_project.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk



Cluster Assignments:

	Product	Cluster
0	Smartphone	2
1	Laptop	1
2	Headphones	0
3	T-Shirt	0
4	Smartphone	0
5	Laptop	1
6	Headphones	0
7	Headphones	0
8	Headphones	3
9	T-Shirt	3

2s completed at 4:46 PM

Application

The e-commerce data analysis project has various practical applications in the context of an e-commerce company. Here are some key applications:

1.Business Performance Evaluation: By analyzing sales data, an e-commerce company can evaluate its overall business performance. This includes assessing total revenue, identifying best-selling products, and understanding sales growth rates. These insights help businesses make informed decisions regarding inventory management, pricing strategies, and resource allocation.

2.Customer Behavior Analysis: Analyzing customer spending patterns and identifying high-spending customers allows e-commerce companies to tailor their marketing strategies, improve customer experiences, and build customer loyalty. It helps in identifying the most profitable customer segments and creating targeted marketing campaigns.

Conclusion

In conclusion,

By performing various data manipulation tasks, we were able to calculate total revenue, identify best-selling products, analyze customer spending patterns, evaluate sales growth rates, understand product category distribution, and utilize clustering algorithms for market segmentation.

The e-commerce data analysis project empowers the company with valuable insights to optimize their business operations, enhance customer experiences, and make data-driven decisions to stay competitive in the dynamic e-commerce landscape.