Data Analysis Report

Introduction

This report demonstrates practical data analysis and visualization using Python libraries pandas, matplotlib, and seaborn. We work with a dataset containing fields like Customer_ID, Product_ID, Purchase_Frequency, Average_Order, etc., focusing on extracting insights through code examples and graphical visualizations. This concise report includes selected examples with explanations.

Section 1: Pandas Examples

Example 1: Load dataset and check summary

```
import pandas as pd
df = pd.read_csv('sales_data.csv')
print(df.info())
```

Explanation: Loads the dataset and prints data types, non-null counts, and memory usage.

Example 2: Filter customers with Purchase_Frequency > 5

```
filtered_df = df[df['Purchase_Frequency'] > 5]
print(filtered_df.head())
```

Explanation: Selects customers with purchase frequency greater than 5 for targeted analysis.

Example 3: Create Total Sales column

```
df['Total_Sales'] = df['Average_Order'] * df['Purchase_Frequency']
print(df[['Customer_ID', 'Total_Sales']].head())
```

Explanation: Calculates total sales per customer by multiplying average order value with frequency.

Example 4: Group by Product_ID to get total sales

```
total_sales = df.groupby('Product_ID')['Total_Sales'].sum()
print(total_sales)
```

Explanation: Aggregates total sales by product to identify best-sellers.

Example 5: Sort customers by Average_Order descending

```
sorted_df = df.sort_values(by='Average_Order', ascending=False)

print(sorted_df.head())
```

Explanation: Sorts customers to find those with highest average order values.

Section 2: Matplotlib Examples

Example 1: Histogram of Purchase Frequency

```
import matplotlib.pyplot as plt

plt.hist(df['Purchase_Frequency'], bins=10, color='skyblue')

plt.title('Purchase Frequency Distribution')

plt.xlabel('Purchase Frequency')

plt.ylabel('Number of Customers')

plt.show()
```

Explanation: Visualizes the distribution of how often customers make purchases.

Example 2: Bar Chart of Total Sales by Product

```
total_sales.plot(kind='bar', color='green')

plt.title('Total Sales per Product')

plt.xlabel('Product ID')

plt.ylabel('Total Sales')

plt.show()
```

Explanation: Displays sales totals for each product to identify top performers.

Example 3: Line Plot of Average Order over Transactions

```
plt.plot(df['Transaction_ID'], df['Average_Order'], color='red')

plt.title('Average Order over Transactions')

plt.xlabel('Transaction ID')

plt.ylabel('Average Order Value')

plt.show()
```

Explanation: Tracks how average order values change across transactions.

Example 4: Scatter Plot of Purchase Frequency vs Average Order

```
plt.scatter(df['Purchase_Frequency'], df['Average_Order'], color='purple')

plt.title('Purchase Frequency vs Average Order')

plt.xlabel('Purchase Frequency')

plt.ylabel('Average Order')
```

```
plt.show()
```

Explanation: Shows relationship between how often customers buy and how much they spend.

Example 5: Pie Chart of Product Sales Distribution

```
product_sales = df.groupby('Product_ID')['Total_Sales'].sum()
plt.pie(product_sales, labels=product_sales.index, autopct='%1.1f%%')
plt.title('Sales Distribution by Product')
plt.show()
```

Explanation: Illustrates the proportion of sales contributed by each product.

Section 3: Seaborn Examples

Example 1: Boxplot of Average Order by Product

```
import seaborn as sns
sns.boxplot(x='Product_ID', y='Average_Order', data=df)
plt.title('Average Order Distribution by Product')
plt.show()
```

Explanation: Displays distribution and outliers of average orders for each product.

Example 2: Heatmap of Correlation Matrix

```
corr = df.corr()
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Feature Correlation Heatmap')
plt.show()
```

Explanation: Visualizes correlations between numerical features to find relationships.

Example 3: Countplot of Customers by Purchase Frequency

```
sns.countplot(x='Purchase_Frequency', data=df)
plt.title('Customer Count by Purchase Frequency')
plt.show()
```

Explanation: Shows how many customers fall into each purchase frequency category.

Example 4: Scatter Plot with Regression Line

```
sns.regplot(x='Purchase_Frequency', y='Average_Order', data=df)
plt.title('Regression: Purchase Frequency vs Average Order')
plt.show()
```

Explanation: Plots the linear relationship and trend between frequency and order value.

Example 5: Violin Plot of Total Sales by Product

```
sns.violinplot(x='Product_ID', y='Total_Sales', data=df)
plt.title('Total Sales Distribution by Product')
```

```
plt.show()
```

Explanation: Combines boxplot and density plot to show distribution shape of sales.

Summary and Conclusion

This report provided selected examples of data analysis and visualization techniques using pandas, matplotlib, and seaborn. Through filtering, grouping, and sorting, pandas helped manipulate data efficiently. Matplotlib and seaborn provided insightful visualizations such as histograms, bar charts, scatter plots, boxplots, and heatmaps. Together, these tools offer a powerful workflow for extracting insights and communicating results effectively from sales datasets or similar data sources.