**Coffee Shop Sales Analysis**

This project is focused on analyzing the sales data of a coffee shop using Python's Pandas, NumPy, Matplotlib, and Seaborn libraries. The dataset contains various details of transactions, including the transaction date, time, store location, product details, and sales amounts. The aim is to perform an exploratory data analysis (EDA), handle missing data, perform data cleaning, and generate visualizations to uncover key insights about sales trends across different store locations and time periods.

**Project Structure**

**Data**

The dataset (Coffee Shop Sales.xlsx) includes the following columns:

* **transaction\_id**: Unique ID for each transaction.
* **transaction\_date**: Date of the transaction.
* **transaction\_time**: Time when the transaction occurred.
* **transaction\_qty**: Quantity of products sold.
* **store\_id**: ID of the store where the transaction occurred.
* **store\_location**: Name of the store's location.
* **product\_id**: ID of the product sold.
* **unit\_price**: Price per unit of the product.
* **product\_category**: Category of the product (e.g., Coffee, Tea, etc.).
* **product\_type**: Type of the product within the category.
* **product\_detail**: Detailed product description.

**Data Analysis Process**

1. **Data Loading**: The dataset is loaded into a Pandas DataFrame for analysis.

df = pd.read\_excel("Coffee Shop Sales.xlsx")

1. **Missing Values**: Checked for any missing data.

missing\_values = df.isnull().sum()

1. **Data Cleaning**:
   * Converted transaction\_time to a proper time format using pd.to\_timedelta().
   * Calculated total sales by multiplying transaction\_qty with unit\_price.
   * Created a new datetime column by combining transaction\_date and transaction\_time.
2. **Descriptive Statistics**: A summary of key statistics like mean, min, max, etc., for numerical columns using .describe().

df.describe()

1. **Outlier Detection**: Identified outliers using the Interquartile Range (IQR) method and visualized them using boxplots.

sns.boxplot(x=df[column])

1. **Data Aggregation**:
   * Grouped the data by store\_location to analyze total sales and number of transactions.
   * Calculated daily sales across different store locations.
2. **Visualizations**:
   * **Boxplot**: Visualized outliers for transaction\_qty, unit\_price, and sales.
   * **Line Plot**: Showed daily sales trends by store location.
   * **Stacked Area Chart**: Visualized cumulative daily sales across different locations.
3. **Time-Based Analysis**:
   * Grouped sales data by day of the week, month, quarter, and year.
   * Created visualizations to explore how sales vary by these time periods.

**Visualizations**

* **Daily Sales by Store Location**: Shows how daily sales fluctuate across different store locations.
* **Weekly, Monthly, Quarterly Sales**: Analyzes how sales change over different time frames to uncover trends.

**Libraries Used**

* **Pandas**: For data manipulation and analysis.
* **NumPy**: For numerical operations.
* **Matplotlib**: For generating plots.
* **Seaborn**: For creating advanced visualizations.

**Running the Project**

1. Clone this repository.
2. Install the required libraries:

pip install pandas numpy matplotlib seaborn

1. Run the analysis script to load the dataset, perform cleaning, analysis, and generate visualizations.

**Key Insights**

* Significant sales differences between various store locations.
* A clear trend in sales growth or decline over certain time periods (e.g., weekly or monthly).
* Identification of peak sales times and store locations with the highest revenue.

**Future Enhancements**

* Incorporate more advanced forecasting techniques.
* Analyze customer purchase patterns and preferences.