**Documentation: Methodology, Analysis, and Insights**

**Methodology for Python and Dataframe:**

**1. Data Collection & Preparation:**

* **Data Source:** The data is sourced from SQL tables in the MySQL database named EXPENSE\_DB. The table expense\_data contains columns such as Date, Category, Payment\_Mode, Description, Amount, and Cashback. Separate sheets are generated from the SQL table data, merged into a single file, and saved as a consolidated CSV file.
* **Data Loading:** The consolidated CSV file is then loaded into a Pandas DataFrame using pandas.read\_csv, allowing for seamless analysis and processing in Python.

**2. Data Cleaning:**

* **Date Formatting:** Convert the Date column to a standardized format (YYYY-MM-DD) using pd.to\_datetime to ensure consistency.
* **Handling Missing Values:** Null or missing values are filled with appropriate defaults or dropped, as necessary, to ensure data integrity.

**3. SQL Query Execution & Transformation:**

* **Predefined Queries:** A set of SQL queries is defined to answer specific business questions such as total expenses, spending trends, top categories, payment modes, etc.
* **Query Execution:** Each query is executed using pandas.read\_sql, fetching results into a DataFrame.
* **Median Calculation:** A custom SQL subquery is used to calculate the median efficiently by ranking the Amount column.

**4. Analysis:**

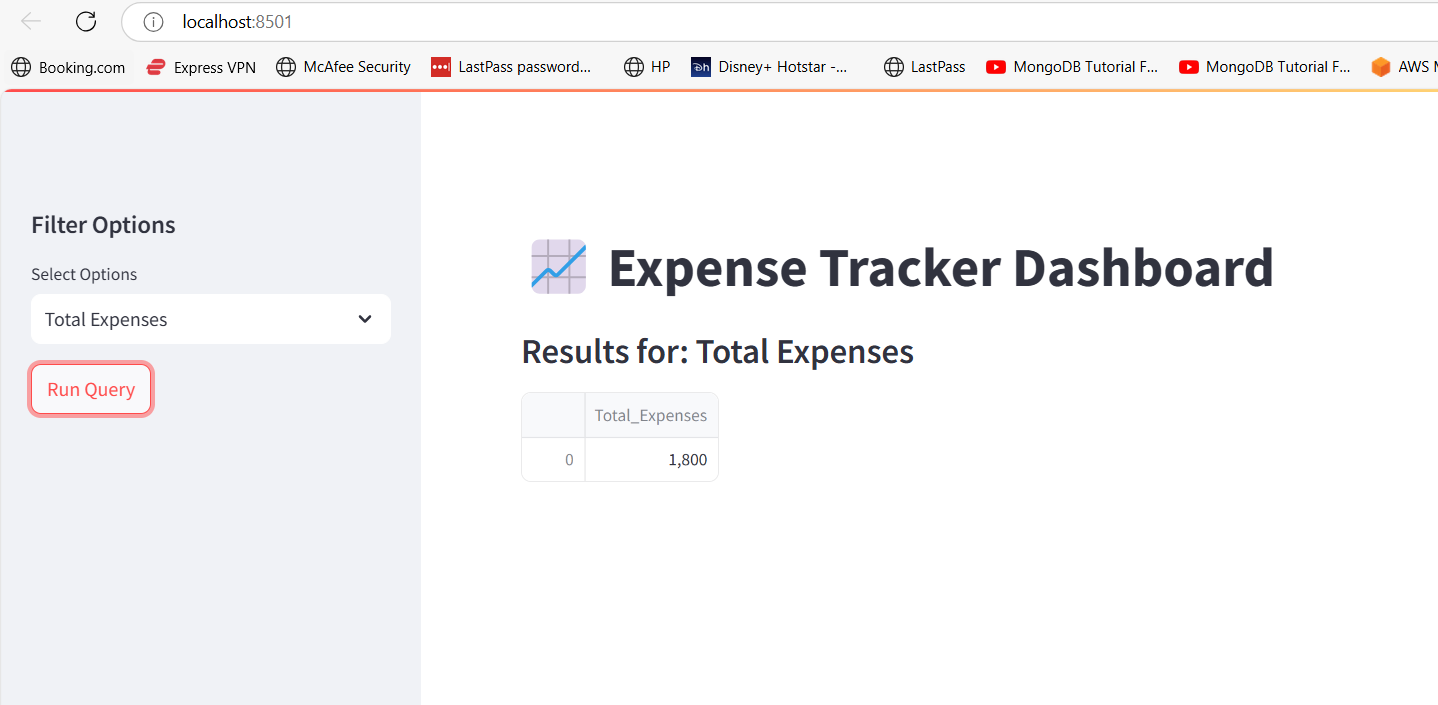
* **Group By & Aggregation:** SQL queries often use GROUP BY to aggregate data by category, payment mode, or date. Sum and average operations are performed to calculate total and average spending.
* **Median Analysis:** A custom median query is used to identify expenses above or below the median, helping distinguish between typical and extreme spending.
  + **Total Expenses:**  
    Displays the total number of records in the expense\_data table.
  + **Total Amount Spent:**  
    Calculates the sum of all Amount fields to show the total money spent.
  + **Average Expense:**  
    Computes the average amount spent per transaction.
  + **Top 5 Categories by Expense:**  
    Provides the top 5 categories contributing the most to the expenses.
  + **Monthly & Daily Spending Trends:**  
    Shows trends in monthly and daily expenses to help track how spending varies over time.
  + **Expenses by Payment Mode:**  
    Identifies which payment modes (e.g., cash, credit card, etc.) are used most frequently.
  + **Highest & Lowest Single Expense:**  
    Identifies the maximum and minimum single expenses to highlight extreme financial transactions.
  + **Expenses in the Last 7 Days:**  
    Shows expenses incurred in the past 7 days to monitor recent spending behaviours.
  + **Most Frequent Categories:**  
    Identifies which categories have the highest frequency of expenses.
  + **Median Analysis (Above and Below Median):**  
    Identifies transactions that are either above or below the median amount to highlight typical and extreme spending behaviours.
  + **Cashback Utilization by Payment Mode:**  
    Analyses how much cashback has been received by different payment methods.

**5. Insights:**

* **Top 5 Categories Insight:**  
  The top 5 categories with the highest total expenses often reflect the largest portions of spending, helping users understand where the majority of their money goes.
* **Payment Mode Analysis:**  
  Payment mode analysis provides insights into how different payment methods contribute to overall spending. For example, if cash is the dominant mode, it could indicate limited use of digital payment options.
* **Spending Trends:**  
  Tracking monthly and daily spending trends helps to identify patterns such as seasonal spending spikes or periods of reduced activity.
* **Median Analysis:**  
  The median analysis helps to distinguish between typical and unusually high or low expenses. Expenses above the median indicate higher-than-usual spending, while those below the median suggest consistent or moderate spending behaviours.
* **Cashback Utilization:**  
  Payment modes with higher cashback usage indicate greater incentives being utilized by the user, potentially influencing their choice of payment methods.

**Screenshots of the Streamlit App:**

* **Total Expenses:**
* **Visualization:** A simple table showing the total number of expenses.



* **Total Amount Spent:**
* **Visualization:** A number showing the total amount spent.

A screenshot of a computer

Description automatically generated

* **Average Expense:**
* **Visualization:** A number showing the total amount spent.

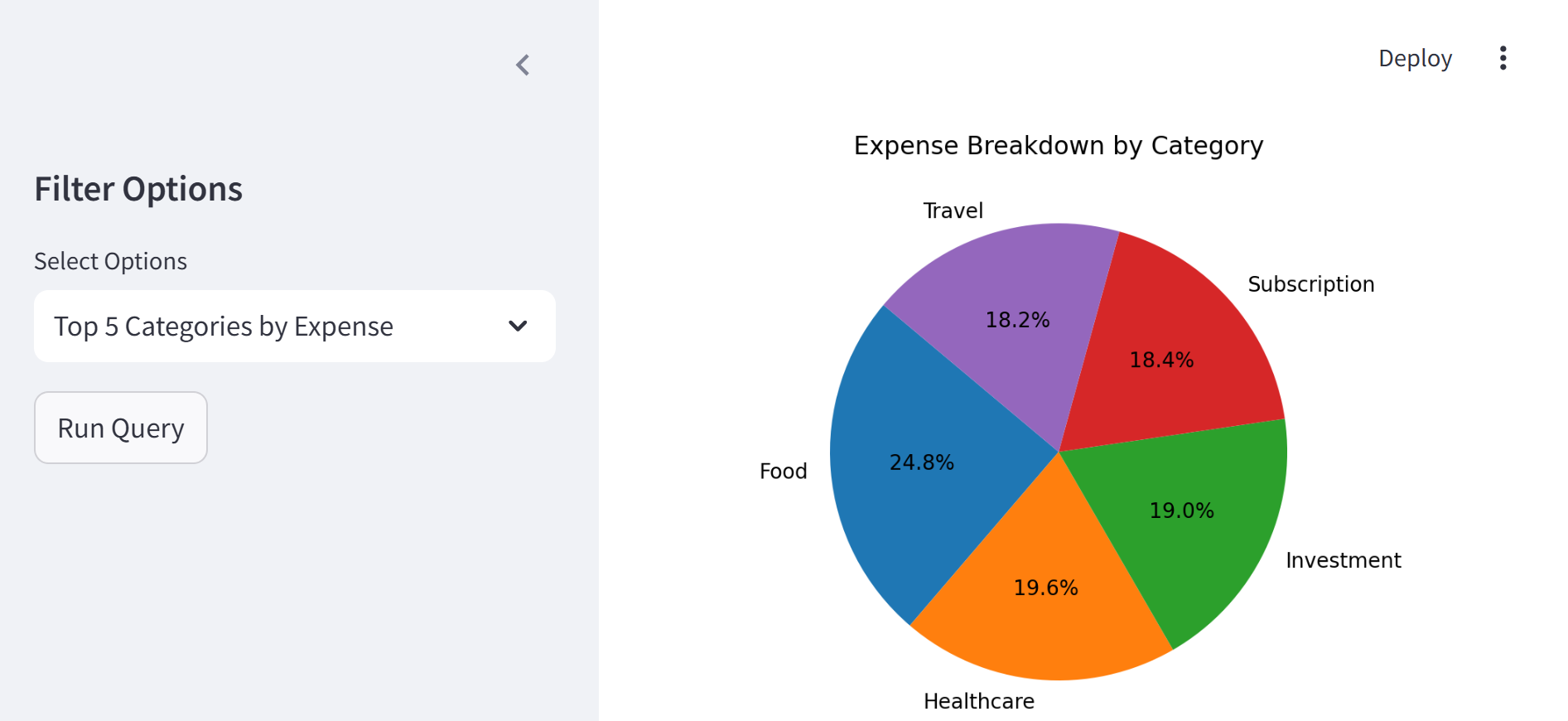
A screenshot of a computer

Description automatically generated

* **Top 5 Categories by Expense:**
* **Visualization:** A pie chart displaying the top 5 expense categories.

A screenshot of a computer

Description automatically generated



* **Monthly Spending Trend:**
* **Visualization:** Line chart showing total monthly spending trends.

A screenshot of a graph

Description automatically generated

* **Daily Spending Trend:**
* **Visualization:** Line chart showing total monthly spending trends.

A screenshot of a graph

Description automatically generated

* **Expenses by Payment Mode:**
* **Visualization:** Bar chart showing expenses categorized by payment modes.

A screenshot of a graph

Description automatically generated

* **Expenses Above Median Amount:**
* **Visualization:** Histogram displaying expenses above the median.

A screenshot of a graph

Description automatically generated

* **Expenses Below Median Amount:**
* **Visualization:** Histogram displaying expenses below the median.

A screenshot of a graph

Description automatically generated

* **Cashback Utilization by Payment Mode:**
* **Visualization:** Bar chart showing cashback usage by payment modes.

A screenshot of a graph

Description automatically generated

* **Yearly Spending Summary:**
* **Visualization:** Bar chart displaying yearly spending trends.

A screenshot of a graph

Description automatically generated

**6. Visualization:**

* **Matplotlib & Seaborn:** The resulting DataFrames are visualized using matplotlib and seaborn to generate appropriate charts like pie charts, bar charts, and line plots.
* **Dynamic Visualization:** Different queries trigger different visualizations based on the structure of the query results.

**7. Result Display:**

* **Displaying Results:** Results are displayed as data tables in Streamlit for quick user reference.
* **Interactive Dashboard:** Streamlit is used to build a web-based dashboard that displays interactive visualizations and insights.

**8. Error Handling:**

* **Exception Handling:** Any exceptions, such as database errors or data formatting issues, are caught and displayed to ensure smooth user interaction.