| **Project Title** | **Analyzing Personal Expenses** |
| --- | --- |
| **Skills take away From This Project** | Python, SQL, and Streamlit |
| **Domain** | Personal Finance and Expense Tracking |

**Problem Statement:**

This project aims to simulate an expense tracker for an individual using the **Faker** library. The project generates realistic monthly expense data, processes and stores it in a SQL database, and creates SQL queries to derive insights into spending behavior. A Streamlit app is developed to visualize these insights and showcase the results of SQL queries. The tracker will highlight expenses across categories like bills, groceries, subscriptions, and personal spending, providing a comprehensive overview of financial habits over a year.

**Business Use Cases:**

* Automating the tracking of personal or business expenses from e-commerce platforms.
* Analyzing and categorizing spending habits to create actionable savings plans.
* Building financial dashboards for tracking income and expenditure trends.Providing businesses insights into procurement and inventory purchasing patterns.

**Approach:**

* **Data Simulation**: Use the **Faker** library to generate a realistic dataset that depicts a person’s expense throughout the month.create 12 different tables for each month.
* **Database Creation**: Create a SQL database schema and load the generated dataset for querying.
* **EDA**: Analyze the dataset using Python libraries to extract insights about spending patterns and trends.
* **Streamlit App**: Develop a user-friendly web application showcasing visualizations and SQL query outputs.
* **Insights & Recommendations**: Provide actionable takeaways based on simulated data analysis.

**Questions to be answered:**

Analyze your spending habits by answering the following 15 questions using SQL queries and visualizing the results in a Streamlit application. **Additionally, formulate your own 10-15 insightful queries to further explore your spending patterns.**

1. What is the total amount spent in each category?
2. What is the total amount spent using each payment mode?
3. What is the total cashback received across all transactions?
4. Which are the top 5 most expensive categories in terms of spending?
5. How much was spent on transportation using different payment modes?
6. Which transactions resulted in cashback?
7. What is the total spending in each month of the year?
8. Which months have the highest spending in categories like "Travel," "Entertainment," or "Gifts"?
9. Are there any recurring expenses that occur during specific months of the year (e.g., insurance premiums, property taxes)?
10. How much cashback or rewards were earned in each month?
11. How has your overall spending changed over time (e.g., increasing, decreasing, remaining stable)?
12. What are the typical costs associated with different types of travel (e.g., flights, accommodation, transportation)?
13. Are there any patterns in grocery spending (e.g., higher spending on weekends, increased spending during specific seasons)?
14. Define High and Low Priority Categories
15. Which category contributes the highest percentage of the total spending?

**Results:**

* A functional Streamlit app showcasing visualizations of spending patterns and outputs of 15 above mentioned and your own 10-15 SQL queries.
* Identification of spending trends, such as top categories and monthly expenditure breakdown.
* Data-driven insights into optimizing expenses and managing finances effectively.

**Project Evaluation metrics:**

* Completeness of data migration and SQL integration.
* Accuracy of the SQL queries and EDA results.
* Functionality and usability of the Streamlit app.
* Visual appeal and informativeness of the charts and tables in the app.
* Adherence to coding and project submission standards.

**Technical Tags:**

Python, SQL, Streamlit, EDA, Financial Analysis, Data Visualization, Expense Tracking

**Data Set:**

1. **Date**: The transaction date.
2. **Category**: Type of expense (Food, Transportation, Bills, etc.).
3. **Payment Mode**: Specifies whether it was a **Cash** or **Online** transaction.
4. **Description**: Details about the expense.
5. **Amount Paid**: Total amount paid for the transaction.
6. **Cashback**: Cashback received (if any) during the transaction.

Example table is given below.



**Data Set Explanation:**

The dataset is a simulated representation of a person’s expenses generated with the Faker library. It is designed for hands-on analysis and visualization.  
Key preprocessing steps include:

* Ensuring realistic date ranges and amounts.
* Categorizing products accurately.
* Structuring data for easy querying and visualization.

**Project Deliverables:**

* Source code for data cleaning, SQL integration, EDA, and Streamlit app.
* SQL scripts for all 20 queries.
* Documentation explaining the methodology, analysis, and insights.
* Screenshots of the Streamlit app with key visualizations and outputs.

**Project Guidelines:**

* Follow Python best practices, including proper naming conventions and commenting.
* Use version control (e.g., GitHub) to track progress and manage the project repository.
* Ensure the Streamlit app is user-friendly and interactive.
* Validate SQL queries to ensure accurate outputs.

**Timeline:**

**Day 1:** Simulate data using the Faker library and design the database schema.

**Day 2:** Load the simulated data into the SQL database and create queries.

**Day 4:** Develop the Streamlit app and integrate SQL queries.

**Day 6:** Finalize documentation, insights, and submit the project.

**Check your mail for the submission deadline of the project.**

**REFERENCES:**

| Project Live Evaluation Metrics | [Project Live Evaluation](https://docs.google.com/document/u/0/d/1QisLD2kqDWFZJG2oDknKn2eMGi-Xq8oFPgA7UWSbcIQ/edit) |
| --- | --- |
| EDA Guide | [Exploratory Data Analysis (EDA) Guide](https://docs.google.com/document/d/1tHiTU1X9UwXSLySpJ-FVCohlf_8xpXwa75vlK9S6wl8/edit?usp=sharing) |
| Capstone Explanation Guideline | [Capstone Explanation Guideline](https://docs.google.com/document/d/1gbhLvJYY7J73lu1g9c6C9LRJvYemiDOdRDAEMe632w8/edit) |
| GitHub Reference | [How to Use GitHub.pptx](https://docs.google.com/presentation/d/1XHCbgUOqbcXNUyQ87vTlKdKRgAbBxtkA/edit?usp=sharing&ouid=109735616107417446342&rtpof=true&sd=true) |
| HOW TO ESTABLISH SQL CONNECTION | [PYTHONSQLCODE\_TAMIL.ipynb](https://drive.google.com/file/d/1ZdgP0OgpxamIKJzAPspZkrExD0JvZb-A/view?usp=sharing) |
| STREAMLIT RECORDING (Tamil) | https://us06web.zoom.us/rec/share/JTr7DywhE1-SarjyIHBSCn4qnl7\_uvJH6IGk06qAlkE0Ny1o\_rqcq5FRFKuo93dm.iyM2o6l0h9aTUkNE |
| STREAMLIT RECORDING (English) | [Special session for STREAMLIT(11/08/2024)](https://docs.google.com/document/d/1aR3pUZFlCi8gicpF6aPHPESeFdOtGMlfob5PckresZk/edit?usp=sharing) |
| STREAMLIT DOCUMENTATION | [Install Streamlit - Streamlit Docs](https://docs.streamlit.io/get-started/installation) |
| FAKER LIBRARY DOCUMENTATION | [Faker’s documentation!](https://faker.readthedocs.io/en/master/)  [Faker · PyPI](https://pypi.org/project/Faker/) |
| Project Orientation (English) | [expense tracker project orientation.mp4](https://drive.google.com/file/d/1cgmwkbFPA0IjTsk4MNrIbz114S-1ZO0T/view?usp=sharing) |
| Project Orientation (Tamil) |  |

**PROJECT DOUBT CLARIFICATION SESSION ( PROJECT AND CLASS DOUBTS)**

**About Session:** The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.

**Note: Book the slot at least before 12:00 Pm on the same day**

**Timing: Monday-Saturday (4:00PM to 5:00PM)**

**Booking link :**[**https://forms.gle/XC553oSbMJ2Gcfug9**](https://forms.gle/XC553oSbMJ2Gcfug9)

**LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)**

**About Session:** The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.

**Note: This form will Open on Saturday and Sunday Only on Every Week**

**Timing:**

**For DS and AIML**

**Monday-Saturday (05:30PM to 07:00PM)**

**Booking link :** [**https://forms.gle/1m2Gsro41fLtZurRA**](https://forms.gle/1m2Gsro41fLtZurRA)

**Evaluation Metrics** : [Project Live Evaluation](https://docs.google.com/document/d/1QisLD2kqDWFZJG2oDknKn2eMGi-Xq8oFPgA7UWSbcIQ/edit?usp=sharing)