Analyzing Personal Expenses

# 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.

# Technologies Used:

* Faker: To generate synthetic data.
* Python: Language for coding.
* Microsoft Vscode: For IDE purpose
* Anaconda: For Environments and packages.
* pandas: To manage data in tabular form and generate DataFrames.
* matplotlib: To create visualizations like bar charts and pie charts.
* mysql.connector: For interacting with the MySQL database.
* streamlit: For creating the interactive web interface.
* MySQL Database: The generated data is imported into a MySQL database (expensetrackerdb), and various SQL queries are executed to analyze and extract insights.

# Methodology

**Data Generation:**

* Faker Library: The Faker library is used to generate synthetic data for the personal expenses tracker. This ensures that the data remains realistic yet entirely fictional, making it perfect for testing and analysis purposes. The randomness in data generation mimics real-world spending patterns while maintaining control over categories, subcategories, and transaction attributes.
* Categories and Subcategories: Expenses are classified into categories such as Groceries, Bills, Travel, Healthcare, etc., each with its own set of subcategories (e.g., Groceries → Fruits, Dairy, etc.). This hierarchical structure helps in organizing transactions more effectively.
* Transactions Generation: For each month, a set number of random transactions (100 in this case) is created, with attributes like:
* Date: Randomly assigned based on the month.
* Payment Mode: Transactions are assigned payment modes such as UPI, Credit Card, etc.
* Description: Descriptions are generated based on the category.
* Amount Paid: Random amounts are assigned depending on the category.
* Cashback: A cashback value is assigned based on the payment mode (e.g., higher cashback for UPI or Credit Card transactions).
* File Storage: The generated data for each month is saved as CSV files and later combined into a single file representing the full year of transactions (2023).

**SQL Integration:**

*SQL Queries:* A set of SQL queries is designed to analyze different aspects of the personal expense data. The queries include:

* Total spending by category.
* Monthly spending trends.
* Spending patterns across different payment modes.
* Transactions with cashback values.
* Category-based average spending.
* Daily, monthly, and seasonal spending trends.
* These queries are executed on a MySQL database, where the data is loaded and analyzed using SQL. For each query, users can select from a list of pre-configured queries, which will be executed and the results displayed.

**Data Visualization:**

*Streamlit:* The Streamlit framework is used to create an interactive web application for data visualization and user interaction.

*Visualizations:*

* Bar Chart: Displays total expenses by category, helping users understand the major contributors to their spending.
* Pie Chart: Illustrates the distribution of payment modes, allowing users to see which modes (Cash, UPI, Credit Card, etc.) are most frequently used.
* Stacked Bar Chart: Shows how different payment modes contribute to expenses in each category, making it easier to analyze spending behavior.
* Line Graph (Cashback): Visualizes the distribution of cashback amounts, showing trends in cashback across all transactions.

# Insights and Analysis:

* Spending Trends: The analysis shows trends over time, such as which categories (e.g., Groceries, Entertainment) dominate a user's spending habits, whether cashback influences the frequency of UPI or Credit Card payments, and how seasonal factors might affect spending.
* Payment Mode Usage: Insights into which payment modes are preferred by users, and whether there’s a correlation between the mode of payment and spending amounts.
* Monthly and Seasonal Trends: By grouping data by month or quarter, the project identifies periods of high spending and how spending patterns fluctuate throughout the year.
* Cashback Utilization: A closer look at cashback trends helps evaluate how cashback offers impact the spending behavior of users.

# Potential Applications

* Personal Finance Management: By analyzing their expenses and payment methods, users can make informed decisions to manage their finances better.
* Business Insights: Businesses can gain insights into consumer spending behavior, payment preferences, and trends that can inform marketing strategies or product development.

# Conclusion

This project provides a comprehensive framework for analyzing personal expenses, combining synthetic data generation, SQL-based analysis, and interactive visualizations to help users track and understand their financial habits over time. It serves as a powerful tool for both individuals looking to manage their finances and businesses seeking to understand consumer spending trends.