GRAPH CHATBOT CODE DOCUMENTATION

* APP.PY

This file contains the front end written in streamlit.

The file contains all the drop-down options to be displayed in the front end along with all the important function calls from different files which are to be made for the generation of the graph. All appropriate prompts for the generation of the corresponding graphs are also included in this file.

* CONFIG.PY

The config.py file contains all the important configurations which include the end points of the local API, the API key and the credentials of the database to be used. These database credentials include the database name, table name, username, password etc. The database used in this code is MySQL and hence a MySQL connector is used, however other database can also be included like PostgreSQL, MS SQL etc.

* DB\_UTILS.PY

The db\_utils.py file includes the code to establish the connection with the MySQL database. The credentials are taken from the config.py file and the connection is established in the function, execute\_sql\_query (). This function also ensures that whenever data has been requested by the user in order for generation of the graph this function ensures that the cursor returns the necessary needed for the generation of the graph.

* QUERY\_GENERATOR.PY

This file includes the necessary prompts of the LLM which will be needed to generate the relevant query based on the user input. The file consists of multi-line prompt which hints to the LLM that it is a master at generating SQL queries which enables for the LLM to generate the queries accurately and once the query is generated the data is fetched form the database by triggering these queries at the backend.

* SCHEMA\_UTILS.PY

This file consists of code in order to load the relevant schema of the database and using the information from that schema generate the SQL queries. The file ‘new\_schema\_2.csv’ is the main excel file which consists of the schema of our database which will be used a basis by our LLM. The function ‘load\_schema’ is used to create the data frame which consists of the table name, database name, column name, data type etc.

* VISUALIZATION.PY

The file visualization.py contains the main logic for creation of the graphs. It has 2 main functions which are used to generate the graphs. The function ‘plot\_100\_percent\_stacked\_bar\_chart’ is used to create and display the stacked bar chart showing the distribution of transaction types per day based on their percentage for a particular month. This function enables for a smoother and faster generation by providing the necessary code without the intervention of pandas AI.

The function ‘generate\_dashboard’ is used to generate all the remaining categories of graphs. This function uses the data which is given by running the generated SQL query at the database and then retrieves the data in the form of suitable data frame and then generates a matplotlib code which in turn generates the required and necessary graph.