Data Engineer Intern Practical Exam

Objective: The objective of this test is to evaluate the candidate's ability to design and implement a data engineering solution that imports customer and order data from CSV files into a MySQL database, and subsequently analyze this data using a Streamlit web application.

CSV files attached to the email itself.

Test Components:

Part 1: Data Preparation

1. **Data Import**:
    - Write a Python script that imports customer and order data from customer.csv and orders.csv files into a MySQL database.
    - The customers table should contain fields for customer_id and customer_name.
    - The orders table should include order_id, customer_id, total_amount, and order_date.
    - **After importing the data app should run using MySQL database not using the csv files**

2. **Database Connection**:
    - Use SQLAlchemy or a similar library to connect to the MySQL database.
    - Ensure that the connection is secure and handles any potential connection errors gracefully.

Part 2: Streamlit App Setup (https://streamlit.io/)

Build a **Streamlit app** with the following components:

1. **Sidebar Filters**:
    - Add a date range filter to allow the user to filter orders by order_date.
    - Add a slider to filter customers by the total amount they've spent (e.g., filter customers who have spent over $1000).
    - Add a dropdown to allow filtering by customers with more than a certain number of orders (e.g., more than 5 orders).

2. **Main Dashboard**:
    - Display the filtered data in a table using st.dataframe().
    - Create a bar chart showing the top 10 customers by total revenue.
    - Create a line chart showing the total revenue over time (grouped by week or month).
    - Add a summary section showing key metrics like:
        - Total revenue.
        - Number of unique customers.
        - Number of orders.

Part 3: Data Analysis

1. **Machine Learning Model** (Bonus):
    - Implement a simple machine learning model (e.g. logistic regression) that predicts whether a customer is a repeat purchaser based on their total orders and revenue.
    - Include validation checks to ensure the model has sufficient data for training and can output the accuracy of predictions.

Part 4: Documentation and Code Quality

1. **Documentation**:
    - Provide clear documentation within the code and an external README file that explains how to set up the MySQL database, run the application, and load the data.

Submission Guidelines:

1. Submit your Streamlit app code as a Python script or Jupyter Notebook.
2. SQL commands used to create the necessary tables in the MySQL database.
3. Ensure the app runs locally, and provide clear instructions in the README on how to install dependencies and run the app (you can use a requirements.txt file).
4. If the app is deployed (on platforms like Streamlit Cloud or Heroku), share the live link as well.

Timeline & Deliverables

Deadline : 27th Of October 2024 - 11.59PM

You have 3 days to complete your work. Your work should be uploaded to a GitHub repository. Once completed, please send your **Document & GitHub repository by email to the following email ID -** indeewari@delivergate.com.

cc : hasitha@delivergate.com , isuru@delivergate.com

You must submit your final deliverables to above mentioned mails.

For any questions related to the assignment email to : hasitha@delivergate.com