Dear Candidate,

Thank you for your interest in joining Accenture coding task

Building a Comprehensive Sales Data Pipeline

Challenge Summary

You are tasked with building a comprehensive sales data pipeline for a retail company. The pipeline should combine generated sales data with data from external sources, perform data transformations and aggregations, and store the final dataset in a database. The aim is to enable analysis and derive insights into customer behaviour and sales performance.

Requirements

Sales data CSV file

- You are provided with a CSV file containing generated sales data (see attachment). **Data Transformation**

   (JSONPlaceholder API - /users)

- Fetch user data from the JSONPlaceholder API endpoint /users using the requests library.
- You can use the endpoint: https://jsonplaceholder.typicode.com/users
- Extract relevant fields such as id, name, username, email, lat and lng from the API response.
- Merge the user data with the sales data based on the customer_id field.
- Include the merged user data in the final dataset, mapping each sale to the respective user. **Data**

   Transformation (OpenWeatherMap API)

- Use the OpenWeatherMap API as an additional data source.
- Get an API key by signing up for a free account.
- You can use the endpoint: https://api.openweathermap.org/data/2.5/weather
- For each sale, extract the location information (e.g., using a fictional store address) and make an API request to fetch the weather data for that location.
- Include relevant weather information (e.g., temperature, weather conditions) in the final dataset, associating it with each sale. **Data Manipulation and Aggregations**

- Perform necessary data manipulations and aggregations on the merged dataset to derive valuable insights.
- Perform aggregations or manipulations based on the data. These are the required aggregations:
   - Calculate total sales amount per customer.
   - Determine the average order quantity per product. ○ Identify the top-selling products or customers.
   - Analyze sales trends over time (e.g., monthly or quarterly sales).
   - Include any other aggregations or data manipulations that you think are relevant.
   - Include weather data in the analysis (e.g., average sales amount per weather condition).

Data Storage

- Design and create an appropriate database schema to store the transformed data.
- Choose a suitable relational database management system and create the necessary tables. (sqlite is recommended as it is easy to install)
- Define the appropriate data types for each column in the database tables.
- Write the logic to store the transformed and aggregated data into the database.

- Provide clear instructions on how to set up and run the data pipeline.

- Include a brief explanation of the data transformation steps and any assumptions made.
- Describe the database schema used to store the transformed data.
- Specify the suggested aggregations and data manipulation tasks to be performed.
- Create some visualizations to present the insights derived from the data (**can be done as bonus point**).
-

Deliverables

Source Code

- Include all the necessary code files and scripts required to run the data pipeline. **Documentation**
- Write a README file that explains the setup and usage instructions for the data pipeline.
- Include a description of the data pipeline components, their functionality, and any dependencies.

Submission

Please submit your code through Github or Gitlab. Kindly provide clear instructions on how to set up and run the data pipeline.

You will be asked to present your solution to the team during the interview – this will be scheduled after the delivery of the challenge which you need to complete within 7 calendar days from receipt.
If you have any questions about the requirements, please feel free to reach out to us.
We wish you the best of luck!