Teamwork Assignment

Starts Nov 6, 2023 12:01 AM

50 % 1 of 2 topics complete

General

PeerScholar Teamwork Assignment

External Learning Tool

Starts Nov 6, 2023 12:01 AM

This is a teamwork assignment where teams will work together to create a Python-based data analysis algorithm to process, analyze, and visualize a given dataset. Each team will consist of three students, with each member assigned to a specific role.

1. The Data Engineer (OOP Specialist) will be responsible for designing classes and objects to handle data cleaning and transformation.

2. The Algorithm Developer (Data Structures and Algorithms Specialist) will be responsible for implementing efficient algorithms to analyze the data and extract meaningful insights.

3. The Visualization Expert will be responsible for creating visual representations of the analyzed data using standard or third-party Python libraries.

The dataset for this assignment is the Superstore Sales Dataset, which contains historical sales data for more than one thousand stores. The dataset includes attributes like Order StoreID, Sales, Date, Customers, OpenStatus, etc.

Tasks:

Data Preprocessing and Cleaning:

1. Design classes to read the dataset from a CSV file.

2. Implement methods to handle missing values and outliers, especially in the "Sales" column.

3. Transform data types, especially the "Date" column, to datetime or epoch timestamp format if necessary.

Data Analysis:

1. Implement an algorithm to identify the top 10 ~~stores~~ sub-categories by sales.

2. Design a data structure to store the monthly sales trend for the entire superstore.

3. Implement a search algorithm to find all days where a promo led to sales above a certain threshold.

4. Design a scenario where a Stack or Queue data structure would be helpful in processing or analyzing the dataset. Implement this scenario and explain your choice.

Efficiency Considerations:

1. Differentiate between writing code that is memory efficient versus code that is time efficient. Provide examples from your implementation.

2. Optimize at least one part of your code for memory efficiency and another part for time efficiency. Document the changes made and their impact.

Data Visualization:

1. Create a bar chart showing the top 10 ~~stores~~ categories by sales.

2. Plot a line chart to display the monthly sales trend for the entire superstore.

3. Design a scatter plot to visualize the relationship between the number of customers and sales for the entire dataset.

Submission:

1. Submit the Python code in IPYNB format (Jupyter Notebook) on **Monday, December 4, 2023.**

2. Within the IPYNB file, use text cells to provide comments and explanations detailing the design decisions made during this phase. This should include design decisions, the approach employed, and any assumptions made.

3. Along with the IPYNB file, submit an exported notebook version in either HTML or PDF format. This exported file should include both code snippets and output.

Here is a link where you can find more information on navigating the assignment phases.

Assignments

Teamwork Assignment

Assignment

Overdue - Dec 11, 2023 11:59 PM

Available on Nov 6, 2023 12:01 AM. **Access restricted before availability starts.**

For this assignment, your team will collaborate to create a Python-based data analysis algorithm that can process, analyze, and visualize a provided dataset. Each team will consist of three students, and each member will be assigned a specific role. You can access the assignment details by clicking on the 'Teamwork Assignment' section under 'Table of Contents' on the course shell and then selecting 'PeerScholar Teamwork Assignment.'