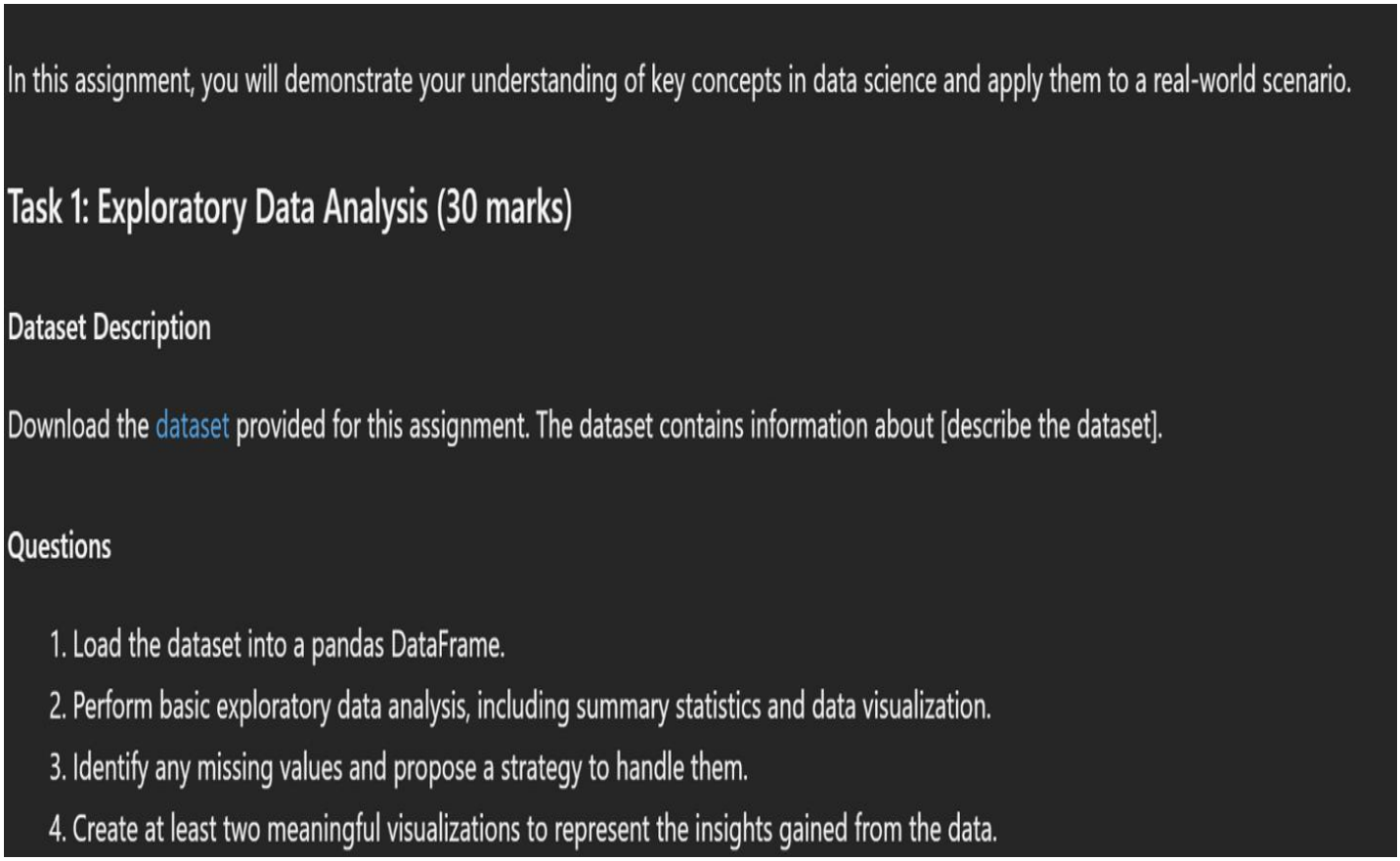
Machine Learning

Q1. I'm capturing the rendered output of the Markdown as a screenshot. Write it in Markdown format and submit your assignment as a Markdown file?



In this assignment, you will demonstrate your understanding of key concepts in data science and apply them to a real-world scenario.

Task 1: Exploratory Data Analysis (30 marks)

Dataset Description

Download the dataset provided for this assignment. The dataset contains information about [describe the dataset].

Questions

1. Load the dataset into a pandas DataFrame.
2. Perform basic exploratory data analysis, including summary statistics and data visualization.
3. Identify any missing values and propose a strategy to handle them.
4. Create at least two meaningful visualizations to represent the insights gained from the data.

Task 2: Data Preprocessing (20 marks)

Data Cleaning

1. Handle missing values based on the strategy proposed in Task 1.
2. Check for and handle any outliers in the dataset.

Feature Engineering

1. Create a new feature that [describe the new feature].
2. Normalize or standardize relevant features.

$$X_{\text{normalized}} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

Discuss the importance of normalization in the context of machine learning.

3. Create a new feature (Z) that represents the interaction between two existing features (X) and (Y) using the formula:

$$Z = \frac{X}{Y}$$

Provide a brief explanation of why you chose these features.

Task 3: Machine Learning (30 marks)

Model Building

1. Split the dataset into training and testing sets.
2. Choose a suitable machine learning algorithm and train a model.
3. Evaluate the model's performance using appropriate metrics.

Hyperparameter Tuning

1. Experiment with hyperparameter tuning to improve the model's performance.
2. Discuss the impact of different hyperparameters on the model.

Task 4: Conclusion and Recommendations (20 marks)

Write a conclusion based on your analysis and propose any recommendations for future work.

Task 5: Advanced Analysis (25 marks)

Question 4.1

Apply a linear regression model to predict a target variable (Y) based on relevant features (X_1, X_2, X_n). Evaluate the model's performance using the Mean Squared Error (MSE) formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Question 5.2

Discuss the strengths and limitations of the linear regression model. Propose potential improvements.

Mathematical Formulas

- **Summation (Sigma):** $\sum_{i=1}^{n} X_i$
- **Theta (θ):** θ
- **Pi (π):** π
- **Logarithm (log):** $\log(x)$

Submission Guidelines

1. Submit your assignment as a markdown file.
2. Include any necessary code snippets in the markdown.
3. Clearly label each section with appropriate headings.
4. Ensure your markdown file is well-formatted for readability.

Evaluation Criteria

Your assignment will be evaluated based on:

- **Completeness:** Did you address all the tasks?
- **Analysis:** How well did you analyze and interpret the data?
- **Clarity:** Is your markdown well-structured and easy to follow?
- **Creativity:** Did you approach the tasks with creativity and critical thinking?
- **Technical Proficiency:** How well did you implement data science techniques?

Good luck!

Deadline: 10-10-24

Note:

- **Assignment Submitted in Pdf Form.**
- **In Pdf form, you must include the screenshot of the code and its output.**
- **If mentioned, then write python script(.py), notebook file(.ipynb) or markdown file(.md) that includes the code for each task.**
- **Include comments in your code to explain the purpose functionality of each step.**

- **If any theoretical question, then explain it briefly with example and diagrams.**
- **Send it in this email id:**

  jtechsolution93@gmail.com
- **You can get help from ChatGPT or any Chatbot, but at the end.**
- **If Any, then It is highly recommended, if you read research papers for assignment.**

  Helping Website for research papers are:

  https://scholar.google.com/

  https://sci-hub.hkvisa.net/
    - **https://sci-hub.hkv**