Intro to Data Science - Final Project Assignment

Course: Intro to Data Science

Project Overview

For your final project, you will be assigned a unique dataset. Your task is to clean the data, fill in missing values, infer key insights, and visualize the data. You will perform these tasks using both Python (with NumPy, pandas, and Matplotlib) and Power BI.

Project Objectives

1. **Data Cleaning and Transformation:**
   - Identify and handle missing values.
   - Clean and preprocess the data to make it suitable for analysis.
2. **Data Analysis and Visualization:**
   - Infer key insights from the data.
   - Visualize the data using appropriate plots and charts.
3. **Tool Utilization:**
   - Perform data cleaning, transformation, and visualization in Python.
   - Perform data cleaning, transformation, and create a dashboard in Power BI.
4. **Presentation:**
   - Prepare a 10-minute presentation on the methodology followed in both Python and Power BI.
   - Present the Power BI dashboard.

Deliverables

1. **Python (Jupyter Notebook):**
   - A `.ipynb` file that contains:
     - Data cleaning and transformation steps.
     - Handling of missing values.
     - Data analysis and insights.
     - Data visualization using Matplotlib.
2. **Power BI:**
   - A `.pbix` file that contains:
     - Data cleaning and transformation steps.
     - Handling of missing values.
     - Data analysis and insights.
     - A comprehensive dashboard with relevant visualizations.
3. **Presentation:**
   - A 10-minute presentation covering:
     - The methodology followed in Python for data cleaning, transformation, analysis, and visualization.

- The methodology followed in Power BI for data cleaning, transformation, analysis, and dashboard creation.
- Presentation and walkthrough of the Power BI dashboard.

Project Guidelines

Python (Jupyter Notebook)

1. **Data Cleaning:**
   - Inspect and understand the structure of the dataset.
   - Handle missing values using appropriate techniques (e.g., imputation, deletion).
   - Remove or correct any inconsistencies or errors in the data.
2. **Data Transformation:**
   - Transform the data as needed to facilitate analysis.
3. **Data Analysis:**
   - Perform data analysis to uncover patterns, correlations, and insights.
4. **Data Visualization:**
   - Use Matplotlib to create meaningful visualizations that support your analysis and insights.
   - Ensure your visualizations are clear, well-labeled, and effectively communicate the findings.

Power BI

1. **Data Cleaning:**
   - Use Power Query to clean and preprocess the data.
   - Handle missing values appropriately.
2. **Data Transformation:**
   - Perform necessary data transformations in Power Query.
3. **Dashboard Creation:**
   - Design a comprehensive dashboard that includes various visualizations to present the key insights from the data.
   - Ensure your dashboard is clear and user-friendly.

Presentation

1. **Methodology Overview:**
   - Provide a clear and concise explanation of the steps followed in Python and Power BI.
   - Discuss the techniques and methods used for data cleaning, transformation, and analysis.
2. **Power BI Dashboard Walkthrough:**
   - Demonstrate the key features and visualizations in your Power BI dashboard.
   - Explain how the dashboard helps in understanding the data and deriving insights.

Submission

- **Python Jupyter Notebook:** Submit the `.ipynb` file via the LMS.
- **Power BI File:** Submit the `.pbix` file via the LMS.
- **Presentation:** Be prepared to present your methodology and Power BI dashboard. Submit the present on the LMS.

Evaluation Criteria

1. **Data Cleaning and Transformation (30%):**
   - Effectiveness of data cleaning and handling of missing values.
   - Appropriateness of data transformations.
2. **Data Analysis and Insights (30%):**
   - Depth of analysis and quality of insights inferred from the data.
3. **Visualization (20%):**
   - Clarity and effectiveness of visualizations in both Python and Power BI.
4. **Presentation (20%):**
   - Quality of the presentation and explanation of the methodology.
   - Demonstration and clarity of the Power BI dashboard.

Tips for Success

- **Start Early:** Begin working on the project as soon as possible to avoid last-minute rush.
- **Understand Your Data:** Spend time understanding the dataset and the context of the data.
- **Document Your Process:** Keep detailed notes on your methodology to make the presentation easier.
- **Seek Help:** Don't hesitate to ask for help or clarification if needed.

Good luck, and I look forward to seeing your insightful analyses and creative visualizations!