Assignment Statement:

Data Processing with Pandas

Objective:

In this assignment, you will apply your knowledge of Pandas to process a real-world dataset that is not clean. You will need to identify and address any missing values, outliers, and other data quality issues. Once the data is clean, you will perform some basic exploratory data analysis (EDA) to identify patterns and trends.

Instructions:

1. Find a dataset that is not clean. You can find datasets on Kaggle, GitHub, or other websites.
2. Load the dataset into Pandas and explore it using the describe() and head() methods.
3. Identify and address any missing values, outliers, and other data quality issues. You can use the dropna(), fillna(), drop(), and IQR() methods to address these issues.
4. Once the data is clean, perform some basic EDA to identify patterns and trends. You can use methods such as groupby(), agg(), hist(), and plot() to perform EDA.

Submission:

Submit a Jupyter Notebook or Python script that contains your code and findings.

Grading:

Your assignment will be graded on the following criteria:

- Correctness of your code
- Completeness of your data processing steps
- Quality of your EDA
- Clarity and organization of your code and findings

Tips:

- When choosing a dataset, look for one that is interesting to you and that you can learn something from.
- Don't be afraid to experiment with different data processing techniques.
- When performing EDA, try to identify patterns and trends that are both statistically significant and meaningful.
- Be clear and organized in your code and findings.