Assignment # 1

As a suggested first step, spend some time finding a data set that you are really passionate about. This can be a data set similar to the data you have available at work or data you have always wanted to analyze.

(You can get data from Kaggle.com https://www.kaggle.com/datasets)

Required

Once you have selected a data set, you will produce the deliverables listed below and submit them. Treat this assignment as an opportunity to produce analysis that are ready to highlight your analytical skills.

Sections required in your report:

- Brief description of the data set and a summary of its attributes
- Initial plan for data exploration
    - **Loading data**
    - **Displaying data**
    - **Review the data set structure and data types**
    - **Finding the features of data**
    - **Finding the total number of records in data set**
    - **Finding distribution of each feature**
    - **Checking the unique values different features**
    - **Finding distribution of categorical features**
    - **Finding descriptive statistics to understand the central tendency and spread of the data**
    - **Finding missing values.**

- Actions taken for data cleaning and feature engineering.
    - **Finding any outliers or anomalies in the data**
    - **Handle missing values by imputing or removing the observations, if necessary.**
    - **Convert categorical variables into numerical variables using one-hot encoding or label encoding.**
    - **3. Normalize the features to ensure that the features are on the same scale.**
    - **Create new features by combining existing features, if necessary.**
    - **Remove any irrelevant or redundant features.**
    - **Visualization of data to find the relationship with different features.**

- Key Findings and Insights, which synthesizes the results of Exploratory Data Analysis in an insightful and actionable manner
- Formulating at least 3 hypotheses about this data
- Conducting a formal significance test for one of the hypotheses and discuss the results
- Suggestions for next steps in analyzing this data
- A paragraph that summarizes the quality of this data set and a request for additional data if needed