

CS667-Fall2025: Practical Data Science

Assignment#2 - TOTAL POINTS: 100

Due Date: 10/25/2025 (Oct 25th)

Team Assignment

S.No.	Questions	Points	Self-Assessment
1	<p>In continuation with your selected dataset in week#5 perform following:</p> <p>Data Understanding:</p> <ul style="list-style-type: none"> Describe the dataset: Include the structure (rows, columns, data types) and its contents. Data types: Identify the types of data (numerical, categorical, text) and any conversion done between types as necessary. 	15	
2	<p>Descriptive Statistics and Summarization:</p> <ul style="list-style-type: none"> Central tendency and spread: What are the mean, median, mode, range, variance, and standard deviation of key numerical variables? Data distribution: Describe the distribution of key variables (use histograms or box plots to illustrate). 	15	
3	<p>Problem Statement:</p> <ul style="list-style-type: none"> Based on the dataset selected, formulate one problem statement for Regression and one for Classification 	10	
4	<p>Model Selection:</p> <ul style="list-style-type: none"> For the Regression problem statement, implement 2 models and compare. For the Classification Problem statement, implement 2 models and compare. 	30	
5	<p>Short Report</p> <ul style="list-style-type: none"> Short report summarizing key findings and insights (1-2 pages) 	30	

Submission:

1. Submit a Word document (include any screenshots) and Python script with the sections mentioned above. **Do not submit zip files.**
2. Answers must include your analysis and not the AI-generated responses. AI-generated responses will attract a penalty.
3. Submit as per Pace University's academic integrity and do not plagiarize. Such submissions will attract a zero grade.
4. There will be a 10% penalty for the submissions past the Due date. And 10% every week after that until the submission is available.