**Individual Project 5**

**DS160**

**Introduction to Data Science**

**Fall 2023**

**Data Science Questions (70 points)**

**Goal:** This project aims to do a basic knowledge check that we covered in this class.

**Instructions:** For this project, create a pdf script titled **IP5\_XXX.pdf**, where **XXX** are your initials. Also create a GitHub repository titled **IP5\_XXX** to which you can **push your pdf file along with the Word file.** Show your best work and keep the document for your future journey.

1. Define the term 'Data Wrangling in Data Analytics.
2. What are the differences between data analysis and data analytics?
3. What are the differences between machine learning and data science?
4. What are the various steps involved in any analytics project?
5. What are the common problems that data analysts encounter during analysis?
6. Which technical tools have you used for analysis and presentation purposes?
7. What is the significance of Exploratory Data Analysis (EDA)?
8. What are the different methods of data collection?
9. Explain descriptive, predictive, and prescriptive analytics.
10. How can you handle missing values in a dataset?
11. Explain the term Normal Distribution.
12. How do you treat outliers in a dataset?
13. What are the different types of Hypothesis testing?
14. Explain the Type I and Type II errors in Statistics?
15. Explain univariate, bivariate, and multivariate analysis.
16. Explain Data Visualization and its importance in data analytics?
17. Explain Scatterplots.
18. **Explain histograms and bar graphs.**
19. **How is a density plot different from histograms?**
20. **What is Machine Learning?**
21. **Explain which central tendency measures to be used on a particular data set?**
22. **What is the five-number summary in statistics?**
23. **What is the difference between population and sample?**
24. **Explain the Interquartile range?**
25. **What is linear regression?**
26. **What is correlation?**
27. **Distinguish between positive and negative correlations.**
28. **What is Range?**
29. **What is the normal distribution, and explain its characteristics?**
30. **What are the differences between the regression and classification algorithms?**
31. **What is logistic regression?**
32. **How do you find Root Mean Square Error (RMSE) and Mean Square Error (MSE)?**
33. **What are the advantages of R programming?**
34. **Name a few packages used for data manipulation in R programming?**
35. **Name a few packages used for data visualization in R programming?**