- **Comprehensive Foundations of Data Science Project Assignment**

Objective: Utilize a selected dataset to conduct a thorough Foundations of Data Science project, covering data exploration, cleaning, visualization, basic inferential statistics, and predictive analytics concepts.

Total Marks: 25

- Project Tasks:

1. **Dataset Selection and Business Problem Definition (1 Mark)**
    - Select a dataset and define a clear business problem or question that your analysis aims to address.
2. **Data Overview (1 Mark)**
    - Conduct a preliminary examination of the dataset. Describe its size, nature of variables (quantitative vs. qualitative), and the potential value it brings to solving the business problem.
3. **Data Types Identification (1 Mark)**
    - Classify the dataset's variables into their respective data types (numerical, categorical, ordinal, etc.).
4. **Missing Values Assessment (1 Mark)**
    - Identify any missing values in the dataset. Provide a summary of the missingness pattern observed.
5. **Data Cleaning Strategy (1 Mark)**
    - Propose a strategy for handling missing data (deletion, imputation, etc.) and justify your approach.
6. **Outliers Detection (1 Mark)**
    - Employ a method to detect outliers in the dataset. Briefly describe how you identified them.
7. **Handling Outliers (1 Mark)**
    - Decide on a strategy for dealing with the detected outliers. Explain your choice and its implications for the analysis.
8. **Data Transformation (1 Mark)**
    - Perform necessary data transformations (normalization, scaling, etc.) to prepare the data for analysis. Justify why these transformations are required.

9. **Categorical Data Encoding (1 Mark)**
    - Encode categorical variables as needed for analysis. Explain your encoding choices.
10. **Exploratory Data Analysis (EDA) Introduction (1 Mark)**
    - Begin EDA by summarizing key statistics (mean, median, mode, standard deviation) for at least three variables.
11. **Visual EDA: Distributions (1 Mark)**
    - Create visualizations (histograms, box plots) to explore the distribution of key variables.
12. **Visual EDA: Relationships (1 Mark)**
    - Generate scatter plots or correlation heatmaps to investigate relationships between variables.
13. **Identify and Interpret Key Relationships (1 Mark)**
    - Based on the visualizations, identify two key relationships in the data. Discuss their potential impact on the business problem.
14. **Descriptive Analytics Summary (1 Mark)**
    - Provide a concise summary of the descriptive analytics findings, highlighting insights relevant to the business problem.
15. **Formulate Statistical Questions (1 Mark)**
    - Based on your EDA, formulate two statistical questions that could further inform the business problem.
16. **Hypothesis Testing Plan (1 Mark)**
    - For one statistical question, outline a plan for a hypothesis test. Specify the null and alternative hypotheses, significance level, and the test statistic to be used.
17. **Inferential Statistics Conceptual Application (1 Mark)**
    - Conceptually apply inferential statistics to estimate a parameter or make predictions about the population based on your sample. No calculations required; explain your thought process.
18. **Introduction to Predictive Modeling (1 Mark)**
    - Identify a potential outcome variable for predictive modeling. Justify its selection based on your business problem and EDA findings.

19. **Feature Selection Rationale (1 Mark)**
    - Discuss how you would select features for the predictive model. Consider correlations, importance, and relevance to the outcome variable.
20. **Feature selection process (1 Mark)**
    - Select the features using at least 3 different model driven techniques
21. **Predictive Model Choice (1 Mark)**
    - Choose a predictive modeling technique suitable for your data and business problem (e.g., linear regression, logistic regression). Explain why this model is appropriate.
22. **Data Visualization for Insights (1 Mark)**
    - Create a dashboard layout or a set of visualizations that could provide actionable insights to a business stakeholder. Describe how these insights address the business problem.
23. **Presentation of Findings (1 Mark)**
    - Summarize the key findings from your project, focusing on how they address the business problem. Include any recommendations or potential actions.
24. **Interpretation of Predictive Model Output (1 Mark)**
    - Once you've conceptually chosen a predictive model, describe how you would interpret the model's output to derive actionable business insights.
25. **Conclusion (1 Mark)**
    - Conclude your project by summarizing the key insights and their implications for the business problem. Reflect on the limitations of your analysis and potential improvements.