Data Analysis Project Guidelines

Techolas Technologies

Objective:

The objective of this project is to enable Techolas students to apply data analysis techniques to a real-world dataset. Students will identify a dataset, analyze it to find and address any issues, and suggest solutions to improve the dataset and extract meaningful insights.

Number of Projects:

Tableau – 1

Excel – 1

Power BI -1

Steps and Guidelines:

1. Choose a Dataset

Select a Dataset: Find a dataset relevant to your interests or field of study. Good sources include Kaggle, UCI Machine Learning Repository, government databases, or other public data sources.

Dataset Description: Provide a brief description of the dataset, including its source, the number of records, and the features included.

2. Understand the Problem

Define the Problem: Clearly state the problem you are addressing with this dataset. This could be related to data quality, missing values, outliers, or specific analysis goals.

Students must address some problems in the data set and should provide the necessary solutions as well.

Research Context: Explain why this problem is important in a real-world context.

3. Data Preprocessing

Data Cleaning: Identify and handle missing values, correct any inconsistencies, and remove duplicates.

Data Transformation: Normalize or standardize the data if necessary. Apply any relevant transformations to the data.

Feature Engineering: Create new features/columns if needed to enhance the dataset and improve analysis.

4. Exploratory Data Analysis (EDA)

Visualizations: Use Tableau, Power BI and Excel for each project's visualizations.

Create various plots (histograms, scatter plots, bar charts, etc.) to understand the distributions and relationships within the data.

Summary Statistics: Calculate and interpret basic statistics (mean, median, standard deviation, etc.) for the dataset.

Insights: Document any interesting patterns or anomalies you discover during EDA.

5. Identify and Address Problems

Data Quality Issues: Identify issues such as missing values, outliers, and data inconsistencies.

Addressing Issues: Explain how you handled these issues (e.g., imputation for missing values, removal or treatment of outliers).

Documentation: Provide a clear and detailed account of your process for addressing these problems.

6. Data Analysis

Analysis Techniques: Apply relevant statistical or machine learning (Optional) techniques to analyze the dataset and address the problem.

Modelling: If applicable, build and evaluate predictive models to derive insights or make predictions.

Validation: Use appropriate validation techniques to assess the performance of your models.

7. Suggest Solutions

Insights and Recommendations: Based on your analysis, provide actionable insights and recommendations to address the identified problems.

Solution Implementation: Suggest how your recommendations can be implemented in a real-world scenario.

8. Conclusion

Summary: Summarize the key findings from your analysis.

Reflection: Reflect on what you learned during the project and any challenges you encountered.

Future Work: Suggest areas for further research or analysis.

9. Report and Presentation

Report: Compile your findings into a well-structured report as per the Techolas Technologies guidelines. Include sections for introduction, methodology, analysis, results, and conclusion.

Presentation: Prepare a presentation summarizing your project. Highlight key insights, problems addressed, and suggested solutions.

Note: Students are encouraged to document their code and steps taken throughout the project to ensure reproducibility and transparency.