



# APPLIED BUSINESS ANALYTICS

**2023-24, Spring Semester**

## Exploratory Data Analysis (EDA) Report Guideline

### Introduction:

#### 1. **Dataset Description:**

- a. Provide a brief overview of the dataset, including the number of observations, variables, and their types (numerical, categorical).
- b. Mention the source of the data and any relevant background information.

#### 2. **Data Cleaning:**

- a. Handling Missing Values:

Describe how missing values were identified and addressed.

Explain the rationale behind the chosen method for handling missing data.

#### 3. **Data Types and Conversions:**

- a. Outline the types of data (categorical, numerical) and any necessary conversions.
- b. Discuss the reasons for encoding categorical variables if applicable.

### Univariate Analysis:

#### 4. **Summary Statistics:**

- a. Present key summary statistics (mean, median, mode, standard deviation) for numerical variables.
- b. Discuss any notable observations or patterns.

#### 5. **Distribution Visualization:**

- a. Utilize visualizations (histograms, box plots) to showcase the distribution of numerical variables.
- b. Describe the shape, central tendency, and spread of the distributions.

#### 6. **Categorical Variables Exploration:**

- a. Use bar charts or pie charts to explore the distribution of categorical variables.
- b. Highlight any dominant categories or patterns.

### Bivariate Analysis:

#### 7. **Correlation Analysis:**

- a. Explore correlations between numerical variables using correlation coefficients or heatmaps.
- b. Discuss the strength and direction of significant correlations.

#### 8. **Scatter Plots and Relationships:**

- a. Create scatter plots to visualize relationships between pairs of numerical variables.

- b. Analyze trends and identify potential outliers.

Multivariate Analysis:

**9. Multivariate Visualizations:**

- a. Utilize tools like pair plots or heatmaps to explore relationships among multiple variables.
- b. Identify patterns or clusters within the data.

Outliers and Anomalies:

**10. Detection and Handling:**

- a. Identify outliers through visualizations or statistical methods. - Discuss the rationale for handling or keeping outliers.

Conclusion:

**11. Key Findings:**

- a. Summarize the main insights gained from the exploratory analysis.
- b. Highlight any unexpected or interesting patterns discovered.

**12. Next Steps:**

- a. Suggest potential areas for further investigation or analysis.
- b. Discuss the implications of the findings for subsequent modeling or decision-making.

Visualization and Reporting:

**13. Clear and Informative Visuals:**

- a. Ensure that all visualizations are appropriately labeled and easy to interpret.
- b. Include captions and titles for each visualization.

**14. Narrative Flow:**

- a. Maintain a logical and clear narrative flow throughout the report.
- b. Use section headings and subheadings to organize the content.