LOAN APPROVAL ANALYSIS

SkillCircle

THINK BEYOND NINE TO FIVE

Introduction

In this assessment, you will perform an Exploratory Data Analysis (EDA) on a dataset related to home loan approval. The dataset is sourced from Skill Circle and contains valuable information for analysis. The primary focus of this assessment is on data exploration and visualization machine learning techniques are not required.

The goals of this assessment are to:

- Gain familiarity with the dataset.
- Perform data exploration and visualization.
- Identify patterns, trends, and potential insights.
- Generate meaningful visualizations to communicate your findings.

Dataset

PLEASE FOLLOW THE LINK TO ACCESS AND DOWNLOAD THE DATASET FROM GOOFLE DRIVE ONLINE.

Project Tasks

Task 1: Data Exploration

- Load the dataset into a Python environment (e.g., Jupyter Notebook).
- Display the first few rows of the dataset to understand its structure.
- Check for missing values and handle them if necessary.
- Summarize basic statistics (mean, median, standard deviation, etc.) for the numeric columns.

Task 2: Data Visualization

2.1 Univariate Analysis

- Explore the distribution of numeric columns using the following visualizations:
- **Histograms**: Plot the frequency distribution of key numeric variables.
- **Box Plots:** Identify potential outliers and visualize the spread of data.
- Analyze categorical variables by creating the following plots:
- **Bar Charts:** Visualize the frequency distribution of categorical variables.
- **Pie Charts:** Represent the composition of categorical variables.

2.2 Bivariate Analysis

- Create scatter plots to explore relationships between pairs of numeric variables.
- Use pair plots (scatter matrix) to visualize interactions between multiple numeric variables simultaneously.
- Investigate the relationship between categorical and numeric variables using box plots or violin plots.

2.3 Multivariate Analysis

- Perform a correlation analysis to identify relationships between numeric variables.
- Visualize correlations using a heatmap.
- Create a stacked bar chart to show the distribution of categorical variables across multiple categories.

Task 3: Geospatial Analysis (Optional)

- If the dataset contains geographical information, visualize data on a map to identify regional trends.
- Use scatter plots or heatmaps to display data patterns across geographic locations.

Submission Guidelines

- Organize your work in a Jupyter Notebook or any preferred environment.
- Include clear explanations and interpretations for each visualization.
- Provide comments and descriptions of your code to make it understandable.
- Include relevant graphs, charts, or plots to support your analysis.
- Submit your completed assessment document.