

Week 2: Data Exploration and Visualization

Tasks:

1. Univariate Analysis

- Create histograms and box plots for numerical variables
- Create bar charts for categorical variables
- Compute and visualize descriptive statistics

2. Bivariate Analysis

- Create scatter plots for pairs of numerical variables
- Create box plots of numerical variables grouped by categorical variables
- Perform and visualize correlation analysis
- Conduct chi-square tests for categorical variables

3. Multivariate Analysis

- Create pair plots
- Perform and visualize principal component analysis (PCA)
- Create parallel coordinates plots

4. Advanced Visualization

- Create interactive visualizations using Plotly or Bokeh
- Develop a dashboard summarizing key insights using Streamlit

5. Insight Generation

- Identify and document key patterns and relationships in the data
- Formulate new hypotheses based on exploratory analysis

Deliverables:

1. Jupyter notebook containing all exploratory data analysis code and visualizations
2. Comprehensive data exploration report (PDF) including:
 - All generated visualizations with detailed interpretations
 - Summary of key findings and their potential impact on dropout prediction
 - List of new hypotheses generated from the analysis
2. Interactive Streamlit dashboard showcasing key visualizations and insights
3. Updated dataset incorporating any new features or transformations identified during EDA