Big Data Analytics ICP3B Report

Name: Akanksha Rasaputhra

Date: 14thJul2024

# Introduction

This report presents the work for the Big Data Analytics ICP3 assignment. It demonstrates practical data visualization and analysis using Matplotlib, NumPy, and Pandas. Tasks include simple plots, subplots, styling, and working with real-world datasets for practice.

# Matplotlib Practice

This section covers basic plots: line plots, scatter plots, bar charts, and histograms using Python lists and arrays.

A screenshot of a computer

AI-generated content may be incorrect.A person standing in front of a white background

AI-generated content may be incorrect.

# NumPy Practice

This section demonstrates generating and manipulating numerical arrays with NumPy, and plotting mathematical functions like squares, exponentials, and sine waves.

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

# Subplots & Plot Styling

Here, I practiced creating multiple subplots and customizing plot styles using Matplotlib's object-oriented interface.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

# Car Sales Dataset Analysis

In this section, I loaded and cleaned the car\_sales CSV file, converted text data to numeric, created cumulative sales columns, and visualized sales trends and price relationships.

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

# Heart Disease Dataset Analysis

Finally, I explored the heart\_disease dataset, created histograms, compared variables for patients over 50, and applied different plot styles and color maps to improve visual clarity.

A screenshot of a computer

AI-generated content may be incorrect.

# Key Highlights

- Practiced Matplotlib’s basic and advanced plotting techniques.  
- Used NumPy for array generation and mathematical transformations.  
- Cleaned and transformed real-world CSV data using Pandas.  
- Visualized trends, distributions, and relationships in datasets.  
- Applied different plot styles and saved figures for sharing.

# Conclusion

This assignment helped strengthen my end-to-end data workflow skills, including data cleaning, transformation, and visualization. It provided a strong foundation for tackling larger data projects and presenting insights clearly.

**## Video Recording Link**

A screen recording of my full notebook walkthrough, including code explanations and output demonstrations, is available here:

<https://drive.google.com/file/d/1kjNilB_xgl4KVUp7fP8MpXLJ4hl4H2ur/view?usp=sharing>