Data and Artificial Intelligence

Cyber Shujaa Program

Week 3 Assignment
Assignment 3: Titanic Exploratory Data Analysis

Student Name: Austin Githinji

Student ID: CS-EH03-25417

Contents

Introduction

1. This project involves performing **Exploratory Data Analysis (EDA)** on the **Titanic dataset**, a classic dataset available on Kaggle.
   The main goal is to explore, clean, and analyze the dataset to uncover hidden patterns and relationships between passenger attributes and survival outcomes.
2. EDA is an essential phase in data science projects as it helps develop a deep understanding of the data's structure, detect anomalies, and prepare it for model training or business insights.
3. Store structured data into a Pandas DataFrame.
4. Export the final dataset to a .csv file.

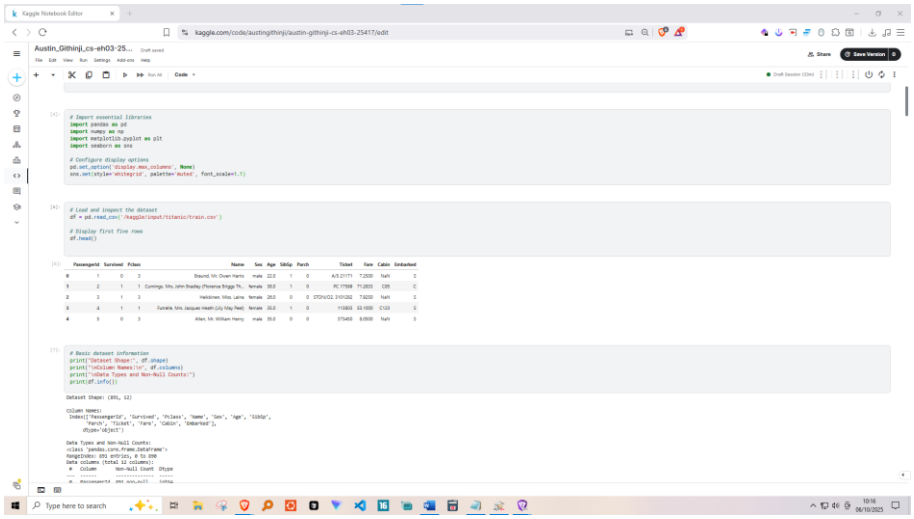
Objectives

The objectives of this assignment are to:

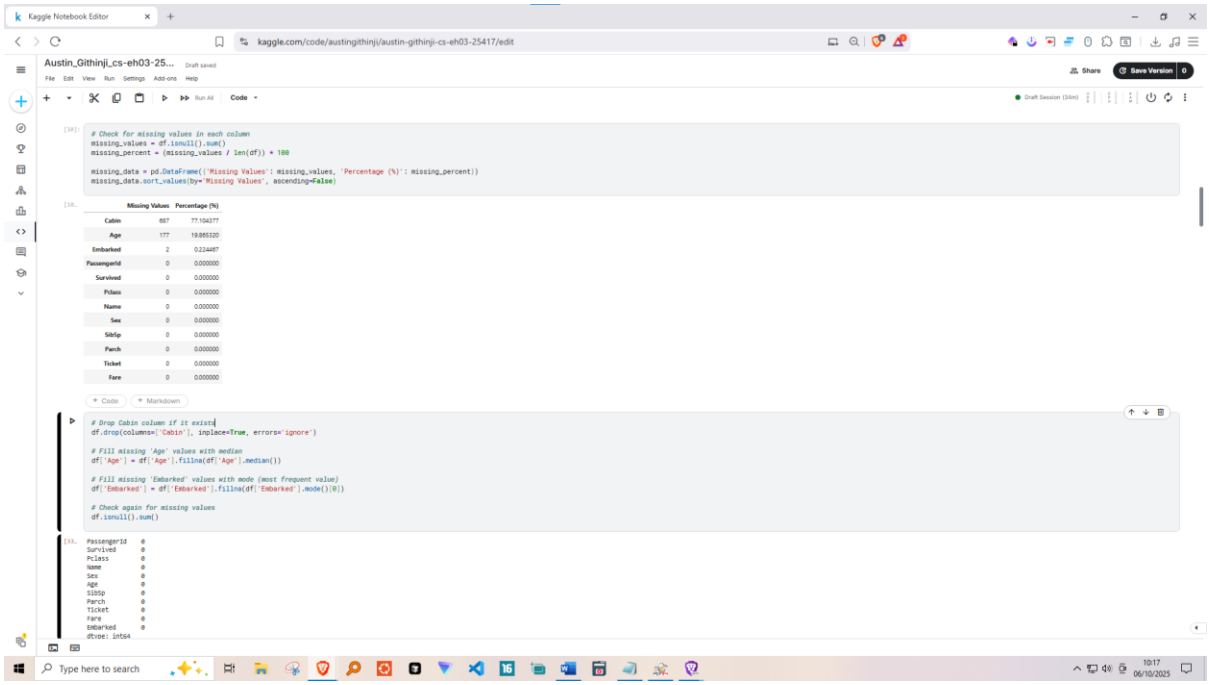
- Load and explore the Titanic dataset using **pandas** and **seaborn**.
- Handle missing values, duplicates, and outliers to clean the data.
- Perform **univariate, bivariate, and multivariate analysis**.
- Visualize relationships among key variables such as **Age**, **Fare**, **Sex**, **Pclass**, and **Survived**.
- Analyze the **target variable (Survived)** and determine factors influencing survival.
- Present findings with supporting visualizations and descriptive insights.

Tasks Completed

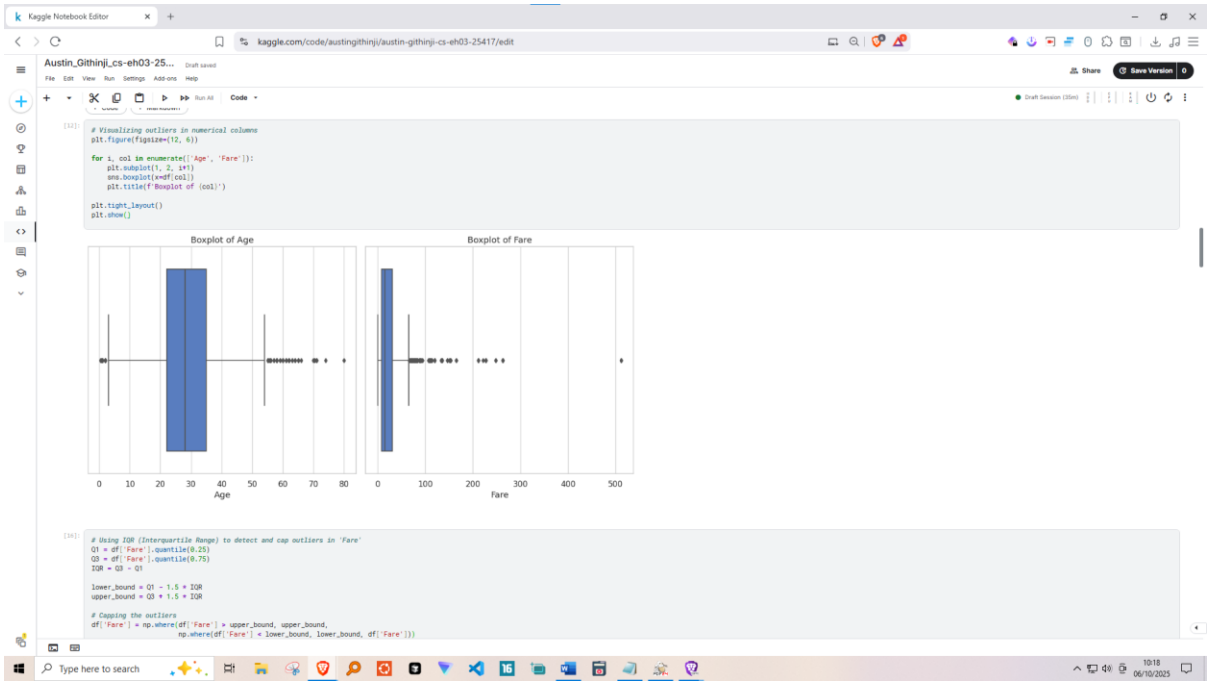
Initial data exploration:



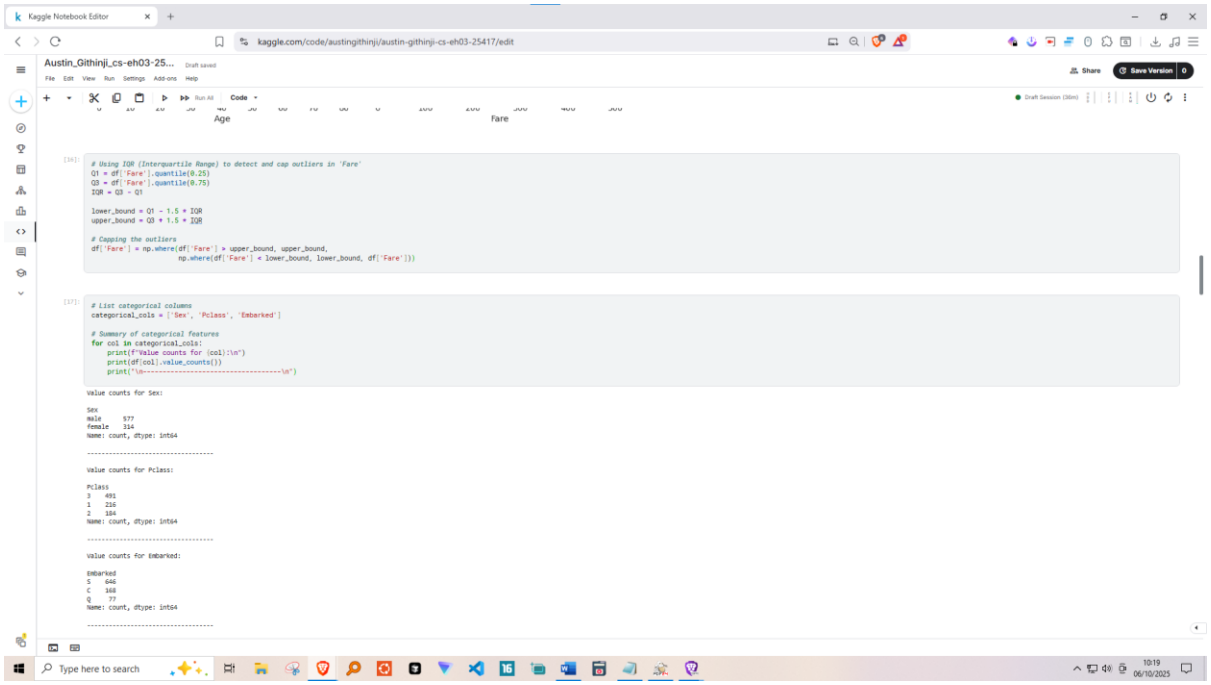
Handling Missing Values and Outliers



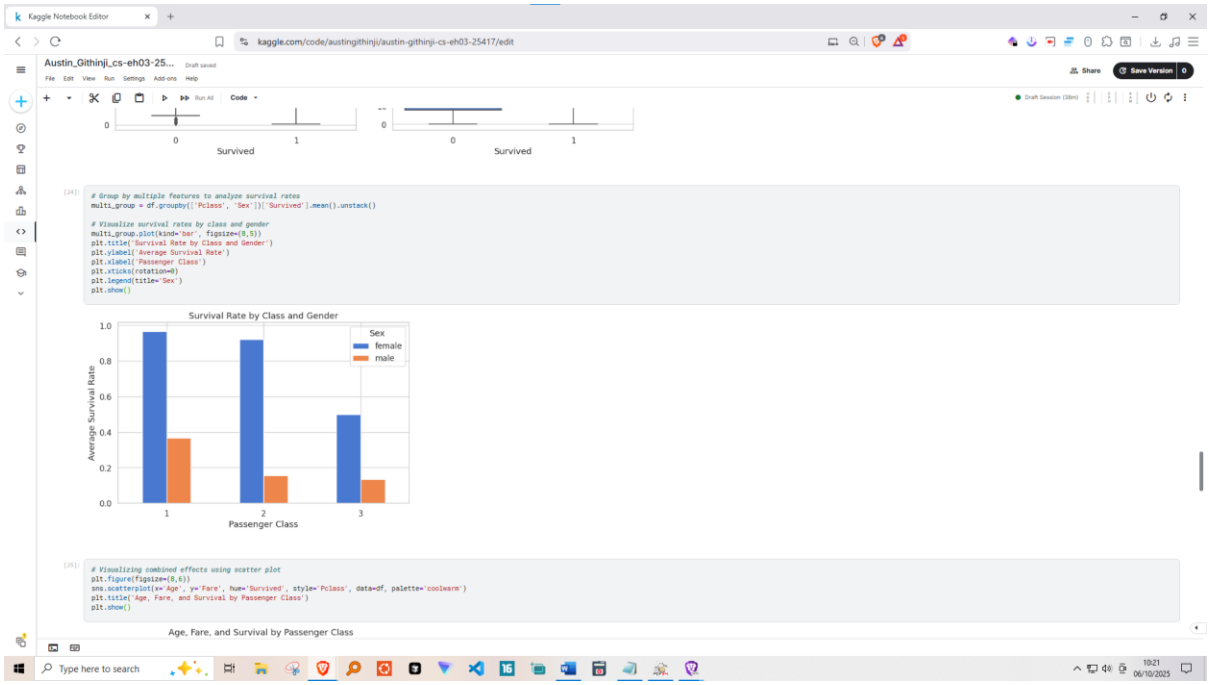
Univariate Analysis

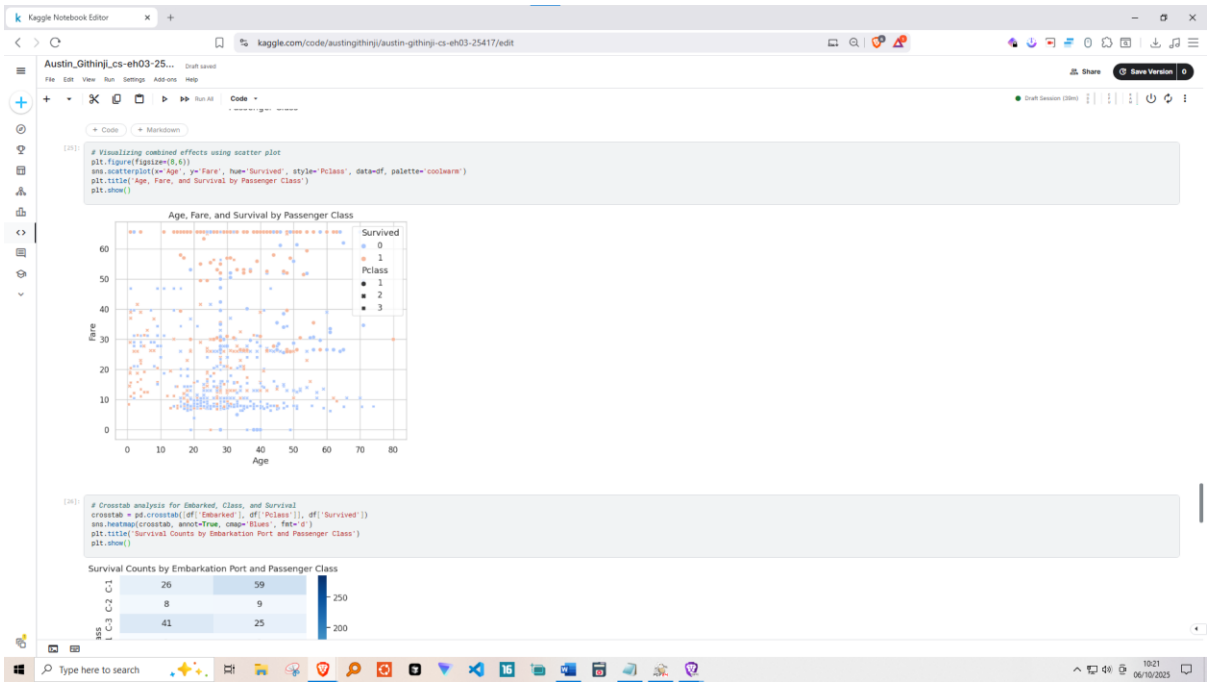


Bivariate Analysis

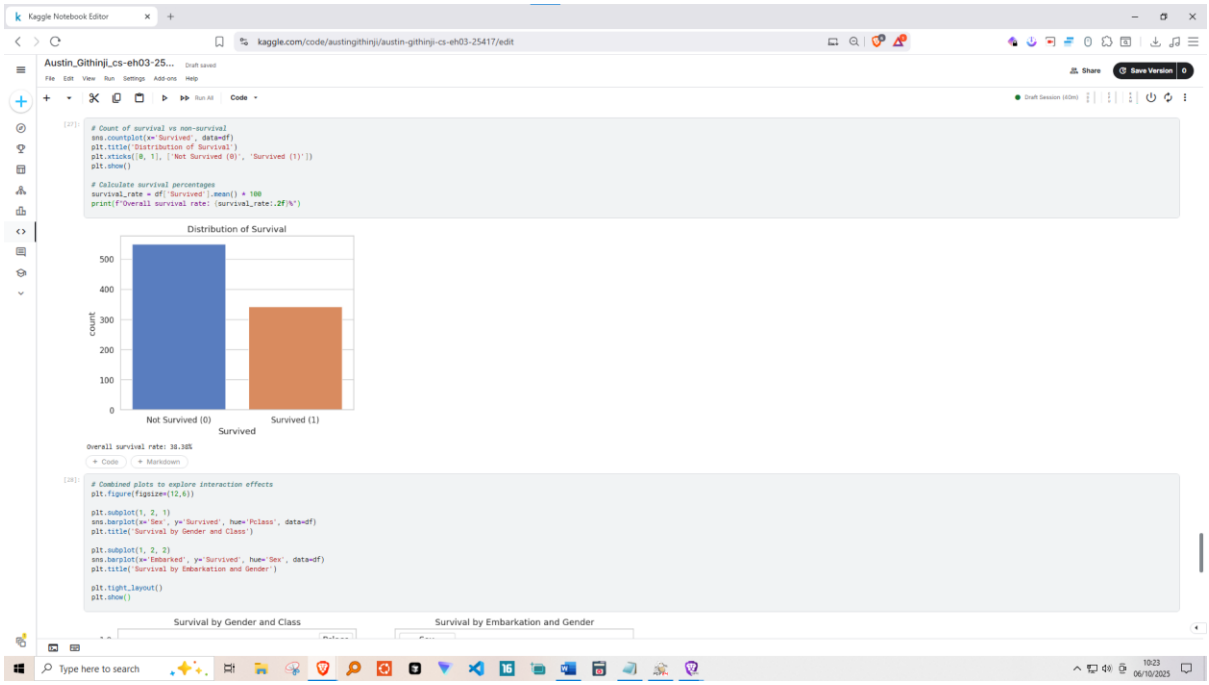


Multivariate Analysis

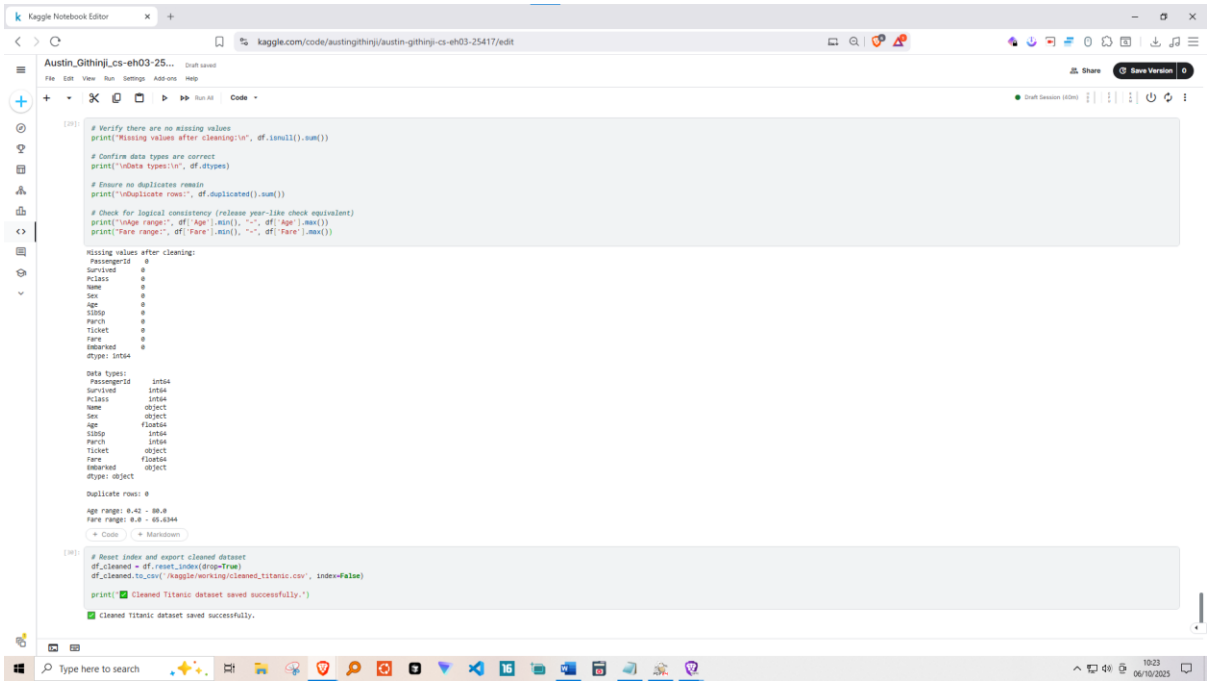




Target Variable Analysis:



Final Steps:



Link to Code: https://www.kaggle.com/code/austingithinji/austin-githinji-cs-eh03-25417

Conclusion

This EDA project successfully demonstrated the data wrangling and analytical process applied to the Titanic dataset.
Through a structured analysis, key patterns affecting survival were identified.
The dataset is now clean, validated, and ready for predictive modeling or deeper statistical inference.

EDA revealed that **social and economic factors** (gender and class) played a crucial role in survival chances.