DATA SCIENCE CAPSTONE PROJECT

IJAZ AHMED
01.09.2025

OUTLINE

Executive Summary

Introduction

Methodology

Results

Conclusion

Appendix

EXECUTIVE SUMMARY

Studied what makes a launch successful by fetching and preparing data for different Predictive Models which and then making a Dashboard for the live analytics

INTRODUCTION

SpaceX is a successful company for making the space travel affordable (62 Million Dollars instead of 165 Million Dollars) because they can reuse the first stage. Using ML Models we want to see what is the best Model to make predictions when they can reuse the first stage.

But

- How often, and what factors are important?

METHODOLOGY

Data Collection using APIs and Scrapping Wikipedia

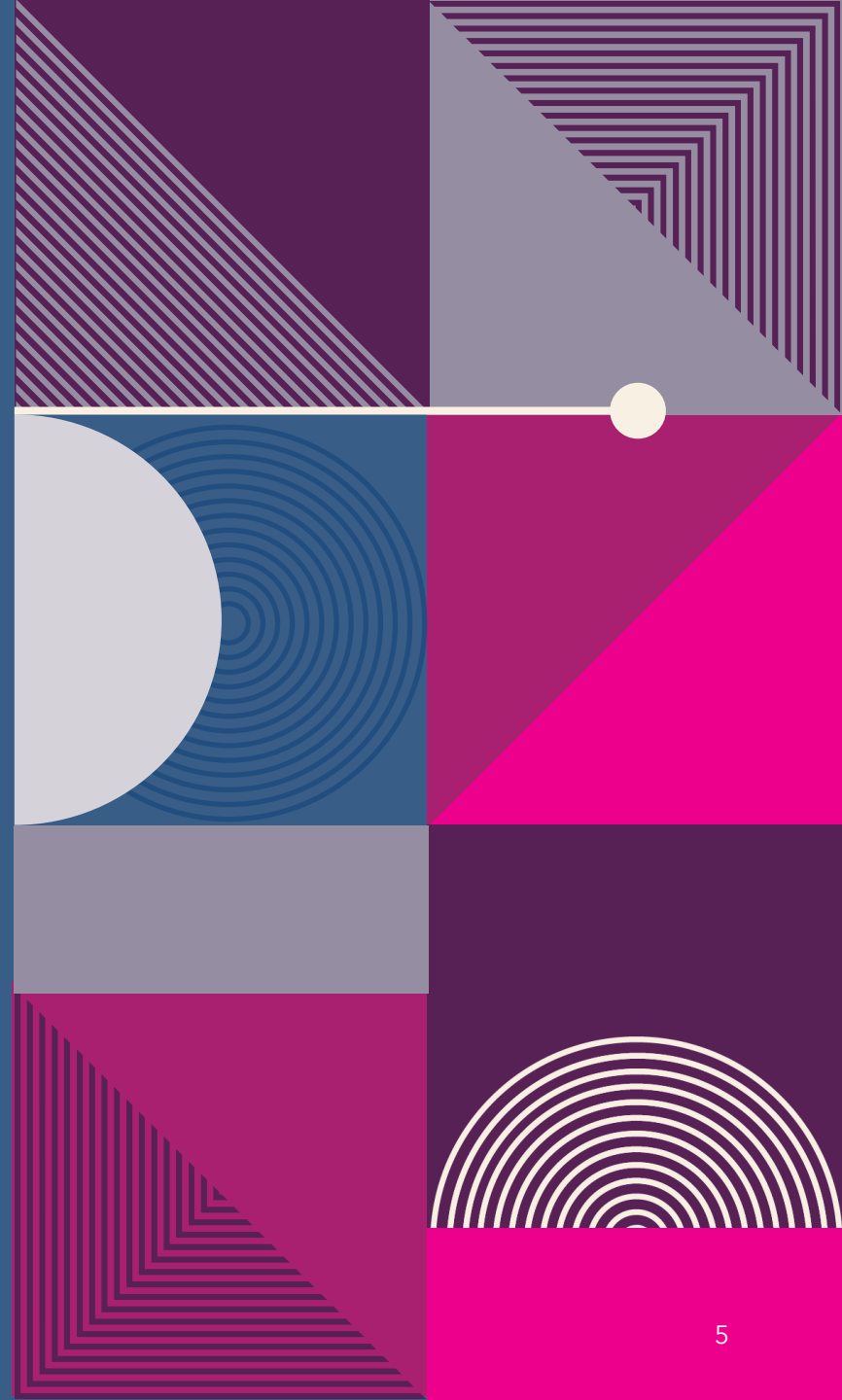
Data Prepearation

EDA with Python and SQL

Predictive Analysis using ML Models

Finding the **Best Model**

Interactive Visuals using Folium and Plotly Dash



DATA COLLECTION

Scrapping:

Using requests module

Using Beautiful Soap and html parser

API-calls:

SpaceX Api

DATA PREPARATION

Focused on Falcon9

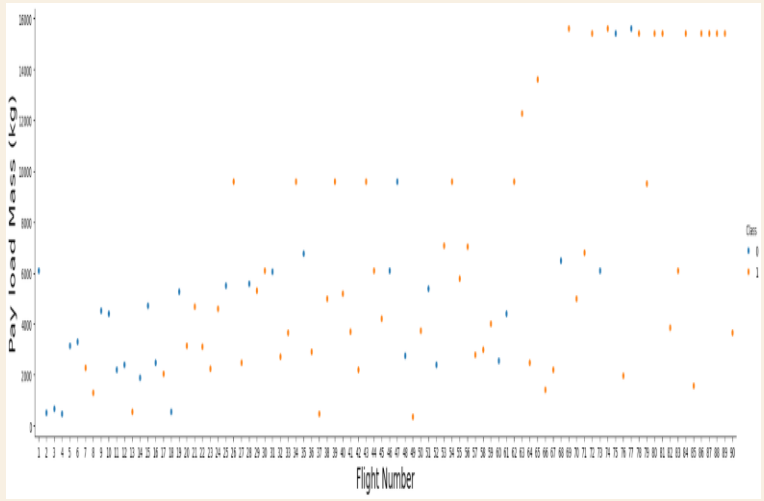
Fixed the null values

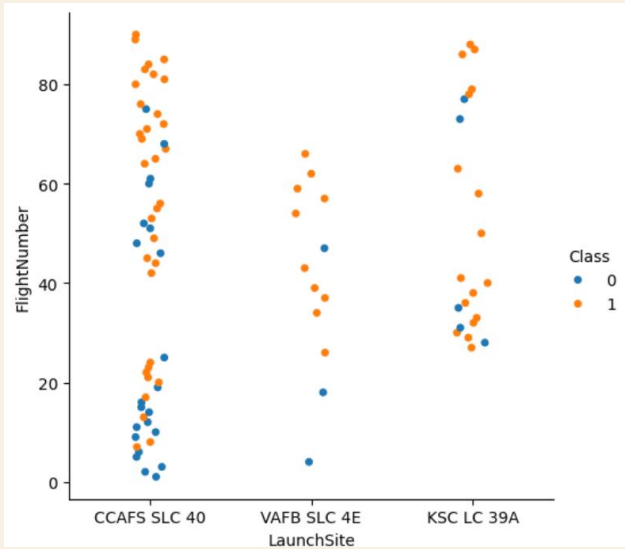
Extracted info for each Orbit, Launch Site Payload

EDA: CHARTS

Scatter Plots were plotted to check the correlation of variables with each other:

- Flight Number
- Payload Mass
- Launch Site
- Orbit
- Success Rate, and Yearly trend for the success





EDA: WITH SQL

Explored further patterns within the data for payload (4000<payload<6000)

Successful missions and their relevant parameters

Ranking the Landing outcomes for drones and groundpad

INTERACTIVE FOLIUM MAP

- The Project also included a map in Folium, to interactively see the sites and their outcomes using Markers

INTERACTIVE DASHBOARD

- Interactive Dashboard for visualization for the factors that can be manually selected

PREDICTIVE ANALYSIS

- Machine Learning Models were trained with GridSearch CV, cv=10 and the models were:
- Logistic Regression
- Decision Tree
- SVM
- KNN

MODELS

- SVM and KNN performed better than others, generally.
- Decision Tree worked best for Validation Data only.

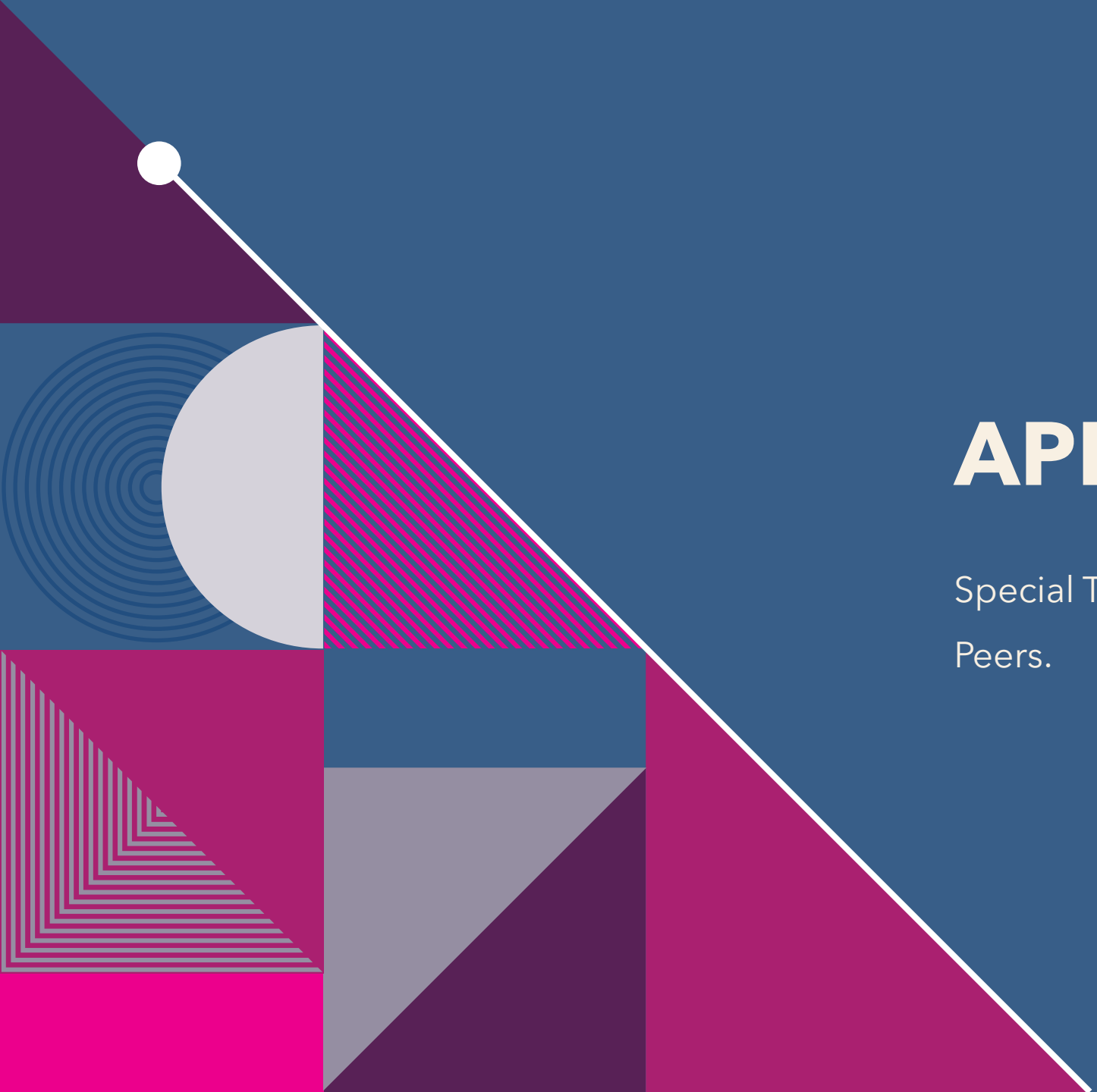
Metric	Test Accuracy	Validation Accuracy
Logistic Regression	0.833	0.8464
SVM	**0.833**	**0.848**
DecisionTree	0.722	0.9017
KNN	**0.833**	**0.848**

CONCLUSIONS

- Best Models were KNN and SVM
- Reused Boosters perform better
- Orbits like LEO, ISS have higher success rates than GTO
- Heavier payloads reduce success probability

SPEAKING ENGAGEMENT METRICS

Impact factor	Measurement	Target	Achieved
Audience interaction	Percentage (%)	85	88
Knowledge retention	Percentage (%)	75	80
Post-presentation surveys	Average rating	4.2	4.5
Referral rate	Percentage (%)	10	12



APPENDIX

Special Thanks to Coursera, Instructors, and Fellow-Peers.