IBM Data Science Certification Project: Falcon 9 Landings

This presentation explores the use of data science and AI tools to predict Falcon 9 first-stage landings, a key factor in SpaceX's cost-effective space travel.

MS **Mariia Slovinska**

Outline

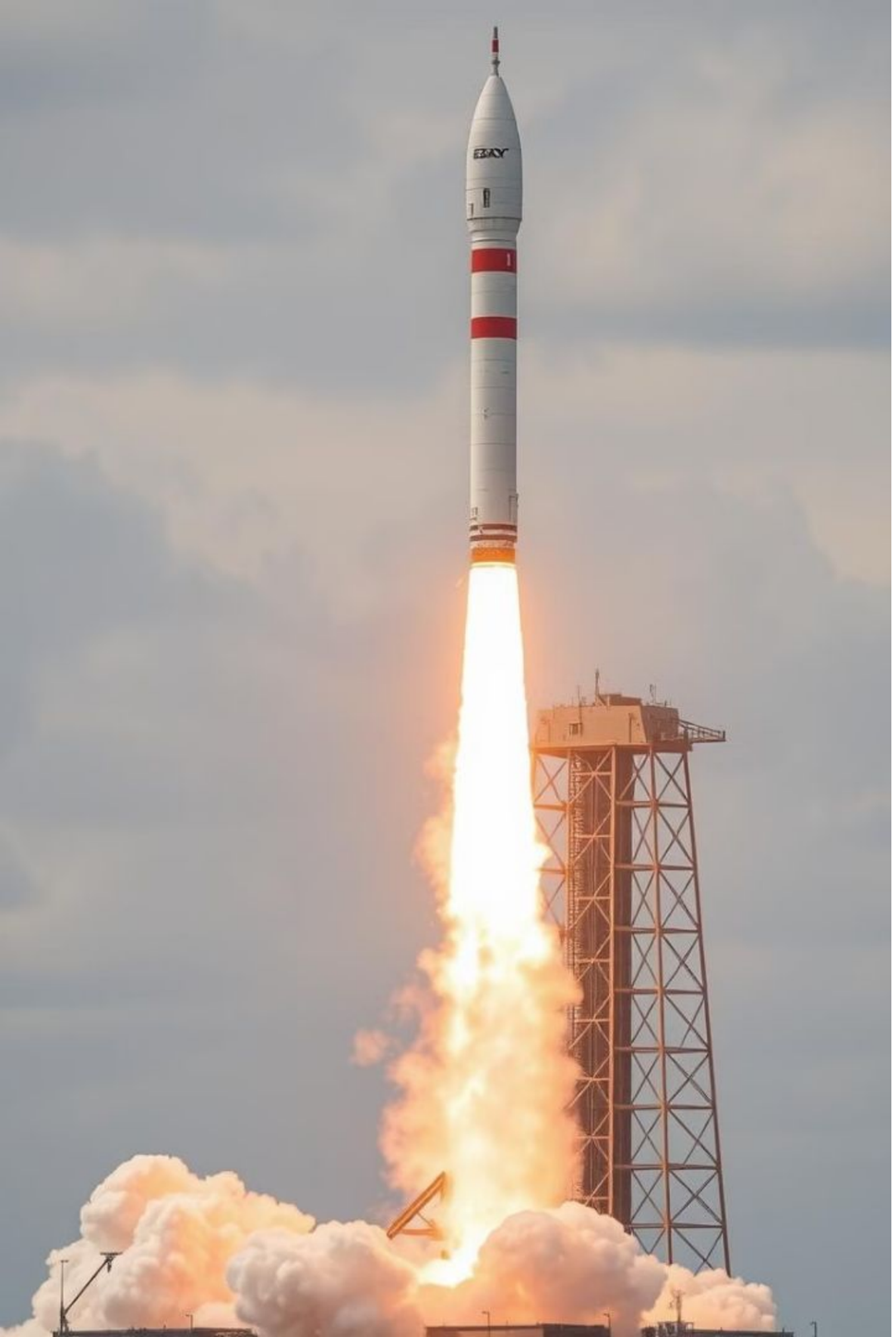
Executive Summary

Methodology

The project used machine learning to predict Falcon 9 landings based on a variety of factors.

Results

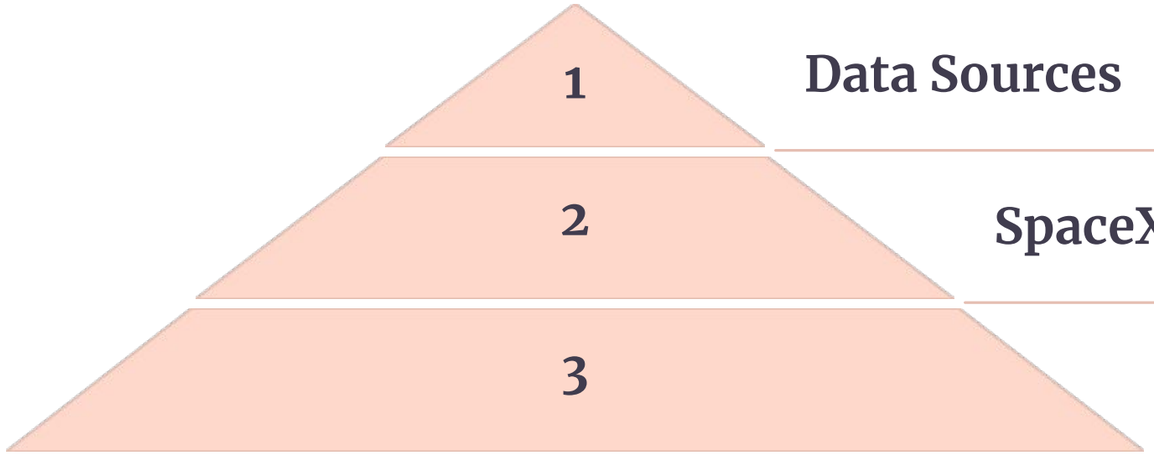
The models achieved significant accuracy in predicting successful landings, providing valuable insights for cost analysis and decision-making.



Introduction

SpaceX has revolutionized space travel by reducing launch costs, a feat largely attributed to its reusable Falcon 9 first stages. This project explores how machine learning can predict the success of these landings, aiding in cost analysis and decision-making for future launches.

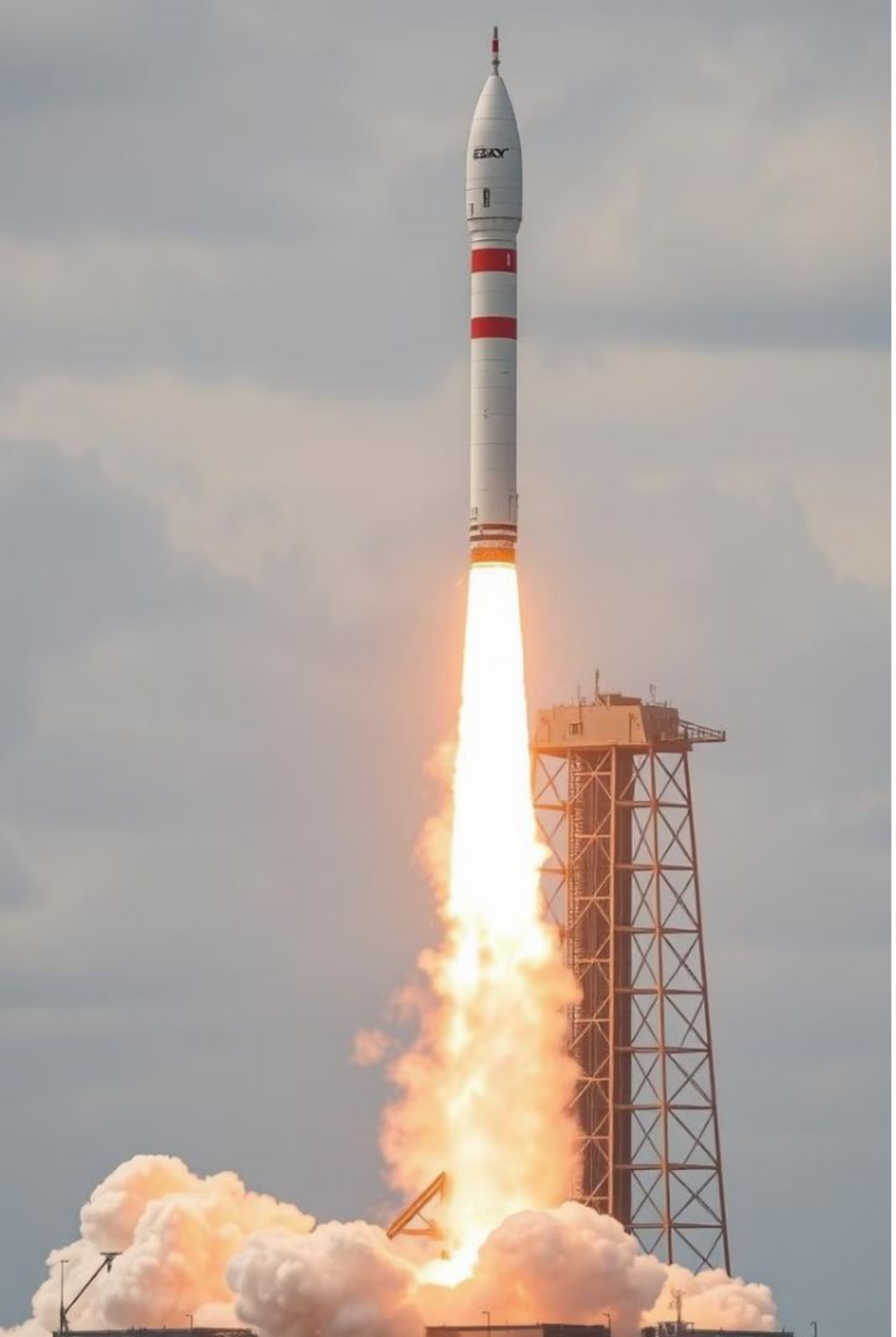
Methodology: Data Collection



Data Sources

SpaceX API

Web Scraping



Complete the Data Collection API Lab

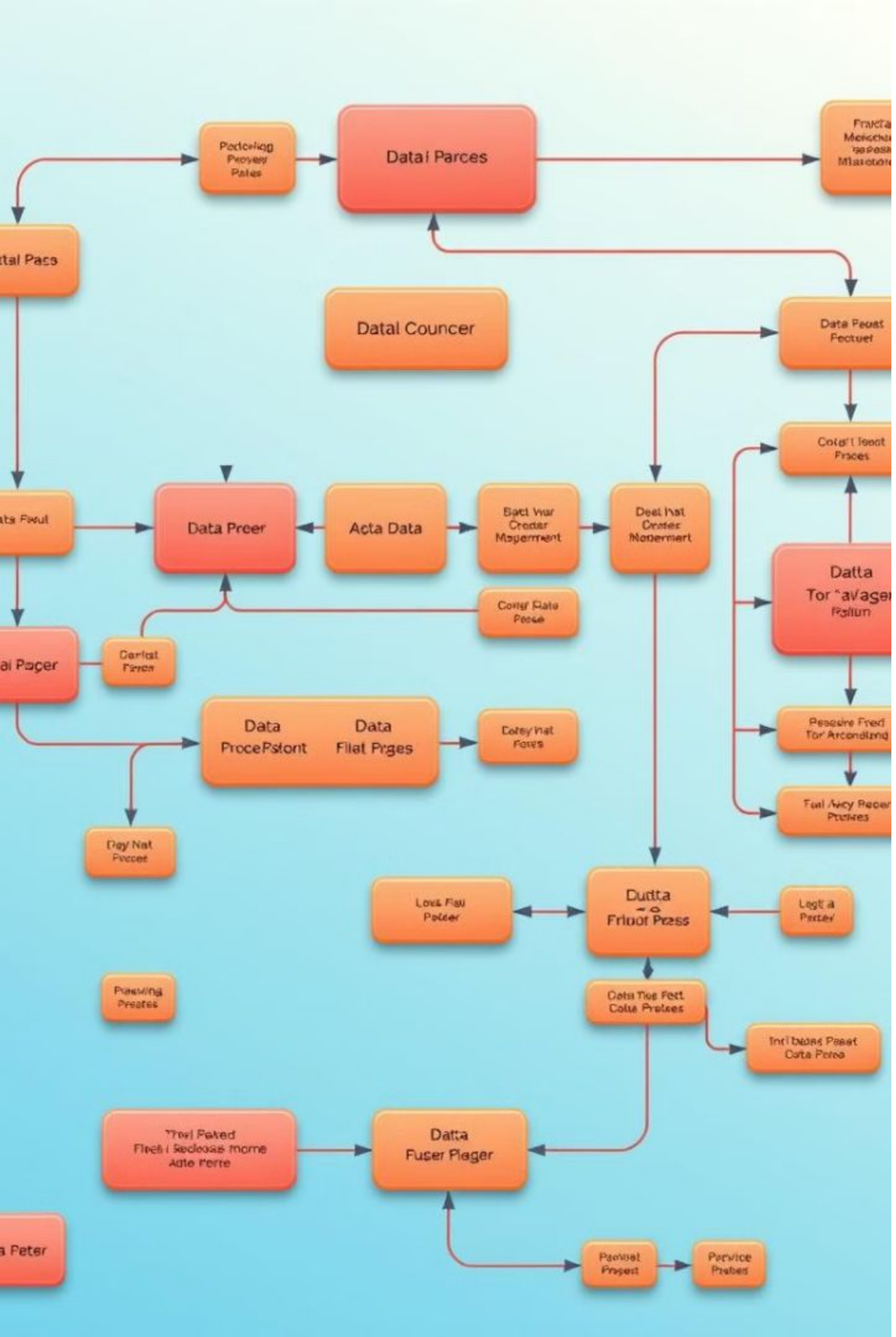
Git: HERE

Complete the Data Collection with Web Scraping lab

Git: 2nd_jupyter-labs-webscraping.ipynb

Data Wrangling

Git: 3rd_labs-jupyter-spacex-Data wrangling-v2.ipynb



Methodology: Data Wrangling

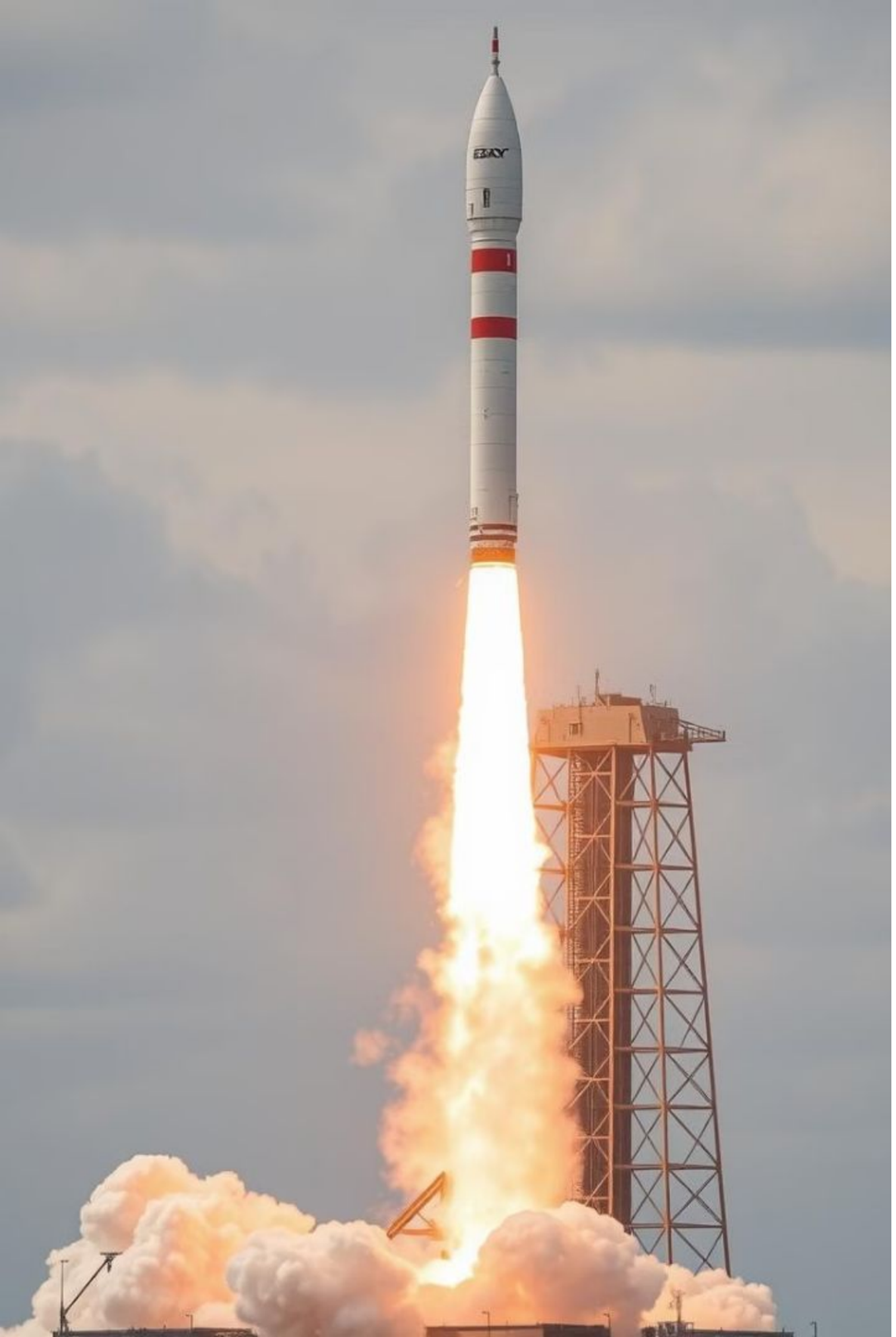
After collecting data, the next step was data wrangling, transforming raw data into a format suitable for analysis and modeling. This involved cleaning, standardizing, and integrating the collected data sets.

Methodology: Exploratory Data Analysis (EDA)

EDA involved using visualizations and SQL queries to explore the collected data and identify patterns and relationships. This helped in understanding the underlying characteristics of successful and unsuccessful landings.

Complete the EDA with SQL

Git: 4rd_jupyter-labs-eda-sql-coursera_sqllite.ipynb



EDA with Visualization Lab

Git: 5th_jupyter-labs-eda-dataviz-v2.ipynb

Interactive Visual Analytics with Folium lab

Git: 6th_lab-jupyter-launch-site-location-v2.ipynb

Build an Interactive Dashboard with Ploty Dash

Git:

7th_Interactive_Dashboard_with_Ploty_Dash.ipynb

SpaceX Launch Records Dashboard

All Sites

Total Success Launches By Sites

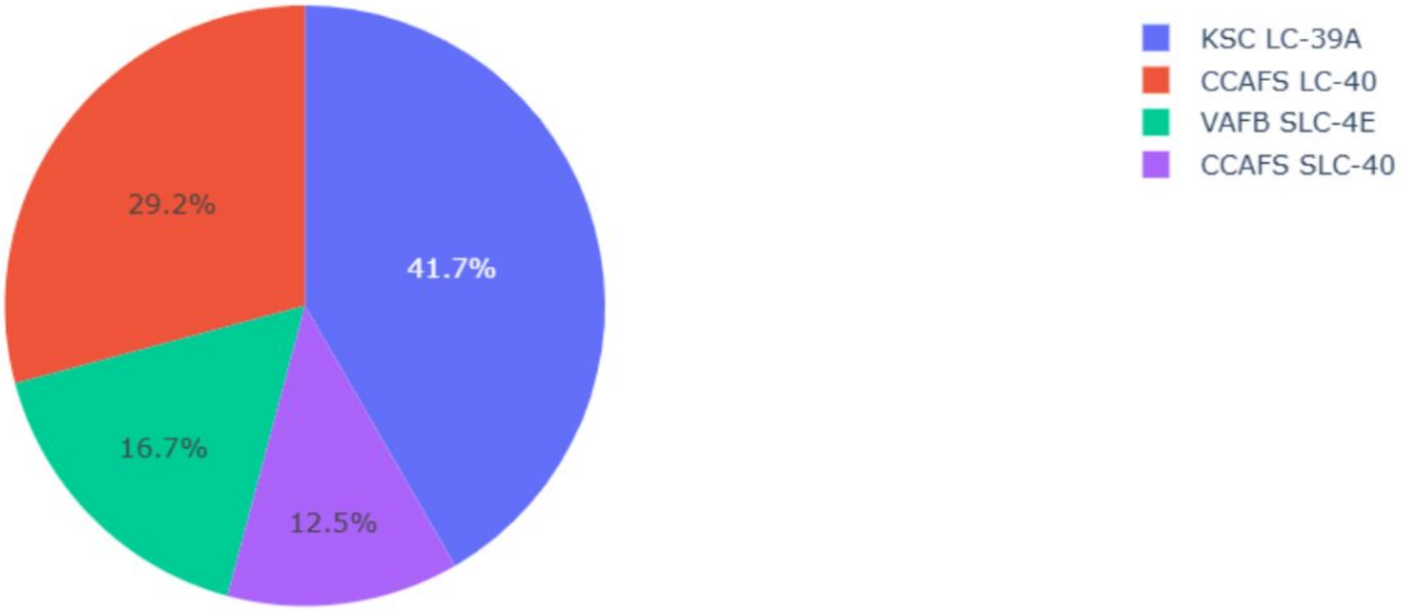
Payload range (Kg):

Payload range (Kg):

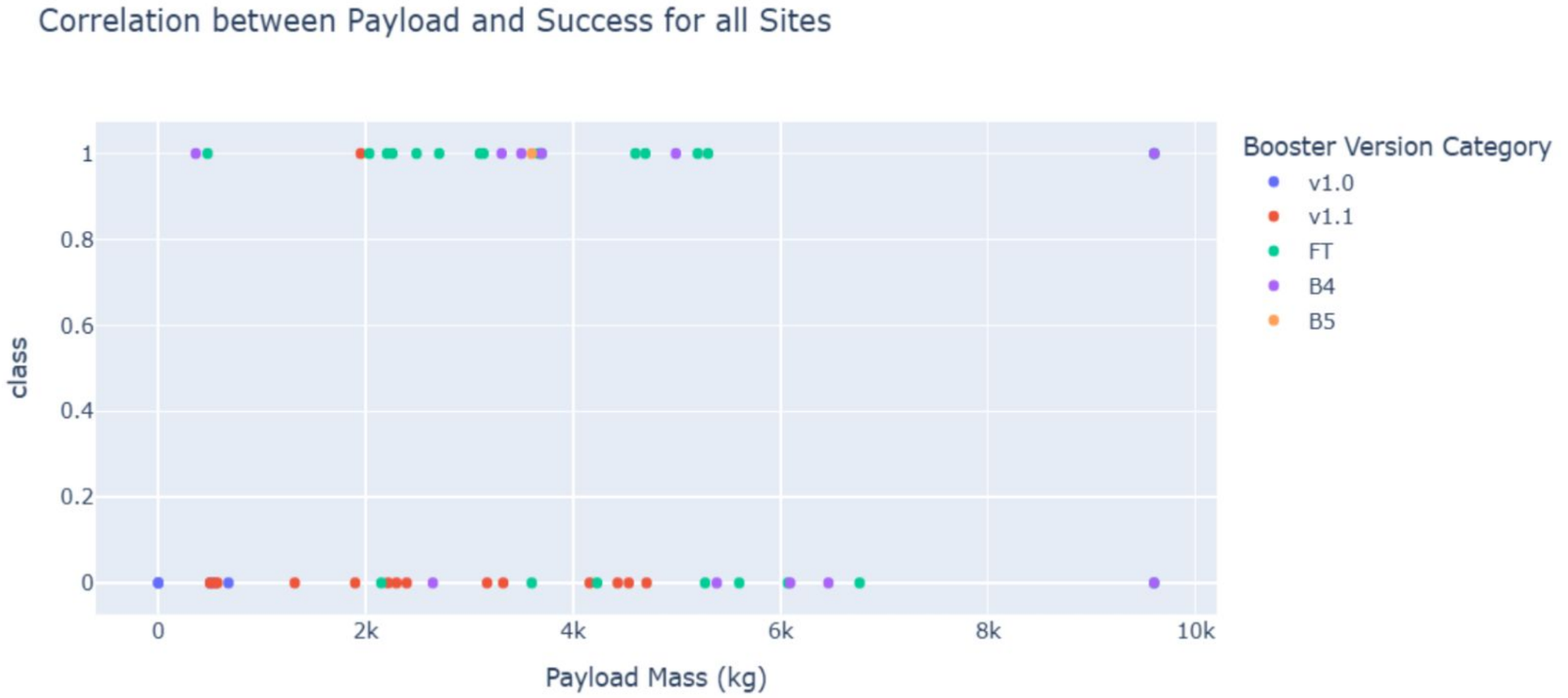
SpaceX Launch Records Dashboard

All Sites

Total Success Launches By Sites



Payload range (Kg):



Methodology: Predictive Analysis

The final step in the methodology involved building and evaluating classification models to predict the likelihood of a successful Falcon 9 landing. This involved training the models on historical data and using them to predict future outcomes.

Predictive Analysis: Falcon 9 First Stage Landing Prediction

GIT: [8th_SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb]

I build a machine learning pipeline to predict the success of Falcon 9's first stage landing.

The process includes data preprocessing, splitting the data into training and testing sets, and applying Grid Search to find optimal hyperparameters.

Test various models - **Logistic Regression, Support Vector Machines, Decision Tree Classifier, and K-Nearest Neighbors**

-to determine the best-performing model.

The final step involves evaluating the model's performance using a confusion matrix to measure accuracy.

Results & Conclusion

Results

The models achieved a high degree of accuracy in predicting successful Falcon 9 landings, demonstrating the potential of data science to inform decision-making in the space industry.

Conclusion

This project highlights the power of data-driven solutions in advancing space exploration and driving cost-efficiency. As the space industry continues to evolve, data science will play a crucial role in enabling more efficient and sustainable ventures.

Test Accuracies and Best Model

4 models were evaluated: Logistic Regression, SVM, Decision Tree, and KNN.

Logistic Regression and KNN achieved the highest test accuracy (94.4%).

Logistic Regression was selected as the best model due to its accuracy and consistency.

GridSearchCV was used to find the best hyperparameters for each model. The table displays the best parameters and validation accuracies achieved.

While Decision Tree had the highest validation accuracy, Logistic Regression proved more reliable on the test data.

Appendix

This repository contains Jupyter Notebooks for the IBM Data Science Capstone project, focusing on applying data science to solve real-world problems. As a Data Scientist for a startup competing with SpaceX, I collected, processed, and analyzed data to predict whether the Falcon 9's first stage would land successfully. The project includes data collection, exploration, predictive modeling, and final deliverables, showcasing practical data science applications.

My Git: slovinskamary/Applied-Data-Science-Capstone/tree/main

Key Findings

Payload mass, orbit type, and launch sites significantly influence launch outcomes.

SpaceX's success rate has steadily improved, showcasing continuous operational enhancements.

Visualization Techniques

Interactive dashboards and maps facilitated the identification of launch success and failure patterns and trends across missions.

Predictive Modeling

Logistic Regression and KNN models achieved 94.4% test accuracy.

Hyperparameter tuning significantly impacted model performance.

Project Takeaways

SpaceX's iterative learning from each launch enhances reliability and reduces costs.

Predictive models are valuable for planning future missions and maintaining a competitive edge.

This project demonstrates the effectiveness of data science and machine learning in addressing complex challenges in space exploration.

Thank you!

Feel free to connect with me on LinkedIn for more updates and insights!

[My linkedin here: Mariia Slovinska]