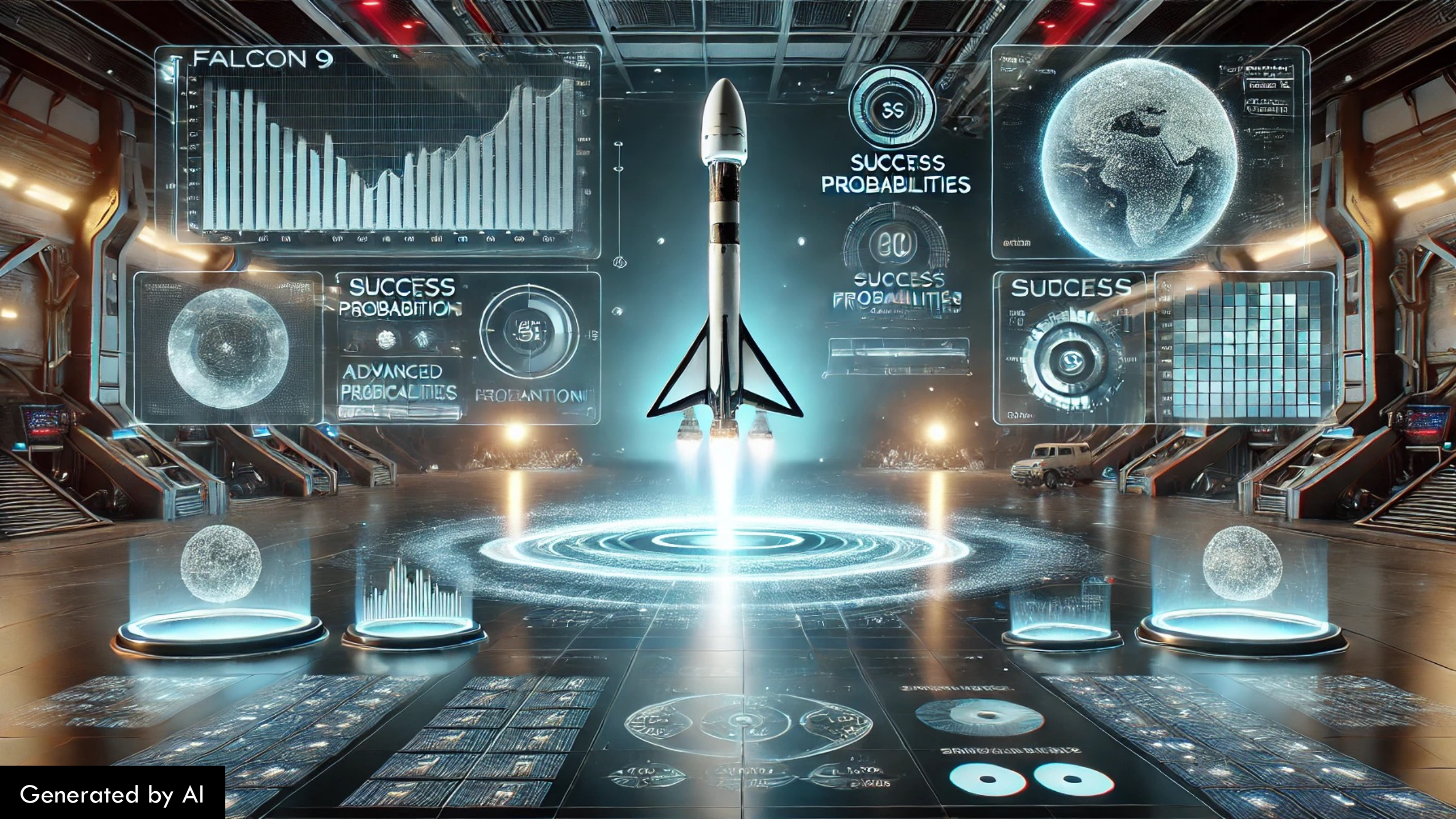

FALCON 9
SUCCESS PROBABILITIES
SUCCESS
ADVANCED
Generated by AI

APPLIED DATA SCIENCE CAPSTONE

Emiljano Agalliu

February 17, 2025

Table of Content

EXECUTIVE SUMMARY

EXECUTION SUMMARY

This Applied Data Science Capstone project focuses on leveraging machine learning techniques to predict the success of SpaceX's Falcon 9 first-stage landings. The goal is to employ classification algorithms to forecast landing outcomes, using data from previous rocket launches. The project follows a well-defined, systematic approach, outlined in the stages below:

Data Acquisition and Preprocessing
We began by sourcing, cleaning, and transforming raw data to ensure its readiness for analysis. This crucial step laid the foundation for robust model development.
first-stage landing would succeed. These models were trained on a diverse set of features and evaluated for their accuracy.

Dynamic Data Visualization
Interactive visualizations were crafted to showcase critical trends and correlations, enhancing the interpretability of the data and providing a clear understanding of underlying patterns.

Exploratory Data Analysis (EDA)
Through comprehensive analysis, we explored the dataset to identify patterns and relationships between key variables. This phase provided valuable insights into the features most influential in predicting landing success.

Predictive Modeling
Various machine learning algorithms were applied to create predictive models, with the aim of determining whether Falcon 9's first-stage landing would succeed. These models were trained on a diverse set of features and evaluated for their accuracy.

Key findings revealed that certain factors, such as launch success and rocket specifications, significantly correlate with landing outcomes. Among the models tested, decision trees emerged as the most accurate in predicting landing success.

Ultimately, this project underscores the potential of machine learning to enhance predictive capabilities within space missions and supports SpaceX's ongoing advancements in rocket technology.

INTRODUCTION

INTRODUCTION

This capstone project focuses on predicting the success of SpaceX's Falcon 9 first-stage landings, a critical factor in reducing launch costs. SpaceX's Falcon 9 rockets are priced at $62 million, a substantial reduction compared to other providers who charge over $165 million. The ability to reuse the first stage of the Falcon 9 is central to this cost-saving approach. By forecasting the likelihood of a successful landing, we can provide key insights into launch cost optimization and offer competitive intelligence for alternative companies considering bids against SpaceX.

While most unsuccessful landings are planned, with some occurring through controlled ocean landings, the project leverages a variety of launch-specific features—such as payload mass, orbit type, and launch site—to model the factors that influence landing success. This predictive capability supports SpaceX's goal of improving efficiency and driving the future of reusable rocket technology.

METHODOLOGY

METHODOLOGY

Overview

The methodology for this project follows a comprehensive and structured approach to ensure accurate predictions of Falcon 9 first-stage landing success. The process is broken down into several key stages:

Data Collection and Preparation

- **SpaceX API:** Data was retrieved directly from SpaceX's API for reliable and real-time information.
- **Web Scraping:** Additional data sources were gathered through web scraping techniques to complement the API data.
- **Data Wrangling and Formatting:** Raw data was cleaned, transformed, and formatted to ensure consistency and usability for analysis.

Exploratory Data Analysis (EDA)

- **Pandas & NumPy:** These libraries were used for data manipulation and statistical analysis to identify key patterns and trends.
- **SQL:** SQL queries were employed to extract specific subsets of data for deeper analysis and feature extraction.

Data Visualization

- **Matplotlib & Seaborn:** These visualization tools were used to create insightful charts and graphs, helping to identify trends and relationships within the dataset.
- **Folium:** Geospatial data was visualized using Folium to map launch sites and landing locations.
- **Dash:** An interactive dashboard was developed using Dash to allow for dynamic exploration of the data.

Machine Learning Prediction

- **Logistic Regression:** Used for binary classification to predict success or failure.
- **Support Vector Machine (SVM):** Applied for classification, offering a robust approach to complex datasets.
- **Decision Tree:** Utilized to build interpretable models for predicting landing outcomes.
- **K-Nearest Neighbors (KNN):** Used to identify patterns based on proximity to similar data points, aiding in prediction accuracy.

RESULTS

RESULTS

Exploratory Data Analysis with SQL

```
Records where launch sites begin with the string 'CCA'
('2010-06-04', '18:45:00', 'F9 v1.0  B0003', 'CCAFS LC-40', 'Dragon Spacecraft Qualification Unit', 0, 'LEO', 'SpaceX', 'Success', 'Failure (parachute)')
('2010-12-08', '15:43:00', 'F9 v1.0  B0004', 'CCAFS LC-40', 'Dragon demo flight C1, two CubeSats, barrel of Brouere cheese', 0, 'LEO (ISS)', 'NASA (COTS) NRO', 'Success', 'Fai
('2012-05-22', '7:44:00', 'F9 v1.0  B0005', 'CCAFS LC-40', 'Dragon demo flight C2', 525, 'LEO (ISS)', 'NASA (COTS)', 'Success', 'No attempt')
('2012-10-08', '0:35:00', 'F9 v1.0  B0006', 'CCAFS LC-40', 'SpaceX CRS-1', 500, 'LEO (ISS)', 'NASA (CRS)', 'Success', 'No attempt')
('2013-03-01', '15:10:00', 'F9 v1.0  B0007', 'CCAFS LC-40', 'SpaceX CRS-2', 677, 'LEO (ISS)', 'NASA (CRS)', 'Success', 'No attempt')
```

```
Names of the unique launch sites in the space mission
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
```

```
Names of the booster_versions which have carried the maximum payload mass.
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7
```

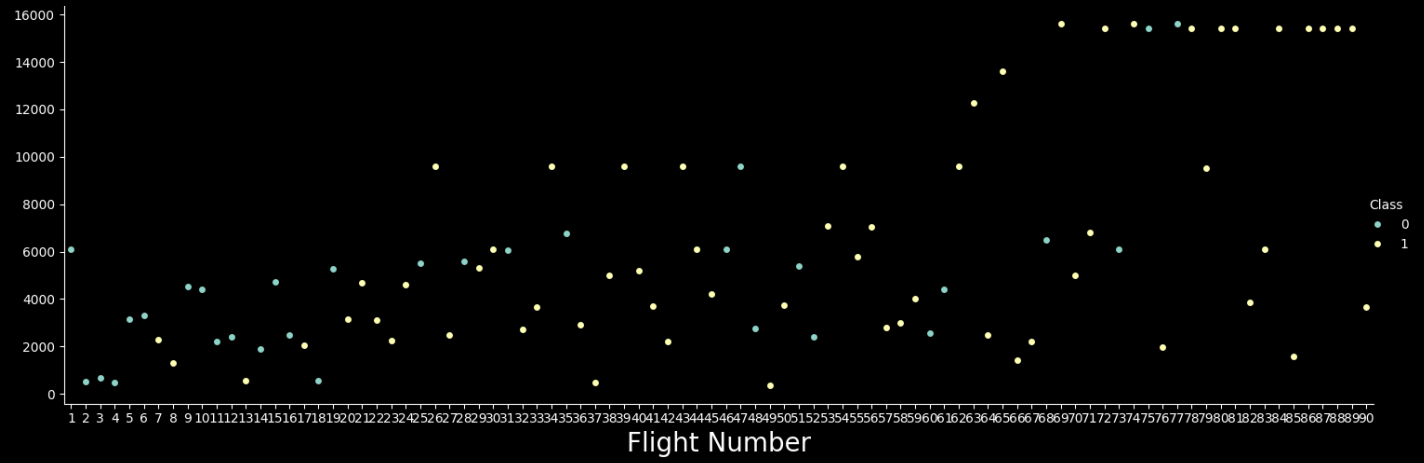
```
Total payload mass carried by boosters launched by NASA (CRS): None
Average payload mass carried by booster version F9 v1.1: 2928.4
The date when the first successful landing outcome in ground pad was achieved: 2015-12-22
Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
F9 FT B1022
F9 FT B1026
F9 FT  B1021.2
F9 FT  B1031.2
Total successful missions: 100
Total failed missions: 1
```

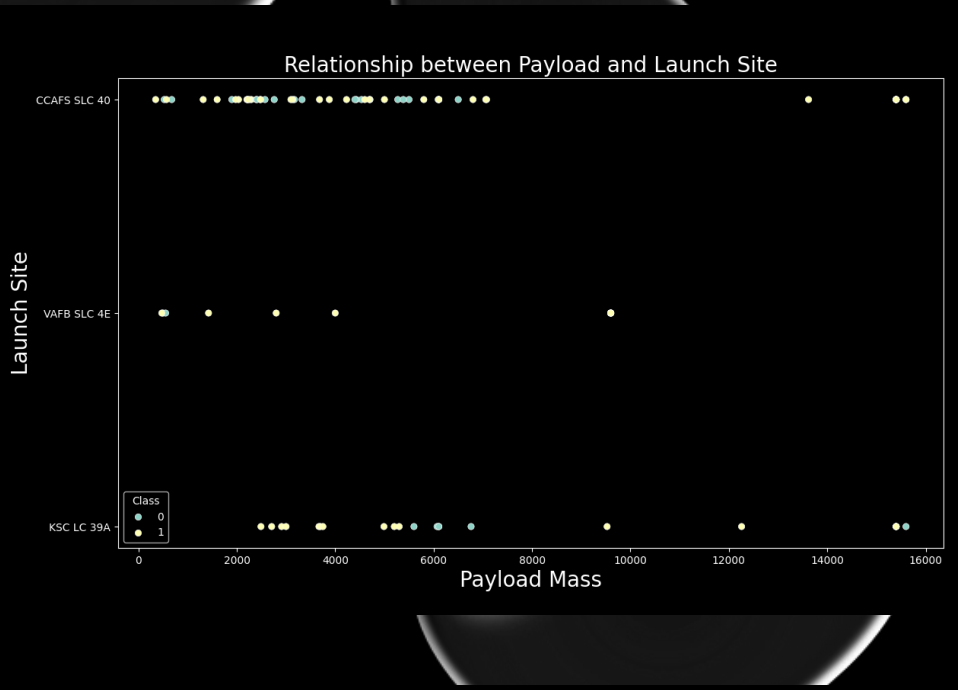
```
Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015.
('January', 'Failure (drone ship)', 'F9 v1.1 B1012', 'CCAFS LC-40')
('April', 'Failure (drone ship)', 'F9 v1.1 B1015', 'CCAFS LC-40')
Landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.
Landing Outcome: No attempt, Count: 10
Landing Outcome: Success (drone ship), Count: 5
Landing Outcome: Failure (drone ship), Count: 5
Landing Outcome: Success (ground pad), Count: 3
Landing Outcome: Controlled (ocean), Count: 3
Landing Outcome: Uncontrolled (ocean), Count: 2
Landing Outcome: Failure (parachute), Count: 2
Landing Outcome: Precluded (drone ship), Count: 1
```

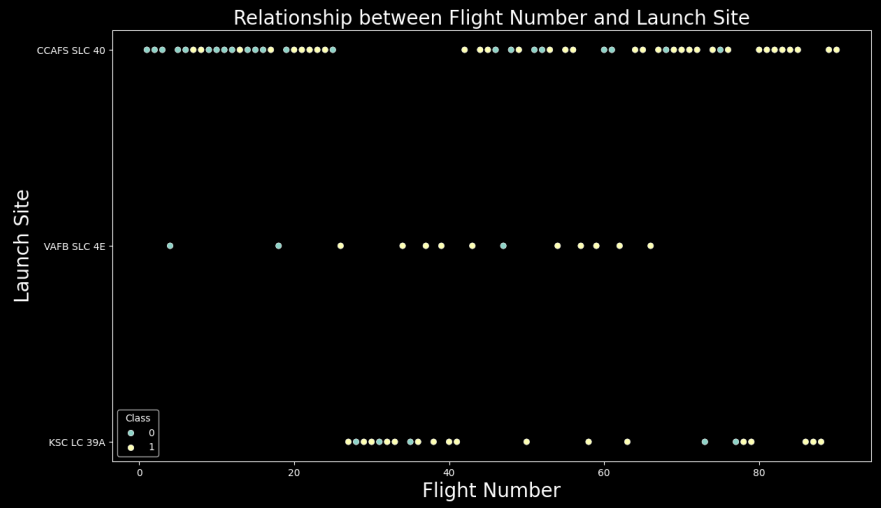
RESULTS

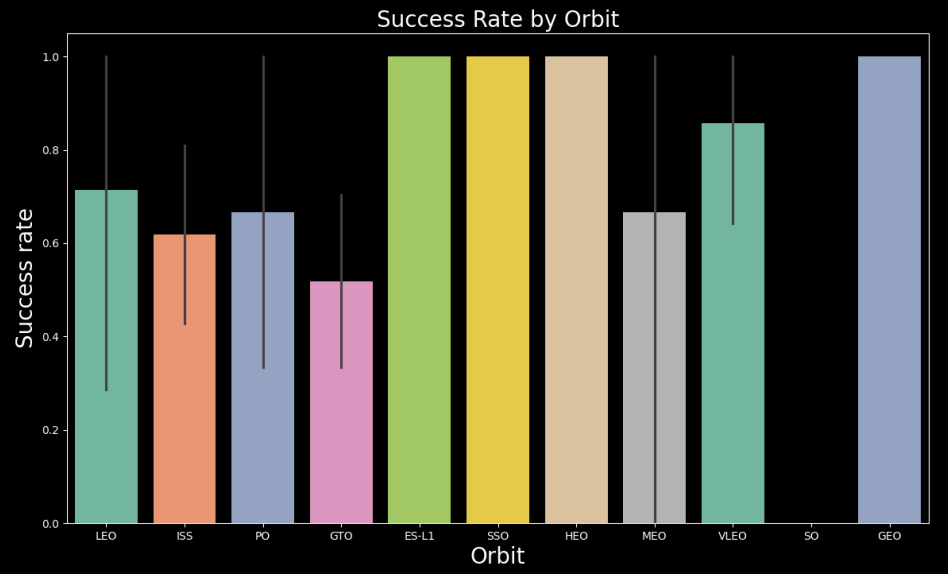
Exploratory Data Analysis with Data Visualization

Matplotlib & Seaborn





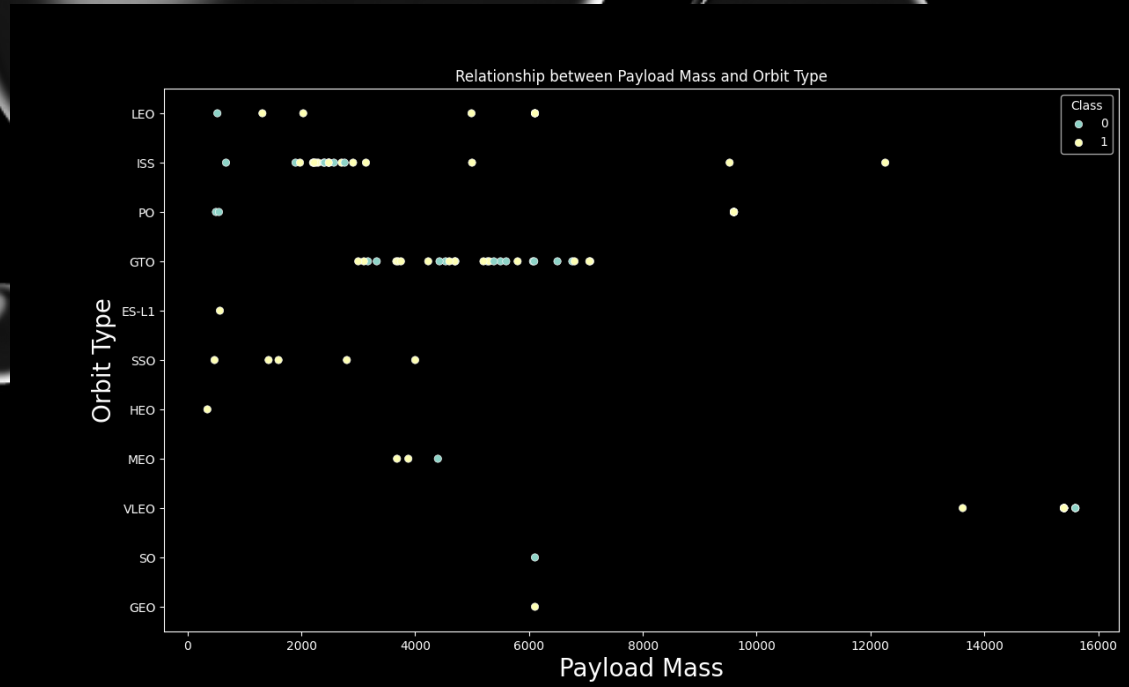


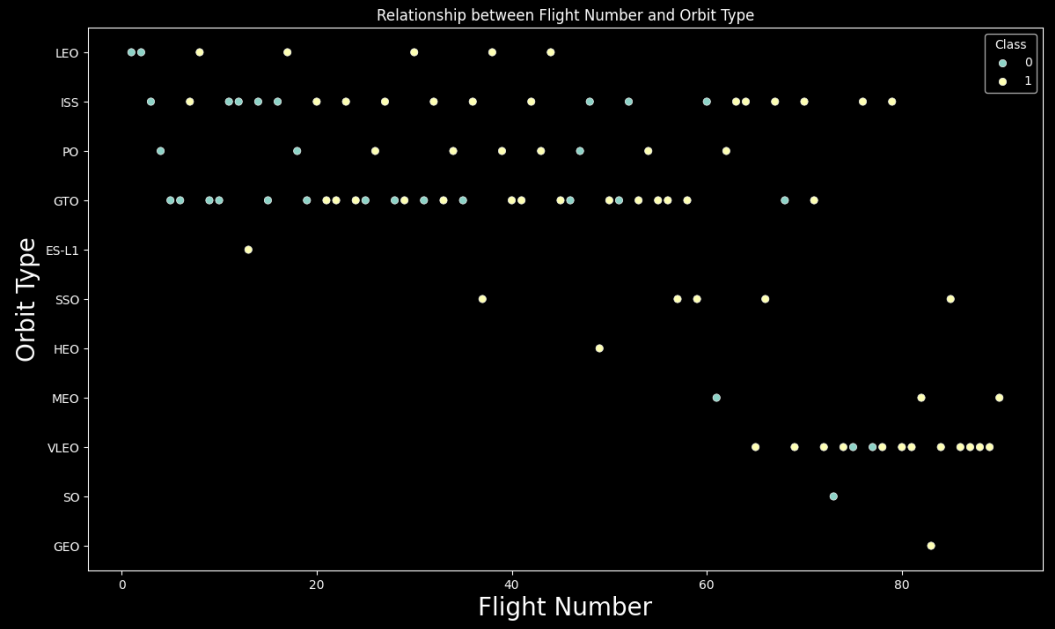


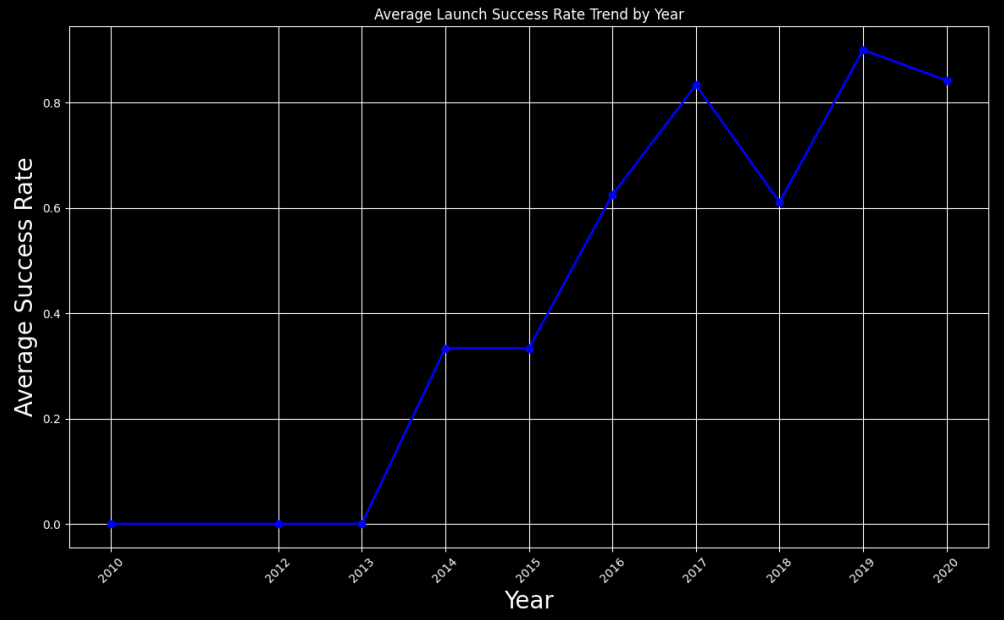
RESULTS

Exploratory Data Analysis with Data Visualization

Matplotlib & Seaborn



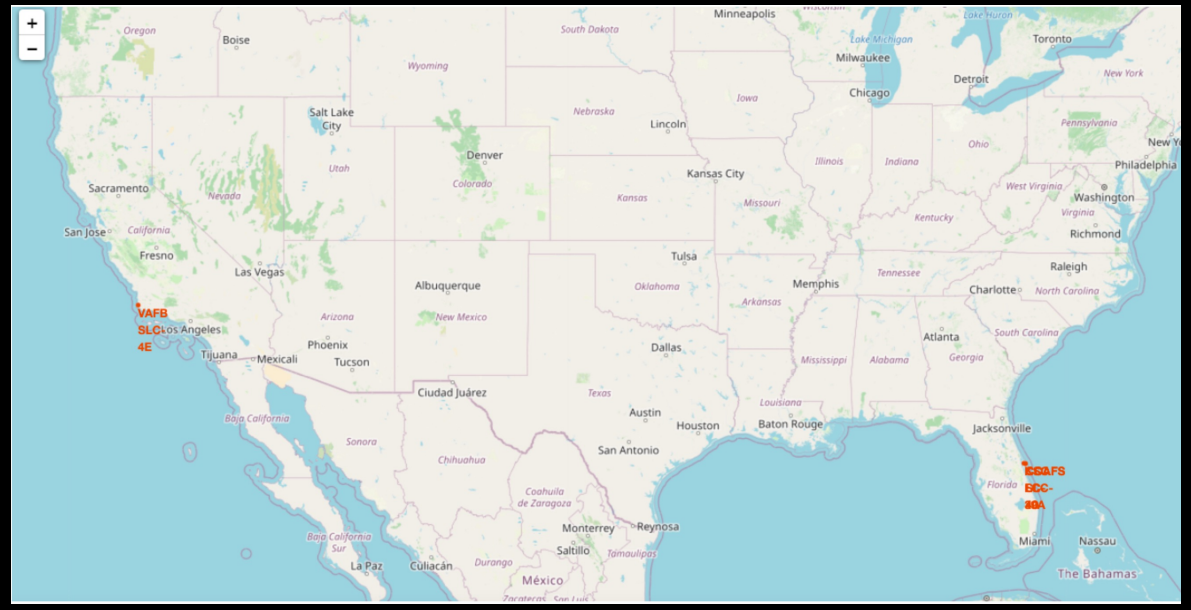


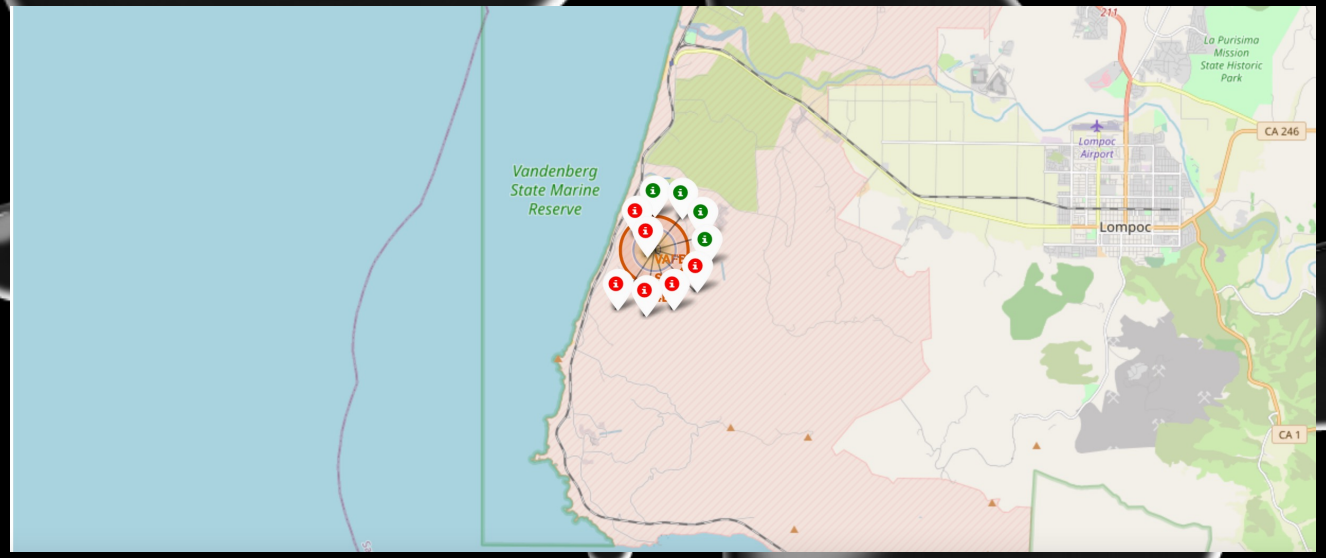


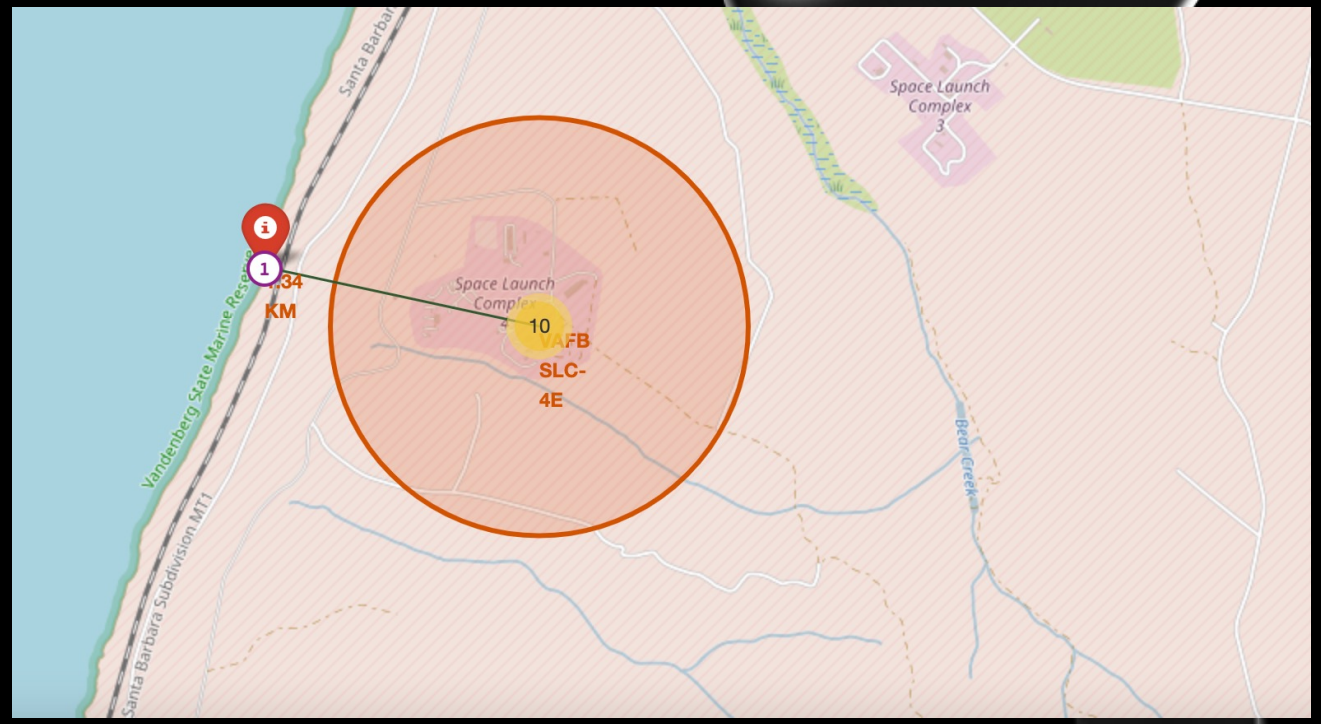
RESULTS

Interactive Visual Analytics and Dashboard

Folium



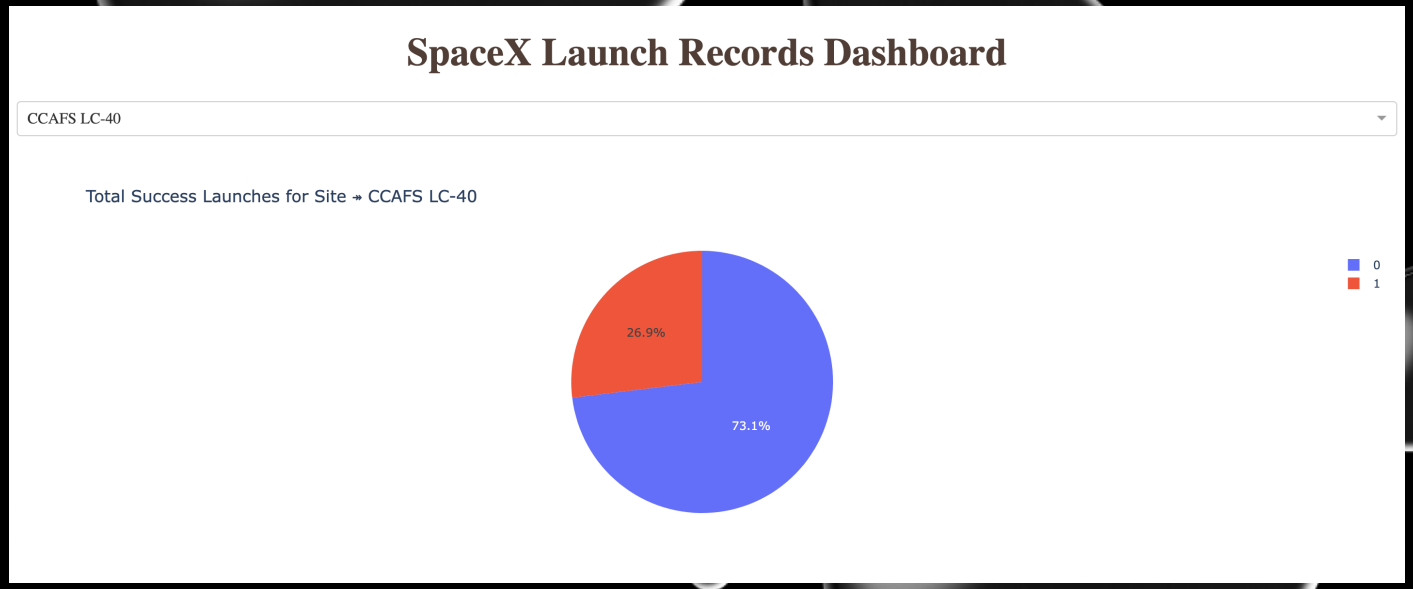


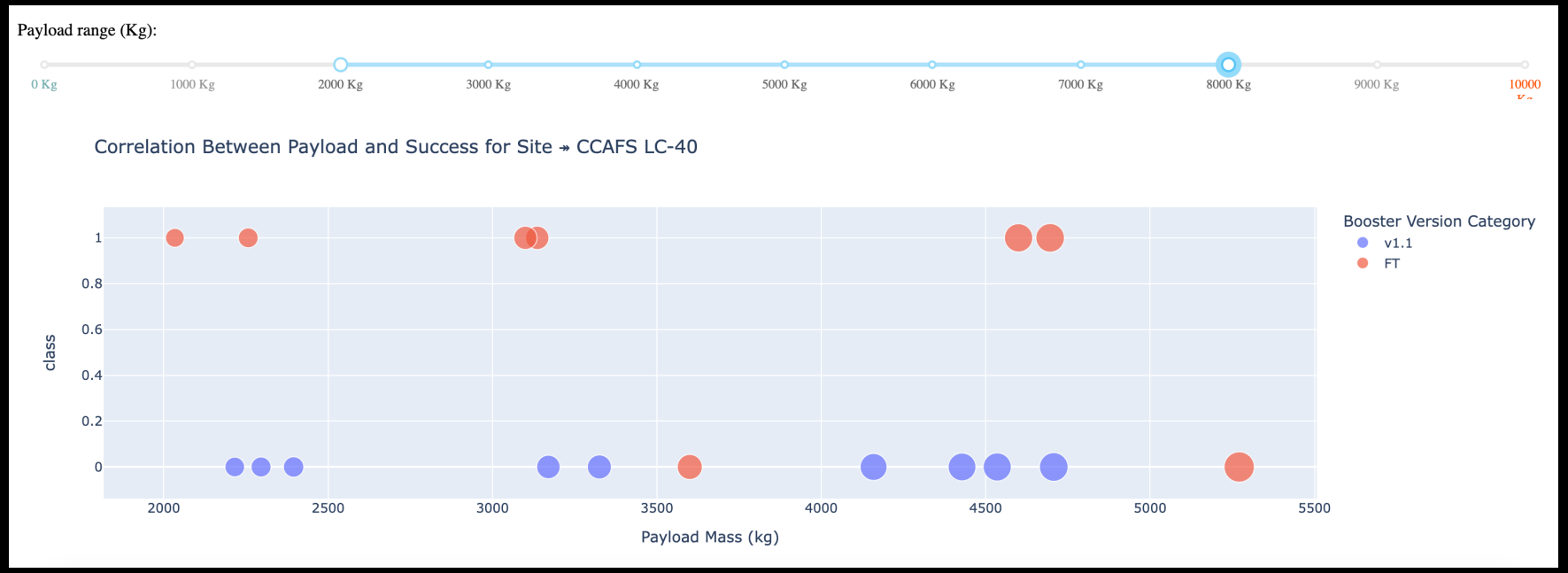


RESULTS

Interactive Visual Analytics and Dashboard

Dash



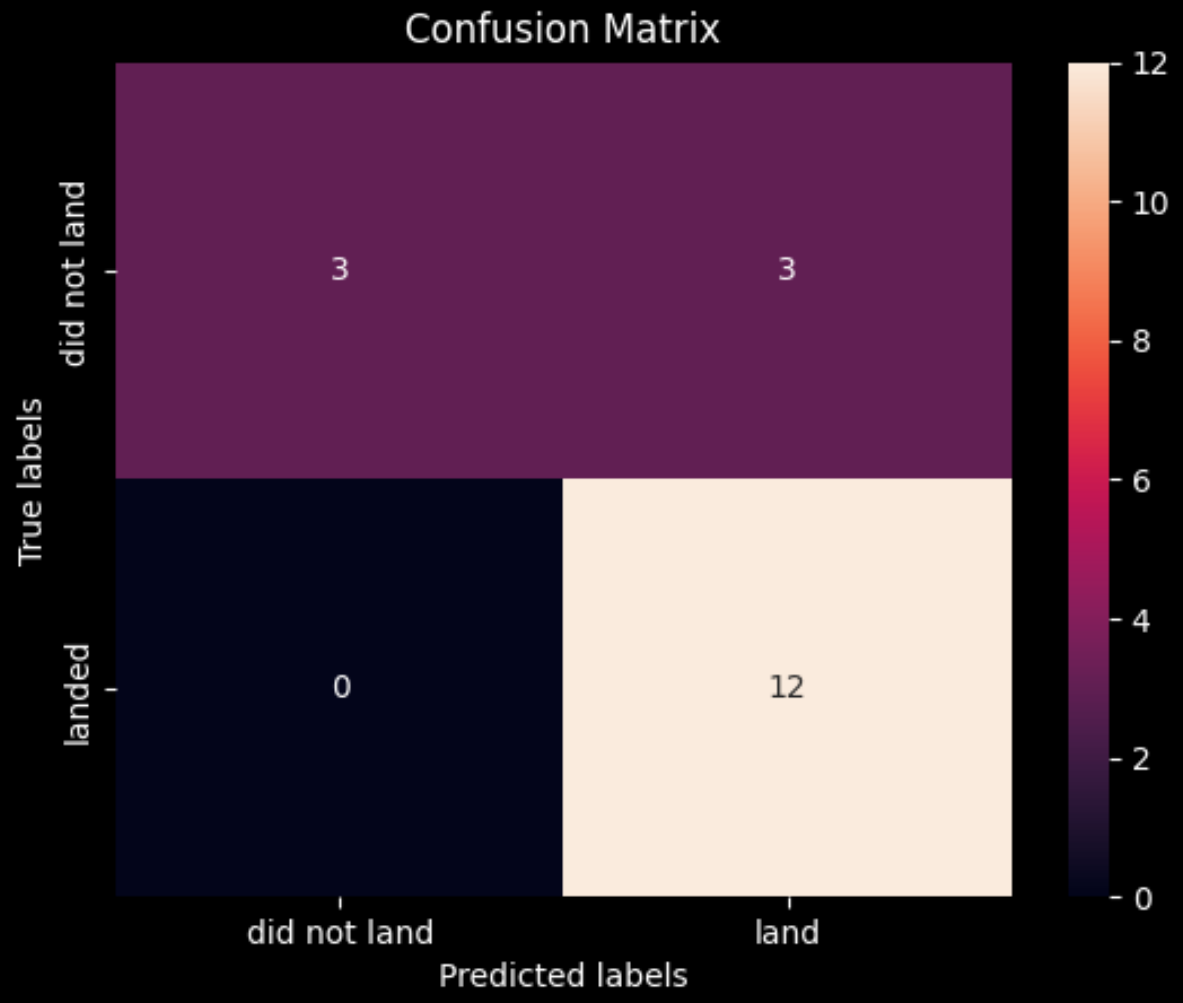


RESULTS

Machine Learning Prediction

- **Logistic Regression**
- **SVM**
- **Decision Tree**
- **KNN**

Confusion Matrix



The confusion matrix from all the models gives us the same results as:

	Predicted: Did Not Land	**Predicted: Land**
True: Did Not Land	**3 (True Negative - TN)**	**3 (False Positive - FP)**
True: Landed	**0 (False Negative - FN)**	**12 (True Positive - TP)**

Interpretation:

- True Positive: Correctly predicted positive 12 cases.
- True Negative: Correctly predicted negative 3 cases.
- False Positive: Incorrectly predicted positive 3 cases - Type I Error.
- False Negative: Incorrectly predicted negative 0 cases - Type II Error.

RESULTS

Machine Learning Prediction

- **Logistic Regression**
- **SVM**
- **Decision Tree**
- **KNN**

Comparison

	Model	Best score
0	Logistic regresssion	0.860714
1	SVM	0.848214
2	Decision tree	0.887500
3	KNN	0.848214

The best model is: Decision Tree with an accuracy of 0.8875

Logistic Regression
Best LogReg parameters found: {'C': 0.1, 'penalty': 'l1', 'solver': 'saga'}
Tuned LogReg Hyperparameters :(best parameters) {'C': 0.1, 'penalty': 'l1', 'solver': 'saga'}
Accuracy LogReg: 0.8607142857142855
Accuracy LogReg on the test data: 0.8333333333333334

SVM
Best SVM parameters found: {'C': np.float64(1.0), 'gamma': np.float64(0.03162277660168379), 'kernel': 'sigmoid'}
Tuned SVM Hyperparameters :(best parameters) {'C': np.float64(1.0), 'gamma': np.float64(0.03162277660168379), 'kernel': 'sigmoid'}
Accuracy SVM: 0.8482142857142856
Accuracy SVM on the test data: 0.8333333333333334

Decision Tree
Best DTree parameters found: {'criterion': 'gini', 'max_depth': 12, 'max_features': 'log2', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'best'}
Tuned DTree Hyperparameters :(best parameters) {'criterion': 'gini', 'max_depth': 12, 'max_features': 'log2', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'best'}
Accuracy DTree: 0.9017857142857142
Accuracy DTree on the test data: 0.7222222222222222

KNN
Best KNN parameters found: {'algorithm': 'auto', 'metric': 'manhattan', 'n_neighbors': 10, 'p': 1, 'weights': 'uniform'}
Tuned KNN Hyperparameters :(best parameters) {'algorithm': 'auto', 'metric': 'manhattan', 'n_neighbors': 10, 'p': 1, 'weights': 'uniform'}
Accuracy KNN: 0.8482142857142858
Accuracy KNN on the test data: 0.8333333333333334

REVIEW

REVIEW

The data visualization analysis suggests that certain features are potentially correlated with the mission outcome, with varying impacts across different scenarios. For example, when analyzing the relationship between payload weight and mission success, we observe that missions with heavier payloads tend to have a higher probability of successful landings, particularly in orbit types such as Polar, LEO, and ISS. This correlation is more apparent in these orbit types, where successful landing rates appear more consistent.

However, the correlation is less evident in missions targeting GTO (Geostationary Transfer Orbit), where both successful and unsuccessful landings are observed, making it difficult to draw clear distinctions based solely on payload or orbit type. This highlights the complexity of predicting mission outcomes, as various factors likely influence the result.

While each feature - whether it be payload weight, orbit type, or other parameters - may have an impact on the final mission outcome, discerning the precise nature and magnitude of these impacts is not straightforward. The interactions between these features and their influence on the success of a mission are multifaceted, and traditional data visualization techniques may not fully capture the underlying patterns.

To gain a deeper understanding and improve prediction accuracy, machine learning algorithms can be employed. These models can analyze historical data to identify complex relationships between the features and the mission outcome, enabling the development of a predictive framework. By learning from past mission data, we can estimate the likelihood of a successful mission based on the given input features, providing valuable insights for future mission planning.

CONCLUSION

CONCLUSION

In this project, the primary objective is to predict whether the first stage of a Falcon 9 launch will successfully land, which is a critical factor in determining the overall cost-efficiency of the launch. Various features of the Falcon 9 mission, such as payload mass, orbit type, and other mission-specific parameters, may influence the likelihood of a successful landing.

To uncover these relationships, several machine learning algorithms were employed to analyze historical Falcon 9 launch data. These algorithms were used to identify patterns and develop predictive models that can forecast the success or failure of a first-stage landing based on the provided features. By leveraging past mission data, the models aim to provide valuable insights for optimizing launch planning and decision-making.

Among the four machine learning algorithms tested, the decision tree algorithm emerged as the most effective, delivering the highest prediction accuracy. This algorithm's ability to handle both categorical and numerical data, while offering interpretable results, made it particularly well-suited for this task. The performance of the decision tree model underscores the potential of machine learning in improving the precision of mission outcome predictions and enhancing cost estimation for Falcon 9 launches.

THANK YOU!