IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nnadozie Ezenekwe
15/04/2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- This project focused on gathering data on Falcon 9 and Falcon Heavy launches using web scraping techniques with Beautiful Soup and the requests library. The collected data was then cleaned and analysed to determine whether a rocket's first-stage landing was successful—a key factor in predicting launch costs. The workflow involved several steps, including Data Wrangling, Exploratory Data Analysis (EDA) with SQL and Python, and applying Machine Learning models to make landing predictions. Four classification models were tested: Logistic Regression, Support Vector Machine (SVM), Decision Tree Classifier, and K-Nearest Neighbours (KNN). The Decision Tree Classifier exhibited overfitting, while SVM and KNN delivered similar performances. Ultimately, Logistic Regression proved to be the most effective model for predicting successful landings.

Introduction

- Falcon 9 is a two-stage, medium-lift launch vehicle developed by SpaceX in the United States. Built for partial reusability, it helps reduce spaceflight costs while delivering payloads into orbit. The rocket's first stage (booster) provides initial thrust, propelling the vehicle to a set speed and altitude before detaching. Once separated, the second stage takes over, carrying the payload to its designated orbital path. Successfully landing the first stage is essential for cost estimation, as reusable boosters make space missions more economical.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Sending a request to retrieve the Falcon 9 Launch Wikipedia page using its URL
- Extracting column headers from the HTML table to identify key data points
- Constructing a structured data frame to store and analyze the extracted information

Data Collection – SpaceX API

- spacex_url="https://api.spacexdata.com/v4/launches/past"

response = requests.get(spacex_url)

https://github.com/nosezen/Coursera-Test/blob/75e2396fce9a8e54c3dafd4c1d36bb20334e303d/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- BeautifulSoups

  https://github.com/nosezen/Coursera-Test/blob/ccb543c4f54eb7a069f985c9257ddeba3cc16d80/jupyter-labs-webscraping.ipynb

Data Collection - Scraping

- A new column, "Class", was created using data from the "Outcome" column, assigning 1 for successful landings and 0 for failures.
- Defined landing categories, grouping outcomes into successful and unsuccessful sets.
- Classified landings by mapping data from "Outcome" to the "Class" column based on predefined success criteria.

https://github.com/nosezen/Coursera-Test/blob/f0b9ed1fddca91746d2c32dd72fbb61c32791d8c/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with SQL

- Retrieve a list of unique launch sites used in space missions.
- Show five records where launch sites start with "CCA".
- Calculate the total payload mass transported by boosters launched under NASA (CRS) missions.
- Determine the average payload mass carried by F9 v1.1 booster version.
- Identify the date of the first successful landing on a ground pad.
- Find the boosters that successfully landed on a drone ship with a payload mass between 4000 and 6000 kg.
- Count the number of successful and failed mission outcomes.
- List all booster versions that have carried the maximum payload mass.
- Extract records showing month names, failed landings on drone ships, booster versions, and launch sites for missions in 2015.Rank landing outcomes—like "Failure (drone ship)" and "Success (ground pad)"—from June 4, 2010, to March 20, 2017, in descending order.

Build a dashboard with Folium

- Visualizing All Launch Sites

Placed a marker at NASA Johnson Space Center, using its latitude and longitude coordinates as the starting location.

Mapped all launch sites with circle markers, popup labels, and text labels, showcasing their geographic positioning relative to the Equator and nearby coastlines.

- Highlighting Launch Outcomes

Used colored markers to indicate launch success (green) and failure (red).

Implemented a Marker Cluster to visually assess which sites have the highest success rates.

- Mapping Distances to Key Locations

- Added colored lines to represent the distance between the KSC LC-39A launch site and surrounding landmarks, such as railways, highways, the coastline, and the nearest city.

Build a dashboard with Plotly Dash

- Launch Site Selection

Implemented a dropdown list allowing users to select a specific launch site for analysis.

- Visualizing Launch Success Rates

Designed a pie chart to display total successful launches across all sites.

When a specific site is chosen, the chart updates to show success vs. failure rates for that location.

- Payload Mass Filtering

Integrated a slider to adjust the payload mass range, enabling users to analyze launches based on payload weight.

- Exploring the Relationship Between Payload and Success

Developed a scatter plot illustrating the correlation between payload mass and launch success, categorized by different booster versions

Predictive Analysis

- Data Preparation & Transformation

Extracted the "Class" column from the dataset and converted it into a NumPy array for efficient analysis.

Standardized the dataset using StandardScaler, ensuring consistent scaling across features before fitting and transforming the data.

- Model Selection & Evaluation

Tested multiple classification models by measuring their Jaccard Score and F1 Score, identifying the most effective approach.

Analyzed each model’s confusion matrix to assess misclassifications and overall performance.

- Training & Performance Assessment

Split the dataset into training and testing sets using train_test_split to validate model accuracy.

Measured each model’s prediction performance by computing test accuracy using the .score() method.

- Hyperparameter Optimization

Leveraged GridSearchCV with 10-fold cross-validation to fine-tune model parameters.

Applied GridSearchCV on Logistic Regression, Support Vector Machines (SVM), Decision Tree, and K-Nearest Neighbors (KNN), selecting the best-performing configurations.

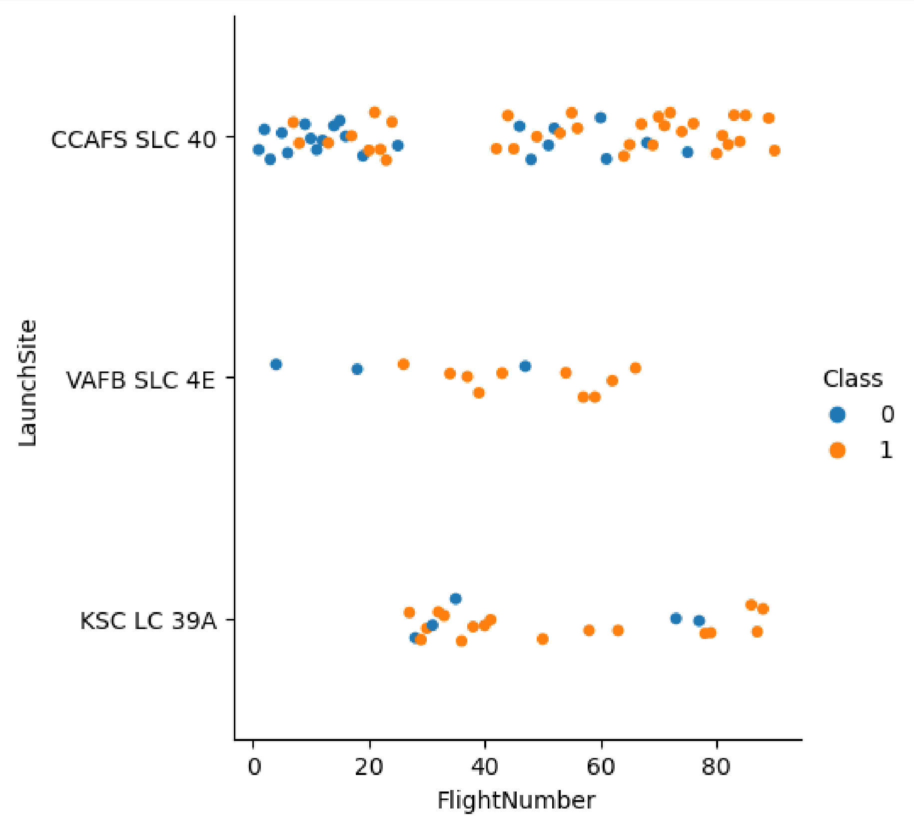
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

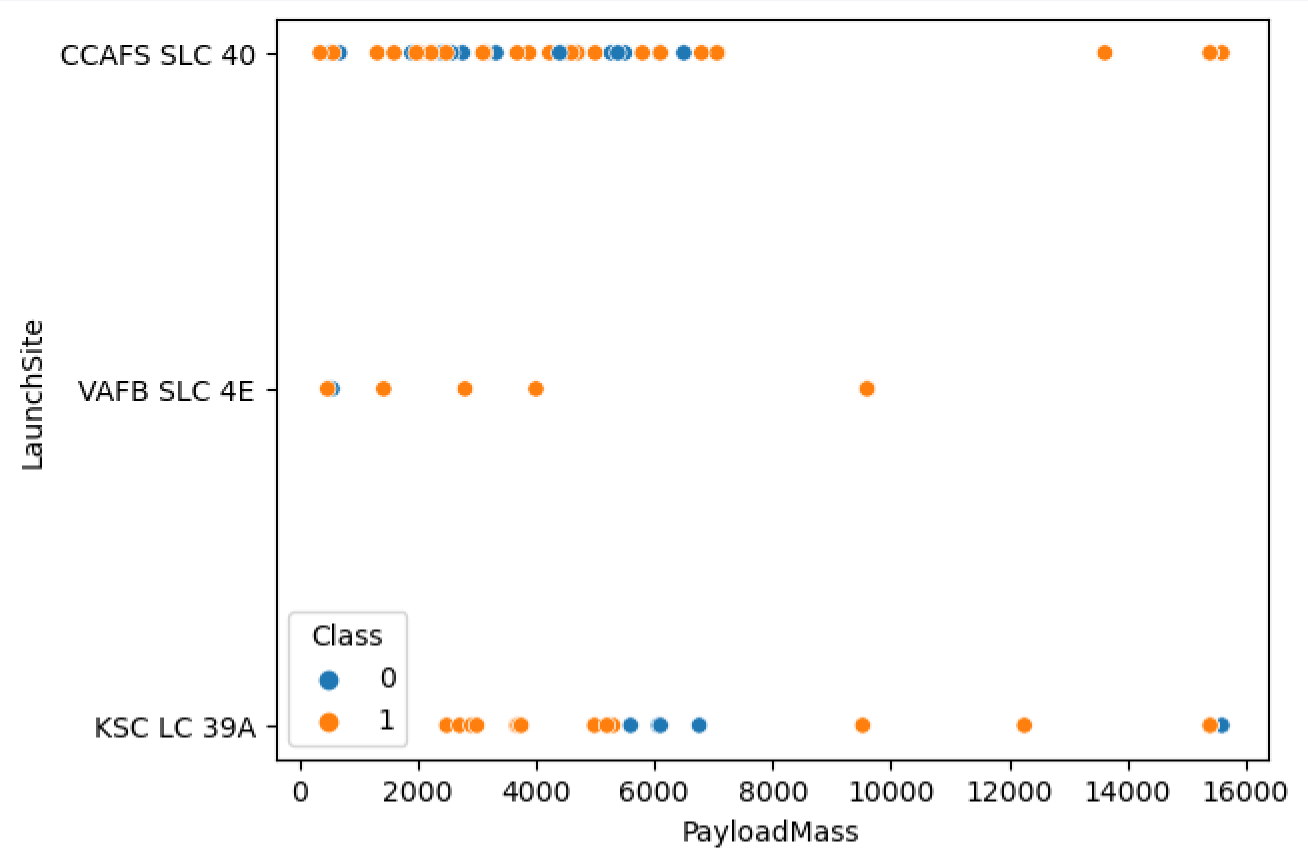
Section 2

Insights drawn from EDA

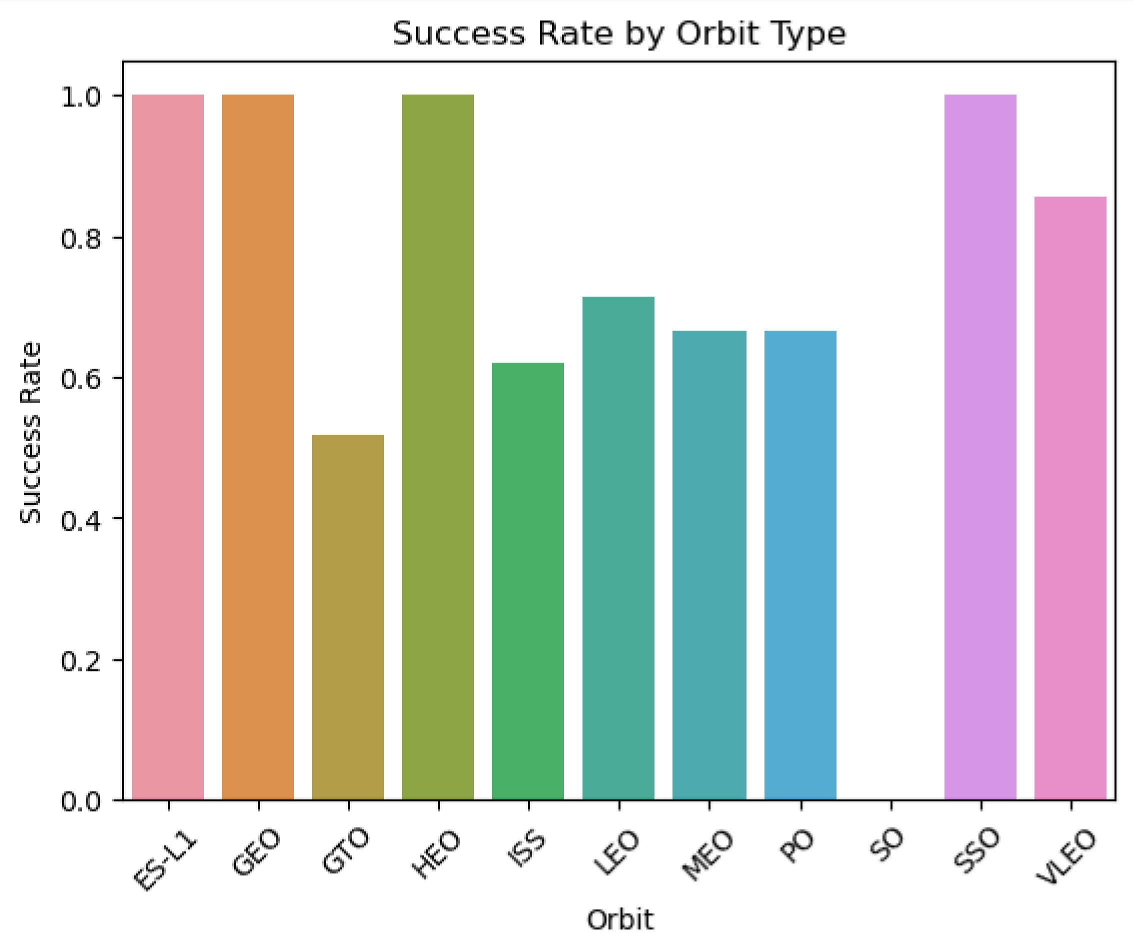
Flight Number vs. Launch Site



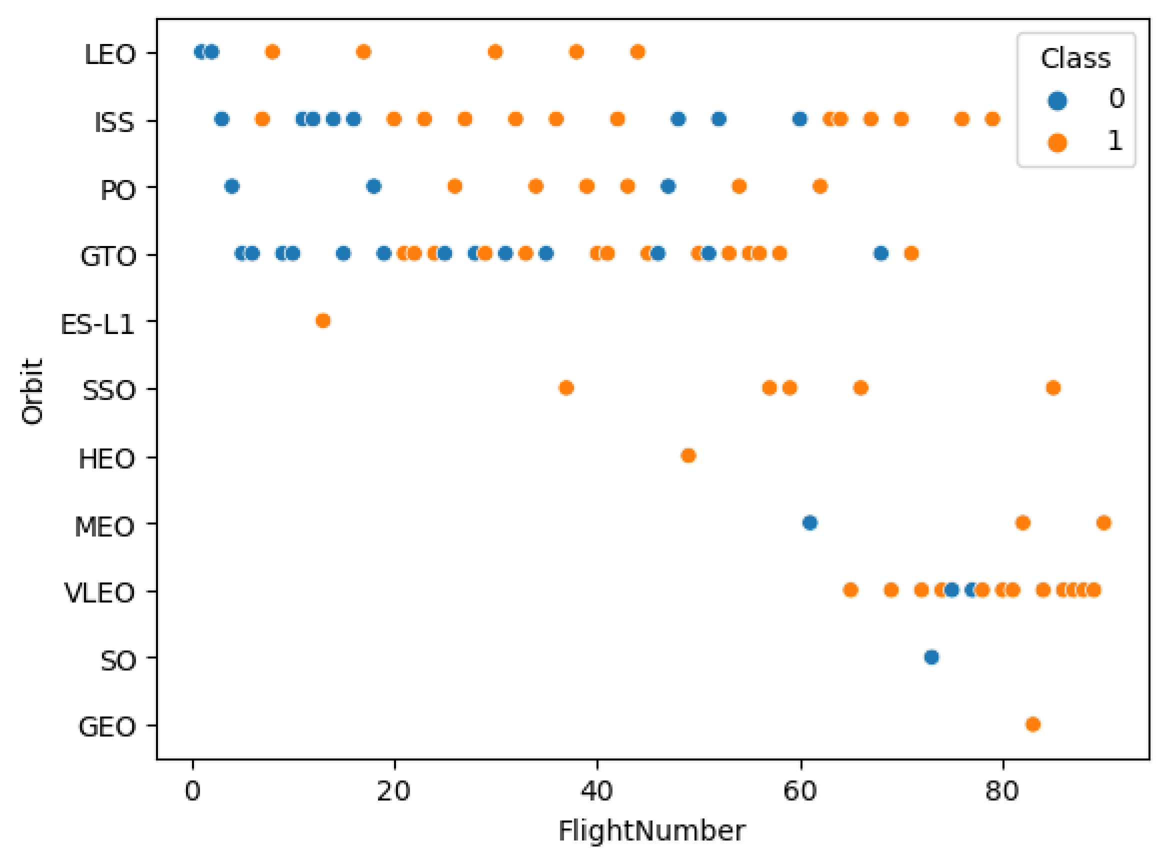
Payload vs. Launch Site



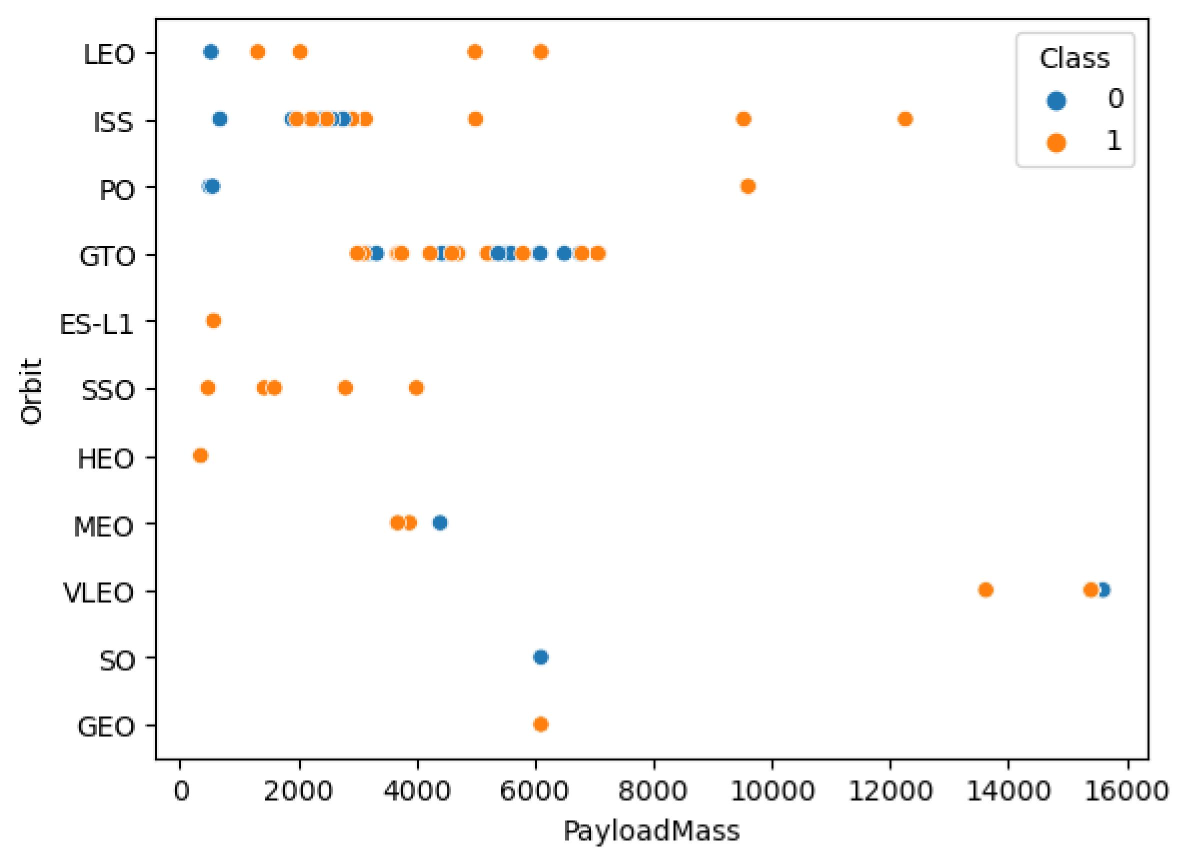
Success Rate vs. Orbit Type



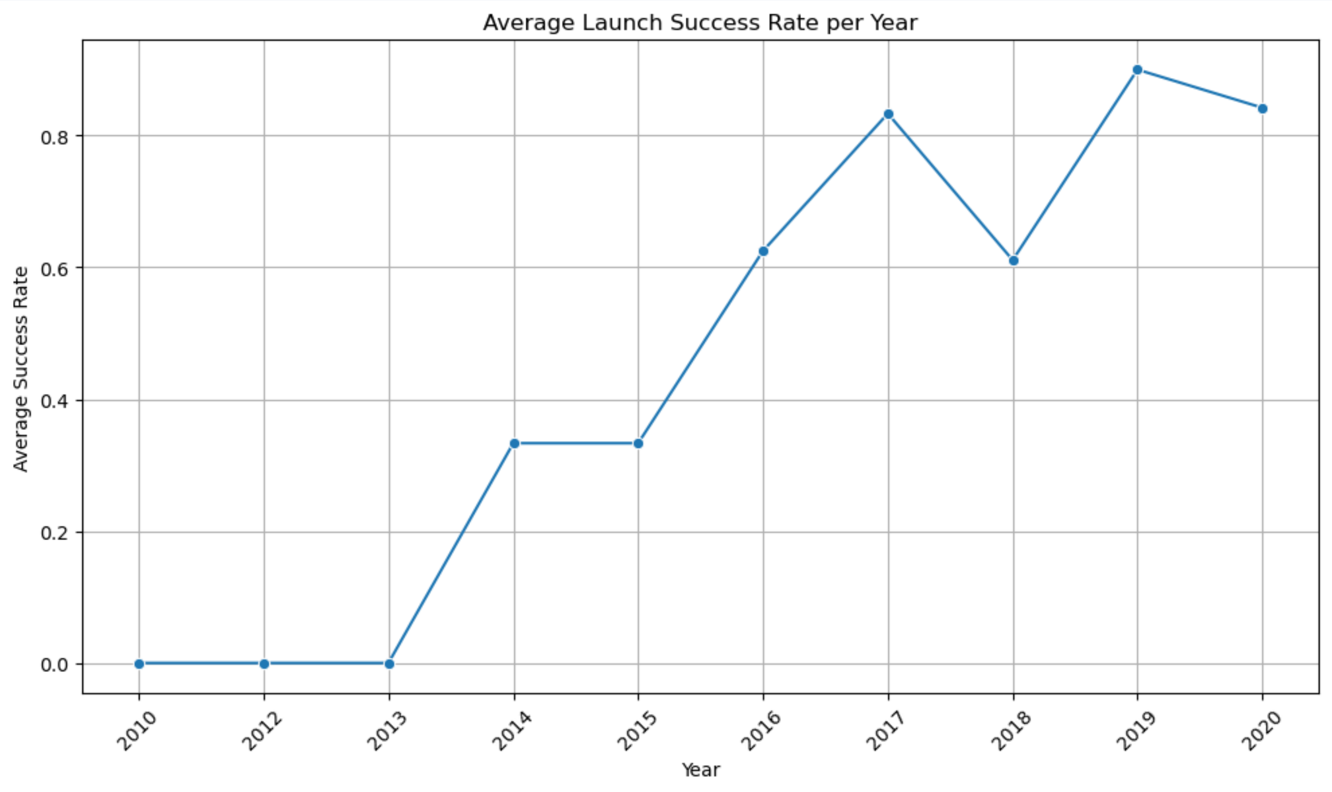
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
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	Failure (parachute)
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

Total Payload Mass

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

MIN(Date)
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

Payload
AsiaSat 8
AsiaSat 6
ABS-3A Eutelsat 115 West B
Turkmen 52 / MonacoSAT
SES-9
JCSAT-14
JCSAT-16
EchoStar 23
SES-10
NROL-76
Boeing X-37B OTV-5
SES-11 / EchoStar 105
Zuma
GovSat-1 / SES-16
SES-12
Merah Putih
Es hail 2
SSO-A
GPS III-01
Nusantara Satu, Beresheet Moon lander, S5
RADARSAT Constellation, SpaceX CRS-18
GPS III-03, ANASIS-II
ANASIS-II, Starlink 9 v1.0
GPS III-04 , Crew-1

Total Number of Successful and Failure Mission Outcomes

Successful_Missions
99

Failed_Mission
1

Boosters Carried Maximum Payload

Booster_Version
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

2015 Launch Records

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

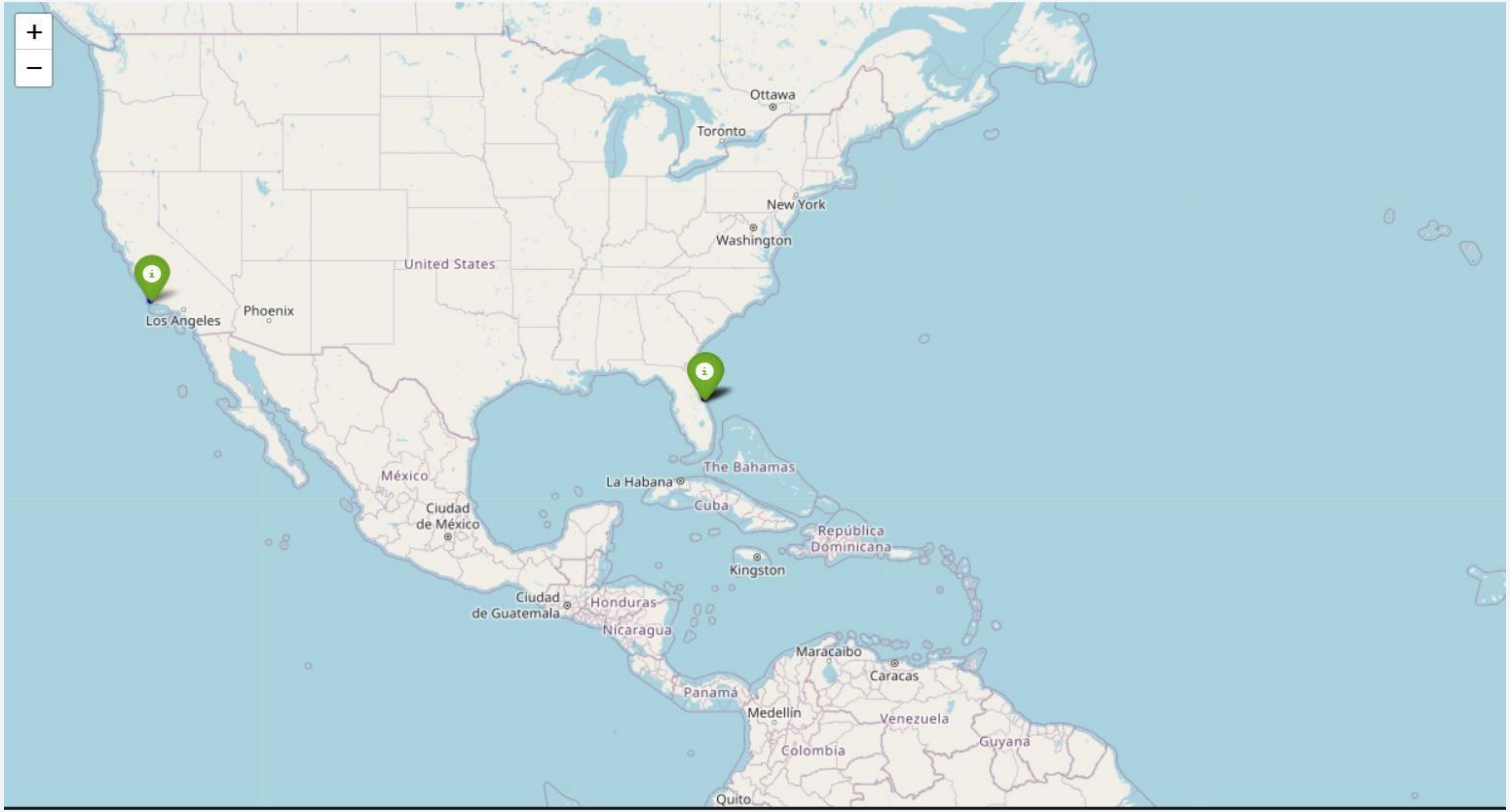
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

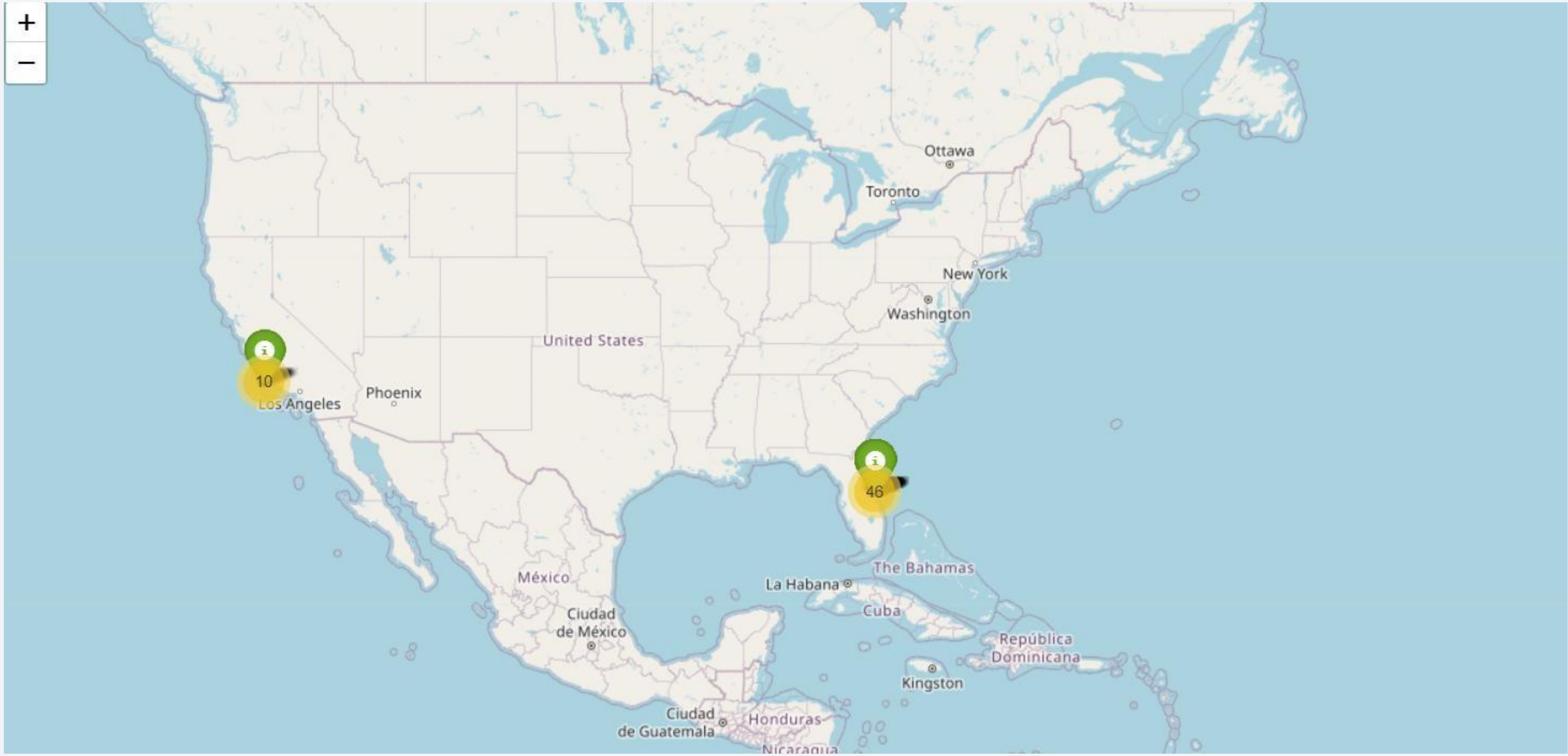
Section 3

Launch Sites Proximities Analysis

Launch


Ottawa
Toronto
New York
Washington
United States
Phoenix
Los Angeles
The Bahamas
La Habana
Cuba
México
Ciudad de México
República Dominicana
Kingston
Ciudad de Guatemala
Honduras
Nicaragua
Maracaibo
Caracas
Panamá
Medellín
Venezuela
Guyana
Colombia
Quito

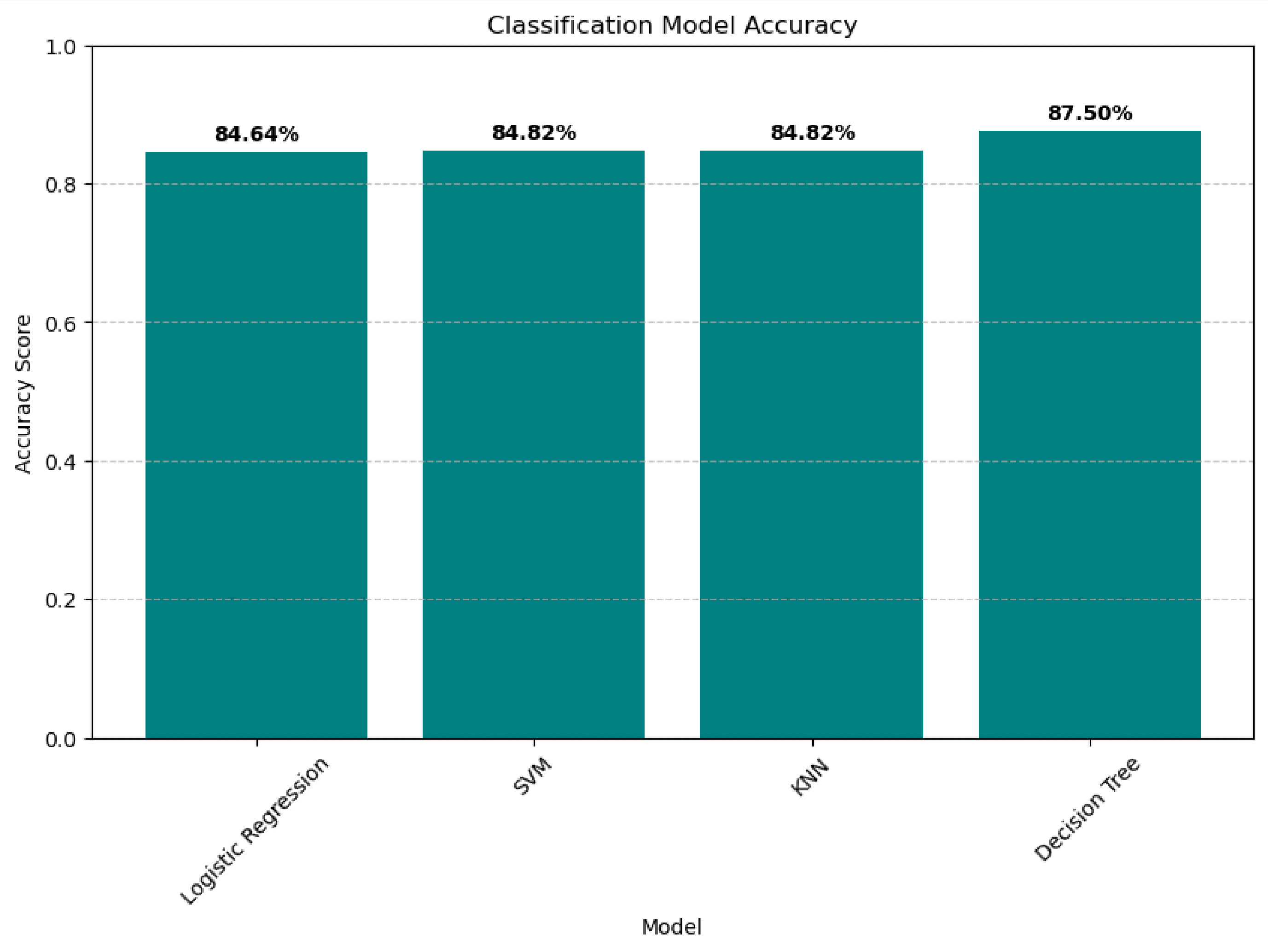
Total missions



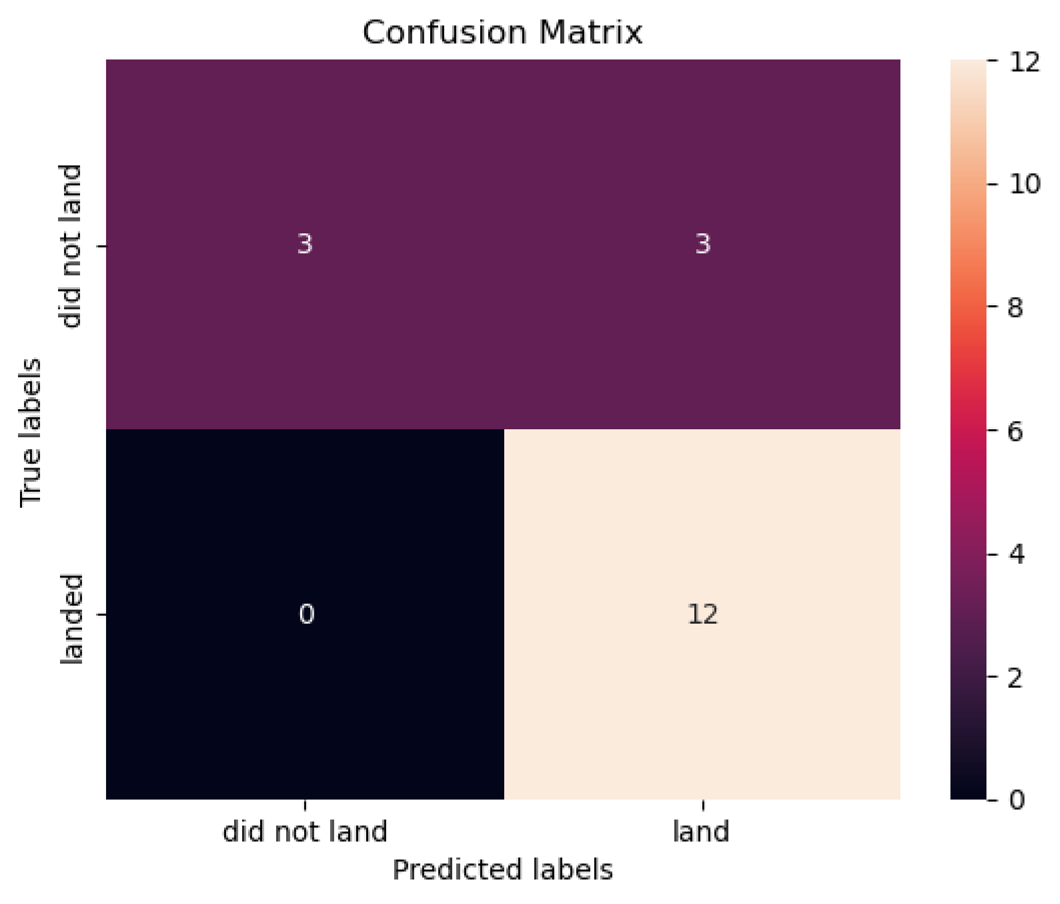
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- The Logistic Regression model delivered the strongest performance in predicting outcomes. Additionally, the Falcon 9 rocket has achieved more successful missions than failed ones.

Thank you!