IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rex Coleman
April 10, 2024

Outline

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

Executive Summary – Predictive Analysis of SpaceX Falcon 9 First Stage Landings

- Background: SpaceX has revolutionized space travel with its Falcon 9 rocket, significantly reducing launch costs through the reusability of its first stage. Accurately predicting these landings is crucial for financial planning and competitive bidding in the space launch market.

- Objective: This project aimed to leverage machine learning techniques to predict the outcome of Falcon 9 first stage landings. By analyzing extensive SpaceX launch data, we sought to identify patterns and factors that influence successful landings.

- Methodology: Utilizing data obtained from SpaceX's API and web scraping, we conducted a thorough data preparation process, including wrangling and cleaning. Through exploratory data analysis, we uncovered key variables impacting landing success. Various classification models were then built, tuned, and evaluated to find the most accurate predictor of landing outcomes.

- Findings: Our analysis revealed significant predictors of landing success, including payload mass, orbit type, and booster version. The best-performing model achieved a high level of accuracy, demonstrating the feasibility of using machine learning for this purpose.

- Implications: These insights can significantly impact the cost-effectiveness of space missions and inform strategic decisions in competitive bidding scenarios. Furthermore, our research contributes to the broader understanding of reusable rocket technology's economic viability.

- Conclusion: Predicting Falcon 9 first stage landings with machine learning offers a promising avenue for optimizing space mission costs and enhancing SpaceX's competitive edge. This study not only underscores the potential of reusability in the space industry but also sets the stage for future innovations in predictive analytics for space travel.

Introduction

- Background:
  - SpaceX's Falcon 9, offered at $62 million per launch, significantly undercuts competitors due to its reusable first stage. Predicting successful landings is pivotal for assessing launch costs and market competition.
- Objective:
  - This study employs machine learning to predict Falcon 9's first-stage landing outcomes. Through comprehensive data analysis—from collection and wrangling to exploratory insights and visualizations—this report aims to uncover what determines landing success.
- Goal:
  - To provide actionable insights for competitive strategies in the space launch industry, enhancing our understanding of reusable rocket technology's economics.

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

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Tools used: Python's Requests, Pandas, NumPy.

- https://github.com/rexcoleman/IBMDataScienceProfessionalCertificate/blob/main/Course-10-AppliedDataScienceCapstone-1-DataCollection-SpaceXAPI.ipynb

STEP 1

Make GET requests to SpaceX API.

STEP 2

Parce data to extract key data points (booster versions, payload mass, launch outcomes, etc.).

Data Collection - Scraping

- Tools used: Python's Requests, Pandas, Beautiful Soup.

- https://github.com/rexcoleman/IBMDataScienceProfessionalCertificate/blob/main/Course-10-AppliedDataScienceCapstone-2-DataCollection-Webscraping.ipynb

STEP 1

Request the Falcon9 Launch Wiki page from its URL

STEP 2

Extract all column/variable names from the HTML table header

STEP 3

Create a data frame by parsing the launch HTML tables

Data Collection - Scraping

- Tools used: Python's Requests, Pandas, Beautiful Soup.

- https://github.com/rexcoleman/IBMDataScienceProfessionalCertificate/blob/main/Course-10-AppliedDataScienceCapstone-2-DataCollection-Webscraping.ipynb

STEP 1

Requested the Falcon9 Launch Wiki page from its URL

STEP 2

Extracted all column/variable names from the HTML table header

STEP 3

Created a data frame by parsing the launch HTML tables

Data Wrangling

STEP 1

Calculated the number of launches on each site.

STEP 2

Calculated the number and occurrence of each orbit.

STEP 3

Calculated the number and occurrence of mission outcome of the orbits.

STEP 4

Create a landing outcome label from Outcome column.

- https://github.com/rexcoleman/IBMDataScienceProfessionalCertificate/blob/main/Course-10-AppliedDataScienceCapstone-3-DataWrangling.ipynb

EDA with SQL

- The Following SQL Queries Were Performed to Better Understand the Dataset as it relates to Falcon 9 Stage 1 landing success.
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
  - Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.
- https://github.com/rexcoleman/IBMDataScienceProfessionalCertificate/blob/main/Course-10-AppliedDataScienceCapstone-4-ExploratoryDataAnalysis-SQL.ipynb

EDA with Data Visualization

- The following charts were plotted with to understand which variables most closely correlate with Falcon 9 Stage 1 landing success.
  - Flight Number vs. Launch Site vs. Landing Success (Scatter Plot)
  - Payload Mass vs. Launch Site vs. Landing Success (Scatter Plot)
  - Orbit Type vs. Landing Success (Bar Plot)
  - Flight Number vs. Orbit Type vs. Landing Success (Scatter Plot)
  - Payload Mass vs. Orbit vs. Landing Success (Scatter Plot)
  - Launch Success Rate Over Time (Line Plot)
- https://github.com/rexcoleman/IBMDataScienceProfessionalCertificate/blob/main/Course-10-AppliedDataScienceCapstone-5-ExploratoryDataAnalysis-Visualization.ipynb

Build an Interactive Map with Folium

- Analyzed an interactive map with Launch Site locations along with proximity to important economic features: highways, coastline, cities, airports, rail. Proximity to cities brings talent, proximity to impactcs cost.
- https://github.com/rexcoleman/IBMDataScienceProfessionalCertificate/blob/main/Course-10-AppliedDataScienceCapstone-6-InteractiveMapWithFolium.ipynb

Build a Dashboard with Plotly Dash

- Created an Interactive Dashboard With Plotly Dash
  - Total Success by Launch Site (Pie Chart)
  - Payload Mass vs. Booster Version vs. Landing Success (Scatter
- These plots were selected to show factors that correlate with Falcon 9 Stage 1 landing success.
- https://github.com/rexcoleman/IBMDataScienceProfessionalCertificate/blob/main/Course-10-AppliedDataScienceCapstone-7-DashboardWithPlotlyDash.py

Predictive Analysis (Classification)

Explore Models

Evaluated Logistic Regression, SVM, Decision Tree and KNN for predicting landing success.

Tune Hyperparameters

Utilized GridSearchCV for optimal model performance.

Select Best Model

Chose best model based off accuracy and derived insights.

https://github.com/rexcoleman/IBMDataScienceProfessionalCertificate/blob/main/Course-10-AppliedDataScienceCapstone-8-MachineLearningPrediction.ipynb

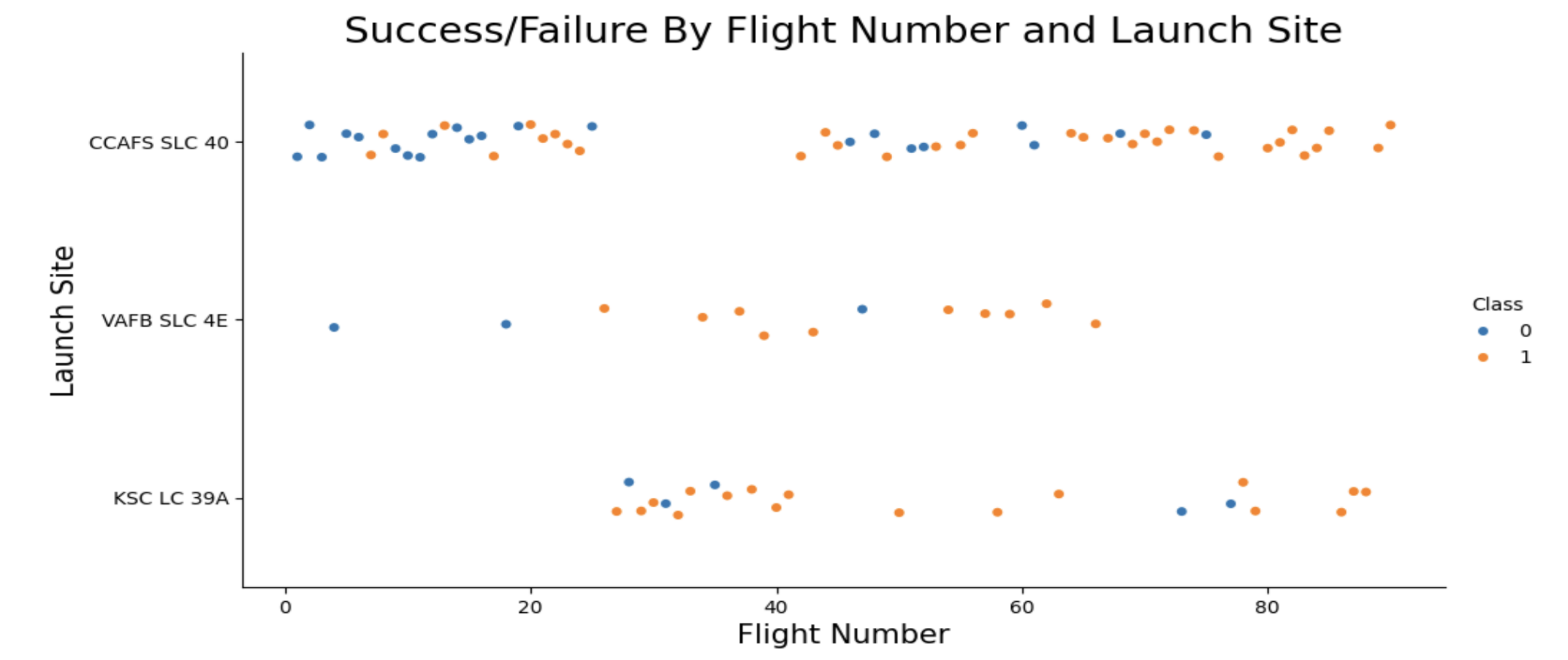
Results

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

Section 2

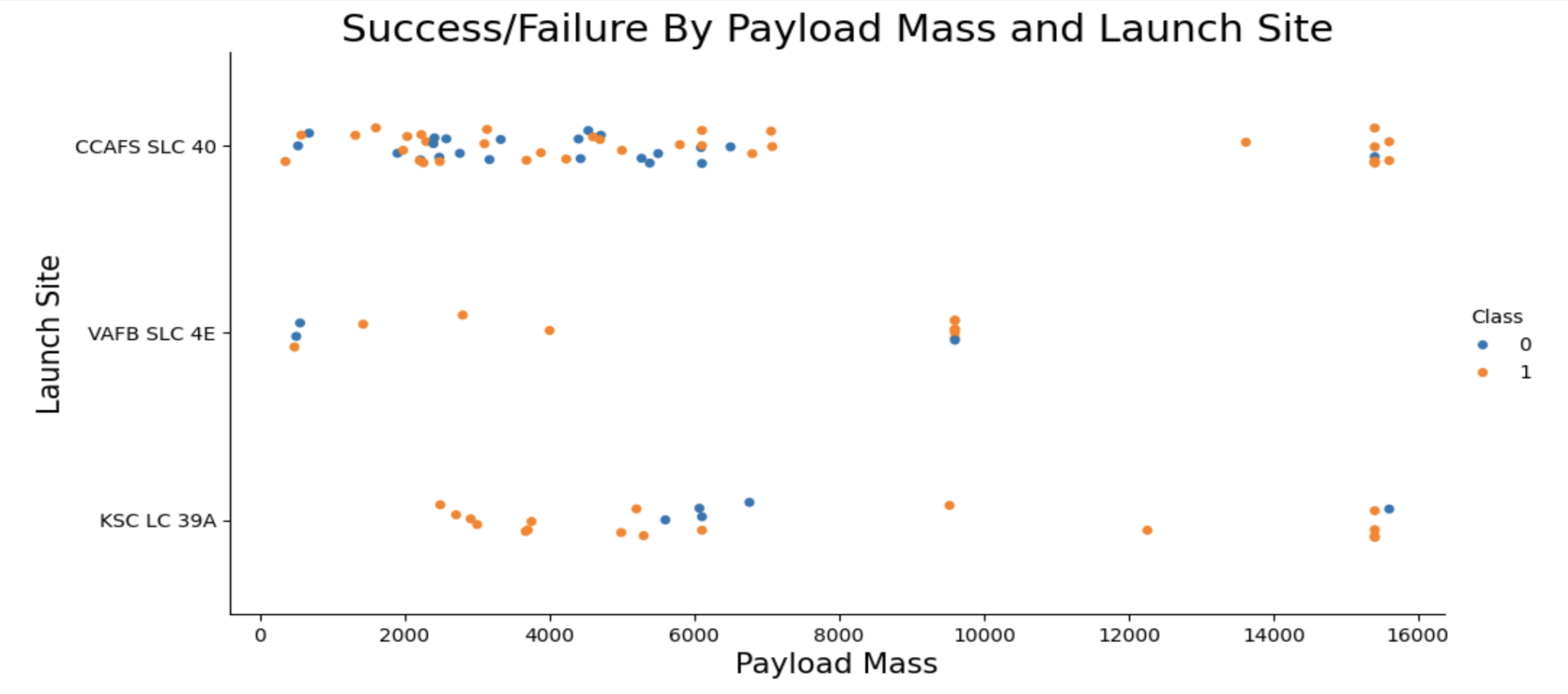
Insights drawn from EDA

Flight Number vs. Launch Site



With experience SpaceX is improving its success rate (Class = 1, Color Orange) although this trend appears to be more pronounced for specific sites (Cape Canaveral and Vandenberg AFB vs. Kennedy Space Center).

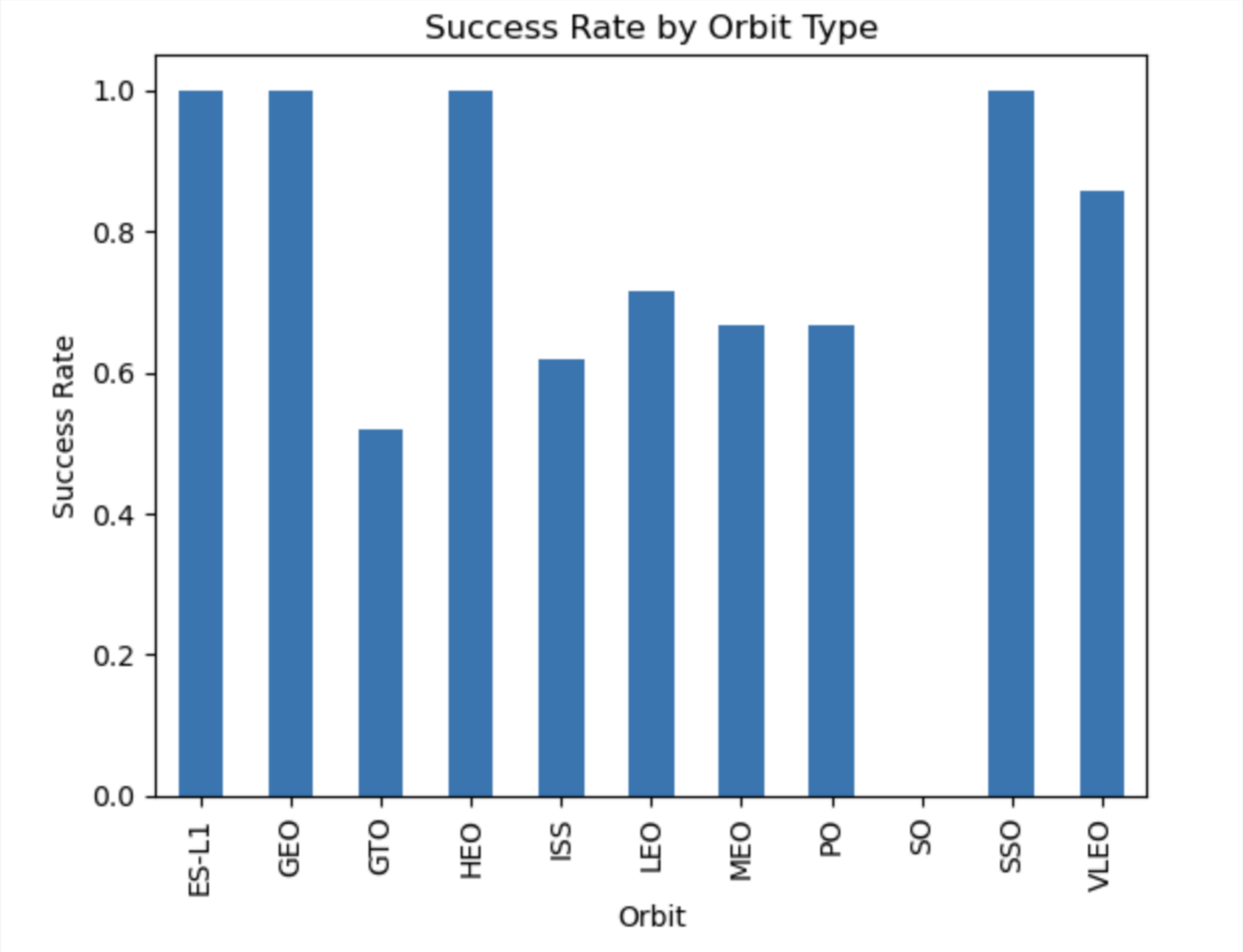
Payload vs. Launch Site



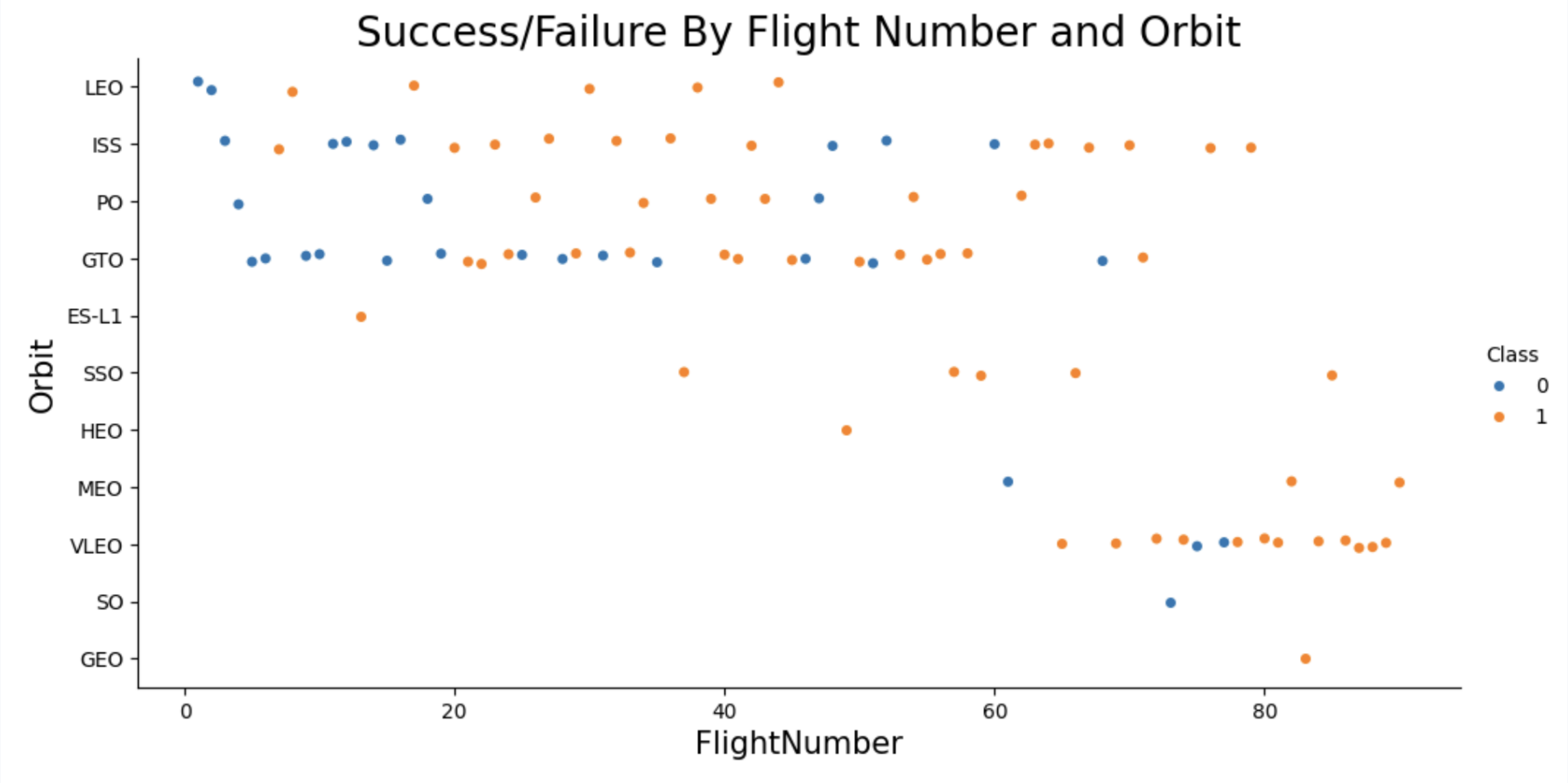
There are no rockets launched for heavypayload mass(greater than 10000) at the VAFB SLC launch site.

Success Rate vs. Orbit Type

- The following orbits have the highest success rates: ES-L1, GEO, HEO, SSO.

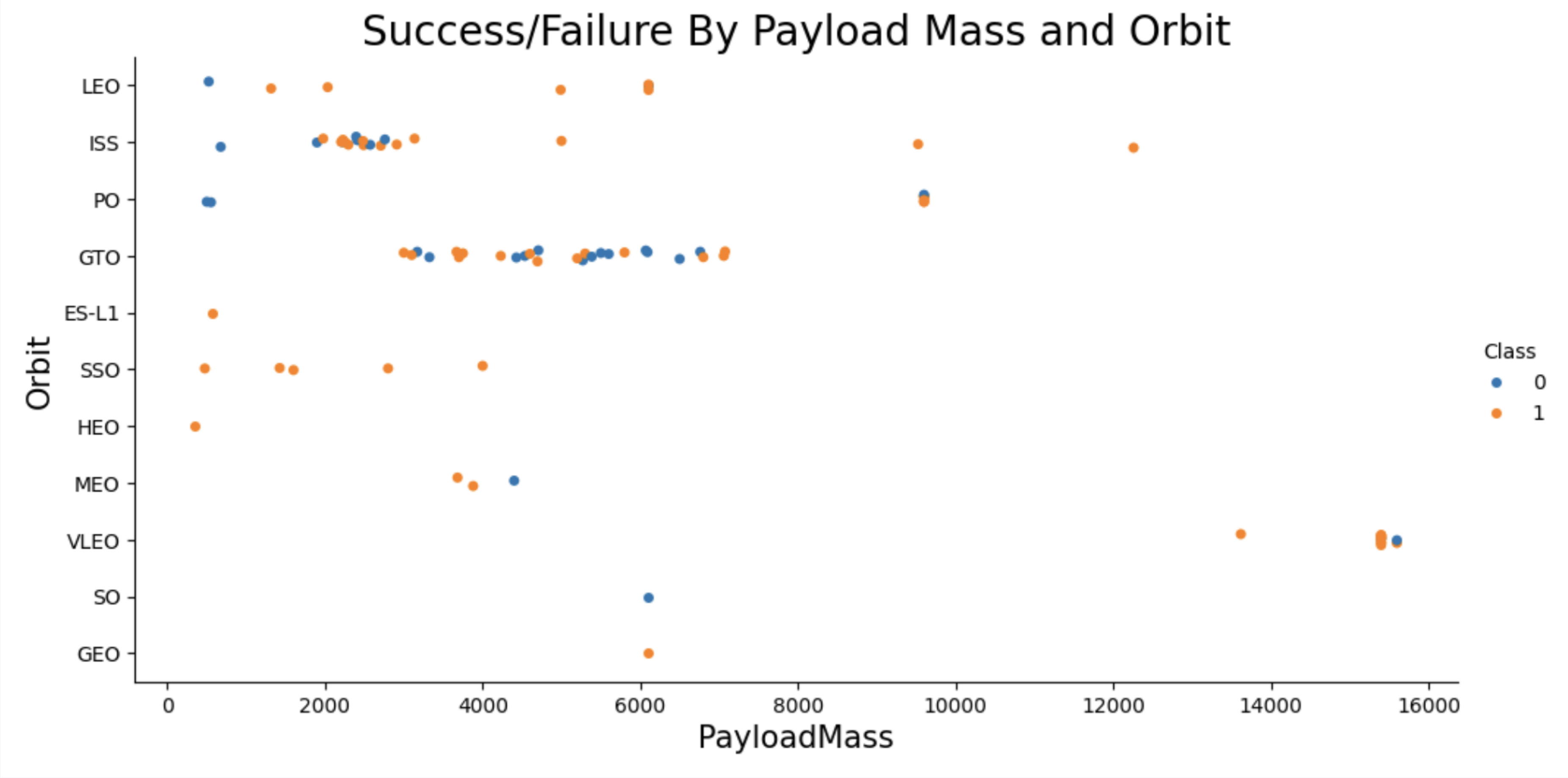


Flight Number vs. Orbit Type



Success rates appear to improve with experience for LEO orbit but not for GTO orbit.

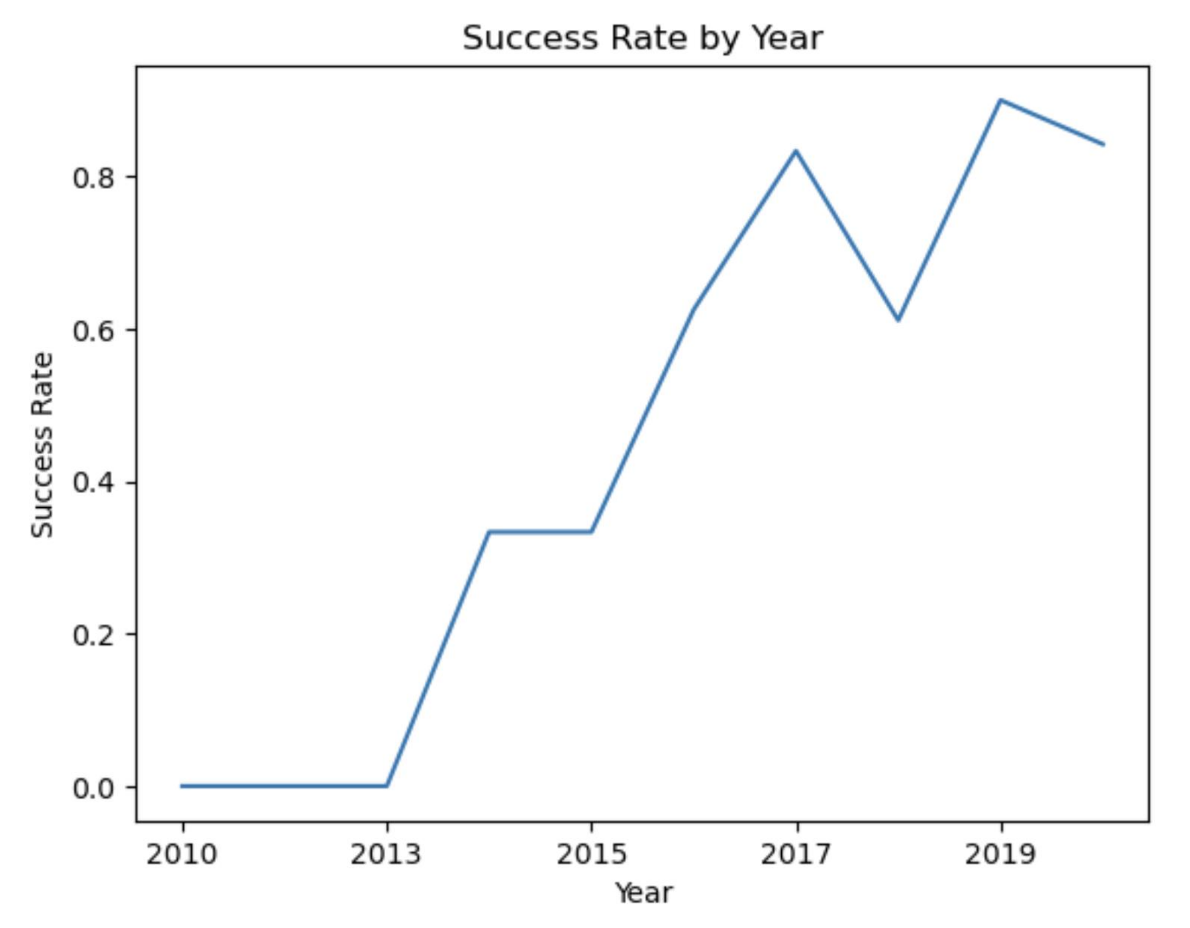
Payload vs. Orbit Type



SpaceX seems to have higher success rates for higher Payload Mass launches on Polar, LEO and ISS Orbits.

Launch Success Yearly Trend

- SpaceX is driving down launch costs with its experience curve.



All Launch Site Names

- SpaceX Launch Sites: Cape Canaveral x 2 (Florida), Vandenberg AFB (California), Kennedy Space Center (Florida).
- %sql SELECT DISTINCT Launch_Site FROM SPACEXTBL;

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

- Find 5 records where launch sites begin with `CCA`.
- %sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;

Total Payload Mass

- Total payload carried by boosters with NASA as the customer.
- %sql SELECT SUM(PAYLOAD_MASS__KG_) as total_mass FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';

total_mass
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1.
- %sql SELECT AVG(PAYLOAD_MASS__KG_) as average_mass FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1';

average_mass
2928.4

First Successful Ground Landing Date

- Dates of the first successful landing outcome on a ground pad.
- %sql SELECT MIN(Date) FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)';

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4,000 KG but less than 6,000 KG.

- %sql SELECT Booster_Version FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes.
- %sql SELECT COUNT(*) AS Success_Count FROM SPACEXTBL WHERE Mission_Outcome LIKE 'Success%' UNION ALL SELECT COUNT(*) AS Failure_Count FROM SPACEXTBL WHERE Mission_Outcome = 'Failure (in flight)';

Success_Count
100
1

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass.
- %sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

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

- 2015 failed drone ship landing outcomes including booster version and launch site name.
- %sql SELECT Booster_Version, Launch_Site, substr(Date, 6, 2) AS Month, substr(Date, 1, 4) AS Year FROM SPACEXTBL WHERE Landing_Outcome = 'Failure (drone ship)' AND substr(Date, 0, 5) = '2015';

Booster_Version	Launch_Site	Month	Year
F9 v1.1 B1012	CCAFS LC-40	01	2015
F9 v1.1 B1015	CCAFS LC-40	04	2015

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.
- %sql SELECT Landing_Outcome, COUNT(Landing_Outcome) \
- AS Outcome_Count \
- FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' \
- GROUP BY Landing_Outcome \
- ORDER BY Outcome_Count DESC;

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

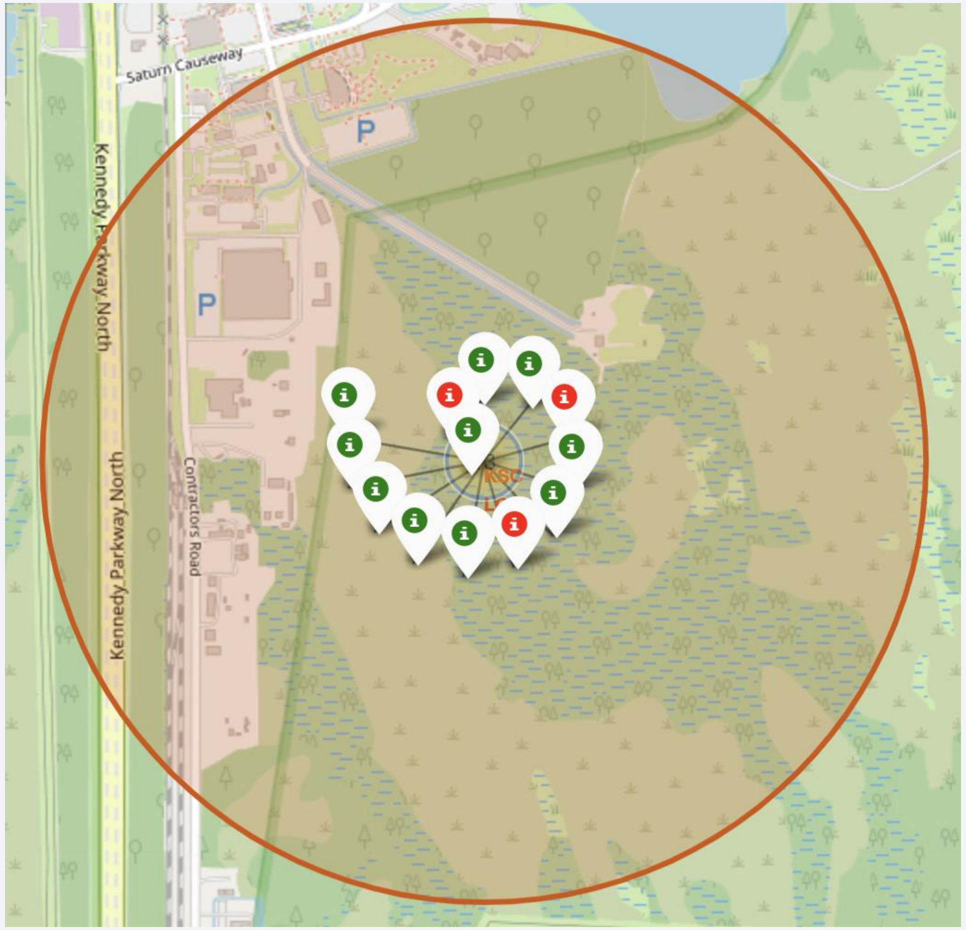
Launch Sites Proximities Analysis

Interactive Map of SpaceX Launch Sites

- SpaceX launch sites. Note: i) proximity to equator (capitalize on Earth's rotational orbit for reduced fuel cost to achieve orbit) and ii) proximity to coast (reduced risk of injury to person or property).

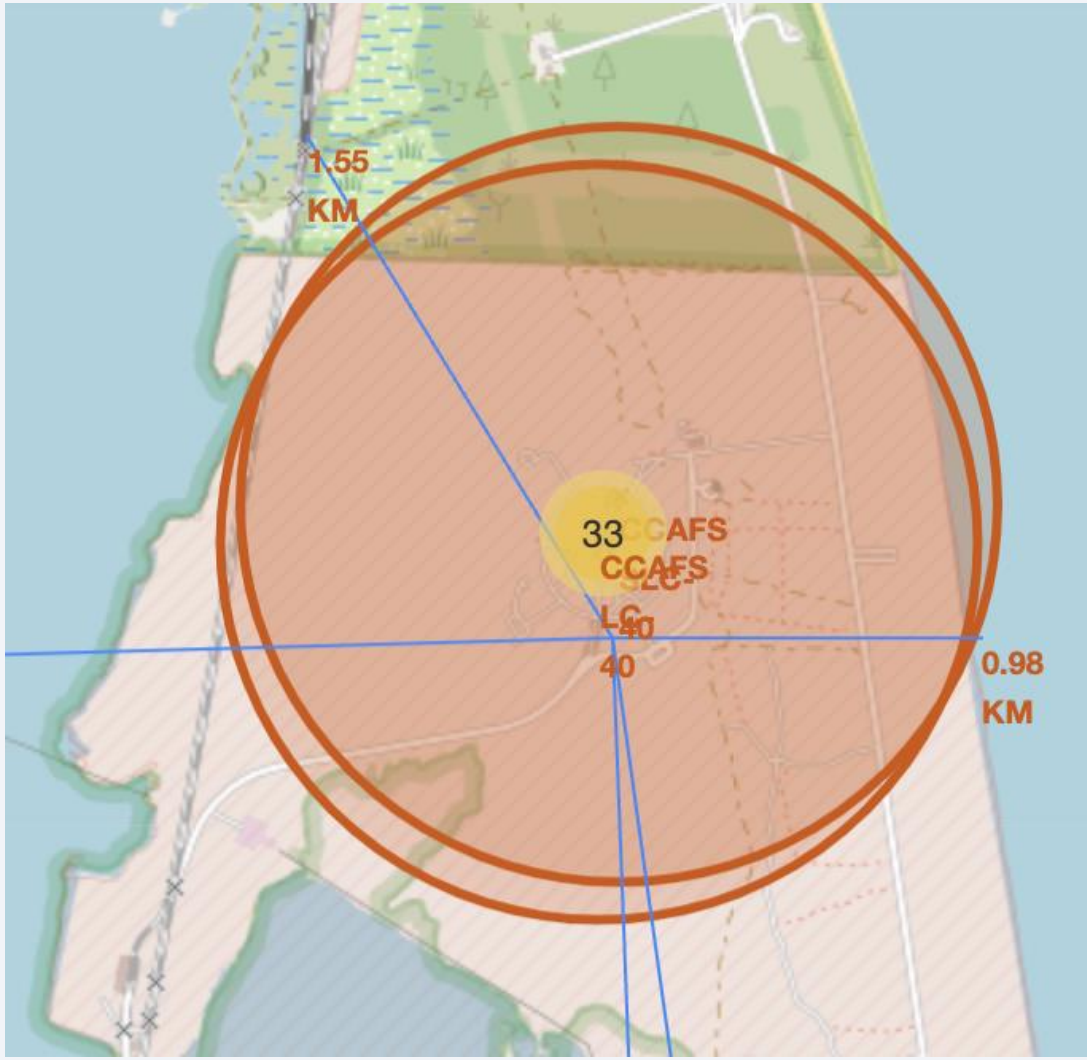
Interactive Map Showing Successes/Failures

- Kennedy Space Center successes (Green) and failures (Red) shown on interactive folium map.



Interactive Map Showing Successes/Failures

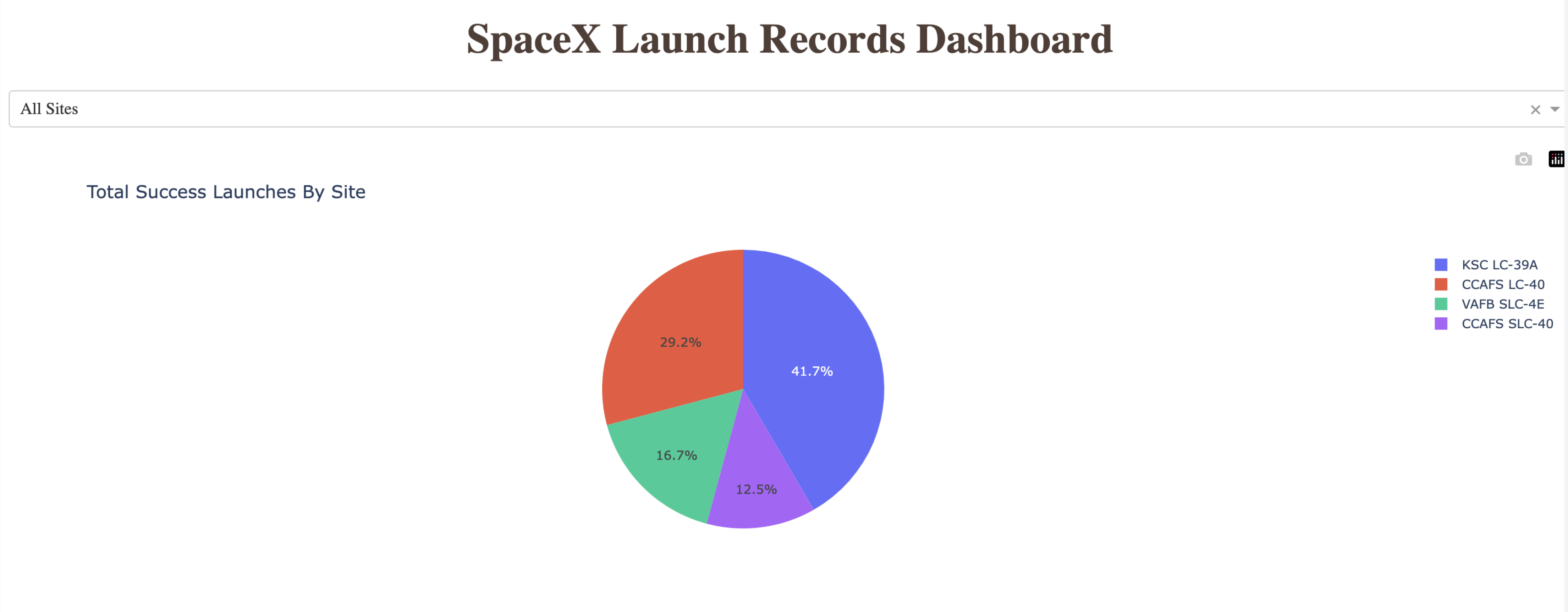
- Interactive folium map shows launch site proximity to important economic features to reduce costs (cities for talent, coast to reduce fallen debris risk, rail way, highway and airport for logistics).



Section 4

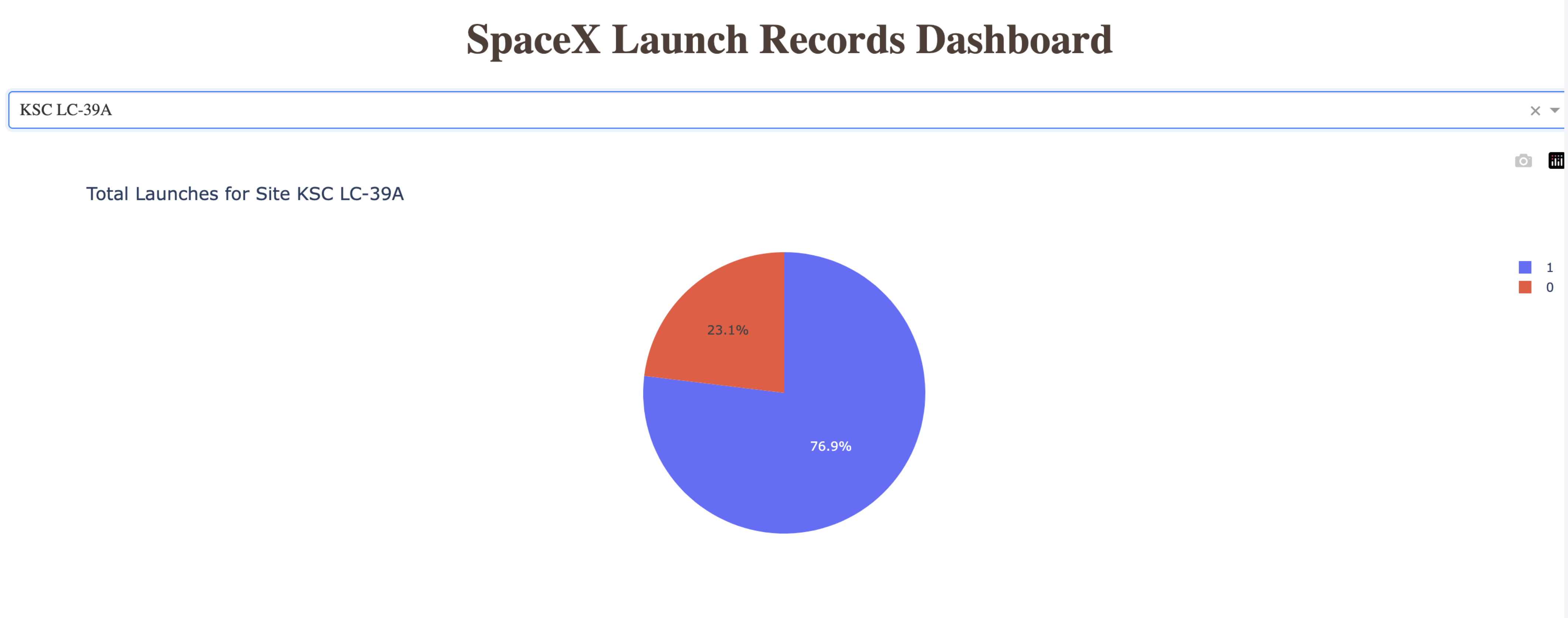
Build a Dashboard with Plotly Dash

Interactive Dashboard w/ Plotly and Dash



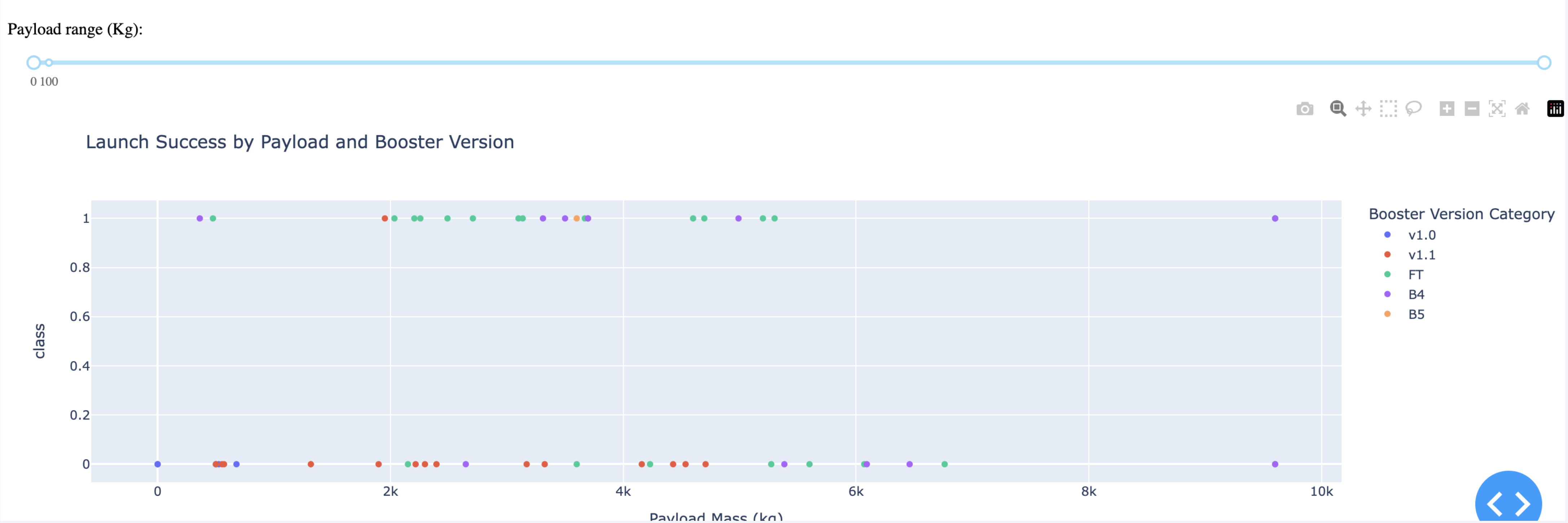
Total successful launches by site. Kennedy Space Center is the winner.

Interactive Dashboard w/ Plotly and Dash



Kennedy Space Center is the winner for the highest success rate.

Interactive Dashboard w/ Plotly and Dash

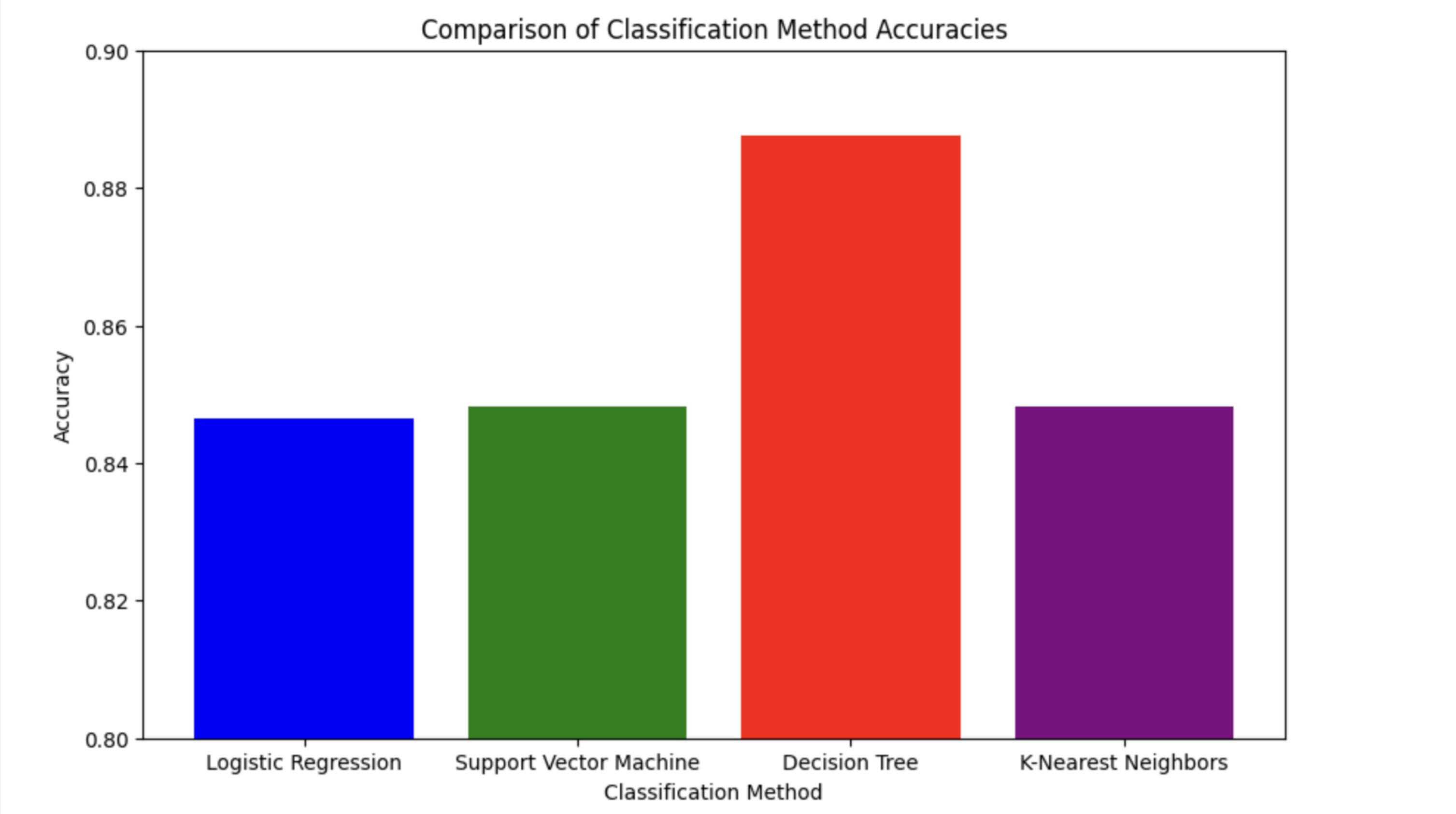


The FT booster appears to have the highest success rate and the v1.1 booster appears to have the lowest success rate.

Section 5

Predictive Analysis (Classification)

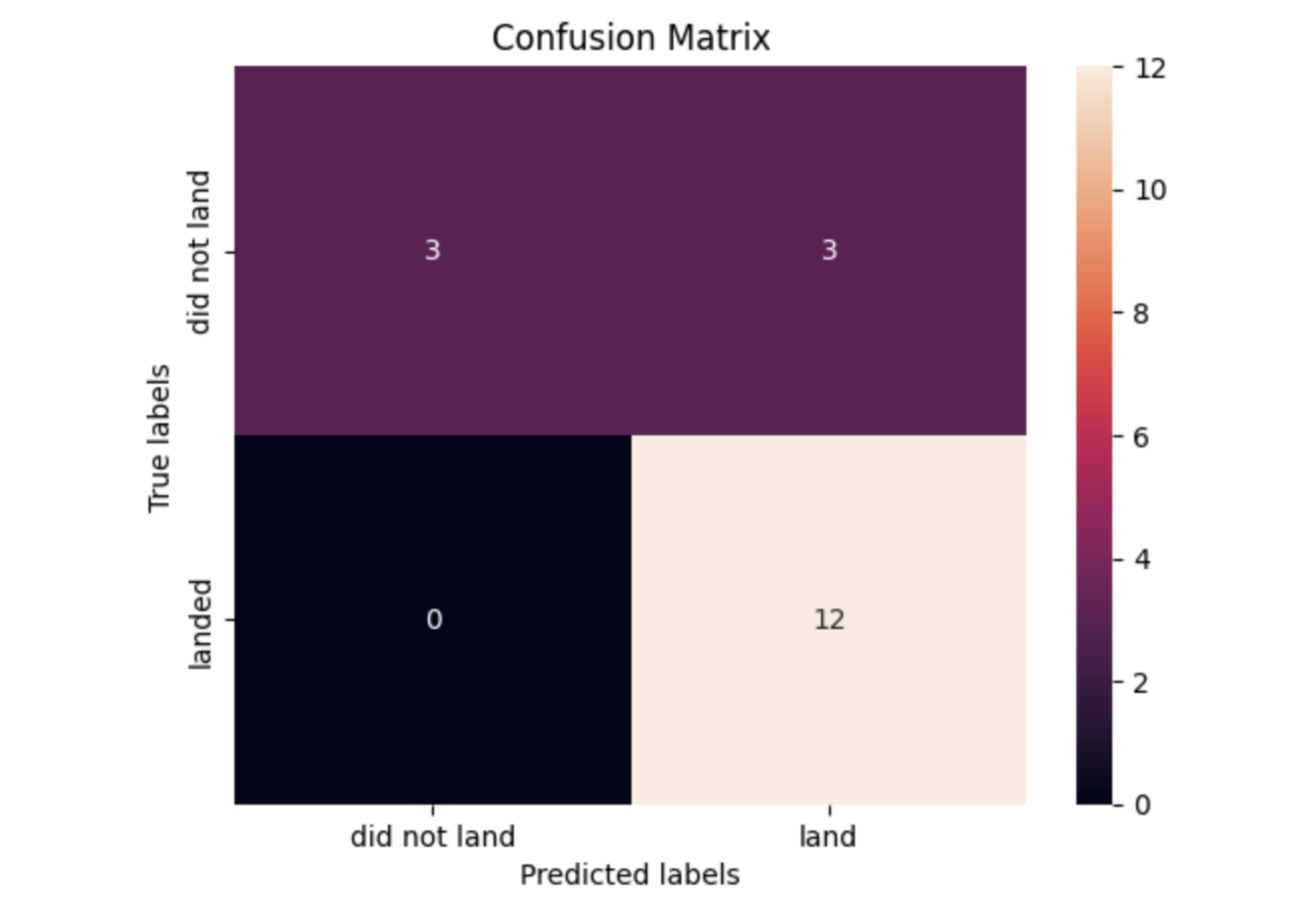
Classification Accuracy



The Decision Tree Model appears to have the highest accuracy.

Decision Tree Confusion Matrix

- A "perfect" model would have all values in the top left and bottom right corners. This model demonstrates false positives (top right) where it predicts 50% (3) landings to be successes while the actual (True Positive) results were 6 (sum of top row).



Conclusions

- Point 1: SpaceX leverages it's experience curve for competitive advantage. This is a significant barrier to entry.
- New entrants may wish to consider a market entry that targets a specific market (that is off SpaceX's radar) and offer with a specific value prop around launch site, orbit and payload mass. Should they find a successful model they can leverage the wins to build experience to drive down costs.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!