IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Laiba Razzaq
15/06/2025

Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

Executive Summary

Web-scraping, data wrangling, data visualization, and machine learning techniques are used to study SpaceX's Falcon 9 rocket launches.

We used exploratory data analysis to identify the features that best predict the success of Falcon 9 launches. Testing different machine learning classifiers, we found that the Decision tree model is the most accurate in predicting the success of Falcon 9 launches.

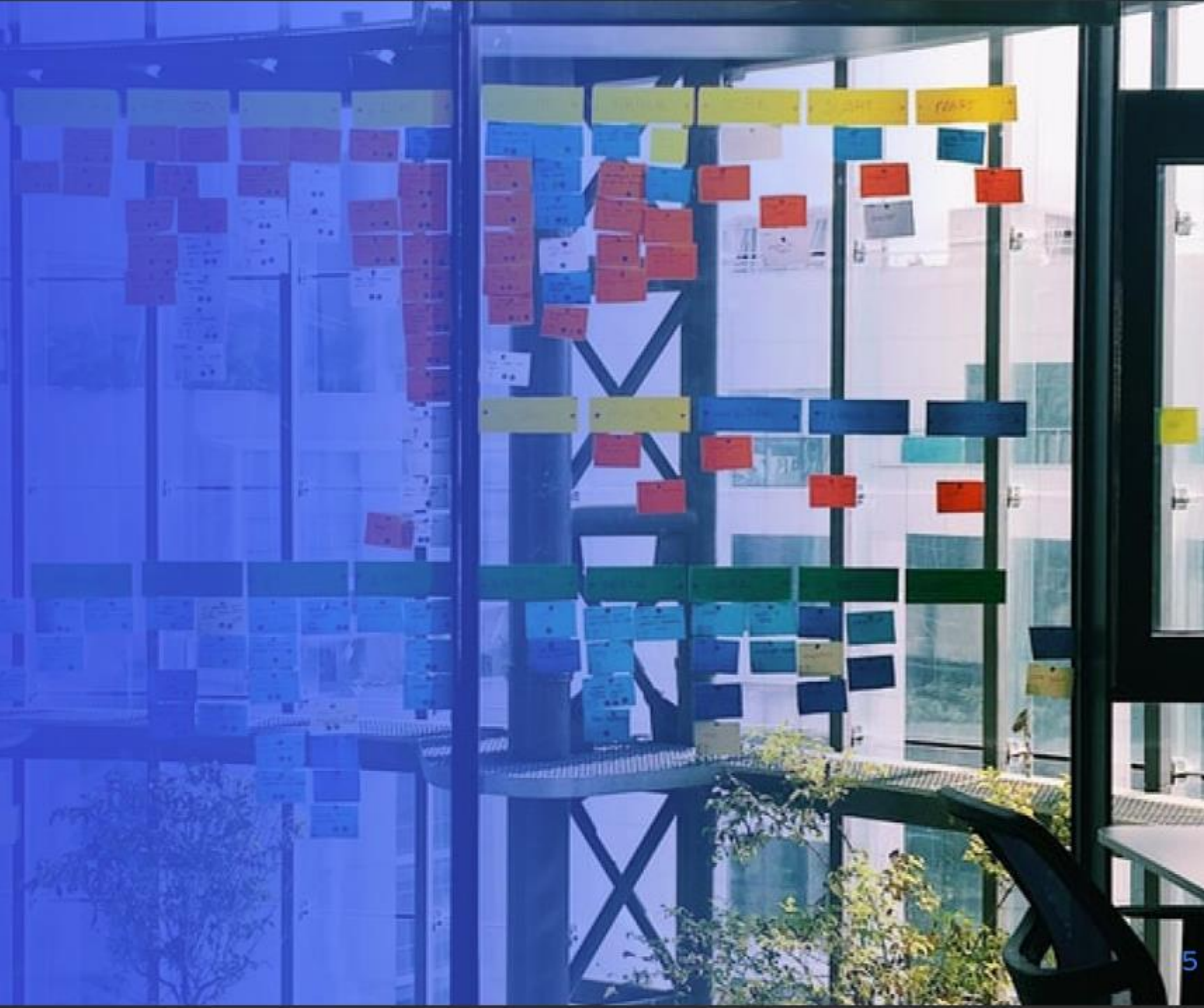
Introduction

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

This project seeks to understand the determinants of Falcon 9 launch success, and tp find the best machine learning model to predict the success of Falcon 9 launches.

Section 1

Methodology



Methodology

Executive Summary

Data collection methodology:

- Space X API and Web-scraping of a Wikipedia page.

Perform data wrangling

- Focus on Falcon 9 launches, and create outcome label based on outcome data.

Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

- Train and tested four classifiers
- Compared model performances using the method score.

Data Collection

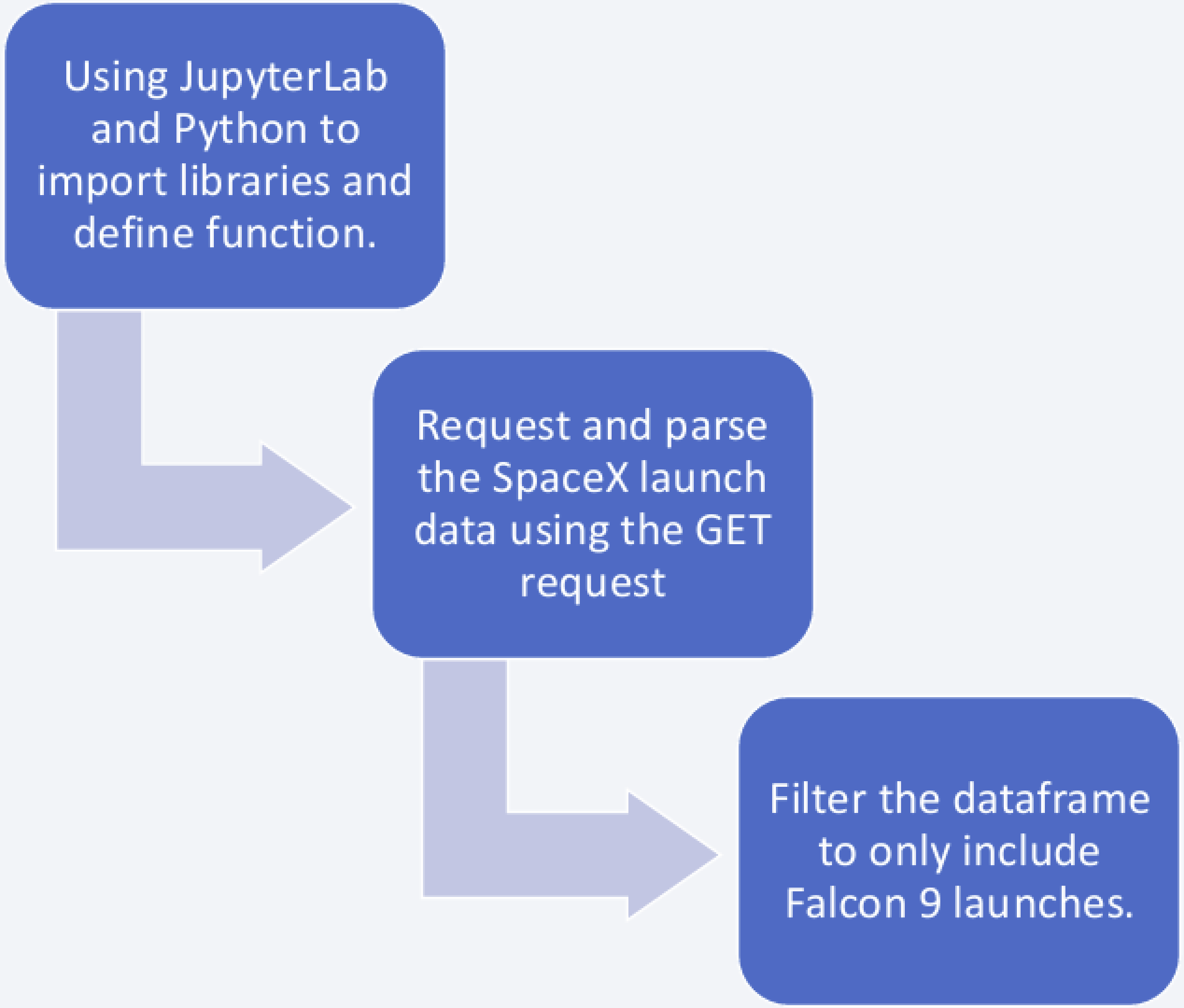
Two main sources of data:

- SpaceX launch data, gathered from SpaceX REST API
- Data on Falcon 9 launches, scraped from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launche)

Data Collection – SpaceX API

- Data is collected through the SpaceX API based on the flow chart:
- Link to the source code of data collection:

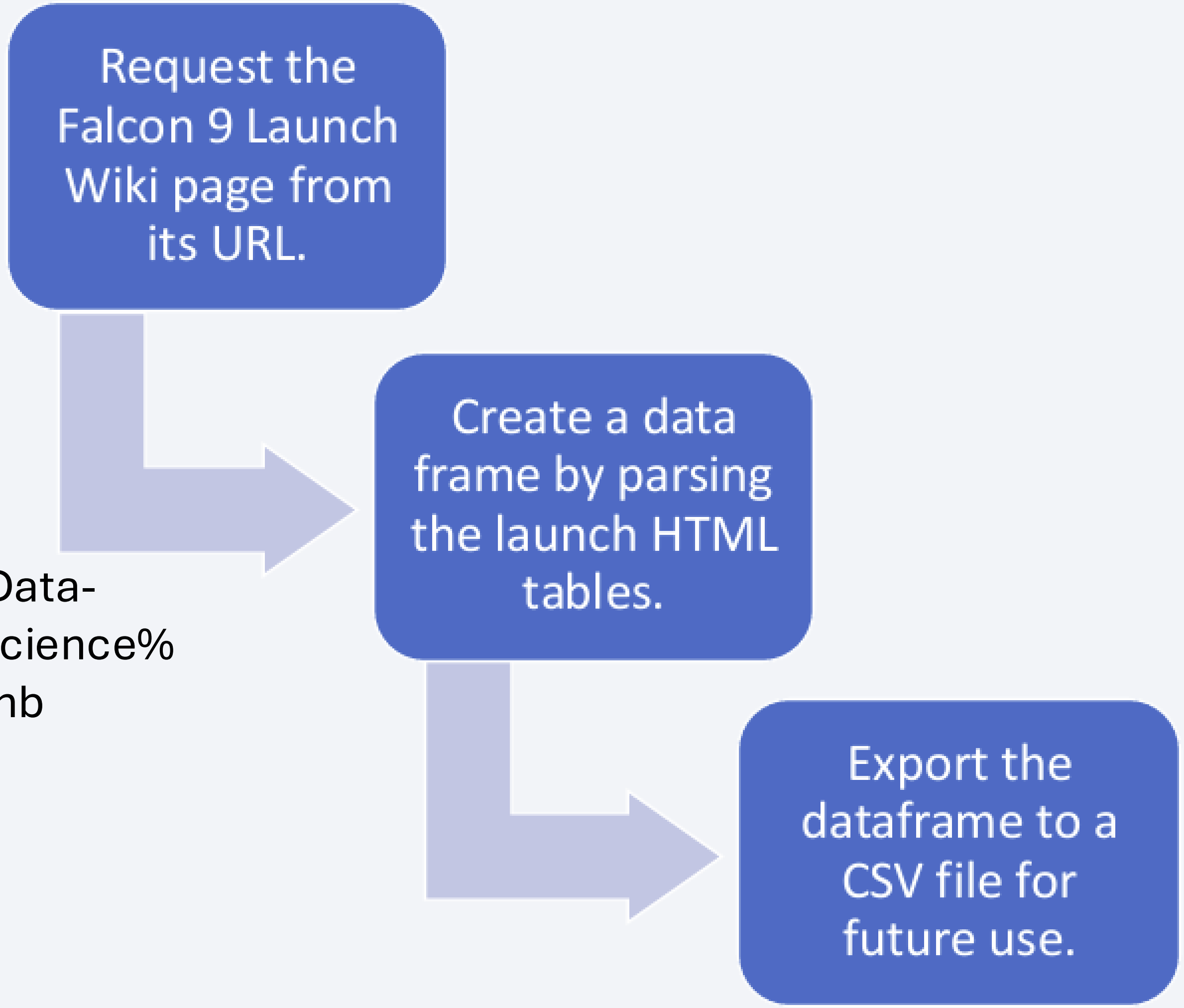
https://github.com/laibarazzaq2010/jupyter-labs-spacex-data-collection=api.ipynb



Data Collection - Scraping

- Data from Wikipedia is scraped based on the flow chart:
- Link to the source code of data scraping:

https://github.com/laibarazaaq2010/IBM-Data-Science/blob/main/Applied%20Data%20Science%20Capstone/jupyter-labs-webscraping.ipynb



Data Wrangling

- Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.
- Key steps are given in the flowchart below:

Calculate the number of launches on each site. → Calculate the number and occurrence of each orbit. → Calculate the number and occurrence of mission outcome of the orbits. → Create a landing outcome label.

- Link to the source code:

https://github.com/laibarazzaq2010/IBM%20Data%20Science/blob/main/Applied%20Data%20Science%20Capstone/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Categorical plots/scatter plots, bar plots, and line plots are used to explore data. Specifically, we used

- Categorical/scatter plots to visualize how factors such as FlightNumber, PayloadMass, LaunchSite, Orbit affect launch success;
- A bar plot to to visualize the relationship between success rate of each orbit type; and
- A line chart to visualize the general trend of success rate across time.

Link to the source code:

https://github.com/laibarzaaq2010/IBM_Data_Science/blob/main/Applied%20Data%20Science%20Capstone/edadataviz.ipynb

EDA with SQL

I performed the following SQL queries:

- Display the names of the unique launch sites
- Display 5 records where launch sites begin with the string "CCA"
- Display the total payload mass carries by boosters launched by NASA(CRS)
- Display the average payload mass carried by booster version F9 v1.1
- List the data when the first successful landing outcome in ground pad was achieved
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but les than 6000
- List the total number of successful and failure mission outcomes
- List all the booster_versions that have carried the maximum payload mass
- Summarize data for year 2015
- Rank the landing outcomes between 2010-06-04 and 2017-03-20

Link to the source code: https://github.com/laibarzaaq2010/IBM%20Data%20Science/blob/main/Applied%20Data%20Science%20Capstone/jupyter-labs-eda-sql-coursera%20sqllite.ipynb

Build an Interactive Map with Folium

Markers, circles, marker clusters, and lines are used in the Folium map. Specifically,

- Markers show specific points, such as launch sites.
- Circles highlight areas around key coordinates (e.g. NASA Johnson Space Center).
- Marker clusters group multiple events at the same location (e.g. launches at a site).
- Lines represent distances between two coordinates.

Link to the source code:

https://github.com/laibarzaaq2010/IBM_Data_Science/blob/main/Applied%20Data%20Science%20Capstone/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- The dashboard application contains input components such as a dropdown list and a range slider for users to interact with a pie chart and a scatter point chart.
- Specifically, there is a Launch Site Drop-down Input Component, a callback function to render success-pie-chart based on selected site dropdown, a Range Slider to Select Payload, and a callback function to render the success-payload-scatter-chart scatter plot
- Link to the source code:

https://github.com/laibarzaaq2010/IBM_Data_Science/blob/main/Applied%20Data%20Science%20Capstone/spacex-dash-app.py

Predictive Analysis (Classification)

- We compared four classification models:
  - Logistic regression
  - Support vector machine
  - Decision tree
  - K-nearest neighbors

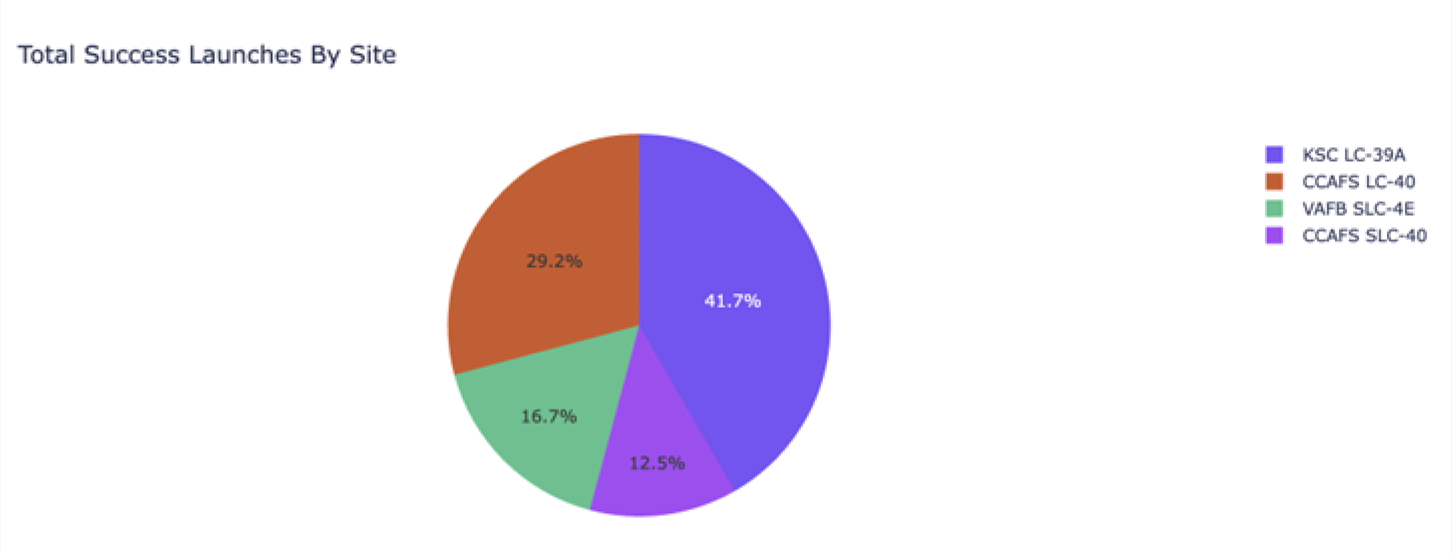
Data preparation and split the data into training and testing sets. → Find the best hyperparameters for each model. → Visualize and compare model performance.

Link to the source code:

https://github.com/laibarzaaq2010/IBM_Data_Science/blob/main/Applied%20Data%20Science%20Capstone/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- Exploratory data analysis results:
  - Space X has 4 launch sites
  - Launch success rate has improved over time
- Interactive analytics demo in screenshots:

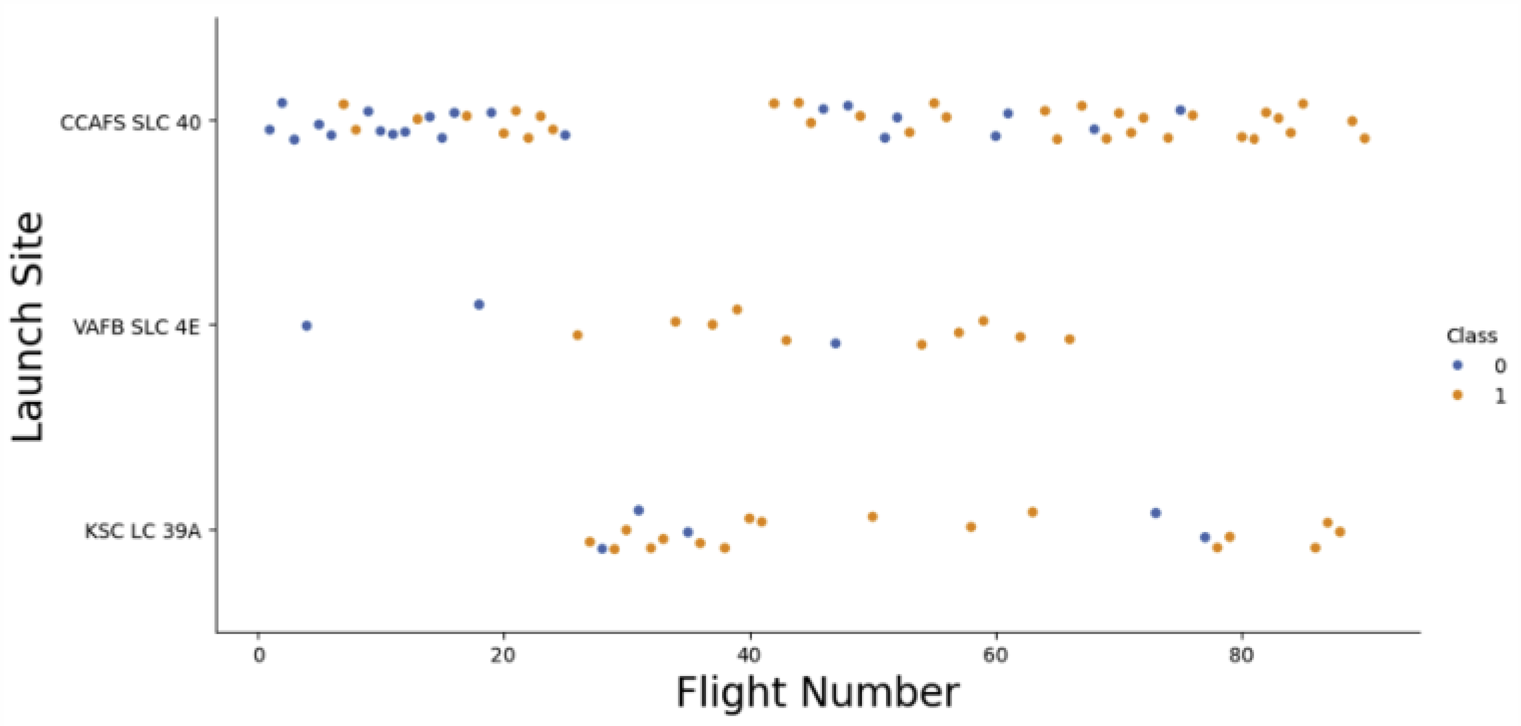


- Predictive Analysis results:
  - Decision tree classifier has the highest accuracy.

Section 2

Insights drawn from EDA

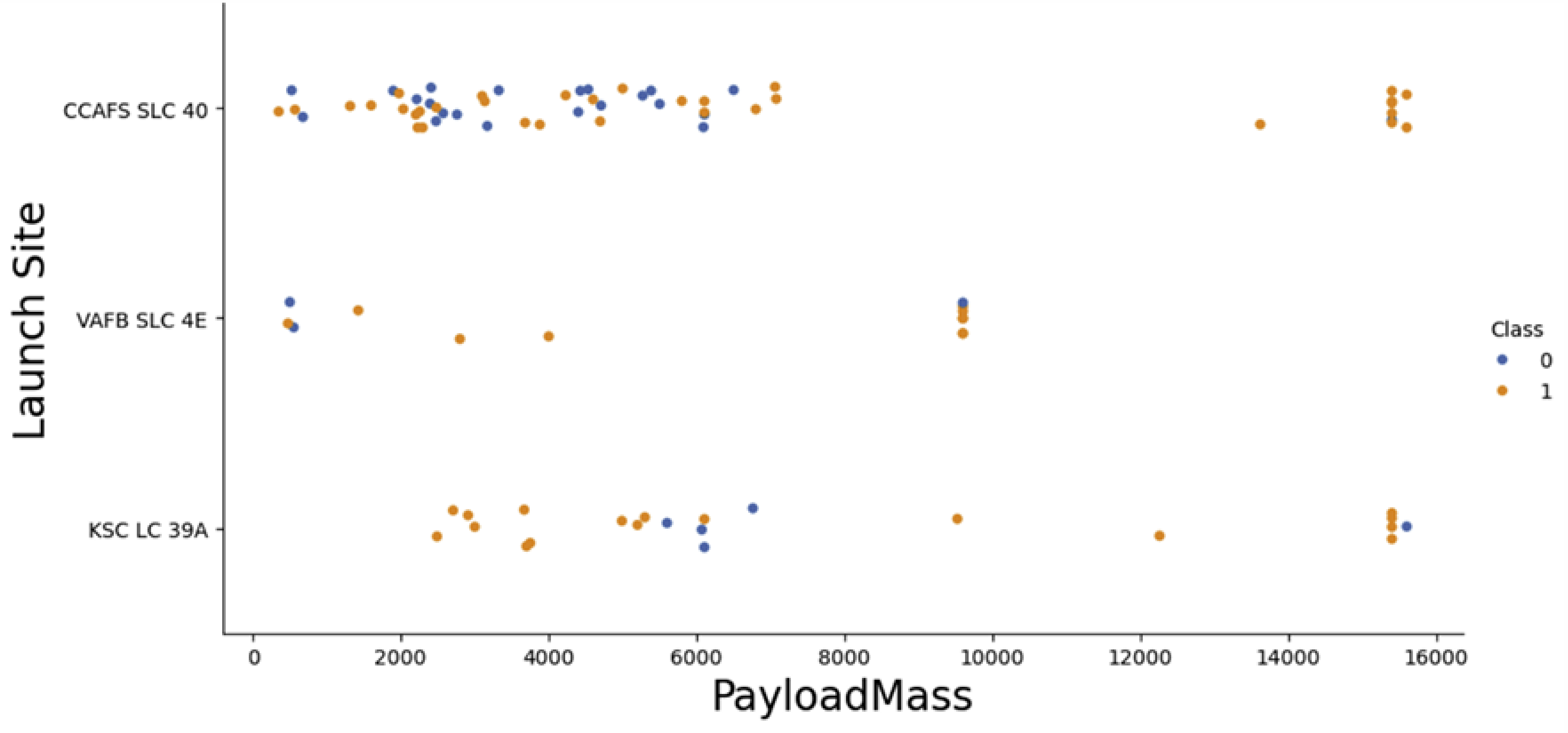
Flight Number vs. Launch Site



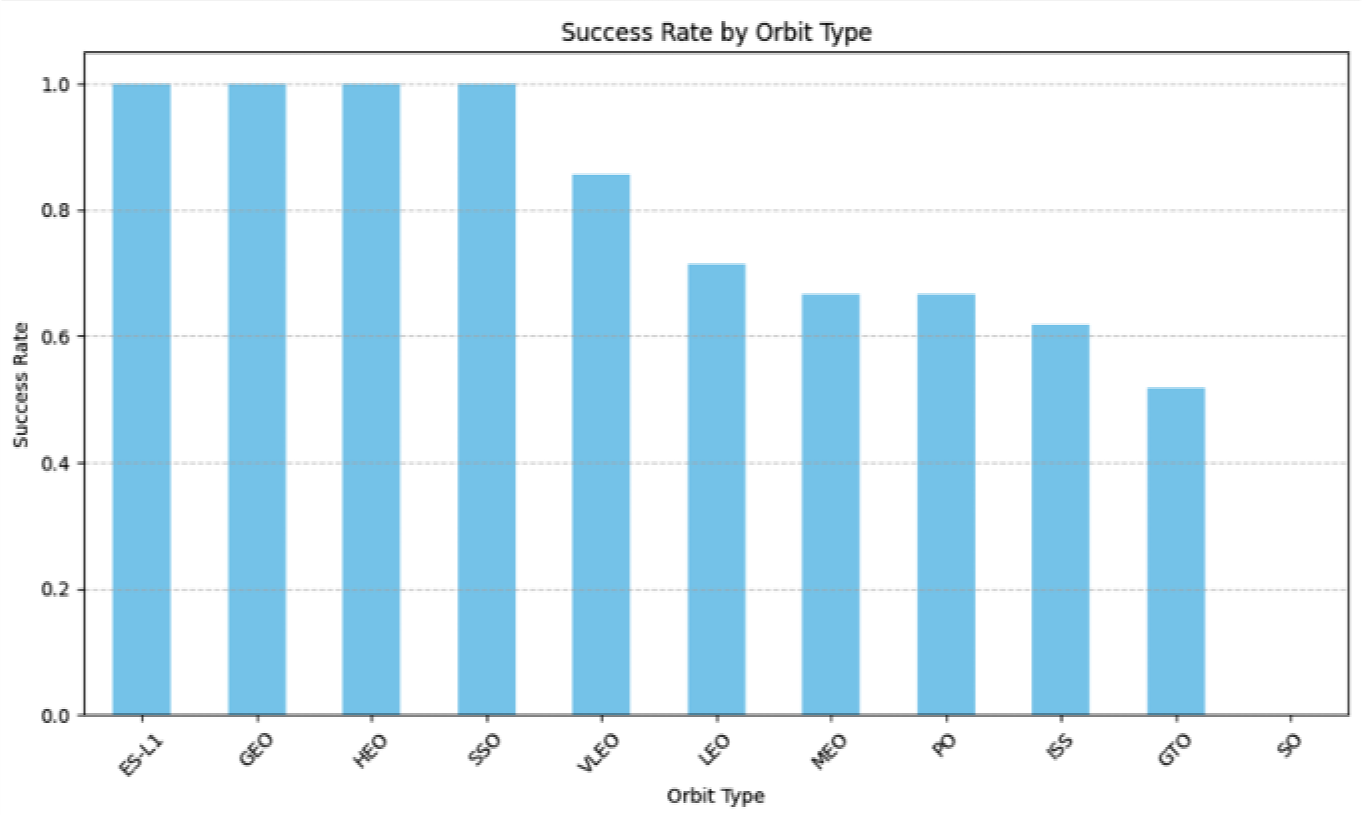
In general, the success rate has improved over time, regardless of the launch site.

Payload vs. Launch Site

- In general, heavier payloads are associated with better success rate.
- Site VAFB SLC 4E's launches have, on average, lower payloads than the other two sites.



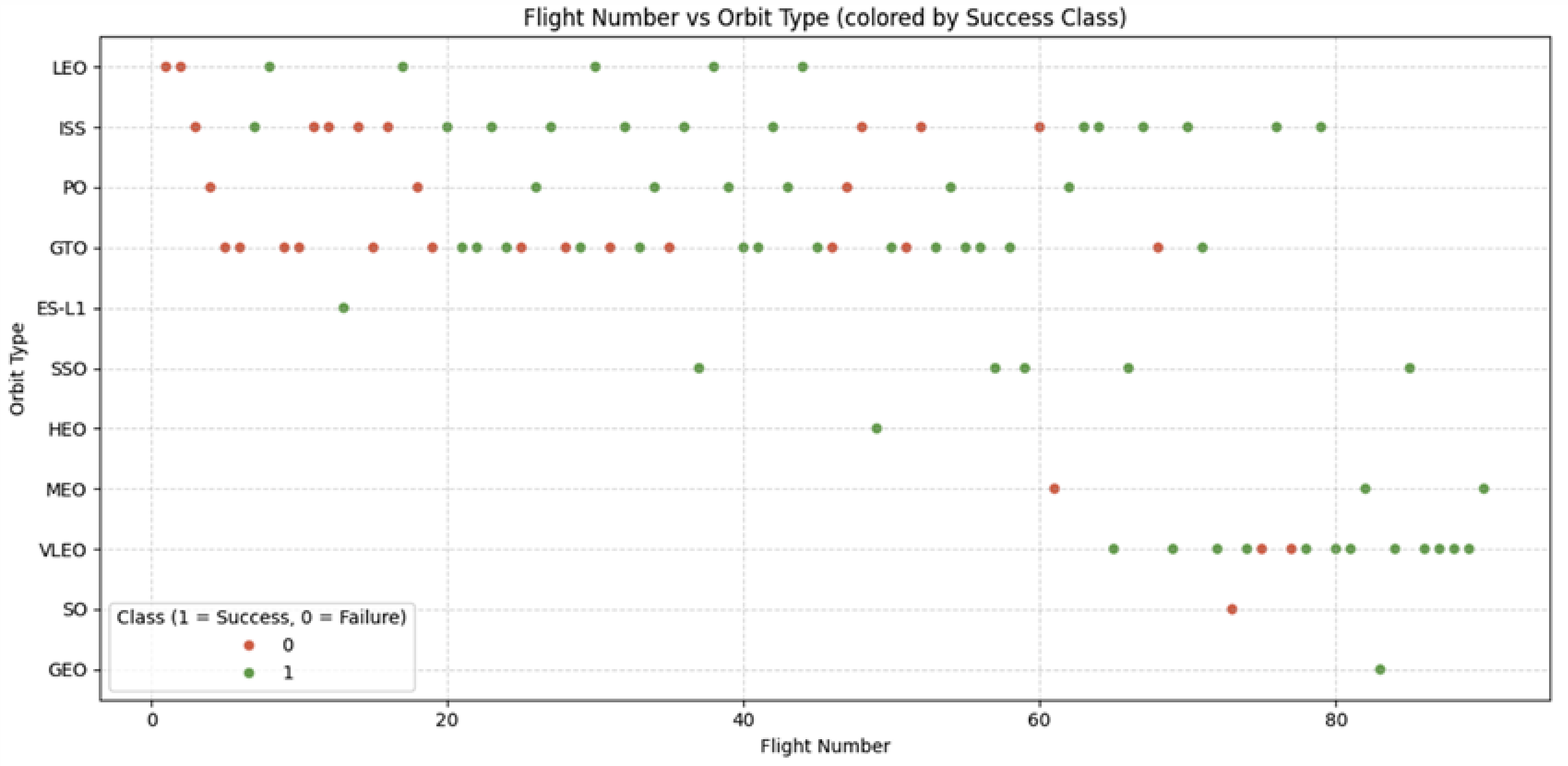
Success Rate vs. Orbit Type



- As we can see in the graph, ES-L1/GEO/HEO/SSO enjoy the highest success rates, while all attemps toward SO have been unsuccessful.

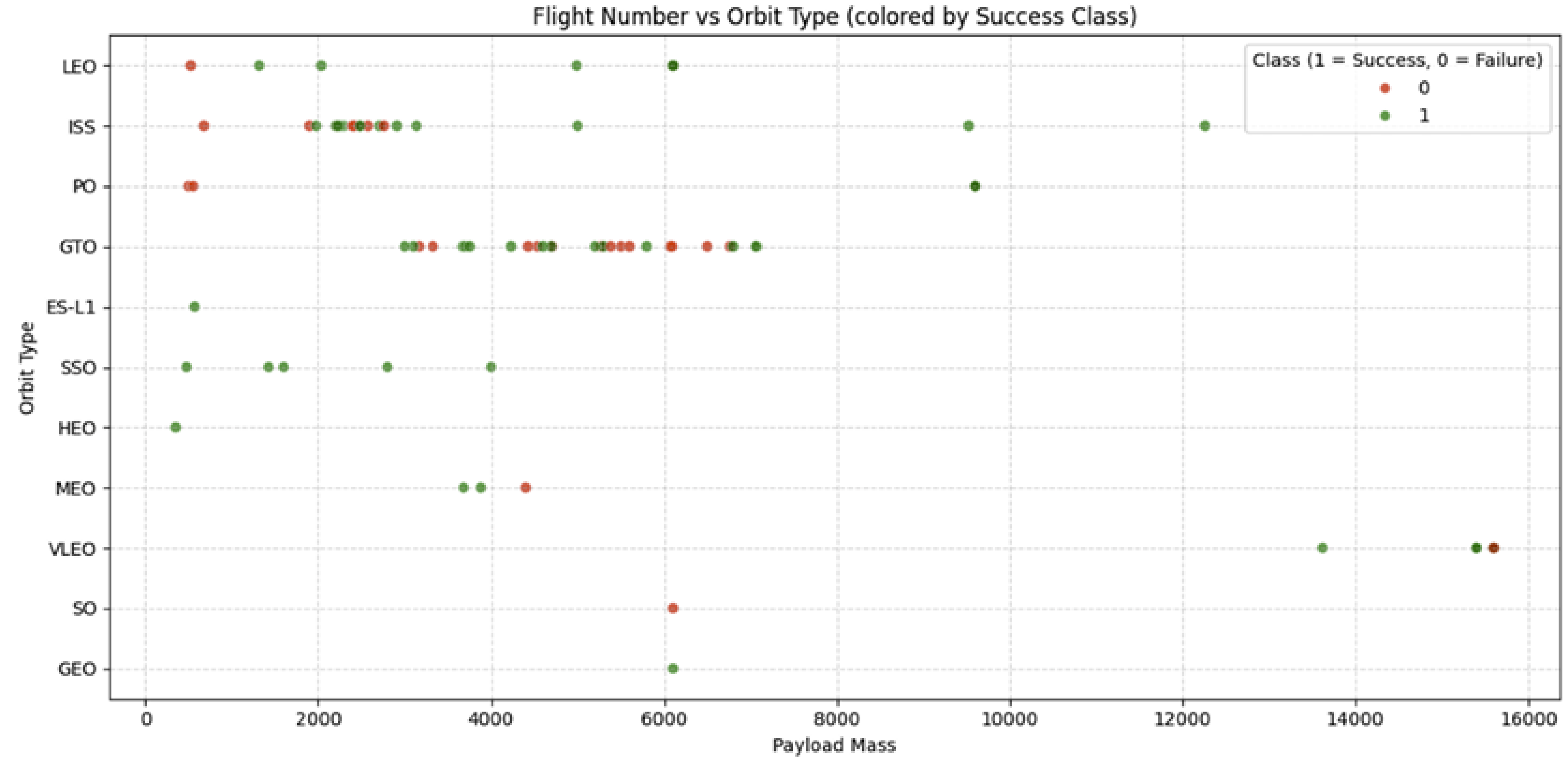
Flight Number vs. Orbit Type

- Succuss rates have improved for all orbit.
- There are more attempt to target SSO, MEO, VLEO and GEO recently, comparing to other orbits.



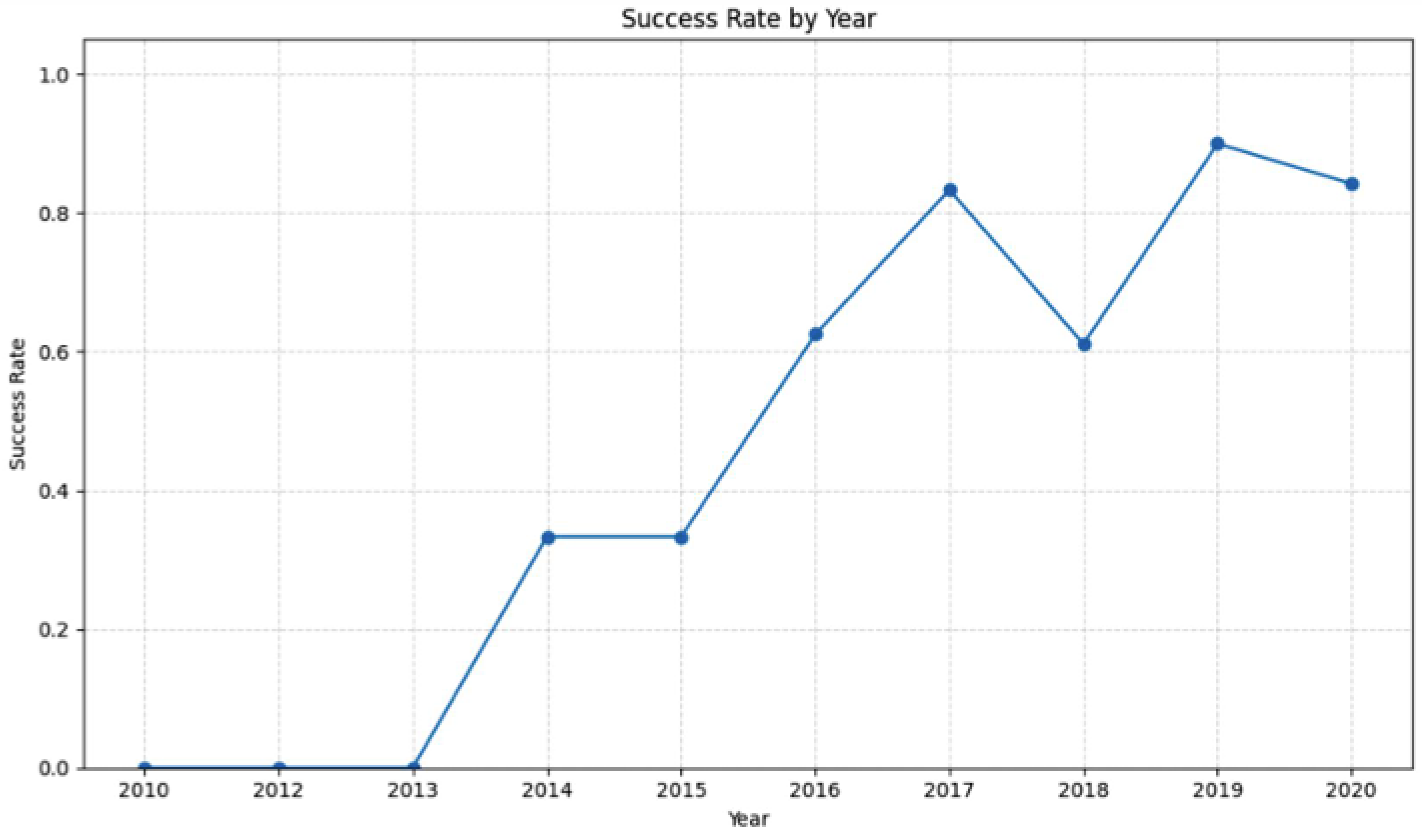
Payload vs. Orbit Type

- For each orbit type, there is no obvious relationship between payload mass and success rate.
- There are many launches toward ISS and GTO but very few launches toward PO, HEO, SO, and GEO.



Launch Success Yearly Trend

- In general, the success rate has increased from 2010 to 2020.
- There was, however, temporary drop in the success rates in 2018 and in 2020.



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- There are four unique launch sites:

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

Launch Site Names Begin with 'CCA'

- The table above shows 5 records where launch sites 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

- The total payload carried by boosters from NASA is 48213kg.

Total_Payload_Mass
48213

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1 is 2928.4kg.

Avg_Payload_Mass
2928.4

First Successful Ground Landing Date

- The first successful landing outcome on ground pad happened on 2015 December 22.

First_Ground_Pad_Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- There are four boosters that have successfully landed on drone ship and had payload mass greater than 4000 but less than 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

- Regarding mission outcomes, there are in total 100 successes and 1 failure:

Mission_Outcome	Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Here are the names of the booster which have carried the maximum payload mass:

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

In 2015, there are two failed launches:

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

- Between 2010-06-04 and 2017-03-20, the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) is ranked below:

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

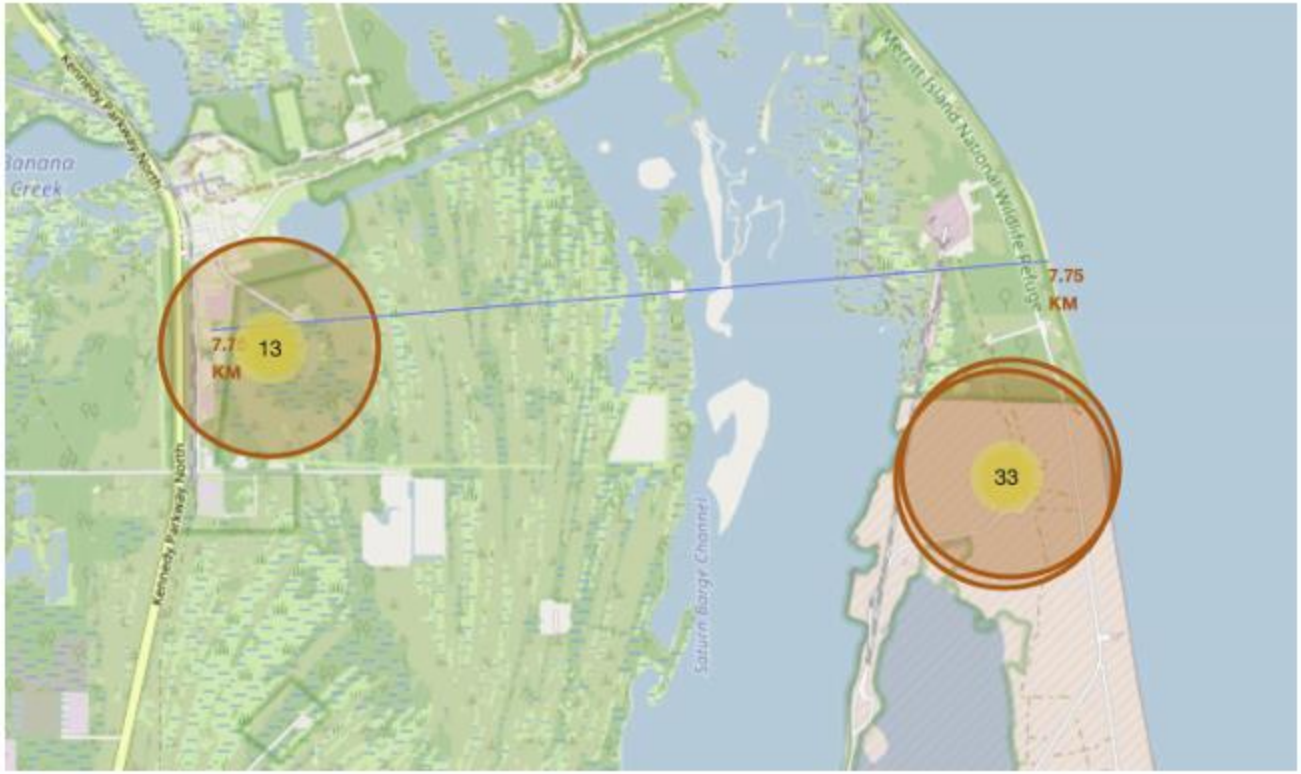
Global Map: All Launch Site

- Globally, the launch sites can be found in California and Florida in the United States.

Launch Outcomes of Site KSC LC-39A

- Here is an example of launch outcomes at KSC LC-39A. Red marks denote failed launches and green marks denote successful launches.

Proximity of Launch Site KSC LC-39A to the Coastline



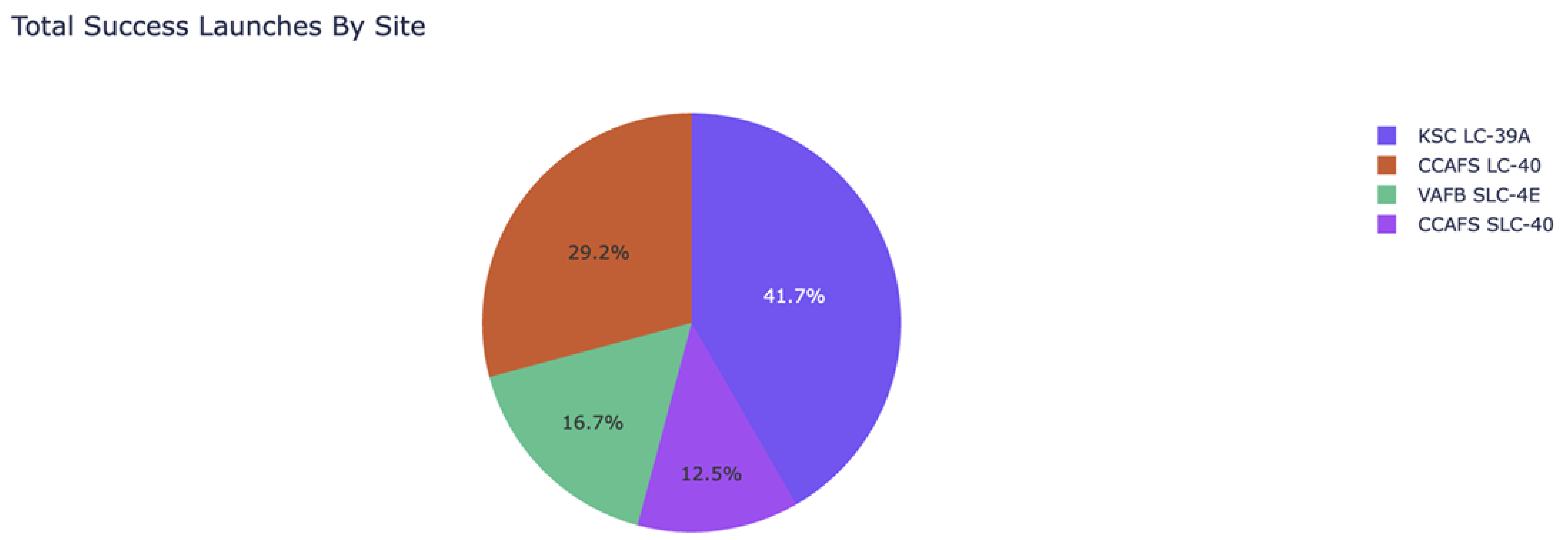
- The launch site KSC LC-39A is 7.75km from the coastline.

Section 4

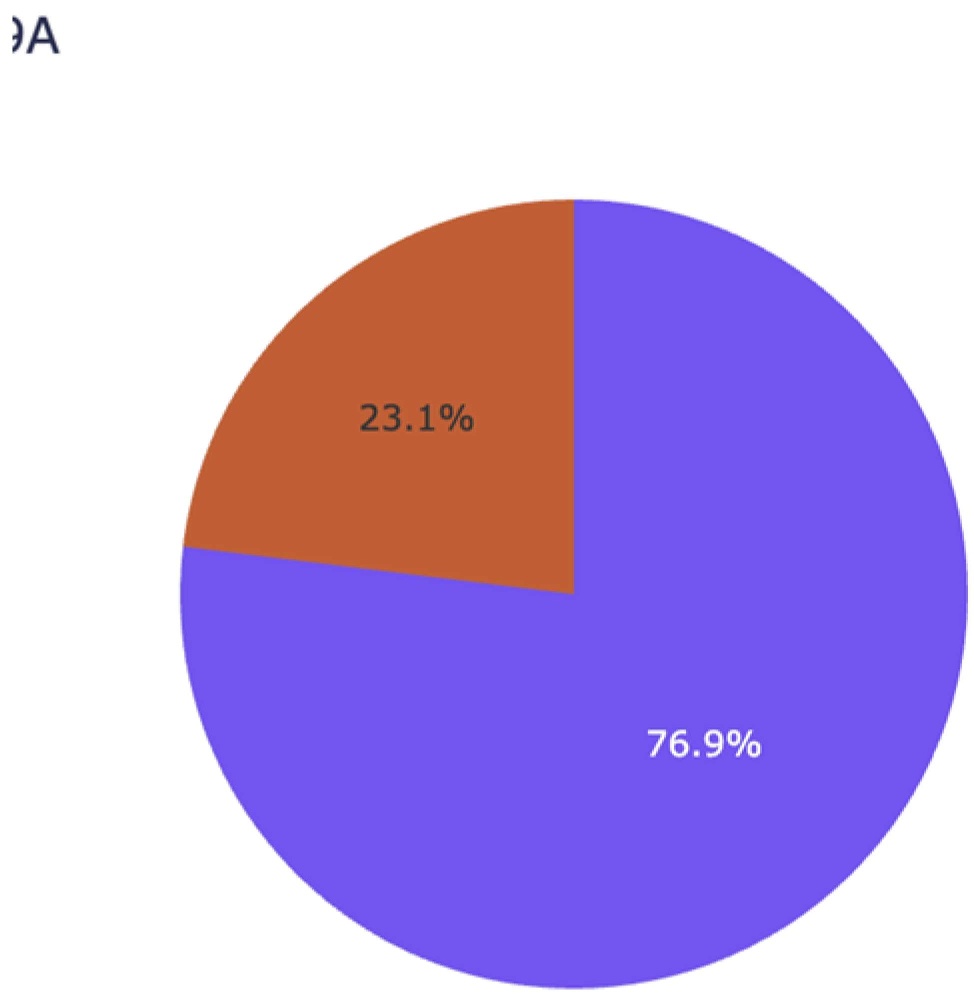
Build a Dashboard with Plotly Dash

Launch Success Counts

- KSC LC-39A has the highest number of successful launches.
- VAFB SLC-4E has the lowest number of successful launches.

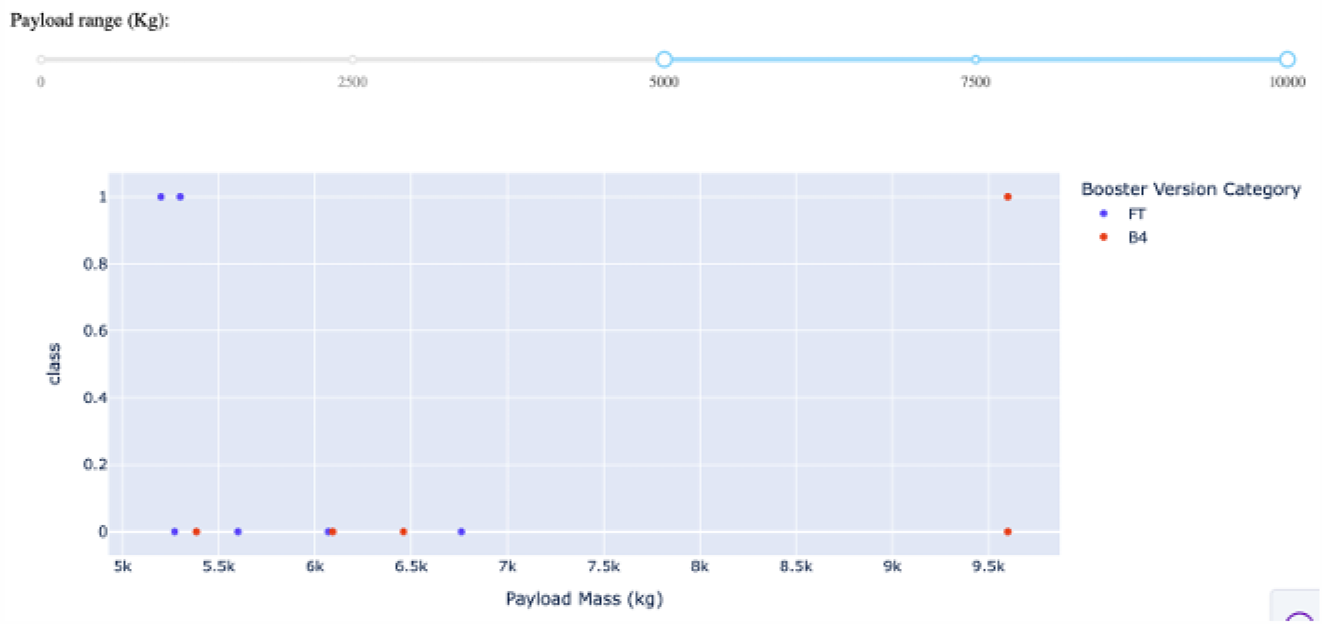


Launch Success Rate of KSC LC-39A



- KSC LC-39A has a success rate of 76.9%.

Relationship between payload and launch outcome



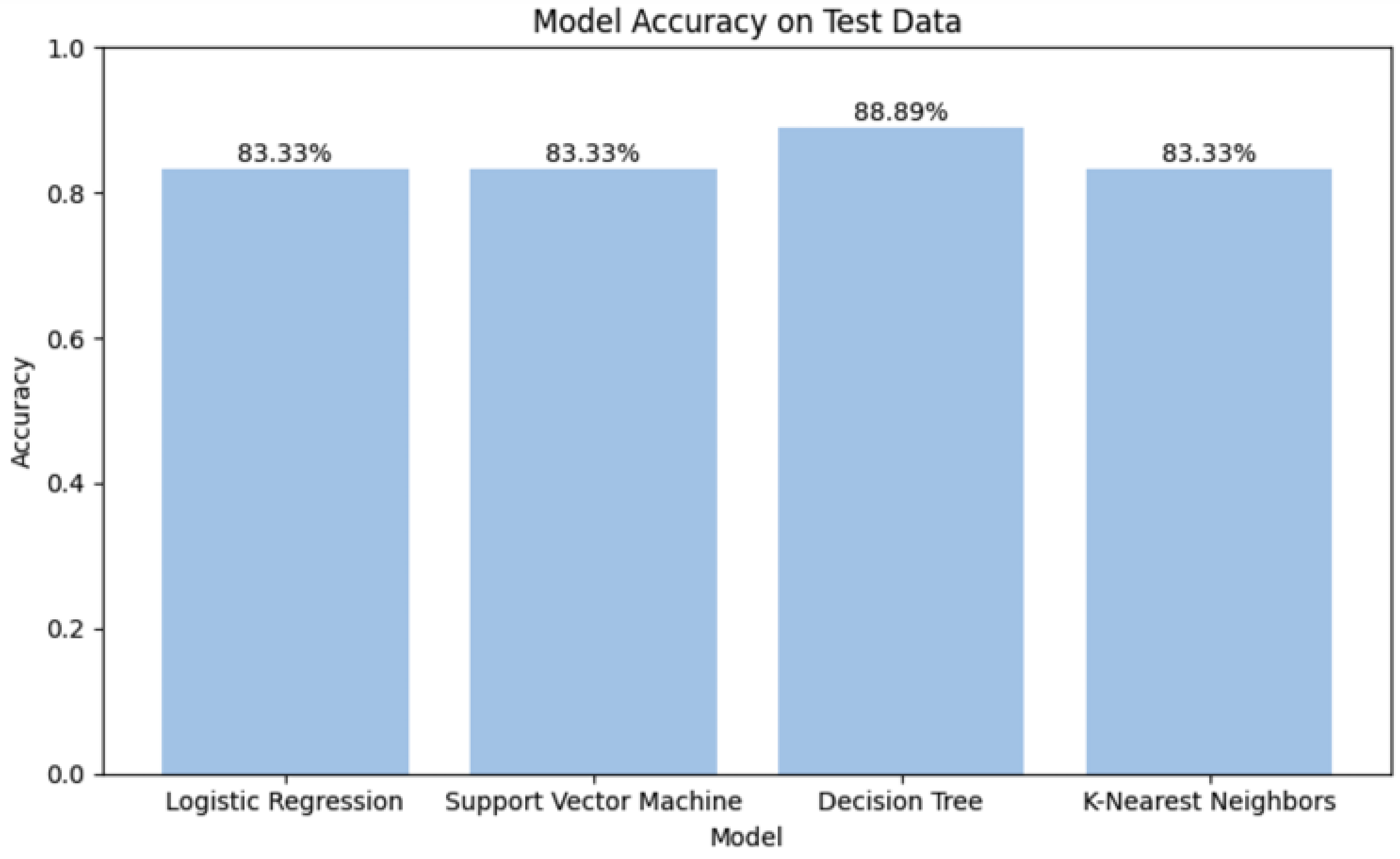
- As is shown on top, overall, booster FT yields the highest success rate.
- As is shown on the bottom, we see that only FT and B4 boosters carry heavy payload mass that are between 5000 and 10000 kg.

Section 5

Predictive Analysis (Classification)

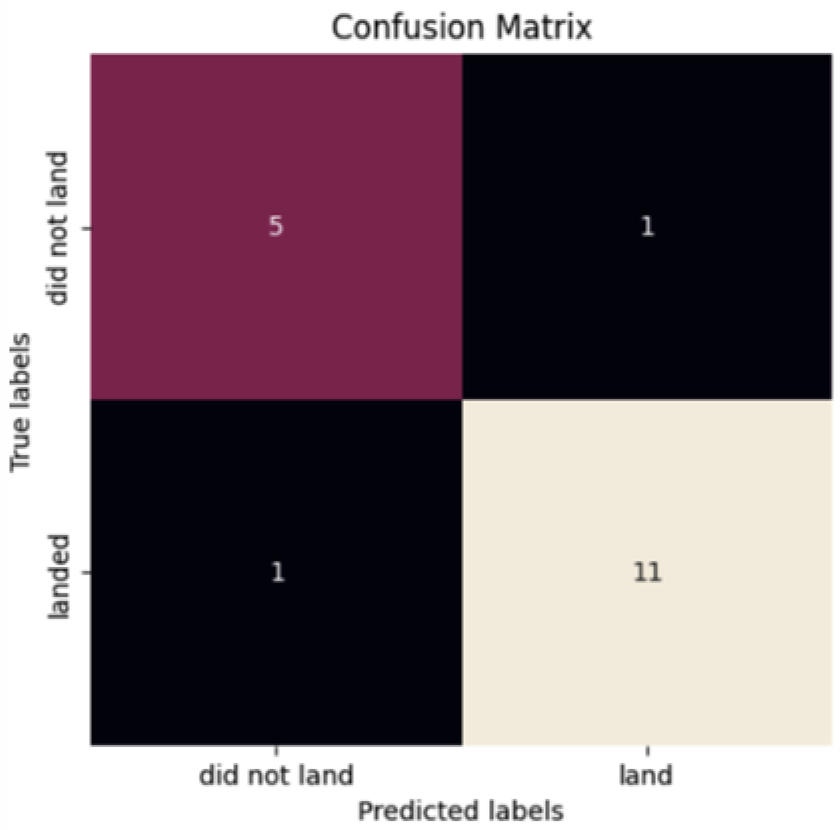
Classification Accuracy

Based on the bar chart, decision tree model has the best accuracy.



Confusion Matrix

- This is the confusion matrix of the Decision Tree model. predicts whether a landing was successful or not. Here's a quick breakdown:
  - **True Negatives (Top-Left): 5** - The model correctly predicted "did not land" 5 times.
  - **False Positives (Top-Right): 1** - The model incorrectly predicted "land" 1 time when it actually "did not land."
  - **False Negatives (Bottom-Left): 1** - The model incorrectly predicted "did not land" 1 time when it actually "landed."
  - **True Positives (Bottom-Right): 11** - The model correctly predicted "landed" 11 times.
- In short, the model performed well, correctly identifying the outcome in 16 out of 18 cases. It made only two errors: one false positive and one false negative.



Conclusions

- The best launch site is KSC LC-39A
- Success rate has improved over time
- Decision tree classifier has the highest accuracy rate in our setting.


Thank you!