IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Raushan Zhandayeva
June 18, 2024

Outline

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

Executive Summary

- **Summary of methodologies:**
  - Data collection: web scraping and SpaceX API
  - Exploratory data analysis: data wrangling, data visualization, interactive dashboard
  - Machine learning: logistic regression, support vector machines, decision tree, kNN

- **Summary of all results:**
  - The launch site location appears to be a very important factor in mission success.
  - All launch sites are located near coastlines, likely for safety reasons.
  - KSC LC-39A has the highest ratio of successful launches at 76.9%.
  - Payloads under 6,000 kg and Falcon 9 FT boosters have the highest success rate.
  - Of the four classification models tested, the Decision Tree Classifier performed best with around 89% accuracy. The confusion matrix validates its strong predictive power.

Introduction

- **Project background and context:**
  - 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
  - The objective is to evaluate the viability of the new company Space Y (rival to SpaceX) to compete with Space X

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology
  - Data from Space X was obtained from 2 sources:
    - Space X API (https://api.spacexdata.com/v4/rockets/)
    - WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

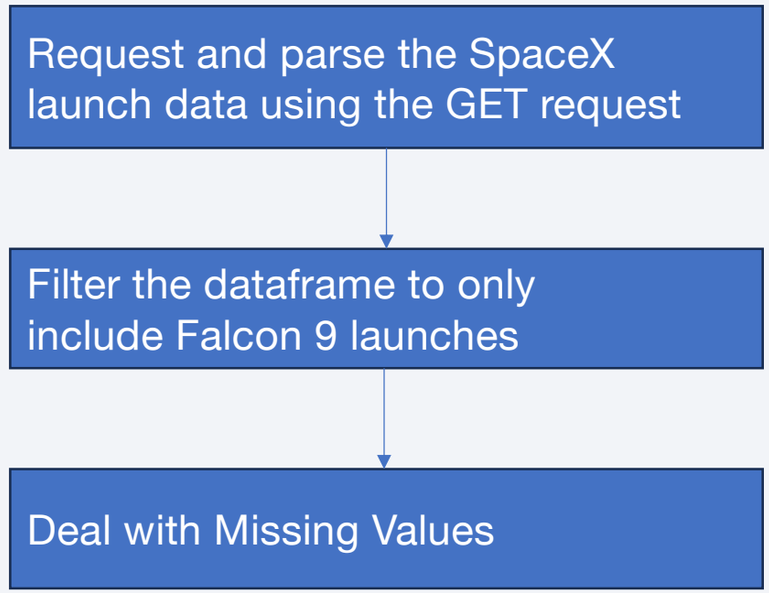
Data Collection

- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches) using web scraping technics.

Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained and then used

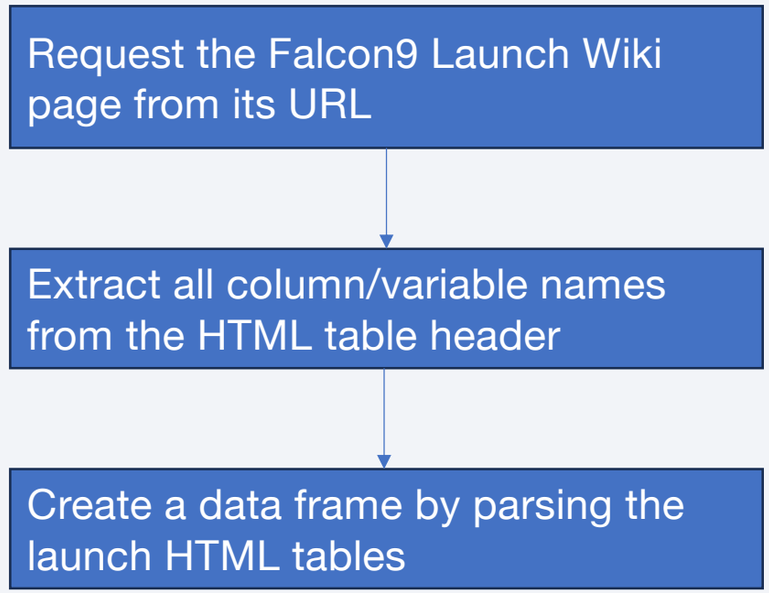
- GitHub URL: https://github.com/raushan-zhandayeva/Data-Science-Capstone/blob/90a07a9356d03bbbba13a13fcc92179756183b86/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia

- GitHub URL: https://github.com/raushan-zhandayeva/Data-Science-Capstone/blob/90a07a9356d03bbbba13a13fcc92179756183b86/jupyter-labs-webscraping.ipynb



Data Wrangling

Two key steps:

1. Exploratory Data Analysis (EDA): calculating the number of launches on each site; calculating the number and occurrence of each orbit; calculating the number and occurrence of mission outcome of the orbits

2. Creation of a landing outcome label from Outcome column

GitHub URL: https://github.com/raushan-zhandayeva/Data-Science-Capstone/blob/90a07a9356d03bbbba13a13fcc92179756183b86/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between different pairs of features

GitHub URL: https://github.com/raushan-zhandayeva/Data-Science-Capstone/blob/main/edadataviz.ipynb

EDA with SQL

SQL queries performed:

- 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 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 less than 6000
- List the names of the booster versions which have carried the maximum payload mass.
- 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.

GitHub URL: https://github.com/raushan-zhandayeva/Data-Science-Capstone/blob/90a07a9356d03bbbba13a13fcc92179756183b86/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Map objects: markers, circles, marker clusters lines

- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates;
- Marker clusters indicate groups of events in each coordinate, like launches in a launch site;
- Lines are used to indicate distances between two coordinates

GitHub URL: https://github.com/raushan-zhandayeva/Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

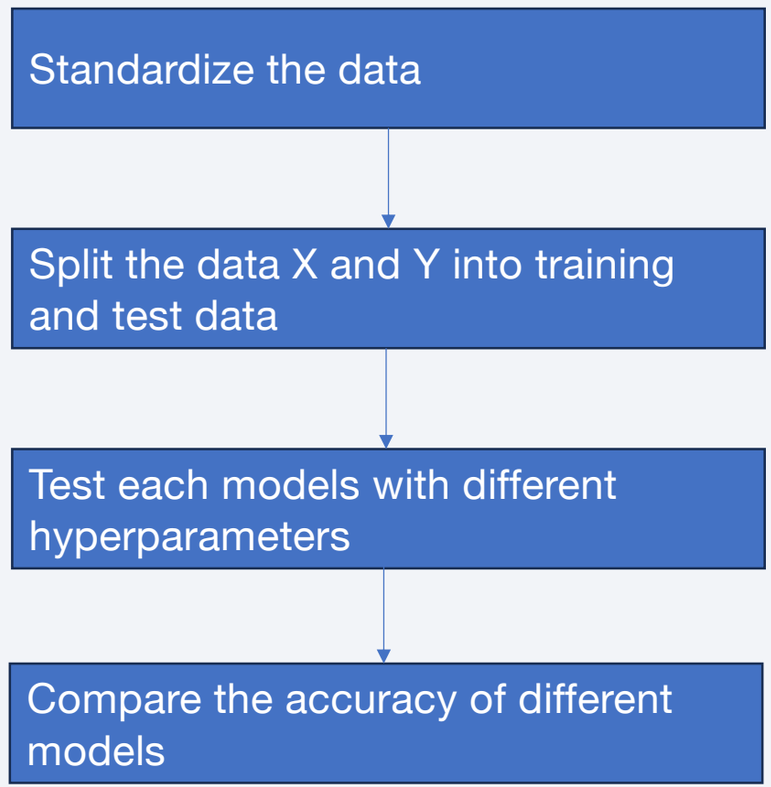
- **The following graphs and plots were used to visualize data:**
  - Percentage of launches by site
  - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads

GitHub URL: https://github.com/raushan-zhandayeva/Data-Science-Capstone/blob/90a07a9356d03bbbba13a13fcc92179756183b86/spacex_dash_app.py

Predictive Analysis (Classification)

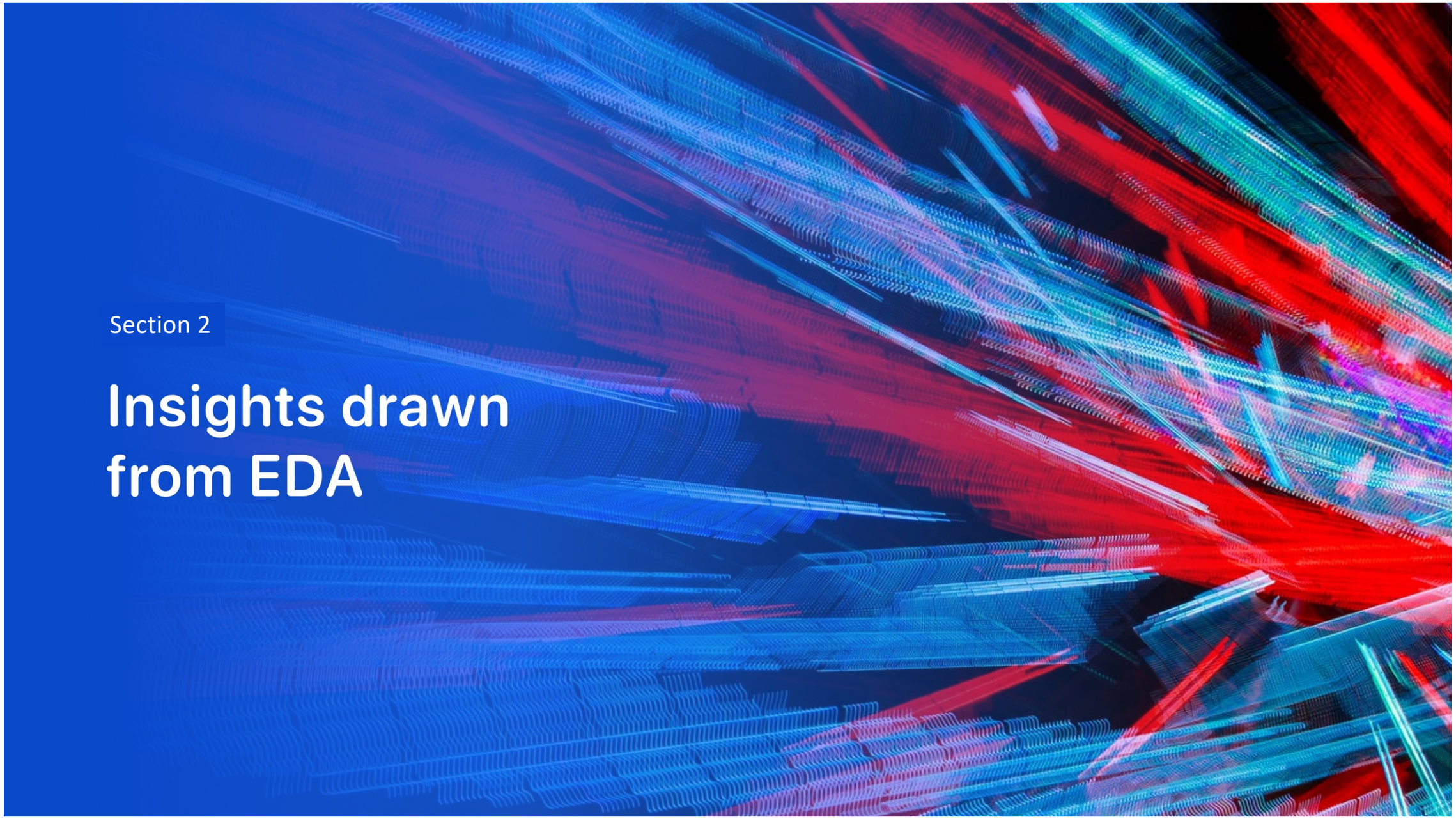
- **Models:** logistic regression, support vector machines, decision tree, kNN

GitHub URL: https://github.com/raushan-zhandayeva/Data-Science-Capstone/blob/90a07a9356d03bbbba13a13fcc92179756183b86/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

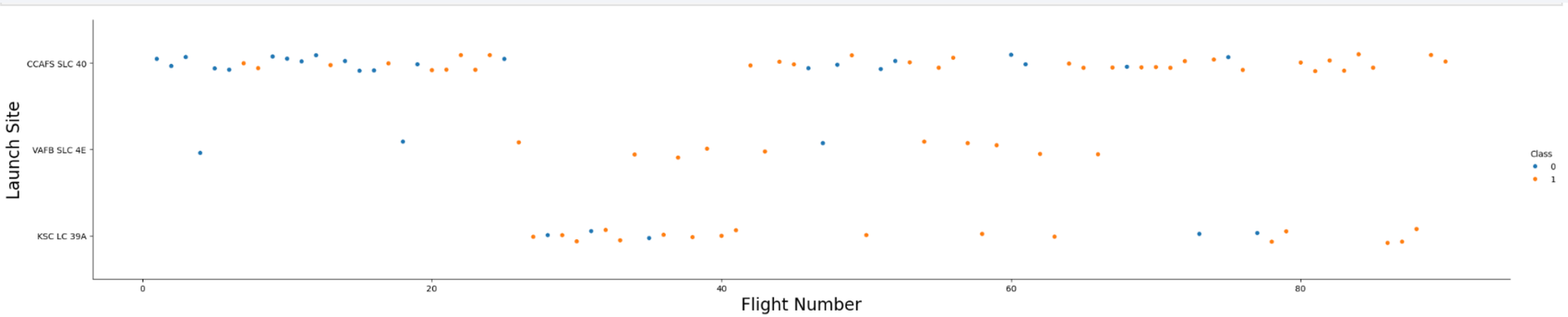


Section 2

Insights drawn from EDA

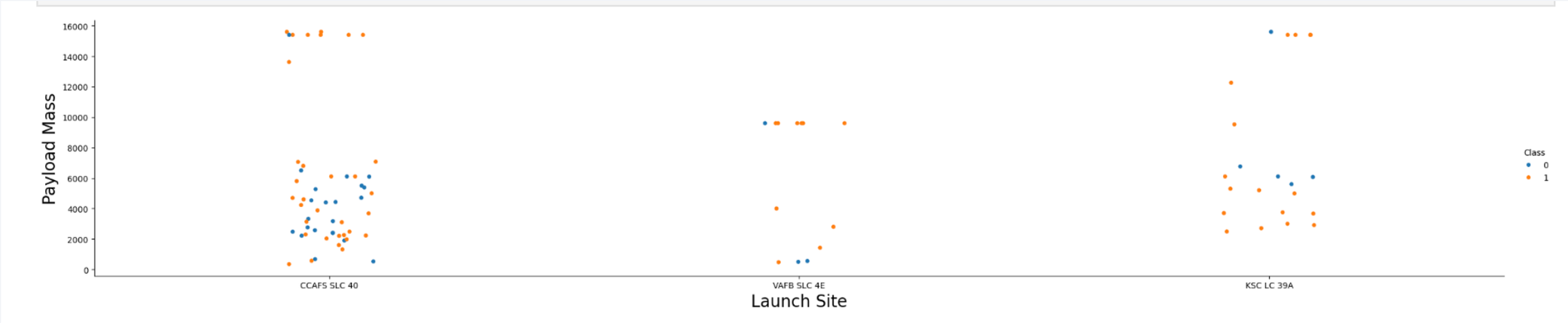


Flight Number vs. Launch Site



- The earliest flights all failed while the latest flights all succeeded.
- The CCAFS SLC 40 launch site has about a half of all launches.
- VAFB SLC 4E and KSC LC 39A have higher success rates.
- It can be assumed that each new launch has a higher rate of success.

Payload vs. Launch Site

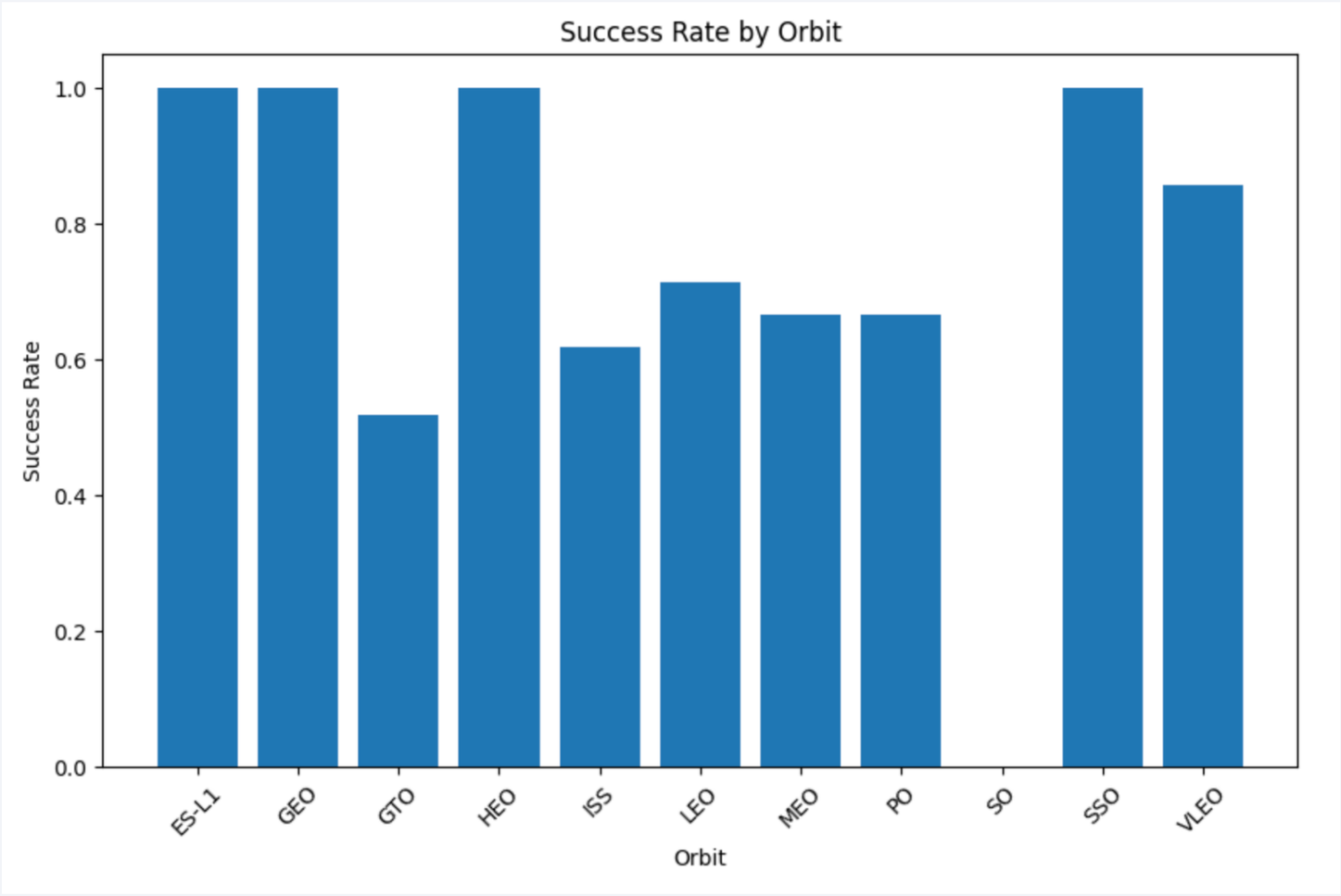


- For every launch site the higher the payload mass, the higher the success rate.
- Most of the launches with payload mass over 7000 kg were successful.
- KSC LC 39A has a 100% success rate for payload mass under 5500 kg too.

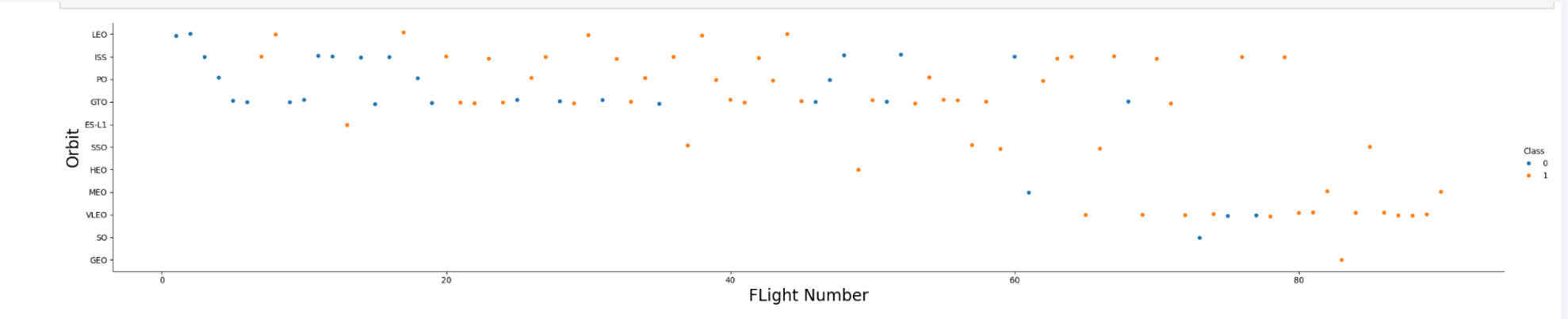
Success Rate vs. Orbit Type

The following orbits have the highest success rate:

- ES-L1;
- GEO;
- HEO; and
- SSO.

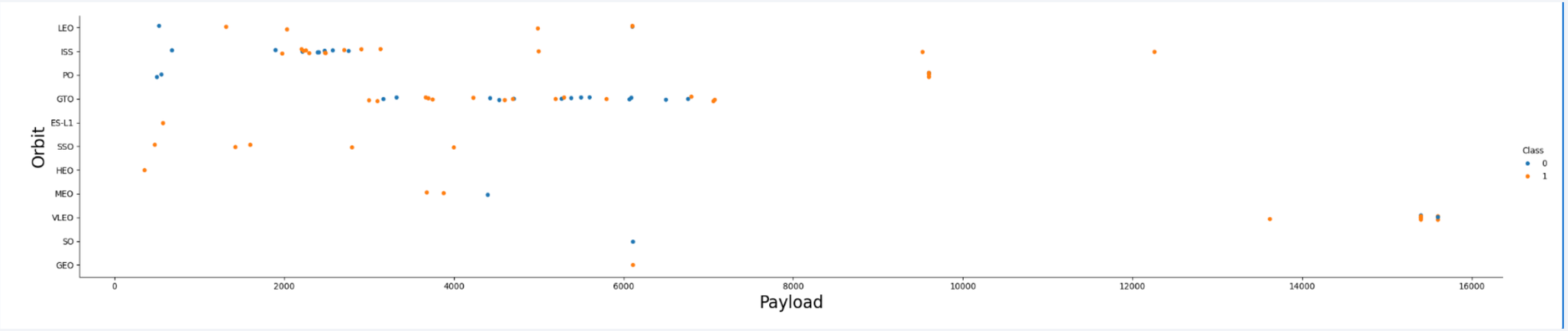


Flight Number vs. Orbit Type



- In the LEO orbit the Success appears related to the number of flights.
- There seems to be no relationship between flight number when in GTO orbit.

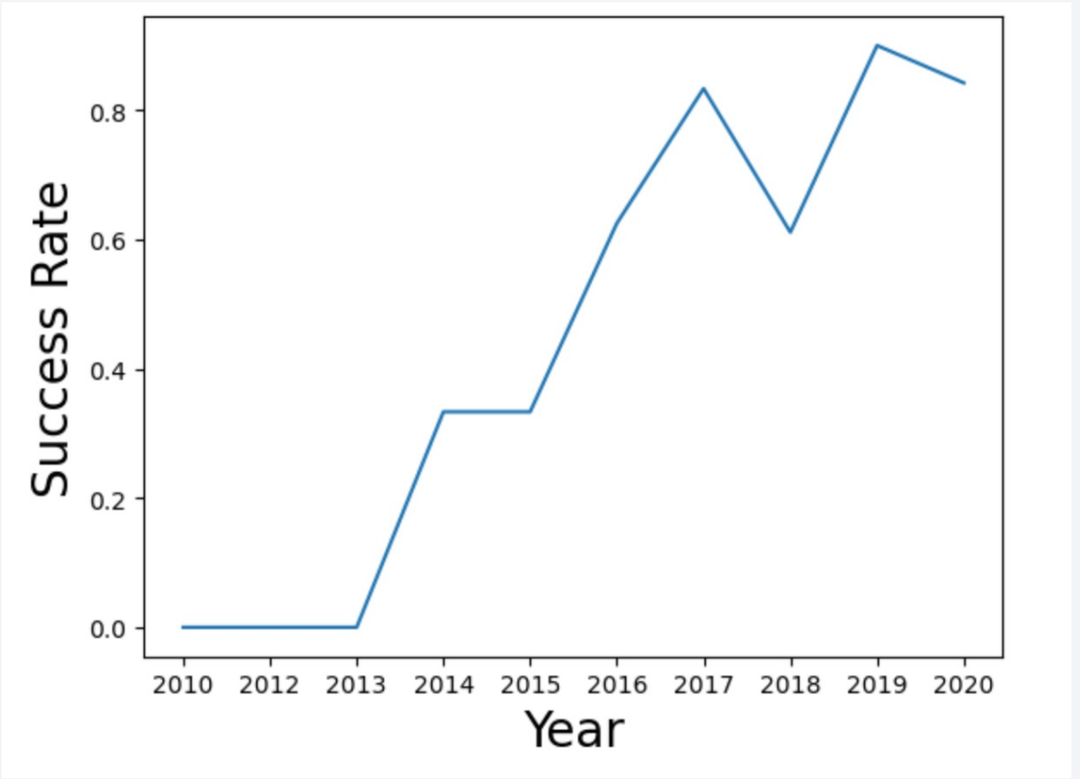
Payload vs. Orbit Type



- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend

The success rate since 2013 kept increasing till 2020.



All Launch Site Names

The names of the unique launch sites:

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

Launch Site Names Begin with 'CCA'

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: 45,596 (KG)

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1: 2928.4 (KG)

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad: 2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

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

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes:

- Successful mission: 100
- Failed mission: 1

Boosters Carried Maximum Payload

The booster which have carried the maximum payload mass: Booster Version (...)

- F9 B5 B1048.4
- F9 B5 B1048.5
- F9 B5 B1049.4
- F9 B5 B1049.5
- F9 B5 B1049.7
- F9 B5 B1051.3
- F9 B5 B1051.4
- F9 B5 B1051.6
- F9 B5 B1056.4
- F9 B5 B1058.3
- F9 B5 B1060.2
- F9 B5 B1060.3

2015 Launch Records

The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

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

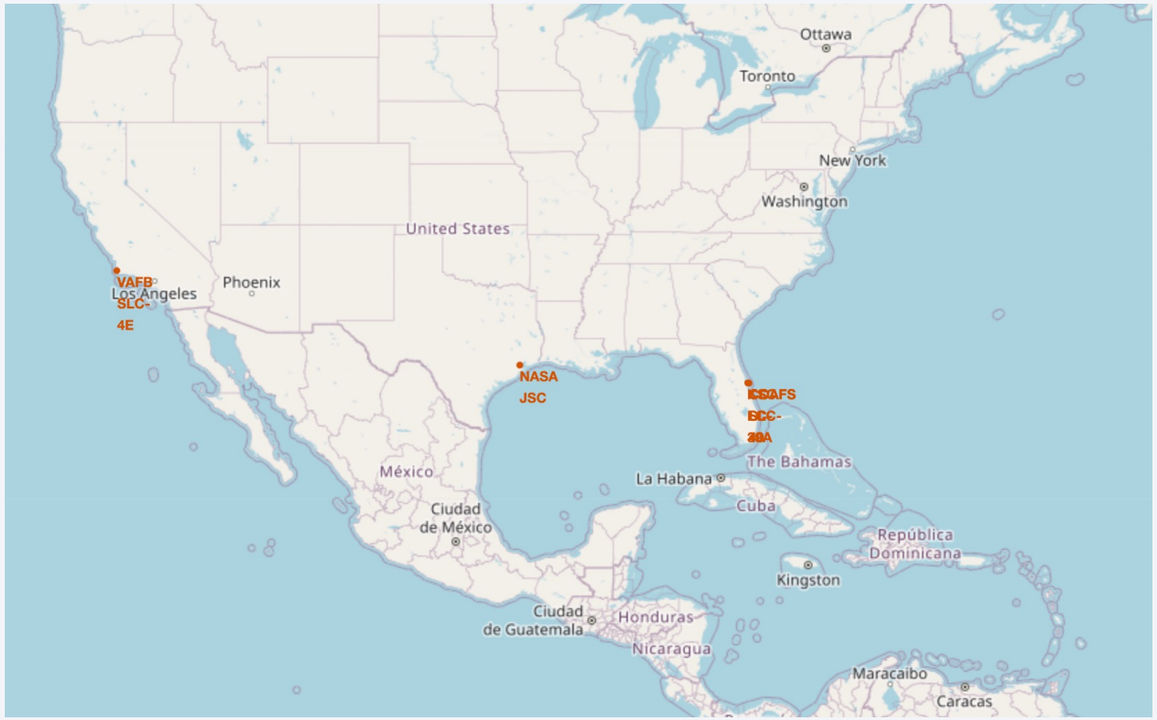
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_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

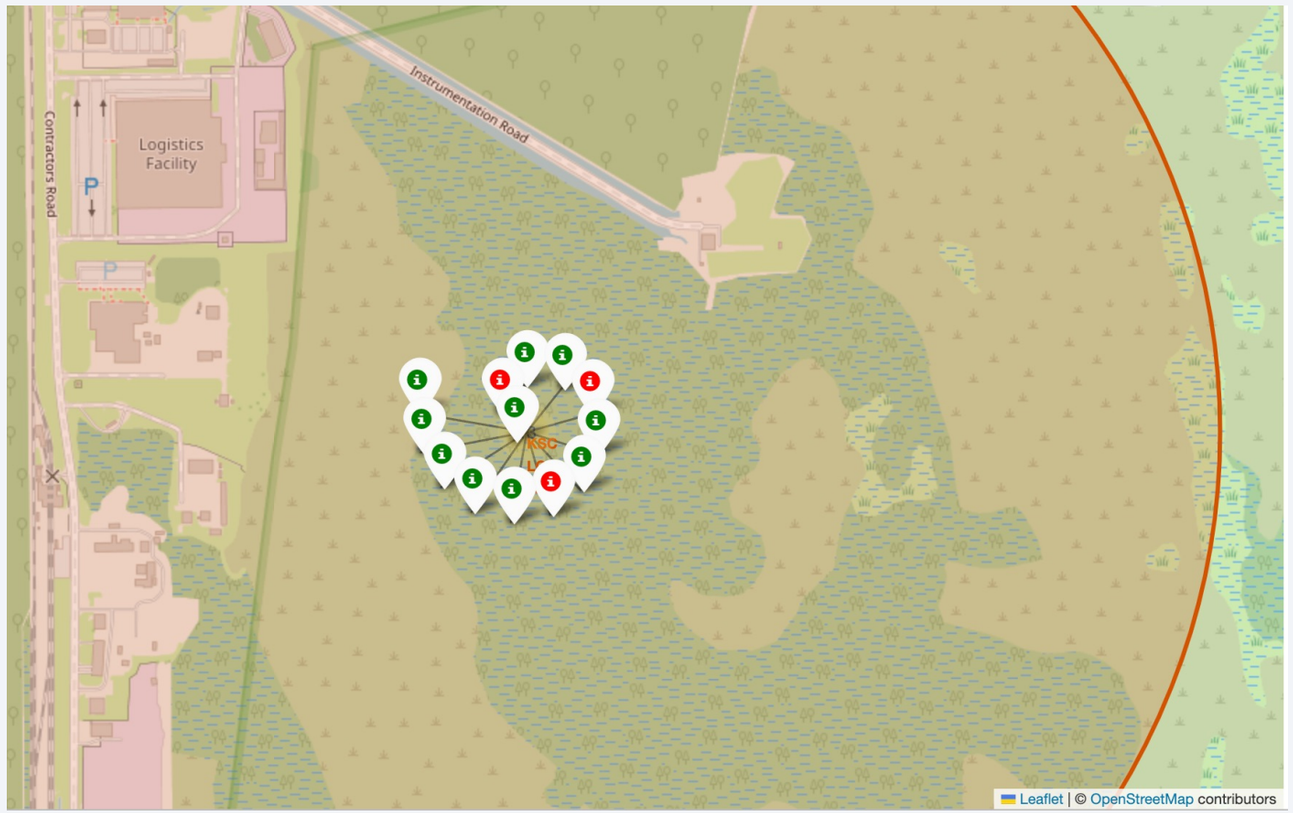
Launch Sites Proximities Analysis

All Launch Sites on a Map



- Launch sites are near the coastline probably due to safety reasons.

Successful/Failed Launches for One of the Launch Sites



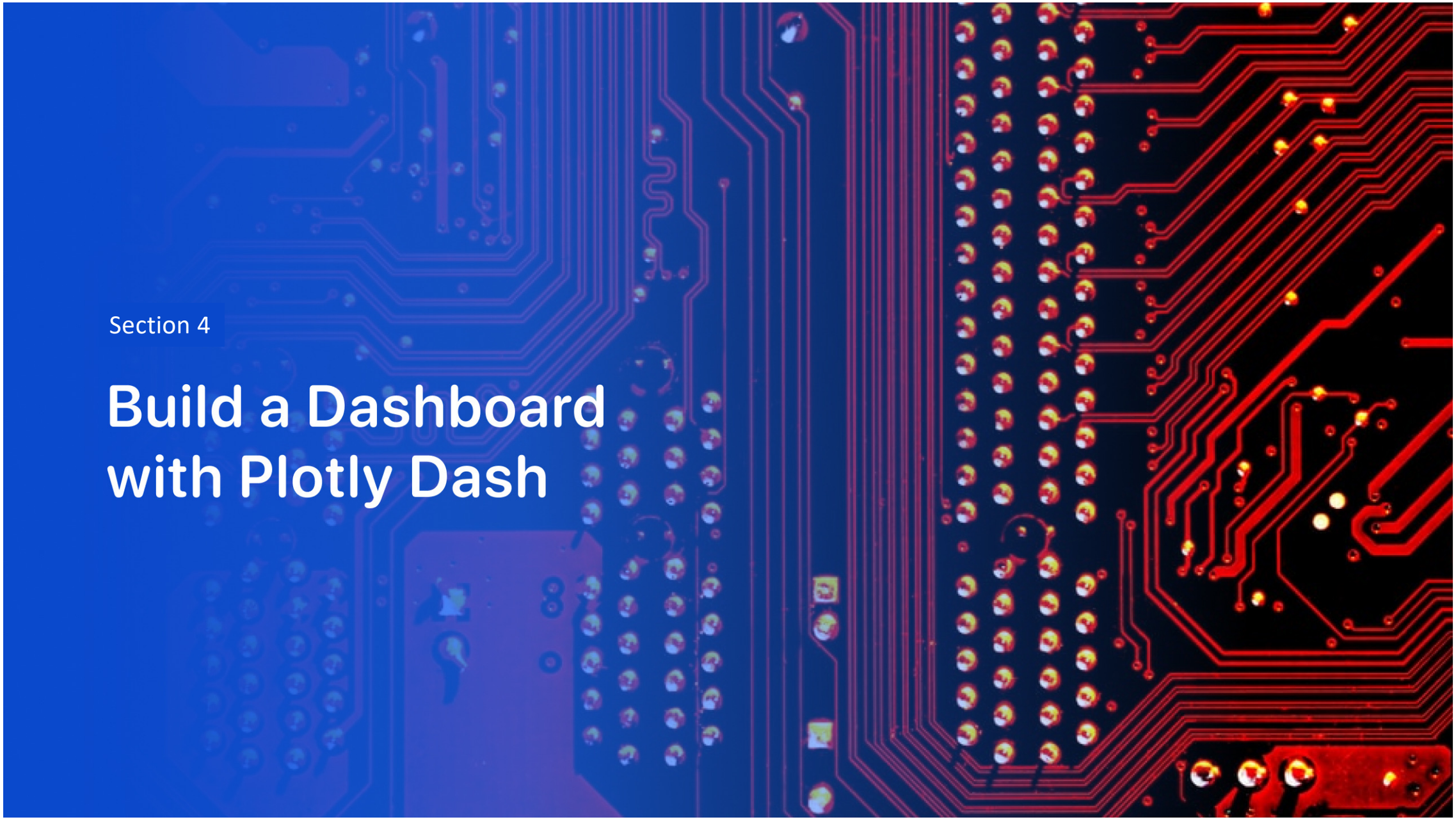
- Example of KSC LC-39A launch site launch outcomes.
- Green markers indicate successful and red ones indicate failure.

Selected Launch Site and Its Proximate Coastline

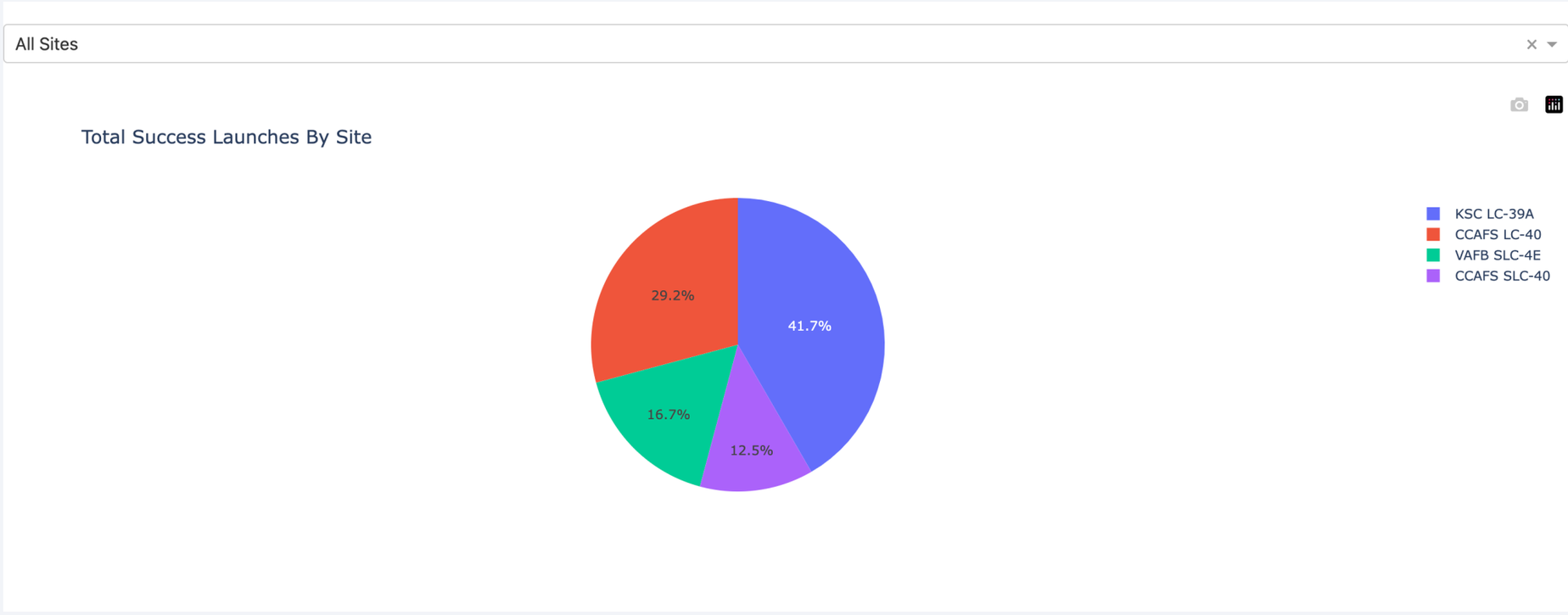
The closest coastline to KSC LC-39A is 3.87 km away.

Section 4

Build a Dashboard with Plotly Dash

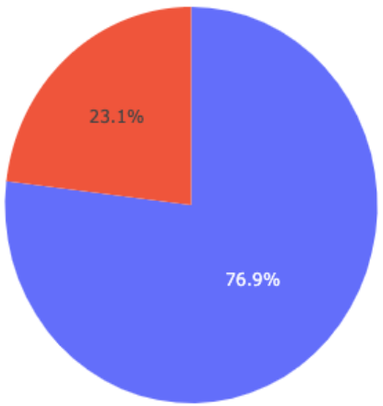


Success Count for All Sites



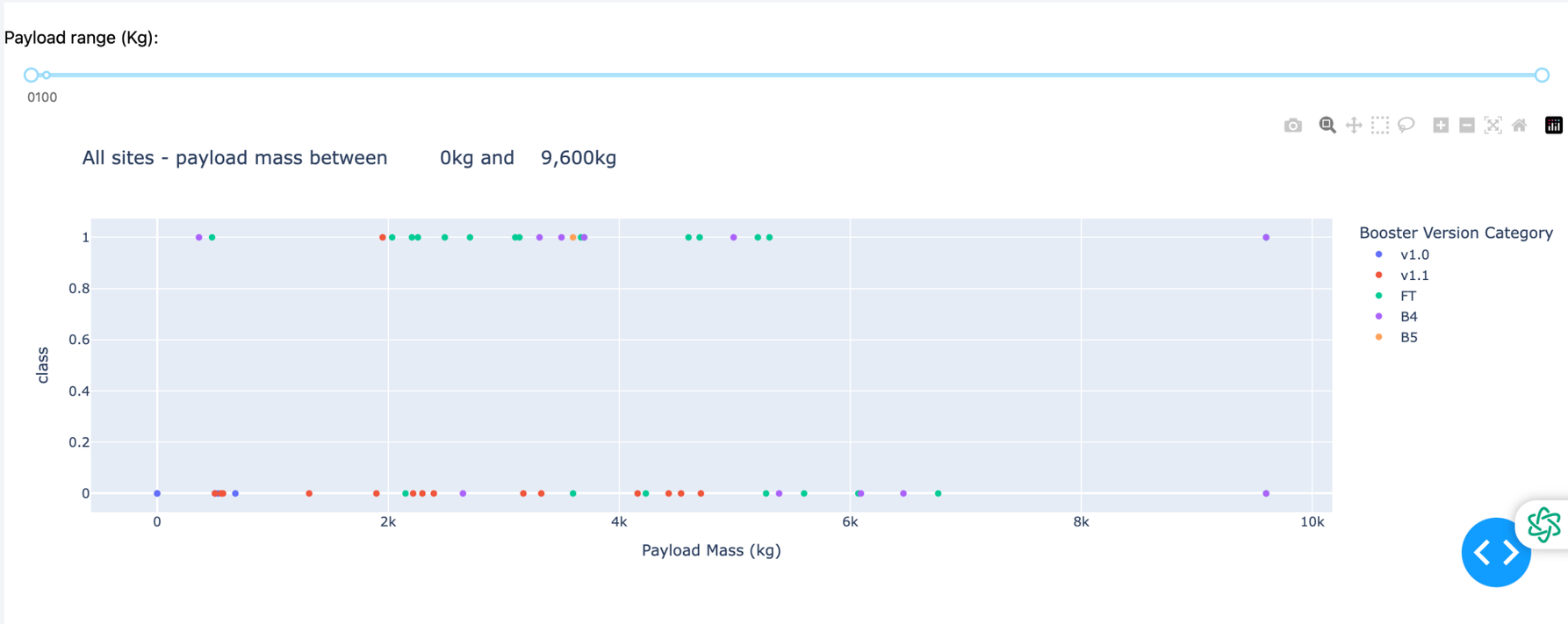
The place from where launches are done seems to be a very important factor of success of missions. The chart clearly shows that from all the sites, KSC LC-39A has the most successful launches.

Launch Site with the Highest Launch Success Ratio



- KSC LC-39A
- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



Payloads under 6,000kg and FT boosters are the most successful combination.

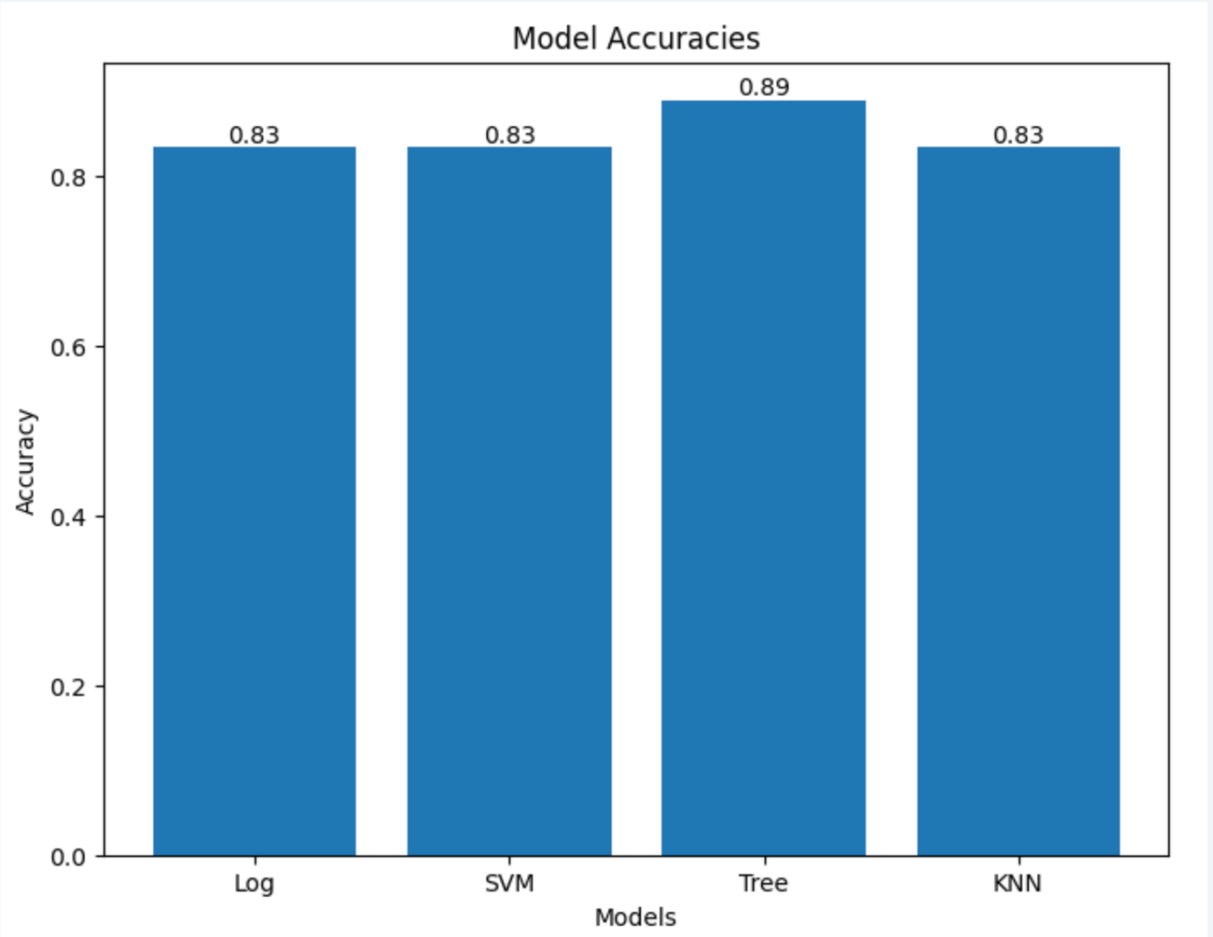
Section 5

Predictive Analysis (Classification)

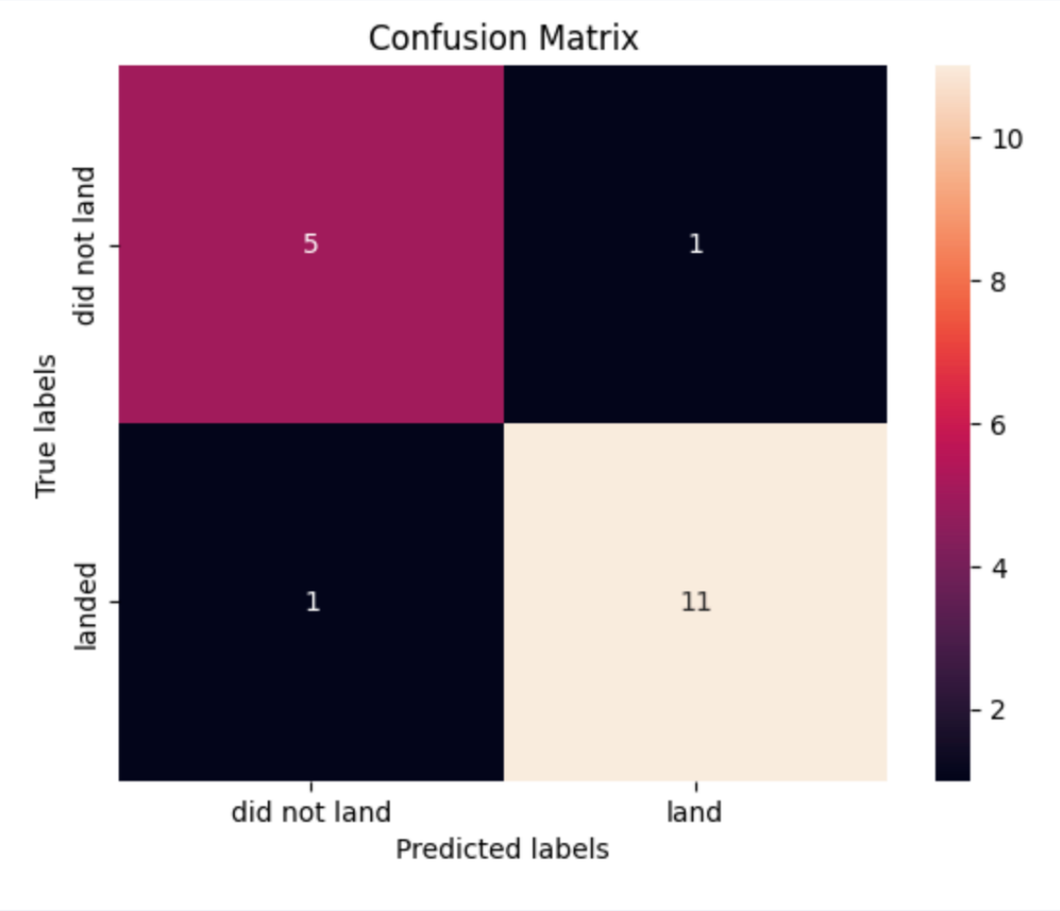
Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside.

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracy around 89%.



Confusion Matrix – Decision Tree



- Confusion matrix of Decision Tree Classifier proves its accuracy by showing a high number of true positive and true negatives compared to the false ones.

Conclusions

- The launch site location appears to be a very important factor in mission success.
- All launch sites are located near coastlines, likely for safety reasons.
- KSC LC-39A has the highest ratio of successful launches at 76.9%.
- Payloads under 6,000 kg and Falcon 9 FT boosters have the highest success rate.
- Of the four classification models tested, the Decision Tree Classifier performed best with around 89% accuracy. The confusion matrix validates its strong predictive power.

Appendix

- Project GitHub: https://github.com/raushan-zhandayeva/Data-Science-Capstone

Thank you!