IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Aaron Zeitlin
09.08.2023

Outline

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

Executive Summary

In response to SpaceX's cost-effective Falcon 9 launches and the importance of first stage landings, we've initiated a data-driven project.

Our aim is to predict successful landings using a machine learning pipeline fueled by publicly available SpaceX launch data. This project addresses crucial questions about landing success and optimization opportunities.

We've employed diverse methodologies, including data collection via API and web scraping, data wrangling, SQL-based exploratory data analysis, and interactive visual analytics.

The project results offer a comprehensive view, spanning exploratory data analysis, interactive analytics, and predictive analytics, positioning us to make data-driven decisions and maintain our industry leadership.

- Summary of methodologies
  - Data Collection with API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Interactive Dashboard
  - Machine Learning Predictions
- Summary of all results
  - Exploratory Data Analysis
  - Interactive Analytics
  - Predictive Analytics

Introduction

- Project background and context

  SpaceX offers Falcon 9 launches for $62 million, a significant cost advantage over competitors, largely due to the first stage's reusability. We aim to predict first stage landings to estimate launch costs accurately. This will be achieved by building a machine learning pipeline using publicly available data on SpaceX launches.

- Problems you want to find answers

  What determines if the first stage rocket will land successfully?

  Can we predict whether the Falcon 9 first stage will land successfully?

  How can we optimize operations to maximize successful landings?

Section 1

Methodology

Methodology

Executive Summary

- Perform Data collection
- Perform data wrangling
- 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

via API [https://api.spacexdata.com/v4/launches/past]

via Web Scraping

[https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches]

Data Collection – SpaceX API

Data was collected from a public API published by SpaceX

[https://api.spacexdata.com/v4/launches/past]

Request and parse the SpaceX launch data using the GET request

Filter the dataframe to only include `Falcon 9` launches

Deal with Missing Values

[https://github.com/aaronzeitlin/Applied-Data-Science-Capstone/blob/master/Week1_2_data-collection-with-web-scraping.ipynb]

Data Collection – Scraping

Data was collected by scraping the Wikipedia article '*List of Falcon 9 and Falcon Heavy launches*'

[https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches]

Request the Falcon9 Launch Wiki page from its URL

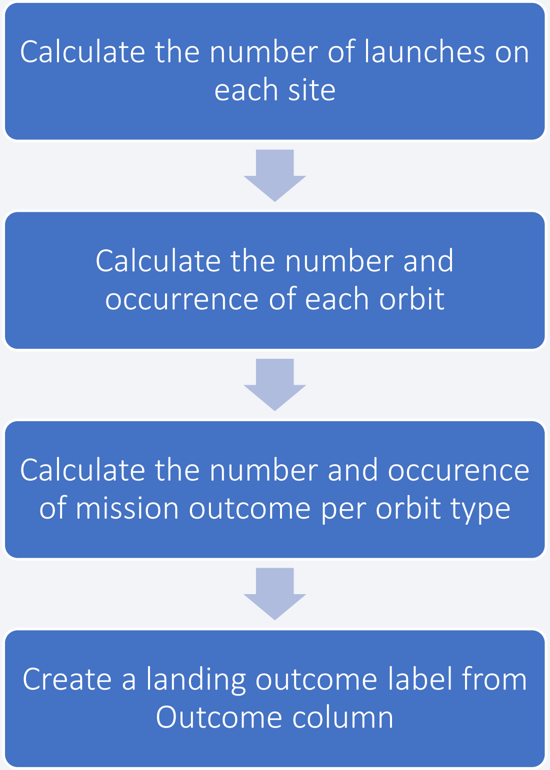
Extract all column/variable names from the HTML table header

Create a data frame by parsing the launch HTML tables

[https://github.com/aaronzeitlin/Applied-Data-Science-Capstone/blob/master/Week1_1_data-collection-with-API.ipynb]

Data Wrangling

Calculated launch site frequencies, orbit occurrences, and mission outcomes by orbit type to introduce a landing outcome label based on the Outcome column for streamlined analysis.



[https://github.com/aaronzeitlin/Applied-Data-Science-Capstone/blob/master/Week1_3_data-wrangling.ipynb]

EDA with SQL

- Display the names of the unique launch sites in the space mission
  - select distinct(Launch_site) from SPACEXTBL
- Display 5 records where launch sites begin with the string 'CCA'
  - select distinct(Launch_site) from SPACEXTBL
- Display the total payload mass carried by boosters launched by NASA (CRS)
  - select sum(PAYLOAD_MASS__KG_) as 'Total Payload Mass launched by NASA (CRS)' from SPACEXTBL where customer is "NASA (CRS)"
- Display average payload mass carried by booster version F9 v1.1
  - select avg(PAYLOAD_MASS__KG_) as 'Average Payload Mass launched by F9 V1.1' from SPACEXTBL where Booster_Version is "F9 v1.1"
- List the date when the first successful landing outcome in ground pad was achieved.
  - select min(Date) as "First Successful Landing Outcome" from SPACEXTBL where Landing_Outcome like "Success%"
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - select Payload from SPACEXTBL where Landing_Outcome is "Success (drone ship)" and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000
- List the total number of successful and failure mission outcomes
  - select count(case when Mission_Outcome like 'Success' then 1 else null end) as "Successful Mission Outcomes", \
  - count(case when Mission_Outcome like 'Failure%' then 1 else null end) as "Unsuccessful Mission Outcomes" \
  - from SPACEXTBL
- List the names of the booster versions which have carried the maximum payload mass. Use a subquery
  - select Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spaceXTBL)
- List the records which will display the month names, failed landing outcomes in drone ship, booster versions, and launch sites for the months in year 2015.
  - select Date, substr(Date, 6, 2) as Month, substr(Date, 0, 5) as Year, Landing_Outcome, Booster_Version, Launch_Site \
  - from SPACEXTBL where substr(Date, 0, 5)='2015' and Landing_Outcome like "Failure%"
- 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.
  - select Landing_Outcome, count(*) as Count from SPACEXTBL \
  - where Date between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by Count desc;

[https://github.com/aaronzeitlin/Applied-Data-Science-Capstone/blob/master/Week2_1_EDA-with-SQL.ipynb]

EDA with Data Visualization

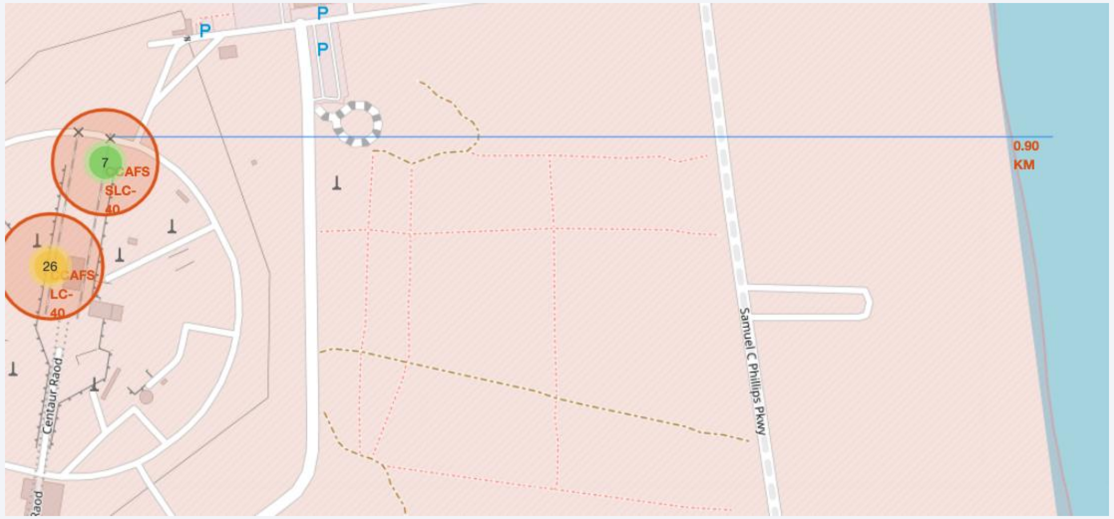
- Relationship between Flight Number and Launch Site
- Payload Mass by Launch Site
- Mean Success Rate by Orbity Type
- Relationship between Flight Number and Orbit Type
- Payload Mass by Orbit Type
- Success Rate by Year

[https://github.com/aaronzeitlin/Applied-Data-Science-Capstone/blob/master/Week2_2_EDL-with-Visualization.ipynb]

Build an Interactive Map with Folium

Objects Added

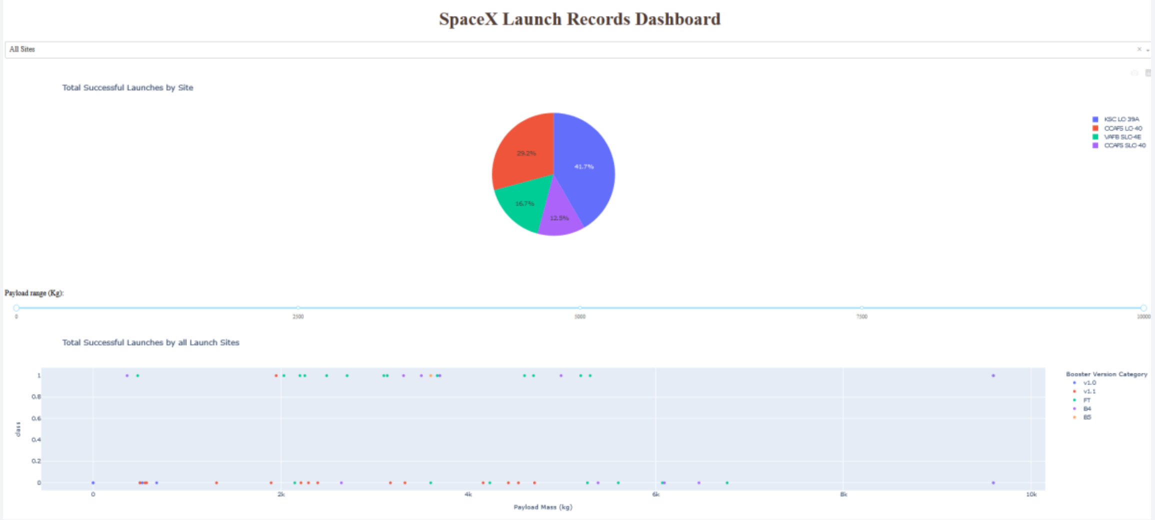
- Launch Sites
- Success / Failed Launches
- Proximities
    - Coastline
    - Cities



[https://github.com/aaronzeitlin/Applied-Data-Science-Capstone/blob/master/Week3_1_Interactive-Visualization-with-Folium.ipynb]

Build a Dashboard with Plotly Dash

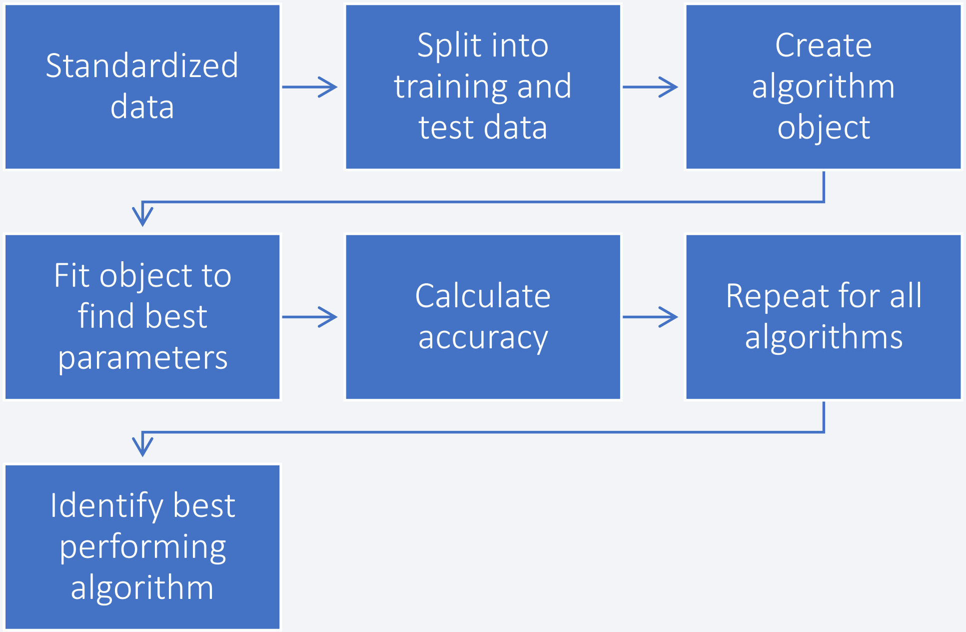
- Plots / Graphs
  - Pie Chart to show the total successful launch count for all sites as well as specific sites.
  - Scatter Plot to show the correlation between payload and launch success.
- Interactions / Filtering
  - Launch Site
  - Payload Mass



[https://github.com/aaronzeitlin/Applied-Data-Science-Capstone/blob/master/Week3_2_spacex_dash_app.py]

Predictive Analysis (Classification)

- Data was standardized before being split into training and test data to be used with four different machine learning algorithms:
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree Classifier
    - K Nearest Neighbor



[https://github.com/aaronzeitlin/Applied-Data-Science-Capstone/blob/master/Week4_1_Machine-Learning-Prediction.ipynb]

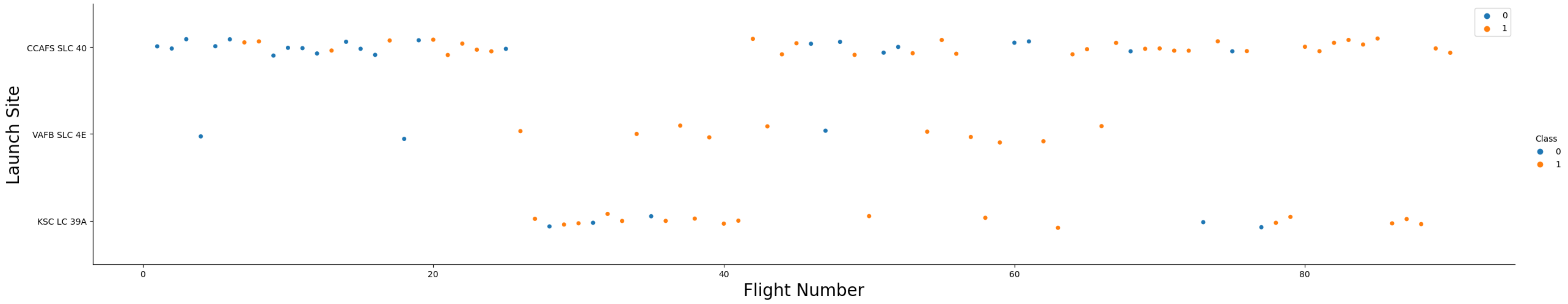
Results

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

Section 2

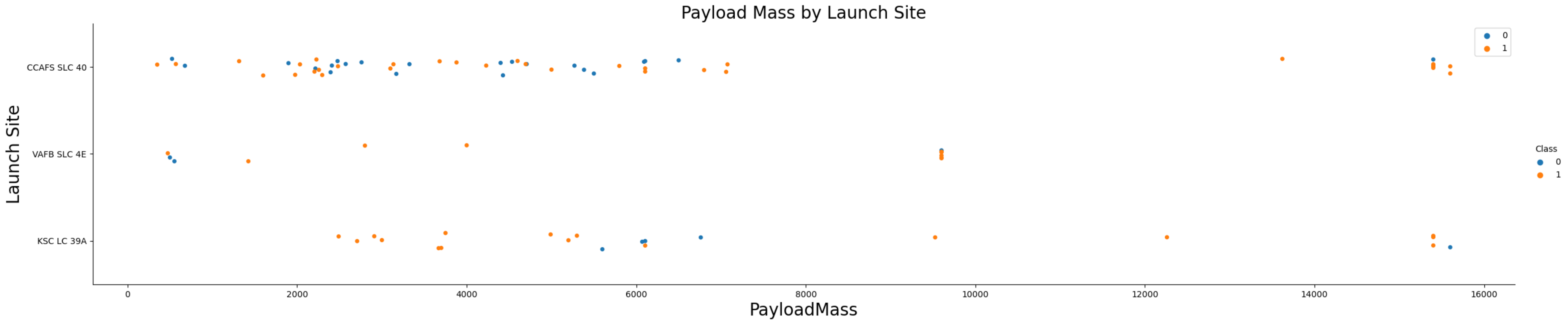
Insights drawn from EDA

Flight Number vs. Launch Site



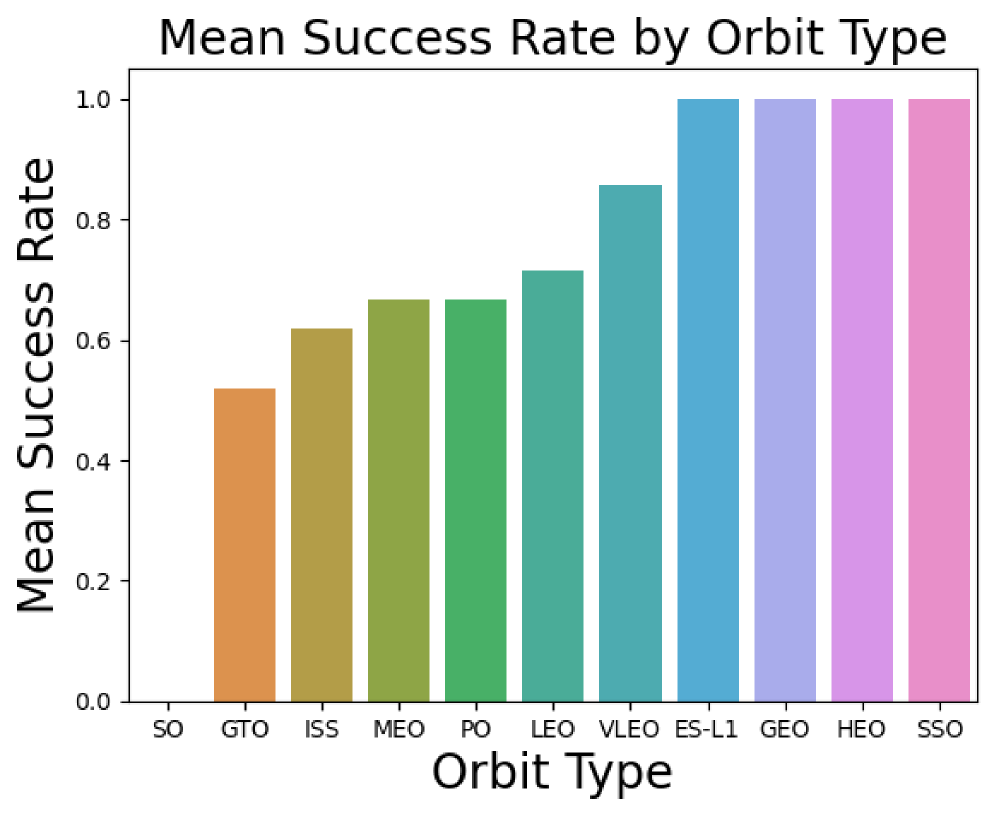
The VAFB SLC 4E launch site has the least number of flights.

Payload vs. Launch Site



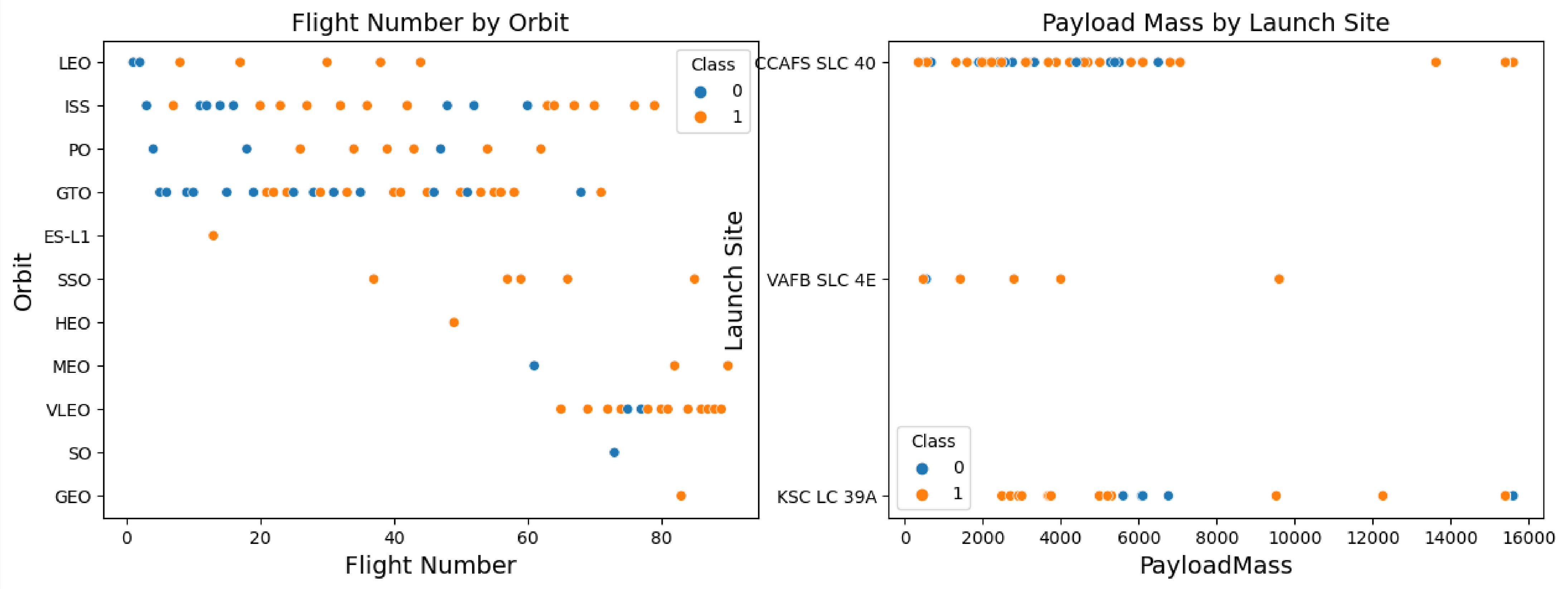
For the VAFB-SLC launch site there are no rockets launched for a payload mass greater than 10000.

Success Rate vs. Orbit Type



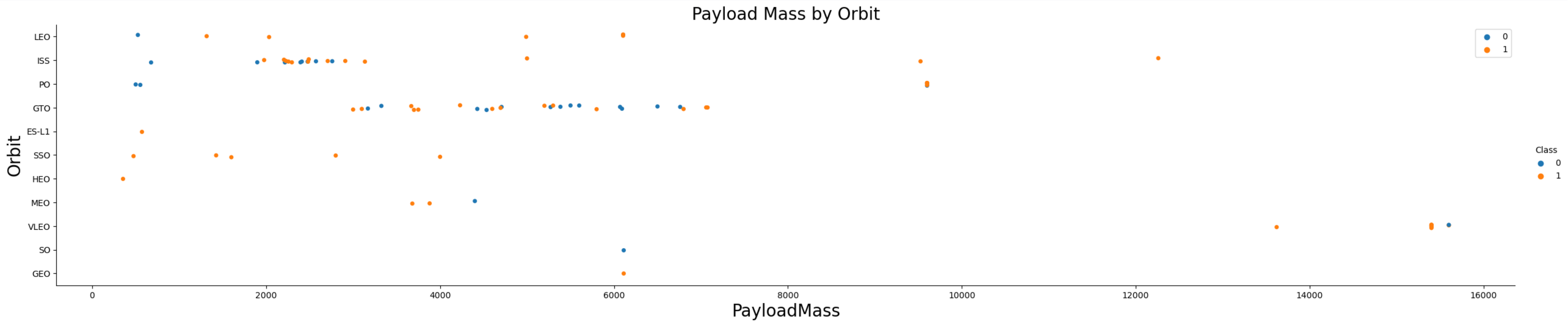
The ES-L1, GEO, HEO, and SSO orbit types are the most successful.

Flight Number vs. Orbit Type



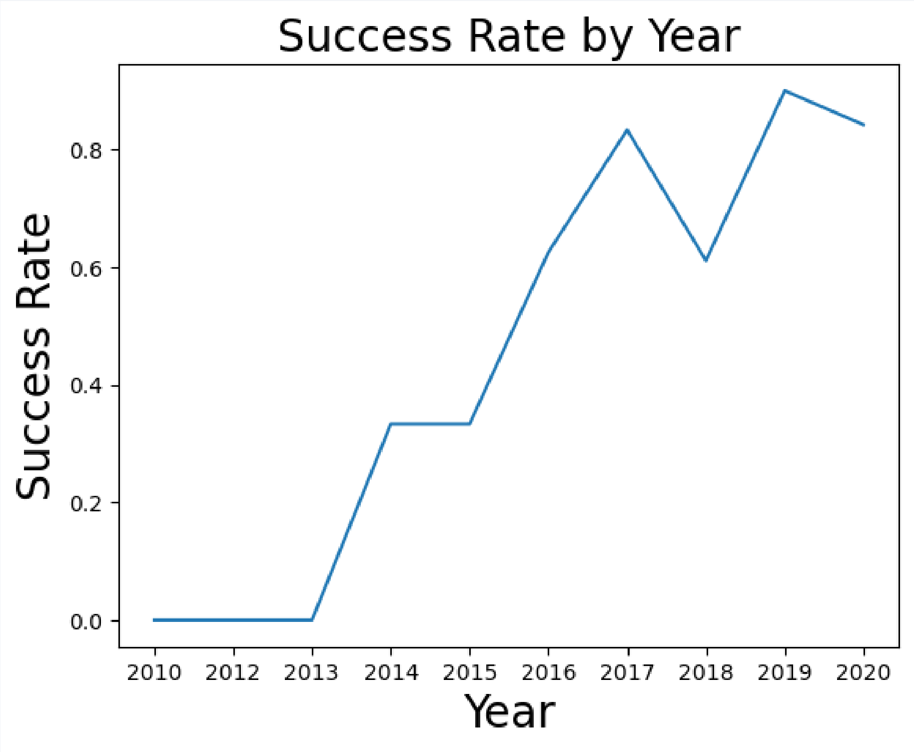
There is no strong relationship between Flight Number and Orbit Type.

Payload vs. Orbit Type



There is no strong relationship between Payload Mass and Orbit Type.

Launch Success Yearly Trend



The success rate increases beginning in 2013.

All Launch Site Names

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

There are four launch sites.

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Total Payload Mass launched by NASA (CRS)
45596

Average Payload Mass by F9 v1.1

Average Payload Mass launched by F9 V1.1
2928.4

First Successful Ground Landing Date

First Successful Landing Outcome
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

Successful Mission Outcomes	Unsuccessful Mission Outcomes
98	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

Date	Month	Year	Landing_Outcome	Booster_Version	Launch_Site
2015-10-01	10	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

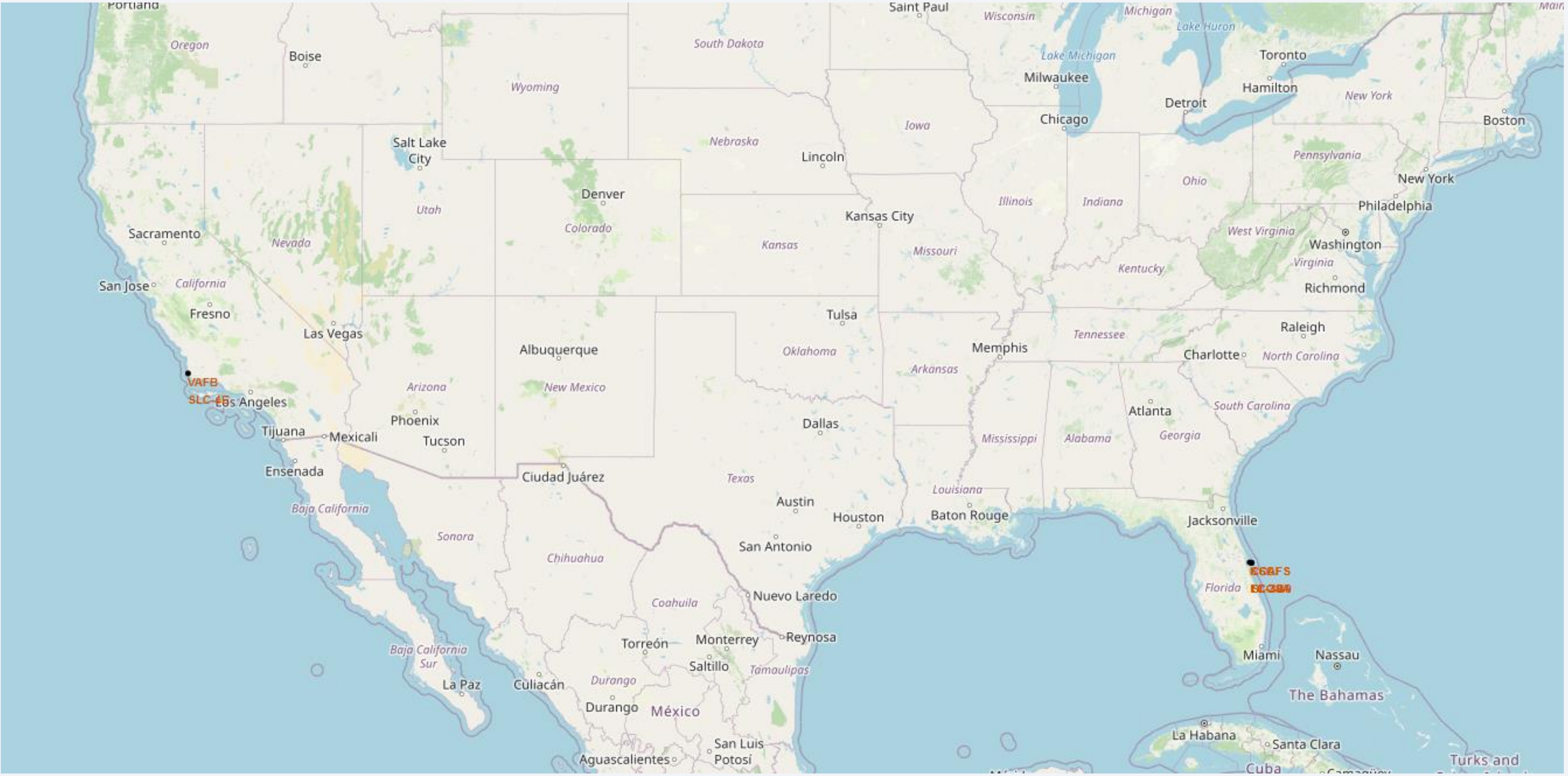
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

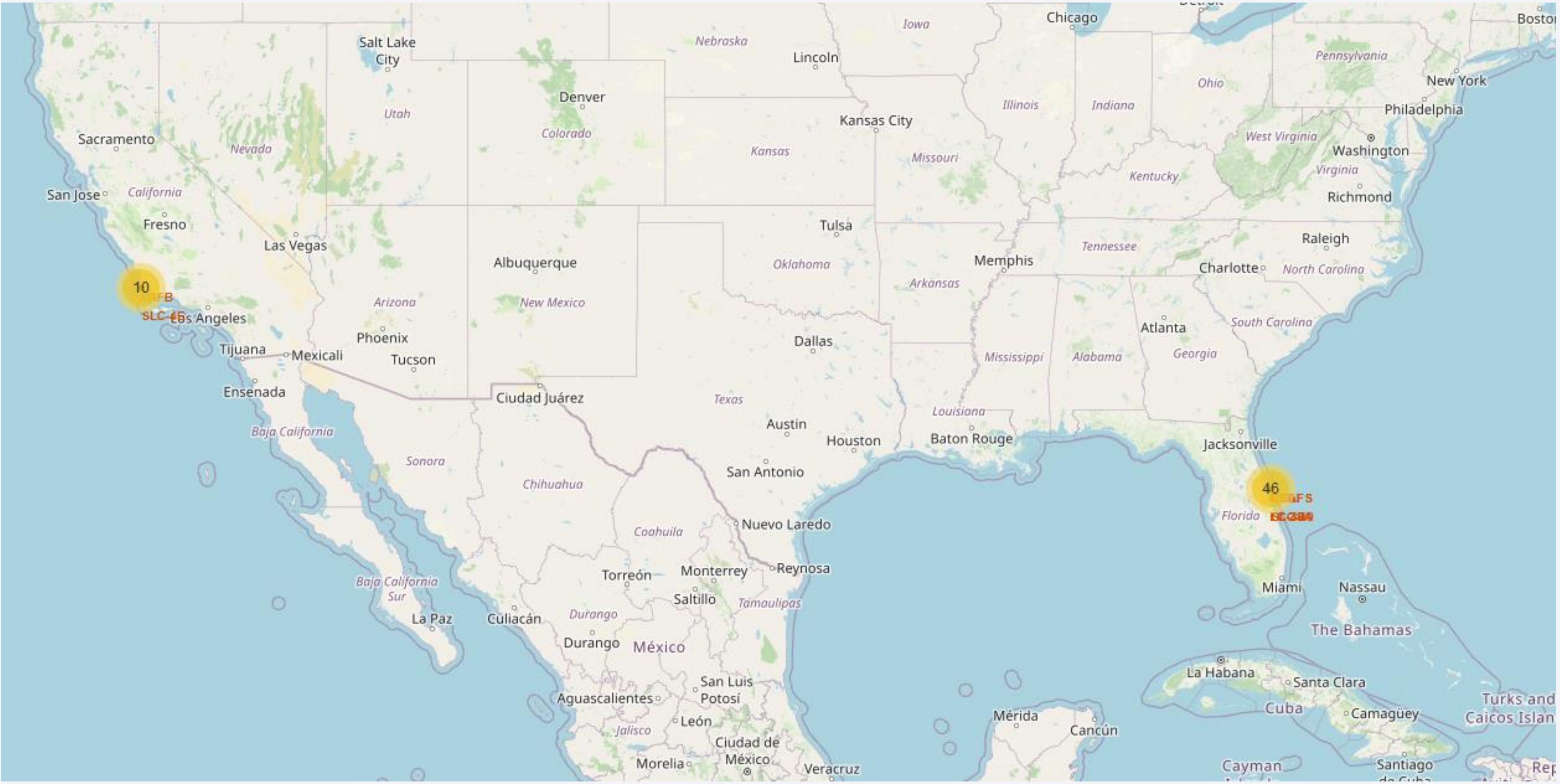
Section 3

Launch Sites Proximities Analysis

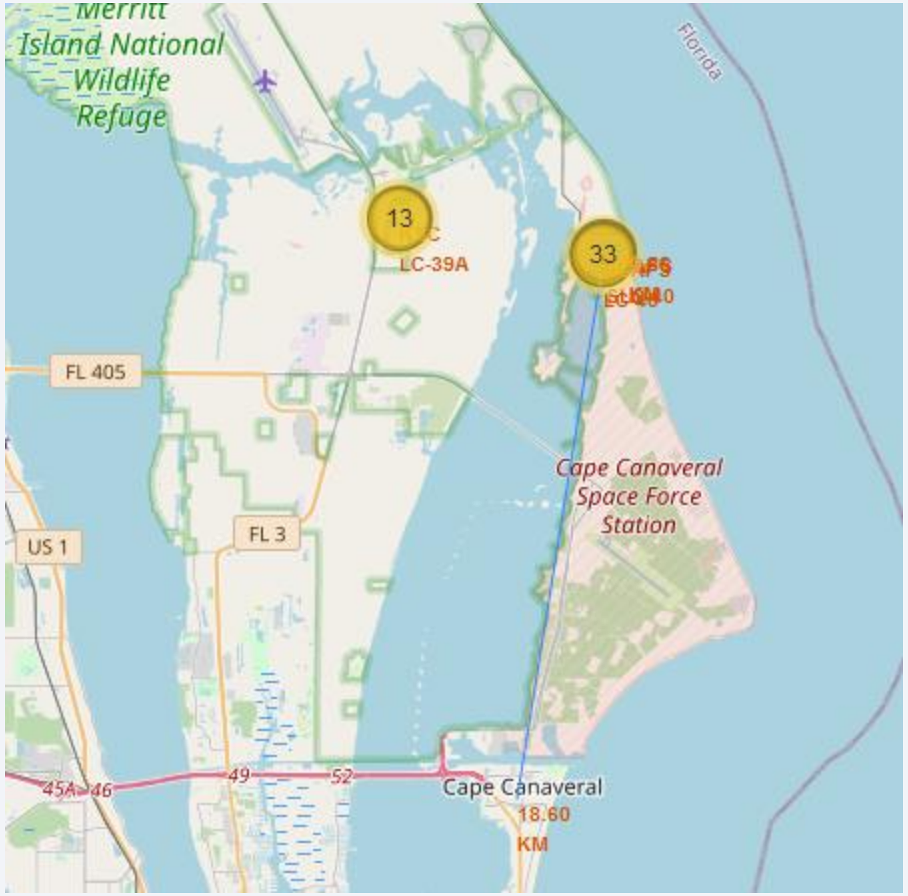
Interactive Map – Launch Sites



Interactive Map – Successful/Failed Launches



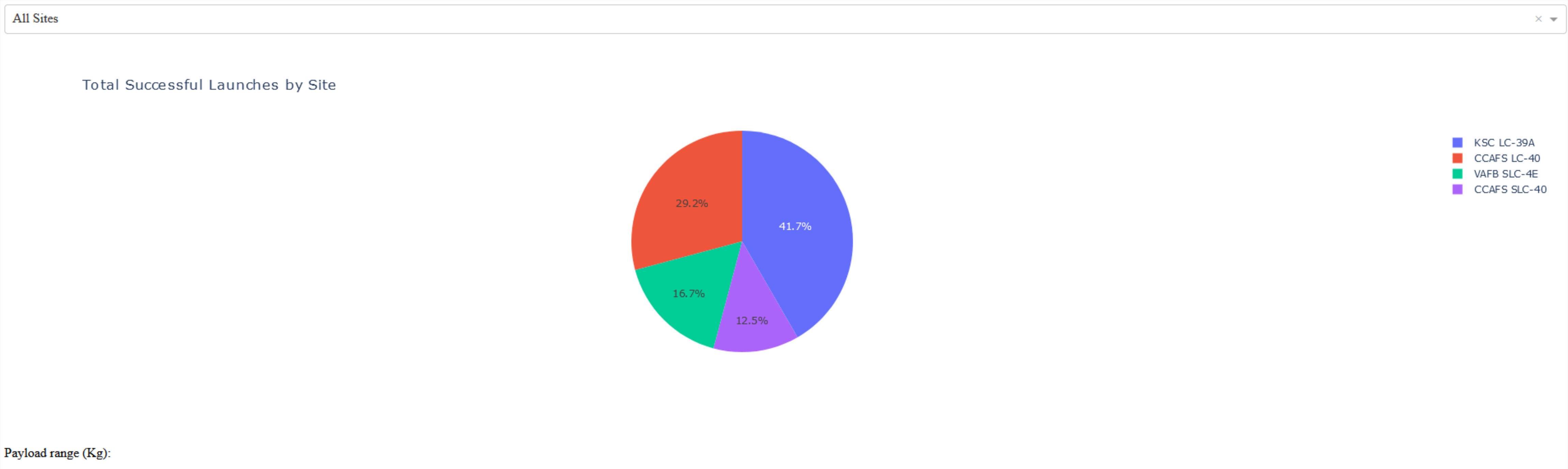
Interactive Map – Proximity to Coastline and City



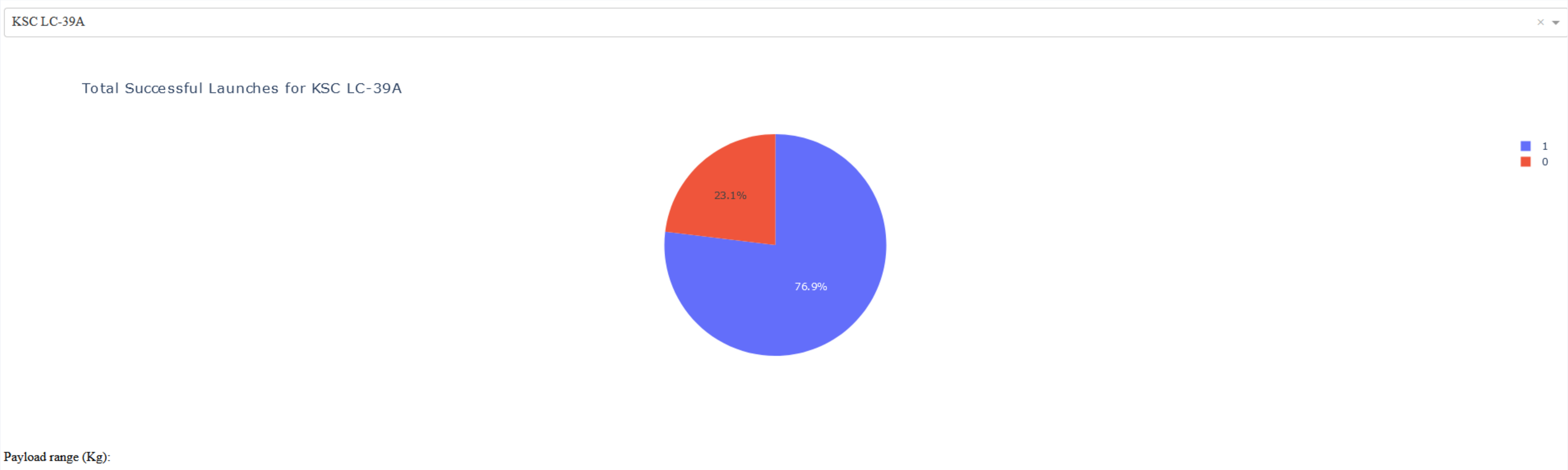
Section 4

Build a Dashboard with Plotly Dash

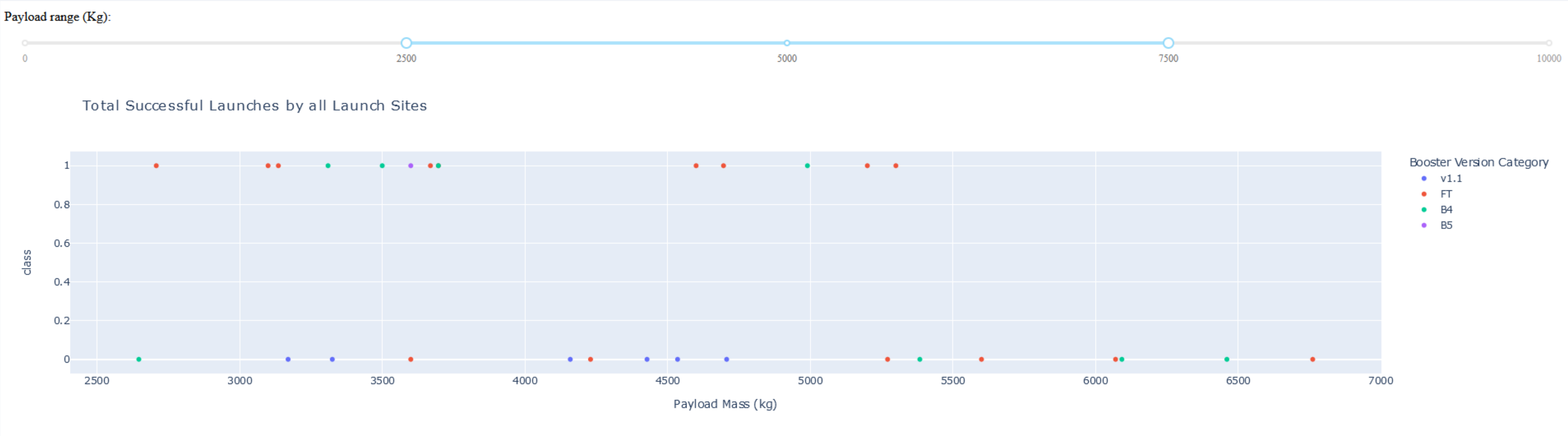
Dashboard – Success Ratio (All Sites)



Dashboard – Success Ratio (KSC LC-39A)



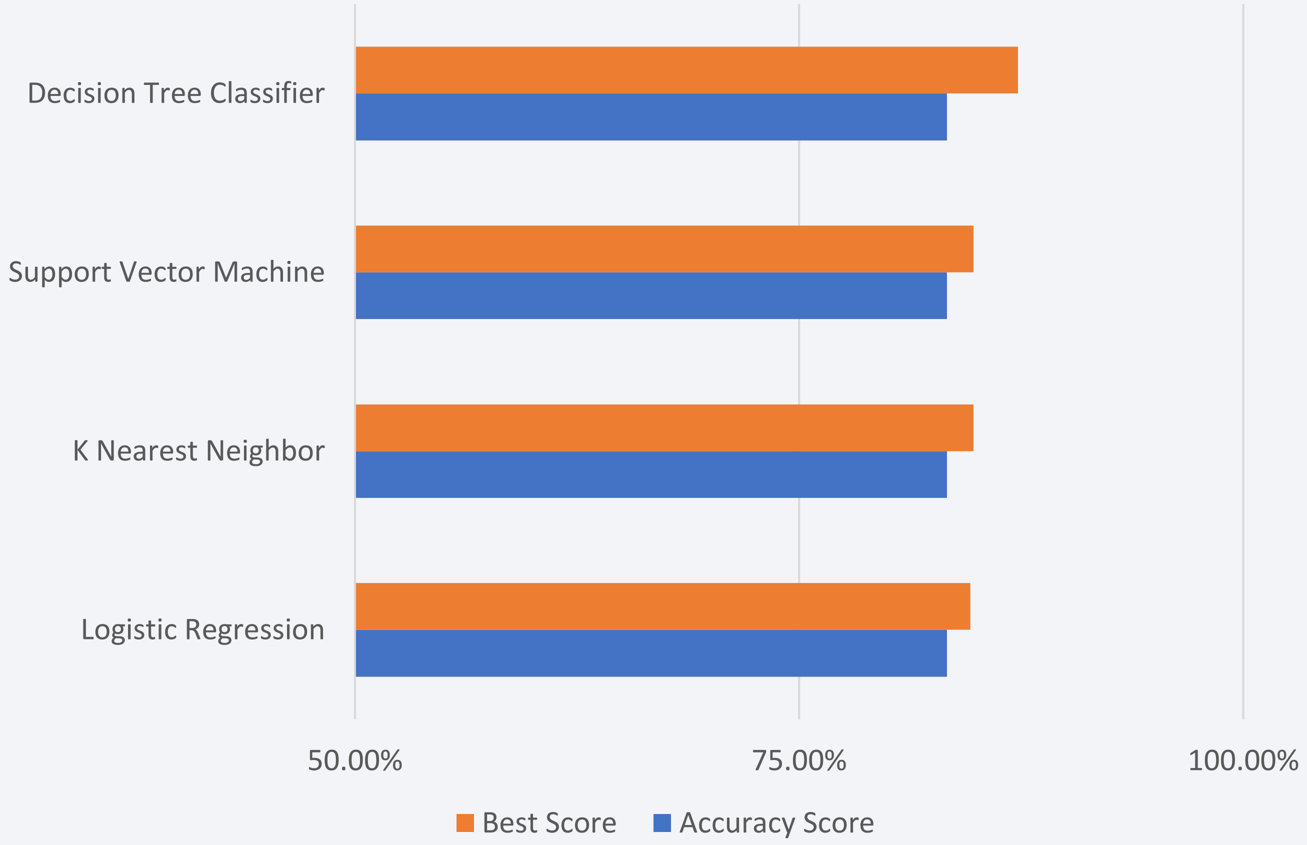
Dashboard – Payload vs. Launch Outcomes



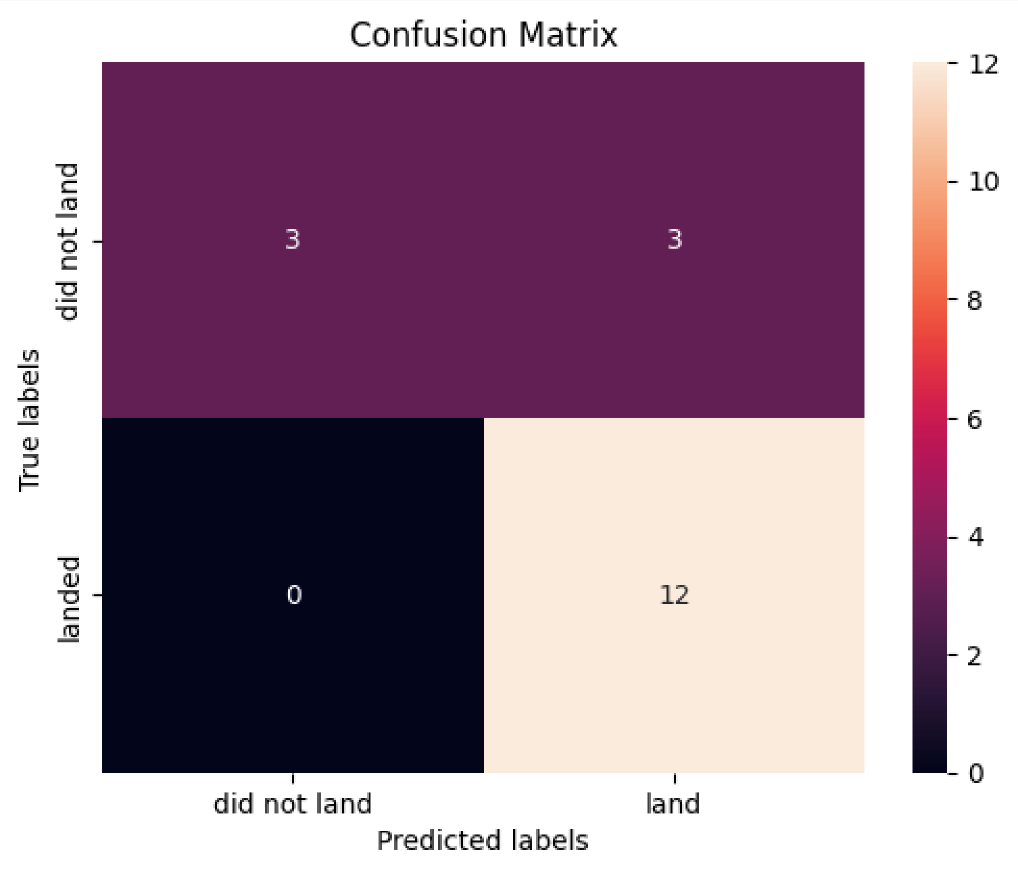
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- The success rate of Falcon 9 first stage landing increases beginning in 2013
- A Decision Tree Classifier is the best performing machine learning algorithm for predicting a successful Falcon 9 first stage landing
- Launch Sites are near coastlines

Appendix

- n/a

Thank you!