IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

Summary of methodologies

- Methodologies used to analyze data were:
  - Web scraping and SpaceX API
  - EDA including data wrangling, data visualization and interactive visual analytics
  - Simple Machine Learning Predictions

Summary of all results

- EDA allowed the ability to identify which features are the best to predict success of launchings
- Machine Learning Predictions showed the best model to predict which characteristics are important to deliver successful launches based on the collected data

Introduction

The project objective was to evaluate the opportunity to compete with Space X with Space Y’s competitive viability based on current Space X data

This was done through doing analysis on figuring out the best estimate for the total cost for launches by predicting success landings of first stage rockets as well as where is the best place to make launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology 2 sources:
  - Space X API (https://api.spacexdata.com/v4/rockets)
  - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- 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 was collected until this step was normalized, dived in trying and test data sets and evaluated by four different models. The accuracy of each model evaluated using different combinations of parameters

Data Collection

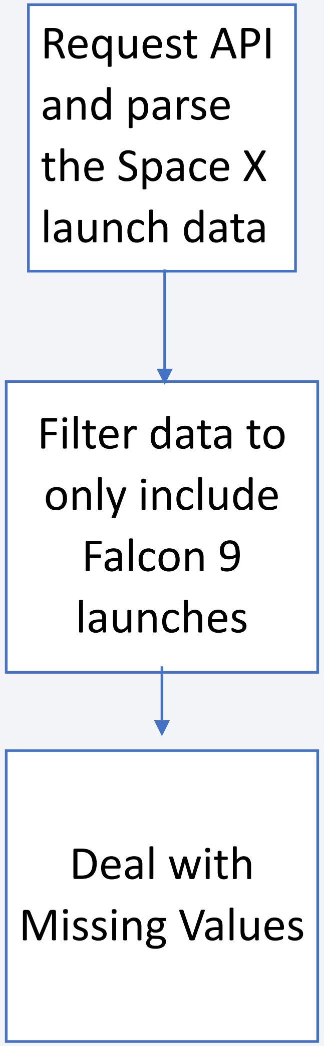
- Data sets were collected from Space X API and from Wikipedia using web scrapping techniques .

Data Collection – SpaceX API

- Space X provides public API from where data can be obtained and then used
- This API was used accordingly to gather data

- Source Code:

https://github.com/bossagawa/CapstoneDataScience/blob/main/Done%20jupyter-labs-spacex-data-collection-api.ipynb

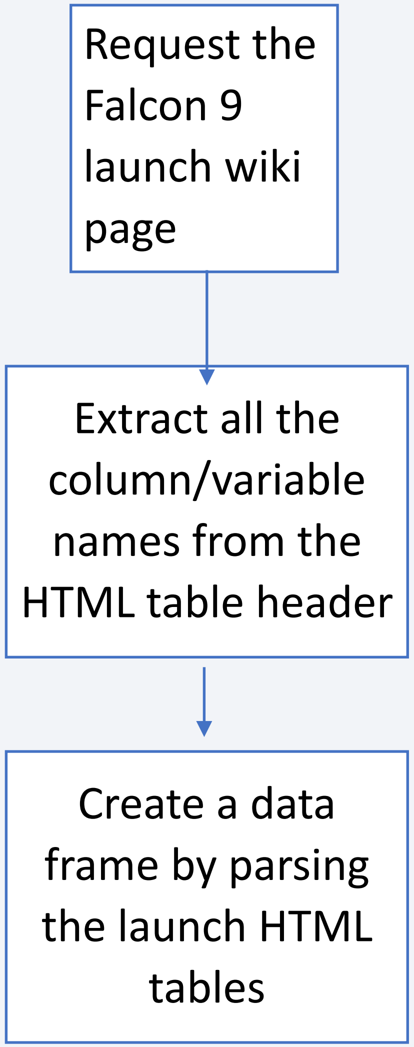


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Source Code:

https://github.com/bossagawa/CapstoneDataScience/blob/main/Done%20jupyter-labs-webscraping.ipynb



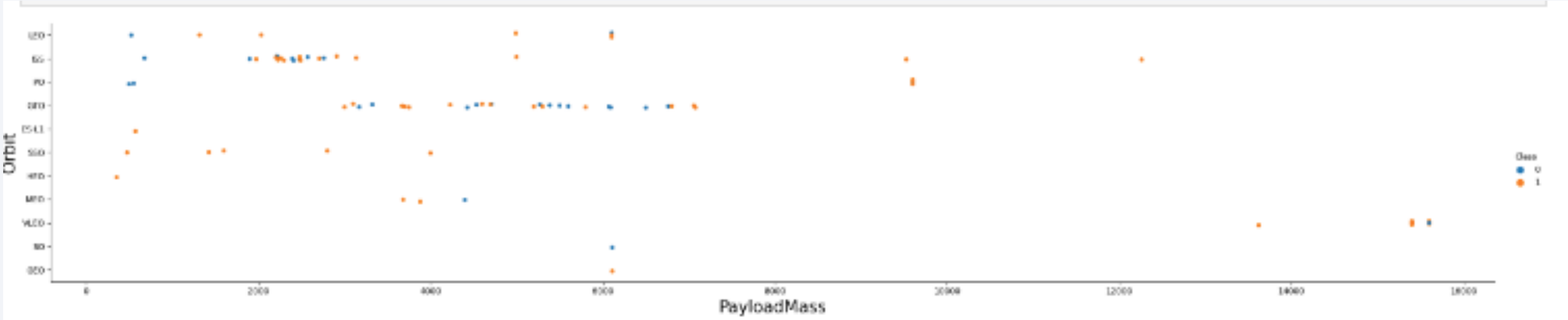
Data Wrangling

- EDA was performed on the data set,
- The summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- The landing outcome label was created from outcome column
- Source Code:

https://github.com/bossagawa/CapstoneDataScience/blob/main/Done%20labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- To explore data, scatterplots and bar plots were used to visualize the relationship between pair of features
- Compared PayloadMass vs orbits to reveal the relationship between payloads and orbit types



Source Code:https://github.com/bossagawa/CapstoneDataScience/blob/main/Done%20jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- The following SQL queries were done:
    - Names of unique launch sites in the space mission
    - Top 5 launch sites whose name begin with ‘CCA’
    - Total payload mass carried by boosters launched by NASA
    - Average payload mass carried by booster version F9 vs 1.1
    - Date when the first successful landing outcome in ground pad was achieved
    - Names of boosters with success in drone ship and have payload mass between 4000 and 6000 kg
    - Total number of successful and failure. mission outcomes
    - Names of the booster versions which have carried the max payload mass
    - Failed Landing outcomes in drone ship, their booster version and launch site names for year 2015
    - Rank of count of landing outcomes between 2010/06/04 to 2017/03/20 based on such as success and failures

Source Code:https://github.com/tflores/applied-data-science-capstone/blob/master/capstone-story-tiagoflores.pdf

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium maps
  - Markers indicate launch sites
  - Circles indicate high lighted areas around specific coordinates like Nasa
  - Lines are used to indicate distances between two coordinates

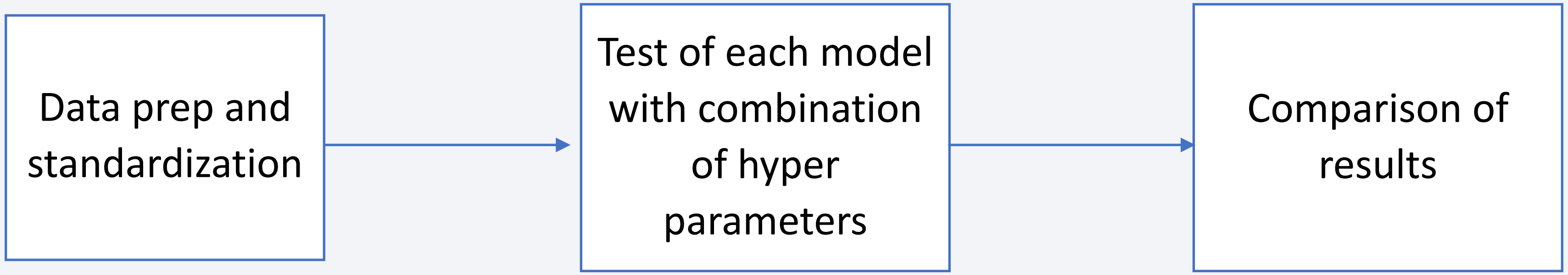
Source code:https://github.com/bossagawa/CapstoneDataScience/blob/main/Done%20IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- The following graphs and plots were use to visualize data
  - Percentage of launches by site
  - Payload range
- This allowed to quickly analyze the relation between payloads and launch sites to help identify where is the best place to launch according to payloads
- Source Code: https://github.com/bossagawa/CapstoneDataScience/blob/main/Plotly_app.py

Predictive Analysis (Classification)

- Four models were compared using: Logistic regression, support vector machine, decision tree and k nearest neighbor
- Source Code:https://github.com/bossagawa/CapstoneDataScience/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite%20(1).ipynb

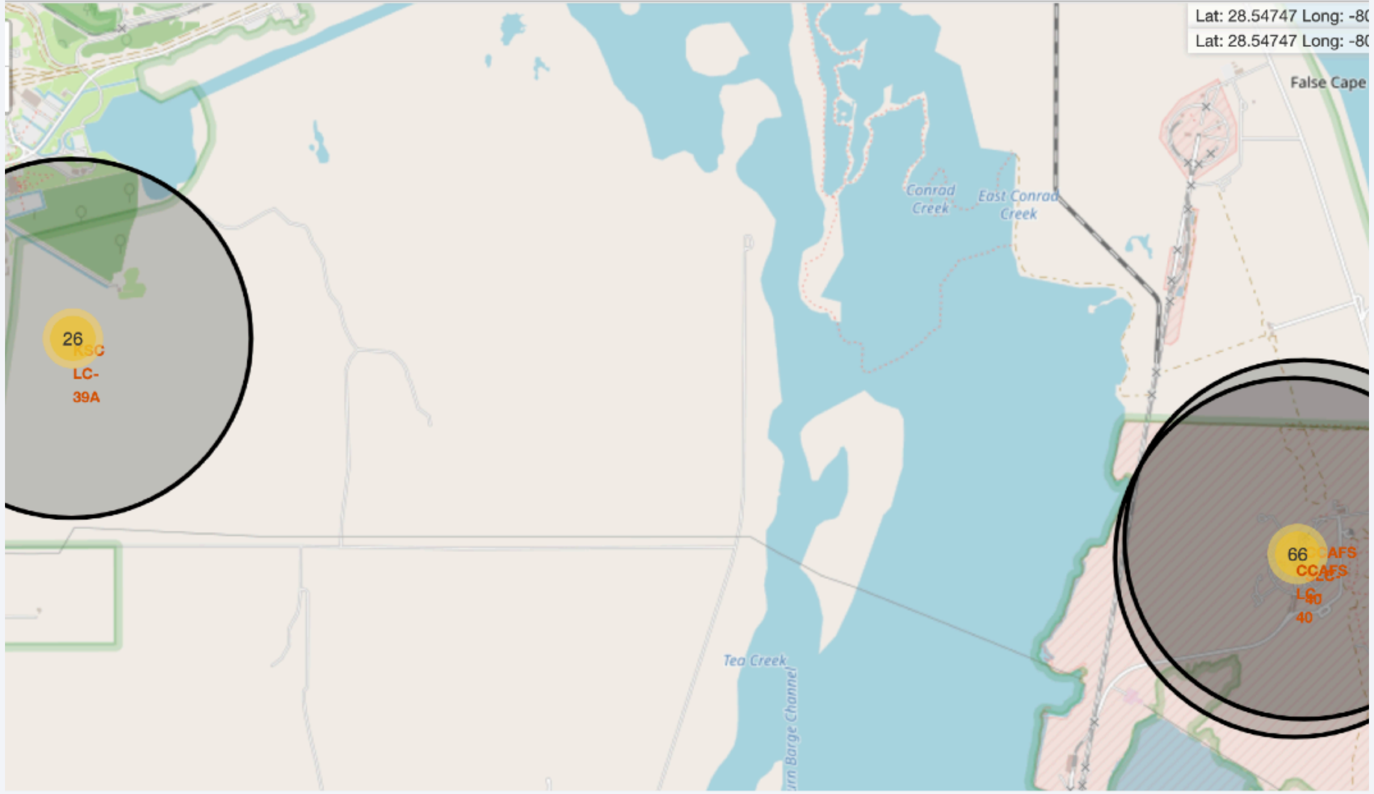


Results

- Exploratory data analysis results
  - Space X uses 4 different launch sites
  - The first launches were done to Space X itself and NASA
  - Average payload for F9 V1.1 booster is 2,928kg
  - The first success landing outcomes happen in 2015 five years after the first launch
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average
  - Almost 100% of mission outcomes were successful
  - Two booster versions failed at alining in done ships in 2015 F9 V1.1 B1012 and F9 v1.1 B1015
  - The number of landing outcomes became better as years passed

Results

- Using interactive analytics was possible to identify launch sites use to be in safety places, near sea for example and have good logistic infrastructure
- Most launches were on the east coast launch sites



Results

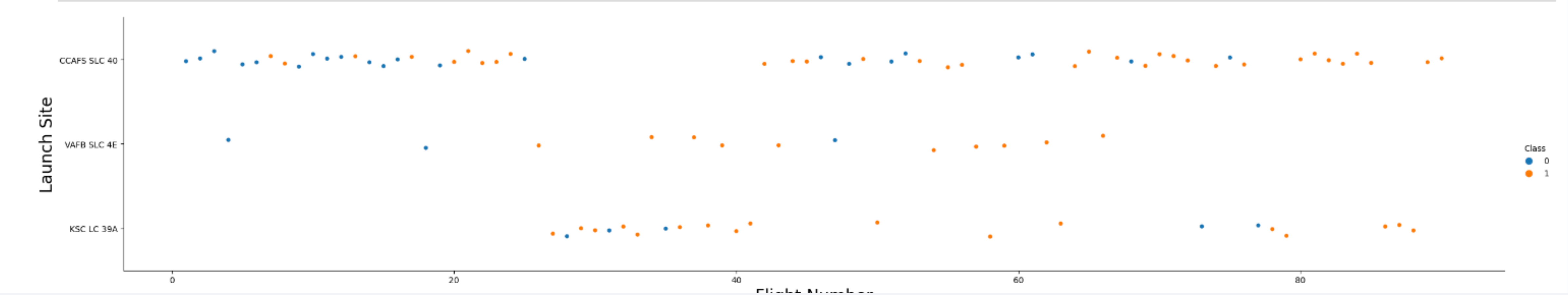
- Predictive Analysis showed the decision tree classifier is the best model to predict success landings having accuracy over 97% and accuracy for test data over 94%

Section 2

Insights drawn from EDA

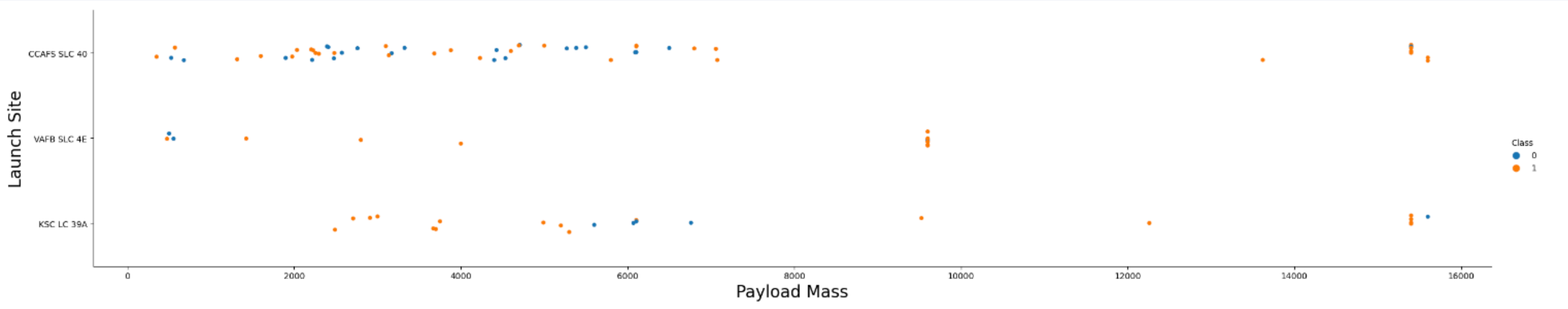
Flight Number vs. Launch Site

- Best launch site is CCAF5 SLC 40 where most were successful
- VAFB SLC 4E then KSC LC 39A were next best
- General success rate improved over time



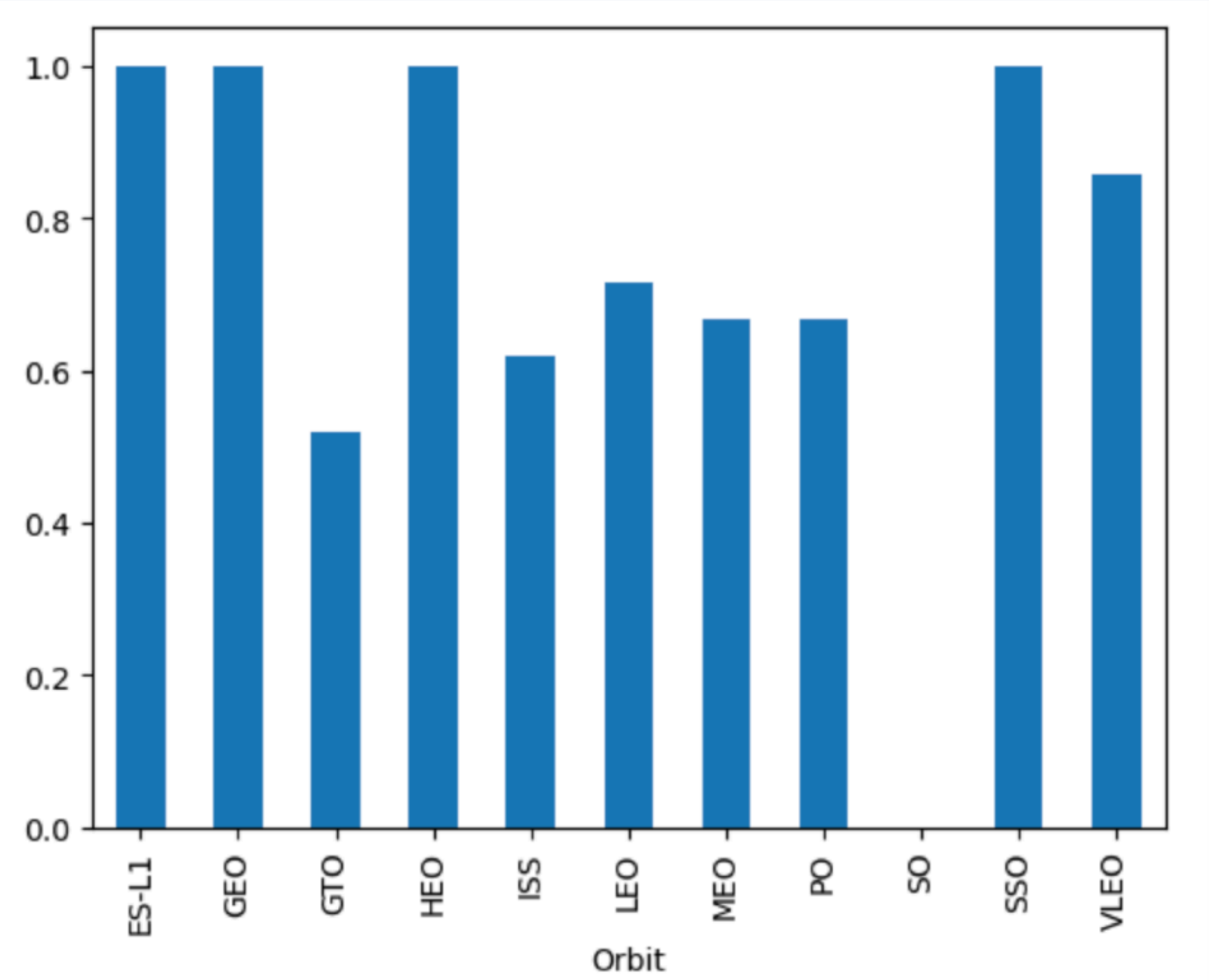
Payload vs. Launch Site

- Payloads over 9kg have excellent success rate
- Payloads over 12kg seem to be possible only for CCAAFS SLC 40 and KSC LC 39A launch sites



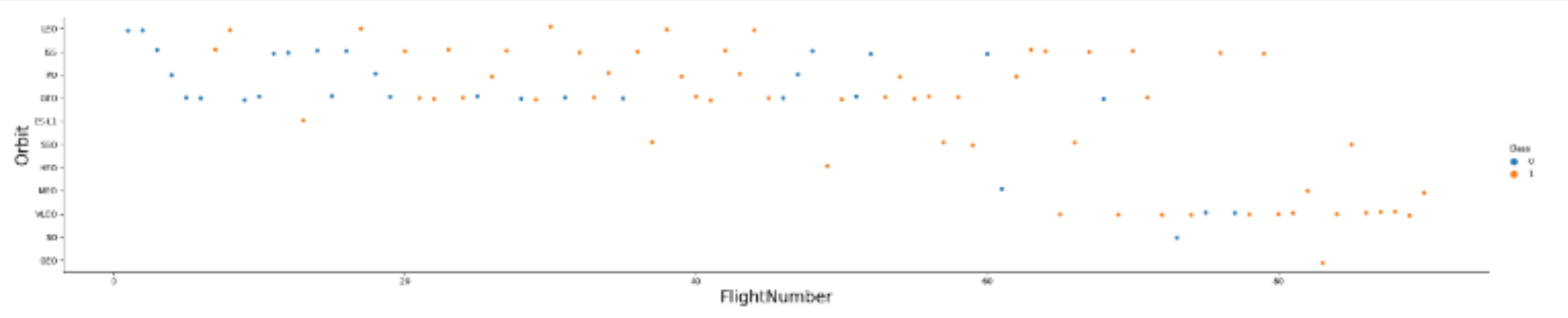
Success Rate vs. Orbit Type

- The best success rates happens to orbits
  - ES-L1
  - GEO
  - HEO
  - SSO
- Next best is
  - VLEO
  - LFO



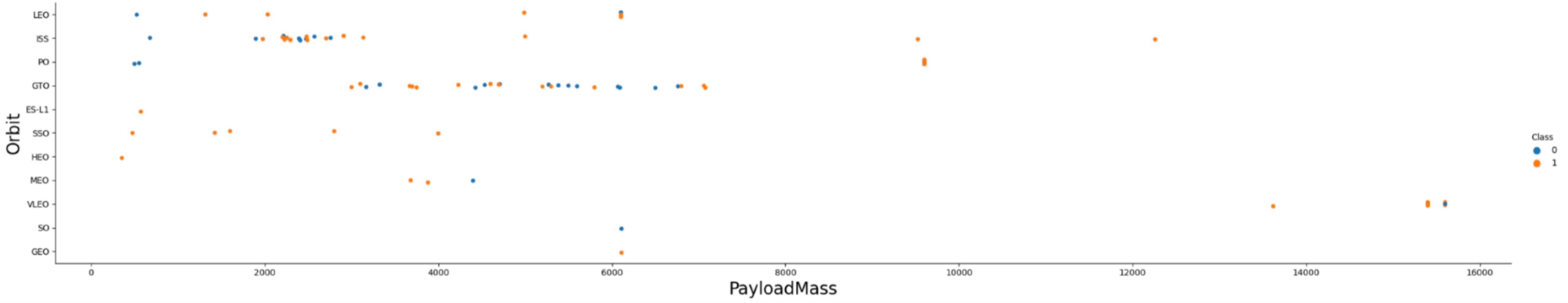
Flight Number vs. Orbit Type

- Success Rate improved over time on all orbits
- VLEO orbit seems a new opportunity due to recent increase in its frequency



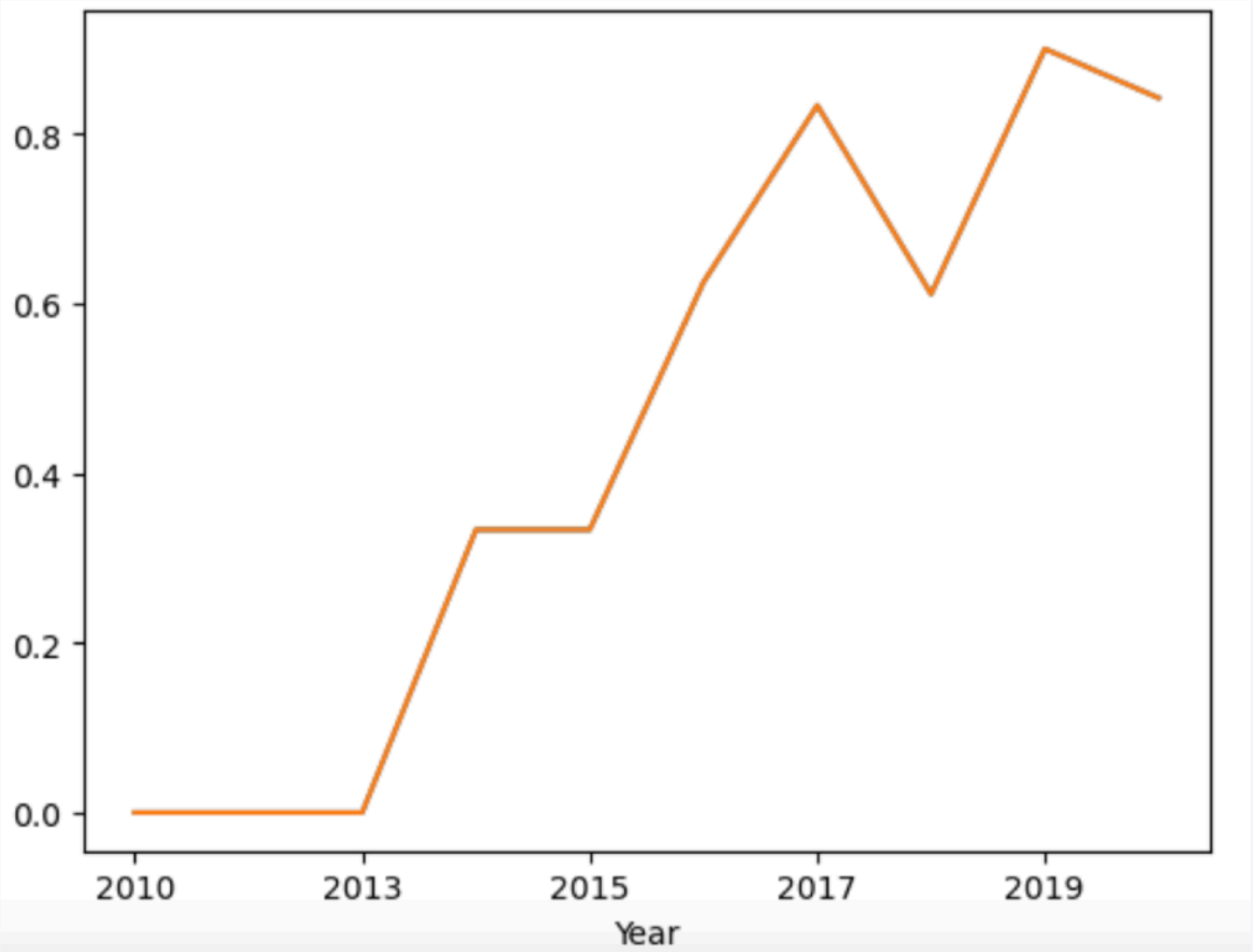
Payload vs. Orbit Type

- There is no relation between payload and success rate to orbit GTO
- ISS orbit has the widest range of payload and a good rate of success
- There are few launches to the orbits SO and GEO



Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until past 2019
- It seems that the first three years was a period of adjustments and improvement of technology



All Launch Site Names

The Launch Site Names are the following:

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

Obtained by selecting unique occurrences of “launch_site” values from the dataset.

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Here is five samples of Cape Canaveral launches

Total Payload Mass

- The total Payload Mass by Nasa was 111268 KG
- Total payload calculated by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4
- The average was calculated by filtering data by the booster version above and calculating the average payload mass

First Successful Ground Landing Date

- 01/08/2018 was the date of the first successful landing outcome on ground pad
- By filtering data by successful landing outcome on ground pad and getting the minimum value for date it's possible to identify the first occurrence

Successful Drone Ship Landing with Payload between 4000 and 6000

- The following list is the names of boosters which 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

- Selecting distinct booster versions according to the filters above, these 4 are the result.

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	QTY
None	898
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- Grouping mission outcomes and counting records for each group led us to the summary above.

Boosters Carried Maximum Payload

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

- These are the boosters which have carried the maximum payload mass registered in the dataset.

2015 Launch Records

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

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 which were found in the database

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

Landing_Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

- This is the Rank 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

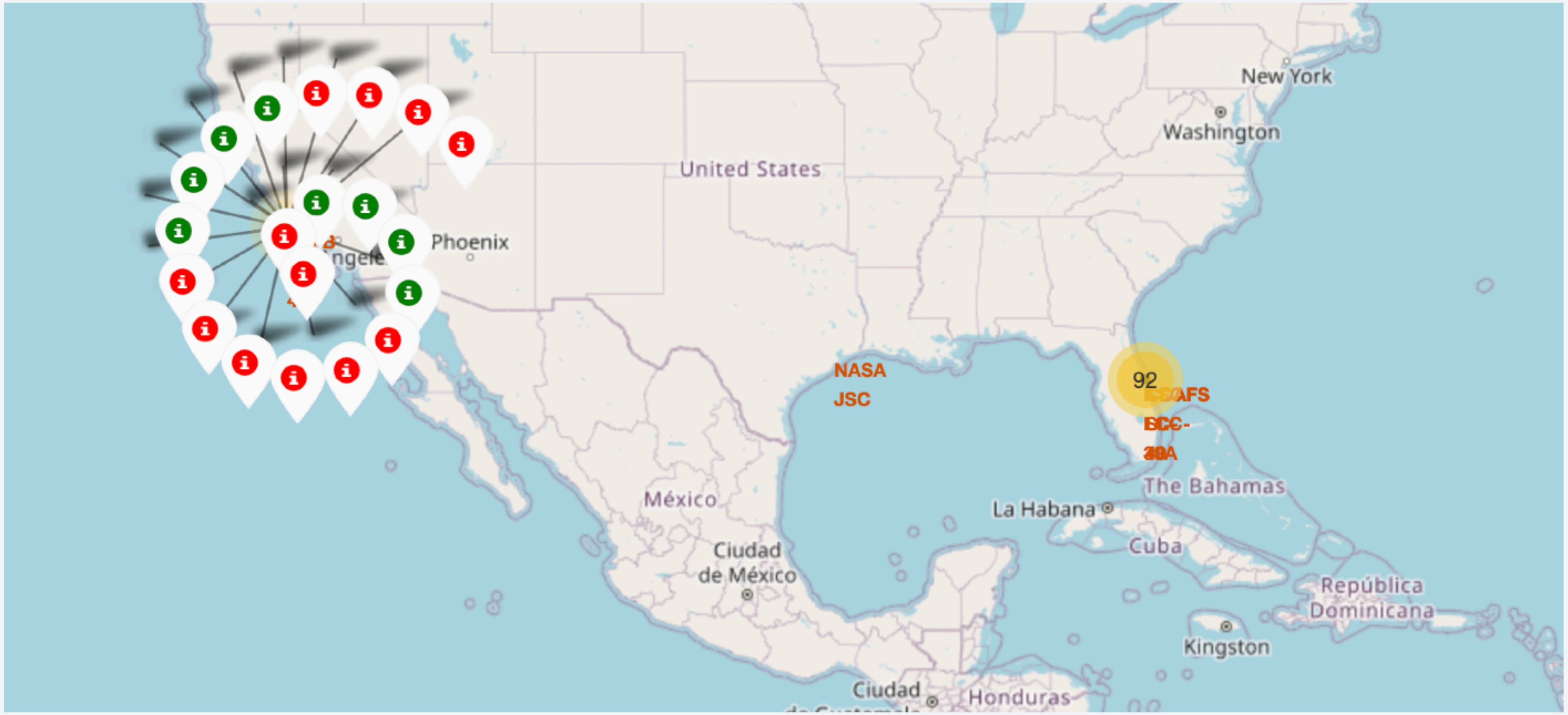
Section 3

Launch Sites Proximities Analysis

All Launch Sites

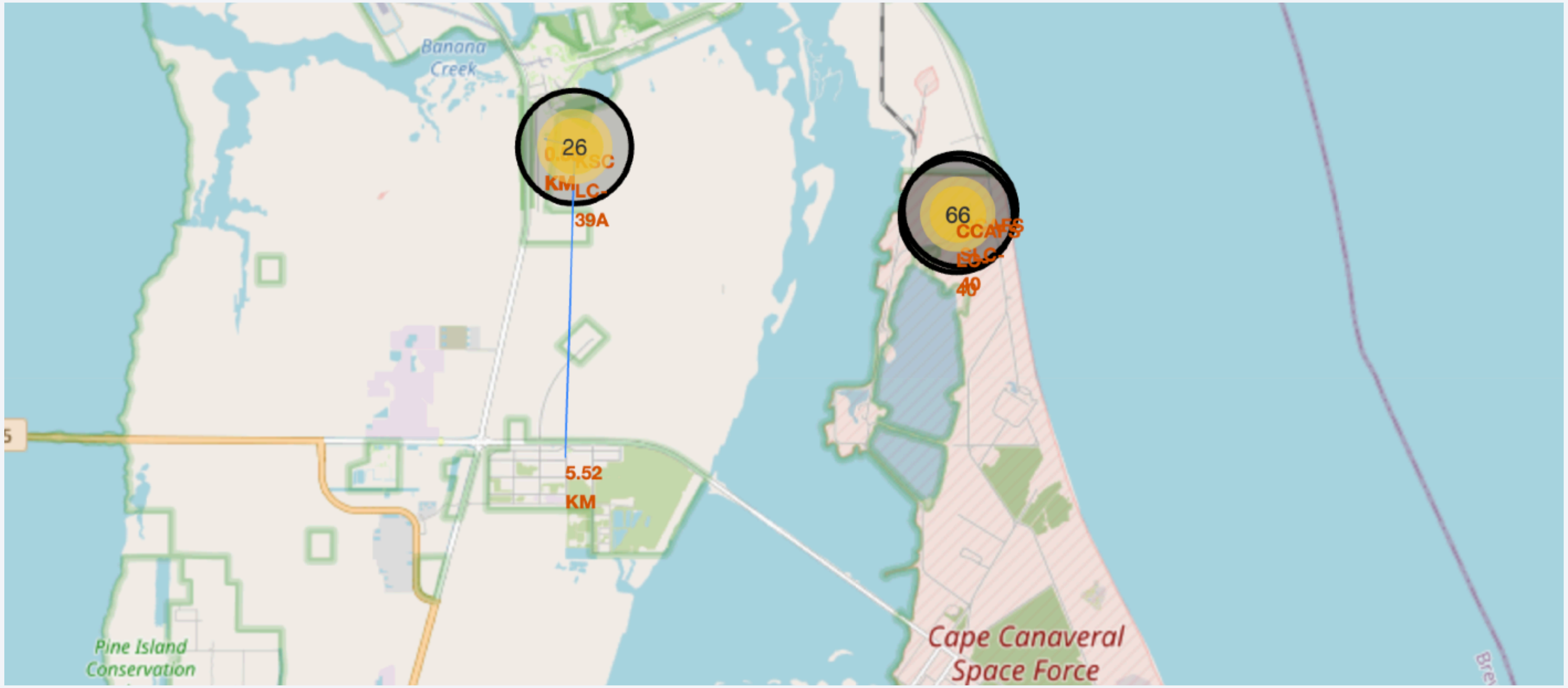
- Launch sites are near sea, probably for safety, but not too far from roads and railroads.

Launch Outcomes by Site



- Green markers indicate successful and red ones indicate failure.

Logistics and Safety

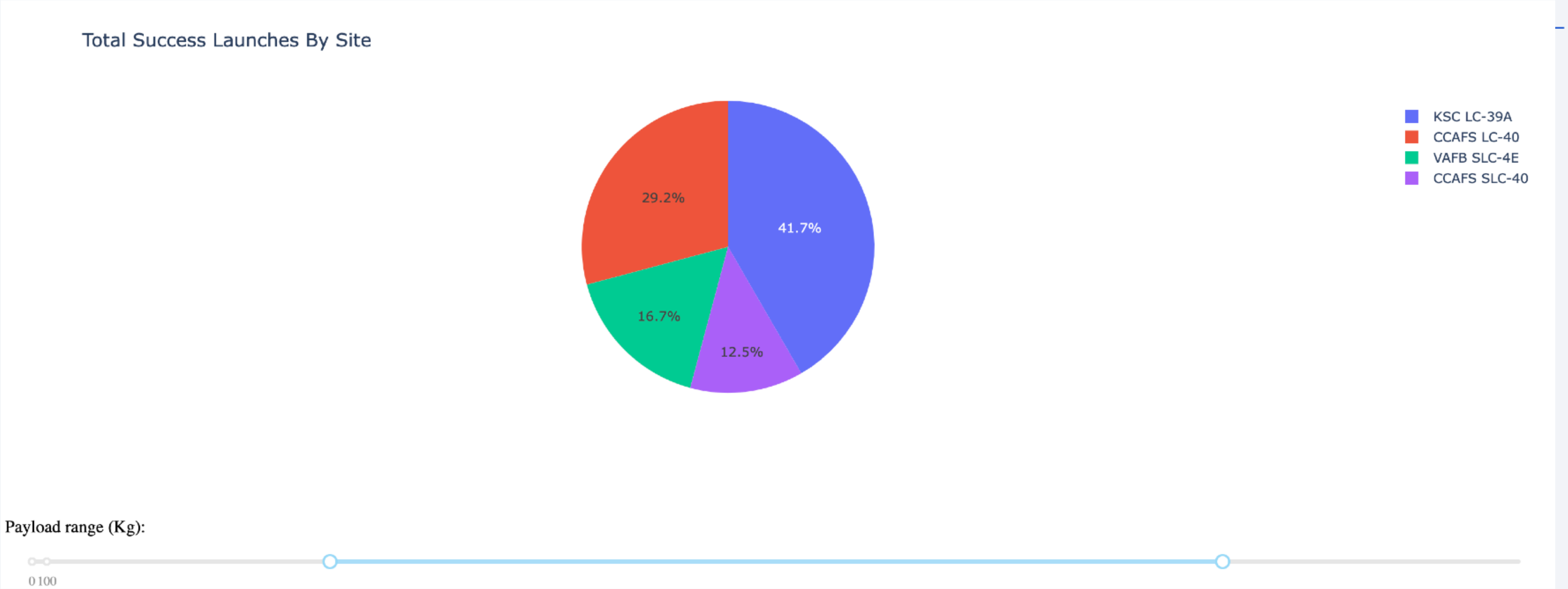


- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas

Section 4

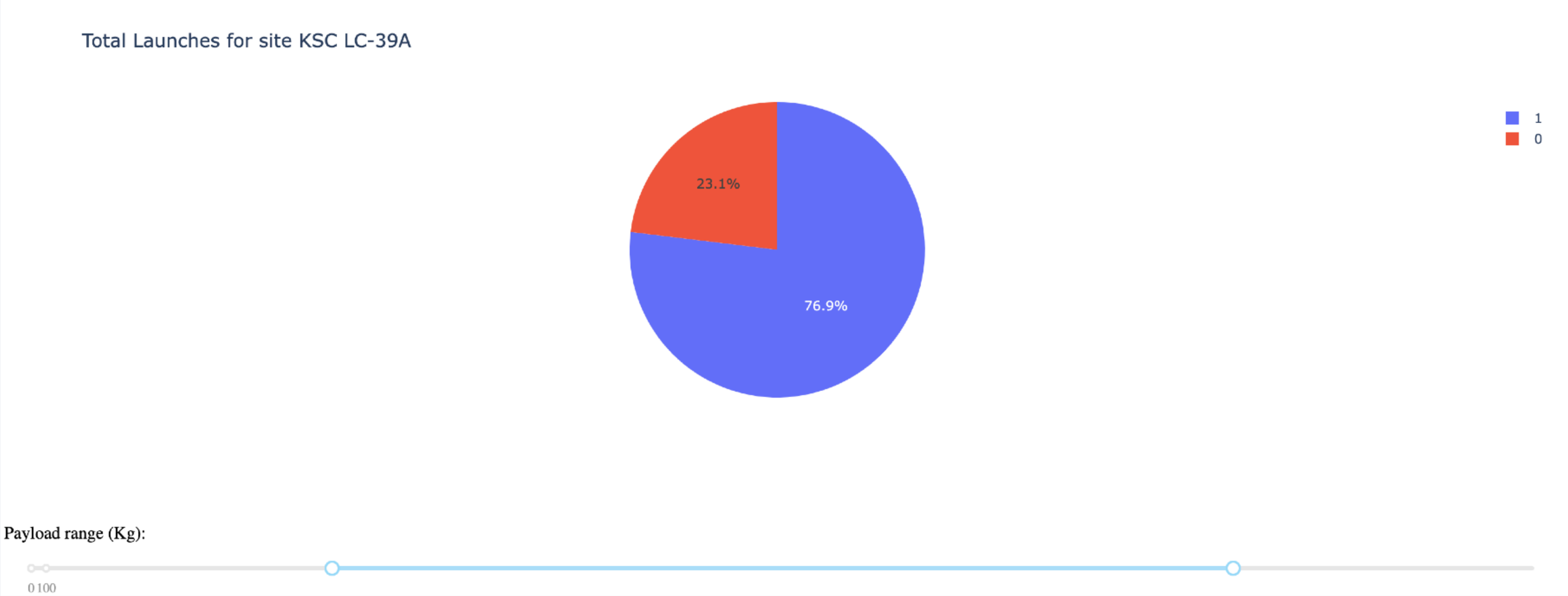
Build a Dashboard with Plotly Dash

Successful Launches by Site



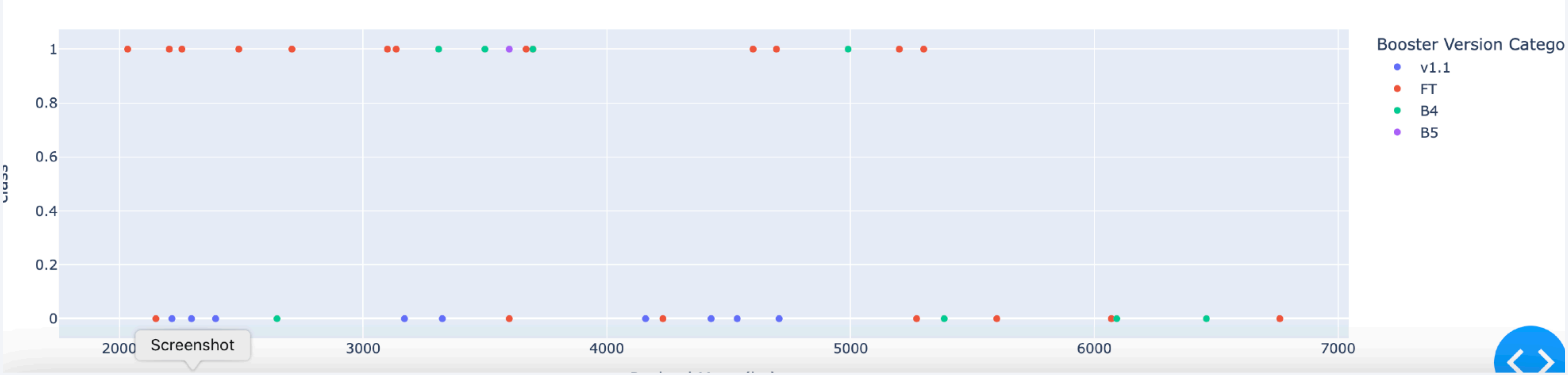
- Launch locations seem to have an impact on success of launches

Highest Launch Success Ratio



- KSC-LC-39A had the highest launch success ratio with 76.9%

Payload vs. Launch Outcome



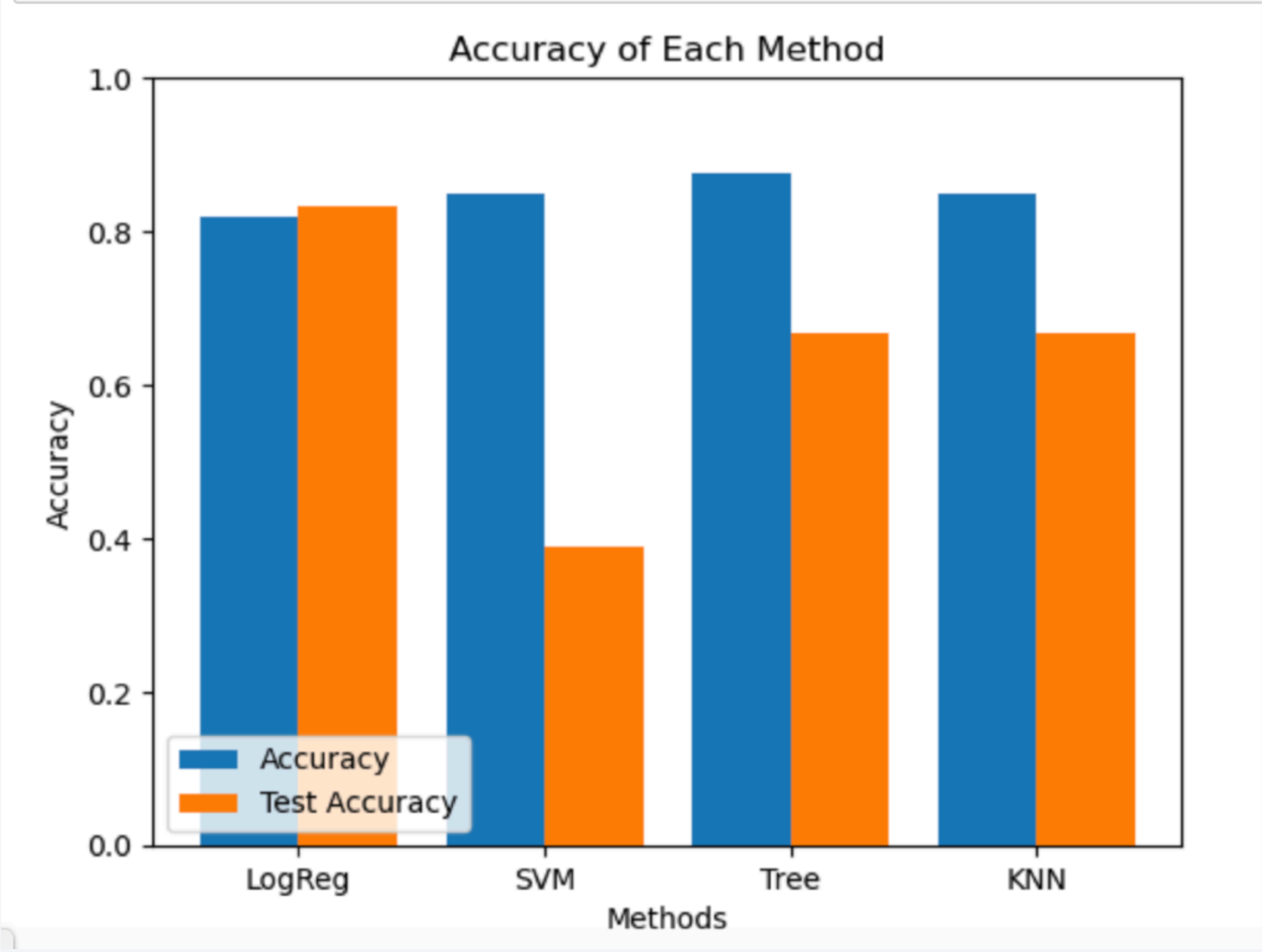
- Payloads under 6,000kg and FT boosters are the most successful combination.

Section 5

Predictive Analysis (Classification)

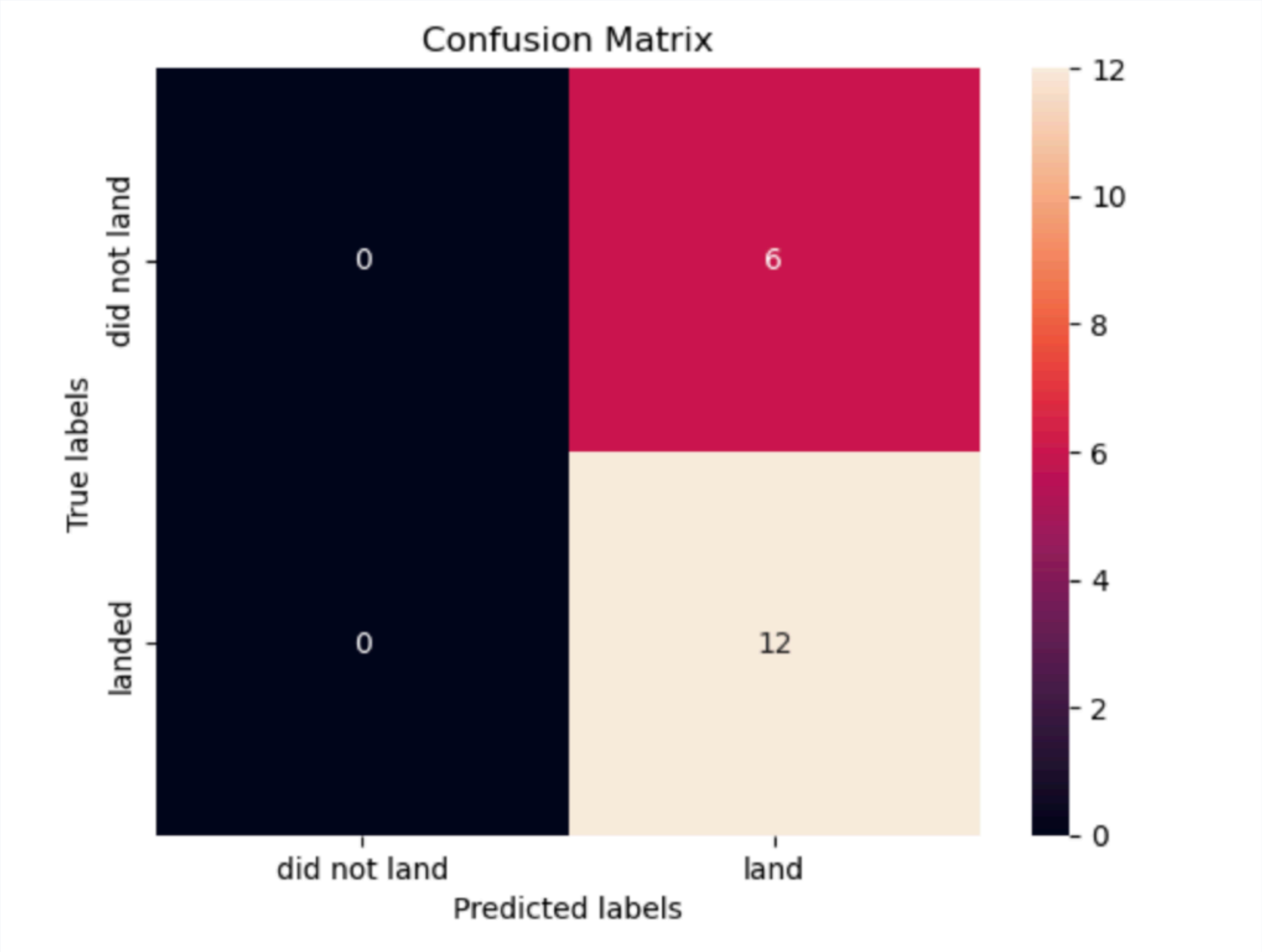
Classification Accuracy

- Decision tree has the highest accuracy with 87%



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.



Conclusions

Different data sources were analyzed conclusions along the process

- The best launch site is KSC LC-39A
- Launches above 7,000kg are less risky
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Thank you!