IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- In this project we have succesfully predicted if the first stage of the falcon 9 rocket will land using several machine learning algorithms
- The main steps include:
  - Data collection
  - Data wrangling
  - EDA
  - Interactive data visualisation
  - ML prediction
- The results show that the probability of success can be predicted using machine learning algorithms
- Out of the algorithms used, decision tree seems to result in the best accuracy for this task

Introduction

- The project is focused on answering two of the most important questions a hypothetical company that wishes to copy the success of spaceX should have:
  - What is the probability that a rocket launch can safely land the first stage of the rocket?
  - What are the best predictors of a successful landing of the first stage?

Section 1

Methodology

Methodology

Executive Summary

- Data collection:
  - Data has been collected from the spaceX API and via web scraping
  - The data has been filtered to use only Falcon 9 launchees
  - Entries with 'None' for PayloadMass have been replaced by the mean value
- Data wrangling:
  - The number of launches for each site was calculated
  - The number of occurrences for each orbit was calculated
  - The number of occurrences of mission outcome for each orbit was calculated
  - A column with the outcome label has been created and filled

Methodology

Executive Summary

- Exploratory Data Analysis (EDA):
  - Established SQLIte database connection
  - Loaded the dataset into SQLIte Table
  - EDA has been performed using SQL queries
- Interactive Visual Analytics Using Folium and Plotly Dash:
  - Marked launch sites on a Folium map
  - Marked the success/failed launches for each site
  - Calculated the distances between a launch site to its proximities
  - Created an interactive Dashboard using Dash

Methodology

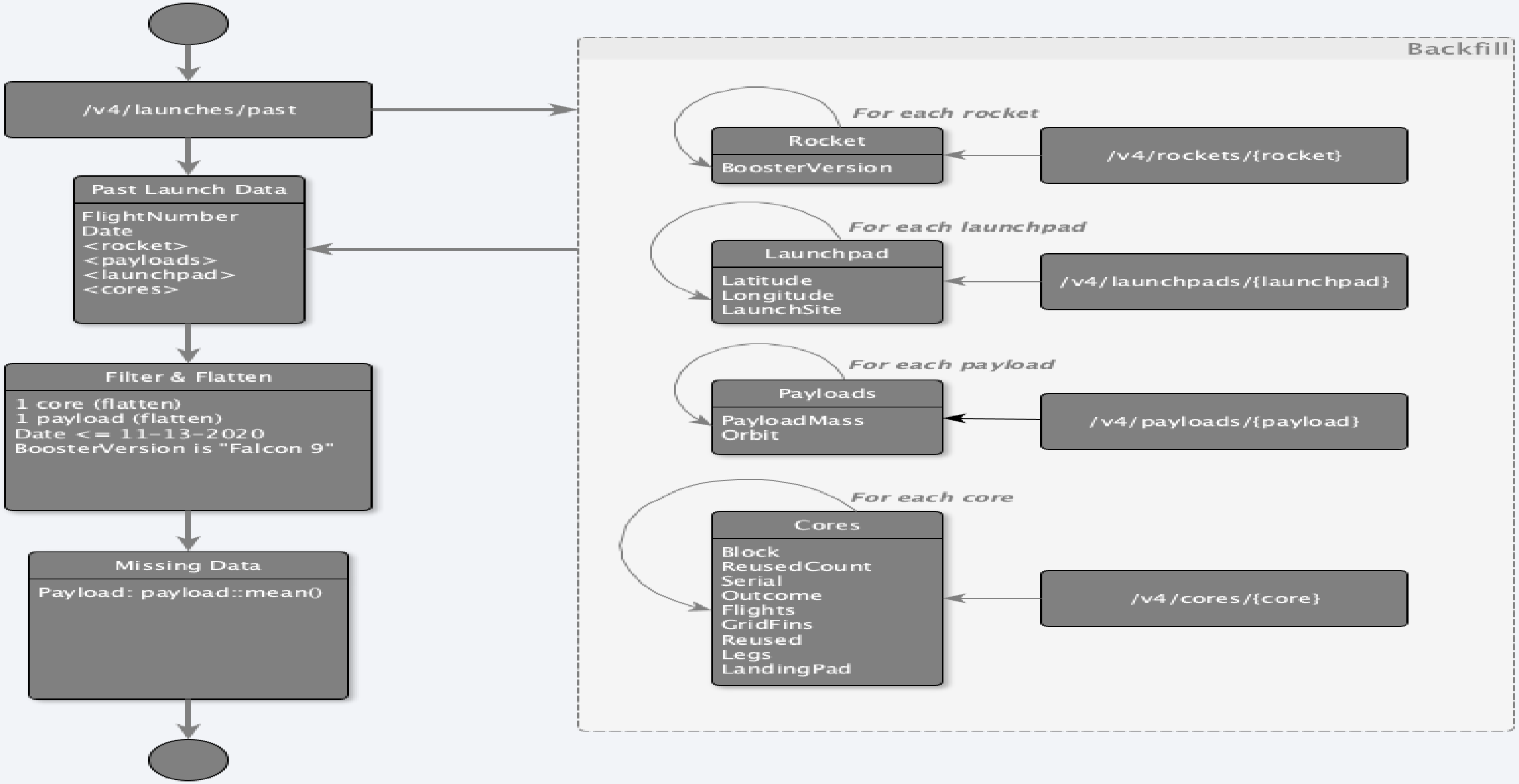
Executive Summary

- Predictive Analysis Using Classification Models:
  - Data has been standardized
  - Data was split into train and test sets
  - Several models have been tested and the accuracy calculated

Data Collection

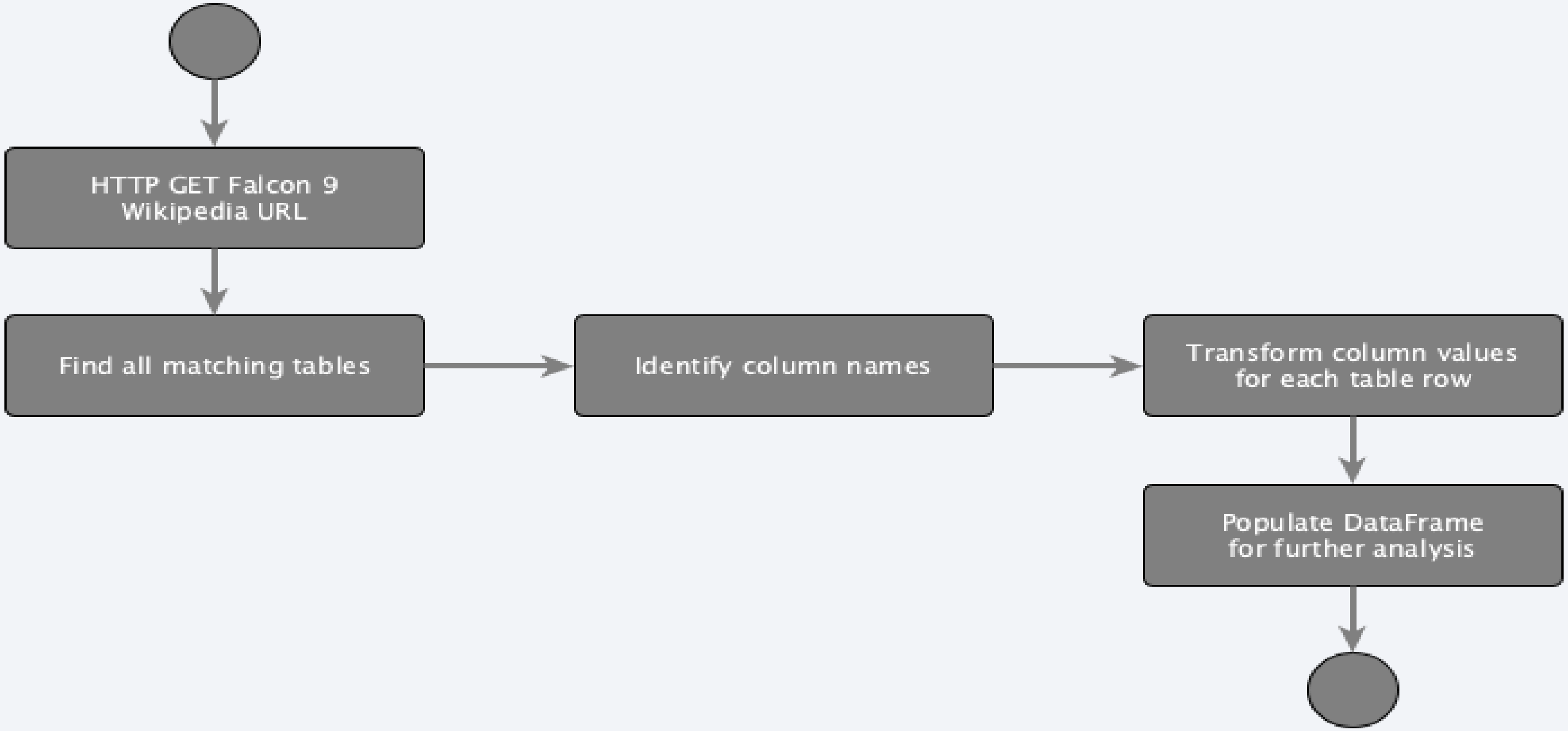
- Data was collected using a combination of API requests and web scraping:
  - Launch data was obtained from: /v4/launches/past
  - Additional data from:
    - /v4/rockets
    - /v4/launchpads
    - /v4/payloads
    - /v4/cores
- Data was scraped from List of Falcon 9 and Falcon Heavy launches Wikipedia page

Data Collection – SpaceX API



[Notebook (Github)]

Data Collection - Scraping



[Notebook (Github)]

Data Wrangling

- Missing values have been found and replaced by the mean for Payload Mass
- The number of launches per Launch Site has been calculated
- The distribution of orbit types in the data set has been shown
- A new column labeled "Class" has been created and filled, to be used as target value in training

- Notebook (Github)

EDA with Data Visualization

- FlightNumber vs. PayloadMass
- FlightNumber vs LaunchSite
- Payload Mass vs Launch Site
- Success Rate of Each Orbit Type
- FlightNumber vs Orbit type
- Payload Mass vs Orbit type
- Launch Success Yearly Trend
- Notebook (Github)

EDA with SQL

- 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.
- 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.
- Notebook (Github)

Build an Interactive Map with Folium

- Launch sites have been added as markers on the map
- Successful and failed launches have been added as green/red circles at each launch site
- The closest coastline to a launch site has been calculated and marked on the map
- Notebook (Github)

Build a Dashboard with Plotly Dash

- A drop-down for launch sites has been added
- A callback function to render the graph based on the selected launch site has been added
- A range slider for the payload has been added
- A callback function to render the graph based on the selected payload mass has been added
- Python Script (Github)

Predictive Analysis (Classification)

- The data has been prepared by taking data['Class'] as the desired outcome, Y
- The data has been split into train and test sets
- Several models have been tested:
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree Classifier
    - K Nearest Neighbours

- Notebook (Github)

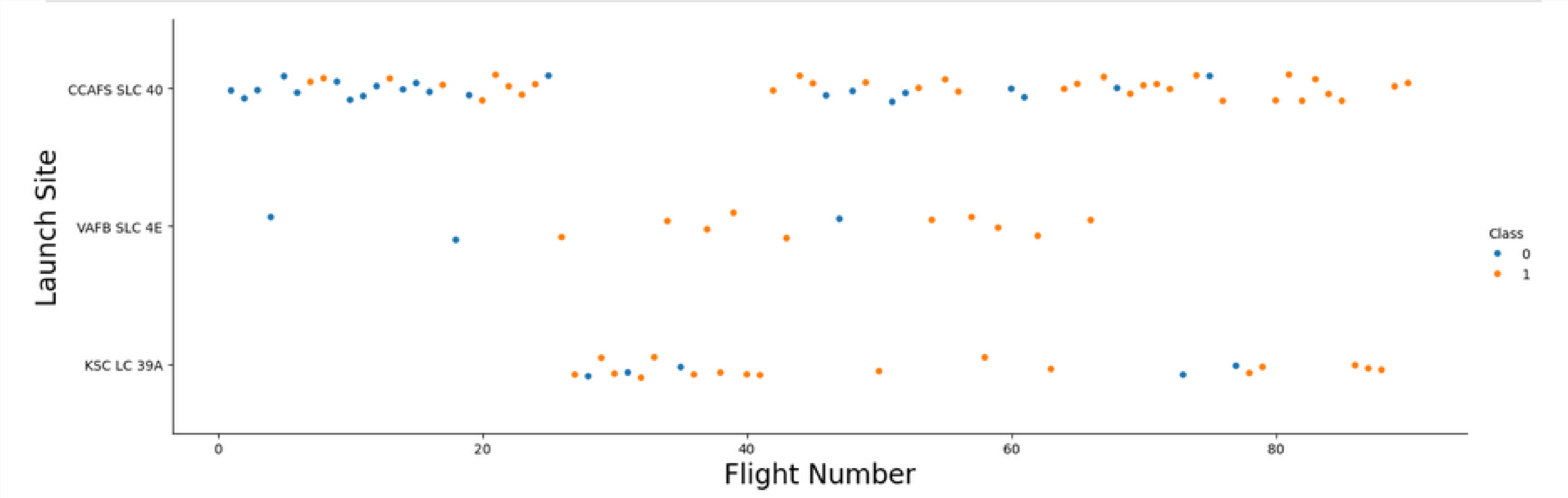
Results

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

Section 2

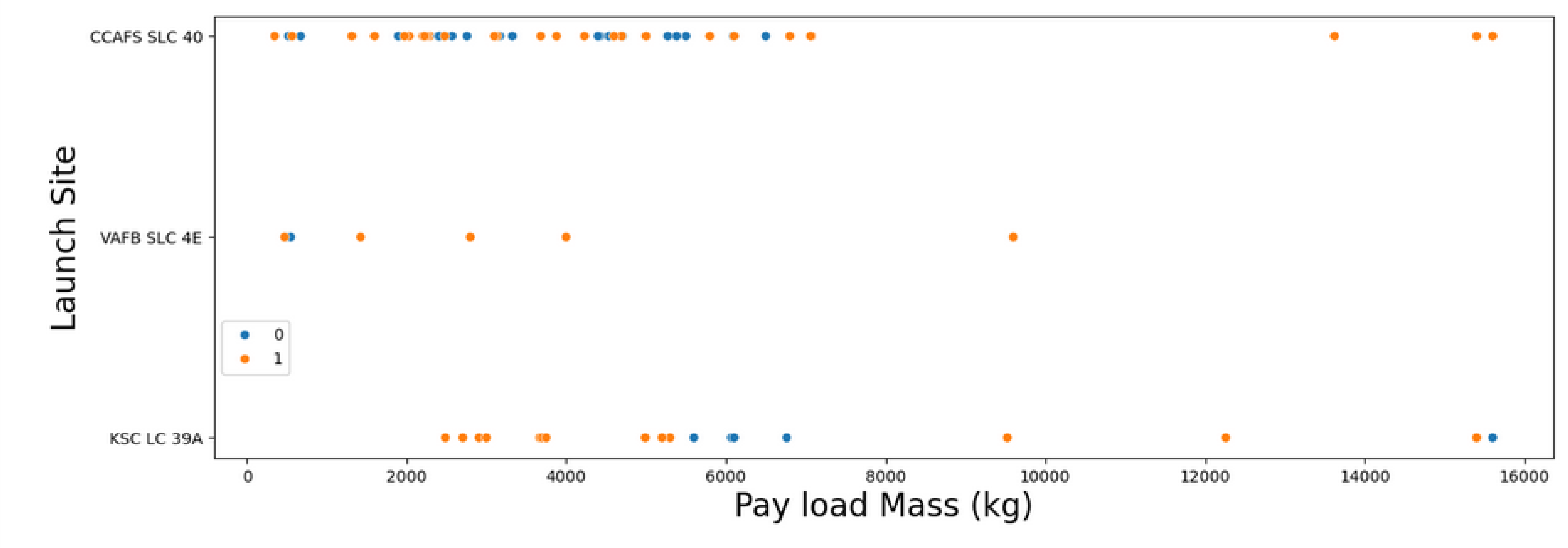
Insights drawn from EDA

Flight Number vs. Launch Site



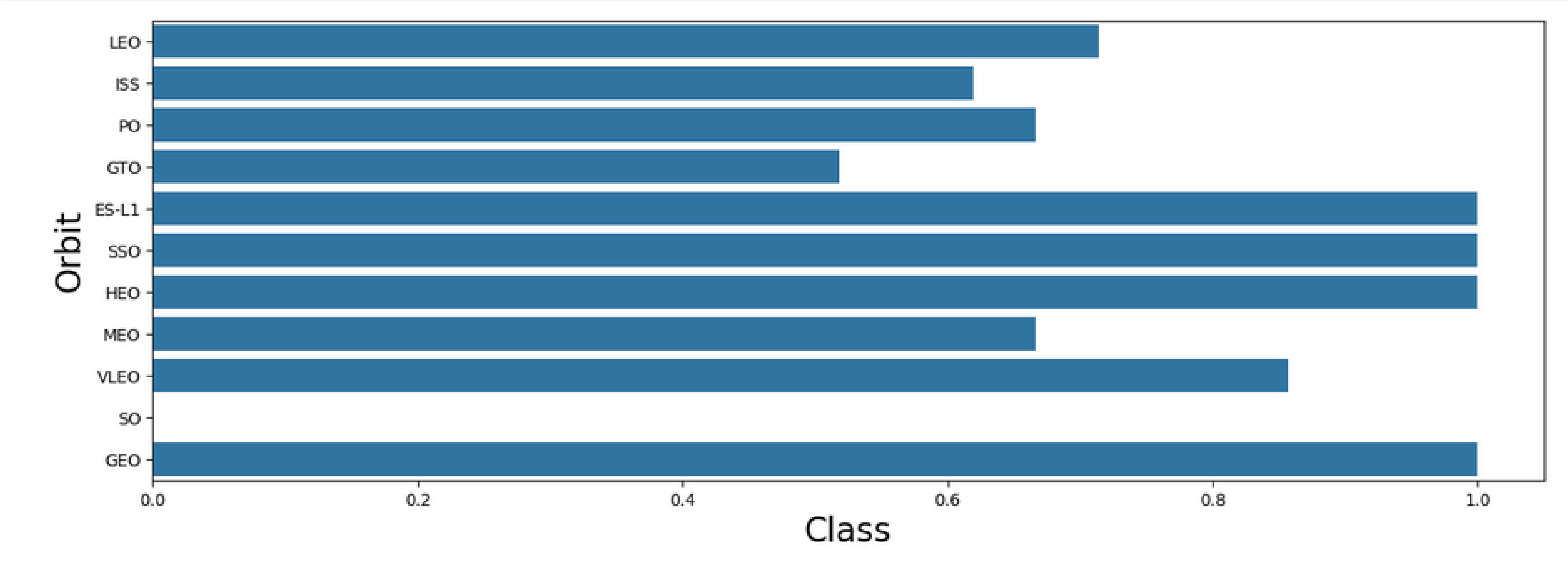
- Launch sites do not seem to influence the success or failure to recover the first stage
- A pattern can be seen as the flight number increases the success rate increases

Payload vs. Launch Site



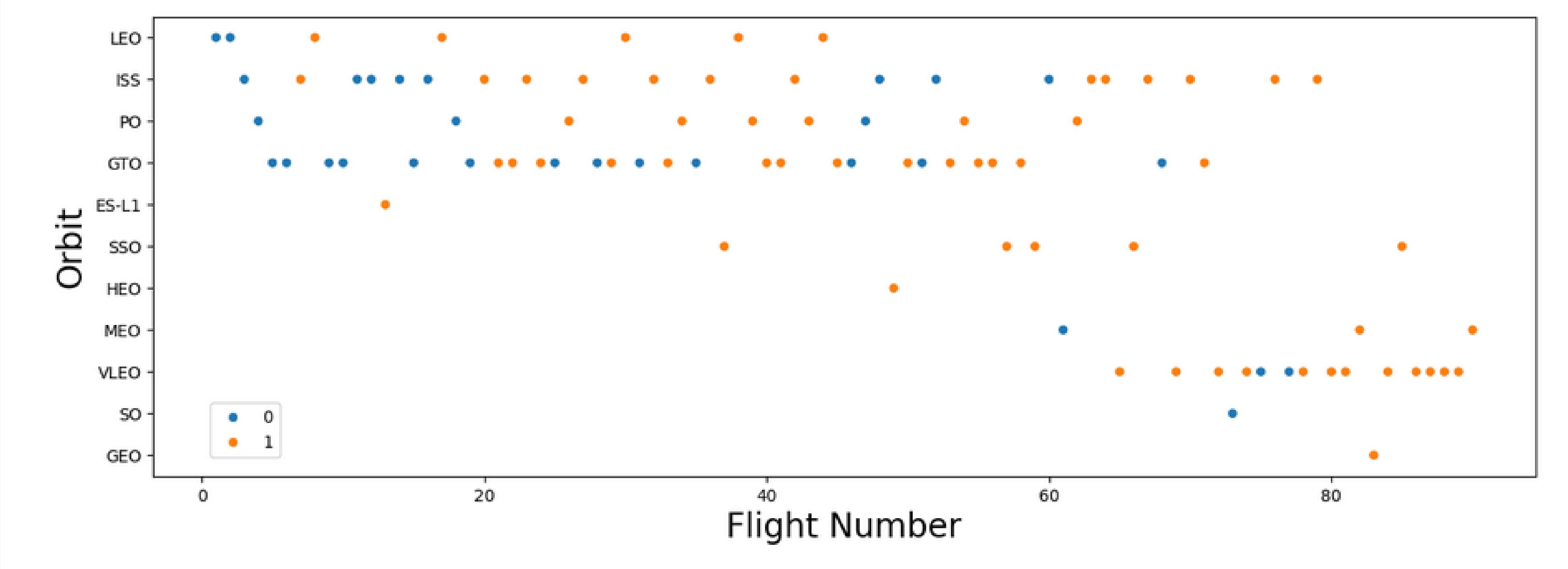
- VAFB SLC 4E has payloads only up to 10000 kg

Success Rate vs. Orbit Type



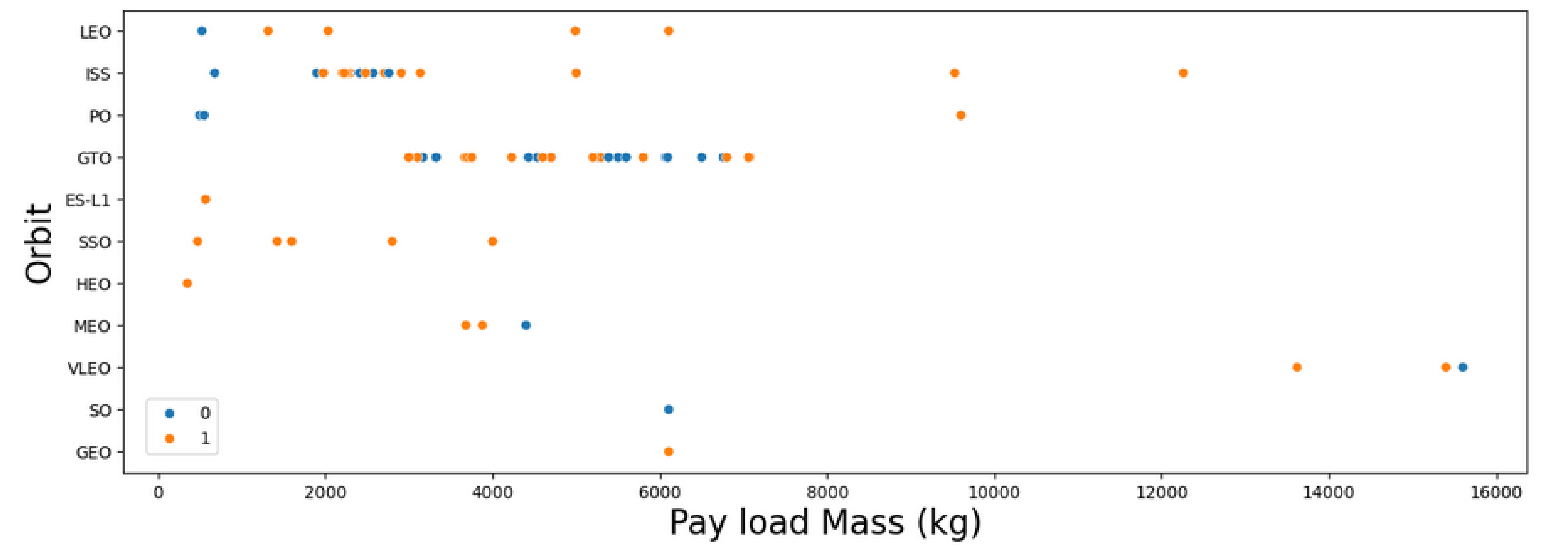
- ES-L1, SSO, HEO and GEO have the maximum success rate

Flight Number vs. Orbit Type



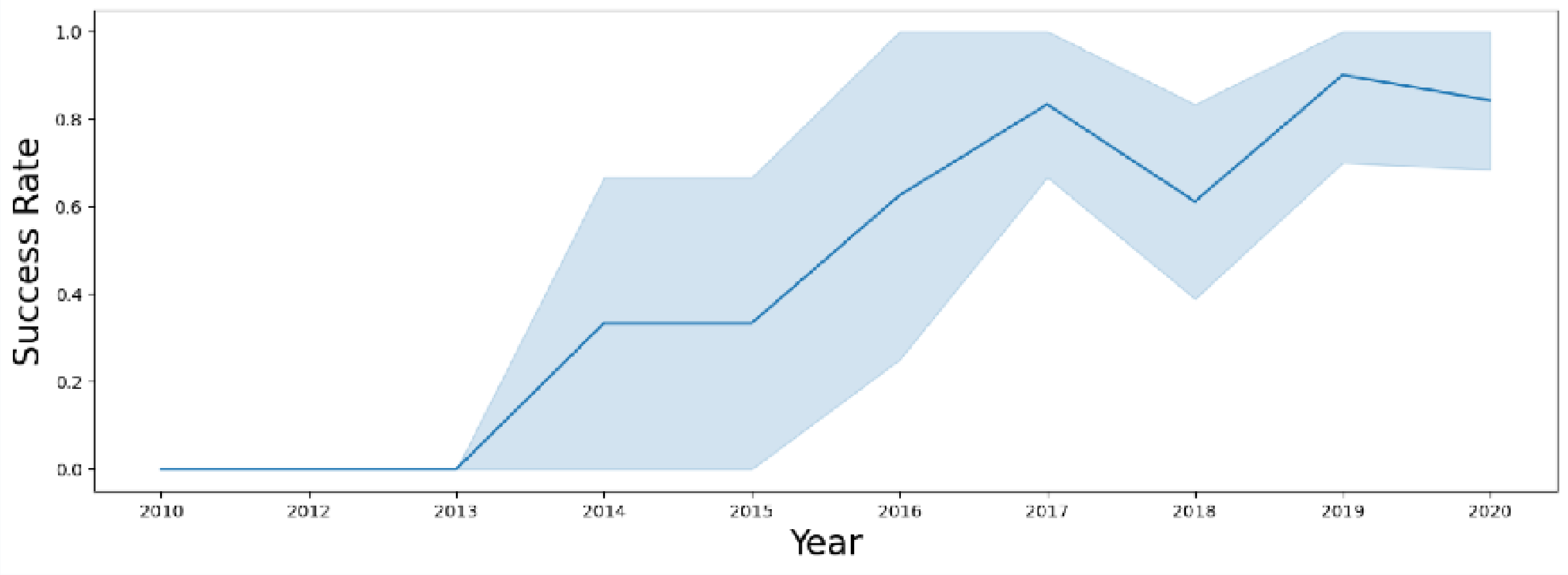
- There is a noticeable trend of higher success rate as the flight number increases
- This is expected as lessons learned in the first flights help improve later flights

Payload vs. Orbit Type



- There seems to be a preference for high payload masses for some orbits like VLEO while others seem to have only small payloads (LEO, SSO, MEO, ES-L1, HEO)

Launch Success Yearly Trend



- The trend of increasing success rate year over year is evident in this graph
- The maximum is reached in 2019 and a plateau above 0.8 could be expected in the future

All Launch Site Names

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

Total Payload Mass

TOTAL_PAYLOAD
45596

Average Payload Mass by F9 v1.1

AVG_PAYLOAD_MASS
2534.6666666666665

First Successful Ground Landing Date

LaunchDate
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version	PAYLOAD_MASS__KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

Mission_Status	COUNT(*)
Failure	1
Success	100

Boosters Carried Maximum Payload

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

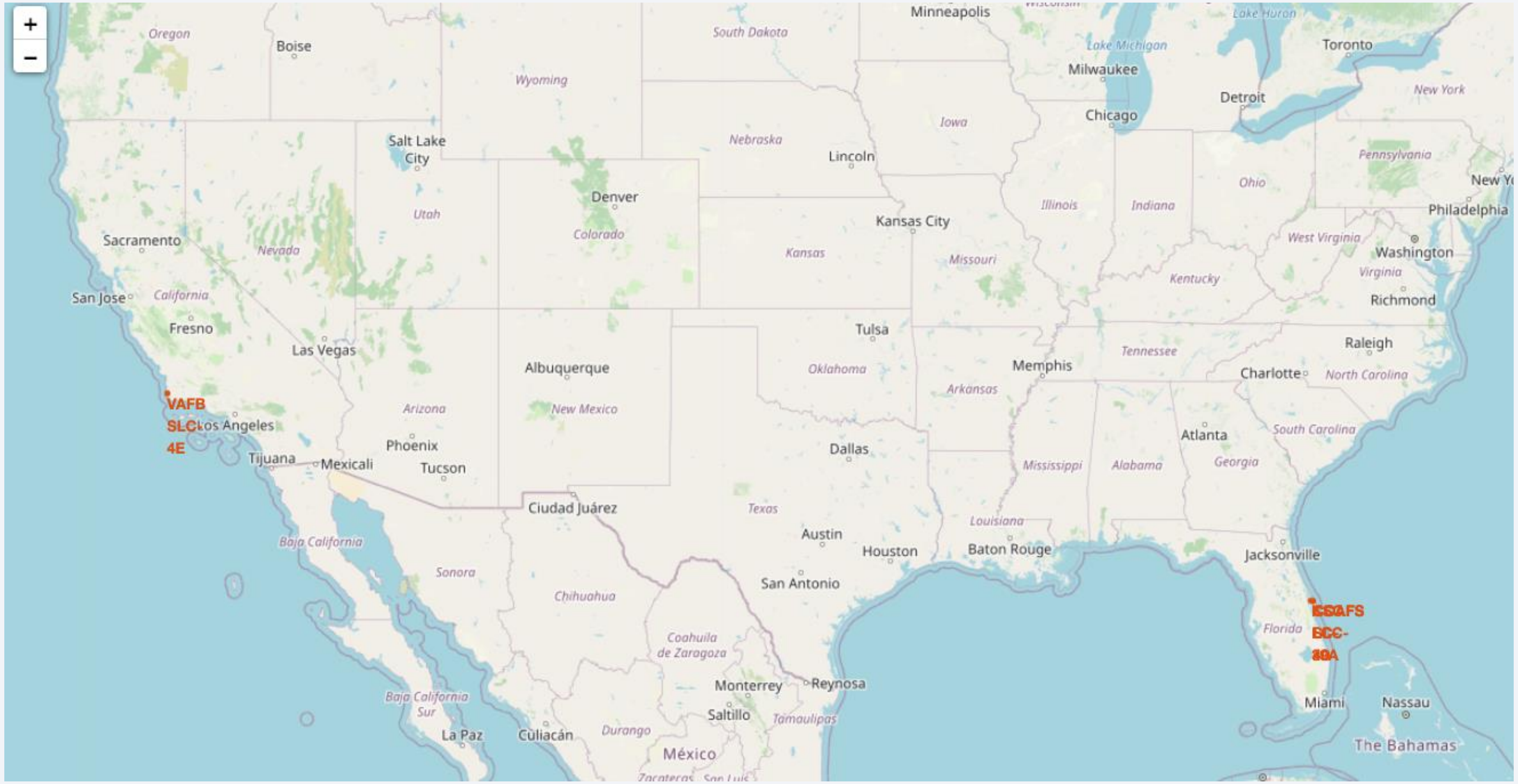
2015 Launch Records

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

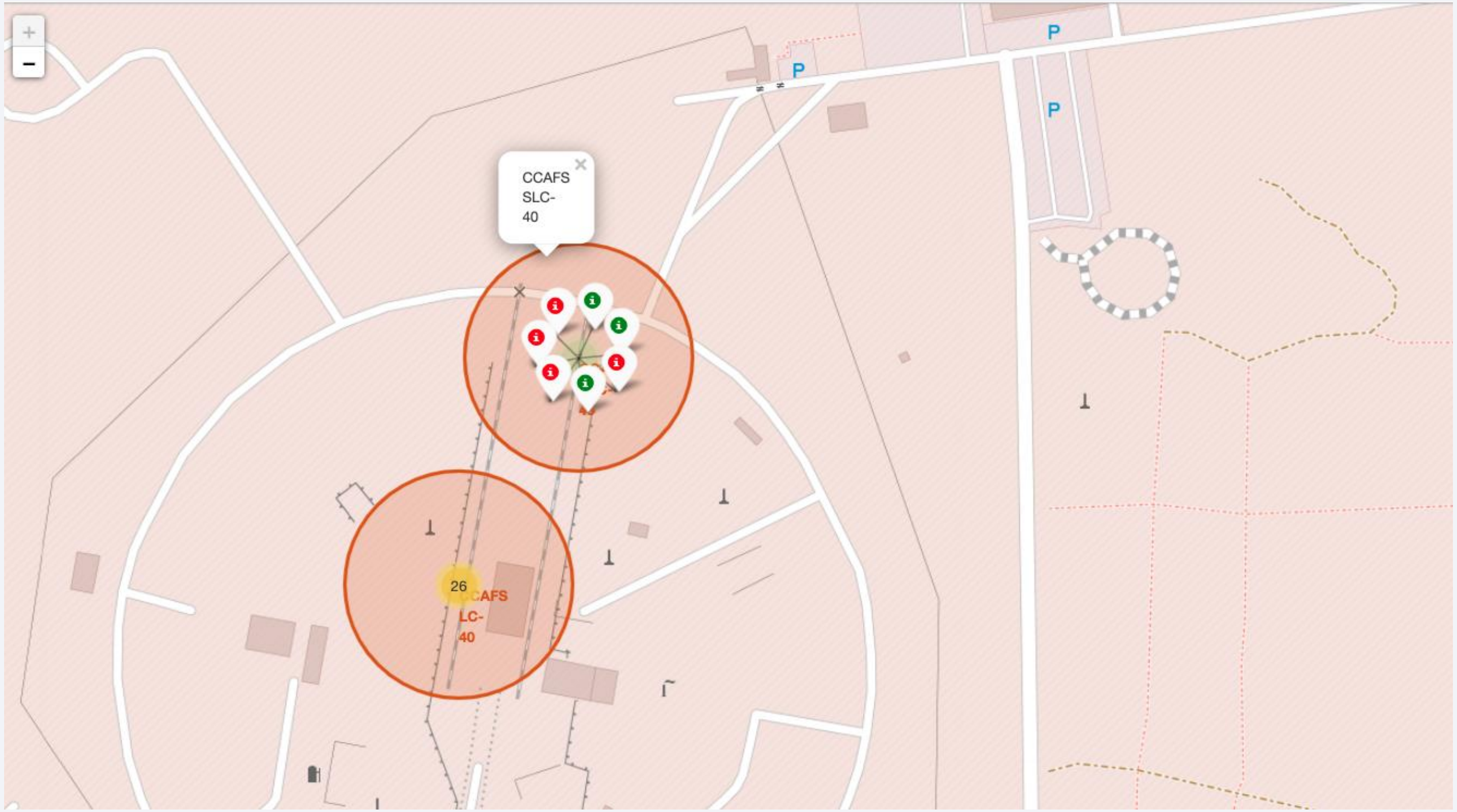
Section 3

Launch Sites Proximities Analysis

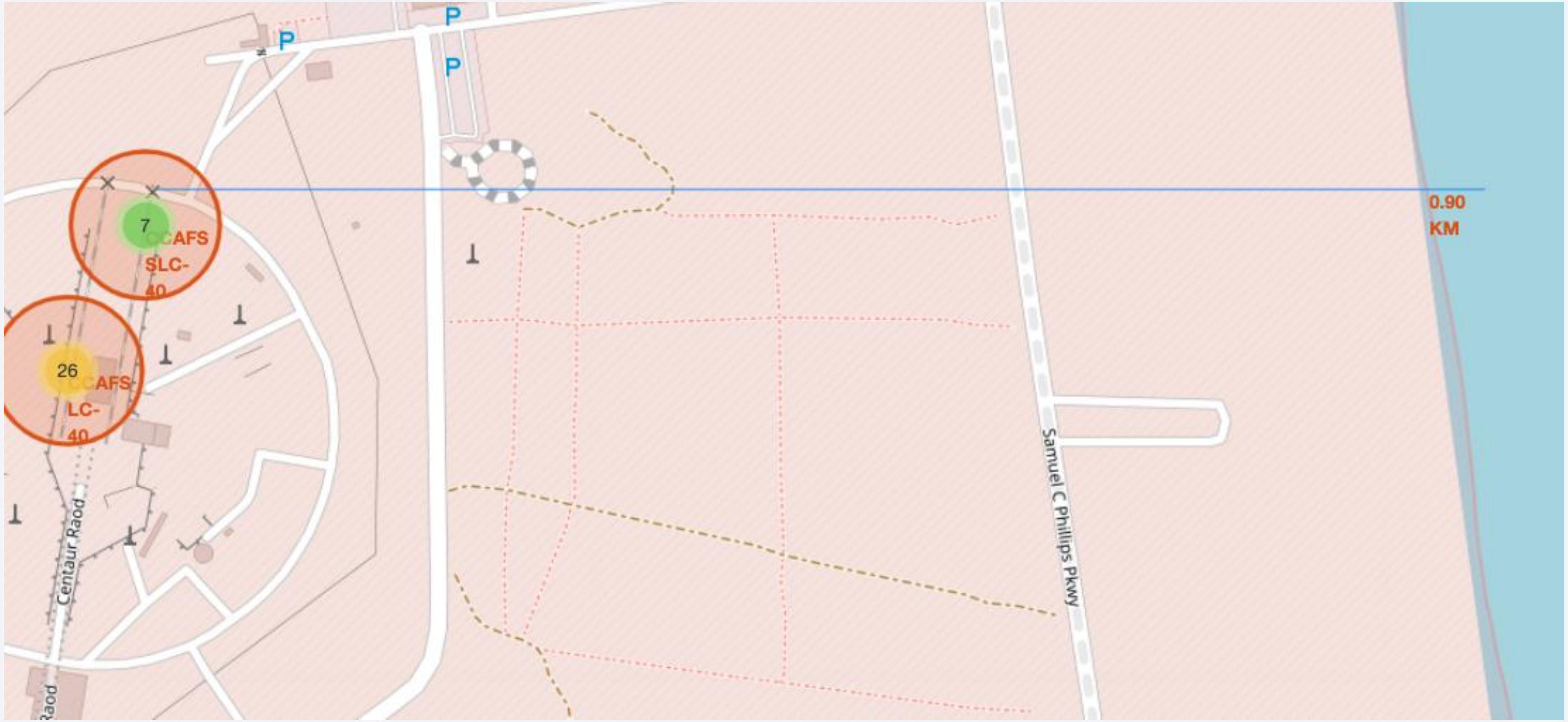
Launch Site Locations



Launch Outcomes



Proximity to Shoreline

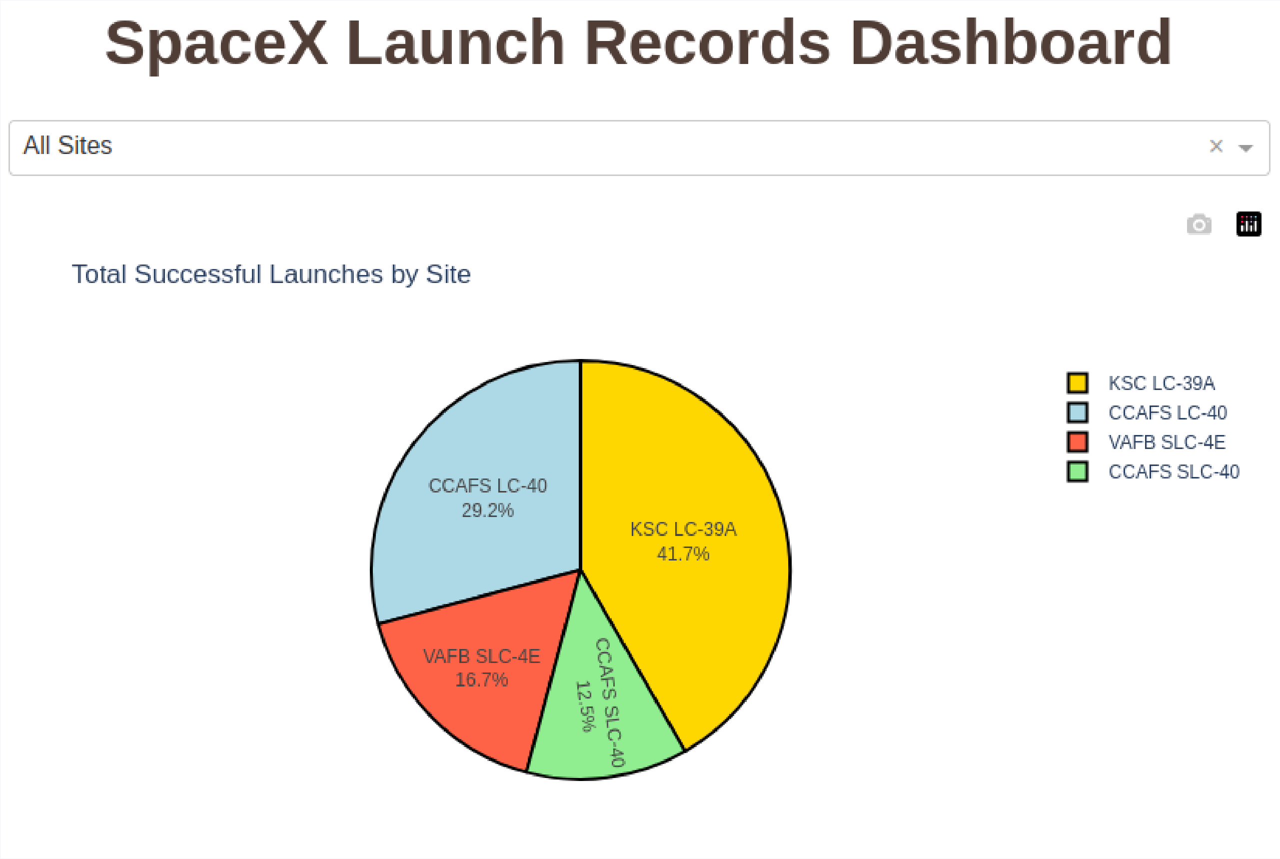


Section 4

Build a Dashboard with Plotly Dash

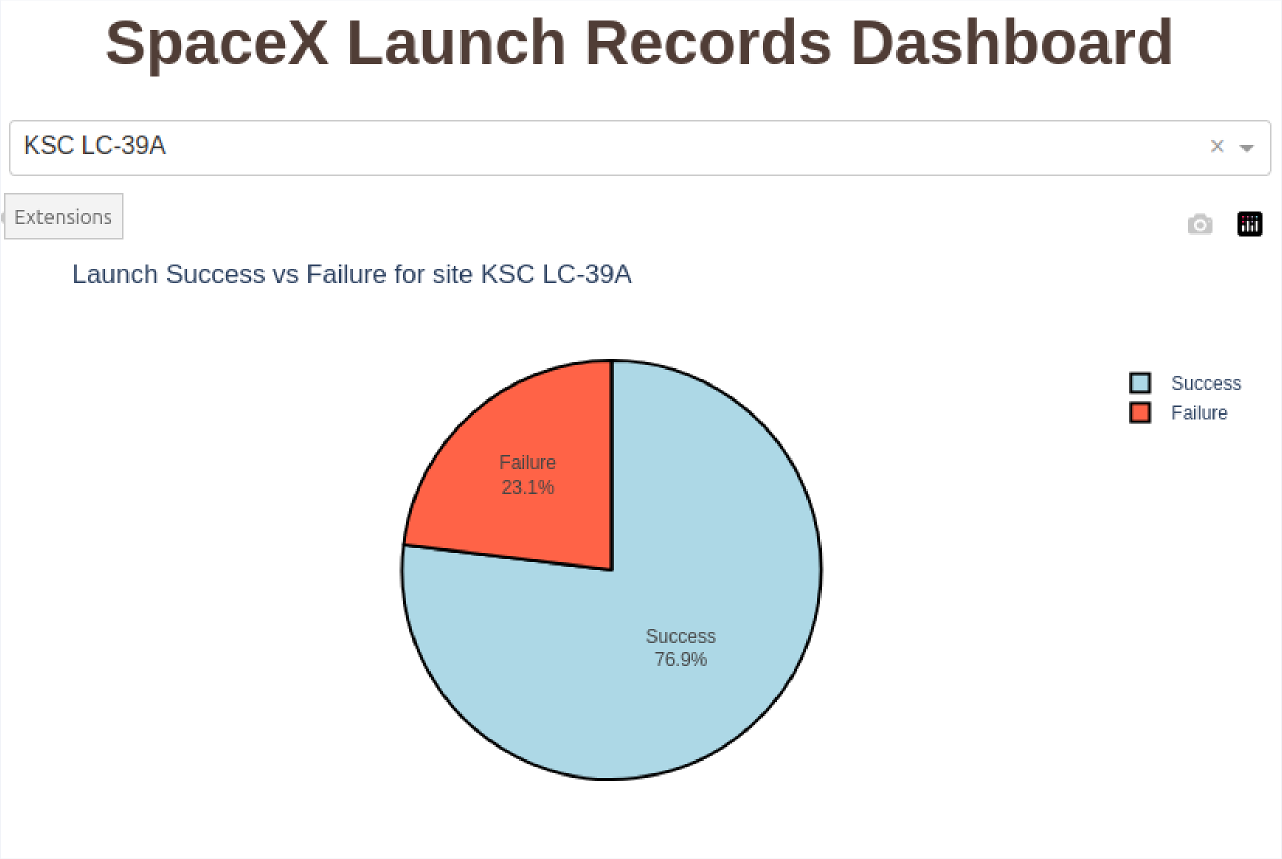
Total Successful Launches by Site

- KSC LC-39A has the highest proportion of successful launches
- CCAFS SLC-40 has the lowest proportion of successful launches

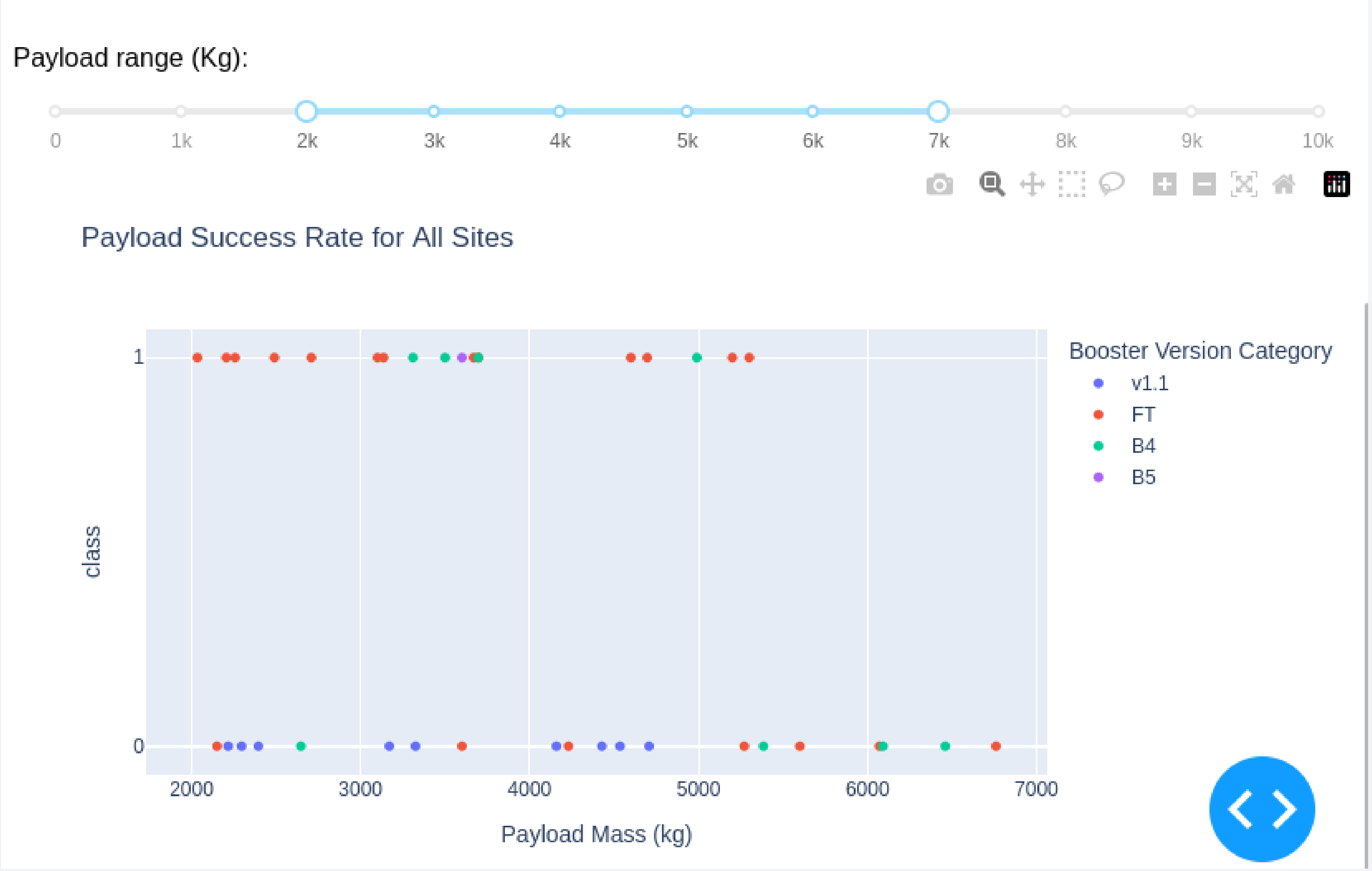


Launch Success vs Failure for site KSC LC-39A

- Success rate is 76.9%



Payload between 2k and 7k Success Rate



Section 5

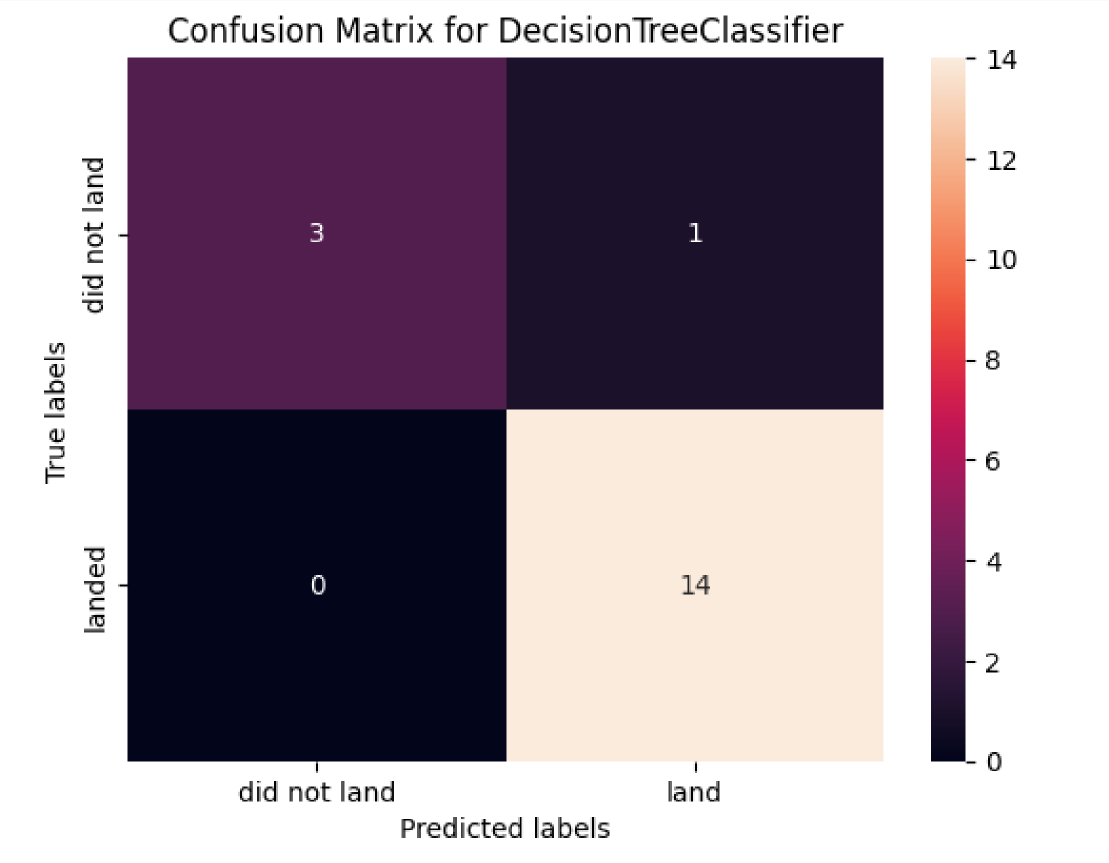
Predictive Analysis (Classification)

Classification Accuracy

Model	Accuracy
Logistic Regression	0.846
Support Vector Machine	0.848
Decision Tree Classifier	0.876
k nearest neighbors	0.848

Confusion Matrix

- 14 observations were correctly predicted as successful landings (true positive)
- 3 observations were correctly predicted as failed landings (true negative)
- 1 observation was incorrectly predicted as a successful landing (false positive)
- No observations were incorrectly predicted as a failed landing (false negative)



Conclusions

- Using EDA the most noticeable trend is that the success rate has increased over time
- Other features such as launch site, orbit or payload mass do not influence the success rate as much
- Several models have been tested to see if the success of a stage 1 landing can be predicted and most models tested have an acceptable level of accuracy
- The best performing model was Decision Tree Classifier with an accuracy of 0.876 and a score of 0.88

Thank you!