IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Steve Kelly
25– JUN - 2022

Outline

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

Executive Summary

- Connecting to SpaceX API, then wrangling the data, and storing into DataFrames to work with is essential for data analysis to be done.
- Exploratory data analysis was done to work with the data and get results through visualizations, in map form, charts, and statistical analysis.
- The findings generated show that SpaceX booster launches have had higher success in recent years. Various payload masses, launch sites, and orbits were tested. Variations in success rate may be due to these factors, or it more than likely is due to the technological advances SpaceX has made with their rocket technology.

Introduction

- As data scientists, we are curious about finding answers to problems. We do this through using data.
- SpaceX has had numerous launches that have failed and succeeded. I wanted to see if a prediction can be made for the Falcon 9 first stage landing successfully.
- If a prediction for the first stage landing is successful, a cost for the launch can be calculated. This can be helpful for alternative companies wanting to bid against SpaceX.

Section 1

Methodology

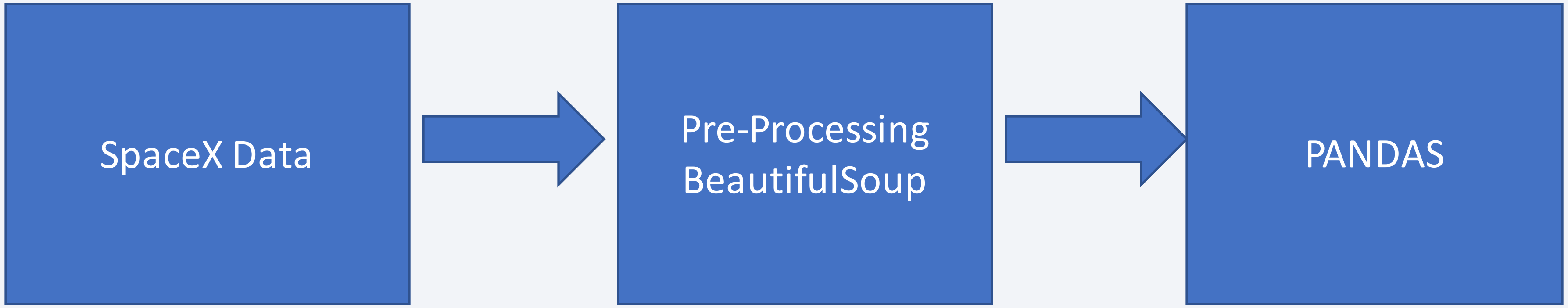
Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from public SpaceX data relating to Falcon 9 and their launches.
- Perform data wrangling
  - Data was processed using Pandas, to create data frames to work with.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

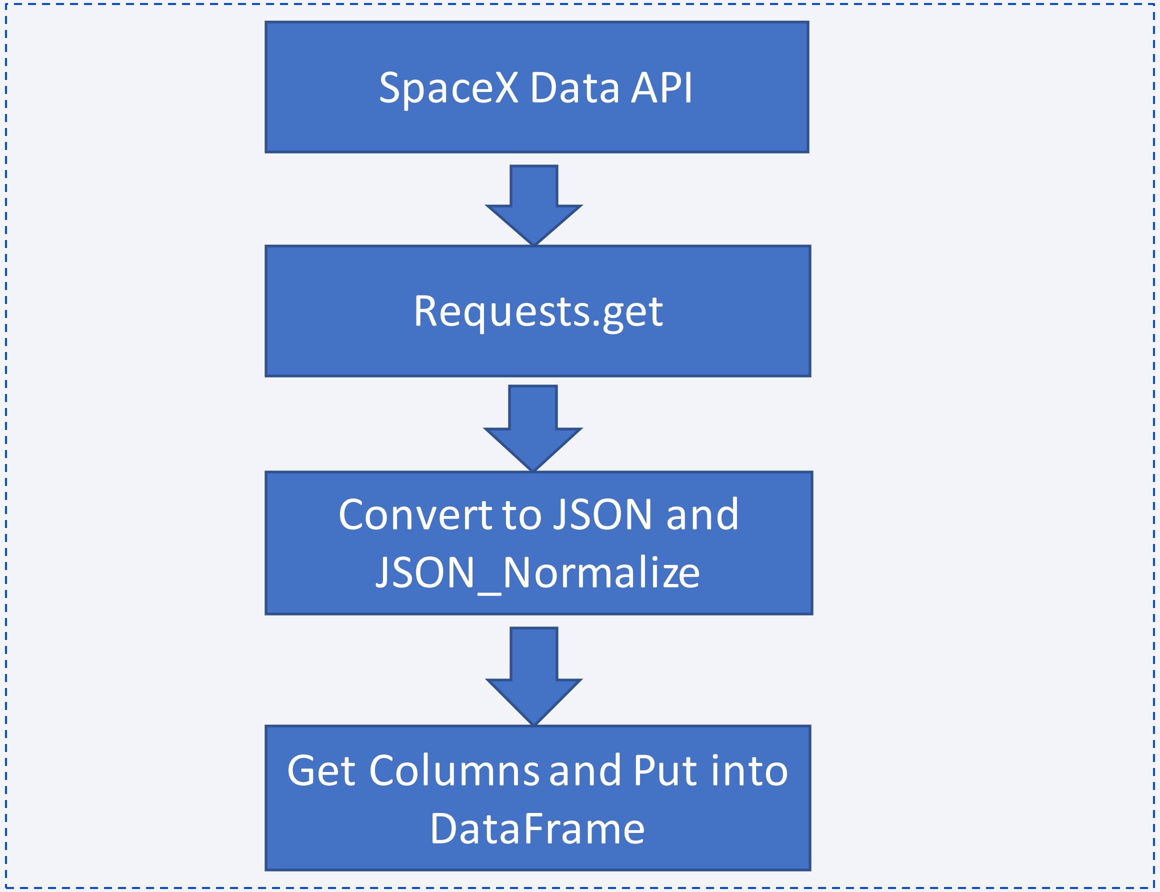
- Data sets were collected from SpaceX Wikipedia Pages listing Falcon 9 launches, and from the SpaceX official website using their SpaceX REST API.
- BeautifulSoup was used to process the HTML data, then PANDAS was used to create a DataFrame from the HTML data.



Data Collection – SpaceX API and Scraping

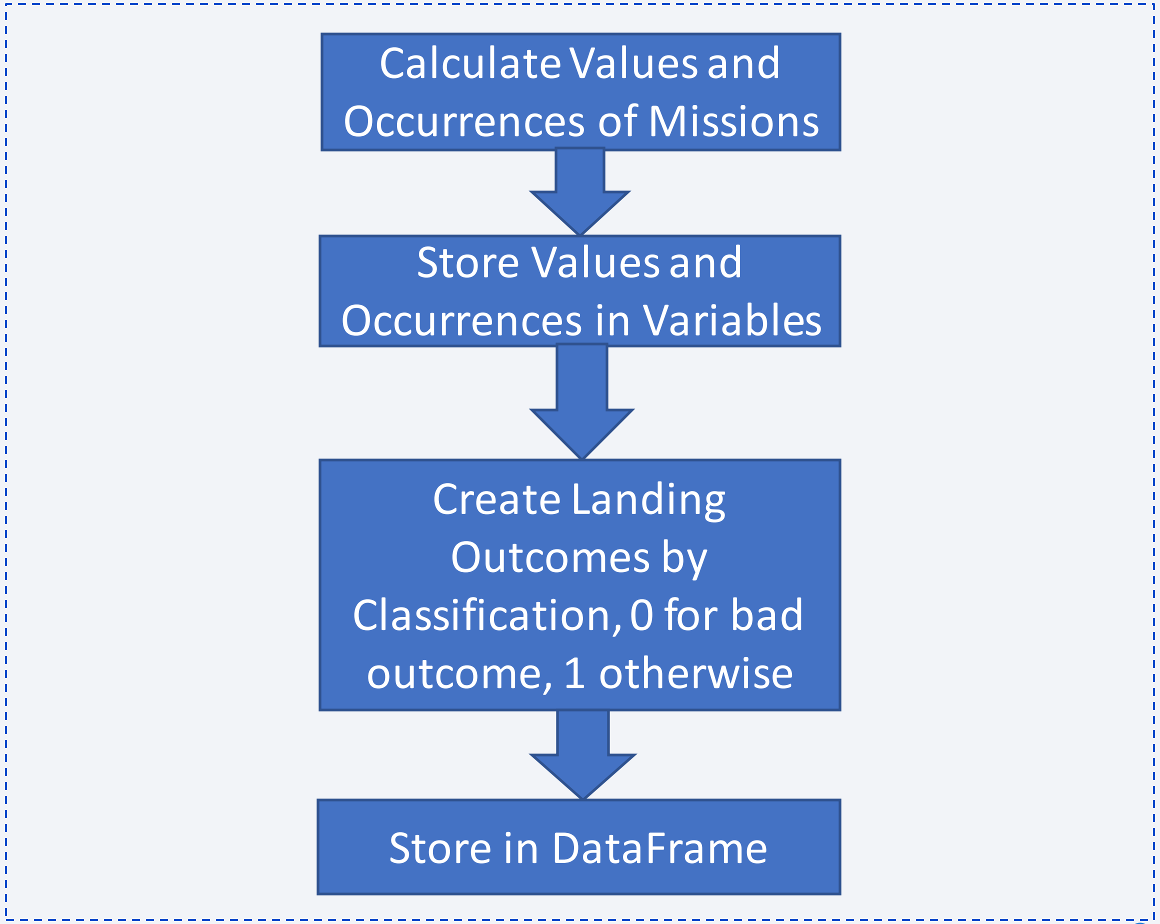
- Call URL of SpaceX Data API
- Use requests.get to store in response variable
- Convert response to json and json_normalize the response
- Get columns from response and put into a dataframe

- github.com/steezkelly/coursera-capstone-project/blob/master/Data_Collection_API%20Completed.ipynb



Data Wrangling

- Calculate Number of Launches on Each Site
- Calculate Number and Occurrence of Each Orbit
- Calculate Number and Occurrence of Each Mission per Orbit Type
- Store in Data Frame as Class (0 bad outcome, 1 otherwise)
- github.com/steezkelly/coursera-capstone-project/blob/master/Data%20Wrangling%20Data%20Science%20Capstone.ipynb



EDA with Data Visualization

- Exploratory Data Analysis was used with Data Visualization
- ScatterPlot was used to see relationship between:
  - Flight Number and Payload Mass
  - Flight Number and Launch Site
  - Flight Number and Orbit type
  - Payload and Orbit Type
- BarPlot was used to see relationship between Orbit and Success (Class)
- LineChart was used to see success of launch per year
- github.com/steezkelly/coursera-capstone-project/blob/master/EDA%20with%20Visualization%20Lab.ipynb

EDA with SQL

- SQL queries were used to:
  - Select the DataSet from CSV Table
  - Unique Launch Sites
  - Total Payload Mass from boosters launched by NASA
  - Average Payload Mass
  - First Successful Launch
  - Successful and Failed Mission Outcomes
  - Boosters which carried maximum payload
- github.com/steezkelly/coursera-capstone-project/blob/master/Complete%20the%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Map objects were created and added to a Folium Map by:
  - Getting coordinates of launch sites
  - Creating folium circles with text markers labeling the site
  - Creating colored marker clusters for class outcome (0 bad outcome, 1 otherwise)
  - Drew lines between launch sites and their proximities

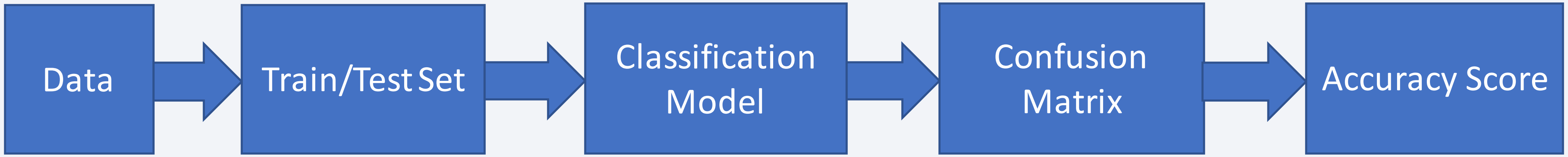
- These objects were added to visually show launch sites and features were located.
- github.com/steezkelly/coursera-capstone-project/blob/master/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- A Plotly Dashboard was created to show a scatterplot and pie chart of each of the launch sites, and all launch sites.
- Scatterplot was used to show the relationship between the booster versions, their payload mass and the outcome of the mission.
- Pie chart was used to show the launch sites and the percentage of successful outcomes.
- A dropdown menu to choose All Sites, or specific launch sites
- github.com/steezkelly/coursera-capstone-project/blob/master/Interactive_Dashboard_Plotly_SpaceX.ipynb

Predictive Analysis (Classification)

- Data was transformed and split into training and test sets. Multiple classification models were created and fed the data. GridSearch and Confusion Matrix were used with models. Calculated score of test set.
- Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors.



- github.com/steezkelly/coursera-capstone-project/blob/master/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

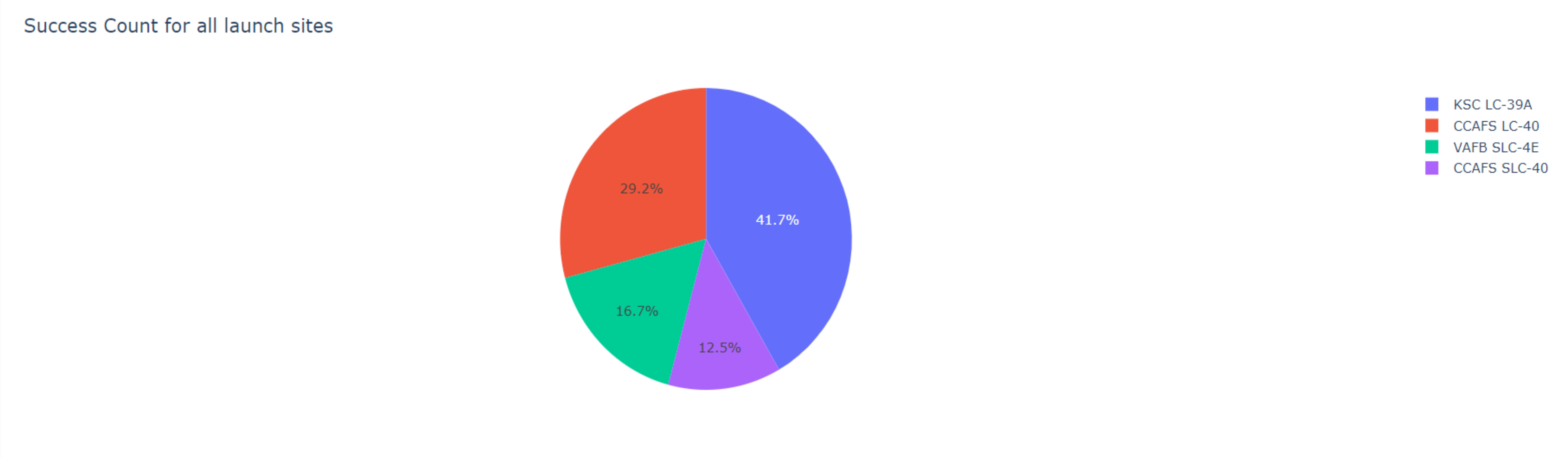
Results

- Orbits had a significant effect on mission outcome for payload masses, and number of mission attempts.
  - GEO, SSO, HHO, and ES-LI had very high success. VLEO and LEO had high success outcome.
  - Heavy payloads were more successful in the Polar, LEO, and ISS orbits.
  - Success rate continuously increased from 2013 to 2020, with 2019 being most successful launch year.

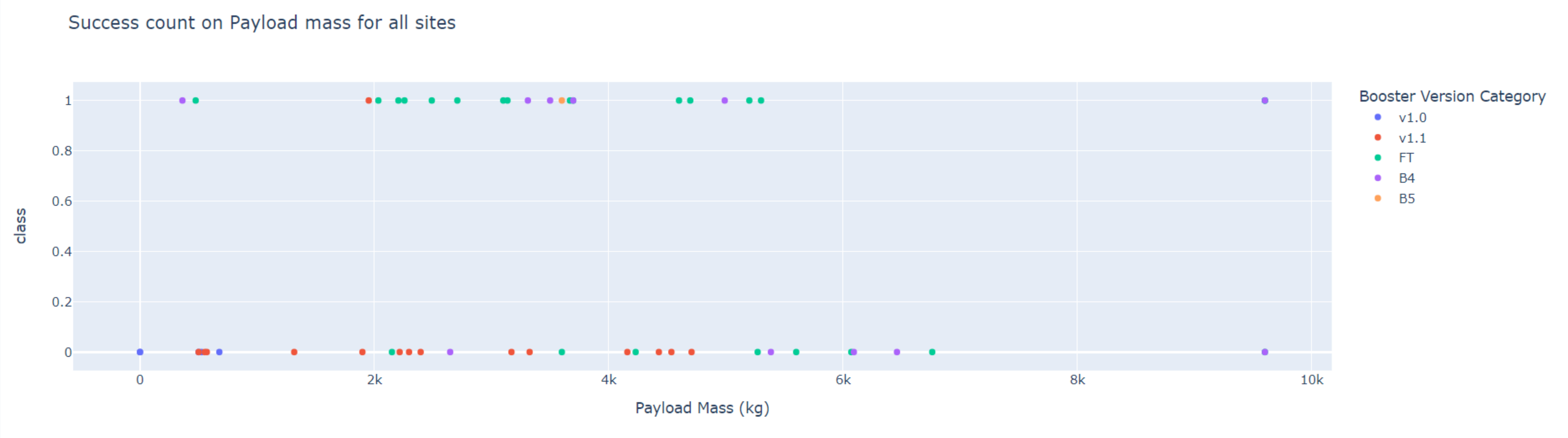
Results

- Launch Sites also had varying results with success as shown in the next slides with our Plotly Dashboard visuals

Results

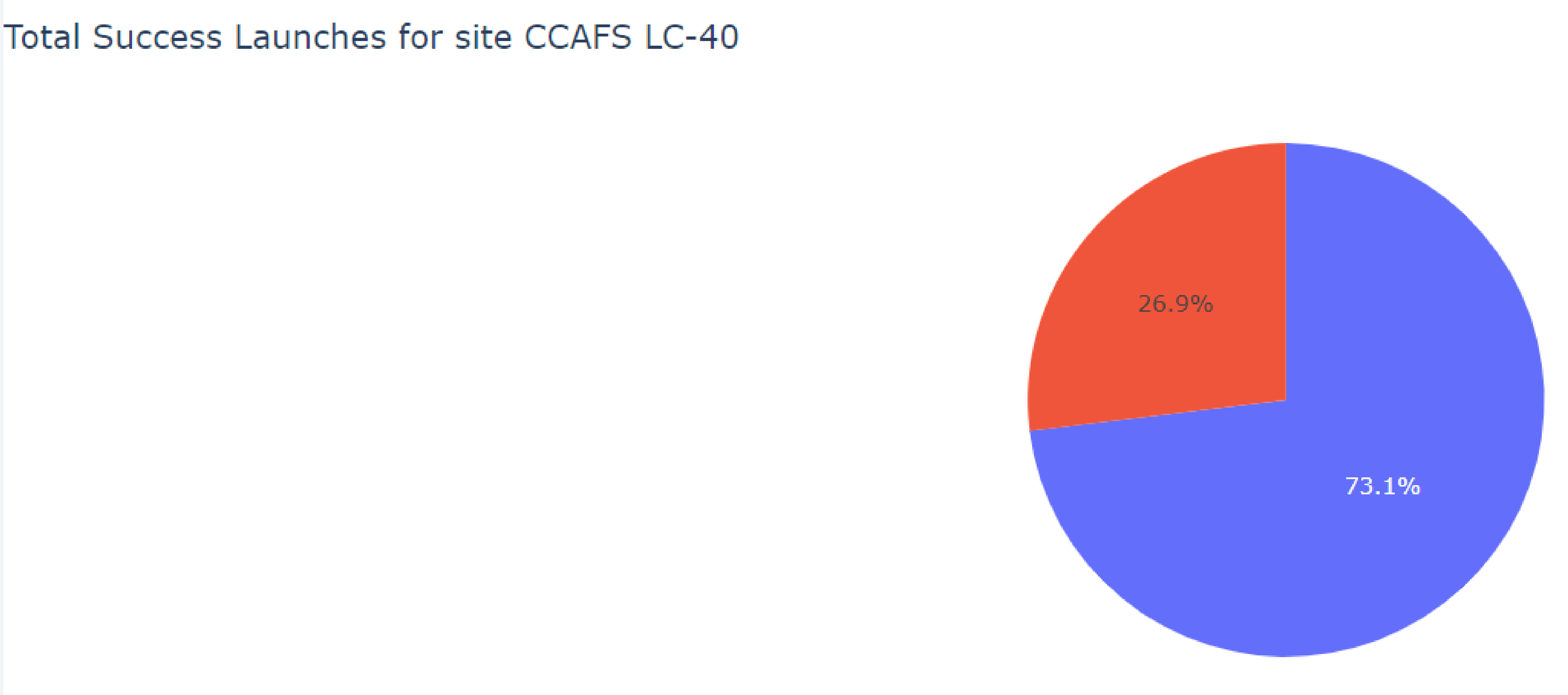

Success Count for all launch sites
41.7%
29.2%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40

Results

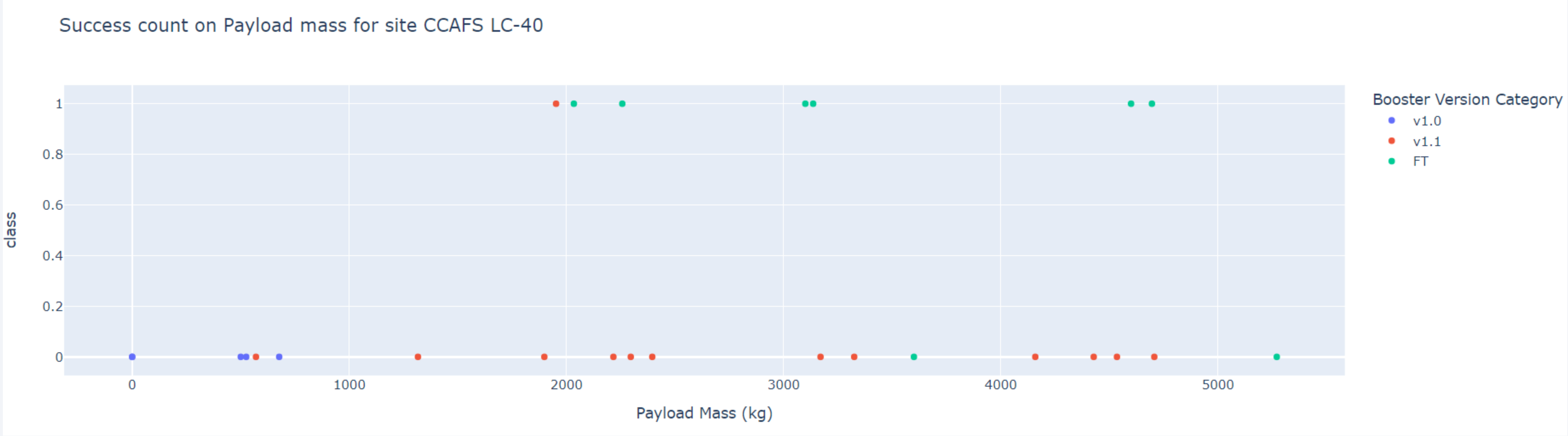


Success count on Payload mass for all sites
class
1
0.8
0.6
0.4
0.2
0
0
2k
4k
6k
8k
10k
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT
B4
B5

Results

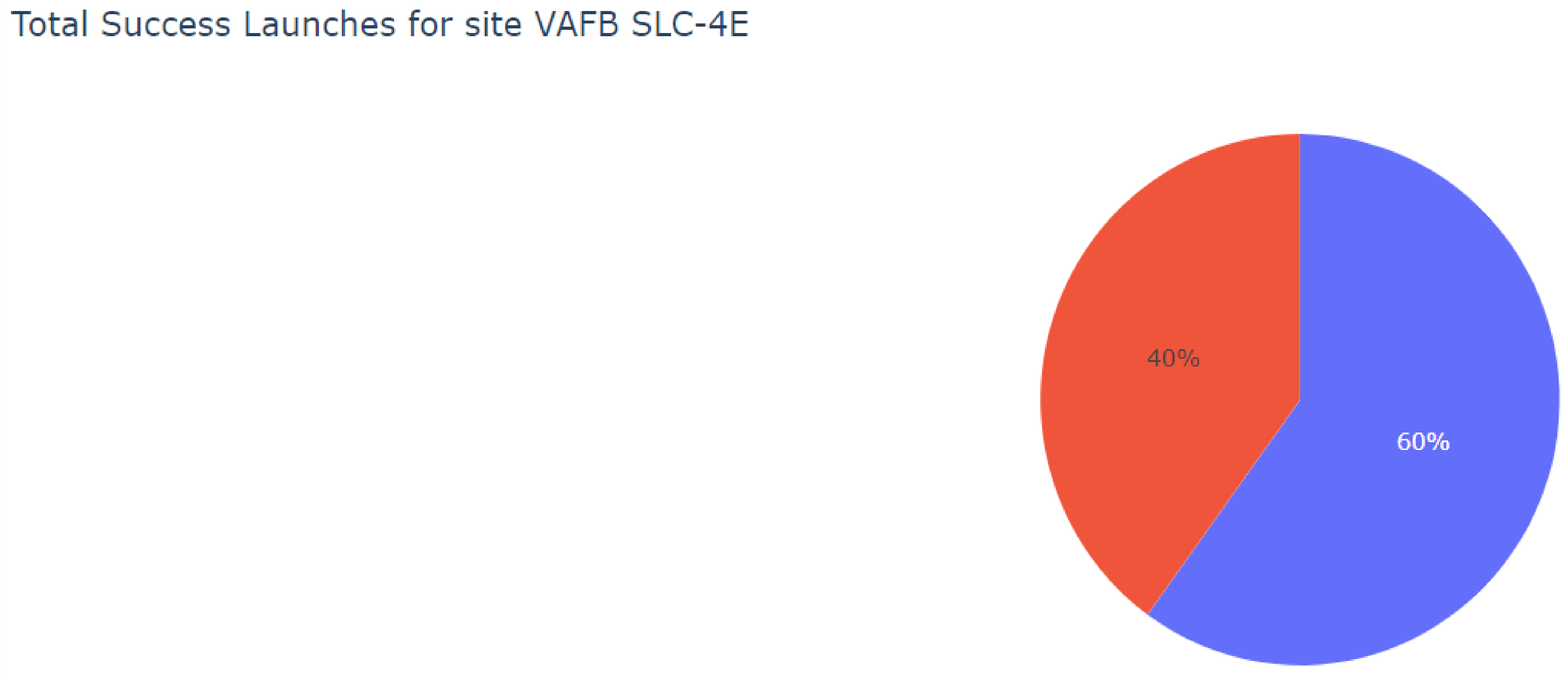


Results

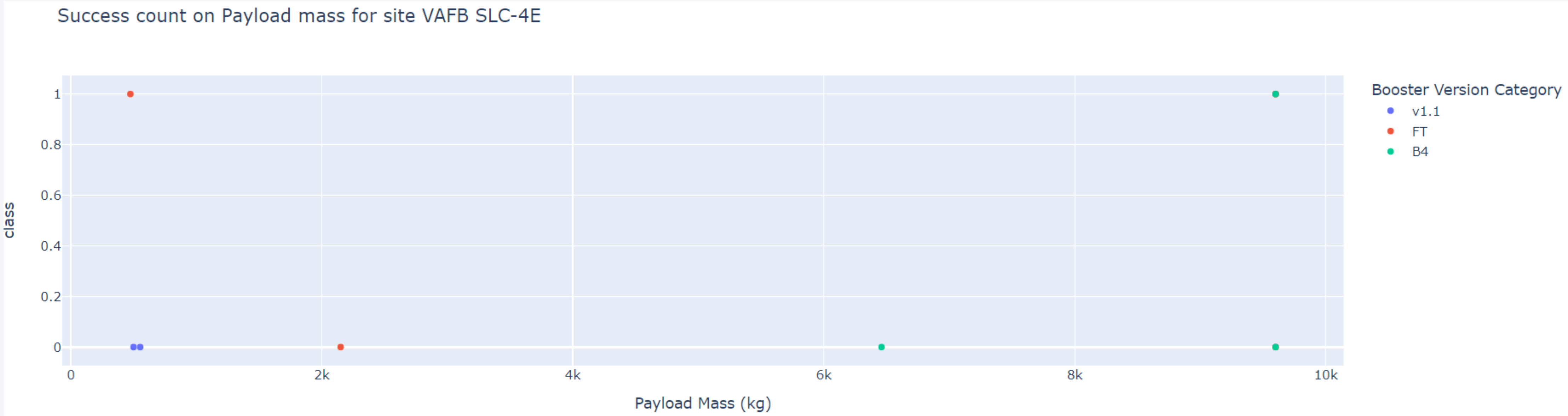


Success count on Payload mass for site CCAFS LC-40
class
1
0.8
0.6
0.4
0.2
0
0
1000
2000
3000
4000
5000
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT

Results



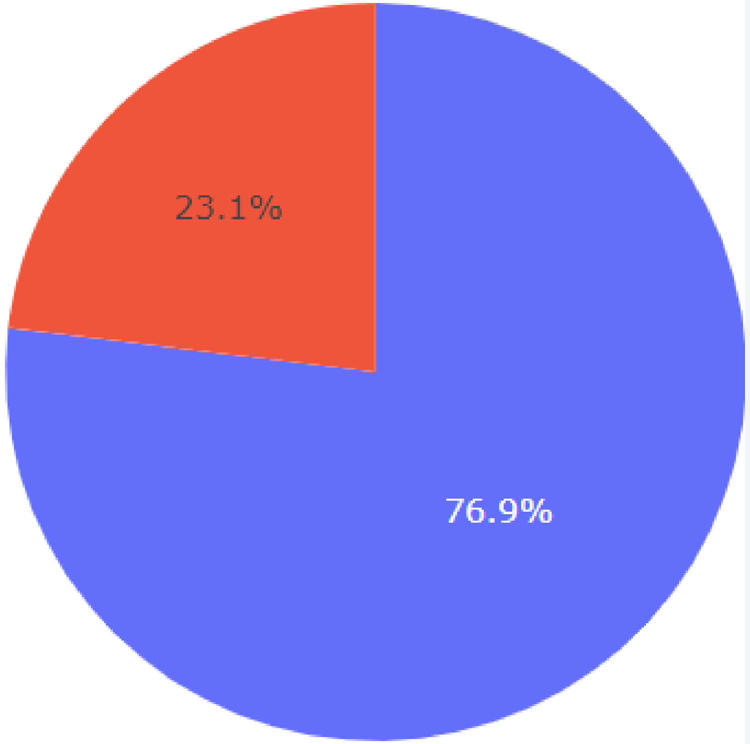
Results



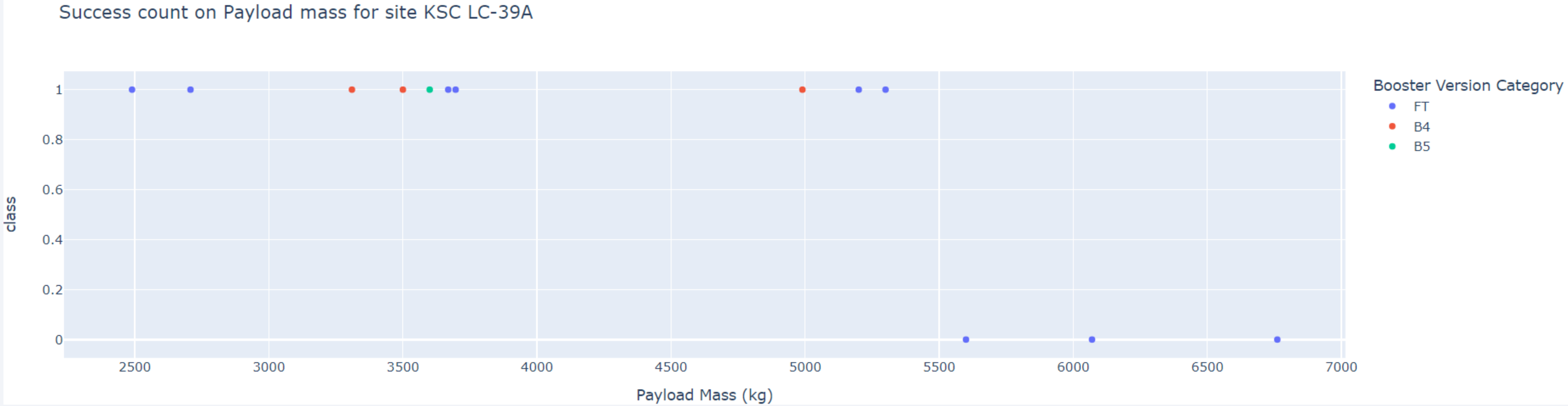
Success count on Payload mass for site VAFB SLC-4E
class
1
0.8
0.6
0.4
0.2
0
0
2k
4k
6k
8k
10k
Payload Mass (kg)
Booster Version Category
v1.1
FT
B4

Results

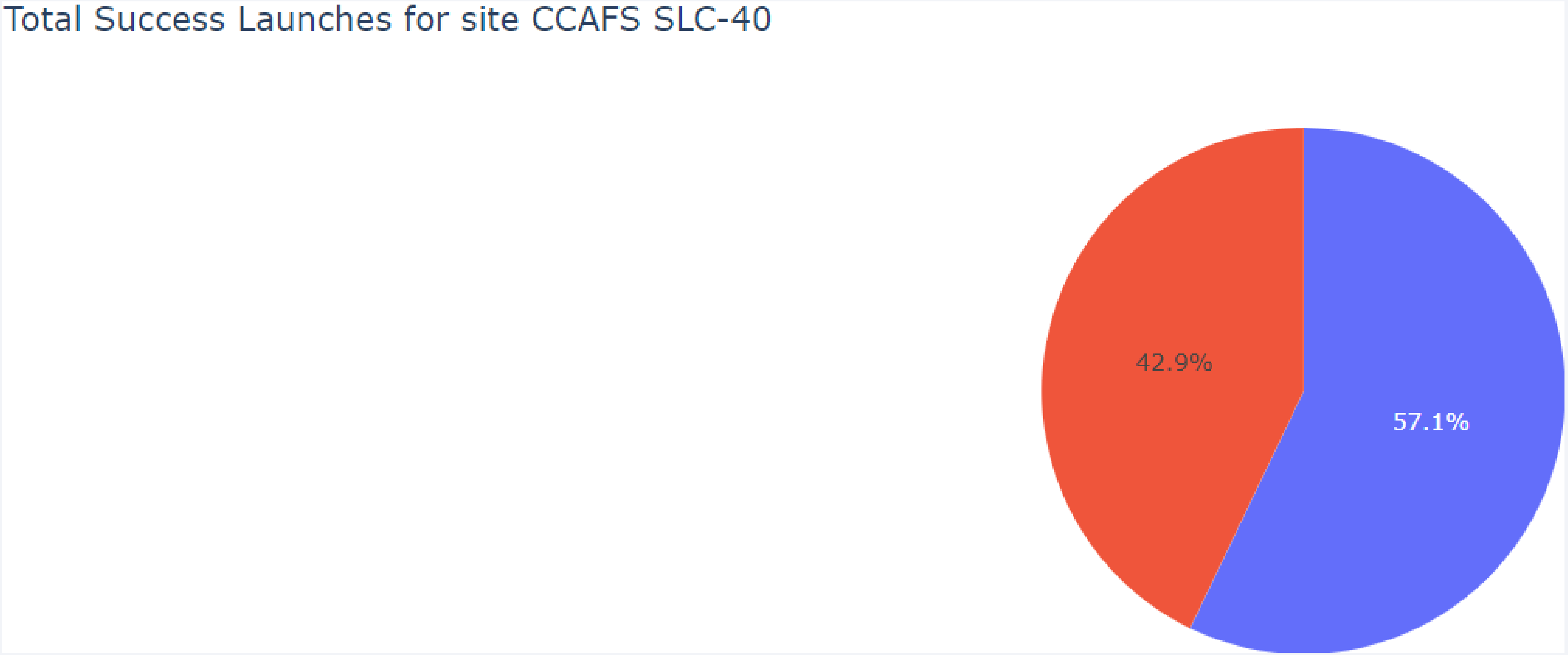
Total Success Launches for site KSC LC-39A



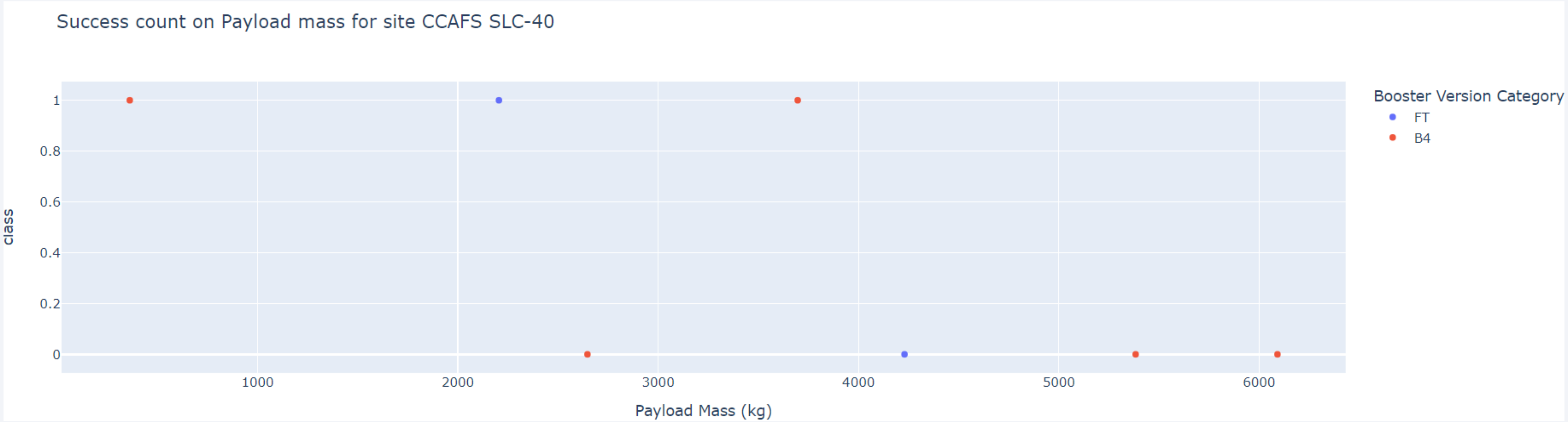
Results


Success count on Payload mass for site KSC LC-39A
class
1
0.8
0.6
0.4
0.2
0
2500
3000
3500
4000
4500
5000
5500
6000
6500
7000
Payload Mass (kg)
Booster Version Category
FT
B4
B5

Results



Results


Success count on Payload mass for site CCAFS SLC-40
class
1
0.8
0.6
0.4
0.2
0
1000
2000
3000
4000
5000
6000
Payload Mass (kg)
Booster Version Category
FT
B4

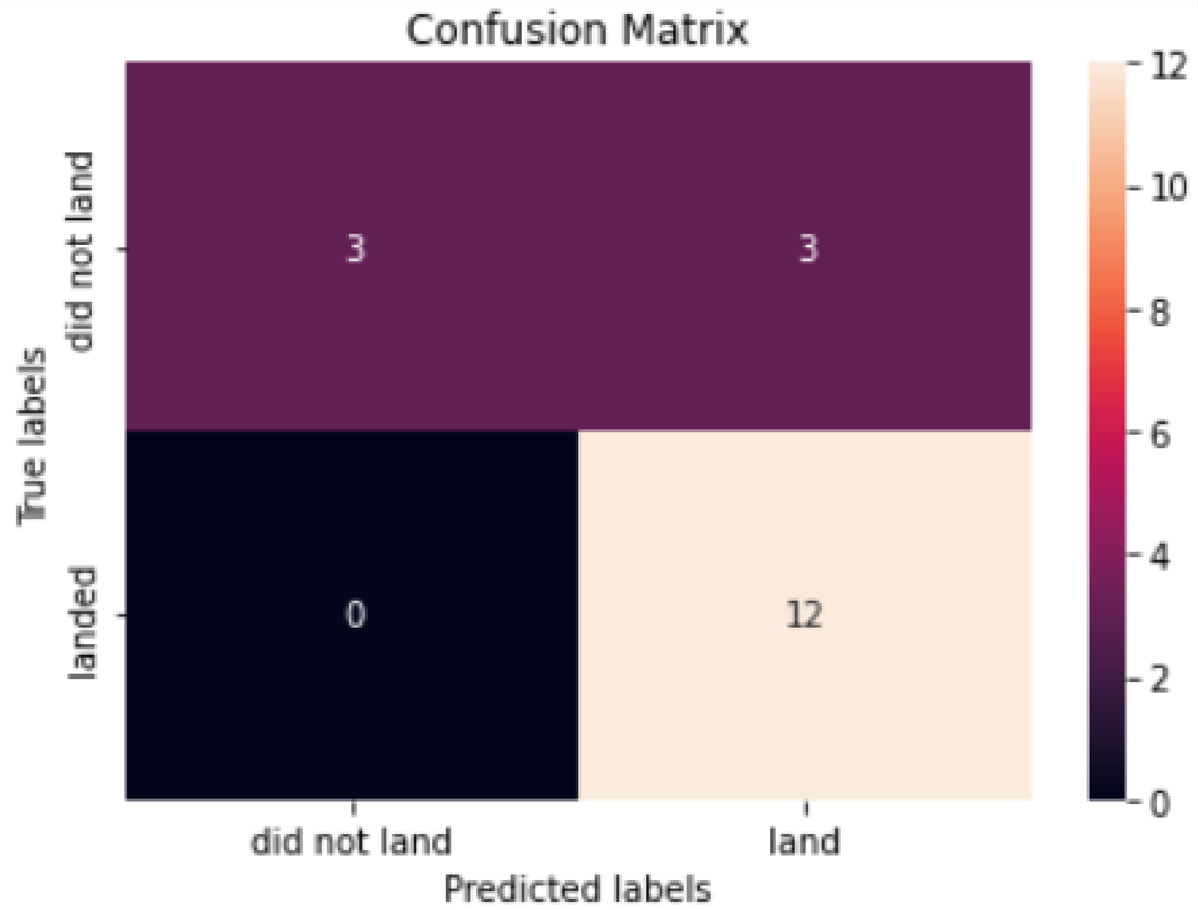
Results

- Using predictive analysis yielded similar results for each of the classification models.

- Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors all yielded the same accuracy score of 83.3%.

Results

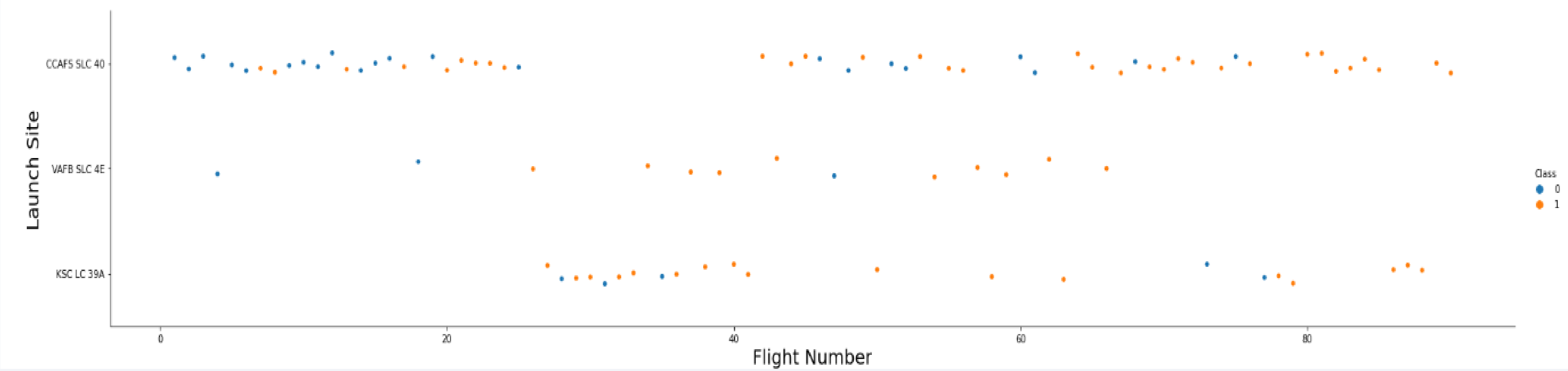
- Similarly, the Confusion Matrix for each model also yielded the same result



Section 2

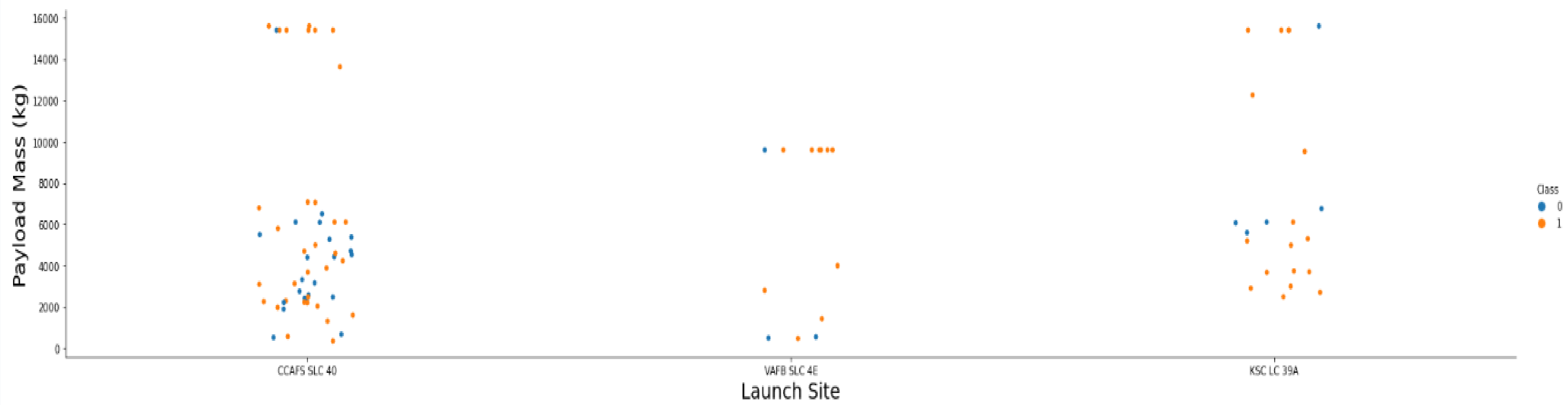
Insights drawn from EDA

Flight Number vs. Launch Site



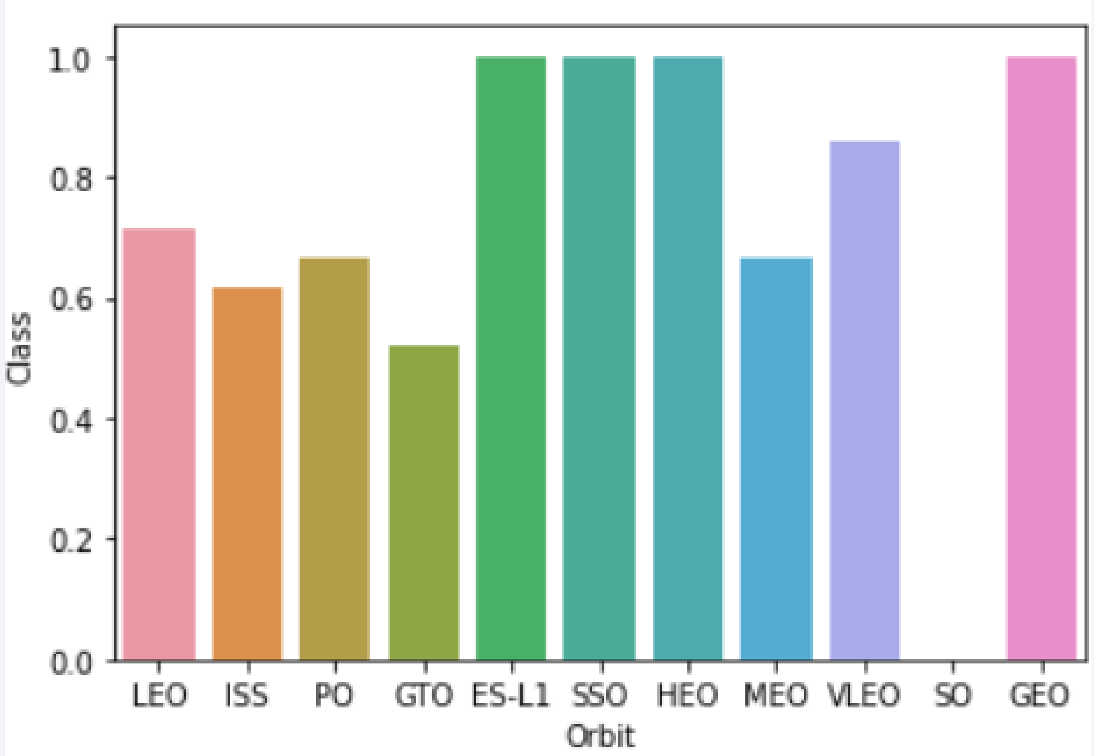
- The three launch sites shown, have higher success (orange) with more flight numbers.

Payload vs. Launch Site



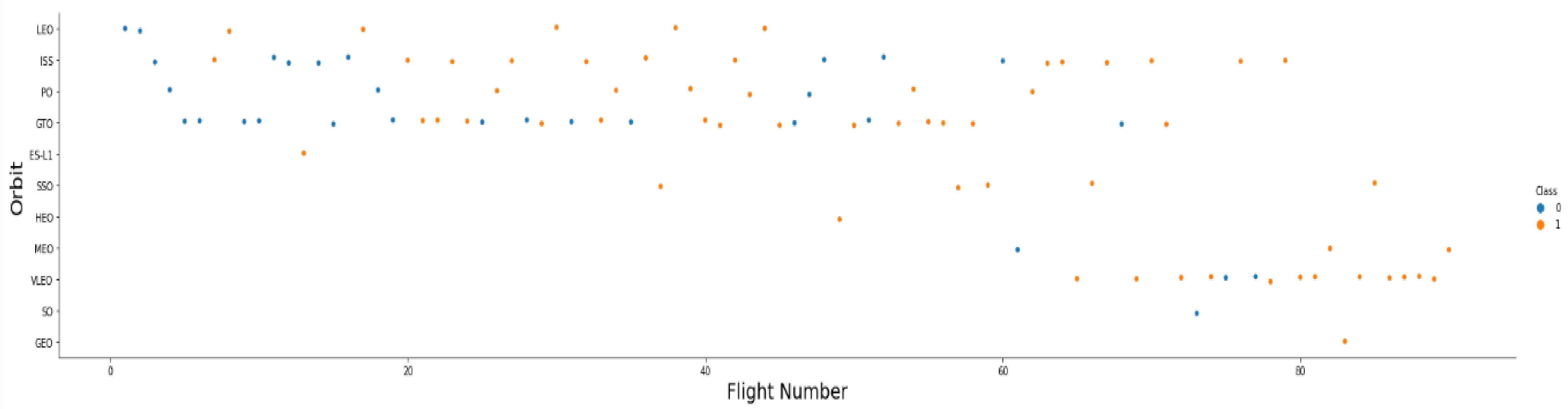
- As the payload mass increases, success (orange) increases in all three launch sites.
- Results vary with lower payload mass, as indicated by CCAFS (launch site on the left in scatterplot above)

Success Rate vs. Orbit Type



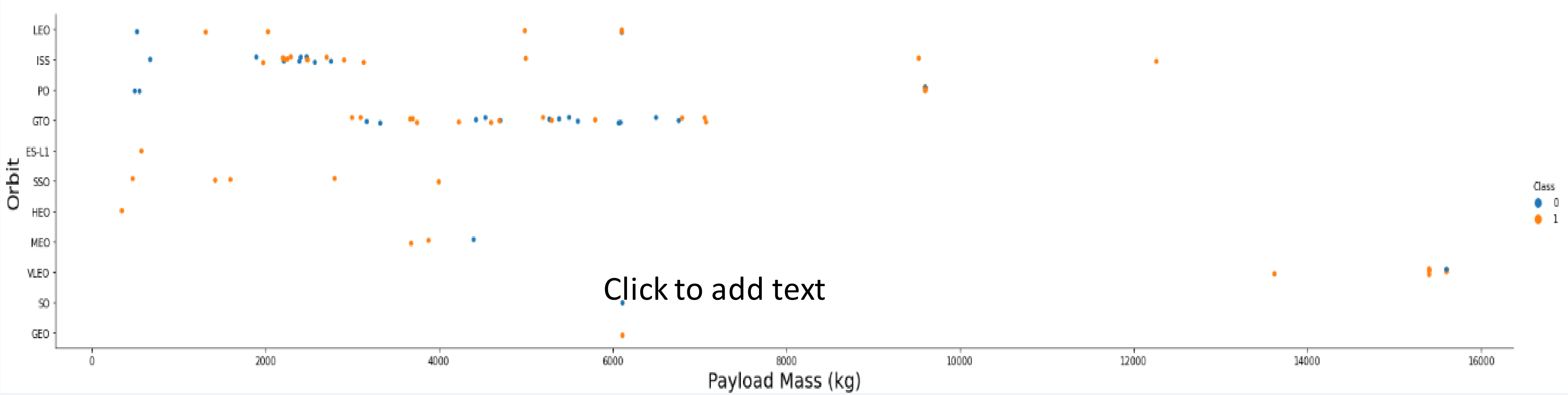
- Success rate was 100% with ES-LI, SSO, HEO, and GEO.
- High success rate achieved with VLEO, LEO, PO, and MEO.

Flight Number vs. Orbit Type



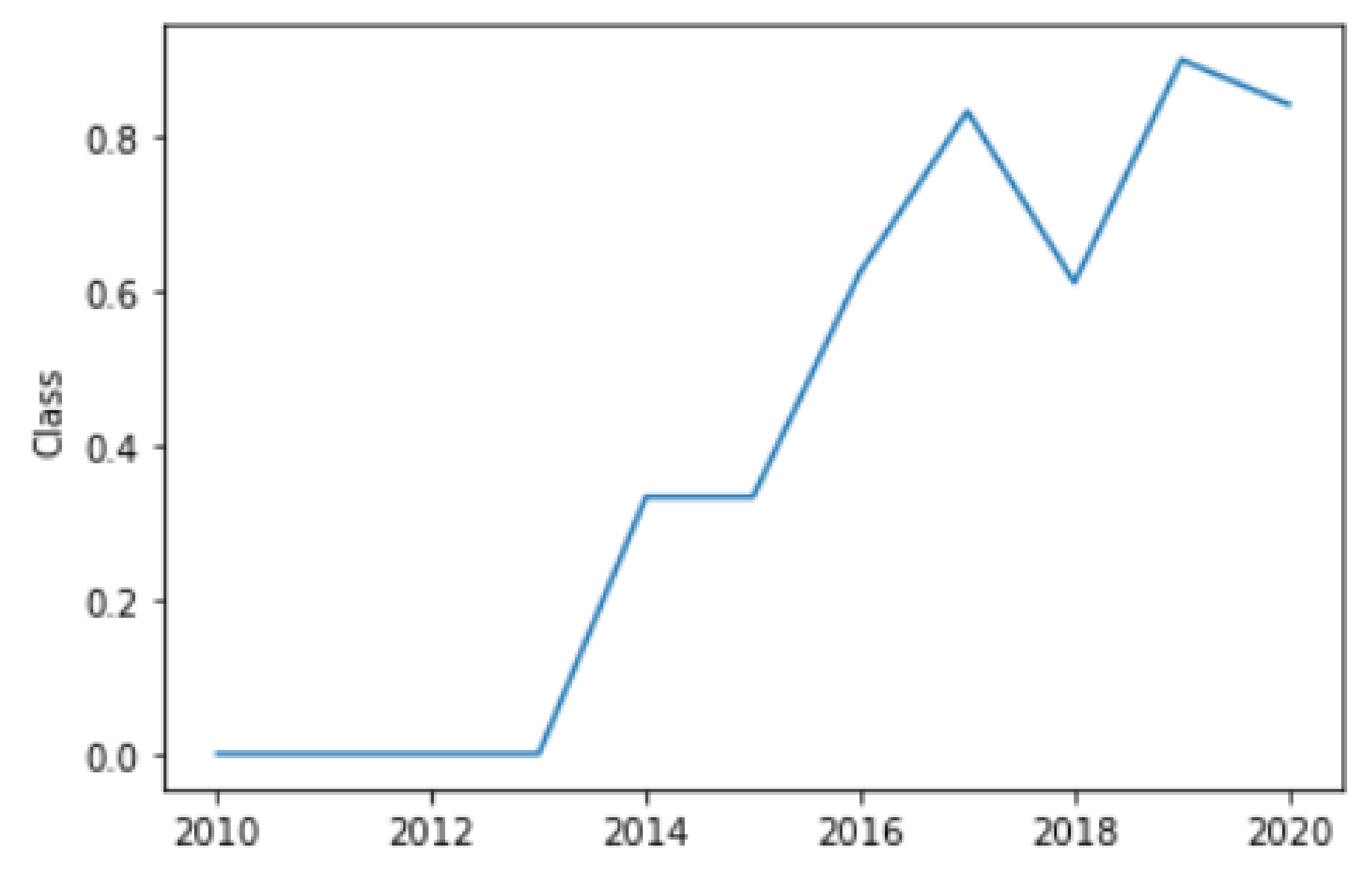
- Orbits generally have higher success rate with flight number
- LEO (top most orbit on scatterplot above) exemplifies this
- GTO does not seem to a relationship with flight number and success rate.

Payload vs. Orbit Type



- Heavy payloads give higher success rate in Polar (PO), LEO, and LSS orbits.
- GTO has no correlation between payload mass and success rate.

Launch Success Yearly Trend



- Success rates increased from 2013 to 2020, with 2019 having the highest success rate.

All Launch Site Names

In [20]:

```
sql SELECT DISTINCT launch_site FROM SPACEXDATABASE;
```

```
 * ibm_db_sa://nqx40319:***@6667d8e9-9d4d-4ccb-ba32-2
Done.
```

Out[20]:

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

- Unique Launch Sites:
- CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E
- This SQL query selects distinct launch site names from the database.

Launch Site Names Begin with 'CCA'

```
sql SELECT * FROM SPACEXDATABASE WHERE launch_site > 'CCA' LIMIT 5 ;
```

```
 * ibm_db_sa://nqx40319:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30376/bludb
Done.
```

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- Five samples were grabbed where launch sites began with "CCA"
- This provides a look at what the data looks like for "CCA".

Total Payload Mass

```
sql SELECT payload_mass__kg_ FROM SPACEXDATABASE WHERE customer = 'NASA (CRS)';
```

payload_mass__kg_
500
677
2296
2216
2395
1898
1952
3136
2257
2490

2708
3310
2205
2647
2697
2500
2495
2268
1977
2972

- The total Payload Mass carried by boosters launched by NASA: 45,596

Average Payload Mass by F9 v1.1

```
sql SELECT avg(payload_mass__kg_) FROM SPACEXDATABASE WHERE booster_version = 'F9 v1.1';

 * ibm_db_sa://nqx40319:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.dat
Done.
```

1
2928

- Average payload mass carried by F9 v1.1: 2928
- This query gets the average of column payload_mass__kg_ for booster version F9 v1.1

First Successful Ground Landing Date

```
sql SELECT min(DATE) FROM SPACEXDATABASE WHERE landing__outcome = 'Success (ground pad)';

 * ibm_db_sa://nqx40319:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.data
Done.
```

1
2015-12-22

- The first successful landing outcome on ground pad was December 12th, 2015
- This query finds the first date for landing outcome where a successful ground landing pad occurred.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
sql SELECT booster_version FROM SPACEXDATABASE WHERE landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ < 60
```

```
 * ibm_db_sa://nqx40319:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30376/bludb
Done.
```

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

- The names of the boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:
- F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2
- This query finds booster version where landing outcome (success drone ship) and weight between 4000kg and 6000kg payload mass

Total Number of Successful and Failure Mission Outcomes

```
sql SELECT COUNT(mission_outcome) FROM SPACEXDATABASE WHERE mission_outcome = 'Success';
```

```
 * ibm_db_sa://nqx40319:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdo
Done.
```

1
99

```
sql SELECT COUNT(mission_outcome) FROM SPACEXDATABASE WHERE mission_outcome = 'Failure (in flight)';
```

```
 * ibm_db_sa://nqx40319:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdo
Done.
```

1
1

- 99 successful missions, and 1 failed mission in flight

Boosters Carried Maximum Payload

```
sql SELECT booster_version FROM SPACEXDATABASE WHERE payload_mass__kg_ =(SELECT MAX(payload_mass__kg_) FROM SPACEXDATABASE);
```

```
 * ibm_db_sa://nqx40319:***@6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30376/bludb
Done.
```

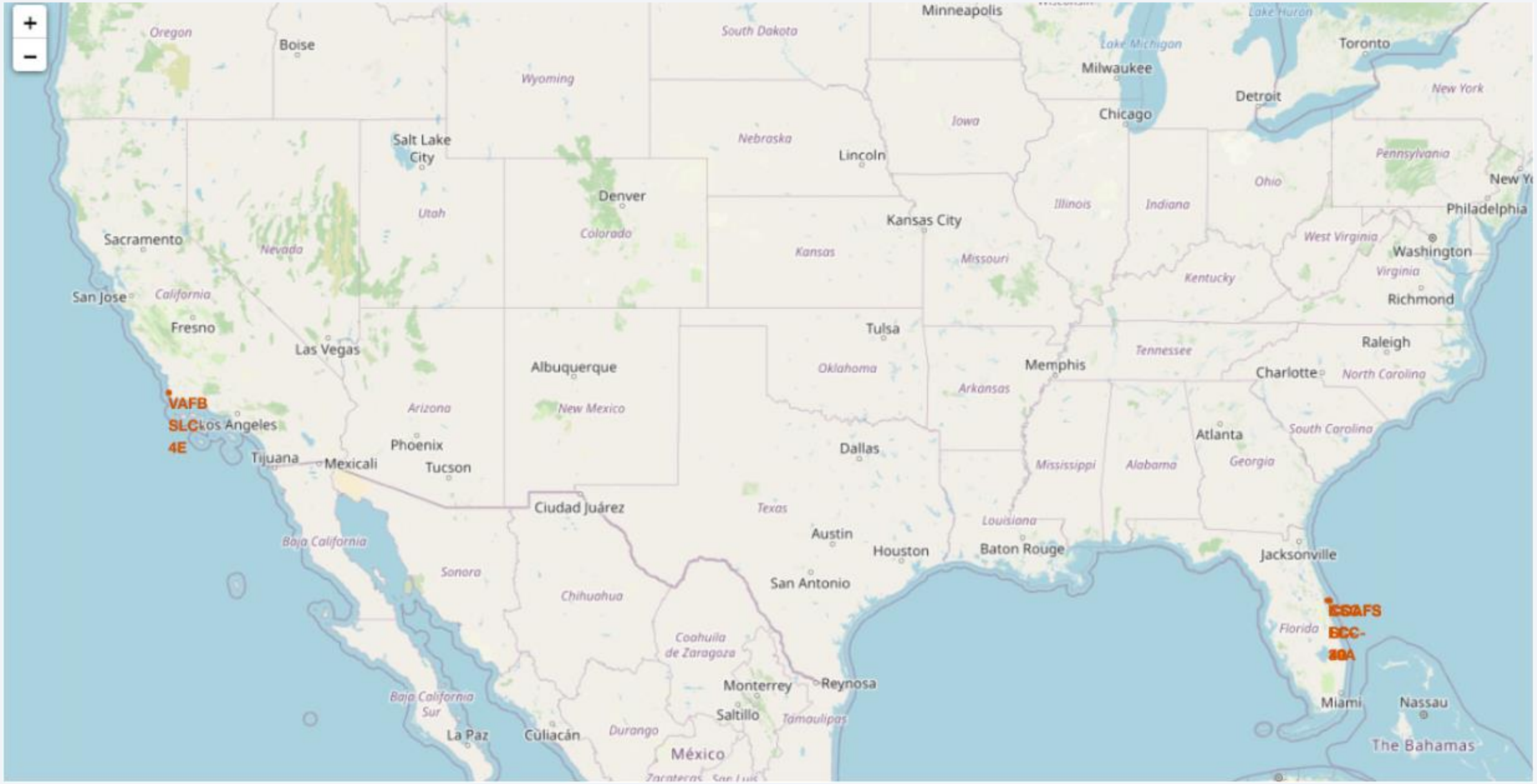
booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

- 12 F9 B5 boosters carried the maximum payload (15600kg)

Section 3

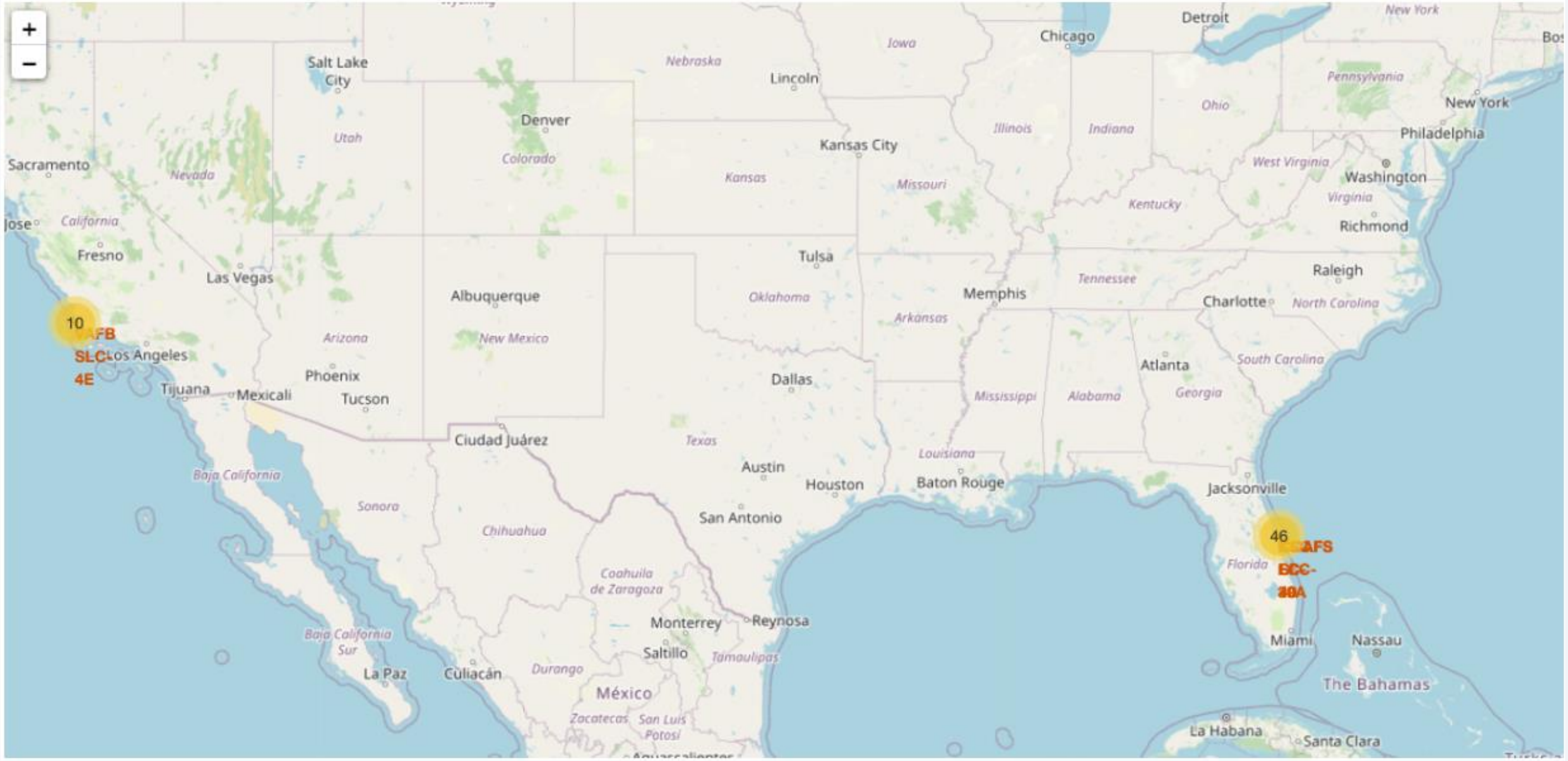
Launch Sites Proximities Analysis

All Launch Sites on Folium Map



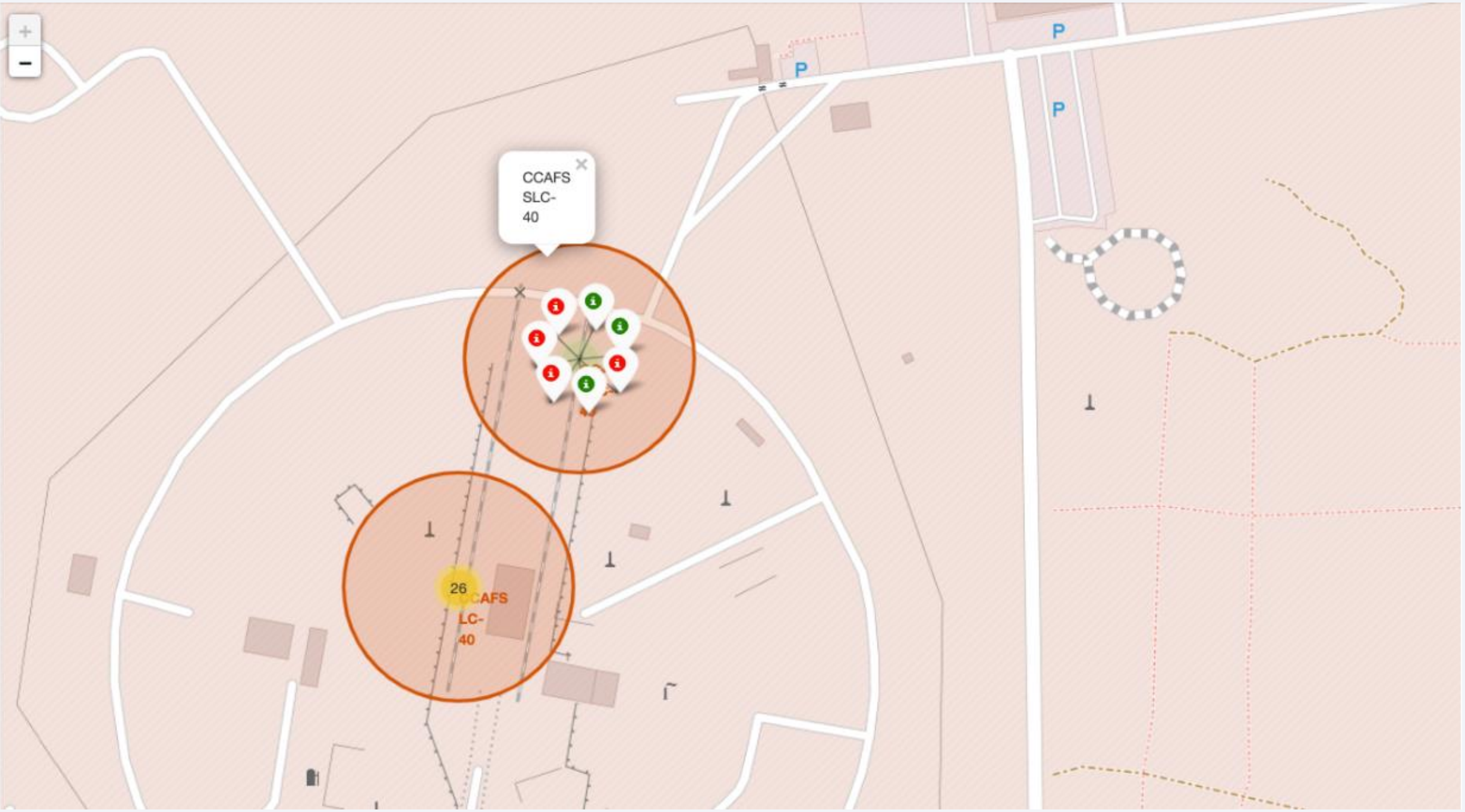
- This screenshot shows the launch sites on a Folium map, Florida and California were SpaceX launch areas.

Number of Launches per State



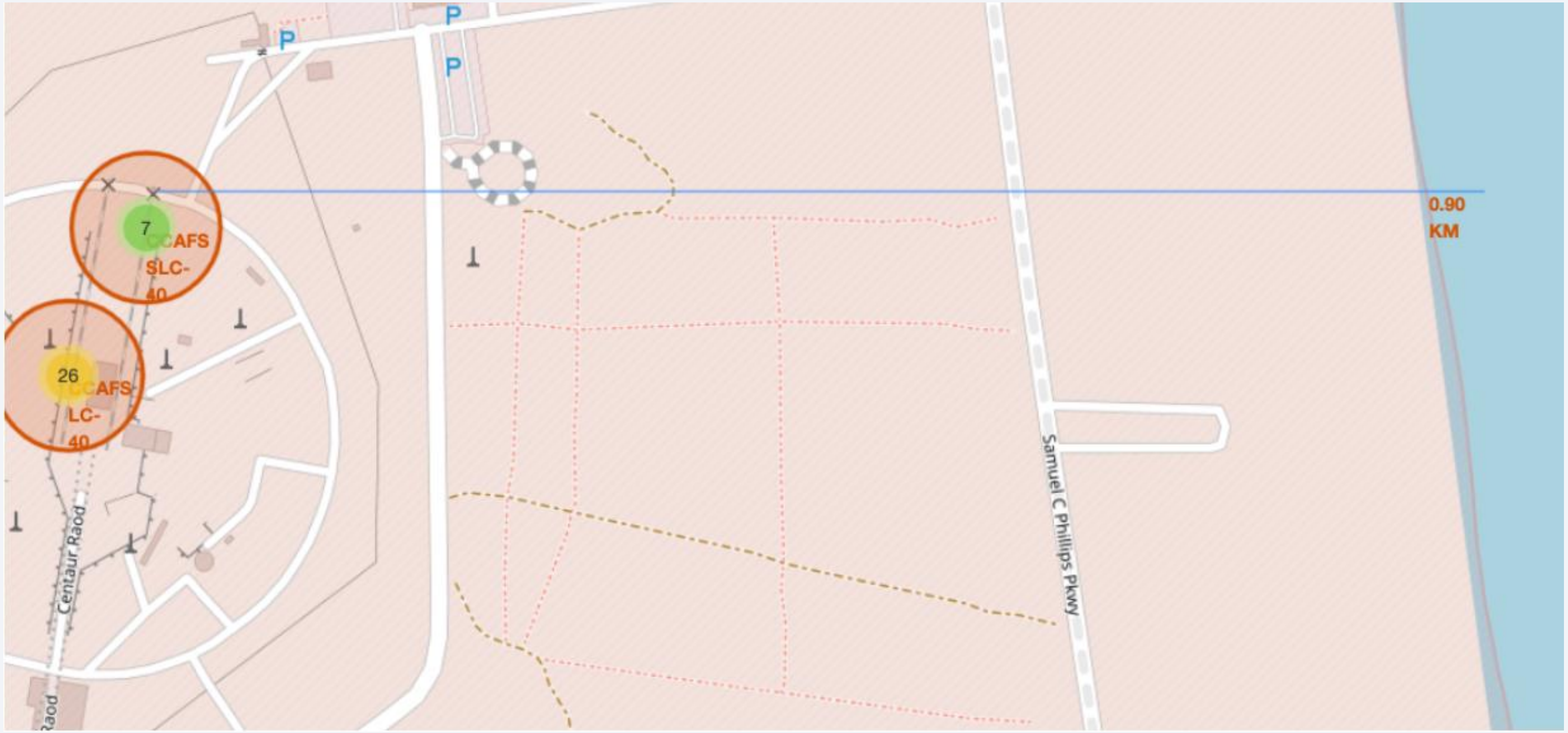
There are 46 launches in Florida, and 10 launches in California.

Zoomed in Launch Sites Success Rate



Zooming in on the map shows which launch sites have relatively high success rates.

Distance from Coastline for Launch Site in Florida



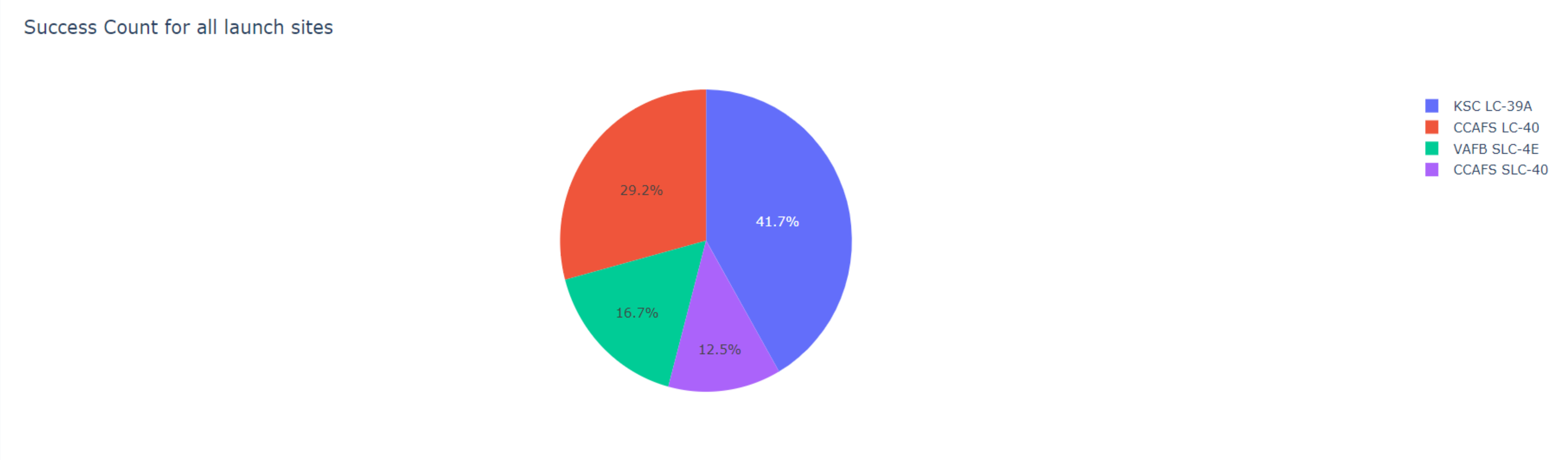
This map shows the distance from the coastline for the launch site area in Florida, 0.90KM.

This helps to show where SpaceX needs to go to retrieve any successful booster landings out at sea.

Section 4

Build a Dashboard with Plotly Dash

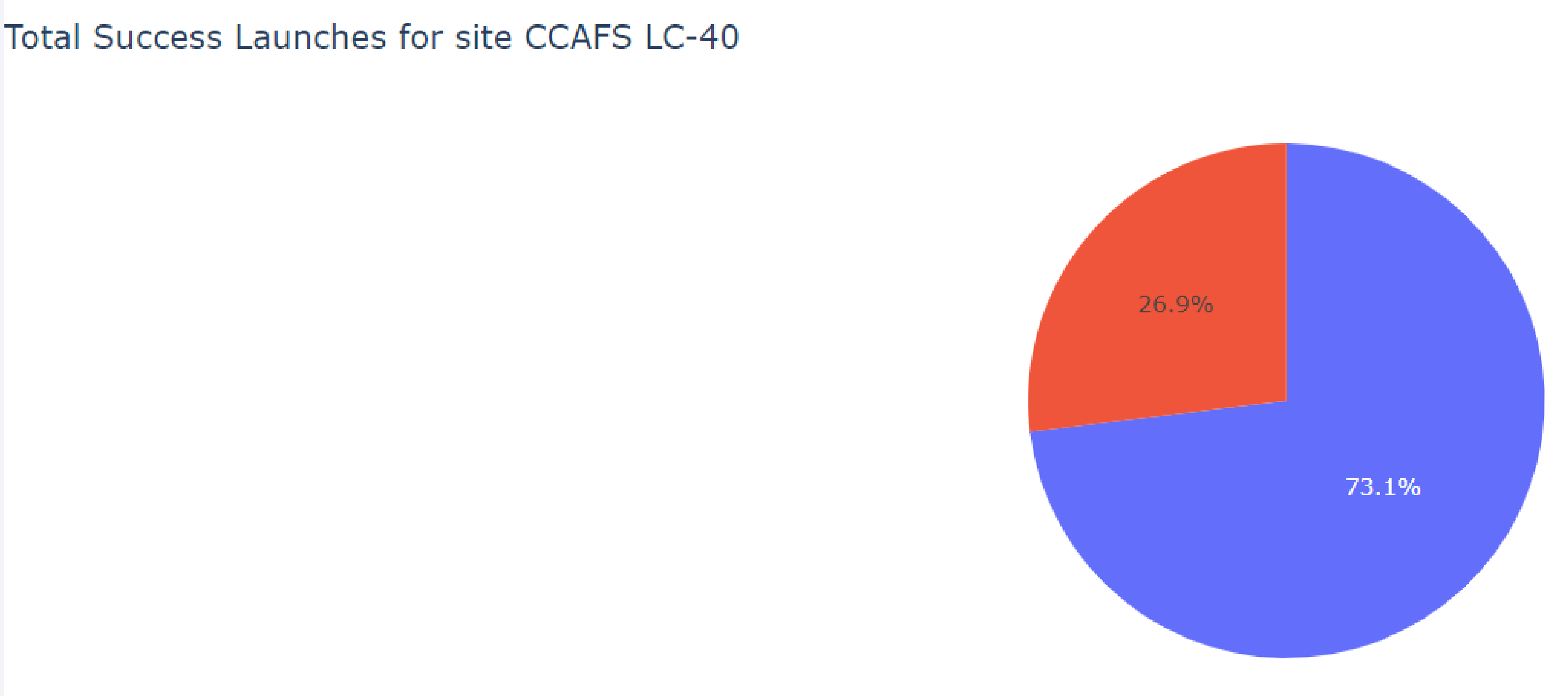
Success Rate for All Launch Sites



This pie chart shows the success rate for all launch sites,

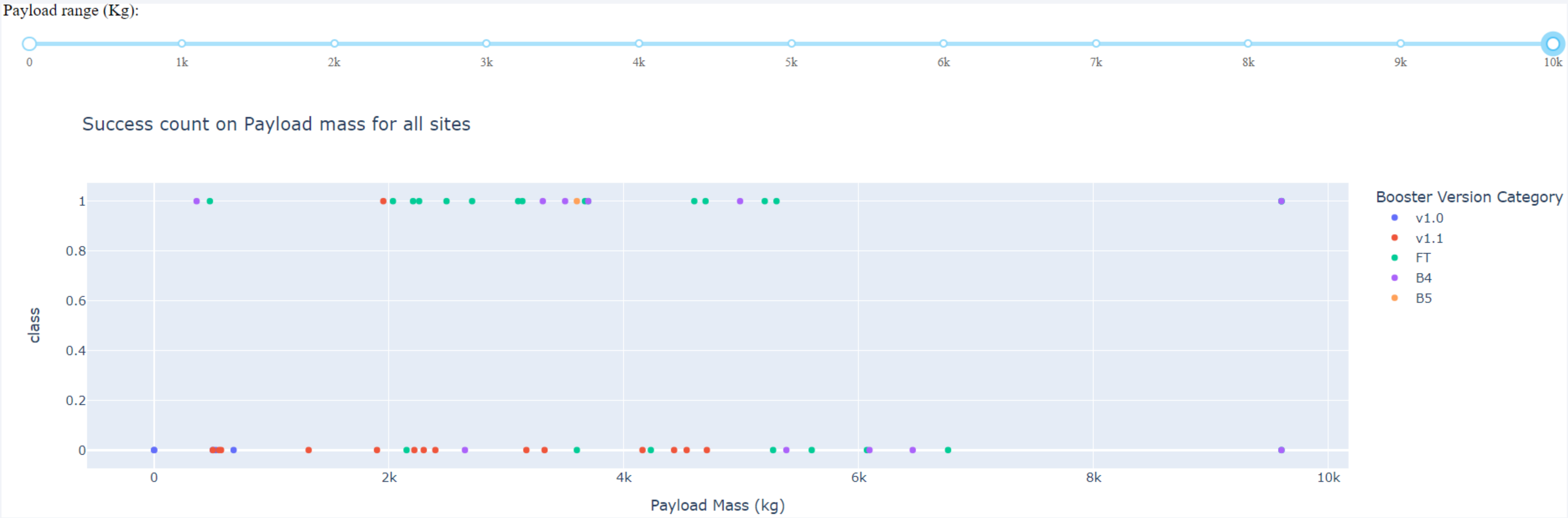
KSC has highest with 41.7%, CCAFS LC-40 has second highest with 29.2%.

Success to Fail Rate for CCAFS LC-40

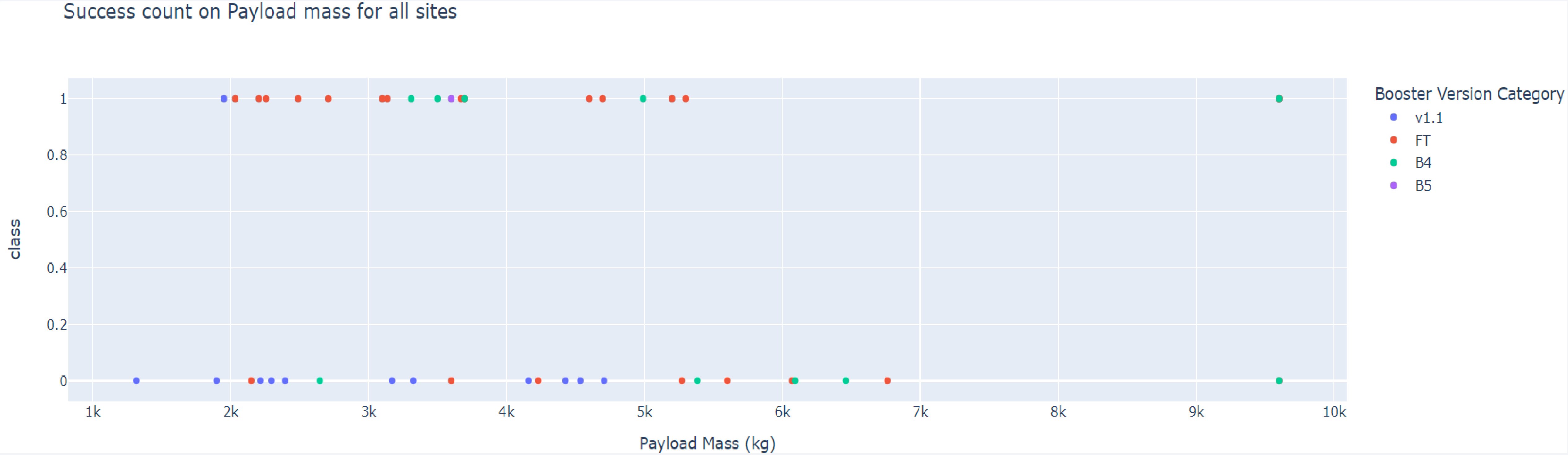


The Success to Fail ratio for CCAFS LC-40 (highest success of all launch sites), is 73.1 success to 26.9% fail.

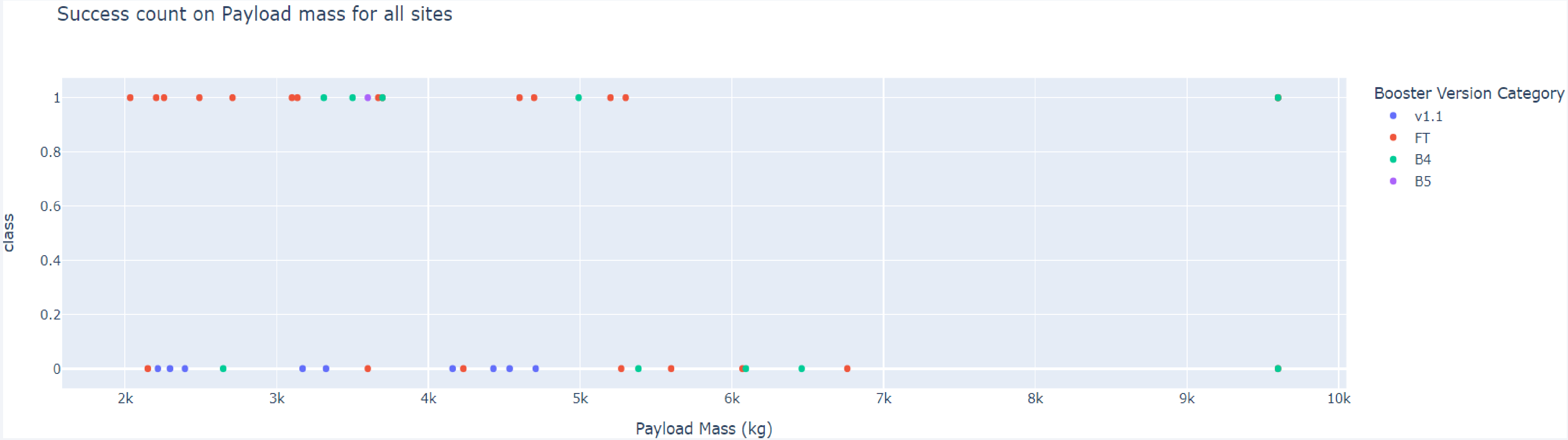
All Payload Masses for All Sites



Payload Mass 1KG to 10KG for All Sites

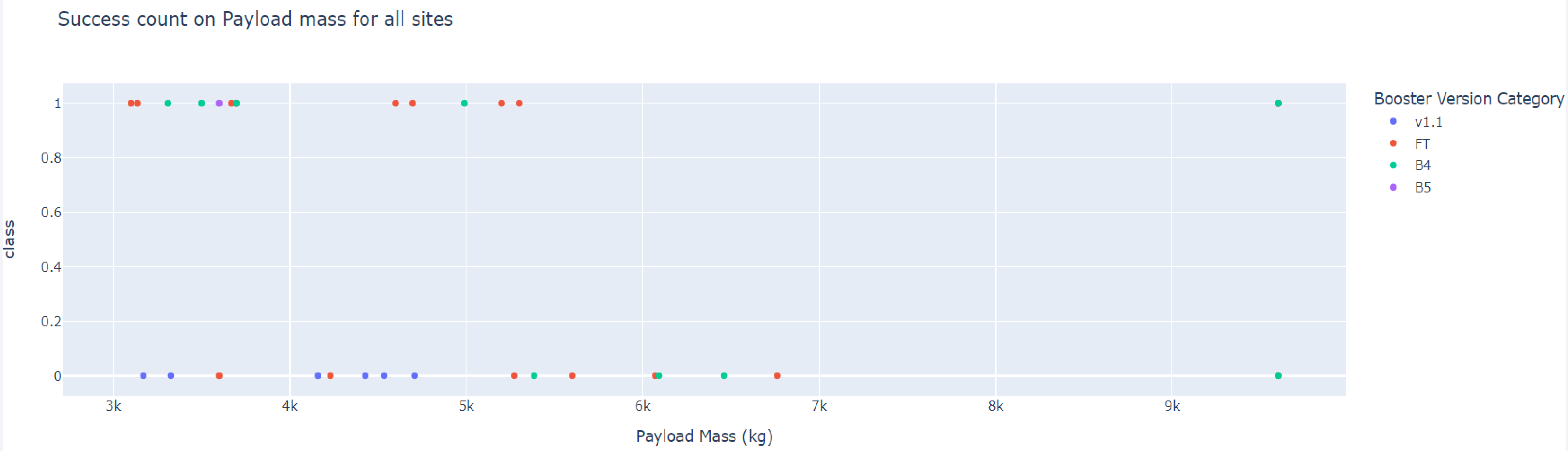


Payload Mass 2KG to 10KG for All Sites



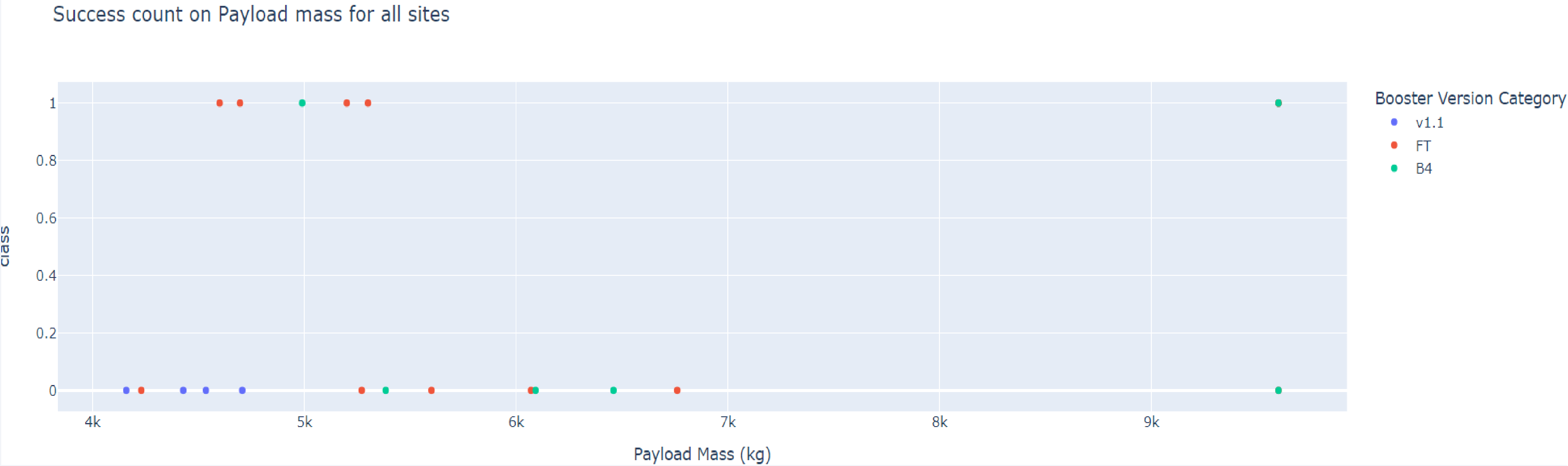
- Starting to see less failed outcomes with higher payload mass.

Payload Mass 3KG to 10KG for All Sites



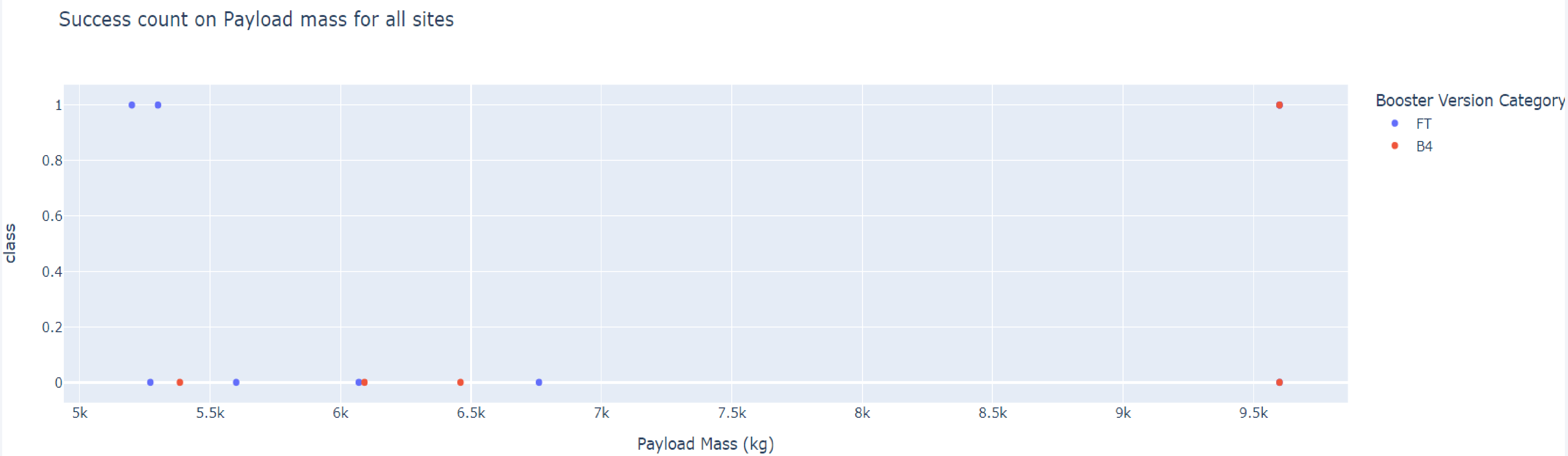
- Starting to see less failed outcomes with higher payload mass.

Payload Mass 4KG to 10KG for All Sites



- As we increase the payload mass range, we now see less success.
- Notably v1.1 has no success outcomes with higher payload masses.

Payload Mass 5KG to 10KG for All Sites

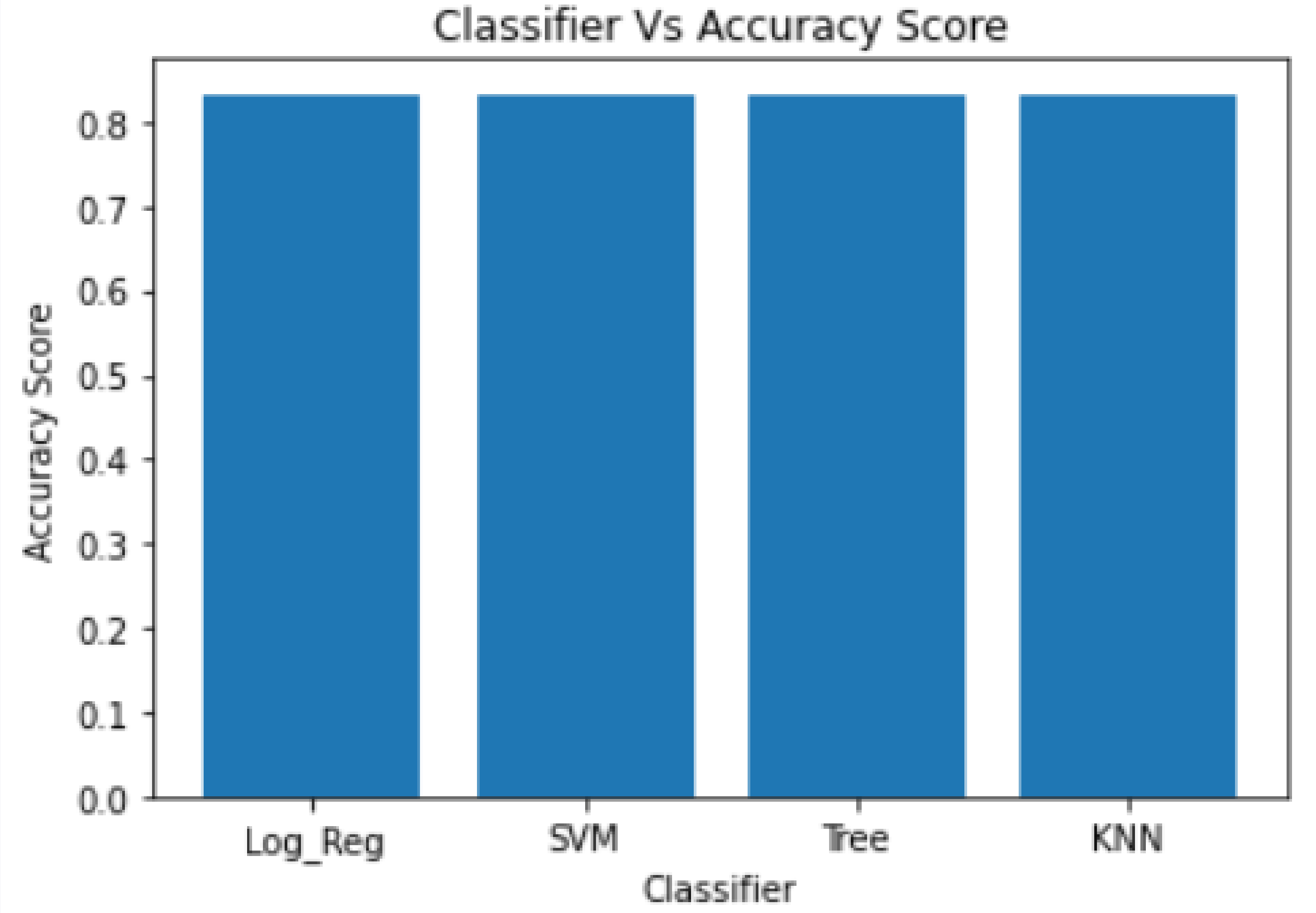


- Boosters are now only FT and B4.
- B4 is the only successful booster above 9KG. Testing not done with FT above 7KG

Section 5

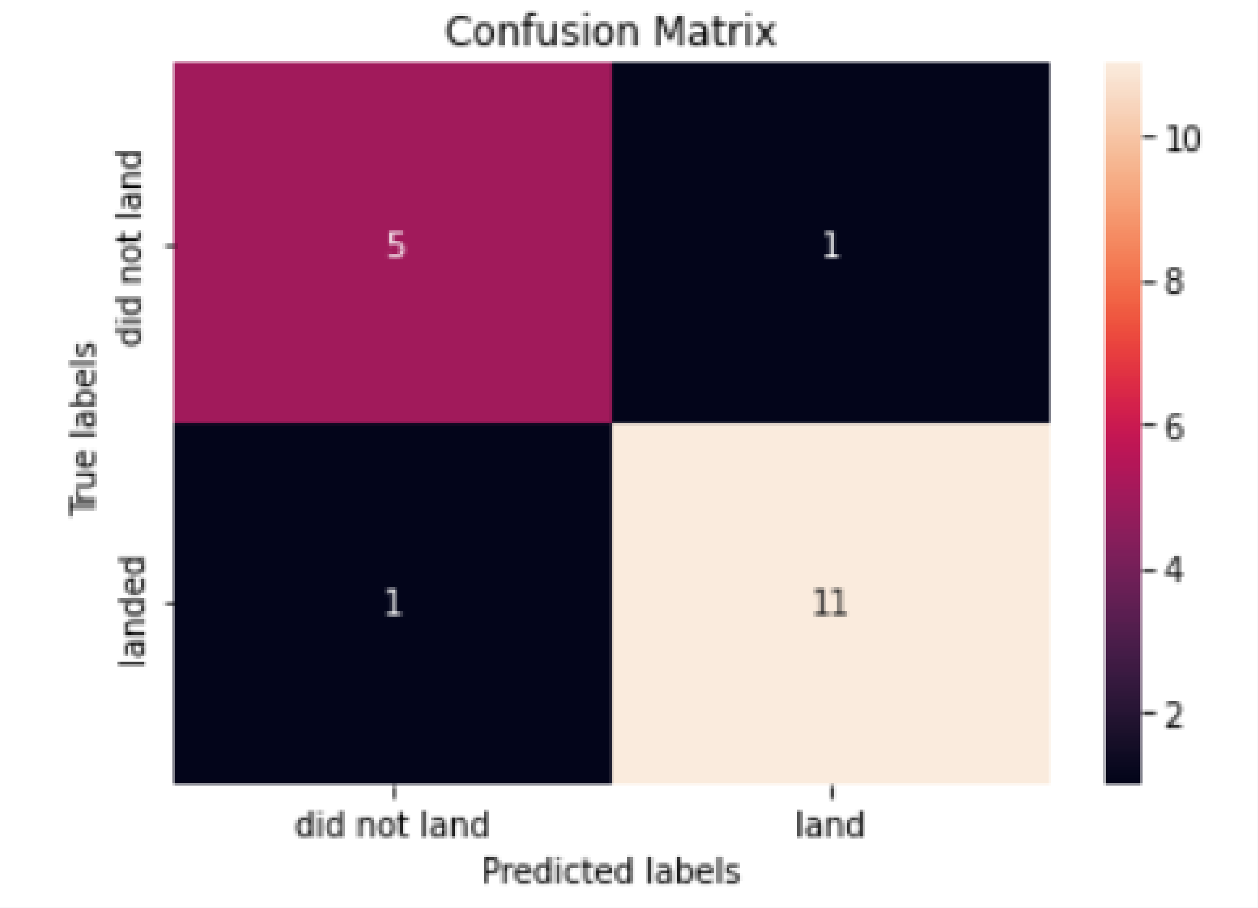
Predictive Analysis (Classification)

Classification Accuracy

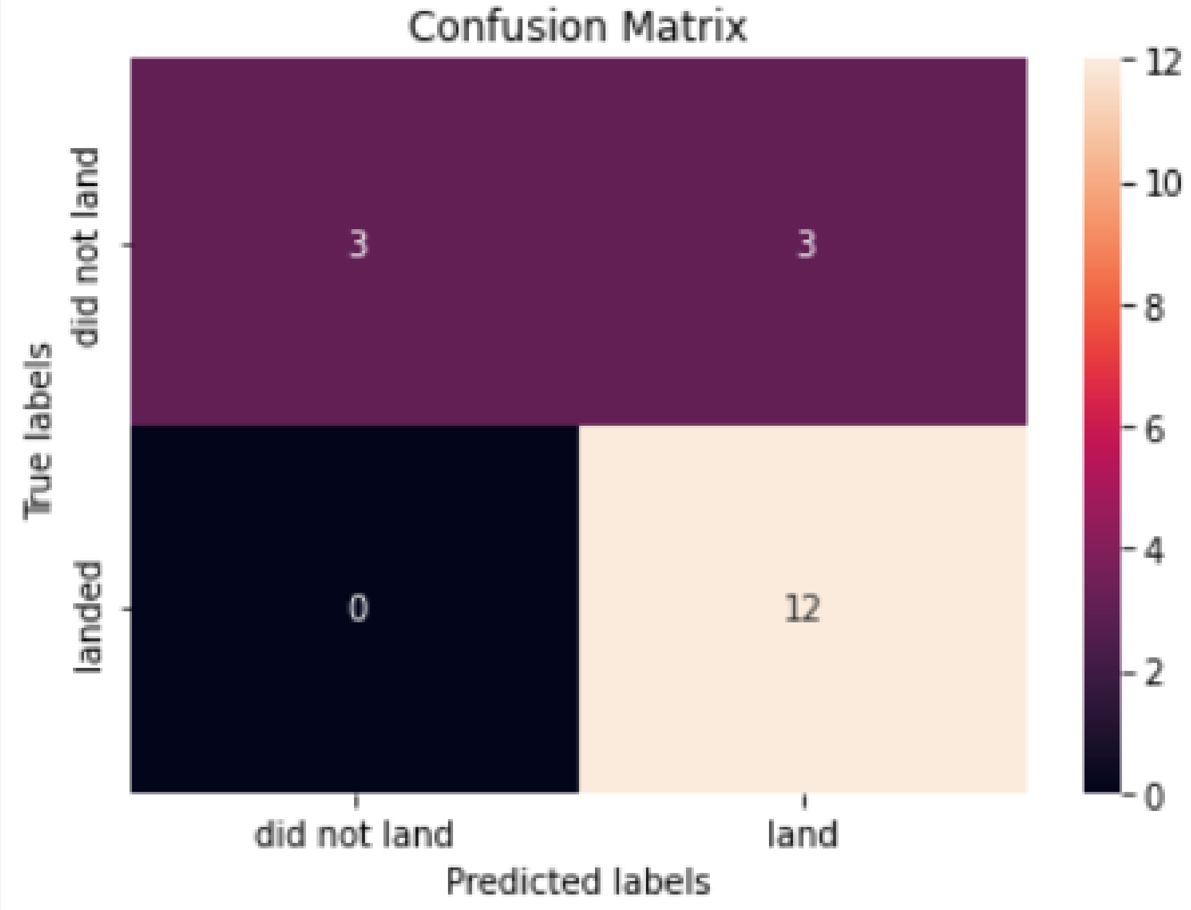


The accuracy score was the same for all classifiers, 83.3%.
Low sample size may be the cause of this.

Confusion Matrix



The Decision Tree had the best Confusion Matrix.



The other classifiers, had the same Confusion Matrix results.

Conclusions

- Orbits had a role to play in successful launches
- Payload Mass mattered for some Orbits but not others.
- Only a couple boosters had high payload masses, limiting results of success and failed launches.
- Different launch sites have varying levels of success.
    - KSC LC-39A and VAFB SLC 4E have success of 77%, while CCAFS LC-40 has success of 60%

Successful launches have increased in recent years.

Due to the cost of launches, launches are expensive to complete.

- More data would provide better prediction success.

Appendix

- For all links to notebooks, please refer to my GitHub below:
- github.com/steezkelly/coursera-capstone-project

Thank you!