IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

SOLOMON KOBBY
EWUDZIE

8 APR 23

Outline

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

Executive Summary

- There is need to determine using Data Science if SpaceX will reuse the first stage of its space rockets to assist Space Y compete.
- Data collection, Data wrangling was performed to acquire workable data. Exploratory data analysis (EDA) was performed using visualization and SQL. Interactive visual analytics was performed using Folium and Plotly Dash. Predictive analysis was performed using classification models.
- It was found that Booster version FT with payload between 2000 and 6000 is more likely to be reused. KSC-LC-39A is the launch site with highest success count and success ratio. Lower orbit travels have higher success rates.
- Predictive models had a good accuracy (83%) in classifying whether a stage 1 would land successfully to be reused.

Introduction

- Stage one does most of the work in moving a spaceship to space. This stage is quite large and expensive. Unlike other rocket providers, SpaceX's Falcon 9 Can sometimes recover the first stage thereby reducing overall costs by half. At other times it will crash
- Space Y that would like to compete with SpaceX.
- Need to gather information about Space X and create dashboards for the team.
- Need to determine using Data Science if SpaceX will reuse the first stage.
- This will assist Space Y to know how to compete in upcoming launches.

Section 1

Methodology

Methodology

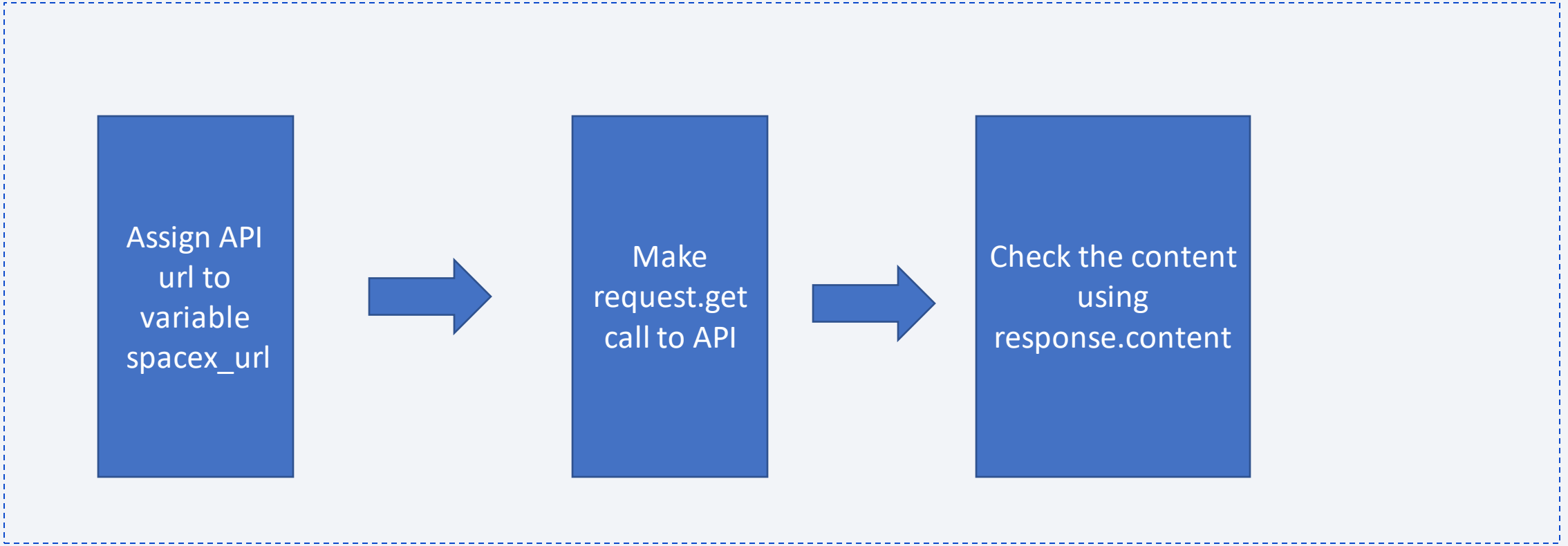
Executive Summary

- Data collection methodology:
  - Describes how data was collected
- Perform data wrangling
  - Describes how data was processed
- 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 classification models were built, tuned and evaluated

Data Collection

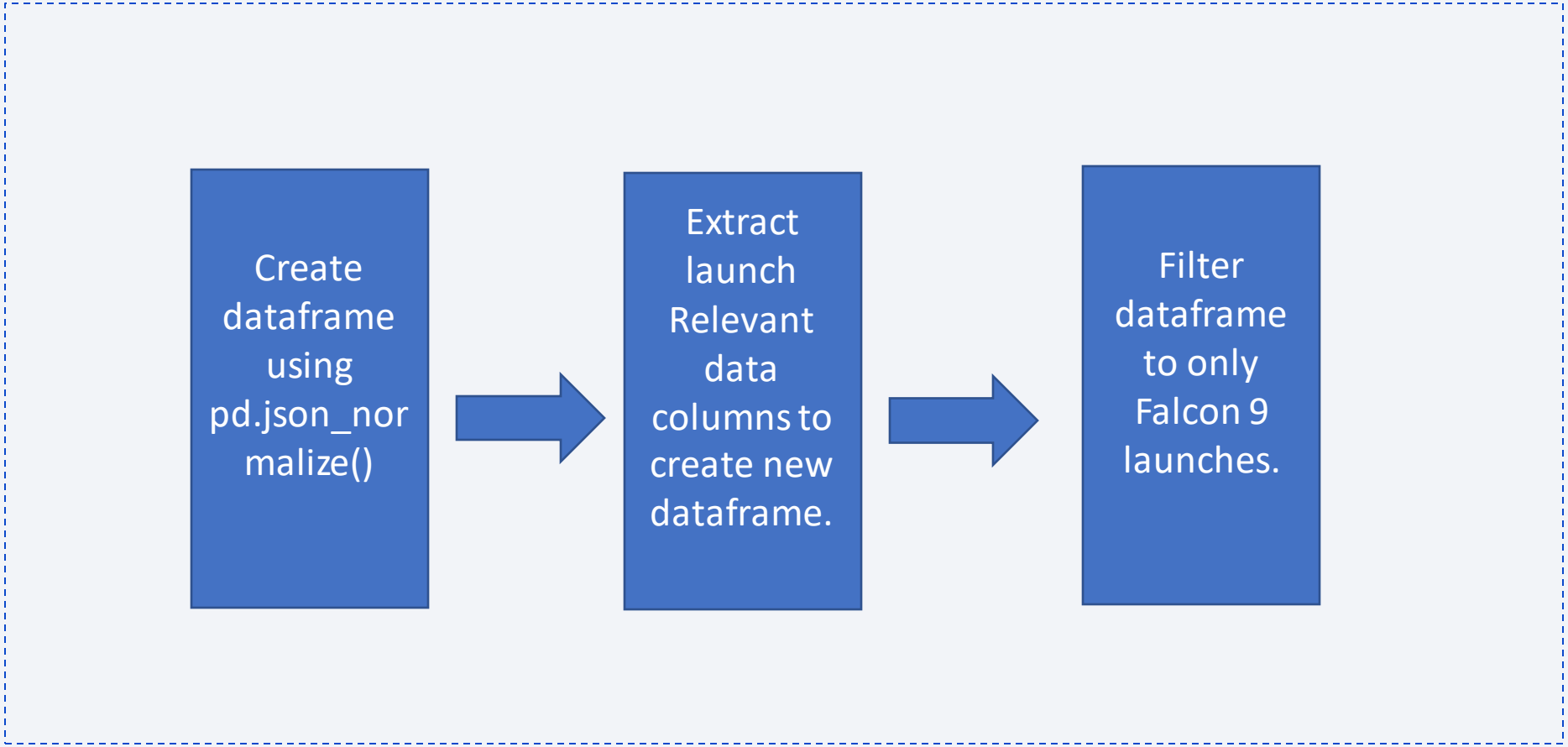
- Done through API calls and webscraping.

Data Collection – SpaceX API



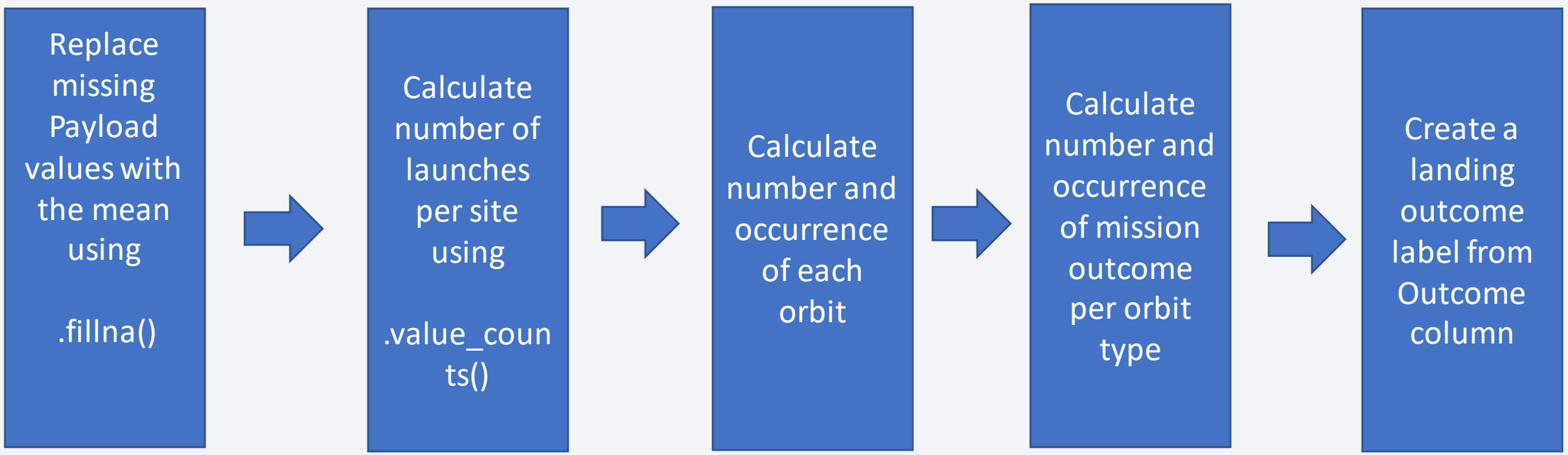
- GitHub URL https://github.com/kobbystorm1/IBM-COURSER/blob/master/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



- GitHub URL https://github.com/kobbystorm1/IBM-COURSER/blob/master/jupyter-labs-spacex-data-collection-api.ipynb

Data Wrangling



- GitHub URL https://github.com/kobbystorm1/IBM-COURSER/blob/master/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Create Scatter point chart of Flight Number vs launch site.
- Create Scatter point chart of relationship between Payload and Launch Site.
- Create Bar chart of success rate of each orbit type.
- Create Scatter point chart of relationship between Flight Number and Orbit type.
- Create Scatter point chart of relationship between Payload and Orbit type.
- Create Line chart of launch success yearly trend.

- GitHub URL https://github.com/kobbystorm1/IBM-COURSER/blob/master/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

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 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 total number of successful and failure mission outcomes

- GitHub URL https://github.com/kobbystorm1/IBM-COURSER/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL (cont)

- 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

- GitHub URL https://github.com/kobbystorm1/IBM-COURSER/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Mark all launch sites on a map.
- Mark the success/failed launches for each launch site.
- Calculate the distances between a launch site to its proximities.

GitHub URL https://github.com/kobbystorm1/IBM-COURSER/blob/master/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- Add a Launch Site Drop-down Input Component to select all or single launch sites.
- Create success-pie-chart based on selected Launch Sites.
- Add a Range Slider to Select Payload.
- Create success-payload-scatter-chart to see Payload success rates.

GitHub URL https://github.com/kobbystorm1/IBM-COURSER/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- Create a NumPy array from the column assign it to Y.
- Standardize the data in X using StandardScaler() transform.
- Use the function to split the data X and Y into training and test data using train_test_split .
- Create a logistic regression, support vector machine, decision tree classifier, k nearest neighbors objects.
- Create a GridSearchCV object of each classification method to find best parameters and Calculate the prediction accuracy on the test data.
- Plot the confusion matrix for each classification method .

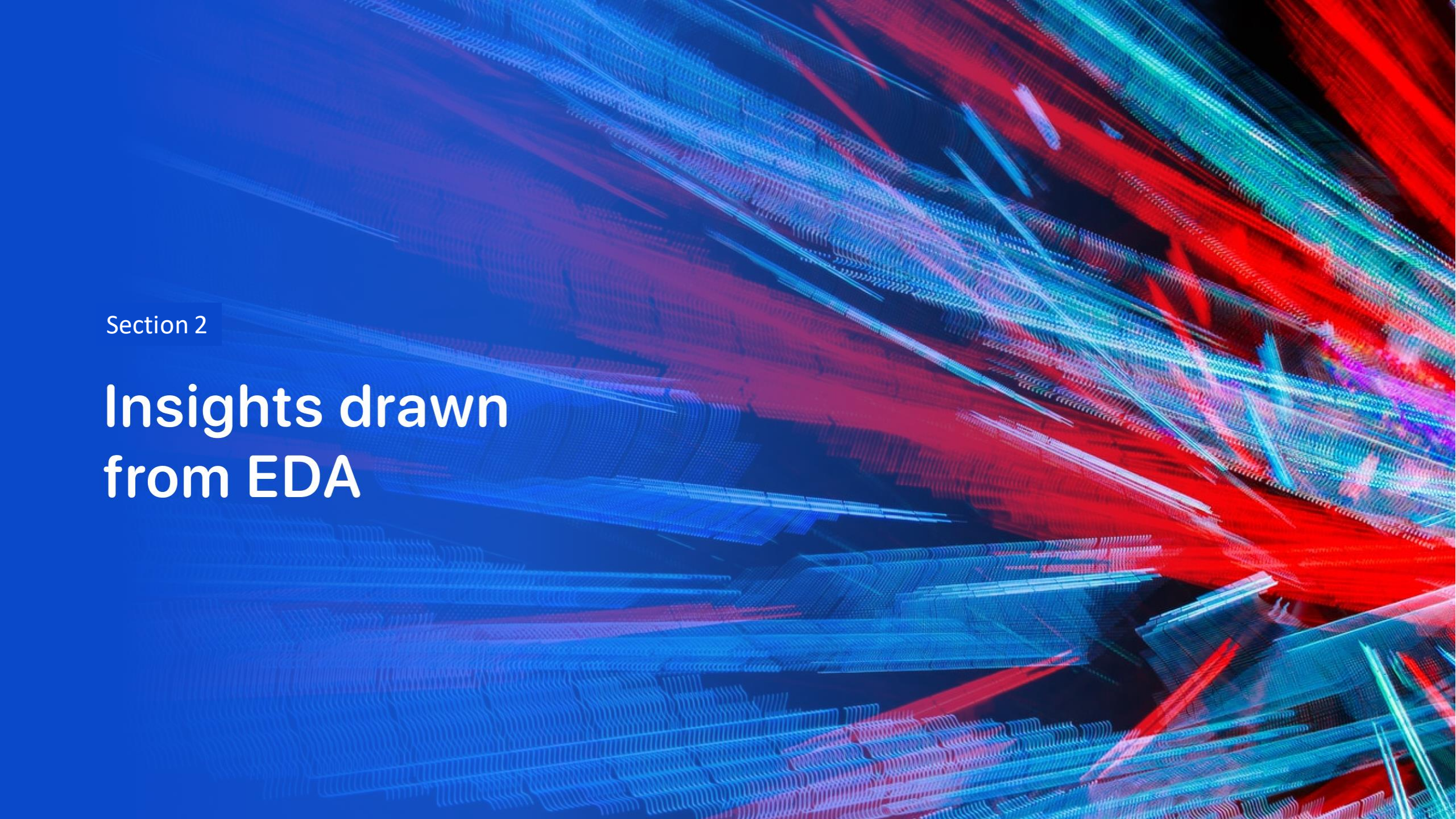
GitHub URL https://github.com/kobbystorm1/IBM-COURSER/blob/master/IBM-_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

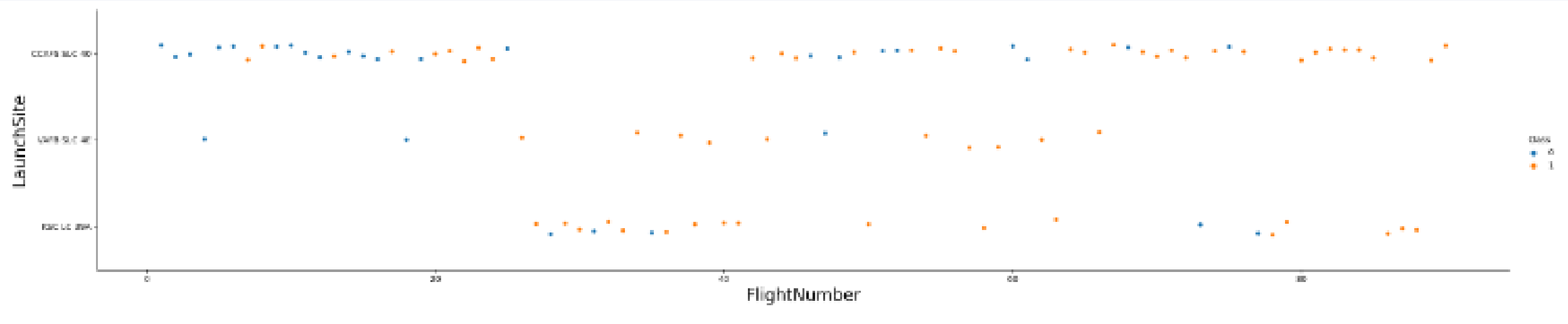
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

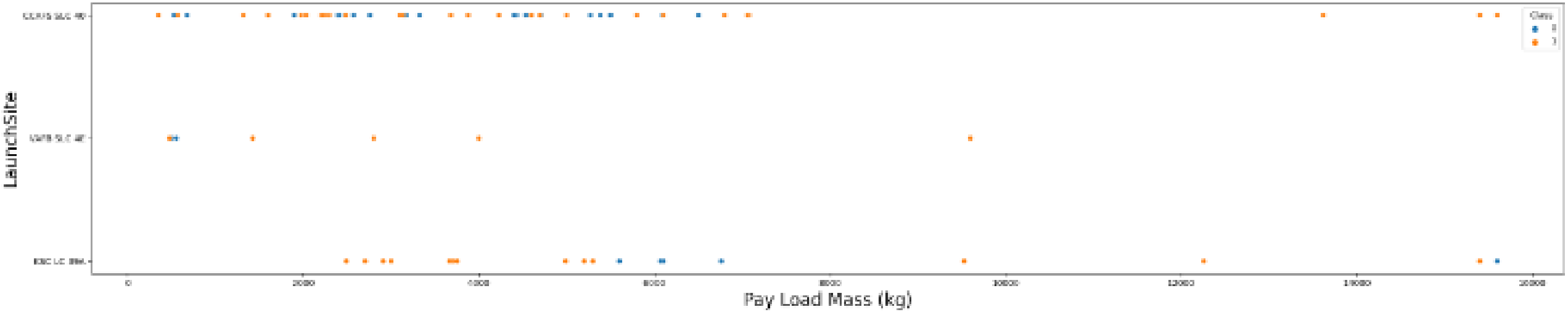


Flight Number vs. Launch Site



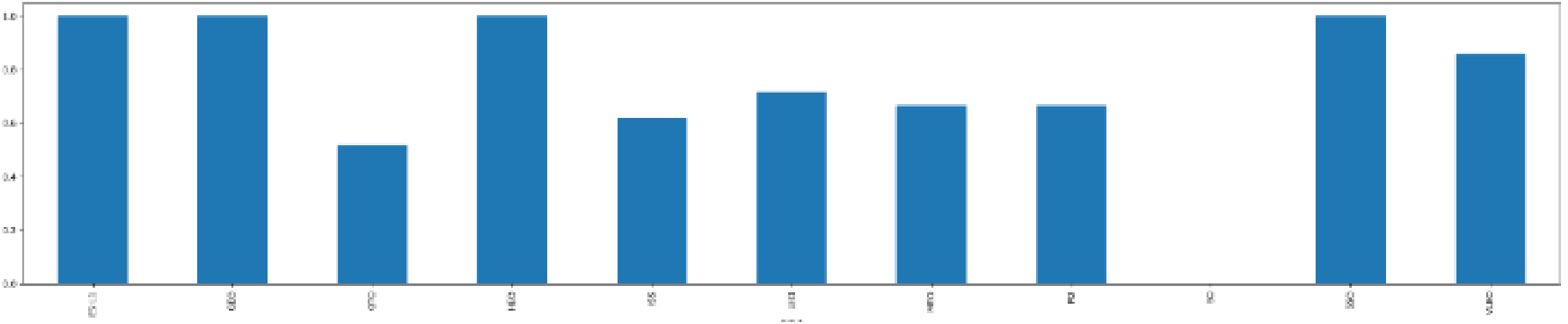
- Scatter plot of Flight Number vs. Launch Site

Payload vs. Launch Site



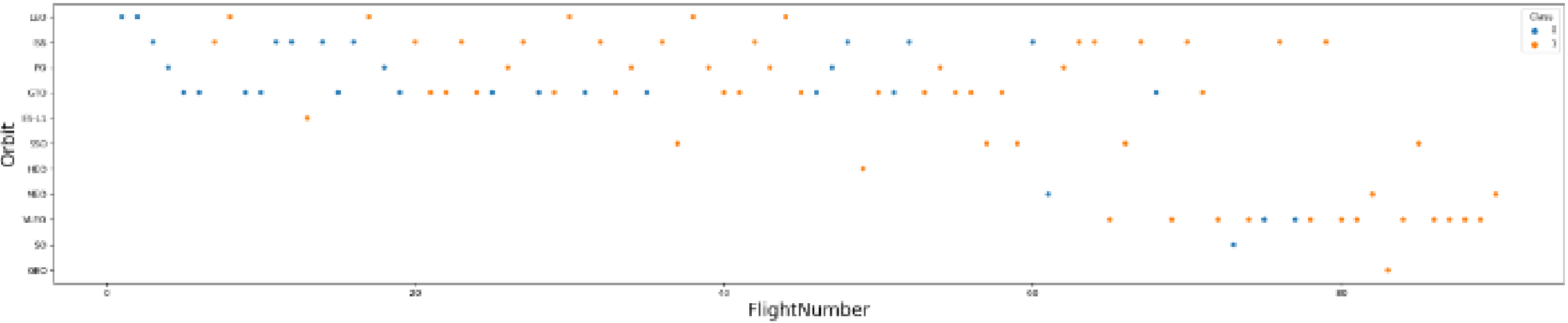
- Scatter plot of Payload vs. Launch Site

Success Rate vs. Orbit Type



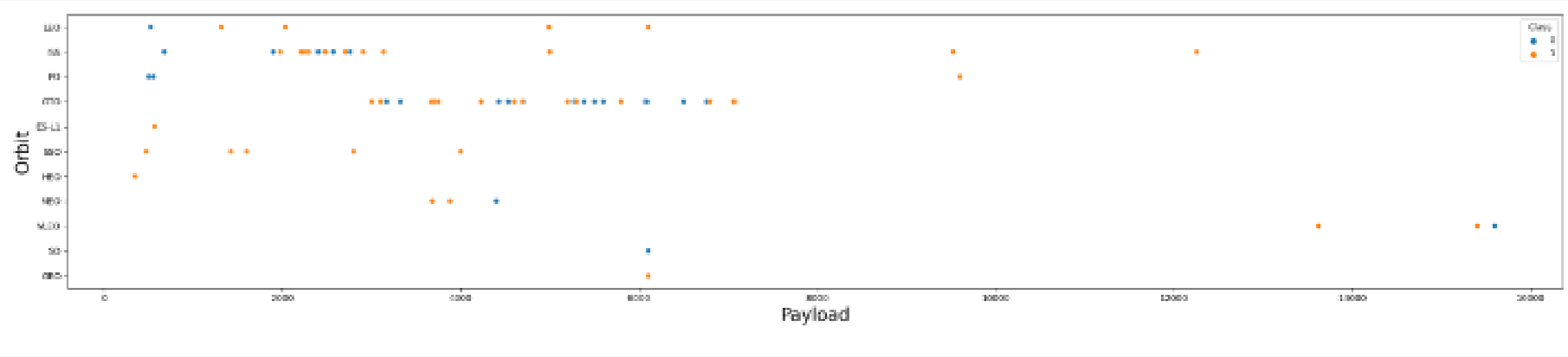
- Lower orbit travels have higher success rates

Flight Number vs. Orbit Type



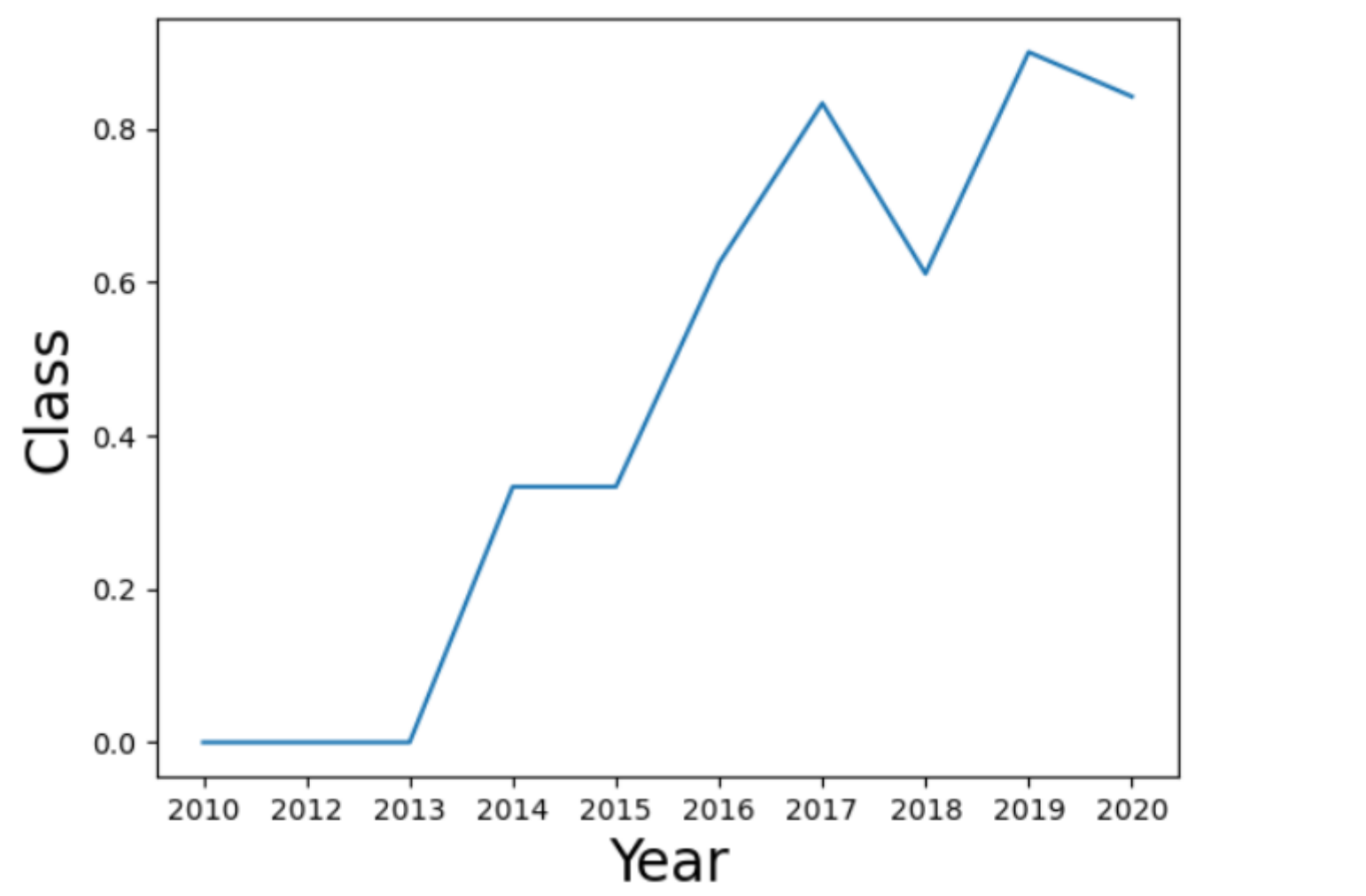
- Show a scatter point of Flight number vs. Orbit type

Payload vs. Orbit Type



- Scatter point of payload vs. orbit type

Launch Success Yearly Trend



Line chart of yearly average success rate

All Launch Site Names

- Names of the unique launch sites

Launch_Site
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
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	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 carried by boosters from NASA

Customer	sum(PAYLOAD_MASS__KG_)
NASA (CRS)	45596

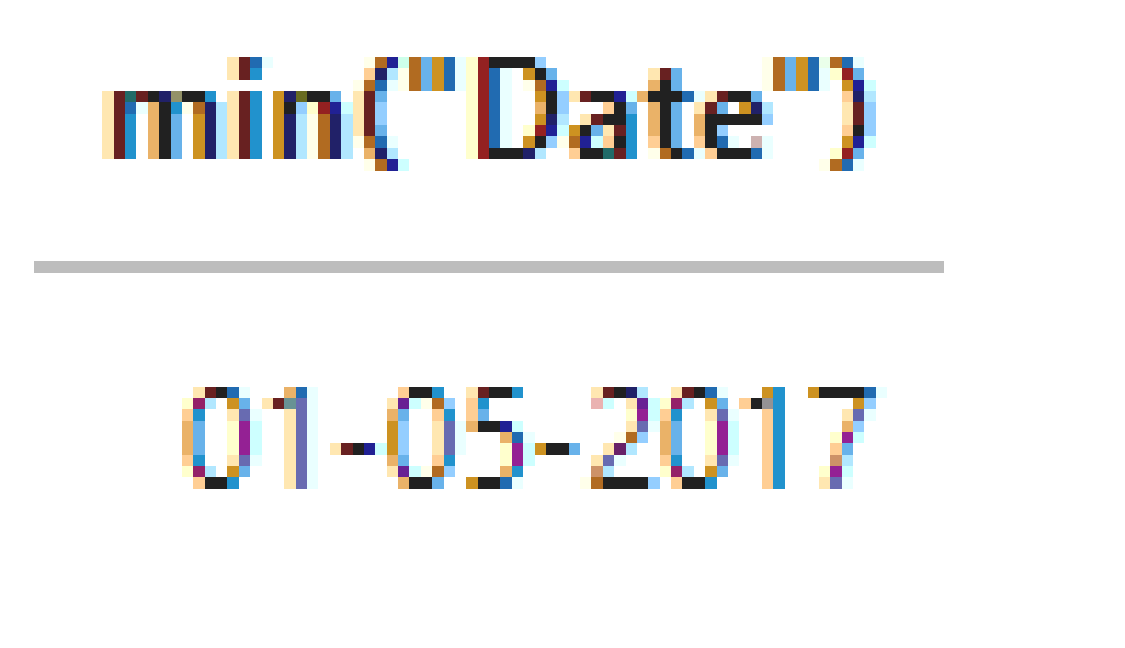
Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

Booster_Version	avg(PAYLOAD_MASS__KG_)
F9 v1.1	2928.4

First Successful Ground Landing Date

Date of the first successful landing outcome on ground pad



Successful Drone Ship Landing with Payload between 4000 and 6000

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

Booster_Version	PAYLOAD_MASS__KG_	Landing _Outcome
F9 FT B1021.1	3136	Success (drone ship)
F9 FT B1022	4696	Success (drone ship)
F9 FT B1023.1	3100	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1029.2	3669	Success (drone ship)
F9 FT B1038.1	475	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)
F9 B4 B1042.1	3500	Success (drone ship)
F9 B4 B1045.1	362	Success (drone ship)
F9 B5 B1046.1	3600	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes.

Mission_Outcome	count("Mission_Outcome")
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

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

2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

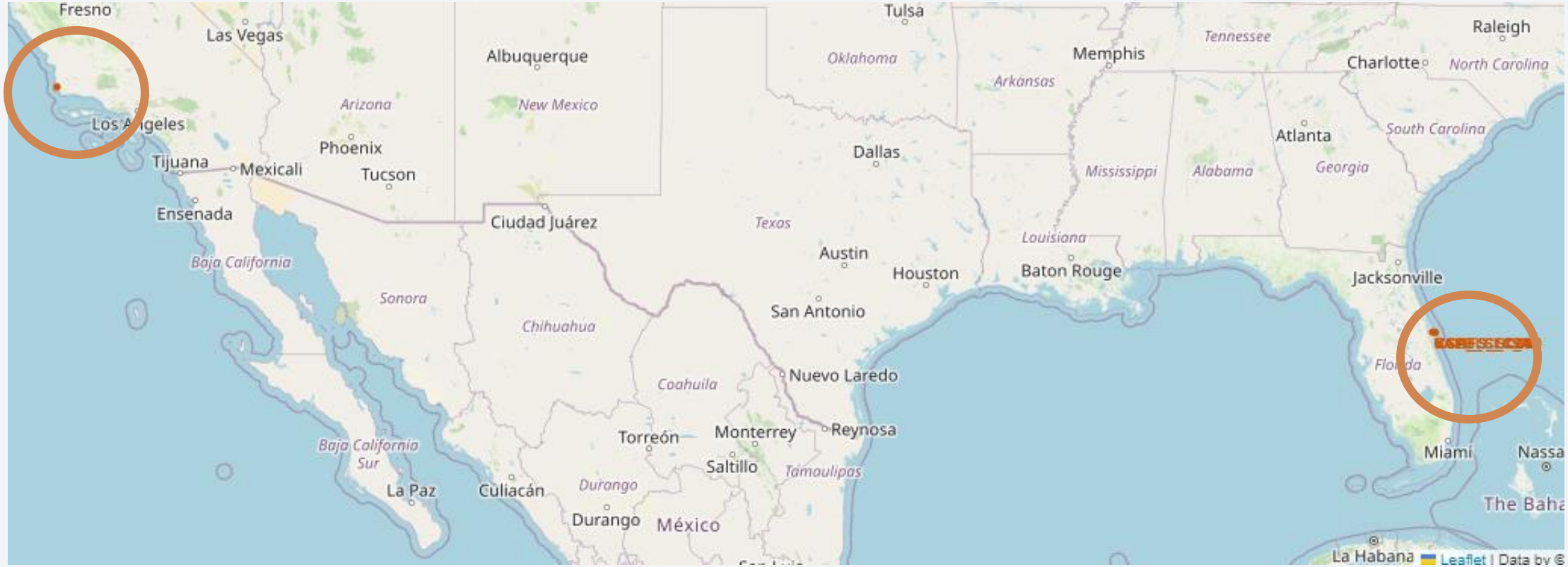
Date	month	year	Landing _Outcome	Booster_Version	Launch_Site
10-01-2015	01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
14-04-2015	04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

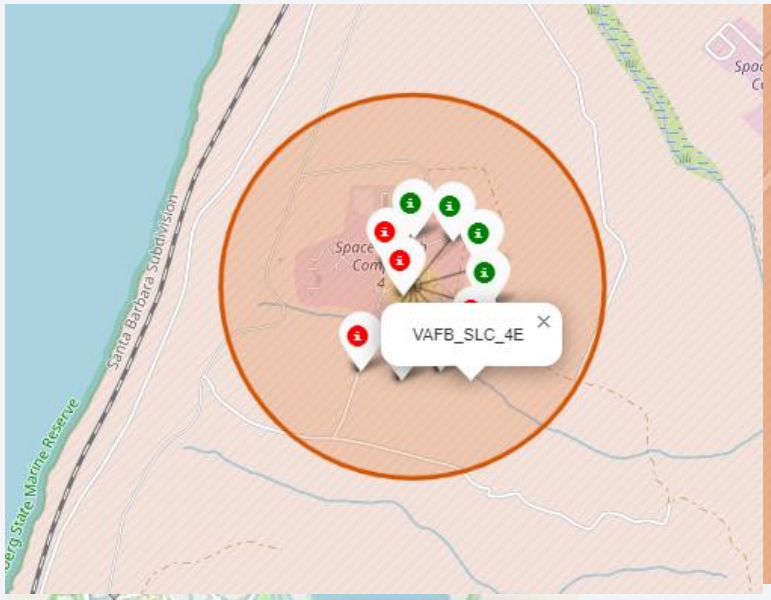

Section 3
Launch Sites
Proximities Analysis

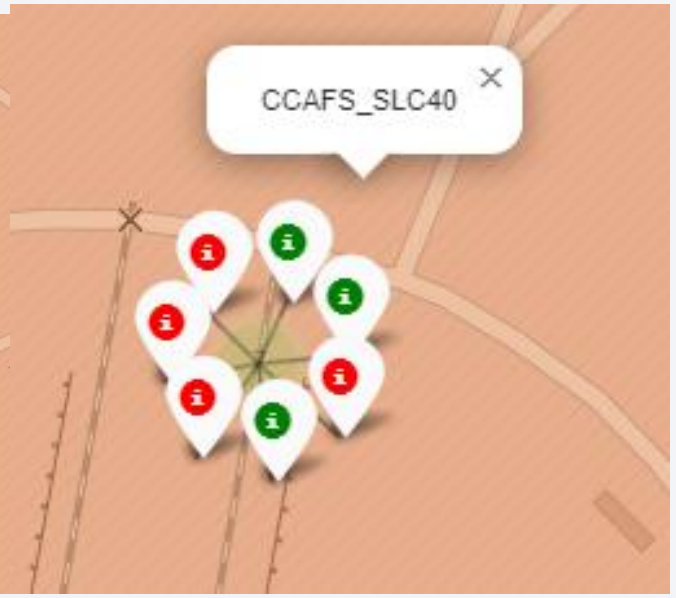
launch sites' location markers on a global map

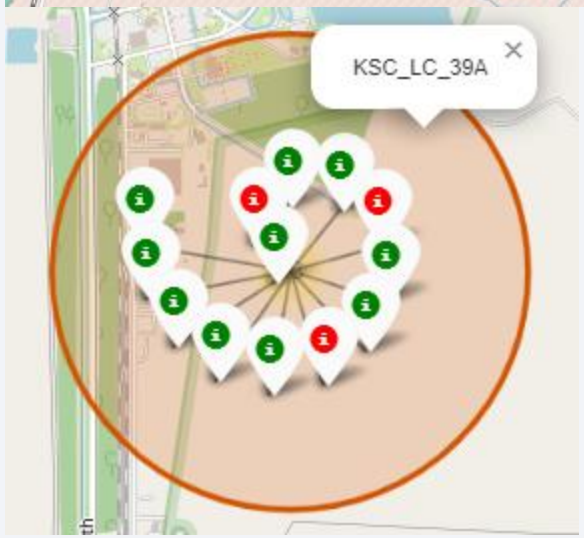


- All Launch sites are by the coast

Color-labeled launch site outcomes

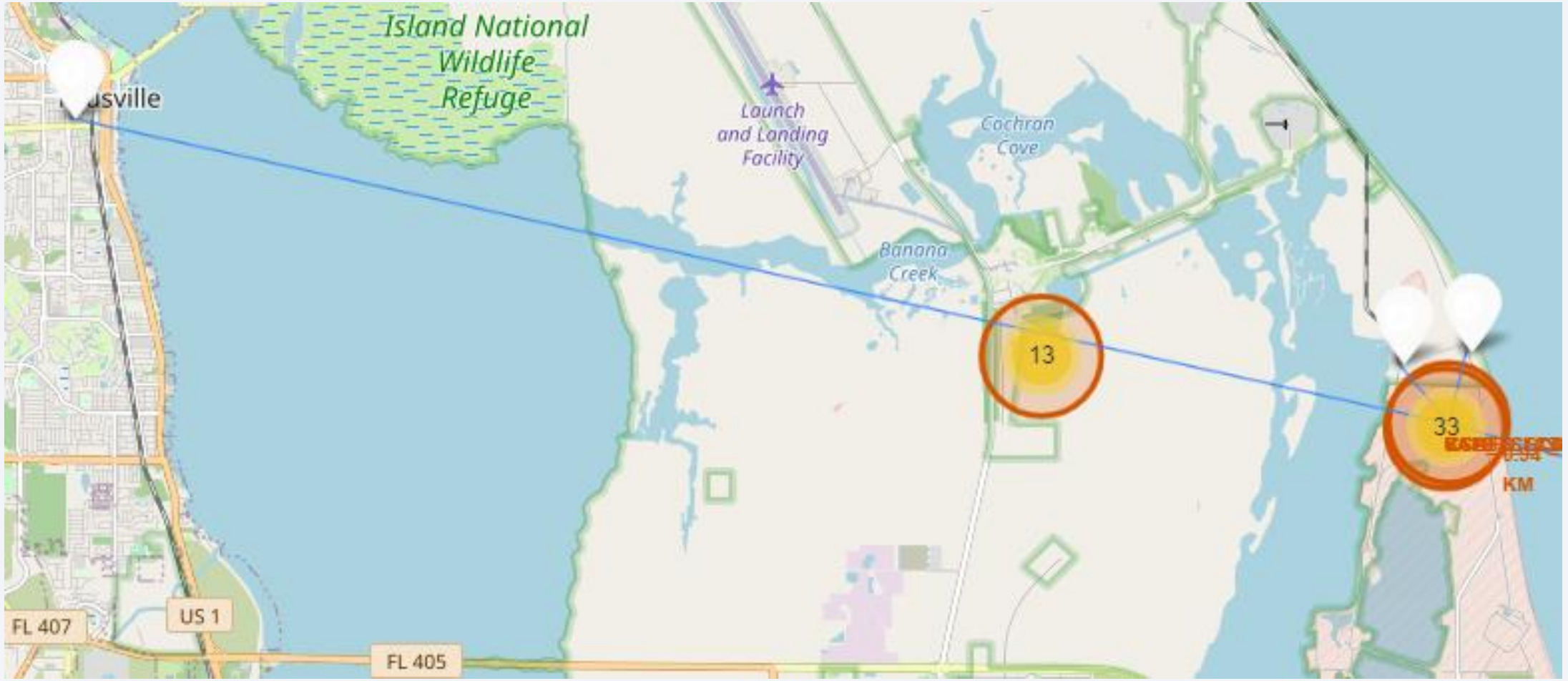




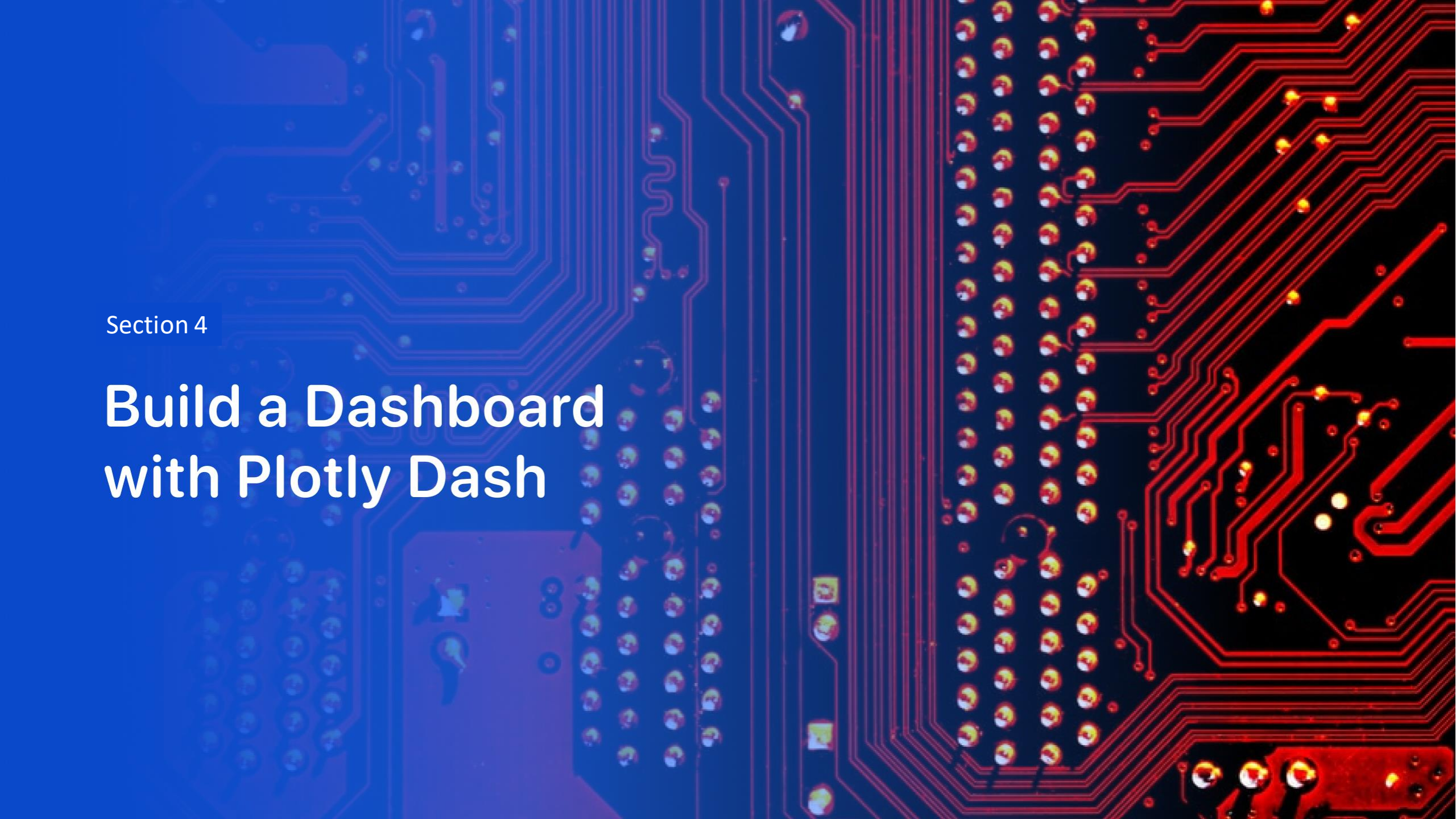


- KSC-LC-39A is the launch site with highest success count

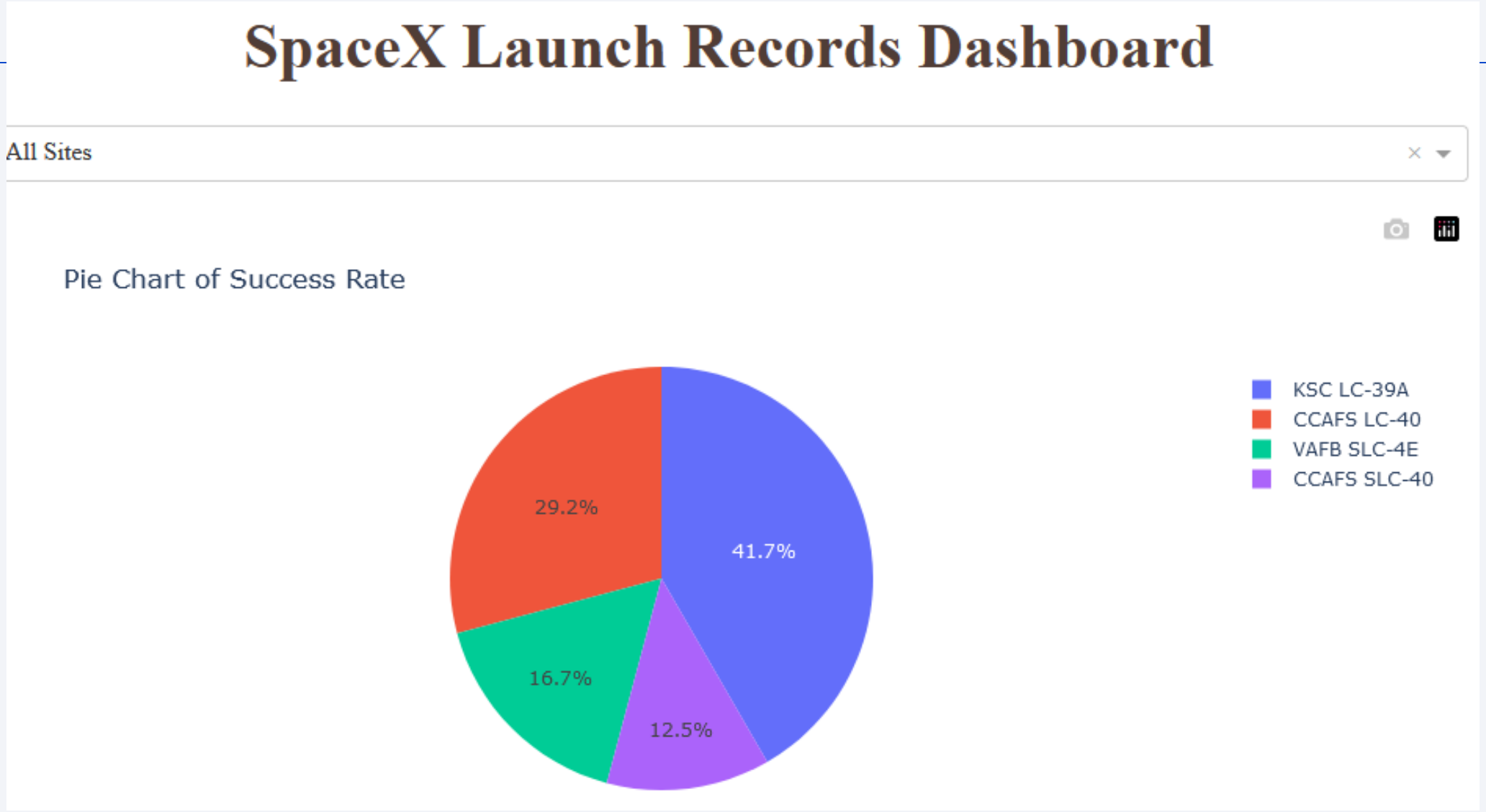
Launch site to its proximities marked



Launch sites are close to the coast, rail and highways for access but far from cities for safety

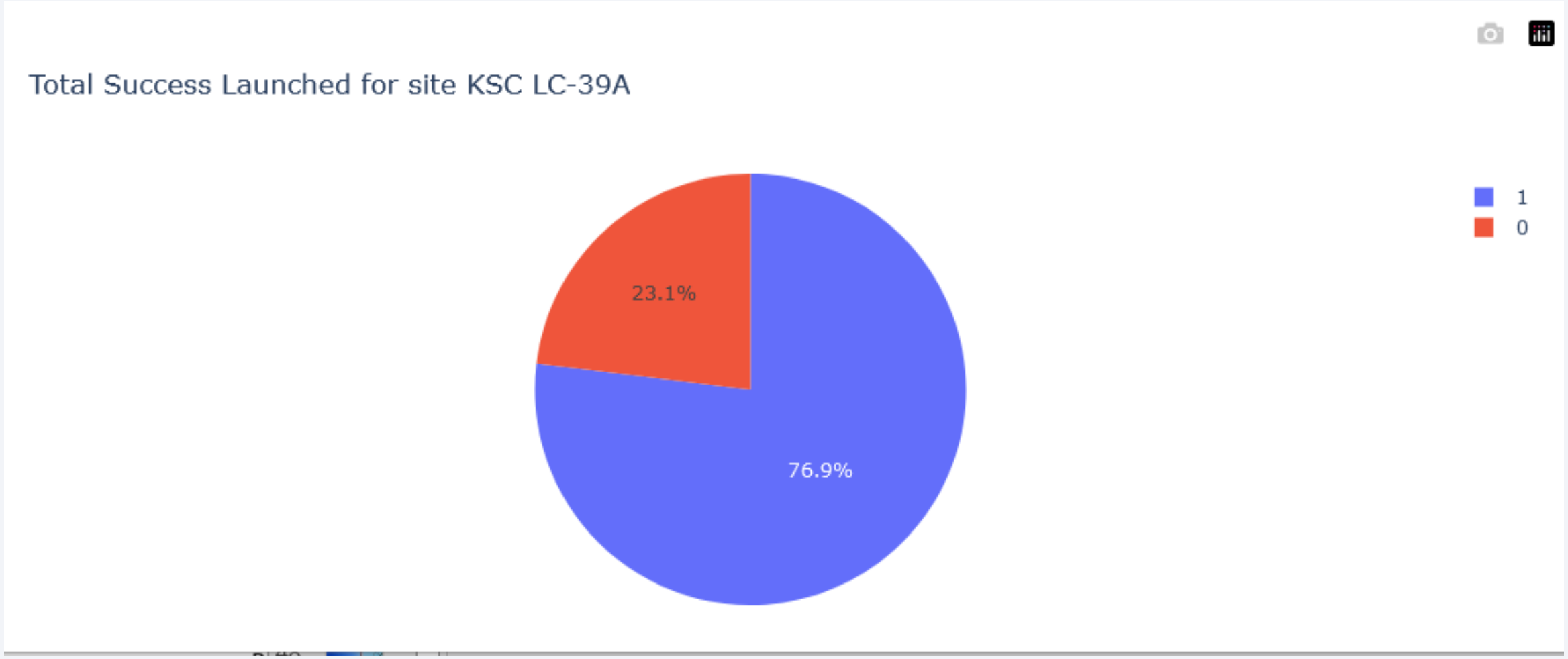

Section 4
Build a Dashboard with Plotly Dash

All launch sites success pie chart



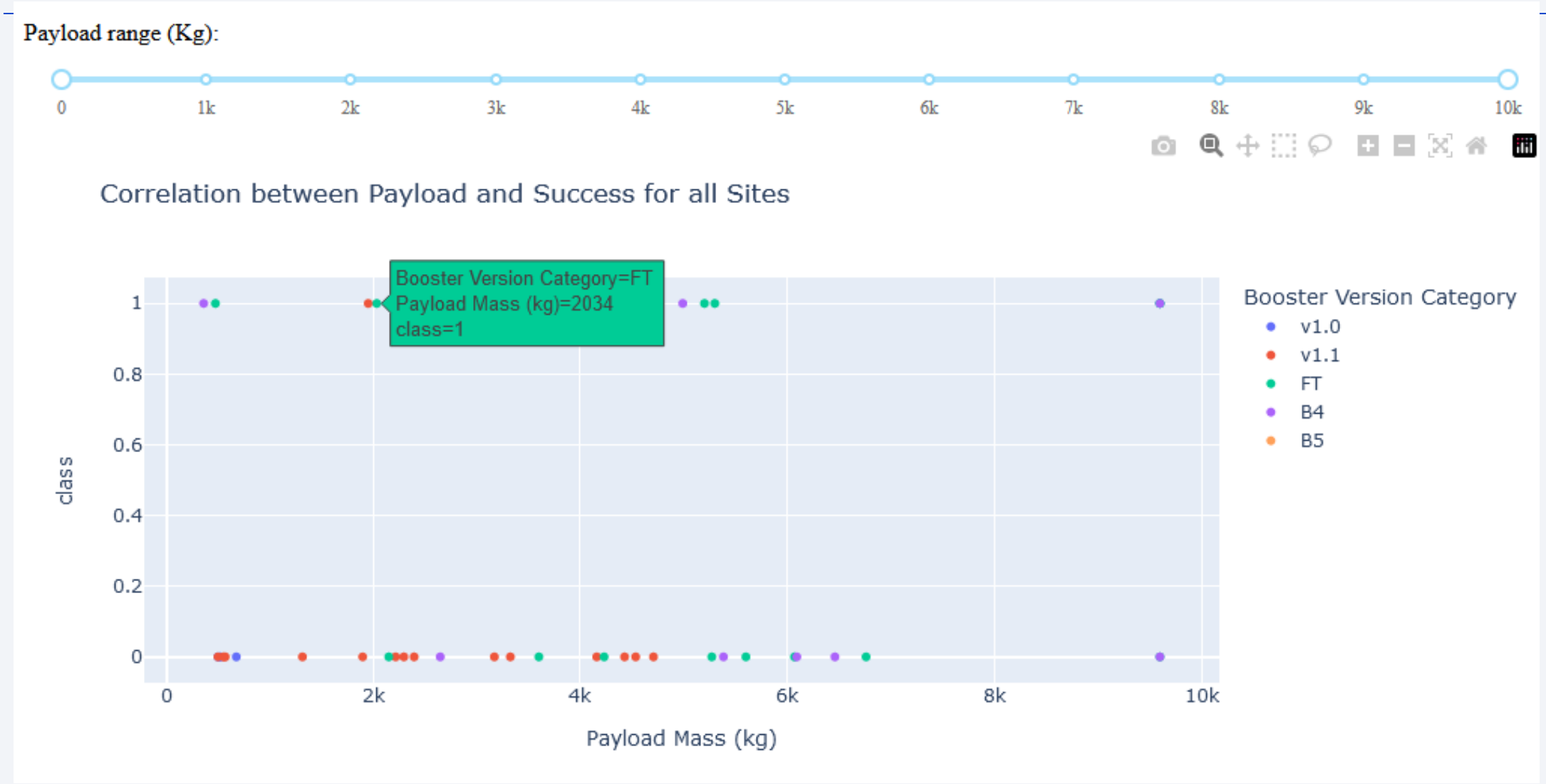
- KSC-LC-39A is the launch site with highest success count

Launch site with highest launch success ratio



KSC-LC-39A has highest success ratio

Payload vs. Launch Outcome scatter plot for all sites

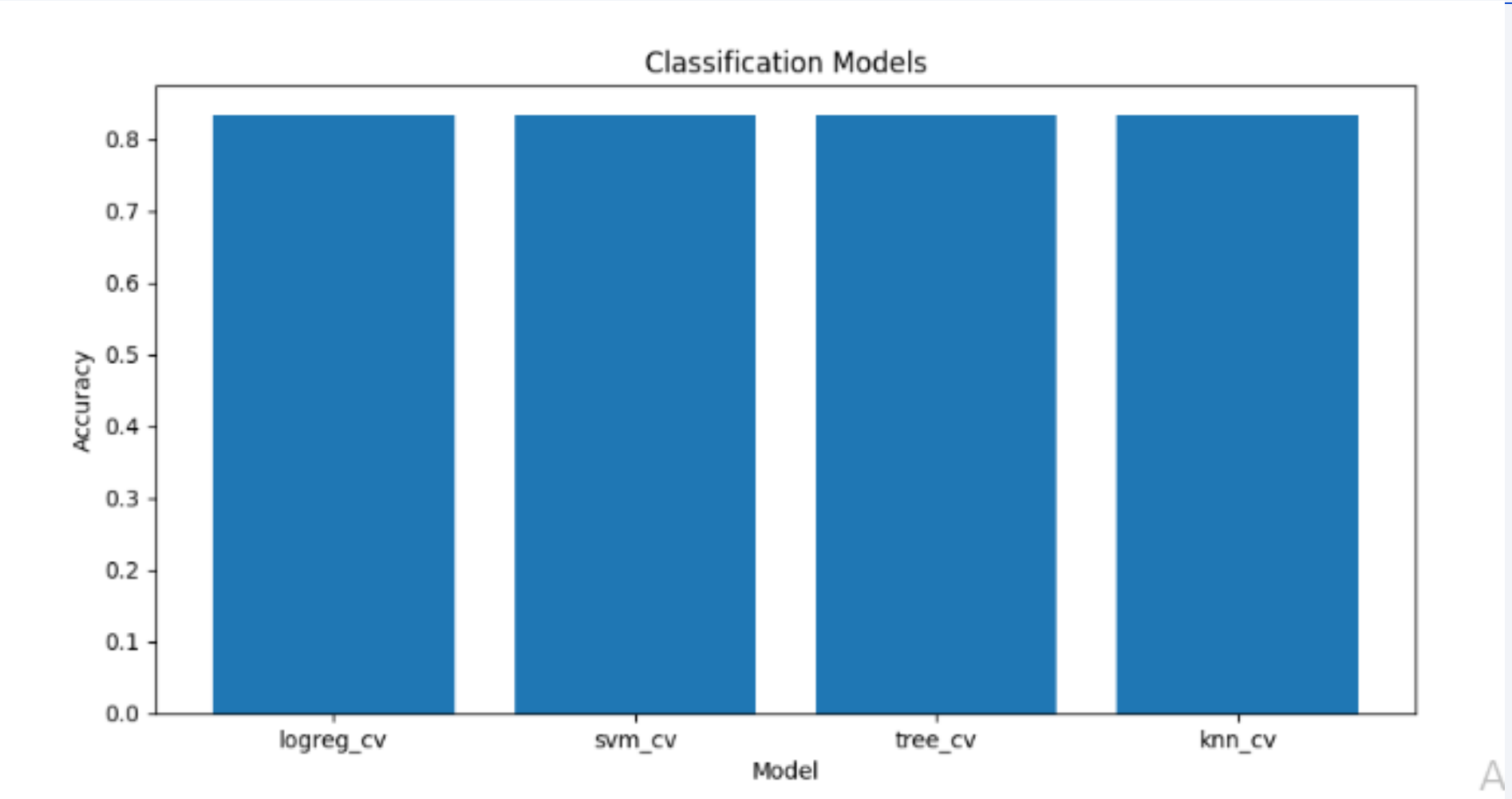


- Booster version FT with payload between 2000 and 6000 is more likely to be reused.

Section 5

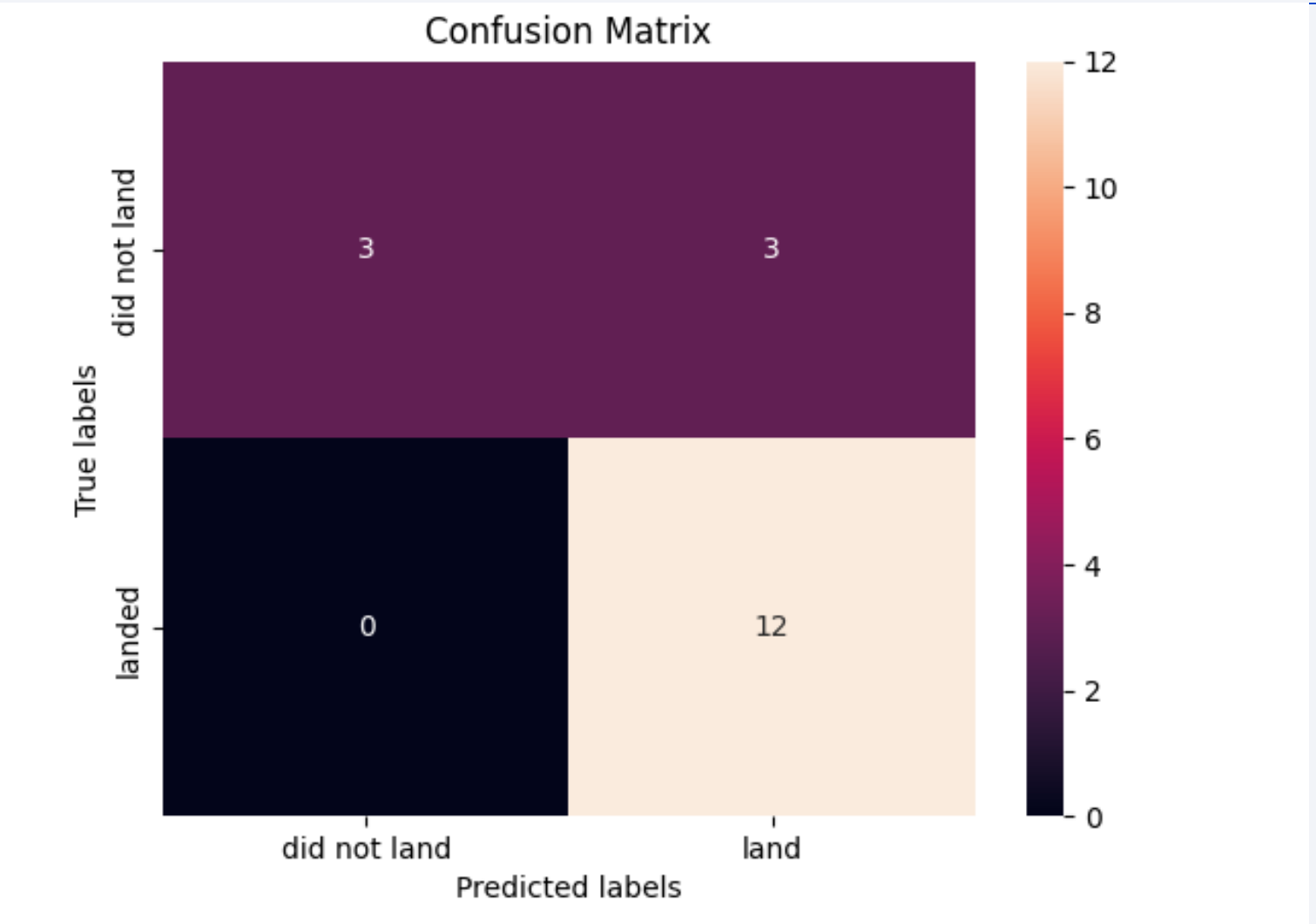
Predictive Analysis (Classification)

Classification Accuracy



- All four models have the same accuracy score

Confusion Matrix



- All models are good at classification but possess a few false positives

Conclusions

- KSC-LC-39A is the launch site with highest success count
- KSC-LC-39A launch site has highest success ratio
- Lower orbit travels have higher success rates.
- Booster version FT with payload between 2000 and 6000 is more likely to be reused/succeed.
- Predictive models had a good accuracy (83%) in classifying whether a stage 1 would land successfully to be reused.

Appendix

- All data available in GitHub repository.
- https://github.com/kobbystorm1/IBM-COURSER

Thank you!