IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Joanne Tsai>
<11/04/2022>

Outline

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

Executive Summary

- Summary of methodologies
  - Collecting data through SpaceX API and web scraping
  - Exploratory Data Analysis through data visualization
  - Analysis the proximity of launch sites through Folium
  - Machine learning model building by svm, classification tree, and logistic regression
- Summary of all results
  - Different features affect successes of launch in different ways and weight.
  - We got the weight of different features, and the prediction of based on new feature fed to the model.

Introduction

- In this project, we analyze the Space X Falcon 9 data set. We aim to find out the relationship between launching features (e.g. launch site, loading mass, orbit…etc) in order to summarize the conditions that maximizes the success rate of landing.

- Our goal is to use Space X Falcon 9 historical launching/landing data to predict whether Space X will attempt to land a rocket or not.

Section 1

Methodology

Methodology

Executive Summary

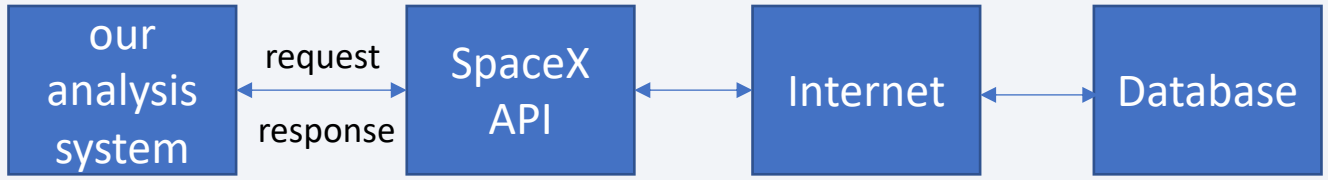
- Data collection methodology:
  - The data was collected through SpaceX API and web scraping from Wikipedia.
- Perform data wrangling
  - The data was preprocessed by removing or filling missing data, removing non-relevant column, selecting needed rows…etc.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use cross validation method to tune the hyperparameters for the models.

Data Collection

- The data set was collected from two resources.
- 1. making get requests to the SpaceX API to get information of launches
- 2. web-scraping to get the launch records of Falcon 9.

Data Collection – SpaceX API

- Used the SpaceX REST API to extract information using identification numbers in the launch data.
- Request rocket launch data from SpaceX API with according URL.
- Parse the launch data using GET request.
- Filter the data frame to only include Falcon 9 launchs.
- https://github.com/JoanneCETsai/IBM_projects/blob/main/spaceX/jupyter-labs-spacex-data-collection-api.ipynb



-From rocket column we got “booster name”

-From launchpad we got “launch site”, “longitude”, “latitude”

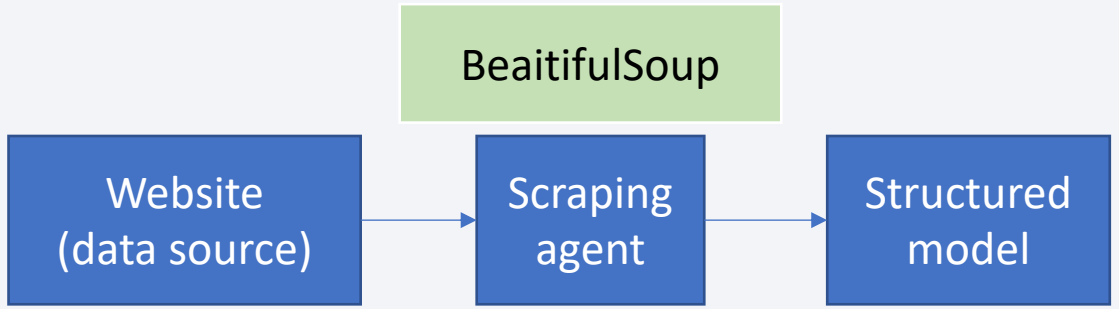
-From payload we got ”mass of payload” and “orbit”

-From cores we got the “outcomes” of the landing, “type” of the landing…etc.

Data Collection - Scraping

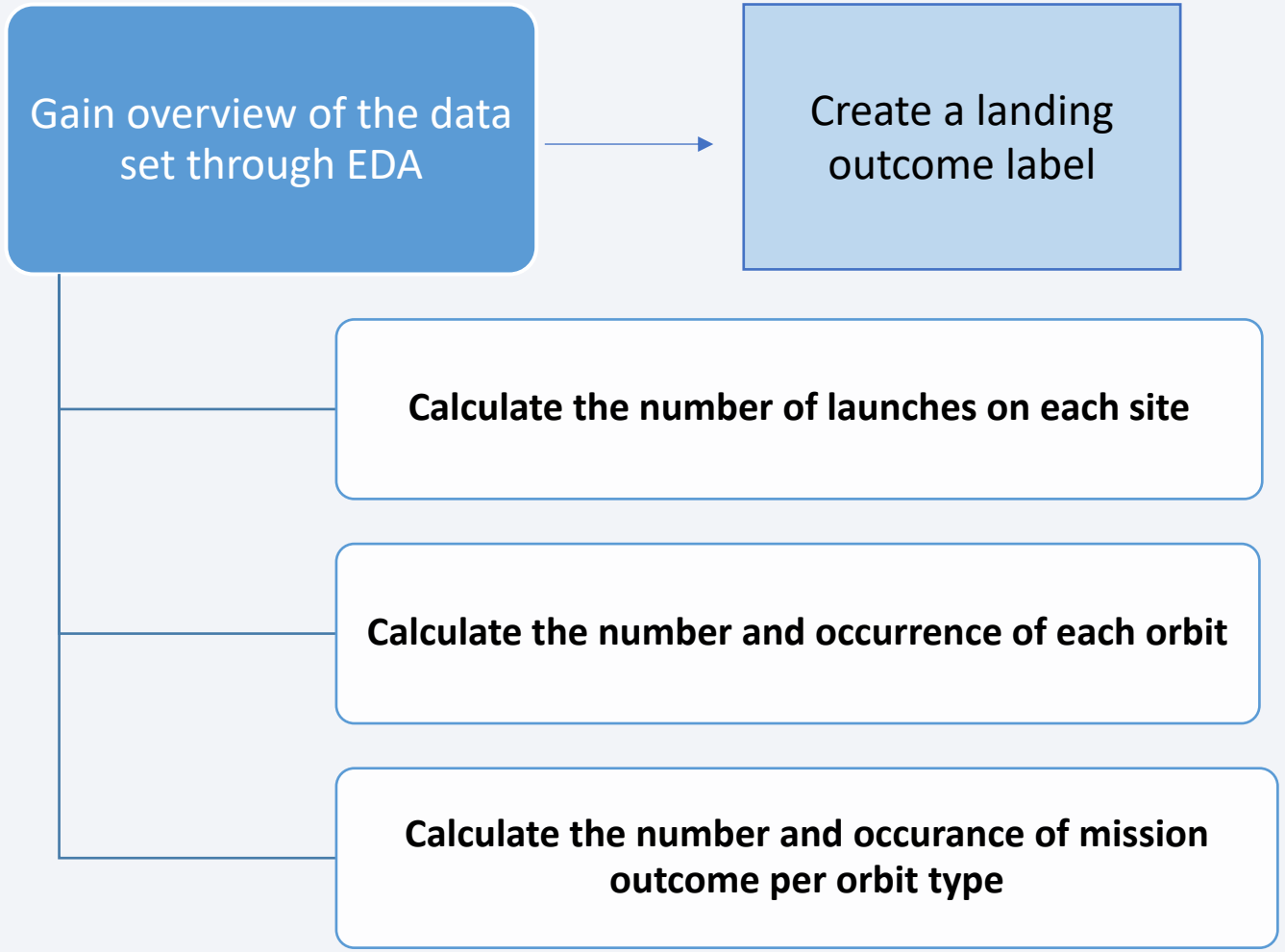
- Web Scrap Falcon 9 launch records with BeautifulSoup.
- Request the Falcon9 Launch Wiki page from its URL
- Extract all column/variable names from the HTML table header
- Create a data frame by parsing the launch HTML tables

- https://github.com/JoanneCETsai/IBM_projects/blob/main/spaceX/jupyter-labs-webscraping.ipynb



Data Wrangling

- Perform Exploratory Data Analysis and determine Training Labels



EDA with Data Visualization

- Scatter plot (with success landing marked with orange, failure landing marked with blue):
  - Flight Number vs Launch Site
  - Payload vs Launch Site
  - Flight Number vs Orbit type
  - Payload vs Orbit type

- Bar chart: success rate of each orbit type
- Line chart: launch success yearly trend

- https://github.com/JoanneCETsai/IBM_projects/blob/main/spaceX/jupyter-labs-eda-dataviz.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

- https://github.com/JoanneCETsai/IBM_projects/blob/main/spaceX/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

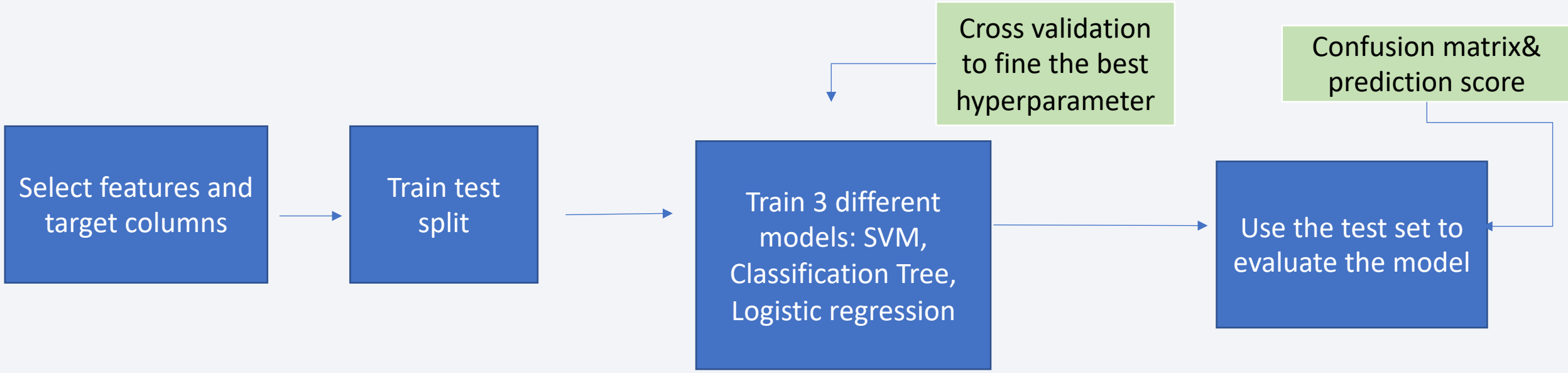
- Mark all launch sites on a map using folium.Circle() and folium.map.Marker
- Mark the success/failed launches for each site on the map using MarkerCluster
- Calculate the distances between a launch site to its proximities
    - Use MousePossition to find the coordinates
    - Use folium.PolyLine to mark lines on the map

- https://github.com/JoanneCETsai/IBM_projects/blob/main/spaceX/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

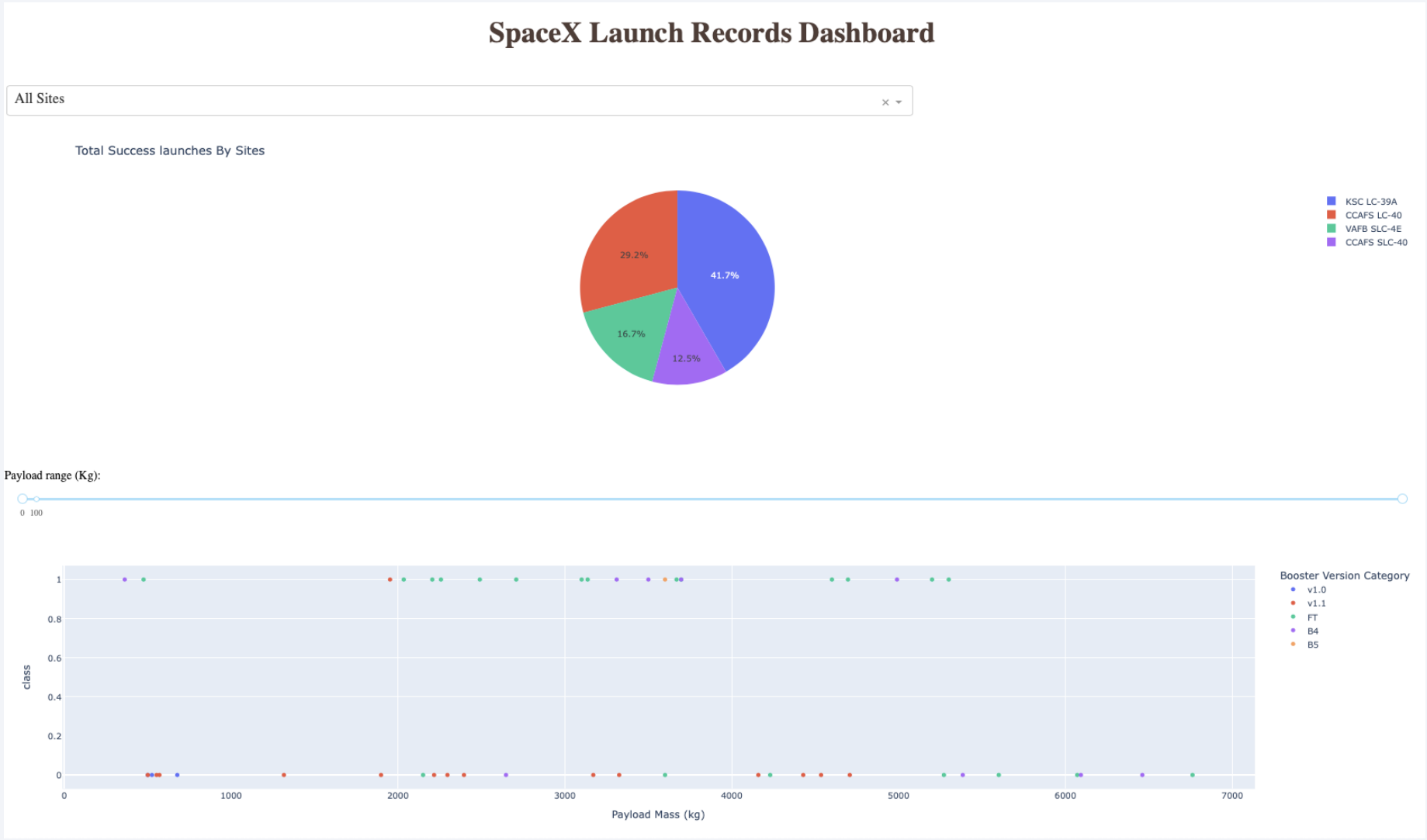
- Pie Chart
  - By selecting the launch sites in the dropdown menu, you can see the success rate of each launch site.
  - If all launch sites are selected, you can see the comparison of success rate of all sites
- Scatter Plot
  - selecting the launch site and loading mass range, you can see the payload mass vs class scatter plot
- https://github.com/JoanneCETsai/IBM_projects/blob/main/spaceX/spacex_dash_app.py

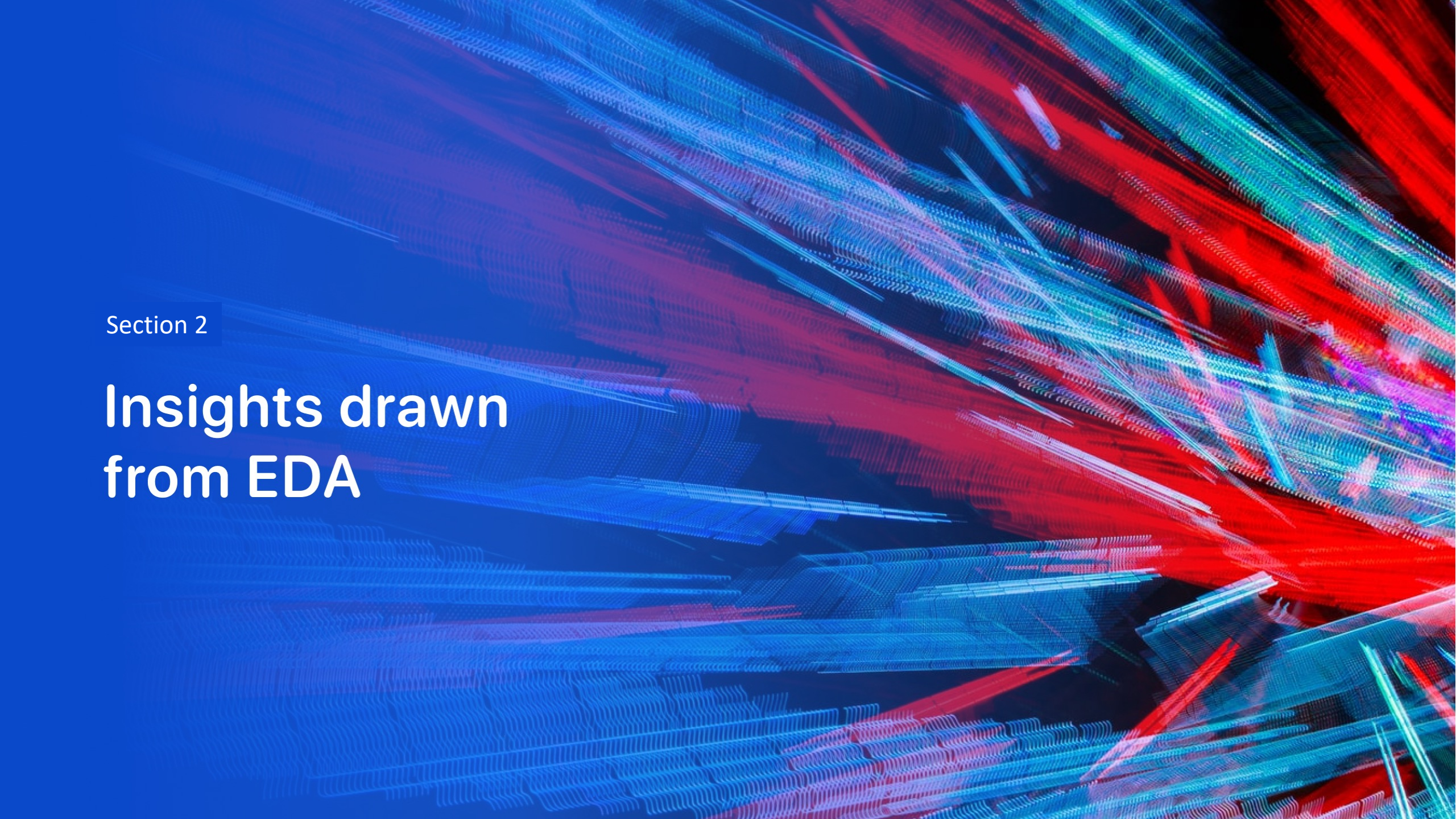
Predictive Analysis (Classification)



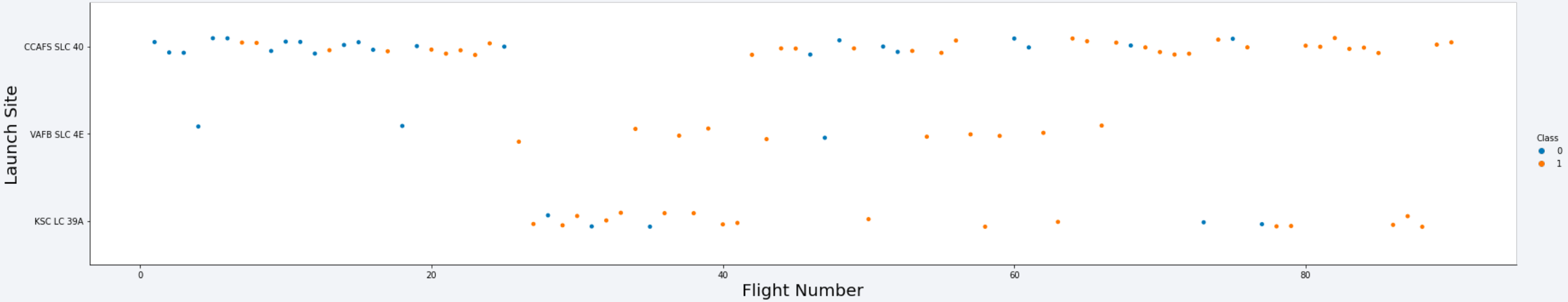
- https://github.com/JoanneCETsai/IBM_projects/blob/main/spaceX/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results


SpaceX Launch Records Dashboard
All Sites
Total Success launches By Sites
29.2%
41.7%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
Payload range (Kg):
0
100
class
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT
B4
B5

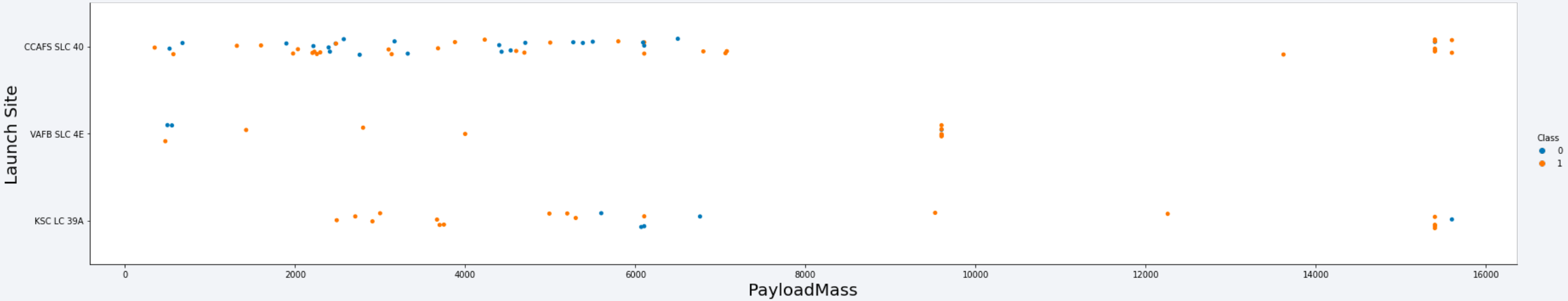

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- For all the boosters, the success rate seems increase with flight number.

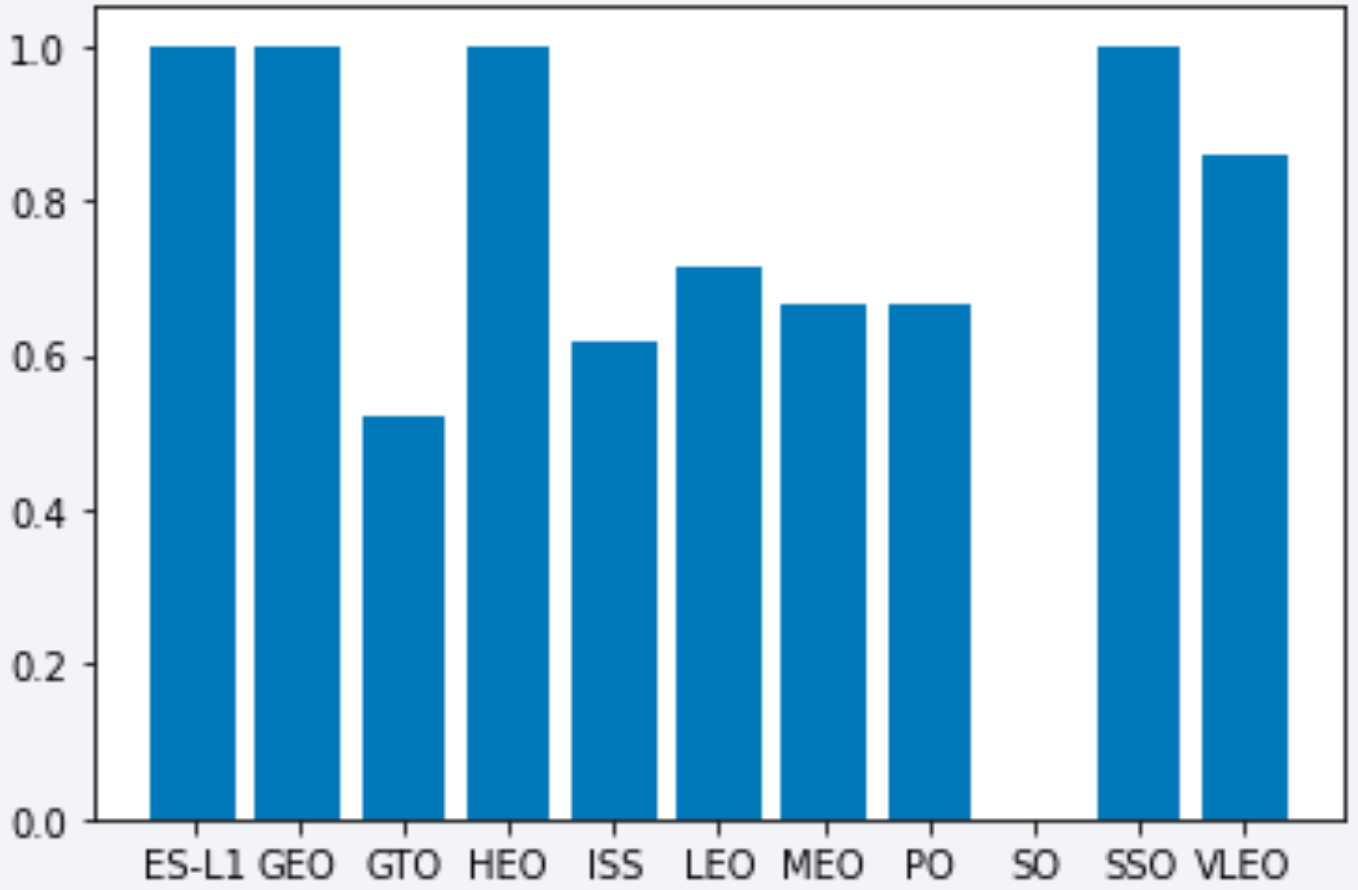
Payload vs. Launch Site



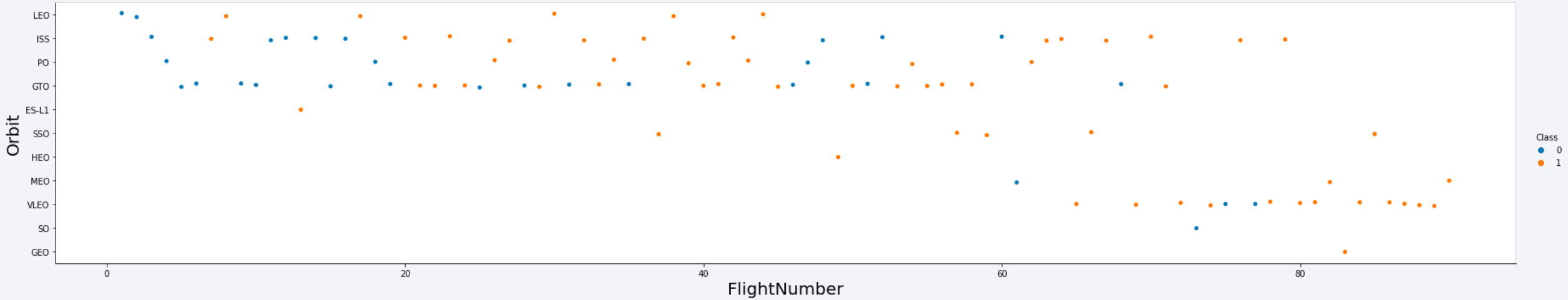
- For CCAFS SLC 40 and VAFB SLC 4E, when payload mass is high, the success rate is high.
- For KSC LC 39A, the relationship between success rate and payload mass is not obvious.

Success Rate vs. Orbit Type

- GTO orbit has the lowest success rate.
- ES-L1, GEO, HEO, and SSO have 100% success rate.
- VLEO also has high success rate.

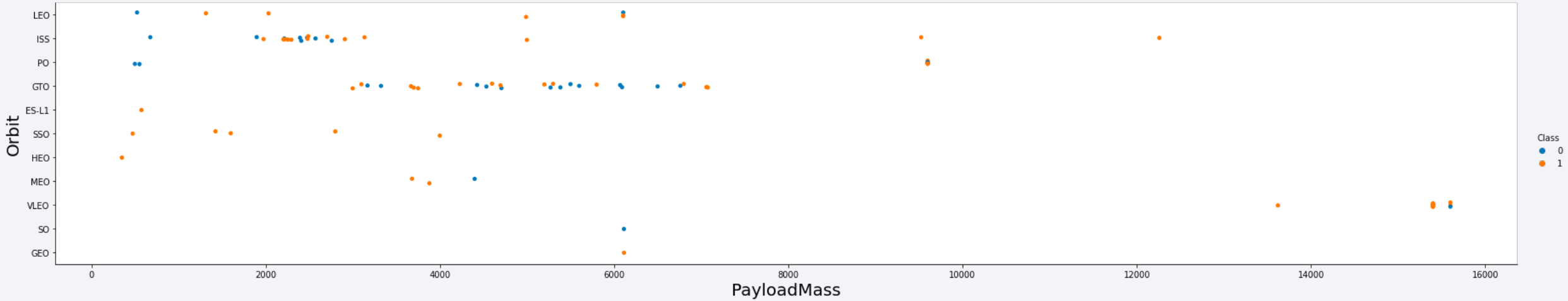


Flight Number vs. Orbit Type



- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

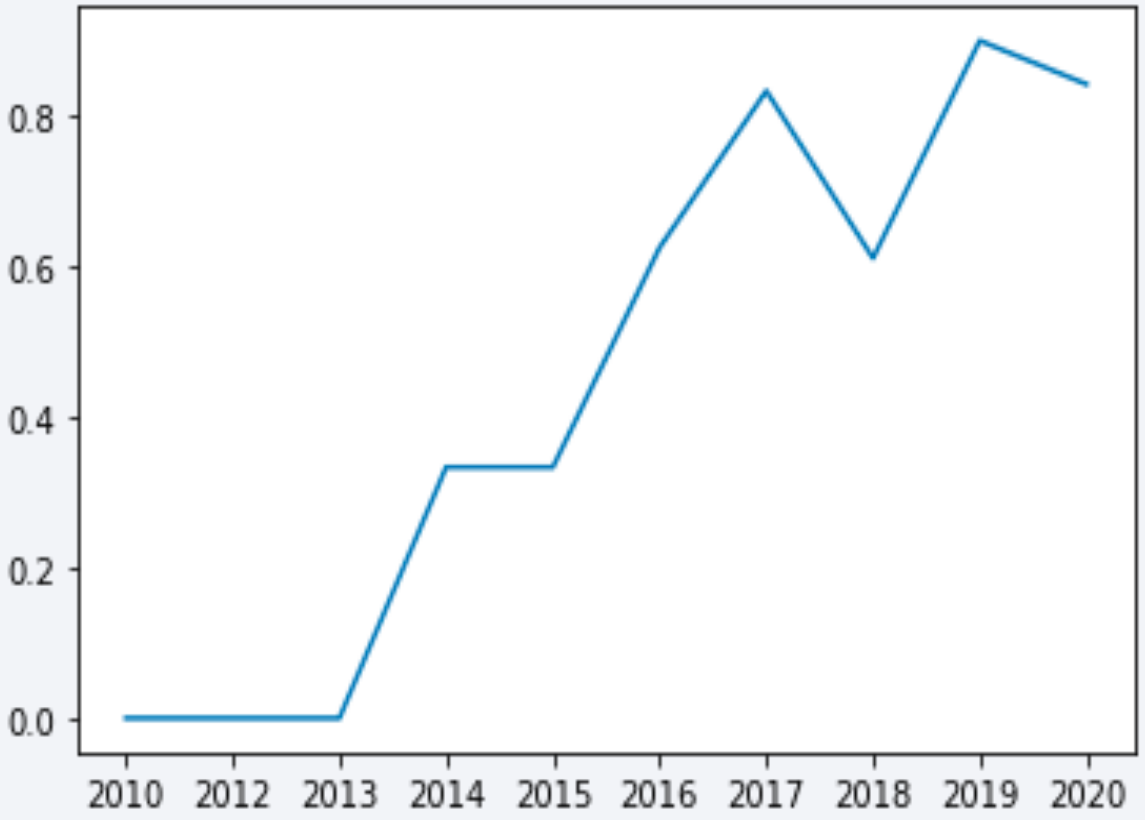
Payload vs. Orbit Type



- With heavy payloads the successful landing rate are more for Polar, LEO and ISS.
- For GTO, we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend

- the success rate since 2013 kept increasing, and fluctuate a bit during 2018-2020.



All Launch Site Names

- The names of the unique launch sites:

  'CCAFS LC-40', 'VAFB SLC-4E', 'KSC LC-39A', 'CCAFS SLC-40'

- There are 4 different launch sites.

Launch Site Names Begin with 'CCA'

	Flight Number	Date	Time (UTC)	Booster Version	Launch Site	Payload	Payload Mass (kg)	Orbit	Customer	Landing Outcome	class	Lat	Long
0	1	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Failure (parachute)	0	28.562302	-80.577356
1	2	2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel o...	0.0	LEO (ISS)	NASA (COTS) NRO	Failure (parachute)	0	28.562302	-80.577356
2	3	2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2+	525.0	LEO (ISS)	NASA (COTS)	No attempt	0	28.562302	-80.577356
3	4	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	No attempt	0	28.562302	-80.577356
4	5	2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	No attempt	0	28.562302	-80.577356

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 = 2928.4 kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

- Present your query result with a short explanation here

Total Number of Successful and Failure Mission Outcomes

- The total number of success : 32
- The total number of failure : 24

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass: F9 FT B1029.1, F9 FT B1036.1, F9 B4 B1041.1, F9 FT B1036.2, and F9 B4 B1041.2

2015 Launch Records

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

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

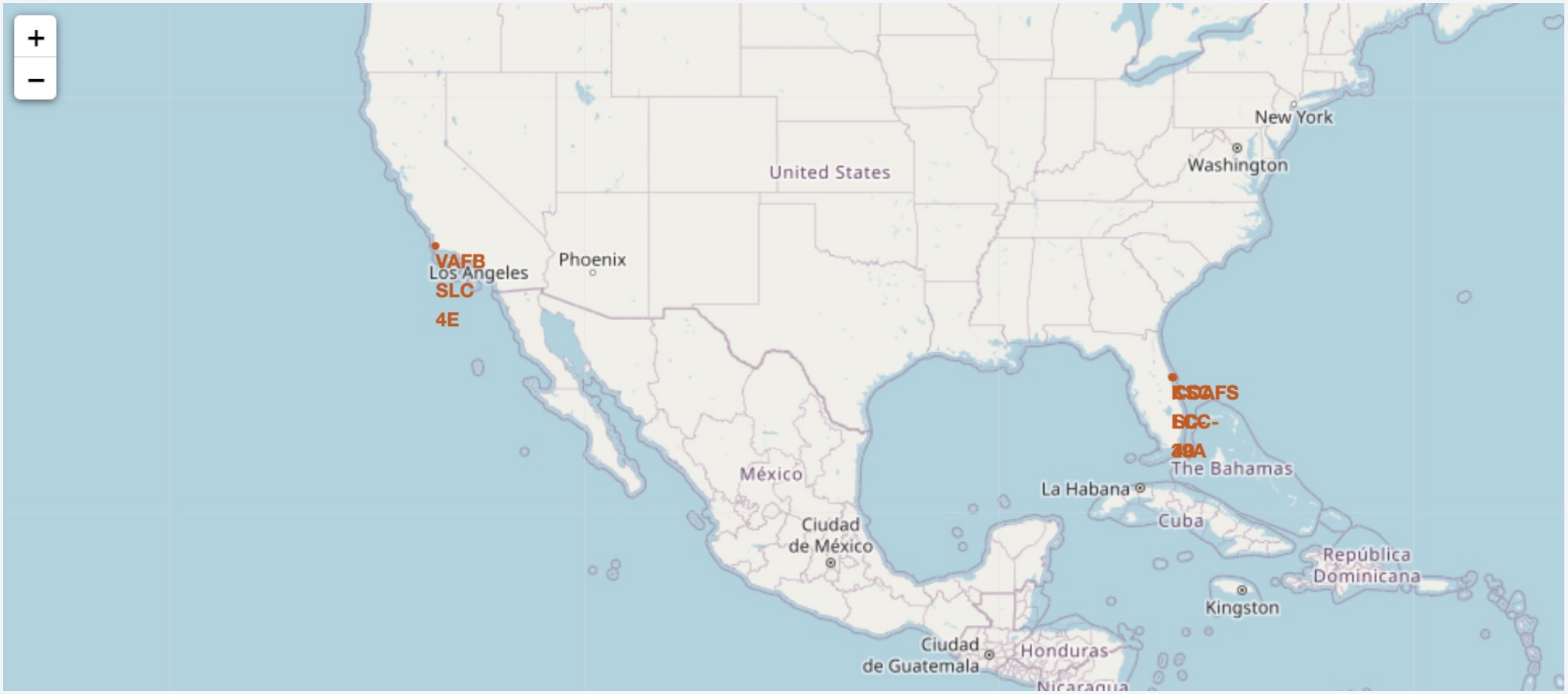
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

- Present your query result with a short explanation here


Section 3
Launch Sites
Proximities Analysis

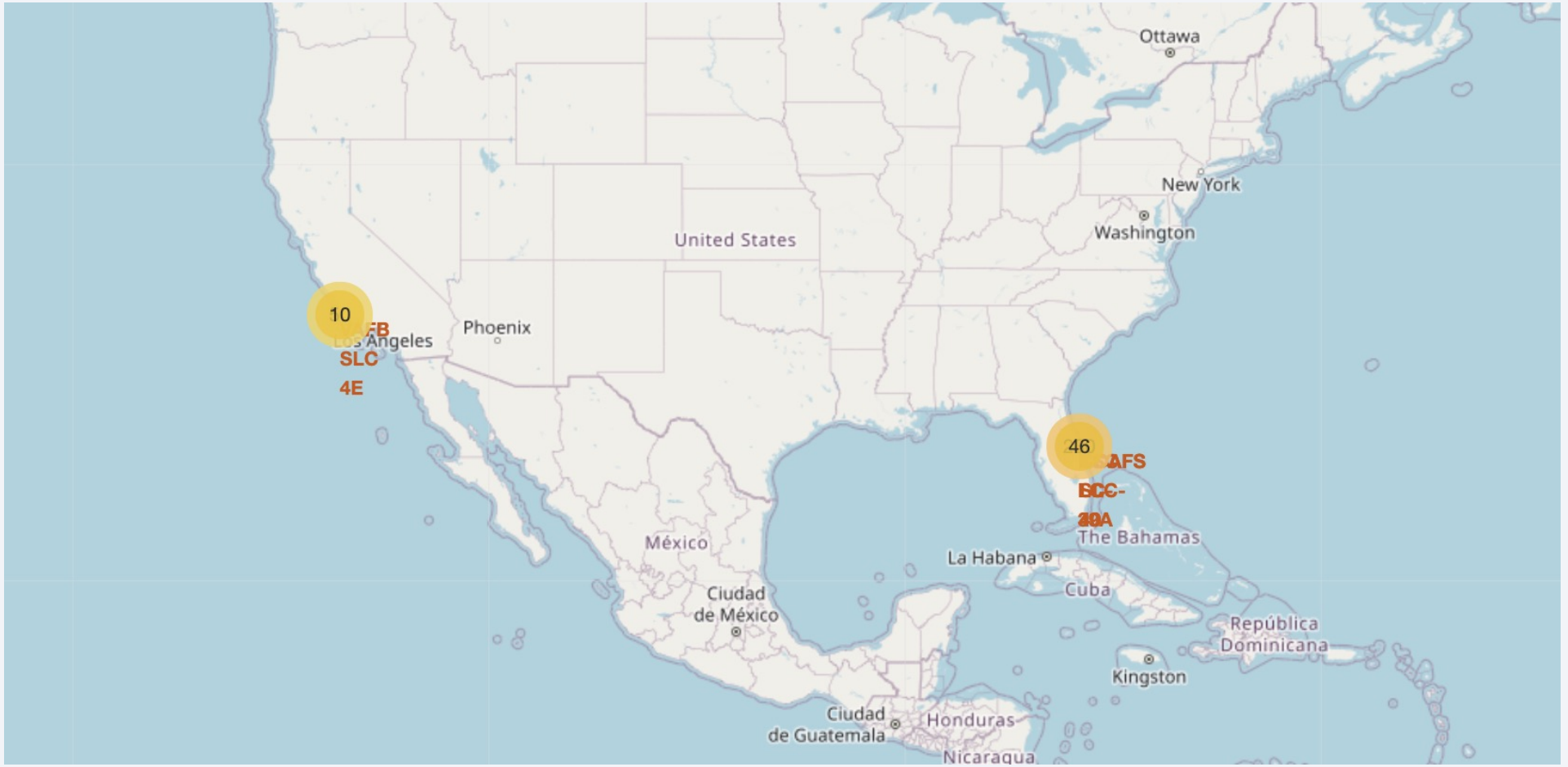
<The location of launch sites>

- We can zoom in to see the location of each launch site.



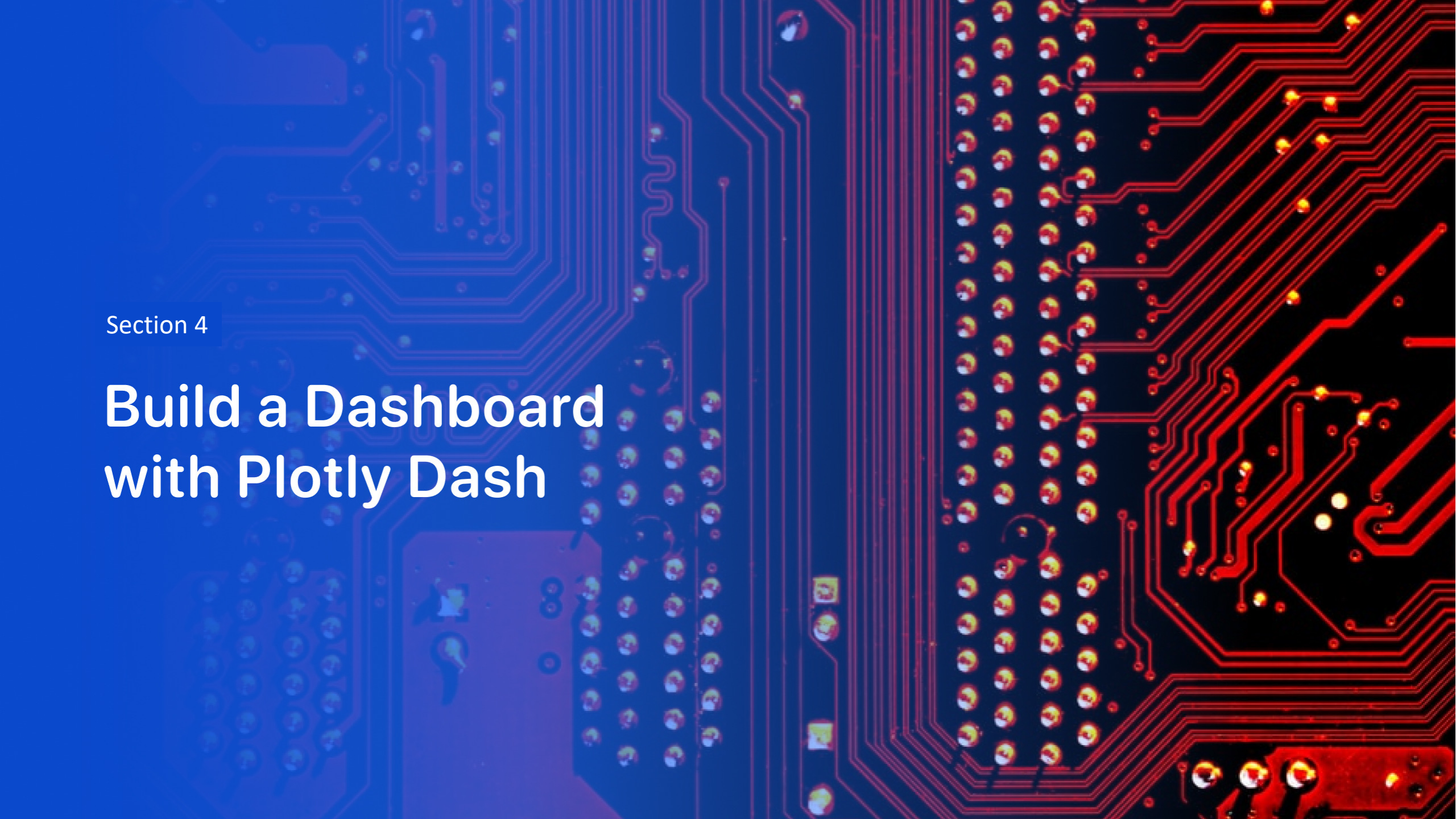
<The success/failed launches for each sites on the map>

- Through this map, we can check the success and failures in each site.

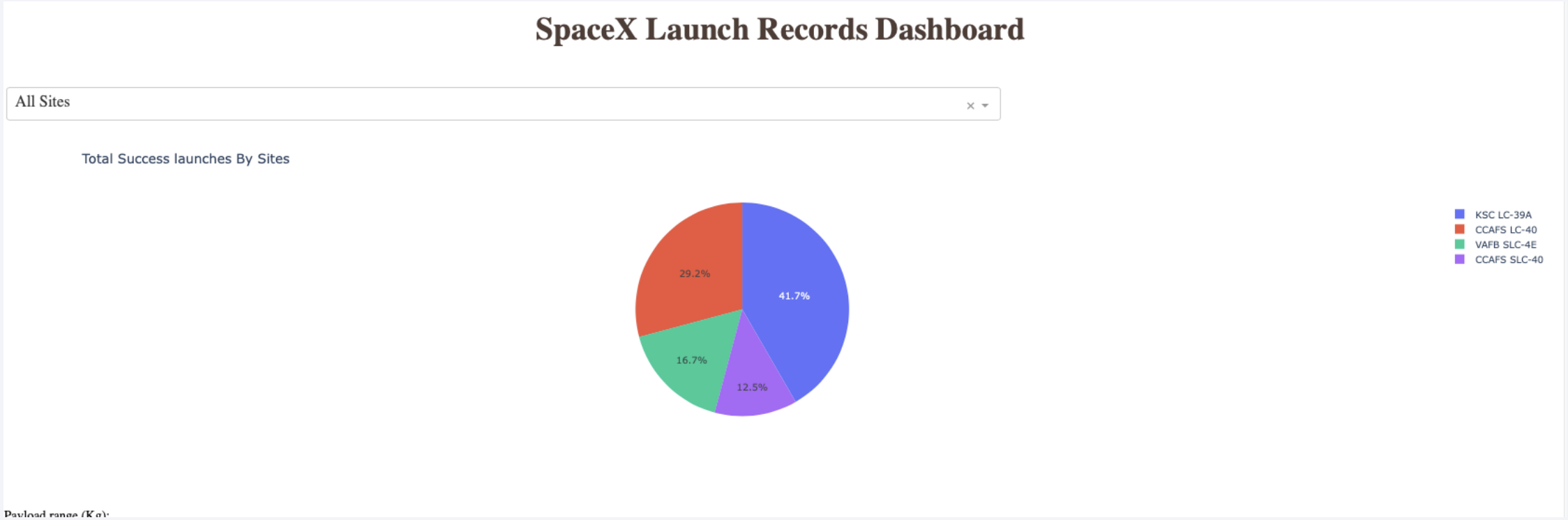


<Distance between a launch site to its proximities>

- Through this part of the map, you can calculate the distance between a launch site and its proximities, eg. Rails, road, and coast.

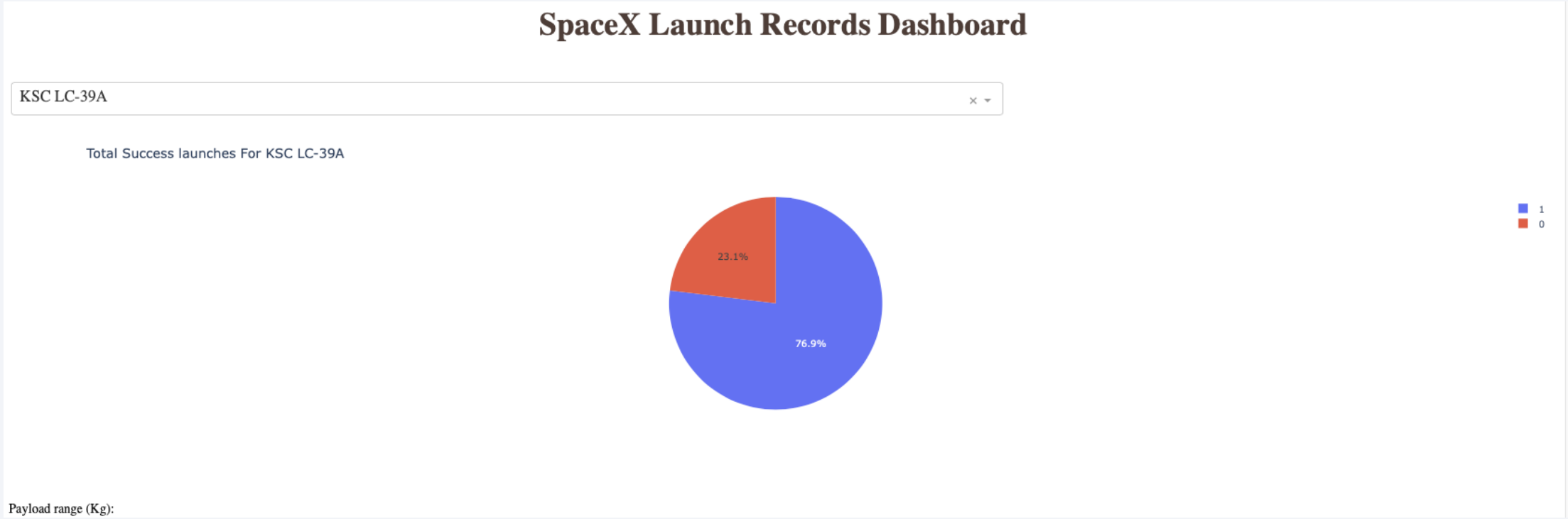

Section 4
Build a Dashboard with Plotly Dash

<Total success launches by sites>



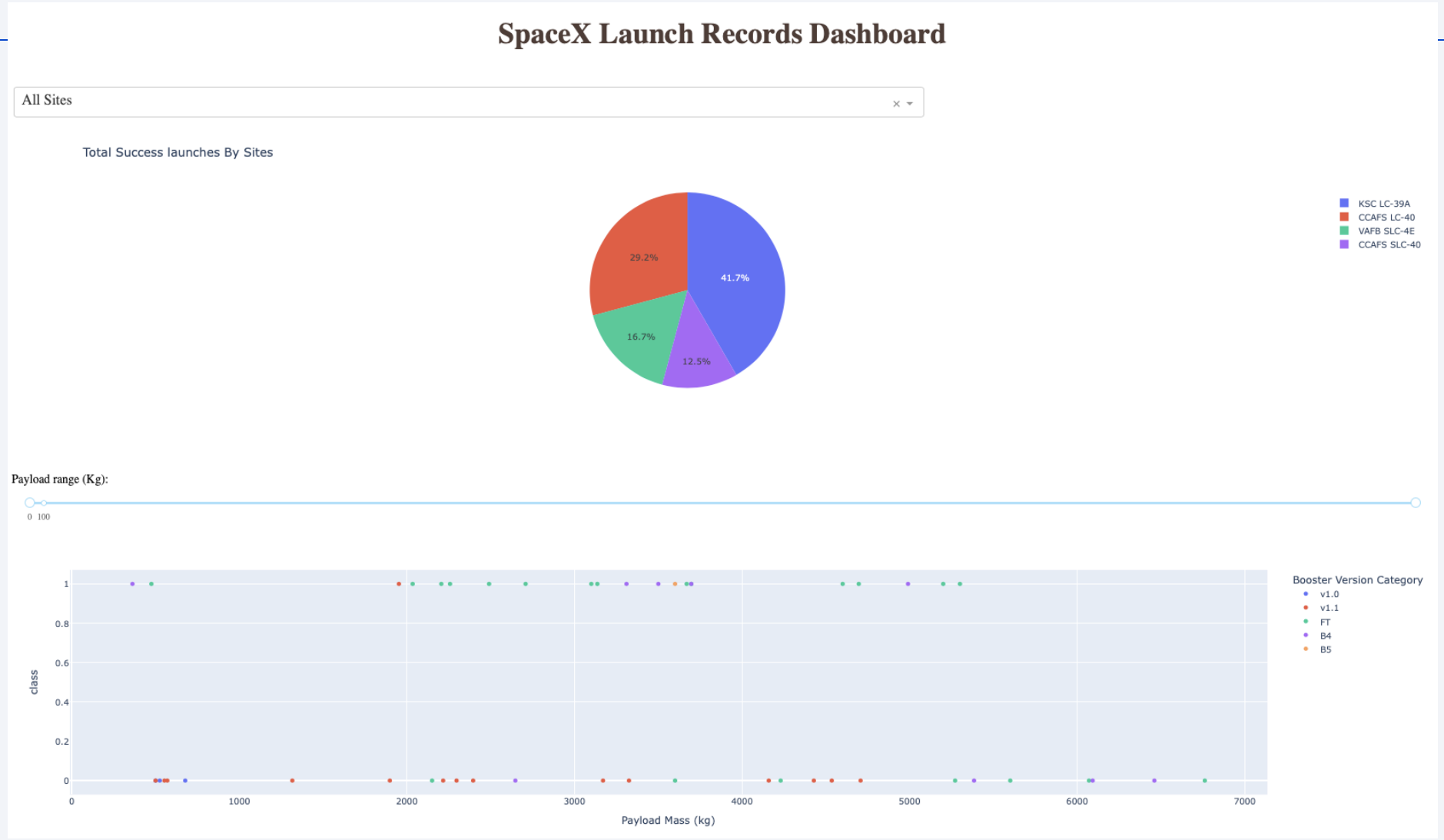
- Launch site KSC LC-39A has the highest success rate of landing.
- Launch site CCAFS SLC-40 has the lowest success rate of landing.

<Success rate of KSC LC-39A >



- Launch site KSC LC-39A has the highest success rate of landing – 76.9%

< Payload vs. Launch Outcome scatter plot for all sites >



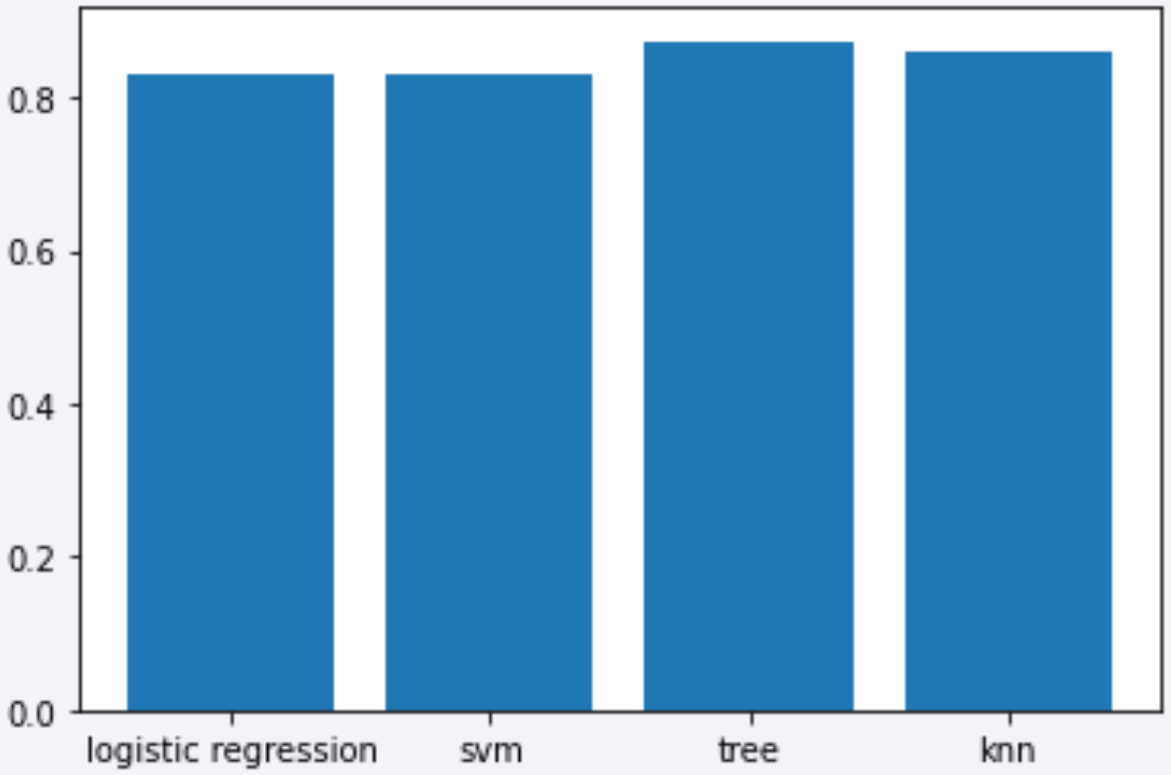
- We can select the range of payload mass and site to see the success.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The results showed that the decision tree classifier has the highest accuracy – 0.874



Confusion Matrix

- This is the confusion matrix of decision tree classifier.
- In the 18 test data point, 3 of them are mis-classified. 2 of them are false negative, 1 of them is false positive.

Conclusions

- The logistic regression model's accuracy is 0.83.
- The svm model's accuracy is 0.83.
- The decision tree model's accuracy is 0.874.
- The KNN model's accuracy is 0.86.

- All 4 models show high accuracy, and they can predict our target well.

Appendix

- Python code reference:

  https://python-visualization.github.io/folium/

  https://plotly.com/python/

  https://realpython.com/python-dash/

  https://plotly.com/python/plotly-express/

Thank you!