IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Xiao Fan
October 4, 2021

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

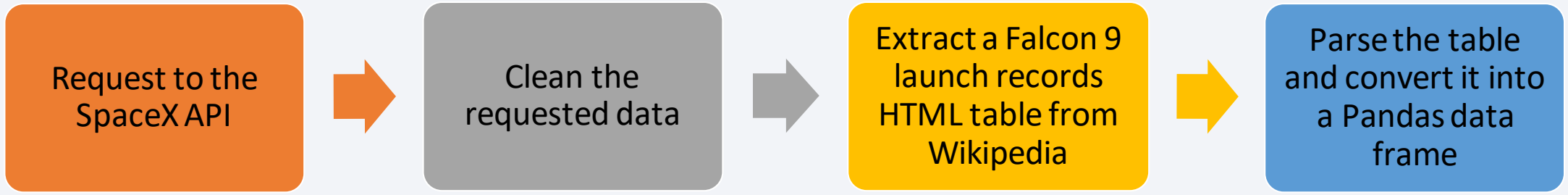
Methodology

Methodology

Executive Summary

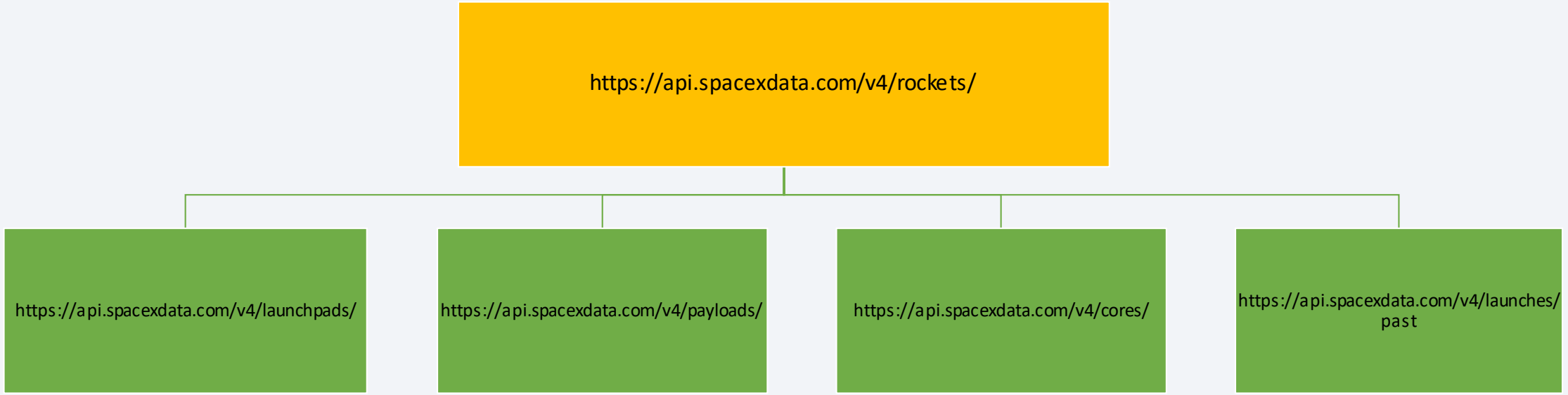
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection



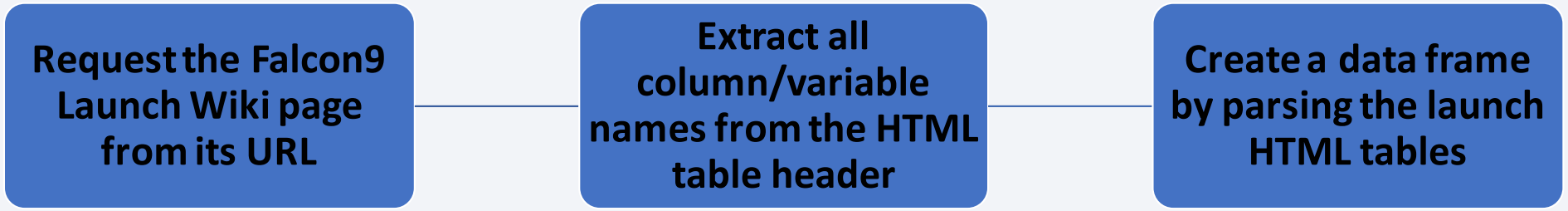
Data Collection – SpaceX API

- GitHub URL: https://github.com/xiaojingfan/testrepo/blob/master/Final_notebook_0.ipynb



Data Collection - Scraping

- GitHub URL:
  https://github.com/xiaojingfan/testrepo/blob/master/Final_Notebook_1.ipynb



Data Wrangling

- In the data set, there are several different cases where the booster did not land successfully.

- We will convert those outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.

- GitHub URL: https://github.com/xiaojingfan/testrepo/blob/master/Final_notebook_2.ipynb

EDA with Data Visualization

- Use scatter plots to show the relationships between Flight Number vs. Launch Site, Payload vs. Launch Site, Flight Number vs. Orbit Type, and Payload vs. Orbit Type.
- Use a bar chart to show the success rate for different orbit types.
- To get the average launch success trend, we use a line chart to plot Year vs. Average Success Rate.
- GitHub URL: https://github.com/xiaojingfan/testrepo/blob/master/Final_notebook_3.ipynb

EDA with SQL

- Use SQL queries to perform the following tasks:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'KSC‘
  - 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 where the successful landing outcome in drone ship was acheived.
  - List the names of the boosters which have success in ground pad 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, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017
  - Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order.
- GitHub URL: https://github.com/xiaojingfan/testrepo/blob/master/Final_notebook_SQL.ipynb

Build an Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities and show it on the map
- GitHub URL:
  https://github.com/xiaojingfan/testrepo/blob/master/Final_Notebook_4.ipynb

Build a Dashboard with Plotly Dash

- Add a Launch Site Drop-down Input Component
- Add a callback function to render success-pie-chart based on selected site dropdown
- Add a Range Slider to Select Payload
- Add a callback function to render the success-payload-scatter-chart scatter plot
- GitHub URL: https://github.com/xiaojingfan/testrepo/blob/master/SpaceX%20Dashboard

Predictive Analysis (Classification)

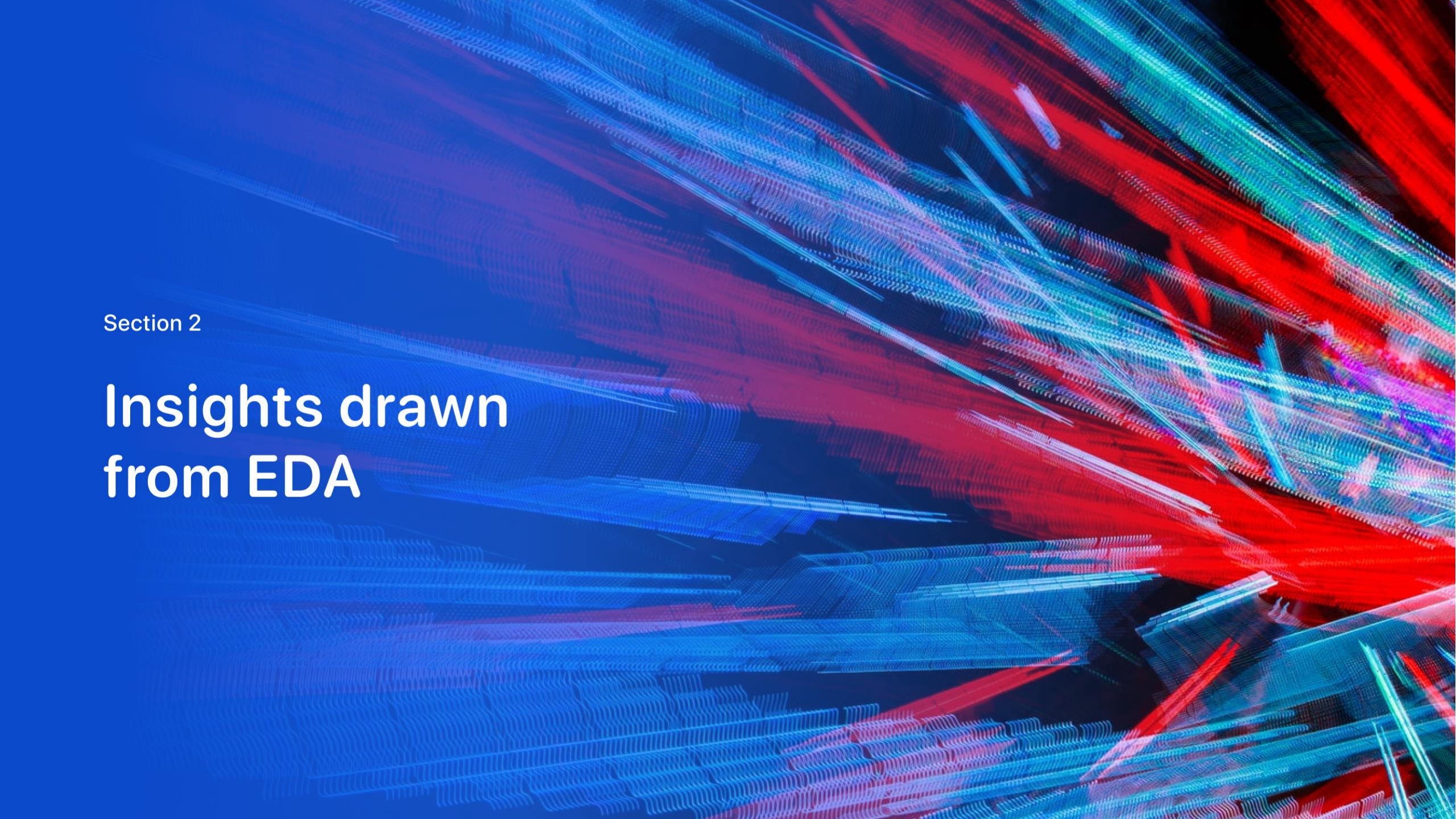
- Create a Numpy array for the column class as the Predicted Variable
- Standardize the predictor variables.
- Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data
- GitHub URL: https://github.com/xiaojingfan/testrepo/blob/master/Final_notebook_5.ipynb

Results

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

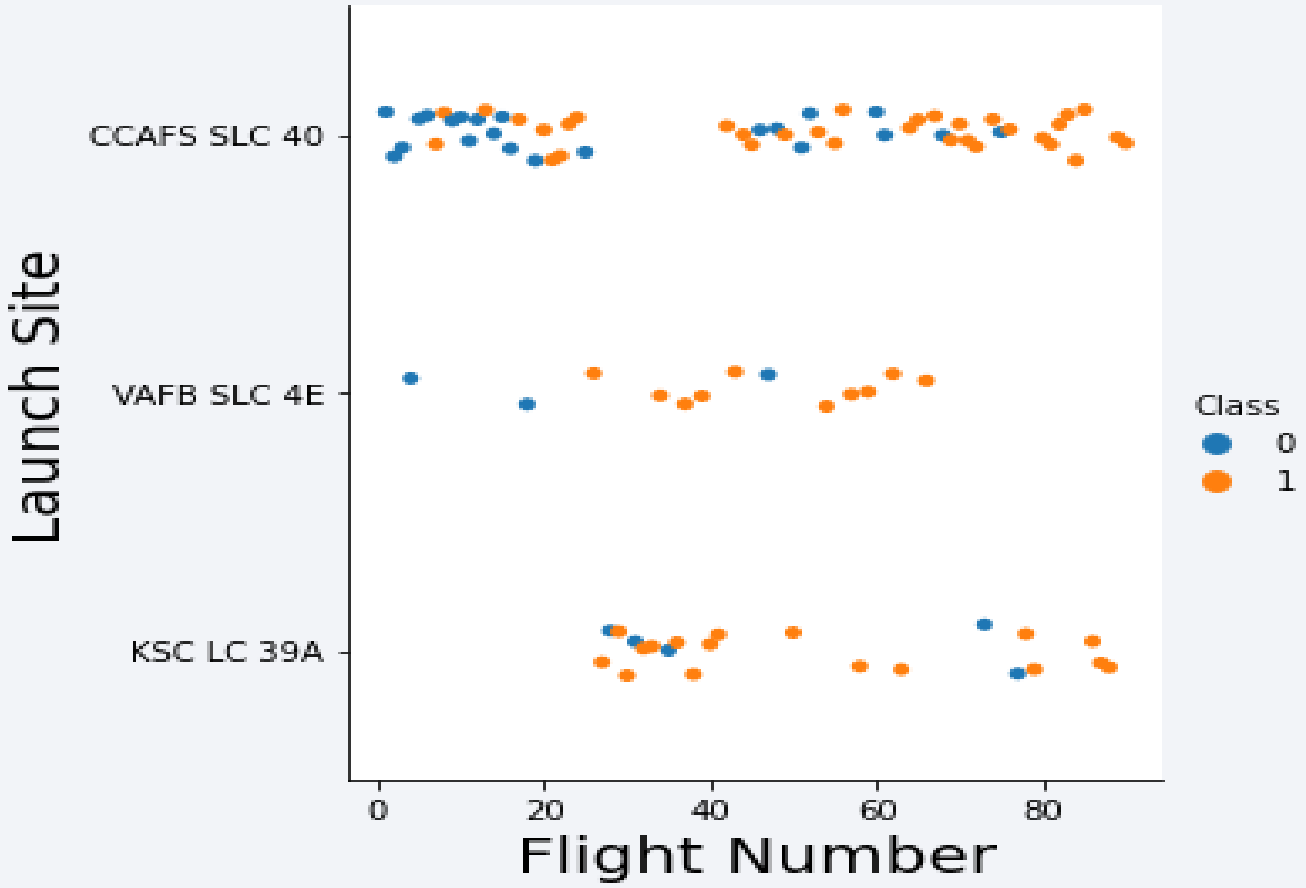
Section 2

Insights drawn from EDA



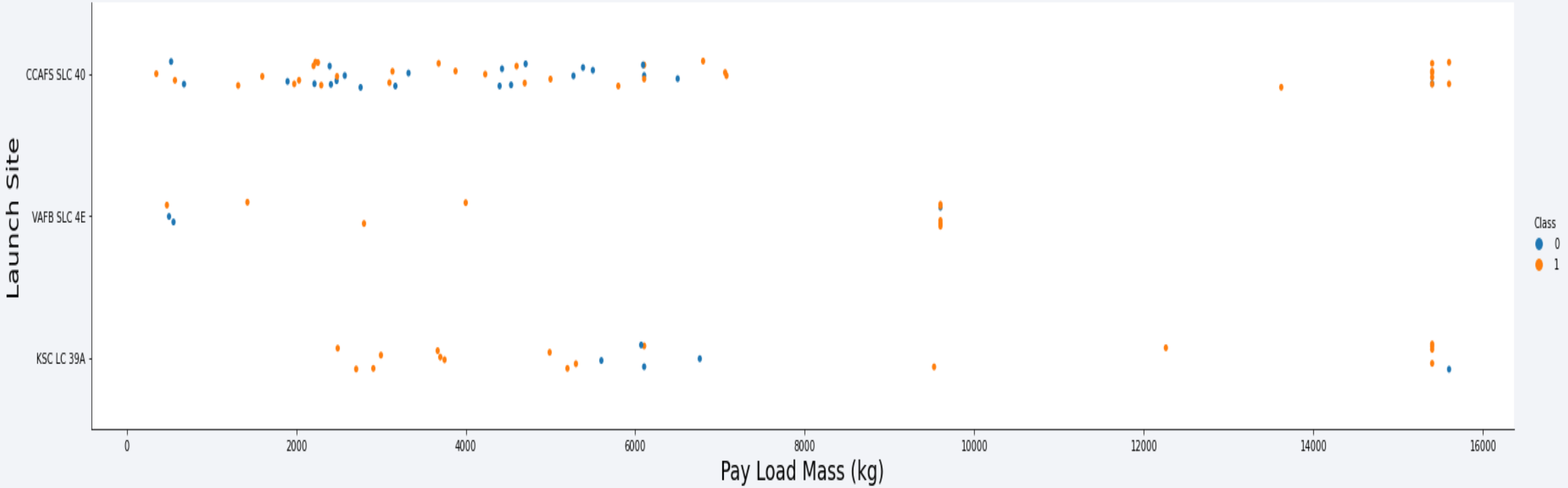
Flight Number vs. Launch Site

- We see that as the flight number increases, the first stage is more likely to land successfully for all launch sites. The launch site KSC LC 39A has the highest success rate.



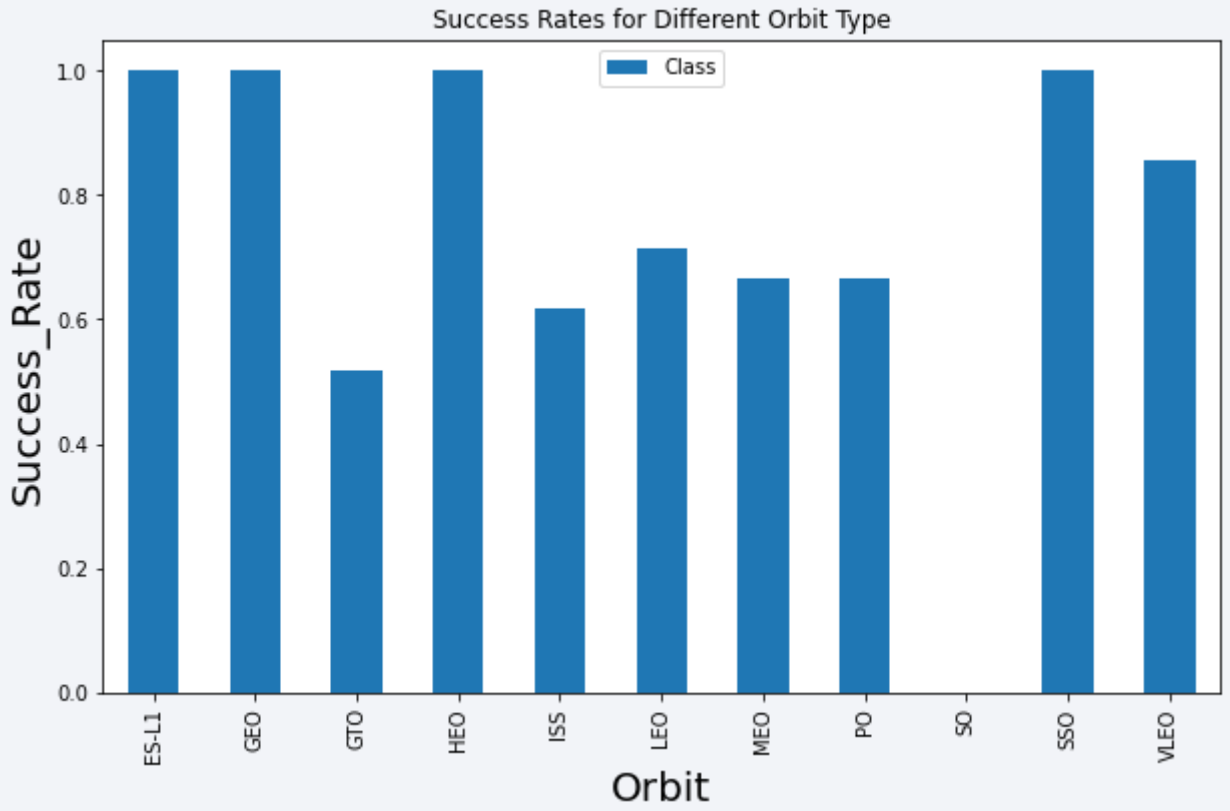
Payload vs. Launch Site

- We see that as the pay load mass increases, the first stage is more likely to land successfully for all launch sites.



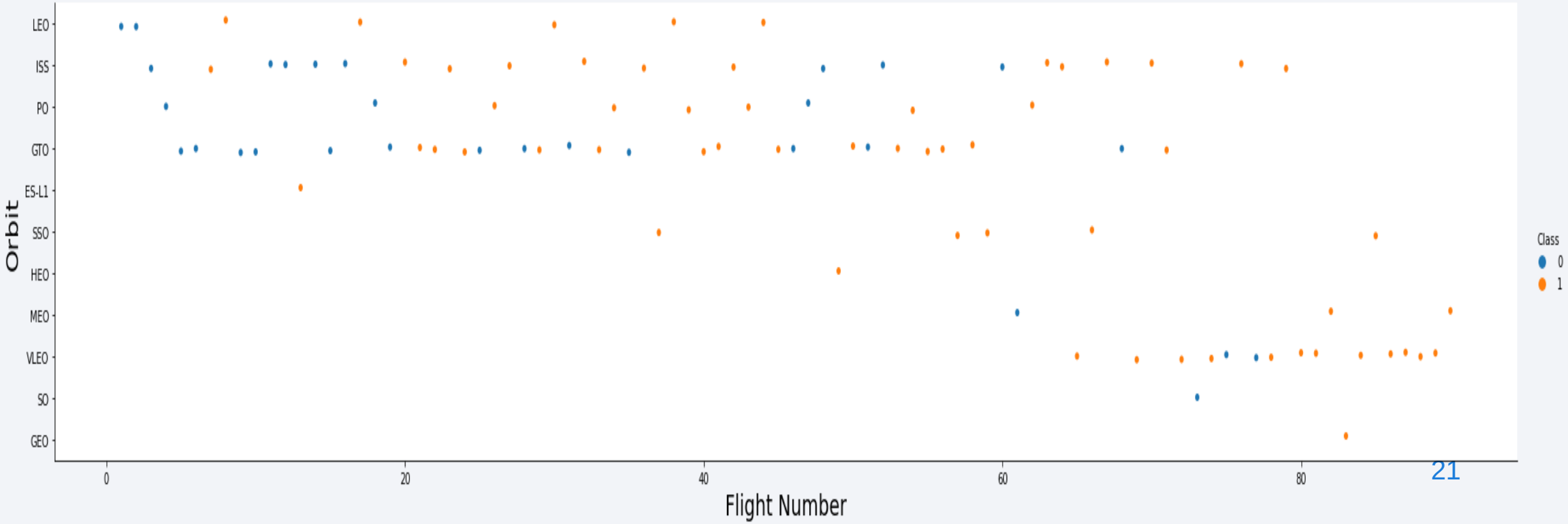
Success Rate vs. Orbit Type

- The orbit types have significant impacts for the success rate. The orbit type ES-L1, GEO, HEO, and SSO has 100% success rate.



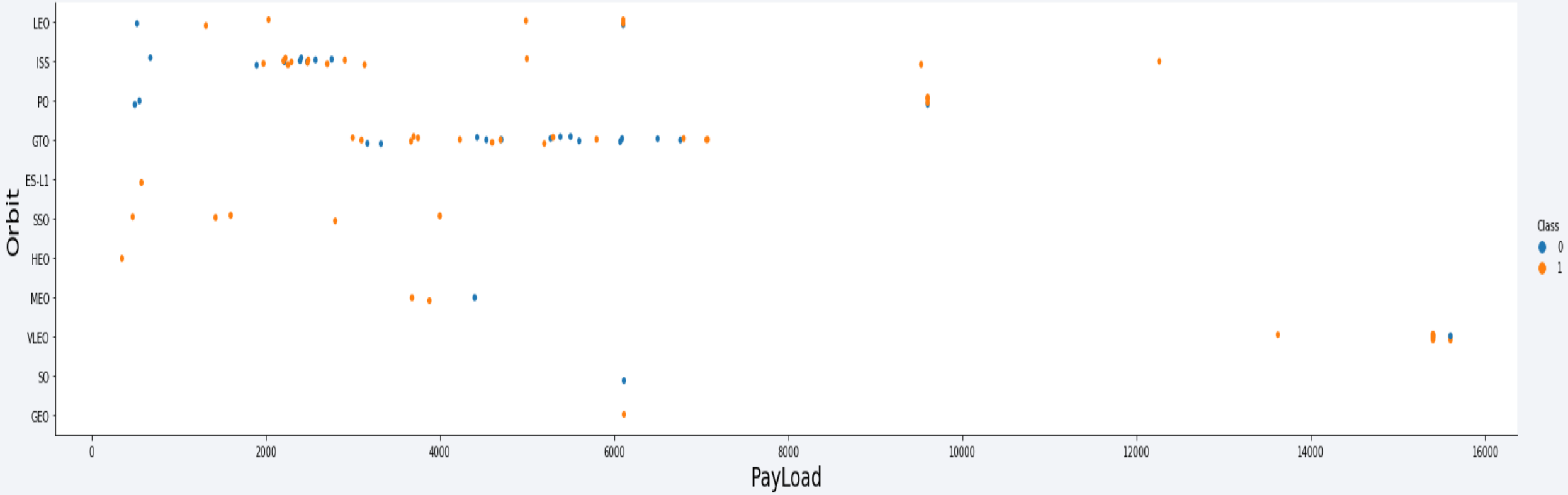
Flight Number vs. Orbit Type

- Flight number and orbit type do not show a strong relationship.



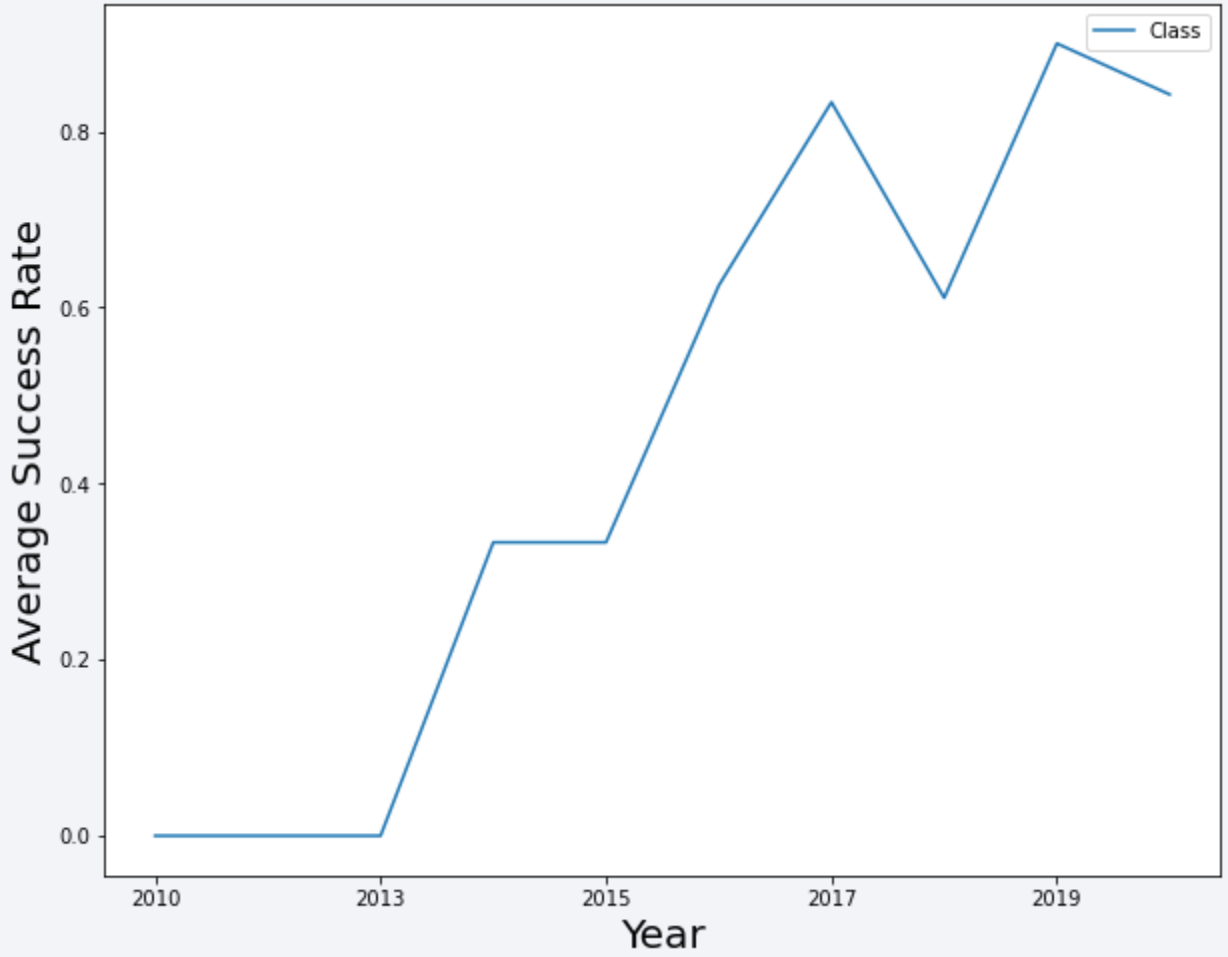
Payload vs. Orbit Type

- Pay load mass and orbit type do not show a strong relationship.



Launch Success Yearly Trend

- We see the success rate generally increases by year, but there is a drop in 2018.



All Launch Site Names

- Find the names of the unique launch sites

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

Launch Site Names Begin with ‘KSC'

- Find 5 records where launch sites begin with `KSC`

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	06:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

average_payload_mass
2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

DATE
2016-04-08

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

booster_version
F9 B4 B1040.1
F9 B4 B1043.1
F9 FT B1032.1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

2015 Launch Records

- List the records which will display the month names, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017

MONTH	landing__outcome	booster_version	launch_site
February	Success (ground pad)	F9 FT B1031.1	KSC LC-39A
May	Success (ground pad)	F9 FT B1032.1	KSC LC-39A
June	Success (ground pad)	F9 FT B1035.1	KSC LC-39A
August	Success (ground pad)	F9 B4 B1039.1	KSC LC-39A
September	Success (ground pad)	F9 B4 B1040.1	KSC LC-39A
December	Success (ground pad)	F9 FT B1035.2	CCAFS SLC

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

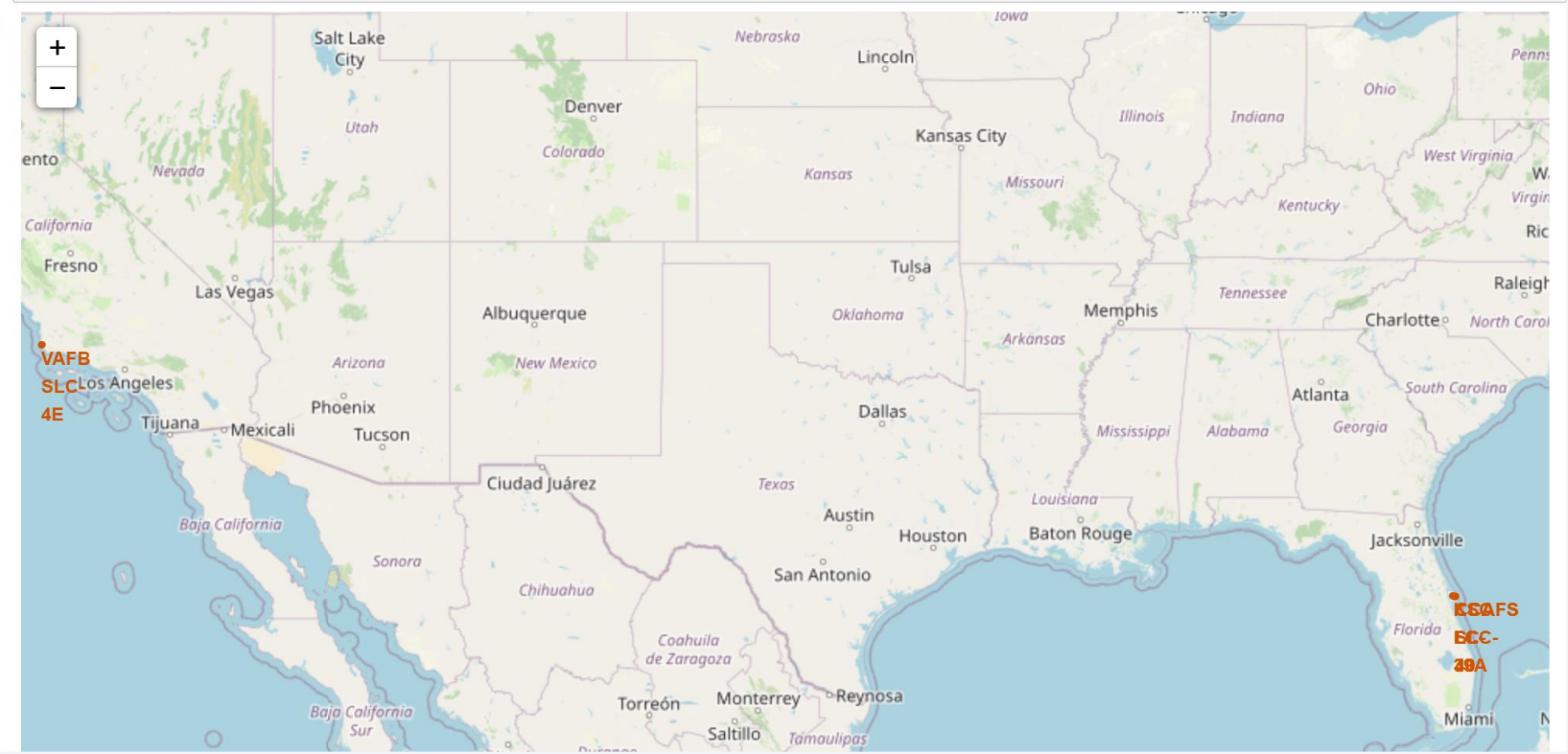
- Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order

DATE	total
2010-06-04	1
2010-12-08	1
2012-05-22	1
2012-10-08	1
2013-03-01	1
2013-09-29	1
2013-12-03	1
2014-01-06	1
2014-04-18	1
2014-07-14	1
2014-08-05	1
2014-09-07	1
2014-09-21	1
2015-01-10	1
2015-02-11	1
2015-03-02	1
2015-04-14	1
2015-04-27	1
2015-12-22	1
2016-01-17	1
2016-03-04	1
2016-04-08	1
2016-05-06	1
2016-05-27	1
2016-06-15	1
2016-07-18	1
2016-08-14	1
2017-01-14	1
2017-02-19	1
2017-03-16	1

Section 4

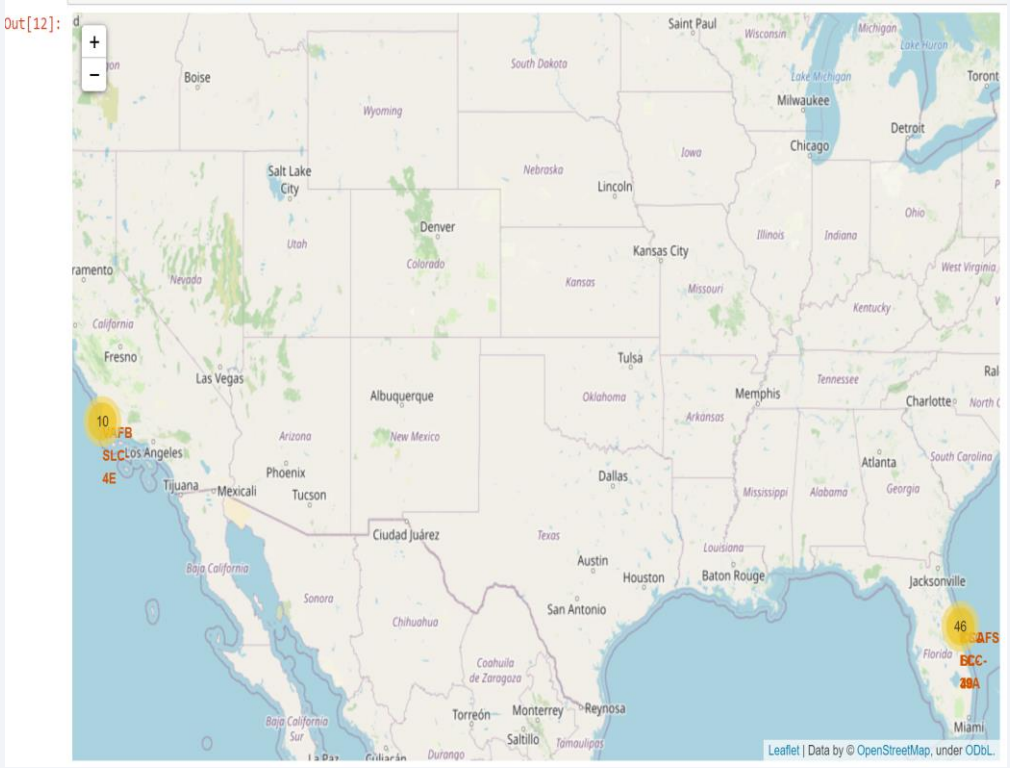
Launch Sites Proximities Analysis

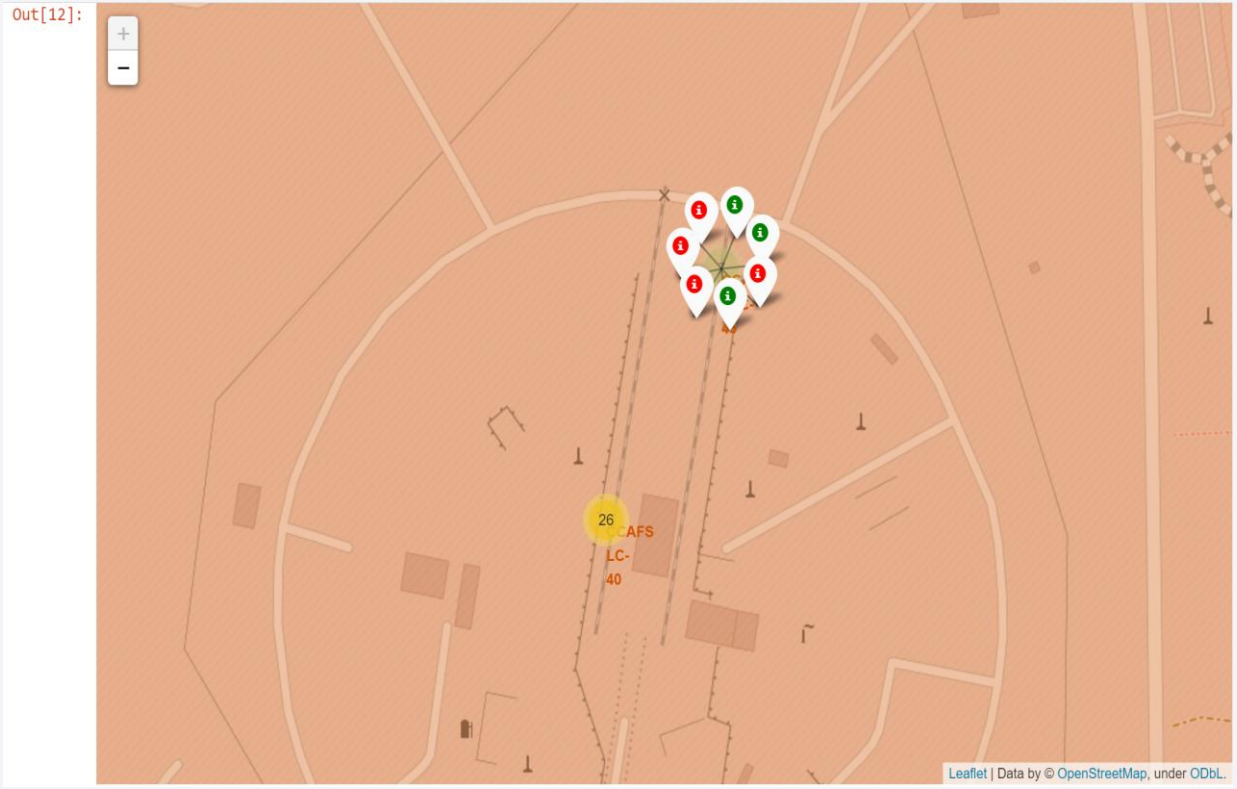
Folium Map with All Launch Sites



The Success/Failed Launches for each site

- The screenshot on the left shows the number of launches on each site.
- The screenshot on the right shows the success (green color) and failed (red color) launches on each site.



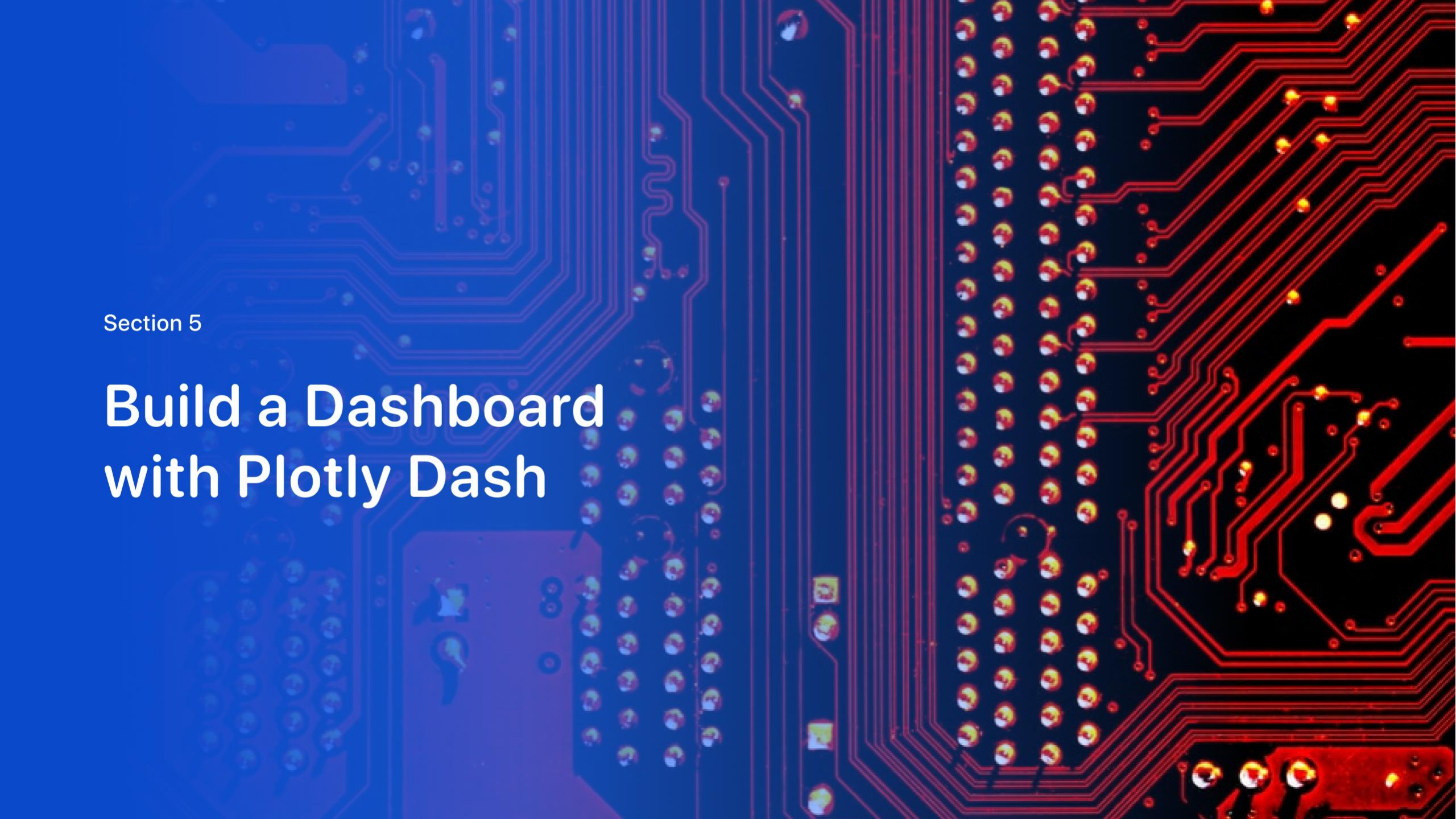


The Distance between a Launch Site to its Proximities

- The screenshot below shows the distance between the launch site CCAFS SLC-40 to the closest railway is 0.90 km.

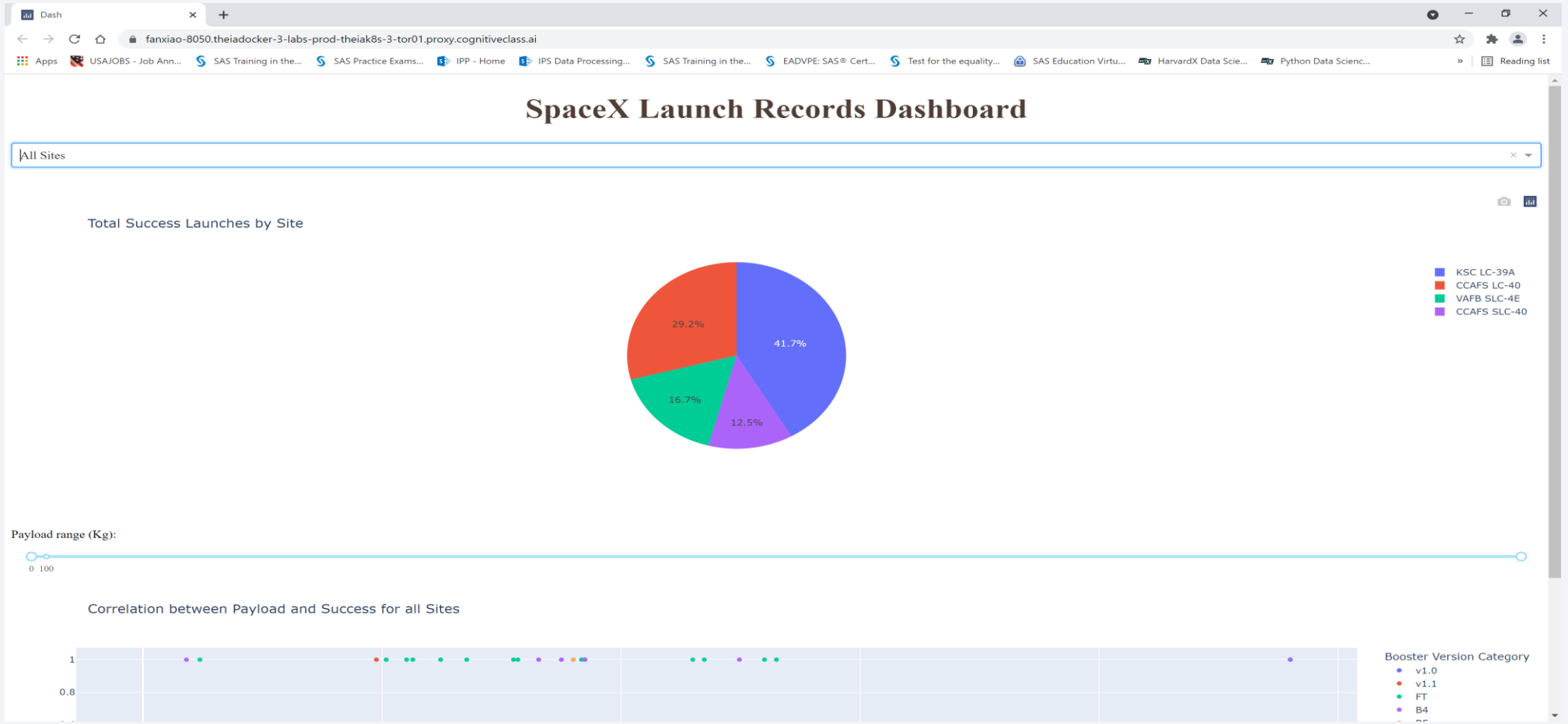
Section 5

Build a Dashboard with Plotly Dash



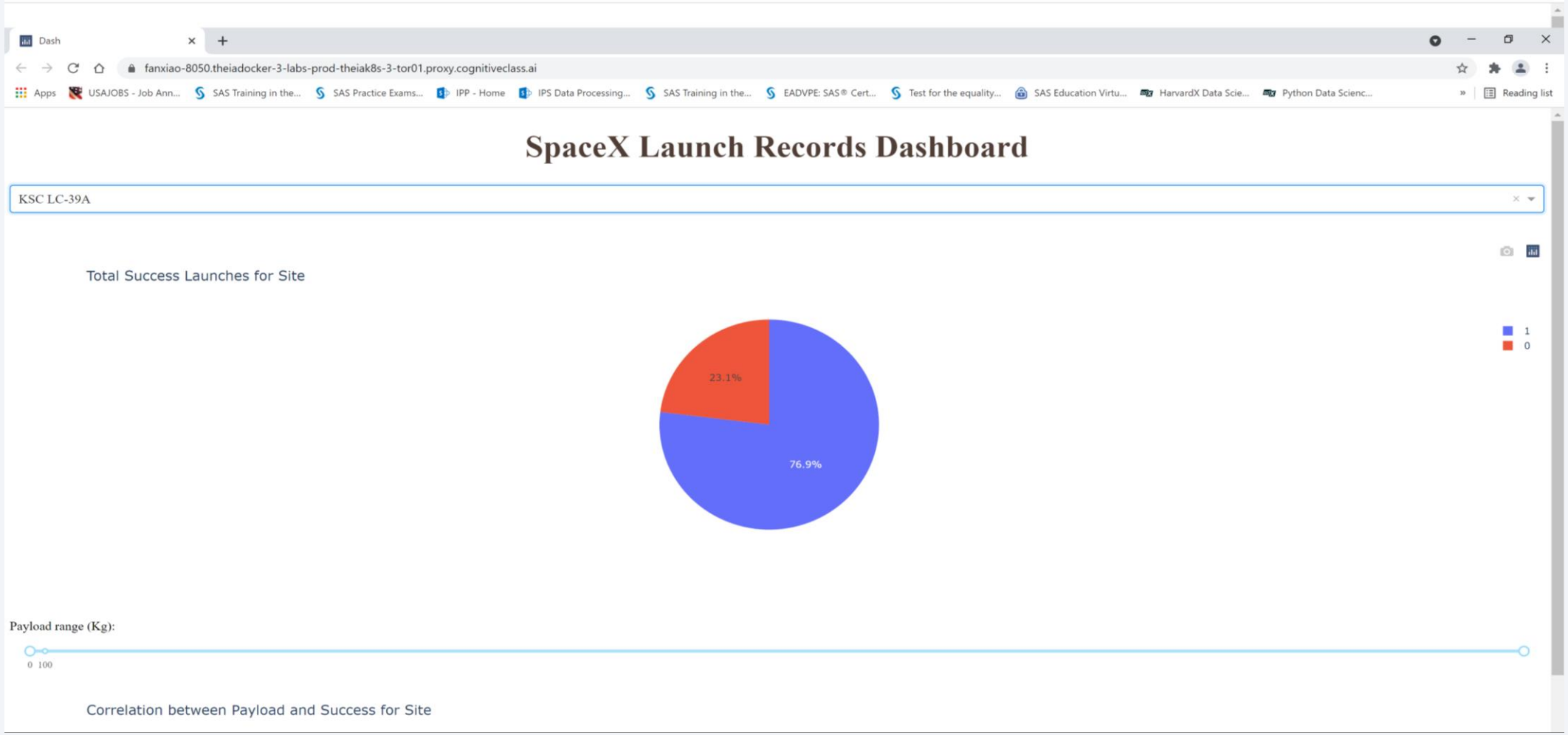
Total Success Launches for all Sites

- Screenshot below shows that the launch site KSC LC-39A has the highest number success launches.



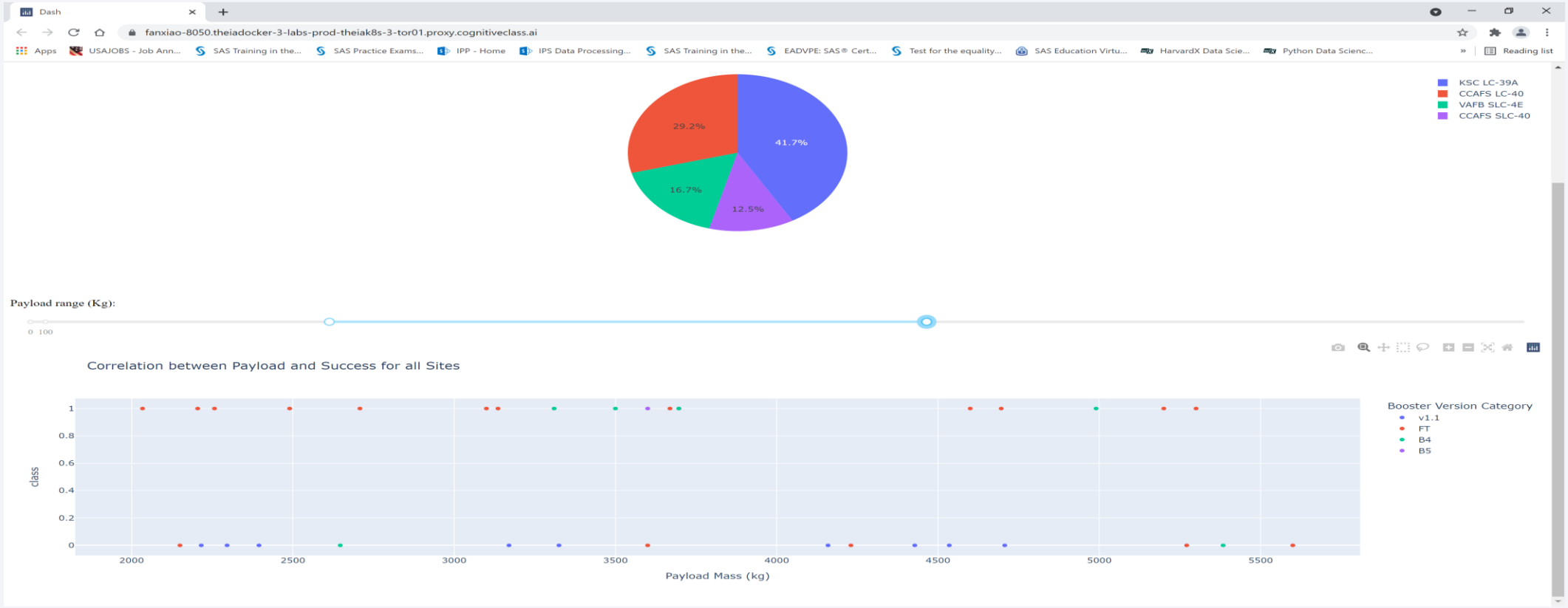
KSC LC-39A Success Rate

- The launch site KSC LC-39A has the highest success rate equal to 76.9%.



Success Rate for Different Booster Versions and Payload Mass

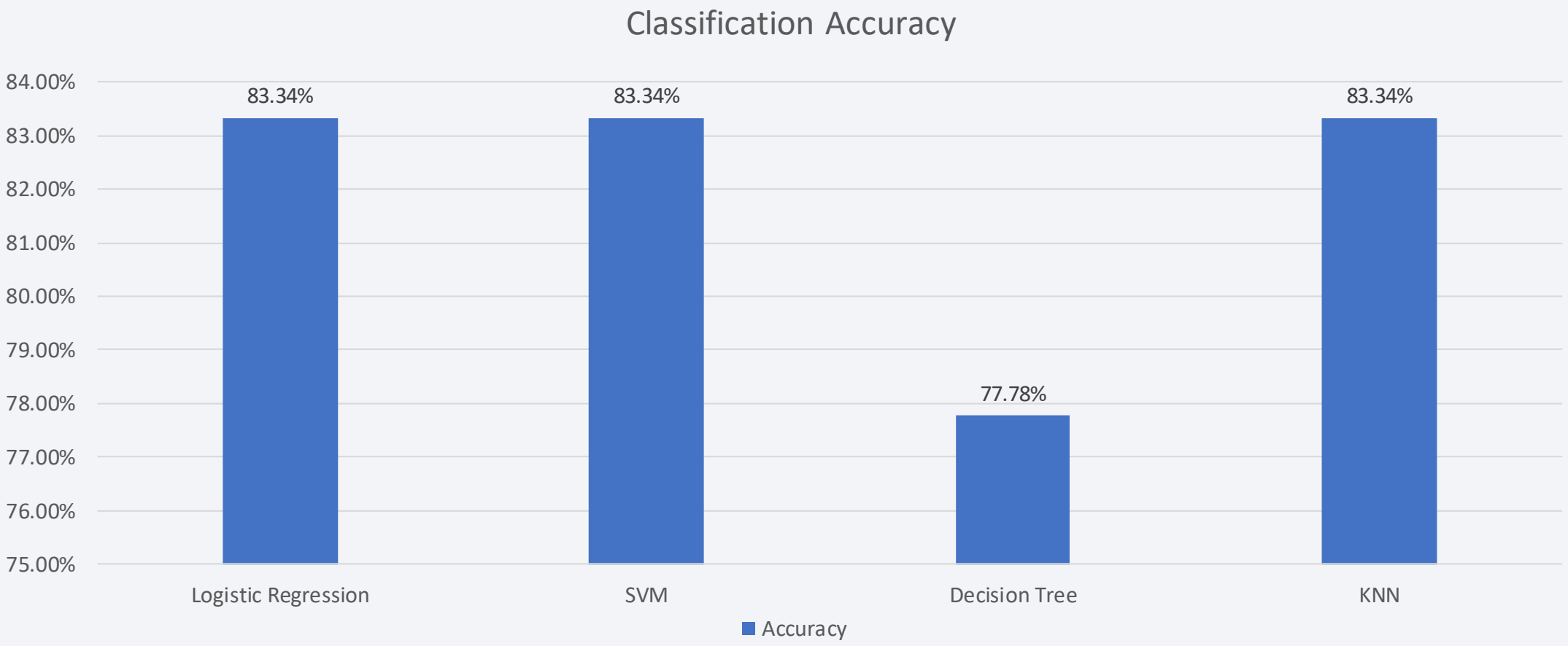
- The booster version FT has the highest success rate.
- The payload mass has highest success rate between 3000kg and 5000kg.



Section 6

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

- All the models perform well with the success launches, but they do not do well with the failed launches. There are 3 failed launches are predicted as successful launches.

Conclusions

- SpaceX average launch rate increases over the years
- In the four launch sites, KSC LC-39A has the highest success rate over the years.
- Payload Mass, Orbit Type, and Booster Version all have significant impact on the launch outcome.
- Predictive models do good jobs to predict the successful launches, but do not have good predictions for the failed launches.

Thank you!