IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Wataru Danjo
2022.10.10

Outline

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

Executive Summary

In this capstone project, we will predict if the Falcon 9 first stage will land successfully.

We investigate the actual landing data of SpaceX and find some relations.

From this project result, we can determine that the best prediction model are SVM, Logistic regression and KNN model.

This prediction model will help us make our strategy for future investment.

Introduction

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

From this points, our question is “Will the first stage land successfully with given Falcon 9?”

Section 1

Methodology

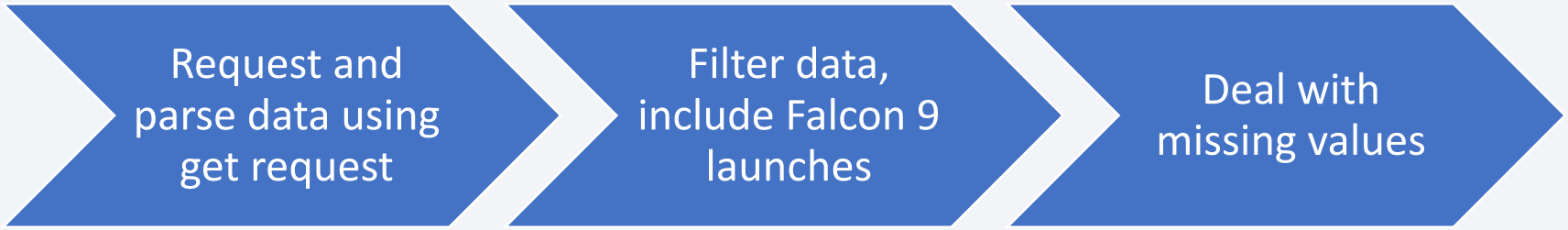
Methodology

Data is collected from the SpaceX API and web scraping from Wikipedia. Then we arrange our data with numerical number which can be used for predicting models.

Using this useful data, we visualize them with Seaborn and SQL. Also, we create map with Folium and Dashboard with Plotly Dash.

Finally, we test some prediction model and decide the best prediction model.

Data Collection – SpaceX API



Git hub URL: https://github.com/watanjo/CapstoneProject/blob/master/1.CollectingTheData.ipynb

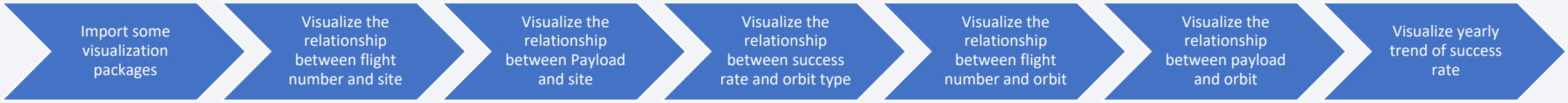
Data Collection - Scraping

Git hub URL: https://github.com/watanjo/CapstoneProject/blob/master/2.DataCollectionWithWebScraping.ipynb

Data Wrangling

Git hub URL: https://github.com/watanjo/CapstoneProject/blob/master/3.DataWrangling.ipynb

EDA with Data Visualization



Git hub URL: https://github.com/watanjo/CapstoneProject/blob/master/5.EDAwithVisualization.ipynb

EDA with SQL

- Displaying the names of the unique launch sites in the space mission **---** Displaying 5 records where launch sites begin with the string 'CCA' --- Displaying the total payload mass carried by boosters launched by NASA (CRS) --- Displaying average payload mass carried by booster version F9 v1.1,
- Listing the date when the first successful landing outcome in ground pad was achieved --- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 --- Listing the total number of successful and failure mission outcomes --- Listing the names of the booster versions which have carried the maximum payload mass --- Listing the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 --- Ranking the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order.

Git hub URL: https://github.com/watanjo/CapstoneProject/blob/master/4.EDAwithSQL.ipynb

Build an Interactive Map with Folium

Using an interactive map with folium, we mark the success / failed launches for each site on the map.

And find these objects.

- Are launch sites in close proximity to railways? : Yes
- Are launch sites in close proximity to highways? : Yes
- Are launch sites in close proximity to coastline? : Yes
- Do launch sites keep certain distance away from cities? : Yes

Git hub URL: https://github.com/watanjo/CapstoneProject/blob/master/5.EDAwithVisualization.ipynb

Build a Dashboard with Plotly Dash

We create two charts in application

- Pie chart : Success rate by each site
- Scatter plot : Relationship between landing outcome and payload mass

From this result, our findings are below

- Which site has the largest successful launches?
- Which site has the highest launch success rate?
- Which payload range(s) has the highest launch success rate?
- Which payload range(s) has the lowest launch success rate?
- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest

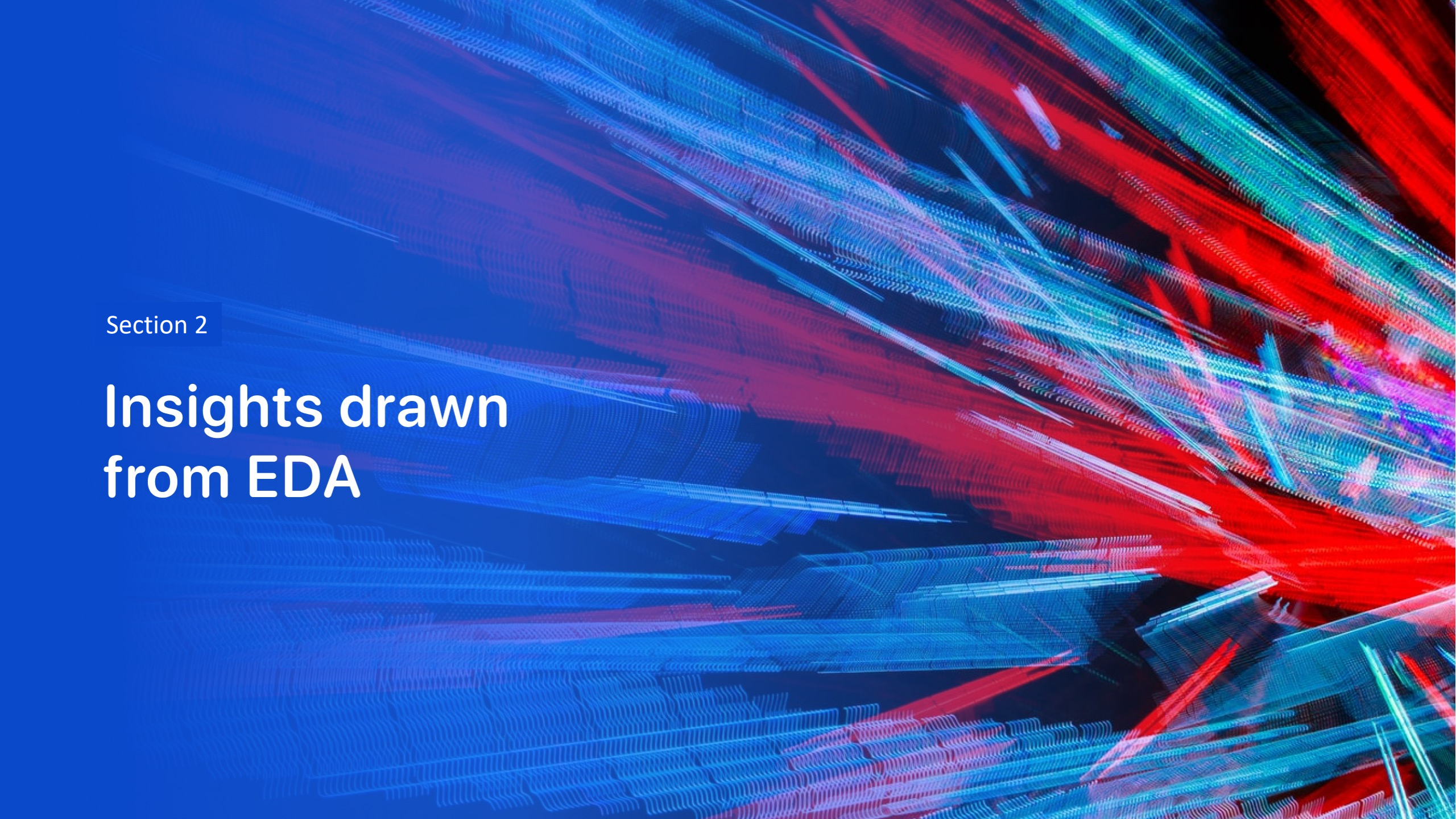
Git hub URL:https://github.com/watanjo/CapstoneProject/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

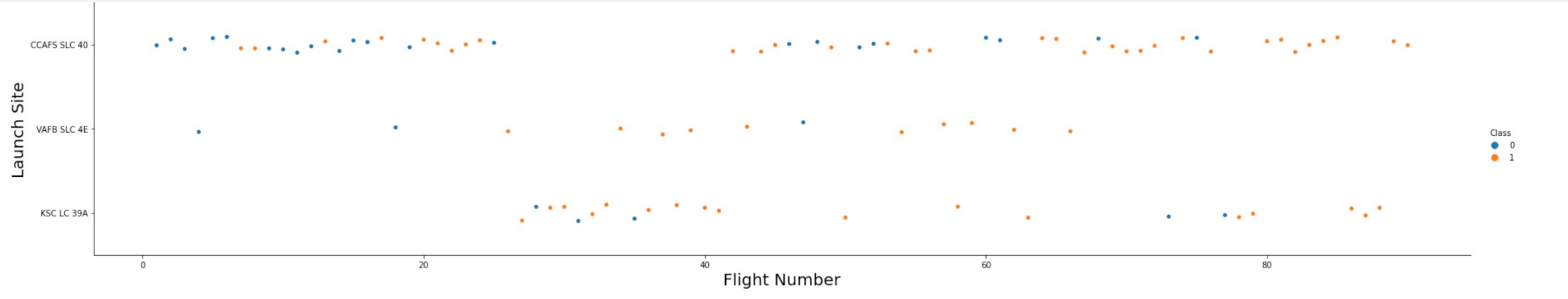
Git hub URL: https://github.com/watanjo/CapstoneProject/blob/master/7.MachineLearningPrediction.ipynb

Results

- The results of the exploratory data analysis revealed that the success rate of the Falcon 9 landings was 66.66%
- The predictive analysis results showed that Logistic regression, SVM and KNN model are the best classification method with an accuracy of 83%

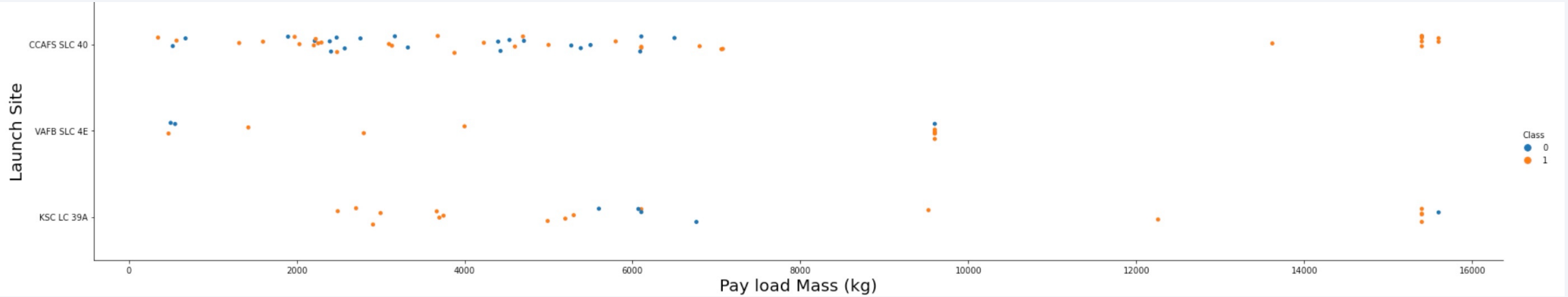

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



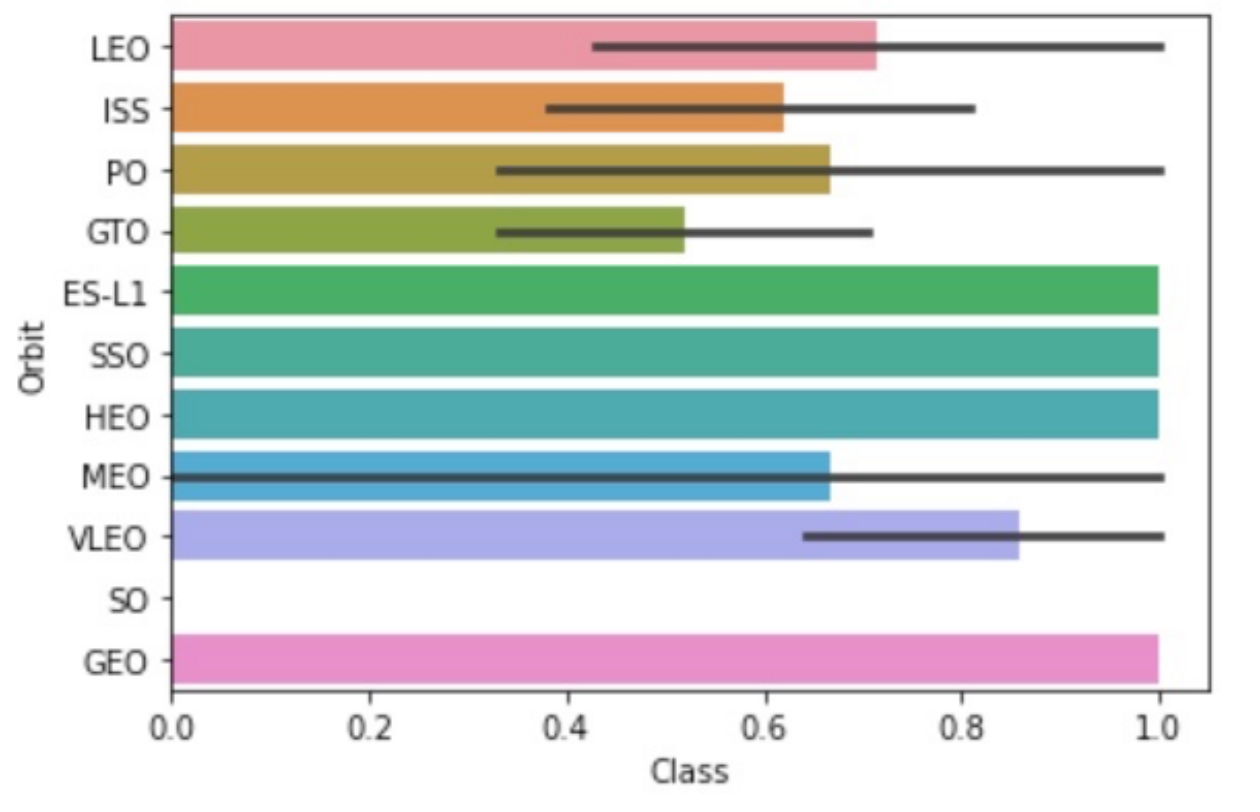
- This figure shows that the success rate increased as the number of flights increased.
- The blue dots represent the successful launches while the orange dot represent unsuccessful luanches.
- There seems to be an increase in successful flights after the 40th launch.

Payload vs. Launch Site



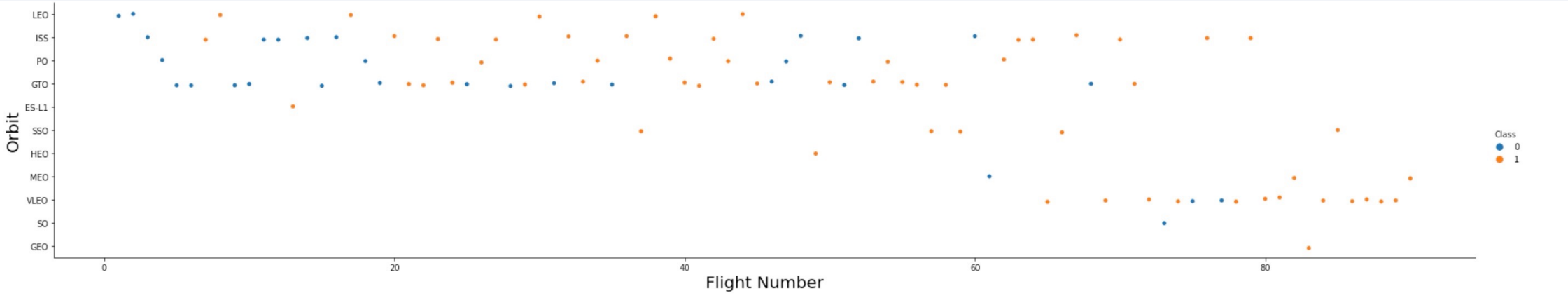
- The blue dots represent the successful launches while the orange dots represent unsuccessful launches.
- For the VAFB-SLC launchsite there are no rockets launched for heavy payload mass
- There seems to be a weak correlation between Payload and Launch Site and therefore decisions cannot be made using this metric.

Success Rate vs. Orbit Type



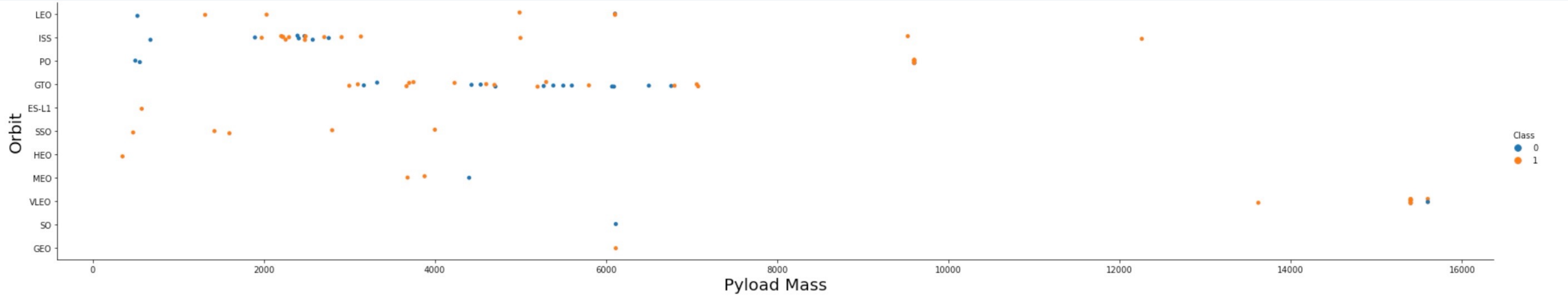
Orbits SSO, HEO, GEO, and ES-L1 have 100% success rates.

Flight Number vs. Orbit Type



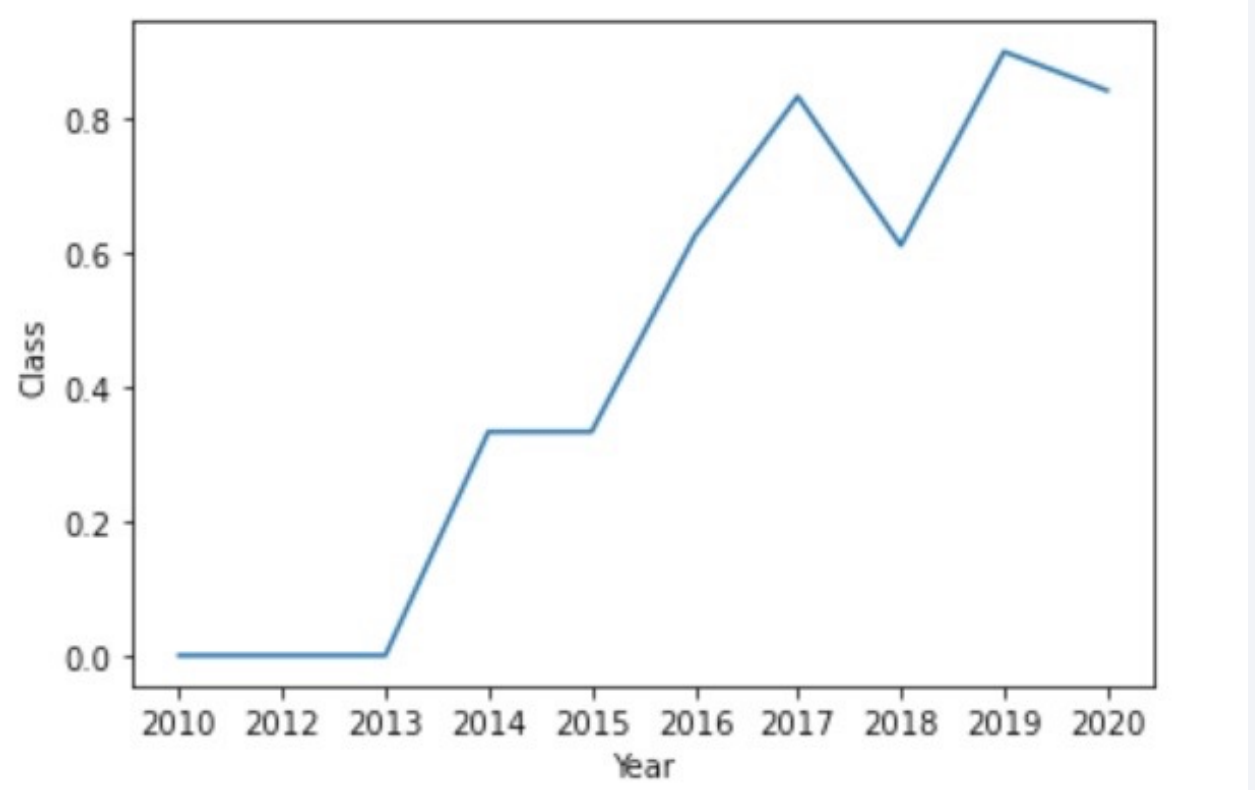
- The blue dots represent the successful launches while the orange dots represent unsuccessful launches.
- In the LEO orbit, the success is positively correlated to the the number of flights.
- There seems to be no relationship between flight number in the GTO orbit.
- The SSO orbit has a 100% success rate however with fewer flights than the other orbits
- FLights numbers greater than 40 have a higher success rate than flight numbers between 0-40.

Payload vs. Orbit Type



- As the payloads get heavier, the success rate increases in the PO, SSO, LEO and ISS orbits.
- There seems to be no direct correlation between obrit type and payload mass for GTO orbit as both successful and failed launches are eually present

Launch Success Yearly Trend



- The general trend of the chart shows an increase in landing success rate as the years pass. There is however a dip in 2018 as well as in 2020.

All Launch Site Names

- Those are all launch site names

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

Launch Site Names Begin with 'CCA'

- Launch sites 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	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

Total Payload Mass

- The total payload mass carried byboosters launched by NASA(CRS)

the total payload mass
45596

Average Payload Mass by F9 v1.1

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

the average payload mass
2928

First Successful Ground Landing Date

- The first successful landing outcome in ground pad was achieived

	1
	2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- The names of the booster_versions 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

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

booster_version	launch_site	landing__outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

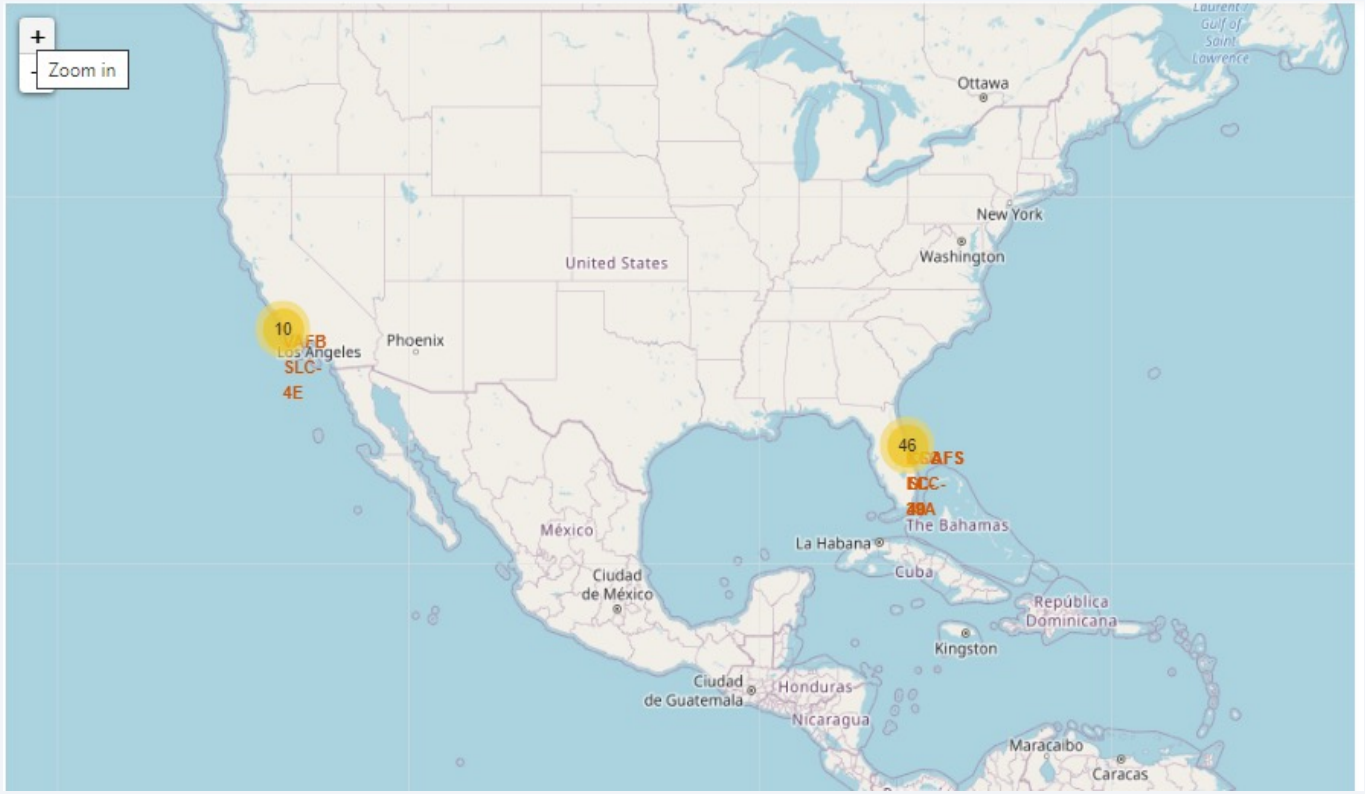
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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

landing__outcome	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

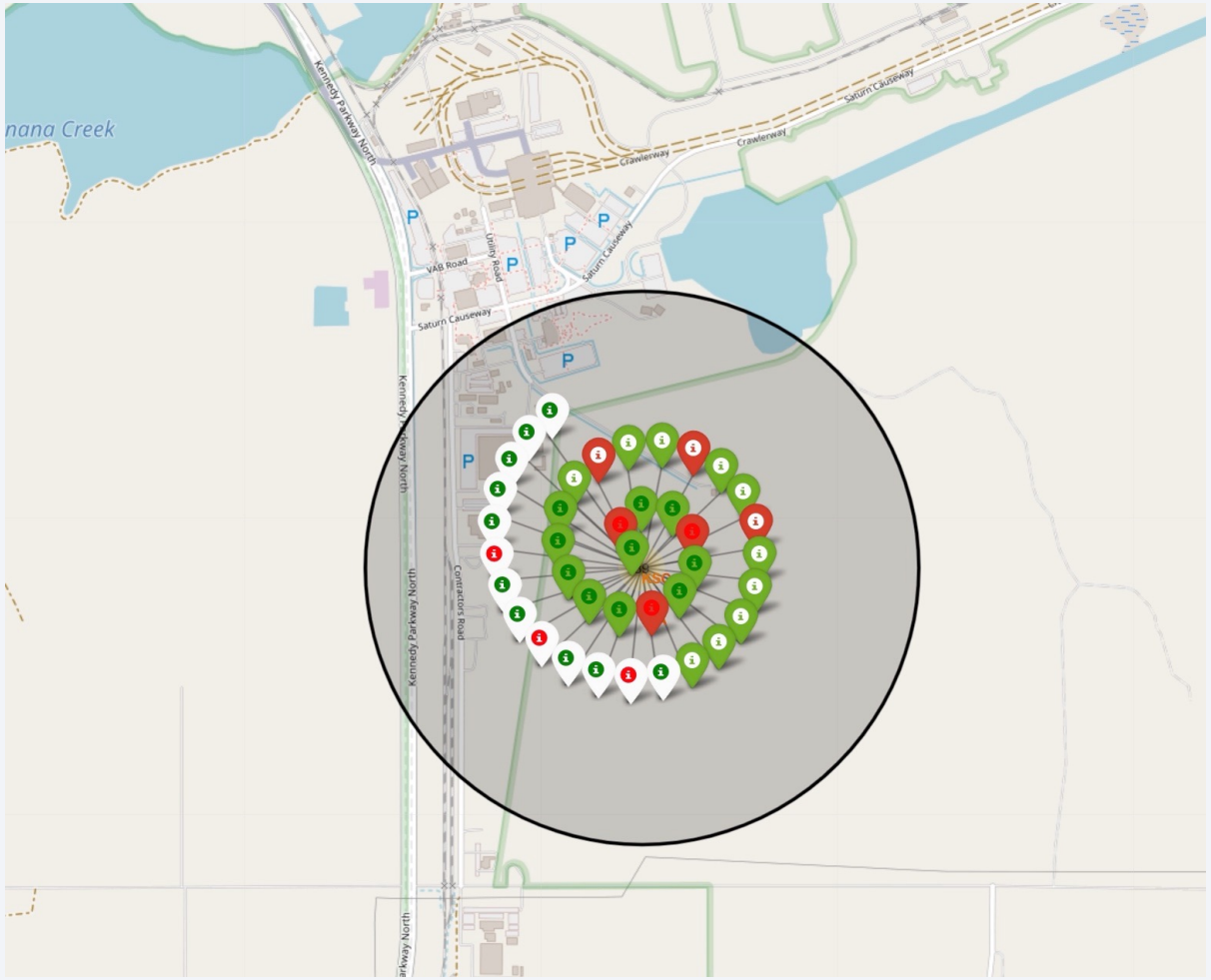

Section 3
Launch Sites
Proximities Analysis

SpaceX Launch Sites Locations



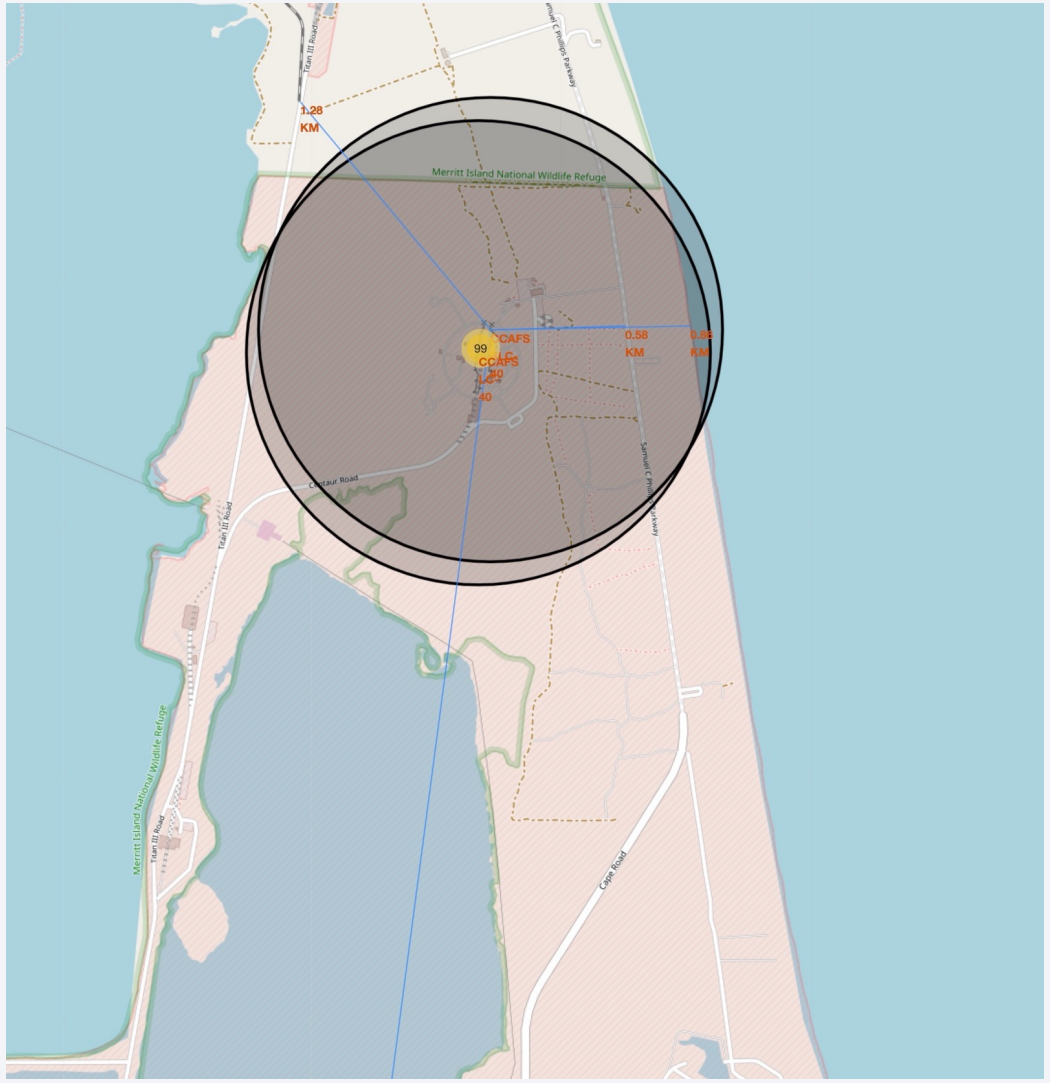
- The yellow markers are indicators of where the locations of all the SpaceX launch sites are situated in the US.
- The launch sites have been strategically placed near the coast

Success or failure point

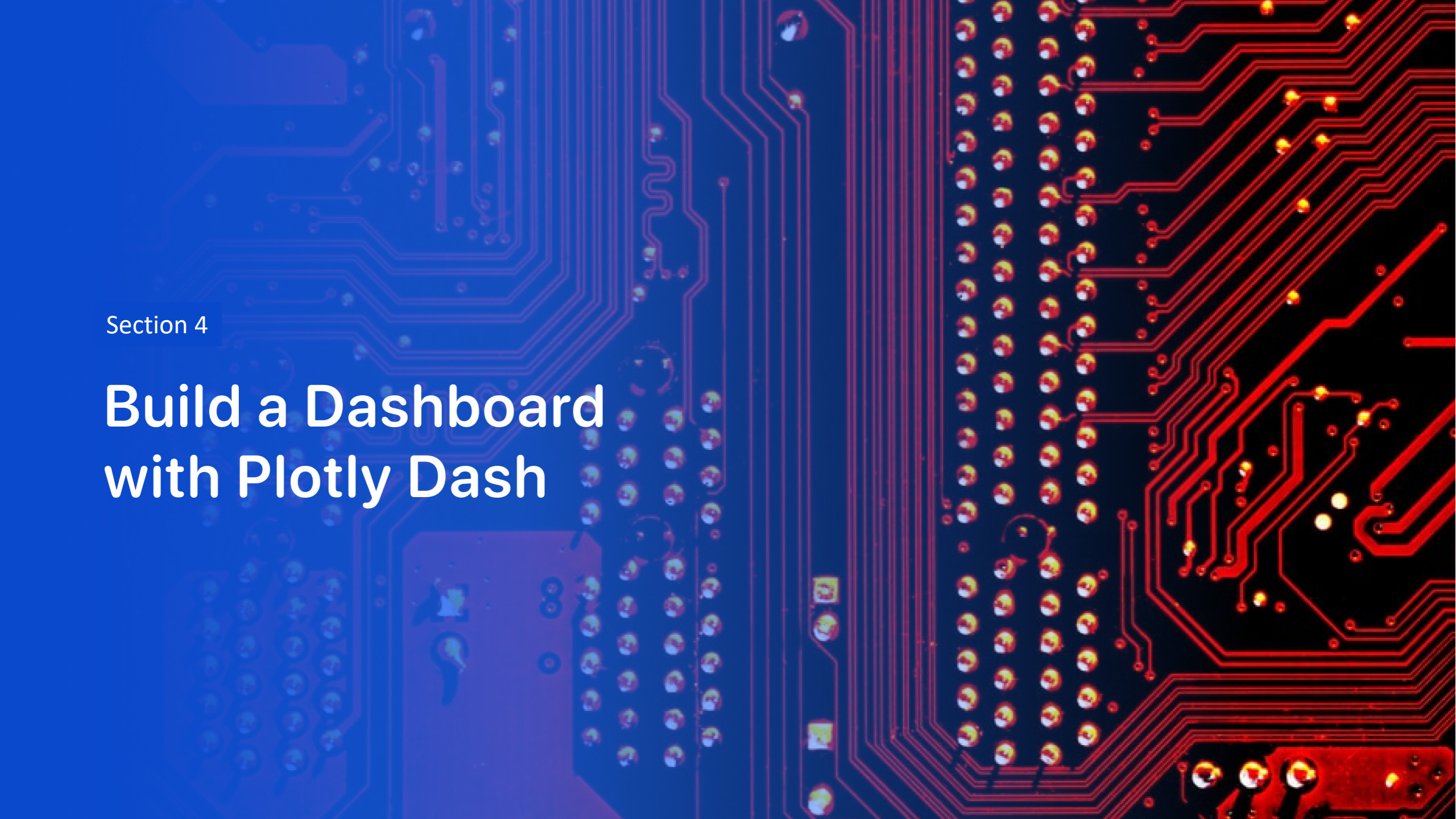


- When we zoom in on a launch site, we can click on the launch site which will display marker clusters of successful landings (green) or failed landing (red).

Location proximities

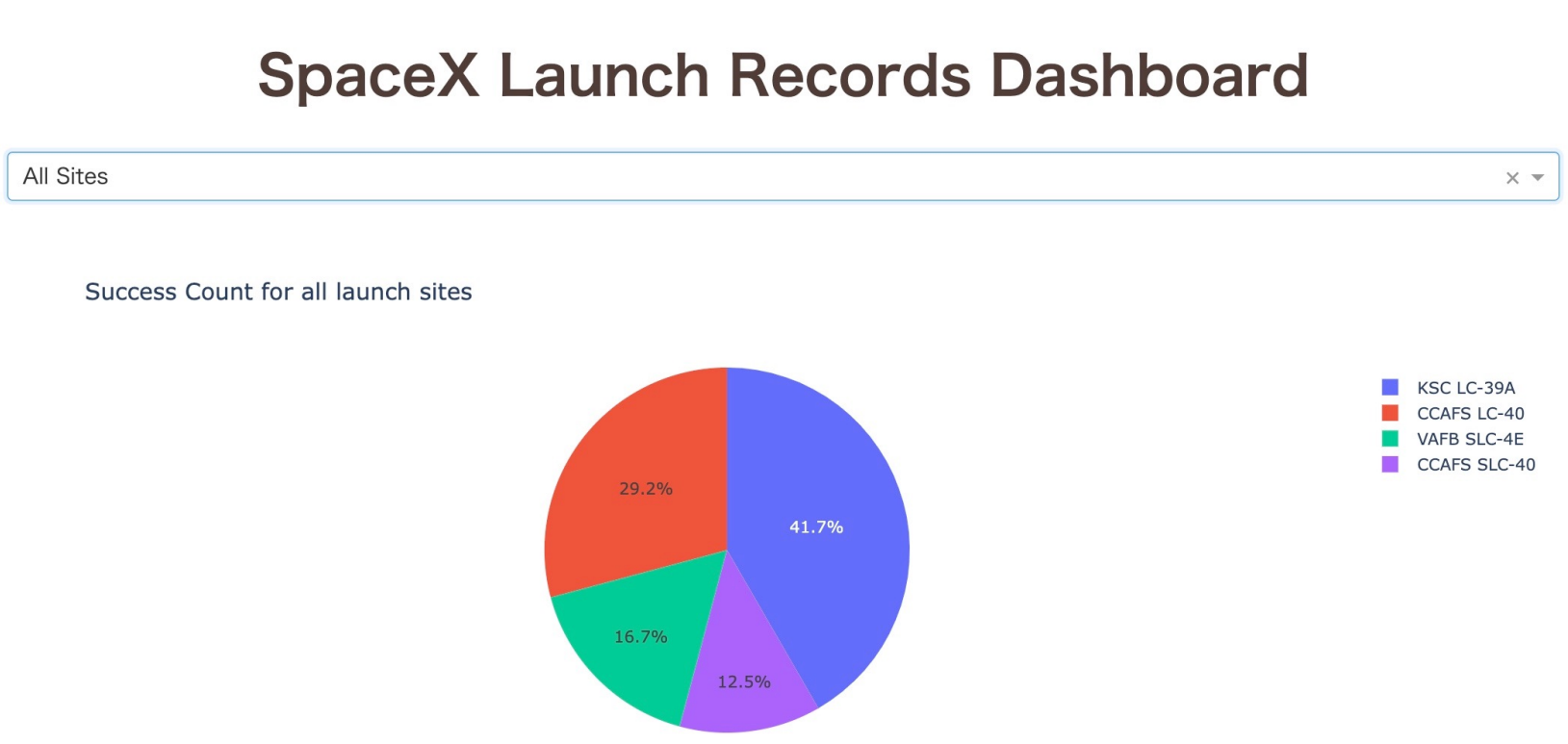


- The generated map shows that the selected launch site is close to a highway for transportation of personnel and equipment. The launch site is also close to the coastlines for launch failure testing.
- The launch sites also maintain a certain distance from the cities. (Can be viewed in notebook).


Section 4
Build a Dashboard with Plotly Dash

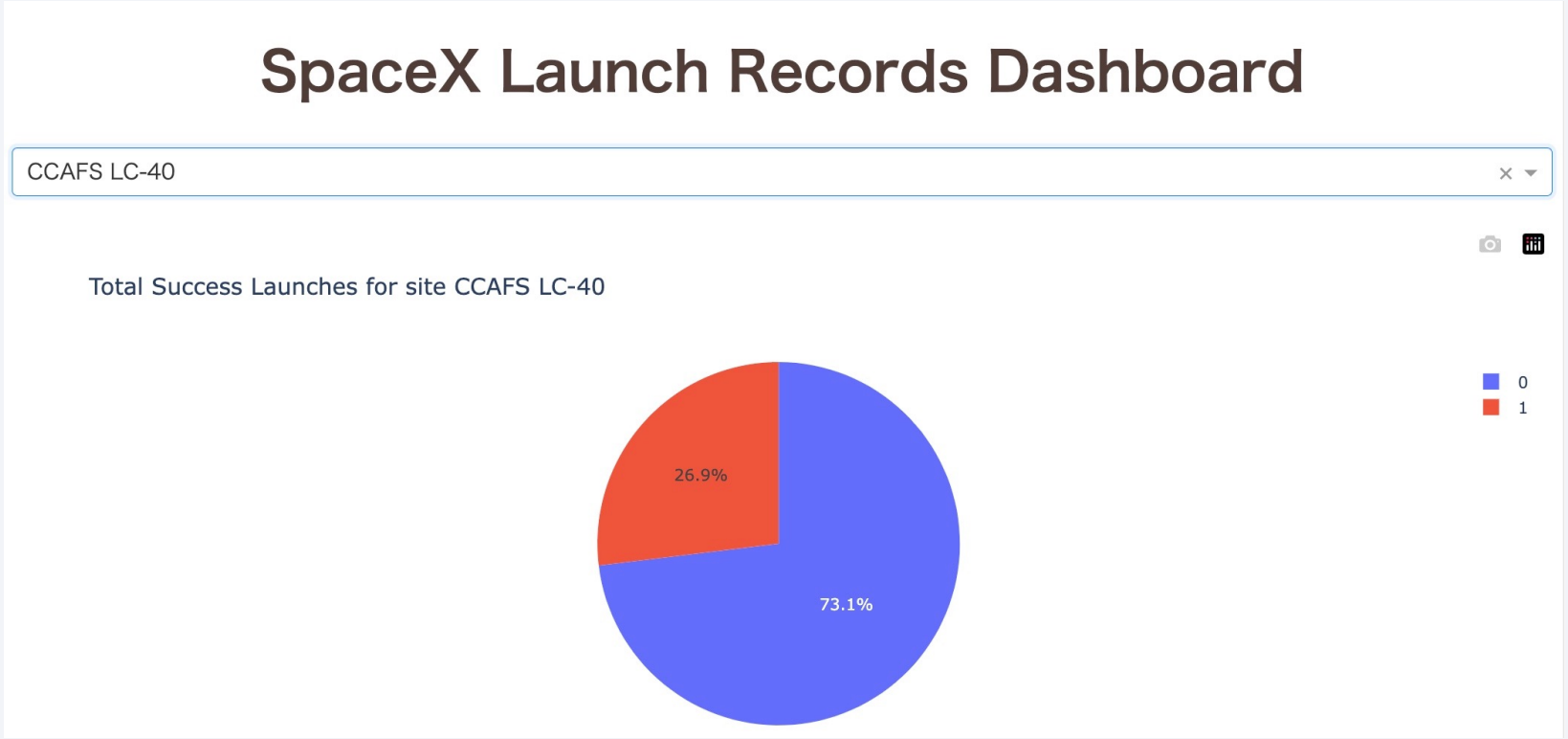
Total Success launches by site

- KSC LC-39A is the best success site, compared to the other sites



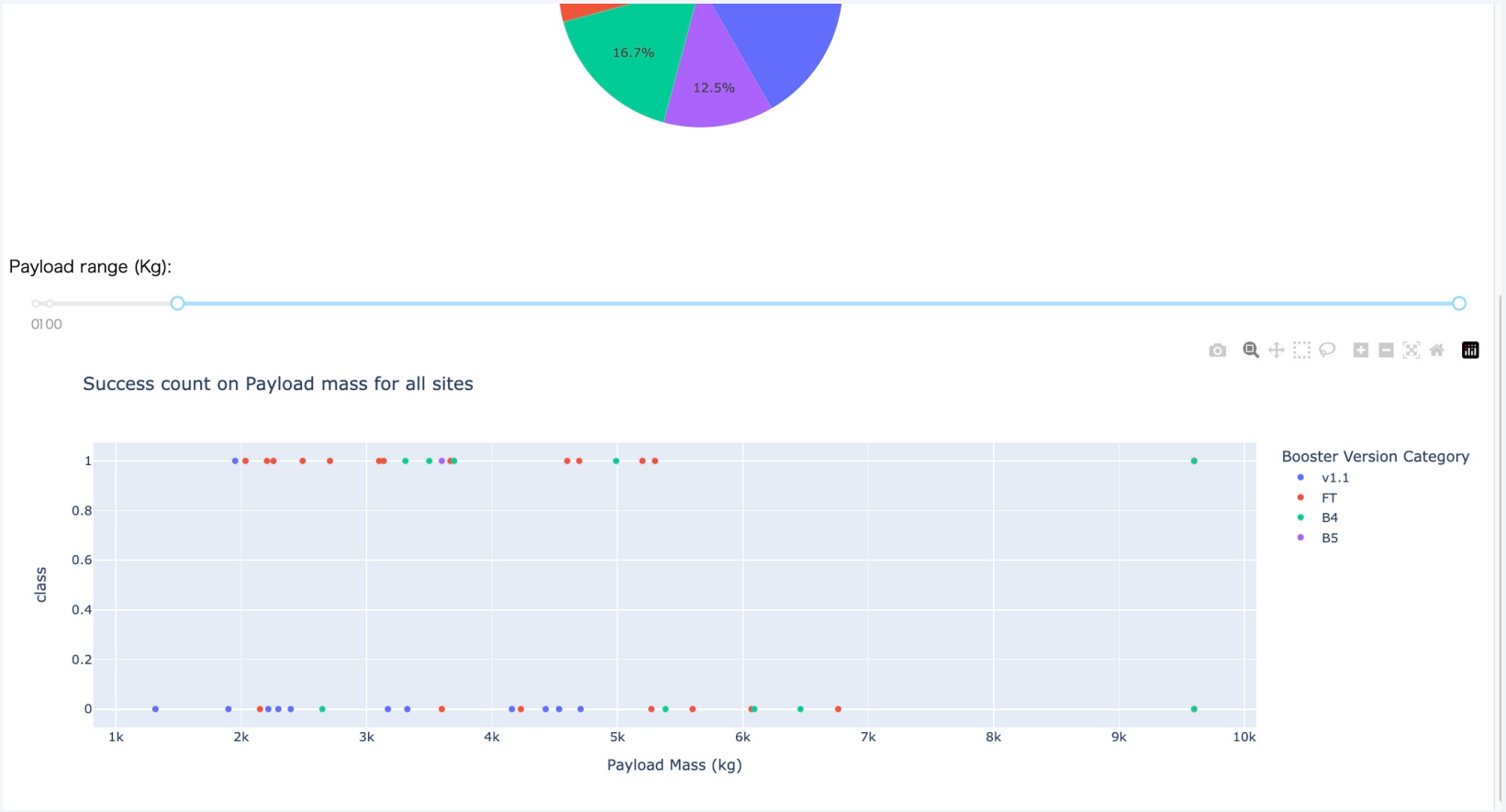
<Dashboard Screenshot 2>

- CCAFS LC-40 success rate is 73.1%



Success vs payload

- The success rate (payload mass over 2,000) is slightly higher than lower mass load



Section 5

Predictive Analysis (Classification)

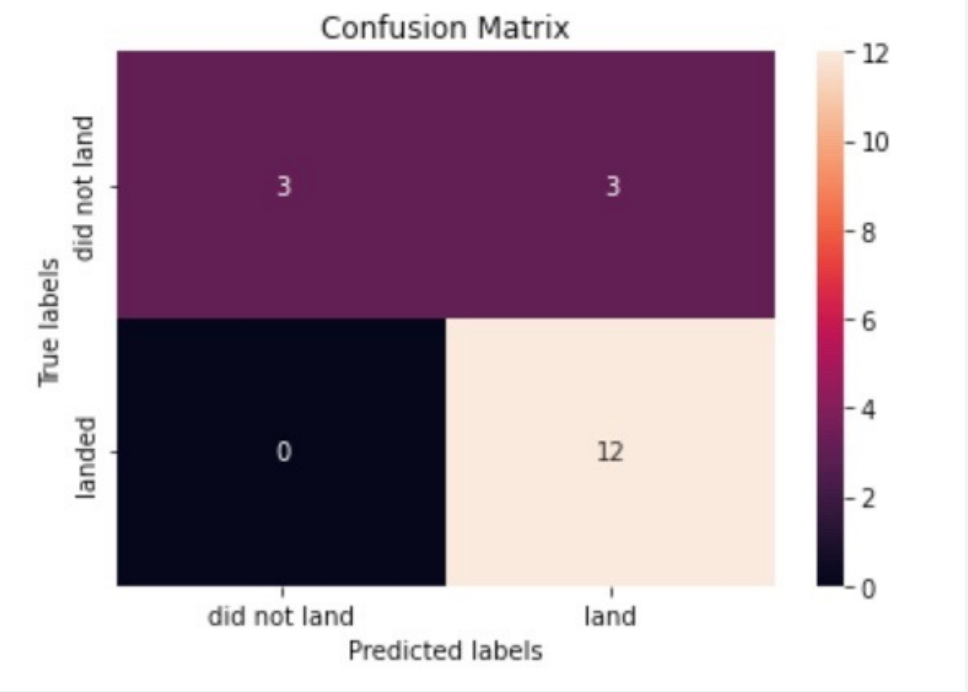
Classification Accuracy

- The best prediction model are SVM, Logistic regression and KNN model.

	accuracy
Logistic regression	0.833333
SVM	0.833333
Decision tree	0.777778
KNN	0.833333

Confusion Matrix

- Confusion Matrix of the best classification model is here



Conclusions

- We can see that there is a positive correlation between number of flights and success rate. Also the success rate has improved over the years.
- Orbits SSO, HEO, GEO, and ES-L1 have 100% success rates.
- Success rate will be improved if the payload mass is higher.
- The best prediction model are SVM, Logistic regression and KNN model.

Appendix

- Github link(main): https://github.com/watanjo/CapstoneProject/tree/master
- SpaceX wiki: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Thank you!