IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kyaw Zin Tun @ Wong Wun Kwint
2023-01-10

Outline

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

Executive Summary

- I use SpaceX open data api to get the data and perform the whole data pipeline to predicting whether the rocket re-landing will be success or not.
- The SpaceX invest in rocket re-landing pays off as the number of success over the years keep increasing which is huge blow to rival companies to play catch-up in the field.

Introduction

- This project is about SpaceX Falcon 9 first stage landing prediction.
- By Knowing whether SpaceX could land its rocket or not, our company can bid against SpaceX for a rocket launch.
- Because if SpaceX could land its rocket, it can cut its rocket price to 62 million dollars per launch which is far cheaper than average 165 million dollars of rival companies.

Section 1

Methodology

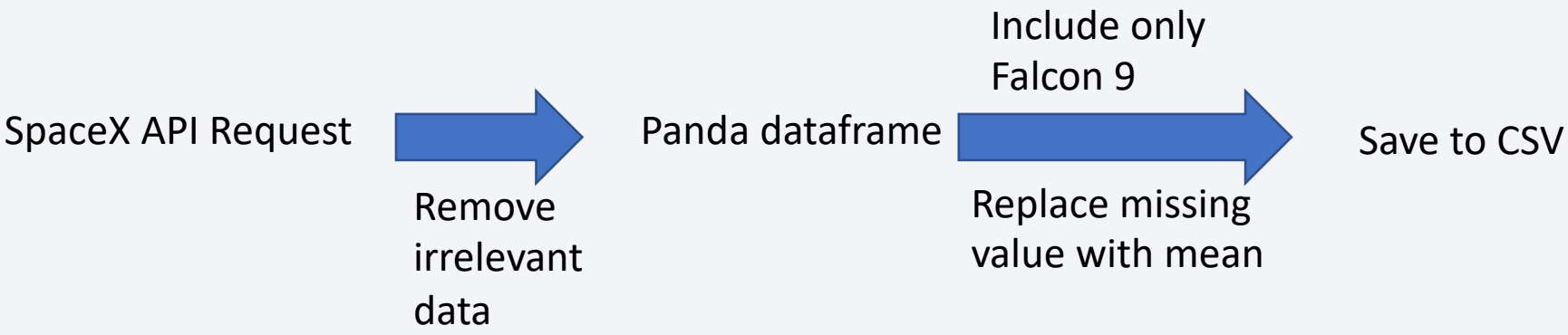
Methodology

Executive Summary

- Data collection methodology:
  - Request the SpaceX launch data with free REST api from SpaceX
- Perform data wrangling
  - Process the data using pandas and numpy
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Train, evaluate different classification models to find the best model for prediction

Data Collection

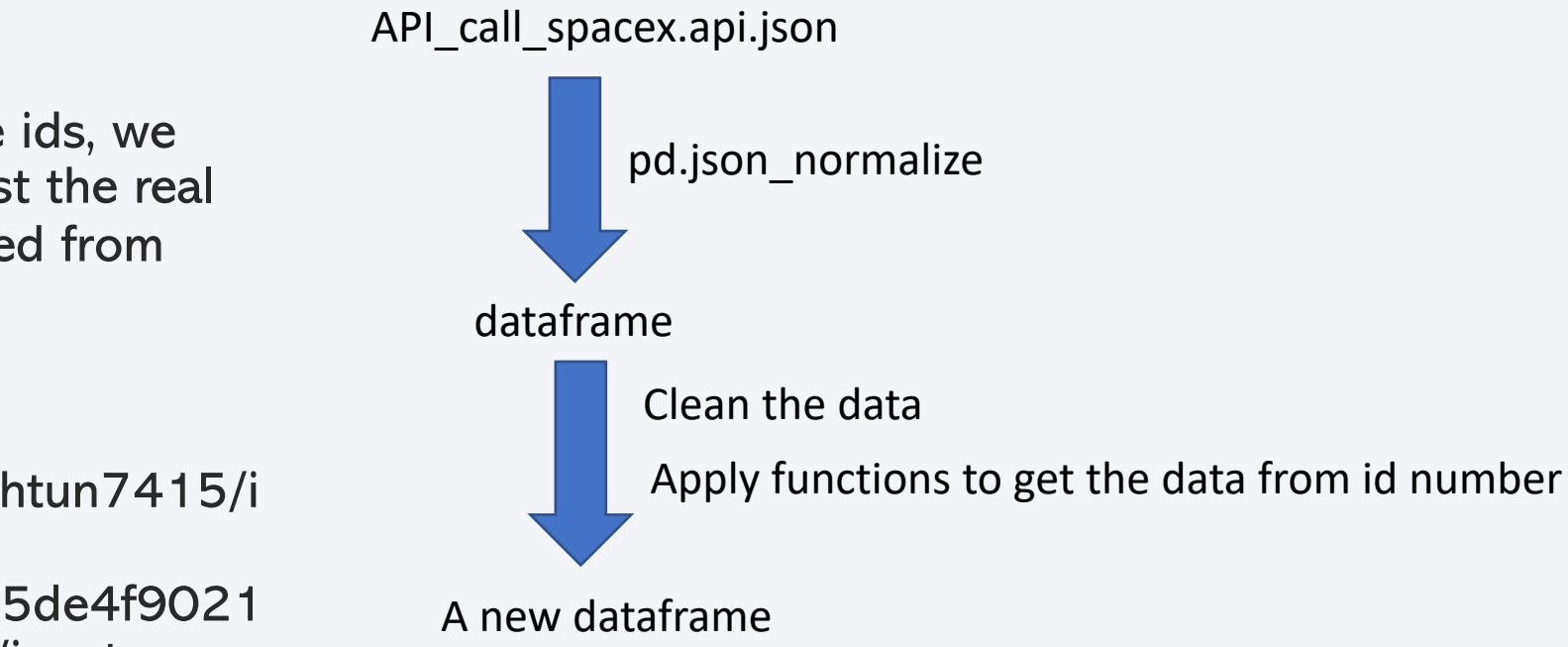
- First, I made a request to SpaceX API and extract relevant data.
- Then, I transformed data to panda dataframe, included only Falcon 9 launches and replaced the missing values with mean value.
- Finally, I saved the data into CSV file.



Data Collection – SpaceX API

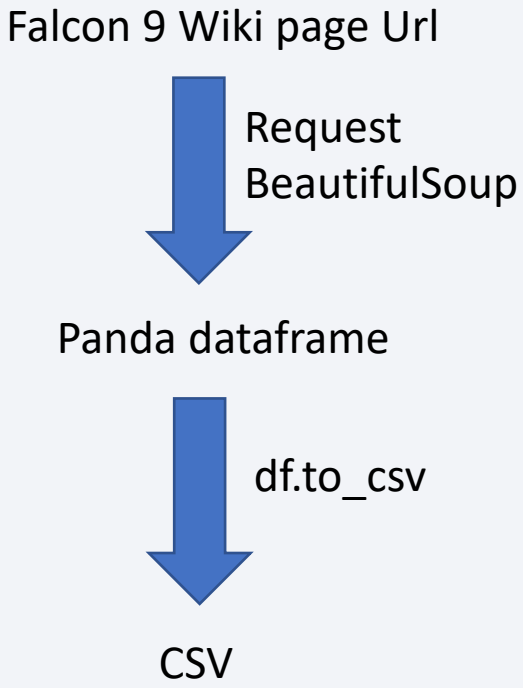
- On the left is my SpaceX API Data Collection Flow.
- Since what we get back are ids, we had to subsequently request the real information with ids returned from above step.

- https://github.com/kyawzinhtun7415/ibm-data-science/blob/98c32f7ec585de4f90218b392cef318b2616b457/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- On the left is the flowchart of Falcon 9 launches records webscraped from Wikipedia
- https://github.com/kyawzinhtun7415/ibm-data-science/blob/98c32f7ec585de4f90218b392cef318b2616b457/jupyter-labs-webscraping.ipynb



Data Wrangling

- First, I did some exploratory data analysis to have an overview of data such as the number of rows and columns, or missing values, etc.
- After that since we have to predict outcomes using machine learning, I standardize the 8 landing outcomes to 0 (failure) and 1 (success) eventually.
- https://github.com/kyawzinhtun7415/ibm-data-science/blob/98c32f7ec585de4f90218b392cef318b2616b457/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- I used the seaborn and matplotlib package of Python to visualize the charts.
- I used the scatter plot to see the relationship of two variables, the bar chart to compare the number of variables and the line chart to see the relationship of the variable over time.
- https://github.com/kyawzinhtun7415/ibm-data-science/blob/98c32f7ec585de4f90218b392cef318b2616b457/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- I used SQL language to further perform my exploratory data analysis.
- I find the average, count of payload_mass by launch site, booster version , etc.
- I mostly use summary functions, groupby and where clauses to further specificities
- https://github.com/kyawzinhtun7415/ibm-data-science/blob/98c32f7ec585de4f90218b392cef318b2616b457/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

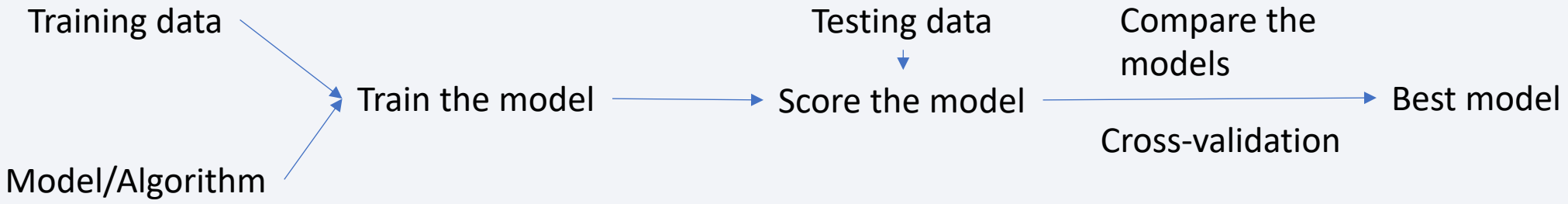
- I built an interactive map to see the launch sites on the map and to find relationship between launching site and outcome of the launch.
- I used the Folium library of Python to build an interactive map.
- Markers and circles are to mark the coordinates of the launch site and have pop-up, color parameter, etc.
- The line function is mainly used to show the distance between the launch site and coastline, railway track, city ,etc to see their relationship on success of launches
- https://github.com/kyawzinhtun7415/ibm-data-science/blob/98c32f7ec585de4f90218b392cef318b2616b457/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- I used Plotly Dash using Python and used pie chart and scatter plot of the library.
- Users can interact with the dashboard to see the success all launch sites or specific launch site.
- I used the pie chart to analyze the proportion of each launch site in terms of success and the scatter plot to analyze relationship between 3 variables.
- https://github.com/kyawzinhtun7415/ibm-data-science/blob/98c32f7ec585de4f90218b392cef318b2616b457/spacex_dash_app.py

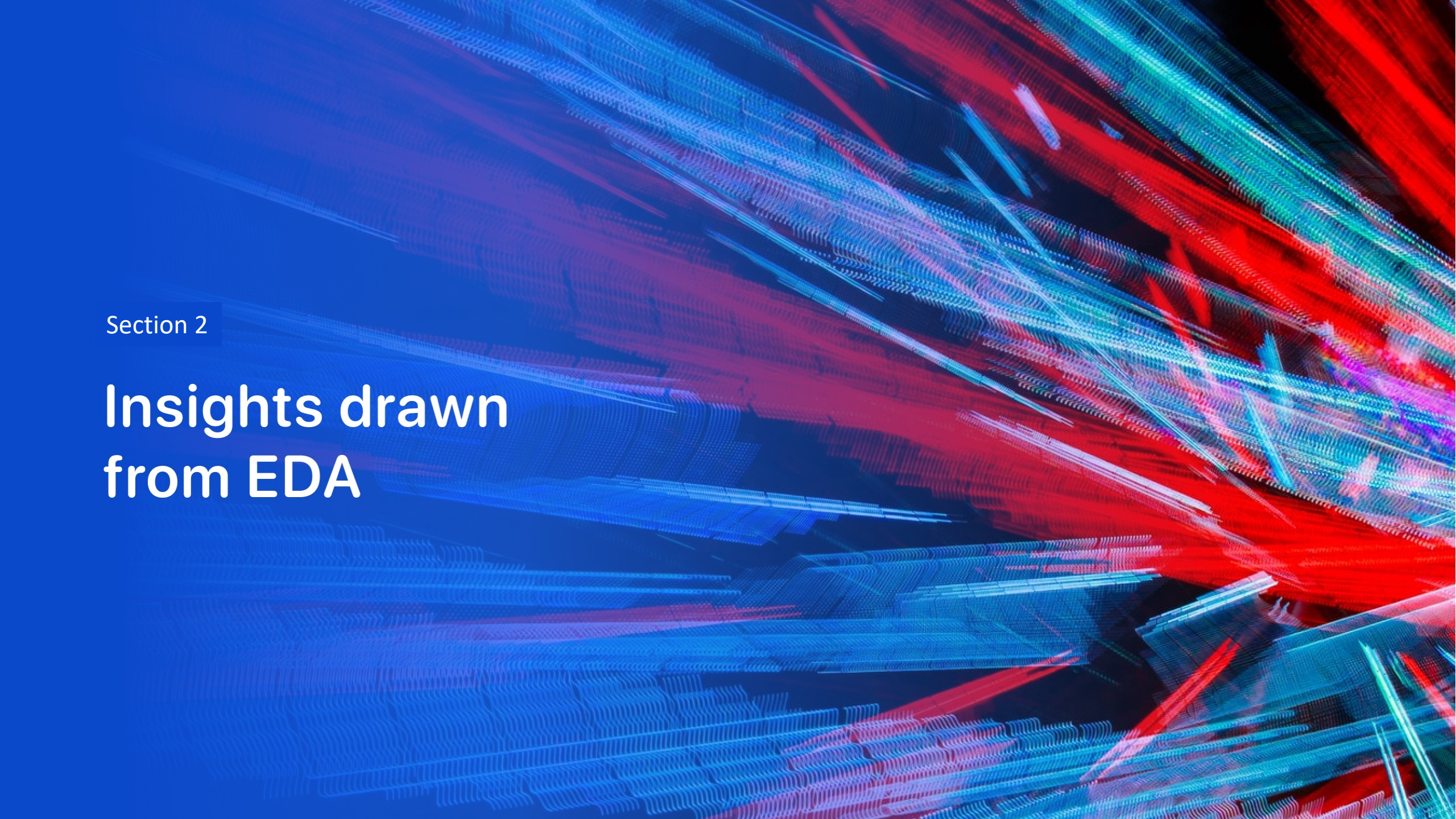
Predictive Analysis (Classification)

- I tested 4 classification models (KNN, logistic regression, SVM, decision tree)
- I used the same parameters to get the same result out of 4 models.
- Out of 4 models, 3 models (KNN, logistic regression, SVM) have approximately the same training and test score so I applied cross-validation and domain-knowledge to choose the simpler model (logistic regression).
- https://github.com/kyawzinhtun7415/ibm-data-science/blob/98c32f7ec585de4f90218b392cef318b2616b457/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

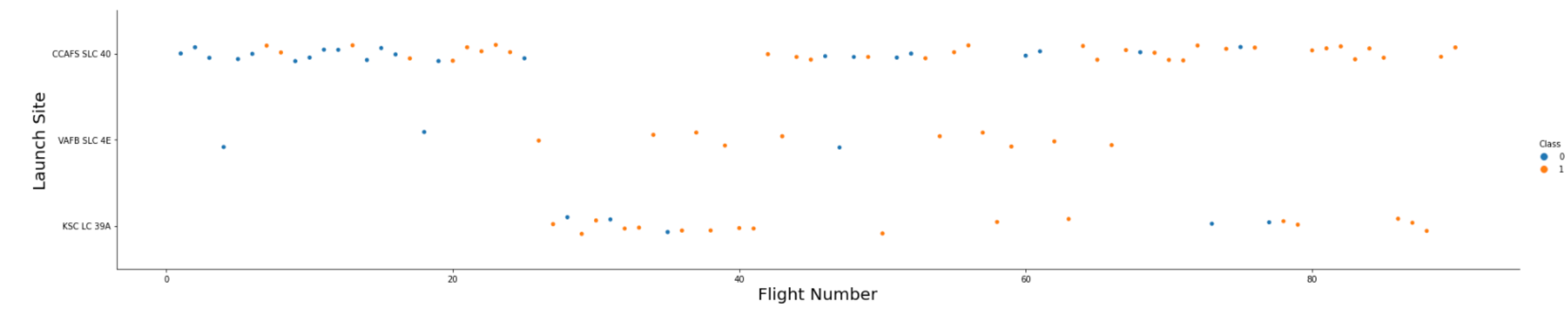


Results

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

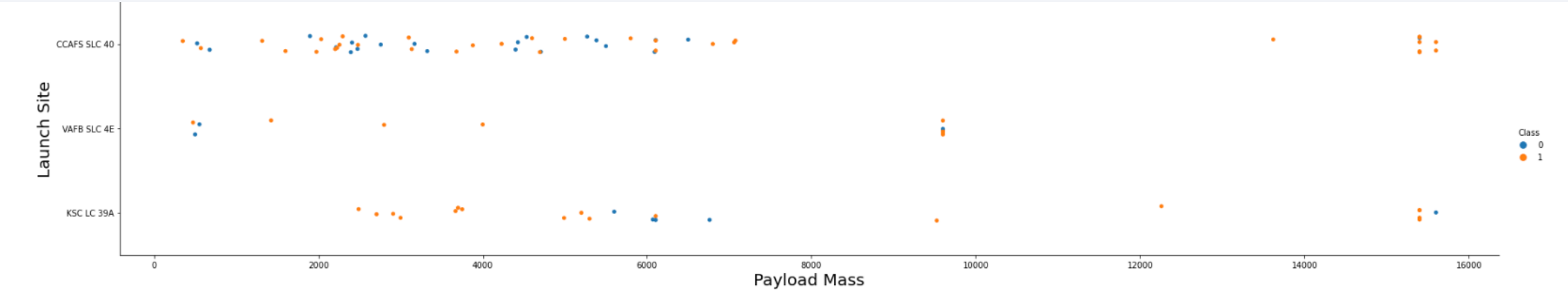

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



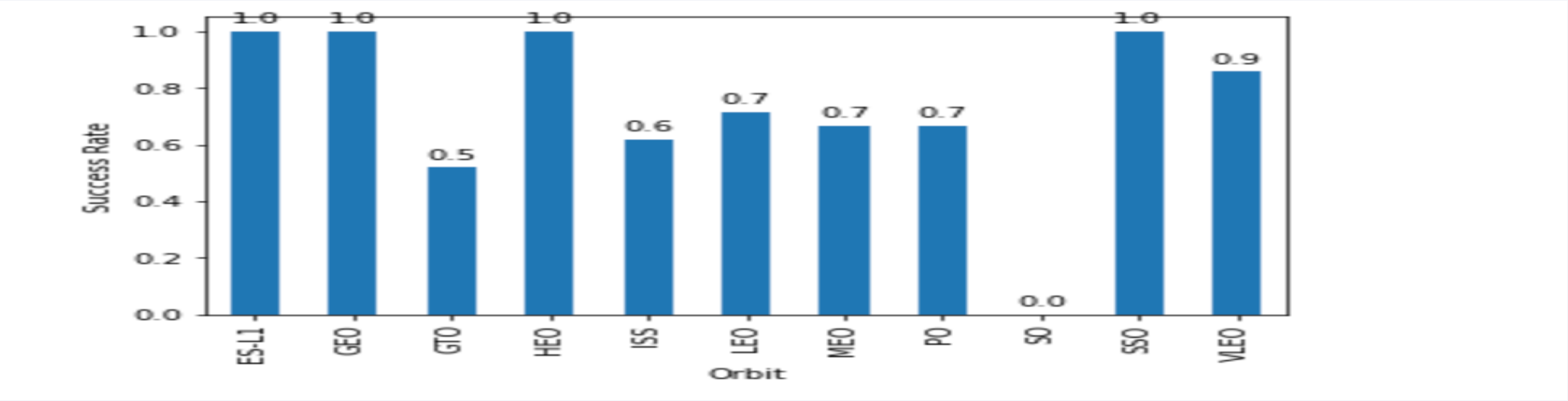
- The plot shows the flight number, launch site and the outcome of the launch in one visualization.
- We can see that the launch site 'CCAFS SLC 40' is most used and approximately from flight number 80, the SpaceX launches have not failed once.

Payload vs. Launch Site



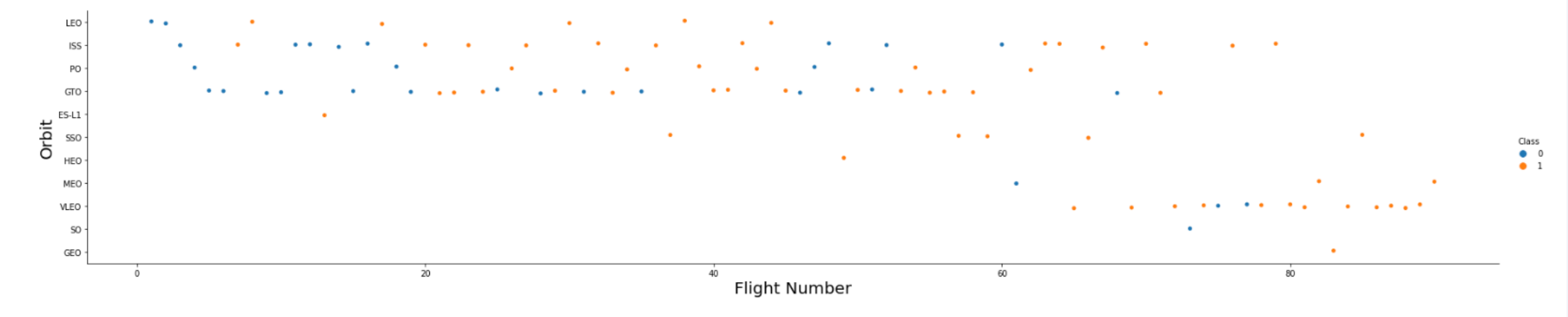
- This plot shows the relationship between the payload and the launch site.
- As we can see, there are no more payload mass heavier than 10,000 on launch site ‘VAFB SLC 4E’.
- The plot cannot clearly describe a linear relationship between payload and launch site.

Success Rate vs. Orbit Type



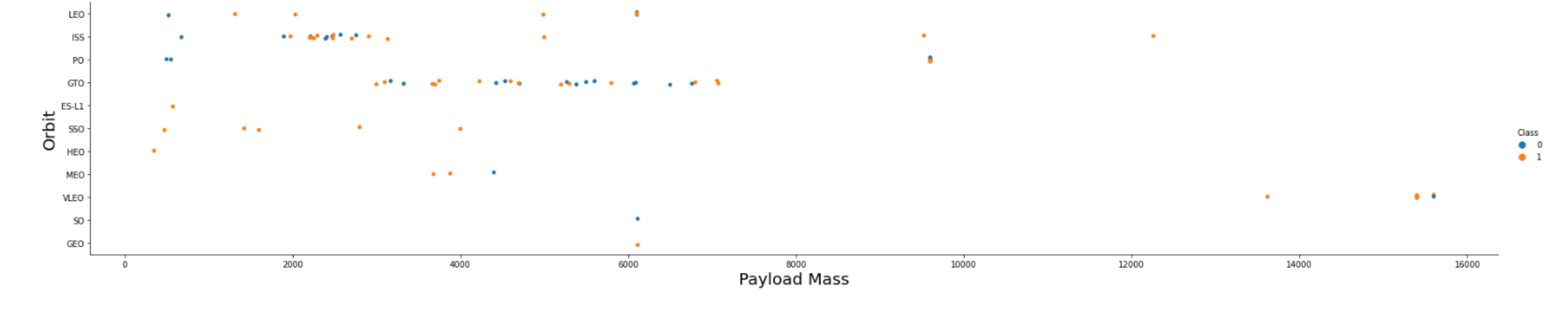
- This bar chart shows the various successful rocket launches by orbit type.
- Both LEO and MEO have moderate success rate whereas HEO and GEO have 100% success rate.

Flight Number vs. Orbit Type



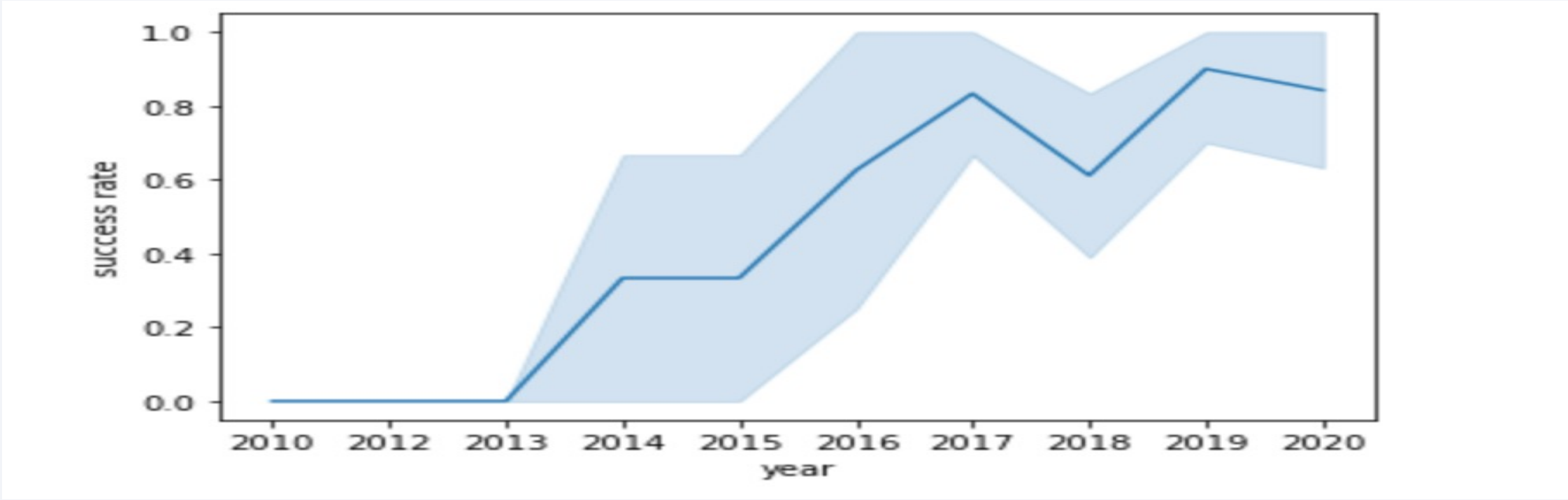
- The above scatter plot describes the relation between orbit type and flight number.
- The later flight number approximately above 80 all encountered successes on all types of orbit types which mean that SpaceX rocket launch re-landing experiments have paid off.

Payload vs. Orbit Type



- The above diagram shows a scatter point of payload vs. orbit type.
- We can see that the rocket launches using orbit SSO are successful 100%.
- Other rocket launches to different types of orbit all have their own failures and successes.

Launch Success Yearly Trend



- The above shows a line chart of yearly average success rate
- The success rate steadily increases as the year progresses except only a small dip in success rate in 2018.

All Launch Site Names

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

- The SQL query on the left displays the names of the unique launch sites.
- The 4 locations are unique launch sites in the space mission.

Launch Site Names 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

- The above shows 5 records where launch sites begin with `CCA`
- We can see all the columns associated with launch site names that begin with ‘CCA’.

Total Payload Mass

SUM(PAYLOAD_MASS__KG_)
45596

- The above shows the total payload carried by boosters from NASA

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

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

First Successful Ground Landing Date

1
2015-12-22

- The query shows the dates of the first successful landing outcome on ground pad

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

Total Number of Successful and Failure Mission Outcomes

successful_missions	failed_missions
100	1

- The above shows the total number of successful and failure mission outcomes
- The number of successful missions is 100 whereas only 1 mission is considered as fail.

Boosters Carried Maximum Payload

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

- These are the names of the booster which have carried the maximum payload mass
- We can see that the maximum payload is carried by Falcon 9 B5 version only.

2015 Launch Records

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

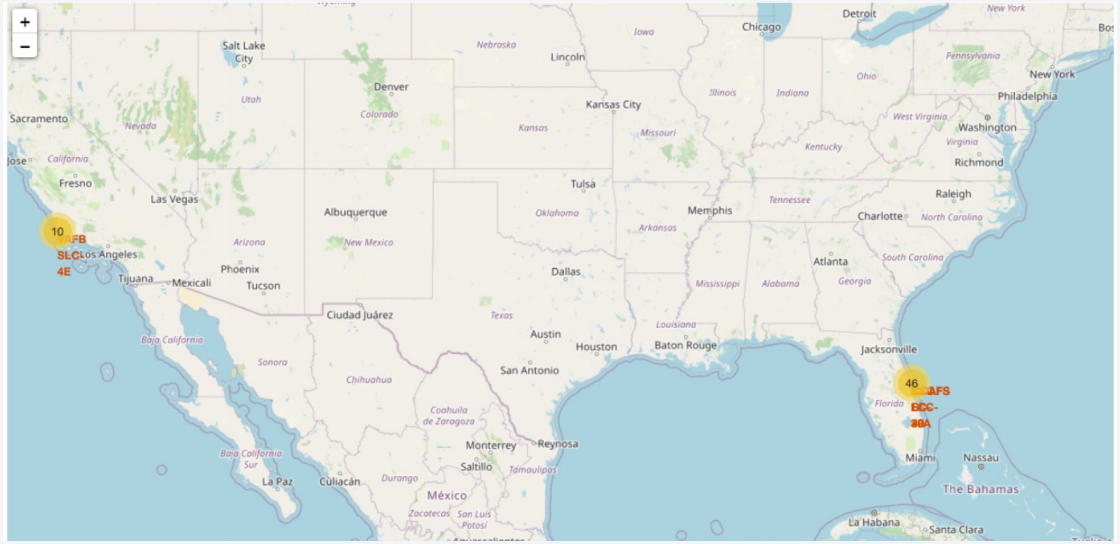
- The above shows the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

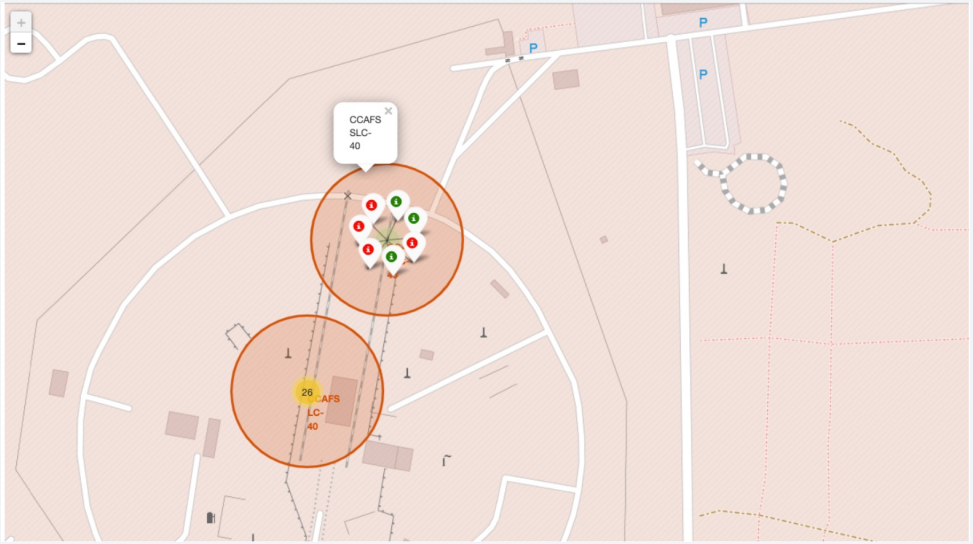

Section 3
Launch Sites
Proximities Analysis

SpaceX Launch Sites

- The launch sites can be divided into two places, east and west coast.

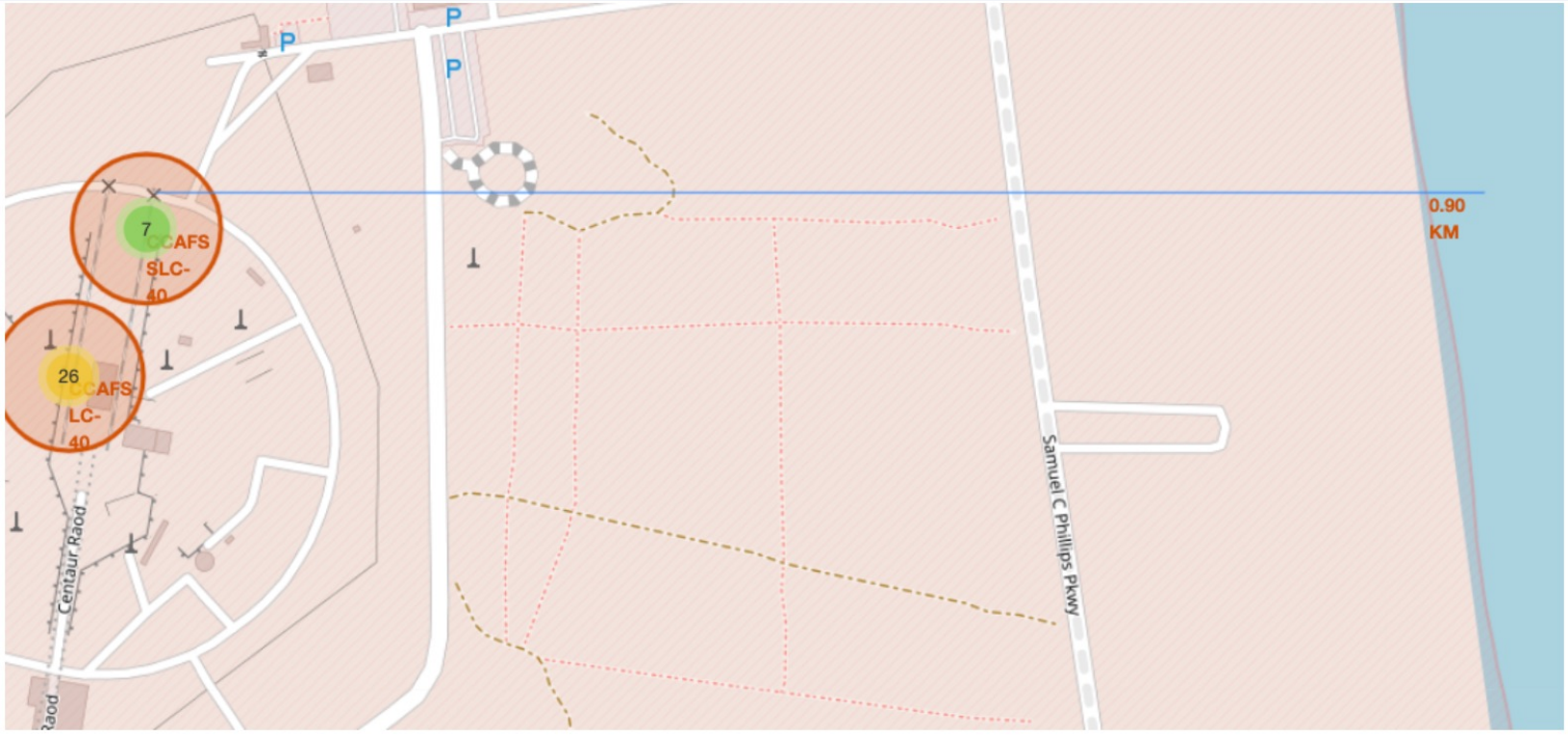
Success/failed launches for each site on the map



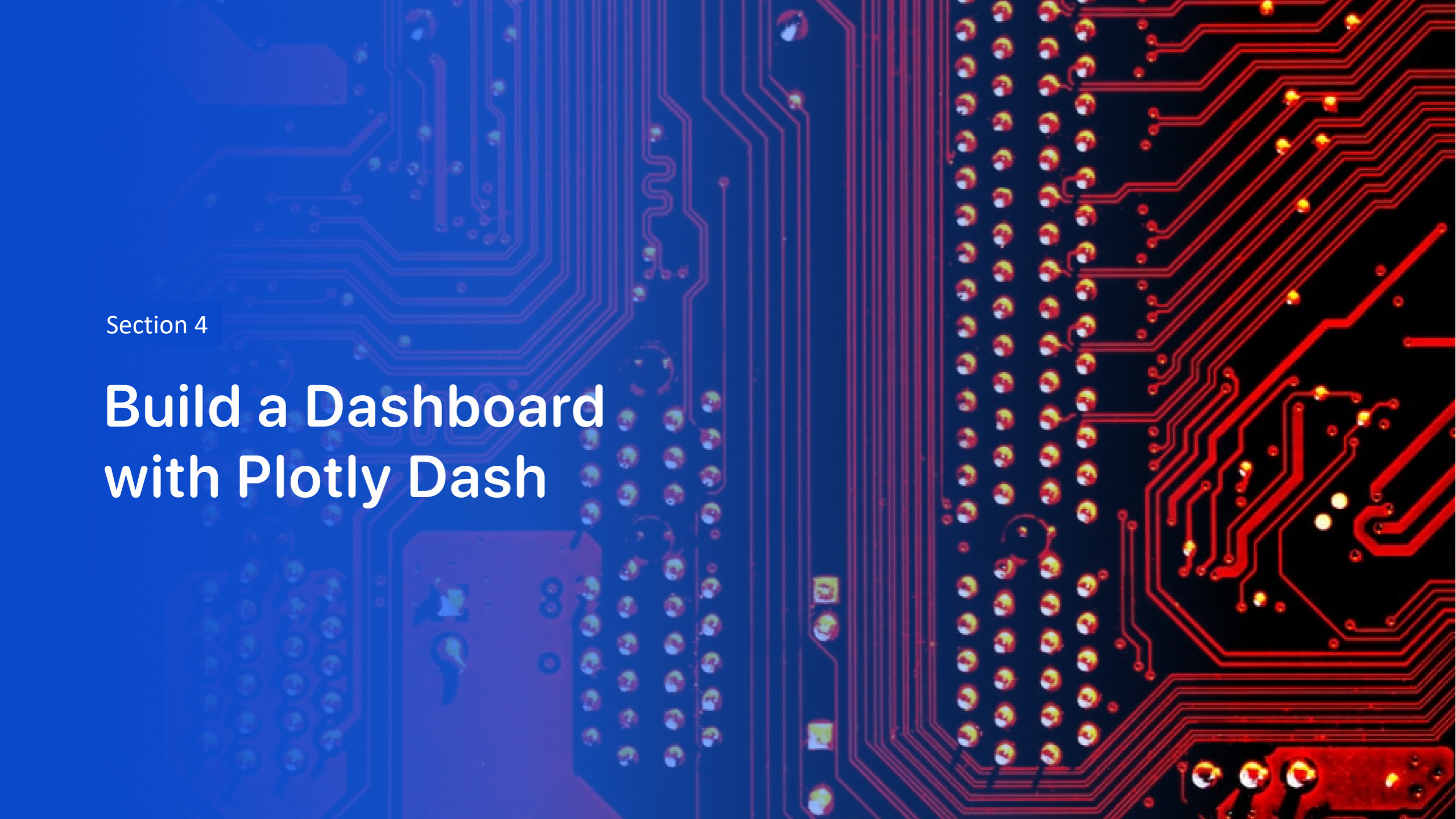


- We can zoom out to see the number of sites and respective names.
- If we zoom in, we can see the green and red marks , each representing success and failed launches on the map.

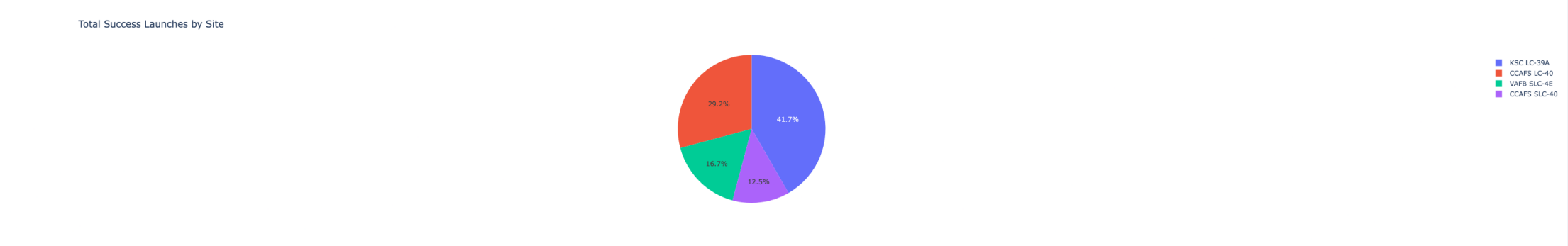
Distances between a launch site to the coastline



- We can use the same type to measure distances to city, railway ,etc to see if the proximity to these areas has effect on mission success or failure.

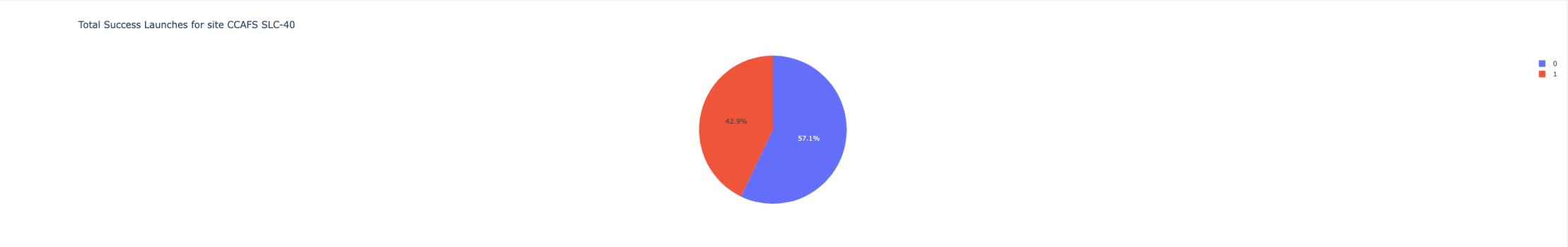

Section 4
Build a Dashboard with Plotly Dash

Launch success count for all sites



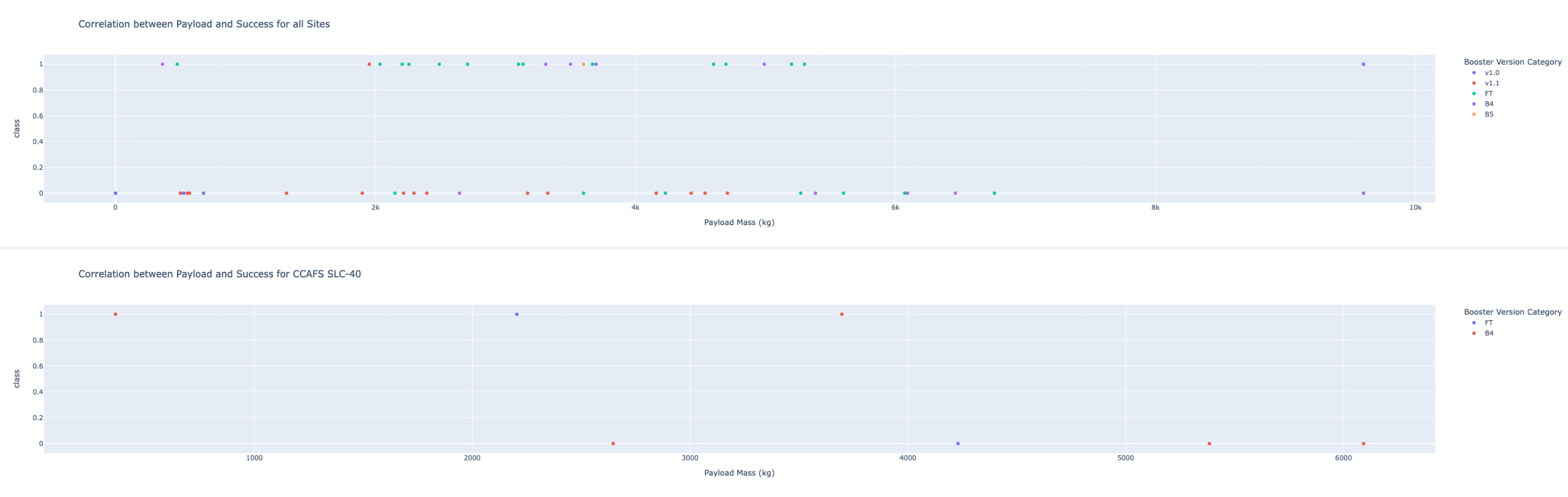
- The launch from KSC LC-39A makes up about 40% of the launch success.
- The runner-up is CCAFS LC-40 with 30% and the rest of the two are around 15%.

Launch site with highest launch success ratio



- The Launch site CCAFS SLC-40 has the highest launch success ratio of 43%.

<Dashboard Screenshot 3>

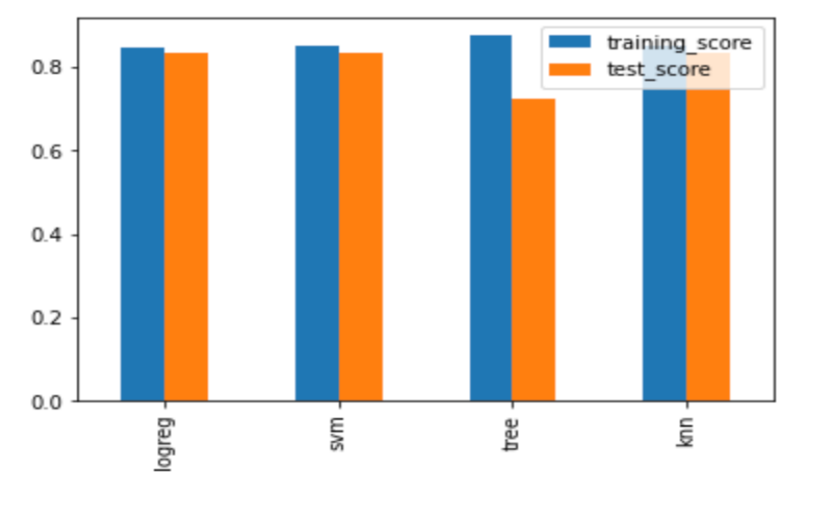


- The above plot shows the correlation between payload and success for all sites and we can observe that FT booster version has the most number of successes.
- Below is an example of the same correlation but for one site only.

Section 5

Predictive Analysis (Classification)

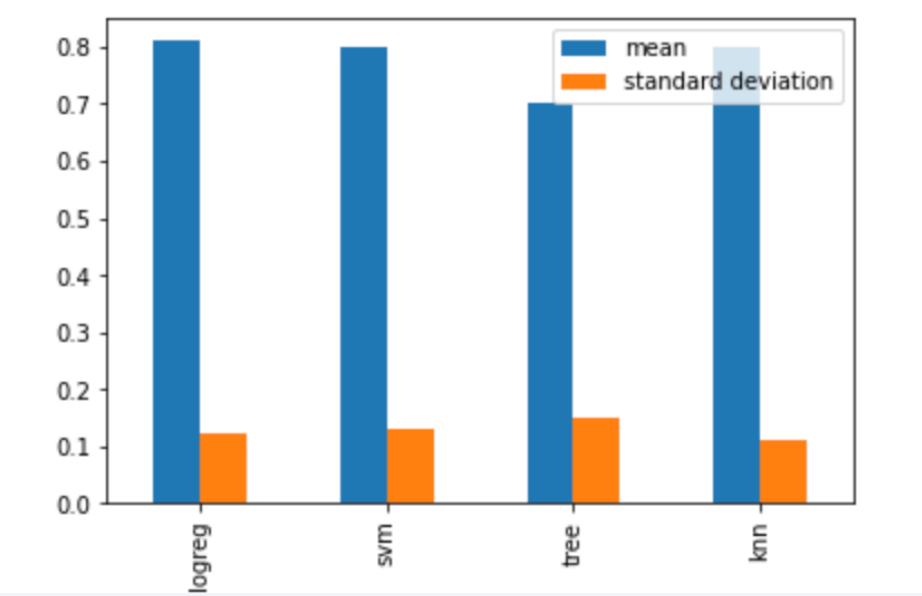
Classification Accuracy

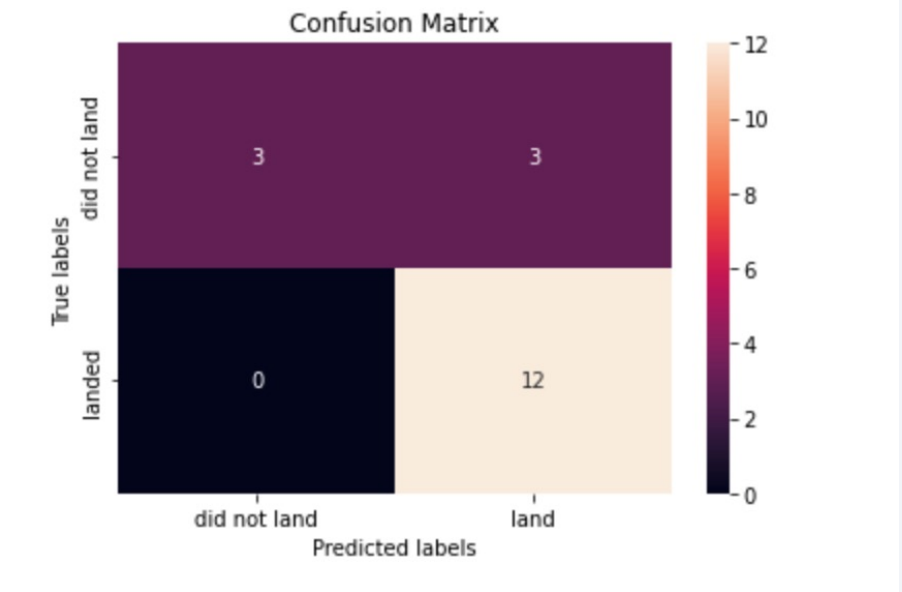


	training_score	test_score
logreg	0.846429	0.833333
svm	0.848214	0.833333
tree	0.875000	0.722222
knn	0.848214	0.833333

- As we can see above, it is difficult to choose between one of the three models (logistic regression, support vector machine, k-nearest neighbors) as they have the same classification accuracy.

Confusion Matrix





- The above is my 10-fold cross-validation result. The blue bar is the mean of the scores and the orange one is the standard deviation of the scores.
- Therefore, after considering mean, standard deviation, complexity, I chose logistic regression as this classification can be solved by logistic regression which contains only 0 and 1.

Conclusions

- SpaceX invest in rocket relanding experiment pays off in the end.
- We can use logistic regression to predict whether SpaceX will land their rocket or not.
- SpaceX can use CCAFS SLC-40 launch site to increase their launch success.

Appendix

Thank you!