IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Zijie Mei
01/17/2024

Outline

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

Executive Summary

For this project, I used API and webscraping to collect data, then cleaned the data. After that, I did exploratory data analysis with both SQL and visualization. Furthermore, I created interactive dashboard with folium and dash. In the end, develped machine learning models to classify outcomes of different launches across the years.

Key Findings:

- I was able to find a good Decision Tree model with 84% in samle accuracy and 83% test accuracy that can effectively predict the launch outcomes.
- I found out that successful rate is going up over the years.

Introduction

Background:

Space X 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 Space X 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 space X for a rocket launch.Problems you want to find answers

Objective:

Provide valueble informations to my company SpaceY, which is to build machine learning models and use public informations to predict if SpaceX will reuse the first stage in order to determine the price of each launch.

Develop dashboards that provides informations about SpaceX that can used by my team.

Section 1

Methodology

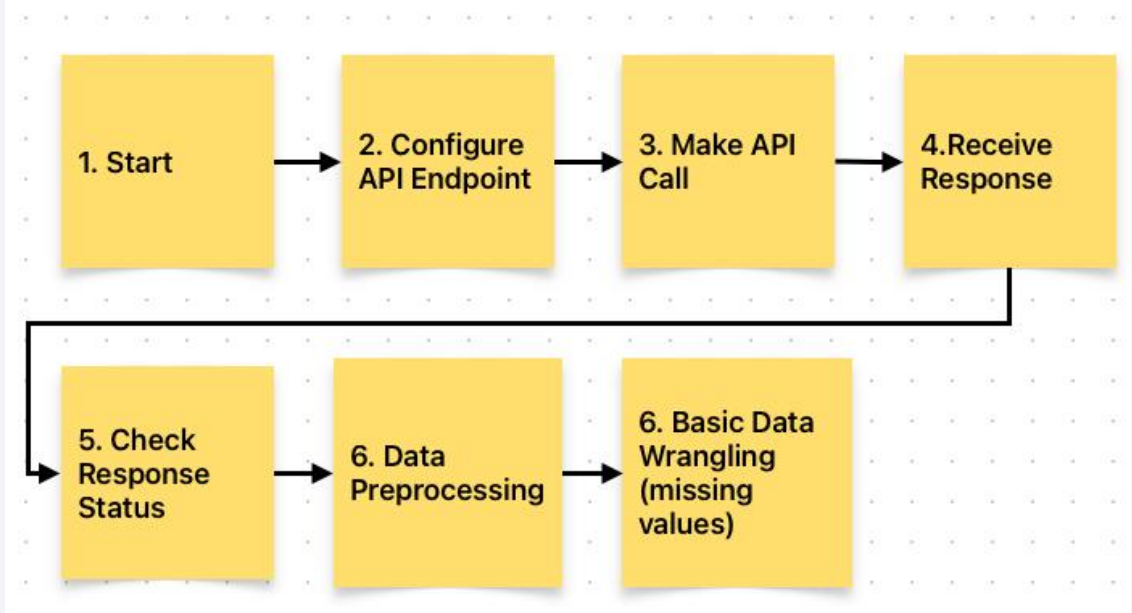
Methodology

Executive Summary

- Data collection methodology:
  - Data was collected by using SpaceX API and webscraping data from Wikipedia table
- Perform data wrangling
  - Some innitial exploration of the data and identified missing values as well as creating label column
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Trained various models through cross validation to find the best parameters

Data Collection – SpaceX API

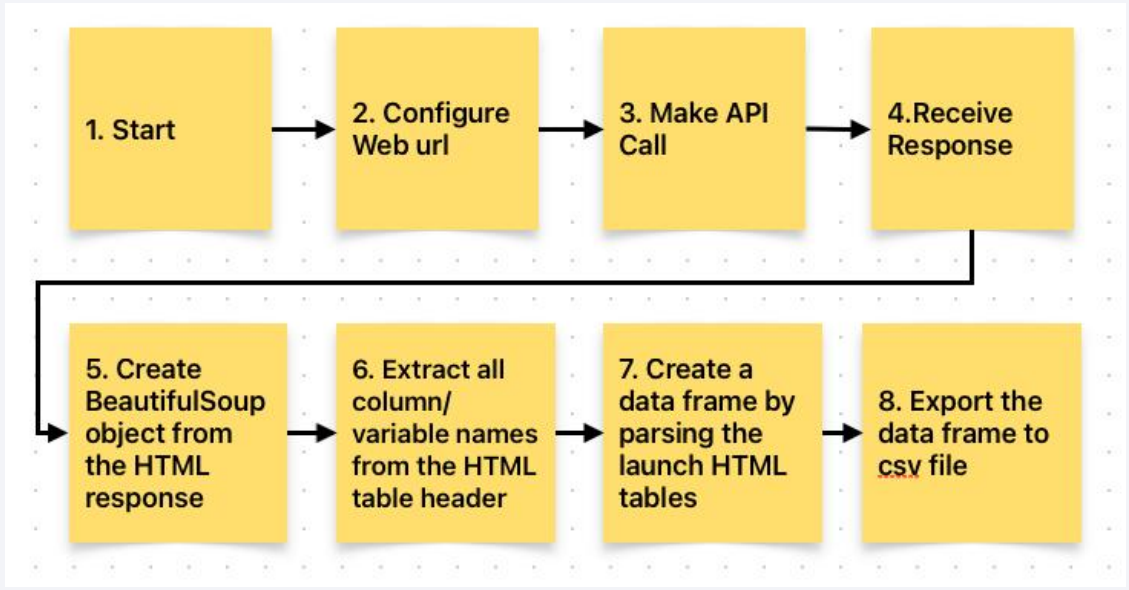
- The picture on the right shows the flowcharts of the API data collection, and below is a hyper link to the github code.
- GitHub Link



Data Collection - Scraping

- The picture on the right shows the flowcharts of the Scraping of the data collection stage, and below is a hyper link to the github code.

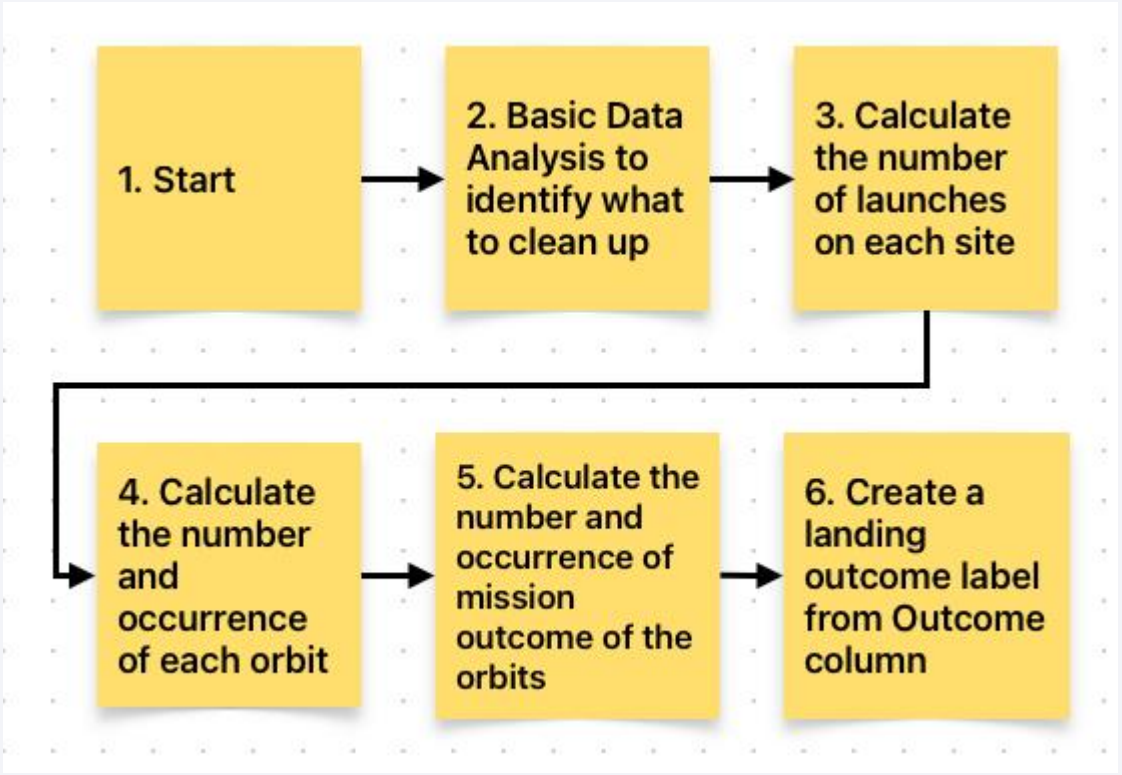
- GitHub Link



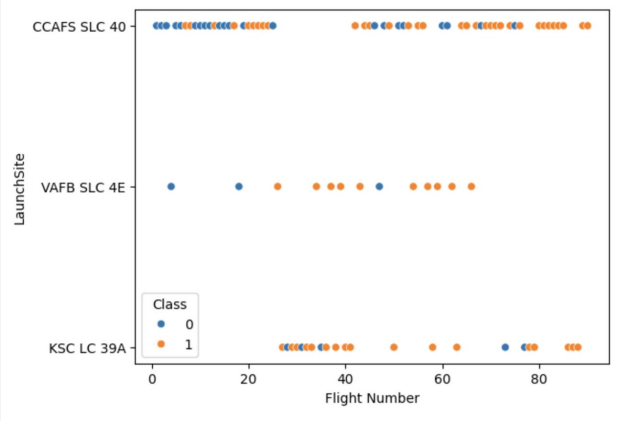
Data Wrangling

- The picture on the right shows the flowcharts of the Scraping of the data collection stage, and below is a hyper link to the github code.

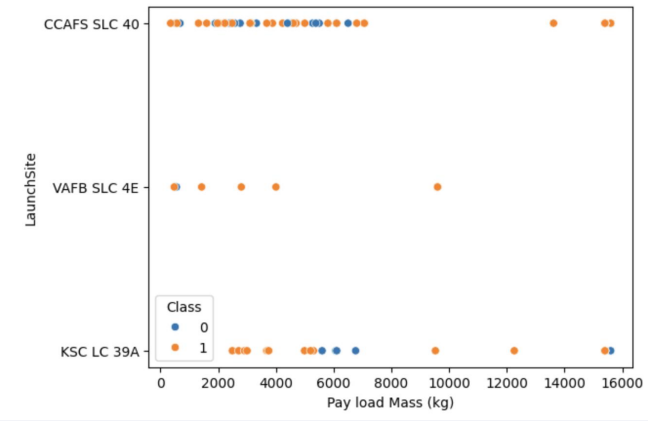
- GitHub Link



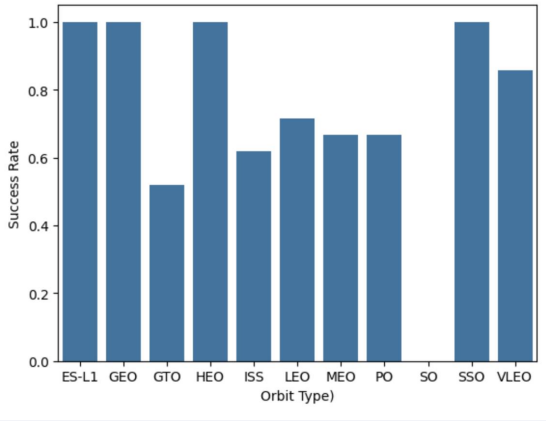
EDA with Data Visualization



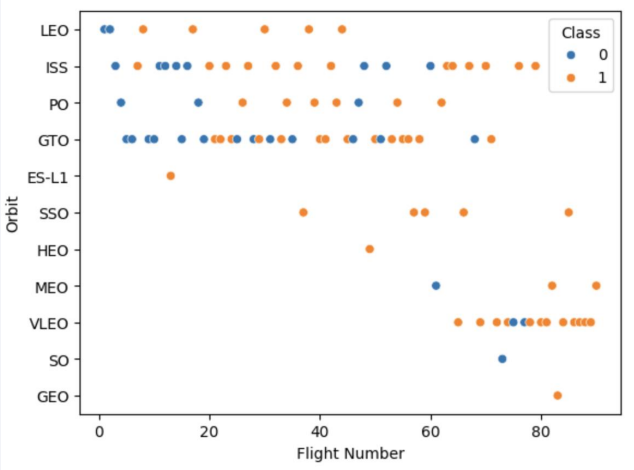
VAFB SLC 4E has the least flight numbers associated with it.



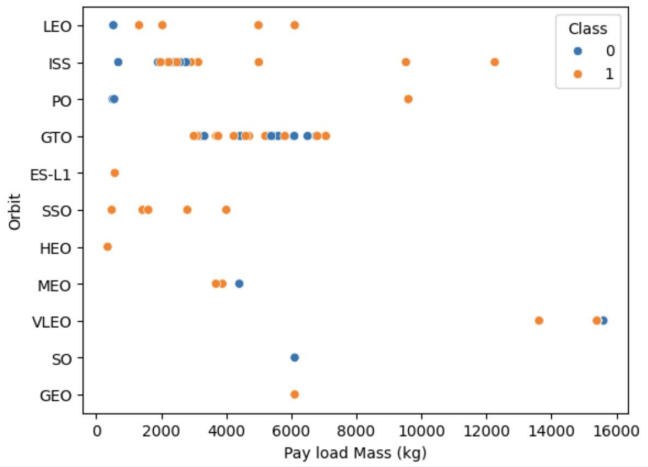
VAFB SLC 4E appears to alwasy have low payload mass.



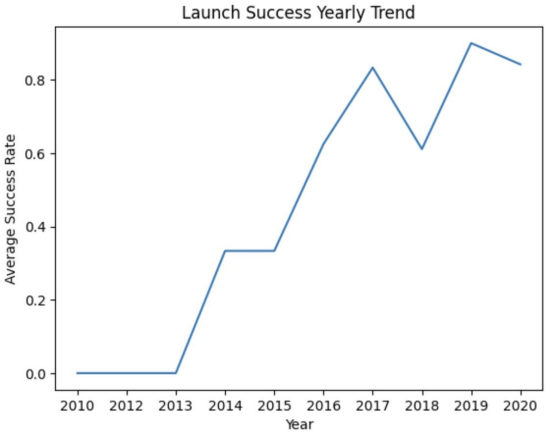
SO and GTO has the least success rate.



ES-L1, GEO and SO only has one Flight Number associated with it.



GTO has a lot of launches with low payload mass



Average success rate has gone up over time.

GitHub Link

EDA with SQL

- Found unique launch sites in the space mission.
- Found 5 records where launch sites begin with the string ‘CCA’
- Total payload mass by NASA(CRS): 45596kg
- Average payload mass carried by booster version F9 v1.1: 2534.7 kg
- First succesful landing outcome in ground pad was in 2015–12–22
- Found the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Found the total number of successful and falure mission outcomes: 71
- Found all the names of the booster versions which have carried the maximum payload mass.
- Found 2 failure landing_outcomes in drone ships in 2015.
- Ranked the count of landing outcomes between the date 2010–06–04 and 2017–03–20
- GitHub Link

Build an Interactive Map with Folium

- Added all launch sites on the map to show where they are.
- Added launch count for each launch site.
- Color coded all successful or failure landings with each launch.
- Made a line from CCAFS LC–40 launch site to the nearest coast line and calculated and marked the distance.
- GitHub Link

Build a Dashboard with Plotly Dash

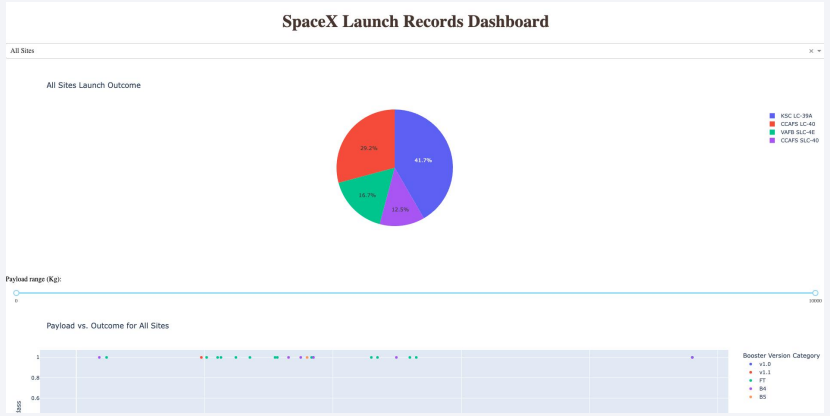
- I only added one dropdown list where options can be picked out of ALL or every launch site.
- Added a pie chart to show the total successful launches count for all sites.
- The pie chart can also show success vs. failed counts for the selected launch site.
- Added a slider to select payload range.
- Added a scatter chart to show the correlation between payload and launch success based on the dropdown pick and the slider range
- GitHub Link

Predictive Analysis (Classification)

- I prepared the data, standardized the features, and splitted data into 80% training and 20% validation, then used cross validation to find the best parameter for each model and eventually compared validation accuracy and confusion matrix for each model to get the best performing model.
- The best perfoming model is decision tree because it has the best in sample in sample accuracy with the same out of sample accuracy of 83.3%.
- It is interesting to notice that all models performed the same in terms of out of sample accuracy which is odd.
- GitHub Link

Results

- Exploratory data analysis results
  - Launch success rate is going up over the years and peaked in 2019.
  - First successful landing happened in Dec 2015
- Interactive analytics demo in screenshots
  - Screenshots demo on the right.
- Predictive analysis results
  - The best perfoming model is decision tree because it has the best confusion matrix since true positive and true nagative is balanced with the best performing out of sample accuracy of 83.3%

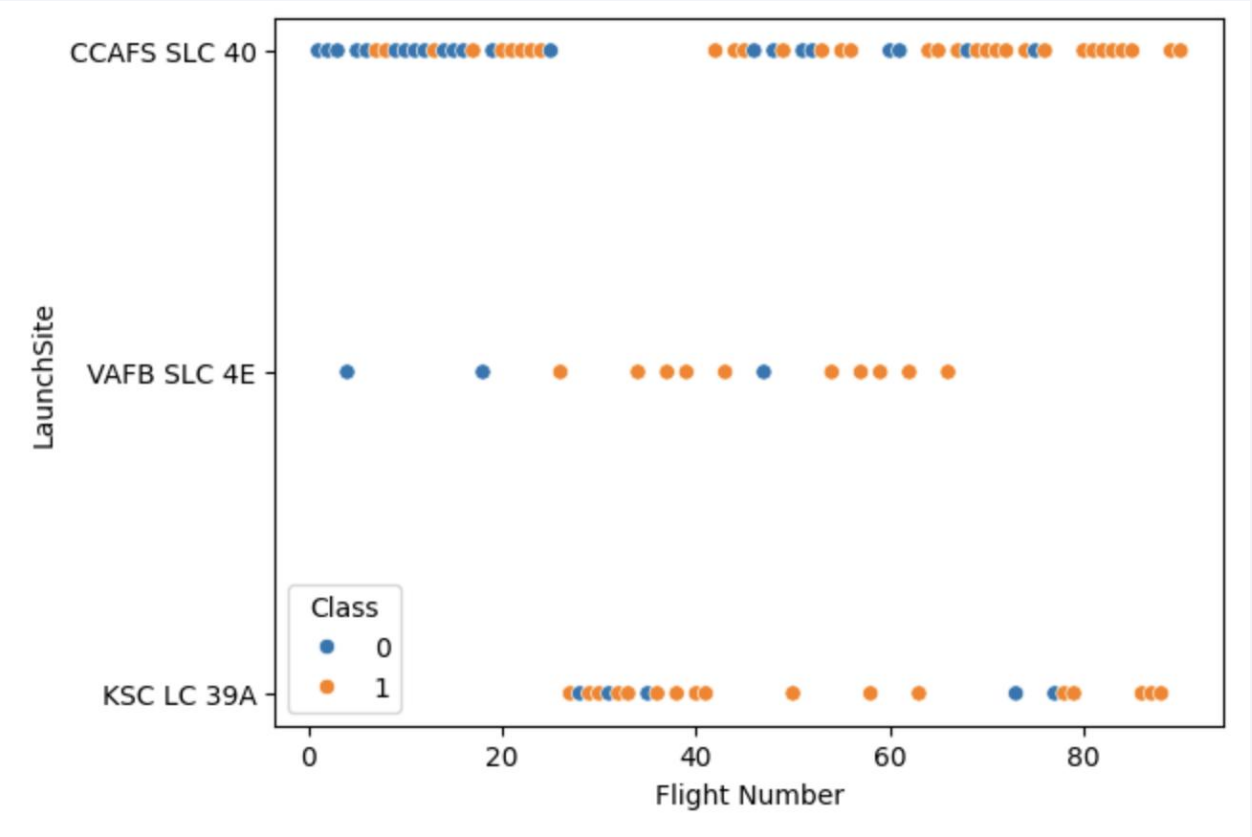


Section 2

Insights drawn from EDA

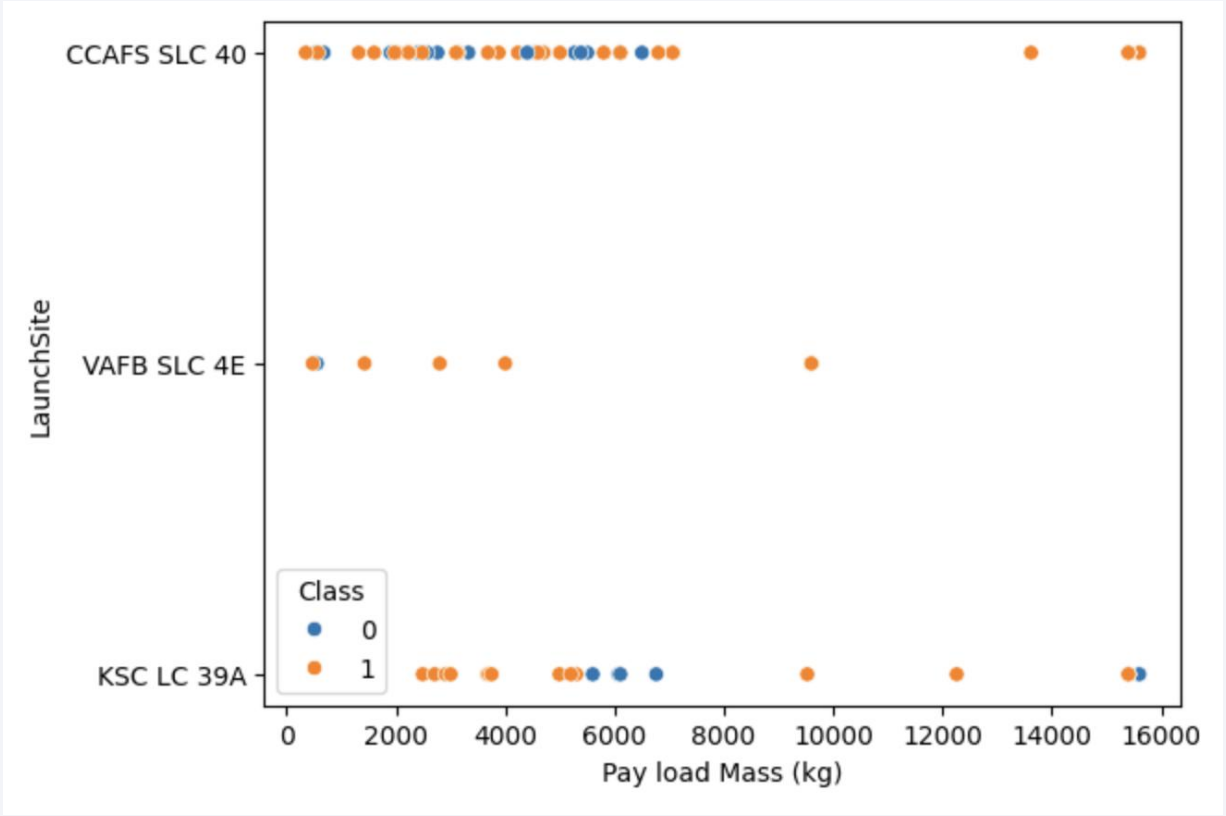
Flight Number vs. Launch Site

- As I can see, higher flight number is showing more successes
- VAFB SLC 4E launch site does not have any flight numbers higher than 65



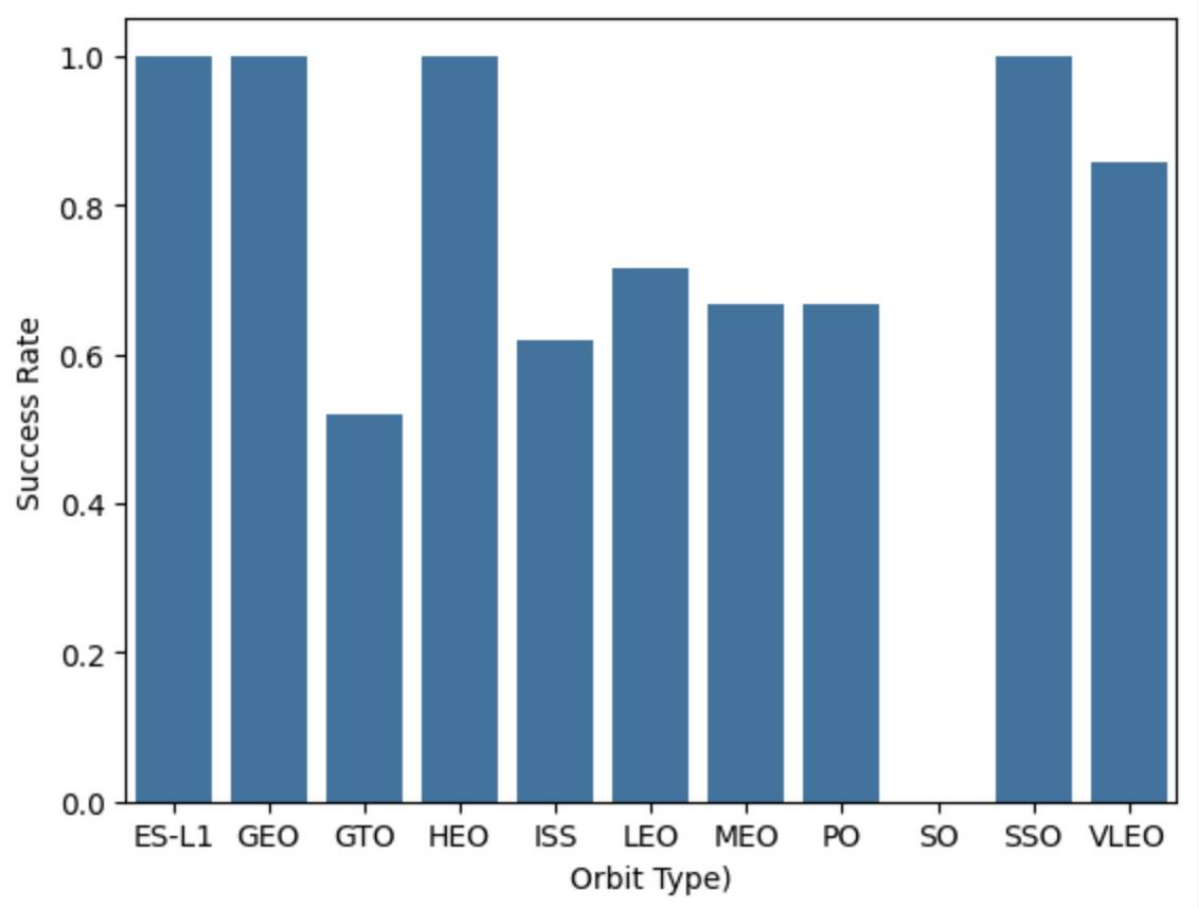
Payload vs. Launch Site

- Most pay load mass are pretty low
- There is no clear correlation with payload mass and success rate as I can see
- Again VAFB SLC 4E does not have much data here and they do not have any higher than 10000kg payload mass launches



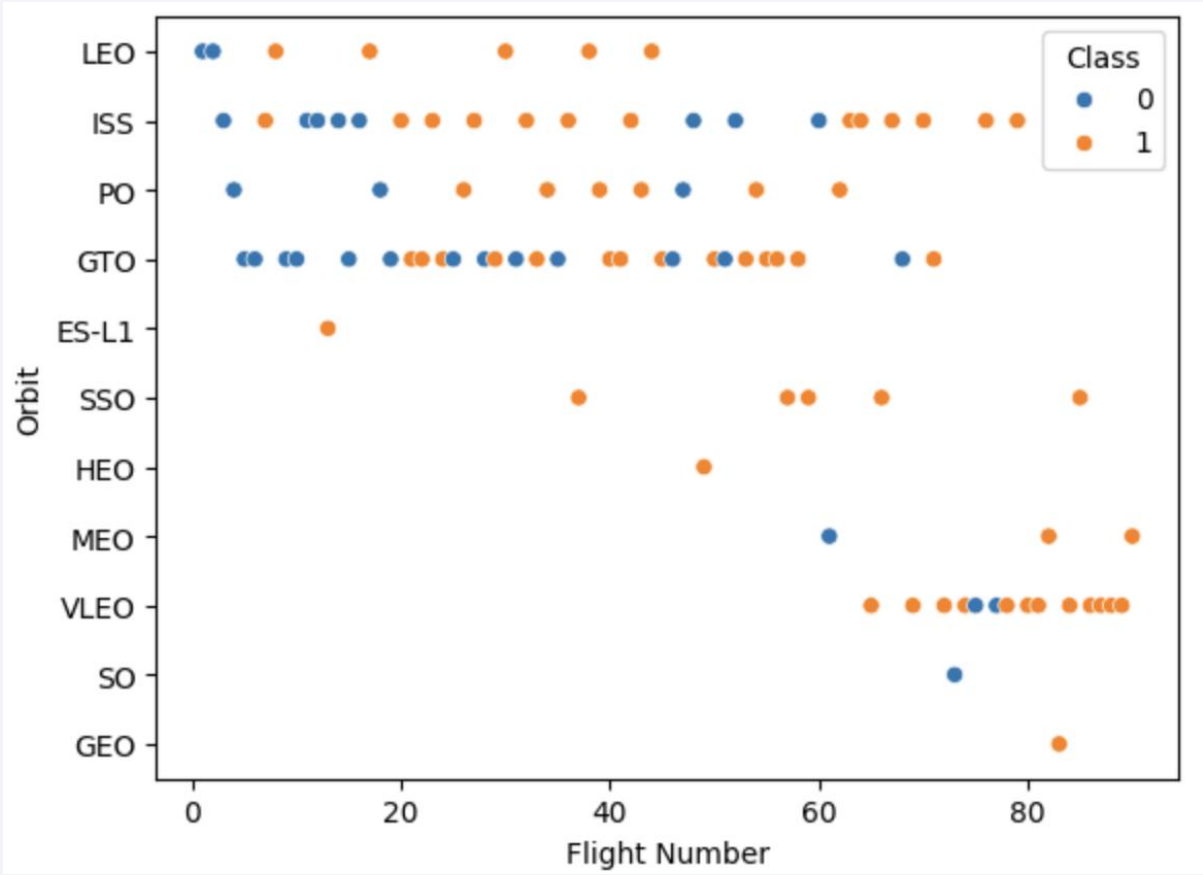
Success Rate vs. Orbit Type

- GTO and SO has the lowest success rate
- ES–L1, GEO, HEO, and SSO has 100% success rate
- The main reason for the stats like this could be because these orbit type does not have much launches with them



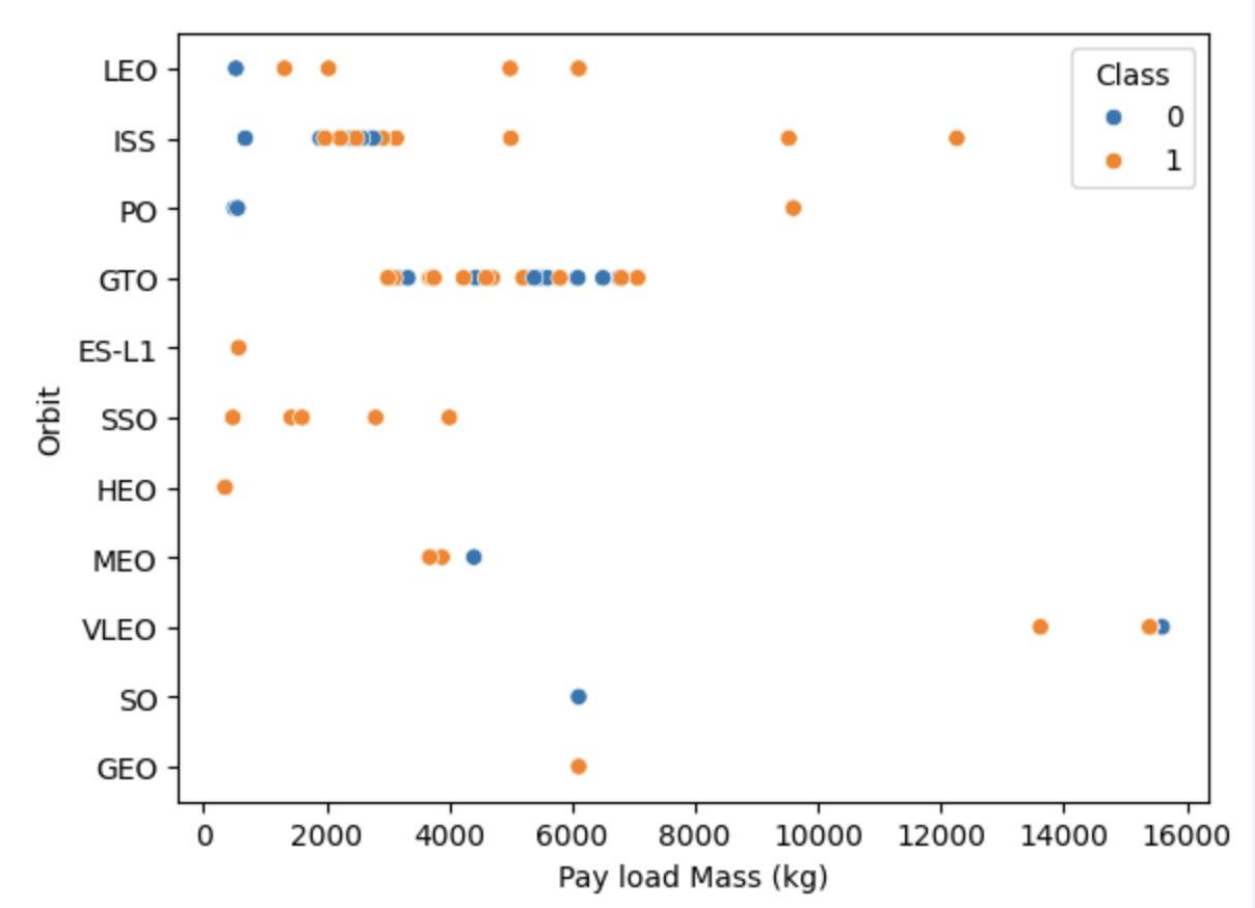
Flight Number vs. Orbit Type

- ISS has launchess across all flight numbers
- VLEO only has high flight number associated with it
- Some of the orbit type indeed only have one launch associated



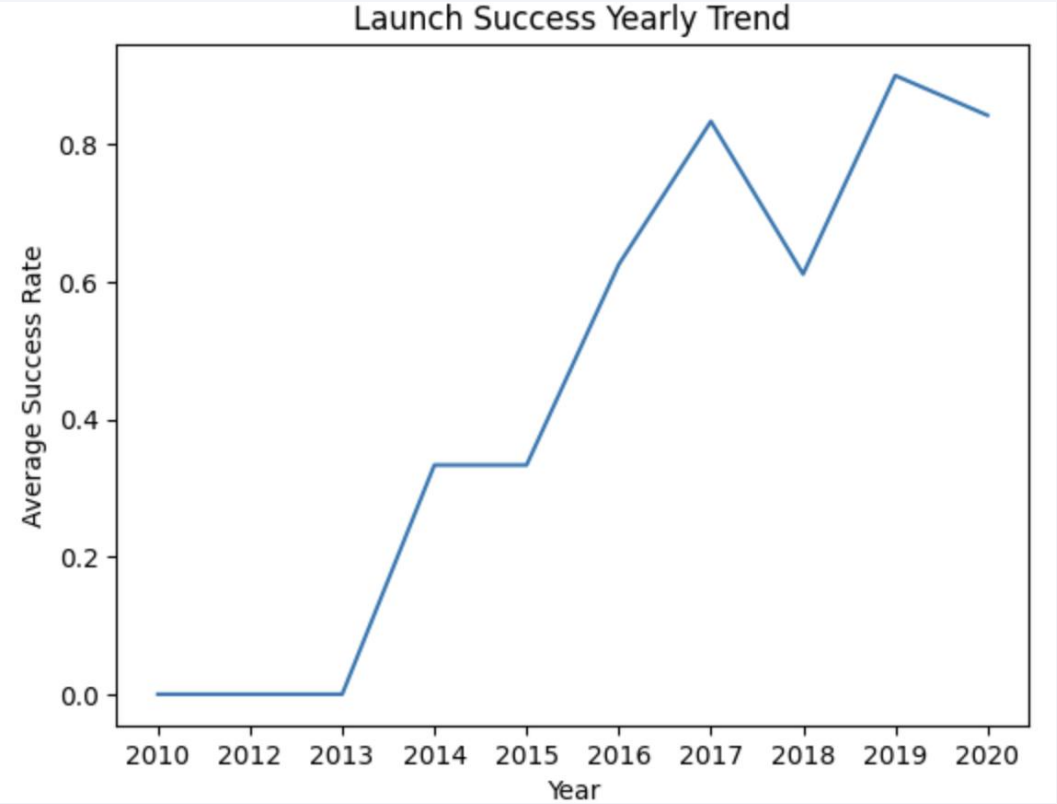
Payload vs. Orbit Type

- SSO only have low payload mass associated with it and they are all succesful
- Most of the launches have payload mass are lower than 8000
- There is no clear visible correlation between payload mass and success rate



Launch Success Yearly Trend

- Overall, success rate is going up
- It droped down by 20% in the year 2018
- It went back up even more in 2019



All Launch Site Names

- There are only 4 different launch sites.

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

Launch Site Names Begin with 'CCA'

- 4 of these records were in the orbit LEO

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- Total Payload Mass is 45596 kg

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass by F9 v1.1 is 2534.7kg
- This is actually pretty low comparing to others

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- The first succesful landing was already pretty late in time
- This can further support my findings where success rate has gone up over the years

MIN(Date)
2015-12-22

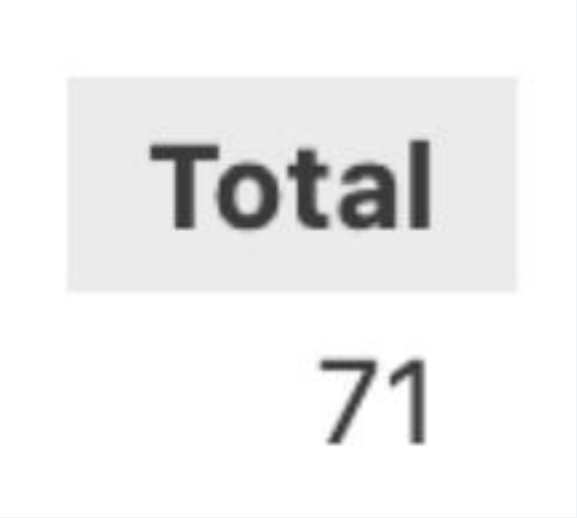
Successful Drone Ship Landing with Payload between 4000 and 6000

- Below are the 4 versions I found

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

Total Number of Successful and Failure Mission Outcomes

- Total outcomes is 71 instances
- I do not have much data to work with here



Boosters Carried Maximum Payload

- Here are all the booster versions I found
- They are all B5 versions

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

- There is only 2 records both v1.1

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

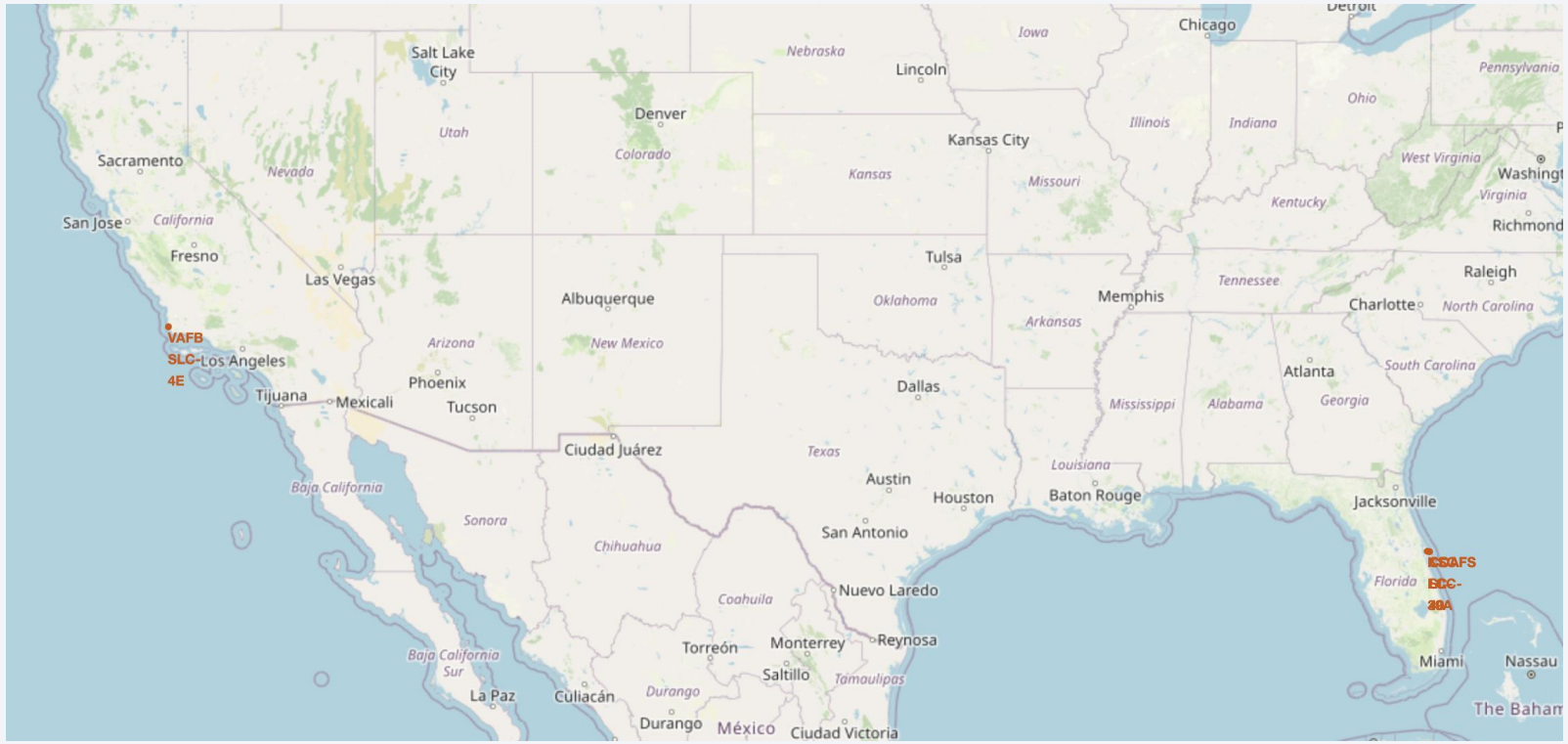
- Here is my findings
- No attempt ranked first with 10 instances

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

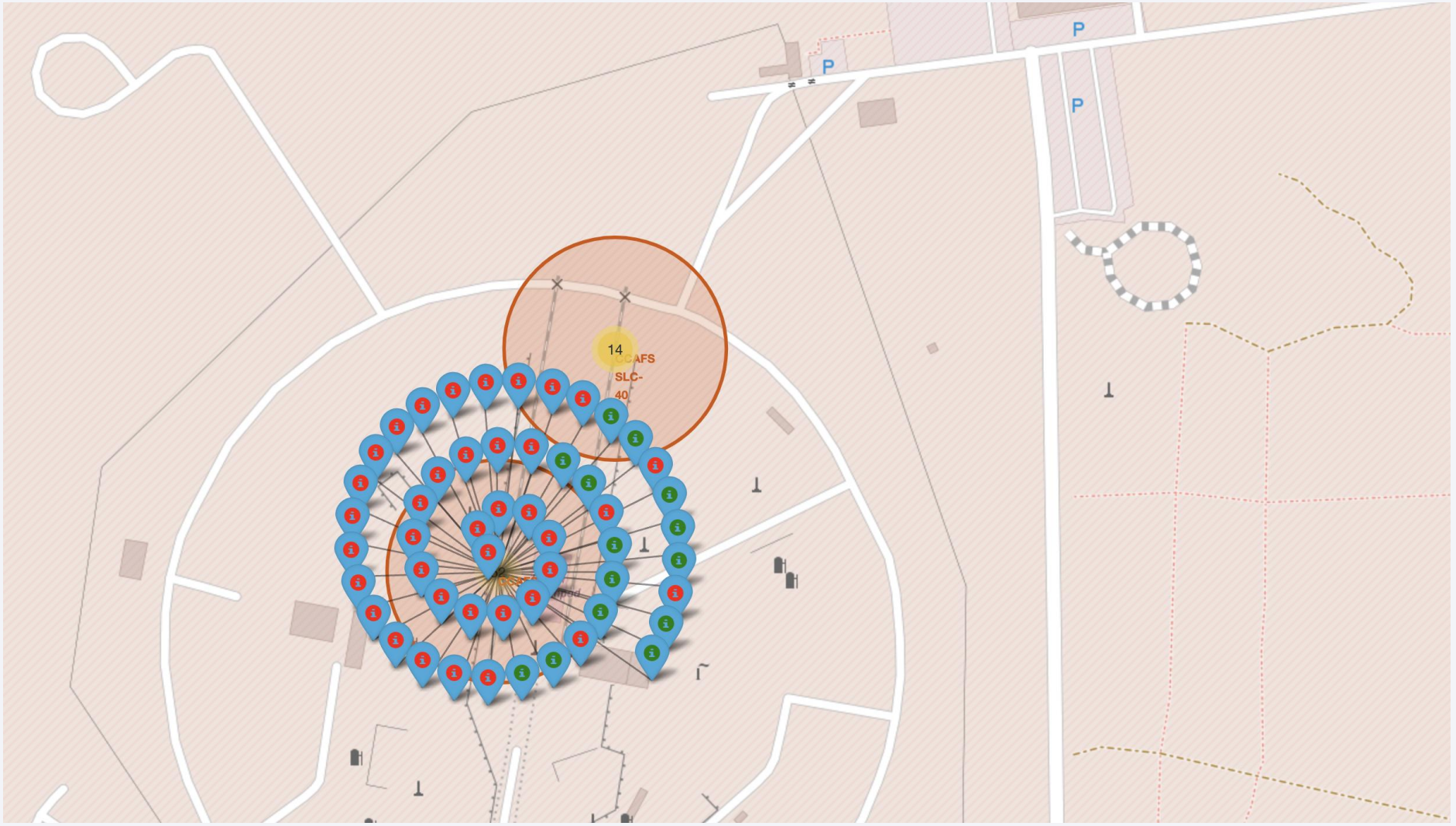

Section 3
Launch Sites
Proximities Analysis

Launch Site Locations

- They are all close to the coast line from the map we can see

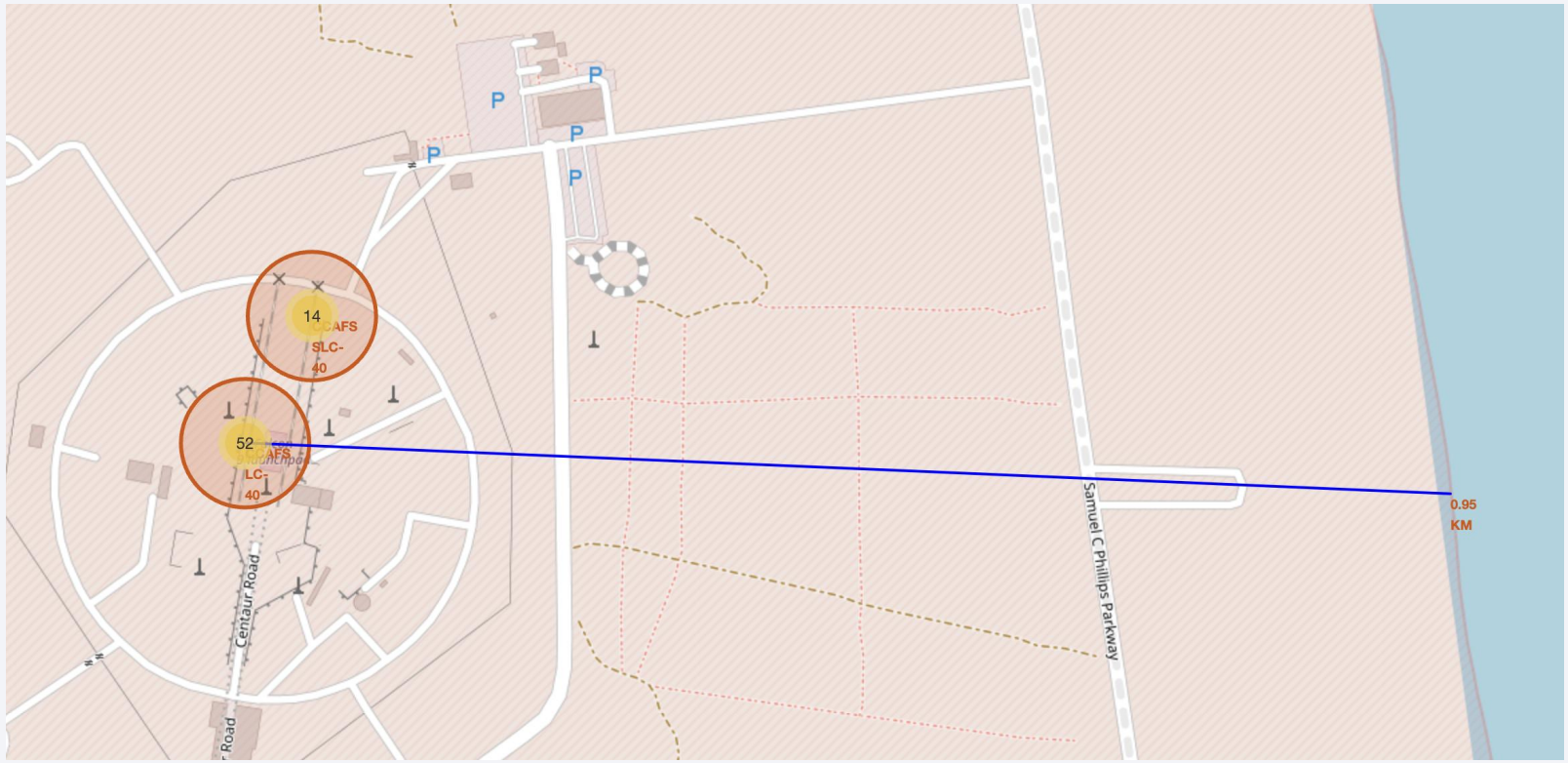


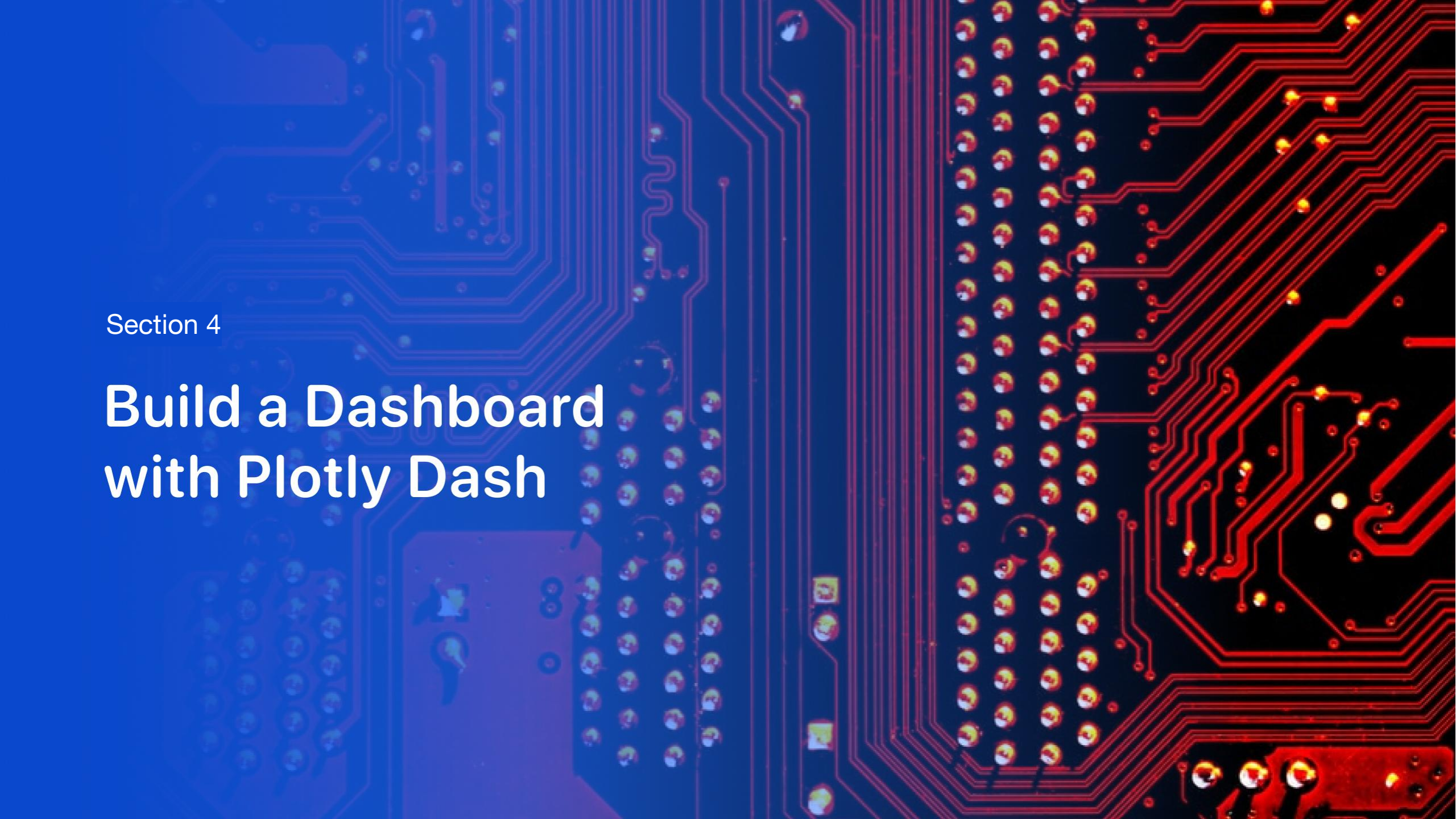
Launch Site with color labled outcomes



Launch site to coast line

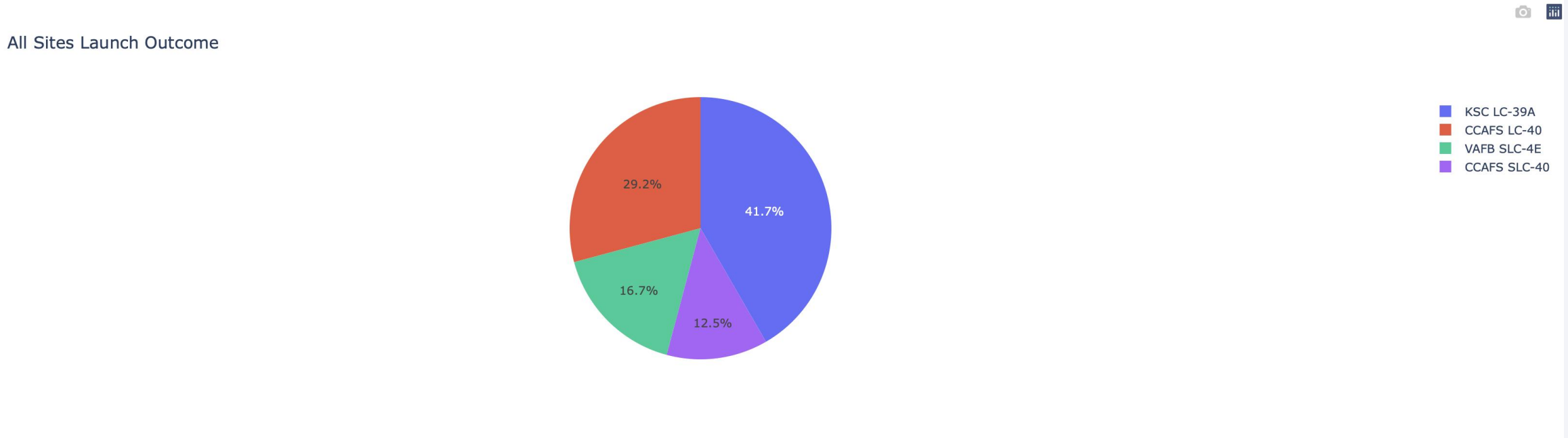
- It is less than 1 km from the coast line




Section 4
Build a Dashboard with Plotly Dash

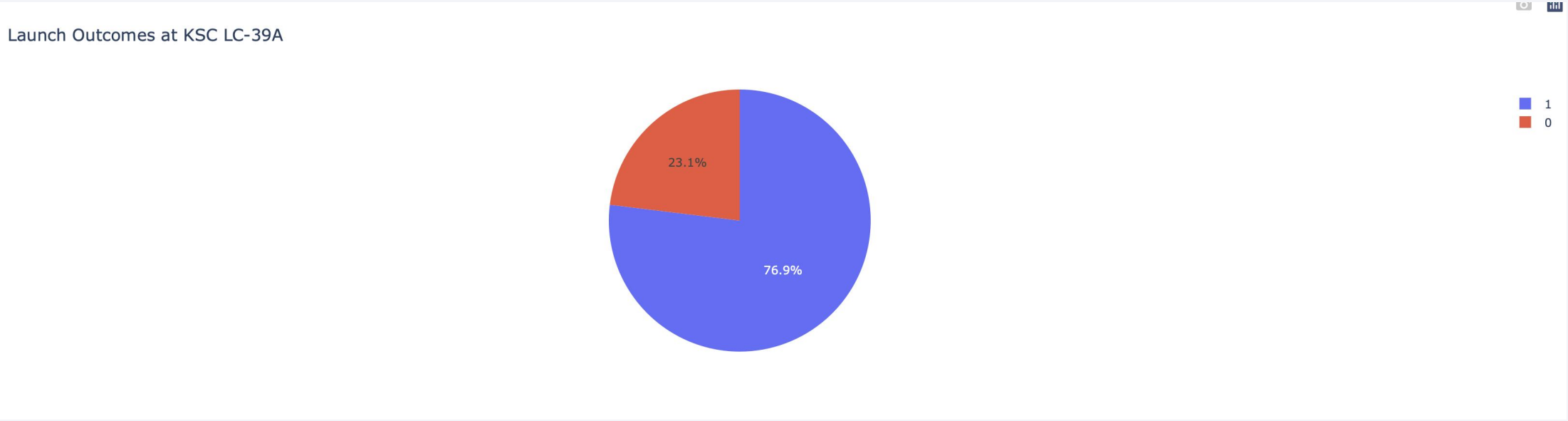
All Sites Launch Outcome

- KSC LC–39A has the most succesful launches
- VAFB SLC–4E has the least succesful launches



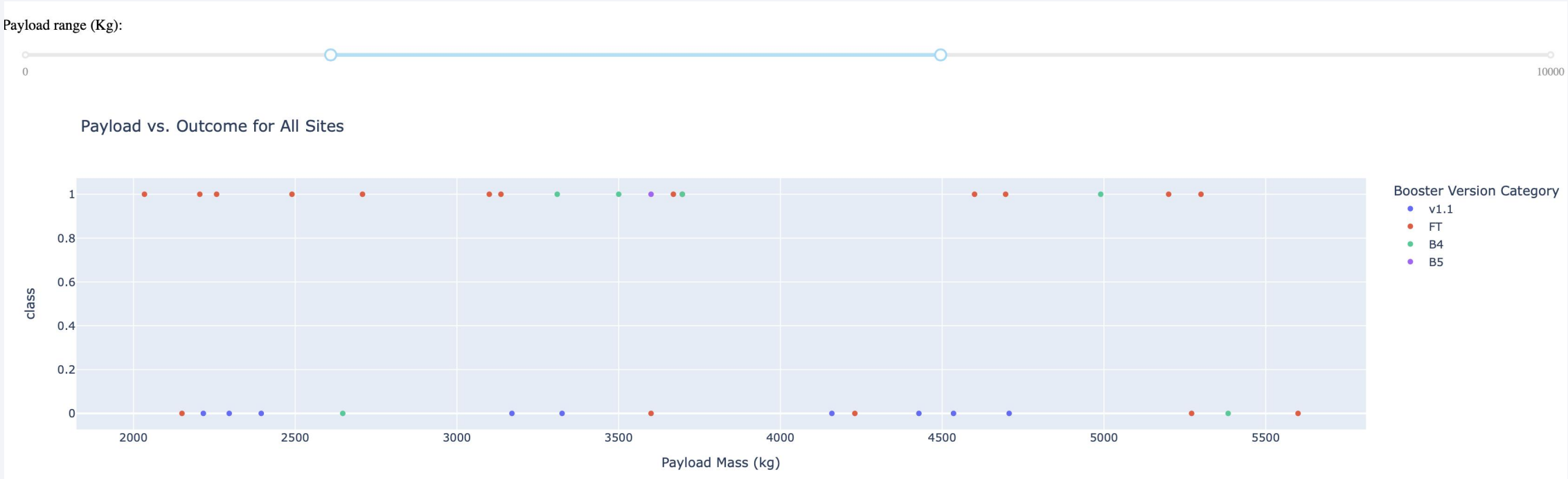
Lauch Outcome at KSC LC-39A

- It has a very high 76.9% succesuful rate
- It only has a 23.1% unsuccesful rate



Payload vs. Outcome for All Sites (2000kg - 6000kg)

- Here is the payload vs. Outcome for all sites between 2000kg and 6000kg
- As you can see, v1.1 all reported unsuccessful
- However, FT were mostly successful

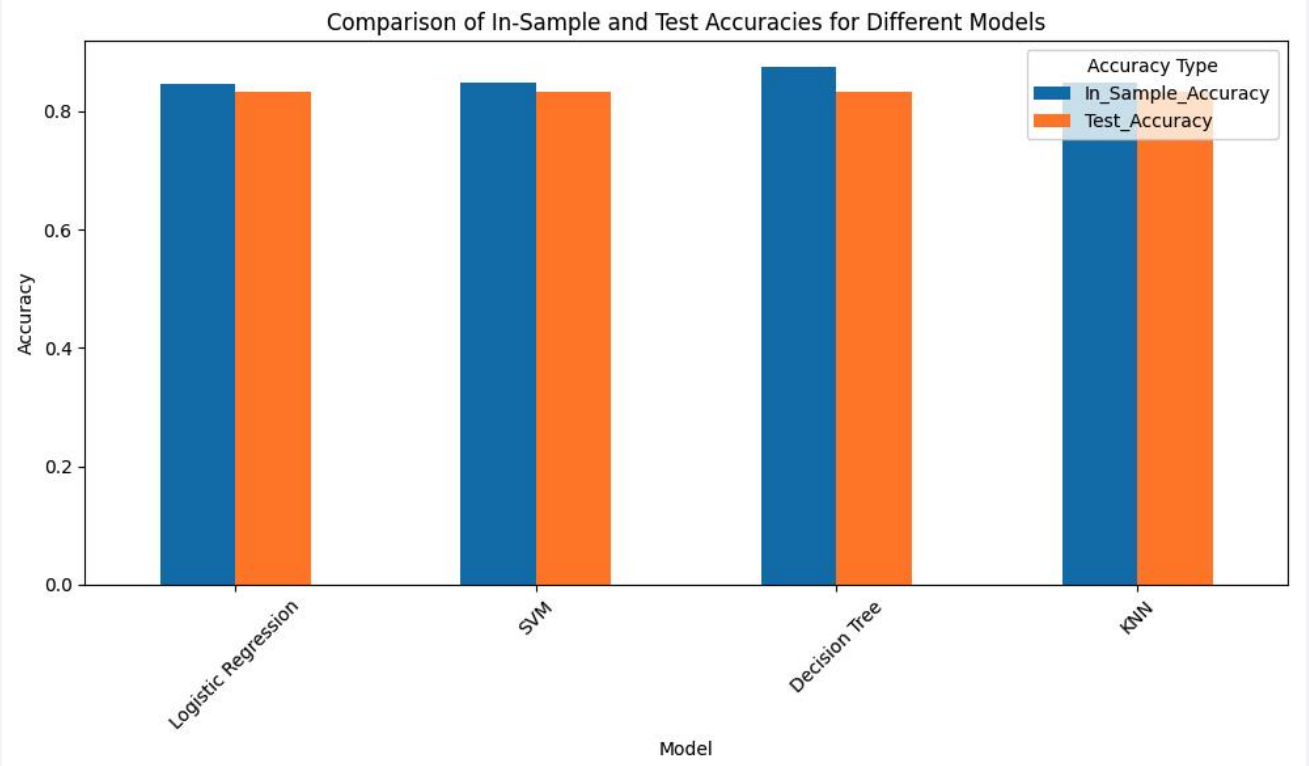


Section 5

Predictive Analysis (Classification)

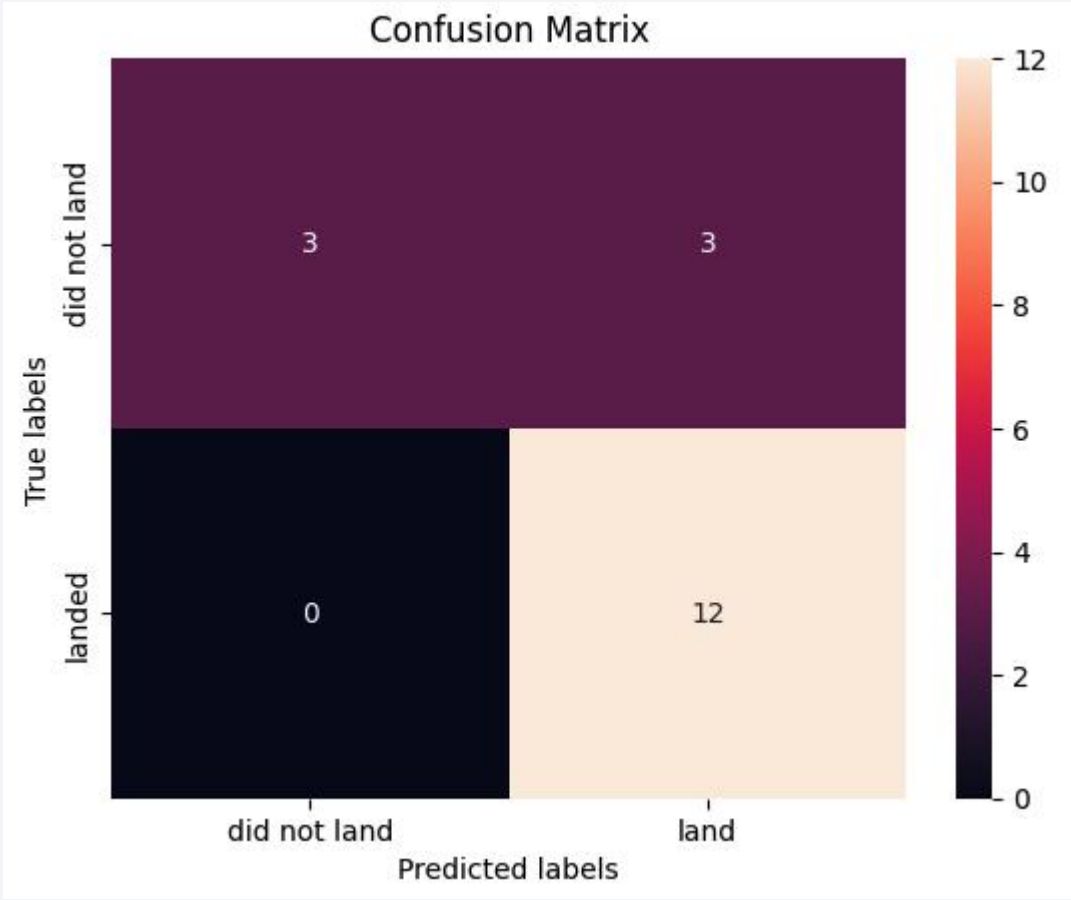
Classification Accuracy

- All 4 models have the same test accuracy, however, decision tree has the highest in sample accuracy making it the best performing model



Confusion Matrix

- Decision tree has the perfect predicition of landed outcomes
- However it had 3 instances where it predicted land when the true out come is did not land
- Resulting that it has 3 False Positives



Conclusions

- I succesfully used API and webscraping to collect data that I need
- I found out that SpaceX success rate is going up over the years through EDA
- There are 4 launch sites that Space X is using where they are all very close to coast line
- I was able to find a good Decision Tree model with 84% in samle accuracy and 83% test accuracy that can effectively predict the launch outcomes

Appendix

I do not have any appendix.

Thank you!