Winning Space Race with Data Science

Oran Szachter
9/2/2023

IBM Developer
SKILLS NETWORK

Outline

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

Executive Summary

- We got data from SpaceX API and Wikipedia
- Using Pandas we performed EDA and showed:
  - successful launch sites
  - Correlation between Payloads and success rate
  - Time evolution of success rate
  - Successful orbits
  - etc.
- Using SQL we got a summary of the data
- Using Folium we got geographical data
- We created a dashboard for interactive exploration
- Using Machine Learning we got 83% accuracy for outcome prediction

Introduction

- Many commercial companies are making space travel affordable
- SpaceX is the most successful one because its rocket launches are relatively inexpensive
- The main reason is that SpaceX can reuse the rocket's first-stage
- We want to predict whether the first stage of the Falcon 9 rocket will land or crash

Section 1

Methodology

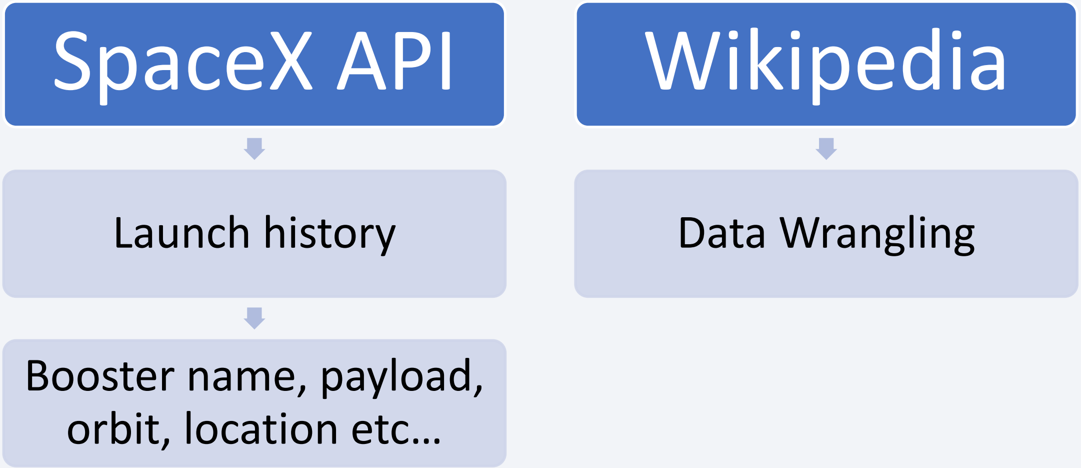
Methodology

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

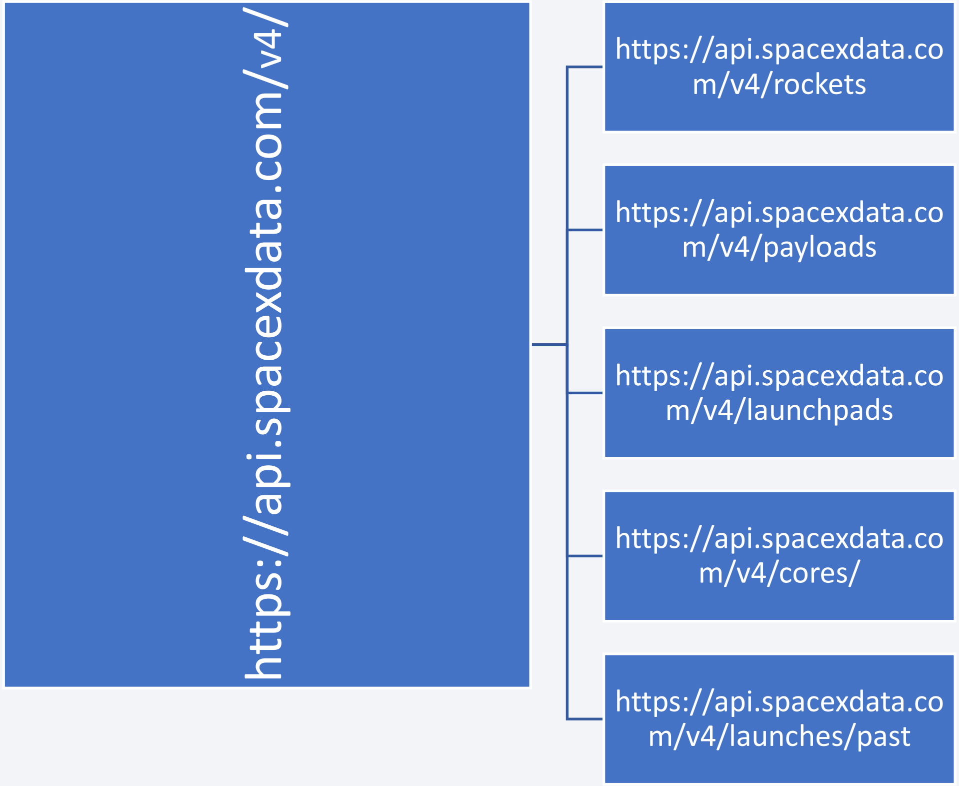
Data Collection

- We used the SpaceX API to get data about launches history
- From Wikipedia we collected Falcon 9 records
- We used data wrangling to present success and failure as binary variables



Data Collection – SpaceX API

- From the rocket column, we get the booster name.
- From the payload, we get the mass of the payload and the orbit it is going to.
- From the launchpad, we get the name of the launch site used, the longitude, and the latitude.
- From cores, we get other data related to the core
- We use this data to interpret the rocket launch data from the past API
- [GitHub of the completed SpaceX API calls notebook]()



Data Collection - Scraping

- We Extract a Falcon 9 launch records HTML table from Wikipedia https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- We Parse the table and convert it into a Pandas data frame
- We saved the data into a CSV file
- GitHub of the web scrapping notebook

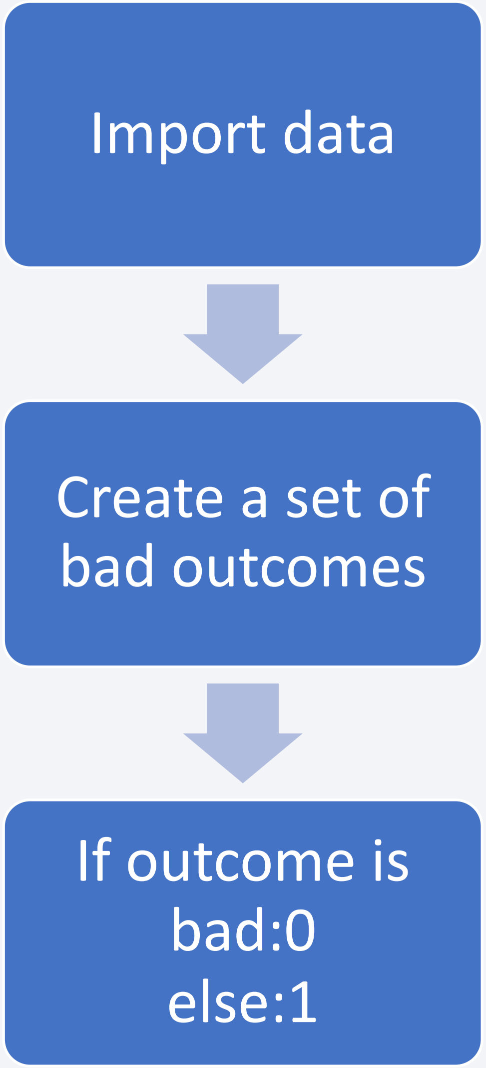
2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[11] in addition to 14 or 15 non-Starlink launches. At 26 launches, 14 of which were for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[12]

[hide] Flight No.	Date and time (UTC)	Version, booster[a]	Launch site	Payload[b]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020 02:19:21[13]	F9 B5 ♺ B1049.4	CCSFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb) [14]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[15]								
79	19 January 2020 15:30[16]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[17] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[18]	NASA (CTS)[19]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[20] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[21] The abort test used the capsule originally intended for the first crewed flight.[22] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[23] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020 14:07[24]	F9 B5 ♺ B1051.3	CCSFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb) [14]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[25]								
81	17 February 2020 15:05[26]	F9 B5 ♺ B1056.4	CCSFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb) [14]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to								

Data Wrangling

- We create a set of outcomes where the second stage did not land successfully
- We create a list where the element is zero if outcome was bad; otherwise, it's one. Then assign it to a new column.
- We can calculate the success rate
- [GitHub of the data wrangling notebook]



EDA with Data Visualization

- We want to see how the Flight Number and Payload affect the success rate
- To eliminate the effect of the Launch Site we show:
  - The relationship between Flight Number and Launch Site
  - The relationship between Payload and Launch Site
- To eliminate the effect of the orbit type we show:
  - The relationship between the success rate of each orbit type
  - The relationship between Flight Number and Orbit type
  - The relationship between Payload and Orbit type
- To show the progression, we show the launch success yearly trend
- GitHub of the EDA with Data Visualization notebook

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 acheived.
- List the names of the boosters which have success in drone ship 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. Use a subquery
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- 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
- [GitHub of the EDA with SQL notebook]()

Build an Interactive Map with Folium

- To mark all launch sites on a map:
  - We used folium.Circle to add a highlighted circle area with a text label on a specific coordinate
  - We used folium.Marker to add text in a particular coordinate
- To Mark the successful/failed launches for each site on the map:
  - We used MarkerCluster to show several launches on the same site
  - We added a folium. Marker to each Cluster; colored green to indicate a success and red to indicate failure
- GitHub of the Folium notebook

Build a Dashboard with Plotly Dash

- A pie chart showing the success ratio for each site
- A scatter plot showing the successful\failed launches and their payload
- This show:
  - 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 has the highest launch success rate?
- GitHub of the complete Plotly Dash file

Predictive Analysis (Classification)

- We split the data into training data (80%) and test data (20%)
- For each model, we found the parameters that optimizes the fit
- We repeated the operations for Logistic Regression, SVM, Tree classifier and KNN
- GitHub of the predictive analysis notebook

Results

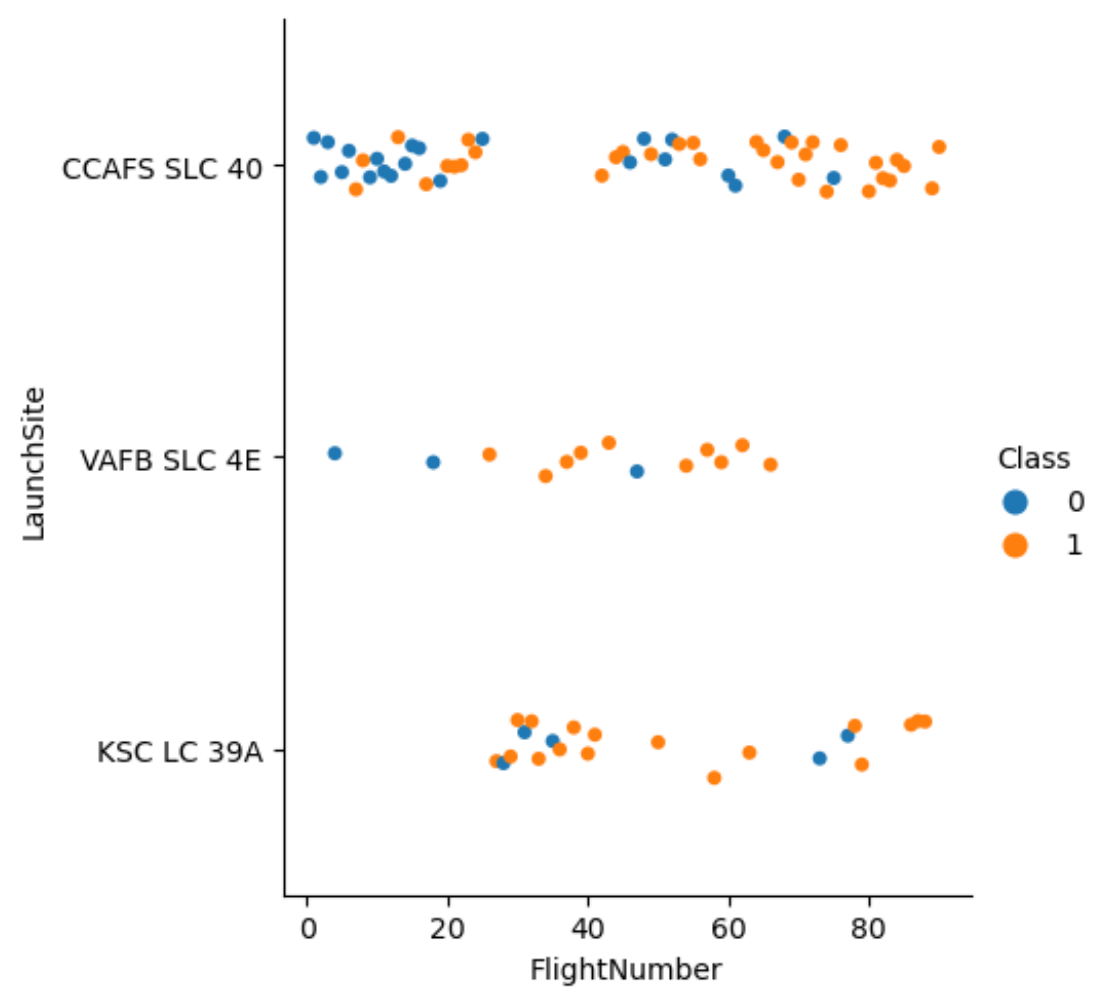
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

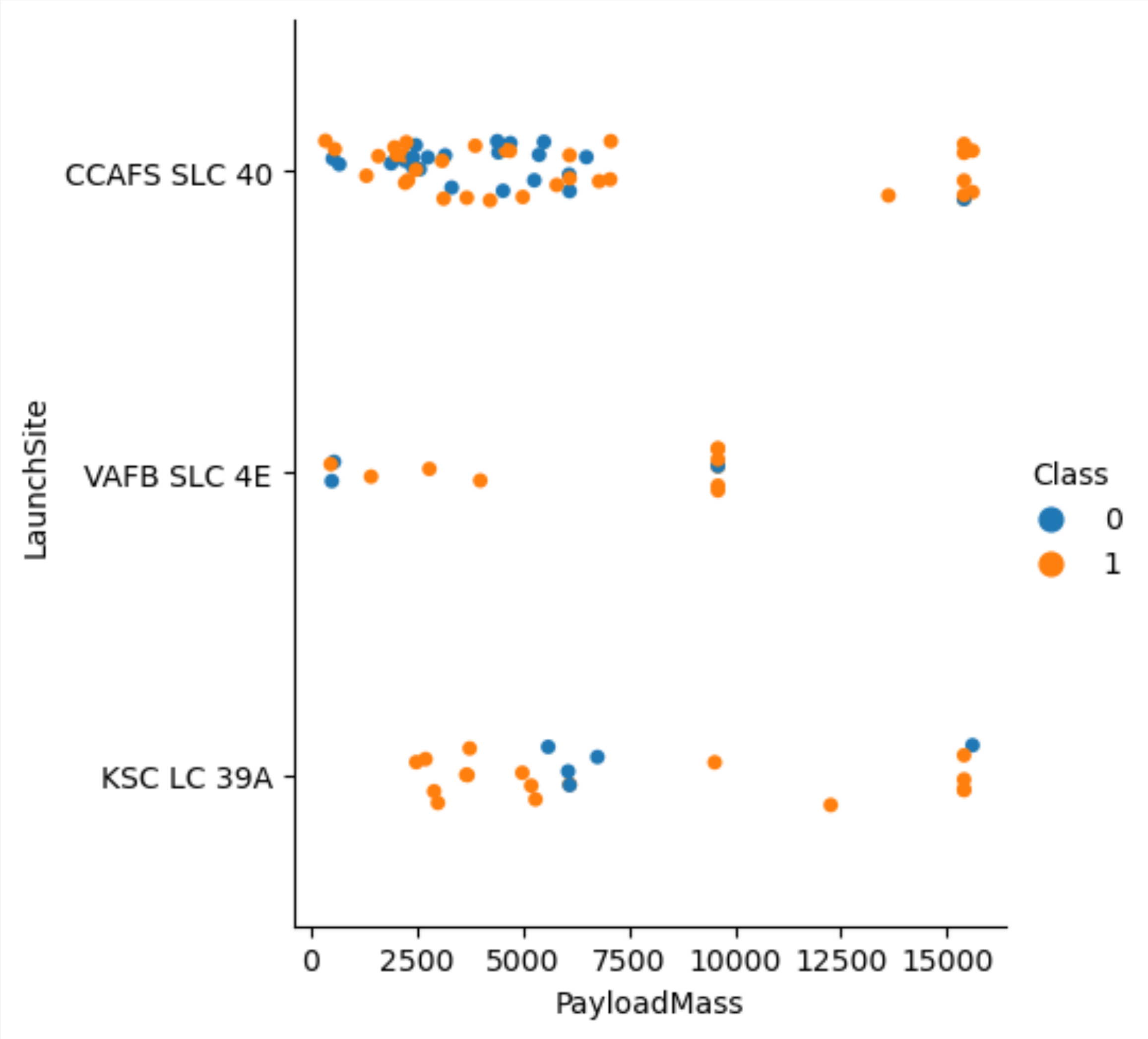
Flight Number vs. Launch Site

- It seems that the early use of CCAFS LC-40 has led to a lower success rate.



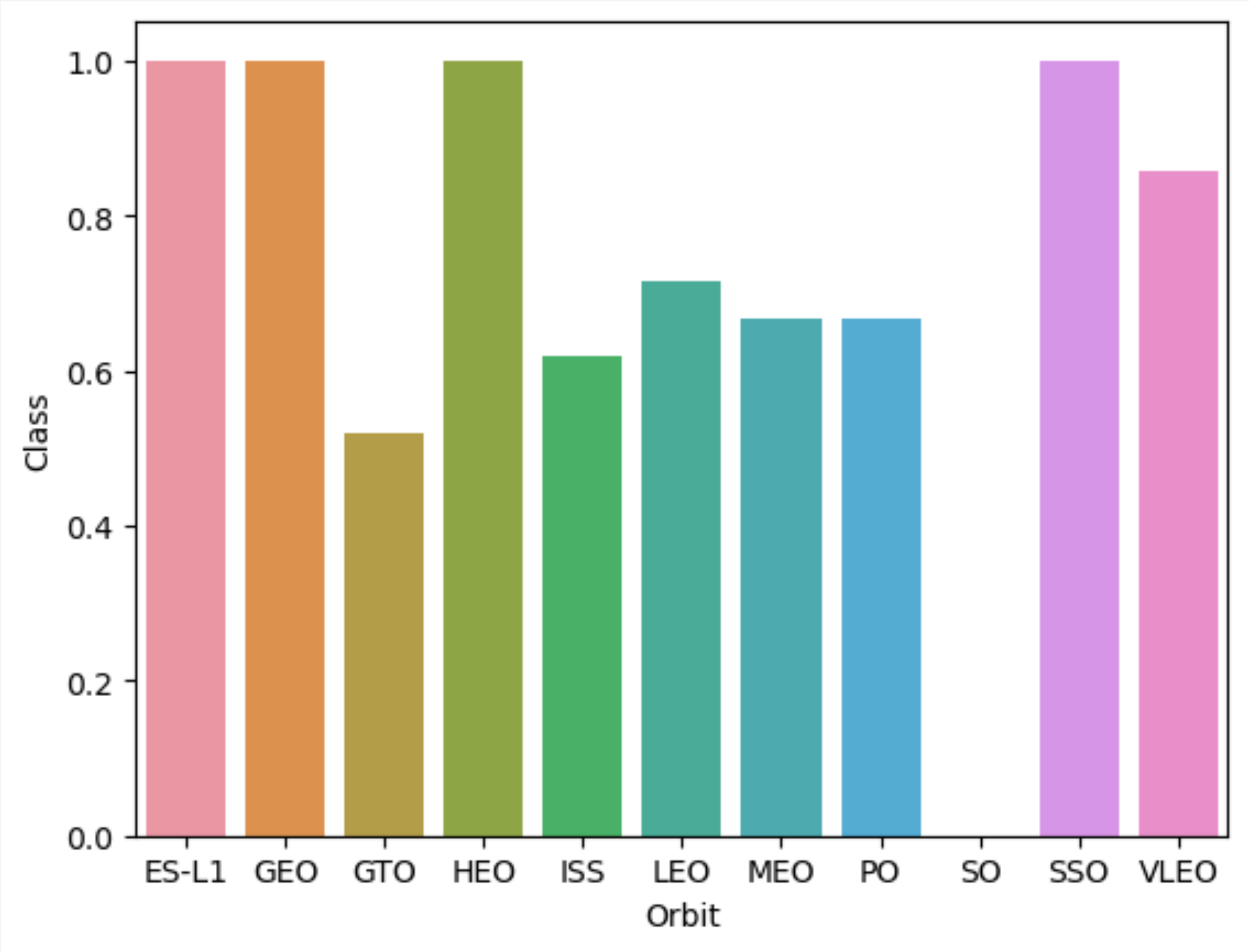
Payload vs. Launch Site

- It seems that the payload mass has little effect at CCAFS SLC-40, while low payload gives higher succes rates at KSC LC-39A



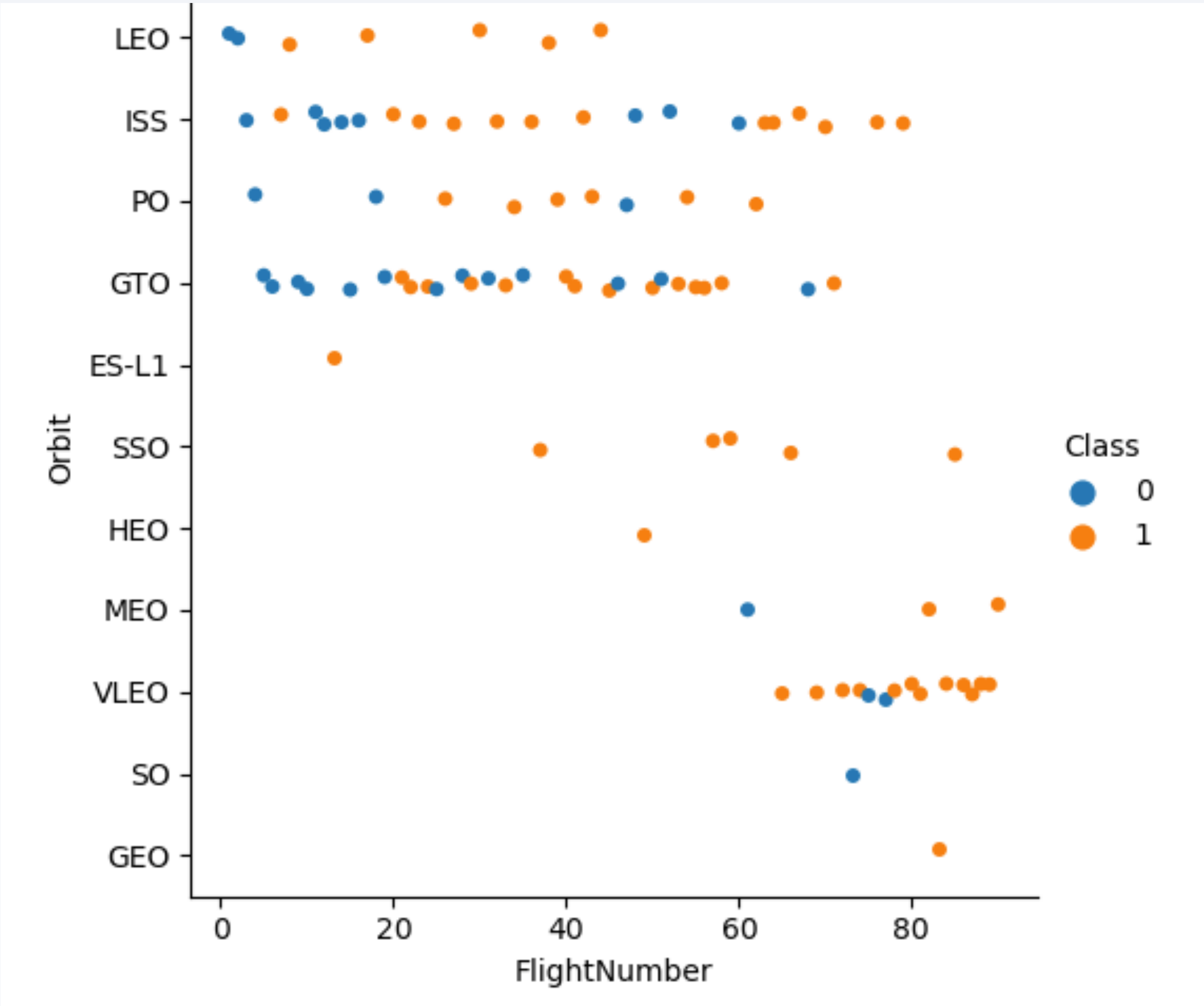
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have higher success rates
- GTO, ISS, LEO, MEO, and PO have lower success rates



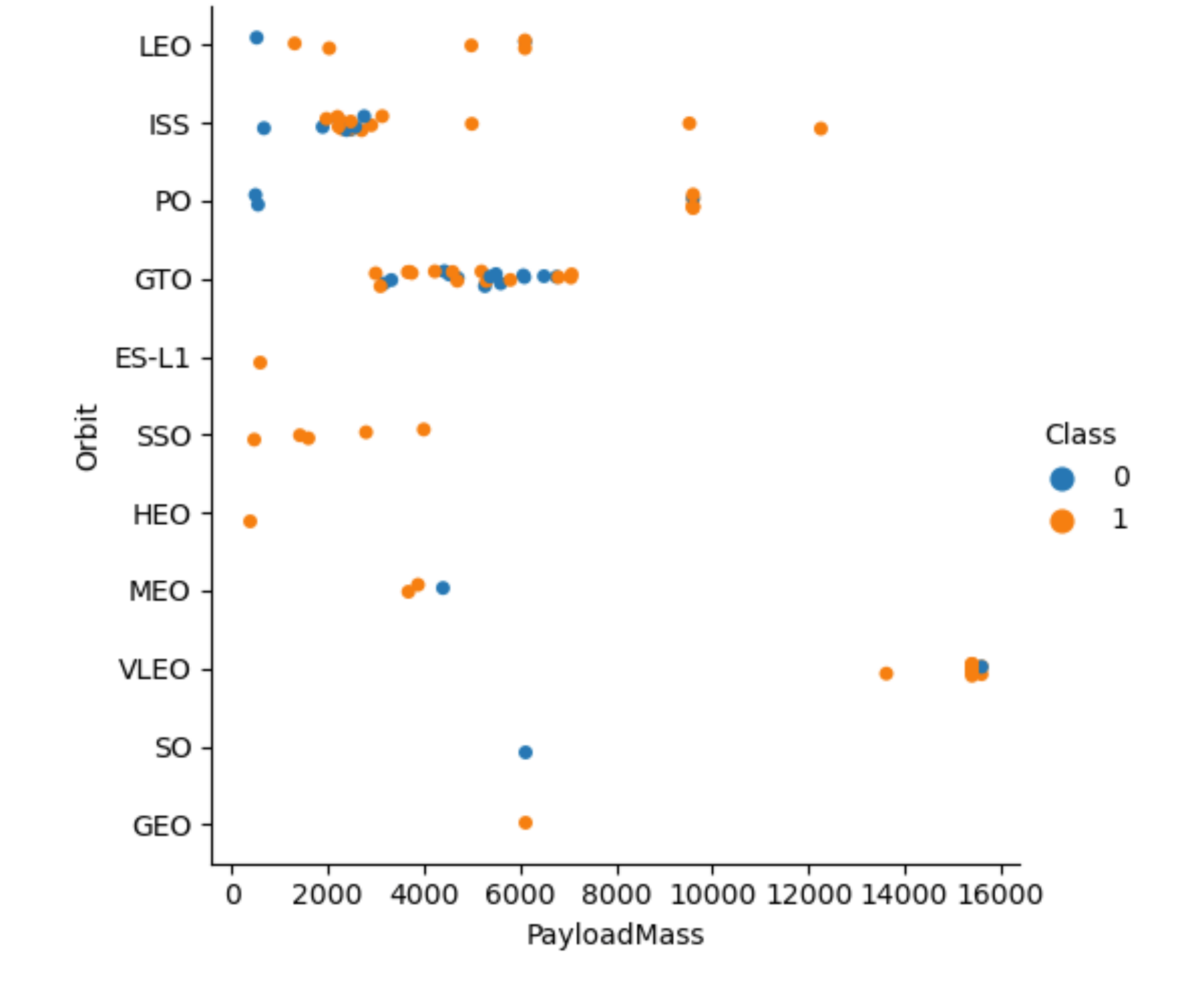
Flight Number vs. Orbit Type

- GTO and ISS show generally lower success rates, while the lower success rates of LEO and PO may be explained by their early use



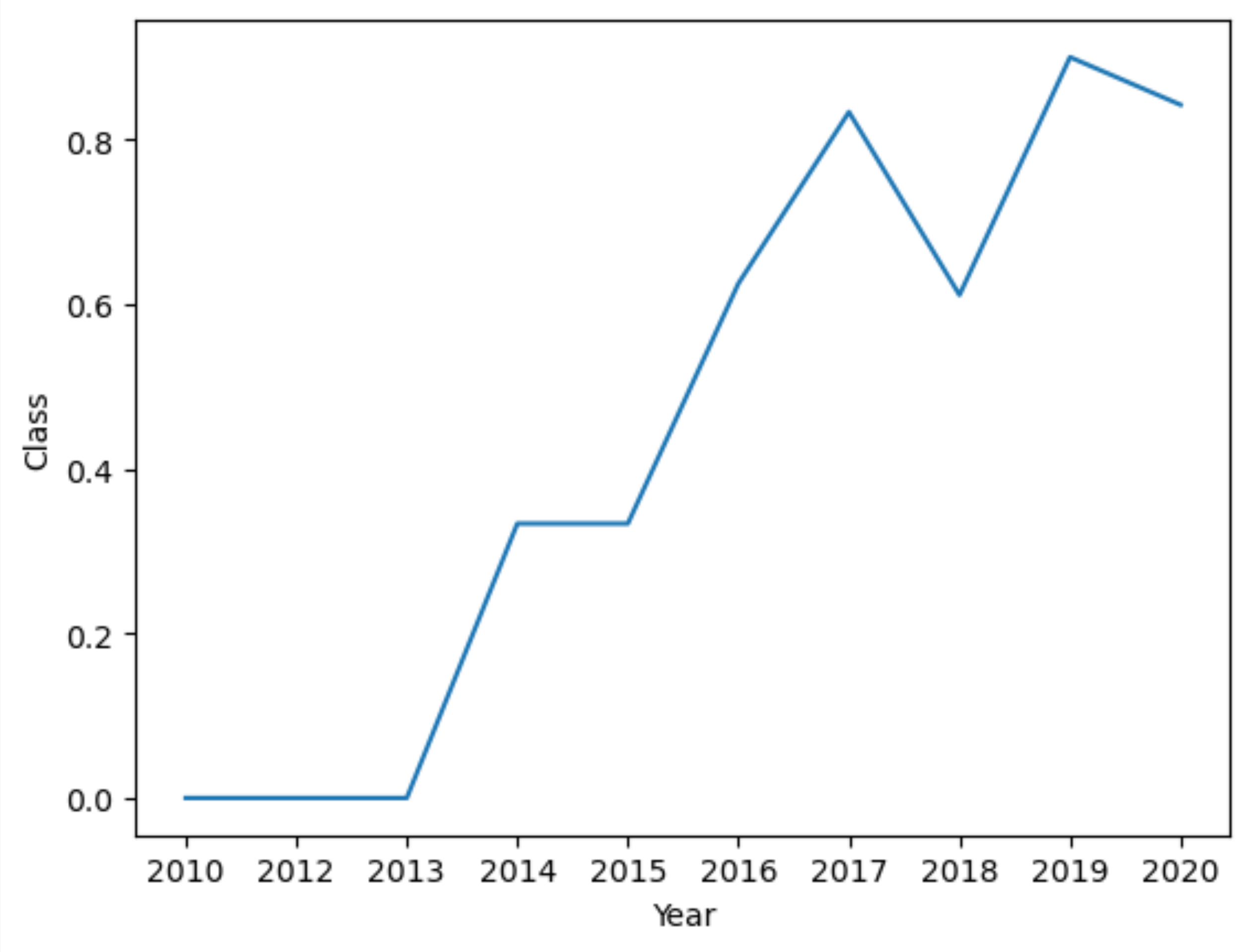
Payload vs. Orbit Type

- LEO, ISS, PO and GTO and MEO were used with lower PayloadMASS.
- However, this cannot explain their lower success rate, because the successful ES-L1, SSO and HEO were also used with lower PayloadMass



Launch Success Yearly Trend

- The success rate increased significantly between the year of 2012 and 2016



All Launch Site Names

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

Launch Site Names 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 was 619967

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 2928

First Successful Ground Landing Date

- The first successful landing outcome on ground pad was on December 22nd 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

F9 FT B1021.1
F9 FT B1023.1
F9 FT B1029.2
F9 FT B1038.1
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

Failure	3
Failure (drone ship)	5
Failure (parachute)	2
Success	38
Success (drone ship)	14
Success (ground pad)	9

Boosters Carried Maximum Payload

F9 B5 B1048.4

F9 B5 B1048.5

F9 B5 B1049.4

F9 B5 B1049.5

F9 B5 B1049.7

F9 B5 B1051.3

F9 B5 B1051.4

F9 B5 B1051.6

F9 B5 B1056.4

F9 B5 B1058.3

F9 B5 B1060.2

F9 B5 B1060.3

2015 Launch Records

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

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

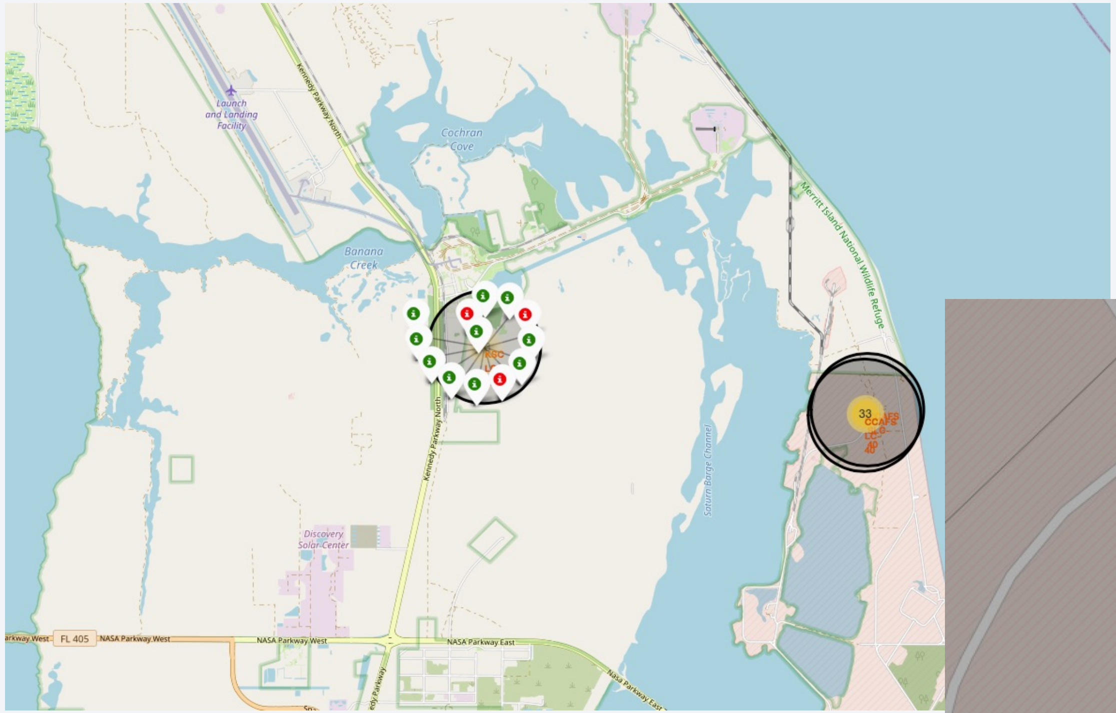
Section 3

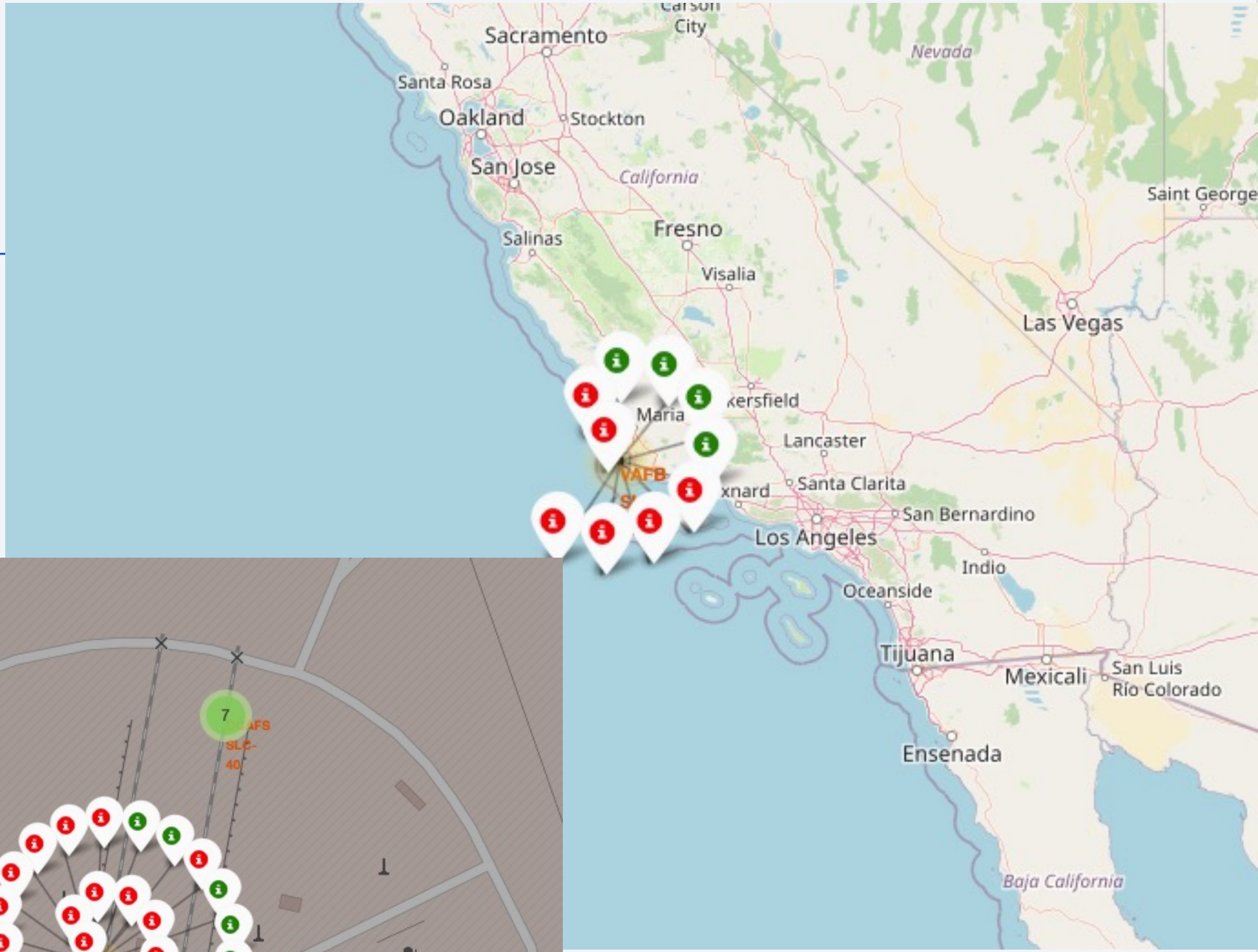
Launch Sites Proximities Analysis

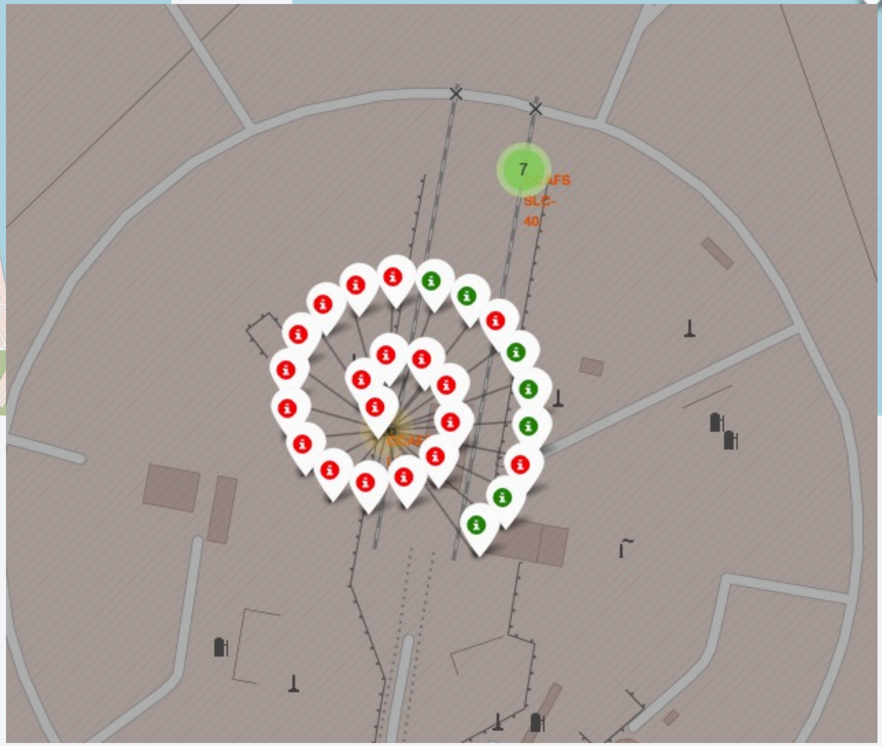
The Four Launch Sites

- The four launch sites are located near the coast of the southern part of the USA

Success/Failed Launches

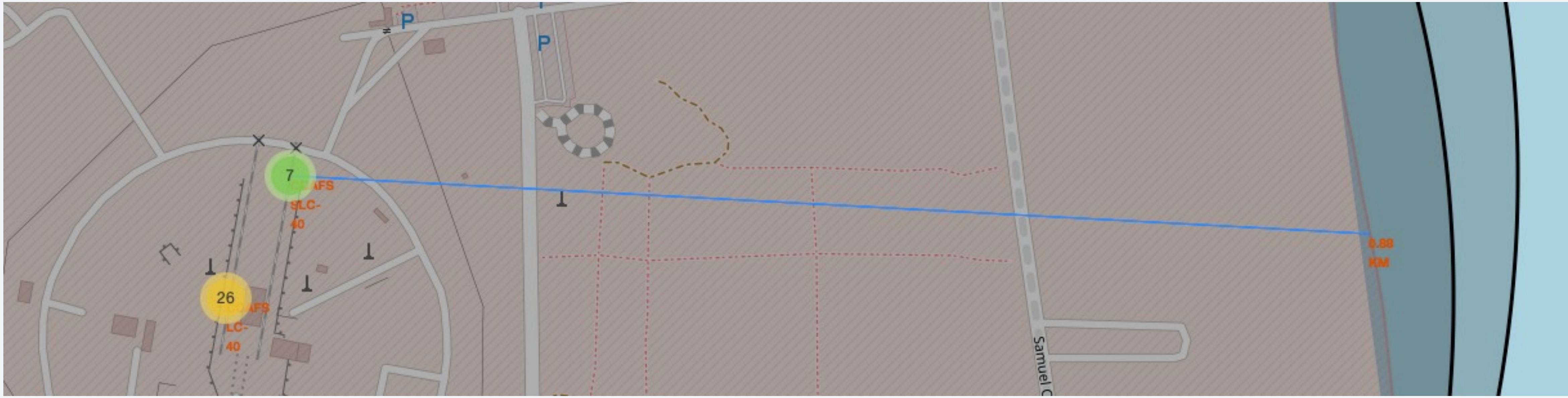






- Both on the east and west coast there are successes and failures

Proximity to Coastline



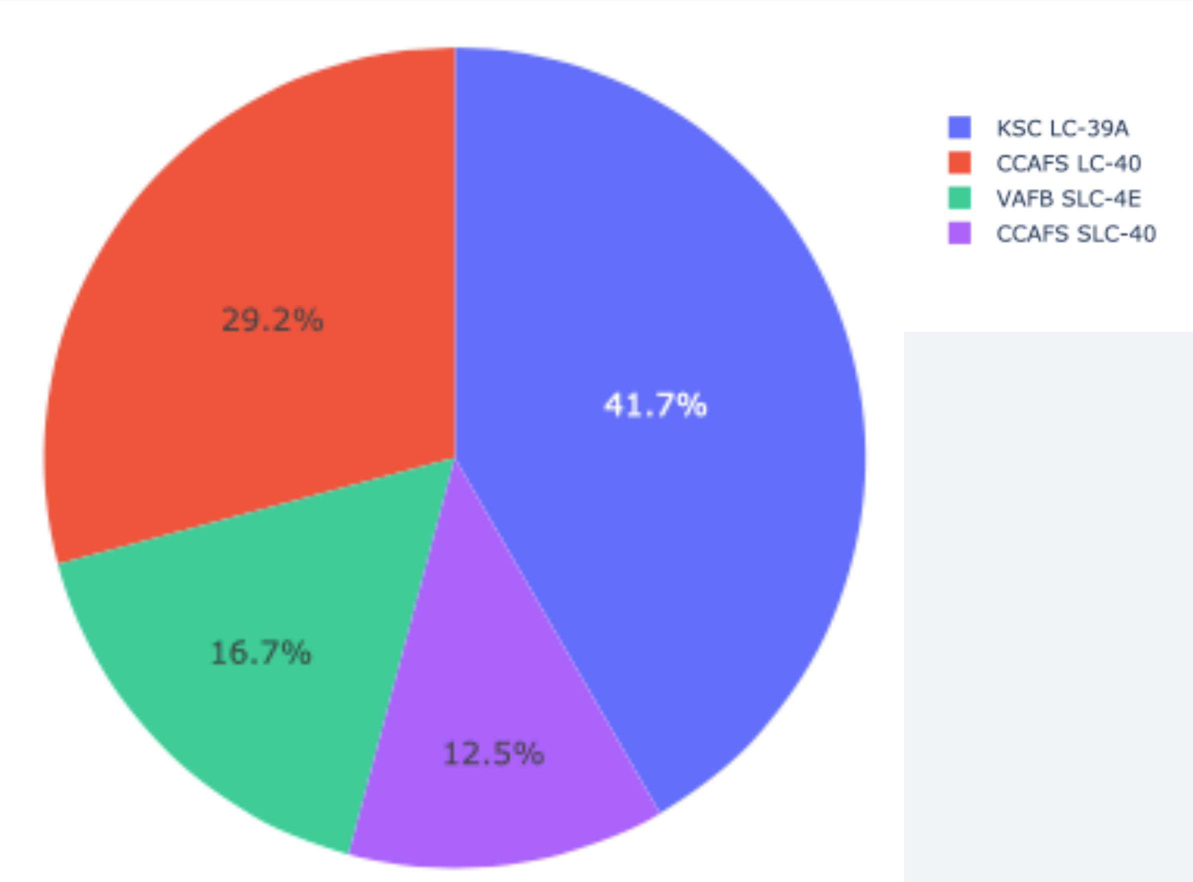
- The launch sites are located near the coast

Section 4

Build a Dashboard with Plotly Dash

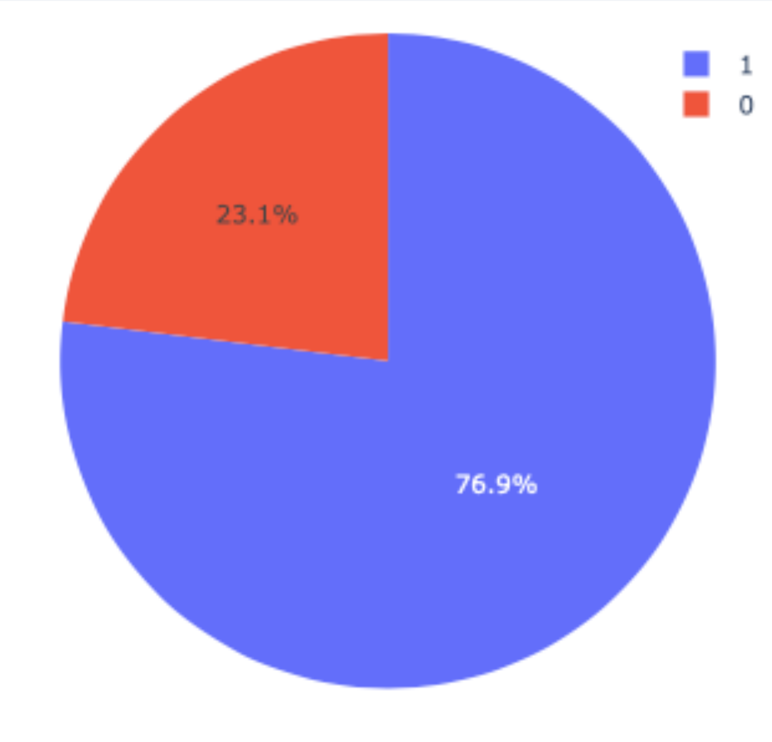
Launch Success Count

- The KSC LC-39A has the largest successful launches

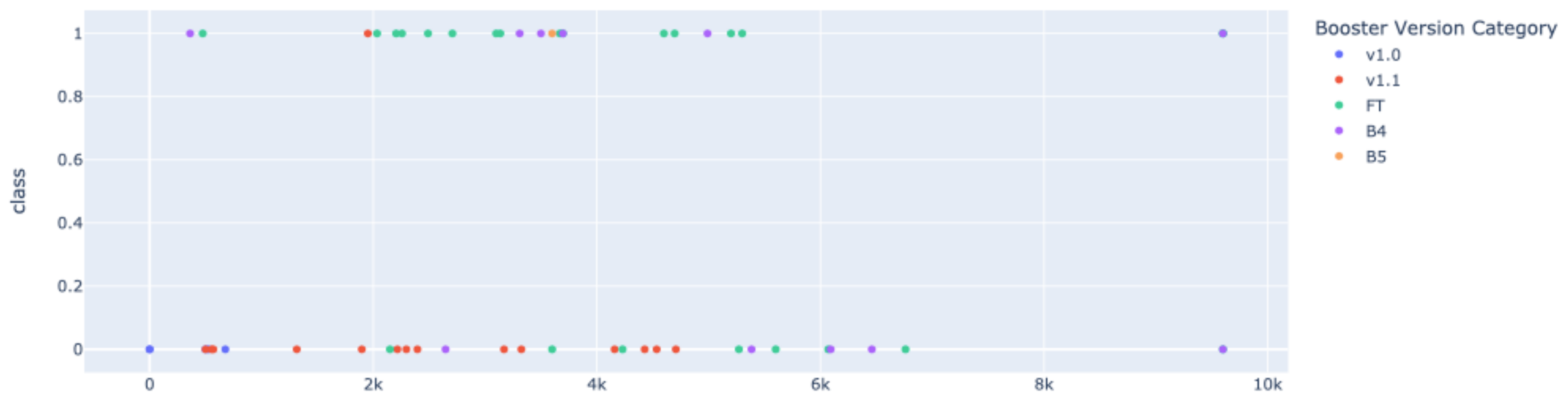


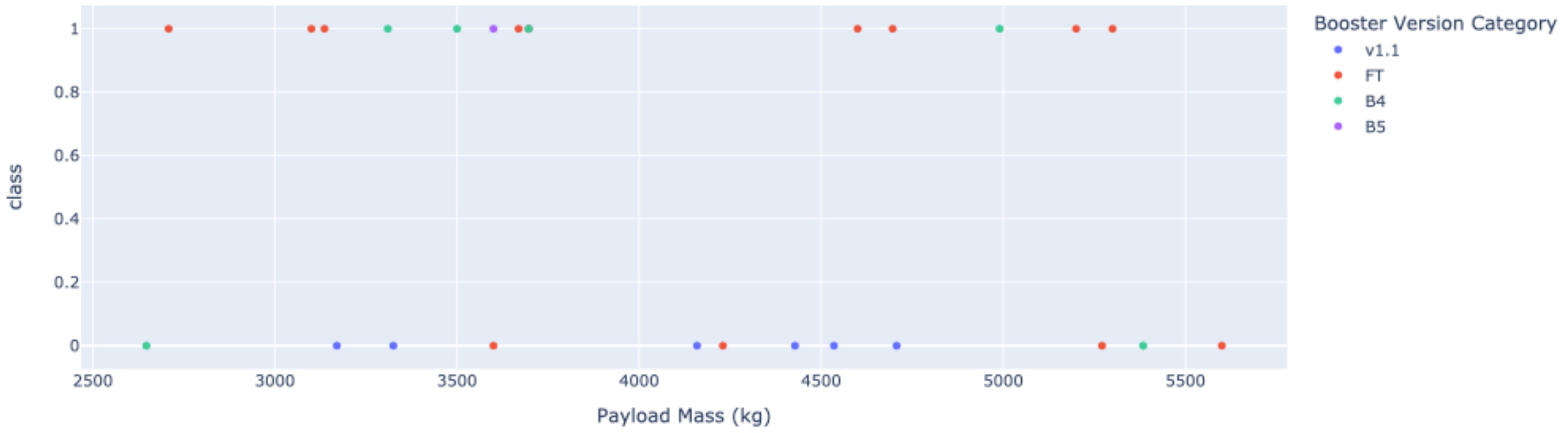
Launch site with highest success rate

- Not only KSC LC-39A has the largest successful launches, it has the largest success rate



Success rate at different Payloads for different Boosters





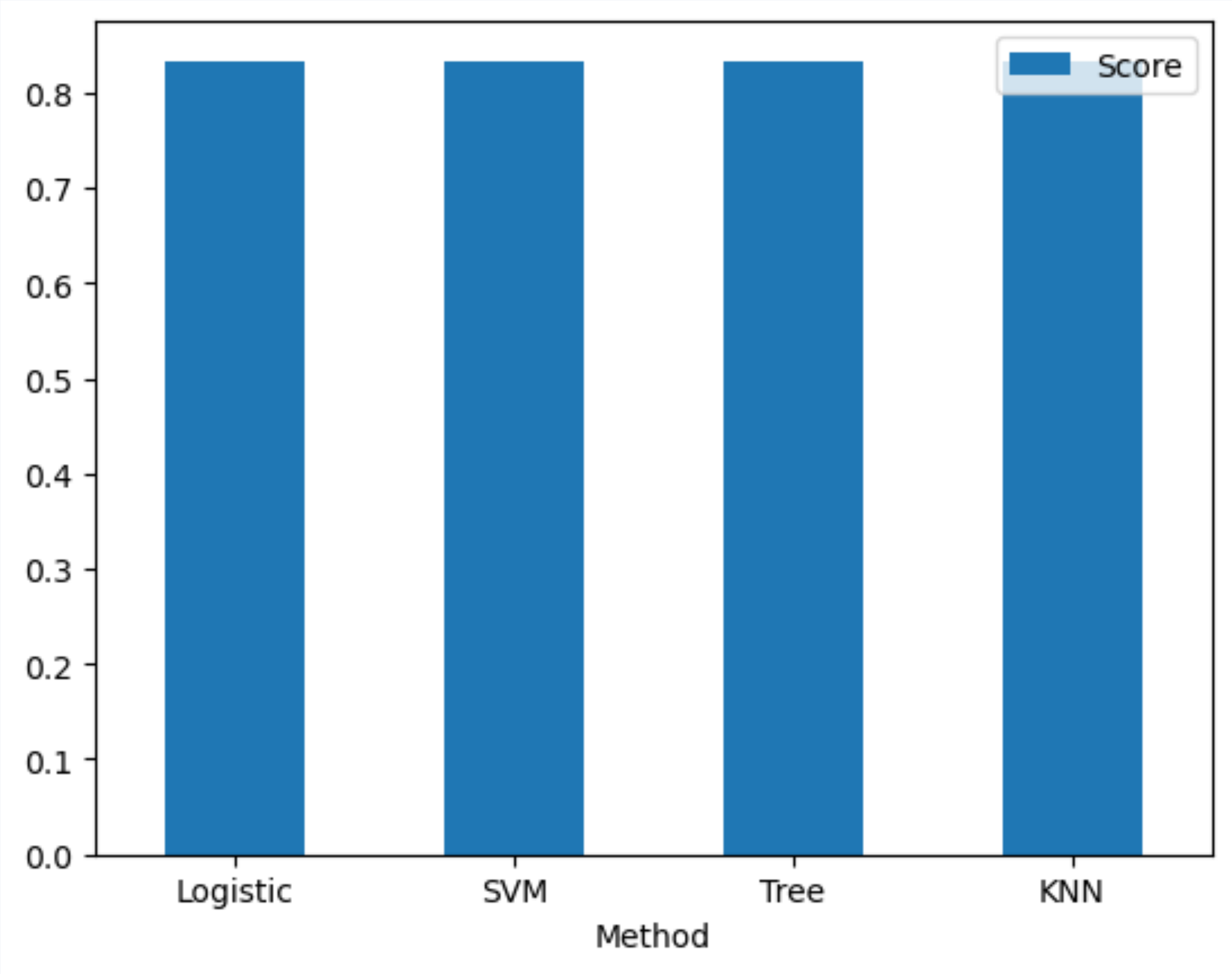
- Most success are between 2K and 6 K
- v1.0 did not operate in that range
- FT and B4 were more successful than v1.1 in that range

Section 5

Predictive Analysis (Classification)

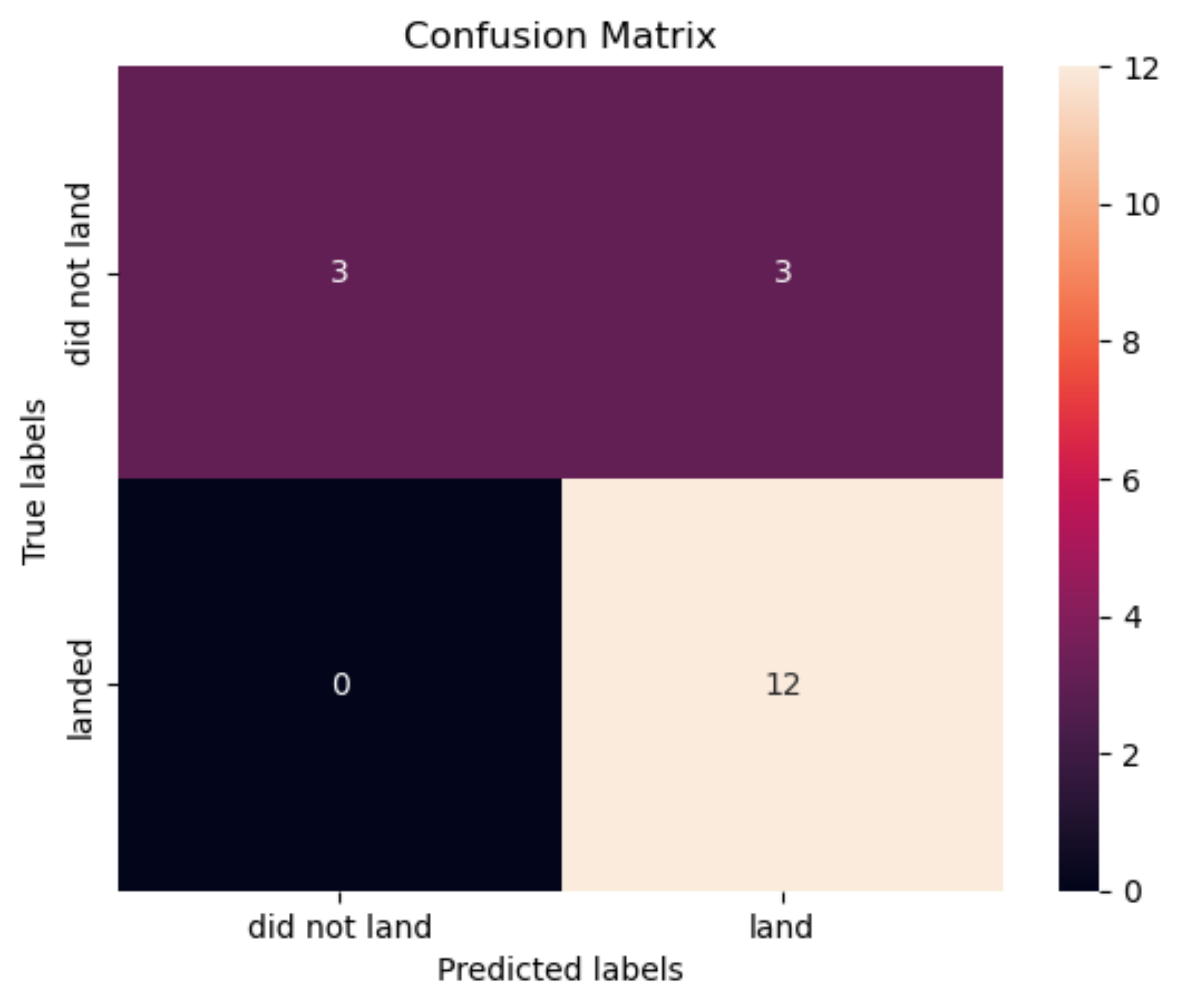
Classification Accuracy

- All the models has the same classification accuracy of 83%



Confusion Matrix

- The model has predicted all the successful outcomes
- Predicted only half the failed outcomes
- If we predict a crash there will be a crash
- If we predict a landing, there is 75% it will land



Conclusions

- The results improve with time
- Lower payloads are correlated with higher fail rates
- KSC LC-39A is the most successful launch site
- We can predict the outcome with an accuracy of 83%
- If we predict a crash, it will probably crash
- If we predict a successful land, the success rate is 75%

Thank you!