Winning Space Race with Data Science

Christopher Garner
October 2021

Outline

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

Executive Summary

	Method	Score
0	logreg	0.666667
1	SVM	0.833333
2	tree	0.791667
3	KNN	0.861111

We have gathered cleaned and analyzed data from SpaceX to reveal that we can predict the likelihood of a successful launch with an 86% success rate by using k nearest neighbor

Introduction

We want to be able to predict how landing performance will be effected by orbit, payload, launch site, landing pad, rocket type, and reused status.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - API calls and web scraping
- Perform data wrangling
  - Success / Failure binary created
- 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

```
def getBoosterVersion(data):
    for x in data['rocket']:
        response = requests.get("https://api.spacexdata.com/v4/rockets/"+str(x)).json()
        BoosterVersion.append(response['name'])
```

Data is collected by making a series of API calls and inserting the results into Data frames.

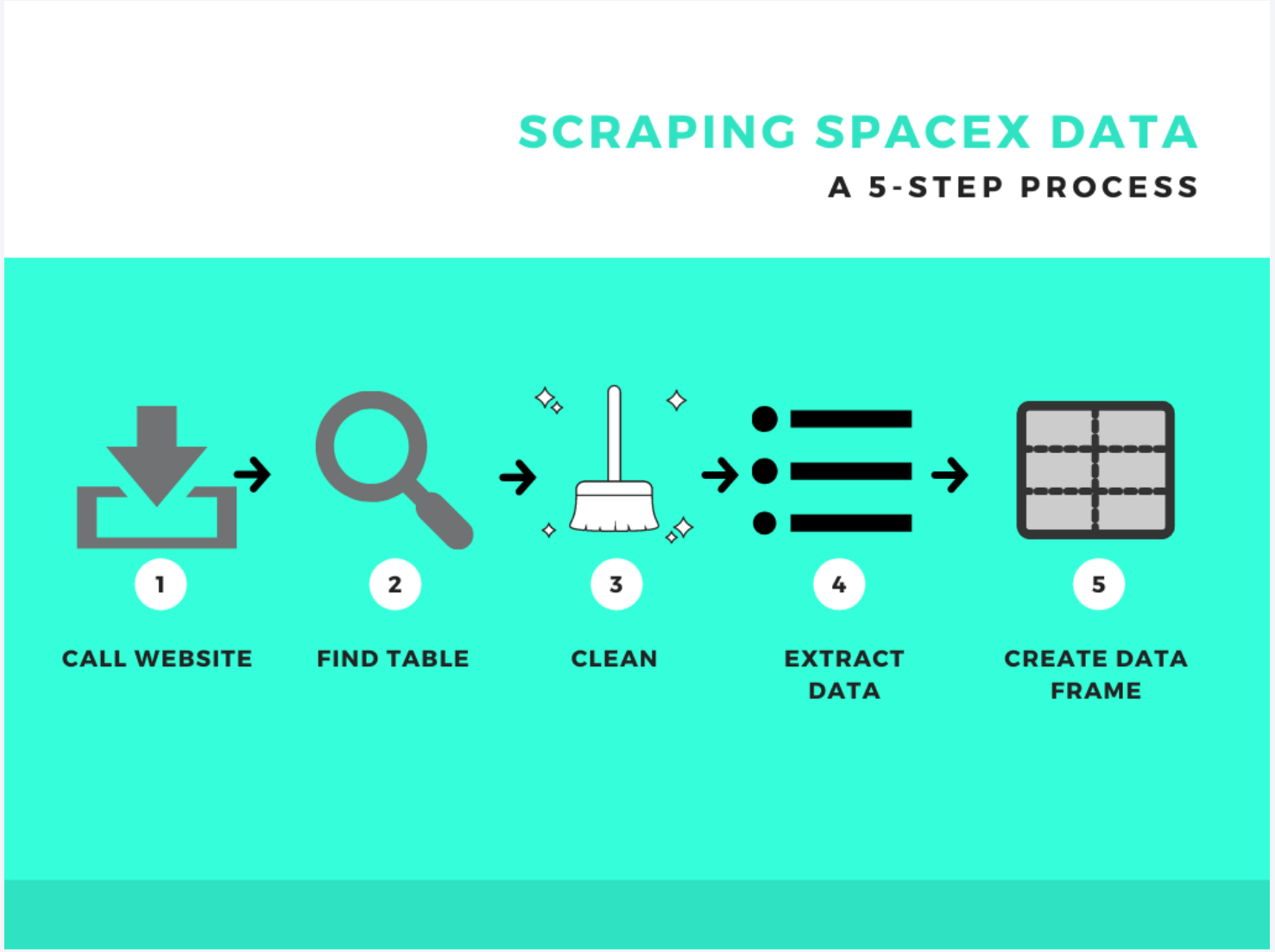
These data frames are then cleaned for the information that we want and formatted.

Data Collection – SpaceX API

https://github.com/undeadcycle/Capstone/blob/main/Data%20Collection.ipynb

Data Collection - Scraping

https://github.com/undeadcycle/Capstone/blob/main/Web_Scrape.ipynb



Data Wrangling

In order to predict a successful outcome, we need to know what a successful outcome is. There are seven possible outcomes in the data that we currently have. If we classify all outcomes with “false” or “none” as a negative outcome, we can then assign everything else a positive outcome; leaving us with a simple one or zero representation.

https://github.com/undeadcycle/Capstone/blob/main/Data%20Wrangling.ipynb

EDA with Data Visualization

Comparisons presented:

- Pay load Mass VS Flight Number (Scatter)
- Launch Site VS Flight Number (Scatter)
- Launch Site VS Pay load Mass (Scatter)
- % Success Rate VS Orbit (Bar)
- Orbit VS Flight Number (Scatter)
- Orbit VS Pay load Mass (Scatter)
- Success rate VS Year (Line)

https://github.com/undeadcycle/Capstone/blob/main/matplotlib.ipynb

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 achieved
- 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

https://github.com/undeadcycle/Capstone/blob/main/SQL.ipynb

Build an Interactive Map with Folium

We have marked all launch sites on the map with the number of launches and which launches were success and failure. We have also drawn lines to nearby locations and calculated that distance.

https://nbviewer.org/github/undeadcycle/Capstone/blob/main/Folium%280%29.ipynb

Build a Dashboard with Plotly Dash

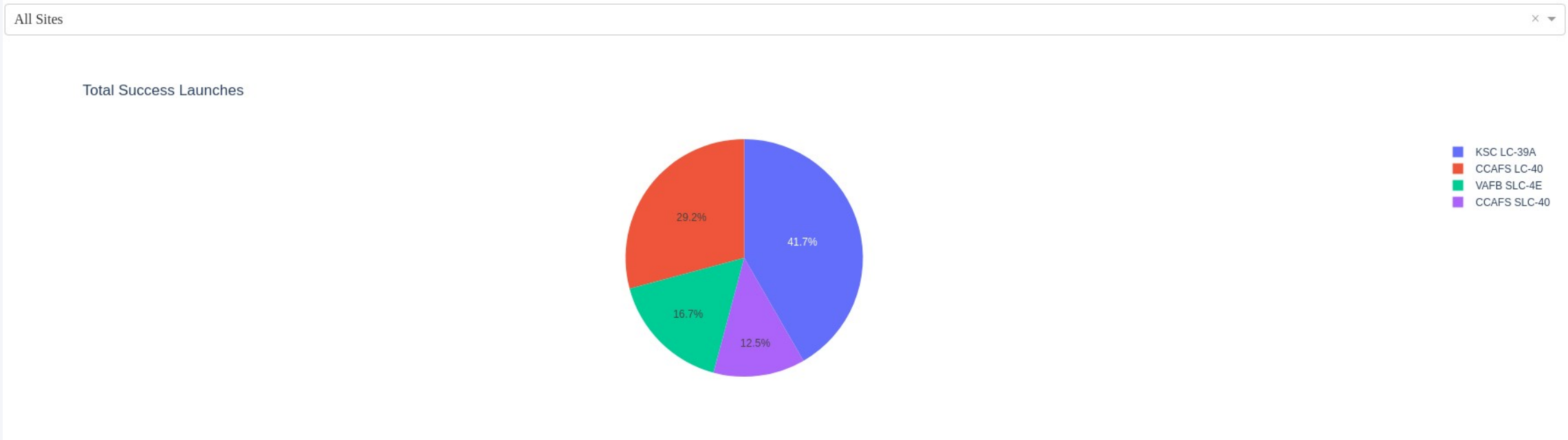
We have created an interactive pie chart that can be modified based on launch site. We also have a success vs payload which can be tuned by site and payload range.

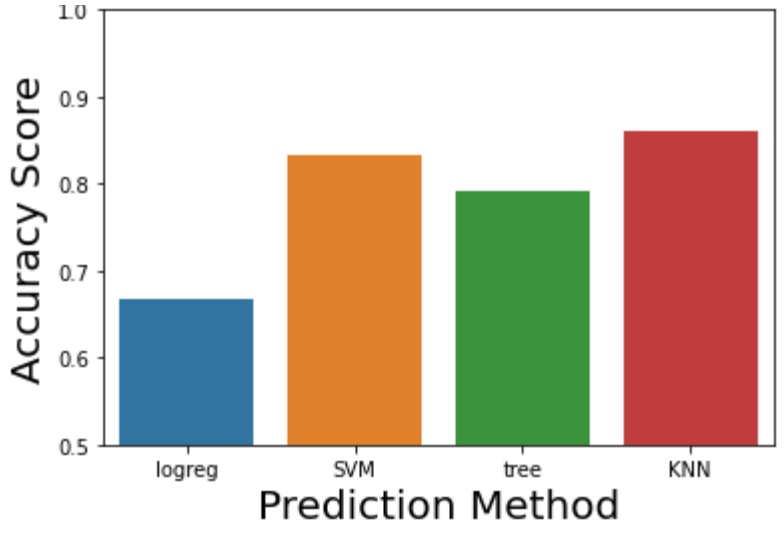
https://github.com/undeadcycle/Capstone/blob/main/dash1.ipynb

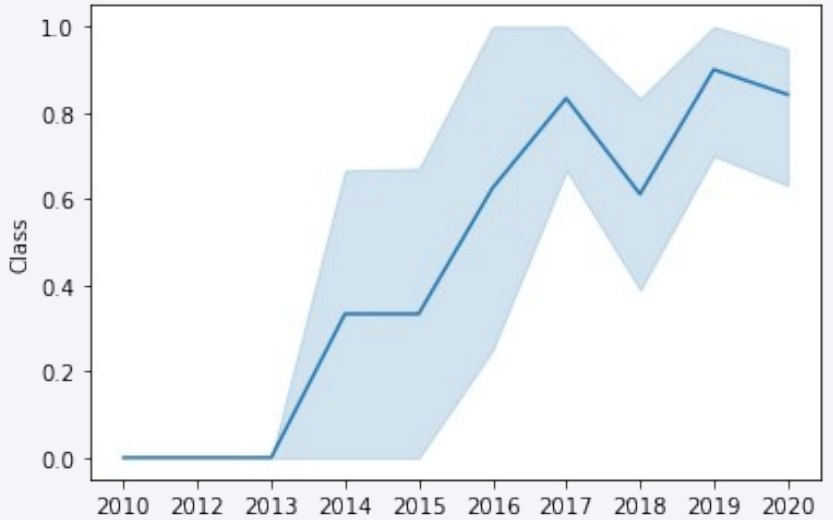
Predictive Analysis (Classification)

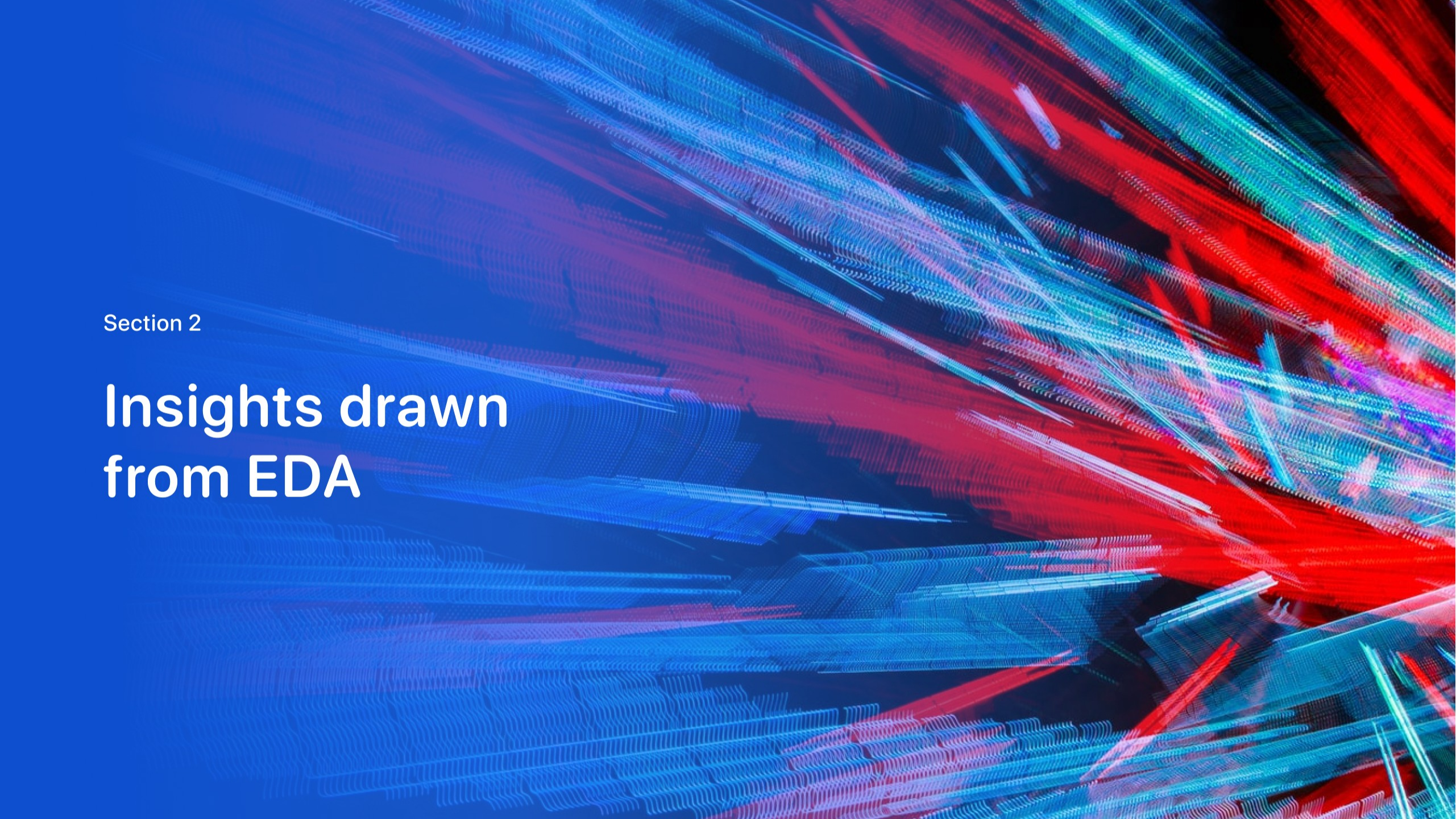
https://github.com/undeadcycle/Capstone/blob/main/ML.ipynb

Results


All Sites
Total Success Launches
29.2%
41.7%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40

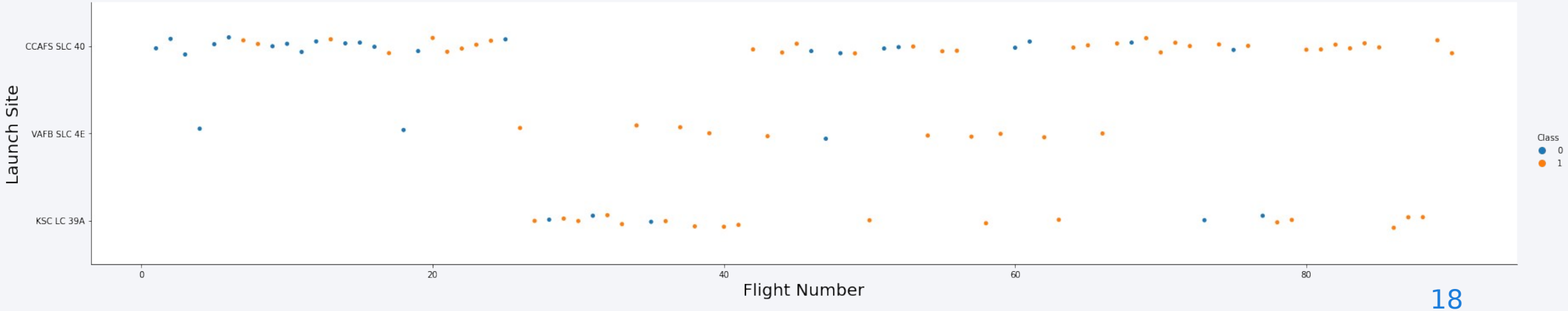

Accuracy Score
1.0
0.9
0.8
0.7
0.6
0.5
logreg
SVM
tree
KNN
Prediction Method


Class
1.0
0.8
0.6
0.4
0.2
0.0
2010
2012
2013
2014
2015
2016
2017
2018
2019
2020


Section 2
Insights drawn from EDA

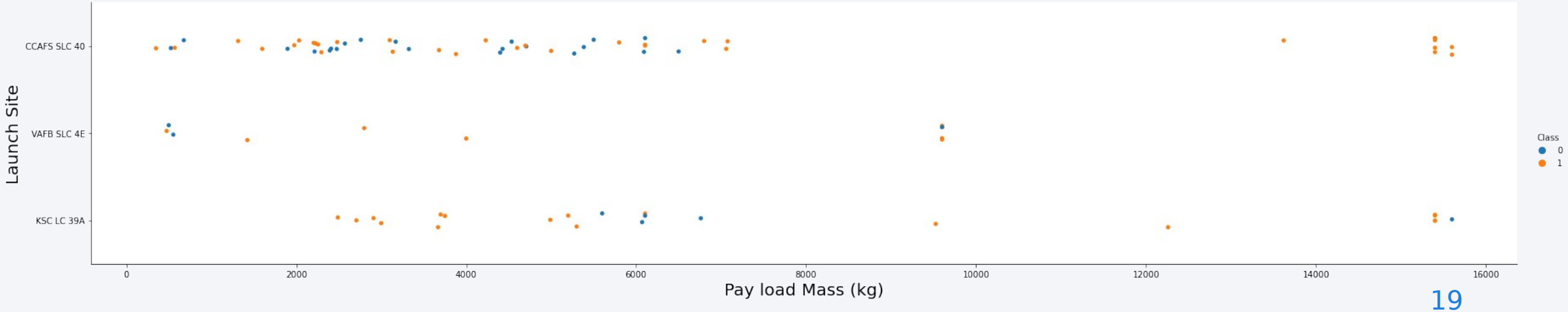
Flight Number vs. Launch Site

It looks as though CCAFS SLC 40 is the main launch site; but for launches around the 30-40 mark some resources may have been diverted to KSC LC 39A



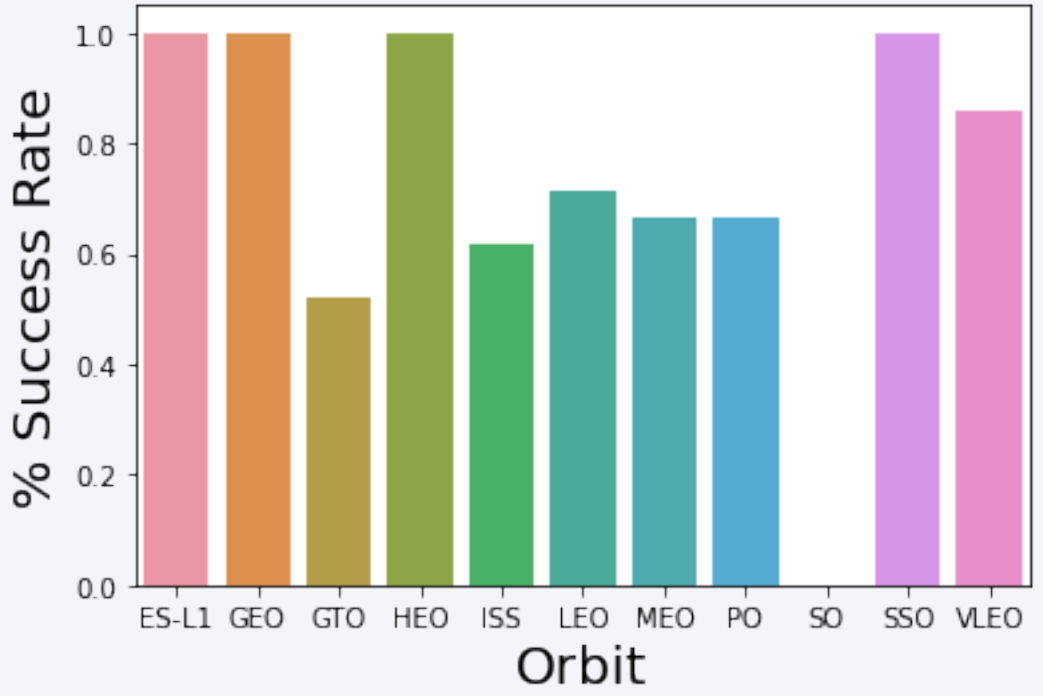
Payload vs. Launch Site

- When the pay load is between 8000kg and 14,000kg site VAFB SLC 4E seems to be preferred
- This range also seems to have fewer launches than the ranges above or below



Success Rate vs. Orbit Type

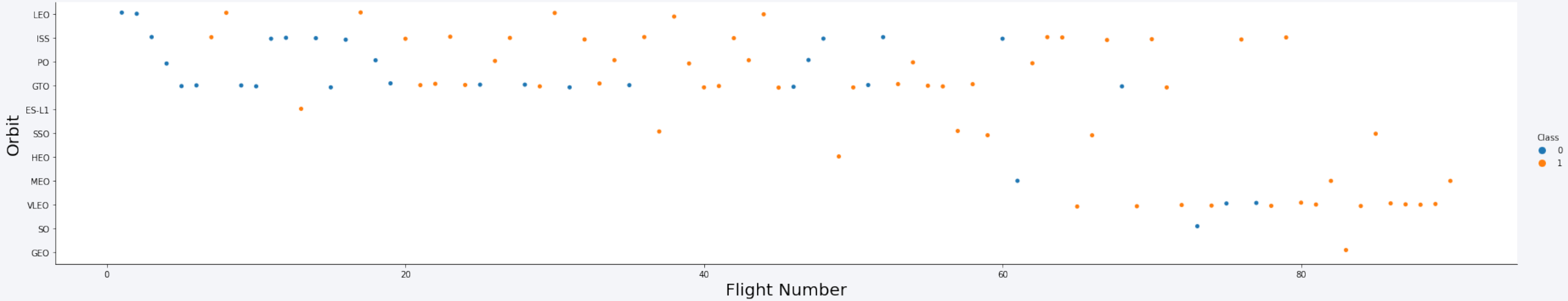
- We can see 4 orbits with a 100% success rate
- We also see 1 orbit with a 0% success rate
- Could the difficulty of each orbit type be signaled by is success rate?



Flight Number vs. Orbit Type

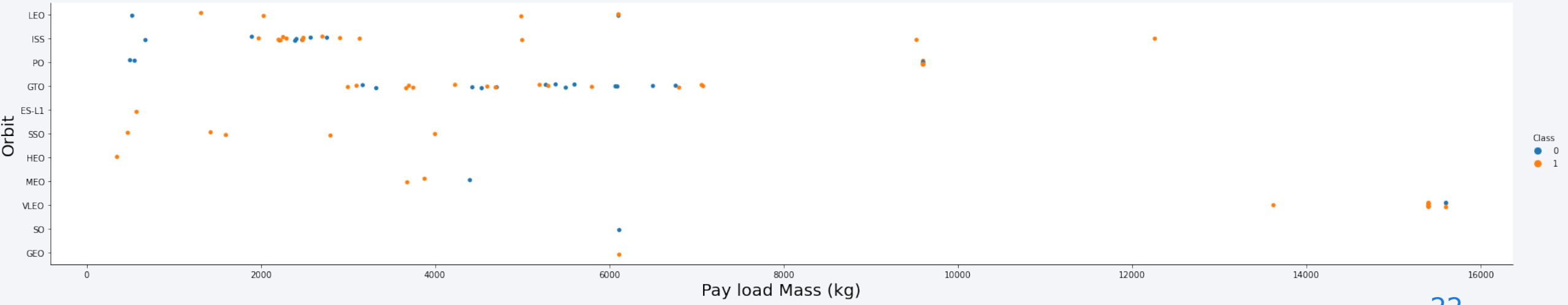
- It seems as though the focus was on 4 orbits, especially towards the beginning

- VLEO seems to be the new interest



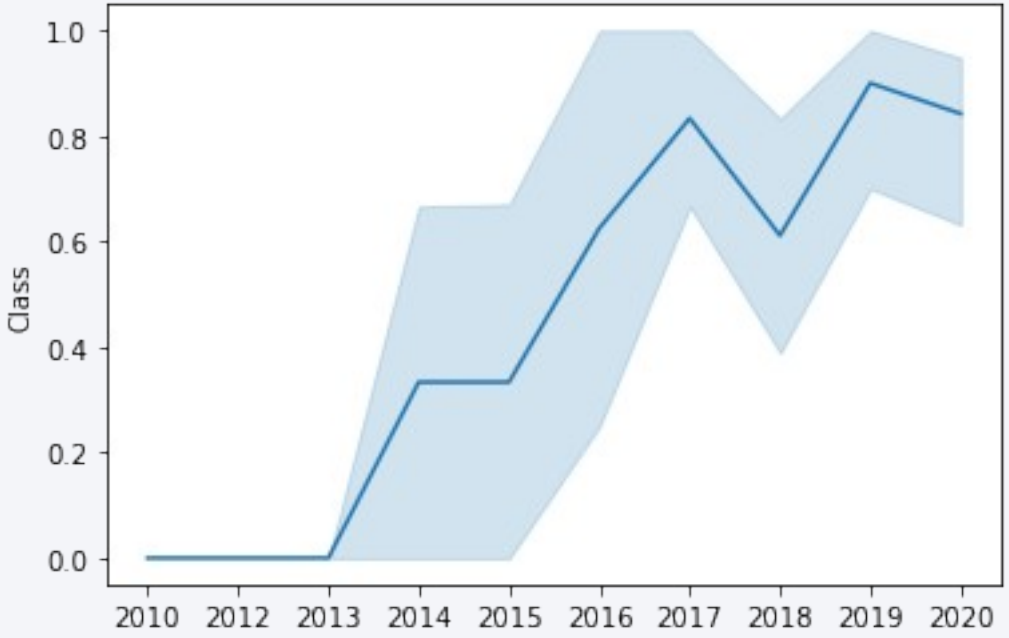
Payload vs. Orbit Type

- Though this does not show time; They may have been consistently pushing the weight limits on orbits like ISS and GTO



Launch Success Yearly Trend

- With the exception of 2018 and 2020 the trend is upward and with the number of orbits having 100% success I would expect this to continue to climb



All Launch Site Names

- The names of 4 launch sites

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with the string 'CCA'

	index	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	0	2010-04-06 00:00:00	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	1	2010-08-12 00:00:00	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	2	2012-05-22 00:00:00	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	3	2012-08-10 00:00:00	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	4	2013-01-03 00:00:00	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 carried by boosters from NASA was 111,268 kg

	SUM(PAYLOAD_MASS__KG_)
0	111268

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 2534.666667 kg

	AVG(PAYLOAD_MASS__KG_)
0	2534.666667

First Successful Ground Landing Date

- The first successful landing outcome on ground pad was 12/22/2015

	MIN(Date)
0	2015-12-22 00:00:00

Successful Drone Ship Landing with Payload between 4000 and 6000

Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are:

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

	Booster_Version
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Success

	COUNT(*)
0	100

Failure

	COUNT(*)
0	1

Boosters Carried Maximum Payload

Boosters which have carried the maximum payload mass

	Booster_Version
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5
5	F9 B5 B1051.4
6	F9 B5 B1049.5
7	F9 B5 B1060.2
8	F9 B5 B1058.3
9	F9 B5 B1051.6
10	F9 B5 B1060.3
11	F9 B5 B1049.7

2015 Launch Records

Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

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

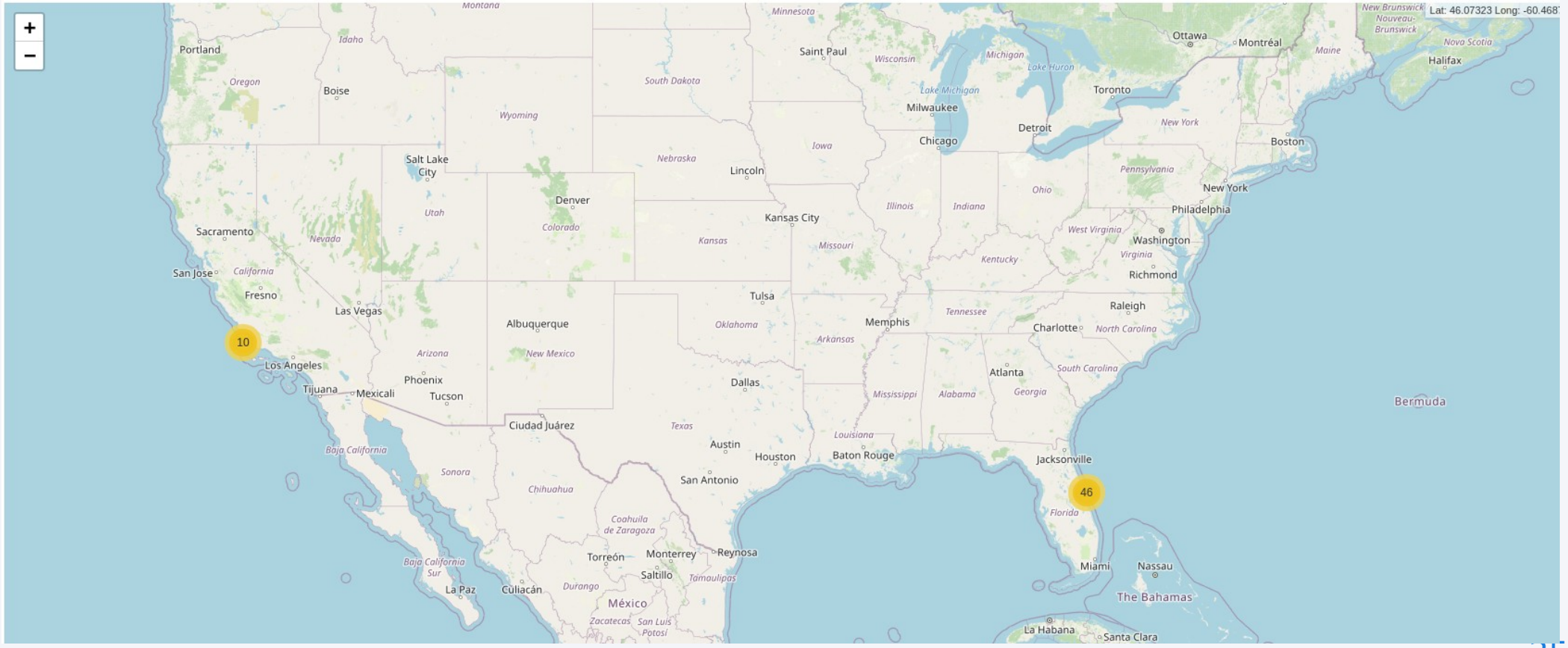
Present your query result with a short explanation here

Section 4

Launch Sites Proximities Analysis

Global Launch sites

Locations and number of launches

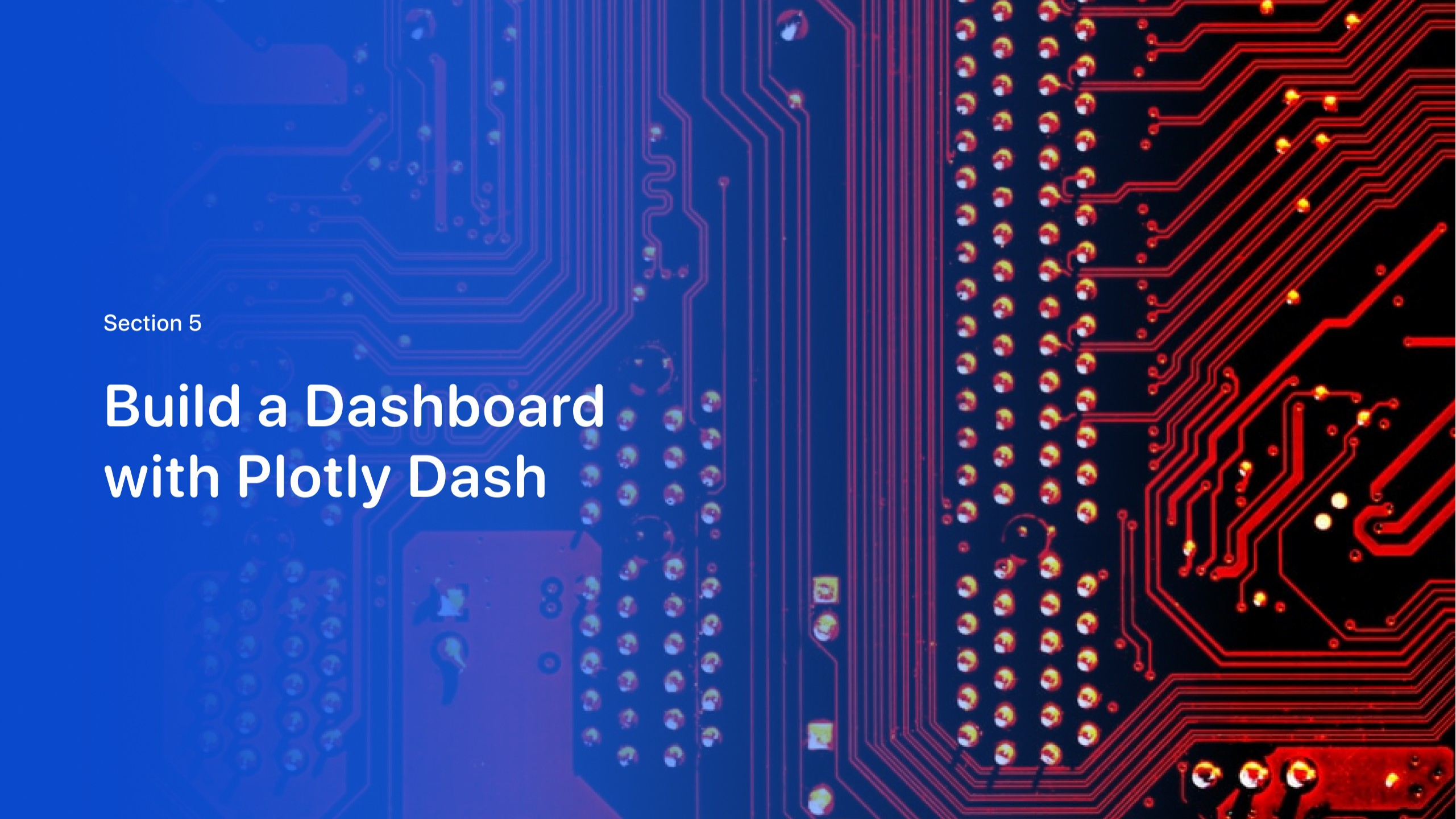


Outcomes

Show success vs failure

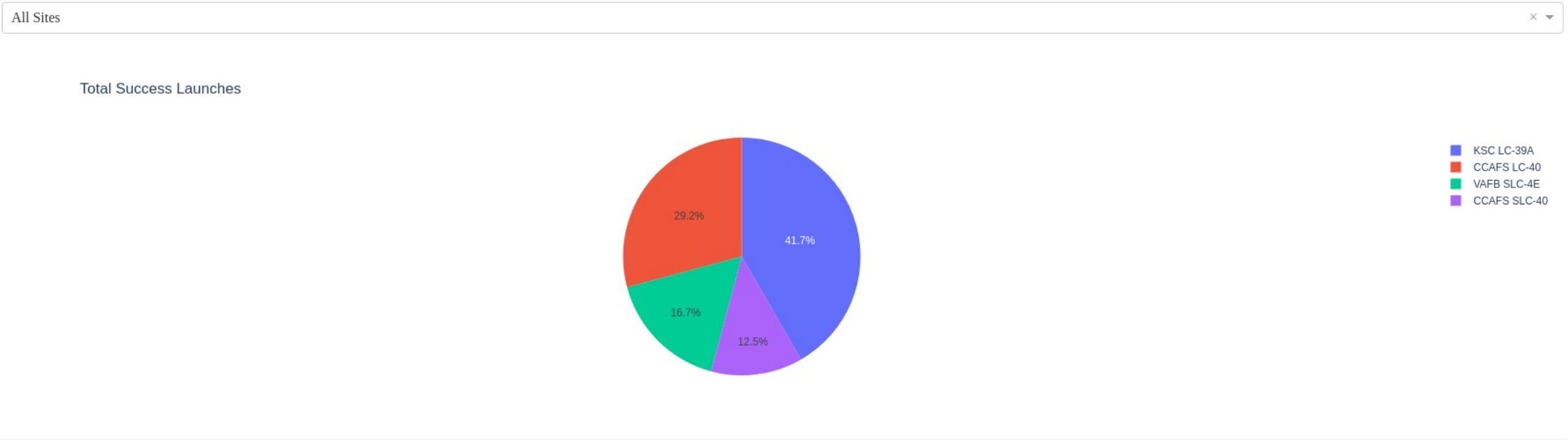
Proximities

Distance to nearby coastline


Section 5
Build a Dashboard with Plotly Dash

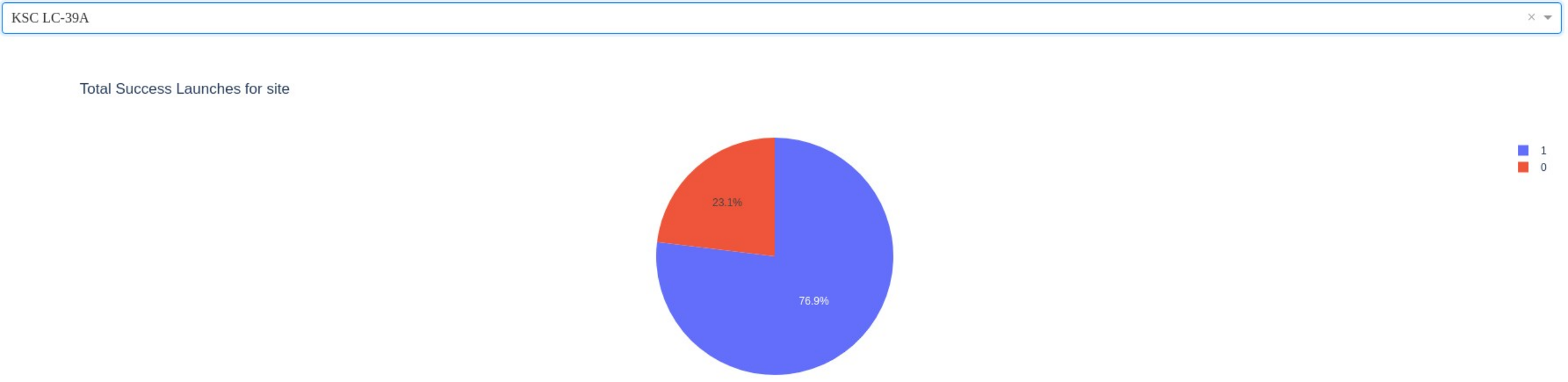
Total Successful Launches

KSC LC 39A has the most successful launces



Site with highest % of success

KSC LC 39A
This is the same site with the highest percentage overall



<Dashboard Screenshot 3>

Replace <Dashboard screenshot 3> title with an appropriate title

Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

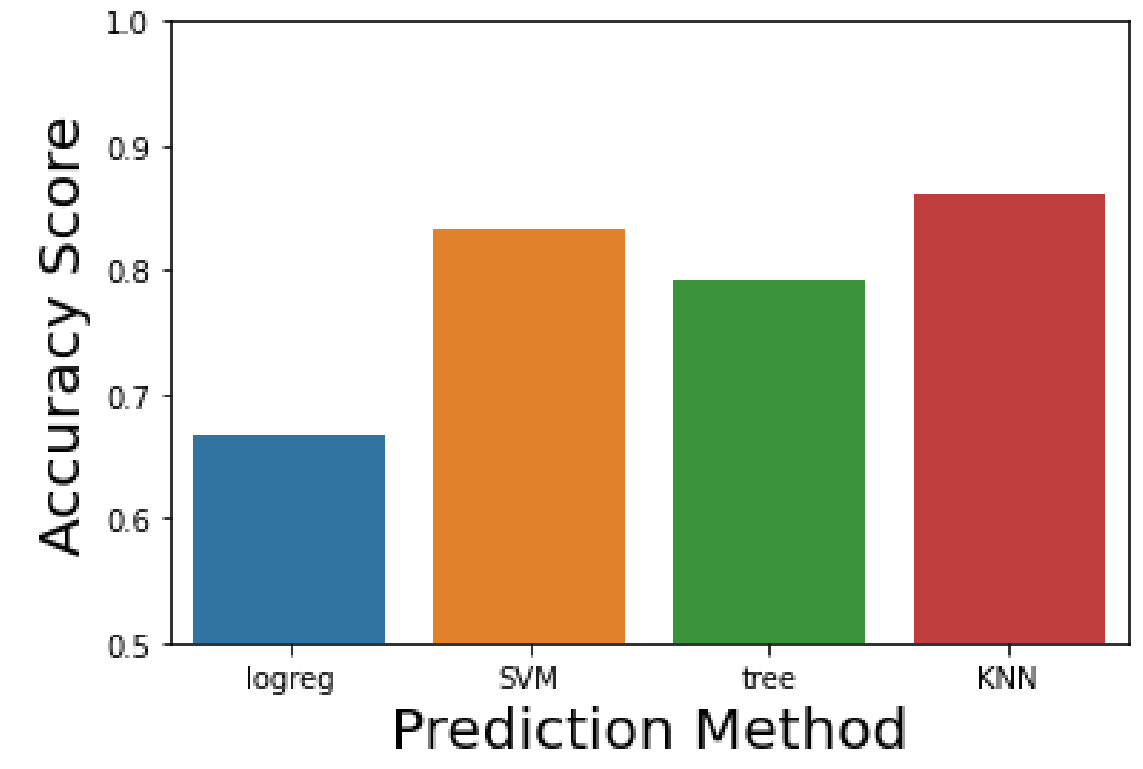
Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 6

Predictive Analysis (Classification)

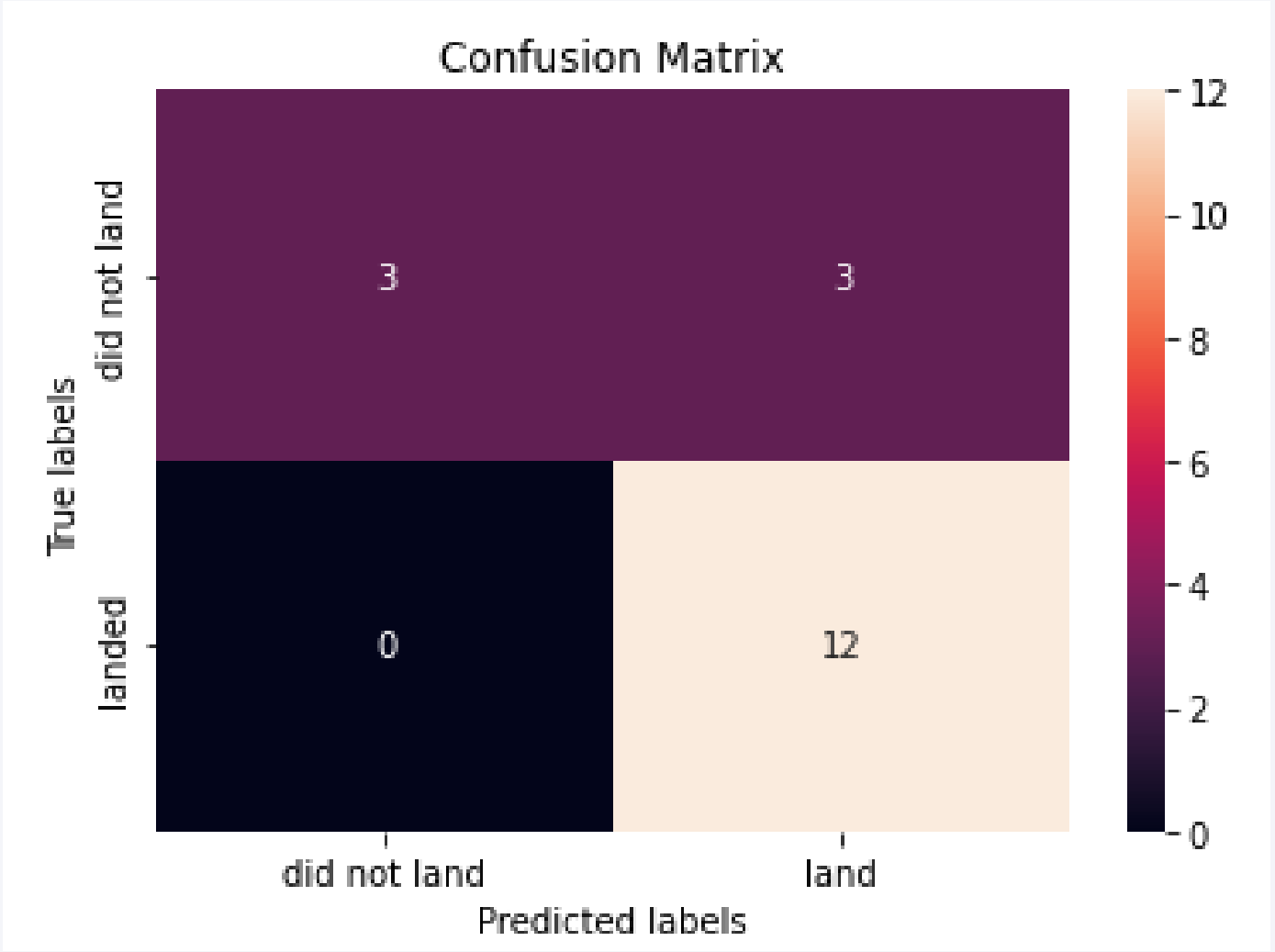
Classification Accuracy

K Nearest Neighbor has the highest classification accuracy



Confusion Matrix

- 3 False positives
- 0 False negatives



Conclusions

Now we can assess the likely success of a mission based on parameters, which gives us the ability to tune parameters to make success more likely.

Appendix

- https://github.com/undeadcycle/Capstone/blob/main/Data%20Collection.ipynb
- https://github.com/undeadcycle/Capstone/blob/main/Data%20Wrangling.ipynb
- https://nbviewer.org/github/undeadcycle/Capstone/blob/main/Folium%280%29.ipynb
- https://github.com/undeadcycle/Capstone/blob/main/ML.ipynb
- https://github.com/undeadcycle/Capstone/blob/main/SQL.ipynb
- https://github.com/undeadcycle/Capstone/blob/main/Web_Scrape.ipynb
- https://github.com/undeadcycle/Capstone/blob/main/dash1.ipynb
- https://github.com/undeadcycle/Capstone/blob/main/matplotlib.ipynb

Thank you!