IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rachel Armstrong
22/05/2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Data was:

- Extracted from SpaceX API or Wikipedia
- Wrangled into a suitable format
- Explored using:
  - Graphs
  - SQL queries
  - Interactive dashboards
- Split into training and testing data
- Created ML models using training data
- Evaluated models using test data

Results:

- All models performed equally well with 83% accuracy
- Models have false positives but not false negatives
- Limited size of dataset makes testing accurately difficult, and further data would give more confidence to results

Introduction

- SpaceX offers rocket launches at a significantly lower cost than competitors
- They are able to do this by re-using part of the rocket
- They can only reuse the rocket part if this part lands successfully
- If we can predict if a rocket part will land successfully then this can be used by competitors to identify cost savings

Section 1

Methodology

Methodology

Executive Summary

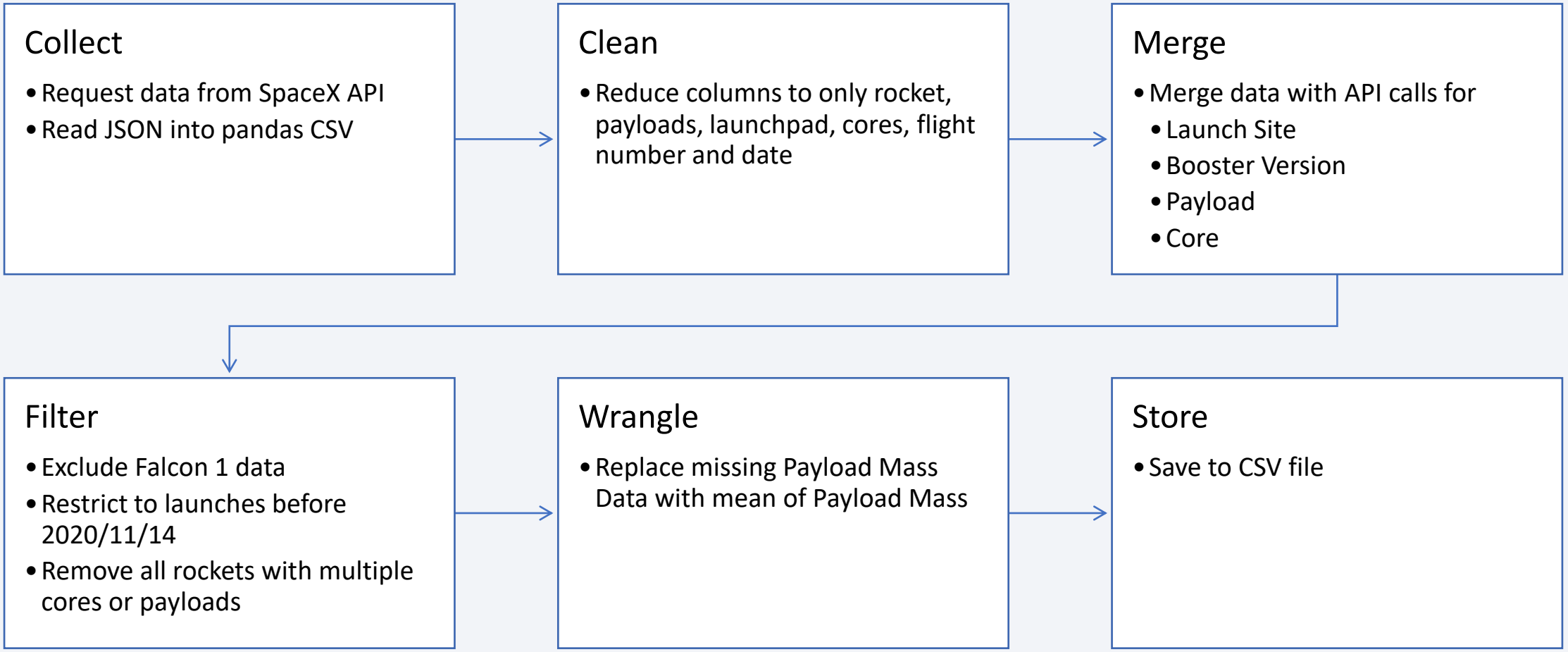
- Data collection methodology:
  - Data collected either via API or web scraping Wikipedia
- Perform data wrangling
  - Outcomes categorized as successful or unsuccessful landings
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built, tuned and evaluated classification models

Data Collection

Data was collected via two different methods:

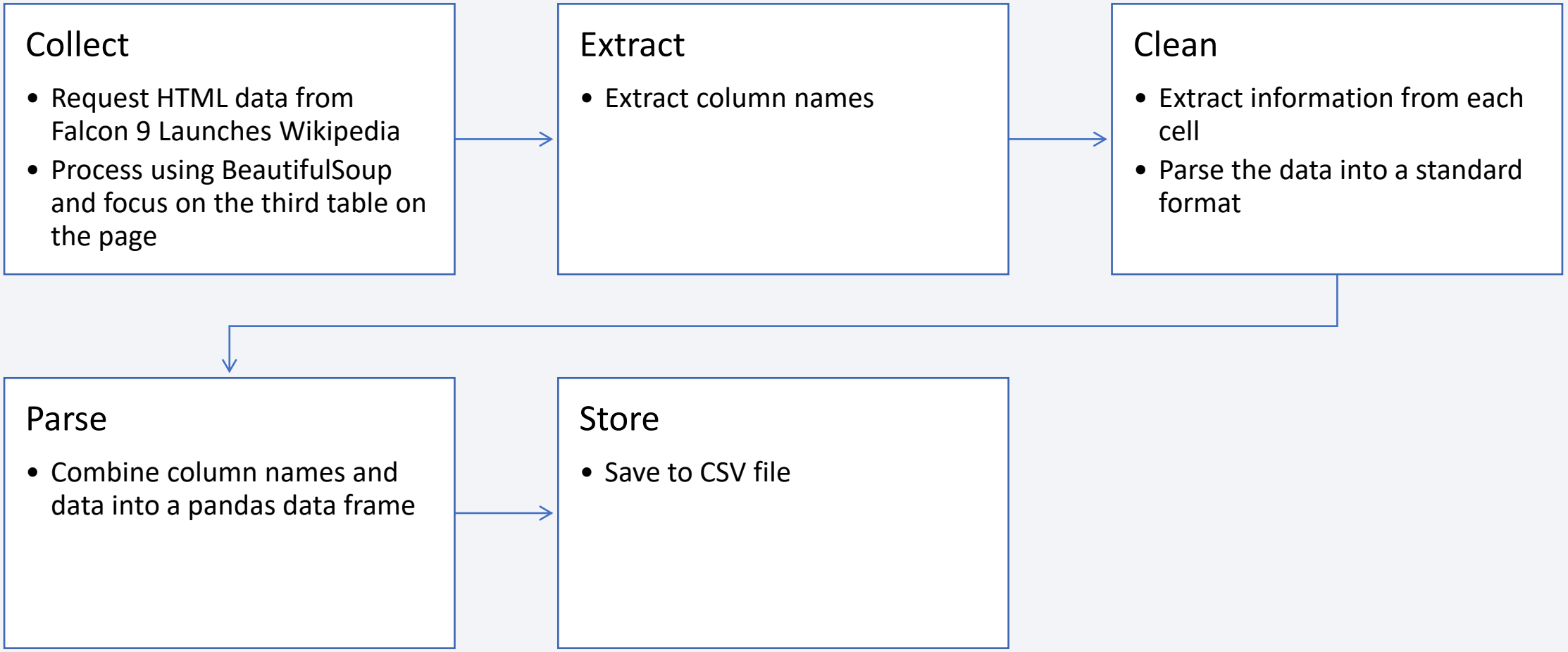
- Scraping the SpaceX Wikipedia website using BeautifulSoup
- Requesting SpaceX data directly from the SpaceX API and merging the multiple requests

Data Collection – SpaceX API



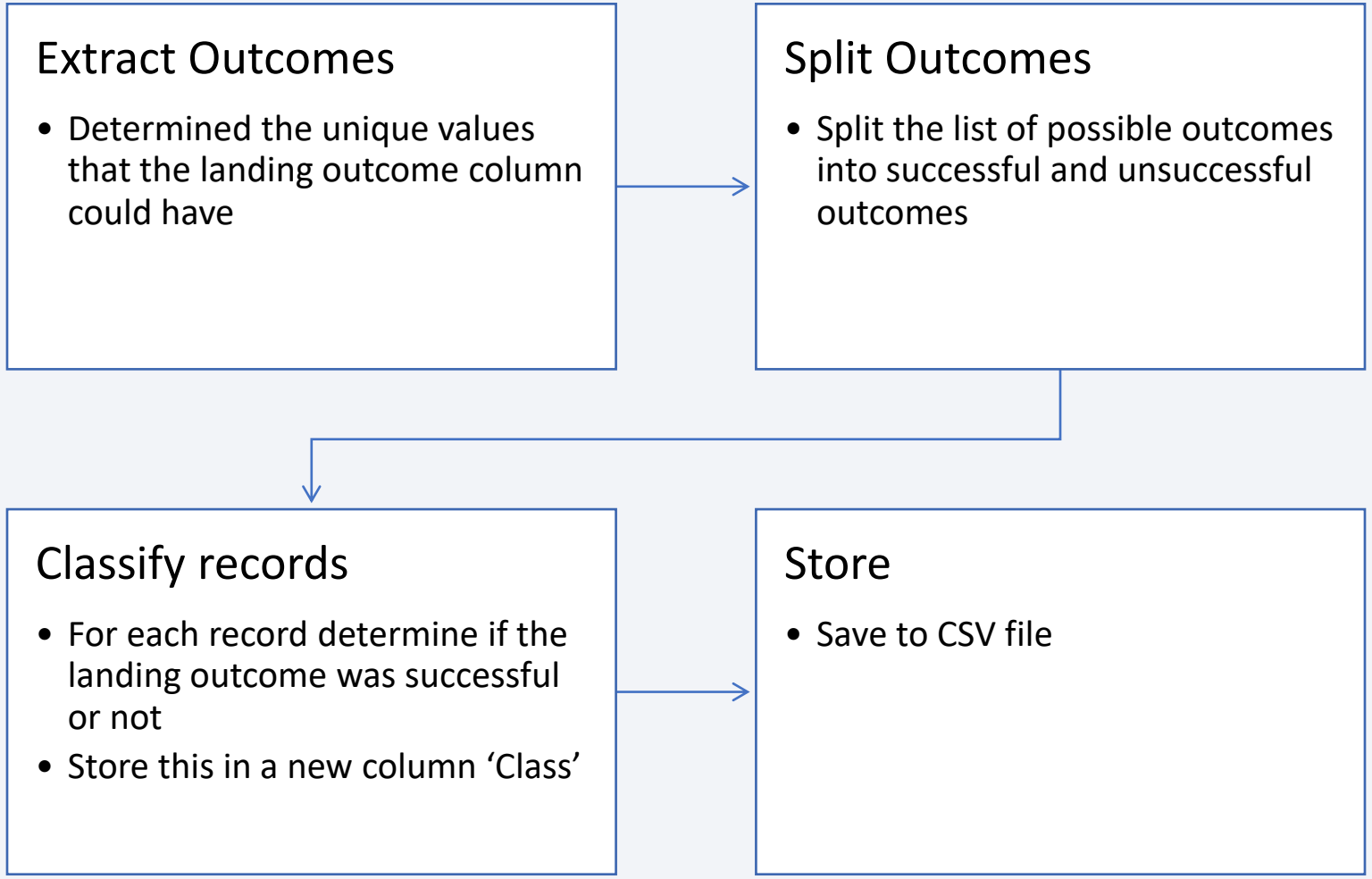
https://github.com/qwerty16/portfolio-data/blob/main/spacex/Data%20Collection%20API.ipynb

Data Collection - Scraping



https://github.com/qwerty16/portfolio-data/blob/main/spacex/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling



https://github.com/qwerty16/portfolio-data/blob/main/spacex/Data%20Wrangling.ipynb

EDA with Data Visualization

Charts were plotted to identify trends and correlations between available features.

Charts plotted:

- Flight number vs launch site
- Payload vs launch site
- Orbit type vs success rate
- Flight number vs orbit type
- Payload vs orbit type
- Success rate over time

https://github.com/qwerty16/portfolio-data/blob/main/spacex/Data%20Exploration%20using%20Matplotlib.ipynb

EDA with SQL

- Names of unique launch sites
- Launch sites which begin with CCA
- Mass of payload launched by NASA
- Average payload carried by F9 v1.1 booster
- First successful ground pad landing date
- Successful boosters used in drone ship landings with payload mass between 4000 and 6000 kg
- Count of launches aggregated by outcome
- Booster versions which have carried maximum payload mass
- Failed drone ship landings in 2015 with launch site and booster version details
- Landing outcomes between 2010-06-04 and 2017-03-20, ranked in descending order

https://github.com/qwerty16/portfolio-data/blob/main/spacex/Data%20Exploration%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Circle added to each launch site labelled with the site name
- Cluster of markers added to each site, one per launch, colour coded according to if the launch was successful or not
- Added a line between a launch site and nearby points of interest

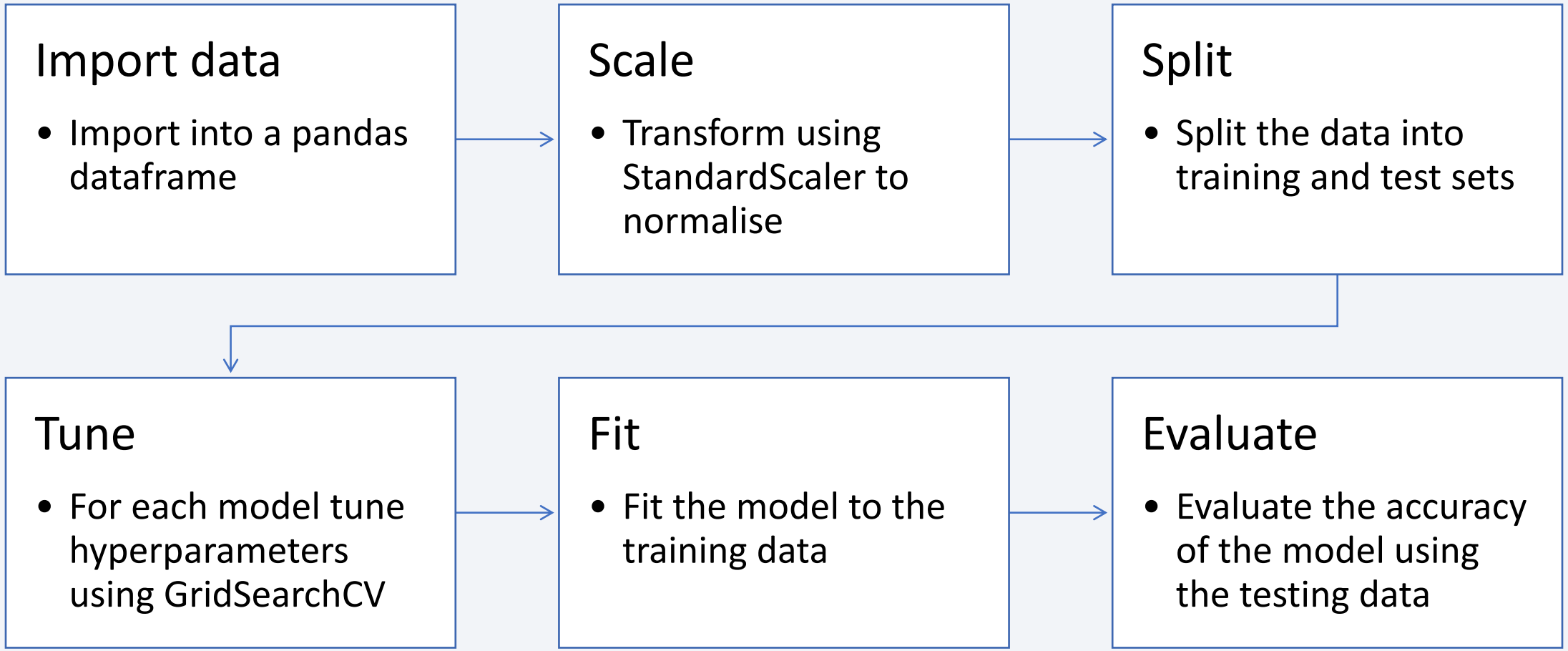
https://github.com/qwerty16/portfolio-data/blob/main/spacex/Data%20Visualisation%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Pie chart which can show successes for each site or successes vs failures for one site (dropdown to select site)
- Scatter chart showing successes vs payload mass, each point was colour coded according to its launch site. The launch site is also affected by the site dropdown. There is a slider filter which allows restriction of the range of payload masses which are included

https://github.com/qwerty16/portfolio-data/blob/main/spacex/spacex_dash_app.py

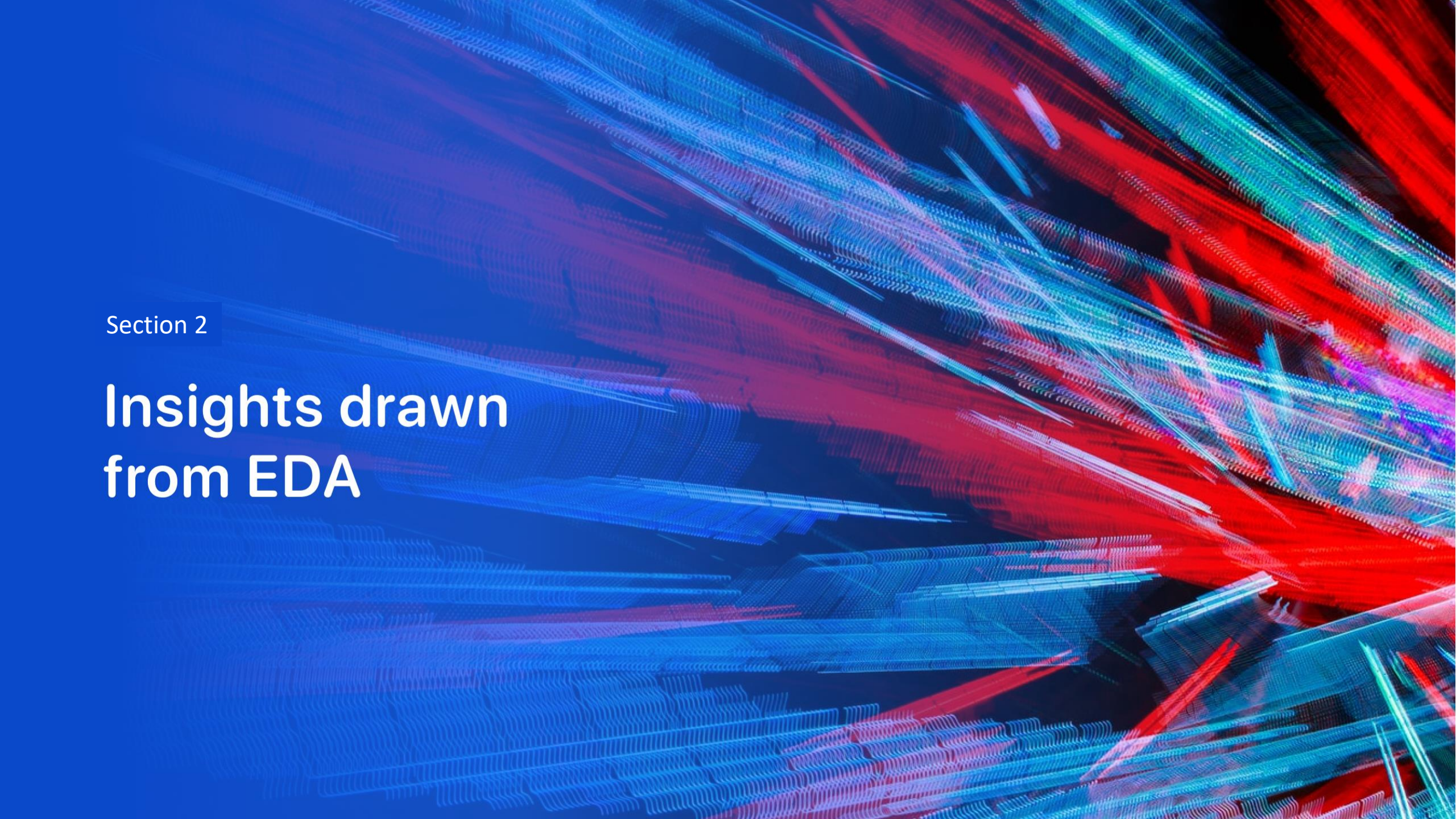
Predictive Analysis (Classification)



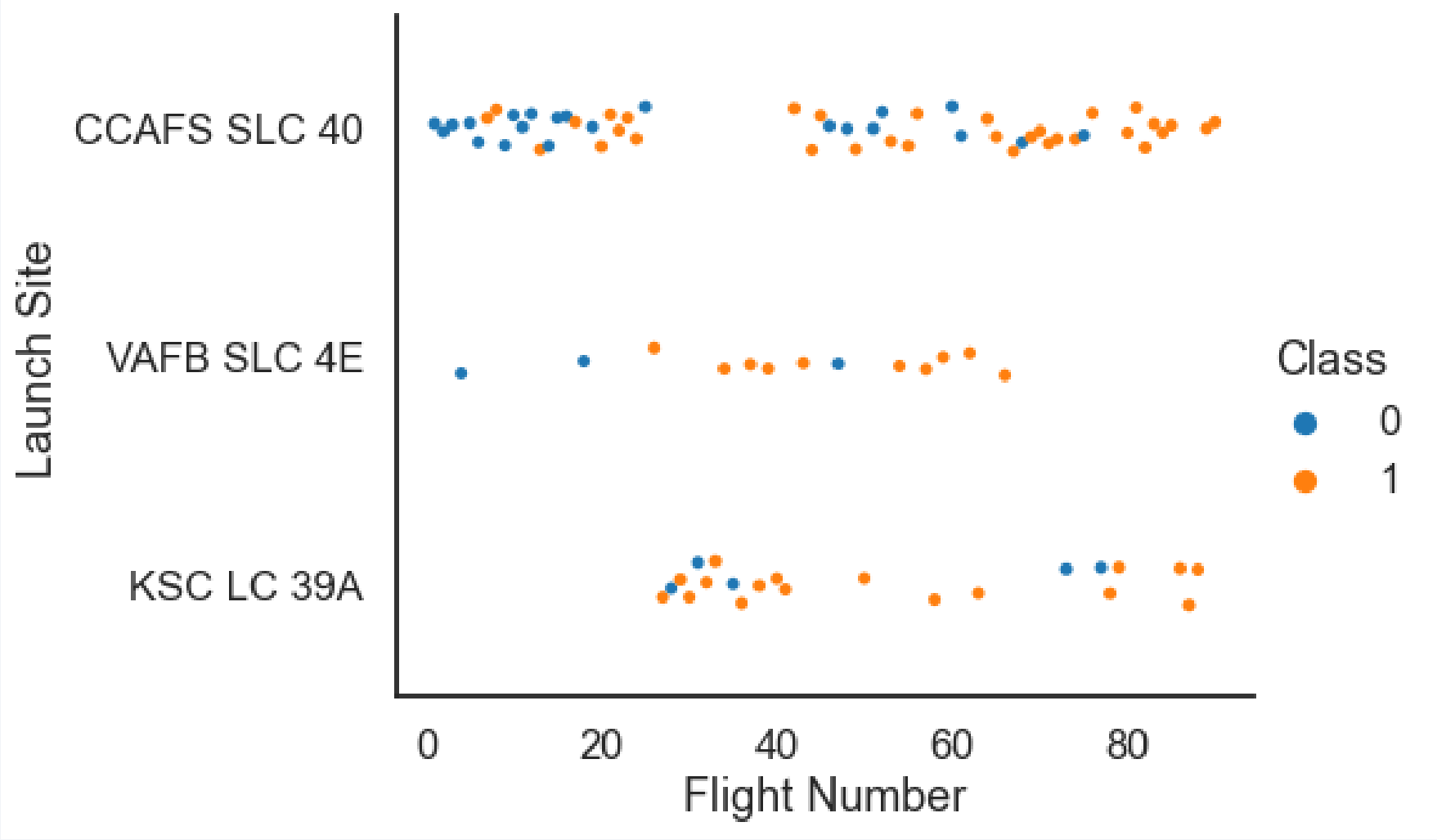
https://github.com/qwerty16/portfolio-data/blob/main/spacex/SpaceX_Machine%20Learning%20Predictions.ipynb

Results

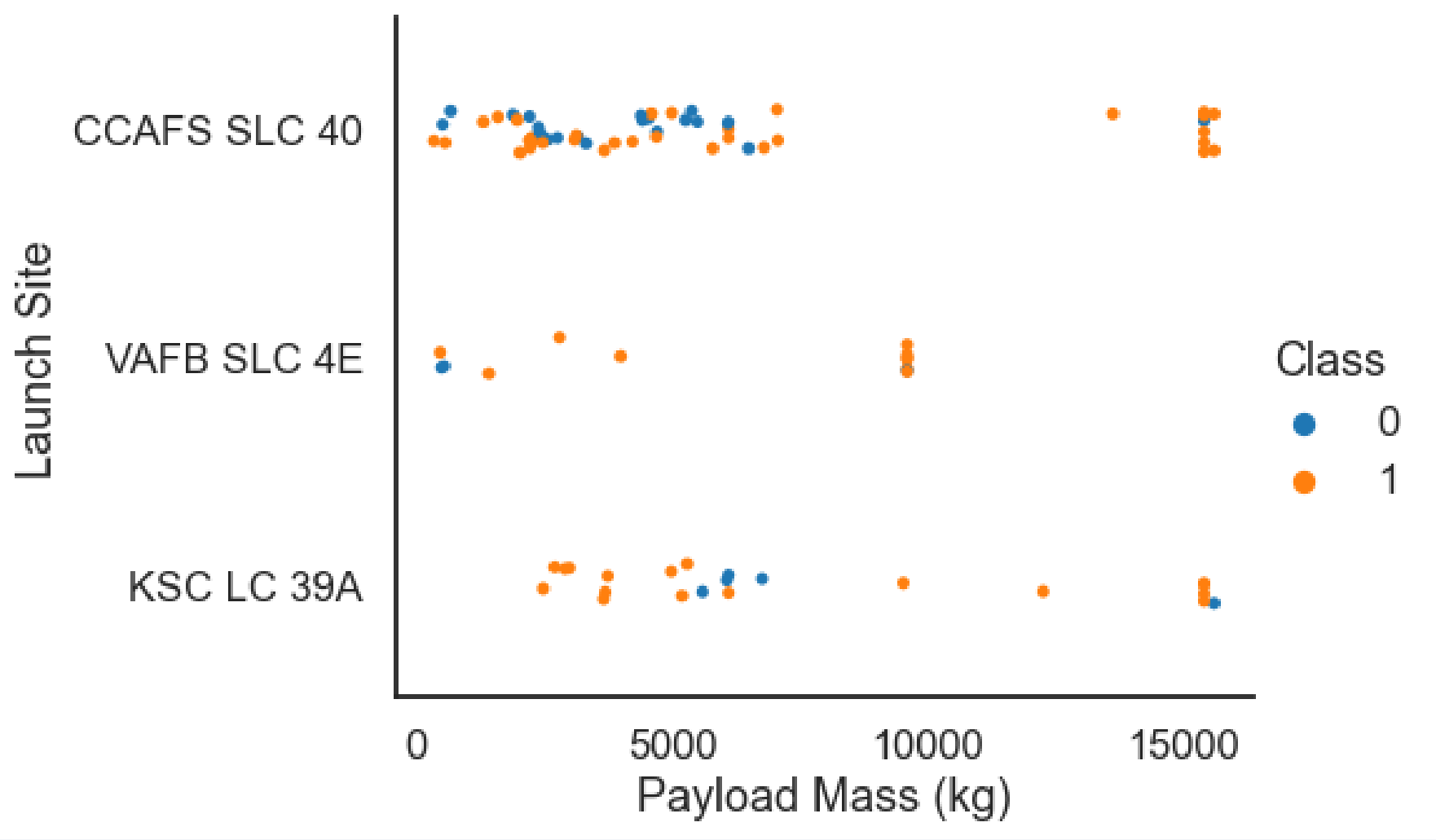
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

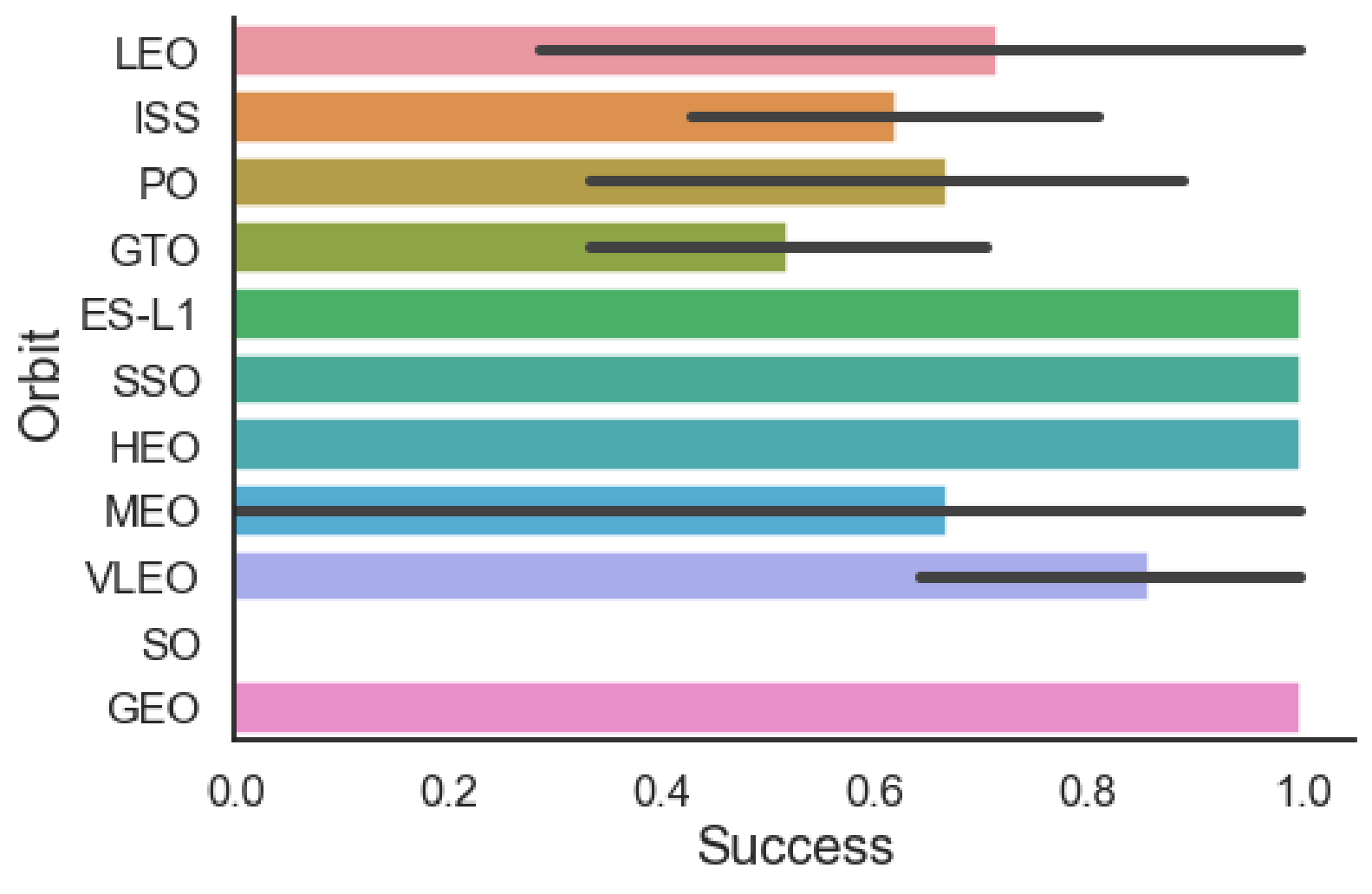
Flight Number vs. Launch Site



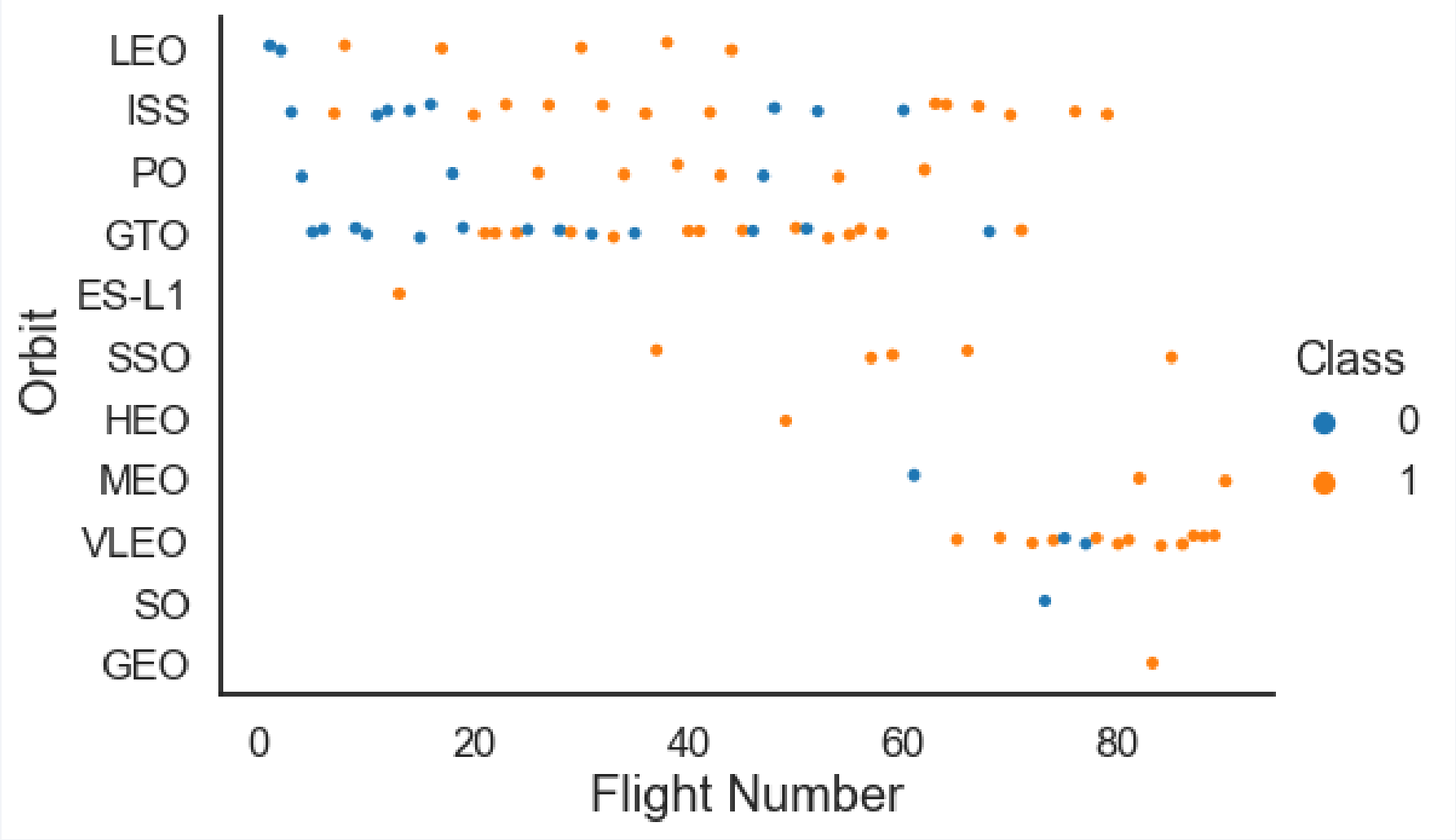
Payload vs. Launch Site



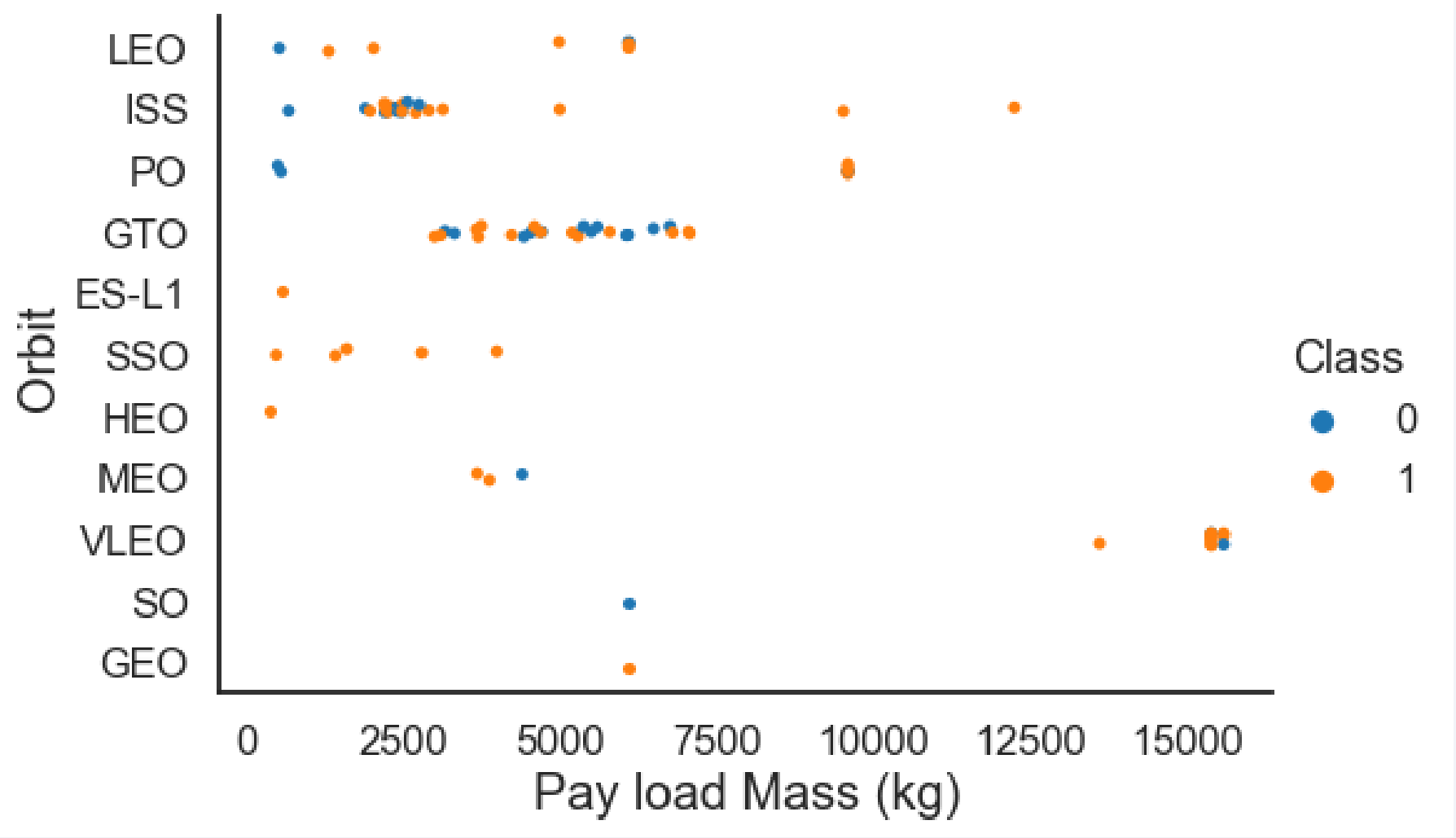
Success Rate vs. Orbit Type



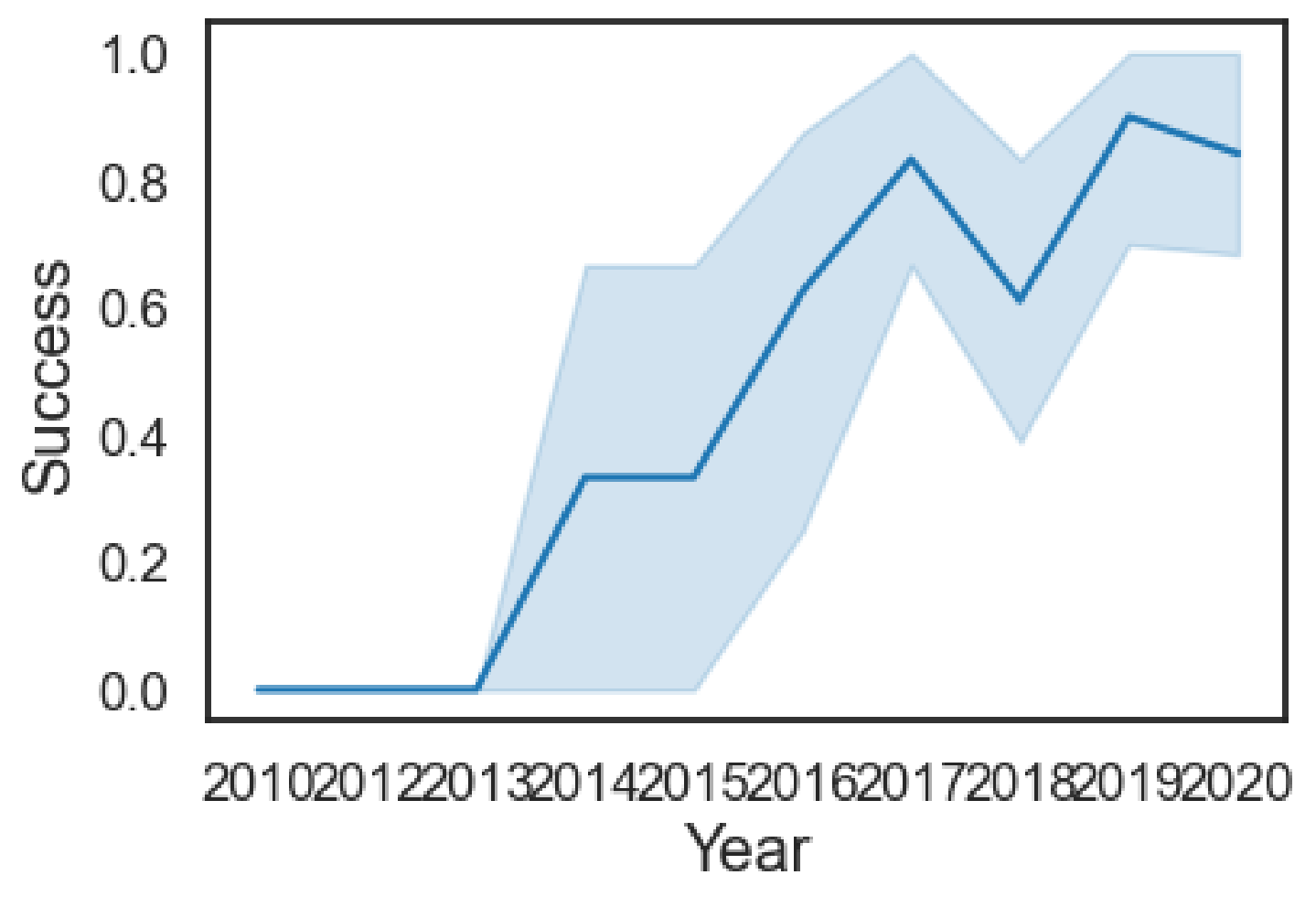
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

	Date	Time (UTC)	Booster Version	Launch Site	Payload	Payload Mass KG	Orbit	Customer	Mission Outcome	Landing Outcome
0	04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	08-12-2010	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	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
3	08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

SELECT sum(Payload_Mass_KG) as Total FROM spacex WHERE Customer = 'NASA (CRS)

	Total
0	45596.0

Average Payload Mass by F9 v1.1

```
select avg(Payload_Mass_KG) as Average from spacex where Booster_Version
like 'F9 v1.1%%'
```

	Average
0	2534.6667

First Successful Ground Landing Date

select min(Date) as First_Date from spacex where Landing_Outcome = 'Success (ground pad)'

	First_Date
0	01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

	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

	Outcome	Count
0	Success	98
1	Failure (in flight)	1
2	Success (payload status unclear)	1
3	Success	1

Boosters Carried Maximum Payload

	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

	Date	Booster_Version	Launch_Site
0	2015-01-10	F9 v1.1 B1012	CCAFS LC-40
1	2015-04-14	F9 v1.1 B1015	CCAFS LC-40

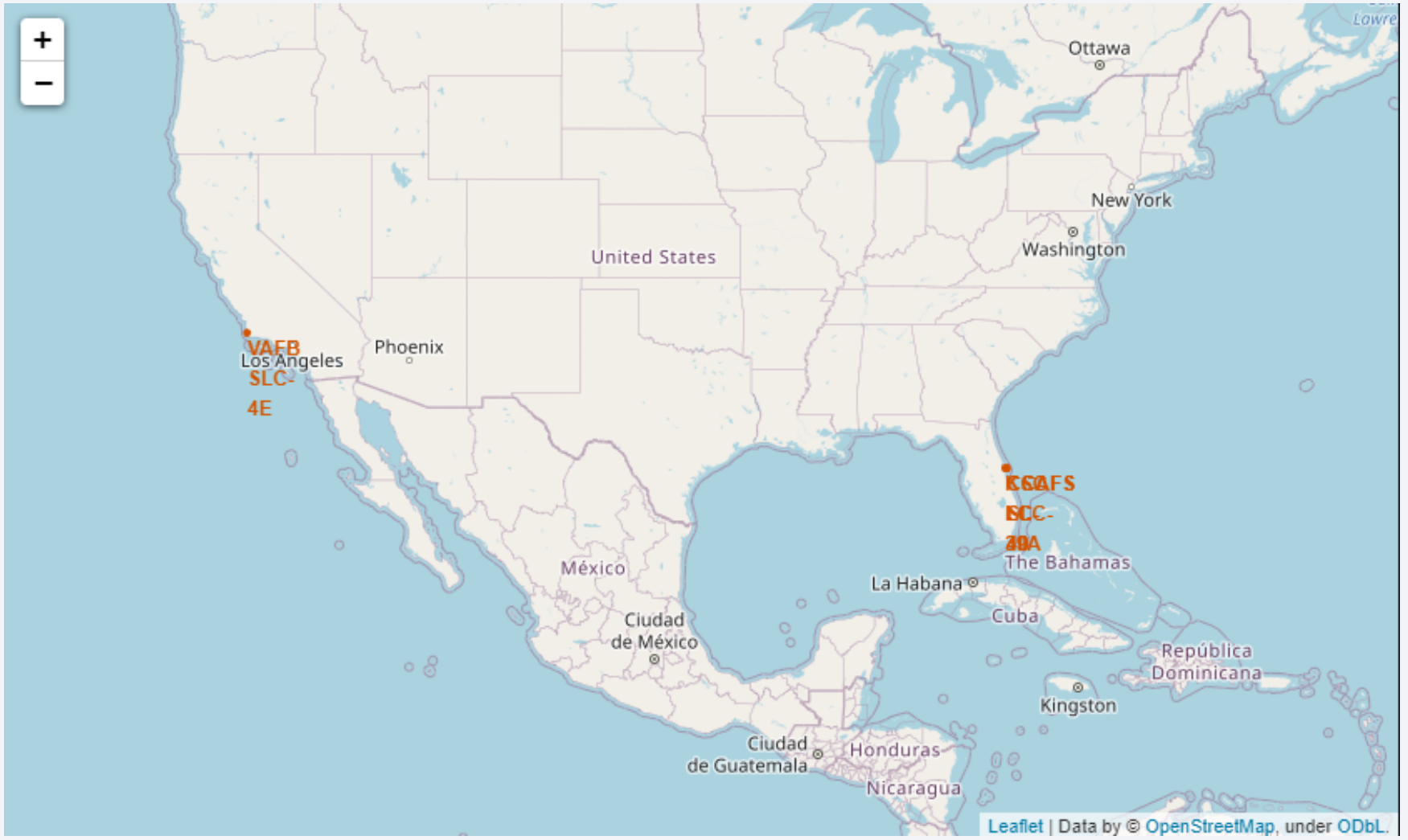
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

	Outcome	Count
0	No attempt	10
1	Failure (drone ship)	5
2	Success (drone ship)	5
3	Controlled (ocean)	3
4	Success (ground pad)	3
5	Failure (parachute)	2
6	Uncontrolled (ocean)	2
7	Precluded (drone ship)	1

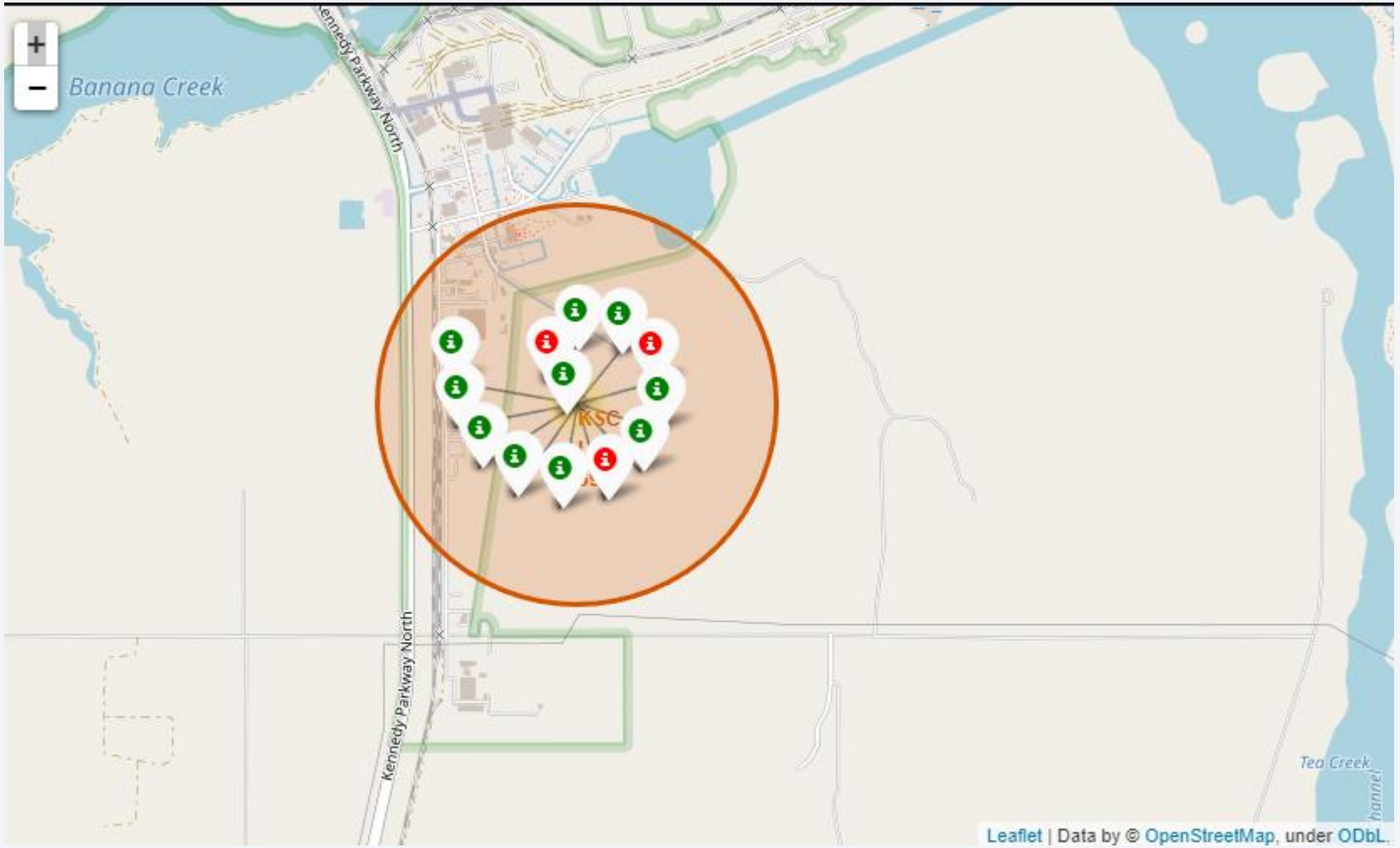
Section 3

Launch Sites Proximities Analysis

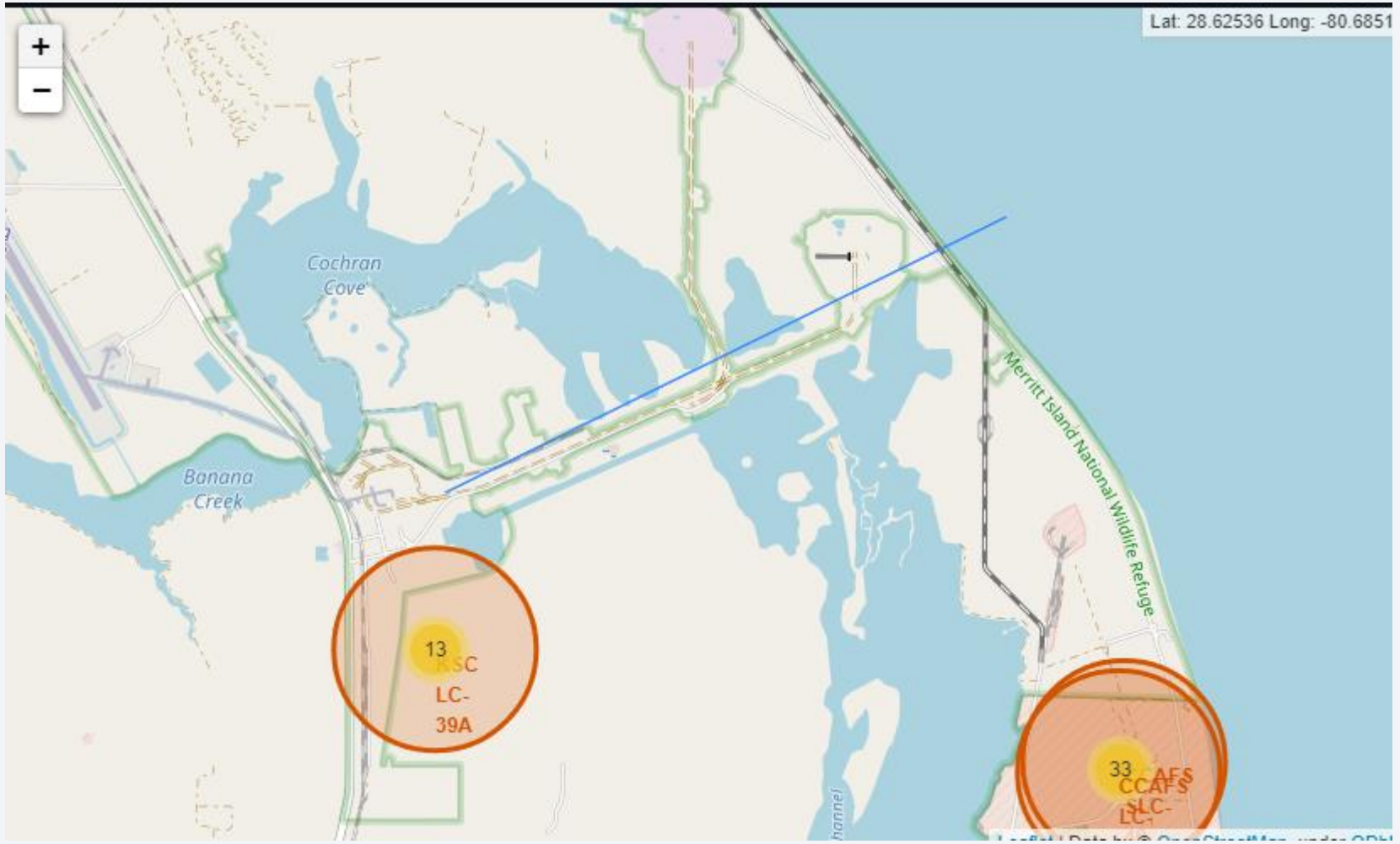
Launch Sites

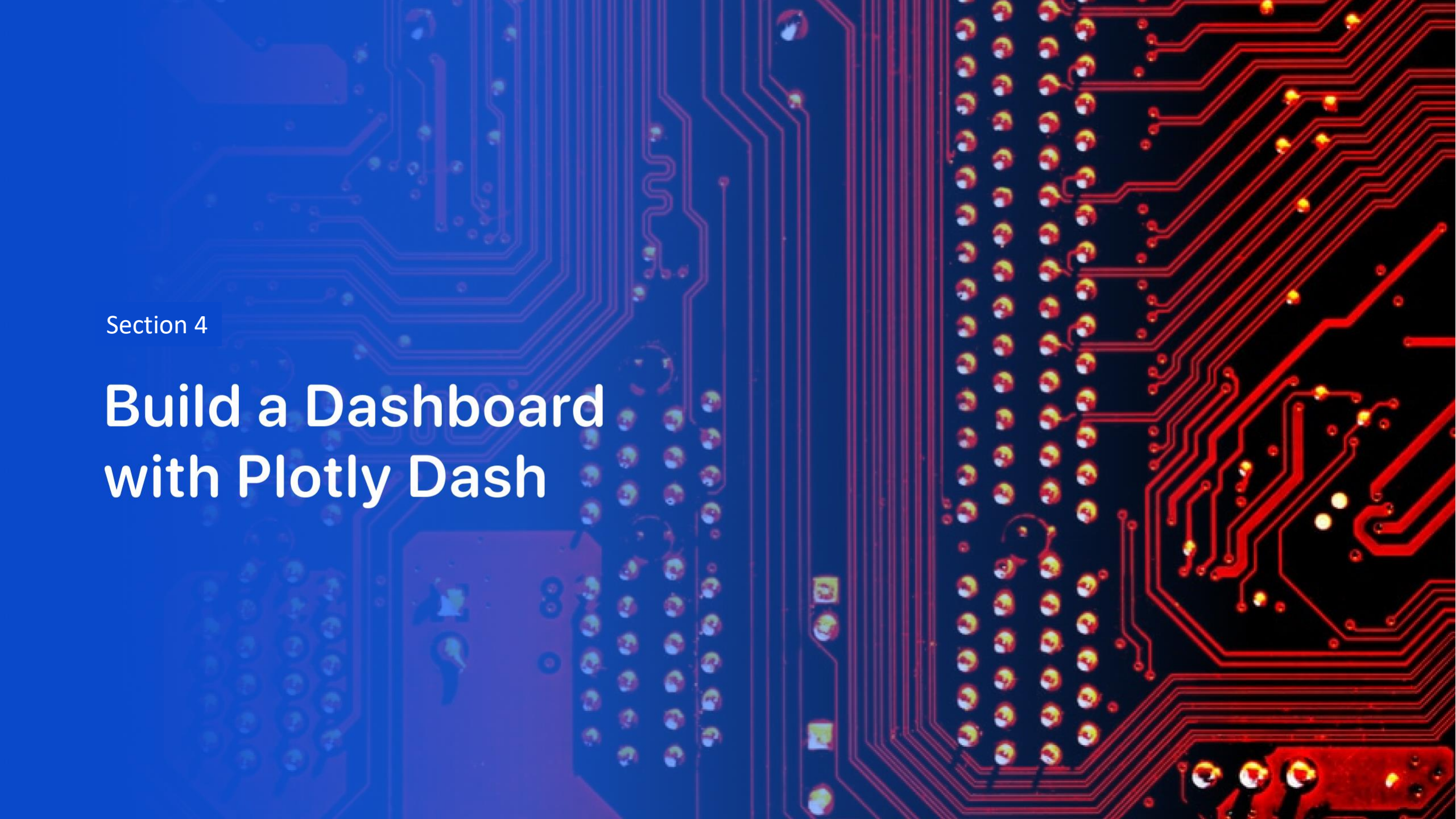


Successes/Failures per launch site

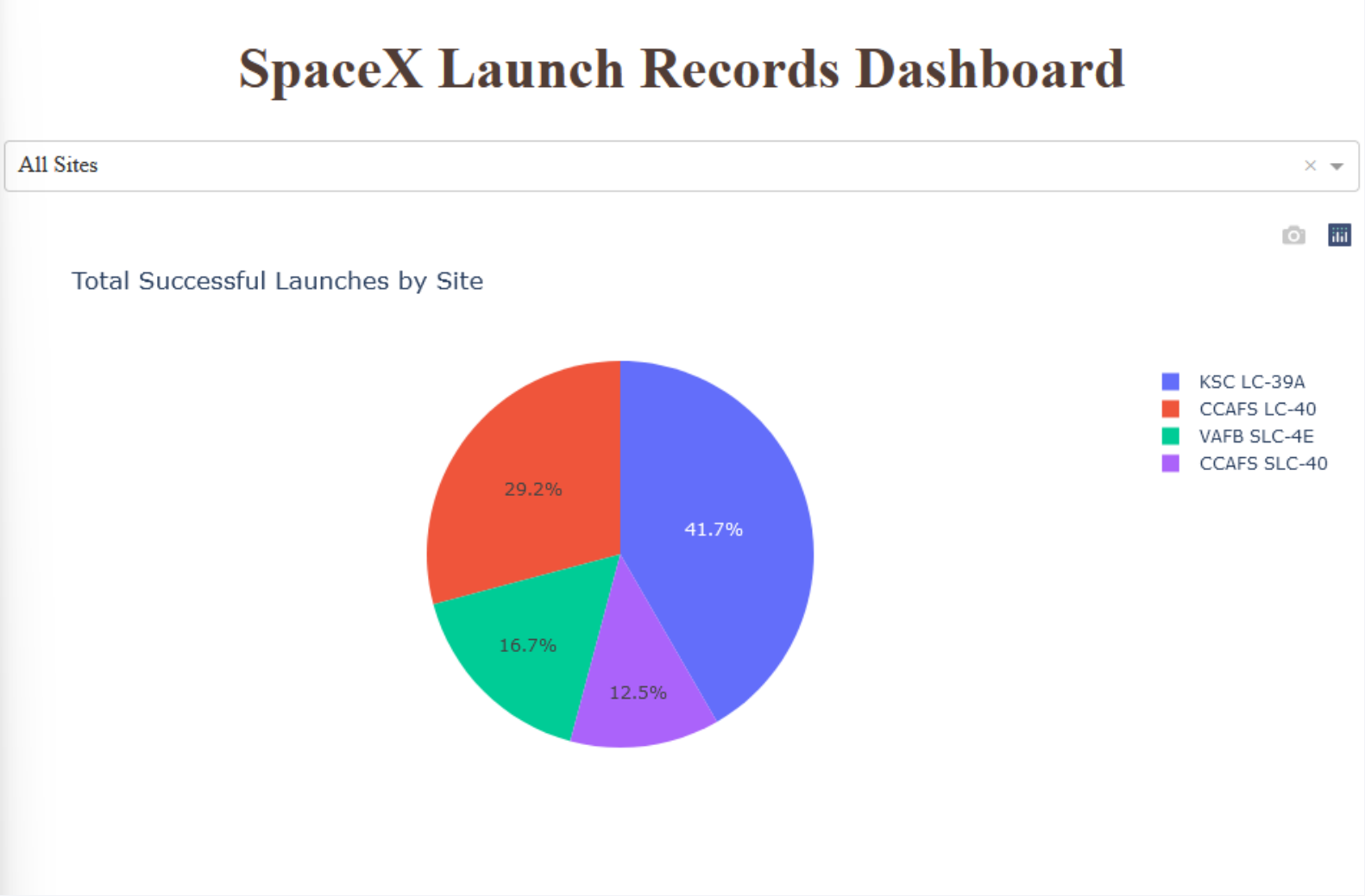


Distance of launch site to coast

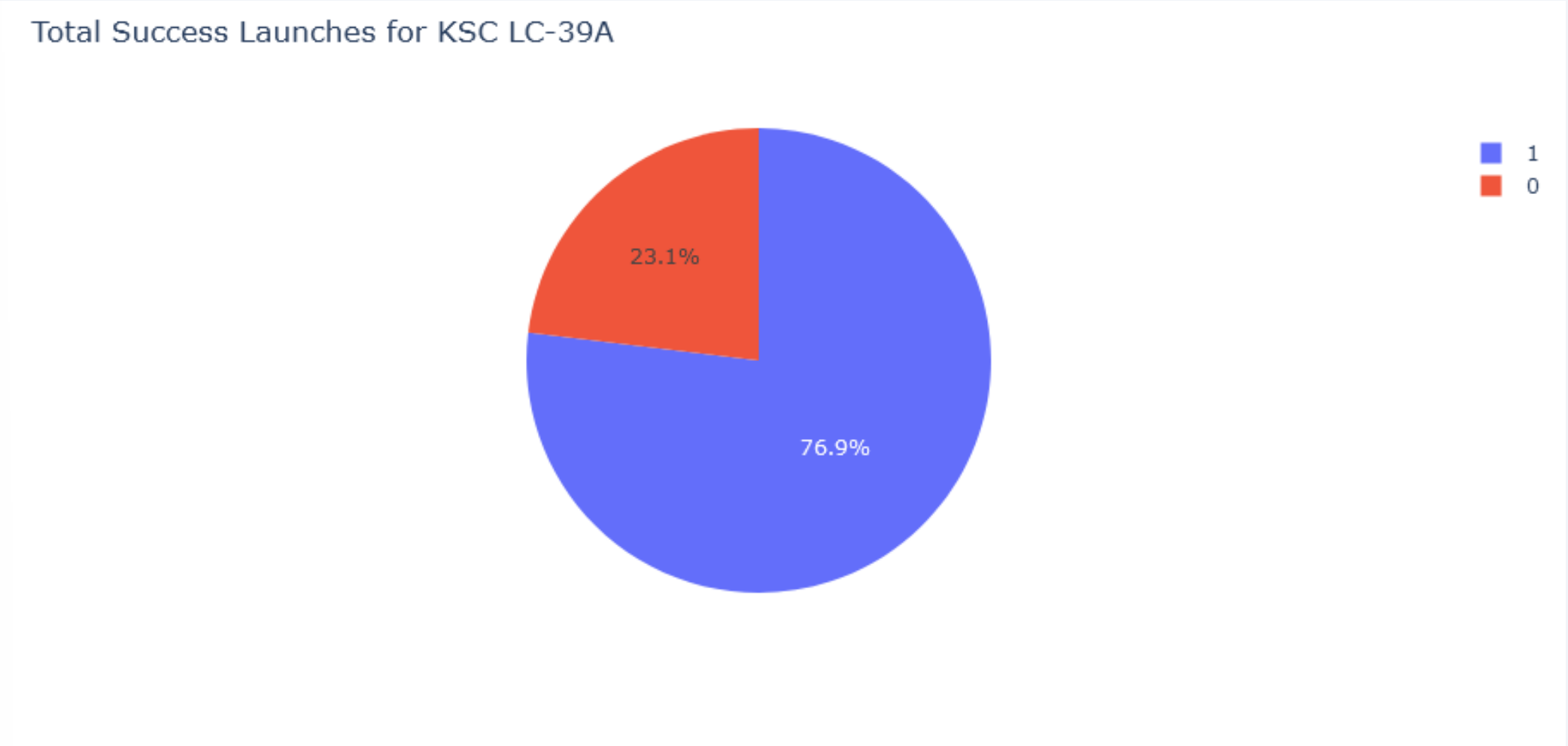



Section 4
Build a Dashboard with Plotly Dash

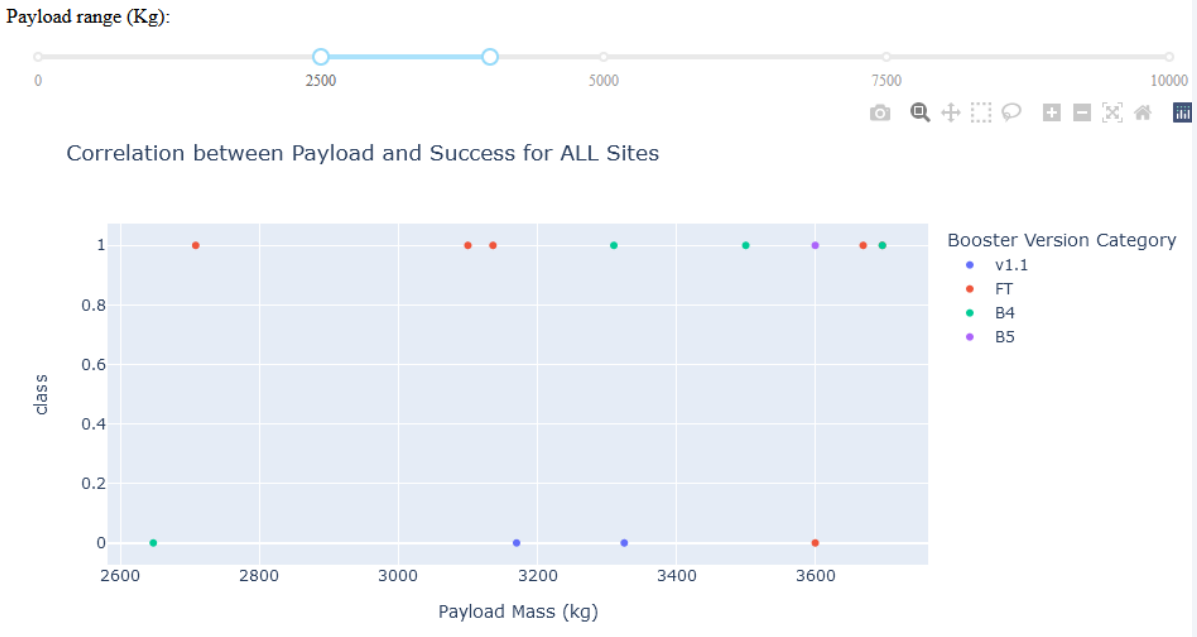
Total Successful Launches by Site



Ratio of successful launches for KSC LC-39A



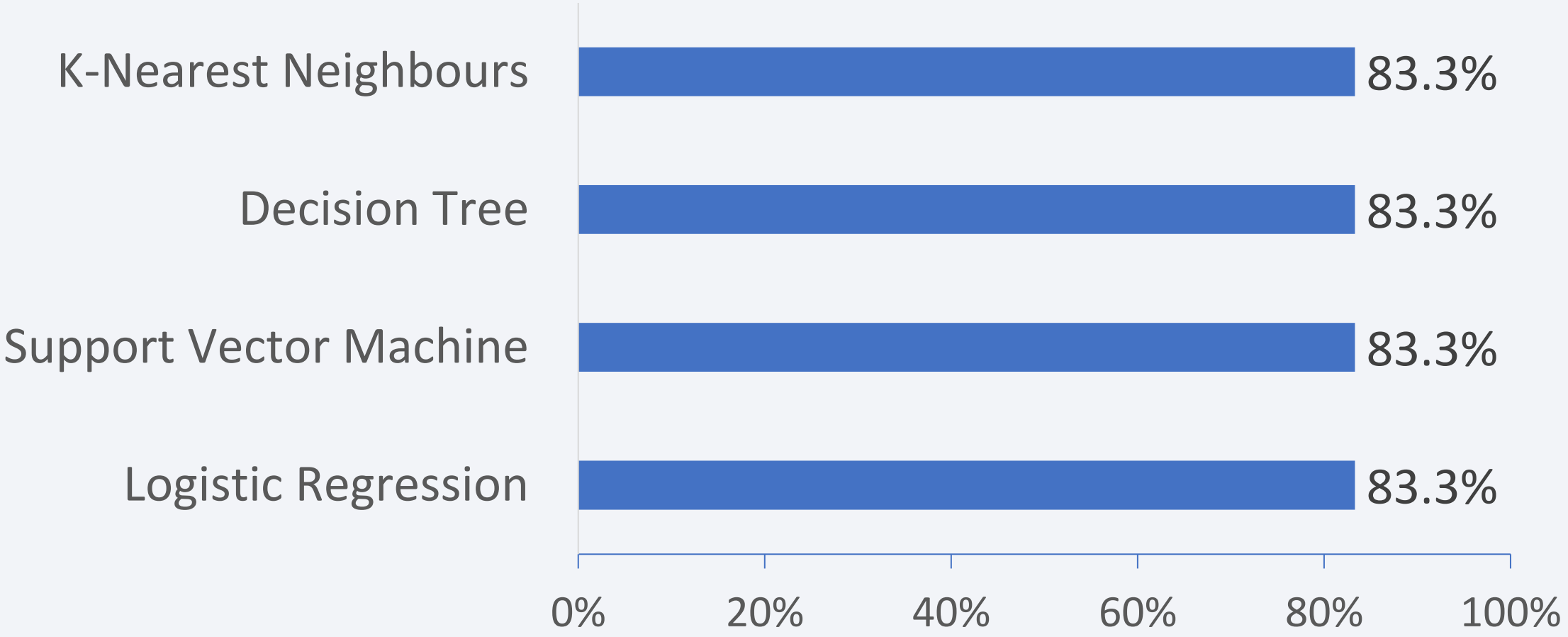
Scatter chart of Payload Mass Vs Success



Section 5

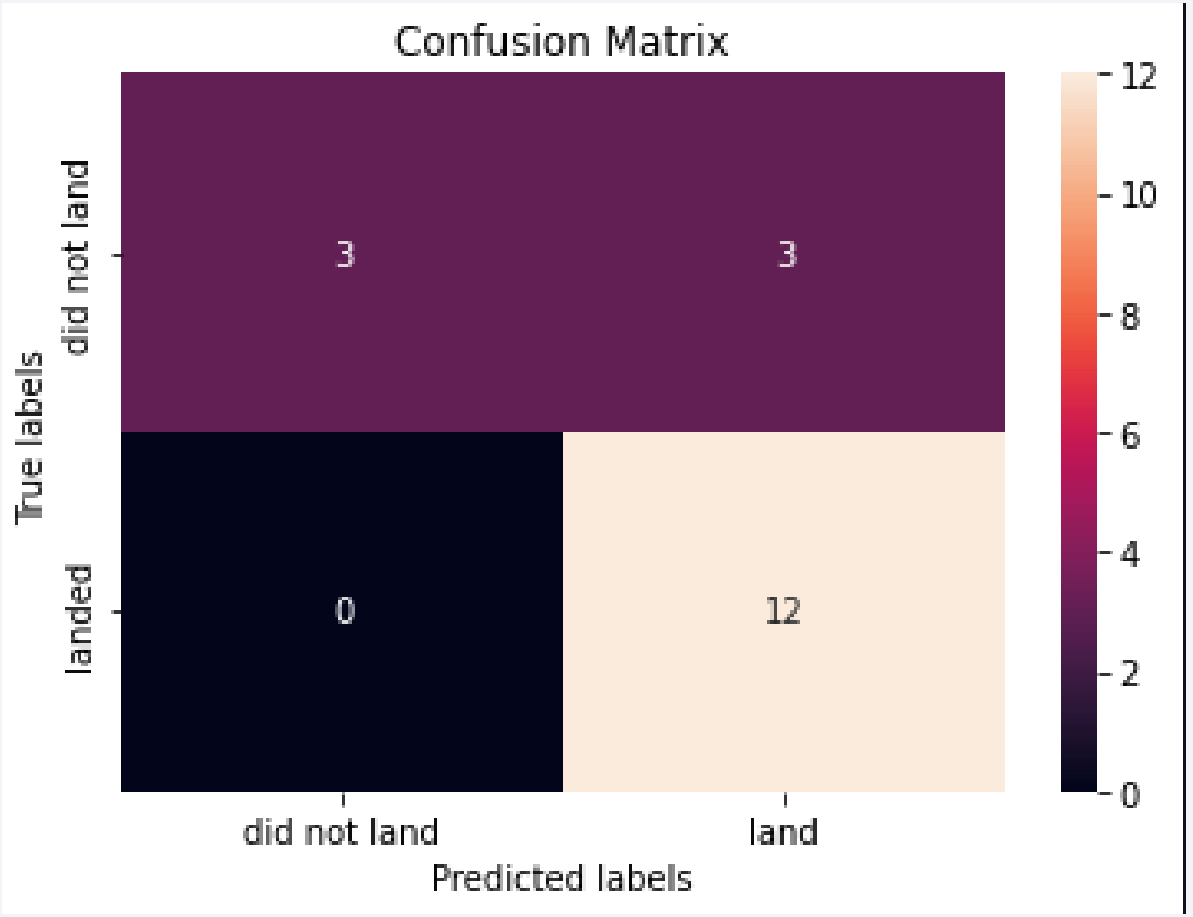
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

- Our model correctly predicted 12 successful landings and 3 unsuccessful landings
- 3 unsuccessful landings were incorrectly predicted to be successful by our model (false positives)
- No landings that were successful were incorrectly predicted to be unsuccessful (false negatives)



Conclusions

- Data about the SpaceX launches can be scraped from Wikipedia or their API
- All models created from this data had the same overall accuracy
- Our models have false positives but not false negatives, therefore it can be used to predict if a stage will fail land
- The lack of overall data points limit the possibilities for testing the model

Thank you!