IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kylee
06 October 2025

Outline

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

Executive Summary

Goal:
Estimates cost-effectiveness of Falcon 9 through prediction of success/failure outcomes

Methodologies

- Data Collection: API, Web Scraping
- Data Wrangling
- Exploration Data Analysis (EDA): SQL, Visualization (Python)
- Interactive Visual Analytics: Folium, Plotly Dash
- Predictive Analysis using Machine Learning

Results Summary

- Final classification model accurately predicts a successful landing with 83% accuracy

Introduction

Project Background & Context:

The space industry is rapidly revolutionizing. SpaceX stands out as a success due to its launch of Starlink and claims of cost efficiency. The relatively inexpensive Falcon 9 rocket launch by SpaceX costs $62 million compared to competitors costs of upwards of $165 million each launch. SpaceX's savings comes from the ability to reuse their first launch.

Objective: Predict the probability of a successful first stage landing of Falcon 9 rocket to inform SpaceY in order to compete in the industry

Problems to Address:

What factors can determine a launch success/fail?

What correlation is between factors in successful launches?

How can a launch be optimized for success?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Public data was retrieved from SpaceX via API
  - Data was supplements by older data scraped from Wikipedia using BeautifulSoup
- Perform data wrangling
  - Data was filtered to only include Falcon 9 launches, features were enriched, and missing values were replace with column means
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Performed hyperparameter tuning using GridSearchCV, compared SLMs, selected best model

Data Collection

Data was collected using both API from SpaceX as well as data scraped from the web to provide historical context. The following data was obtained for utilization in this modeling:

SpaceX API:

FlightNumber, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, FLights, GridFins,ReUsed, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

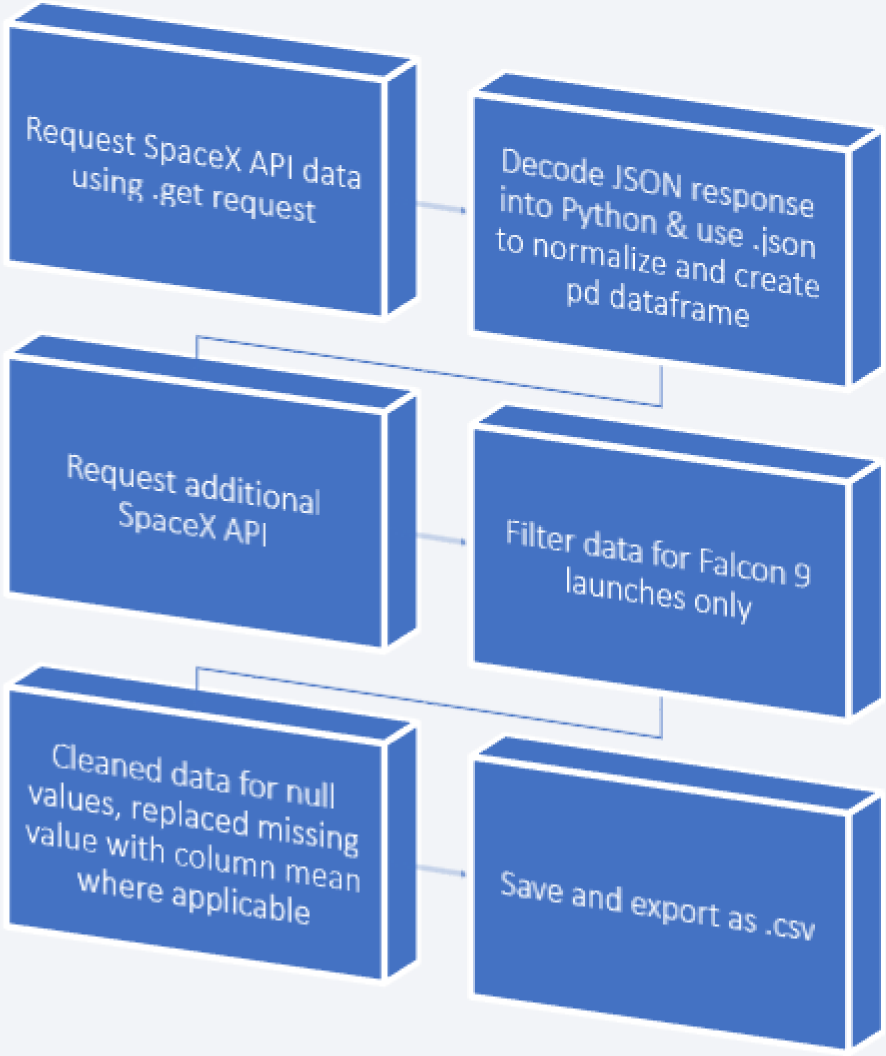
Web Scraping:

Flight #, Launch SIte, Payload, Payload Mass, Orbit, Customer, Launch outcome, Version Booster, Booster Landing, Date, Time

Data Collection – SpaceX API

- The flowchart details the collection of data with python and SpaceX API

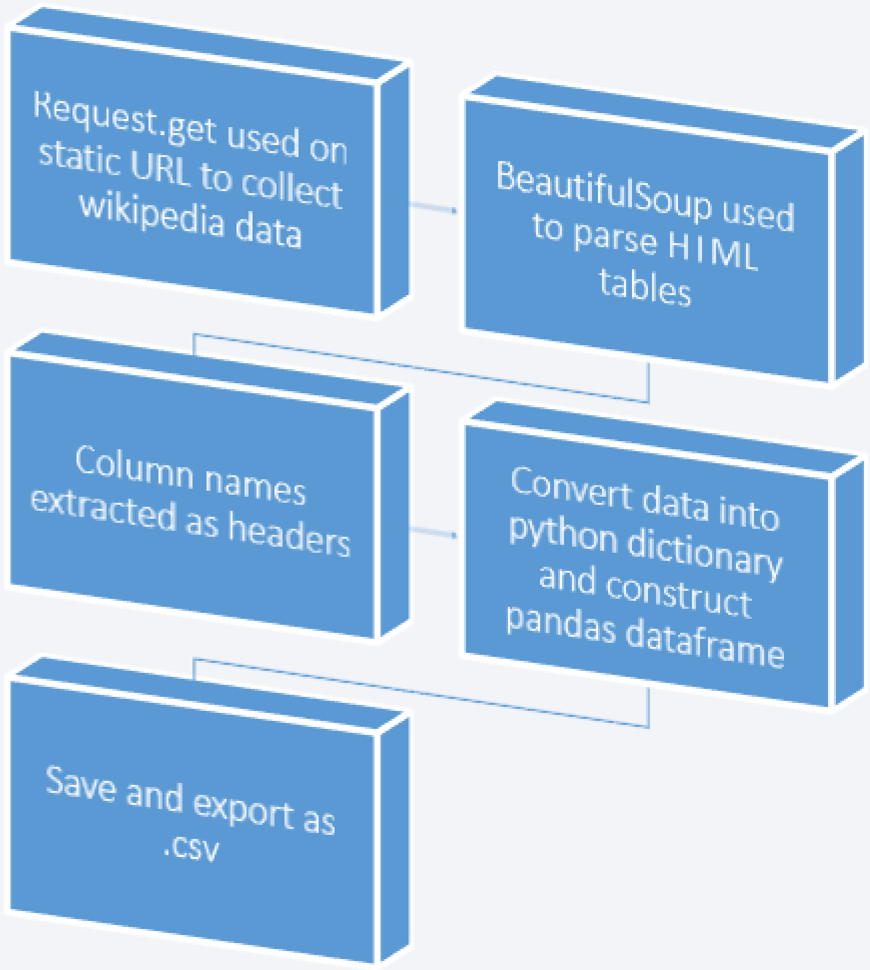
- Implementation of the process can be viewed here



Data Collection - Scraping

- The flowchart details the collection of supplemental data using web scraping from wikipedia using BeautifulSoup then parsed

- Implementation of the process can be viewed here



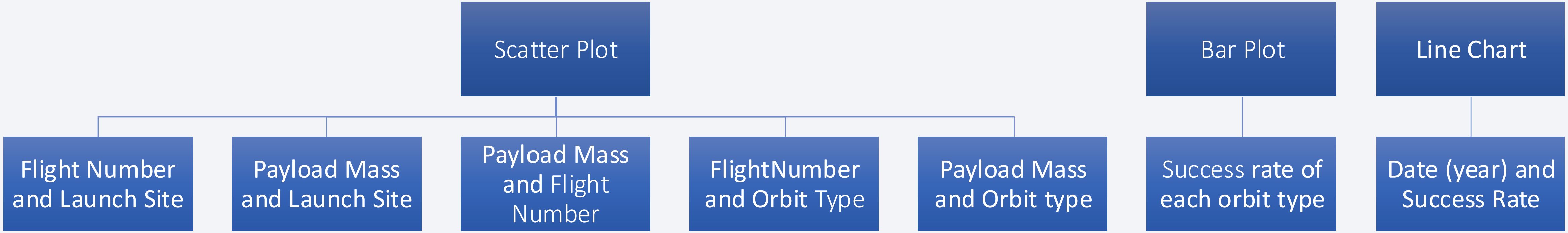
Data Wrangling

- **Import Data**: Use pd.read_csv to import data
- **Identify & Calculate**: Identify missing values, Identify column types using .dtypes
- **Perform Value Counts**: Take value counts to define model features such as Launch Sites & Orbit
- **Create New Feature**: Create "Class" feature to represent landing outcomes as 1-Succes and 0-Other
- **Calculate success rate**: Use mean of Class to create target value
- **Save & Export**: Store and export data as CSV

Process implementation available [here]

EDA with Data Visualization

Charts Used



EDA with SQL

Github file

- SQL queries performed:
  - Display unique launch sites
  - Display 5 sites beginning with 'CCA'
  - Display Payload Mass of boosters launched by NASA
  - Display Avg Payload carried by F9 v1.1
  - List date of first success landing in ground pad outcome
  - List successful landing in drone ship w/payload mass >4,000 kg and <6,000kg
  - List total success, total failures
  - List all the booster versions that have carried the maximum payload mass
  - List the records which will display the month names, failure landing outcomes in drone ship , booster versions, launch site for the months 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

Build an Interactive Map with Folium

GitHub Files

Map objects such as markers, circles, lines, etc. added to a folium map

- Launch Sites
  - Highlights all launch sites and allows for identification of geographical location in relations to coasts, equator, ect.
- Launch Outcomes
  - Cluster groups identified through green (success) and red (failure) marking to easily identify which sites have high rates of successful launches
- Distance Lines
  - Displays lines showing proximity of launch sites to nearby infrastructure such as railways, highways, cities which highlights essential launch site location infrastructure

Build a Dashboard with Plotly Dash

Github File

Drop Down List
Allows filtering of data by launch site

Slider
Allows selection of Payload Mass range to evaluate for correlation between mass and launch outcome (success/failure)

Pie Chart
Allows for easy view of success/failure rates

Scatter Chart
Allows for visual evaluation of correlation between Payload Mass and Launch Outcome for different sites

Predictive Analysis (Classification)

GitHub File

Development & Evaluation of Predictive Model

Load & Prepare Data

- Import data, load into a NumPy array, Standardize

Split into Test/Train

- Split data into test and train sets

Develop model

- Apply GridSearchCV across various algorithms to determine best parameters

Evaluate

- Use test set to evaluate each model

Model Selection

- Plot confusion matrix to visualize and compare accuracy of each to select best score

Results

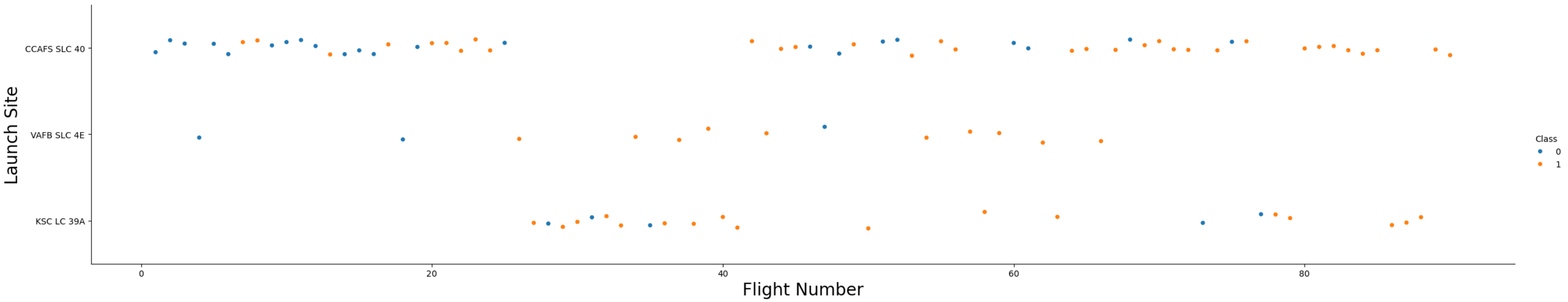
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

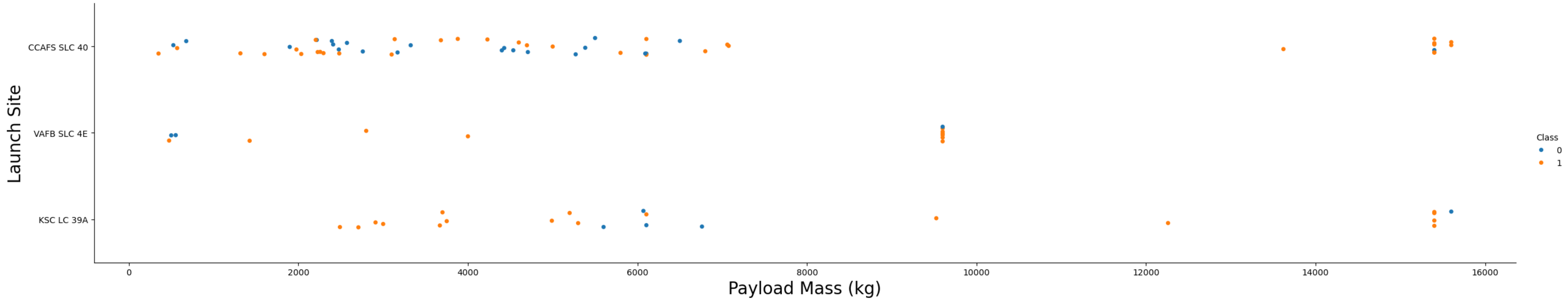
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

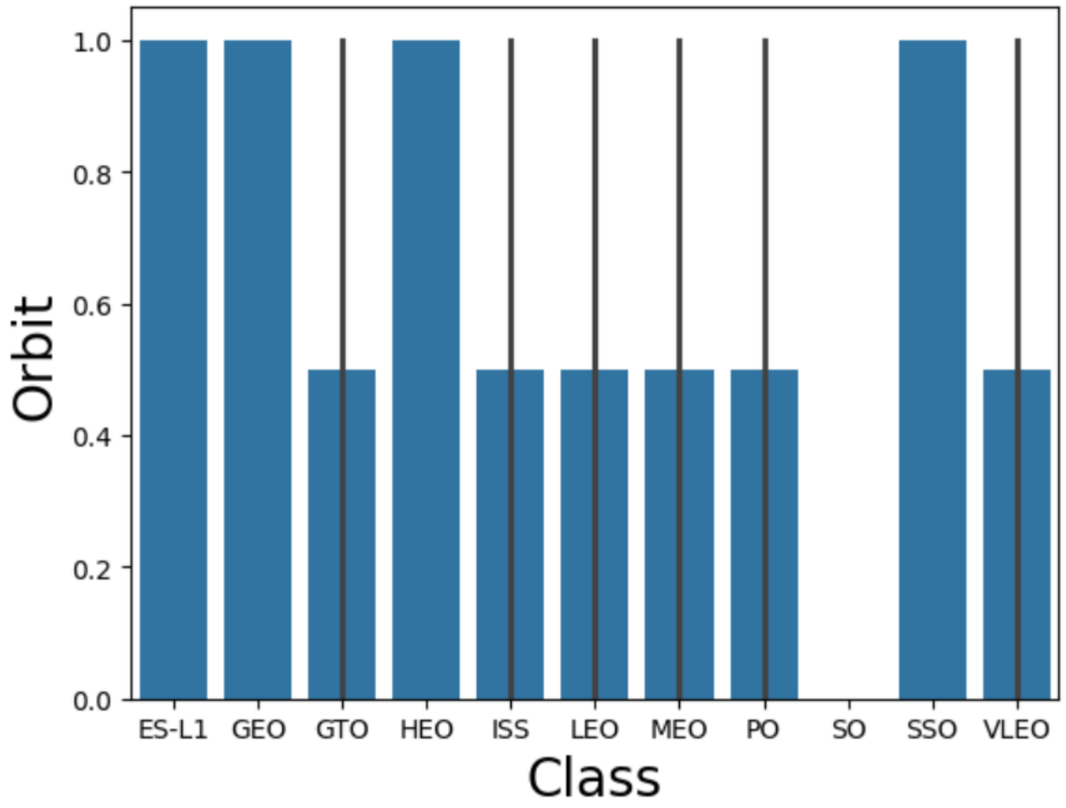
- CCAFS is the most used Launch Site
- VAFB SLC 4E had fewer launches overall, but with mostly successful outcomes
- Suggests SpaceX's reliability improved over time (as flight numbers increased, failures decreased)





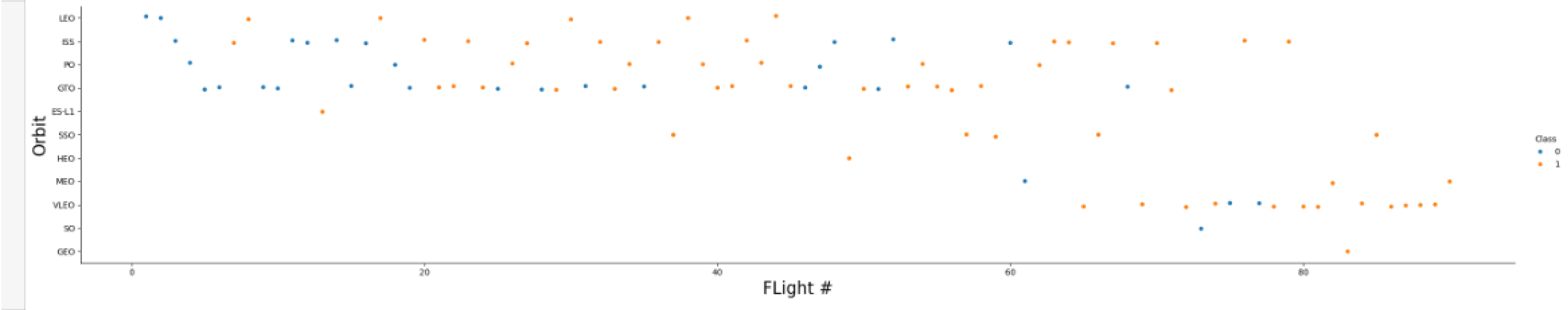
Payload vs. Launch Site

- Launch success doesn’t depend strongly on payload mass, but some launch sites handle different payload ranges more frequently.
- The higher payloads (above ~10,000 kg) mostly come from KSC LC 39A, suggesting that site handles heavier missions.



Success Rate vs. Orbit Type

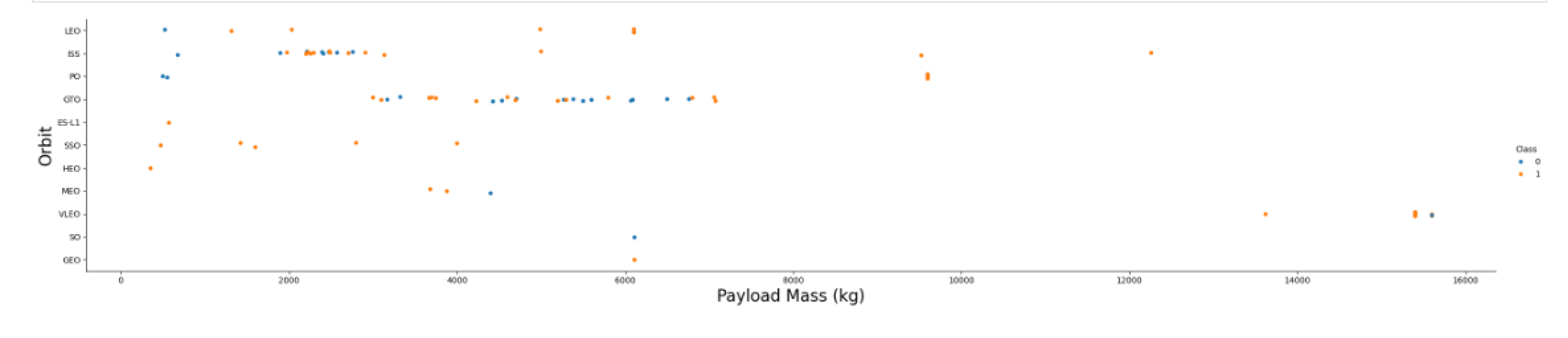
- SpaceX had perfect success rates for some higher orbits like GEO and SSO, but mixed results for lower orbits (LEO, ISS, etc.)



Flight Number vs. Orbit Type

- Success seems to be related to the number of flights in LEO orbit. While in the GTO orbit, there appears to be no relationship between flight number and success.

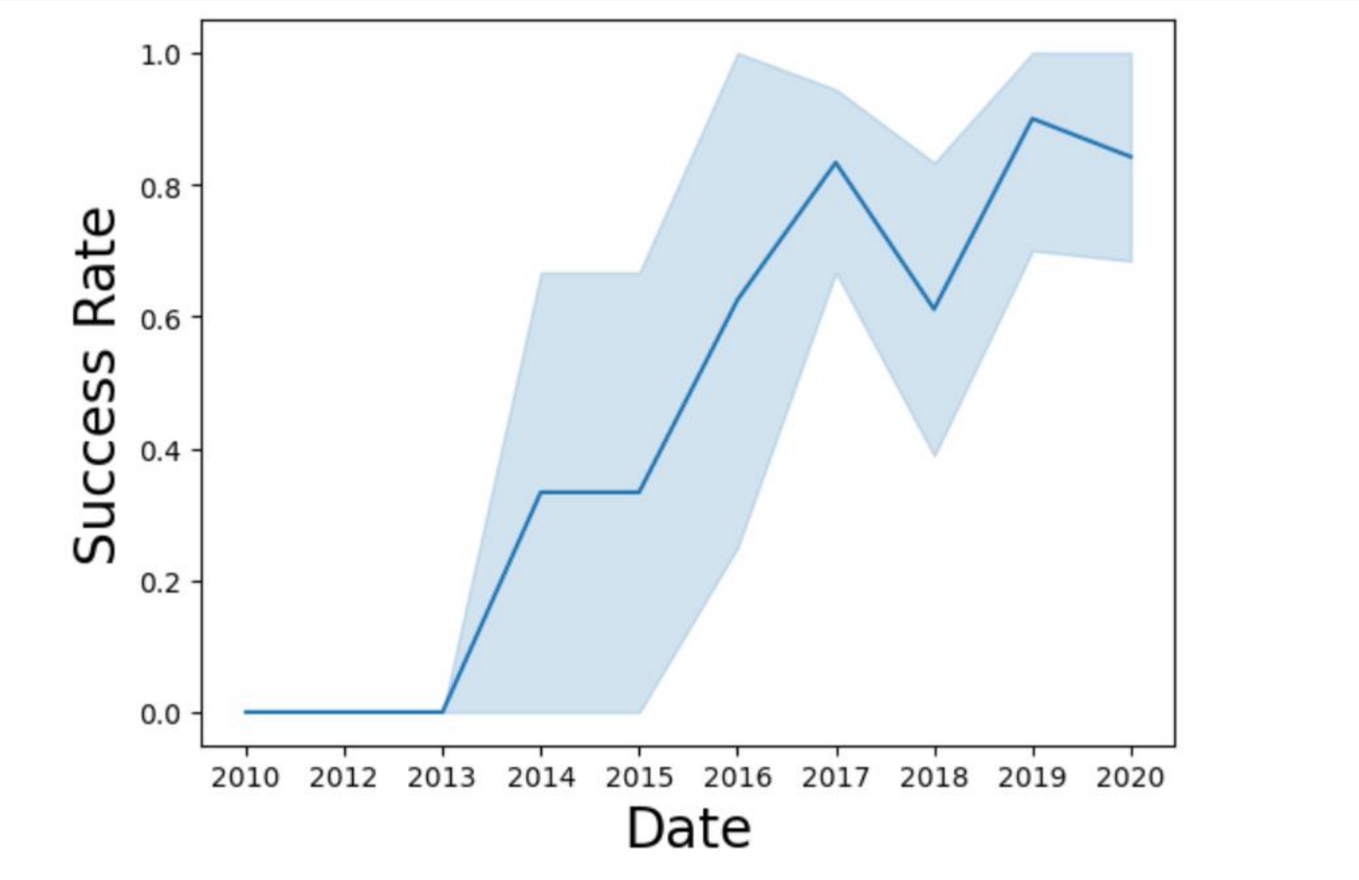
Payload vs. Orbit Type



- PO, LEO and ISS are where higher Payloads have success
- GTO is inconclusive as it has both failure and success across Payload Masses

Launch Success Yearly Trend

- The Success Rate increase each year from 2013-2019.
- There is a slight dip in 2018, and the Success Rate is decreasing once it hits 2020



All Launch Site Names

- There are 4 unique Launch Sites
- This can be found using command:
    - %sql select distinct(LAUNCH_SITE) from SPACEXTBL

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

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

Launch Site Names Begin with 'CCA'

- The command *%sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5* will pull Launch Site names beginning with CCA

Total Payload Mass

- Sum function used to calculate total Payload Mass for NASA (CRS) who is a customer
- Total is 45,596 kg

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Average Payload Mass by F9 rocket is 3,928.4 kg

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
```

```
 * sqlite:///my_data1.db
Done.
```

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

First successful ground landing is 22 December 2015

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

```
 * sqlite:///my_data1.db
Done.
```

min(DATE)
2015-12-22

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

Boosters with successful drone ship landings between 4000 and 6000 Payload can be found by:

%sql select BOOSTER_Version from SPACEXTBL where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000

Total Number of Successful and Failure Mission Outcomes

- There were 100 success and 1 failure

```
%sql select count(Mission_Outcome) from SPACEXTBL GROUP BY 'Mission_Outcome' = 'Success' or Mission_Outcome = 'Failure (in flight)'
```

```
 * sqlite:///my_data1.db
Done.
```

count(Mission_Outcome)
100
1

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

- 12 Boosters carried the Max Payload
- *Code used:*

  %sql select Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)

2015 Launch Records

- There were 2 failed landing_outcomes in drone ship in year 2015 both at CCAFS LC-40

```
%sql SELECT SUBSTR(Date,6,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Failure%drone%' AND SUBSTR(Date,0,5) = '2015'
```

```
 * sqlite:///my_data1.db
Done.
```

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

- %sql select Landing_Outcome, COUNT(*) AS Numbers FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'GROUP BY Landing_Outcome ORDER BY Numbers DESC;

Landing_Outcome	Numbers
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

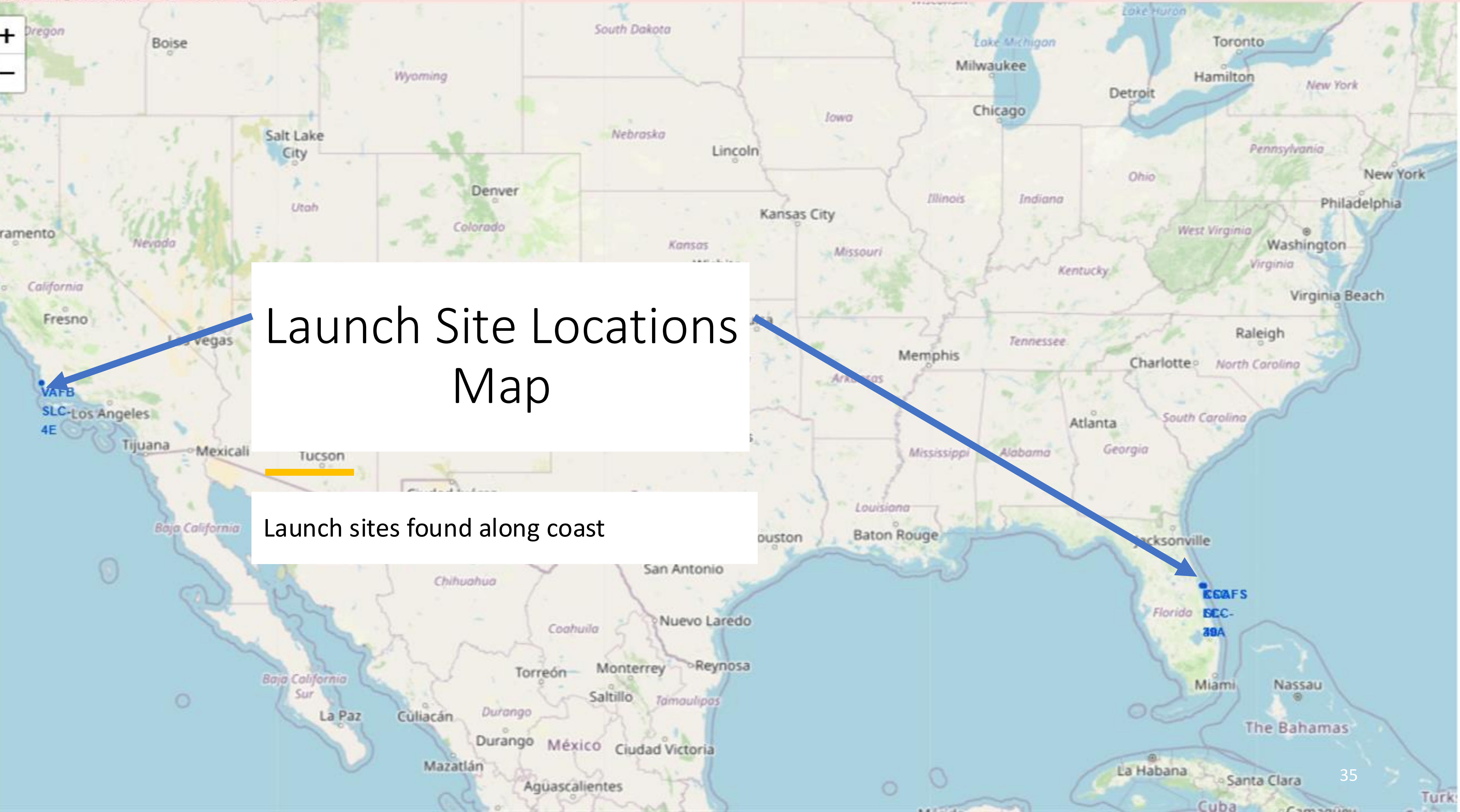
Section 3

Launch Sites Proximities Analysis

```
return [float(x) for x in coords]
```

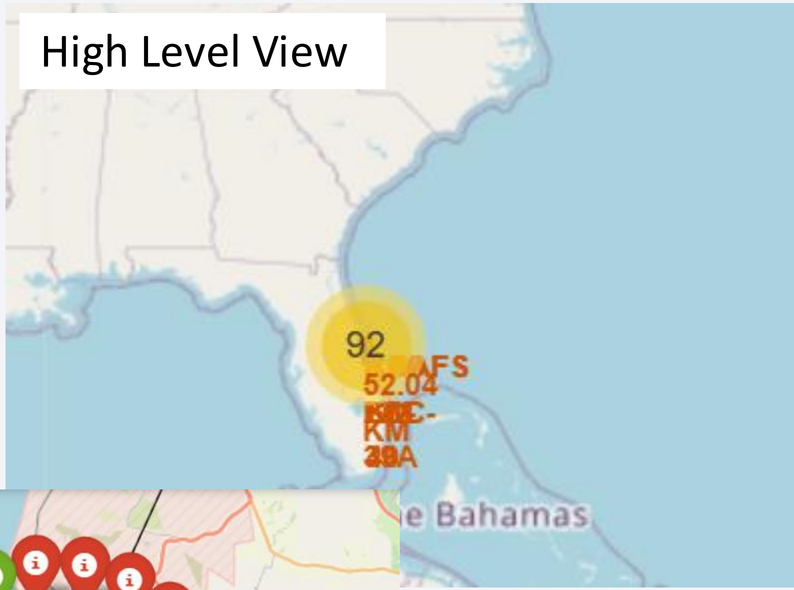
Launch Site Locations Map

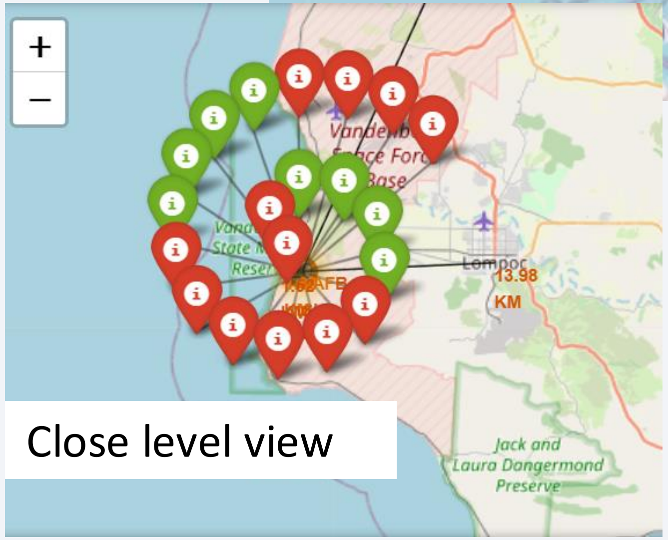
Launch sites found along coast



Success/Failure Launch Markers Map

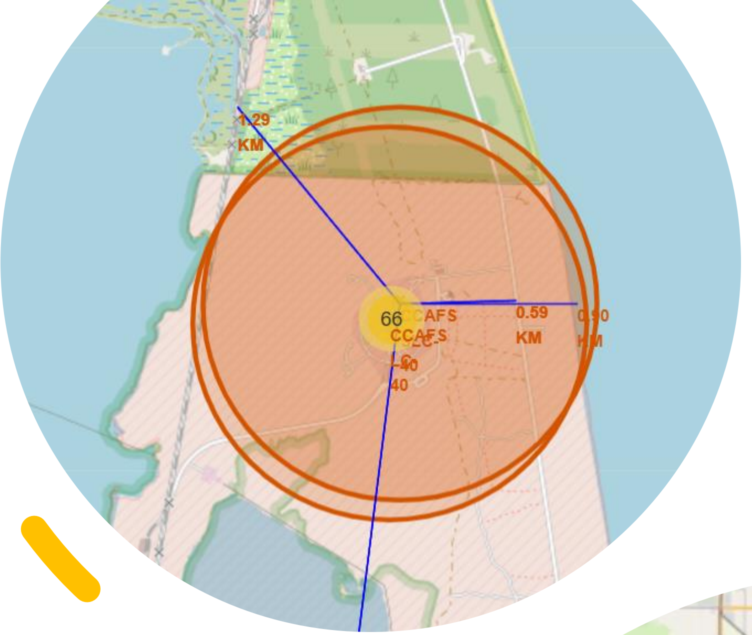
- Maps show the color-labeled launch outcomes on the map (green = success, red = fail)
- A high level view shows number of launches
- Close level view allows easy identification of which site have high rates of successful launches

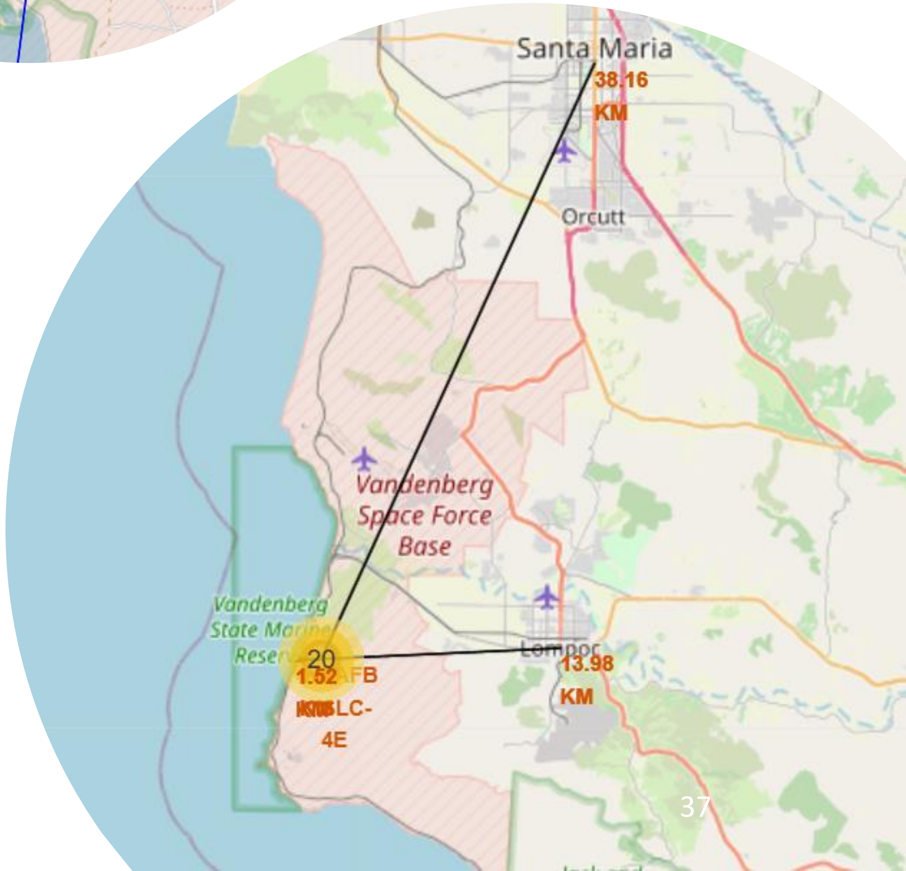




Proximity Markers Map

- Represents the selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed



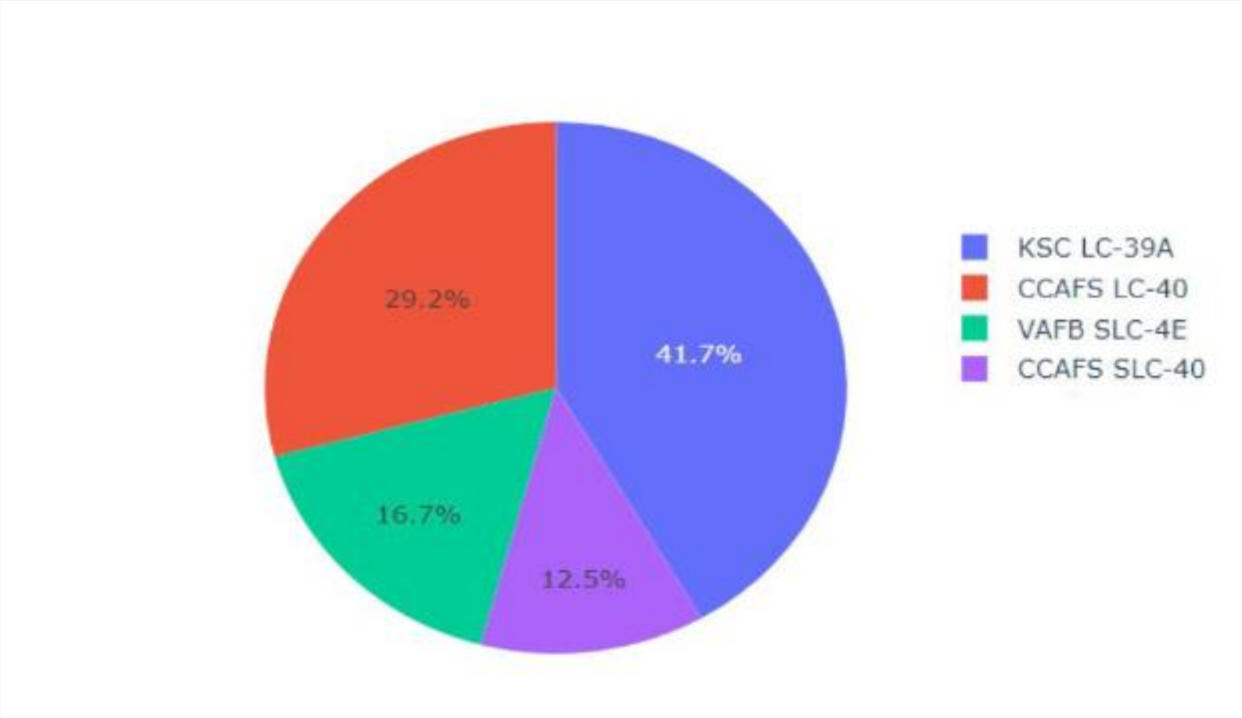


Section 4

Build a Dashboard with Plotly Dash

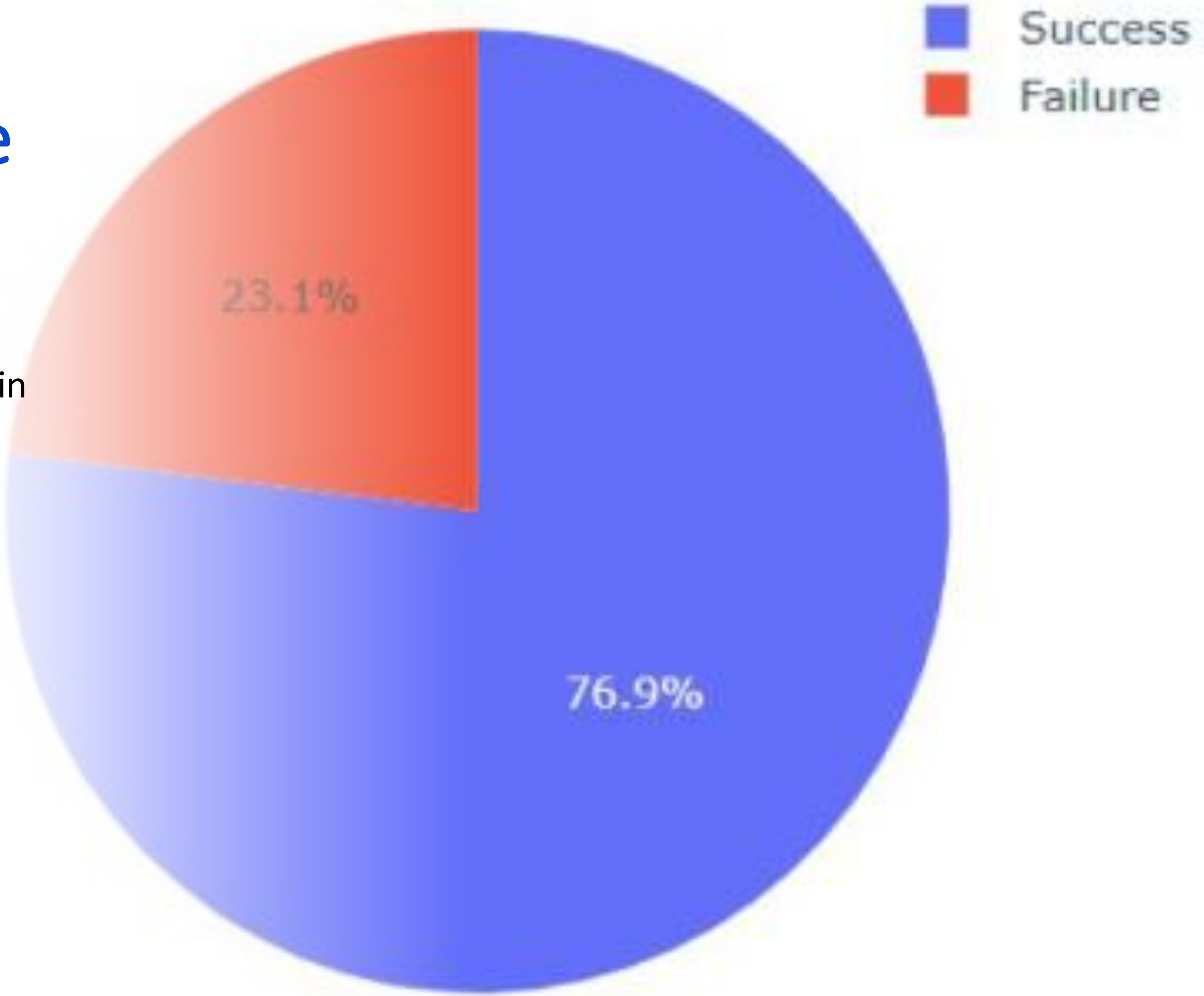
All Launch Sites Pie Chart

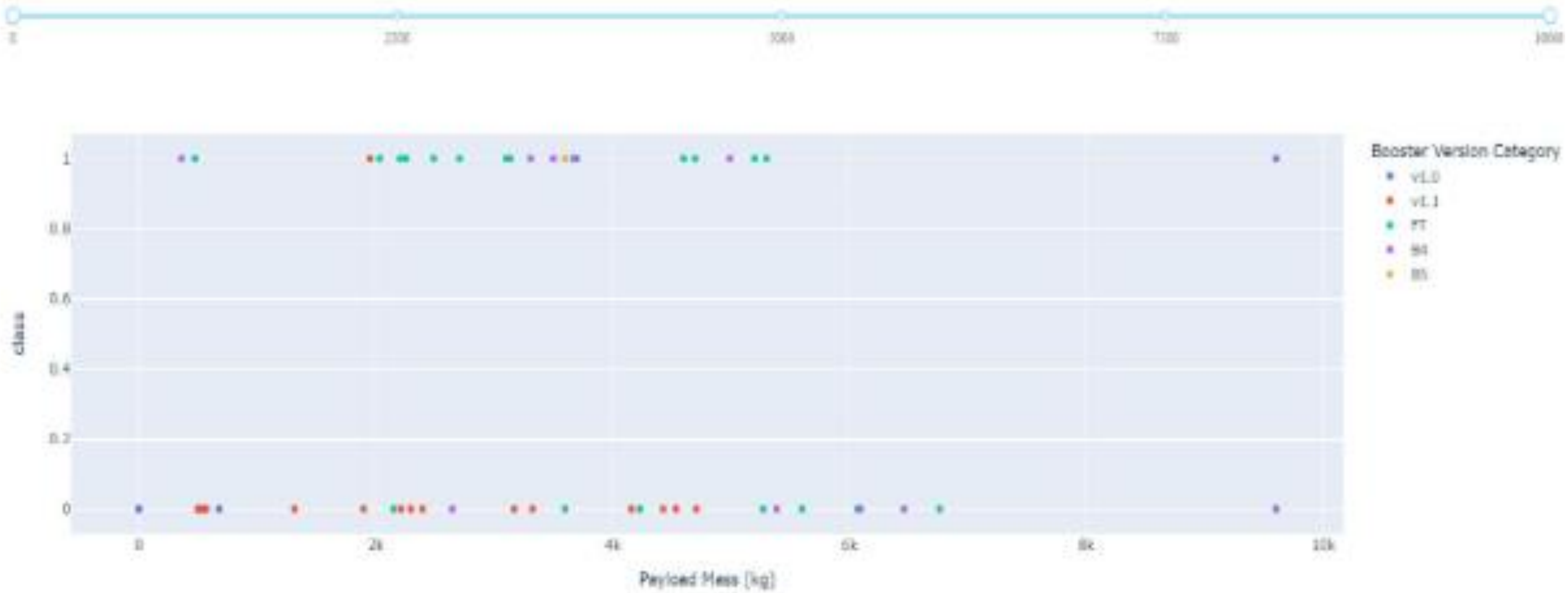
KSC LC-39A has the most launches at 41.7%, which is nearly half of all launches



Success vs Failure of KSC LC 39A

- KSC LC 39A is 76.9% successful in launches

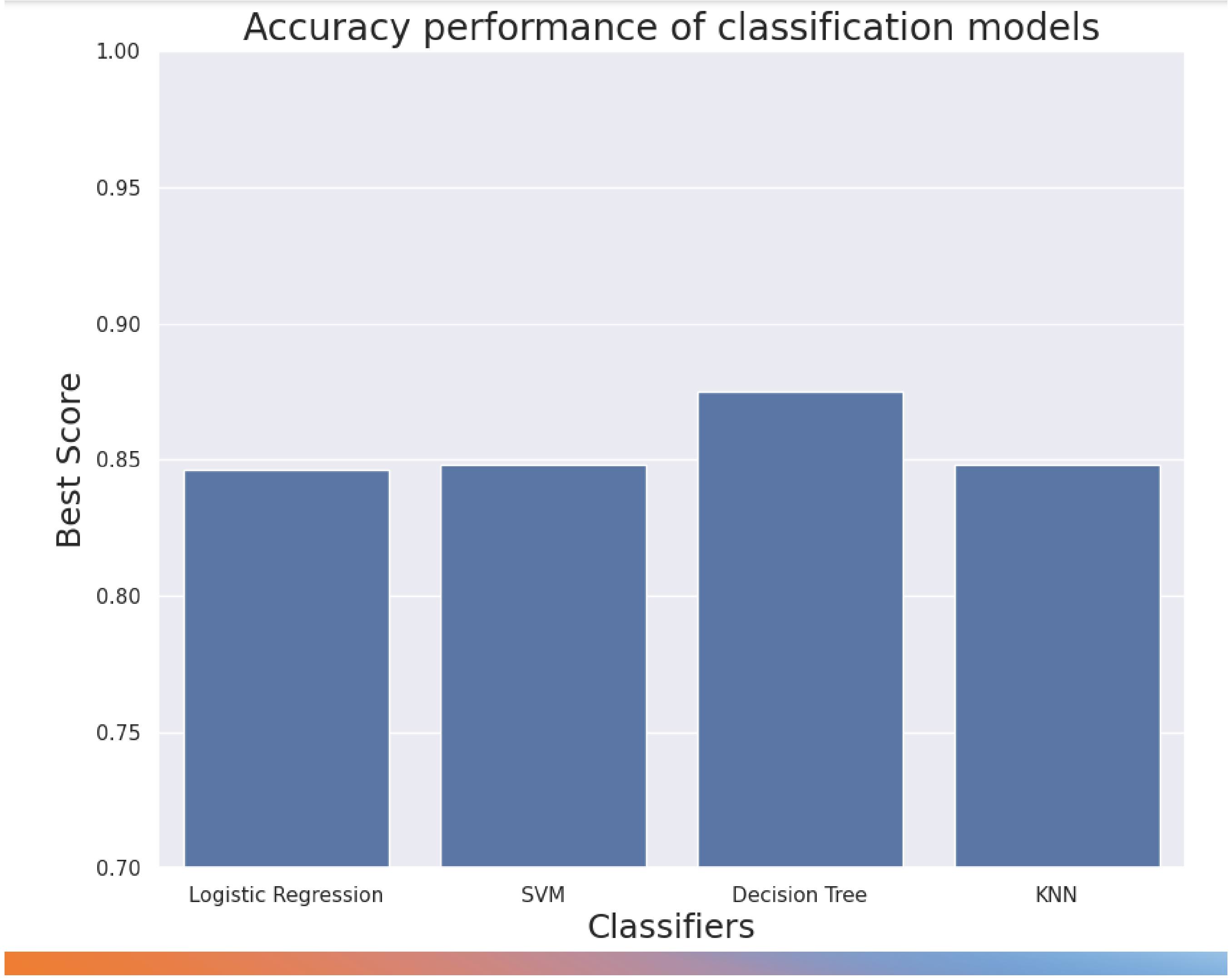




Payload vs. Launch Outcome for Different Payloads

Section 5

Predictive Analysis (Classification)



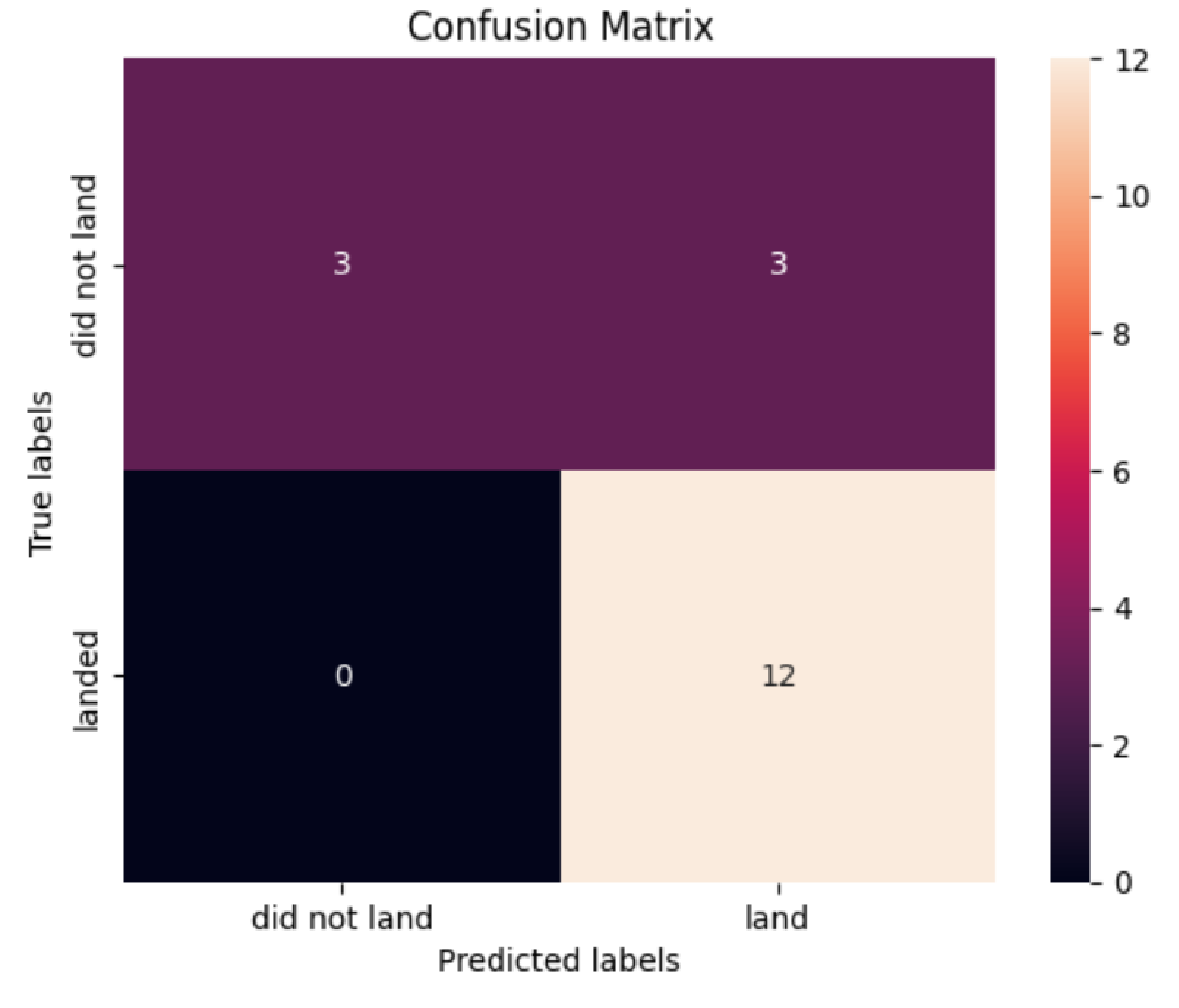
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- The accuracy of the Decision Tree model is ~86%

Confusion Matrix

- The test set had 18 samples, of those 18 the decsion tree model correctly predicted 12 Landings (True Positive) and 3 Did Not Lands (True Negative)
- There were no False Negatives
- The model predicted 3 false positives



Conclusions

- Payload Mass and Orbit seem to play a role into successful launch and landings, so therefore should be taken into consideration
- Launch sites located along the coast and away from civilian infrastructure are optimal
- With accuracy of 87% we can predict success of a launch leading to confidence in outcomes
- Over time, the success of launches has improved

Image Source

Thank you!

https://github.com/kylee-ka/Data-Science-Capstone-SpaceX-Falcon9-Launch-Success