IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Redha Zidan
19 JAN 2024

Outline

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

Executive Summary

- The following data is sourced from Wikipedia and spaceX api. In this project we did extensive data analysis on the given data. Machine learning models were also created to give the best accuracy results.
- Four machine learning models were produced: Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K Nearest Neighbors.

Introduction

- SpaceX is one of the leading space programs company which has surpassed all the past technologies for space.
- We must find predict if space Y can perform stage 1 recovery like Space X.

Section 1

Methodology

Methodology

Executive Summary

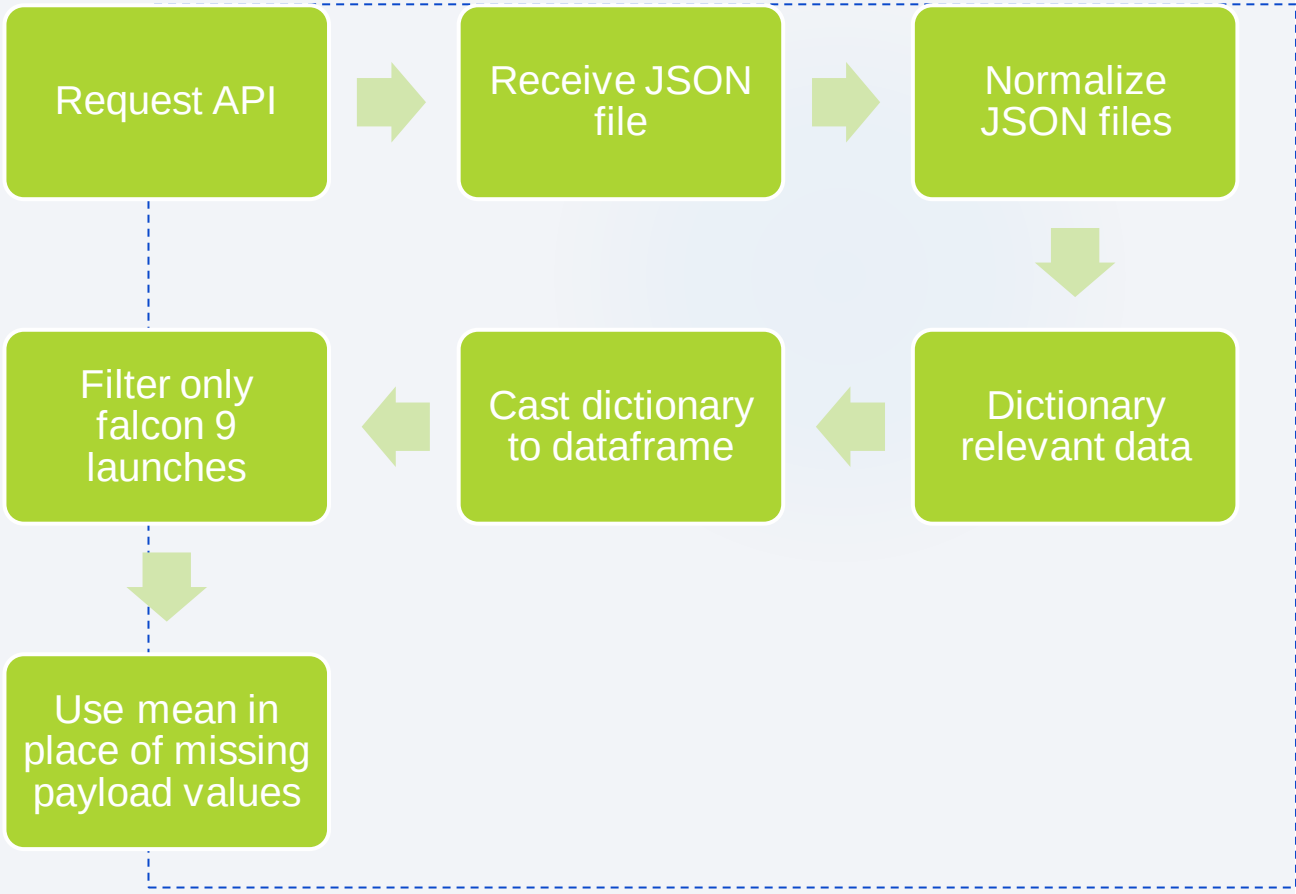
- Data collection methodology:
  - We successfully sourced our data from the SpaceX api.
- Perform data wrangling
  - Classifying data as successful and unsuccessful.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- We have sourced our data from SpaceX Wikipedia page and the SpaceX API by requesting data.
- We have created two data frames,
- SpaceX API:- FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins,Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude
- Wikipedia:-Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Collection – SpaceX API

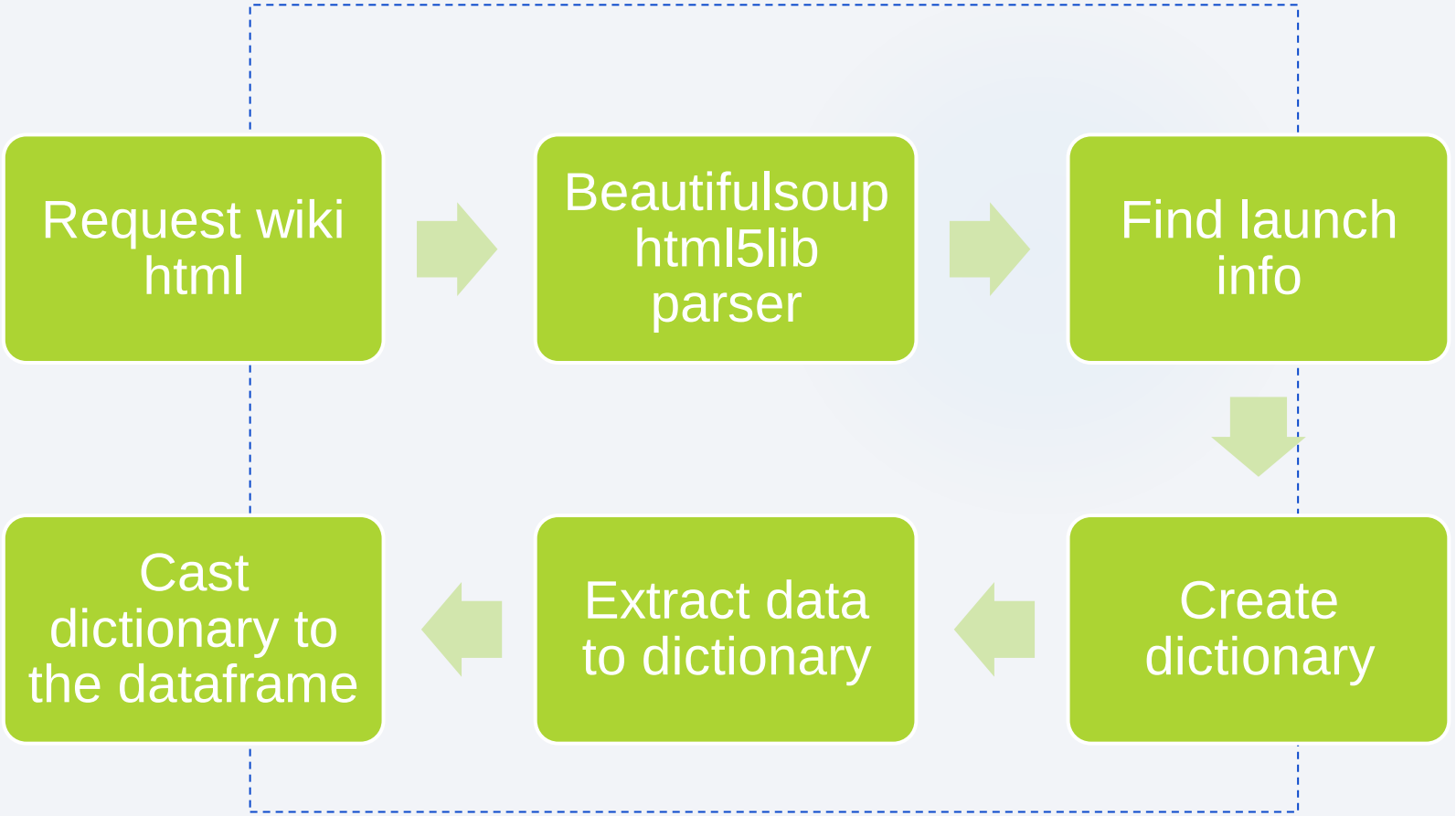
https://github.com/zidanredha2/Capstone/blob/713c4885b523d6585f42d13c6160d5cd1ecc98c2/Data%20Collection%20Api%20.ipynb



Data Collection - Scraping

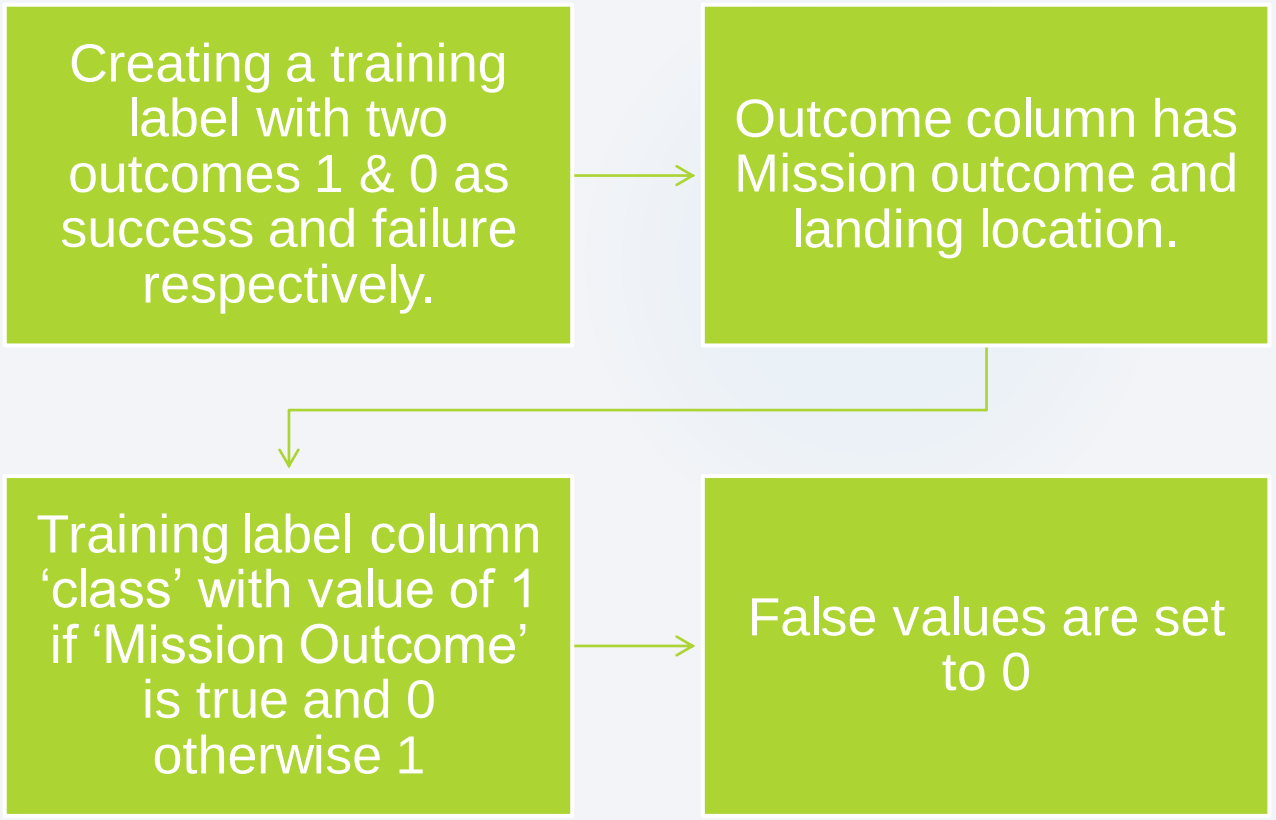
url-

https://github.com/zidanredha2/Capstone/blob/713c4885b523d6585f42d13c6160d5cd1ecc98c2/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- We convert and format the givien data into more useable format and filter out the columns and that we require.
- https://github.com/zidanredha2/Capstone/blob/713c4885b523d6585f42d13c6160d5cd1ecc98c2/Data%20wrangling%20.ipynb



EDA with Data Visualization

- Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit vs. Success Rate, Flight Number vs. Orbit, Payload vs Orbit, and Success Yearly Trend
- Scatter plots, line charts, and bar plots were used to compare relationships between variables to
- decide if a relationship exists so that they could be used in training the machine learning model.
- https://github.com/zidanredha2/Capstone/blob/713c4885b523d6585f42d13c6160d5cd1ecc98c2/EDA%20with%20Visualization.ipynb

EDA with SQL

- Loaded data set into IBM DB2 Database.
- Queried using SQL Python integration.
- Queries were made to get a better understanding of the dataset.
- Queried information about launch site names, mission outcomes, various pay load sizes of customers and booster versions, and landing outcomes
- https://github.com/zidanredha2/Capstone/blob/713c4885b523d6585f42d13c6160d5cd1ecc98c2/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

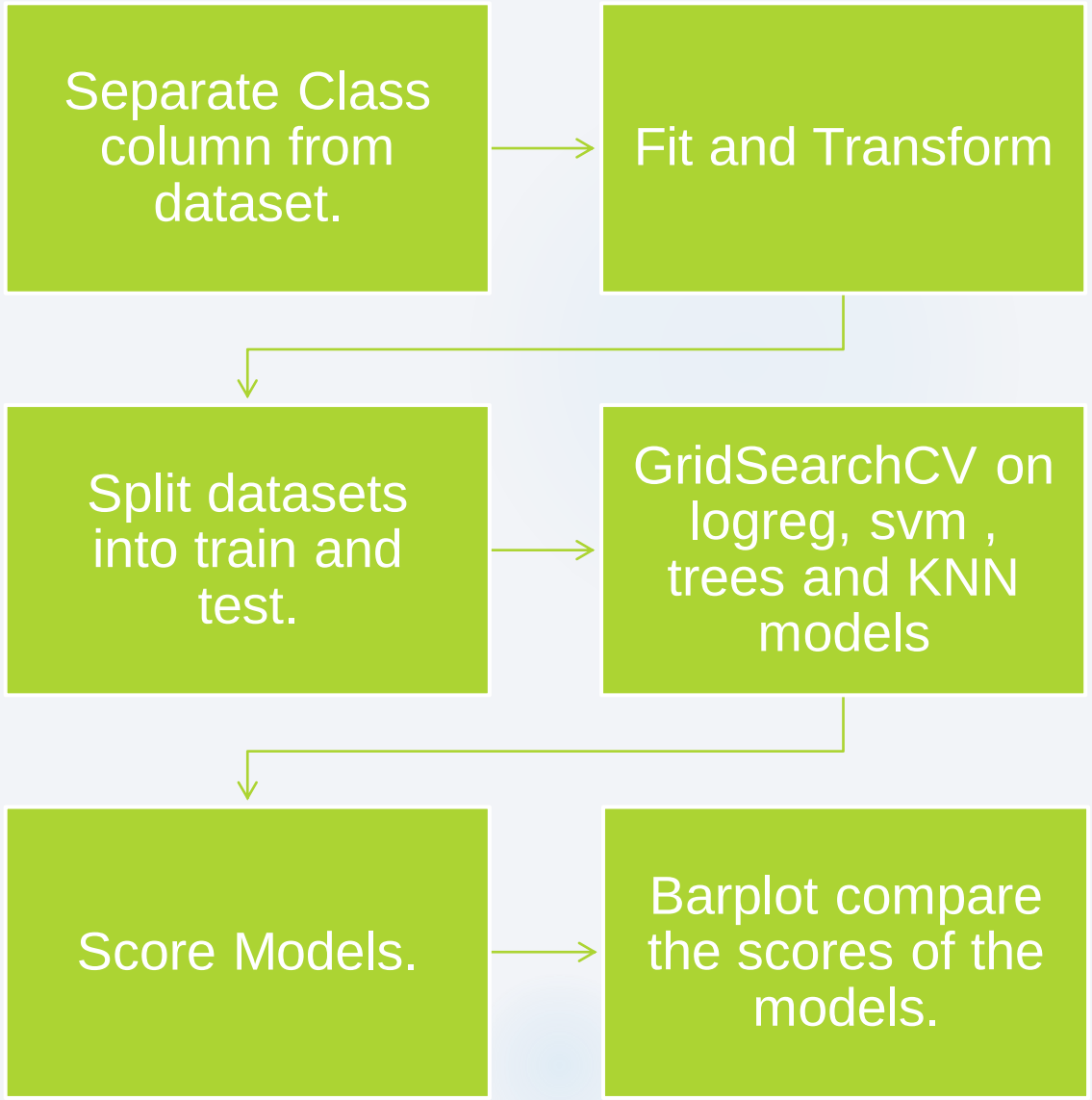
- Folium maps mark Launch Sites, successful and unsuccessful landings, and a proximity example to key locations: Railway, Highway, Coast, and City.
- We do this so that we can understand why launch sites may be located where they are. Also visualizes successful landings relative to location.
- https://github.com/zidanredha2/Capstone/blob/713c4885b523d6585f42d13c6160d5cd1ecc98c2/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

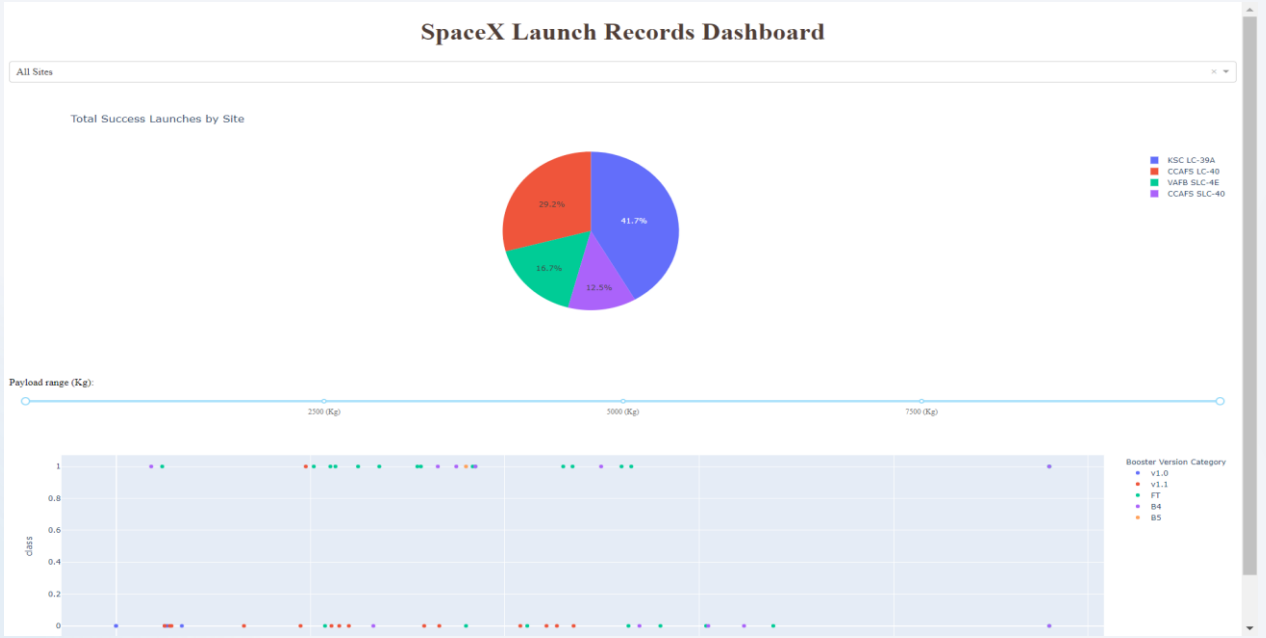
- Dashboard includes a pie chart and a scatter plot.
- Pie chart can be selected to show distribution of successful landings across all launch sites and can be selected to show individual launch site success rates.
- Scatter plot takes two inputs: All sites or individual site and payload mass on a slider between 0 and 10000 kg.
- The pie chart is used to visualize launch site success rate.
- The scatter plot can help us see how success varies across launch sites, payload mass, and booster version category.
- https://github.com/zidanredha2/Capstone/blob/5b8f17fcde46ba223a92c43ca526b4bd17ca65a9/spacex_dash_app.py

Predictive Analysis (Classification)

- We split our datasets and use fit and transform features and test and train datasets to predict the outcomes using various algorithms like knn models and trees.
- https://github.com/zidanredha2/Capstone/blob/5b8f17fcde46ba223a92c43ca526b4bd17ca65a9/Machine%20Learning%20Prediction.ipynb



Results

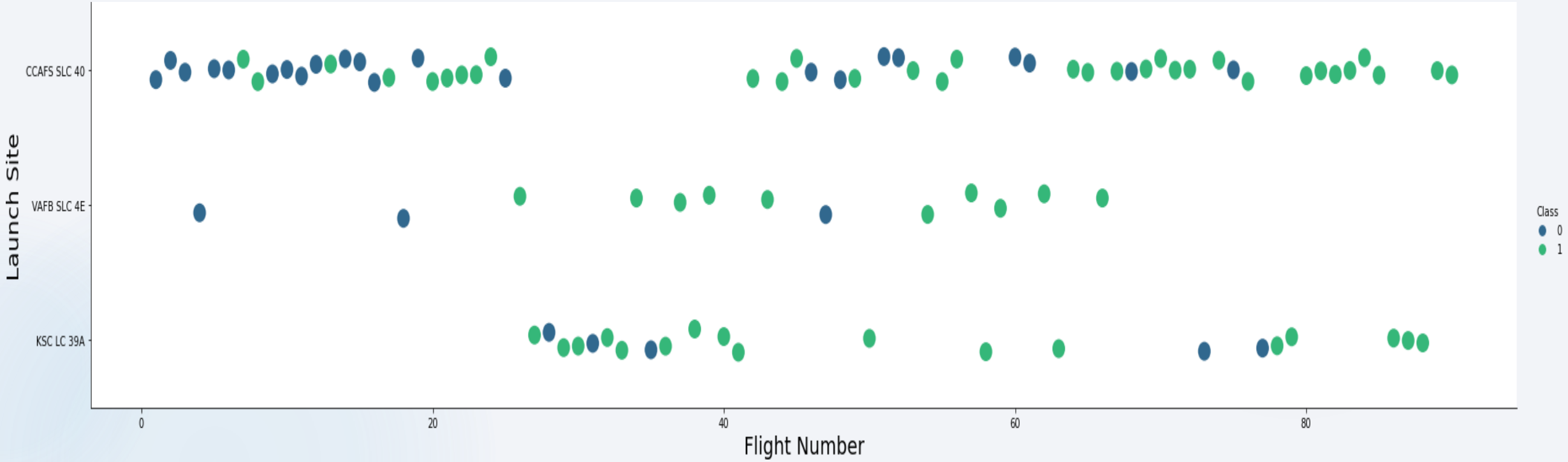


The results of EDA with visualization, EDA with SQL, Interactive Map with Folium, and finally the results of our model with about 83% accuracy.

Section 2

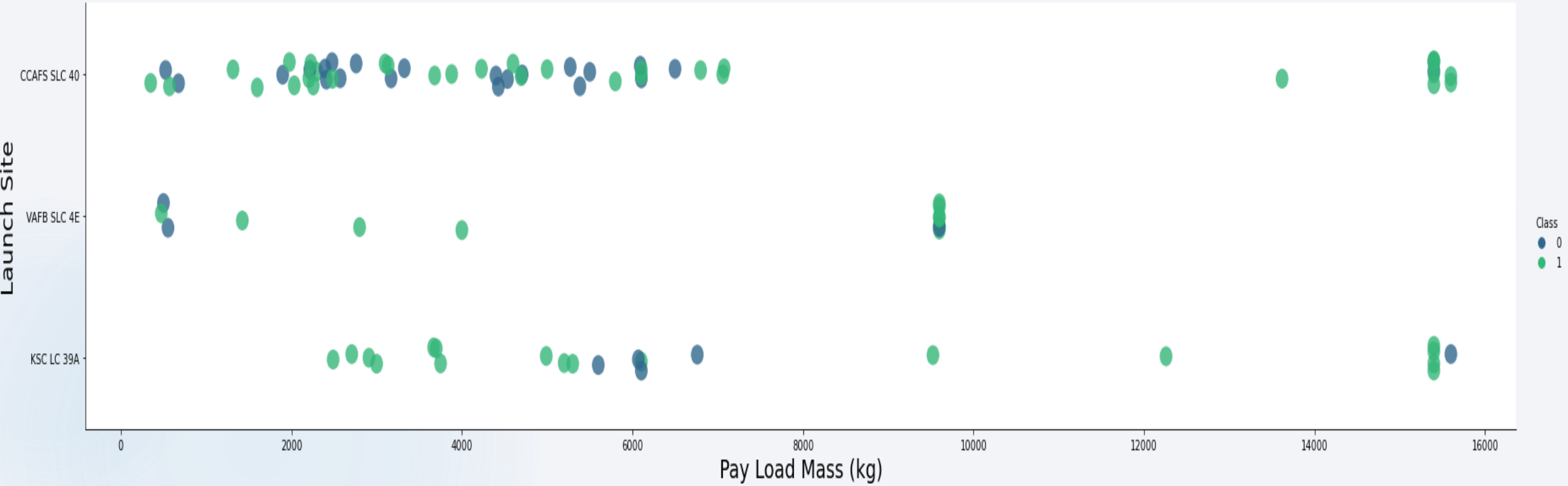
Insights drawn from EDA

Flight Number vs. Launch Site



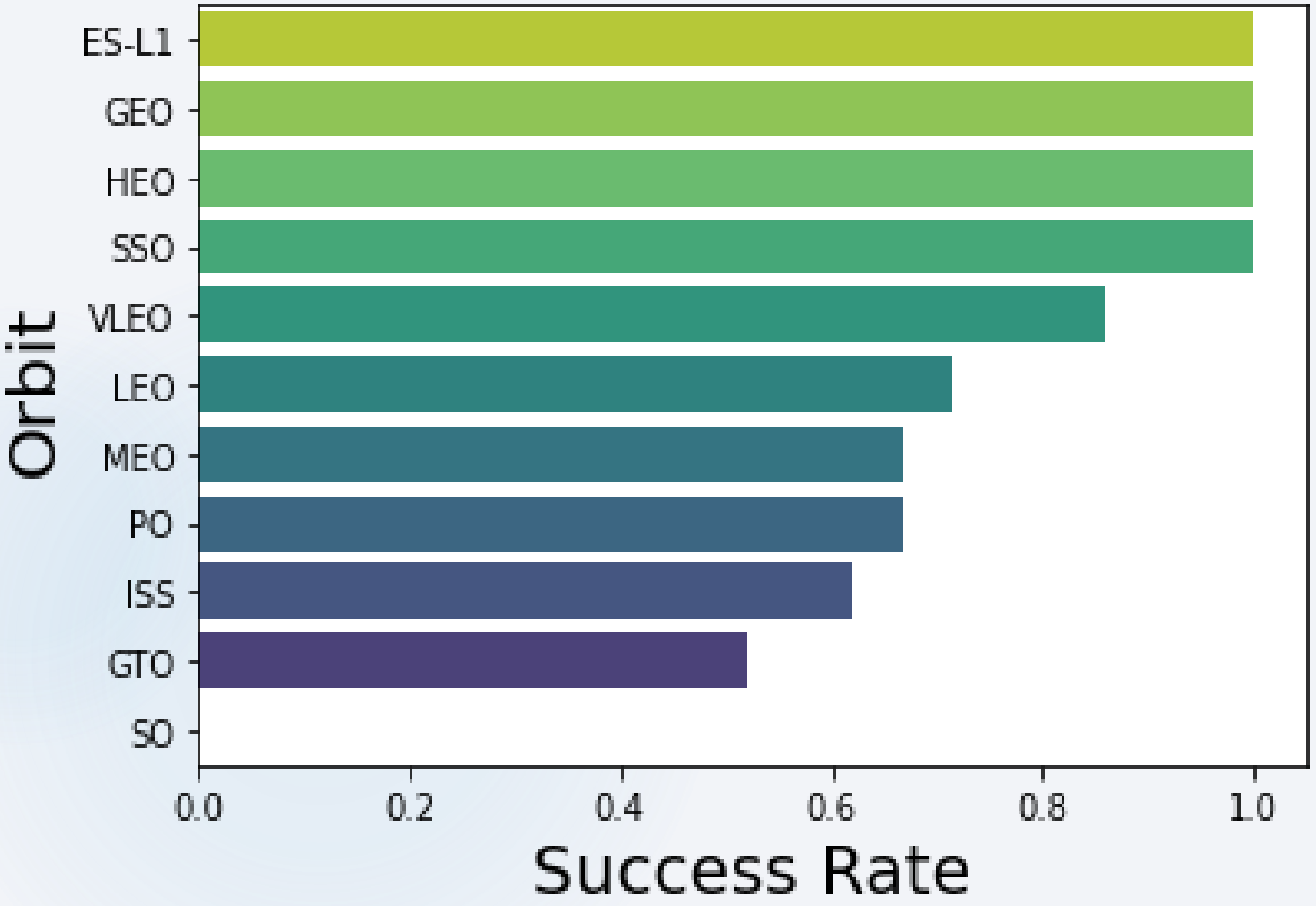
an increase in success rate over time (indicated in Flight Number). Likely a big Leap around flight 20 which significantly increased success rate. CCAFS appears to be the main launch site as it has the most volume.

Payload vs. Launch Site



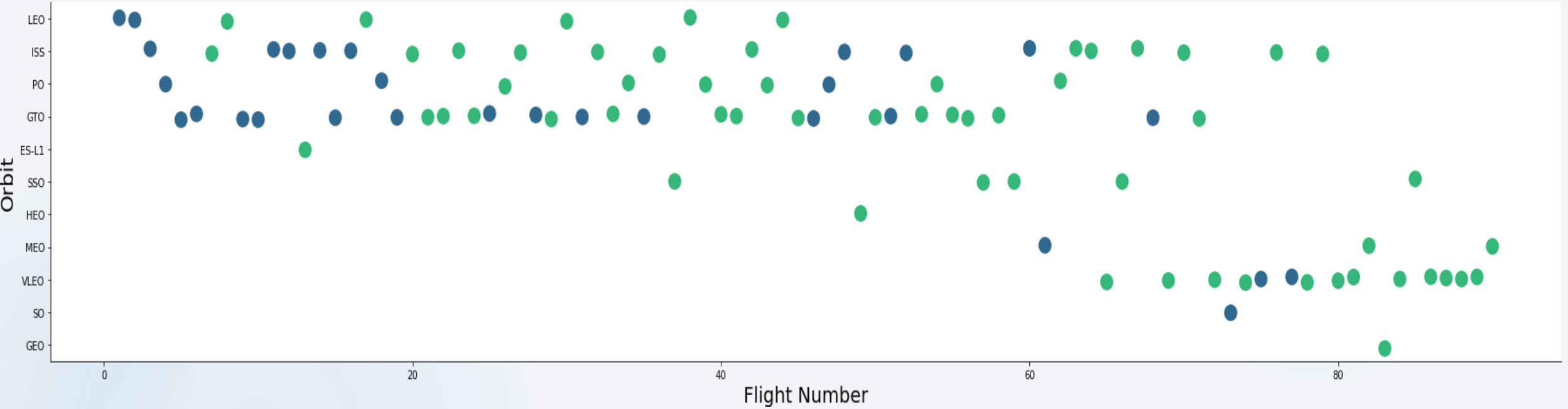
Payload mass appears to fall mostly between 0-6000 kg. Different launch sites also seem to use different payload mass.

Success Rate vs. Orbit Type



The given graph indicates the successrates of all the given flights with 0 being failure and 1.0 being 100 so the scale of successrate is supposed to be multiplied by 100 to know the percentage of successrate.

Flight Number vs. Orbit Type



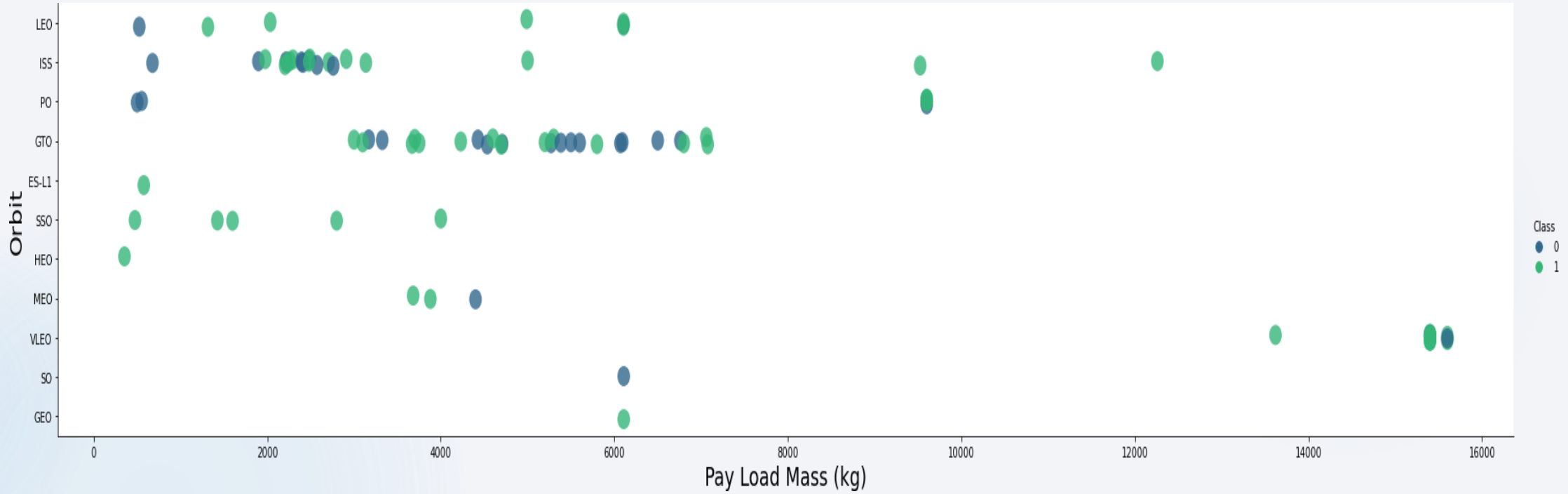
Launch Orbit preferences changed over Flight Number. Launch

Outcome seems to correlate with this preference.

SpaceX started with LEO orbits which saw moderate success LEO and returned to VLEO in recent launches

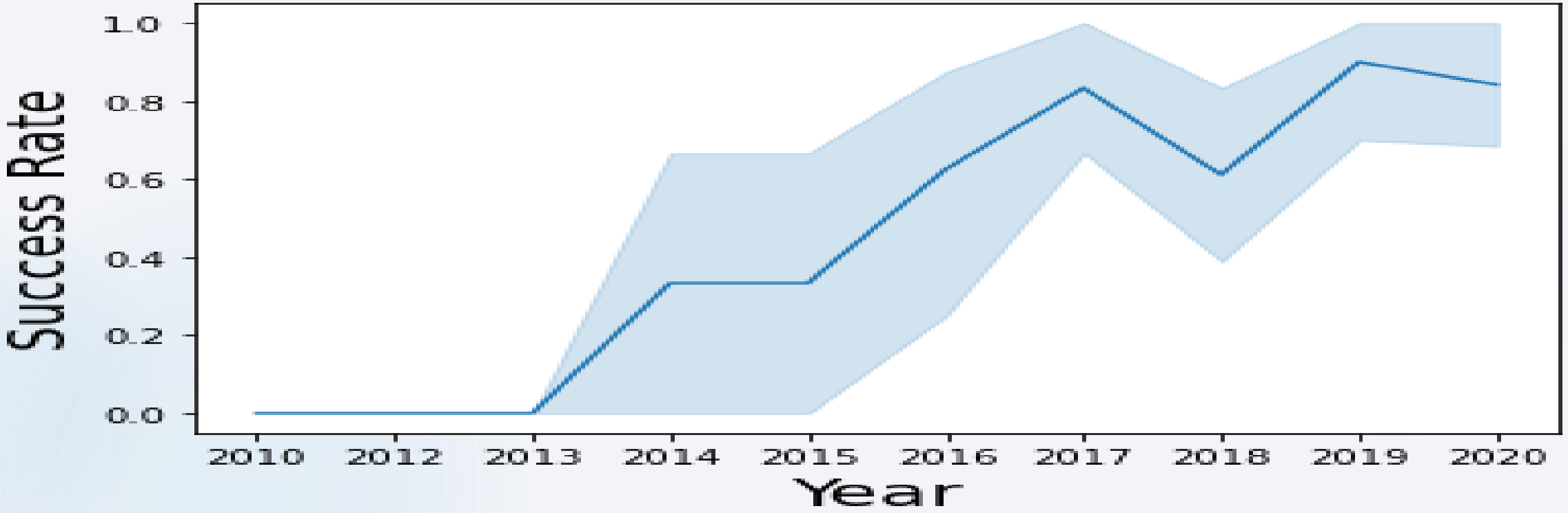
SpaceX appears to perform better in lower orbits or Sun-synchronous orbits

Payload vs. Orbit Type



- Payload mass seems to correlate with orbit.
- LEO and SSO seem to have relatively low payload mass.
- The other most successful orbit VLEO only has payload mass values in the higher end of the range.

Launch Success Yearly Trend



As we notice over the years the success rates have highly increased ever since 2013 and recent times touching almost 90%.

All Launch Site Names

In [4]:
```
%%sql
SELECT UNIQUE LAUNCH_SITE
FROM SPACEXDATASET;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2
Done.
```

Out[4]:

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

CCAFS SLC-40 and CCAFSSLC-40 likely all represent the same launch site with data entry errors.

CCAFS LC-40 was the

previous name.

Likely only 3 unique

launch_site values:

CCAFS SLC-40, KSC

LC-39A, VAFB SLC-4E

Launch Site Names Begin with 'CCA'

In [5]:
```
%%sql
SELECT *
FROM SPACEXDATASET
WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[5]:

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


```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) AS SUM_PAYLOAD_MASS_KG
FROM SPACEXDATASET
WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86
Done.
```

sum_payload_mass_kg
45596

Maximum payload is 45596 kgs which was sent to the International Space Station.

Average Payload Mass by F9 v1.1


```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD_MASS_KG
FROM SPACEXDATASET
WHERE booster_version = 'F9 v1.1'
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-8(
Done.
```

avg_payload_mass_kg
2928

The avg payload mass is 2928 KGS in Falcon 9.

First Successful Ground Landing Date


```
%%sql
SELECT MIN(DATE) AS FIRST_SUCCESS
FROM SPACEXDATASET
WHERE landing__outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81
Done.
```

first_success
2015-12-22

As we can see in the given output that the first successful landing took place on 22nd of December in 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000


```
%%sql
SELECT booster_version
FROM SPACEXDATASET
WHERE landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4001 AND 5999;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.database
Done.
```

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

The Output list in the given picture consists of the list of the successful drone ship landing with payload between 4000 and 6000 kgs.

Total Number of Successful and Failure Mission Outcomes


```
%%sql
SELECT mission_outcome, COUNT(*) AS no_outcome
FROM SPACEXDATASET
GROUP BY mission_outcome;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-
Done.
```

mission_outcome	no_outcome
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

As we can see in the obtained outcome the count of the number of successful and failure missions in dataframe of SpaceX.

Boosters Carried Maximum Payload


```
%%sql
SELECT booster_version, PAYLOAD_MASS__KG_
FROM SPACEXDATASET
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXDATASET);
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1
Done.
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

The output has the list of max. payload mass in kgs of each one of the boosters.

2015 Launch Records

- There were total of 7 launches of the falcon 9.
- Out of 7, only one ended in failure while the rest of the six had a successful mission outcome with some having failed booster recovery and few having controlled recovery while only one booster recovery was successful on the ground pad.

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


```
%%sql
SELECT landing__outcome, COUNT(*) AS no_outcome
FROM SPACEXDATASET
WHERE landing__outcome LIKE 'Succes%' AND DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing__outcome
ORDER BY no_outcome DESC;
```

```
 * ibm_db_sa://ftb12020:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg
Done.
```

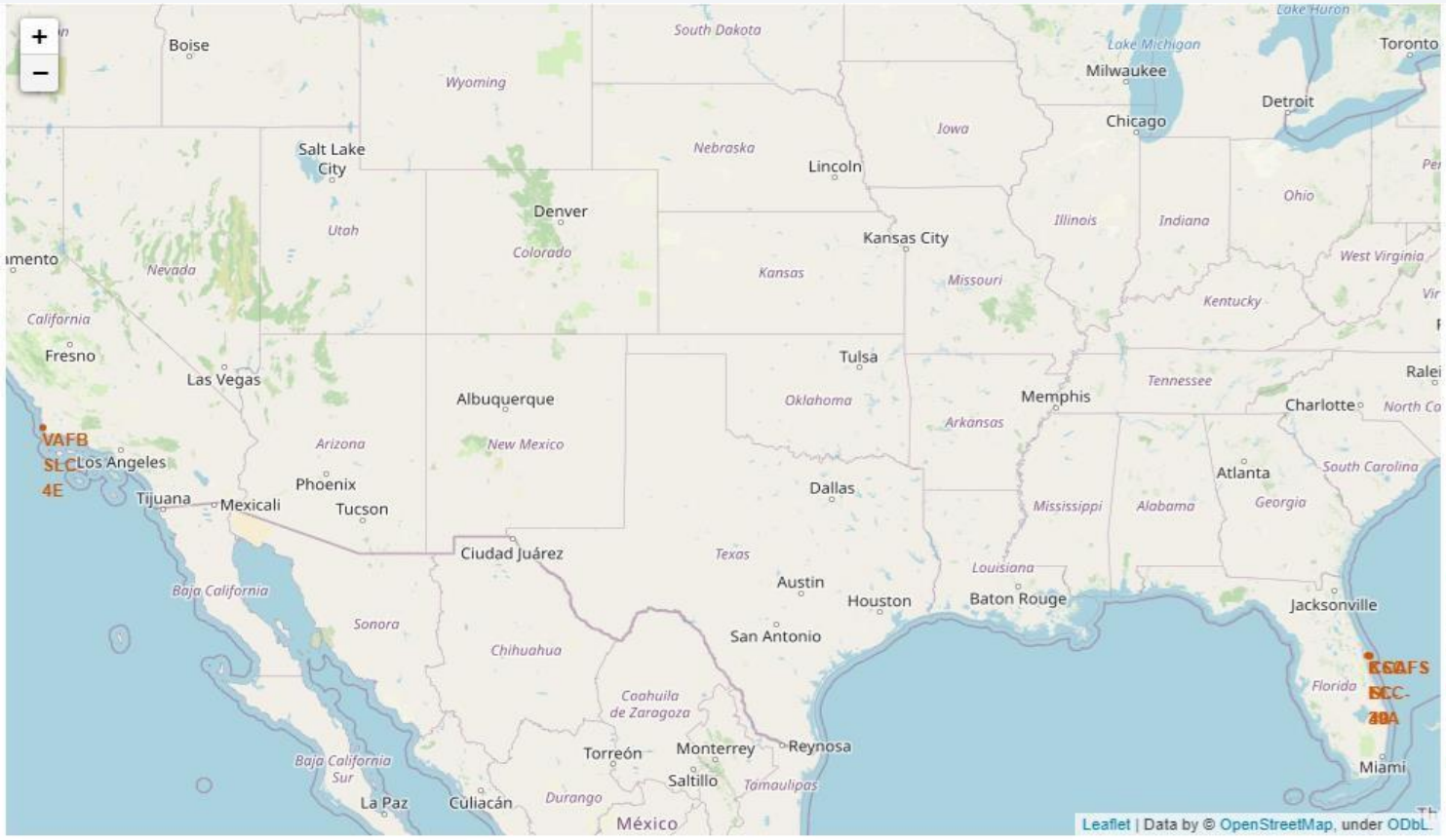
landing__outcome	no_outcome
Success (drone ship)	5
Success (ground pad)	3

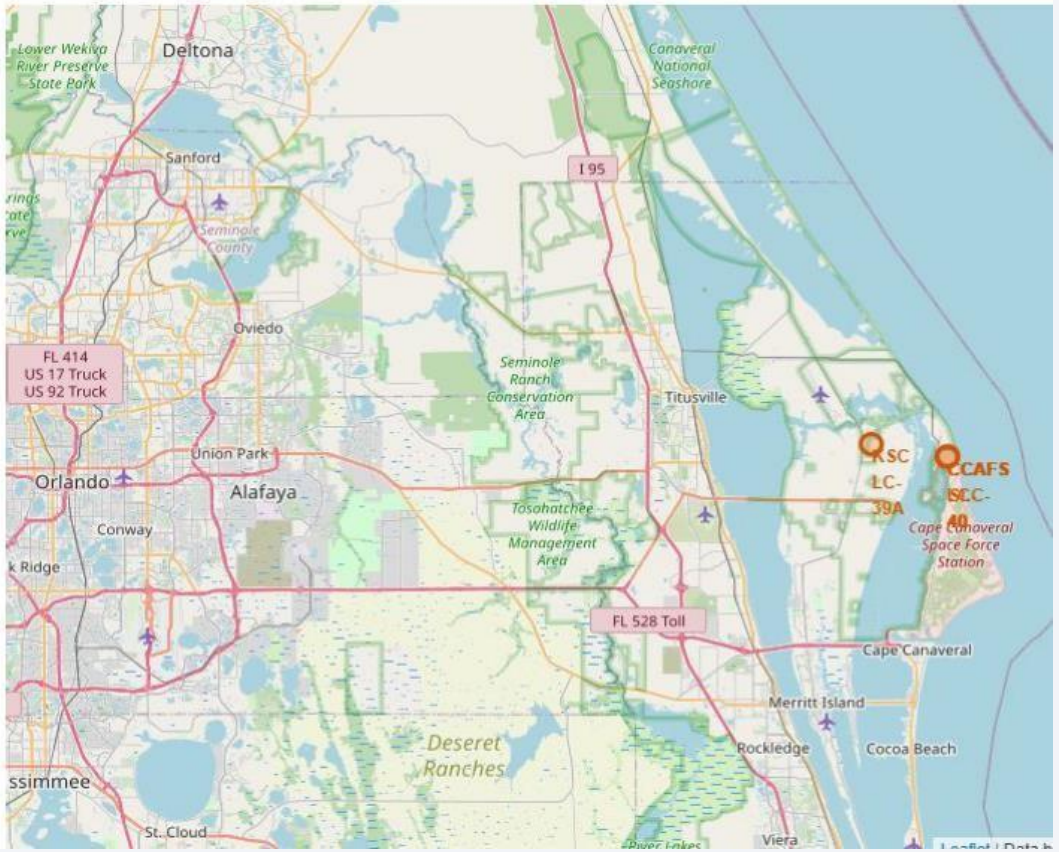
We can see from the output that 5 drone ship landing were successful while 3 ground pad landing was made in span of 2010 to 2017.

Section 3

Launch Sites Proximities Analysis

Launch Sites of Falcon 9



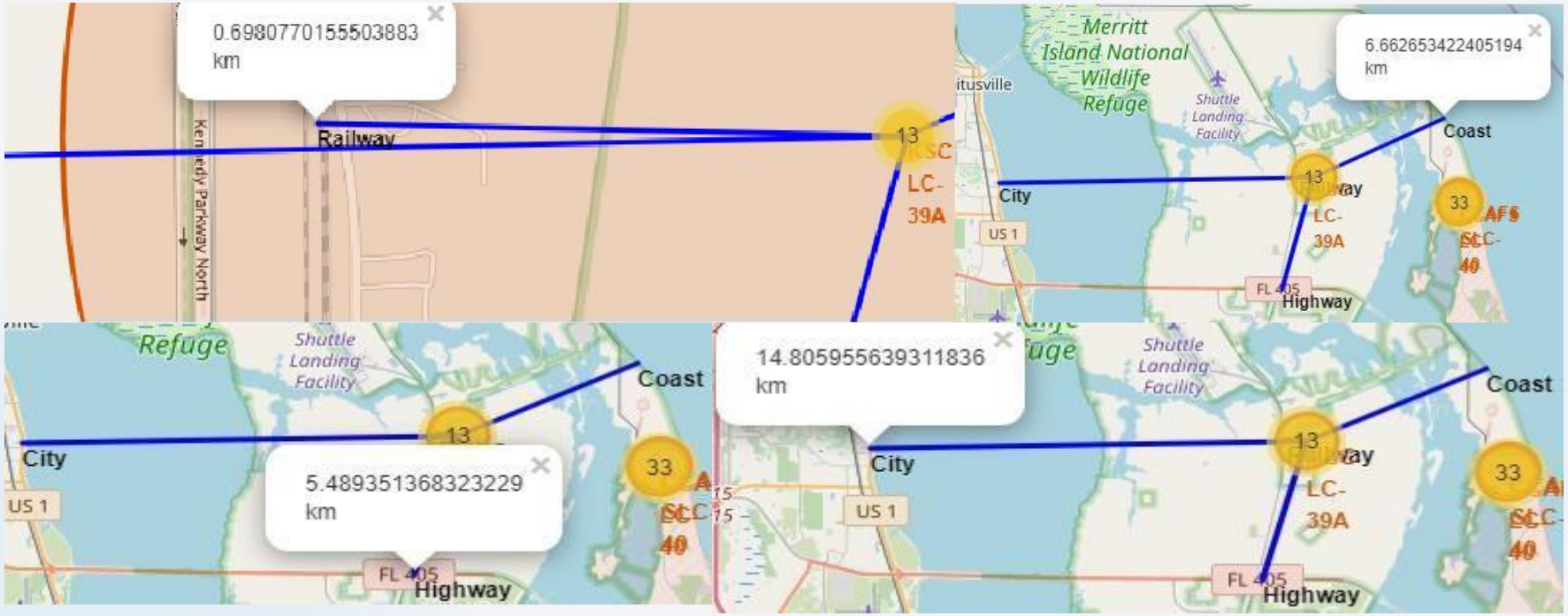


The above folium map is of USA and the right map consists of the launch sites of falcon 9 in Florida.

Color-Coded Launch Markers

The above map shows successful landings in green and failure landings in red.

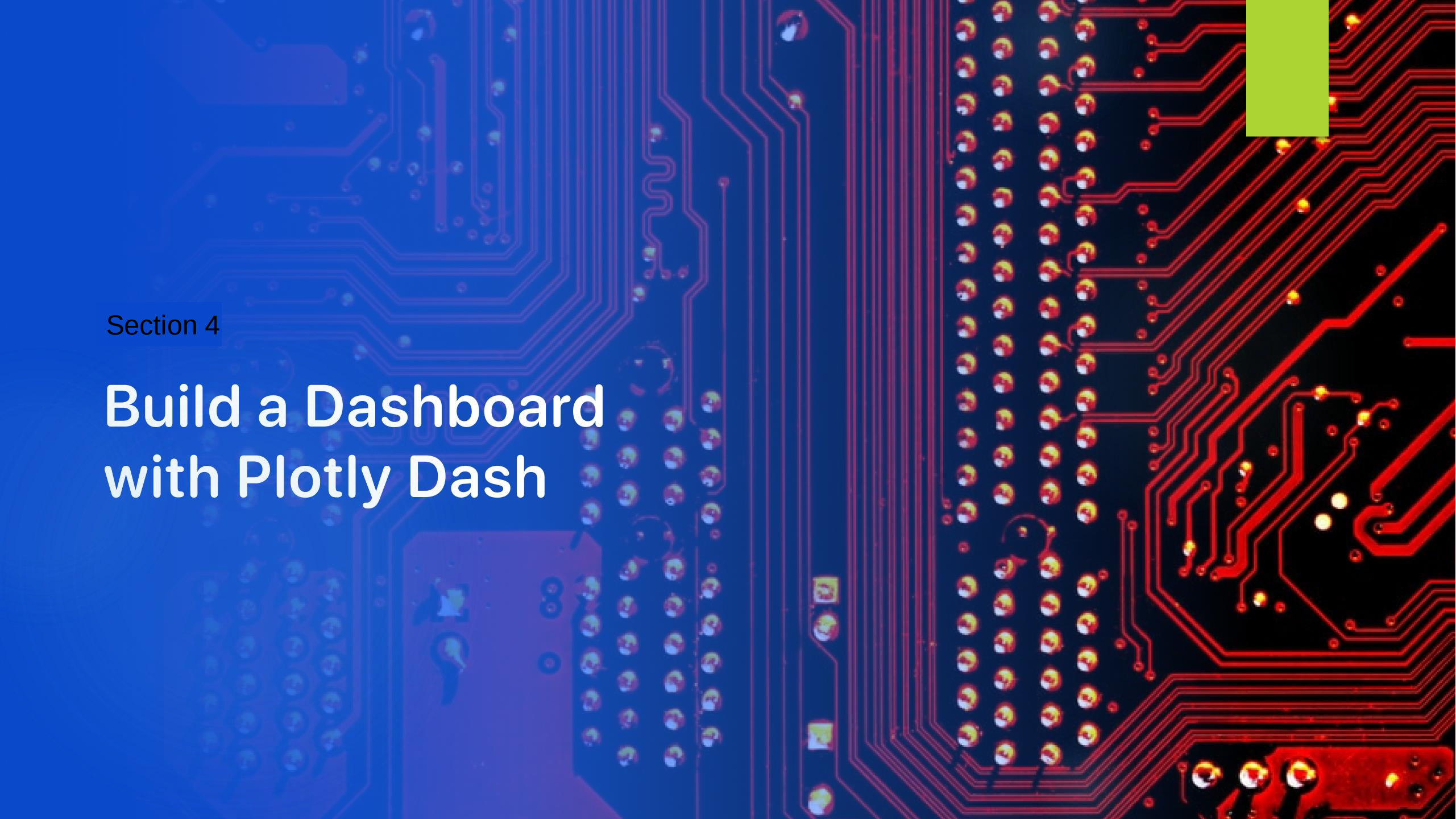
Key Location Proximities



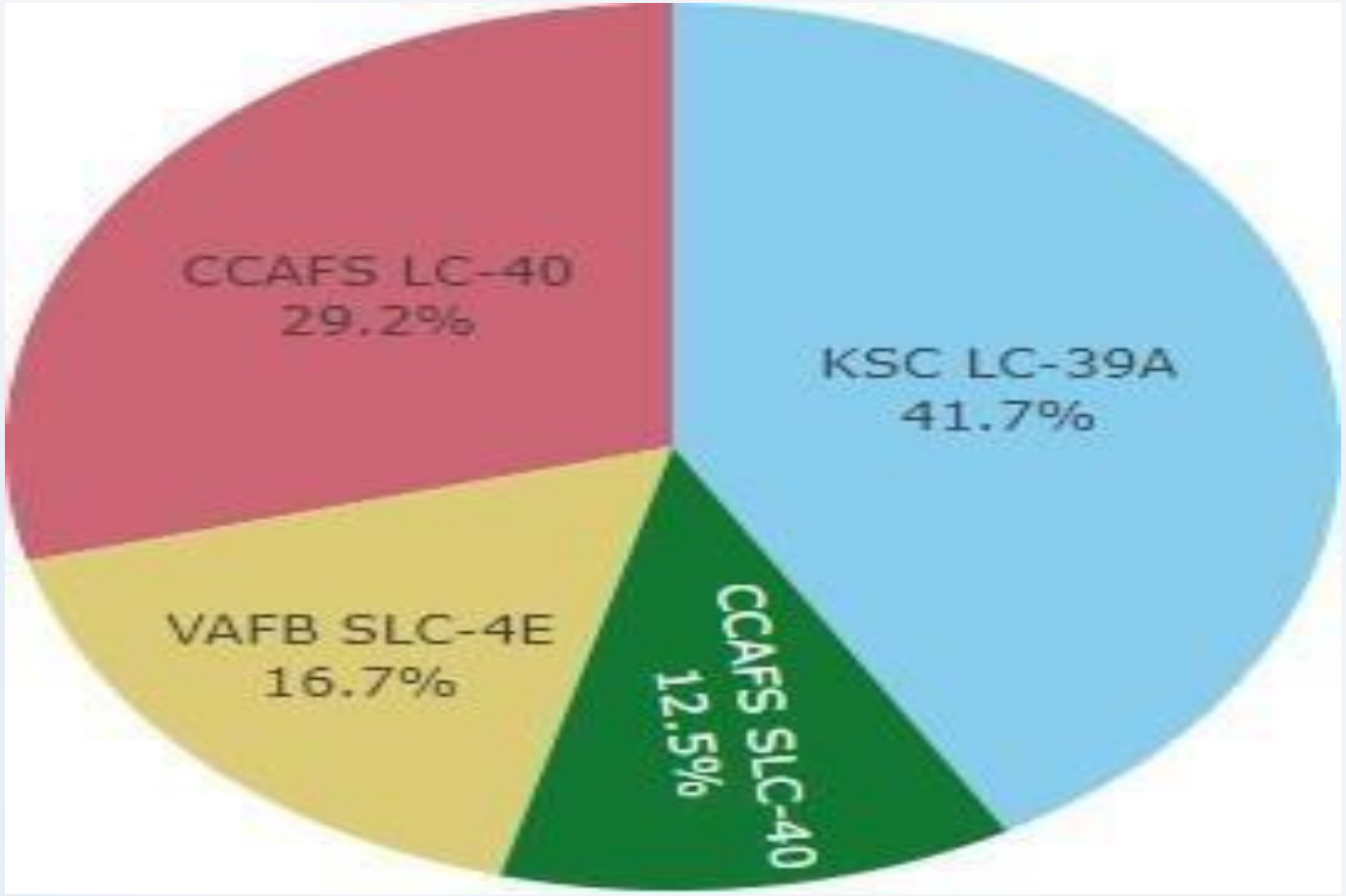
As we can see from the above pictures that the launch sites are close to highways and railway for supply and human transport. And the landing sites are close to sea so the failure landing pod can land in the sea than in popularity dense area.

Section 4

Build a Dashboard with Plotly Dash

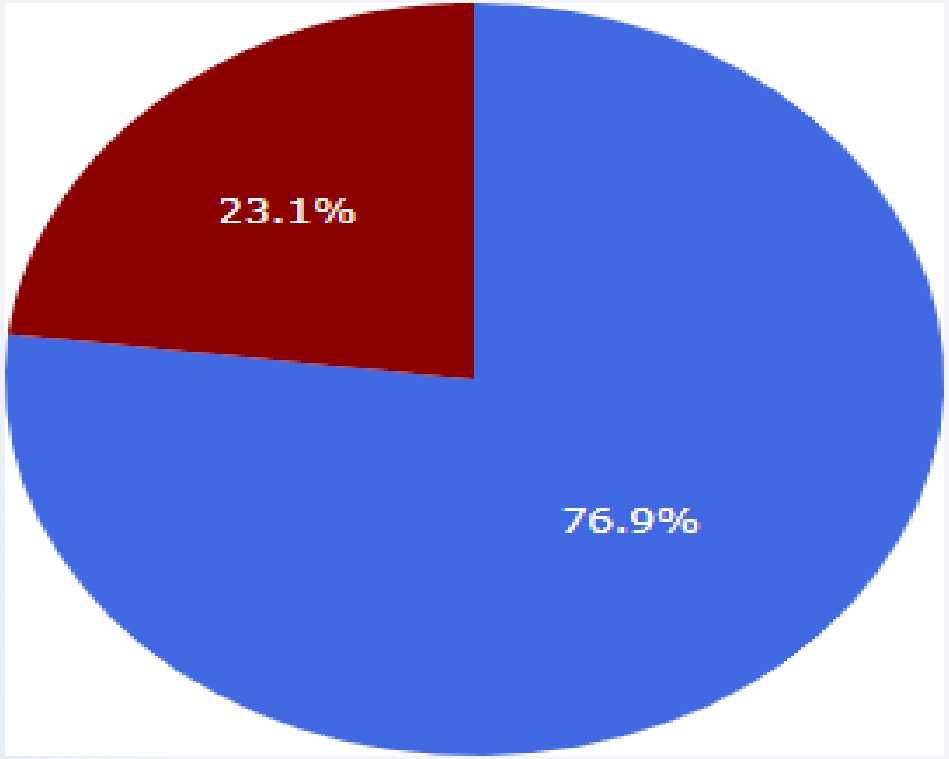


Successful Launches Across Launch Sites



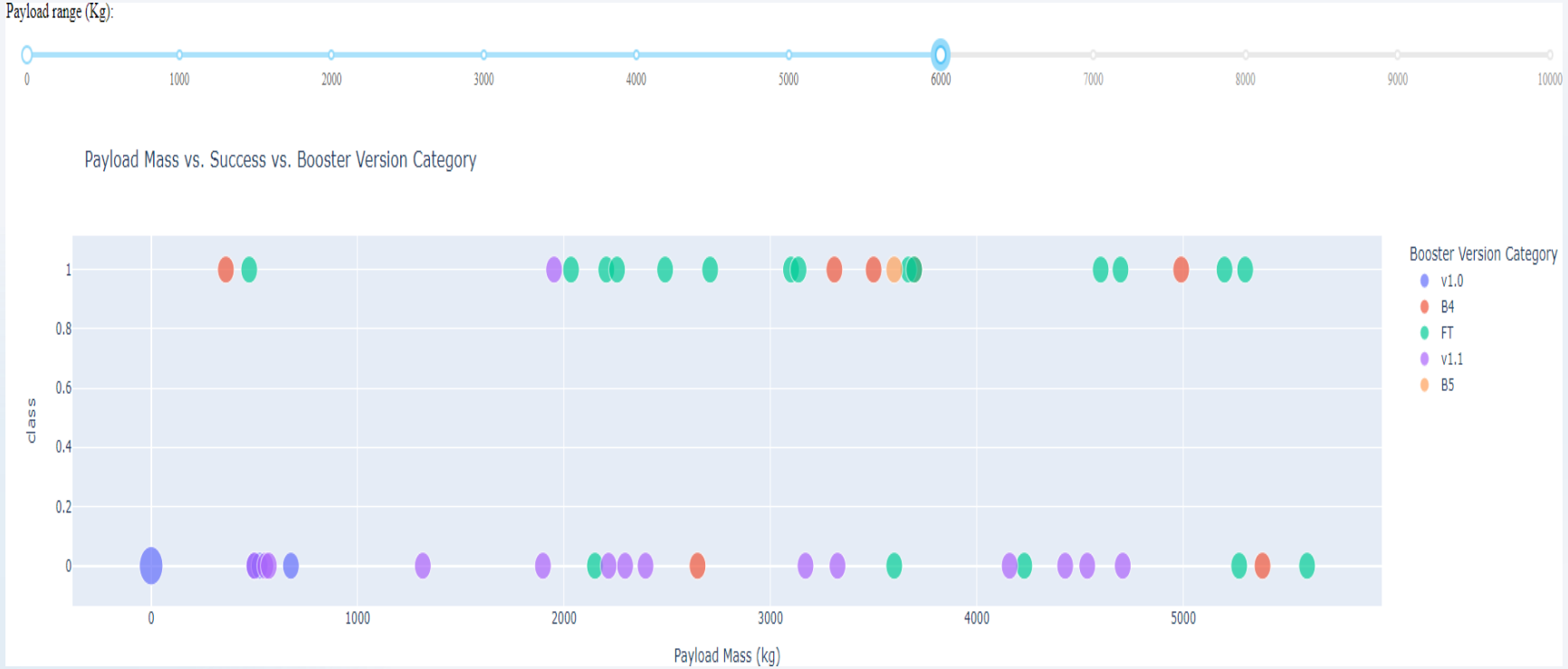
This pie chart above shows us the distribution of successful launch sites out of 4.

Highest Success Rate Launch Site



The above pie chart shows the probability of success of the launch site with blue color of success and brown of failure.

Payload Mass vs. Success vs. Booster Version Category

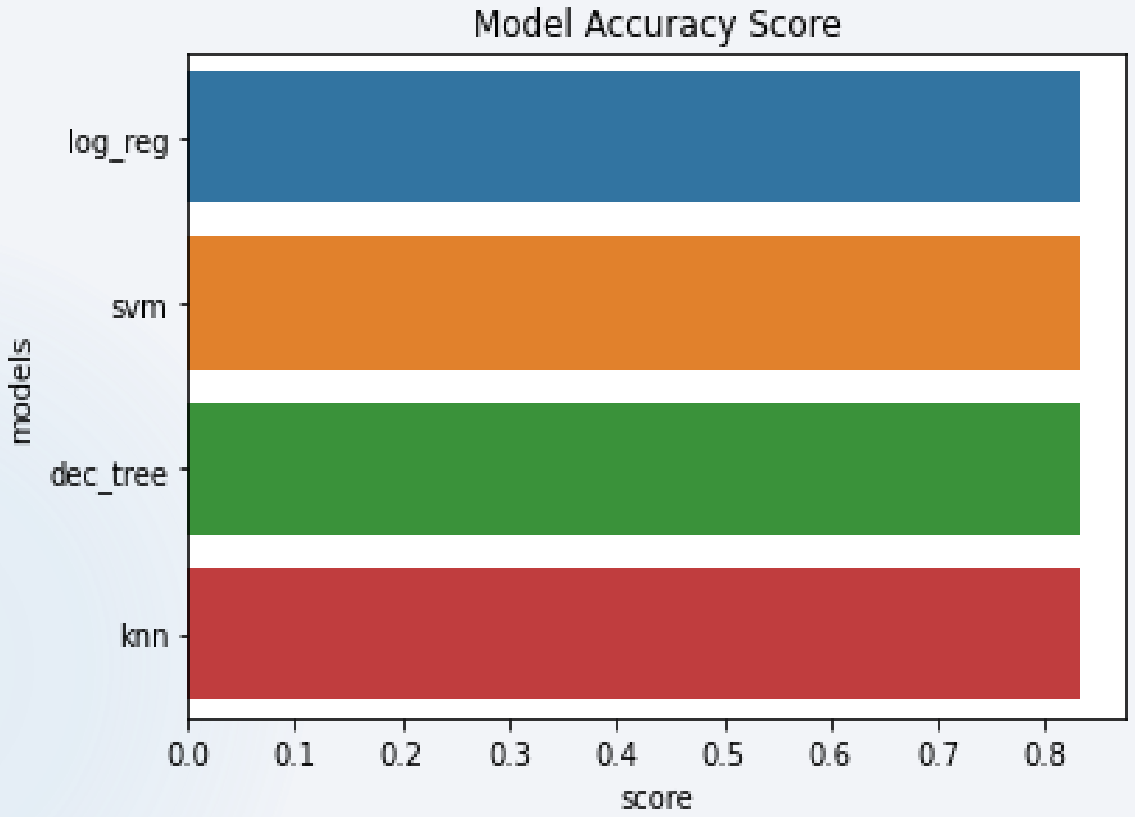


The above dashboard allows us to select payload upto 10000. We can see the surprisingly the 0 payload has 2 failure landings.

Section 5

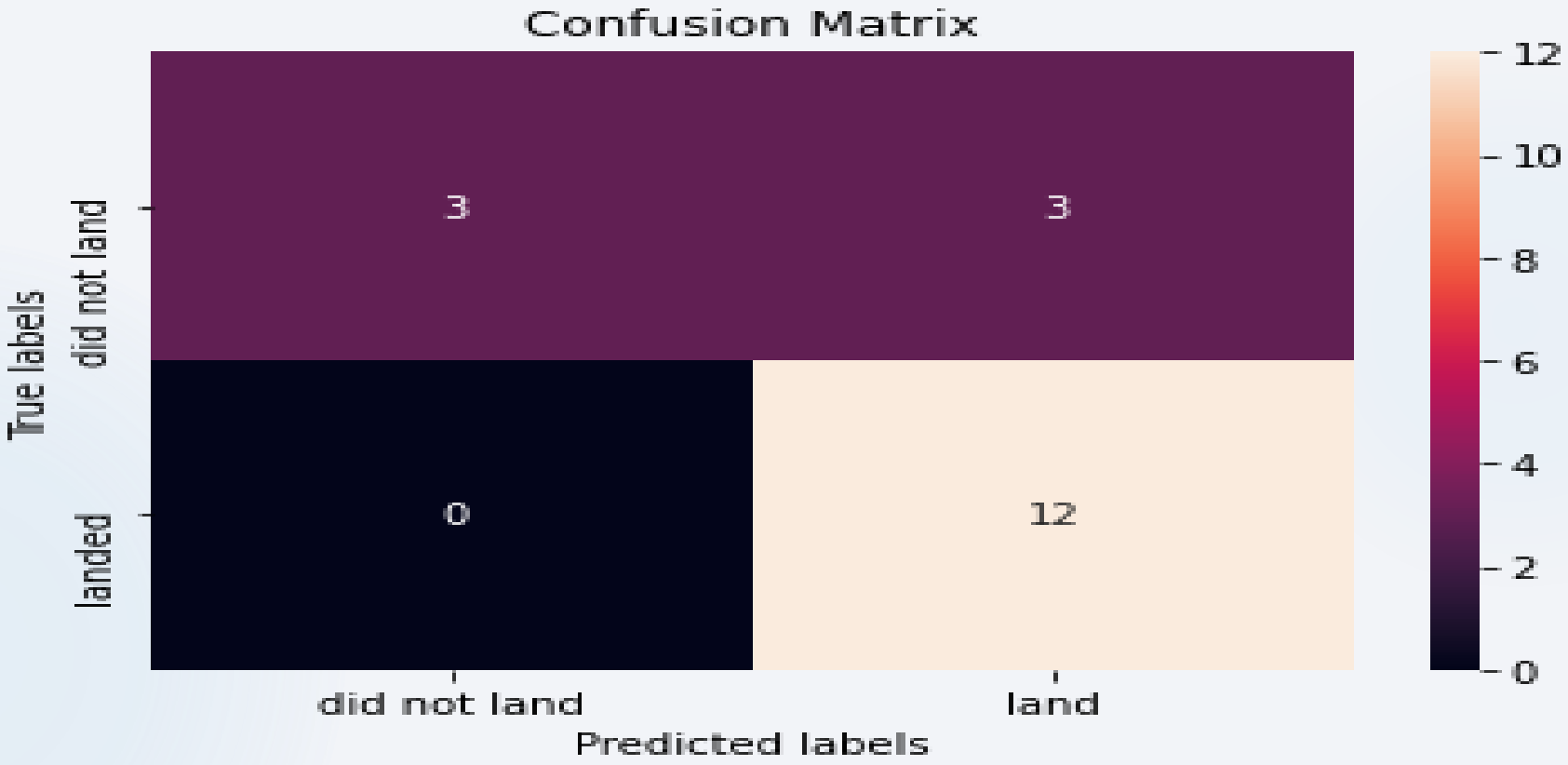
Predictive Analysis (Classification)

Classification Accuracy



We can note that all the models we have texted has the exact same accuracy of 83.3%.

Confusion Matrix



We can note that all models have the same predictions.

Conclusions

- We have successfully created a ML model for SpaceY to go up against SpaceX.
- SpaceY can use our data analysis the tweak and make changes to their plans if any.
- We have also made an interactive dashboard with DASH so SpaceY can calculate their chances upto 10000 kgs.
- Our models have performed with an accuracy rate of 83.3%.
- The more and the better data we collect, the better our machine learning models shall perform.

Appendix

- Github URL,

https://github.com/zidanredha2/Capstone

Thank you!