IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Adonis Miliano
08/30/2024

Outline

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

Executive Summary

- **Goal:** To learn which features have the largest impact on successful launch landings for SpaceX.
  - *Data Collection* from SpaceX launch data.
  - *Data Cleaning and Preprocessing* was completed on Jupyter Notebooks using Python.
  - *Exploratory Data Analysis* through data visualization.
  - *Model selection* using various machine learning algorithms.
- Result Summary:
  - Most successful launches were from launch site KSC-LC 39A
  - Payload Mass and Booster Versions were impactful when determining launch landing.

Introduction

Background:
- Our client "SpaceY" is starting a rocket launch company
- Company SpaceX spends less than 40% in rocket launches compared to other companies due to reusing their first stage rocket.

Problem to Solve:
- We will find out what **successful** launch landings have in common to increase the probability of a successful launch landing for SpaceY.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected via web-scraping from Falcon 9 Launch data on wikipedia.
- Perform data wrangling
  - Data was checked for null values.
  - New "class" column for a numerical representation if that launch was as success or failure.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology (Continued)

- Perform predictive analysis using classification models
  - ***Building the Model:***
    - First we needed to create a testing and data set
    - Multiple algorithms used to test (Logistic Regression, SVM, KKN etc.)
  - ***Tuning the Model:***
    - Multiple parameters were tested using GridSearchCV and training data.
    - Different parameters tested for each model.
  - ***Evaluating the Model:***
    - The parameters yielding the highest accuracy scores were used for said model.
    - Compared the accuracy score of the different machine learning algorithms to ascertain which algorithm would yield the most accurate results.

Data Collection - SpaceX API

- Data was initially collected using SpaceX REST API.
- The API allows us to exchange data between SpaceX client and our Python Applications.

Request and **Parse** data using GET request

↓

Filter the data frame to only include Flacon 9 Launches

↓

Begin Data Wrangling

Github Link to SpaceX API Call Code: Completed SpaceX API

Data Collection - Scraping

Falcon 9 Launch Records were **web scraped** using *"BeautifulSoup"*, a Python library for data collection.

→

Extracted column and variable names.

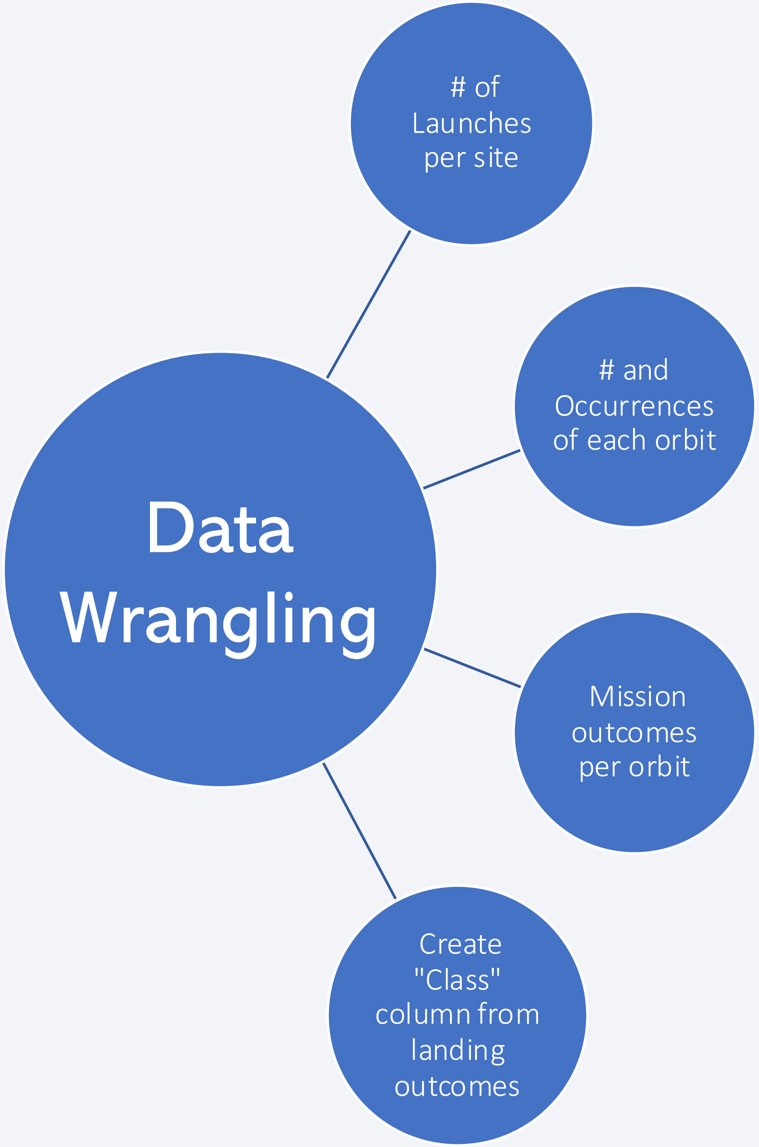
→

Created a data frame for preprocessing.

Historical data was collected from a wikipedia page titled "List of Falcon 9 and Falcon Heavy launches"

Github Link to Web Scraping Code: Completed Web Scraping Code

Data Wrangling



- The data was processed using Pandas and Numpy libraries on Python.
    - Caclulated the number of launches on each site.
    - Calculated the number and occurrences of each orbit.
    - Calculate the number and occurrences of mission outcome of the orbits.
    - Create a landing outcome label from "Outcome" column called "Class".

GitHub Link: Data Wrangling

EDA with Data Visualization

- **Scatter plots** were utilized to visualize any potential relationship between two variables.
- **Pie Charts** were used for a visual understanding of the success ratio for launch sites.
- **Line plot** was used to display any change that could have occurred during the years the launches were performed.
- **Bar Graph** to compare the success rate of different oribt launches.

GitHub Link: EDA with Data Visualization

EDA with SQL

- SQL Queries Performed:
  - Display the names of the unique launch sites in the space mission
  - 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 total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass.
  - Rank the count of landing outcomes between the 2010-06-04 and 2017-03-20.

GitHub Link: [EDA with SQL]()

Build an Interactive Map with Folium

- **Markers** were created on the folium map to highlight the labels for name of the launch cite.
- **Circles** were were created on the folium map to designate the launch cite locations for the Falcon 9 boosters.
- **Lines** made it easier to visualize the distance between launch sites and important landmarks like city centers, railroads or coast lines.

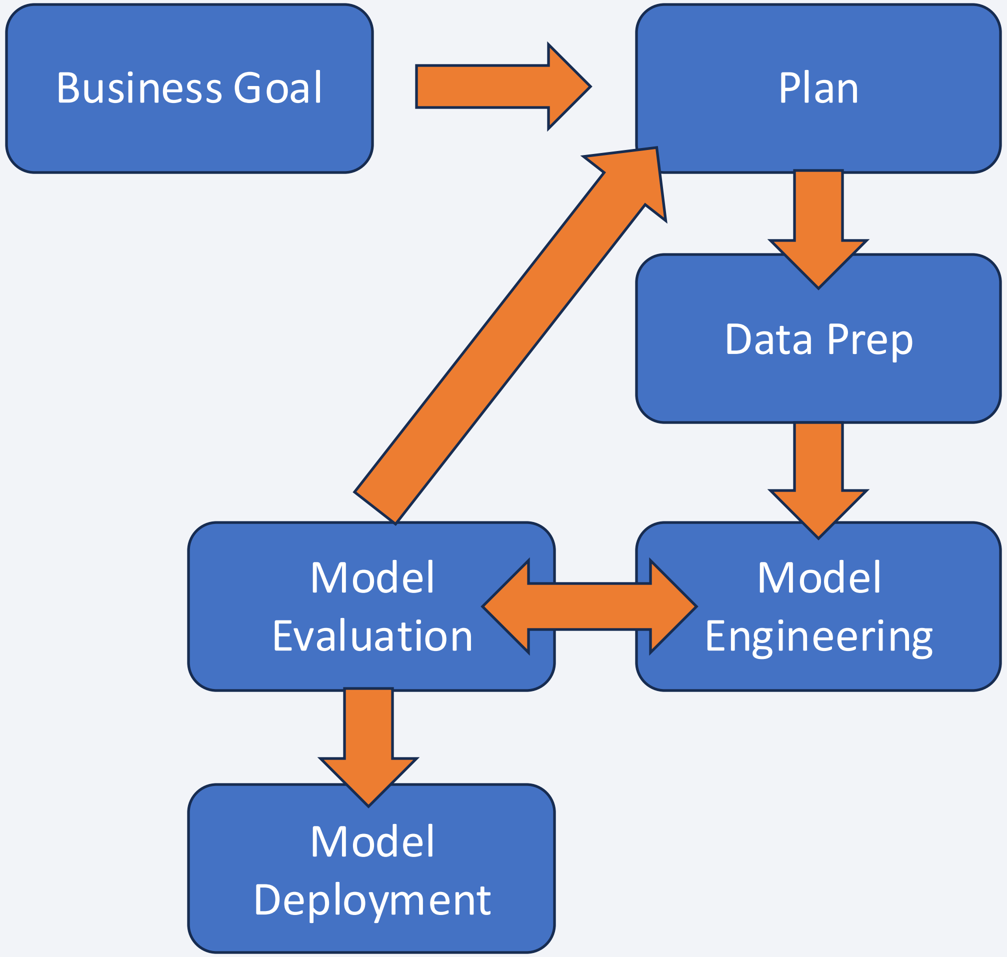
- GitHub Link: Interactive Map with Folium

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Added a dropdown to select a launch site to see the success ratio for the launch site.
- Selecting "All" would show all launch sites and show the ratio of successes.
  - A pie chart will then generate depending on the selection
- Added a slider to select the payload mass range to see how the mass could effect landing success.
  - A scatter plot would be generated depending on the range.
- GitHub Link: Plotly Dash

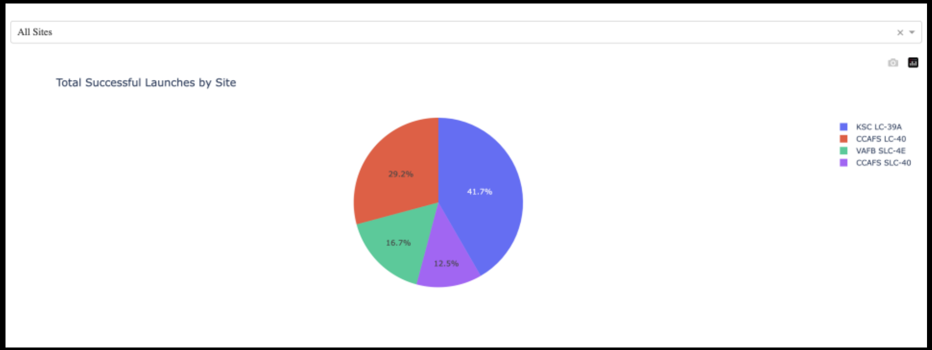
Predictive Analysis (Classification)

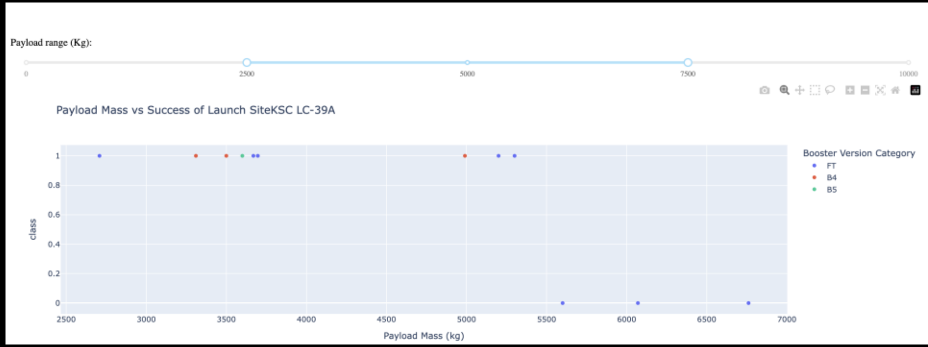
- Building the Model:
  - First we decide on the type of Model
  - SVM, LR, KNN, Tree etc.
- Evaluating the Model:
  - Parameters were tested using GridSearchCV
- Improving the Model:
  - Improving the model to optimize for out of sample testing
- Best Performance:
  - Score metric was the ***Accuracy Score*** of the model, comparing the predicted outcomes to the expected outcomes.
- GitHub Link: Predictive Analysis



Results

- We can see the launch site that had the most successful landings.
- We can also see that there are specific Payload Mass ranges that have more successful landings.
- With our predictive analysis, all models had an accuracy score above 80% when predicting the likelihood of successful launches given a set of features.



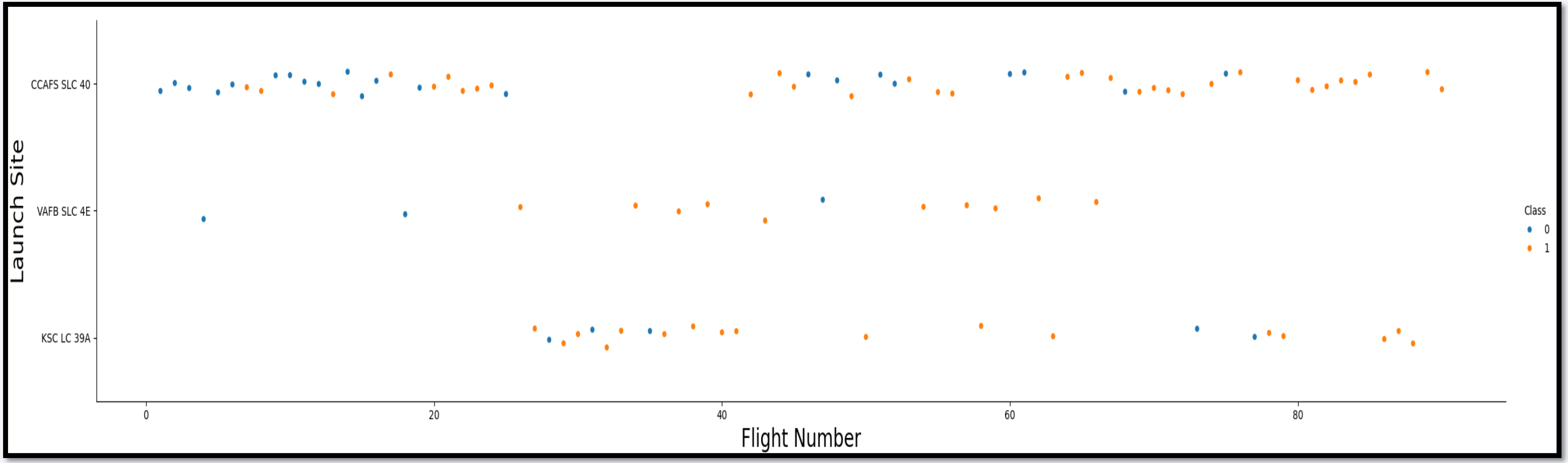


Section 2

Insights drawn from EDA

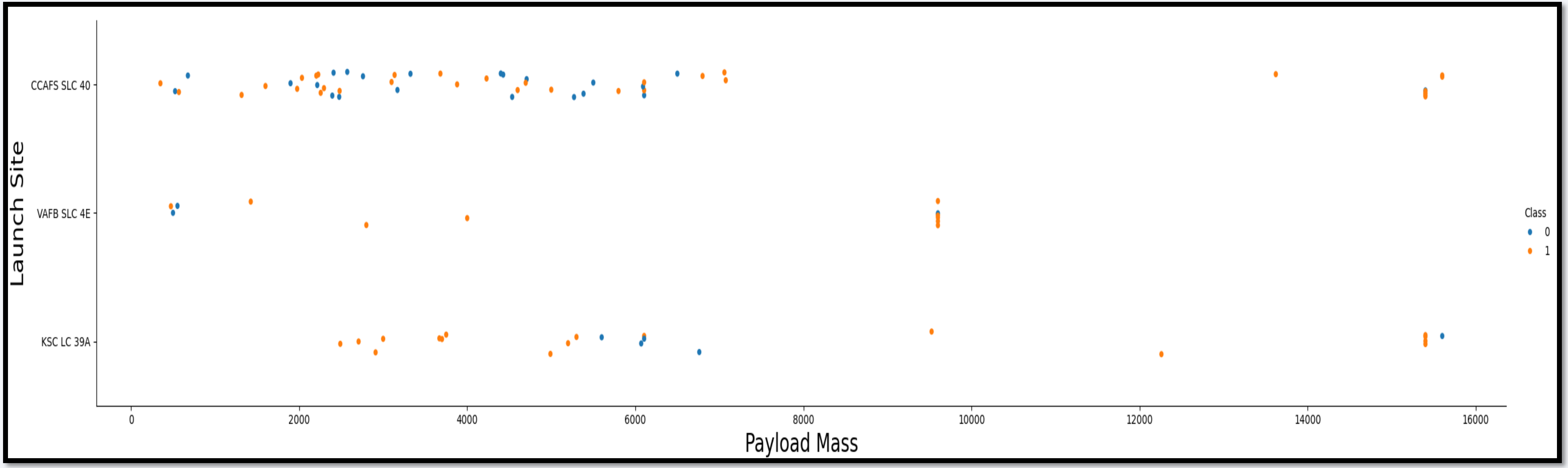
Flight Number vs. Launch Site

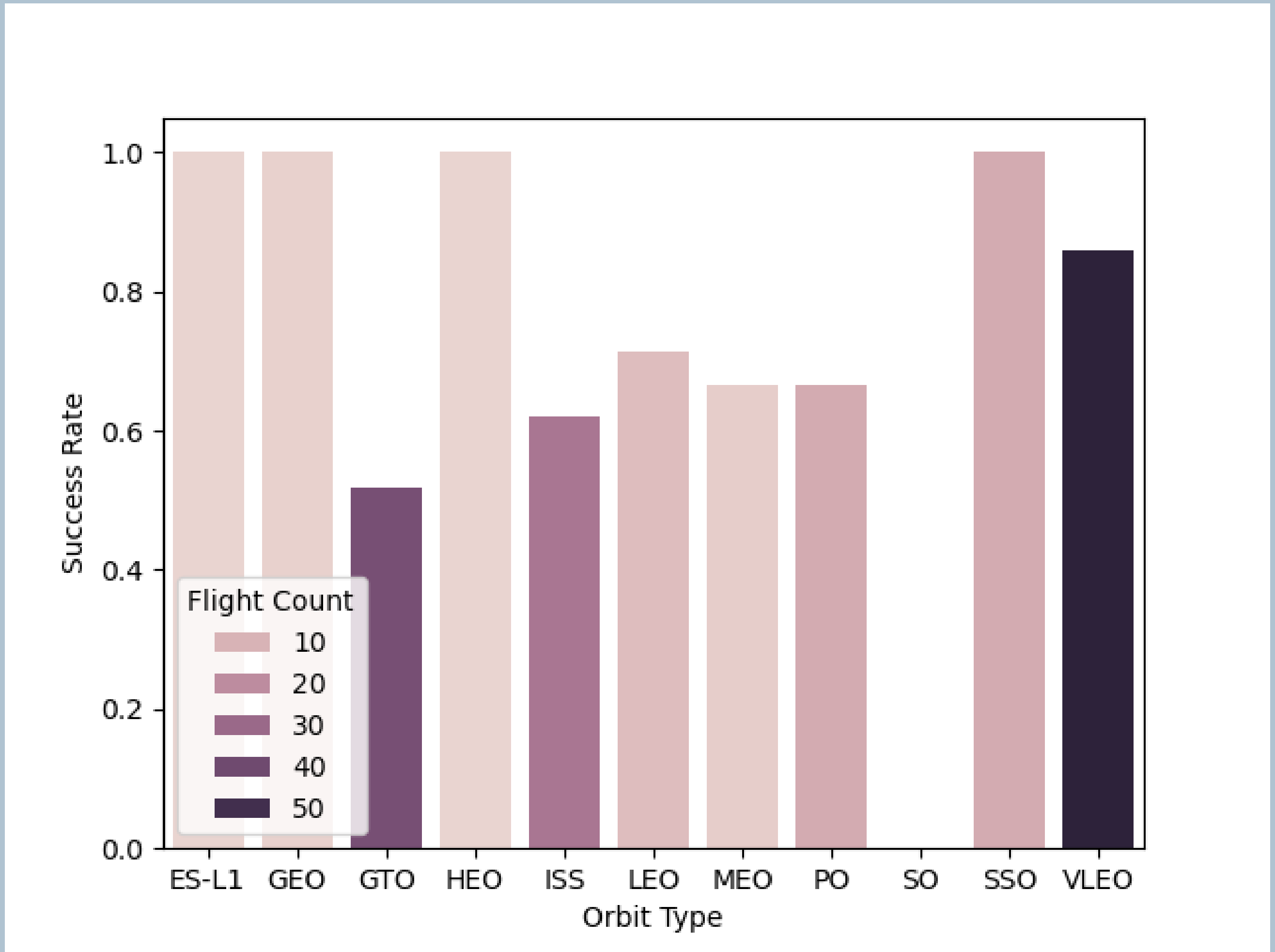
We can see in the image below that as the Flight Number increased, the number of successful launches increased, especially for CCAFS SLC 40 launch site.



Payload vs. Launch Site

We can see here that VAFB SLC 4E doesn't have a payload mass over 10,000Kg, while the other two launch sites have payload masses even around 16,000Kg.



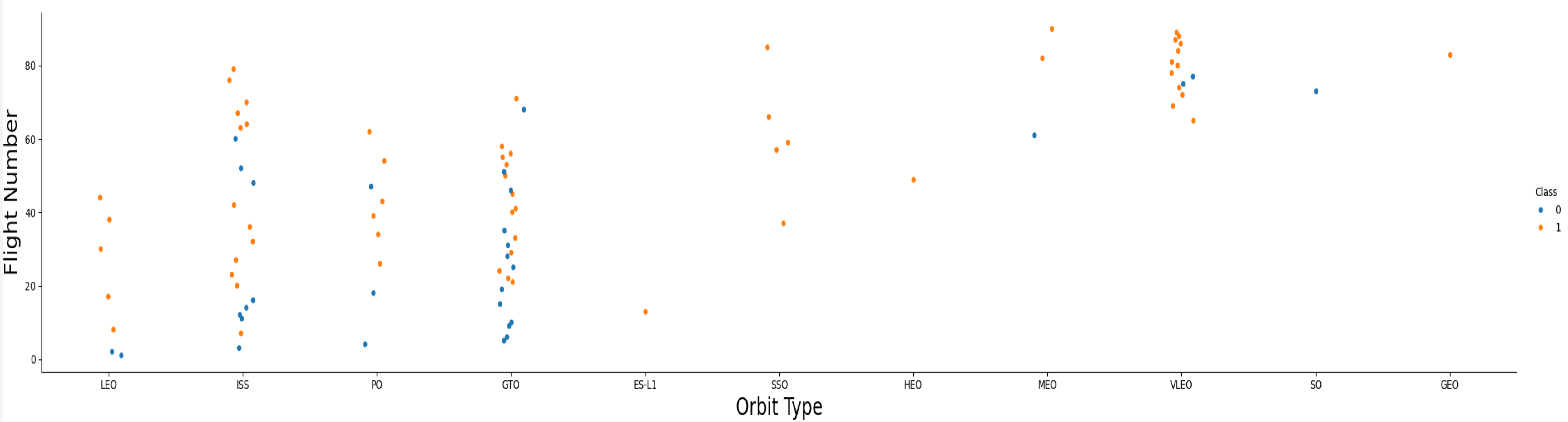


Success Rate vs. Orbit Type

- *We can see in the bar chart that the orbit type with the highest count of launches with a high success rate is VLEO.*
- *Although HEO and other orbits have a higher success rate, they have lower launch counts.*

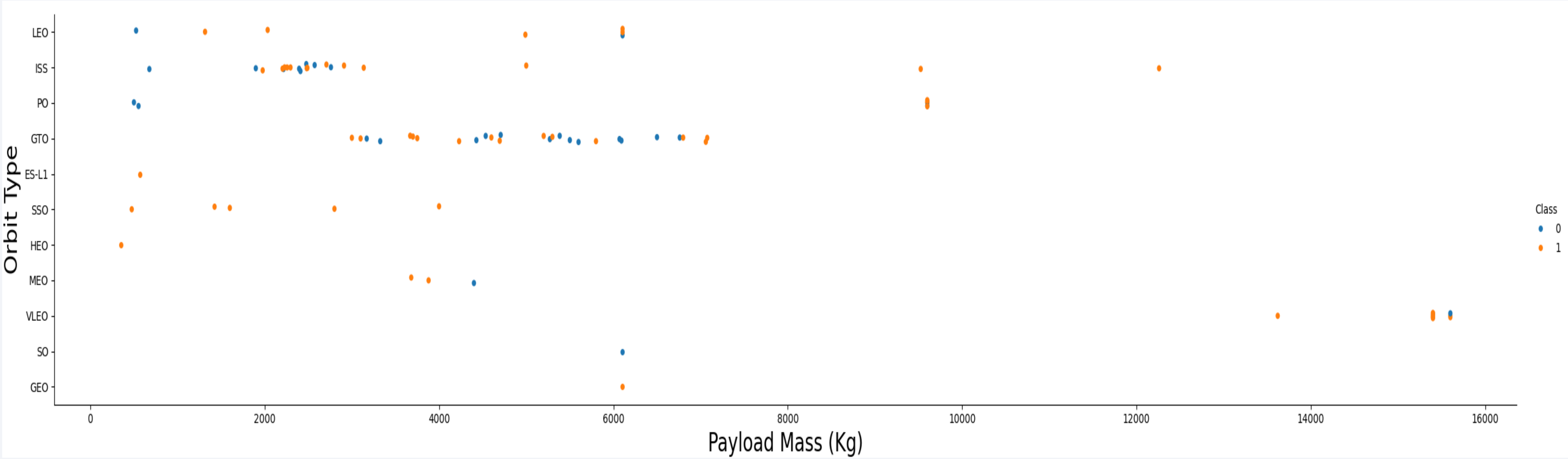
Flight Number vs. Orbit Type

There seems to be increased success for LEO orbit as flight number increases; however, for GTO orbit there is no clear relationship between fligh number and success rate.



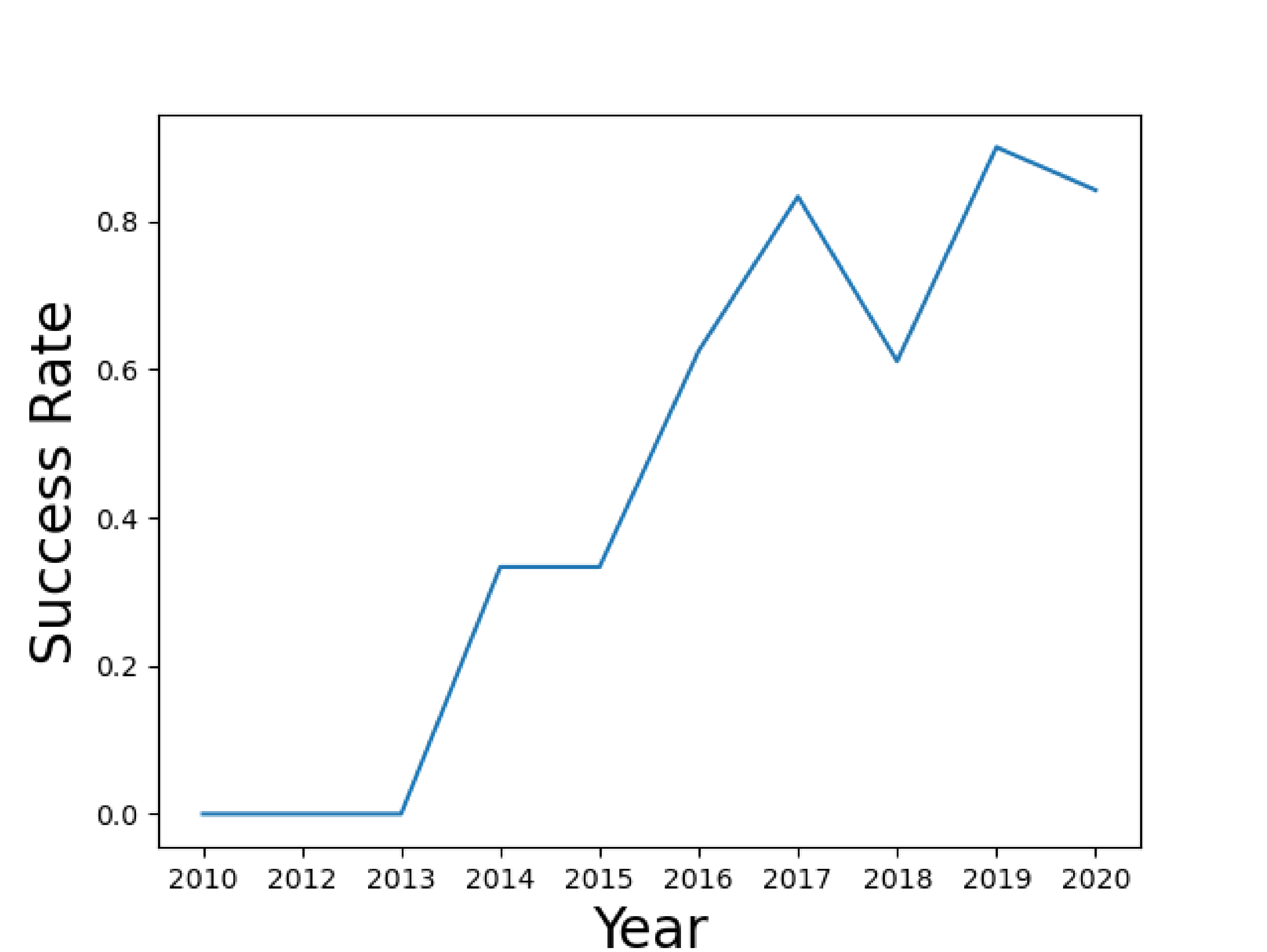
Payload vs. Orbit Type

With heavy payloads the successful landing rate are more prevalent in Polar, LEO, and ISS. However, there is not a strong correlation between payload mass and success rate for GTO.



Launch Success Yearly Trend

We can see here that there was increase in Launch Succes from 2010-2017. A dip in success from 2017-2018 but increasing once again after 2018.



All Launch Site Names

Names of Each Launch Site:

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

Task 1

Display the names of the unique launch sites in the space mission

[8]:
```
%%sql
SELECT DISTINCT("Launch_Site")
FROM SPACEXTABLE
```

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

[8]:

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

Launch Site Names Begin with 'CCA'

- Here are 5 records that start with 'CCA'
- One of many methods to filter our data and investigate specific rows.

Task 2

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

[9]:
```
%%sql
SELECT *
FROM SPACEXTABLE
WHERE "Launch_Site" LIKE 'CCA%'
LIMIT 5
```

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

[9]:

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

Total Payload Mass

Here we see the total payload mass carried by Nasa is approx. 45596, giving us an idea of the overall mass carried by NASA throughout the years.

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

[10]:

```
%%sql
SELECT "Customer", SUM(PAYLOAD_MASS__KG_) AS "TOTAL_PAYLOAD_MASS"
FROM SPACEXTABLE
WHERE "Customer" = "NASA (CRS)"
GROUP BY "Customer"
```

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

[10]:

Customer	TOTAL_PAYLOAD_MASS
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

Out of the total payload mass carries by NASA, the average amount carries by Booster Version F9 v1.1 is 2,938.4 Kg.

Task 4

Display average payload mass carried by booster version F9 v1.1

[85]:

```
%%sql
SELECT "Booster_Version", AVG(PAYLOAD_MASS__KG_) AS "AVERAGE_PAYLOAD_MASS"
FROM SPACEXTABLE
WHERE "Booster_Version" = "F9 v1.1"
```

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

[85]:

Booster_Version	AVERAGE_PAYLOAD_MASS
F9 v1.1	2928.4

First Successful Ground Landing Date

The first successful ground landing was December 22nd, 2015

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

[103]:

```
%%sql
SELECT *, MIN("Date")
FROM SPACEXTABLE
WHERE "Landing_Outcome" = "Success (ground pad)"
```

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

[103]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome	MIN("Date")
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)	2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Here we can see that there were 4 Drone Ships that landed successfully that were between 4,000 kg and 6,000 kg

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[109]:

```
%%sql
SELECT "Booster_Version", "Landing_Outcome", "PAYLOAD_MASS__KG_"
FROM SPACEXTABLE
WHERE "Landing_Outcome" = "Success (drone ship)" AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000
```

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

[109]:

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

There were a total of 101 mission outcomes, only one of which was a failure in flight.

Task 7

List the total number of successful and failure mission outcomes

[115]:
```
%%sql
SELECT "Mission_Outcome", COUNT(*) AS OCCURENCES
FROM SPACEXTABLE
GROUP BY "Mission_Outcome"
```

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

[115]:

Mission_Outcome	OCCURENCES
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- These are the boosters that carried the maximum payload mass.
- This shows us which boosters have a record of heavy payload support.

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

[117]:
```
%%sql
SELECT "Booster_Version", "PAYLOAD_MASS__KG_" AS "MAX_PAYLOAD_MASS"
FROM SPACEXTABLE
WHERE "MAX_PAYLOAD_MASS" = (SELECT MAX("PAYLOAD_MASS__KG_")
                            FROM SPACEXTABLE)
```

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

[117]:

Booster_Version	MAX_PAYLOAD_MASS
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

2015 Launch Records

- In 2015, there were two failures.
- Both were two different Booster version.
- Both failures occurred in the first trimester.

Task 9

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

[130]:
```
%%sql
SELECT substr("Date",6,2) AS "Month", substr("Date",1,4) AS "Year", "Landing_Outcome", "Booster_Version", "Launch_Site"
FROM SPACEXTABLE
WHERE "Landing_Outcome" = "Failure (drone ship)" AND "Year" = "2015"
```

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

[130]:

Month	Year	Landing_Outcome	Booster_Version	Launch_Site
01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

The "Outcome Rank" here is the number of times an outcome occurred between June 4th, 2010 and March 20th, 2017

Task 10

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.

[139]:

```
%%sql
SELECT "Landing_Outcome", COUNT("Landing_Outcome") AS "Outcome Rank"
FROM SPACEXTABLE
WHERE "Date" > "2010-06-04" AND "Date" < "2017-03-20"
GROUP BY "Landing_Outcome"
ORDER BY "Outcome Rank" DESC
```

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

[139]:

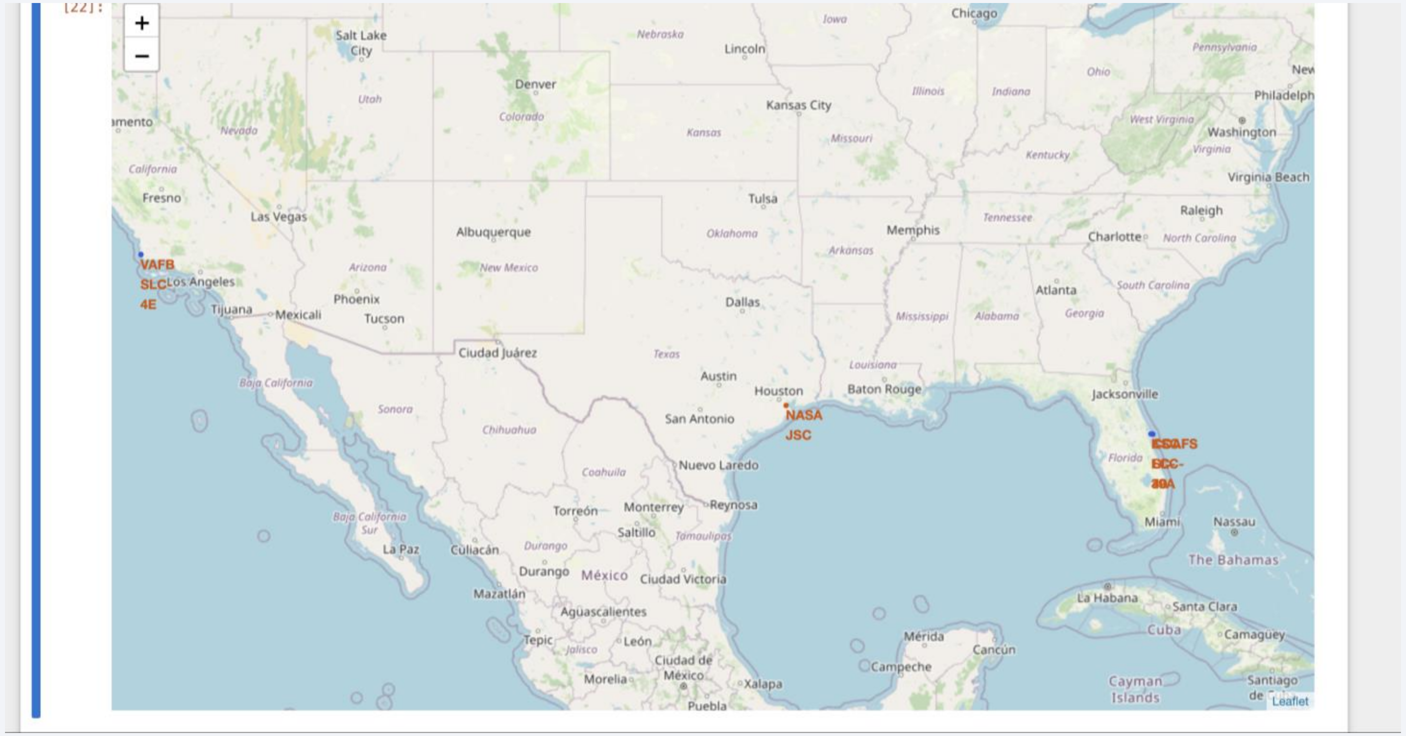
Landing_Outcome	Outcome Rank
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

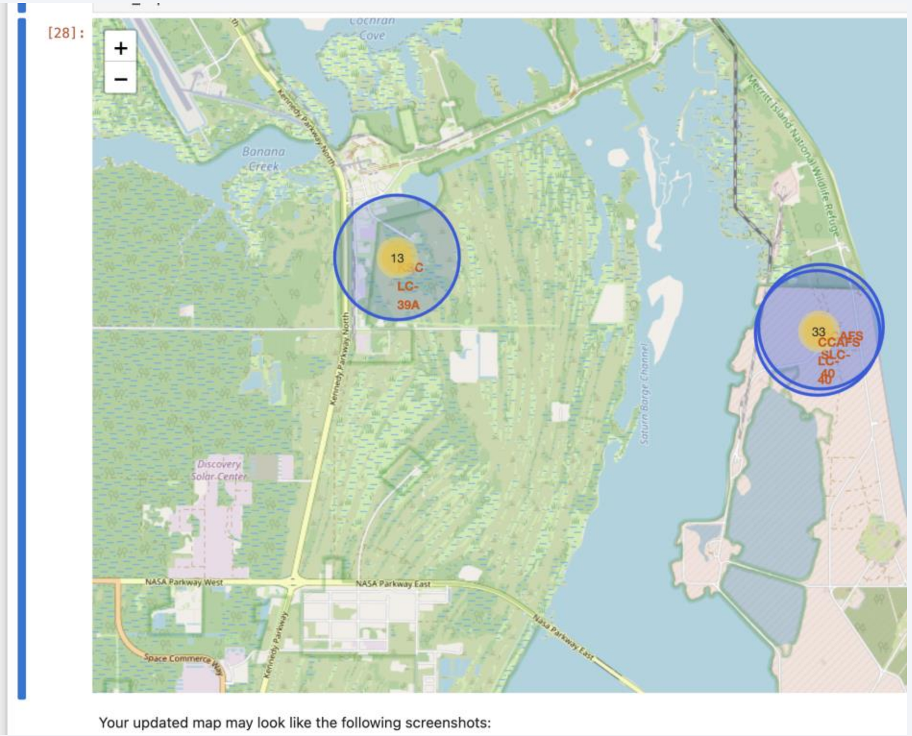
Launch Sites – Folium Map

- Here I added *red labels* as marker objects within Folium.
- Each launch site has been labeled at their respective coordinates.



Labeled Launch Outcomes – Folium Map

- The total outcomes will portray in the yellow circles
- When zooming in, we can see the separation between successful landings (green) and failures (red).
- Simple visual guide to easily discern number of successful launches.

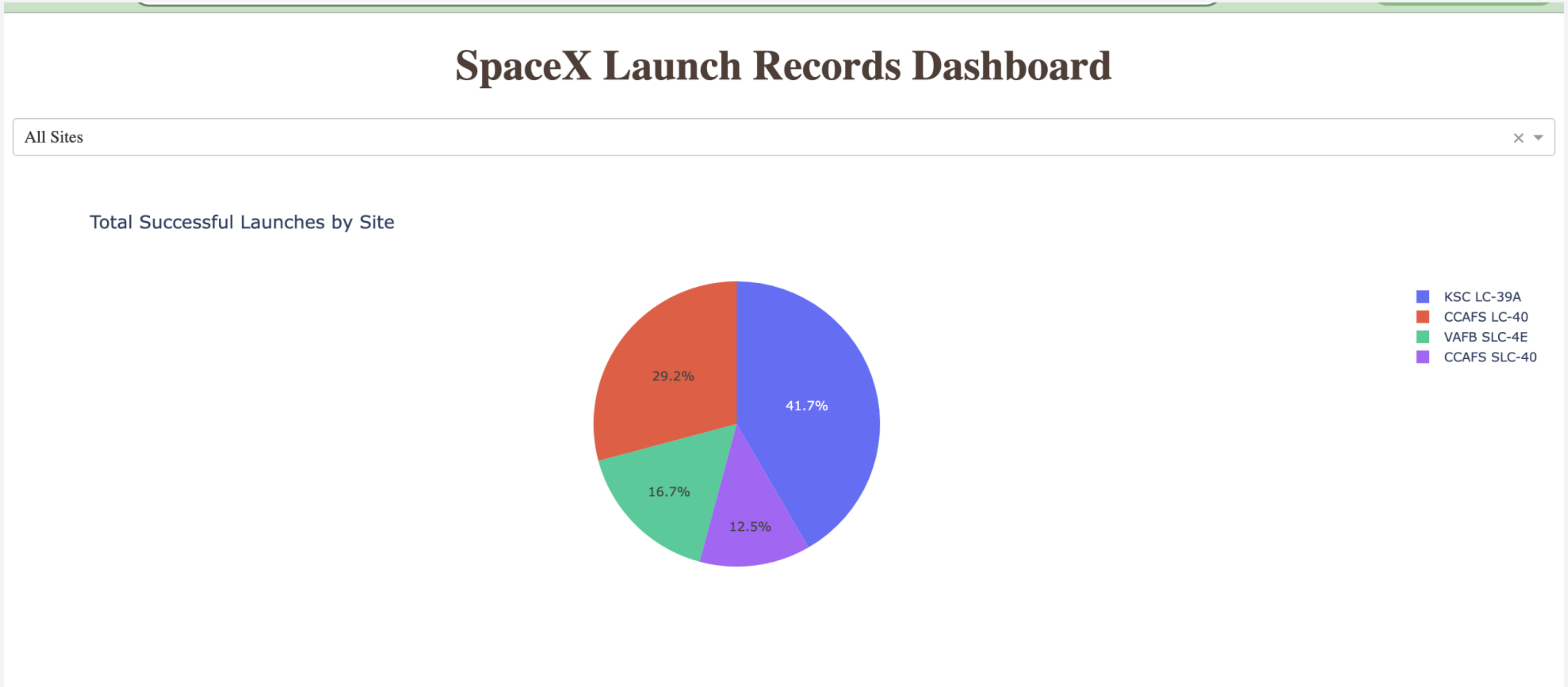


Section 4

Build a Dashboard with Plotly Dash

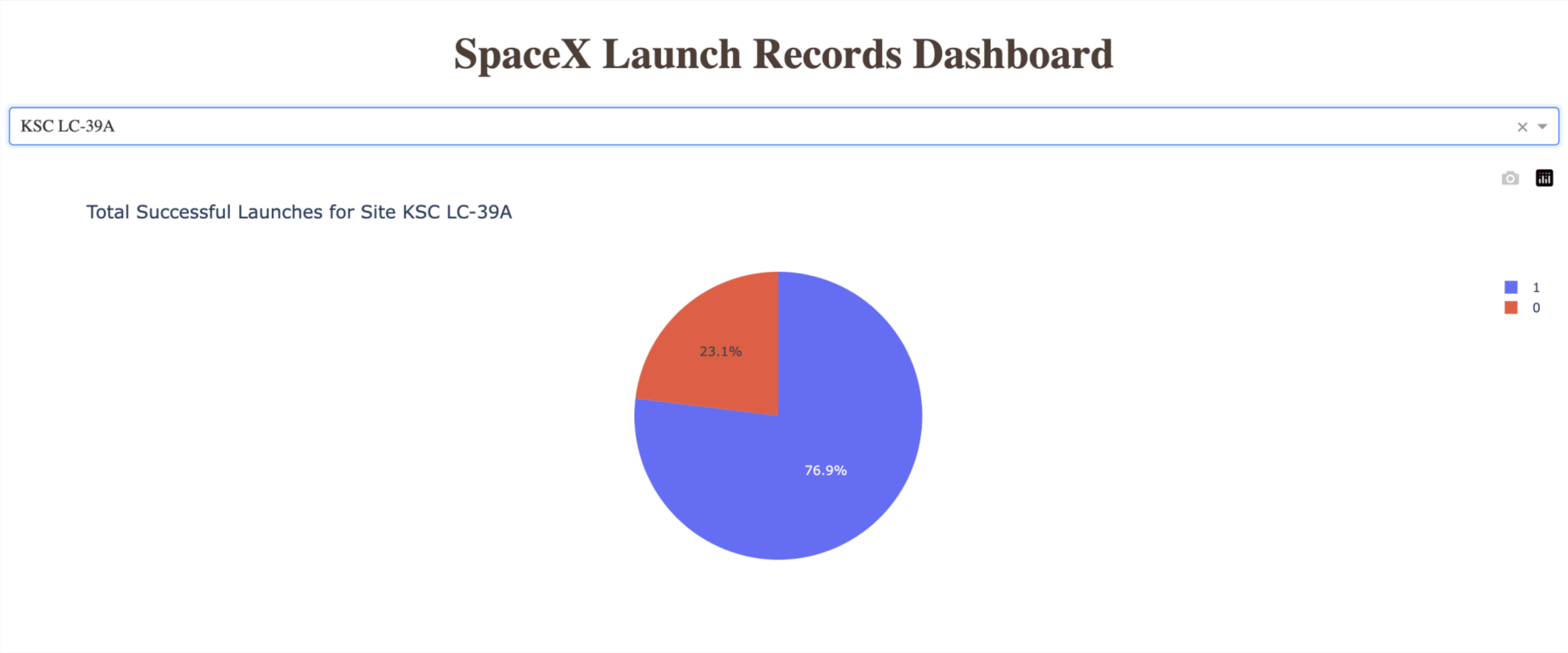
Launch Success Count – Plotly Dash

- Below is a pie chart that is partitioned based on the success count of each launch site from the total
- We can visually see that Launch Site with the greatest amount of success is KSC LC-39A



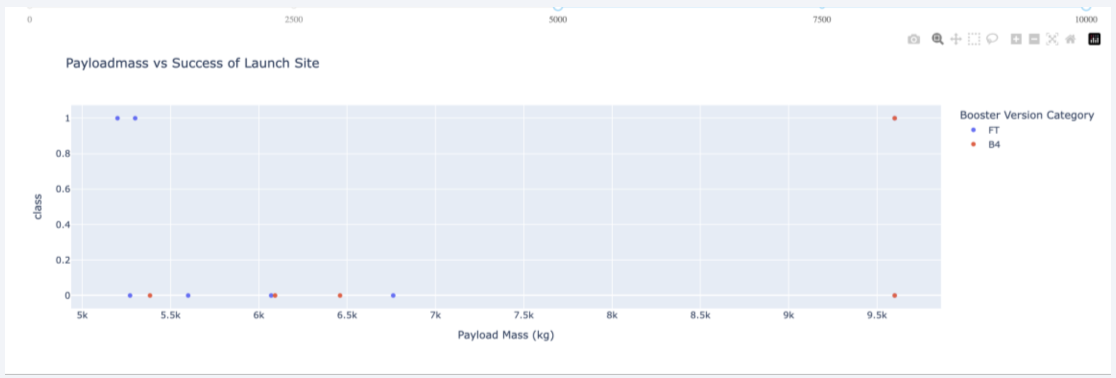
Highest Launch Success – Ploty Dash

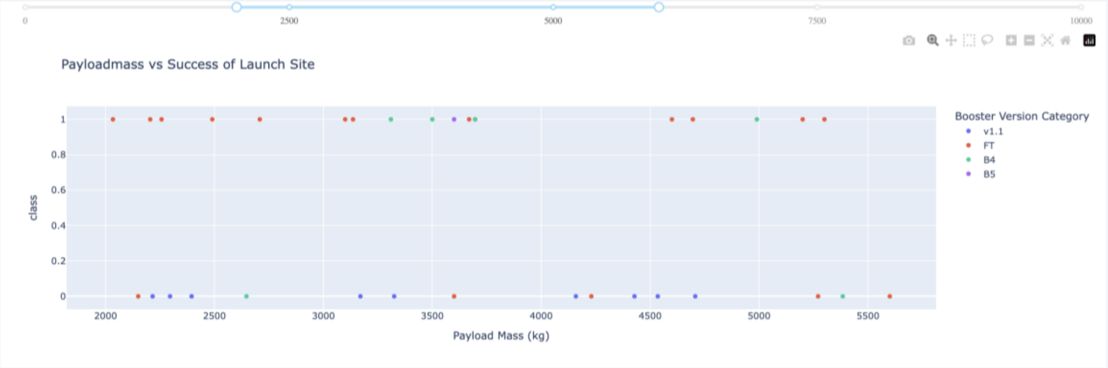
- We see here that launch site KSC LC-39A has the highest success ratio with 76.9%



Payload vs Launch Outcome – Ploty Dash

- The payload range with greater success is between (1952kg to 3696kg) and (4600kg - 5300kg)
- The booster version with the highest success rate also seems to be B5 but FT has a larger number of launches



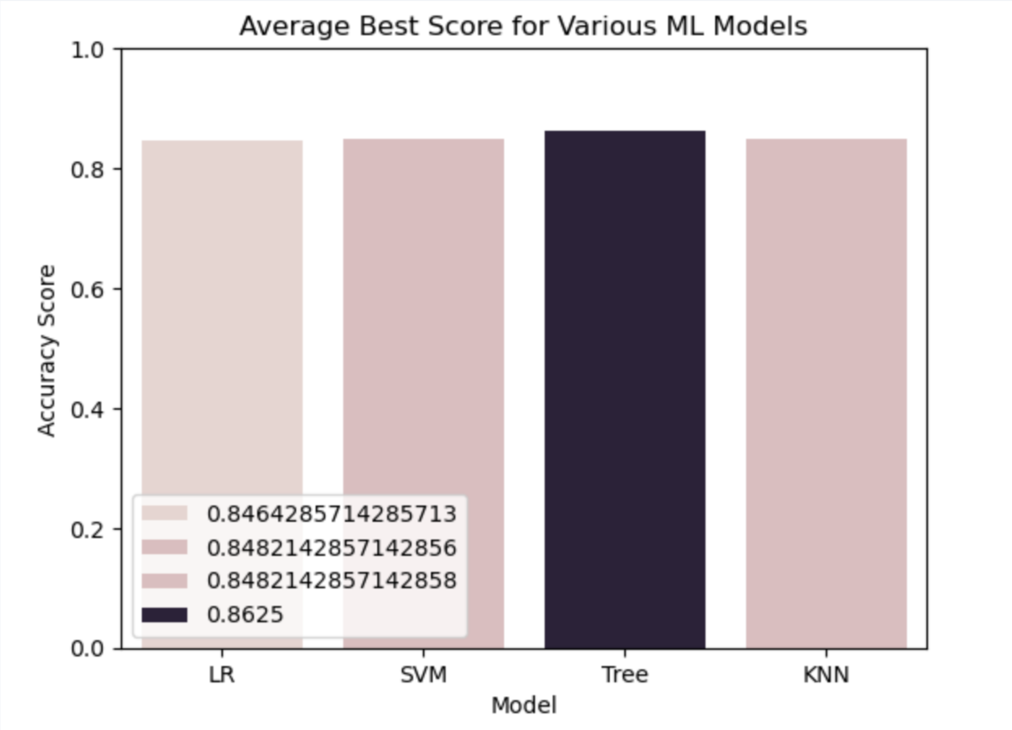


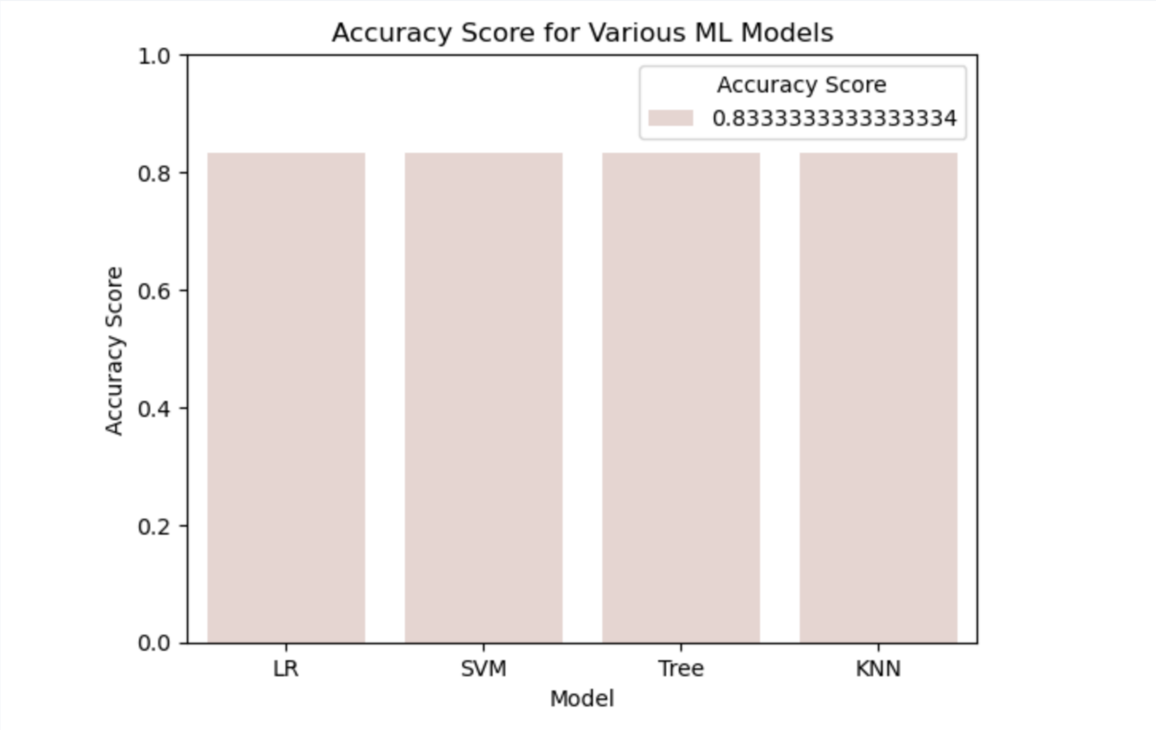
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- When looking at test sample accuracy, all models performed the same with 83% accuracy.
- When looking at average best accuracy score when training, the Tree model seemed to have performed the best.

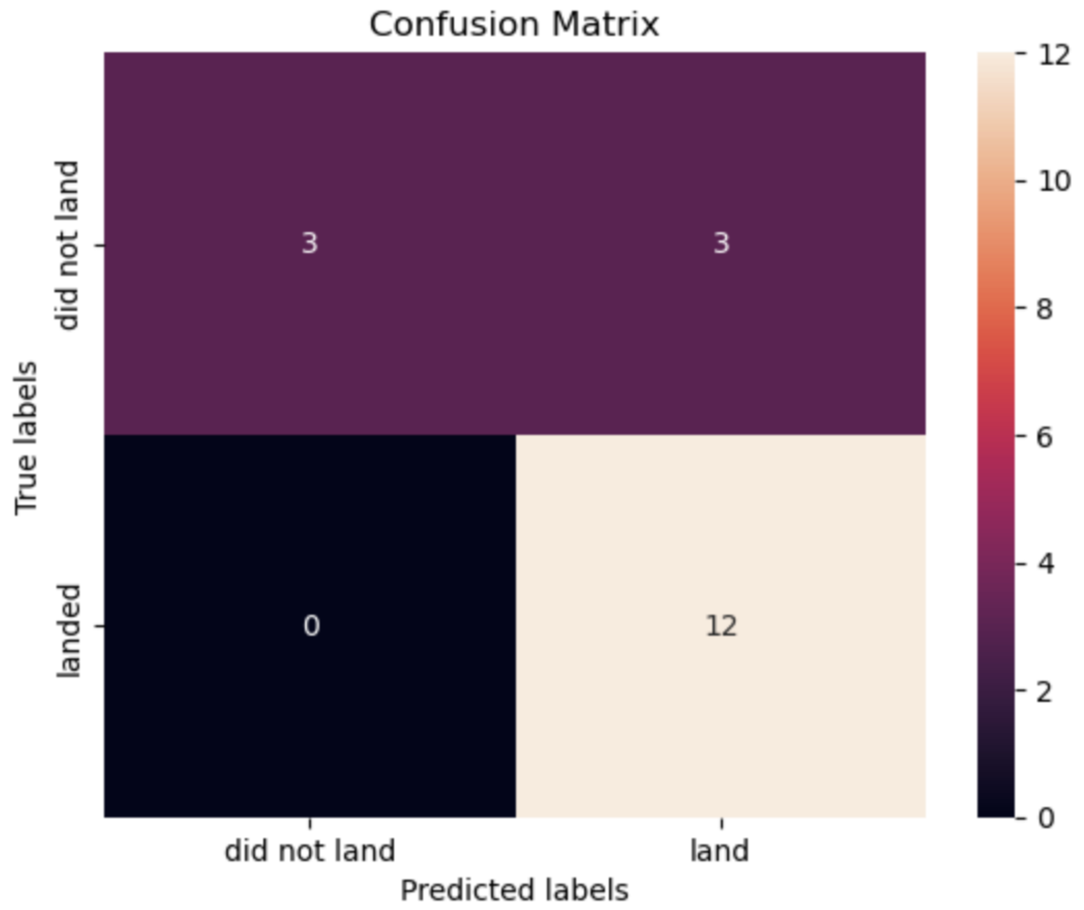




Confusion Matrix

- All confusion matrices had the same output, here is the one generated by the Decision Tree Model.
- We see that we have 12 true positives, but we also have 3 false positives where our model predicted "land" but the true label was "did no land"

[51]:
```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Conclusions

Orbit type is not a major factor for success.

There are payload ranges with higher levels of success

Specific launch sites have higher levels of success

Machine Learning algorithms can aid is predicting success, but there is about a 20% chance for error.

Appendix

Link to GitHub Repository containing all code and data used for this presentation:

LINK TO CAPSTONE PROJECT

Thank you!