IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jiyoon Shin
24.02.25.

Outline

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

Executive Summary

- This study delves into the prediction of Falcon 9 rocket first-stage landing outcomes, leveraging machine learning techniques. The analysis encompasses methodologies spanning data collection, preprocessing, model training, hyperparameter tuning, and evaluation.
- the findings underscore the potential of machine learning in forecasting Falcon 9 first-stage landings, offering valuable implications for decision-making in the aerospace industry.

Introduction

- Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.
- Essentially, this means that if we can determine whether the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against Space X for a rocket launch.
- So, in this lab, I created a machine learning pipeline to predict if the first stage will land given the data from the preceding labs.

Section 1

Methodology

Methodology

Executive Summary

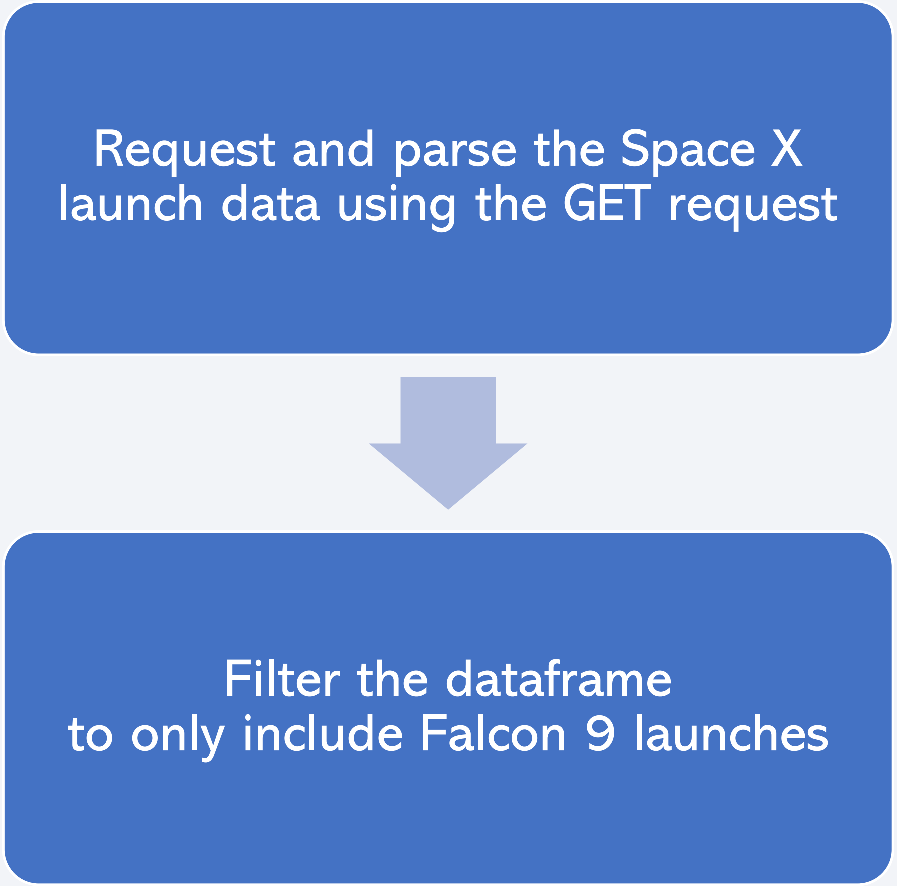
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Gather data from an Space X API
- Perform web scrapping to collect historical launch records of Falcon 9 from Wikipedia page titled ‘List of Falcon 9 and Falcon Heavy launches'

Data Collection – SpaceX API

- GitHub URL of the completed SpaceX API calls notebook:
  - https://github.com/jineoni/Coursera-ds-capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- GitHub URL of the completed web scraping notebook:
  - https://github.com/jineoni/Coursera-ds-capstone/blob/main/jupyter-labs-webscraping.ipynb

Request the Falcon9 Launch Wiki page from its URL

↓

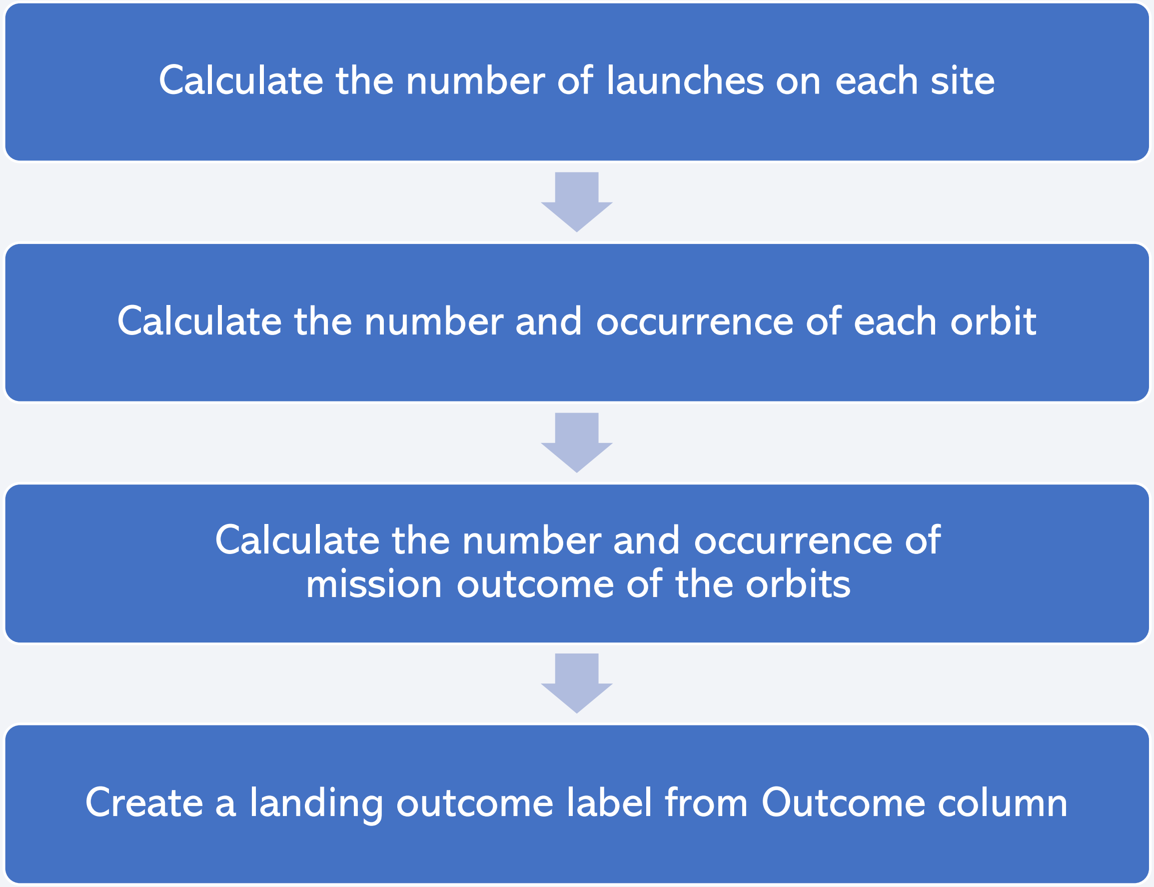
Extract all column/variable names from the HTML table header

↓

Create a data frame by parsing the launch HTML tables

Data Wrangling

- Convert various scenarios where the booster failed to land successfully into training labels: 1 indicates successful landing, while 0 indicates failure
- GitHub URL of completed data wrangling related notebooks:
  - https://github.com/jineoni/Coursera-ds-capstone/blob/main/jupyter-labs-spacex-data-wrangling.ipynb



EDA with Data Visualization

- Perform Exploratory Data Analysis and Feature Engineering
- GitHub URL of completed EDA with data visualization notebook:
  - https://github.com/jineoni/Coursera-ds-capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA

- Visualize the relationship between Flight Number and Launch Site
- Visualize the relationship between Payload and Launch Site
- Visualize the relationship between success rate of each orbit type
- Visualize the relationship between FlightNumber and Orbit type
- Visualize the relationship between Payload and Orbit type
- Visualize the launch success yearly trend

Feature Engineering

- Create dummy variables to categorical columns
- Cast all numeric columns to float64

EDA with SQL

- Some SQL queries I performed:
  - SELECT DISTINCT Launch_Site FROM SPACEXTBL
  - SELECT Landing_Outcome, COUNT(*) AS “total number“ FROM SPACEXTBL GROUP BY Landing_Outcome
- GitHub URL of completed EDA with SQL notebook:
  - https://github.com/jineoni/Coursera-ds-capstone/blob/main/jupyter-labs-eda-sqllite.ipynb

Build an Interactive Map with Folium

- Mark the success/failed launches for each site on the map
  - to see which sites have high success rates
- Calculate the distances between a launch site to its proximities
  - identified the regional characteristics of launch sites with high success rates by calculating the distances to the coastline or to urban areas
- GitHub URL of completed interactive map with Folium map:
  - https://github.com/jineoni/Coursera-ds-capstone/blob/main/jupyter-labs-launch-site-location.ipynb

Build a Dashboard with Plotly Dash

- Pie chart to illustrate the proportion of landing success and failure for the selected launch site chosen by the user.
- Scatter chart to visualize the relationship between payload mass and success rate.
- GitHub URL of your completed Plotly Dash lab:
  - https://github.com/jineoni/Coursera-ds-capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Conduct hyperparameter tuning for SVM, Decision Tree, Logistic Regression, KNN.
- Determine the best performing method using test data.
- GitHub URL of completed predictive analysis lab:
  - https://github.com/jineoni/Coursera-ds-capstone/blob/main/jupyter-labs-spacex-machine-learning-prediction.ipynb

Pre-processing

- Standardize the data.
- Use the function train_test_split to split the data X and Y into training and test data.

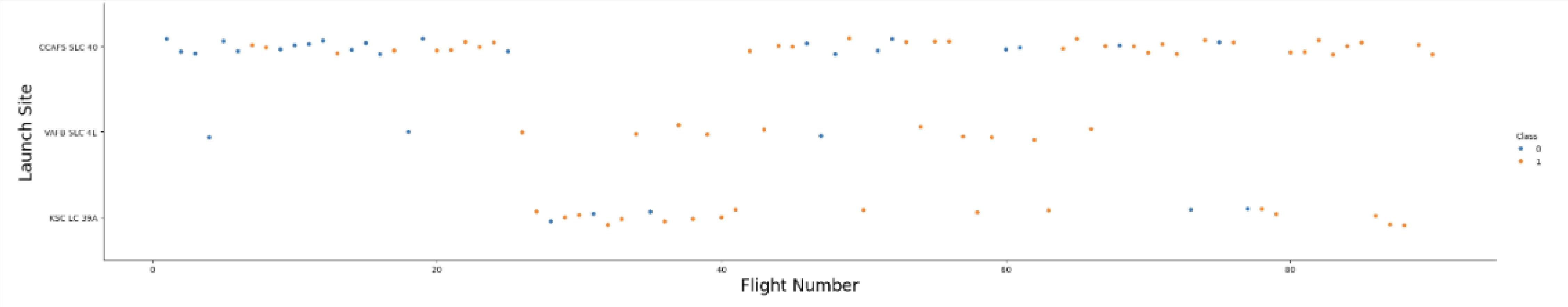
For each of the four models

- Create a model object then create a GridSearchCV object with cv = 10. Fit the object to find the best parameters from the dictionary .
- Calculate the accuracy on the test data using the method.

Section 2

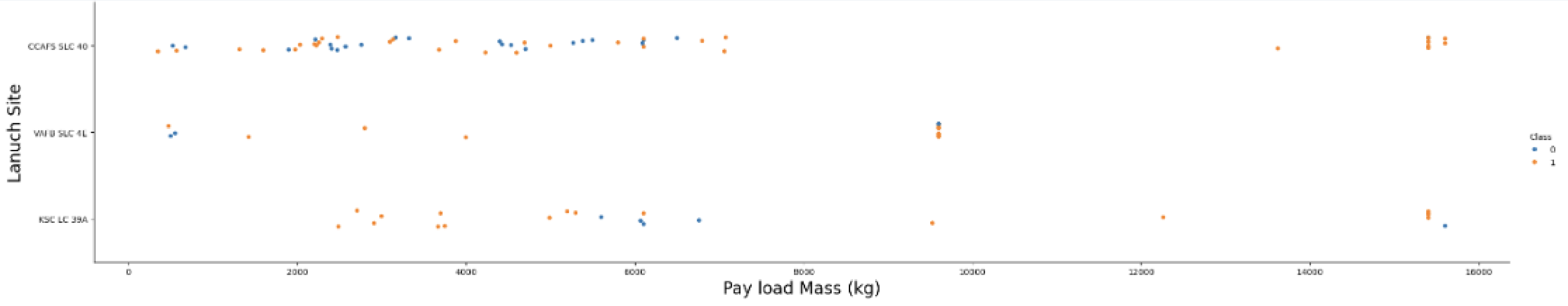
Insights drawn from EDA

Flight Number vs. Launch Site



- All three launch sites, especially CCAFS SLC 40, show that as the flight number increases, the success rate of landing also increases. Considering that the flight number is indexed when sorting the data for each launch site based on dates, it can be interpreted that the success rate of landing tends to increase over time.

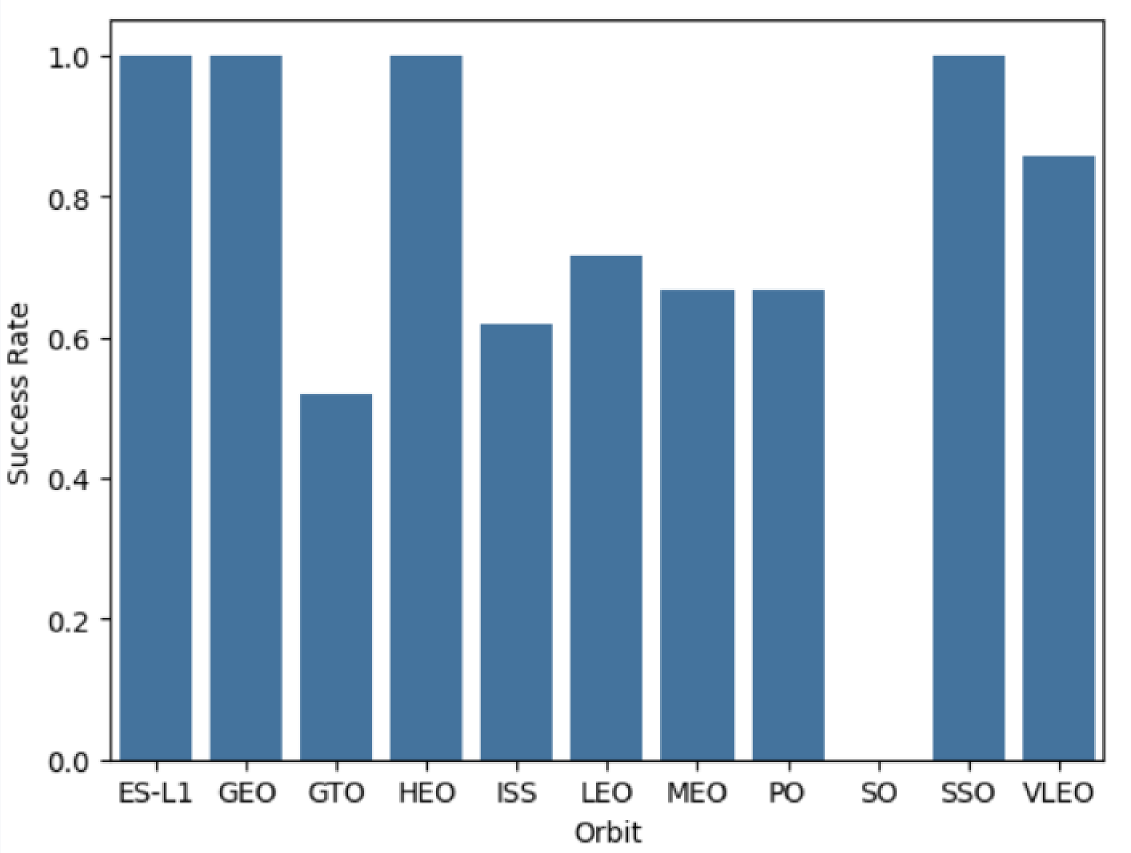
Payload vs. Launch Site



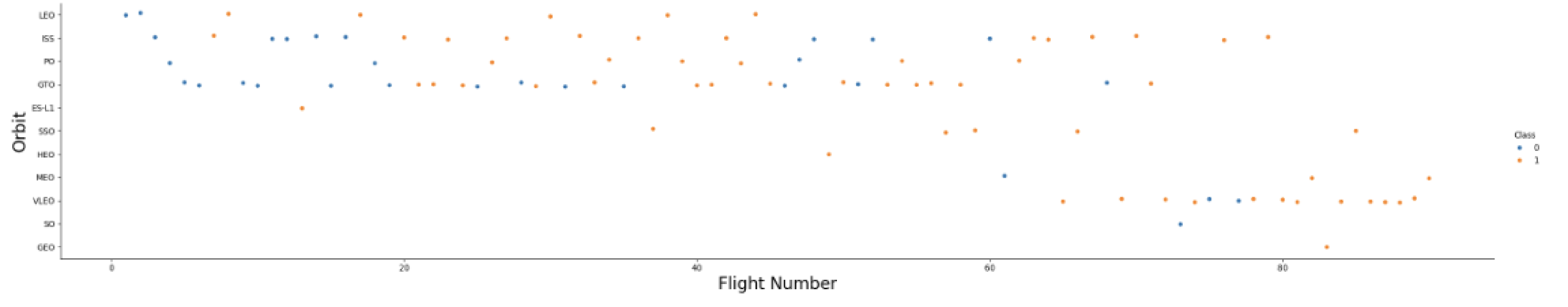
- The trend generally suggests that as the payload mass increases, the probability of landing success also increases. However, considering the rarity of attempts with a payload mass exceeding 10000, such an interpretation lacks strong credibility. Moreover, at the VAFB SLC launch site, there have been no attempts with a payload mass exceeding 10000.

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO orbits have recorded a success rate of 1. On the other hand, SO has recorded a success rate of 0. If one intends to attempt a launch, focusing on orbits with ES-L1, GEO, HEO, and SSO could increase the likelihood of success.

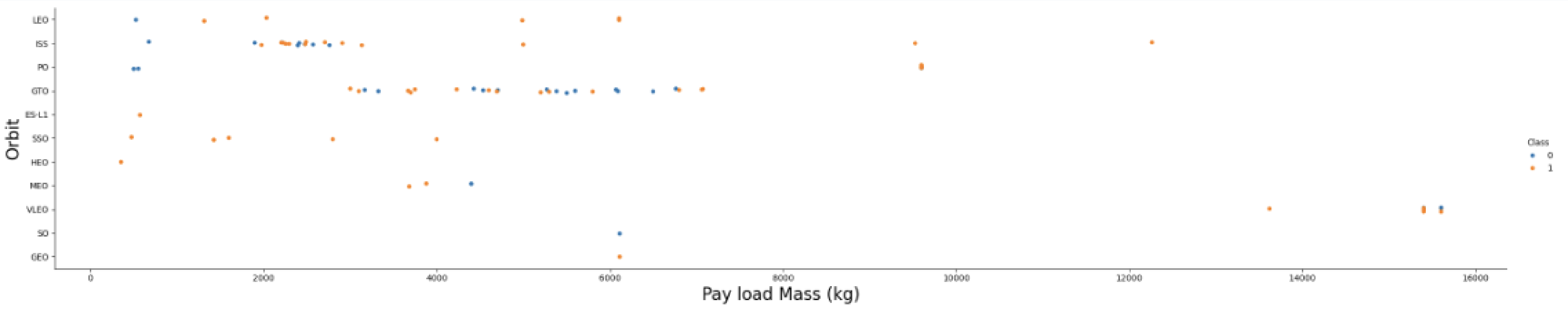


Flight Number vs. Orbit Type



- In the LEO orbit, the number of successes tends to increase as the flight number grows. In contrast, the GTO orbit experiences repeated successes and failures regardless of the flight number. This confirms that the correlation between flight number and success can vary depending on the type of orbit.

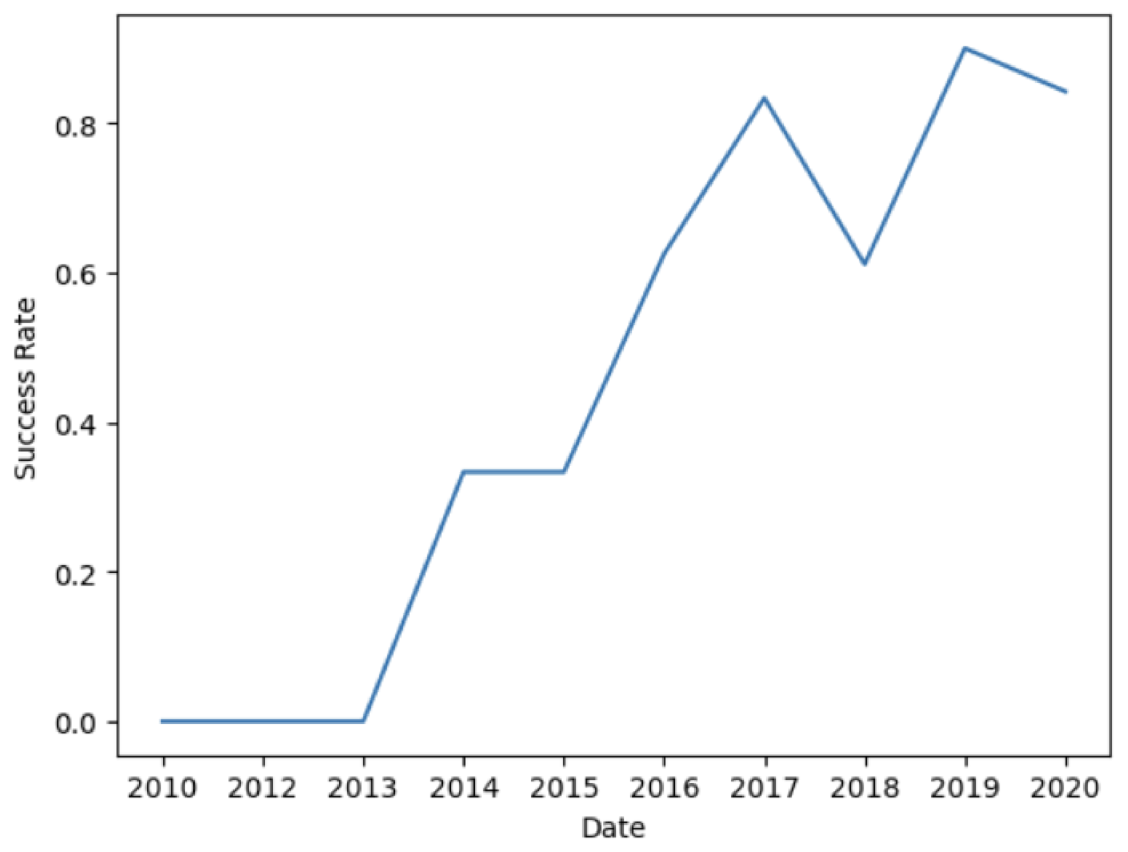
Payload vs. Orbit Type



- PO, LEO, and ISS orbits show an increasing success rate as the payload mass increases.
- For ES-L1, SSO, and HEO orbits, all attempts with lighter payloads have been successful, but there have been no attempts with heavier payloads. It is reasonable to speculate that these orbits may only be suitable for launches with lighter payloads, given the absence of attempts with heavier masses.

Launch Success Yearly Trend

- The graph on the right represents the average launch success rate by year. Overall, there is an increasing trend, but it’s noticeable that the performance was not as good in 2018 compared to the previous year.



All Launch Site Names

In [10]:
```
%%sql
SELECT DISTINCT Launch_Site
    FROM SPACEXTBL
```

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

Out[10]:

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

Launch Site Names Begin with 'CCA'

Out[13]:

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 attemp
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attemp
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Total Payload Mass

```
In [17]:  %%sql
          SELECT SUM(PAYLOAD_MASS__KG_) AS "total payload mass"
              FROM SPACEXTBL
              WHERE Customer = 'NASA (CRS)'

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

Out[17]:

total payload mass
45596

Average Payload Mass by F9 v1.1

```
In [19]: %%sql
         SELECT AVG(PAYLOAD_MASS__KG_) AS "average payload mass"
             FROM SPACEXTBL
             WHERE Booster_Version LIKE 'F9 v1.1%'

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

Out[19]:

average payload mass
2534.6666666666665

First Successful Ground Landing Date

In [22]:
```
%sql SELECT DISTINCT Landing_Outcome FROM SPACEXTBL
```

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

Out[22]:

Landing_Outcome
Failure (parachute)
No attempt
Uncontrolled (ocean)
Controlled (ocean)
Failure (drone ship)
Precluded (drone ship)
Success (ground pad)
Success (drone ship)
Success
Failure
No attempt

In [24]:
```
%%sql
SELECT MIN(Date) AS "first successful landing outcome in ground pad"
    FROM SPACEXTBL
    WHERE Landing_Outcome = "Success (ground pad)"
```

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

Out[24]:

first successful landing outcome in ground pad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

In [25]:

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

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

Out[25]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

In [28]:

```
%%sql
SELECT Landing_Outcome, COUNT(*) AS "total number"
    FROM SPACEXTBL
    GROUP BY Landing_Outcome
```

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

Out[28]:

Landing_Outcome	total number
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

```
In [29]: %%sql
SELECT Booster_Version
    FROM SPACEXTBL
    WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)

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

Out[29]:

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

2015 Launch Records

In [30]:

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

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

Out[30]:

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

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

In [31]:

```
%%sql
SELECT Landing_Outcome, COUNT(Landing_Outcome)
    FROM SPACEXTBL
    WHERE Date BETWEEN "2010-06-04" AND "2017-03-20"
    GROUP BY Landing_Outcome
    ORDER BY COUNT(Landing_Outcome) DESC
```

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

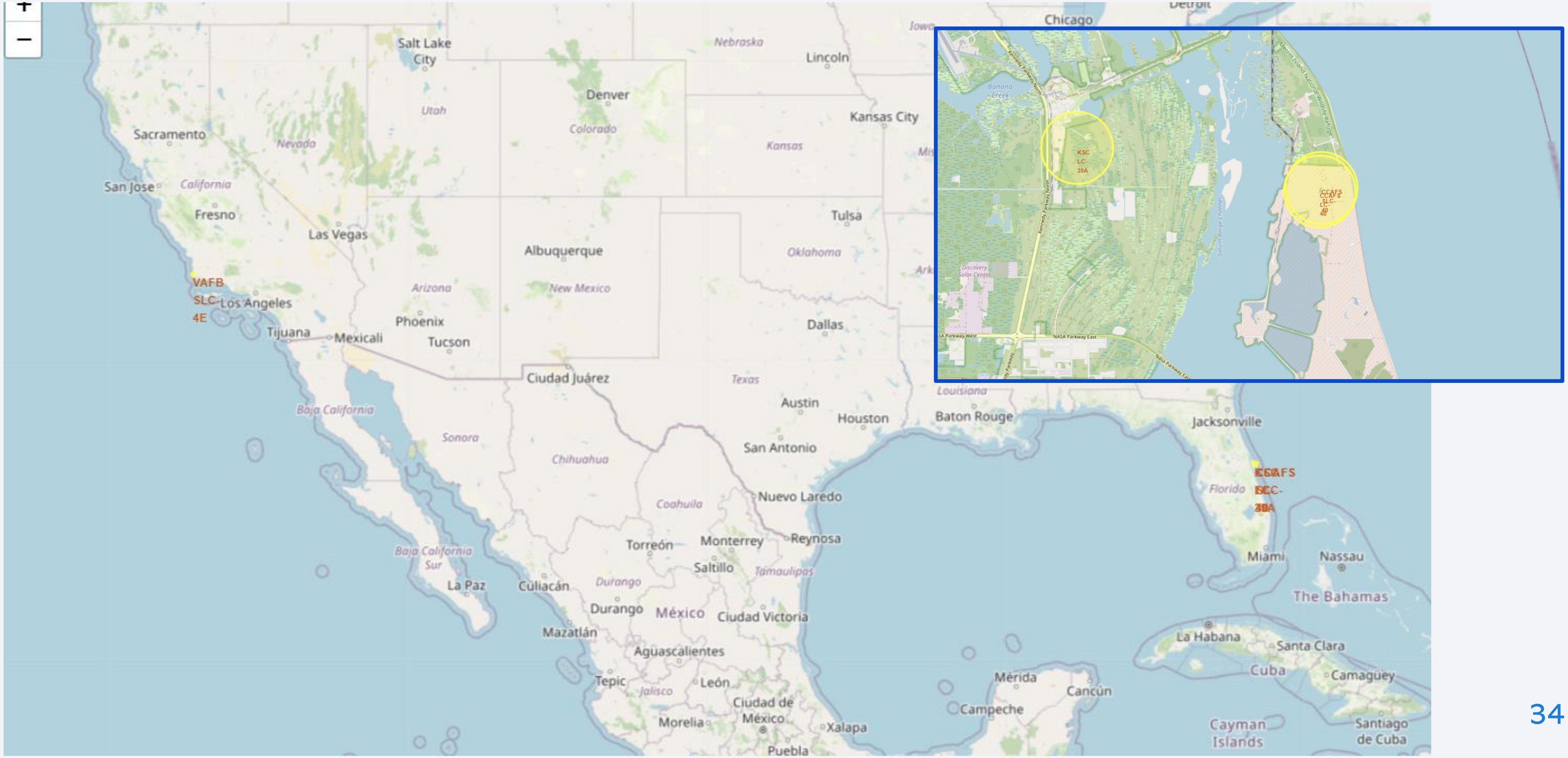
Out[31]:

Landing_Outcome	COUNT(Landing_Outcome)
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

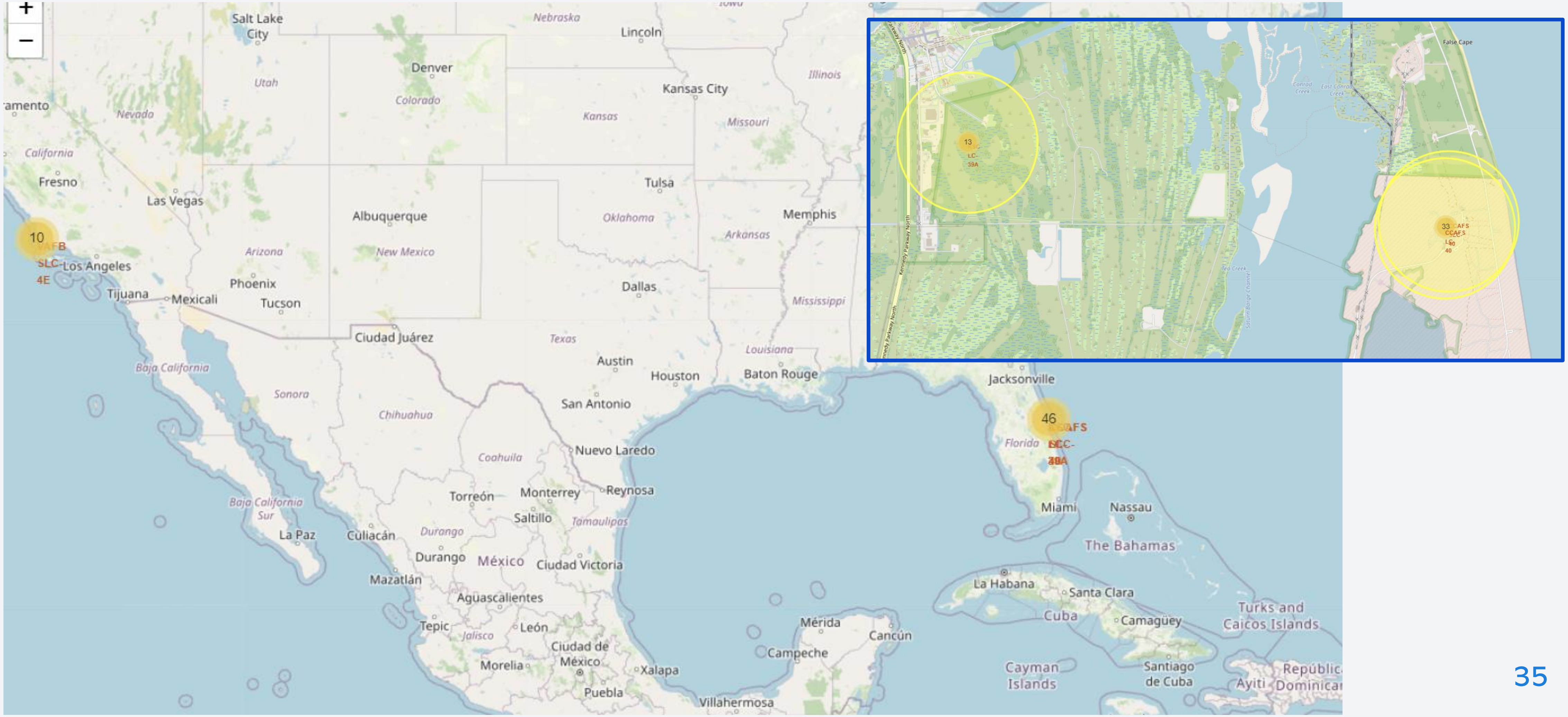
Section 3

Launch Sites Proximities Analysis

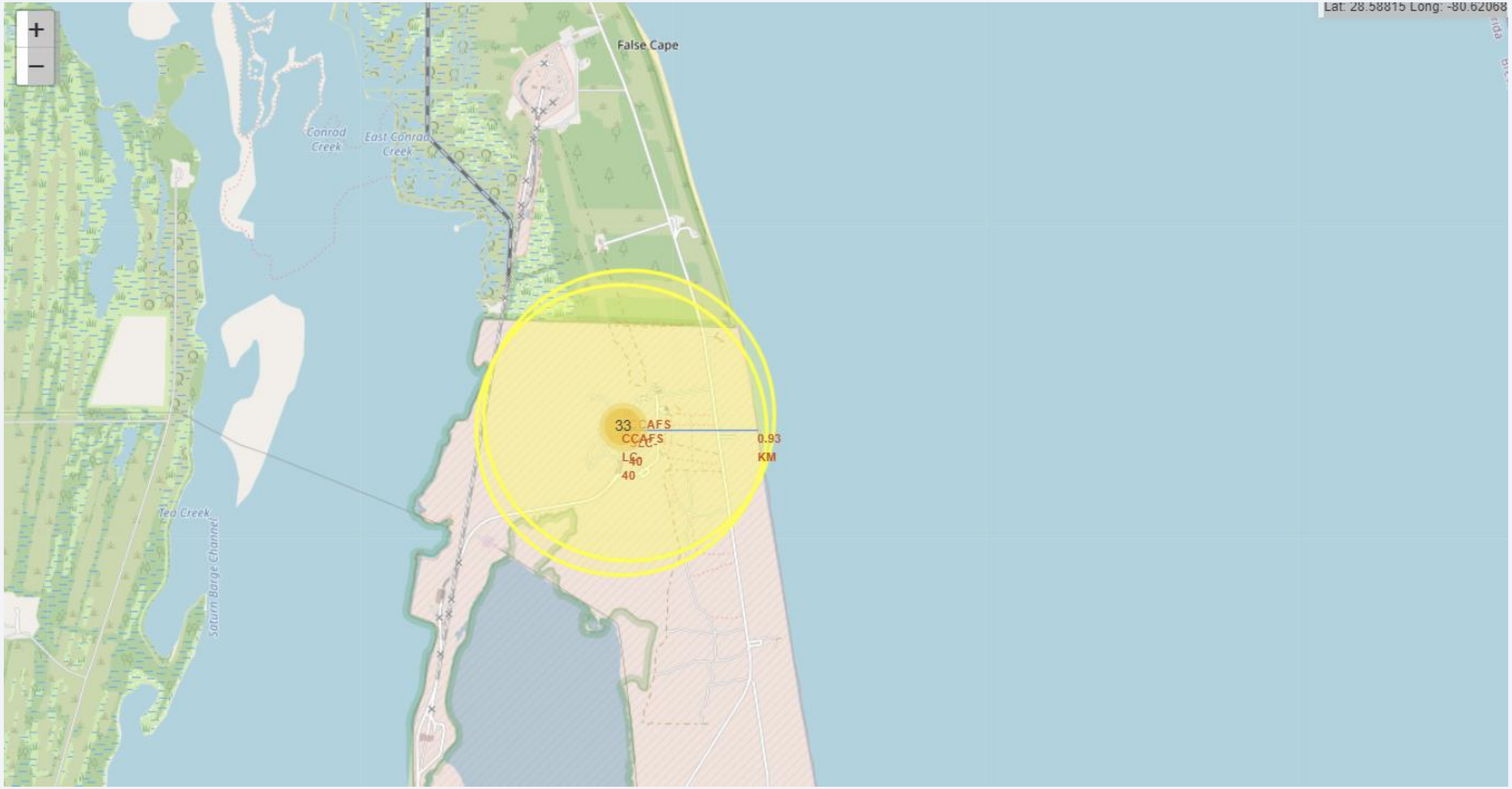
Launch Site Locations



Success/Failed Launch Site Locations



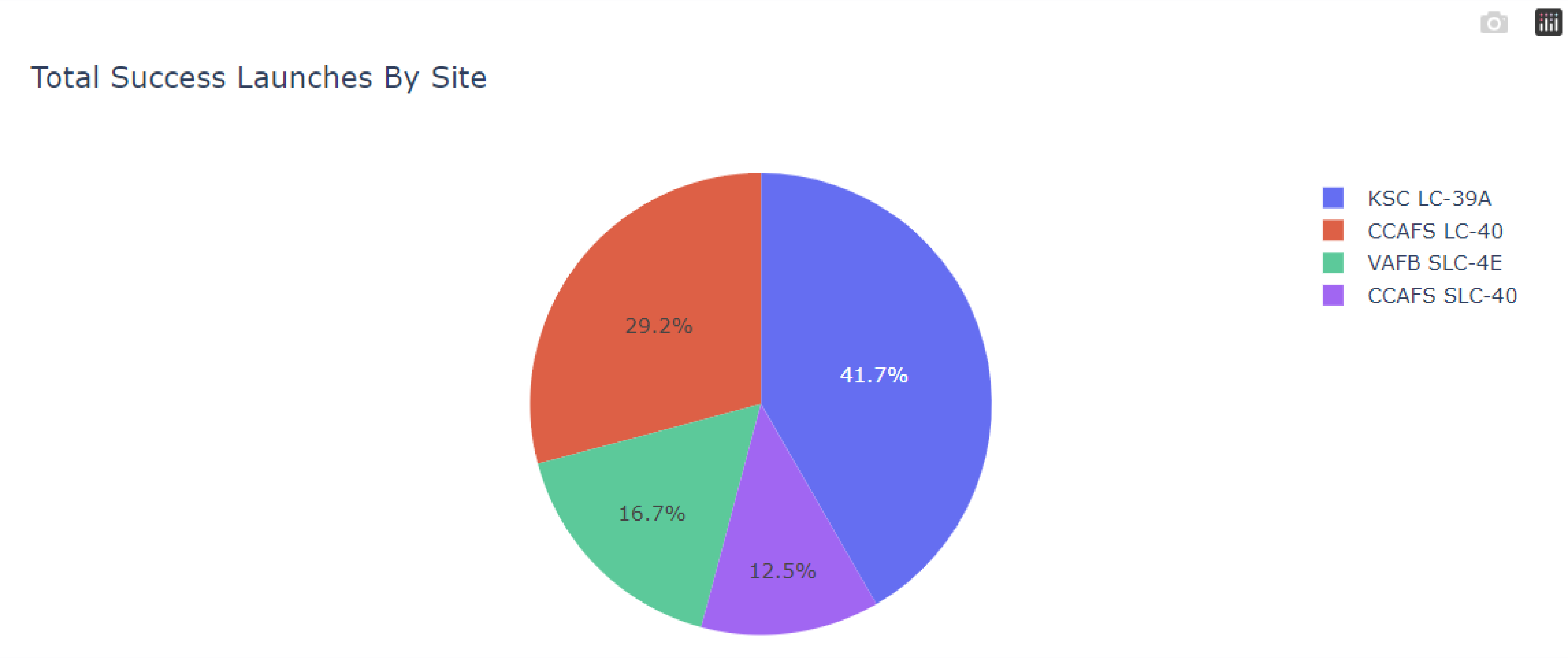
The distances between a launch site to its proximities



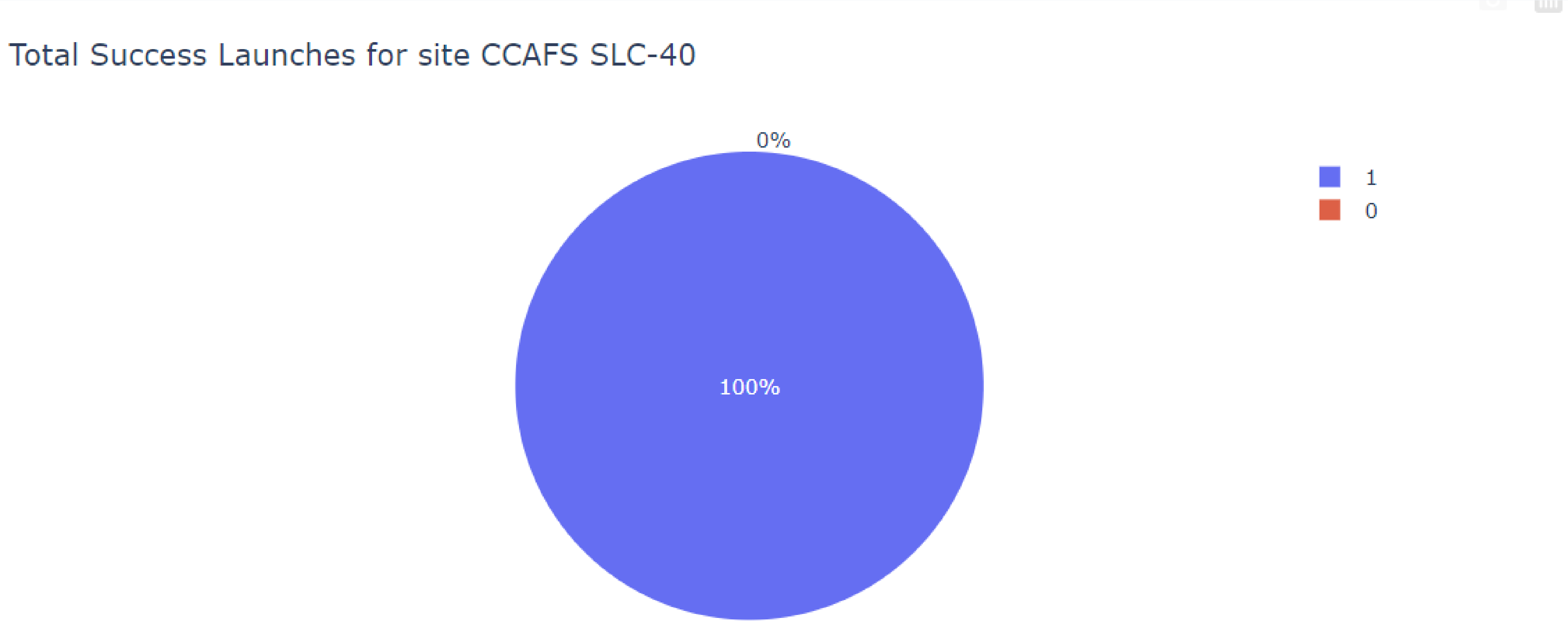
Section 4

Build a Dashboard with Plotly Dash

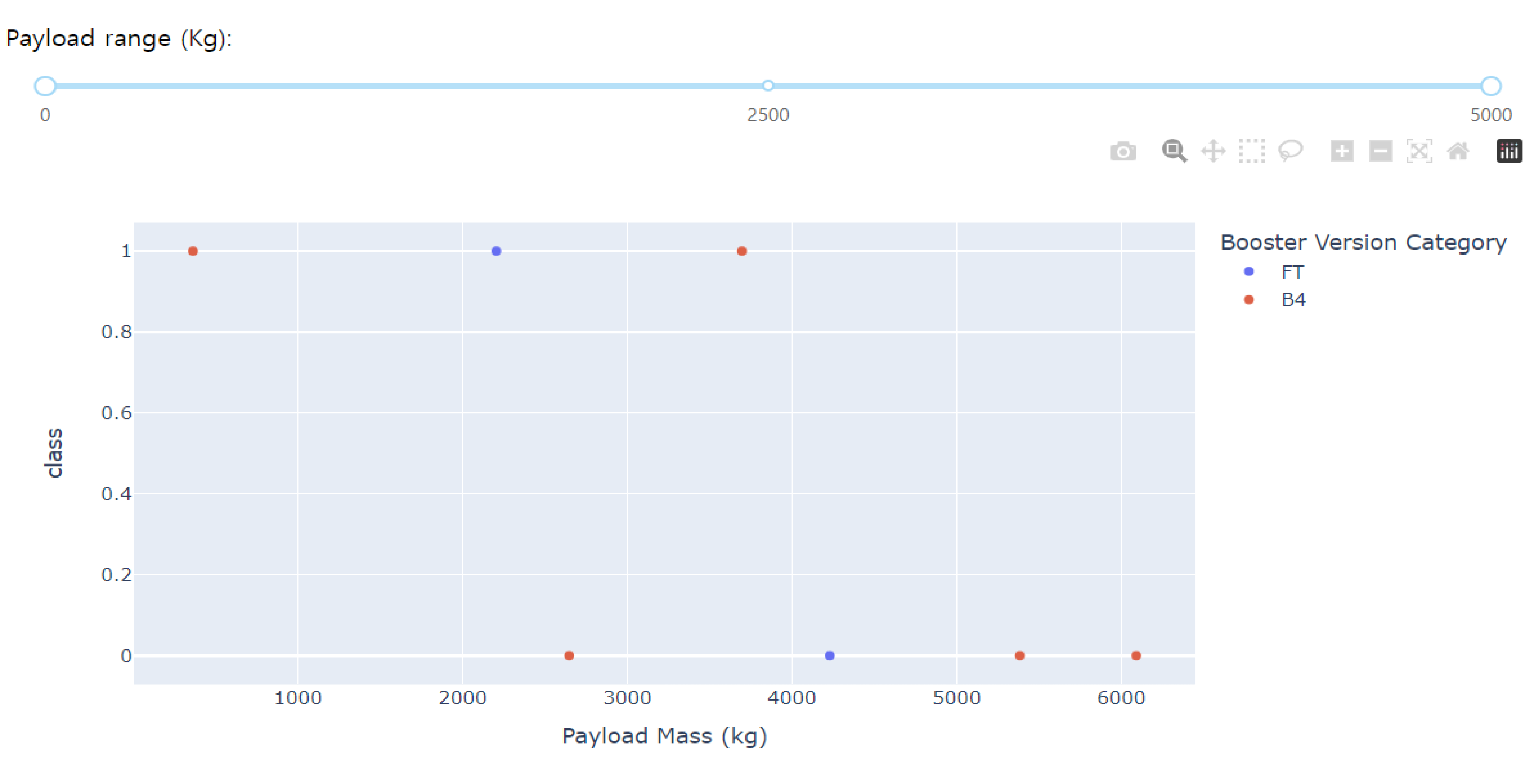
Launch Success Count for all cities



A Launch Site with Highest Launch Success Ratio



Payload vs. Launch Outcome for all cities

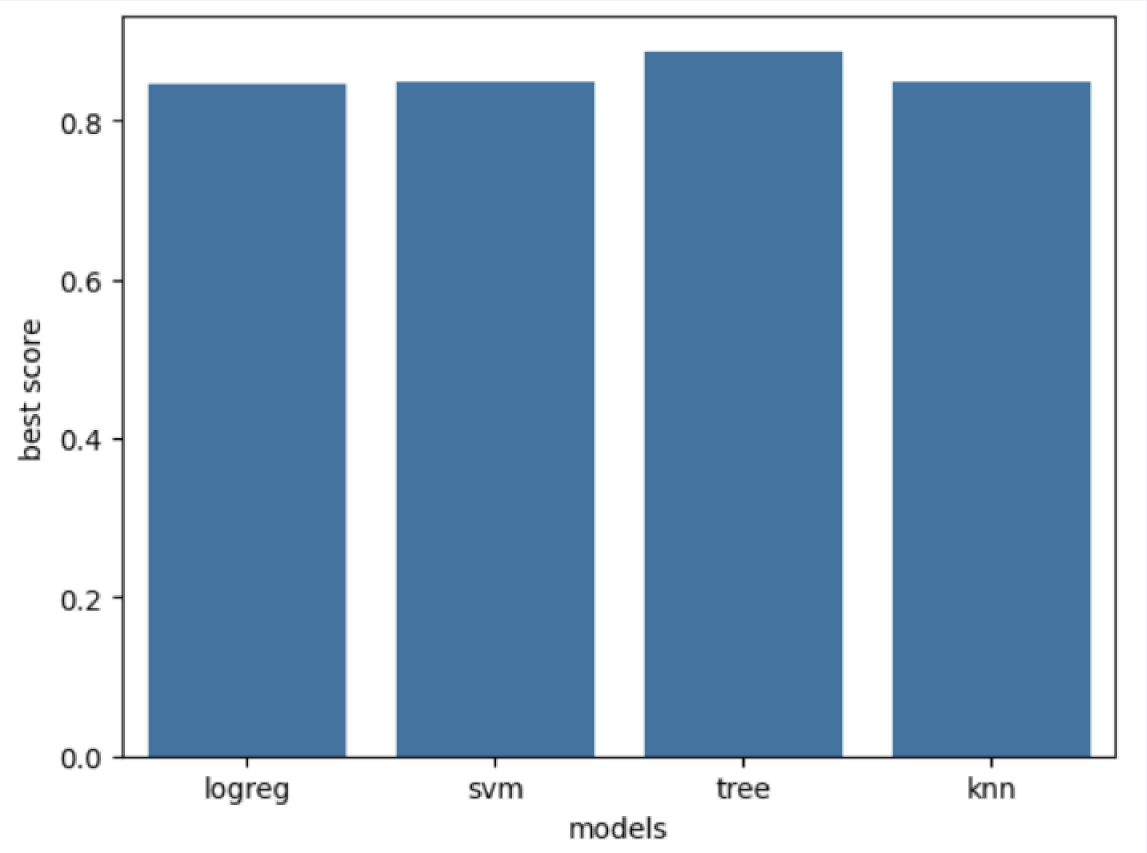


Section 5

Predictive Analysis (Classification)

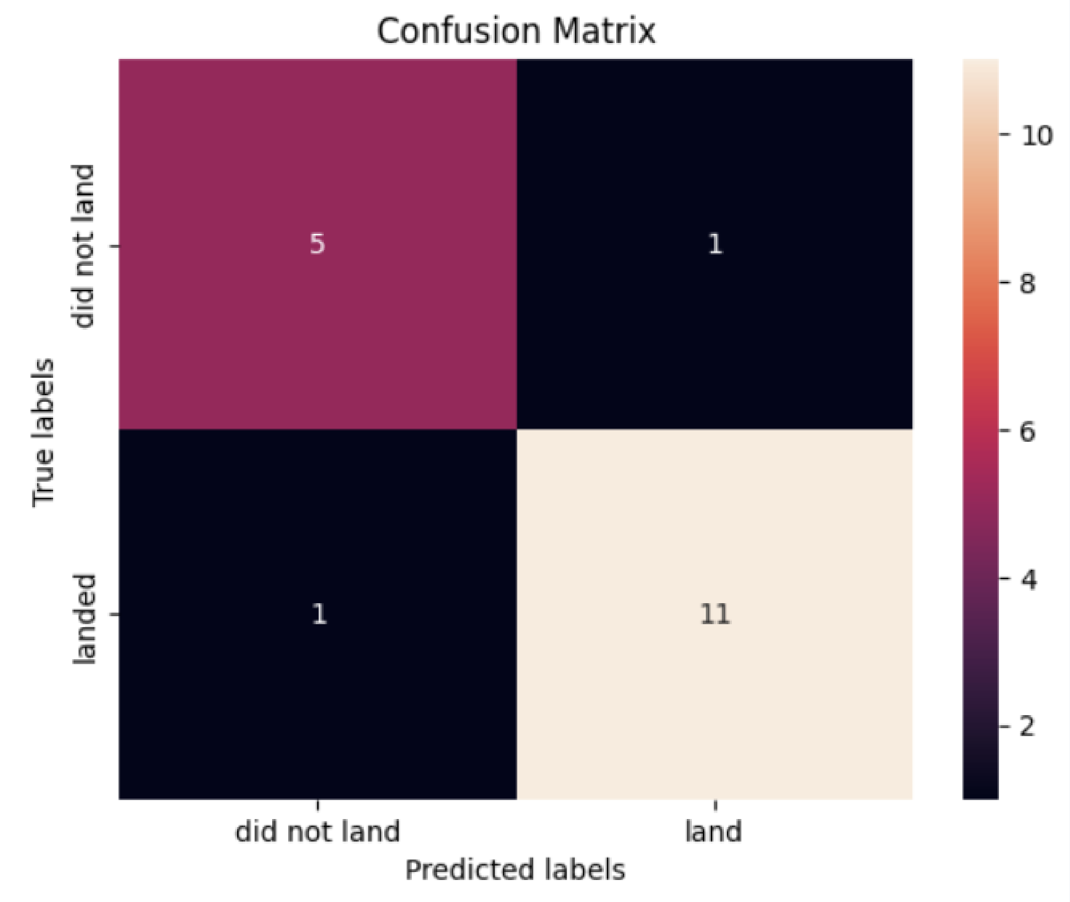
Classification Accuracy

- The scores on the test data were the same for all four models. However, the best score during the model training process was highest for the decision tree model. Therefore, it can be said that the most suitable classification model is the decision tree.



Confusion Matrix

- The tree model failed in prediction a total of two instances.
  - One instance was where ‘landed’ was predicted as ‘do not land’,
  - And the other was where ‘did not land’ was predicted as ‘land’.
- Nevertheless, it can still be considered as performing reasonably well in its predictions.



Thank you!