SPACEX

IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ikenna Uwaezuoke

September 5th, 2023

Outline

Executive Summary

1. Summary of methodologies
2. - Data Collecting & Data Wrangling
3. - Exploratory Data Analysis
4. - Interactive Visual Analytics
5. - Machine Learning Prediction
6. Summary of all results
7. A predictive analytics using machine learning to predict if the first stage of the Falcon 9 would land successfully

3

Introduction

- **Project background and context**

SpaceX 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 SpaceX can reuse the first stage. This stage does most of the work and is much larger than the second stage. However, sometimes it does not land. Other times, Space X will sacrifice the first stage due to the mission parameters like payload, orbit, and customer.

- **Problems you want to find answers**

To determine if the first stage will land, in order to be able to determine the cost of a launch.

Section 1

Methodology

Methodology

- Data collection methodology:
  - Data was collected from SpaceX API and Wiki page through the web scraping method.

- Perform data wrangling:
  - Perform EDA to find some patterns
  - Determine what would be the label for training supervised model

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Compare logistic regression model, support vector machine, decision tree classification, KNN by using GridSearchCV to select the best fit model.

Data Collection

1 API

spacex_url="https://api.spacexdata.com/v4/ launches/past"

- Collecting the data from SpaceX API
- Clean the data

2 Web page

https://en.wikipedia.org/wiki/List_of_Falcon_9 _and_Falcon_Heavy_launches

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

Data Collection-SpaceX API

- Request data from SpaceX API
- Convert the JSON result into a dataframe
- Filter dataframe to only Falcon 9' launches and data wrangling
- Export to csv

Link to GitHub for requesting data from SpaceX API

GitHub URL for completed SpaceX API request

Data Collection - Scraping

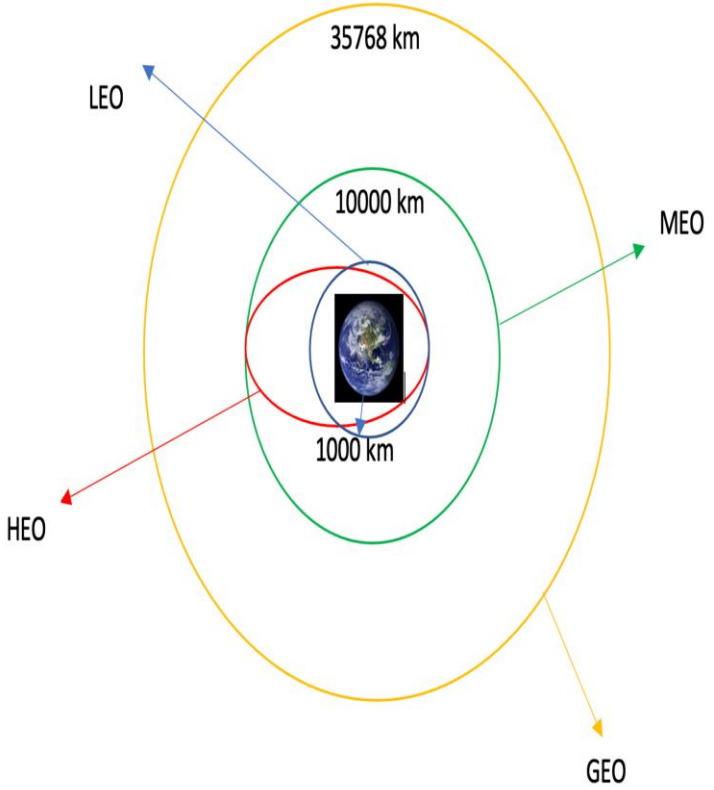
- Request the Falcon 9 Launch Wiki page from its URL
- Extract all column/variable names from the HTML table header
- Create a data frame by parsing the launch HTML tables
- Export to csv

Link:

- GitHub for Wiki web scraping

Data Wrangling

- Identify and calculate the percentage of the missing values in each attribute
- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome per orbit type
- Create a landing outcome label from Outcome column



Link: GitHub link for Data Wrangling

EDA with Data Visualization

- Visualize the relationship between Flight Number and Launch Site -> scatter plot
- Visualize the relationship between Payload and Launch Site -> scatter plot
- Visualize the relationship between success rate of each orbit type -> bar plot
- Visualize the relationship between Flight Number and Orbit type -> scatter plot
- Visualize the relationship between Payload and Orbit type -> scatter plot
- Visualize the launch success yearly trend -> line chart
- Create dummy variables to categorical columns -> Features Engineering
- Cast all numeric columns to float64 -> using astype function

- The scatter plot is the best to describe the relation between two categorical data
- The bar plot is the best to compare several categorical data
- The line plot is the best to show the time series data

Link: GitHub link for EDA DataViz

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- 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 date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Link: GitHub Link for EDA with SQL

Build an Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities
  - To determine if launch site is close to the railway
  - To determine if launch site is close to the highway
  - To determine if launch site is close to the city
  - To determine if launch site is close to the coastline

Link: GitHub link for building an interactive map with folium

Build a Dashboard with Plotly Dash

- Add a drop-down list to enable launch site input component
- Add a pie-chart to show success vs failed launch counts based on the selected site dropdown
- Add a slicer to select payload range
- Add success-payload-scatter-chart scatter plot based on the selected site dropdown
- Add a callback function

These actions were performed to inspect the relationship of success rate between launch site and payload.

Link: Github link for building a dashboard with plotly dash

Predictive Analysis (Classification)

Data wrangling

Data standarization

Split into traning and test datasets

Predictive model evalutation

- Logistic regression
- Support vector machine
- Decision tree classifer
- K-nearest neighnors

Predictive model selection

- K-nearest neighnors
- SVM
- Logistics regression

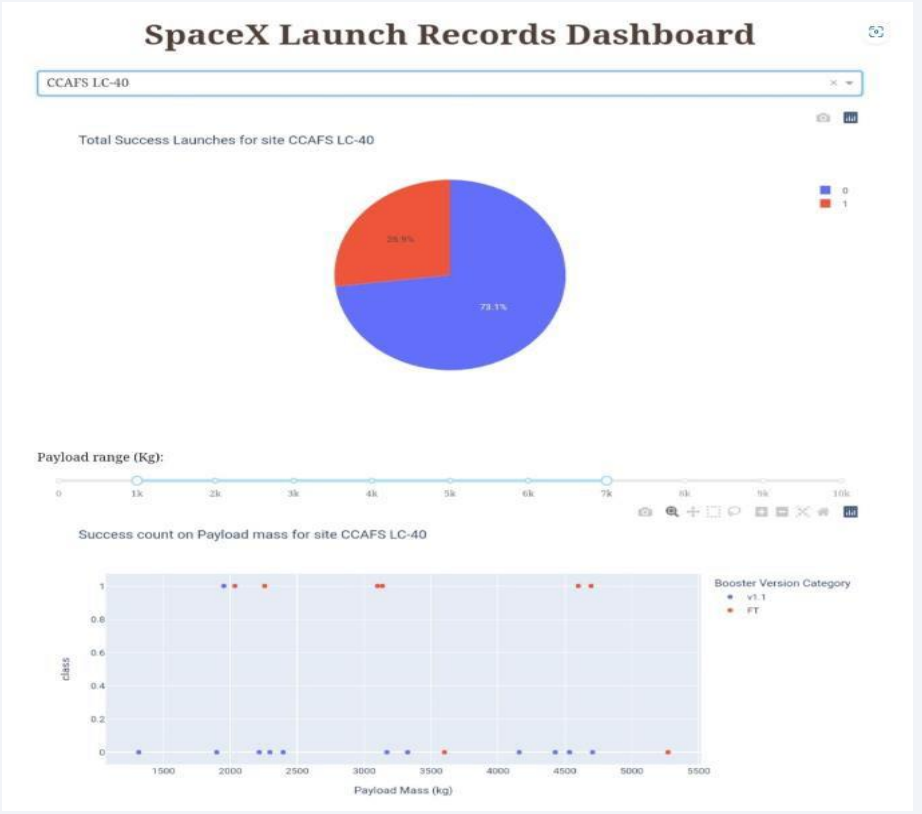
Link: GitHub link for Machine Learning Prediction

Results

Interactive analytics demo in screenshots

- Exploratory data analysis results

During the exploratory data analysis: • we have visualized the SpaceX launch dataset using matplotlib and seaborn and as a result we discovered some preliminary correlations between the launch site and success rates. • we have visualized the SpaceX launch dataset using Plotly Dash and as a result could define which site has the largest successful launches and has the highest launch success rate; which payload range(s) has the highest launch success rate and the lowest launch success rate; which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate.



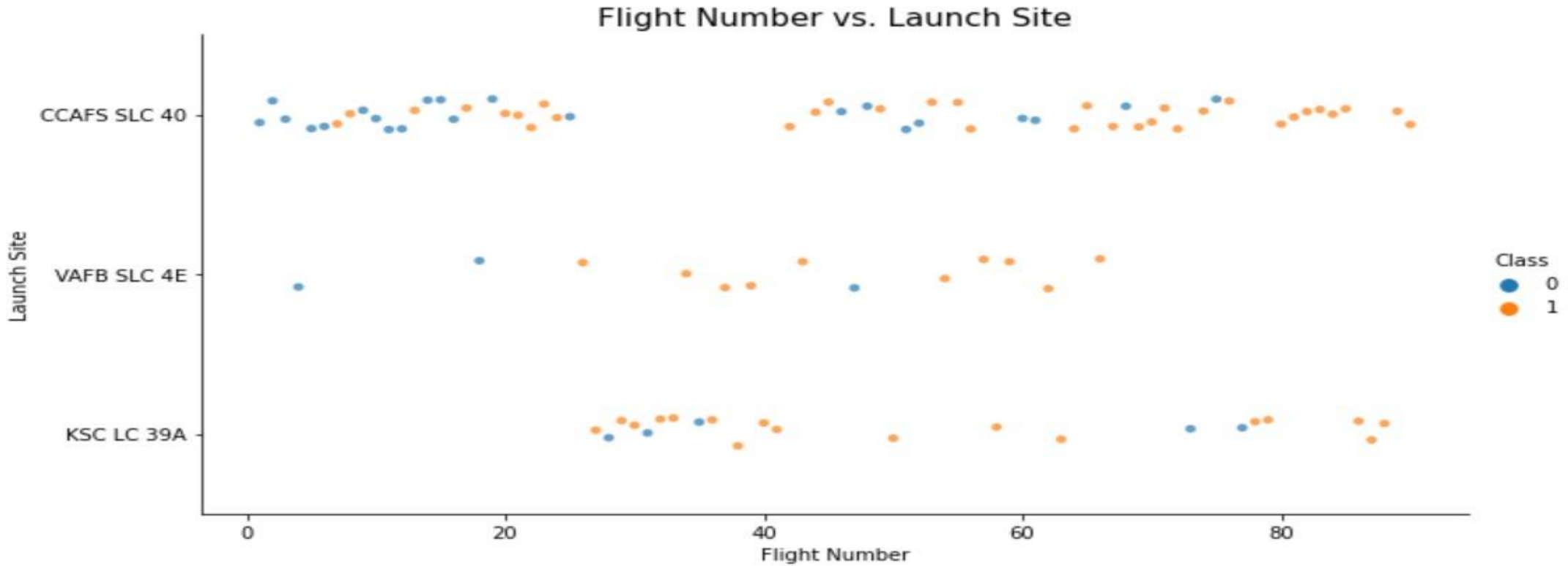
- Predictive analysis results

During the predictive analysis we used the best hyperparameter values and as a result we could determine the model with the best accuracy using the training data

Section 2

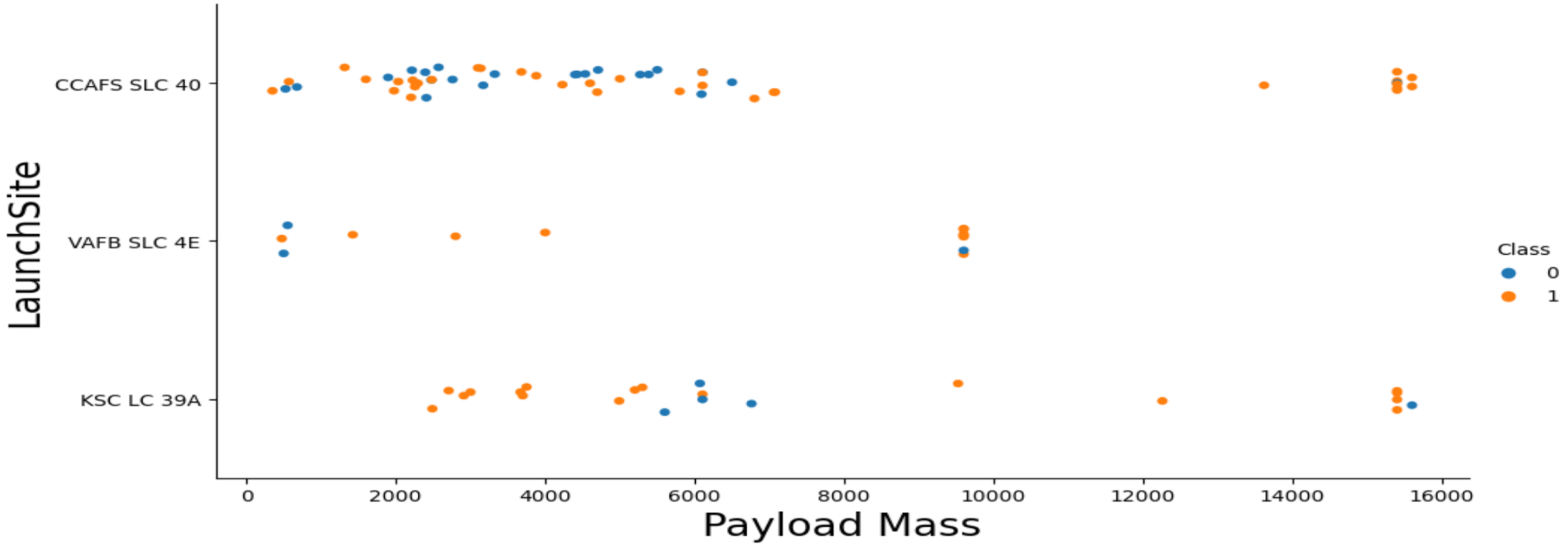
Insights drawn from EDA

Flight Number vs. Launch Site



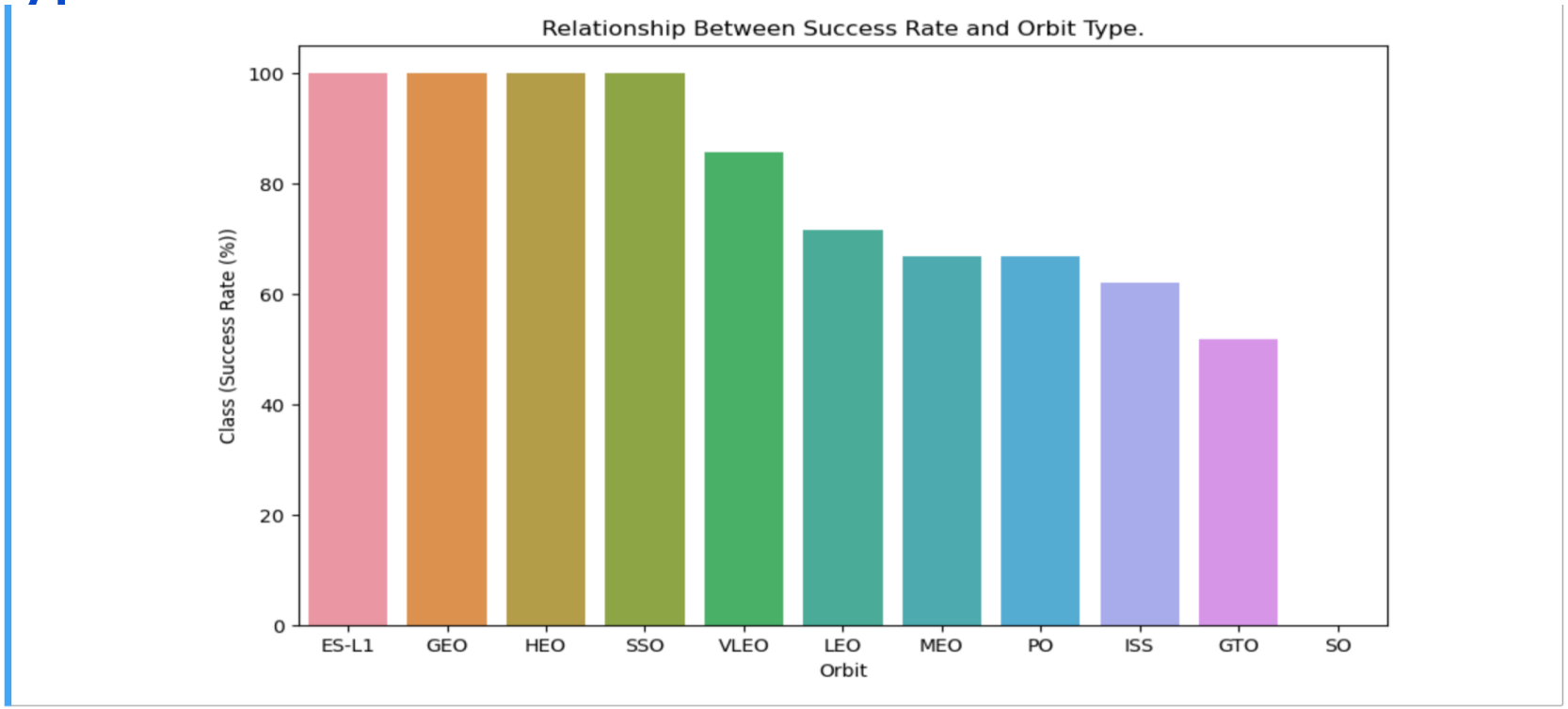
-> KSC LC 39A has the highest sucessful rate, vice verce CCAFS SLC 40 has the lowest

Payload vs. Launch Site



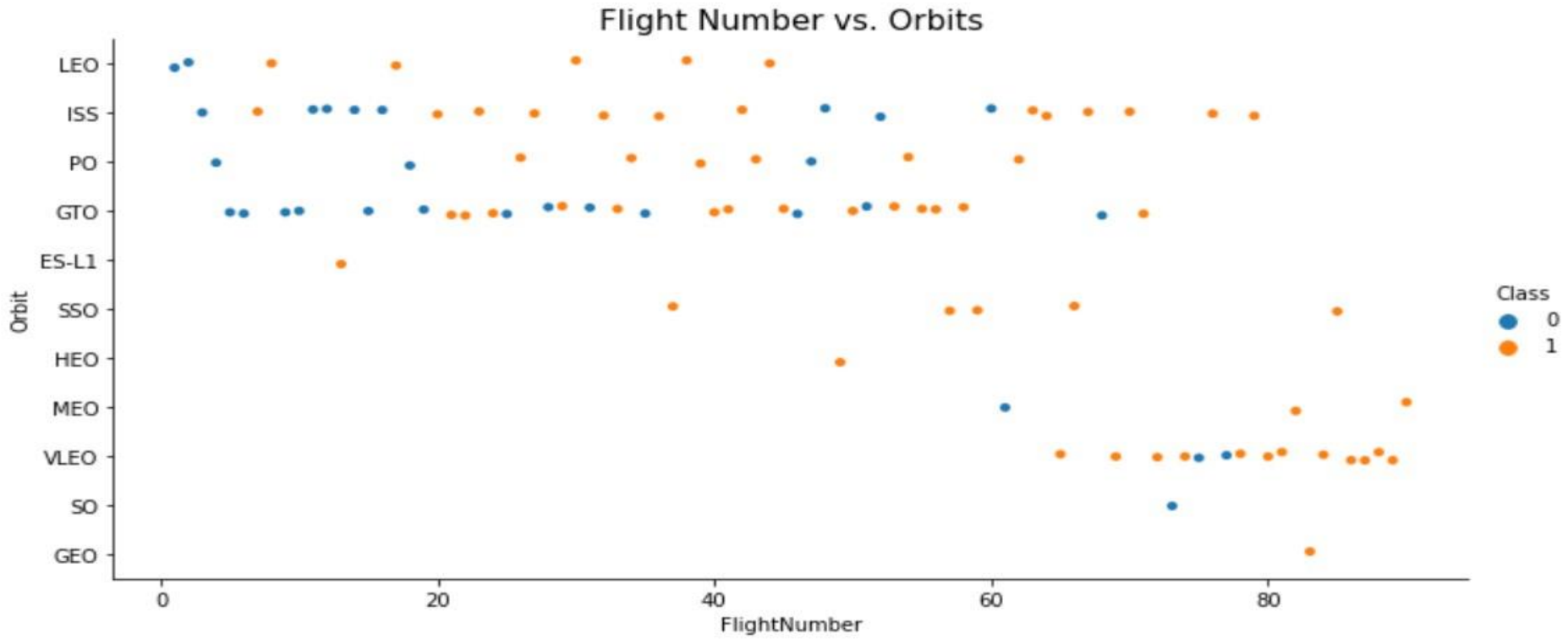
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



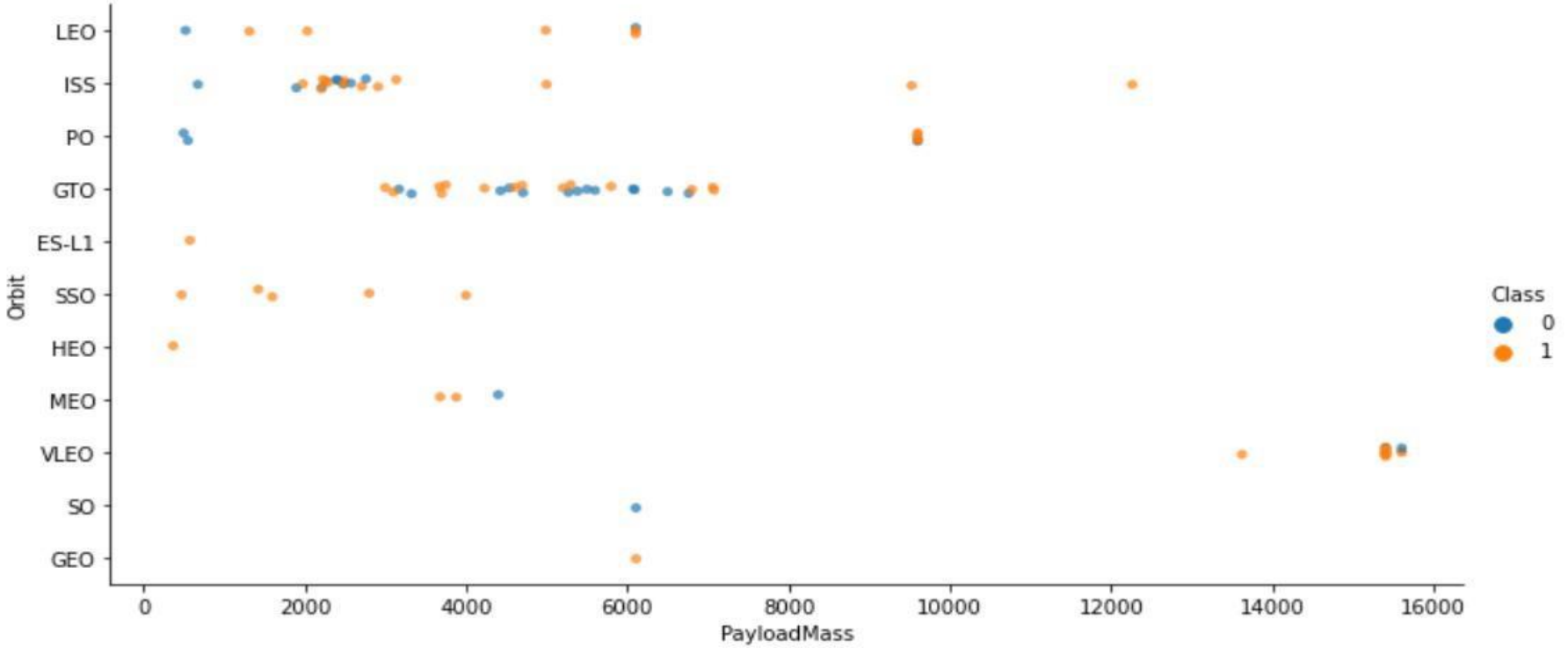
As can be seen, Orbit types ES-L1, GEO, HEO, and SSO have 100% success rate while ISS, and GTO have lower success rate. SO has 0

Flight Number vs. Orbit Type



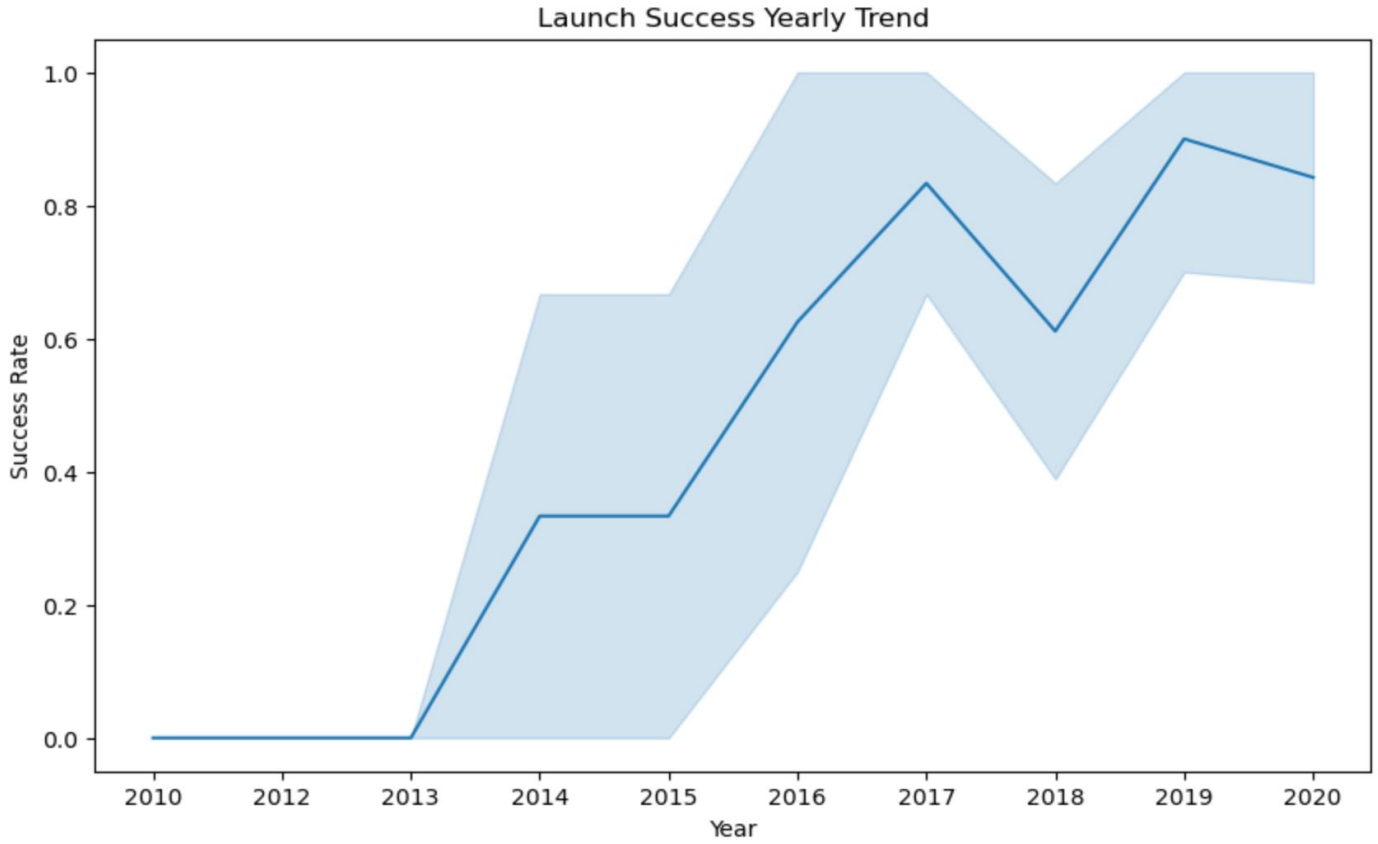
->In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



->With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

There are 4 different launch sites in total

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTBL;
```

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

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

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

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

Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
5:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
3:00	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)
4:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
5:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
0:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Total Payload mass for NASA (CRS) is 45,596 kg

In [11]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as "TOTAL PAYLOAD MASS (KGS)", Customer FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)'
```

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

Out[11]:

TOTAL PAYLOAD MASS (KGS)	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

In [23]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as 'AVERAGE PAYLOAD MASS', CUSTOMER, BOOSTER_VERSION FROM 'SPACEXTBL' WHERE Booster_Versi
```

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

Out[23]:

AVERAGE PAYLOAD MASS	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

First Successful Ground Landing Date

In [58]:
```
%sql SELECT MIN(DATE) AS "First Successful Landing Date" FROM 'SPACEXTBL' WHERE Landing_Outcome = "Success (ground pad)"
```

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

Out[58]:

First Successful Landing Date
2015-12-22

-> The first ground landing successful is on 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT Customer, Landing_Outcome,PAYLOAD_MASS__KG_ FROM SPACEXTBL
WHERE Landing_Outcome ='Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

Customer	Landing_Outcome	PAYLOAD_MASS__KG_
SKY Perfect JSAT Group	Success (drone ship)	4696
SKY Perfect JSAT Group	Success (drone ship)	4600
SES	Success (drone ship)	5300
SES EchoStar	Success (drone ship)	5200

-> The most successful landing is by drone ship.

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT Mission_Outcome, Count(*) AS Numbers FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

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

-> There are 1 failure in flight, 99 successes and 1 success with unclear payload status.

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version, Max_Payload FROM (SELECT Booster_Version, MAX(PAYLOAD_MASS__KG_)
                                                AS Max_Payload FROM SPACEXTBL GROUP BY Booster_Version)
```

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

Booster_Version	Max_Payload
F9 B4 B1039.2	2647
F9 B4 B1040.2	5384
F9 B4 B1041.2	9600
F9 B4 B1043.2	6460
F9 B4 B1039.1	3310
F9 B4 B1040.1	4990
F9 B4 B1041.1	9600
F9 B4 B1042.1	3500
F9 B4 B1043.1	5000
F9 B4 B1044	6092

2015 Launch Records

```
%sql SELECT SUBSTR(Date,4,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL
WHERE Landing_Outcome LIKE 'Failure%drone%' AND SUBSTR(Date,7,4) = '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

-> In January and April, 2015 there are launch failure by booster B1012 and B1015

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

In [106…

```
%sql SELECT Landing_Outcome, COUNT(*) AS Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Lan
```

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

Out[106…

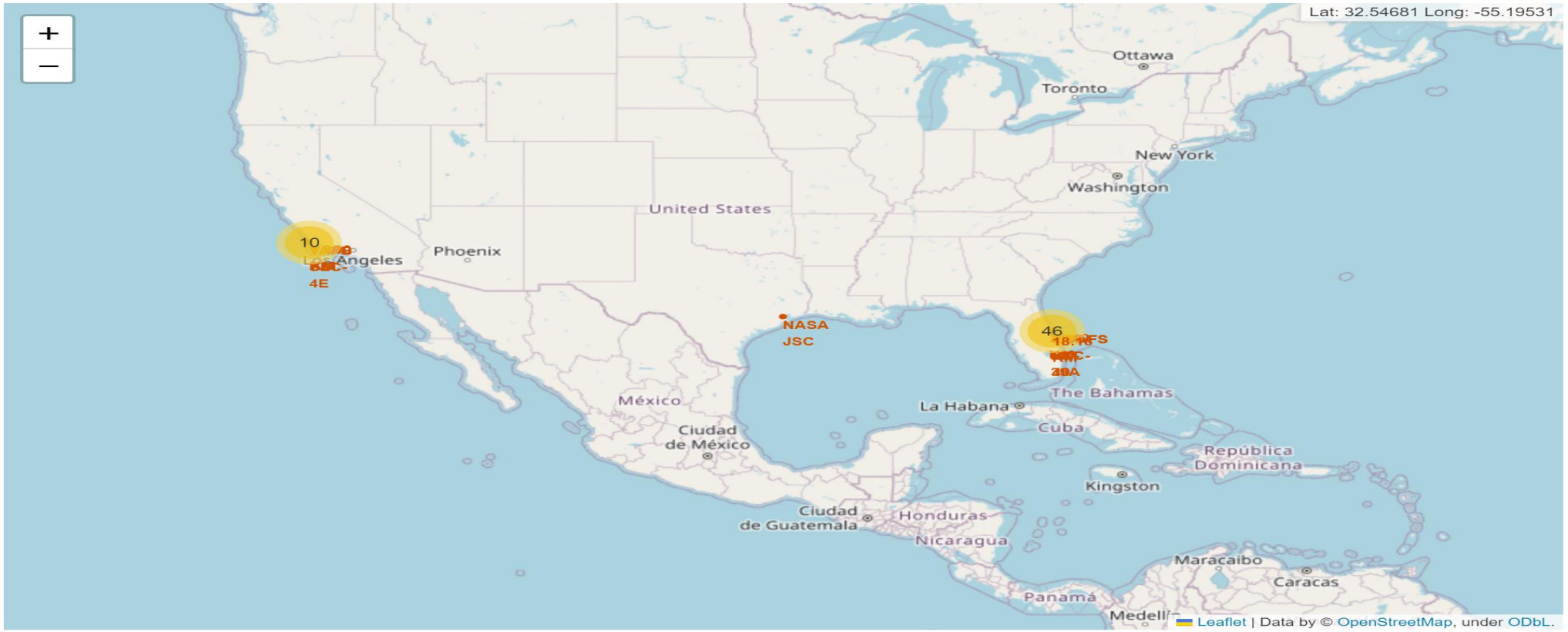
Landing_Outcome	Count
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

<Folium map showing all launch sites>

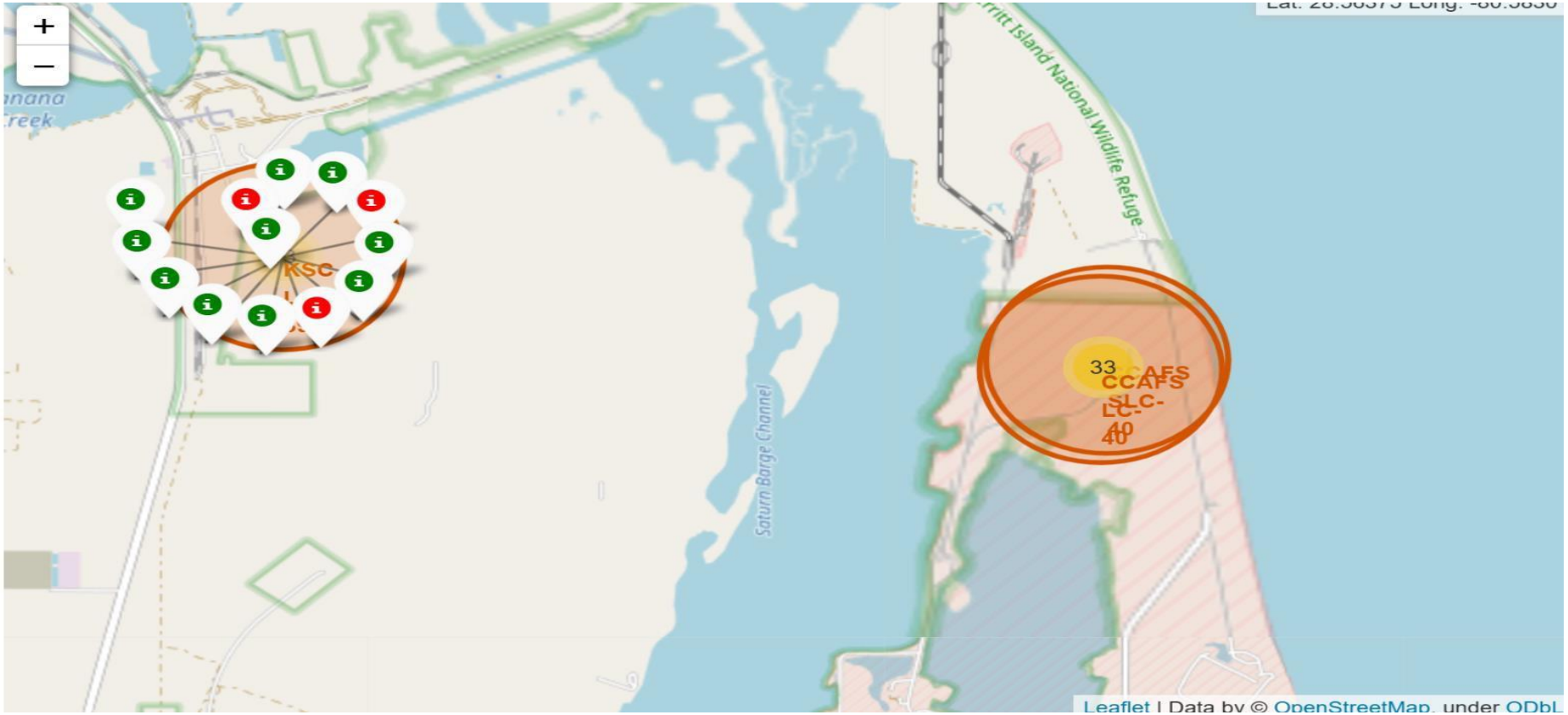
Out[114]:



and you should find a small yellow circle near the city of Houston and you can zoom-in to see a larger circle.

<Launch outcome of different sites>

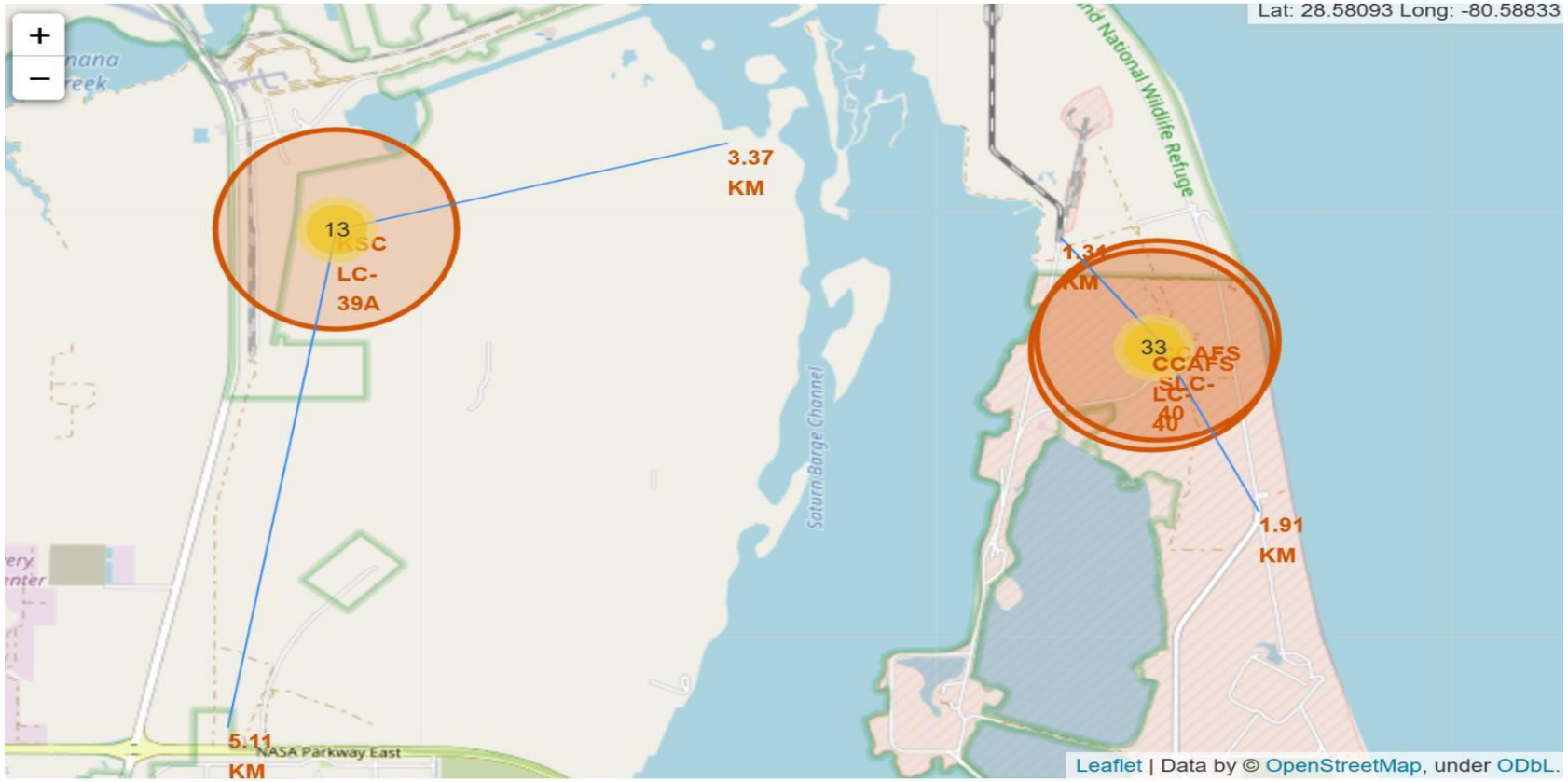
- Left coast site has 10 trails and right coast site has 46 trails

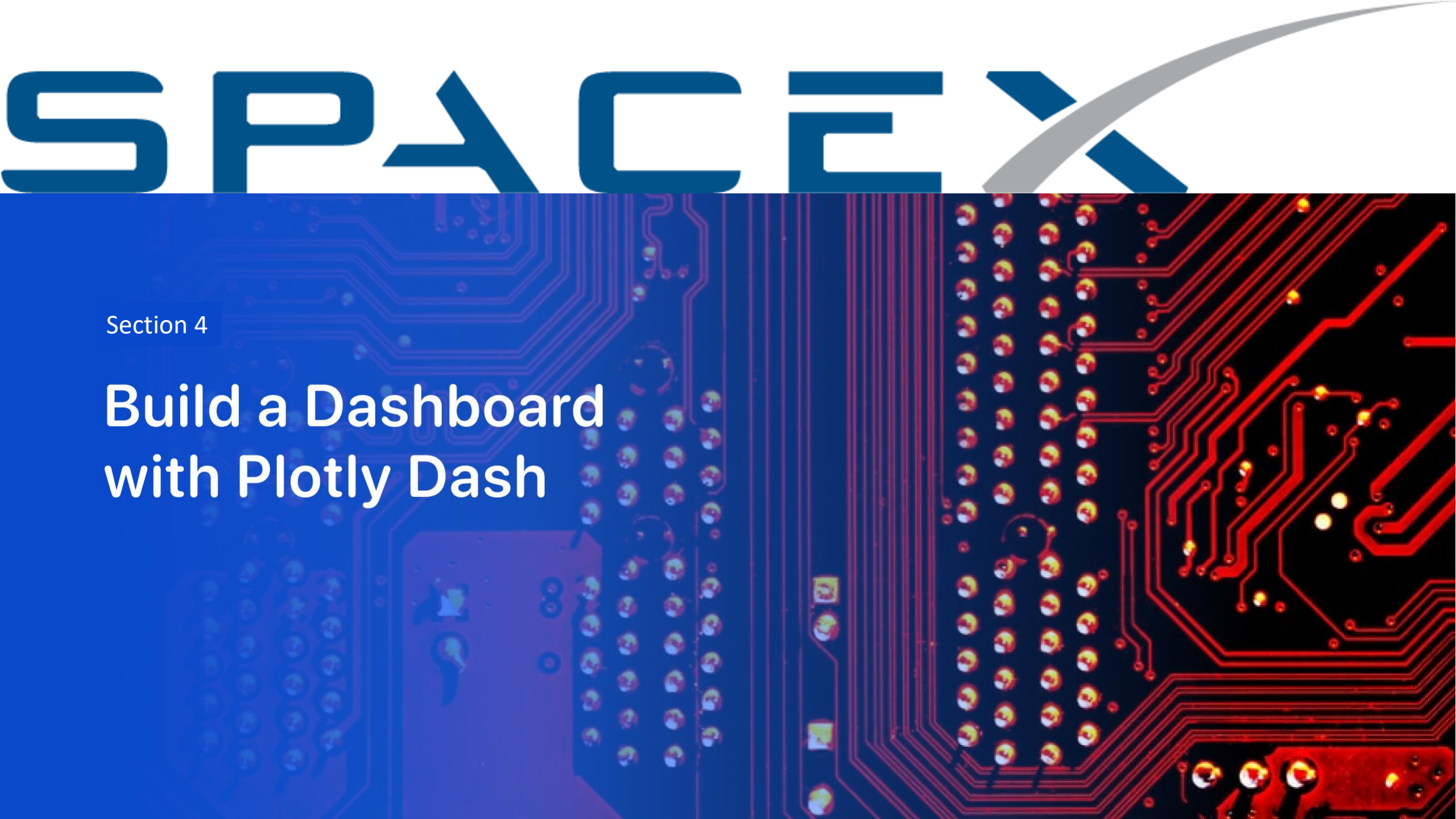


<Launch Sites Proximity To Landmarks>

- KSC LC-39 A is
  3.37 km far from the coast, and
  5.11 km from the city

- CCAFS LC-40 is
  1.91 km from the highway
  and
  1.34km from the railway





Section 4

Build a Dashboard with Plotly Dash

All site launch

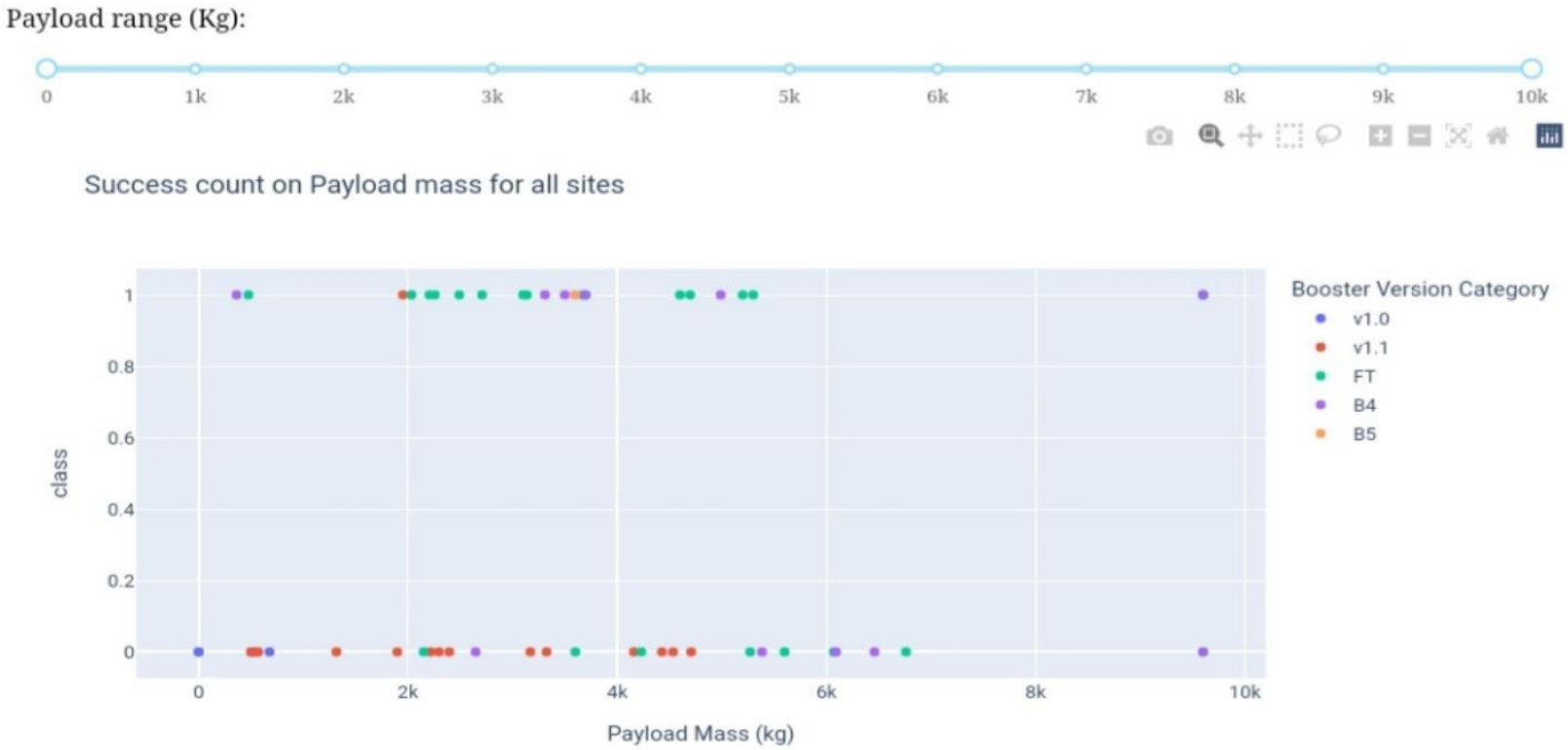
- KSC LC 39A: 41.7%
- CCAFS LC-40: 29.2%
- VAFB SLC -4E: 16.7%
- CCAFS SLC-40: 12.5%

Highest success launch ratio

- Success launch ratio: 73.1%

Payload vs. Launch Outcome

- V1.0 can take heaviest payload
- The success land happens between payload from 2k to 5k
- FT has the highest success rate



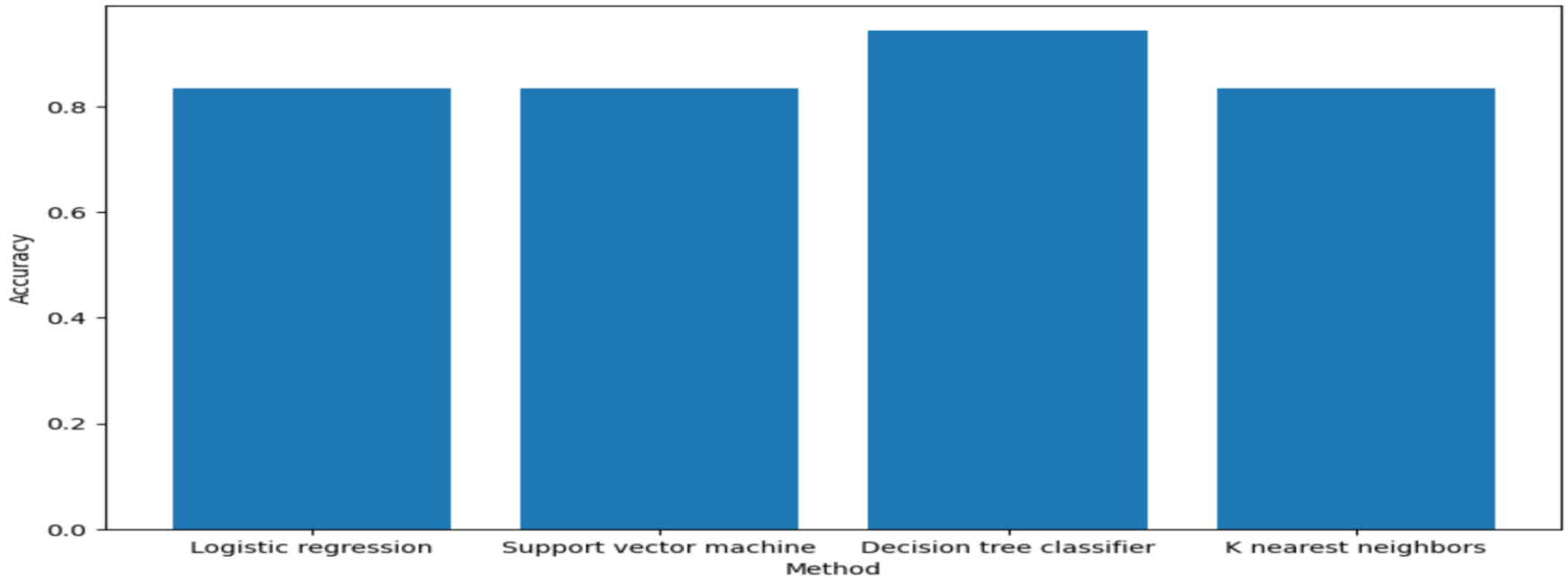
Section 5

Predictive Analysis (Classification)

Classification Accuracy

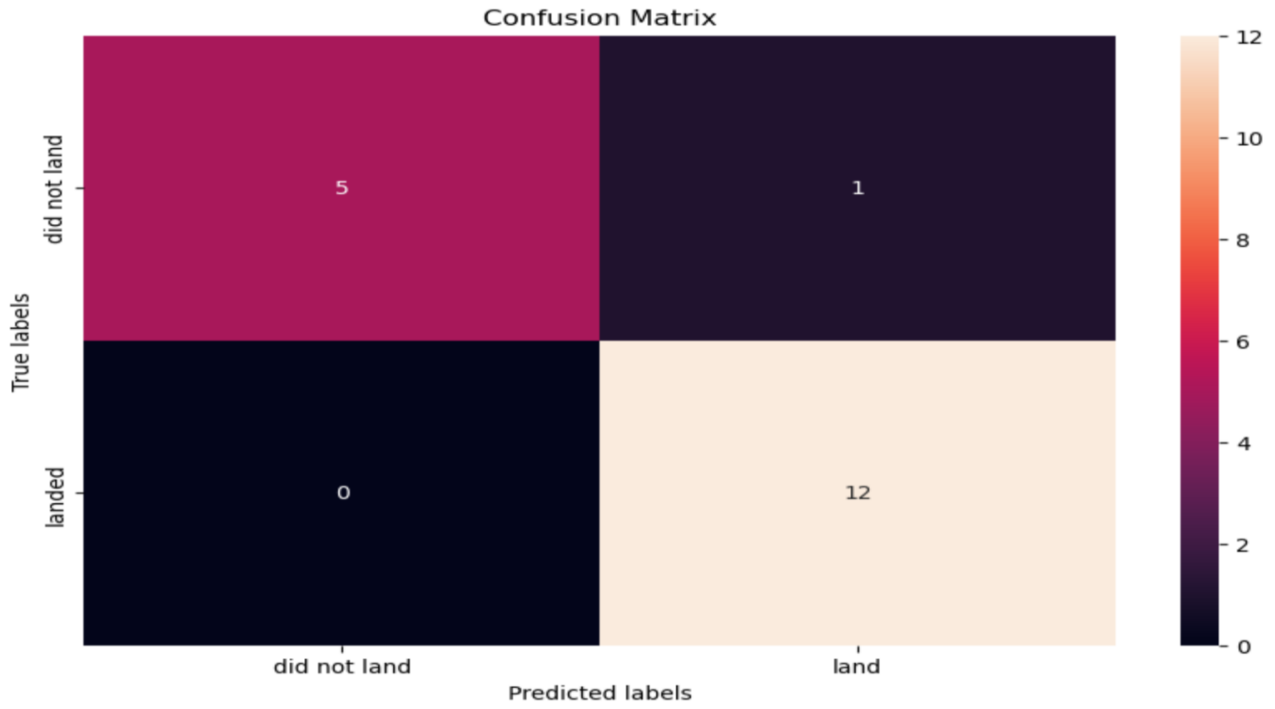
Plot a bar chart to show accuracy of the methods

In [93]:
```
plt.figure(figsize=(10, 6))
plt.bar(method, accuracy)
plt.xlabel('Method')
plt.ylabel('Accuracy')
plt.show()
```



Confusion Matrix

```
In [87]:
plt.figure(figsize=(10, 6))
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Conclusions

- There is a correlation between launch site and success rate Payload mass is also associated with the success rate.: the more massive the payload, the less likely the first stage will return
- For orbit type, SO has the least success rate while ES-L1, GEO, HEO and SSO have the highest success rate According to the yearly trend
- There has been an increase in the success rate since 2013 kept increasing till 2020.

Appendix

- https://github.com/Amazing-Ike/SpaceX-Falcon-9-first-stage-Landing-Prediction

SPACEX

Thank you!