IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Chobtham Thammakrang
5 April 2566

Outline

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

Executive Summary

- **Summary of methodologies**

  - Data Collection using SpaceX API & Web Scraping

  - Data Wrangling

  - EDA

  - Visualization Analysis

  - ML Prediction

- **Summary of all results**

  - Create a machine learning pipeline to predict if

    The Falcon 9 first stage will land successfully

- **Project background and context**
- - This project will predict if the Falcon 9 first stage will land successfully. 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. Therefore if we can determine if 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 SpaceX for a rocket launch.
- **Problems you want to find answers**
- - Determine if the first stage of Falcon 9 will land successfully

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
    - using SpaceX API & Web Scraping
- Perform data wrangling
    - Using pandas & numpy library
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

- Data Collection – SpaceX API
- - requesting rocket launch data from SpaceX API with the URL.
- - make the requested JSON results more consistent, will use the static response object.
- - decode the response content as a Json and turn it into a Pandas dataframe.
- Data Collection – Scraping
- - perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.
- - Create a BeautifulSoup object from the HTML response.
- - Extract all column/variable names from the HTML table header.
- - Create a data frame by parsing the launch HTML table.

Data Collection – SpaceX API

In [9]:
```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

We should see that the request was successfull with the 200 status response code

In [10]:
```
response.status_code
```

Out[10]: 200

- GitHub URL of the completed SpaceX API calls notebook (https://github.com/Ab1u3M8/Predict-if-SpaceX-Falcon-9-first-stage-will-land-successfully/blob/main/spacex-data-collection-api.ipynb)

In [11]:
```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

Using the dataframe `data` print the first 5 rows

In [12]:
```
# Get the head of the dataframe
data.head()
```

Out[12]:

	static_fire_date_utc	static_fire_date_unix	net	window	rocket	success	failures	details	crew	ships	capsules	payloads	
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}]	Engine failure at 33 seconds and loss of vehicle	[]	[]	[]	[5eb0e4b5b6c3bb0006eeb1e1]	5e9e45(

Data Collection - Scraping

- GitHub URL of the completed web scraping notebook(https://github.com/Ab1u3M8/Predict-if-SpaceX-Falcon-9-first-stage-will-land-successfully/blob/main/SpaceX-data-webscraping.ipynb)

```
# use requests.get() method with the provided static_url
# assign the response to a object
html_response = requests.get(static_url)
```

Create a `BeautifulSoup` object from the HTML `response`

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(html_response.text, 'html.parser')
```

Data Wrangling

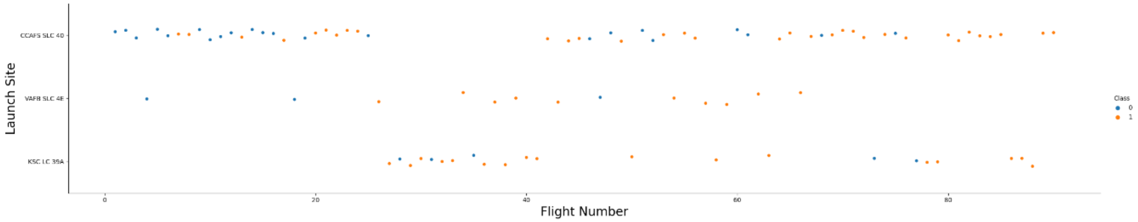
```
In [7]: # Apply value_counts() on column LaunchSite
        df['LaunchSite'].value_counts()

Out[7]: CCAFS SLC 40    55
        KSC LC 39A      22
        VAFB SLC 4E     13
        Name: LaunchSite, dtype: int64
```

- 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 occurence of mission outcome per orbit type
- Create a landing outcome label from Outcome column
- GitHub URL of completed data wrangling notebooks(https://github.com/Ab1u3M8/Predict-if-SpaceX-Falcon-9-first-stage-will-land-successfully/blob/main/module_1_spacex-data_wrangling.ipynb)

EDA with Data Visualization

```python
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(data = df, x = "FlightNumber", y = "LaunchSite", hue = "Class", aspect = 5)
plt.xlabel("Flight Number", fontsize = 20)
plt.ylabel("Launch Site", fontsize = 20)
plt.show()
```



- 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
- Create dummy variables to categorical columns
- Cast all numeric columns to float64
- GitHub URL of completed EDA with data visualization notebook
- (https://github.com/Ab1u3M8/Predict-if-SpaceX-Falcon-9-first-stage-will-land-successfully/blob/main/EDA_for_Data_Visualization.ipynb)

EDA with SQL

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

```
%%sql
SELECT "Landing _Outcome",count("Landing _Outcome")as LANDING_OUTCOME_COUNT,DATE
from SPACEXTBL where substr(Date,7,4) || substr(Date,4,2) || substr(Date,1,2) between '20100604'
and '20170320'
group by "Landing _Outcome" order by count("Landing _Outcome") desc
```

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

Landing _Outcome	LANDING_OUTCOME_COUNT	Date
No attempt	10	22-05-2012
Success (drone ship)	5	08-04-2016
Failure (drone ship)	5	10-01-2015
Success (ground pad)	3	22-12-2015
Controlled (ocean)	3	18-04-2014
Uncontrolled (ocean)	2	29-09-2013
Failure (parachute)	2	04-06-2010
Precluded (drone ship)	1	28-06-2015

- List the date when the first succesful landing outcome in ground pad was acheived.
- 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. Use a subquery
- 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.
- GitHub URL of completed EDA with SQL notebook,(https://github.com/Ab1u3M8/Predict-if-SpaceX-Falcon-9-first-stage-will-land-successfully/blob/main/EDA-sql.ipynb)

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
- GitHub URL of completed interactive map with Folium map,(https://github.com/Ab1u3M8/Predict-if-SpaceX-Falcon-9-first-stage-will-land-successfully/blob/main/module_3_launch_site_location.ipynb)

Predictive Analysis (Classification)

- create a column for the class
- Standardize the data
- Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data
- [GitHub URL](https://github.com/Ab1u3M8/Predict-if-SpaceX-Falcon-9-first-stage-will-land-successfully/blob/main/module_4_SpaceX_Machine_Learning_Prediction.ipynb) of completed predictive analysis, (https://github.com/Ab1u3M8/Predict-if-SpaceX-Falcon-9-first-stage-will-land-successfully/blob/main/module_4_SpaceX_Machine_Learning_Prediction.ipynb)

Find the method performs best:

```
Report = pd.DataFrame({'Method' : ['Accuracy']})

knn_accuracy=knn_cv.score(X_test, Y_test)
Decision_tree_accuracy=tree_cv.score(X_test, Y_test)
SVM_accuracy=svm_cv.score(X_test, Y_test)
Logistic_Regression=logreg_cv.score(X_test, Y_test)

Report['Logistic_Reg'] = [Logistic_Regression]
Report['SVM'] = [SVM_accuracy]
Report['Decision Tree'] = [Decision_tree_accuracy]
Report['KNN'] = [knn_accuracy]

Report.transpose()
```

	0
Method	Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.666667
KNN	0.833333

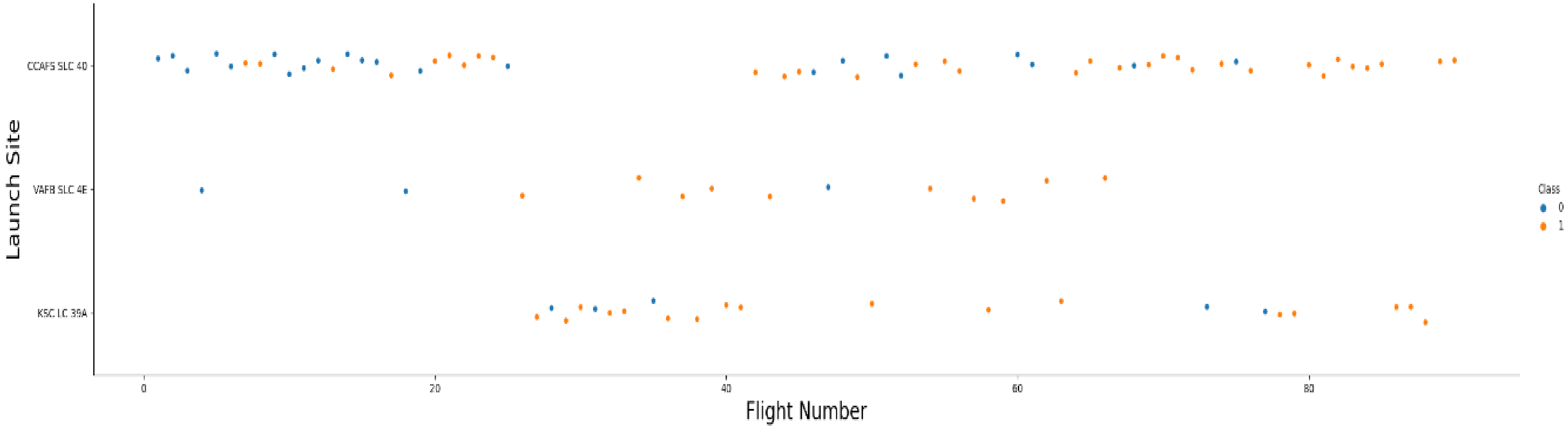
Results

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

Section 2

Insights drawn from EDA

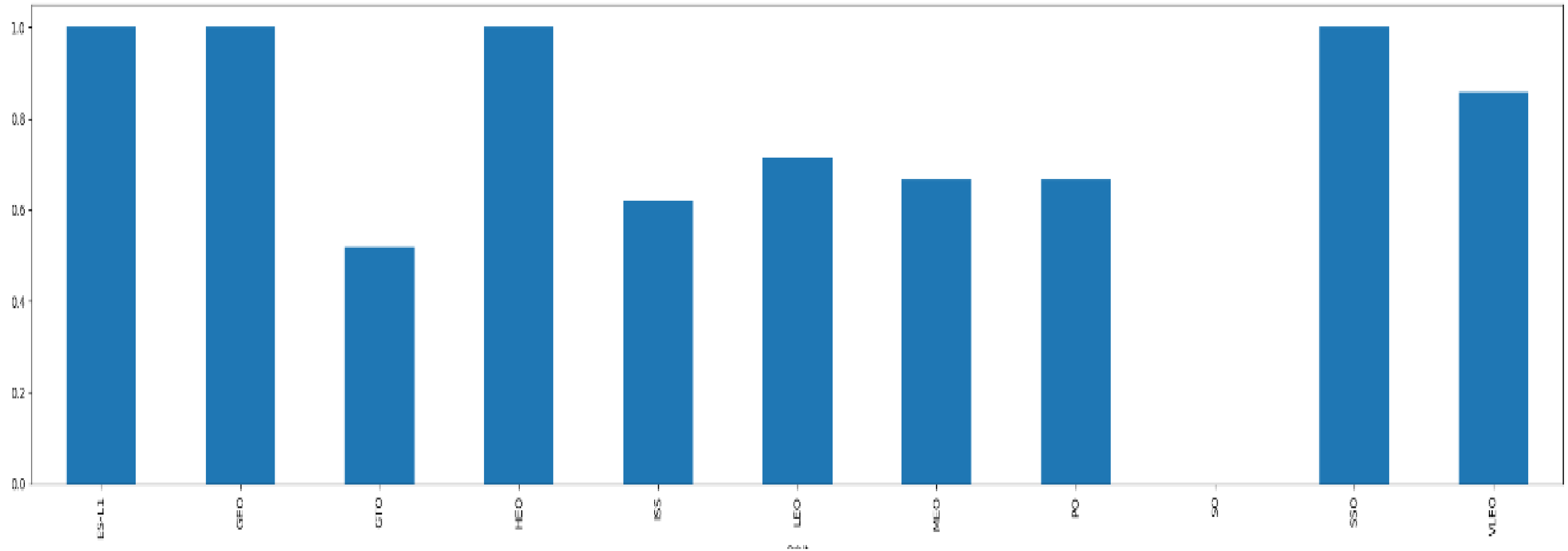
Flight Number vs. Launch Site



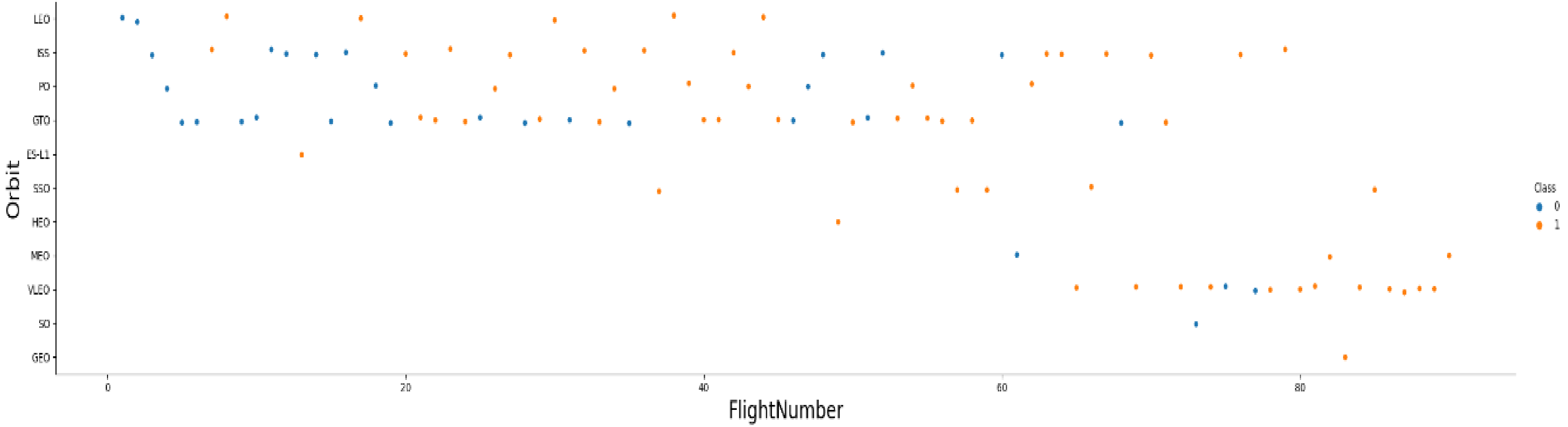
Payload vs. Launch Site

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

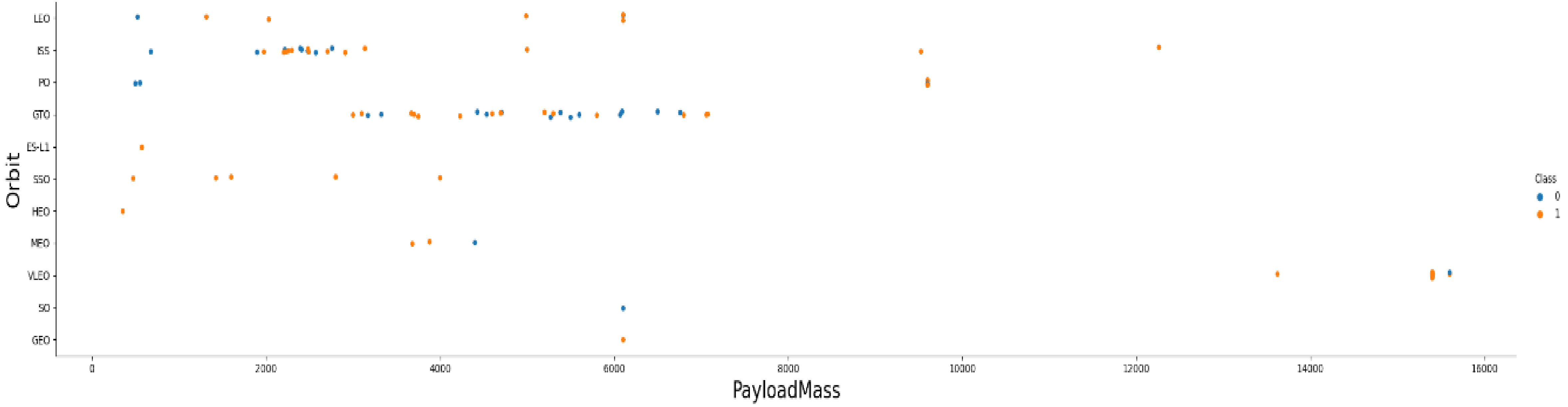


Flight Number vs. Orbit Type



You should see that 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.

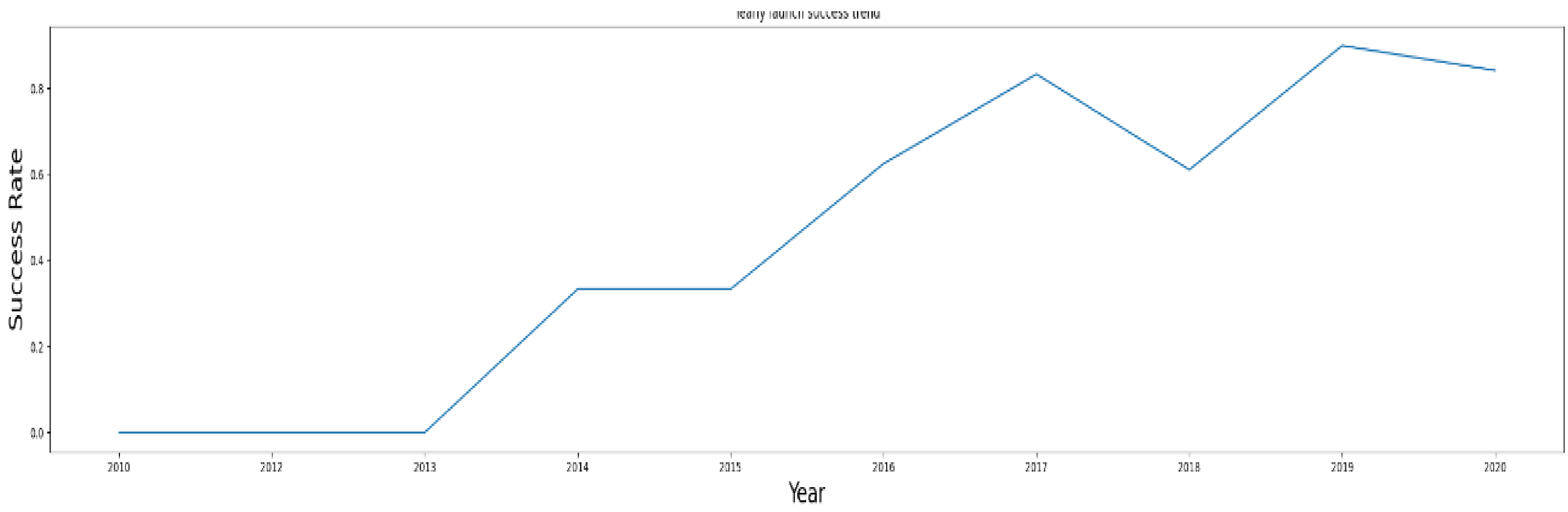
Payload vs. Orbit Type



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

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

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	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 from SPACEXTBL where customer like 'NASA (CRS)'
```

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

sum
45596

Average Payload Mass by F9 v1.1

```
%sql select avg(payload_mass__kg_) as average from SPACEXTBL where booster_version like 'F9 v1.1%'
```

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

average
2534.6666666666665

First Successful Ground Landing Date

```
#%sql select min(date) as date from SPACEXTBL where mission_outcome like 'Success'
%sql SELECT * FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (ground pad)"
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
22-12-2015	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
01-05-2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
03-06-2017	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14-08-2017	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
07-09-2017	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990	LEO	U.S. Air Force	Success	Success (ground pad)
15-12-2017	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
08-01-2018	01:00:00	F9 B4 B1043.1	CCAFS SLC-40	Zuma	5000	LEO	Northrop Grumman	Success (payload status unclear)	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Total Number of Successful and Failure Mission Outcomes

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

2015 Launch Records

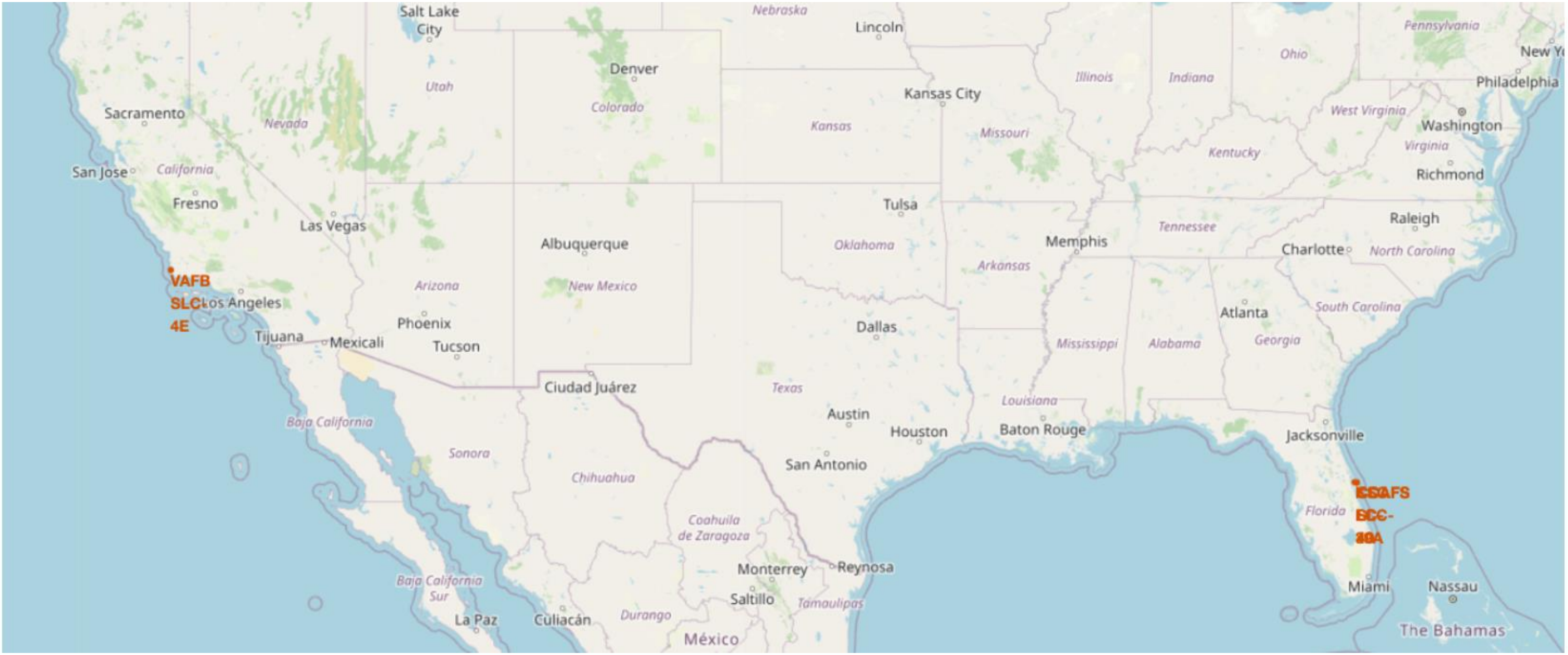
month	Booster_Version	Landing _Outcome
01	F9 v1.1 B1012	Failure (drone ship)
04	F9 v1.1 B1015	Failure (drone ship)

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

Landing _Outcome	LANDING_OUTCOME_COUNT	Date
No attempt	10	22-05-2012
Success (drone ship)	5	08-04-2016
Failure (drone ship)	5	10-01-2015
Success (ground pad)	3	22-12-2015
Controlled (ocean)	3	18-04-2014
Uncontrolled (ocean)	2	29-09-2013
Failure (parachute)	2	04-06-2010
Precluded (drone ship)	1	28-06-2015

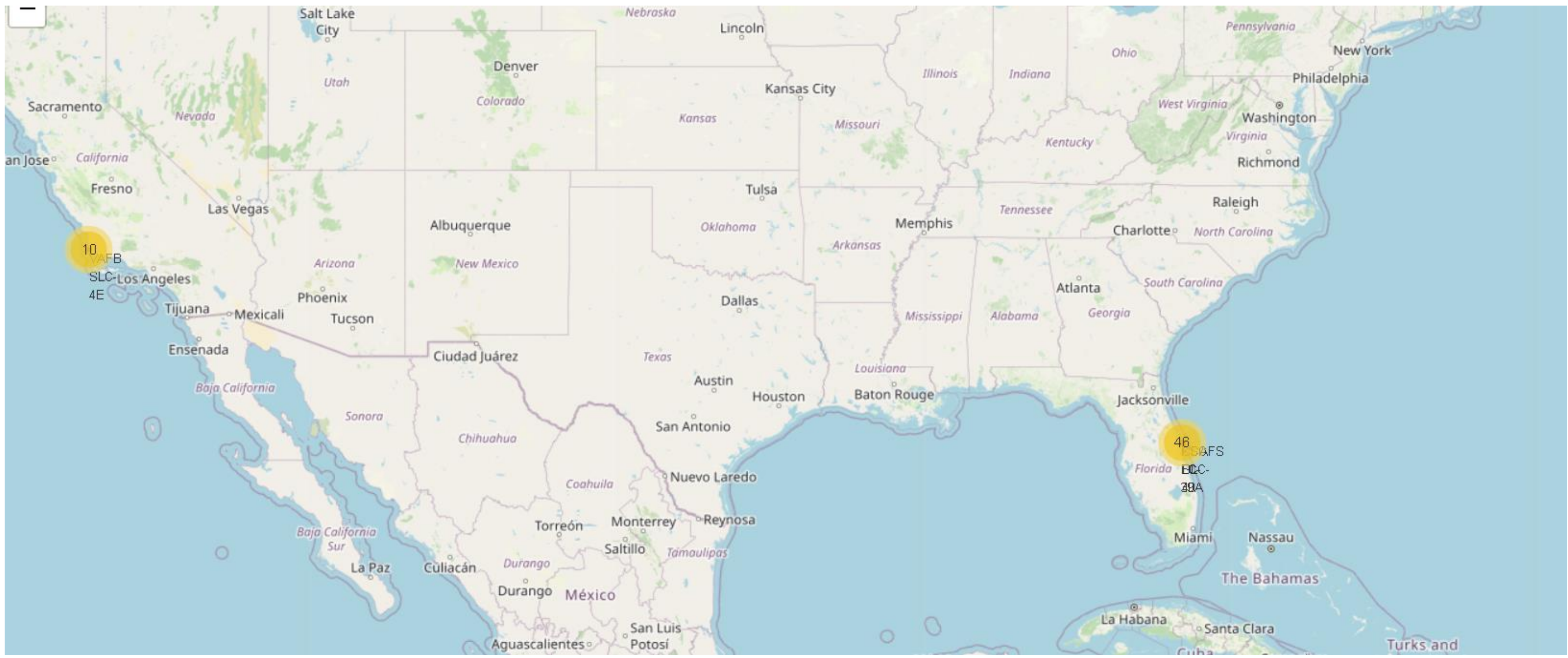
Section 3

Launch Sites Proximities Analysis

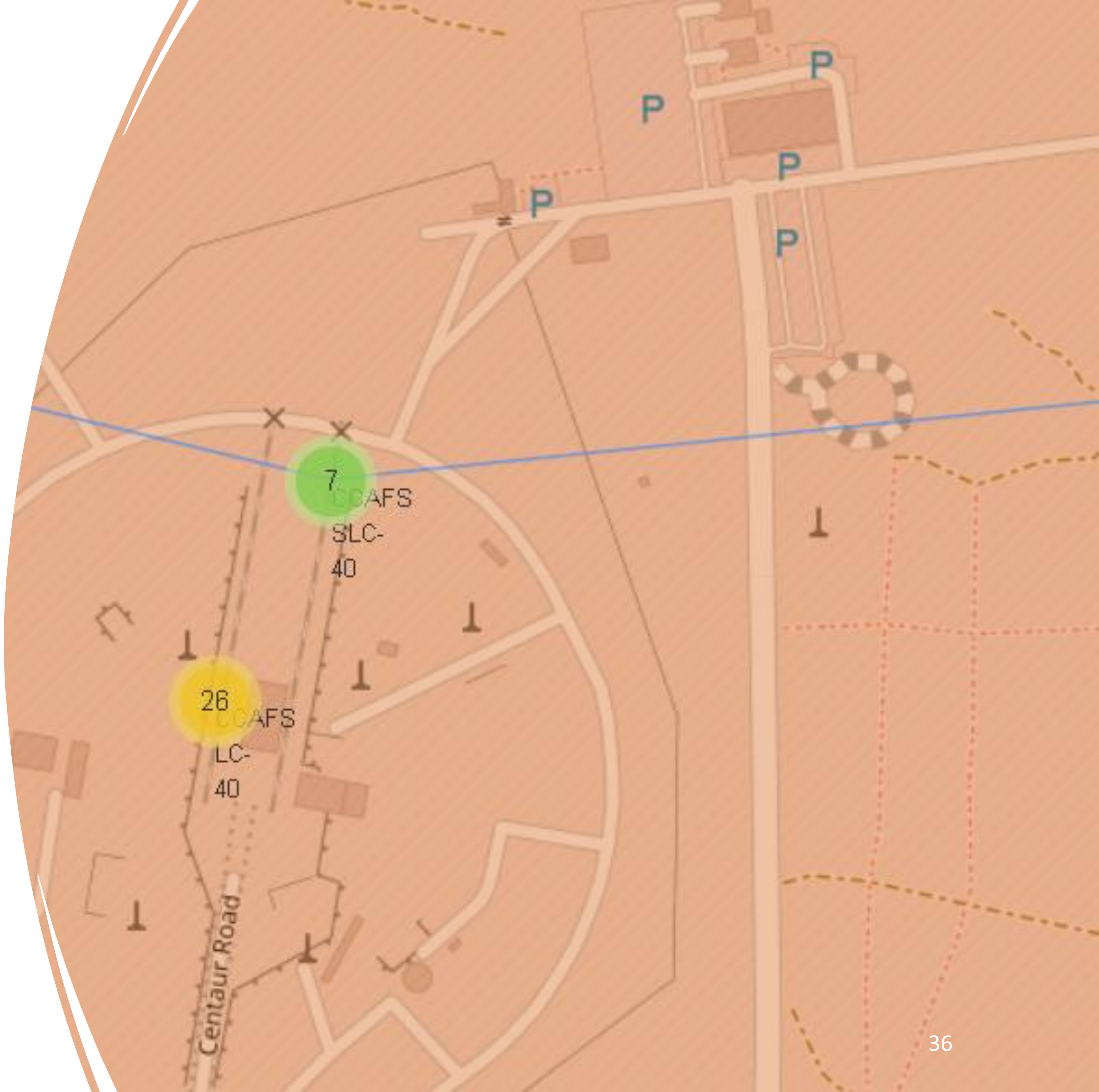


map with marked launch sites

Map marker with success & failed launch sites



map with the distances line between a launch site



Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

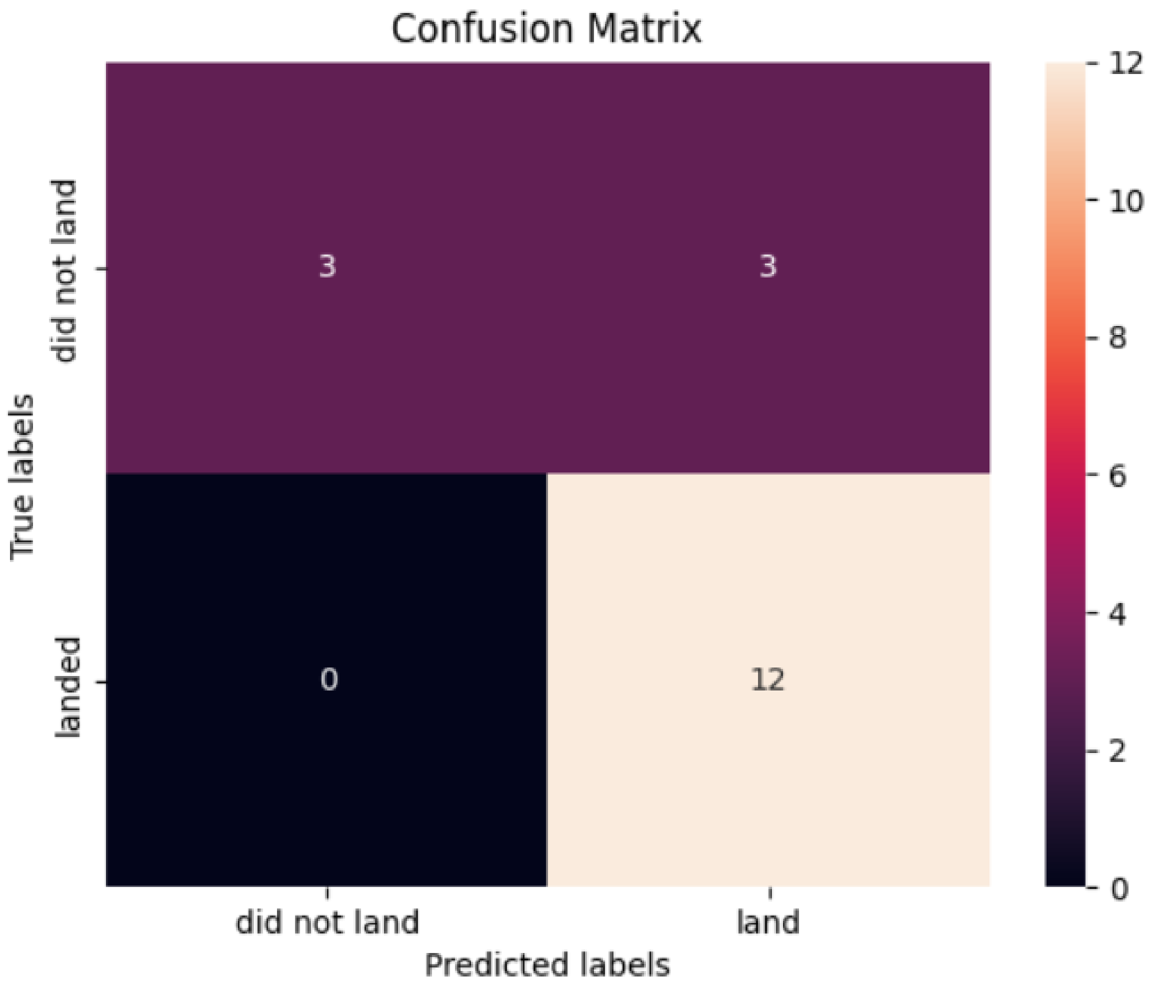
Predictive Analysis (Classification)

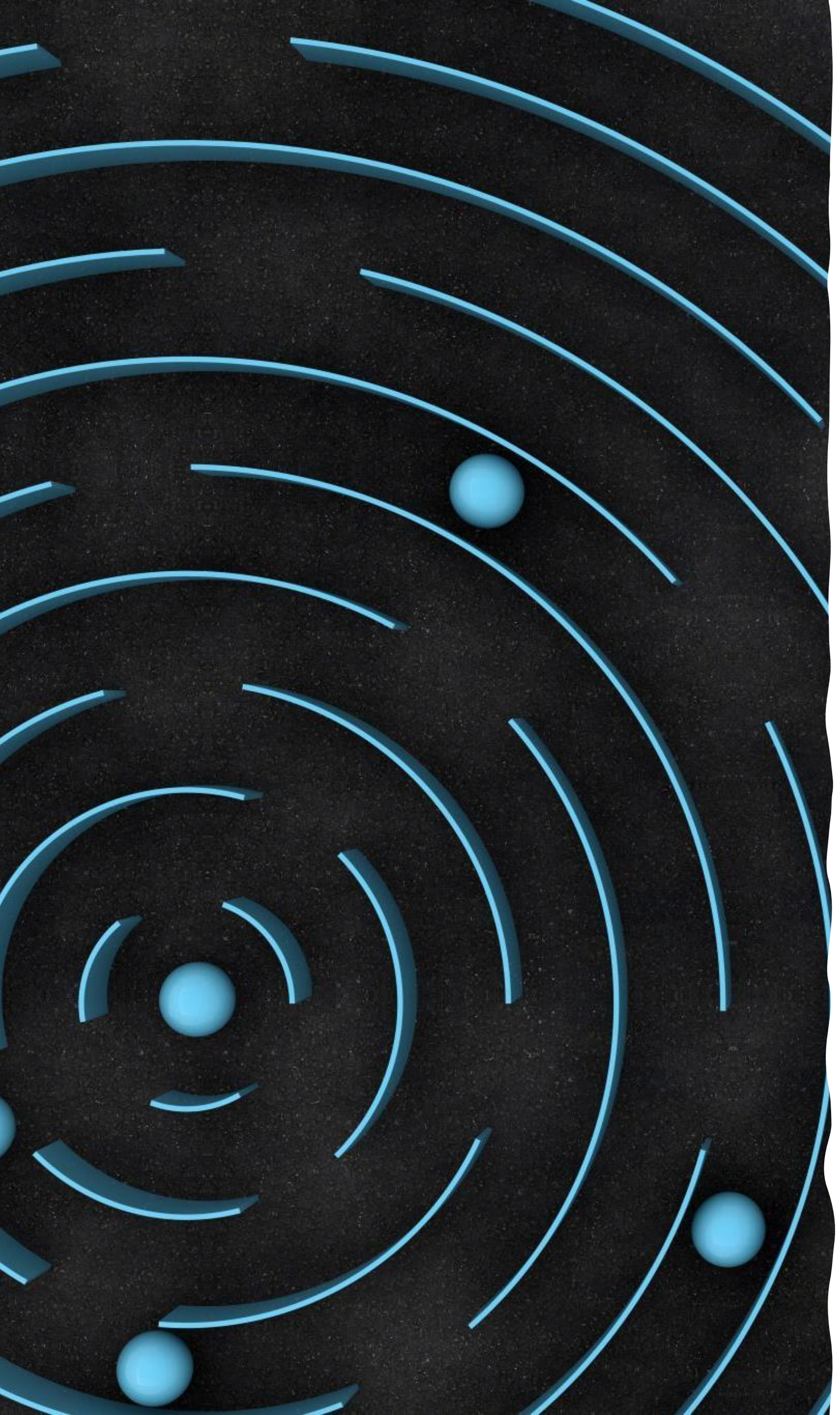
Classification Accuracy

Method	Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.666667
KNN	0.833333

Confusion Matrix

- The confusion matrixes of Knn, SVM, Logistic_Reg are the same accuracy.





Conclusions

- To increase the probability of a successful launch, orbits ES-L1, GEO, HEO, SSO as they have 100% successful flights.
- KSC LC-39A launch site has the highest ratio of successful.
- Weights over 2000kg and under 5500kg have a higher chance of successful.

Appendix

- GitHub(https://github.com/Ab1u3M8/Predict-if-SpaceX-Falcon-9-first-stage-will-land-successfully)

Thank you!