Predicting Falcon 9 Success:

A Data-Driven Approach to Rocket Landing Forecasting

Natasa Vakouftsi

Jan 2025

OUTLINE

- **Executive Summary**
- **Introduction**
- **Methodology**
  - Data collection and data wrangling methodology
  - EDA and interactive visual analytics methodology
  - Predictive analysis methodology
- **Results**
  - EDA with visualization results
  - EDA with SQL results
  - Interactive map with Folium results
  - Plotly Dash dashboard results
  - Predictive analysis (classification)
- **Conclusion**
- **Discussion**

EXECUTIVE SUMMARY

In this project, we leveraged data science techniques to predict whether the first stage of SpaceX's Falcon 9 rocket would land successfully. The ability to accurately predict the success of these landings is crucial for a competing startup seeking to make informed bids for rocket launches. By analyzing the historical dataset of Falcon 9 launches, we built predictive models that can help assess landing outcomes with high accuracy.

INTRODUCTION

The goal of this project is to predict the success of the first-stage landing of SpaceX's Falcon 9 rocket. Accurately predicting whether the rocket will land successfully is a critical component for SpaceX in terms of mission success, cost optimization, and resource allocation.

This prediction model serves the purpose of assisting in making more informed decisions on future launches. By analyzing historical launch data, we aim to determine the factors that influence landing success, thus improving the operational strategy and enhancing the company's overall performance.

The insights gained from this model can help optimize the design, testing, and overall planning for future missions, ultimately contributing to reliable and cost-effective space transportation.

METHODOLOGY

- Data collection and data wrangling methodology
- EDA and interactive visual analytics methodology
- Predictive analysis methodology

Data collection

```
# Show the head of the dataframe
print(launch_df.head())
```

```
   FlightNumber        Date BoosterVersion  PayloadMass Orbit  \
0             1  2006-03-24       Falcon 1         20.0   LEO
1             2  2007-03-21       Falcon 1          NaN   LEO
2             4  2008-09-28       Falcon 1        165.0   LEO
3             5  2009-07-13       Falcon 1        200.0   LEO
4             6  2010-06-04       Falcon 9          NaN   LEO

        LaunchSite    Outcome  Flights  GridFins  Reused   Legs LandingPad  \
0  Kwajalein Atoll  None None        1     False   False  False       None
1  Kwajalein Atoll  None None        1     False   False  False       None
2  Kwajalein Atoll  None None        1     False   False  False       None
3  Kwajalein Atoll  None None        1     False   False  False       None
4     CCSFS SLC 40  None None        1     False   False  False       None

   Block  ReusedCount    Serial   Longitude   Latitude
0    NaN            0  Merlin1A  167.743129   9.047721
1    NaN            0  Merlin2A  167.743129   9.047721
2    NaN            0  Merlin2C  167.743129   9.047721
3    NaN            0  Merlin3C  167.743129   9.047721
4    1.0            0     B0003  -80.577366  28.561857
```

```
data_falcon9 = launch_df[launch_df['BoosterVersion'] != 'Falcon 1']
print(data_falcon9.head())
```

```
   FlightNumber        Date BoosterVersion  PayloadMass Orbit    LaunchSite  \
4             6  2010-06-04       Falcon 9          NaN   LEO  CCSFS SLC 40
5             8  2012-05-22       Falcon 9        525.0   LEO  CCSFS SLC 40
6            10  2013-03-01       Falcon 9        677.0   ISS  CCSFS SLC 40
7            11  2013-09-29       Falcon 9        500.0    PO   VAFB SLC 4E
8            12  2013-12-03       Falcon 9       3170.0   GTO  CCSFS SLC 40

       Outcome  Flights  GridFins  Reused   Legs LandingPad  Block  \
4    None None        1     False   False  False       None    1.0
5    None None        1     False   False  False       None    1.0
6    None None        1     False   False  False       None    1.0
7  False Ocean        1     False   False  False       None    1.0
8    None None        1     False   False  False       None    1.0

   ReusedCount Serial   Longitude   Latitude
4            0  B0003  -80.577366  28.561857
5            0  B0005  -80.577366  28.561857
6            0  B0007  -80.577366  28.561857
7            0  B1003 -120.610829  34.632093
8            0  B1004  -80.577366  28.561857
```

- We made a get request to the SpaceX API in order to obtain the data.
- The dataframe was filtered to include only Falcon 9 launches

Data collection

- Web scraping was performed for Falcon 9 and Falcon Heavy Launches Records from Wikipedia and the HTML table was used.
- We parsed the launch record values and created a dataframe.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...

Data Wrangling

The missing values in our dataset were addressed and replaced with the mean of the payload mass.

```
data_falcon9.isnull().sum()
```

```
FlightNumber       0
Date               0
BoosterVersion     0
PayloadMass        5
Orbit              0
LaunchSite         0
Outcome            0
Flights            0
GridFins           0
Reused             0
Legs               0
LandingPad        26
Block              0
ReusedCount        0
Serial             0
Longitude          0
Latitude           0
dtype: int64
```

EDA and interactive visual analytics methodology

We used the dataset from the EDA with SQL step to visualize the relationship between the following:

- Flight Number and Launch Site
- Payload Mass and Launch Site
- Success rate of each orbit type
- FlightNumber and Orbit type
- Payload Mass and Orbit type

and observed the launch success yearly trend.

```
from js import fetch
import io

URL = "https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/data
resp = await fetch(URL)
dataset_part_2_csv = io.BytesIO((await resp.arrayBuffer()).to_py())
df=pd.read_csv(dataset_part_2_csv)
df.head(5)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False

Predictive analysis

- The categorical variables (e.g., launch site, payload, and booster version) were cleaned, encoded and split the data into training and testing sets.
- We built multiple Classification Models (Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)) and compared the accuracy of tuned models.

```python
data = pd.read_csv(text1)
```

```python
data.head()
```

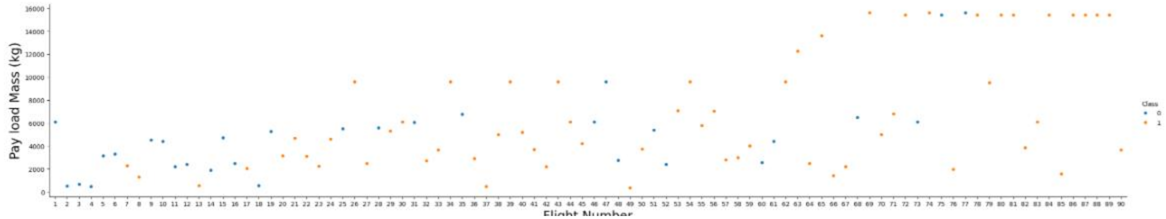
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False

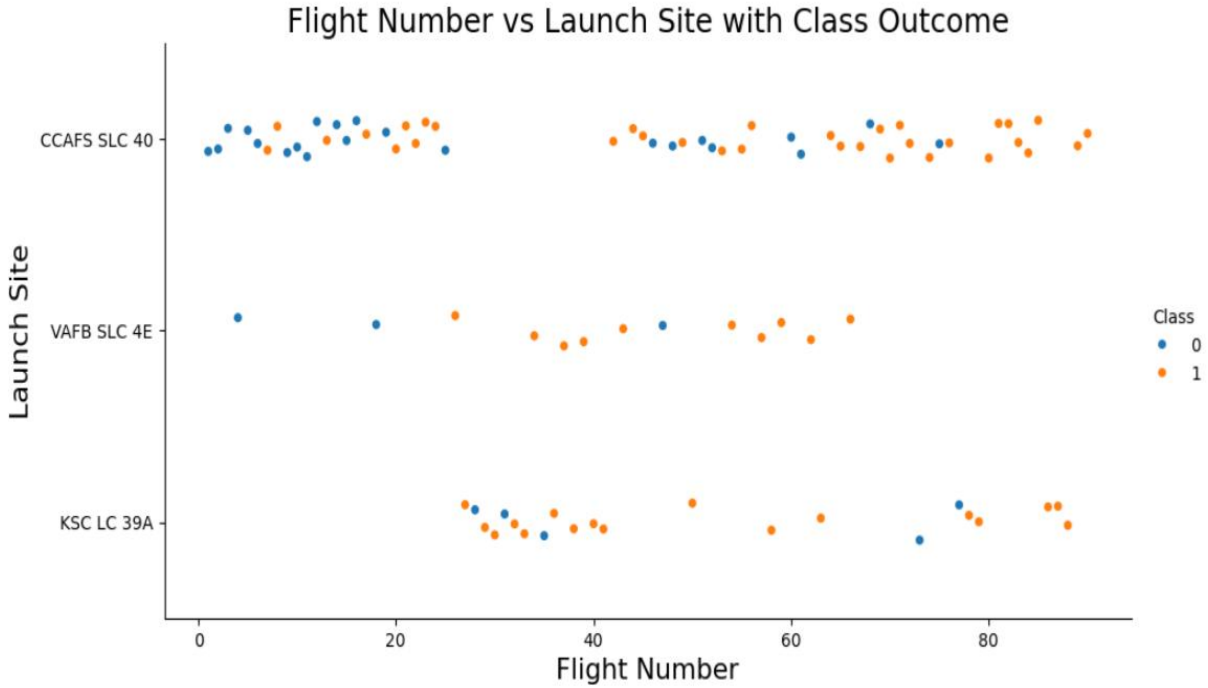
Results

- EDA with visualization results
- EDA with SQL results
- Interactive map with Folium results
- Plotly Dash dashboard results
- Predictive analysis (classification)

EDA with visualization results

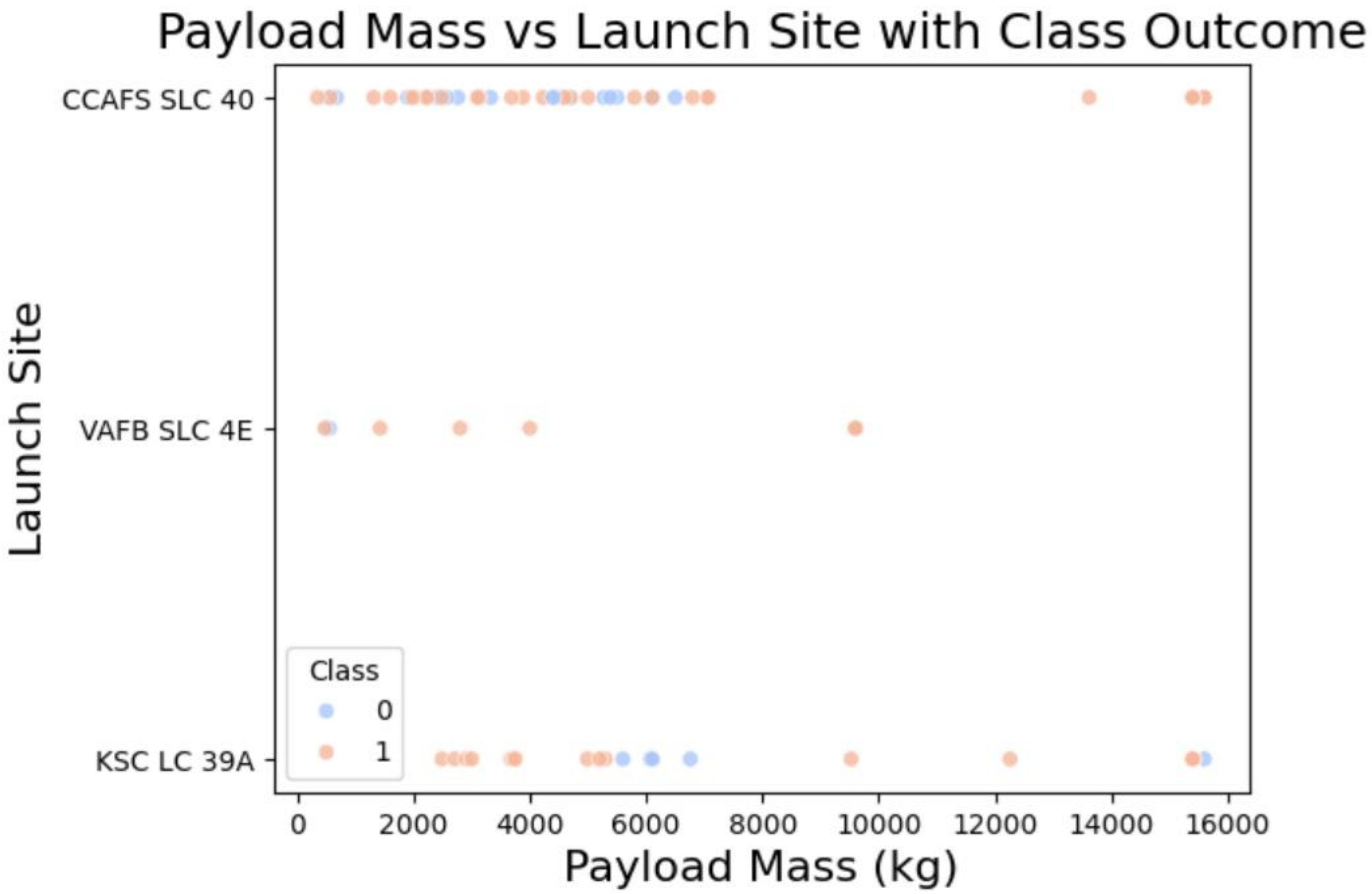
```
sns.catplot(y="PayloadMass", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Pay load Mass (kg)",fontsize=20)
plt.show()
```



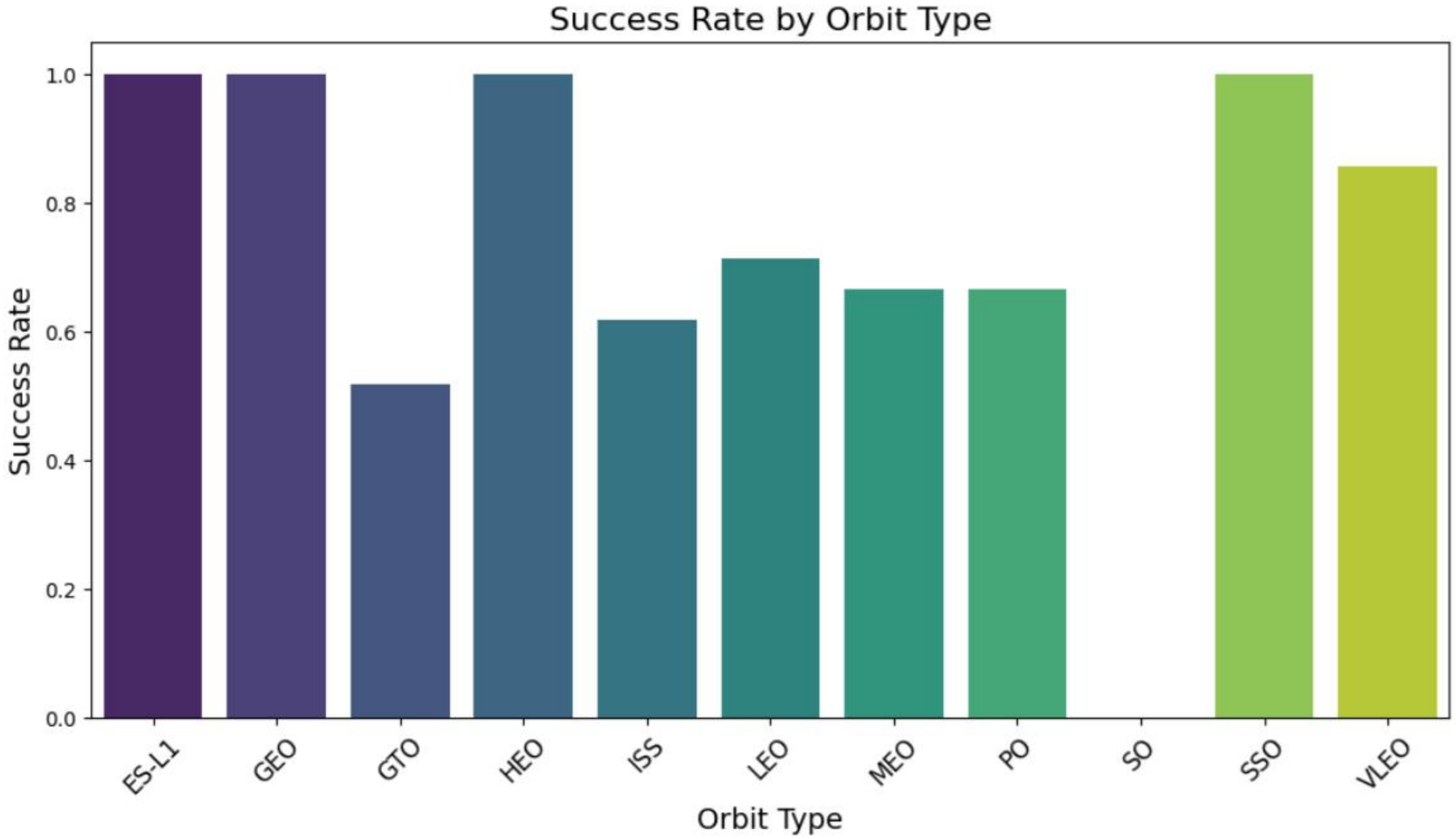


EDA with visualization results

For launch site VAFB SLC 4E there are no rockets launched with payload mass greater than 10000kg.

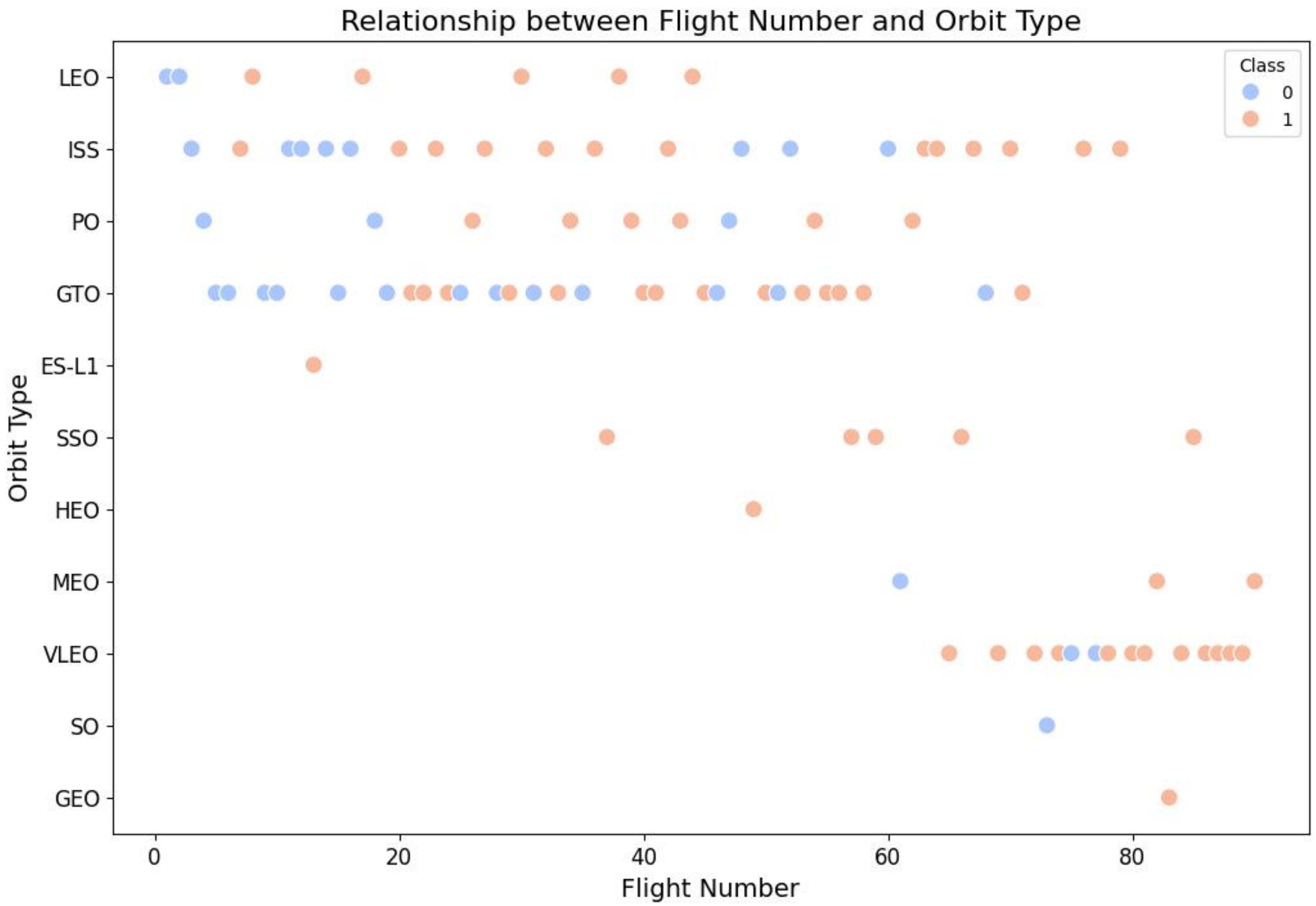


EDA with visualization results



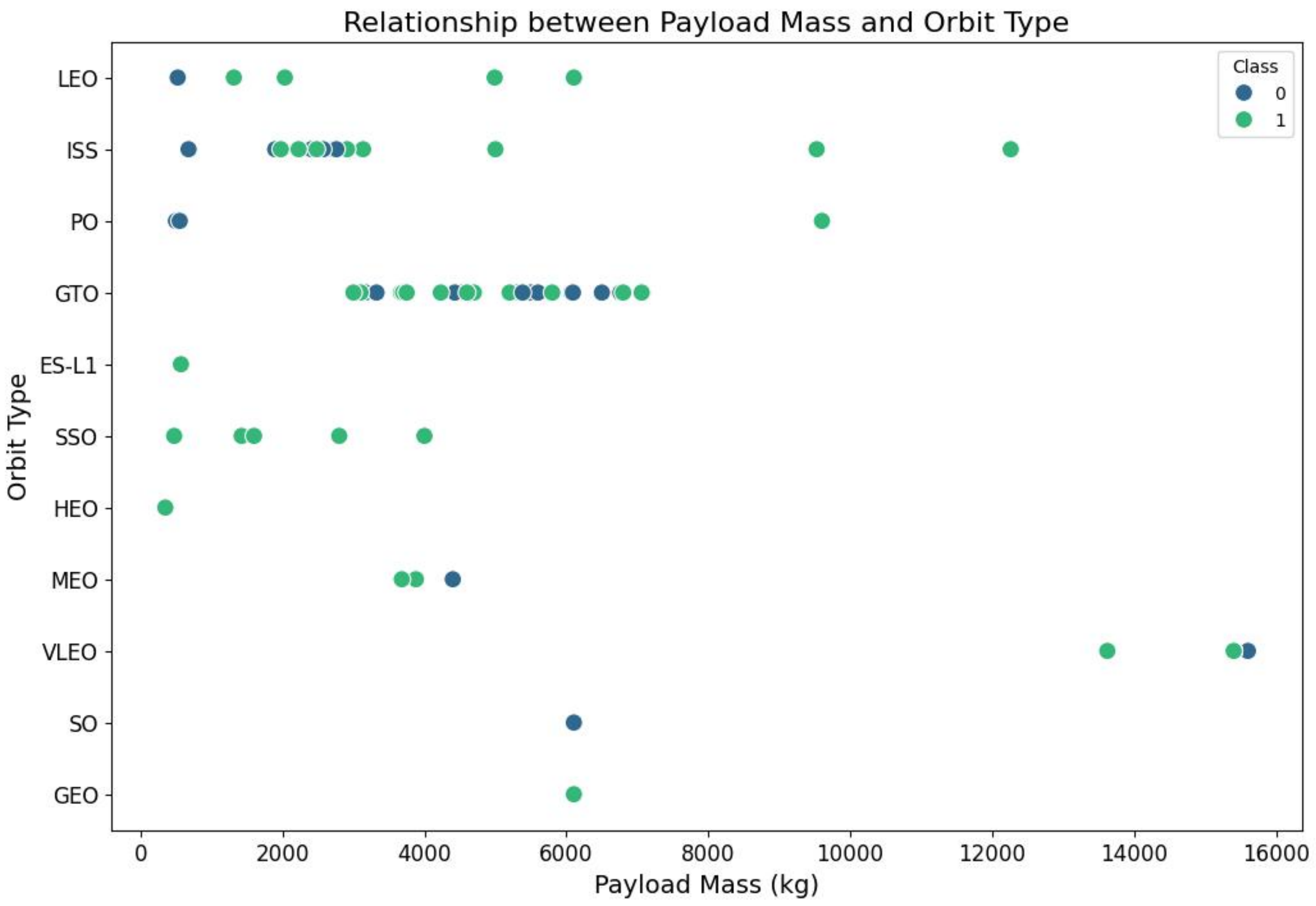
Orbits ES-L1, GEO, HEO and SSO appear to have higher success rates.

EDA with visualization results



Success in the LEO orbit appears to be related to flight number.
For SO orbit that doesn't seem to be the case.

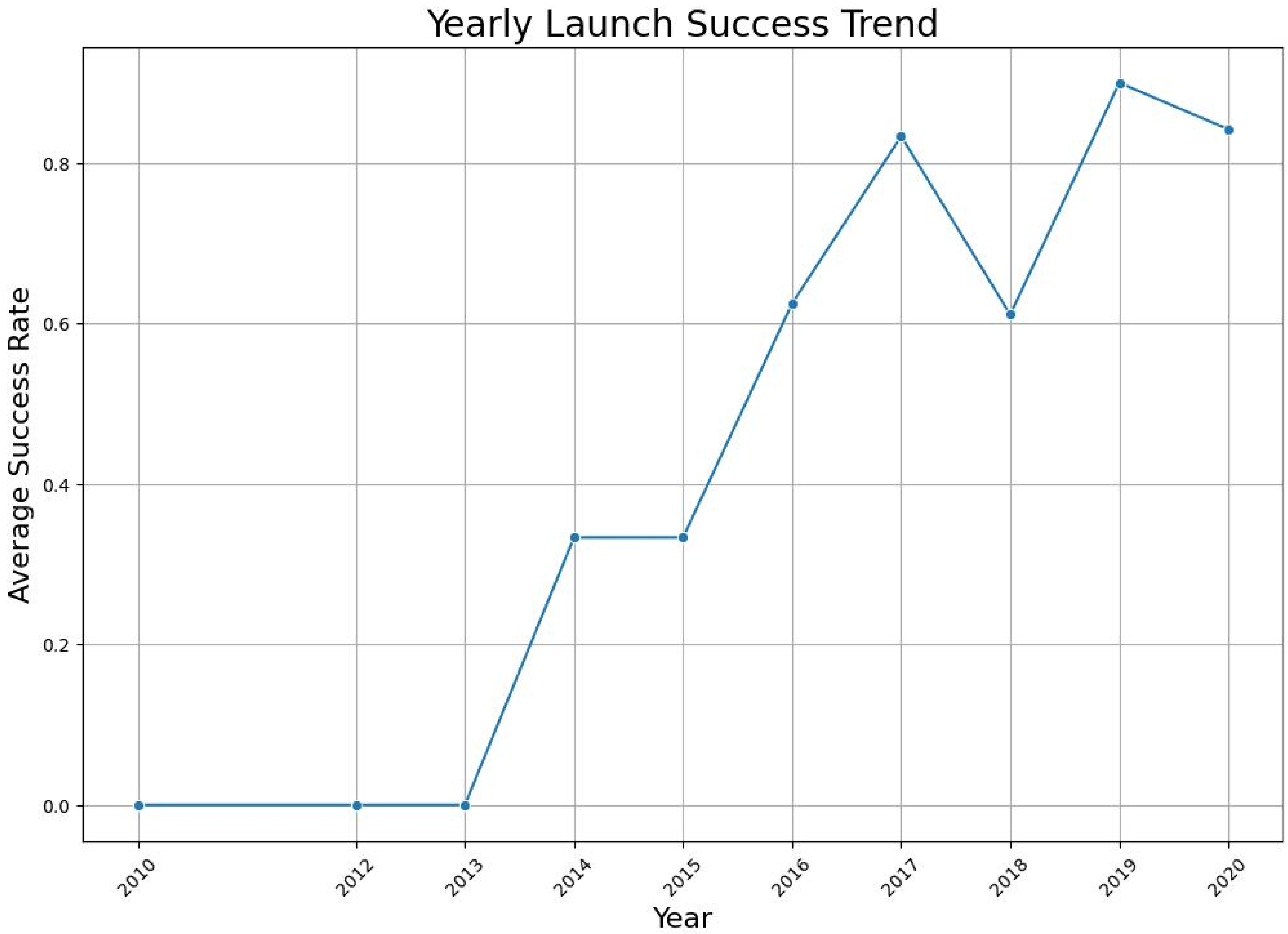
EDA with visualization results



For Polar,LEO and ISS the positive landing rate is greater with heavy payloads.

EDA with visualization results

The success rate from 2013 to 2010 kept increasing.



EDA with SQL results

- The unique launch sites are displayed.
- We found 5 records where the launch sites begin with the string 'CCA'

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

EDA with SQL results

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

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';
```

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

Total_Payload_Mass
45596

Task 4

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

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS Average_Payload_Mass FROM SPACEXTABLE WHERE "Booster_Version" = 'F9
```

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

Average_Payload_Mass
2928.4

We found the total payload mass carried by boosters launched by NASA (CRS) and the average payload mass carried by booster version F9 v1.1

EDA with SQL results

The date the first succesful landing outcome in ground pad was acheived and the total number of successful and failure mission outcomes were determined.

First_Successful_Landing_Date
2018-07-22

Outcome	Count
Failure	1
Success	100

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

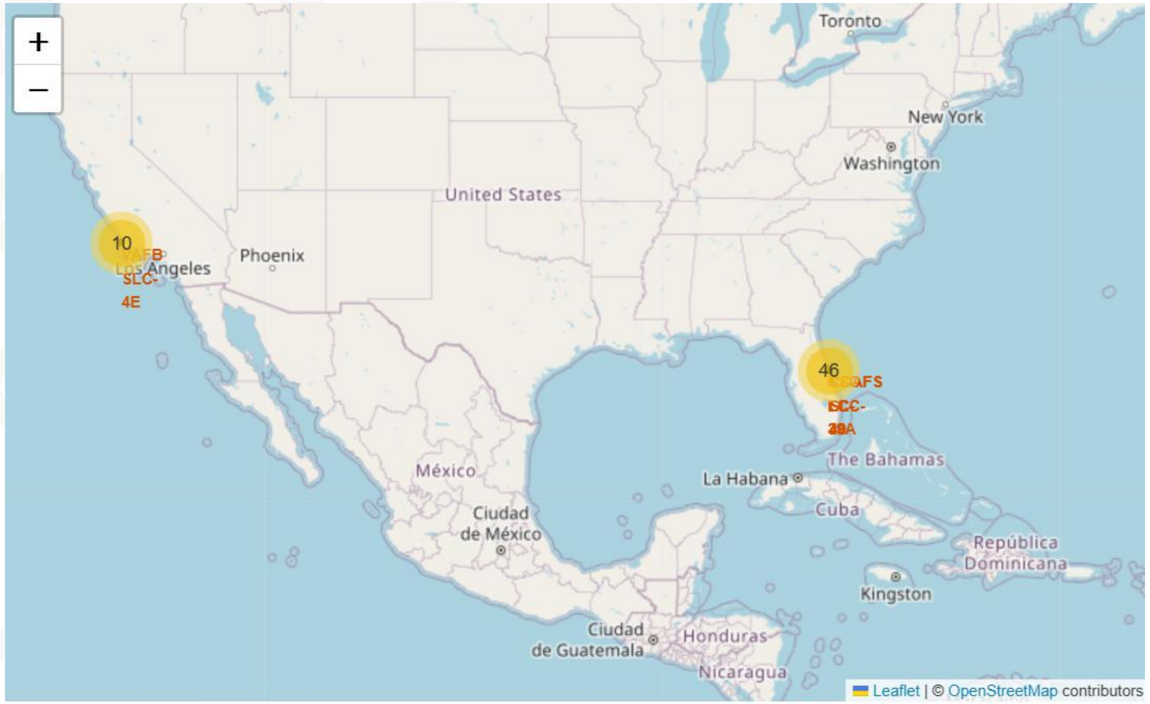
```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)
```

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

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

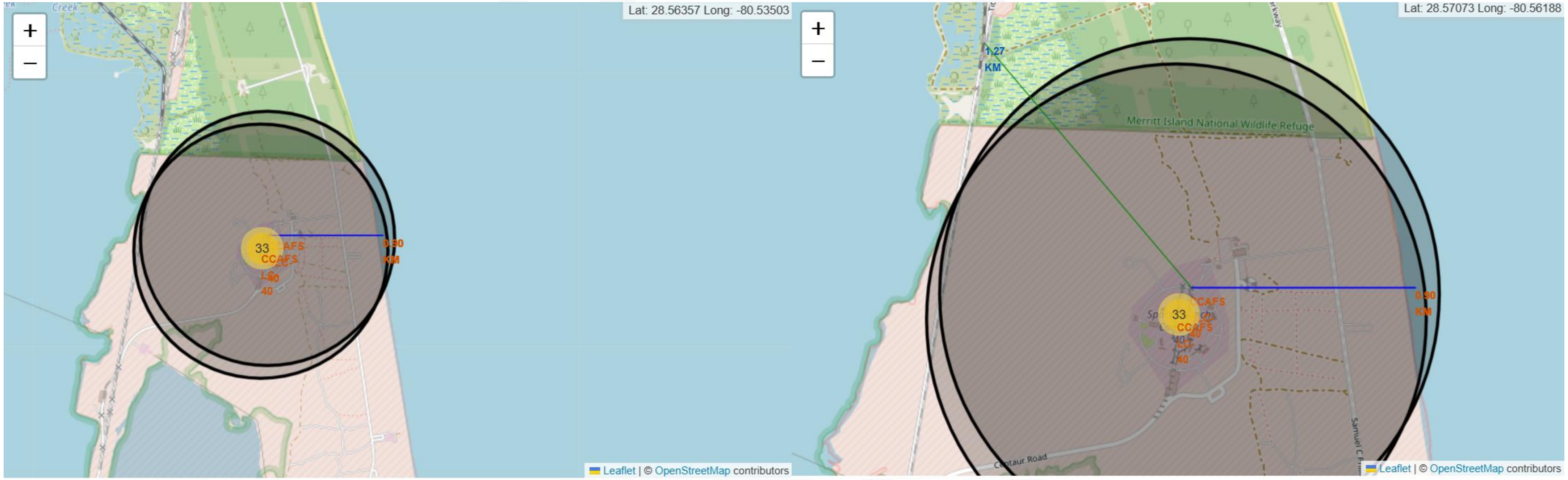
Interactive map with Folium results

We marked all launch sites on a map as well as success/failed launches for each site.



Interactive map with Folium results

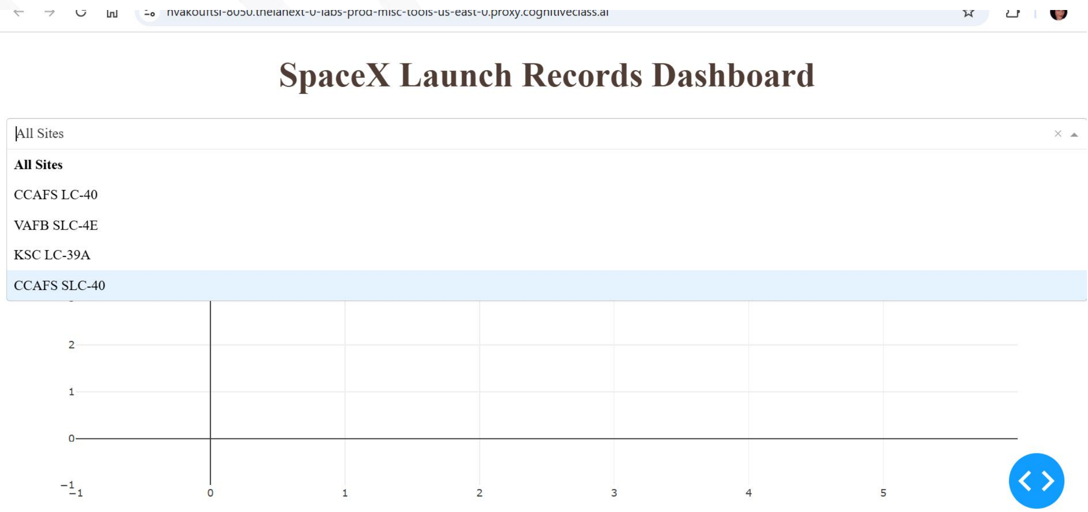
The distance for each launch site to its proximities was determined.



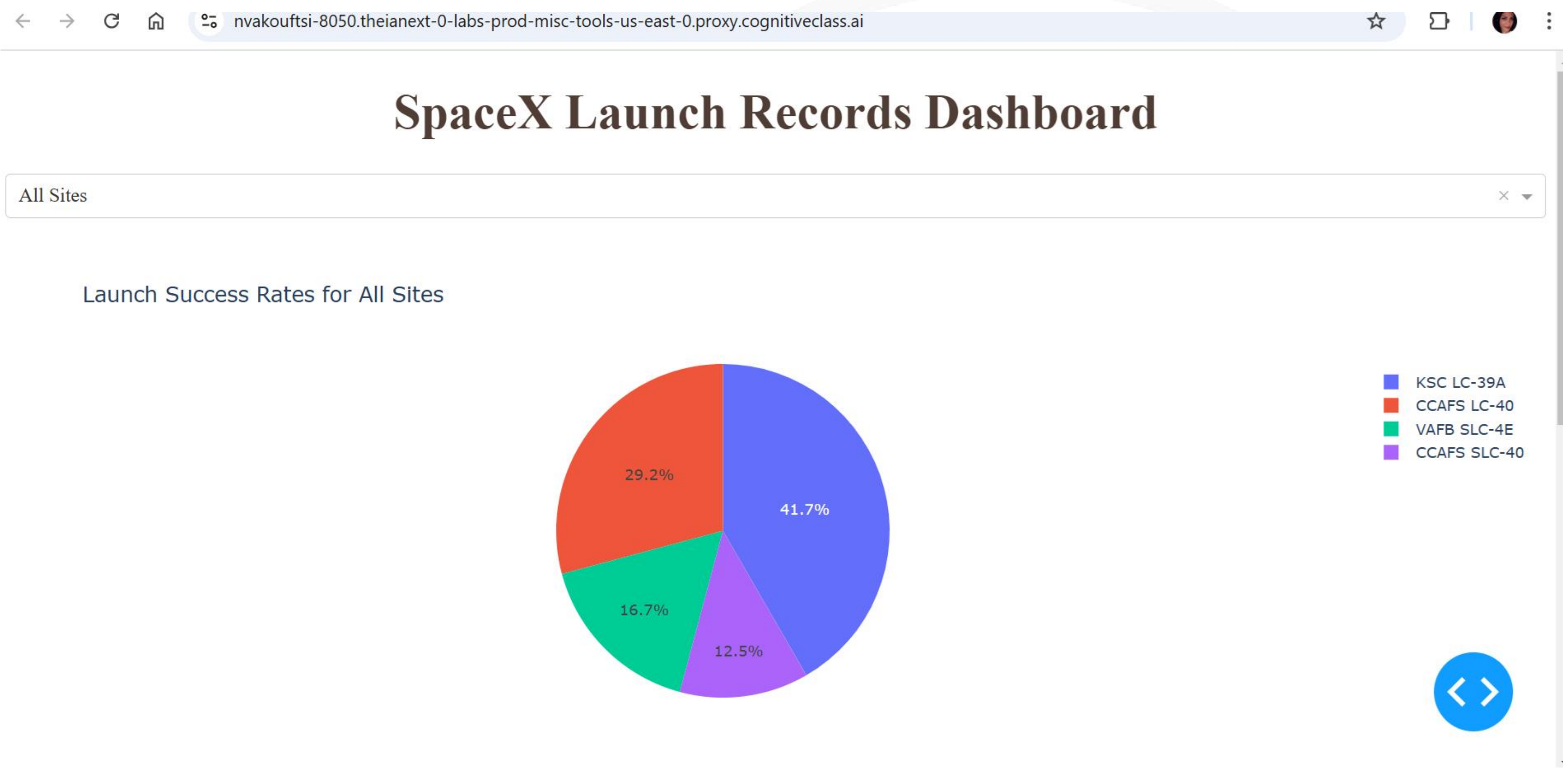
Launch sites appear to be close to their proximities.

Plotly Dash dashboard results

- We read the SpaceX launch data into a Pandas DataFrame to analyze payload, launch site, and success rates.
- We developed a dashboard which included a dropdown menu for selecting launch sites, a range slider to filter payload mass and Incorporated interactive visualizations

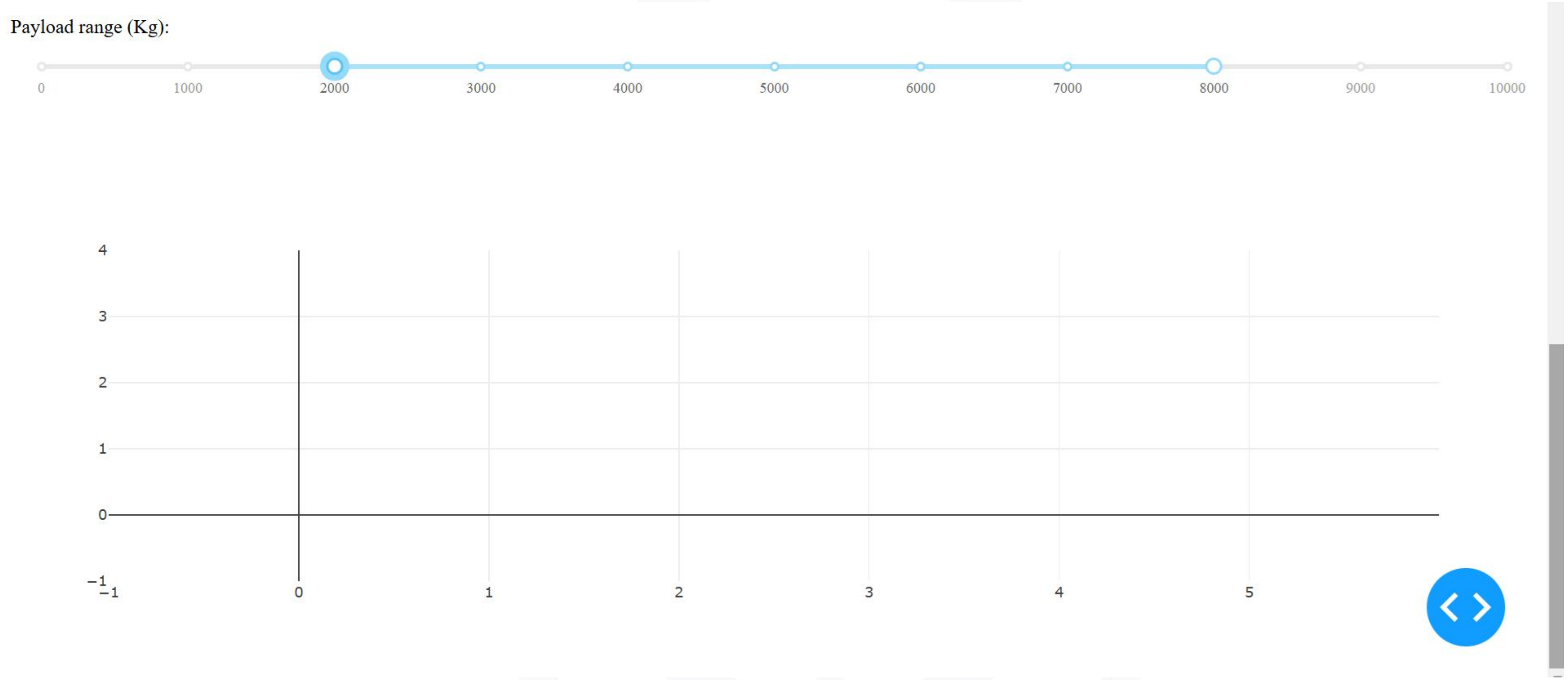


Plotly Dash dashboard results



KSC LC-39A had the most successful launches

Plotly Dash dashboard results



Plotly Dash dashboard results

The FT booster appears to have the highest launch success rates

Predictive analysis (classification) results

- The accuracy of tuned models was compared.
- We plotted confusion matrices to analyze classification results and identified key insights.

```
X_train, X_test, Y_train, Y_test
```

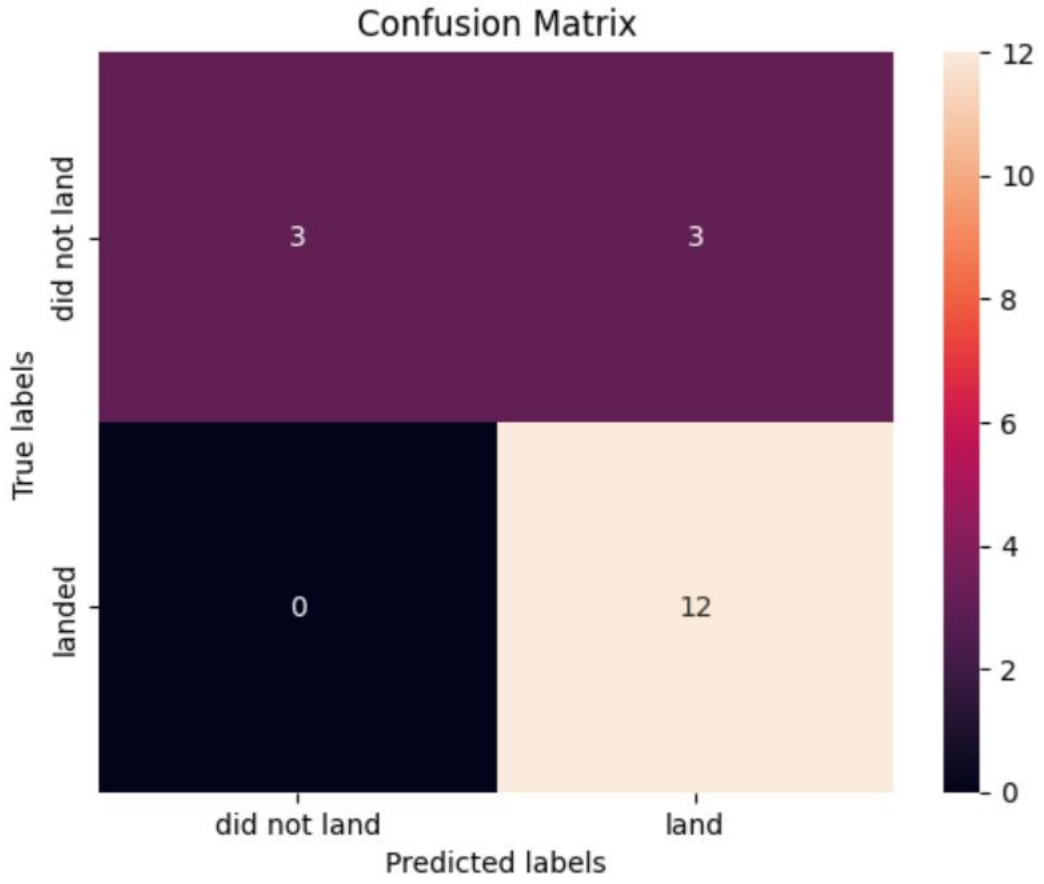
```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
```

we can see we only have 18 test samples.

```
Y_test.shape
```

```
(18,)
```

Predictive analysis (classification) results



Accuracy of each model used:

- Logistic Regression **0.83**
- Support Vector Machine (SVM) **0.83**
- Decision Tree **0.61**
- K-Nearest Neighbors (KNN) **0.83**

Decision Tree had lower accuracy. For the rest of the models the results are practically the same.

Conclusion

- Data exploration and analysis revealed key factors influencing the success of Falcon 9 rocket landings, such as payload mass, launch site, and booster version.
- Machine learning models were developed to predict landing success with similar accuracy levels.
- These insights can guide future rocket launch planning and optimization

Discussion

By incorporating additional data points, such as weather conditions and manufacturing details we can achieve deeper analysis.