IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abdukhakimov Asatilla
13.10.2024

Outline

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

Executive Summary

This report presents a machine learning approach to predicting the success of SpaceX Falcon 9 first stage landings.

Summary of Methodologies:

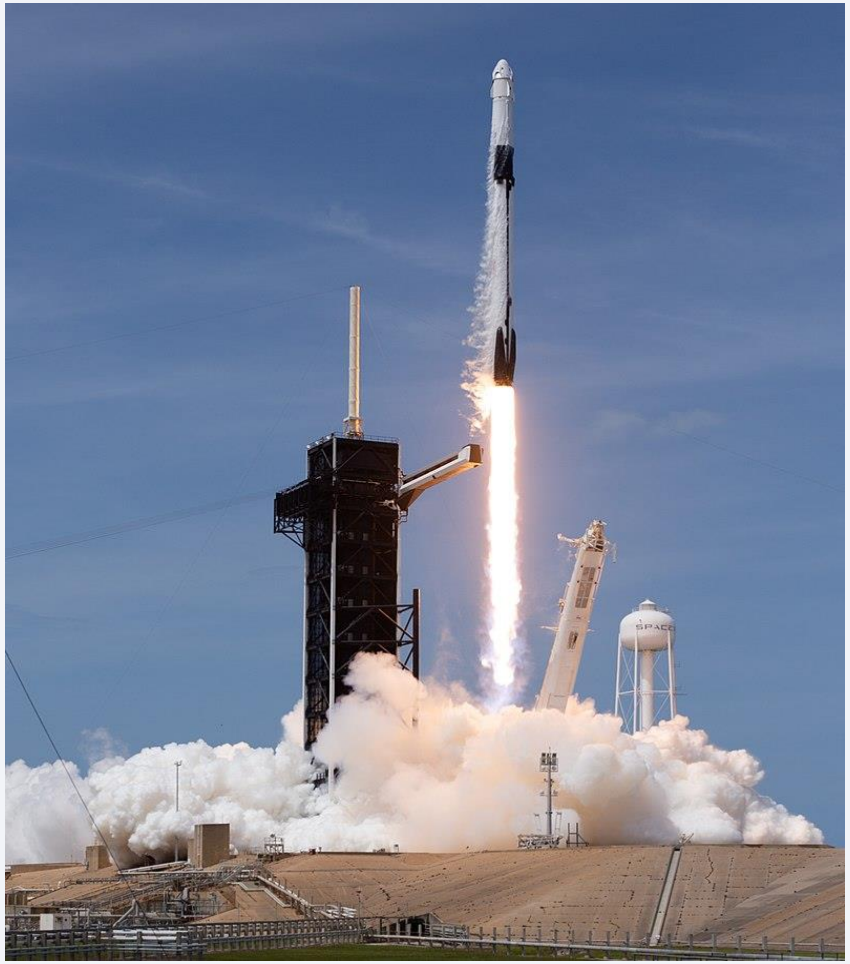
- Data Collection and Wrangling
- Exploratory Data Analysis
- Interactive Visual Analytics
- Machine Learning Prediction

Summary of Results:

The selected model achieved 83% accuracy, offering a powerful tool for understanding and potentially improving the success rate of Falcon 9 first stage landings. Continuous refinement of the model with new mission data will further enhance its accuracy and utility for future SpaceX missions.

Introduction

This project aims to predict the successful landing of **SpaceX's Falcon 9 first stage rocket**. SpaceX offers significantly cheaper rocket launches (at $62 million) compared to competitors (upwards of $165 million) primarily due to their ability to reuse the first stage of the Falcon 9. By accurately predicting the landing success of the first stage, we can estimate the overall cost of a launch. This information can be leveraged by other companies competing with SpaceX for rocket launch contracts.



Introduction

- This project will develop a machine learning model to predict whether SpaceX will reuse the first stage of a Falcon 9 rocket for a given launch. Using publicly available data about past SpaceX missions, the model will learn to identify patterns and relationships between various factors and the decision to reuse or expend the first stage.

- This trained model can then be used to predict the likelihood of first stage reuse for future missions, offering valuable insights into SpaceX's launch strategies and potential cost savings.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from SpaceX API and a Wikipedia website
- Perform data wrangling
  - Perform Data Analysis to Find Some Patterns & Determine Training Labels
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Find the best performing model by comparing Logistic Regression, Support Vector Machines, and K-Nearest Neighbors using GridSearchCV to fine-tune their parameters.

Data Collection

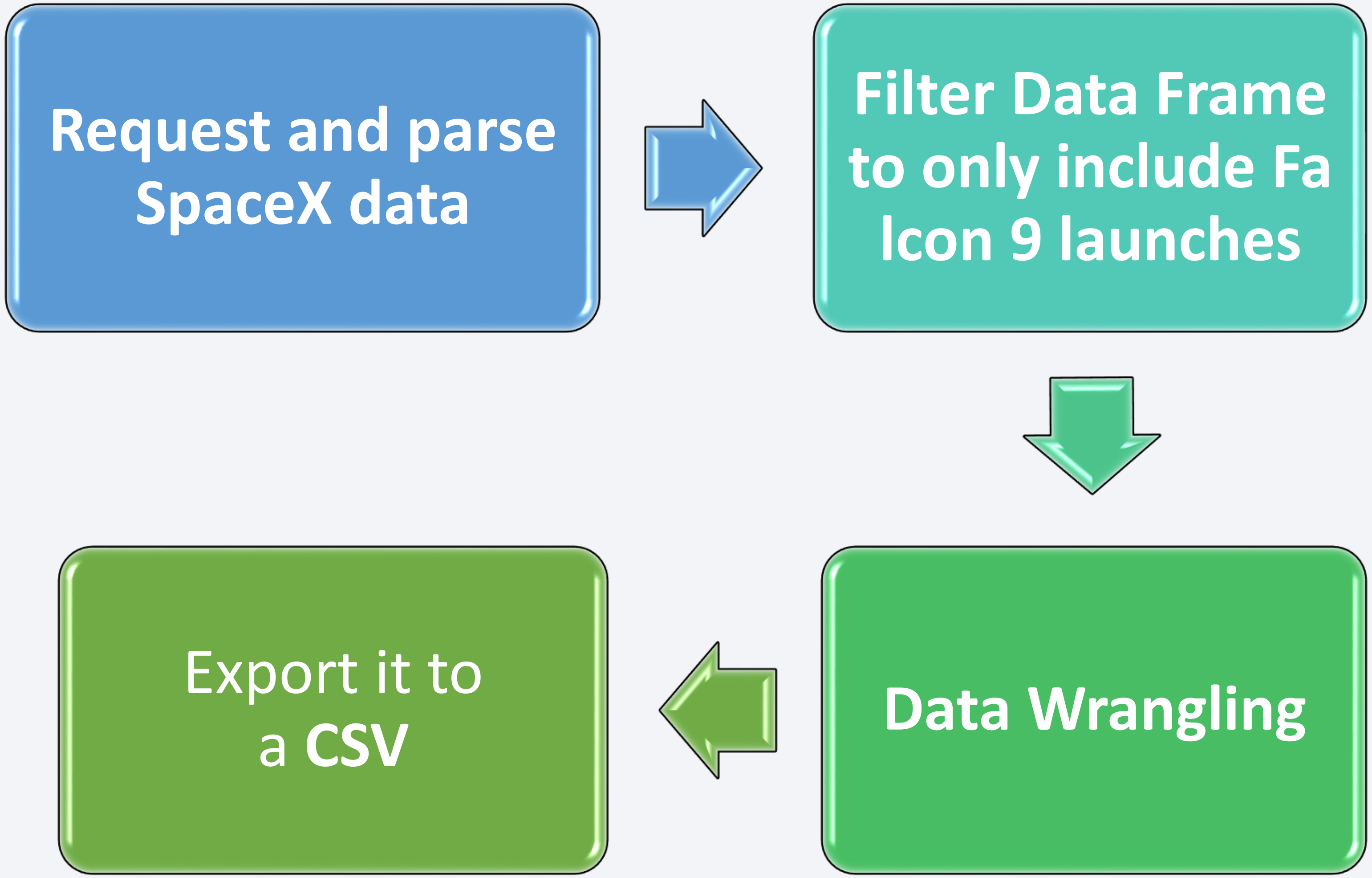
Data was collected from:

- SpaceX API
- Wikipedia website

https://api.spacexdata.com/v4/launches/past

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection - SpaceX API



Details can be found on GitHub link here:

https://github.com/asat94/Applied-Data-Science-for-SpaceX/blob/main/1.%20Data%20Collection%20with%20API%20-%20SpaceX.ipynb

Data Collection - Scraping

Request Falcon9 Launch Wiki page

Extract all variable names from HTML table header

Create a data frame by parsing the laun ch HTML tables

Export it to a **CSV**

Details can be found on GitHub link here:

https://github.com/asat94/Applied-Data-Science-for-SpaceX/blob/main/2.%20Data%20Collection%20with%20Web%20Scraping%20-%20%20-%20SpaceX.ipynb

Data Wrangling

Identify Missing Values

- Replace Missing values for Payload Mass using its mean value

Calculate:

- Number of launches on each site
- Number and occurrence of each orbit
- The number and occurrence of mission outcome of the orbits

Create:

- Classification variable , in which,
  0 means failure
  and
  1 means successful

Link: https://github.com/asat94/Applied-Data-Science-for-SpaceX/blob/main/3.%20Data%20Wrangling%20-%20SpaceX.ipynb

EDA with Data Visualization

Exploratory Data Analysis using Matplotlib and Seaborn Libraries:

- Flight Number vs. Payload Mass
- Flight Number vs. Launch Site
- Payload Mass and Launch Site
- Success rate of each orbit type
- Flight Number vs. Orbit type
- Payload Mass and Orbit type
- Landing success yearly trend

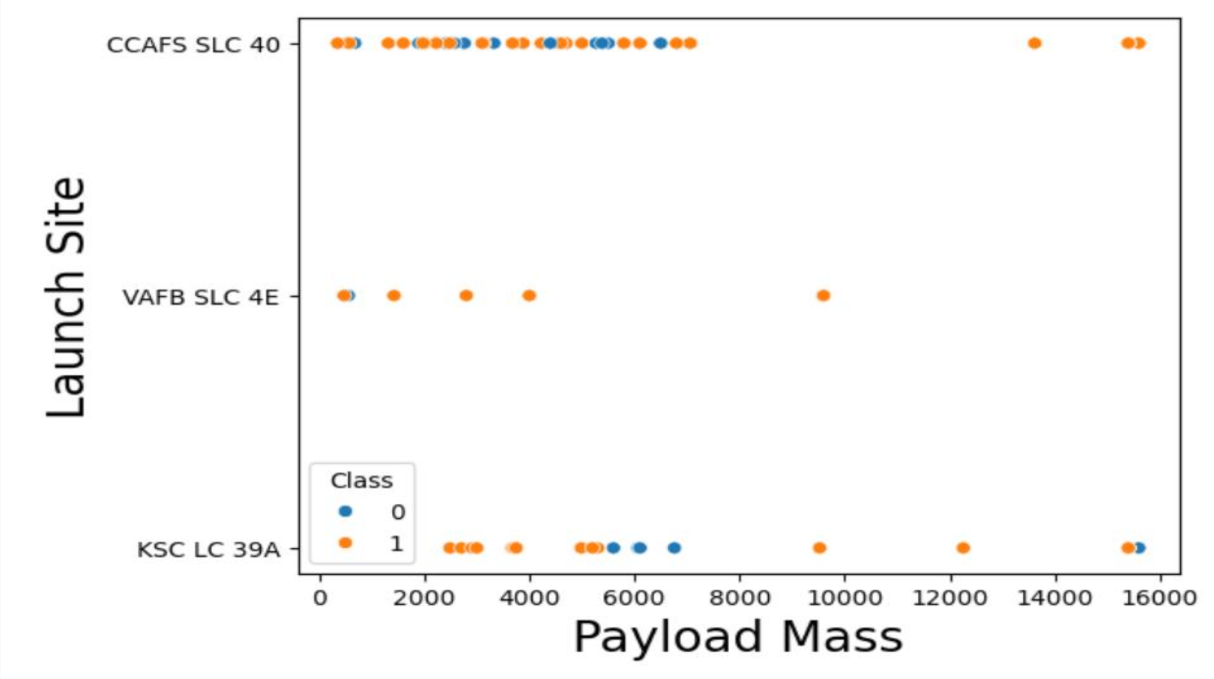
https://github.com/asat94/Applied-Data-Science-for-SpaceX/blob/main/5.%20Exploratory%20Data%20Analysis%20for%20Data%20Visualization%20-%20SpaceX.ipynb

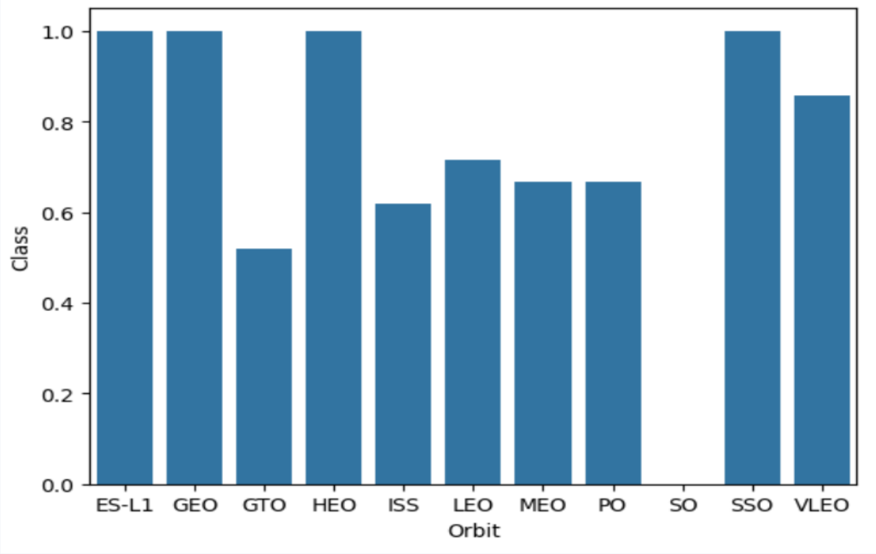
Preparing Data Feature Engineering

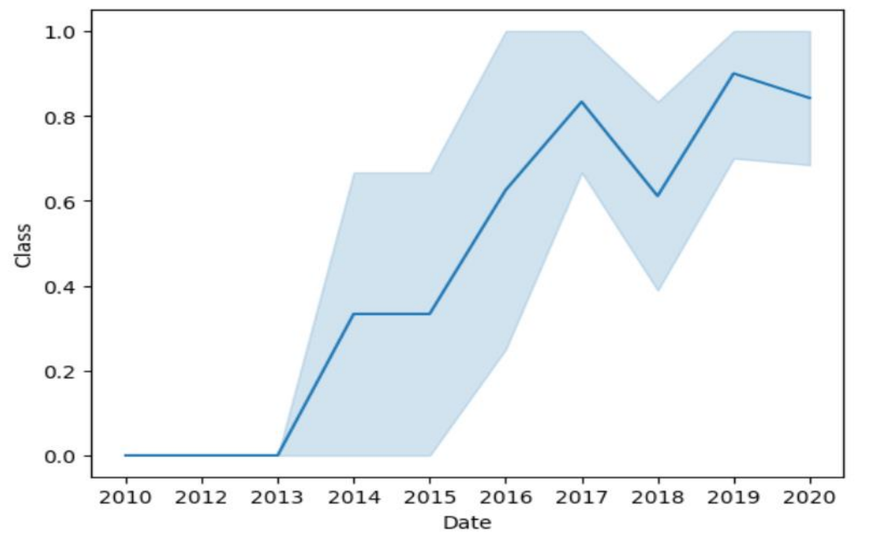
- Select features that are used in prediction

EDA with Data Visualization

Visualize how Flight Number, Payload Mass, Launch Site, Orbit Type variables would affect the launch outcome using scatter, bar and line graphs







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 between 4000 and 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.
- 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.

https://github.com/asat94/Applied-Data-Science-for-SpaceX/blob/main/4.%20Exploratory%20Data%20Analysis%20using%20SQL%20-%20SpaceX.ipynb

Build an Interactive Map with Folium

The success rate may depend on the location and proximities of a launch site

- 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

https://github.com/asat94/Applied-Data-Science-for-SpaceX/blob/main/6.%20Interactive%20Visual%20Analytics%20with%20Folium%20-%20SpaceX.ipynb

Build a Dashboard with Plotly Dash

To Identify the Relationship of Success Rate between Launch Site and Payload, an Interactive Dashboard is Built with Plotly Dash that includes:

- A dropdown list to enable Launch Site selection
- A pie chart to show the total successful launches count for all sites
- A slider to select payload range
- A scatter chart to show the correlation between payload and launch success

https://github.com/asat94/Applied-Data-Science-for-SpaceX/blob/main/7.%20Interactive%20Dashboard%20with%20Ploty%20Dash%20-%20SpaceX.py

Predictive Analysis (Classification)

Standardization

- preprocessing. StandardScaler()
- fit_transform(X)

Split Data

- Train set 80%
- Test set 20%

Model Evaluation

- Logistic Regression
- SVM
- Decision Tree
- KNN

Find the Best method

- Calculate Accuracy Score for all models

https://github.com/asat94/Applied-Data-Science-for-SpaceX/blob/main/8.%20Machine%20Learning%20Prediction%20-%20SpaceX.ipynb

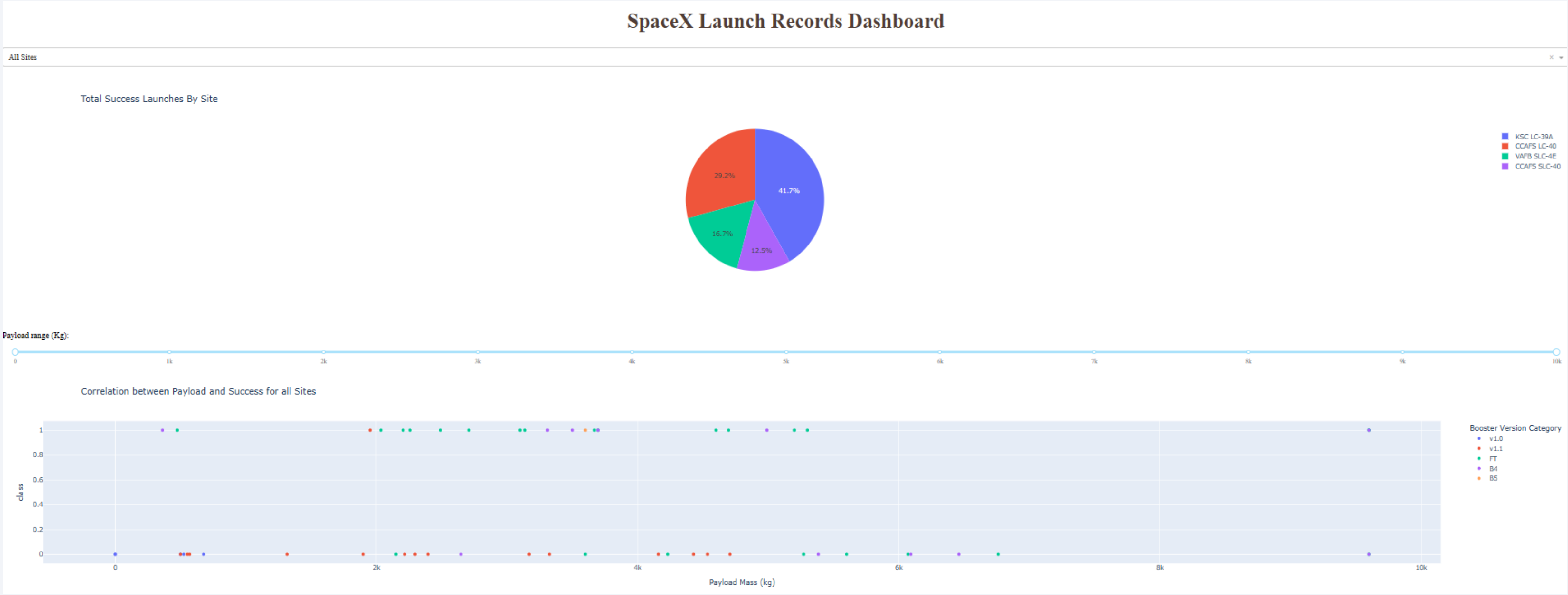
Results

Exploratory data analysis results

- In the LEO orbit, the success seems to be related to the number of flights
- With heavy payloads the successful landing rate are more for Polar, LEO and ISS
- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(> 10t). CCAFC-SLC shows better results than KSC-LC
- The success rate since 2013 kept increasing till 2020

Results

- **Interactive analytics demo in screenshots**



Results

- **Predictive analysis results**

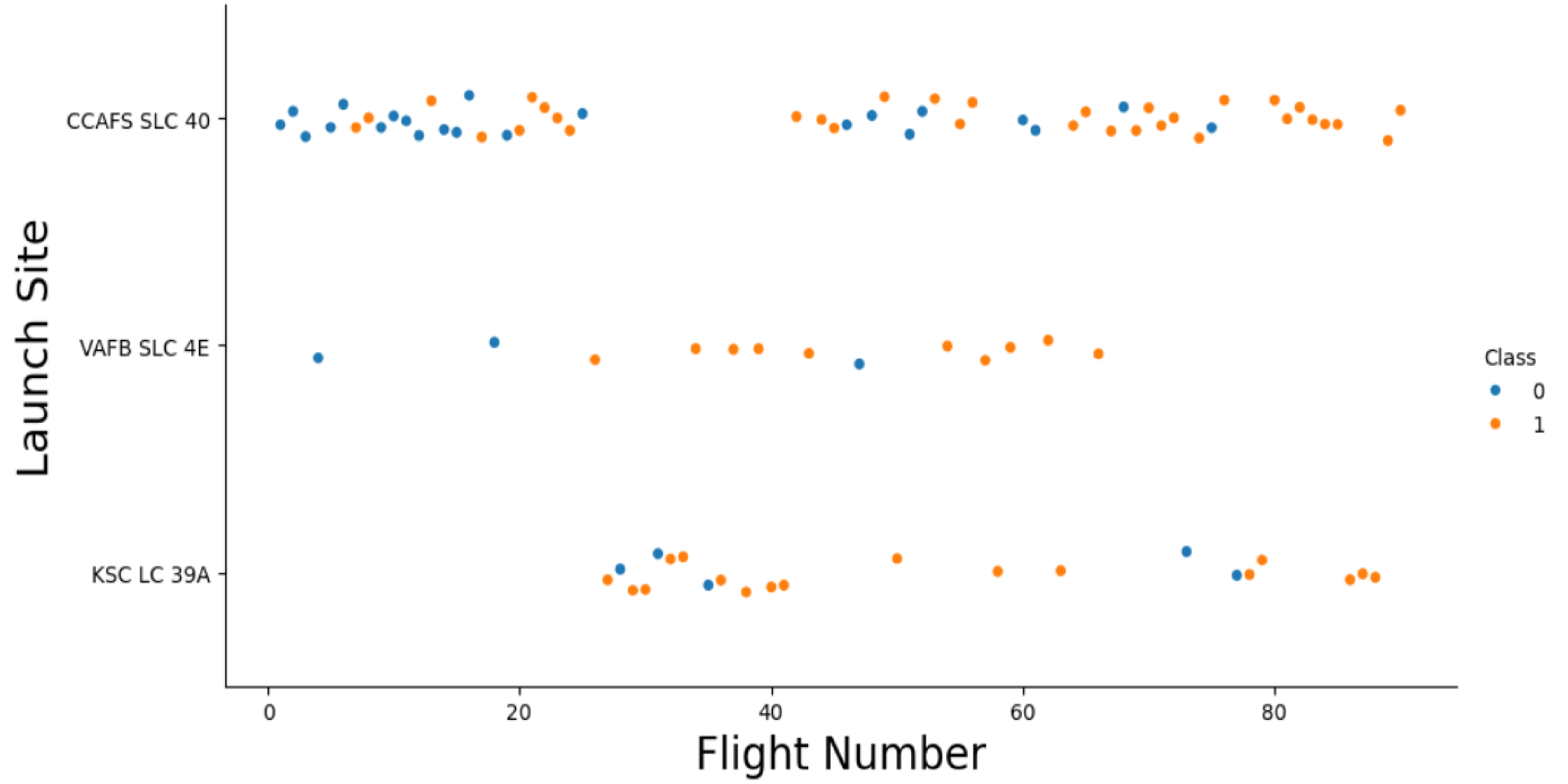
The below table shows the calculated Accuracy Score for Logistic Regression, SVM, Decision Tree and KNN models. The result shows the same accuracy rate for all methods.

	Logistic Regression	SVM	Decision Tree	KNN
Accuracy on Test Data	83.33%	83.33%	83.33%	83.33%

Section 2

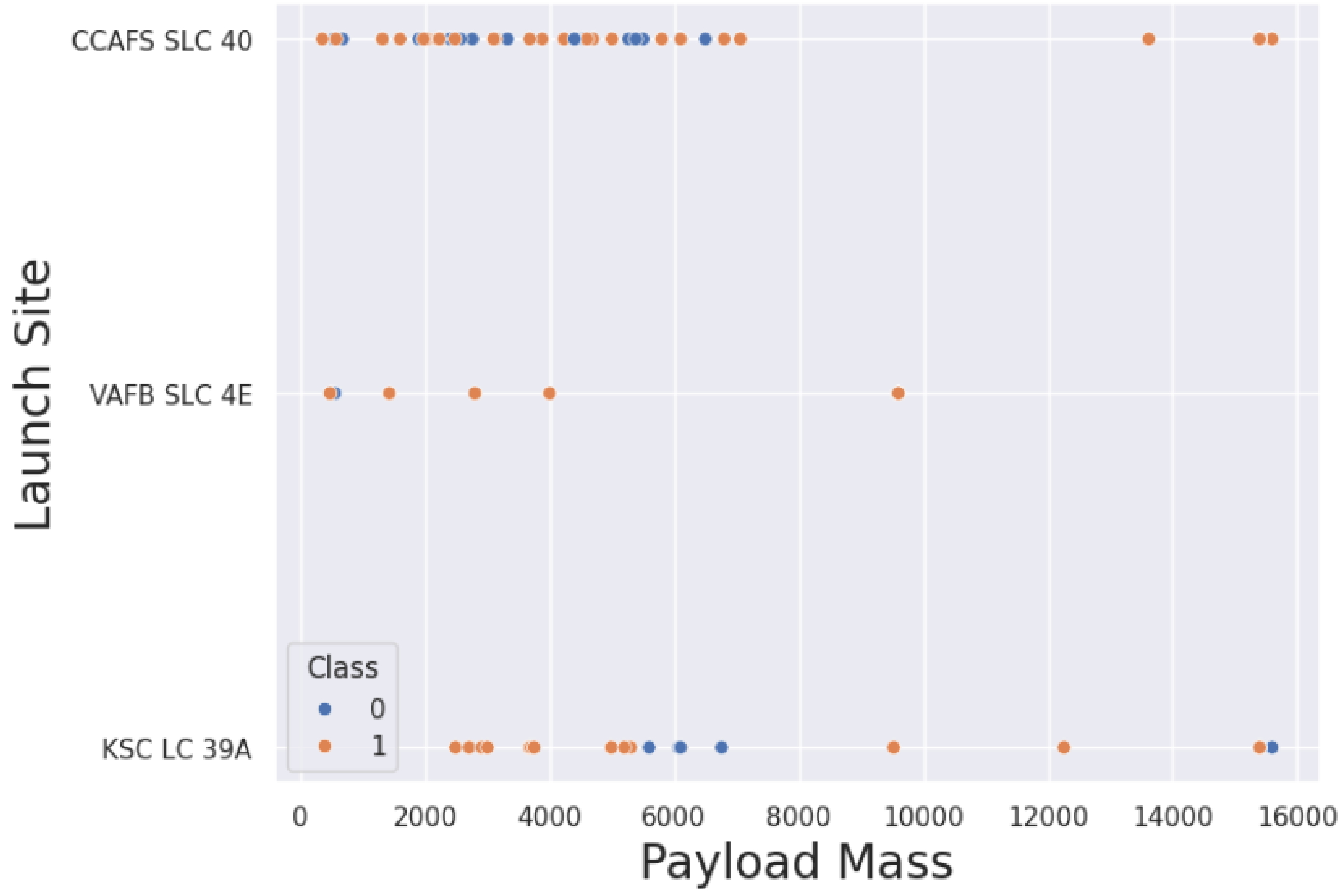
Insights drawn from EDA

Flight Number vs. Launch Site



CCAFS SLC 40 launch site shows most of failure rates in the beginning of the project. However, after flight number No. 80, we may observe that first stage landind is more likely to land successfully.

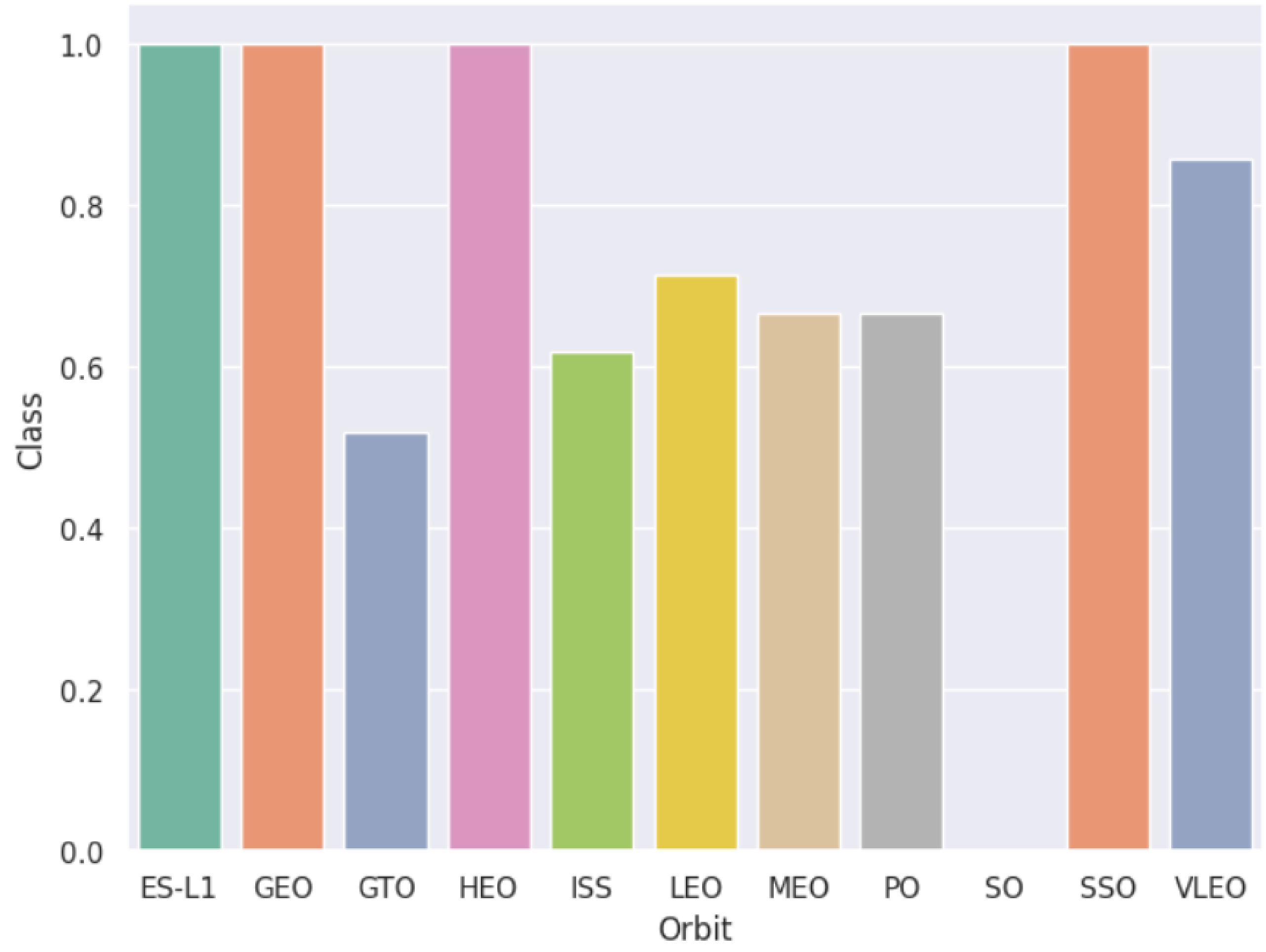
Payload vs. Launch Site



Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find

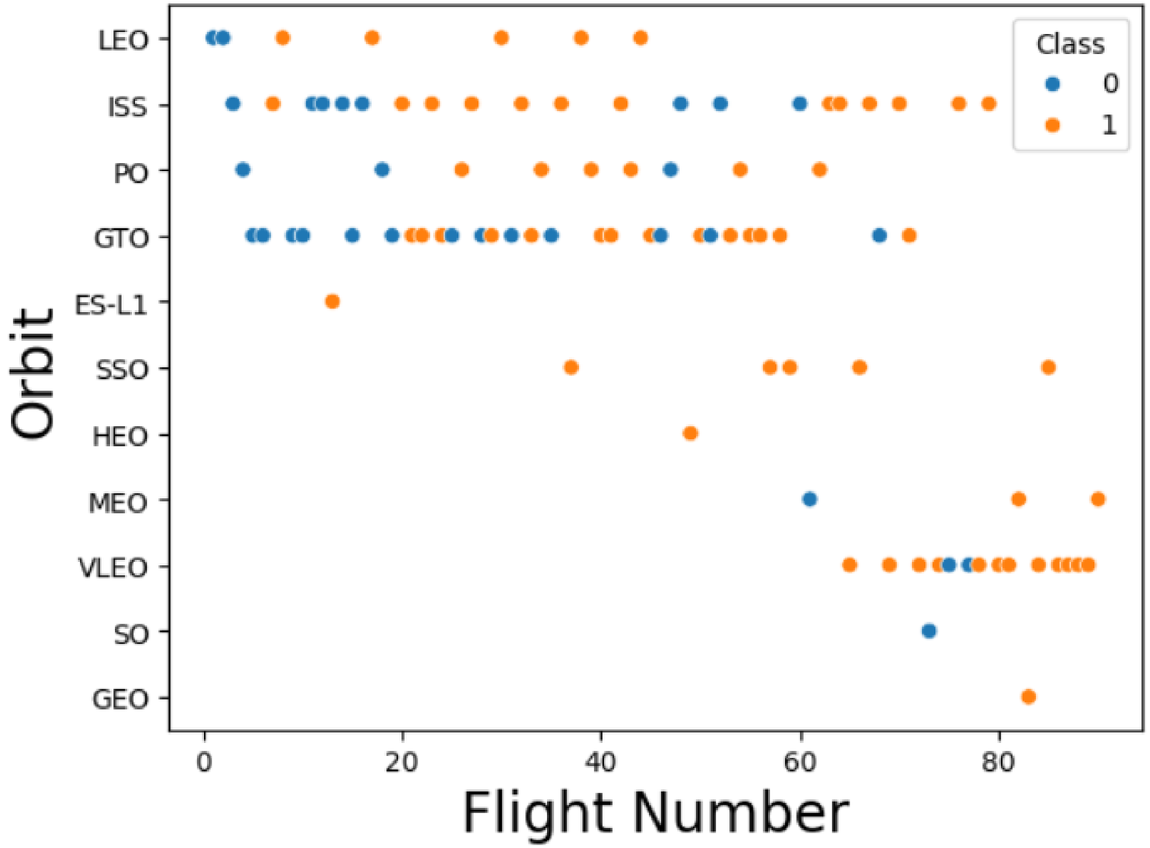
for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



ES-L1, GEO, HEO, SSO orbits have the highest (100%) success rate, followed by VLEO. However, SO orbit has 0% successful landing.

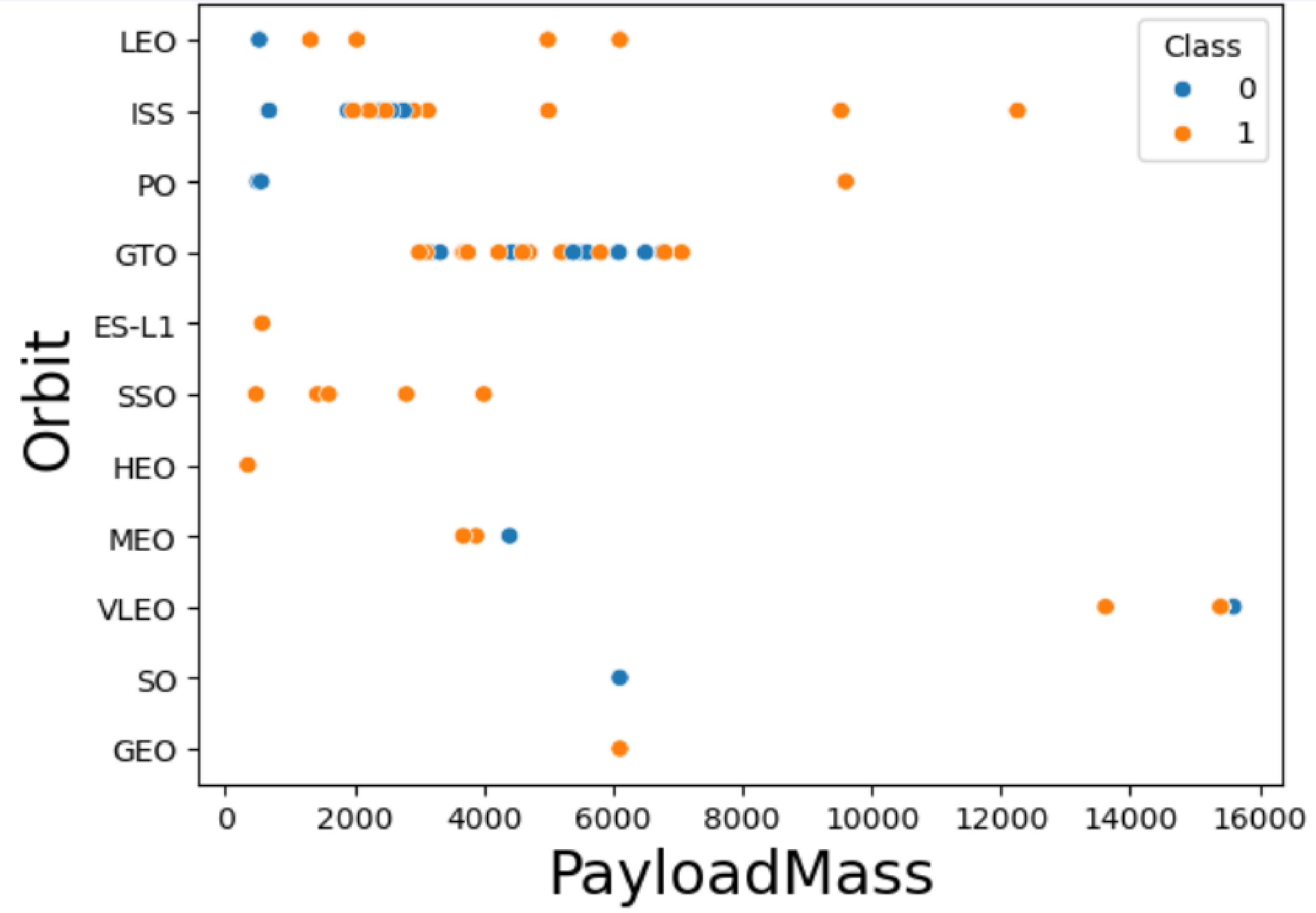
Flight Number vs. Orbit Type



You can observe that in the LEO orbit, success seems to be related to the number of flights.

Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

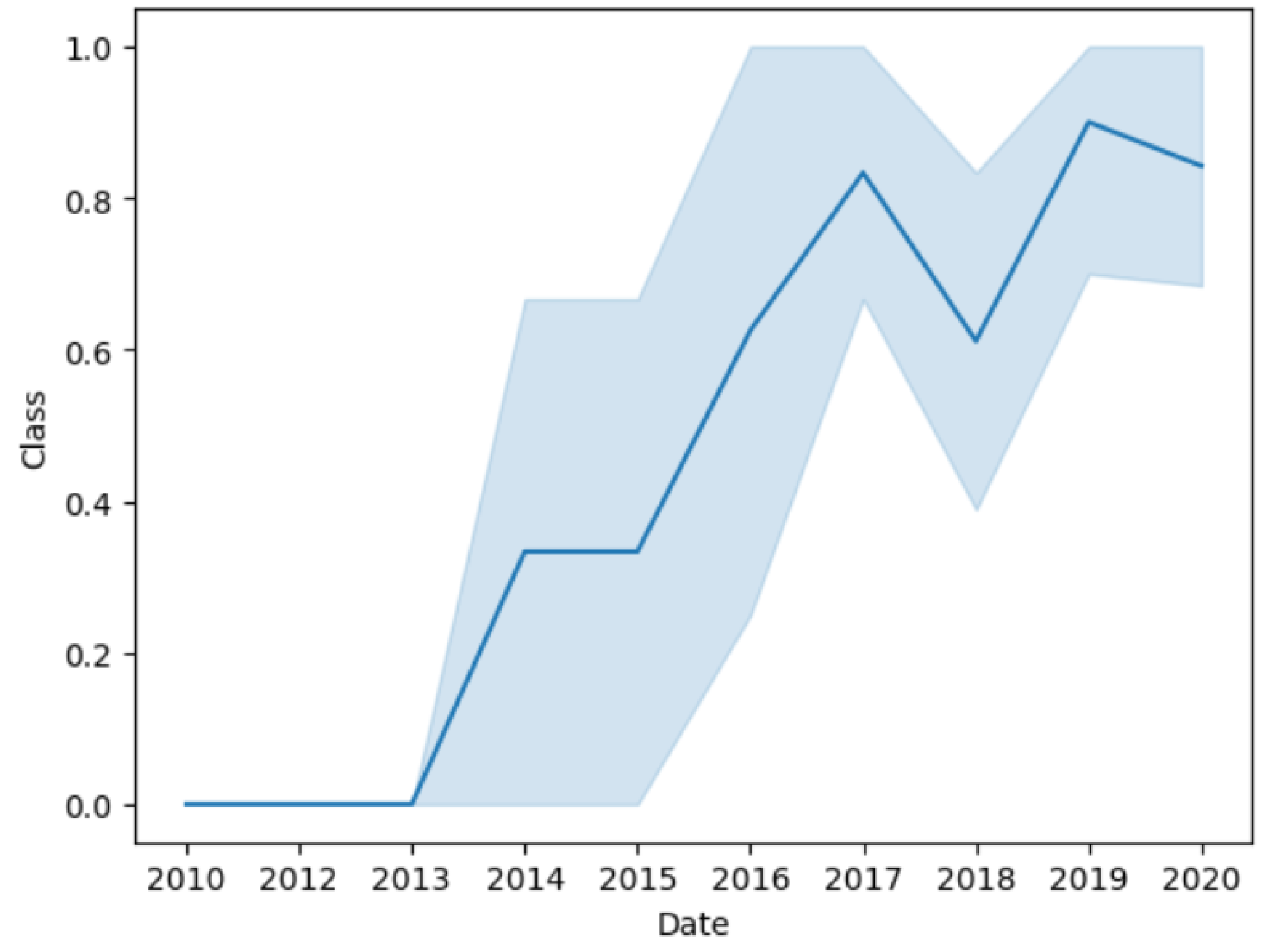
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



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

All Launch Site Names

```
[11]: %sql select distinct("Launch_Site") from SPACEXTABLE

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

[11]:

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

4 Unique Launch Sites have been found with Distinct Statement using SQL

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5
```

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

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

All 5 records where launch sites begin with `CCA` have failure outcomes or no attempt at all

Total Payload Mass

Task 3

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

```
[13]: %sql select sum("PAYLOAD_MASS__KG_") from SPACEXTABLE where Customer='NASA (CRS)'
```

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

[13]:

sum("PAYLOAD_MASS__KG_")
45596

The total payload mass carried by boosters launched by NASA is 45596KG

Average Payload Mass by F9 v1.1

Task 4

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

[14]:
```
%sql select avg("PAYLOAD_MASS__KG_") from SPACEXTABLE where Booster_Version ='F9 v1.1'
```

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

[14]:

avg("PAYLOAD_MASS__KG_")
2928.4

The average payload mass carried by booster version F9 v1.1 is 2928.4KG

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

[15]:
```
%sql SELECT MIN(DATE) \
FROM SPACEXTBL \
WHERE Landing_Outcome = 'Success (ground pad)'
```

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

[15]:

MIN(DATE)
2015-12-22

2015-12-22 was the date of the first successful landing outcome on ground pad

Successful Drone Ship Landing with Payload between 4000 and 6000

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

[16]:
```
%sql select Booster_Version  from SPACEXTABLE \
WHERE Landing_Outcome = 'Success (drone ship)' and "PAYLOAD_MASS__KG_" between 4000 and 6000
```

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

[16]:

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

Above is the list of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

[17]:
```
%sql select Mission_Outcome, Count(Mission_Outcome) from SPACEXTABLE group by Mission_Outcome
```

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

[17]:

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

100 successful mission outcomes and 1 failure (in flight) mission outcome were observed

Boosters Carried Maximum Payload

```
%sql select Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTABLE \
    where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from SPACEXTABLE)
```

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

Booster_Version	PAYLOAD_MASS__KG_
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

Here is the list of the names of boosters which have carried the maximum payload mass (15600 KG)

2015 Launch Records

```
%sql select substr(Date, 6,2) as Month, Date, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE \
where Landing_Outcome = 'Failure (drone ship)' and substr(Date,0,5)='2015'
```

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

Month	Date	Landing_Outcome	Booster_Version	Launch_Site
01	2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

This is the list of failed landing outcomes in drone ship, their booster versions, and launch site names for the months in 2015

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

```
%sql SELECT "Landing_Outcome", Date, Count("Landing_Outcome") as count FROM SPACEXTABLE \
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' \
group by "Landing_Outcome"  \
order by count DESC
```

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

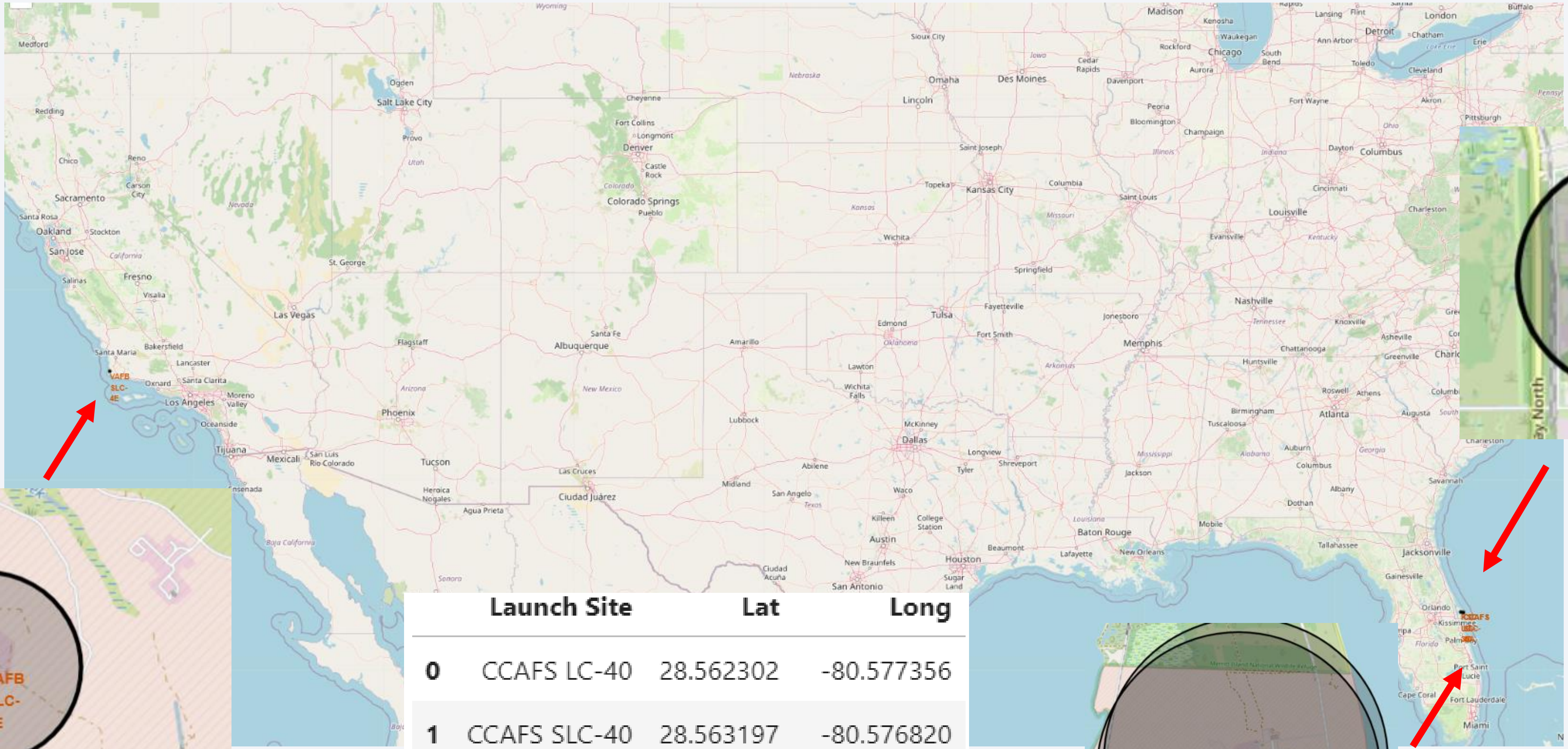
Landing_Outcome	Date	count
No attempt	2012-05-22	10
Success (drone ship)	2016-04-08	5
Failure (drone ship)	2015-01-10	5
Success (ground pad)	2015-12-22	3
Controlled (ocean)	2014-04-18	3
Uncontrolled (ocean)	2013-09-29	2
Failure (parachute)	2010-06-04	2
Precluded (drone ship)	2015-06-28	1

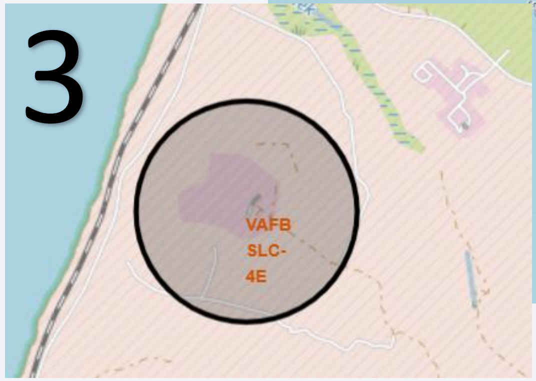
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Section 3

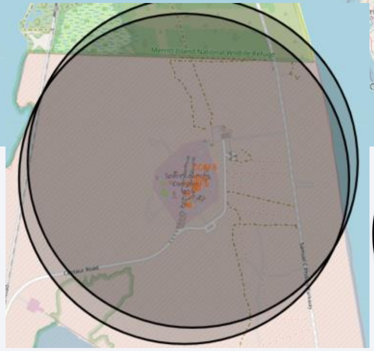
Launch Sites Proximities Analysis

All launch sites on a map

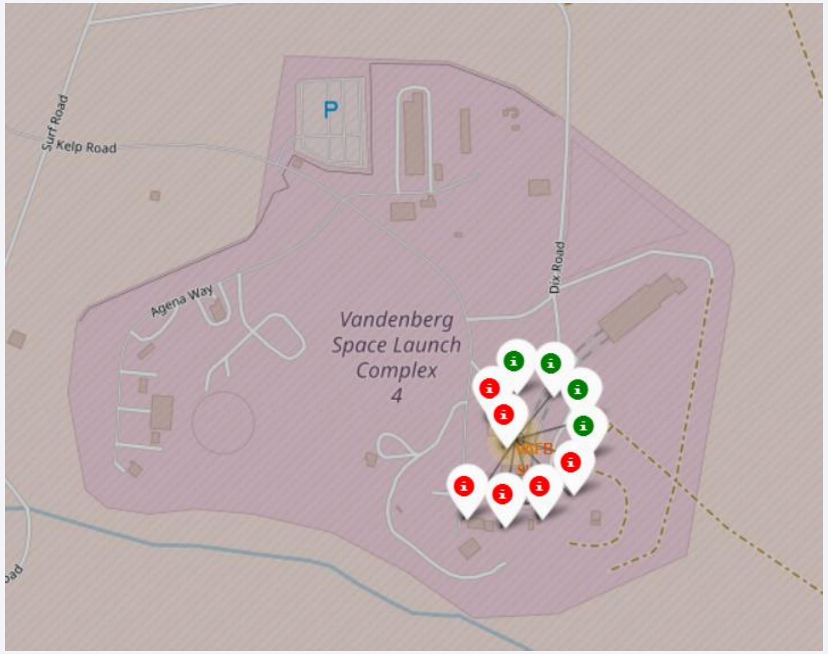




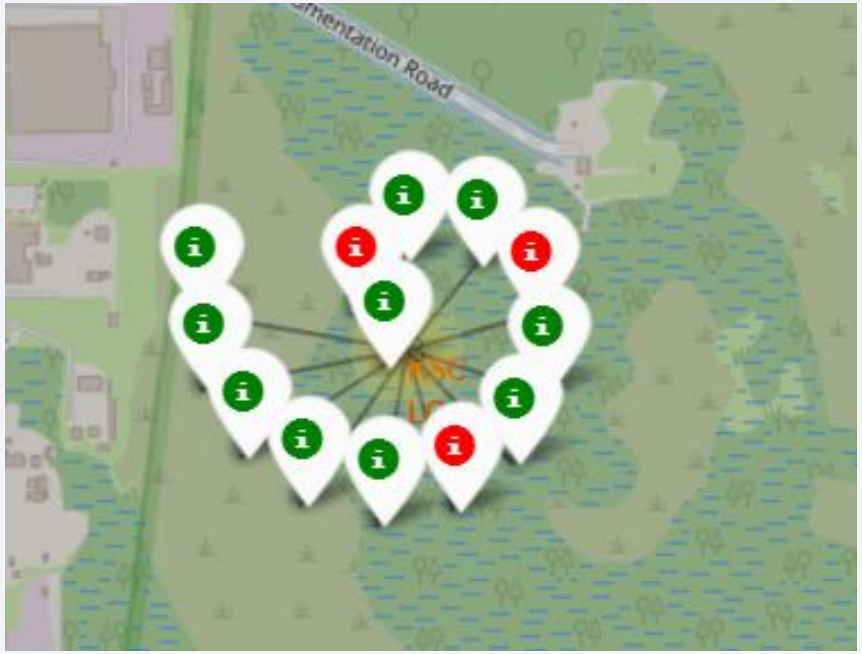
	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745



Success/failed launches for each site on the map



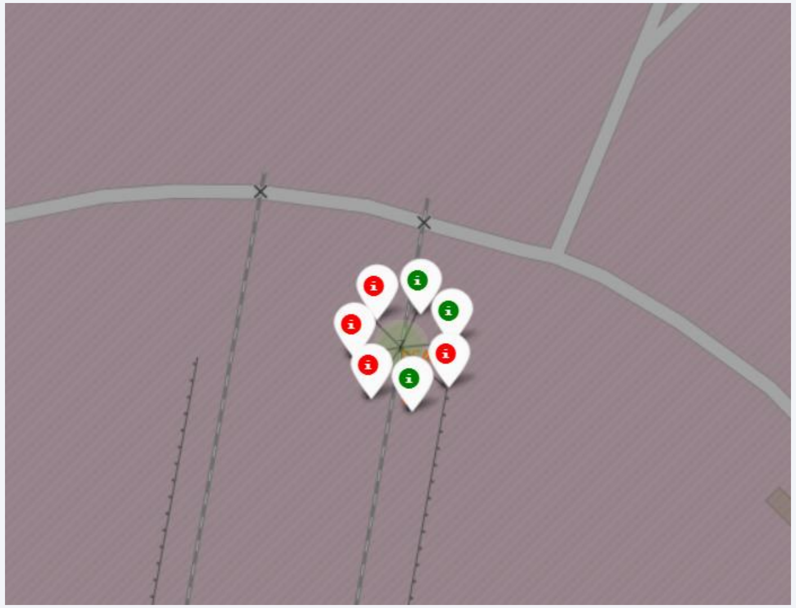
VAFB SLC-4E
Vandenberg Space Launch Complex 4



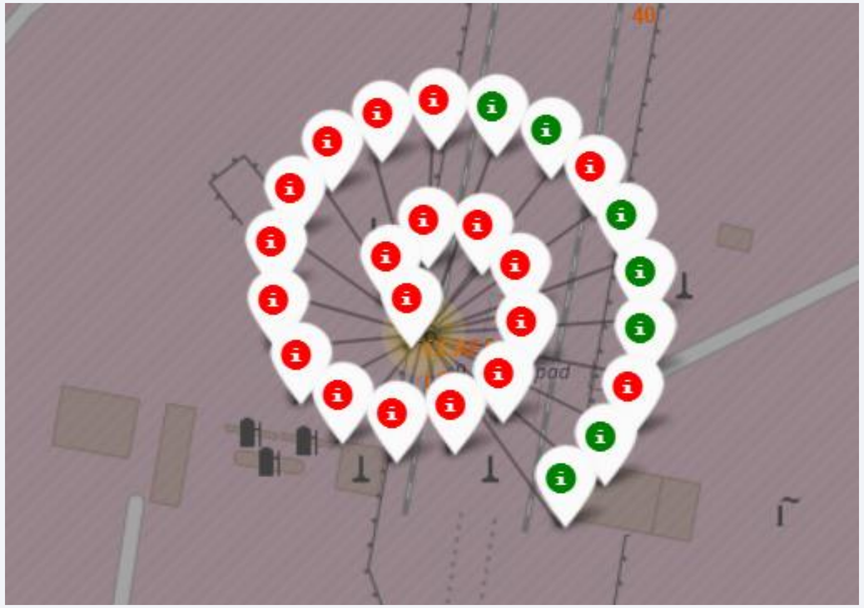
KSC LC-39A
Kennedy Space Center Launch Complex 39A

Class = 1 (Green Color) = Successful Launch
Class = 0 (Red Color) = Failed Launch

Success/failed launches for each site on the map



CCAFS SLC-40



CCAFS LC-40

KSC LC-39A launch site has relatively higher success rates compared to other launch sites

Class = 1 (Green Color) = Successful Launch
Class = 0 (Red Color) = Failed Launch

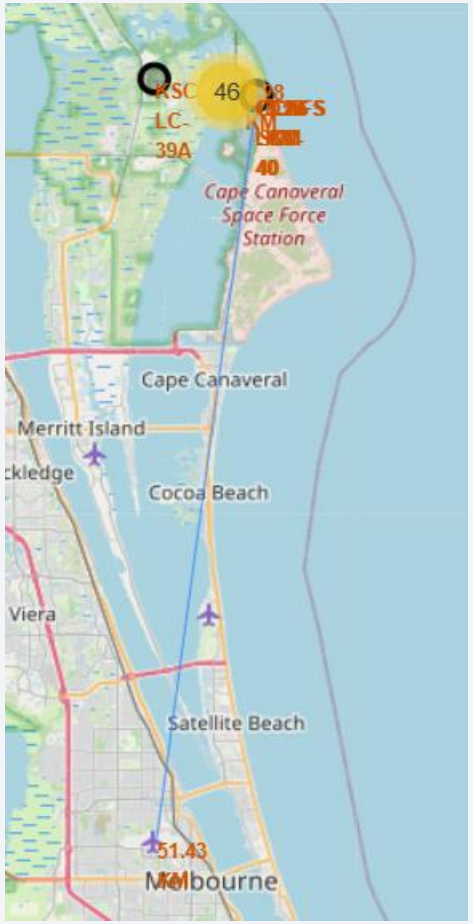
Distances between a launch site to its proximities

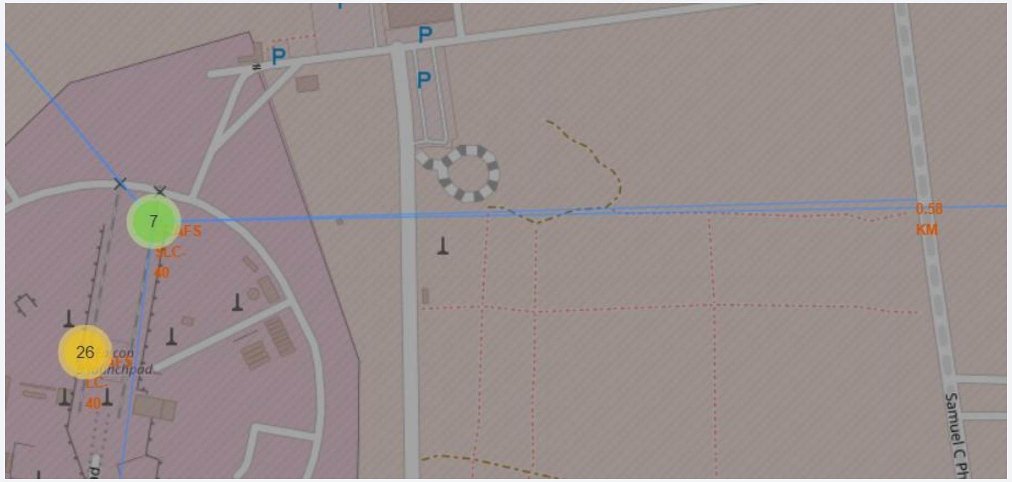
The distance between the closest coastline and the launch site is 0.86km

Distances between a launch site to its proximities

- Distance between a launch site to its closest city, railway, highway:

```
distance_highway = 0.5834695366934144  km
distance_railroad = 1.2845344718142522  km
distance_city = 51.434169995172326  km
```



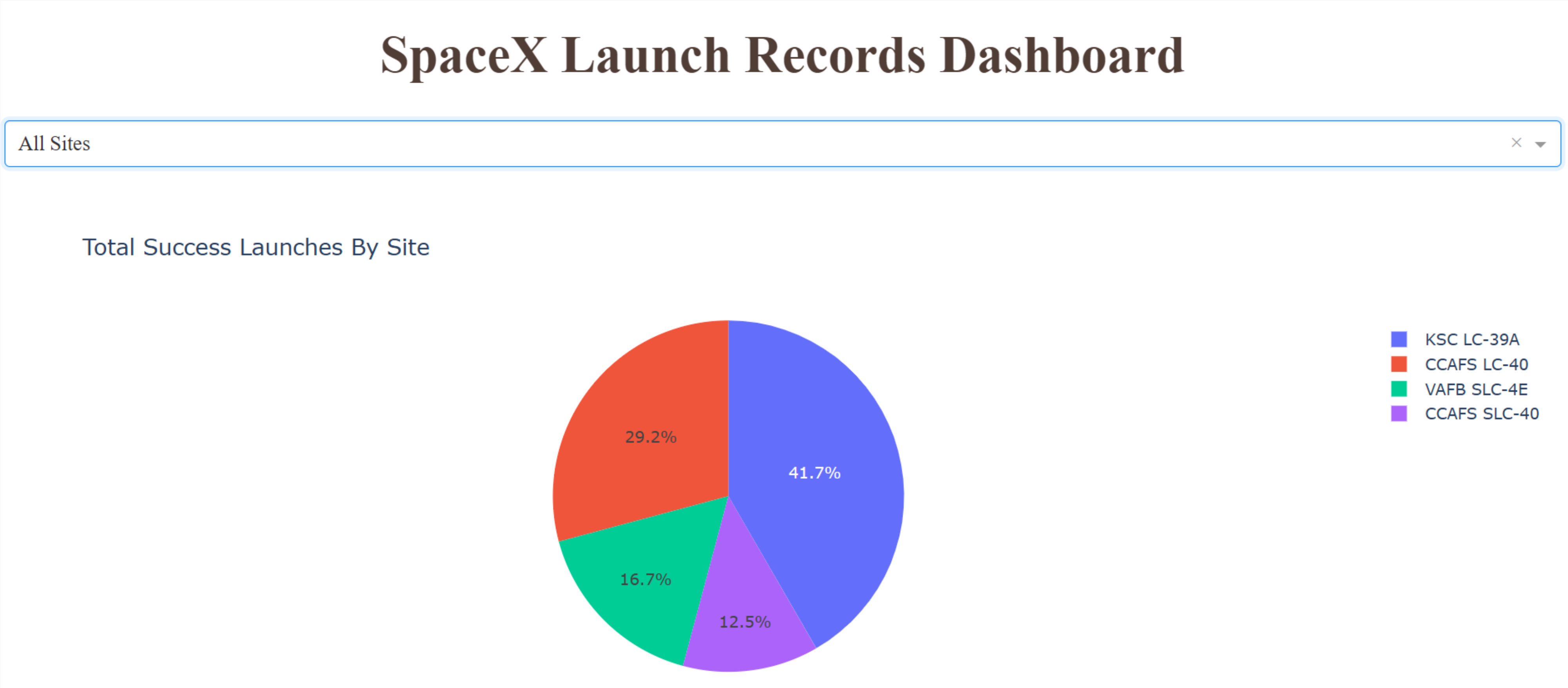


Launch sites are in close proximity to coastline and keep certain distance away from cities

Section 4

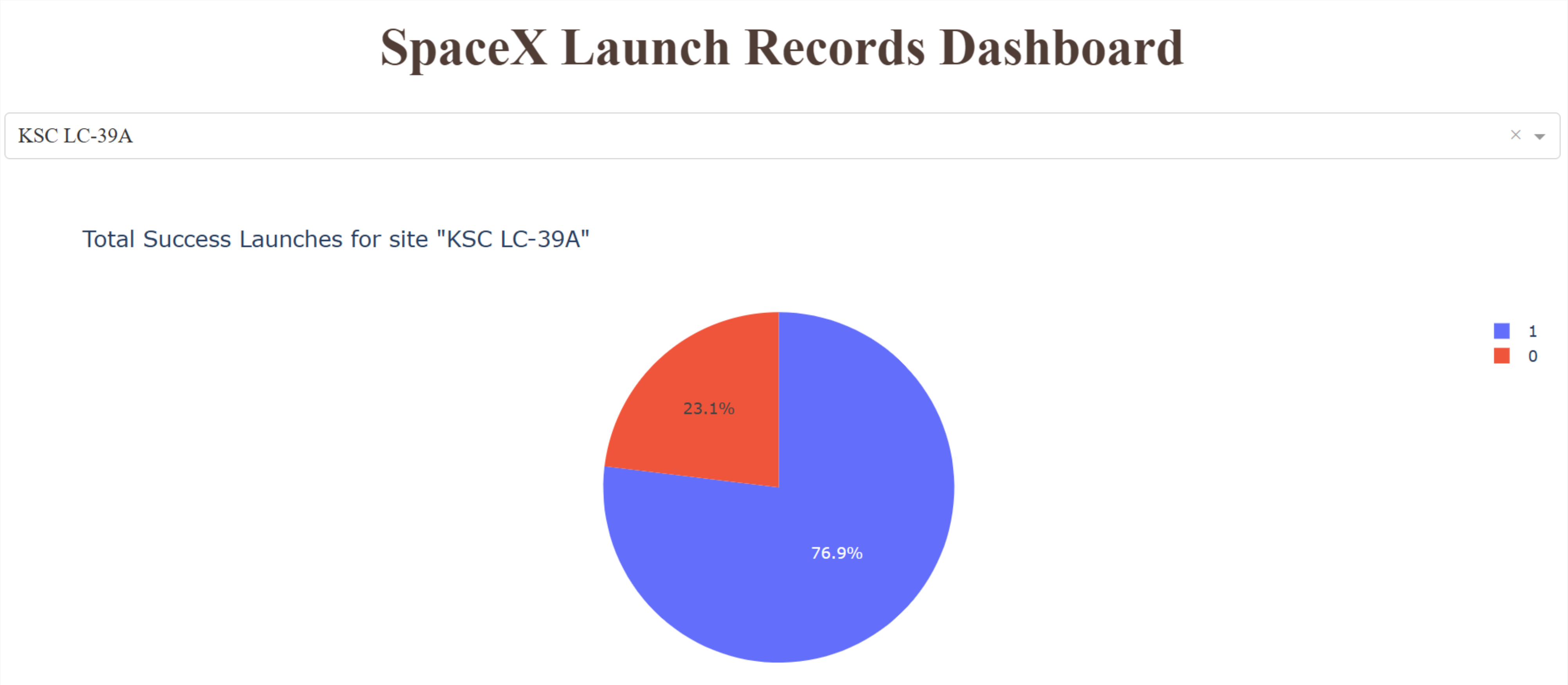
Build a Dashboard with Plotly Dash

Total Success Launches By Site



KSC LC-39A launch site has the highest success rate at almost 42%, followed by CCAFS LC-40 at 29%, VAFB SLC-4E at 16.7% and CCAFS SLC-40 at 12.5%

Highest Success Launch By KSC LC-39A



The total success launch rate for KSC LC-39A was 76.9% with10 /13 successful launches.

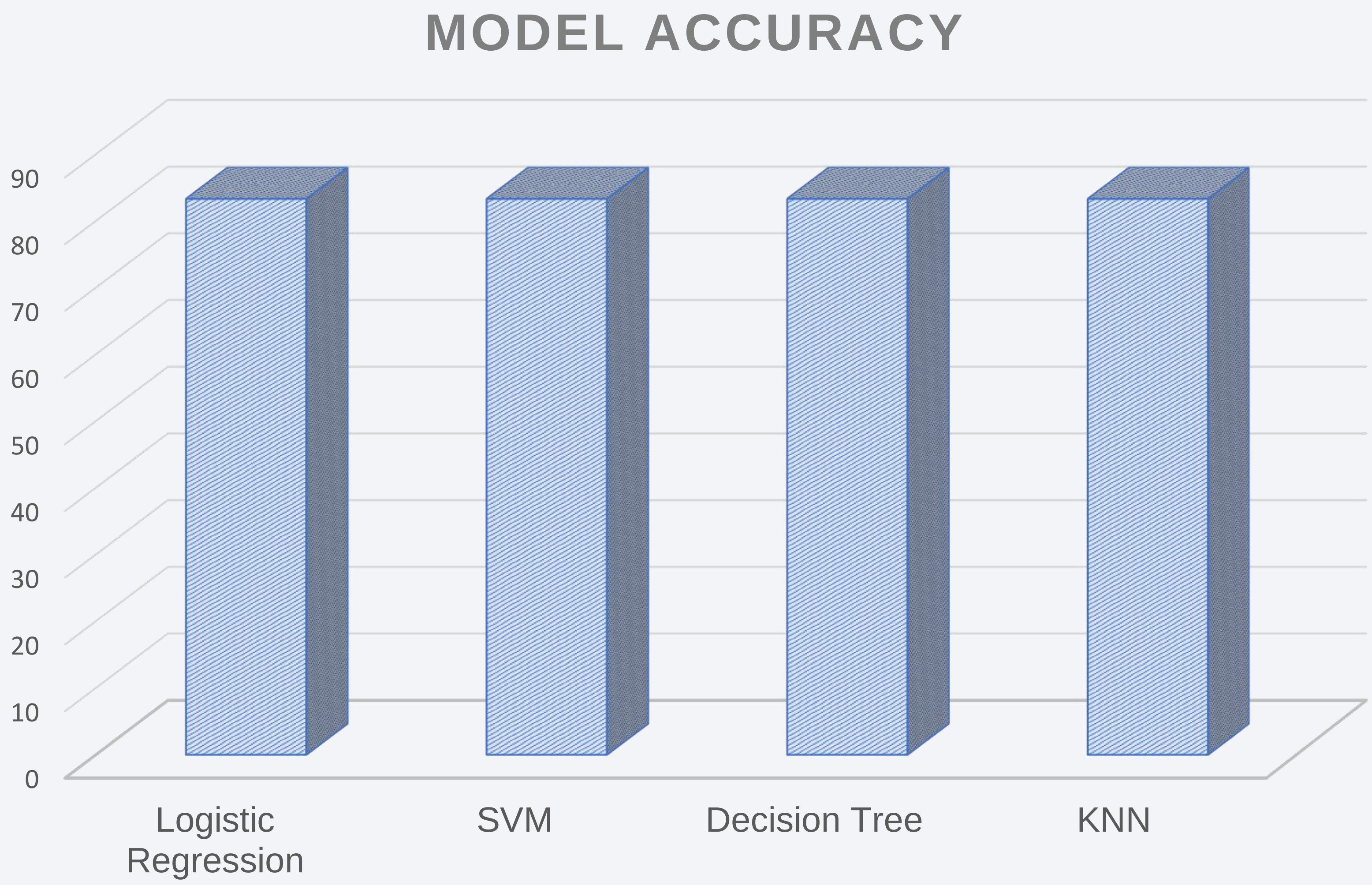
<Dashboard Screenshot 2>

FT Booster Version has the largest success rate. Most of successful launches have payload mass between 1900kg and 5300kg

Section 5

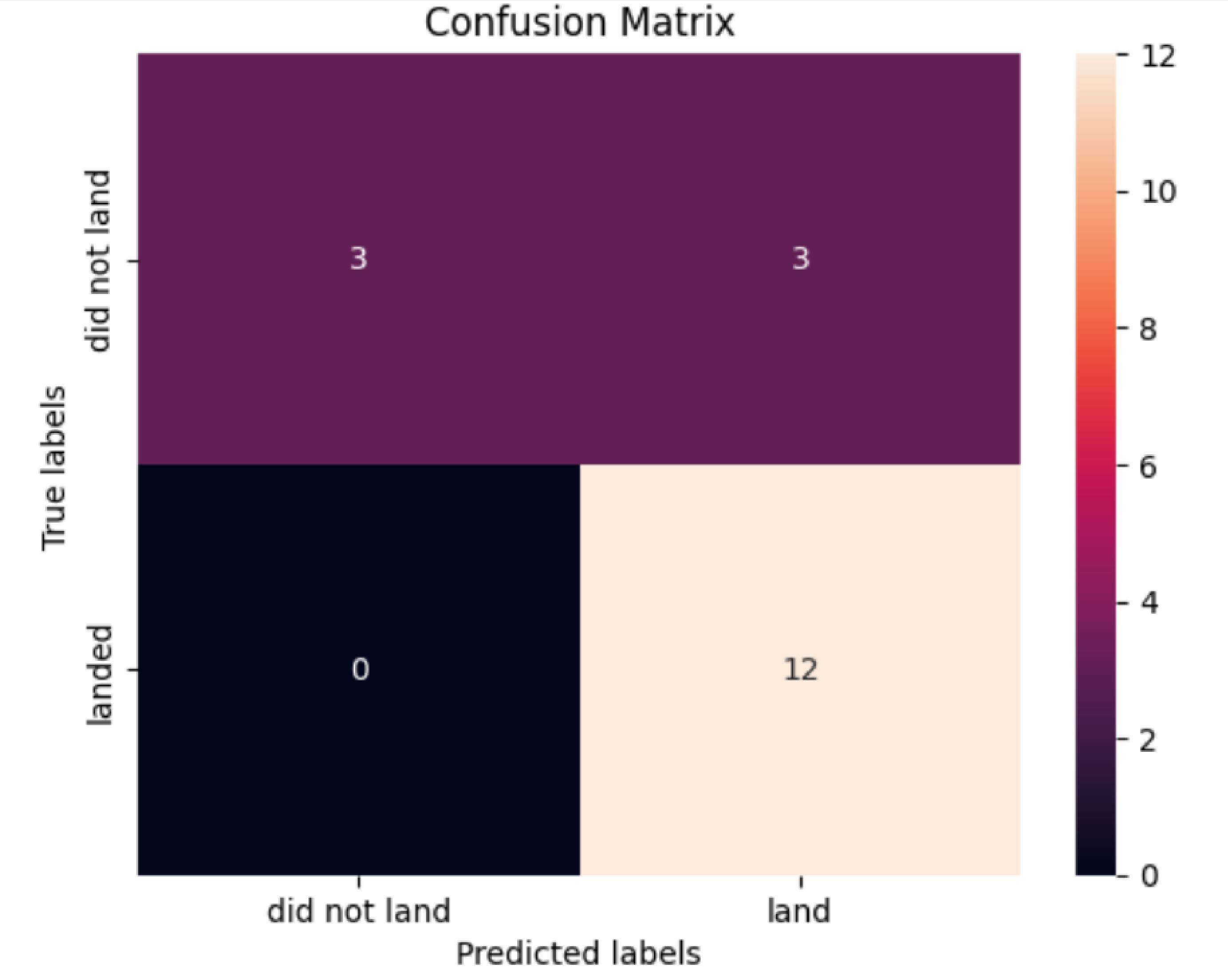
Predictive Analysis (Classification)

Classification Accuracy



Accuracy Score was calculated for Logistic Regression, SVM, Decision Tree and KNN models. The result shows the same accuracy rate: 83.3 % for all methods

Confusion Matrix



All Logistic Regression, SVM, Decision Tree and KNN models depicted the same confusion matrix which was able to predict all landed results successfully, however it poorly predicts mission failure.

Conclusions

- The success rate since 2013 kept increasing till 2020
- With heavy payloads the successful landing rate are more for Polar, LEO and ISS.
- ES-L1, GEO, HEO, SSO orbits have the highest success rate.
- KSC LC-39A launch site has the highest success rate compared to other 3 sites.
- FT Booster Version has the largest success rate. Most of successful launches have payload mass between 1900kg and 5300kg
- All predictive models showed 83.3% accuracy rate along with the same confusion matrix.
- The confusion matrix could predict all landed results successfully, however it poorly predicted mission failure.
- Future work could explore alternative predictive models to increase the accuracy rate.

Appendix

- https://github.com/asat94/Applied-Data-Science-for-SpaceX/tree/main

Thank you!