IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gianluca Barbiere
18/09/2023

Outline

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

Executive Summary

- **Summary of methodologies**
  - Gathering Data, Web Scraping
  - Exploratory Data Analysis using Data Visualization
  - Exploratory Data Analysis using SQL
  - Dynamic Map using Folium
  - Control Panels developed with Plotly Dash
  - Forecasting Analysis
- **Summary of all results**
  - Findings from Exploratory Data Analysis
  - Maps and interactive control panel
  - Prognostic outcome

Introduction

- **Project background and context**
  The objective of this undertaking is to anticipate the successful landing of the initial stage of the Falcon 9. As per information available on their official site, SpaceX states that the expense of propelling the Falcon 9 rocket amounts to 62 million dollars. In contrast, alternate providers charge a minimum of 165 million dollars per launch.

  The variation in cost is elucidated by the capability of SpaceX to recycle the initial stage of the rocket. Evaluating the prospective landing of the stage allows us to establish the expenses linked to a launch.
  This data holds significance for any other enterprise aspiring to rival SpaceX in the domain of rocket launches.

- **Problems you want to find answers**
  We want to calculate the price of each SpaceX space launch, in particular we will predict if the Falcon 9 first stage will land successfully, and what are the interconnections among the rocket factor that influence the landing outome

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:**
  - The requested data was collected using the SpaceX REST API and from Wikipedia via Web Scraping..
- **Perform data wrangling**
  - The data has been analyzed with different Python libraries and handling null values where they were present and performed an One Hot Encoding for the classification models.
- **Perform exploratory data analysis (EDA) using visualization and SQL**
- **Perform interactive visual analytics using Folium and Plotly Dash**
- **Perform predictive analysis using classification models**
  - It was created a machine learning pipeline to predict if the first stage will land with different classification model. (LR, SVM, DTREE, KNN)

Data Collection

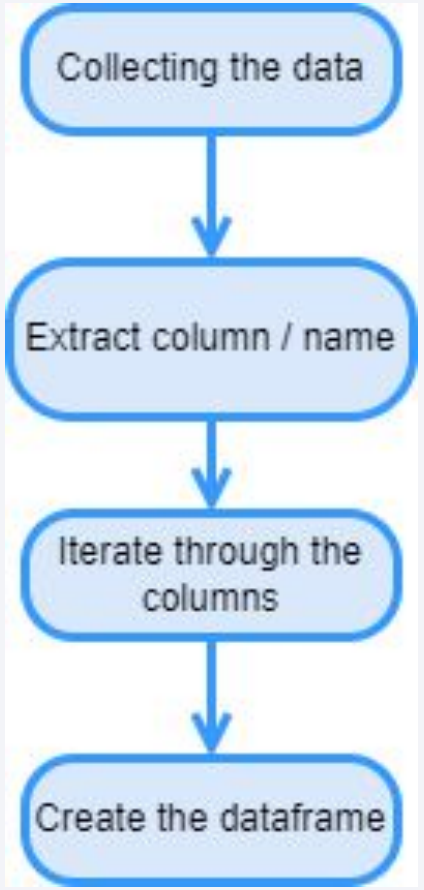
- As mentioned before, the data has been collected using the SpaceX Rest API and Wikipedia via Web Scarping.
  We collected the datas that have meaningful information about rockets, launchpads, payloads, cores

Data Collection – SpaceX API

- Initially, has been requested and parsed the SpaceX launch data using the corresponding request.
- We created a subset of our dataframe with the columns rocket, payloads, launchpad and cores
- After this, we filtered the resulted dataset for a specific voosted version, which is Falcon 9

Data Collection - Scraping

- We requested the required data from the Falcon9 Launch Wiki page and created a BeautifulSoup object to store the data
- With this new object, the specific table that includes the required information was searched
- After the table was founded, a dataframe with those information was created, ready to be analyzed.



Data Wrangling

- Using the dataset that we created in the previous steps, we analyzed the dataset
- We were able to see the number of launches for each site and the number of launches on each orbit.
- Also, we were able to determine the number of landing outcome per orbit type and calculate the success rate

EDA with Data Visualization

- We plotted various relationship between the data as :
  Flight Number, Payload Mass, Launch Site and Success rate per Orbit

- This plots give us information about the Orbit success over the last decade, useful information about launch sites and their rocket launches, and the relationship between the payload mass and success for the return of the first stage

EDA with SQL

- Displayed the launch sites of the dataset
- Displayed the total payload mass carried by a specific booster
- Displayed the average payload mass carried by a specific booster version
- Listed when the first successful landing
- Listed the names of the boosters which have success in drone ship and have a payload mass between 4000 Kg and 6000 Kg
- Listed the number of success and failure missions
- Listed the names of booster version which have carried the max payload
- Listed specific columns, such as, month, landing outcome, booster version and launch site for the year 2015
- Ranked the count of landing outcome between 2010-06-04 and 2017-03-20

Build an Interactive Map with Folium

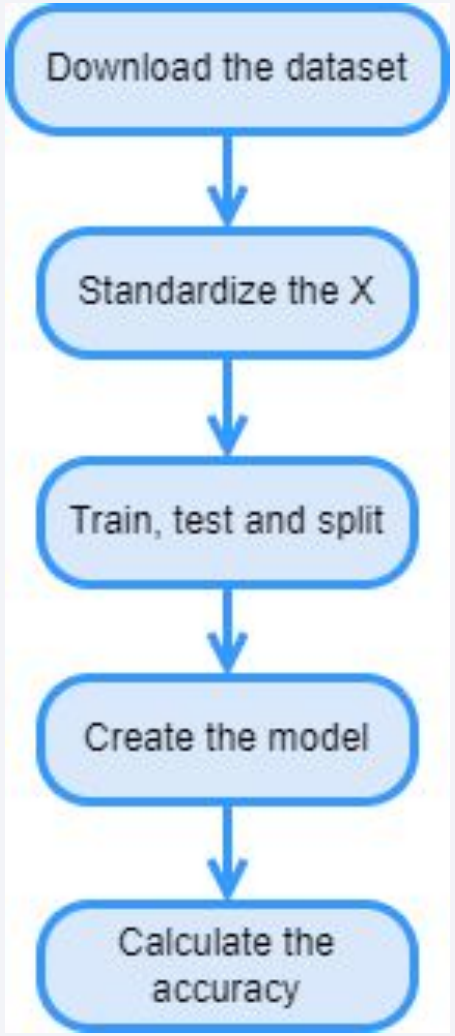
- In this map we added circles to highlight the location of the launch sites, the markers are used to highlight the launches performed by the launch sites which are colored in green or red based on whether they have been successful or not, and the lines were used to see the distance between different key locations

Build a Dashboard with Plotly Dash

- We display 2 plots that shows the ratio of success for every launch site and the correlation between the payload mass and the success outcome for for every launch site
- Thanks to this 2 plots, we can clearly see which site has the best ratio and the importance of the payload mass in every launch

Predictive Analysis (Classification)

- We downloaded the dataset created in the previous steps and defined what is the class to predict Y and the features X used for fit the model
- We splitted the X and Y in the train and test set and created 4 different models to compare which one perform best using metrics and confusion matrix
- At the end, we can see which model did better comparing the accuracy score of each model

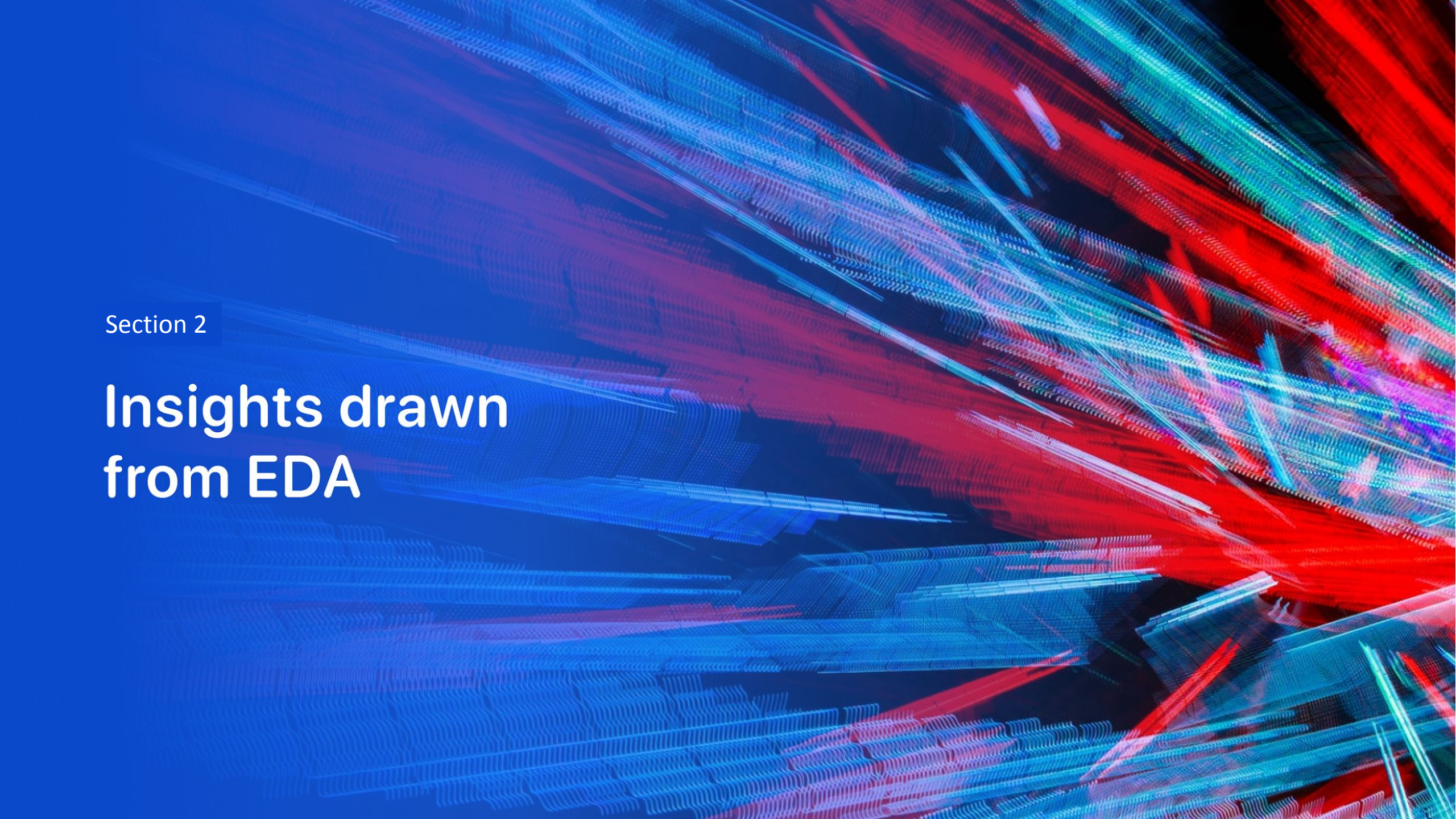


Results

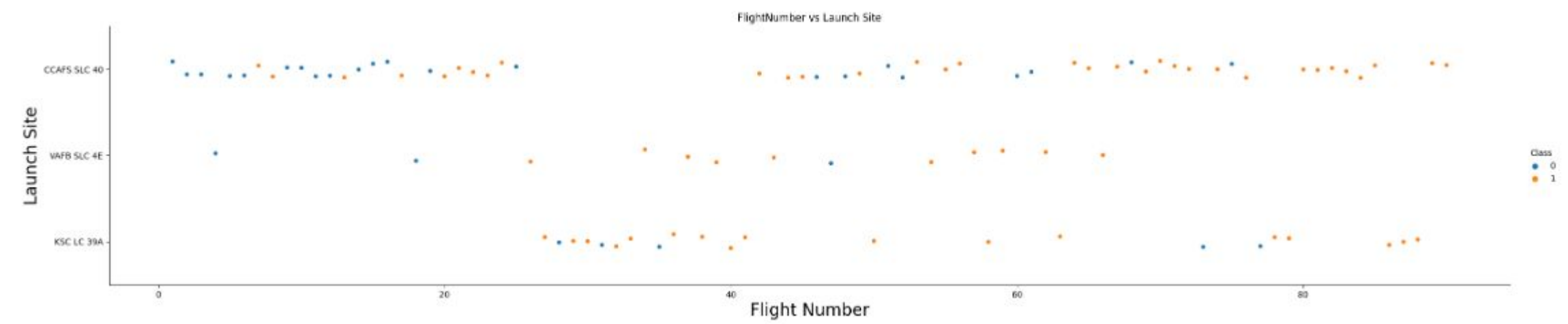
- **Exploratory data analysis results**
- **Interactive analytics demo in screenshots**
- **Predictive analysis results**

Section 2

Insights drawn from EDA

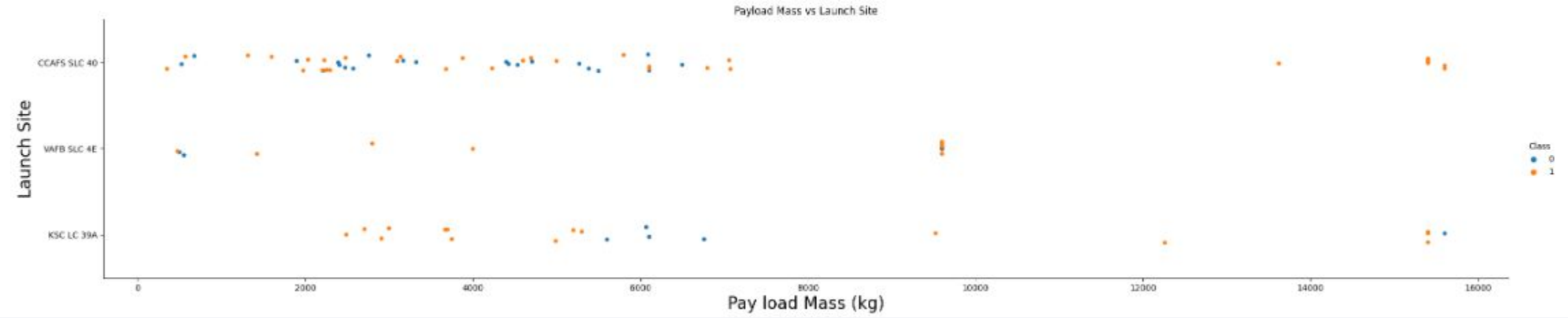


Flight Number vs. Launch Site



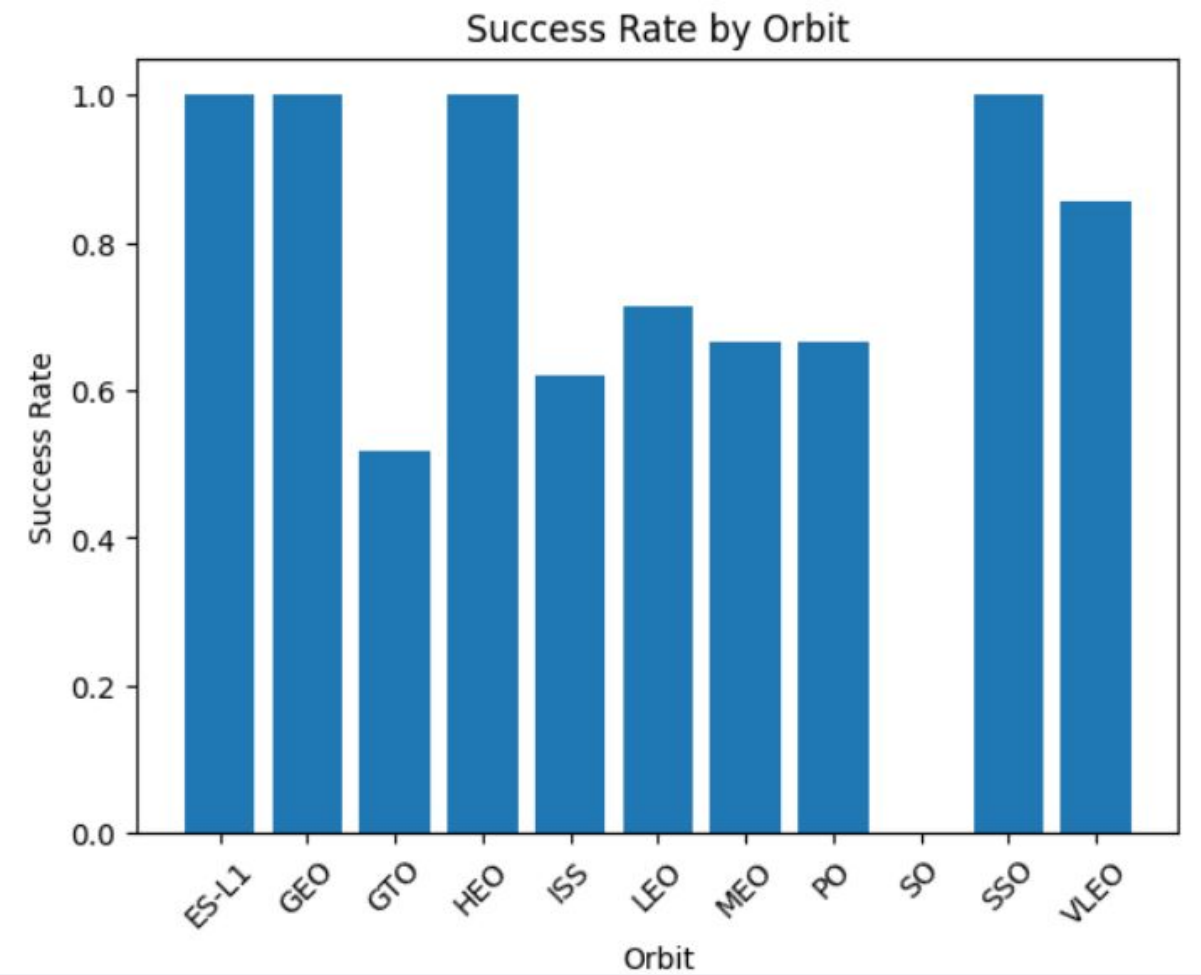
We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC, LC-39A and VAFB SLC 4E has a success rate of 77%

Payload vs. Launch Site



As we can see, for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)

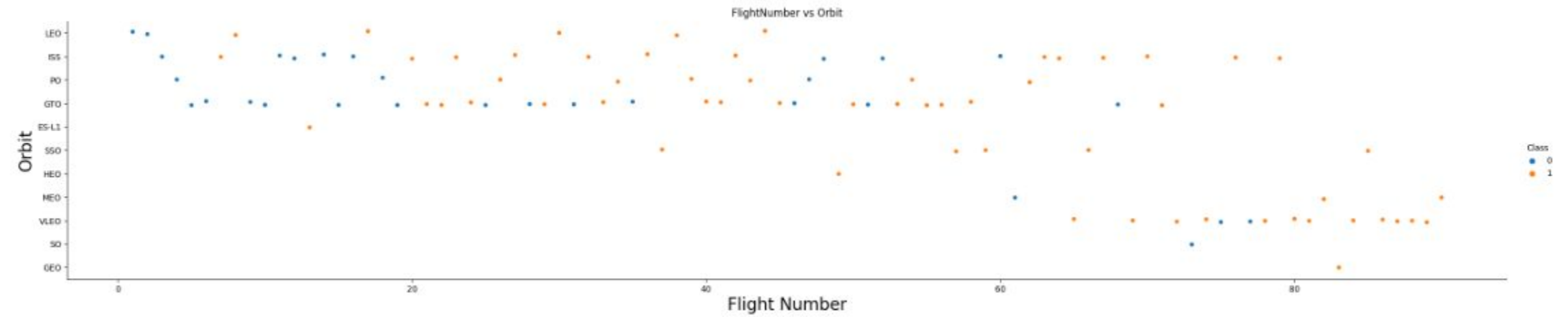
Success Rate vs. Orbit Type



As shown in the plot.
The orbits that have the highest success rate are :

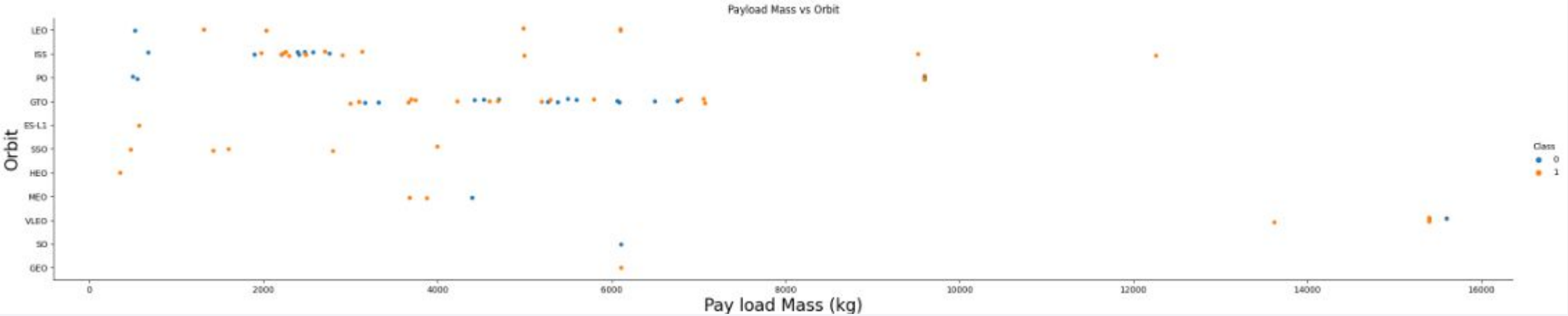
ES-L1, GEO, HEO, SSO

Flight Number vs. Orbit Type



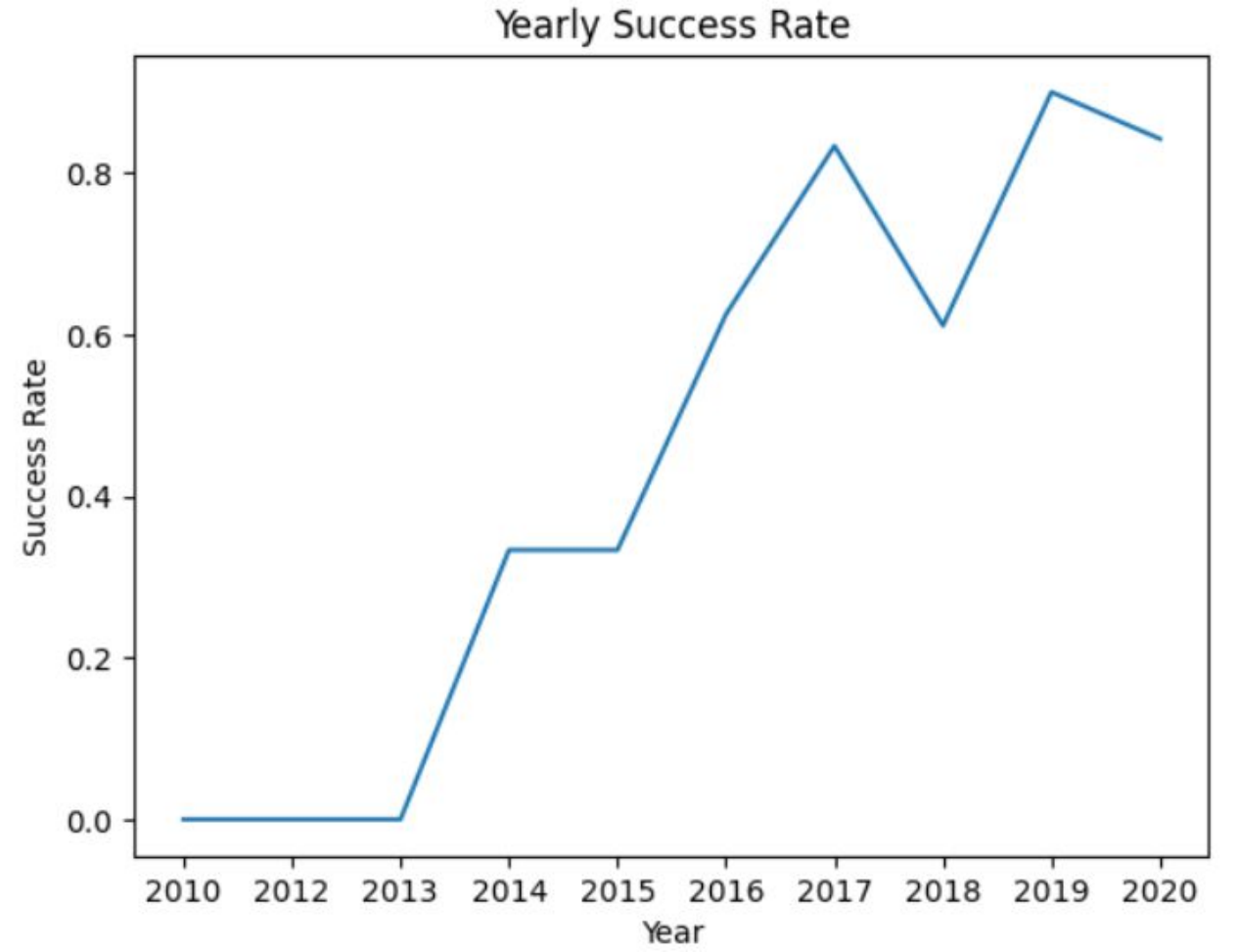
We can see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

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



The plot shows the success rate over the past decade.

We can see that the success rate constantly increases

All Launch Site Names

```
%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

Thanks to this query, we displayed the launch sites from the dataset

Launch Site Names Begin with 'CCA'

```
%%sql select * from SPACEXTBL
    where Launch_Site LIKE "CCA%" limit 5;
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit
2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2

Thanks to this query, we displayed the first 5 launch sites from the dataset that starts with “CCA”

Total Payload Mass

```
%%sql select sum(PAYLOAD_MASS__KG_) as "Payload Carried by NASA (CRS)" from SPACEXTBL
    where Customer = "NASA (CRS)"
```

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

Payload Carried by NASA (CRS)
45596

Thanks to this query, we displayed the total payload carried by boosters from NASA from the dataset

Average Payload Mass by F9 v1.1

```
%%sql select avg(PAYLOAD_MASS__KG_) as "Payload Avg by F9 v1.1" from SPACEXTBL
    where Booster_Version like "F9 v1.1"
```

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

Payload Avg by F9 v1.1
2928.4

Thanks to this query, we displayed the average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

```
%%sql select min(Date) as Date, landing_outcome from SPACEXTBL
    where landing_outcome = "Success (ground pad)"
```

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

Date	Landing_Outcome
2015-12-22	Success (ground pad)

Thanks to this query, we displayed the date of the first successful landing outcome on ground pad

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql select Booster_Version, PAYLOAD_MASS__KG_, landing_outcome from SPACEXTBL
    where PAYLOAD_MASS__KG_ between 4000 and 6000
    and landing_outcome like "Success (drone ship)"
```

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

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)

Thanks to this quey, we displayed 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

```
%%sql select landing_outcome, count(*) as count from SPACEXTBL
    where landing_outcome like "Success%" or landing_outcome like "Failure%"
    group by landing_outcome
```

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

Landing_Outcome	count
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
Success	38
Success (drone ship)	14
Success (ground pad)	9

Thanks to this quey, we calculated the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

```
%%sql select distinct(Booster_Version), PAYLOAD_MASS__KG_ as Max_Payload from SPACEXTBL
    where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

Booster_Version	Max_Payload
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

Thanks to this quey, we listed the names of the booster which have carried the maximum payload mass, which is 15600

2015 Launch Records

```
%%sql select case
    when substr(Date, 6, 2) = '01' then 'January'
    when substr(Date, 6, 2) = '02' then 'February'
    when substr(Date, 6, 2) = '03' then 'March'
    when substr(Date, 6, 2) = '04' then 'April'
    when substr(Date, 6, 2) = '05' then 'May'
    when substr(Date, 6, 2) = '06' then 'June'
    when substr(Date, 6, 2) = '07' then 'July'
    when substr(Date, 6, 2) = '08' then 'August'
    when substr(Date, 6, 2) = '09' then 'September'
    when substr(Date, 6, 2) = '10' then 'October'
    when substr(Date, 6, 2) = '11' then 'November'
    when substr(Date, 6, 2) = '12' then 'December'
    else 'Unknown'
  end as Month,
  landing_outcome, booster_version, launch_site from SPACEXTBL
    where landing_outcome like "Failure (drone ship)"
    and substr(Date, 1, 4) = '2015';
```

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

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

Thanks to this quey, we listed the failed landing outcomes in drone ship, their booster versions, month,
and launch site names for in year 2015

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

```
%%sql select landing_outcome, count(*) as Count FROM SPACEXTBL
    where Date between '2010-06-04' and '2017-03-20'
    and landing_outcome = 'Failure (drone ship)' OR landing_outcome = 'Success (ground pad)'
    group by landing_outcome
    order by Count desc
```

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

Landing_Outcome	Count
Success (ground pad)	9
Failure (drone ship)	5

Thanks to this quey, we ranked the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010/06/04 and 2017/03/20

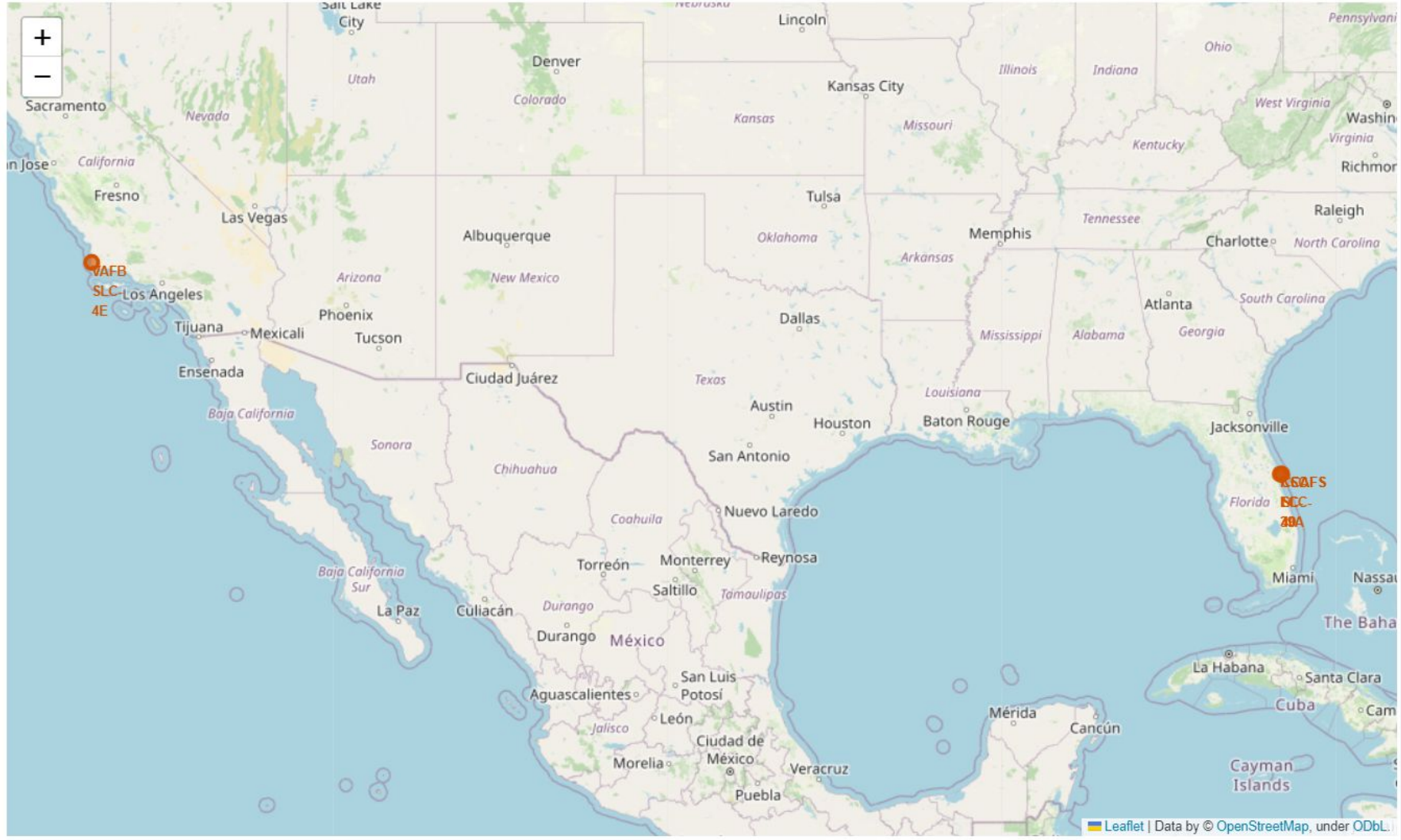

Section 3
Launch Sites
Proximities Analysis

Folium Instructions

Due to a GitHub error that I couldn't resolve, for proper viewing of the .ipynb file, please use the following link: https://nbviewer.org/

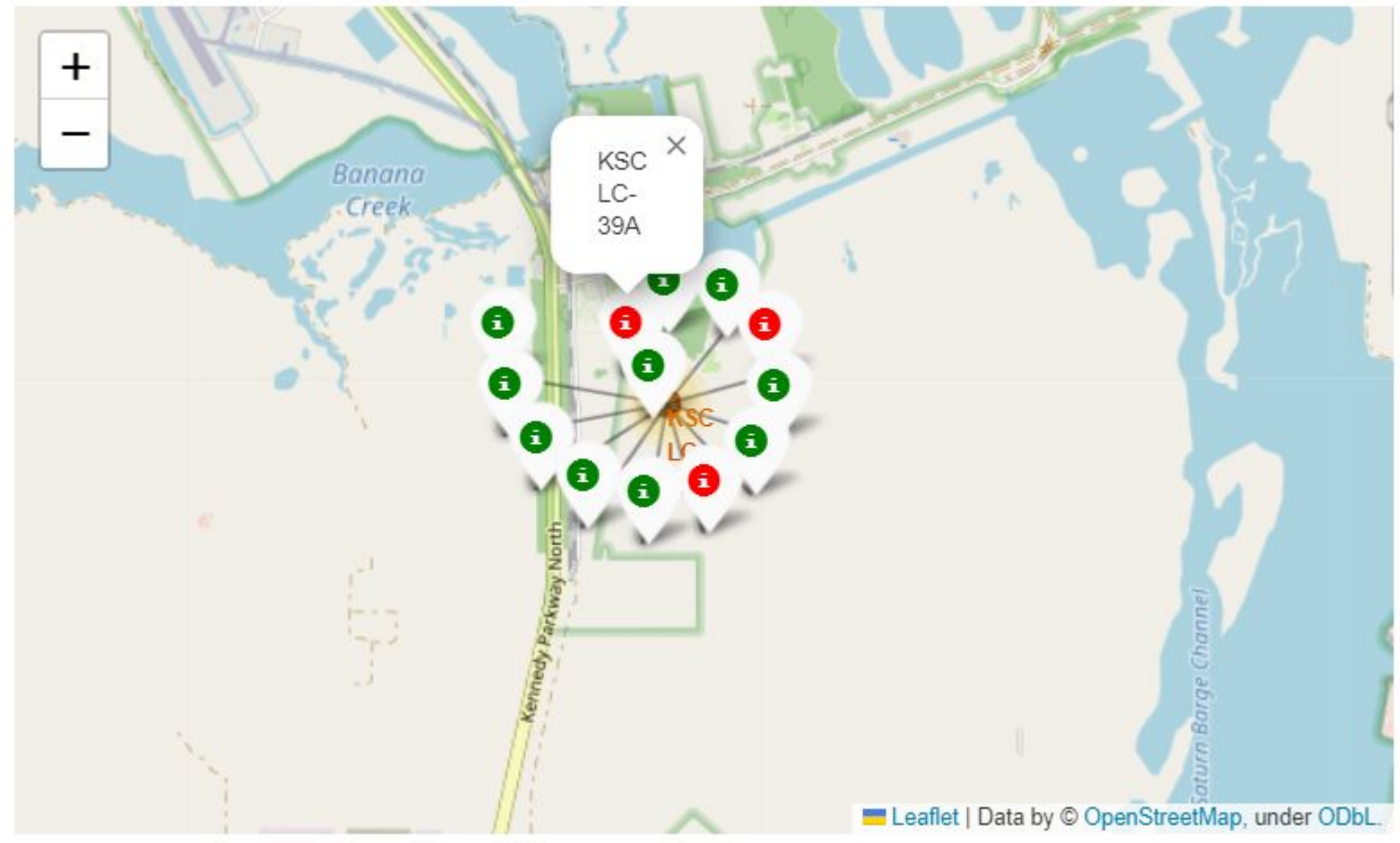
nbviewer allows you to correctly view the file with the generated maps. Just enter the GitHub file path in the search bar, and click 'Go!'

Launch Sites location



The map shows where the launch sites of VAFB - SLC4E, KSC LC-39A and CCAF SLC-40 are in the United States

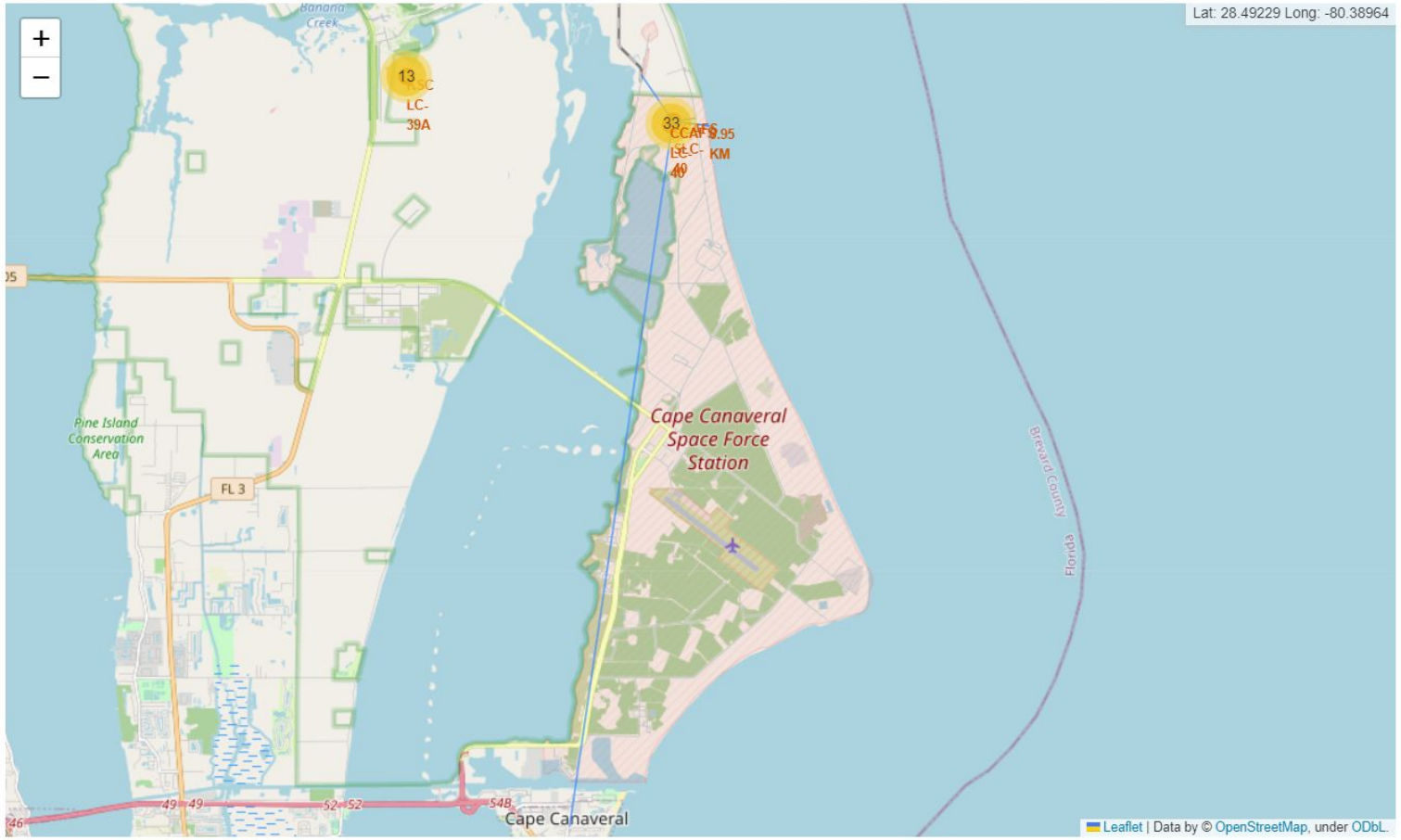
Map of Success/Failed launches for KSC LC-39A



The map aside, shows the number of launches that have been done in the launch site KSC LC-39A.

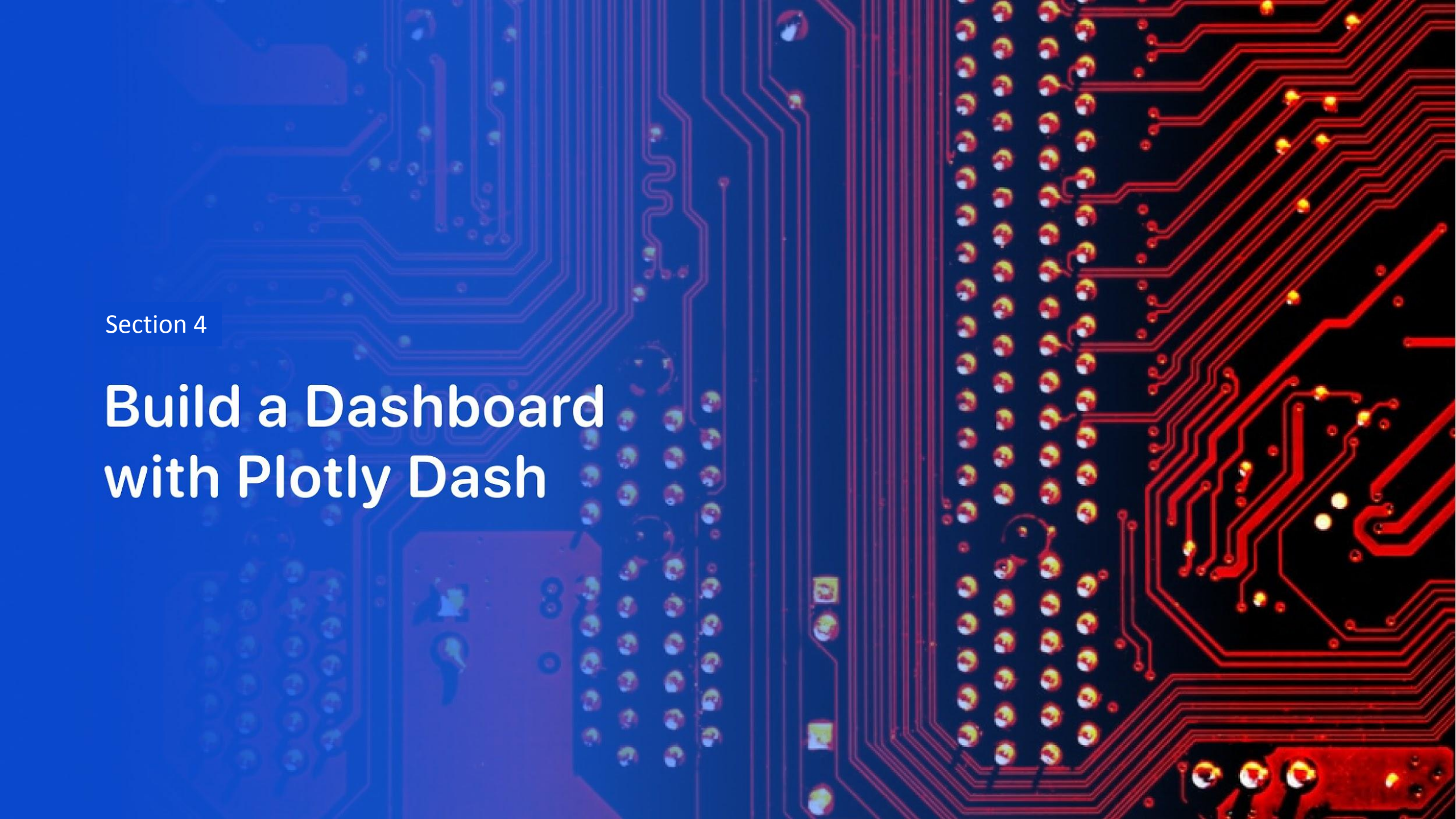
The launches that are marked by the color green indicates that the first stage landed, instead the color red indicates that it does not landed

Demographic information of CCAFS SLC-40

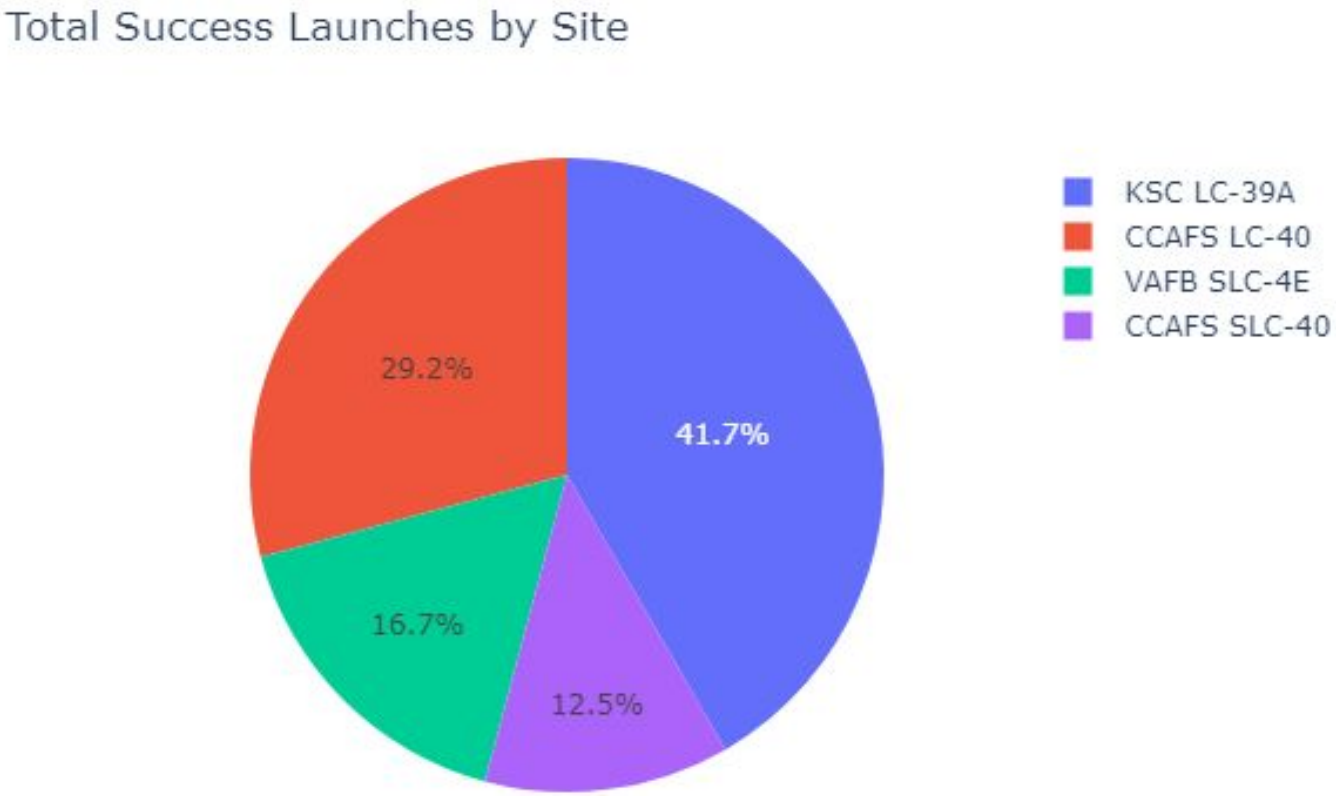


As shown in the mapp aside, we can see the distance between launch site and the coast, which is 0.95KM.

You can also view the close proximity to railways, highways and the city nearby, which is Cape Canaveral

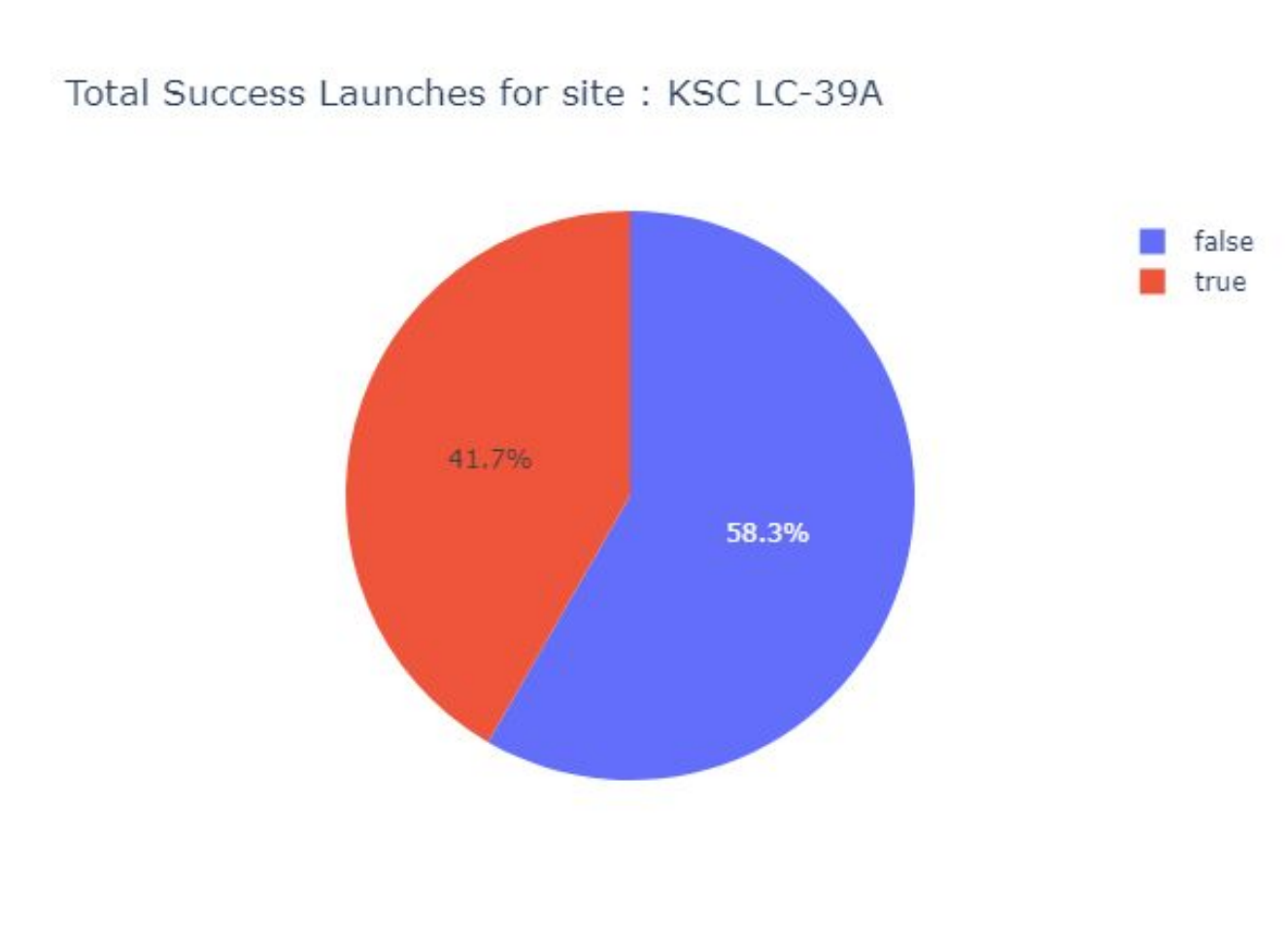

Section 4
Build a Dashboard with Plotly Dash

Total Success Launches by Site



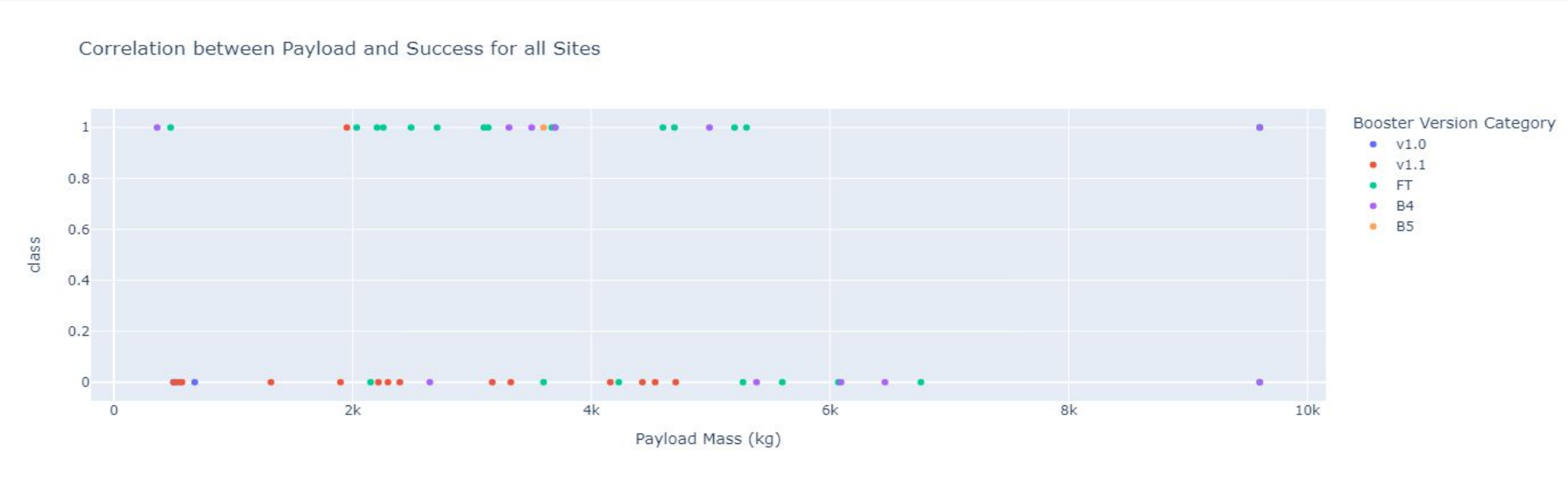
As shown in the figure, we can see that the site KSC LC-39A has the highest success launches

Total Success Launches for KSC LC-39A



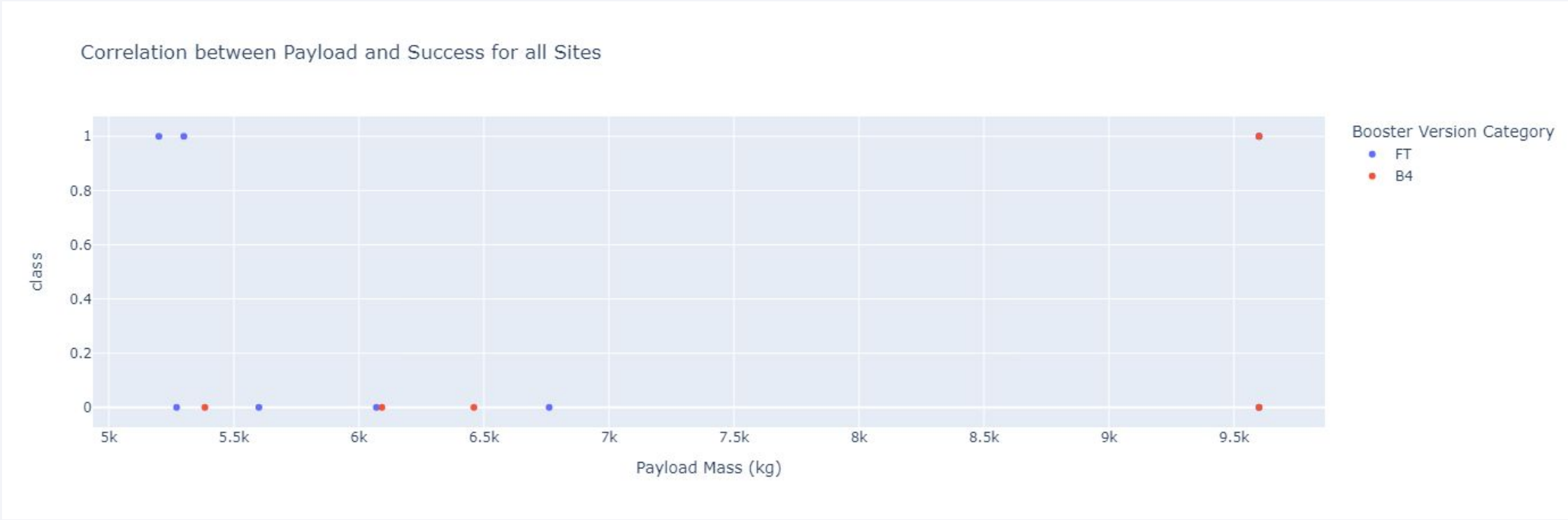
The site KSC LC-39A has the success ratio with 58.3% of first stage success landing

Payload mass vs Success outcome



This plot show the correlation between Payload mass, with range 0 to 10K, and Success outcome for all Sites.As shown in the figure, as long as the payload mass increase, the success is likely to happen, Except for the booster version v1.0 and v1.1

Payload mass vs Success outcome



This plot show the correlation between Payload mass, with range 5K to 10K, and Success outcome for all Sites.As shown in the figure, as long as the payload mass increase, the success is likely to happen, Except for the booster version v1.0 and v1.1

Section 5

Predictive Analysis (Classification)

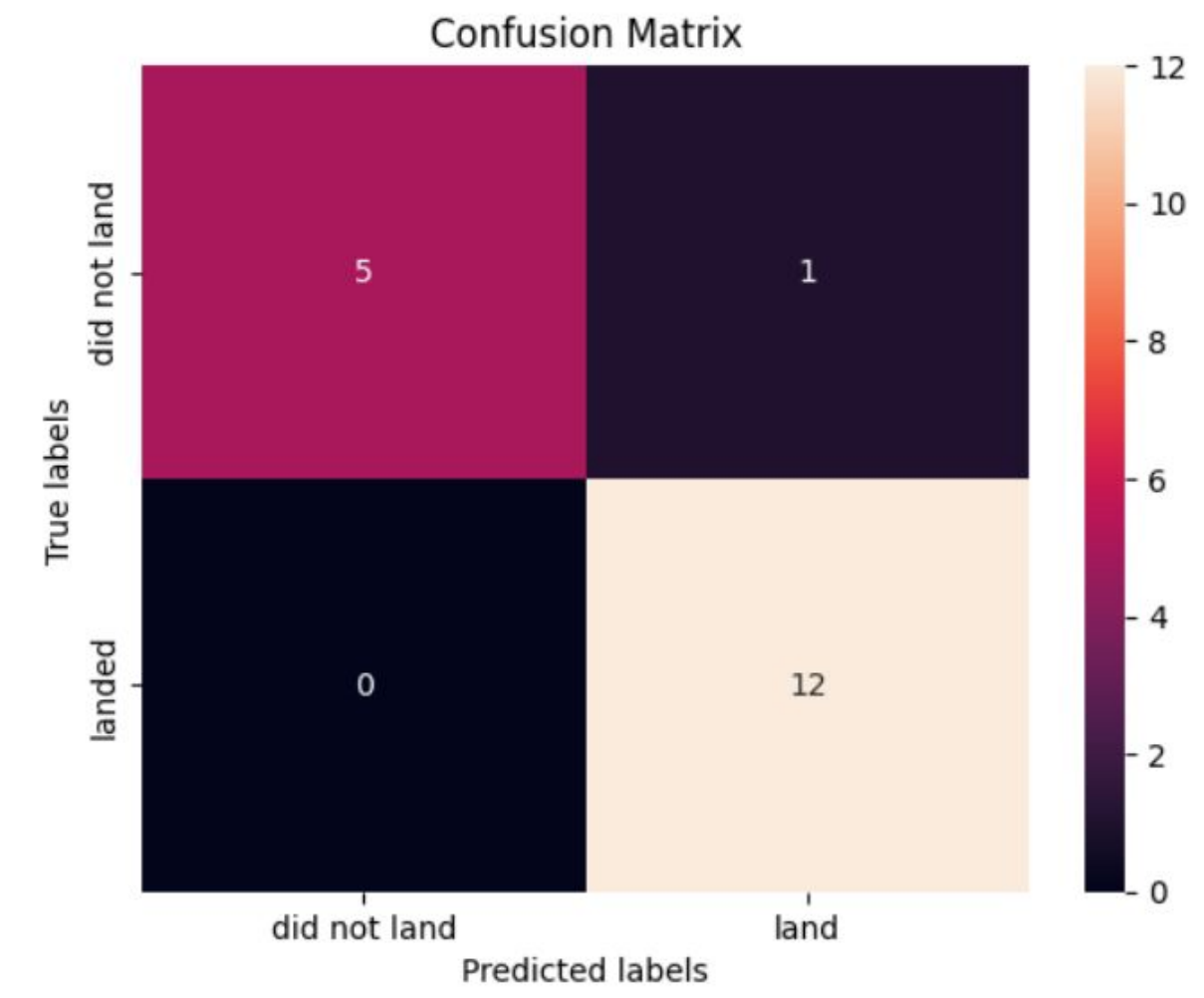
Classification Accuracy

```
print("Logistic Regression accuracy :",logreg_cv.best_score_)
print("SVM accuracy :",svm_cv.best_score_)
print("Decision Tree accuracy :",tree_cv.best_score_)
print("KNN accuracy :",knn_cv.best_score_)
```

```
Logistic Regression accuracy : 0.8222222222222222
SVM accuracy : 0.8222222222222223
Decision Tree accuracy : 0.9
KNN accuracy : 0.8444444444444444
```

The Decision Tree classifier performs better than the other ones, with an accuracy of 0.9

Confusion Matrix



The model predicts correctly that in 5 launches the first stage does not land and in 12 launch does it land.

The model has wrongly predicts that in 1 launch the first stage does land and in 0 launches does not land

Conclusions

- We have analyzed and preprocessed the data from the SpaceX dataset, dealing with null values and we have shown what the major features of interest were in the dataset
- We have displayed the success rate of each launch site and the relationship between the most important variables that ensure the success of the first stage landing.
- We also have shown some demographic information in the map about the launch sites.
- We have trained various classification algorithms to determine whether a launch can successfully land the first stage or not.

Thank you!