IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

André Chrisostomo
14/02/2025

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- This project aims to visualize SpaceX flight data and learn how to predict certain events. I.e: predict if the first stage of the booster will successfully return to the base for reuse
- Can we determine the chance for a booster to successfully land back on Earth?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected using SpaceX's Rest API
- Perform data wrangling
  - We used SpaceX's Rest API to convert the given hashes to readable data
- Perform exploratory data analysis (EDA) using visualization and SQL
  - We used SQL to get an understanding about our data
- Perform interactive visual analytics using Folium and Plotly Dash
  - We used Folium and Plotly Dash to interact with our data in an intuitive way
- Perform predictive analysis using classification models
  - We built classification models to predict our chosen dependent variable

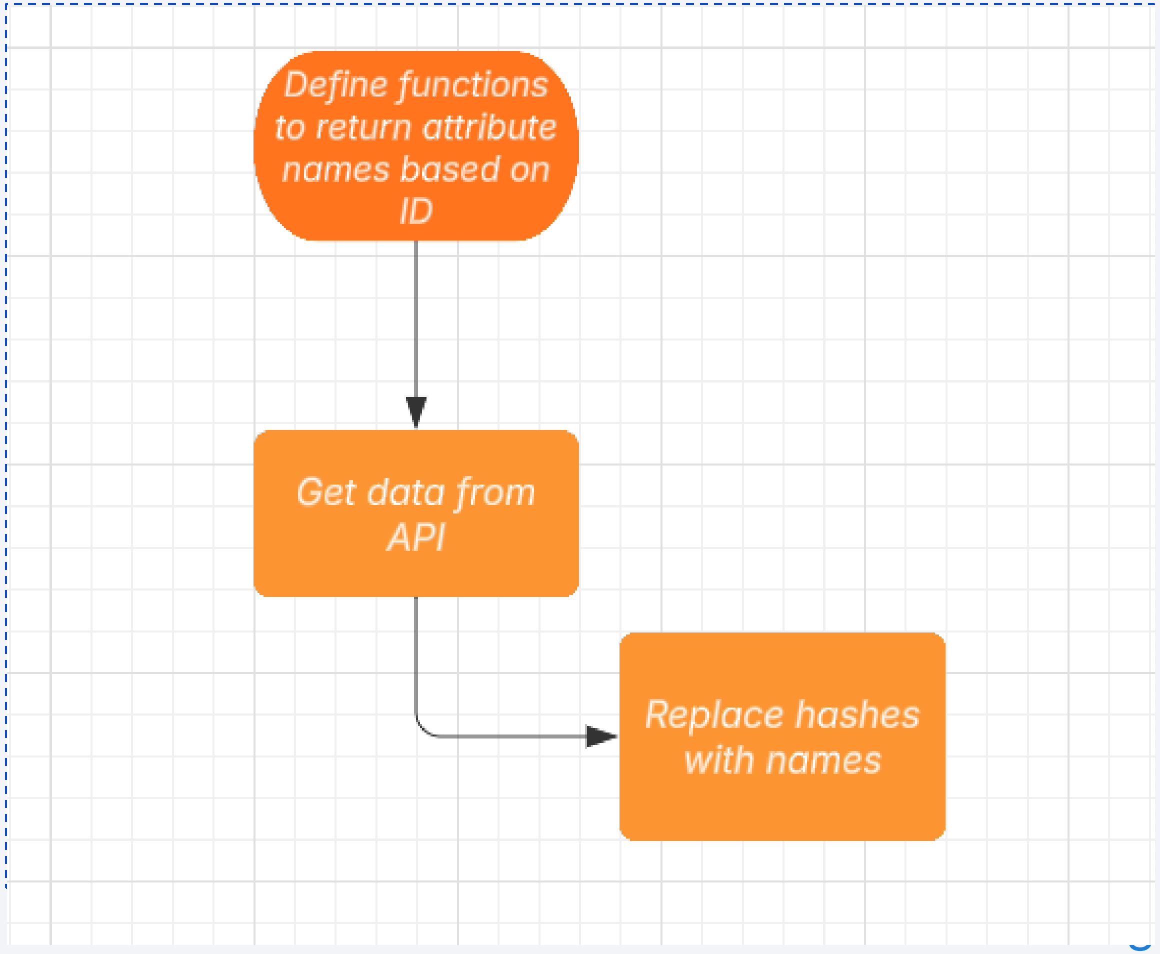
Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

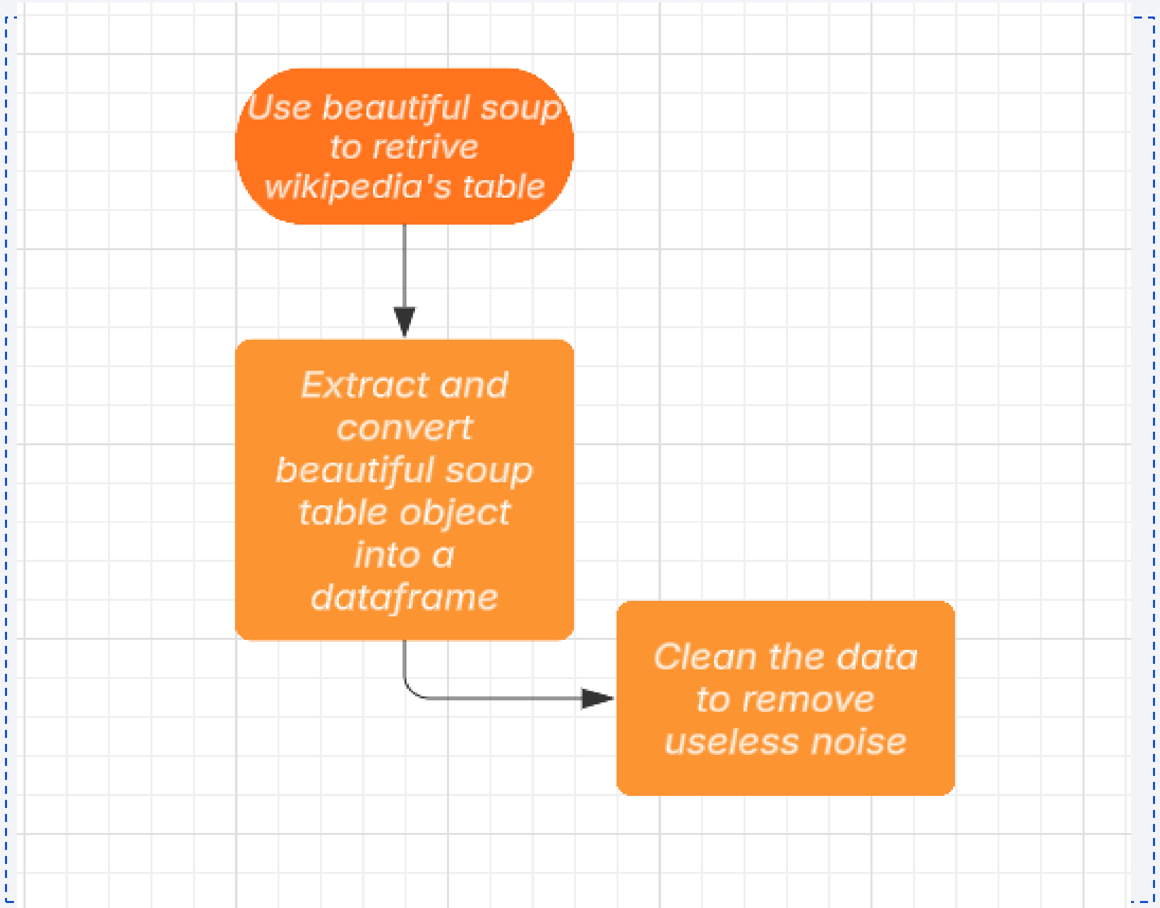
- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- https://github.com/AndreNalinChrisostomo/Data-Science/blob/main/SPACE_X/LAB%201/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

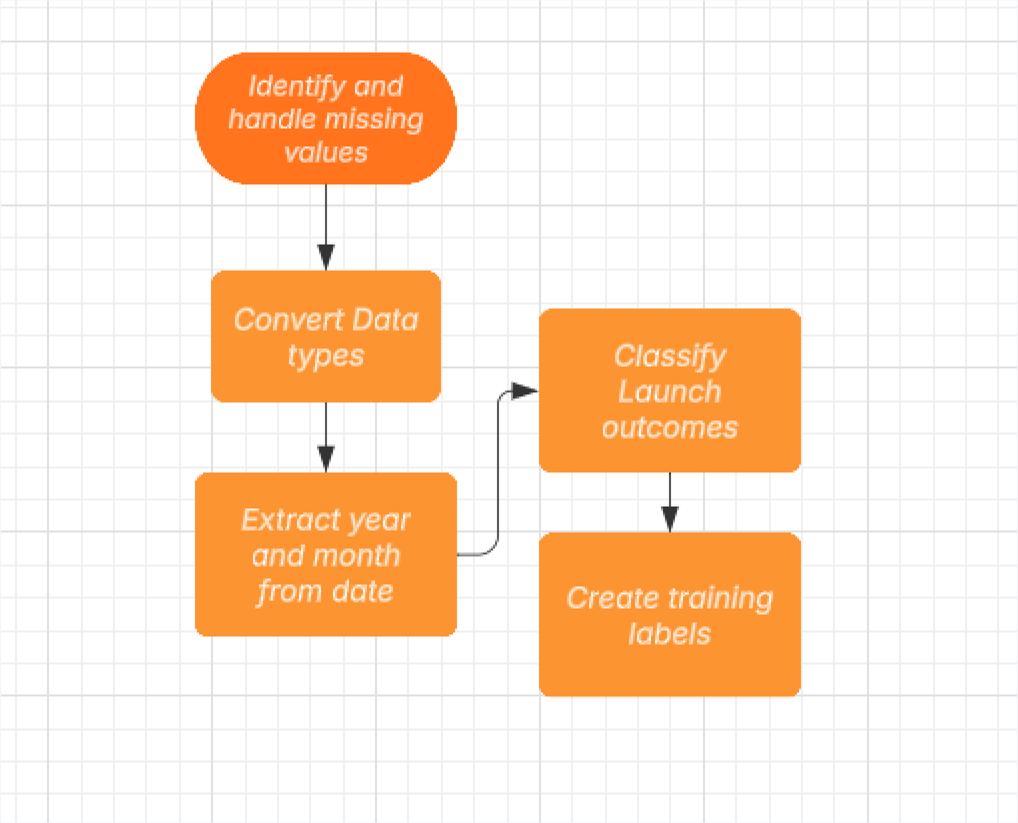
- Present your web scraping process using key phrases and flowcharts
- https://github.com/AndreNalinChrisostomo/Data-Science/blob/main/SPACE_X/LAB%202/jupyter-labs-webscraping.ipynb



Data Wrangling

- Perform exploratory data analysis to identify patterns, calculate missing values, classify launch outcomes, and create training labels for SpaceX Falcon 9 first stage landing prediction.

- https://github.com/AndreNalinChrisostomo/Data-Science/blob/main/SPACE_X/LAB%203/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- We used bar charts to visualize the count of successful and failed launches, line charts to visualize launch outcomes over time, scatter plots to identify relationships between features, histograms to visualize distribution of numerical features, box plots to visualize to identify outliers and heatmaps to visualize correlations between features.
- https://github.com/AndreNalinChrisostomo/Data-Science/blob/main/SPACE_X/LAB%203/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - SELECT DISTINCT
  - SELECT SUM
  - SELECT MIN
  - SELECT AVG
  - SELECT COUNT
  - SELECT MAX
- https://github.com/AndreNalinChrisostomo/Data-Science/blob/main/SPACE_X/LAB%204/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

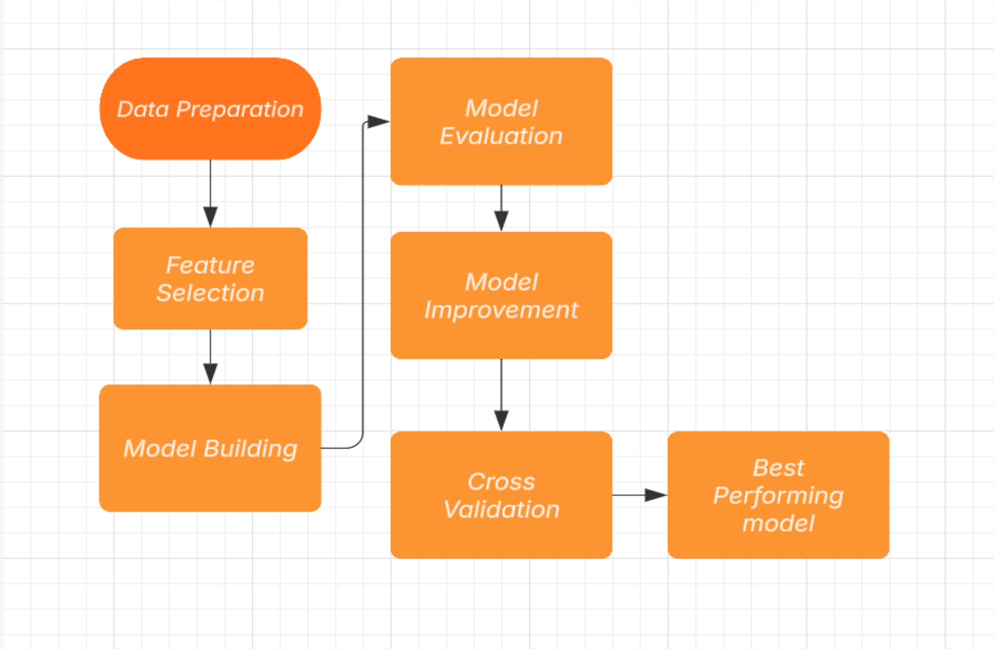
- Circles, Markers, MarkerClusters, PolyLines
- These objects were added to better visualize the position of the bases on the map
- https://github.com/AndreNalinChrisostomo/Data-Science/blob/main/SPACE_X/LAB%206/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- I used Pie charts and scatter plots were used to visualize data in plotly
- I added pie charts to better visualize the proportion of successful and non-successeful launches, and scatter plots to visualize interactions between features and their respective launch status

Predictive Analysis (Classification)

- We prepared our data, selected the features, built our model using Logistic Regression, SVM and decision trees. Evaluated each model and hypertuned the parameters to find the better model.
- https://github.com/AndreNalinChrisostomo/Data-Science/blob/main/SPACE_X/LAB%207/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

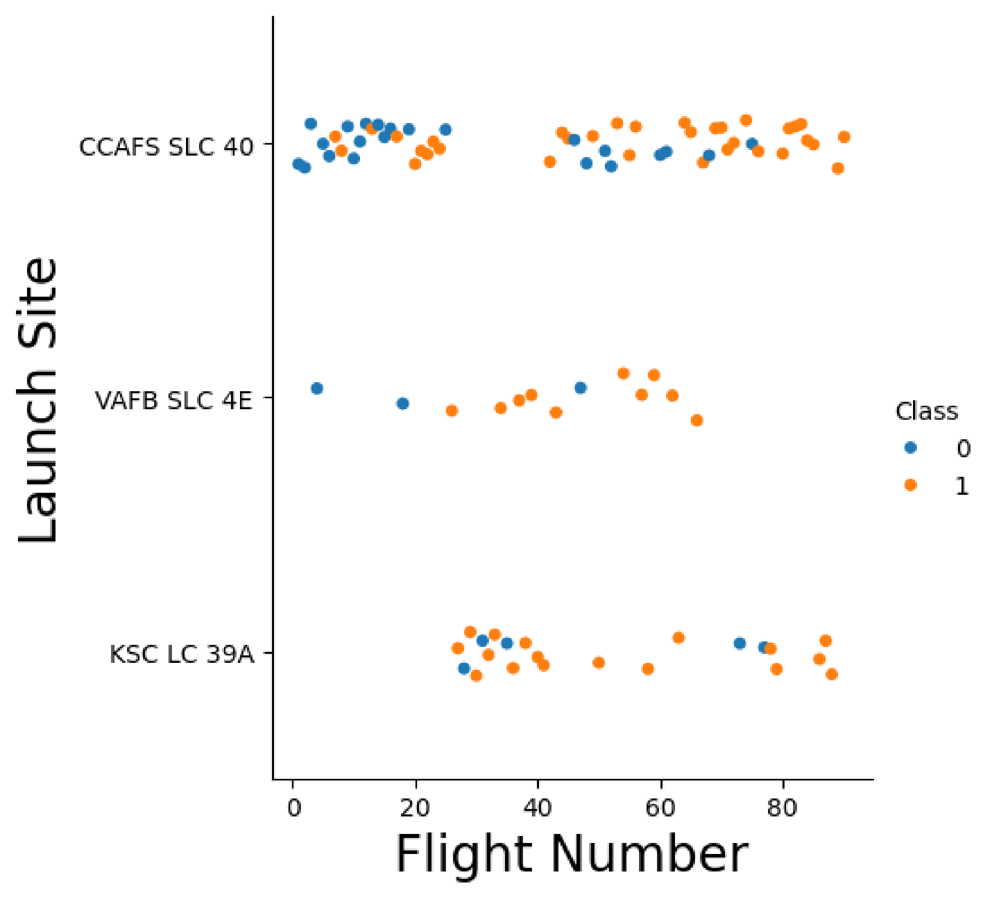
From the scatter plot of Flight Number vs. Launch Site, we can observe the following patterns:

1. **CCAFS SLC 40**: This launch site has a mix of successful and unsuccessful landings across different flight numbers. However, as the flight number increases, the number of successful landings also increases.

2. **KSC LC 39A**: This launch site shows a higher number of successful landings compared to unsuccessful ones, especially as the flight number increases.

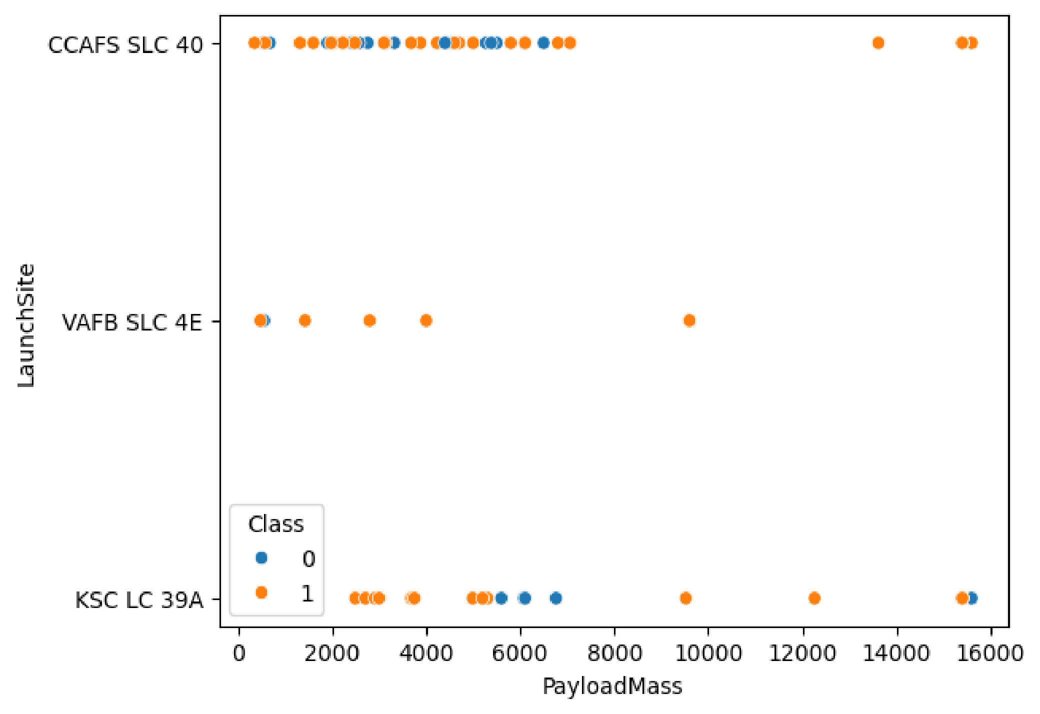
3. **VAFB SLC 4E**: This launch site has fewer data points, but it shows a trend of successful landings with increasing flight numbers.

Overall, it appears that as the flight number increases, indicating more experience and possibly improved technology, the likelihood of successful landings also increases across different launch sites.



Payload vs. Launch Site

In the scatter plot on the right we can se a variety of launches both successful and unsuccessful. With a tendency of most successful launches with heavier payloads. This might be because of the evolution in technologies and the growing confidence as the lauches became more successful.



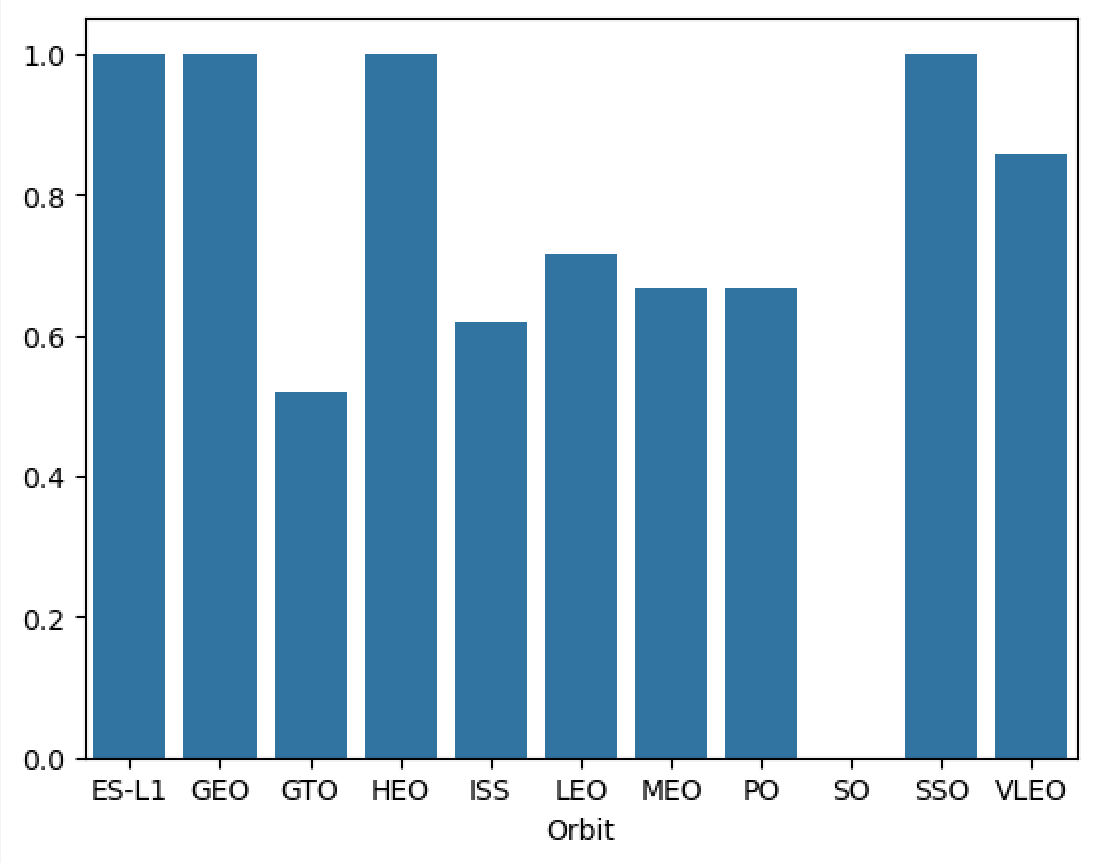
Success Rate vs. Orbit Type

In the bar chart, we can see that the most successful orbit types are:

- ES-L1
- GEO
- HEO
- SSO

And the least successful orbit types are:

- GTO
- ISS
- LEO
- MEO
- PO



Flight Number vs. Orbit Type

From the scatter plot of Payload Mass vs. Orbit type, we can observe the following patterns:

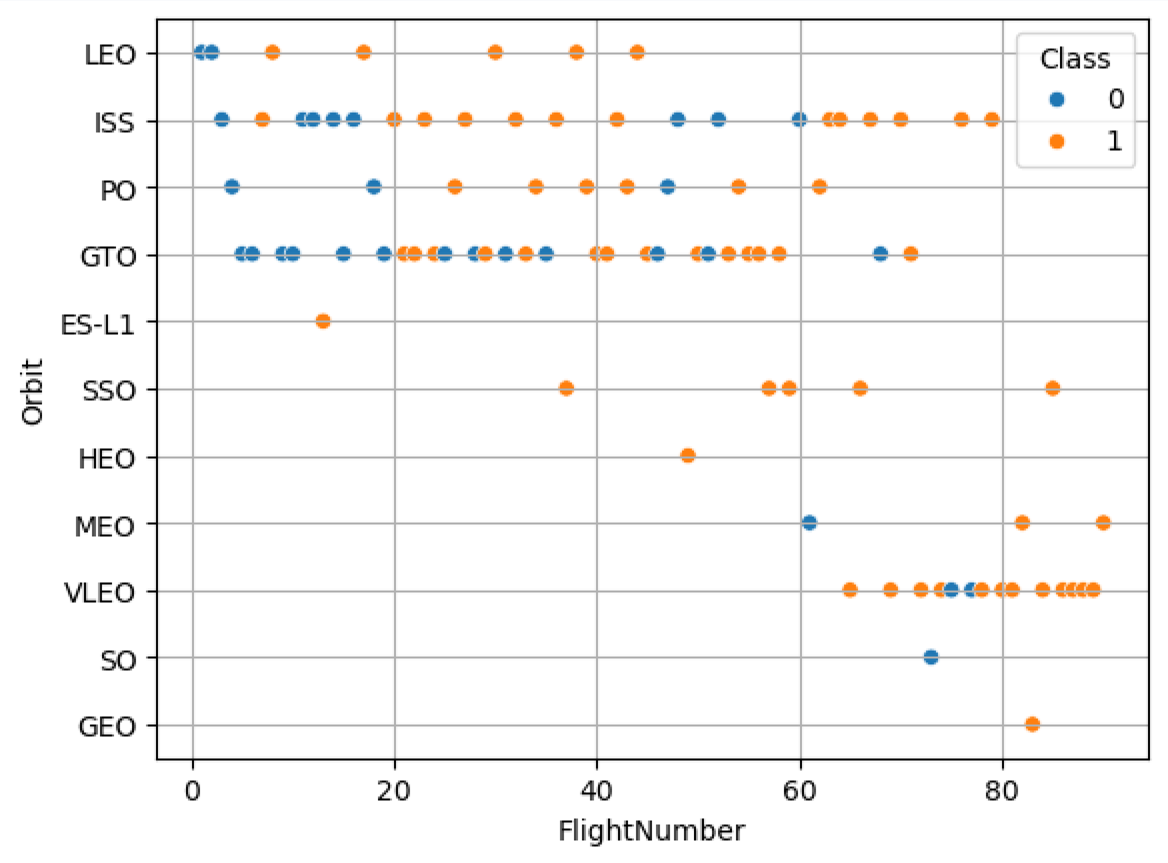
1. LEO (Low Earth Orbit): There is a higher success rate for heavier payloads in LEO. This indicates that SpaceX has a reliable track record for launching heavier payloads to LEO.

2. Polar: Similar to LEO, Polar orbits also show a higher success rate for heavier payloads.

3. ISS (International Space Station): The success rate for ISS orbits is also higher for heavier payloads.

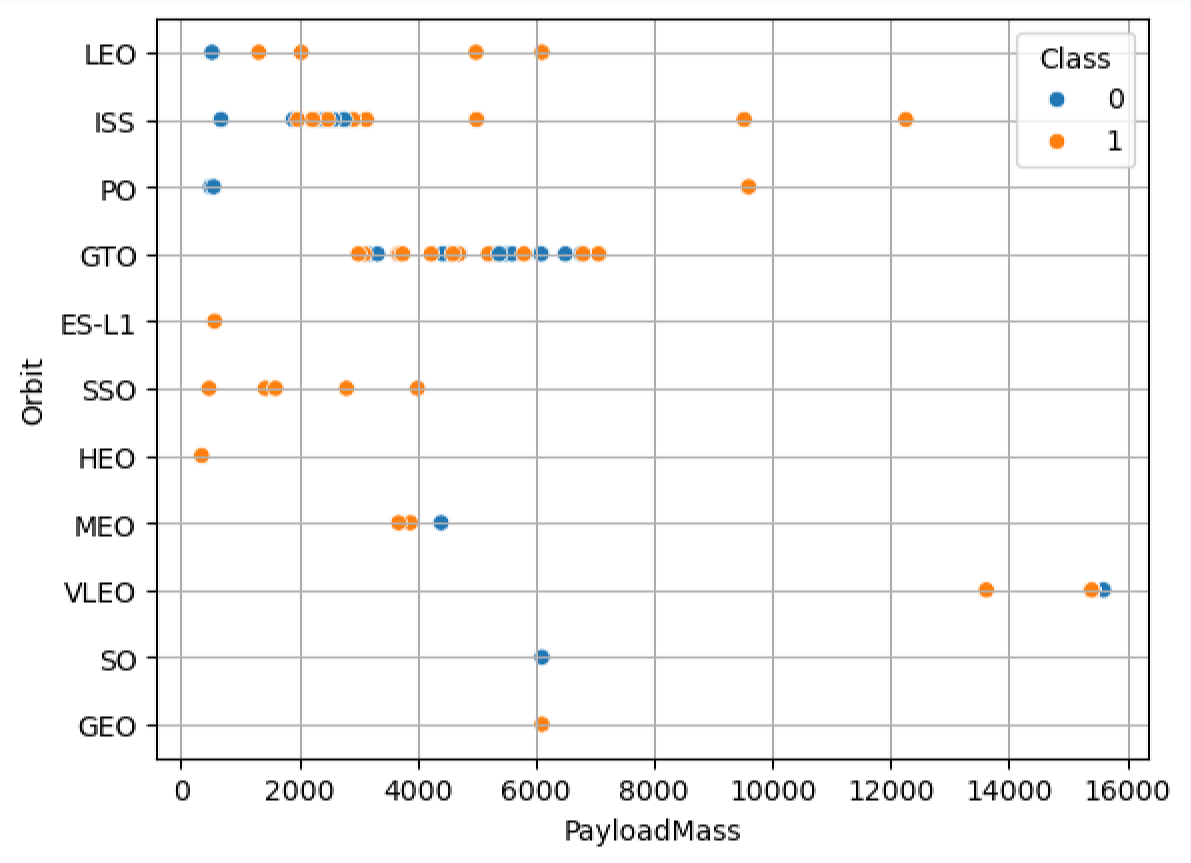
4. GTO (Geostationary Transfer Orbit): For GTO, it is difficult to distinguish between successful and unsuccessful landings as both outcomes are present for various payload masses.

Overall, it appears that for most orbits, heavier payloads tend to have a higher success rate, except for GTO where the success rate is not clearly related to the payload mass.



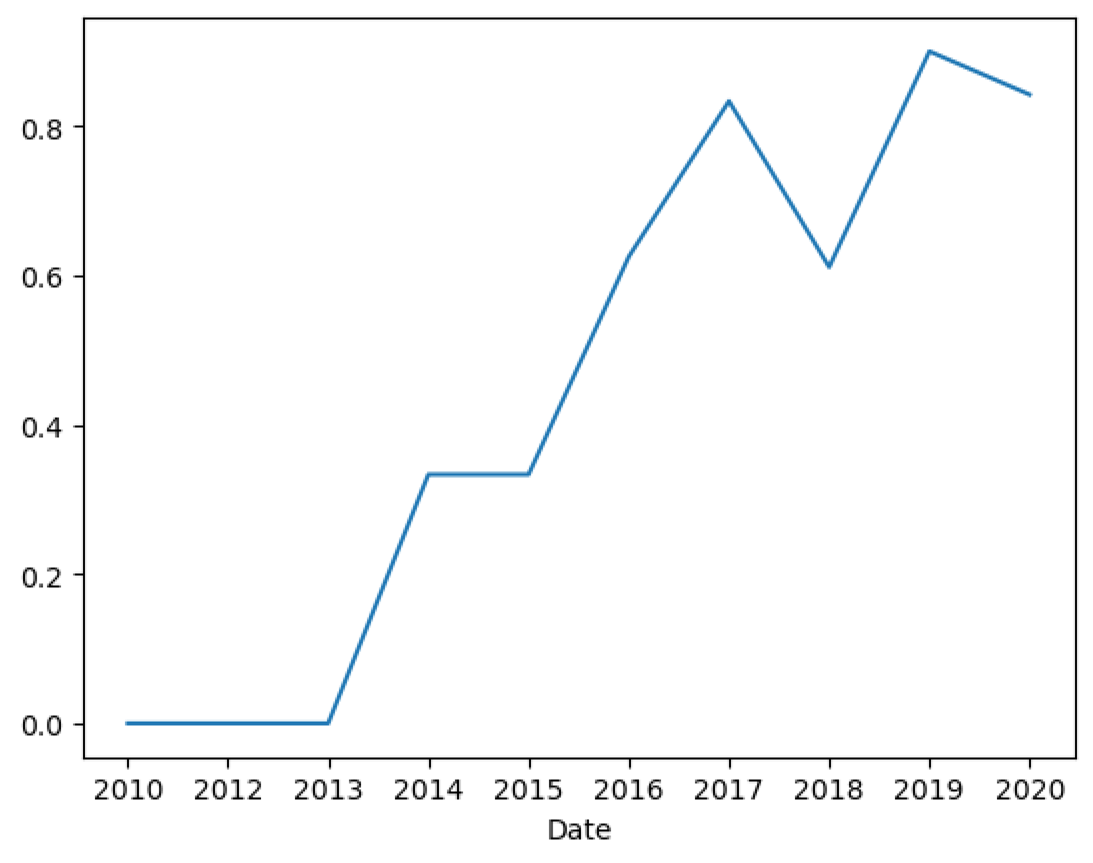
Payload vs. Orbit Type

With heavier payloads, the ISS, Polar and LEO orbits have more successful rate than the others.



Launch Success Yearly Trend

Successful rate has been increasing since 2013 until 2020



All Launch Site Names

Using this SQL query we can find the unique launch sites

```sql select DISTINCT "Launch_Site" from SPACEXTABLE```

Query result:

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40
```

Launch Site Names Begin with 'CCA'

With the following query we can find the 5 first records whose Launch Sites begin with "CCA"

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

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

Total Payload Mass

With the following query, we can calculate the toal mass carried by NASA

```
sql select sum("PAYLOAD_MASS__KG_") as 'Total Mass Carried by NASA' from
spacextable where "Customer" = 'NASA (CRS)'
```

Total Mass Carried by NASA
45596

Average Payload Mass by F9 v1.1

With the following query we can calculate the average payload mass carried by F9 v1.1

sql select avg("PAYLOAD_MASS__KG_") as 'Average Payload Mass Carried by F9 v1.1 booster' from spacextable where "Booster_Version" = 'F9 v1.1'

Average Payload Mass Carried by F9 v1.1 booster
2928.4

First Successful Ground Landing Date

With the following query, we can see the date of the first successful ground landing

sql select min(Date) as 'Date of first spacex launch' from spacextable where "Landing_Outcome" like '%Success%'

Date of first spacex launch
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- With the following query, we can see the successful drone ship landing which had a payload mass between 4000 and 6000 kg

```
sql select "Booster_Version", "PAYLOAD_MASS__KG_" from spacextable where
"Landing_Outcome" = 'Success (drone ship)' and "PAYLOAD_MASS__KG_" > 4000
and "PAYLOAD_MASS__KG_" < 6000
```

Booster_Version	PAYLOAD_MASS__KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

With the following query we can take a look at the successful and failed landings and the occurrences of each one.

sql select case when "Landing_Outcome" like 'Success%' then 'Success' when "Landing_Outcome" like 'Failure%' then 'Failure' end as 'Landing_Class', count(*) as ocorrencias from spacextable group by Landing_Class

Landing_Class	ocorrencias
None	30
Failure	10
Success	61

Boosters Carried Maximum Payload

With the following query, we can visualize every booster version that carried the maximum payload mass

```
sql select "Booster_Version" from spacextable where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from spacextable)
```

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

- With the following query, we can list the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

sql select substr(Date, 6, 2) as Month, "Landing_Outcome", "Booster_Version", "Launch_Site" from spacextable where "Landing_Outcome" like 'Failure (drone ship)' and substr(Date, 0, 5) = '2015'

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

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

- With the following query, we can 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

sql select "Landing_Outcome", count(*) as Count from spacextable where Date between '2010-06-04' and '2017-03-20' group by "Landing_Outcome" order by Count desc

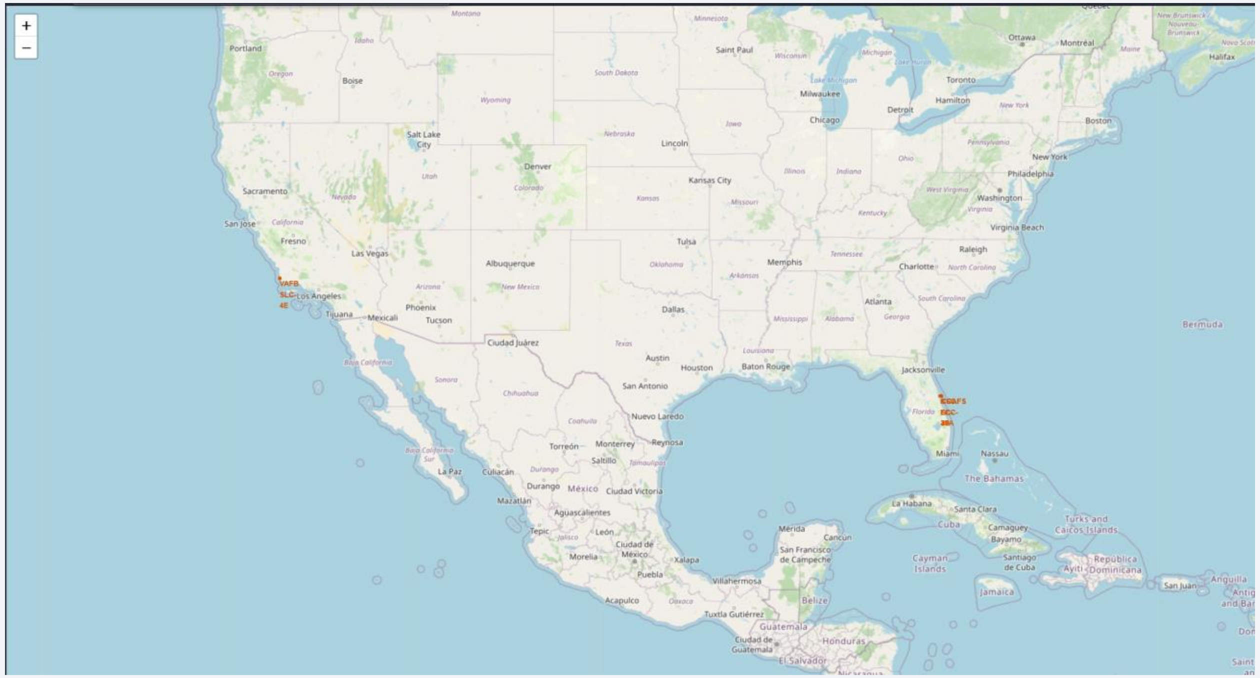
This query extracts and orders the Landing Outcomes by the count of ocurrences.

Section 3

Launch Sites Proximities Analysis

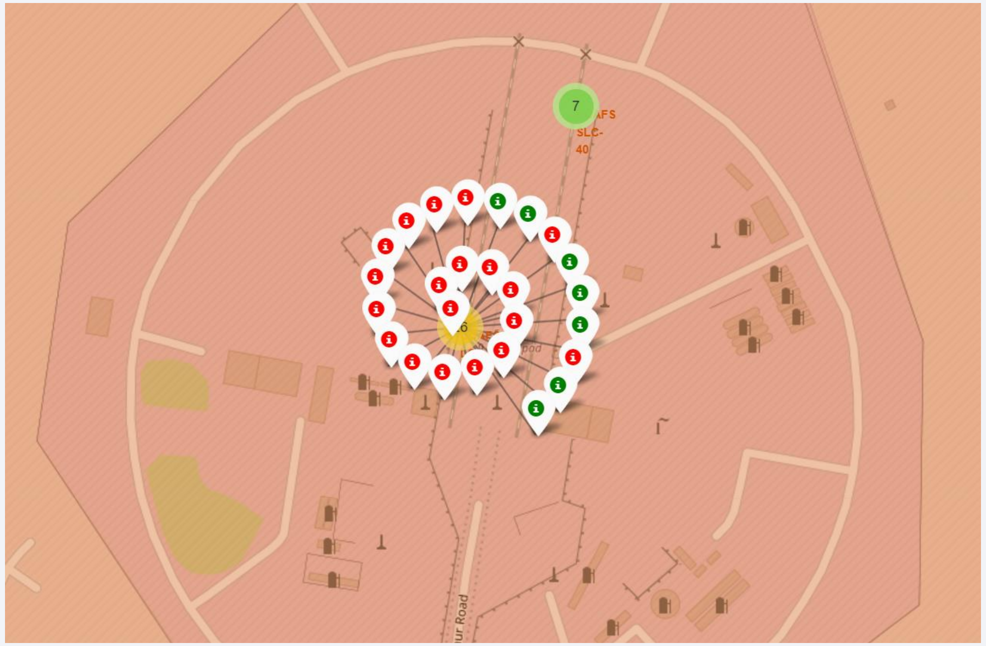
Launch Sites location

In the following image, we can visualize on the Map, the location of each launch site. Its Important to note that each launch site lays On each extrimity of United States.
Each base is strategically placed to optimize Mission security and efficiency.



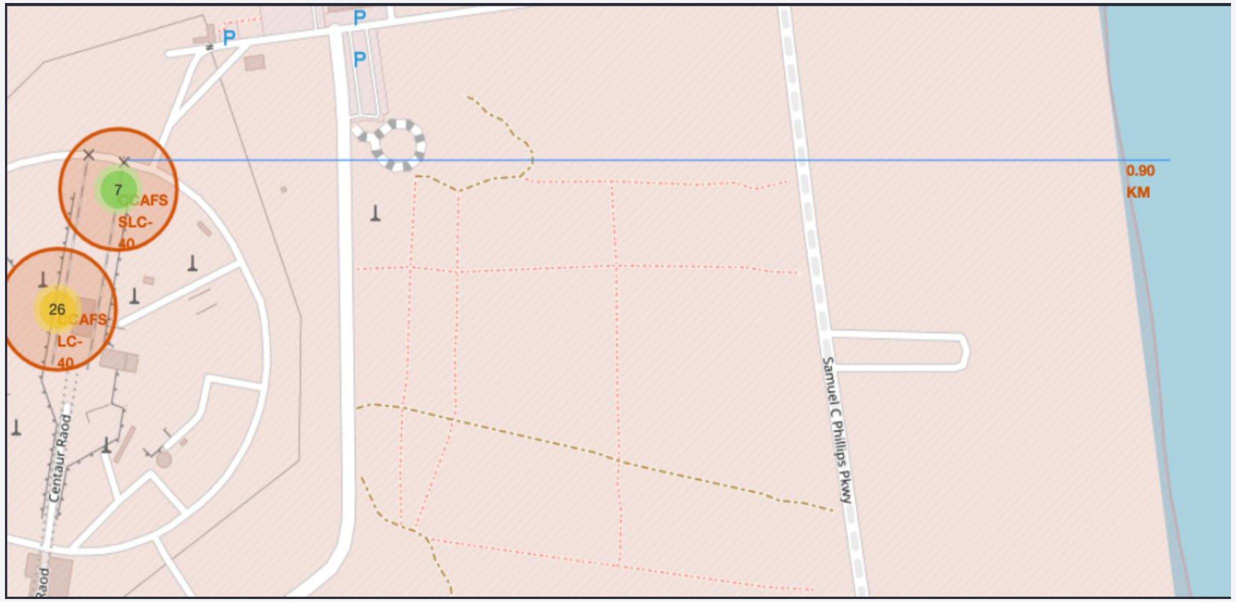
Launch Outcomes

In this map we can navigate trough the launch sites and visualize the successful and failed launches that came from each one of them.



Launch Site proximity to coastline

In the following image we can see the distance from the launch site to the coastline

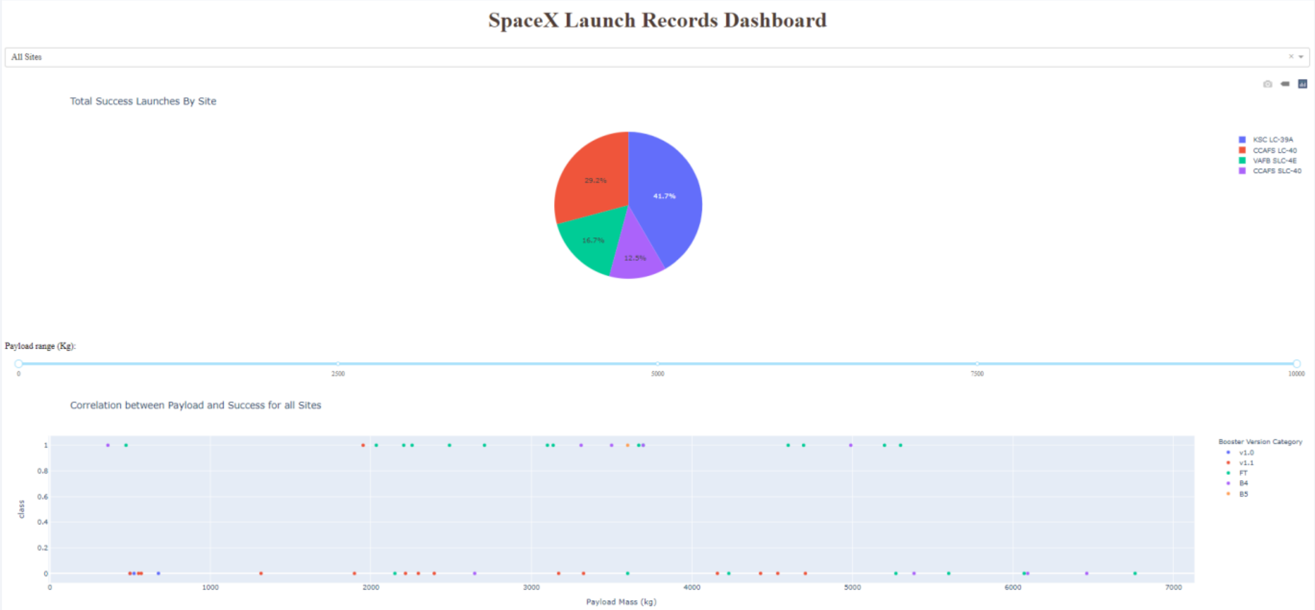


Section 4

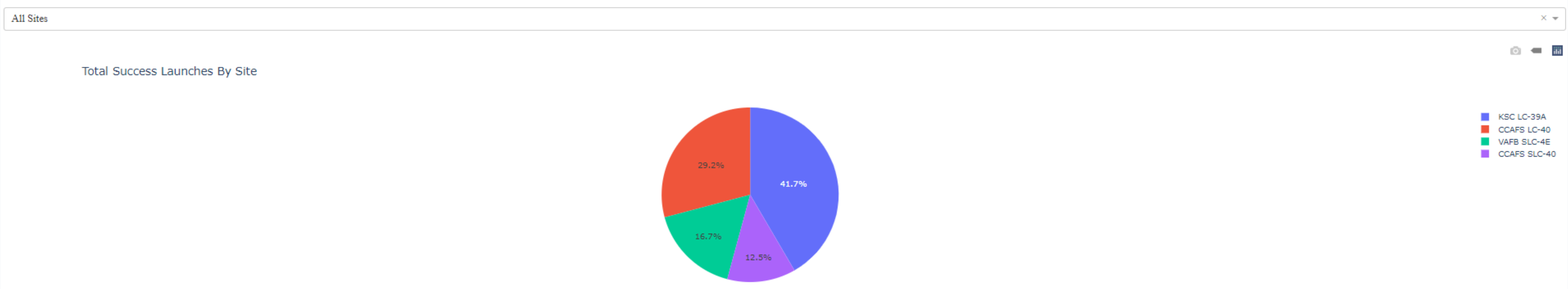
Build a Dashboard with Plotly Dash

Successful launches by launch site

In the image on the right we can see the successful launches count on each launch site represented by a pie chart

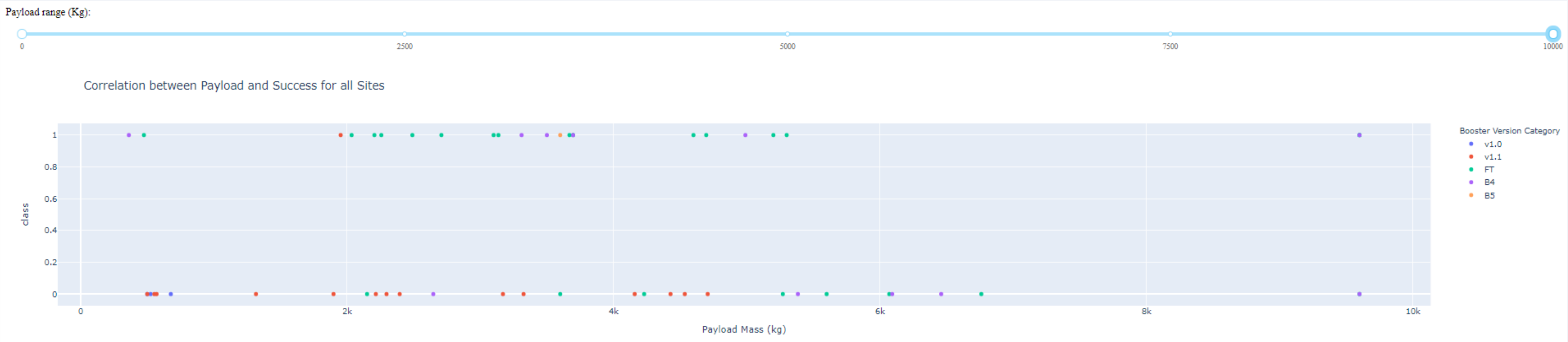


Most successful Launch Site



In the pie chart we can see the launch sites and visualize the most successful launch site

Payload Mass vs Launch Outcome



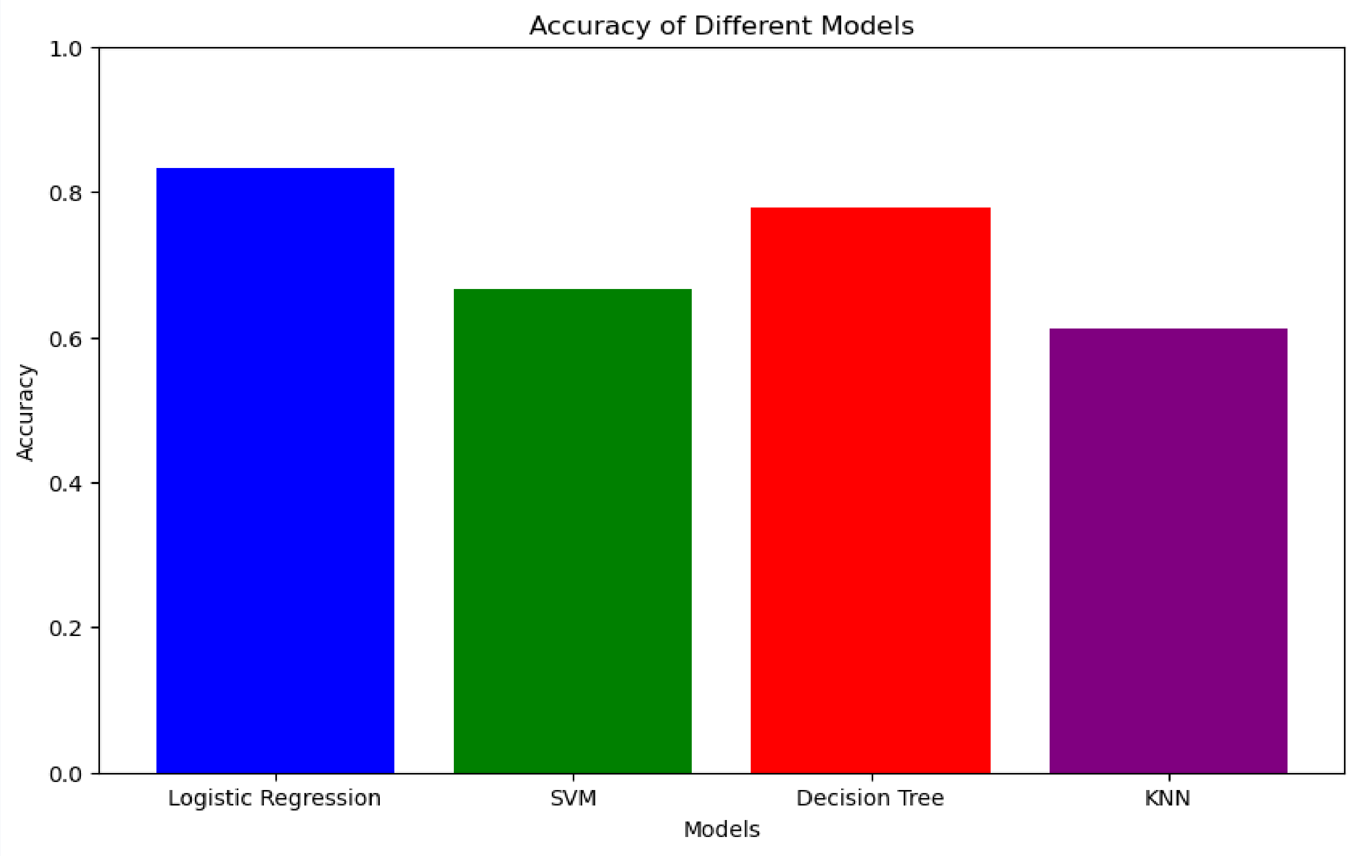
In the scatter plot, we can see the relation between payload mass and launch outcome

Section 5

Predictive Analysis (Classification)

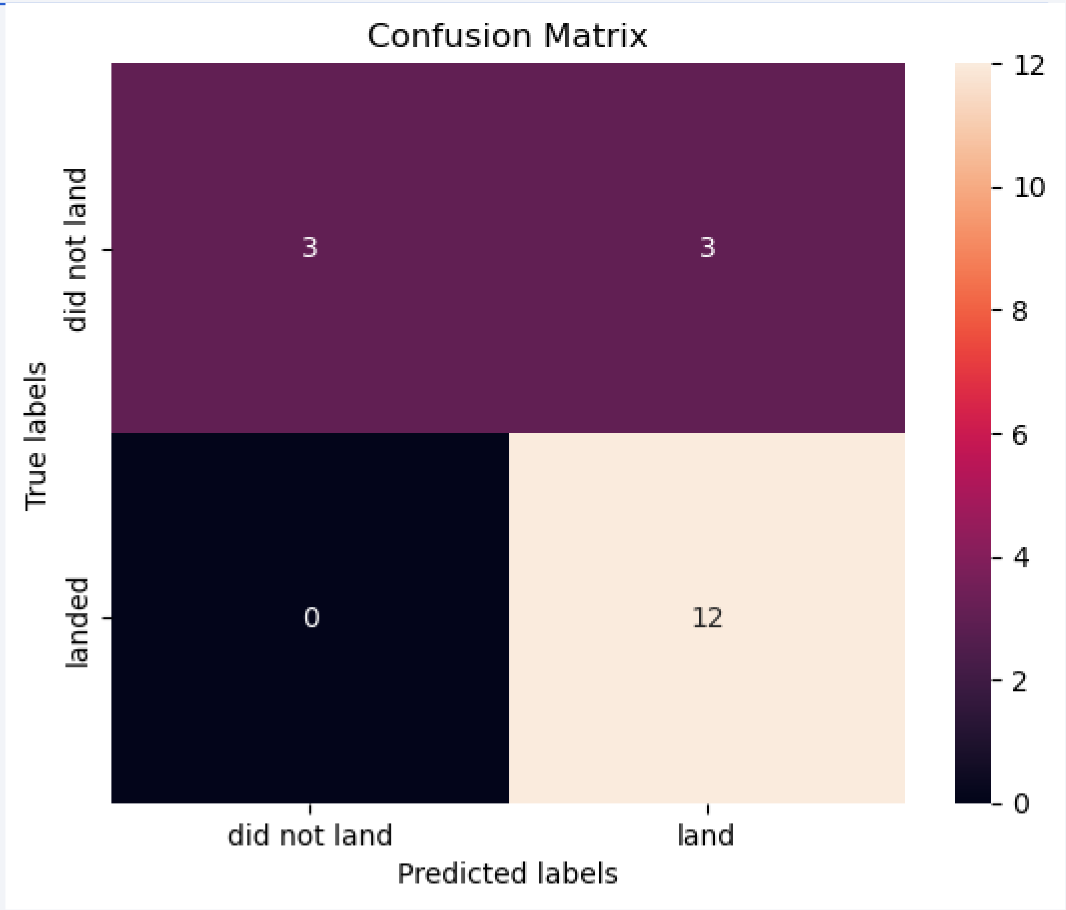
Classification Accuracy

In the bar chart we can see that the best performing model is Logistic regression



Confusion Matrix

Logistic Regression's confusion matrix



Conclusions

- We can see that the overall success rate of launches has been increased due to advancements in technology
- We can observe that as successful launches became more and more frequent, spacex developed a growing confidence and started sending heavier payloads
- Each Launch site is strategically placed on the map

Thank you!