EXECUTIVE SUMMARY: PREDICT THE REUSE OF FIRST STAGE BY SPACEX

- Project Overview: To predict if SpaceX will reuse the first stage, we aimed to use a trained machine learning model and public information to determine the cost of a launch.
  - Sub Point 1
  - Sub Point 2

- Approach:
  - Data collection and Wranglinfd: We collected data using spaceX API and webscraping
  - Exploratory Data Analysis: we conducted EDA to identify patterns and key features within the data
  - Predictive Analysis: We engineered new features, and evaluated various machine learning algorithms including logistic regression, support vector machine(SVM), Decision tree, KNN.

- Results:
  - EDA: Launch success improved over time, KSC LC-39A had the most successful landing rate and Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate
  - Visualization: Most Lauch sites were closer to the equator and to the coast
  - Predictive Analysis:: All models performed well on the test set.

Introduction

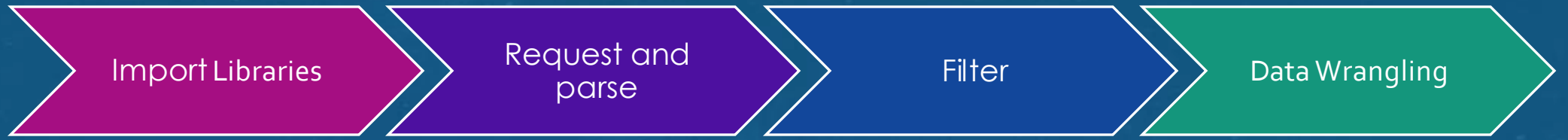
SpaceX, led by Elon Musk, has revolutionized aerospace with reusable rocket technology, notably reusing Falcon 9 first stages. Data science plays a pivotal role in predicting first stage reusability by analyzing historical data, telemetry, and machine learning. This predictive capability impacts space industry finances and environmental sustainability.

In this study, we embark on a journey to harness the power of machine learning models to predict whether SpaceX will achieve successful first stage recovery and reuse. By examining historical data on SpaceX's launches, mission details, and outcomes, we aim to build a predictive model that can inform stakeholders, enthusiasts, and decision-makers about the likelihood of successful reuse.

This research is not only a testament to the growing synergy between advanced data analytics and space technology but also an attempt to provide valuable insights into SpaceX's pioneering efforts in rocket recovery and reuse. As we delve into the methodology, data collection, feature engineering, and model development, we anticipate that our findings will shed light on the future of sustainable space travel and inspire further innovations in the aerospace industry.

METHODOLOGY

DATA COLLECTION - API



- Import libraries and define Auxiliary functions
- Requests
  - Pandas
  - Numpy
  - Datetime

- Request and parse SpaceX launch date using GET request
  - Request data from API
  - Decode response using .json() and convert to data frame using pd.json_normalize
  - Create dictionary
  - Create data frame from dictionary

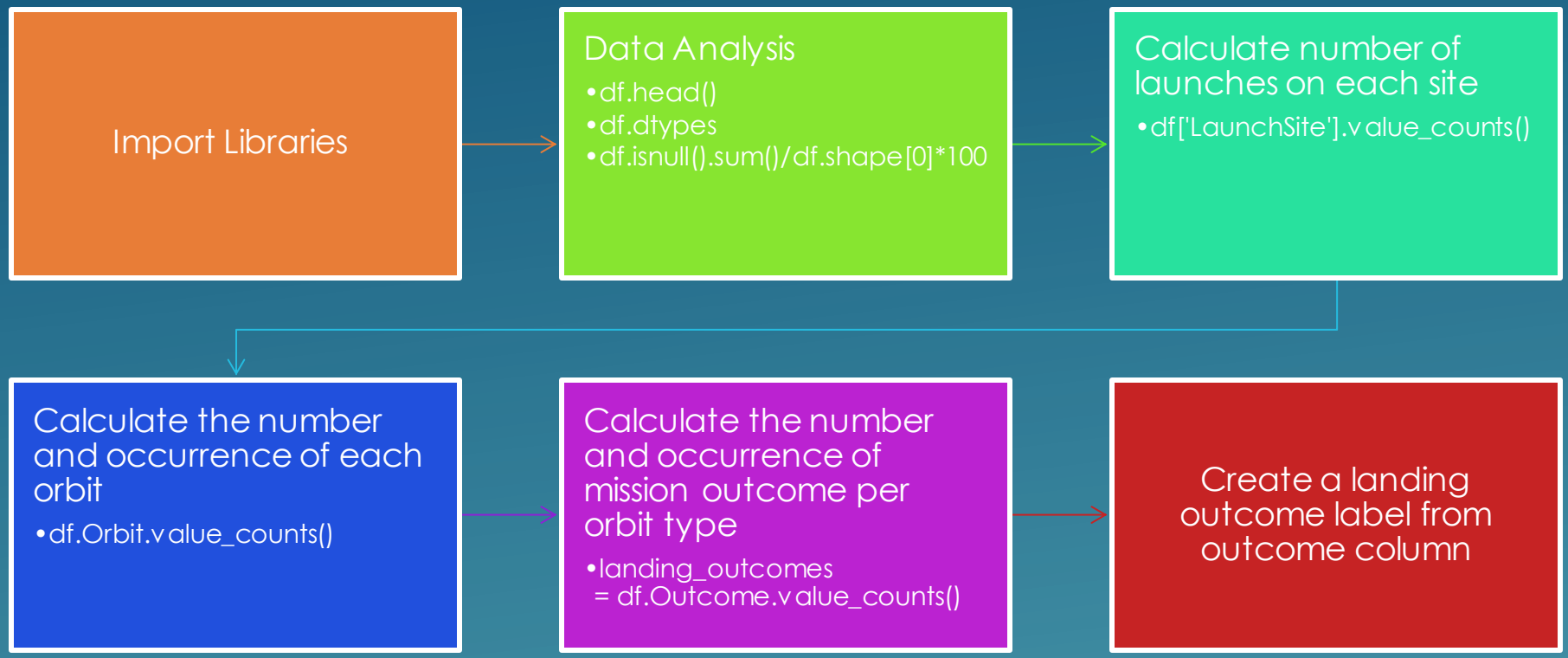
- Filter the Data frame only to include Falcon9 launches

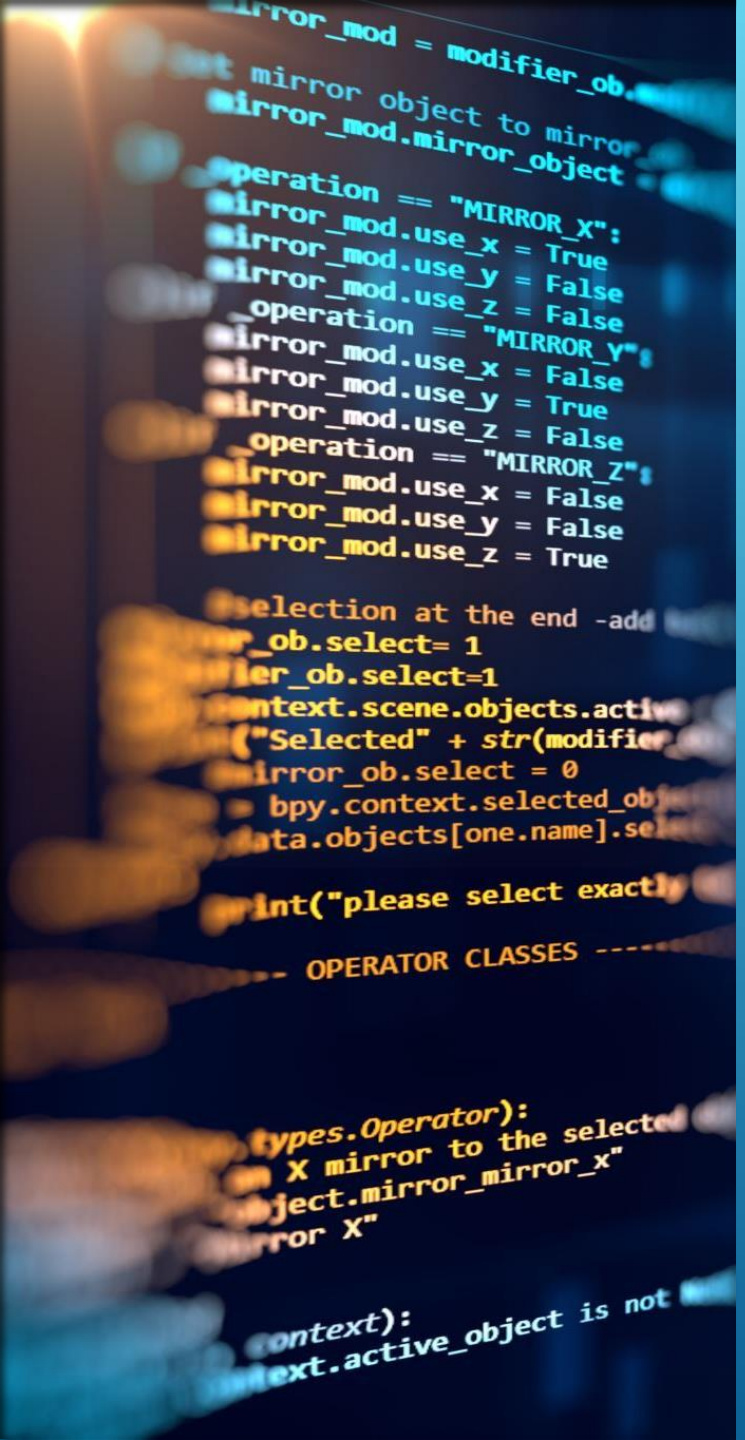
- Dealing with missing values
  - Isnull().sum()
  - .mean()
  - np.nan

DATA COLLECTION – WEBSCRAPING

Import	•Sys •BeautifulSoup •Unicodedata •Pandas
Request	•Perform HTTP GET method •Create BeautifulSoup Object from HTML response
Extract Columns	•Use find_all function •Extract column names from HTML table header
Create Data frame	•Create an empty dictionary from data •Fill in dictionary with extracted records from table rows •Create data frame from the dictionary

DATA WRANGLING





EDA WITH SQL

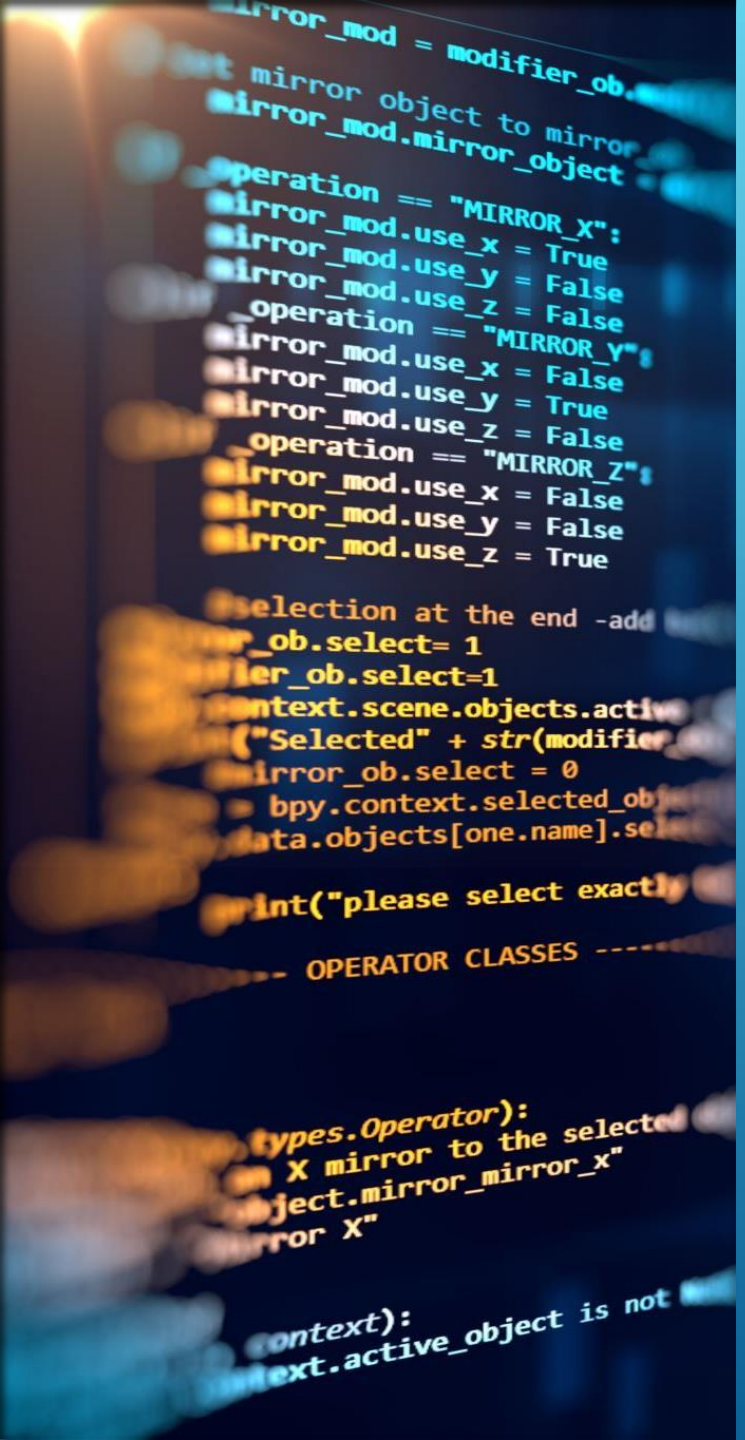
We connected to the database

Displayed

- unique launch sites
- 5 records where launch sites began with 'CCA'
- the total payload mass carried by boosters launched by NASA (CRS)
- average payload mass carried by booster version F9 v1.1

Listed

- Date of first successful landing on ground pad
- Names of boosters which had success landing on drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failed missions
- Names of booster version which have carried max payload
- Failed landing outcomes on drone ship, their booster version and launch site for the months in the year 2015
- Count of landing outcomes between 2010-06-04 and 2017-03-20 (desc)



EDA WITH VISUALIZATION

- To understand how certain features affected the launch, we visualized the relationships between
  - Flight number and Payload Mass
  - Flight number and launch site
  - Success rate of each orbit using the group by function
  - Flight number and orbit type
  - Payload and orbit type

- Visualized the launched success yearly trend
- Performed features engineering by using one hot encoding
- Casted all numeric columns to float64 using .astype()

INTERACTIVE VISUALIZATION WITH FOLIUM

Marked all launch sites

- Created a blue circle at NASA Johnson Space Centre's coordinate with a popup label
- Added red circles at all launch sites coordinates with a popup label

Coloured Markers of Launch outcome

- Added coloured markers of successful and unsuccessful launches at each launch site

Distances between launch sites and proximities

- Added lines to indicate the distances between launch sites and its proximity to the nearest coastline, railway and city.

DASHBOARD WITH PLOTLY DASH

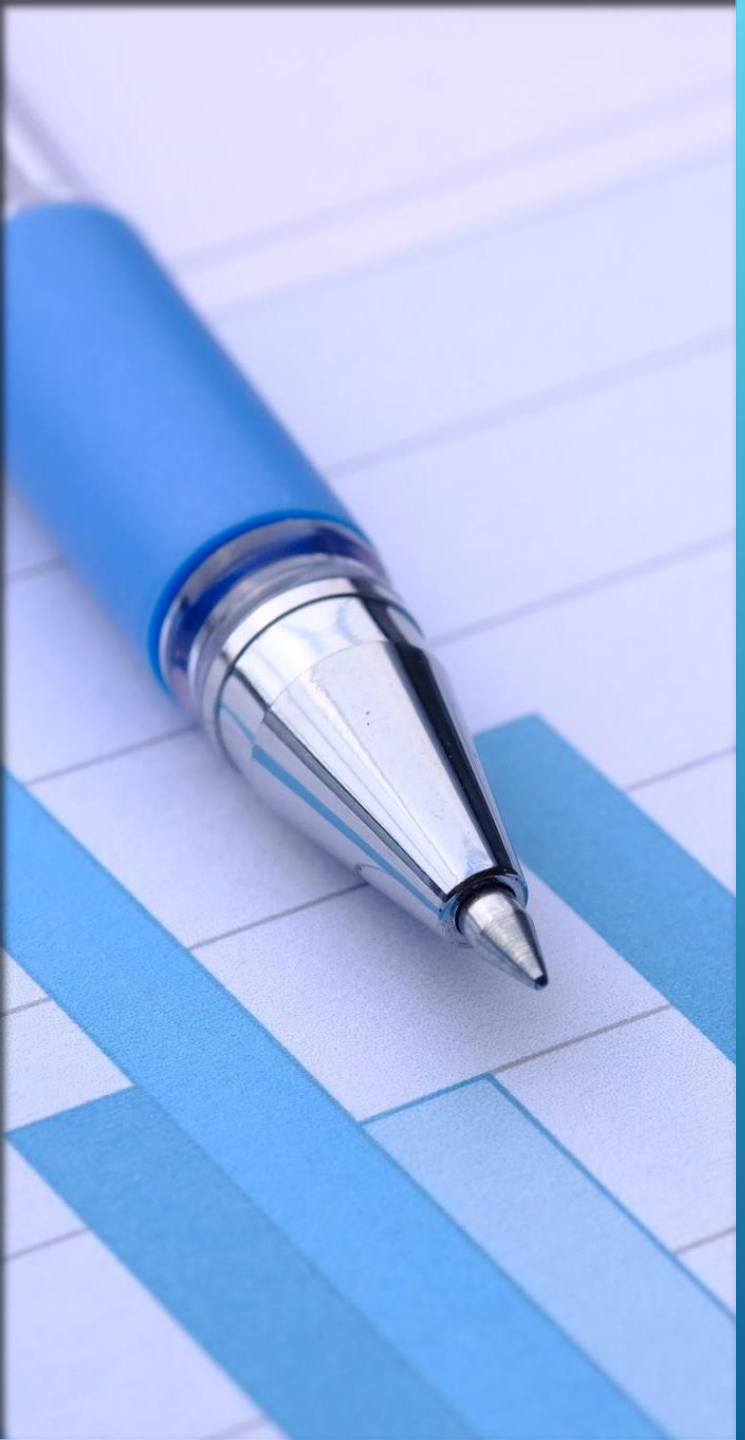
Drop down list with launches
- Allow users to select all launch sites or certain launch site
- Added red circles at all launch sites coordinates with a popup label

Pie chart showing successful launches
- Allow users to visualize successful and unsuccessful launches

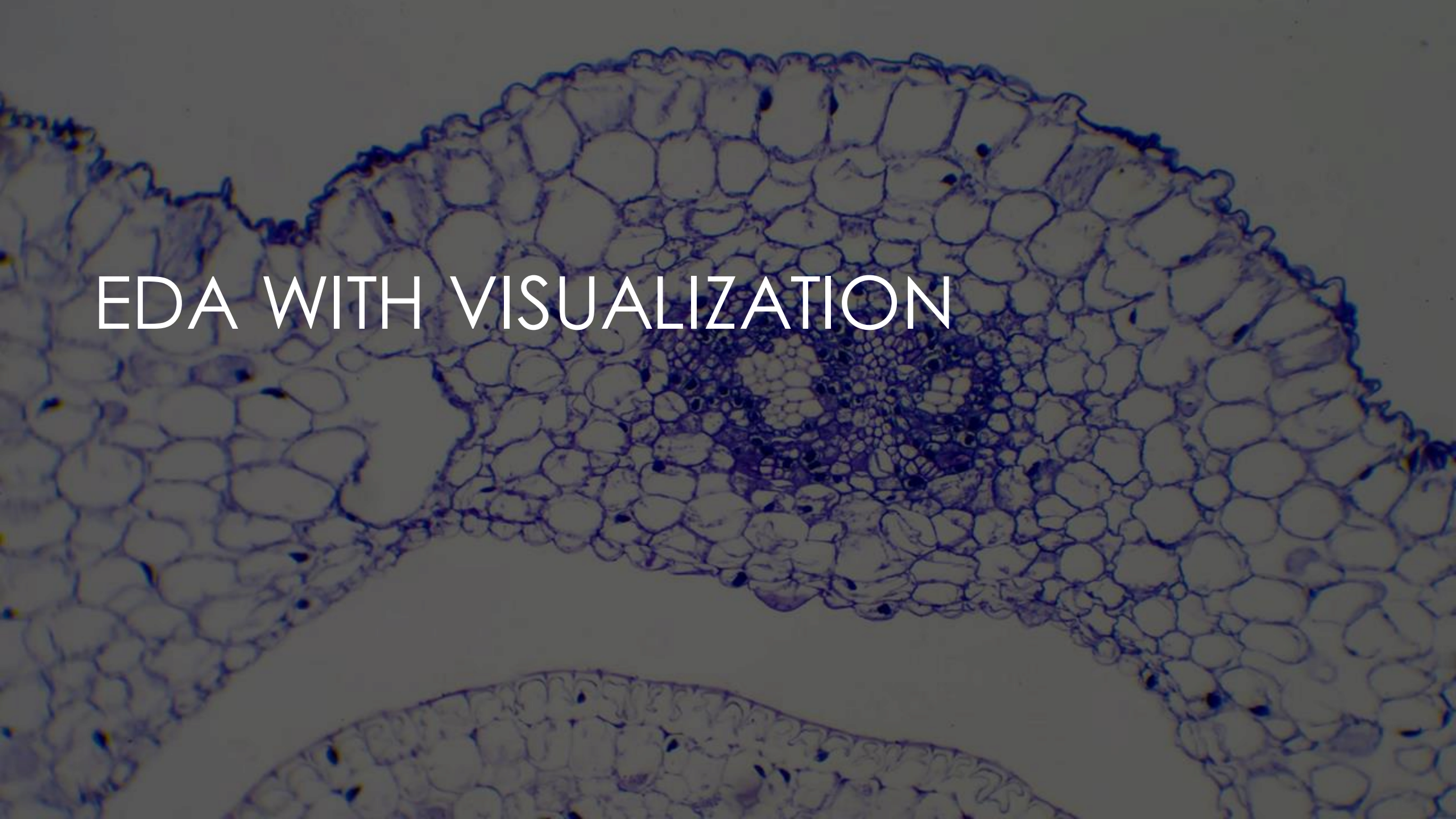
Slider of payload mass range
- Allow users to select payload mass range

Scatter chart
- Allow users to visualize the relationship between payload mass and launch success

PREDICTIVE ANALYSIS

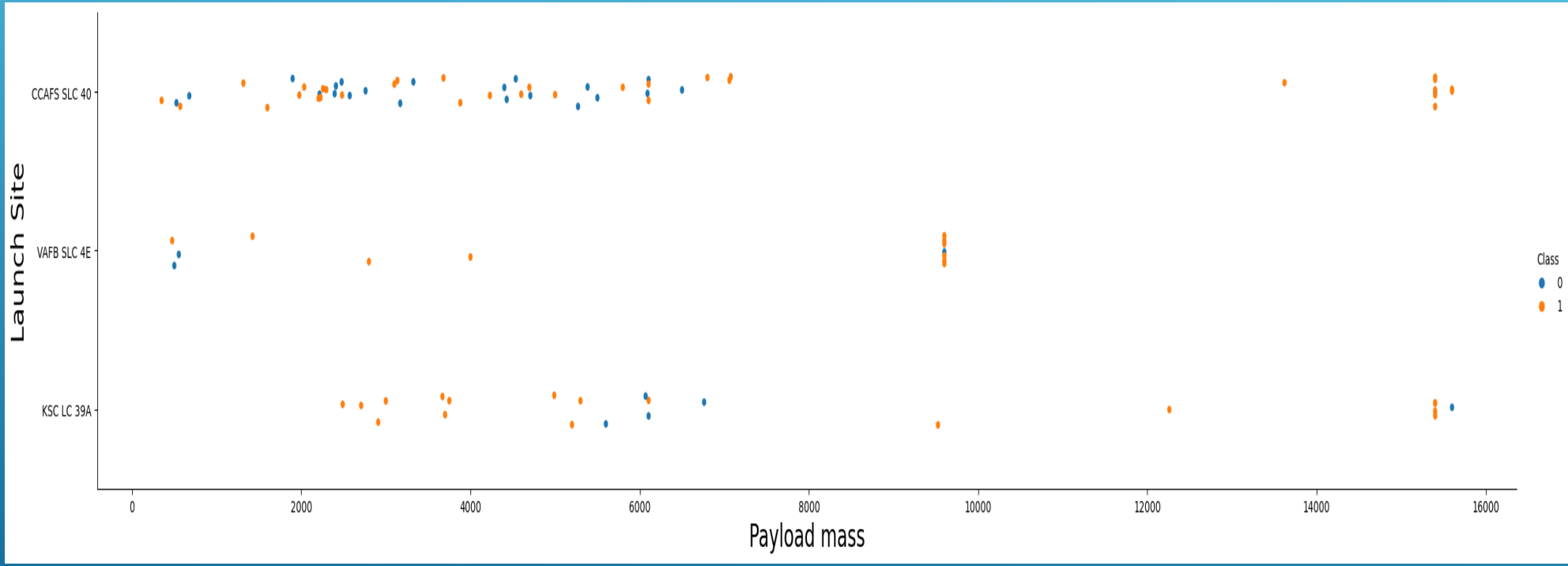


RESULTS

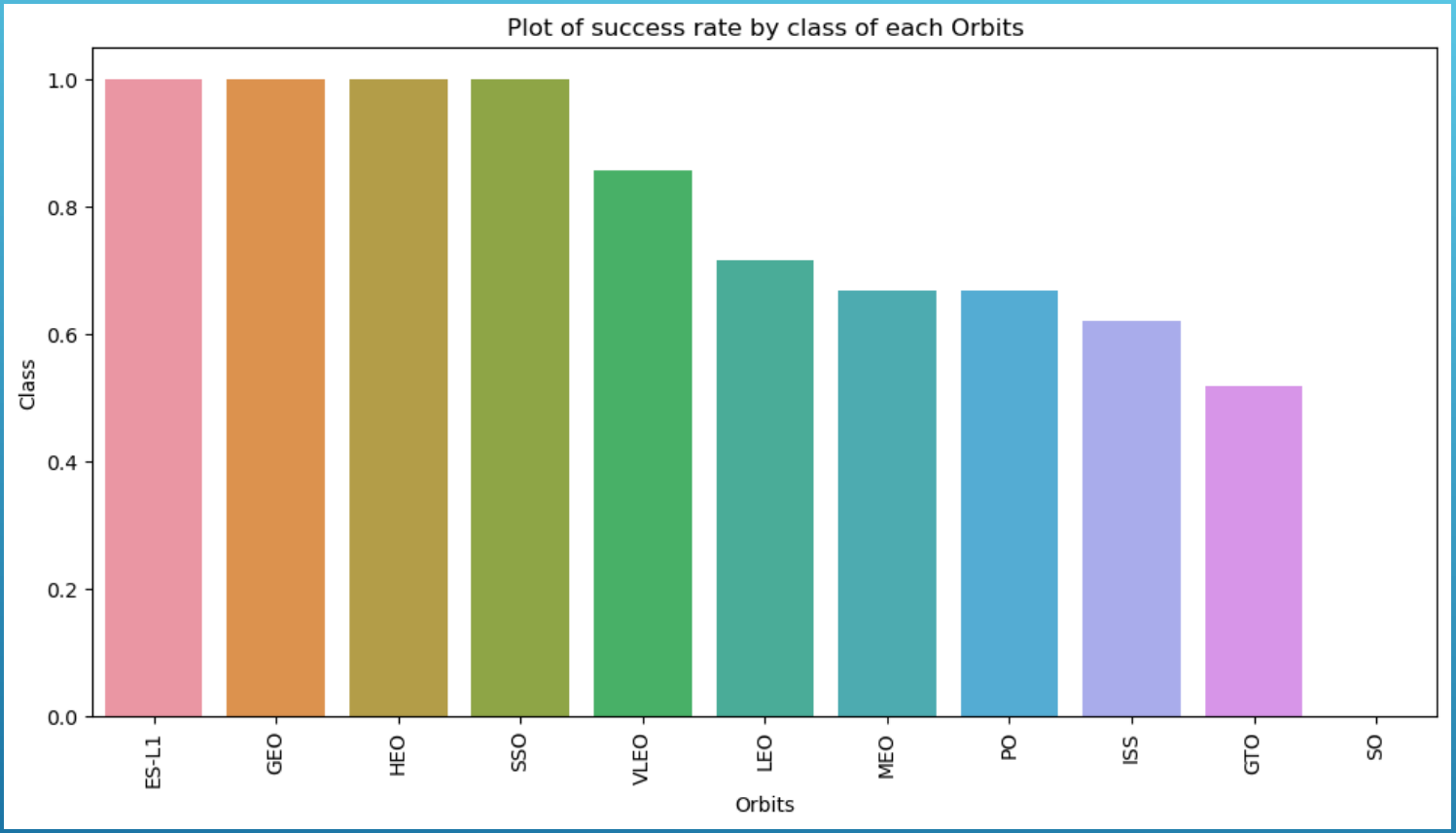


EDA WITH VISUALIZATION

KSC LC-39A and CCAFS SLC-40 had more success when the payload mass increased



EDA WITH VISUALIZATION

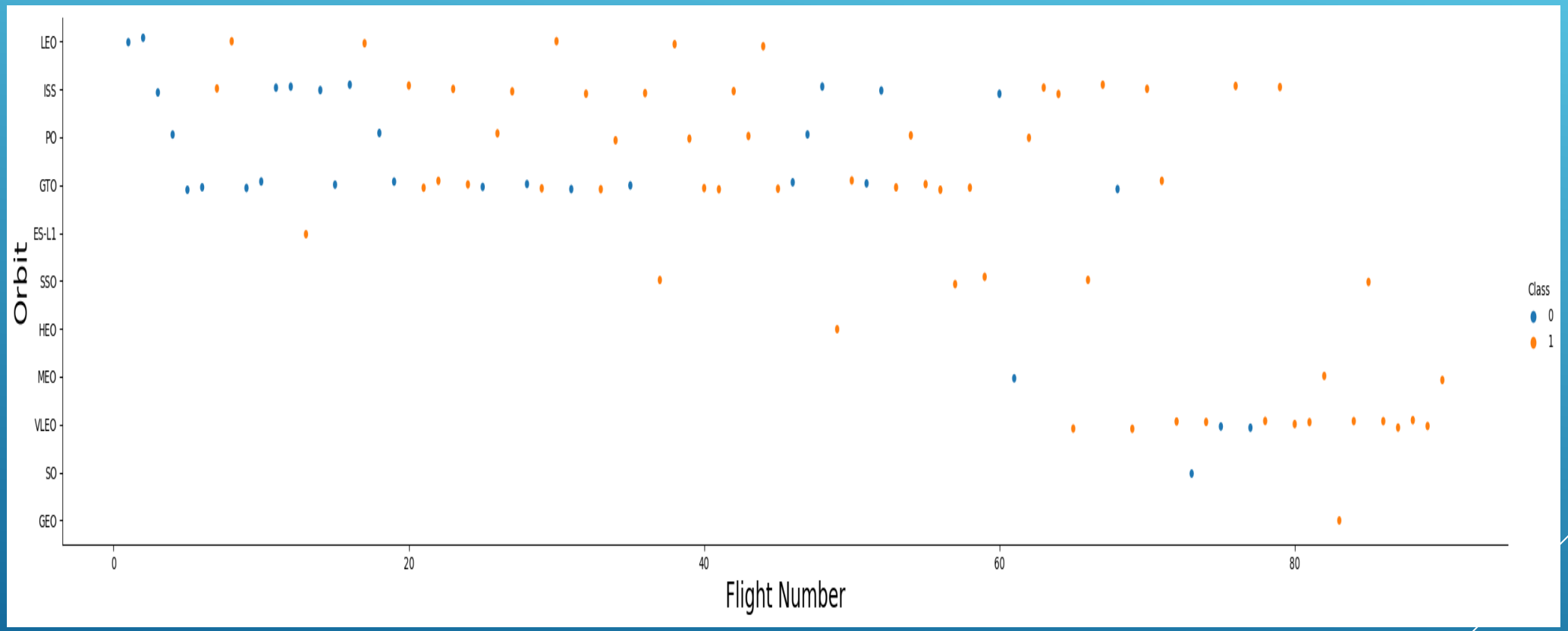


Orbits Success Rates

- ES-L1, GEO, HEO and SSO had 100% success rate
- VLEO, MEO, PO, ISS and GTO had above 50% success rate
- SO had a 0% success rate

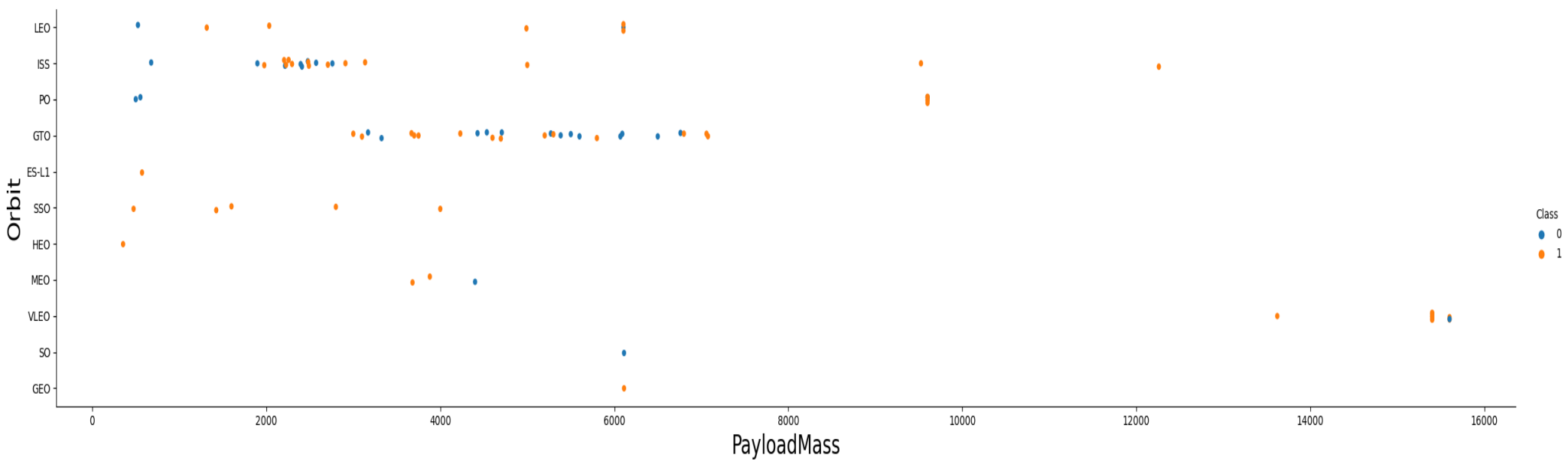
EDA WITH VISUALIZATION

Higher flight number influenced increased the success rates of some orbits while lower flight number influenced the success rates of other orbits. It can be that flight number is not a factor that influenced success rates in general.

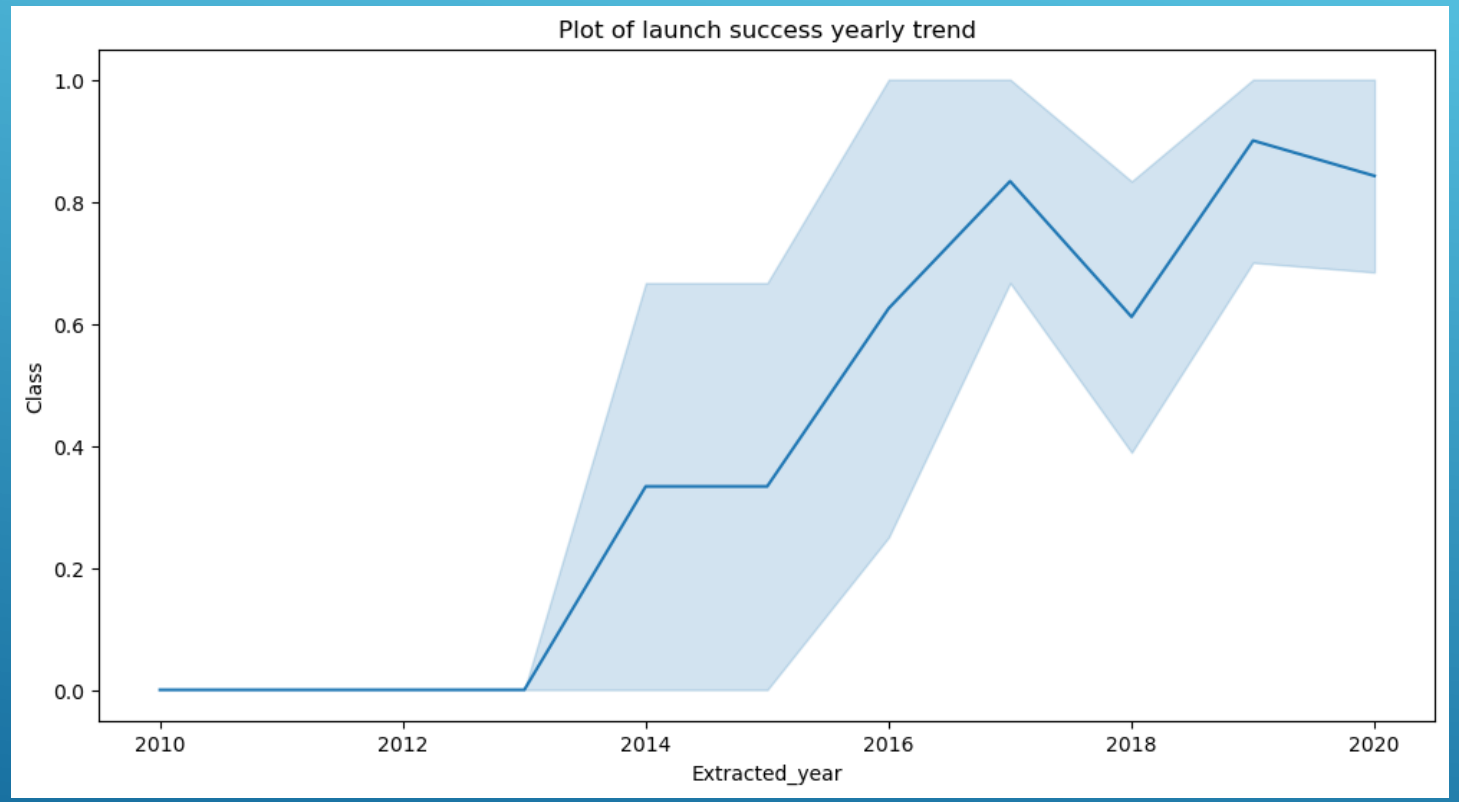


EDA WITH VISUALIZATION

Orbits had more success rate when the payload mass was between 2000 and 6000 kg.

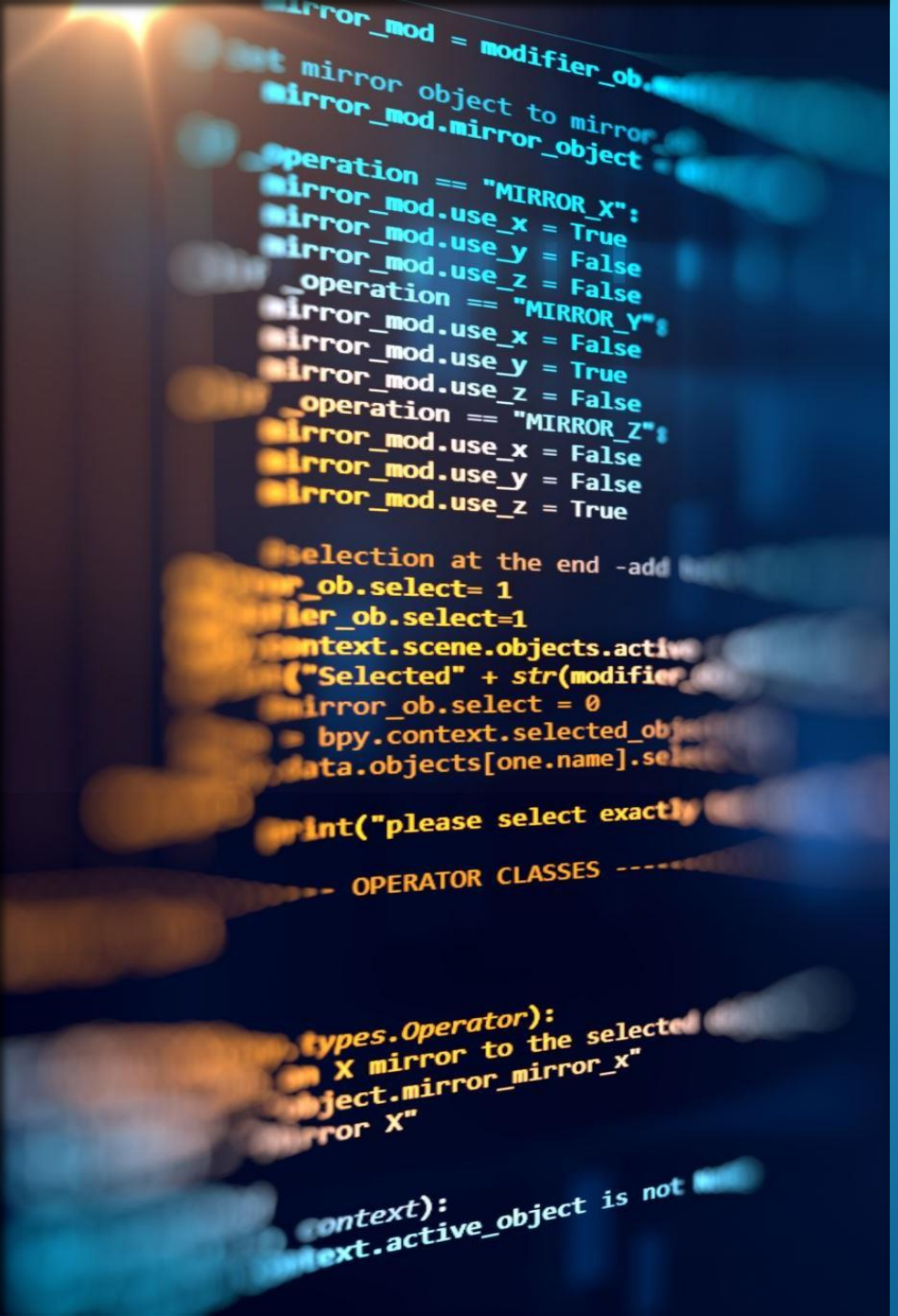


EDA WITH VISUALIZATION



There was an increase in successful launches as the year increases. This good be due to advancement in technology and infrastructure.

EDA WITH VISUALIZATION



EDA RESULTS WITH SQL

EDA RESULTS WITH SQL

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

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

- The launch sites used in this research were
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

TotalPayloadMass	Customer
48213	NASA (CRS)

AveragePayloadMass
2534.6666666666665

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

Booster_Version	Landing_Outcome	PAYLOAD_MASS_KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

- Total payload mass was 48, 213kg carried by boosters launched by NASA (CRS)
- Average payload mass was 2534.67kg carries by booster version F9 v1.1
- First successful landing on ground pad was 22-12-2015
- Boosters that had successful outcome on drone ship with payload mass greater that 4000 but less than 6000 are
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

EDA RESULTS WITH SQL

EDA
RESULTS WITH SQL

- There were 1 failure(in-flight), 99 successful outcomes and 1 success (unclear payload status)

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

[7]:

	month	Date	Booster_Version	Launch_Site	Landing_Outcome
5-		2015-10-01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
5-		2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- There were two failure outcomes in drone ship in 2015 with boosters
  - F9 v1.1 B1012
  - F9 v1.1 B1015

EDA RESULTS WITH SQL

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

Booster Version carrying maximum payload mass

INTERACTIVE MAP WITH FOLIUM

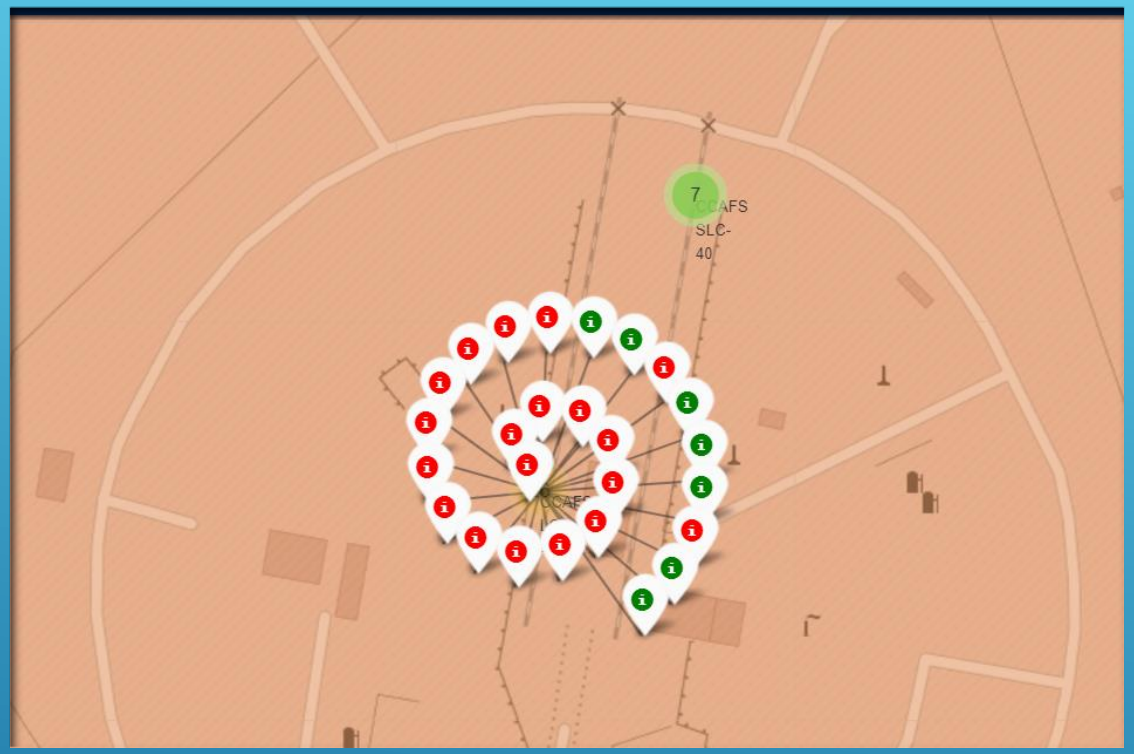
INTERACTIVE MAP WITH FOLIUM

	Launch Site	Lat	Long	class
46	KSC LC-39A	28.573255	-80.646895	1
47	KSC LC-39A	28.573255	-80.646895	1
48	KSC LC-39A	28.573255	-80.646895	1
49	CCAFS SLC-40	28.563197	-80.576820	1
50	CCAFS SLC-40	28.563197	-80.576820	1
51	CCAFS SLC-40	28.563197	-80.576820	0
52	CCAFS SLC-40	28.563197	-80.576820	0
53	CCAFS SLC-40	28.563197	-80.576820	0
54	CCAFS SLC-40	28.563197	-80.576820	1
55	CCAFS SLC-40	28.563197	-80.576820	0

	Launch Site	Lat	Long	class	marker_color
46	KSC LC-39A	28.573255	-80.646895	1	green
47	KSC LC-39A	28.573255	-80.646895	1	green
48	KSC LC-39A	28.573255	-80.646895	1	green
49	CCAFS SLC-40	28.563197	-80.576820	1	green
50	CCAFS SLC-40	28.563197	-80.576820	1	green
51	CCAFS SLC-40	28.563197	-80.576820	0	red
52	CCAFS SLC-40	28.563197	-80.576820	0	red
53	CCAFS SLC-40	28.563197	-80.576820	0	red
54	CCAFS SLC-40	28.563197	-80.576820	1	green
55	CCAFS SLC-40	28.563197	-80.576820	0	red

INTERACTIVE MAP WITH FOLIUM

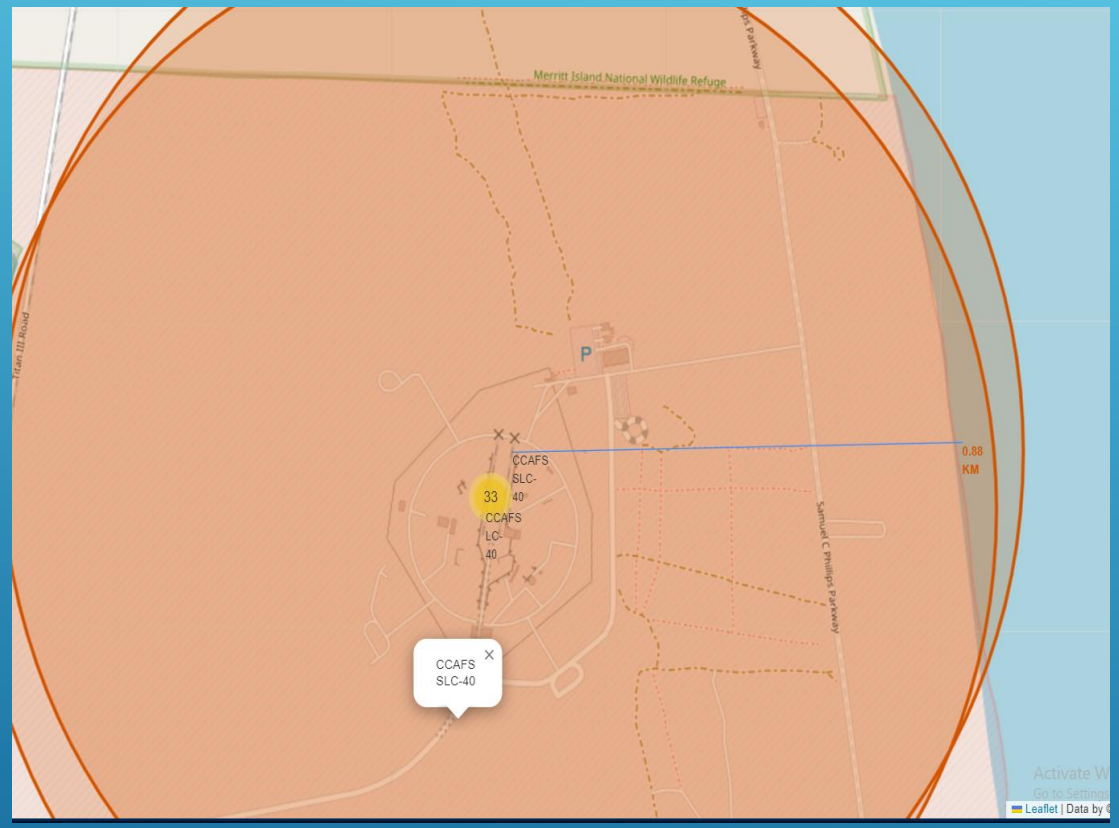
- The closer the launch site is to the equator, the easier it is to launch to equatorial orbit.



- Green markers indicate a successful launch
- Red marker indicate that the launch was unsuccessful
- Launch site CCAFS SLC-40 had low success rates

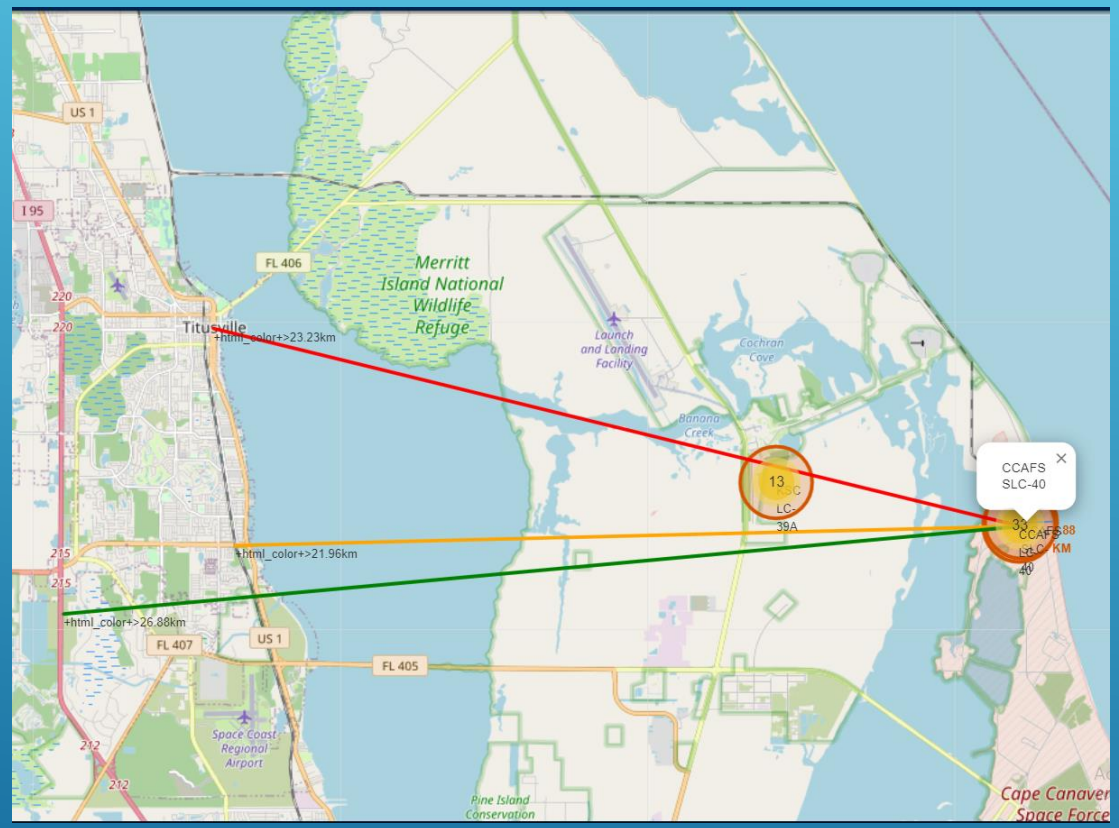
INTERACTIVE MAP WITH FOLIUM

INTERACTIVE MAP WITH FOLIUM



The distance from the launch site CCAFS-SLC-40 to the nearest coastline is 0.88km.

INTERACTIVE MAP WITH FOLIUM

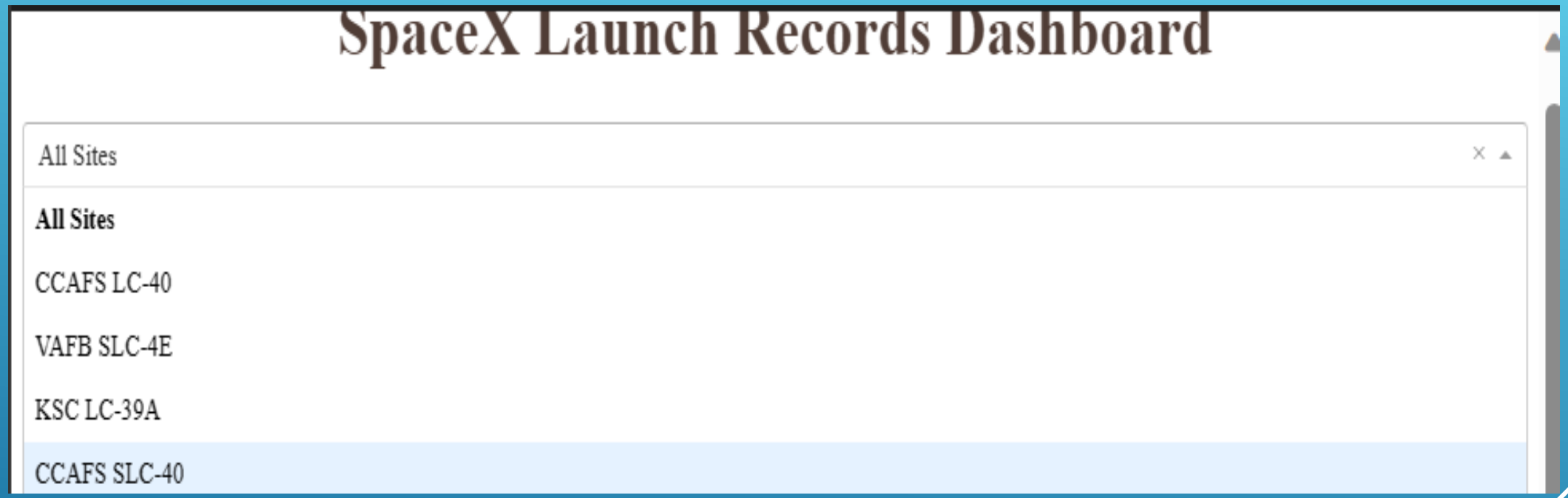


- CCAFS SLC-40
  - .88KM to nearest coastline
  - 21.96km to nearest railway
  - 23.3km to nearest city
  - 26.88km to nearest railway

Overall, the launch sites are closest to coastlines but farther away from cities, railways and highways.

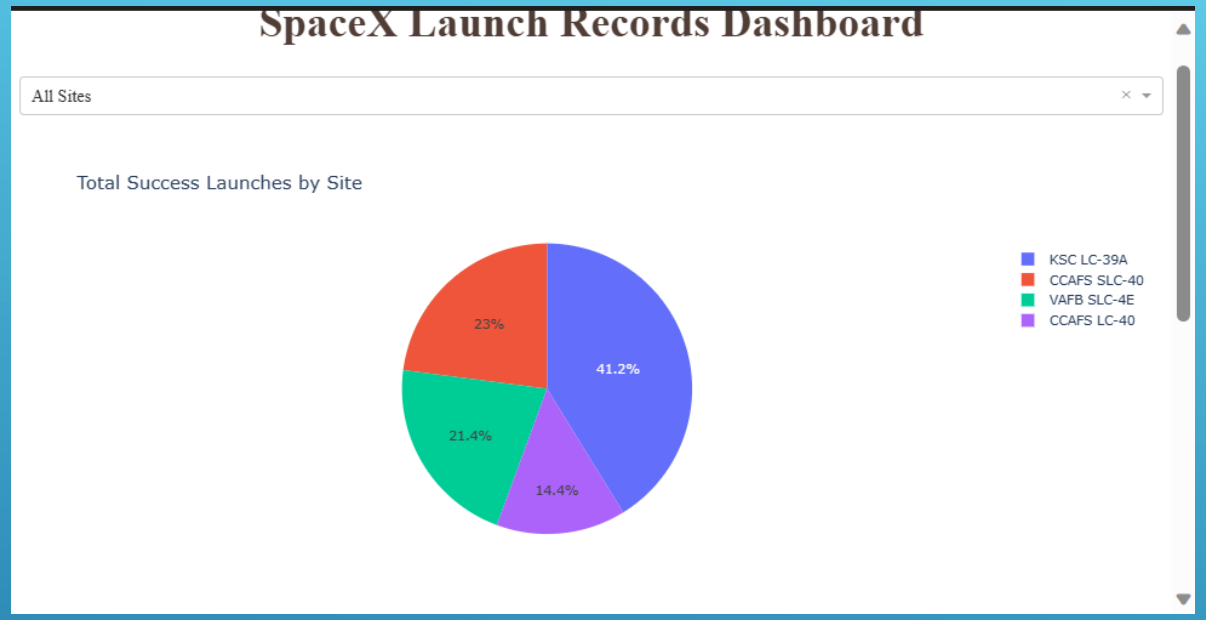
PLOTLY DASHBOARD

PLOTLY DASHBOARD RESULTS



A drop-down menu showing an option of whether to look at all site or individual site.

PLOTLY DASHBOARD RESULTS

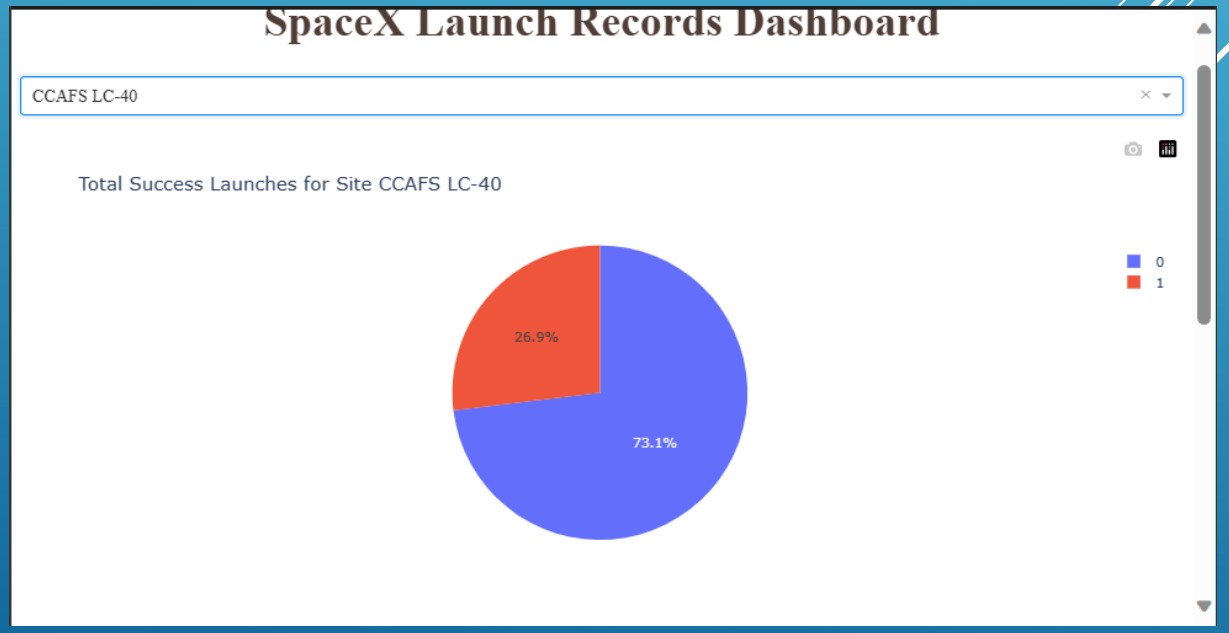


KSC LC-39A had the most successful launch having a 41.2%

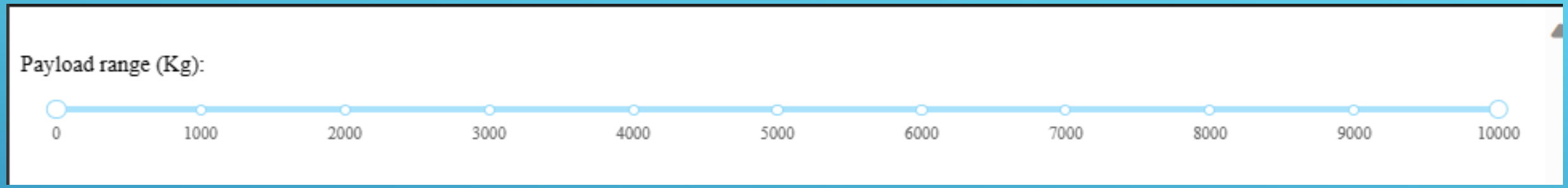
KSC LC-39A had the highest percentage of successful launches (76.9)

10 successful launches

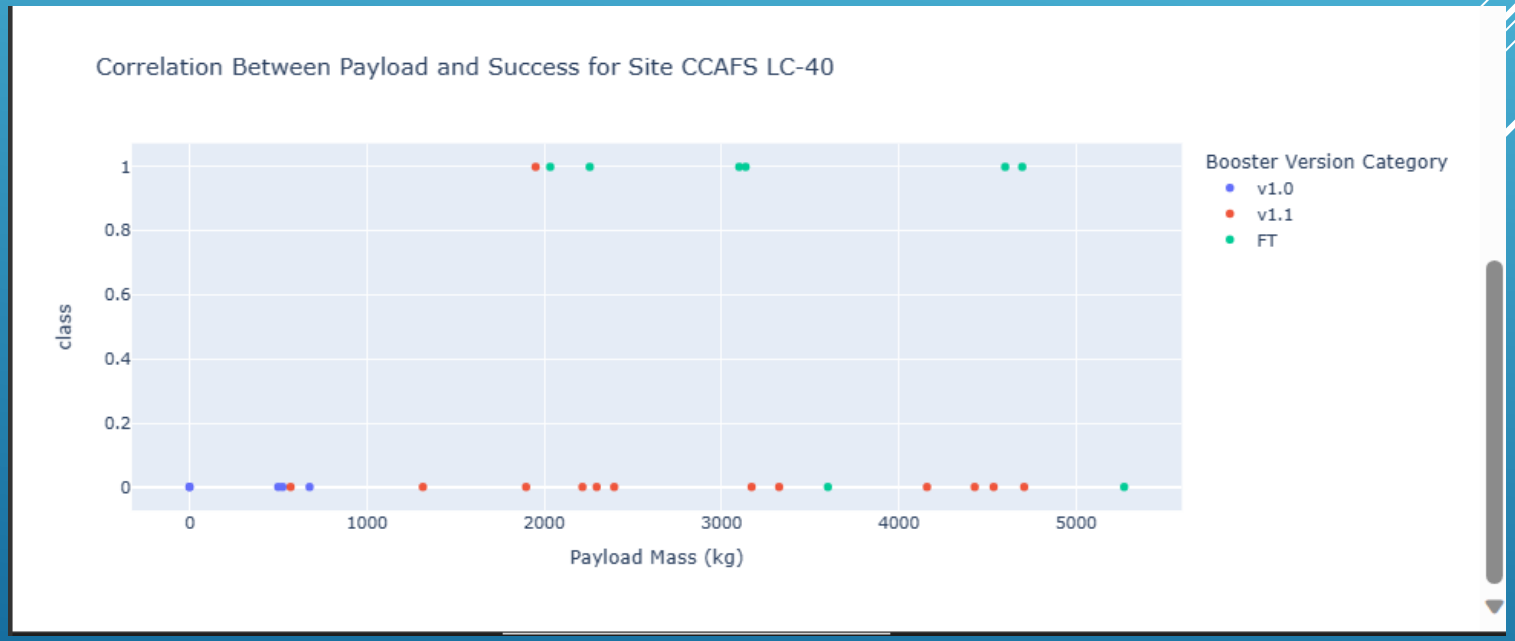
3 failed launches



PLOTLY DASHBOARD RESULTS



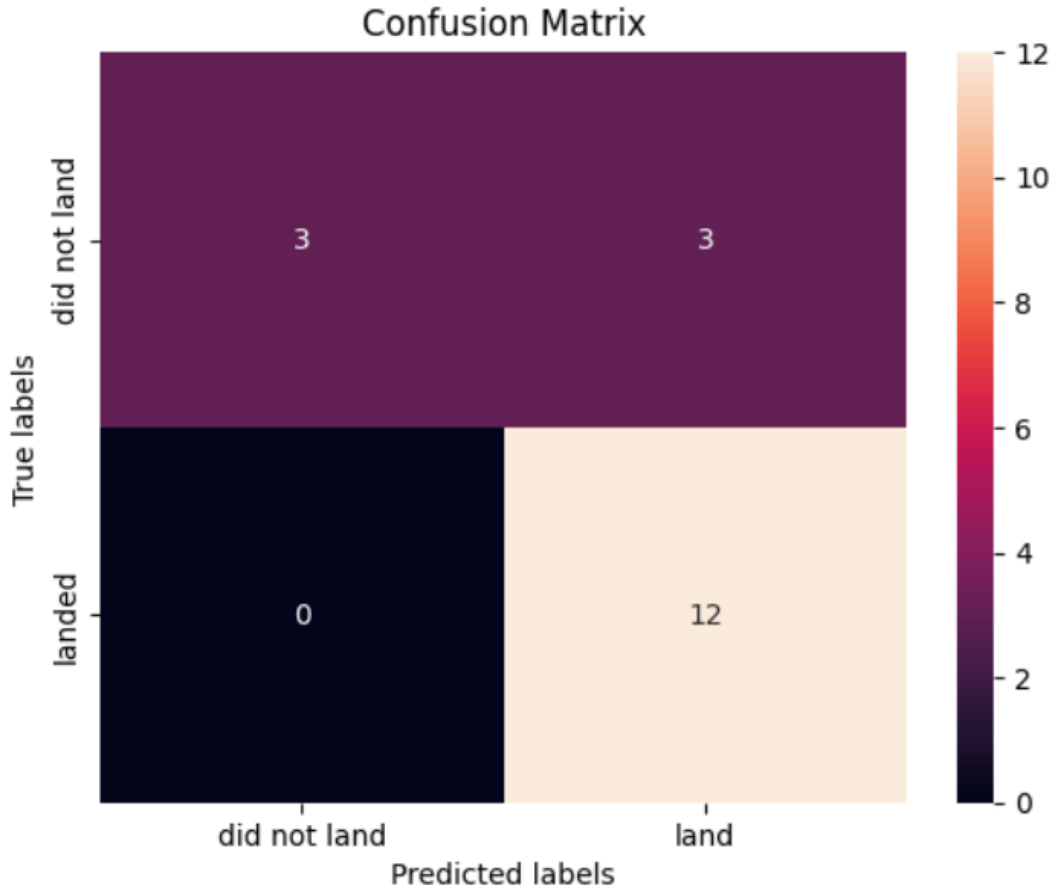
Payload masses between 2000kg and 5000kg had the highest success rate



PREDICTIVE ANALYSIS

```
Best model is DecisionTree with a score of 0.9017857142857144
Best params is : {'criterion': 'gini', 'max_depth': 4, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split':
10, 'splitter': 'random'}
```

The decision tree algorithm outperformed the other algorithms by a slightly smaller margin.



PREDICTIVE ANALYSIS

- Confusion matrix gives us an overview of the performance of any classification algorithm.
- The confusion matrix for all models were the same

CONCLUSION

Conclusions

- Launch success increased over time
- Flight number is may not necessarily be a factor that influences the success rates in orbits.
- Payload masses between 2000kg and 5000kg had the highest success rate
- The decision tree algorithm outperformed the other algorithms by a slightly smaller margin.
- KSC LC-39A had the most successful launch having a 41.2%
- Launch sites are closest to coastlines but farther away from cities, railways and highways.