IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Arijit Gogoi
12th December, 2024

# Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

# Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - EDA with data visualization
  - EDA with SQL
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all results
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

# Introduction

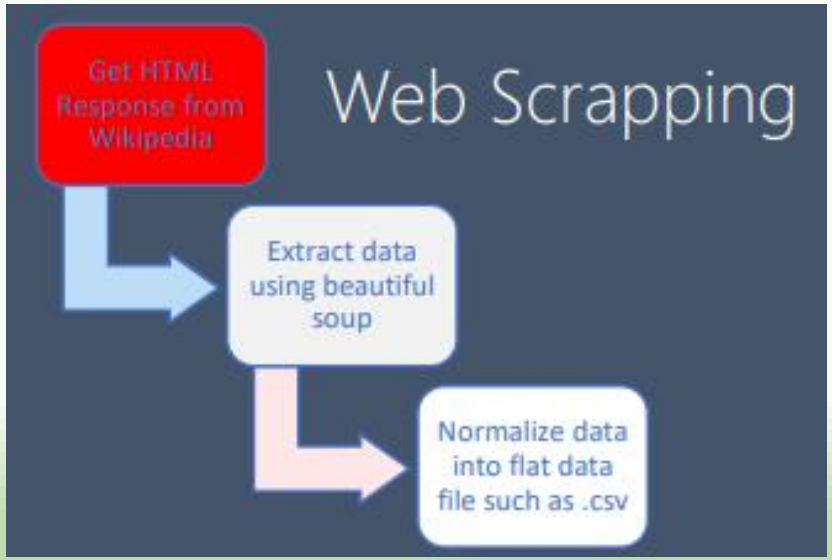
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars.
- Other providers cost upward of 165 million dollars each.
- The savings is because SpaceX can reuse the first stage.
- If we can determine if the first stage will land, we can determine the cost of a launch.
- This information can be used by an alternate company to bid against SpaceX for a rocket launch.
- So, we want to weaknesses of SpaceX, its launch sites, payload mass failure rate, etc.

# Methodology

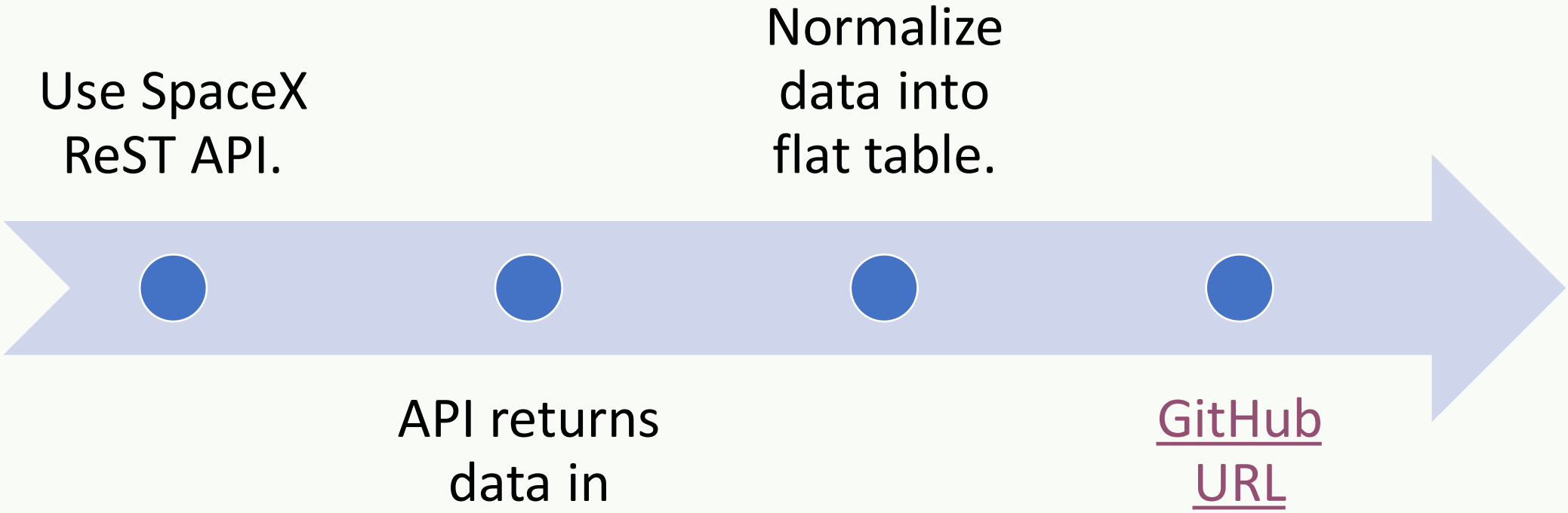
- Two ways of data collection:
  - SpaceX ReST API, processed using Pandas.
  - Web scraping SpaceX data from Wikipedia, processed using BeautifulSoup.
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash.
- Predictive analysis using classification models.
  - Scikit learn is a powerful machine learning tool in our hands.

# Data Collection

- SpaceX ReST API.
  - GET request to the API endpoint: *https://api.spacexdata.com/v4/launches/past*
  - Receive response in JSON.
  - Using Pandas, read into a DataFrame to normalize data and make it flat.
- Scrape Wikipedia
  - Using BeautifulSoup.
  - Data Wrangling in Python.

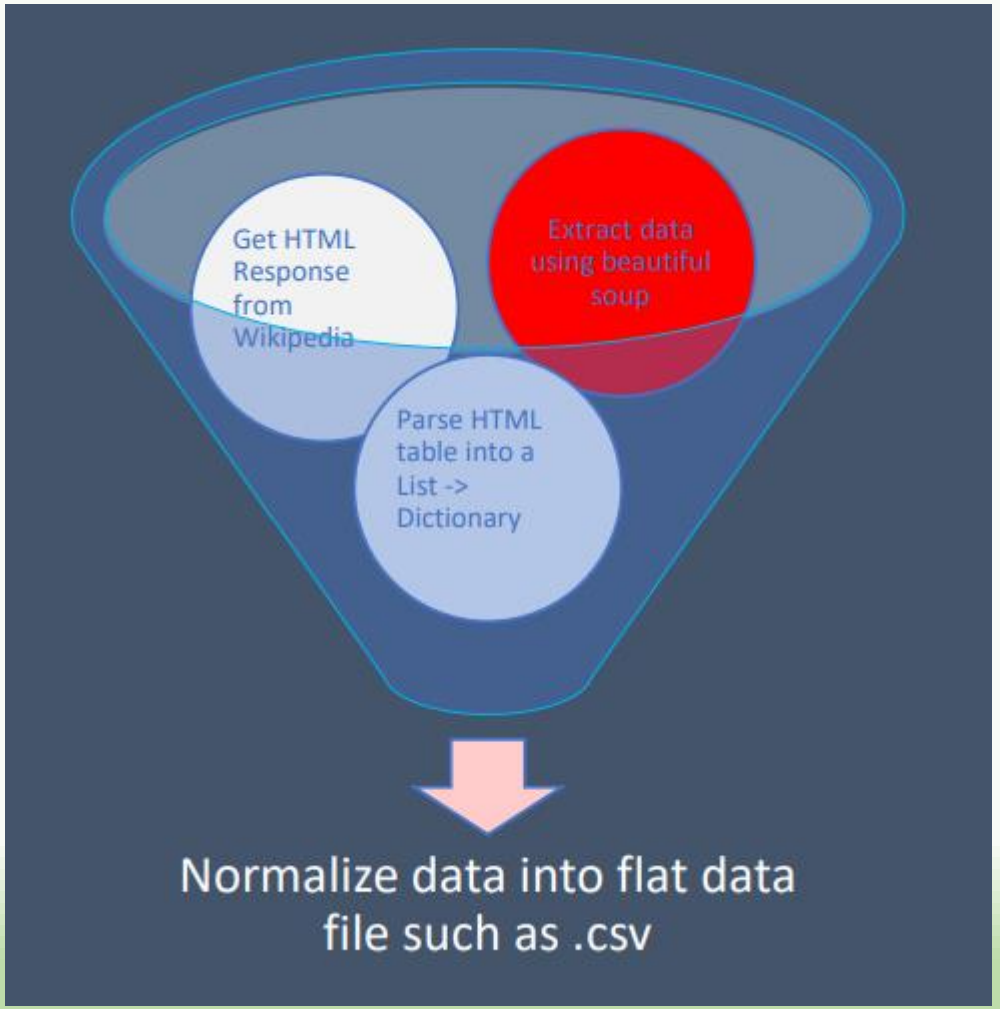


# Data Collection – SpaceX API



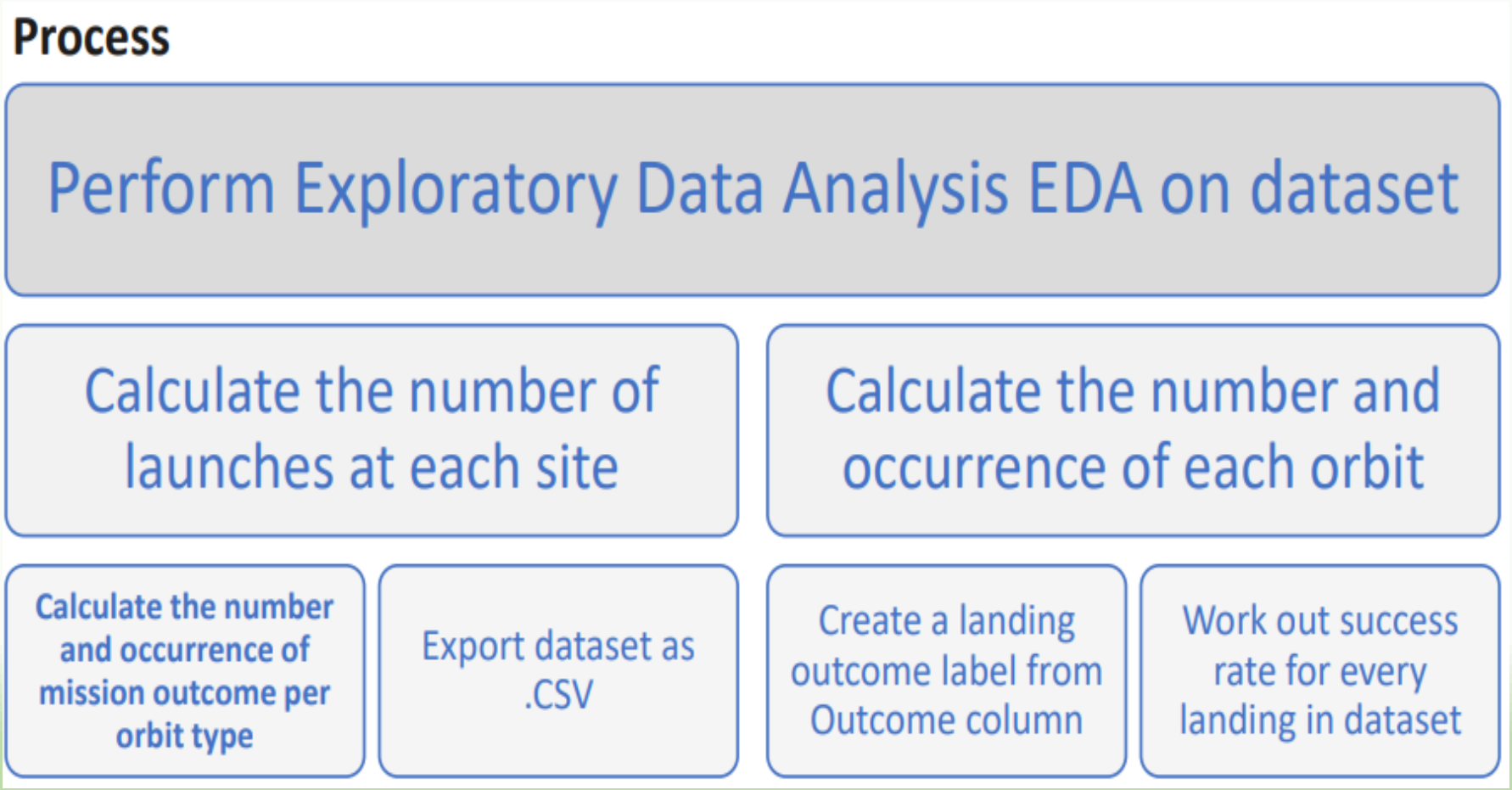
# Data Collection - Scraping

- Web scraping involves extracting data from tabular data.
- Parse HTML tags into BeautifulSoup objects.
- Use vanilla Python to clean the extracted data into Pandas DataFrame.
- GitHub



# Data Wrangling

- GitHub URL



# EDA with Data Visualization

- GitHub URL

## Scatter Graphs being drawn:

Flight Number VS. Payload Mass

Flight Number VS. Launch Site

Payload VS. Launch Site

Orbit VS. Flight Number

Payload VS. Orbit Type

Orbit VS. Payload Mass

Scatter plots show how much one variable is affected by another. The relationship between two variables is called their correlation . Scatter plots usually consist of a large body of data.

## Bar Graph being drawn:

Mean VS. Orbit

A bar diagram makes it easy to compare sets of data between different groups at a glance. The graph represents categories on one axis and a discrete value in the other. The goal is to show the relationship between the two axes. Bar charts can also show big changes in data over time.

## Line Graph being drawn:

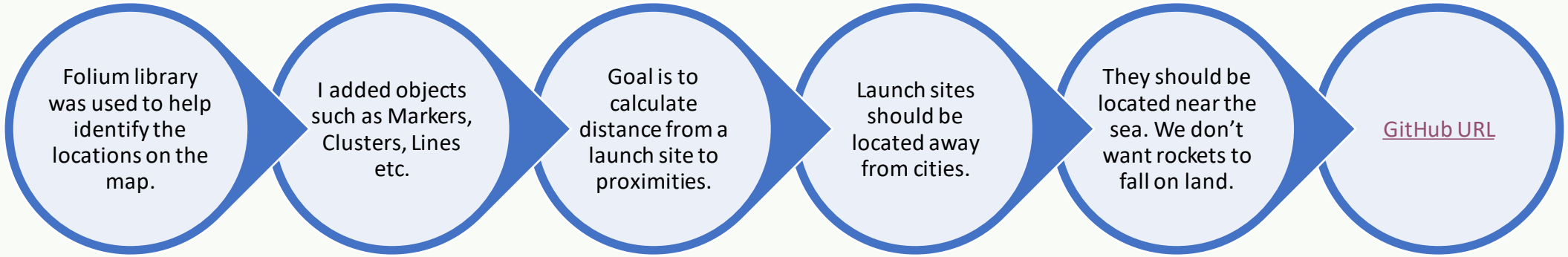
Success Rate VS. Year

Line graphs are useful in that they show data variables and trends very clearly and can help to make predictions about the results of data not yet recorded

# EDA with SQL

- GitHub URL
- Displaying the names of the unique launch sites in the space mission .
- Displaying 5 records where launch sites begin with the string ‘CCA’.
- Displaying the total payload mass carried by boosters launched by NASA.
- Displaying average payload mass carried by booster version F9 v1.1.
- Listing the date where the successful landing outcome in drone ship was achieved.
- Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000.
- Listing the total number of successful and failure mission outcomes.
- Listing the names of the booster_versions which have carried the maximum payload mass.
- See full slides.

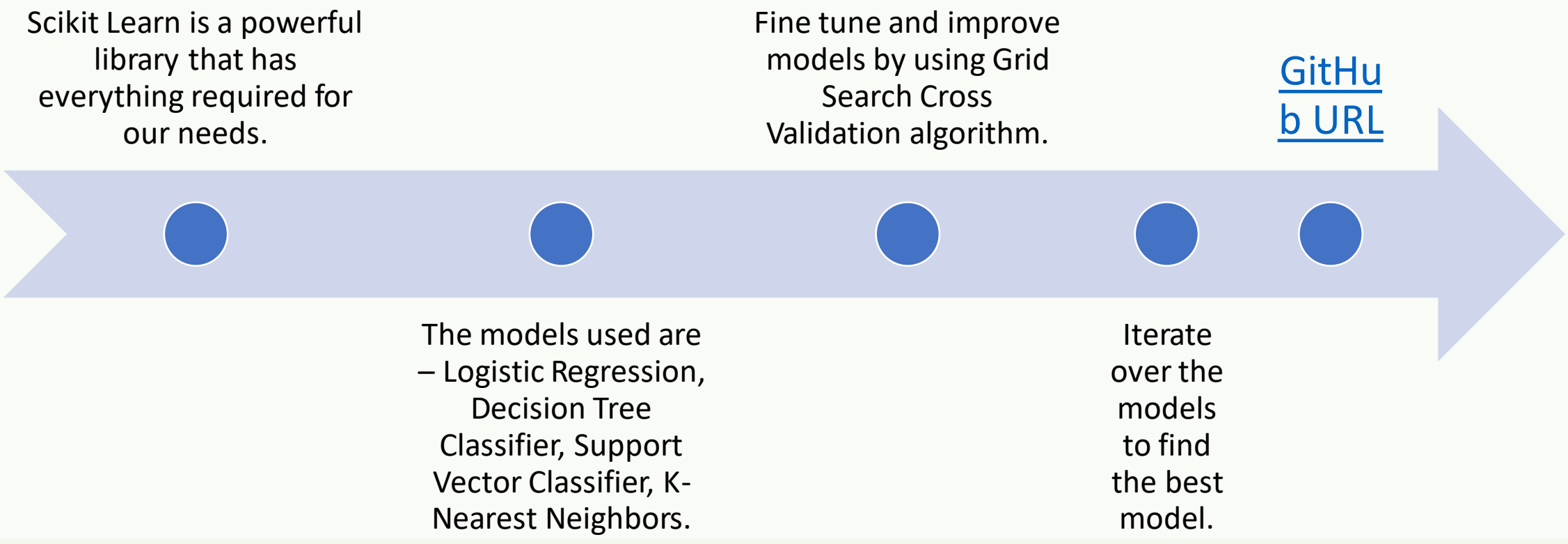
# Interactive Map with Folium



# Dashboard with Plotly Dash

- I created a dashboard that shows the following –
- Overall success rate of all launch sites.
- Success rate by individual launch sites. This helps identify the successful launch sites.
- Success count on Payload mass for all sites.
- GitHub URL

# Predictive Analysis (Classification)

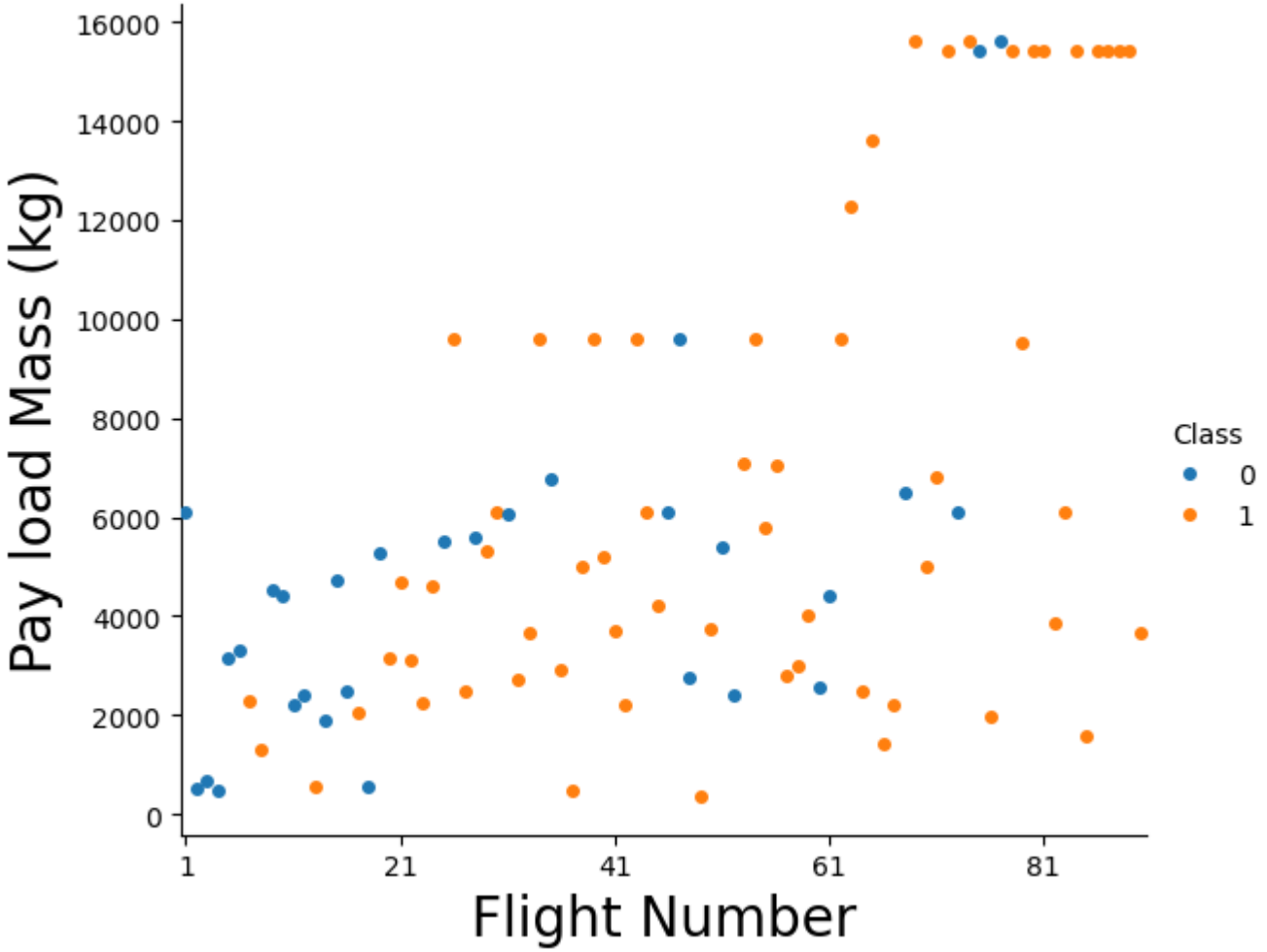


# RESULTS

- EDA with Visualization.
- EDA with SQL.
- Folium analytics.
- Predictive Analysis Results.

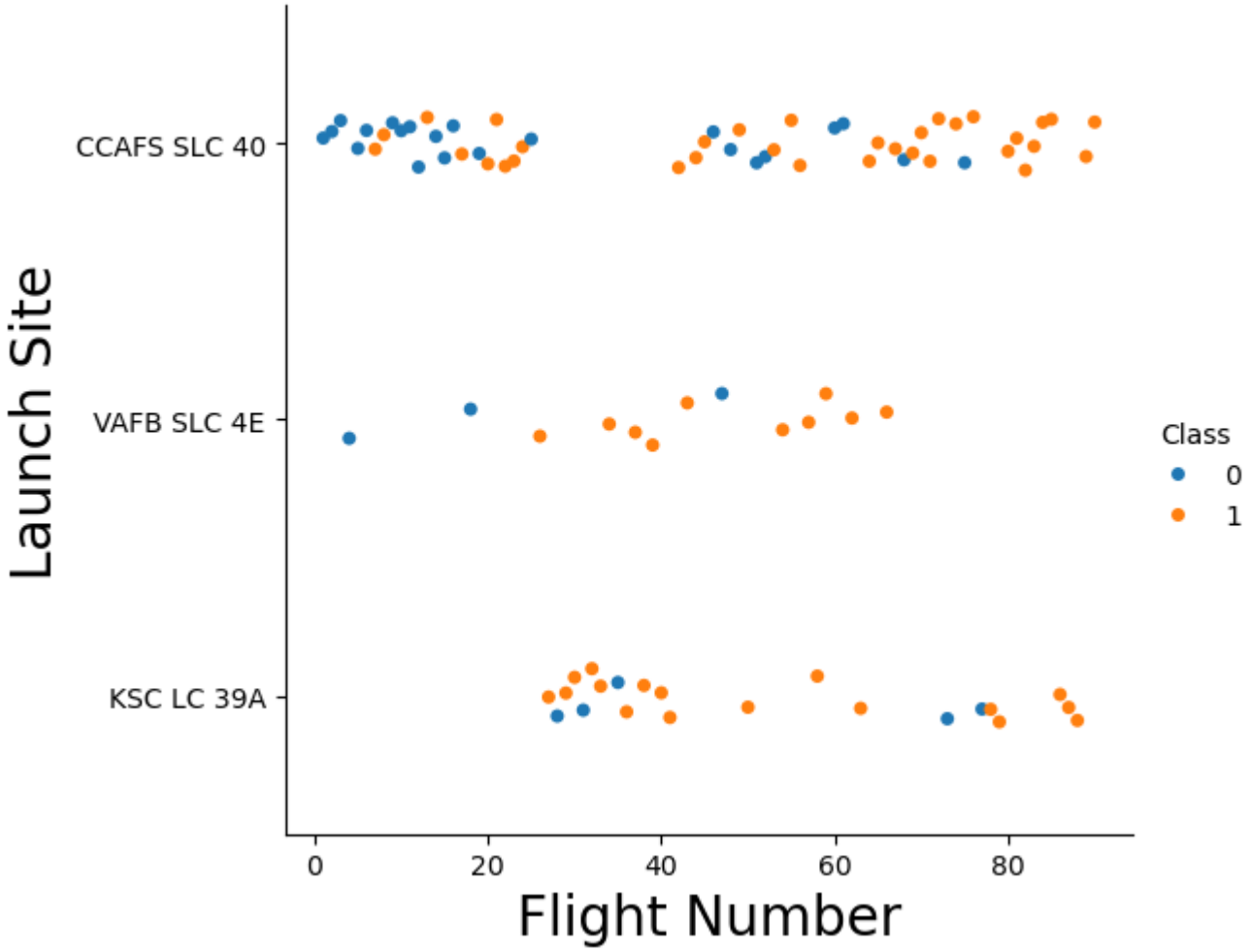
# EDA with Visualization

# Flight Number vs. Payload Mass



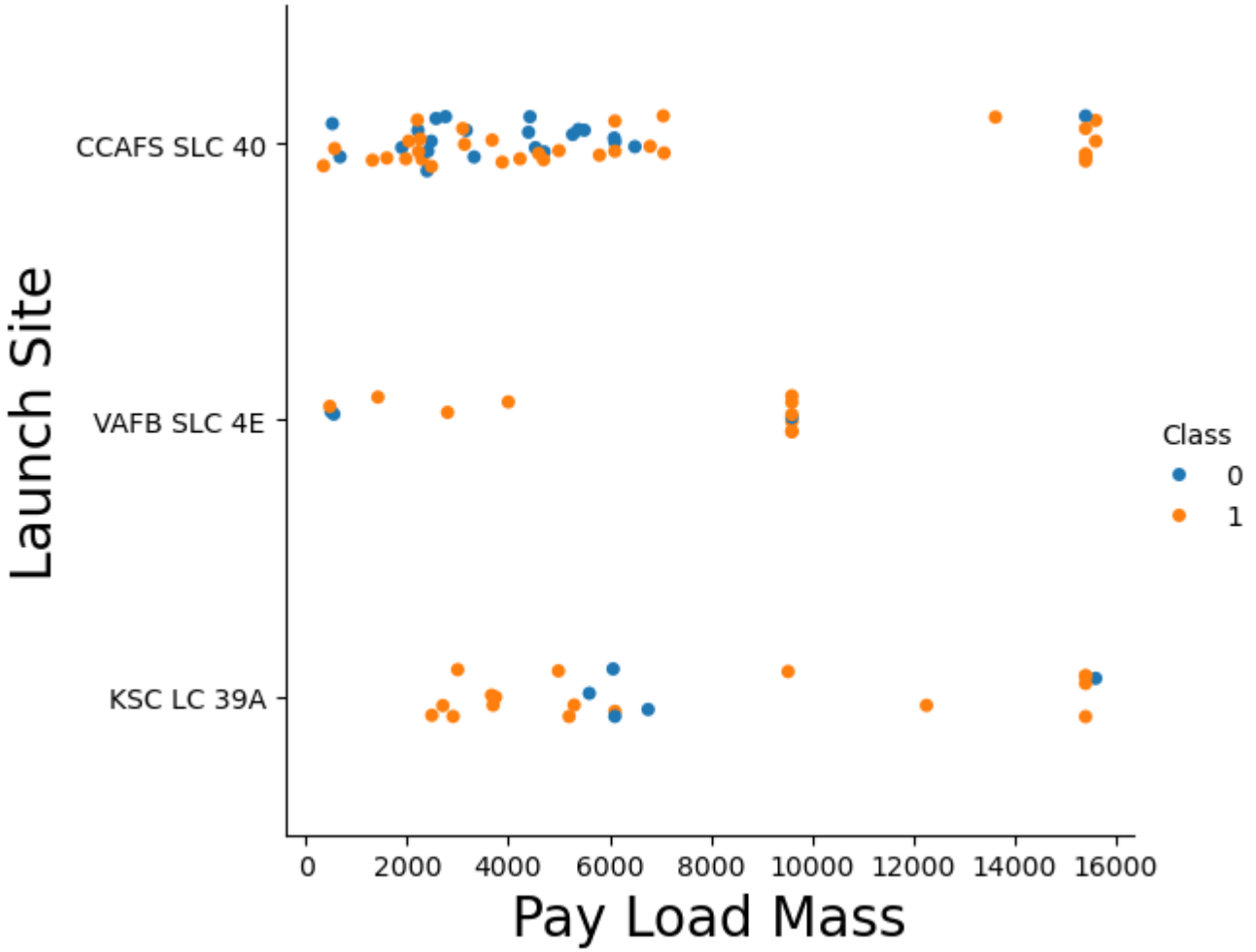

- More payloads tend to mean better success (see orange dots)
- More payloads mean bigger rockets.
- Hence better results.

# Flight Number vs. Launch Site




- The launch site CCAFS SLC 40 has more attempts and more successes.
- KSC LC 39A has better success rate.

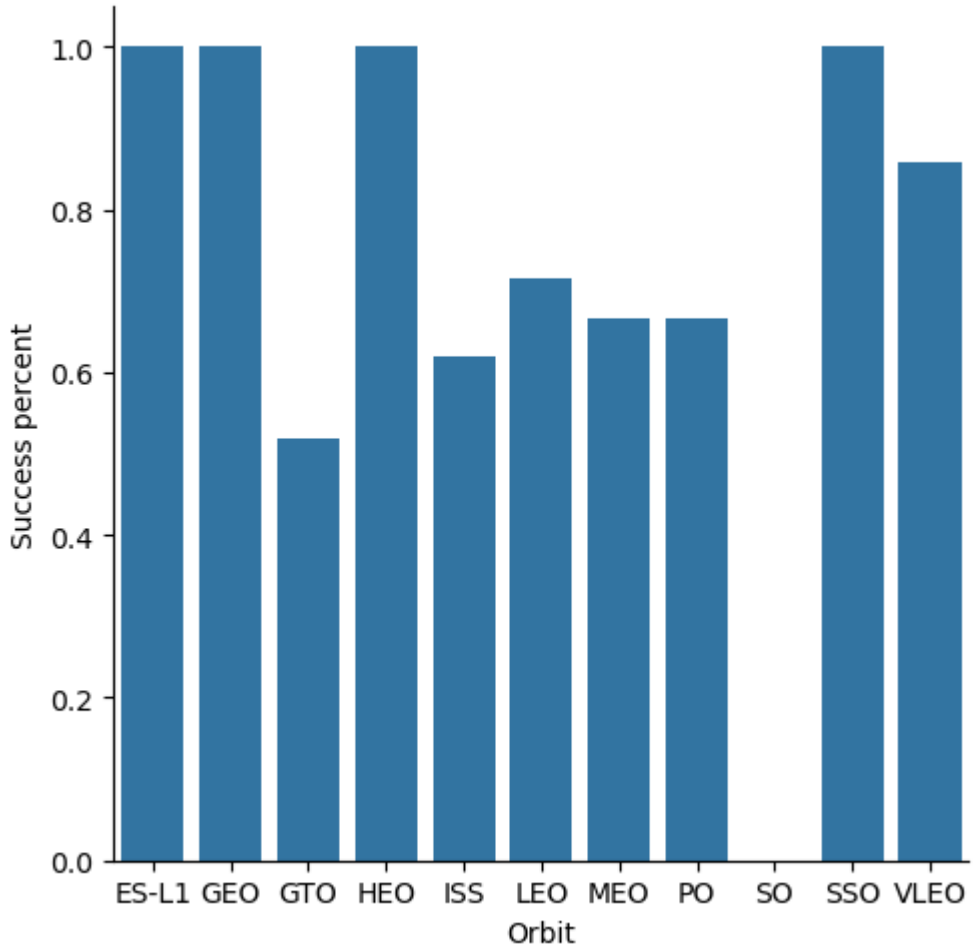
# Launch Site vs. Payload Mass




- VAFB-SLC launch site has no rockets launched for heavy payload mass (greater than 10000).

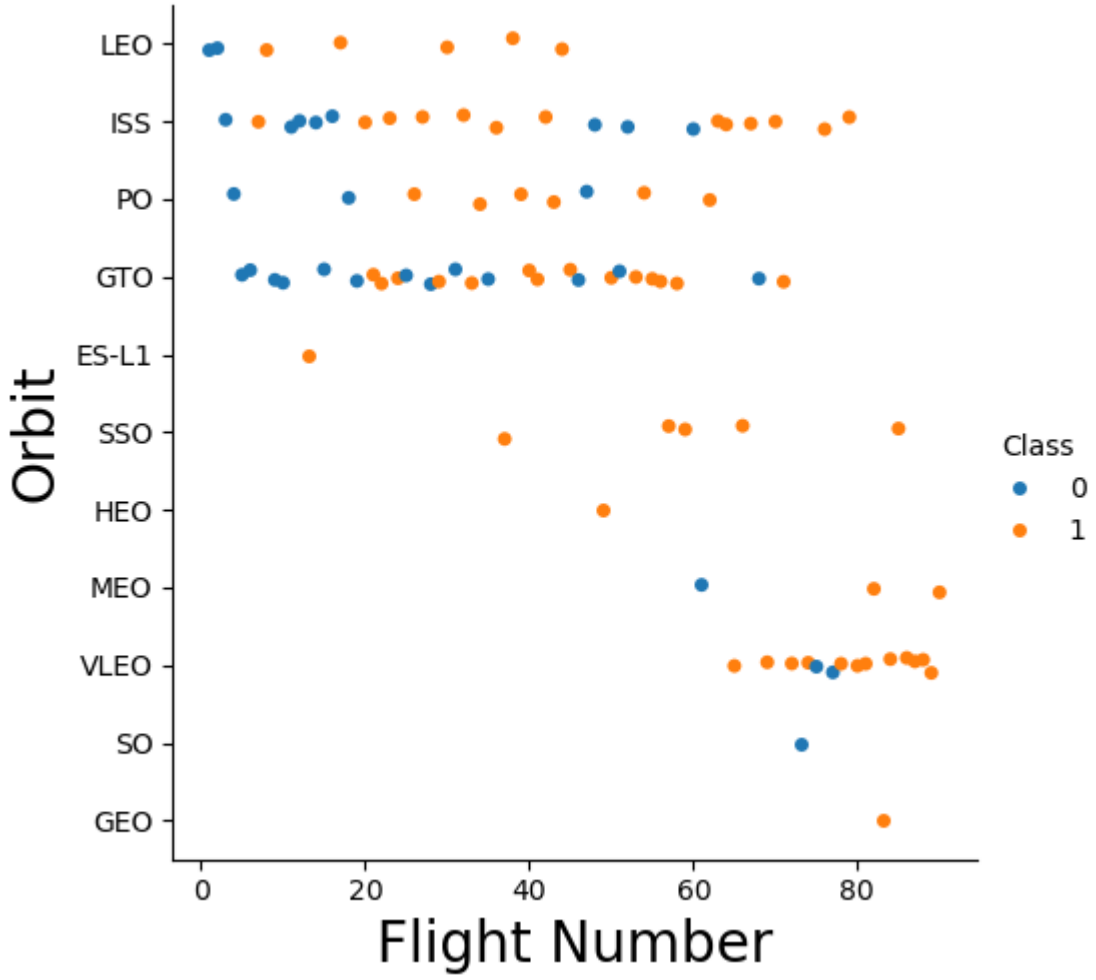
# Success Rate vs. Orbit Type

- Orbits that have full success rates - ES-L1, GEO, HEO and SSO.

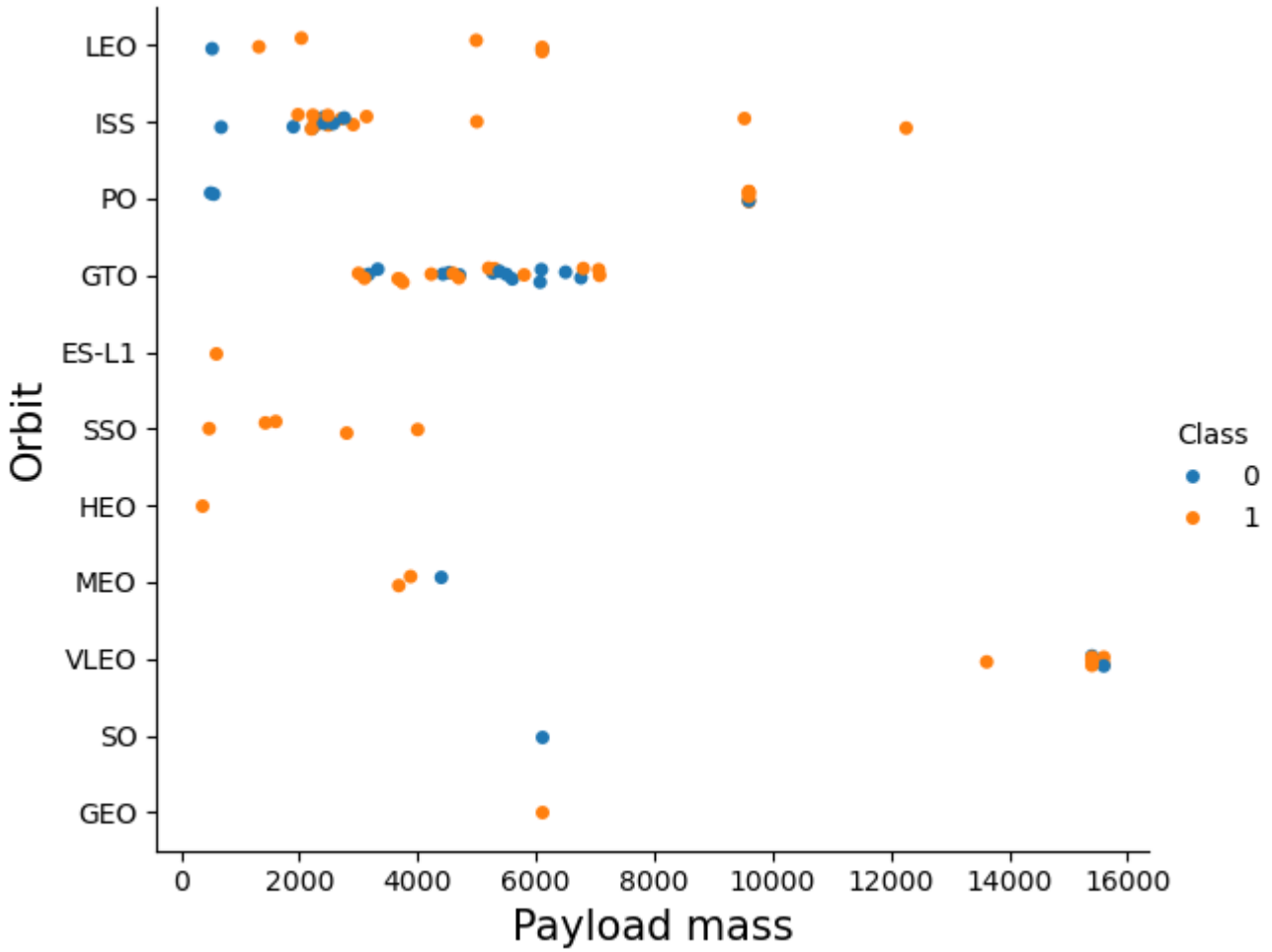


# Flight Number vs. Orbit Type

- SSO and LEO have relationship with flight number.



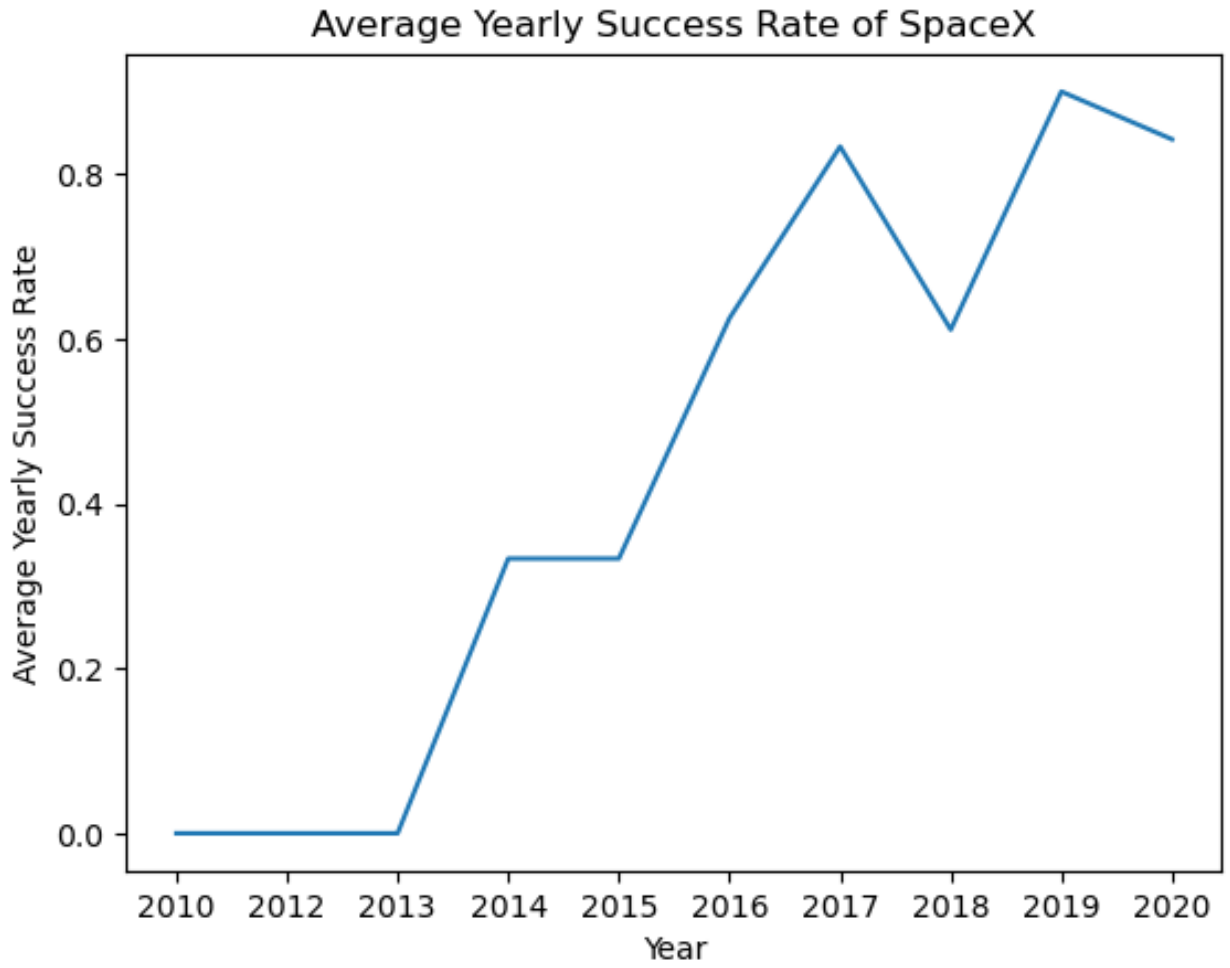
# Payload vs. Orbit Type




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

# Launch Success Yearly Trend

- Success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.



# EDA with SQL

# All Launch Site Names

- These are all the launch sites.

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

# Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

| 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

- Total payload carried by boosters for customer NASA.

| Customer | Total Payload Mass |
|---|---|
| NASA (CRS) | 45596 |

# Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

| Booster_Version | Average Payload Mass of F9 v1.1 |
|---|---|
| F9 v1.1 | 2928.4 |

# First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad.

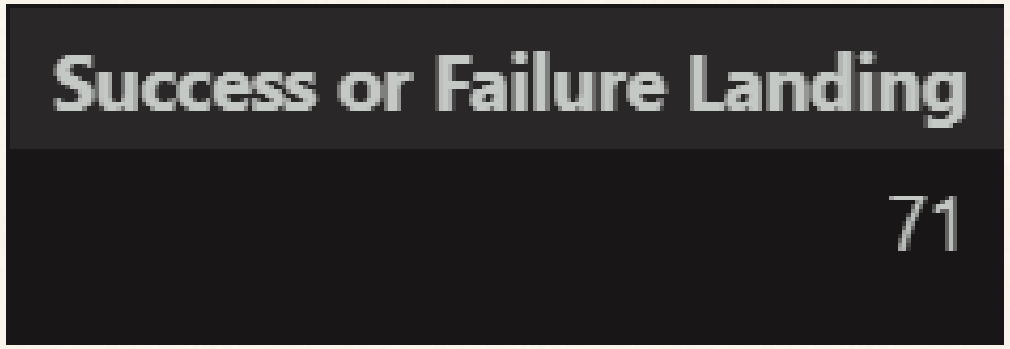
| First date of successful landing in ground pad. | Landing_Outcome |
|---|---|
| 2015-12-22 | Success (ground pad) |

# Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000.

| 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 Number of Successful and Failure Mission Outcomes




- Total number of successful and failure mission outcomes.
- Note that 31 outcomes are not listed as either success or failure.
- Total records in database is 100.

# Boosters Carried Maximum Payload

- Maximum Payload Mass is 15600 kg. These are all the Boosters which have carried max payload.

| 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 |

# 2015 Launch Records

- Failed landing outcomes in drone ship, their booster versions, and launch site names for the year 2015.
- Only 2 records.

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

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

- Ranking of count of landing outcomes between the dates 2010-06-04 and 2017-03-20.

| Landing_Outcome | Count of Landing Outcome |
|---|---|
| No attempt | 10 |
| Success (drone ship) | 5 |
| Failure (drone ship) | 5 |
| Success (ground pad) | 3 |
| Controlled (ocean) | 3 |
| Uncontrolled (ocean) | 2 |
| Failure (parachute) | 2 |
| Precluded (drone ship) | 1 |

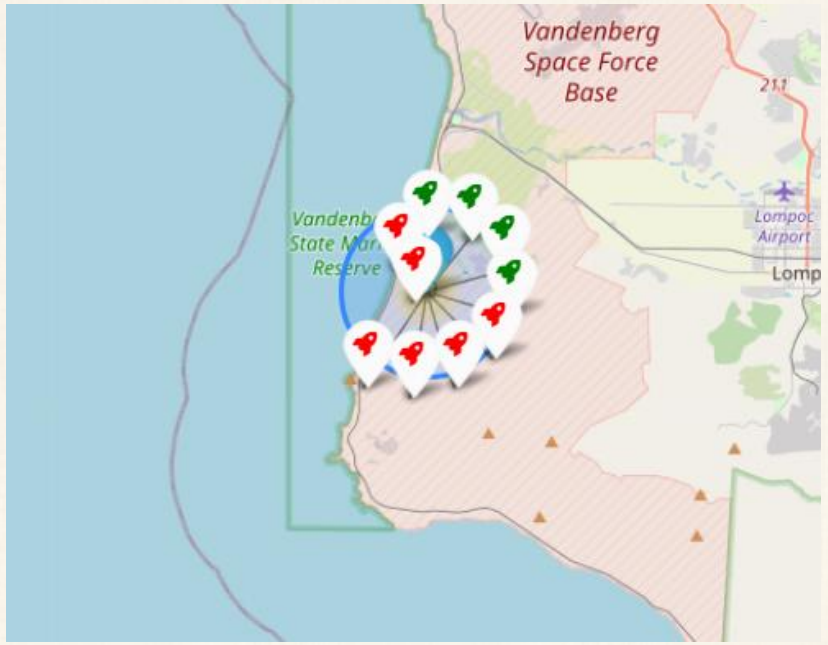
# Folio Analytics

# All Launch Sites Marked on Map

- SpaceX has launch sites in Florida and California.

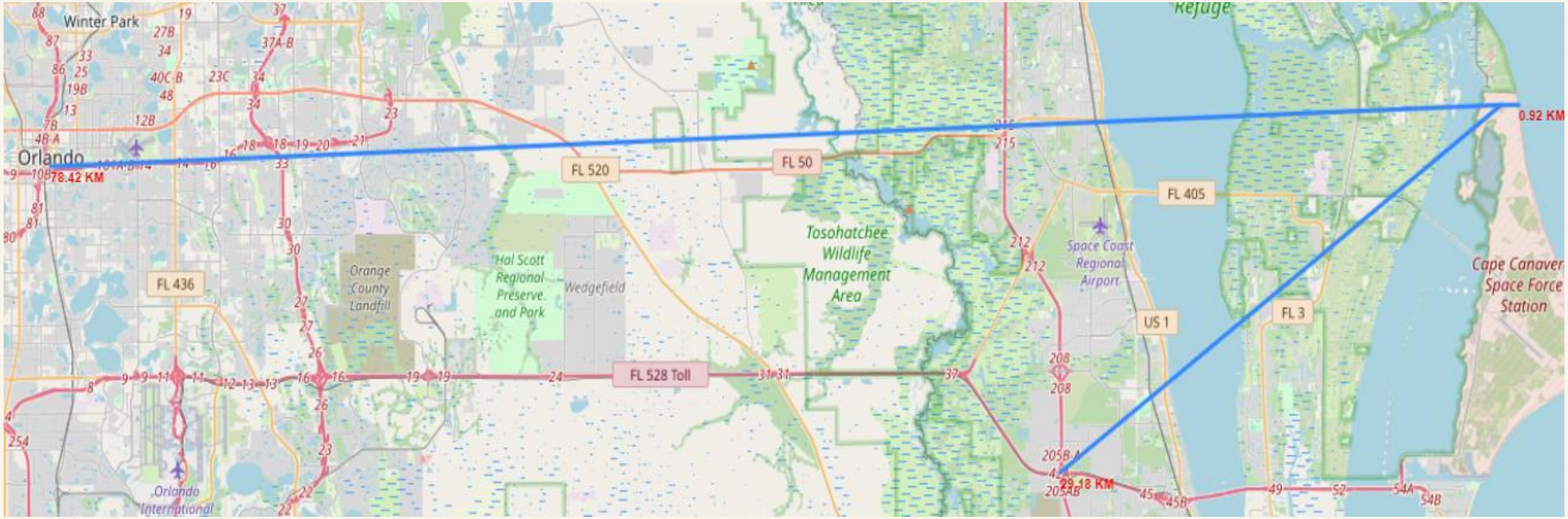
# Success and Failure of Launch Sites

- Green rocket means landing success. Red rocket means landing failure.

# Proximities of CCAFS Launch Site

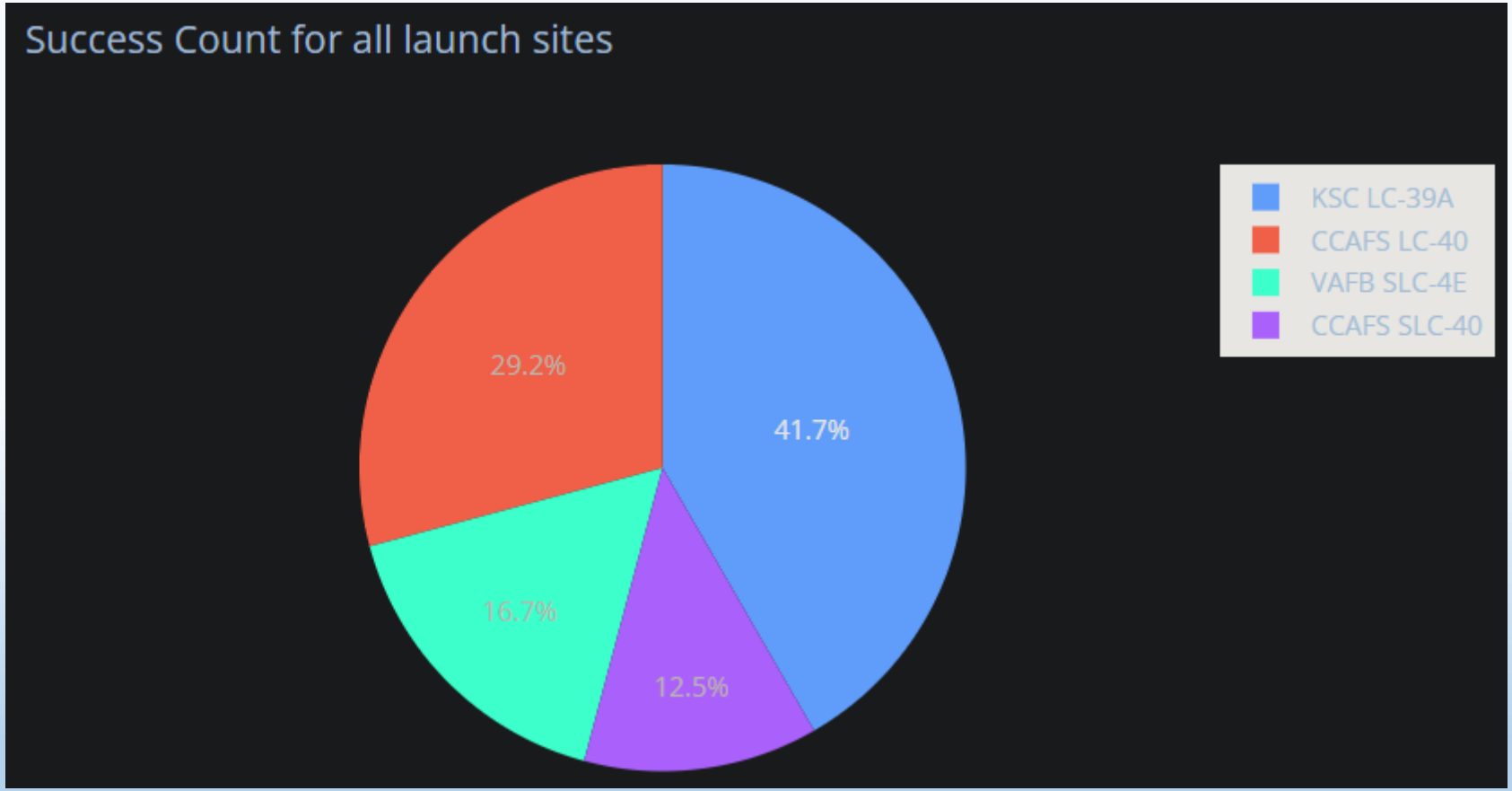
- CCAFS is situated in an ideal location and hence has the best success rates.
- Its very close to the sea on an island and far from civilians.



# Launch Site and Paylod Success Rate Dashboard

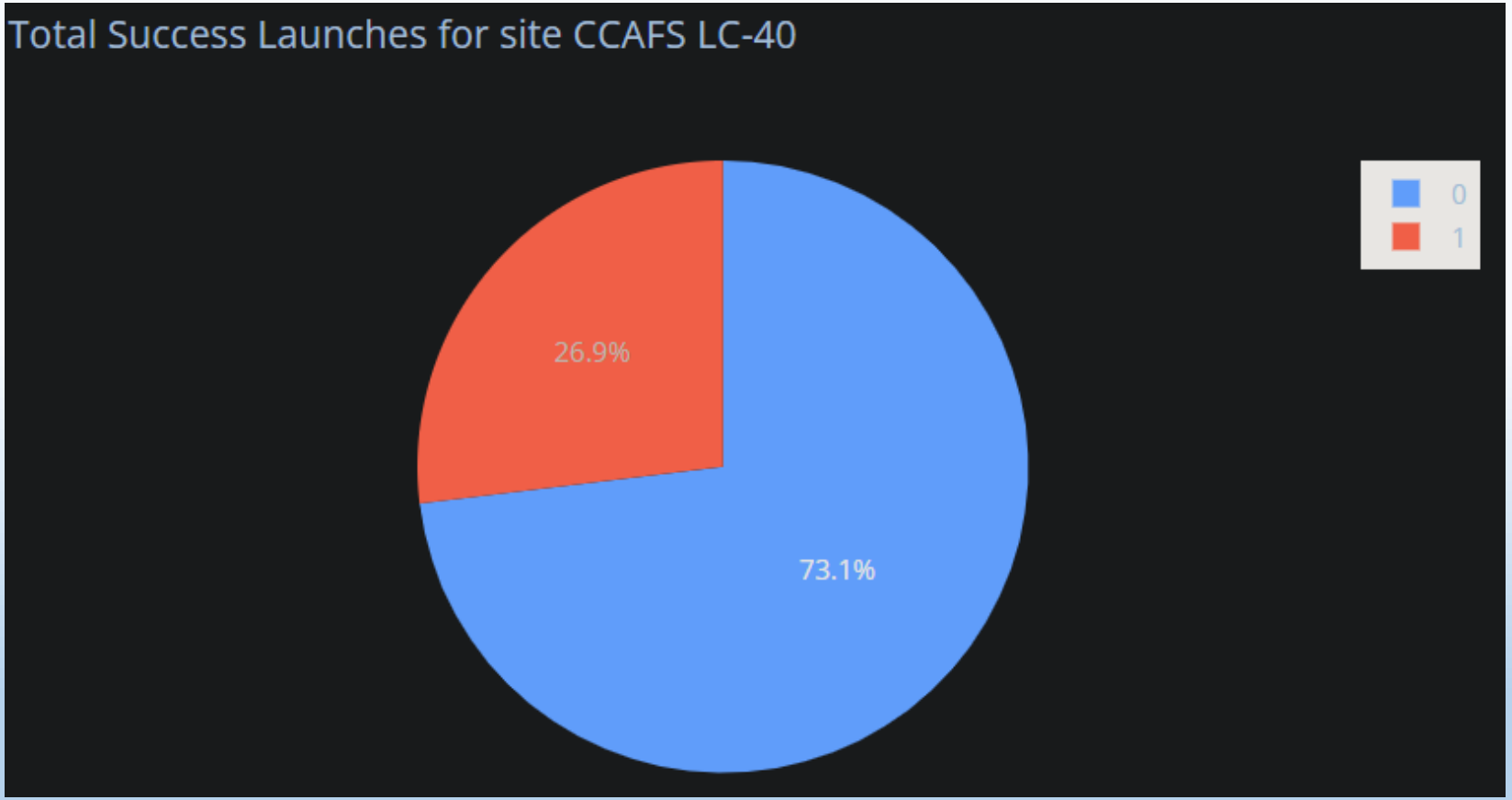
# Success proportion of sites

- The CCAFS site is the least successful.



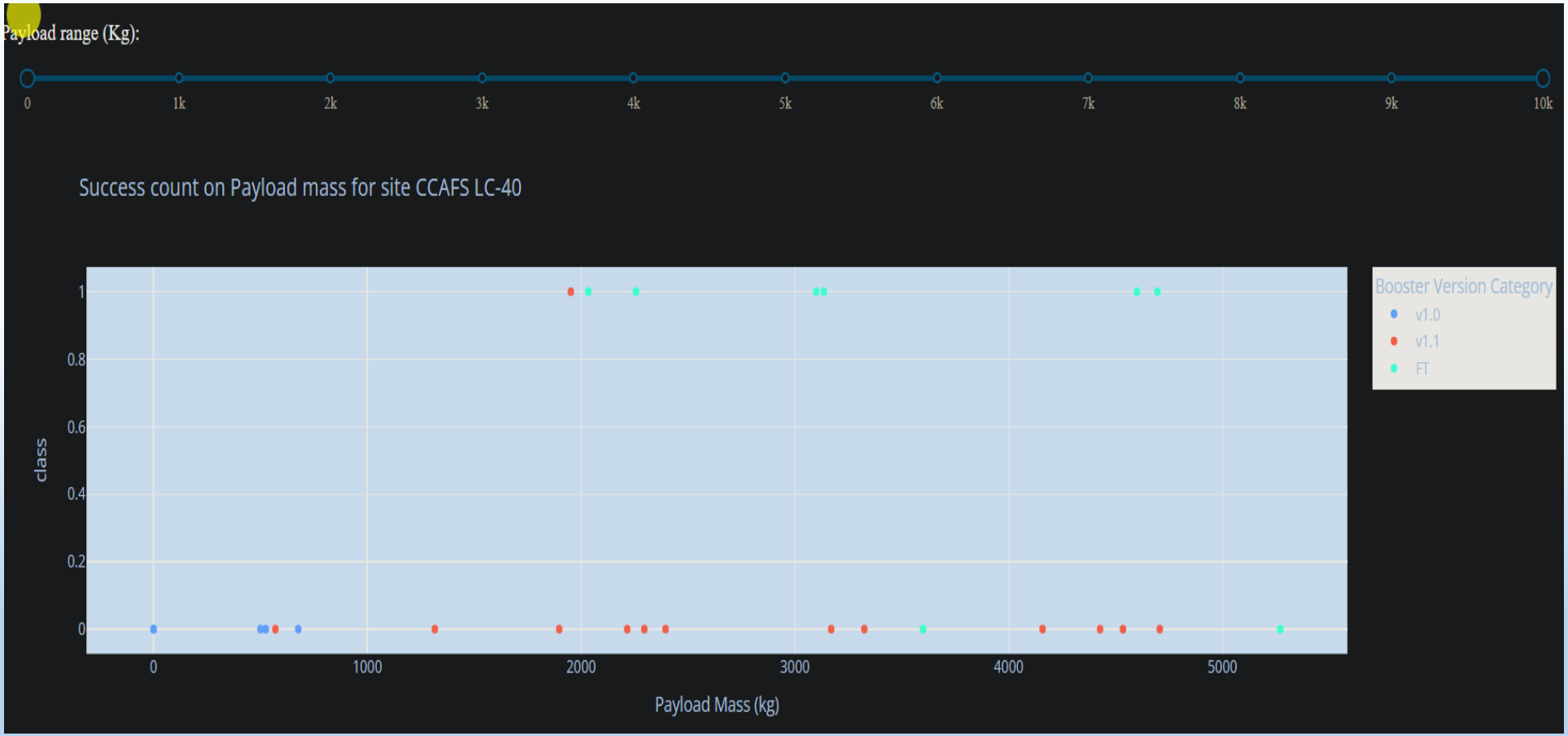
# Success ratio of CCAFS

- We are competitors of SpaceX. We want to bid against CCAFS launch site.
- CCAFS has an abysmal success ratio of 27%. That's 73% failure.



# Success ratio of Payload Mass

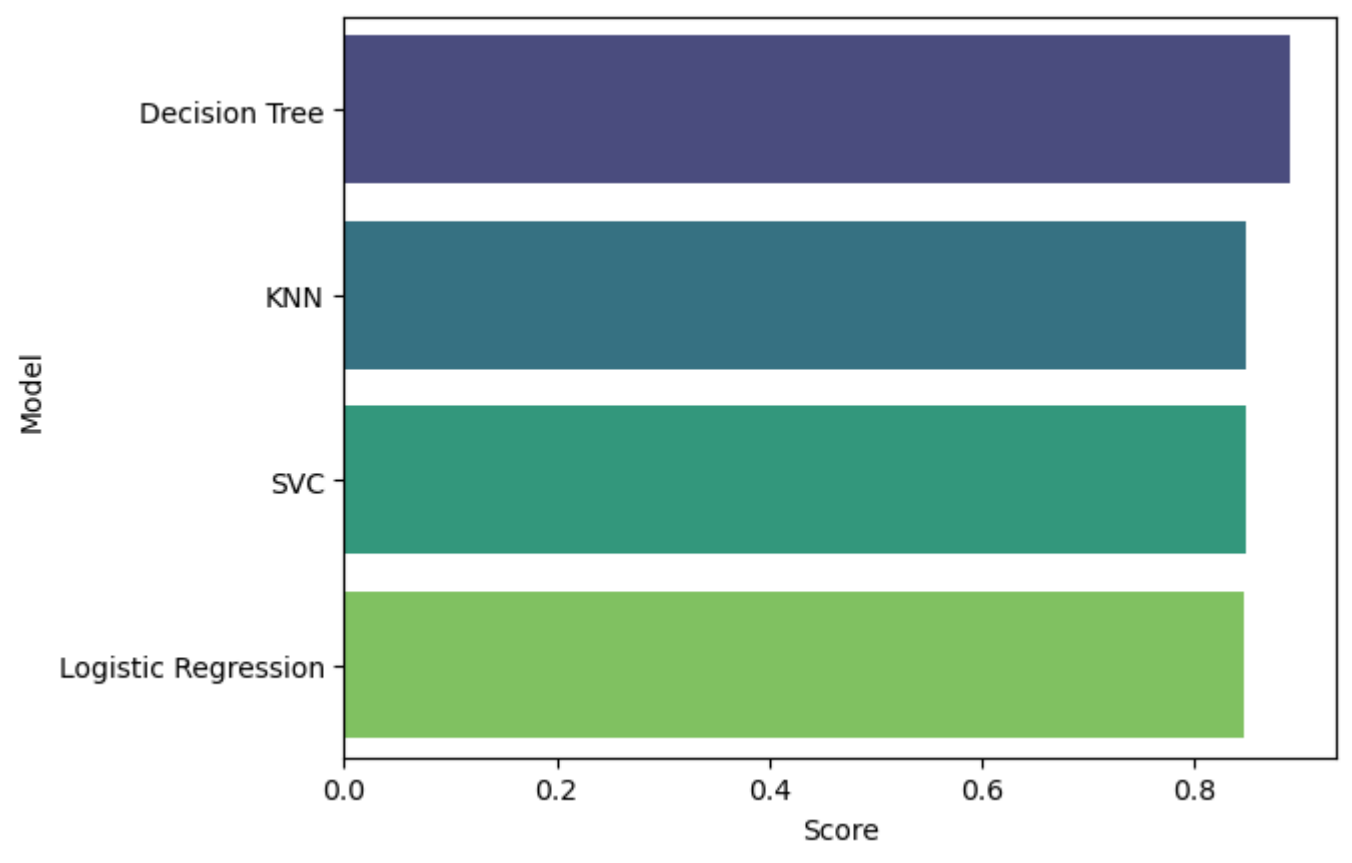
- I also created a slider for Payload vs. class for your perusal.



# Predictive Analysis Results

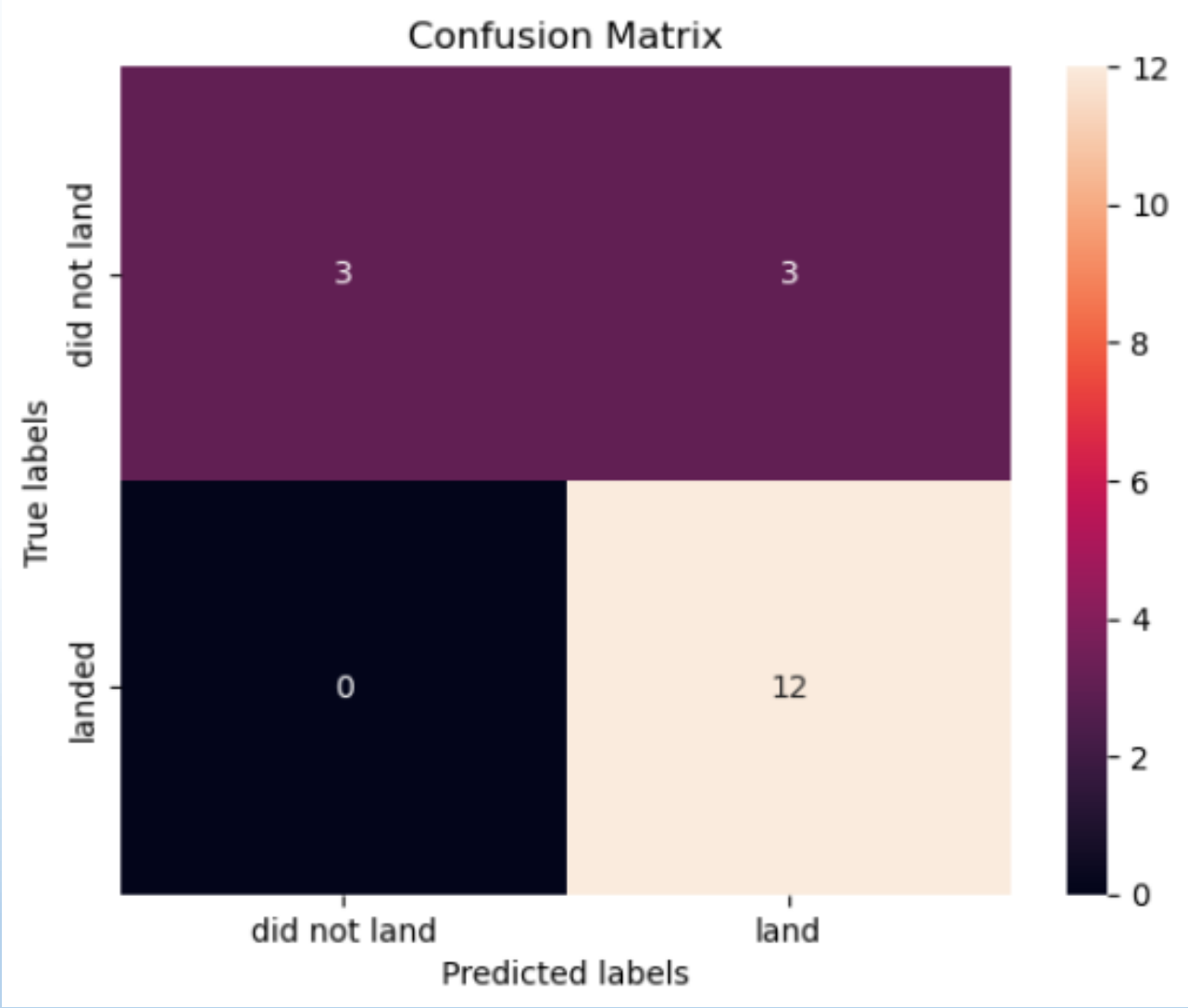
# Accuracy Score of Decision Tree (Best Model)

- Out of all the models tested, Decision Tree had marginally better score than the rest.




| | Model | Score |
|---|---|---|
| 0 | Decision Tree | 0.889286 |
| 1 | KNN | 0.848214 |
| 2 | SVC | 0.848214 |
| 3 | Logistic Regression | 0.846429 |

# Confusion Matrix of Decision Tree Classifier



# Best parameters for Decision Tree Classifier

```
The best parameters for Decision Tree are:
{    'criterion': 'gini',
     'max_depth': 12,
     'max_features': 'sqrt',
     'min_samples_leaf': 4,
     'min_samples_split': 2,
     'splitter': 'random'}
```

# Conclusions

- Key Insight: SpaceX is getting better with time.
- The Decision Tree Model is the best predicts success of launches.
- Low weighted payloads perform better than the heavier payloads.
- KSA facility had the most successful launches amongst all the sites.
- Orbits GEO, HEO, SSO, ES-L1 have the best success rates.

# Thank you!