IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Waleed Qadir Khan

2023-11-22

Outline

Executive Summary

- Summary of methodologies
    - Data collection
    - Data wrangling
    - Data Visualization
    - Data Analysis with SQL
    - Building interaction with Folium
    - Predictive analysis
- Summary of all results
    - Exploratory Data Analysis results
    - Interactive Dashboard
    - Predictive Analysis

Introduction

- Project Statement
  - The project's goal is to predict the successful landing of Falcon 9's first stage, crucial for cost estimation.
  - Falcon 9 launches cost $62 million, significantly lower than competitors' prices of over $165 million, due to the reusability of its first stage.
  - Accurate predictions of landing success are valuable for companies competing with SpaceX in rocket launches.
- Problems you want to find answers
  - The key focus of the investigation is to identify the main characteristics that differentiate a successful landing from a failed one for SpaceX's rockets. This encompasses understanding how various rocket variables interact and influence the outcome of a landing, and determining the optimal conditions that would enable SpaceX to achieve the highest landing success rate.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Query SpaceX API for Falcon 9 launch data.
  - Scrape Wikipedia for Falcon 9 launch records.
- Perform data wrangling
  - filtering data, cleaning inconsistencies, and one-hot encoding categorical variables
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Data collection combined SpaceX REST API requests and Wikipedia web scraping to acquire comprehensive launch details for in-depth analysis.
- SpaceX API
    - FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Longitude, Latitude etc.
    - https://api.spacexdata.com/v4/launches/past
- Wikipedia
    - Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Date, Time etc.
    - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

- GitHub URL

- 1. SpaceX Rest API call
- 2. Decode to json
- 3. create dataframe
- 4. clean data
- 5. export to csv

Data Collection - Scraping

- [GitHub URL]

- Get call to acquire data
- Create a BeautifulSoup object from response
- Extract relevant details from tables
- Create dictionary from data
- Convert to pandas dataframe
- Export to csv

Data Wrangling

- Conduct exploratory data analysis (EDA) to discern patterns and establish training labels.
- Quantify launch frequencies by site.
- Enumerate orbit types and their corresponding launch instances.
- Assess mission outcomes in relation to different orbit types.
- Generate binary landing outcome labels from the 'Outcome' column.
- Export the refined dataset to a CSV file for further use.
- GitHub URL

EDA with Data Visualization

- Scatter Plots – to show the relationship between variables.
    - 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
- Bar Graphs – to show relationship between categoric values
    - Success rate vs. Orbit
- Line Graph – to show trend over time
    - Success rate vs. Year
- GitHub URL

EDA with SQL

- 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 (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved
- Listing the names of the boosters which have success in drone ship 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
- Listing the failed landing outcomes in drone ship, their booster versions and launch site names in year 2015
- Ranking 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
- [GitHub URL]

Build an Interactive Map with Folium

- Map 1 – Mark all launch sites on the Map
  - Mark the NASA Johnson Space Center and all launch sites with coordinates, popups, and labels.
- Map 2 – Marking successful and failed launches for each site on the map
  - Use green markers for successful launches and red for failures, clustered by launch site.
- Map 3 – Distances between launch site to its proximities
  - Display distances from Launch Site KSC LC-39A to nearby railways, highways, coastlines, and the closest city with colored lines.
- GitHub URL

Build a Dashboard with Plotly Dash

- Add a Launch Site Drop-down input component
  - Added a dropdown list to allow for Launch Site selection.
- Adding a callback to render success-pie-chart based on selected site dropdown
  - Incorporated a pie chart displaying total and site-specific launch success rates.
- Adding a Range Slider to Select Payload
  - Implemented a payload range selector slider.
- Add a callback function to render the success-payload-scatter-chart scatter plot
  - Presented a scatter plot to analyze payload versus launch success correlation.
- GitHub URL

Predictive Analysis (Classification)

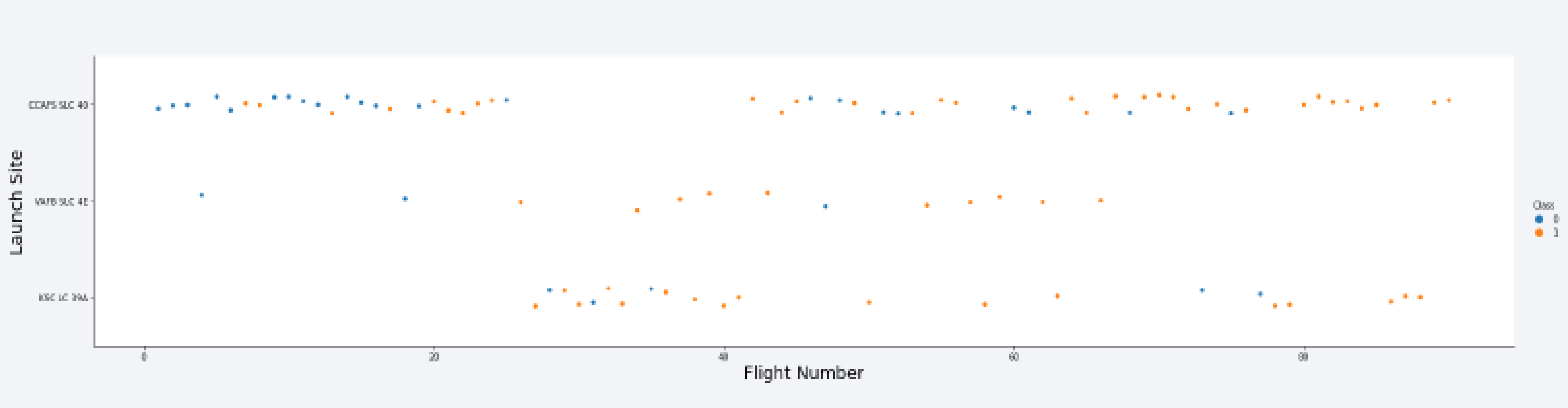
- Data Preparation
  - Load and normalize dataset.
  - Partition into training and test sets.
- Model Preparation
  - Choose machine learning algorithms.
  - Configure GridSearchCV parameters.
  - Train models on the training set.
- Model Evaluation
  - Determine optimal hyperparameters.
  - Measure model accuracy on test set.
  - Visualize with Confusion Matrix.
- Model Comparison
  - Assess model performance by accuracy.
  - Select the highest-accuracy model.
- GitHub URL

Results

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

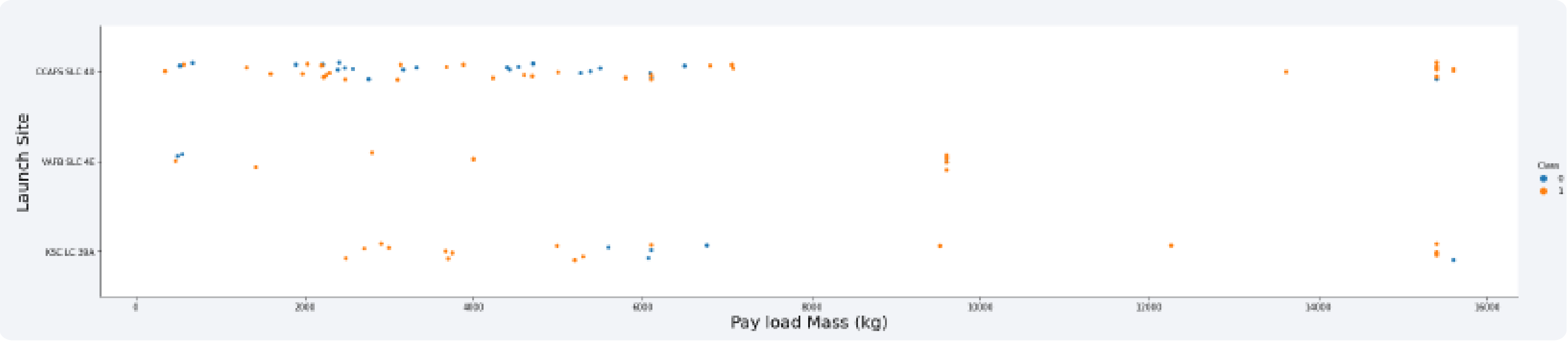
Section 2

Insights drawn from EDA



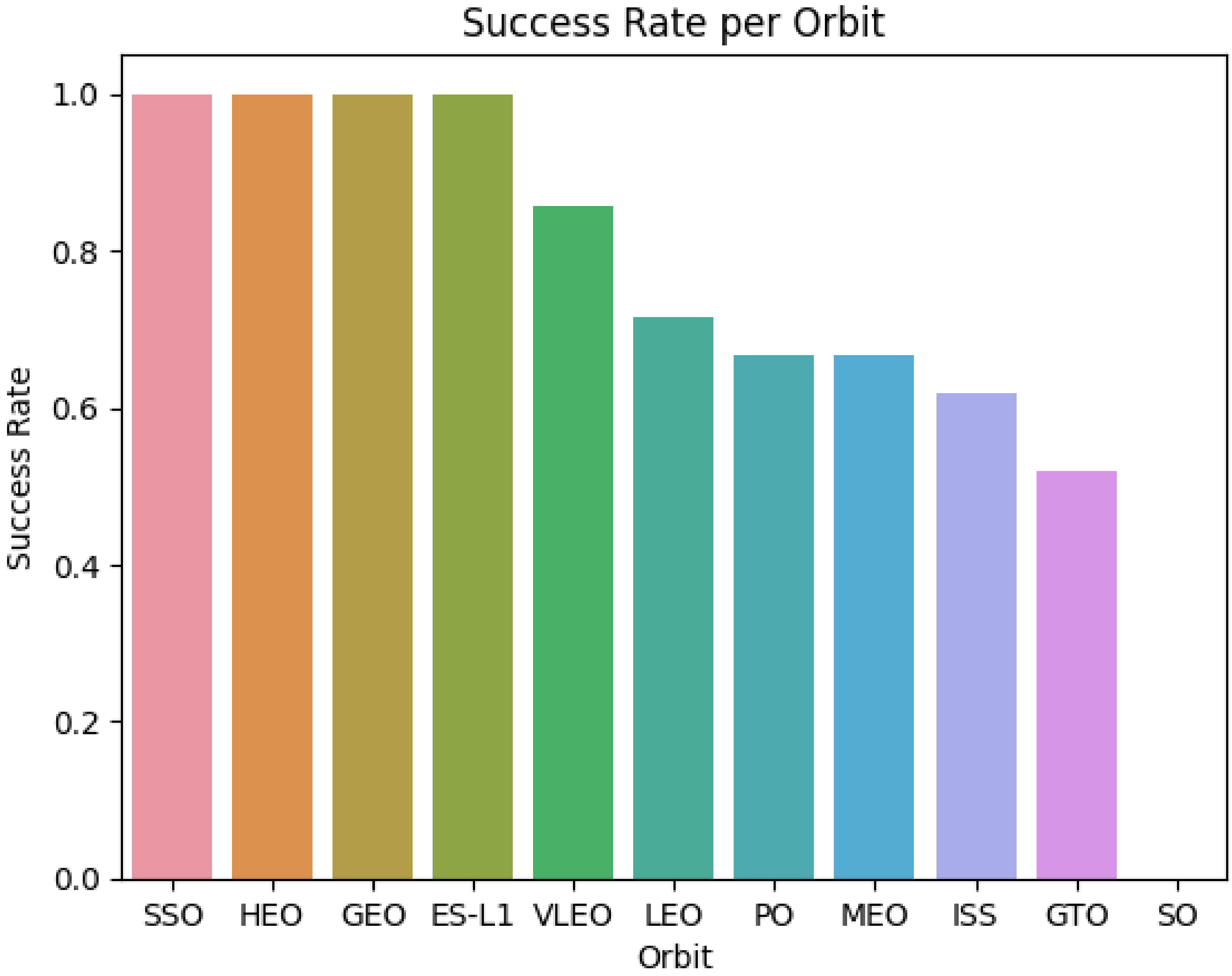
Flight Number vs. Launch Site

- We notice that the success rate for each site is increasing.
- We notice that some sites launch more rockets than the others.



Payload vs. Launch Site

- Higher payload masses generally increase launch success rates, although excessively heavy payloads can adversely affect landings, varying by launch site.



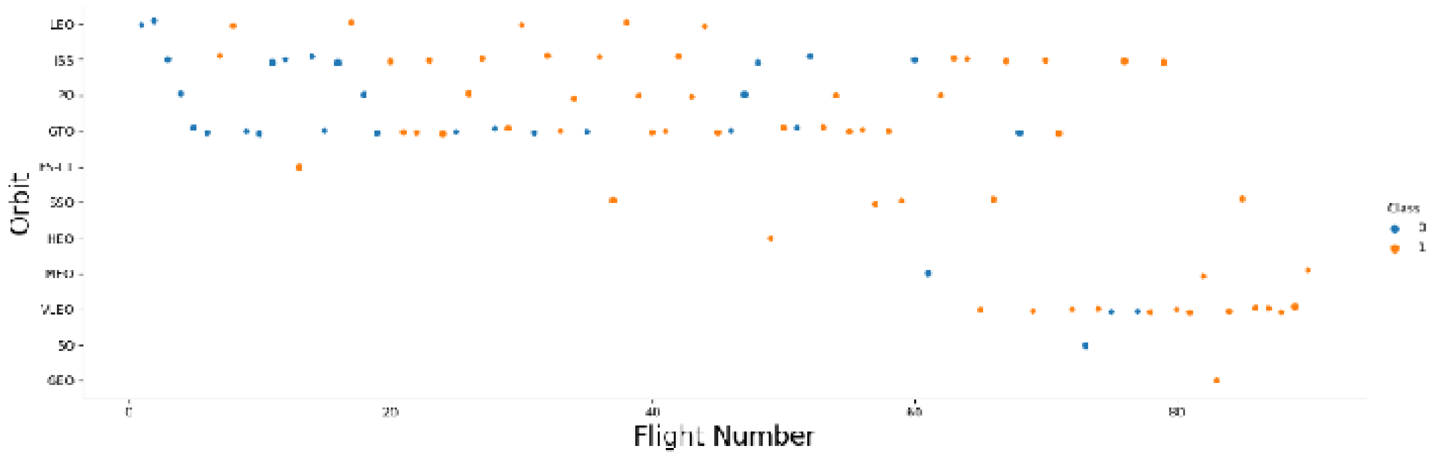
Success Rate vs. Orbit Type

- SO has 0% success rate
- ES-L1, GEO, HEO, SSO have the highest success rate.

Flight Number vs. Orbit Type

- The success rate for LEO orbit missions correlates positively with an increasing number of flights.

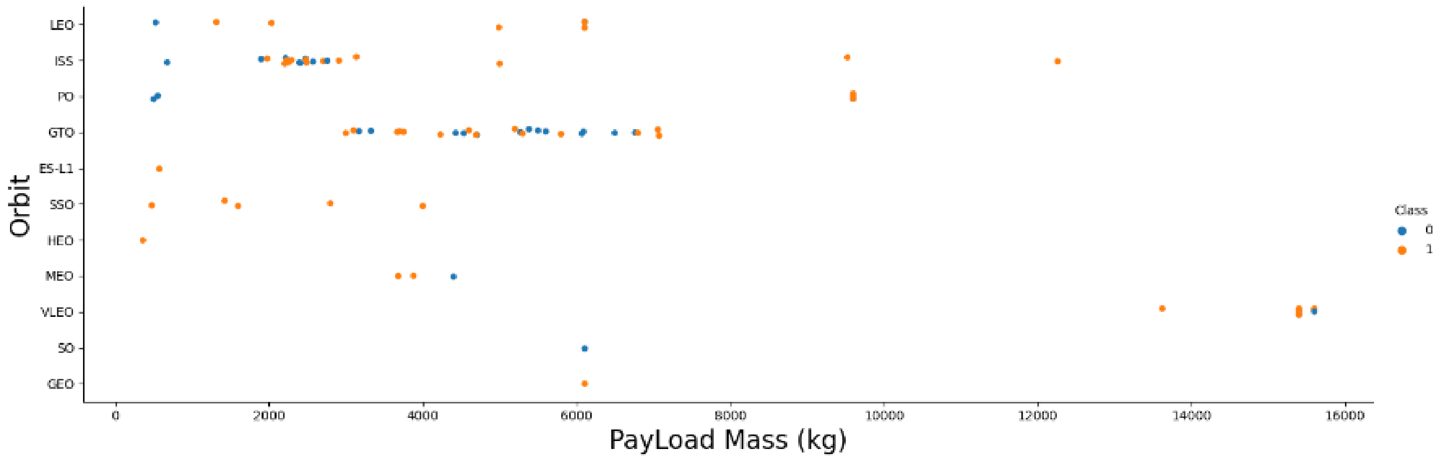
- No clear success rate correlation for GTO orbit flights, while knowledge from previous missions may contribute to higher success rates in SSO and HEO orbits.



Payload vs. Orbit Type

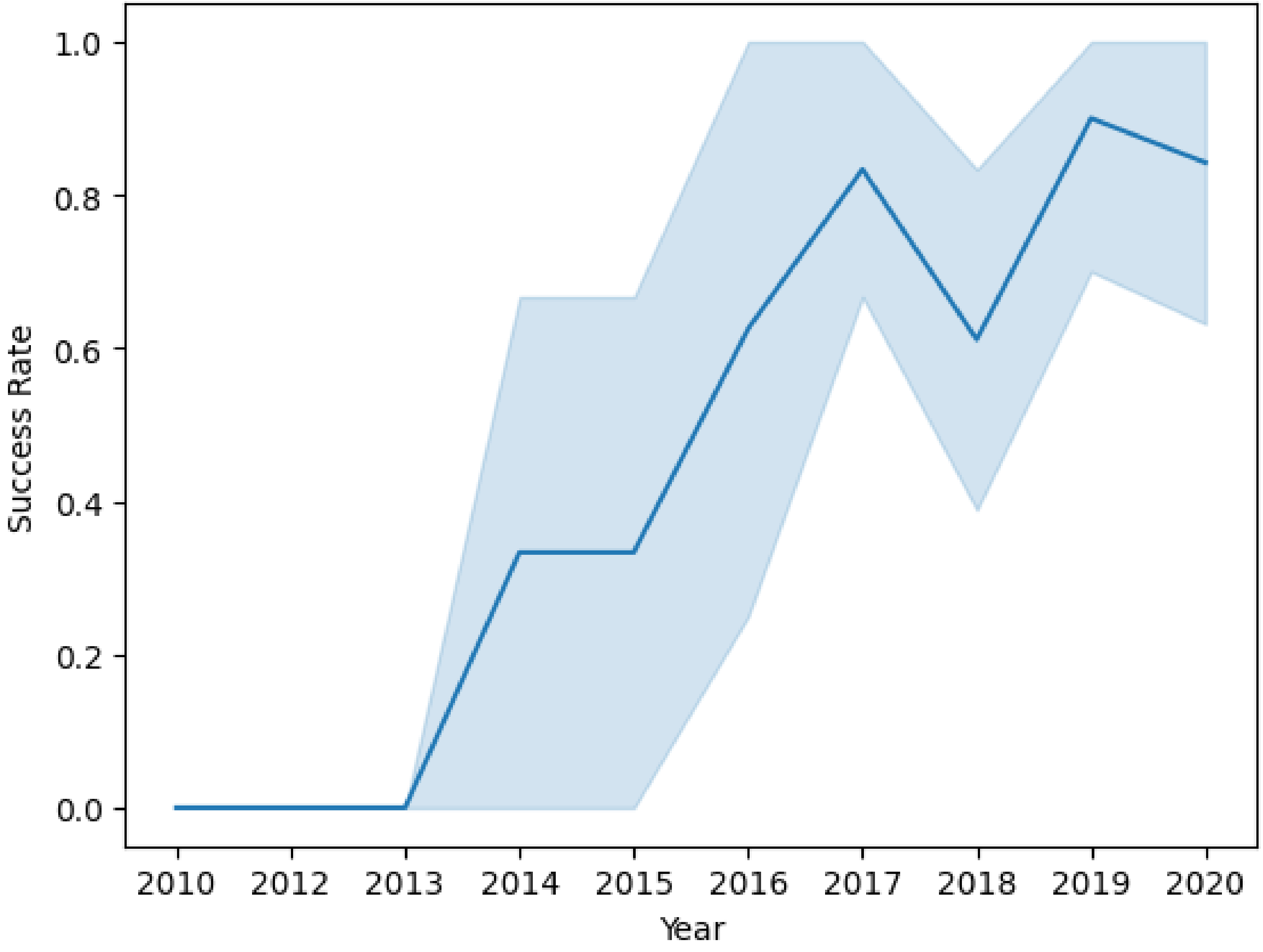
- Heavier payloads enhance launch success rates in LEO orbits, while lighter payloads benefit GTO orbit launches.

- Payload weight impacts GTO orbits negatively but contributes positively to GTO and Polar LEO (ISS) orbit success.



Launch Success Yearly Trend

- The SpaceX rocket success rate has been on an upward trend from 2013 through 2020.



All Launch Site Names

Task 1

Display the names of the unique launch sites in the space mission

```
%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

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string '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	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

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

```
%sql select SUM(PAYLOAD_MASS__KG_) as Total_payload from SPACEXTBL where customer = 'NASA (CRS)' ;
```

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

Total_payload
45596

Average Payload Mass by F9 v1.1

Calculate the average payload mass carried by booster version F9 v1.1

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

```
%sql select AVG(PAYLOAD_MASS__KG_) as Avg_payload_mass from SPACEXTBL where booster_version ='F9 v1.1';
```

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

Avg_payload_mass
2928.4

First Successful Ground Landing Date

Find the dates of the first successful landing outcome on ground pad

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

Hint:Use min function

```
%sql select MIN(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)';
```

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

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select DISTINCT(Booster_Version) from SPACEXTBL where Landing_Outcome = 'Success (drone ship)' and (PAYLOAD_MASS__KG_ 
```

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

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

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

List the total number of successful and failure mission outcomes

```
%sql select count(*) from SPACEXTBL as Total_successful_missions where mission_outcome = 'Success'
```

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

count(*)
98

Boosters Carried Maximum Payload

List the names of the booster which have carried the maximum payload mass

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql select BOOSTER_VERSION from SPACEXTBL where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL)
```

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

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql select landing__outcome, BOOSTER_VERSION, LAUNCH_SITE from SPACEXTBL where (LANDING__OUTCOME = 'Failure (drone ship)') and (YEAR(date)=2015)
```

```
 * ibm_db_sa://zbn26140:***@dashdb-txn-sbox-yp-dal09-11.services.dal.bluemix.net:50000/BLUDB
Done.
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- 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

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(*) from SPACEXTBL where date between '2010-06-04' and '2017-03-20' group by (Landing_Out
```

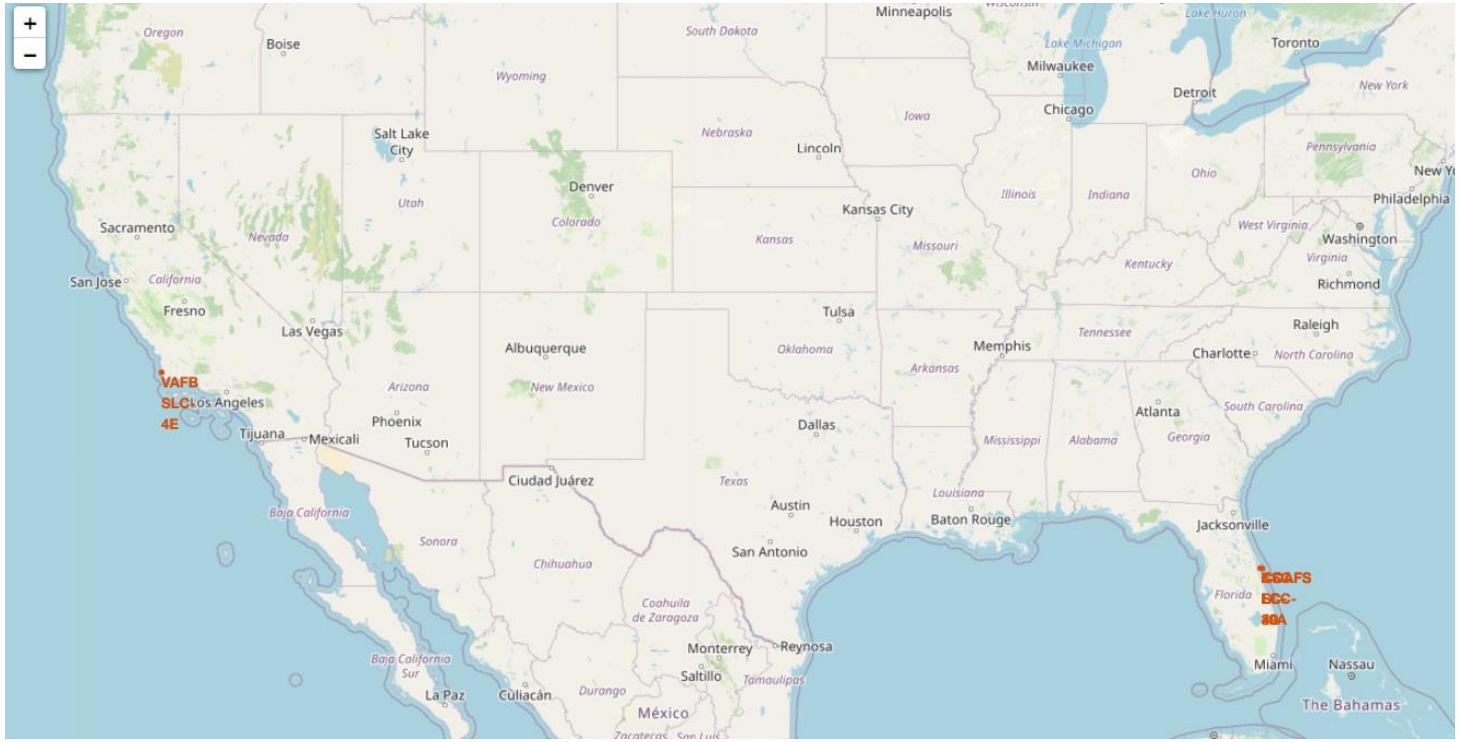
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	count(*)
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

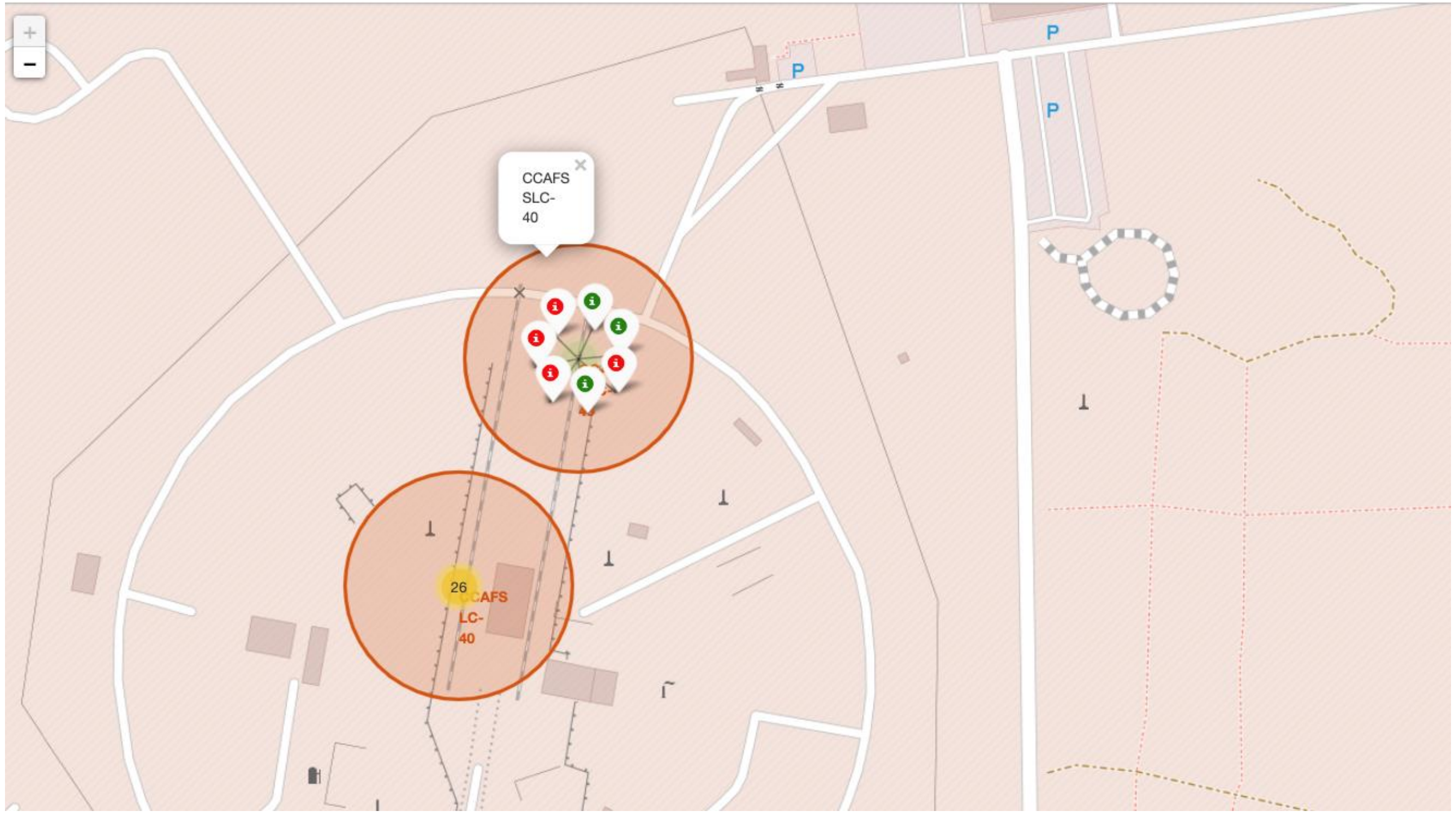
Folium - Launch Sites



- Launch sites are strategically located near the equator to utilize the Earth's rotational speed for orbital launches
- Launch sites are close to coasts to reduce debris risks over populated areas.

Successful / Failed Launches

- The map shows the successful and failed launches for the site.



Folium Map – Site Proximity

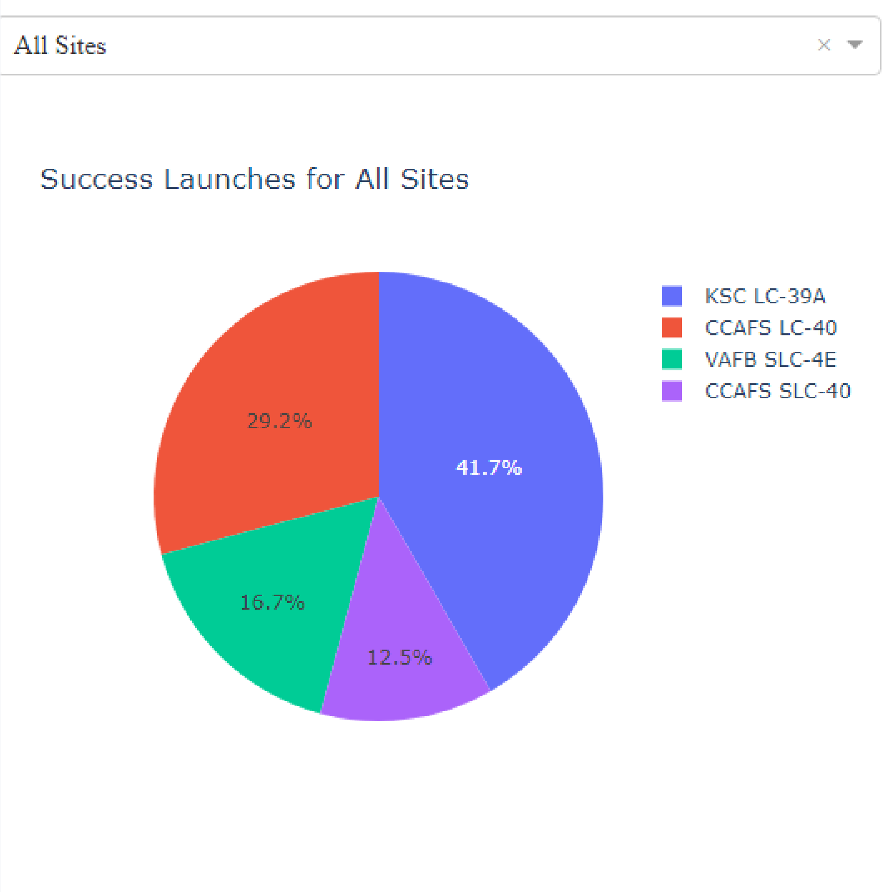
- The map shows the proximity of the launch site to the nearest coastline

Section 4

Build a Dashboard with Plotly Dash

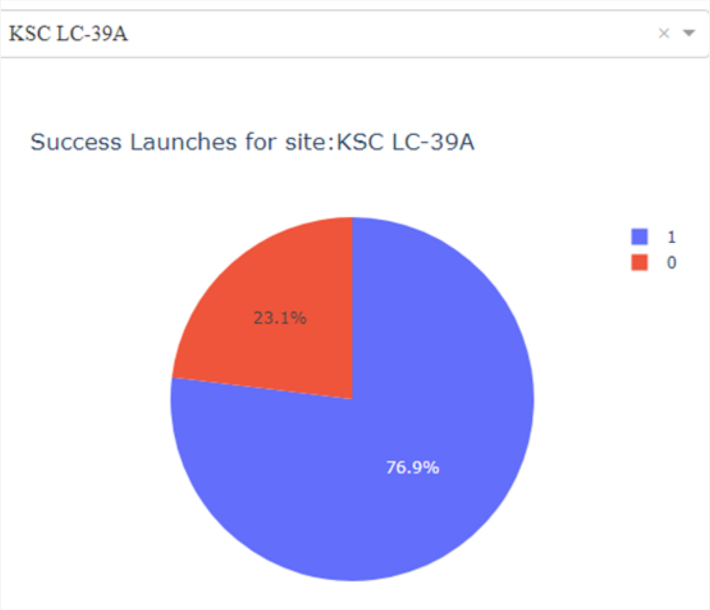
Space X Launch Record Dashboard

- KSC LC-39A has the best success rate of launches.



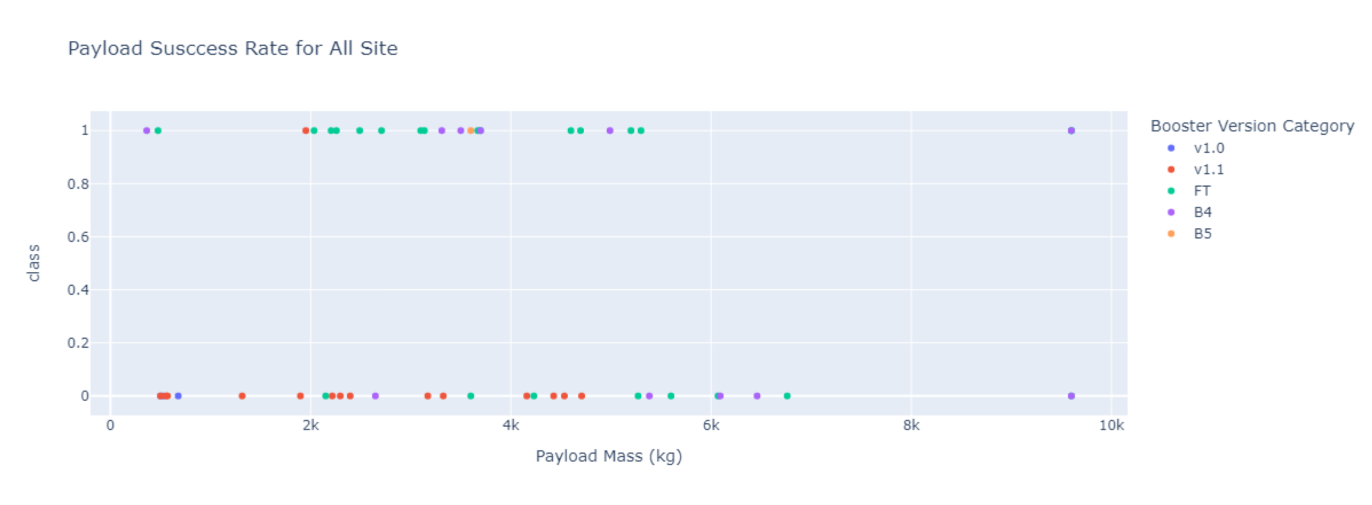
Space X Launch Record Dashboard – KSC LC-39A

- KSC LC-39A has a success of 76.9% and failure of 23.1 %.

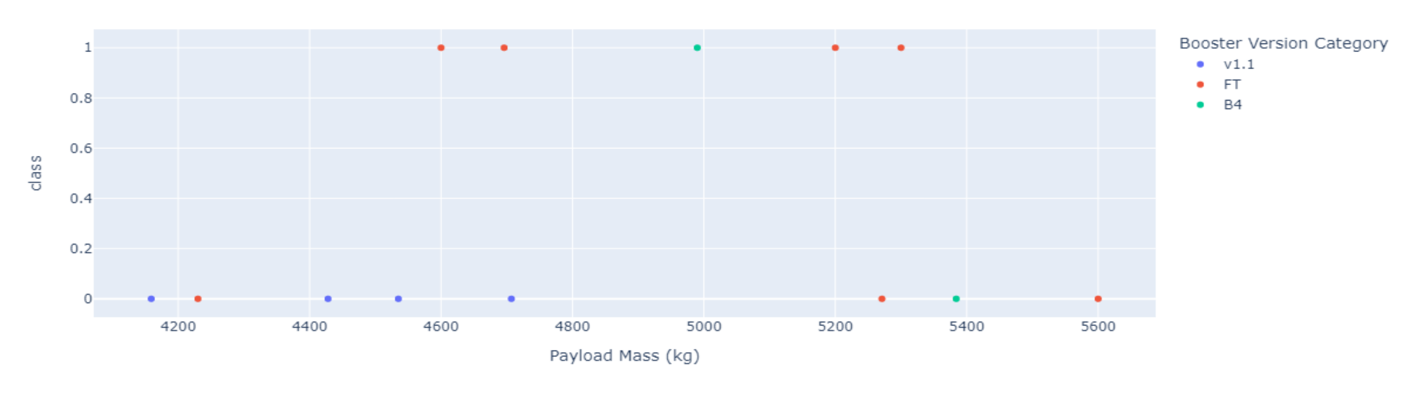


Payload success rate for all sites

- The plot shows the success rate vs payload mass for each booster version.



- The slider allows us to change the payload range to limit payloads between 4000 and 5600 kgs



Section 5

Predictive Analysis (Classification)

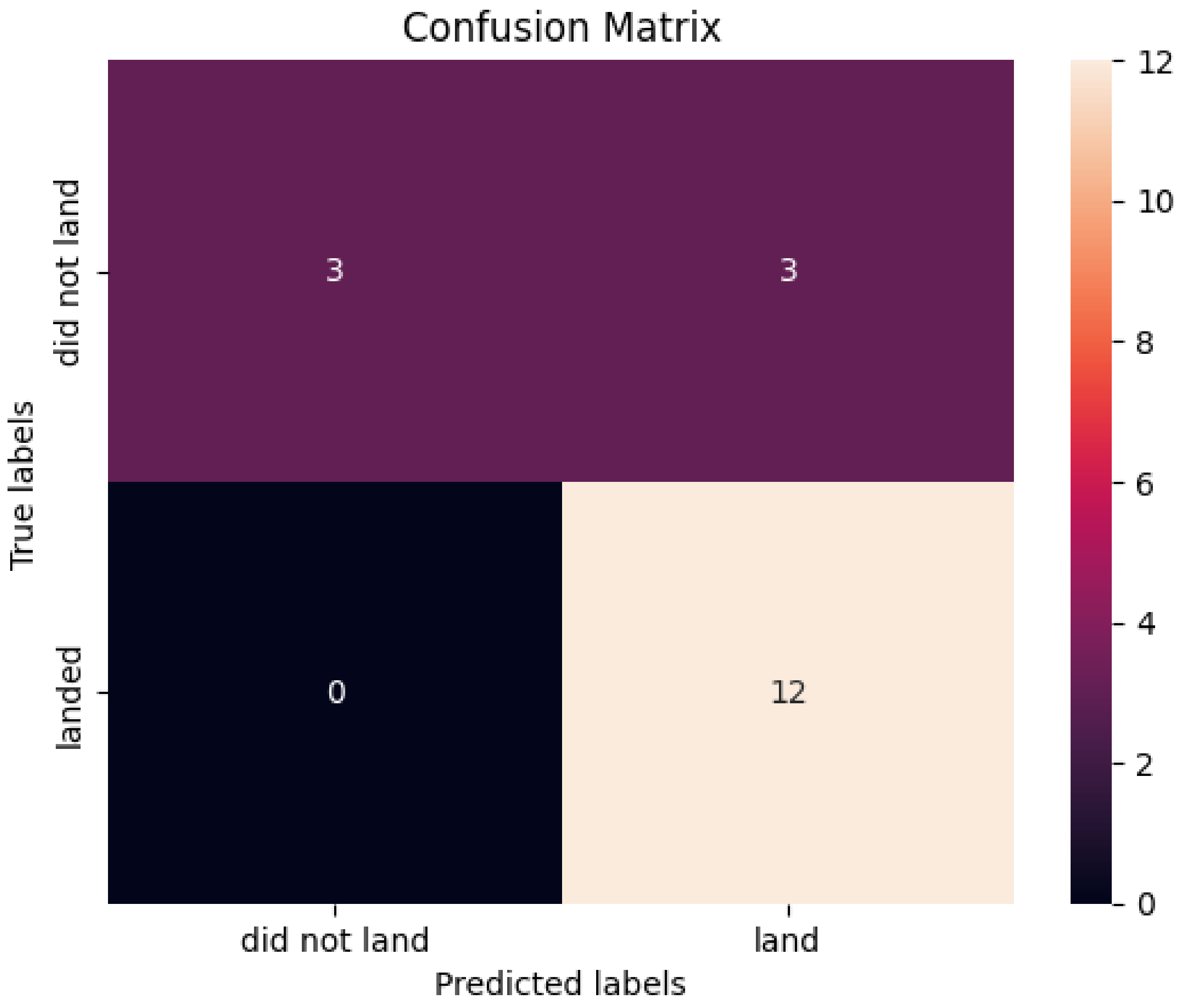
Classification Accuracy

- Upon analyzing the performance metrics derived from the Test Set, it is challenging to definitively ascertain the superior method due to closely aligned results.
- This ambiguity in performance superiority may stem from the limited scope of the Test Set —too small a number to yield statistically significant conclusions.
- To circumvent this limitation and obtain a more reliable evaluation, we expanded our testing to encompass all available methods across the entirety of the Dataset, allowing for a more comprehensive and robust assessment of each method's efficacy.

	Accuracy Train	Accuracy Test
Tree	0.876786	0.833333
Knn	0.848214	0.833333
Svm	0.848214	0.833333
Logreg	0.846429	0.833333

Confusion Matrix

- KNN gave the best performance with the highest true positives and the lowest false negatives



Conclusions

- Launch success rates vary by site, payload, and orbit.
- SpaceX's launch success rates are trending upwards.
- Launch sites are equatorial with rail and coastal access.
- The KNN classifier most accurately predicts launch failures.
- GEO, HEO, SSO, and ES L1 orbits have the highest success rates.
- Optimal payload mass varies by orbit, with lighter payloads generally more successful.
- Launch success rates have increased yearly.
- KSC LC-39A boasts the highest launch success rate.

Appendix

- GitHub URL

Thank you!