IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Arun Asok, PhD>
<09-22-2023>

Outline

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

Executive Summary

- Overview: The commercialization of space flight has reached a tipping point in recent years. Here, we sought to determine if SpaceX will reuse the first stage of their rocket using public information.
- Methods: We use a variety methods to gather launch and rocket data from the SpaceX REST API endpoints and analyzed these data with advanced graphical and machine learning methods.
- Conclusion: In general, CCAFS LS-40 appears to be the optimal site for success across multiple parameters, however more data is needed to understand why in relation to the Falcon 9 booster.

Introduction

- Background: Several private companies including Virgin Galactic, Rocket Labs, Blue Origin, and Space X have chartered the privatization of space flight. SpaceX is unique relative to other companies given that they can reuse the first stage booster of their Falcon 9 rocket which in turn substantially reduces the cost (~62 million) relative to competitors (~high of 165 million).
- We sought to ask: (1) What is the relative success of landing the stage 1 booster as (2) a proxy for cost efficiency (in spite of several variables which may dictate reuse). Furthermore, we (3) also wanted to know if there were any unique characteristics the contributed to success across geographic, booster version, and time domains that may contribute to increased success for SpaceY.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- Perform data wrangling
- 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 were collected from two freely available public sources: (1) the SpaceX REST API endpoints and (2) web-scrapped from Wikipedia.
- key phrases and flowcharts on the following pages

Data Collection - SpaceX API

- https://github.com/aasok11/Public/blob/7adc3d9b84b508ec14c0b3e82baead7e0396d572/jupyter-labs-spacex-data-collection-api.ipynb

1. Define URL of SpaceX API containing Falcon 9 launches
2. Extract and download the requested data
3. Transform into a Pandas DF.
4. Normalize data and clean missing values.
5. Convert and save to useable csv file.

lace your flowchart of web scraping here

Data Collection – Scraping

- https://github.com/aasok11/Public/blob/7adc3d9b84b508ec14c0b3e82baead7e0396d572/jupyter-labs-webscraping.ipynb

1. Define URL of Wikipedia site to scrape Falcon 9 launches
2. Use Beautiful Soup to extract from html site.
3. Parse data and transform into a Pandas DF.
4. Label/Extract Relevant information Columns and clean missing columns
5. Convert and save to useable csv file.

Data Wrangling

- https://github.com/aasok11/Public/blob/7adc3d9b84b508ec14c0b3e82baead7e0396d572/jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

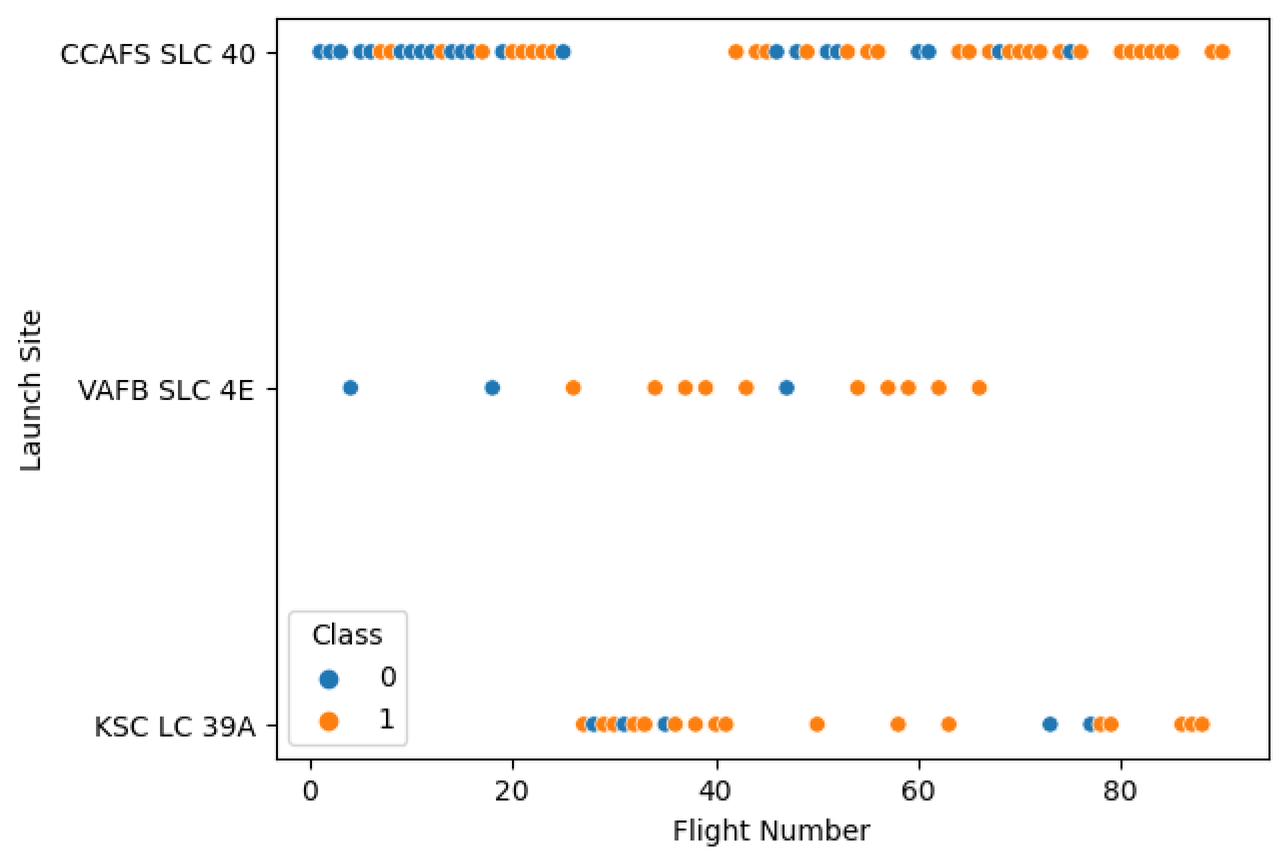
1. Load the API/scraped data.
2. Determine missing or null information
3. Replace null information with the mean and transform variables into integer objects.
4. Label/Extract Relevant information Columns and clean missing columns
5. Create relevant new summary variables (e.g., landing statistics, etc.)

Section 2

Insights drawn from EDA

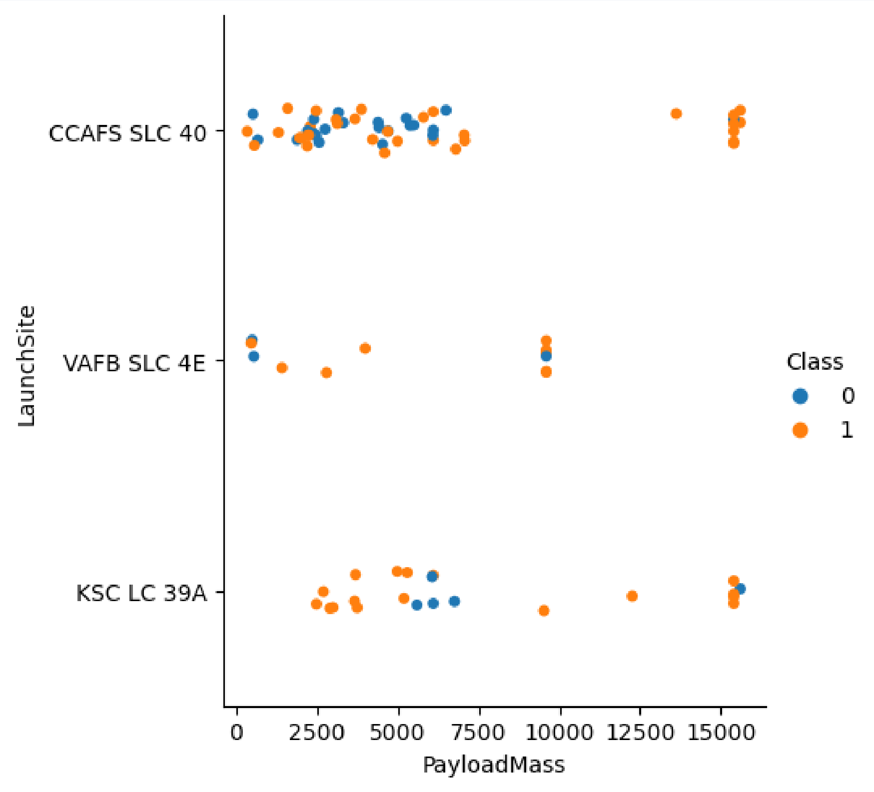
Flight Number vs. Launch Site Scatter Plot

- Figure 1. The scatter plot represents Flight success (orange circles, class 1) compared against flight failures (blue circles, class 2) as a function of flight number (x-axis) across launch sites (y-axis).
- ([Clickable Notebook Link]())

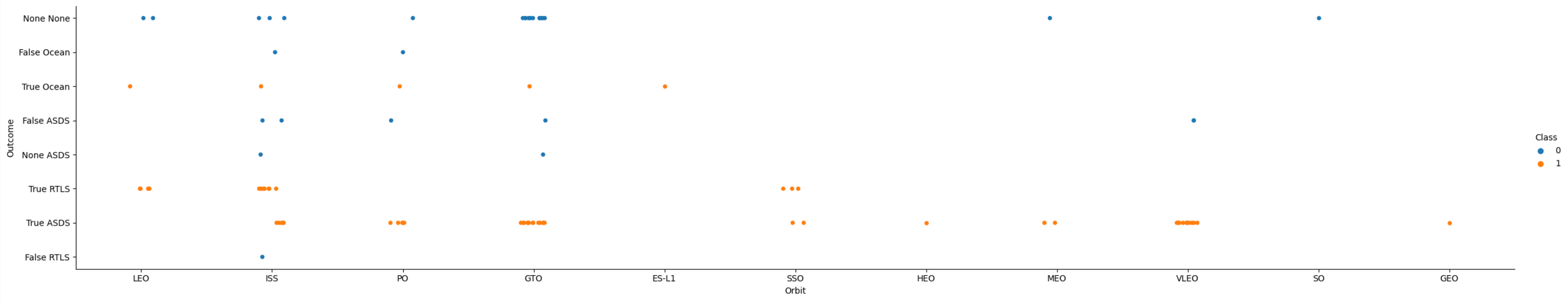


Payload vs. Launch Site

- Figure 2. The scatter plot represents Flight success (orange circles, class 1) compared against flight failures (blue circles, class 2) as a function of Payload mass (x-axis) across launch sites (y-axis).
- ([Clickable Notebook Link]())

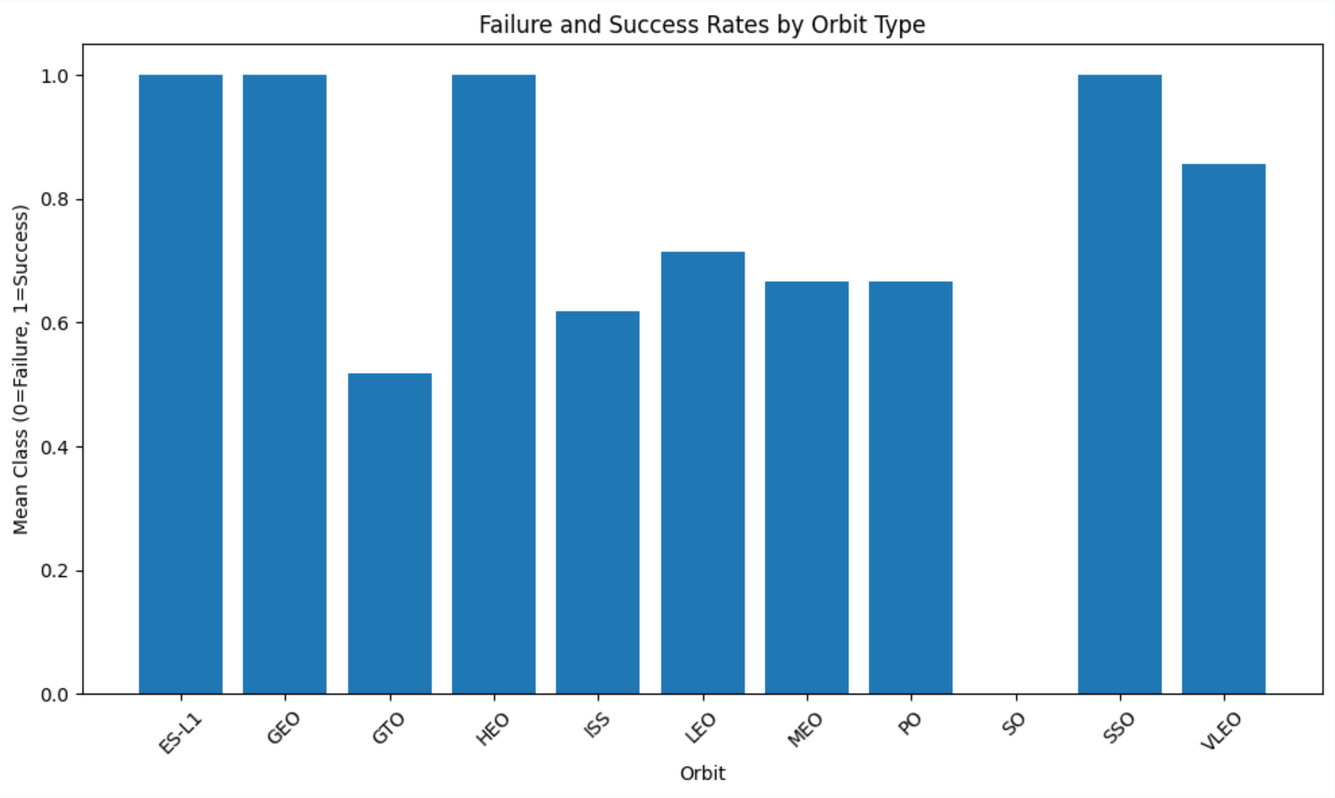


Success Rate vs. Orbit Type



- Figure 3. The scatter plot represents flight success (orange circles, class 1) compared against flight failures (blue circles, class 2) as a function of the type of orbit reached (x-axis) and the type of success or failure (y-axis).
- (Clickable Notebook Link)

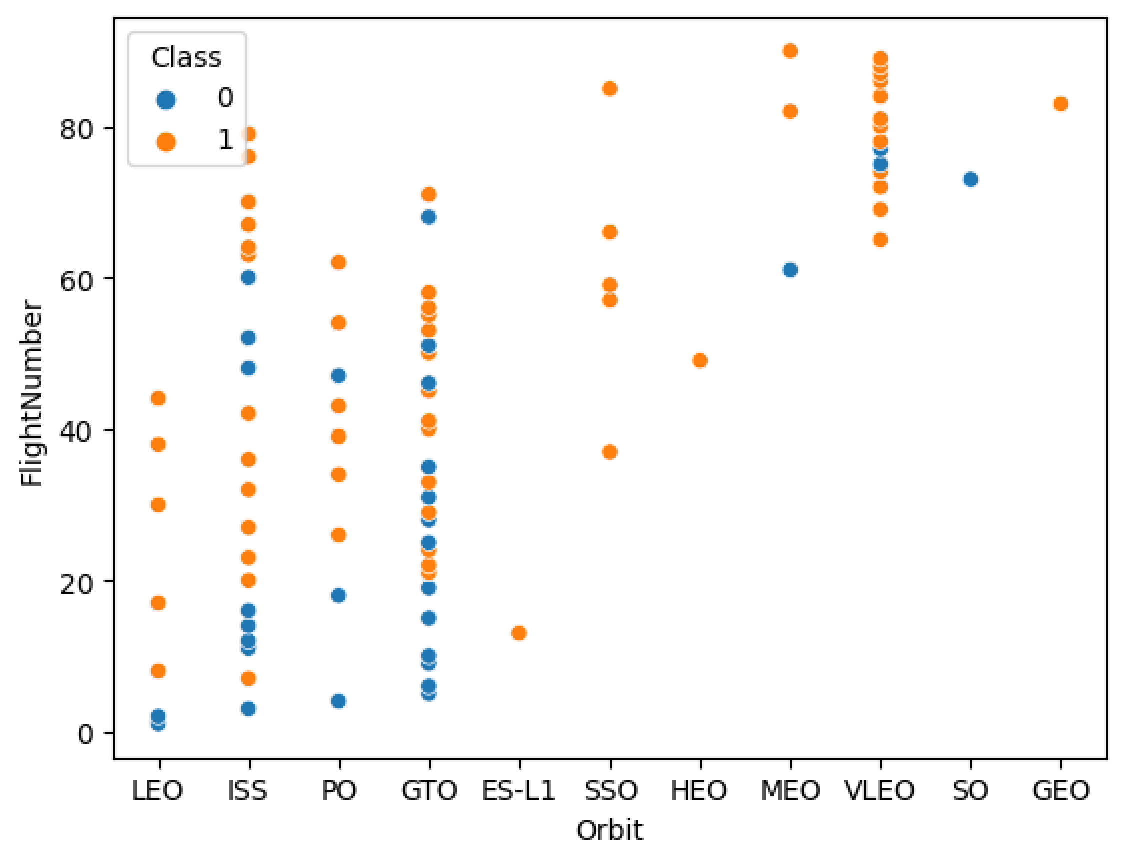
Success Rate vs. Orbit Type



- Figure 4. The box plot represents mean success rate (y-axis, 0 or 1) as a function of the type of orbit reached (x-axis).
- (Clickable Notebook Link)

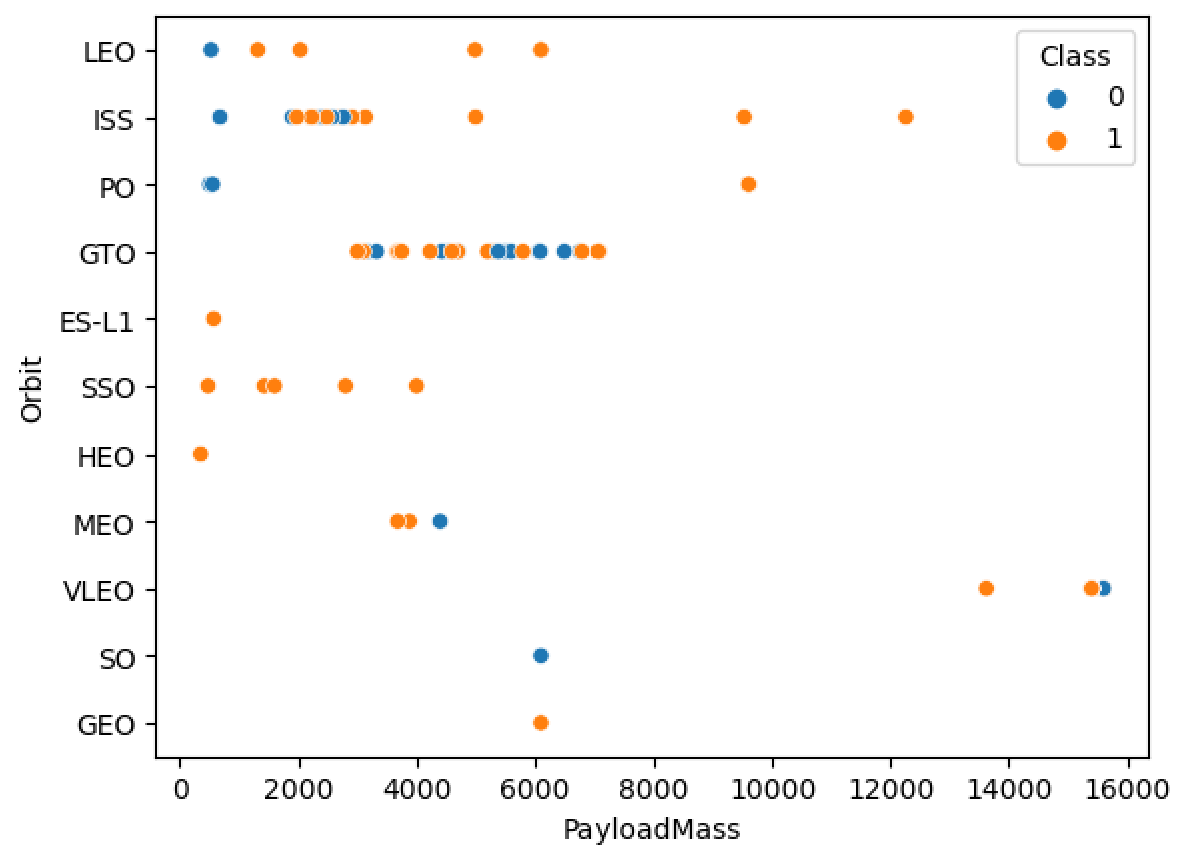
Flight Number vs. Orbit Type

- Figure 5. The scatter plot represents flight success (orange circles, class 1) compared against flight failures (blue circles, class 2) as a function of the type of orbit reached (x-axis) and the type of success or failure (y-axis).
- ([Clickable Notebook Link]())

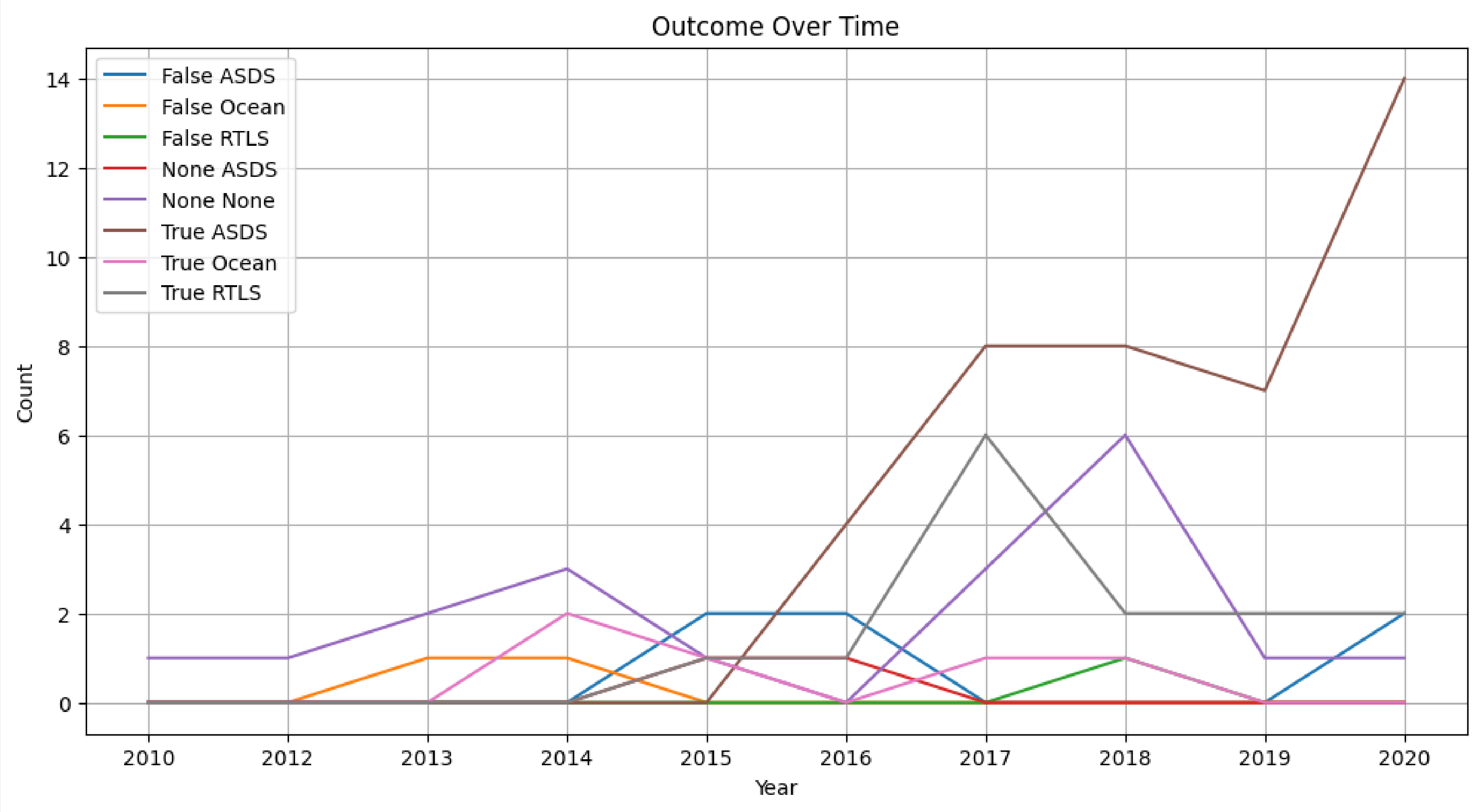


Payload vs. Orbit Type

- Figure 6. The scatter plot represents flight success (orange circles, class 1) compared against flight failures (blue circles, class 2) as a function of the type of payload mass (x-axis) and the orbit reached (y-axis).
- ([Clickable Notebook Link]())



Launch Success Yearly Trend



- Figure 6. The line chart represents the specific type of flight success or failure (individually colored lines) across years (x-axis) and the overall number (y-axis).
- ([Clickable Notebook Link])

All Launch Site Names

- This SQL query represents each of the launch sites used that are available in the database.
- (Clickable Notebook Link)

Task 1

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

In [10]:
```
%sql select distinct "Launch_Site" from SPACEXTABLE
```

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

Out[10]:

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

Launch Site Names Begin with 'CCA'

Task 2

Display 5 records where launch sites begin with the string 'CCA'

[10]:
```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

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

[10]:

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

- This SQL query represents the first five launch sites in the database that start with ‘CCA’.
- (Clickable Notebook Link)

Total Payload Mass

Task 3

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

```
[14]: %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Customer LIKE 'NASA%'
 * sqlite:///my_data1.db
Done.
```

[14]:

sum(PAYLOAD_MASS__KG_)
99980

- This SQL query represents the total payload mass carried by boosters launched by NASA.
- (Clickable Notebook Link)

Average Payload Mass by F9 v1.1

Task 4

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

```
[17]: %sql select avg(PAYLOAD_MASS__KG_) as average_value from SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'

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

[17]:

average_value
2928.4

- This SQL query represents the average payload mass carried by Falcon 9 v1.1 boosters.
- (Clickable Notebook Link)

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

```
[24]: %sql SELECT MIN(Date) AS first_successful_ground_pad_landing_date FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)'

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

[24]:

first_successful_ground_pad_landing_date
2015-12-22

- This SQL query represents the first date from the database of launches in which a successful landing on a ground pad were achieved.
- (Clickable Notebook Link)

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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

[23]:
```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

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

[23]:

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

- This SQL query represents which specific Falcon 9 boosters with a payload mass between 4000-6000 have successfully landed on a drone ship.
- (Clickable Notebook Link)

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%sql SELECT Landing_Outcome, COUNT(*) as total_count FROM SPACEXTABLE WHERE Landing_Outcome LIKE 'Success%' OR Landing_Outcome LIKE 'Failure%' GROUP BY Landing_Outcom
```

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

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

- This SQL query represents the total number of landing success or failures stratified against type the type of target (e.g., drone ship, etc.) outcome.
- (Clickable Notebook Link)

Boosters Carried Maximum Payload

- This SQL query lists the Falcon 9 booster versions that have carried the maximum payload mass.
- ([Clickable Notebook Link])

Task 8

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

```
[28]: %sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
```

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

[28]:

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

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

Task 10

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.

[61]:
```
%sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count, RANK() OVER (ORDER BY COUNT(*) DESC) AS Outcome_Rank FROM SPACEXTABLE WHERE Landing_Outcome IN ('Failure (dr
```

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

[61]:

Landing_Outcome	Outcome_Count	Outcome_Rank
Success (ground pad)	5	1
Failure (drone ship)	5	1

Would you like to receive official Jupyter

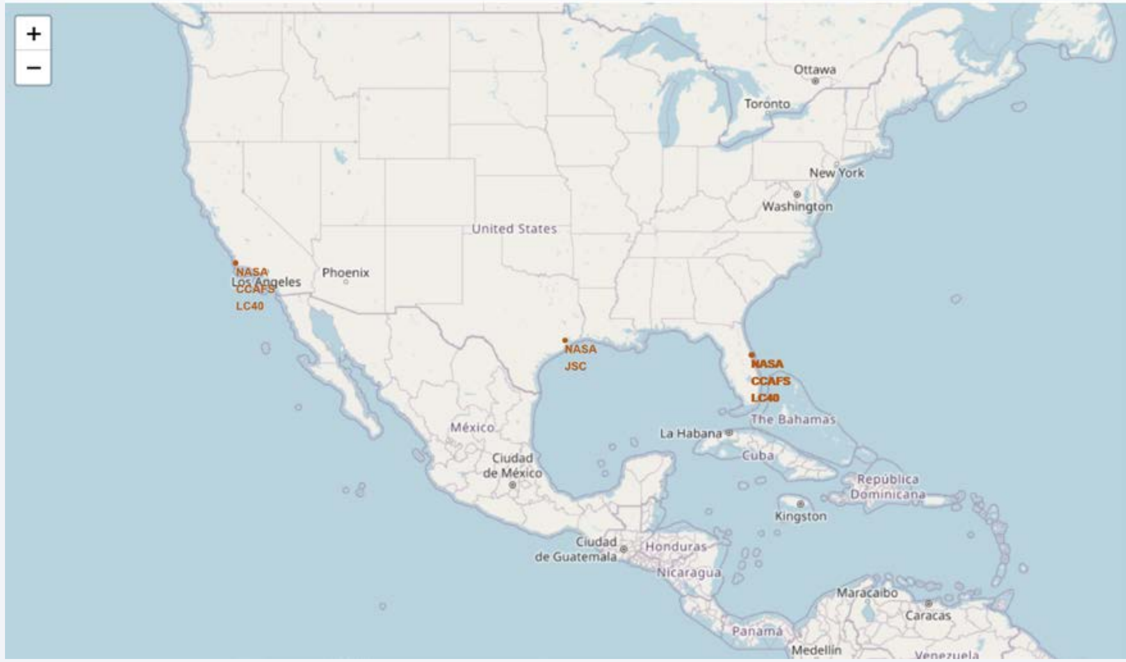
- This SQL query Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- (Clickable Notebook Link)

Section 3

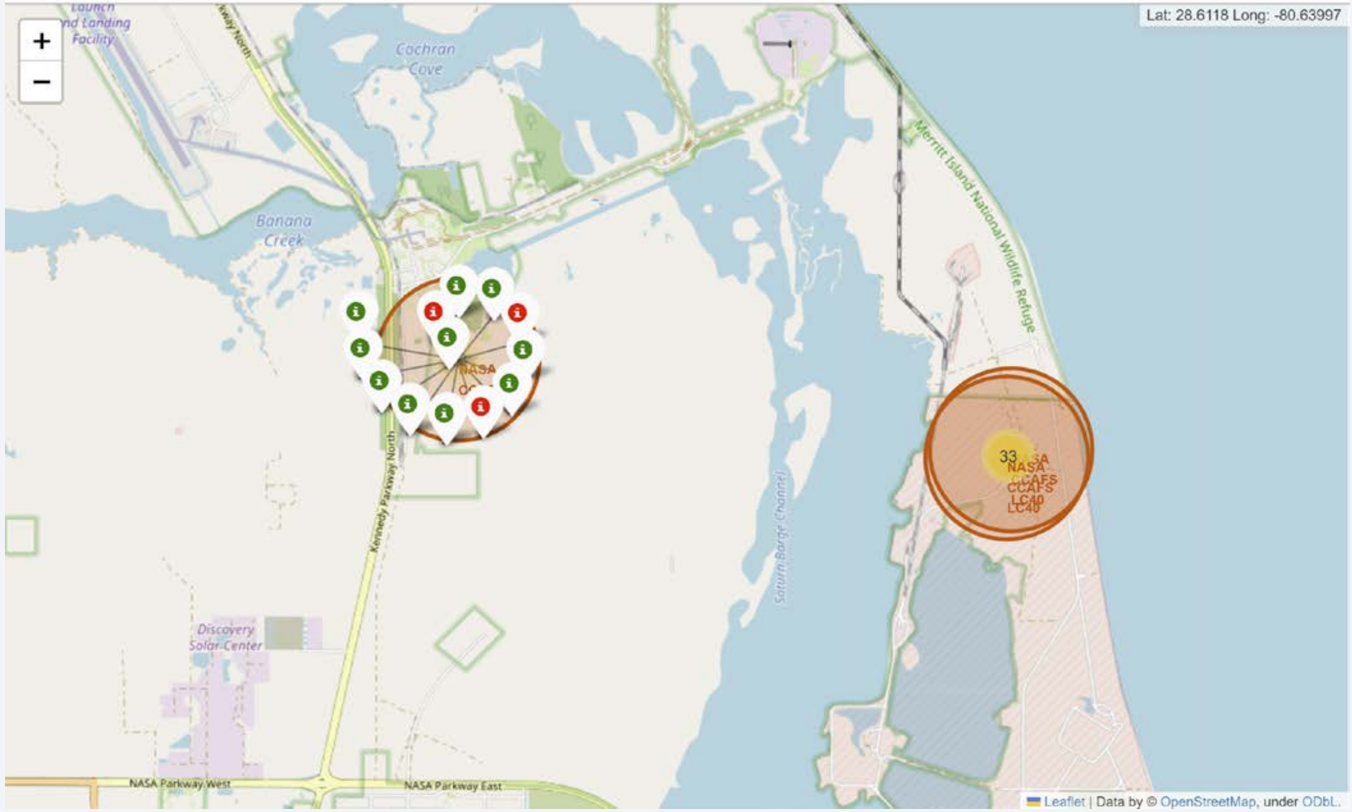
Launch Sites Proximities Analysis

SpaceX Launch Sites

- This Folium map represents all launch sites which happen to be restricted to US locations. NASA CCAFS contains 2 sites in very close proximity.
- ([Clickable Notebook Link](#))

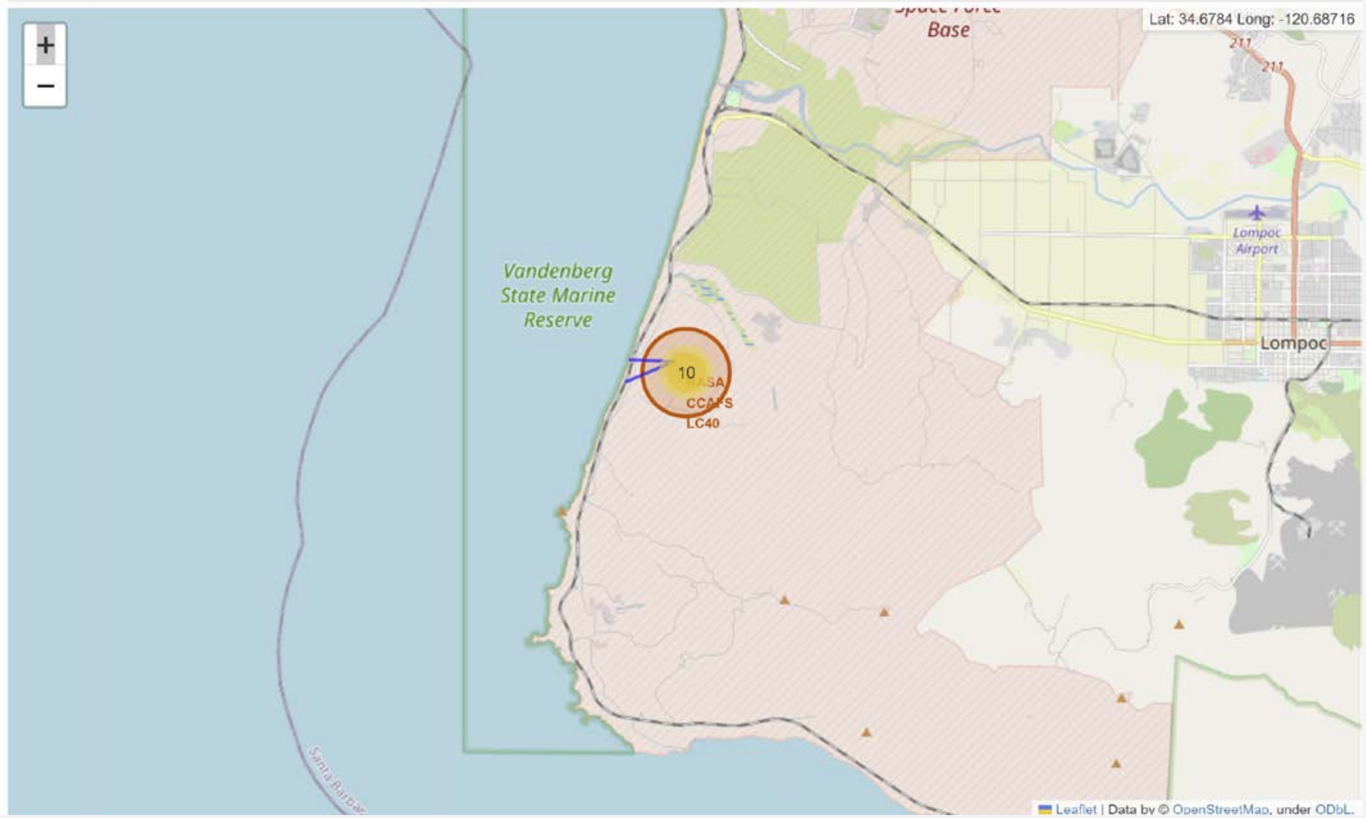


Launch Site Success and Failure Rate



- This Folium map represents a drilldown of the success (green markers) and failures (red markers) for an individual launch site using NASA CCAFS as a representative site.
- (Clickable Notebook Link)

Launch Site Proximity to Coastlines and Railways

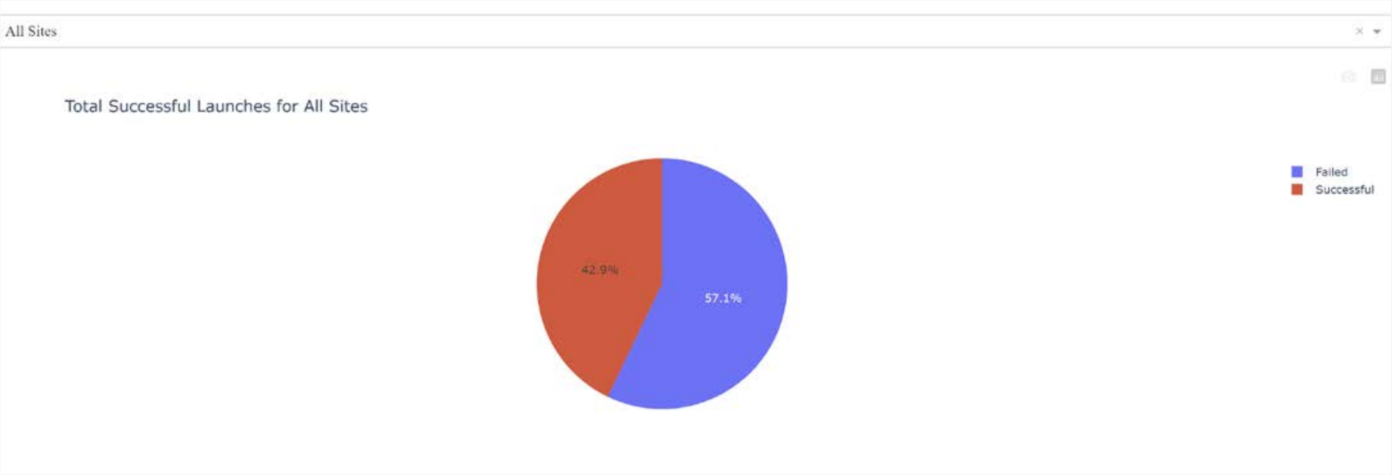


- This Folium map represents the proximity (via blue lines) of a representative launch site (Vandenburg) to both the coastline and a railway.
- (Clickable Notebook Link)

Section 4

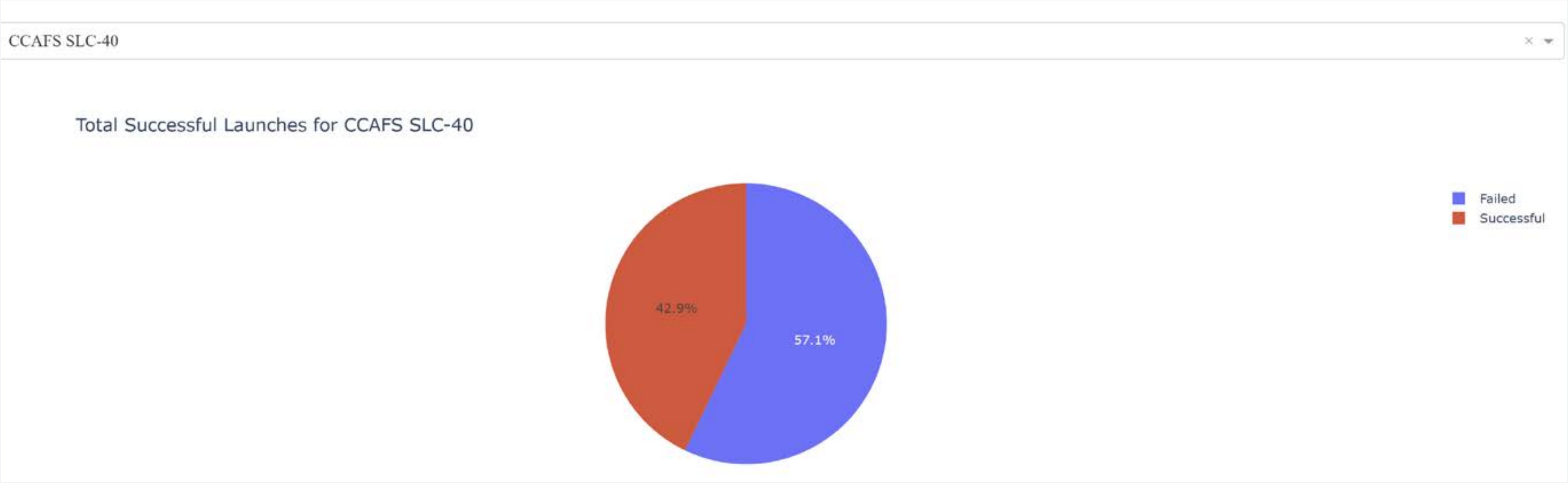
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard



- Figure. A plotly dashboard pie chart showing the percentage of successful (red) launches compared against failed (blue) launches.
- ([Clickable Notebook Link])

SpaceX Launch Records Dashboard



- Figure. A plotly dashboard pie chart showing the percentage of successful (red) launches compared against failed (blue) launches for the most successful launch site.
- ([Clickable Notebook Link]())

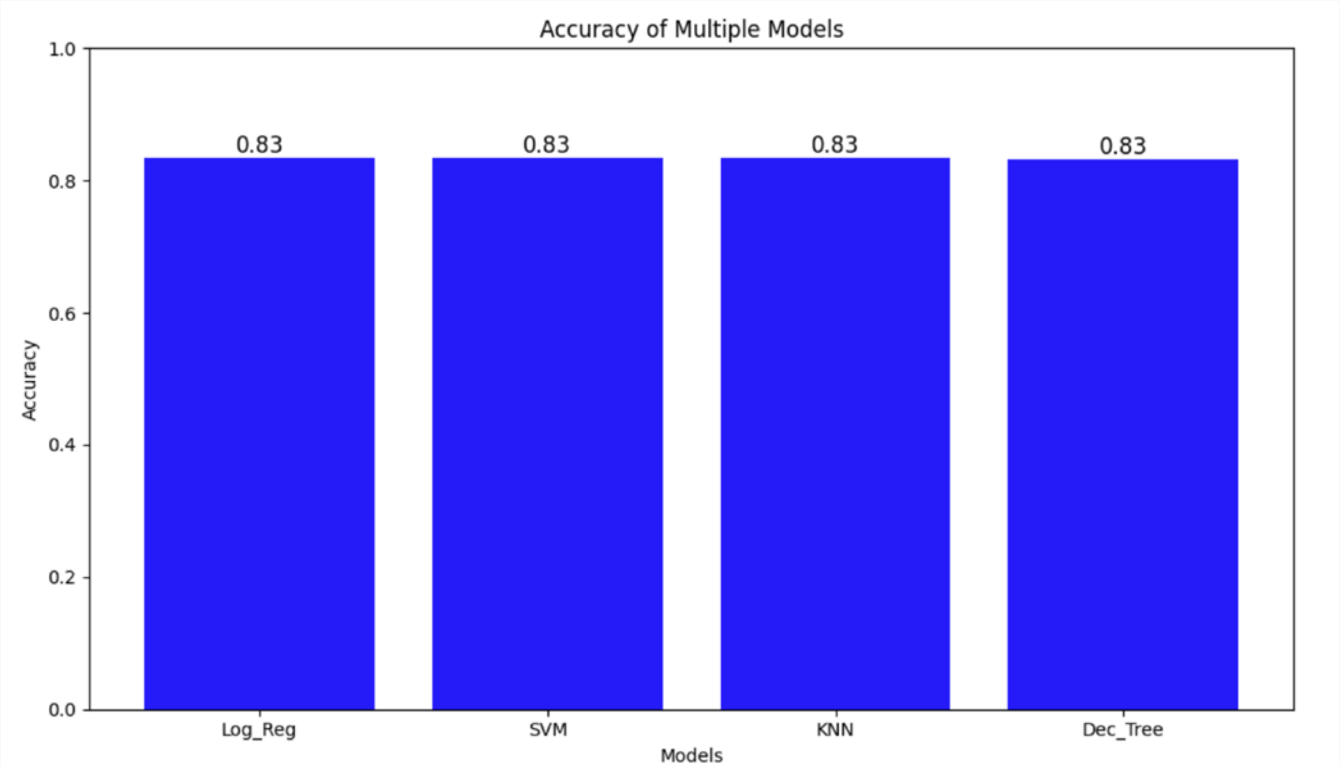
< SpaceX Launch Records Dashboard >

- Figure. A plotly dashboard showing the correlation between launch success (y-axis) against payload mass (x-axis) for each booster version (legend; circle colors).
- ([Clickable Notebook Link](#))

Section 5

Predictive Analysis (Classification)

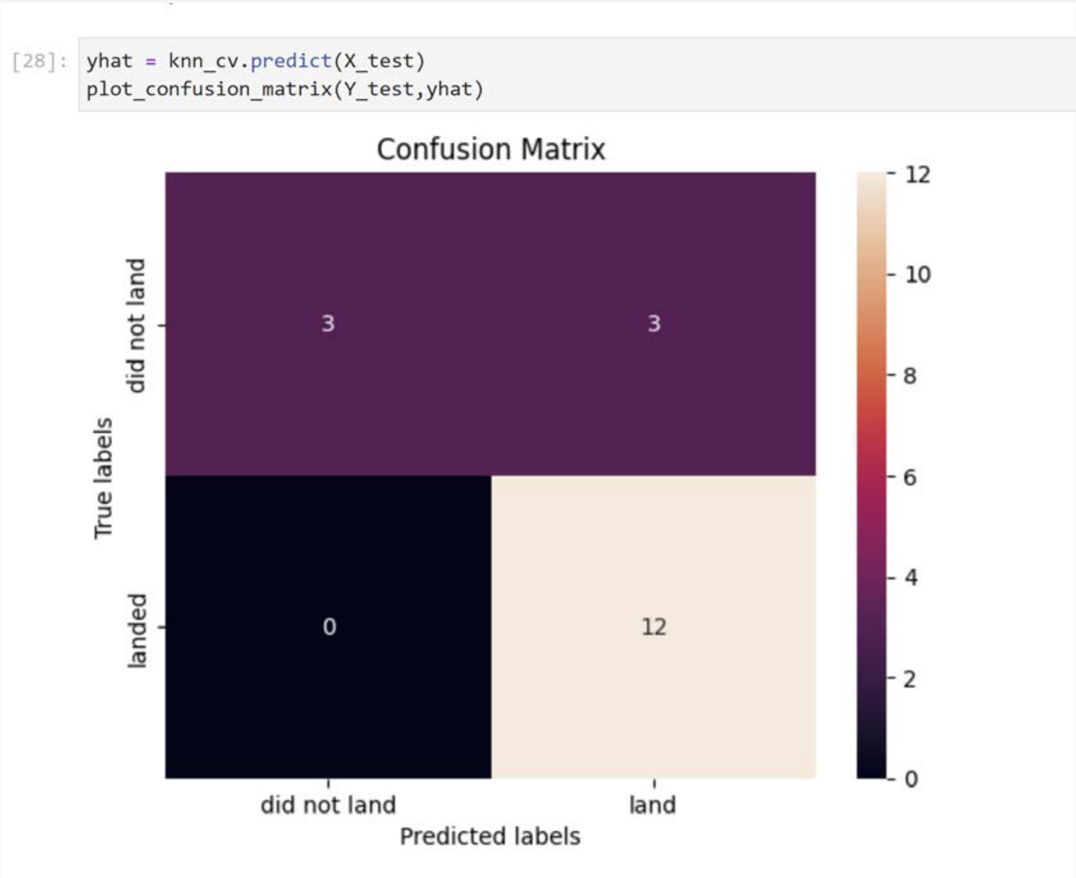
Classification Accuracy



- Figure. A bar chart of each machine learning model (x-axis) based on its accuracy (y-axis).
- (Clickable Notebook Link)

Confusion Matrix

- Figure. A confusion matrix of the k-nearest neighbor model for predicting values.
- (Clickable Notebook Link)



Conclusions

- CCAFS LS-40 is the launch site with greatest rate of launch success.
- VLEO orbits have the greatest success.
- Multiple Falcon 9 boosters have carried maximum payloads.
- Launch sites are in close proximity to coastlines and railways.
- Across various Machine Learning models, the specific hyperparameters contained in the dataset and used across models do not seem to dramatically increase accuracy over one another.
- More data on other variables should be examined to increase the overall success rate across rockets.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!