IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Pham Thi Ngoc Ha
December 3rd, 2024

Outline

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

Executive Summary

- Summary of methodologies
  - SpaceX Data Collection using SpaceX API
  - SpaceX Data Collection with Web Scraping
  - SpaceX Data Wrangling
  - SpaceX Exploratory Data Analysis using SQL
  - SpaceX EDA DataViz Using Python Pandas and Matplotlib
  - SpaceX Launch Sites Analysis with Folium-Interactive Visual Analytics and Plotly Dash
  - SpaceX Machine Learning Landing Prediction
- Summary of all results
  - EDA results
  - Interactive Visual Analytics and Dashboards
  - Predictive Analysis (Classification)

Introduction

- Project background and context:
  - SpaceX advertises Falcon 9 rocket launches on its website for $62 million, significantly lower than the $165 million charged by other providers. This cost efficiency is largely due to SpaceX's ability to reuse the rocket's first stage. By determining whether the first stage will successfully land, we can estimate the cost of a launch. This insight could be valuable for competitors aiming to bid against SpaceX for rocket launch contracts.
- Problems you want to find answers:
  - The goal of this capstone project is to predict whether the Falcon 9 first stage will successfully land using data from its rocket launches as advertised on SpaceX's website.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

- How data sets were collected:
  - The dataset was collected by making GET requests to the SpaceX API to retrieve historical launch data. Static JSON responses were used for consistency and processed into a structured format for analysis.
  - The web scraping process involves sending an HTTP GET request to retrieve the Falcon 9 Wiki page, parsing the HTML with BeautifulSoup to locate the target table, and extracting column names and data rows. The data is cleaned, stored in a dictionary, and converted into a Pandas DataFrame for analysis.

Data Collection – SpaceX API

```
[Start]
  ↓
[Request: /v4/launches/past] → [Response: Launch Data JSON]
  ↓
[Extract IDs: Rocket, Launchpad, Payloads, Cores]
  ↓
+-----------------+   +---------------------+   +-------------------+   +-------------------+
| Rocket API:     |   | Launchpad API:      |   | Payloads API:     |   | Cores API:        |
| /v4/rockets/id  |   | /v4/launchpads/id   |   | /v4/payloads/id   |   | /v4/cores/id      |
| BoosterVersion  |   | Launch Site,        |   | Mass, Orbit       |   | Landing Outcome,  |
|                 |   | Longitude,          |   |                   |   | Grid Fins, etc. |
|                 |   | Latitude            |   |                   |   |                   |
+-----------------+   +---------------------+   +-------------------+   +-------------------+
       ↓                      ↓                         ↓                         ↓
  [Append to Lists]   [Append to Lists]         [Append to Lists]         [Append to Lists]
            ↓
         [Create DataFrame]
            ↓
          [End]
```

GitHub URL: https://github.com/ngocha2612/falcon-9/blob/main/1.%20Space-X%20Data%20Collection%20API.ipynb

Data Collection - Scraping

```
[Start]
  ↓
[Send HTTP GET Request]
  ↓
[Parse HTML with BeautifulSoup]
  ↓
[Identify All Tables]
  ↓
[Locate Target Table (3rd table)]
  ↓
[Extract Column Names]
  ↓
[Iterate Over Rows in Target Table]
  ↓
+--------------------------------------+
| For Rows with Valid Flight Numbers: |
|    - Extract Data                    |
|    - Append to Dictionary            |
+--------------------------------------+
  ↓
[Create DataFrame from Dictionary]
  ↓
[End]
```

- GitHub URL: https://github.com/ngocha2612/falcon-9/blob/main/2.%20Space-X%20Web%20scraping%20Falcon%209%20and%20Falcon%20Heavy%20Launches%20Records%20from%20Wikipedia.ipynb

Data Wrangling

- How data were processed: I analyzed SpaceX Falcon 9 launch data to understand landing outcomes and create a classification label. Missing values in the dataset were identified, and the mission outcomes were categorized as successful (1) or unsuccessful (0) based on the Outcome column. The resulting classification variable represents whether the first-stage booster landed successfully. The cleaned and labeled dataset was exported for further analysis.
- GitHub URL: https://github.com/ngocha2612/falcon-9/blob/main/3.%20Space-X%20Data%20Wrangling%20spacex.ipynb

Data Import → Identify Missing Values → Explore Attributes → Label Transformation → Export Cleaned Data

EDA with Data Visualization

1. Scatter Plot: Flight Number vs. Payload Mass

- To visualize the relationship between flight numbers (experience) and payload mass.

2. Bar Chart: Success Rate by Launch Site

- To compare success rates across launch sites (e.g., VAFB SLC 4E, KSC LC 39A, CCAFS SLC 40). Bar charts clearly show categorical data variations and are effective for summarizing patterns by site.

3. Bar/Column Chart: Success Rate by Orbit Type

- To analyze success rates for different orbit types (e.g., ES-L1, GEO, GTO). This chart visually demonstrates which orbits are more challenging or easier for successful launches.

4. Line Chart: Yearly Success Rate Trends

- To show the progression of success rates over time (2013–2020). A line chart effectively captures trends, making it easy to identify improvements or fluctuations across years.

- GitHub URL: https://github.com/ngocha2612/falcon-9/blob/main/5.%20Space-X%20EDA%20DataViz%20Using%20Pandas%20and%20Matplotlib%20-%20SpaceX.ipynb

EDA with SQL

- Summarization of the SQL queries performed
  - Retrieve the distinct names of all launch sites in the dataset.
  - Select the first 5 records where the launch site name starts with “CCA”.
  - Retrieve the total number of successful landings (class = 1).
  - Find the launch outcome that resulted in the highest payload mass.
  - Calculate the average payload mass for each launch site.
  - List the flight numbers of all unsuccessful launches (class = 0).
  - Identify the booster version used for the launch with the highest payload mass.
  - Count the total number of records in the dataset.
  - Retrieve the flight number and launch site for the record with the smallest payload mass.
- GitHub URL: https://github.com/ngocha2612/falcon-9/blob/main/4.%20Space-X%20EDA%20Using%20SQL.ipynb

Build an Interactive Map with Folium

- Summarization of what map objects such as markers, circles, lines, etc. I created and added to a folium map
  - 1. **Markers for Launch Sites**:
    - To clearly pinpoint the exact location of each SpaceX launch site on the map for visual clarity and intuitive understanding.
  - 2. **Outcome Markers**:
    - To visualize the success and failure distribution of launches at each site. This helps identify patterns, such as which sites have higher success rates.
  - 3. **Proximity Markers**:
    - To analyze the geographical relationships between launch sites and nearby infrastructure (e.g., coastlines, highways, railways, cities). This aids in understanding logistical considerations and safety measures.
  - 4. **Circles**:
    - Added for better visual emphasis on the location of each launch site, making them stand out against the background.
  - 5. **Distance Lines**:
    - To represent the physical distances between launch sites and important geographical features or infrastructure. This provides insights into proximity and safety considerations.
  - 6. **Marker Clusters**:
    - To avoid cluttering the map when multiple launches occurred at the same location. Clustering provides a cleaner and more organized view.
  - 7. **Mouse Position Plugin**:
    - To facilitate precise measurement of coordinates for proximity analysis, enabling users to dynamically explore the map and identify relevant geographical points.

- GitHub URL: https://github.com/ngocha2612/falcon-9/blob/main/6.Space-X%20Launch%20Sites%20Locations%20Analysis%20with%20Folium-Interactive%20Visual%20Analytics.ipynb

Build a Dashboard with Plotly Dash

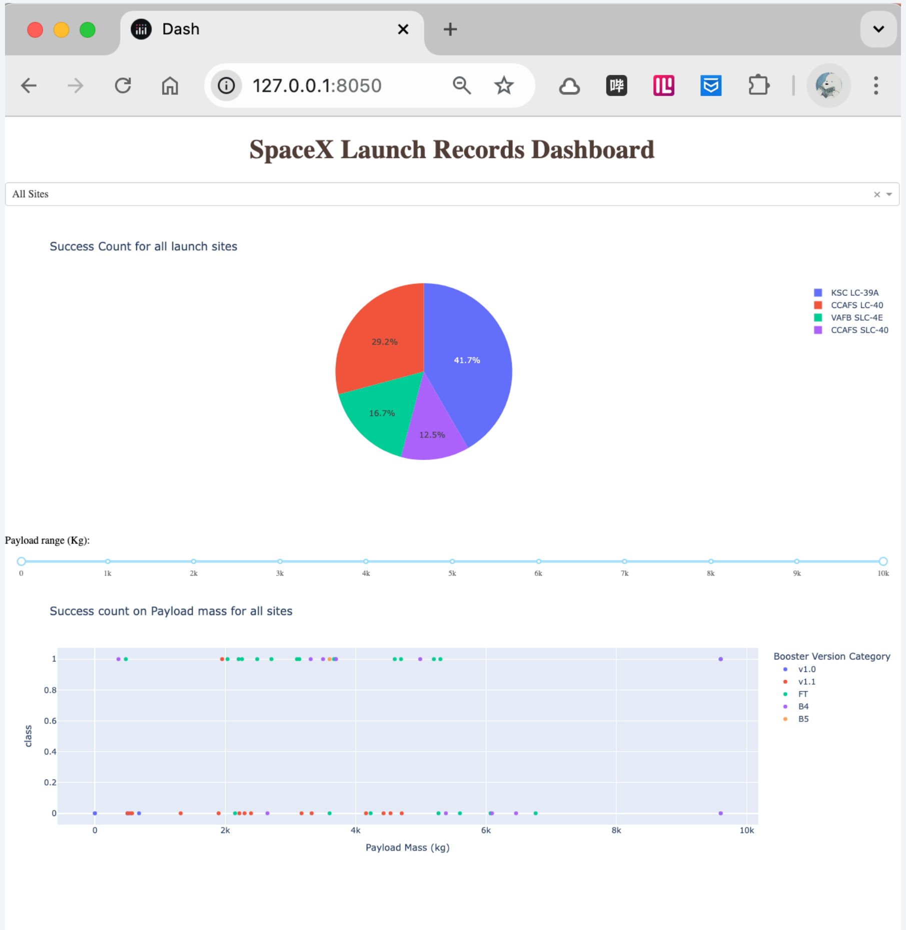
- Summarization of what plots/graphs and interactions have been added to a dashboard
- 1. **Dropdown for Launch Site Selection**
  - Allows users to select a specific launch site or view data for all sites.
- 2. **Pie Chart for Launch Success Counts**
  - Provides a clear overview of the success rate of launches either across all sites or for a specific site.
- 3. **Range Slider for Payload Selection**
  - Enables users to filter the data by payload mass to explore how payload weight correlates with launch outcomes.
- 4. **Scatter Plot for Payload vs. Launch Success**
  - Visualizes the relationship between payload mass and launch success rate, highlighting any trends or anomalies.
  - Scatter plots are effective for revealing patterns and correlations between two continuous variables. This plot helps users explore whether payload mass impacts success rates and how booster versions play a role.
- GitHub URL: https://github.com/ngocha2612/falcon-9/blob/main/7.%20Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash%20-%20spacex_dash_app.py

Predictive Analysis (Classification)

- Summarization of how I built, evaluated, improved, and found the best performing classification model
- **1. Data Preprocessing**
  - Standardized features to ensure consistency across all variables.
  - Split data into training (80%) and testing (20%) subsets to evaluate model performance.
- **2. Model Building and Hyperparameter Tuning**
  - **Logistic Regression:** Tuned regularization strength (C) and solver using GridSearchCV.
  - **Support Vector Machine (SVM):** Tuned kernel type, regularization (C), and kernel coefficient (gamma).
  - **Decision Tree:** Tuned splitting criteria, maximum depth, and minimum samples for splitting/leaf nodes.
  - **K-Nearest Neighbors (KNN):** Tuned the number of neighbors (n_neighbors), distance metric (p), and search algorithm.
- **3. Model Evaluation**
  - Used GridSearchCV to identify the best hyperparameters and cross-validated accuracy for each model.
  - Evaluated models on the test set for final accuracy and plotted confusion matrices to identify classification issues (e.g., false positives).
- **4. Comparison of Methods**
  - Compiled test accuracies into a summary report. All models performed equally well, achieving 83.33% accuracy on the test set.
- GitHub URL: https://github.com/ngocha2612/falcon-9/blob/main/8.%20SpaceX%20Machine%20Learning%20Prediction.ipynb

Results

- Exploratory data analysis results
  - As the flight number increases in each of the 3 launch sites, so does the success rate.
  - For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
  - Orbits ES-L1, GEO, HEO & SSO have the highest success rates at 100%, with SO orbit having the lowest success rate at ~50%. Orbit SO has 0% success rate.
  - In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
  - With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
  - However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.
  - The success rate since 2013 kept increasing till 2020
- Interactive analytics demo in screenshots
- Predictive analysis results
  - All models performed equally well, achieving 83.33% accuracy on the test set.

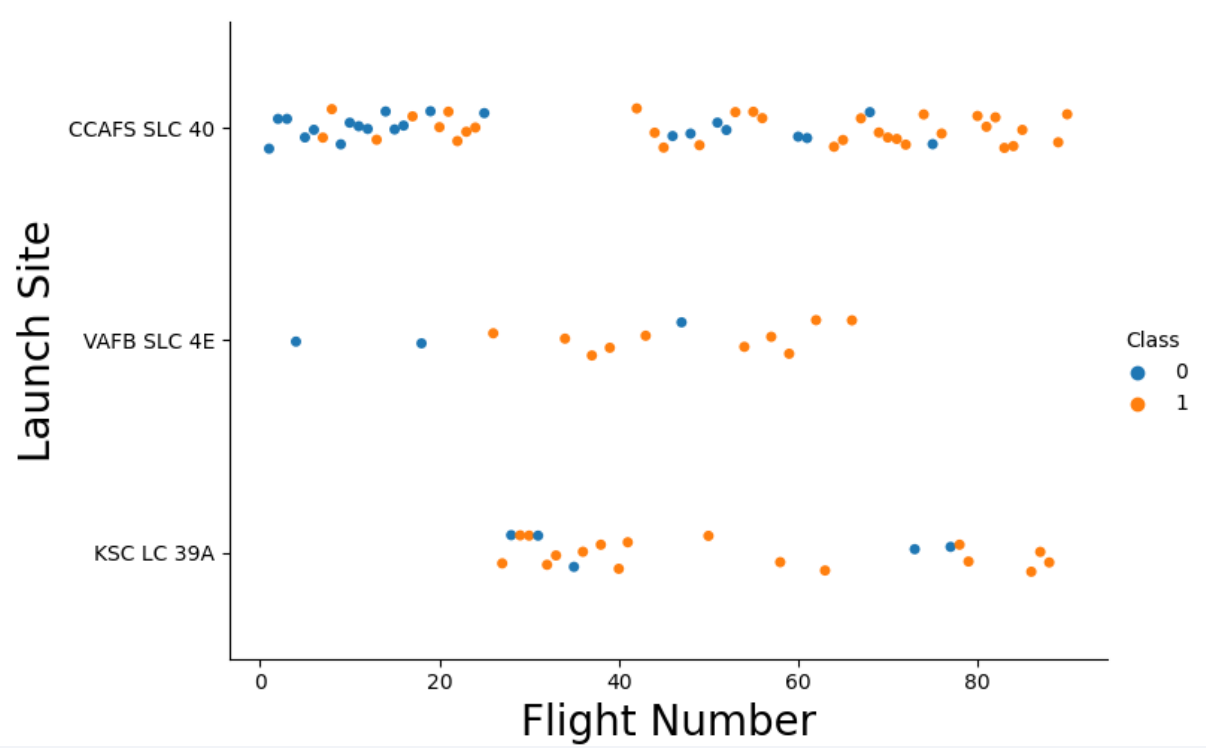


Section 2

Insights drawn from EDA

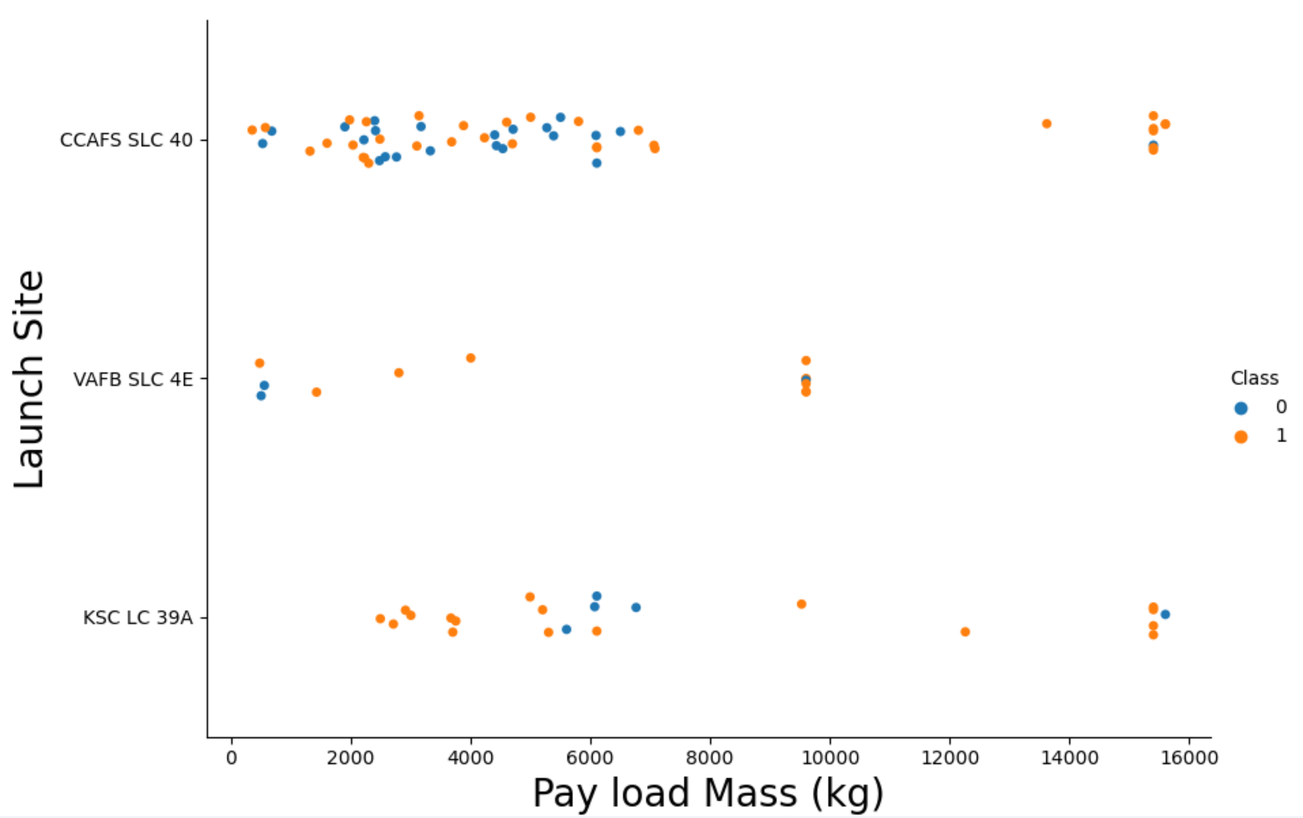
Flight Number vs. Launch Site

- We can deduce that, as the flight number increases in each of the 3 launch sites, so does the success rate. For instance, the success rate for the VAFB SLC 4E launch site is 100% after the Flight number 50. Both KSC LC 39A and CCAFS SLC 40 have a 100% success rates after 80th flight.



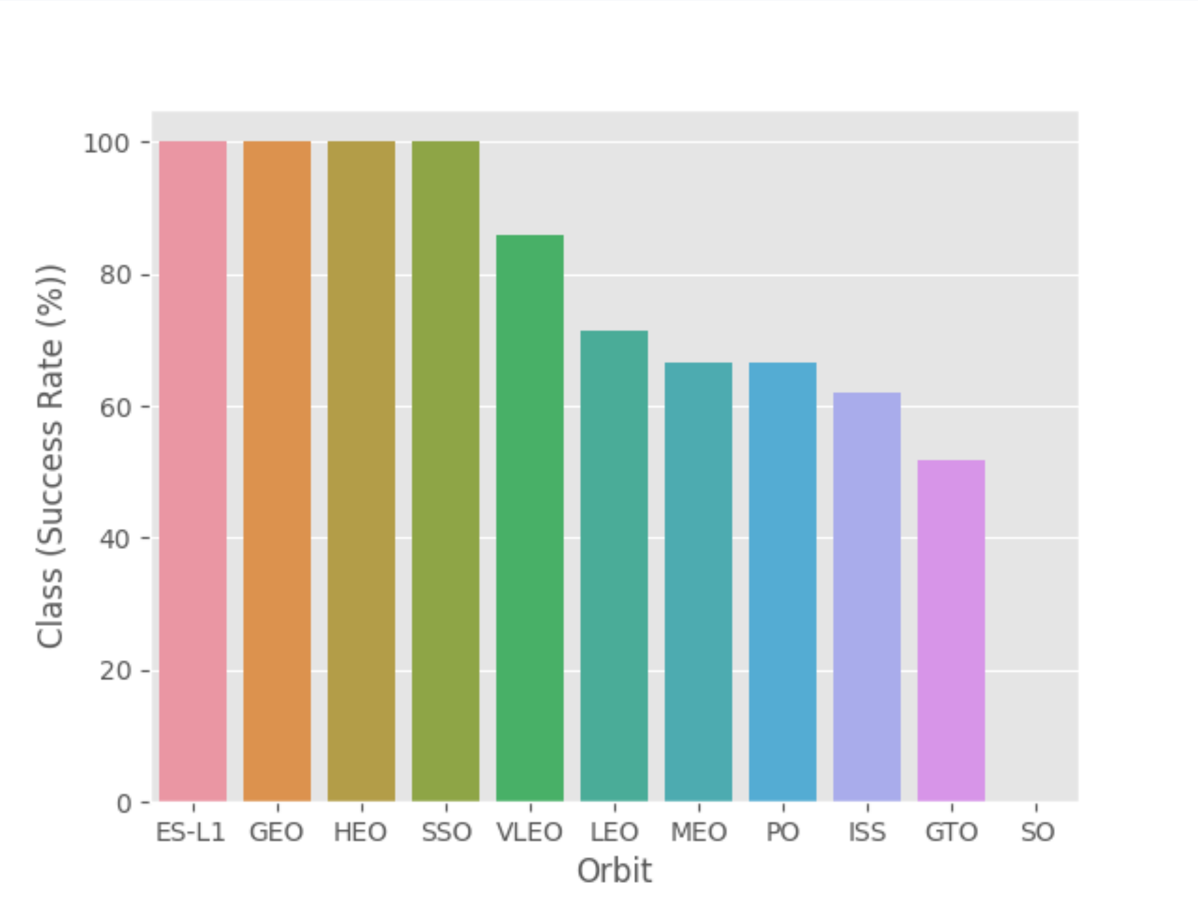
Payload vs. Launch Site

- Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).



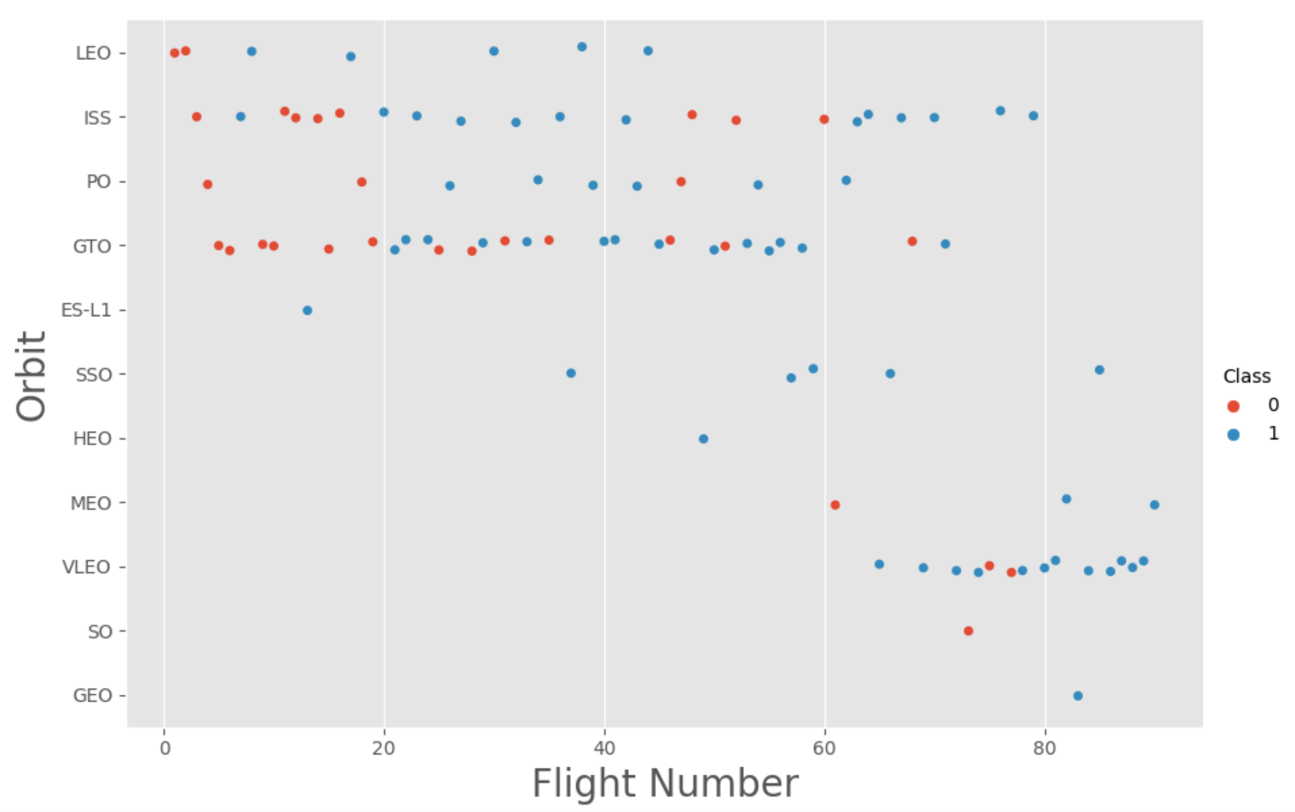
Success Rate vs. Orbit Type

- Orbits ES-L1, GEO, HEO & SSO have the highest success rates at 100%, with SO orbit having the lowest success rate at ~50%. Orbit SO has 0% success rate.



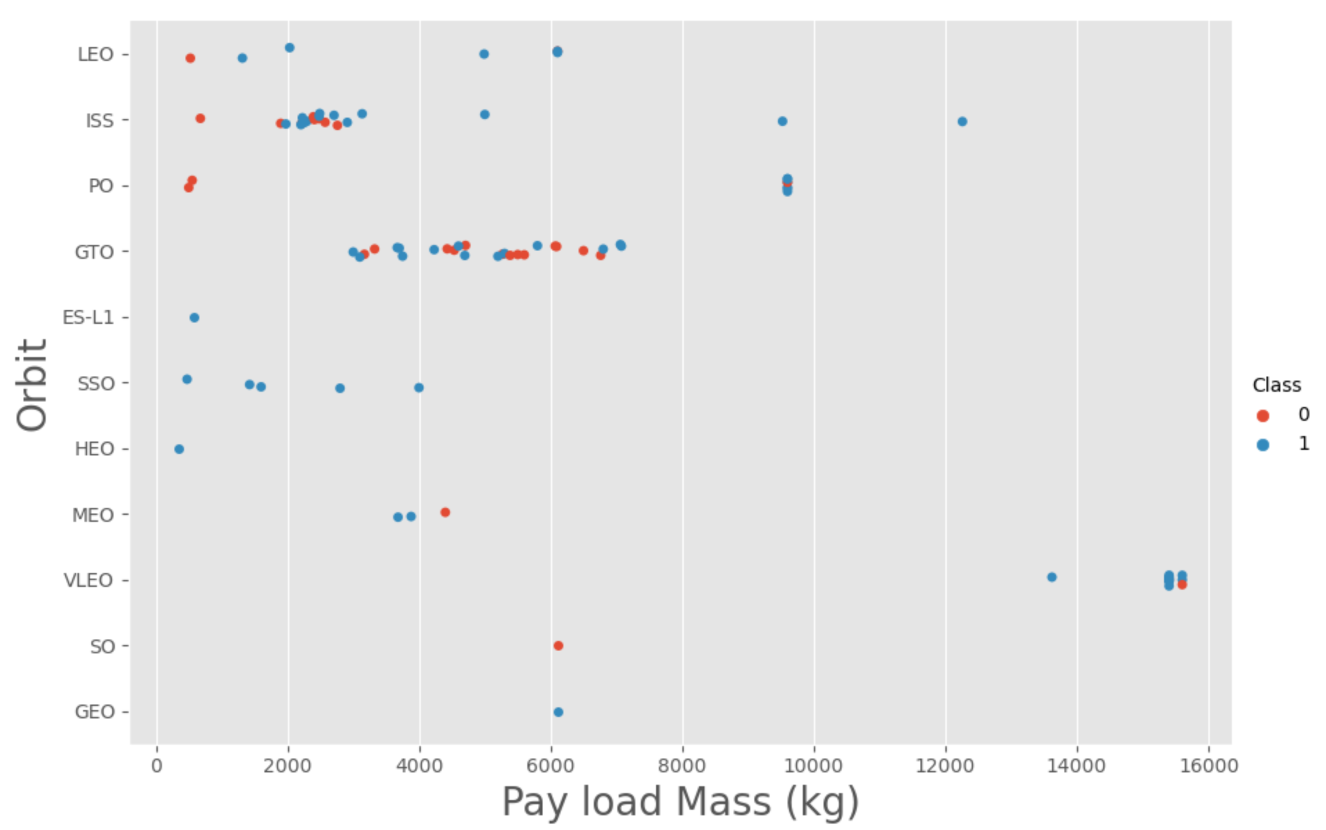
Flight Number vs. Orbit Type

- You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



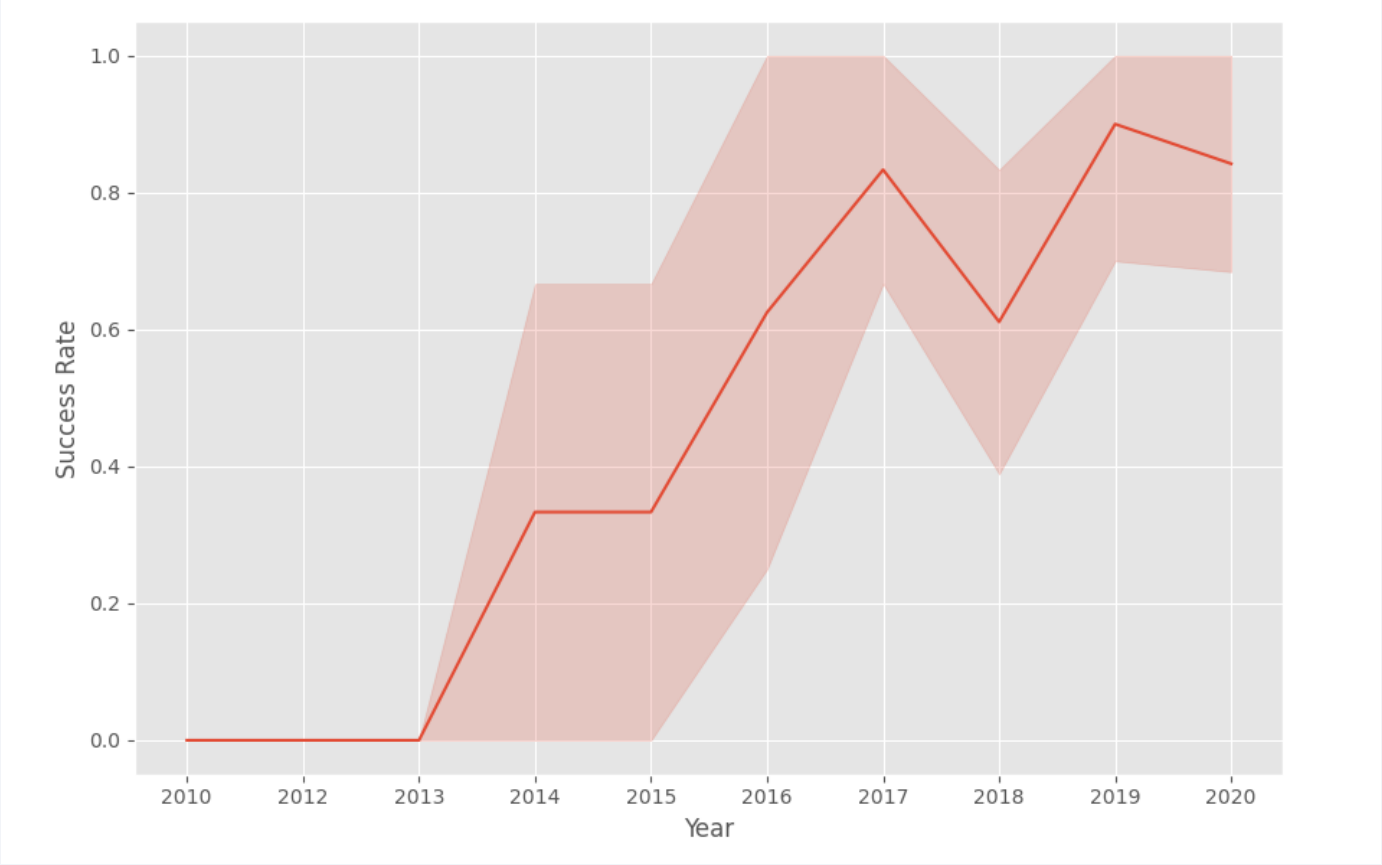
Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



Launch Success Yearly Trend

- You can observe that the success rate since 2013 kept increasing till 2020



All Launch Site Names

- The four unique launch sites identified are:
  1. **CCAFS LC-40**
  2. **VAFB SLC-4E**
  3. **KSC LC-39A**
  4. **CCAFS SLC-40**
- These are key locations used in SpaceX's space missions.

```
In [31]:
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;

 * sqlite:///my_data1.db
Done.
Out[31]:
```

Launch_Sites
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`
- All records are associated with **CCAFS LC-40** (Cape Canaveral Air Force Station).

In [72]:
```
%sql SELECT * FROM 'SPACEXTBL' WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

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

Out[72]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landin _Outcom
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failu (parachut
08-12-2010	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	Failu (parachut
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attem
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attem
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attem

Total Payload Mass

- The total payload carried by boosters from NASA
- **Key Result:**
    - **Total Payload Mass**: **45,596 kg**.
    - **Customer**: The launches were conducted for NASA under the **Commercial Resupply Services (CRS)** program.
- This result demonstrates SpaceX's significant contribution to NASA's CRS missions, which primarily focus on delivering cargo to the International Space Station (ISS).

```
In [19]: %sql SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Version FROM 'SPACEXTBL' WHERE Booste

 * sqlite:///my_data1.db
Done.
Out[19]:
```

Payload Mass Kgs	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1
- **Key Result:**
  - **Average Payload Mass**: Approximately **2534.67 kg**.
  - **Customer**: The missions were conducted for **MDA**.
  - **Booster Version**: **F9 v1.1 B1003**.
- This result highlights the payload capability of the **F9 v1.1** booster version for specific missions targeting MDA's requirements. The payload mass reflects the average weight of cargo delivered during these launches.

In [19]:

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Version FROM 'SPACEXTBL' WHERE Booste
```

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

Out[19]:

Payload Mass Kgs	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad
- The query identifies the **earliest date** when SpaceX achieved its first **successful landing on a ground pad** by using the MIN function and filtering for the **Landing Outcome** equal to "Success (ground pad)".
- **Key Result: Date of First Successful Ground Pad Landing**: **01-05-2017**

```
In [21]:
%sql SELECT MIN(DATE) FROM 'SPACEXTBL' WHERE "Landing _Outcome" = "Success (g

 * sqlite:///my_data1.db
Done.
Out[21]:
```

MIN(DATE)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- **Booster Versions** with successful drone ship landings and payload mass between 4000 kg and 6000 kg:
  - F9 FT B1022 (Payload: JCSAT-14)
  - F9 FT B1026 (Payload: JCSAT-16)
  - F9 FT B1021.2 (Payload: SES-10)
  - F9 FT B1031.2 (Payload: SES-11 / EchoStar 105)
- These boosters represent missions where SpaceX successfully utilized drone ship landings to recover the rocket after delivering payloads in the specified mass range.

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

In [26]:

```
# %sql SELECT * FROM 'SPACEXTBL'
```

In [27]:

```
%sql SELECT DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE "Landing _
```

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

Out[27]:

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- The query lists the total number of mission outcomes, categorized by their result (success or failure). Here are the outcomes and their respective counts:
  - **Failure (in flight)**: 1 mission
  - **Success**: 98 missions
  - **Success** (with 1 mission outcome under this category)
  - **Success (payload status unclear)**: 1 mission
- The query suggests that the majority of missions have been successful, with a total of 99 successful missions and 1 failure during flight. However, it seems that some entries might represent the same outcome (success), but are categorized differently.

In [28]:

```
%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM SPACEXT
```

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

Out[28]:

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

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass
- The table presents a list of booster versions, their corresponding payloads, and the payload mass in kilograms for each mission. All the payloads listed have a mass of **15,600 kg**.
- The pattern here shows that these missions primarily involve **Starlink satellite launches** with a consistent payload mass of **15,600 kg**. Some of the missions also include notable events such as **Crew Dragon Demo-2** and **SpaceX CRS missions**.

In [30]:

```
%sql SELECT "Booster_Version",Payload, "PAYLOAD_MASS__KG_" FROM SPACEXTBL WHE
```

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

Out[30]:

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

2015 Launch Records

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Both of these entries show successful missions but with failures in landing on the drone ship. The query provides a clear indication of when these failures occurred, along with the associated mission details.

In [68]:
```
%sql SELECT substr(Date,7,4), substr(Date, 4, 2),"Booster_Version", "Launch_Site", Payload, "PAYLOAD_MASS__KG_",
```

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

Out[68]:

substr(Date,7,4)	substr(Date, 4, 2)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Mission_Outcome	Landing _Outcome
2015	01	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	Success	Failure (drone ship)
2015	04	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	Success	Failure (drone ship)

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

- These records include only successful landing outcomes (e.g., “Success (ground pad)” or “Success (drone ship)”). Each entry demonstrates the evolving success of SpaceX’s reusable booster technology during this period, particularly between 2010 and 2017. The missions listed are foundational milestones in proving SpaceX’s reliability in achieving precise landings for boosters after launch.

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

In [74]:

```
%sql SELECT * FROM SPACEXTBL WHERE "Landing _Outcome" LIKE 'Success%' AND (Date BETWEEN '04-06-2010' AND '20-03-2
```

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

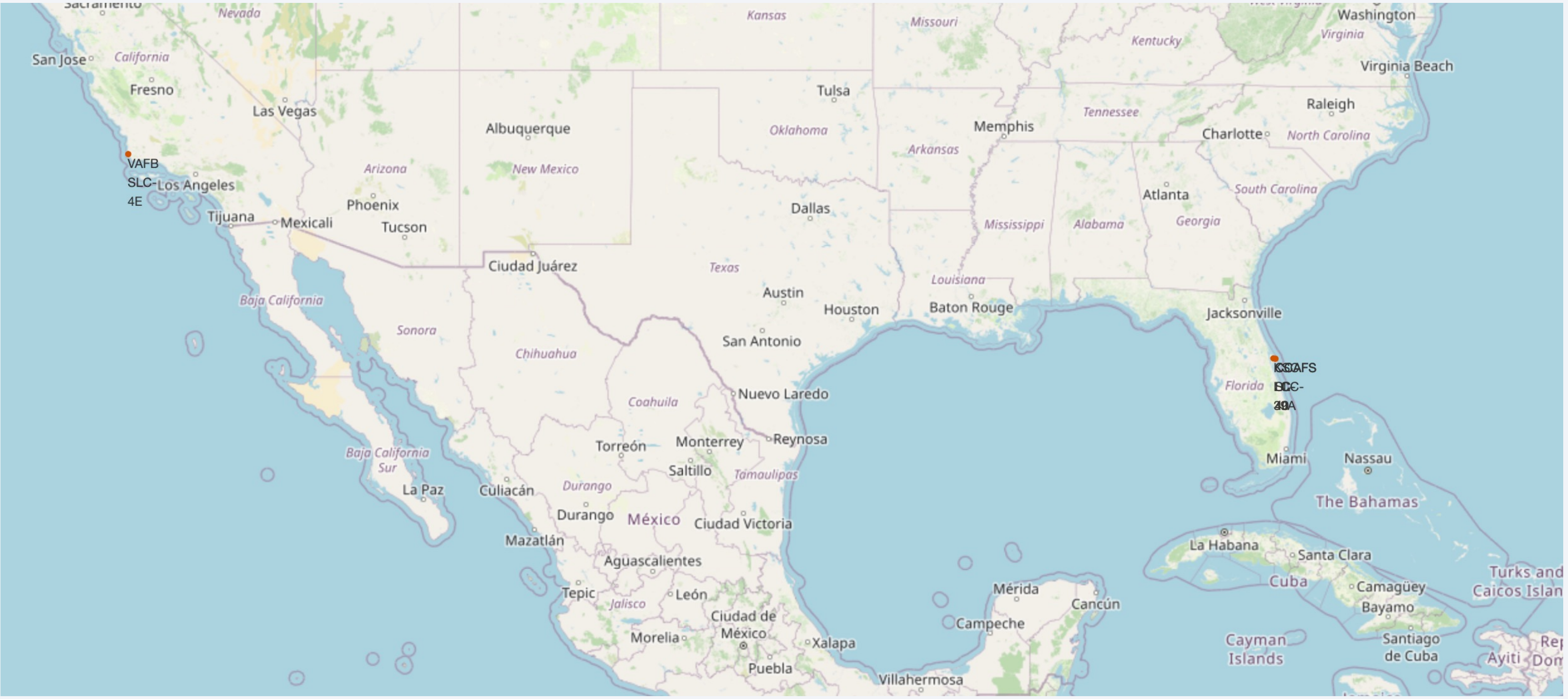
Out[74]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success
18-10-2020	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success
18-08-2020	14:31:00	F9 B5 B1049.6	CCAFS SLC-40	Starlink 10 v1.0, SkySat-19, -20, -21, SAOCOM 1B	15440	LEO	SpaceX, Planet Labs, PlanetIQ	Success
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success
18-04-2018	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362	HEO	NASA (LSP)	Success
17-12-2019	00:10:00	F9 B5 B1056.3	CCAFS SLC-40	JCSat-18 / Kacific 1, Starlink 2 v1.0	6956	GTO	Sky Perfect JSAT, Kacific 1	Success
16-11-2020	00:27:00	F9 B5B1061.1	KSC LC-39A	Crew-1, Sentinel-6 Michael	12500	LEO (ISS)	NASA (CCP)	Success

Section 3

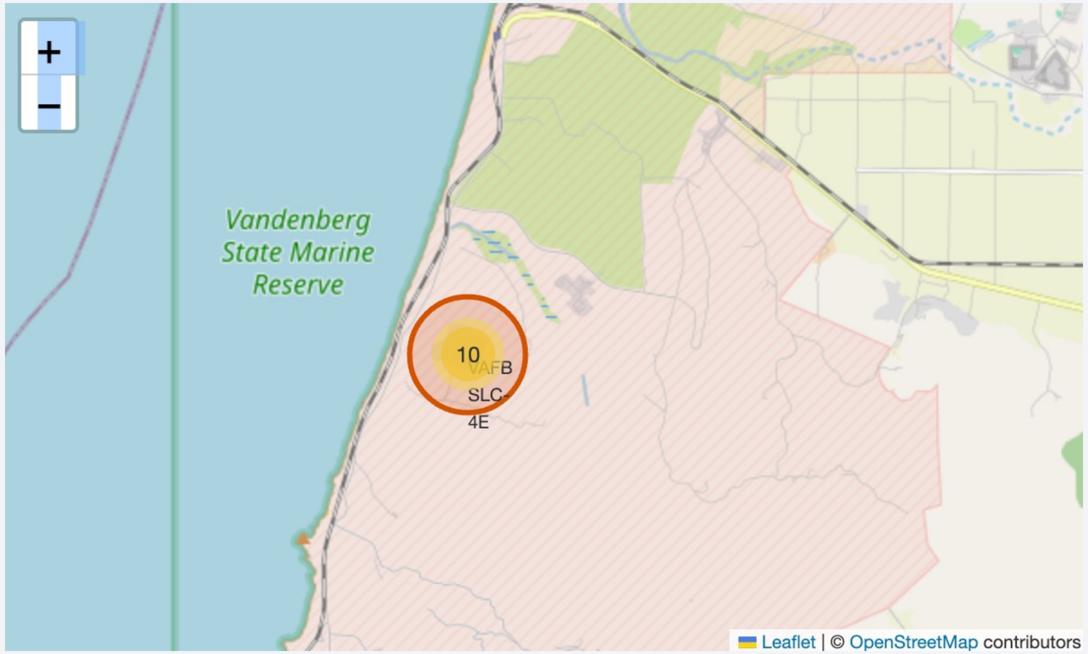
Launch Sites Proximities Analysis

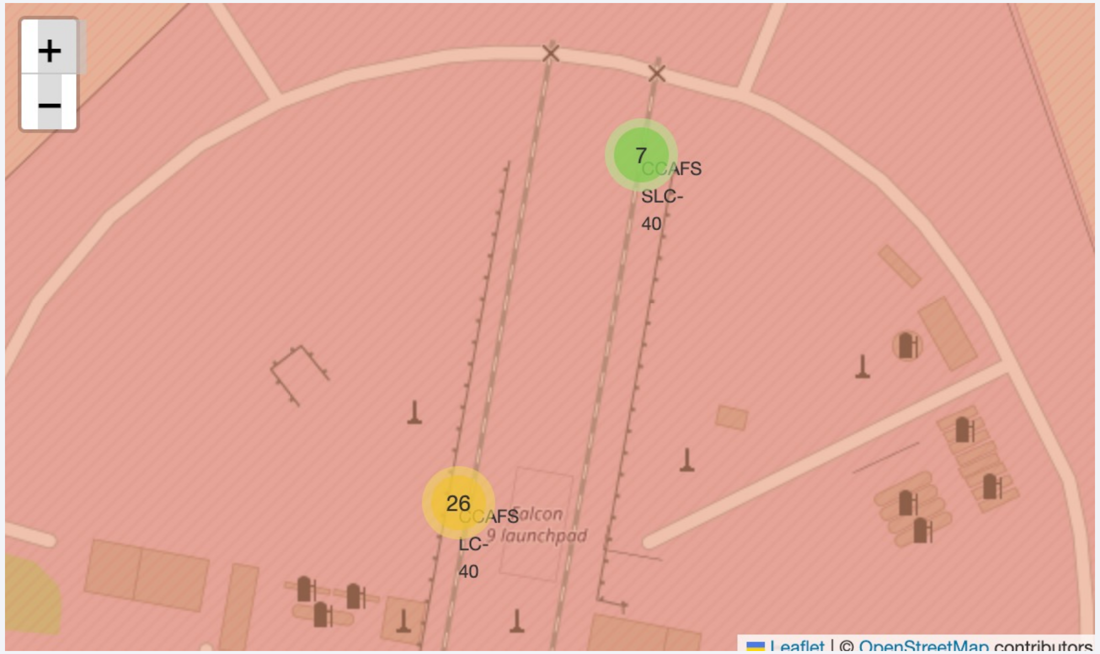
Launch Site Locations



Launch sites are in proximity to the equator and the coast. This makes sense as it takes less fuel to get into space from the equator due to the physics of Earth's rotation. The launch sites in close proximity to the coast are also logical for safety reasons.

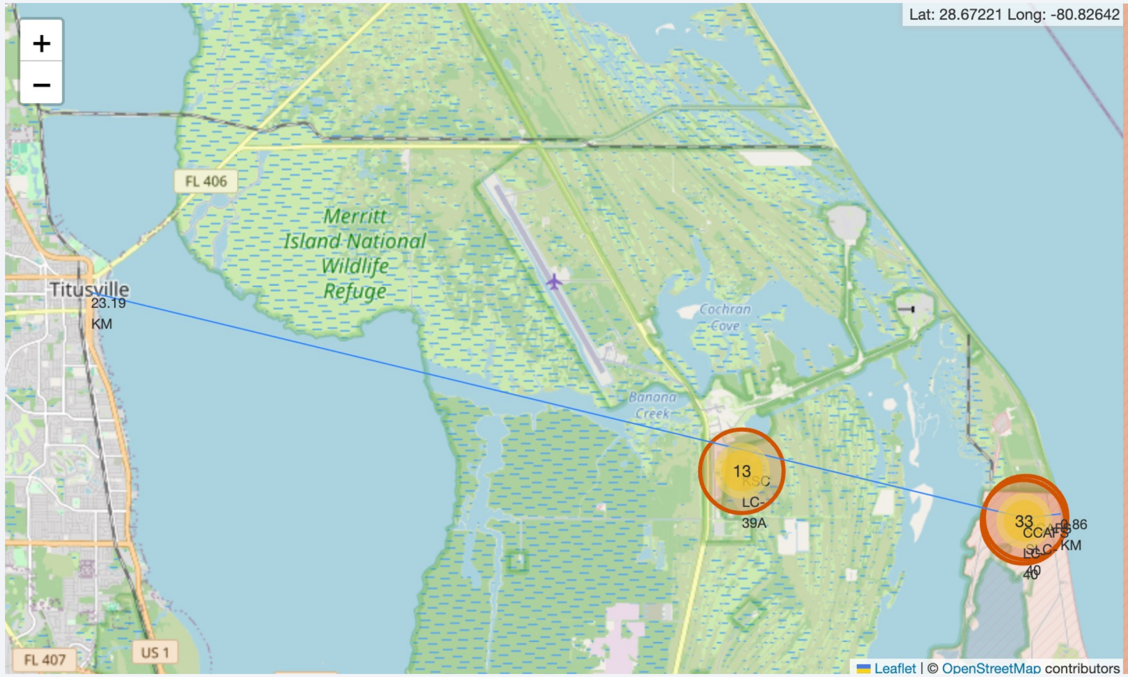
The success/failed launches for each site

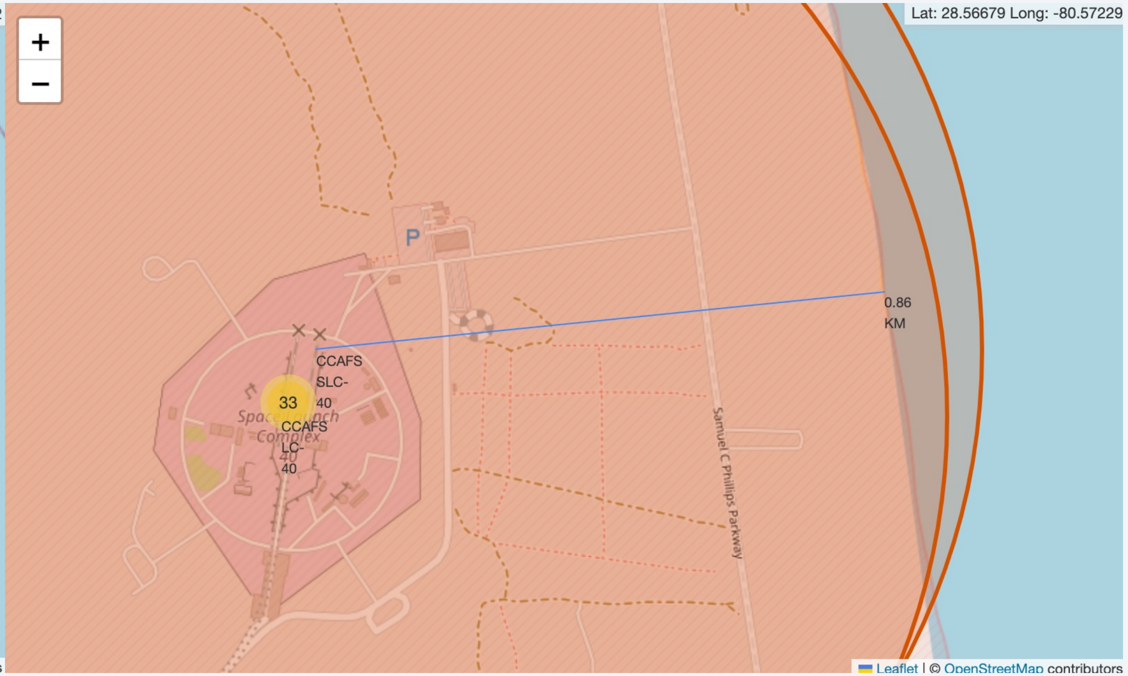




From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.

Distances between a launch site to its proximities





1. Proximity to Railways: Launch sites are often built near railways for logistical ease in transporting heavy equipment, like rocket components, to the launch facility.

2. Proximity to Highways: Highways allow for quick road transportation of equipment and personnel, making proximity beneficial.

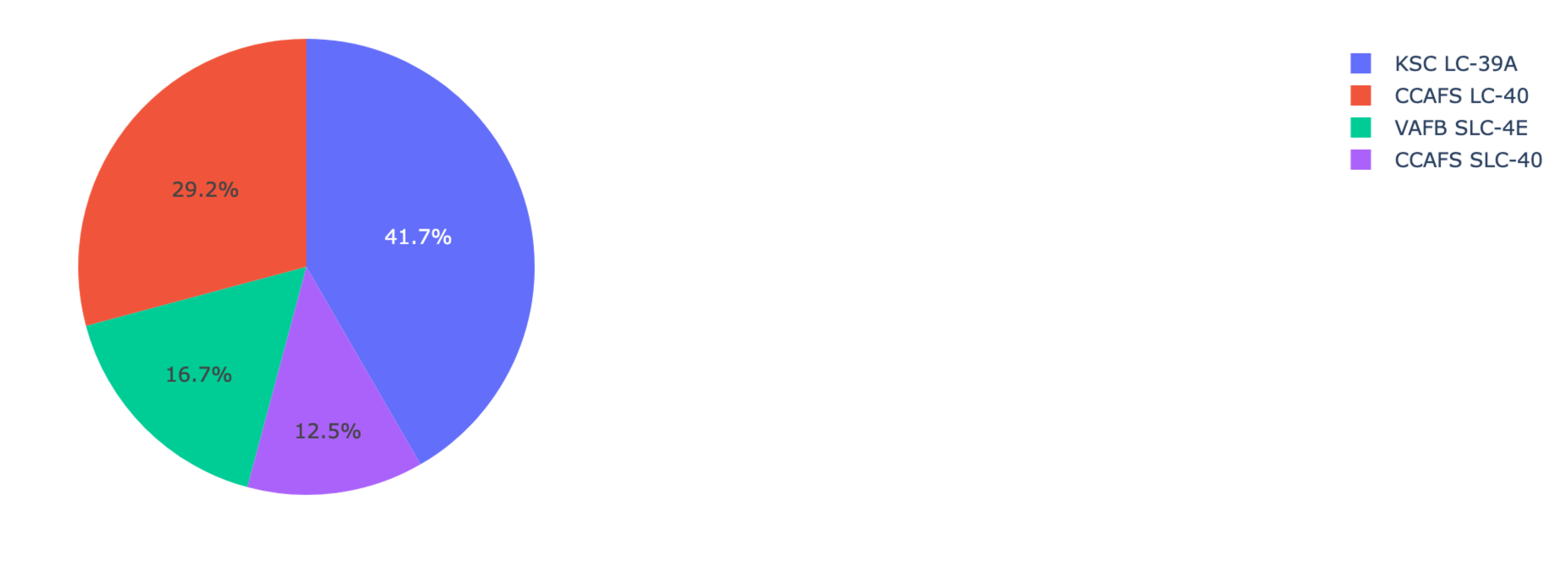
3. Proximity to Coastlines: Launch sites are typically near coastlines to reduce risk in the event of a failure and allow rockets to fly over open water during early ascent.

4. Distance from Cities: Launch sites are often located far from densely populated areas to minimize risks associated with launches, such as noise, debris, or accidents.

Section 4

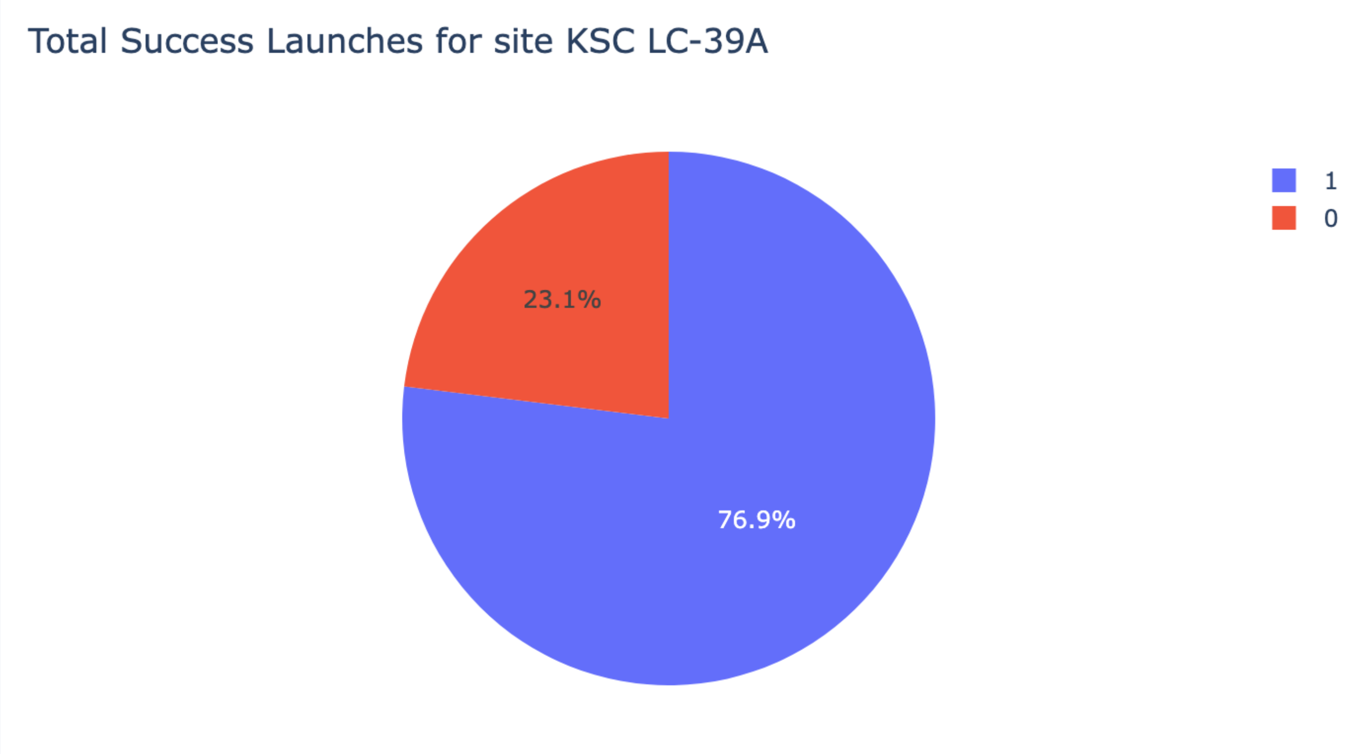
Build a Dashboard with Plotly Dash

Success Count for all Launch Sites



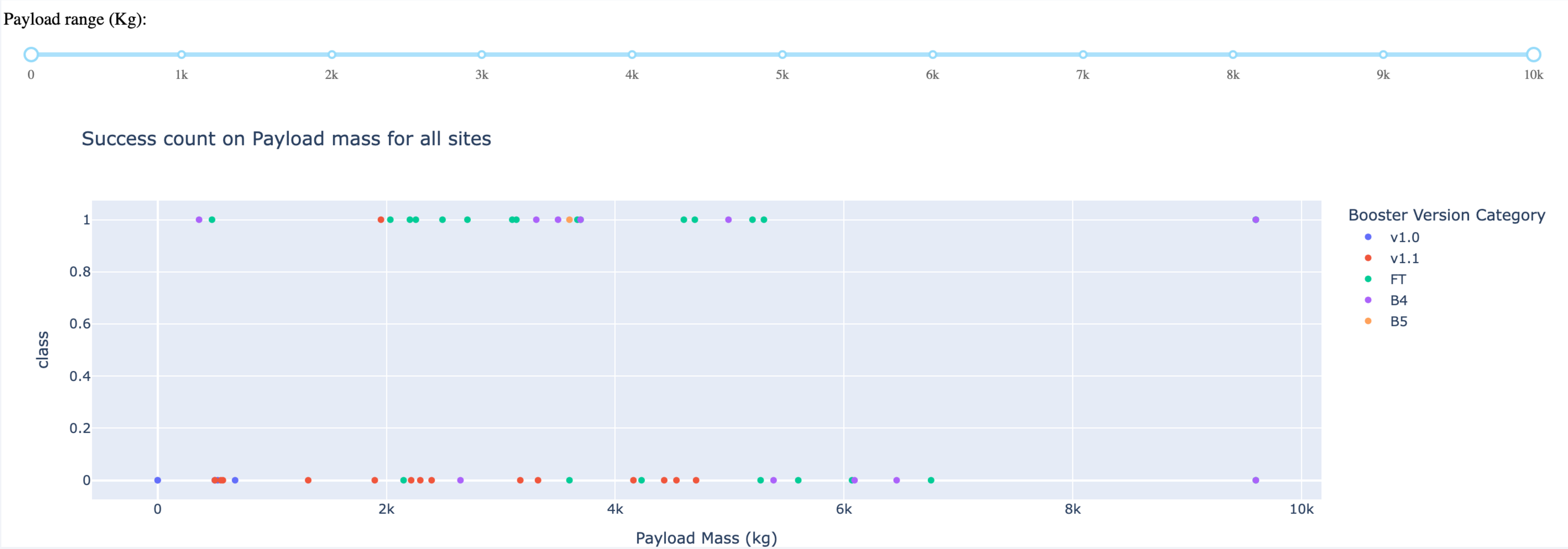
- Launch site KSC LC-39A has the highest launch success rate at 42% followed by CCAFS LC-40 at 29%, VAFB SLC-4E at 17% and lastly launch site CCAFS SLC-40 with a success rate of 13%.

Launch Site with Highest Launch Success Ratio



- Launch site KSC LC-39A had the highest success ratio of 76.9% success against 23.1% failed launches

Success count on Payload mass for all sites



Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All the methods perform equally on the test data: i.e. They all have the same accuracy of 0.83 on the test data.

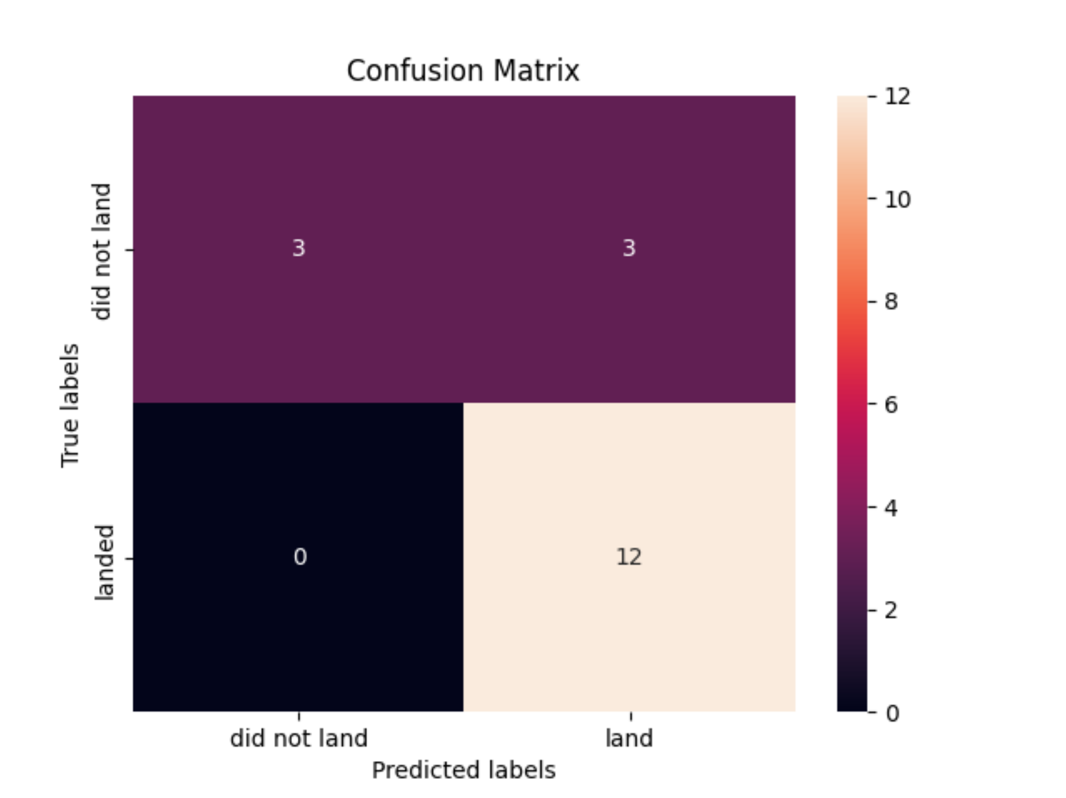
Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

Confusion Matrix

- 1. **True Negatives (TN)**: **3**
- 2. **False Positives (FP)**: **3**
- 3. **False Negatives (FN)**: **0**
- 4. **True Positives (TP)**: **12**

Insights:

- **Strengths**: The model is excellent at identifying rockets that landed (100% recall), meaning it rarely misses a successful landing.
- **Weaknesses**: There are a few false positives (predicting "Landed" for rockets that did not land), lowering precision slightly.
- This matrix suggests the model performs well overall, especially in identifying successful landings, but there is room for improvement in reducing false positives.



Conclusions

- Different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- We can deduce that, as the flight number increases in each of the 3 launch sites, so does the success rate. For instance, the success rate for the VAFB SLC 4E launch site is 100% after the Flight number 50. Both KSC LC 39A and CCAFS SLC 40 have a 100% success rates after 80th flight.
- If you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000).
- Orbits ES-L1, GEO, HEO & SSO have the highest success rates at 100%, with SO orbit having the lowest success rate at ~50%. Orbit SO has 0% success rate.
- LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.
- And finally the success rate since 2013 kept increasing till 2020.

Thank you!