IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rebecca Kimble

April 28, 2025

Outline

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

Executive Summary

Summary of methodologies:

1. Data was collected from multiple sources, including web-based charts.
2. Exploratory data analysis was done using SQL and visualizations.
3. A Dashboard was created to highlight mission success rates.
4. Predictive analysis was created to better understand the likelihood of successful mission completion.

Summary of all results:

- Generally, missions have grown more successful over time
- Most missions are conducted east coast of Florida at a significant distance from highways, railways, and towns.
- Mission success can be predicted using logistic regression with at least 83% accuracy.

Introduction

- Elon Musk has been sending rockets into space for over two decades. As he recently played a key role in Trump’s coup and the successful commandeering of American democracy, it is critical to evaluate whether his most expensive personal games—his ventures with rockets—are truly successful endeavors.
- Musk’s company, SpaceX, offers discount rocket launches because it reuses its first-stage rockets. This equipment must successfully return to Earth for the savings to be realized.
- We examined success rate elements to create a means of predicting the likelihood of being able to reuse a first-stage rocket after initial launch.

Section 1

Methodology

Methodology - Executive Summary

Data collection methodology

- Data was collected from 5 sources.
- Four sources were from SpaceXdata.com and included details on booster versions, launch sites, payload weights, and other rocket-specific details.
- The final source was taken from Wikipedia's summary of rocket launches using a specific booster, the Falcon 9

Perform data wrangling

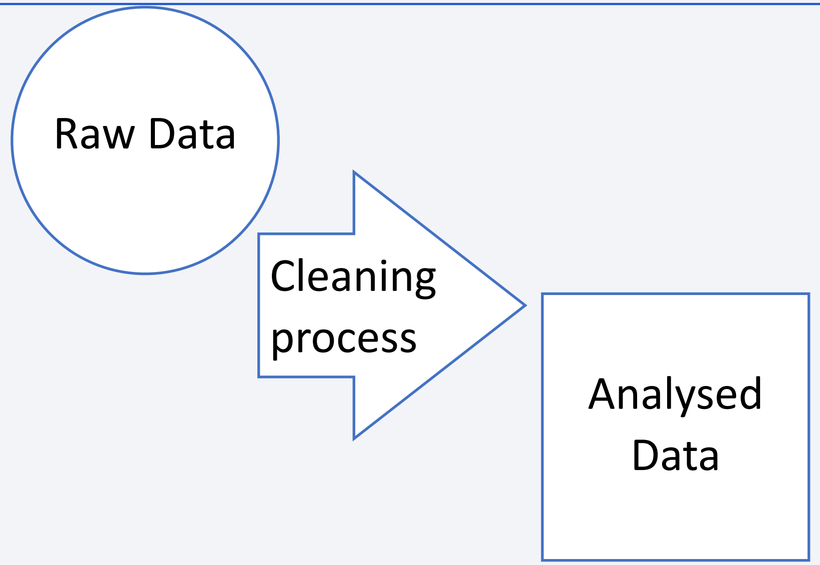
- SpaceX launches different types of rockets into different orbits around the Earth, and they then return to the ground in different ways. We looked at the frequency of each of these variables as it related to the success or failure of the mission to reclaim the initial rocket.
    - Exploratory data analysis (EDA) using visualization and SQL
    - Interactive visual analytics using Folium and Plotly Dash
    - Predictive analysis using classification models

Data Collection

- Data was downloaded from SpaceX using Application Programming Interface to obtain the required fields.
    - Booster information was obtained from a dataset regarding SpaceX rocket information
    - Launch Site locations (latitude and longitude) were obtained from a dataset describing SpaceX launch pads. Similarly, payload and orbit information were obtained from specialized datasets.
    - Finally, core data regarding outcomes, core usage, gridfins, legs, origin, and final descent was taken from a fourth database.
- Additional data regarding a specific rocket, the Falcon 9, was gathered from Wikipedia using the webscraping techniques in BeautifulSoup.

Data Collection – SpaceX API

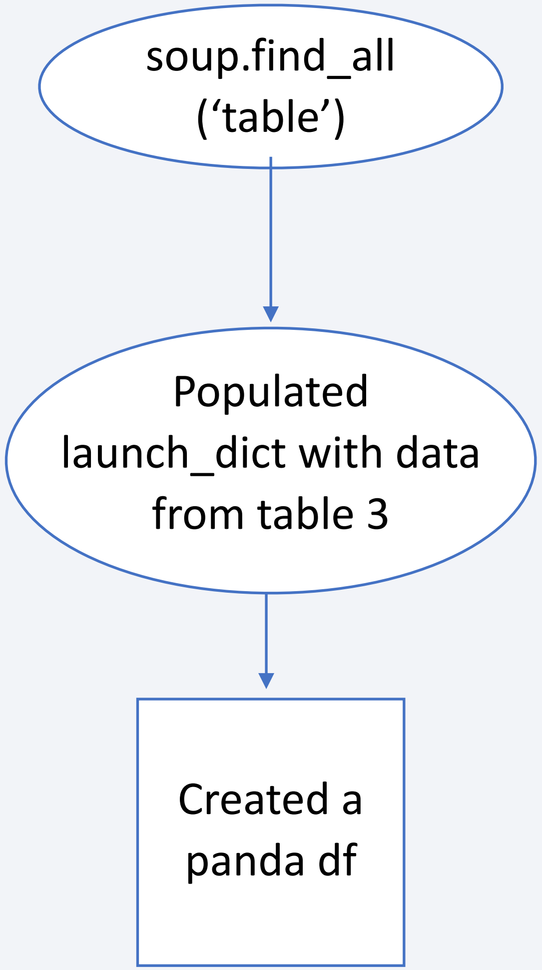
- Booster versions were limited to the Falcon 9
- Missing payload data was replaced with mean values
- Fields were combined into one dataset
- https://github.com/cronecoding/IBM_DS_Final/jupyter-labs-spacex-data-collection-api-v2.ipynb



Data Collection - Scraping

Wikipedia offers an in-depth chart of Falcon 9 launches. This chart was used in our analysis through BeautifulSoup code.

- requests.get("https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches")
- soup = BeautifulSoup(response.text, 'html.parser')
- https://github.com/cronecoding/IBM_DS_Final/blob/main/jupyter-labs-webscraping.ipynb



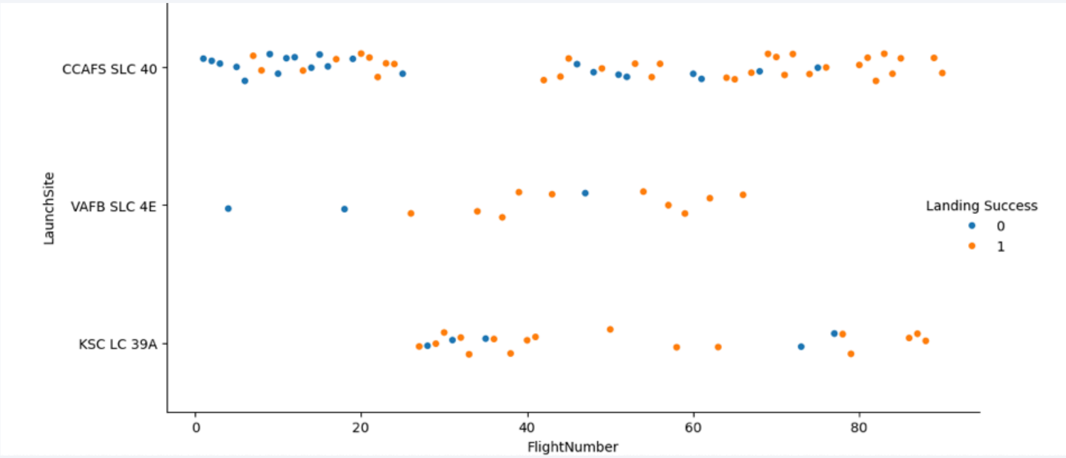
Data Wrangling

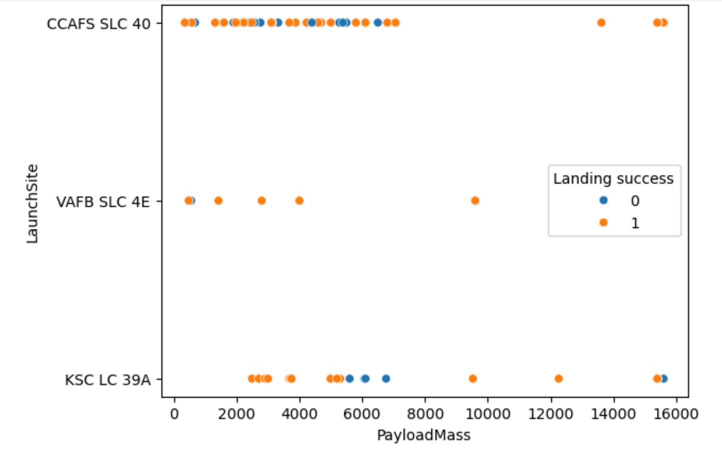
- Mission success was simplified to a binary. Zero was a missed recovery, and one represents a successful recovery.
- A separate file entitled “bad outcomes” was created
- Total launches from each site were calculated.
- The types of orbit missions reached were tallied.

- https://github.com/cronecoding/IBM_DS_Final/blob/main/labs-jupyter-spacex-Data wrangling-v2.ipynb

EDA with Data Visualization

- Each visualization success (orange) or failure (blue) was noted by selecting the hue in the chart code.
- First, a category plot of payload vs flight number showed changes over time.
- The data was then separated into launch sites. The majority of launches, successes and failures, were done from Florida.
- This remained true when evaluating payload differences. California launches were lower weight and less frequent.





- https://github.com/cronecoding/IBM_DS_Final/blob/main/jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

- SELECT DISTINCT “Launch_Site” FROM SPACEXTBL;
- SELECT FROM SPACEXTBL WHERE “Launch_Site” LIKE “CCA%” LIMIT 5;
- SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = “NASA (CRS)”;
- SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE “Booster_Version” LIKE “F9 v1.1%”;
- SELECT DATE FROM SPACEXTBL WHERE “Landing_Outcome” LIKE “Success%” ORDER BY DATE LIMIT 1;
- SELECT DISTINCT “Booster_Version” FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
- SELECT “Landing_Outcome”, COUNT(*) AS TOTAL FROM SPACEXTBL GROUPBY “Landing_Outcome”
- SELECT DISTINCT “Booster_Version” FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
- SELECT SUBSTR(“Date”, 6,2) AS Month, “Landing_Outcome”, “Booster_Version”, “Launch_Site” FROM SPACEXTBL WHERE “Landing_Outcome” LIKE “%Failure%” AND “Landing_Outcome” LIKE “%drone ship%” AND SUBSTR(“DATE”, 1.4) = “2015”;
- https://github.com/cronecoding/IBM_DS_Final/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Launch locations were highlighted on a map of the United States of America, which included the number of launches at each site and whether they were a success or not.
- A mouse-based coordinate generator and distance calculator allowed us to measure to the nearest coastline, railway, town, or major highway to explore reasons behind variations in success. All launch sites are by the coast, but Florida is more remote from human population and transport elements.

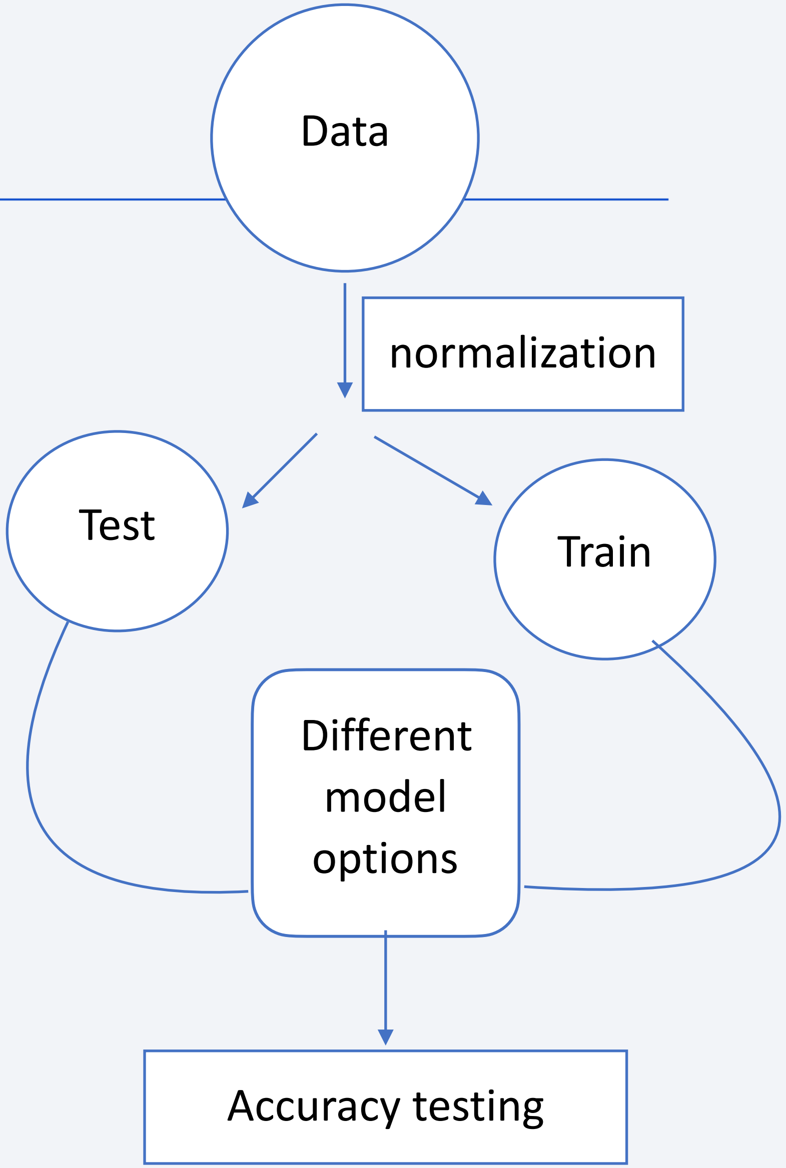
- https://github.com/cronecoding/IBM_DS_Final/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- The dashboard shows the percentage of success via a pie chart for each launch site, plus the total of all sites. This shows clearly that the Kennedy Space Center Launch Complex is the most successful location to launch, with 77% success.
- A slide bar allows the payload to be adjusted in a scatter plot chart, highlighting success vs failure for different boosters. Payloads greater than 6000kg are much less likely to succeed, regardless of launch location.
- This allows one to easily see the impact that payload has on mission success
- https://github.com/cronecoding/IBM_DS_Final/blob/main/spacex-dash-app.py

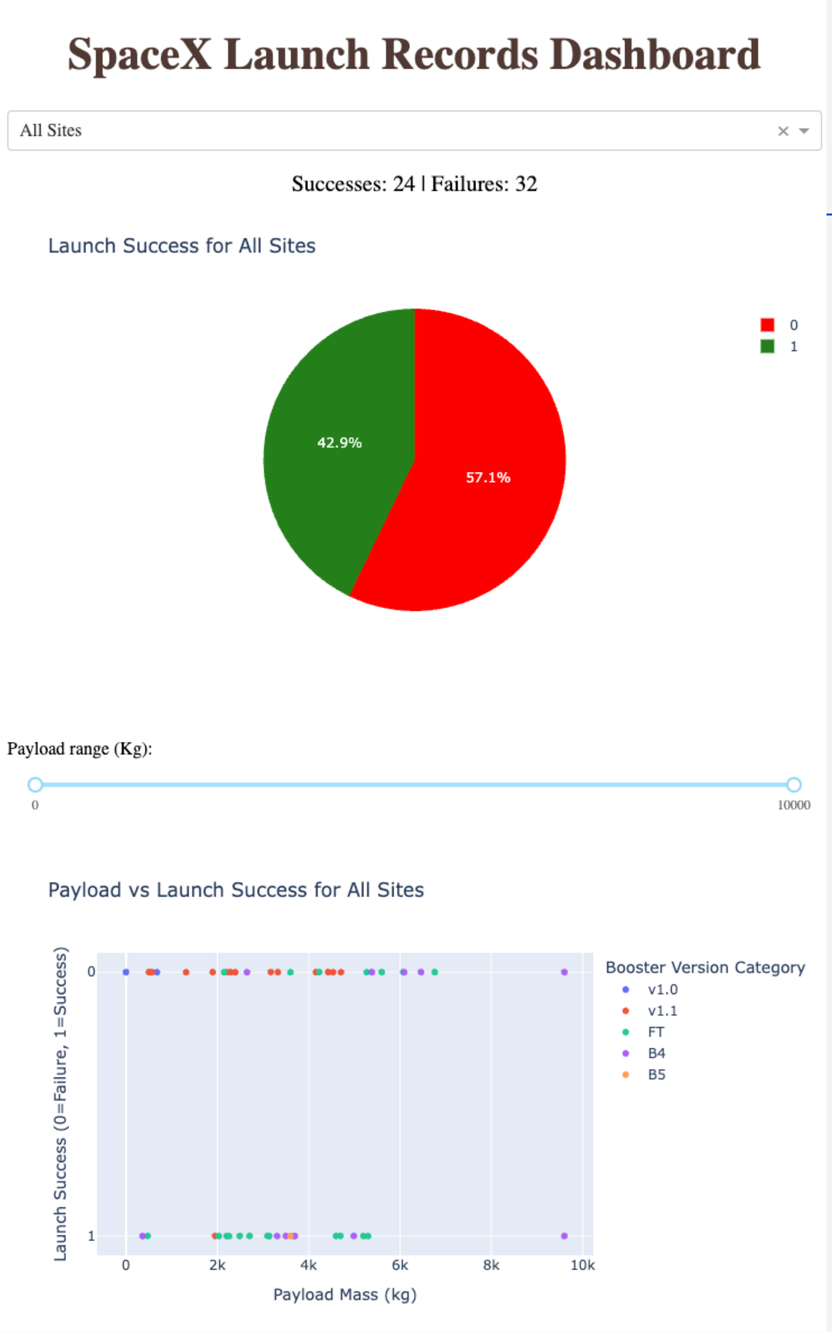
Predictive Analysis (Classification)

- Data was transformed using StandardScaler and split into train/test data sets. The data sets were systematically evaluated for best fit to predict a successful landing.
    - Logistic regression had an accuracy of 83-94.
    - GridSearchCV was used for support vector machine modeling with an accuracy of 83-88.
    - Decision tree modeling had an accuracy of 61-87.
    - K nearest neighbors yielded accuracy levels of 81-94
- https://github.com/cronecoding/IBM_DS_Final/blob/main/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb



Results

- Exploratory data analysis results:
  - Launch location and Payload differences play a role in predicting the likelihood of success of a given mission.
  - Lighter payloads launched from Florida do best.
- Predictive analysis found that logistic regression had the highest accuracy in predicting a Falcon 9 landing success (94.44%)

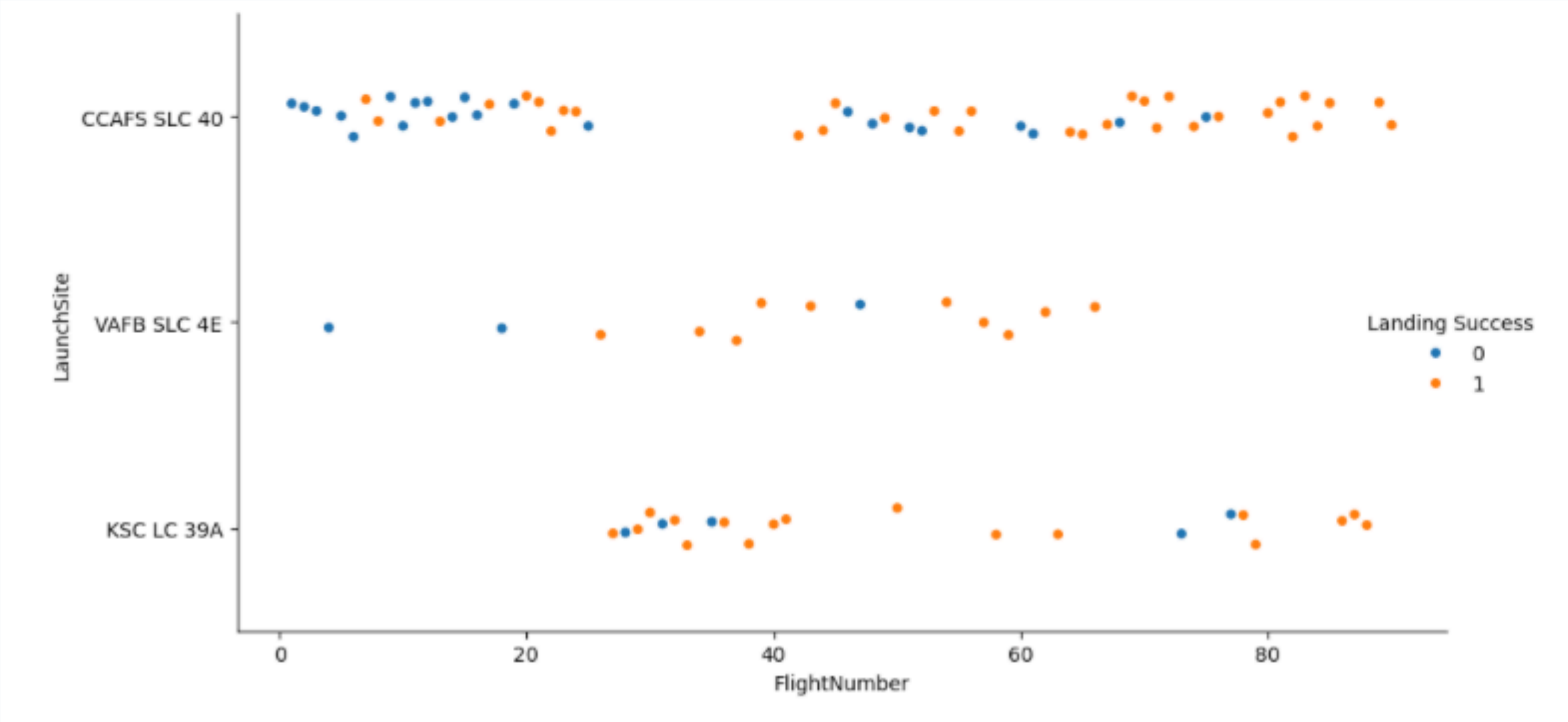


Section 2

Insights drawn from EDA

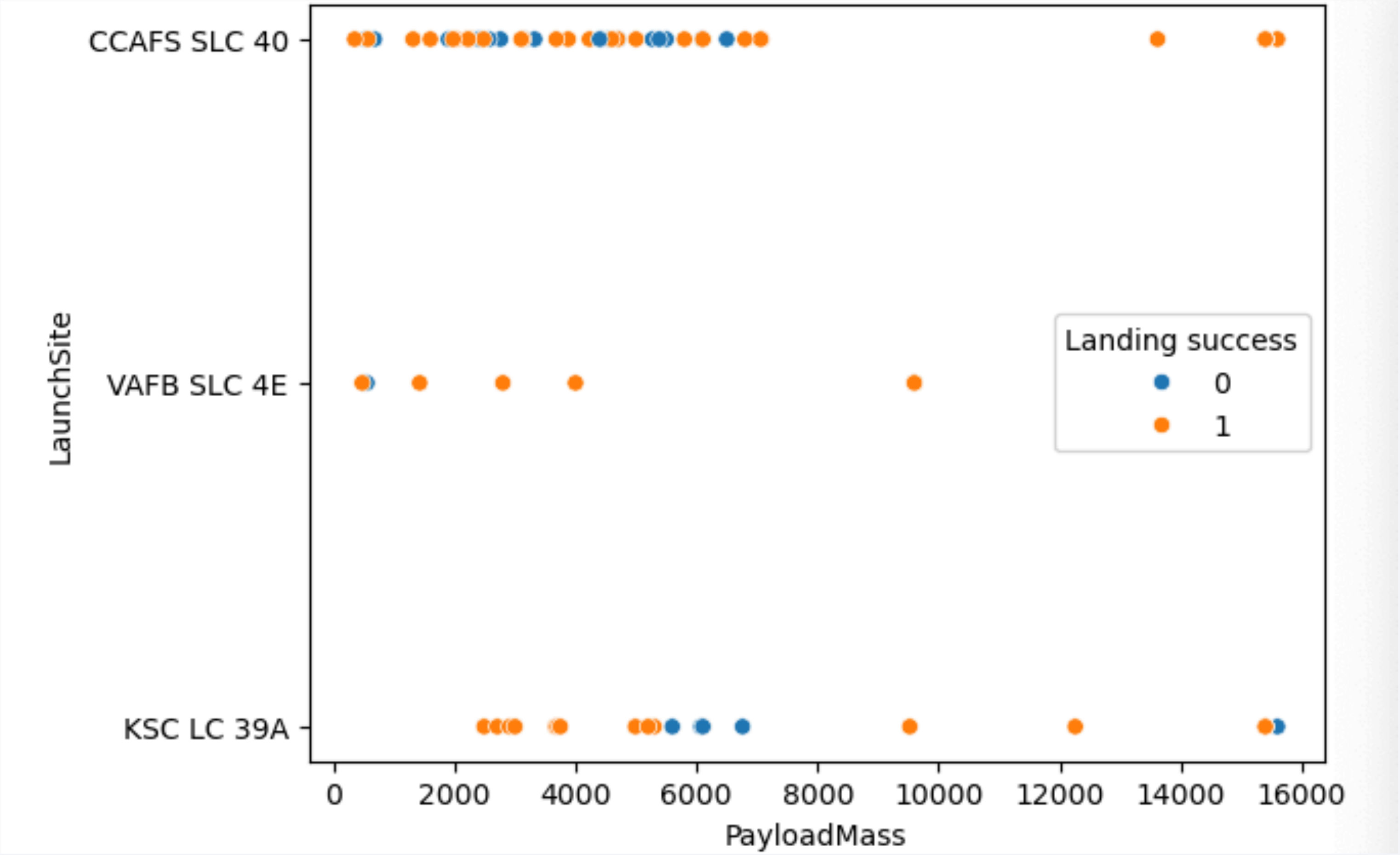
Flight Number vs. Launch Site

Overall, success improves with time at all locations.



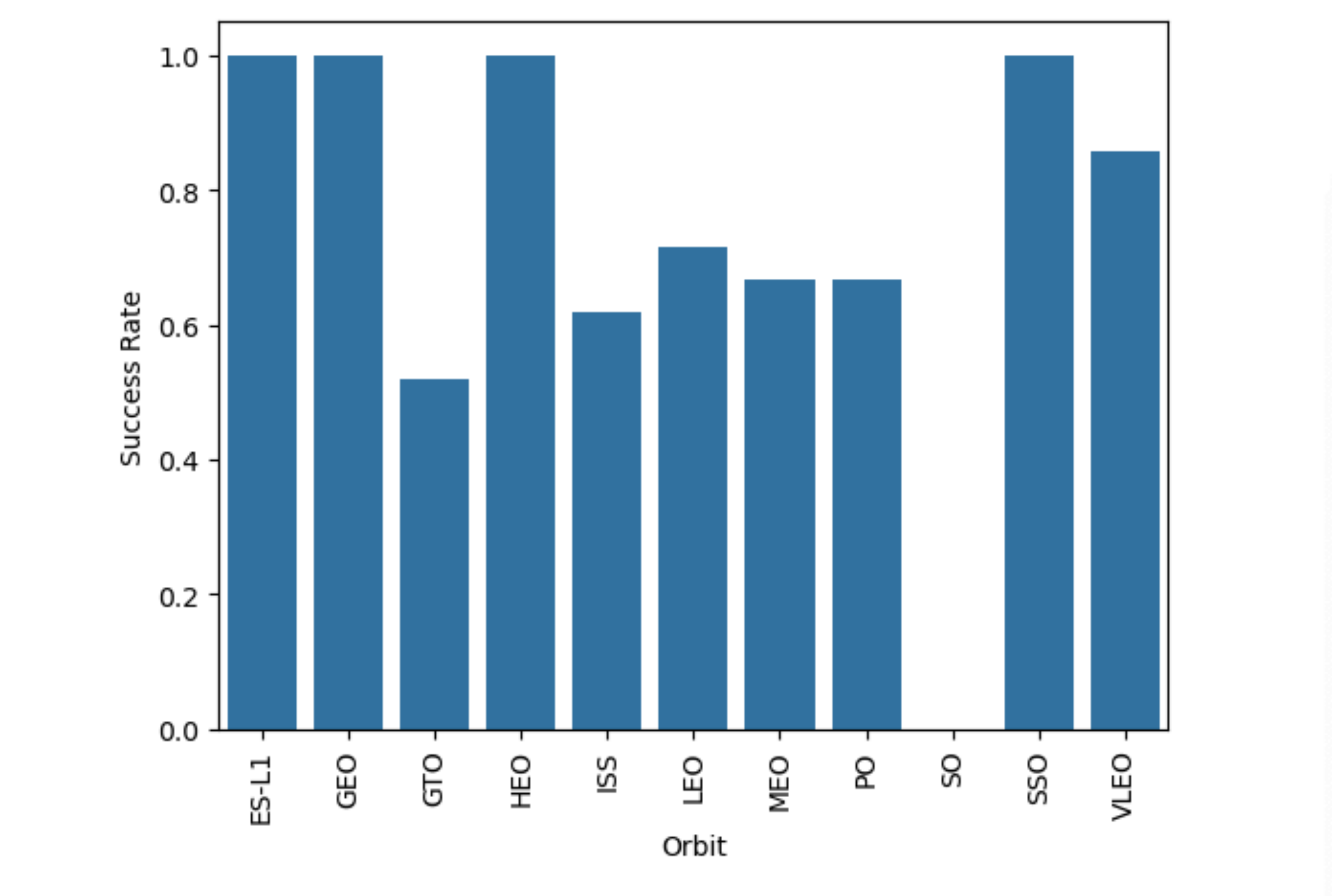
Payload vs. Launch Site

Not all sites use heavy payloads.



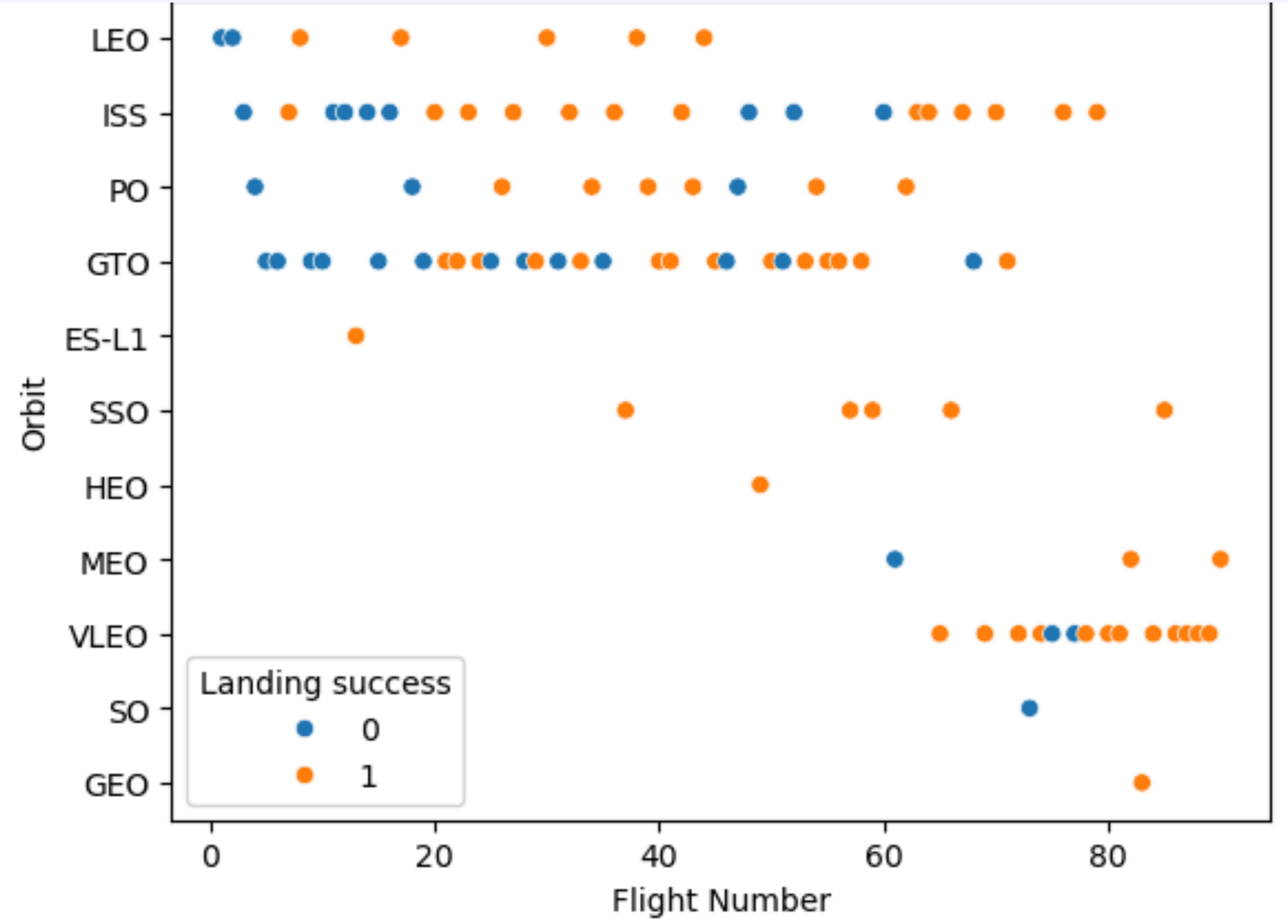
Success Rate vs. Orbit Type

ES-L1, GEO, HEO, and SSO orbits have excellent success rates. Higher and lower orbits are more variable.



Flight Number vs. Orbit Type

Flight success has improved. Many recent successes have used the VLEO orbit.

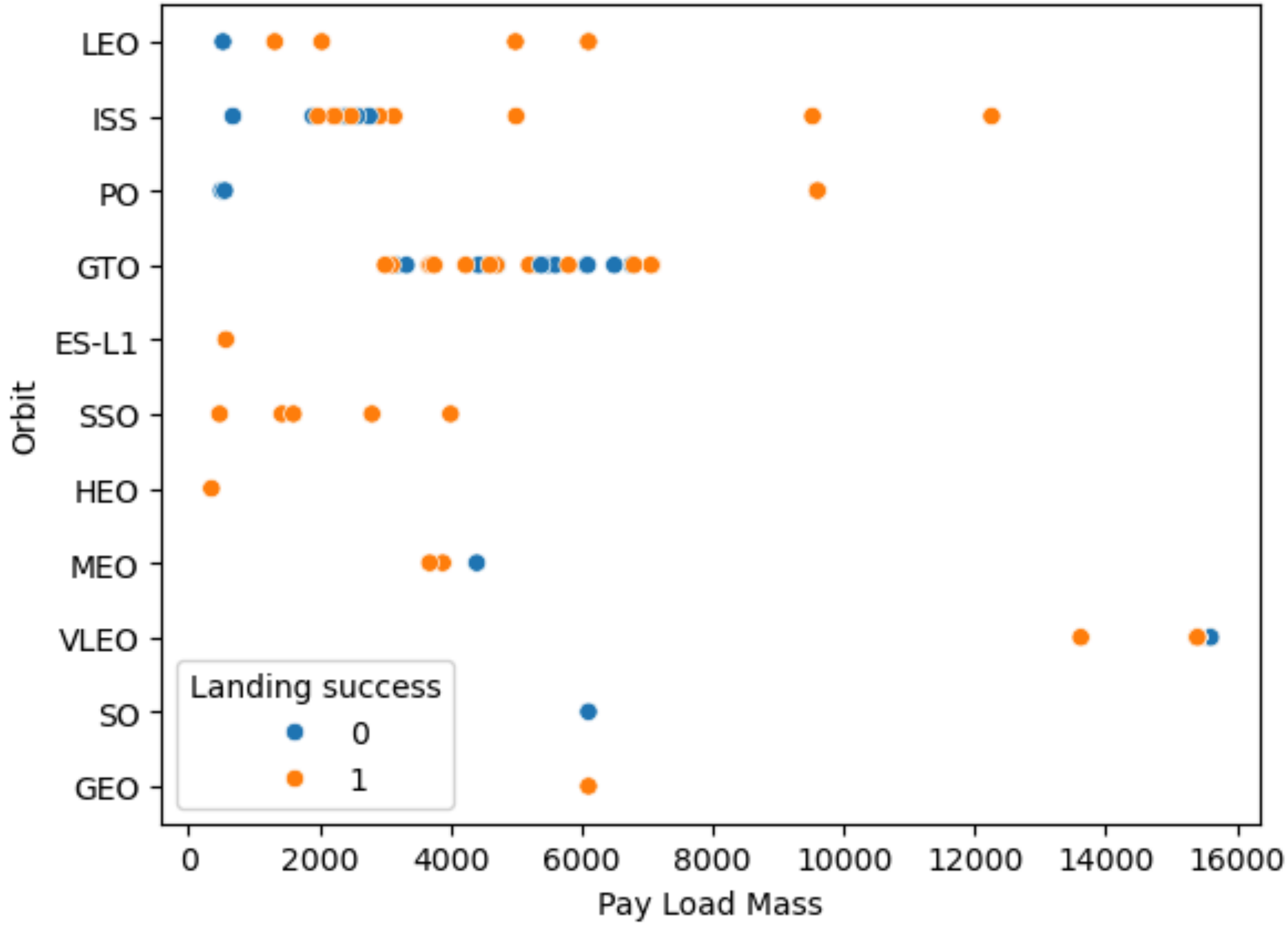


Payload vs. Orbit Type

The ISS orbit has a high success rate even with higher payloads.

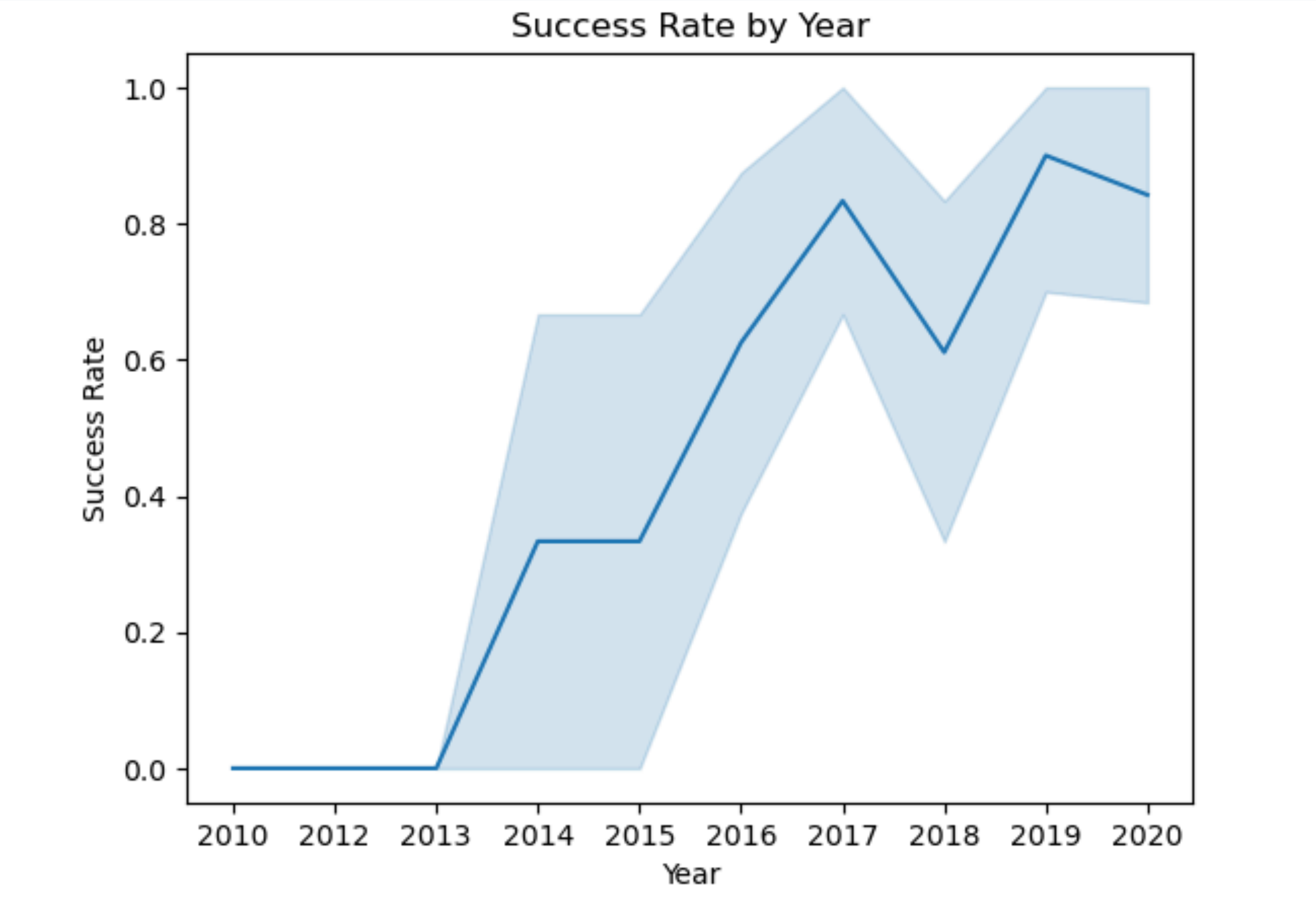
SSO has only been used for lower payload masses

The heaviest payloads have a 67% success in the VLEO orbit



Launch Success Yearly Trend

After initial struggles, launch success rates have increased.



All Launch Site Names

Selecting only distinct entries summarizes the launch sites used by SpaceX.

```
%%sql
SELECT DISTINCT "Launch_Site"
FROM SPACEXTBL
```

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

Launch Site Names Begin with 'CCA'

By using LIKE and “%” all of the Cape Canaveral sites can be seen. This is the first five entries.

```
%%sql
SELECT *
FROM SPACEXTBL
WHERE "Launch_Site" LIKE "CCA%"
LIMIT 5
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)

Total Payload Mass

- Calculating the total payload carried by boosters from NASA uses the SELECT SUM() command and limits the results to those WHERE the customer is NASA.

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE CUSTOMER = "NASA (CRS)"
```

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Calculating the average payload mass carried by booster version F9 v1.1 uses the SELECT AVG() command. This was limited to Booster versions that included F9 version 1.1.

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_)
FROM SPACEXTBL
WHERE "BoOster_Version" LIKE "F9 v1.1%"
```

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

Finding the dates of the first successful landing outcome on the ground pad required limiting dates to only successful landing outcomes and ordering by ascending order.

```
%%sql
SELECT DATE
FROM SPACEXTBL
WHERE "Landing_Outcome" LIKE "Success%"
ORDER BY DATE
limit 1
```

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Listing the names of boosters that have successfully landed on a drone ship with a payload mass greater than 4000 but less than 6000, required multiple specifications within the WHERE command.

```
%%sql
SELECT DISTINCT "Booster_Version"
FROM SPACEXTBL
WHERE "Landing_Outcome" LIKE "Success%"
  AND "Landing_Outcome" LIKE '%drone ship%'
  AND PAYLOAD_MASS__KG_ > 4000
  AND PAYLOAD_MASS__KG_ < 6000;
```

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

Total Number of Successful and Failure Mission Outcomes

Calculated the number of successful and failure mission outcomes by selecting outcomes classified using success or failure. However, other descriptions were also included, for example, “No attempt” or “controlled Ocean landing.” As these were not concluded with the definition of success, they were not analyzed as either successes or failures.

```
%%sql
SELECT
  CASE
    WHEN "Landing_Outcome" LIKE '%Failure%' THEN 'Failure'
    WHEN "Landing_Outcome" LIKE '%Success%' THEN 'Success'
    ELSE 'Other'
  END AS Outcome,
  COUNT(*) AS Total
FROM SPACEXTBL
GROUP BY Outcome;
```

Outcome	Total
Failure	10
Other	30
Success	61

Boosters Carried Maximum Payload

Listing the names of the boosters that have carried the maximum payload mass used a subquery.

```
%%sql
SELECT DISTINCT "Booster_Version"
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL
);
```

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 required using the SUBSTR command and multiple WHERE statements.

```
%%sql
SELECT
    SUBSTR("Date", 6, 2) AS Month,
    "Landing_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM SPACEXTBL
WHERE "Landing_Outcome" LIKE '%Failure%'
  AND "Landing_Outcome" LIKE '%drone ship%'
  AND SUBSTR("Date", 1, 4) = '2015';
```

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

```
%%sql
SELECT "Landing_Outcome", COUNT(*) AS OutcomeCount
FROM SPACEXTBL
WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY OutcomeCount DESC;
```

Landing_Outcome	OutcomeCount
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

Section 3

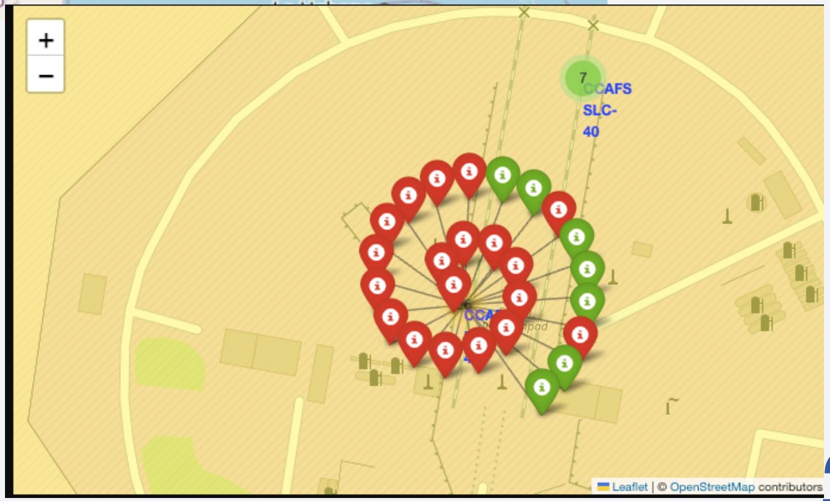
Launch Sites Proximities Analysis

Folium Map of SpaceX Launch Sites

Launch sites are situated on the east and west coast. Both are close to the ocean. The Florida location has been the site of the majority of launches.

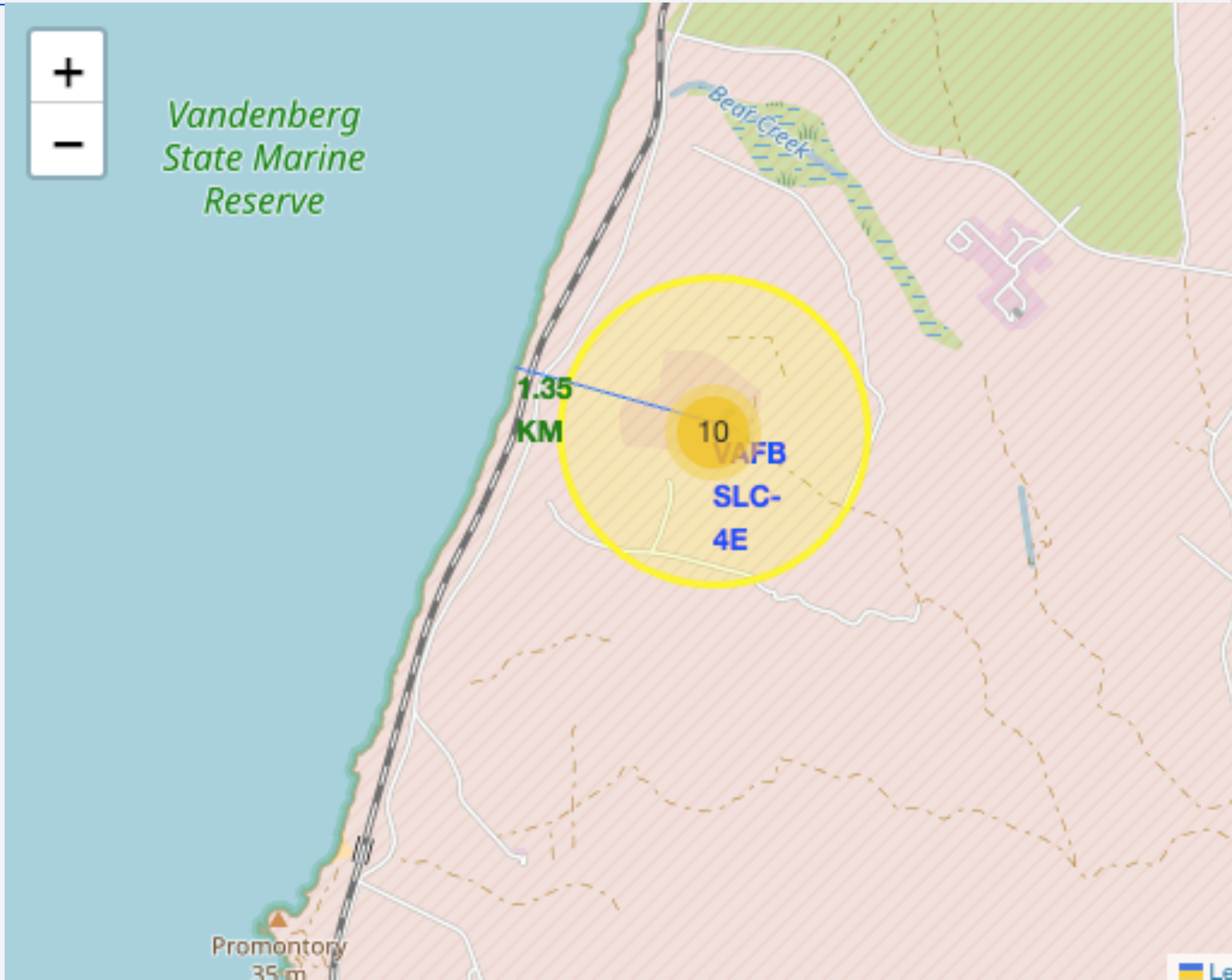
Launch successes via Folium map view

Click on the map to see mission success (green) or failures (red). The map can be zoomed in for greater detail.



Using Folium Map to calculate distance between objects

Using this tool, one can see that the SpaceX West Coast launch site is near key transport lanes, including Highway 1 and a central rail line. It is only 1.35km from the Pacific Ocean.



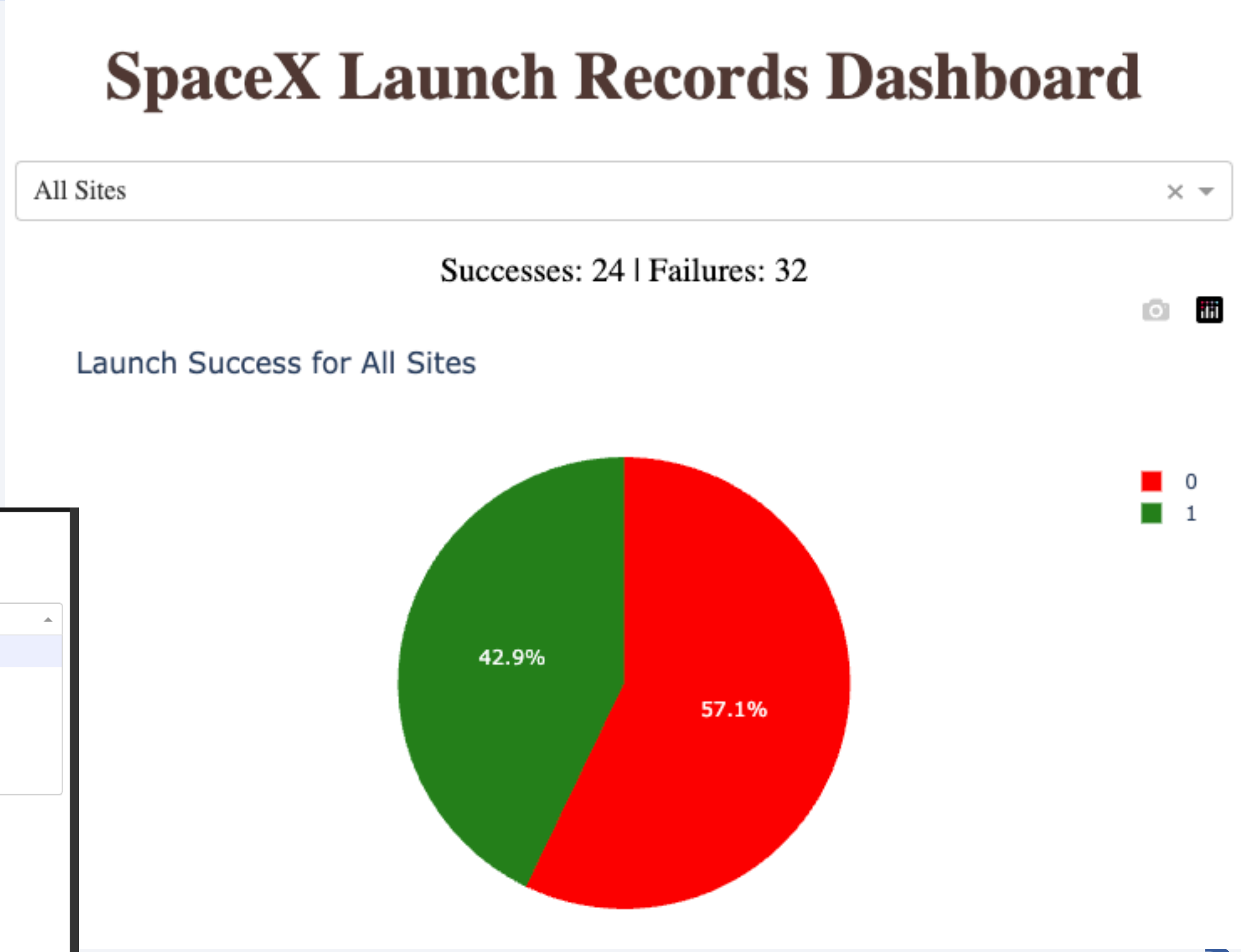
Section 4

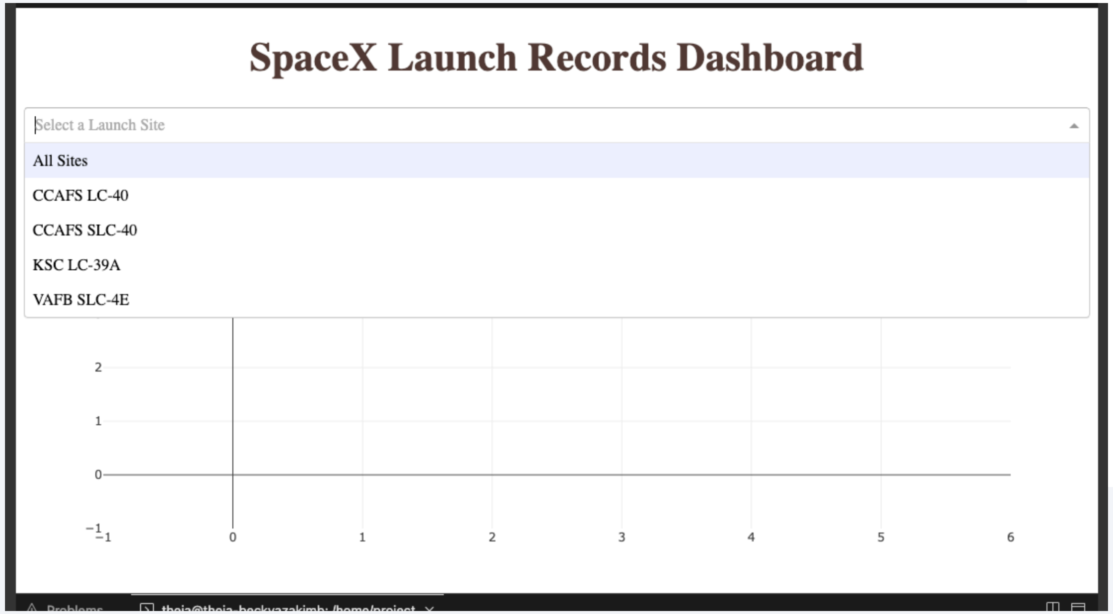
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard

Across all launch sites, SpaceX Falcon 9 rockets have a 42.9% success rate for returning to Earth in a usable fashion.

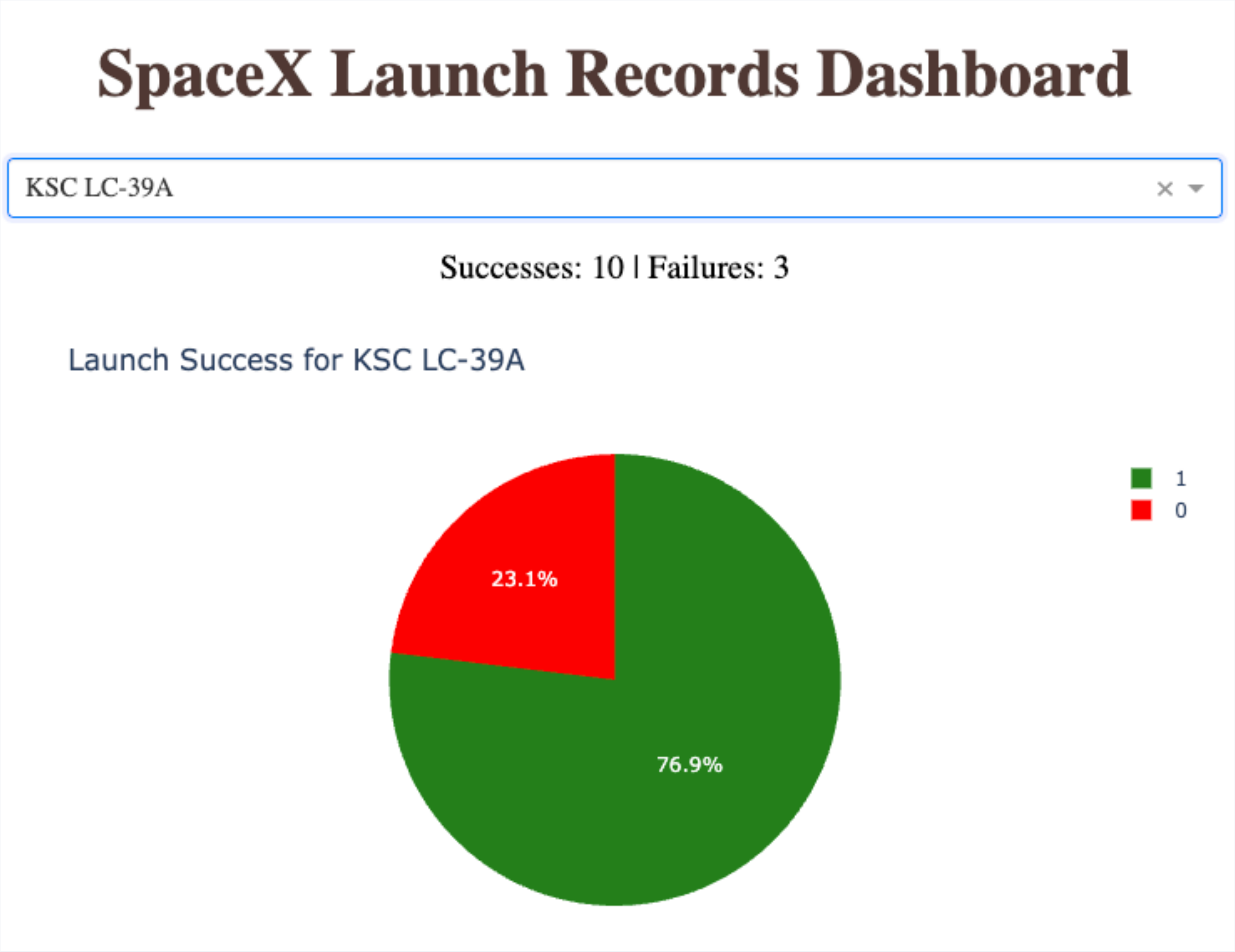
The dashboard drop-down menu allows for viewing specific site success patterns as well.





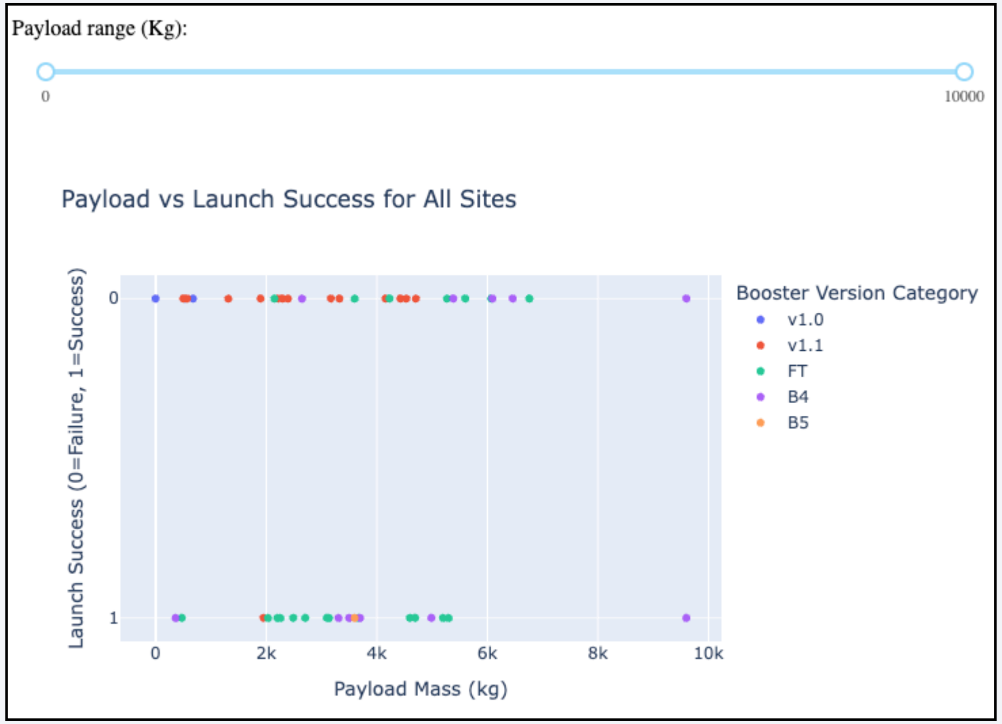
Dashboard for Individual Launch Sites

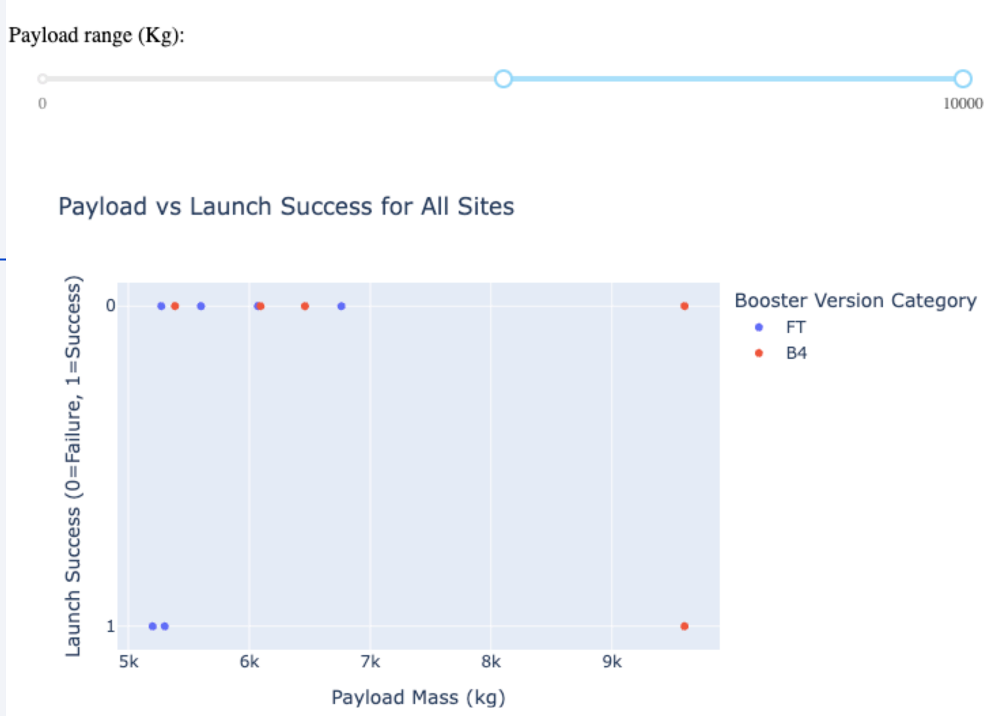
KSC LC-39A (Kennedy Space Center) has the highest success rate with 77%.



Payload slider feature

- The Payload slider allows one to easily change the data shown. It highlights the challenges of heavy (>5000 kg) payloads versus lighter loads.





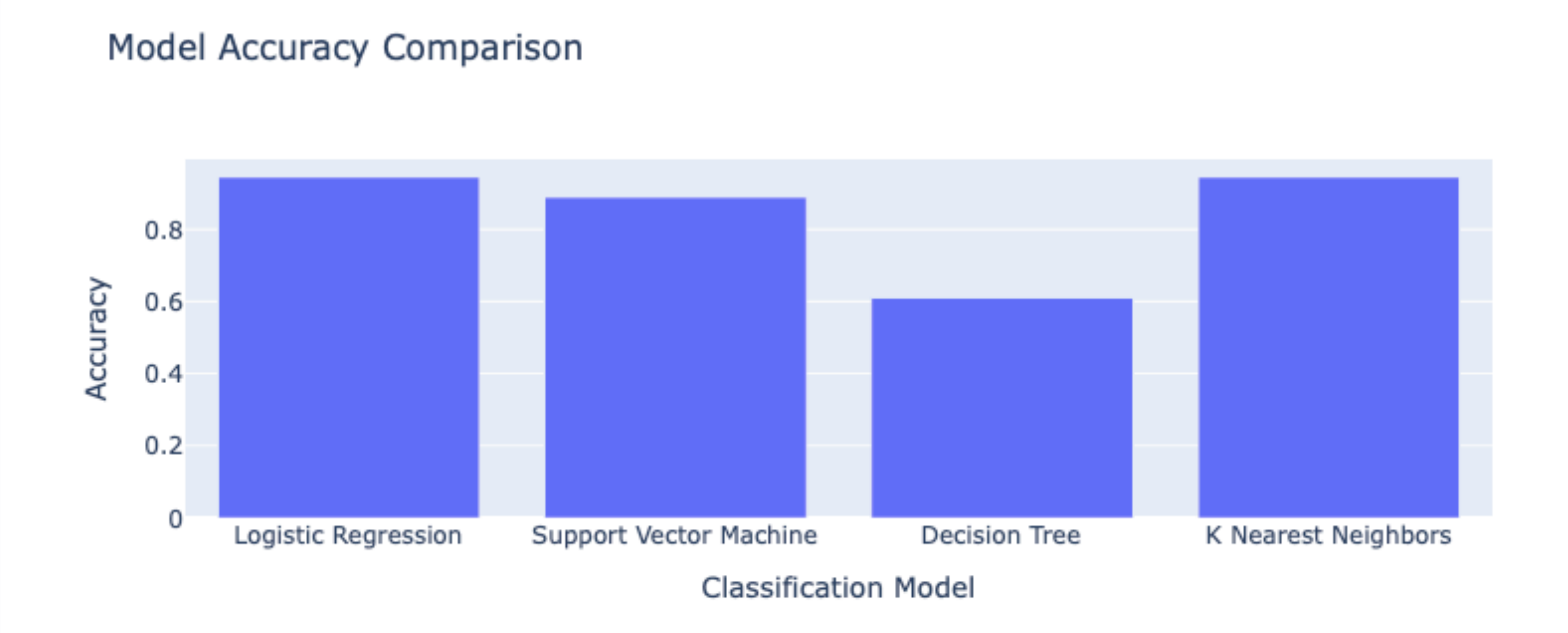
Section 5

Predictive Analysis (Classification)

Classification Accuracy

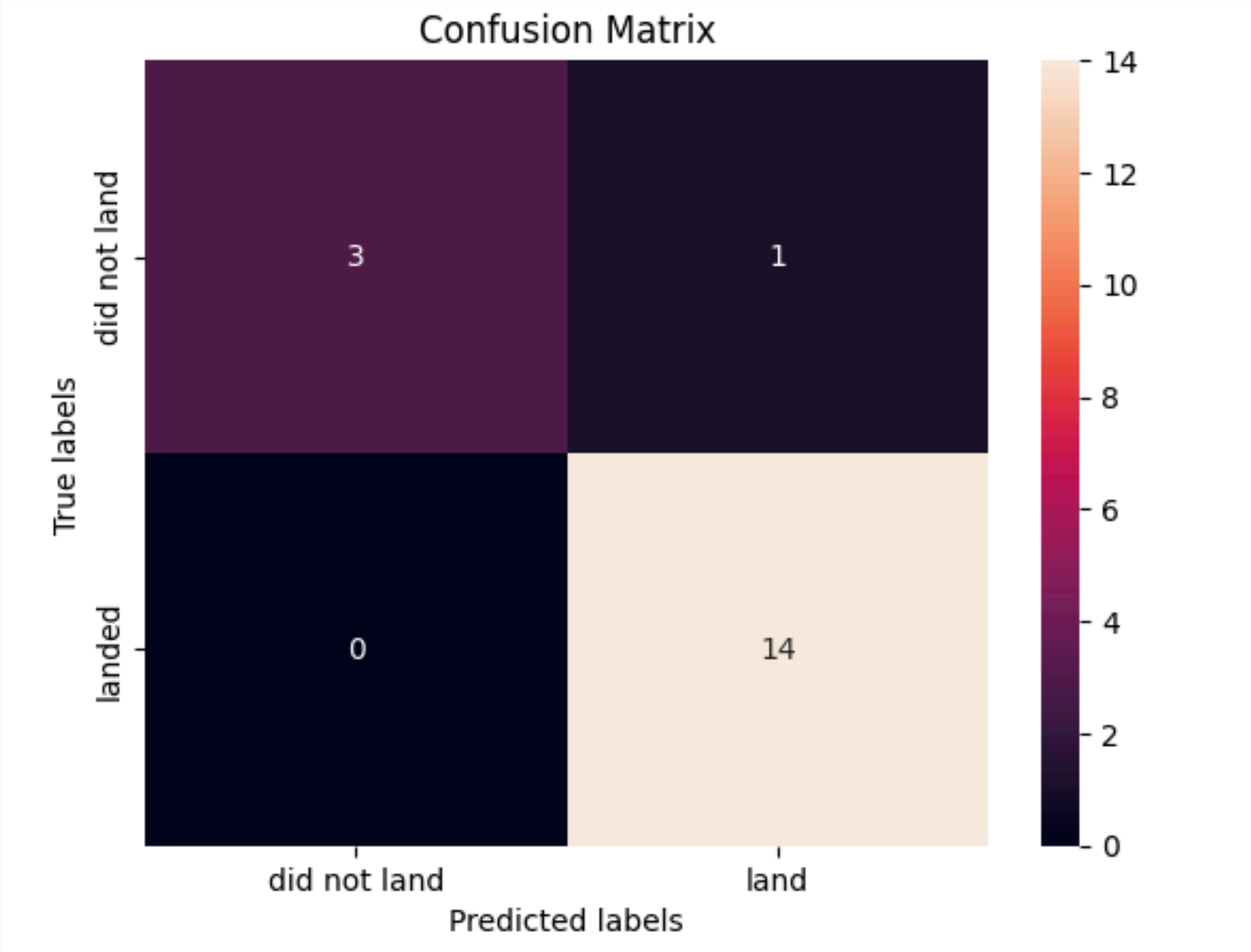
Logistic Regression achieved the highest Accuracy.

KNN and SVM were close seconds.



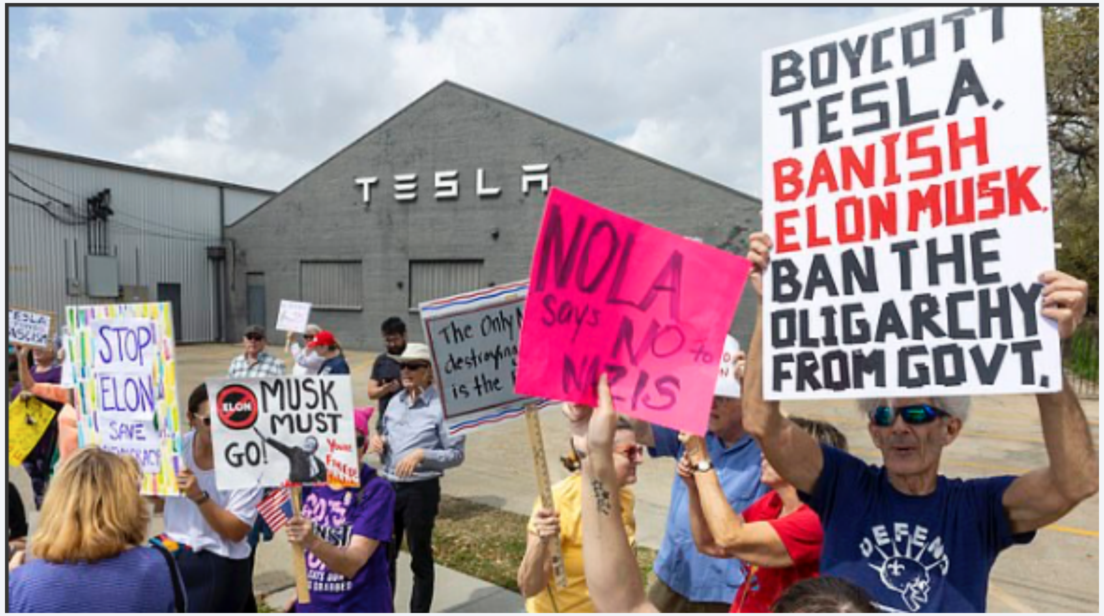
Confusion Matrix

The Logistic Regression confusion matrix shows excellent prediction of both accurate landings (14/15) and failures (3/3).



Conclusions

1. Elon Musk’s signature enterprise, SpaceX, has a less than 50% success rate, failing to retrieve a usable rocket after launch most of the time.
2. Logistic Regression can make accurate predictions of the success or failure of a mission using a Falcon 9 booster rocket based on orbit, launch site, and payload mass.
3. The crooked bastard needs to get his fancy fingers out of our government.



Appendix

- The views of this project are strictly my own and are not (necessarily) endorsed by any other group I am a part of.
- All code should be available on my GitHub site (https://github.com/cronecoding) along with other projects I am working on.
- Thank you to my husband and dogs for putting up with this intellectual exploration without financial compensation.
- If you like my work, please hire me.

Thank you!