IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

JP Galik
May 16, 2025

Outline

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

Executive Summary

Capstone Summary: Predicting SpaceX Falcon 9 Landings

This project focused on predicting the success of Falcon 9 first-stage landings using machine learning and data visualization. I collected and analyzed historical launch data to uncover patterns and build predictive models.

- Data sourced from SpaceX API and Wikipedia (web scraping)
- Cleaned and explored with SQL and EDA
- Engineered new features and standardized the dataset
- Built interactive visualizations using Folium and Plotly Dash
- Applied SVM, KNN, and Decision Tree models; tuned with GridSearchCV
- Launch site and payload mass were key predictors of success
- Decision Tree identified as the most effective model

Introduction

- Project background and context:
  - This project predicts whether the Falcon 9 first stage will land successfully. SpaceX lowers launch costs by reusing this stage, offering launches at $62M versus $165M+ from competitors. Accurately predicting landing success helps estimate costs and provides insights for companies competing with SpaceX.

- Problems you want to find answers
  - What launch conditions and features influence the successful landing of the Falcon 9 first stage?
  - Among the models tested with machine learning, algorithm performs best in predicting landing success?

Section 1

Methodology

Methodology

Executive Summary: This project uses a full data science pipeline—including collection, processing, analysis, visualization, and modeling—to predict Falcon 9 first-stage landing success.

- Data collection methodology:
  - Launch data was obtained from the SpaceX API, offering detailed information on Falcon 9 missions, including launch dates, sites, payloads, and outcomes.
- Perform data wrangling
  - Data was cleaned for consistency, with key features extracted and new ones engineered to enhance predictive value.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Used SQL queries to find data driven answers from the data
  - Visualized relationships between the launch sites and success rates

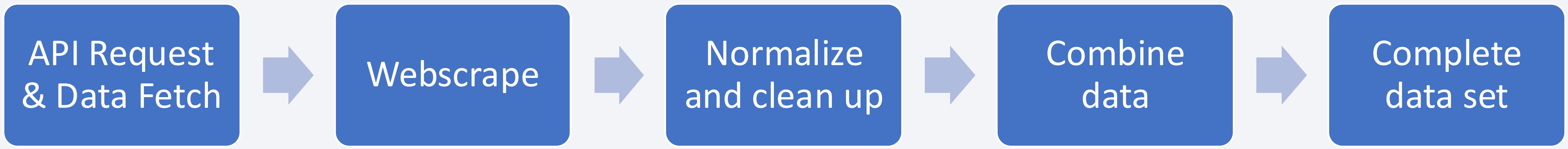
Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
  - Made interactive maps displaying launch sites using Folium
  - Showed the successful and unsuccessful launch as well as distances between their locations
- Perform predictive analysis using classification models
  - Developed and assessed several predictive classification models.
  - Identified the best-performing model for predicting landing success

Data Collection

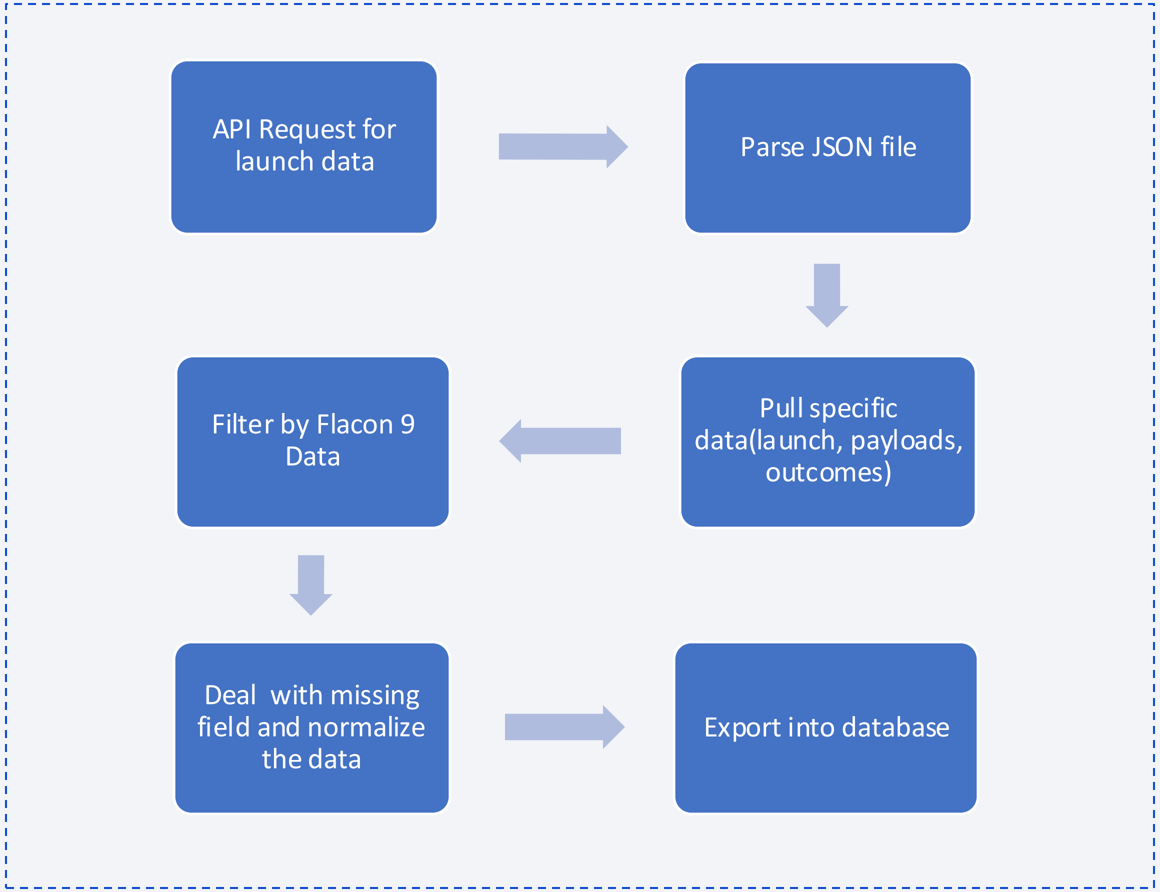
- Data collection was doing from an HTTP request from SpaceX APIs. Additionally, Web Scraping was done to extract more info. The data was stored, cleaned up, and combined into a final set.



Data Collection – SpaceX API

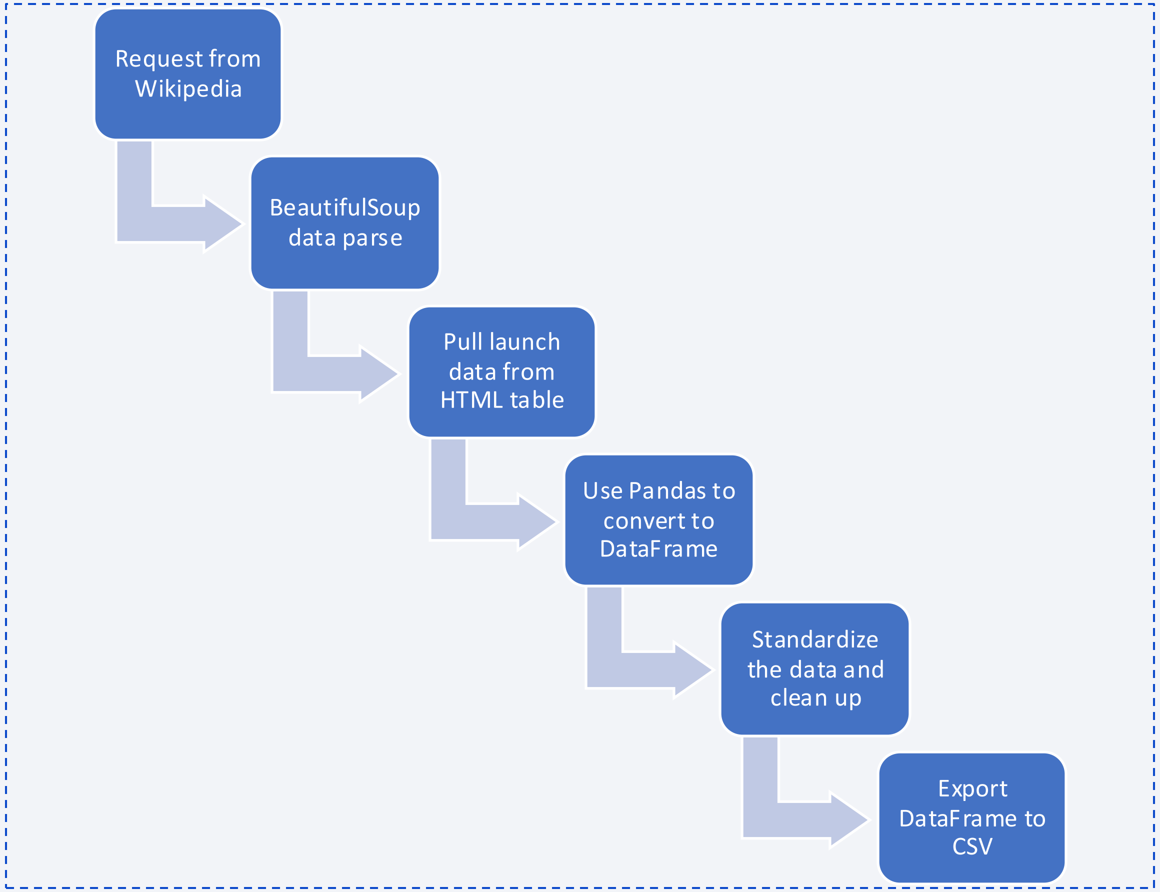
- API request using Python from SpaceX: v4/launches
- The response came back as a JSON file, so I converted it to Python
- The applicable fields needed were extracted to build a proper data frame
- Falcon 9 launch data was filtered into the final set and exported and stored for this exercise.

- Github URL: https://github.com/jgalik182/tesrepo/blob/main/01_Space_X_Data_Collection.ipynb



Data Collection - Scraping

- Initiate webscraping by fetching content from Wikipedia on the Falcon 9 list and Heavy Launches.
- Parse the data using BeautifulSoup by extracting the HTML table and its' content.
- Using Pandas, convert the table in order to build a DataFrame.
- Clean up the DataFrame and export to CSV to be used in the next steps.
- Github URL: https://github.com/jgalik182/tesrepo/blob/main/02_Space_X_Web_Scrapping.ipynb



Data Wrangling

- First the data needed to be cleaned up and normalized
  - Find missing values or rows and delete ones that did not have enough relevant information
  - Whitespace, standardization of words, numbers, and dates
- Exploratory Data Analysis (EDA) and Integration
  - Launches, sites, orbits occurrence, and outcomes from the mission were explored.
  - The dataset was merged into a consistent DataFrame
- Final cleaned up document
  - The DataFrame was converted to CSV.
- GitHub URL: https://github.com/jgalik182/tesrepo/blob/main/03_Space_X_Data_Wrangling.ipynb

EDA with Data Visualization

- To understand the data, various charts and visuals were used to display summarized characteristics of the data.
- Plots show us a key insights as to whether there was a strong relationship between two variables such as flight number vs payload mass vs launch site etc.
- Histograms shows us the spread and tendency of the data as well as any outliers that can skew the results.
- Line charts shows us trends over time of the success rate of the Falcon 9 launches over time which increased since 2013.
- GitHub URL: https://github.com/jgalik182/tesrepo/blob/main/05_Space_X_EDA_Data_Visualization.ipynb.ipynb

EDA with SQL

- Unique launch sites
- The first 5 records that were launch sites that began with “CCA”
- Total payload mass carried by boosters launched by NASA
- Average payload mass carried by Falcon 9 v1.1 booster
- Successful landing date with the outcome of the grounded landing
- Successful booster outcomes with payload mass between 4000kg and 6000kg
- List of failed and successful missions
- Booster versions and their max payload mass
- Failed outcomes on drone ship in 2015 with relevant information including month, booster, and launch site
- GitHub URL: https://github.com/jgalik182/tesrepo/blob/main/04_Space_X_EDA_SQL.ipynb

Build an Interactive Map with Folium

- An interactive map was made to show geographical information on the launch sites:
  - Launch site info labeled with markers, popup and text labels
  - Longitude and latitude coordinates for each launch site as well as distance between them
  - Landing outcomes dictated in color; green for a success and red for failure.

- GitHub URL: https://github.com/jgalik182/tesrepo/blob/main/06_Space_X_Visuals_Folium.ipynb.ipynb.ipynb

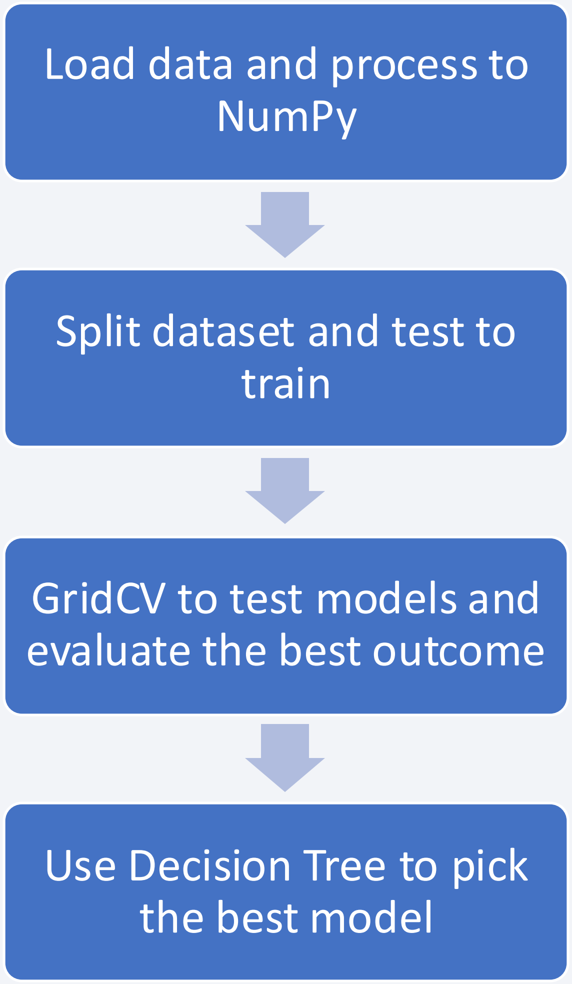
Build a Dashboard with Plotly Dash

- An interactive dashboard was made using Plotly Dash to show Launch site data
  - Up top is a drop down to filter out each launch site to look at individually
  - A pie chart below that displays all or the selected site to show a breakdown of successful launches
  - Beneath that is a scatter plot which shows color coded outcomes for the payload mass and booster version.

- GitHub URL: https://github.com/jgalik182/tesrepo/blob/main/07_Space_X_Visuals_Plotly.py

Predictive Analysis (Classification)

- The DataFrame was loaded and converted to a NumPy array
- The datasets were split using split training and testing data
- GridSearchCV was utilized to test hyperparameters including Logistic Regression, Support Vector Classifier, Decision Tree Classifier, and KNN Classifier
- GitHub URL: https://github.com/jgalik182/tesrepo/blob/main/08_Space_X_Predictive_Analytics.ipynb

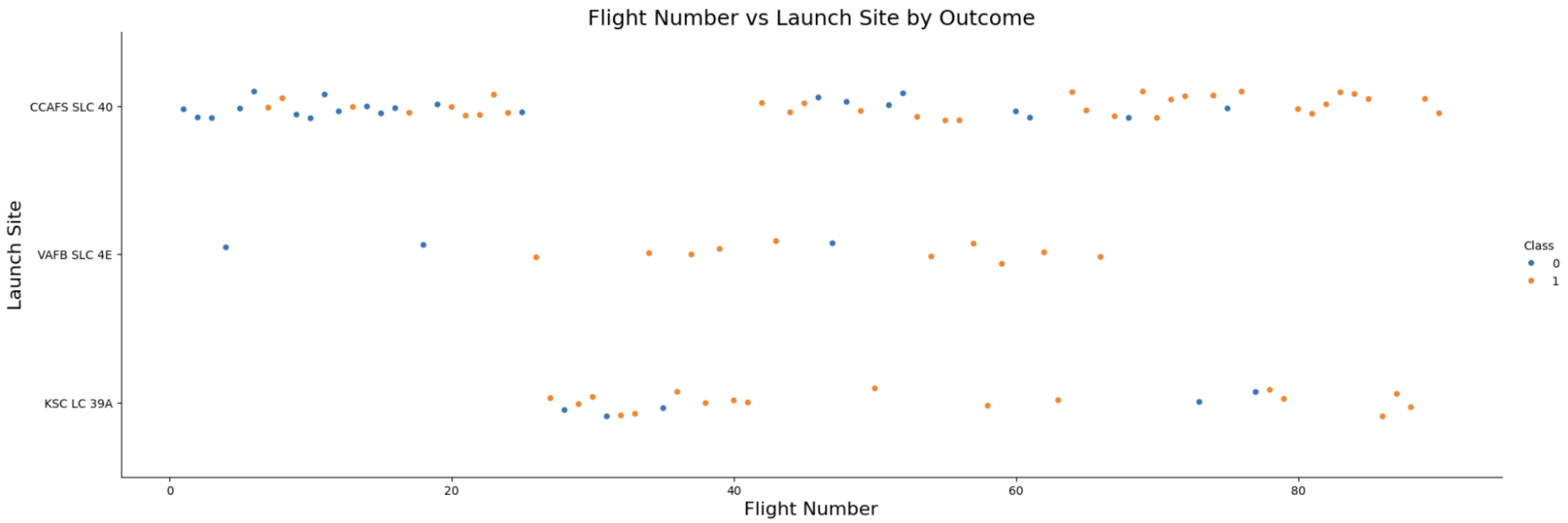


Results

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

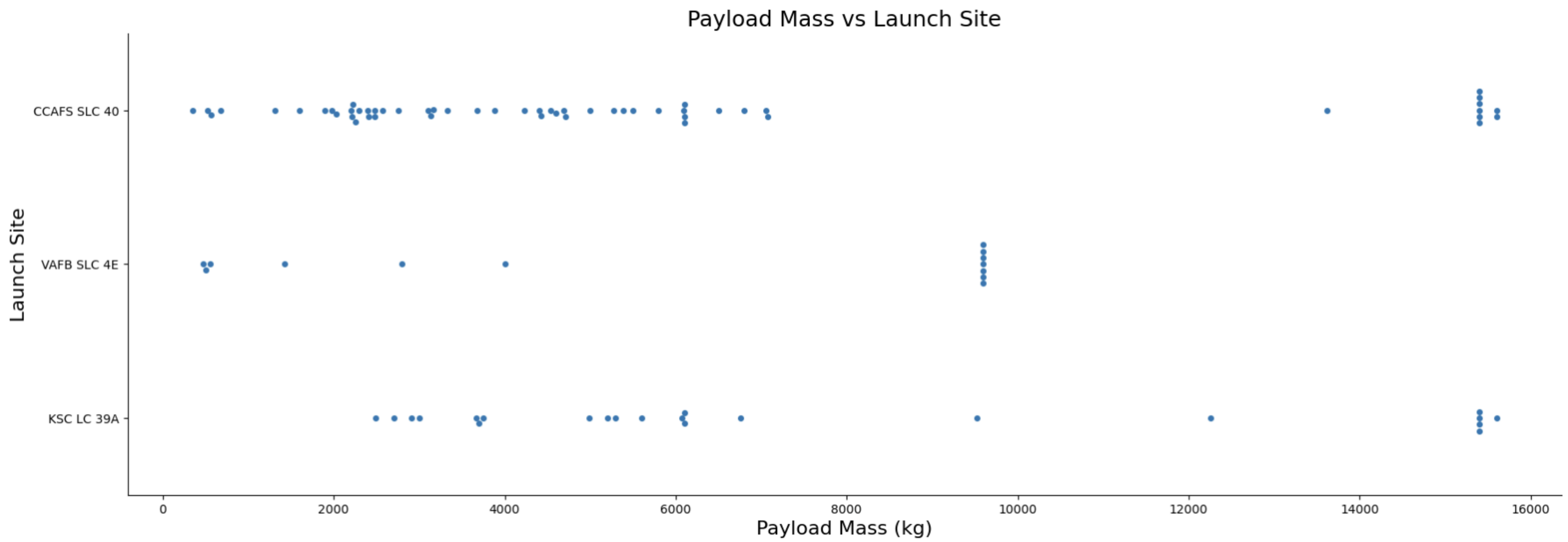
Section 2

Insights drawn from EDA



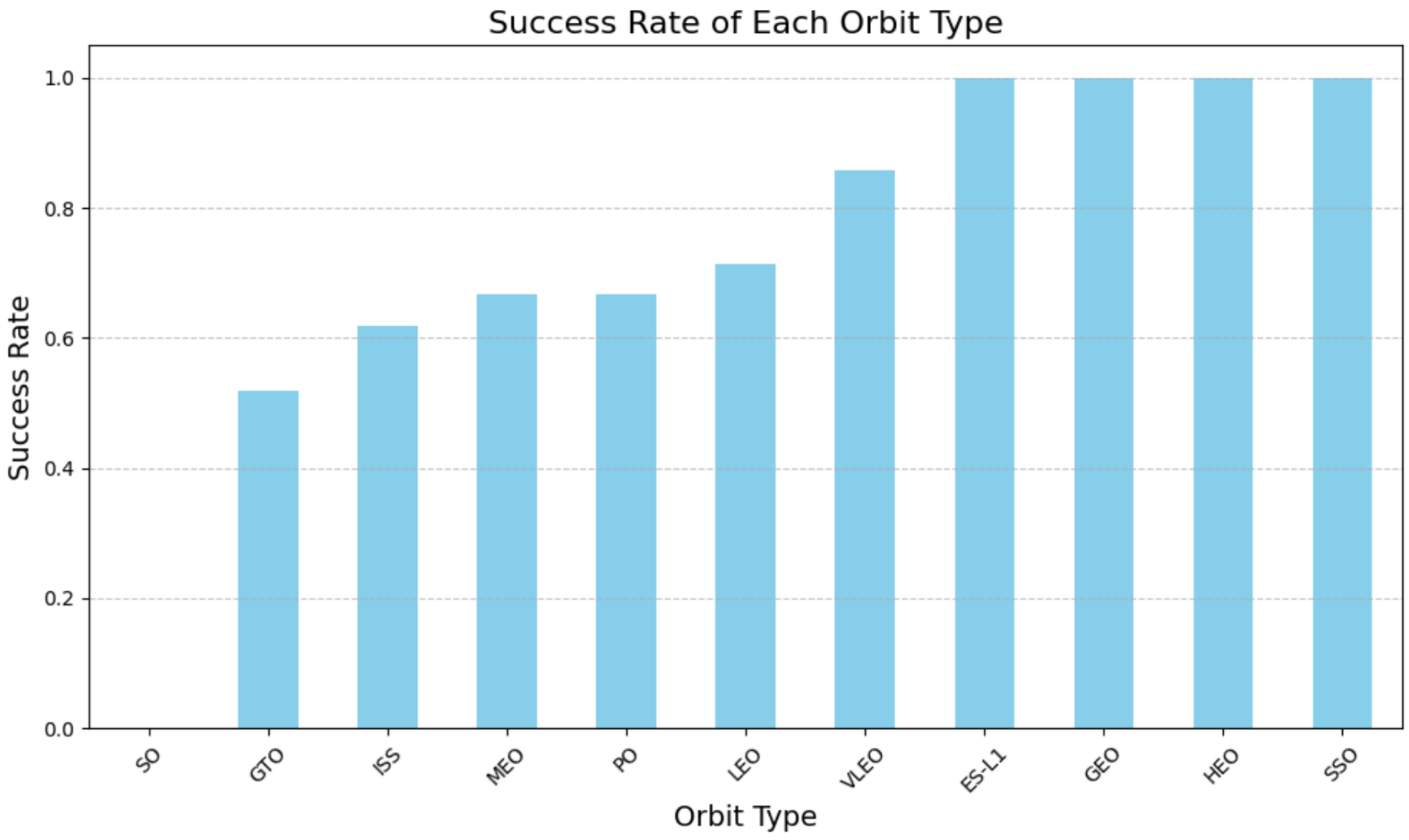
Flight Number vs. Launch Site

- Early on, all the sites showed success and failures, however, a clear upward trend in successful landings happened over time.
- Most of the launches happen at the CCAFS SLC 40 site.
- Even with less launches, VAFB SLC 4E & KSC LC 39A sites have less failures and a higher successful landing outcome.



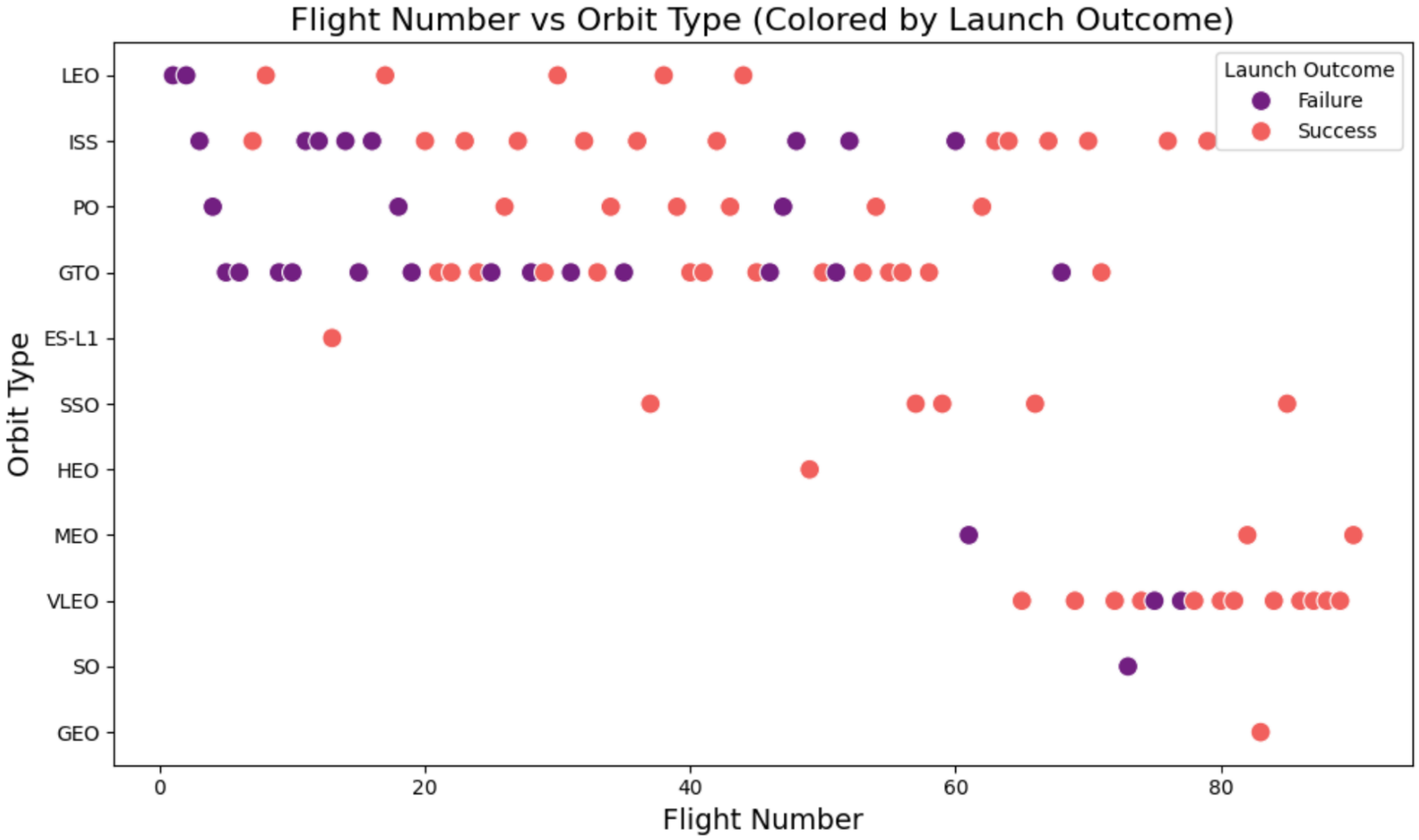
Payload vs. Launch Site

- The CCAFS SLC 40 site handles most of the launches with payloads under 8,000 KG while the other sites can handle higher ranged payloads.
- KSC LC 39A handles higher payloads while site VAFB SLC 4E handles numerous medium payloads for their missions.



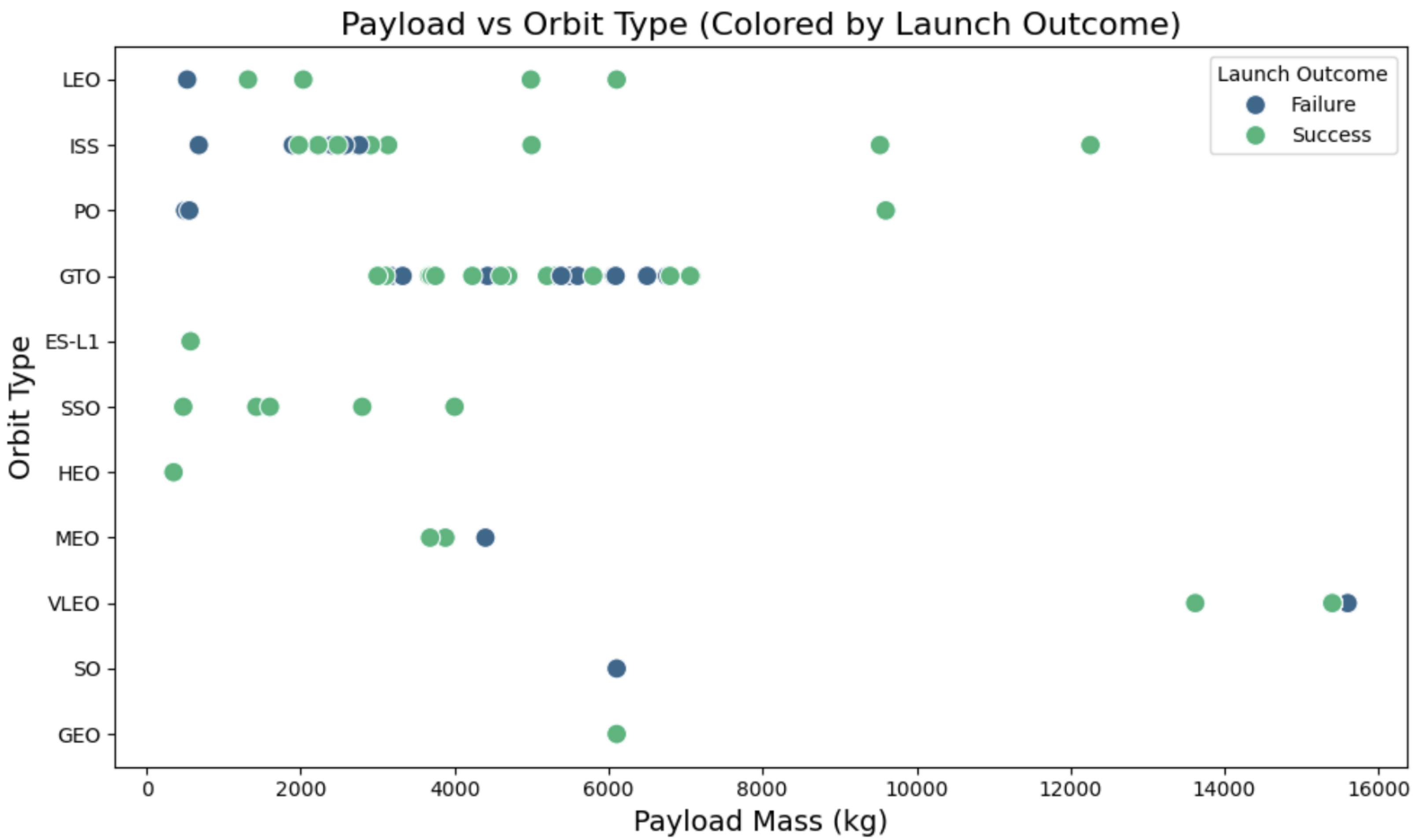
Success Rate vs. Orbit Type

- The orbit types that received the highest success rate include ES-L1, GEO, HEO, and SSO
- The orbit types that received moderate success include GTO, ISS, MEO, PO, LEO, and VELO
- The single orbit type that had failed was SO.



Flight Number vs. Orbit Type

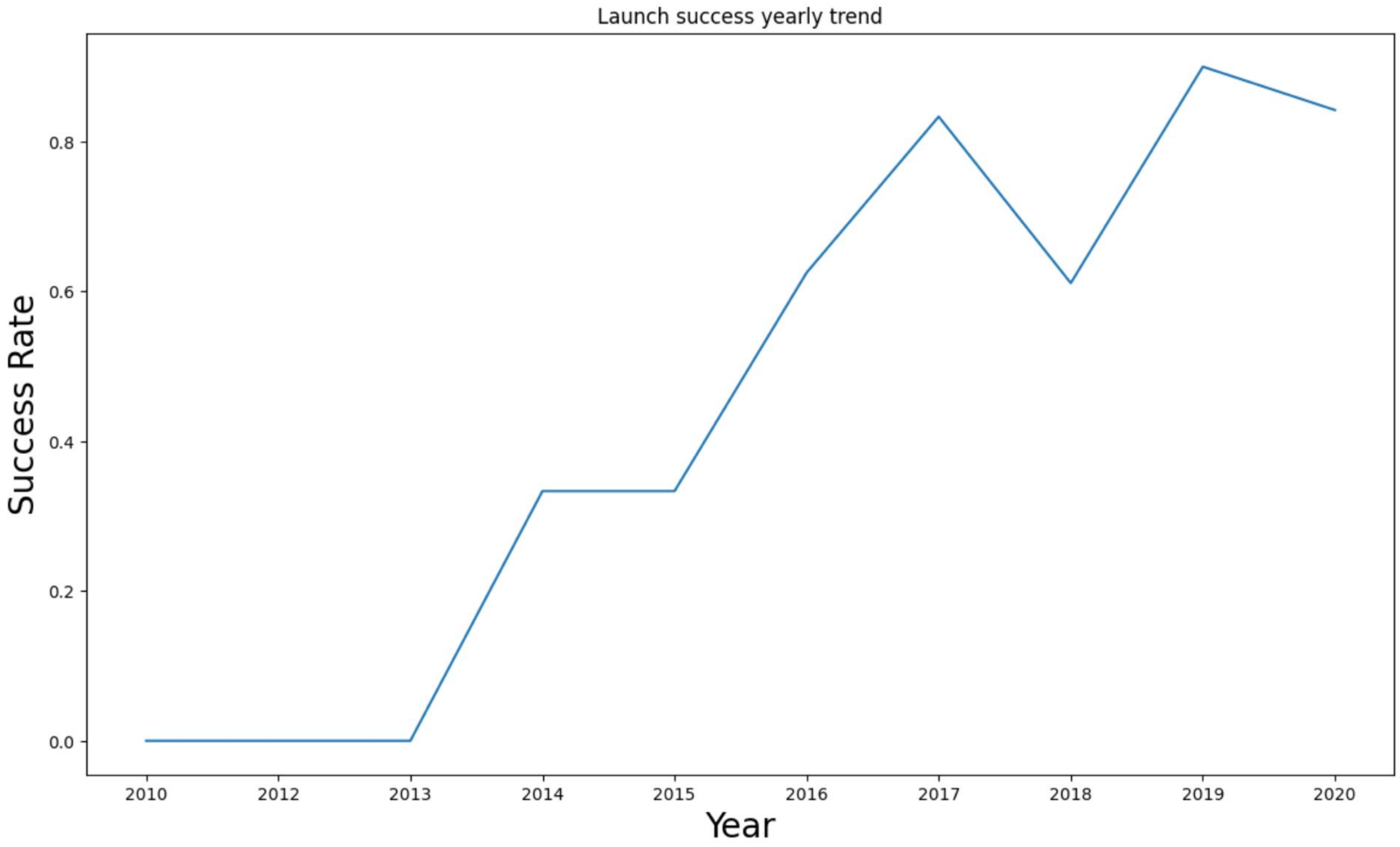
- Early on, the Falcon 9 launches failed mostly, then, over time there was a 50/50 ratio of success, and eventually after 70 launches, missions were 100% a success.
- In addition to the orbits LEO, ISS, PO, & GTO being used for flights, orbit VELO was deployed after 60 flights with mostly successful outcomes.



Payload vs. Orbit Type

- The most common pay load mass for flights is under 8,000 kg with about 25% failure.

- Flights with a higher payload above 8,00 kg are less frequent, but all but one outcome resulted in a success.



Launch Success Yearly Trend

- Since 2013, the success rate for each launch has increased steadily year over year.
- 2018 showed a slip decrease in success, but in 2019, the rate climbed back up, showing that the Falcon 9 team learned from what happened the previous year.

All Launch Site Names

Task 1

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

```
%sql select distinct "Launch_Site" FROM SpaceXtable;
```

```
* 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 SPACEXTABLE 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_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
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 (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

Total Payload Mass

Task 3

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

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';
```

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

SUM("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

Task 4

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

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

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

AVG("PAYLOAD_MASS__KG_")
2928.4

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

```
%sql SELECT MIN("Date") FROM SPACEXTABLE 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

- Query: **%sql** SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (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

```
%sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTABLE WHERE "Mission_Outcome" IN
    ('Success', 'Failure') GROUP BY "Mission_Outcome";
```

Mission_Outcome	Total
Success	98

Boosters Carried Maximum Payload

- **%sql** SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);

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

- **%%sql SELECT CASE WHEN** substr("Date", 6, 2) = '01' **THEN** 'January' **WHEN** substr("Date", 6, 2) = '02' **THEN** 'February' **WHEN** substr("Date", 6, 2) = '03' **THEN** 'March' **WHEN** substr("Date", 6, 2) = '04' **THEN** 'April' **WHEN** substr("Date", 6, 2) = '05' **THEN** 'May' **WHEN** substr("Date", 6, 2) = '06' **THEN** 'June' **WHEN** substr("Date", 6, 2) = '07' **THEN** 'July' **WHEN** substr("Date", 6, 2) = '08' **THEN** 'August' **WHEN** substr("Date", 6, 2) = '09' **THEN** 'September' **WHEN** substr("Date", 6, 2) = '10' **THEN** 'October' **WHEN** substr("Date", 6, 2) = '11' **THEN** 'November' **WHEN** substr("Date", 6, 2) = '12' **THEN** 'December' **ELSE** 'Unknown' **END AS** "Month_Name", "Mission_Outcome", "Booster_Version", "Launch_Site" **FROM** SPACEXTABLE **WHERE** substr("Date", 0, 5) = '2015';

Month_Name	Mission_Outcome	Booster_Version	Launch_Site
January	Success	F9 v1.1 B1012	CCAFS LC-40
February	Success	F9 v1.1 B1013	CCAFS LC-40
March	Success	F9 v1.1 B1014	CCAFS LC-40
April	Success	F9 v1.1 B1015	CCAFS LC-40
April	Success	F9 v1.1 B1016	CCAFS LC-40
June	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
December	Success	F9 FT B1019	CCAFS LC-40

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.

```
%%sql

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

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

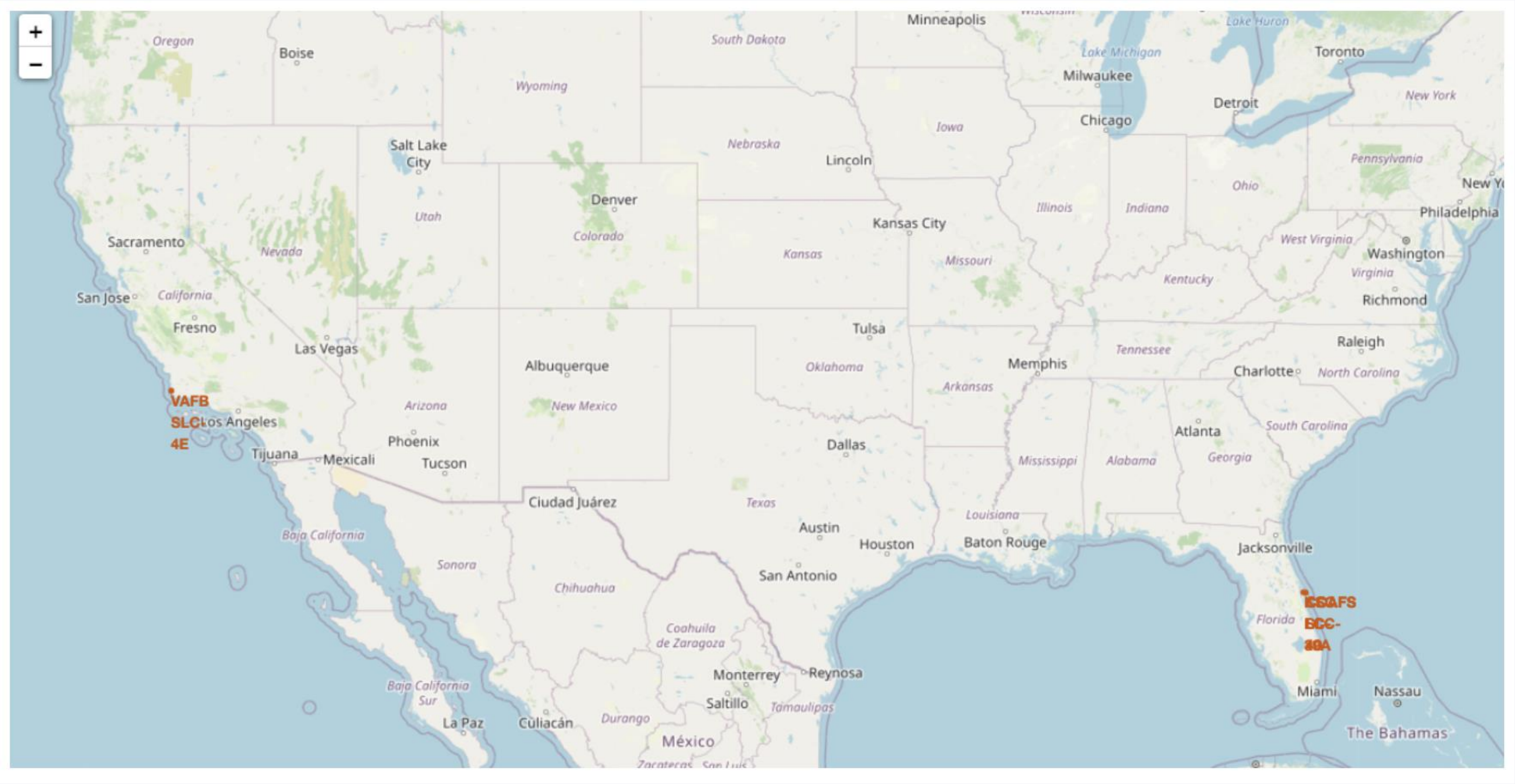
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

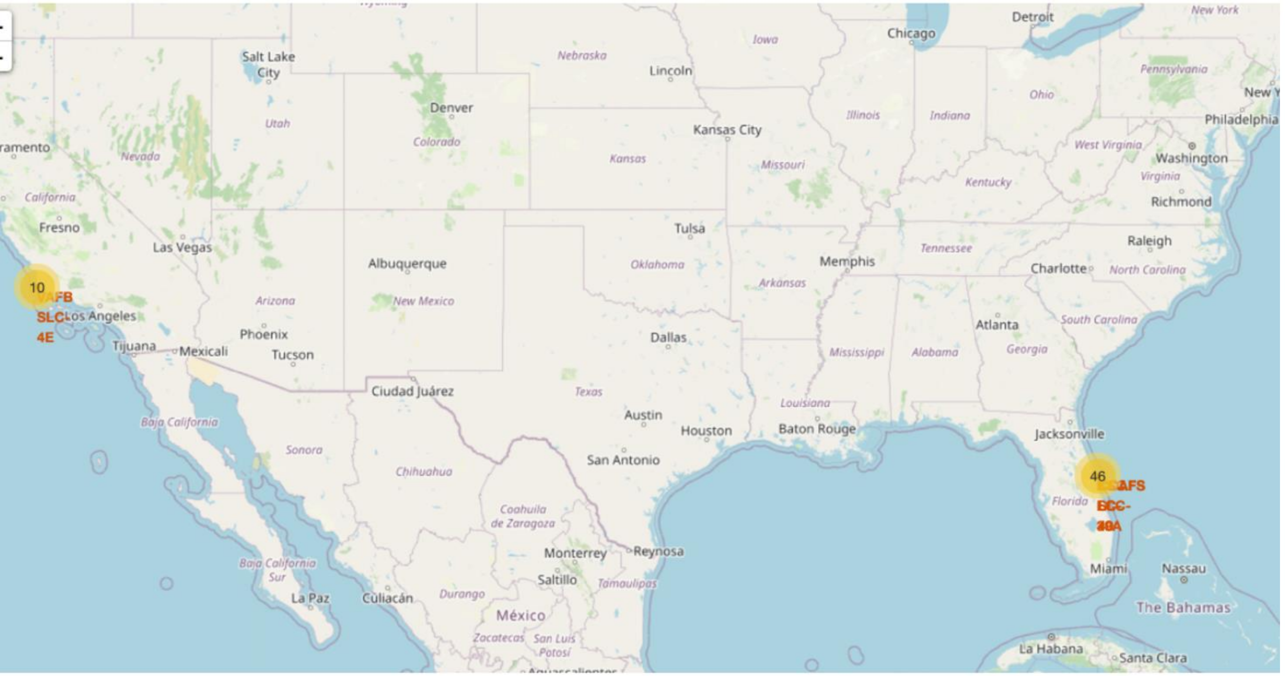
Locations of Launch Sites

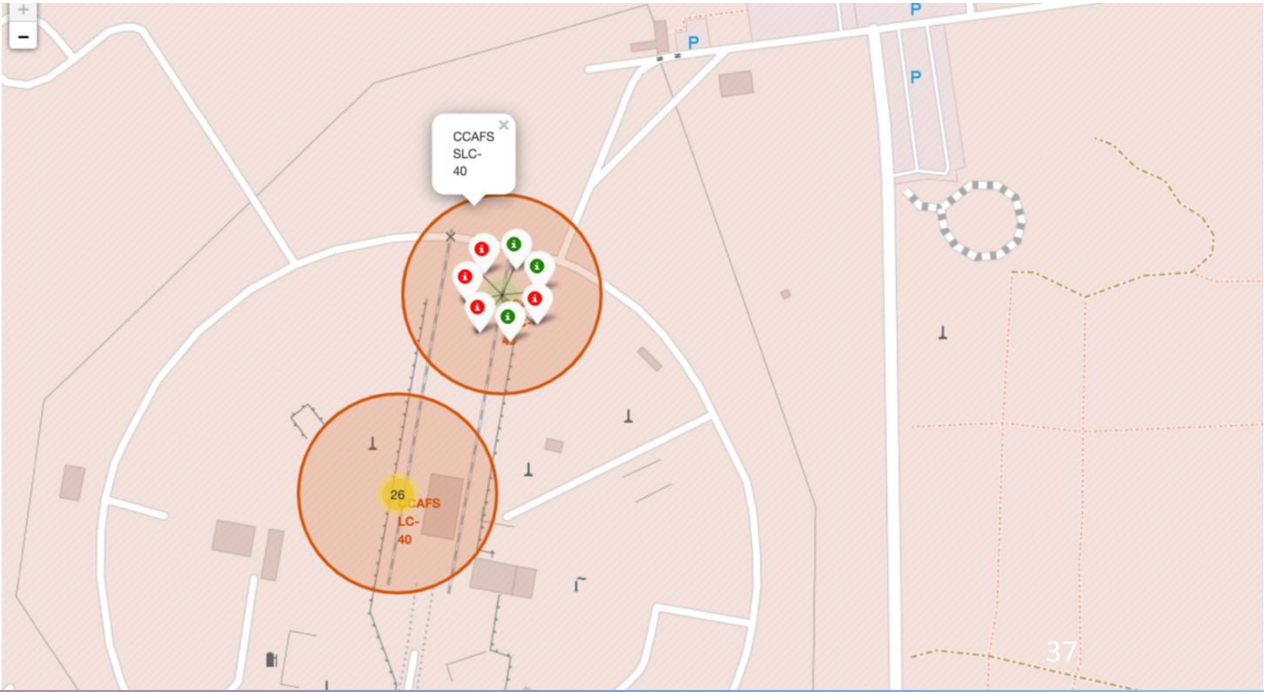
- Launch sites are placed near the coastal areas to minimize casualties in the event of failure.
- Launch sites are also placed nearest to the equator to due to the rotation of the Earth's axis.



Launch Outcomes Mapped

- The map on the left shows the number of launches at each launch site
- The map on the right drills down the launches in a color-coded sequence with the green dots representing success and the red dots meaning failure.





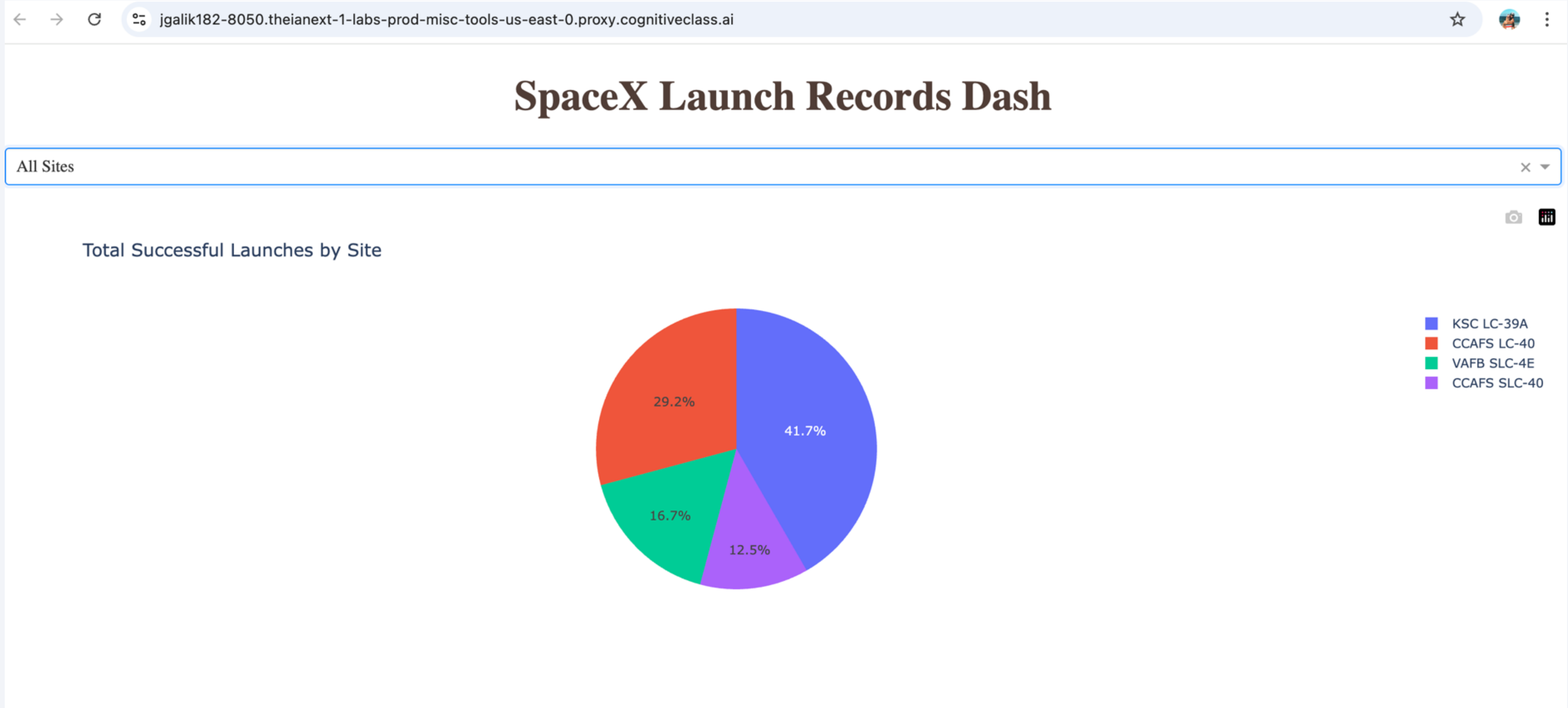
Proximate Locations in Relation to Launch Site

- The map shows some notable locations in relation to the launch site:
  - Railroad 1.02KM
  - Roadway 0.51KM

Section 4

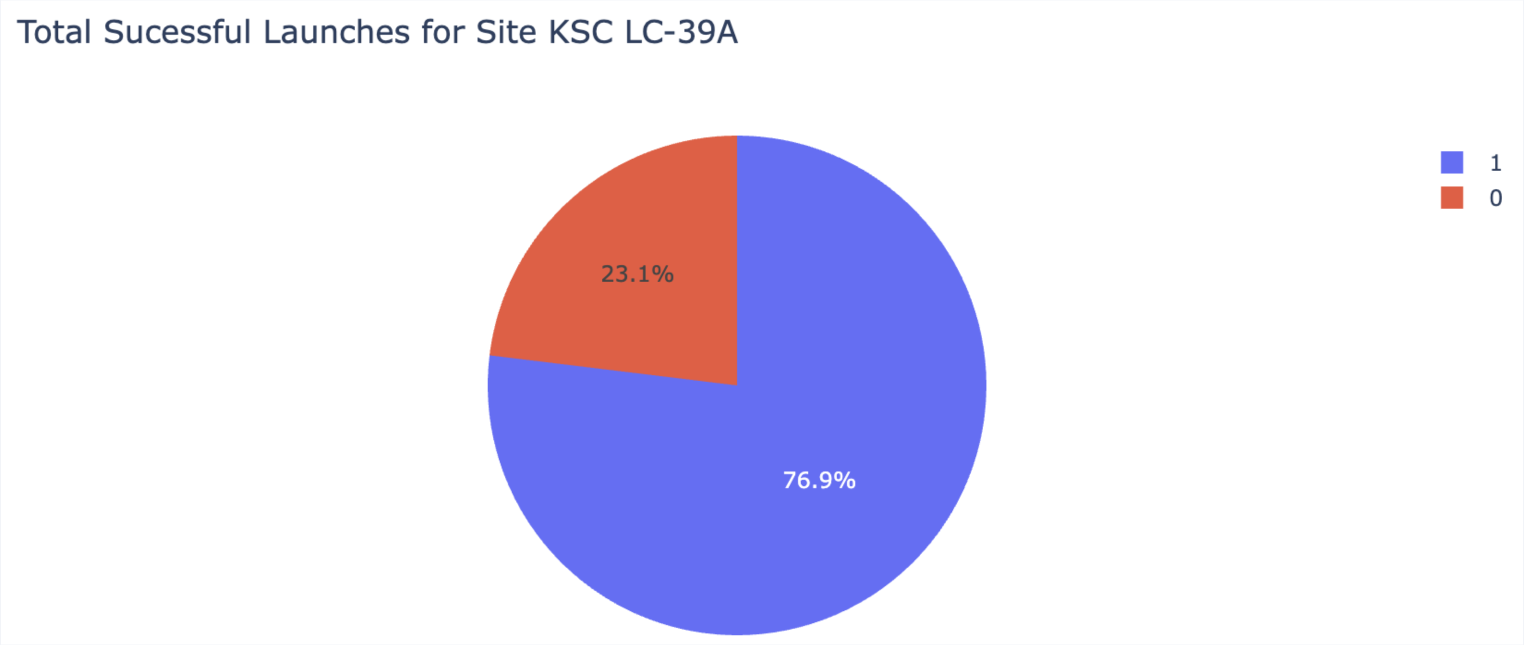
Build a Dashboard with Plotly Dash

Successful Landings by Launch Site

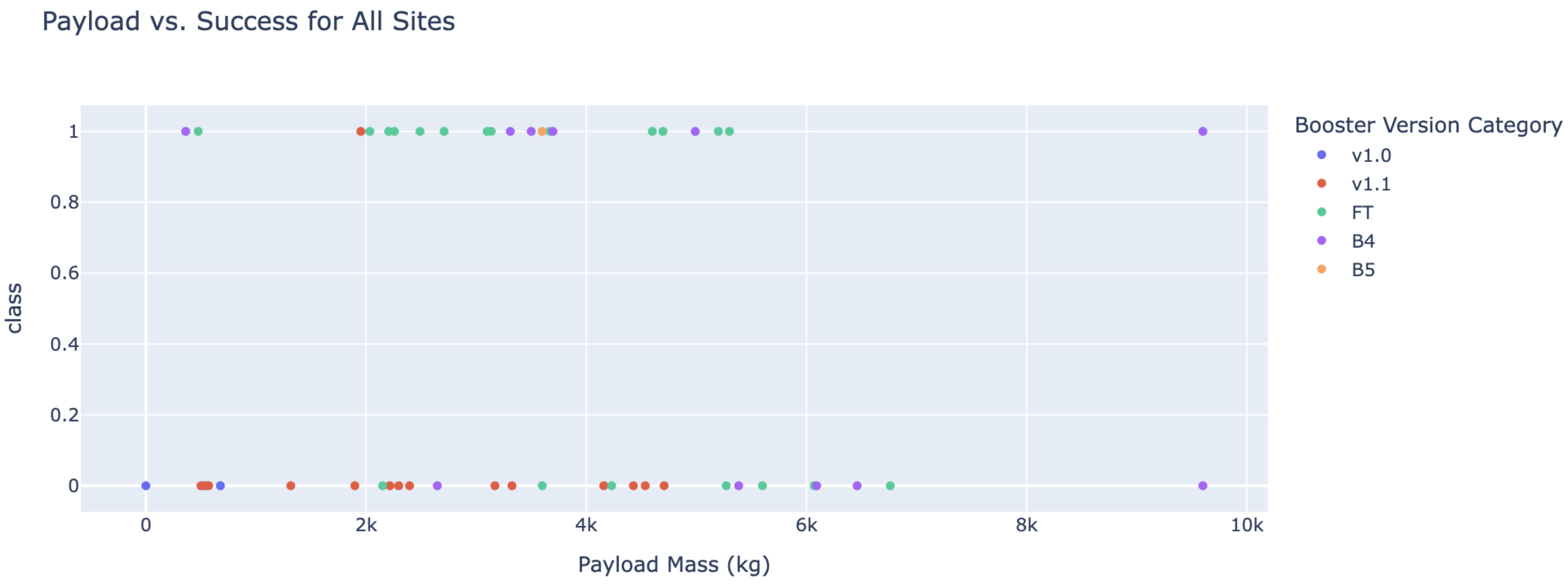


- KSC LC-39A made up for the most successful launch site, 41.7%, followed by CCAFS LC-40, VAFB SLC-4E, & CCAFS SLC-40.

Pie Chart with highest Launch Site Success



- This chart shows that launch site KSC LC-39A had the highest success rate ratio.
- 76.9% (represented as Class 1) of launches were successful here.



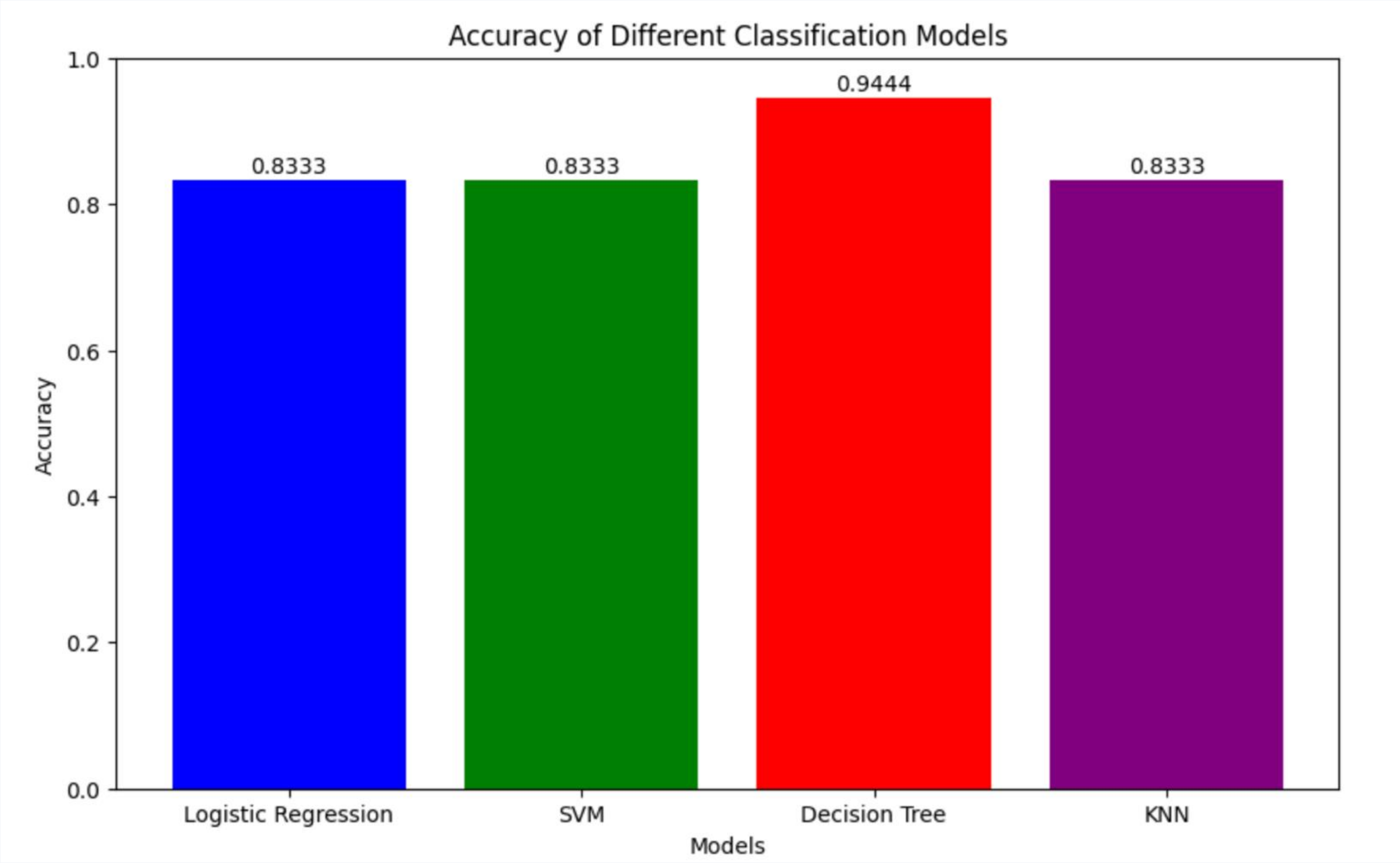
- The green dot representing booster version FT shows to have the most success across middle tier payload masses.
- Booster version v1.1 only had one success and many failures.
- Booster version B5 only had 1 launch, and it was a success, further analysis on this version should be studied to see if it can lead to more successes.

Section 5

Predictive Analysis (Classification)

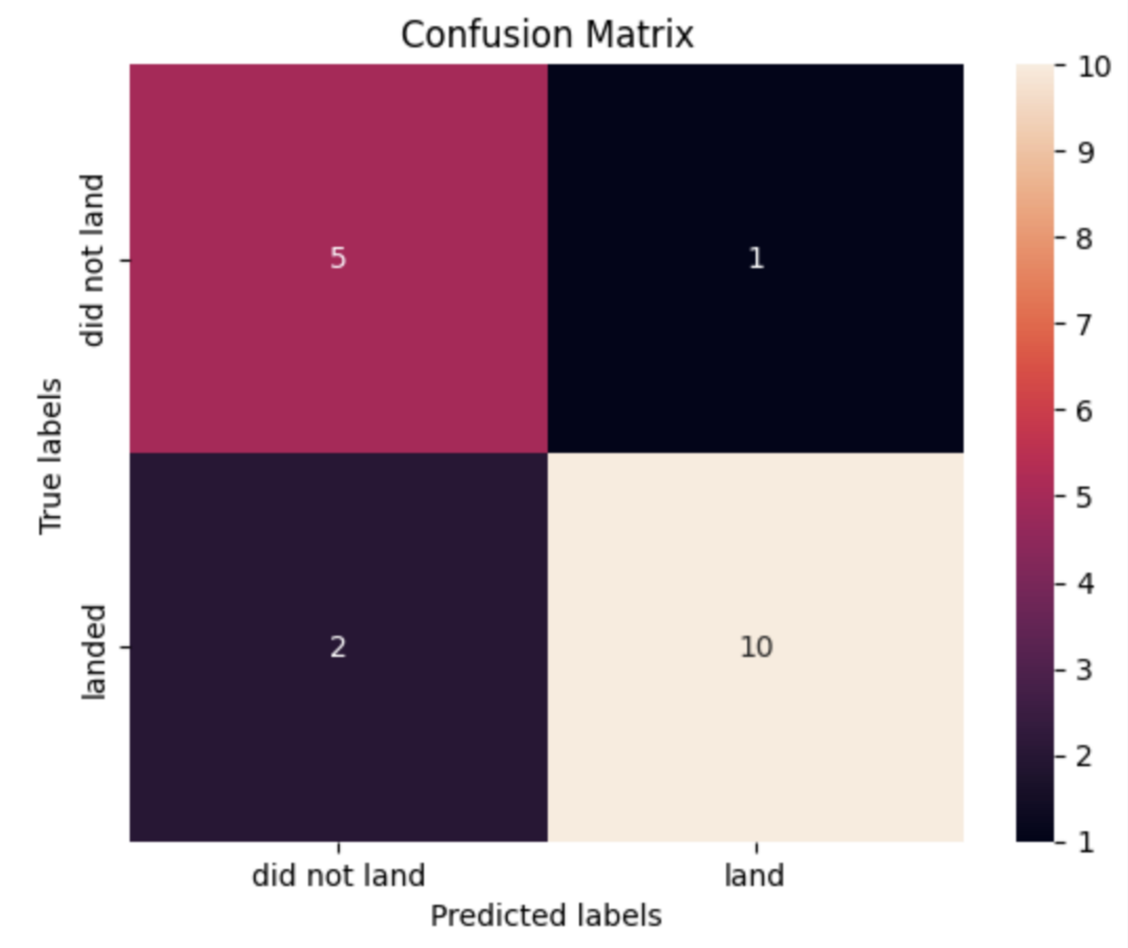
Classification Accuracy

- Of the classification models tested, the Decision Tree models showed it achieved the highest accuracy with 0.94.
- Although Logistic Regression, SVM, and KNN showed good scores of 0.83, the Decision Tree was more accurate.



Confusion Matrix

- The confusion matrix for the decision tree classifier shows 10 cases were a true positive.
- 1 false positive was predicted.
- There were 5 true negatives that predicted failed landings.
- There were 2 instances that resulted in a false negative.



Conclusions

- Over time, success rates have increased, this shows us that the Falcon 9 team is learning from their launch failures and improving on what went wrong.
- Even though there are some higher payload launches, most of them end up successful.
- Factors like orbit and the location of the launch site played a key role in the success of the mission with ES-L1, GEO, HEO, & SSO orbits and the KSC LC39A launch site being key pieces in a successful launch.
- Of the predictive models used, the one with the most accuracy was the Decision Tree Classifier displaying a high yield of successful outcomes for launch missions for the Falcon 9 team.

Appendix

- All supporting charts and code can be found here: https://github.com/jgalik182/tesrepo/tree/main

Thank you!