IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Max Henderson
12/7/2025

Outline

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

Executive Summary

- Project goal: Determine probability of Falcon 9 first-stage landing success to estimate launch cost savings.
- Why it matters: First-stage reuse drives ~$100M savings vs. competitors.
- Key findings:
    - Overall historical success rate: ~66%
    - Payload mass, launch site, booster version, and orbit strongly influence landing success.
- Deliverables: Predictive model + interactive Dash dashboard + analytical insights.

Introduction

- SpaceX flight performance is a key benchmark for commercial space reliability and risk assessment.
- Mission success is critical for reliability, cost efficiency, and customer confidence—especially as reusable boosters continue to evolve.
- The purpose of this research is to understand how various factors, such as launch site, payload mass, and booster variant relate to mission success.
- Outcomes of this study can inform:
  - **Reliability forecasting** for future missions
  - **Risk assessment** for payload customers
  - **Operational planning** regarding boosters, sites, and payload limits

Section 1

Methodology

Methodology

Executive Summary

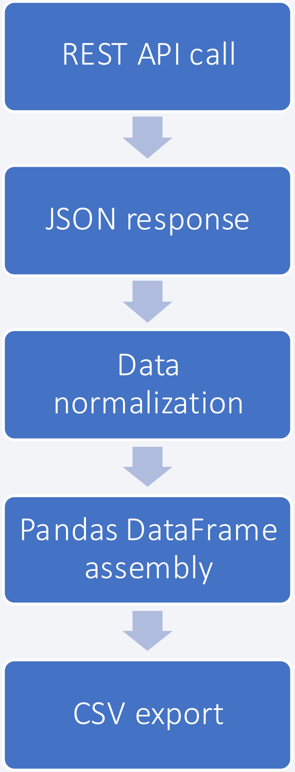
- Data collection methodology:
  - Via web scraping and SpaceX API
- Perform data wrangling
  - Clean inconsistent naming, extract outcome as binary, and merge API and web scraped data into a unified dataset
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build, tune, evaluate classification models

Data Collection

- Data was collected via two methods:
    1. Web scraping data from a table on a webpage
    2. Calling on data via the SpaceX API (Application Programing Interface)
- GitHub:
    1. SpaceX API Lab
    2. Scraping Lab

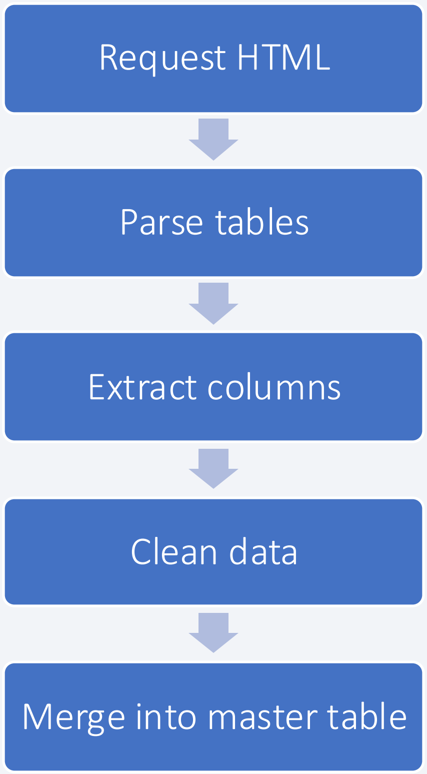
Data Collection – SpaceX API

- Queried SpaceX API endpoints for launch data
- Parsed JSON payloads for mission metadata
- Extracted booster version, block, reuse count, etc.
- Cleaned nested structures into tabular format
- GitHub: SpaceX API Lab



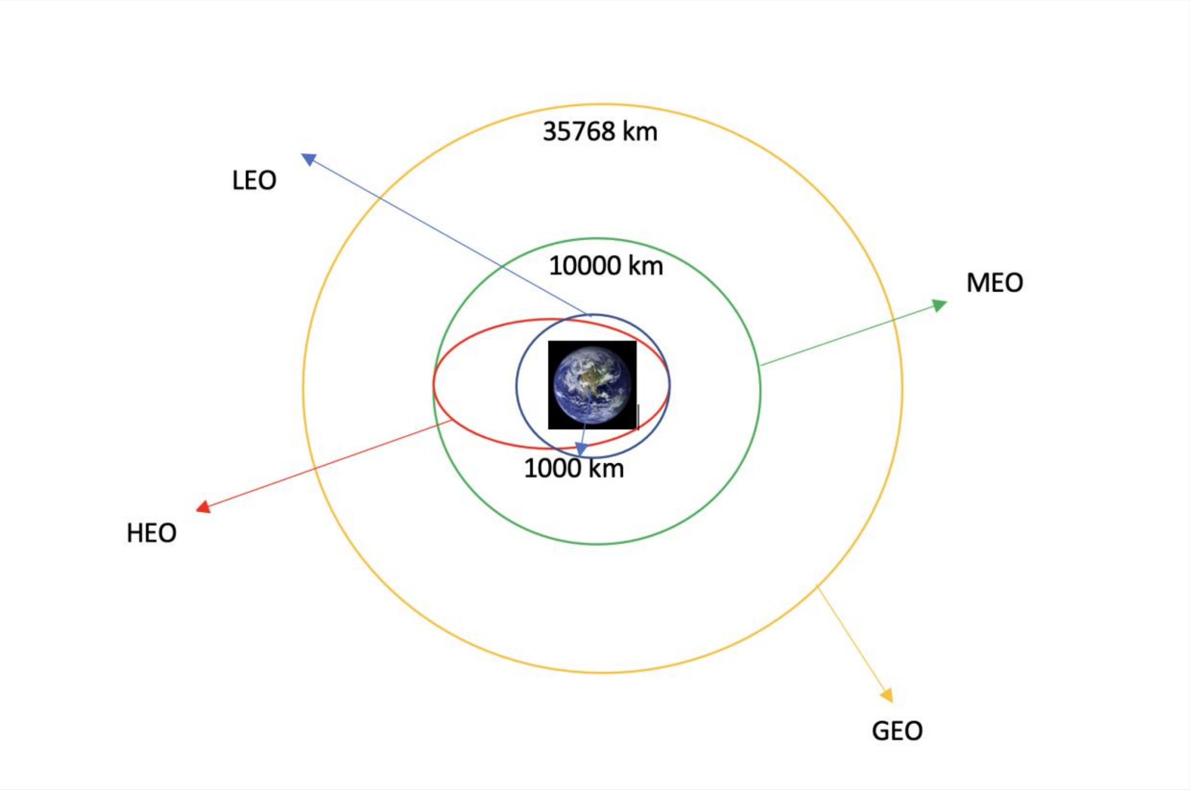
Data Collection - Scraping

- Used BeautifulSoup to scrape Wikipedia Falcon 9 booster tables
- Extracted: booster version, flight number, reuse status, etc.
- Joined Scraped data with API data
- GitHub: Web Scraping Lab



Data Wrangling

- Main Tasks:
  - Standardized column names and data types
  - Identified columns containing missing values
  - Categorized various outcomes as either "good" or "bad"
  - Created binary **Class** column for landing success results based on categorized outcomes
- GitHub: Data Wrangling Lab



EDA with Data Visualization

- Utilized a variety of plotting methods to visualize trends in the data
- Scatter plots:
  - Flight Number vs. Payload Mass (color-coded by launch outcome)
  - Flight Number vs. Launch Site (color-coded by launch outcome)
  - Payload Mass vs. Launch Site (color-coded by launch outcome)
  - Flight Number vs. Orbit (color-coded by launch outcome)
  - Payload Mass vs. Orbit (color-coded by launch outcome)
- Bar plot: Orbit vs. Success Rate (launch outcome)
- Line Plot: Year vs. Success Rate (launch outcome)

GitHub: EDA Data Visualization

EDA with SQL

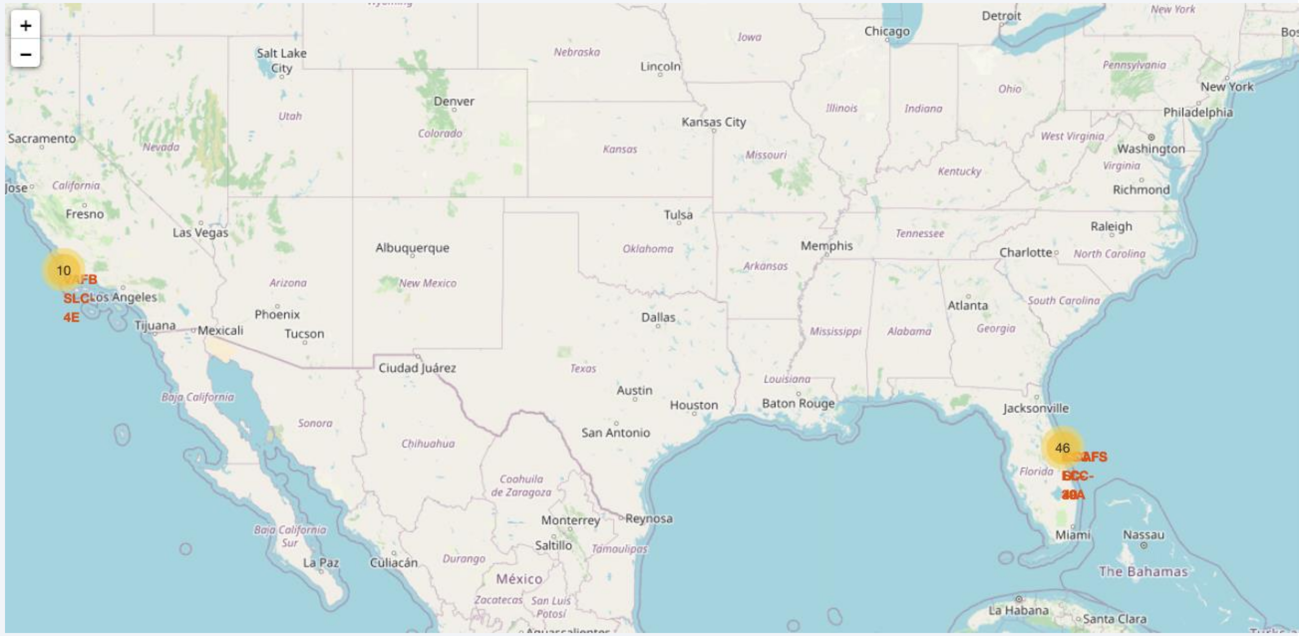
- Queries Performed
    - Unique launch sites
    - Launch sites beginning with "CCA"
    - Total NASA booster payload mass
    - Average mass for booster version "F9 v1.1"
    - First successful ground landing date
    - Boosters with drone-ship success + 4000–6000 kg payload
    - Mission outcome counts
    - Boosters carrying maximum payload
    - 2015 failed drone-ship landings with booster + launch site
    - Ranked landing outcomes (2010–2017)

GitHub: EDA with SQL

Build an Interactive Map with Folium

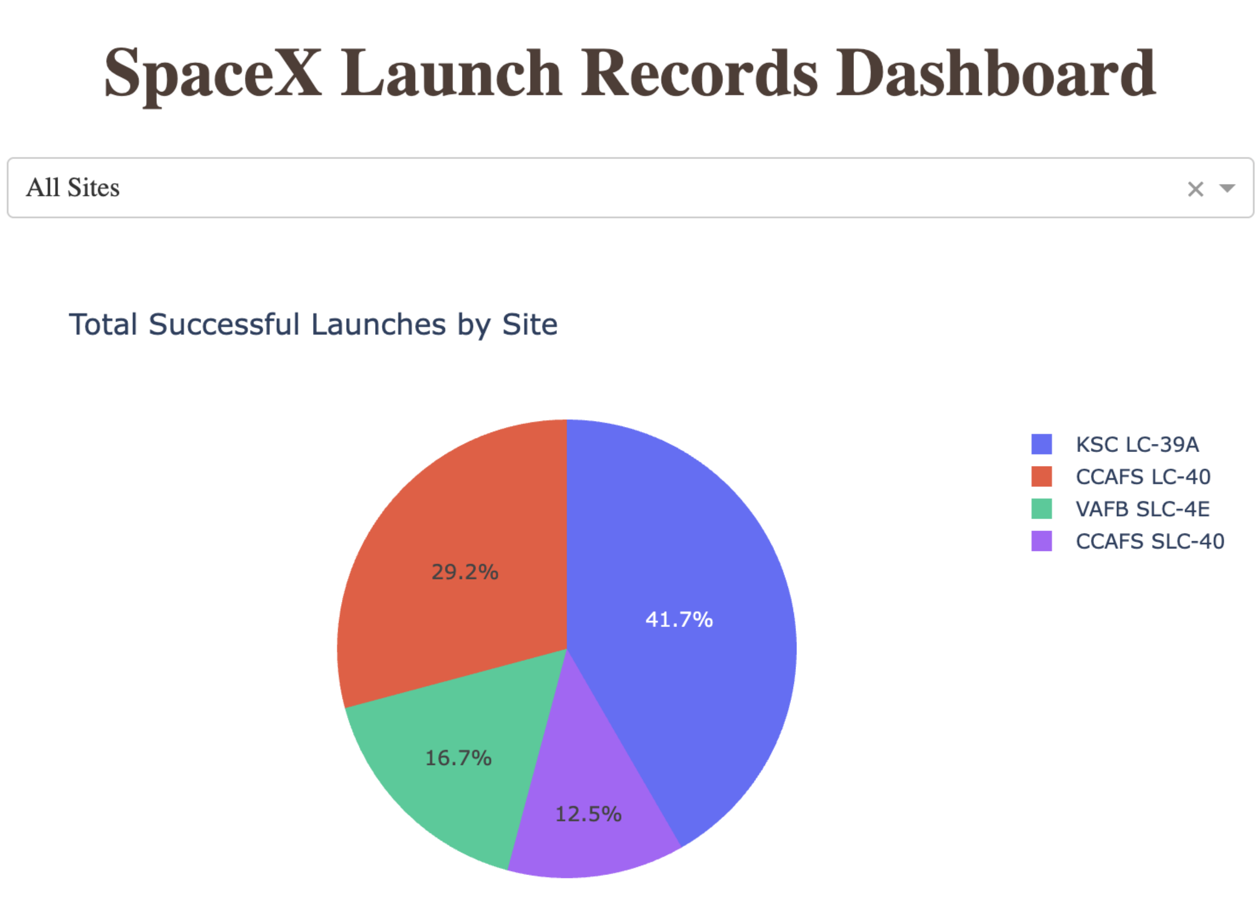
- Objects Used
  - Launch site markers with popups
  - Circle markers showing payload mass ranges
  - Color-coded success/failure layers
  - Distance line from site to nearest coastline or highway
- Insights
  - Coastal proximity is a clear pattern in launch site selection
  - Sites differ in surrounding infrastructure (roads, rail lines)

GitHub: Folium Lab



Build a Dashboard with Plotly Dash

- Launch Site dropdown
  - Success Pie Chart
  - Payload range slider
  - Payload vs. Launch Outcome scatter
- Key Interactions
  - Selecting specific launch sites reveals success vs. failure ratios
  - Adjusting payload slider highlights payload ranges with highest success
  - KCS LC-39A launch site has highest ratio of successful launches
- GitHub: Plotly Dash Lab



Predictive Analysis (Classification)

- Models Built
  - Logistic Regression
  - SVM (Support Vector Machine)
  - KNN (K-Nearest Neighbors)
  - Decision Tree
- Final Test Accuracy
  - Logistic Regression: ~83%
  - SVM: ~83%
  - KNN: ~83%
  - Decision Tree: ~94% ← Best model
- Confusion Matrix Interpretation
  - High true-positive and true-negative rates
  - Primary errors occur in borderline payload/booster combinations
  - Indicates model is reliable for predicting landing success
- GitHub: Predictive Analysis Lab

Results

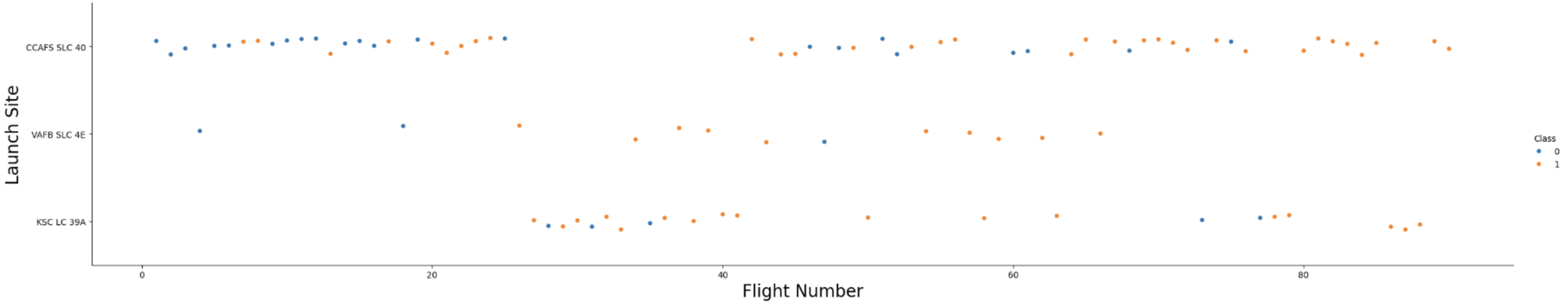
- Landing success has improved significantly over time, with post-2015 missions showing a sharp increase in first-stage recovery success.
- Launch site impacts success rates, with KSC LC-39A and CCAFS SLC-40 showing higher proportions of successful landings than VAFB SLC-4E.
- Payload mass has a non-linear relationship with success:
  - Moderate payload masses show higher landing success
  - Extremely heavy payloads increase landing difficulty and failure risk
- Booster reuse is correlated with higher success, indicating operational learning and engineering maturity.
- Geospatial analysis shows launch sites are strategically located near coastlines, enabling safe downrange landings and recovery logistics.
- Interactive Dash visualizations reveal:
  - Clear payload ranges associated with higher success
  - Significant differences in success by launch site and booster version

Section 2

Insights drawn from EDA

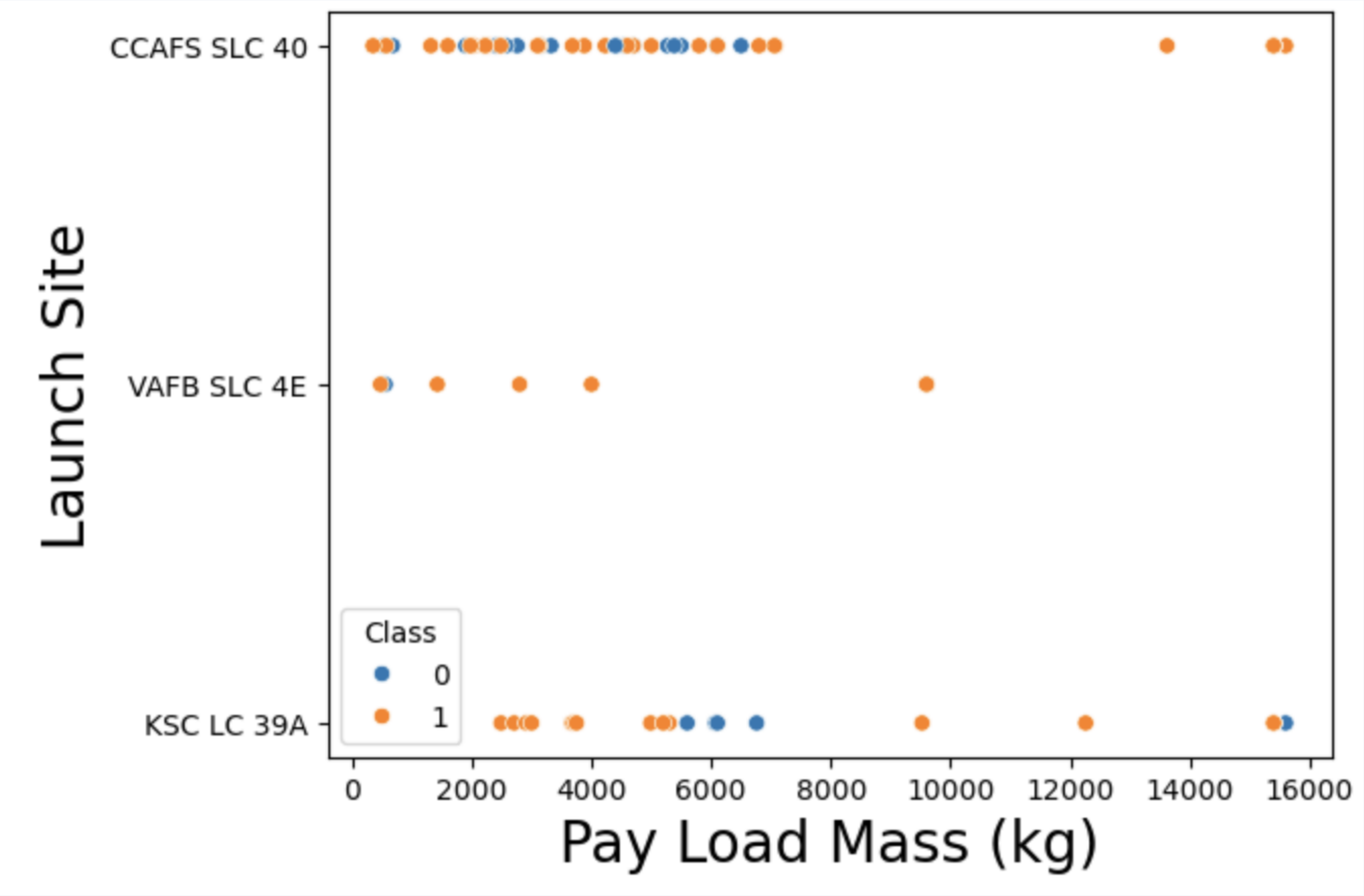
Flight Number vs. Launch Site

- Early Falcon 9 launches (lower flight numbers) show a higher concentration of **unsuccessful landings**, reflecting SpaceX's initial learning and testing phase.
- As flight number increases, the proportion of **successful landings increases**, indicating clear performance improvement over time.
- Certain launch sites (e.g., **KSC LC-39A and CCAFS SLC-40**) exhibit a higher frequency of successful landings in later missions.
- This trend suggests that **operational experience and launch site maturity** are strong contributors to first-stage landing success.



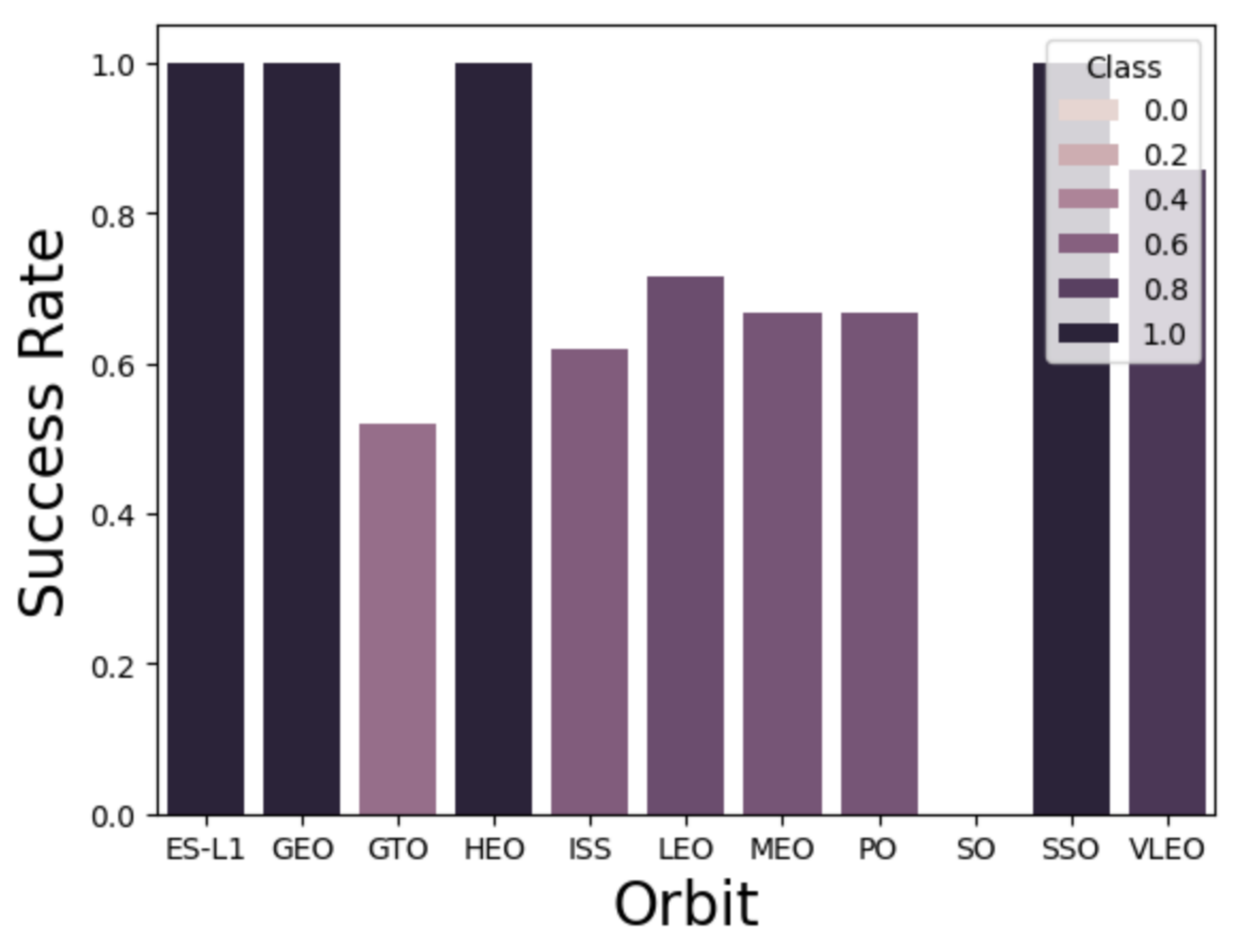
Payload vs. Launch Site

- Certain launch sites (notably KSC LC-39A and CCAFS SLC-40) support successful landings across a wider range of payload masses.
- Launch site VAFB SLC 4E appears to have a smaller maximum payload mass capacity



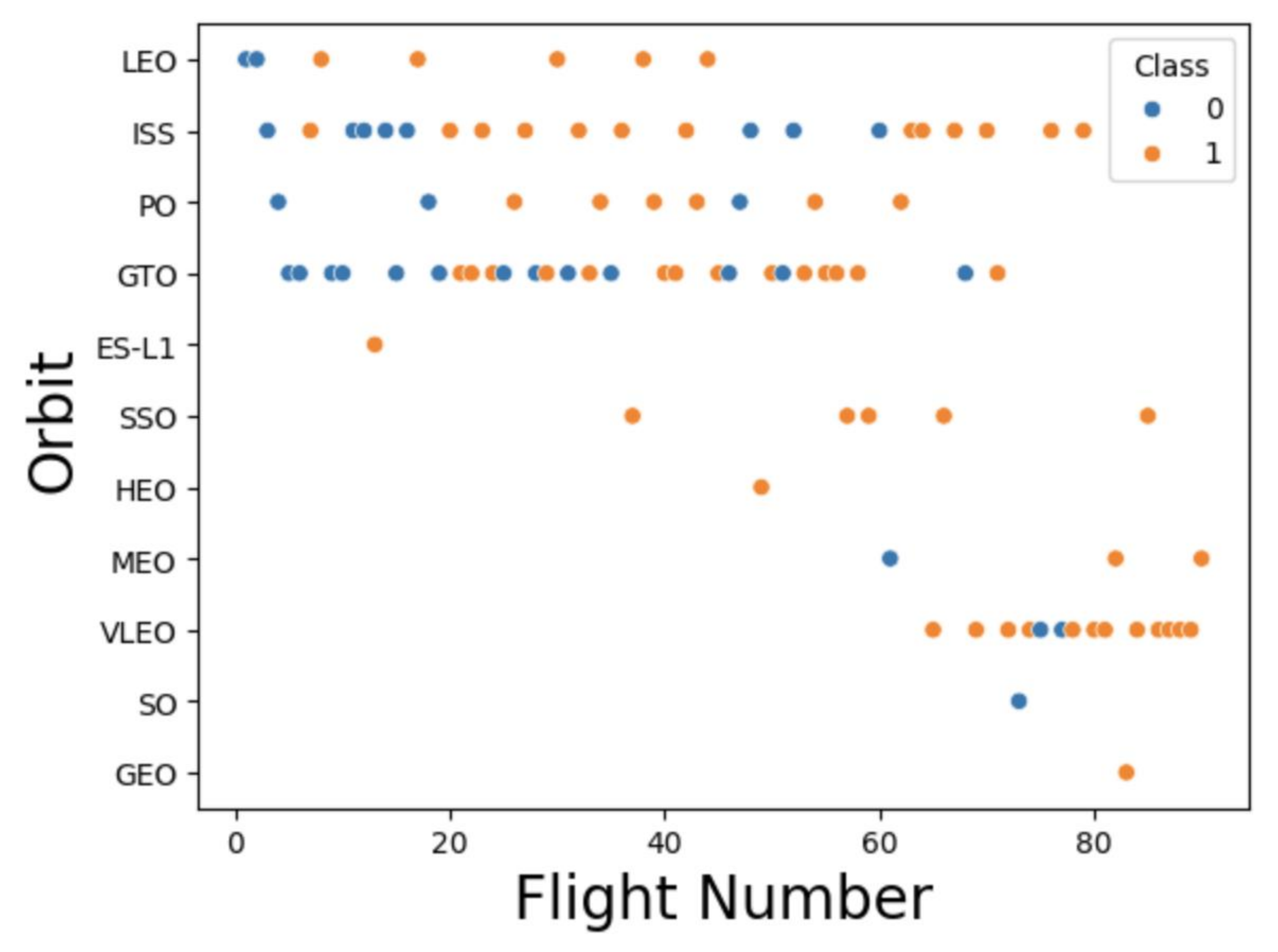
Success Rate vs. Orbit Type

- Several orbit types appear to have only successful launch results (ES-L1, GEO, HEO, SSO)
- The orbit type with the lowest success rate is SO, followed by GTO
- This suggests that orbit type may impact success rate, as certain orbit types have significantly higher historical success rates than others



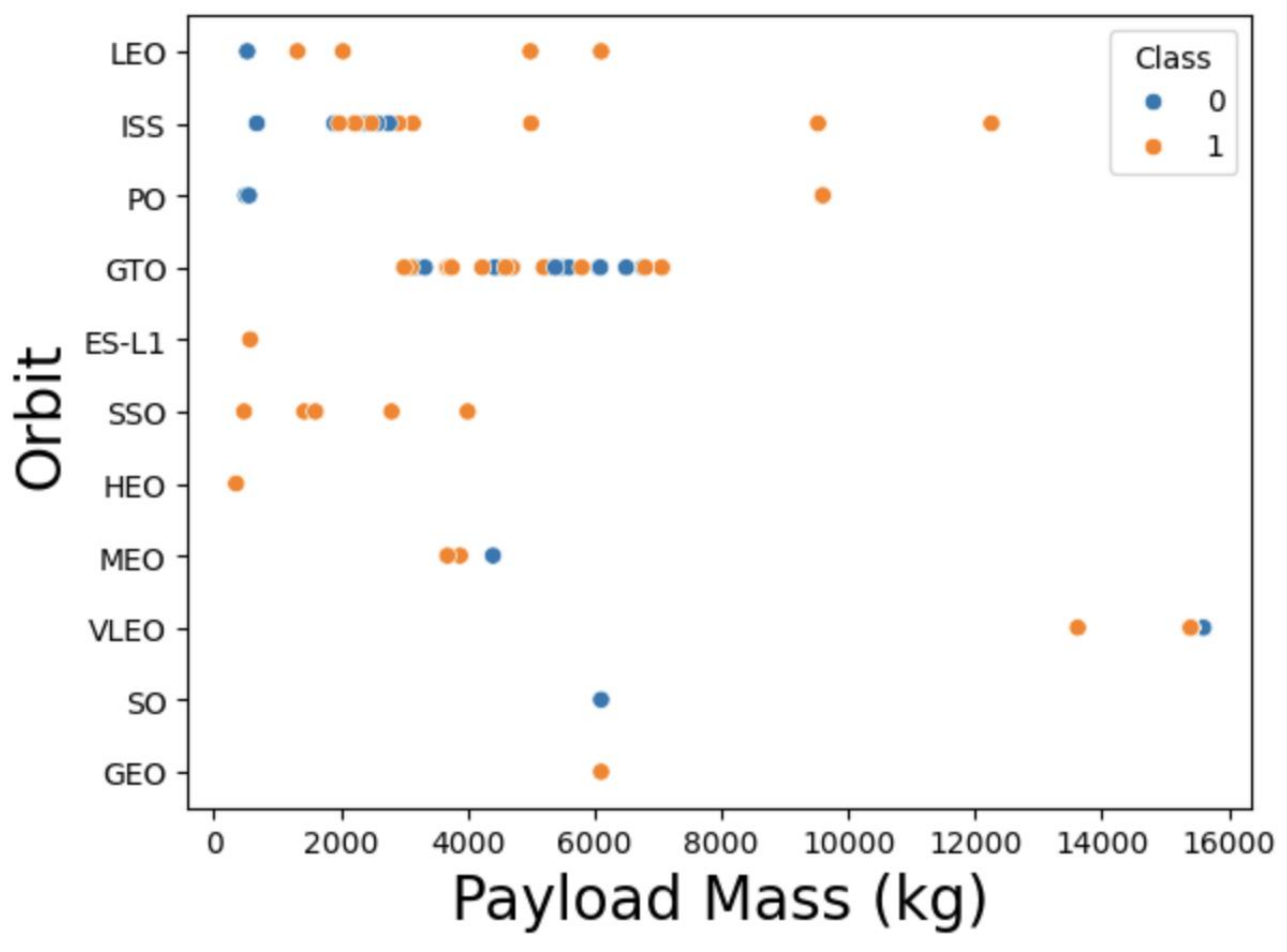
Flight Number vs. Orbit Type

- This graph appears to be more insightful than the bar chart shown in the previous slide
- While SO was shown to be the lowest success rate in the previous slide, you can now see that there was only one attempted launch at this orbit
- Certain orbit types, LEO and ISS for example, seem to become more consistently successful as flight number increases



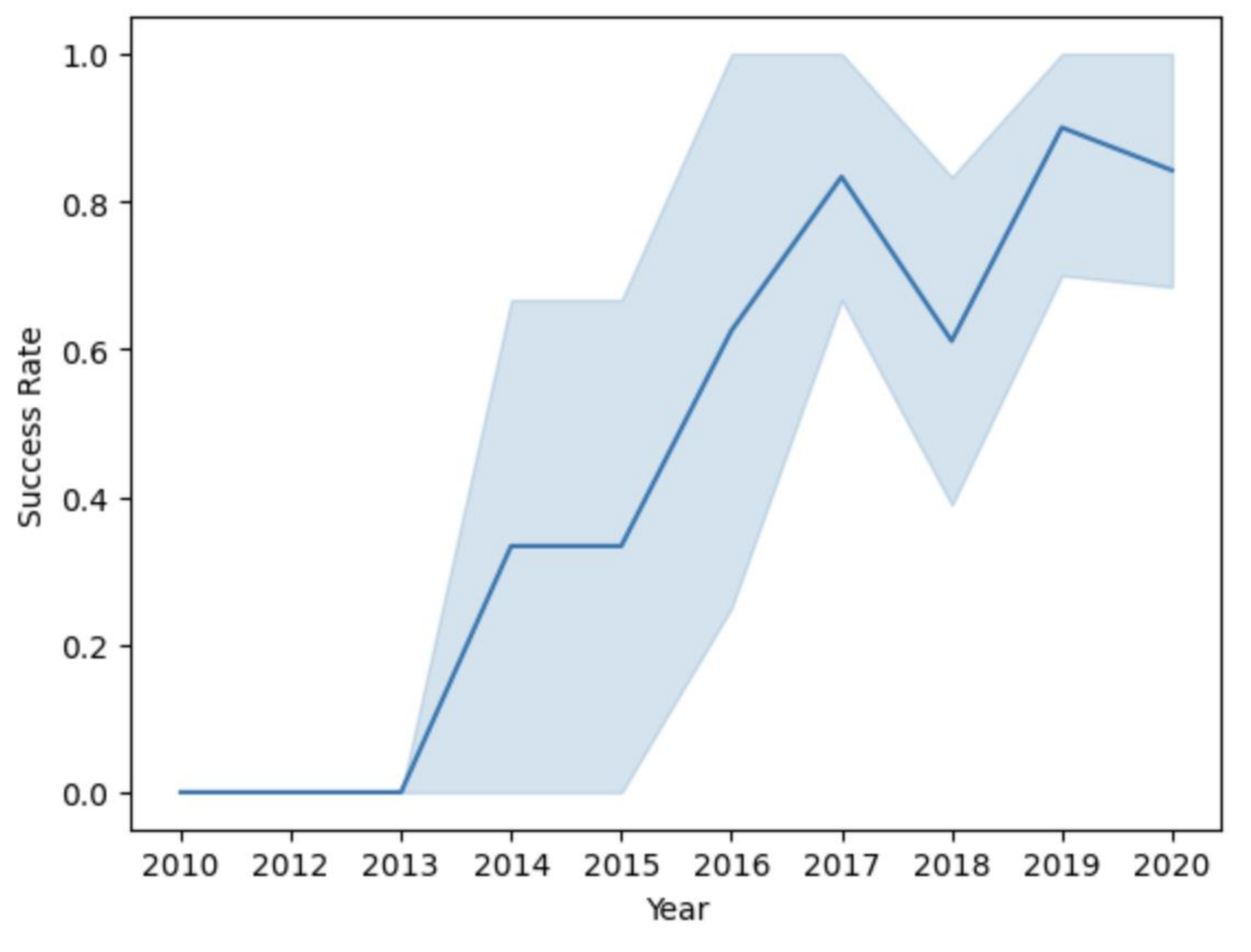
Payload vs. Orbit Type

- LEO and ISS orbits show a strong concentration of successful landings (orange), especially at lower to mid-range payload masses, with relatively few failures.
- GTO missions exhibit a mix of successes and failures, with blue points appearing more frequently as payload mass increases, indicating higher landing risk for heavier GTO payloads.
- Some orbit types (e.g., HEO, MEO, GEO) have very limited data points, making definitive conclusions less reliable for those categories.



Launch Success Yearly Trend

- A clear upward trend begins around 2014, coinciding with the introduction and refinement of landing technologies such as grid fins and drone ship landings.
- Success rates increase sharply between 2015 and 2017, indicating rapid operational learning and improved booster recovery reliability.
- By 2019–2020, success rates consistently exceed ~80–90%, demonstrating a mature and reliable first-stage landing process.
- First-stage landing success improves steadily over time, reflecting SpaceX's rapid learning curve and maturation of reusable launch technology.



All Launch Site Names

- Used SQL queries to analyze the data
- As shown in the query results displayed in this screenshot, there are four unique launch sites:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

Task 1

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

In [10]:

```
%sql select distinct "Launch_Site" from SPACEXTBL
```

```
 * 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'

- Used an SQL query to find the first five launch records from a launch site beginning with "CCA"
- These launches occurred over the course of several years
- All of these launches were performed using F9 boosters

```
%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_
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

- Found the sum of the payload mass column, and labeled the result in a new column titled, "total_payload_mass"
- From this SQL query, we can see the total payload mass carried by boosters launched by NASA is 99,980 kg

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

```
%sql select SUM(PAYLOAD_MASS__KG_) as total_payload_mass from SPACEXTBL where Customer like 'NASA%'
```

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

total_payload_mass
99980

Average Payload Mass by F9 v1.1

- Found the average of the payload mass column for all F9 v1.1 boosters, and labeled the result in a new column titled,"avg_payload_mass"
- From this SQL query, we can see the average payload mass carried by F9 v1.1 boosters is ~2,534.67 kg

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 like 'F9 v1.1%'
```

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

avg_payload_mass
2534.6666666666665

First Successful Ground Landing Date

- From the query below, we can see that the first successful ground landing date was December 22nd, 2015
- We used the min(Date) query to find the first date in which the Landing_Outcome column displayed "Success (ground pad)"

```
%sql select min(Date) as first_launch from SPACEXTBL where Landing_Outcome = "Success (ground pad)"
```

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

first_launch
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- By using the query below, we can see that the booster versions with successful drone ship landings with a payload between 4000 and 6000 kg includes F9 FT B1002, B1026, B1021.1, and B1031.2

```
ACEXTBL where Landing_Outcome = "Success (drone ship)" and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000;
```

```
 * 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

- The total number of successful and failed mission outcomes can be seen displayed by the query below

```
%sql SELECT Mission_Outcome, COUNT(*) AS total_count FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

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

Boosters Carried Maximum Payload

- The results of the query below are displayed in the right-hand screenshot
- Query: `%sql SELECT DISTINCT booster_version, payload_mass__kg_ FROM SPACEXTBL WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_ ) FROM SPACEXTBL);`

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- Using the query below, we can see that there were two failed launches in January and April of 2015
- Query: `%sql SELECT substr(Date, 6, 2) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL WHERE Landing_Outcome LIKE '%Failure (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

- Using the query below, we found 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
- Query: `%sql SELECT Landing_Outcome, COUNT(*) AS outcome_count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY outcome_count DESC;`

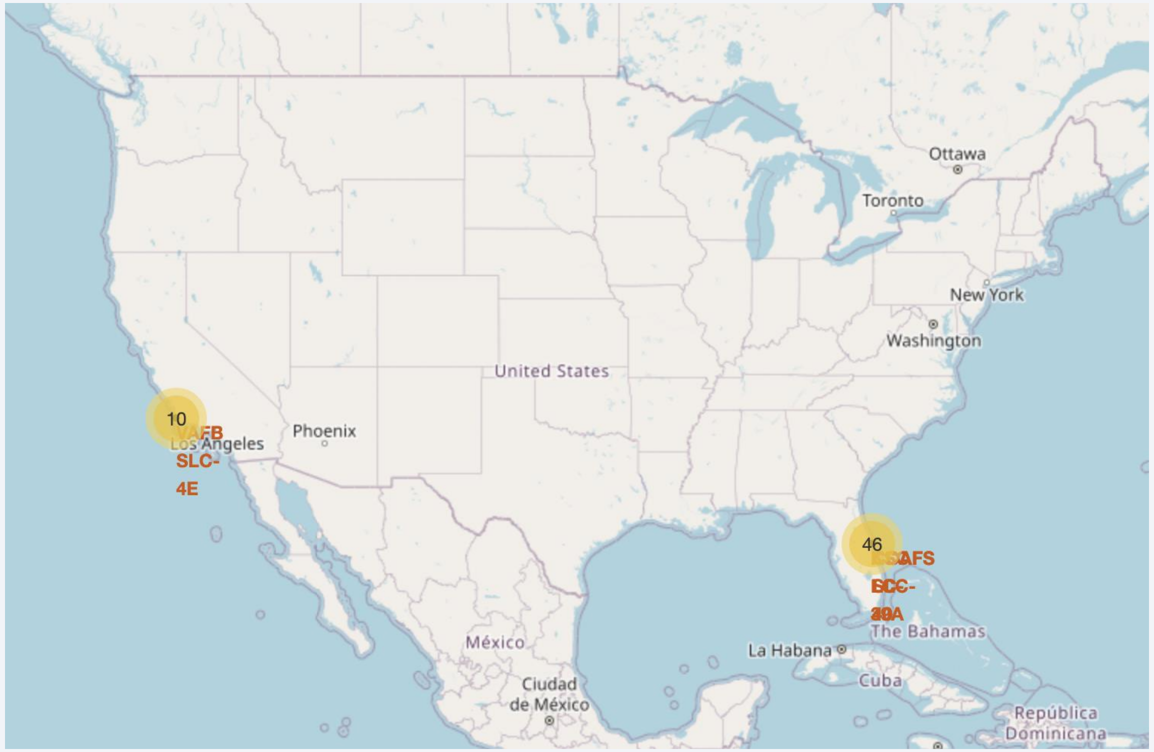
Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

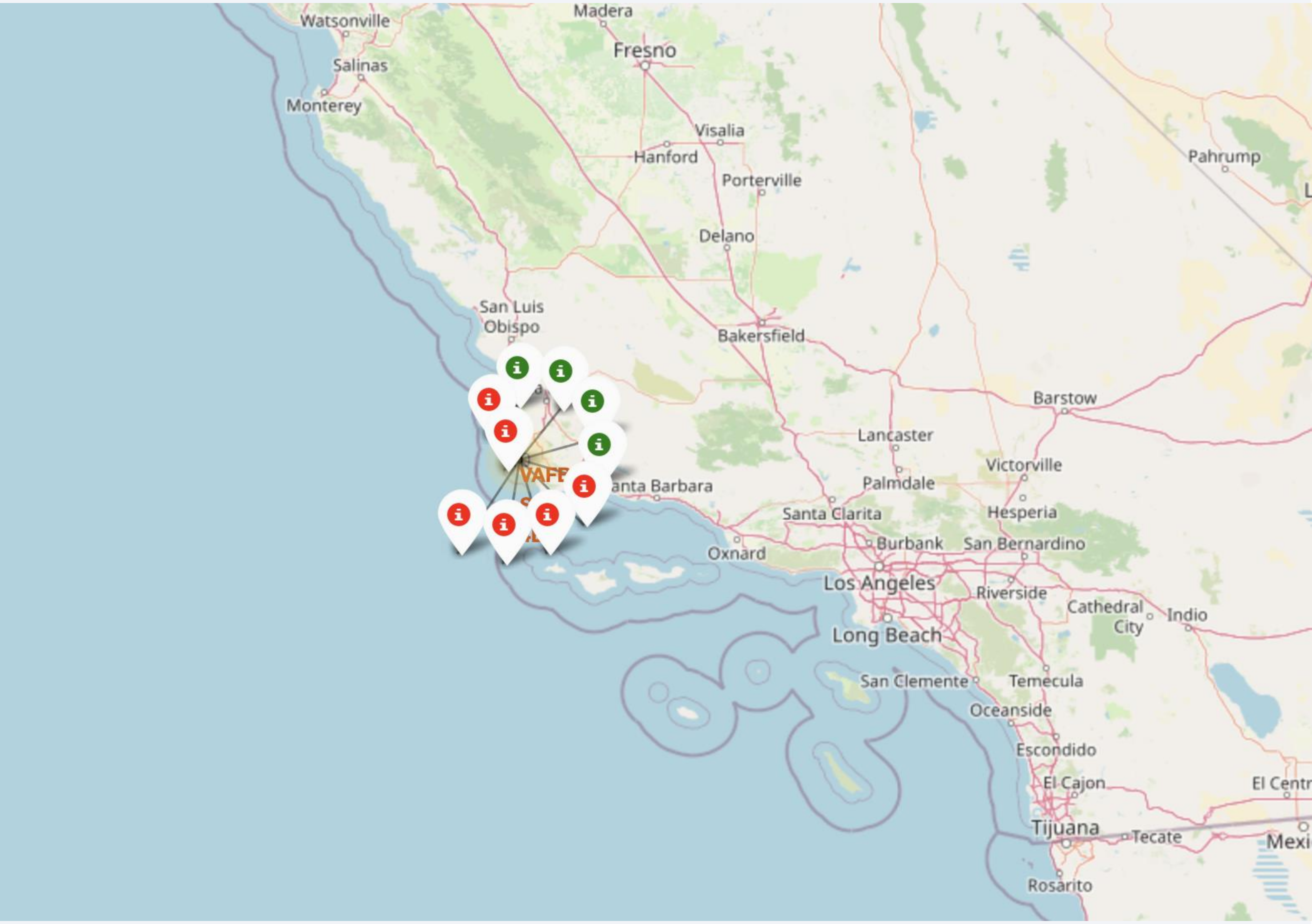
Geographic Locations of All Launch Sites

- Using Folium, we created an interactive map that allowed further analysis of geographic influences on launch outcomes
- As you can see, all of the launch sites used by SpaceX are located on a coastline
    - This likely helps to mitigate risk, as failed launches are more likely to end up in the ocean than on land



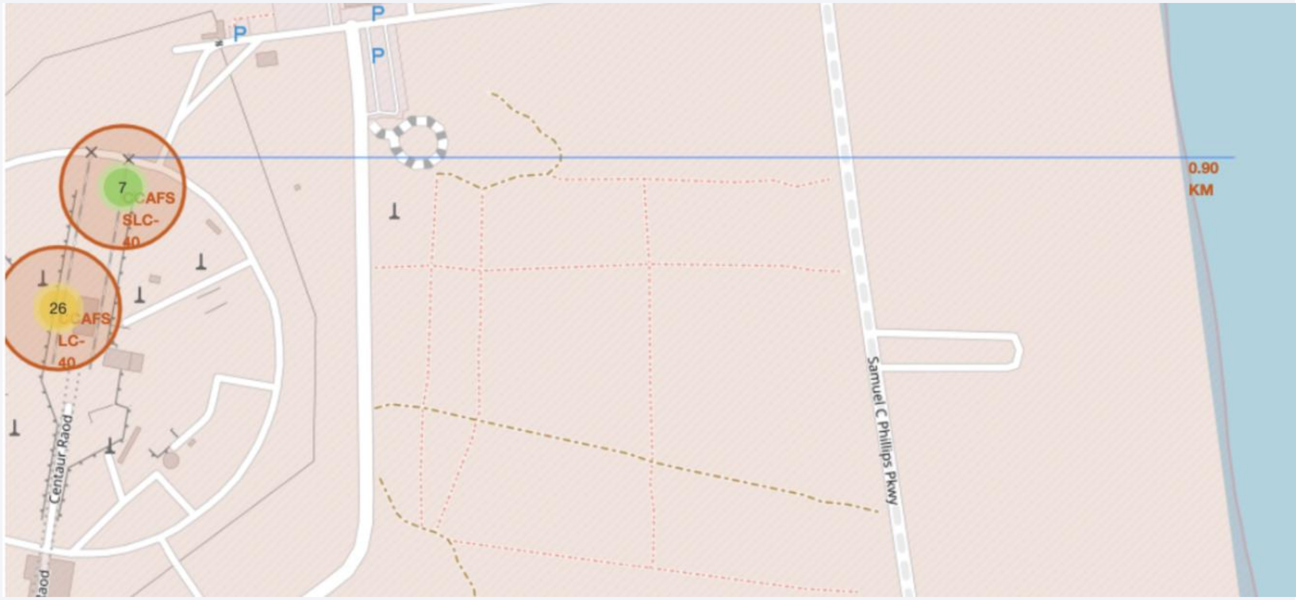
Launch Outcomes in Folium

- Within this interactive map, clicking on the yellow launch site markers allows you to easily view individual launch outcomes
- These pop-ups (red for unsuccessful launches and green for successful launches) allow us to view success rates at each launch site on a per-launch basis



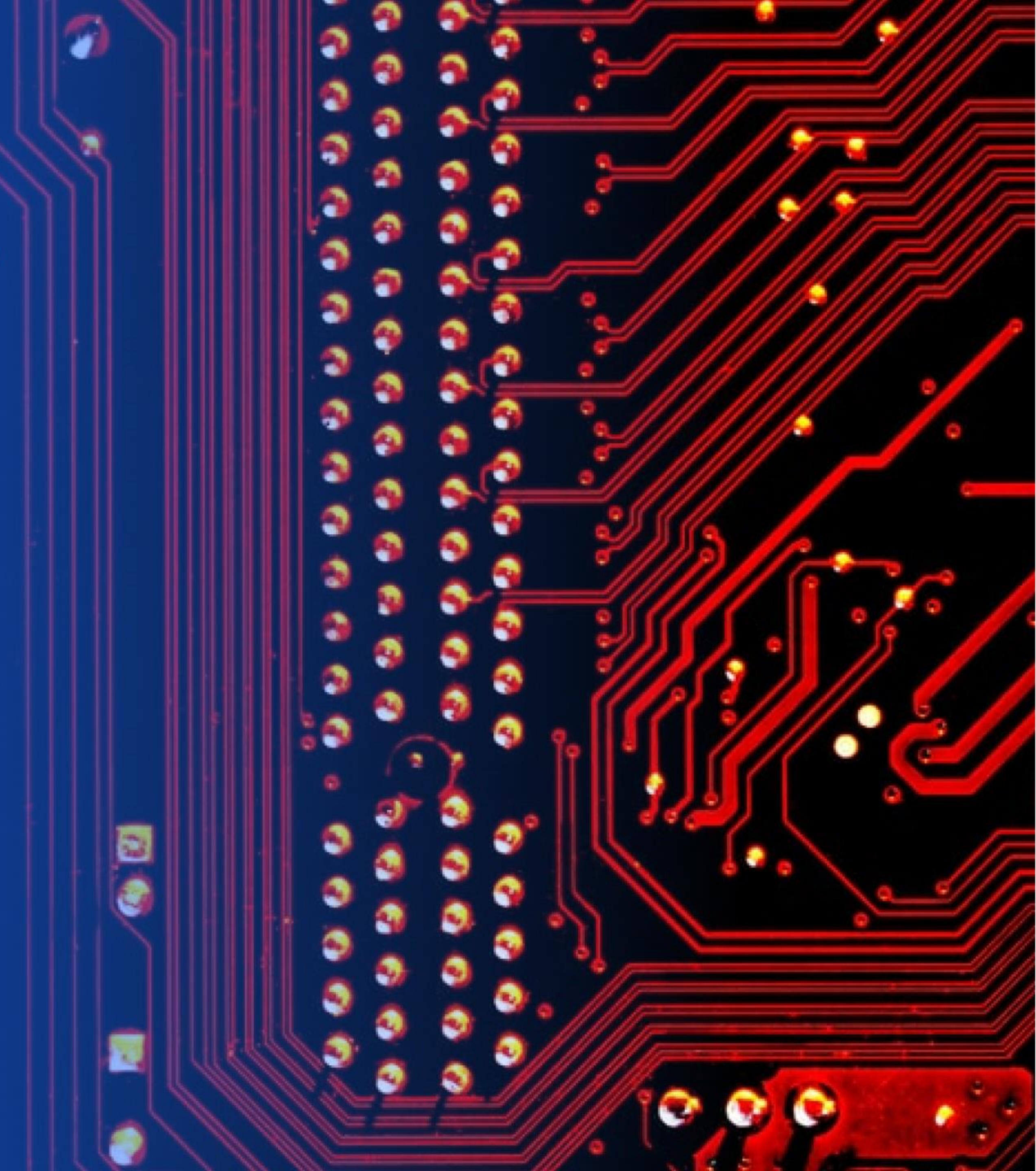
Launch Site Proximities

- Through this interactive map, we were also able to analyze the proximity of each launch site to various nearby locations
- As you can see in the right-hand screenshot, this launch site was located less than 1km from the coastline
- Additionally, launch sites were located relatively far from public roadways and infrastructure, as to avoid any potential harm to the public in the event of a failed launch



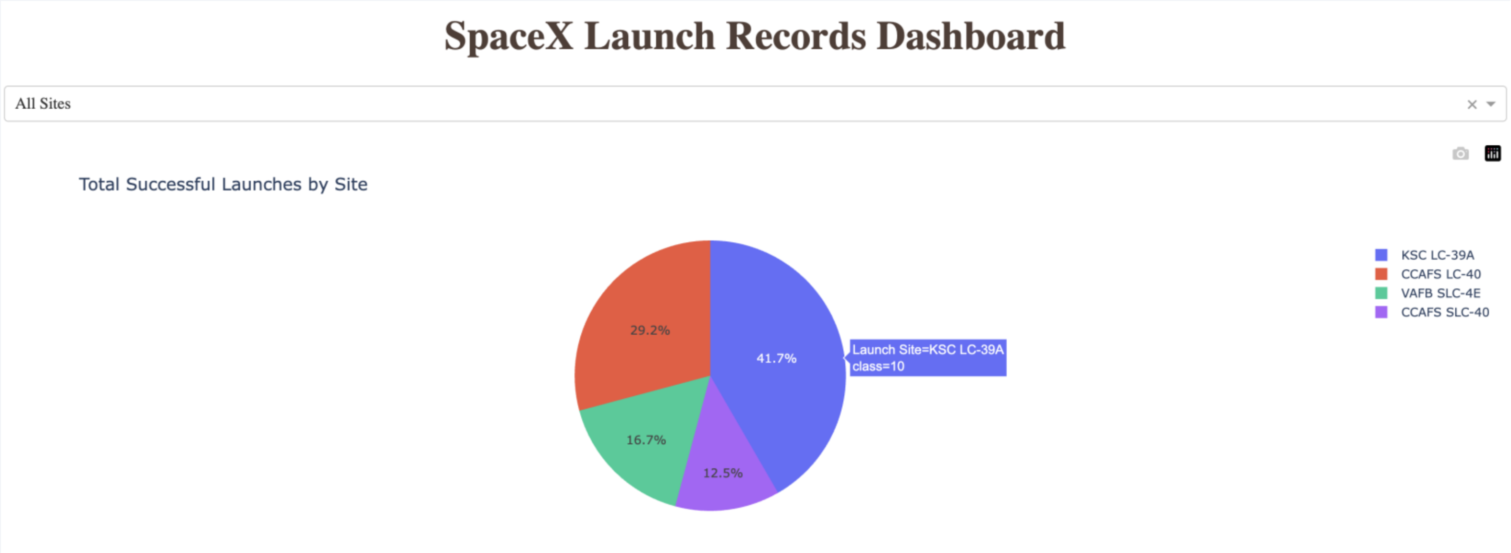
Section 4

Build a Dashboard with Plotly Dash



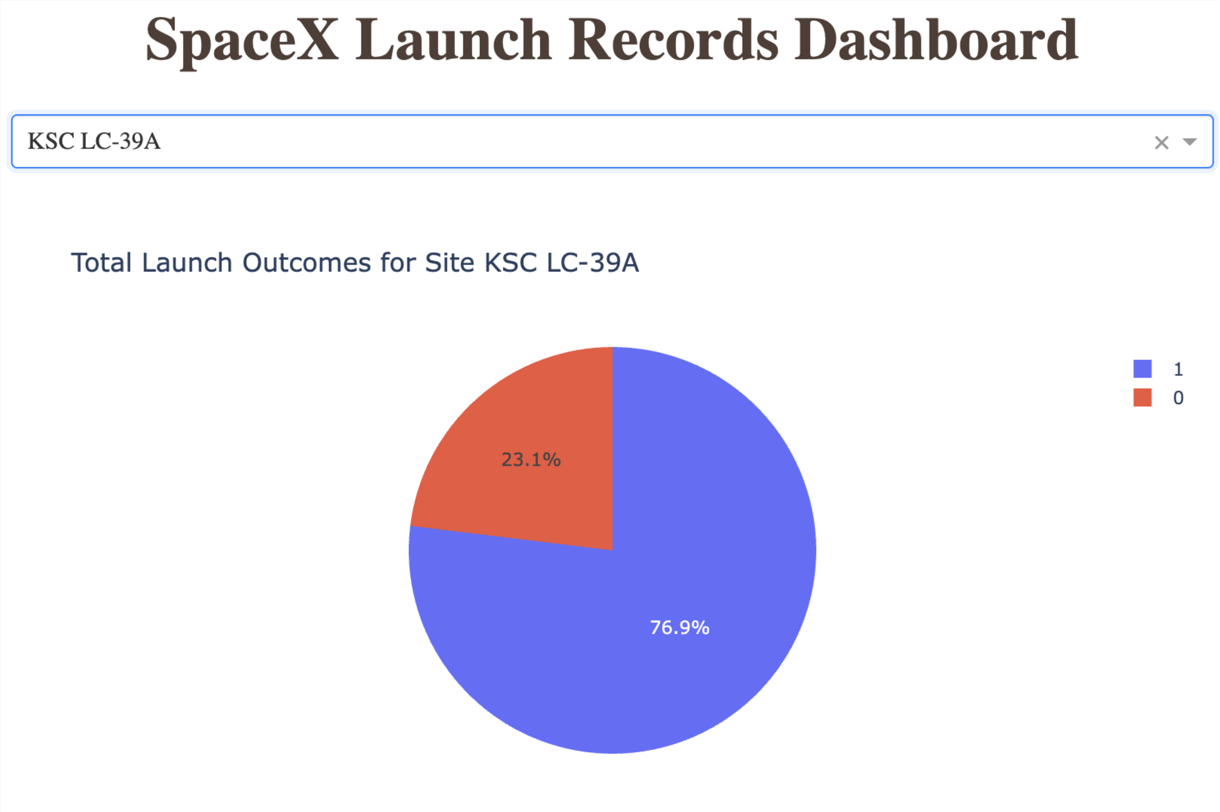
Interactive Dash Application

- Next, we created an interactive dashboard, allowing users to filter the data and see updated graphs in real time
- By filtering the data to all sites, we can view a pie chart of launch success count for all sites, as shown to the right
- This can help show us which launch sites had the highest proportion of the successful launches



Launch Site with Highest Launch Success Ratio

- Using the dropdown menu we created, we can view total launch outcomes of individual sites
- By comparing the pie charts of each site, we can see that **KSC LC-39A** is the launch site with the highest success ratio
- This was the only launch site with a success ratio greater than 50%



Payload vs. Launch Outcome

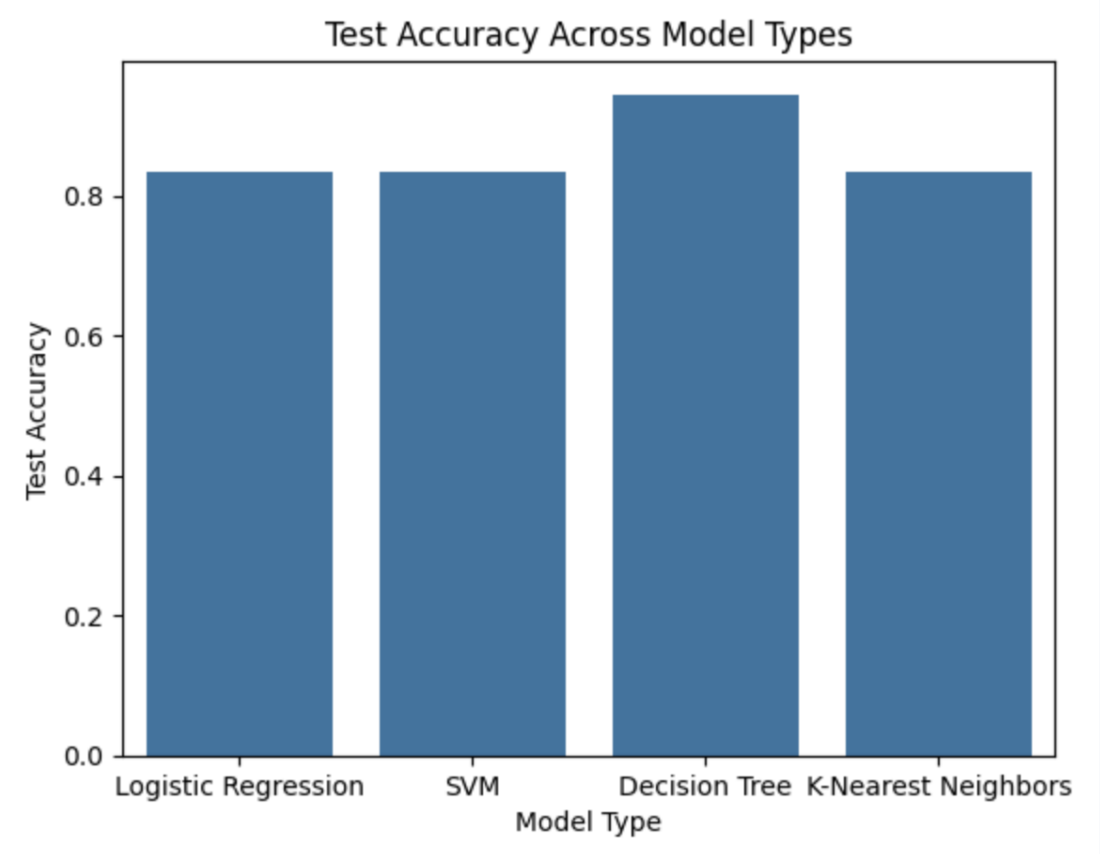
- We also created a scatter plot which plots Payload vs. Launch Outcome
  - This can be further filtered per launch site or via including or excluding certain ranges of the payload mass
- By analyzing this graph, we can see that the proportion of failed launches increases at the higher end of the payload mass spectrum
- Additionally, we can see that some booster versions tend to be more successful than others, such as the FT boosters

Section 5

Predictive Analysis (Classification)

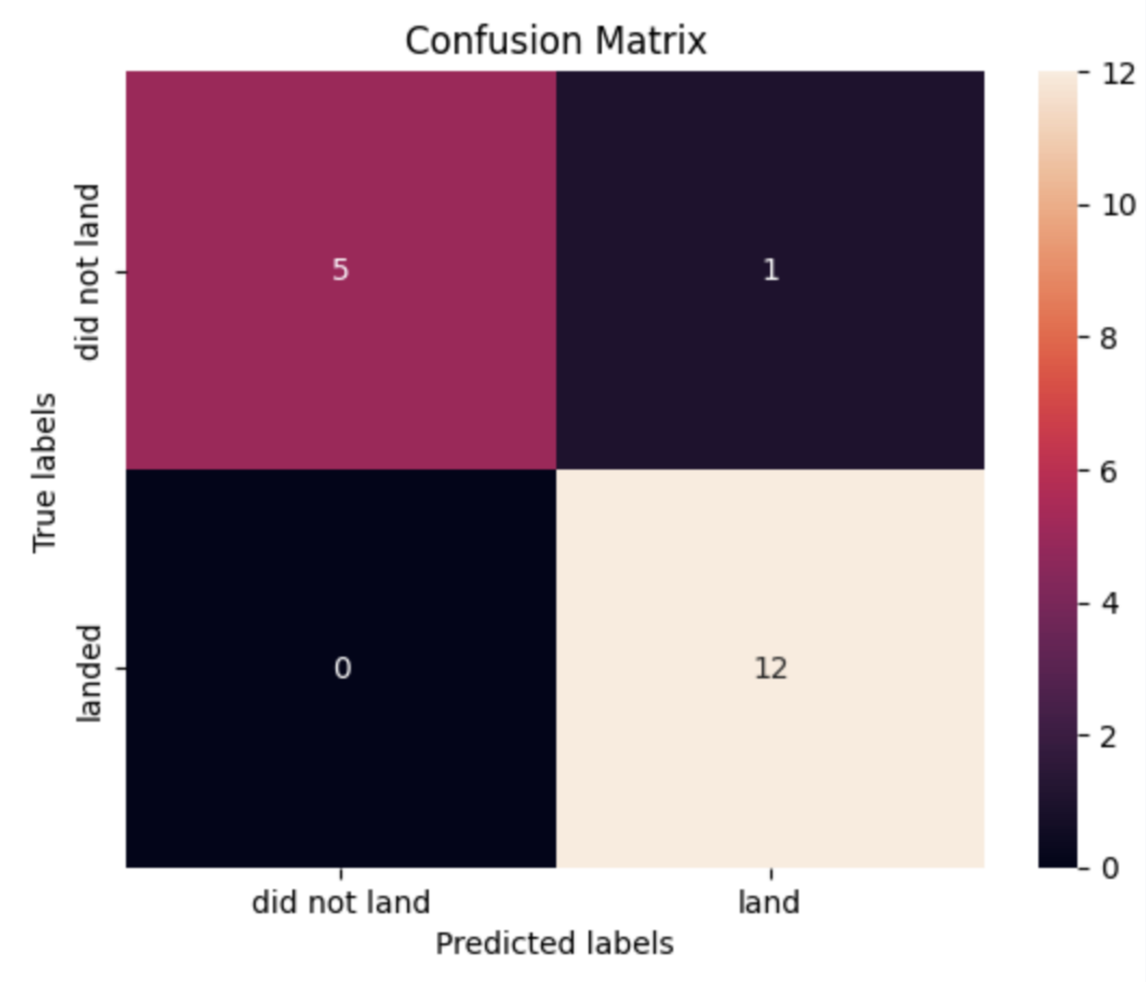
Classification Accuracy

- Finally, we tested 4 different predictive models on the data to find the one that performed the best, splitting the data into training and testing sets to help us with this determination
- From the test data, all of the model types had the same test accuracy (~83%), except for the Decision Tree classifier, which had ~94% test accuracy
- Therefore, the Decision Tree classifier was the best predictive model



Confusion Matrix

- For the decision tree classifier model we selected on the previous slide, we can further confirm this being a good predictive model by analyzing its confusion matrix
- From this, we can see only one instance in which the predicted labels don't match the true labels
    - In this case, the model predicted a landing when the booster did not land



Conclusions

- First-stage landing success has improved dramatically over time
- Launch site and landing infrastructure affect success probability
- Reusability and booster experience strongly correlate with success
- Payload mass and mission type significantly influence landing outcomes
- SpaceX's competitive cost advantage is **data-driven and engineering-driven**, supported by measurable improvements in reusability, launch infrastructure, and operational decision-making. Competitors must either match this learning curve or adopt alternative strategies to compete on launch cost.

Thank you!