IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Clara Tan Xin Yue
July 24 2025

Find all the documents here:
https://github.com/Claratxy/DataScienceAssignments/tree/a17c0bf72ed0f5890d186e541fbcec67e16f3a32/Applied%20Data%20Science%20Capstone

Outline

Executive Summary

Summary of methodologies

This project explores how we can predict whether SpaceX's Falcon 9 rocket first-stage will land successfully. These landings are important because reusing rockets can lower the cost of space launches.

To do this, we collected data using APIs and web scraping. We cleaned and prepared the data, then explored it using SQL and charts. We also used interactive maps to make the data easier to understand. Finally, we used machine learning to make predictions.

Summary of all results

The results showed that the rocket's payload size, orbit type, and launch site were important factors that affect landing success. The interactive visuals helped us see patterns more clearly, and our machine learning models were quite accurate. This kind of analysis could help companies better estimate costs and compete more effectively with SpaceX.

Introduction

Project background and context

SpaceX's Falcon 9 rockets can be reused after launch, which helps lower the cost. While other rocket launches can cost up to 165 million dollars, SpaceX does it for just 62 million because it reuses the rocket's first stage. That's why predicting whether the first stage will land safely is so important—it can help us estimate launch costs better.

This project was designed to figure out what affects the success of those landings. If other companies want to compete with SpaceX, knowing this information can help them plan smarter and offer better bids.

Steps to reach that goal

- Collected launch data from SpaceX using APIs and web scraping
- Cleaned and combined the data so we could analyze it
- Used charts, maps, and SQL to look for patterns and insights
- Built and tested a machine learning model to predict if the rocket will land successfully

Key problems to answer

- What affects the rocket's ability to land safely?
- How do these factors work together?
- What conditions lead to a successful landing?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API and Web Scraping from Wikipedia
- Perform data wrangling
  - Explore, clean, transform, integrate and validate data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression, Decision Tress, KNN and SVM

Data Collection

Data was gathered from two main sources:

1. **SpaceX API**

    - Provided real-time data about rockets, launch sites, payloads, and missions
    - Data was requested using GET, then converted into pandas DataFrames using JSON functions
    - This ensured accuracy and consistent structure for analysis

2. **Wikipedia Web Scraping**

    - Extracted historical launch records and extra mission info not found in the API
    - Used BeautifulSoup to collect data from HTML tables
    - This helped fill gaps and verify the API data

Data Collection – SpaceX API

Using REST API calls:

- Made GET requests to fetch JSON responses.
- Parsed the responses using .json() and normalized with pandas.json_normalize().
- Transformed the structured data into pandas DataFrames, ready for analysis.

Find this notebook here:

https://github.com/Claratxy/DataScienceAssignments/blob/a17c0bf72ed0f5890d186e541fbcec67e16f3a32/Applied%20Data%20Science%20Capstone/1.%20jupyter-labs-spacex-data-collection-api-v2.ipynb

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNet
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Use json_normalize meethod to convert the json result into a dataframe
data=pd.json_normalize(response.json())
```

Task 2: Filter the dataframe to only include `Falcon 9` launches

Finally we will remove the Falcon 1 launches keeping only the Falcon 9 launches. Filter the data dataframe using the `BoosterVersion` column to only keep the Falcon 9 launches. Save the filtered data to a new dataframe called `data_falcon9`.

```
# Hint data['BoosterVersion']!='Falcon 1'
data_falcon9 = df2[df2['BoosterVersion'] == "Falcon 9"]
data_falcon9.head()
```

Task 3: Dealing with Missing Values

Calculate below the mean for the `PayloadMass` using the `.mean()`. Then use the mean and the `.replace()` function to replace `np.nan` values in the data with the mean you calculated.

```
# Calculate the mean value of PayloadMass column
mean_payload_mass = data_falcon9['PayloadMass'].mean()

# Replace np.nan values with the calculated mean
data_falcon9['PayloadMass'].fillna(mean_payload_mass, inplace=True)
mean_payload_mass
```

Data Collection - Scraping

Used **BeautifulSoup** to scrape Falcon 9 and Falcon Heavy launch data from Wikipedia to capture historical landing outcomes and mission details that were not available through the API.

- Parsed HTML tables into **pandas DataFrames** for analysis
- Filled data gaps and validated API results for accuracy
- Enhanced feature diversity, boosting the quality of our predictive model

Find this notebook here:

https://github.com/Claratxy/DataScienceAssignments/blob/a17c0bf72ed0f5890d186e541fbcec67e16f3a32/Applied%20Data%20Science%20Capstone/2.%20jupyter-labs-webscraping.ipynb

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

```
# use requests.get() method with the provided static_url
html_data = requests.get(static_url)
print("Status code:", html_data.status_code)  # should be 200 if successful
```

```
Status code: 200
```

Create a `BeautifulSoup` object from the HTML `response`

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(html_data.text, 'html.parser')
```

TASK 2: Extract all column/variable names from the HTML table header

```
# Use the find_all function in the BeautifulSoup object, with element type `table`
# Assign the result to a list called `html_tables`
html_tables = soup.find_all('table')
```

TASK 3: Create a data frame by parsing the launch HTML tables

We will create an empty dictionary with keys from the extracted column names in the previous task. Later, this dictionary will be converted into a Pandas dataframe

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
```

Data Wrangling

Applied a full data wrangling workflow to clean, transform, and prepare the dataset for machine learning. Using **pandas**, **numpy**, and **sklearn**, we ensured consistency and accuracy throughout the process.

Key steps included:

- Exploring and cleaning the raw data
- Combining and validating multiple sources
- Calculating launch counts by site and orbit types
- Creating a landing outcome label for training
- Exporting the cleaned results to CSV for modeling

Find this notebook here:

https://github.com/Claratxy/DataScienceAssignments/blob/a17c0bf72ed0f5890d186e541fbcec67e16f3a32/Applied%20Data%20Science%20Capstone/3.%20labs-jupyter-spacex-Data%20wrangling-v2.ipynb

TASK 1: Calculate the number of launches on each site

```
# Apply value_counts() on column LaunchSite
df['LaunchSite'].value_counts()
```

TASK 2: Calculate the number and occurrence of each orbit

Use the method `.value_counts()` to determine the number and occurrence of each orbit in the column `Orbit`

```
# Apply value_counts on Orbit column
df['Orbit'].value_counts()
```

TASK 3: Calculate the number and occurence of mission outcome of the orbits

Use the method `.value_counts()` on the column `Outcome` to determine the number of `landing_outcomes`. Then assign it to a variable landing_outcomes.

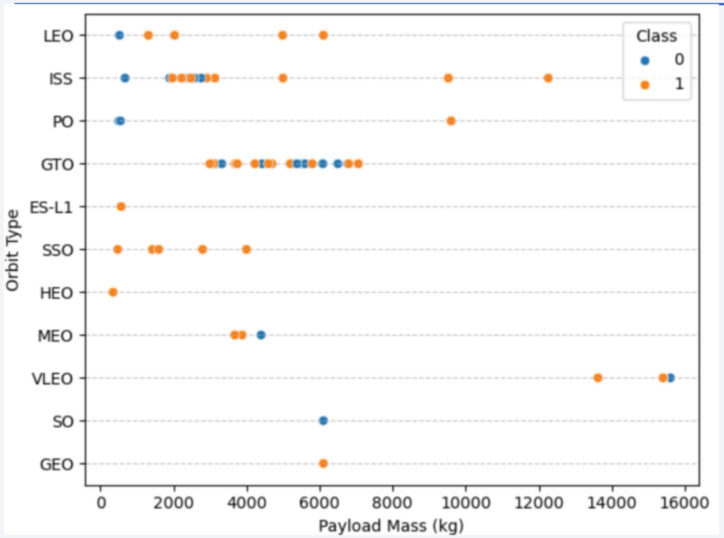
```
# landing_outcomes = values on Outcome column
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes
```

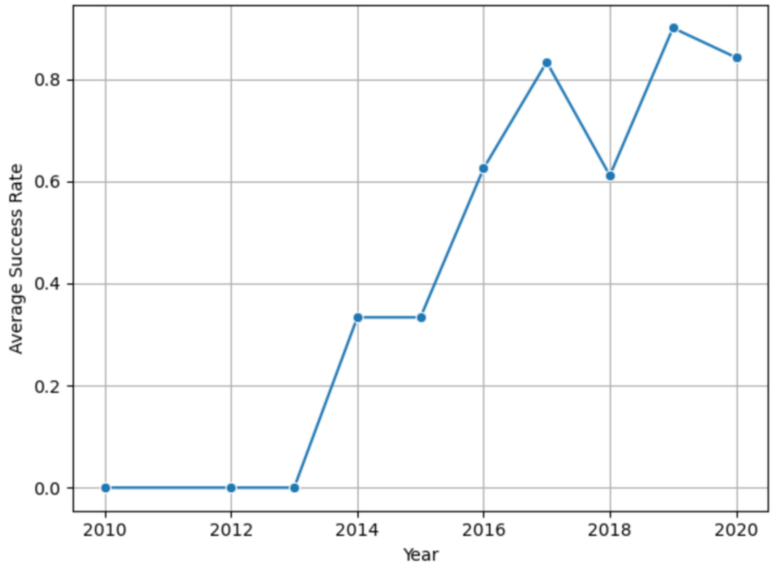
TASK 4: Create a landing outcome label from Outcome column

Using the `Outcome`, create a list where the element is zero if the corresponding row in `Outcome` is in the set `bad_outcome`; otherwise, it's one. Then assign it to the variable `landing_class`:

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
bad_outcomes = set([
    'None None', 'None failure', 'Crash', 'Ocean failure',
    'Landing pad failure', 'Uncontrolled', 'Precluded',
])
landing_class = [0 if outcome in bad_outcomes else 1 for outcome in df['Outcome']]
df['Class'] = landing_class
```

EDA with Data Visualization





Summary of Charts Plotted:

- Scatter Plots: Showed relationships between flight number, payload mass, orbit type, and launch site to spot trends and anomalies
- Bar Charts: Compared success rates across orbit types to identify which were more reliable
- Line Charts: Tracked yearly launch success trends, highlighting SpaceX's improvement over time

Find this notebook here:

https://github.com/Claratxy/DataScienceAssignments/blob/a17c0bf72ed0f5890d186e541fbcec67e16f3a32/Applied%20Data%20Science%20Capstone/5.%20jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

SQL-based analysis was performed on the SpaceX dataset directly within the Jupyter notebook using a PostgreSQL database. This allowed us to uncover valuable insights and answer specific business questions.

Key steps included:

- Identified unique launch sites and summarized mission outcomes
- Calculated payload totals and averages based on mission type and booster versions
- Filtered records to analyze failed drone ship landings and their related booster and site details
- Used grouping and subqueries to track performance trends, milestones, and technical benchmarks

Find this notebook here:

https://github.com/Claratxy/DataScienceAssignments/blob/a17c0bf72ed0f5890d186e541fbcec67e16f3a32/Applied%20Data%20Science%20Capstone/4.%20jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Map Objects Added:

- Markers: Plotted each launch site to show geographic distribution
- Color-coded Circles: Represented launch outcomes (success or failure) at each site
- Marker Clusters: Grouped densely located sites (e.g., in Florida) to avoid overlap and improve clarity
- Distance Lines: Drew lines from launch sites to nearby coastlines to study safety and proximity

Reasons for adding those objects:

- Helped distinguish between successful and failed launches visually
- Made it easy to spot sites with high success rates
- Enabled location-based insights—such as proximity to cities, highways, or coastlines—that may influence landing performance

Find this notebook here:

https://github.com/Claratxy/DataScienceAssignments/blob/a17c0bf72ed0f5890d186e541fbcec67e16f3a32/Applied%20Data%20Science%20Capstone/6.%20lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

Charts and Graphs Added:

- Pie Charts: Show launch success counts by site and overall outcomes (success vs. failure)
- Scatter Plot: Displays the relationship between payload mass, launch outcomes, and booster version

Interactive Features:

- Dropdown Menu: Allows users to select specific launch sites, dynamically updating all visualizations
- Payload Range Slider: Filters the scatter plot by payload mass in real time for targeted exploration

Find this notebook here:

https://github.com/Claratxy/DataScienceAssignments/blob/a17c0bf72ed0f5890d186e541fbcec67e16f3a32/Applied%20Data%20Science%20Capstone/7.%20Interactive%20Dashboard%20with%20Plotly%20Dash.py

Predictive Analysis (Classification)

Workflow Overview:

- Loaded and transformed data using numpy and pandas
- Split into training and testing sets for evaluation
- Applied models including Logistic Regression, SVM, Decision Tree, and KNN
- Tuned hyperparameters with GridSearchCV and cross-validation
- Used accuracy as our performance metric

After testing, we selected the best-performing model based on accuracy, supported by feature engineering and algorithm tuning.

Find this notebook here:

https://github.com/Claratxy/DataScienceAssignments/blob/a17c0bf72ed0f5890d186e541fbcec67e16f3a32/Applied%20Data%20Science%20Capstone/8.%20SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

Results

Exploratory Data Analysis (EDA):

- Most launches occurred at two major sites in Florida and California
- Charts and maps helped visualize payload mass and launch frequency trends
- Launch success rates varied by site and orbit type

Descriptive Analytics:

- Dashboards and maps allowed users to explore outcomes by site, booster type, and payload
- Larger payloads didn't always lead to lower success—trend and outlier detection was made easy

Predictive Analysis:

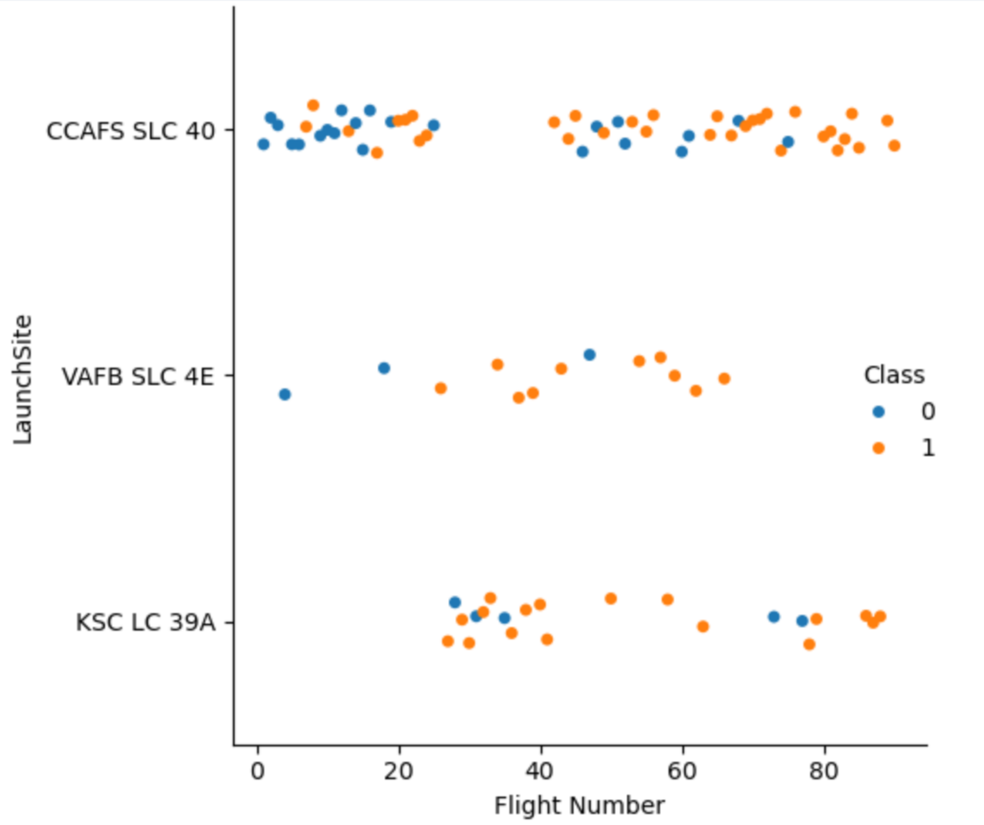
- Built and evaluated multiple machine learning models: Logistic Regression, SVM, Decision Tree, and KNN
- All models reached around **83% accuracy**, with no standout winner
- Showed strong potential for forecasting landing success and informing cost predictions

Section 2

Insights drawn from EDA

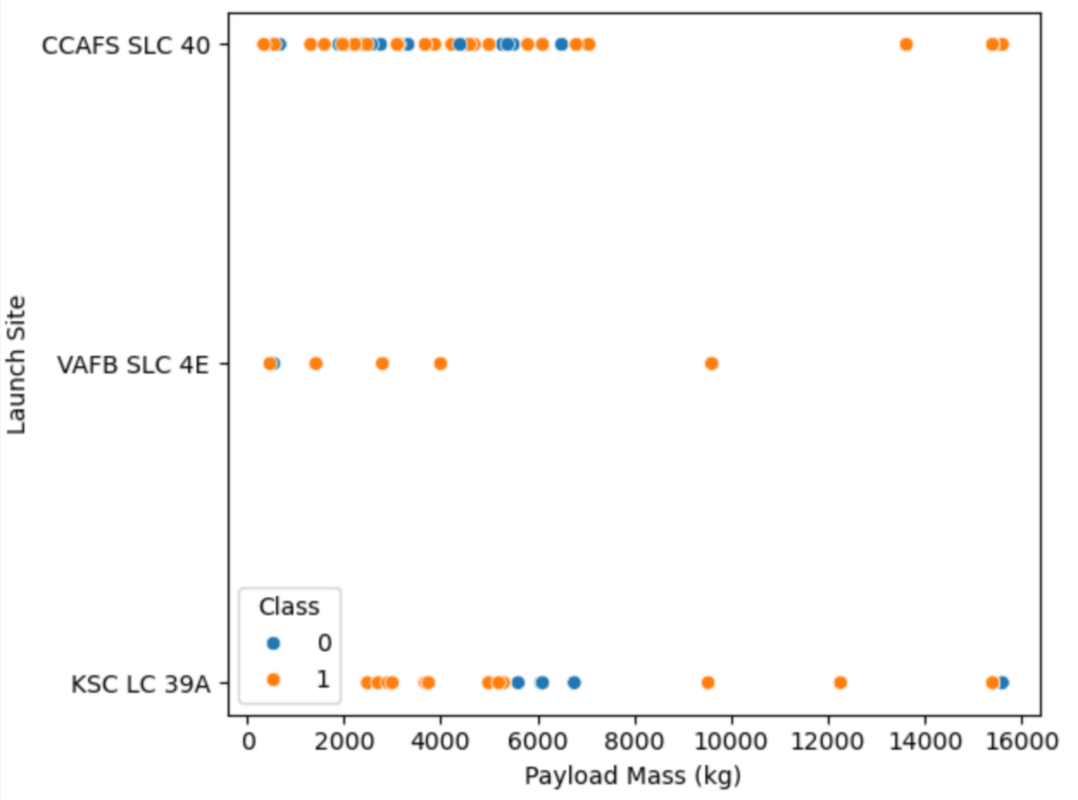
Flight Number vs. Launch Site

- CCAFS SLC-40 in Florida shows consistent launch activity across a wide range of flight numbers, making it the most frequently used site
- KSC LC-39A flights are mostly at higher flight numbers, showing more recent and increasing use
- VAFB SLC-4E in California has fewer launches, with flight numbers more widely distributed
- As flight numbers rise, success rates (Class I landings) improve across all sites—pointing to operational learning and refinement over time
- The data suggests a direct link between higher launch frequency at a site and better landing outcomes



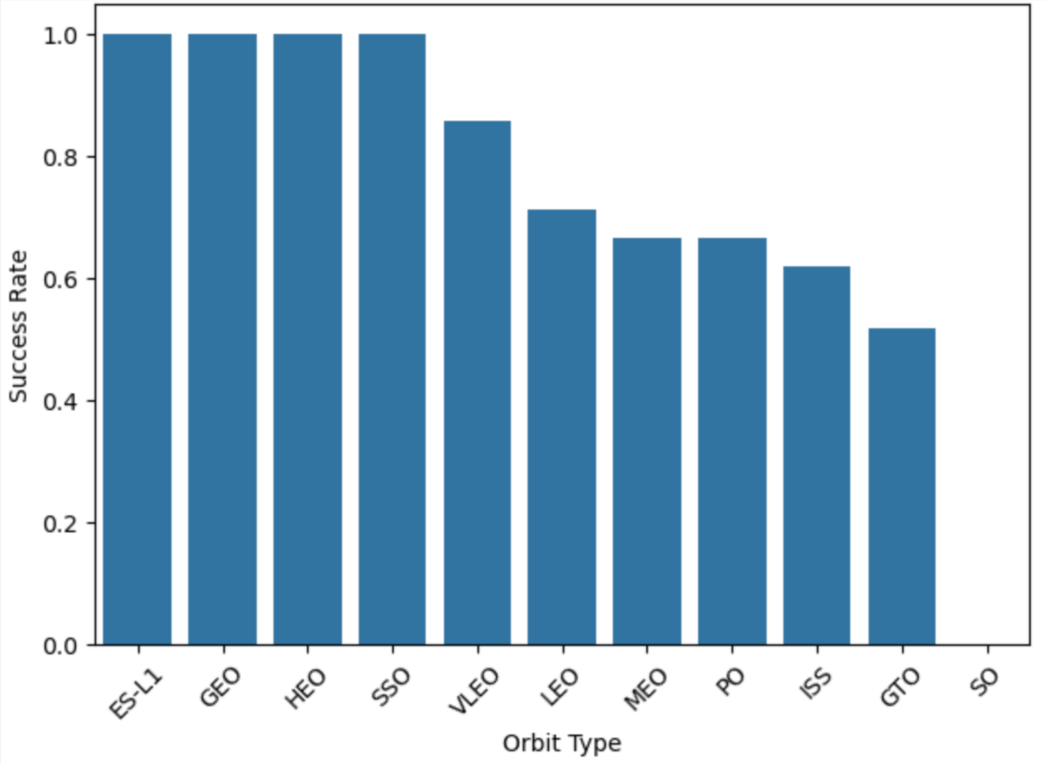
Payload vs. Launch Site

- SpaceX's launch trends show how different sites handle varying payload sizes and landing success:
- CCAFS SLC-40 mostly launches payloads under 6,000 kg, with frequent successful landings at these lower weights
- KSC LC-39A supports a wide range of payloads—from light to heavy—with successful landings across all mass categories
- VAFB SLC-4E handles fewer launches, mainly with lighter payloads; none exceed 10,000 kg, but all have landed successfully
- Although rare, launches with very heavy payloads (10,000–16,000 kg) are generally successful and mainly occur at CCAFS SLC-40 and KSC LC-39A
- At CCAFS SLC-40, higher payload masses are linked to increased landing success, suggesting improved capability over time



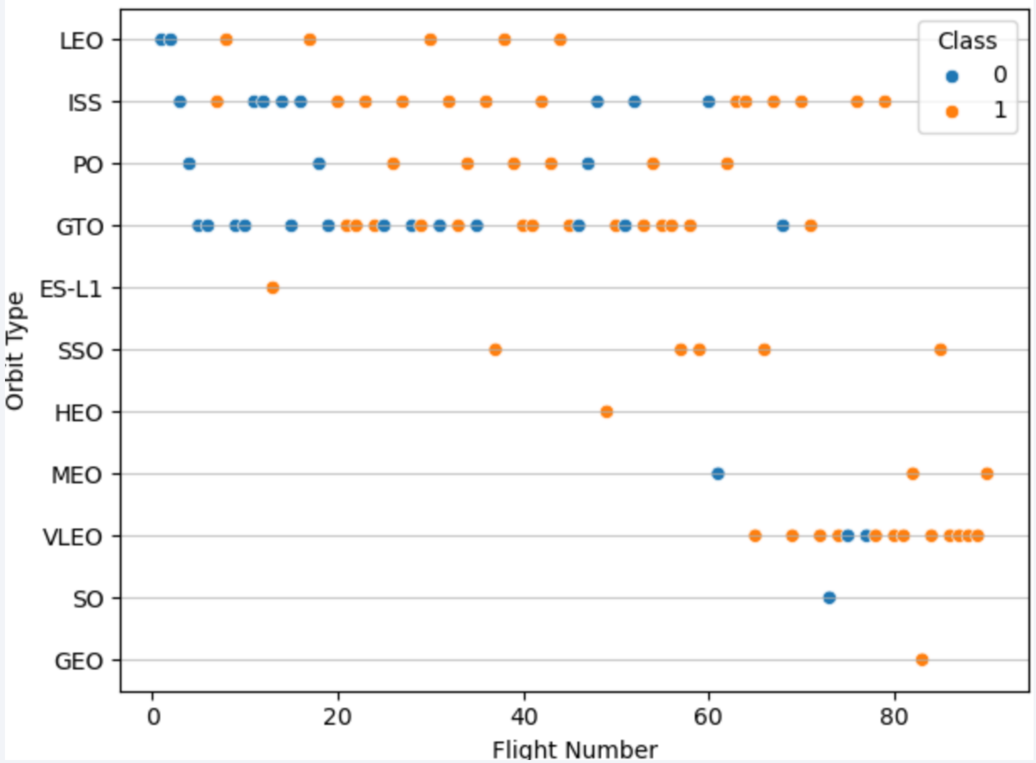
Success Rate vs. Orbit Type

- Orbit types like ES-LI, GEO, HEO, SSO, and VLEO show consistently high success rates—often with 100% successful landings
- More frequently used orbits such as LEO and PO have lower landing success, likely due to complex mission demands or higher operational frequency
- GTO (Geostationary Transfer Orbit) has the lowest landing success rate among all orbit categories studied
- These differences suggest that certain orbit missions present greater technical challenges, affecting recovery performance



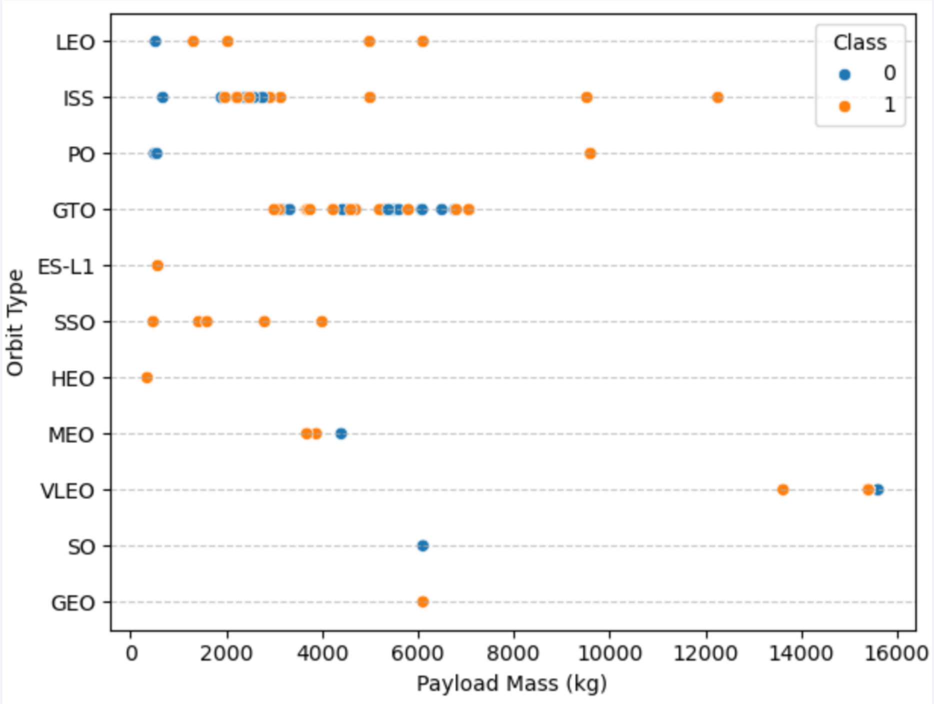
Flight Number vs. Orbit Type

- Missions to **VLEO** and **MEO** mostly appear at higher flight numbers
- **LEO** and **GTO** missions are spread across the entire flight range, but LEO missions show improved success rates at higher flight numbers
- In contrast, **GTO** launches do not display a clear link between flight number and landing success
- **ISS** and **PO** missions show steady launch frequency, though their landing outcomes vary
- Rare orbits like **ES-LI**, **GEO**, **HEO**, and **SSO** are limited to select flight numbers, and all resulted in successful landings



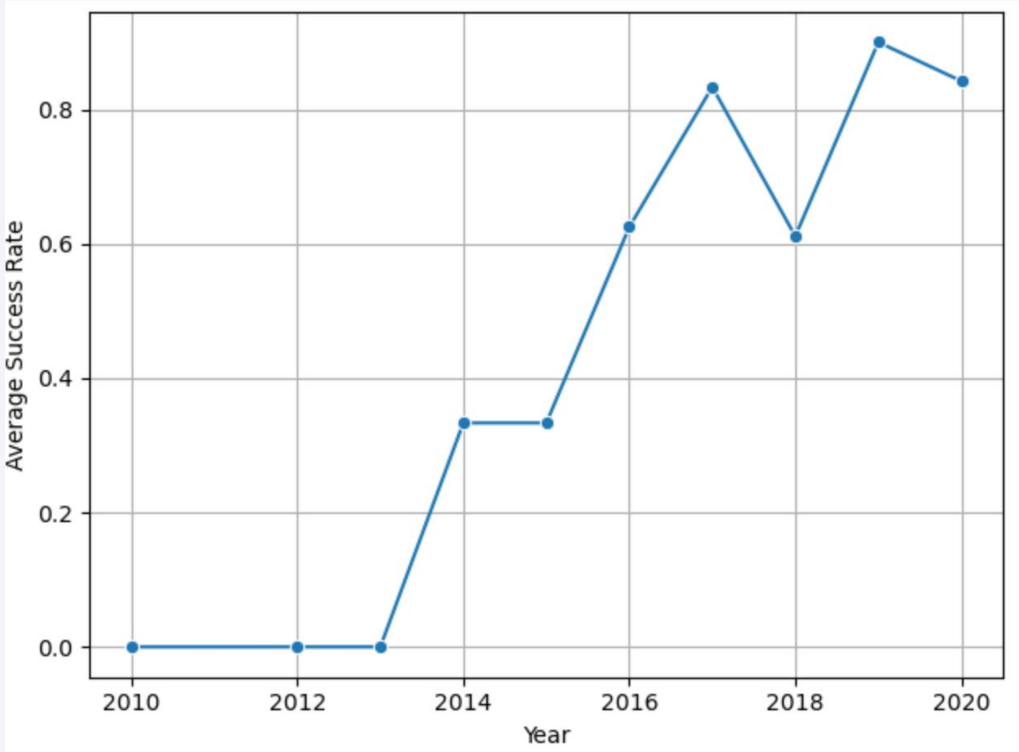
Payload vs. Orbit Type

- ISS Missions Most missions carry payloads ranging from 2,000 to 4,000 kg, showing a high rate of successful landings.
- GTO Missions Typically involve payloads of 5,000 to 6,000 kg, but outcomes are mixed. Success rate does not clearly correlate with payload mass.
- High Payload Missions (10,000+ kg) Launches carrying large payloads across orbit types are relatively uncommon.
- VLEO and SSO Missions Despite a wide range of payload masses, these orbits consistently show strong landing success.
- Polar, LEO, and ISS Missions with Heavy Payloads These orbit types show higher landing success rates when launching heavier payloads.
- GTO Missions with Heavy Payloads GTO missions do not show a clear relationship between payload size and landing success.



Launch Success Yearly Trend

- From **2010 to 2013**, very few missions achieved successful landings, highlighting the early experimental stage.
- **2014** was a stable year, showing slight improvement but no major change in overall success.
- Starting in **2015**, the success rate began to climb significantly.
- Between **2015 and 2017**, performance improved rapidly, with the average success rate surpassing **80% in later years**.
- This upward trend reflects SpaceX's growing expertise in first-stage landings, likely driven by advanced technology and increased operational experience.



All Launch Site Names

The dataset identifies three primary launch locations used by SpaceX:

- CCAFS LC-40, VAFB SLC-4E, and KSC LC-39A are the main launch sites listed.
- CCAFS LC-40 appears more than once, which could indicate inconsistent labeling or duplication in the data.
- To ensure only unique site names were shown in the dataset, the SQL keyword DISTINCT was used during data processing.

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

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

```
 * 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'

- All five launch records shown originate from CCAFS LC-40.
- The missions dated 2010 to 2013 mostly involved demo flights or payloads with zero mass.
- Every mission listed resulted in successful landings, demonstrating a strong performance record from this launch site during the displayed timeframe.
- The SQL query used included the keyword 'LIKE 'CCA%' to filter and display only those records where the launch site name begins with "CCA".

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

```
%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_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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 (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- NASA's Commercial Resupply Service (CRS) missions have transported a total payload mass of 45,596 kg using SpaceX launch vehicles.
- This figure highlights the strength of the collaboration between SpaceX and NASA, especially for servicing the International Space Station.
- The substantial total payload reflects both the frequency and reliability of these missions over time.

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" = 'NASA (CRS)';
```

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

TOTAL_PAYLOAD_MASS
45596

Average Payload Mass by F9 v1.1

- The Falcon 9 v1.1 booster has an average payload capacity of 2,928.4 kg.
- This value serves as a reference point for planning missions that use this specific booster model.
- It helps evaluate changes in payload capabilities across different versions of Falcon 9.
- Valuable for monitoring how upgrades have improved booster performance over time.

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"= 'F9 v1.1';
```

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

Avg_Payload_Mass
2928.4

First Successful Ground Landing Date

- SpaceX achieved its first successful ground landing on December 22, 2015.
- This marked a major milestone in the company's efforts to develop reusable launch technology.
- In the dataset, the landing date was identified using filtering logic to match the first successful landing on a ground pad.

```
%sql SELECT MIN(Date) AS First_Succesful_landing FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

First_Succesful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Four distinct Falcon 9 Full Thrust boosters (B1022, B1026, 31021.2, and 31031.2) successfully landed on drone ships while carrying payloads in the 4,000 to 6,000 kg range.
- Success within this mass range highlights SpaceX's capability to handle technically demanding recovery operations.
- The result was obtained using an SQL query that filtered boosters where the landing platform was a drone ship and the payload mass was greater than 4000 kg but less than 6000 kg.

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

```
%sql SELECT "Booster_version" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND "Payload_Mass__kg_" > 4000 AND "Payload_Mass__kg_" <
```

```
 * 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

- Out of all recorded missions, 98 were successful, with only a few failures or ambiguous results.
- Only one in-flight failure was observed.
- Some outcomes labeled as "Success (payload status unclear)" point to minor gaps in mission reporting.
- To generate these counts, a SQL query filtered records where MissionOutcome included either success or failure using wildcard matching.

List the total number of successful and failure mission outcomes

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

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

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

Boosters Carried Maximum Payload

- Several Falcon 9 Block 5 boosters reached the maximum recorded payload capacity of 15,600 kg.
- Consistent performance across multiple Block 5 units reflects design robustness and reliable engineering.
- The boosters were identified by applying a subquery with the MAX() function and filtering with the WHERE clause.

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate funct

```
%sql SELECT "Booster_Version", "Payload_Mass__kg_" FROM SPACEXTBL WHERE "Payload_Mass__kg_" = (SELECT MAX("Pa
```

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

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

- In 2015, two drone ship landing failures occurred: one in January and one in April, both at CCAFS LC-40.
- These missions used the Falcon 9 v1.1 booster, suggesting potential technical or environmental challenges during that period.
- The records were filtered using a combination of SQL conditions: WHERE, LIKE, AND, and BETWEEN to isolate failed drone ship landings in 2015.

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql SELECT SUBSTR(Date, 6, 2) AS Month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Fa
```

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

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

- The most common landing outcome during this period was "No attempt" (10 instances), typical for early missions that either served as test flights or were not intended to recover the booster.
- Drone ship landings had an even balance of outcomes, with 5 successful recoveries and 5 failures.
- Ground pad landings saw 3 successes, while ocean recoveries (both controlled and uncontrolled) made up the remaining outcomes.

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-

```
%sql SELECT "Landing_Outcome", Count(*) AS Outcome_Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20
```

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

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

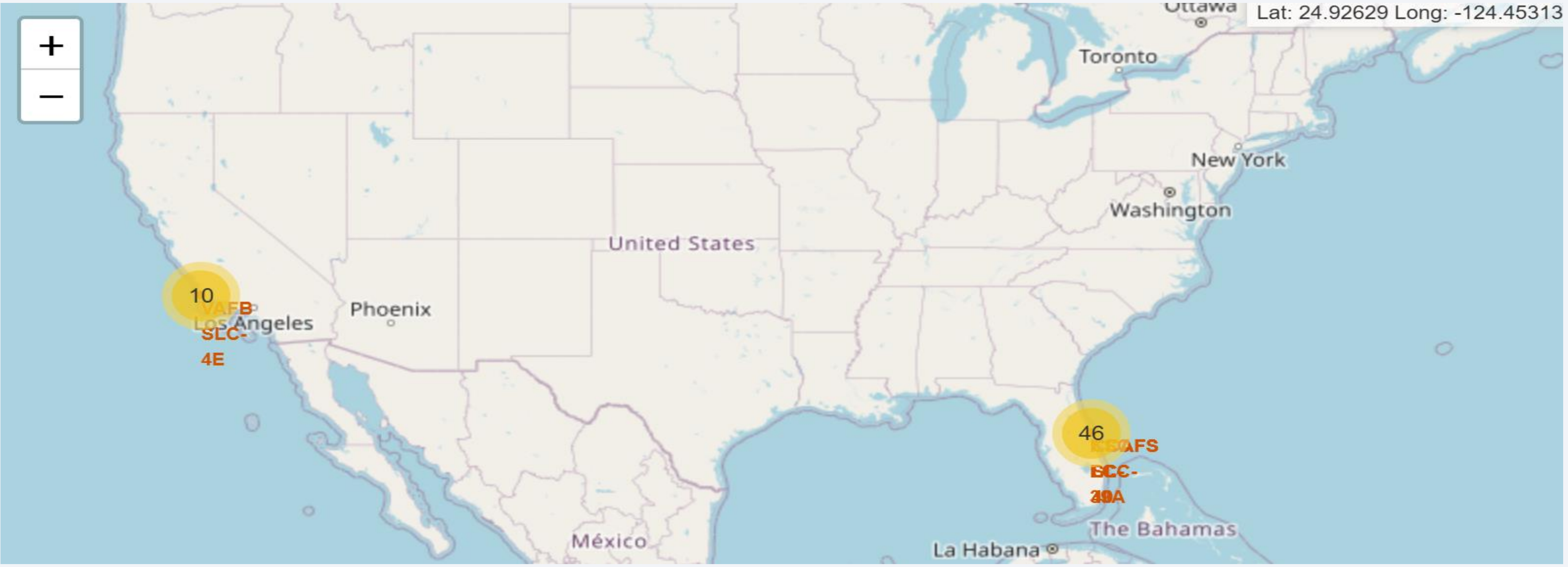
Section 3

Launch Sites Proximities Analysis

SpaceX Launch Site Overview

SpaceX operates several launch sites across the United States, each serving different mission needs:

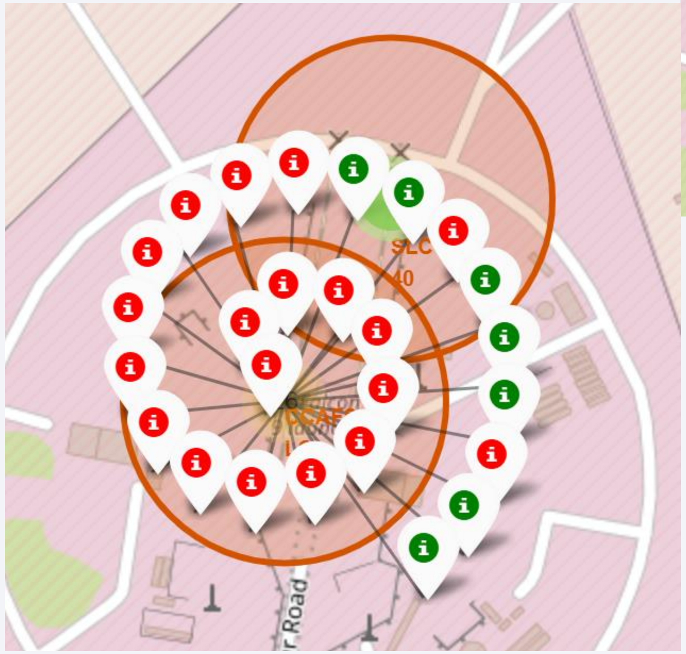
- Most launches happen at CCAFS LC-40 in Florida, making it the busiest site
- Other key sites include KSC LC-39A (Florida) and VAFB SLC-4E (California)
- Florida's East Coast is SpaceX's main launch hub due to high activity
- The California site (VAFB SLC-4E) is used less often, mainly for polar or sun-synchronous missions
- By spreading launch sites geographically, SpaceX can target a wide range of orbits and meet different customer needs

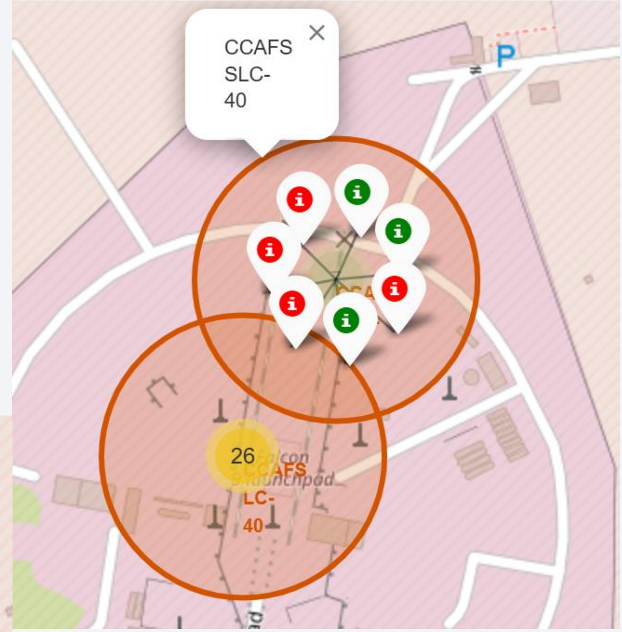


Launch and Landing Patterns at CCAFS SLC-40

SpaceX's activity at CCAFS Launch Complex 40 highlights key recovery strategies:

- Multiple landings—both successful and failed—have occurred at the same location, reflecting SpaceX's trial-and-error approach to perfecting rocket recovery.
- Visual markers show progress over time: green for successful landings and red for failures, making performance trends easy to identify.
- The pad functions as both a launch site and a landing zone, maximizing the use of existing infrastructure.
- Landings happen within a tightly controlled area, helping SpaceX manage risk while fine-tuning precision.

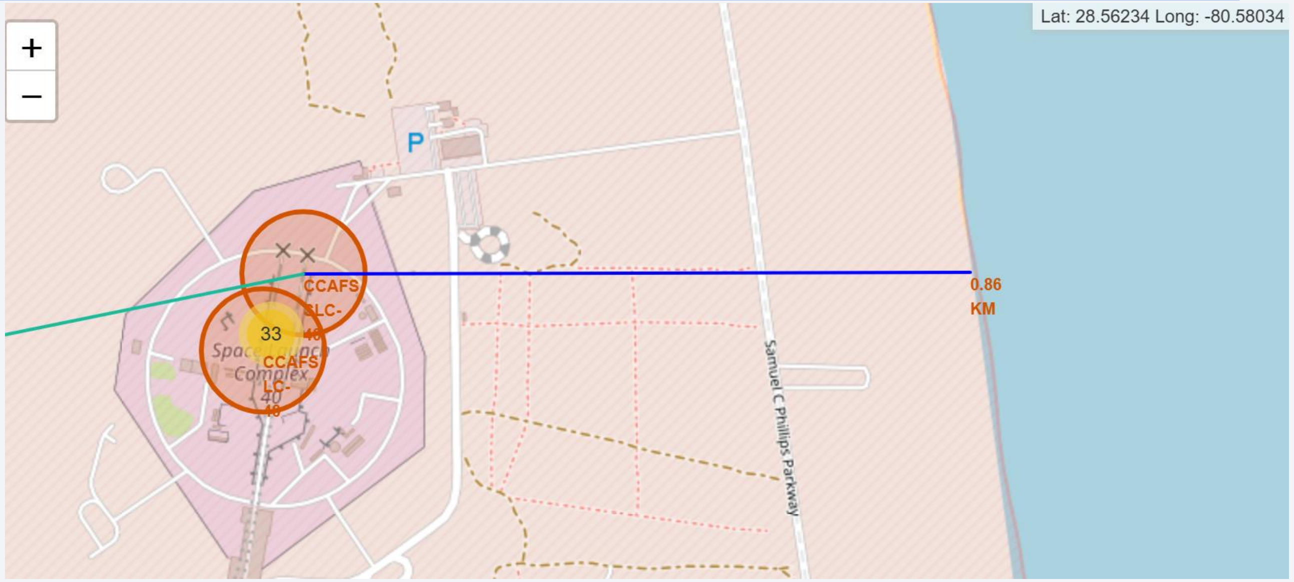


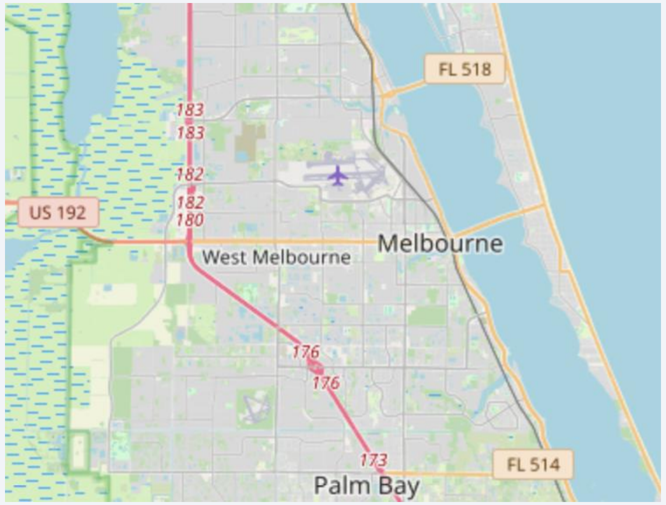


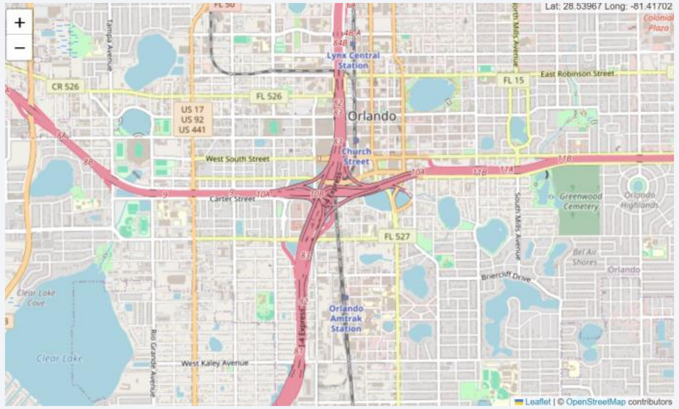
Coastal Advantage of CCAFS SLC-40

The SpaceX launch pad at Cape Canaveral Space Force Station is strategically located close to the ocean:

- Just 0.9 km from the coastline, the site reduces risks to land-based areas and enables direct launches over water
- Coastal access improves safety and helps meet regulatory requirements
- The short distance also facilitates efficient drone ship recovery of rocket boosters at sea
- This location demonstrates how natural geography plays a key role in selecting launch sites





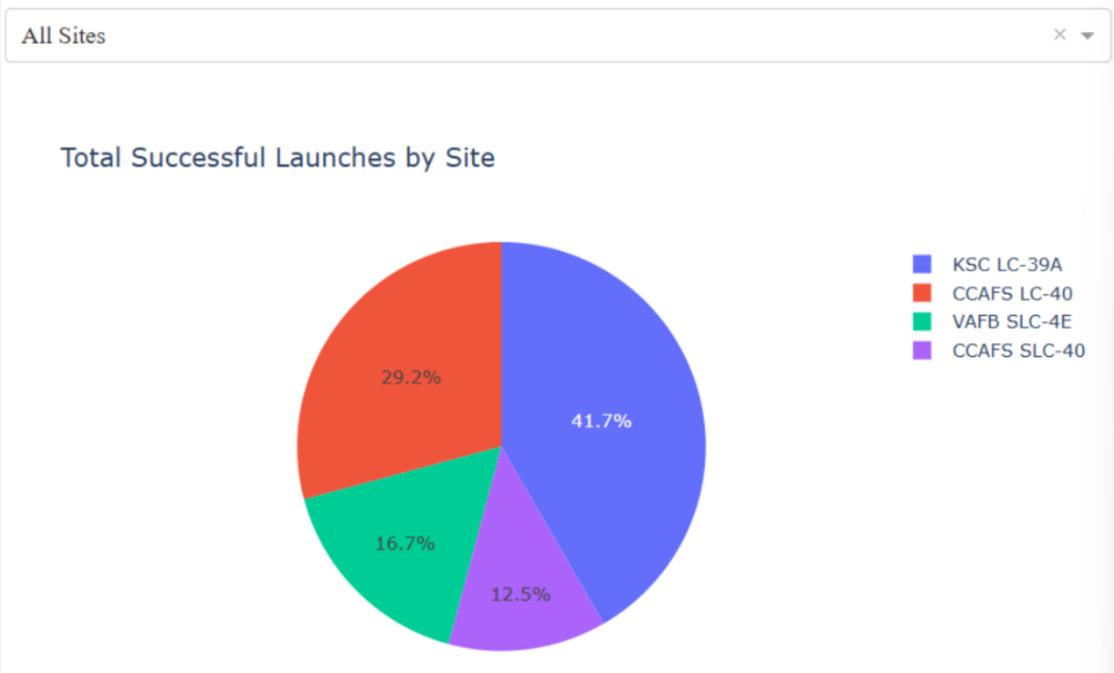


Section 4

Build a Dashboard with Plotly Dash

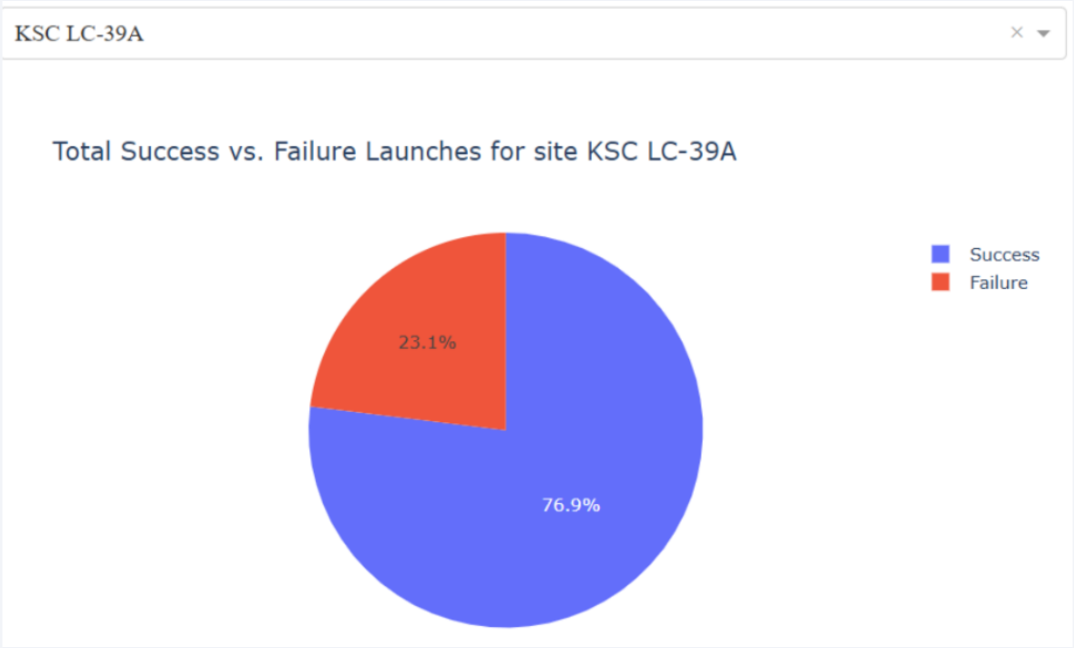
Distribution of Successful Launches by Site

- The launch site KSC LC-39A contributed the largest share of successful missions, accounting for 41.7%.
- It is followed by CCAFS LC-40 with 29.2%, VAFB SLC-4E with 16.7%, and CCAFS SLC-40 with 12.5%.
- The presence of both “LC-40” and “SLC-40” suggests potential inconsistencies or duplication in site naming within the data.



Launch Success vs. Failure at KSC LC-39A

- At the site KSC LC-39A, 76.9% of launches were successful, while 23.1% resulted in failure.
- The high success ratio confirms this site’s importance for high-stakes missions and its overall dependability.
- Although some failures occurred, they do not significantly affect the site's reputation for reliability.



Relationship Between Payload and Landing Success

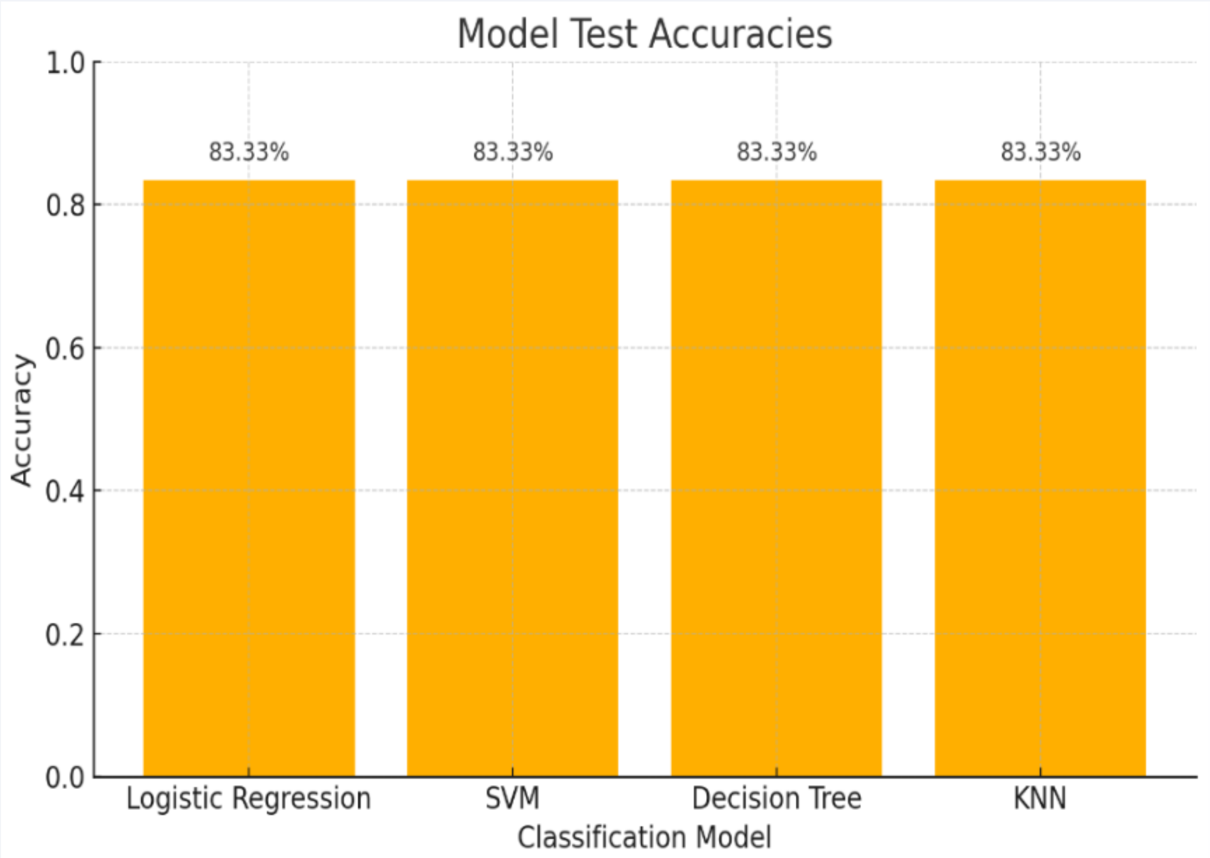
- A scatter plot was used to explore how payload mass correlates with launch outcomes, across various booster types.
- Most successful missions (classified as '1') fall in the lower to mid payload range, typically up to around 6000 kg.
- The FT booster variant shows the highest concentration of successful outcomes.
- Failures (classified as '0') are more frequently tied to older booster versions like v1.0 and v1.1, mostly with lighter payloads.

Section 5

Predictive Analysis (Classification)

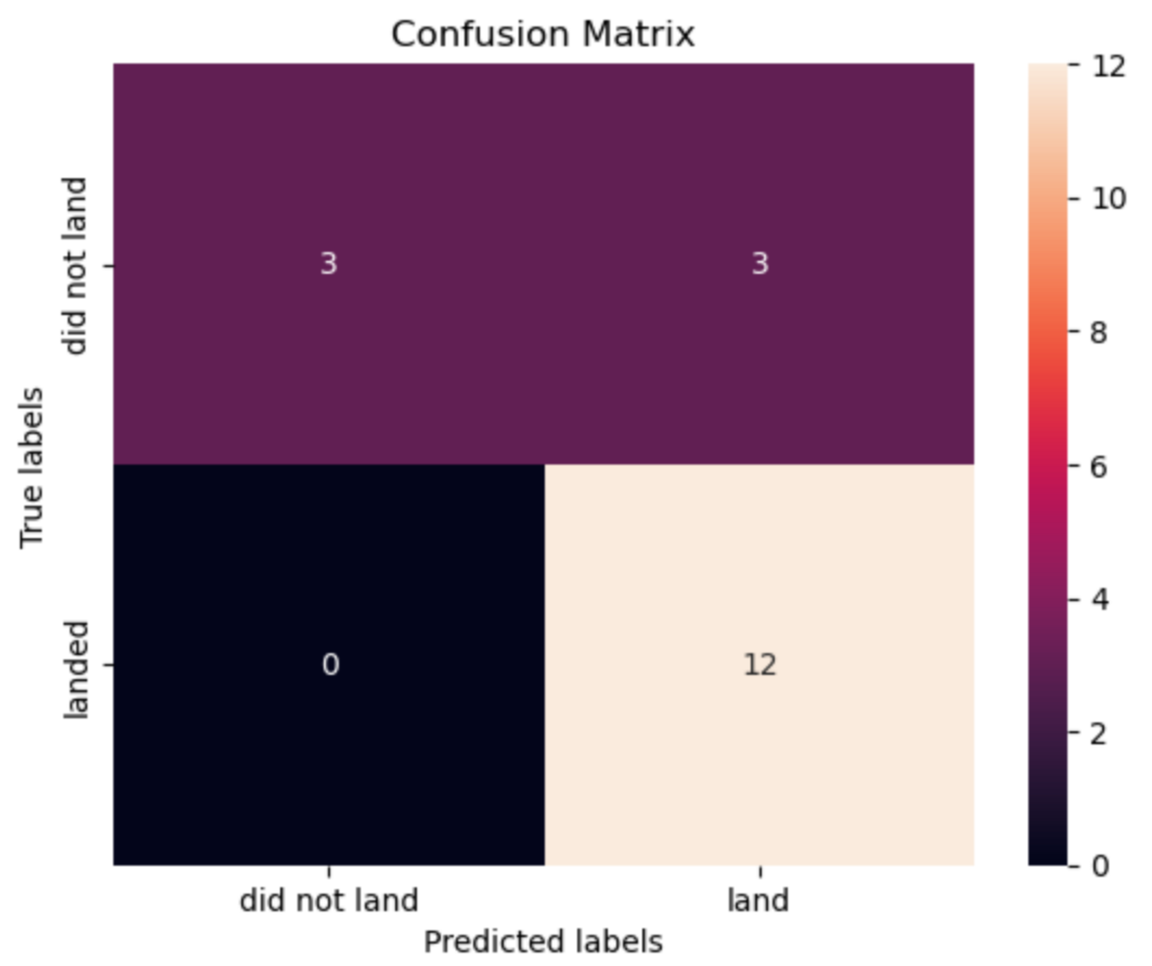
Classification Accuracy

- All models—Logistic Regression, SVM, Decision Tree, and KNN—achieved 83.33% accuracy.
- The equal accuracy suggests that changing models yields limited gains without further tuning or improved features.
- Although all models performed equally in terms of accuracy, the Decision Tree classifier was highlighted for interpretability and clear classification output.



Confusion Matrix

- The model achieved zero false negatives, correctly identifying all successful landings.
- False positives were present (3 cases), where failed landings were incorrectly predicted as successful, indicating some optimism in the model's predictions.
- Specificity was balanced at 50%, showing room for refining failure detection.



Conclusions

- All four models delivered consistent performance, driven by strong core features like payload mass, orbit type, and launch site.
- The error profile leans toward minimizing missed opportunities rather than avoiding unnecessary reuse attempts.
- Logistic Regression is preferred for its transparency, helping validate the impact of features like payload size and orbit type.
- To improve accuracy, next steps should include hyperparameter tuning, exploring feature interactions, and using ensemble methods.
- Additional findings reinforce broader trends:
  - Higher launch volume at a site correlates with improved success rates
  - Success rates steadily increased from 2013 to 2020Orbits such as ES-LI, GEO, HEO, SSO, and VLEO had the highest success rates
  - KSC LC-39A was the most reliable launch site overall
  - The Decision Tree model showed strong classification performance for landing outcomes

Appendix

All code, notebooks and supporting documents for this projects can access through:

https://github.com/Claratxy/DataScienceAssignments/tree/1121c1a4e6df975f3df6498e86925679da3aeed2/Applied%20Data%20Science%20Capstone

Thank you!