IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Duguma Yeshitla
2025.07.31

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

This project employed a comprehensive data science workflow to analyze SpaceX launch records. Data was collected through web scraping and API access, then cleaned and structured for analysis. Exploratory Data Analysis (EDA) and SQL queries were used to identify key trends in booster performance and payload distribution. Geospatial visualization with Folium provided insights into launch site proximities, while an interactive Plotly Dash dashboard enabled intuitive exploration of mission outcomes.

Multiple classification models were trained to predict landing success, with Support Vector Machines (SVM) achieving the highest accuracy at 86%. The findings highlight improvements in launch success post-2015, with NASA and F9 v1.1 boosters playing a major role in heavy-payload missions.

Introduction

- SpaceX has attracted global attention through several historic achievements.
- It became the first private company to bring a spacecraft back from low-Earth orbit, achieving this in December 2010.
- On its website, SpaceX lists the price of a Falcon 9 rocket launch at 62 million dollars, while other providers charge upwards of 165 million dollars. A major factor in the lower cost is SpaceX's ability to reuse the first stage.
- Thus, if we can predict whether the first stage will land, we can estimate the launch cost.
- This insight is valuable for any competing company aiming to challenge SpaceX in rocket launch bids.

Section 1

Methodology

Methodology

Executive Summary

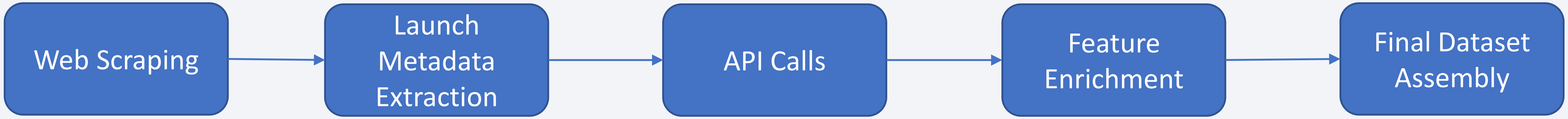
- Data collection methodology:
    - Data was collected using the Space-X open API and scrapping Wikipedia page regarding Space-X launches.
- Perform data wrangling
    - The target Landing outcome class was generated based on the cleaning & transformation of existing data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

The dataset was collected through a two-stage pipeline combining web scraping and API integration.

First, we scraped historical SpaceX launch data from the official SpaceX website using the BeautifulSoup library to retrieve launch details like mission name, date, and outcome.

Then, we enhanced this data by querying the SpaceX REST API and the OpenNotify API to gather additional features, including launch site coordinates, payload mass, and orbit details. This hybrid approach ensured both completeness and real-time accuracy.



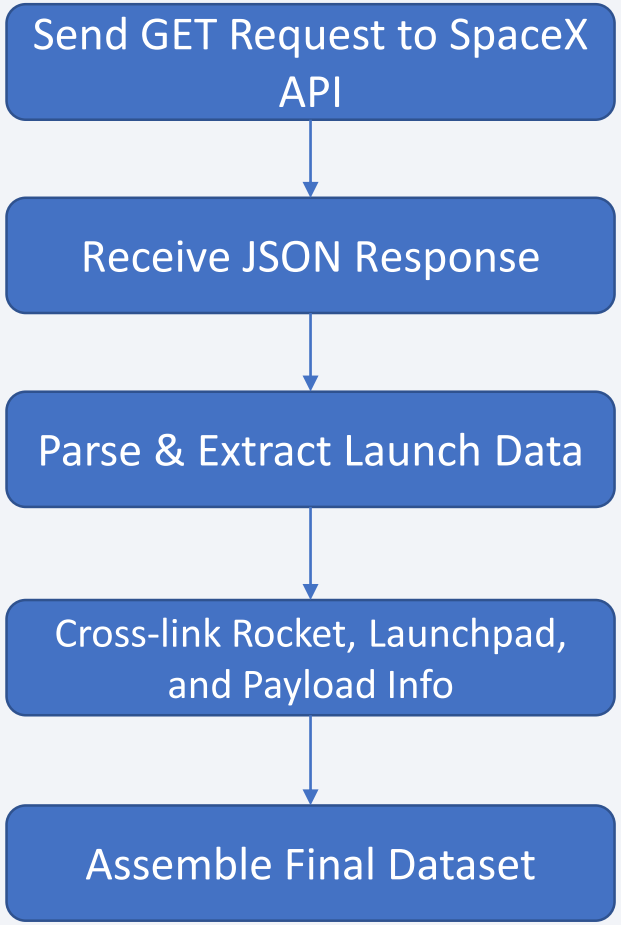
Data Collection – SpaceX API

We collected historical launch data using the SpaceX REST API by sending HTTP GET requests to endpoints such as:

- ***/v4/launches,***
- ***/v4/rockets,***
- ***/v4/launchpads, and***
- ***/v4/payloads.***

This allowed us to programmatically retrieve structured JSON data containing launch outcomes, rocket configurations, payload details, and launchpad locations.

Key steps included parsing API responses, normalizing nested data, and merging datasets using unique IDs (e.g., rocket_id, launchpad_id).



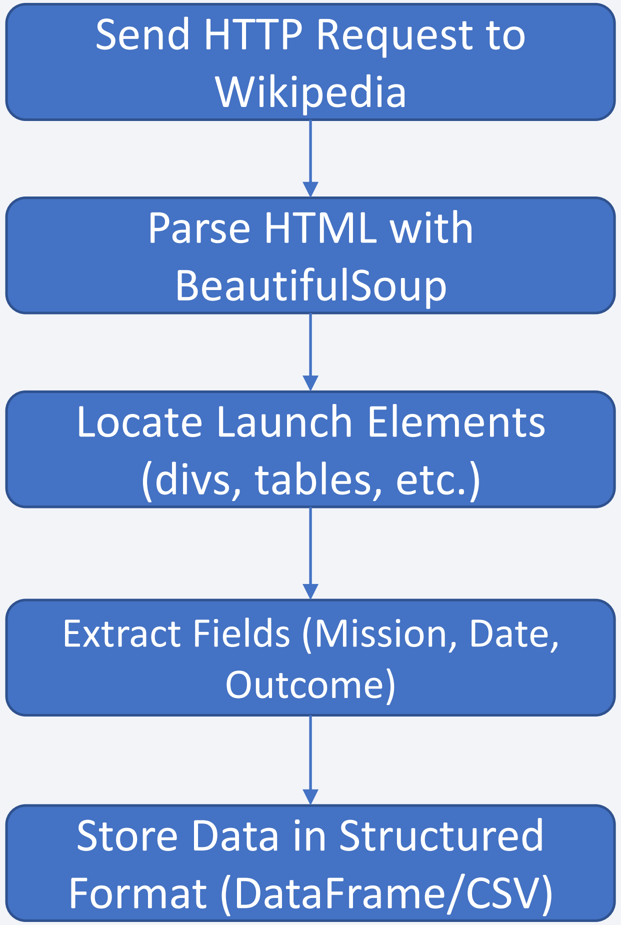
https://github.com/panderior/IBM-DS-Capstone-proj/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

We performed web scraping using Python's BeautifulSoup and requests libraries to extract SpaceX launch information from a Wikipedia page related to it.

The process involved sending HTTP GET requests, parsing the HTML structure, and extracting key data such as mission names, launch dates, and success status.

We then stored the extracted data into structured formats like Pandas DataFrames for further cleaning and integration. This approach was used where REST APIs lacked historical or formatted data.



https://github.com/panderior/IBM-DS-Capstone-proj/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

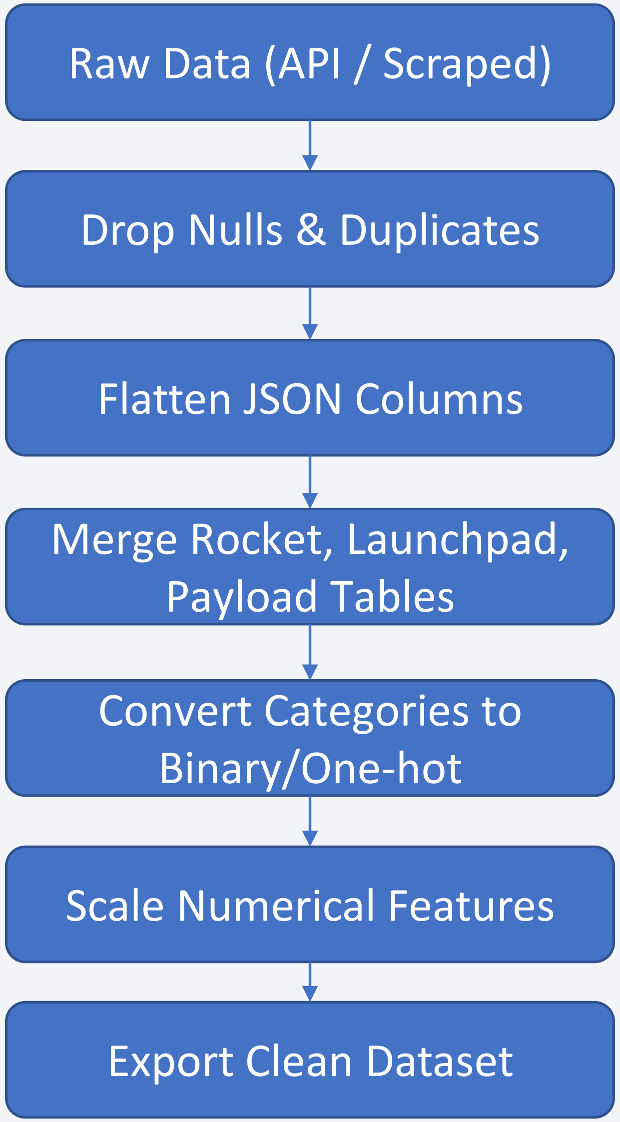
After collecting raw launch data, we performed comprehensive data wrangling to prepare it for analysis and modeling.

This involved removing null entries, normalizing nested JSON structures, and converting categorical values (e.g., landing outcomes) into binary labels.

We used Pandas for merging datasets by keys like launch_id, rocket_id, and payload_id. Irrelevant features were dropped, and numerical values like payload mass were scaled for model compatibility.

The final dataset was clean, structured, and ML-ready.

https://github.com/panderior/IBM-DS-Capstone-proj/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

We used visualizations to uncover patterns, trends, and relationships in the SpaceX launch dataset.

- Bar plots were used to compare success rates across launch sites, highlighting which locations were most reliable.
- Pie charts helped visualize the proportion of successful vs. failed landings.
- Scatter plots explored the correlation between payload mass and landing outcome, helping assess whether heavier payloads affect success.
- We also used histograms to examine the distribution of payloads and heatmaps to detect potential multicollinearity among numerical features.

These charts guided feature selection and model design by revealing influential variables and outliers.

https://github.com/panderior/IBM-DS-Capstone-proj/blob/main/edadataviz.ipynb

EDA with SQL

Summary of Performed Queries

- Selected key columns such as launch site, payload mass, orbit, and landing outcome from the launch dataset.
- Filtered successful Falcon 9 launches using WHERE clauses on booster_version and landing_success.
- Aggregated launch outcomes by site using GROUP BY to identify top-performing locations.
- Used ORDER BY with LIMIT to rank launch sites based on the number of successful landings.
- Applied JOIN queries to merge tables like launch data with payload and booster metadata.
- Queried records with maximum payload mass to explore correlation with landing success.
- Used nested subqueries to extract missions with extreme values (e.g., heaviest, lightest).
- Counted the number of landings per orbit type using GROUP BY orbit

https://github.com/panderior/IBM-DS-Capstone-proj/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

We used Folium to build an interactive map visualizing SpaceX launch sites and their geographic context. Markers were placed on each known launch site to show the location and name of the site.

We added circle markers around the launch points to represent operational zones, using different radii for visual emphasis. To analyze proximity, lines were drawn between launch sites and nearby infrastructure (e.g., coastlines or airports), and popups were included to show site details like launch success rate.

These map objects helped us understand the spatial distribution of launches and their potential impact zones, supporting site selection and risk assessment.

https://github.com/panderior/IBM-DS-Capstone-proj/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

We built an interactive dashboard using Plotly Dash to explore and visualize SpaceX launch data. The dashboard includes a dropdown menu for selecting a specific launch site and a range slider to filter payload mass.

Based on user selections, we dynamically update two visualizations:

- a pie chart showing launch success counts by site and
- a scatter plot displaying the relationship between payload mass and landing success, colored by booster version.

These components provide an intuitive interface for users to investigate patterns and draw insights about mission outcomes across different sites and payload sizes.

https://github.com/panderior/IBM-DS-Capstone-proj/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)

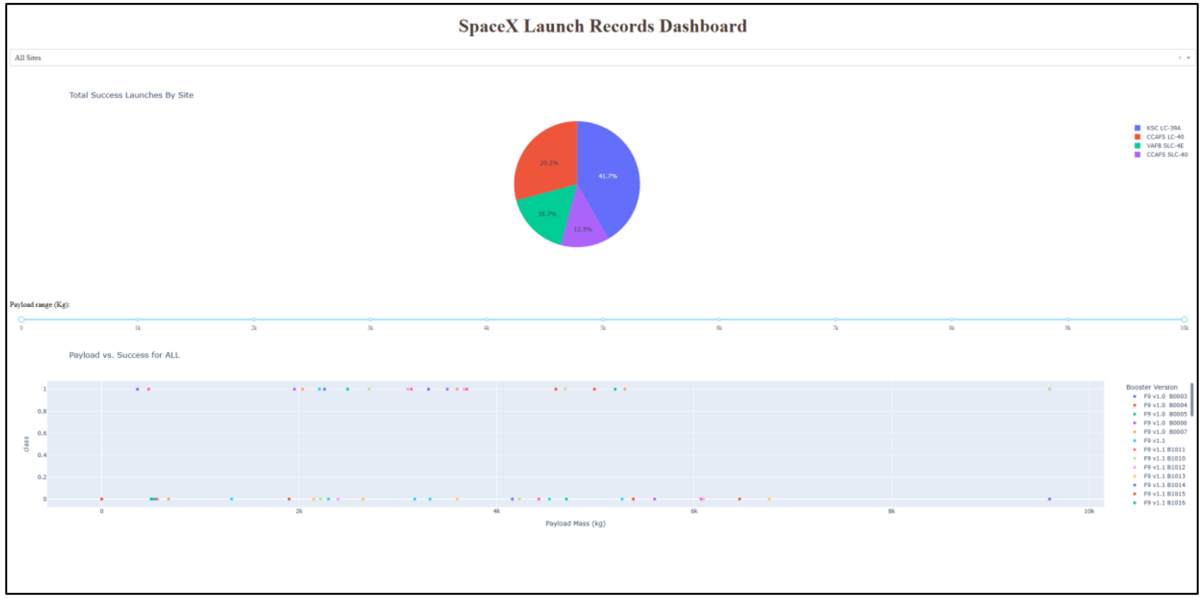
To predict whether a Falcon 9 first stage would successfully land, we built and compared multiple classification models. The process began with feature selection and data normalization, followed by splitting the dataset into training and testing sets.

We trained Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN) models.

Each model was evaluated using accuracy. To improve performance, we applied GridSearchCV for hyperparameter tuning. The best-performing model was chosen based on the highest test accuracy and balanced metrics across classes.

https://github.com/panderior/IBM-DS-Capstone-proj/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

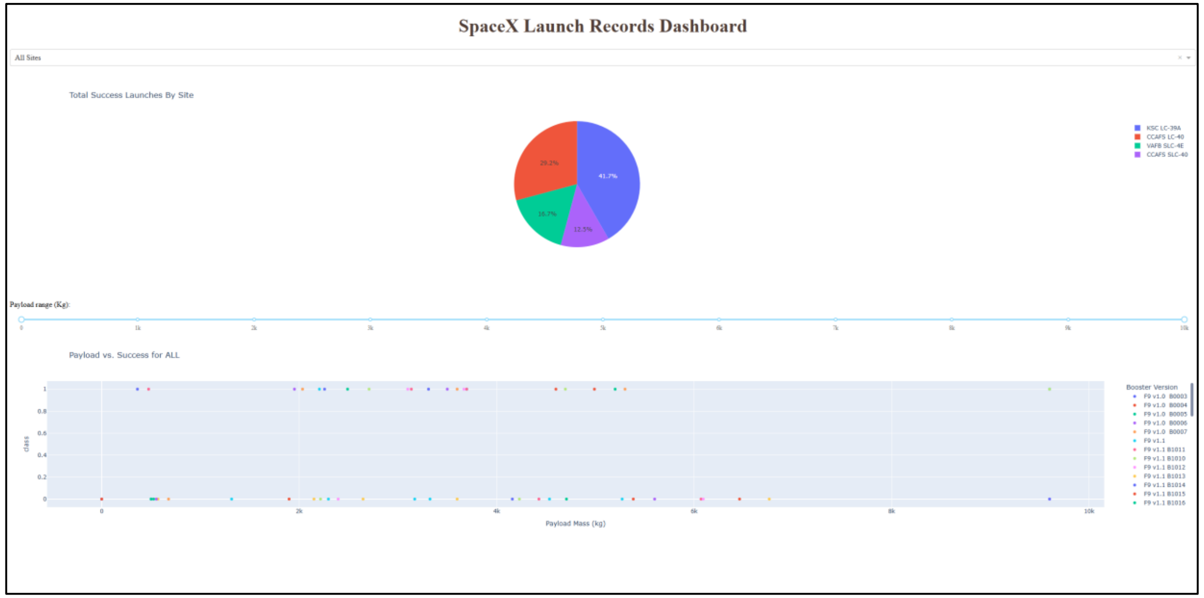
Results



Exploratory data analysis results

- EDA showed that successful landings increased over time, especially for later flight numbers with higher payloads. Launch sites like KSC LC-39A and CCAFS SLC-40 had higher success rates than VAFB SLC 4E.
- Missions to orbits like GEO and ES-L1 achieved consistently high success, while GTO and LEO had more variability. Successful landings were typically associated with mid-to-high payload ranges. A time trend revealed a sharp rise in success post-2014, reflecting improvements in rocket reuse and reliability.

Results



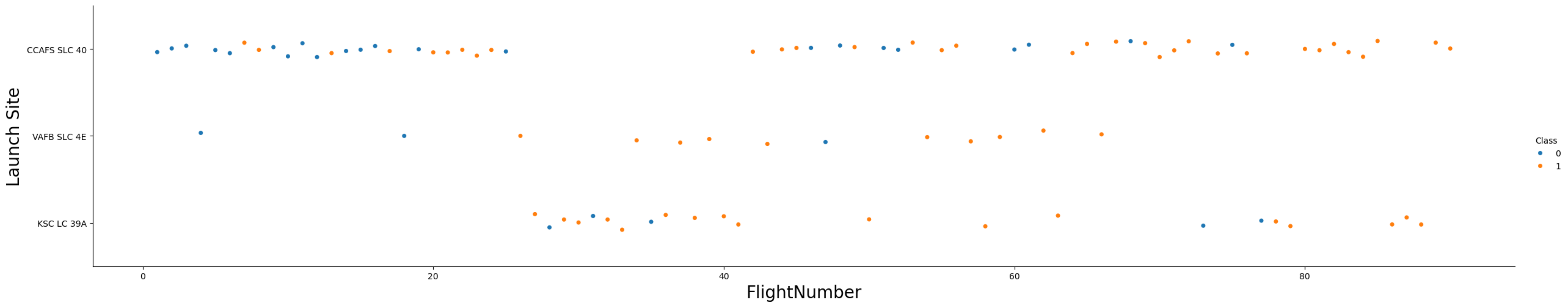
Predictive analysis results

- We trained and evaluated four classification models—Logistic Regression, SVM, Decision Tree, and KNN—to predict Falcon 9 first-stage landing success. Logistic Regression, Decision Tree, and KNN each achieved the highest accuracy of 94.4%, while SVM followed closely with 88.9%.
- Confusion matrices showed strong performance across models, with minimal false predictions. Logistic Regression, Decision Tree, and KNN correctly predicted all landing outcomes except for one false positive, making them the best performers in terms of both accuracy and reliability.

Section 2

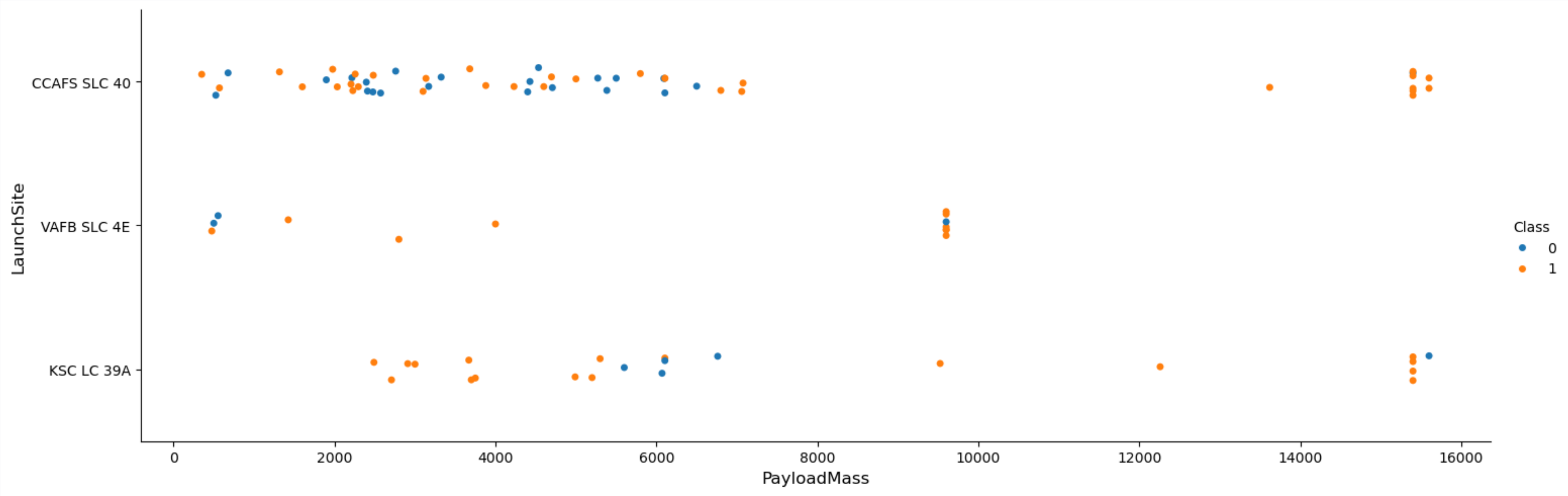
Insights drawn from EDA

Flight Number vs. Launch Site



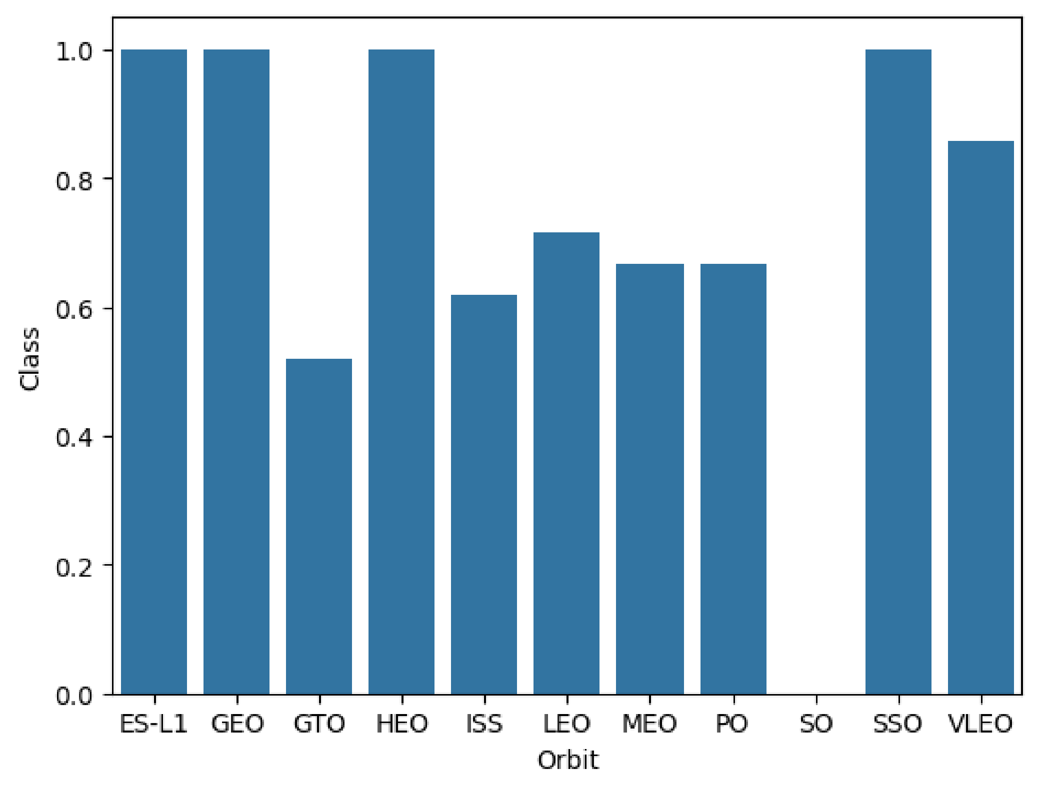
- This scatter plot shows the distribution of Falcon 9 launches across different **launch sites** over time (represented by flight number), with colors indicating **landing success (1)** or **failure (0)**.
- The plot reveals that **CCAFS SLC 40** and **KSC LC 39A** have had the most frequent launches and a **higher proportion of successful landings** in recent flights.
- In contrast, **VAFB SLC 4E** has fewer launches and more variability in outcome. Overall, the trend suggests **improved success rates** over time, especially at high-traffic sites.

Payload vs. Launch Site



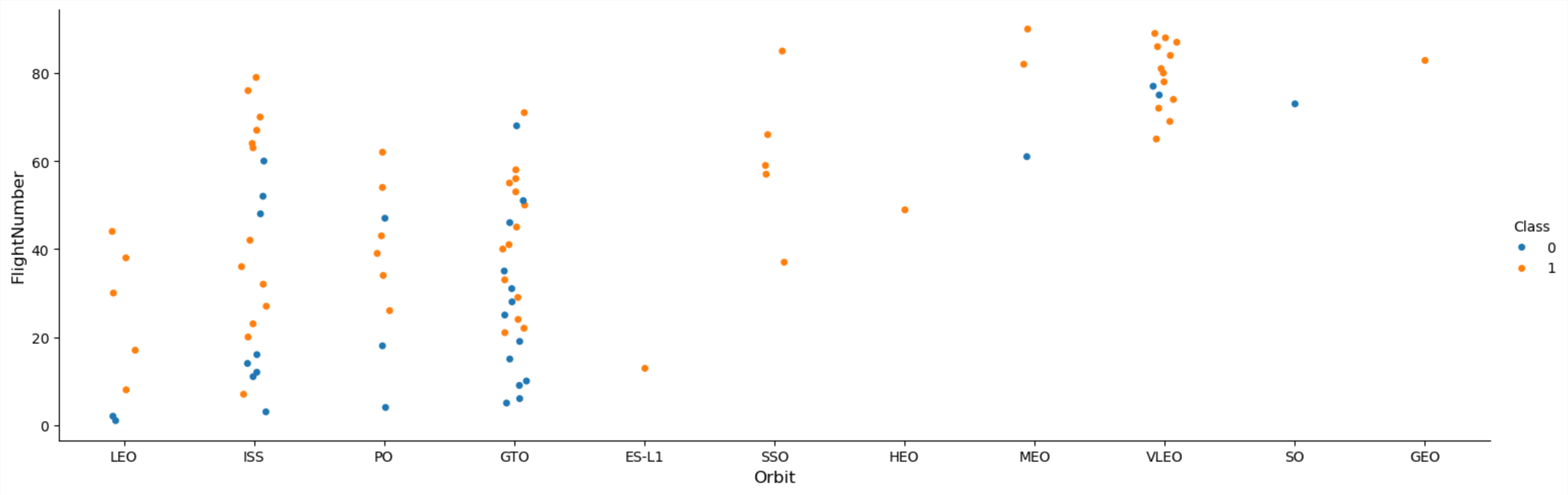
- This scatter plot explores the relationship between **launch sites**, **payload mass**, and **landing success**. Each point represents a launch, colored by outcome (1 = success, 0 = failure).
- The plot shows that **CCAFS SLC 40** handled a wide range of payloads and achieved many successful landings, even at higher masses. **KSC LC 39A** also demonstrated strong performance with heavier payloads. In contrast, **VAFB SLC 4E** had fewer launches and more scattered outcomes.
- Overall, higher payload missions were more common at major sites and still saw high success rates.

Success Rate vs. Orbit Type



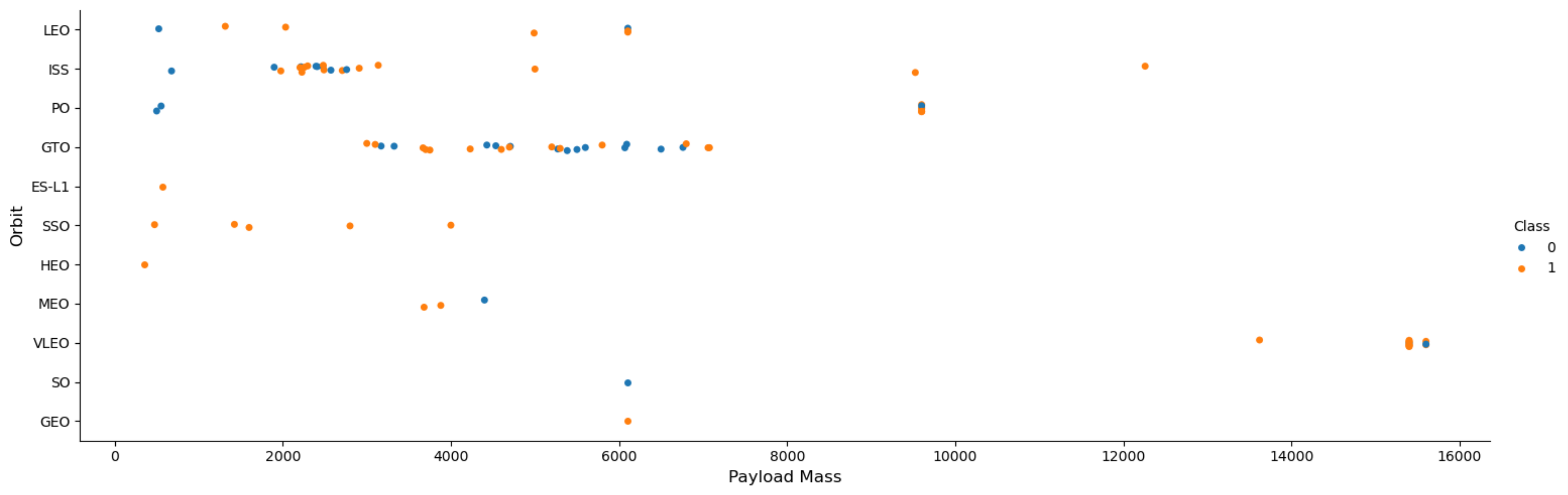
- This bar chart displays the average landing success rate across different **orbit types**. Missions targeting **ES-L1**, **GEO**, **HEO**, and **SSO** achieved a **100% success rate**, indicating strong reliability for these orbits.
- In contrast, missions to **GTO** had the **lowest success rate**, suggesting higher technical challenges or risk. Other orbits like **LEO**, **MEO**, and **VLEO** showed moderate but consistent performance.
- These insights help identify which mission profiles are more favorable for achieving successful landings.

Flight Number vs. Orbit Type



- This scatter plot illustrates the distribution of missions across **orbit types** over increasing **flight numbers**, with color indicating **landing success**.
- It shows that **more recent missions** (higher flight numbers) to orbits like **ISS**, **GTO**, **VLEO**, and **SSO** have seen **more consistent success** (orange). In contrast, **earlier missions** and some targeting **GTO** or **LEO** had a higher rate of failures (blue).
- The trend suggests that **operational improvements over time** have led to better success rates, especially for high-frequency orbit.

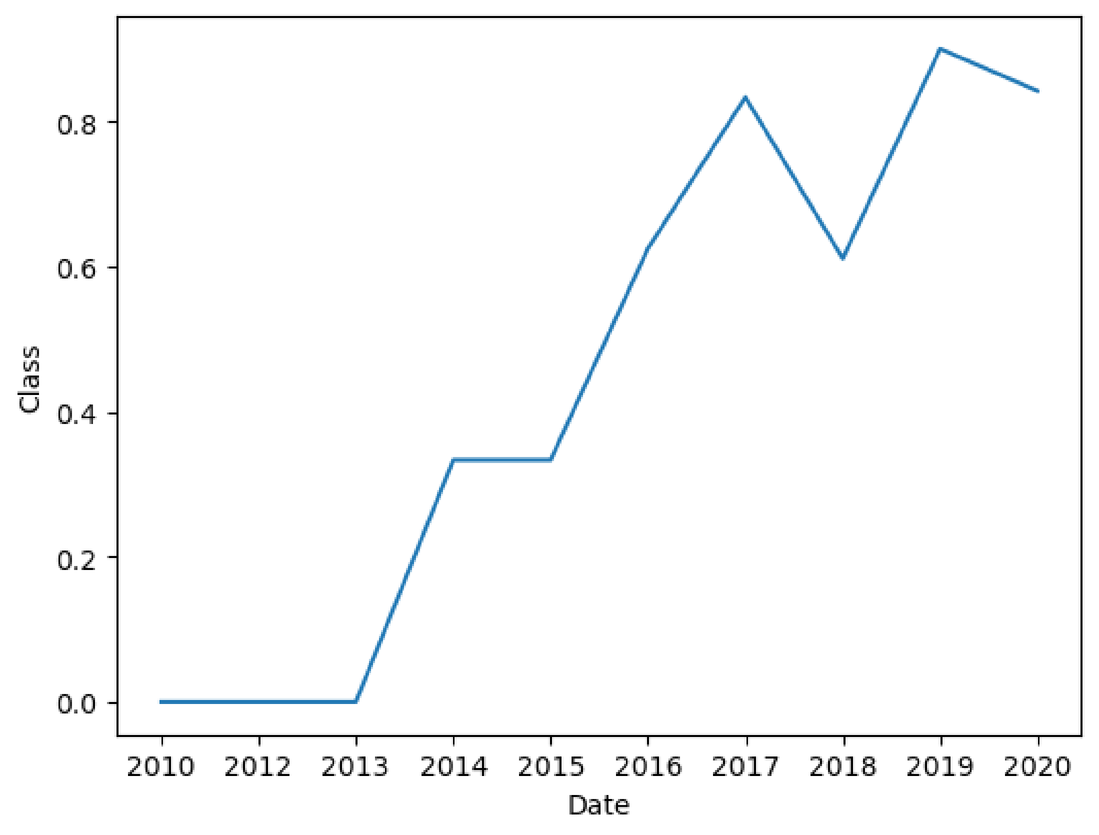
Payload vs. Orbit Type



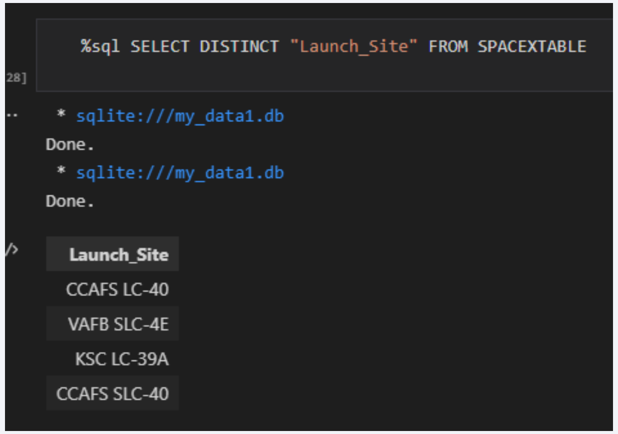
- This scatter plot shows the relationship between **payload mass**, **orbit type**, and **landing success**. Each dot represents a launch, color-coded by success (1 = orange, 0 = blue).
- The chart reveals that **GTO**, **ISS**, and **LEO** missions span a wide payload range with mixed outcomes, while orbits like **SSO**, **HEO**, and **VLEO** generally see high success rates regardless of payload.
- Notably, **heavier payloads (≥10,000 kg)** are mostly associated with successful landings, highlighting the capability of Falcon 9 to reliably deliver larger payloads to specific orbits.

Launch Success Yearly Trend

- The line chart illustrates a clear upward trend in **Falcon 9 landing success rates over time**.
- From 2010 to 2013, success was minimal, but starting in **2014**, there was a steady improvement, reaching over **90% success by 2019**.
- This trend reflects **technological advancements and operational refinements** in SpaceX’s reusable launch systems, highlighting their growing reliability and cost-efficiency year over year.



All Launch Site Names



- Using an SQL query on the SpaceX dataset, we identified four distinct launch sites:
  - **CCAFS LC-40**
  - **VAFB SLC-4E**
  - **KSC LC-39A**
  - **CCAFS SLC-40**
- These sites represent the geographical locations used for SpaceX launches, which are crucial for analyzing launch outcomes, regional performance, and site-specific trends in mission success.

Launch Site Names Begin with 'CCA'

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

Python

```
 * 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

A filtered SQL query retrieved the first 5 launches from sites beginning with CCA, specifically CCAFS LC-40. All five missions were successful, showcasing early operational reliability. However, landing outcomes varied — two had parachute failures and the rest recorded no landing attempts. Payloads were primarily for NASA (COTS/CRS) missions, mostly to LEO (ISS) orbits, indicating SpaceX's early focus on cargo delivery for government contracts.

Total Payload Mass

```
[25]  %sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer"=="NASA (CRS)"

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

SUM("PAYLOAD_MASS__KG_")
45596

An SQL SUM on the PAYLOAD_MASS__KG_ for all records where Customer = 'NASA (CRS)' returns 45,596 kg.

This figure represents the cumulative mass of cargo that SpaceX has delivered under NASA contracts.

Average Payload Mass by F9 v1.1

```
[27]  %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

An SQL AVG query on the PAYLOAD_MASS__KG_ for records where Booster_Version = 'F9 v1.1' yields 2,928.4 kg.

This value represents the typical payload capacity carried by the F9 v1.1 boosters and serves as a benchmark for comparing performance across different booster versions.

First Successful Ground Landing Date

```
[32]  %sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" == "Success (ground pad)" AND "Mission_Outcome" == "Success"

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

MIN("Date")
2015-12-22

- An SQL MIN(Date) query filtered for records where both Mission_Outcome = 'Success' and Landing_Outcome = 'Success (ground pad)' returns 2015-12-22.
- This date marks SpaceX's first successful return and landing of the Falcon 9 first stage on a ground pad.

Successful Drone Ship Landing with Payload between 4000 and 6000

[33]

```
%sql SELECT * FROM SPACEXTABLE WHERE "Landing_Outcome" == "Success (drone ship)"AND "Mission_Outcome" == "Success" AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Four boosters—**F9 FT B1022**, **B1026**, **B1021.2**, and **B1031.2**—achieved successful drone ship landings while carrying payloads between 4000 and 6000 kg.

This demonstrates SpaceX's capability to land boosters reliably under moderate payload stress during GTO missions.

Total Number of Successful and Failure Mission Outcomes

[37]

```
%sql SELECT "Mission_Outcome", COUNT(*) FROM SPACEXTABLE GROUP BY "Mission_Outcome"
```

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

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

We analyzed the mission outcomes of SpaceX launches. The data shows that:

- **98 missions were successful**
- **1 mission failed in flight**
- **2 additional entries were labeled as “Success” and “Success (payload status unclear)”**

This indicates an **overwhelming success rate**, highlighting the operational reliability of SpaceX launch missions, with only one recorded in-flight failure.

Boosters Carried Maximum Payload

```
[42]
%sql SELECT * FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_"==(SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE)
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2019-11-11	14:56:00	F9 B5 B1048.4	CCAFS SLC-40	Starlink 1 v1.0, SpaceX CRS-19	15600	LEO	SpaceX	Success	Success
2020-01-07	2:33:00	F9 B5 B1049.4	CCAFS SLC-40	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600	LEO	SpaceX	Success	Success
2020-01-29	14:07:00	F9 B5 B1051.3	CCAFS SLC-40	Starlink 3 v1.0, Starlink 4 v1.0	15600	LEO	SpaceX	Success	Success
2020-02-17	15:05:00	F9 B5 B1056.4	CCAFS SLC-40	Starlink 4 v1.0, SpaceX CRS-20	15600	LEO	SpaceX	Success	Failure
2020-03-18	12:16:00	F9 B5 B1048.5	KSC LC-39A	Starlink 5 v1.0, Starlink 6 v1.0	15600	LEO	SpaceX	Success	Failure
2020-04-22	19:30:00	F9 B5 B1051.4	KSC LC-39A	Starlink 6 v1.0, Crew Dragon Demo-2	15600	LEO	SpaceX	Success	Success
2020-06-04	1:25:00	F9 B5 B1049.5	CCAFS SLC-40	Starlink 7 v1.0, Starlink 8 v1.0	15600	LEO	SpaceX, Planet Labs	Success	Success
2020-09-03	12:46:14	F9 B5 B1060.2	KSC LC-39A	Starlink 11 v1.0, Starlink 12 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-06	11:29:34	F9 B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-18	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-24	15:31:34	F9 B5 B1060.3	CCAFS SLC-40	Starlink 14 v1.0, GPS III-04	15600	LEO	SpaceX	Success	Success
2020-11-25	2:13:00	F9 B5 B1049.7	CCAFS SLC-40	Starlink 15 v1.0, SpaceX CRS-21	15600	LEO	SpaceX	Success	Success

- All listed boosters carried the maximum payload of **15,600 kg**. These missions were primarily Starlink launches and were conducted by **F9 B5** boosters across various block versions (e.g., B1048.4, B1049.4, etc.). Most missions were successful in both delivery and landing.
- The **F9 B5** series consistently supports the heaviest payloads, showcasing its significance in SpaceX's heavy-lift operations.

2015 Launch Records

```
[49] %sql SELECT substr(Date, 6,2) AS "Month", Booster_Version, Launch_Site, Landing_Outcome  FROM SPACEXTABLE WHERE substr(Date,0,5)='2015' AND Landing_Outcome LIKE "Failure (drone ship)"
                                                                                                        Python
 * sqlite:///my_data1.db
Done.
```

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

In 2015, two missions experienced **drone ship landing failures**.
Both launches occurred at **CCAFS LC-40**, using booster versions:

- **F9 v1.1 B1012** (January)
- **F9 v1.1 B1015** (April)

This insight highlights early-stage recovery challenges SpaceX faced with drone ship landings during that year.

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

[56]
```
%sql SELECT Landing_Outcome, COUNT(Landing_Outcome) 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

- The most common result was “No attempt” (10 times), indicating early missions lacked recovery attempts.
- “Success (drone ship)” and “Failure (drone ship)” were tied with 5 occurrences each, showing the evolving but still maturing recovery efforts on sea platforms.
- Other outcomes included ground pad successes (3), controlled/uncontrolled ocean landings, and parachute failures, each occurring 1–3 times.

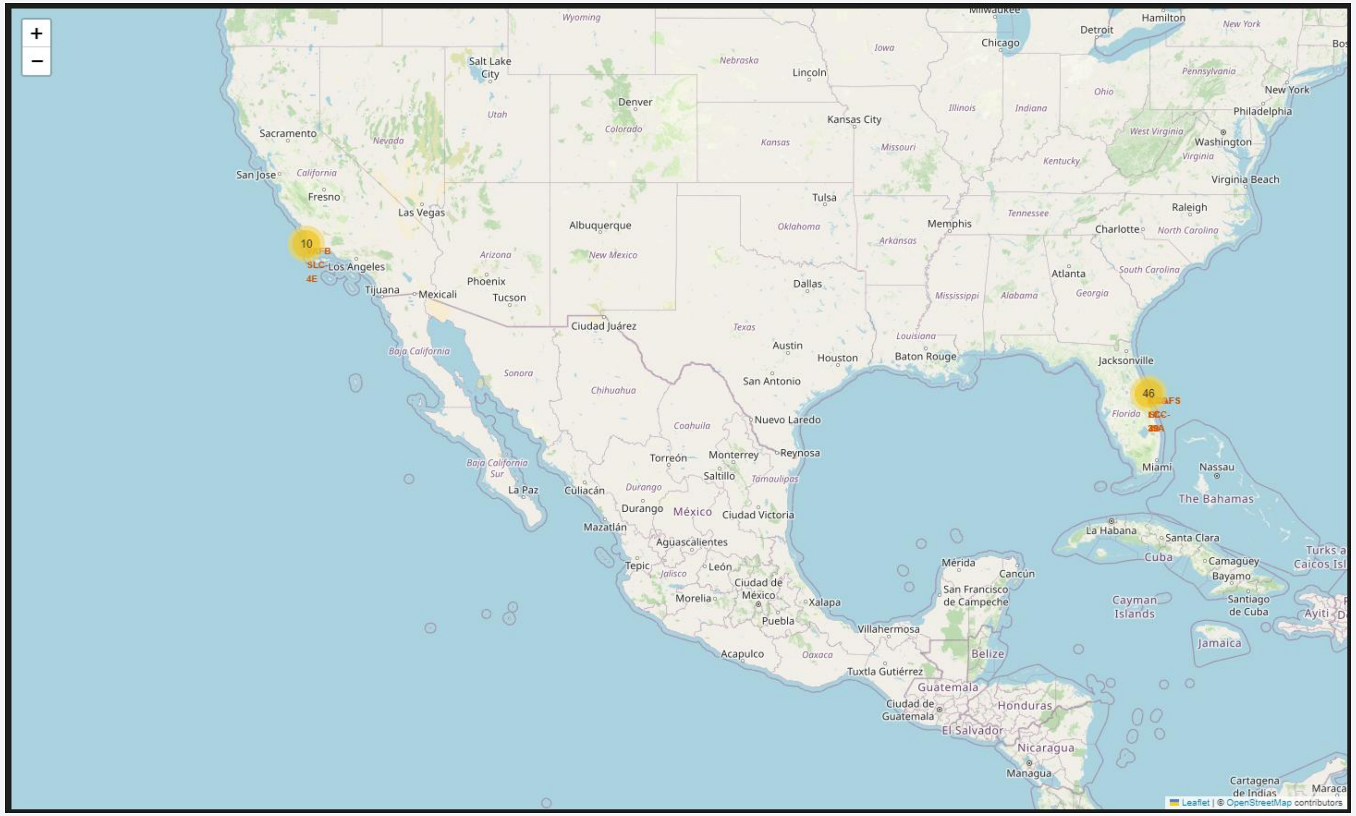
Recovery reliability improved over time, but in this period, drone ship landings had mixed outcomes, and recovery was not attempted in several early missions.

Section 3

Launch Sites Proximities Analysis

Launch Site Distribution Overview

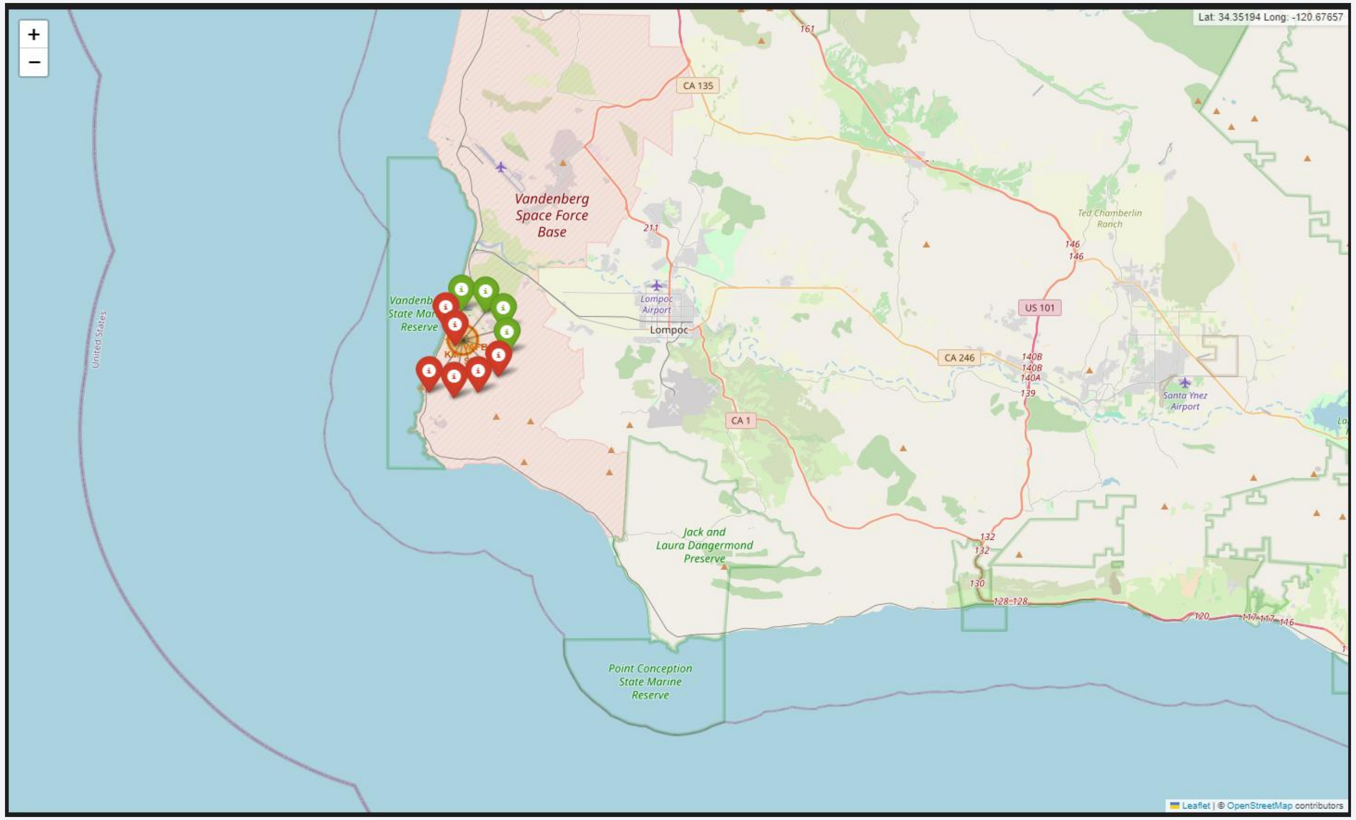
- The folium map displays the geographical distribution of SpaceX launch sites across the U.S. Each marker on the map represents a launch site, clustered to show density and labeled for identification.
- High concentration of launches from Florida (marked with “46”), specifically around the Cape Canaveral and Kennedy Space Center areas.
- Secondary cluster in California (marked with “10”), likely representing Vandenberg Air Force Base.
- Marker clustering helps distinguish frequently used sites from less active ones.



Launch Outcomes at Vandenberg Space Force Base (VAFB)

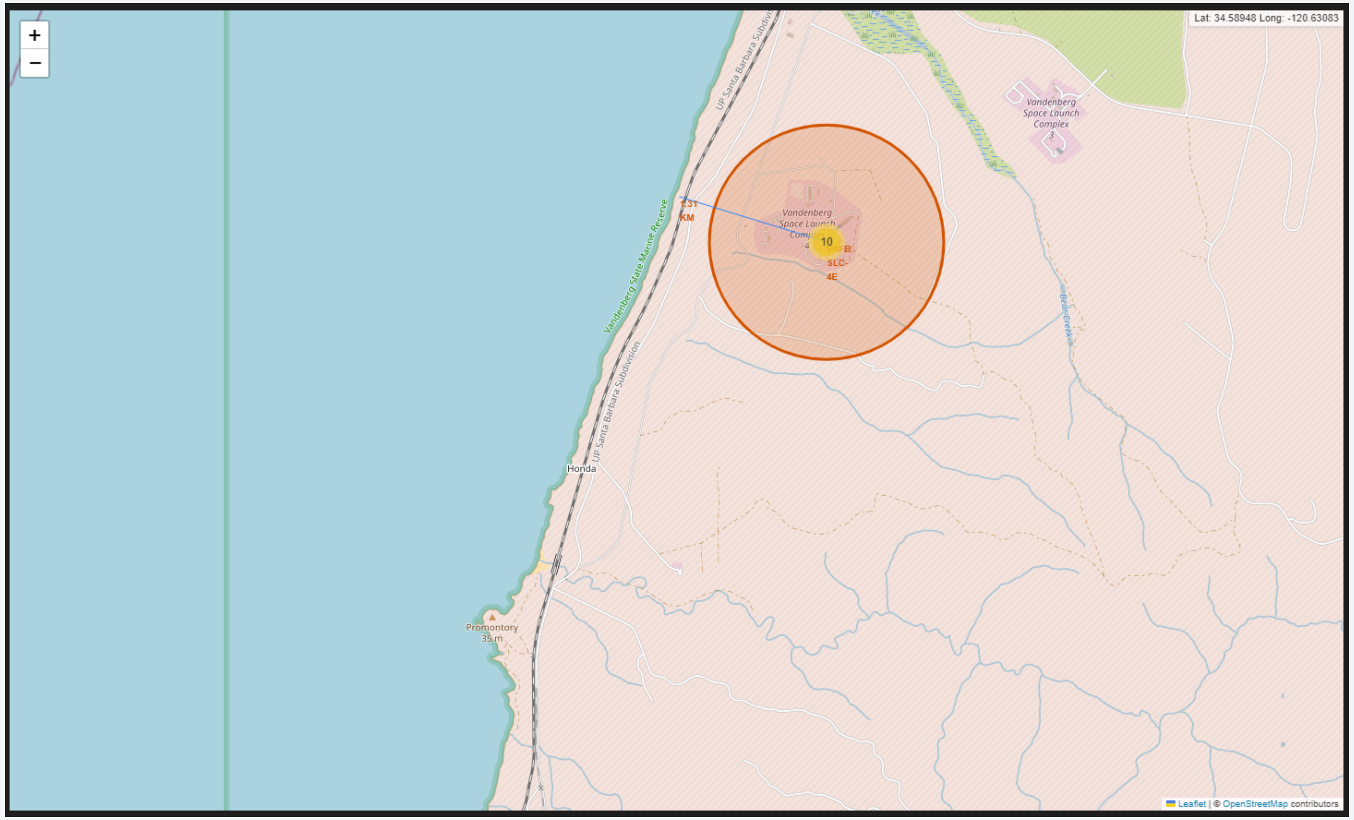
- The map highlights the launch activities concentrated around VAFB in California.

- Green markers represent successful landings, while red markers indicate failures.



Launch Site and Nearby Coastline Proximity

- The Vandenberg launch site is strategically located near the coast (~1.3 km), which is beneficial for over-ocean launches (minimizing risk to populated areas).
- The map demonstrates the site's accessibility and suitability for orbital launches toward the Pacific.
- This proximity analysis helps in assessing safety zones, potential recovery areas, and environmental considerations.

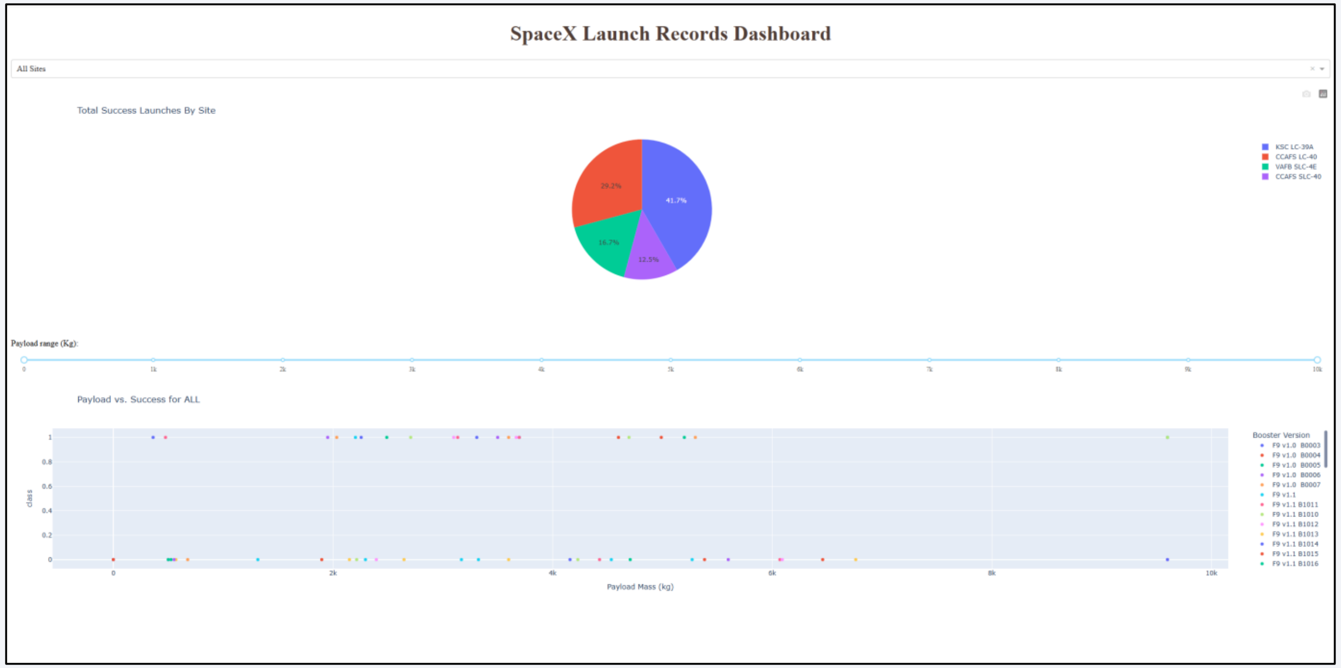


Section 4

Build a Dashboard with Plotly Dash

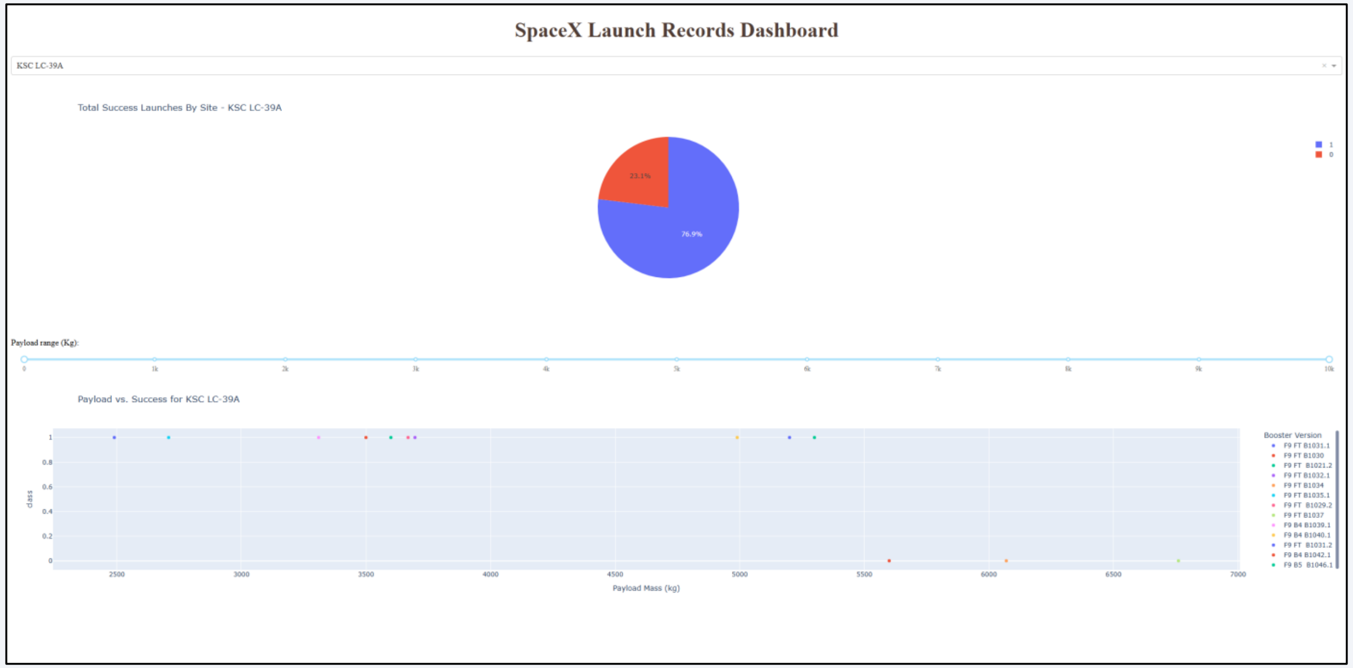
Plotly Dashboard

- Interactive Plotly Dashboard provides dynamic filtering of launch success data by site and payload range, enhancing exploratory data analysis.
- Combines pie chart and scatter plot views to visually correlate launch site performance and payload success patterns, making trends easier to interpret.
- KSC LC-39A accounts for the highest share (41.7%) of successful launches, followed by CCAFS LC-40 (29.2%), indicating strong performance at these sites.
- Scatter plot shows most payloads (0–10,000 kg) tend to succeed, with some failure variation across different booster versions.



Dashboard: Launch Site with Highest Success Ratio (KSC LC-39A)

- The interactive dashboard filters SpaceX data by launch site and payload range, providing instant insights via pie and scatter plots.
- For KSC LC-39A, the pie chart shows a ~77% success rate, confirming it as the site with the highest launch success ratio.

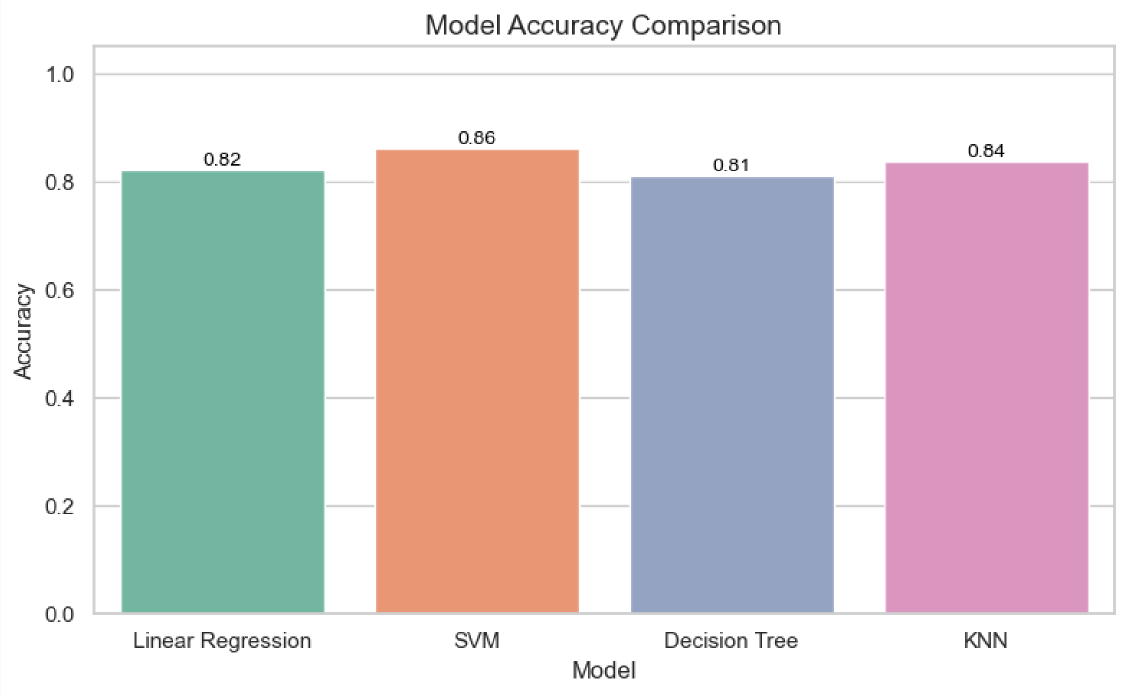


Section 5

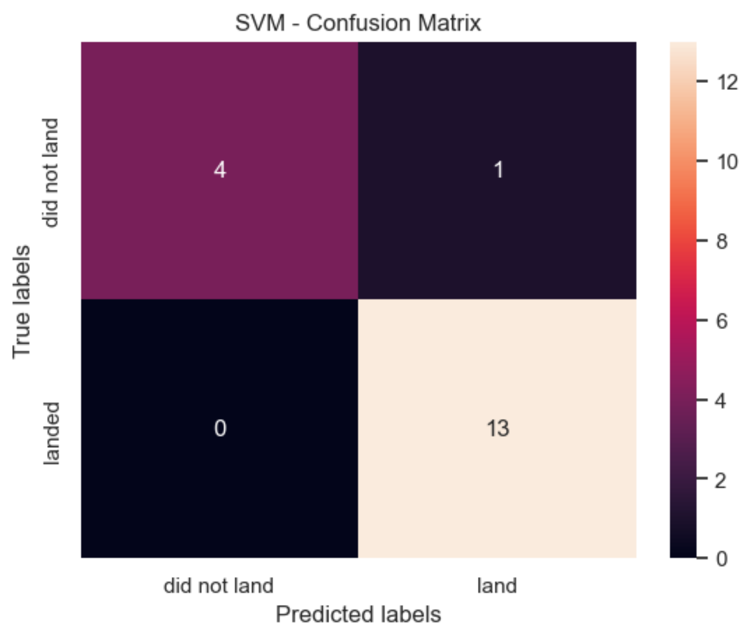
Predictive Analysis (Classification)

Classification Accuracy

- A bar chart compares the accuracy of four classification models: Linear Regression, SVM, Decision Tree, and KNN.
- Among them, SVM achieved the highest classification accuracy of 0.86, making it the most effective model for this dataset.



Confusion Matrix



- The SVM model, which achieved the highest accuracy (0.86), shows strong classification performance with only one misclassified case.
- The confusion matrix indicates 13 correct land predictions and 4 correct 'did not land' predictions, suggesting high precision and recall for both classes.

Conclusions

- Data scraping and wrangling involved cleaning SpaceX launch records, extracting launch outcome details, and harmonizing payload data for effective analysis.
- EDA revealed key insights such as booster version F9 v1.1 frequently carrying payloads above 4000 kg and NASA contributing significantly to total payload mass.
- Launch outcome analysis showed increasing landing success over time, with most failures concentrated before 2015 and successes dominated by ground pad and drone ship recoveries.
- Among all ML models, SVM achieved the highest classification accuracy (86%), making it the best performer for predicting launch success.
- Folium maps visualized global launch site locations and their proximities to coastlines and infrastructure, aiding logistical insights.
- The Plotly dashboard allowed interactive exploration of launch outcomes by site and payload, making trends and patterns easily understandable.

Thank you!

https://github.com/panderior/IBM-DS-Capstone-proj