IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Srinath Yellepeddi
December 10th, 2025

Outline

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

Executive Summary

- **Summary of methodologies:**
  - Data Collection using SpaceX REST API and Web Scraping.
  - Data Wrangling using Pandas and Numpy.
  - Exploratory Data Analysis (EDA) using SQL queries and data visualization techniques.
  - Interactive Visual Analytics and Dashboards using Folium and Plotly Dash.
  - Machine Learning Prediction using classification models.
- **Summary of all results:**
  - Data was successfully collected from sources.
  - Final processed dataset provided a clean, structured foundation for effective EDA and machine learning.
  - Exploratory Data Analysis (EDA) identified the most influential features for predicting launch success.
  - Interactive maps and visual dashboards were successfully created using Folium and Plotly Dash.
  - Model evaluation confirmed the machine learning algorithm that delivered the strongest predictive accuracy on the test data.

Introduction

- In this capstone project, I take on the role of a data scientist for a new rocket company, Space Y, aiming to compete with SpaceX.
- My responsibilities include estimating launch costs by collecting and analyzing SpaceX data and developing dashboards to present key insights.
- I also build a machine learning model using publicly available SpaceX data to predict whether SpaceX will successfully reuse the first-stage booster for each launch.

Section 1

Methodology

Methodology

Executive Summary

- Data Collection Methodology:
  - Data was collected from the SpaceX REST API and supplemented with launch details scraped from the public Wikipedia page (List of Falcon 9 and Falcon Heavy launches).
- Performed Data Wrangling:
  - Data was cleaned, merged, and transformed using Pandas and NumPy to handle missing values and prepare a structured dataset for analysis.
- Performed exploratory data analysis (EDA) using visualization and SQL.
- Performed interactive visual analytics using Folium and Plotly Dash.
- Performed predictive analysis using classification models.
  - Classification models were trained, hyperparameters tuned via GridSearchCV, and evaluated on test data to predict first-stage landing success.

Data Collection

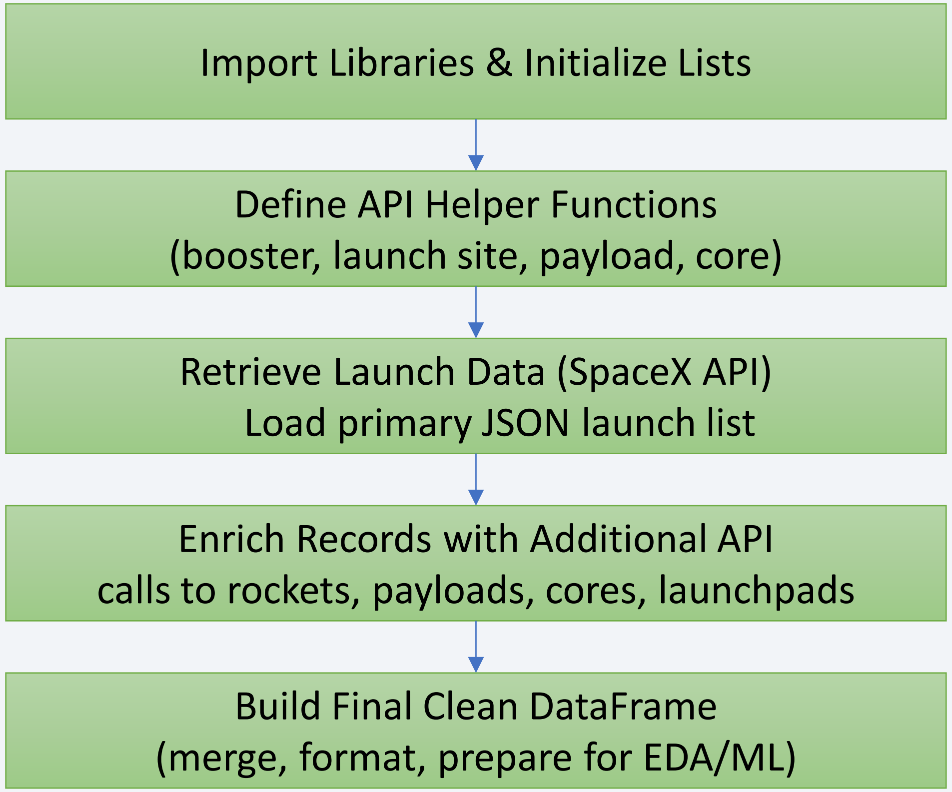
- **SpaceX REST API:** Retrieved structured JSON datasets for launches, rockets, payloads, cores, and landing outcomes using the official SpaceX API.
- **Web Scraping:** Extracted supplementary launch and booster details from public SpaceX/Wikipedia pages using Python tools such as BeautifulSoup.
- **Data Integration:** Merged API and scraped datasets, standardized feature formats (numeric fields, categories, timestamps), and resolved inconsistencies.
- **Data Storage:** Loaded all processed data into Pandas DataFrames for downstream wrangling, visualization, and machine learning tasks.

Data Collection – SpaceX API

Data was retrieved through REST API calls to the SpaceX API by following below data collection process, as outlined in the flowchart:

- Initialized environment by importing required Python libraries and setting up empty lists to store launch-related data.
- Created reusable API helper functions to retrieve booster, launch site, payload, and core details for each launch record.
- Fetched the primary launch dataset from the SpaceX REST API, obtaining the full JSON list of Falcon launches.
- Enriched each launch entry by making additional API calls to gather detailed information on rockets, payloads, cores, and launchpads.
- Merged and formatted all retrieved data into a unified Pandas DataFrame with consistent structure and clean field types.
- Prepared the final dataset for downstream tasks such as data wrangling, EDA, visual analytics, and machine learning.

- [GitHub URL of the completed SpaceX API calls Notebook]

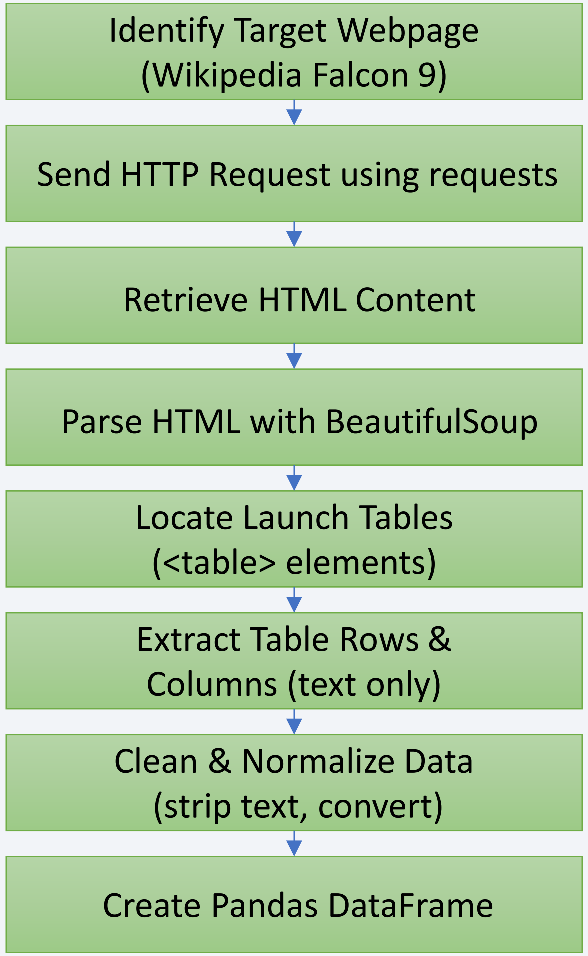


Data Collection – Web Scraping

Followed below Web Scraping Process as shown in flow chart:

- **Identify target webpage**: Selected the Wikipedia page containing Falcon 9 & Falcon Heavy launch tables.
- **Send HTTP request**: Retrieved the page's HTML content using requests.
- **Parse HTML structure**: Used **BeautifulSoup** to navigate the DOM and locate launch tables.
- **Extract relevant data**: Parsed table rows and cells to capture mission dates, flight numbers, outcomes, booster versions, landing results, etc.
- **Clean raw text**: Removed HTML tags, standardized text, and fixed formatting inconsistencies.
- **Build structured dataset**: Converted extracted lists into a clean **Pandas DataFrame** ready for merging with API data.

- GitHub URL of the completed Web Scraping Notebook

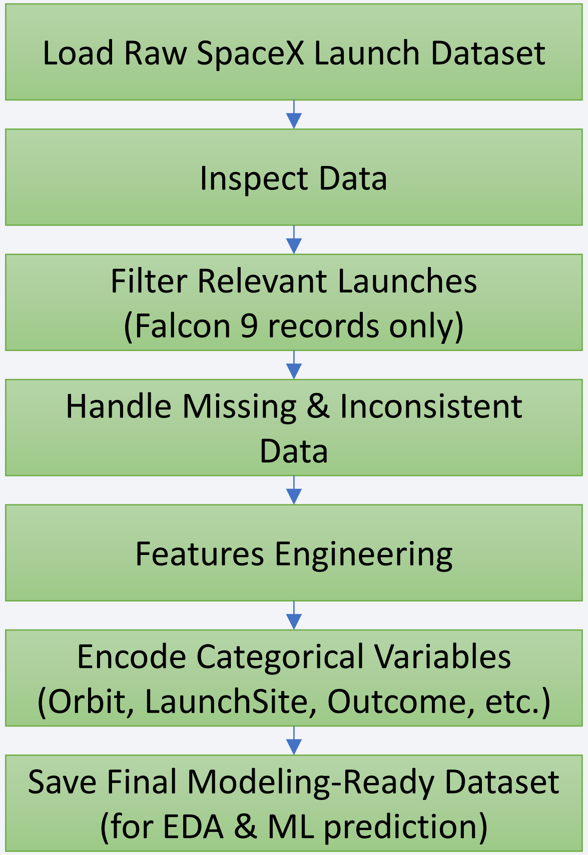


Data Wrangling

Following is the Data Wrangling process as outlined in flowchart:

- Loaded SpaceX launch data into Pandas DataFrames.
- Cleaned inconsistent values and standardized column formats.
- Removed unneeded columns and filtered only Falcon 9 launches.
- Handled missing values using imputation and logical rules.
- Extracted new engineered features (e.g., Landing Outcome, Launch Site).
- Converted categorical variables into numerical encodings for modeling.
- Merged API and scraped datasets into one unified analytical dataset.

- [GitHub URL of completed Data Wrangling Notebook]()



EDA with Data Visualization

- **Following charts were plotted:**
  - **Catplot** to analyze the relationship between Flight Number and Payload Mass.
  - **Catplot** to visualize the relationship between Flight Number and Launch Site.
  - **Catplot** to examine the relationship between Payload Mass and Launch Site.
  - **Bar chart** to compare launch success rates across different orbit types.
  - **Catplot** to analyze the relationship between Flight Number and Orbit Type.
  - **Catplot** to examine the relationship between Payload Mass and Orbit Type.
  - **Line chart** to visualize the yearly trend of launch success.
- GitHub URL of completed EDA with Data Visualization Notebook

EDA with SQL

- **Key SQL analyses performed:**
    1. Retrieved all unique launch sites used in SpaceX missions.
    2. Filtered the first 5 launches where the launch site started with 'CCA’.
    3. Calculated the total payload mass carried by boosters for NASA CRS missions.
    4. Calculated the average payload mass carried by booster F9 v1.1.
    5. Found the earliest date of a successful ground pad landing.
    6. Listed boosters that successfully landed on drone ships with payload mass between 4000–6000 kg.
    7. Computed total counts of successful and failed missions.
    8. Found booster versions that carried the highest recorded payload mass.
    9. Displayed 2015 drone-ship failure missions with month name, booster version, and launch site.
    10. Ranked landing outcomes between 2010-06-04 and 2017-03-20 in descending order of occurrence.
- GitHub URL of completed EDA with SQL Notebook

Build an Interactive Map with Folium

- **Following map objects were created and added to the folium map:**

  - **Marker Clusters** to group nearby launch site markers and reduce map clutter.
  - **Markers** to display individual SpaceX launch site locations.
  - **Circle Markers** to visualize launch success vs. failure at each site.
  - **Distance Lines** to show proximity between launch sites and nearby infrastructure.

- GitHub URL of completed interactive map with Folium

Build a Dashboard with Plotly Dash

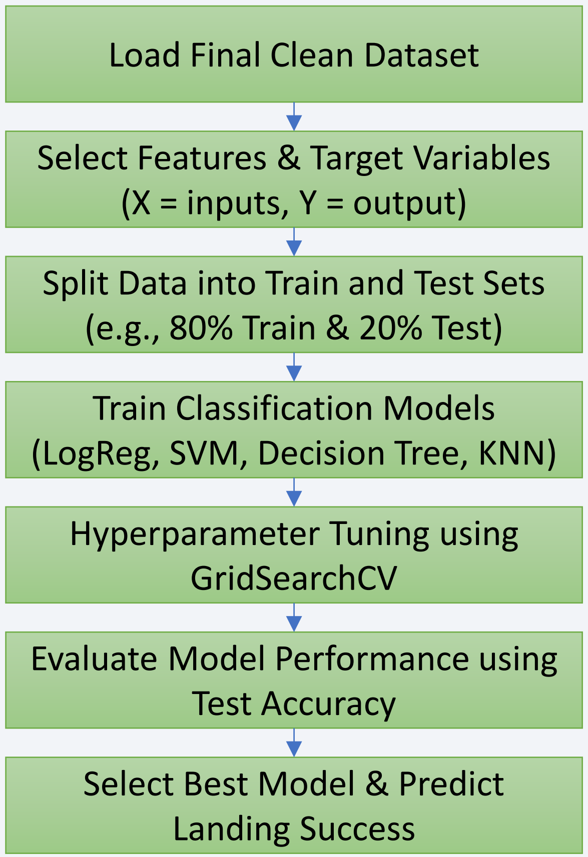
- **Following are plots, graphs and interactions that were added to the Plotly dashboard:**

  - Dropdown menu to select launch site.
  - Pie chart to examine success rate of each launch site.
  - Slider to select payload range.
  - Scatter Plot to display relationship between launch site, payload, and booster version.

- GitHub URL of completed Plotly Dash Lab

Predictive Analysis (Classification)

- Prepared the final cleaned dataset for machine learning by selecting relevant features and the target variable.
- Split the data into training and testing sets to evaluate model performance.
- Logistic Regression, SVM, Decision Tree, and KNN models were trained on the prepared dataset.
- Used **GridSearchCV** for hyperparameter tuning and cross-validation.
- Evaluated all four models using test accuracy and selected the best-performing model for final prediction.
- [GitHub URL of completed predictive analysis lab]

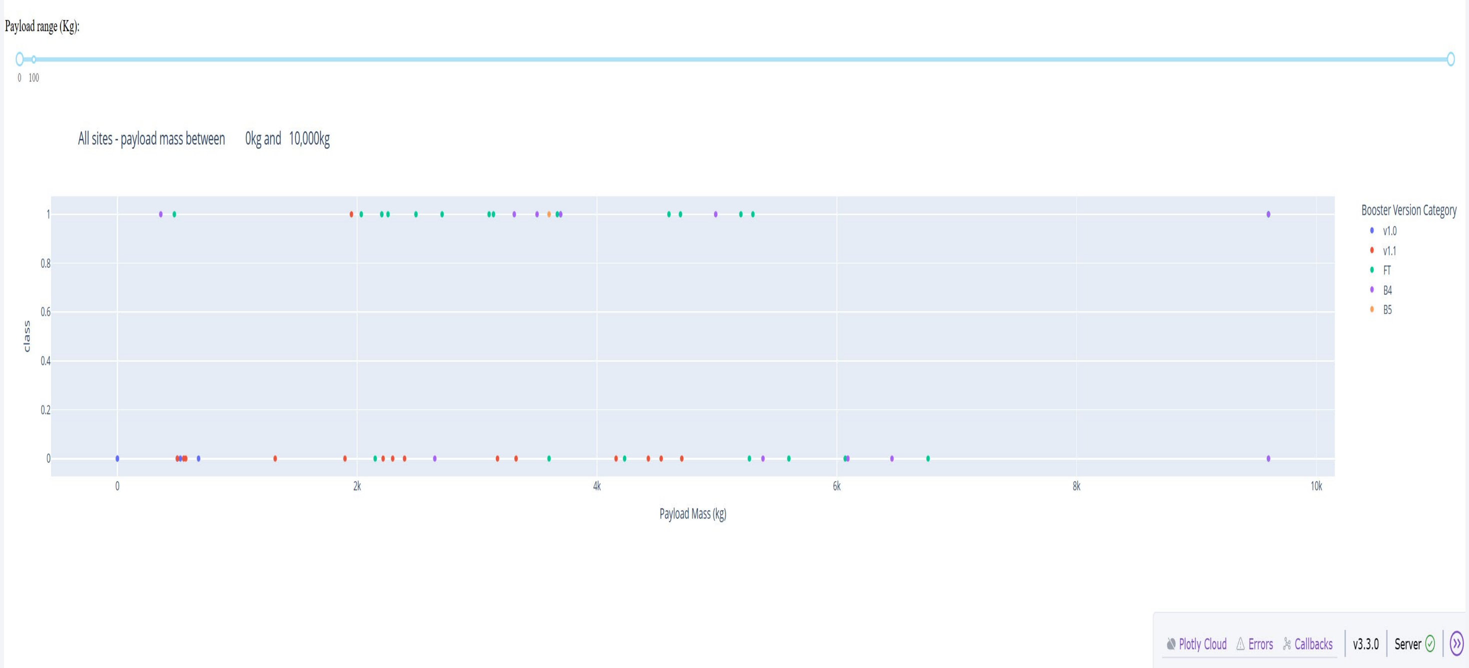


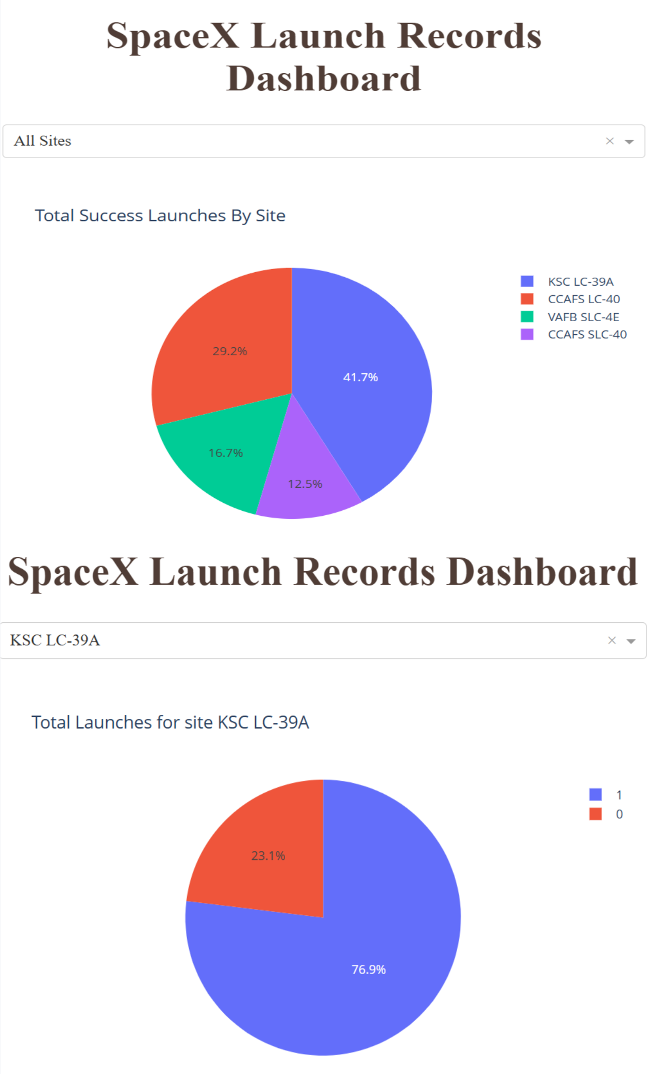
Results

- **Exploratory data analysis results:**
  - Launch success rate shows a clear upward trend with increasing flight number, indicating improved reliability over time.
  - Lower to mid-range payload masses showed higher success probability, while very heavy payloads exhibited mixed outcomes.
  - **KSC LC-39A** and **CCAFS SLC-40** emerged as the **most successful launch sites** based on both visual and SQL analysis.
  - **ES-L1, GEO, HEO,** and **SSO** orbits showed the **highest launch success rates**, while GTO had the lowest success rate.
  - Later Falcon 9 booster versions showed significantly better landing performance, confirming the impact of reusability improvements.
  - Drone ship landing success increased over time, especially after early operational failures.
  - SQL analysis confirmed that successful missions outnumbered failures overall, validating the trend observed in visual EDA.
  - The highest payload missions were carried by specific booster versions, as identified using SQL aggregation.

Results

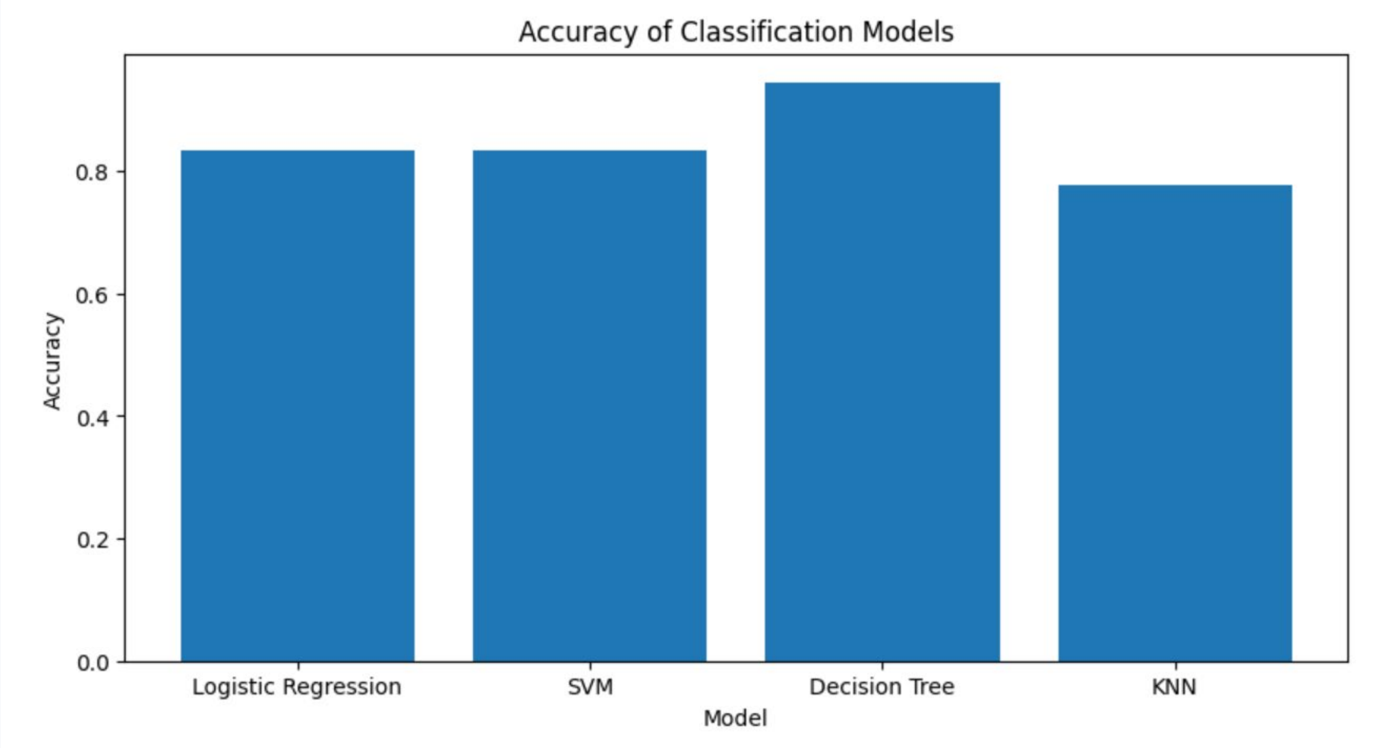
- Interactive analytics demo screenshots:





Results

- **The Decision Tree classifier** outperformed all other models, **achieving over 94% accuracy** in predicting successful landings on test data.

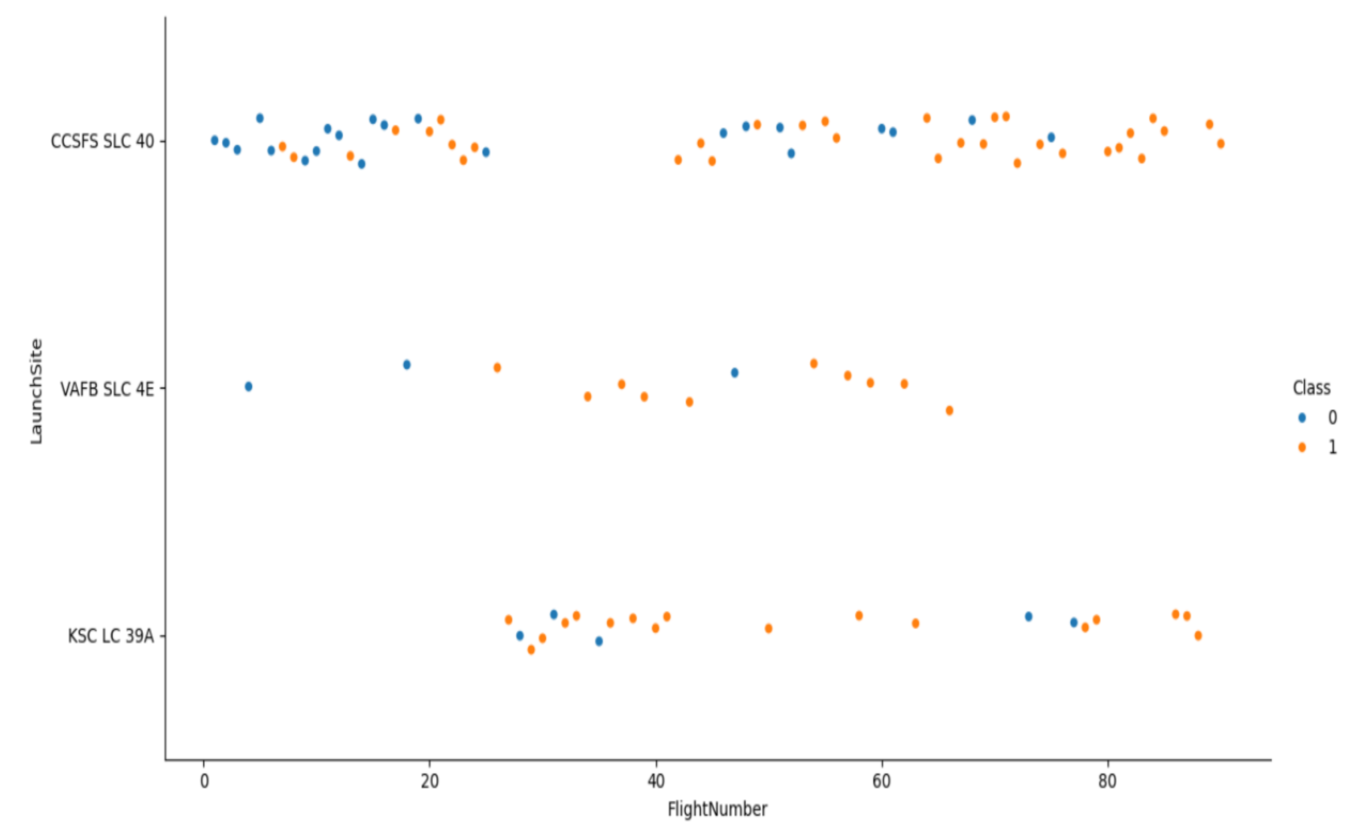


Section 2

Insights drawn from EDA

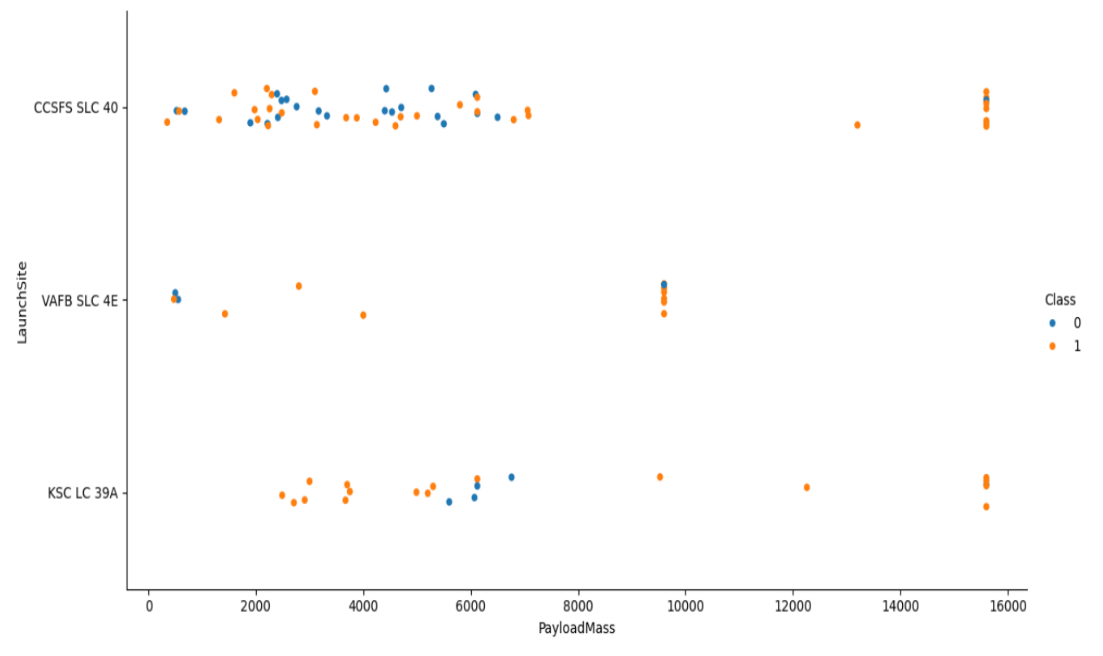
Flight Number vs. Launch Site

- **CCAFS SLC-40** handled the highest number of launches, making it SpaceX's **most frequently used launch site**.
- KSC LC-39A and VAFB SLC-4E show high success rates, but with significantly fewer launches compared to CCAFS SLC-40.
- **Launch success clearly improves with increasing flight number**, indicating operational learning and growing reliability over time.
- **Later missions across all launch sites exhibit fewer failures**, reinforcing SpaceX's performance improvements.



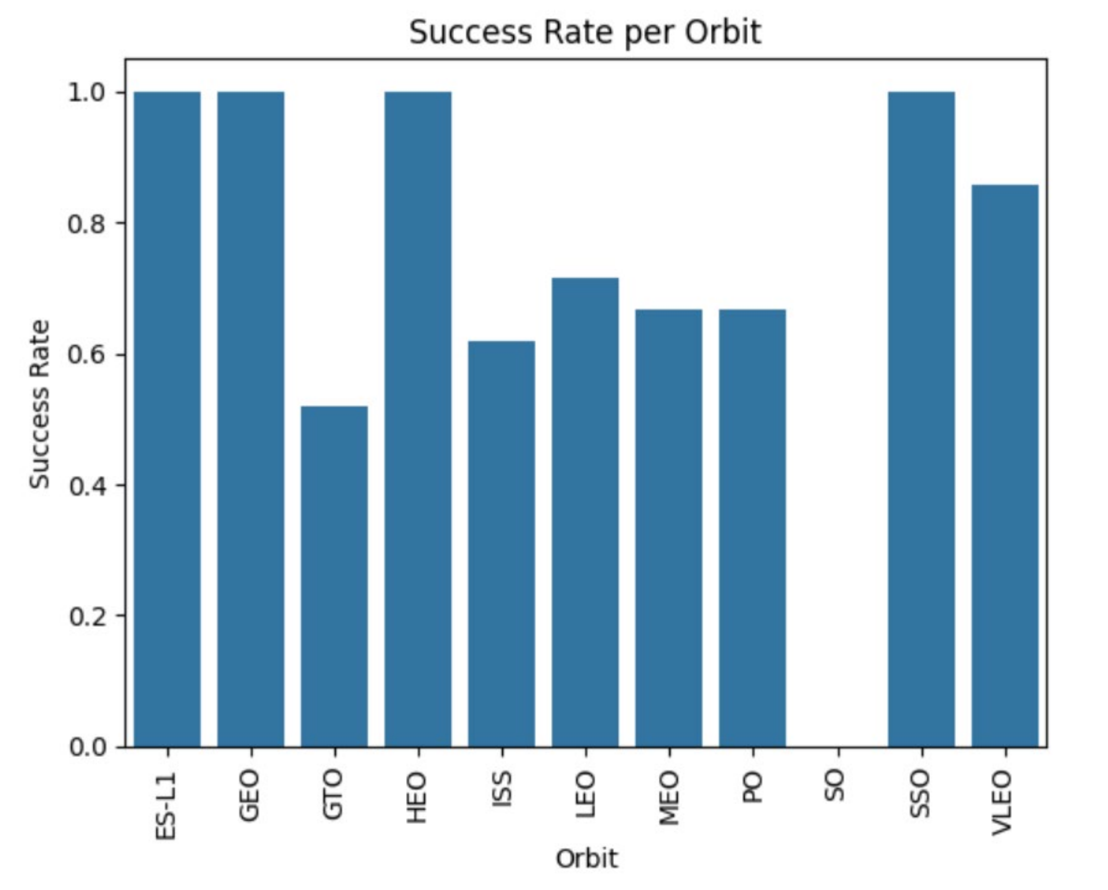
Payload vs. Launch Site

- CCAFS SLC-40 handles the widest range of payload masses, from small missions (~350 kg) to very heavy missions (>15,000 kg), making it the most versatile launch site.
- VAFB SLC-4E launch site is limited to low-to-medium payload missions and shows no launches above ~10,000 kg, indicating payload constraints.
- KSC LC-39A launch site is primarily used for medium and heavy payload missions, with most launches occurring above ~4,000 kg.
- The heaviest payload missions (>12,000 kg) are exclusively launched from launch sites CCAFS SLC-40 and KSC LC-39A, not from VAFB SLC 4E.

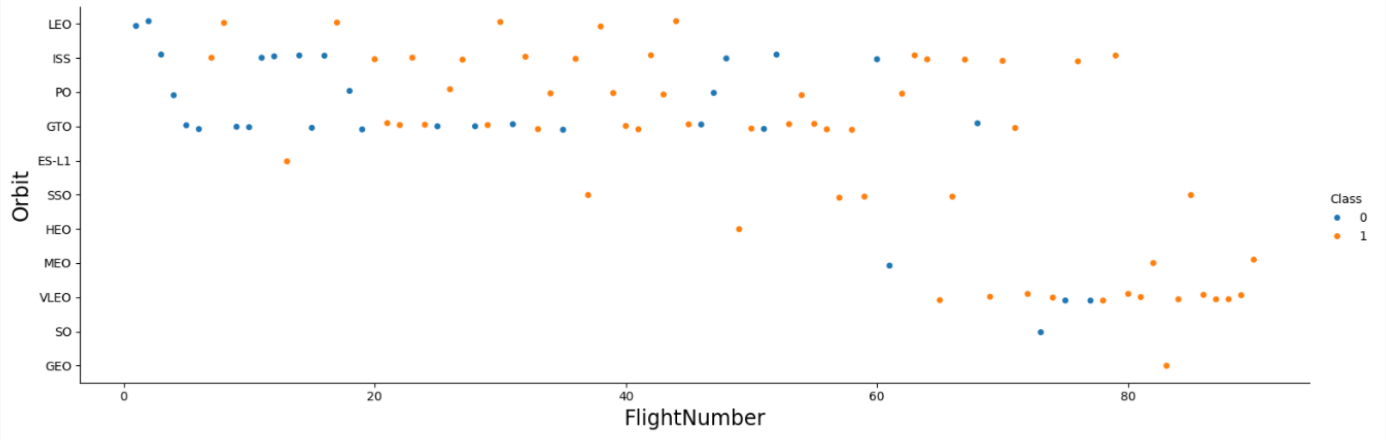


Success Rate vs. Orbit Type

- **ES-L1, GEO, HEO, and SSO** achieved **~100% mission success**, indicating peak launch reliability.
- **GTO** recorded the **lowest success rate** (~52%), making it the **highest-risk orbit**.
- LEO, MEO, PO, and ISS showed moderate success (≈60–70%), reflecting stable but variable performance.

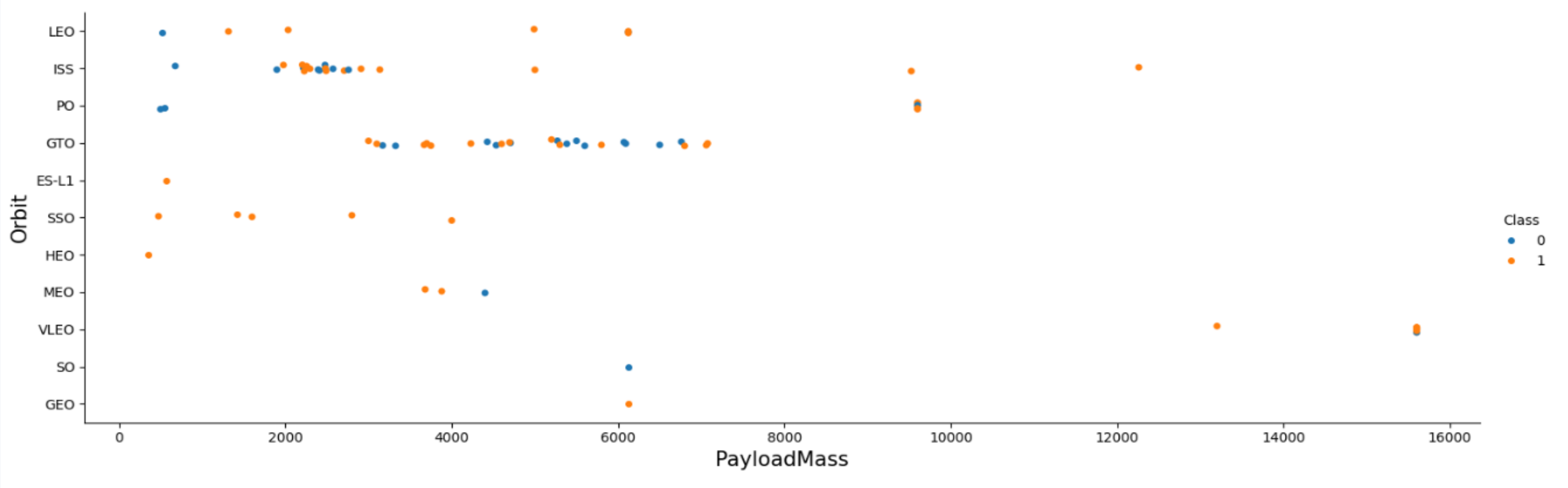


Flight Number vs. Orbit Type



- Launch success generally increases with flight number, indicating improved reliability and operational learning over time across most orbits.
- LEO and ISS missions show the most consistent success patterns, making them the most stable and frequently used operational orbits.
- GTO and PO missions display mixed success outcomes, suggesting higher mission complexity and risk compared to LEO/ISS.
- VLEO launches appear mostly in later flight numbers and are predominantly successful, highlighting this orbit as a new and rapidly maturing commercial opportunity.
- Orbits with fewer total missions (ES-L1, HEO, SO) show limited data but tend to be high-success, low-volume missions.

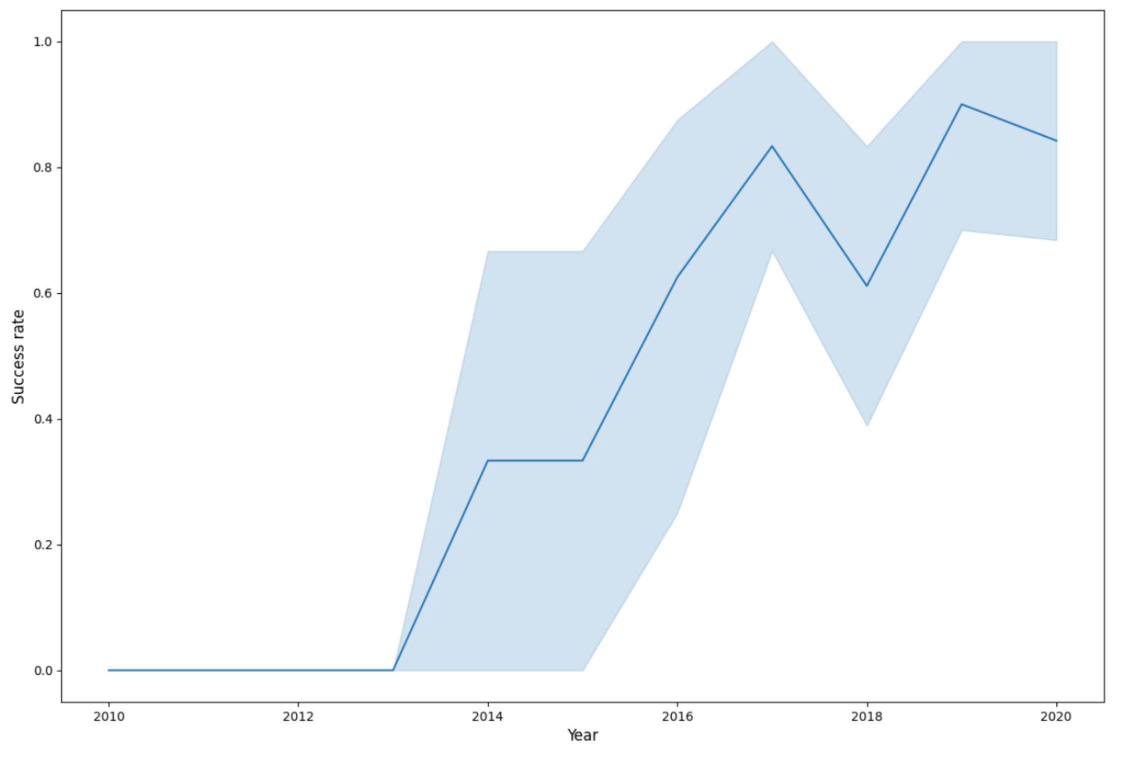
Payload vs. Orbit Type



- LEO and ISS missions dominate the low-to-mid payload range (≈1,000–4,000 kg) and show a high concentration of successful landings, indicating stable operational performance.
- For GTO missions, payload mass does not show a clear or consistent relationship with landing success, as both successful and unsuccessful landings occur across similar payload ranges.
- Ultra-heavy payload missions are rare and primarily appear in newer orbits like VLEO, reflecting recent expansion into high-capacity deployments.
- SO and GEO orbits have very limited launch frequency, making it difficult to draw strong statistical conclusions for these orbits.

Launch Success Yearly Trend

- The early years (2010–2013) represent a technology development and experimentation phase before consistent performance was achieved.
- **Success rate begins to improve significantly after 2013** and remains high through 2020, with minor fluctuations.
- The brief dip around 2018 is **followed by a rapid recovery**, demonstrating **resilience** and **fast learning** from failures.



All Launch Site Names

- The following SQL query extracts all distinct launch sites, showing that SpaceX conducts launches from four major operational facilities.

SQL Query Screenshot:

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

In [35]:

```
%%sql
SELECT DISTINCT "Launch_Site"
FROM SPACEXTABLE ORDER BY 1;
```

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

Out[35]:

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

Launch Site Names Begin with 'CCA'

- The query shown in below screenshot retrieves the first five launch records where the launch site name starts with “**CCA**”, highlighting early missions conducted from **Cape Canaveral** launch facilities (**CCAFS LC-40**).

SQL Query Screenshot:

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

In [36]:
```
%%sql

SELECT * FROM SPACEXTABLE
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5;
```

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

Out[36]:

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

- The total payload mass carried by boosters from **NASA** is **48,213 kg**.
- This total payload mass represents the combined cargo delivered by **SpaceX** for **NASA's** Commercial Resupply Services (**CRS**) missions.

SQL Query Screenshot:

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

In [40]:

```
%%sql
SELECT SUM(CAST("Payload_Mass__kg_" AS REAL)) AS total_payload_mass_kg
FROM SPACEXTABLE
WHERE "Customer" LIKE '%NASA (CRS)%';
```

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

Out[40]:

total_payload_mass_kg
48213.0

Average Payload Mass by F9 v1.1

- The average payload mass carried by the **Falcon 9 v1.1** booster is **~2,928 kg**, indicating this version was primarily used for **medium-weight missions** during its operational period.

SQL Query Screenshot:

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

In [49]:

```
%%sql
SELECT AVG(CAST("Payload_Mass__kg_" AS REAL)) AS avg_payload_mass_kg
FROM SPACEXTABLE
WHERE "Booster_Version" = 'F9 v1.1';
```

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

Out[49]:

avg_payload_mass_kg
2928.4

First Successful Ground Landing Date

- The **first successful ground pad landing** was achieved on **December 22, 2015**, marking a major milestone in SpaceX's reusable rocket program.

SQL Query Screenshot:

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

In [50]:

```
%%sql
SELECT MIN("Date") AS first_success_ground_pad
FROM SPACEXTABLE
WHERE "Landing_Outcome" LIKE 'Success (ground pad)%';
```

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

Out[50]:

first_success_ground_pad
2015-12-22

Successful Drone Ship Landing with Payload between 4000kg and 6000kg

SQL Query Screenshot:

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

In [51]:
```
%%sql
SELECT DISTINCT "Booster_Version"
FROM SPACEXTABLE
WHERE "Landing_Outcome" LIKE 'Success (drone ship)%'
  AND CAST("Payload_Mass__kg_" AS REAL) > 4000
  AND CAST("Payload_Mass__kg_" AS REAL) < 6000;
```

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

Out[51]:

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

- The query identifies **four Falcon 9 booster variants (B1022, B1026, B1021.2, B1031.2)** that successfully landed on drone ships while **carrying medium-heavy payloads between 4,000 kg and 6,000 kg**, demonstrating their proven reliability in this demanding payload range.

Total Number of Successful and Failure Mission Outcomes

- The results show that **SpaceX** achieved 99 successful missions, with only 1 in-flight failure, and 1 mission marked as successful but with unclear payload status, highlighting **an overall extremely high mission success rate**.

SQL Query Screenshot:

List the total number of successful and failure mission outcomes

In [65]:
```
%%sql
SELECT TRIM("Mission_Outcome") AS Outcome,
       COUNT(*) AS QTY
FROM SPACEXTABLE
GROUP BY Outcome;
```

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

Out[65]:

Outcome	QTY
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- The SQL query shown in screenshot uses a subquery with the **MAX() aggregate function** to identify all Falcon 9 booster versions that carried the maximum recorded payload mass.
- The result shows that multiple Block 5 boosters have successfully handled the heaviest payload missions, demonstrating strong performance consistency across the fleet.
- This highlights the reliability and reusability of **Falcon 9 Block 5 boosters** for high-capacity launches, even under maximum load conditions.

SQL Query Screenshot

In [66]:

```
%%sql
SELECT DISTINCT "Booster_Version"
FROM SPACEXTABLE
WHERE CAST("Payload_Mass__kg_" AS REAL) = (
  SELECT MAX(CAST("Payload_Mass__kg_" AS REAL)) FROM SPACEXTABLE
) ORDER BY "Booster_Version";
```

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

Out[66]:

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

2015 Launch Records

- In **2015** there were only two "Failure (drone ship)" landings, occurred in **January** and **April**.
- Both failures happened at the same launch site, **CCAFS LC-40**, indicating that all drone-ship failures that year were associated with this facility.
- **After April 2015**, no additional drone-ship landing failures are reported for that year in the dataset, hinting at improvements in landing performance as the year progressed.

SQL Query Screenshot

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.

In [22]:

```
%%sql
SELECT
  CASE substr("Date",6,2)
    WHEN '01' THEN 'January'   WHEN '02' THEN 'February'
    WHEN '03' THEN 'March'     WHEN '04' THEN 'April'
    WHEN '05' THEN 'May'       WHEN '06' THEN 'June'
    WHEN '07' THEN 'July'      WHEN '08' THEN 'August'
    WHEN '09' THEN 'September' WHEN '10' THEN 'October'
    WHEN '11' THEN 'November'  WHEN '12' THEN 'December'
  END AS Month,
  "Landing_Outcome",
  "Booster_Version",
  "Launch_Site"
FROM SPACEXTABLE
WHERE substr("Date",1,4) = '2015'
  AND "Landing_Outcome" LIKE 'Failure (drone ship)%';
```

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

Out[22]:

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

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

- The **SQL query** shown in screenshot filters all launches between June 4, 2010, and March 20, 2017, using the **BETWEEN** condition on the Date column. It then groups the data by landing outcome. **COUNT(*)** is used to calculate how many times each landing outcome occurred. Finally, results are ranked in descending order of frequency using **ORDER BY** qty **DESC**.
- The query result showed which landing outcomes were most common during SpaceX's early landing years.
- "**No attempt**" is the **most frequent outcome** (10 launches), showing that in early years, many missions did not even try to recover the booster.

SQL Query Screenshot

In [67]:

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

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

Out[67]:

Landing_Outcome	qty
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

All Launch Sites

❑Although none of the U.S. launch sites are located near the Equator, Florida is positioned farther south than California, allowing rockets to benefit more from Earth's rotational speed.

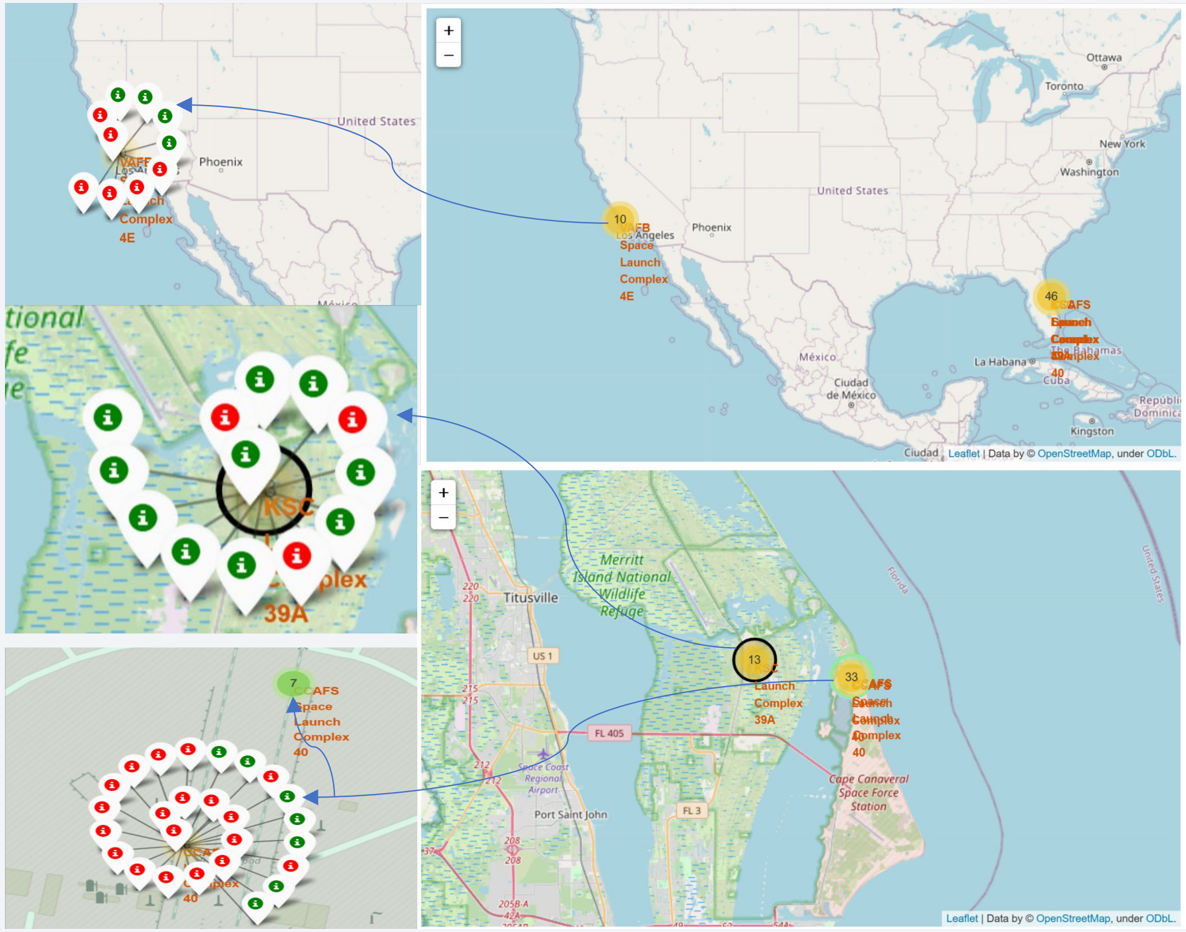
❑All launch sites are intentionally built along the coastline so that rockets can launch safely over open water, reducing risks from falling debris and enabling safe booster recovery operations.

Success & Failed Launches for each site on the map

- The maps show launch activity and landing outcomes across California and Florida launch sites using clustered and color-coded markers.

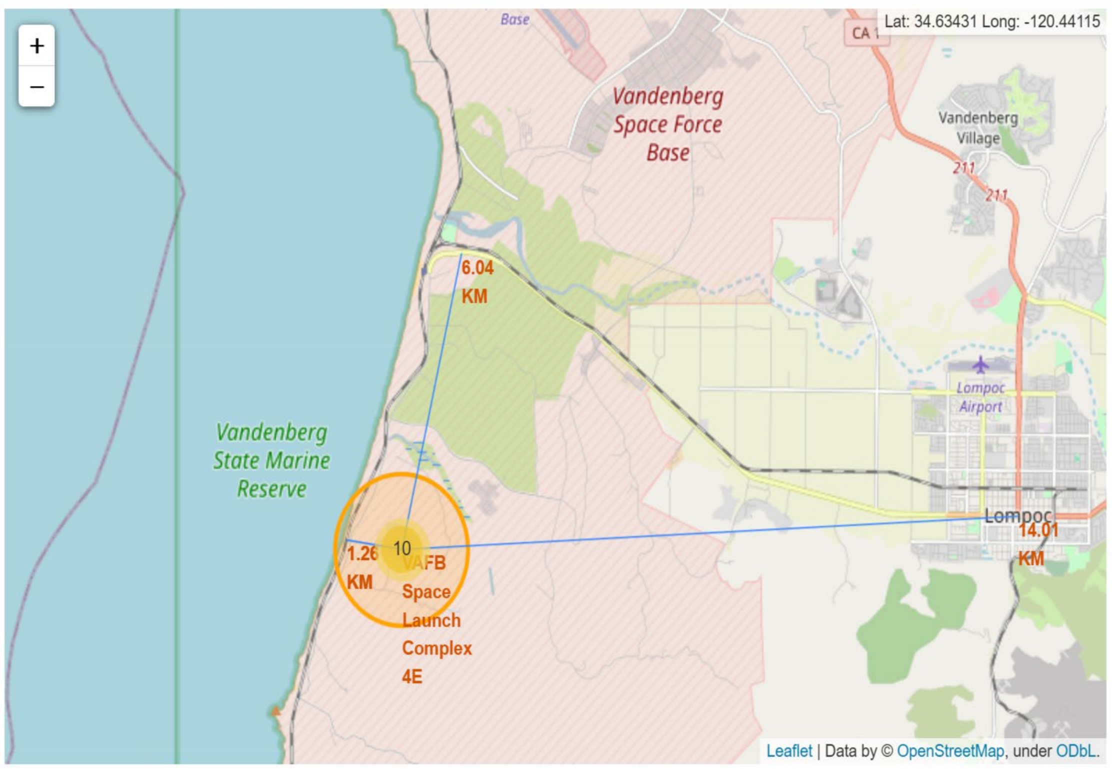
Launch Sites	Successful Landings (green markers)	Failed Landings (red markers)	Total Launches
CCAFS LC-40 & CCAFS SLC-40	10	23	33
KSC LC-39A	10	3	13
VAFB SLC-4E	4	6	10

- Florida dominates overall launch activity with a total of 46 launches.
- California (VAFB SLC-4E) recorded 10 launches, with 4 successful landings (green markers) and 6 failed landings (red markers).
- The higher density of markers in Florida reflects SpaceX's primary operational focus, while California is used more selectively, likely for polar and specialized orbits.



Proximity Analysis of Launch Site Using Folium Map

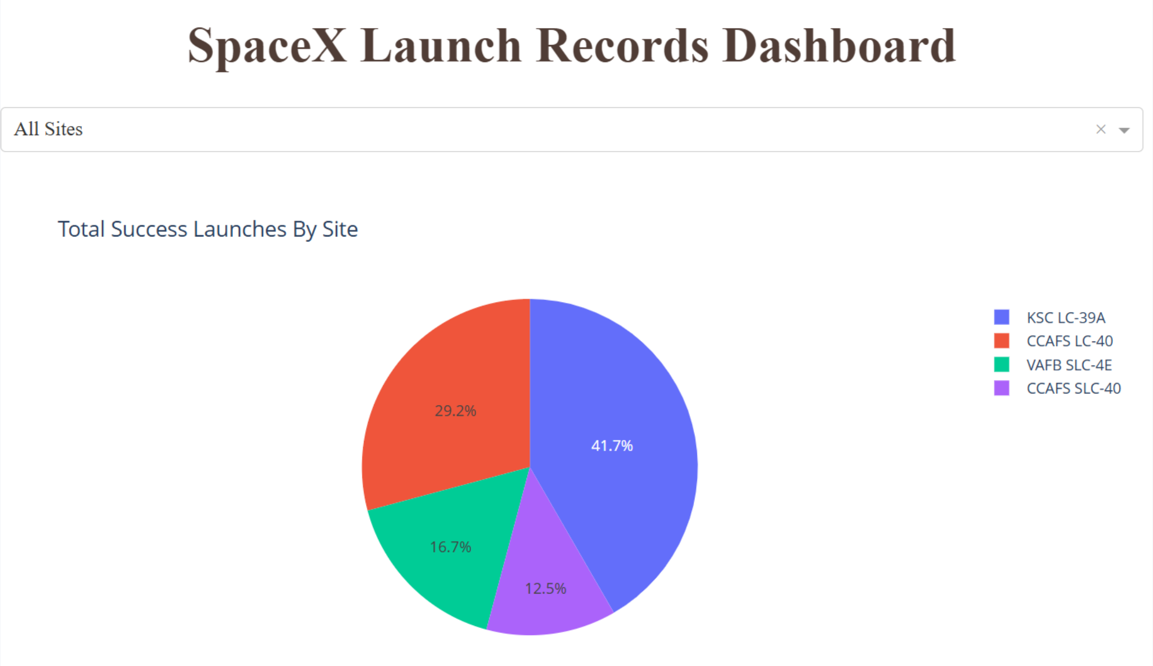
- Launch sites are intentionally located **near coastlines** to ensure rockets launch safely over the ocean.
- They maintain **moderate distance from highways**, close enough for equipment transport but far for safety.
- **Railways are not directly adjacent** to launch pads, reducing explosion-related risks.
- Nearby cities remain **over 10 km away**, maintaining a safe buffer zone while supporting workforce access.
- The map demonstrates these principles clearly, showing measured distances from launch site to the coast, roadways, and Lompoc city.



Section 4

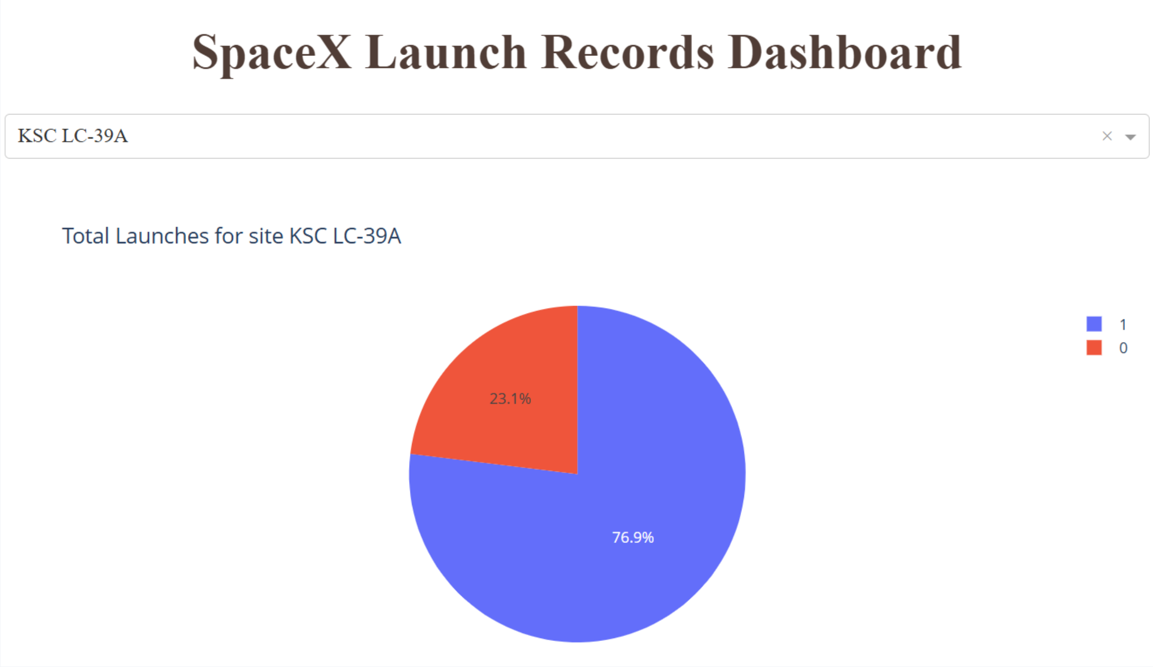
Build a Dashboard with Plotly Dash

Total Successful Launches by Site



Most of **SpaceX’s** successful missions are concentrated at **KSC LC-39A** and **CCAFS LC-40** launch sites located in **Florida**, highlighting it as the primary hub for reliable launch operations.

Launch Site with Highest Launch Success Ratio



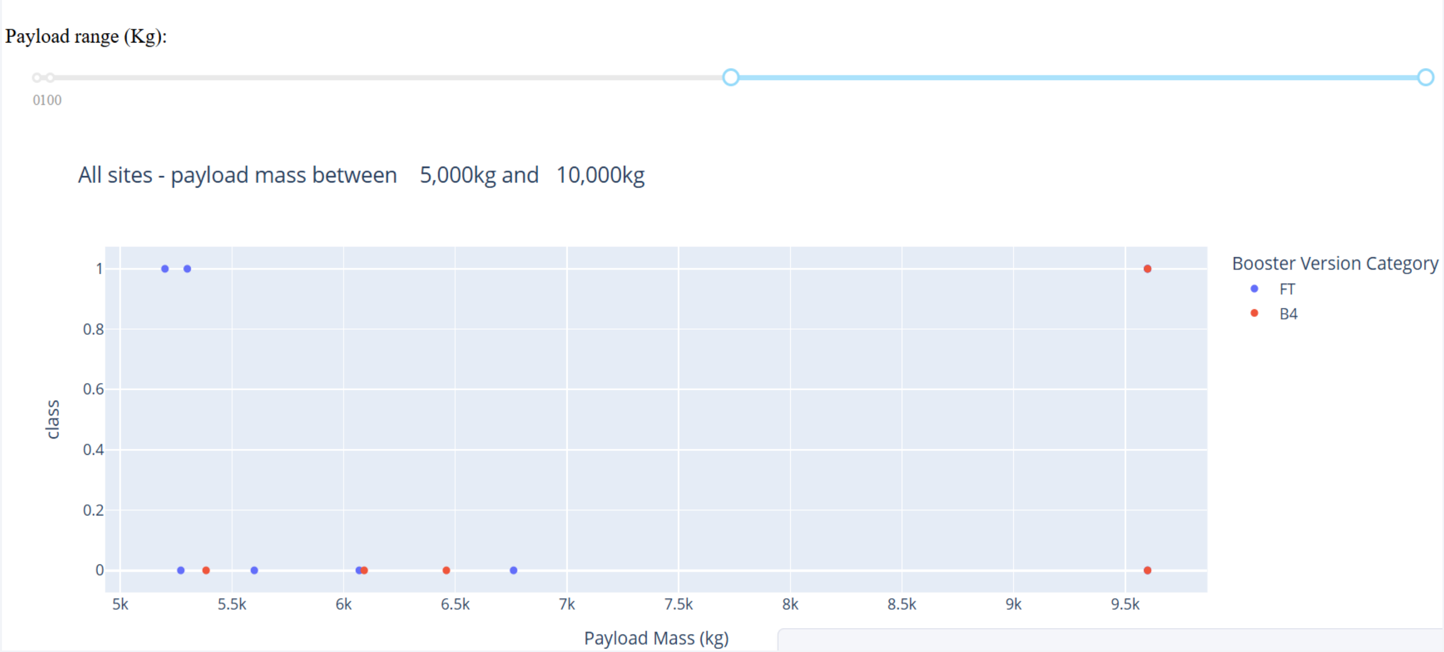
With nearly **77% successful** launches, **KSC LC-39A** demonstrates the highest reliability among all SpaceX launch sites.

Payload vs. Launch Outcome

- Below plot shows the relationship between Payload Mass and Launch Outcome (class: 1 = success, 0 = failure) across all SpaceX launch sites.
- Most successful launches are clustered between ~2,000 kg and ~6,000 kg, indicating this is the most reliable payload range for Falcon 9 missions.

Payload vs. Launch Outcome

- **FT** and **B4 booster versions dominate** the successful launches, indicating higher reliability compared to **older versions (v1.0 and v1.1**).
- **The heaviest mission (~9,600 kg)** was successful with **booster B4**, demonstrating the increased capability of later booster versions.

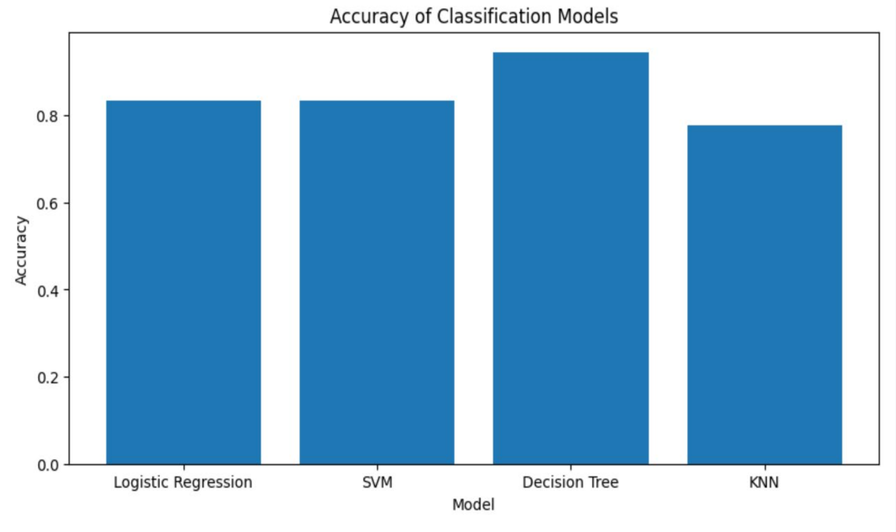


Section 5

Predictive Analysis (Classification)

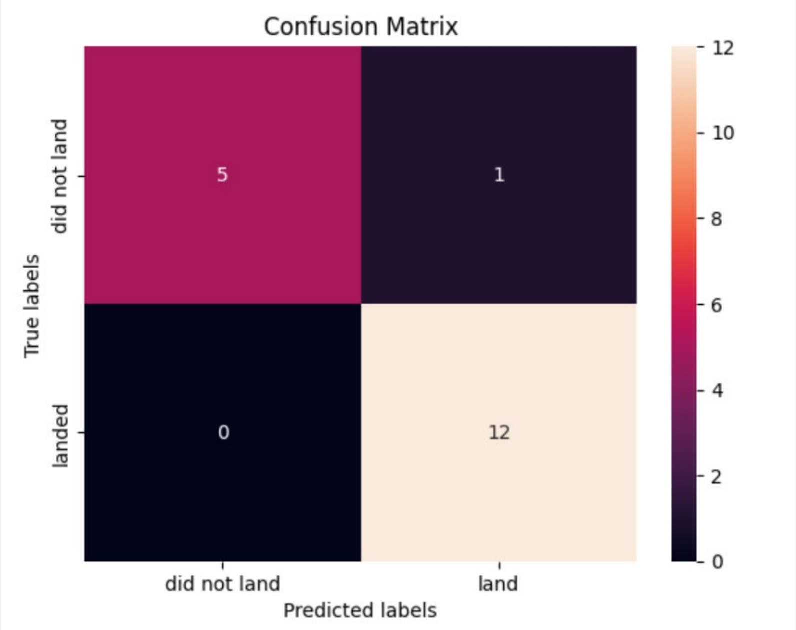
Classification Accuracy

- **Four** machine learning classification models were evaluated: **Logistic Regression**, **Support Vector Machine (SVM)**, **Decision Tree**, and **K-Nearest Neighbors (KNN)**, and their accuracies are compared in the bar chart.
- The **Decision Tree model achieved the highest accuracy of 94.4%** indicating it performs the best in predicting launch success.
- **Logistic Regression** and **SVM** show very similar and stable performance, each achieving **83.3% accuracy**.
- The **KNN model recorded the lowest accuracy** of **~78%**, suggesting it is less effective for this dataset compared to the other methods.
- Overall, the results show that **tree-based models are better suited for capturing the complex patterns in SpaceX launch data**.



Confusion Matrix

- The Decision Tree model demonstrates **excellent predictive reliability**, especially in identifying **successful landings with 100% recall**.
- With only **one false alarm and zero missed successes**, this model is **well-suited for mission-critical SpaceX launch outcome predictions**.



Conclusions

- Collected and integrated **SpaceX** launch data using **APIs, web scraping, and structured datasets**, capturing mission details such as **payload mass, orbit, booster version, launch site, and landing outcomes**.
- Performed **data cleaning, feature engineering, and exploratory analysis**, revealing that **launch success steadily improved after 2013**, and **Florida launch sites dominate both total and successful launches**.
- **Geospatial analysis using Folium maps** confirmed that all launch sites are strategically placed **near coastlines, away from cities, and close to transport infrastructure** for safety and logistics.
- Built and evaluated **four machine learning models** (Logistic Regression, SVM, KNN, Decision Tree) to predict landing success. **Decision Tree Classifier emerged as the best model** with **~94% accuracy and zero missed successful landings**, making it the most reliable choice for operational prediction.
- This project demonstrates how **data science can accurately predict rocket landing outcomes**, enabling better **mission planning, cost efficiency, and risk reduction**.

Appendix

- Folium maps are not supported for direct display on GitHub, so screenshots were used for visualization.
- All relevant assets like Python code snippets, SQL queries, charts, data sets, dashboards, and Jupyter Notebooks are available at following GitHub Repository.

View the full project on GitHub

Thank you!