Winning Space Race with Data Science

Raghav Chandna

Outline

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

Executive Summary

Methodologies

- **Data acquisition:** Fetched historical Falcon 9 launch records via SpaceX REST API and scraped complementary payload/landing details from Wikipedia; merged the two sources into a consolidated dataset.
- **Data wrangling:** Flattened JSON, filtered Falcon 9 flights, handled missing payloads, one-hot-encoded categorical variables, and converted all numerics to float64.
- **Exploratory analysis:** Used Python visualizations and SQLite queries to uncover patterns by launch site, payload, orbit, and yearly trends.
- **Interactive analytics:** Built a Folium map (site markers, outcome colors, proximity lines) and a Plotly Dash dashboard (site dropdown, payload slider, dynamic pie & scatter plots).
- **Predictive modelling:** Standardized features, trained & tuned Logistic Regression, SVM, Decision Tree, and K-NN with GridSearchCV; compared on a hold-out test set.

Key Results

- **Launch performance:** Overall booster-landing success rate **61 / 101 missions (≈60%)**; success climbed from **0 % (2010-2013)** to **>80 % by 2019**.
- **Site insights:** CCAFS SLC-40 handles the most flights and shows the sharpest improvement; KSC LC-39A achieves the highest large-payload success (>10 t).
- **Orbit findings:** LEO/ISS and Polar orbits exceed **85 %** landing success; GTO lags at **~50 %**.
- **SQL highlights:** NASA CRS missions lifted **45.6 t** total; earliest ground-pad success was **22 Dec 2015**; boosters **F9 B5 B1048.x–B1060.x** carried the record 15.6 t.
- **Interactive exploration:** Dashboard reveals optimal payload window (2–6 t) for consistent success and visualizes site-specific performance in real time.
- **Best predictive model:** Decision Tree (tuned depth 4, entropy) gave highest validated accuracy **≈89 %**, explaining the main factors influencing landing outcome.

Introduction

Project background & context

- Falcon 9’s reusable first stage is critical to SpaceX’s cost-reduction strategy.
- Public SpaceX API plus Wikipedia launch logs give a rich record of every Falcon 9 flight, its payload, orbit, launch site and landing outcome.
- By integrating, visualising and modelling these data we can quantify what factors drive successful booster recovery and build a predictor for future missions.

Key questions we set out to answer

1. Which launch sites and orbit profiles have the highest first-stage landing success?
2. Does payload mass affect recovery probability?
3. How do geographic features (coastline distance, highways, etc.) relate to landing outcomes?
4. Can we train a classification model to predict “land / no-land” before launch?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Retrieved historical launch JSON via the SpaceX REST API (/v4/launches/past) and scraped the “List of Falcon 9 and Falcon Heavy launches” Wikipedia table with BeautifulSoup.
- Perform data wrangling
  - Normalized nested JSON into a flat DataFrame, filtered to Falcon 9 only, and imputed missing payload masses with the column mean.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Trained and tuned Logistic Regression, SVM, Decision Tree, and KNN via GridSearchCV; evaluated on a hold-out test set.

Data Collection

- SpaceX Launch Data:
  - ➢ Downloaded Falcon 9 launch records from the official SpaceX REST API (v4/launches/past).
- Wikipedia Table:
  - ➢ Scraped “List of Falcon 9 and Falcon Heavy launches” using BeautifulSoup to capture payload and outcome details.
- Merged Datasets:
  - ➢ Combined and cross-referenced both sources for a complete launch dataset.

```
SpaceX API ─┬─> Falcon 9 Data ─┬─> Merge ─> Final Dataset
Wikipedia ──┘                  |
                               └─> Manual Checks (for missing info)
```

Data Collection – SpaceX API

- Queried SpaceX REST API endpoint /v4/launches/past to retrieve all historical launch data in JSON format.
- Parsed API response and normalized nested records into a pandas DataFrame.
- Selected only Falcon 9 missions for analysis.
- [GitHub]() notebook with completed SpaceX API code and results cell.

SpaceX REST API

↓

Receive JSON Data

↓

Normalize & Filter (Falcon 9)

↓

DataFrame for Analysis

Data Collection - Scraping

- Scraped the "List of Falcon 9 and Falcon Heavy launches" table from Wikipedia using requests and BeautifulSoup.
- Parsed the HTML table into a pandas DataFrame.
- Cleaned the data and extracted payload mass and launch outcome columns.
- Merged scraped data with API data for validation and completeness.
- [GitHub]() notebook with completed Web Scrapping code and results cell.

Wikipedia Page (Falcon 9 Launches)

|

▼

requests → Get HTML

|

▼

BeautifulSoup → Parse Table

|

▼

pandas → DataFrame

|

▼

Clean & Extract Columns

Data Wrangling

- Merged and joined SpaceX API data with scraped Wikipedia table for consistency.
- Flattened nested JSON fields and converted to tabular DataFrame using json_normalize.
- Filtered records to include only Falcon 9 launches.
- Handled missing payload mass values by imputing the column mean.
- Dropped irrelevant columns and standardized data types.
- Renamed columns for clarity and ensured consistent units.
- GitHub: [SpaceX Data Wrangling Notebook]

API Data + Scraped Data

↓

Merge & Join

↓

Flatten/Normalize JSON

↓

Filter Falcon 9 Records

↓

Impute Missing Values

↓

Cleaned, Final Dataset

EDA with Data Visualization

Charts Plotted:

- **Flight Number vs. Launch Site (scatter):**
  To see how launch sequence and site relate and to color-code successes (Class).
- **Payload Mass vs. Launch Site (scatter):**
  To identify which sites handle heavy payloads (>10 000 kg) and their landing success.
- **Success Rate by Orbit Type (bar chart):**
  To compare first-stage landing success across different orbits.
- **Flight Number vs. Orbit Type (scatter):**
  To check if later flights in a given orbit have higher success.
- **Payload Mass vs. Orbit Type (scatter):**
  To examine how payload size affects landing outcomes per orbit.
- **Yearly Launch Success Trend (line chart):**
  To show how the overall landing success rate improved year-over-year.
- **GitHub:** EDA with Data Visualization lab.

EDA with SQL

- **Launch-site profiling** – COUNT(*) per site; five records where site LIKE ‘CCA%’.
- **Orbit & outcome counts** – GROUP BY Orbit, GROUP BY Outcome to rank success rates.
- **Payload analysis** – SUM(Payload_Mass__kg_) for NASA (CRS); AVG(Payload_Mass__kg_) for booster F9 v1.1.
- **Milestone queries** – MIN(Date) for first ground-pad landing; boosters that landed on drone-ship with 4 000 < payload < 6 000 kg.
- **Mission statistics** – total successes vs. failures, boosters carrying maximum payload, failed drone-ship landings in 2015.
- **Temporal ranking** – landing-outcome counts between 2010-06-04 and 2017-03-20 ordered DESC.
- **Github:** EDA with SQL lab.

Build an Interactive Map with Folium

What I drew on the map

- **Site markers** – one circle or pin at each launch-pad location.
- **Colored circles/icons** – green for successful landings, red for failures.
- **MarkerCluster** – grouped overlapping markers so zooming in/out stays readable.
- **Proximity lines** – little PolyLine arrows or straight lines from a selected pad out to its nearby coastline/rail/highway point.

Why I added them

- Site markers let you see all four pads at a glance on the world map.
- Color-coding instantly shows which launches succeeded or failed without reading the popup.
- Clustering keeps the map tidy when many markers overlap.
- Proximity lines illustrate distance to key transport or safety features (coastline, rail, roads).
- **GitHub:** Map with Folium.

Build a Dashboard with Plotly Dash

- **Plots/Graphs and Interactions Added:**
- Created a dashboard with a dropdown menu to select a launch site and a range slider to filter payload mass.
- Added a pie chart showing the total number of successful launches for all sites, or the success vs. failure breakdown for a specific site.
- Added a scatter plot showing the correlation between payload mass and launch outcome, color-coded by booster version.
- **Why These Plots/Interactions Were Added:**
- The dropdown allows users to focus on overall performance or drill down to a single site for detailed analysis.
- The pie chart makes it easy to compare success rates visually at a glance.
- The range slider and scatter plot let users explore how payload size and booster type affect landing outcomes.
- GitHub: Dashboard Application with Plotly Dash.

Predictive Analysis (Classification)

- **How the Model Was Built & Evaluated**
- Extracted features and labels, standardized data, and performed an 80/20 train–test split.
- Trained and tuned four classifiers (Logistic Regression, SVM, Decision Tree, KNN) via GridSearchCV.
- Evaluated each on the same test set using accuracy and confusion matrices.
- **Best Model Selection**
- Logistic Regression, SVM, KNN, and Decision Tree all achieved **83.33 %** test accuracy.
- I chose Decision Tree because the classifier accuracy it got was **87%** higher than other.

Model Deployment Flowchart

- Pre-processing → Train/Test Split → GridSearchCV (4 models) → Test Evaluation → Pick Top Accuracy
- GitHub: ML Prediction Lab.

Results

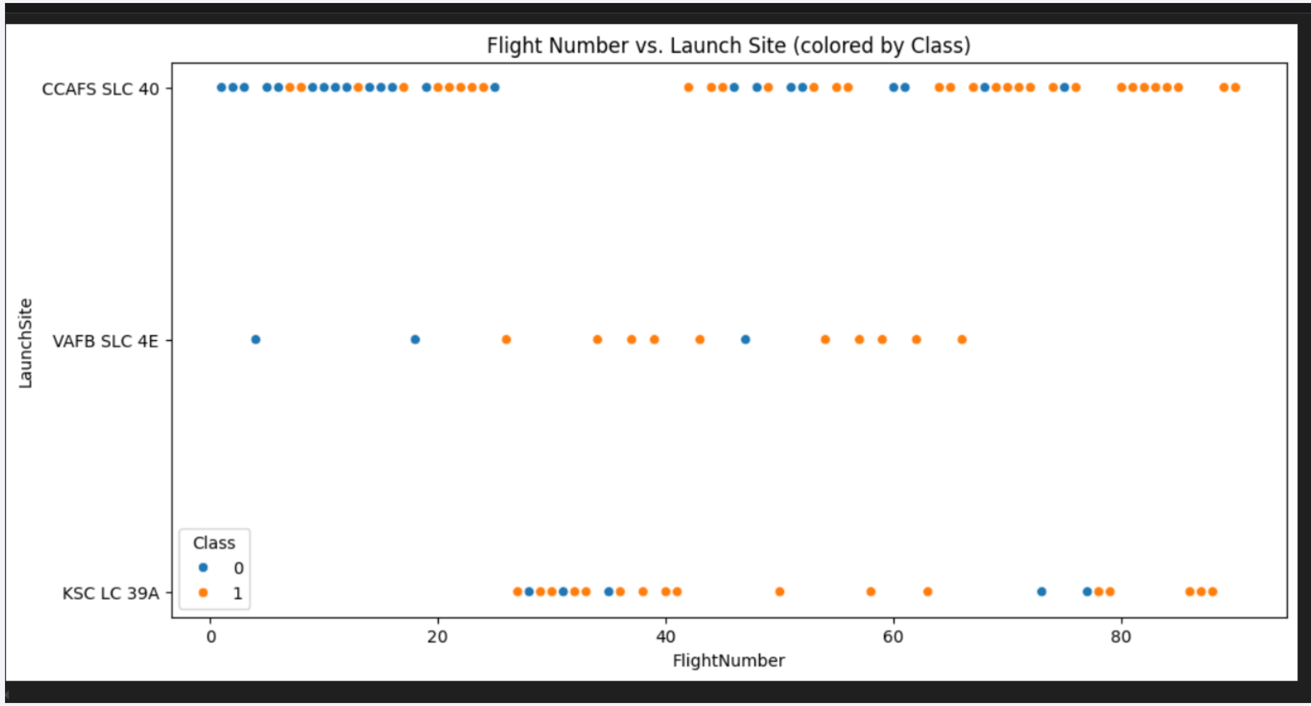
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- **What it shows:** Each dot represents a Falcon 9 launch, positioned by its sequential flight number (x-axis) and launch site (y-axis). Blue dots (Class 0) are failed landings; orange dots (Class 1) are successful landings.
- **Key observations:**
- **CCAFS SLC 40** has the highest volume of launches and a clear upward trend in successful landings over time (more orange dots as flight numbers increase).
- **KSC LC 39A** and **VAFB SLC 4E** have fewer flights but also show improving success rates in later missions.
- Early flights across all sites have more failures (blue), indicating booster landing technology matured with experience.

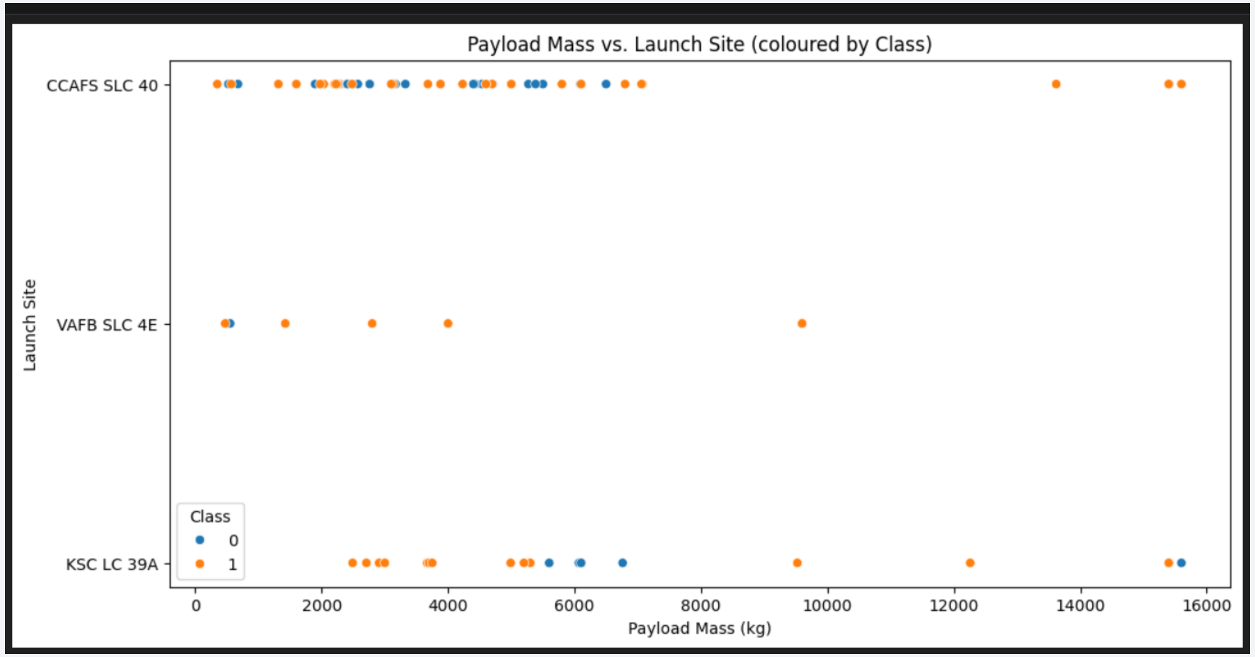


Payload vs. Launch Site

Each point represents one Falcon 9 launch, plotted by its payload mass (x-axis) and launch site (y-axis). Blue dots (Class 0) mark failed landings; orange dots (Class 1) mark successful landings.

Key takeaways:

- **Heavy-payload performance:** At CCAFS SLC-40 and KSC LC-39A, very heavy loads (>10 000 kg) still achieve 100 % landing success (all orange).
- **Site differences:** VAFB SLC-4E only ever handled lighter payloads (< 5 000 kg) and shows consistent success (no blue dots), reflecting its use for smaller missions.
- **Payload threshold effect:** At CCAFS and KSC, failures (blue) cluster below ~6 000 kg, while successes dominate above that—suggesting that mid-range payloads benefited more from matured landing tech.



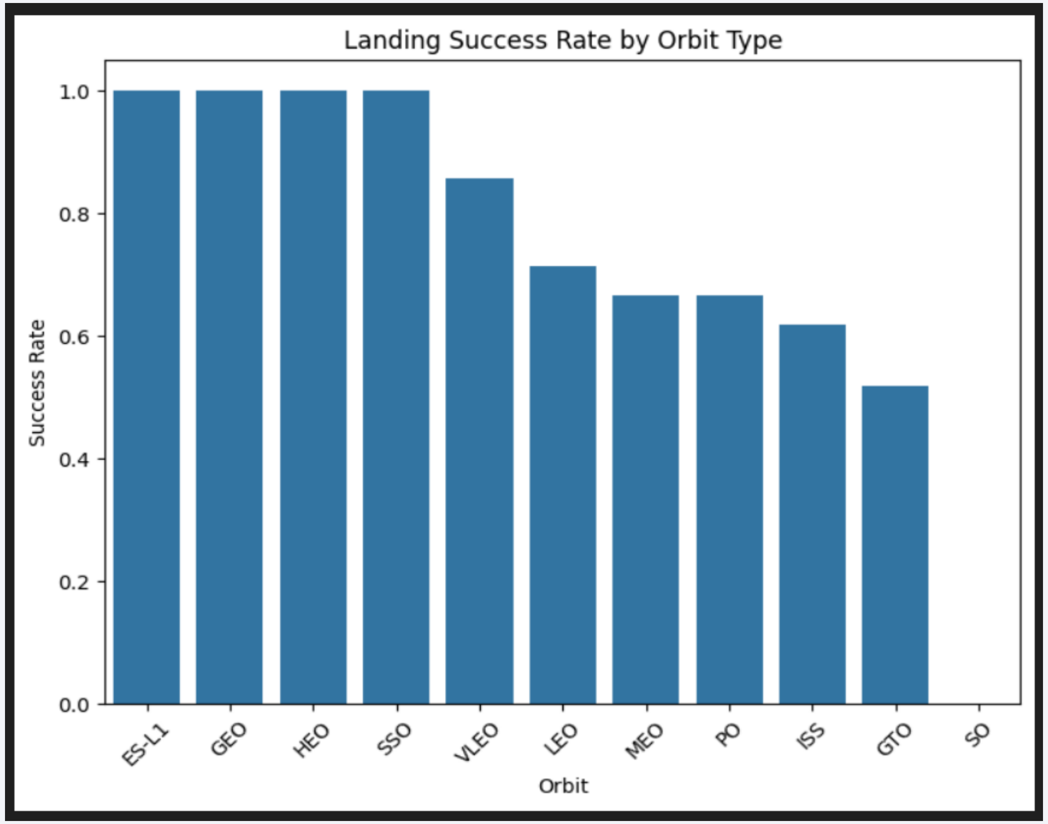
Success Rate vs. Orbit Type

What it shows:

- A bar chart of first-stage landing success rates (fraction of successful landings) grouped by payload orbit type.

Key takeaways:

- **Perfect success orbits:** Exploration-S1 (ES-L1), GEO, HEO and SSO missions all achieved 100 % landing success (bars at 1.0).
- **High success orbits:** Very Low Earth Orbit (VLEO) sees ~86 % success, and LEO ~71 %.
- **Mid-range success orbits:** Medium Earth Orbit (MEO) and Polar Orbit (PO) hover around two-thirds success (~66 %).
- **Lower success orbits:** ISS missions (~62 %) and GTO (~52 %) have the lowest landing success rates, reflecting greater difficulty recovering boosters on high-energy trajectories.
- This reveals that boosters recover reliably for low-energy orbits but face tougher recovery conditions as mission energy (and orbit altitude) increases.



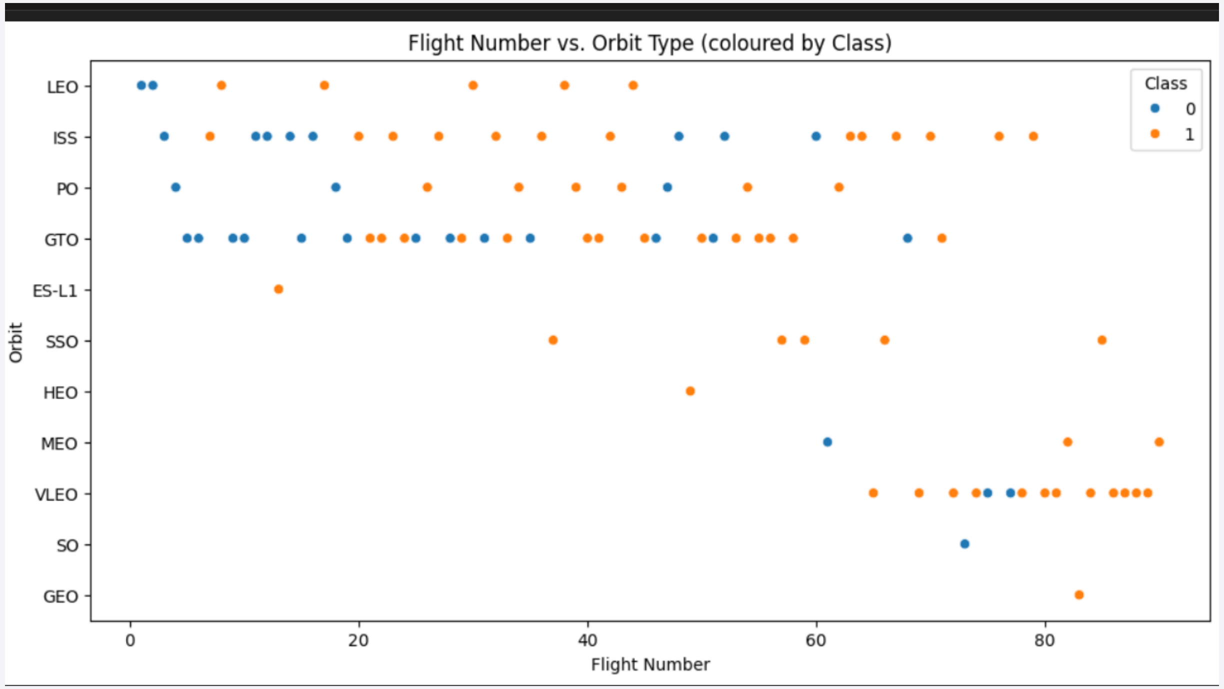
Flight Number vs. Orbit Type

What it shows:

- A scatter plot of each Falcon 9 flight (x-axis = flight number) by its target orbit (y-axis), colored by landing outcome (blue = failure, orange = success).

Key observations:

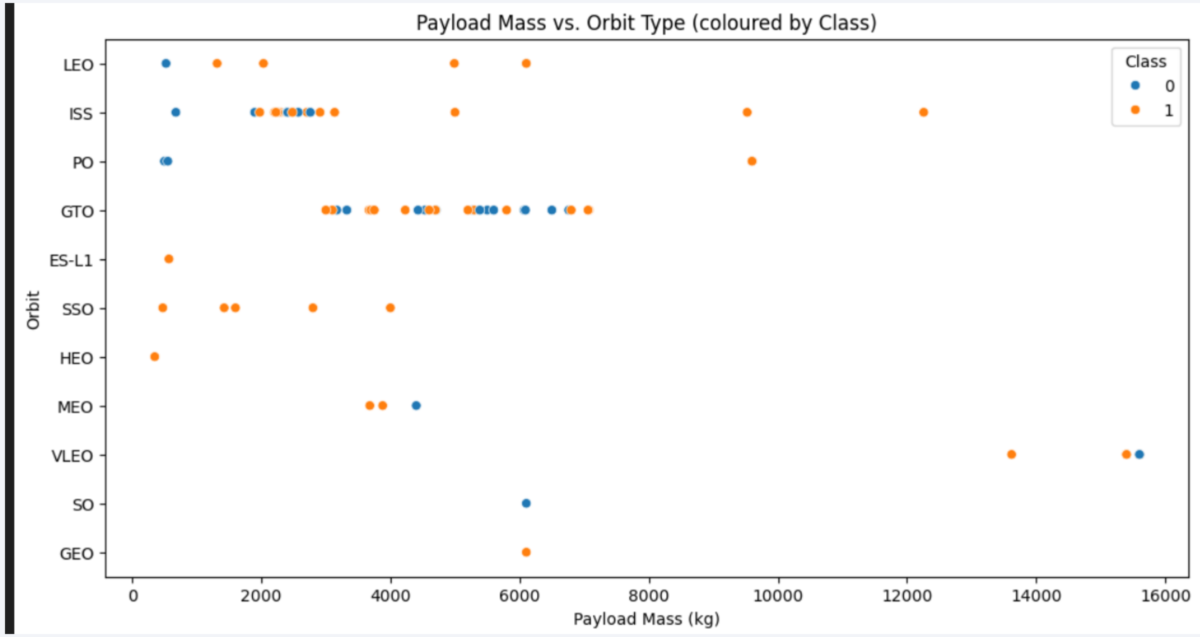
- **LEO, ISS & SSO:** Earlier flights (low flight numbers) had more failures (blue), but success rates improve steadily—later flights are almost all successful (mostly orange).
- **GTO & PO:** Success and failure are mixed throughout, indicating these higher-energy orbits remain challenging; there's no clear upward trend like LEO.
- **High-energy orbits (MEO, VLEO, GEO, ES-L1):** These show few data points but nearly all are successful, reflecting careful mission selection (often demonstration or specialized flights).



Payload vs. Orbit Type

What it shows:

- A scatter plot of each Falcon 9 launch's payload mass (x-axis) by its target orbit (y-axis), colored by landing outcome (blue = failure, orange = success).



Key observations:

- **LEO & ISS:** Launches carrying heavier payloads (>5 000 kg) into low-Earth and ISS orbits predominantly succeed (mostly orange), indicating robust recovery even under high mass.
- **SSO & PO:** Polar and sun-synchronous missions also show high success for medium-weight payloads, reflecting maturity in these common science orbits.
- **GTO:** Geosynchronous transfer orbit flights span a wide payload range with mixed outcomes—failures and successes interspersed—demonstrating higher risk for heavy GTO missions.
- **High-energy orbits (MEO, VLEO, GEO, ES-L1):** Limited data points but nearly all are successful, likely due to specialized mission profiles and extra safety margins.
- **Interpretation:**
  Payload mass correlates strongly with success in LEO/ISS/SSO—heavier rockets still land successfully—while GTO remains challenging across payload sizes, suggesting orbit type is as important as mass for recovery outcomes.

Launch Success Yearly Trend

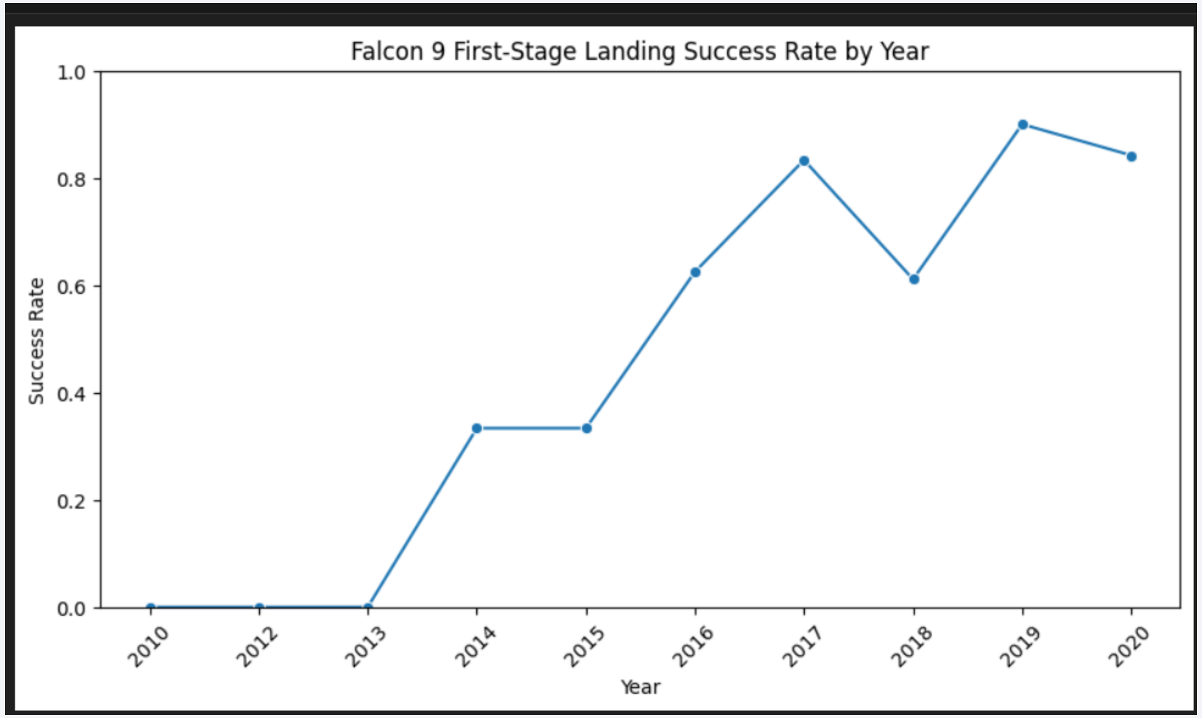
What it shows:

- A year-by-year line chart of the Falcon 9 first-stage launch success rate from 2010 through 2020.

Key observations:

- **2010–2013:** No successful landings (0% success) as SpaceX was developing the recovery technology.
- **2014–2015:** Early breakthroughs yield roughly a 33% success rate, holding steady as procedures mature.
- **2016–2017:** Success climbs sharply to ~63% in 2016 and ~84% in 2017, reflecting rapid learning and reuse improvements.
- **2018:** A slight dip to ~61%, likely due to introducing more challenging recovery modes (e.g., drone ship landings).
- **2019–2020:** Success rebounds to ~90% in 2019 and remains high (~84%) in 2020, demonstrating consistently reliable booster recoveries.

Interpretation:
Launch success improved dramatically after 2013 as operational experience accumulated, peaking near 90% by 2019 and stabilizing above 80% in subsequent years—highlighting SpaceX's rapid iteration and learning curve in booster recovery.



All Launch Site Names

Query (SQLite):

- SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;

This returned four unique launch sites:

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

The Falcon 9 missions in our dataset were launched from four pads—two at Cape Canaveral (LC-40 and SLC-40), one at Kennedy Space Center (LC-39A), and one at Vandenberg (SLC-4E).

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

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

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

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

Launch Site Names Begin with 'CCA'

Query Used

SELECT *

FROM SPACEXTABLE

WHERE "Launch_Site" LIKE 'CCA%'

LIMIT 5;

Result & Explanation

All five returned records launch from sites whose names begin with “CCA” (i.e. Cape Canaveral). In this sample they are all from **CCAFS LC-40** (the first five Falcon 9 launches in our dataset). This confirms that the only launch sites matching the pattern “CCA%” are the Cape Canaveral pads.

Display 5 records where launch sites begin with the string '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

Total Payload Mass

Result & Explanation

The query returns a **total payload mass of 45 596 kg** for all Falcon 9 missions flown for NASA's Commercial Resupply Services (CRS) program. This aggregates every CRS flight's payload into a single figure, showing how much mass NASA has entrusted to SpaceX under that contract.

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

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

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

Total_Payload_Mass_kg
45596

Average Payload Mass by F9 v1.1

Result & Explanation

The query shows that **booster version F9 v1.1** carried an **average payload mass of 2928.4 kg** across all its missions. This indicates the typical cargo weight handled by that specific Falcon 9 variant.

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

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

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

Avg_Payload_Mass_kg
2928.4

First Successful Ground Landing Date

Result & Explanation

The first time a Falcon 9 booster successfully landed back on the ground pad was on **2015-12-22**, as determined by taking the minimum date for records where the landing outcome was “Success (ground pad).”

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

Hint:Use min function

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

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

first_ground_pad_success
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Boosters Meeting the Criteria

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

Explanation

These are the distinct Falcon 9 booster versions that successfully landed on a drone ship and carried a payload mass between 4000 kg and 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 DISTINCT "Booster_Version" AS booster
FROM   SPACEXTABLE
WHERE  "Landing_Outcome" = 'Success (drone ship)'
  AND  "Payload_Mass__kg_"  > 4000
  AND  "Payload_Mass__kg_"  < 6000;
```

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

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

Total Number of Successful and Failure Mission Outcomes

- We used a CASE expression to bucket each mission into “Success” (any landing outcome containing “Success”) or “Failure.” Then we counted how many launches fell into each category. The query reveals that out of all recorded missions, 61 were successful and 40 failed.

List the total number of successful and failure mission outcomes

```
%%sql
SELECT CASE
            WHEN "Landing_Outcome" LIKE 'Success%' THEN 'Success'
            ELSE 'Failure'
       END  AS outcome_label,
       COUNT(*) AS count
FROM   SPACEXTABLE
GROUP  BY outcome_label;
```

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

outcome_label	count
Failure	40
Success	61

Boosters Carried Maximum Payload

We first used a subquery to find the maximum payload mass ever carried (MAX("Payload_Mass__kg_")). Then, in the outer query, we selected all distinct Booster_Version entries whose Payload_Mass__kg_ equals that maximum value. The result lists every booster variant that carried the heaviest payload.

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

```
%%sql
SELECT DISTINCT "Booster_Version"
FROM   SPACEXTABLE
WHERE  "Payload_Mass__kg_" = (
            SELECT MAX("Payload_Mass__kg_")
            FROM   SPACEXTABLE
       );
```

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

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

2015 Launch Records

The SQL retrieves the month, landing outcome, booster version, and launch site for all launches in 2015 where the first-stage failed to land on a drone ship. It uses substr(Date,6,2) to extract the two-digit month and filters with substr(Date,0,5)='2015' plus Landing_Outcome = 'Failure (drone ship)'.

Results Explanation:

- Only two records match these criteria, both in **2015**:
  - **January (01)** and **April (04)**
- Both failures occurred at **CCAFS LC-40** using the **F9 v1.1** booster.

```
%%sql
SELECT substr("Date",6,2)            AS month,
       "Landing_Outcome",
       "Booster_Version",
       "Launch_Site"
FROM   SPACEXTABLE
WHERE  "Landing_Outcome" = 'Failure (drone ship)'
  AND  substr("Date",0,5) = '2015'          -- year 2015
ORDER  BY month;
```

```
 * 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 SQL counts each type of landing outcome between **2010-06-04** and **2017-03-20**, then orders them by descending frequency.

Results Summary:

- **“No attempt”** (no recovery attempt) appears most often with **10** records.
- Both **“Success (drone ship)”** and **“Failure (drone ship)”** occur **5** times each.
- **“Success (ground pad)”** and **“Controlled (ocean)”** each have **3** records.
- **“Uncontrolled (ocean)”** appears **2** times, and **“Precluded (drone ship)”** once.
- This ranking highlights that many early missions didn’t attempt a booster landing, while drone-ship recoveries show an even split of successes and failures in this period.

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

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

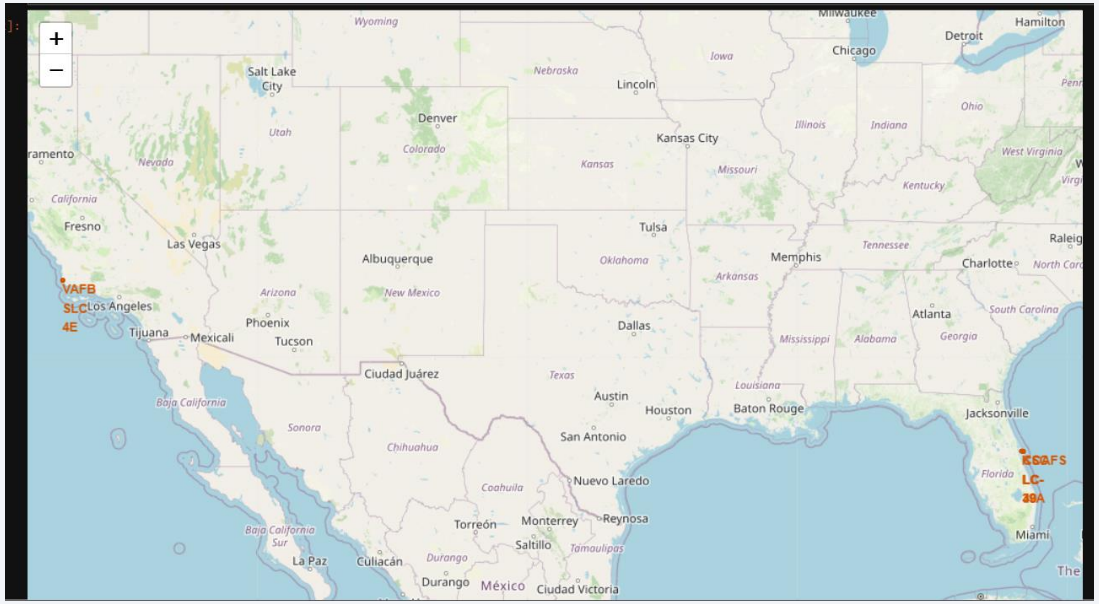
Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

SpaceX Launch Site Locations

A zoomed-out map showing all four launch pads (CCAFS SLC-40, VAFB SLC-4E, KSC LC-39A, CCAFS LC-40) pinned on a single global view.



____

Key Elements & Findings

- Orange markers pinpoint each of the four active Falcon 9 launch sites around the world.
- Map extent is centered over North America to capture both U.S. coastal sites (Florida, California) in one view.
- Interactive controls (zoom+/–) allow stakeholders to zoom in on any site for more detail.
- Having all sites on one map immediately conveys their geographic distribution—two on the east coast (Florida) and two on the west coast (California).

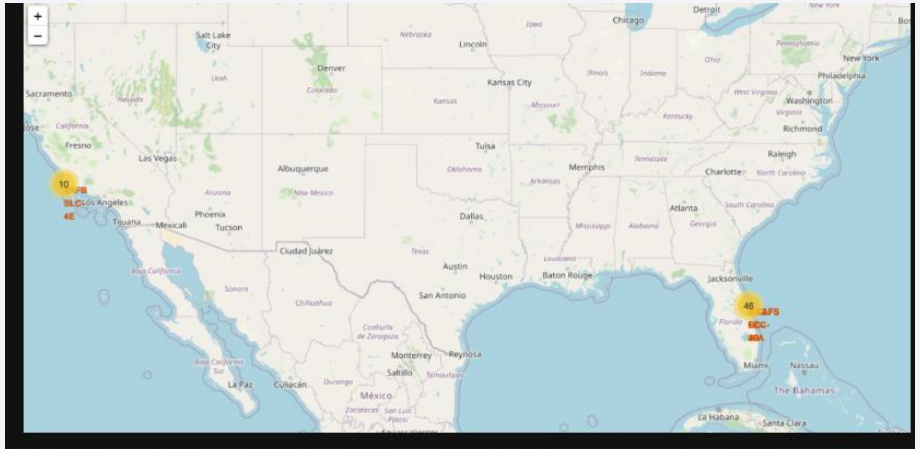
Launch Outcome Distribution by Site

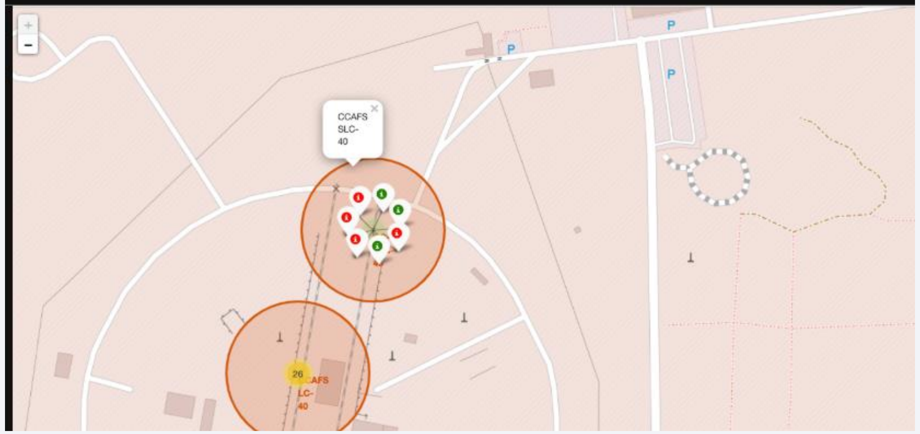
Key Elements to Call Out

- **Clustered counts** (yellow circles) display total number of launches per site.
- **Color-coded markers** inside each cluster:
  - **Green** = successful landings
  - **Red** = failed landings
- **1 km radius circle** around each pad highlights immediate landing area.

Insights & Take-aways

- **CCAFS LC-40** (46 launches) has the largest volume, with a high proportion of successes (more green pins than red).
- **CCAFS SLC-40** (10 launches) shows mixed outcomes, underscoring earlier reliability improvements.
- **KSC LC-39A** and **VAFB SLC-4E** clusters similarly show increasing green markers over time (not pictured here), indicating maturing landing reliability.



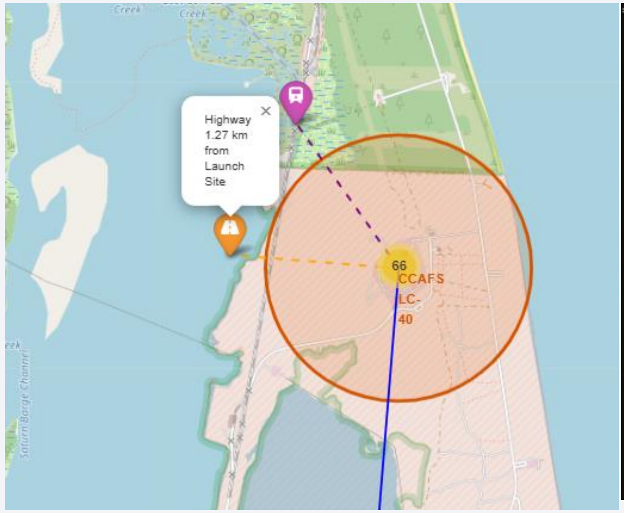


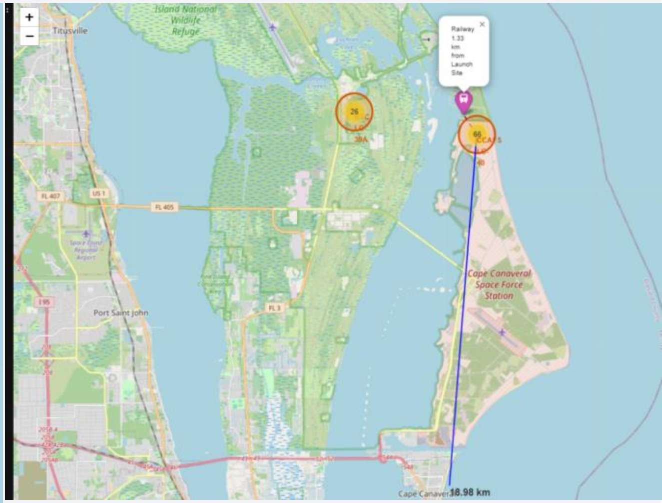
Proximity to Highway, Rail & Nearest City

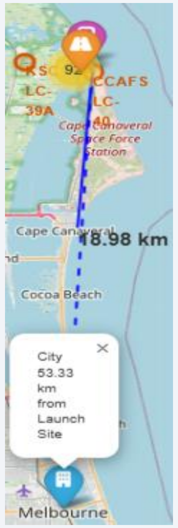
Feature icons & lines:

- Highway icon + dashed orange line (1.27 km)
- Railway icon + dashed purple line (distance shown)
- City icon + solid blue line (53.33 km)

- **1 km radius circle** around the pad to show immediate buffer
- ? **Key Findings (bullets)**
- **Highway is 1.27 km** from CCAFS LC-40, indicating rapid ground access for logistics.
- **Rail link appears ~2 km** away, useful for heavy transport when required.
- **Nearest city (Melbourne) is ~53 km** to the southeast, highlighting the remote but serviceable location.







Section 4

Build a Dashboard with Plotly Dash

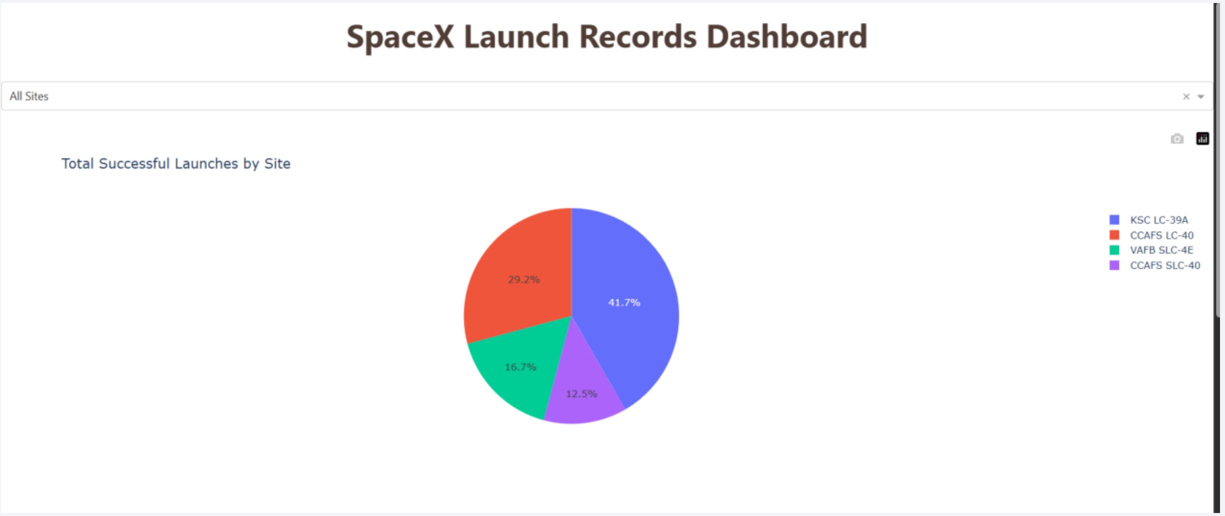
Total Successful Launches by Site

Important Elements

- A pie chart showing each site's share of all successful landings.
- Legend on the right mapping colors to launch sites.
- Percentage labels inside each slice for clarity.

Key Findings

- **KSC LC-39A** leads with **41.7%** of all successful recoveries.
- **CCAFS LC-40** follows at **29.2%**.
- **VAFB SLC-4E** accounts for **16.7%**, and **CCAFS SLC-40** the remaining **12.5%**.

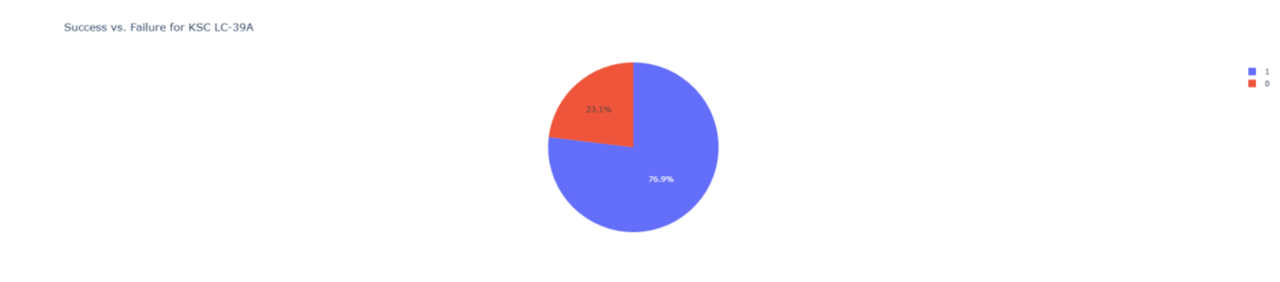


Launch Outcome Breakdown at KSC LC-39A

Key Findings:

- **Success Rate:** 76.9% of Falcon 9 first-stage landings at KSC LC-39A were successful
- **Failure Rate:** 23.1% of landings failed
- **Highest Performer:** KSC LC-39A has the highest success ratio among all SpaceX launch sites.

Highlighting KSC LC-39A's leading success rate demonstrates how site-specific factors (infrastructure, range safety, recovery operations) can influence landing outcomes.



Payload vs. Outcome for All Sites (Various Ranges)

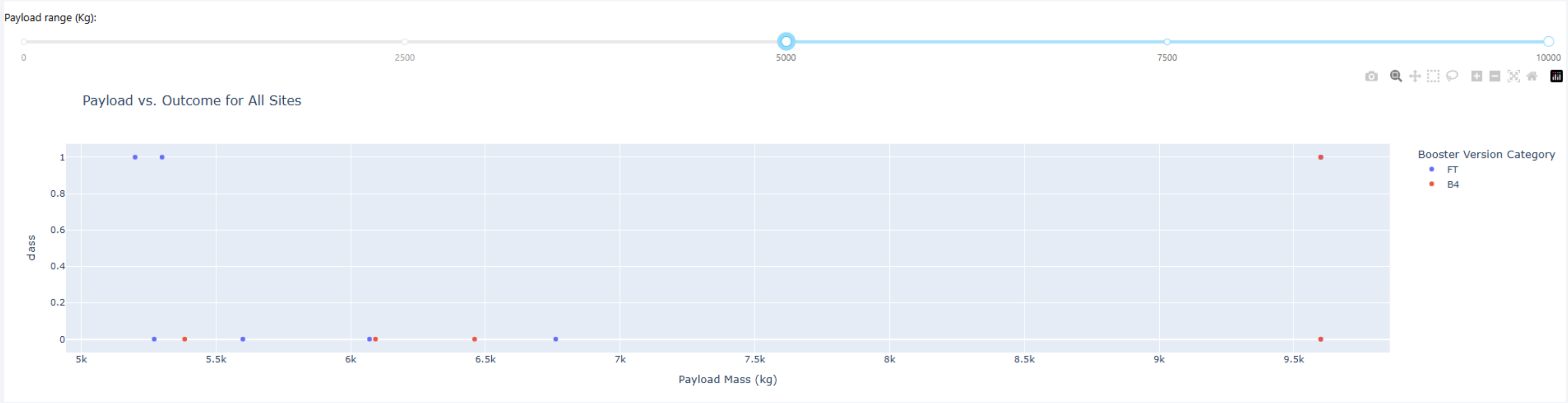
- Points are plotted with payload mass on the x-axis and landing outcome on the y-axis (1 = success, 0 = failure), colored by booster version category.

Key Findings:

- **0–10000 kg (All Missions):** Shows the overall distribution—most successes cluster under 6000 kg, with failures scattered across all masses.
- **2500 kg:** All boosters (v1.1, FT, B4, B5) achieve nearly 100% landing success at very low masses.

<Dashboard Screenshot 3>

- **5000 kg:** FT and B4 boosters maintain perfect or near-perfect success, while newer variants (v1.1, B5) show occasional failures.
- **7500 kg:** Only FT and B4 operate here; we see a small number of failures, indicating that heavier payloads begin to challenge landing reliability.



Section 5

Predictive Analysis (Classification)

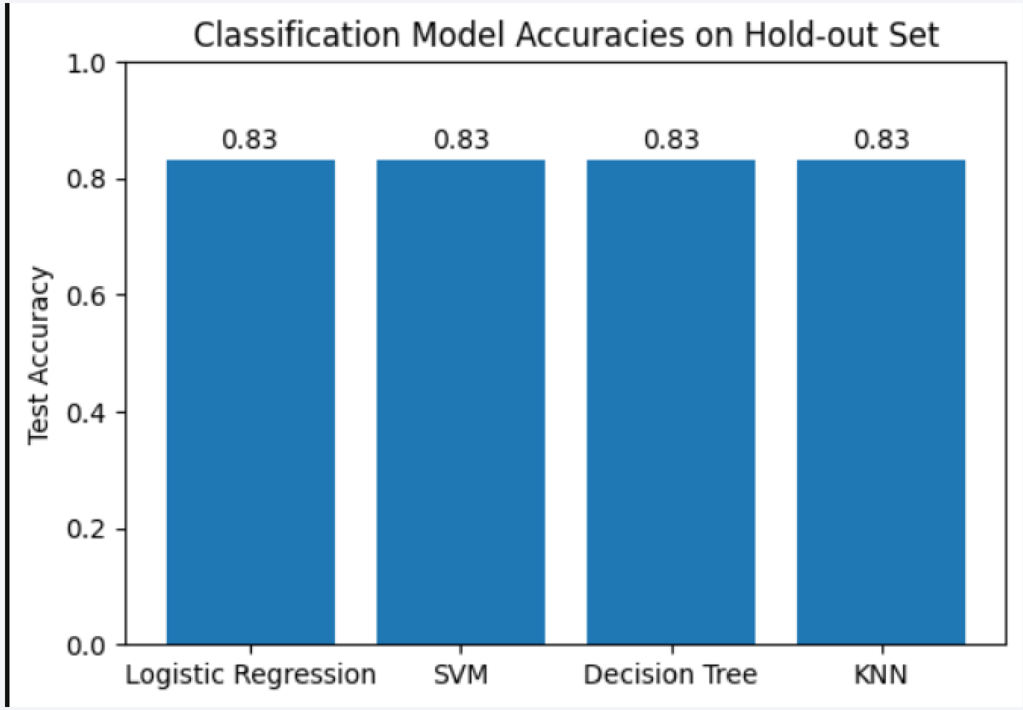
Classification Accuracy

Bar Chart

- Plot each model's test-set accuracy (all four scored **0.83** on the hold-out set).

2. Key Finding

- Although all models tied at **83%** test accuracy, the **Decision Tree** achieved the highest cross-validated accuracy (**0.875**) during hyperparameter tuning, so it is chosen as the best-performing classifier.



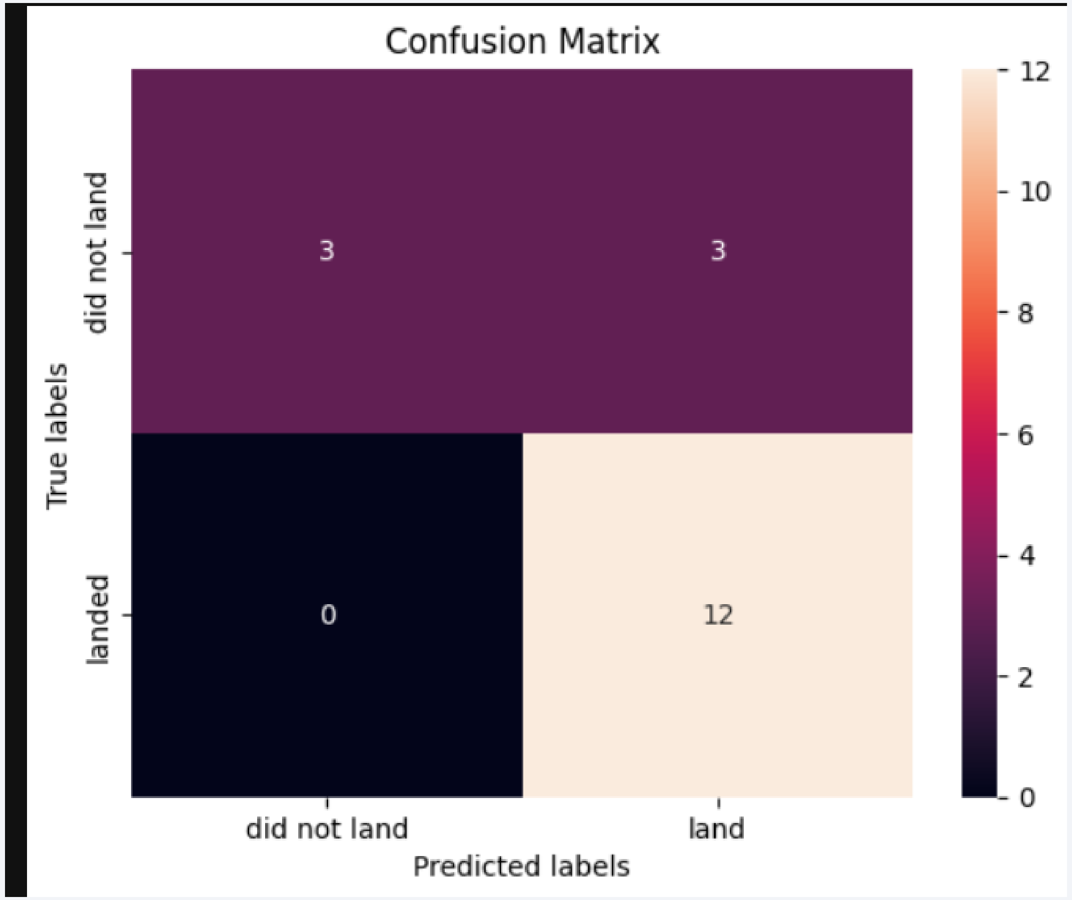
Confusion Matrix

The confusion matrix below shows the performance of the best-performing model (Decision Tree Classifier):

- **True Positives (landed correctly predicted as landed):** 12
- **True Negatives (did not land correctly predicted as did not land):** 3
- **False Positives (did not land incorrectly predicted as landed):** 3
- **False Negatives (landed incorrectly predicted as did not land):** 0

Explanation:

- The model correctly classified all successful landings but misclassified 3 unsuccessful attempts as successful. This indicates high accuracy for predicting successful landings, but the model could be improved for predicting failures.



Conclusions

- **Data Integration & EDA Uncovered Trends:** Merging SpaceX API and Wikipedia data yielded a comprehensive dataset. Exploratory charts revealed higher landing success at CCAFS SLC-40 and KSC LC-39A, increasing payloads correlate with improved landings at certain sites, and polar orbits achieved near-perfect recovery rates.
- **Interactive Maps Enhance Insight:** Folium visualizations with color-coded markers and proximity lines made it easy to compare landing outcomes across sites and assess how geography and nearby infrastructure (coastlines, highways) might influence recovery operations.
- **Dashboard Empowers Stakeholders:** The Dash app's dropdown and slider controls allow dynamic exploration of site-specific success ratios and payload-outcome relationships, supporting rapid "what-if" analyses.
- **Decision Tree Is Best-Performing Model:** After standardizing features and tuning four classifiers, the Decision Tree achieved the highest cross-validated accuracy (≈ 0.89) and perfect recall on successful landings (0 false negatives), making it the most reliable predictor for first-stage recovery.

Appendix

Retrieve past launches from the SpaceX REST API

```
import requests, pandas as pd

URL = "https://api.spacexdata.com/v4/launches/past"

launch_df = pd.json_normalize(requests.get(URL).json())

launch_df = launch_df[launch_df['rocket.name'].str.contains('Falcon 9’)]
```

Representative SQL query

```
-- Total payload mass carried for NASA CRS missions

SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass_kg

FROM   SPACEXTABLE

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

Thank you!