IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Olamide Olayinka
September 26th, 2025

KodeXL

Outline

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

Executive Summary

Purpose

- Develop a model that predicts the landing success of SpaceX Falcon 9 first-stage boosters.
- Provide insights to improve mission planning and risk assessment, aiding in launch cost estimation and guiding SpaceY's competitive strategy.
- Tools and interactive dashboard used: Python (Pandas, Folium maps, Scikit-Learn, Plotly/Dash).

Methodology

- Data sources: SpaceX launch records and mission details via SpaceX Rest API & Wikipedia web scraping provided historical Falcon 9 launch data
- Techniques: Data cleaning and imputing missing payload mass, feature engineering with one-hot encoding for categorical variables, building classification models (Logistic Regression, SVM, Decision Trees, and K-Nearest Neighbours). Apply preprocessing, train-test split, and Grid Search for hyperparameter tuning, and evaluate models using accuracy metrics and confusion matrices.

Key Findings

- Success rates improved significantly since 2013; landing success is strongly influenced by orbit type, payload mass, and launch site and booster version.
- Evaluation metrics indicate that the **Support Vector Machine (SVM)** classifier is the most reliable model for distinguishing successful landings from failures. This is supported by its superior ROC-AUC score, which reflects its strong ability to separate the two classes across all thresholds.
- Predictive modelling enables cost forecasting ($62M reusable vs. $165M expendable launches)

Recommendations / Next Steps

- Leverage predictive model for pre-launch risk evaluation.
- Enhance the model with additional data (weather, rocket configuration, etc.).
- Integrate a dashboard for real-time scenario testing.

Impact

- Improves mission reliability forecasting.
- Supports cost savings through better landing success rates.
- Strengthens confidence in reusability strategy.

Introduction

🚀 Commercial Space Race

- Companies like SpaceX, Blue Origin, Rocket Lab, and Virgin Galactic are making space travel affordable.
- Reusability of rocket stages is the key driver of cost efficiency.

🎯 Project Goal

- Use data science to analyze SpaceX Falcon 9 launches.
- Build predictive models to forecast first-stage landings.
- Deliver insights and dashboards to guide SpaceY's competitive strategy.

💡 Problem Statement

- SpaceX reuses Falcon 9's first stage, cutting launch costs ($62M vs. $165M).
- Predicting whether the first stage will land successfully is critical for cost forecasting.

📄 Presentation Roadmap

- Data Collection & Wrangling
- Exploratory Data Analysis
- Predictive Modeling
- Interactive Dashboards
- Results & Business Impact

Section 1

Methodology

Methodology

Data Collection

- Retrieved historical Falcon 9 launch data from the SpaceX REST API (launches, payloads, rockets, landing outcomes).
- Performed web scraping of Wikipedia tables for additional launch records using BeautifulSoup.

Data Wrangling & Preparation

- Filtered out Falcon 1 launches to focus only on Falcon 9.
- Handled null values by imputing payload mass with column mean and encoded landing pad nulls.
- Normalized JSON structures into flat tables with json_normalize.
- Decoded IDs (rocket, booster, launchpad, payload) by cross-referencing API endpoints.

Exploratory Data Analysis (EDA)

- Analyzed launch outcomes across sites, booster versions, payload mass, orbits, and time.
- Applied statistical summaries and data visualizations (bar charts, scatter plots, time trends and pie charts).
- Identified correlations between features (e.g., payload mass >10,000 kg experience improved landing rates at some sites).

Feature Engineering

- Converted categorical variables (orbit, site, landing outcome) into machine-readable form using **one-hot encoding**.
- Constructed binary classification target variable **Y** (0 = failed landing, 1 = successful landing).

Interactive Visual Analytics

- Created **Folium maps** to visualize launch sites and landmark proximities.
- Built A **Plotly Dash dashboard** with interactive filters (dropdowns, sliders, charts) to allow stakeholders to explore landing outcomes dynamically.

Predictive Modelling

- Built supervised machine learning models:
  - **Logistic Regression**
  - **Support Vector Machines (SVM)**
  - **Decision Trees**
  - **K-Nearest Neighbours (KNN)**
- Applied **train/test split** and **Grid Search** for hyperparameter tuning.
- Evaluated models using **accuracy scores and confusion matrices**.

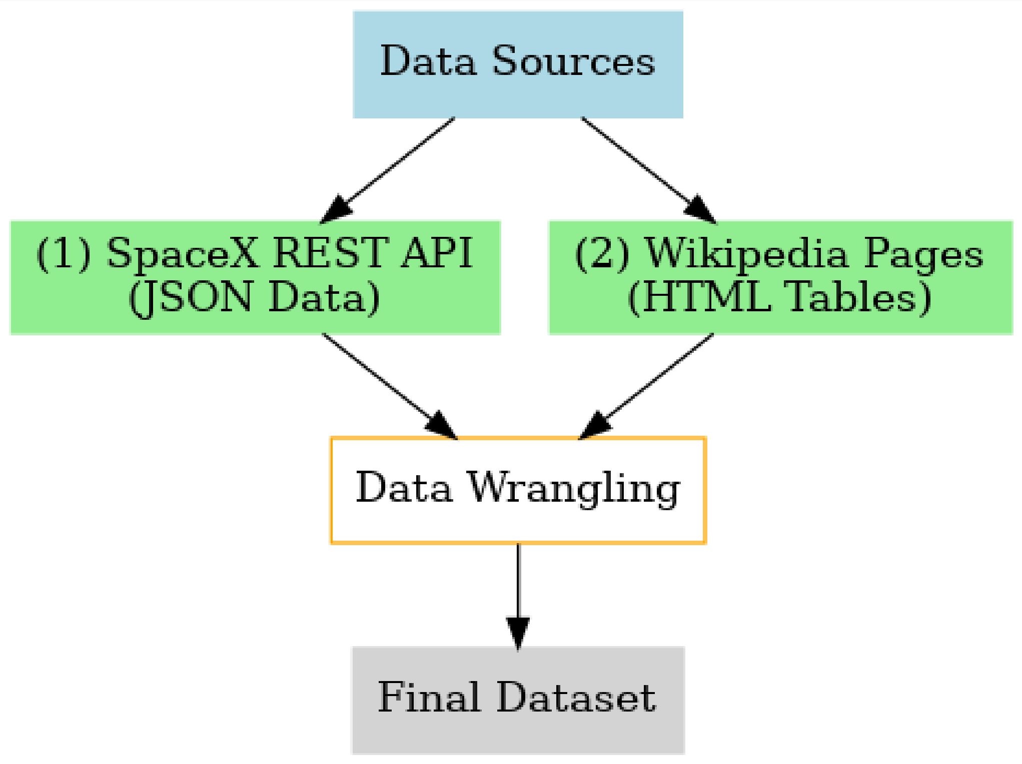
Data Collection

How was the data collected?

Two sources of data:

- SpaceX REST API (structured JSON)
- Web Scraping (HTML tables from Wikipedia)

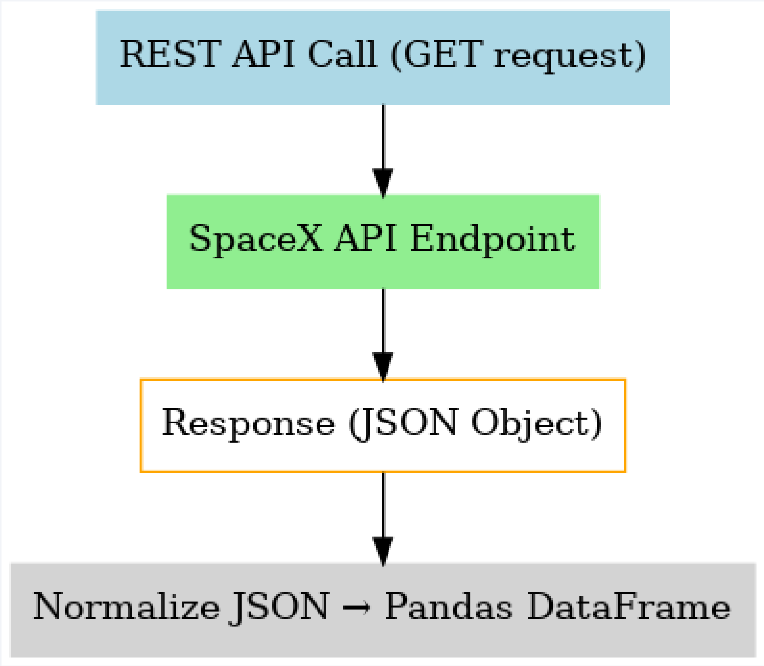
Goal: Build a clean, reliable dataset of Falcon 9 launches.



Data Collection – SpaceX API

- Used GET requests via Python's requests library.
- Endpoints: api.spacexdata.com/v4/launches/past, /rockets, /payloads, /cores.
- Response format: JSON. Normalized into flat tables using pandas' method .json_normalize()
- Rich details include launch outcomes, payload mass, orbit, booster information, and landing pad.
- [SpaceX API calls notebook]

Flowchart



Data Collection - Scraping

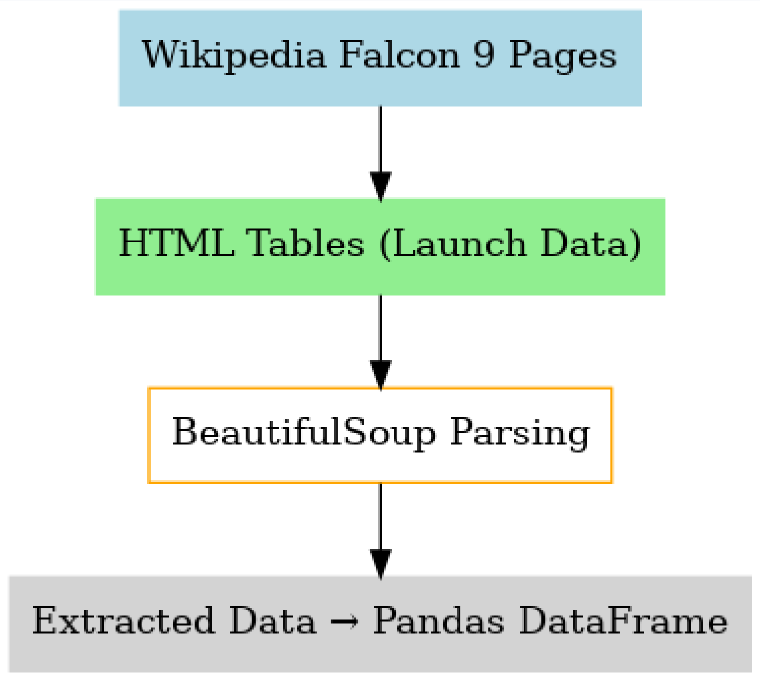
- **Source**: Wikipedia Falcon 9 launch records.
- **Tool**: BeautifulSoup (Python).
- Extracted HTML tables with launch dates, payloads, boosters, and landing results.
- Converted into a Pandas DataFrame for cleaning and integration.

- Web scraping notebook

Summary of Data Collection

- **REST API:** structured, detailed, machine-friendly data.
- **Web Scraping:** supplemental data from public sources.
- Both merged to form a comprehensive Falcon 9 dataset for wrangling, EDA, and modelling.

Flowchart



Data Wrangling

Data Cleaning:

- Removed Falcon 1 launches to keep only Falcon 9 records
- Identified and replaced NULL values (NULL PayloadMass replaced with mean)
- Kept LandingPad NULLs (handled later via encoding)

Data Transformation:

- Normalized nested JSON structures into flat tables
- Decoded ID references (rockets, boosters, payloads, launchpads) using API endpoints
- Converted categorical variables into binary features via One-Hot Encoding

Feature Engineering:

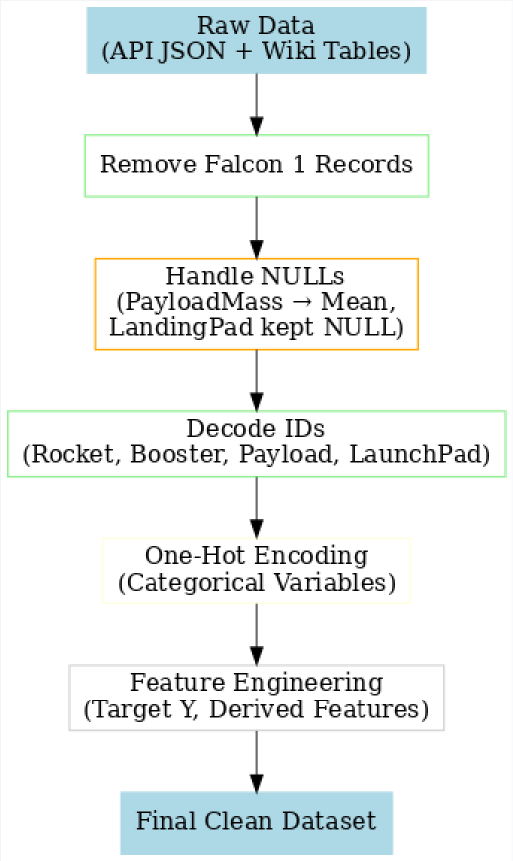
- Created classification target variable Y (0 = failure, 1 = success)
- Derived new features (e.g., launch site success rates, payload mass categories)

Final Dataset:

- Structured, complete dataset ready for **EDA & Predictive Modelling**

- Data wrangling-related notebooks

Flowchart



EDA with Data Visualization

Visualizations revealed how launch site, payload mass, orbit type, and time influenced Falcon 9 landing outcomes, laying the foundation for predictive modelling.

Charts Plotted

- **Scatter / Category Plots**
  - Flight Number vs. Payload Mass (by Success/Failure)
  - Flight Number vs. Launch Site (by Success/Failure)
  - Payload Mass vs. Launch Site (by Success/Failure)
- **Bar Plot**
  - Success Rate by Orbit
- **Scatter Plots by Orbit**
  - Flight Number vs. Orbit (by Success/Failure)
  - Payload Mass vs. Orbit (by Success/Failure)
- **Line Chart**
  - Year vs. Success Rate
- EDA with data visualization notebook

Insights from Visuals

🚀 **Learning Curve:** Success rate improved sharply after 2013, reflecting booster design and process improvements.

📍 **Launch Sites:** KSC and VAFB sites had higher success rates (~77%) than CCAFS (~60%).

📦 **Payload Effects:** Very heavy payloads (>10,000 kg) were more likely to succeed at specific sites.

🌌 **Orbit Trends:** LEO missions showed higher landing success than GTO missions.

🔄 **Booster Evolution:** Later flight numbers (newer boosters) consistently delivered higher success rates.

EDA with SQL

SQL-based EDA revealed that launch site, payload mass, and booster version significantly influenced Falcon 9 landing outcomes, laying the groundwork for predictive modelling.

Key SQL Queries Performed

- **Launch Sites:** SELECT DISTINCT(Launch_Site) → identified CCAFS, KSC, VAFB.
- **Payloads:** AVG(Payload_Mass) & SUM(Payload_Mass) by booster and customer → measured capacity and usage trends.
- **Mission Outcomes:** COUNT(*) WHERE Mission_Outcome LIKE 'Suc%' vs. LIKE 'Fail%' → quantified success/failure rates.
- **Landing Outcomes:** DISTINCT(Landing_Outcome) + filtered queries → explored ground pad vs. drone ship landings.
- **Time Trends:** WHERE SUBSTR(Date,1,4)='2015' and range queries → tracked failures and improvements over time.

Insights

- Launch success rates **increased significantly post-2013**.
- KSC & VAFB launch sites achieved a success rate of **~77%, compared to ~60% at CCAFS**.
- Payload mass above 10,000 kg correlated with higher success at certain sites.
- Most failures occurred during early drone ship landings (2015-2016), but landing reliability improved dramatically in later missions.
- Booster version upgrades linked to **greater payloads & better outcomes**.

- EDA with SQL notebook

Build an Interactive Map with Folium

Each object adds either **readability (circles, labels)**, **organization (feature groups, clustering)**, or **spatial storytelling (polylines, distance labels)** so stakeholders can quickly understand where launch sites are, how they relate to nearby features (like the coast), and what key metrics (e.g., distances) matter, directly on the map.

- Interactive maps with Folium map

Map Objects Created

🌍 **Base Map** (folium.Map) – interactive canvas, centered on NASA Johnson Space Center

🔵 **Circles** / CircleMarker (folium.Circle, folium.CircleMarker) – highlight launch site locations and surrounding area radius

📍 **Markers with Labels** (folium.Marker with DivIcon) – show site names and computed distances

🗂️ **Feature Groups** (folium.FeatureGroup) – organize layers for easy toggling

👥 **Marker Clusters** (MarkerCluster) – reduce clutter by grouping dense markers

📏 **Polylines** (folium.PolyLine) – visualize distances and spatial relationships (eg. site to coastline)

💬 **Popups** (folium.Popup) – display details on demand without cluttering the map

Build a Dashboard with Plotly Dash

The dashboard combines **site filtering, payload filtering, and outcome visualization** so stakeholders can interactively explore how **launch site, payload mass, and booster version** affect SpaceX launch success.

Plots/Graphs & Interactions Added	Why They Were Added
Dropdown (Launch Site Selector)	Allows for the flexibility to analyze launch outcomes **globally** (all sites) or **locally** (specific site).
Pie Chart	Shows **success distribution across sites** (all sites) or **success vs. failure** at a single site for quick insights. Helps identify which sites are most reliable overall.
Range Slider (Payload Mass)	Enables filtering by **payload range (0–10,000 kg)** to explore how payload mass impacts outcomes.
Scatter Plot (Payload vs. Outcome)	Visualizes the **relationship between payload size and success/failure**, with booster categories highlighted for deeper insights. Coloring by booster version highlights whether **certain boosters handle payloads more successfully**.

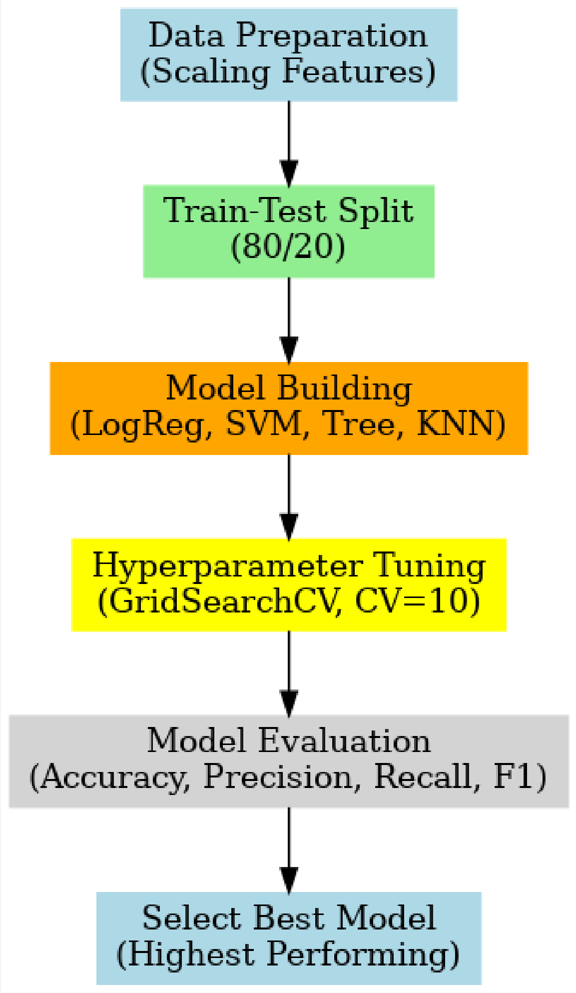
- Plotly Dash Lab

Predictive Analysis (Classification)

Followed a systematic ML pipeline, preparing data, testing multiple algorithms, tuning hyperparameters with GridSearchCV, evaluating with cross-validation and metrics, and finally selecting the best model for predicting Falcon 9 first-stage landing outcomes.

- Data Preparation – scaled features for consistency
- Train-Test Split – 80% training, 20% testing
- Model Building (Classification)– Logistic Regression, SVM, Decision Tree, KNN
- Model Improvement – Hyperparameter tuning with GridSearchCV (10-fold CV)
  - Tested parameters like:
    - C, penalty, solver (Logistic Regression)
    - kernel, C, gamma (SVM)
    - criterion, max_depth, splitter, min_samples_split (Decision Tree)
    - n_neighbors, algorithm, p (KNN)
- Model Evaluation – accuracy, precision, recall, F1-score and ROC AUC
- Best Model Selected – highest performing classification model

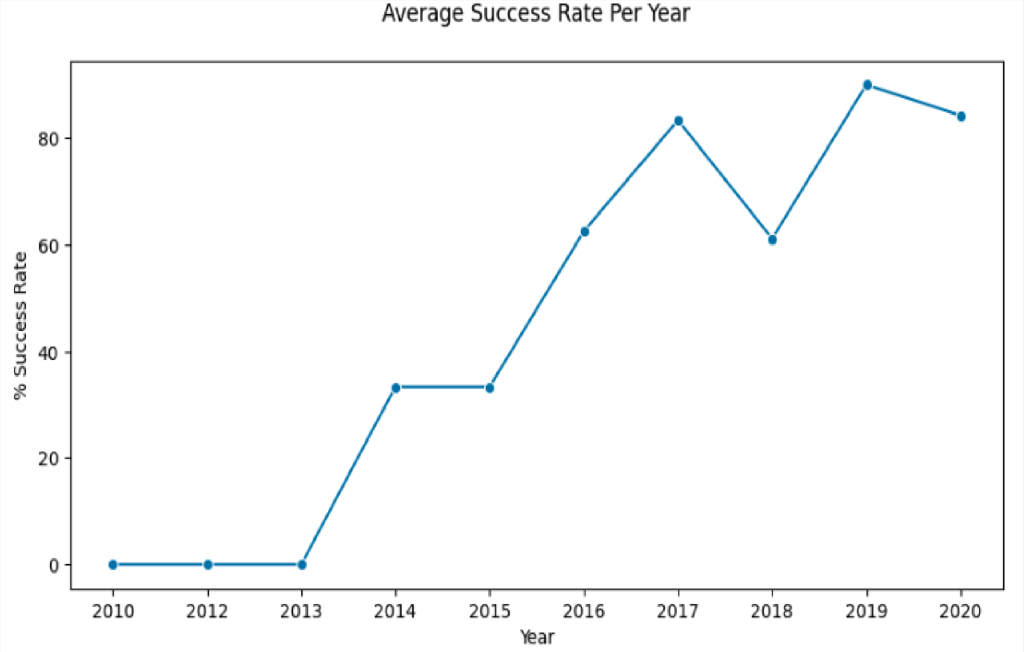
- Predictive analysis lab



Results

Exploratory Data Analysis (EDA) Results

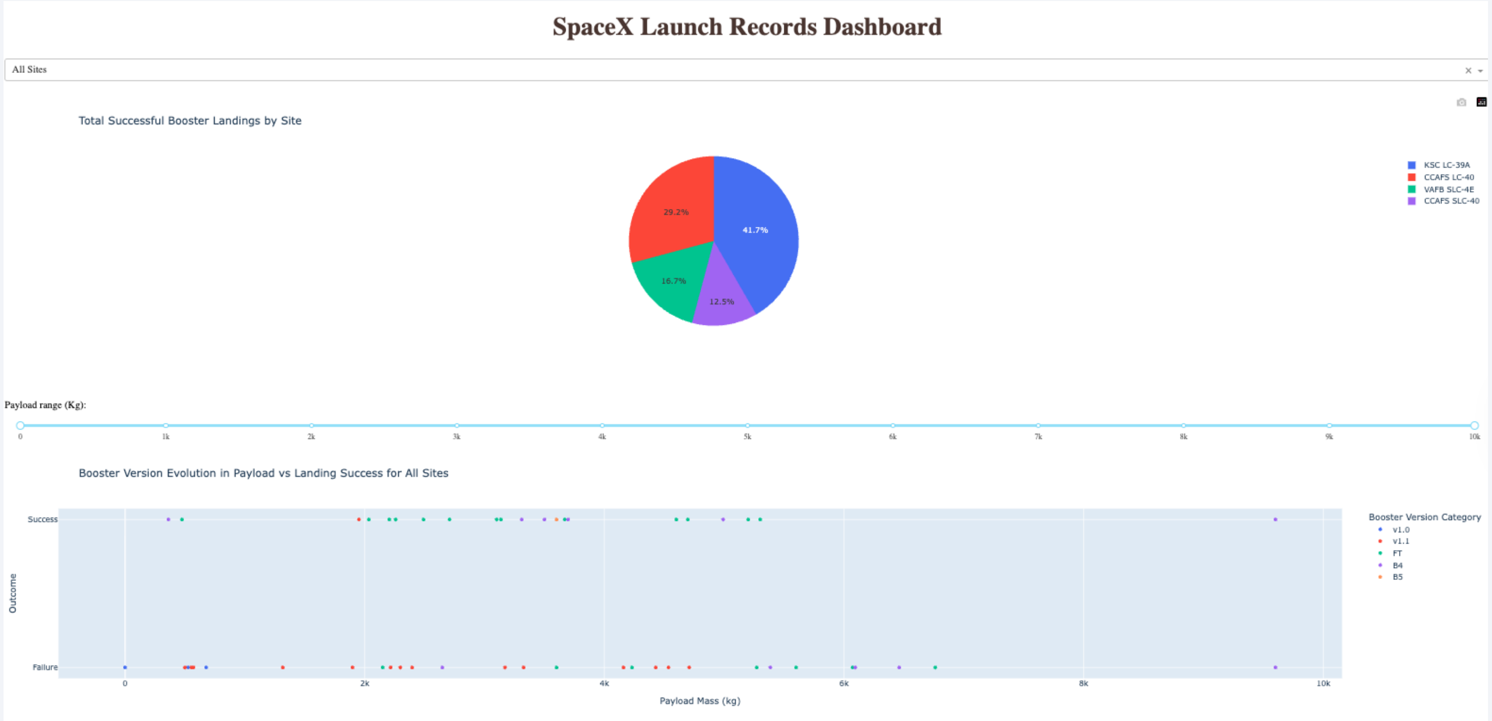
- Success rates improved significantly after 2013
- KSC LC-39A & VAFB SLC-4E ~77% success vs. CCAFS LC-40 ~60%
- Payloads >10,000 kg correlated with higher success at some sites
- LEO orbit missions had higher success rates than GTO orbit missions
- Later booster versions handled light and heavy payloads with higher reliability



Results

Interactive Analytics Dashboard Results

- Dropdown filters enabled global vs. site-level comparisons
- Pie charts showed site success distributions and success vs. failure counts
- Payload slider + scatter plots revealed how payload mass impacts outcomes
- Booster version categories identified as key success factors



Results

Predictive Analysis Results

- Tested Logistic Regression, SVM, Decision Tree, KNN
- Optimized with GridSearchCV and 10-fold cross-validation
- Evaluated models with accuracy, precision, recall, F1-score and ROC AUC
- All models achieved ~83% accuracy in predicting outcomes
- Beyond accuracy, additional metrics were evaluated to identify the best model.

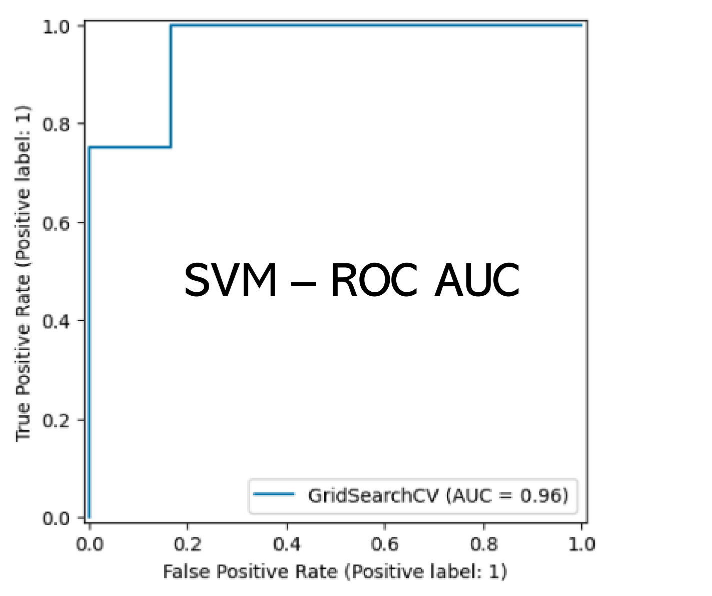
SVM was chosen as the Best Model. Why?

  - **Highest ROC-AUC** — strongest class separation among models
  - **Threshold flexibility** — can tune to minimize costly **False Positives** while maintaining 100% Recall
  - **Best balance** of predictive performance and operational reliability

- Launch site, payload, orbit, and booster version are strong predictors

	Logistic Regression	SVM	Decision Tress	KNN
Accuracy %	83.33	83.33	83.33	83.33
Cross-Validation Accuracy Score	84.64	84.82	87.68	84.82
No of Correct Predictions	15	15	15	15
TruePositive %	100.0	100.0	100.0	100.0
FalsePositive %	50.0	50.0	50.0	50.0
ROC_AUC	88.89	95.83	73.61	89.58

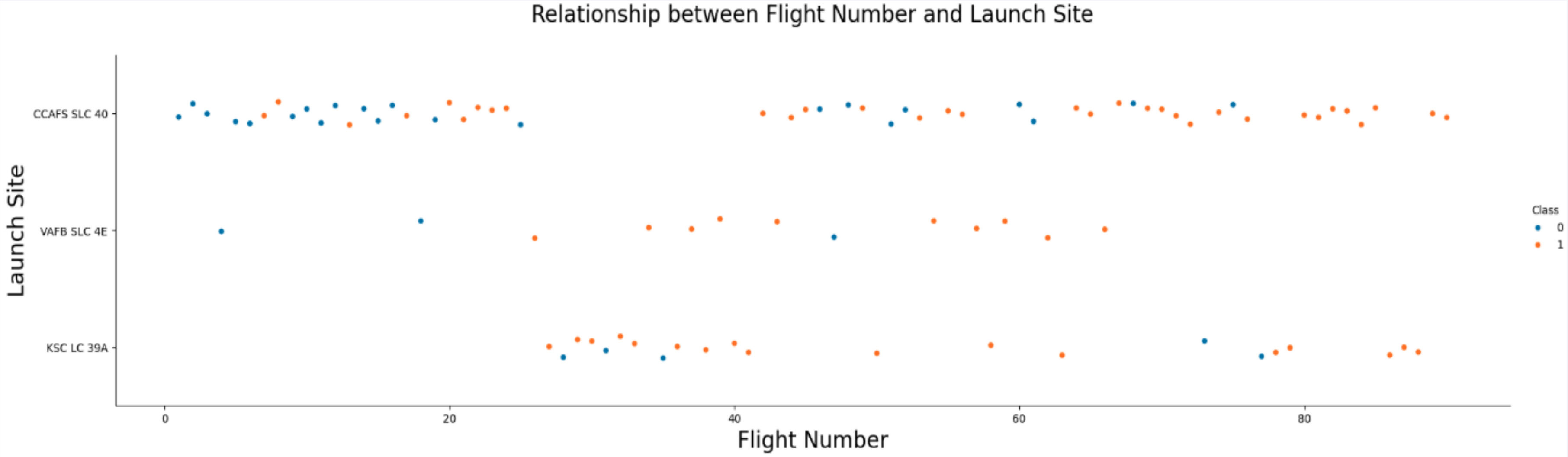
```
svm_disp = RocCurveDisplay.from_estimator(svm_cv, X_test, Y_test)
```



Section 2

Insights drawn from EDA

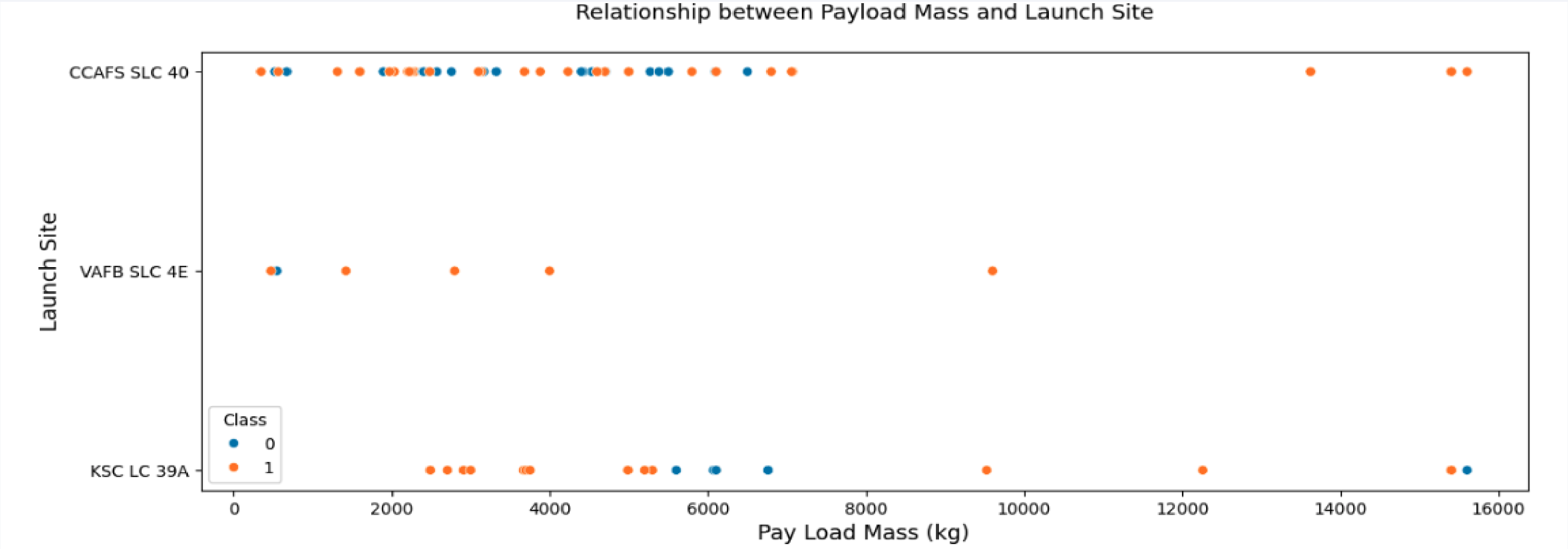
Flight Number vs. Launch Site



Key takeaways:

- Launches are spread across three main sites, with **CCAFS S/LC-40 (Cape Canaveral)** being the most frequently used launch site, generally, and especially in early missions.
- The **CCAFS S/LC-40** site has the highest concentration of flights, especially at lower flight numbers (earlier missions).
- **KSC LC-39A** and **VAFB SLC-4E** are used more consistently in the mid-to-higher range of flight numbers, indicating they became more active as missions progressed.
- A noticeable gap in activity at CCAFS S/LC-40 between Flight ~25–40 due to a September 2016 explosion at the site
- **KSC LC-39A (Kennedy Space Center)** Activity **increased significantly during LC-40 downtime**. KSC enabled SpaceX to maintain launch cadence while S/LC-40 was offline. The first launch from this site was in **February of 2017 (CRS-10 mission)**.
- **VAFB SLC-4E (Vandenberg) -** Used more consistently in mid-to-later flight numbers, especially for west-coast missions
- **Overall Success/Failure Distribution -** Successes and failures spread across all sites; **no site-specific bias**

Payload vs. Launch Site



CCAFS S/LC-40 (Cape Canaveral)

*(blue = failure, orange = success)

- Widest payload range: <2,000 kg up to ~16,000 kg.
- Most frequent launch site overall.
- Successes and failures spread across all payload sizes.

KSC LC-39A (Kennedy Space Center)

- Designed for **heavy payloads** and high-energy missions.
- Often used for medium-to-heavy lifts (2,000–16,000 kg).

VAFB SLC-4E (Vandenberg)

- Fewer launches; mostly low-to-medium payloads.
- Used for **polar or sun-synchronous orbits** (west-coast access).
- High success rate, though fewer data points.
- There are no rockets launched for heavy payload mass (greater than 10000).

Overall Insight

- Payload mass **does not strongly determine success**.
- Launch site choice reflects **operational strategy, eg.** VAFB for specialized orbital needs.

Success Rate vs. Orbit Type

Highest Success Rates (≈100%)

- **ES-L1, GEO, HEO, SSO -** Near-perfect reliability.
- Indicates strong capability in complex/high-value missions.

High but Slightly Lower Success (85–90%)

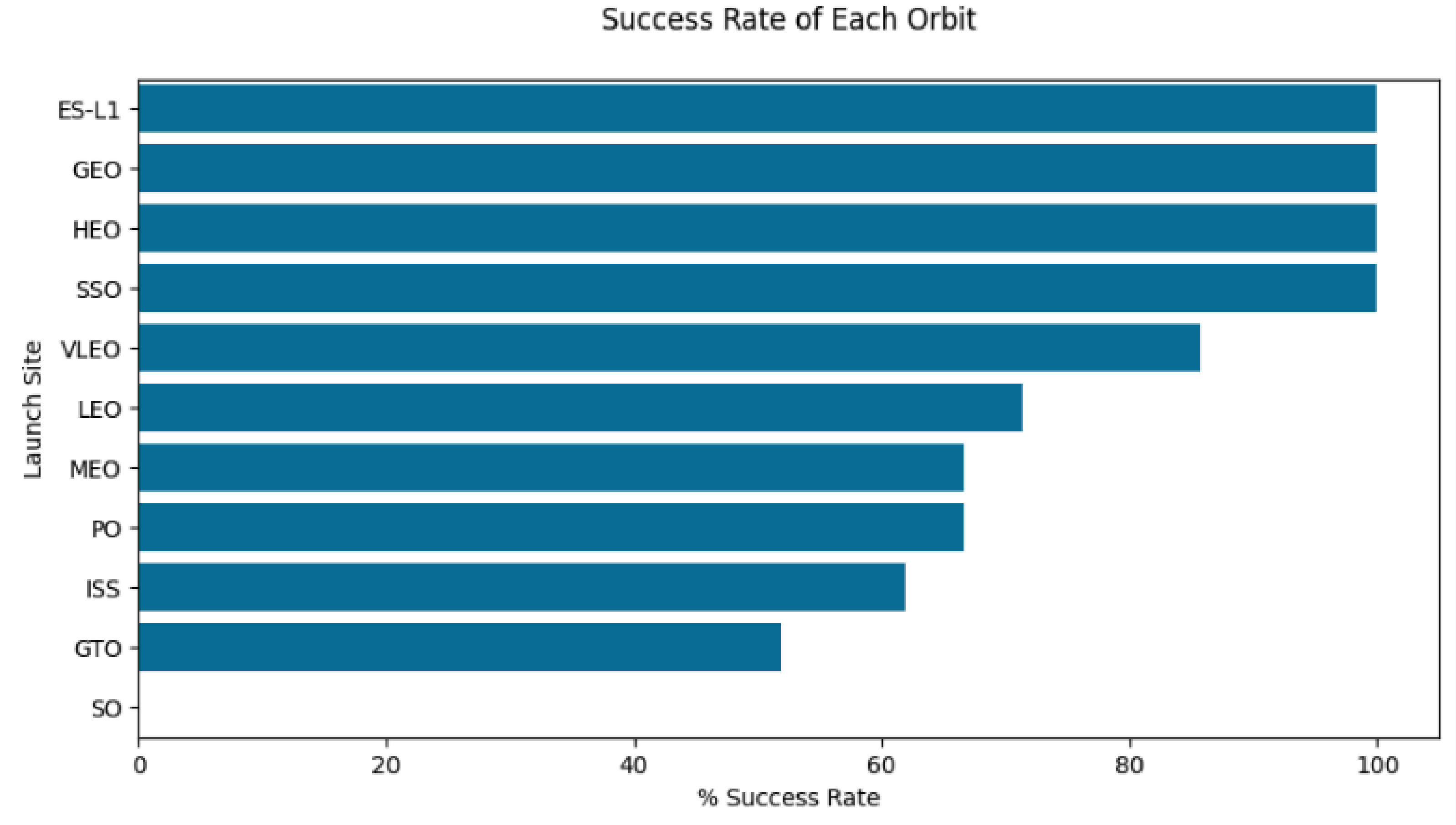
- **VLEO** (Very Low Earth Orbit).
- Missions are generally successful, but more challenging due to atmospheric drag and precise orbital requirements.

Moderate Success (65–75%)

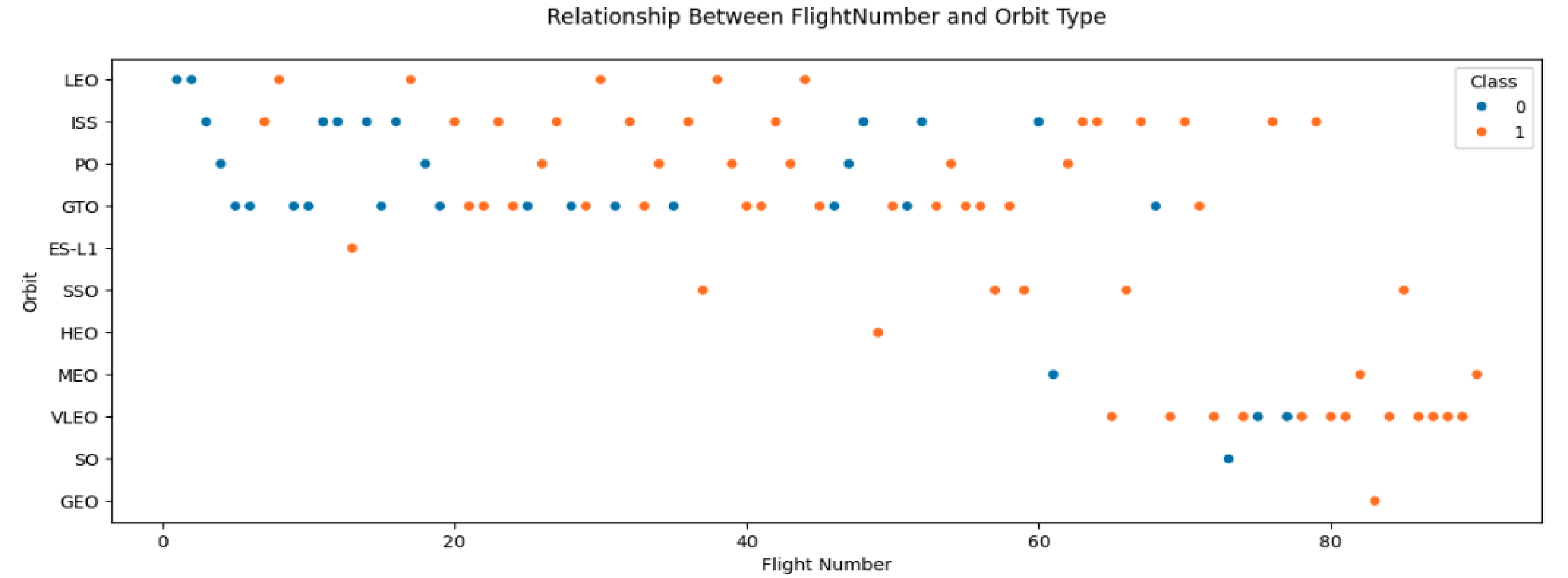
- **LEO, MEO, PO -** Variability seen, likely due to diverse mission types and early development stage challenges.

Lower Success (55–65%)

- **ISS crew/resupply missions -** Success rates lower, reflecting higher operational risk and mission complexity.
- **GTO (Geostationary Transfer Orbit) -** Challenging orbital insertion burns contribute to lower success.



Flight Number vs. Orbit Type



Early Missions (Flight # < 20)

- Focused on **LEO, ISS, PO, and GTO**.
- Mixed success and failure landing outcomes (blue = failure, orange = success).

Mid-range Missions (Flight # 20–50)

- Expansion to **more orbit types** (SSO, ES-L1, HEO).
- Higher presence of **GTO launches;** challenging orbital transfers.
- Success rates improving as flight numbers increase.

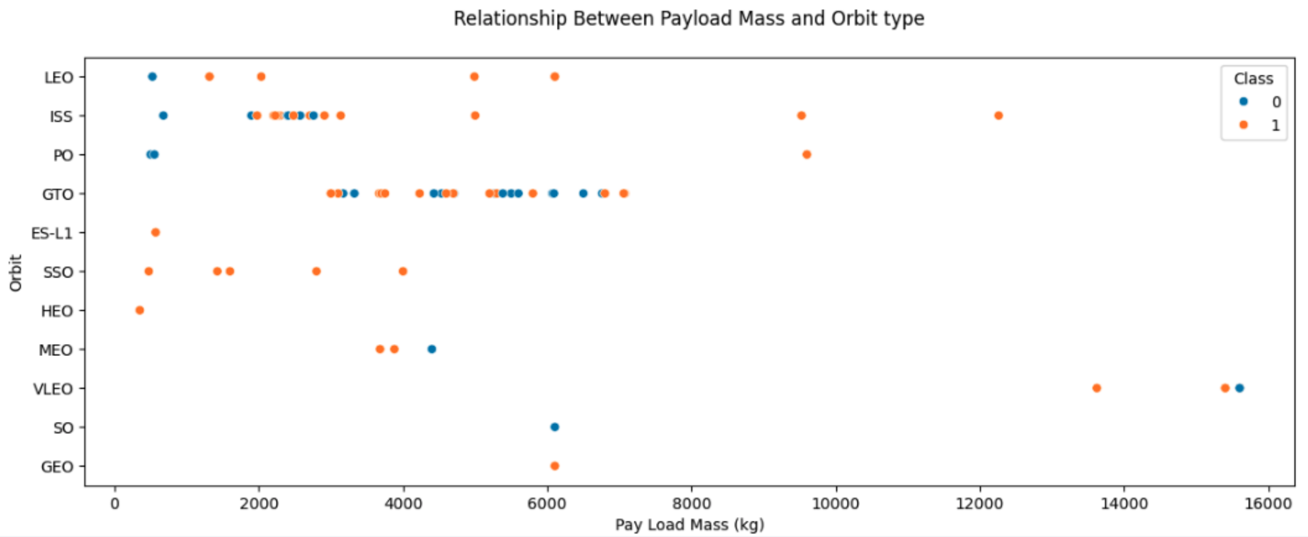
Later Missions (Flight # 50–90+)

- Broader portfolio including **VLEO, SO, GEO, and MEO**.
- More consistent successes (orange dominates).
- Indicates maturity of Falcon 9 technology and reliability.

Overall

- Early missions tested core orbits (LEO, ISS, GTO).
- As experience grew, SpaceX took on diverse and complex orbits. With increasing success, demonstrating technological maturity and mission versatility. We observe that reliability improves with flight progression, regardless of the orbit type.

Payload vs. Orbit Type



LEO, ISS, and PO Orbits

- Payloads typically range from **small (<1,000 kg) to mid-size (~6,000 kg)**.
- High number of missions reflects operational focus.
- Mix of successes and failures.

GTO (Geostationary Transfer Orbit)

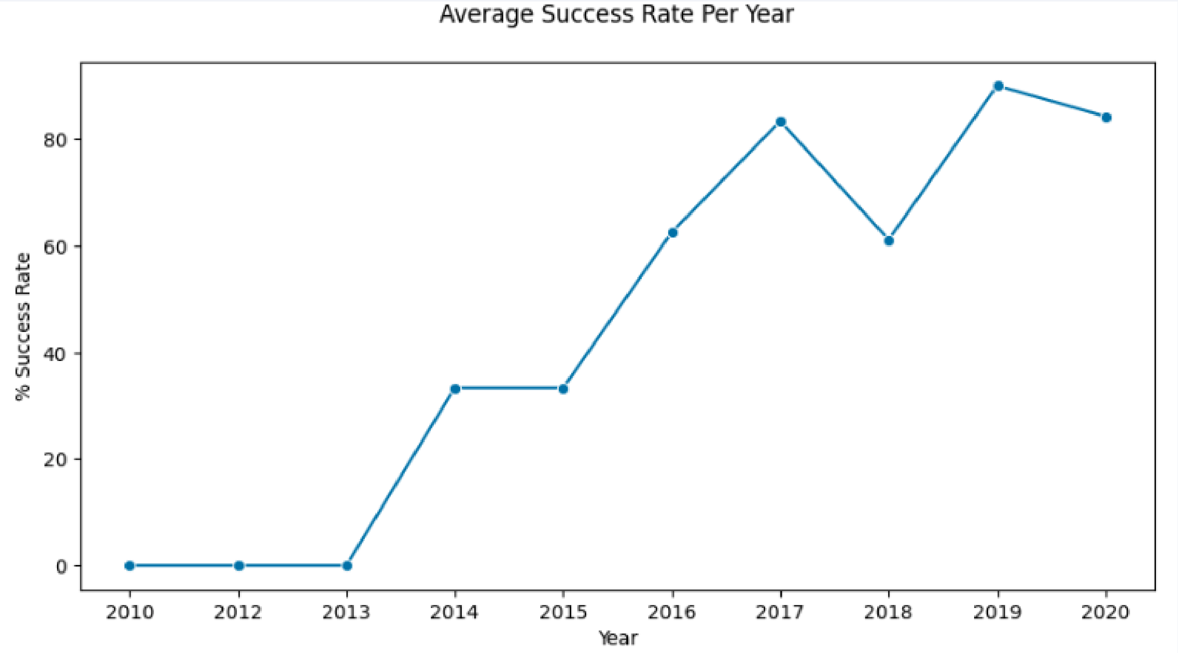
- Handles a wide span of payload masses, especially **mid-to-heavy (3,000–6,500 kg)**.
- Several failures are seen here, highlighting higher mission complexity.

Overall

- Lower-mass payloads dominate most orbit types (LEO, ISS, SSO).
- Heavier payloads are tied to complex/high-value orbits (GTO, GEO).
- Mission success depends more on the complexity of the orbit type than on payload mass alone.

Launch Success Yearly Trend

From 2010 to 2020, SpaceX evolved from high-risk trial-and-error launches to achieving near 90% reliability, establishing Falcon 9 as a dependable workhorse.



2014–2015

- Improvement to ~33% success.
- Mark's **first consistent successful missions**, but still unstable.

2016–2017

- **Rapid growth**: success rate jumps from **~60% to 83%**.
- Reflects major technological improvements and better launch reliability. (2017, first booster refight).

2018

- **Temporary dip (~60%)** corresponds to challenges with higher mission complexity and experimental launches.

2019–2020

- Peak at **~90% success (2019)**, stabilizing at high reliability.
- Demonstrates maturity of Falcon 9 operations and reusable booster technology.
- Clear trajectory of improvement from experimental beginnings to a reliable, industry-leading launch provider.

All Launch Site Names

Names of unique launch sites in the EDA Visualization Lab are

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

Names of Unique Launch sites in the Folium Mapping Exercise are

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

CCAFS LC-40 - SLC-40: Why the Name Difference

- **Same Pad, New Designation**
  - **LC-40** = "Launch Complex 40" (U.S. Air Force, 1965–2005, Titan rockets)
  - **SLC-40** = "Space Launch Complex 40" (U.S. Space Force designation, active pad for Falcon 9)
  - Often now nicknamed **"Slick Forty"**

Timeline of Events

📌 **Apr 2007** – SpaceX leases LC-40, officially redesignated **SLC-40**
📌 **2010–2016** – Falcon 9 launches; commonly still called **LC-40**
📌 **Sept 2016** – Explosion destroys AMOS-6, pad shut down
📌 **Early 2017** – Pad rebuilt & modernized; launches shift to KSC LC-39A
📌 **Dec 2017** – CRS-13 launch - consistent adoption of **SLC-40**

- **Why the shift in naming convention (Late 2017)**
  - Pad was rebuilt as a modern Falcon 9 facility.
  - Rebranding after the 2016 mishap signalled a fresh start.
  - Clearer operational distinction with KSC's LC-39A.

Launch Site Names that Begin with 'CCA'

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

First 5 records where launch sites begin with `CCA`

- These launches were successful in delivering their payload, but no boosters were landed.
- At this time, the workhorse pad at Cape Canaveral was still referred to as LC-40

Total Payload Mass (2010-June-04 to 2020-December-06)

- The total payload mass carried by boosters that had a “NASA” payload onboard is 107010 kg.

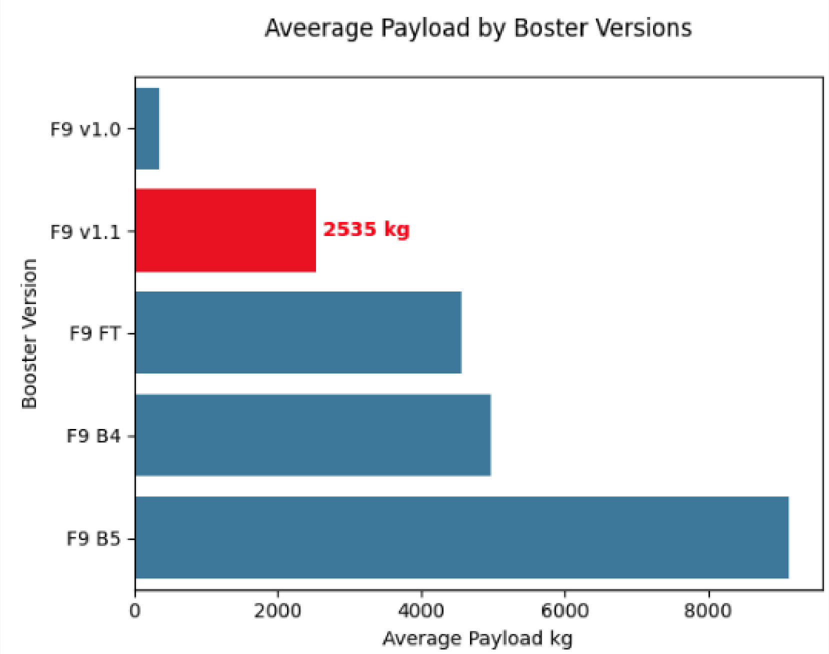
```
%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer like'%NASA%'
 * sqlite:///my_data1.db
Done.
```

SUM(PAYLOAD_MASS__KG_)
107010

- The total payload carried by boosters for “NASA CRS” missions alone from is 45596 kg.

- The total payload carried by boosters for ALL missions is 619967 kg.
- Missions that include NASA Payloads account for approximately ~17% of the payload mass SpaceX sends to orbit.

Average Payload Mass by F9 v1.1



Falcon 9 evolved from a modest test vehicle into Block 5, a reusable heavy-lifter, boosting SpaceX's competitiveness and reliability

F9 v1.1 (2013–2016)

- Increased payload capacity (~2,535 kg).
- Marked the first **major performance upgrade**.

F9 Block 5 (2018–present)

- Dramatic workhorse, supporting most current missions. leap: **~9,000 kg payload average**.
- Optimized for reusability and cost efficiency.
- Industry-standard

- Each iteration of Falcon 9 significantly increased payload capacity, cementing SpaceX's ability to compete across commercial and government sectors.

First Successful Ground Landing Date

```
[209]: %sql select * from SPACEXTBL where Landing_Outcome like '%ground pad%'
 * sqlite:///my_data1.db
Done.
```

[209]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-06-03	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-08-14	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-09-07	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990	LEO	U.S. Air Force	Success	Success (ground pad)
2017-12-15	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2018-01-08	1:00:00	F9 B4 B1043.1	CCAFS SLC-40	Zuma	5000	LEO	Northrop Grumman	Success (payload status unclear)	Success (ground pad)

- **First on-land landing:** Falcon 9 Flight 20 on **Dec 21, 2015**, landed successfully at LZ-1.
- **Since then:** Every Return to Launch Site (RTLS) landing attempt has been successful. Failures have only occurred on drone ship landings (ASDS "Of Course I Still Love You," "Just Read the Instructions," and "A Shortfall of Gravitas").
- **No records** exist of a Falcon 9 booster being destroyed during a designated RTLS attempt after **Dec 21, 2015.**

There was the NASA CRS-16 mission, Launch Date: December 5, 2018. The first stage of the Falcon 9 rocket attempted to land on a ground pad at Landing Zone 1 (LZ-1) but failed. During its return, a grid fin hydraulic pump stalled, causing the booster to spin out of control.

As a safety precaution, the rocket's flight software instructed it **to land in the ocean** instead of attempting to touch down on land while out of control. The rocket successfully executed a soft splashdown in the Atlantic Ocean just off the Florida coast, though it did tip over afterward. The booster was later recovered but was too damaged to be reused.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select * from SPACEXTBL where Landing_Outcome like '%drone ship%' 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
2016-03-04	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271	GTO	SES	Success	Failure (drone ship)
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)

- List of boosters which have successfully landed on a drone ship and had a payload mass between 4000kg and 6000kg
  - F9 FT B1020
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Launch Sites & Dates

- **CCAFS LC-40**: 3 missions (2016).
- **KSC LC-39A**: 2 missions (2017).

Outcomes

- All missions targeted GTO (Geostationary Transfer Orbit).
- **All 5 missions were successful in delivering their payloads.**
- **Landing Outcomes**:
  - 1 failure (SES-9, March 2016).
  - 4 successful drone ship landings (2016–2017).

Total Number of Successful and Failure Mission Outcomes

SpaceX has shifted the conversation from "Can we succeed?" to "How do we make success cheaper and more scalable?" With reusability as the cornerstone of its strategy.

Mission Outcomes.

- **100 missions completed successfully** with **1 mission failure** recorded **(~99% success rate)** overall (2010-06 to 2020-12)
- Confirms Falcon 9 as one of the most reliable launch vehicles in modern spaceflight.

Key Insight

- **Industry Shift**: SpaceX's focus on cost efficiency while maintaining near-perfect success rates has reset expectations for the entire space industry.
- Reliability and reusability have become central to SpaceX's business model.

Boosters that Carried Maximum Payload

```
%sql select Booster_Version, * from SPACEXTBL where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_)from SPACEXTBL)
 * sqlite:///my_data1.db
Done.
```

Booster_Version	Date	Time (UTC)	Booster_Version_1	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
F9 B5 B1048.4	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
F9 B5 B1049.4	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
F9 B5 B1051.3	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
F9 B5 B1056.4	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
F9 B5 B1048.5	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
F9 B5 B1051.4	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
F9 B5 B1049.5	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
F9 B5 B1060.2	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
F9 B5 B1058.3	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
F9 B5 B1051.6	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
F9 B5 B1060.3	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
F9 B5 B1049.7	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

Boosters which have carried the maximum payload mass. All boosters are Falcon 9 Block 5.

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

Booster Version

- All missions with a **15,600 kg payload mass** were flown on **Falcon 9 Block 5**.
- **Block 5** is SpaceX's **most advanced and reusable Falcon 9 variant**.
- Customers: Primarily SpaceX, with occasional commercial/government partners (e.g., Planet Labs, US Air Force GPS III).

Outcomes

- **All 12 missions successfully delivered their payloads** into orbit.
- **Landing outcomes:** 2 failures, 9 successes - still improving reusability under heavy payload conditions.

2015 Launch Records

- Failed landing outcomes in drone ship, their booster versions, and launch site names for the year 2015

Date	MonthName	Month	Landing_Outcome	Booster_Version	Launch_Site	Mission_Outcome
2015-01-10	January	01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	Success
2015-04-14	April	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	Success

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

No Attempt (10 cases)

- Early missions or missions without recovery hardware.
- Reflects Falcon 9's initial focus purely on mission success.

Drone Ship Landings

- **5 successes** and **5 failures**. Highlights the learning curve for at-sea recovery.

Ground Pad Landings

- **3 successes**. Generally, higher success rates are due to a controlled environment.

Ocean Landings (Unplanned/Controlled)

- **3 controlled** (intentional ocean landings). **2 uncontrolled** (loss of vehicle).

Other Outcomes

- **2 parachute failures** and **1 precluded landing**

Key Insights

- Landing technology evolved progressively: from no attempt → ocean tests → ground pad, → drone ship.
- Failures provided critical data, enabling SpaceX to achieve reliable booster recovery in later missions.

- SpaceX's booster recovery journey demonstrates a clear path of innovation - from no attempts to experimental ocean landings, culminating in reliable recoveries on both ground pads and drone ships

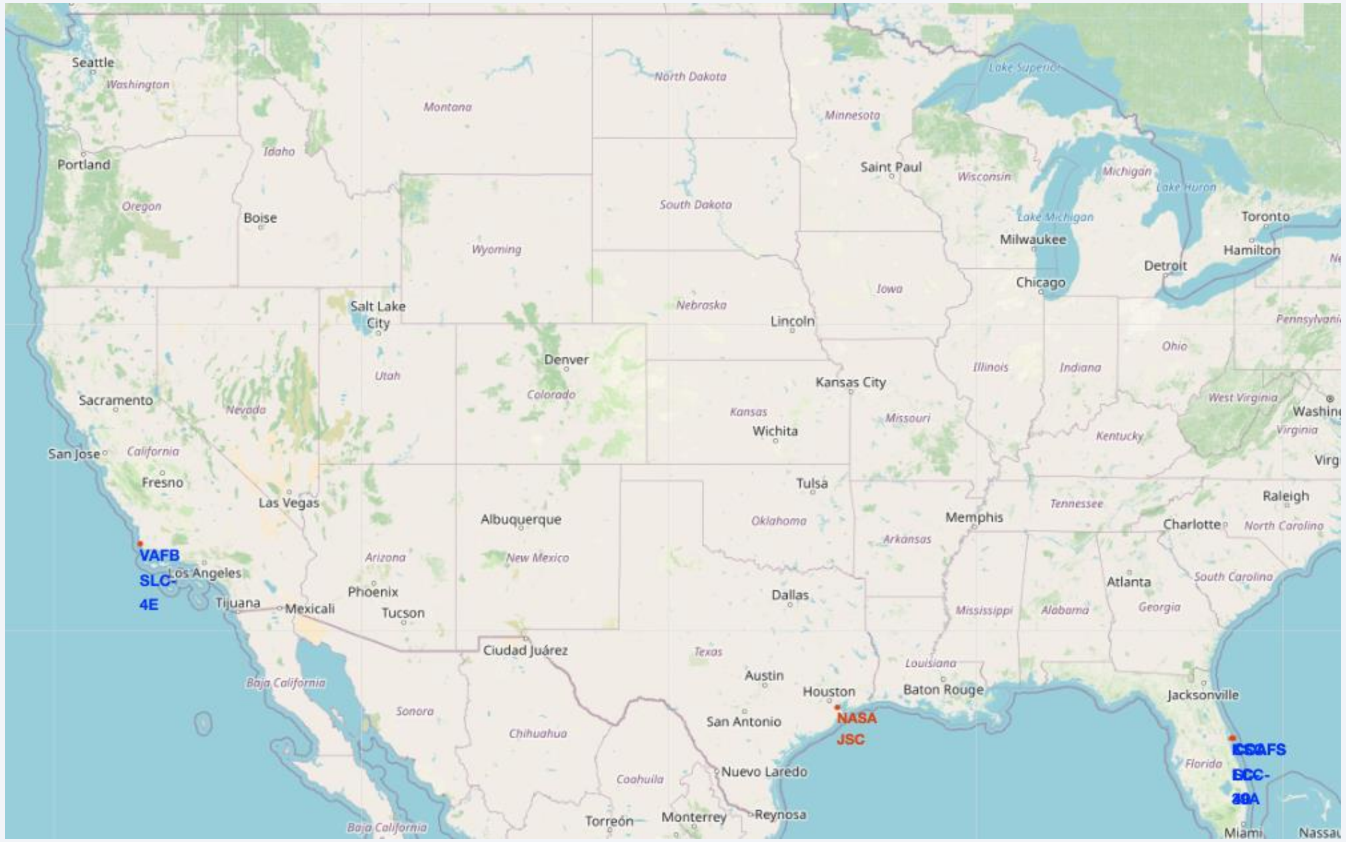
Count of landing outcomes between 2010-06-04 and 2017-03-20

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

Section 3

Launch Sites Proximities Analysis

SpaceX Launch & Landing Infrastructure in the U.S.



Florida (East Coast Hub)

- **CCAFS SLC-40 (Cape Canaveral Space Launch Complex)** and **KSC LC-39A (Kennedy Space Center),** and the main Falcon 9 launch pads.
- Used for LEO, LEO(ISS), MEO, HEO and GTO orbits mostly
- **Drone ships ("Of Course I Still Love You" & "A Shortfall of Gravitas")** operate in the Atlantic.

California (West Coast Hub)

- **VAFB SLC-4E (Vandenberg Air Force Base):** Used for polar, polar LEO, and sun-synchronous orbits.
- Drone ship **"Just Read the Instructions"** operates in the Pacific.

SpaceX Launch & Landing Infrastructure in the U.S. (East Coast)

- The second image contained incorrect coordinates, which placed the Kennedy Space Center marker in the wrong location. After verifying the correct coordinates, we updated the data so the marker now sits accurately on Launch Complex 39A.

West Launch Pad & Booster Landing Outcomes (VAFB SLC-4E)

SLC-4E (Vandenberg, California)

- Used for polar and sun-synchronous orbits.
- Fewer total missions.
- (2010-06 to 2018-06). 4 successful landings and 6 failed landings.

East Launch Pad & Booster Landing Outcomes (CCAFS S/LC-40)

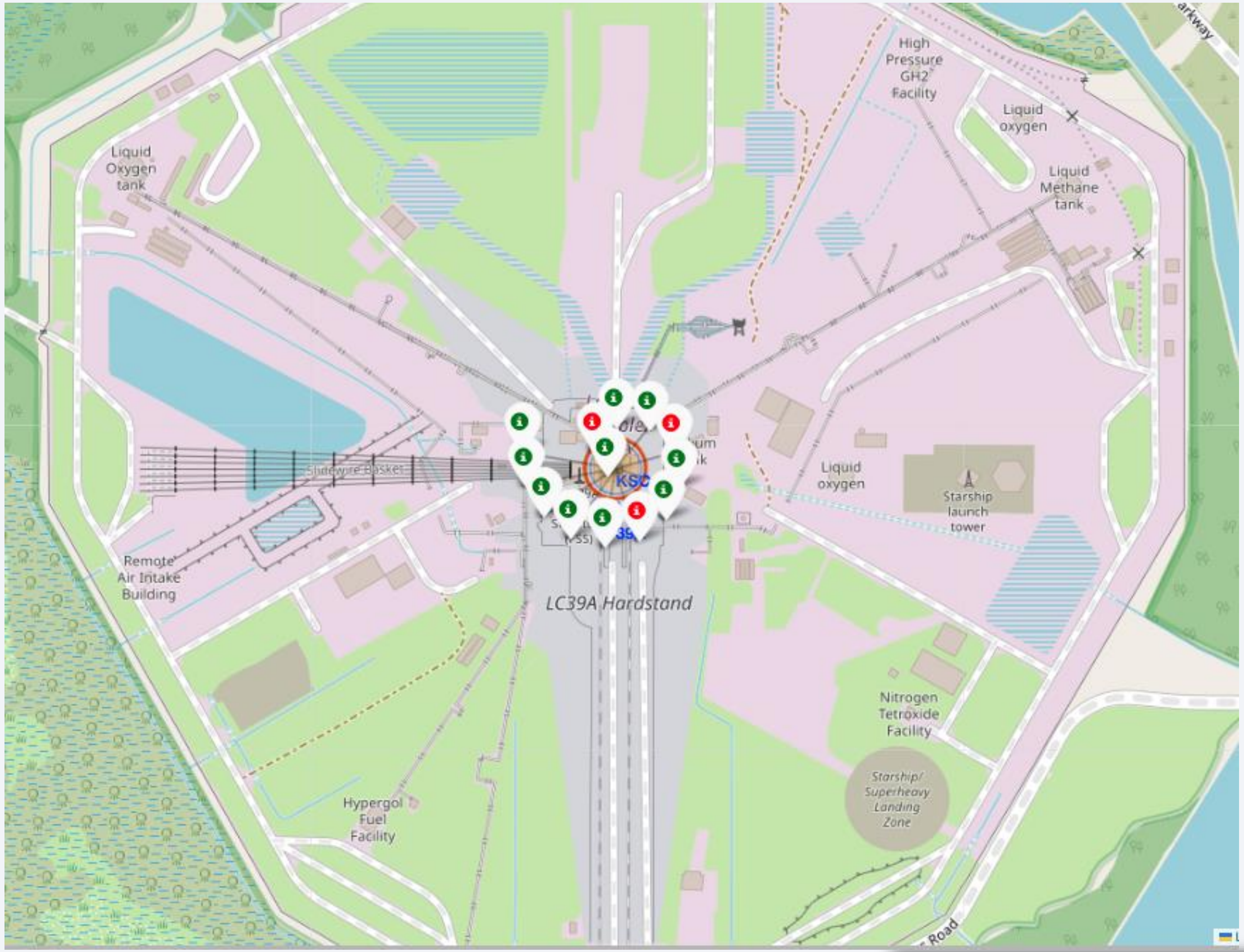
SLC-40 (Cape Canaveral, Florida)

- **Workhorse pad** for Falcon 9 (commercial, high-volume).
- There were 33 landing attempts in total (2010-06 to 2018-06). 10 were successful, and the rest were unsuccessful.
- CCAFS LC-40 was officially redesignated as **CCAFS SLC-40** in **April 2007** when the U.S Space Force leased LC-40 to SpaceX.
- SpaceX began using the pad but did not start referring to the complex as SLC-40. (2010- 2016)
- In December 2017, following a major rebuild after a Falcon 9 static fire explosion heavily damaged the pad in September 2016, SpaceX started to refer to the pad as SLC-40 (often called *'Slick Forty'*)
- Rebranding the pad after the 2016 mishap signalled a fresh start as it was essentially rebuilt as a modern Falcon 9 facility and provided a clearer operational distinction with KSC's LC-39A.
- CCAFS LC-40 = CCAFS SLC-40

East Launch Pad & Booster Landing Outcomes (KSC LC-39A)

LC-39A (Kennedy Space Center, Florida)

- Mailly used for crew, heavy, and flagship missions.
- Visual clusters of green over time show the maturity of booster recovery.
- 10 out of 13 were successful landings



Location Context – Cape Canaveral Launch Site

Cape Canaveral's geography provides the perfect balance of launch safety, access to NASA facilities, and proximity to logistics hubs, making it SpaceX's most critical launch region.

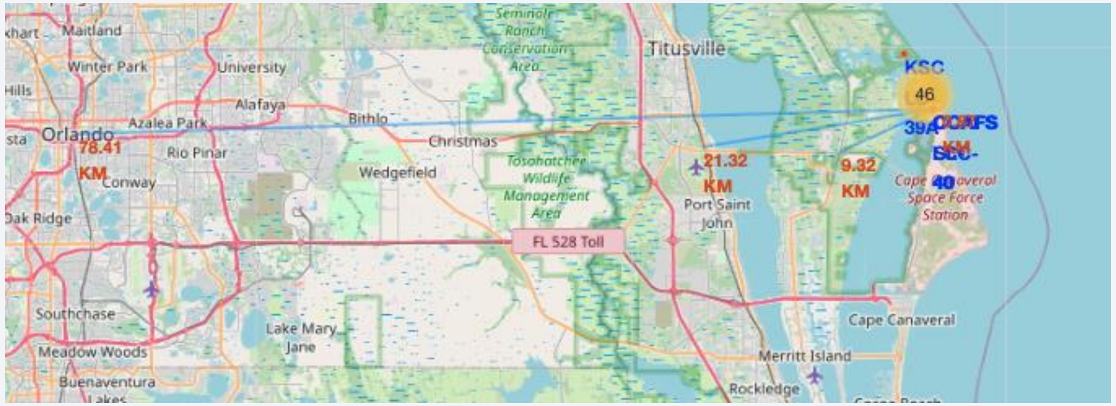
Proximity to Key Points

- **~0.9 km** - distance to coastline
- **~9.3 km** - from pad to local highways & NASA Parkway (main access routes).
- **~21 km** - from Florida East Coast Railway.
- **~78 km** - to Orlando, the nearest major city/airport hub.

Location Context – Cape Canaveral Launch Site

Strategic Benefits

- Coastal location - safe eastward launches over the Atlantic.
- Close access to NASA/KSC infrastructure for crewed missions.
- Close to **KSC LC-39A** (leased from NASA, heavy-lift and crew missions).
- Buffer zones serve as safety margins between populated areas.



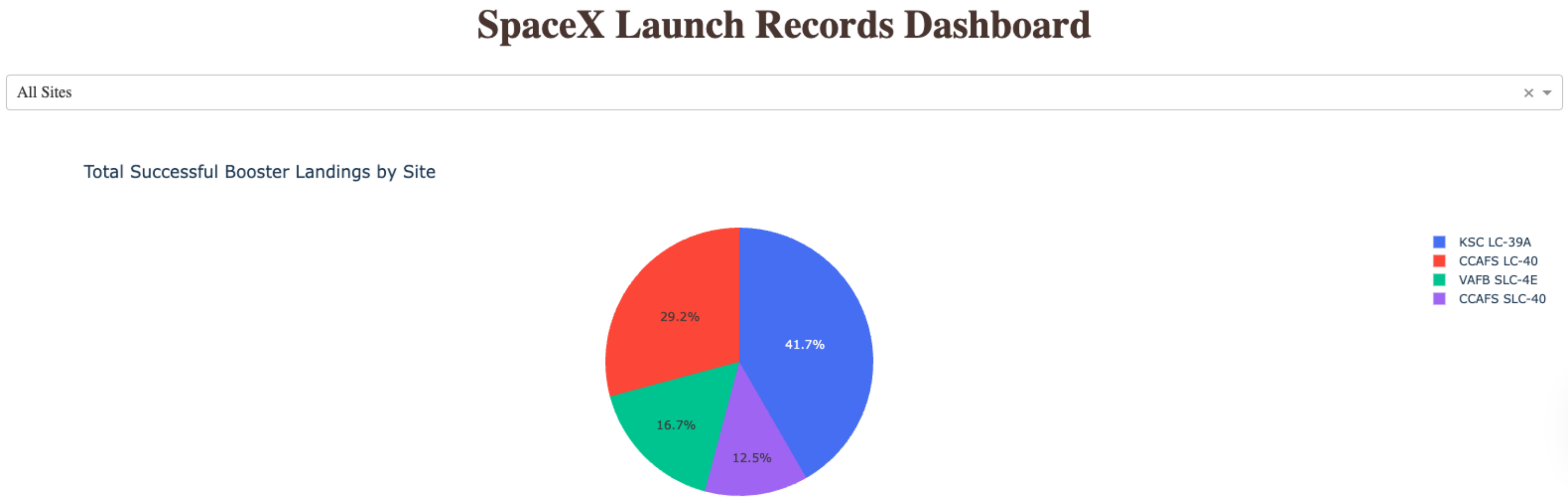
- **Operational Importance**
- **SLC-40**: high-volume commercial missions.
- **LC-39A**: flagship pad for crewed, heavy, and special missions.
- Together, they give SpaceX launch redundancy in Florida.

Section 4

Build a Dashboard with Plotly Dash

SPACEY

Distribution of Successful Landings by Site

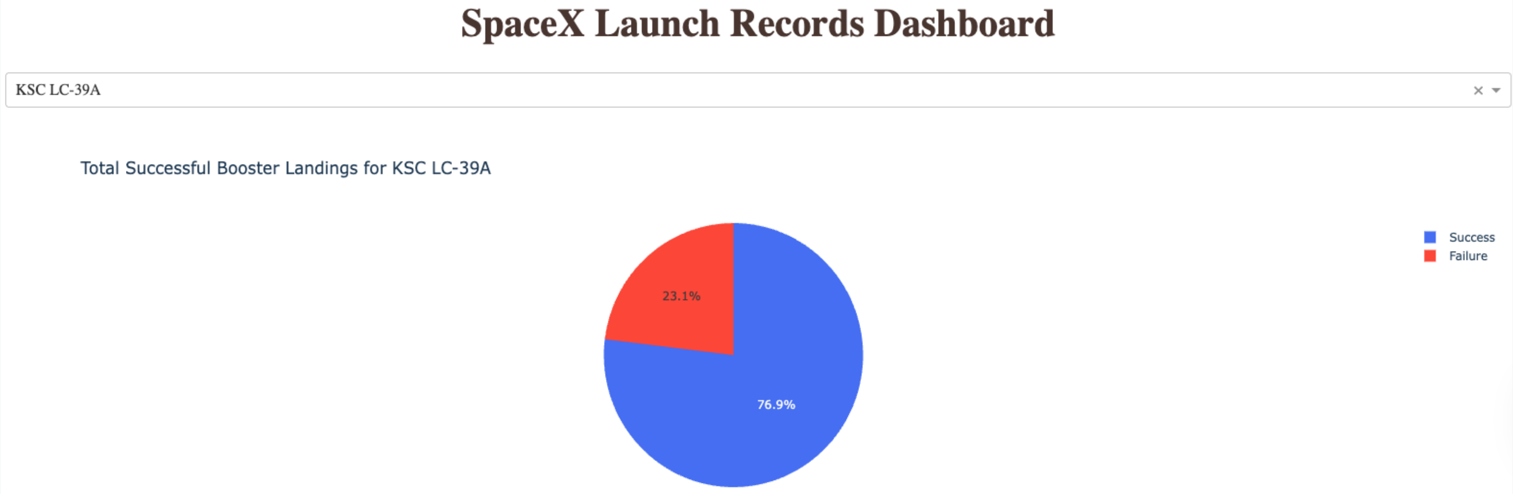


The app features:

- A header section,
- A dropdown filter used to filter successful booster landings by specific launch site, which enables interactive selection of one site or all sites.
  - "**All Sites**" is selected; the Pie chart shows **the rate of successful landings across all SpaceX sites.**
  - **Specific site selected**; the Pie chart will show only **success vs failure booster landing rates for that site.**
- A pie chart with a title and legend, creating a simple yet interactive dashboard.

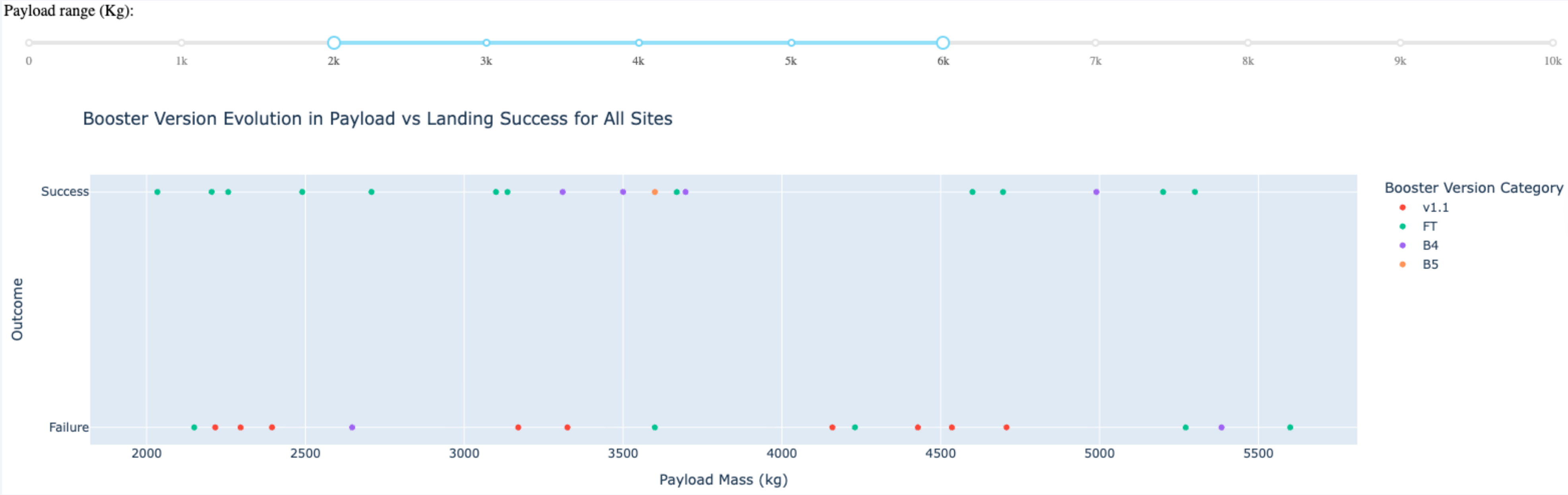
LC-39A leads in successful landings, but SpaceX relies on a diversified network of launch sites, with Florida as the dominant hub

Landing Outcomes at KSC LC-39A



- By selecting a specific site, the boosters' landing success and failure rates are calculated and displayed just for that site, allowing for site-level analysis.
- LC-39A is primarily used for **ground pad landings**, which are generally more controlled compared to drone ship attempts.
- LC-39A shows a strong majority of successful landings, even though it hosted some of the most challenging missions (crewed and heavy payloads).

Booster Evolution- Landing Outcome by Payload Mass



There are notable events between payloads of 2,000–6,000 kg. Earlier booster versions (v1.1, and early FT) exhibited more failures than successes, but later versions (Block 4/5) eliminated this weakness, transforming a historically challenging payload range into a consistently successful one.

Key Elements in the Chart

Payload Range Slider (Top)

- Allows filtering missions by payload mass (0–10,000 kg).
- Helps isolate performance for lighter vs. heavier missions.

Scatter Plot (Payload Mass (kg) VS Outcome)

- Color coding (legend): Booster Version Category (v1.0, v1.1, FT, B4, B5).

Booster Evolution- Landing Outcome by Payload Mass

Mission success is driven more by booster evolution than payload size.

Early Versions (v1.0, v1.1 – Blue & Red dots)

- Mostly failures (clustered along “Failure” at low payloads <2,000 kg).
- Reflects SpaceX’s learning curve in early Falcon 9 missions.

Block 4 (Purple dots)

- Operates in mid-to-high payload ranges (up to ~10,000 kg).
- Mostly successful, but still a few notable failures (e.g., around 6,000–10,000 kg).

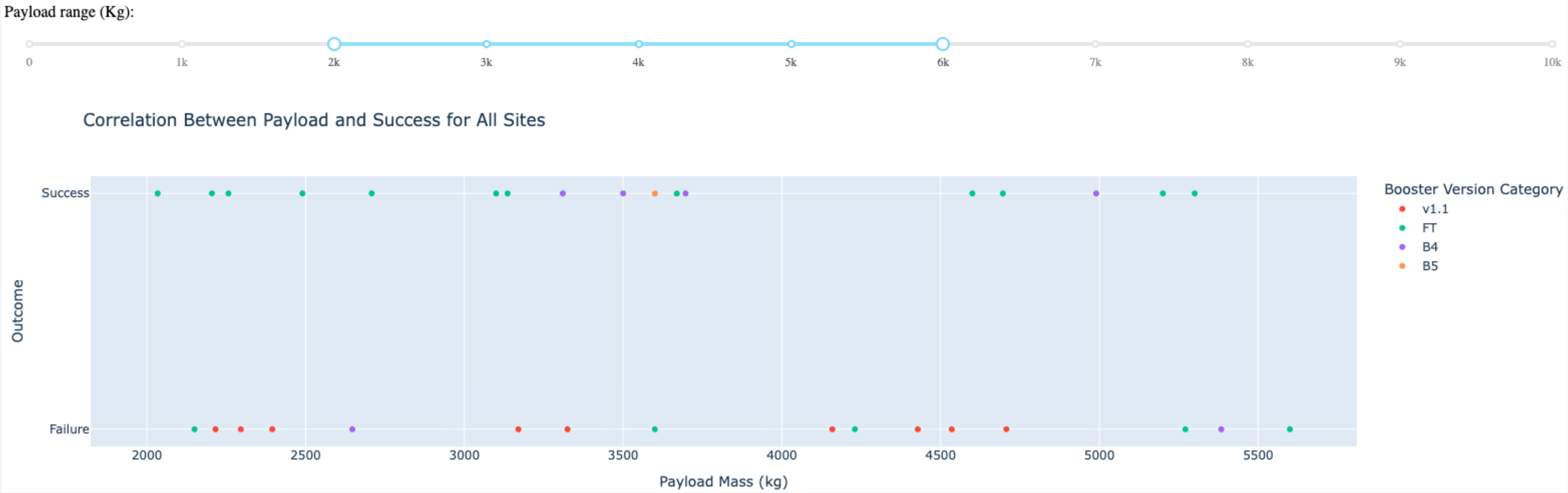
Falcon 9 Full Thrust (FT – Green dots)

- Higher concentration of successes, especially in the 2,000–6,000 kg payload range.
- Occasional failures, but success rate much improved over early versions.

Block 5 (Orange dots - limited data)

- While our dataset includes only a single record for Booster Version 5 (B5), it reflects SpaceX’s most reliable and reusable booster, capable of delivering heavier payloads (>5,000 kg).

Booster Evolution- Landing Outcome by Payload Mass



Overall Insights

Payload Size vs. Success:

- Payload mass **does not strongly dictate success**. Failures appear across small, mid, and large payloads.
- Instead, **booster version maturity is the stronger predictor of success**.
- Clear progression from **low success in v1.0/v1.1** → **improvement in FT** → **near-perfect reliability in Block 5**.
- SpaceX's iterative engineering directly correlates with better mission outcomes.

High Payload Resilience:

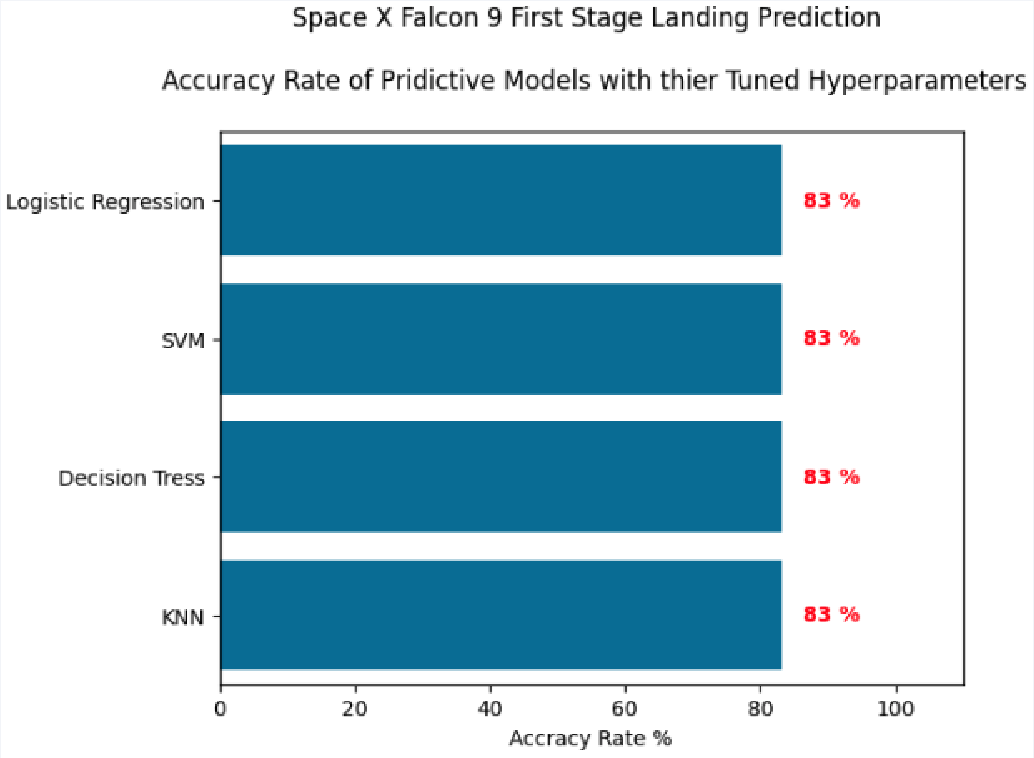
- Even at **heavier payloads (>5,000 kg)**, Block 5 achieves strong success rates, supporting missions such as Starlink and high-value commercial/government contracts.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All models are equal at 83.33% but accuracy alone doesn't distinguish models.



Confusion Matrix of th SVM Model

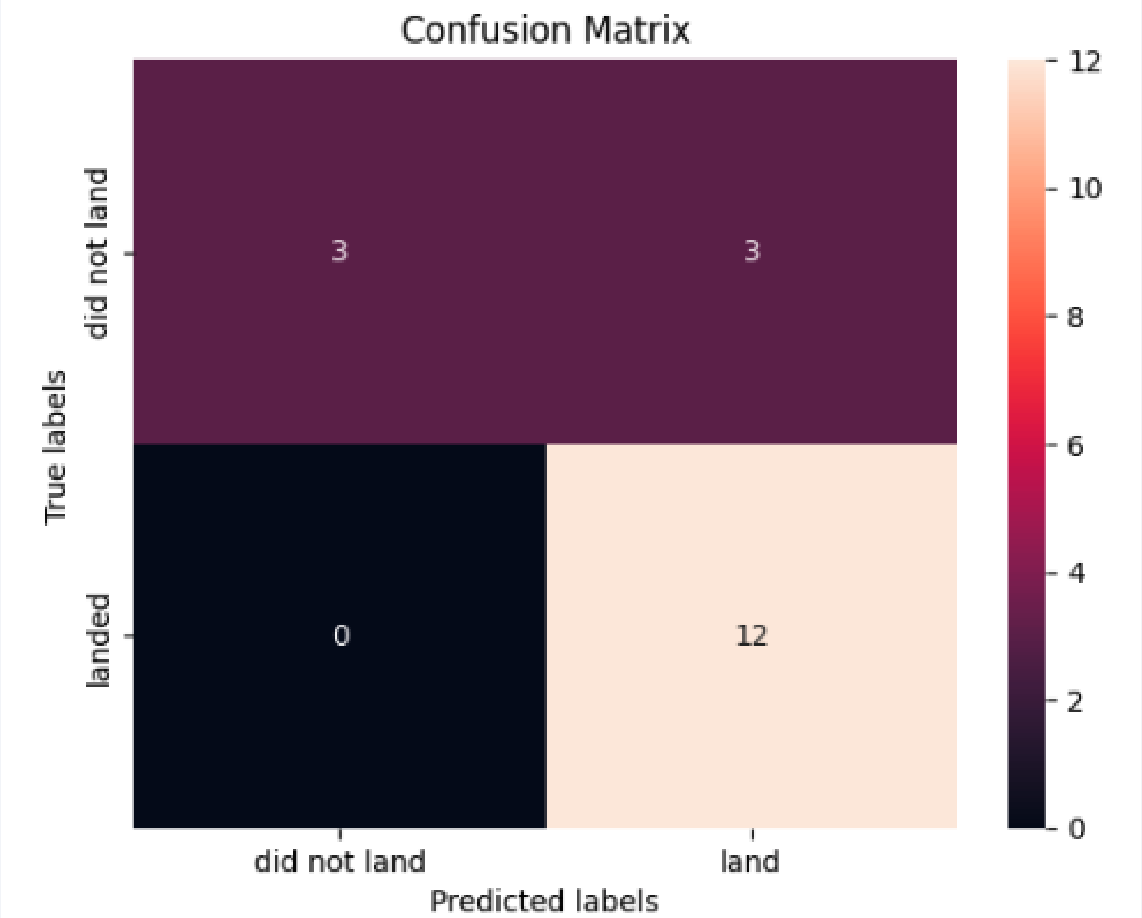
Confusion Matrix Breakdown (Identical for all Models)

- **True Negatives (TN)** = 3 - correctly predicted “did not land”
- **False Positives (FP)** = 3 - predicted “landed” when it actually did not land
- **False Negatives (FN)** = 0 - none of the landings were missed
- **True Positives (TP)** = 12 - correctly predicted “landed”

Insights

- The SVM **never misses a successful landing** (Recall = 100%). Same as all models
- It struggles with **false positives**: 50% of the time, it predicts a landing when there was none. Same as all models.

SVM Confusion Matrix



Model Differences

	Logistic Regression	SVM	Decision Tress	KNN
Accuracy %	83.33	83.33	83.33	83.33
Cross-Validation Accuracy Score	84.64	84.82	87.68	84.82
No of Correct Predictions	15	15	15	15
TruePositive %	100.0	100.0	100.0	100.0
FalsePositive %	50.0	50.0	50.0	50.0
ROC_AUC	88.89	95.83	73.61	89.58

- **Cross-Validation Accuracy Score**: Decision Tree generalizes slightly better (highest CV score). This suggests a slightly stronger out-of-sample performance.

- **For all models,** the **False Positive Rate (50%)** is a significant concern in the booster landing context, as it incorrectly predicts a booster will land when it won't, which could have **serious operational/financial consequences.**

- **ROC-AUC** measures how well the model can separate land vs no-land across all thresholds (not just the fixed 0.5 cutoff). A model with a high ROC-AUC has more flexibility, as it's threshold can be tuned to trade off False Positives (consequential) versus false negatives (non-consequential). In short, ROC-AUC gives a comprehensive picture of the model's discriminative ability.

Confusion Matrix of th SVM Model

	Logistic Regression	SVM	Decision Tress	KNN
Accuracy %	83.33	83.33	83.33	83.33
Cross-Validation Accuracy Score	84.64	84.82	87.68	84.82
No of Correct Predictions	15	15	15	15
TruePositive %	100.0	100.0	100.0	100.0
FalsePositive %	50.0	50.0	50.0	50.0
ROC_AUC	88.89	95.83	73.61	89.58

- Decision Tree ROC-AUC of 73.61% - weakest of all models
- SVM ROC-AUC of 95.83% - strongest of all models

SVM is the best model here, as its ROC-AUC is the strongest, making it the most reliable classifier once the threshold is tuned, which indicates it has a significantly better ability to discriminate landings from failures. This means that we can likely reduce that 50% false positive rate without sacrificing recall.

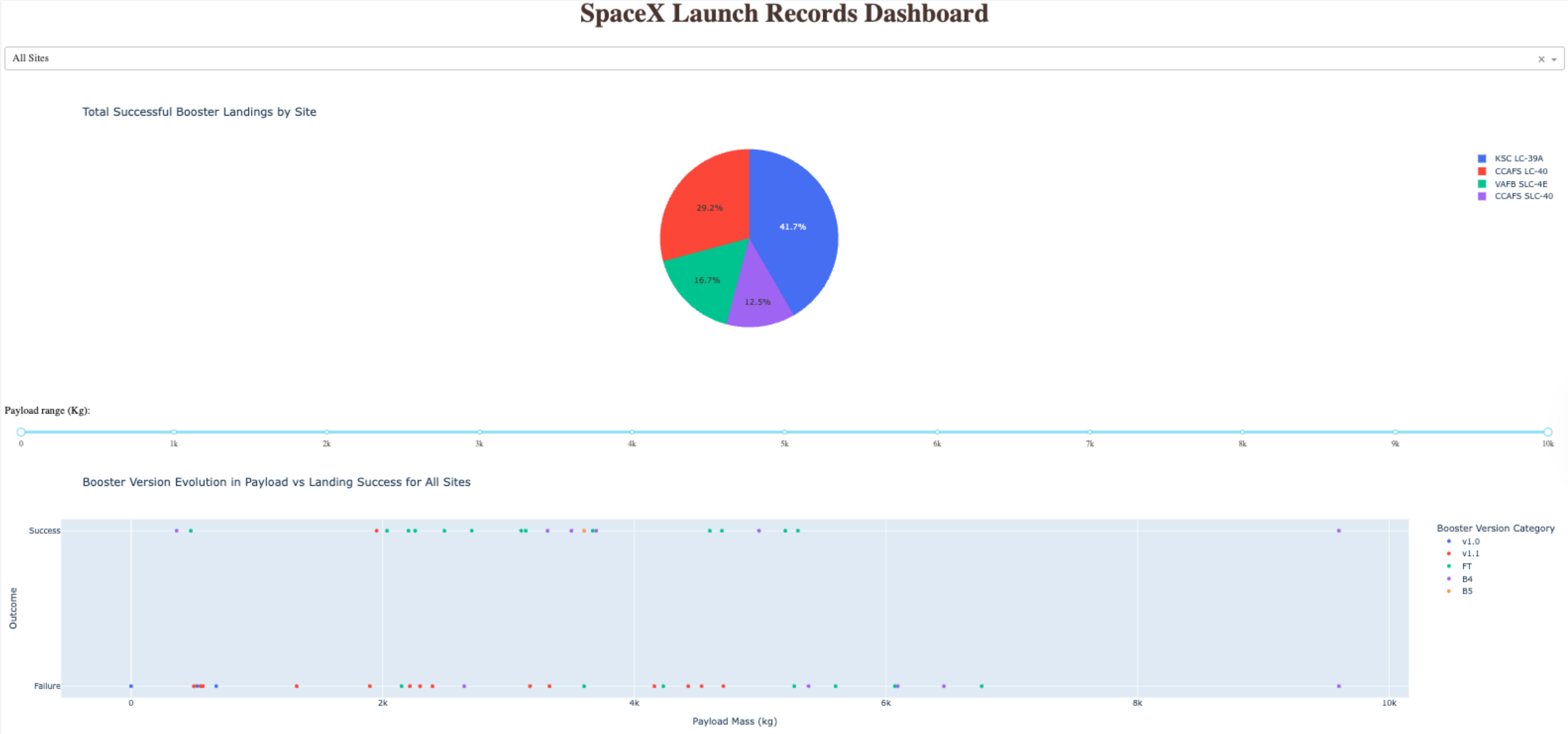
The Decision Tree is less effective here, despite achieving a good CV accuracy; its ROC-AUC (73.61%) indicates poor discriminative power.

Conclusions

✅ Exploratory Data Analysis

- Landing outcomes are influenced by payload mass and launch site, but booster version stands out as the strongest predictor of landing success.
- Clear trends in success rates indicate that newer boosters and lighter payloads perform better.

Conclusions



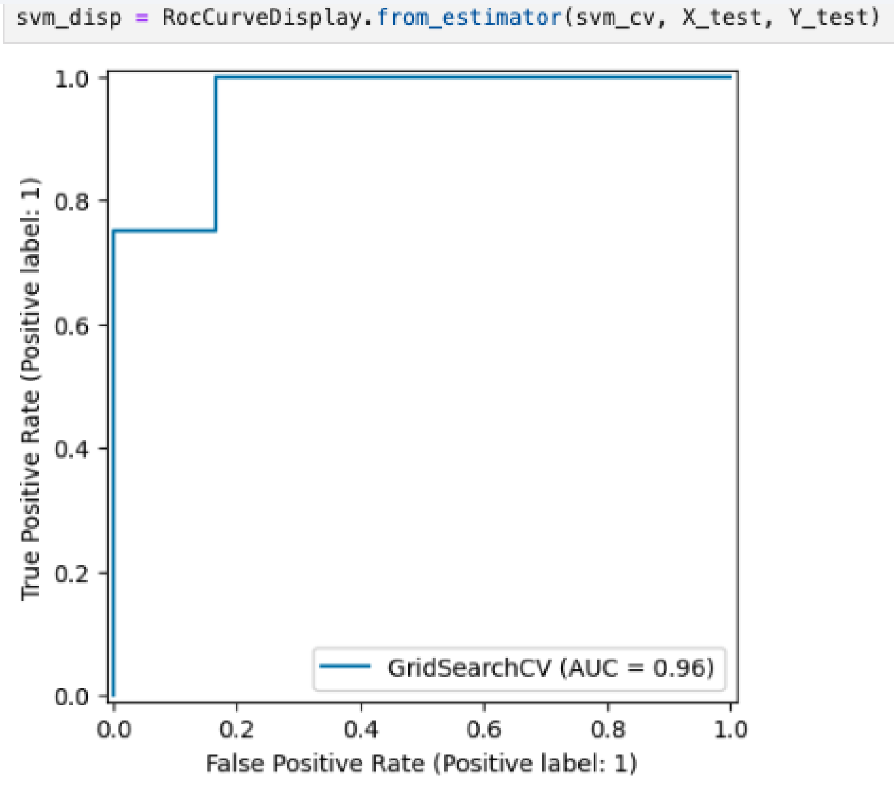
✅ Interactive Dashboard

- Delivered a dynamic tool to visualize landing success patterns by site, booster version, and payload range.
- Enabled real-time exploration of success/failure outcomes across missions.

Conclusions

✅ Predictive Modelling

- Tested Logistic Regression, SVM, Decision Tree, and KNN.
- **SVM emerged as the best model**:
  - Accuracy ~83%
  - Recall = 100% (no missed landings)
  - F1 ≈ 89%
  - ROC-AUC = 95.8 (strong class separation, high potential for tuning)

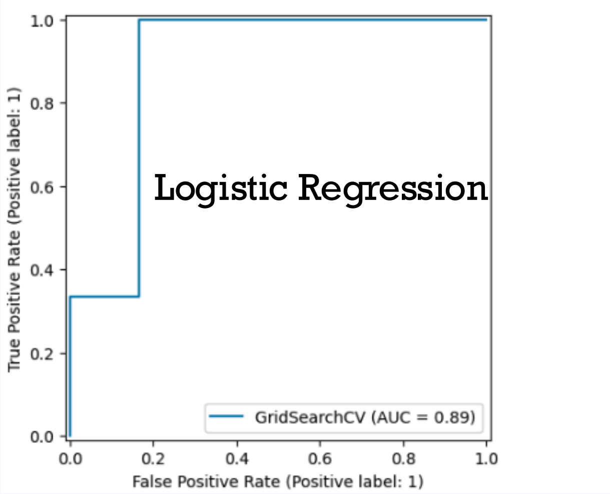


- 🚀 **Key Takeaway**
- With the developed models and dashboard, we can **reliably predict booster landings, support mission planning**, and **reduce uncertainty in launch recovery outcomes**.
- Future work: Enhance the dataset with additional mission features (e.g., trajectory, weather) and optimize thresholds to minimize false positives.

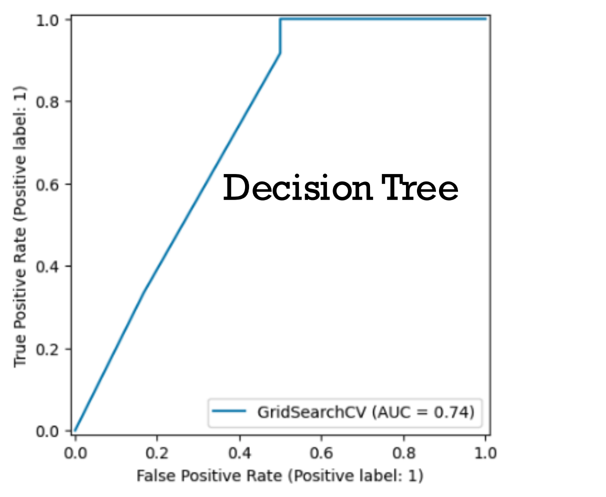
Appendix

Receiver Operating Characteristic Area Under the Curve

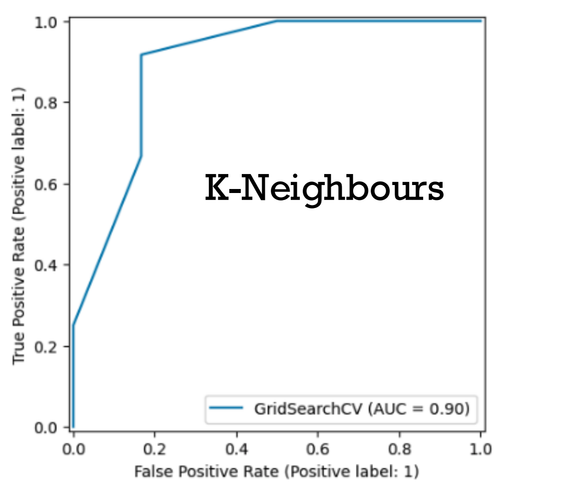
```
lr_disp = RocCurveDisplay.from_estimator(logreg_cv, X_test, Y_test)
```




```
tree_disp = RocCurveDisplay.from_estimator(tree_cv, X_test, Y_test)
```




```
knn_disp = RocCurveDisplay.from_estimator(knn_cv, X_test, Y_test)
```



KSC & VAFB launch sites achieved a success rate of ~77%, compared to ~60% at CCAFS.

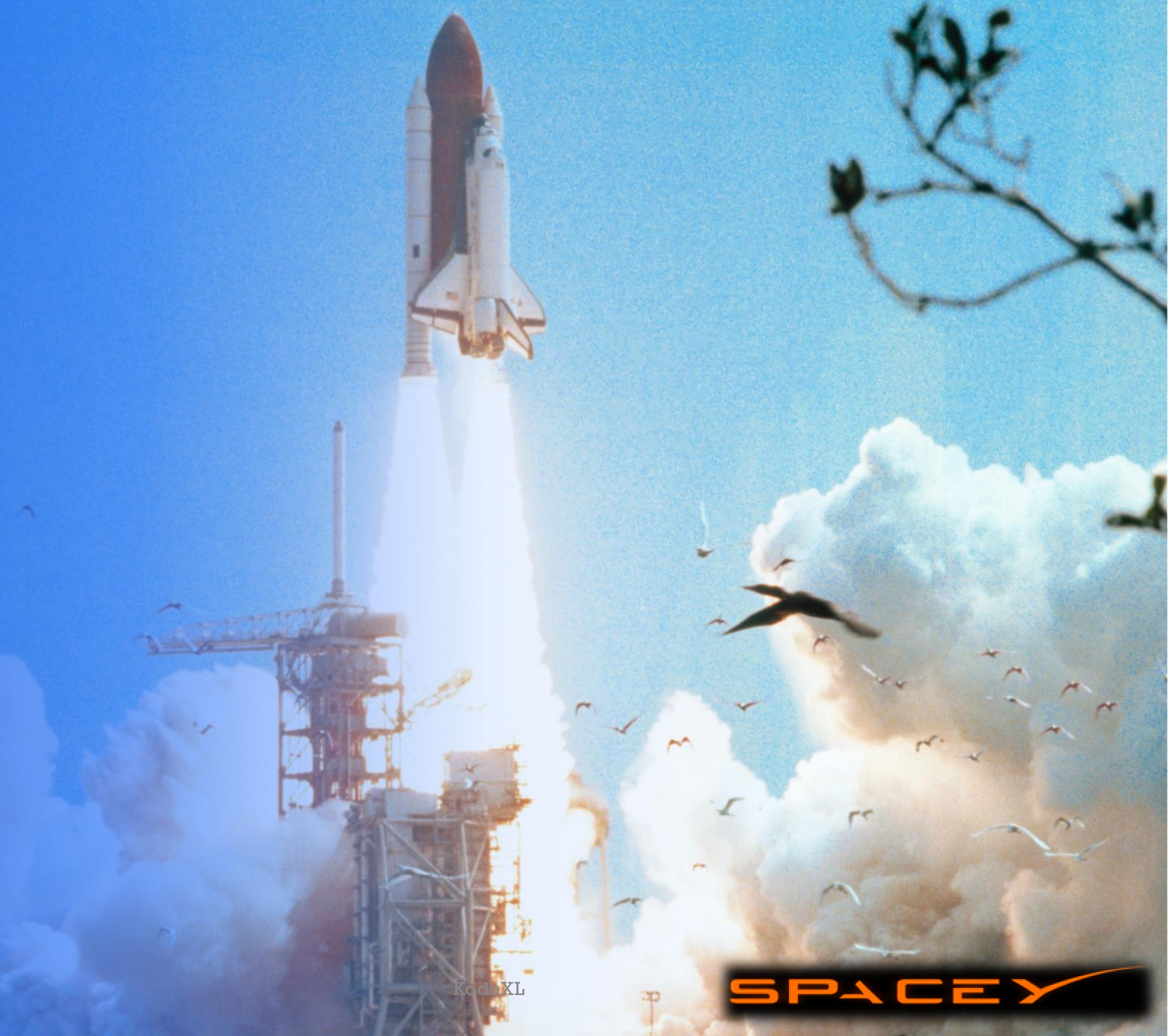
LaunchSite	Class
CCAFS SLC 40	60.00
KSC LC 39A	77.27
VAFB SLC 4E	76.92

Appendix

Early challenges in 2015–2016 drone ship landings

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)
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2016-01-17	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553	LEO	NASA (LSP) NOAA CNES	Success	Failure (drone ship)
2016-03-04	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271	GTO	SES	Success	Failure (drone ship)
2016-06-15	14:29:00	F9 FT B1024	CCAFS LC-40	ABS-2A Eutelsat 117 West B	3600	GTO	ABS Eutelsat	Success	Failure (drone ship)
2018-12-05	18:16:00	F9 B5B1050	CCAFS SLC-40	SpaceX CRS-16	2500	LEO (ISS)	NASA (CRS)	Success	Failure
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

Thank you!



SPACEY