IBM Developer
SKILLS NETWORK

Winning the Space Race with Data Science

Natasha Leask
17 May 2025

Outline

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

Executive Summary

This analysis provides Space Y with a data-driven foundation for strategic planning in the competitive launch market. The 94.44% accurate Decision Tree model offers reliable predictions of SpaceX's operational capabilities, particularly regarding first stage landing success, the key factor in SpaceX's cost advantage. By understanding the patterns and factors driving landing success, Space Y can develop targeted technologies, optimal operational procedures, and competitive pricing strategies to effectively compete in the commercial space market.

Key findings

1. **Model performance:** Decision Tree classifier achieved superior results (94.44% accuracy, 92.31% precision, 100% recall) compared to other algorithms (83.33% accuracy)
2. **Temporal pattern:** Landing success rates improved significantly after the first 30 flights, indicating a clear learning curve in SpaceX operations
3. **Launch site analysis:** KSC LC-39A demonstrated the highest success rate (76.9%), though all sites showed improvement over time
4. **Orbit influence:** Certain orbits (ES-L1, GEO, HEO, SSO) achieved 100% success rates, while GTO missions saw substantially lower success (52%)
5. **Payload impact:** No direct correlation between payload mass and landing success, with even heavy payloads (15,000+ kg) achieving successful landings in later flights

Business implications

The Decision Tree model provides reliable predictions of landing outcomes, offering valuable intelligence for competitive strategy development, cost estimation, and technology prioritisation in the commercial space launch industry.

Introduction

In the rapidly evolving commercial space industry, Space Y aims to gain competitive insights into SpaceX's launch operations through comprehensive data analysis. This project leverages advanced data science techniques to understand and predict critical aspects of rocket launches.

Key Objectives

- Predict first stage rocket landing probability
- Understand factors influencing launch success
- Provide actionable insights for Space Y's launch strategy

Expected Outcomes

A predictive model that offers insights into rocket launch dynamics, enabling more informed decision-making in the commercial space industry.

Section 1

Methodology

Methodology

1. **Data collection**

Primary data sourced from the SpaceX REST API (launches/past endpoint)

Supplementary data collected through web scraping of HTML tables (Wikipedia)

2. **Data wrangling**

Wrangling and cleaning data in preparation for visualisations and machine learning model training.

3. **Performing exploratory data analysis (EDA) using visualisations and SQL**

4. **Creating interactive data visualisations with Folium and Plotly Dash**

5. **Performing predictive analysis using classification models**

Comprehensive model evaluation

Rigorous hyperparameter tuning for:Support Vector Machines (SVM), Classification Trees, Logistic Regression

Systematic model performance comparison

Data Collection

Data sources

SpaceX REST API | Wikipedia | IBM-supplied datasets

1 Data collection

a) API: SpaceX API (JSON) → dataset_part_1.csv (CSV)

b) Web scraping: Wikipedia (HTML) → spacex_web_scraped.csv (CSV)

2 Data wrangling

a) dataset_part_1.csv (CSV) → dataset_part_2.csv (CSV)

3 EDA with SQL

a) Spacex.csv (CSV)

4 EDA visualisations

a) dataset_part_2.csv (CSV) → dataset_part_3.csv (CSV)

5 Folium visualisation

a) spacex_launch_geo.csv (CSV)

6 Interactive dashboard

a) spacex_launch_dash.csv (CSV)

7 Machine learning

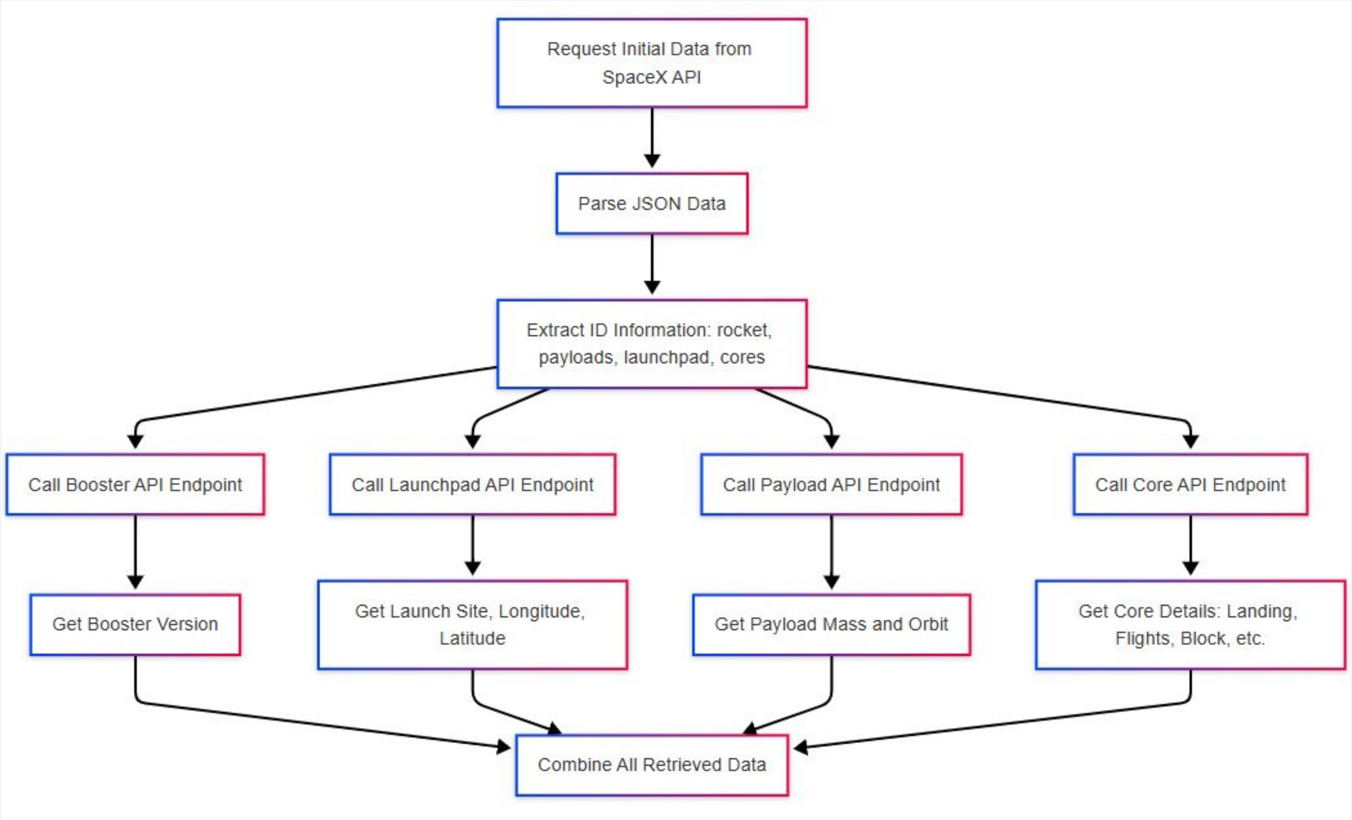
a) dataset_part_2.csv (CSV)

b) dataset_part_3.csv (CSV)

Data Collection – SpaceX API

Primary data was sourced from the SpaceX REST API (launches/past endpoint).

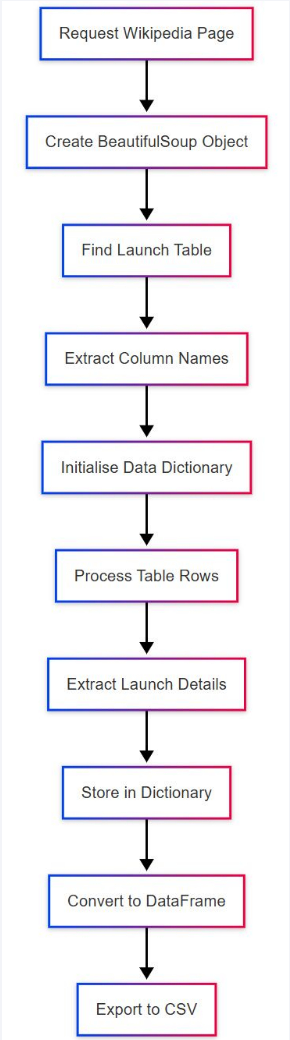
Notebook available on Github:
https://github.com/natashaleask/IBM-Applied-Data-Science-Capstone/blob/main/1_Data_collection_API.ipynb



Data Collection - Scraping

Supplementary data was collected through web scraping of HTML tables (Wikipedia).

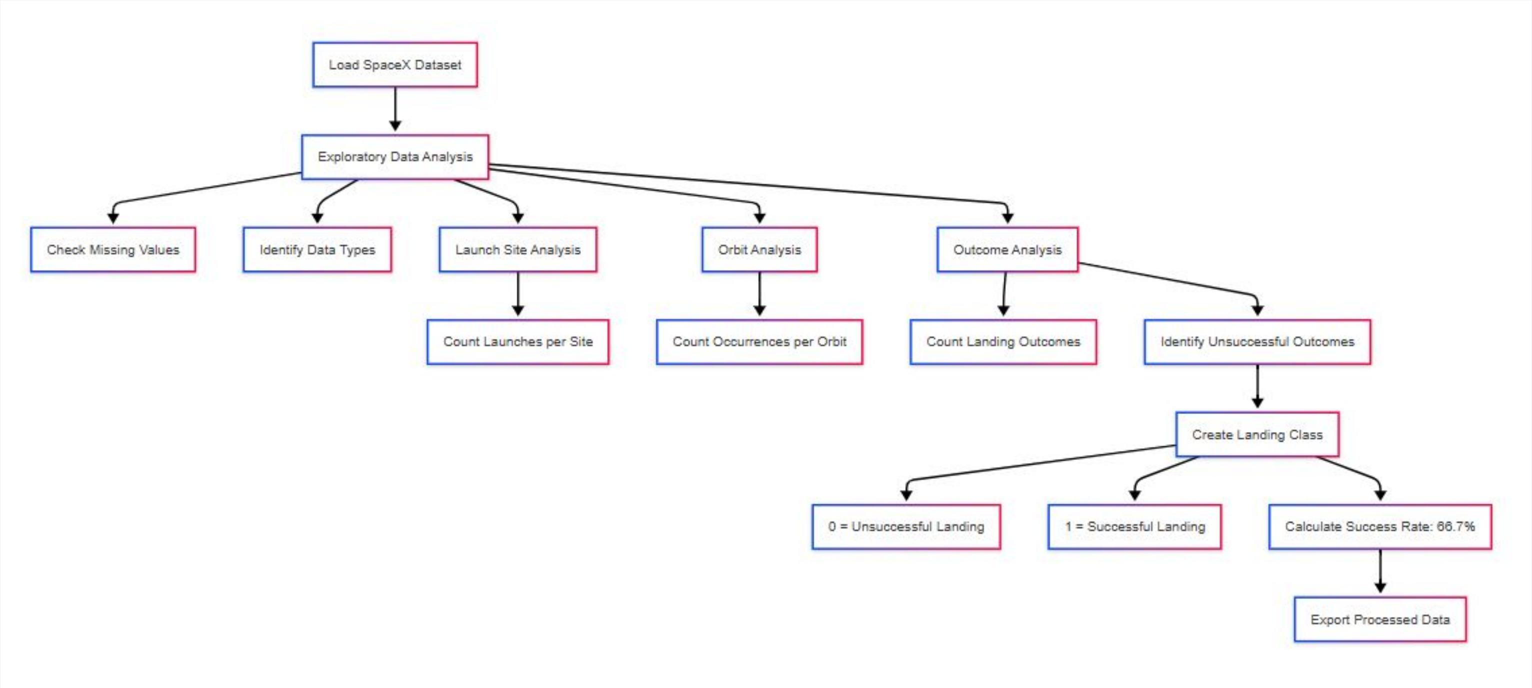
Notebook available on Github:
https://github.com/natashaleask/IBM-Applied-Data-Science-Capstone/blob/main/2_Data_collection_Web_scraping.ipynb



Data Wrangling

Exploratory Data Analysis (EDA) was performed to identify patterns in the data and determine appropriate labels for supervised model training. Various landing outcomes (such as "True Ocean" for successful ocean landings, "False Ocean" for unsuccessful ocean landings, "True RTLS" for successful ground pad landings, "False RTLS" for unsuccessful ground pad landings, "True ASDS" for successful drone ship landings, and "False ASDS" for unsuccessful drone ship landings) were converted into binary training labels, where "1" represented successful booster landings and "0" represented unsuccessful landings.

Notebook available on Github: https://github.com/natashaleask/IBM-Applied-Data-Science-Capstone/blob/main/3_Data_wrangling.ipynb



EDA with Data Visualisation

These visualisations were chosen to reveal patterns and relationships between various factors (flight experience, payload mass, launch site, orbit type) and the success rate of Falcon 9 first stage landings.

1. **Scatter Chart: FlightNumber vs. PayloadMass**: Revealed payload mass alone, does not determine success or failure. Landing success may be related to other improvements over time.
2. **Categorical Plot: FlightNumber vs. LaunchSite**: Used catplot to examine relationships between flight numbers, launch sites, and mission success class.
3. **Scatter Chart: LaunchSite vs. PayloadMass**: Investigated potential relationships between launch locations and payload weights.
4. **Bar Chart: Success Rate by Orbit Type**: Showed the success rate for different orbit types, helping identify which orbits have higher success percentages.
5. **Categorical Plot: FlightNumber vs. Orbit Type**: Examined trends between flight numbers and specific orbit types.
6. **Scatter Chart: Payload vs. Orbit Type**: Analysed the relationship between payload mass and orbit types.
7. **Line Chart: Yearly Success Trend**: Plotted average success rates by year to visualise how landing success has improved over time.

Notebook available on Github:
https://github.com/natashaleask/IBM-Applied-Data-Science-Capstone/blob/main/5_EDA_Dataviz.ipynb

EDA with SQL

I used various SQL techniques including filtering, aggregation functions, subqueries, string manipulation, and date functions, all applied to analyse SpaceX launch data.

Notebook available on Github: https://github.com/natashaleask/IBM-Applied-Data-Science-Capstone/blob/main/4_EDA_SQL.ipynb

Task 1: Query to display unique launch sites Retrieves all distinct launch locations used by SpaceX	`SELECT DISTINCT "LAUNCH_SITE" FROM SPACEXTABLE`
Task 2: Query to find all missions from launch site CCAFS LC-40 Returns the first 5 records where the launch site name begins with "CCA"	`SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE "CCA%" LIMIT 5`
Task 3: Query to calculate total payload mass for NASA (CRS) missions Sums up the payload mass across all NASA CRS missions (45,596 kg)	`SELECT SUM(PAYLOAD_MASS__KG_) AS "Total Payload Mass (kg)" FROM SPACEXTABLE WHERE Customer LIKE "NASA (CRS)"`
Task 4: Query to find average payload mass for F9 v1.1 boosters Calculates the average payload carried by the F9 v1.1 booster version (2,928.4 kg)	`SELECT AVG(PAYLOAD_MASS__KG_) AS "Average Payload Mass (kg)" FROM SPACEXTABLE WHERE Booster_Version LIKE "F9 v1.1"`
Task 5: Query to find the first successful ground pad landing Determines the earliest date when SpaceX achieved a successful ground pad landing (2015-12-22)	`SELECT MIN(Date) AS "First Successful Landing" FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (ground pad)"`

EDA with SQL

Task 6: Query to find boosters with successful drone ship landings and specific payload range Lists boosters that successfully landed on drone ships with payloads between 4000-6000 kg	`SELECT Booster_Version FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000`
Task 7: Query to count mission outcomes by success/failure Groups and counts missions by their outcome status	`SELECT Mission_Outcome, COUNT(*) FROM SPACEXTABLE GROUP BY Mission_Outcome`
Task 8: Query using a subquery to find boosters carrying maximum payload. Identifies which booster versions carried the heaviest payloads	`SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)`
Task 9: Query to find drone ship failures in 2015, showing month information Lists failed drone ship landings in 2015 with month details	`SELECT STRFTIME("%m", Date) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE STRFTIME("%Y", Date) = "2015" AND Landing_Outcome = 'Failure (drone ship)'`
Task 10: Query to rank landing outcomes by frequency Counts and ranks different landing outcomes between specified dates	`SELECT Landing_Outcome, COUNT(*) AS OutcomeCount FROM SPACEXTABLE WHERE Date BETWEEN "2010-06-04" AND "2017-03-20" GROUP BY Landing_Outcome ORDER BY OutcomeCount DESC`

Interactive Map: Folium

The overall purpose of these visualisations is to discover geographic factors that might influence launch success rates and to understand considerations for optimal launch site placement.

Task 1: Mark all launch sites on the map
This task involved creating a Folium map showing all SpaceX launch sites. The goal was to visualise the geographic distribution of these sites using their latitude and longitude coordinates.

Task 2: Mark success/failed launches for each site
This task enhances the map by visualising the success rate of launches at each site.
Green markers were used to represent successful launches (class=1)
Red markers were used to represent failed launches (class=0)
A MarkerCluster object was implemented to handle multiple launches at the same coordinates
This visualisation helped identify which sites have higher success rates

Task 3: Calculate distances between launch sites and proximities
This task explores the geographic relationships between launch sites and nearby features.

Task 4: Visualise the relationships with PolyLines
This task involves drawing lines between launch sites and nearby features.

Notebook available on Github:
https://github.com/natashaleask/IBM-Applied-Data-Science-Capstone/blob/main/6_Launch_site_locations_analysis_with_Folium.ipynb

Dashboard: Plotly Dash

I added two primary visualisations to the SpaceX Launch Records Dashboard:

1. **Success-Pie-Chart**: This visualisation shows either the overall success rate by launch site or detailed success/failure breakdown for a specific site.
2. **Success-Payload-Scatter-Chart**: This scatter plot displays the relationship between payload mass and mission success, color-coded by booster version.

These visualisations are connected to two key interactive components:

1. **Launch Site Dropdown**: Allows users to filter data by specific launch sites or view all sites.
2. **Payload Range Slider**: Enables users to focus on specific payload mass ranges.

The pie chart provides a clear picture of success rates across different launch sites, helping to identify the most reliable locations.

The scatter plot reveals critical payload-to-success relationships, allowing engineers to determine optimal payload ranges for future missions.

Notebook available on Github:
https://github.com/natashaleask/IBM-Applied-Data-Science-Capstone/blob/main/7_Dash.py

Predictive Analysis (Classification)

1. **Data collection**: The Dataset contains information about previous Falcon 9 launches, including technical specifications and outcomes.
2. **Data preparation**: Created the target variable (Y) from the 'Class' column, where 1 indicates a successful landing and 0 indicates an unsuccessful landing.
3. **Feature engineering**: Standardised the features using StandardScaler to ensure all features contribute equally to the model and to improve model convergence.
4. **Data splitting**: Created training and test sets with an 80/20 split, maintaining the class distribution to ensure representative evaluation.
5. **Model selection**: Four classification algorithms were selected to compare their performance on this specific prediction task.
6. **Hyperparameter tuning**: For each model, used GridSearchCV with 10-fold cross-validation to find the optimal hyperparameters, exploring various combinations systematically.
7. **Model training**: Trained each model with their optimal hyperparameters on the training data.
8. **Model evaluation**: Evaluated each model's performance on the test set using multiple metrics to get a comprehensive view of model performance.

Github: https://github.com/natashaleask/IBM-Applied-Data-Science-Capstone/blob/main/7_Machine_Learning_Prediction.ipynb

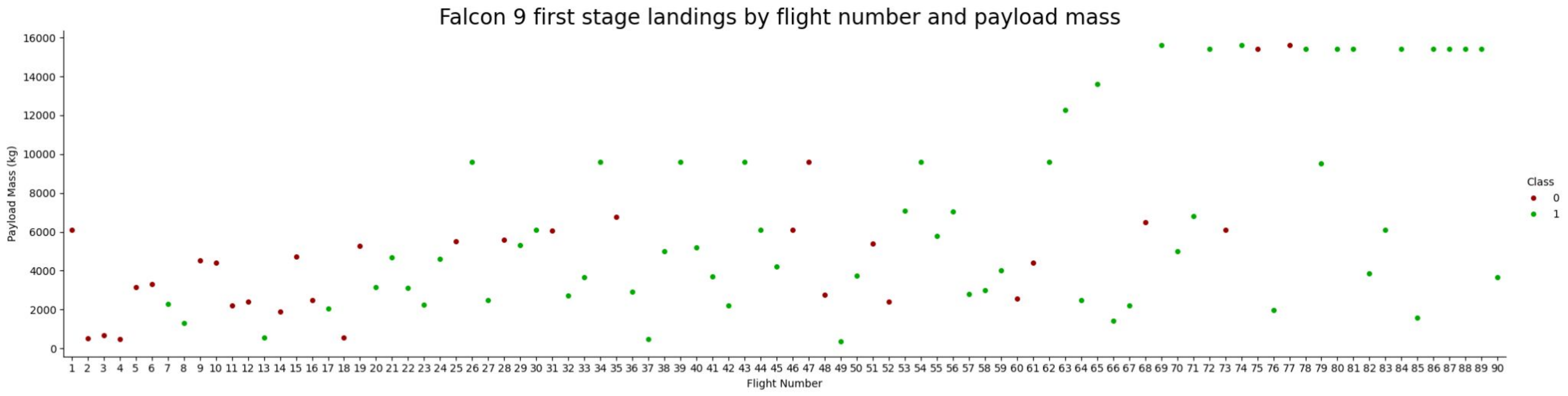
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Findings

1. Most of the failures (red dots) are clustered in the early flight numbers (1 - 30).
2. After flight 30, we see more green dots, indicating improved success rates over time.
3. There is no clear correlation between payload mass and success rate. Payload mass alone, does not determine success or failure. Landing success may be related to other improvements over time.
4. Payload mass does not prevent successful landings. Some heavier missions have landed successfully.
5. There is cluster of very high mass payloads (15,000 kg +) in later flights (65-90), almost all of which were successful landings. This might indicate that as SpaceX gained experience, they became more confident in landing heavier payloads.



Payload vs. Launch Site

Findings

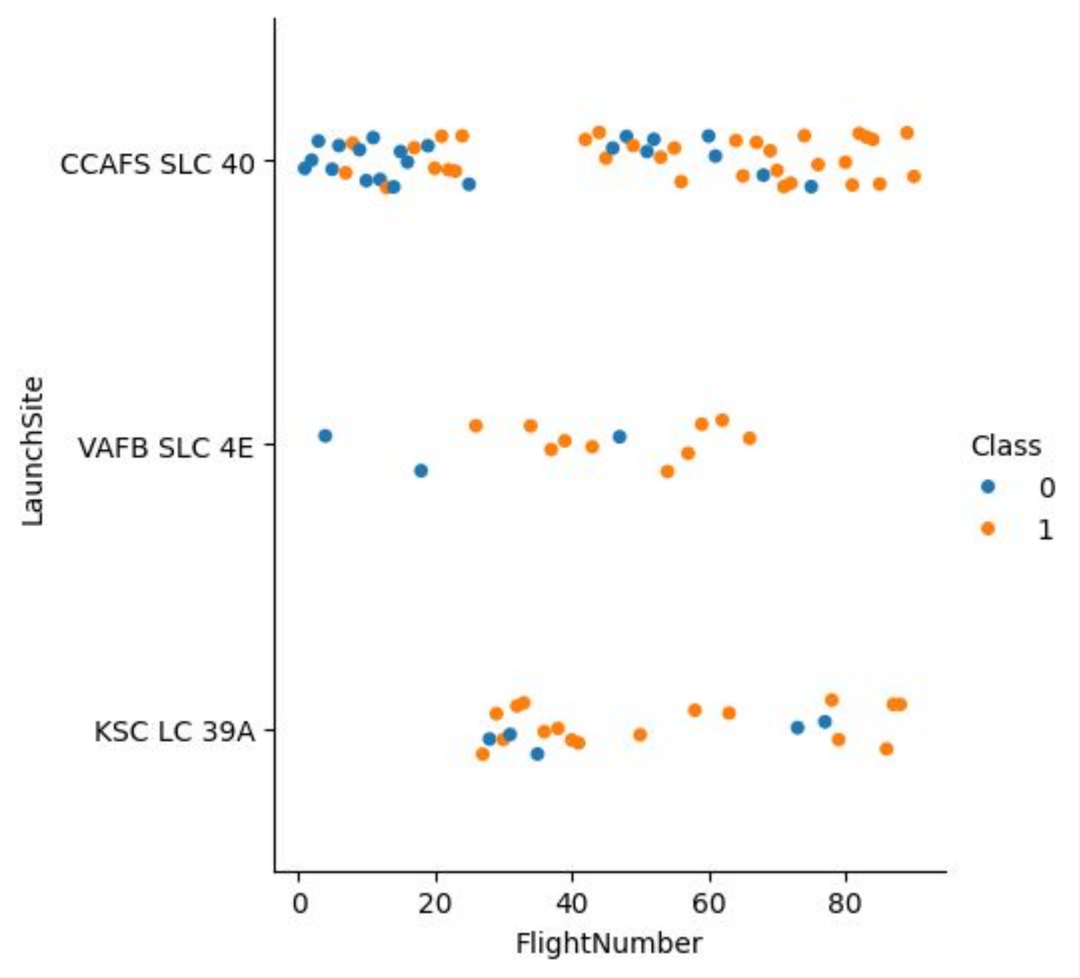
CCAFS SLC 40 (Cape Canaveral) Used most frequently in early flights and consistently over time.
This site had early failures but became very reliable over time.

VAFB SLC 4E (Vandenberg AFB, California) Fewer launches overall. Still a mix of failures and successes, but appears to show more consistent success than failures over time.

KSC LC 39A (Kennedy Space Center) Starts getting used only after Flight 25. A few early failures, but mostly successful launches overall.

Launch site matters in context of time.

All sites indicate improved success rates as SpaceX gains experience. No site inherently performs worse — earlier launches were just riskier regardless of location. CCAFS SLC 40 had the roughest start but turned into a reliable workhorse. KSC LC 39A and VAFB SLC 4E were used more once SpaceX had more experience, so they have fewer failures overall.



Success Rate vs. Orbit Type

Findings

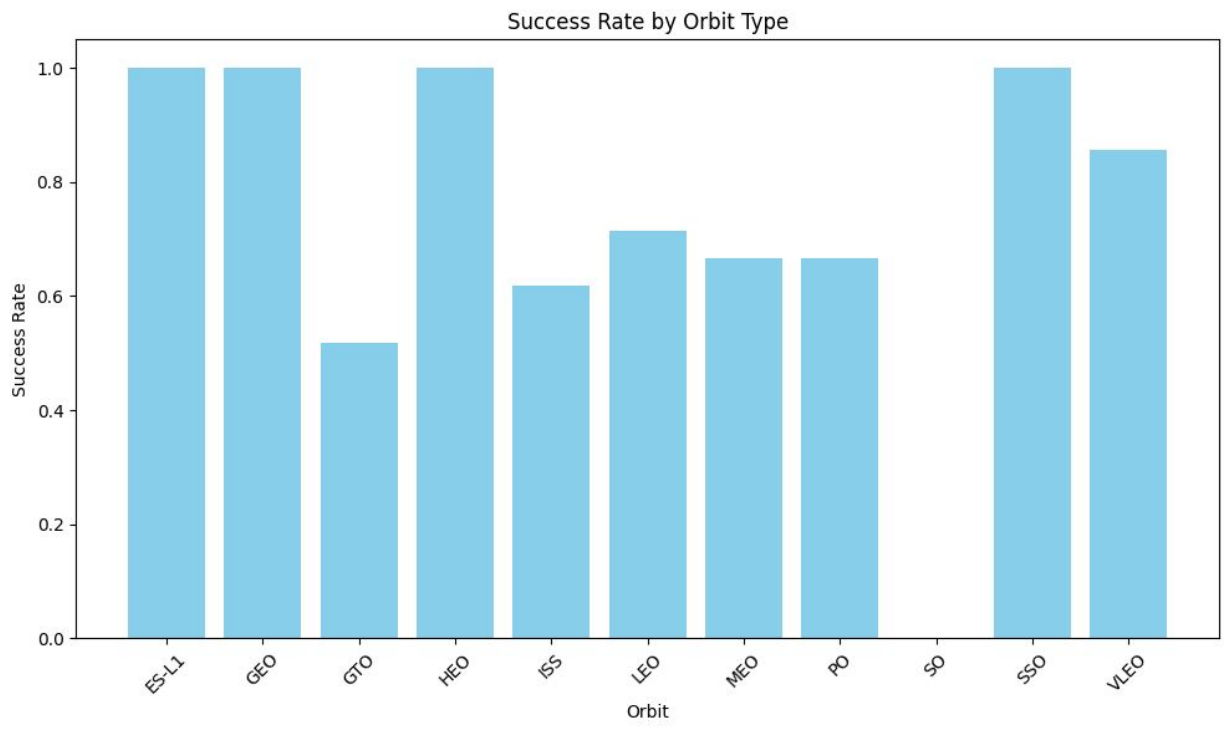
These orbits show a success rate of 1.0, meaning they had no failed missions in the dataset:ES-L1, GEO, HEO, SSO

These are the most reliable orbit types based on the data.

These orbits have success rates between ~0.6 and ~0.85, indicating a mix of successful and failed missions: VLEO, LEO, MEO, PO, SO

These are somewhat reliable but not perfect. There's room for improvement or further investigation.

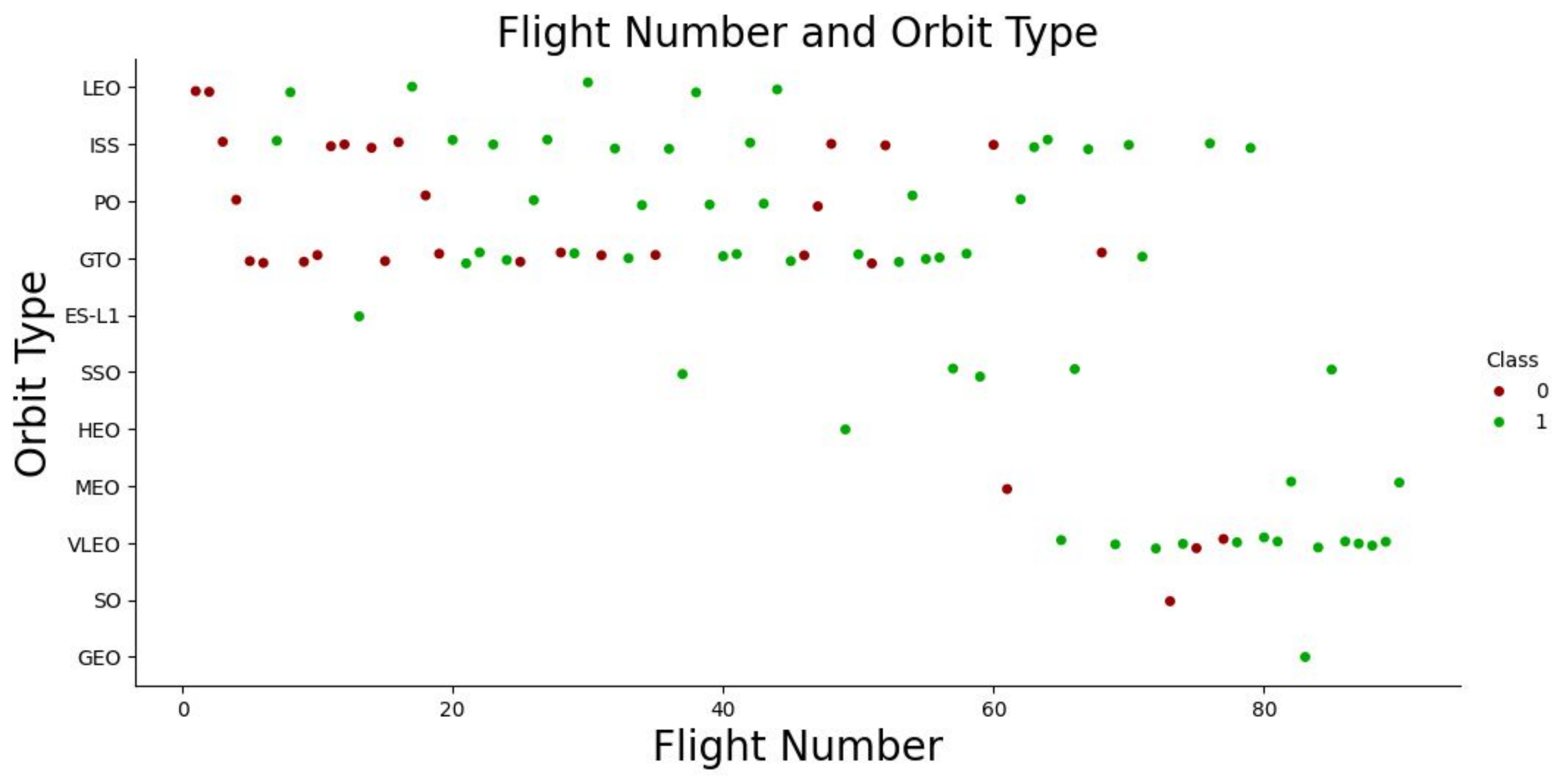
This orbit has the lowest success rate in the chart: GTO (around 0.52 or 52%)



Flight Number vs. Orbit Type

Findings

In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

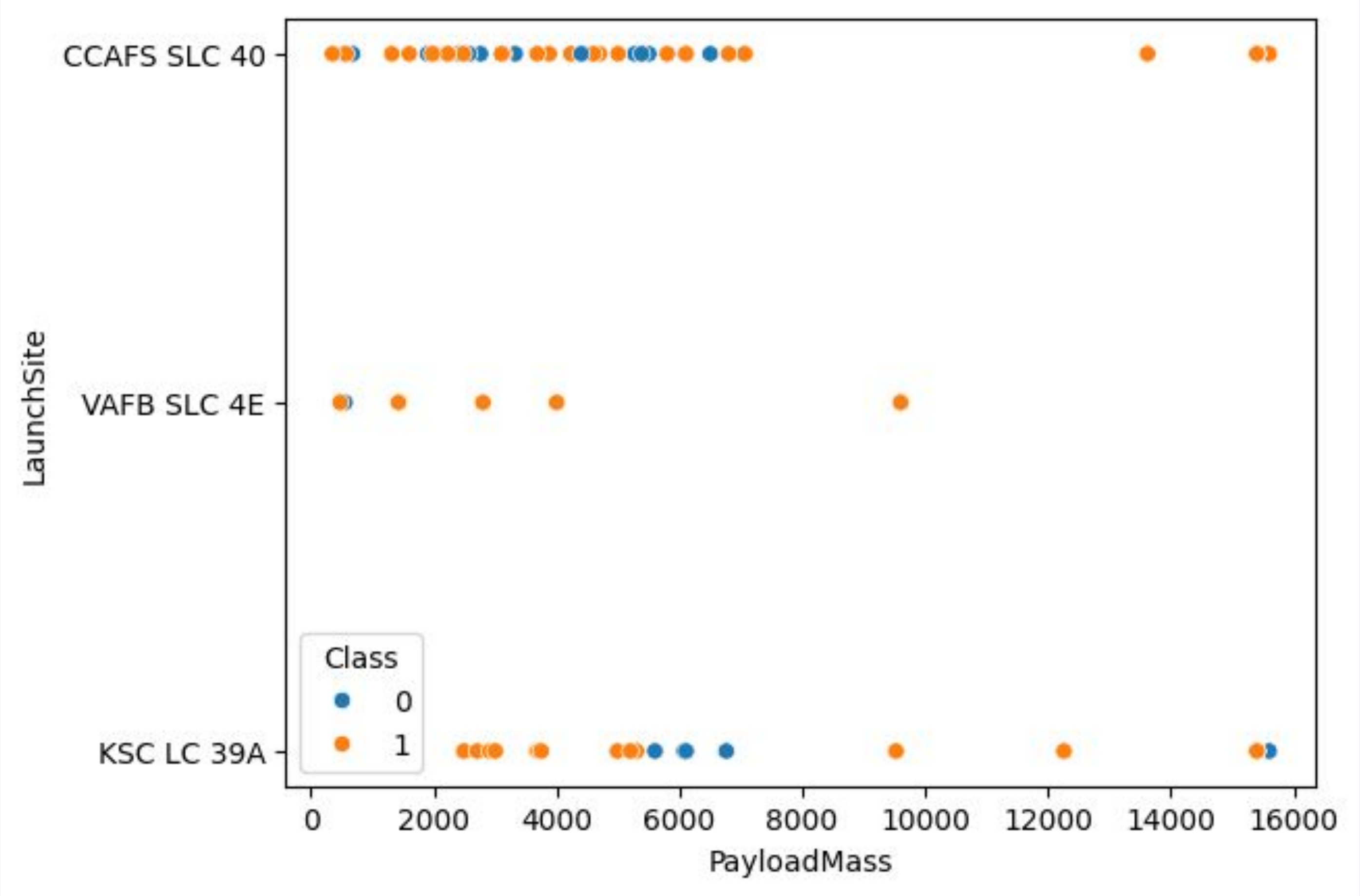


Payload Mass

Findings

VAFB SLC 4E (Vandenberg AFB, California)

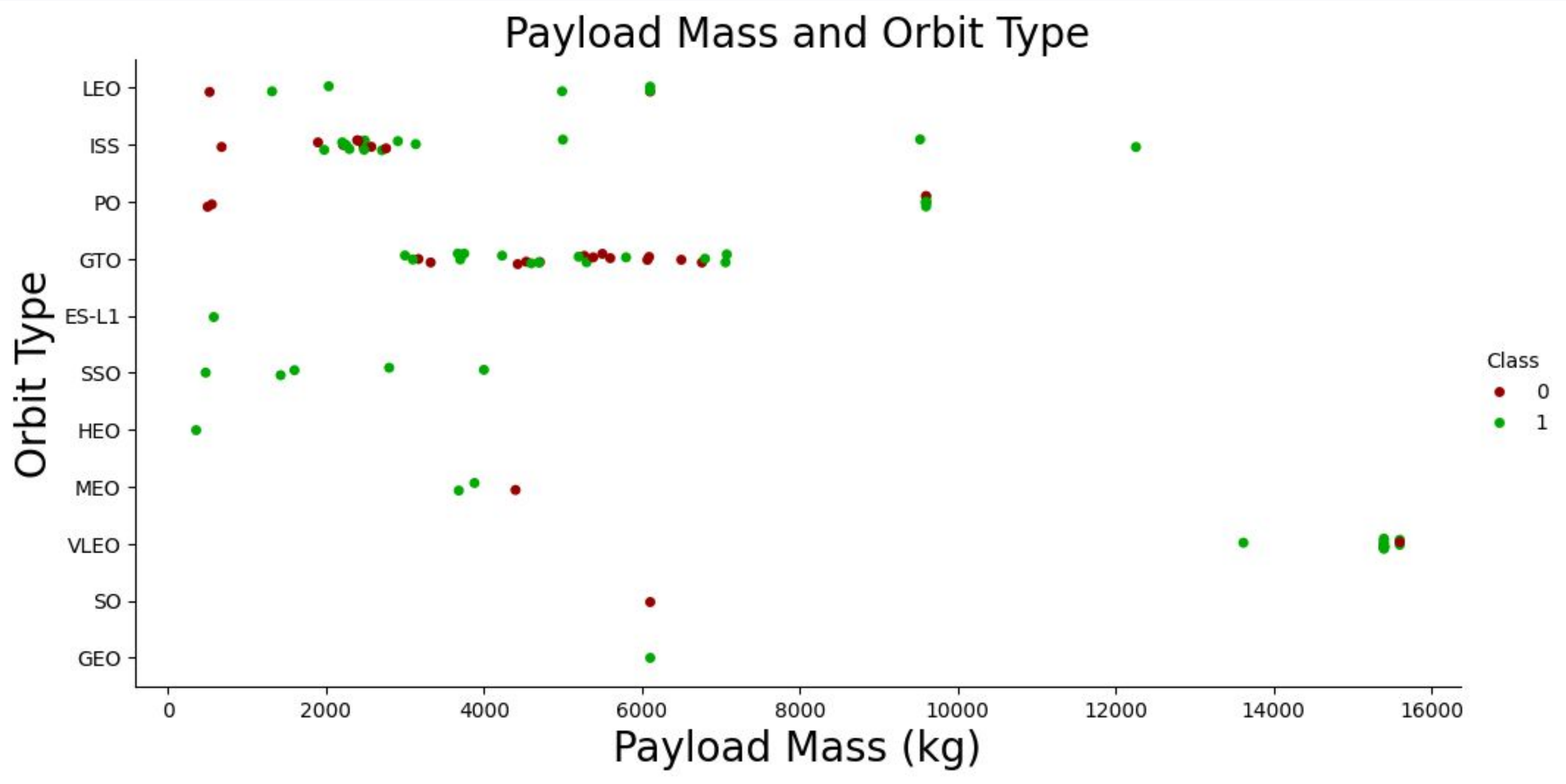
For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass (greater than 10000).



Payload vs. Orbit Type

Findings
With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

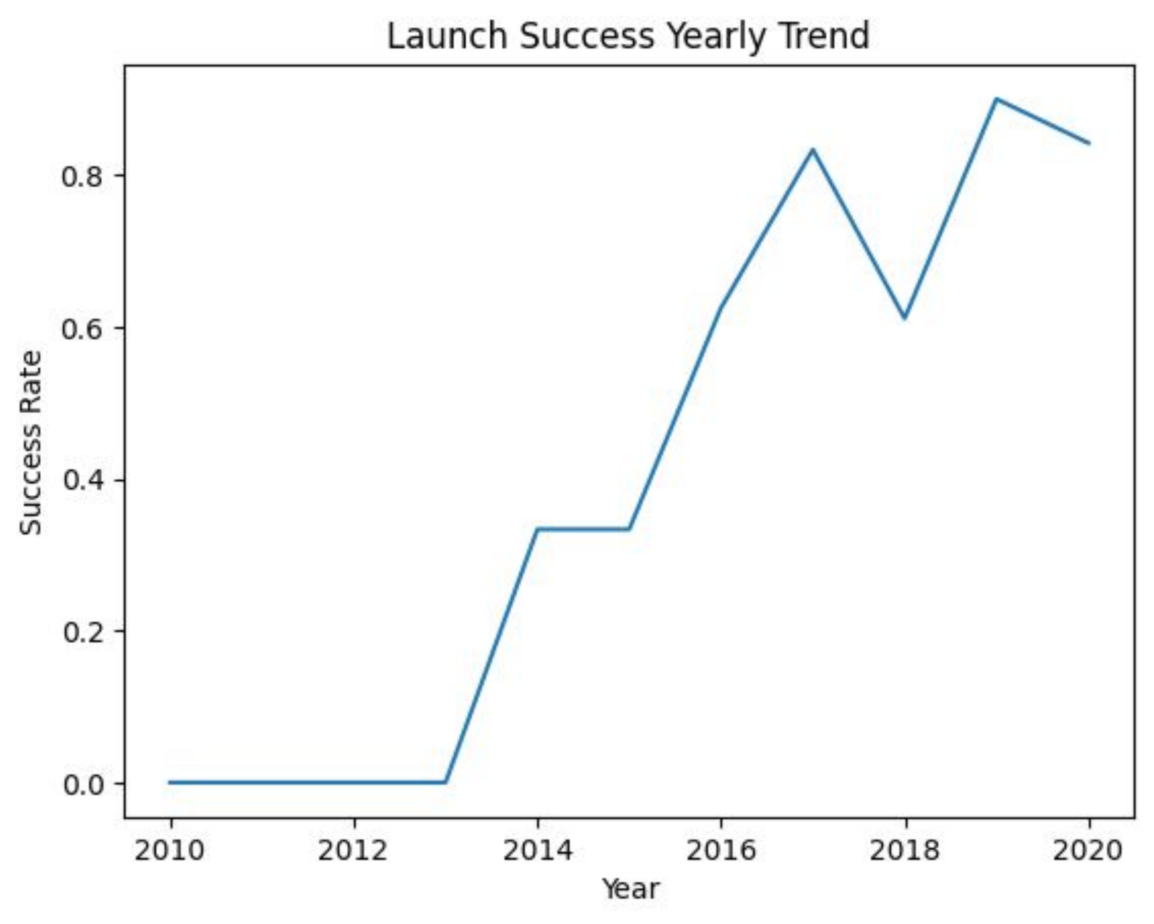
However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



Launch Success Yearly Trend

Findings

The success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.



All Launch Site Names

Task 1: Query to display unique launch sites

- `SELECT DISTINCT "LAUNCH_SITE" FROM SPACEXTABLE`
- Retrieves all distinct launch locations used by SpaceX

Query Result:

This result shows the four distinct launch sites used by SpaceX for their missions. The query used `SELECT DISTINCT "LAUNCH_SITE" FROM SPACEXTABLE` to eliminate duplicate values and show only unique launch locations. Three of the sites are located in Florida (the CCAFS sites at Cape Canaveral Air Force Station and KSC at Kennedy Space Center), while one is in California (VAFB at Vandenberg Air Force Base). SpaceX strategically uses different launch sites depending on the desired orbit and mission requirements.

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

Launch Site Names Begin with 'CCA'

Task 2: Query to find all missions from launch site CCAFS LC-40

- `SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE "CCA%" LIMIT 5`
- Returns the first 5 records where the launch site name begins with "CCA"

Query Result:

This query result shows the first 5 SpaceX missions launched from sites beginning with "CCA" (Cape Canaveral Air Force Station). The data reveals:

- All 5 missions used the earliest Falcon 9 booster version (F9 v1.0)
- They represent SpaceX's earliest launches from 2010-2013
- All launches were from the CCAFS LC-40 pad specifically
- The first two missions ended with parachute landing failures, while the later three had no landing attempts
- These early missions included Dragon qualification tests, demo flights, and the first Commercial Resupply Service (CRS) missions to the International Space Station

	Date	Time_UTC	Booster_Version	Launch_Site
0	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40
1	2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40
2	2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40
3	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40
4	2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40

Total Payload Mass

Task 3: Query to calculate total payload mass for NASA (CRS) missions

- `SELECT SUM(PAYLOAD_MASS__KG_) AS "Total Payload Mass (kg)" FROM SPACEXTABLE WHERE Customer LIKE "NASA (CRS)"`
- Sums up the payload mass across all NASA CRS missions (45,596 kg)

Query Result:

This query result shows the total combined payload mass (45,596 kg) carried by all SpaceX boosters for NASA's Commercial Resupply Services (CRS) missions.

Total Payload Mass (kg)
45596

Average Payload Mass by F9 v1.1

Task 4: Query to find average payload mass for F9 v1.1 boosters

- `SELECT AVG(PAYLOAD_MASS__KG_) AS "Average Payload Mass (kg)" FROM SPACEXTABLE WHERE Booster_Version LIKE "F9 v1.1"`
- Calculates the average payload carried by the F9 v1.1 booster version (2,928.4 kg)

Query Result:

This query result shows that the Falcon 9 v1.1 booster version carried an average payload mass of 2,928.4 kg across all its missions.

Average Payload Mass (kg)
2928.4

First Successful Ground Landing Date

Task 5: Query to find the first successful ground pad landing

- `SELECT MIN(Date) AS "First Successful Landing" FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (ground pad)"`
- Determines the earliest date when SpaceX achieved a successful ground pad landing (2015-12-22)

Query Result:

This query result shows that 22 December 2015 was the date of SpaceX's first successful landing on a ground pad.

First Successful Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6: Query to find boosters with successful drone ship landings and specific payload range

- `SELECT Booster_Version FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000`
- Lists boosters that successfully landed on drone ships with payloads between 4000-6000 kg

Query Result:

This query result shows four specific Falcon 9 boosters that successfully landed on drone ships while carrying payloads weighing between 4,000 and 6,000 kg.

All of these boosters belong to the "Full Thrust" version of the Falcon 9, which had significantly improved performance over earlier versions. Interestingly, two of these boosters (B1021.2 and B1031.2) have the ".2" suffix, indicating they were on their second flight - demonstrating successful reuse of rockets that had previously flown and landed.

	Booster_Version
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Task 7: Query to count mission outcomes by success/failure

- `SELECT Mission_Outcome, COUNT(*) FROM SPACEXTABLE GROUP BY Mission_Outcome`
- Groups and counts missions by their outcome status

Query Result:

The data shows:

- SpaceX has had an overwhelmingly successful track record with 98 clear successes
- There was only 1 in-flight failure
- There's 1 mission marked with a space in the success field (likely a data entry inconsistency)
- There was 1 mission classified as "Success (payload status unclear)" where the rocket performed correctly but there was uncertainty about the payload's final status

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

Boosters Carried Maximum Payload

Task 8: Query using a subquery to find boosters carrying maximum payload

- `SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)`
- Identifies which booster versions carried the heaviest payloads

Query Result:

All identified boosters are from the "Block 5" (B5) variant - SpaceX's final and most advanced version of the Falcon 9.

Many of these boosters show numbers after the decimal point (like B1051.6, indicating they were on their 6th flight), demonstrating SpaceX's reusability program.

	Booster_Version
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5
5	F9 B5 B1051.4
6	F9 B5 B1049.5
7	F9 B5 B1060.2
8	F9 B5 B1058.3
9	F9 B5 B1051.6
10	F9 B5 B1060.3
11	F9 B5 B1049.7

2015 Launch Records

Task 9: Query to find drone ship failures in 2015, showing month information

- `SELECT STRFTIME("%m", Date) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE STRFTIME("%Y", Date) = "2015" AND Landing_Outcome = 'Failure (drone ship)'`
- Lists failed drone ship landings in 2015 with month details

Query Result:

The results show that SpaceX experienced two drone ship landing failures in early 2015:

- One in January using booster F9 v1.1 B1012
- One in April using booster F9 v1.1 B1015

Both failures occurred at the Cape Canaveral Air Force Station LC-40 launch site.

	Month	Landing_Outcome	Booster_Version	Launch_Site
0	01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
1	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Task 10: Query to rank landing outcomes by frequency

- `SELECT Landing_Outcome, COUNT(*) AS OutcomeCount FROM SPACEXTABLE WHERE Date BETWEEN "2010-06-04" AND "2017-03-20" GROUP BY Landing_Outcome ORDER BY OutcomeCount DESC`
- Counts and ranks different landing outcomes between specified dates

Query Result:

The results show that in SpaceX's early years:

- "No attempt" was the most common outcome (10 missions), reflecting that landing and recovery weren't initially attempted for many flights
- Drone ship landings were attempted frequently, with equal numbers of successes and failures (5 each)
- Ground pad landings were less common but more successful (3 successes)
- Various ocean landing attempts (controlled and uncontrolled) were also part of SpaceX's development process
- The early parachute landing approach (used in the very first missions) was attempted twice, both failing

	Landing_Outcome	OutcomeCount
0	No attempt	10
1	Success (drone ship)	5
2	Failure (drone ship)	5
3	Success (ground pad)	3
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Failure (parachute)	2
7	Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch sites’ locations

Are all launch sites in proximity to the Equator line?

These launch sites are not particularly close to the Equator (0° latitude). They're all in the northern hemisphere between 28° and 35° North latitude. The first three sites (CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A) are all in Florida at around 28.5°N, while VAFB SLC-4E is in California at about 34.6°N.

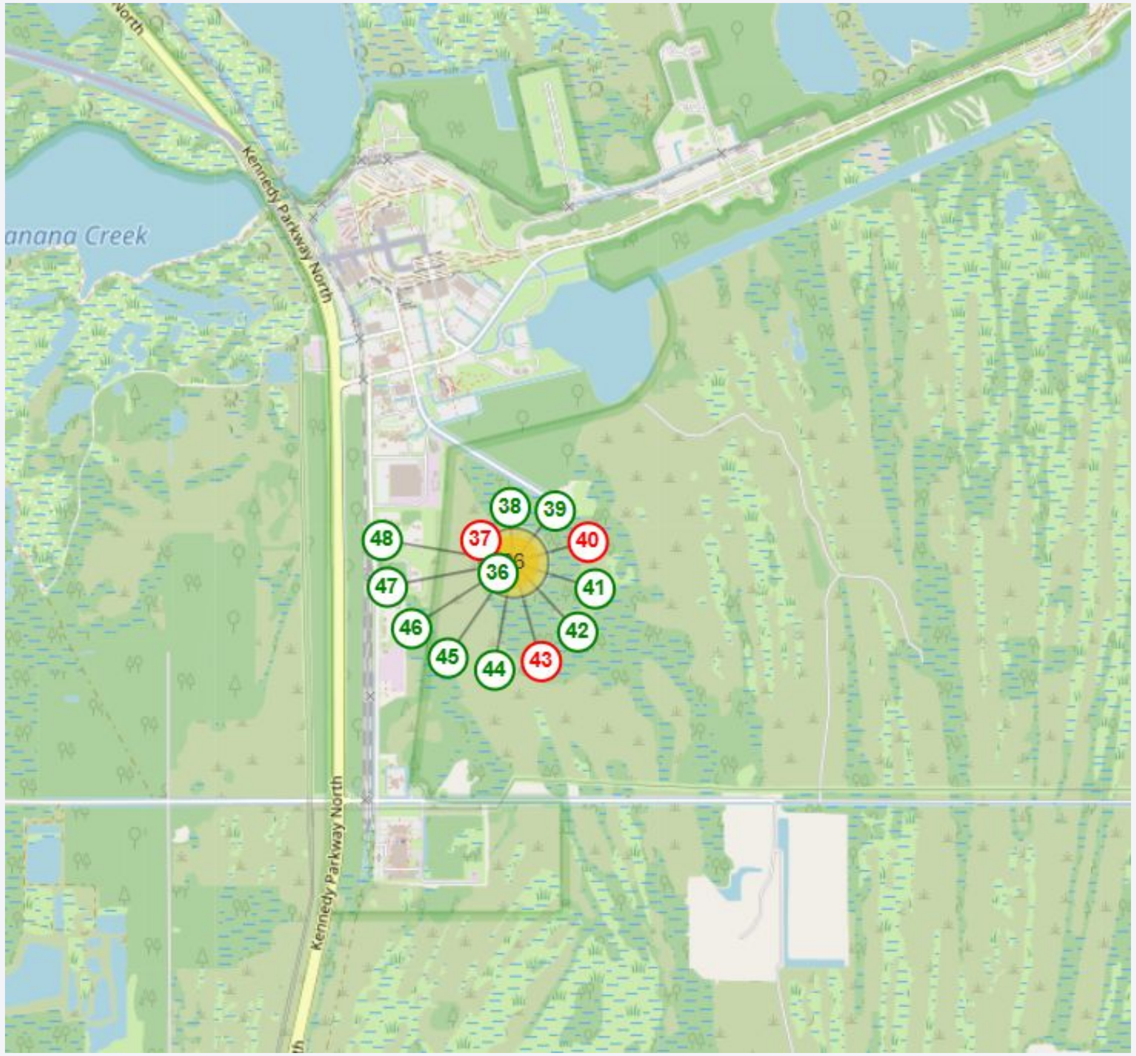
Are all launch sites in very close proximity to the coast?

Yes, all of these launch sites are located very close to coastlines. The three Florida sites (CCAFS and KSC) are located along the Atlantic coast at Cape Canaveral The VAFB site (Vandenberg Air Force Base) is located on the Pacific coast in California

Launch outcomes

KSC LC 39A (Kennedy Space Center)

KSC LC-39A has a launch success rate of **76.9%**, with **10 successful launches out of 13**.

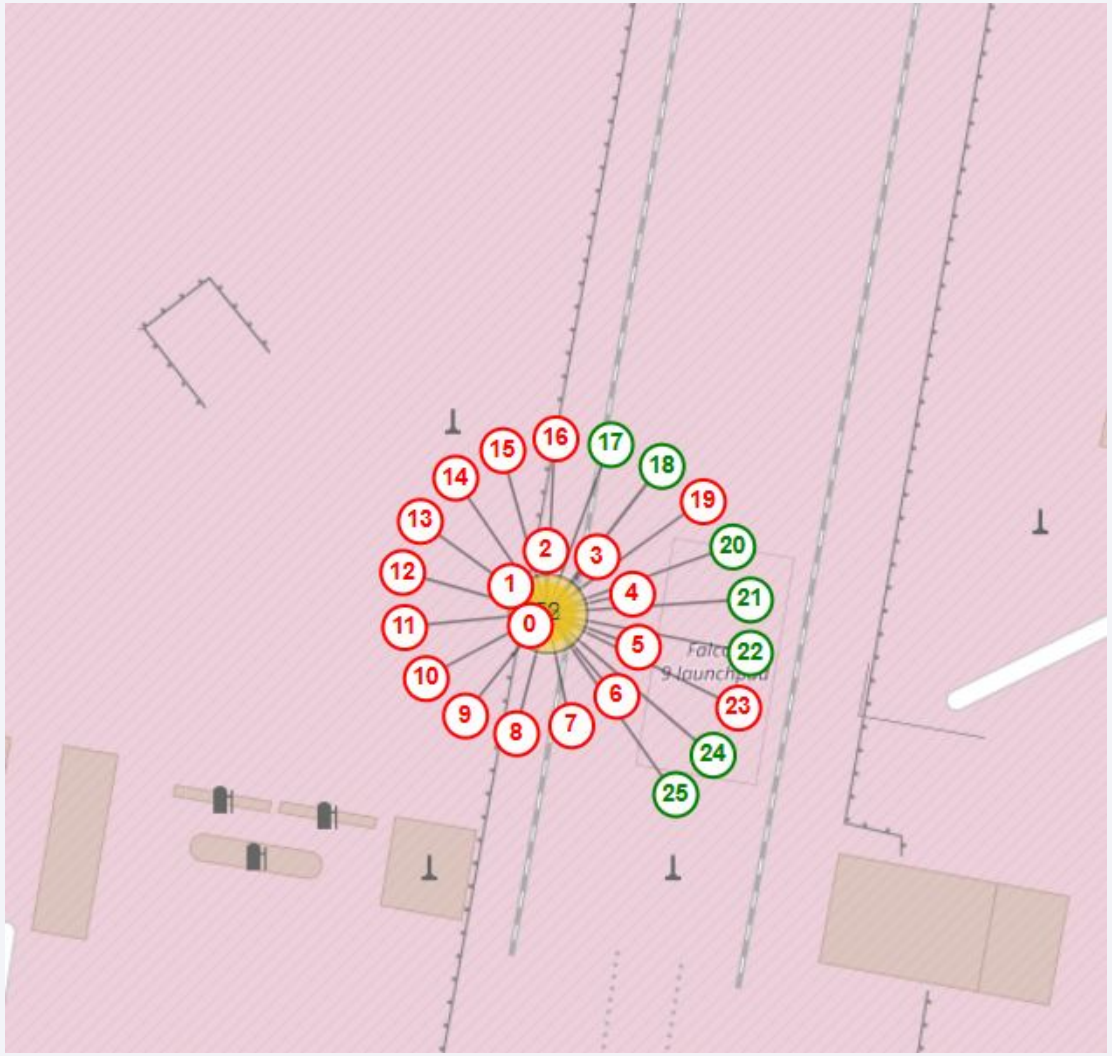


Launch outcomes

CCAFS SLC 40 (Cape Canaveral)

Cape Canaveral's SLC-40 launch complex has been an important facility for SpaceX operations, hosting numerous Falcon 9 launches for both commercial and government missions.

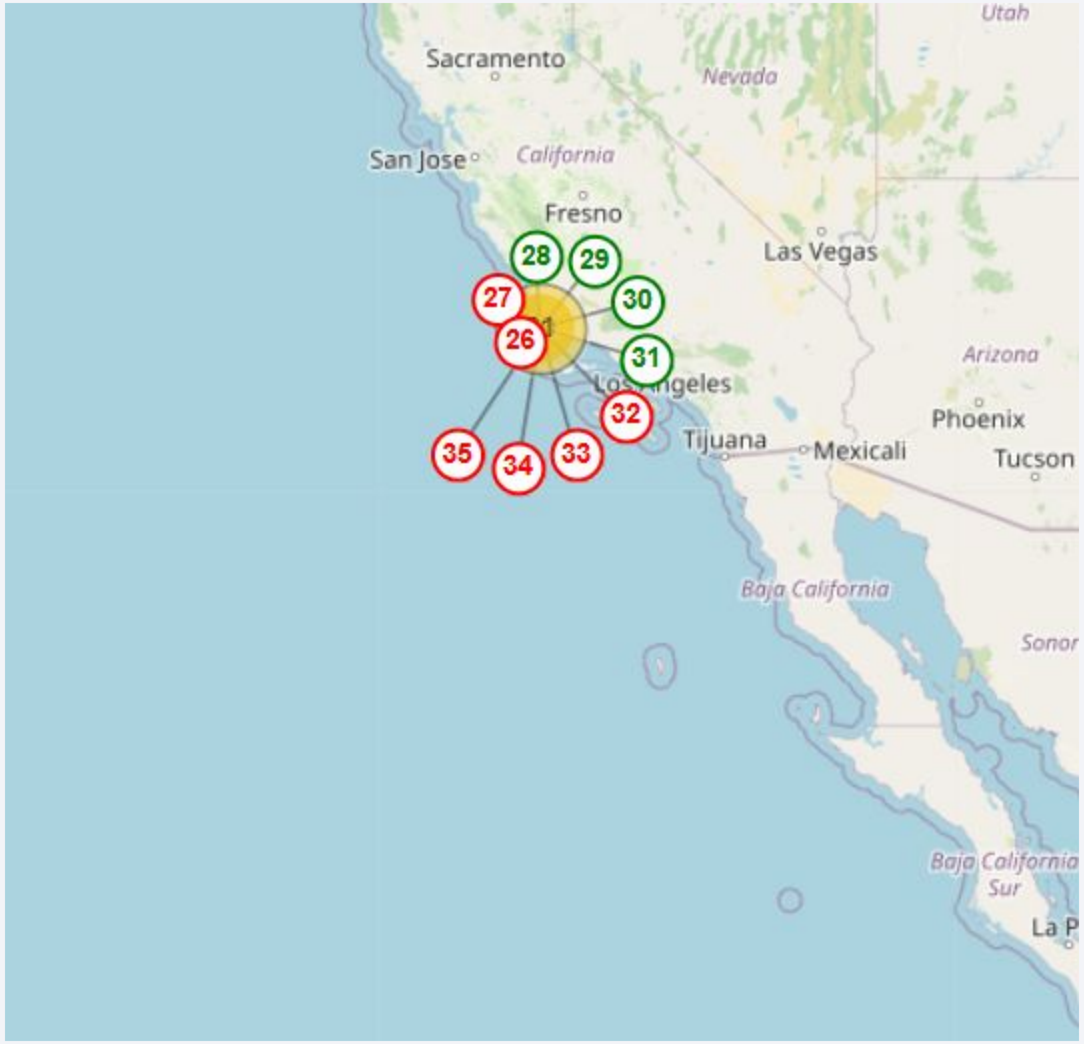
The data suggests it's a reliable launch facility, though not the highest performing one.



Launch outcomes

VAFB SLC 4E (Vandenberg AFB, California)

VAFB SLC-4E is used less frequently and has a launch success rate of 16.7%.



Launch outcomes

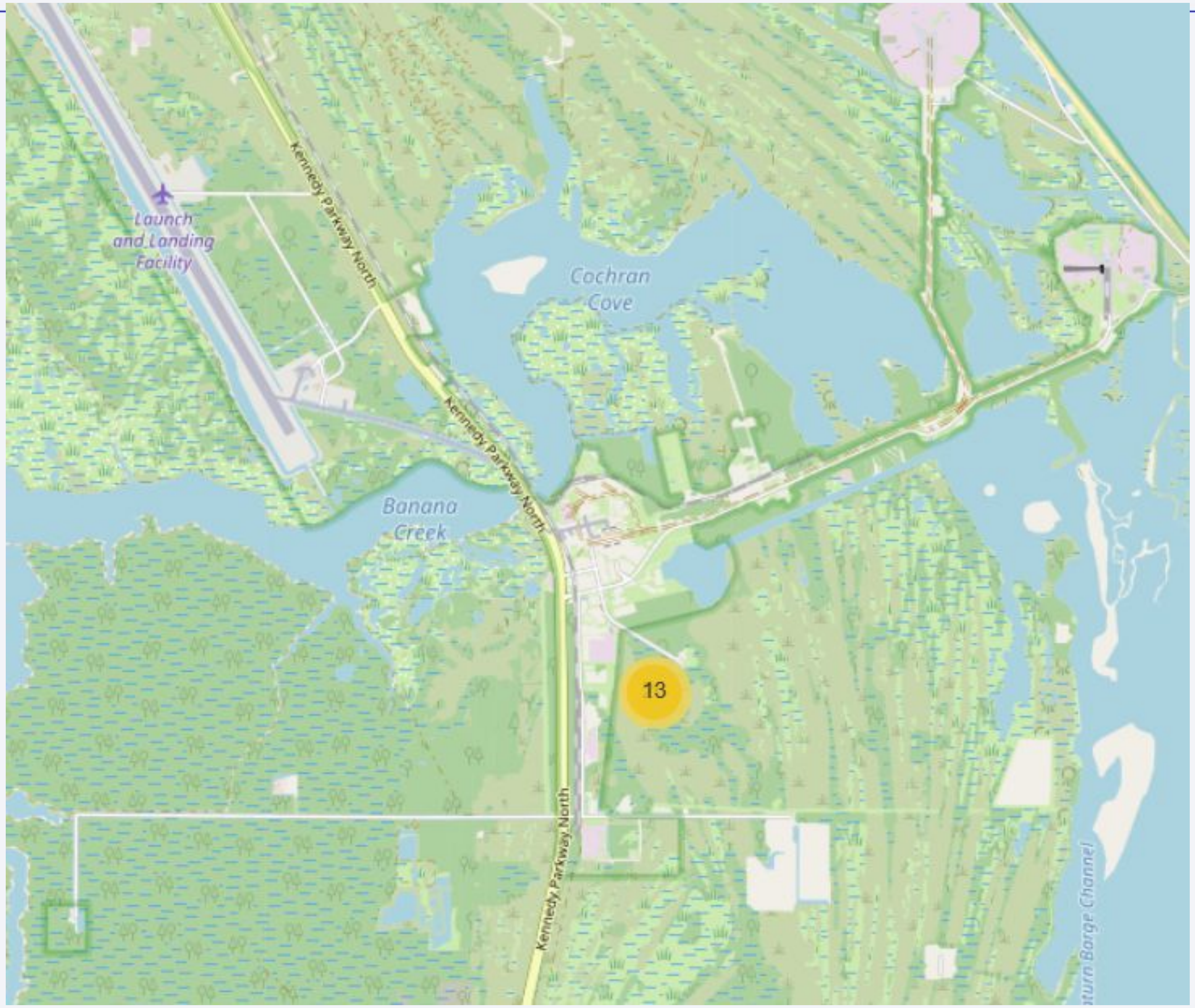
KSC LC 39A (Kennedy Space Center)

Proximity to railways:
A rail spur connects KSC to the mainline that runs parallel to NASA Parkway.

Proximity to highways:
The complex is served by Kennedy Parkway and NASA Parkway, linking it to major road networks.

Proximity to coastline:
Launch Complex 39A sits just a few hundred meters from the Atlantic shoreline.

Distance from cities:
KSC is isolated from cities like Titusville and Cocoa Beach, separated by rivers and wildlife preserves.



Launch outcomes

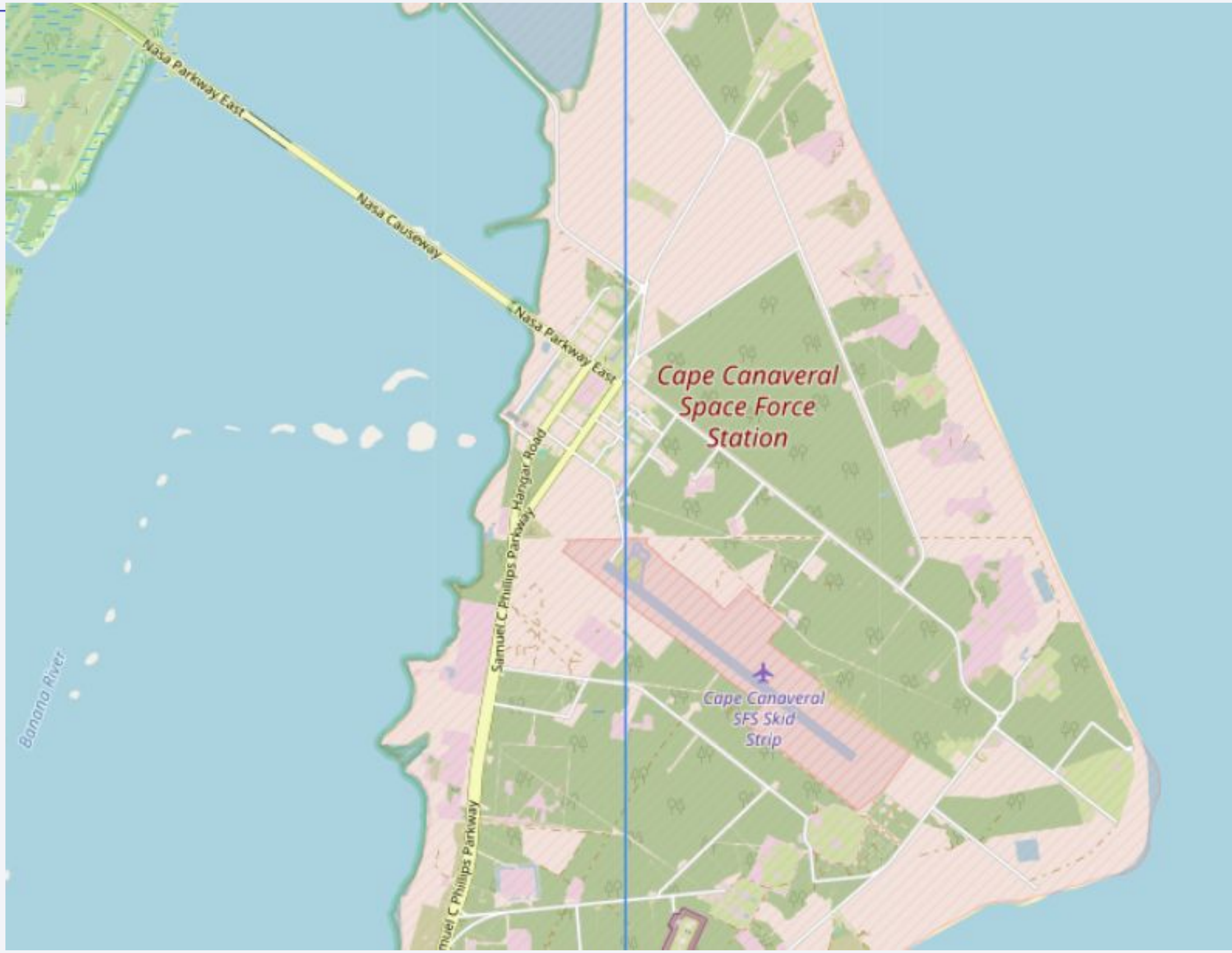
CCAFS SLC 40 (Cape Canaveral)

Proximity to railways:
A rail line enters CCSFS from the mainland, running parallel to Samuel C. Phillips Parkway.

Proximity to highways:
Connected to NASA Parkway and Max Brewer Memorial Parkway, facilitating access from Titusville and surrounding regions.

Proximity to coastline:
Launch pads are directly on the Atlantic coast, pointing east over open ocean.

Distance from cities:
The station is buffered by waterways and natural areas from nearby cities like Titusville (~10–15 km away).



Launch outcomes

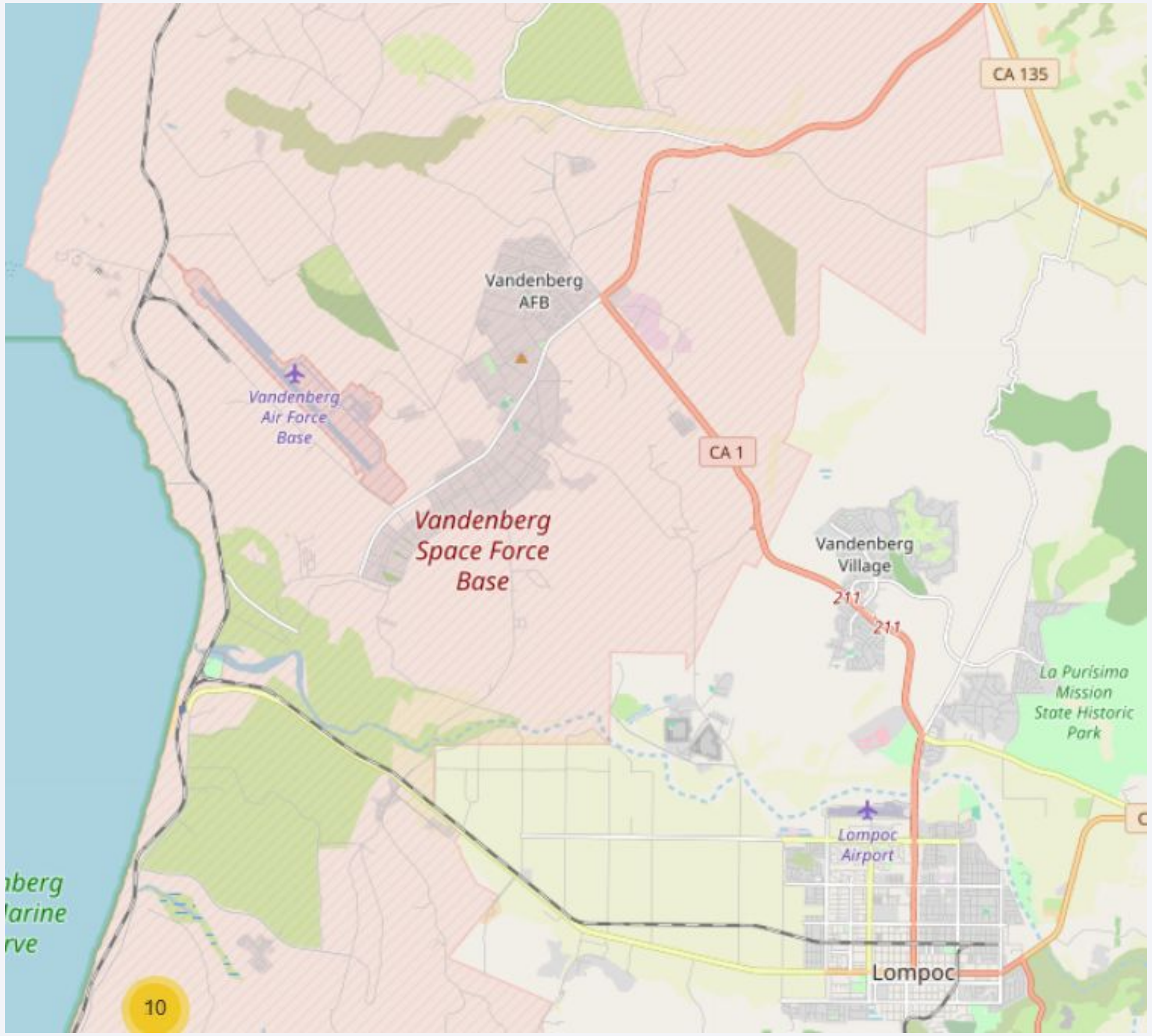
VAFB SLC 4E (Vandenberg AFB, California)

Proximity to railways:
A railway line runs directly alongside the base, visible on the map entering from the southeast.

Proximity to highways:
Located near California State Route 1 (CA-1) and Highway 246, which connect the base to Lompoc and surrounding areas.

Proximity to coastline:The base is directly on the Pacific coastline, ideal for polar and sun-synchronous launches.

Distance from cities:
VSFB is about 5–10 km from the nearest urban area, Lompoc.

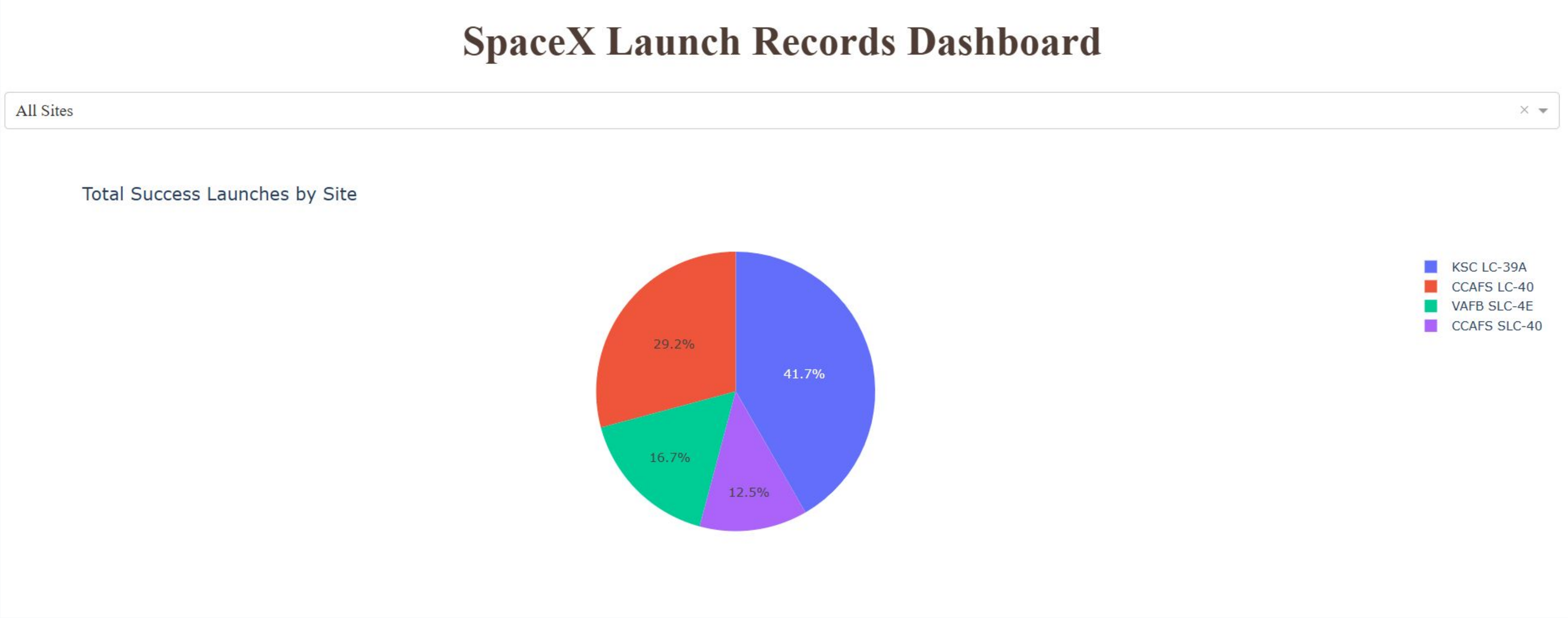


Section 4

Build a Dashboard with Plotly Dash

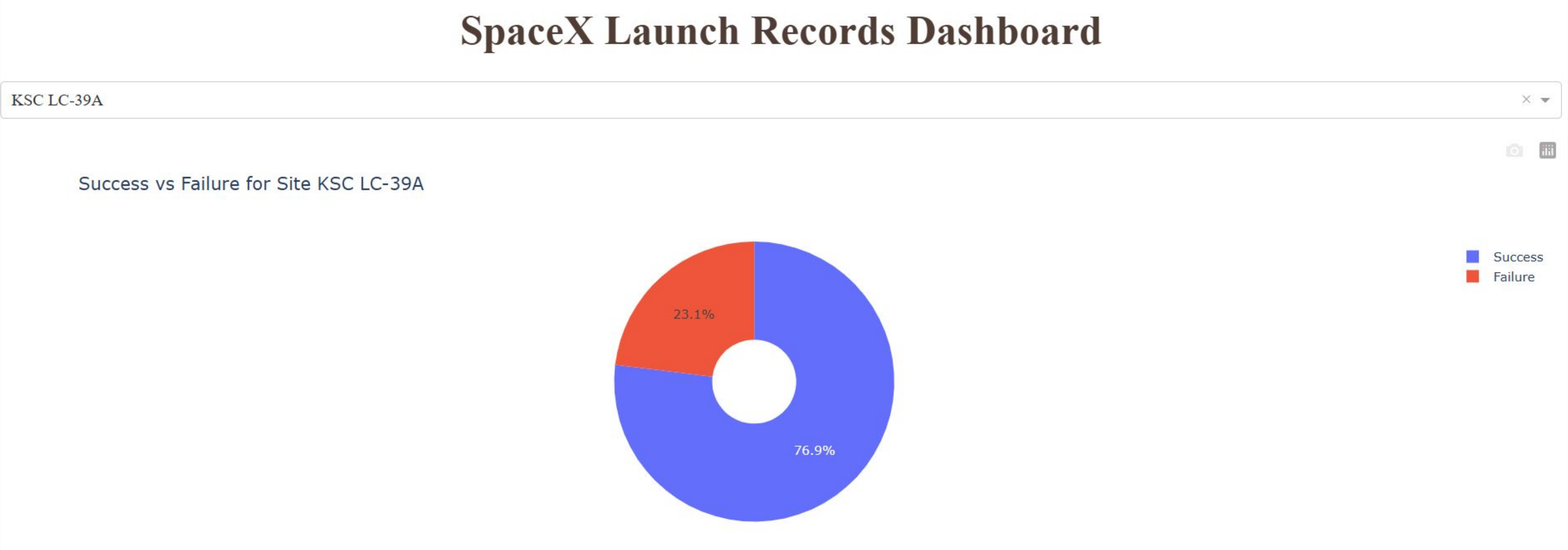
Launch success count for all sites

The pie chart displays SpaceX launch success distribution across four sites, with KSC LC-39A having the highest success rate at 41.7%, followed by CCAFS LC-40 at 29.2%, VAFB SLC-4E at 16.7%, and CCAFS SLC-40 at 12.5%.



Highest launch success ratio

The KSC LC-39A launch site has achieved a high success rate of 76.9% with only 23.1% of launches resulting in failure.



Payload vs. Launch Outcome

Payload Range Analysis:

- The scatter plots show that payloads in the mid-range (approximately 3000-5000 kg) have a higher concentration of successful launches (value of 1)
- Lower payload ranges (under 2500 kg) and higher payload ranges (above 6000 kg) show more mixed results with both successes and failures

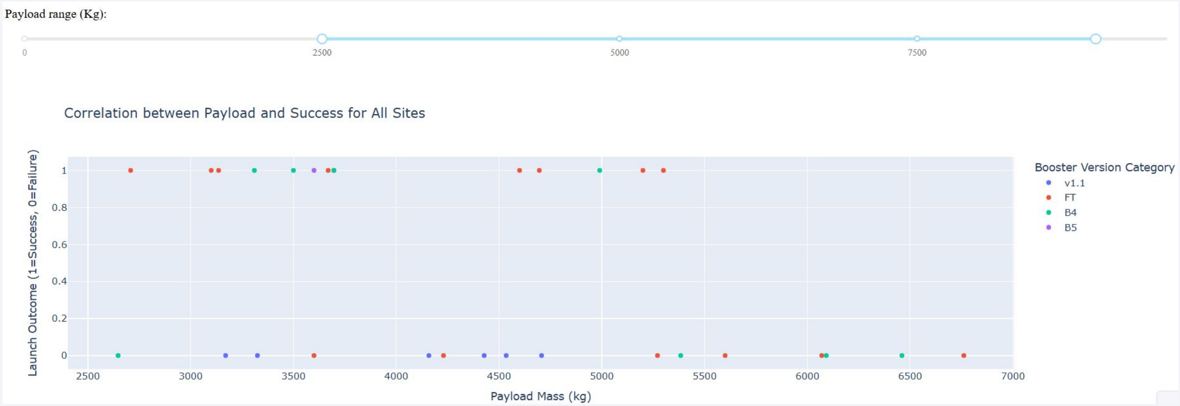
Booster Version Performance:

- The FT booster version appears to have the most consistent success rate across different payload masses
- B4 booster version also shows good performance, particularly in the mid-range payloads
- V1.0 and v1.1 versions show more mixed results

Payload vs. Launch Outcome

Key Findings:

- The most optimal payload range for success appears to be between 3000-5000 kg where there is a higher density of successful launches
- When filtering to the higher payload ranges (Image 2, 5200-6800 kg), there are fewer successful launches overall
- Image 3 shows the full payload range (0-7000 kg) and confirms that the middle ranges have better success rates than the extremes
- There's a clear pattern showing that very light payloads (under 1000 kg) and very heavy payloads (over 6000 kg) have more failures (value of 0) compared to the medium payload ranges



Section 5

Predictive Analysis (Classification)

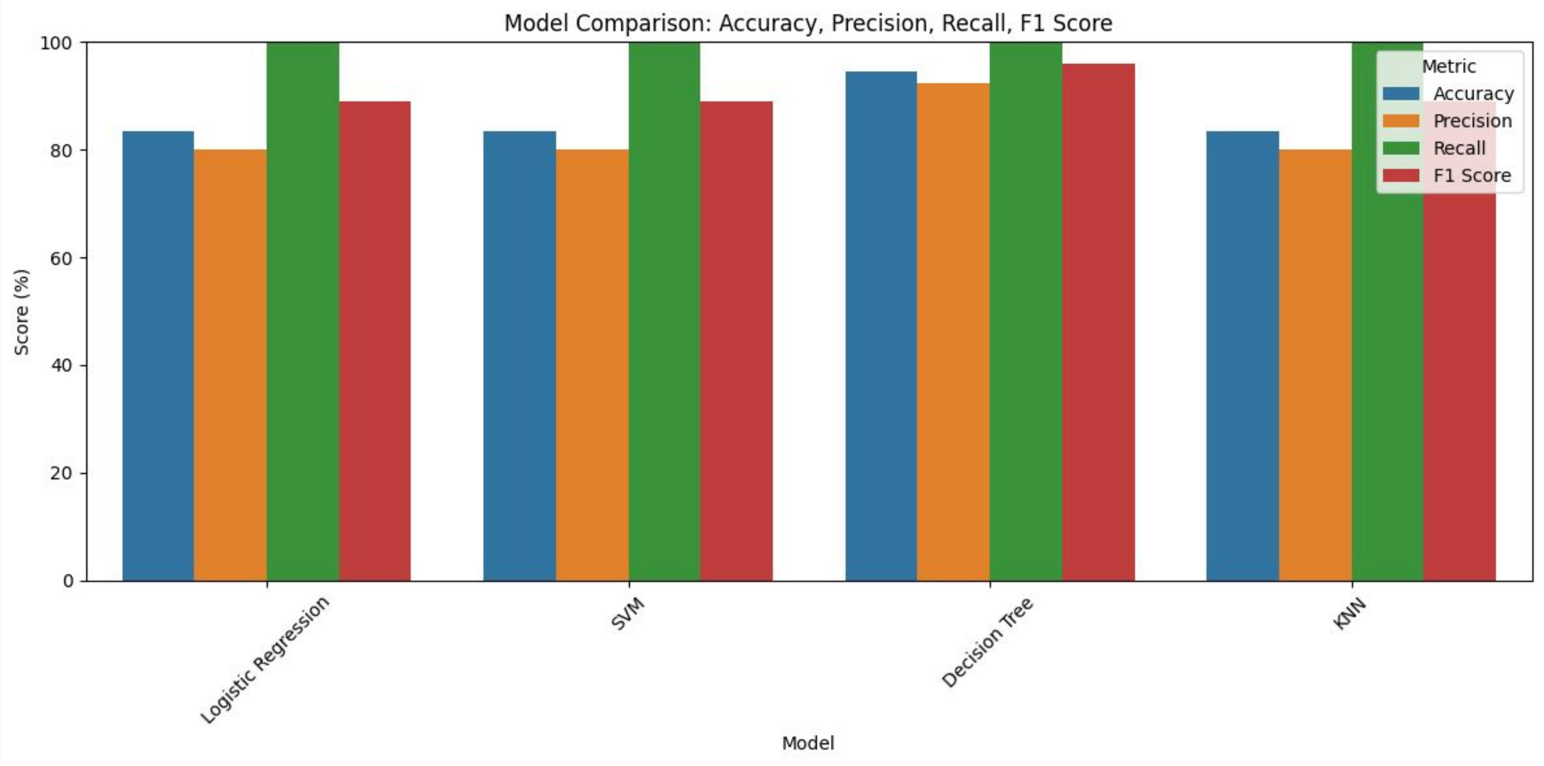
Classification Accuracy

Key Findings:

Decision Tree: 94.44% accuracy, 92.31% precision, 100% recall, 96% F1 score. Outperforms other models across all metrics except recall (where all models achieve 100%)

Other models: All demonstrate 83.33% accuracy, 80% precision, 100% recall, 88.89% F1 score

	Model	Accuracy	Precision	Recall	F1 Score
0	Logistic Regression	83.333333	80.000000	100.0	88.888889
1	SVM	83.333333	80.000000	100.0	88.888889
2	Decision Tree	94.444444	92.307692	100.0	96.000000
3	KNN	83.333333	80.000000	100.0	88.888889

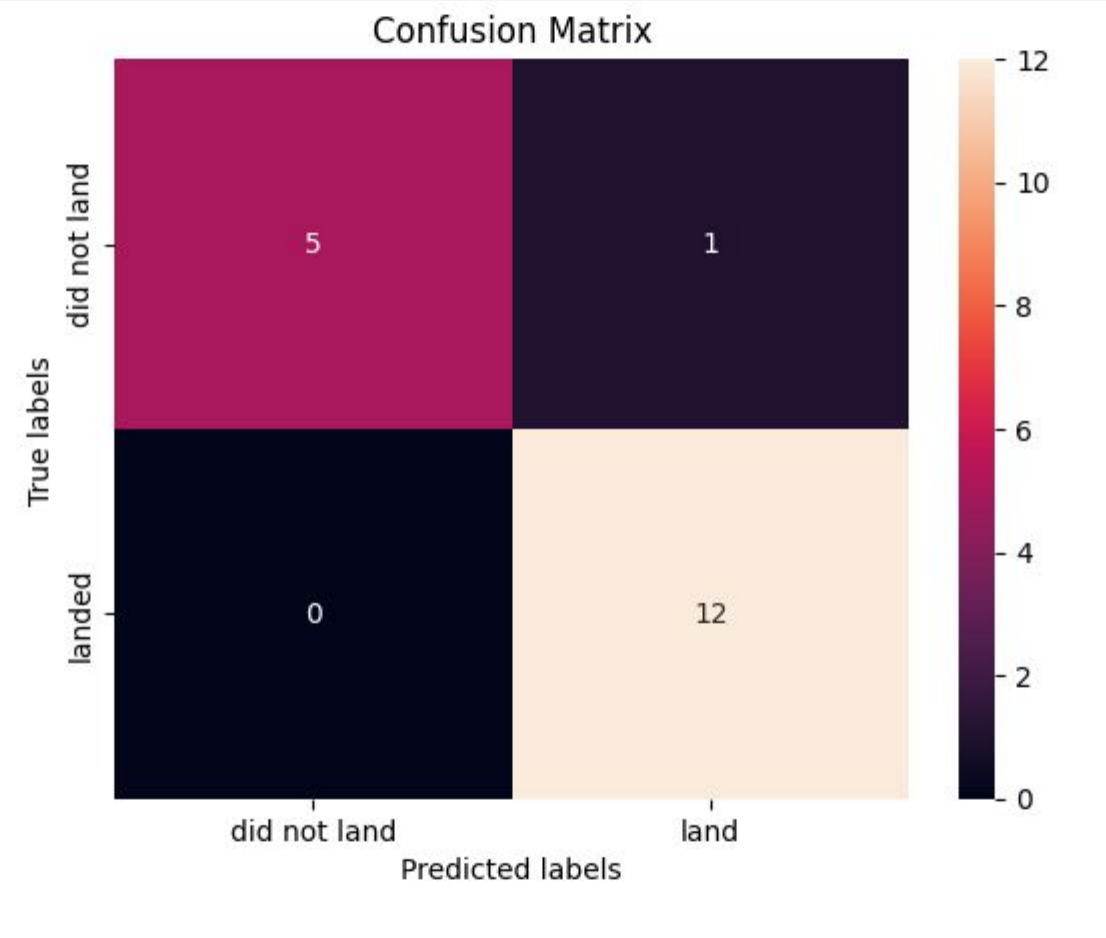


Confusion Matrix

Key Findings:

Reading the matrix (True labels on y-axis, Predicted labels on x-axis):

- **True Positives (bottom right)**: 12 cases where the model correctly predicted a successful landing
- **True Negatives (top left)**: 5 cases where the model correctly predicted an unsuccessful landing
- **False Positives (top right)**: 1 case where the model incorrectly predicted a successful landing when it actually failed
- **False Negatives (bottom left)**: 0 cases where the model incorrectly predicted a failure when it actually landed



Section 6

Conclusions

Key findings and recommendations

1. Success rates improved considerably after the first 30 flights, with consistent success in later missions.

Strategic implication:

Space Y should expect a similar learning curve in its operations
The analysis suggests that approximately 30 launches may be required to achieve consistent landing success
Early business planning should account for higher failure rates and associated costs
Investment in extensive ground testing and simulation could potentially accelerate this learning curve

2. No clear correlation between payload mass and success rate, with even very heavy payloads (15,000+ kg) achieving successful landings in later flights.

Strategic implication:

Space Y need not restrict its business model to lighter payloads
The company can confidently develop heavy-lift capabilities knowing that payload mass alone doesn't determine landing success
Focus should be on technological improvements rather than payload limitations
This opens market opportunities for larger satellite deployments and potential lunar/Mars missions

Key findings and recommendations

3. All sites demonstrate improved success rates over time, with KSC LC-39A having the highest success rate (76.9%).

Strategic implication:

Launch site selection is less critical than operational experience
Space Y should prioritise securing access to KSC LC-39A for critical missions if possible
Alternative sites remain viable with proper operational protocols
Site selection should consider mission-specific requirements (orbit type, payload mass) rather than success probability alone

4. ES-L1, GEO, HEO, and SSO orbits show 100% success rates, while GTO has the lowest rate at approximately 52%.

Strategic implication:

Space Y should price GTO missions with higher margins to account for increased landing failure risk
Early missions should target high-success orbits to build reputation and operational confidence
The company could develop specialised expertise in GTO missions as a market differentiator
R&D investment in GTO-specific landing technologies could provide competitive advantage

Key findings and recommendations

Decision Tree Model Value

The superior performance of the Decision Tree model (94.44% accuracy) over other approaches offers several specific advantages:

1. **Interpretability:** The decision paths can be traced to understand exactly which factors most influence landing success
2. **Feature importance:** The model clearly identifies which variables have the greatest predictive power, guiding R&D prioritisation
3. **High precision:** The 92.31% precision means Space Y can confidently identify when SpaceX will achieve successful landings, enabling accurate cost estimation for competitive bidding
4. **Perfect recall:** With 100% recall, the model never misses successful landing scenarios, ensuring Space Y won't miss opportunities to adjust pricing strategies

Finding: The confusion matrix shows only one false positive, indicating robust classification of landing outcomes based on mission parameters.

Strategic Implication:

- Space Y can reliably estimate SpaceX's costs and pricing strategies
- The identified features driving successful landings should inform technology development priorities
- Understanding the conditions of the single misclassified case provides insight into edge-case scenarios
- The model offers a framework for continuous monitoring of SpaceX's performance

Key findings and recommendations

Recommended Strategic Actions for Space Y

Based on this analysis, Space Y should:

1. **Develop a phased approach:** Plan for a deliberate development curve with expected higher failure rates in early launches
2. **Target orbit strategy:** Initially focus on the highly successful orbit types (ES-L1, GEO, HEO, SSO) while building capabilities
3. **Progressive payload scaling:** Begin with mid-range payloads (3000-5000 kg) where success rates are highest, then expand to heavier payloads
4. **Continuous monitoring:** Implement ongoing analysis of SpaceX launches to update the predictive model and refine competitive strategies
5. **Site selection:** Develop capabilities compatible with multiple launch sites, with preference for KSC LC-39A for critical missions
6. **Pricing strategy:** Incorporate landing probability into mission cost structure, with higher margins for GTO missions and other higher-risk profiles

Conclusion

This analysis provides Space Y with a data-driven foundation for strategic planning in the competitive launch market. The 94.44% accurate Decision Tree model offers reliable predictions of SpaceX's operational capabilities, particularly regarding first stage landing success—the key factor in SpaceX's cost advantage. By understanding the patterns and factors driving landing success, Space Y can develop targeted technologies, optimal operational procedures, and competitive pricing strategies to effectively compete in the commercial space market.

Thank you!