

Winning Space Race with Data Science

Jovanna Garza
April 1st 2025



Outline

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



Executive Summary

- Summary of methodologies
 - Collected data via SpaceX API and web scraping
 - Cleaned and wrangled data; created binary landing outcome variable
 - Performed EDA using SQL, visualizations, and geospatial mapping (Folium)
 - Built interactive dashboard with Plotly Dash
 - Trained and evaluated classification models: Logistic Regression, SVM, Decision Tree, KNN
- Summary of all results
 - Launch success increased over time
 - KSC LC-39A had the highest success rate
 - Heavy payloads and orbits like GEO, ES-L1, and SSO had 100% success
 - All models achieved ~83% accuracy
 - Decision Tree slightly outperformed other models

Introduction

- **Background**

- The rise of commercial spaceflight has made launches more frequent and accessible.
- SpaceX leads the industry by reducing costs through the reuse of the Falcon 9 first stage, bringing launch prices down to ~\$62 million.
- The ability to **predict first stage landing success** is essential for cost estimation and planning.
- This project simulates working as a data scientist at a fictional company, **SpaceY**, aiming to compete with SpaceX.
- Using public data and machine learning, we analyze and predict whether the **Falcon 9's first stage** will land successfully.

- **Problem Statement:**

- Can we predict whether the first stage of a SpaceX Falcon 9 rocket will land successfully using publicly available launch data?



Section 1

Methodology

Methodology



Executive Summary

- Data Collection
 - Retrieved Falcon 9 launch data via SpaceX API and Wikipedia web scraping
- Data Wrangling & Processing
 - Cleaned data, handled missing values, applied one-hot encoding, and created binary landing outcome
- Exploratory Data Analysis (EDA)
 - Explored trends using SQL and visualizations (payload, orbit, launch site, flight number)
- Interactive Visual Analytics
 - Visualized launch sites with Folium and built an interactive dashboard using Plotly Dash
- Predictive Modeling
 - Trained classification models (Logistic Regression, SVM, Decision Tree, KNN)
 - Decision Tree slightly outperformed other models

Data Collection

- How data sets were collected:
 - Queried SpaceX API to obtain structured launch data
 - Applied `.json_normalize()` to flatten nested data
 - Scrapped Wikipedia to supplement launch outcomes and booster details
 - Combined both sources for a more complete dataset

Data Collection

Data Sources & Flow



Request JSON

Parse with
.json()

Flatten with
.json_normalize()

Convert to
DataFrame

Filter for Falcon
9

Request HTML

Parse with
BeautifulSoup

Extract Falcon
9 Table

Convert to Dictionary
→ DataFrame

Export to CSV

Combined both CSVs
Cleaned & merged for modeling

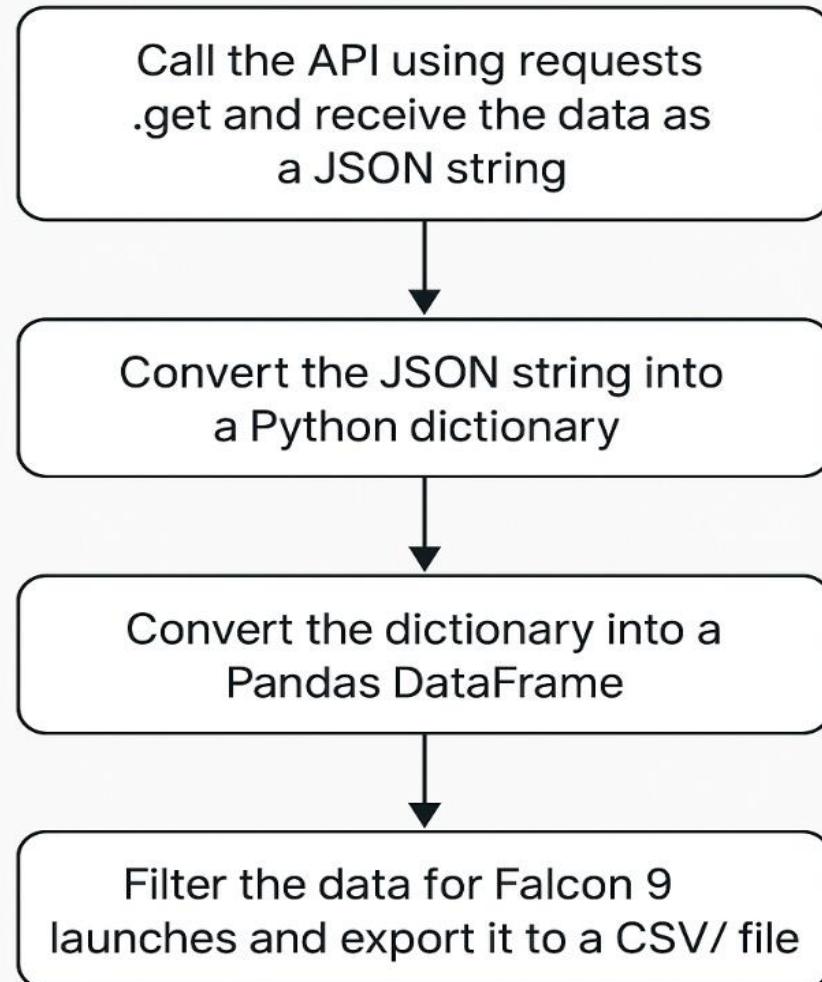
Data Collection – SpaceX API

- Requested data using `requests.get()` from: <https://api.spacexdata.com/v4/launches>
- Parsed JSON data with `.json()`
- Normalized nested fields using `pandas.json_normalize()`
- Converted data into a Pandas DataFrame
- Filtered for Falcon 9 launches only
- Exported cleaned data to CSV for further processing

SpaceX API calls notebook GitHub

- 1) <https://github.com/jovannagarza/Spacex-Applied-Data-Science-Capstone/blob/main/1%20API%20Data%20Collection.ipynb>

Flowchart of SpaceX API Calls



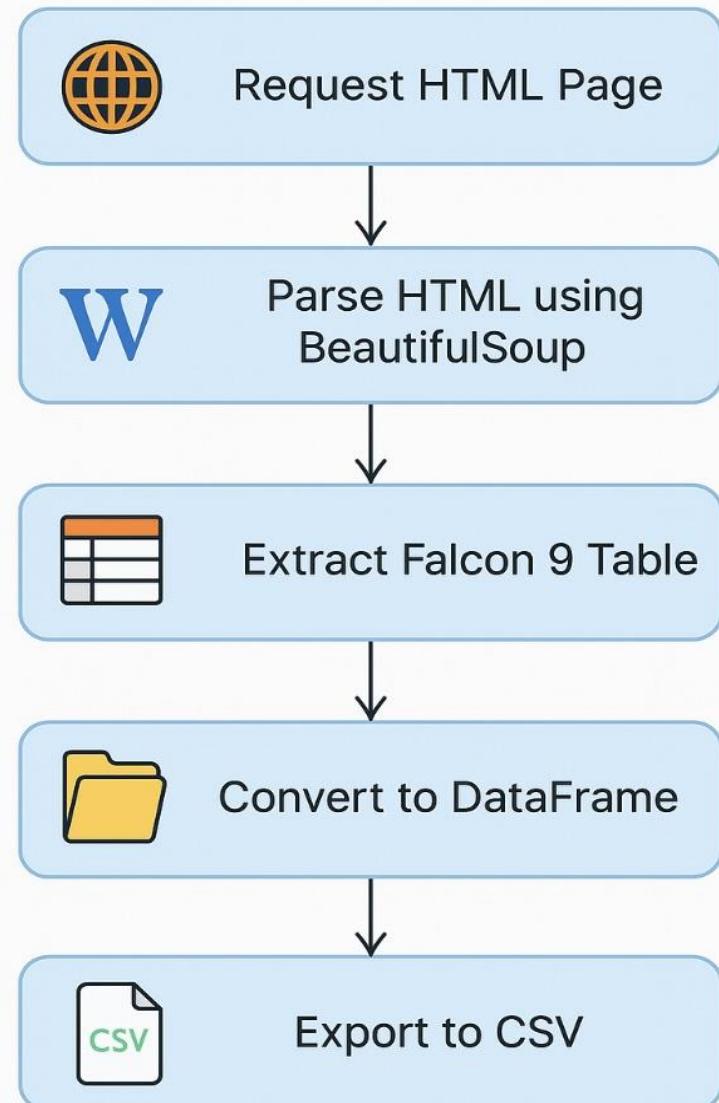
Data Collection - Scraping

- Retrieved Falcon 9 launch data directly from **Wikipedia**
- Fetched the **HTML content** of the page using Python
- Used **BeautifulSoup** to parse the HTML structure
- Located the **launch table** containing Falcon 9 missions
- Extracted table data and organized it into a **dictionary**
- Converted the dictionary into a **Pandas DataFrame**
- Exported the cleaned data to a **CSV file** for further use
- Output: Well-structured Falcon 9 dataset ready to be analyzed

Web scraping notebook GitHub

- 2) <https://github.com/jovannagarza/Spacex-Applied-Data-Science-Capstone/blob/main/2%20Web%20Scraping.ipynb>

Data Collection – Web Scraping



Data Wrangling

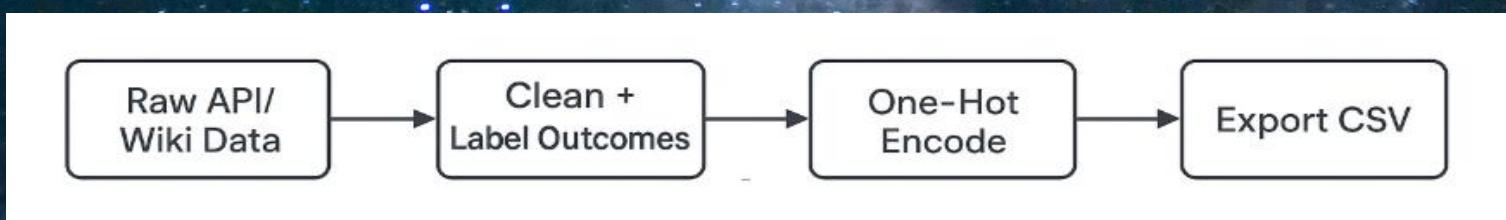
- Filtered DataFrame to include only Falcon 9 launches
- Handled missing values (e.g., filled in payload mass with calculated mean)
- Created binary outcome column based on landing success:
 - 1 = Successful landing (ASDS or RTLS)
 - 0 = Unsuccessful landing or no attempt
- Mapped outcome types (e.g., True ASDS, False Ocean) to binary class
- Performed feature engineering to prepare for ML models
- Applied one-hot encoding to categorical columns (launch site, orbit, etc.)
- Cleaned and exported final dataset to CSV for analysis and modeling

Data Wrangling notebooks

- 3) [https://github.com/jovannagarza/Spacex-Applied-Data-Science-Capstone/blob/main/3\)%20Data%20Wrangling.ipynb](https://github.com/jovannagarza/Spacex-Applied-Data-Science-Capstone/blob/main/3)%20Data%20Wrangling.ipynb)



ThePhoto by PhotoAuthor is licensed under CC BY-SA.



EDA with SQL

- Queried unique launch sites used across all missions
- Retrieved total payload mass carried by boosters launched by NASA (CRS)
- Calculated the average payload mass for booster version F9 v1.1
- Identified the first successful landing on a ground pad
- Listed boosters that successfully landed on drone ships with payloads between 4,000–6,000 kg
- Counted total number of successful vs. failed missions
- Found booster versions that carried the maximum payload
- Retrieved failed drone ship landings in 2015, including launch site and booster version
- Ranked landing outcomes by frequency between 2010–2017

EDA with SQL notebook GitHub

- 4) <https://github.com/jovannagarza/Spacex-Applied-Data-Science-Capstone/blob/main/4%20EDA%20SQL.ipynb>



The Photo by PhotoAuthor is licensed under CC BY SA.

EDA with Data Visualization

- Flight Number vs. Launch Site
 - Used a scatter plot to identify launch success trends across different sites over time
- Payload Mass (kg) vs. Launch Site
 - Plotted to examine how payload size varies by site and how it relates to landing outcomes
- Payload Mass (kg) vs. Orbit Type
 - Used a scatter plot to explore relationships between orbit and payload success
- Success Rate by Orbit
 - Displayed with a bar chart to highlight which orbits had higher success rates
- Launch Success Over Time
 - A line chart visualizing how success rates have improved with more launches
- Charts were chosen to:
 - ✓ Compare continuous vs. categorical variables
 - ✓ Reveal patterns and relationships relevant for predictive modeling
 - ✓ Help visualize improvements over time

EDA with Data Visualization notebook GitHub

- 5) <https://github.com/jovannagarza/Spacex-Applied-Data-Science-Capstone/blob/main/5%20EDA%20Data%20Visualization.ipynb>



Build an Interactive Map with Folium

- Added red circle markers at all launch site coordinates
→ To clearly visualize where launches took place
- Labeled each marker with the launch site name using popup labels
→ Helps identify launch locations at a glance
- Plotted green and red markers for successful and failed landings
→ Green = success, Red = failure — shows performance by site
- Drew colored lines from CCAFS SLC-40 to nearby proximity points:
→ City, coastline, highway, railway
→ Demonstrates how launch sites are strategically located for safety and access
- These visual elements made it easy to:
 - ✓ Compare launch site performance
 - ✓ Understand geographic advantages of each location
 - ✓ Support site analysis for future launch planning

Interactive map with Folium map GitHub

- + 6) <https://github.com/jovannagarza/Spacex-Applied-Data-Science-Capstone/blob/main/6%20Interactive%20Visual%20Analytics%20with%20Folium.ipynb>



Build a Dashboard with Plotly Dash

- Added a dropdown menu to select either all launch sites or an individual site
 - Allows users to filter data by site and view launch-specific insights
- Included a payload mass slider
 - Enables dynamic filtering based on payload weight for custom analysis
- Pie chart showing launch outcomes (success vs. failure)
 - Helps users visualize overall success rate or by selected site
- Scatter plot: Payload Mass vs. Launch Success
 - Displays correlation between payload size and landing success across booster versions
- These elements were added to:
 - ✓ Make the dashboard interactive and user-friendly
 - ✓ Allow quick exploration of launch performance
 - ✓ Highlight trends that may influence landing outcomes

Plotly Dash lab GitHub

- 7) <https://github.com/jovanngarza/Spacex-Applied-Data-Science-Capstone/blob/main/7%20Interactive%20Visual%20Analytics%20Plotly.py>



ThePhoto by PhotoAuthor is licensed under CC BY SA.



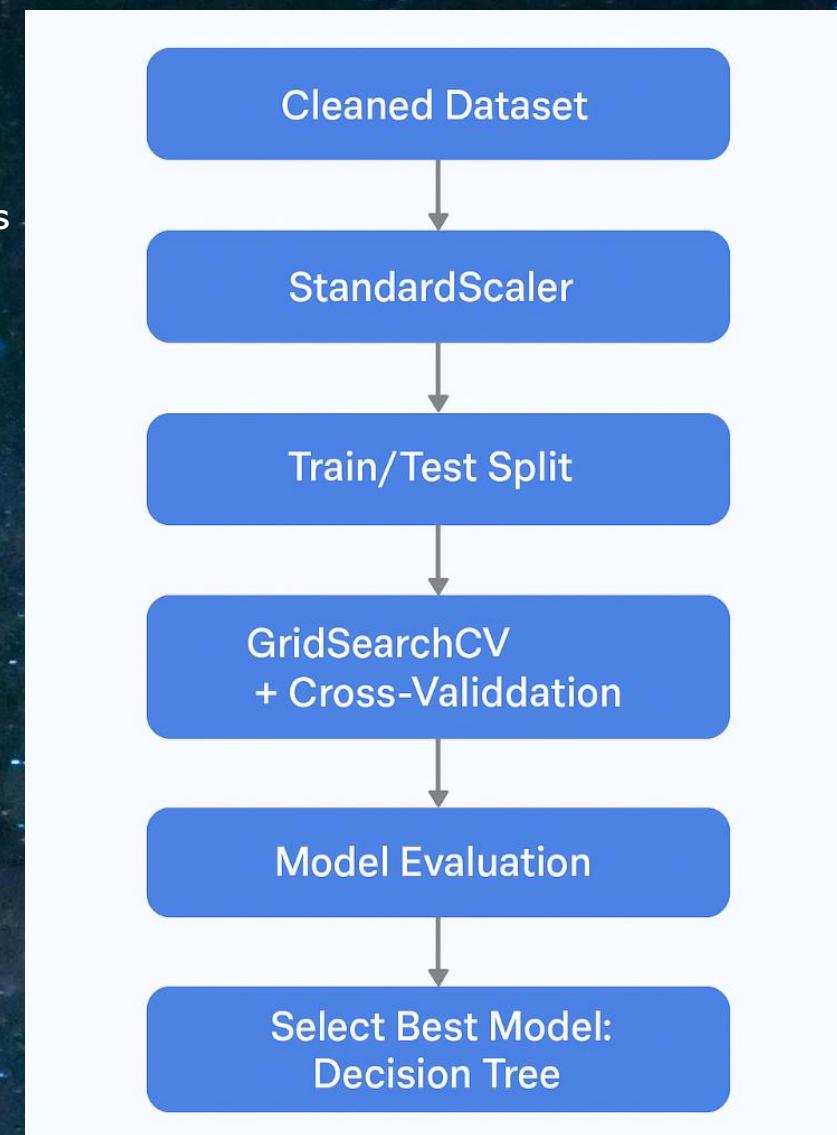
ThePhoto by PhotoAuthor is licensed under CC BY SA.

Predictive Analysis (Classification)

- Selected classification models:
 - Logistic Regression, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors (KNN)
- Preprocessed data:
 - Standardized features using StandardScaler
 - Split dataset into training and test sets
- Model tuning:
 - Used GridSearchCV with cross-validation to optimize hyperparameters
- Model evaluation:
 - Compared models using accuracy, precision, recall, F1-score, and confusion matrix
 - All models had similar test accuracy (~83%), but Decision Tree had the highest GridSearchCV score
- Best model identified:
 - Decision Tree performed best overall and was selected for final predictions

Predictive Analysis GitHub

- 8) <https://github.com/jovannagarza/Spacex-Applied-Data-Science-Capstone/blob/main/8%20Machine%20Learning%20Prediction%20lab.ipynb>

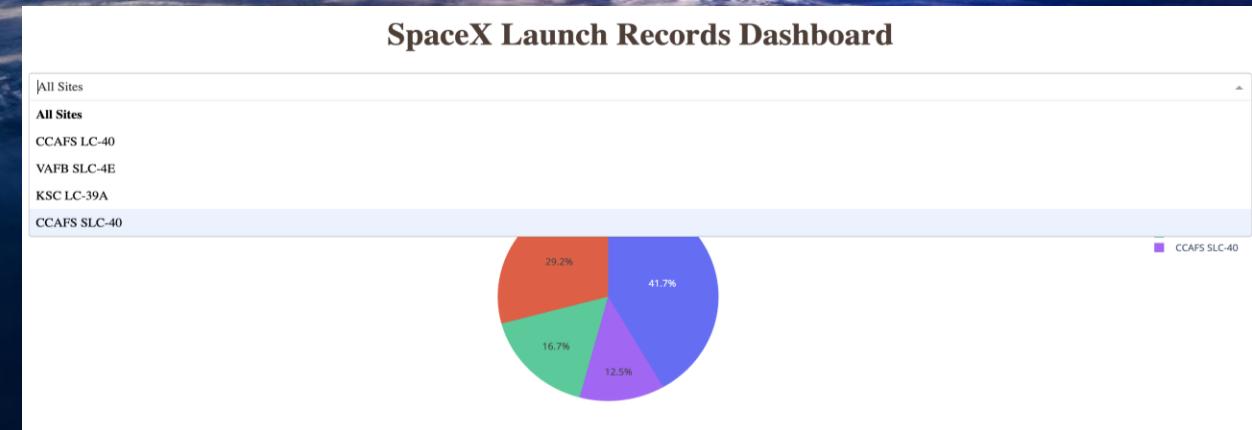


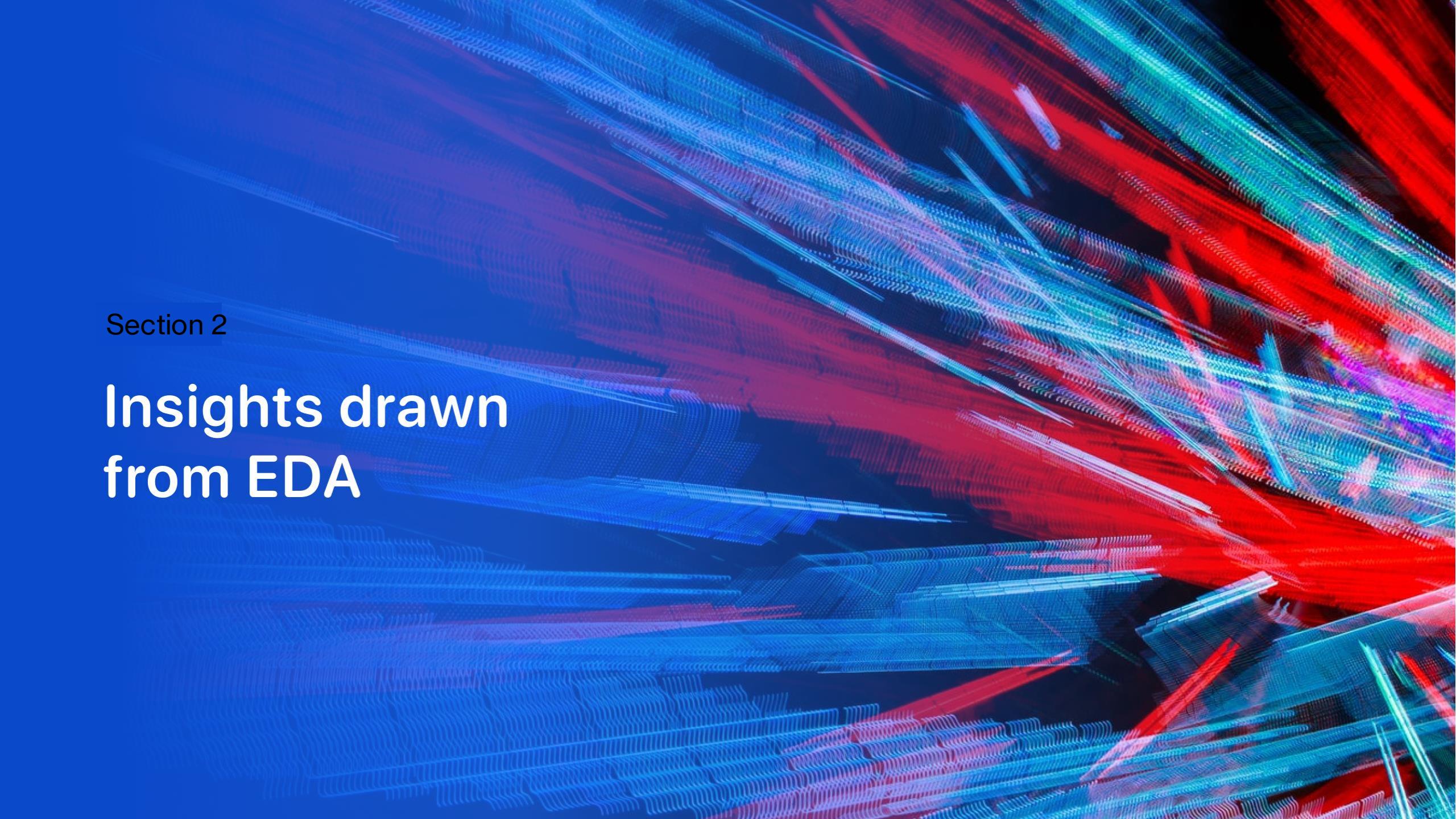
Results

- Exploratory Data Analysis (EDA)
 - Launch success increased over time, especially after early missions
 - KSC LC-39A had the highest success rate of all launch sites
 - Certain orbits (ES-L1, GEO, HEO, SSO) had 100% landing success
 - Heavier payloads generally had a higher success rate, especially for LEO and ISS orbits
- Predictive Analysis
 - All models achieved ~83% test accuracy
 - Decision Tree slightly outperformed others with the best GridSearchCV score
 - Final metrics:
 - Precision: 0.80
 - Recall: 1.0
 - F1 Score: 0.89
 - Confusion Matrix: 12 TP, 3 TN, 3 FP, 0 FN

```
7]: # Select relevant sub-columns: 'Launch Site', 'Lat(Latitude)', 'Long(Longitude)', 'class'
spacex_df = spacex_df[['Launch Site', 'Lat', 'Long', 'class']]
launch_sites_df = spacex_df.groupby(['Launch Site'], as_index=False).first()
launch_sites_df = launch_sites_df[['Launch Site', 'Lat', 'Long']]
launch_sites_df
```

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610745



The background of the slide features a complex, abstract digital visualization. It consists of numerous thin, glowing lines that create a sense of depth and motion. The lines are primarily blue and red, with some green and purple highlights. They form a grid-like structure that curves and twists across the frame, resembling a 3D wireframe or a network of data points. The overall effect is futuristic and dynamic, suggesting concepts like data flow, digital communication, or complex systems.

Section 2

Insights drawn from EDA

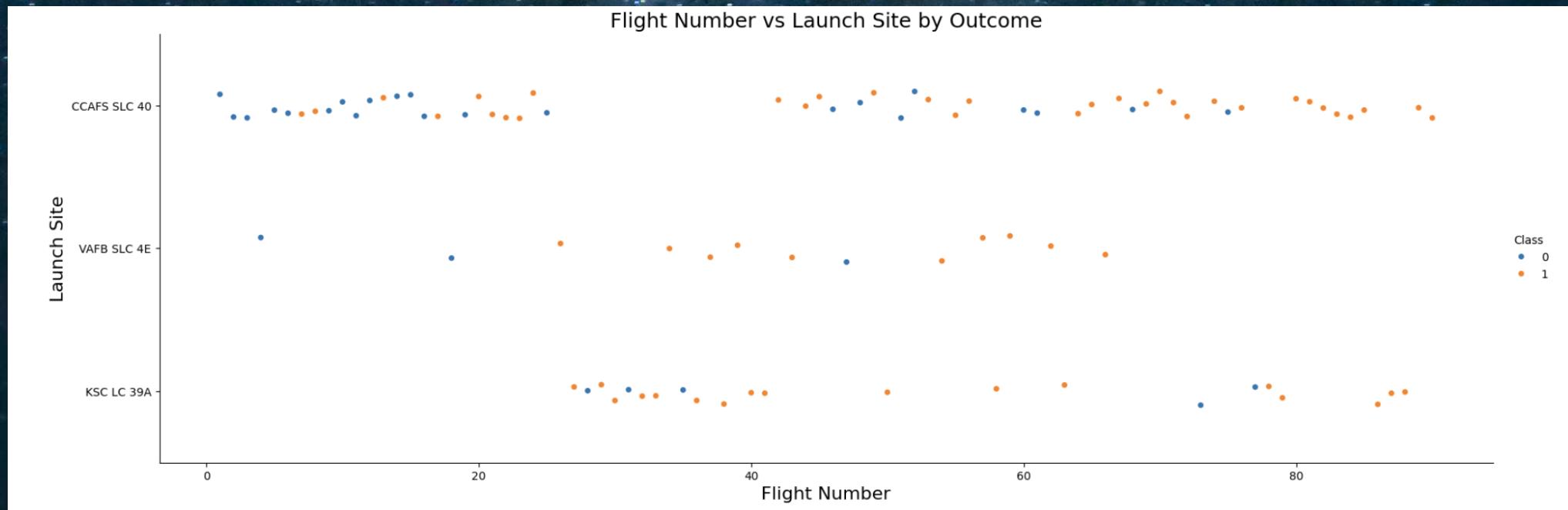
Flight Number vs. Launch Site

- **Explanation:**

- This scatter plot visualizes how **launch success evolves over time** across different launch sites.
- **Orange** = Successful landings
- **Blue** = Failed landings

- **Insights:**

- **Earlier flights** had a **lower success rate**, especially at CCAFS SLC-40
- Sites like **KSC LC-39A** and **VAFB SLC-4E** show higher success rates as time progressed



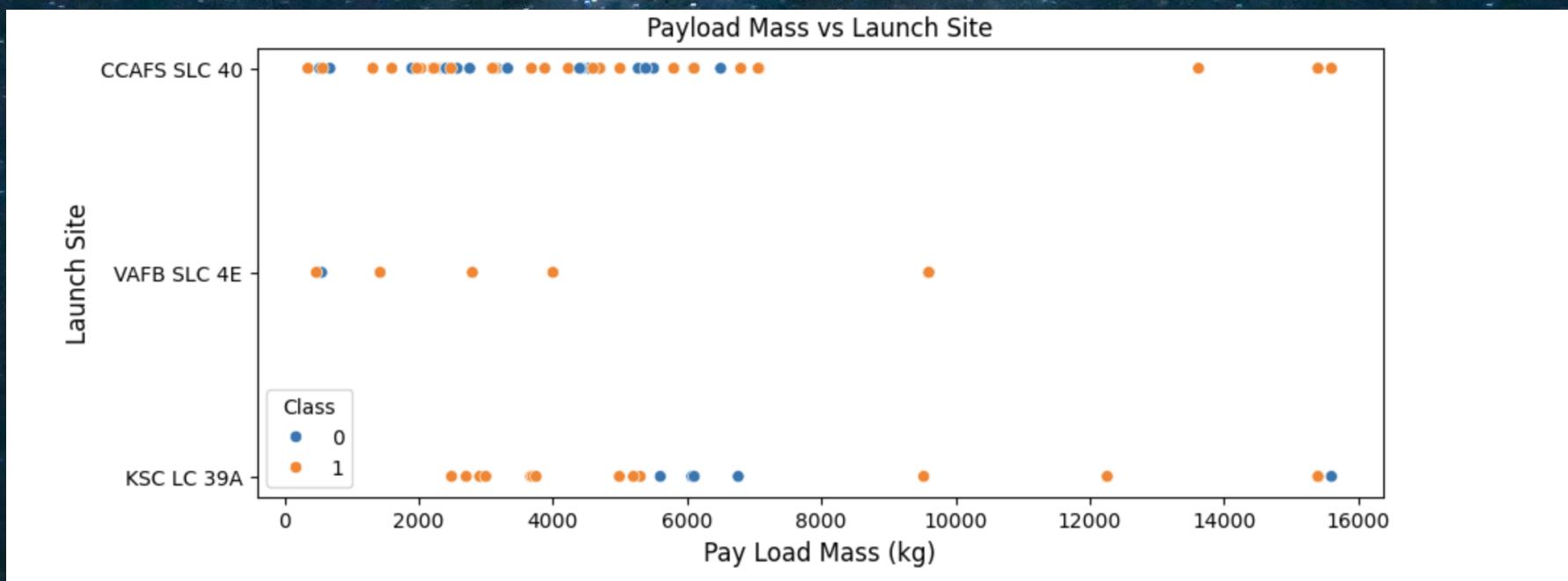
Payload vs. Launch Site

- **Explanation**

- This scatter plot shows how **payload mass (kg)** varies across different **launch sites**, and how that relates to **landing outcomes**.
- **Orange** = Successful landings
- **Blue** = Failed landings

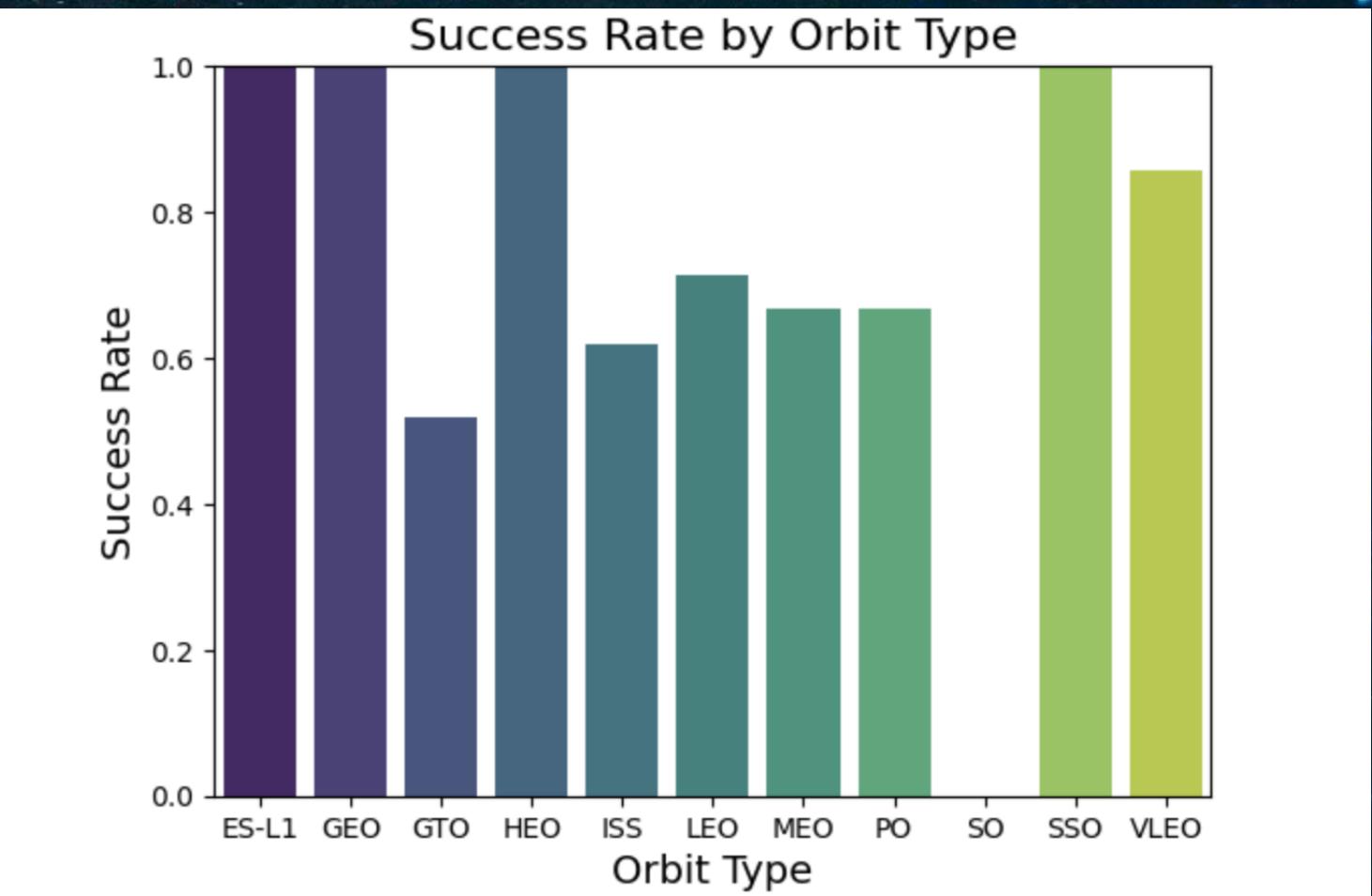
- **Insights:**

- **KSC LC-39A** shows **100% success** for payloads under **5,500 kg**
- **CCAFS SLC-40** had a wider range of payloads but a **lower success rate**
- **VAFB SLC-4E** did not launch payloads greater than ~10,000 kg
- Launch site and payload capacity both impact **first-stage landing success**



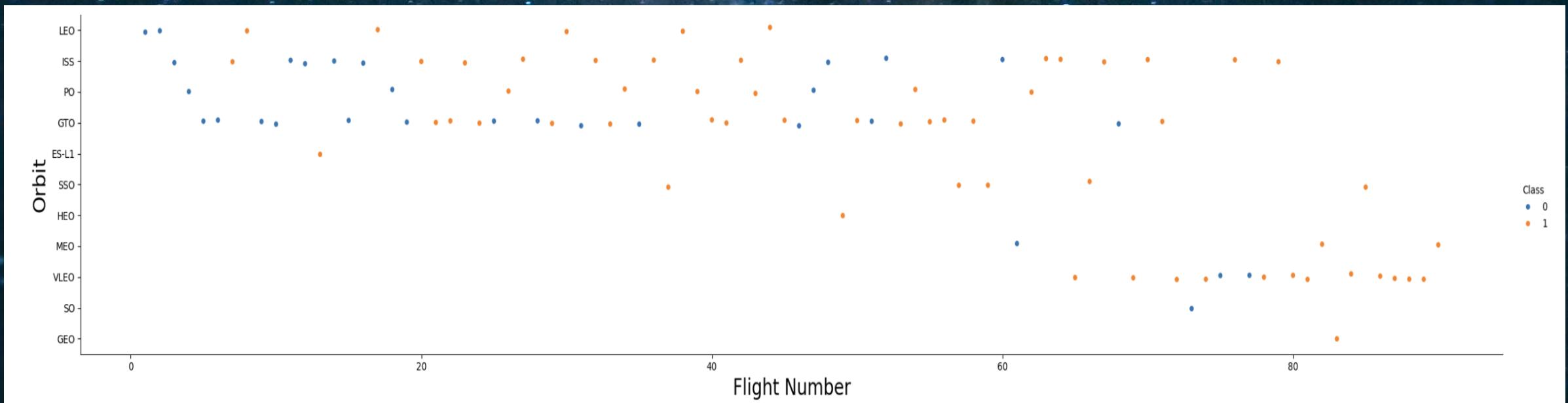
Success Rate vs. Orbit Type

- Insights:
 - + 100% Success:
 - ES-L1, GEO, HEO, SSO — every mission targeting these orbits landed successfully
 - + Moderate Success (50–80%):
 - GTO, ISS, LEO — common but more challenging orbits
 - + 0% Success:
 - SO — missions targeting this orbit failed to land
- Orbit type has a clear impact on the likelihood of first stage recovery, likely due to distance, trajectory, and mission parameters



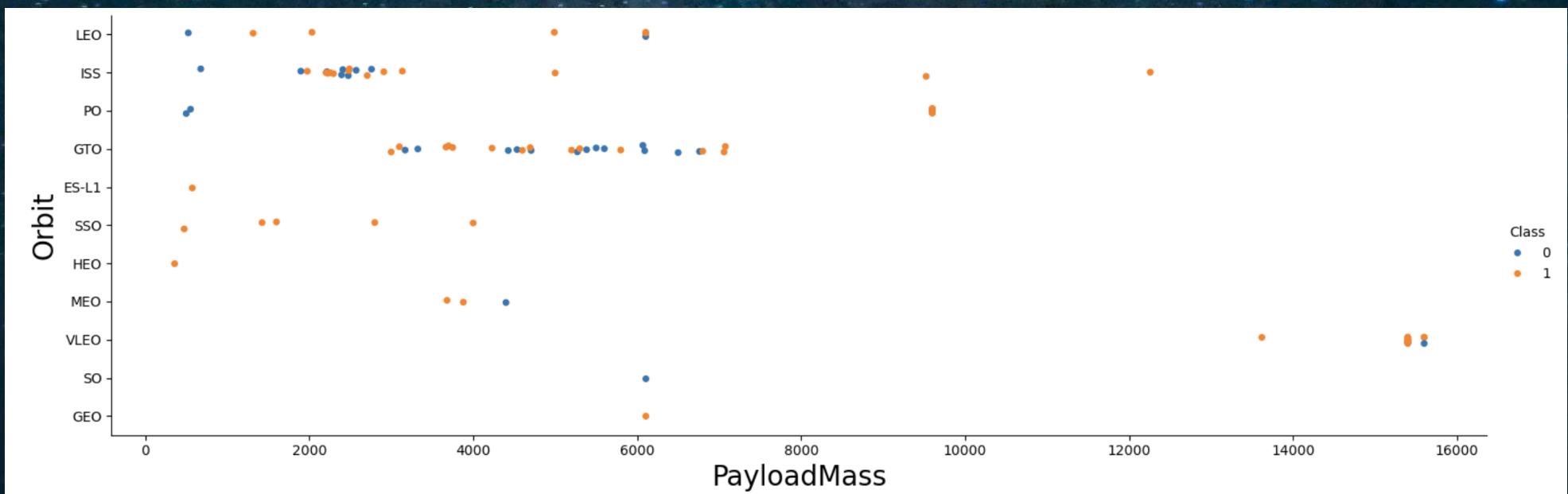
Flight Number vs. Orbit Type

- This scatter plot visualizes the **relationship between flight number (launch order) and the orbit type** for each mission.
- Later flights generally show **more successful landings**, especially in **LEO and ISS orbits**
- **GTO orbit shows mixed results**, with both successes and failures across all launch stages
- This visualization demonstrates how **experience (higher flight number)** correlates with **improved mission outcomes** in certain orbits



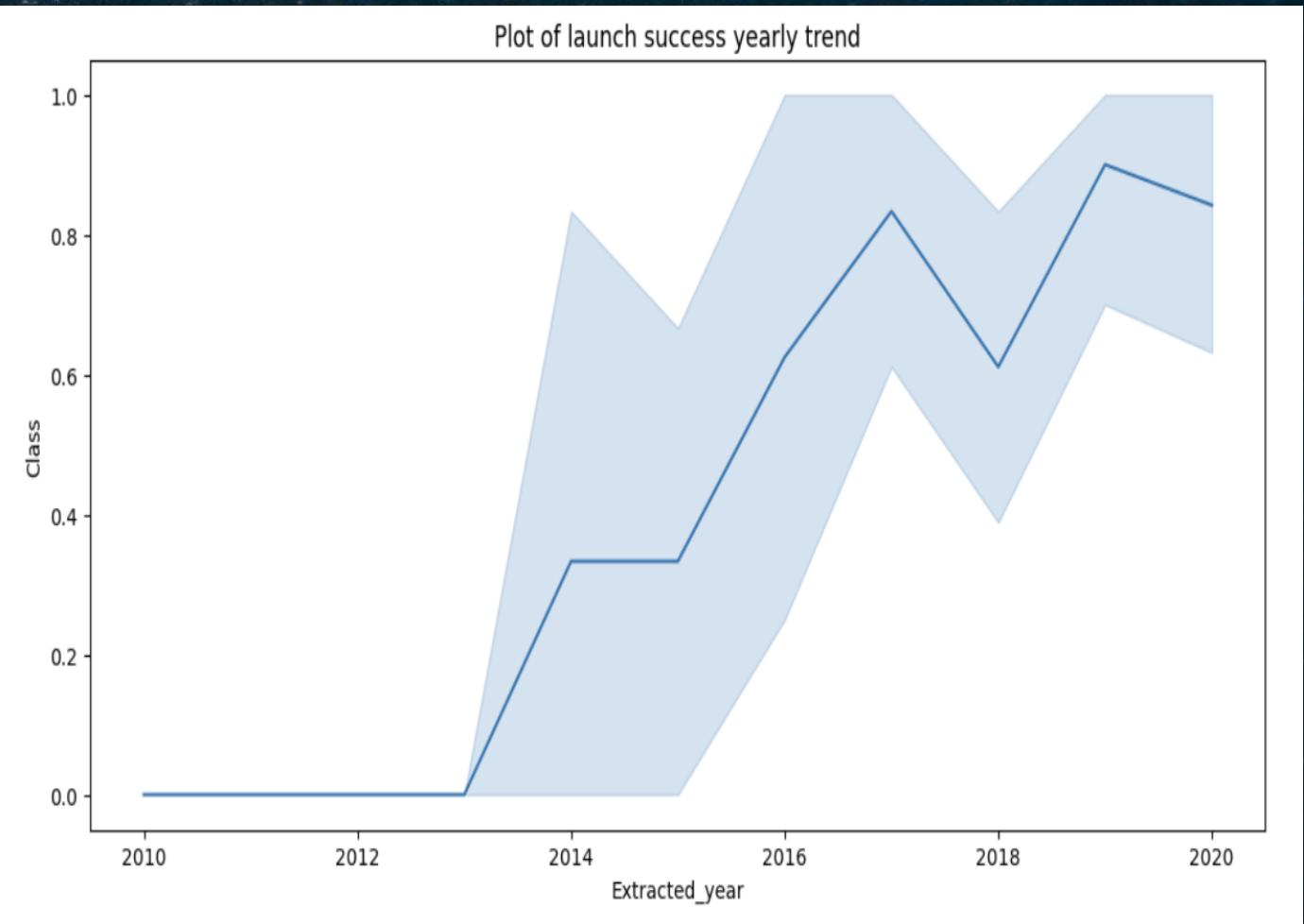
Payload vs. Orbit Type

- This scatter plot shows the relationship between payload mass (kg) and orbit type, and how they relate to landing success.
- LEO, ISS, and PO orbits typically carry heavier payloads and tend to have higher success rates
- GTO orbit shows a mix of successes and failures with heavier payloads
- High success at lower payloads for orbits like SSO
- This highlights how orbit choice and payload size can impact mission planning and landing success



Launch Success Yearly Trend

- This line chart shows the **average success rate** of Falcon 9 first stage landings by **year**.
- **Significant improvements** from 2013 to 2017
- Minor drop between **2017–2018** and again in **2019–2020**
- Overall, launch success has **steadily increased** as SpaceX gained more experience
- Reflects the **evolution of technology and landing reliability** over time



All Launch Site Names

- These represent the four primary locations where Falcon 9 rockets have been launched. Identifying the unique sites is essential for analyzing site-specific success rates and for creating interactive filtering options in visual dashboards.

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

[11]:

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

* sqlite:///my_data1.db
Done.

[11]:

Launch_Site

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- This query filters launch site names that start with 'CCA', returning sites located at the Cape Canaveral Air Force Station (CCAFS) in Florida. These locations are key launch sites for SpaceX and represent a significant portion of the Falcon 9 missions in the dataset.

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

```
[42]: %sql SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

* 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

- The query calculates the **total payload mass** for missions where NASA was listed as the customer
- NASA-related missions carried a **combined payload of 45,596 kg**, reflecting SpaceX's key partnership with government space efforts

Task 3

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

```
: %sql SELECT SUM("PAYLOAD_MASS__KG_") AS "Total Payload Mass (kg)" FROM SPACEXTBL WHERE "Customer" = 'NASA (CRS)';

* sqlite:///my_data1.db
Done.

: Total Payload Mass (kg)
: 45596
```

Average Payload Mass by F9 v1.1

- This query calculates the average payload mass carried by the F9 v1.1 booster version
- Results show that F9 v1.1 missions carried an average of 2,928 kg, helping us compare performance across booster types

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

```
: %sql SELECT AVG(PAYLOAD_MASS__KG_) AS "Average payload mass by Booster Version F9 v1.1" FROM SPACEXTBL WHERE BOOSTER_VERSION =  
* sqlite:///my_data1.db  
Done.  
: Average payload mass by Booster Version F9 v1.1  
-----  
2928.4
```

First Successful Ground Landing Date

- This query finds the earliest date of a successful landing on a ground pad (RTLS = Return to Launch Site)
- The result shows that the first ground landing occurred on December 22, 2015, marking a major milestone in SpaceX's reusability efforts

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

Hint: Use min function

```
: %sql SELECT MIN("Date") AS "First Successful Ground Pad Landing" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)'
```

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

```
Done.
```

```
: First Successful Ground Pad Landing
```

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- This query returns booster versions that had successful drone ship landings with medium-range payloads
- These boosters demonstrated reliability under moderate payload conditions, contributing to reusable launch success

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" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS_KG_"  
* sqlite:///my_data1.db  
Done.  
] : Booster_Version  
-----  
F9 FT B1022  
F9 FT B1026  
F9 FT B1021.2  
F9 FT B1031.2
```

Total Number of Successful and Failure Mission Outcomes

- This query counts the number of missions with **successful (1)** and **failed (0)** first-stage landings
- The result shows **98 successful landings** and **26 failures**, highlighting SpaceX's overall improvement in reliability

List the total number of successful and failure mission outcomes

```
%sql SELECT TRIM("Mission_Outcome") AS Cleaned_Outcome, COUNT(*) AS Outcome_Cou
```

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

```
Done.
```

Cleaned_Outcome	Outcome_Count
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- This query identifies booster versions that carried the heaviest payload mass in the dataset
- These F9 B5 series boosters show SpaceX's ability to support heavy-lift missions with reusable hardware

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS_KG_" = ( SELECT
```

```
* 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

- This query lists failed drone ship landings that occurred in 2015, including the booster version and launch site
- These early failures helped SpaceX refine drone ship landing technology, leading to future success

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql SELECT strftime('%m', "Date") AS Month, "Booster_Version", "Launch_Site", "Landing_Outcome" FROM SPACEXTBL WHERE "Land
```

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

```
Done.
```

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

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

- This query ranks different landing outcomes by their frequency during early Falcon 9 missions
- The most frequent outcome was Failure on drone ships, showing the learning curve before SpaceX perfected reusability

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

```
%sql SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
```

```
* 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

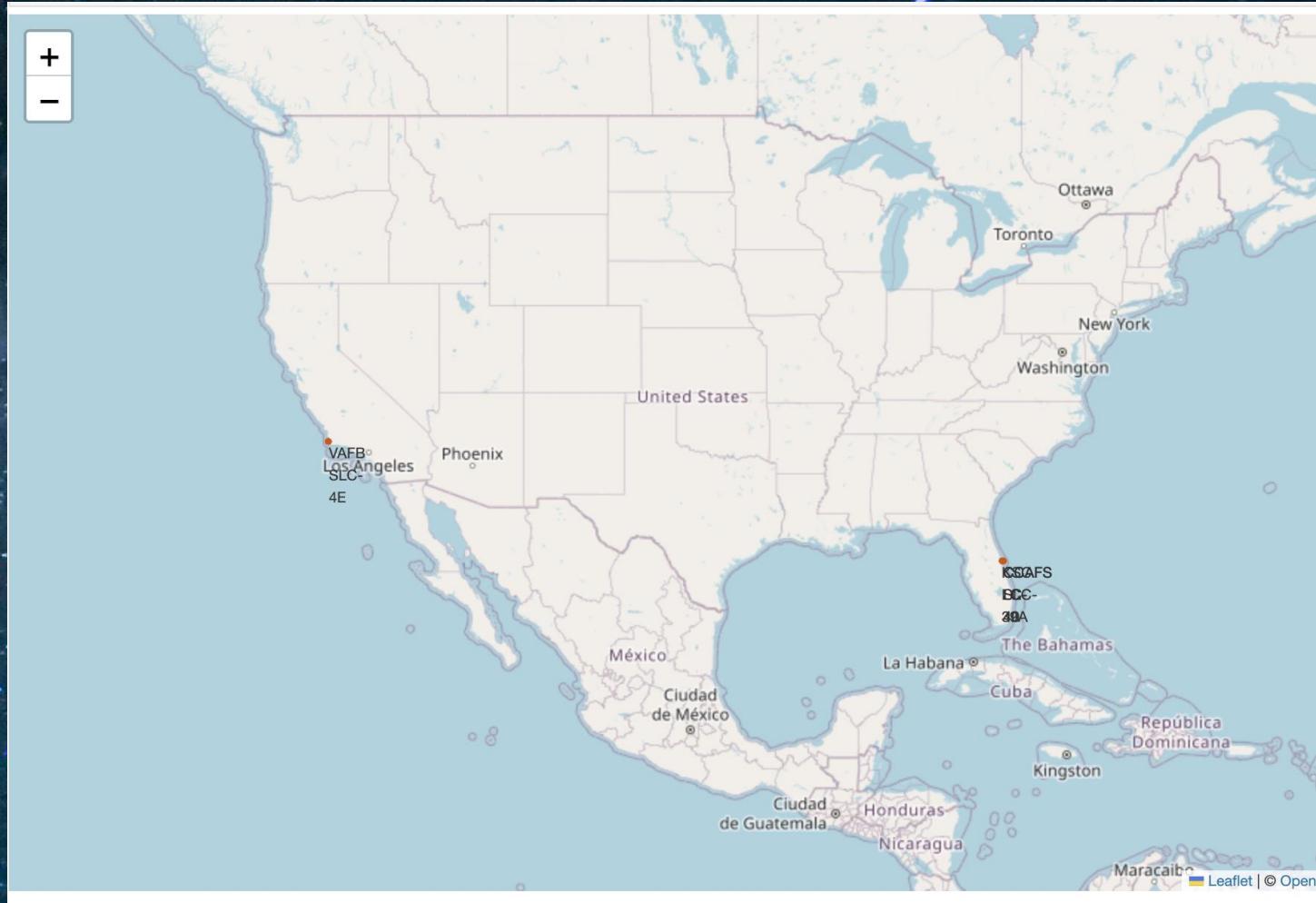
The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth's horizon against a dark blue sky. City lights are visible as numerous small white and yellow dots, primarily concentrated in the lower right quadrant where the United States appears. In the upper right, there are bright green and yellow bands of light, likely the Aurora Borealis or Australis. The overall atmosphere is dark and mysterious.

Section 3

Launch Sites Proximities Analysis

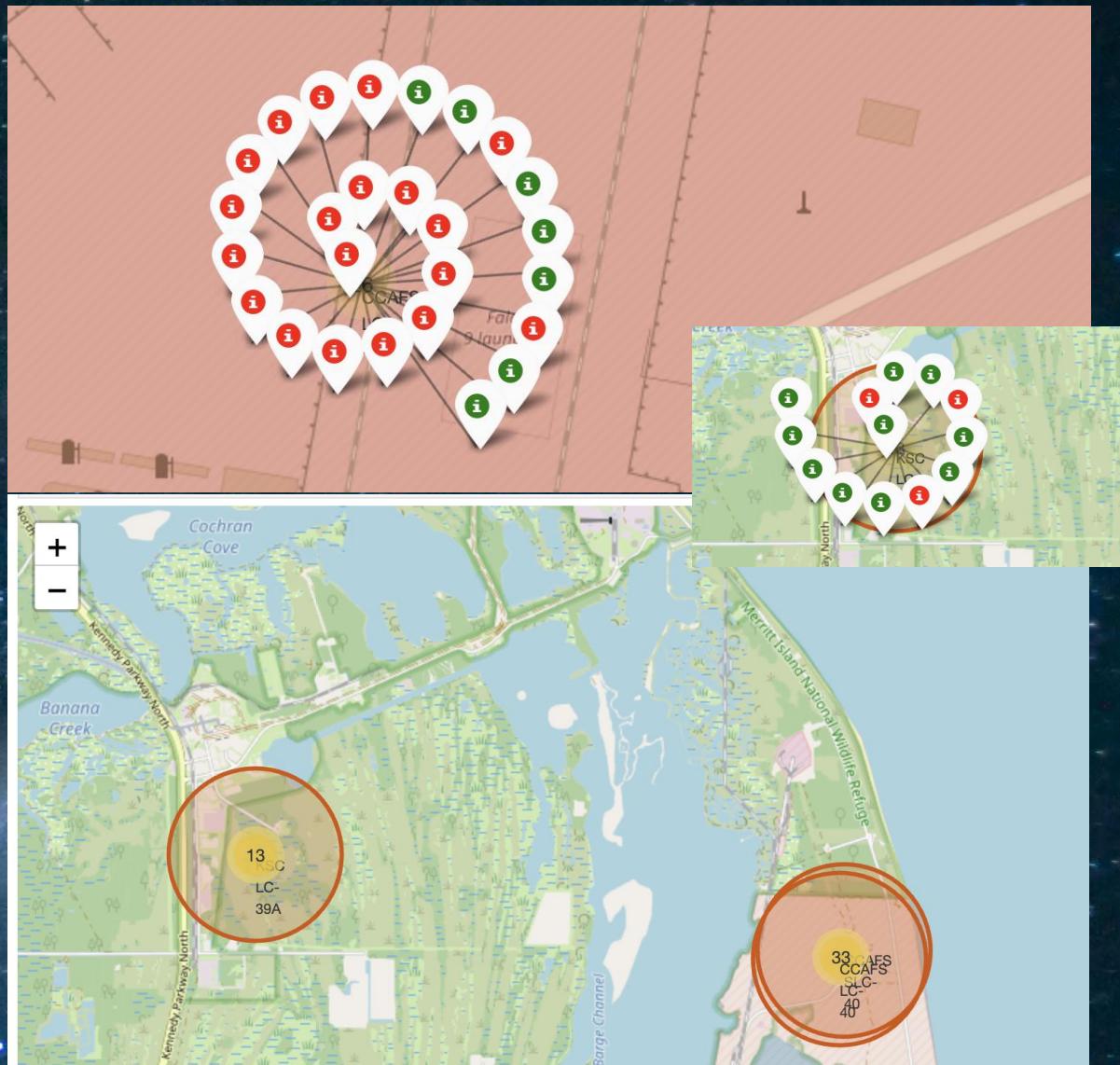
Launch Site Locations (Global Folium Map)

- This interactive Folium map displays all Falcon 9 launch sites, marked with red circular markers
- All launch sites are located near coastlines to allow safe stage recovery in case of failure
- Most launch sites are in the U.S., clustered around Cape Canaveral, Florida, and Vandenberg AFB, California
- Their proximity to water and the equator offers fuel efficiency and safety advantages



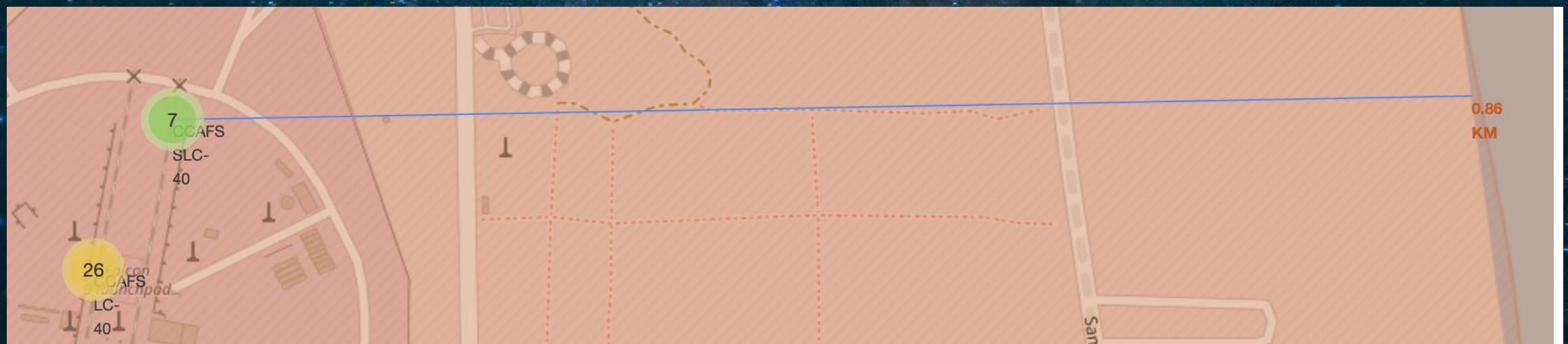
Launch Outcomes by Site (Success vs. Failure)

- This Folium map visualizes launch outcome markers:
- Green = Successful landings
- Red = Failed landings
- Markers are plotted at each launch site location
- Helps identify which sites had higher success rates
- KSC LC-39A shows mostly green markers, indicating a strong success history
- CCAFS SLC-40 includes both green and red, reflecting earlier development stages
- This view highlights how success rates vary by site, supporting analysis for site reliability and mission planning



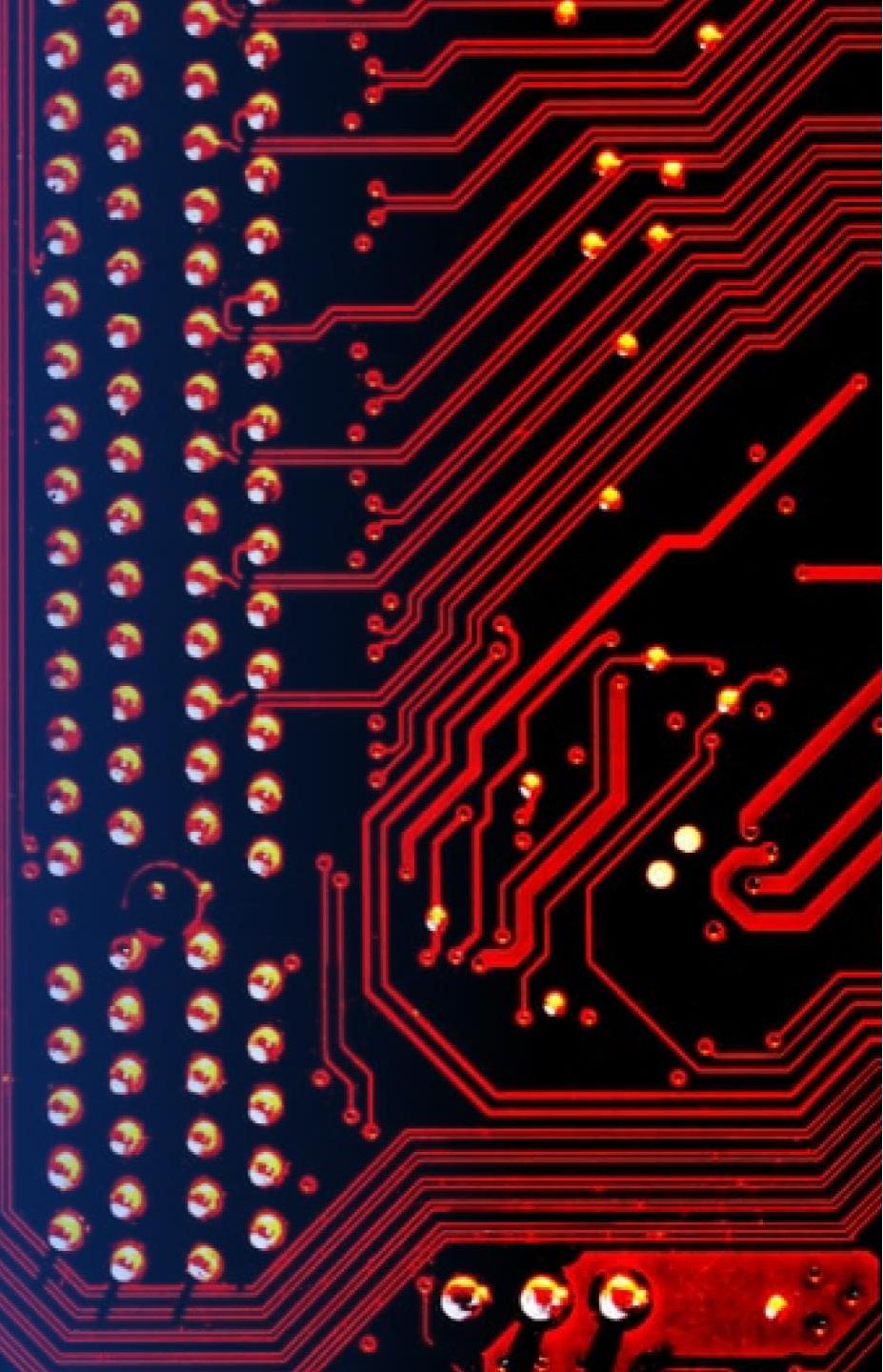
CCAFS SLC-40 Proximity to Key Infrastructure

- This map shows the **distances from CCAFS SLC-40 to important nearby infrastructure:**
 - Coastline: ~0.86 km
 - Railway: ~21.96 km
 - Highway: ~26.88 km
 - City: ~23.23 km
- The site is **close to the coast**, minimizing risk during booster separation
- **Distant from urban areas**, ensuring safety in case of failure
- Still **accessible** via road and rail for logistics and support



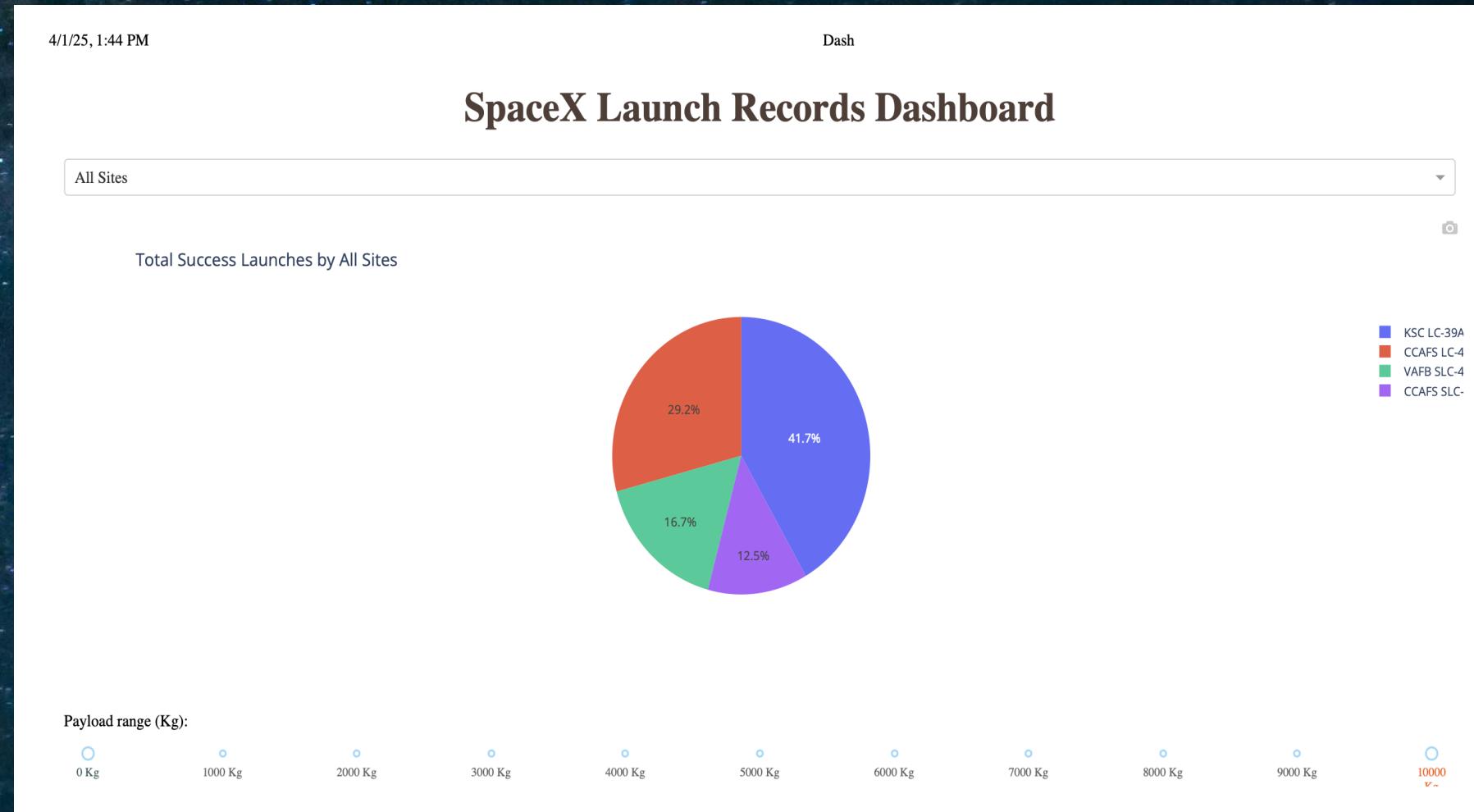
Section 4

Build a Dashboard with Plotly Dash



Launch Success Distribution by Site (Pie Chart)

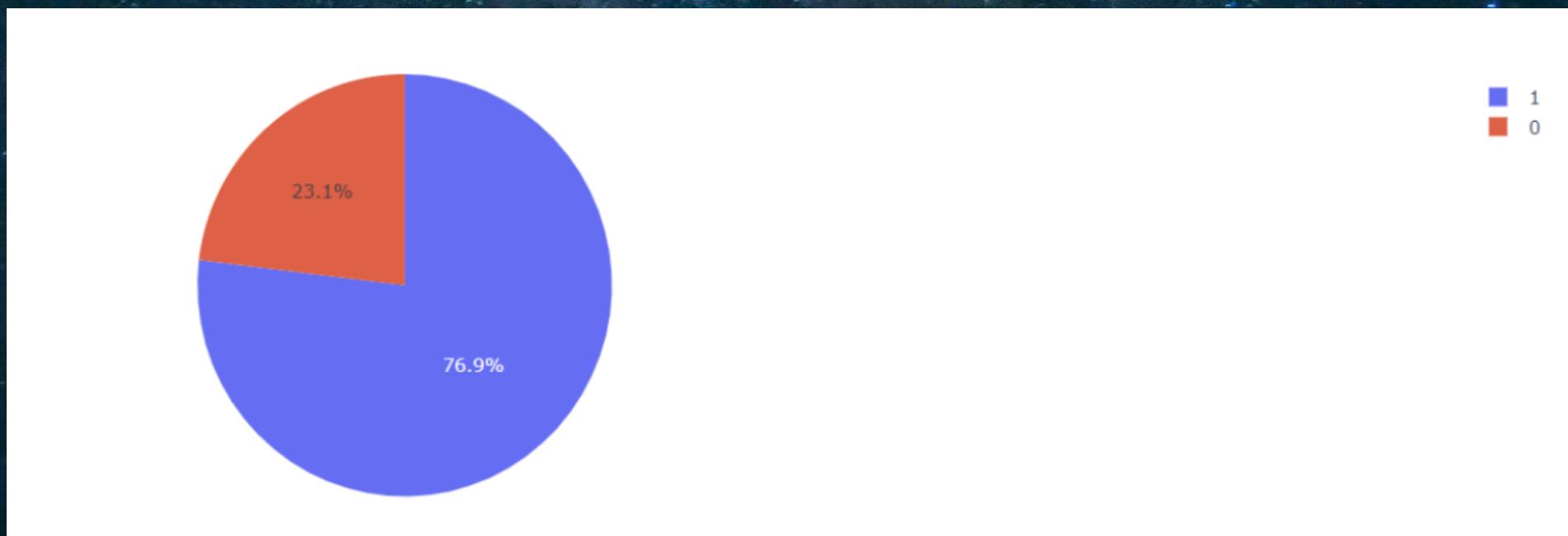
- KSC LC-39A and CCAFS LC-40 have the largest shares, indicating they are SpaceX's most frequently used and reliable sites
- VAFB SLC-4E and CCAFS SLC-40 has a smaller share, showing it is less frequently used
- KSC LC-39A – 41.7%
- CCAFS SLC-40 – 12.5%
- VAFB SLC-4E – 16.7%
- CCAFS LC-40 – 29.2%



KSC LC-39A Launch Success Ratio (Pie Chart)

- Key Points:

- Success: 76.9% (1)
- Failure: 23.1% (0)
- Confirms KSC LC-39A as a high-performing launch site for SpaceX.



Payload Mass vs. Launch Outcome (All Sites)

- The scatter plot shows how payload mass impacts launch outcome across all sites
- Success rate is highest for payloads between 2,000–6,000 kg
- Booster version FT consistently shows strong performance across payload ranges
- Very high or very low payloads show more variability in outcomes, possibly due to mission complexity



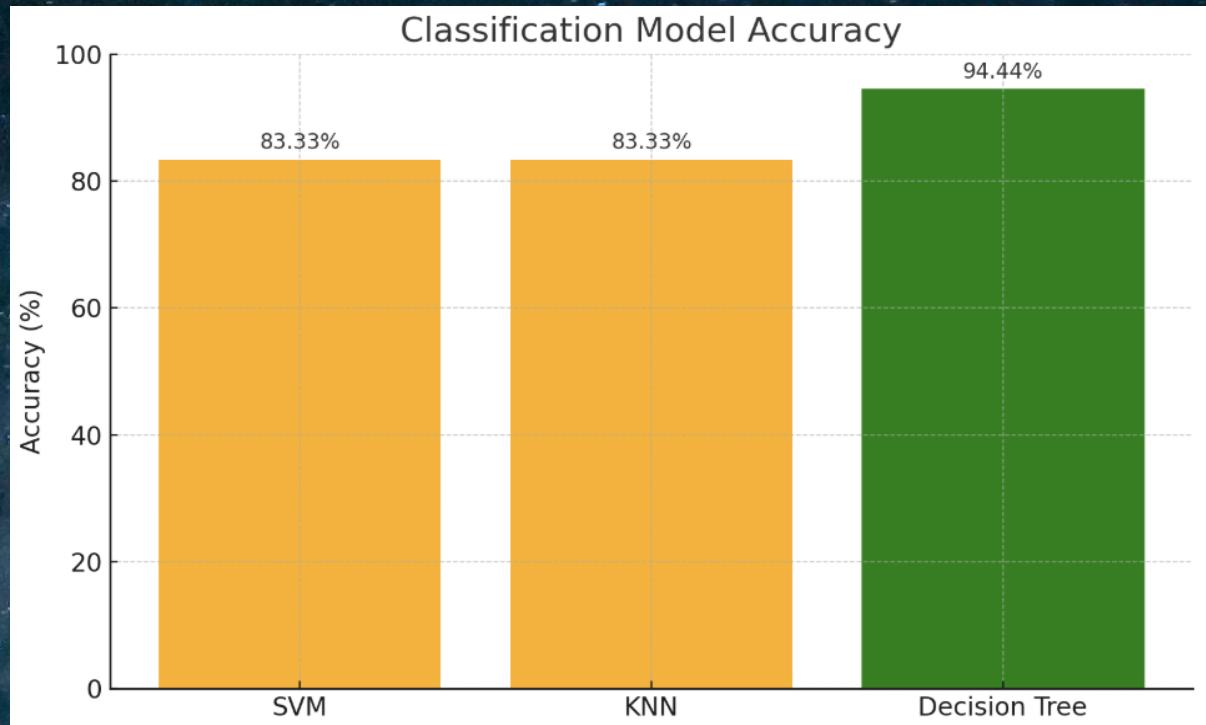
The background of the slide features a dynamic, abstract design. It consists of several thick, curved lines in shades of blue and yellow, creating a sense of motion and depth. The lines curve from the bottom left towards the top right, with some lines being more prominent than others. The overall effect is reminiscent of a tunnel or a high-speed journey through a digital space.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

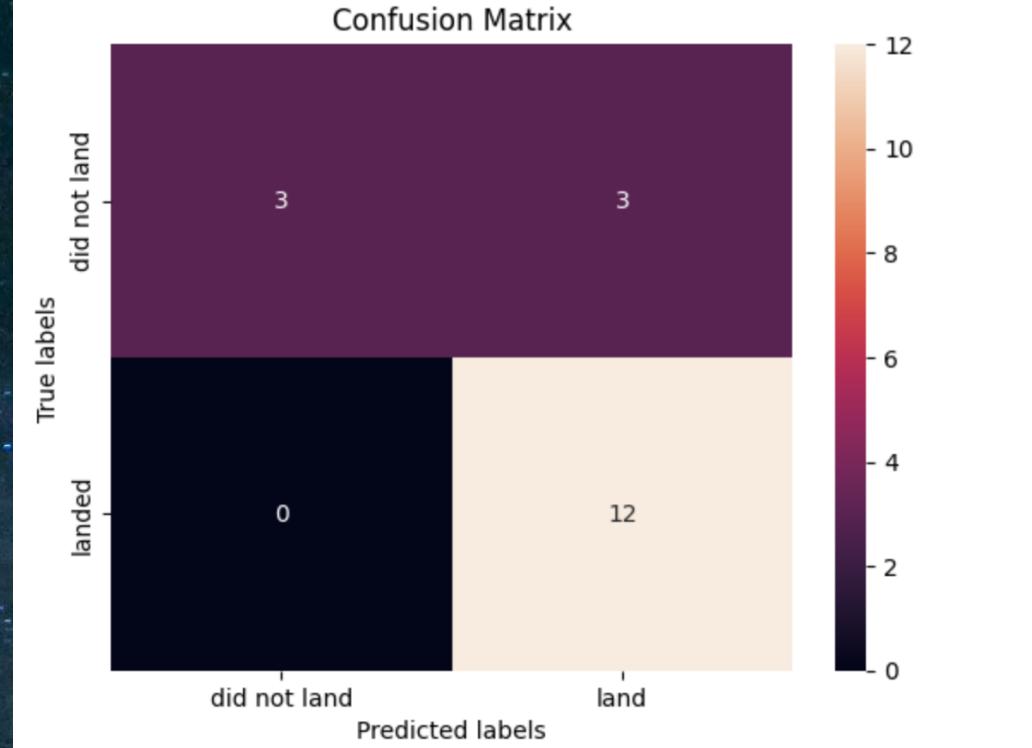
- Accuracy Scores:
 - + Support Vector Machine (SVM): 83.33%
 - + K-Nearest Neighbors (KNN): 83.33%
 - + Decision Tree: 94.44%
- Finding:
 - + The Decision Tree model achieved the highest classification accuracy, making it the best-performing model for predicting Falcon 9 first stage landing success.



Confusion Matrix

- **True Positives (TP):** 12 — Correctly predicted successful landings
- **True Negatives (TN):** 3 — Correctly predicted failures
- **False Positives (FP):** 3 — Predicted success, but actually failed
- **False Negatives (FN):** 0 — No missed successful landings
- **Key Insight:**
- With no false negatives, the model reliably identifies every actual success, making it very effective for mission planning and safety.

```
yhat = tree_cv.predict(X_test)  
plot_confusion_matrix(Y_test,yhat)
```



```
GridSearchCV  
GridSearchCV(cv=10, estimator=DecisionTreeClassifier(),  
param_grid={'criterion': ['gini', 'entropy'],  
'max_depth': [2, 4, 6, 8, 10, 12, 14, 16, 18],  
'max_features': ['log2', 'sqrt', None],  
'min_samples_leaf': [1, 2, 4],  
'min_samples_split': [2, 5, 10],  
'splitter': ['best', 'random']})  
estimator: DecisionTreeClassifier  
DecisionTreeClassifier()  
DecisionTreeClassifier()
```

Conclusions

- Built a machine learning model to predict Falcon 9 first stage landing success
- Identified payload mass, orbit type, and launch site as key success factors
- KSC LC-39A had the highest launch success ratio across all sites
- The Decision Tree model achieved the best accuracy (94.44%) with no false negatives
- Interactive dashboards and maps provided valuable visual insights into launch patterns and performance



Thank you!

