SpaceX Falcon 9 First Stage Landing Prediction

Przemyslaw Marcowski

Outline

1. Executive Summary
2. Introduction
3. Methodology
4. Results
5. Conclusion
6. Appendix

Executive Summary

- **Objective:** Predict SpaceX Falcon 9 first stage landing success

- **Key Elements:**
  - Data collection via SpaceX API and web scraping
  - Exploratory data analysis using visualization and SQL
  - Interactive visual analytics with Folium and Plotly Dash
  - Predictive analysis using classification models

- **Key Outcome:**
  Developed models with ~90% test accuracy for predicting launch success

Introduction

- **Background:**
  - SpaceX aims to revolutionize space transportation with reusable rockets
  - Falcon 9 launches cost ~$62 million vs. $165+ million for competitors
  - Predicting successful landings is critical for cost reduction and mission planning

- **Problem Statement:**
  Determine first stage landing probability to estimate launch costs

- **Key Questions:**
  1. Can we predict future launch success based on historical data?
  2. What factors contribute most to a successful launch?

Methodology

Data Collection

- **SpaceX API:**
  - Endpoint: https://api.spacexdata.com/v4/launches/past
  - Collected launch details in JSON format
  - Date cutoff: 2020-11-13

- **Web Scraping:**
  - Source: Wikipedia's List of Falcon 9 launches
  - Extracted additional launch information

SpaceX API Data Collection

- **Process:**
  - Used SpaceX REST API to retrieve launch data
  - Collected data on rockets, launch sites, payloads, and cores

Fetch Launch Data from API → JSON → DataFrame → Extract Relevant Fields → Preprocess and Clean Data → Save Data to Flat File

GitHub URL: [Data Collection API]

Web Scraping

- **Process:**
  - Used **BeautifulSoup** to scrape launch data from Wikipedia.
  - Extracted launch details, including date, booster version, launch site, payload, and landing outcomes.

Request HTML Content From Wikipedia → Parse HTML with BeautifulSoup → Locate Target Tables → Extract and Clean Data → Save Data to Flat File

GitHub URL: [Web Scraping]

Data Wrangling

- **API Data:**
  - Handled missing values *via* imputation
  - Filtered data to include only Falcon 9 launches
  - Created binary landing outcome classification (1=landed, 0=did not land)
  - Performed feature engineering

- **Web-Scraped Data:**
  - Transformed from raw into structured format
  - Cleaned

GitHub URL: Data Wrangling

Exploratory Data Analysis (EDA)

- **Visualization Techniques:**
  - Scatter plots: Flight Number vs. Orbit type, Payload vs. Orbit type
  - Bar charts: Success rate by orbit type
  - Line charts: Yearly average launch success trend

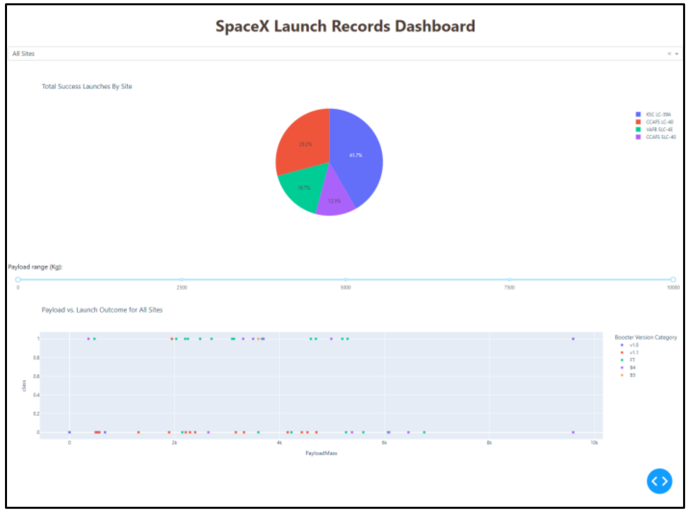
- **SQL Queries:**
  - Identified unique launch sites
  - Calculated total payload for NASA missions
  - Analyzed success/failure ratios

GitHub URL: EDA with Visualization

GitHub URL: EDA with SQL

Interactive Visual Analytics

▪ **Folium Maps:**
- Plotted launch site locations
- Color-coded markers for success/failure
- Calculated distances to nearby facilities



▪ **Plotly Dash Dashboard:**
- Success rate distribution by launch site
- Payload mass vs. launch success correlation

GitHub URL: [Geographical Analysis]

GitHub URL: [Dashboard]

Predictive Analysis

- **Preprocessing:**
  Data cleaning, transformation, and splitting

- **Model Selection:**
  Tested multiple classification algorithms

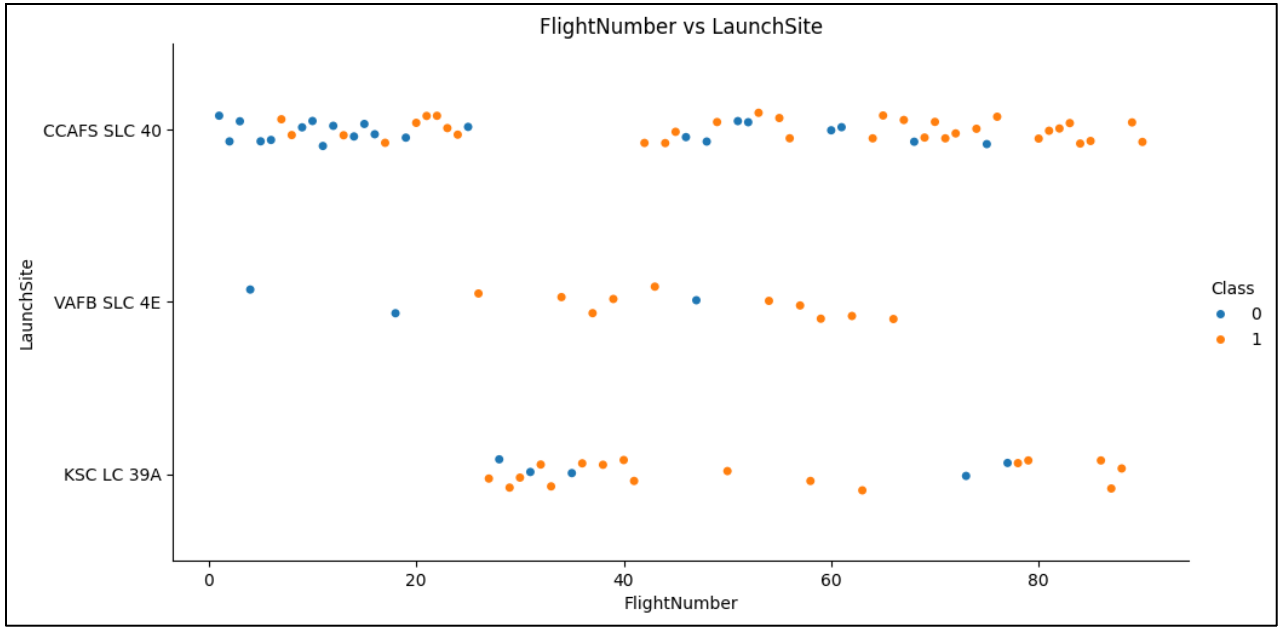
- **Hyperparameter Tuning:**
  Used grid search with stratified *k*-fold cross-validation strategy for optimization

- **Evaluation:** Compared model performances
  using confusion matrices and multiple performance metrics

GitHub URL: [Landing Prediction]

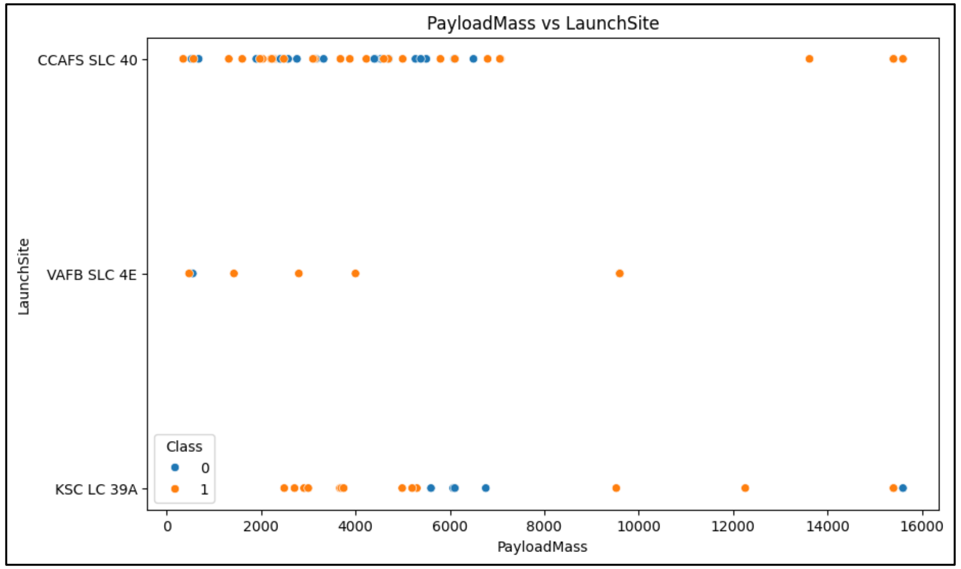
EDA Results

Flight Number vs. Launch Site



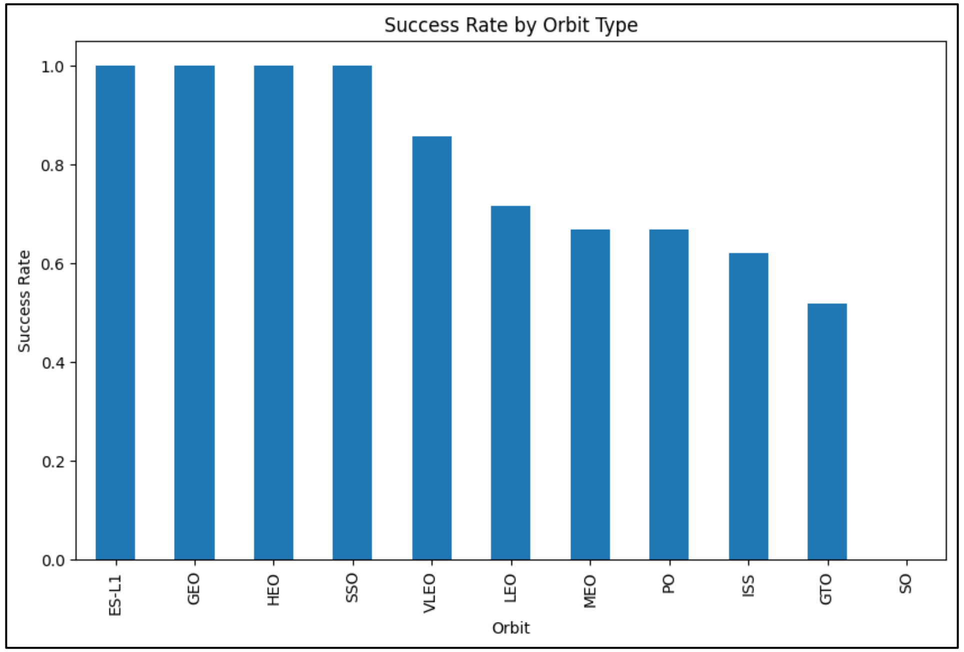
- **Launch Site Performance:**
  - KSC LC 39A shows higher success rates (Class 1) as flight numbers increase
  - CCAFS SLC 40 has a balanced mix of successes and failures across all flight numbers
  - VAFB SLC 4E has fewer launches but tends toward successful outcomes
- **Improvement Over Time:**
  Increase in successful launches with higher flight numbers, suggesting improvement with experience.

Payload Mass vs. Launch Site



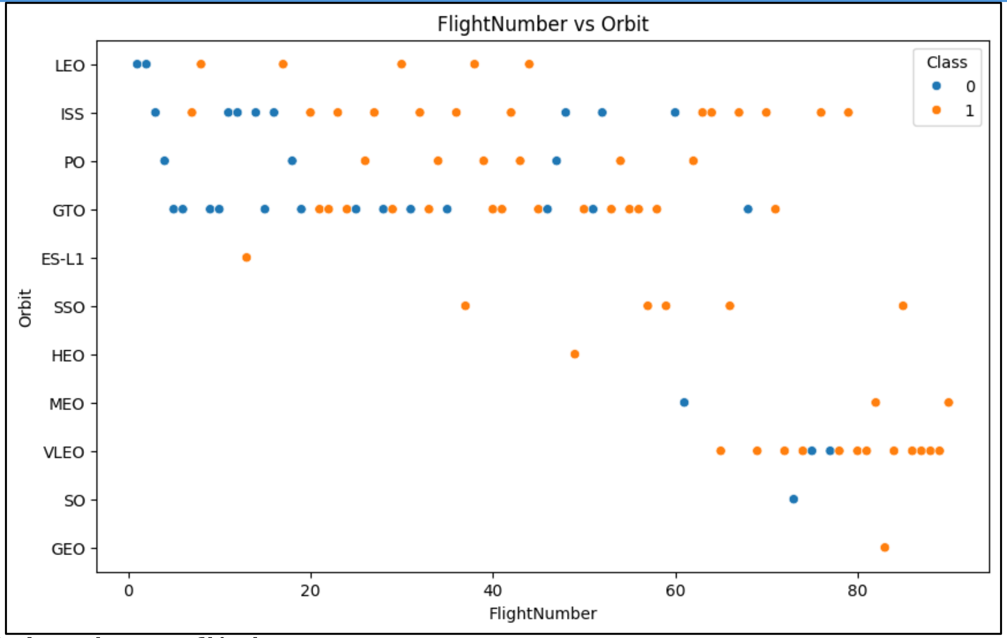
- **Launch Sites:**
  - 3 on the east coast, 1 on the USA west coast
  - KSC LC-39A showed highest success rate
- **Payload Mass Distribution:**
  - CCAFS SLC-40 has the most launches, with payload masses concentrated between 2,000 to 8,000 units
  - VAFB SLC-4E has fewer launches, with payloads mostly below 2,000 units
  - KSC LC-39A shows a wide range of payload masses, from very light to the heaviest (up to 16,000 units)

Success Rate vs. Orbit Type



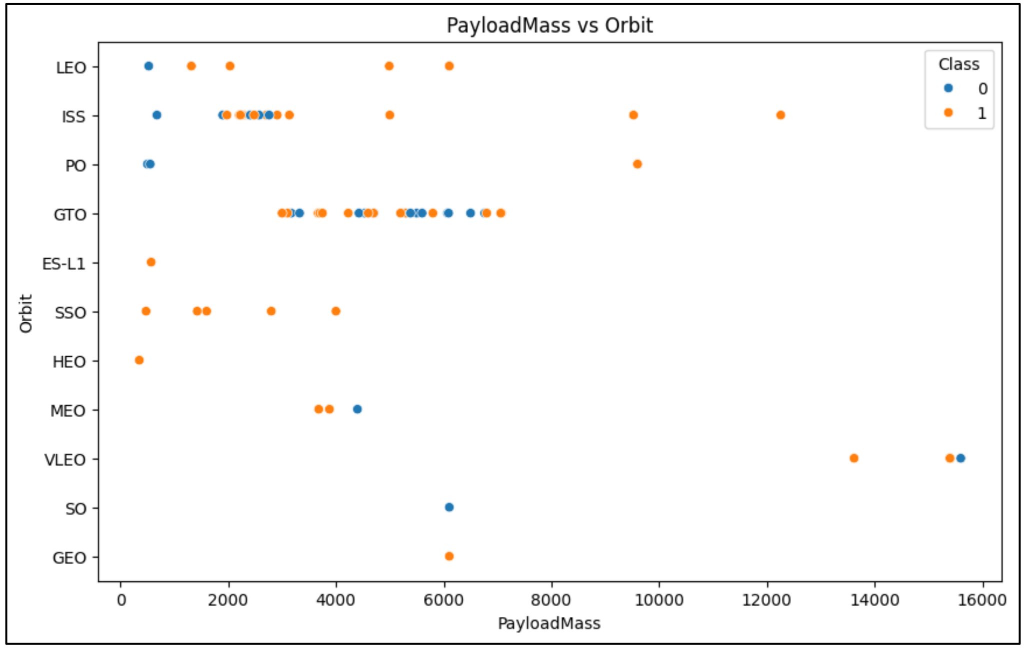
- **High Success Rates:**
  - Orbits like ES-L1, GEO, HEO, and SSO have a 100% success rate, indicating reliable performance in these missions
  - VLEO (Very Low Earth Orbit) also shows a high success rate, just below the top tier
- **Moderate Success Rates:**
  LEO (Low Earth Orbit), MEO (Medium Earth Orbit), and PO (Polar Orbit) have moderate success rates, ranging between 60% to 80%
- **Lower Success Rates:**
  Missions to the ISS (International Space Station) and GTO (Geostationary Transfer Orbit) have lower success rates, with GTO being notably the least successful

Flight Number vs. Orbit Type



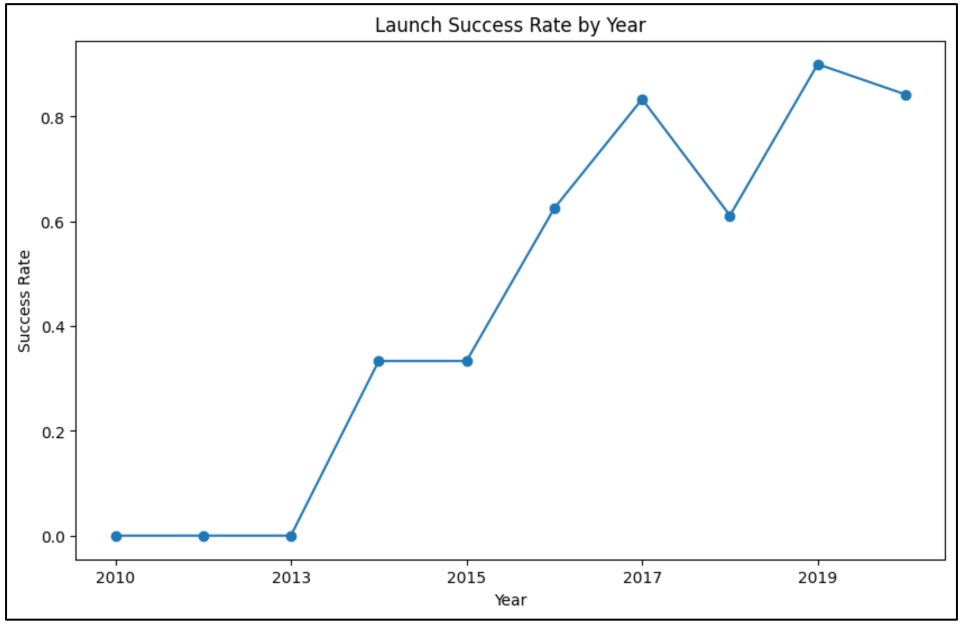
- **Success Trends Over Time:**
  - For orbits like GEO, SSO, and ES-L1, later flights tend to have more successes (Class 1), suggesting an improvement over time
  - VLEO and MEO show a strong trend toward successful launches as flight numbers increase, reflecting improved reliability
- **Consistent Performance:**
  - Orbits such as HEO, SSO, and GEO have more consistent successes (Class 1) across different flight numbers
- **Challenges with Specific Orbits:**
  - GTO and LEO have a higher presence of failures (Class 0), especially in earlier flight numbers, indicating these orbits were more challenging initially
  - ISS-related missions also show a mix of success and failure, though performance improves with higher flight numbers

Payload Mass vs. Orbit Type



- **Payload Mass and Success:**
  - Higher payloads (over 10,000 kg) are generally associated with successful launches (Class 1), especially in GTO and GEO
  - Lower payloads have a mix of successes and failures across various orbits
- **Orbit-Specific Trends:**
  - GTO shows a range of payloads with both successes and failures, suggesting it's challenging for medium-range payloads
  - ISS missions mostly have smaller payloads, with mixed outcomes
- **High Payload Success:**
  Success rates improve with larger payloads in orbits like GEO, indicating better handling of heavier loads over time

Launch Success Yearly Trend



- **Overall Improvement:**
  Significant increase in success rate from 2013 onwards, showing improved reliability over time
- **Steady Growth:**
  Rapid growth between 2014 and 2017, with a success rate climbing from 0.4 to over 0.8
- **Fluctuations:**
  Dip in success around 2018, followed by recovery and stabilization at a high success rate in 2019
- **Early Challenges:**
  No successful launches before 2014, indicating early challenges in achieving consistent outcomes

All Launch Site Names

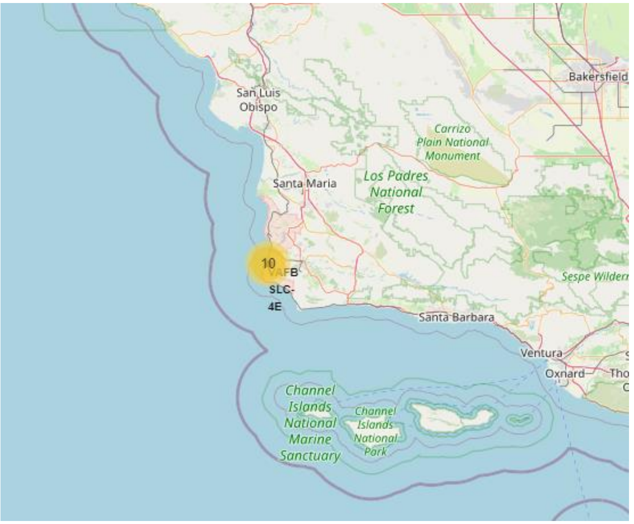
```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEX_TABLE;
```

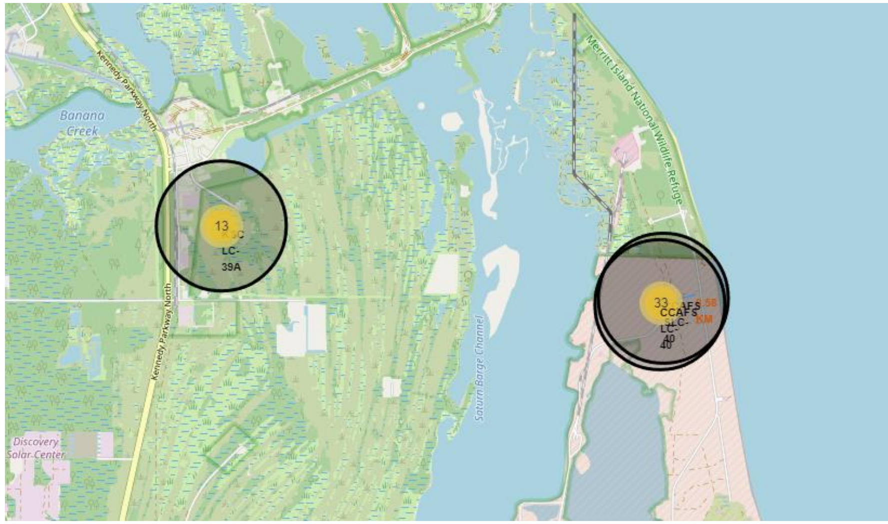
```
 * sqlite:///spacex_data.db
Done.
```

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

- **Four Launch Sites:**

- **3** on the USA west coast:
  - **KSC LC-39A**
  - **CCAFS SLC-40**
  - **CCAFS-LC40**





Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEX_TABLE WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///spacex_data.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

- Launches from CCAFS LC040 site
- All LEO orbit
- NASA was the customer in 4 of them

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS payloadmass FROM SPACEX_TABLE WHERE Customer LIKE 'NASA%';
```

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

payloadmass
99980

- Launches from CCAFS LC040 site
- All LEO orbit
- NASA was the customer in 4 of them

Average Payload Mass by F9 v1.1

```
%sql SELECT ROUND(AVG(PAYLOAD_MASS__KG_), 2) AS avg_payload_mass FROM SPACEX_TABLE WHERE Booster_Version LIKE 'F9 v1.1%';
```

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

avg_payload_mass
2534.67

Average payload mass carried by F9 booster v1.1 is **2534.67**

First Successful Ground Landing Date

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

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

MIN(Date)
2015-12-22

First successful ground pad landing on **22-12-2015**

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT BOOSTER_VERSION FROM SPACEX_TABLE WHERE LANDING_OUTCOME='Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

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

Boosters which successfully landed on drone ship with 4000-6000 unit payloads:

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

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS missionoutcomes FROM SPACEX_TABLE GROUP BY MISSION_OUTCOME;
```

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

Mission_Outcome	missionoutcomes
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Total Outcomes:

- Success: **100**
- Failure: **1**

Boosters Carried Maximum Payload

Booster names which carried maximum payload:

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

```
[10] %sql SELECT BOOSTER_VERSION FROM SPACEX_TABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEX_TABLE);
```

```
 * sqlite:///spacex_data.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

```
%sql SELECT SUBSTR(Date, 6, 2) AS Month, "Landing_Outcome", Booster_Version, Launch_Site FROM SPACEX_TABLE WHERE SUBSTR(Date, 1, 4) = '2015' AND "Landing_Outcome" = 'Failure (drone ship)';
```

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

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

Failed landing outcomes in drone ship with their booster versions and launch site names for in year 2015:

Month	Booster Version	Launch Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

```
%sql SELECT LANDING_OUTCOME, COUNT(*) AS Count FROM SPACEX_TABLE WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY Count DESC;
```

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

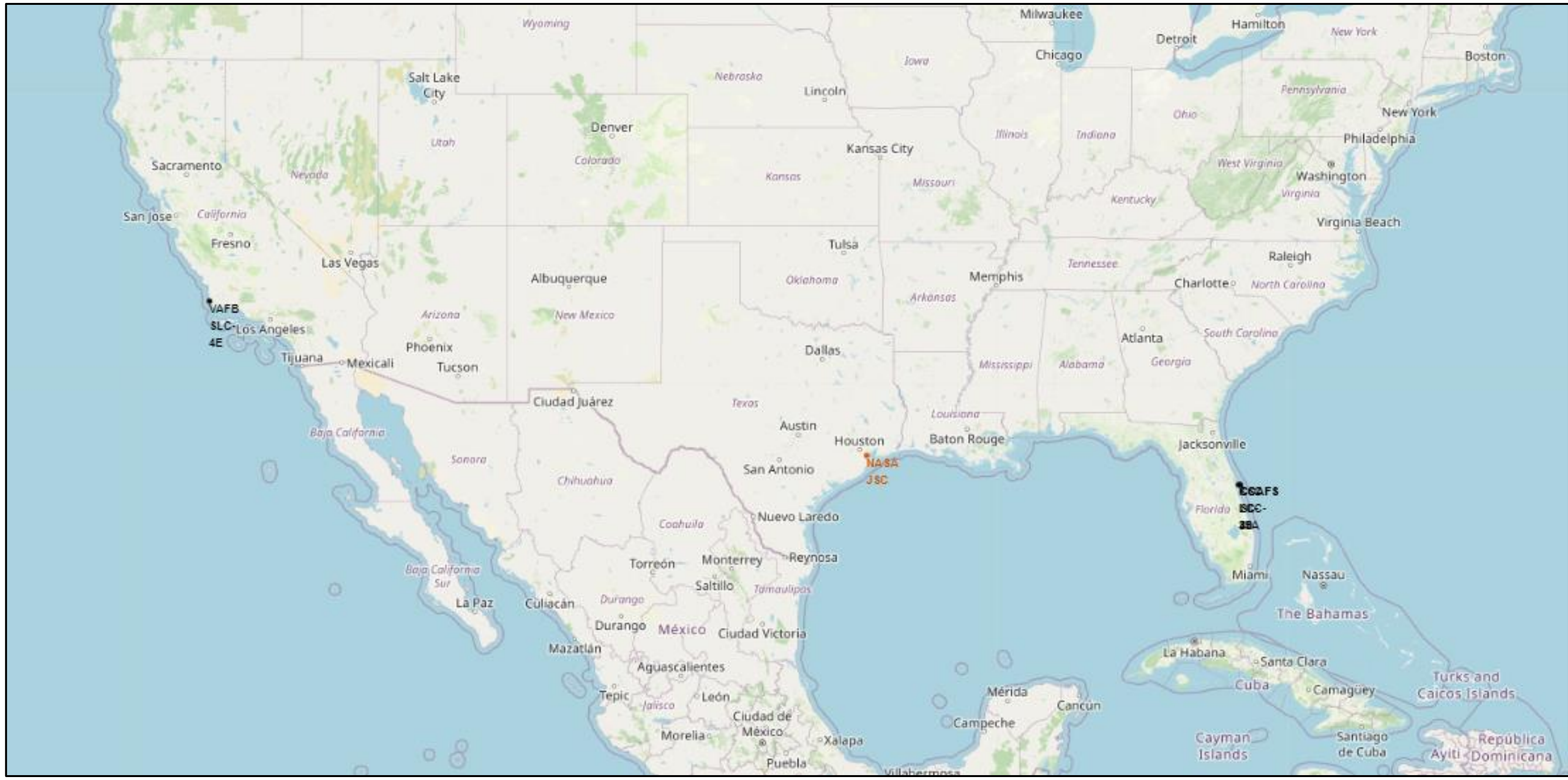
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- **No Attempt Dominance:**
  10 landings had no attempt, making it the most common outcome in the specified period
- **Successes:**
  - 5 successful landings on a drone ship
  - 3 successful landings on a ground pad
- **Failures and Challenges:**
  - 5 failures on a drone ship and 2 parachute-related failures
  - Controlled and uncontrolled landings in the ocean occurred 3 and 2 times, respectively
- **Rare Outcomes:**
  Only 1 instance of a precluded landing on a drone ship.

Launch Sites Proximity Analysis

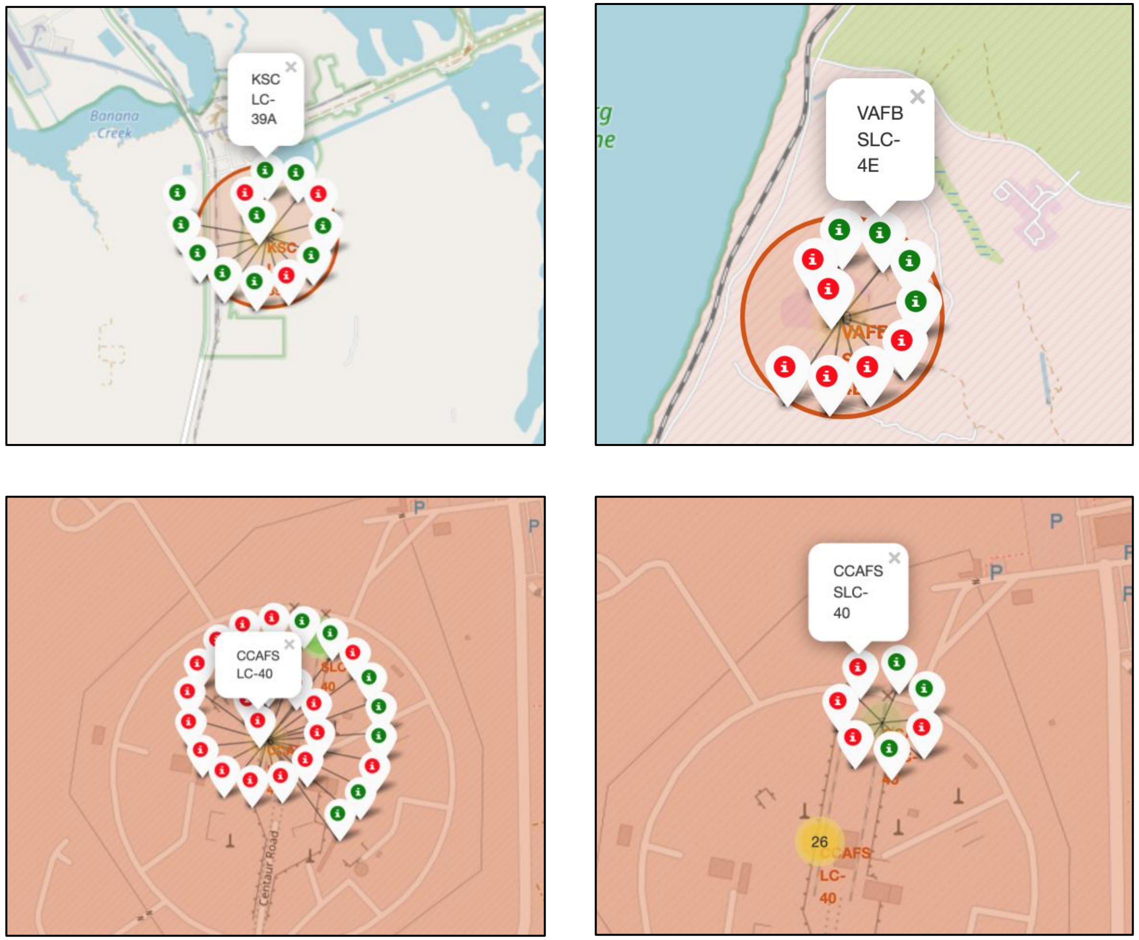
Locations of Launch Sites on Maps

Launch Site	Latitude	Longitude
CCAFS LC-40	28.562302	-80.577356
VAFB SLC-4E	34.632834	-120.610745
KSC LC-39A	28.573255	-80.646895
CCAFS SLC-40	28.563197	-80.576820

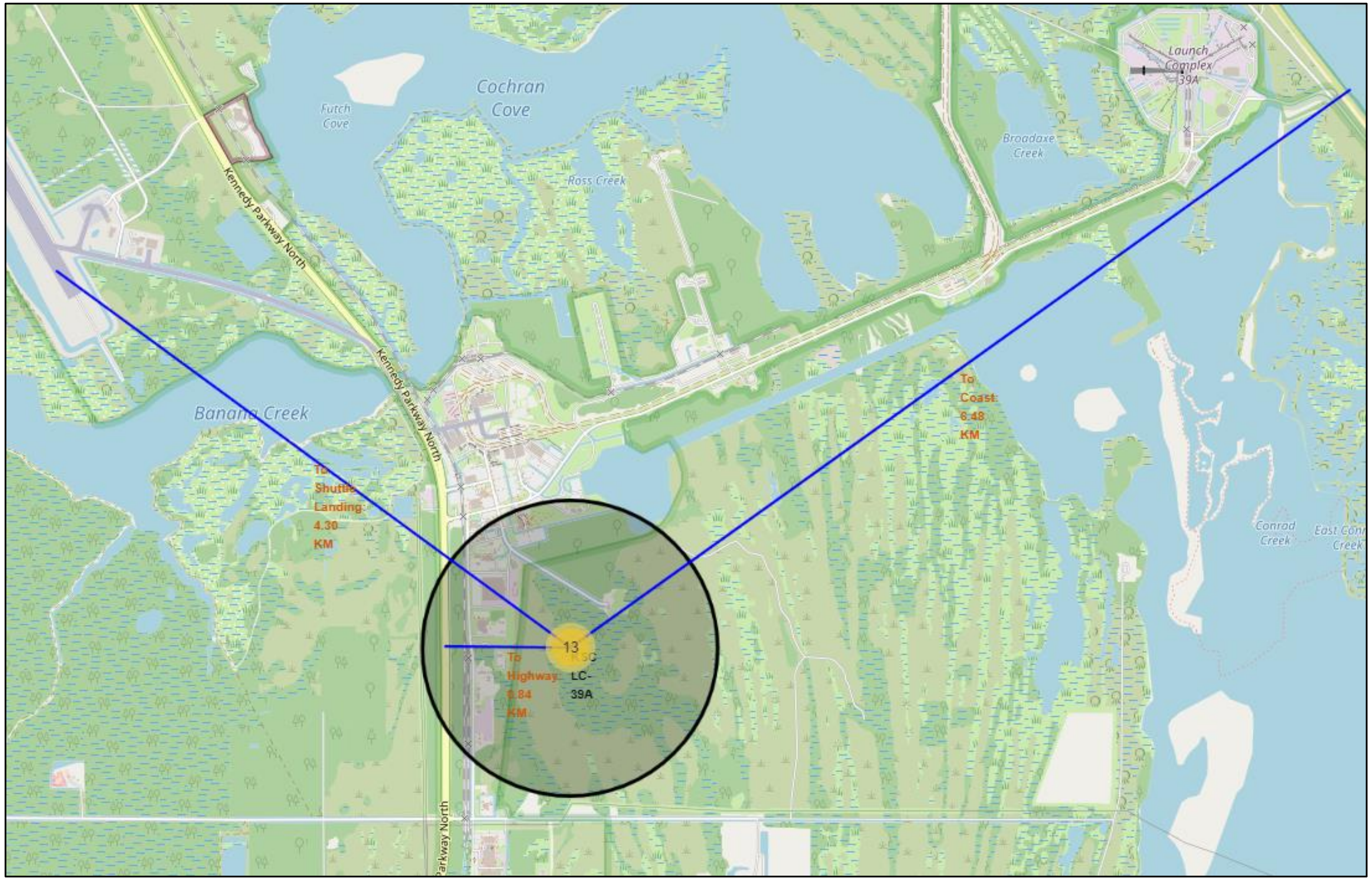


Display Launch Outcome by Color

KSC LC-39A has relatively **higher success rate**
CCAFS LC-40 and **CCAFS SLC-40** have **lower rates**



Show Distance to Proximities



KSC LC-39A Proximities:

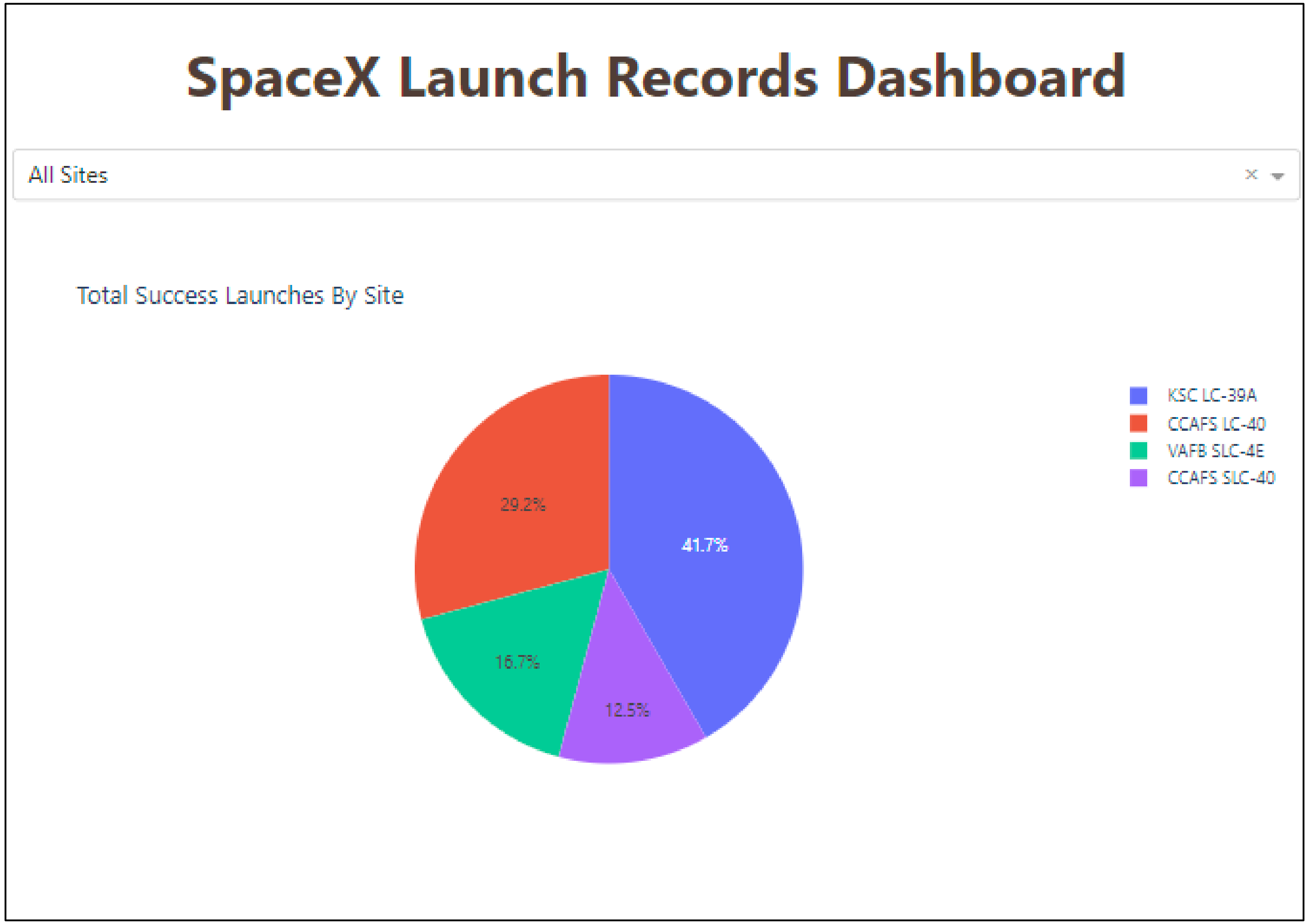
- **Shuttle Landing: 4.30 KM**
- **Highway: 0.84 KM**
- **Coastline: 6.48 KM**

Build a Dashboard with Plotly Dash

Total Successful Launches for All Sites

Success Rate for All Launch Sites:

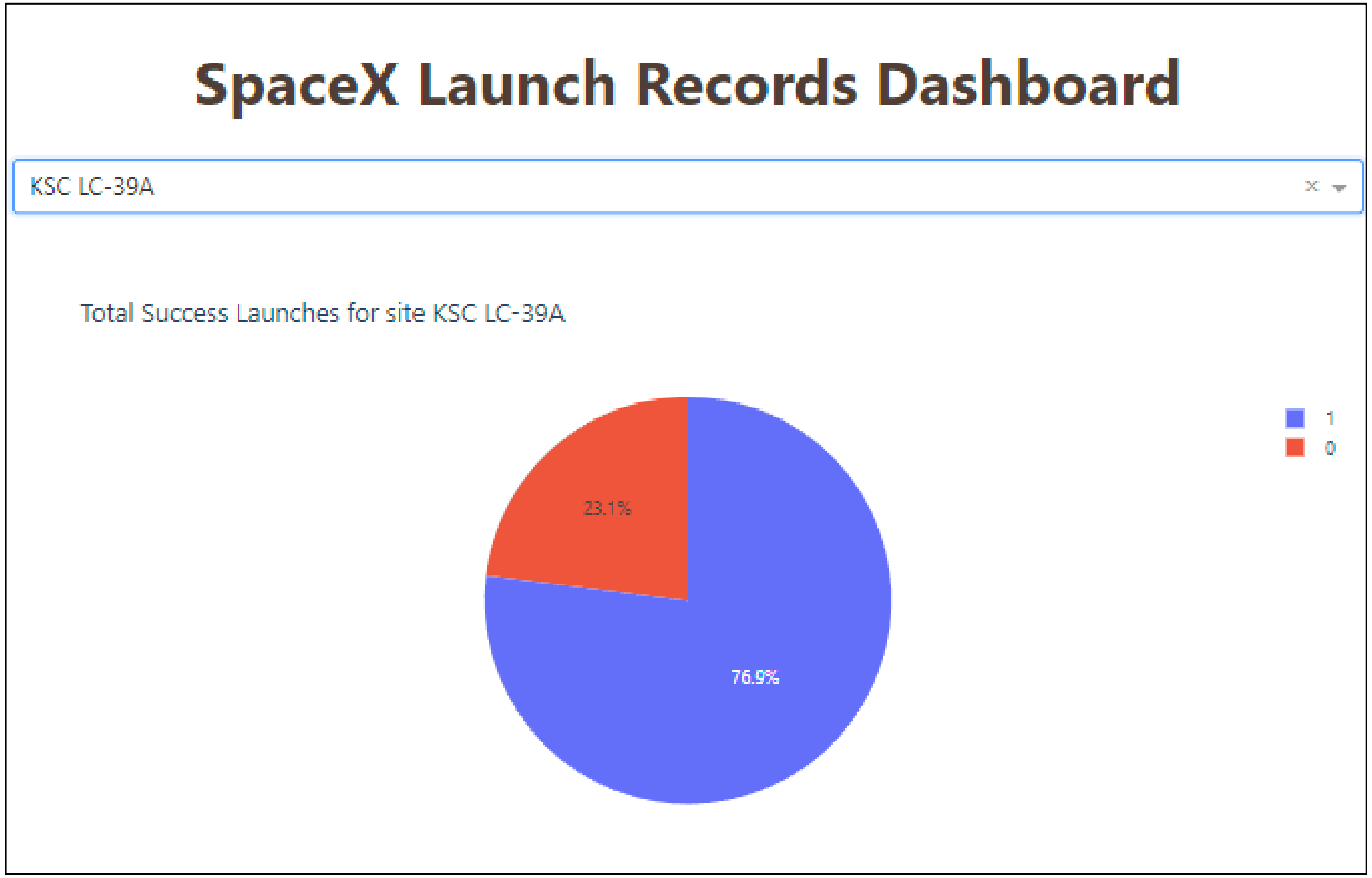
- **CCAFS LC-40: 29.2%**
- **VAFB SLC-4E: 16.7%**
- **KSC LC-39A: 41.7%**
- **CCAFS SLC-40: 12.5%**



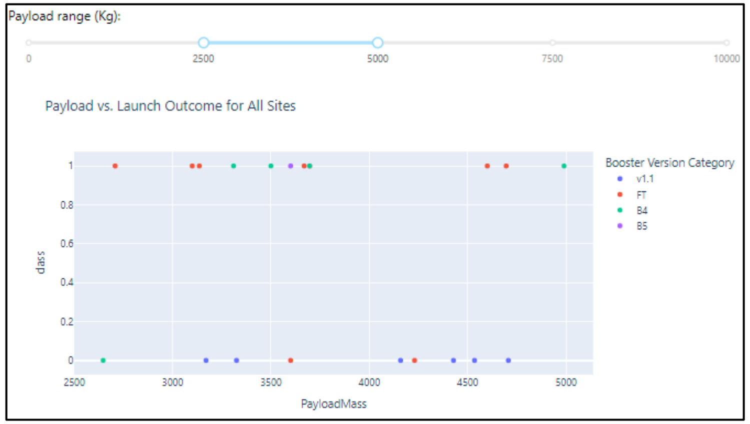
Success Ratio for KSC LC-39A

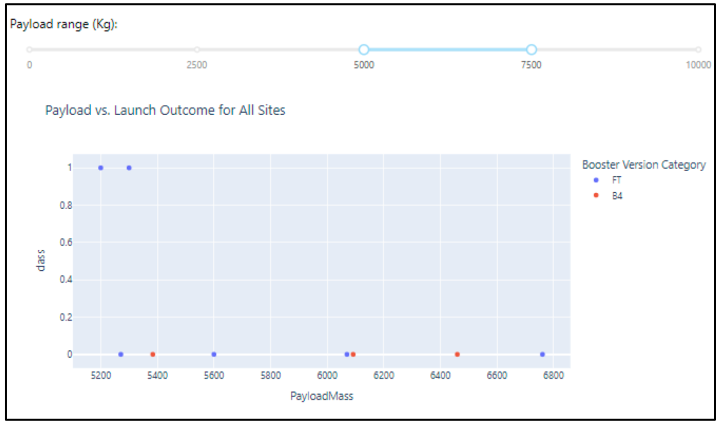
Success Rate for All Launch Sites:

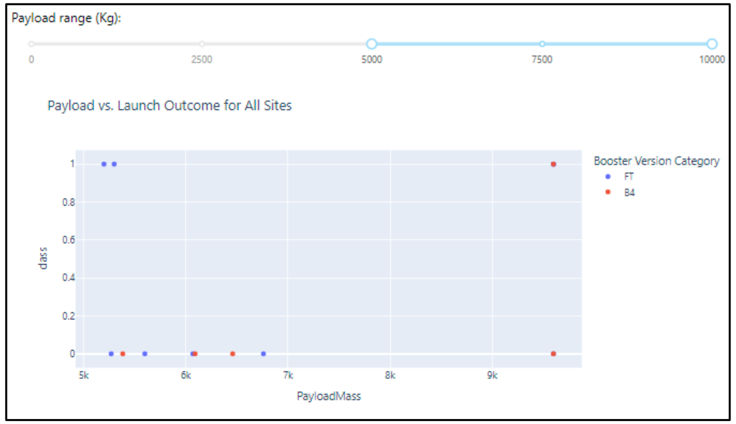
- **CCAFS LC-40: 29.2%**
- **VAFB SLC-4E: 16.7%**
- **KSC LC-39A: 41.7%**
- **CCAFS SLC-40: 12.5%**



Success Rate vs. Payload Mass







- **Lighter Payloads (2,500–5,000 kg):** Higher success rates across all boosters
- **Medium Payloads (5,000–7,000 kg):** FT version performs best; B4 is less consistent
- **Heavy Payloads (>7,000 kg):** FT still succeeds; B4 struggles with variability
- **Booster Versions:**
  - **FT:** Consistent success across payloads
  - **B4 & B5:** Mixed outcomes, especially with heavier payloads
  - **v1.1:** Less reliable with lighter payloads.FT is the most versatile and reliable booster

Predictive Analysis

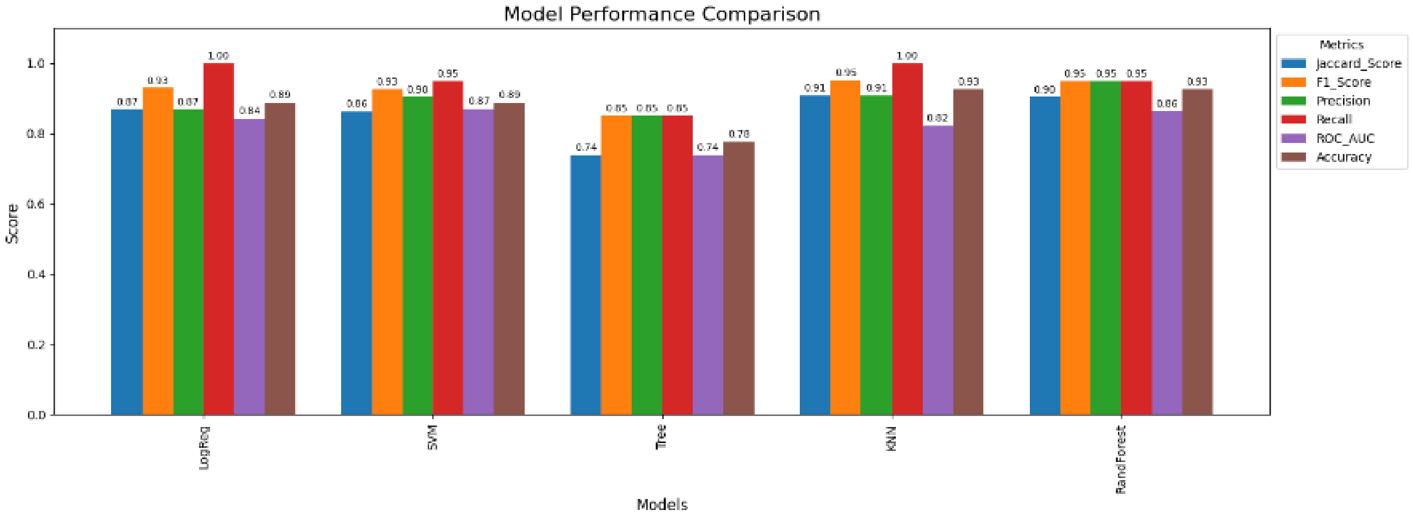
Classification Accuracy

- Models evaluated on multiple metrics
- SVM selected as best solution using ROC AUC (Score: 0.8714)
- Best parameters:
  - 'C': 1.0,
  - 'gamma': 0.03162277660168379,
  - 'kernel': 'sigmoid'

```
Model Performance Metrics:
```

	Jaccard_Score	F1_Score	Precision	Recall	ROC_AUC	Accuracy
LogReg	0.869565	0.930233	0.869565	1.00	0.842857	0.888889
SVM	0.863636	0.926829	0.904762	0.95	0.871429	0.888889
Tree	0.739130	0.850000	0.850000	0.85	0.739286	0.777778
KNN	0.909091	0.952381	0.909091	1.00	0.821429	0.925926
RandForest	0.904762	0.950000	0.950000	0.95	0.864286	0.925926

```
Best model for each metric:
Jaccard_Score: KNN (Score: 0.9091)
F1_Score: KNN (Score: 0.9524)
Precision: RandForest (Score: 0.9500)
Recall: LogReg (Score: 1.0000)
ROC_AUC: SVM (Score: 0.8714)
Accuracy: KNN (Score: 0.9259)
```



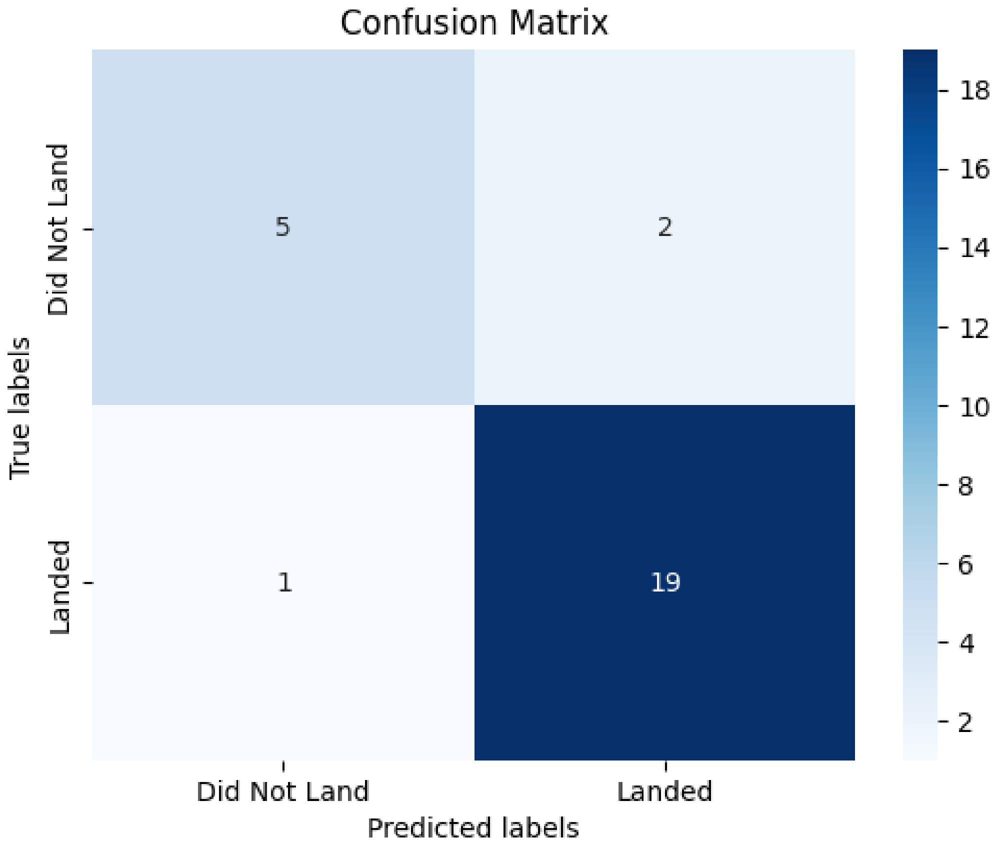
Confusion Matrix

- **Best Model:** SVM (ROC_AUC: 0.87)
- **Confusion Matrix:**
  - 19 True Positives, 5 True Negatives
  - 2 False Positives, 1 False Negative
- **Metrics:**
  - Precision: 0.90
  - Recall: 0.95 for “Landed”
  - Overall Accuracy: 0.89

SVM performs well
at predicting successful landings
given limited training data

```
Best model based on ROC_AUC: SVM (Score: 0.8714)

Confusion Matrix for SVM:
```




```
Classification Report for SVM:

              precision    recall  f1-score   support

Did Not Land       0.83      0.71      0.77         7
      Landed       0.90      0.95      0.93        20

    accuracy                           0.89        27
   macro avg       0.87      0.83      0.85        27
weighted avg       0.89      0.89      0.89        27
```

Conclusions

- **Consistent Improvement:** SpaceX's success rate has steadily increased, showcasing effective learning and reliability over time
- **FT Booster Advantage:** The FT booster excels across various payloads, handling heavier missions where older versions faced challenges, highlighting engineering progress
- **Experience Matters:** Higher flight numbers correlate with better outcomes, demonstrating that experience is key to tackling complex missions like GTO and ISS
- **Data-Driven Optimization:** SpaceX’s tailored booster strategies have improved reliability, ensuring success across a range of missions
- **SVM Model Insights:** Predictive modeling with SVM achieved high accuracy (0.89), underscoring the value of data in optimizing launch outcomes
- **Leading the Field:** SpaceX’s advancements and data-driven approach position it as a leader in space exploration, ready for future challenges

Appendix

- **Links:**
  - Project Repository
  - SpaceX API Documentation

- **References:**
  - **Wikipedia:** List of Falcon 9 and Falcon Heavy launches
  - **Python Libraries:** pandas, numpy, matplotlib, seaborn, scikit-learn, folium, plotly, dash

- **Acknowledgments**
  - Special thanks to the SpaceX API team for providing access to launch data
  - Acknowledgment to the open-source community for tools and libraries used

Thank You