IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Albeiro Gonzalez>
<01-01-2026>

Outline

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

Executive Summary

This project analyzes historical data of launches of rockets model Falcon 9 by SpaceX to determine factors that influence the success of landings of reusable rockets allowing understanding the possible reduction of costs in the launches.

Using data analysis techniques such as data collection(via SQL, APIs, and web scraping), data wrangling(with Python packages), data visualization(with Python libraries and Plotly for interactive visualization), and Machine Learning for predictions this project intends to show useful insights to make decisions for future launches.

The analysis revealed that payload mass, launch site, and booster version significantly influence landing success. Several machine learning models were evaluated, all achieving similar accuracy (~83%),demonstrating that historical launch data can be effectively used to predict landing outcomes and support future operational decisions.

Introduction

SpaceX is the most successful private company in the modern space race. It has revolutionized the aerospace industry by reusing rockets in the first stage of the landings, dramatically reducing the costs of launches to USD 67 million. In this project, we aim to predict the success rate of Falcon 9 boosters' first-stage landing. The underlying hypothesis is: Accurate prediction of the first-stage landing can predict cost reduction in future launches.

To test this hypothesis, we would like to answer the following questions:

- What is the overall success rate of SpaceX Falcon 9's first-stage landings?
- Does the launch site affect the success rate of the first-stage landing ?
- Does the payload mass influence the success rate of the first-stage landing ?
- Do orbits impact the success of the first-stage landing ?
- Does the first-stage landing success rate improve over time?
- Does the recovery of boosters contribute to cost reduction in missions?

Section 1

Methodology

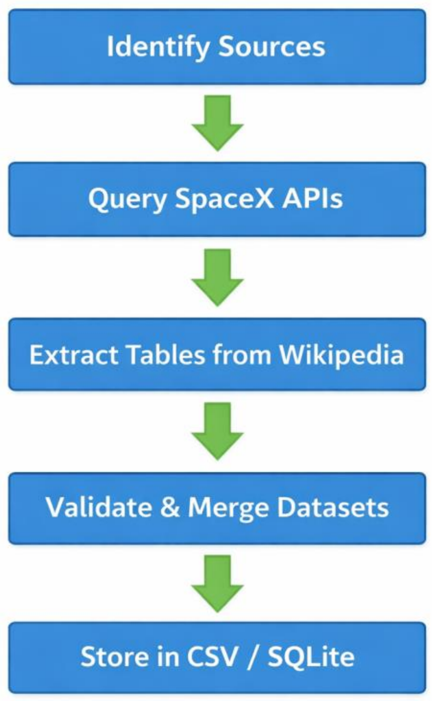
Methodology

Executive Summary

- Data collection methodology: Historical SpaceX launch data was gathered from public sources such as Wikipedia and the SpaceX API. The dataset includes vehicle characteristics, payload mass, launch site information, and landing outcomes.
- Data wrangling : The data was cleaned and transformed by handling missing values, removing inconsistencies, standardizing formats, and preparing categorical and numerical variables for analysis.
- Exploratory data analysis (EDA) using visualization and SQL : Exploratory analysis was made by querying with SQL and visualizing patterns with Python libraries.
- Interactive visual analytics using Folium and Plotly Dash : Folium, a Python librarie, was used to see the launch sites on an interactive map. Plotly Dash was used to build a dashboard application to show the relation between payload mass and success rates of first-stage landings.
- Predictive analysis using classification models : The following ML algorithms were used to predict the successful first-stage landing of the Falcon 9 boosters: Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K-nearest neighbors.

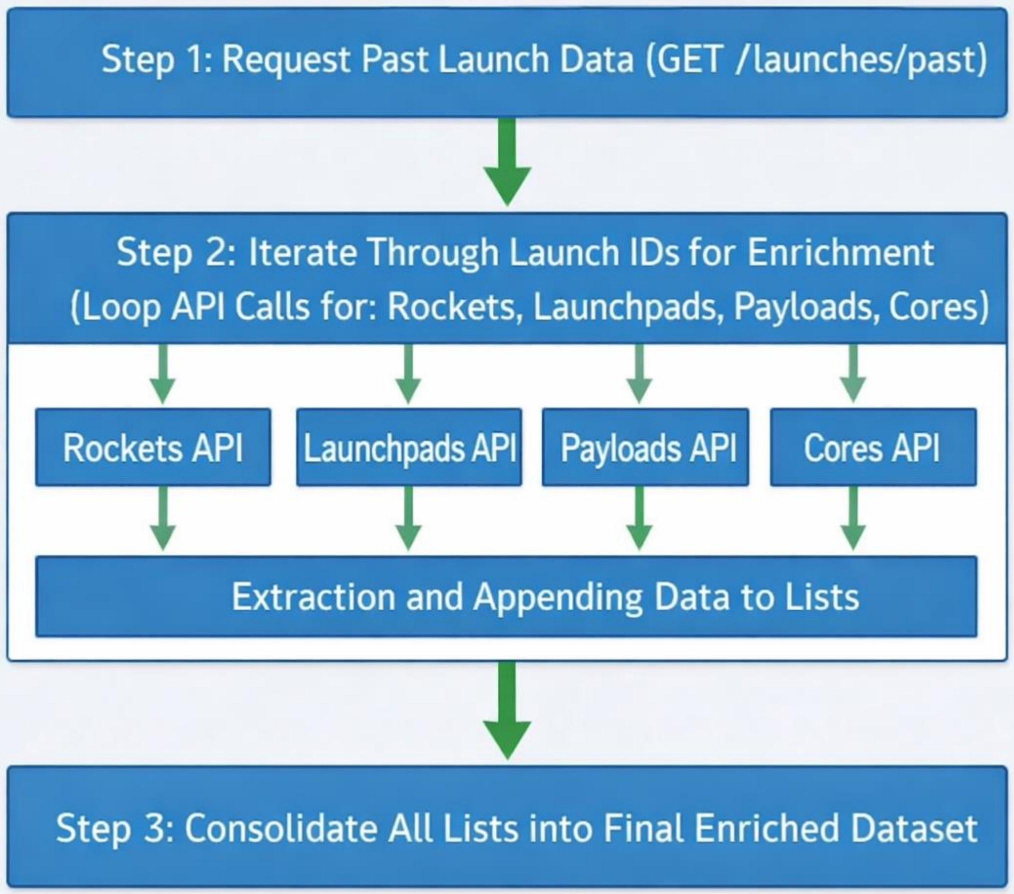
Data Collection

- SpaceX launch data was collected from publicly available sources, including:
  • Wikipedia launch histories
  • SpaceX API datasets
  • Online repositories curated for SpaceX mission analysis
- **Data included:** FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude, and Class.
- **Collection process overview:**
  Identify sources → Query SpaceX APIs→Extract raw tables with BeautifulSoup and Pandas from wikipedia→ Validate & merge datasets → Store in CSV/SQLite for analysis



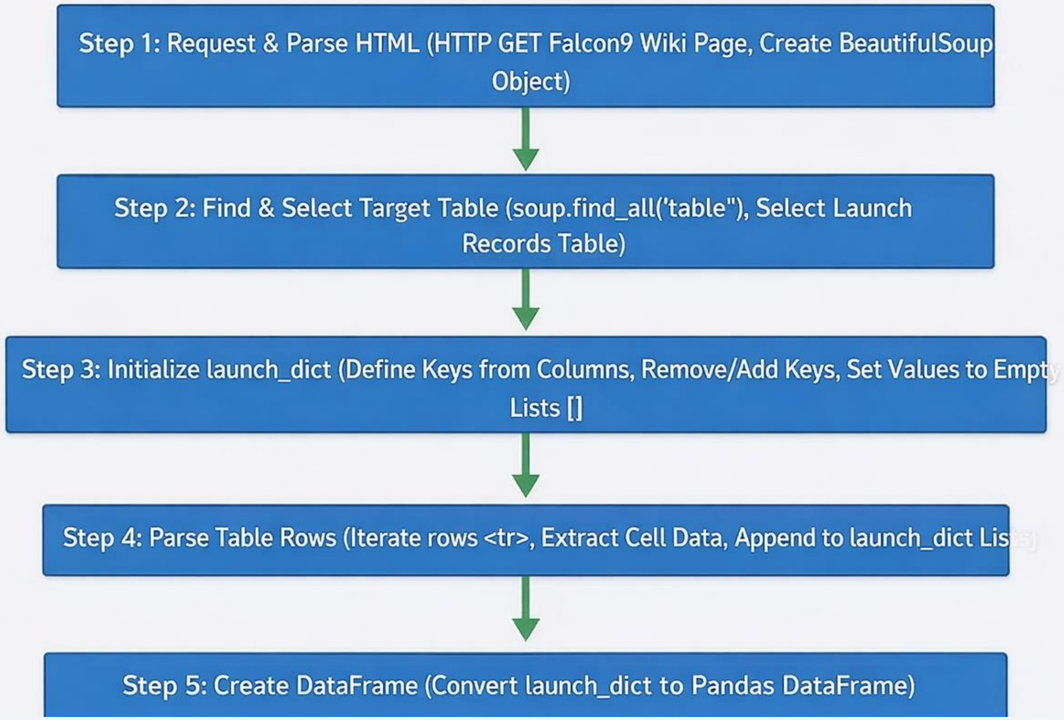
Data Collection – SpaceX API

- GitHub URL of the completed SpaceX API calls notebook



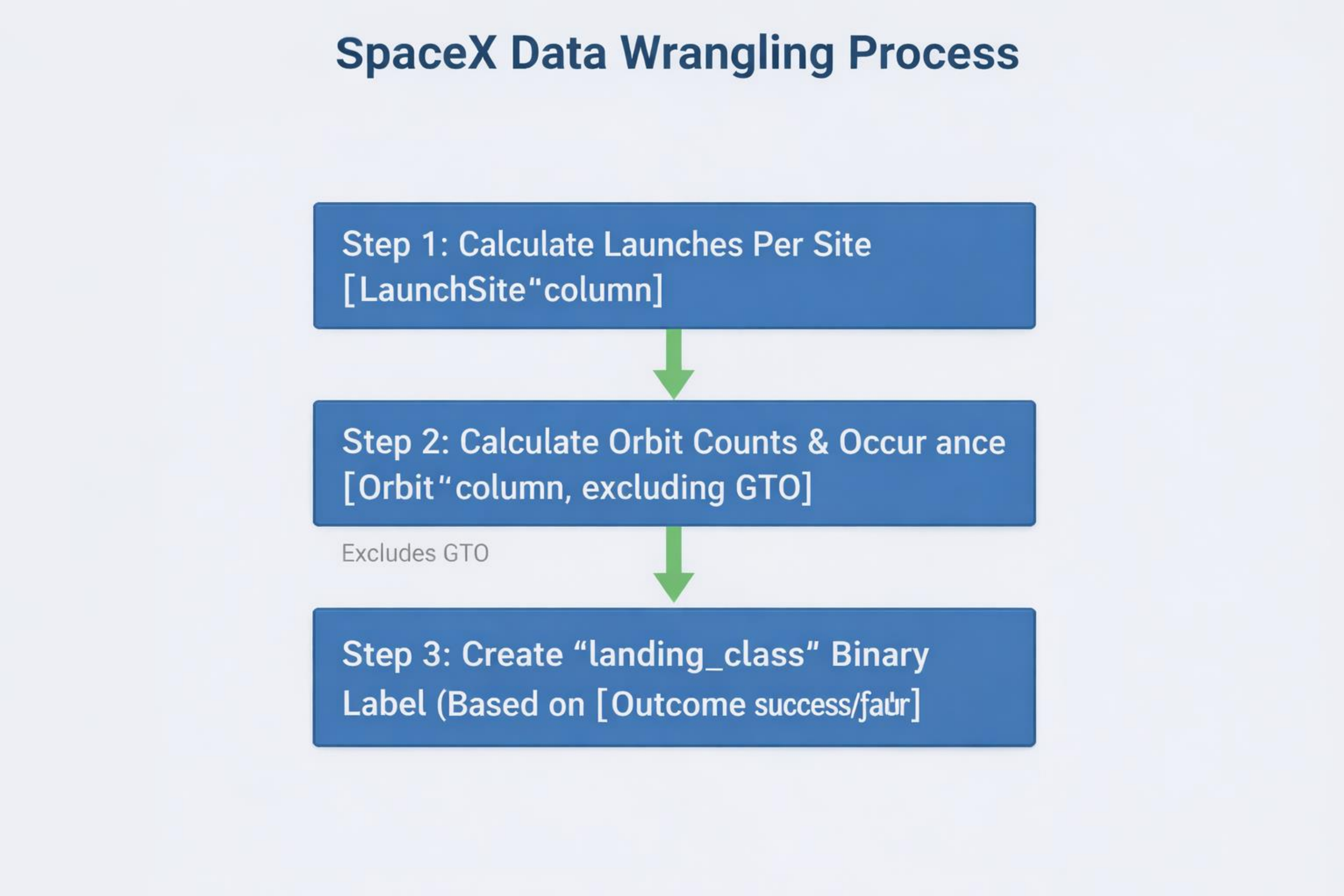
Data Collection – Web Scraping

- GitHub URL of the completed web scraping notebook.



Data Wrangling

- GitHub URL completed data wrangling notebook.



EDA with Data Visualization

- Using Scatter plots, we visualized the relation between the following features:
  - Flight number and payload mass
  - Flight number and launch site
  - Payload and launch site
  - Success rate and orbit type
  - Flight number and orbit type
  - Payload and orbit type
- Using a line graph, we showed the trend of success in launches over time.
  - Year vs success in launches over time.

GitHub URL of completed EDA with data visualization notebook.

EDA with SQL

- After storing the dataframe in a SQL DB, we extracted the following tables using SQL queries.

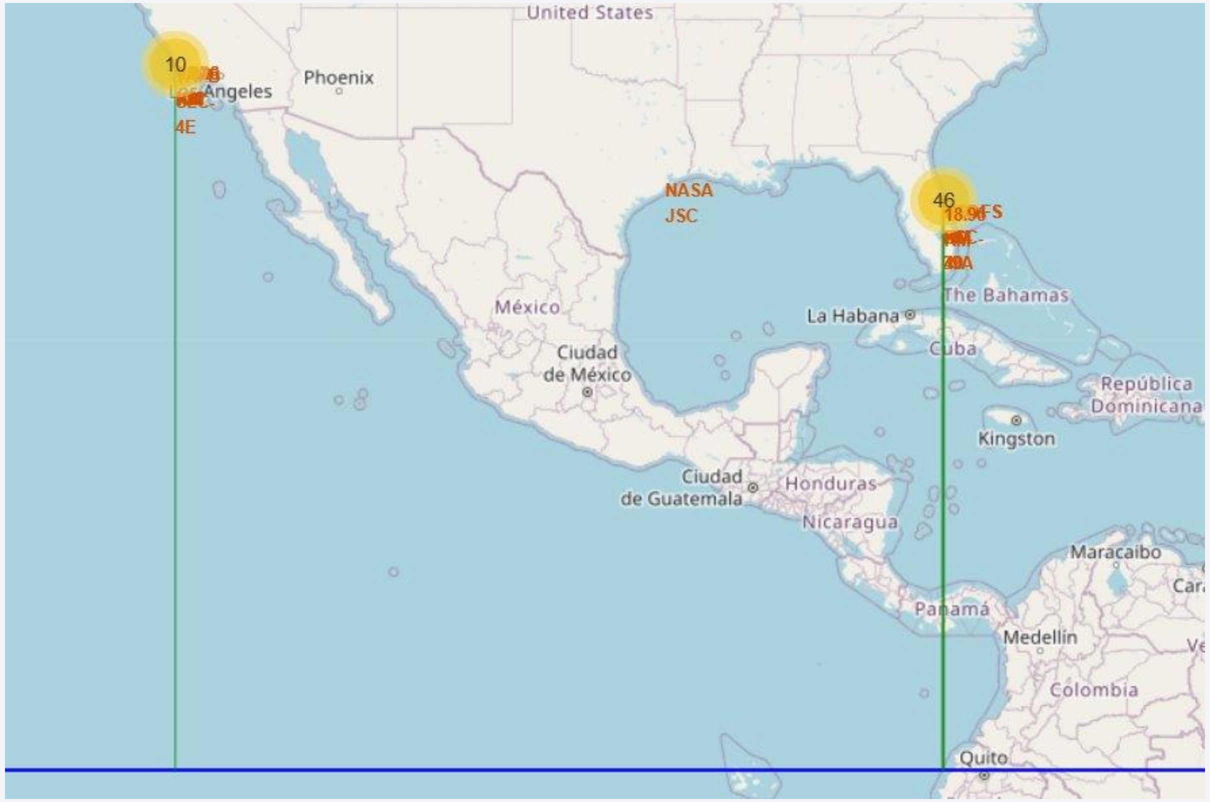
  - Names of the unique launch sites in the space mission
  - Launch sites begin with the string 'CCA' corresponding to the launch site CCAFS LC-40
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date when the first successful landing outcome in the ground pad was achieved.
  - Names of the boosters with successful landings in drone ship and payload mass greater than 4000 but less than 6000
  - Total number of successful and failed mission outcomes
  - Booster versions that have carried the maximum payload mass
  - Failure landing outcomes in drone ships with booster versions and launch sites in the year 2015.
  - Landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20.

- GitHub URL of your completed EDA with SQL notebook.

Build an Interactive Map with Folium

Using Folium, a powerful Python library, an interactive web map was created that:

- Marks the four launch sites using the function marker. Shows a line from the launch sites to the equator line to determine the influence of this proximity on the launches. The distance is displayed as a marker along the lines.
- A cluster of markers to display the success of each launch at each site.
- Lines to show the distance of launch sites to railways, highways, and coastlines using markers and lines functions.

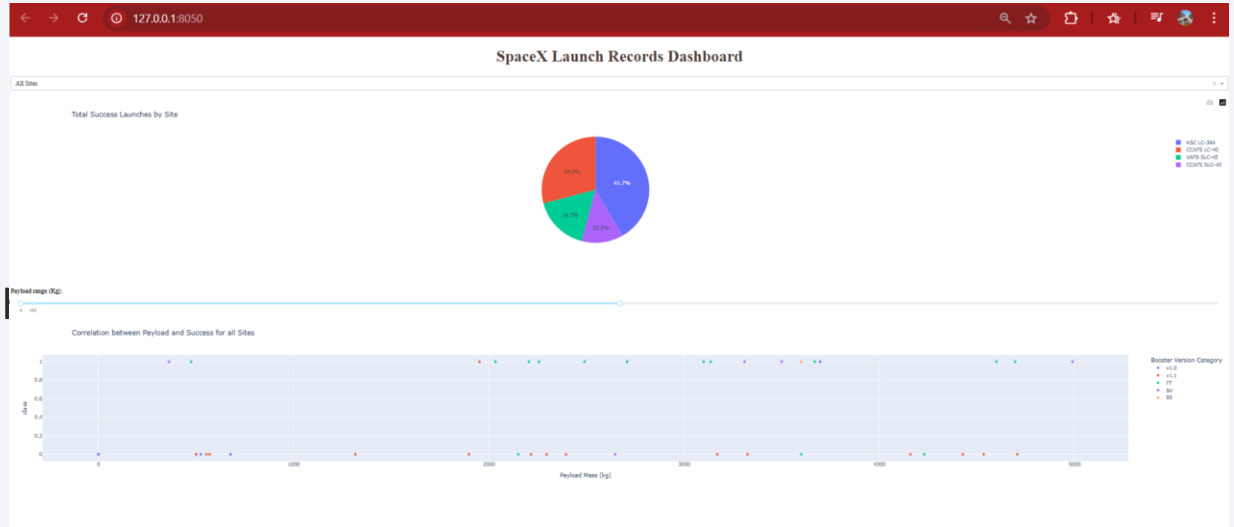


- GitHub URL completed interactive map with Folium map notebook.

Build a Dashboard with Plotly Dash

- Using Plotly Dash, a powerful open-source Python framework for building interactive web-based data applications and dashboards, a dashboard was created with the following graphics:

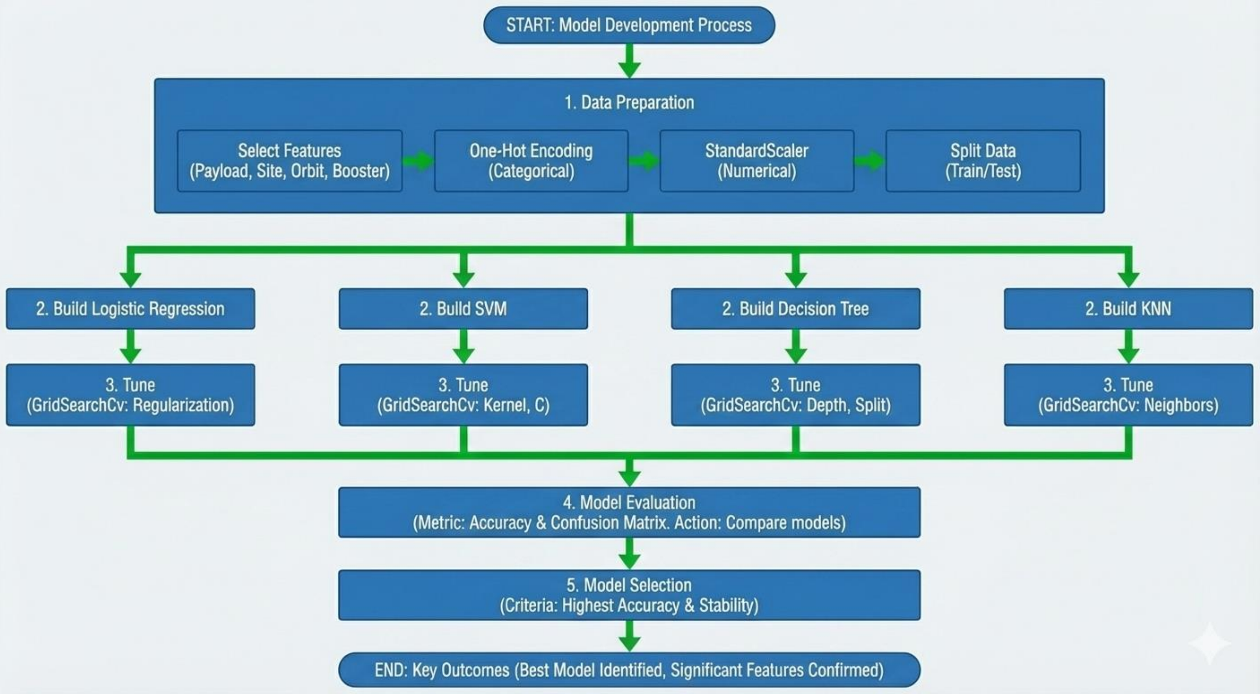
- A pie chart that displays the success rate for all and every one of the launch sites. This visualization allows us to quickly observe the first-stage landing success rate for each launch site.
- A scatter plot that displays the relation between Payload mass and success rate, colored by booster version. Using the payload range, we can determine the influence of the booster and the payload mass on the first-stage landing success rate.



- GitHub URL of the completed Plotly Dash application.

Predictive Analysis (Classification)

- Built classification models to predict Falcon 9 first-stage landing success
- Prepared data using feature encoding and standardization
- Split data into training (80%) and test (20%) sets
- Trained and tuned multiple models using GridSearchCV (cv = 10):
    - Logistic Regression
    - Support Vector Machine (SVM)
    - Decision Tree
    - K-Nearest Neighbors (KNN)
- Evaluated models using accuracy and confusion matrices
- Selected the best-performing model based on test accuracy and stability



GitHub URL of completed predictive analysis lab.

Results

- The first-stage landing success rate improves as the number of flights increases, with the most noticeable improvement observed at CCAFS SLC 40 (Cape Canaveral), which also hosted the highest number of launches.
- The overall landing success rate increased steadily from 2013 to 2017 (remaining relatively stable in 2014) and continued to improve after 2015, reflecting SpaceX's learning curve and operational maturity.
- Launches with payloads greater than 10,000 kg are primarily conducted from East Coast launch sites, and missions targeting LEO, ISS, and polar orbits show higher first-stage landing success rates.
- Missions to ES-L1, GEO, HEO, SSO, and VLEO orbits exhibit the highest first-stage landing success rates, indicating greater likelihood of successful booster recovery for these orbital profiles.
- As the total number of flights increases, VLEO appears to become a more frequently targeted orbit, suggesting a preference toward missions where booster recovery is more feasible.

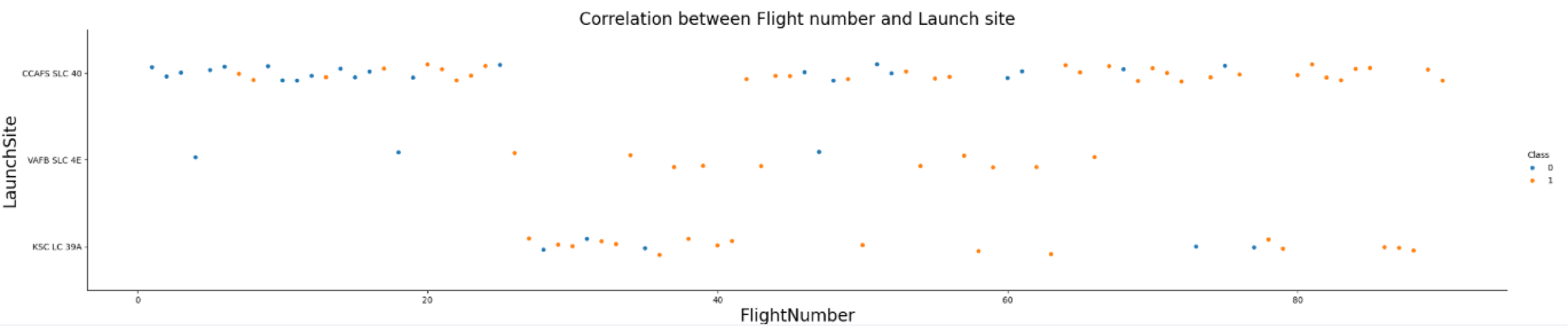
These exploratory insights guided the feature selection and modeling approach used in the predictive analysis phase.

Section 2

Insights drawn from EDA

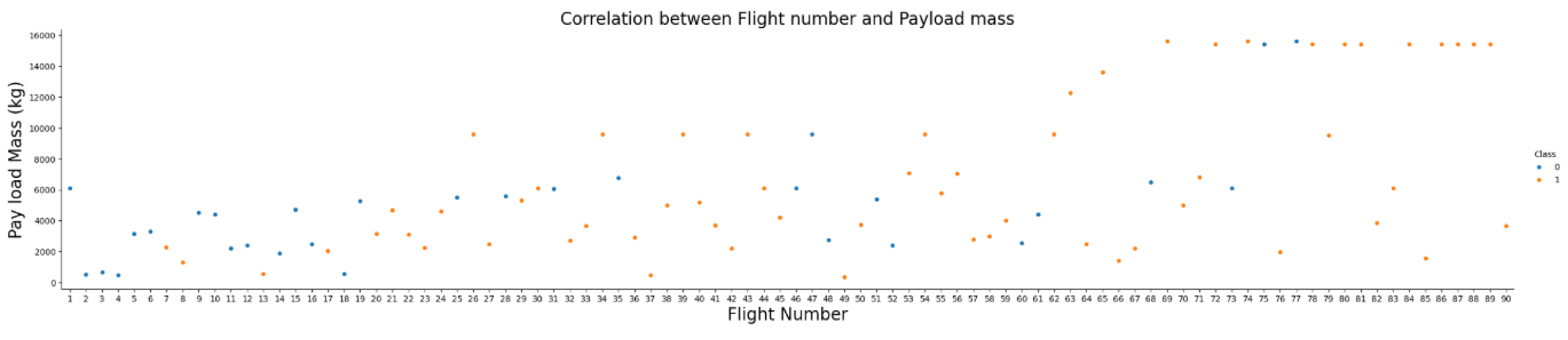
Flight Number vs. Launch Site

- Launches from CCAFS SLC 40, Cape Canaveral, are most frequent.
- The first-stage landing success rate increases as the flights increase, showing maturity in the process and noticeable improvements.
- Launches are mainly made from the East Coast (CCAFS SLC 40 and KSL LC-39).



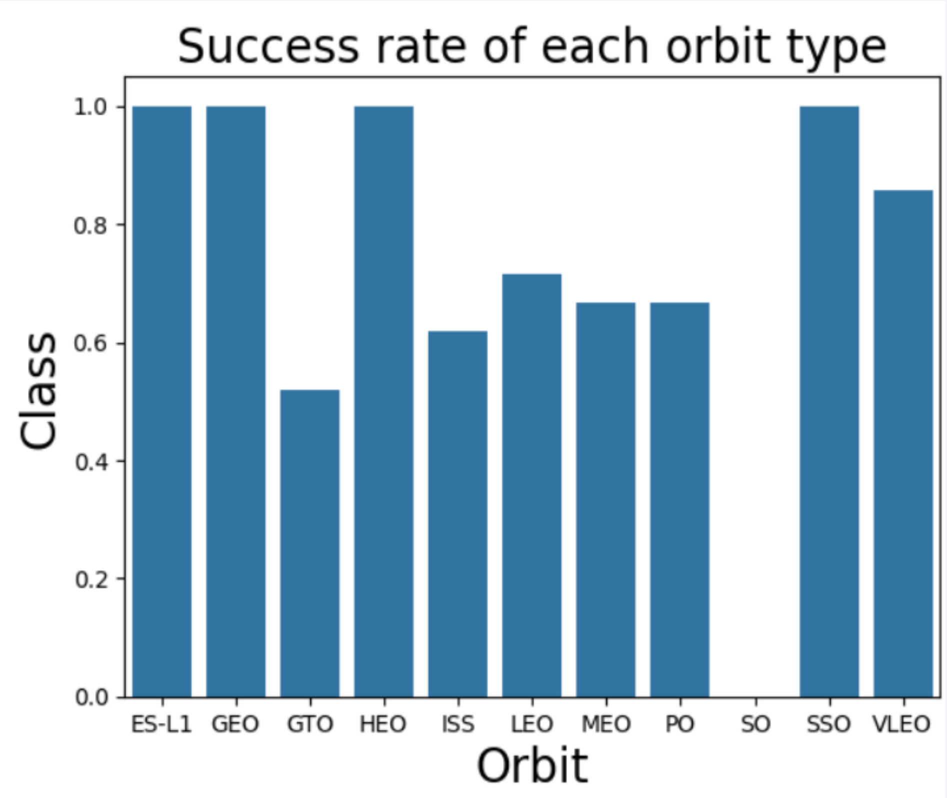
Payload vs. Launch Site

- We see that as the flight number increases, the first stage is more likely to land successfully.
- The latest flights seem to demonstrate the capacity to carry a heavier payload mass



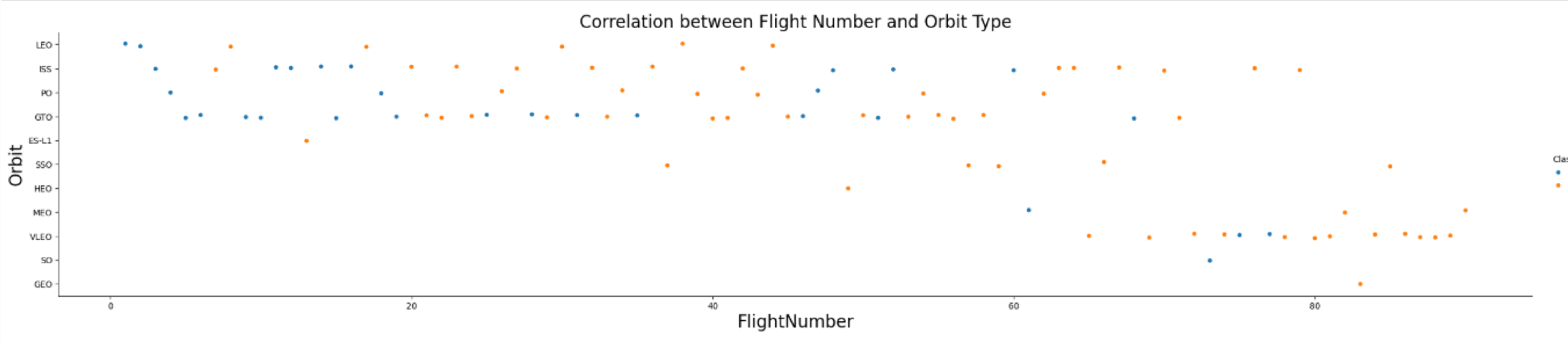
Success Rate vs. Orbit Type

- Missions to ES-L1, GEO, HEO, SSO, and VLEO orbits exhibit the highest first-stage landing success rates, indicating greater likelihood of successful booster recovery for these orbital profiles.



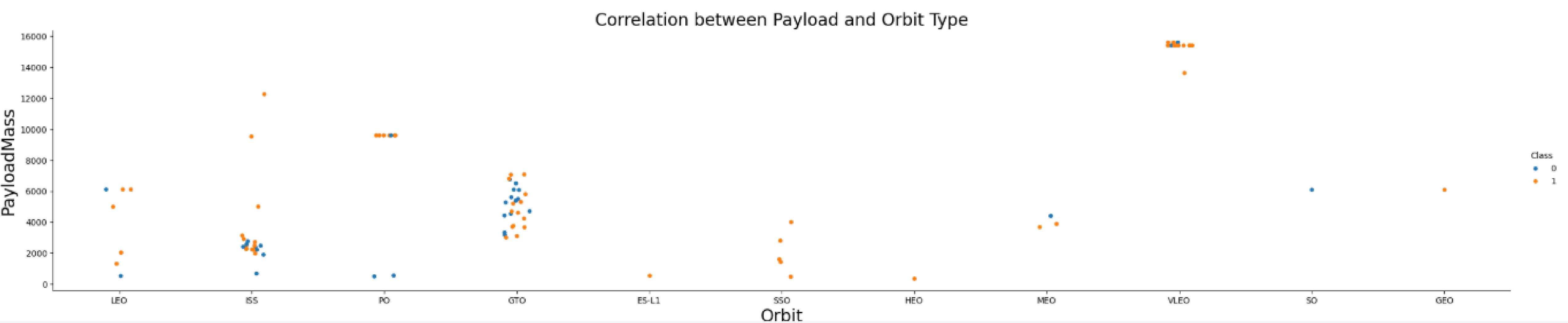
Flight Number vs. Orbit Type

- In the LEO orbit, the Success appears to be related to the number of flights
- As for the other Orbits, there doesn't seem to be a relation between Falcon 9 launches and their success in these orbits.
- LEO appears to be the preferred orbit for Falcon 9 launches in the latest flights.



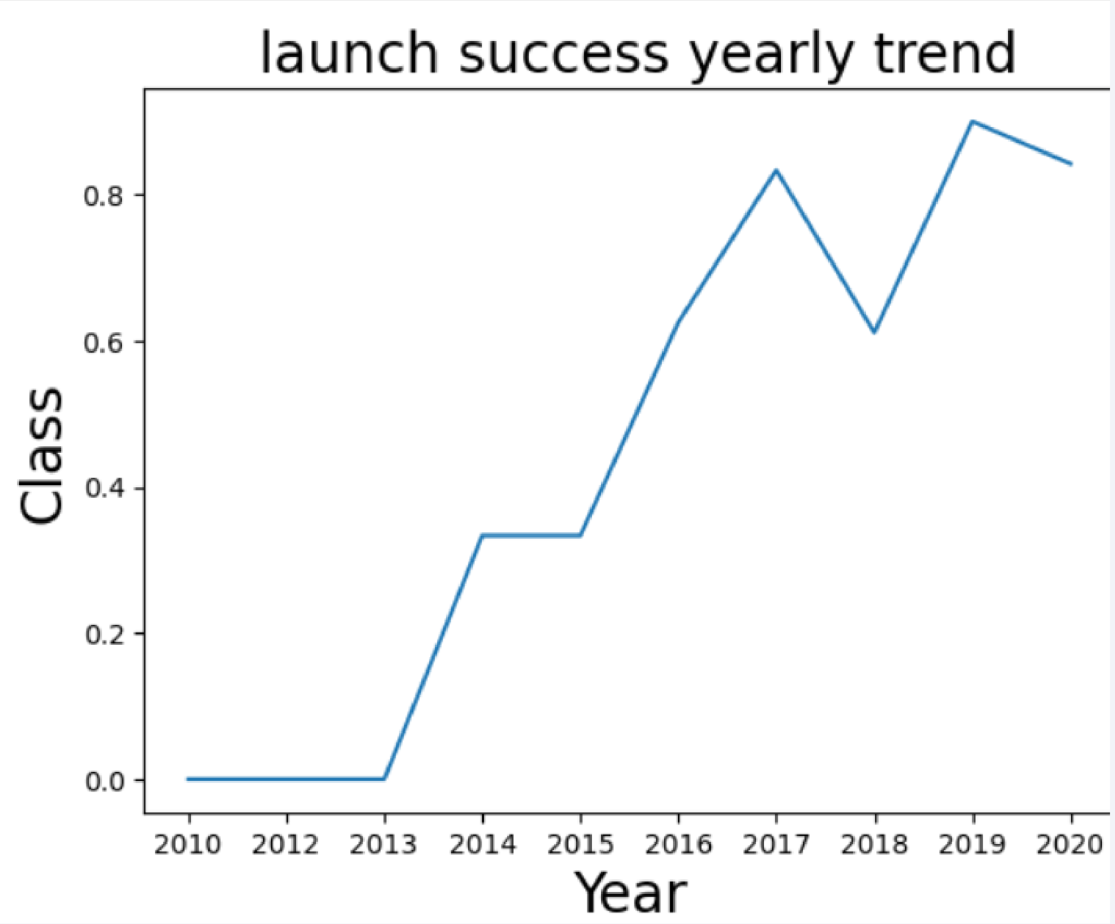
Payload vs. Orbit Type

- With heavy payloads of more than 10,000, the successful landing or positive landing rate is higher for Polar, LEO, and ISS orbits.
- VLEO is the orbit where the majority of Falcon 9 rockets have been sent with heavier payload masses.



Launch Success Yearly Trend

- The success rate since 2013 has kept increasing till 2017 (stable in 2014), and after 2015 it started increasing, showing the maturity of the company with the launches.



All Launch Site Names

- List of the four launch sites. The launch site, KSC LC-39A, is the only one on the West Coast.

```
%sql Select DISTINCT Launch_Site from SPACEXTABLE;
```

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

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

Launch Site Names Begin with 'CCA'

- Sample of records for the launch site CCAFSLC-40.

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

```
%sql Select * from SPACEXTABLE 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

- 48,213 kg is the total payload mass carried by NASA boosters.

[14]:
```
%sql Select sum(cast("PAYLOAD_MASS__KG_" as float)) as suma from SPACEXTABLE where Customer like '%NASA (CRS)%';
```

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

[14]:

suma
48213.0

Average Payload Mass by F9 v1.1

- 2534,6 kg is the average payload mass carried by booster version Falcon 9, F9 v1.1.

Task 4 ¶

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

15]:
```
%sql Select AVG(cast("PAYLOAD_MASS__KG_" as float)) as AVG from SPACEXTABLE where Booster_Version like '%F9 v1.1%';
```

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

15]:

AVG
2534.6666666666665

First Successful Ground Landing Date

- The first time a Falcon 9 landed on a ground pad was on December 22, 2015, at 01:29 UTC (Universal Coordinated Time). The launch occurred on the evening of December 21st local time in Florida.

```
[16]: %sql Select Date,Mission_Outcome,Landing_Outcome from SPACEXTABLE where Mission_Outcome=='Success' and Landing_Outcome like '%ground%' limit 1;
```

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

[16]:

Date	Mission_Outcome	Landing_Outcome
2015-12-22	Success	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- The query identifies Falcon 9 boosters (FT variants) that successfully landed on drone ships while carrying payloads between 4,000 and 6,000 kg.
  The results show that multiple boosters were able to achieve successful drone ship landings under medium–heavy payload conditions, highlighting SpaceX's capability to recover first-stage boosters even with substantial payload masses.

```
•[17]: %sql Select DISTINCT Booster_Version,Mission_Outcome,Landing_Outcome,PAYLOAD_MASS__KG_
from SPACEXTABLE where (PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000) and(Landing_Outcome like '%Success (drone ship)%');
```

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

[17]:

Booster_Version	Mission_Outcome	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success	Success (drone ship)	4696
F9 FT B1026	Success	Success (drone ship)	4600
F9 FT B1021.2	Success	Success (drone ship)	5300
F9 FT B1031.2	Success	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

- 101 Falcon 9 missions were analyzed, including both successful and failed mission outcomes.
- This dataset provides a sufficient historical basis to evaluate mission reliability and landing performance trends.

Task 7

List the total number of successful and failure mission outcomes

[18]:
```
%sql Select count(Mission_Outcome) from SPACEXTABLE where Mission_Outcome like'Success%' or Mission_Outcome like'Failure%';
```

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

[18]:

count(Mission_Outcome)
101

Boosters Carried Maximum Payload

- Multiple Falcon 9 Block 5 boosters achieved the maximum recorded payload mass of 15,600 kg. This indicates that SpaceX reused several high-performance boosters for the heaviest missions, highlighting the reliability and operational capability of the Block 5 configuration.

```
%sql select Booster_Version,PAYLOAD_MASS__KG_ from SPACEXTABLE where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- SpaceX conducted the first two attempts to land Falcon 9 v1.1 boosters on an Autonomous Spaceport Drone Ship (ASDS) in 2015. Both failed, yet provided critical data for reusability development:
- January 10, 2015 (CRS-5 Mission, Booster B1012):
  - Hard landing and explosion on the Just Read the Instructions drone ship.
  - Issue: Ran out of hydraulic fluid for the grid fins, resulting in loss of control and excessive lateral velocity
- April 14, 2015 (CRS-6 Mission, Booster B1015):
  - Hard landing and explosion on the same drone ship.
  - Issue: An engine throttle valve failed to respond fast enough during the landing burn maneuver.

```
.]: %sql select substr(Date,6,2) as Month,substr(Date,0,5) as Year, Booster_Version,Landing_Outcome,Launch_Site
from SPACEXTABLE where substr(Date,0,5)='2015' and Landing_Outcome like '%Failure%'
```

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

.]:

Month	Year	Booster_Version	Landing_Outcome	Launch_Site
01	2015	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
04	2015	F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40

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

- In 2012, the recovery of the booster was not an option.
- We can see the first successful booster recovery in April 2016 on a drone ship.
- In December 2015, we saw the first successful booster recovery on ground pad
- 3 years of missions before the first booster success recovery.

Task 10

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.

[22]:
```
%sql select Date, Landing_Outcome, count(*) as frequency
from SPACEXTABLE where substr(Date,0,10) >'2010-06-04' and substr(Date,0,10) <'2017-03-20' group by Landing_Outcome order by frequency desc
```

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

[22]:

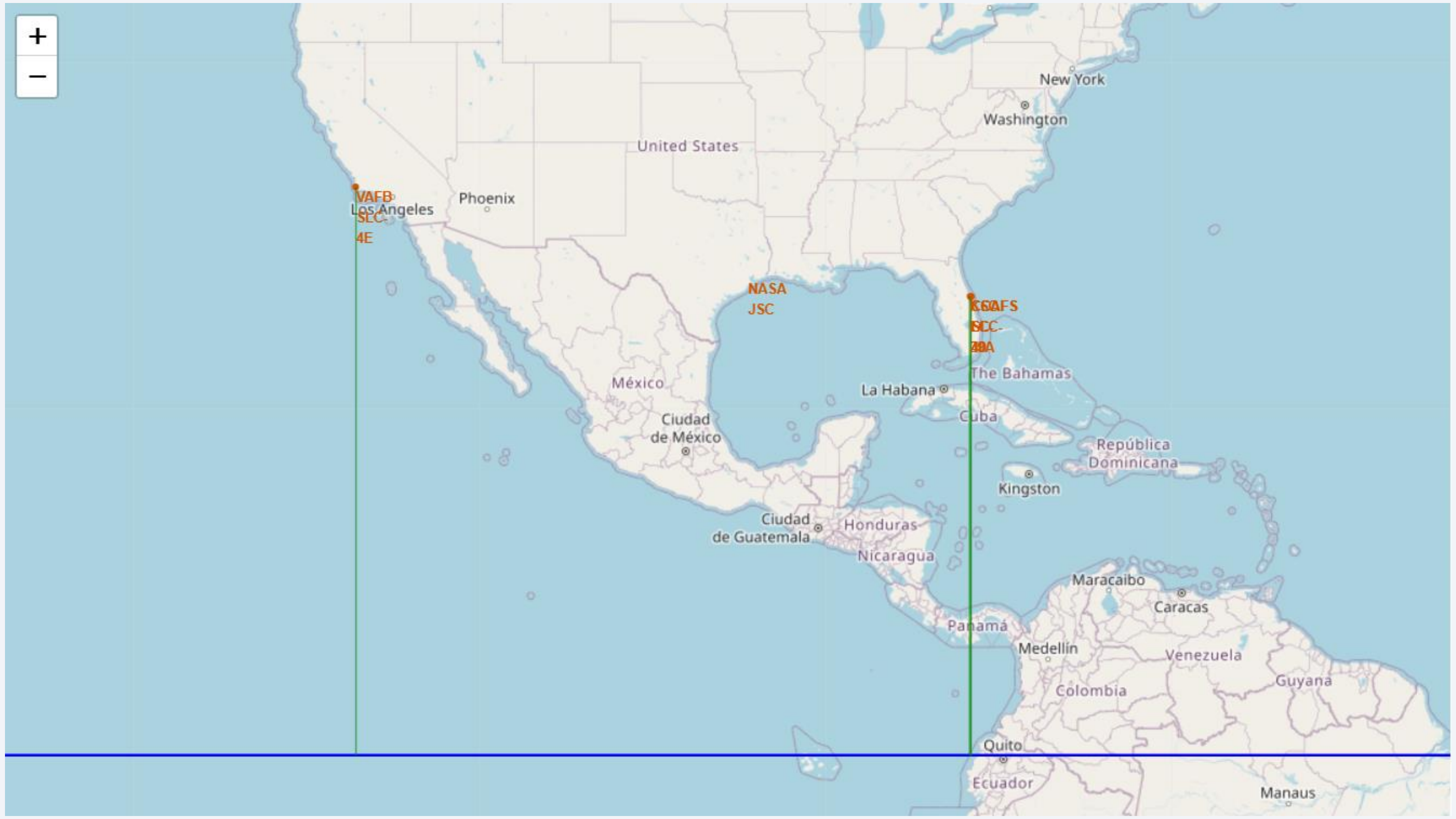
Date	Landing_Outcome	frequency
2012-05-22	No attempt	10
2016-04-08	Success (drone ship)	5
2015-01-10	Failure (drone ship)	5
2015-12-22	Success (ground pad)	3
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2015-06-28	Precluded (drone ship)	1
2010-12-08	Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

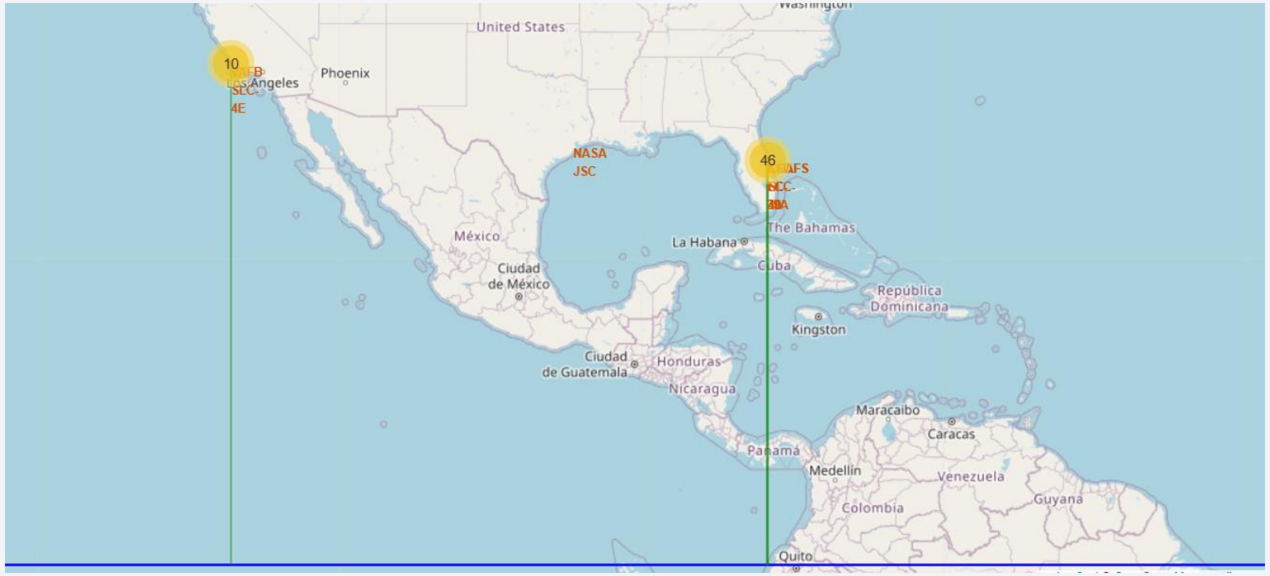
<Space X Launch Sites>

- We can explore on a map the location of the launch sites

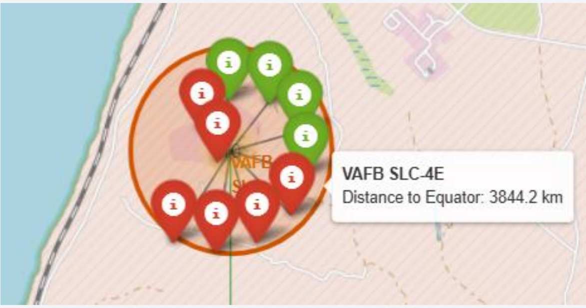


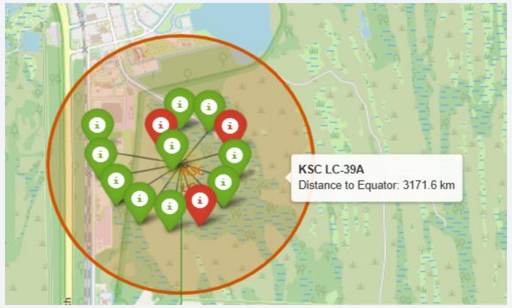
- We can see the VAFB SLC-4E launch site in the west.
- CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A on the Florida coast in the east.
- The blue line shows the equator line.
- Green lines show the distance of the launch sites to the equator line.

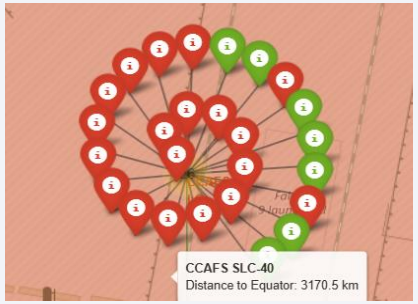
Number of Launches

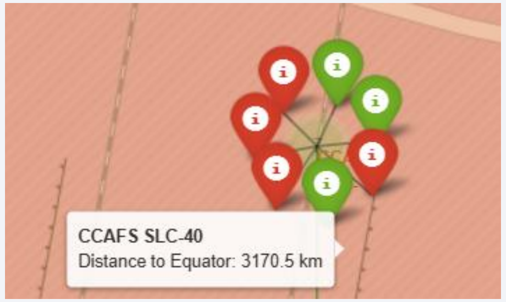


- 10 launches were made from the launch site VAFB SLC-4E
- 46 from the launch sites CCAFS LC-40,CCAFS SLC-40,and KSC LC-39A
- We can visually conclude that the launch site KSC LC-39A has the highest success rate

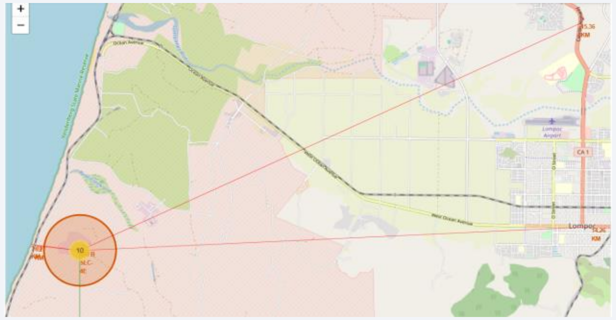


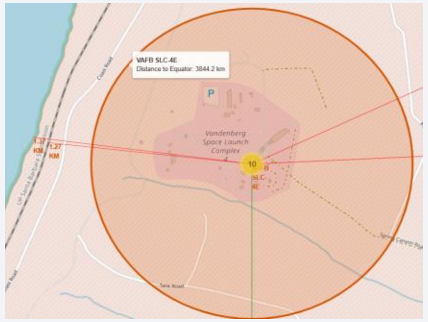






Proximities of Launch Sites





VAFB SLC-4E is approximately 1.2 km from the nearest railway.

CCAFS LC-40 is less than 1 km (0.9 km) from the nearest railway. This proximity facilitates the transportation of heavy equipment and rocket components.

VAFB SLC-4E is approximately 15.3 km from the closest highway.

CCAFS LC-40 is approximately 0.6 km from the closest highway. This allows efficient ground access and logistical support.

The distance to the closest coastline is approximately 1.3 km for the west coast site (VAFB).

The distance is approximately 0.9 km for the east coast site (CCAFS). This proximity is likely due to safety reasons, allowing rocket stages and debris to fall into the ocean.

VAFB SLC-4E is approximately 14.2 km from the closest city, Lompoc.

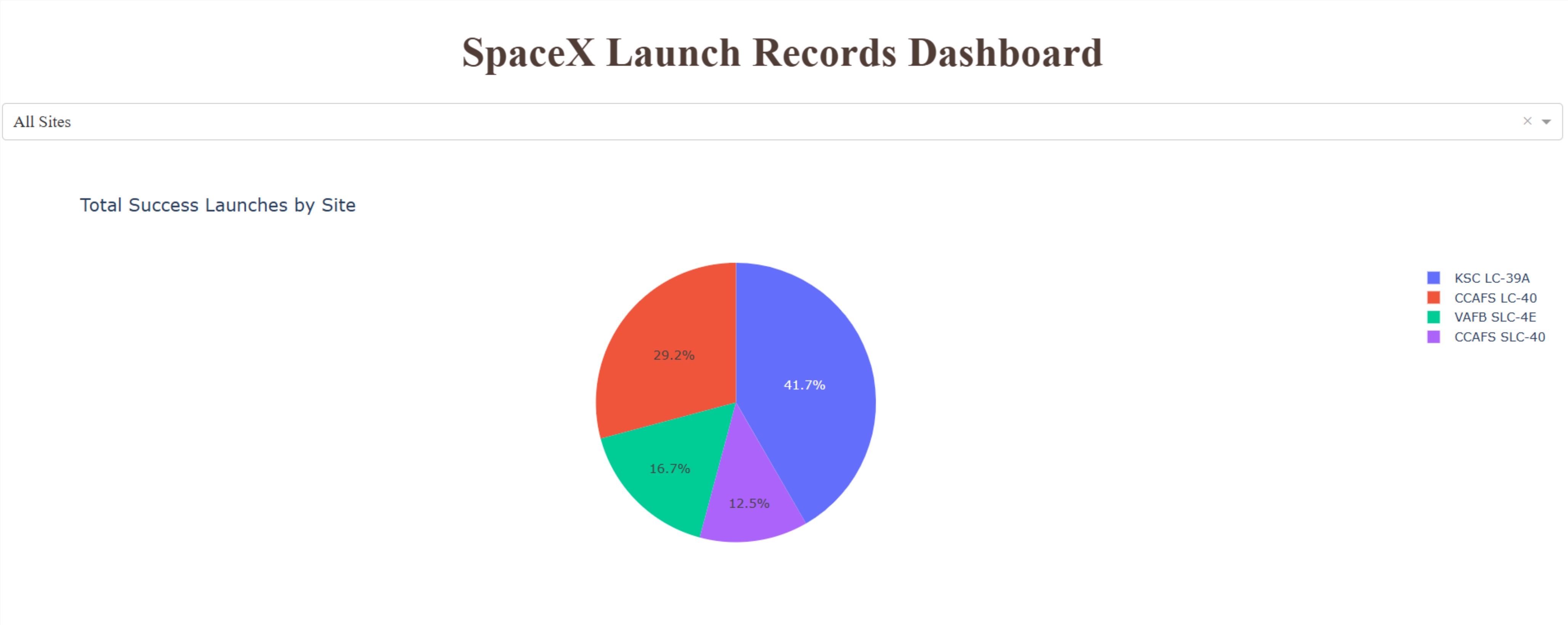
CCAFS LC-40 is approximately 18.9 km from the closest city, Cape Canaveral. This helps minimize risks to populated areas in case of launch failures.

Section 4

Build a Dashboard with Plotly Dash

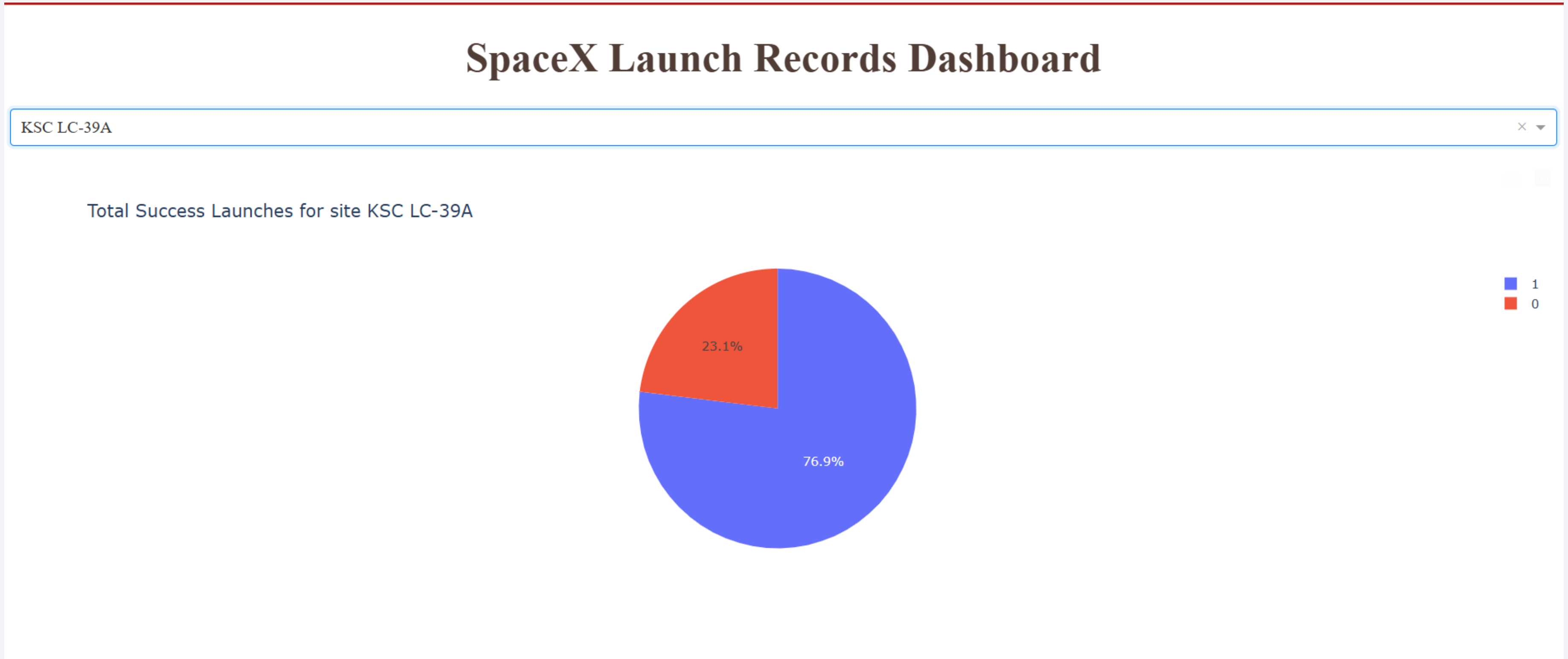
First-Stage Landing Success Rate per Launch Site

- Over 80% of launches originated from CCAFS LC-40, KSC LC-39A, and CCAFS SLC-40, making them SpaceX's most critical launch sites.

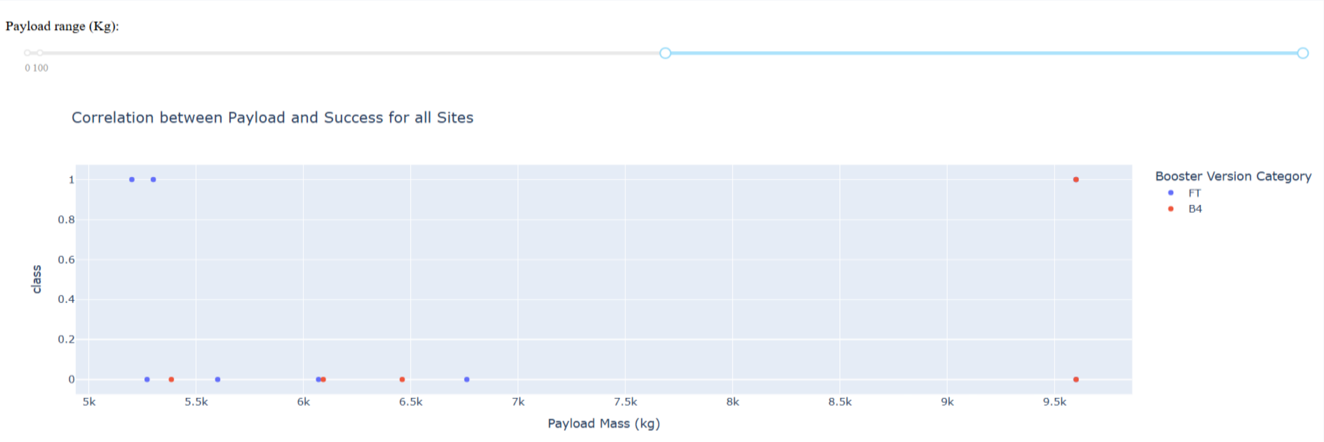


Importance of Launch Site KSC LC-39A

- First-stage landing success rate for the launch site KSC LC-39A(Kennedy Space Center in Florida) is more than 76 percent.
- Demonstrating strategic importance for SpaceX in reusing boosters to reduce mission costs.



Falcon 9 Booster Versions and Success Rate



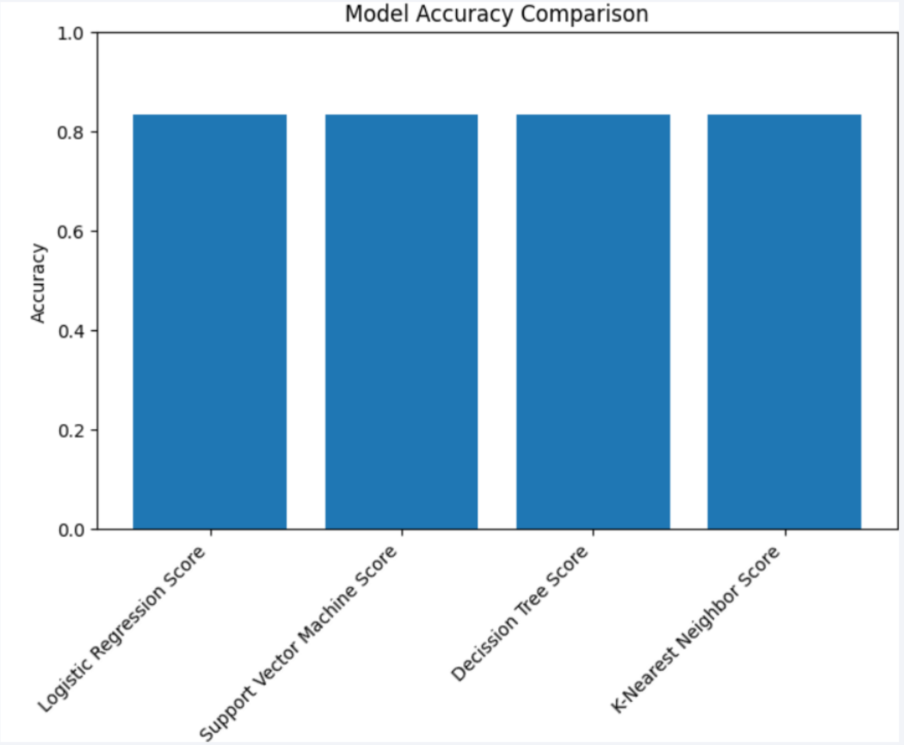
- For payloads heavier than 5000 km, the number of launches is smaller as well as the success rate.
- We can see that SpaceX uses for these payloads the recent versions of Falcon9 model.

- From booster version FT, we can observe a strong landing success rate for payloads between 1000 and 5000 kg. Older versions (v1.0 and v1.1) have a very low landing success rate.

Section 5

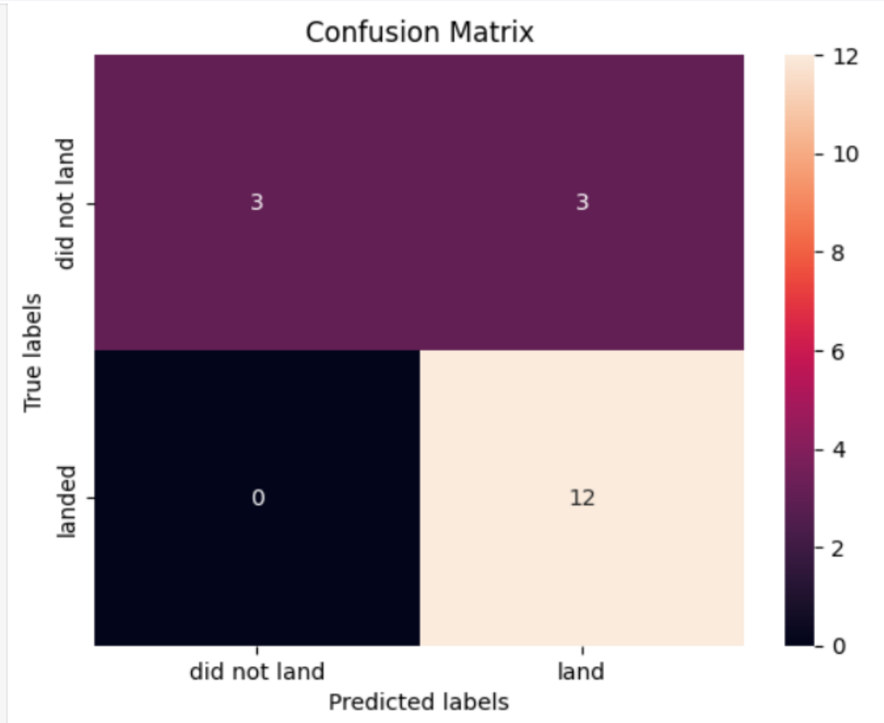
Predictive Analysis (Classification)

Classification Accuracy



- All ML models(Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors) achieved a similar accuracy of 83%.
- There doesn't seem to be major differences in the accuracy of the models, most likely due to the size of the data sample used for the prediction analysis.
- However, the results indicate that the dataset with historical Falcon 9 launch records can reasonably predict first-stage landing outcomes, providing useful insights to determine booster reusability and potential cost reduction in future missions.

Confusion Matrix



- Using SVM, K-means, and Logistic Regression algorithms, we could correctly predict 12 true positives: 12 success launches and 3 true negatives, 3 unsuccessful launches.
- The models(whose score was 83%)also fail in predict 3 unsuccessful launches.(false positives)

Conclusions

- This Applied Data Science Capstone Project successfully predicted first-stage landing outcomes of SpaceX Falcon 9 boosters, highlighting SpaceX's operational maturity and explaining how booster reusability contributes to cost reduction in space missions. The project also helps demonstrate how booster reusability has positioned SpaceX as the leader in the modern space race.
- Orbit, Payload mass, booster version, and Launch site were identified as key factors influencing successful first-stage landing outcomes.
- Interactive visualizations using Plotly or Folium amplify the context of the influence of geographical locations and factors such as payload mass on first-stage landing success.

Appendix

- Selected SQL queries used for mission outcome and payload analysis
- Python code snippets for data preprocessing and model tuning
- Confusion matrix outputs for selected classification models
- Screenshots of interactive Dash and Folium visualizations
- External data sources: SpaceX API and Wikipedia launch records
    - https://api.spacexdata.com/v4/rockets/
    - https://api.spacexdata.com/v4/launchpads/
    - https://api.spacexdata.com/v4/payloads/
    - https://api.spacexdata.com/v4/cores/
    - https://api.spacexdata.com/v4/launches/past
    - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Thank you!