IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ken Ely
27 October 2025

Outline

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

Executive Summary

- Collected and cleaned Falcon 9 launch data (sourced from API and Wikipedia)
- Conducted EDA using Python, SQL, and interactive dashboards
- Built classification models (LR, SVM, Decision Tree, KNN) with hyperparameter tuning
- Successfully identified parameters that influenced landing success
- Developed machine learning models that could predict landing success from those parameters

Introduction

- SpaceX's Falcon 9 launches cost $62M, much cheaper than competitors ($165M+)
- Savings come from reusing the first stage of the rocket
- Thus, predicting first-stage landing success helps estimate launch costs
- **Can we predict whether the Falcon 9 first stage will land successfully?**

Section 1

Methodology

Methodology

Executive Summary

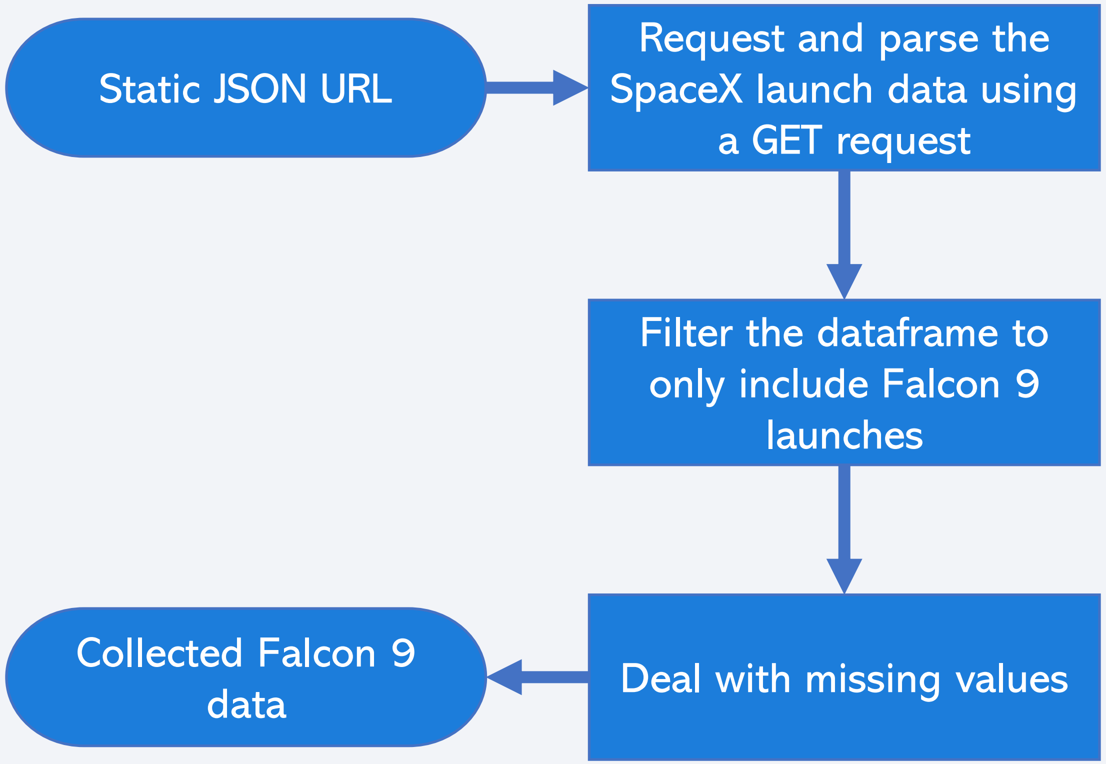
- Data collection methodology
  - Data gathered from SpaceX REST API
- Perform data wrangling
  - Landing outcomes converted to classes y (either 0 or 1)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data preprocessed and split into training and testing sets
  - Grid search used to find hyperparameters for each model
  - Confusion matrices used to evaluate models

Data Collection

- Two means of data collection were explored for this project:
  - Gathering data from the SpaceX REST API (i.e., Representational State Transfer Application Programming Interface)
  - Webscraping the “List of Falcon 9 and Falcon Heavy launches” Wikipedia page

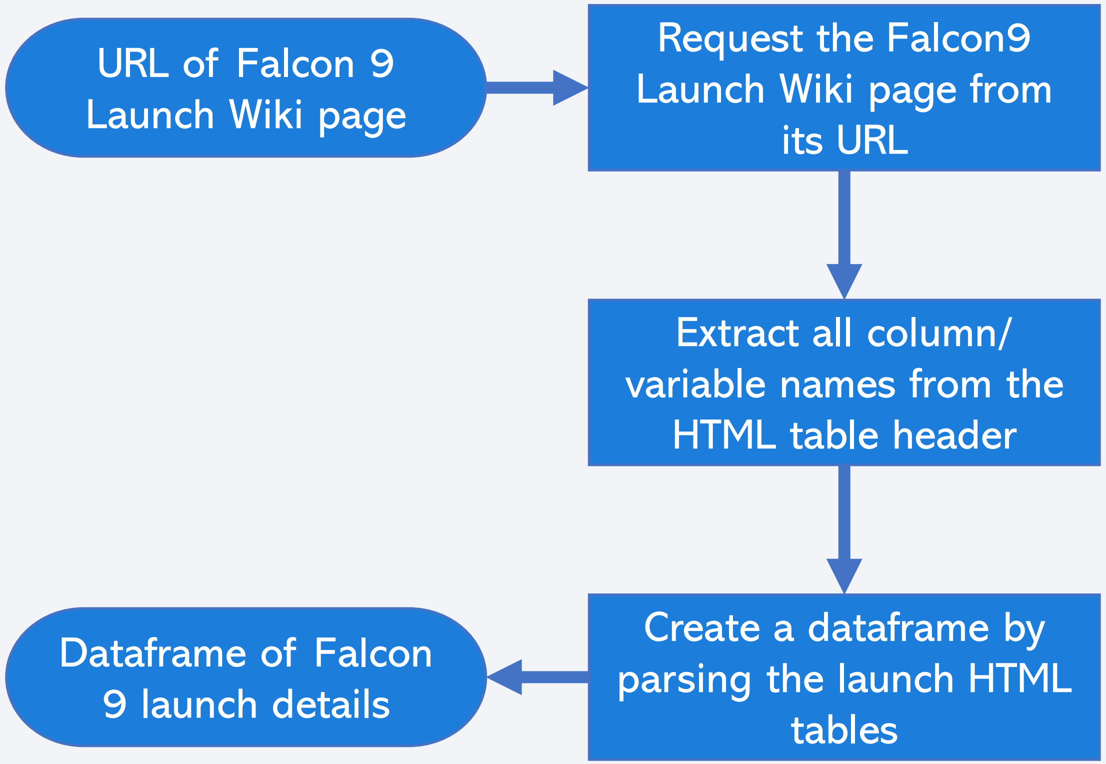
Data Collection – SpaceX API

- Data obtained from the SpaceX REST API using a GET request to https://api.spacexdata.com/v4/launches/past
- Since there was data from both Falcon 1 and Falcon 9 launches, the data had to be filtered
- Link to completed SpaceX API calls notebook on GitHub



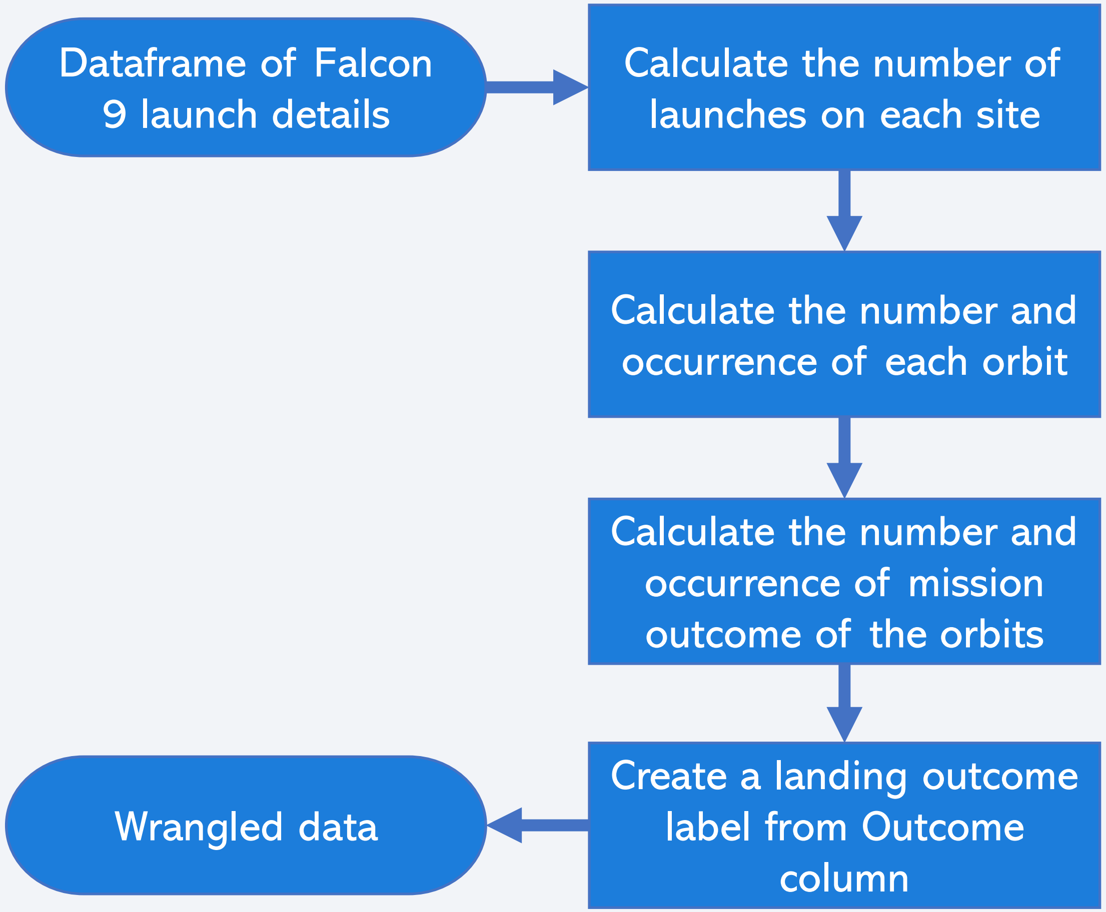
Data Collection - Scraping

- "BeautifulSoup" Python library used to scrape the "List of Falcon 9 and Falcon Heavy launches" Wikipedia page
- Table containing launch records was identified, and was converted into a Pandas dataframe
- Link to completed web scraping notebook on GitHub



Data Wrangling

- Using the data collected earlier and the .value_counts() method, the following were evaluated:
  - Number of launches on each site
  - Occurrence of each orbit type
  - Occurrence of each mission outcome
- An additional variable “Class” was appended to the dataframe, which indicates whether the first stage of a particular launch landed successfully
- [Link] to completed data wrangling notebook on GitHub



EDA with Data Visualization

- Charts plotted:
    - Flight Number vs Launch Site – shows launch site used for each flight
    - Payload vs Launch Site – shows masses of payloads launched from each launch site
    - Success Rate vs Orbit Type – shows how successful flights are by the orbit they used
    - Flight Number vs Orbit Type – shows how success rates with different orbit types have changed with time
    - Payload vs Orbit Type – shows how successful each orbit rate is with different masses of payloads
    - Launch Success Yearly Trend – shows the average launch success trend
- Link to completed EDA with data visualization notebook on GitHub

EDA with SQL

- Queries performed:
  - Names of the unique launch sites
  - Records for CCAFS launch sites
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date of first successful landing outcome in ground pad
  - Boosters which have success in drone ship and have payload mass between 4-6 tonnes
  - Total number of successful and failure mission outcomes
  - Booster versions which have carried the maximum payload mass
  - Failed landings in 2015
  - Count of landing outcomes between 4 June 2010 and 20 March 2017
- Link to completed EDA with EDA with SQL notebook on GitHub

Build an Interactive Map with Folium

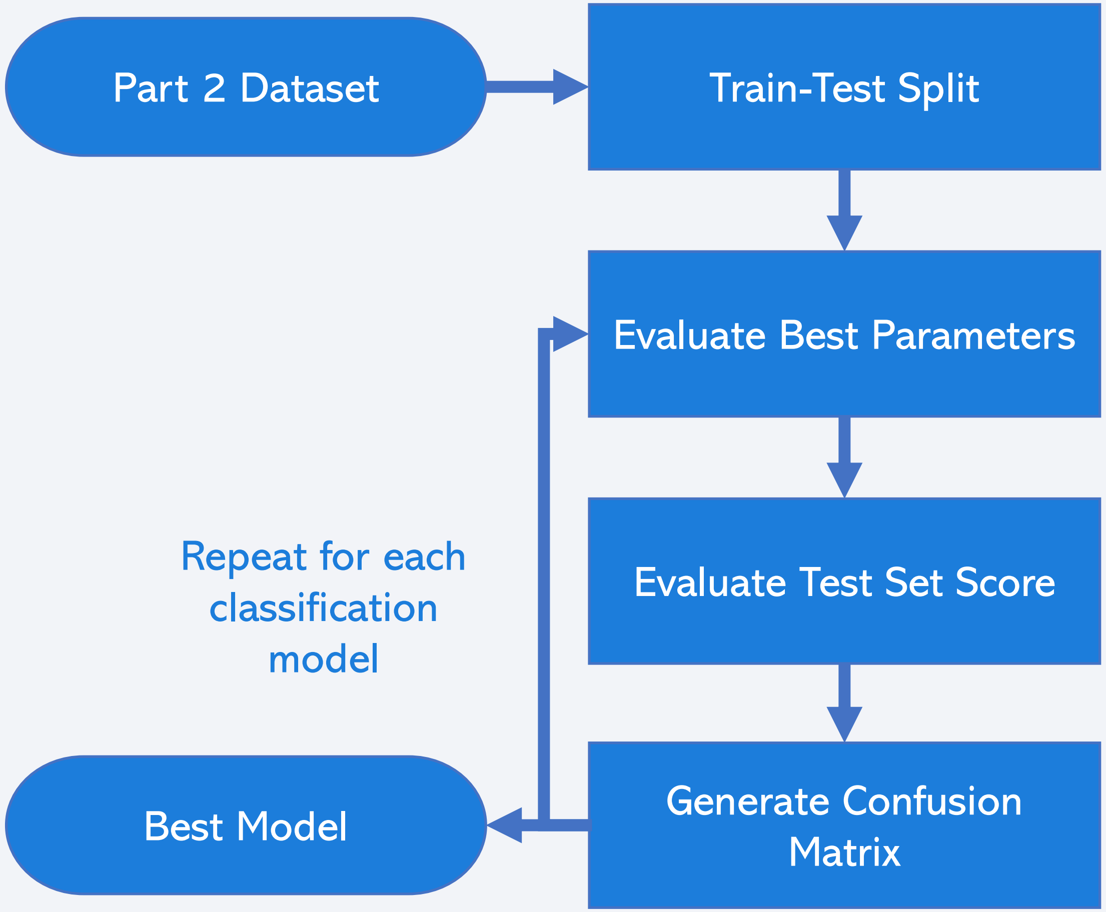
- Added circles to indicate locations of launch sites
- Added colour-coded markers to indicate individual rocket launches
- Added lines to show proximity of launch site CCAFS SLC-40 to its nearest coastline, railway, highway, and city
- Link to completed interactive map notebook on GitHub

Build a Dashboard with Plotly Dash

- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose
- Added a pie chart which shows the ratio of successful to failed launches from a specified launch site
  - Choosing "All Sites" will show the number of successful launches by site
  - Allows the success rates of different launch sites to be compared
- Added a scatter plot which shows the correlation between payload and success for all sites
  - Range slider above it allows the user to narrow down the range of payloads they want to view
  - Dots colour-coded based on booster version used on a particular launch
  - Allows the success rates of different payload ranges and also booster versions to be compared
- Link to completed Plotly Dash lab on GitHub

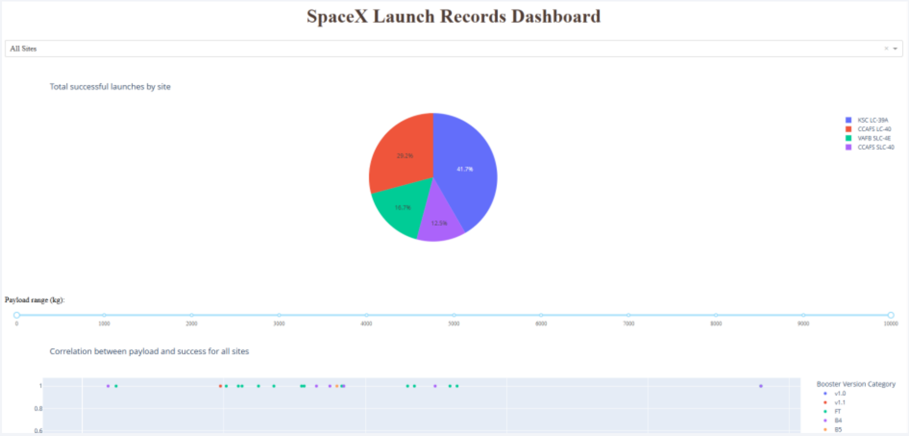
Predictive Analysis (Classification)

- Data is first preprocessed and split into a "train" set and a "test" set
- Logistic Regression, Support Vector Machine, Decision Tree Classifier, and k-Nearest Neighbours classification models were investigated
- For each model:
  - The best parameters for the model are evaluated using GridSearchCV
  - The score method is used to calculate the accuracy of the model on the test data
  - A confusion matrix is generated to allow the effectiveness of the model to be evaluated
- Link to completed predictive analysis lab on GitHub

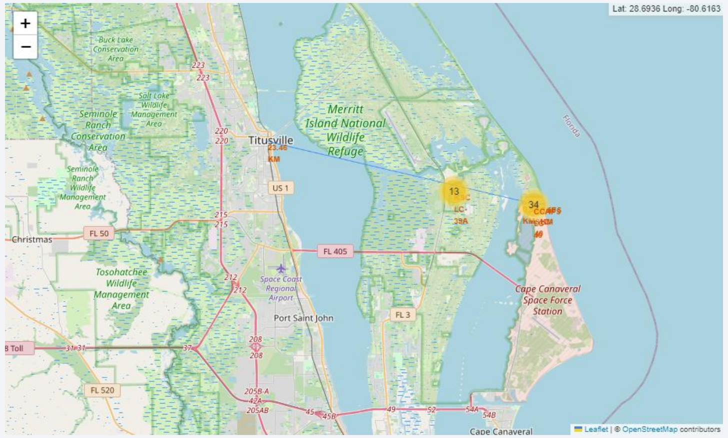


Results

- Exploratory Data Analysis (EDA)
  - Launch success highest at KSC LC-39A (76.9%)
  - Success rate increased steadily from 2013–2017
  - Optimal payload range: 2–4 tonnes
  - Best orbits: ES-L1, GEO, HEO, SSO (100% success)



- Interactive Analytics (Dash & Folium)
  - Folium map: Visualized launch sites, success/failure by location
  - Plotly Dash dashboard: Compared success rates by site, payload, and booster version



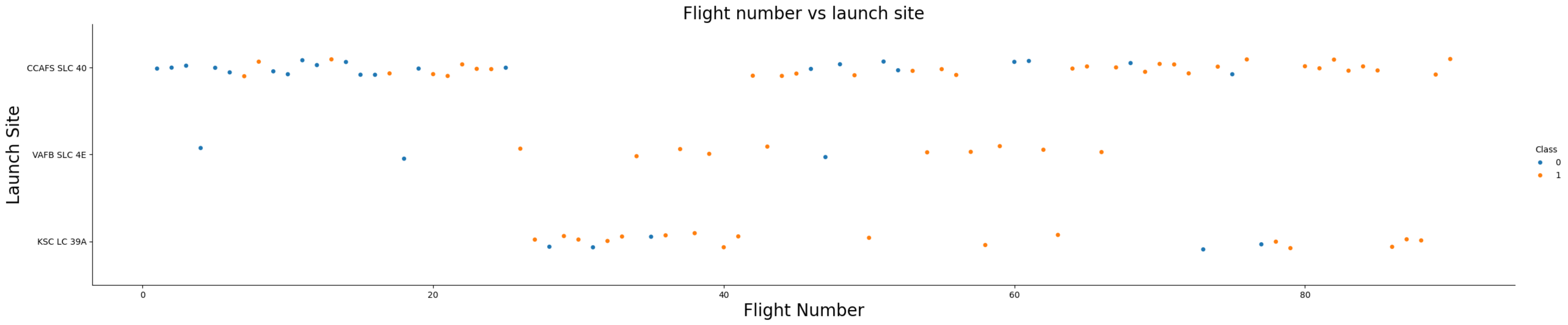
- Predictive Analysis
  - Tested LR, SVM, Decision Tree, KNN models
  - SVM and Logistic Regression achieved highest accuracy
  - Models effectively predicted landing success using launch features

Section 2

Insights drawn from EDA

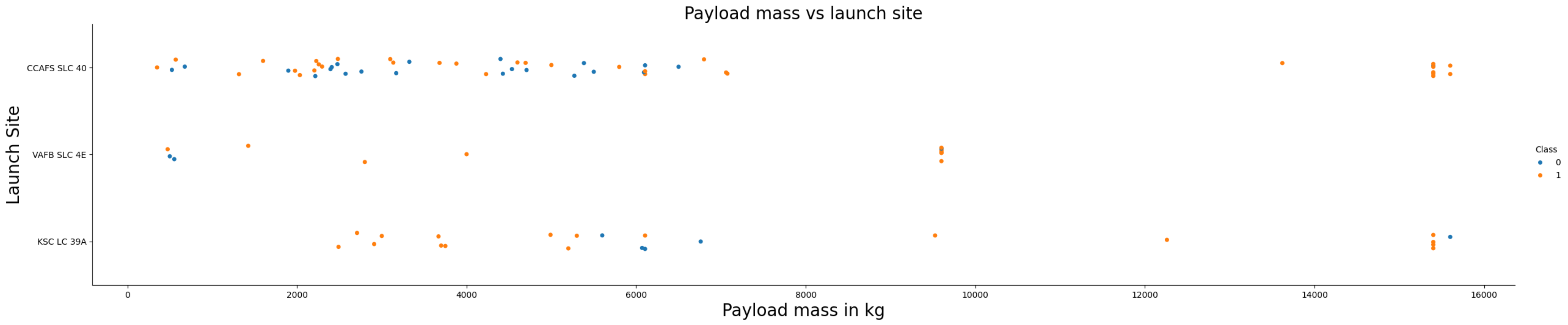
Flight Number vs. Launch Site

- CCAFS SLC 40 is the most used launch site
- For flights 27-41, KSC LC 49A temporarily became the main launch site used
- VAFB SLC 4E used occasionally



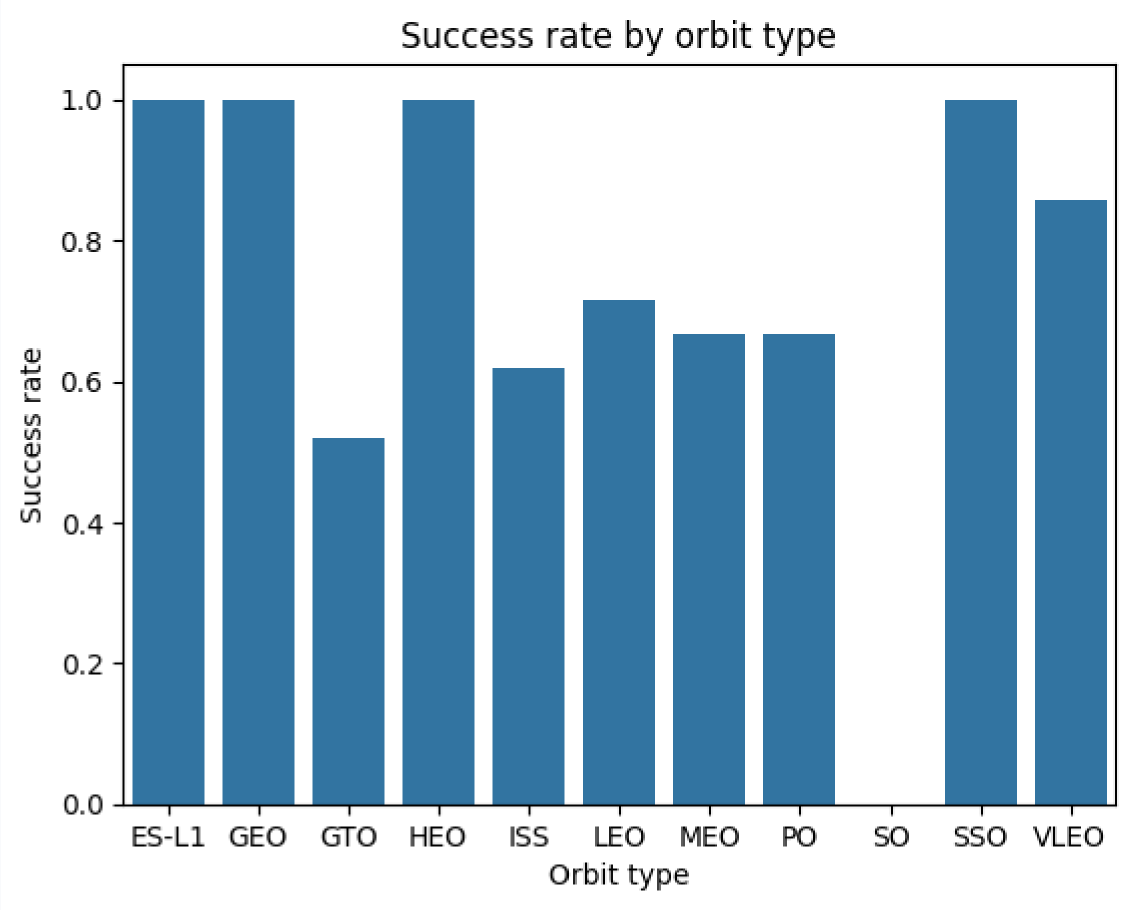
Payload vs. Launch Site

- Most payloads are under 8 tonnes
- VAFB SLC 4E not used for payloads greater than 10 tonnes
- KSC LC 39A not used for payloads less than 2 tonnes



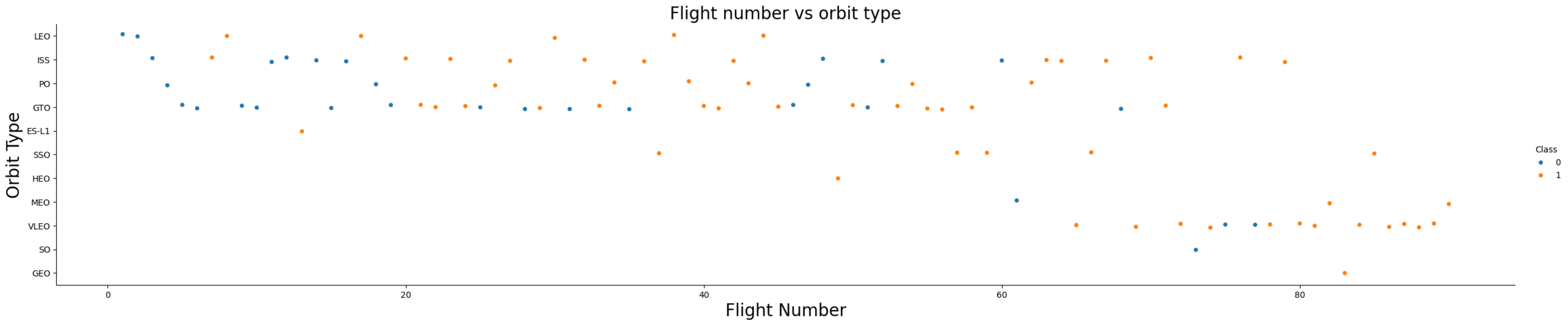
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO orbits have the highest success rates (100%)
- VLEO also comes in close (85.7%)



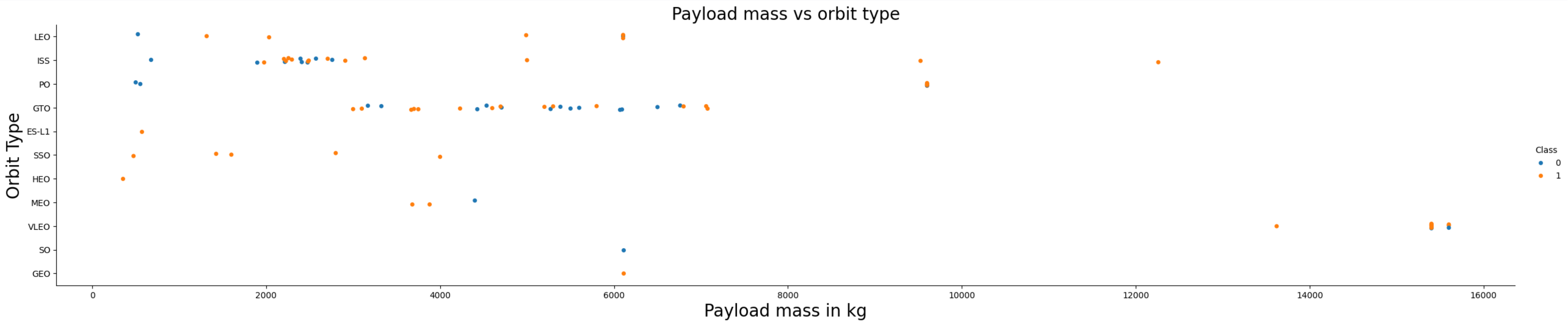
Flight Number vs. Orbit Type

- Rate of success increases with flight number in LEO orbit
- Conversely, no relationship between success rate and flight number in GTO orbit



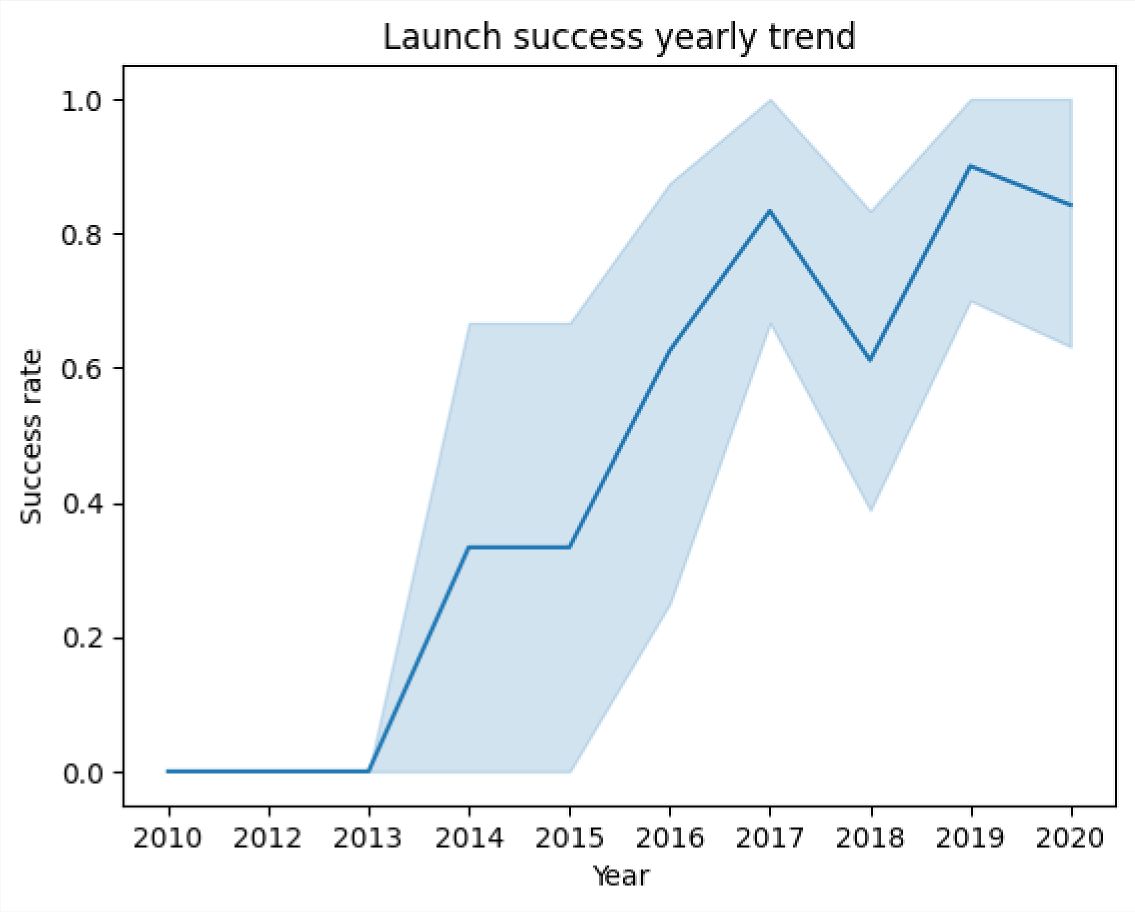
Payload vs. Orbit Type

- Heavier payloads have a higher success rate for PO, LEO and ISS orbits



Launch Success Yearly Trend

- Success rate increased continuously from 2013-2017
  - Stable between 2014-2015



All Launch Site Names

- Found the names of all unique launch sites in the dataset
- Launch sites:
    - CCAFS LC-40
    - VAFB SLC-4E
    - KSC LC-39A
    - CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Extracted 5 records where launch sites begin with ‘CCA’:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
4/06/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
8/12/2010	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)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
8/10/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
1/03/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried by boosters from NASA evaluated to be **45,596 kg**

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 evaluated to be **2928.4 kg**

First Successful Ground Landing Date

- First successful landing outcome on ground pad was found to have occurred on **22 December 2015**

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed drone ships with payload masses between 4-6 tonnes:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Evaluated the frequency of each mission outcome:

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

Boosters Carried Maximum Payload

- Found the boosters which have carried the maximum payload mass:

  - 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

- Found the failed drone ship landings from 2015:

substr(Date,6,2)	Landing_Outcome	Booster_Version	Launch_Site
1	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
4	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Evaluated the frequency of landing outcomes for flights between 4 June 2010 and 20 March 2017 (and listed them in descending order):

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

Section 3

Launch Sites Proximities Analysis

Map of launch sites

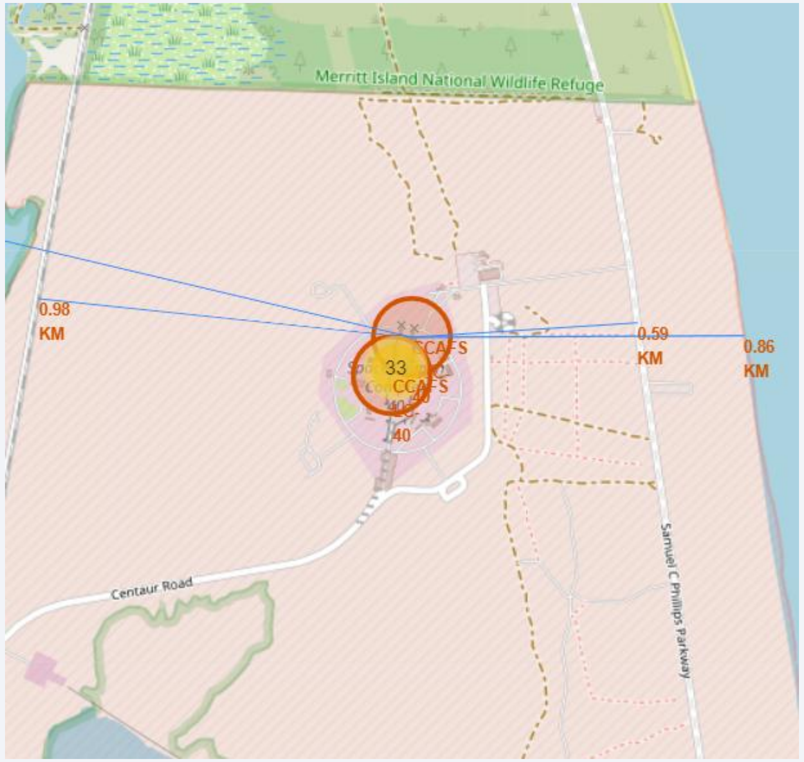
- All launch sites are in proximity to the equator line
  - Allows rockets to harness the Earth's rotation better
- All launch sites are near coasts
  - Gives rockets an uninhabited place to travel over (safer in the event the rocket fails)

Map of rocket launches

- Each marker indicates the location of a rocket launch
- Makers colour-coded to indicate the success/failure of the launch (red = failure, green = success)
- From this map, it was evident that KSC LC-39A had the highest launch success rate

Proximity of launch sites to key features

- Launch sites are close to railways and highways
  - Allows rocket components and other goods to be transported onto the launch sites
  - Also allows staff and visitors to commute to and from the site
- Launch sites are also close to coastlines
  - Gives rockets an uninhabited place to travel over (safer in the event the rocket fails)
- Launch sites keep a distance from cities, but are also not too far away from them
  - They are noisy and hazardous sites that need large amounts of land
  - However, they also depend on a large labor force to operate

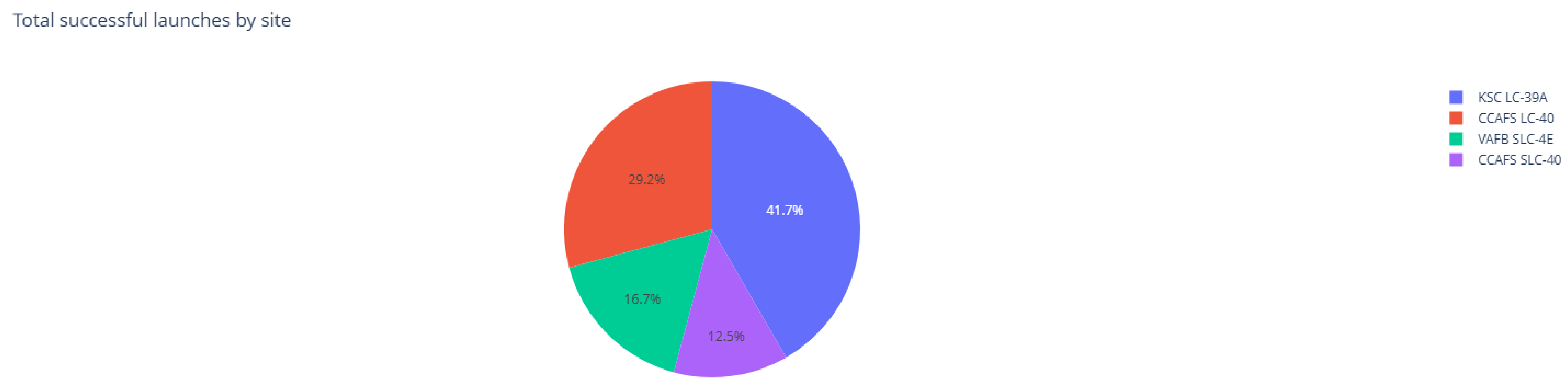


Section 4

Build a Dashboard with Plotly Dash

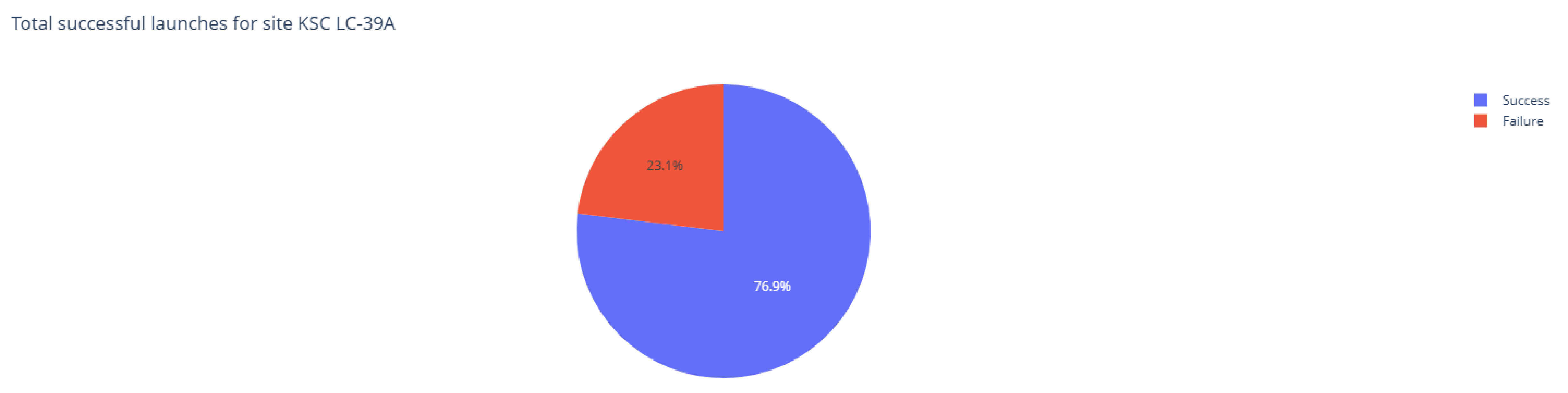
Total successful launches by site

- Most successful launches occurred at KSC LC-39A
- CCAFS SLC-40 has the least number of successful launches



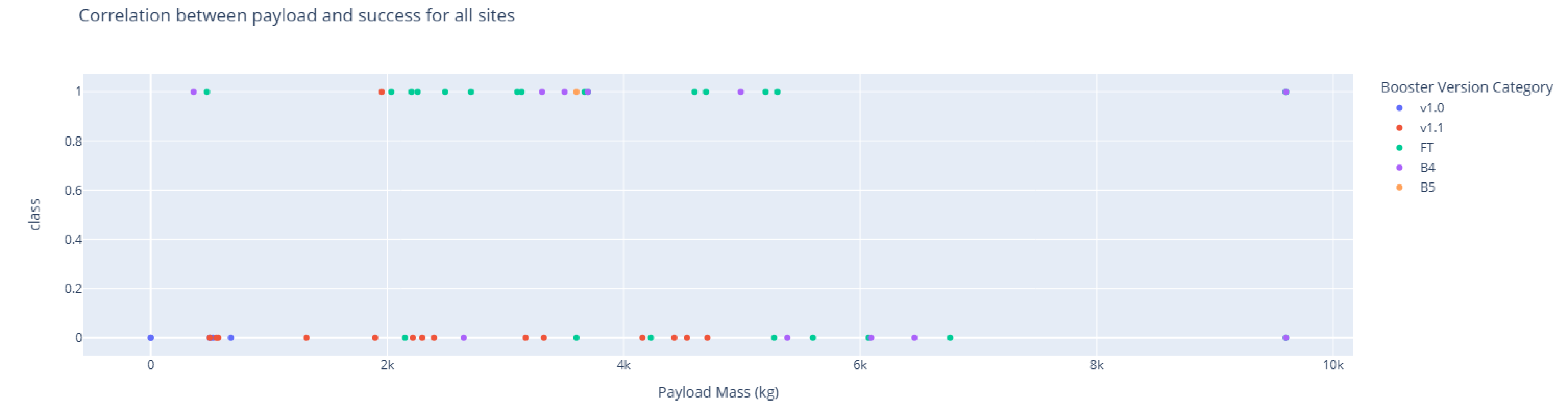
Highest launch success ratio

- KSC LC-39A has the highest launch success ratio of all launch sites investigated
- 76.9% of launches at the site were successful



Payload vs launch outcome

- The 2–4 ton payload range has the highest launch success rate
- The 0–2 ton and >6 ton payload ranges had the lowest launch success rates
- Of all the booster version categories, “FT” has the highest launch success rate

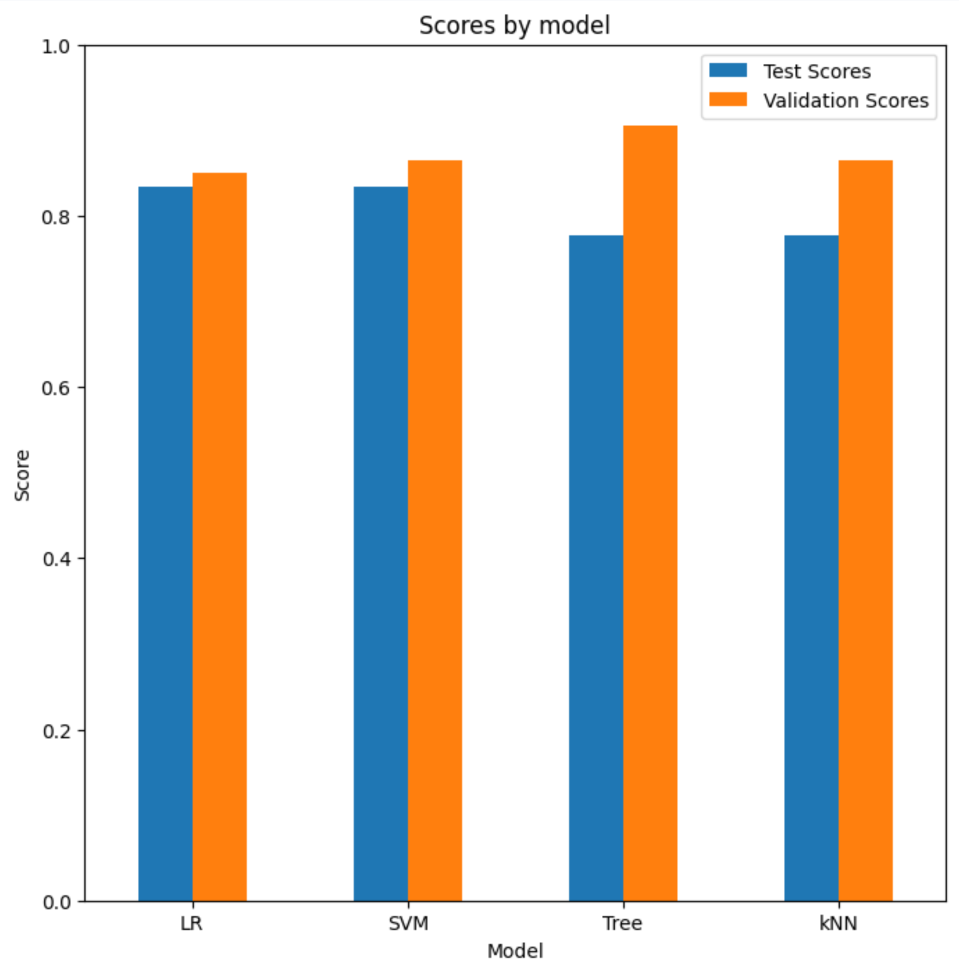


Section 5

Predictive Analysis (Classification)

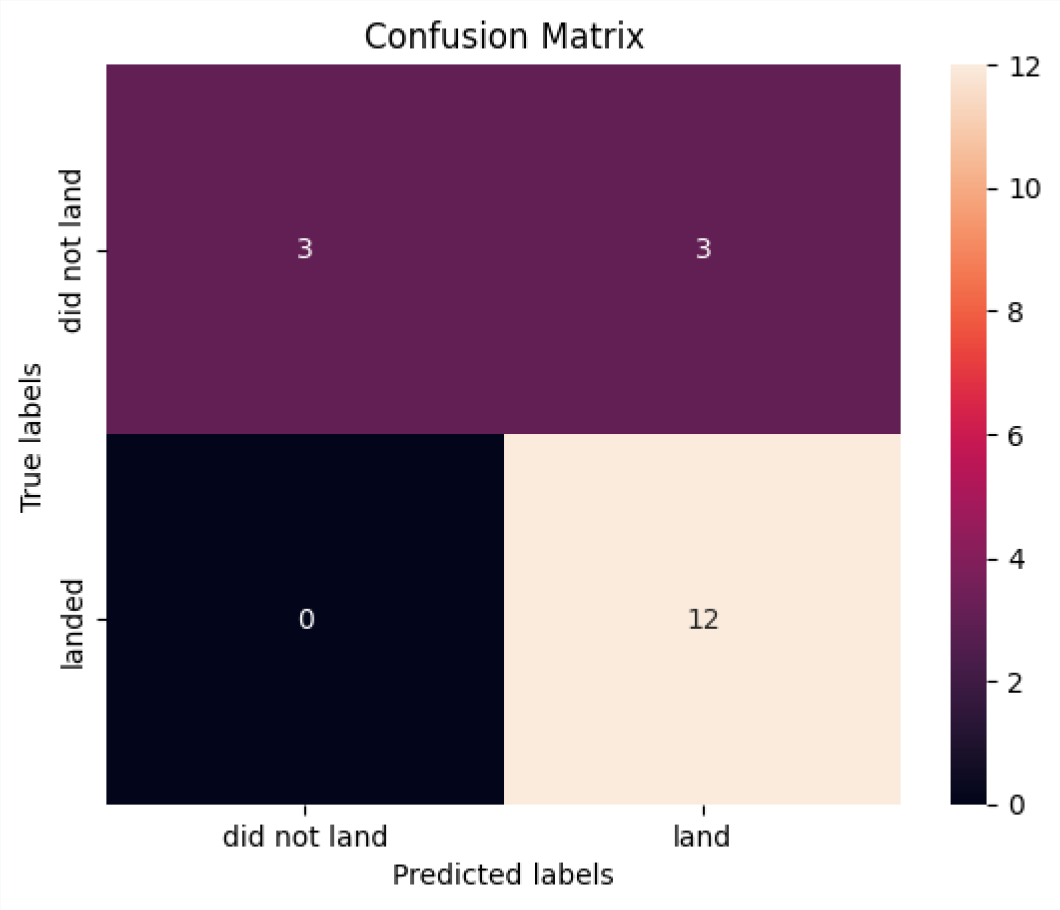
Classification Accuracy

- Support Vector Machine (SVM) was the machine learning model with the highest classification accuracy
  - Test score tied with Logistic Regression (LR)
  - SVM had a slightly higher validation score



Confusion Matrix

- Both LR and SVM achieved the results shown on the right
- Both of these models successfully predicted all the successful landings in the dataset, but only correctly predicted half of the failed landings.



Conclusions

- Landing success can be effectively predicted using machine-learning models
- SVM and Logistic Regression were the most reliable classifiers
- Launch success is influenced by orbit type, payload mass, and booster version
- Predictive insights can help estimate launch costs and guide competitive pricing strategies

Appendix

- GitHub repository containing all files
- SpaceX REST API (Accessed 29 Sept. 2025)
- List of Falcon 9 and Falcon Heavy launches, Wikipedia (Accessed 29 Sept. 2025)

Thank you!