Winning Space Race with Data Science

Meks Egbuchunam
08/15/2025

Outline

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

Executive Summary

- Collected and analyzed public SpaceX launch data to study Falcon 9 first-stage recovery
- Cleaned data and conducted an EDA for better understanding and subsequent modeling
- Visualized data through conventional graphing and interactive maps
- Applied multiple predictive models to estimate recovery success
- Results showed equal accuracy among models

Introduction

- Project background and context
  - The commercial space industry is growing quickly with companies like SpaceX reducing launch costs through reusable rockets
  - Falcon 9 first stage recovery lowers costs from about 165 million dollars to approximately 62 million dollars per launch
  - Recovery is not guaranteed and depends on payload, orbit, and mission requirements
- Problems you want to find answers to
  - We want to analyze public launch data to predict whether SpaceX's Falcon 9 first stage will be recovered
  - We also want to use these predictions to help Space Y estimate launch costs and compete more effectively in the market

Section 1

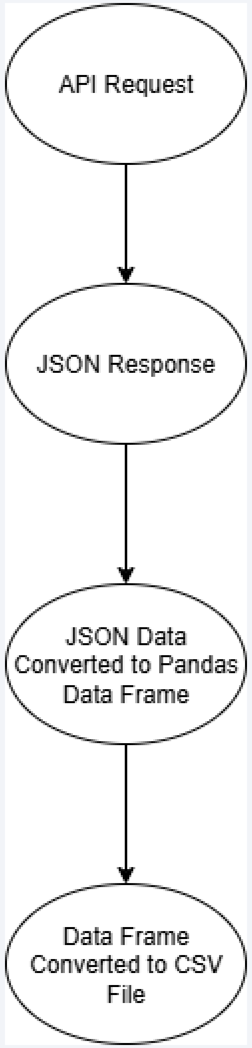
Methodology

Methodology

- Data collection methodology:
  - Gathered past launch data from the SpaceX REST API and Wikipedia tables
  - Retrieved rocket, payload, and landing details from multiple API endpoints
  - Converted JSON responses into a Pandas DataFrame
- Perform data wrangling
  - Filtered data to include only Falcon 9 launches
  - Replaced missing PayloadMass values with the column mean
  - Kept LandingPad null values for later encoding
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

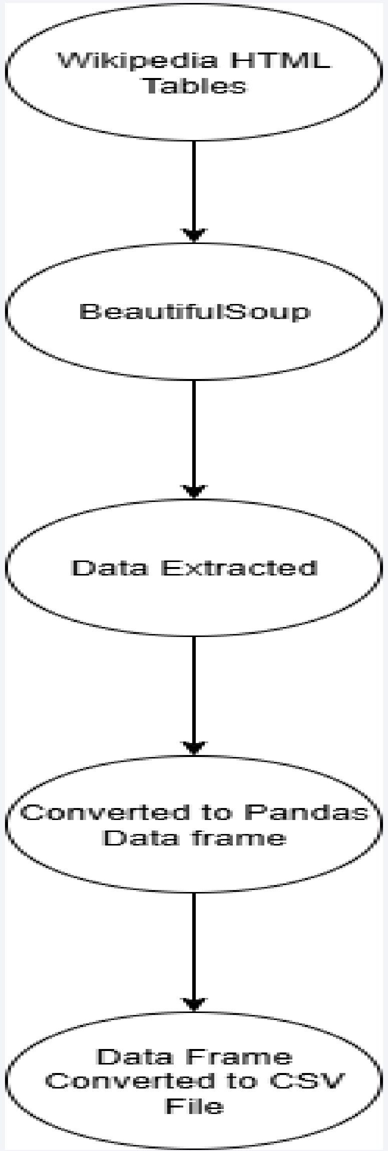
Data Collection – SpaceX API

- Obtained Falcon 9 launch data using the SpaceX REST API
- Used Python requests to make GET calls and received JSON responses
- Converted JSON data into a Pandas DataFrame with json_normalize
- Extracted extra details (rocket, payload, launch site, core) by matching ID codes with other parts of the API
- Exported the dataset into CSV format for later analysis
- Github URL: https://github.com/Meks7/SpaceX-Launch-Predictions/blob/main/Lab%201%20-%20data%20collection.ipynb



Data Collection - Scraping

- Collected Falcon 9 launch records from Wikipedia pages
- Parsed the tables and extracted launch data using Python's BeautifulSoup
- Stored the extracted data into lists
- Converted lists into a Pandas data frame for further cleaning and analysis
- Saved the scraped dataset to CSV format
- GitHub URL: https://github.com/Meks7/SpaceX-Launch-Predictions/blob/main/Lab%202%20-%20data%20collection%20(web%20scraping).ipynb



Data Wrangling

- Combined datasets into a single data frame
- Filtered out Falcon 1 records
- Replaced null values where appropriate (left LandingPad column null values unchanged to represent trials with no Landing Pad used).
- You need to present your data wrangling process using key phrases and flowcharts
- Converted Data to CSV
- Github URL: https://github.com/Meks7/SpaceX-Launch-Predictions/blob/main/Lab%203%20-%20Data%20Wrangling.ipynb

EDA with Data Visualization

- Plotted Flight Number vs Launch Site to see how launch experience affects success
- Graphed Payload Mass vs Launch Site to check if heavier payloads lower success rates
- Looked at success rate by orbit type to compare different mission goals
- Plotted Flight Number vs Orbit and Payload Mass vs Orbit for patterns in outcomes
- Tracked yearly success rate with a line chart to spot performance trends over time
- Github URL: https://github.com/Meks7/SpaceX-Launch-Predictions/blob/main/Lab%205%20-%20EDA%20with%20Visualization.ipynb

EDA with SQL

- Found the unique launch site names
- Selected 5 records where launch sites begin with “CCA”
- Calculated total payload mass carried by NASA missions
- Computed average payload mass for booster version F9 v1.1
- Retrieved the date of the first successful landing on a ground pad
- Listed boosters that landed on a drone ship with payloads between 4000–6000 kg
- Counted the total number of successful vs failed mission outcomes
- Found the booster that carried the maximum payload mass
- Listed failed drone ship landings in 2015 with booster version and launch site
- Ranked landing outcomes between 2010–06–04 and 2017–03–20
- Github URL: https://github.com/Meks7/SpaceX-Launch-Predictions/blob/main/Lab%204%20-%20complete%20the%20EDA%20(SQL).ipynb

Build an Interactive Map with Folium

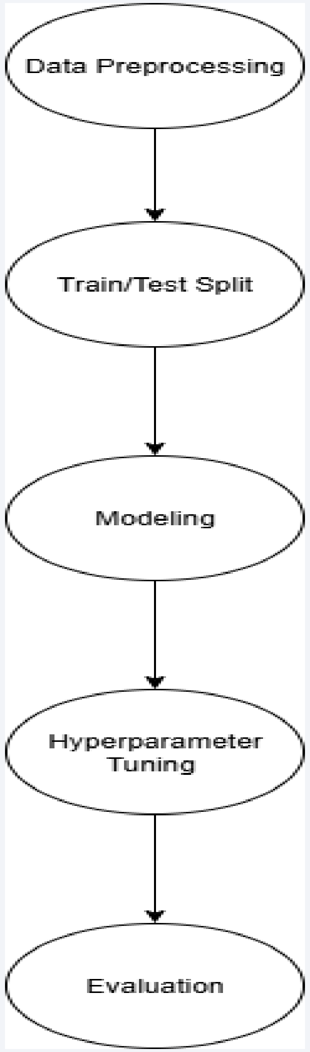
- Added markers to show all launch site locations
- Used circles to display success/failure outcomes by color
- Drew lines from launch sites to nearby landmarks (e.g., coastlines, highways) to measure distance
- Those objects were added in order to give a clear visual of where SpaceX launches occur, how outcomes vary by site, and how launch locations relate to surrounding geography
- Github URL: https://github.com/Meks7/SpaceX-Launch-Predictions/blob/main/Lab%206%20-%20Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Added a dropdown menu to filter results by launch site
- Created a pie chart showing total successes by site, and success vs failure when a specific site is selected
- Added a payload range slider to adjust the range of payload masses displayed
- Built a scatter plot to show the relationship between payload mass and launch success, colored by booster version
- Those objects were added in order to allow interactive exploration of launch success across sites, payloads, and booster types
- Github URL: https://github.com/Meks7/SpaceX-Launch-Predictions/blob/main/Lab%207%20-%20SpaceX%20dash%20app.py

Predictive Analysis (Classification)

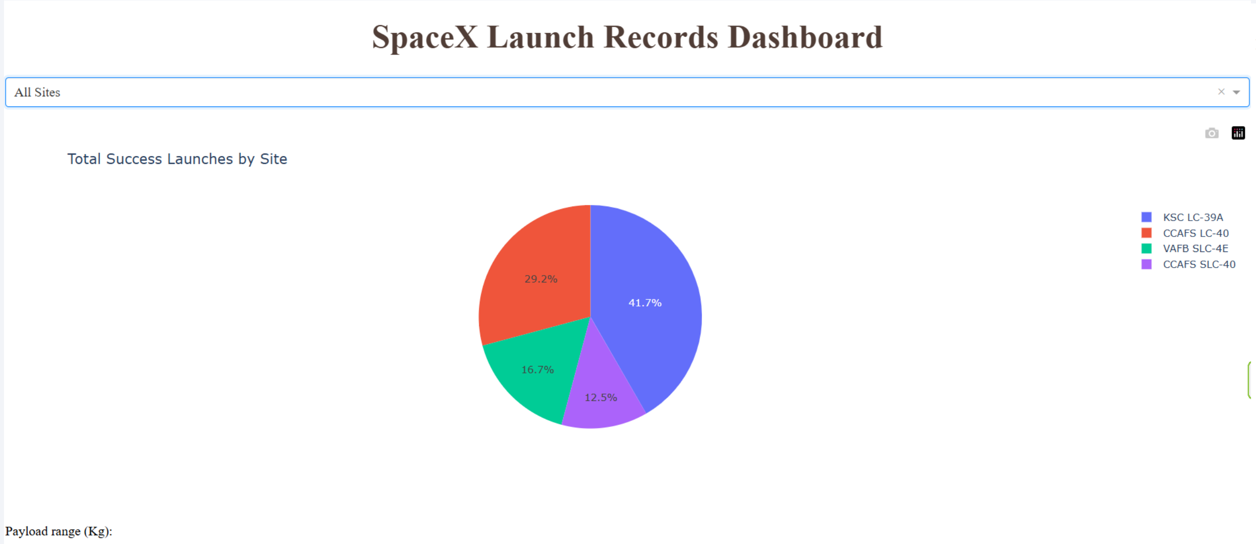
- Built multiple classification models: Logistic Regression, SVM, Decision Tree, and KNN
- Split the dataset into training and test sets, then trained each model
- Tuned hyperparameters (e.g., grid search for Decision Tree and SVM) to improve performance
- Evaluated models using accuracy scores and confusion matrices
- Found the best performing model based on accuracy and interpretability
- You need present your model development process using key phrases and flowchart
- Github URL: https://github.com/Meks7/SpaceX-Launch-Predictions/blob/main/Lab%208%20-%20Machine%20Learning%20Prediction.ipynb

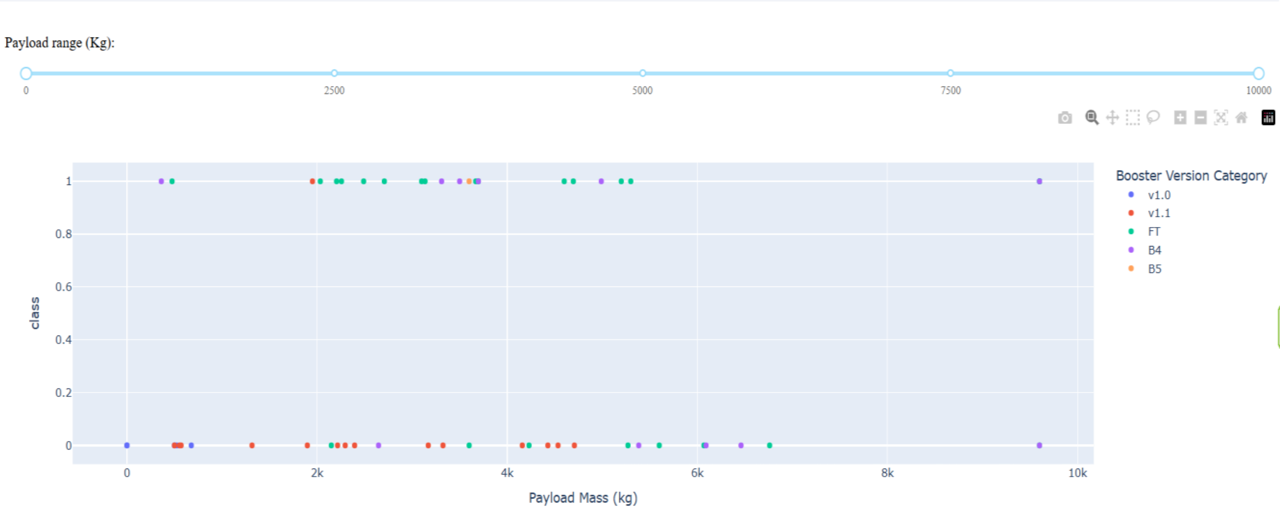


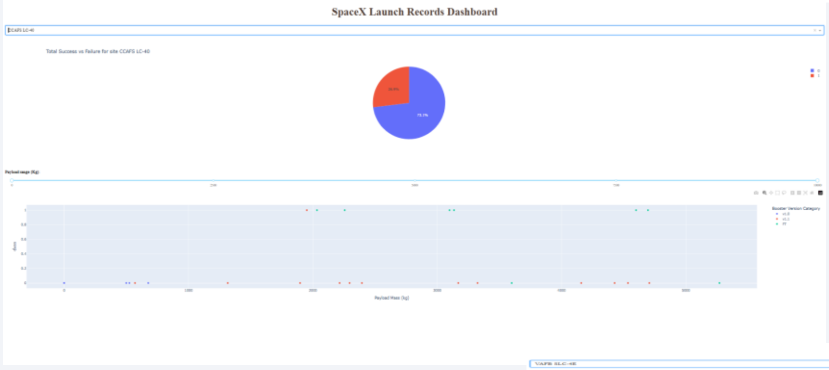
Results

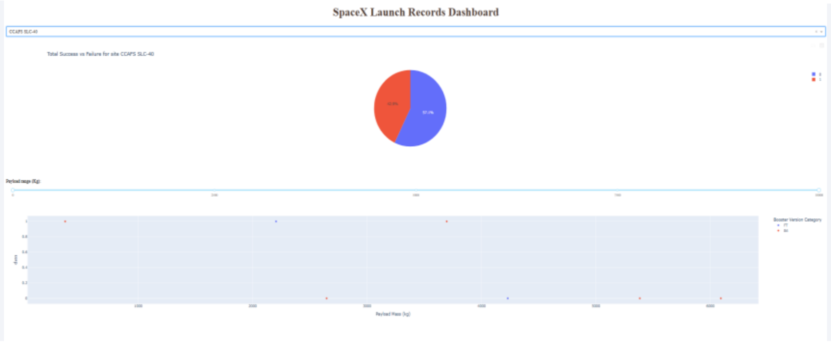
- Exploratory Data Analysis Results
  - Launches came from four main sites (CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E)
  - Average payload mass for F9 v1.1 was calculated, and NASA missions carried the largest total payloads
  - Success rates varied by orbit type and launch site
  - Success improved over time as flight numbers increased
- Predictive Analysis Results
  - All models achieved similar accuracy scores (around the low–mid 80% range)
  - Logistic Regression performed the best (though accuracy was near identical across all subjects) and was chosen as the best model

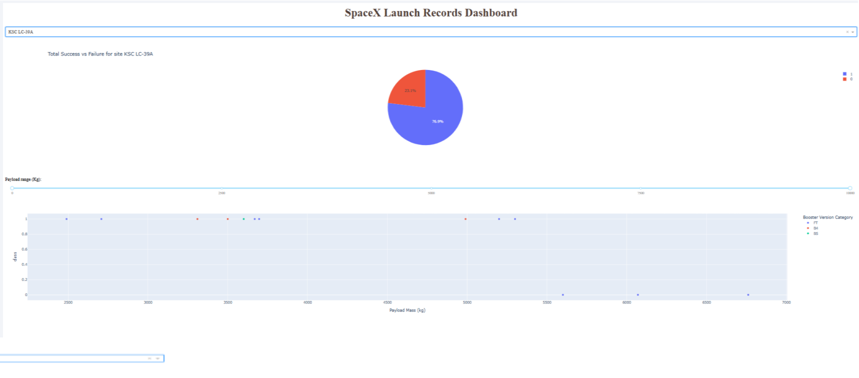
Interactive Analytics Demo

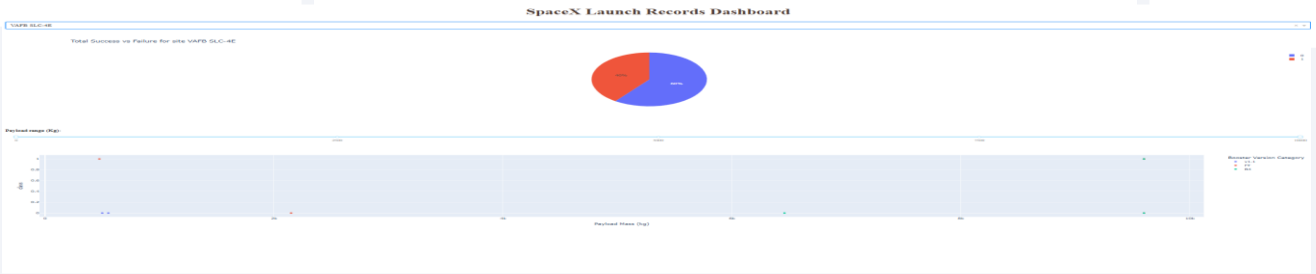










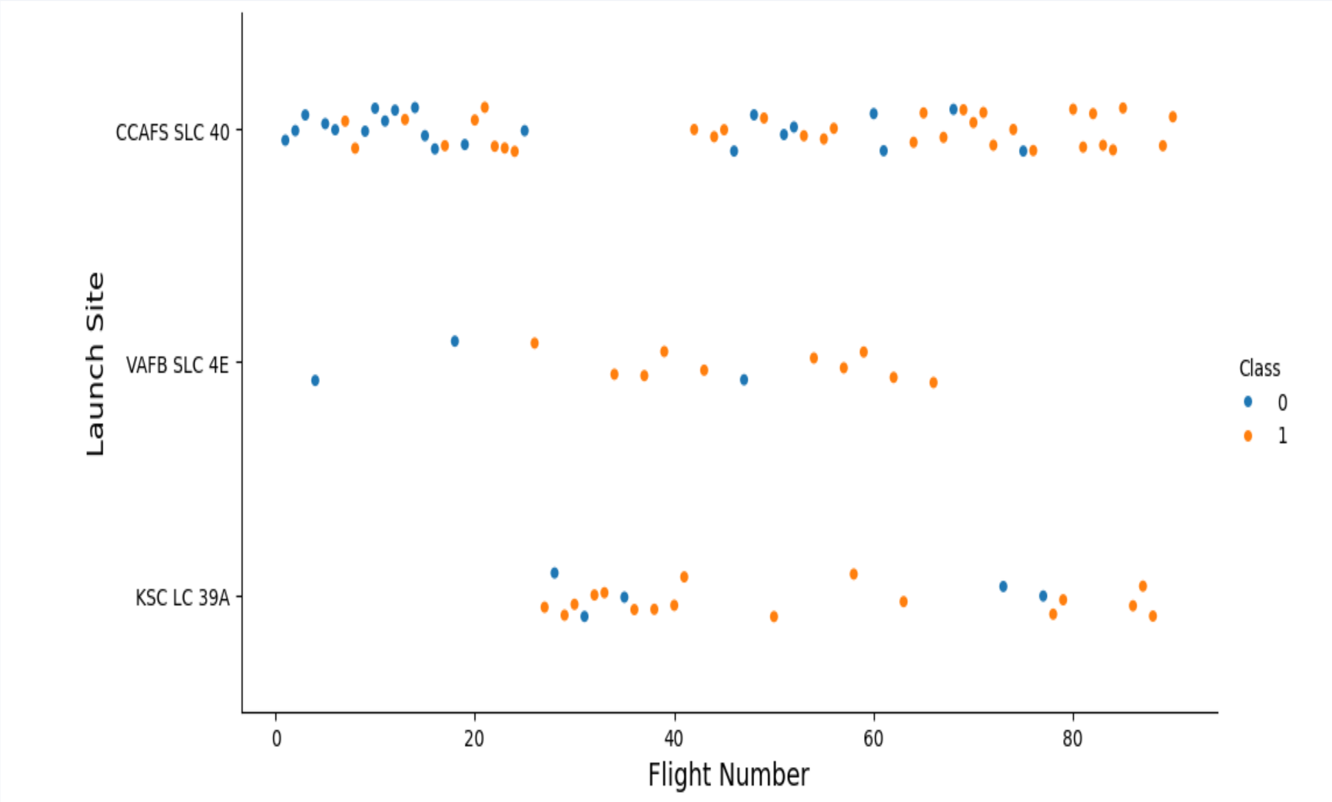


Section 2

Insights drawn from EDA

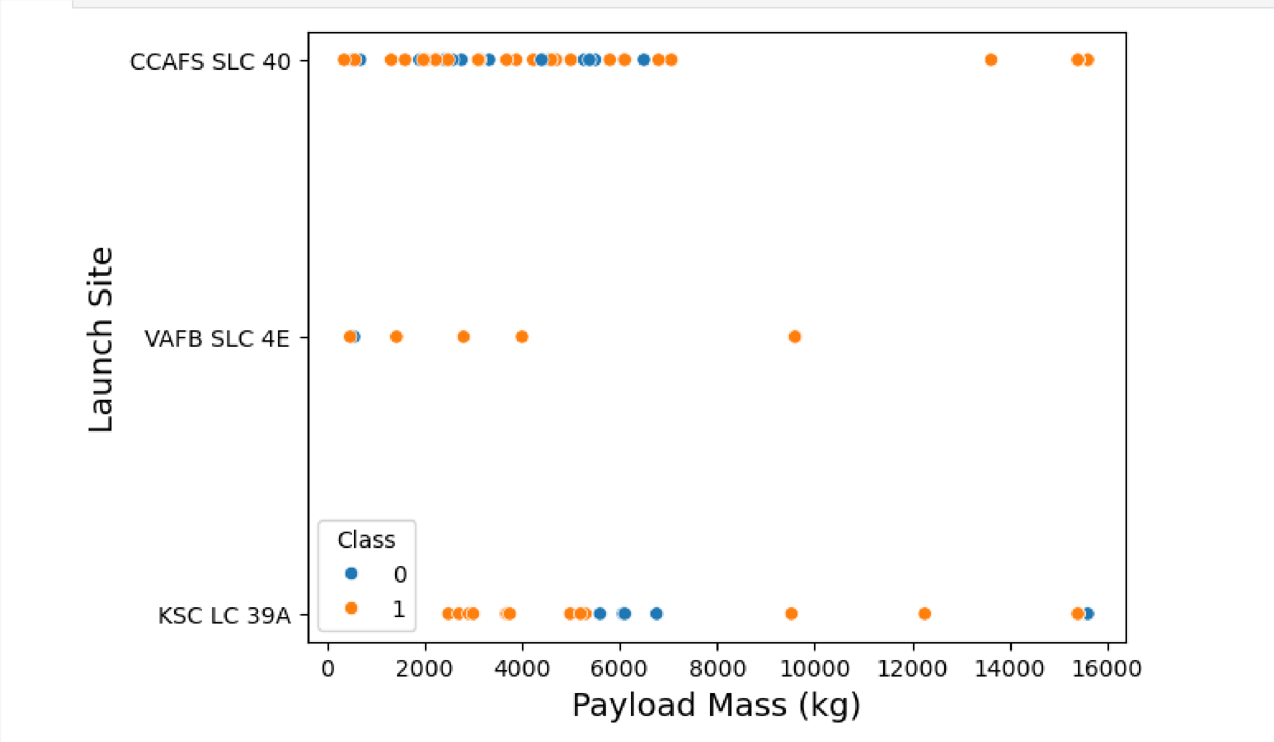
Flight Number vs. Launch Site

- At lower flight numbers (early launches), there are more failures across sites
- As flight numbers increase (later launches), the number of successful landings increases
- The plot suggests that both site choice and accumulated flight experience are linked to higher landing success rates.



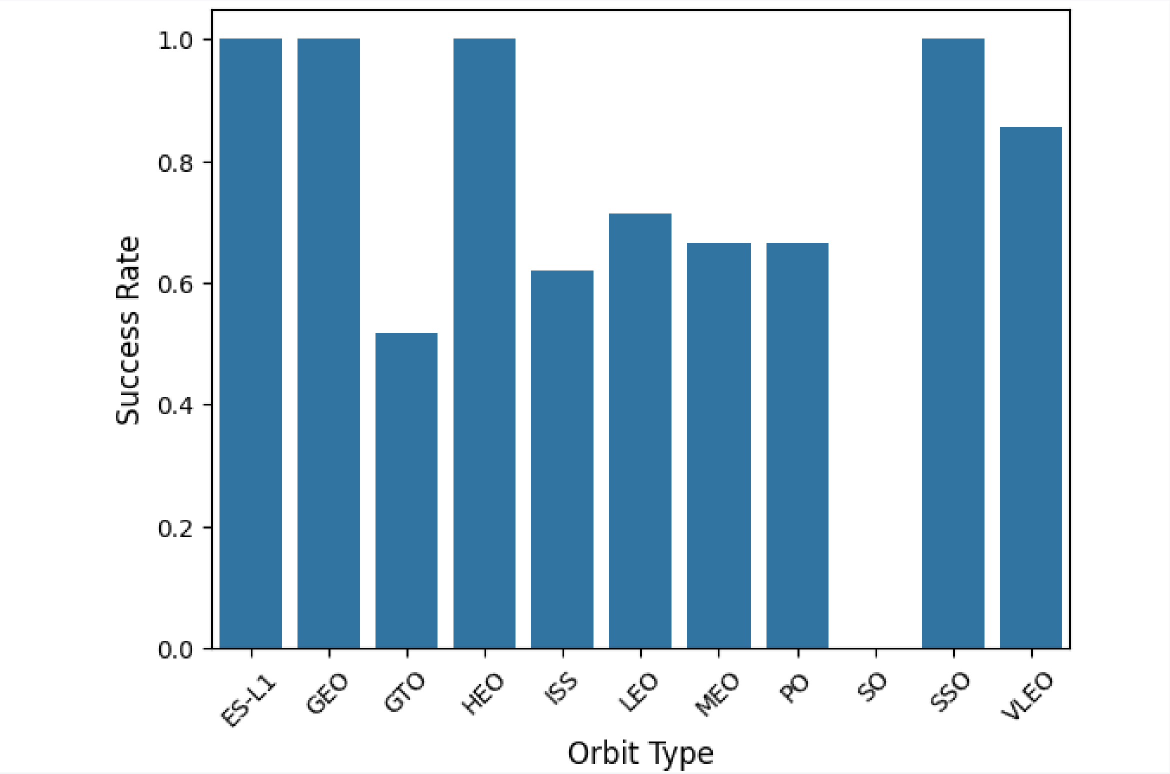
Payload vs. Launch Site

- At KSC LC-39A, most payloads were heavier and had a high success rate.
- At CCAFS SLC-40, launches cover a wide payload range but with more mixed outcomes.
- At VAFB SLC-4E, payloads were lighter overall, with fewer launches and a mix of success/failure.
- Overall, success is possible across all payload ranges
- Heavier payloads have had some failures in the past, but success is still possible even at high payload masses.



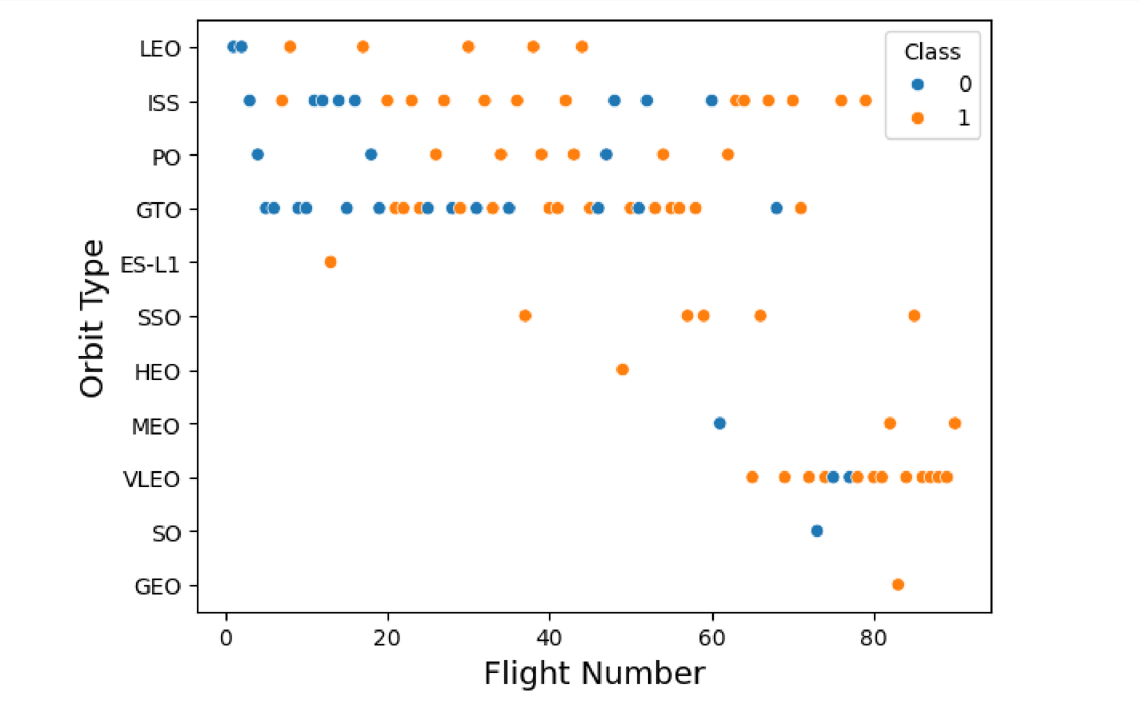
Success Rate vs. Orbit Type

- Certain orbits, such as ES-L1 and GEO, have near perfect success rates
- GTO has the lowest success rate
- Other orbits, such as MEO, LEO, and PO are middle of the pack
- This suggests that mission orbit strongly affects landing success, with more demanding orbits (like GTO) being riskier.



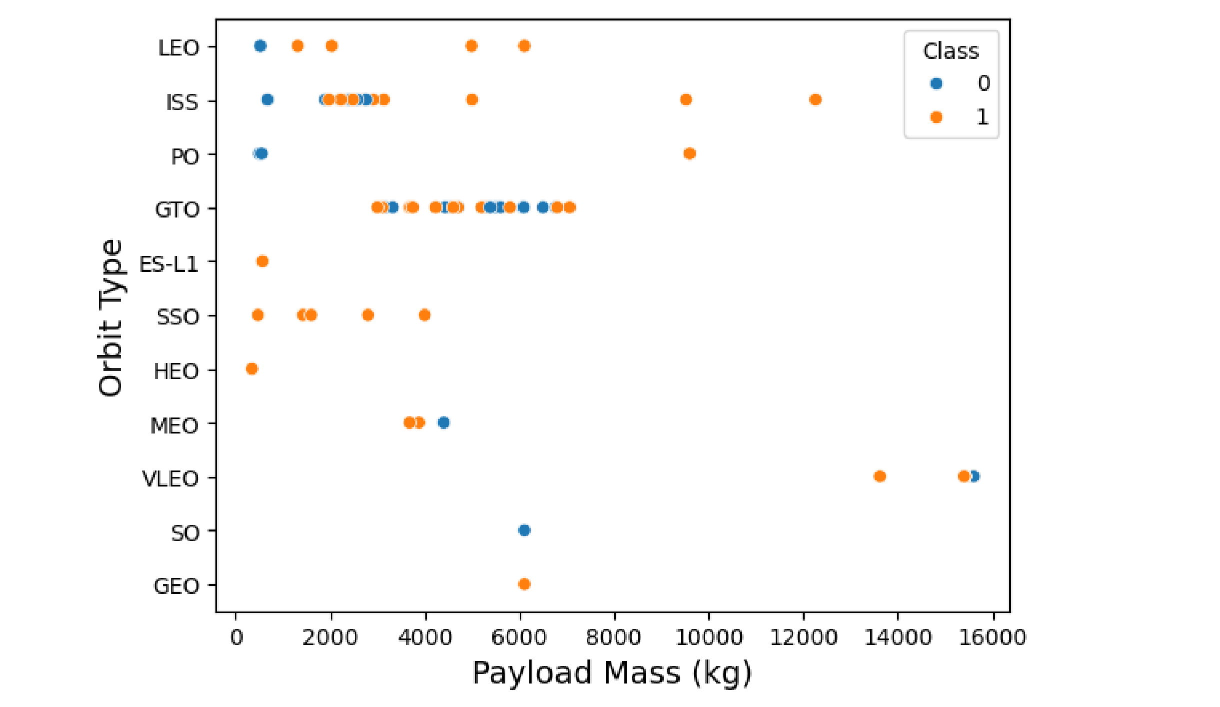
Flight Number vs. Orbit Type

- Like the previous scatterplots, failures are more common across orbits at in earlier stages (lower flight numbers).
- Likewise, As flight numbers increase, the proportion of successful launches increases.
- As flights increase, success rates do as well



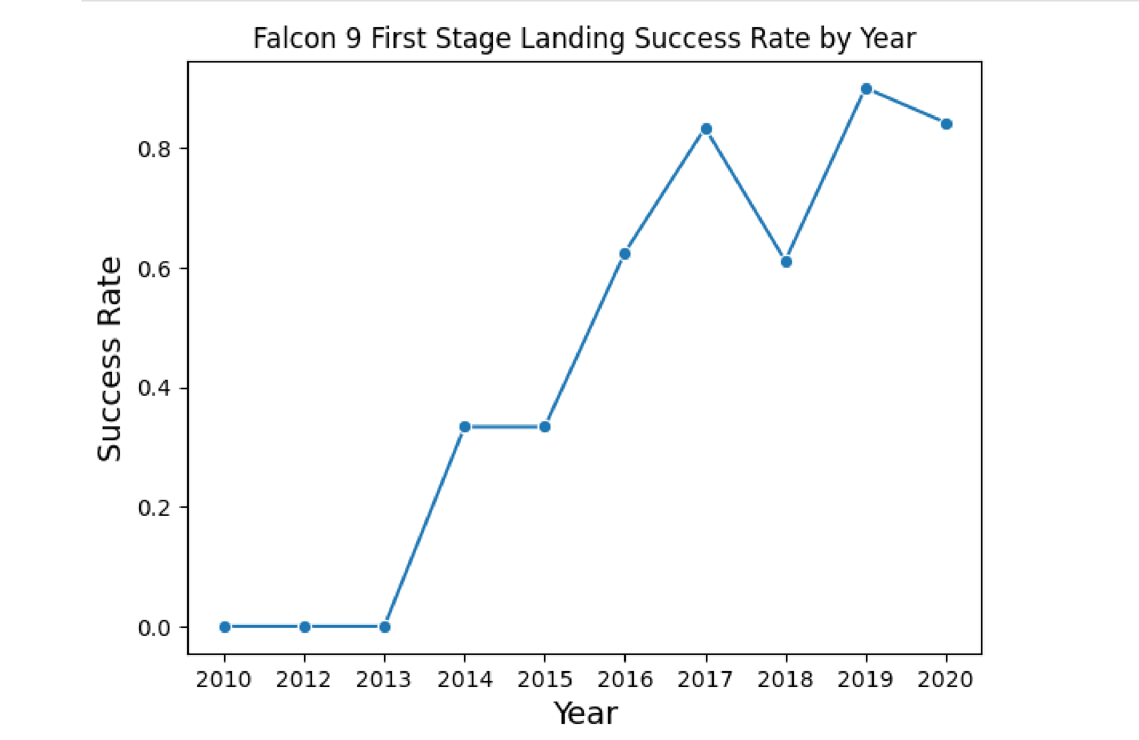
Payload vs. Orbit Type

- VLEO covers heavier payloads.
- GTO clusters around payloads of approximately 4000-6000 kg
- No strong interaction seems to be present between payload and orbit



Launch Success Yearly Trend

- From 2010–2013, Falcon 9 first stage landings had little success.
- Starting in 2014, success rates improved steadily, reaching over 80% by 2017.
- Despite a dip in 2018, SpaceX quickly recovered, achieving nearly 90% success in 2019–2020.
- Overall, the trend shows rapid progress and reliability improvements



All Launch Site Names

- The names of the different Launch Sites

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

Launch Site Names Begin with 'CCA'

- Query filtered launch sites beginning with “CCA” (Cape Canaveral).
- Verified at least 5 records exist for Cape Canaveral–based launches.

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 boosters used for NASA's Commercial Resupply Services (CRS) missions carried a combined payload of 45,596 kg.

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- Booster version F9 v1.1 carried an average payload of approximately 2,928 kg per mission.

avg_payload_mass
2928.4

First Successful Ground Landing Date

- Date: 2015-12-22 (December 22nd, 2015)
- This was the first time a Falcon 9 booster successfully landed on solid ground, proving reusability was possible

first_successful_ground_pad_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Four Falcon 9 Full Thrust boosters achieved success within the conditions: B1022, B1026, B1021.2, and B1031.2.

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

Total Number of Successful and Failure Mission Outcomes

- The overwhelming majority of Falcon 9 missions were successful (approximately 98%)

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

Boosters Carried Maximum Payload

- The maximum payload capacity was achieved by multiple Falcon 9 Block 5 boosters.

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

- Two failed attempts occurred in January and April 2015, both with Falcon 9 v1.1 boosters at Cape Canaveral (LC-40).

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

- In SpaceX's early years, many missions had no landing attempts (10) or ended in ocean landings as tests.
- By this period, successes and failures on drone ships were evenly split (5 each), while 3 ground pad landings succeeded.

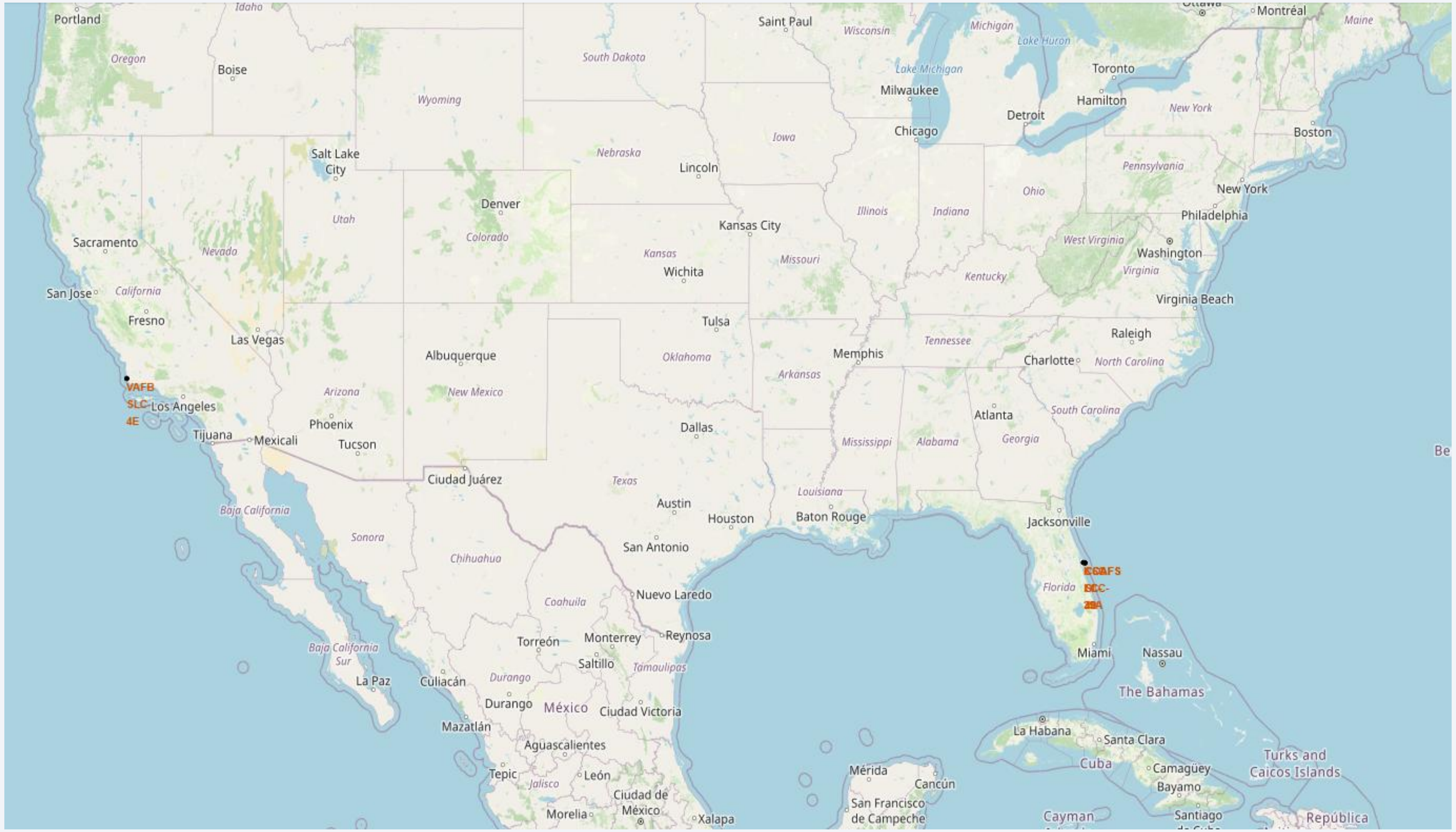
Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch Site Locations

- Two Regions/States where the launch sites reside
  - Florida (Cape Canaveral, Kennedy Space Center) supports equatorial & GTO missions.
  - California (Vandenberg AFB) supports polar & sun-synchronous missions.

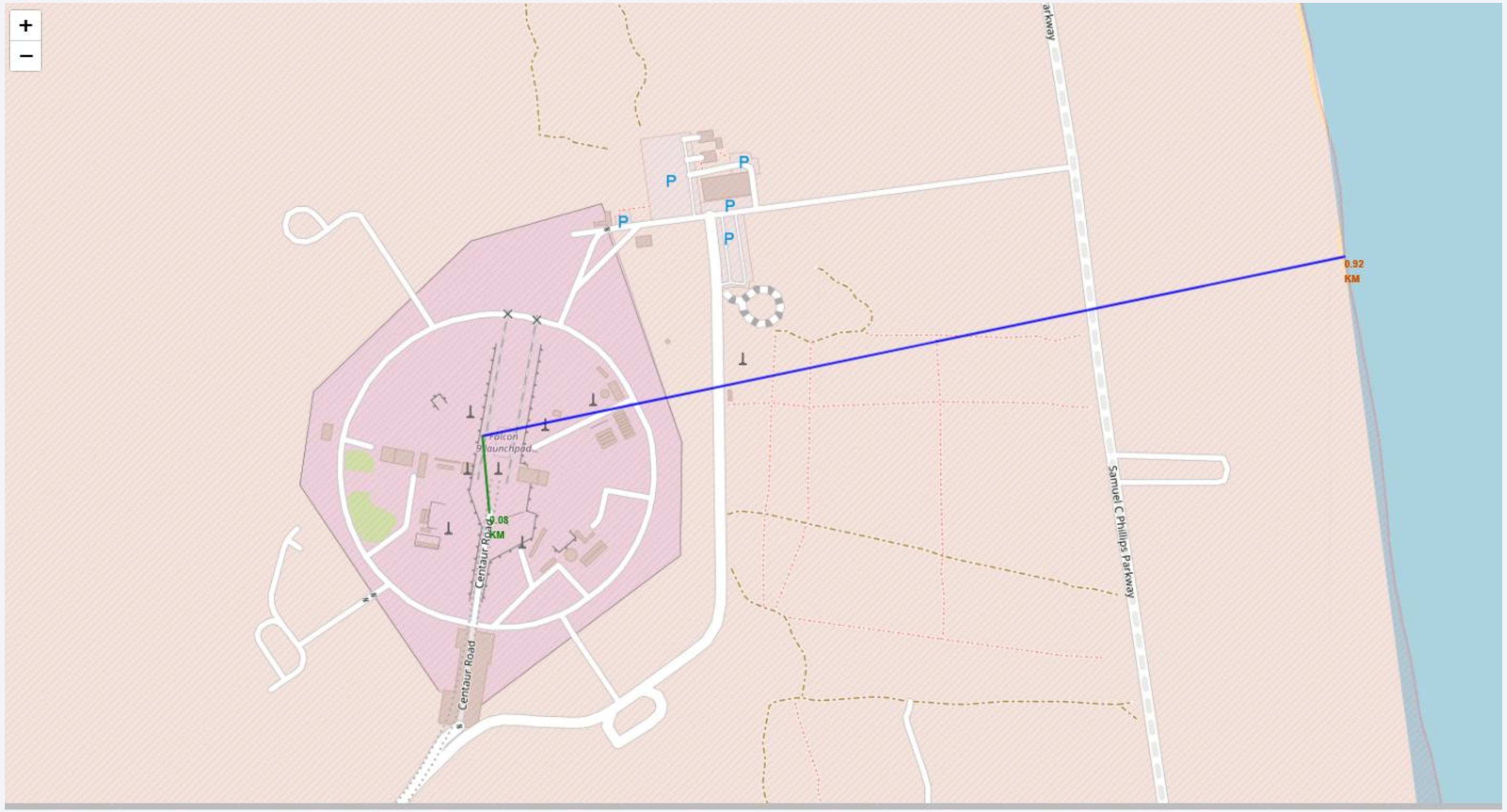


Launching Outcomes By Sight

- Green markers = successful landings, red markers = failed attempts.
- Florida (CCAFS LC-40, CCAFS SLC-40, KSC LC-39A) shows a high density of launches, with mostly successful outcomes in later years.
- California (VAFB SLC-4E) had fewer launches, also showing a mix of early failures and later successes.

Launch Site Proximity

- Map shows distance from the Cape Canaveral launch site to coastline (blue line, 0.92 km) as well as from the launch site to centaur road (green line, 0.08 km)

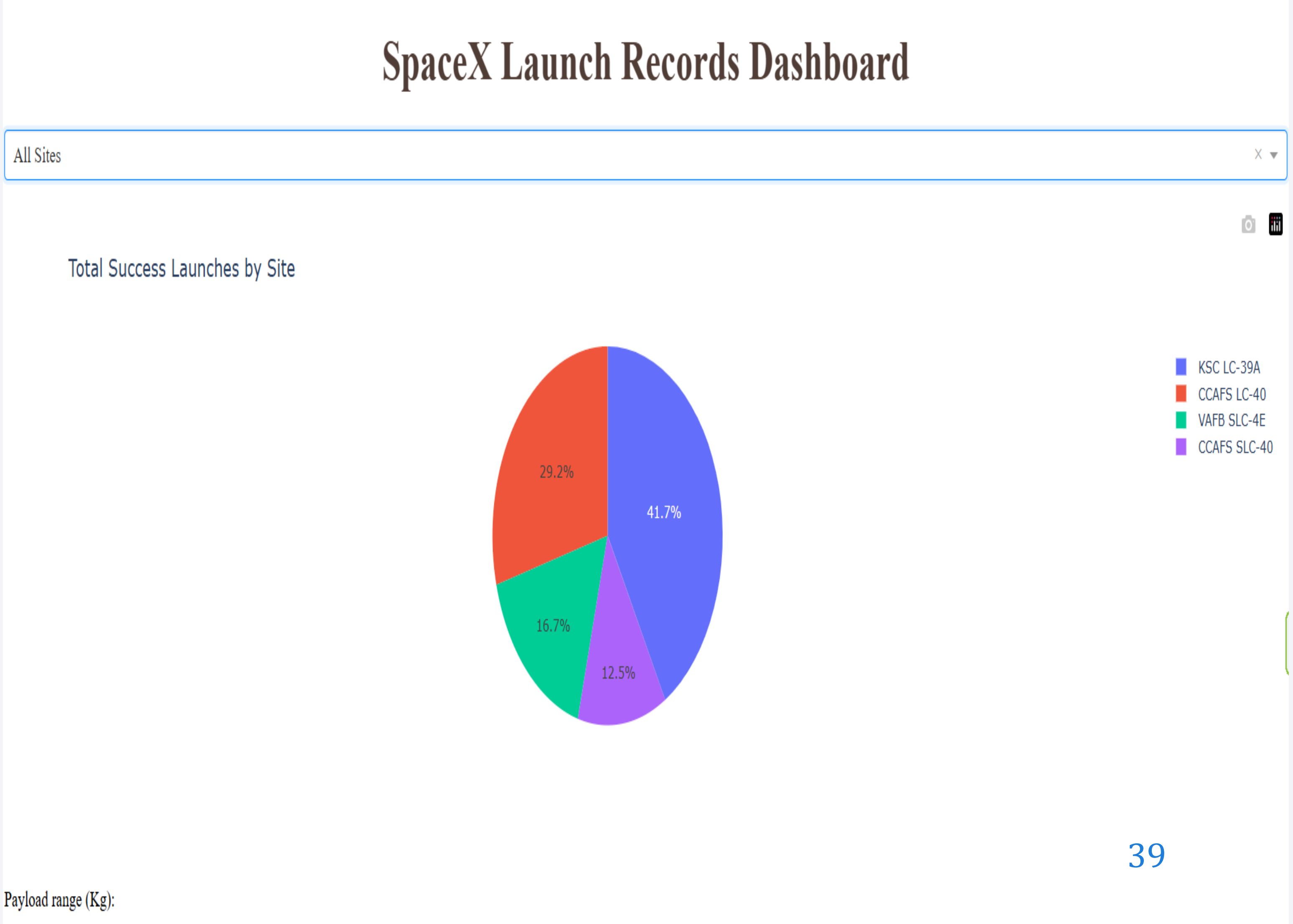


Section 4

Build a Dashboard with Plotly Dash

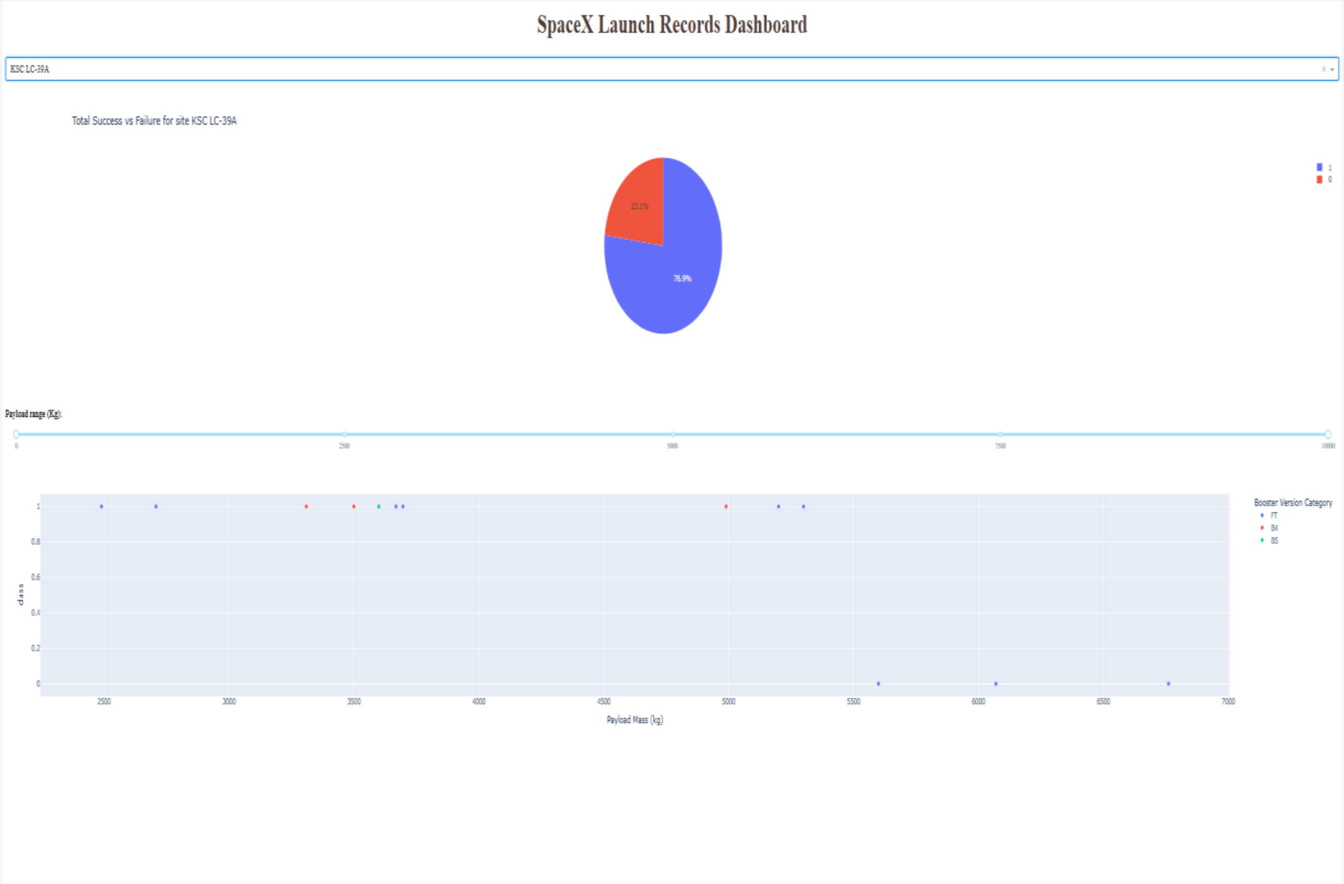
Total Success Launches by Site

- The pie chart shows the Distribution of successful launches across the four sites.
- KSC LC-39A has the highest proportion of successful launches (approximately 42%).
- VAFB SLC-4E and CCAFS SLC-40 have fewer successes at approximately 17% and 13% respectively.



Success v. Failure for KSC LC-39A

- The Pie Chart shows launch success vs failure for Kennedy Space Center Launch Complex (KSC LC-39A).
    - The chart shows that about 77% of the launches were successful, while about 23% of the launches were unsuccessful
    - This is the highest among all the launch sites
- The scatter plot displays the relationship between Payload Mass (kg) and Launch Outcome (Class) for this site.
    - Most payload ranges, especially between 2500–6000 kg, show successful outcomes (Class = 1).



Payload v. Launch Outcome Scatterplot

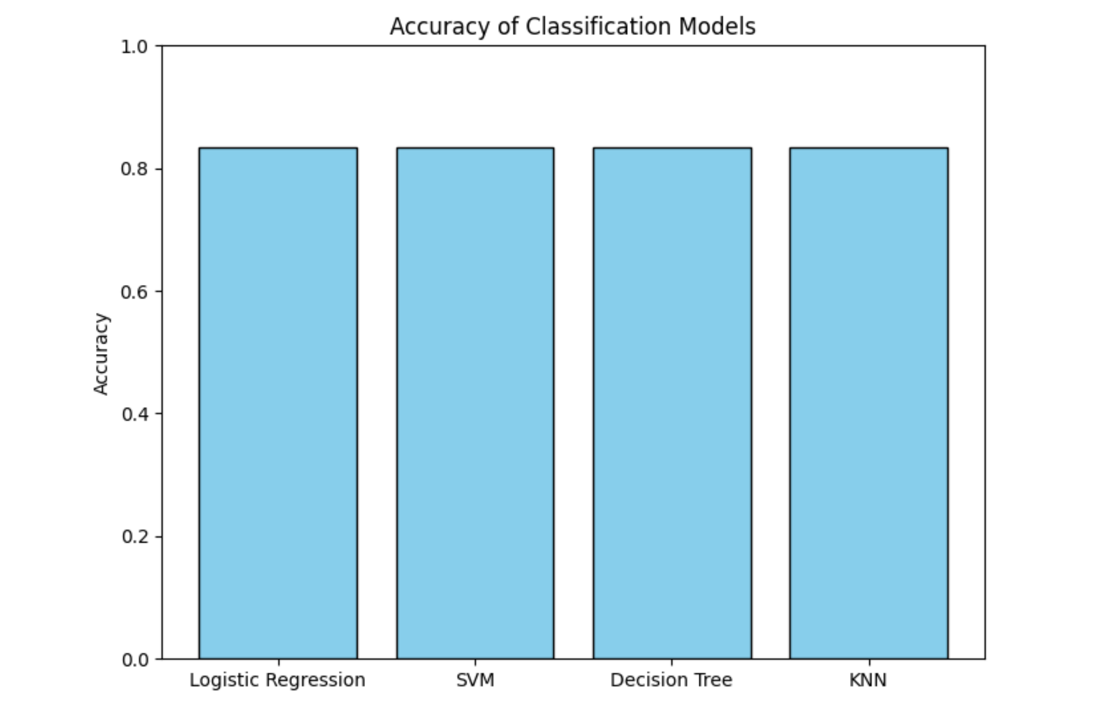
- The scatterplot shows the Payload Mass (kg) vs. Launch Outcome (Success/Failure) across all sites
  - Payload slider allows filtering by mass range (0–10,000 kg).
- Most successful launches are clustered between 2,000–6,000 kg payloads.

Section 5

Predictive Analysis (Classification)

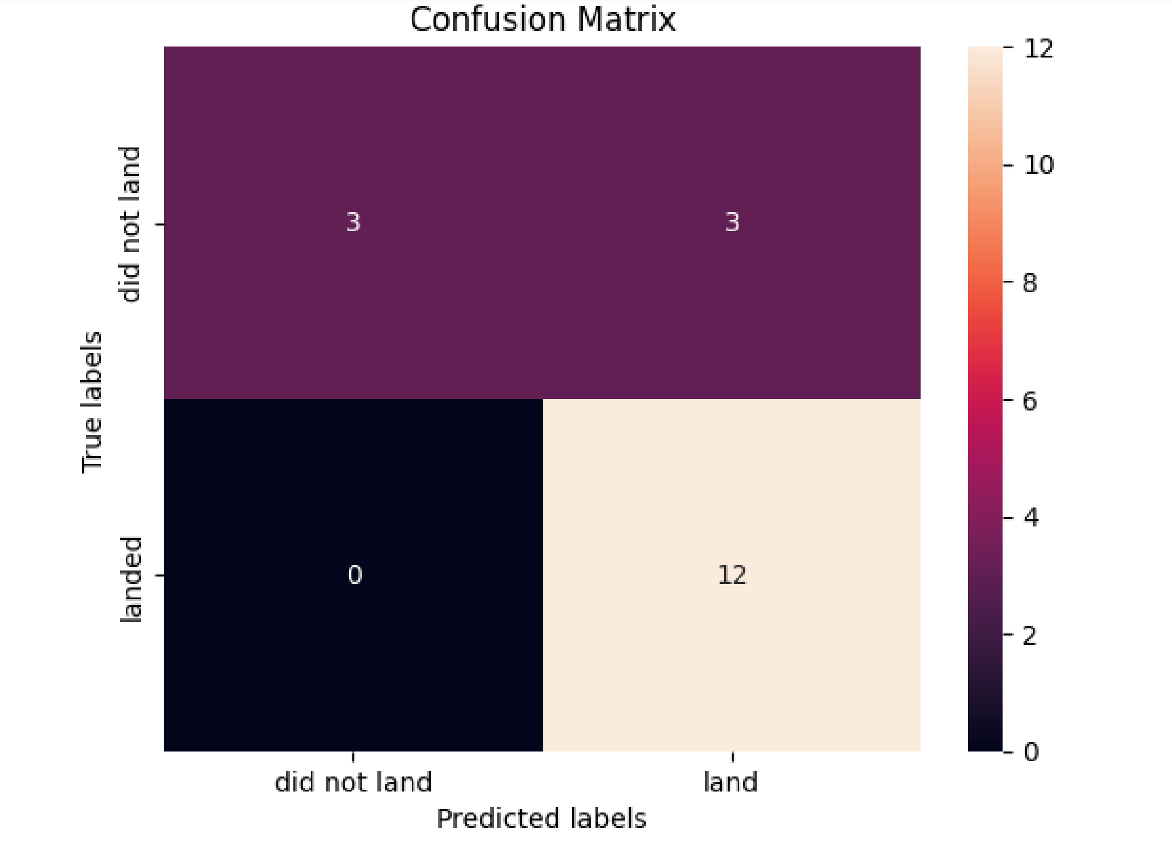
Classification Accuracy

- All four classification models showed near equal accuracy at approximately 83.33%.



Confusion Matrix

- The model correctly classified 15 out of 18 launches
- Misclassified failed landings as successful 50% of the time. That area may need more training.
- Perfectly predicted all successful landings
- All classification models performed equally.



Conclusions

- Falcon 9 recovery success is strongly influenced by payload mass, orbit, and launch site.
- SpaceX has achieved steadily increasing landing success rates, reaching approximately 90% in recent years.
- Among launch sites, KSC LC-39A demonstrated the highest success ratio.
- Machine learning models were able to predict landing outcomes with ~83% accuracy.
    - The models were able to predict successful landings perfectly (100% rate)
    - However, they predicted failed landings with only a 50% rate (more training may be necessary)
- These predictions can help the company estimate launch costs more accurately and strengthen market competitiveness.

Thank you!