IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Michal Precikowski
08/19/2025

Outline

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

Executive Summary

- Methods in brief: I combined targeted web scraping with the SpaceX REST API to create a consolidated dataset, which I then standardized using data wrangling. I developed interactive folium maps and a Plotly Dash dashboard, used plots and SQL to investigate patterns, and trained and adjusted classification models to forecast launch success.
- Brief results: Lighter payloads performed well, and success rates tended to be higher for specific orbits and launch locations. EDA helped to understand which features are the best option to define the success of launchings. The best model used payload, orbit, booster version, and site to accurately estimate the success probability.

Introduction

- To compete with the well-known SpaceX, the new company Space Y is taking action. By reducing marginal launch costs, Falcon 9's reusable architecture has changed the way people can access orbit. I convert mission-level records into useful insight for this project: Which elements are most beneficial to success? What is the relationship between launch site, orbit, and payload mass? Can those relationships be predicted to new launches using a predictive model?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The SpaceX API was used to request rocket launch data
  - Falcon 9 historical launch records were scraped from Wikipedia
- Perform data wrangling
  - Based on the collected data, a landing outcome label was created to represent the outcome of each launch
- 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 Collection

- Data sets were collected using:
  - The Space X API available at https://api.spacexdata.com
  - Web scraping from a Wikipedia page titled “List of Falcon 9 and Falcon Heavy Launches”

Data Collection – SpaceX API

- The SpaceX REST API allows to request launch, rocket, core, capsule, starlink, launchpad, and landing pad data
- HTTP GET requests were sent to request data from SpaceX API
- Dataset was constructed using the extracted properties for Falcon 9 launches, including the booster name, name of the launch site and others
- Dataset was exported to a CSV file
- Completed notebook available at https://github.com/MagicMagic99/DataCollectionAPI-Michal/blob/main/jupyter-labs-spacex-data-collection-api-Michal%20(1).ipynb

Data Collection – SpaceX API

Step 1	Step 2	Step 3
Collect data from the SpaceX REST API and parse the response into a Pandas dataframe	Filter data for Falcon 9 launches and extract meaningful attributes	Deal with missing values and save the dataset to a CSV file

Data Collection - Scraping

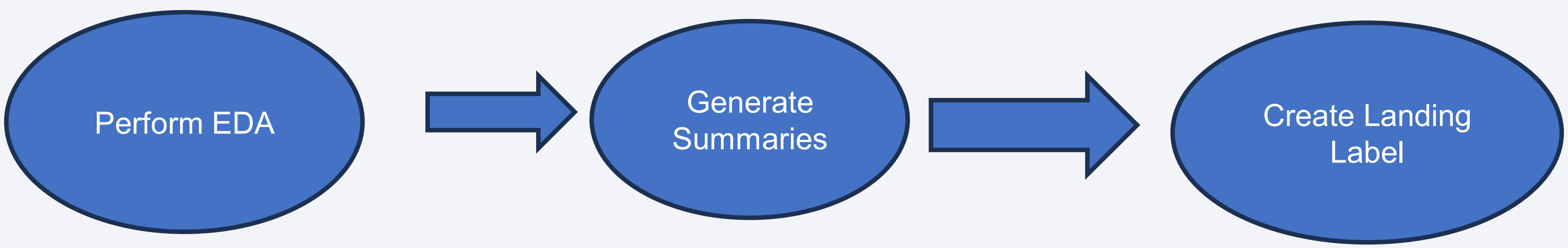
- Falcon 9 historical launch data were collected by performing web scraping a Wikipedia page
- Data was extracted with BeautifulSoup library and converted into a Pandas data frame
- Data was saved to a CSV file
- Completed notebook available at https://github.com/MagicMagic99/Webscraping-Michal/blob/main/jupyter-labs-webscraping-Michal.ipynb

Data Collection - Scraping

Step 1	Step 2	Step 3
Collect data from a Wikipedia page	Extract data using the BeautifulSoup library	Export data to a CSV file

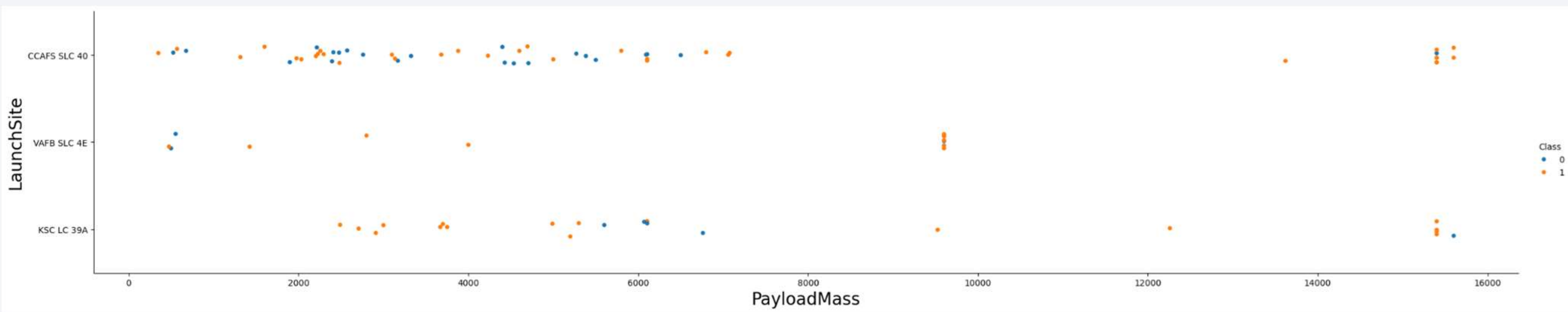
Data Wrangling

- Exploratory Data Analysis(EDA) was performed to find patterns in the data
- Number of launches on each site was determined, occurrences of each orbit and mission outcome per orbit were calculated
- A landing outcome label was created, representing the outcome of each launch
- Completed data wrangling related notebook available at https://github.com/MagicMagic99/Data-wrangling-Michal/blob/main/labs-jupyter-spacex-Data%20wrangling-Michal.ipynb



EDA with Data Visualization

- For data visualization, scatter point charts, bar chart and line chart were used
- The scatter point charts allowed to visualize the relationships between pair of values, e.g. payload mass and launch site



- Completed EDA with data visualization notebook available at https://github.com/MagicMagic99/edadataviz-Michal/blob/main/edadataviz-Michal.ipynb

EDA with SQL

- Set of SQL queries were performed to select:
    - Names of the unique launch sites in the space mission
    - Launch sites begin with the string 'CCA' limited to 5 records
    - 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 ground pad was achieved
    - Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
    - Total number of successful and failure mission outcomes
    - booster_versions that have carried the maximum payload mass
    - Records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
    - 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.
- Completed EDA with SQL notebook available at https://github.com/MagicMagic99/eda-sql-coursera_sqllite-Michal/blob/main/jupyter-labs-eda-sql-coursera_sqllite-Michal%20(1).ipynb

Build an Interactive Map with Folium

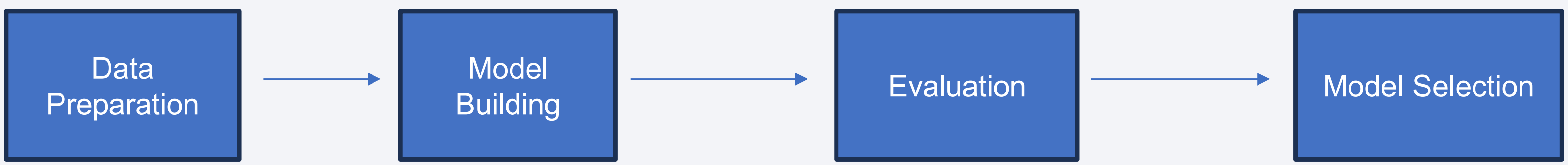
- Circles, markers, marker clusters and lines were used
  - Circle to highlight area with a text label on a specific coordinate
  - Marker to indicate a point on map
  - MarkerCluster to simplify a map containing many markers having the same coordinate
  - Lines to indicate distances between two coordinates
- Completed interactive map with Folium map available at https://github.com/MagicMagic99/folium_lab_launch_site_location-Michal/blob/main/lab_jupyter_launch_site_location-Michal.ipynb

Build a Dashboard with Plotly Dash

- Pie chart and scatter chart were used to visualize data
  - Pie chart to show the total successful launches count for a site or all sites
  - Scatter chart to show the correlation between payload and launch success
- Those plots allowed to identify the most successful launch sites according to payload mass
- Completed Plotly Dash lab available at https://github.com/MagicMagic99/spacex-dash-Michal/blob/main/spacex-dash-app%20(1).py

Predictive Analysis (Classification)

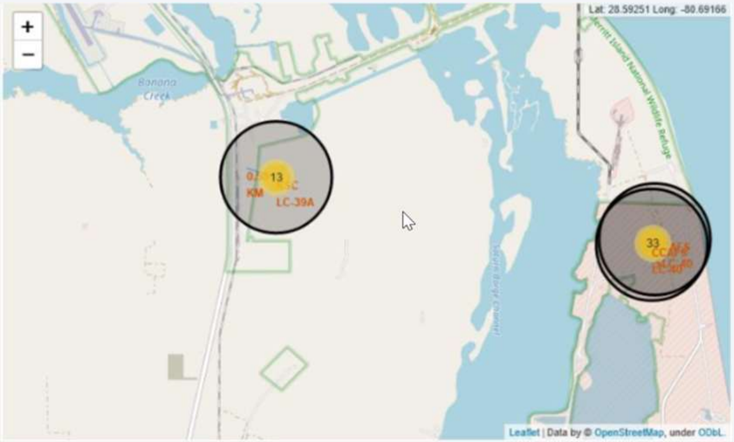
Building several classification models (Logistic Regression, Decision Trees, and SVM) using features like payload, orbit, and launch site. Each model was tuned and evaluated on test data with metrics such as accuracy and confusion matrices. After comparison, the best-performing model was selected for its balanced accuracy and ability to generalize.



https://github.com/MagicMagic99/SpaceX_Machine-Learning-Prediction_Part_5-Michal/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5-Michal.ipynb

Results

Exploratory Data Analysis
1.Lighter payloads → higher success rates.
2.Heavy payloads more likely to fail.
3.LEO orbits showed the strongest success ratios.
4.Success rates improved steadily year by year.
Interactive Analytics
5. Folium map displayed all launch sites with success/failure markers.
6. Zoomed maps showed sites in relation to highways, railways, and coastlines.
7. Plotly Dash dashboard allowed filtering by payload range, orbit, and site.
8. Interactive tools confirmed patterns seen in SQL queries and plots.
Predictive Analysis
9. Models tested: Logistic Regression, Decision Tree, SVM.
10. Best-performing model achieved high accuracy and balanced outcomes.
11. Key predictors: payload mass, orbit type, and launch site.

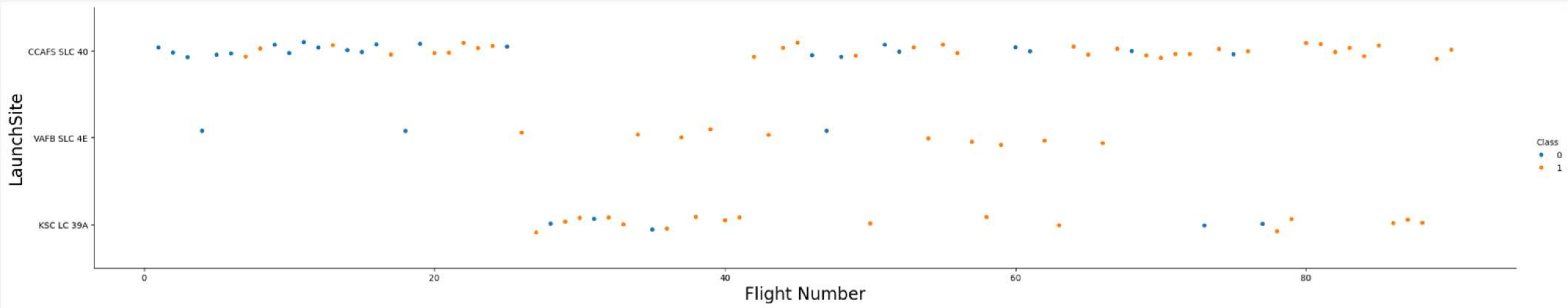


Section 2

Insights drawn from EDA

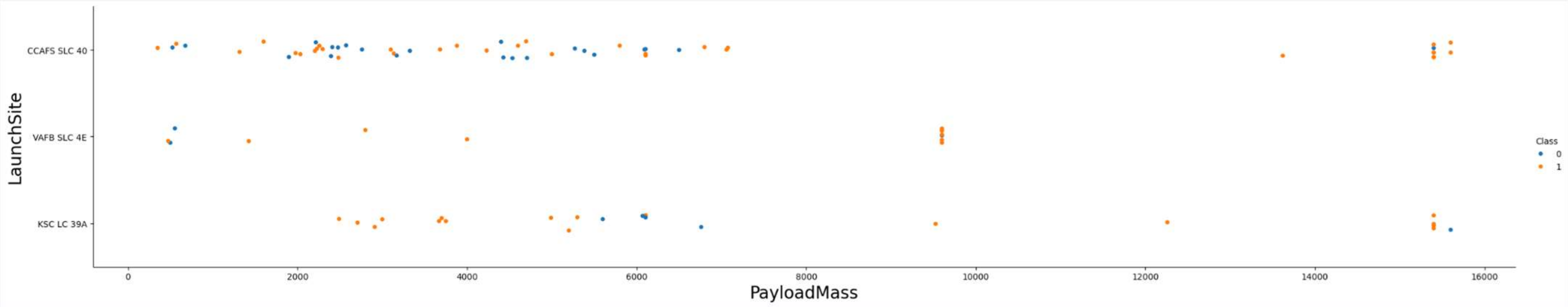
Flight Number vs. Launch Site

- The plot shows that the CCAF5 SLC 40 launch site had the highest number of successful launches recently
- The success ratio has improved over time, at the beginning majority of the launches were unsuccessful



Payload vs. Launch Site

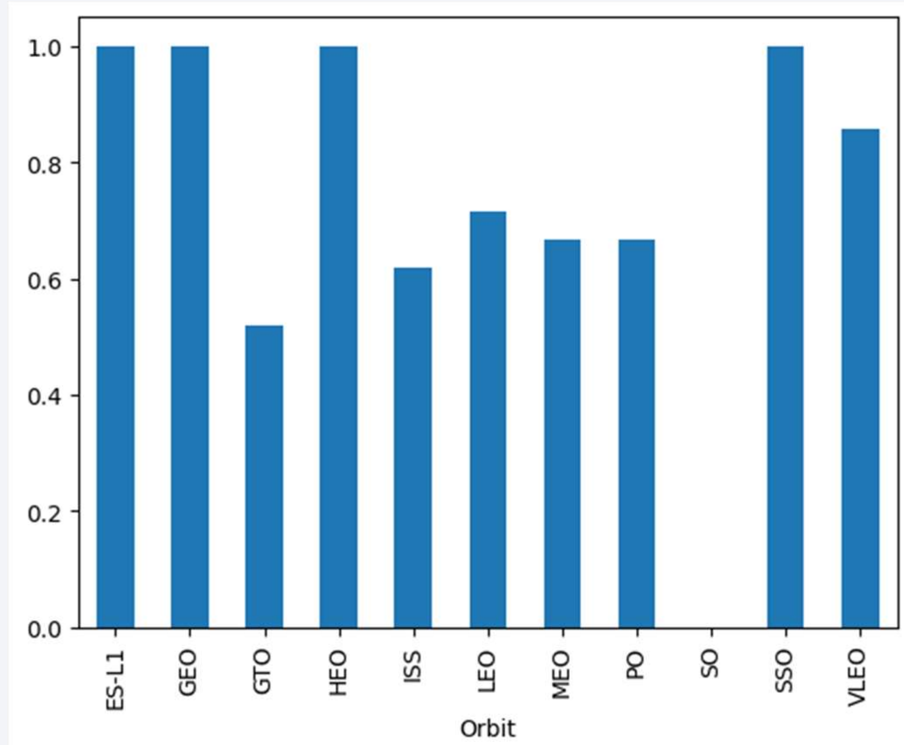
- Payloads around 4,000kg, 9,000kg and 16,000 kg have very good success rate on more than one launch site
- For the VAFB-SLC launchsite there are no rockets launched for heavy payload mass (greater than 10,000kg)



Success Rate vs. Orbit Type

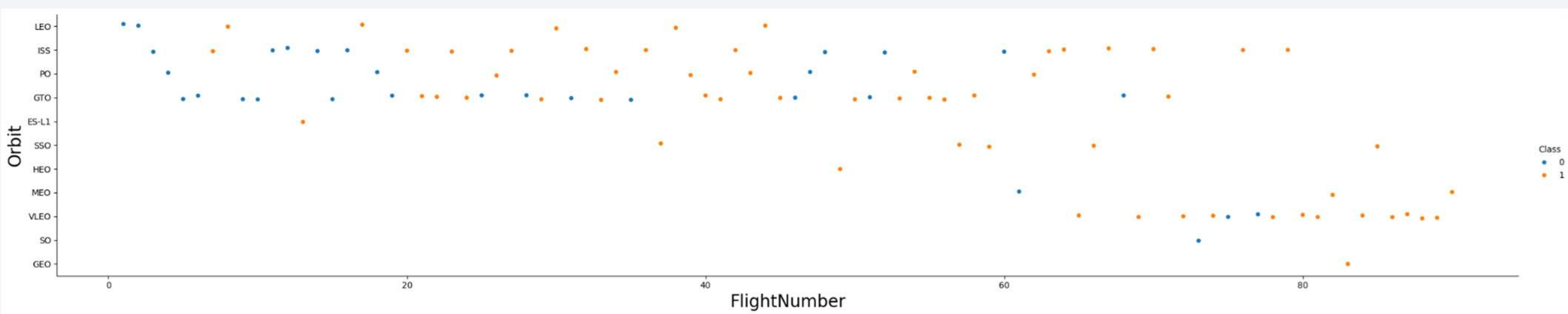
- The highest success rate were seen on:
  - ES-L1
  - GEO
  - HEO
  - SSO

- SO orbit success rate is 0 which means there were no launches related to that orbit type or all launches were unsuccessful



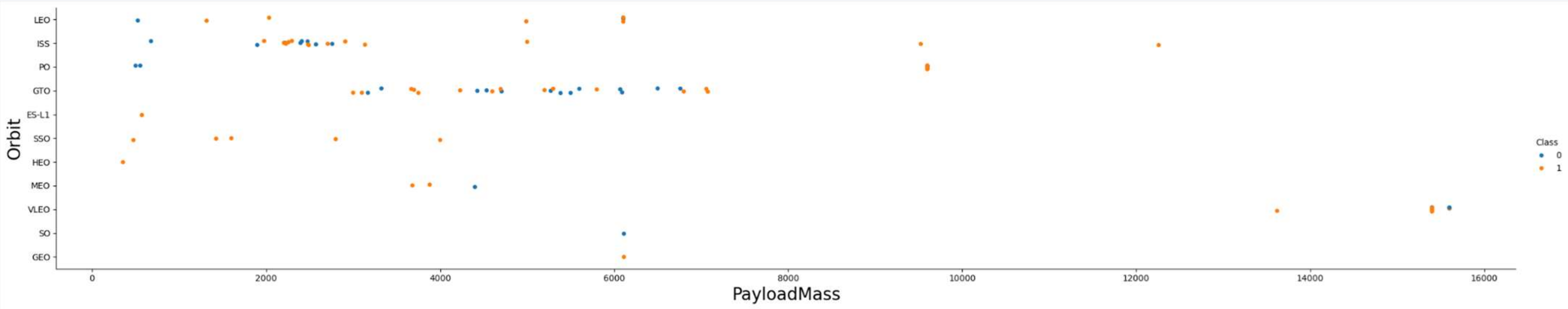
Flight Number vs. Orbit Type

- VLEO orbit was related to high number of recent successful flights
- Four orbit types (LEO, ISS, PO, GTO) were mainly picked for first 60 missions



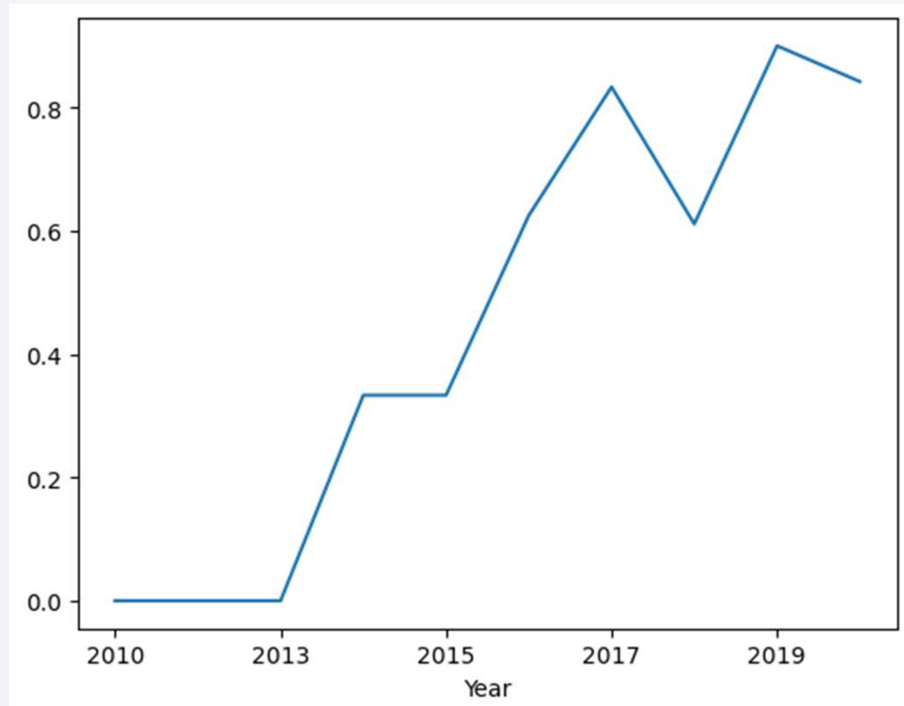
Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- No visible relation between payload and success rate to orbit GTO
- Impossible to indicate payload relation with success rate to some orbits, e.g. SO, GEO, due to few launches



Launch Success Yearly Trend

- No successful launches until 2013
- Success rate increased significantly from 2013 to 2017
- Decline in 2017 could indicate testing launches to new orbit types or different payloads



All Launch Site Names

- There are four lauching sites:

	Launch Site
0	CCAFS LC-40
1	CCAFS SLC-40
2	KSC LC-39A
3	VAFB SLC-4E

Listing the distinct launch sites shows all the locations SpaceX has used, and it forms the base for doing site-by-site comparisons later in the analysis.

Launch Site Names Begin with 'CCA'

- 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
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

- Filtering with the CCA% pattern pulls up the Cape Canaveral pads, which are some of the most frequently used launch sites in the dataset.

Total Payload Mass

- Total payload carried by boosters from NASA

TOTAL_PAYLOAD
111268

Adding up payload mass for NASA missions shows how much cargo the agency has launched on Falcon 9.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1

AVG_PAYLOAD
2928.4

Taking the average payload for the F9 v1.1 booster shows what that version could handle compared to newer ones.

First Successful Ground Landing Date

- First successful landing outcome on ground pad

FIRST_SUCCESS_GP
2015-12-22

Finding the date of the first successful ground landing is important because it marks the point when SpaceX proved boosters could be reused. This milestone helps set the timeline for how their reusability efforts developed.

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Checking for successful drone-ship landings in the 4–6 ton payload range shows how SpaceX managed heavier missions while still recovering boosters. It highlights the balance between carrying larger loads and maintaining reusability.

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

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

Counting all successful and failed launches gives a clear picture of overall performance. It shows how many missions went as planned and how many didn't, giving context for SpaceX's progress.

Boosters Carried Maximum Payload

- Names of the boosters which have carried the maximum payload mass

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

Listing boosters that carried the heaviest payloads showcases edge-case performance and informs payload-risk hypotheses.

2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

A focused 2015 slice for drone-ship failures, including booster versions and sites, helps diagnose early-phase learning areas.

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

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

Landing_Outcome	QTY
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

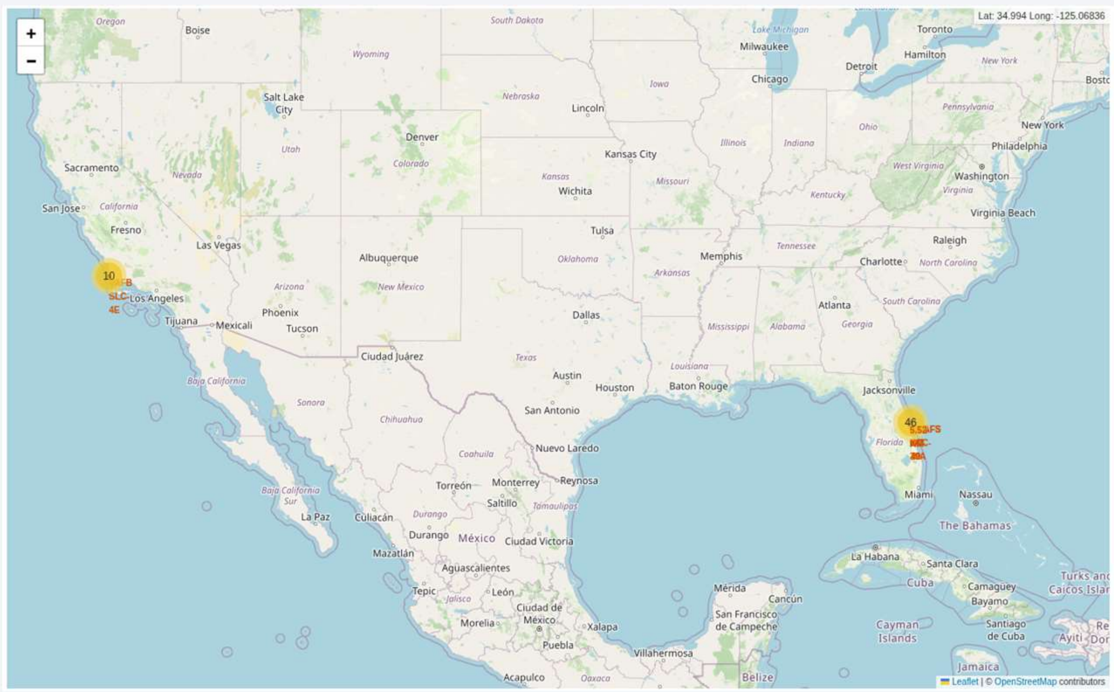
Ranking the landing outcomes from this period shows the balance between successes and failures in SpaceX's early years, giving context to how their recovery methods improved over time.

Section 3

Launch Sites Proximities Analysis

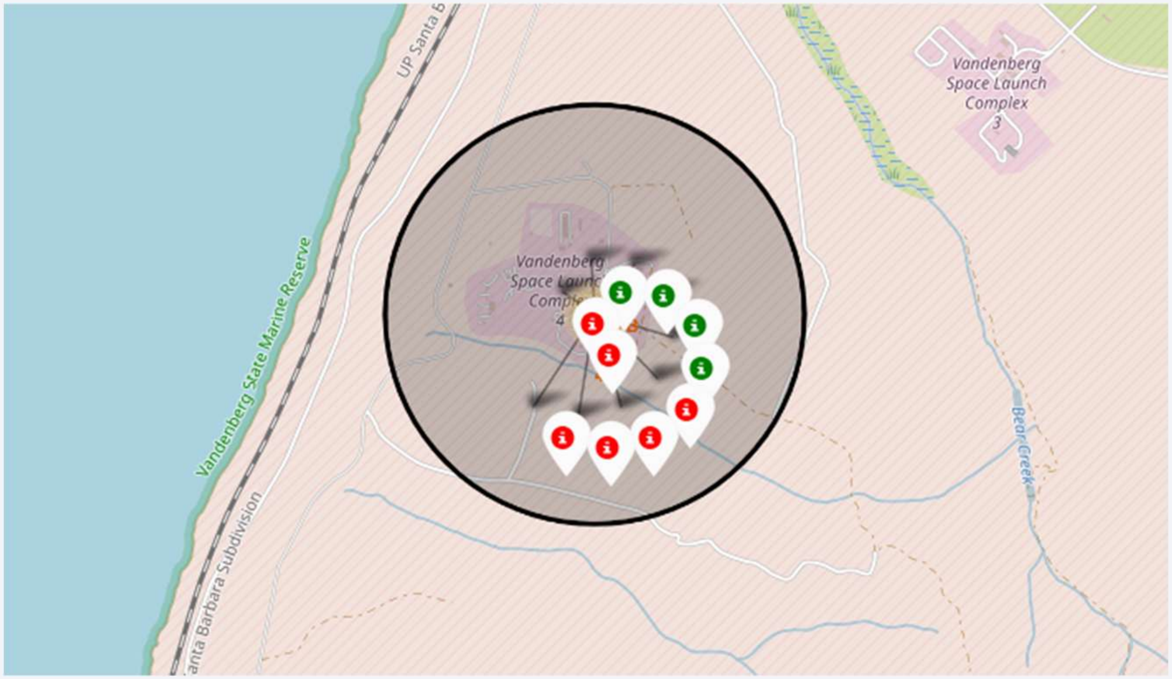
All launch sites

- Launch sites are located near sea for safety, as rocket failures are less catastrophic over open water, and for logistics, as rocket components are easier to transport by ships.

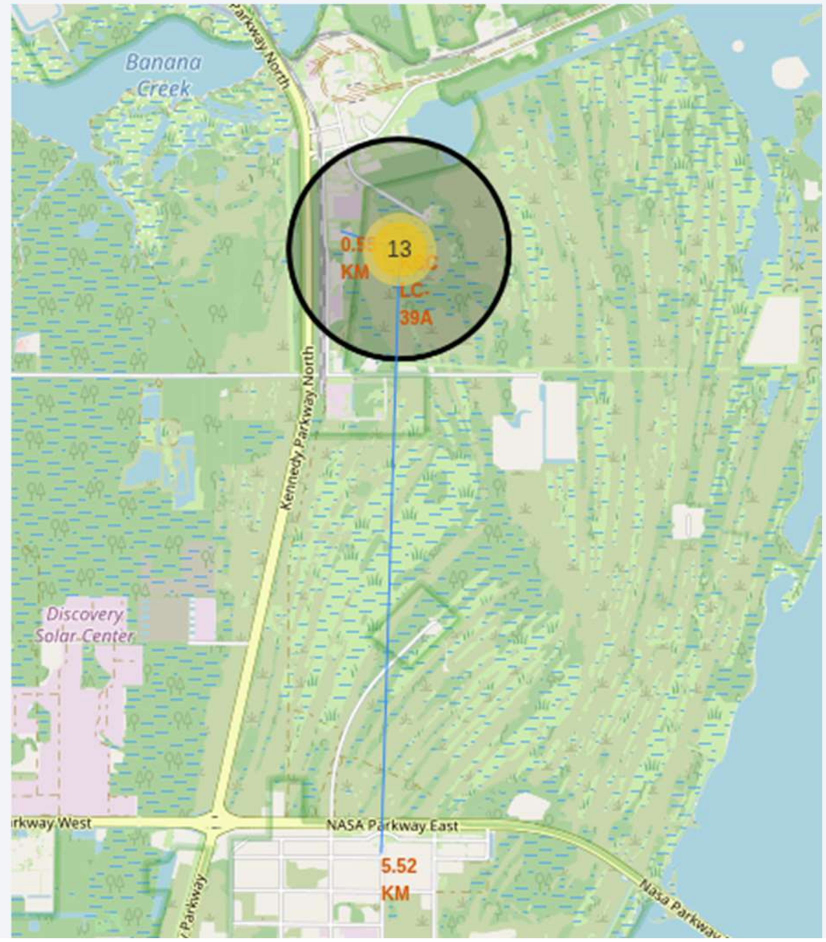


Launch outcomes by site

- Launch outcomes for VAFB SLC-4E launch site
- Green markers indicate successful launches and red markers indicate failures.



Logistics for KSC LC-39A launch site



Logistics for KSC LC-39A launch site

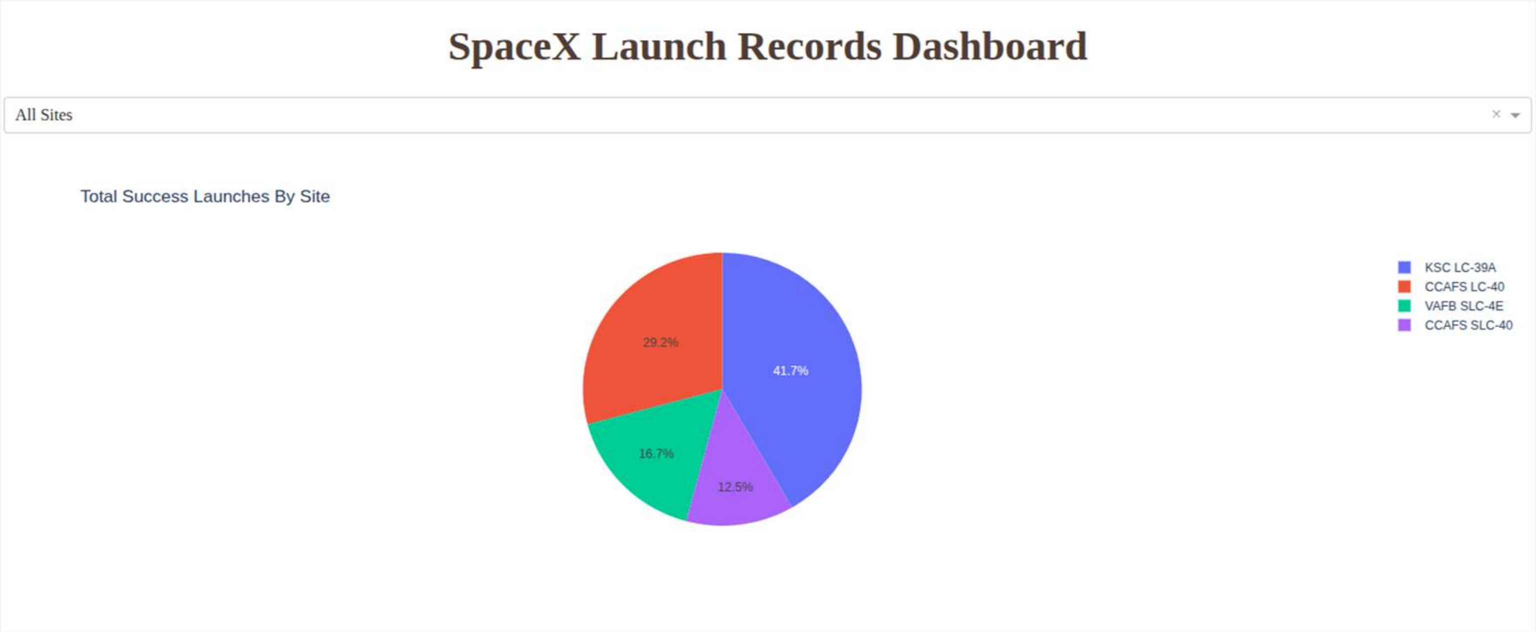
- Launch site KSC LC-39A is located near railroad
- Kennedy Parkway road facilitates the movement of equipment and personnel
- To ensure safety, the launch site location is relatively far from houses

Section 4

Build a Dashboard with Plotly Dash

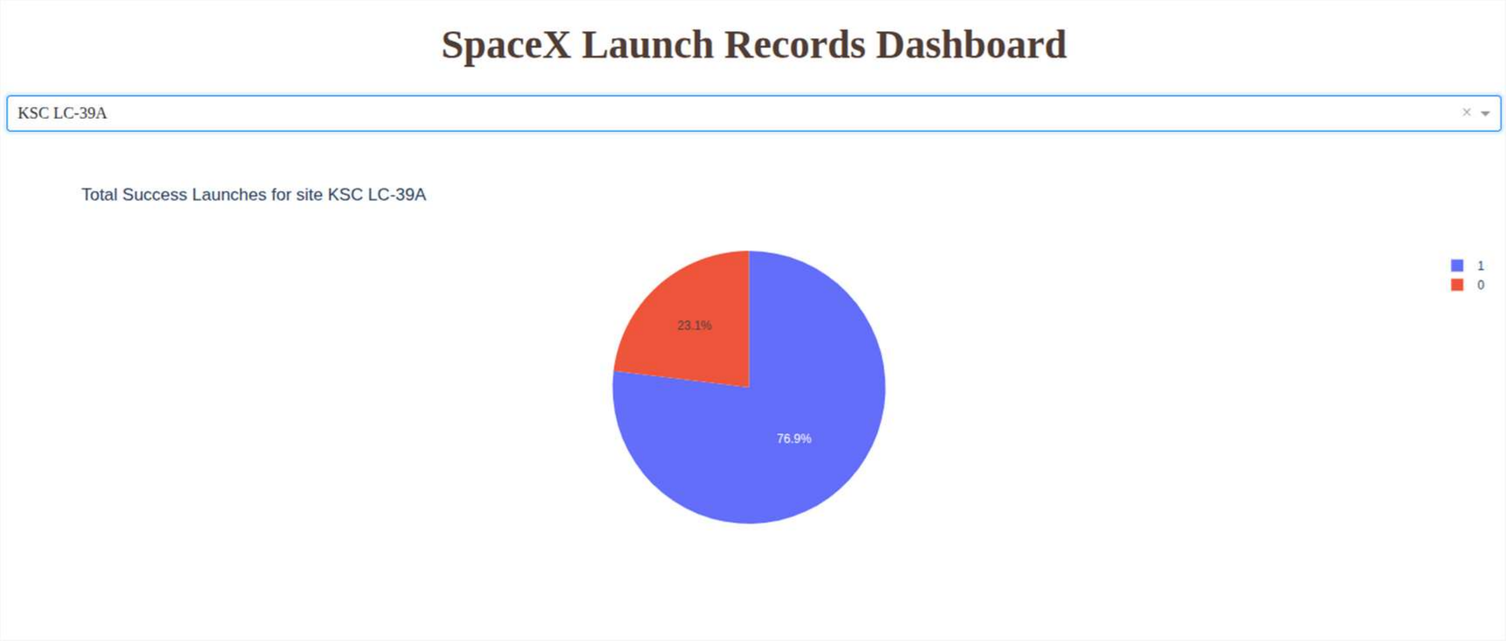
Successful launches for all sites

- Success of SpaceX rocket launch depends significantly on the launch site, success rate ranges from 12.5% to 41.7%



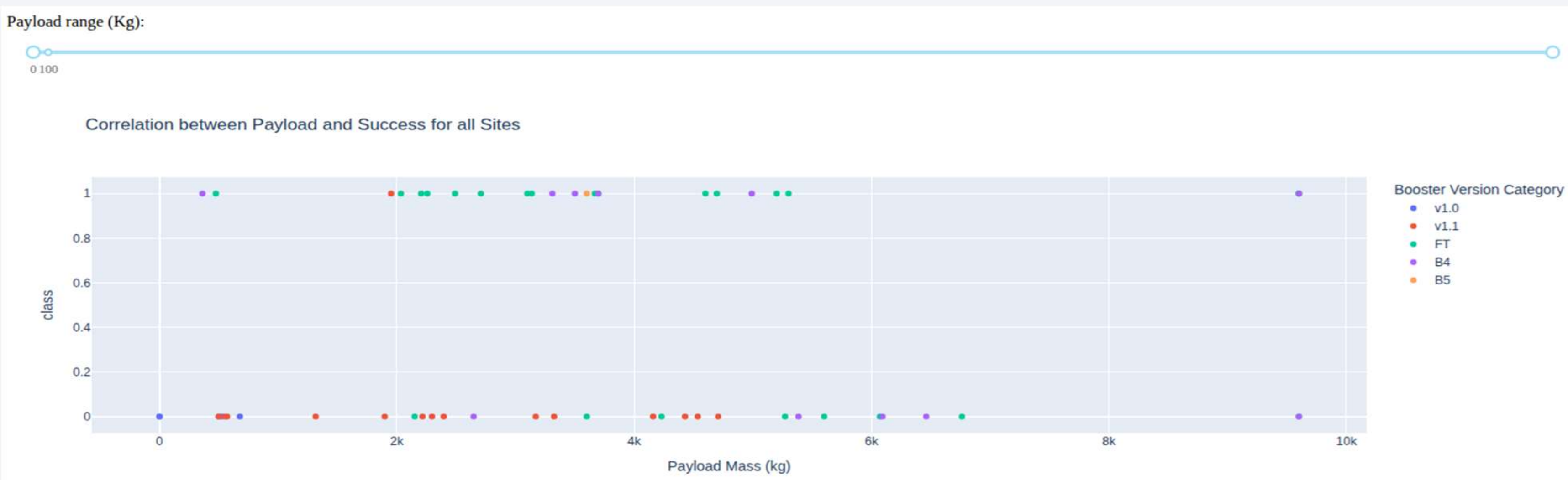
Launch site with highest launch success ratio

- Kennedy Space Center (KSC LC-39A) launch site exhibits the highest launch success ratio at 76.9 %



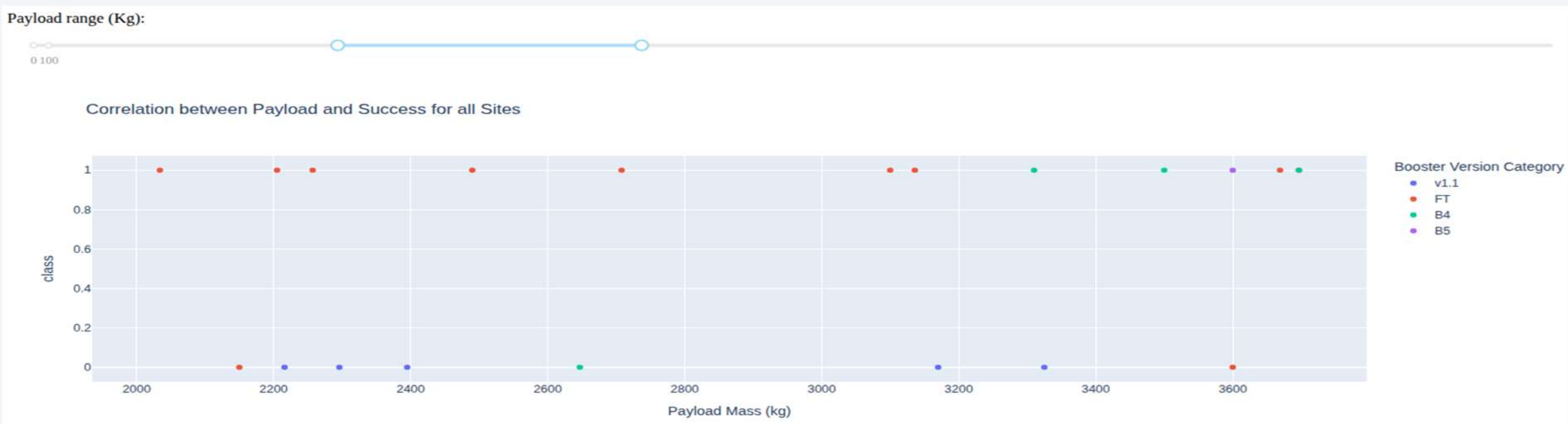
Payload vs launch outcome

- Launches with payloads in range 2,000-4,000kg with FT boosters were the most successful among different payload and booster combinations



Payload vs launch outcome

- For payloads over 2,000kg and under 3,200 kg FT boosters were the only successful ones
- All launches with v1.1 boosters ended up with failure

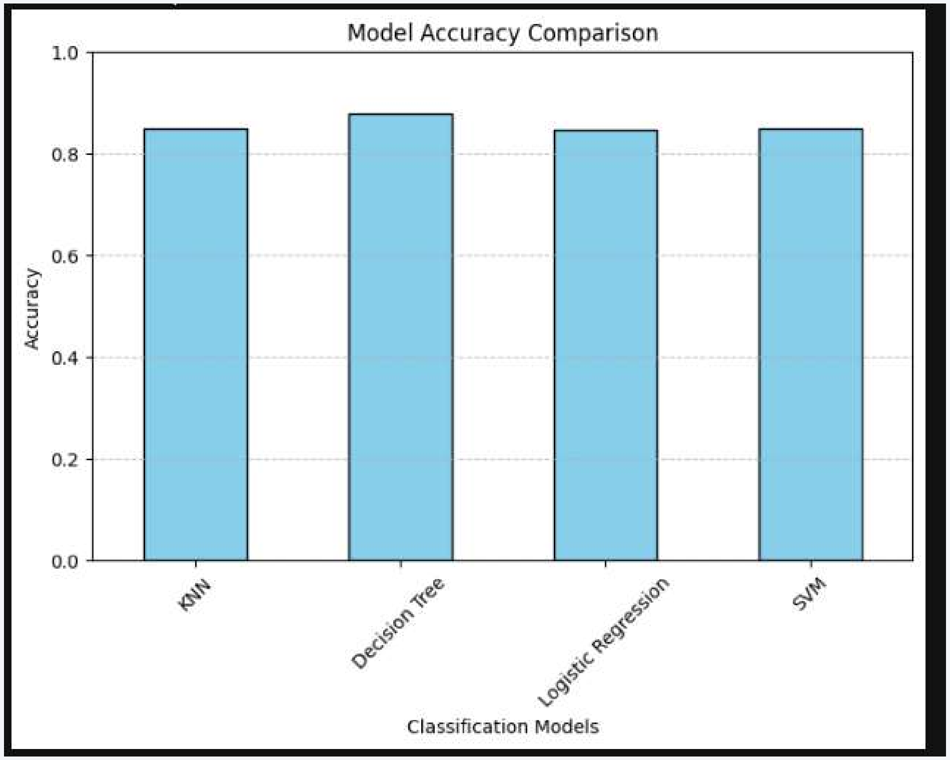


Section 5

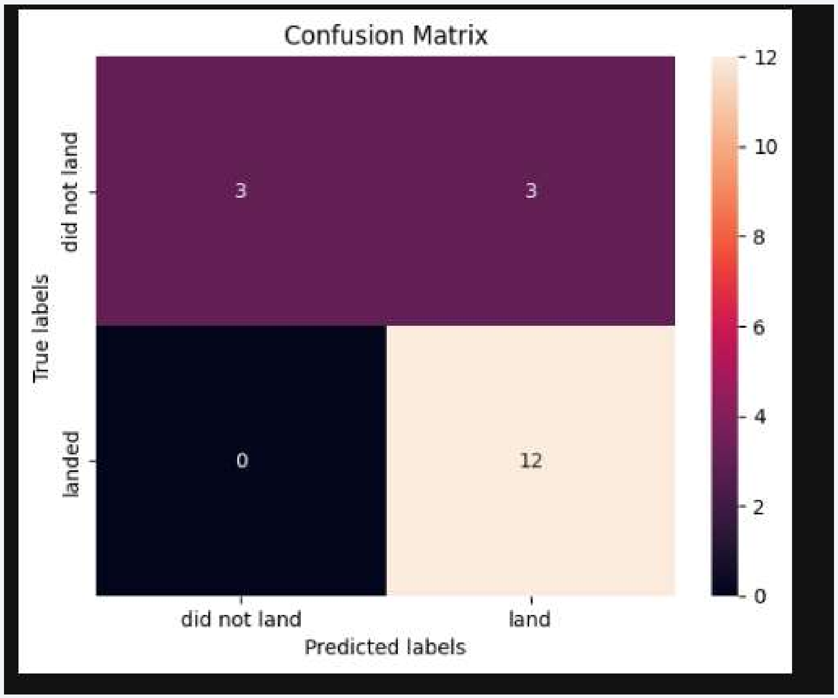
Predictive Analysis (Classification)

Classification Accuracy

- The method that performs best is " Decision Tree " with a score of 0.8767857142857143



Confusion Matrix



The confusion matrix shows balanced performance across classes, with most errors occurring in edge cases (e.g., heavy payloads or challenging orbits). This aligns with domain expectations and suggests useful decision support for planning.

Conclusions

- SpaceX’s launch data shows steady improvement in success rates year after year.
- Lighter payloads and certain orbits (like LEO) consistently had higher chances of success.
- Frequent use of the same launch sites led to more reliable results over time.
- Booster reusability played a big role in driving down failures and improving consistency.
- The predictive model confirmed these trends and proved useful for estimating future outcomes.
- Overall, the project shows how data science can explain past results and help forecast what’s ahead.


Thank you!