IBM Developer
SKILLS NETWORK

IBM DATA SCIENCE CAPSTONE PROJECT

Space X Falcon 9 Landing Analysis

Borja Canto
December 10th, 2024

Outline

Executive Summary

In this capstone project, we will predict if the SpaceX Falcon 9 first stage will land successfully using several machine learning classification algorithms.

Summary of Methodologies:

This project follows these steps:

- Data Collection
- Data Wrangling
- Exploratory Data Analysis
- Interactive Visual Analytics
- Predictive Analysis (Classification)

Summary of Results:

This project produced the following outputs and visualizations:

- Exploratory Data Analysis (EDA) results
- Geospatial analytics
- Interactive dashboard
- Predictive analysis of classification models

Introduction

SpaceX, a leader in the space industry, strives to make space travel affordable for everyone. Its accomplishments include sending spacecraft to the international space station, launching a satellite constellation that provides internet access and sending manned missions to space.

SpaceX can do this because the rocket launches are relatively inexpensive ($62 million per launch) due to its novel reuse of the first stage of its Falcon 9 rocket. Other providers, which are not able to reuse the first stage, cost upwards of $165 million each.

Other companies launch cost
$165 million

SpaceX launch cost
$62 million

SpaceX launches
119
in 2024

NASA launches
135
Since 1994

Introduction

By determining if the first stage will land, we can determine the price of the launch. To do this, we can use public data and machine learning models to predict whether SpaceX – or a competing company – can reuse the first stage.

Problems we want to find answers

What are the main characteristics of a successful or failed landing ?

What are the conditions which will allow SpaceX to achieve the best landing success rate ?

How can we accurately predict the landing outcome using machine learning models?

Section 1

Methodology

Methodology

The overall methodology includes:

1. **Collect** data using SpaceX REST API and web scraping techniques
2. **Wrangle** data by filtering the data, handling missing values and applying one hot encoding, to prepare the data for analysis and modeling.
3. **Explore** data via EDA with SQL and data visualization techniques
4. **Visualize** the data using Folium and Plotly Dash
5. **Build Models** to predict landing outcomes using classification models. Tune and evaluate models to find best model and parameters

Data Collection

The SpaceX API gave the primary data resource for the analysis, but an initial data collection process was made using Falcon 9 data available on Wikipedia.

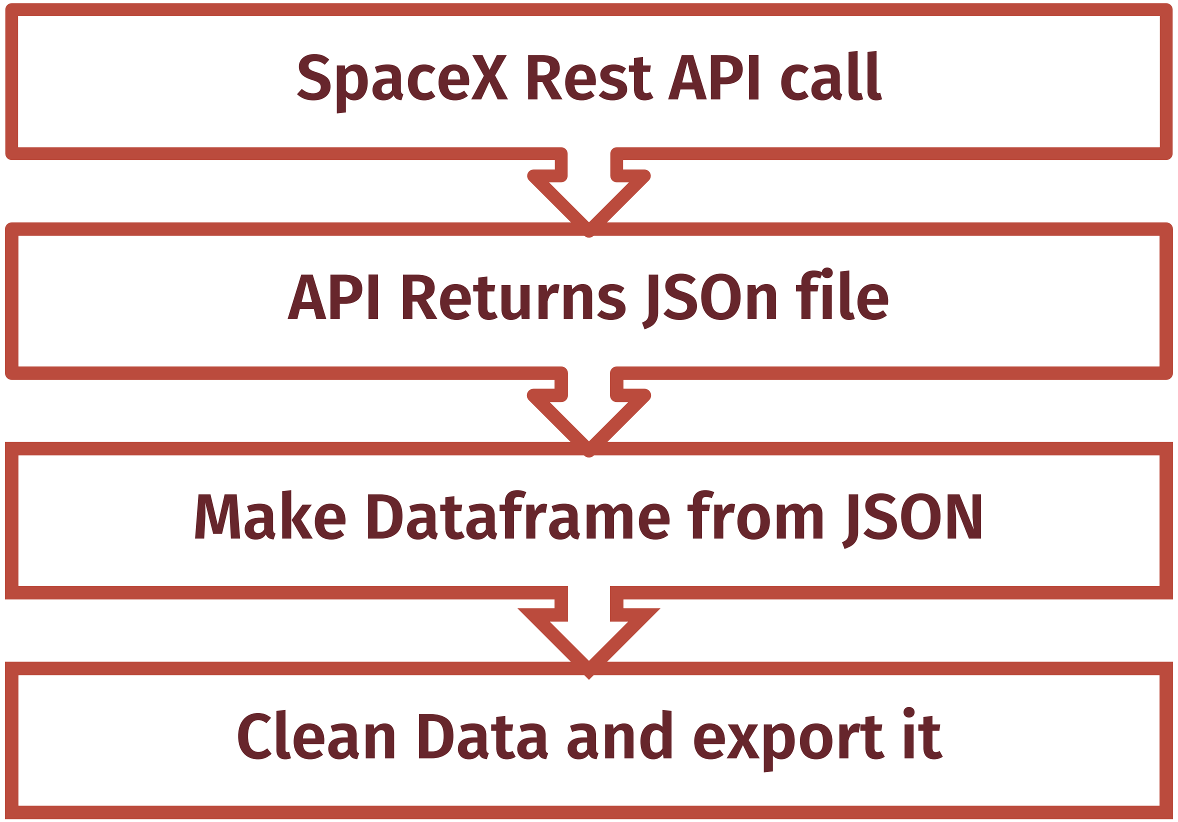
Data sets were collected:

- from Space X API (https://api.spacexdata.com/v4/rockets/)
- from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

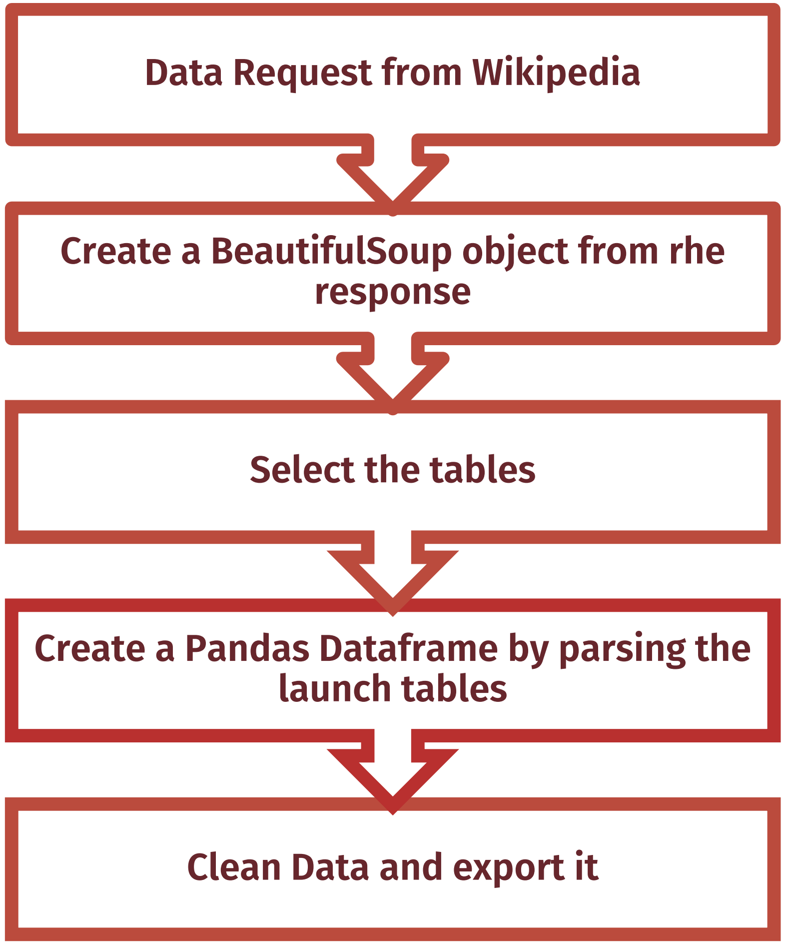
Although the data collection processes were significantly different, a general schema can be considered, which is described in the following flowchart

Data Collection – SpaceX API

- The API used is https://api.spacexdata.com/v4/rockets .
- The API provides data about many types of rocket launches done by SpaceX, the data is therefore filtered to include only Falcon 9.
- Data was extracted from the response from the API and loaded into a Pandas



Data Collection – Scraping

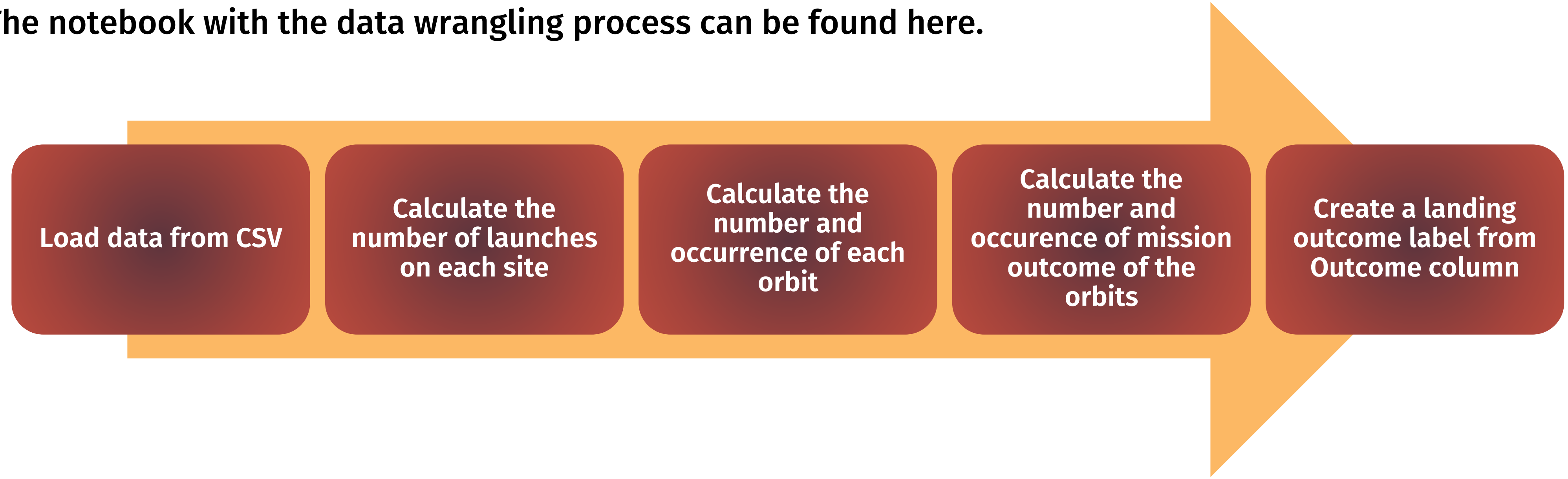


- The data is scraped from https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- The website contains only the data about Falcon 9 launches.
- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

Data Wrangling

GitHub

- The CSV file from the SpaceX API data collection contained the data in need of cleaning/wrangling.
- The launch sites, orbit types and mission outcomes were processed and reformatted.
- The mission outcome types were converted to a binary classification (onehot encoding) where 1 represented the Falcon 9 first stage landing being a success and 0 represented a failure.
- The new mission outcome classification column was added to the DataFrame.
- Also, from the analysis was possible to identify the number of launches per site, the name and number of orbits in which the launch was made, and the types of possible outcomes for a Falcon 9 mission.
- The notebook with the data wrangling process can be found here.



EDA with Data Visualization

GitHub

Exploratory Data Analysis (EDA) involves visually exploring and summarizing the main characteristics of a dataset. The goal is to understand the data's distribution, identify patterns, and uncover relationships between variables.

Scatter Graphs:

Scatter charts are useful to observe relationships, or correlations, between two numeric variables.

- Flight Number and Launch Site
- Payload and Launch Site
- Orbit Type and Flight Number
- Payload and Orbit Type

Bar Graph:

Bar charts are used to compare a numerical value to a categorical variable.

- Success rate vs. Orbit

Line Graph:

Line charts contain numerical values on both axes, and are generally used to show the change of a variable over time.

- Success rate vs. Year

EDA with SQL

GitHub

To gather some information about the dataset, some SQL queries were performed.

The SQL queries performed on the data set were used to:

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display the average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome on a ground pad was achieved
- List the names of the boosters which had success on a drone ship and a payload mass between 4000 and 6000 kg
- List the total number of successful and failed mission outcomes
- List the names of the booster versions which have carried the maximum payload mass
- List the failed landing outcomes on drone ships, their booster versions, and launch site names for 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

Build an Interactive Map with Folium

These objects are created in order to understand better the problem and the data. We can show easily all launch sites, their surroundings and the number of successful and unsuccessful landings.

Markers Indicating Launch Sites

- Added blue circle at NASA Johnson Space Center's coordinate with a popup label showing its name using its latitude and longitude coordinates
- Added red circles at all launch sites coordinates with a popup label showing its name using its name using its latitude and longitude coordinates

Colored Markers of Launch Outcomes

- Added colored markers of successful (green) and unsuccessful (red) launches at each launch site to show which launch sites have high success rates
- Distances Between a Launch Site to Proximities
- Added colored lines to show distance between launch site CCAFS SLC40 and its proximity to the nearest coastline, railway, highway, and

Build a Dashboard with Plotly Dash

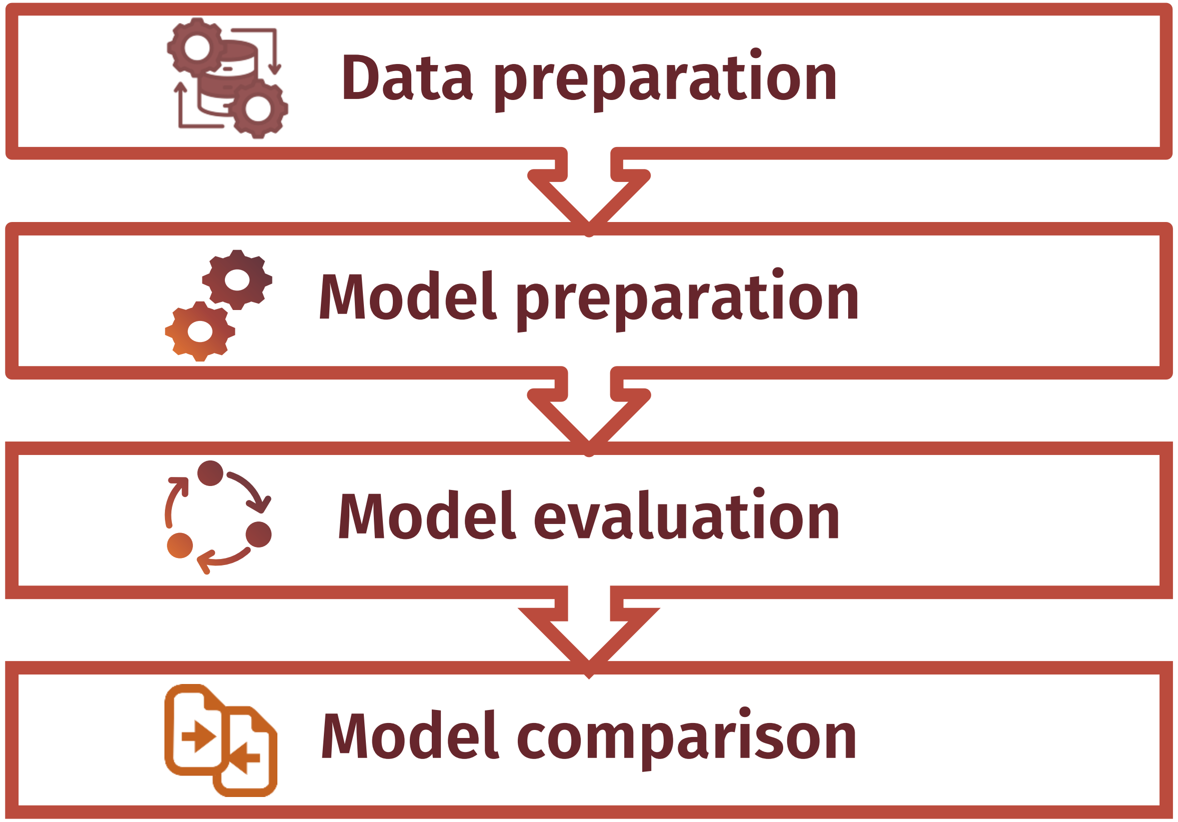
GitHub

A plotly dashboard with interactive elements were made to have an interactive visualization of the data.

1. **Dropdown List with Launch Sites**
   - Enables users to select specific launch sites for analysis.
   - Facilitates filtering and focused exploration based on geographical locations.
2. **Slider of Payload Mass Range**
   - Allows users to adjust payload mass ranges dynamically.
   - Offers flexibility in examining launch success concerning payload mass variations.
3. **Pie Chart Showing Successful Launches**
   - Displays the distribution of successful and failed launches.
   - Helps visualize the overall success rate and performance trends.
4. **Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version**
   - Shows the relationship between payload mass and launch success.
   - Allows users to explore how payload mass influences mission outcomes.

Predictive Analysis (Classification)

1. Standardize the data with StandardScaler. Fit and transform the data
2. The dataset was split into training and testing sets.
3. The following machine learning models were trained on the training data set:
   1. Logistic Regression
   2. SVM (Support Vector Machine)
   3. Decision Tree
   4. KNN (k-Nearest Neighbors)
4. Hyper-parameters were evaluated using GridSearchCV() and the best was selected using the best_params method.
5. Using the best hyper-parameters, each of the four models were scored on accuracy by using the testing data set



Results Summary

Exploratory Data Analysis

- We can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.
- Orbits ES-L1, GEO, HEO and SSO have a 100% success rate
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS

Visual Analytics

- Most launch sites are near the equator, and all are close to the coast
- Launch sites are far enough away from anything a failed launch can damage (city, highway, railway), while still close enough to bring people and material to support launch activities
- KSC LC-39A has the highest success rate among landing sites

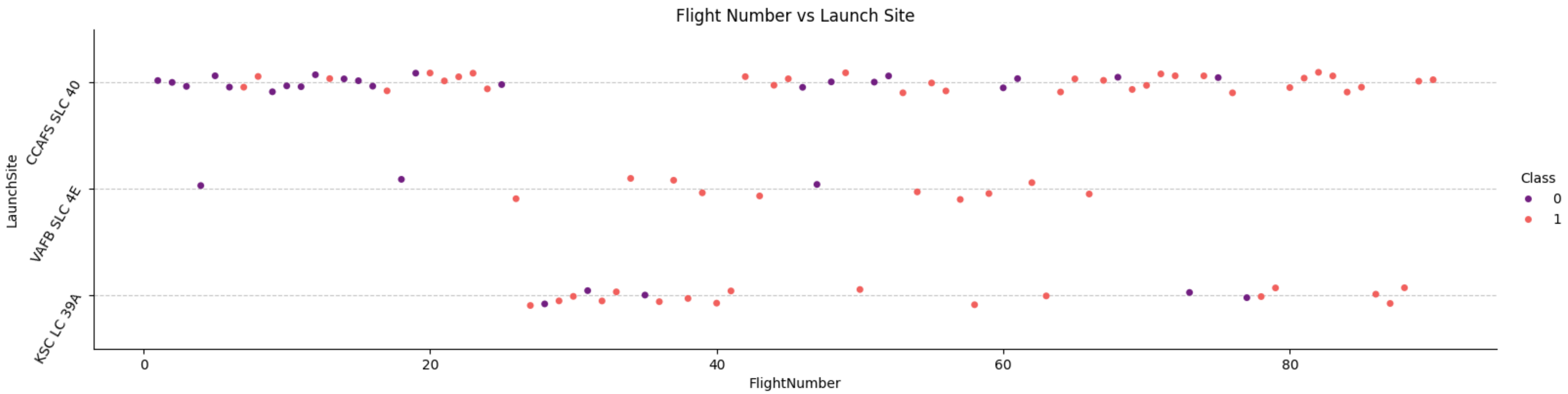
Predictive Analytics

- The results are practically the same. This is because the dataset is small and having lesser values.

Section 2

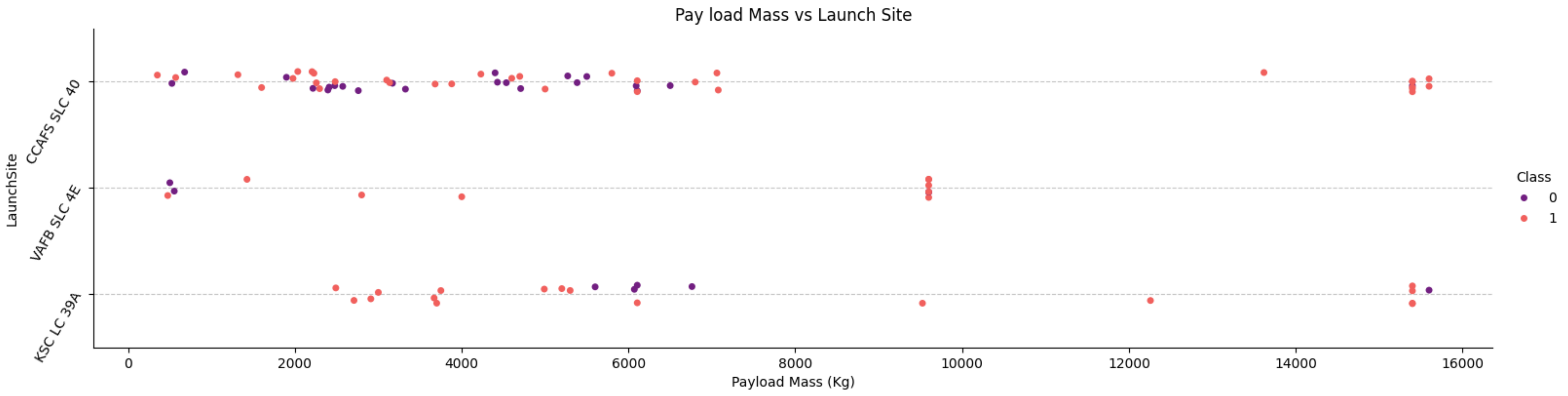
Insights drawn from EDA

Flight Number vs. Launch Site



- As the number of flights increases, the rate of success at a launch site increases.
- Most of the early flights (flight numbers < 30) were launched from CCAFS SLC 40, and were generally unsuccessful.
- The flights from VAFB SLC 4E also show this trend, that earlier flights were less successful.
- No early flights were launched from KSC LC 39A, so the launches from this site are more successful.
- Above a flight number of around 30, there are significantly more successful landings (Class = 1).

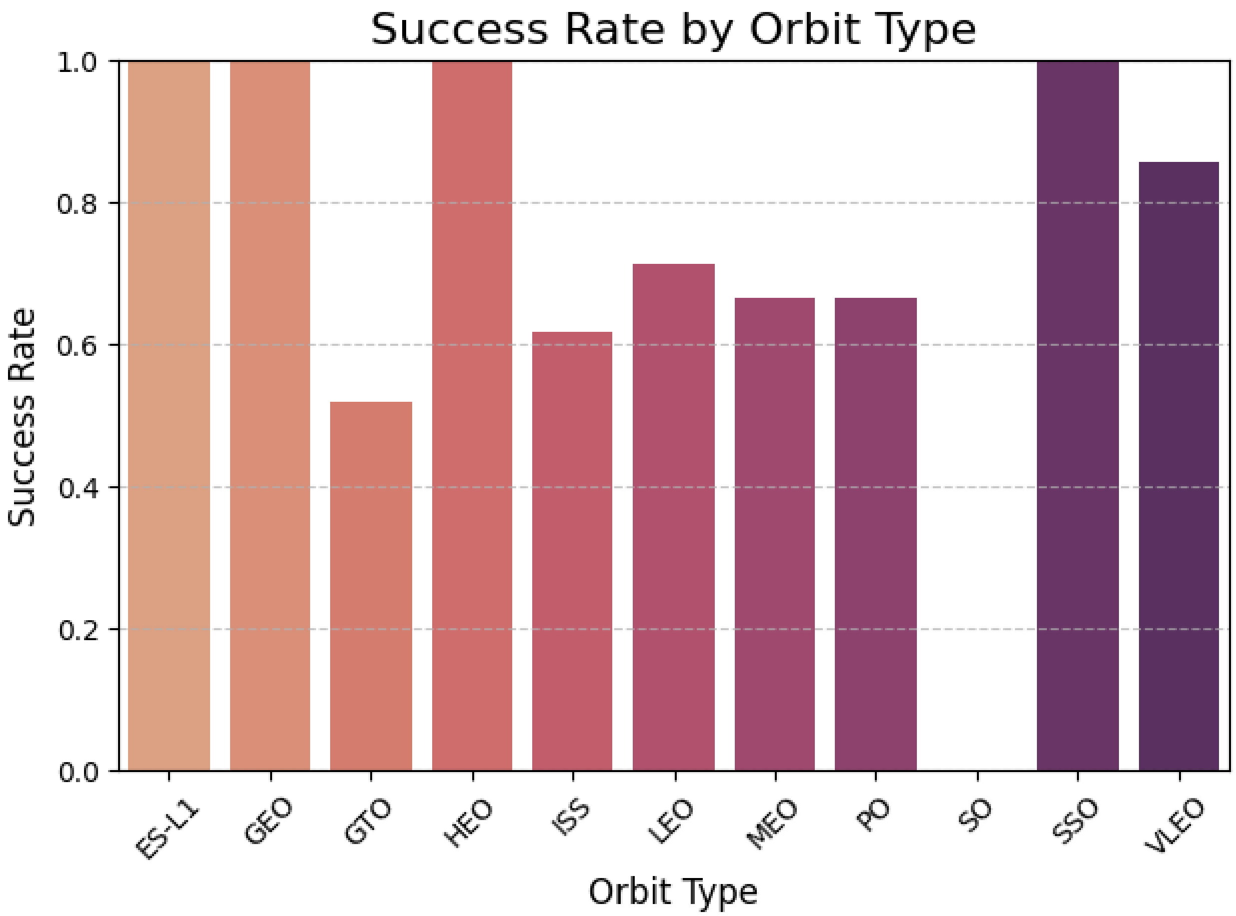
PayLoad Mass vs. Launch Site



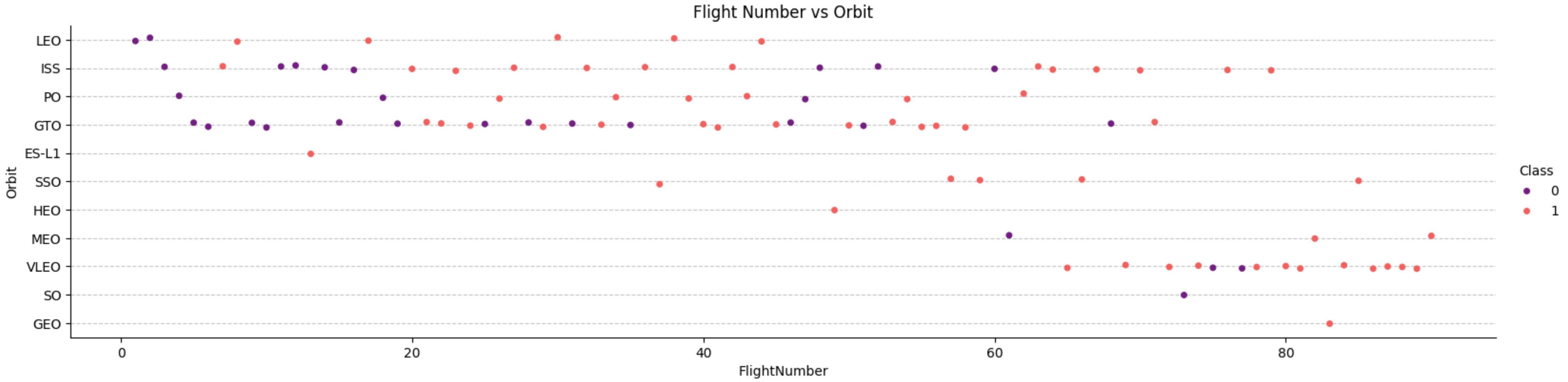
- Above a payload mass of around 7000 kg, there are very few unsuccessful landings, but there is also far less data for these heavier launches.
- VAFB SKC 4E has not launched anything greater than ~10,000 kg
- All sites launched a variety of payload masses, with most of the launches from CCAFS SLC 40 being comparatively lighter payloads (with some outliers
- KSC LC 39A has a 100% success rate for launches less than 5,500 kg

Success Rate vs. Orbit Type

- **100% success rate:**
  - ES-L1 (Earth-Sun First Lagrangian Point)
  - GEO (Geostationary Orbit)
  - HEO (High Earth Orbit)
  - SSO (Sun-synchronous Orbit)
- **80%-99% Success rate:**
  - VLEO (Very Low Earth Orbit)
- **50%-80% success rate:**
  - GTO (Geosynchronous Orbit)
  - ISS (habitable artificial satellite in low Earth Orbit)
  - LEO (Low Earth Orbit)
  - MEO (Medium Earth orbit)
  - PO (Polar Orbit)
- **0% Success rate:**
  - SO (Heliocentric Orbit)



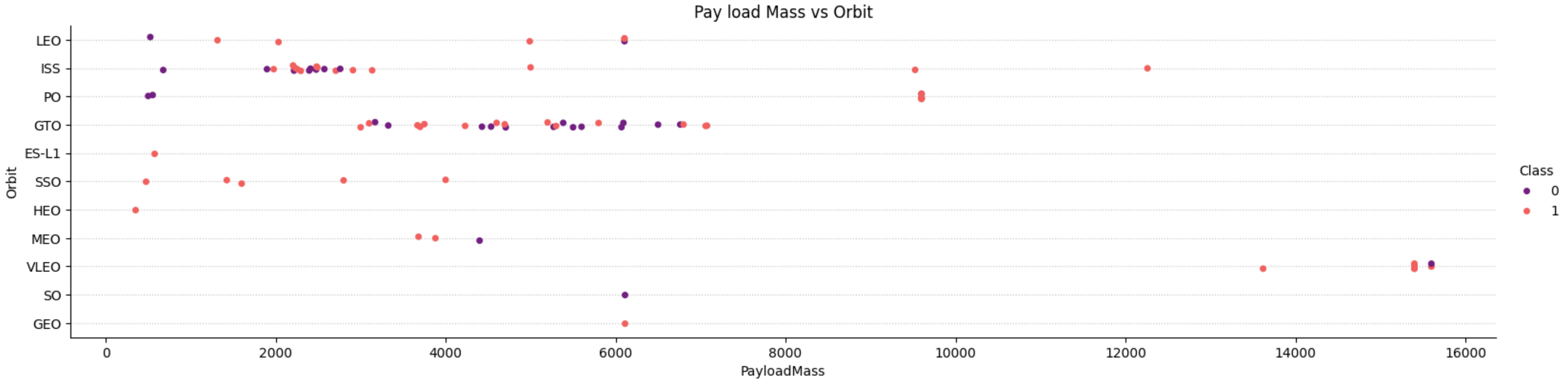
Flight Number vs. Orbit Type



This scatter plot of Orbit Type vs. Flight number shows a few useful things that the previous plots did not, such as:

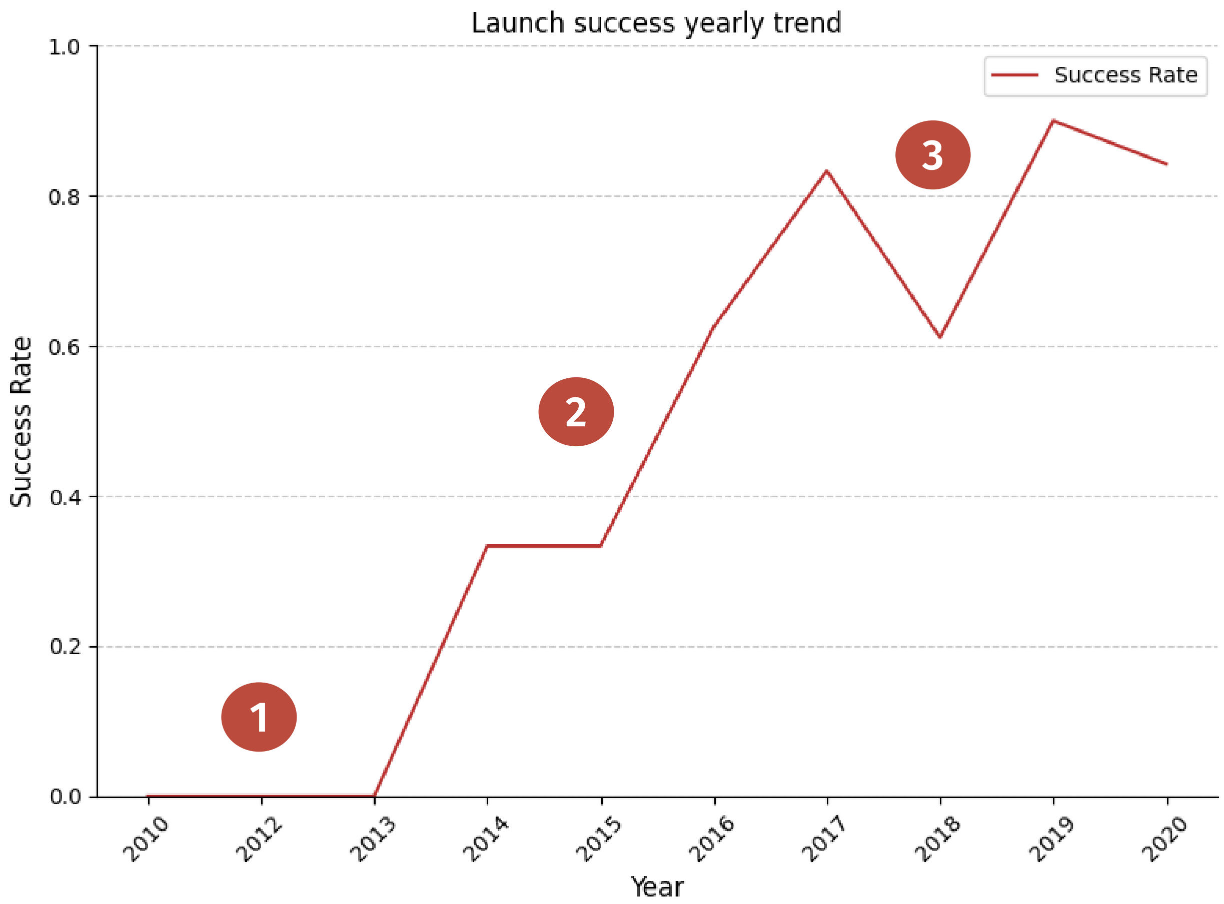
- **The 100% success rate of GEO, HEO, and ES-L1 orbits can be explained by only having 1 flight into the respective orbits.**
- **The 100% success rate in SSO is more impressive, with 5 successful flights.**
- **Generally, as Flight Number increases, the success rate increases. This is most extreme for LEO, where unsuccessful landings only occurred for the low flight numbers (early launches).**

Payload vs. Orbit Type



- PO, ISS, LEO orbit types have more success with heavy payloads
- For GTO, the relationship between payload mass and success rate is unclear.
- VLEO (Very Low Earth Orbit) launches are associated with heavier payloads, which makes intuitive sense.

Launch Success Yearly Trend



1. Between 2010 and 2013, all landings were unsuccessful (as the success rate is 0).
2. After 2013, the success rate generally increased, despite small dips in 2018 and 2020.
3. After 2016, there was always a greater than 50% chance of success.

Launch Site Information

There are four unique launch sites.

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

They are obtained by selecting unique occurrences of "launch_site" values from the dataset.

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

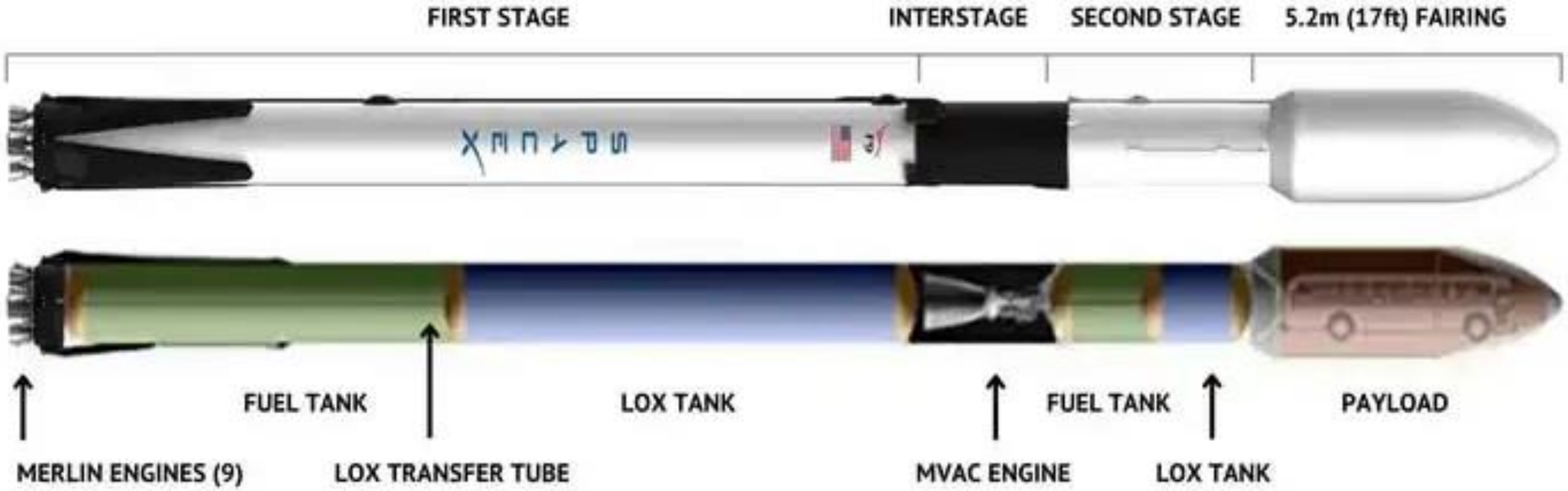
Payload Mass

total payload mass carried by NASA (CRS) boosters

45,596Kg

average Payload Mass by F9 v1.1

2,534Kg



Landing & Mission Info

Successful Drone Ship Landing with Payload between 4000 and 6000 Kg

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

Selecting booster versions according to the filters above, these 4 are the result.

first successful landing in ground pad

December 22th, 2015

Successful and Failure Mission Outcomes

- 99 Success
- 1 Success (payload status unclear)
- 1 Failure in Flight

Grouping mission outcomes and counting records for each group led us to the summary above.

Failed Landings on Drone Ship in 2015

Month	Booster_Version	Launch_Site	Date	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	2015-01-10	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	2015-04-14	Failure (drone ship)

Successful Drone Ship Landing with Payload between 4000 and 6000 Kg

- 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

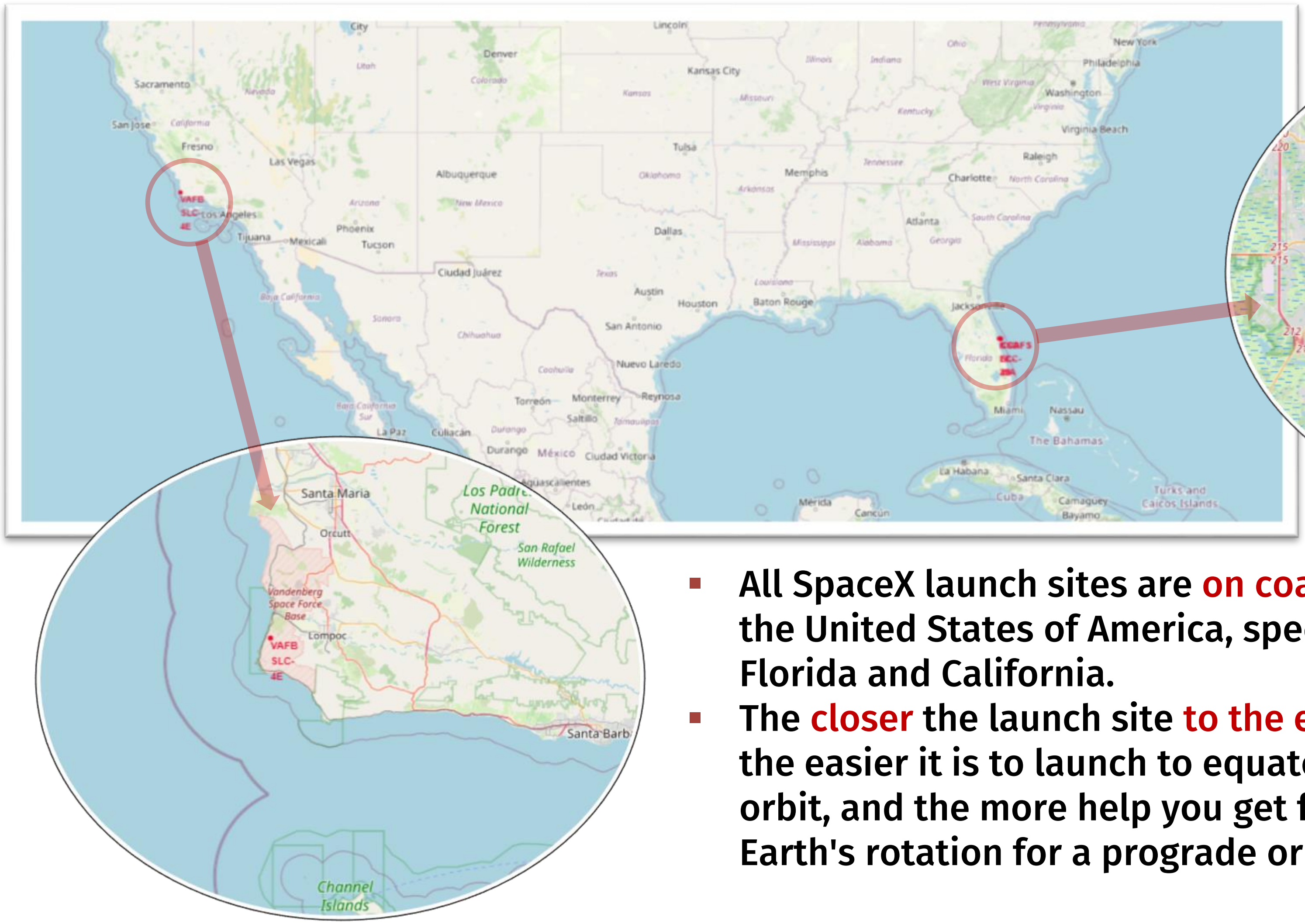
Count of landing outcomes between 2010-06-04 and 2017-03-20

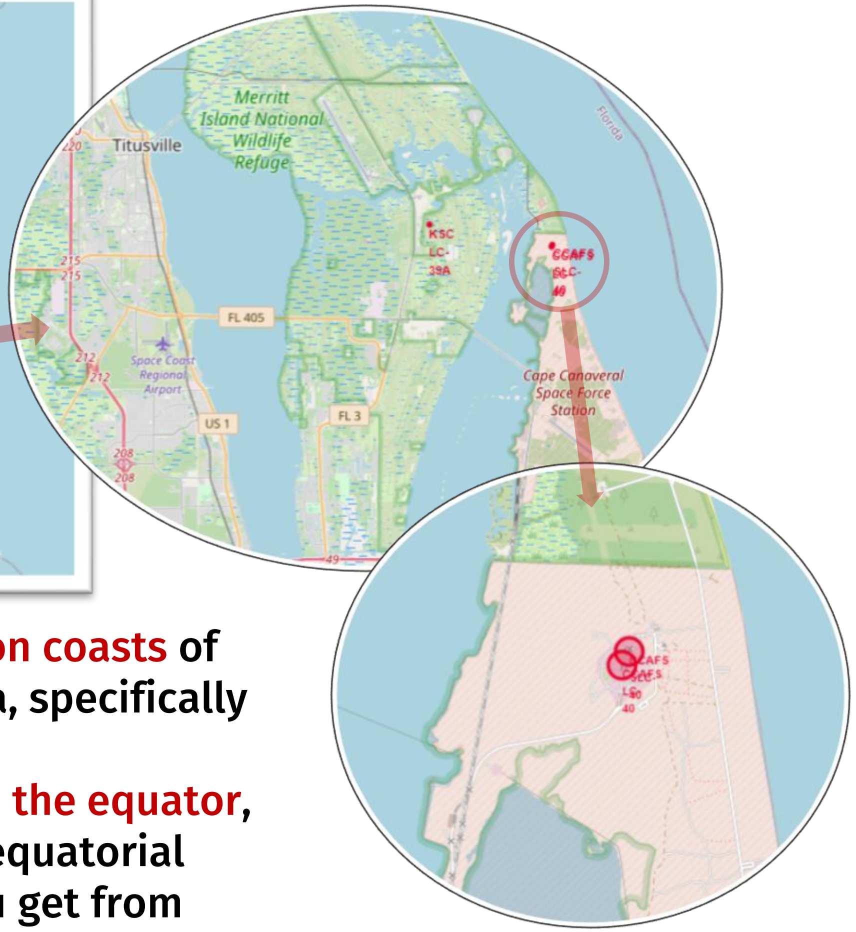
Landing Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Site Analysis

Falcon 9 Launch Site Locations





- All SpaceX launch sites are on coasts of the United States of America, specifically Florida and California.
- The closer the launch site to the equator, the easier it is to launch to equatorial orbit, and the more help you get from Earth's rotation for a prograde orbit.

Launch Outcomes

- The markers display the mission outcomes (Success/Failure) for Falcon 9 first stage landings. They are grouped on the map to be associated with the geographical coordinates for the launch.
- We note that KSC LC-39A has a higher launch success rate.

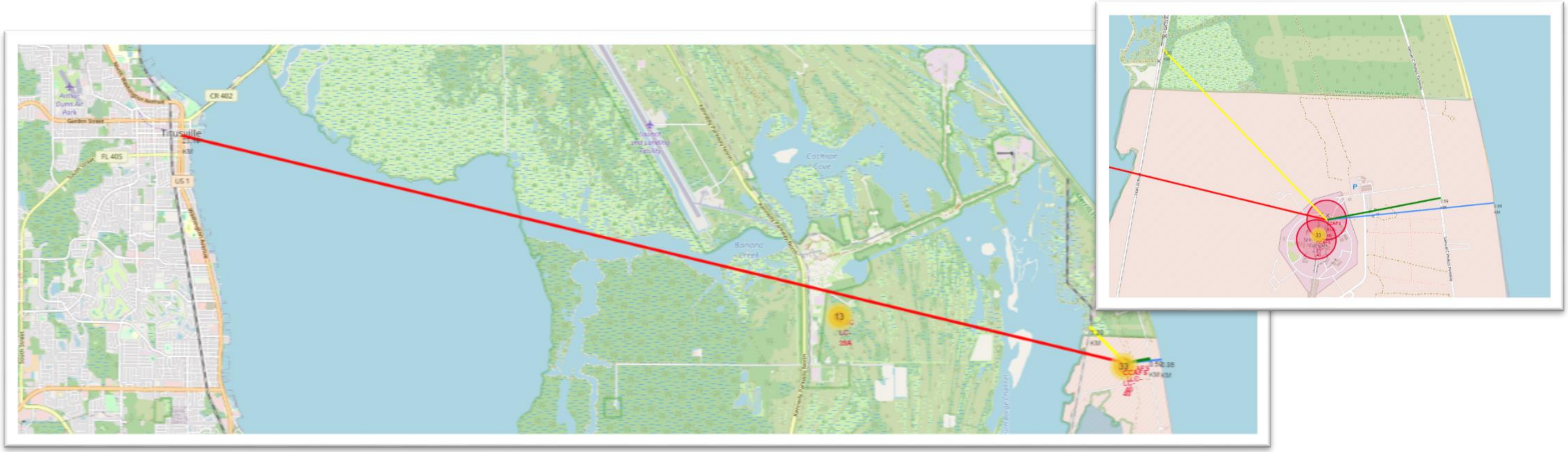
VAFB SLC-4E

KSC LC-39A

CCAFS LC-40

CCAFS SLC-40

Distance to Proximities CCAFS SLC-40



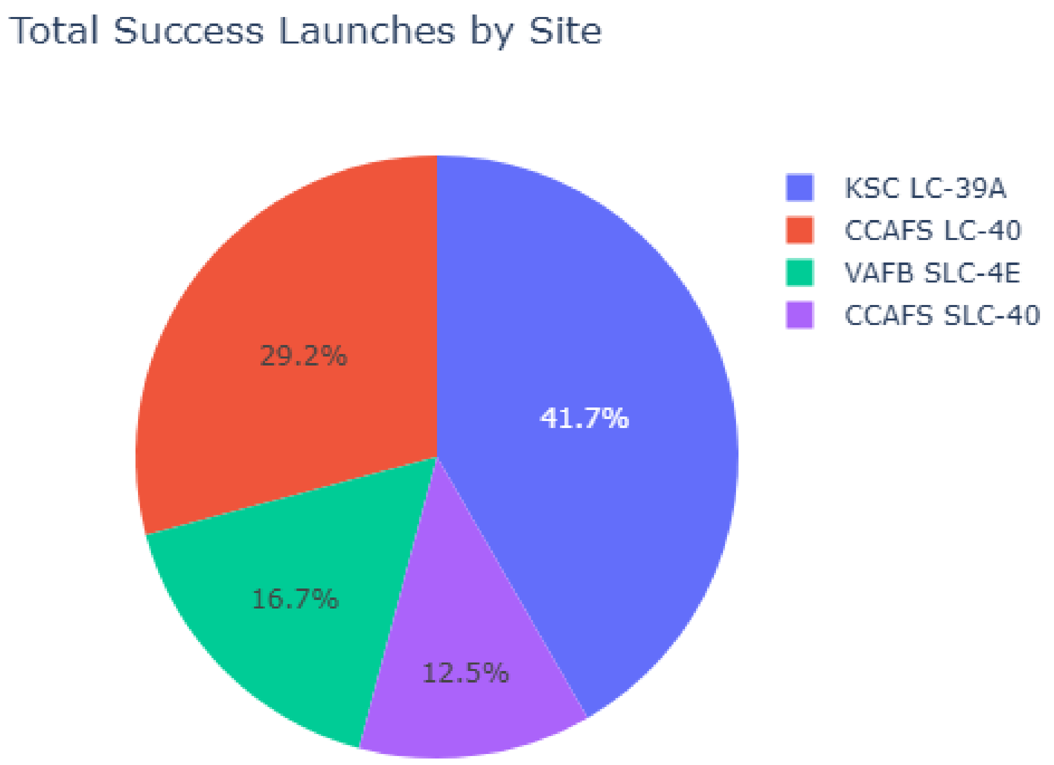
- 0.85 Km to the coastline: Help ensure that spent stages dropped along the launch path or failed launches don't fall on people or property.
- 23.19 Km to the nearest city and : Safety - Security, needs to be an exclusion zone around the launch site to keep unauthorized people away and keep people safe.
- 1.20 km to the nearest railway and 0.59 Km to the nearest highway : Transportation/Infrastructure and Cities, need to be away from anything a failed launch can damage, but still close enough to roads/rails/docks to be able to bring people and material to or from it in support of launch activities.

Section 3

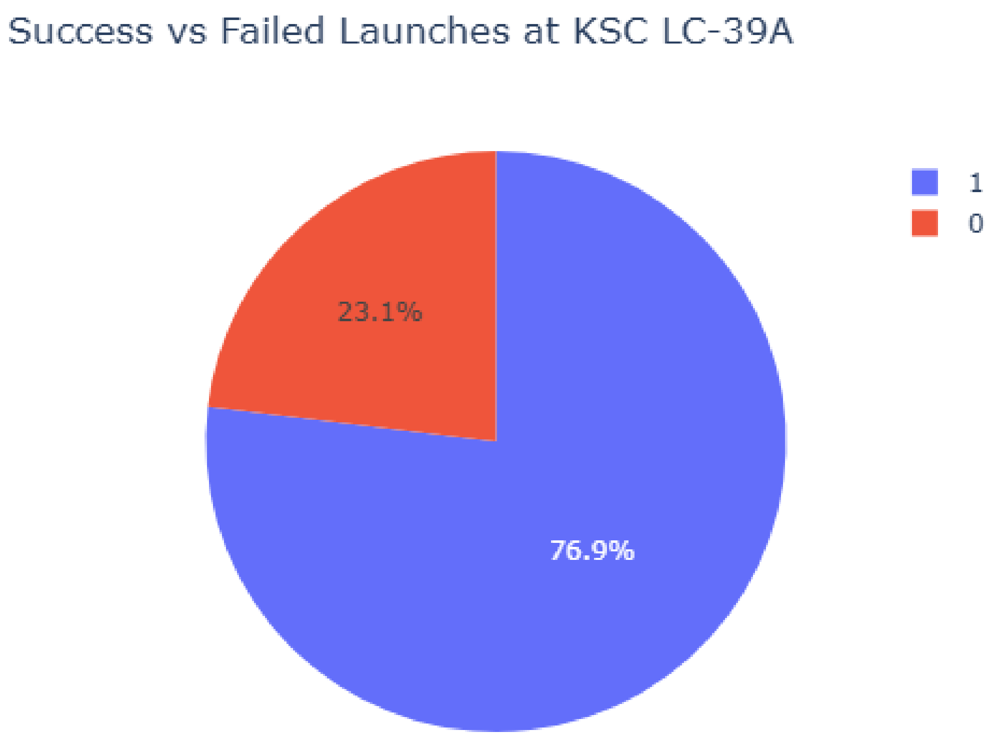
SpaceX Launch Records Dashboard

Launch Success by Site

- With all launch sites selected, the pie chart displayed the distribution of successful Falcon 9 first stage landing outcomes between the different launch sites.
- The place from where launches are done seems to be a very important factor of success of missions.
- The launch site KSC LC-39A had the most successful launches, with 41.7% of the total successful launches



Launch Success by Site

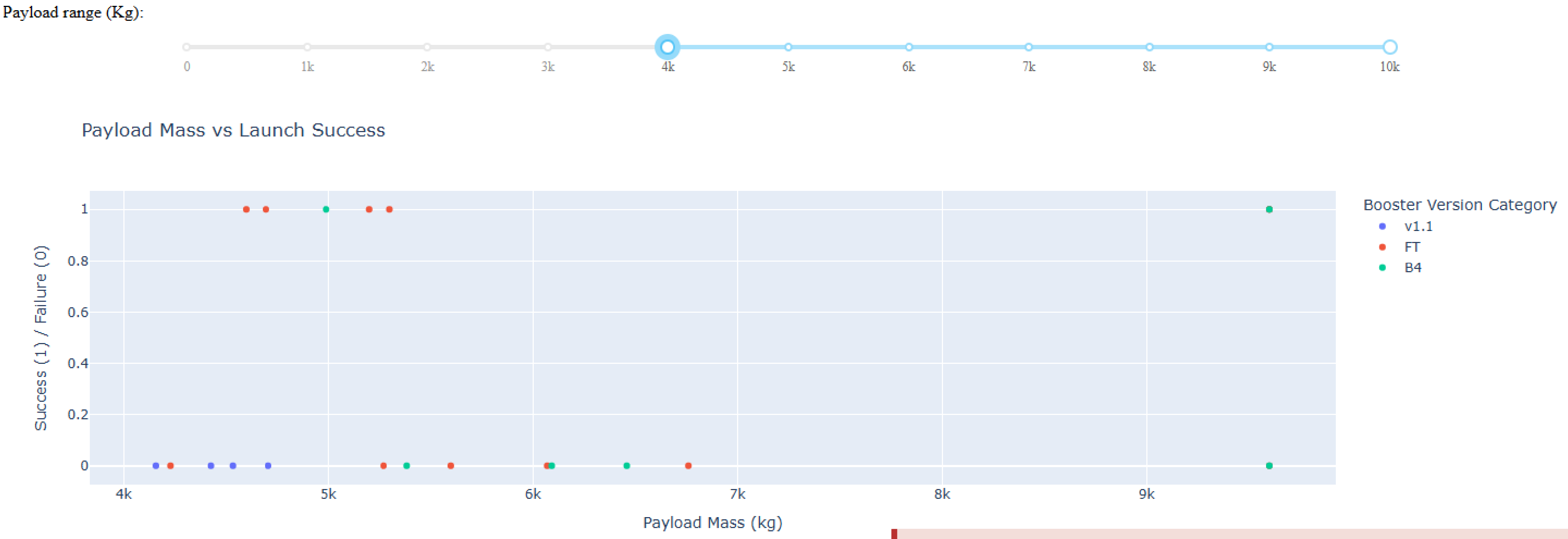


- Falcon 9 first stage failed landings are indicated by the ‘0’ Class (■ red wedge in the pie chart) and successful landings by the ‘1’ Class (■ blue wedge in the pie chart).
- KSC LC-39A has the highest success rate amongst launch sites (76.9%)
- 10 successful launches and 3 failed launches

Payload Mass and Success

- **Payloads between 2,000 kg and 4,000 kg and FT Booster Version have the highest success rate**

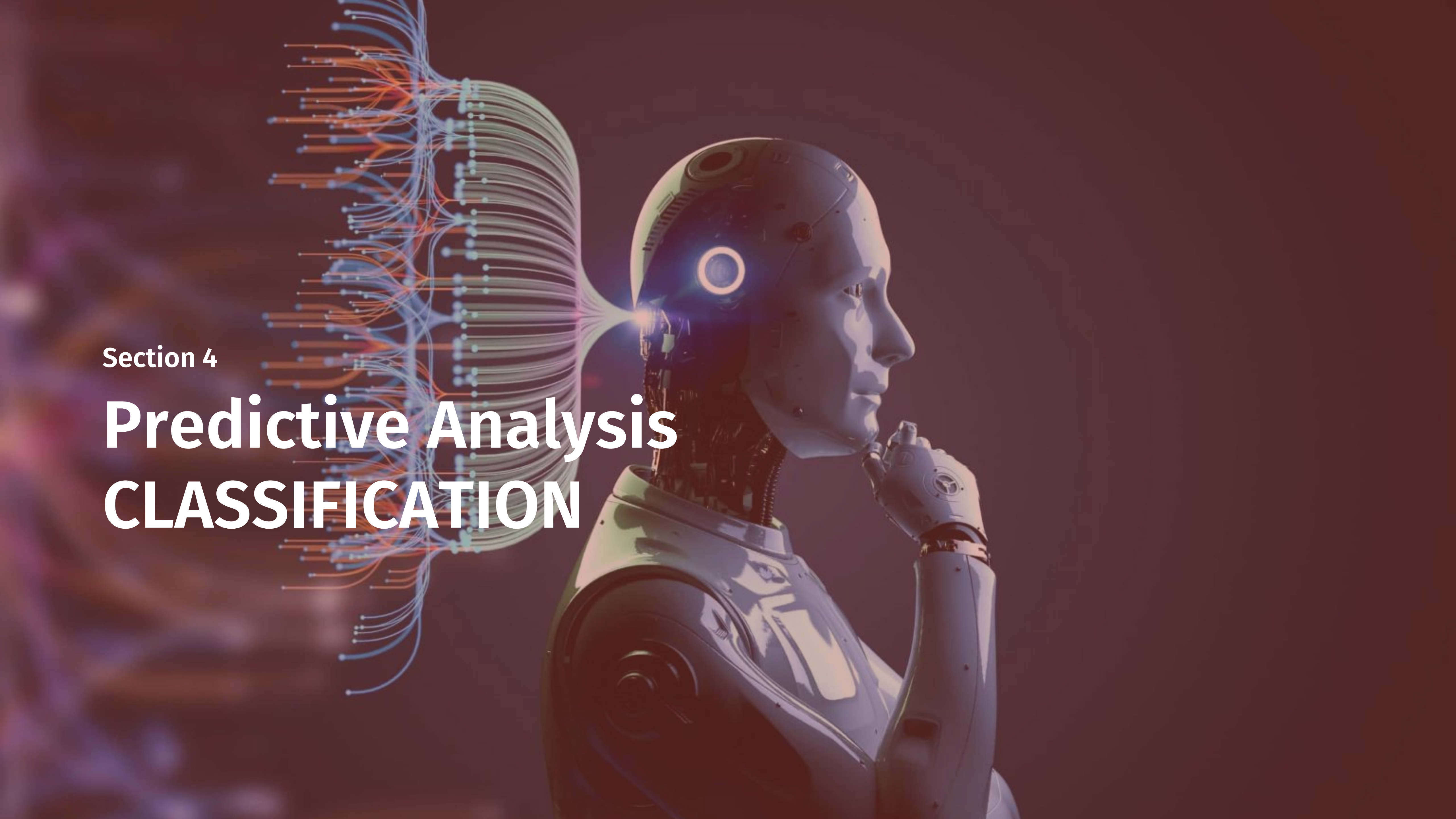
Payload Mass and Success



- **With heavier payload mass it appears to end in an unsuccessful landing**

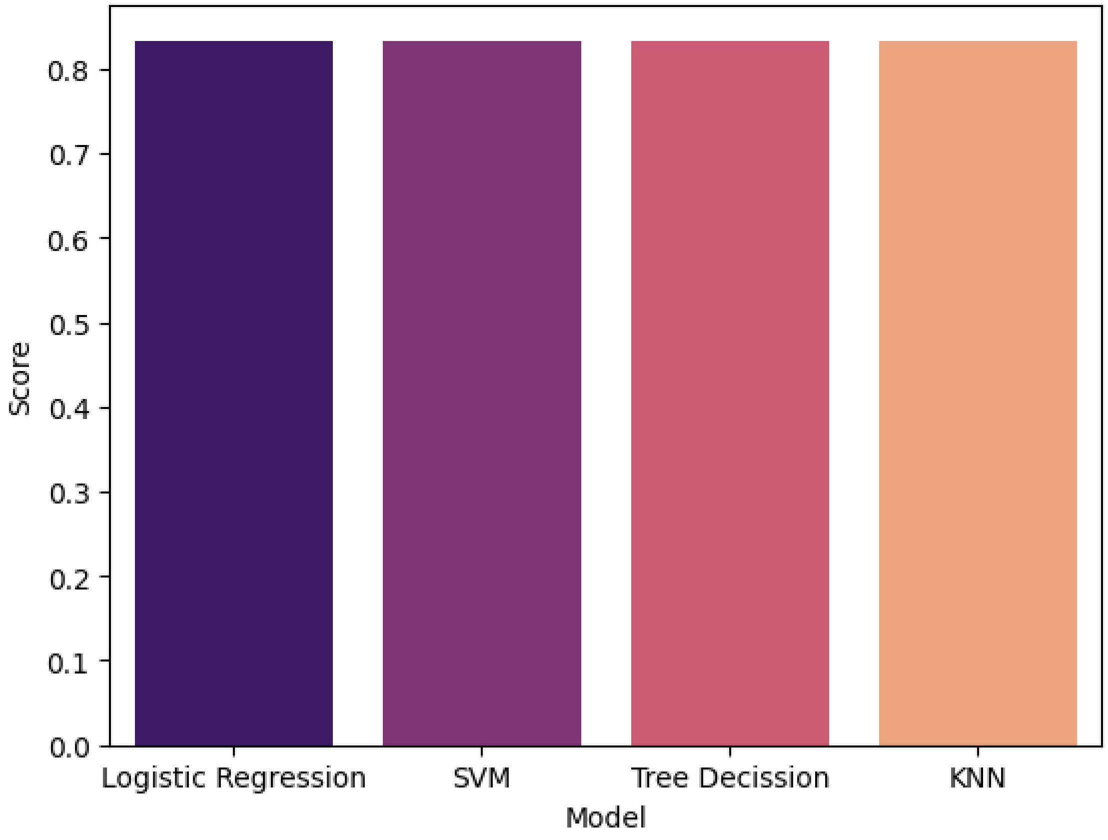
Section 4

Predictive Analysis CLASSIFICATION

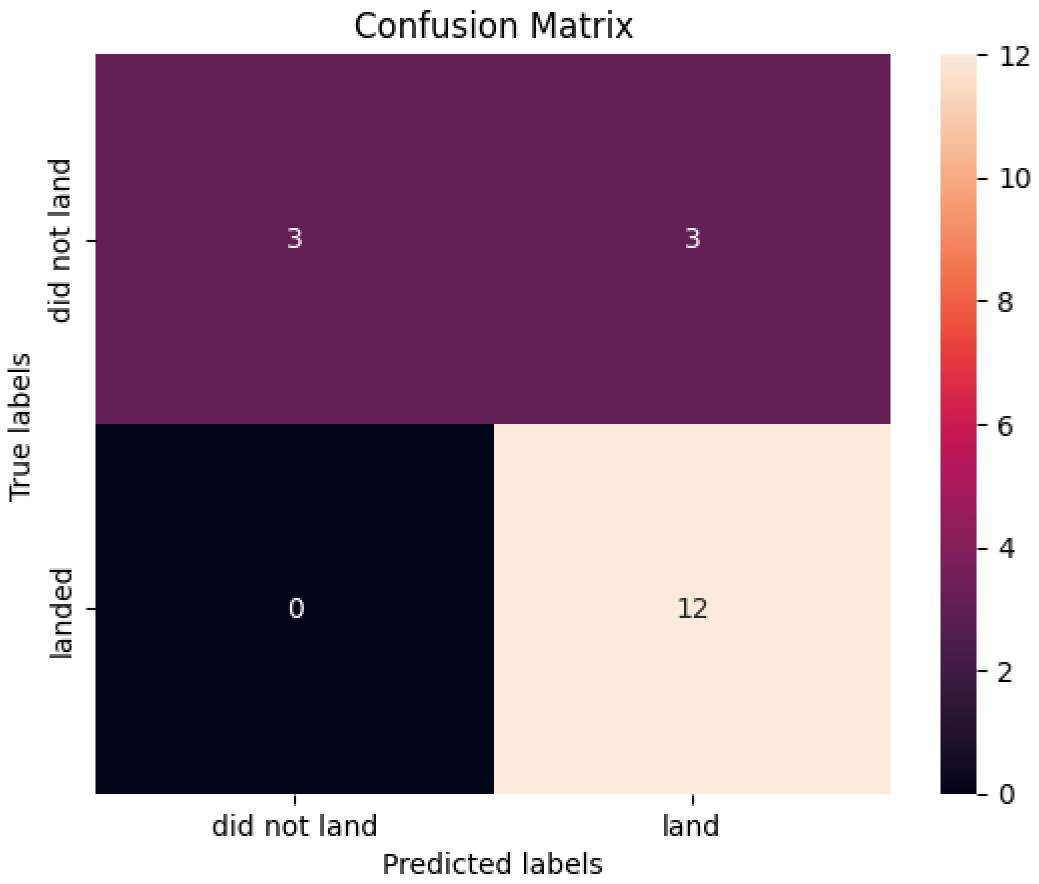


Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside
- All the models performed at about the same level and had the same scores. This is likely due to the small dataset.



Confusion Matrix



- A confusion matrix summarizes the performance of a classification algorithm
- All the confusion matrices were identical
- The fact that there are false positives (Type 1 error) is not good
- Confusion Matrix Outputs:
  - 12 True positive
  - 3 True negative
  - 3 False positive
  - 0 False Negative

Conclusions

- The success of a mission can be explained by several factors such as the launch site, the orbit and especially the number of previous launches. Indeed, we can assume that there has been a gain in knowledge between launches that allowed to go from a launch failure to a success.
- The scatter plot analysis showed that the "FT" booster version has a high success rate across various payload masses, demonstrating its reliability and robustness compared to other booster versions. This suggests that future missions might benefit from utilizing this booster version for improved success rates.
- The success for massive payloads (over 4000kg) is lower than that for low payloads.
- With the current data, we cannot explain why some launch sites are better than others. To get an answer to this problem, we could obtain atmospheric or other relevant data.
- The results for Classification Models are practically the same. This is because the dataset is small and having lesser values. A larger dataset will help build on the predictive analytics results to help understand if the findings can be generalizable to a larger data set

Appendix

- **SpaceX API (JSON):** https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- **Wikipedia (Webpage):** https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- **Launch Geo (CSV):** https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv
- **Launch Dash (CSV):** https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_dash.csv
- **GitHub URL (API Data Collection):** https://github.com/Borguas/IBMDataScienceCapstoneSpaceX/blob/main/01-Spacex-data-collection-api-v2.ipynb
- **GitHub URL (Web Scraping):** https://github.com/Borguas/IBMDataScienceCapstoneSpaceX/blob/main/02-Wikipedia-webscraping.ipynb
- **GitHub URL (Data Wrangling):** https://github.com/Borguas/IBMDataScienceCapstoneSpaceX/blob/main/03-Spacex-Data%20wrangling-v2.ipynb
- **GitHub URL (EDA with SQL):** https://github.com/Borguas/IBMDataScienceCapstoneSpaceX/blob/main/05-eda-sql-coursera_sqllite.ipynb
- **GitHub URL (EDA with Data Visualization):** https://github.com/Borguas/IBMDataScienceCapstoneSpaceX/blob/main/04-eda-dataviz-v2.ipynb
- **GitHub URL (Folium Maps):** https://github.com/Borguas/IBMDataScienceCapstoneSpaceX/blob/main/06-launch-site-location-v2.ipynb
- **GitHub URL (Dashboard File):** https://github.com/Borguas/IBMDataScienceCapstoneSpaceX/blob/main/07-spacex_dash_app.py
- **GitHub URL (Machine Learning):** https://github.com/Borguas/IBMDataScienceCapstoneSpaceX/blob/main/08-SpaceX-Machine-Learning-Prediction-v1.ipynb


Thank You!
NASA
T+00:01:58
NASA
NASA