IBM Developer
SKILLS NETWORK

**Bhundiya Bansiben Rameshbhai**
23/08/2024

SPACEX

# Outline

- Executive Summary

- Introduction

- Methodology

- Results

- Conclusion

- Appendix

# Executive Summary

- Summary of methodologies

• Data was collected from the SpaceX public API and publically available data on Wikipedia. Data wrangling include extracting launch outcome information to serve as the dependent variable in the Machine Learning models.

 • SQL queries and data visualizations (static plots, interactive maps, and an interactive dashboard) were created to discover insights about the data set and answer questions.

• Predictive analysis was pursued using Logistic Regression, SVM (Support Vector Machine), Decision Tree, and KNN    (k-Nearest Neighbors) Machine Learning models.

• Summary of all results

• Launch data include info about flight number, date of launch, payload mass, orbit type, launch site, mission outcome and other variables.

• Logistic Regression, SVM (Support Vector Machine), and KNN (k-Nearest Neighbors) all perform equally well for Machine Learning models on this dataset.

# Introduction

• In competition with SpaceX, a rival rocket launch company wants to make predictions about the success/failure
  SpaceX Falcon 9 rocket first stage landings.

• What is the nature and extent of the data that we have on SpaceX Falcon 9 first stage landings?

• Which machine learning model would work best (have the highest accuracy) to predict the outcome of a Falcon -
  9 first stage landing from a future launch?

• Will a future Falcon 9 first stage landing be successful?
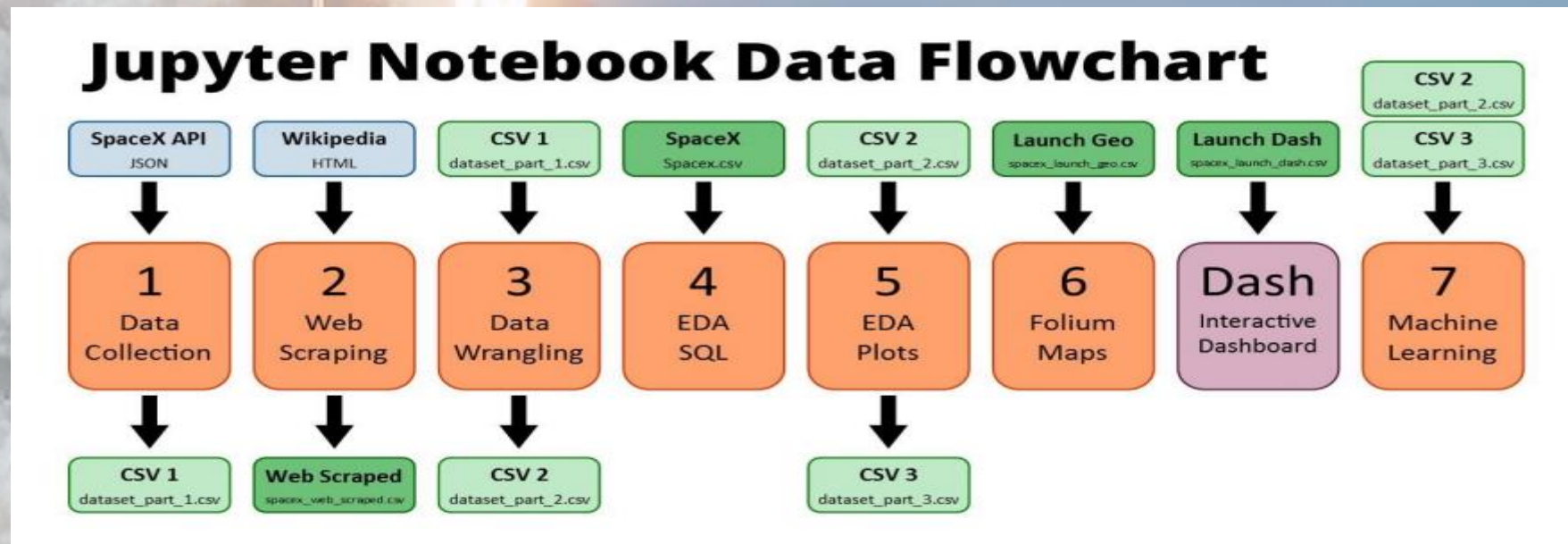
# Section 1

# Methodology

## Executive Summary

• SpaceX API and Wikipedia launch table data was collected.

• Data was cleaned in preparation for visualizations, queries and machine learning model creation.

• Exploratory data analysis (EDA) was done using visualization and SQL.

• Interactive visual analytics were created using Folium and Plotly Dash.

• Predictive analysis using classification models was done.

# Data Collection

- The data sets were collected from:

- An IBM copy of a call to the publically accessible SpaceX API with launch data in JSON format.

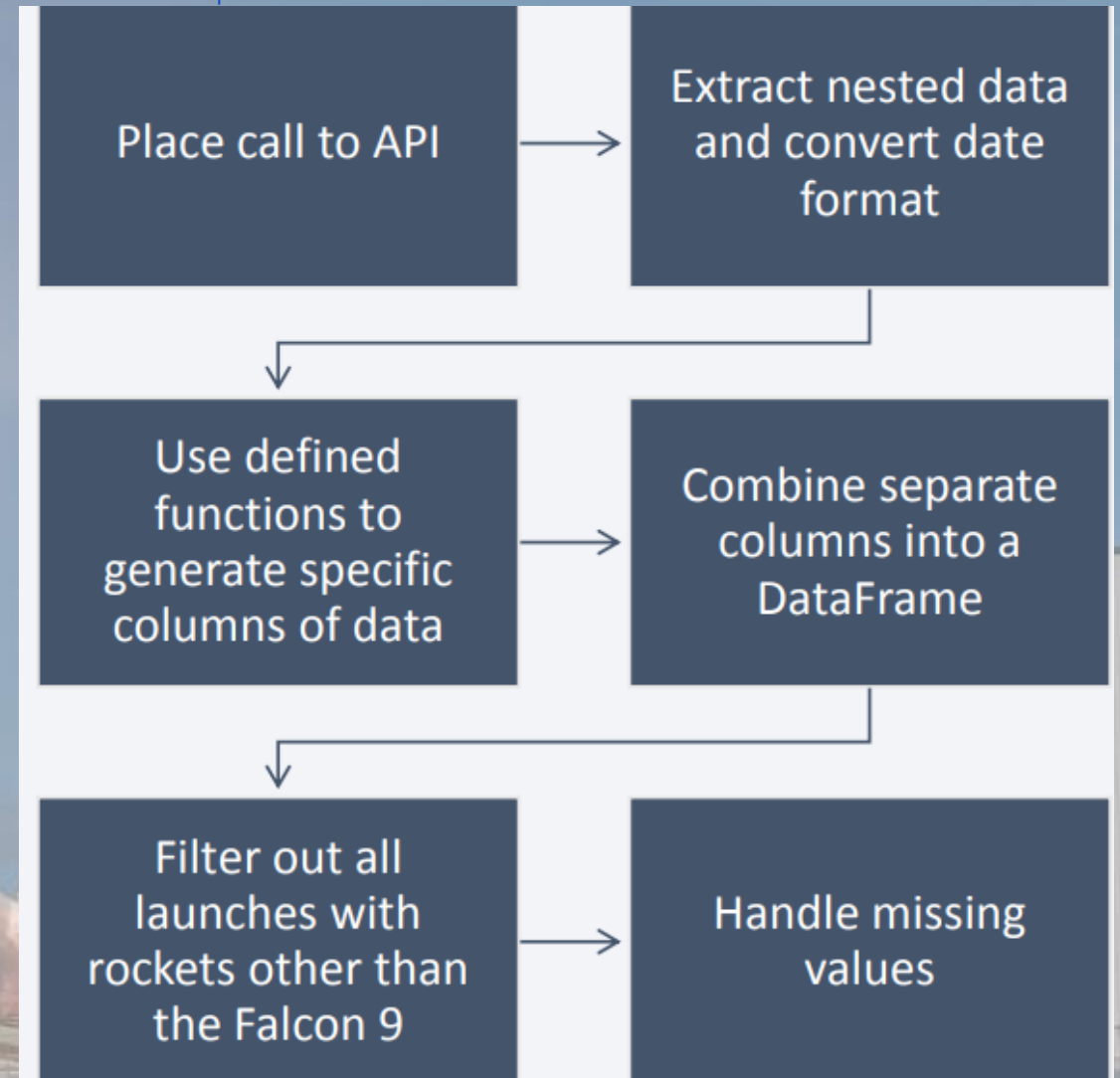- A permanently linked Wikipedia page with launch data in HTML tables

# Data Collection – SpaceX API

- The SpaceX API has data available publically.

- Once a GET request has been made to the SpaceX API and the response received, the data can be placed into a Pandas DataFrame for further analysis

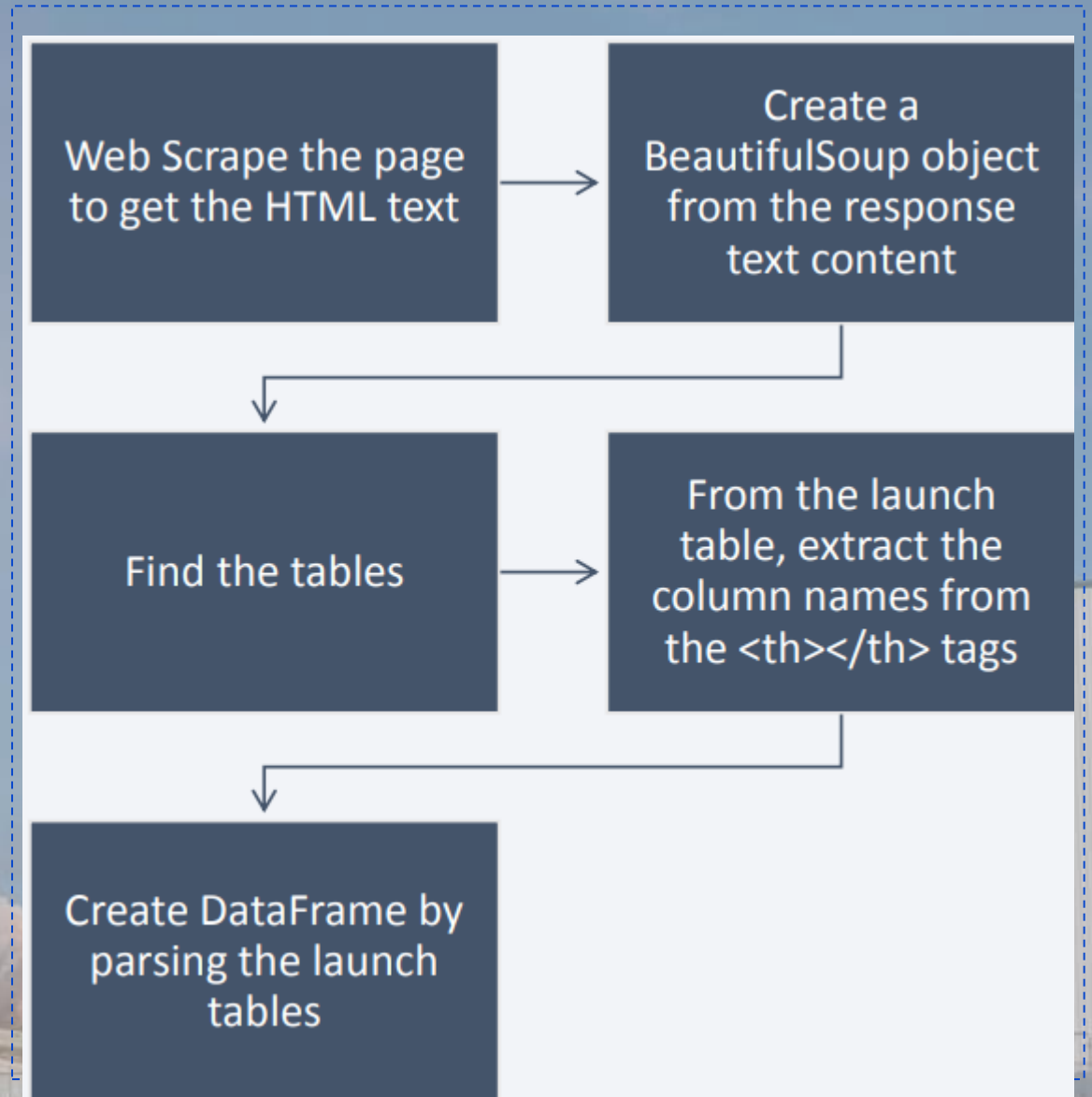- GitHub URL (Data Collection):

IBM-Applied-Data-Science-capstonee/jupyter-labs-spacex-data-collection-api (1).ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

| | |
|---|---|
| Place call to API | Extract nested data and convert date format |
| Use defined functions to generate specific columns of data | Combine separate columns into a DataFrame |
| Filter out all launches with rockets other than the Falcon 9 | Handle missing values |

# Data Collection - Scraping

- Wikipedia has a page that has tables of data about SpaceX launches

- These tables can be scraped to extract launch data that can be put into a Pandas DataFrame for further analysis.

- GitHub URL (Web Scraping):

IBM-Applied-Data-Science-capstonee/Spacex_Data_Collection_Web_Scraping.ipynb.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

# Data Wrangling

- The .csv file from the first section contains the data that needed to be cleaned.

- The launch sites, orbit types and mission outcomes were cleaned up.
  - The handful of mission outcome types were converted to a binary classification where 1 means that the Falcon 9 first stage landing was a success and 0 means
- The new classification was added to the DataFrame for further analysis that it was a failure.
  - GitHub URL (Data Wrangling):IBM-Applied-Data-Science-capstonee/Spacex-Data wrangling.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

# EDA with Data Visualization

- The following charts were created to look at Launch Site trends

- Scatterplot to see mission outcome relationship split by Launch Site and Flight Numbe

- Scatterplot to see mission outcome relationship split by Launch Site and Payload.

- The following charts were created to look at Orbit Type trends

- Bar chart to see mission outcome relationship with Orbit Type.

- Scatterplot to see mission outcome relationship split by Orbit Type and Flight Number.

- Scatterplot to see mission outcome relationship split by Orbit Type and Payload.

- The following chart was created to look at trends based on time

- Line plot to see mission outcome trend by year.

- GitHub URL (EDA with Data Visualization): IBM-Applied-Data-Science-capstonee/SpaceX_Falcon9_EDA_Vizulization.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

# EDA with SQL

- Queries were written to extract information about:

- Launch sites

- Payload masses

- Dates

- Booster types

- Mission outcomes

- GitHub URL (EDA with SQL): IBM-Applied-Data-Science-capstonee/SpaceX_Falcon9_EDA_SQL.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

# Build an Interactive Map with Folium

• Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map

• Markers were added for launch sites and for the NASA Johnson Space Center

• Circles were added for the launch sites.

• Lines were added to show the distance to the nearby features:

• Distance from CCAFS LC-40 to the coastline

• Distance from CCAFS LC-40 to the rail line

• Distance from CCAFS LC-40 to the perimeter road

• GitHub URL (Folium Maps): IBM-Applied-Data-Science-capstone/SpaceX_Falcon9_Interactive_Folium.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstone (github.com)

# Build a Dashboard with Plotly Dash

- The input dropdown is used to select one or all launch sites for the pie chart and scatterplot.
  - The pie chart displays one of two things:
  - For All Sites – the distribution of successful Falcon 9 first stage landings between the sites
  - For One Site – the distribution of successful and failed Falcon 9 first stage landings for that site
  - The input slider is used to filter the payload masses for the scatterplot.
  - The scatterplot displays the distribution of Falcon 9 first stage landings split by payload mass, mission outcome and by booster version category.
- GitHub URL (Dashboad File): IBM-Applied-Data-Science-capstonee/07_SpaceX_Interactive_Visual_Analytics_Plotly.py at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

# Predictive Analysis (Classification)

- The dataset was split into training and testing sets.

- Logistic Regression, SVM (Support Vector Machine), Decision Tree, and KNN (k-Nearest Neighbors) machine learning models were trained on the training data set.

- Hyper-parameters were evaluated using GridSearchCV() and the best was selected 'best_params'

- Using the best hyper-parameters, each of the four models were scored on accuracy by using the testing data set.

- GitHub URL (Machine Learning):  IBM-Applied-Data-Science-capstonee/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)
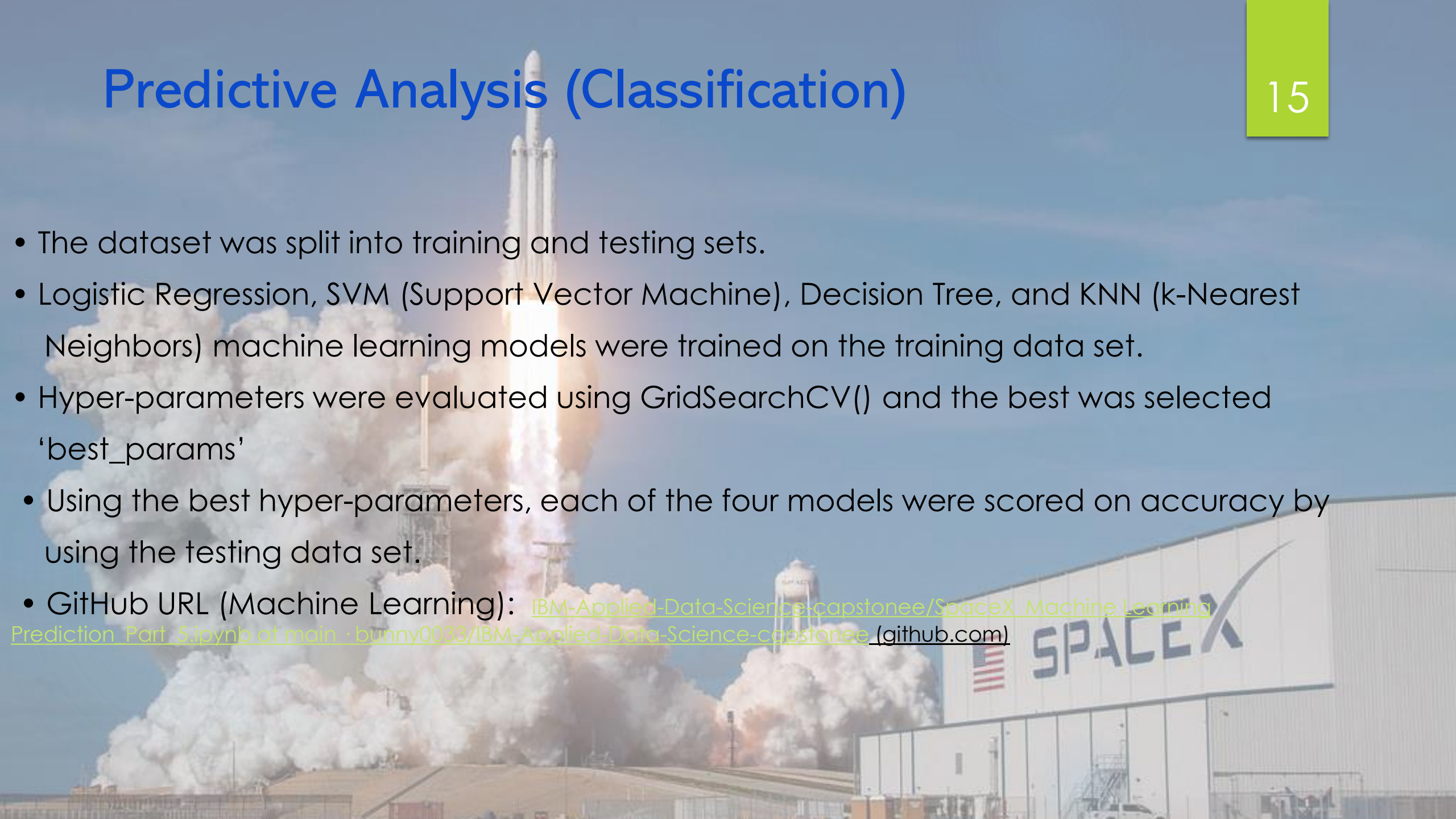
# Results

- Exploratory data analysis results

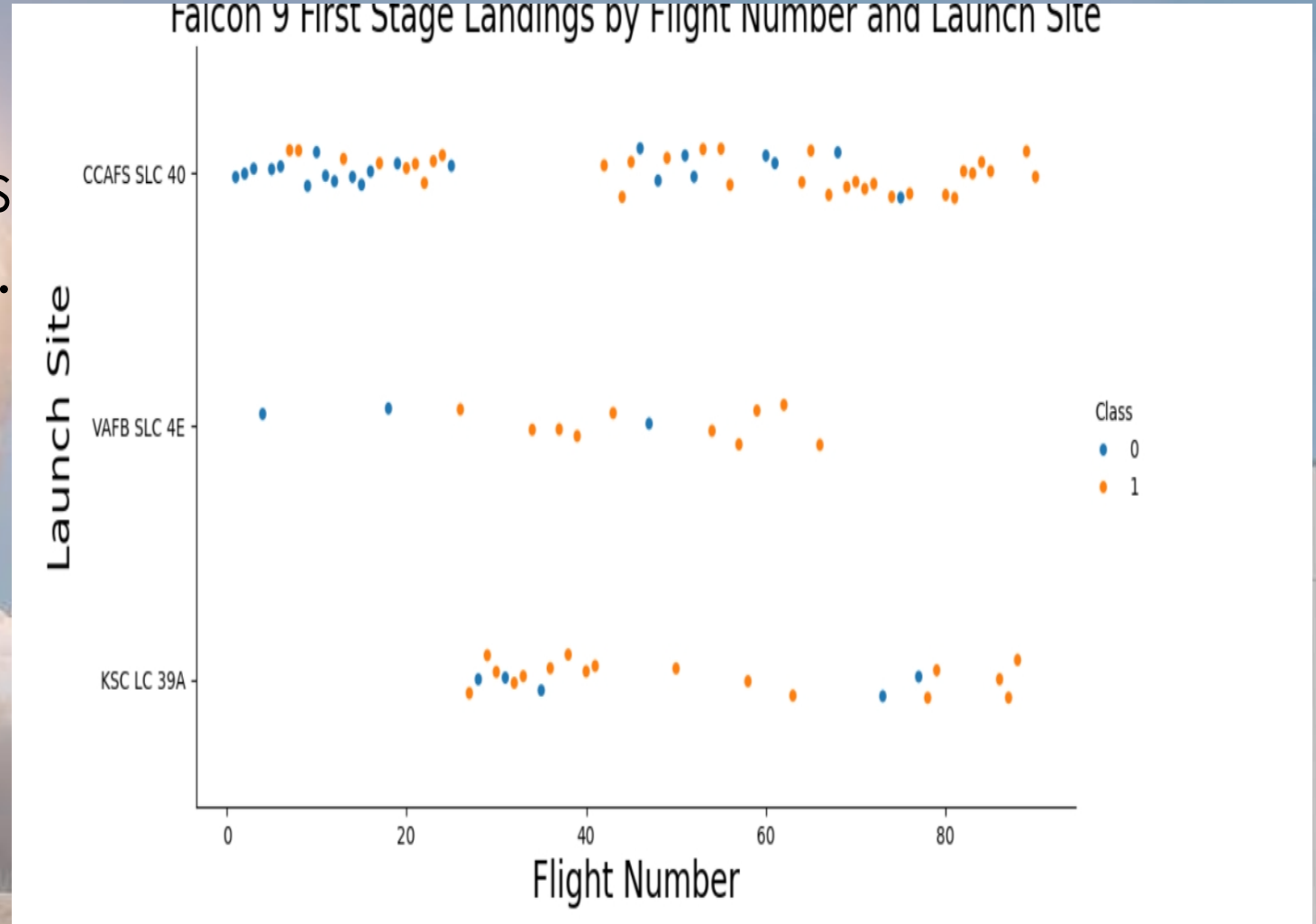- Interactive analytics demo in screenshots

- Predictive analysis results

Section 2

# Flight Number vs. Launch Site

• Successful Falcon 9 first stage landings appear to become more prevalent as the flight number increases.



Falcon 9 First Stage Landings by Flight Number and Launch Site

# Payload vs. Launch Site

• For the CCAFS SLC 40 launch site, the payload mass and the landing outcome appear to not be strongly correlated.

• The failed landings at the KSC LC 39A launch site are all grouped around a narrow band of payload masses.



Falcon 9 First Stage Landings by Payload Mass (kg) and Launch Site
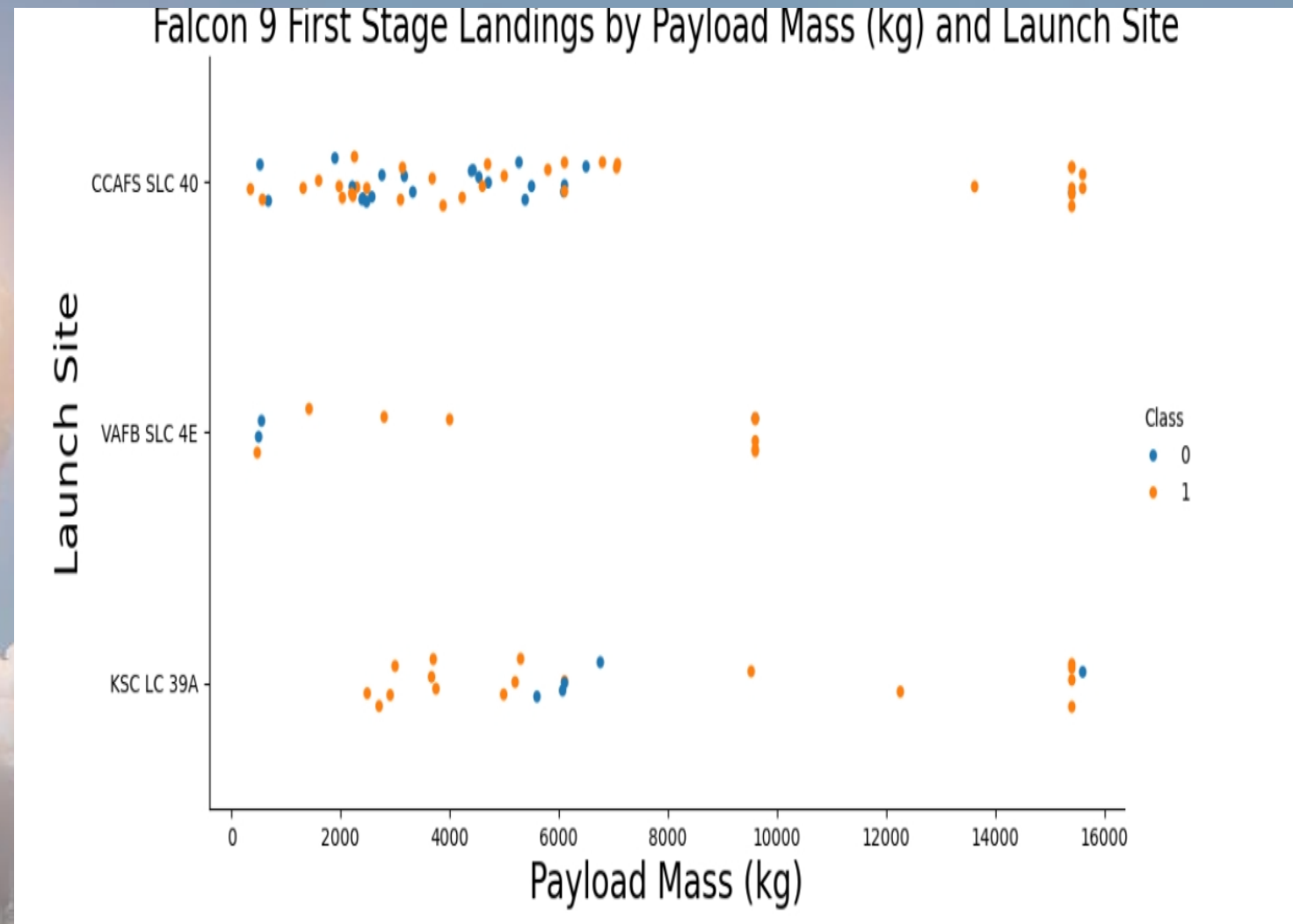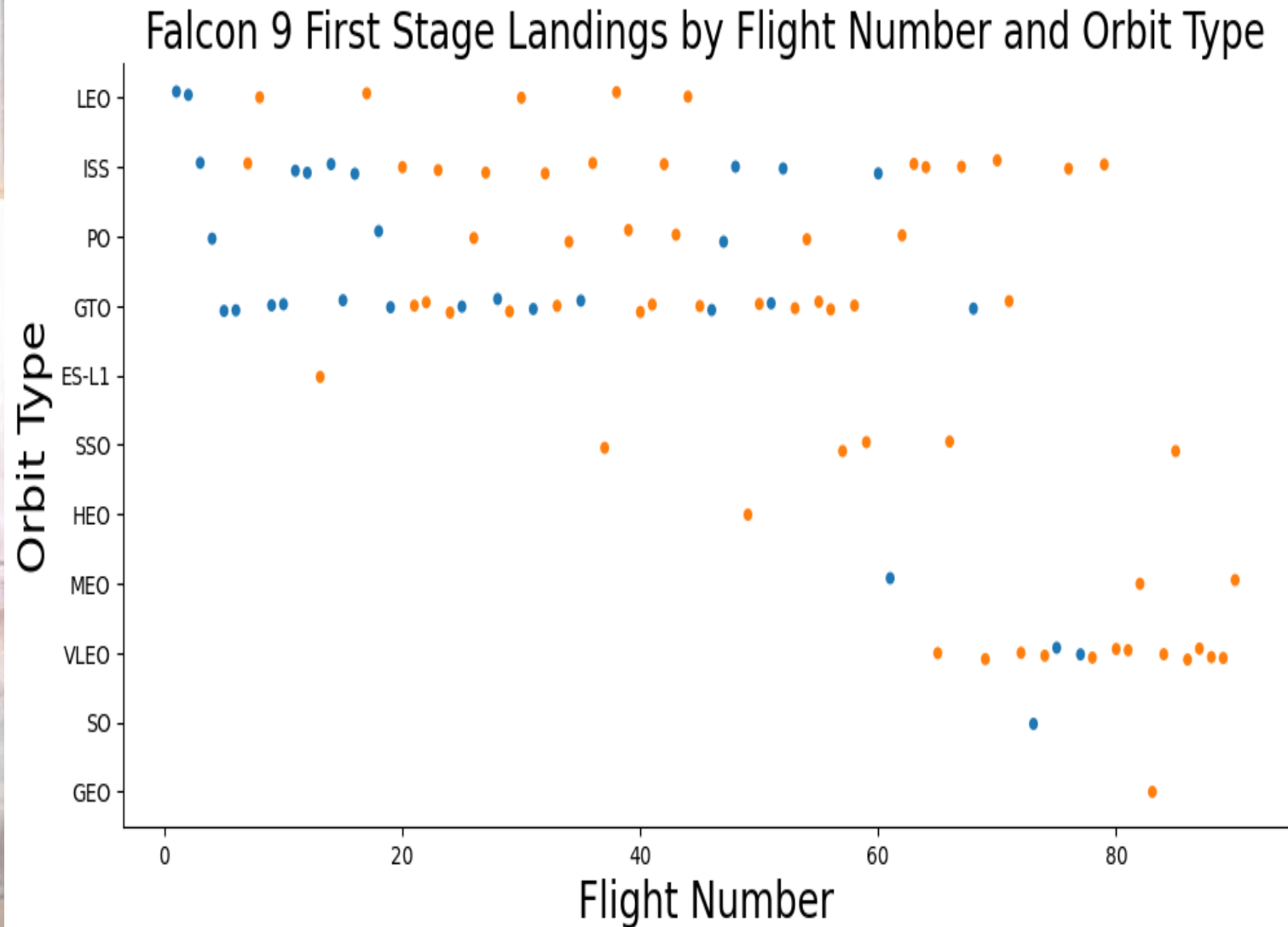
# Success Rate vs. Orbit Type

- ES-L1, SSO, HEO and GEO orbits have no failed first stage landings.

- SO orbits have no successful first stage landings.
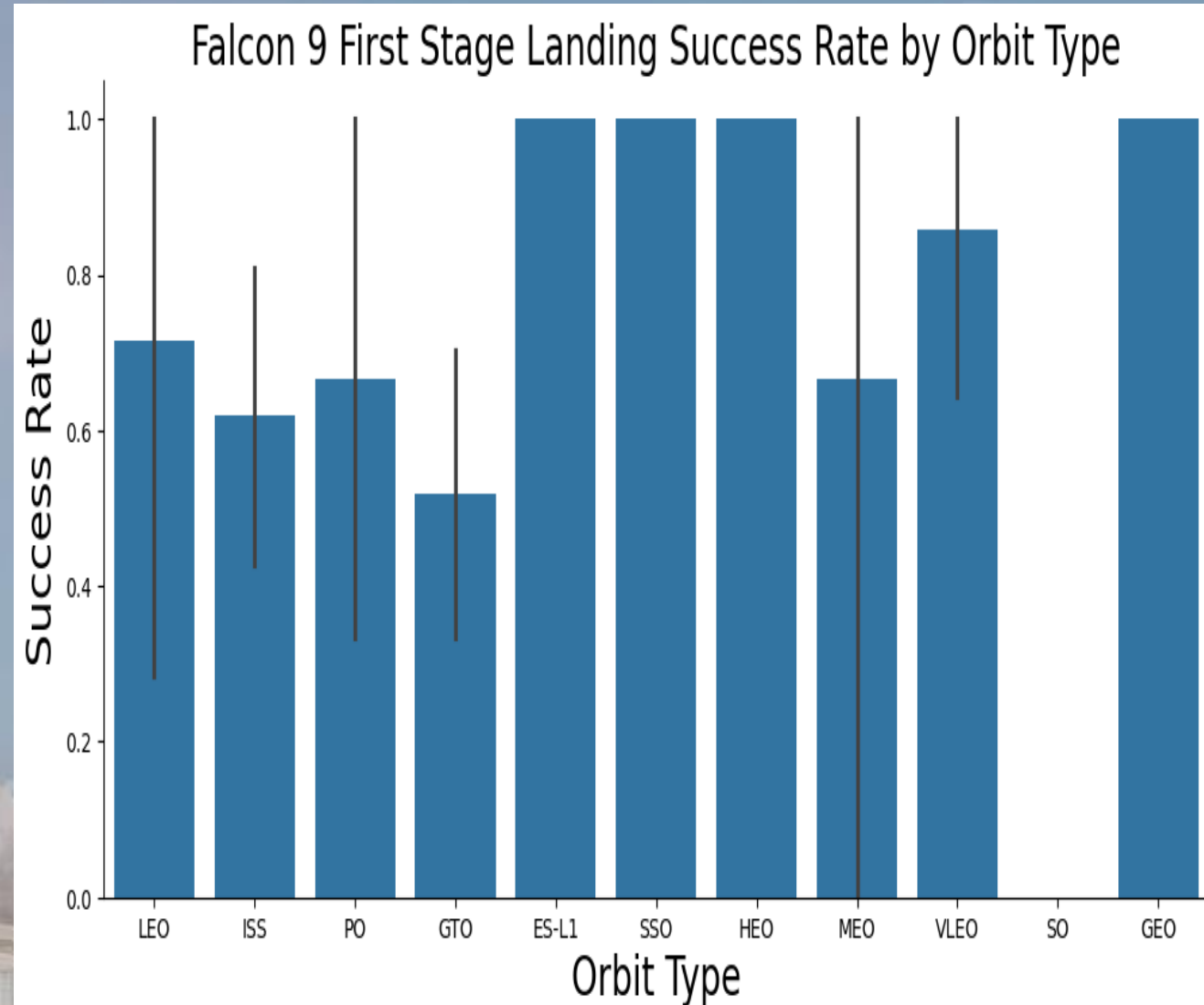
# Flight Number vs. Orbit Type

• There is a correlation between flight number and success rate with larger flight numbers being associated with higher success rates.



Falcon 9 First Stage Landings by Flight Number and Orbit Type

# Payload vs. Orbit Type

- Some orbit types have better success rates than others

- Success rate appears to have no obvious correlation with payload mass


Falcon 9 First Stage Landings by Payload Mass and Orbit Type
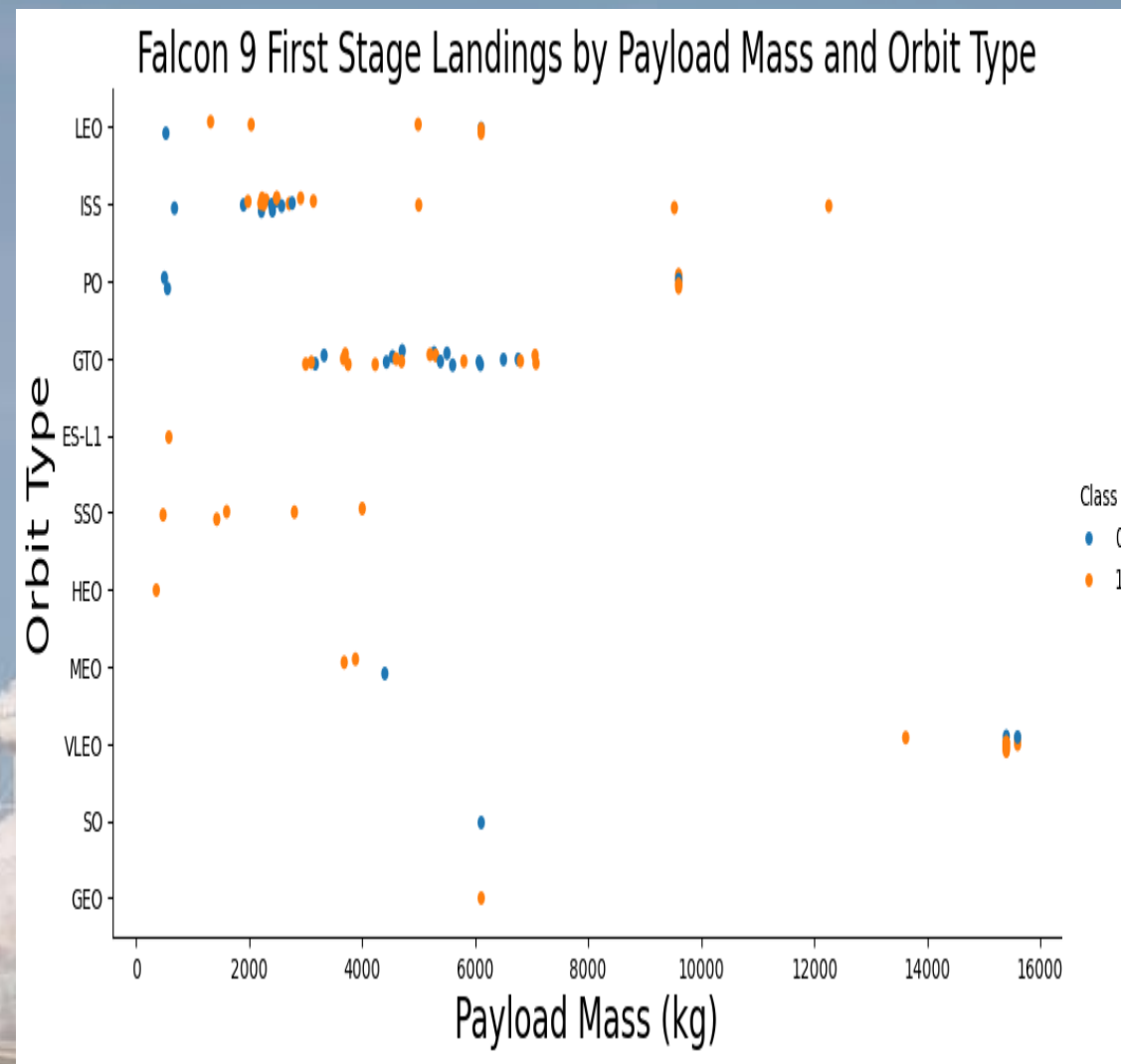
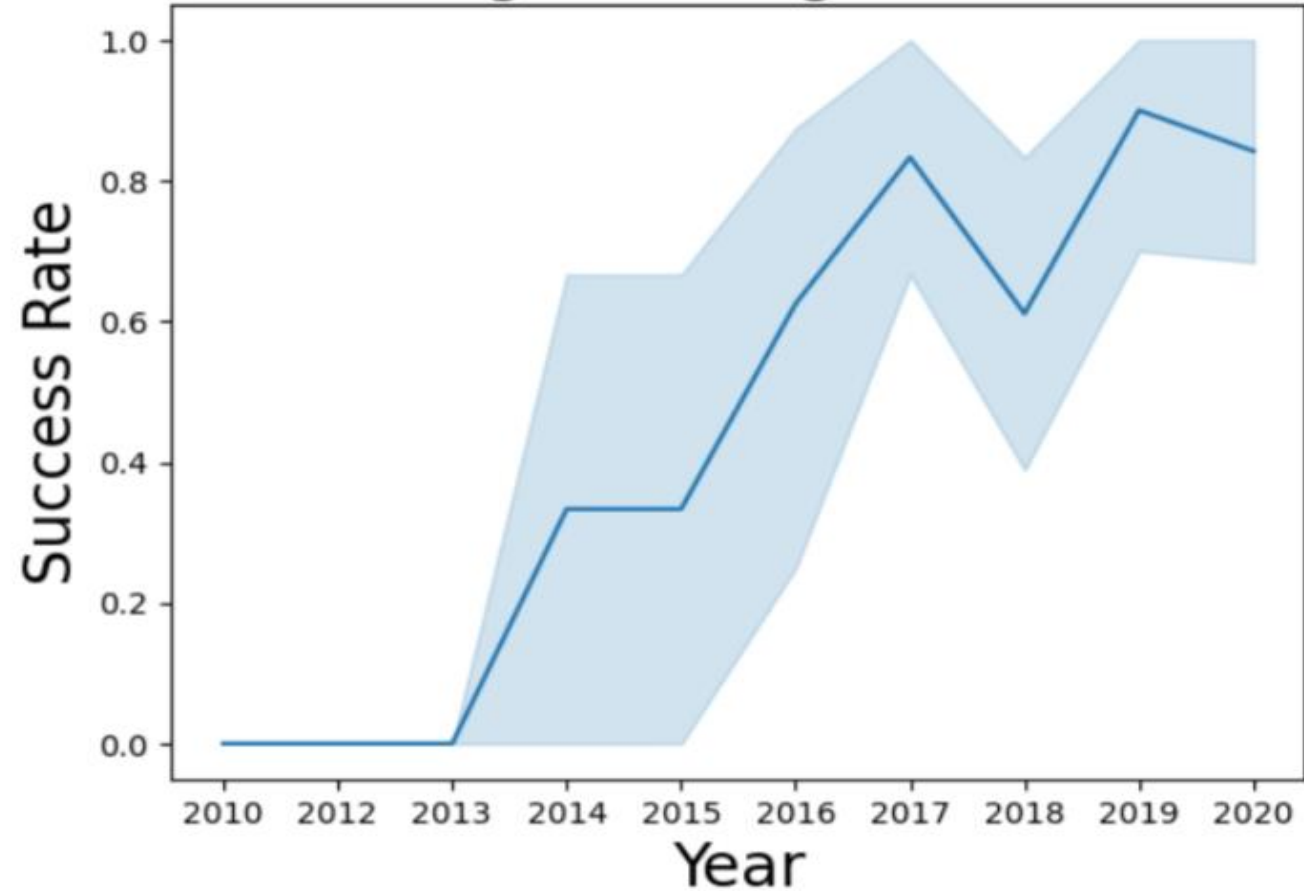# Launch Success Yearly Trend

▶ Show a line chart of yearly average success rate

▶ Show the screenshot of the scatter plot with explanations



Falcon 9 First Stage Landing Success Rate by Year

# All Launch Site Names

- Question: What are the names of the unique launch sites?

- Query: SELECT DISTINCT LAUNCH_SITE FROM SPACEXDATASET;

- Result:

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

- Explanation: There are four unique launch sites. All Launch Site

# Launch Site Names Begin with 'CCA'

- Task: Find 5 records with launch sites that begin with `CCA'.

- Query: SELECT * FROM SPACEXDATASET WHERE launch_site LIKE 'CCA%' LIMIT 5

- Result:

| 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 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 2012-10-08 | 00:35:00 | 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 |

- Explanation: This is a fairly straightforward sampling mechanism used to gain a sense of the data contained in the database table. Launch Site Names That Begin with cca
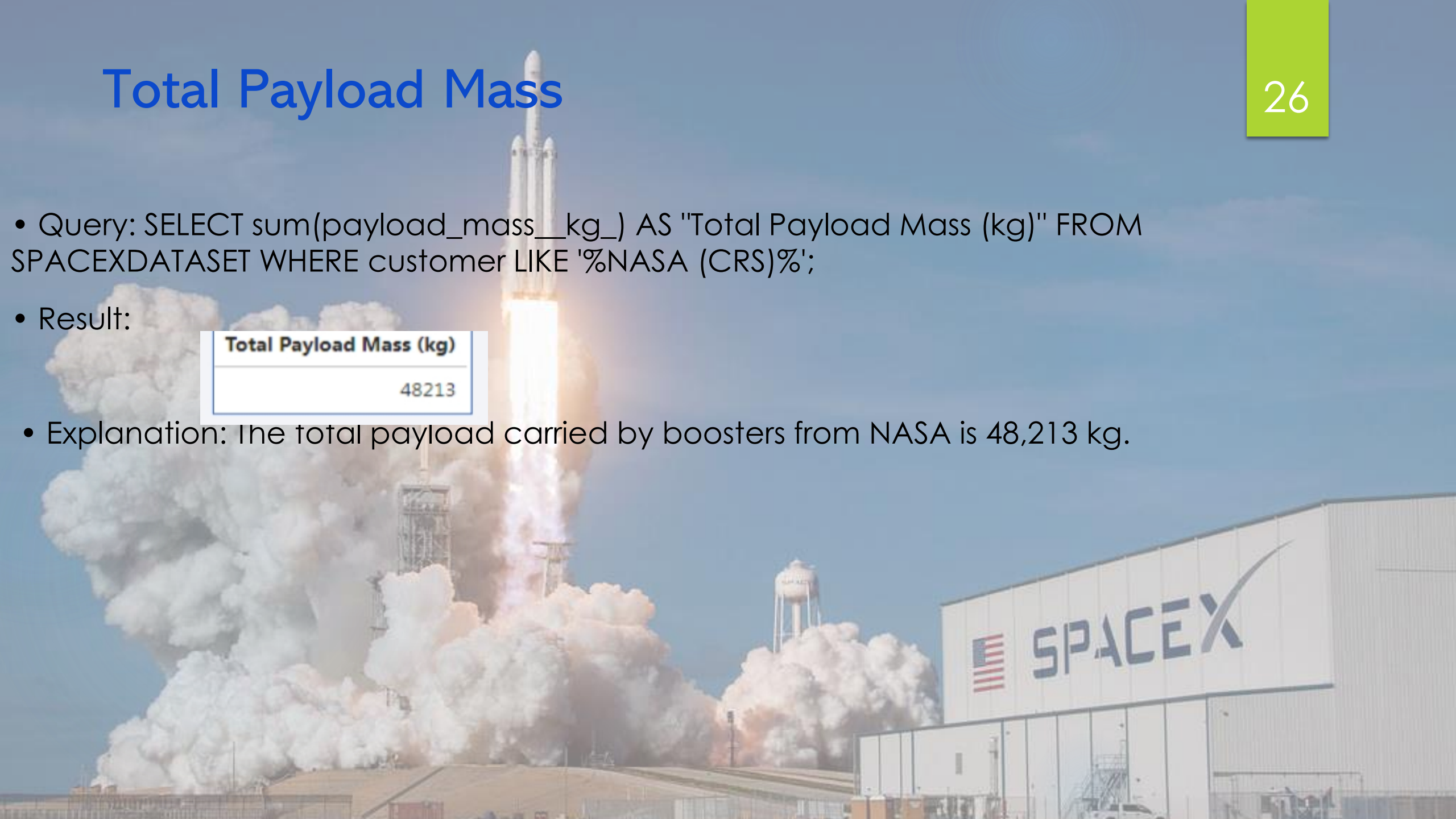
# Total Payload Mass

• Query: SELECT sum(payload_mass__kg_) AS "Total Payload Mass (kg)" FROM SPACEXDATASET WHERE customer LIKE '%NASA (CRS)%';

• Result:

| Total Payload Mass (kg) |
|---|
| 48213 |

• Explanation: The total payload carried by boosters from NASA is 48,213 kg.
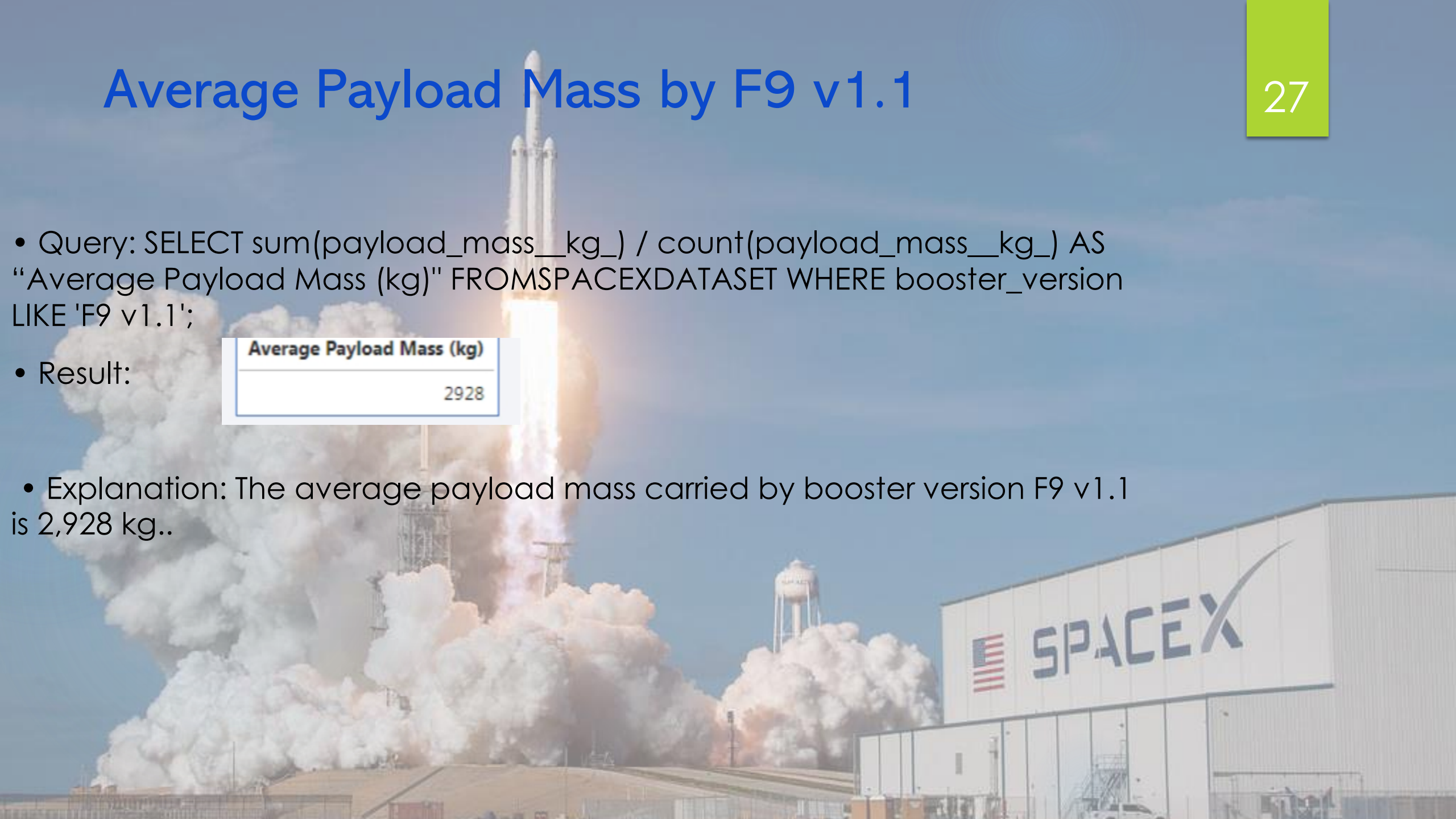
# Average Payload Mass by F9 v1.1

• Query: SELECT sum(payload_mass__kg_) / count(payload_mass__kg_) AS "Average Payload Mass (kg)" FROMSPACEXDATASET WHERE booster_version LIKE 'F9 v1.1';

• Result:

| Average Payload Mass (kg) |
|---|
| 2928 |

• Explanation: The average payload mass carried by booster version F9 v1.1 is 2,928 kg..

# First Successful Ground Landing Date

- Query: SELECT min(DATE) AS "First Successful Landing Outcome Date" FROM SPACEXDATASET WHERE landing__outcome LIKE 'Success (ground pad)'

- Result

| First Successful Landing Outcome Date |
|---|
| 2015-12-22 |

- Explanation: The first successful landing outcome on ground pad occurred on December 22, 2015.

Query: SELECT DISTINCT booster_version FROM SPACEXDATASET WHERE landing__outcome = 'Success (drone ship)' and payload_mass__kg_ BETWEEN 4000 and 6000;

- Result:

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

- Explanation: The four booster versions that have successfully landed on drone ship with a payload mass greater than 4,000 kg but less than 6,000 kg are listed above. Successful Drone Ship Landing with Payload between 4000 and 6000

# Total Number of Successful and Failure Mission Outcomes

- Query: SELECT (SELECT count(*) FROM SPACEXDATASET WHERE lcase(landing__outcome) LIKE '%success%') AS "Success", count(*) AS "Failure" FROM SPACEXDATASET WHERE lcase(landing__outcome) NOT LIKE '%success%

- Result:

| Success | Failure |
|---------|---------|
| 61 | 40 |

- Explanation: There were 61 successful and 40 failed mission outcomes. Total Number of Successful and Failure Mission Outcome

# Boosters Carried Maximum Payload

► Query: SELECT booster_version, payload_mass__kg_ FROM SPACEXDATASET WHERE payload_mass__kg_ = (SELECT max(payload_mass__kg_) FROM SPACEXDATASET);

• Result

| 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

▶ Query: Records SELECT MONTHNAME(DATE) AS "Month", landing__outcome, booster_version, launch_site FROM SPACEXDATASET WHERE landing__outcome = 'Failure (drone ship)' AND YEAR(DATE) = 2015;

• Result:

| Month | landing__outcome | booster_version | launch_site |
|---|---|---|---|
| January | Failure (drone ship) | F9 v1.1 B1012 | CCAFS LC-40 |
| April | Failure (drone ship) | F9 v1.1 B1015 | CCAFS LC-40 |

• Explanation: There were two failed landing outcomes with a drone ship in 2015. Both launched from CCAFS LC-40. One occurred in January and the other in April.

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

▶ Query: SELECT landing__outcome, count(landing__outcome) AS "Count" FROM SPACEXDATASET WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing__outcome ORDER BY count(landing__outcome) DESC

• Result:

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

• Explanation: The most common landing outcome was 'not attempted'. Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Section 3

# Falcon 9 Launch Site Locations

- VAFB SLC-4E (California, USA)

- Vandenberg Air Force Base Space Launch Complex 4E

- KSC LC-39A (Florida, USA)

- Kennedy Space Center Launch Complex  39A
- CCAFS LC-40 (Florida, USA)

- Cape Canaveral Air Force Station Launch Complex 40

- CCAFS SLC-40 (Florida, USA)

- Cape Canaveral Air Force Station Space Launch Complex 40
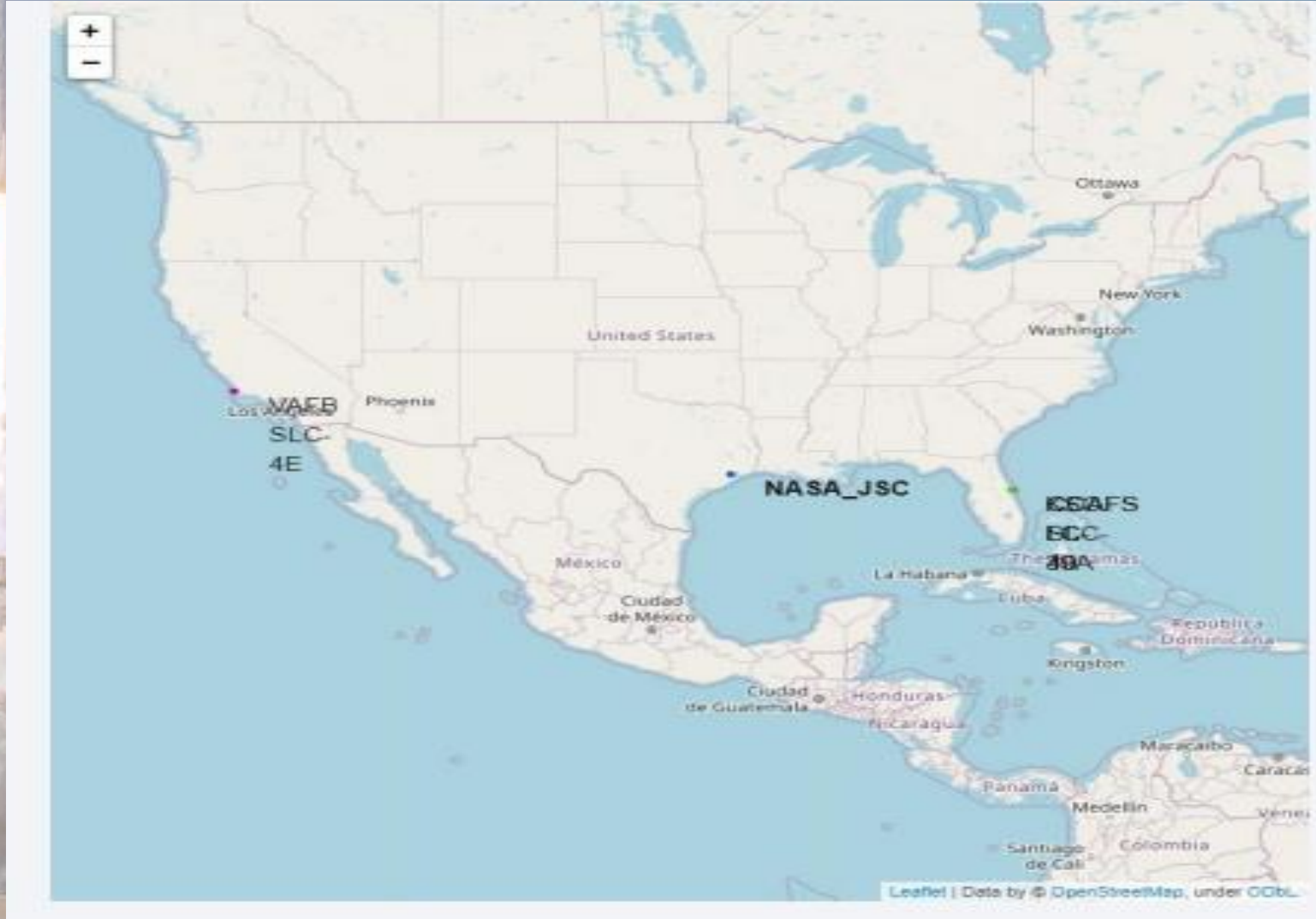
# Markers of Success/Failed Landings

- 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 site.
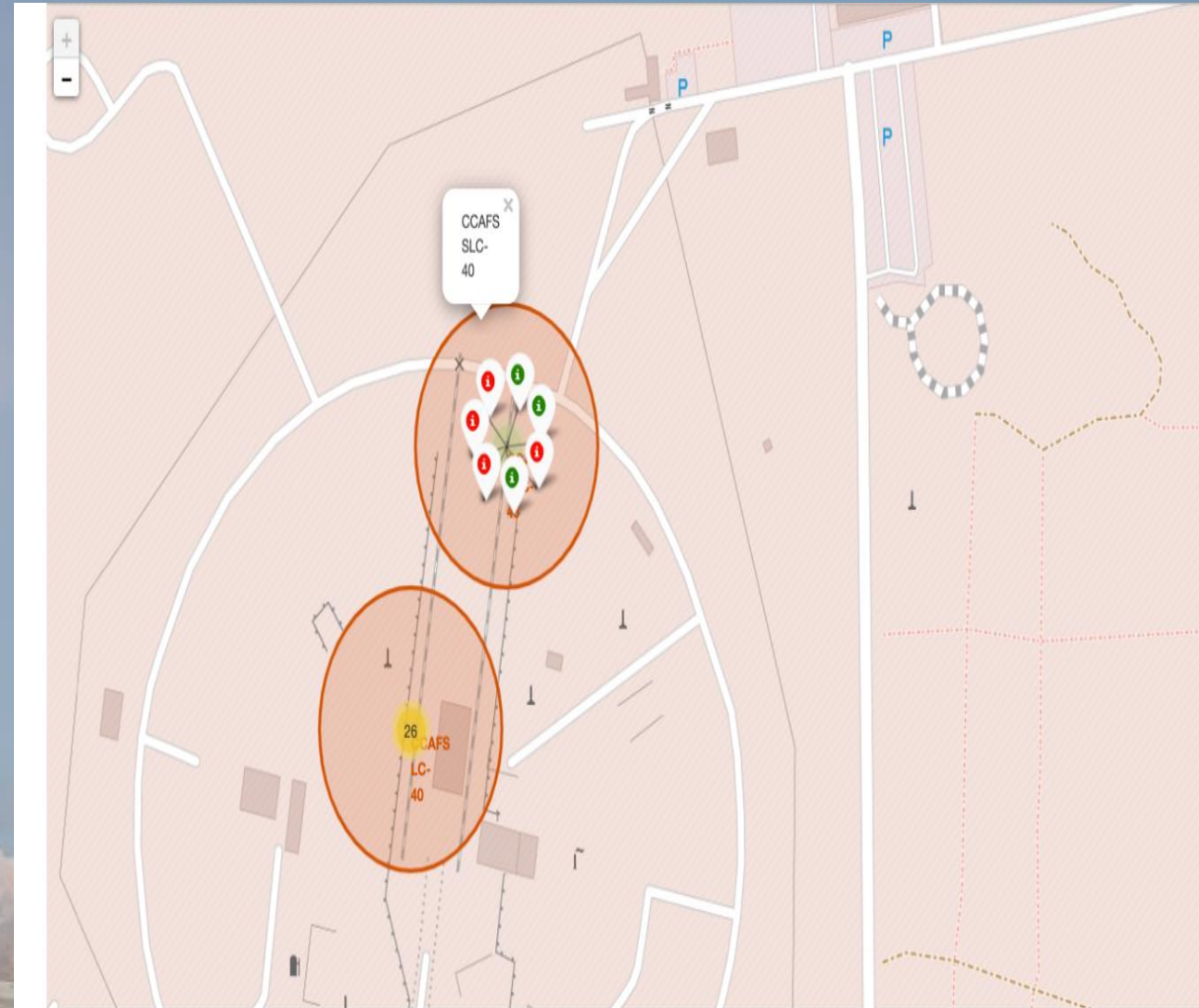
• The CCAFS LC-40 and CCAFS SLC-40 launch sites have coordinates that are close to being, but are not exactly, right on top of each other.

• The perimeter road around CCAFS LC-40 is 0.19 km away from the launch site coordinates.

• The coastline is 0.92 km away from CCAFS LC-40.

• The rail line is 1.33 km away from CCAFS LC-40.

Section 4

• The dropdown menu allows the selection of one or all launch sites.

• With all launch sites selected, the pie chart displays the distribution of successful Falcon 9 first stage landing outcomes between the different launch sites.

• The greatest share of successful Falcon 9 first stage landing outcomes (at 41.7% of the total) occurred at KSC LC-39A.

## SpaceX Launch Records Dashboard

All Sites

Total Success Launches By Site

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

29.2%
41.7%
16.7%
12.5%
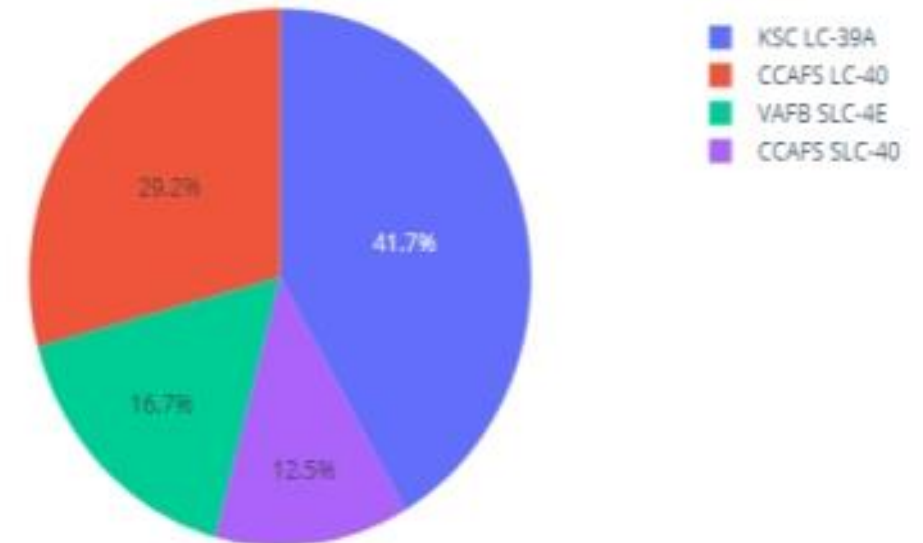
# Launch Site with Highest Launch Success Ratio

• Falcon 9 first stage failed landings are indicated by the '0' Class (■ blue wedge in the pie chart) and successful landings by the '1' Class (■ red wedge in the pie chart).

• CCAFS SLC-40 was the launch site that had the highest Falcon 9 first stage landing success rate (42.9%).



SpaceX Launch Records Dashboard

CCAFS SLC-40
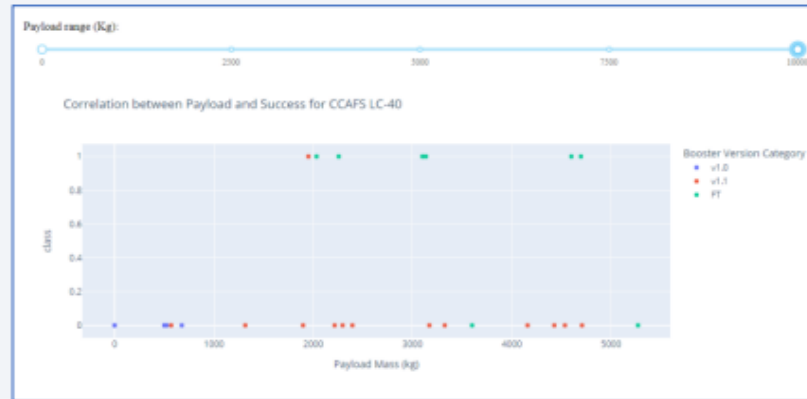
Total Success Launches for site CCAFS SLC-40

42.9%   57.1%
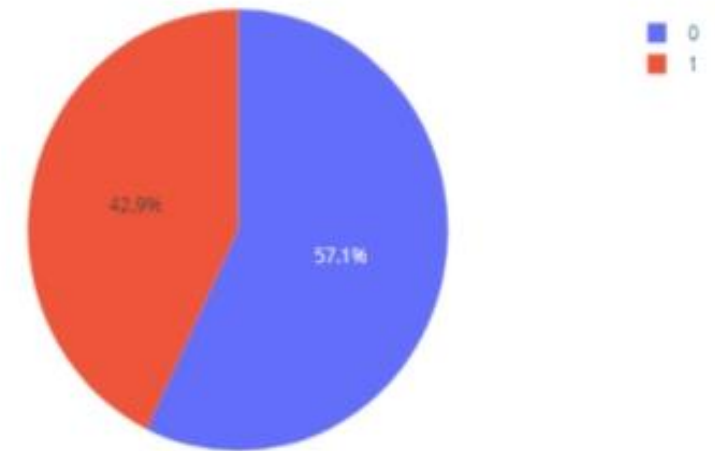
0
1

# <Dashboard Screenshot 3>

• These screenshots are of the Payload vs. Launch Outcome scatter plots for all sites, with different payload selected in the range slider.

• The payload range from about 2,000 kg to 5,000 kg has the largest success rate.

• The 'FT' booster version category has the largest success rate.



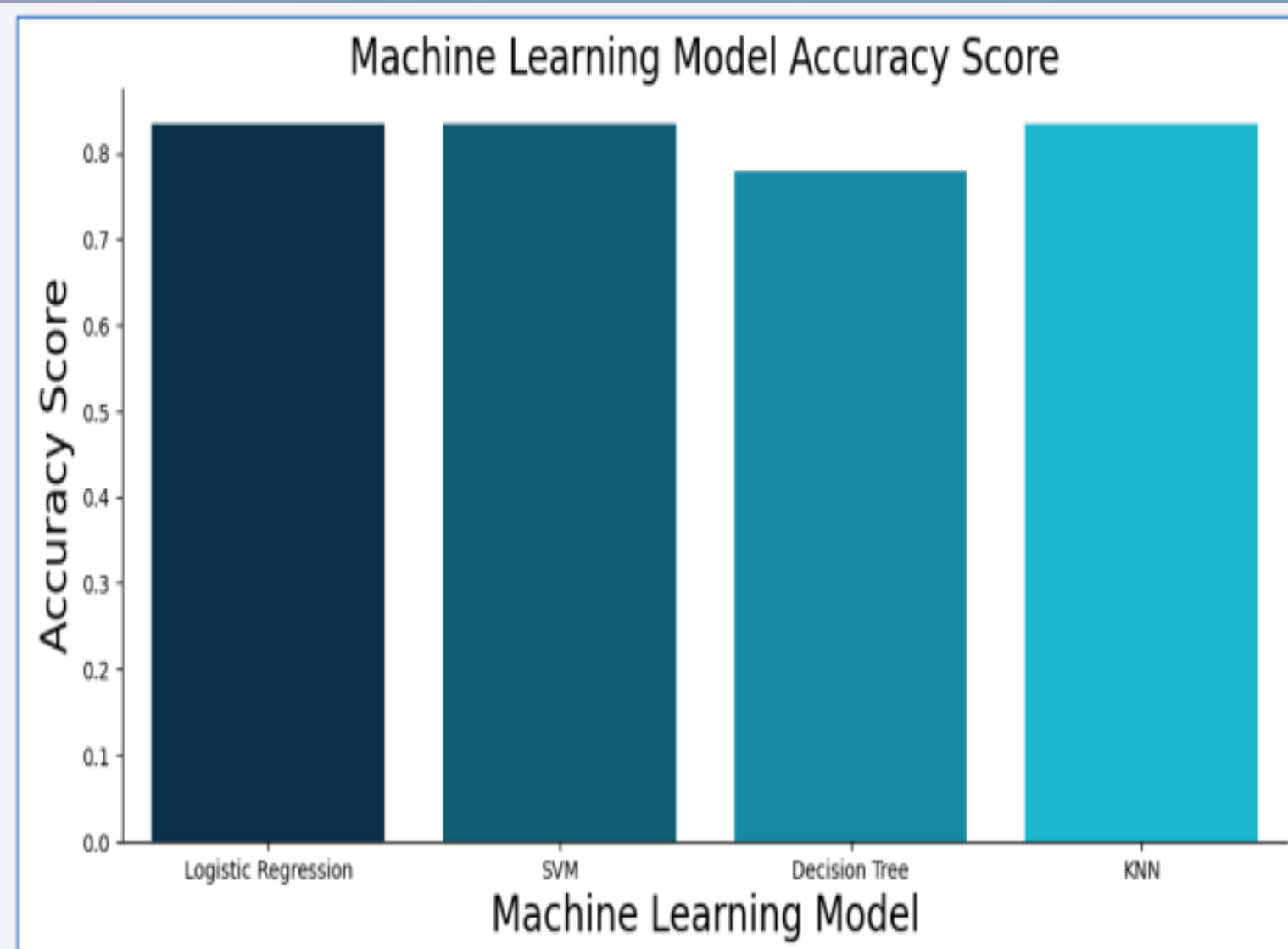CCAFS LC-40



CCAFS SLC-40



KSC LC-39A



VAFB SLC-4E

# Section 5

# Classification Accuracy

• All models performed equally well except for the Decision Tree model which performed poorly relative to the other models.
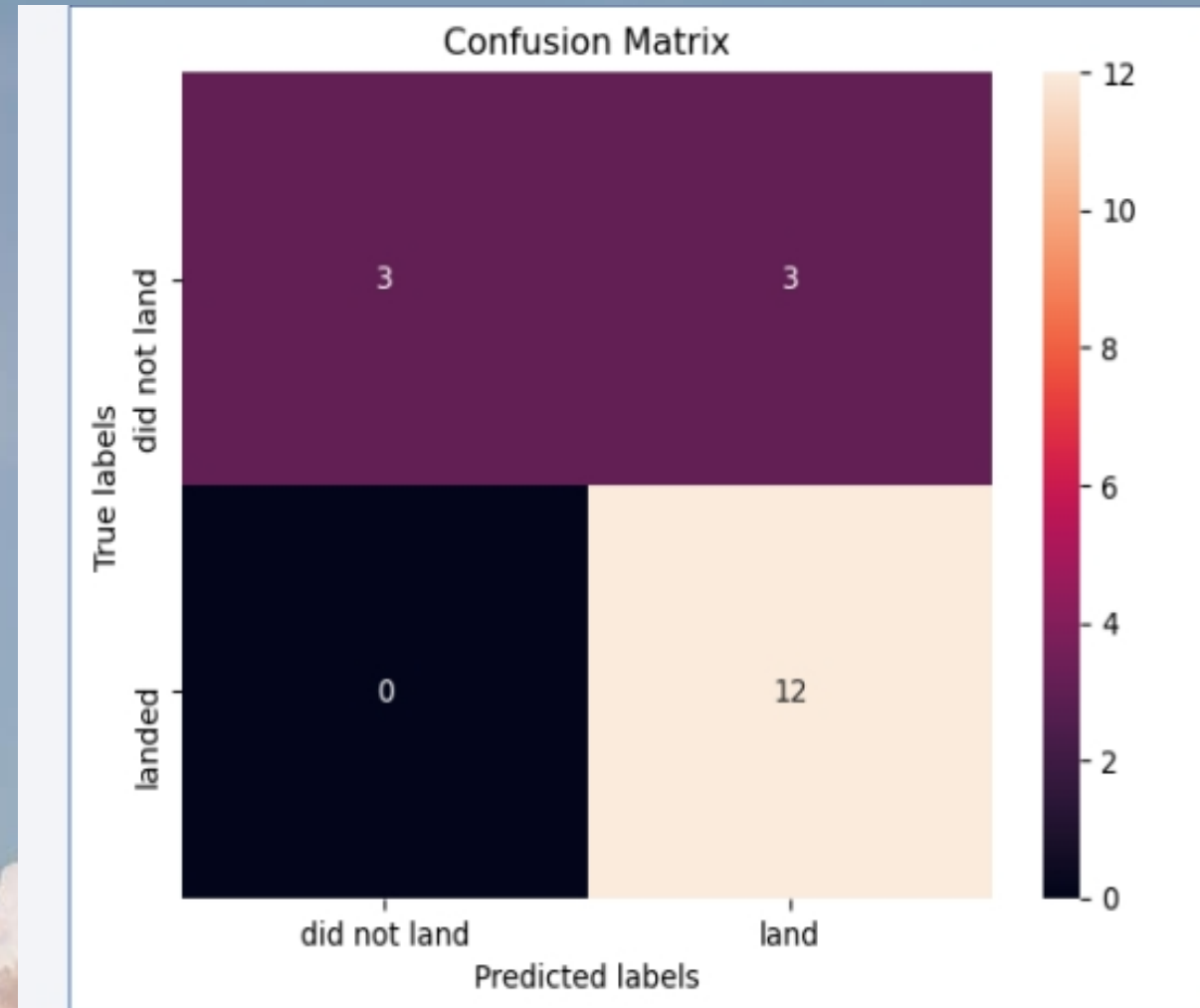
# Confusion Matrix

• Shown here is the confusion matrix for the Logistic Regression model.

• Confusion matrices can be read as:



• Prediction Breakdown:

• 12 True Positives and 3 True Negatives

• 3 False Positives and 0 Falsze Negatives

# Conclusions

- SpaceX does not have a perfect track record of Falcon 9 first stage landing outcomes

- SpaceX's Falcon 9 first stage landing outcomes have been trending towards greater success as more launches are made

- The machine learning models can be used to predict future SpaceX Falcon 9 first stage landing outcomes.

# Appendix

Jupyter notebooks and Dashboard python file

• GitHub URL (Data Collection):IBM-Applied-Data-Science-capstonee/jupyter-labs-spacex-data-collection-api (1).ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

• GitHub URL (Web screpping): IBM-Applied-Data-Science-capstonee/Spacex_Data_Collection_Web_Scraping.ipynb.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com

• GitHub URL (Data Wrangling):IBM-Applied-Data-Science-capstonee/Spacex-Data wrangling.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

• GitHub URL (EDA with SQL): IBM-Applied-Data-Science-capstonee/SpaceX_Falcon9_EDA_SQL.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

• GitHub URL (EDA with Data Visualization): IBM-Applied-Data-Science-capstonee/SpaceX_Falcon9_EDA_Vizulization.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

• GitHub URL (Maps from Folium): ): IBM-Applied-Data-Science-capstonee/SpaceX_Falcon9_Interactive_Folium.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

• GitHub URL (Dashboard): ): IBM-Applied-Data-Science-capstonee/07_SpaceX_Interactive_Visual_Analytics_Plotly.py at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)

• GitHub URL (Machine Learning):IBM-Applied-Data-Science-capstonee/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · bunny0033/IBM-Applied-Data-Science-capstonee (github.com)