IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nick Bignell
20250411

Outline

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

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.
- The main steps in this project include:
  - Data collection, wrangling, and formatting
  - Exploratory data analysis
  - Interactive data visualization
  - Machine learning prediction
- Our graphs show that some features of the rocket launches have a correlation with the outcome of the launches, i.e., success or failure.
- It is also concluded that decision tree may be the best machine learning algorithm to predict if the Falcon 9 first stage will land successfully.

Introduction

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch.
- Most unsuccessful landings are planned. Sometimes, SpaceX will perform a controlled landing in the ocean.
- The main question that we are trying to answer is, for a given set of features about a Falcon 9 rocket launch which include its payload mass, orbit type, launch site, and so on, will the first stage of the rocket land successfully?

Section 1

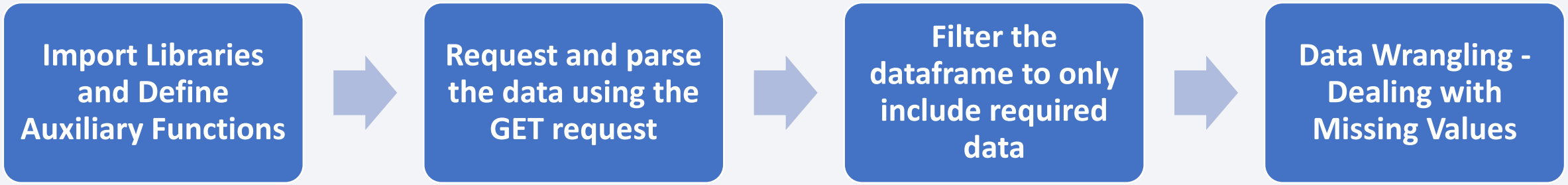
Methodology

Methodology

- The overall methodology includes:
  1. Data collection, wrangling, and formatting, using:
     - SpaceX API
     - Web scraping
  2. Exploratory data analysis (EDA), using:
     - Pandas and NumPy
     - SQL
  3. Data visualization, using:
     - Matplotlib and Seaborn
     - Folium
     - Dash
  4. Machine learning prediction, using
     - Logistic regression
     - Support vector machine (SVM)
     - Decision tree
     - K-nearest neighbors (KNN)

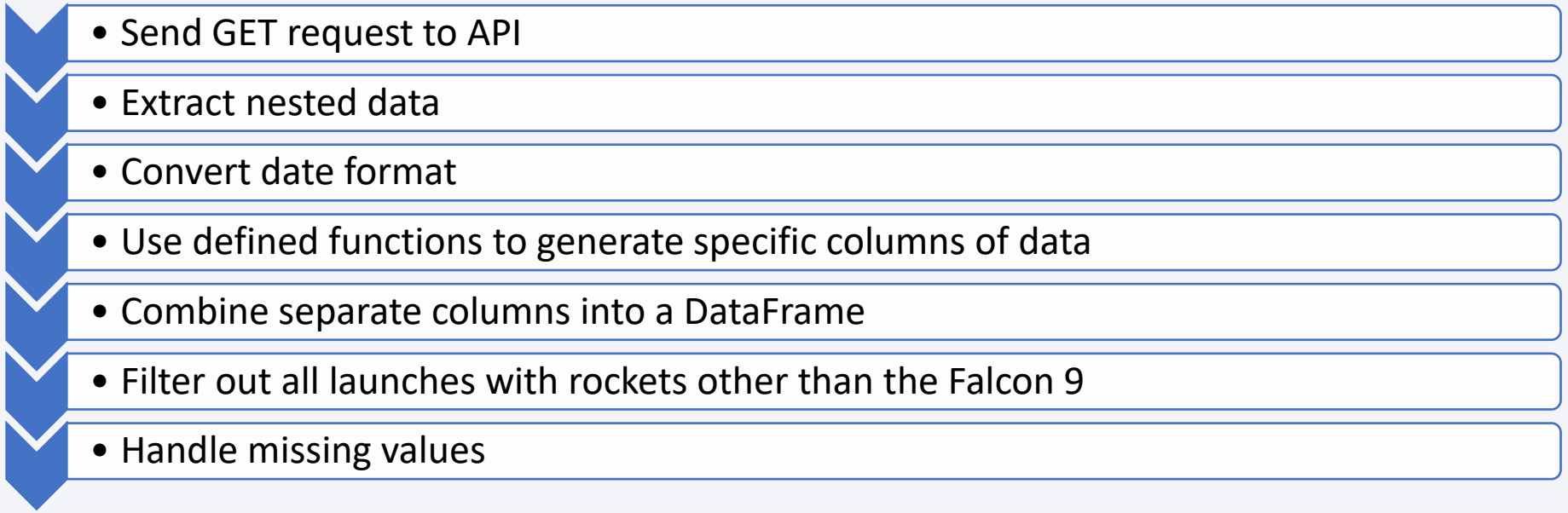
Data Collection

- The data sets were collected from:
- •An IBM copy of a response from a publicly accessible API with launch data in JSON format.
- •A permanently-linked Wikipedia page with launch data in HTML tables.
- •Additional data sets were provided with the course in CSV file format



Data Collection – SpaceX API

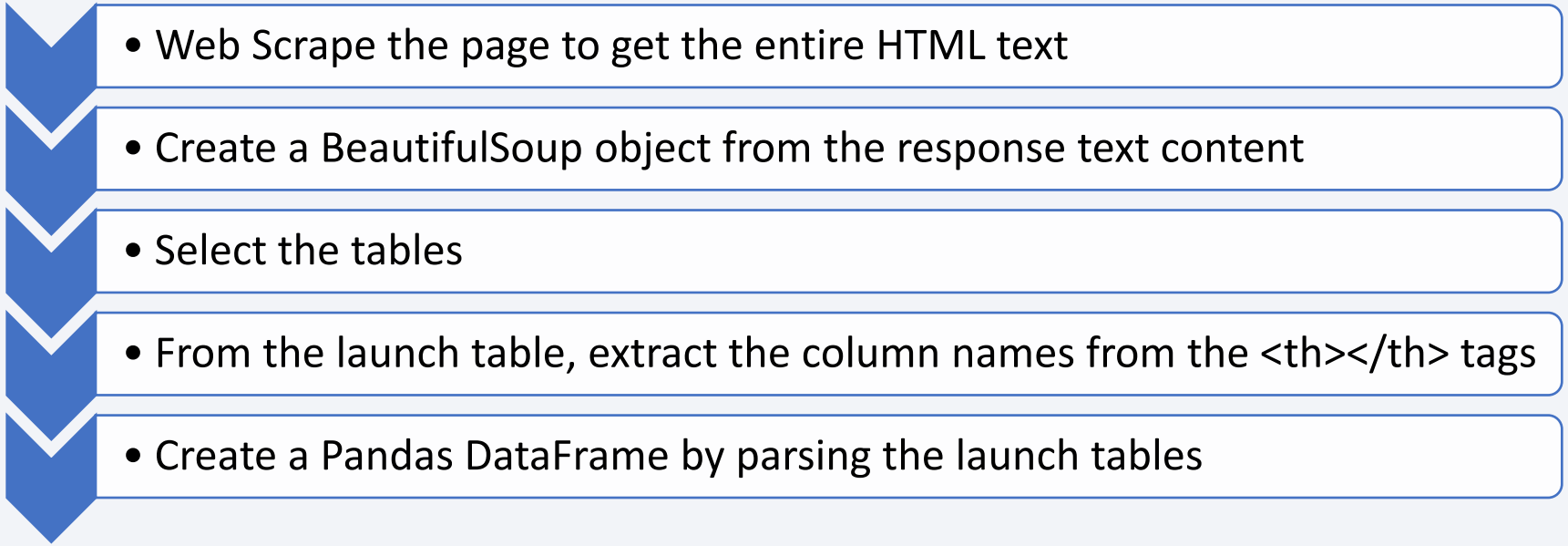
- SpaceX API
  - The API used is https://api.spacexdata.com/v4/rockets/.
  - The API provides data about many types of rocket launches done by SpaceX, the data is therefore filtered to include only Falcon 9 launches.
  - Every missing value in the data is replaced the mean the column that the missing value belongs to.



GitHub URL of completed data wrangling related notebooks

Data Collection - Scraping

- Web scraping
  - The data is scraped from https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
  - Launch data was extracted from these tables and loaded into a Pandas DataFrame for further analysis.



GitHub URL of completed data wrangling related notebooks

Data Wrangling

- The data is later processed so that there are no missing entries and categorical features are encoded using one-hot encoding.
- An extra column called ‘Class’ is also added to the data frame. The column ‘Class’ contains 0 if a given launch is failed and 1 if it is successful.

- Load data from CSV file from earlier
- Find the number of launches at each site
- Find the quantity of each type of orbit
- Find the quantity of each type of mission outcome
- Create a DataFrame column from the mission outcome data
- Compile data into a single DataFrame

GitHub URL of your completed data wrangling related notebooks

EDA with Data Visualization

- Matplotlib and Seaborn
  - Functions from the Matplotlib and Seaborn libraries are used to visualize the data through scatterplots, bar charts, and line charts.
  - The plots and charts are used to understand more about the relationships between several features, such as:
    - The relationship between flight number and launch site
    - The relationship between payload mass and launch site
    - The relationship between success rate and orbit type
- GitHub URL of your completed EDA with data visualization notebook

EDA with SQL

- SQL
  - The data is queried using SQL to answer several questions about the data such as:
    - The names of the unique launch sites in the space mission
    - The total payload mass carried by boosters launched by NASA (CRS)
    - The average payload mass carried by booster version F9 v1.1
- GitHub URL of your completed EDA with SQL notebook

Build an Interactive Map with Folium

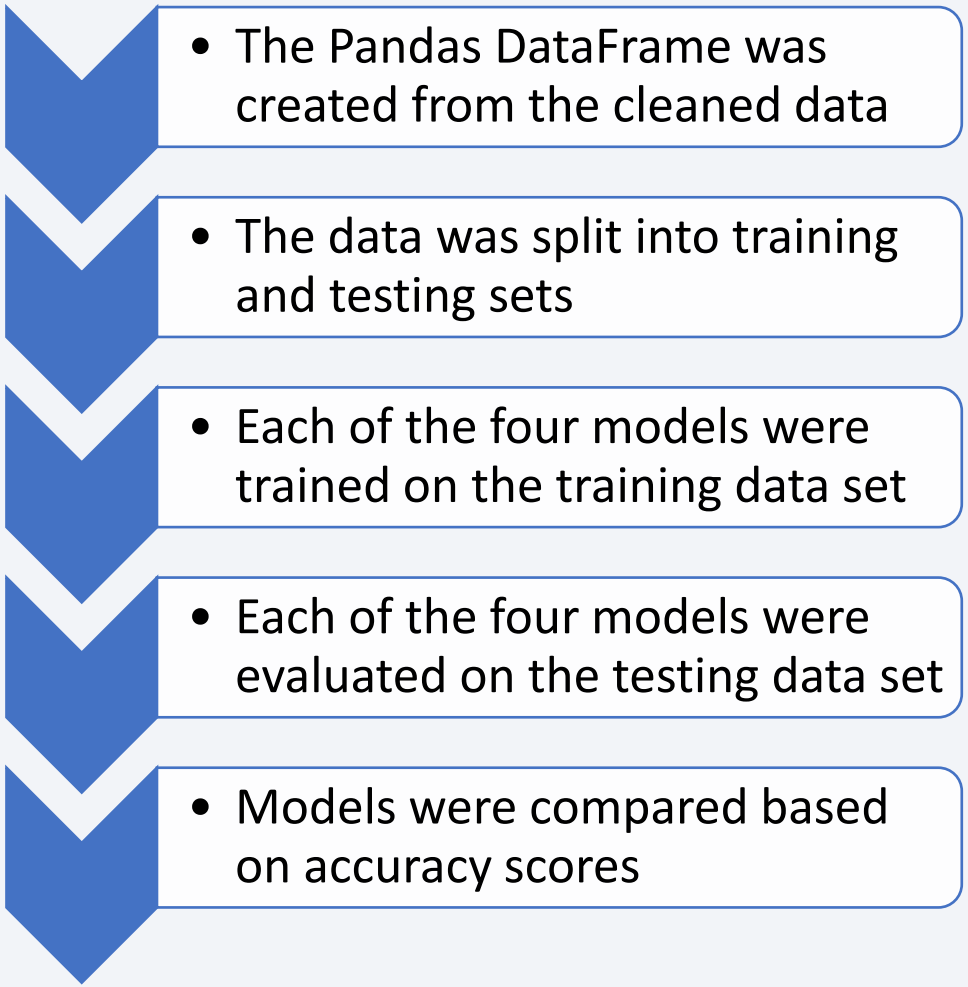
- Folium
  - Functions from the Folium libraries are used to visualize the data through interactive maps.
  - The Folium library is used to:
    - Mark all launch sites on a map
    - Mark the succeeded launches and failed launches for each site on the map
    - Mark the distances between a launch site to its proximities such as the nearest city, railway, or highway
- GitHub URL of your completed interactive map with Folium map

Build a Dashboard with Plotly Dash

- Dash
  - Functions from Dash are used to generate an interactive site where we can toggle the input using a dropdown menu and a range slider.
  - Using a pie chart and a scatterplot, the interactive site shows:
    - The total success launches from each launch site
    - The correlation between payload mass and mission outcome (success or failure) for each launch site
- GitHub URL of your completed Plotly Dash lab

Predictive Analysis (Classification)

- Functions from the Scikit-learn library are used to create our machine learning models.
- The machine learning prediction phase include the following steps:
  - Standardizing the data
  - Splitting the data into training and test data
  - Creating machine learning models, which include:
    - Logistic regression
    - Support vector machine (SVM)
    - Decision tree
    - K nearest neighbors (KNN)
  - Fit the models on the training set
  - Find the best combination of hyperparameters for each model
  - Evaluate the models based on their accuracy scores and confusion matrix
- GitHub URL of your completed predictive analysis lab



Results

- The results are split into 5 sections:
  - SQL (EDA with SQL)
  - Matplotlib and Seaborn (EDA with Visualization)
  - Folium
  - Dash
  - Predictive Analysis
- In all of the graphs that follow, class 0 represents a failed launch outcome while class 1 represents a successful launch outcome.

Nick Bignell

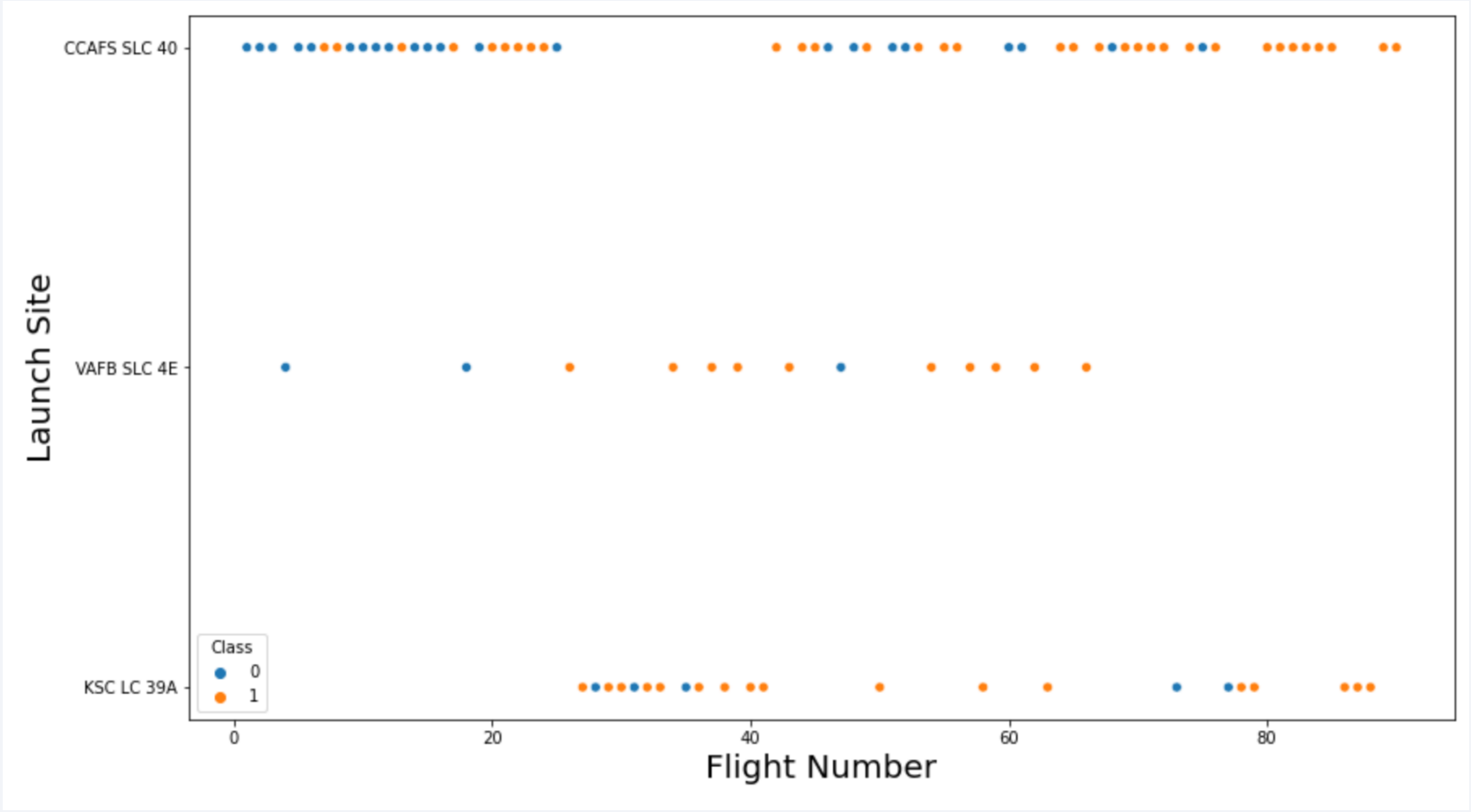
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

The relationship between flight number and launch site

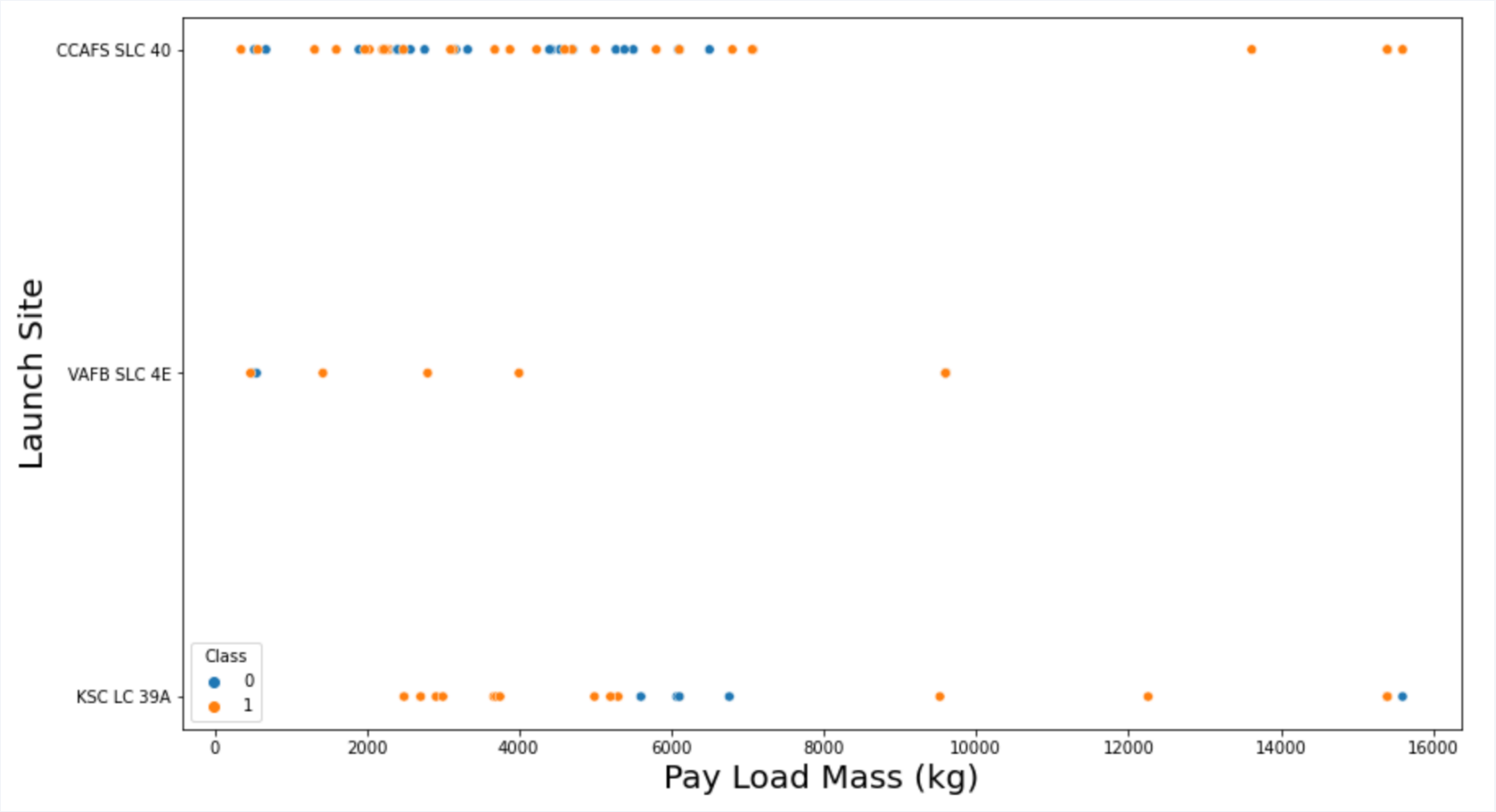
Success rate varied noticeably with launch site. Successful Falcon 9 first stage landings appear to become more prevalent as the flight number increases.



Payload vs. Launch Site

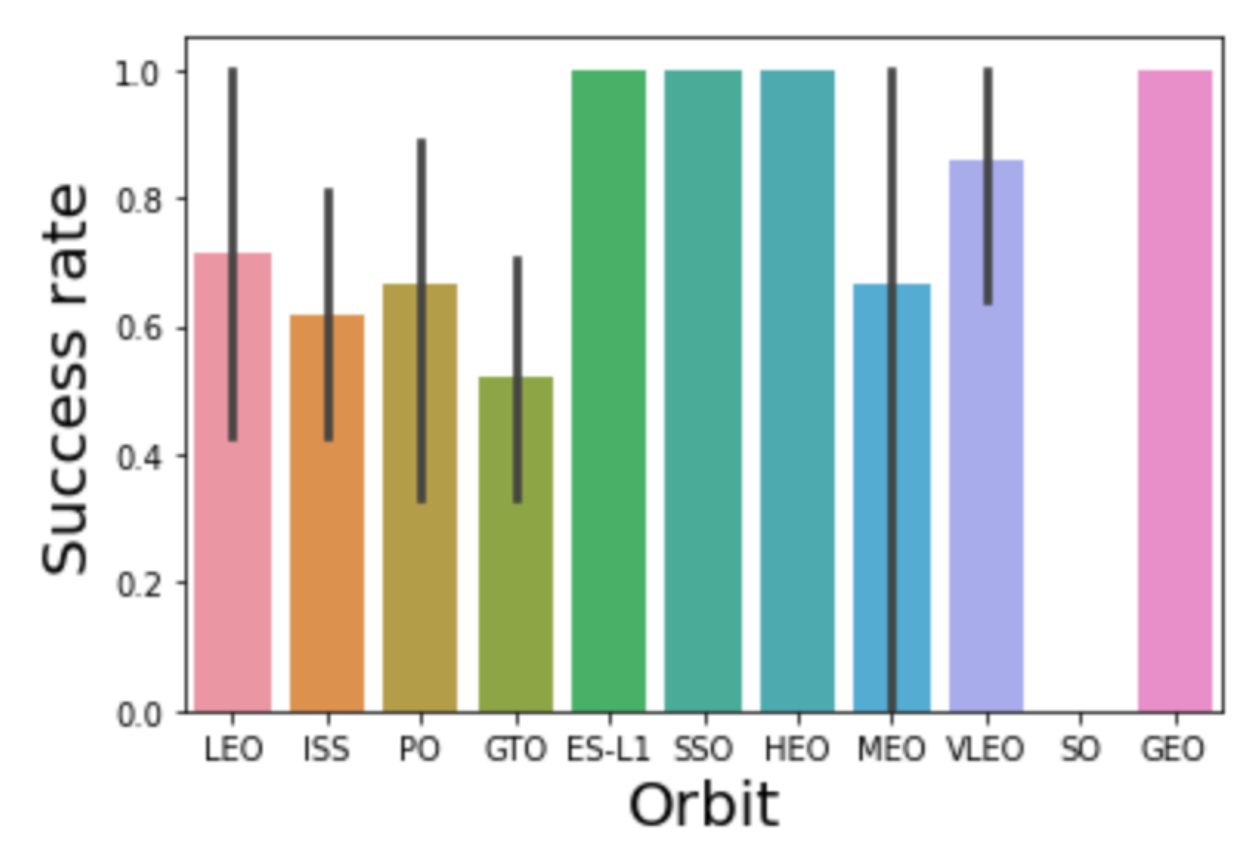
- For the CCAFS SLC 40 launch site, the payload mass and the landing outcome appear to not be strongly correlated.
- The failed landings at the KSC LC 39A launch site are mostly grouped around a narrow band of payload masses.

The relationship between payload mass and launch site



Success Rate vs. Orbit Type

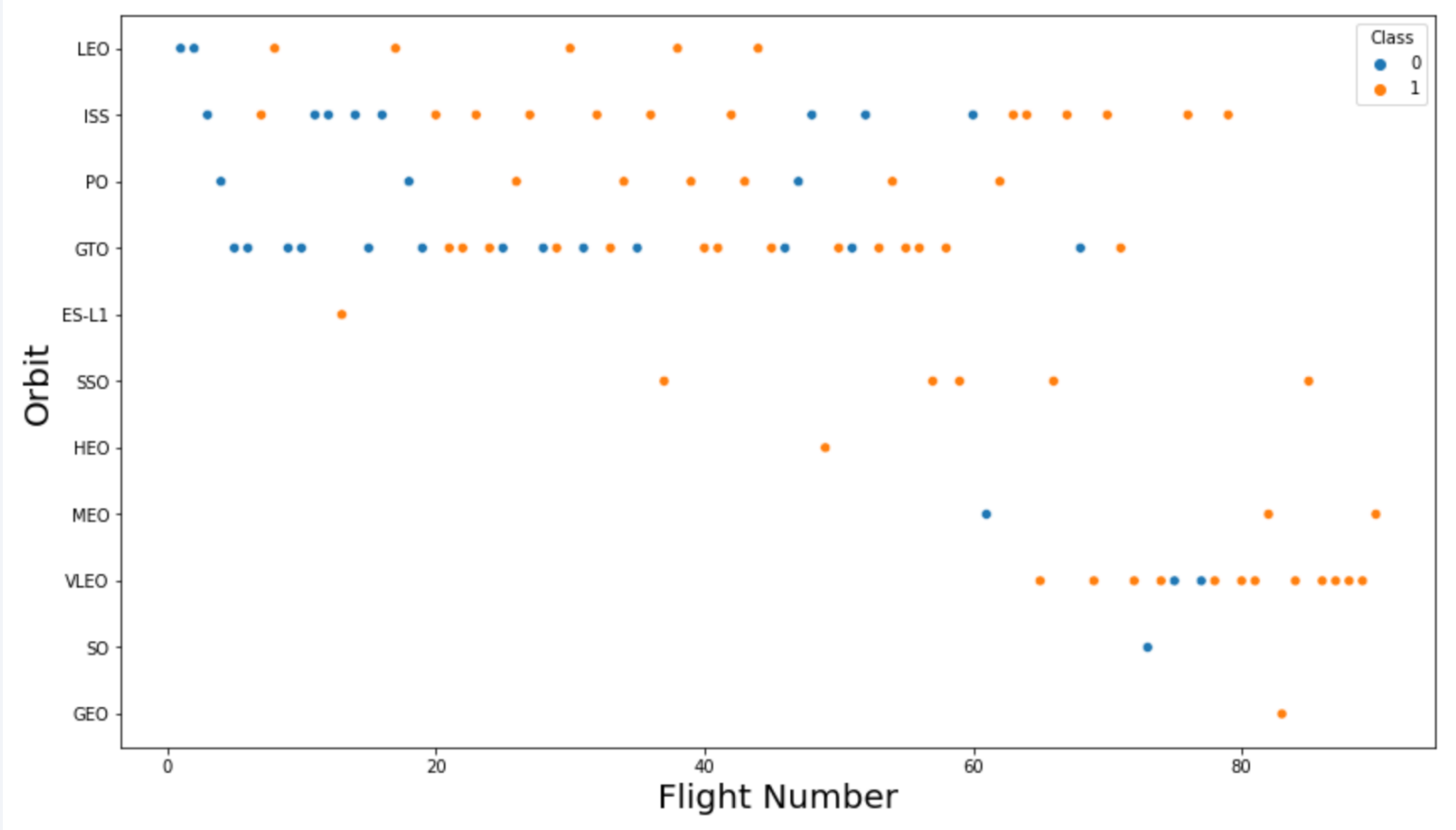
- ES-L1, SSO, HEO and GEO orbits have no failed first stage landings.
- SO orbits have no successful first stage landings.



Flight Number vs. Orbit Type

- There is a positive correlation between flight number and success rate. (I.e. Larger flight numbers were associated with higher success rates.)

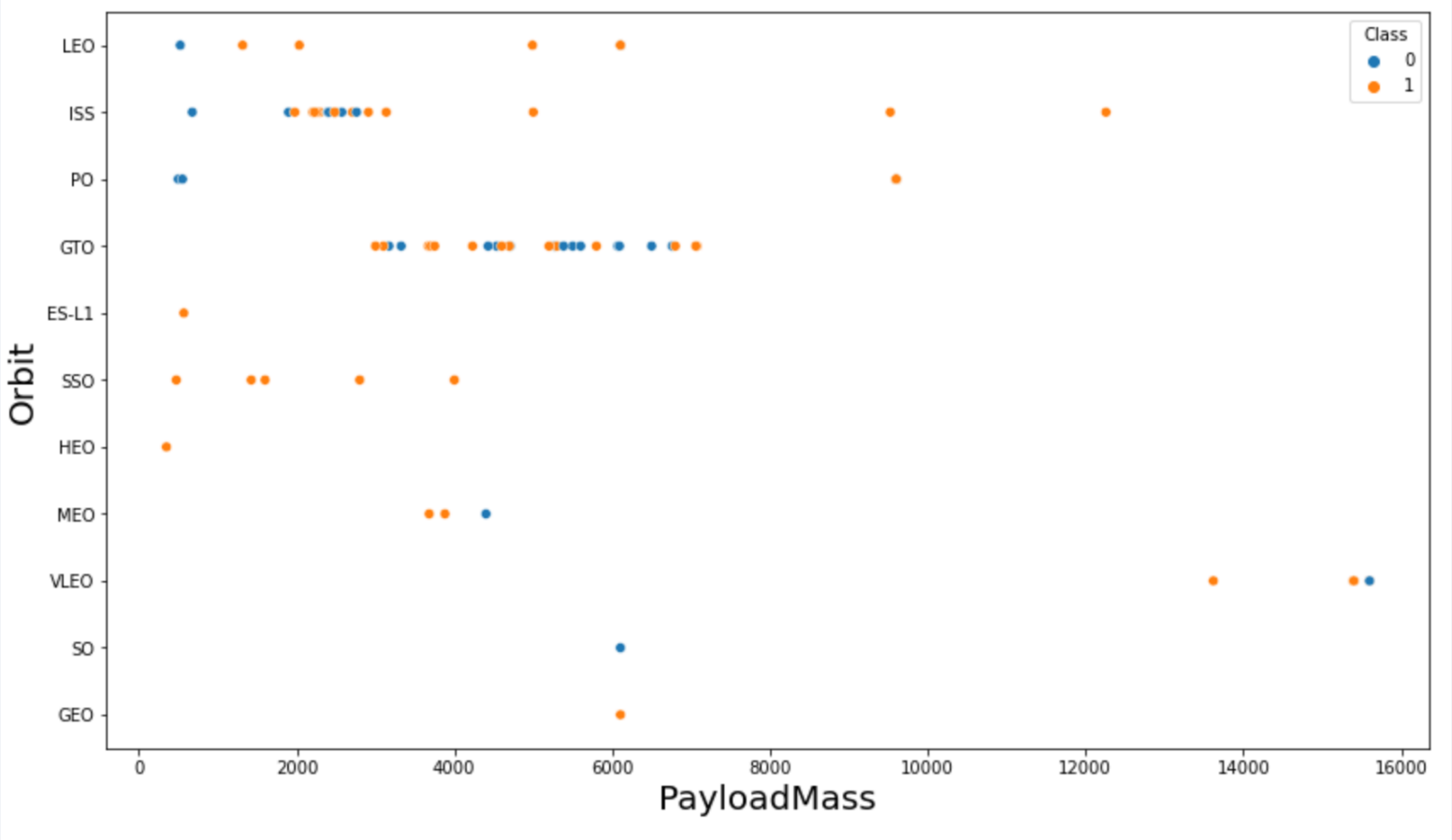
The relationship between flight number and orbit type



Payload vs. Orbit Type

- Some orbit types showed higher success rates than others.
- Success rate appeared to have no obvious correlation with payload mass.

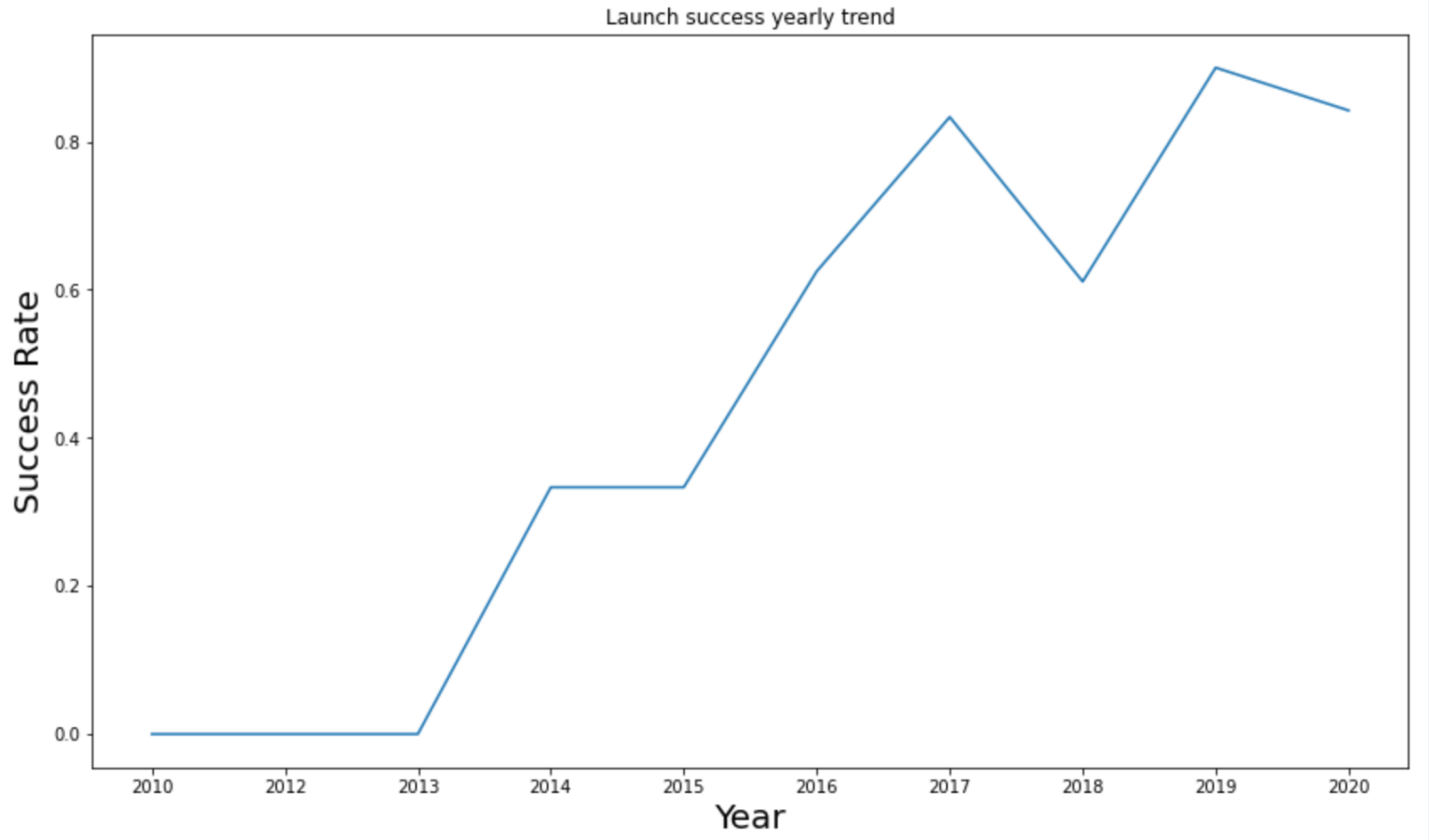
The relationship between payload mass and orbit type



Launch Success Yearly Trend

- The success rate of the Falcon 9 first stage landings has increased significantly over the selected interval of years.

The launch success yearly trend



All Launch Site Names

- 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.

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

- Query: SELECT* FROM`SPACEXDATASET` WHERE`launch_site` LIKE'CCA%' LIMIT5;
- 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.

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- 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

- Calculate the average payload mass carried by booster version F9 v1.1
- Query:SELECT sum(`payload_mass__kg_`) / count(`payload_mass__kg_`) AS“Average Payload Mass (kg)"FROM `SPACEXDATASET` 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

- Find the dates of the first successful landing outcome on ground pad
- Query:SELECTmin(`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 22nd December 2015.

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
- Query:
```
SELECT DISTINCT `booster_version` FROM`SPACEXDATASET`
WHERE`landing__outcome` = 'Success (drone ship)'AND`payload_mass__kg_`
BETWEEN4000 AND6000;
```
- 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.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

- Query:SELECT (SELECTcount(*) FROM`SPACEXDATASET` WHERElcase(`landing__outcome`) LIKE'%success%') AS"Success“, count(*) AS"Failure"FROM`SPACEXDATASET` WHERElcase(`landing__outcome`) NOTLIKE'%success%';
- Result:

Success	Failure
61	40

- Explanation: There were 61 successful and 40 failed mission outcomes.

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Query: SELECT`booster_version`, `payload_mass__kg_`FROM`SPACEXDATASET` WHERE`payload_mass__kg_` = (SELECTmax(`payload_mass__kg_`) FROM`SPACEXDATASET`);
- Result:
- Explanation: The maximum payload mass carried in this dataset is 15,600 kg. Twelve (12) separate Falcon 9 boosters carried this amount of payload mass.

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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Query: SELECTMONTHNAME(`DATE`) AS 'Month', `landing__outcome`, `booster_version`, `launch_site` FROM`SPACEXDATASET` WHERE`landing__outcome` = 'Failure (drone ship)' ANDYEAR(`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

- 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
- 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 ‘No attempt’.

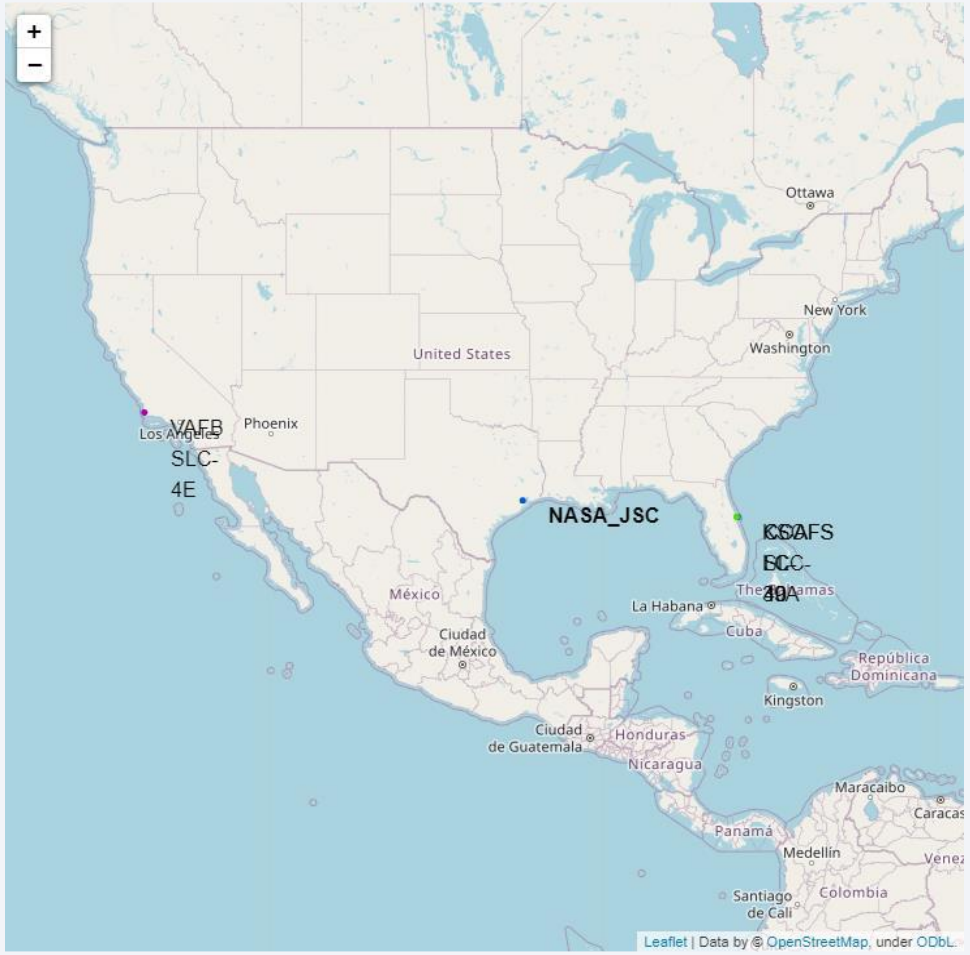
Nick Bignell

Section 3

Launch Sites Proximities Analysis

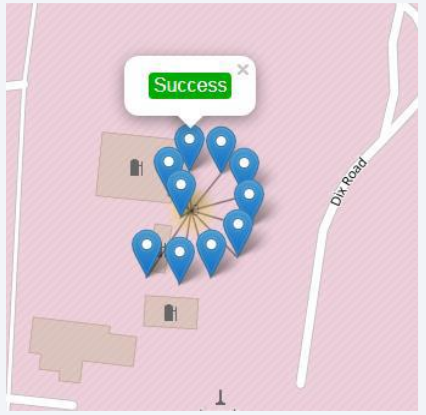
Falcon 9 Launch Site Locations

- California, USA
  - VAFB SLC-4E | Vandenberg Air Force Base Space Launch Complex 4E
- Florida, USA
  - KSC LC-39A | Kennedy Space Center Launch Complex 39A
  - CCAFS LC-40 | Cape Canaveral Air Force Station Launch Complex 40
  - CCAFS SLC-40 | Cape Canaveral Air Force Station Space Launch Complex 40

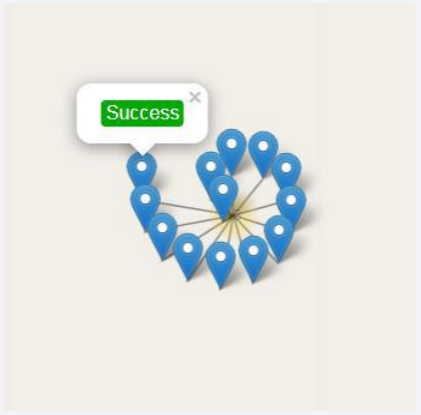


Map Markers of Success/Failed Landings

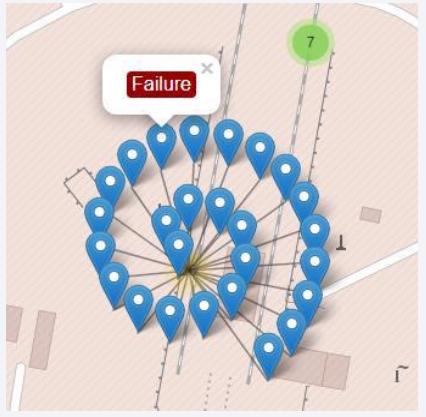
- The markers display the mission outcomes (Success/Failure) for Falcon 9 first stage landings. They are grouped on the map to be associated with the geographical coordinates for the launch site.
- A sense of a launch site's success rate for Falcon 9 first stage landings can be gleaned from the relative number of green success markers to red failure markers.



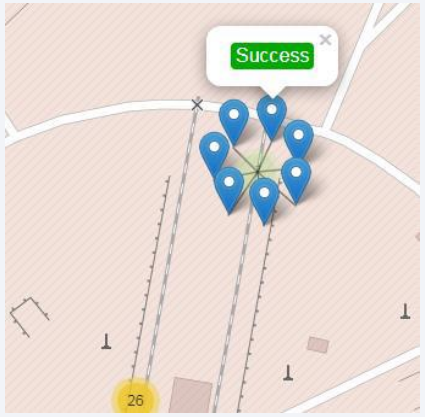
VAFB SLC-4E



KSC LC-39A



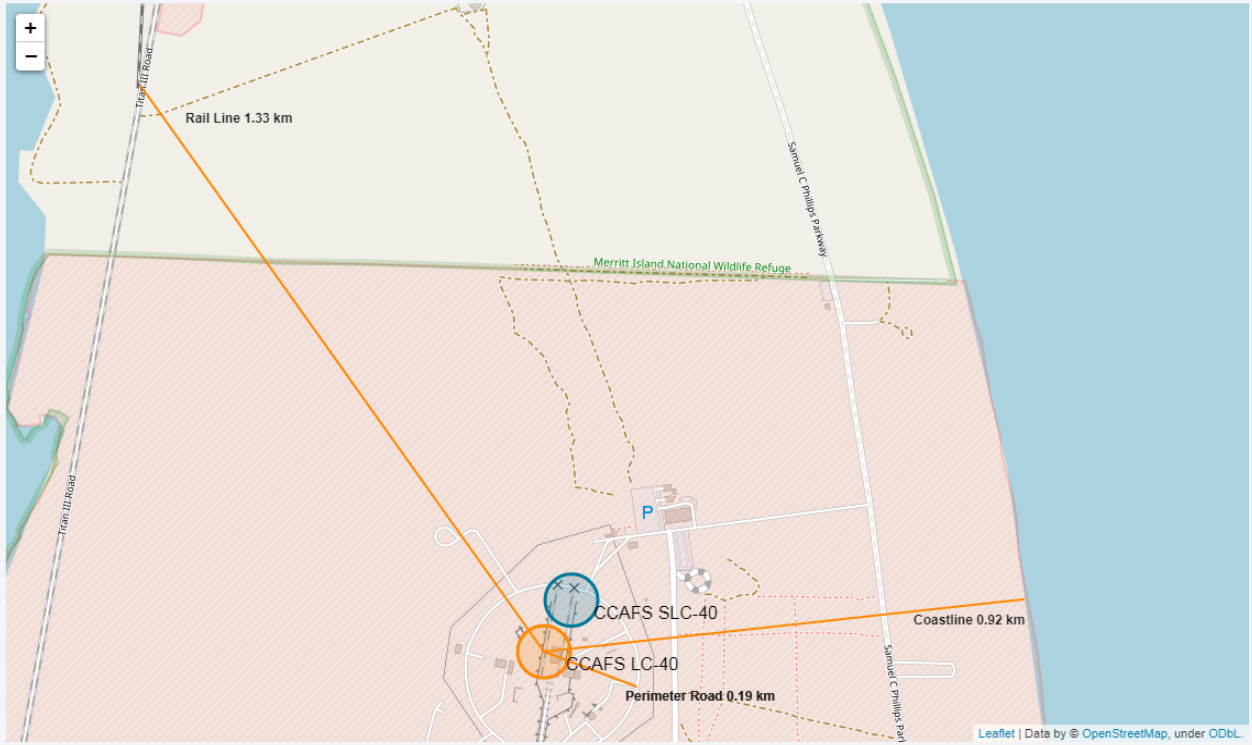
CCAFS LC-40

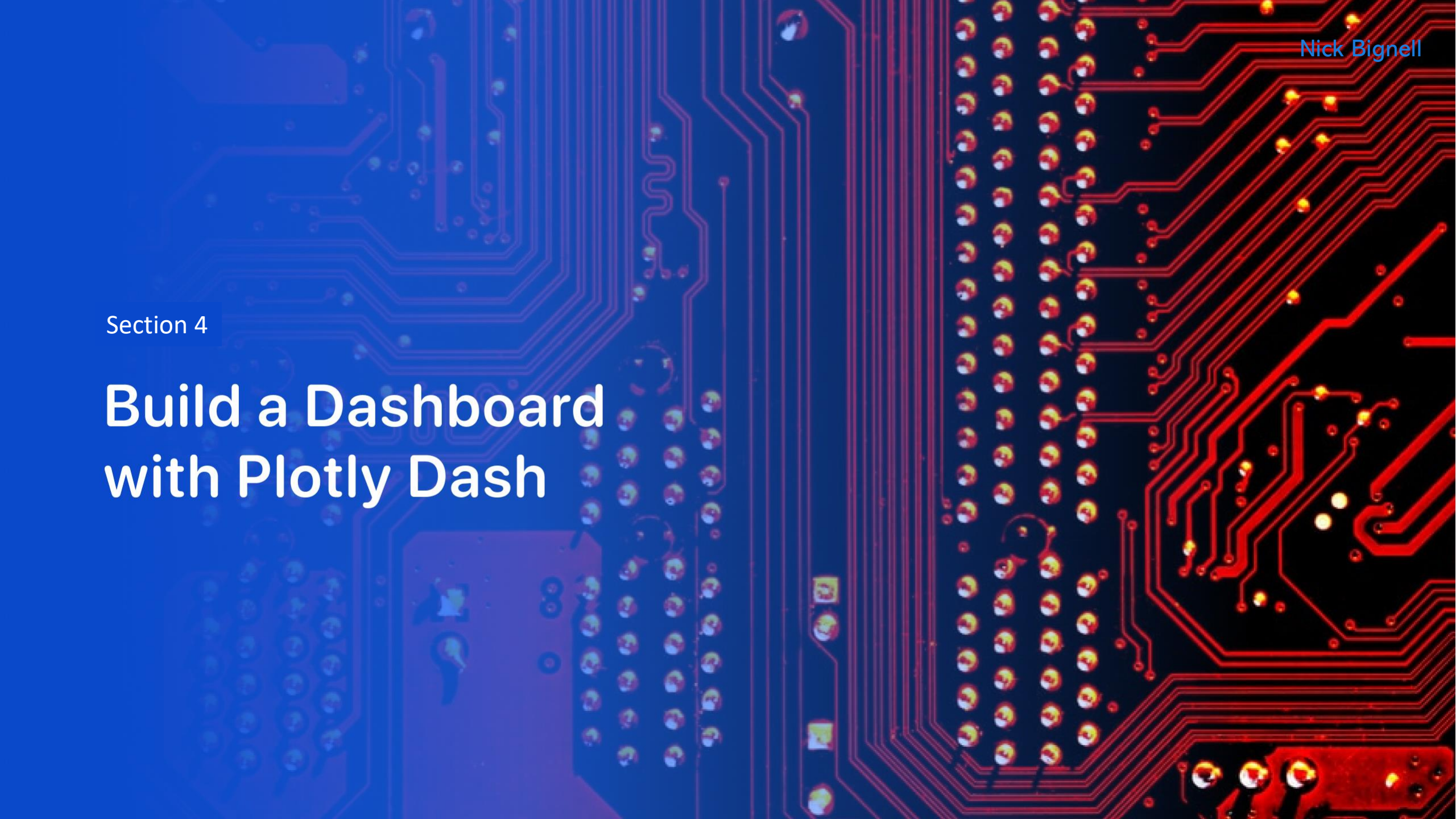


CCAFS SLC-40

Distance from Launch Site to Proximities

- 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.





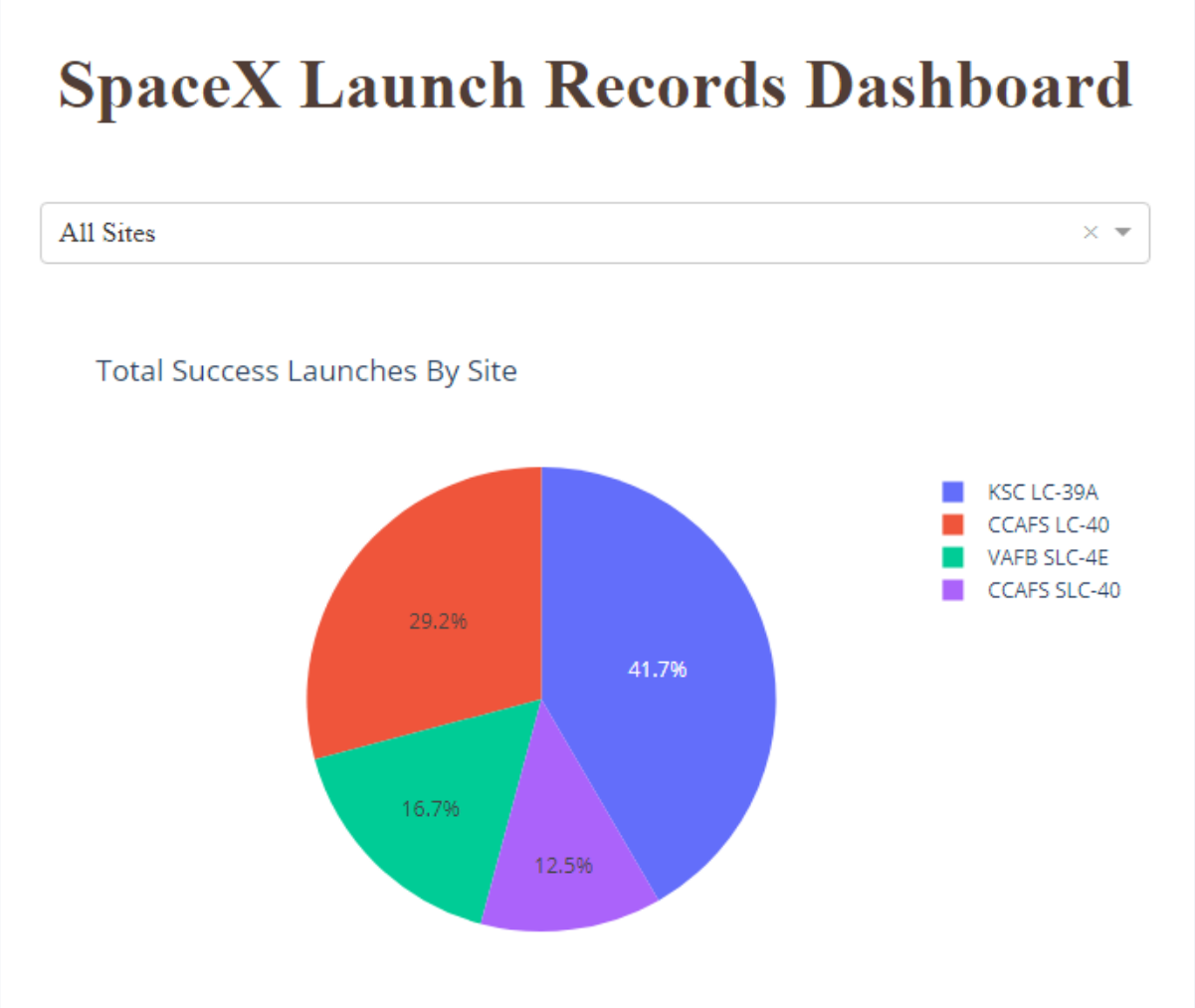
Nick Bignell

Section 4

Build a Dashboard with Plotly Dash

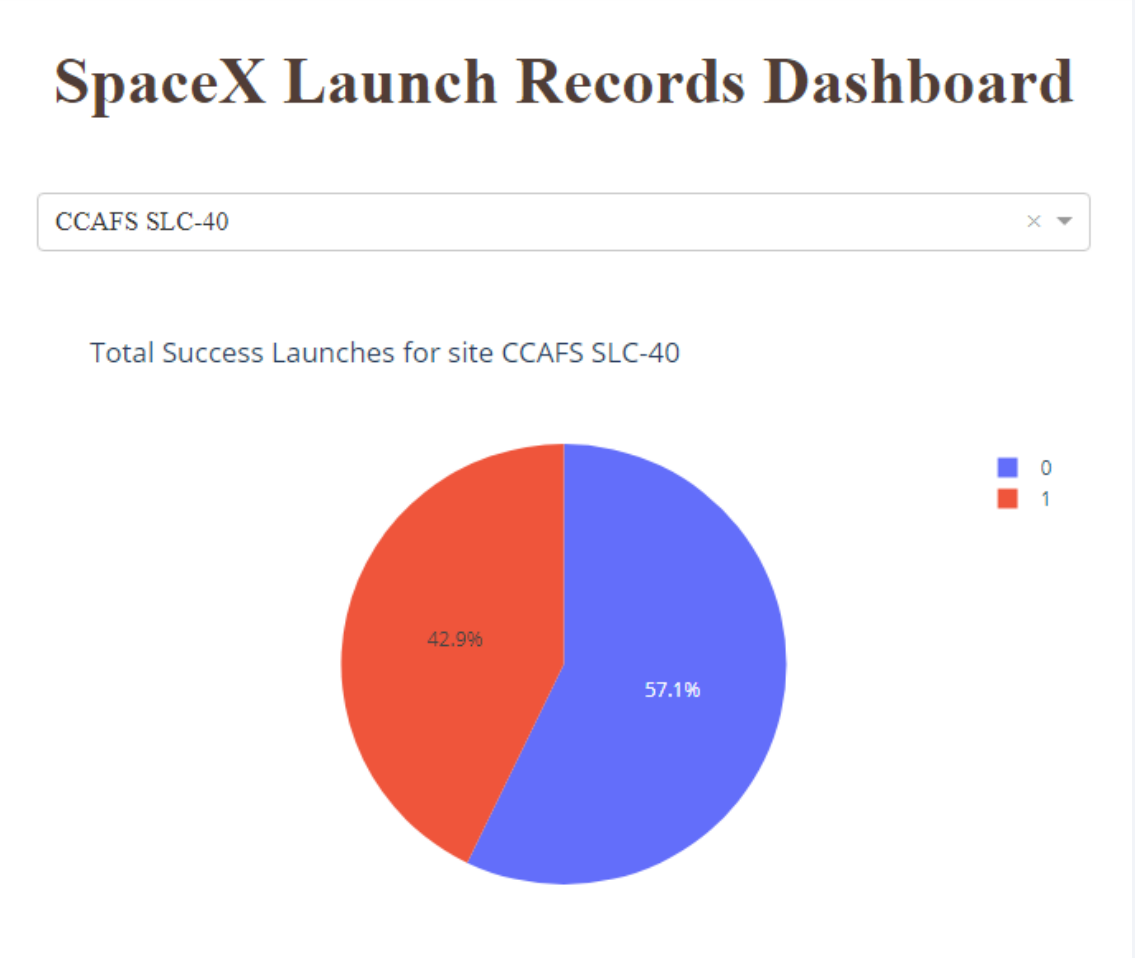
<Dashboard Screenshot 1>

- The dropdown menu allowed the selection of one or all launch sites.
- 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 greatest share of successful Falcon 9 first stage landing outcomes (at 41.7% of the total) occurred at KSC LC-39A.



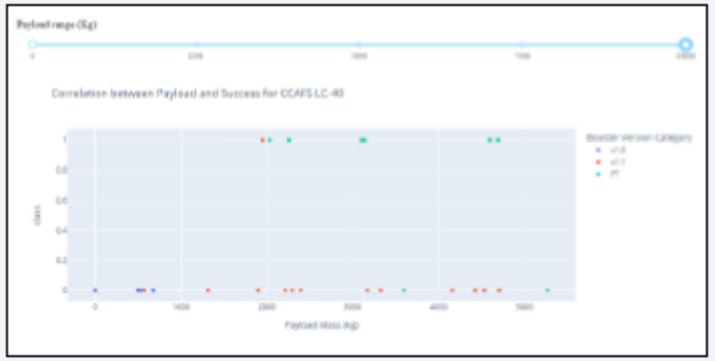
Launch Site with Highest Launch Success Ratio

- •Falcon 9 first stage failed landingsare 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%).

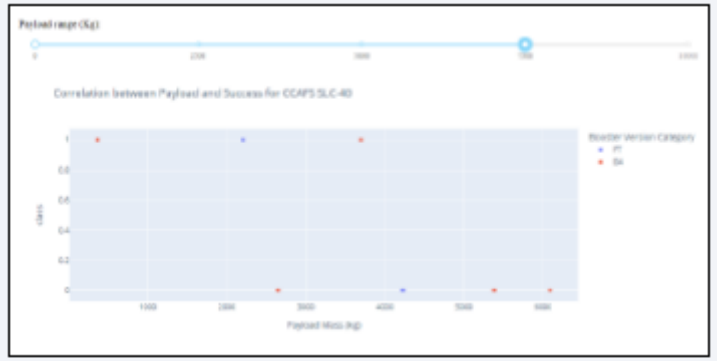


Payload vs. Launch Outcome

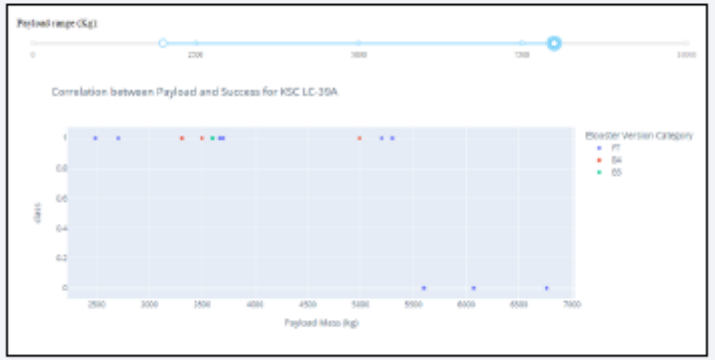
- •These screenshots are of the Payload vs. Launch Outcome scatter plots for all sites, with different payload selected in the range slider.
- The payload range from about 2,000 kg to 5,000 kg has the largest success rate.
- The ‘FT’ booster version category has the largest success rate.



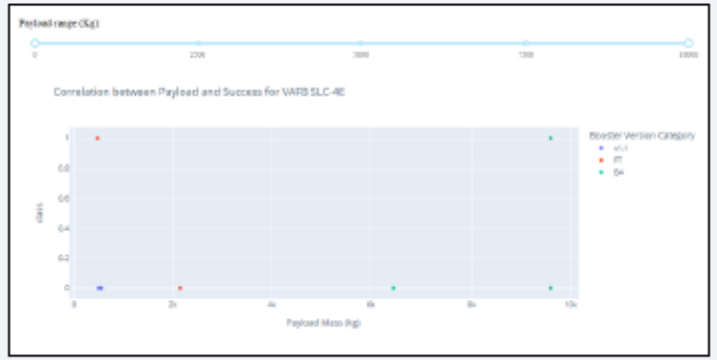
CCAFS LC-40



CCAFS SLC-40



KSC LC-39A



VAFB SLC-4E

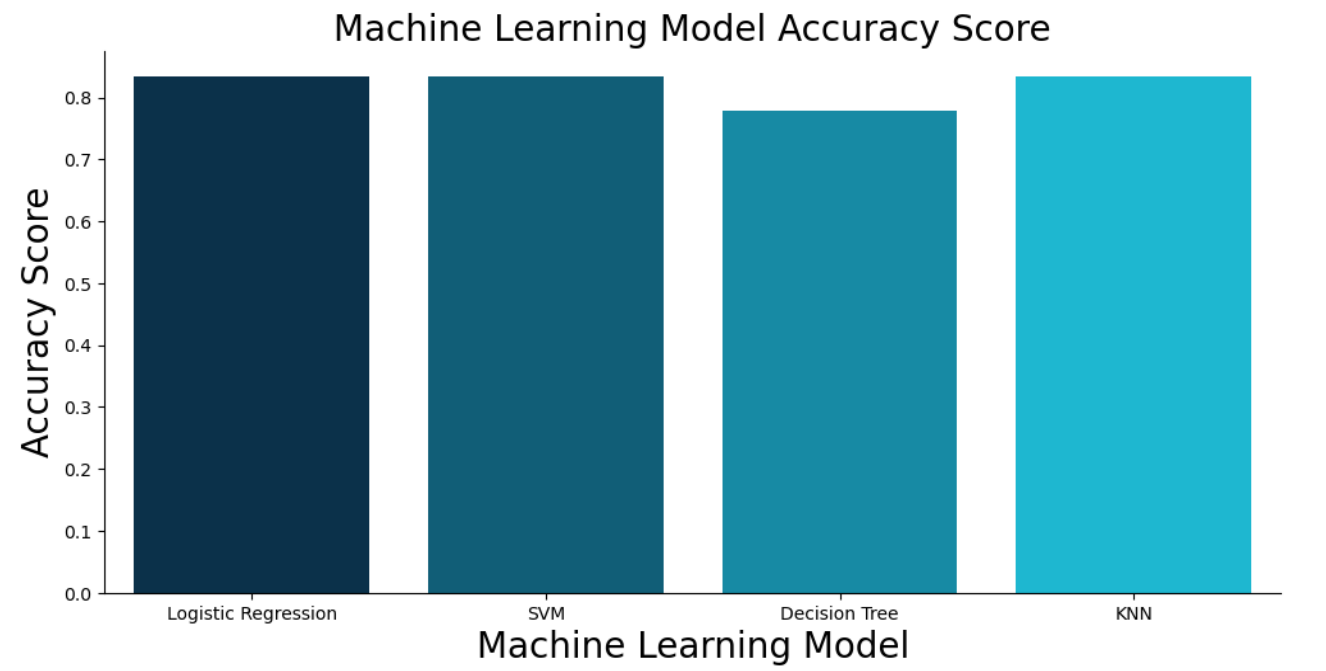
Nick Bignell

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All models performed equally well except for the Decision Tree model which performed poorly relative to the other models.

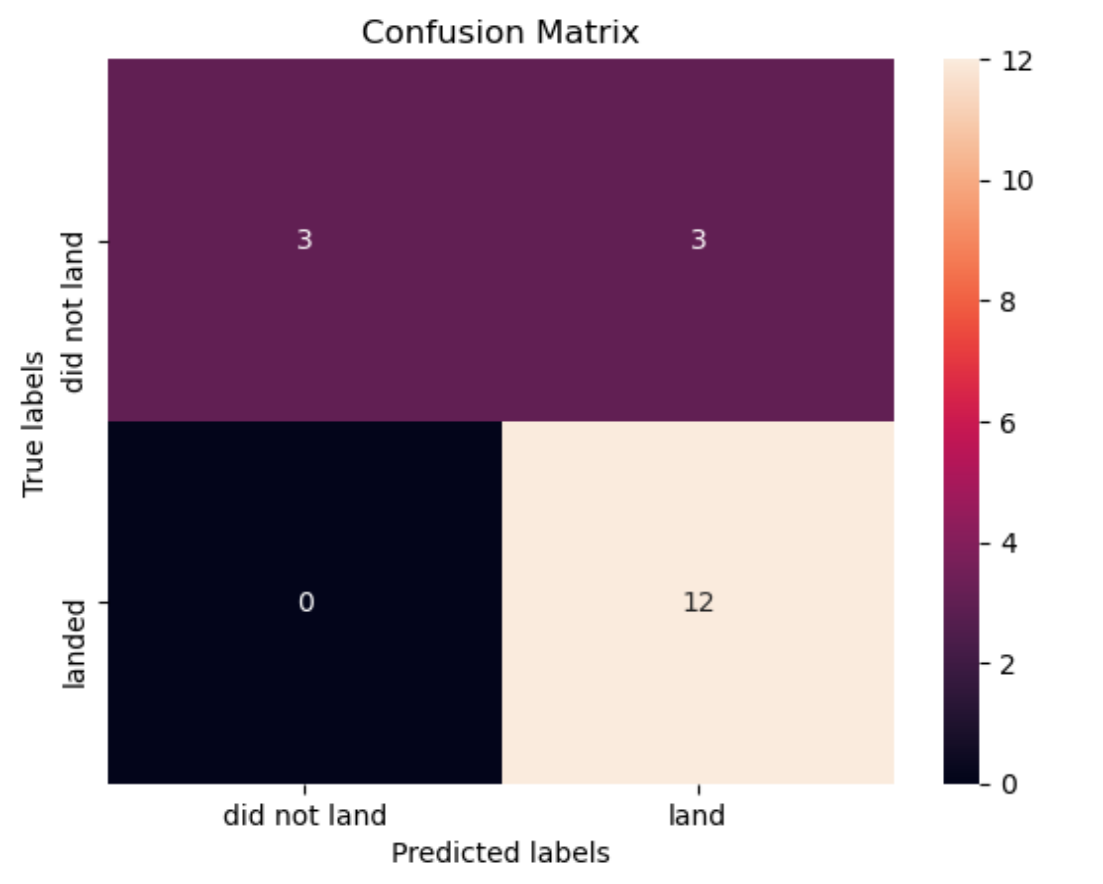


Confusion Matrix

- Shown here is the confusion matrix for the Logistic Regression model.
- Confusion matrices can be read as:

True Negative	**False Positive**
False Negative	**True Positive**

- Prediction Breakdown:
  - 12 True Positives and 3 True Negatives
  - 3 False Positives and 0 False Negatives



Conclusions

- SpaceX’s record for Falcon 9 first stage landing outcomes has improved.
- The trend is toward better performance and 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

- 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
- SpaceX (CSV): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDS0321ENSkillsNetwork26802033-2022-01-01
- 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

Thank you!