On behalf of Space

Exploratory Data Analysis of Falcon 9 Rocket

Brent Allard
January 30, 2022

Table of Contents

Executive Summary

Summary of methodologies

- Data Collection
- Data Wrangling
- EDA with SQL
- EDA with Data Visualization
- Interactive Map with Folium
- Dashboard with Plotly Dash
- Predictive Analysis

Summary of all results

- Exploratory Data Analysis results
- Interactive Analytics Demo (Plotly Dash)
- Predictive Analysis Results

Introduction

Project Background

To compete with SpaceX on cost and to deliver reliable and reusable rocket payloads.

By using the **Falcon 9 Rocket** as the basis for our analysis. We can determine the cost of each launch.

Trained machine learning models with public information from SpaceX, will accurately predict if the first stage will be reused.

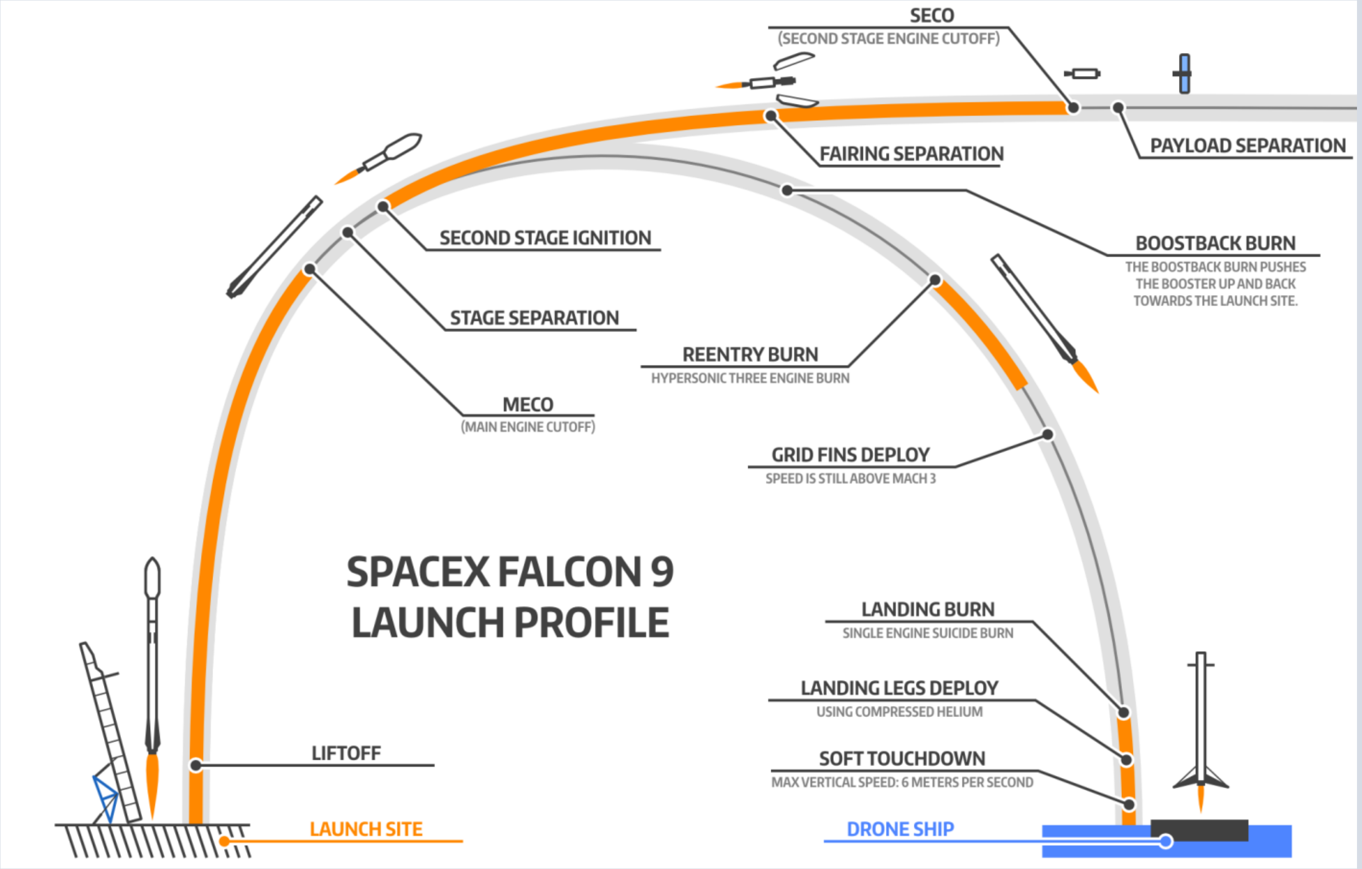
Project Constraints

Determine key factors in predicting first stage recovery.

- Launch location
- Payload
- Orbital altitude

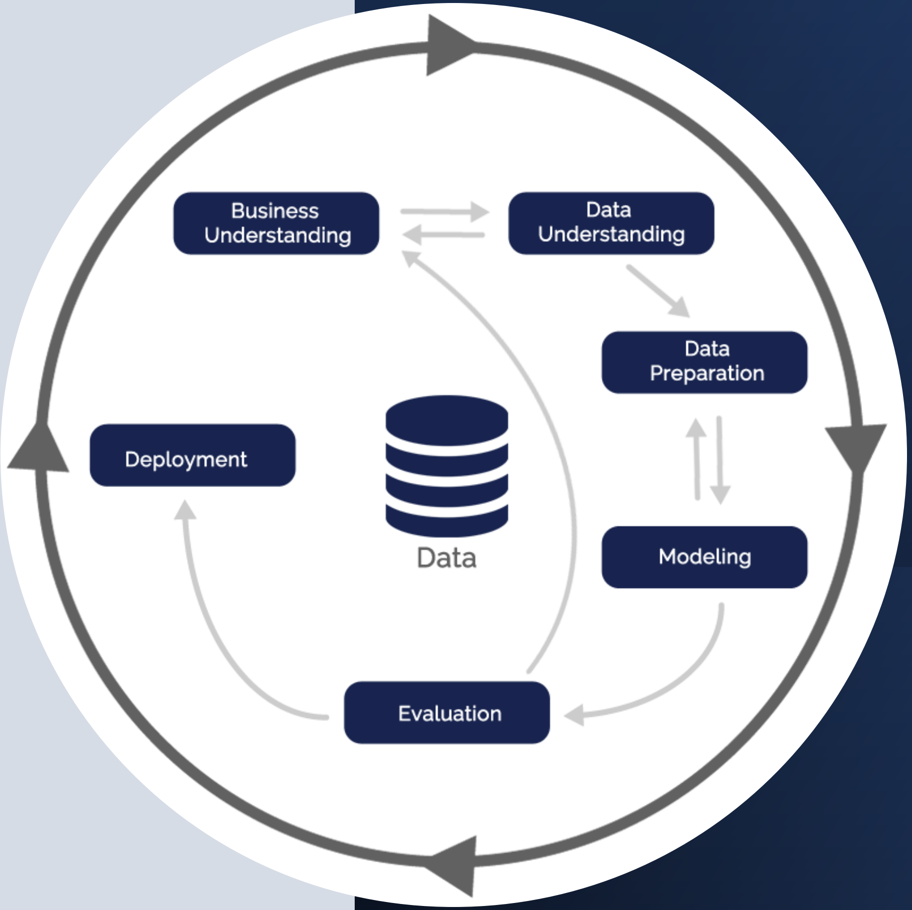
Introduction

Project Concept



Section One

Methodology



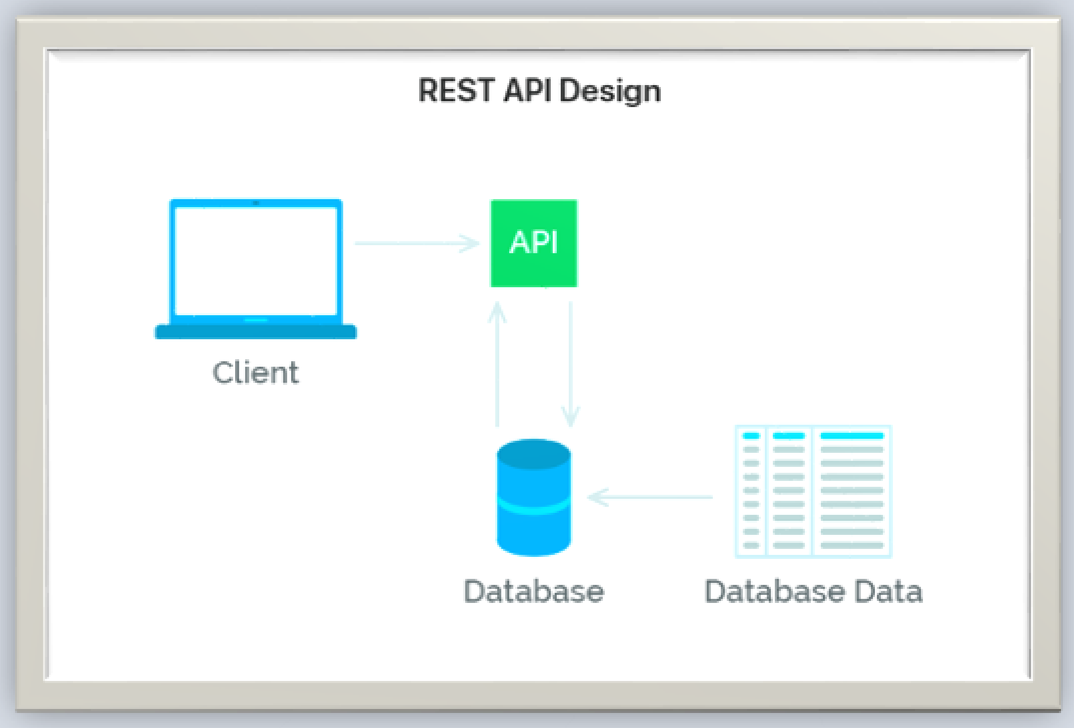
Data Requirement	Resource
Data Collection	• SpaceX API (REST API) • Web Scraping (Wikipedia)
Data Wrangling	• IBM Watson Studio Notebook • Python Libraries: • Requests, Beautiful Soup, NumPy, Pandas, Matplotlib, Plotly
EDA with SQL	• IBM db2 Database • SQL queries to assess feature selection for ML Model

Data Collection – Space X API

SpaceX launch data is gathered from an API, specifically the SpaceX REST API.

Our goal is to use this data to predict whether SpaceX will attempt to land a rocket or not.

This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.



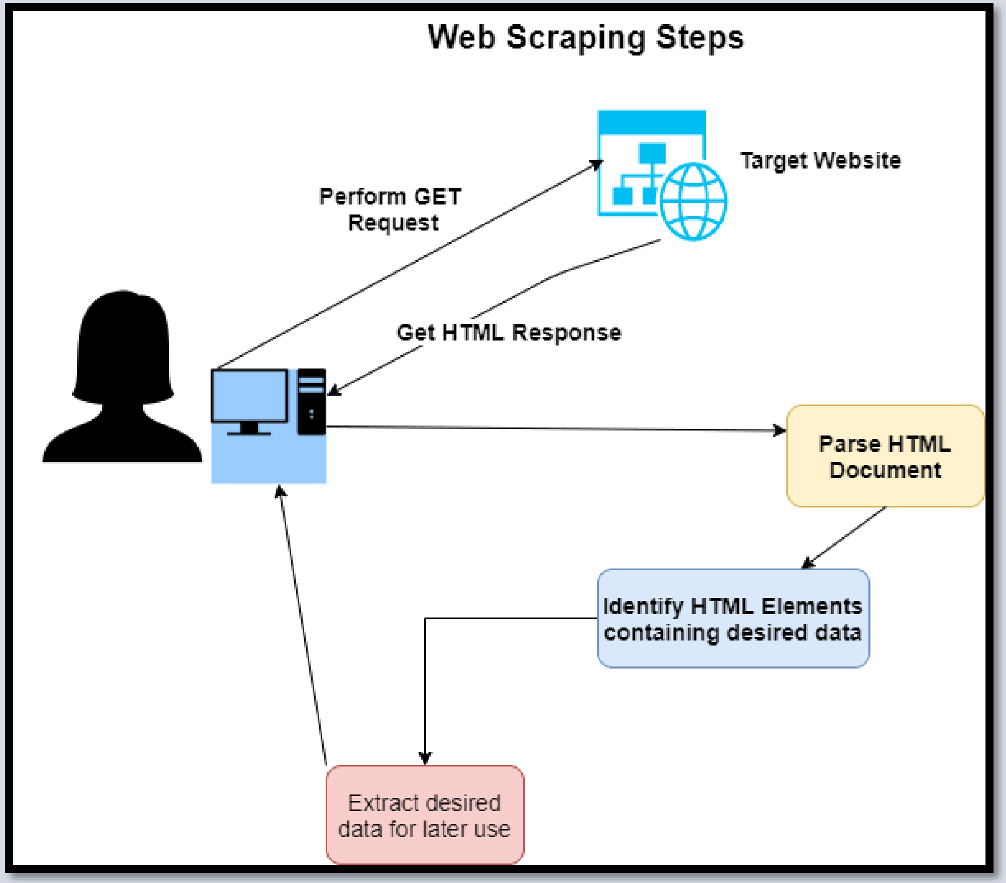
API (REST) process flow diagram

Data Collection – Web Scraping

Using the Python BeautifulSoup package to web scrape some HTML tables that contain valuable Falcon 9 launch records.

We will then parse the data from those tables and convert them into a Pandas data frame for further visualization and analysis

Our goal is to use this data to select the best feature labels to train our supervised machine learning model.



Web scraping process diagram

Data Wrangling

To utilize the data in our supervised machine learning model. We need to perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what labels we should use for training our supervised model.

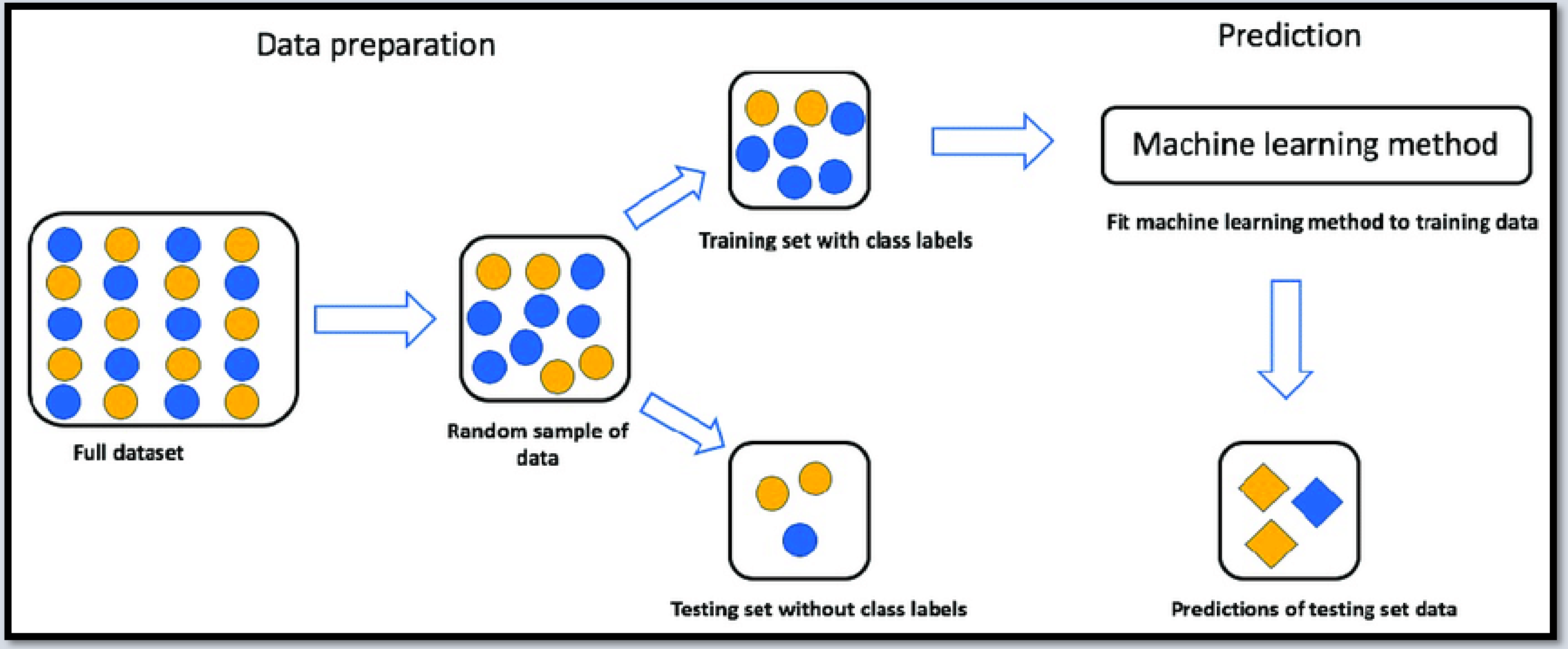
Organize and put data into consistent formats. Assign quantitative and qualitative data types.

Remove data errors, duplicate data and formatting discrepancies for data consistency.

Explore relationship between data to find patterns to be used in feature selection

Data Wrangling

After analyzing and selecting the best features we split the data into two sets to train and test our supervised machine learning model to make prediction on the outcome of a successful rocket landing.



https://github.com/brent-allard/IBM_DATASCIENCE_CAPSTONE/blob/master/Web%20Scraping%20and%20Data%20Collection.ipynb

EDA with Data Visualization Summary

Scatter Plot

Scatter Plots provide a better understanding of the correlation between two variables

- Flight Number vs Launch Site
- Flight Number and Orbit type
- Payload and Orbit type

Bar Graph

A bar diagram makes it easy to compare sets of data between different groups.

- Success Rate by Launch Site

Line Graph

Line graphs provide a good overview of the trends in the data points

- Success Rate by Year

EDA with SQL

Overview of steps

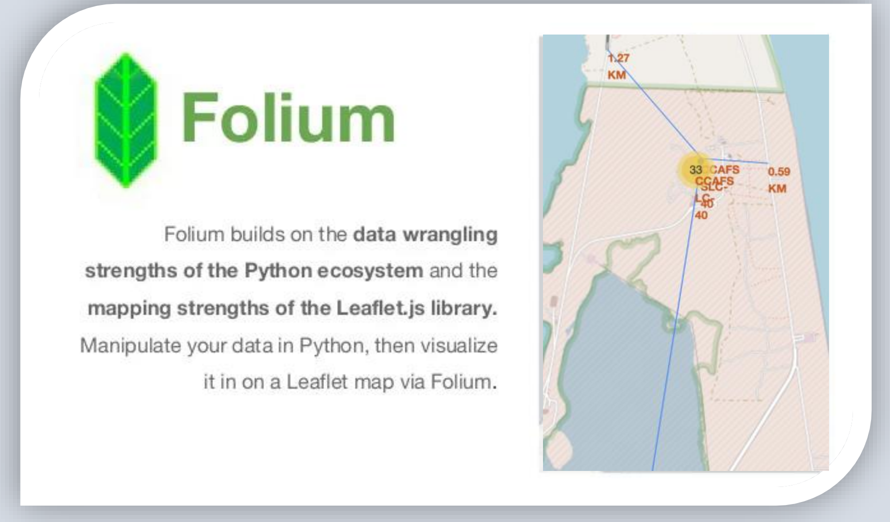
1. Load the dataset into the corresponding table in a Db2 database
2. Standardize date types to facilitate import.
3. Execute SQL queries to explore and understand data features.

SQL Query summary

- *Names of the unique launch sites in the space mission*
- *5 records where launch sites begin with the string 'CCA'*
- *Total payload mass carried by boosters launched by NASA (CRS)*
- *Average payload mass carried by booster version F9 v1.1*
- *List the 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*
- *List the total number of successful and failure mission outcomes*
- *List the names of the booster_versions which have carried the maximum payload mass*
- *List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015*
- *Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20,*

https://github.com/brent-allard/IBM_DATASCIENCE_CAPSTONE/blob/master/EDA%20with%20SQL.ipynb

Interactive Map
with Folium



<u>Summary of steps</u>

- Mark all launch sites on a map by using the latitude and longitude coordinates.
- Mark the success/failed launches for each site with Green and Red on the map with a circle marker and cluster object for ease of identification.
- Calculate the distances between a launch site to its proximities and drawn lines to indicate distance to launch site.

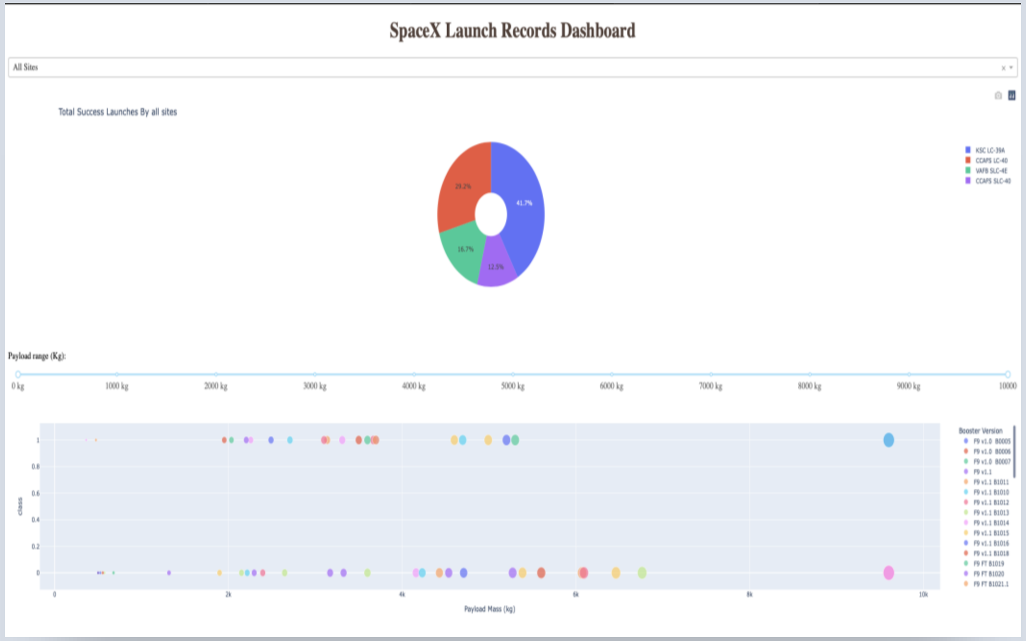
The interactive map provides insight into the features of the dataset in color to easily identify the relationships between the launch site success and distances to various landmarks.

Dashboard

A Plotly Dash application is provided to perform interactive visual analytics on SpaceX launch data in real-time.

This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart

> The Pie chart helps to determine the distribution of success launches by site.
>
> The Scatter Plot provides insight into how the payload may be correlated with mission outcomes for selected site(s).



plotly | Dash

Predictive Analysis (Classification)

Objectives

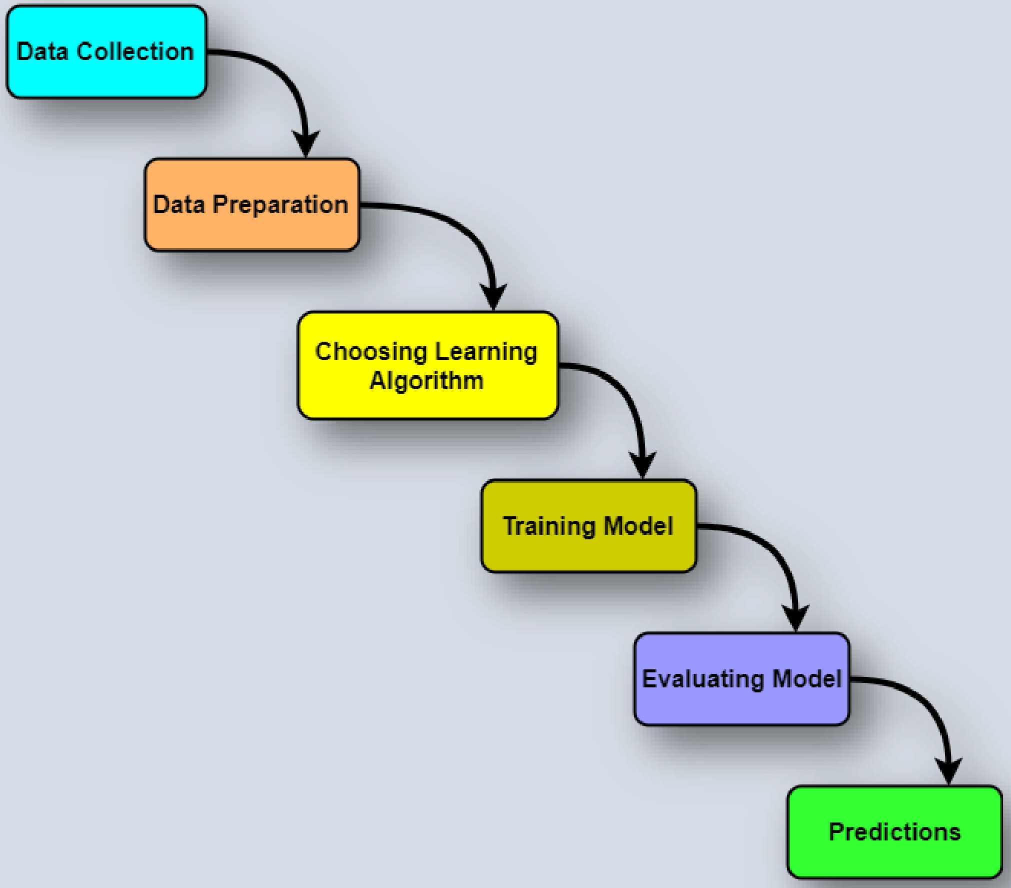
- Identify best machine learning algorithm to provide predictive analysis results.

Summary

- Create a column for the class (launch outcome)
- Normalize the data
- Split into training data and test data
- Determined best parameters using the data attribute `best_params\_` and the accuracy on the validation data using the data attribute `best_score\_`.

Best Hyperparameter were determined for the following Algorithm:

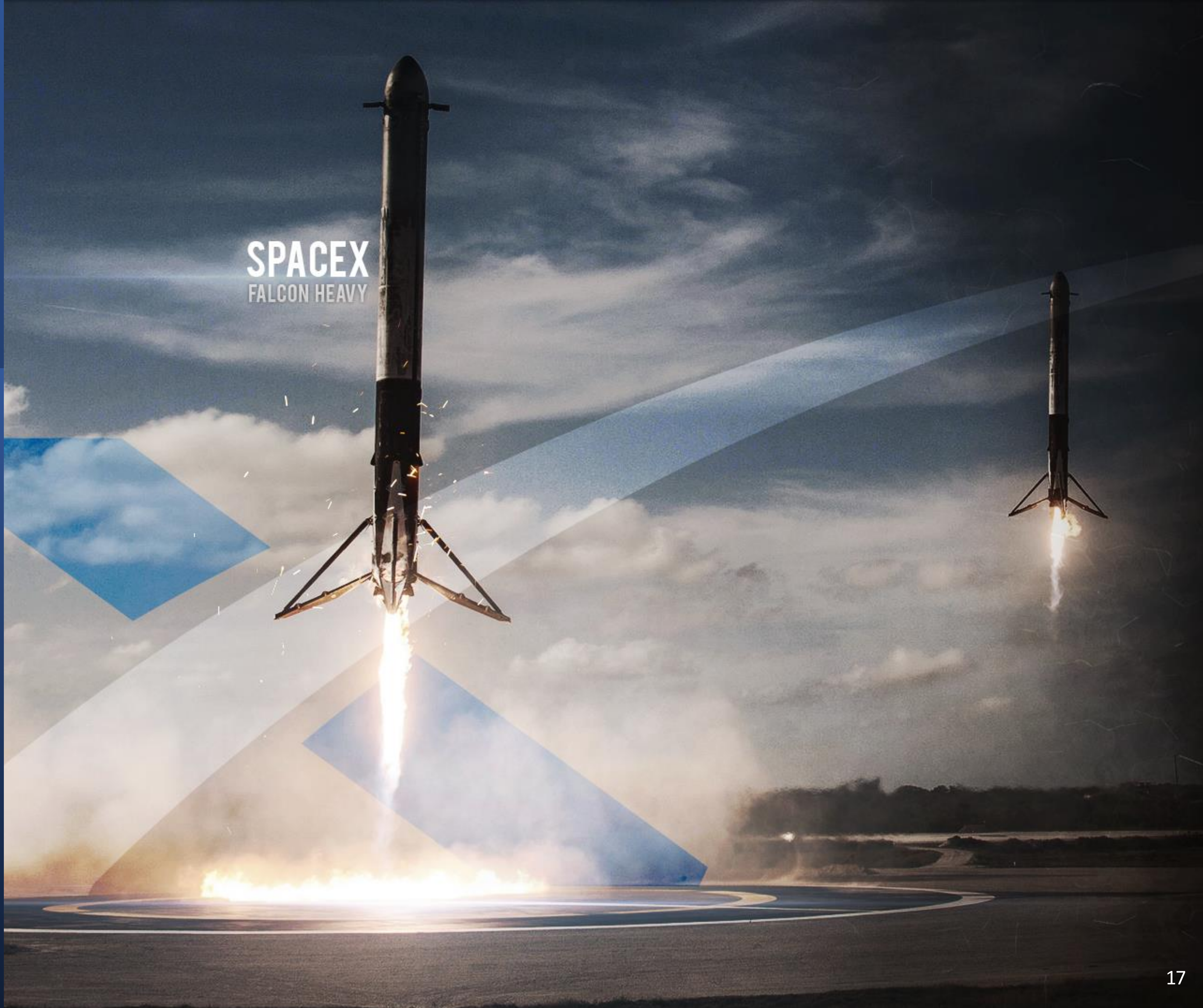
- Support Vector Machine
- Classification Trees
- Logistic Regression
- K Nearest Neighbor



Machine Learning Workflow

Results

Section two



Results - Content

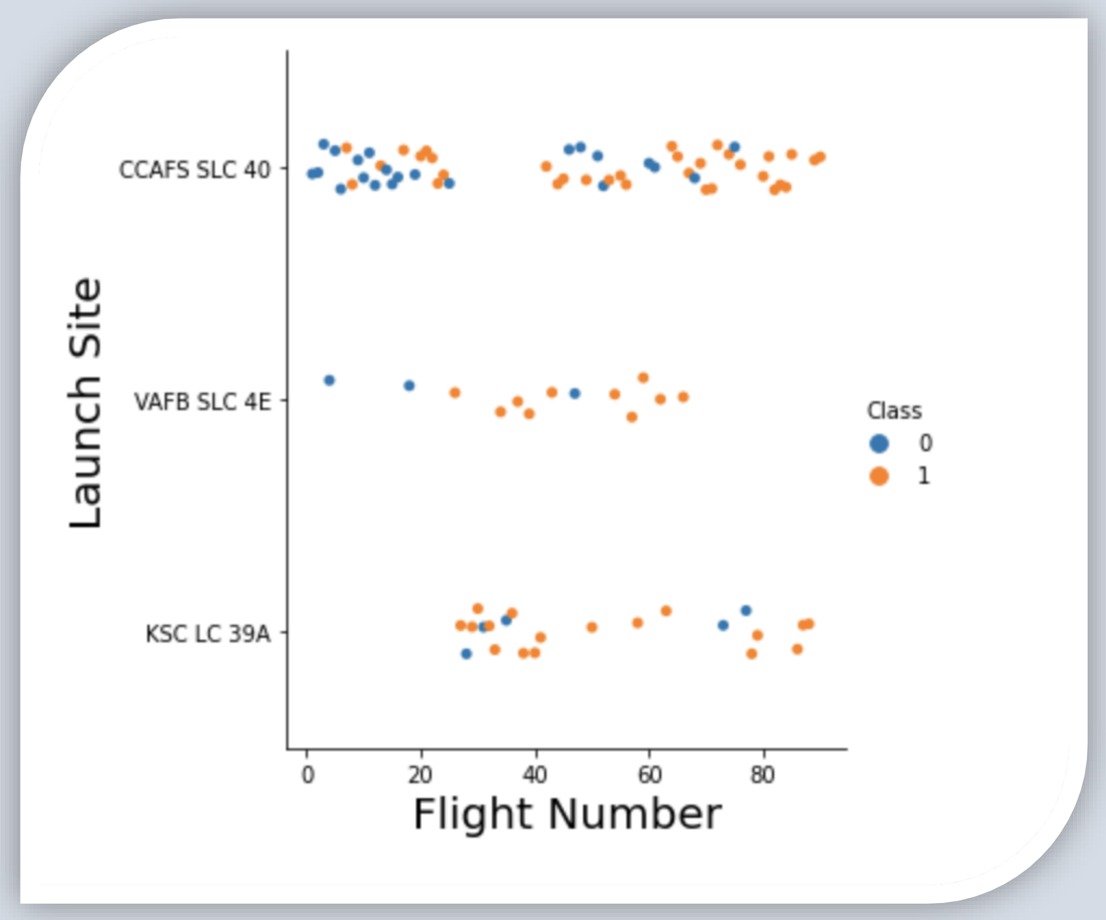
EXPLORATORY DATA ANALYSIS RESULTS

INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS

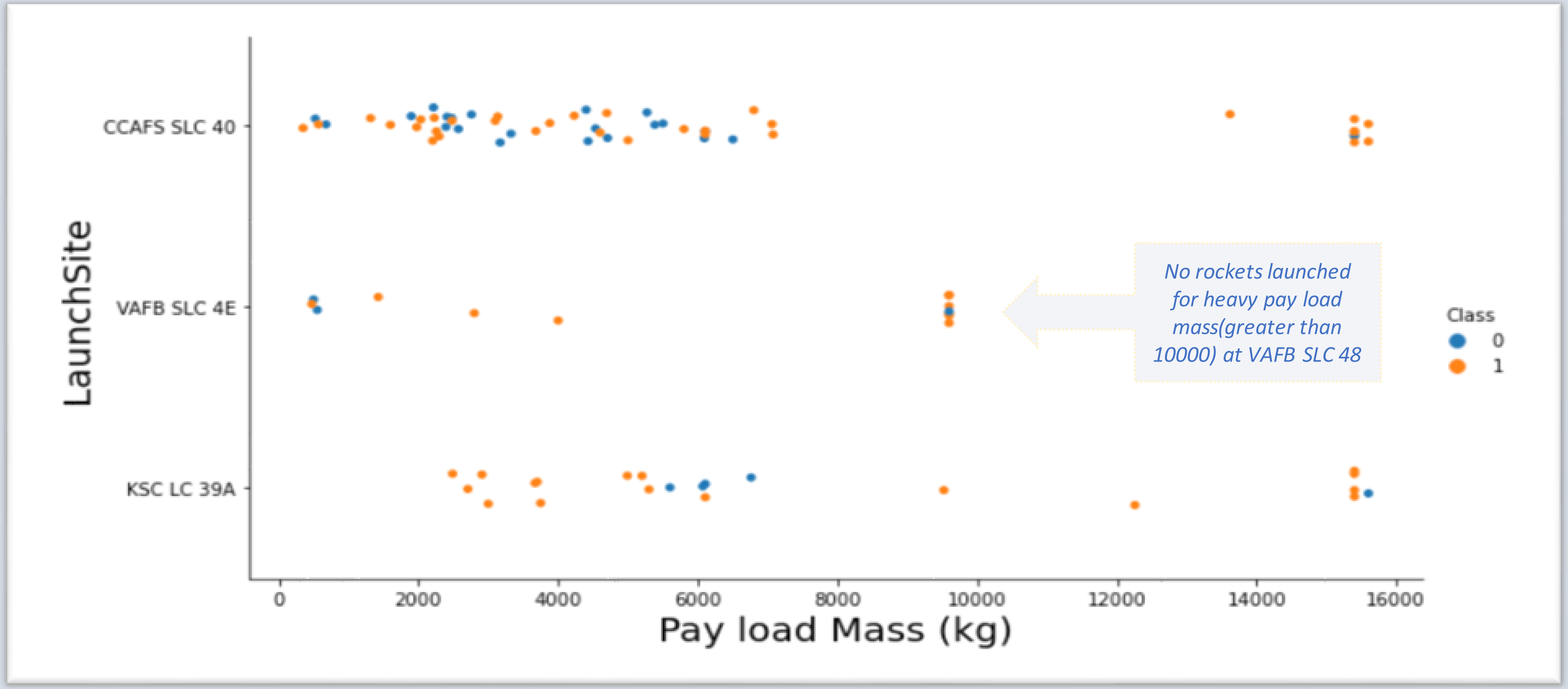
PREDICTIVE ANALYSIS RESULTS

Flight Number vs. Launch Site

- CCAFS LC-40, has a success rate of 60 %, while VAFB SLC 4E and KSC LC-39A has a success rate of 77%.
- Increasing success rate for all 3 launch sites based on number of flights attempted.

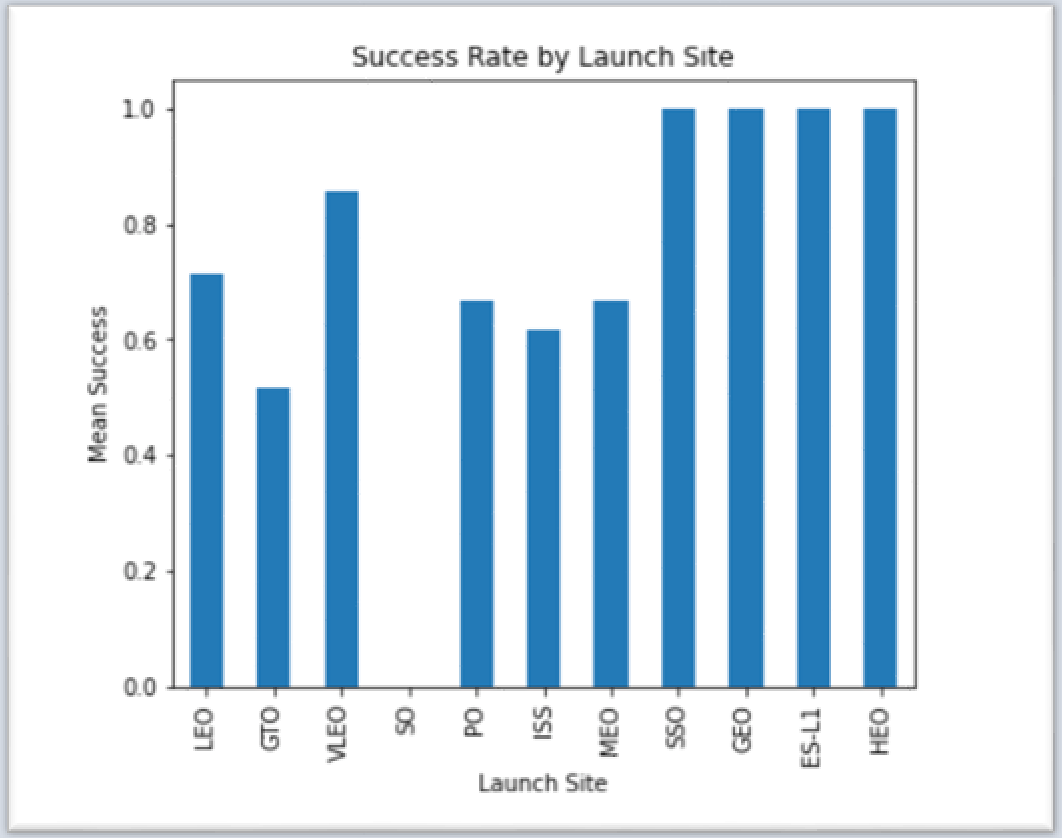


Payload vs. Launch Site



- Significantly more data is available for predicting the success rate below a payload of 10000kg.

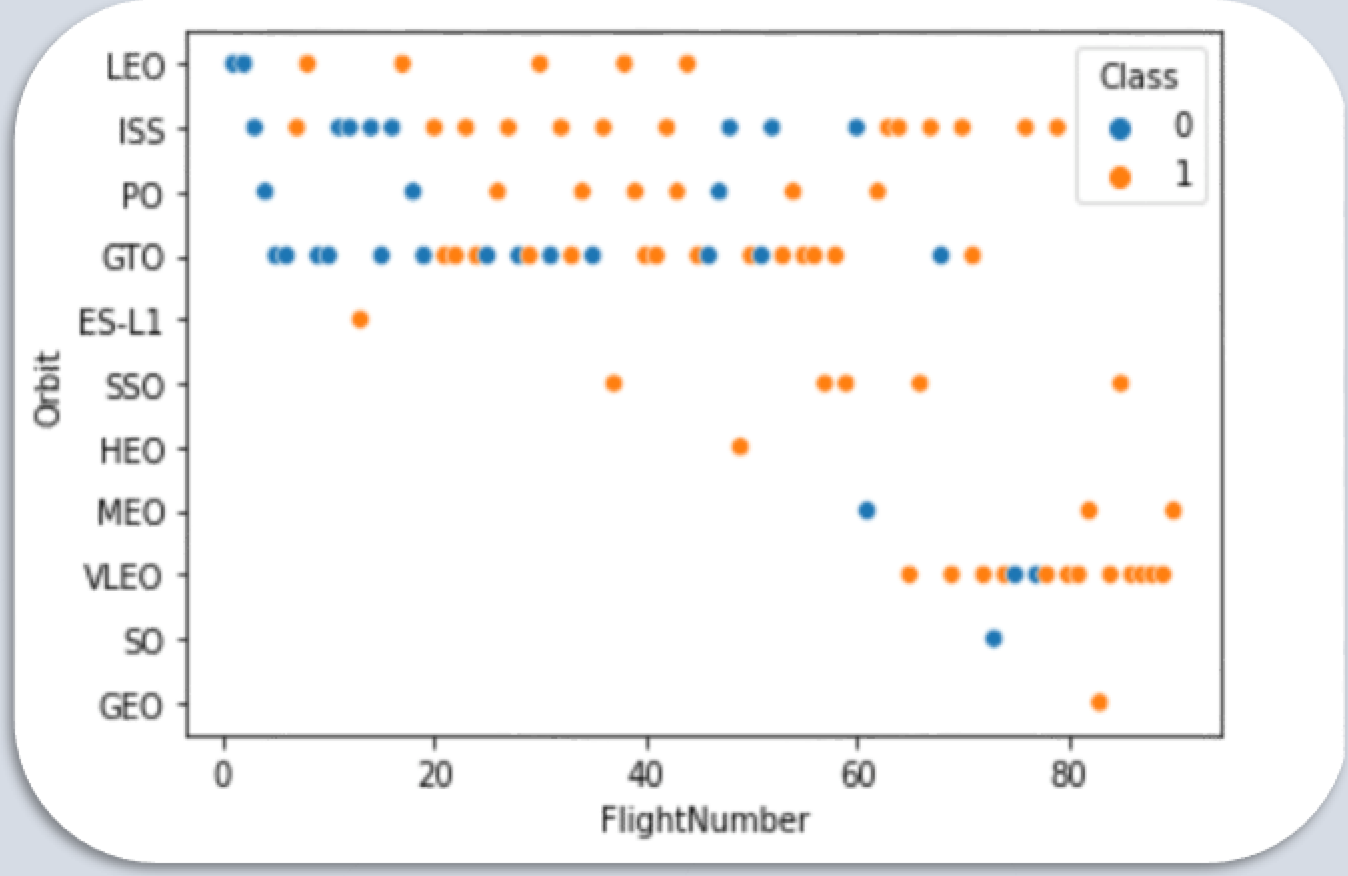
Success Rate vs. Orbit Type



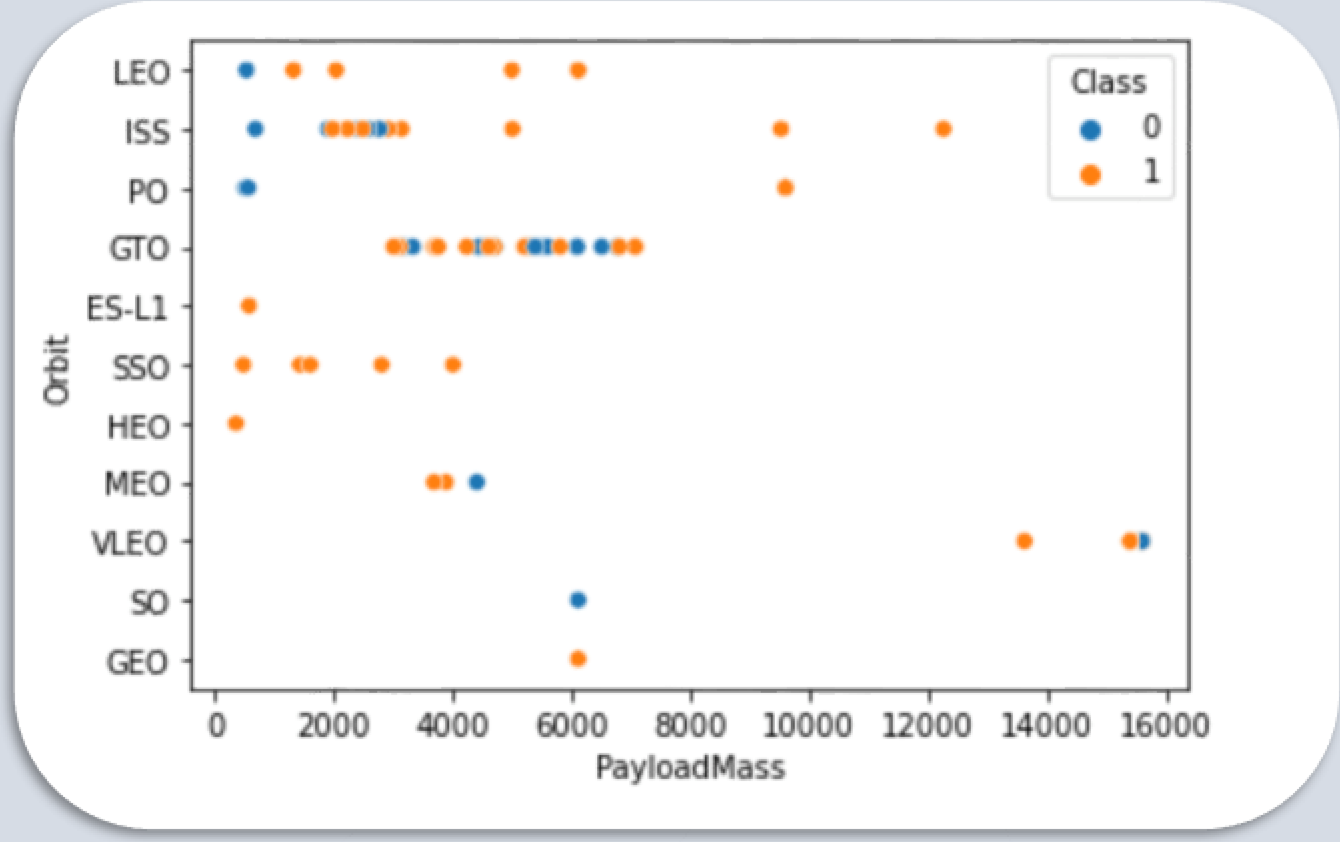
- SSO, GEO, ES-L1, HEO have a mean success rate of 100%.
- VLEO mean success rate of 85%

Flight Number vs. Orbit Type

- The LEO orbit success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

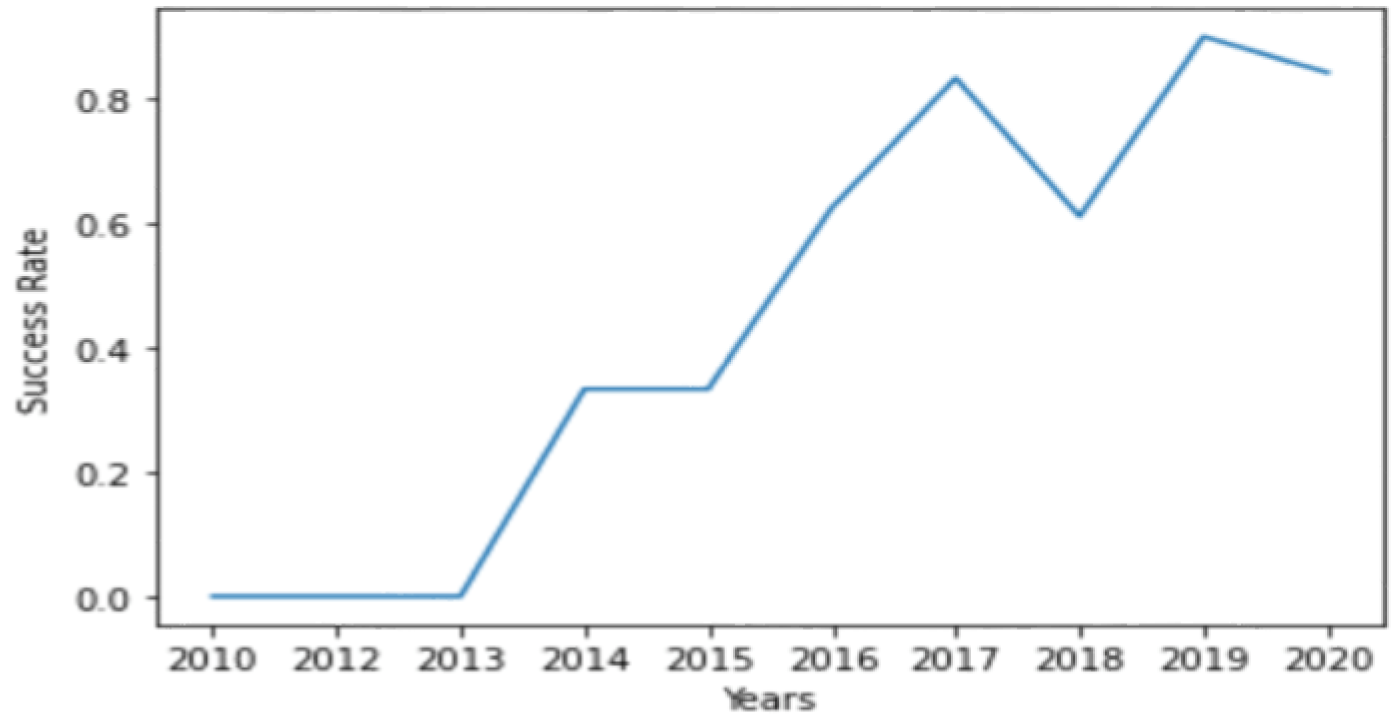


Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However, for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are not delineated.

Launch Success Yearly Trend



The sucess rate since 2013 has increased till 2017. With a small decrease in success followed by an increase right up until 2020

EDA with SQL (Results)

SQL QUERIES

select unique(launch_site) from SPACEXTBL

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

select * from SPACEXTBL
where launch_site like 'CCA%'
limit 5

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

SQL QUERIES cont'd

```
select sum(payload_mass__kg_) as
"TOTAL_PAYLOAD_MASS (kg)" from SPACEXTBL
where customer = 'NASA (CRS)'
```

TOTAL_PAYLOAD_MASS (kg)
45596

```
select avg(payload_mass__kg_) as
"AVERAGE_PAYLOAD_MASS (kg)" from SPACEXTBL
where booster_version = 'F9 v1.1'
```

AVERAGE_PAYLOAD_MASS (kg)
2928

```
select min(DATE) as
"FIRST_SUCCESSFUL_LANDING" from SPACEXTBL
where landing__outcome = 'Success (ground pad)'
```

first_successful_landing
2015-12-22

```
select unique(booster_version) from SPACEXTBL
where landing__outcome = 'Success (drone ship)'
and payload_mass__kg_ >4000 and
payload_mass__kg_ <6000
```

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

SQL QUERIES cont'd

```
select mission_outcome, count(mission_outcome) as "TOTAL MISSION OUTCOMES"
from SPACEXTBL where mission_outcome in ('Success','Failure (in flight)') group by
mission_outcome order by mission_outcome desc
```

mission_outcome	TOTAL MISSION OUTCOMES
Success	99
Failure (in flight)	1

```
select booster_version from SPACEXTBL
where payload_mass__kg_ = (select
max(payload_mass__kg_) from
SPACEXTBL)
```

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

SQL QUERIES cont'd

```
select DATE ,booster_version , launch_site , landing__outcome
from SPACEXTBL where landing__outcome = 'Failure (drone
ship)' and DATE >= '2015-01-01'  and DATE <= '2015-12-31'
```

DATE	booster_version	launch_site	landing__outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

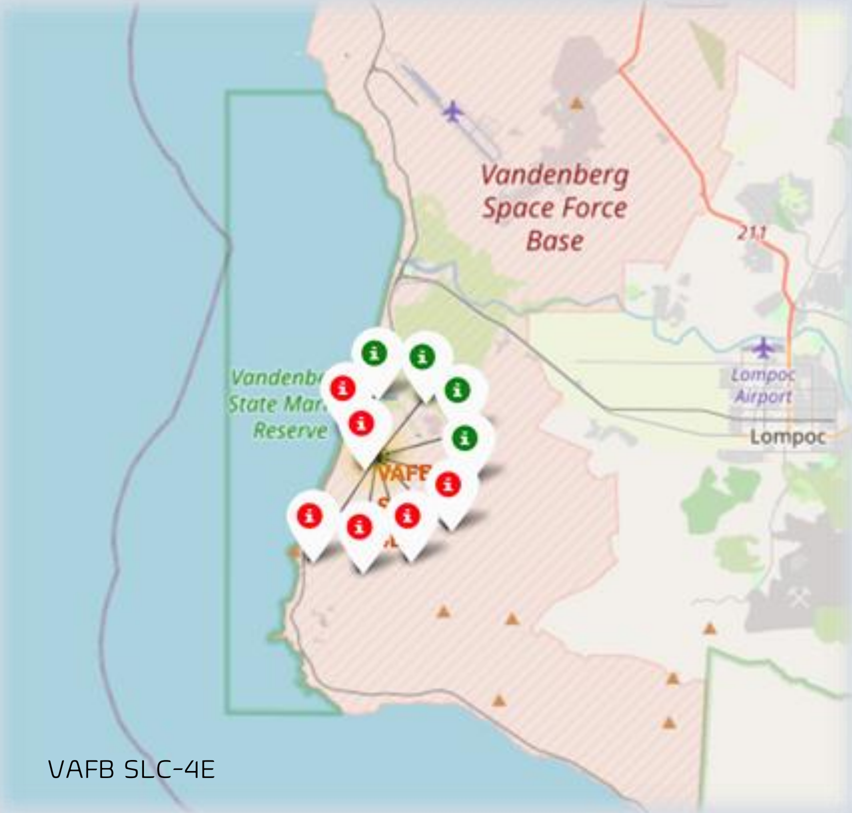
```
select landing__outcome,
count(*)as "TOTAL_OUTCOMES",
rank() over (order by count(*)
desc) as "LANDING_RANK" from
SPACEXTBL where DATE >= '2010-
06-04' and DATE <= '2017-03-20'
group by landing__outcome order
by LANDING_RANK
```

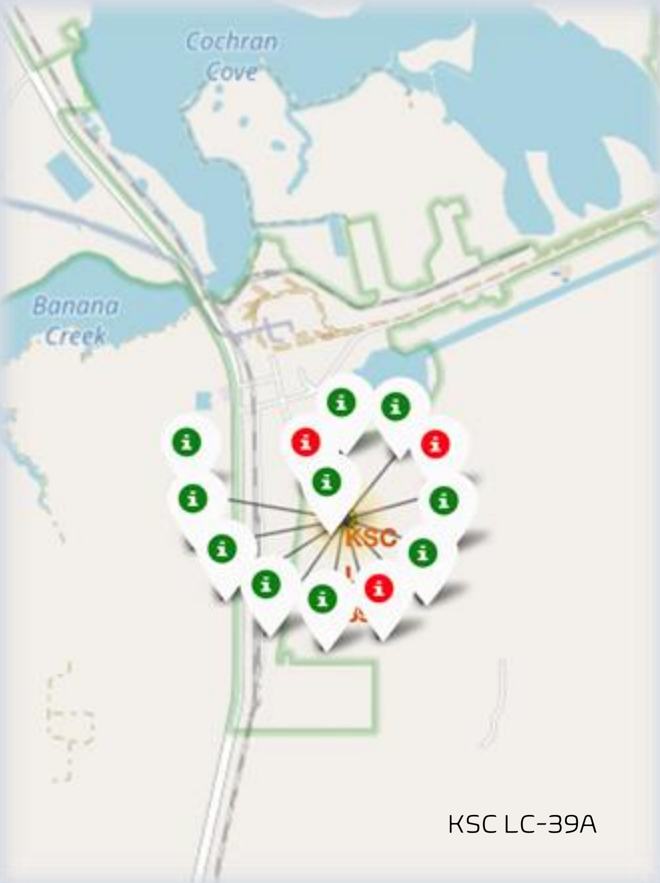
landing__outcome	total_outcomes	landing_rank
No attempt	10	1
Failure (drone ship)	5	2
Success (drone ship)	5	2
Controlled (ocean)	3	4
Success (ground pad)	3	4
Failure (parachute)	2	6
Uncontrolled (ocean)	2	6
Precluded (drone ship)	1	8

4 Launch sites located on the Pacific and Altantic coasts of United States.

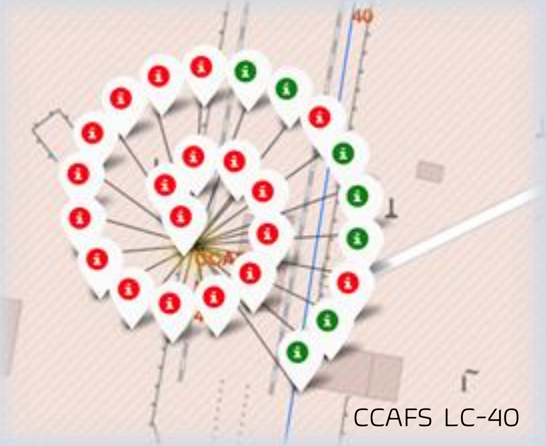
Space X Global Launch Site locations - Folium

California





Florida

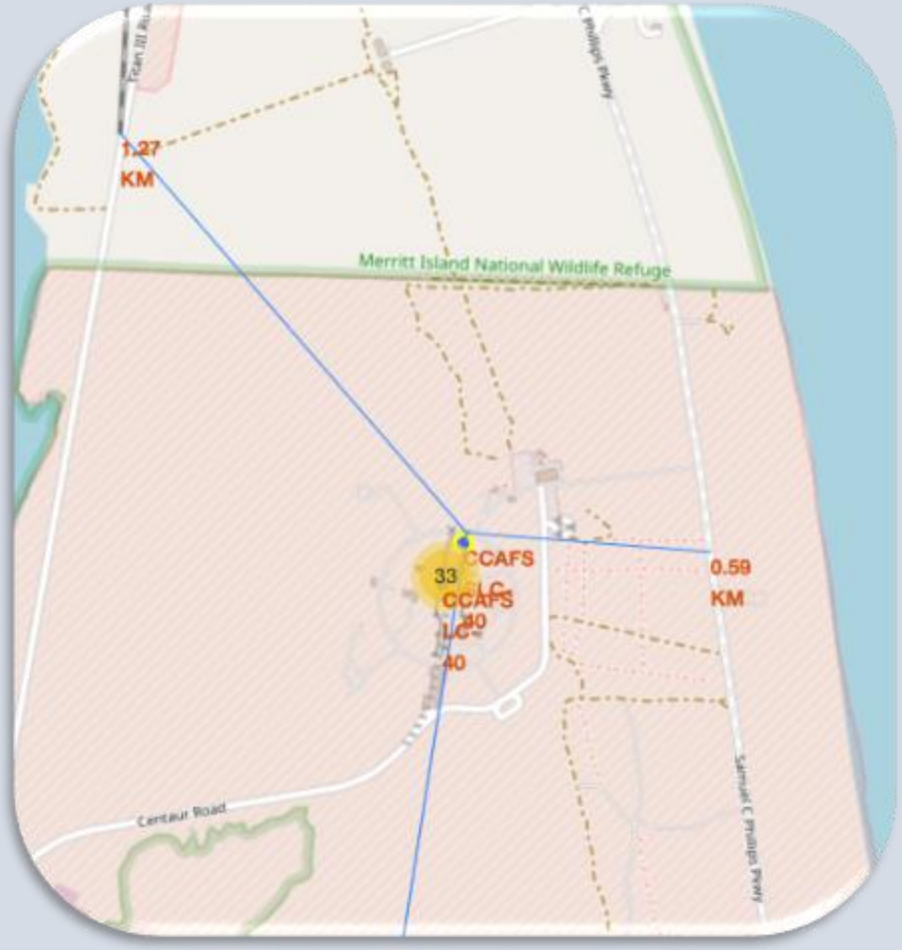


Space X Launch Site Outcomes - Folium

Green Markers indicate successful outcome
Red Markes indicate unsuccessful outcome

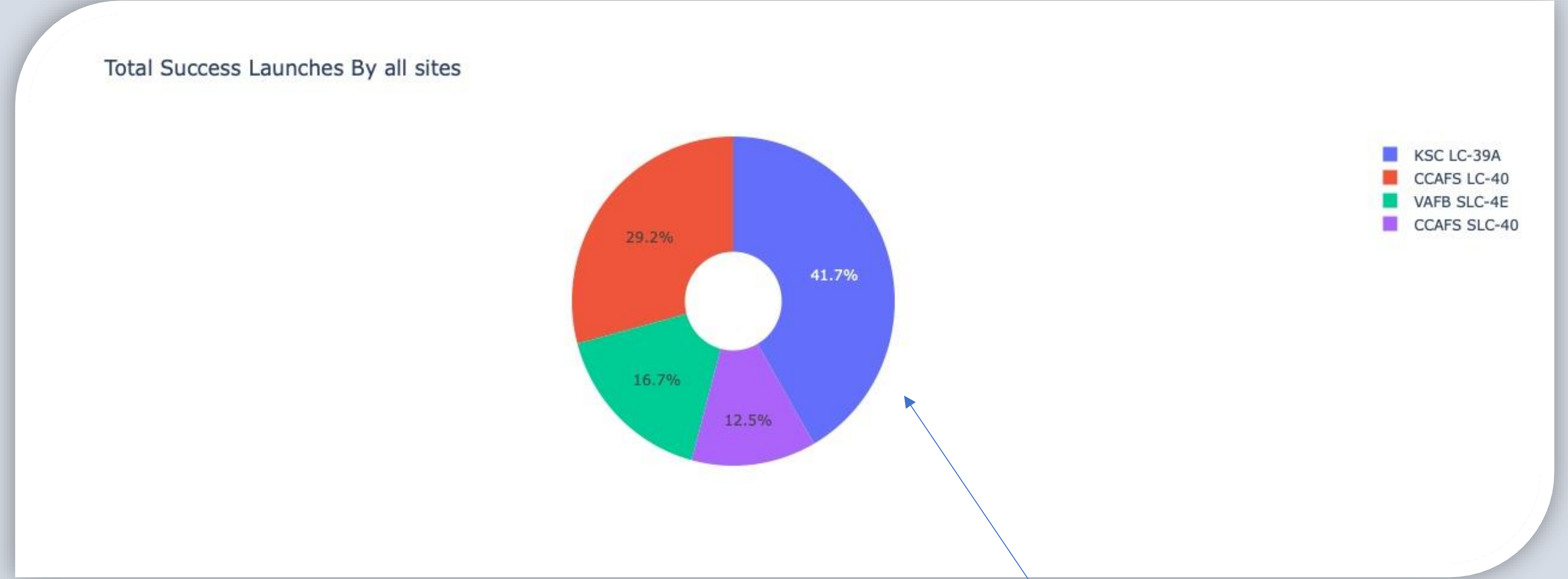
Calculated distance from CCAFS SLC-40 to nearest:

- Railroad = 1.27 km
- Highway = 0.59 km
- City = 18.21 km



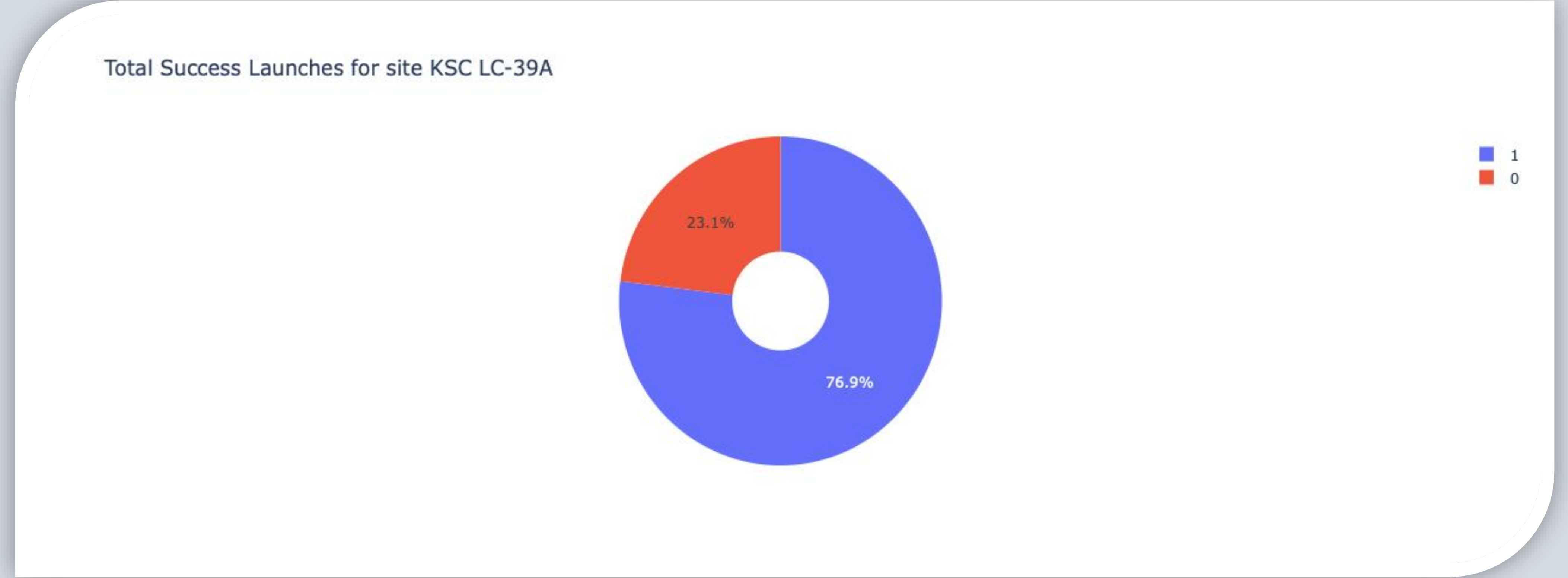
Space X Global Launch Site Proximity - Folium

DASHBOARD



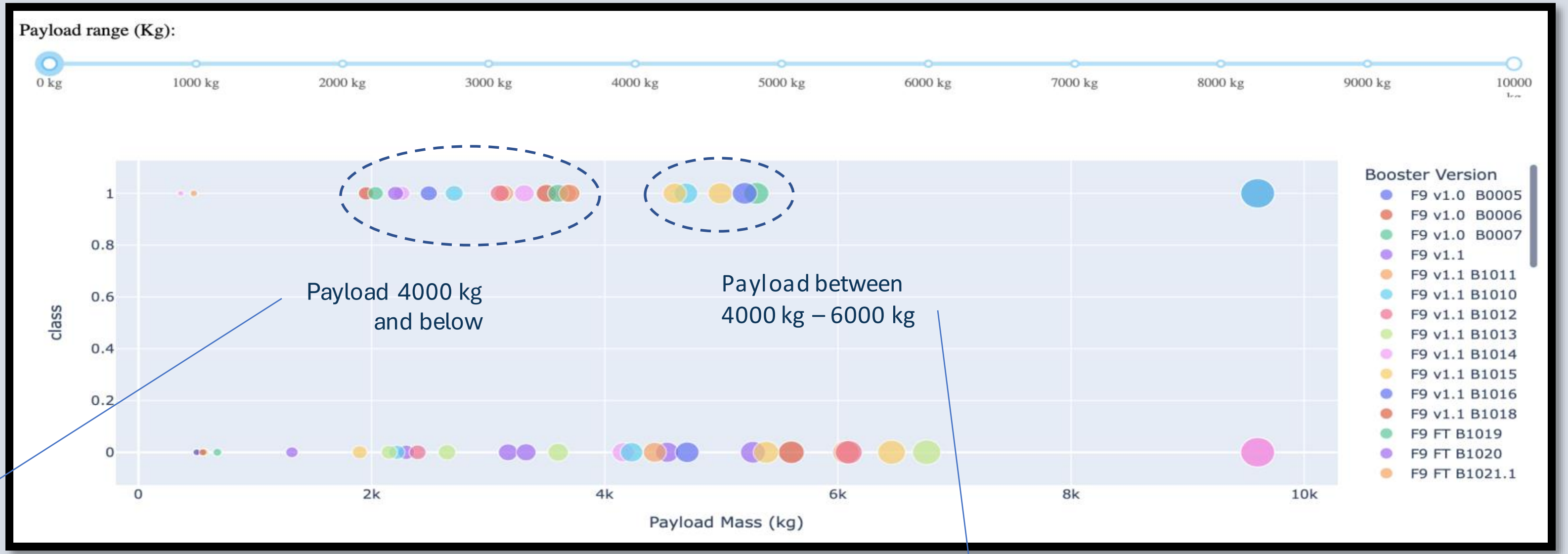
KSC LC-39 accounts for the largest percentage of launch success of the 4 sites.

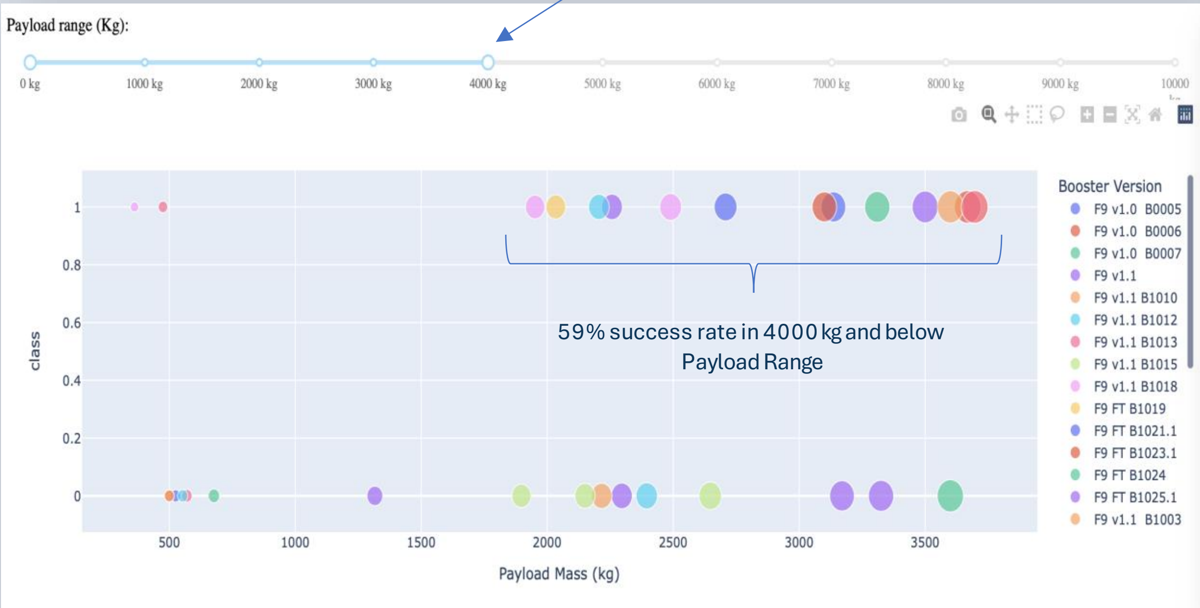
DASHBOARD



KSC LC-39A site did produced a success rate of over 3:1. Similar results were reported at CCAFC LC-40 (2.7:1)

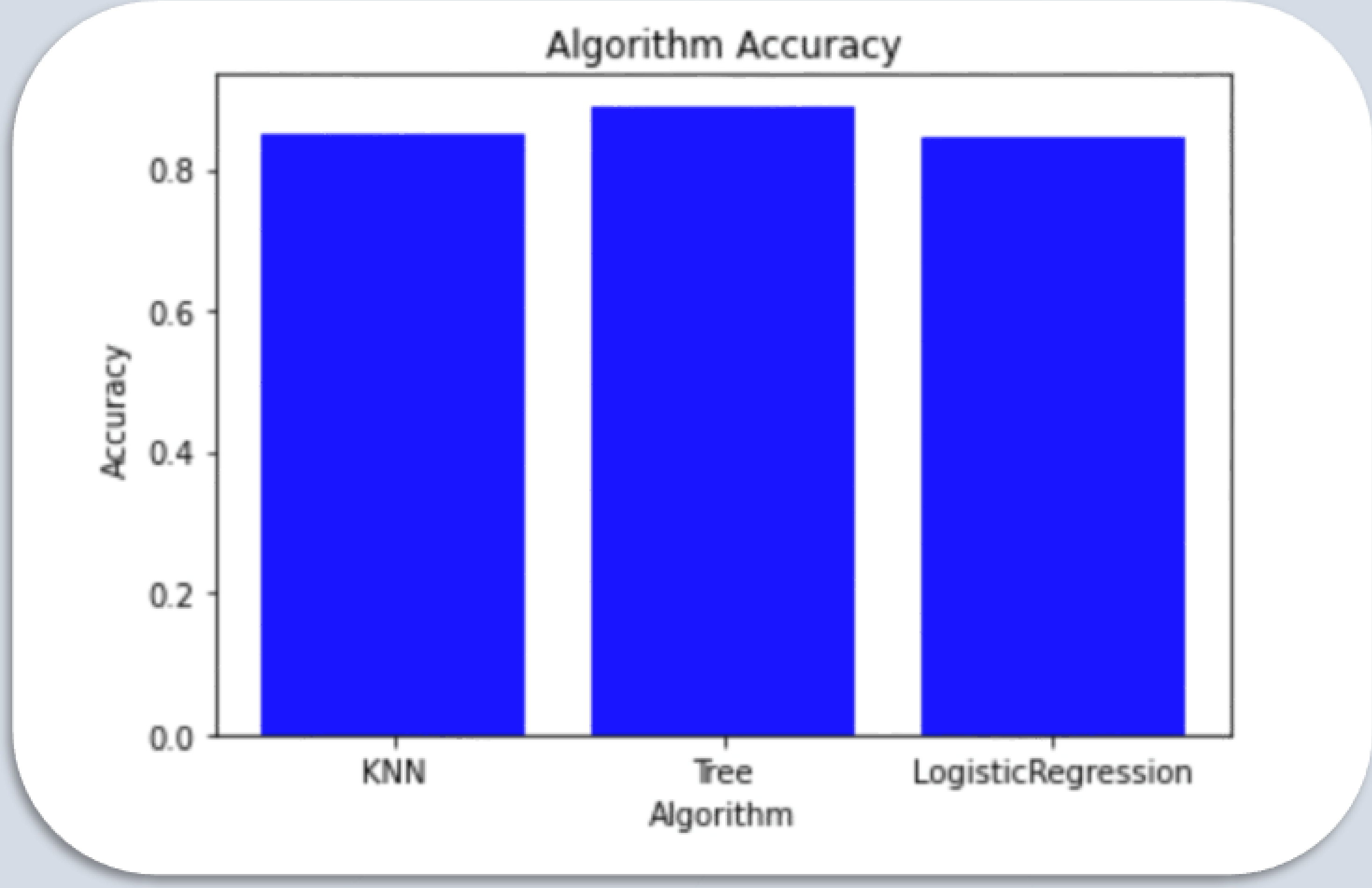
DASHBOARD


Payload range (Kg):
Payload 4000 kg
and below
Payload between
4000 kg – 6000 kg
class
Payload Mass (kg)
Booster Version


Payload range (Kg):
59% success rate in 4000 kg and below
Payload Range
class
Payload Mass (kg)
Booster Version

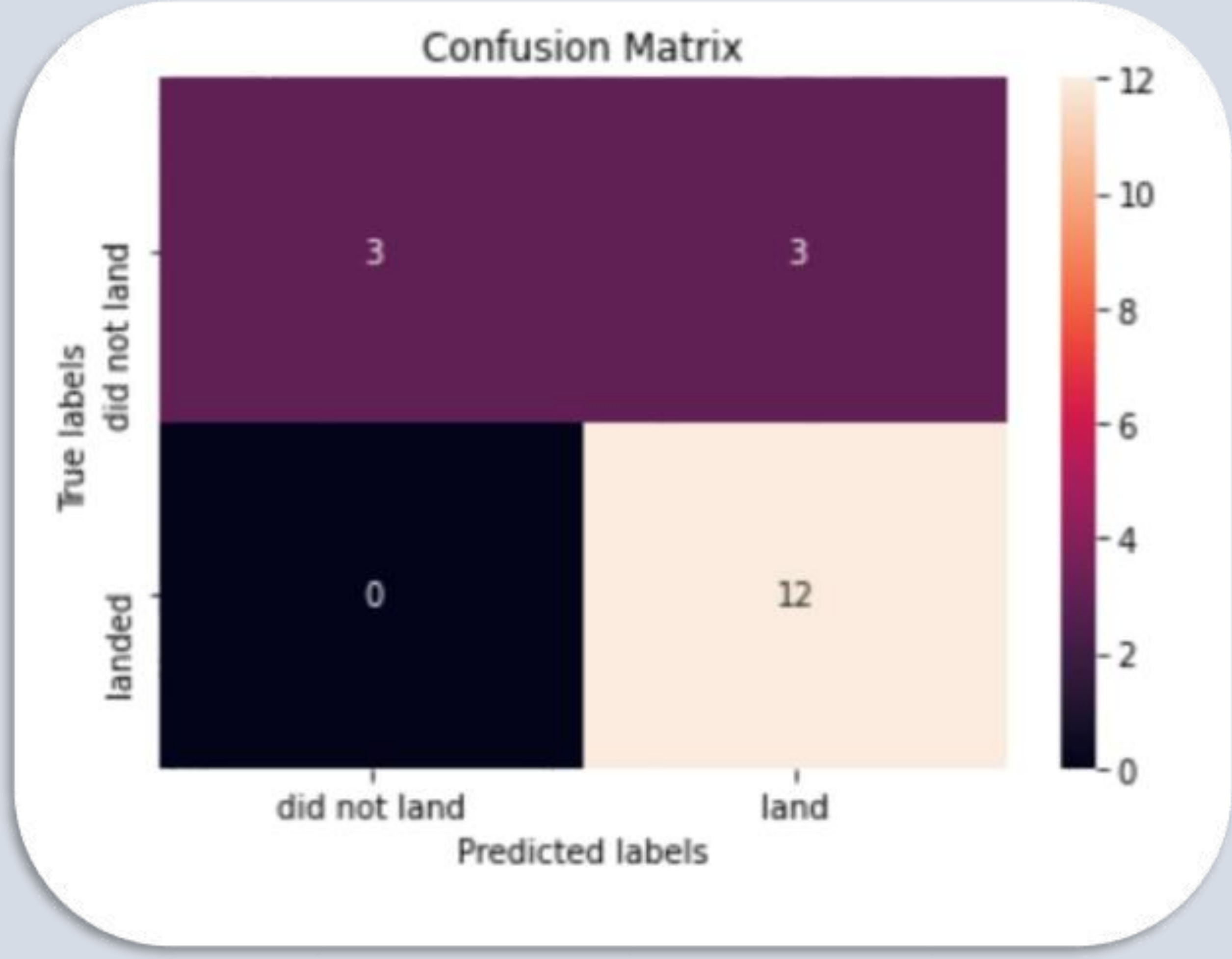

Payload range (Kg):
Version FT B1021.2, FT B1031.2, B4 B1040.1, FT B1022, FT B1026 were
successful in this Payload Range.
class
Payload Mass (kg)
Booster Version

Classification Accuracy

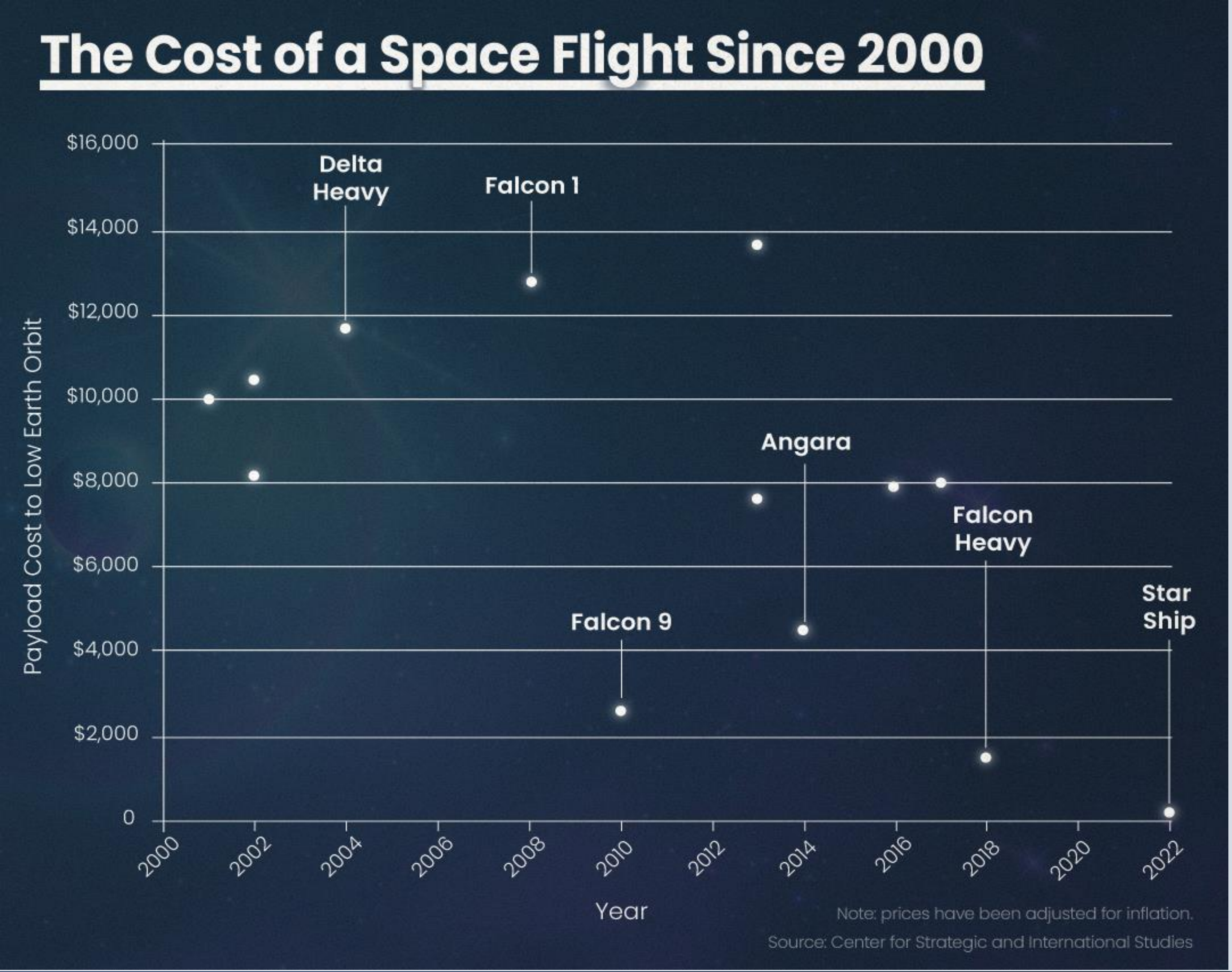


The most accurate model for prediction was the Tree Classifier with a score of 0.8892857142857145

Confusion Matrix



The Tree Algorithm was able to predict with a high degree of certainty the number of booster modules that would land successfully.



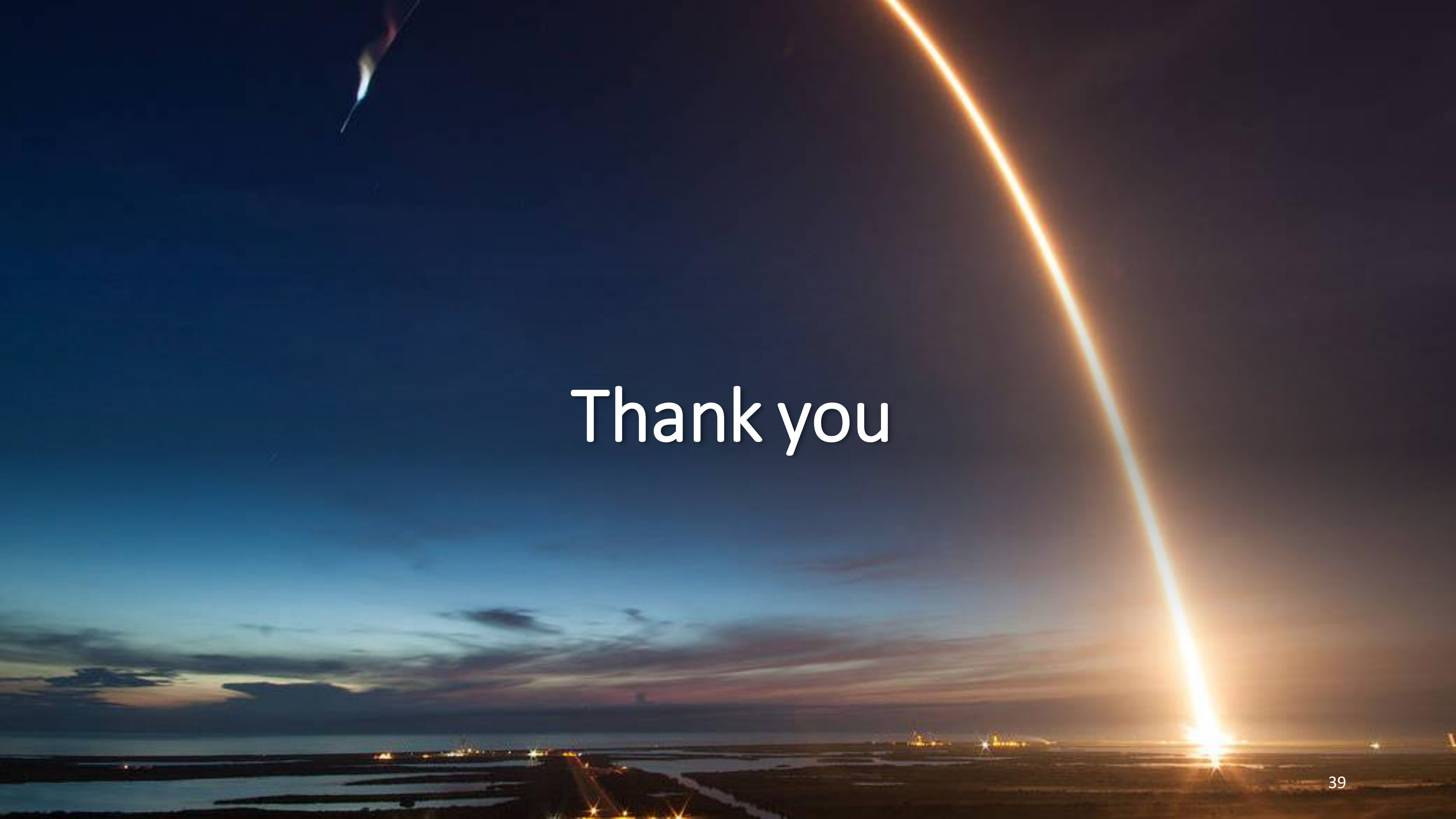
Conclusions

- Based on SpaceX data and using the Tree Classifier Model Predictor and Data Analysis,

- Best Launch location
  - KSC LC-39
- Best Payload
  - 4000 kg – 6000 kg
- Orbital altitude
  - Polar, LEO and ISS.

Appendix

https://github.com/brent-allard/IBM_DATASCIENCE_CAPSTONE/tree/master



Thank you