Predicting Falcon 9 Space First Stage Landing using Data Science

ANGELA RODRIGUEZ

25.07.2022

IBM Developer

SKILLS NETWORK

OUTLINE

EXECUTIVE SUMMARY

- SpaceX Falcon 9 rocket launches: ca 62 million dollars.
- Other providers > ca 165 million dollars.
- SpaceX can reuse the (costly) first stage.

- Train a ML model that can predict if the first stage of falcon 9 will land successfully.
    1. Data Collection - SpaceX API and Web scrapping
    2. Exploratory Data Analysis (EDA) using data visualization and SQL
    3. Interactive analytics using Plotly Dash and Folium
    4. Predictive analysis using ML models:
        *Divide dataset into training and testing sets
        *Find the best hyperparameter sets and evaluate the accuracy of the models

EXECUTIVE SUMMARY

What did we obtain?

- SpaceX's successful recoveries have the following trends:
- Light payload (in the range of 2000-4000kg)
- Launched from site KSC LC-39A
- Successfully recovered via drone ship

- Our ML model can predict the outcome of a given recovery with a reasonable accuracy of 83%.

INTRODUCTION

INTRODUCTION

- SpaceX can reuse the first stage, then giving cheaper launches.
- Although this is not always guaranteed.

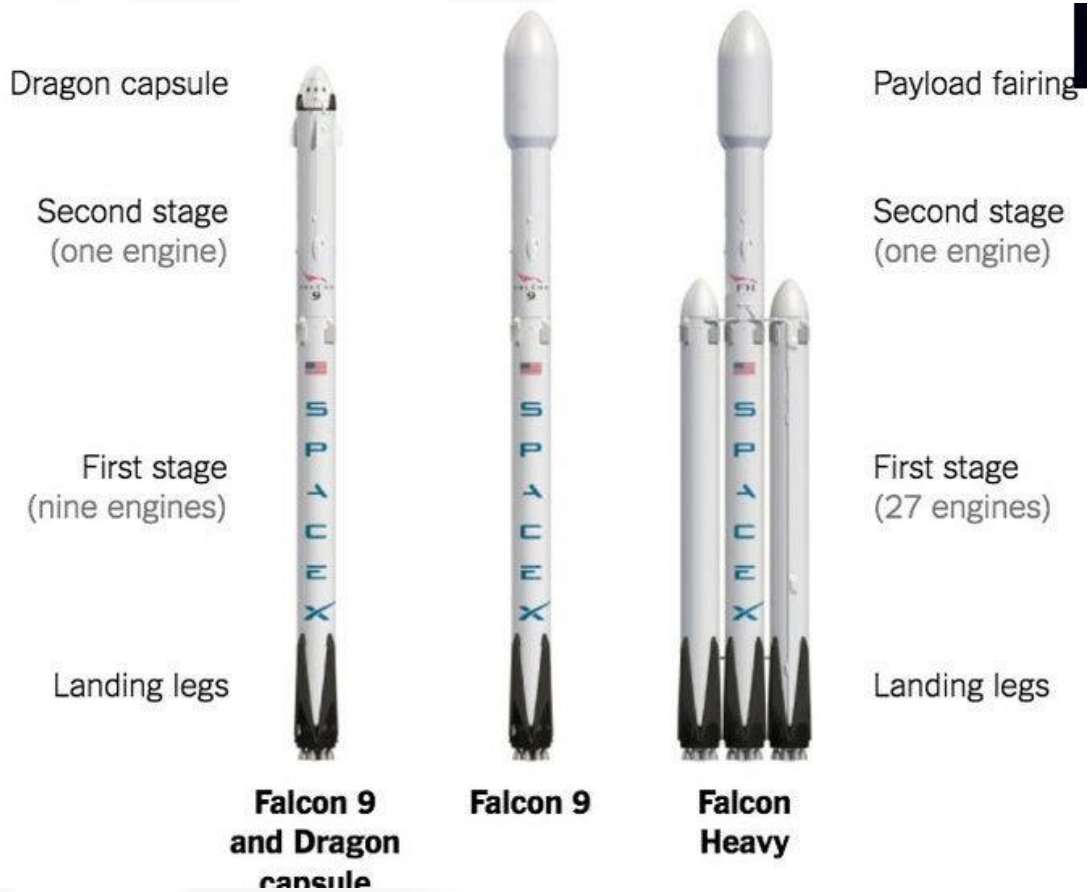
The launch success rate may depend on many factors, for example:

*payload mass and type,

*orbit type

*location (launch site)

*Initial position of rocket trajectories.



INTRODUCTION

- We created a machine learning pipeline to predict if the first stage will land given the data available.
- Then, we will be able to understand the aspects of the successful recoveries and apply them to optimize our own SpaceY rockets.
- Tons of money can be saved!

METHODOLOGY

METHODOLOGY

- Data collection
- Perform data wrangling
- EDA analysis using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

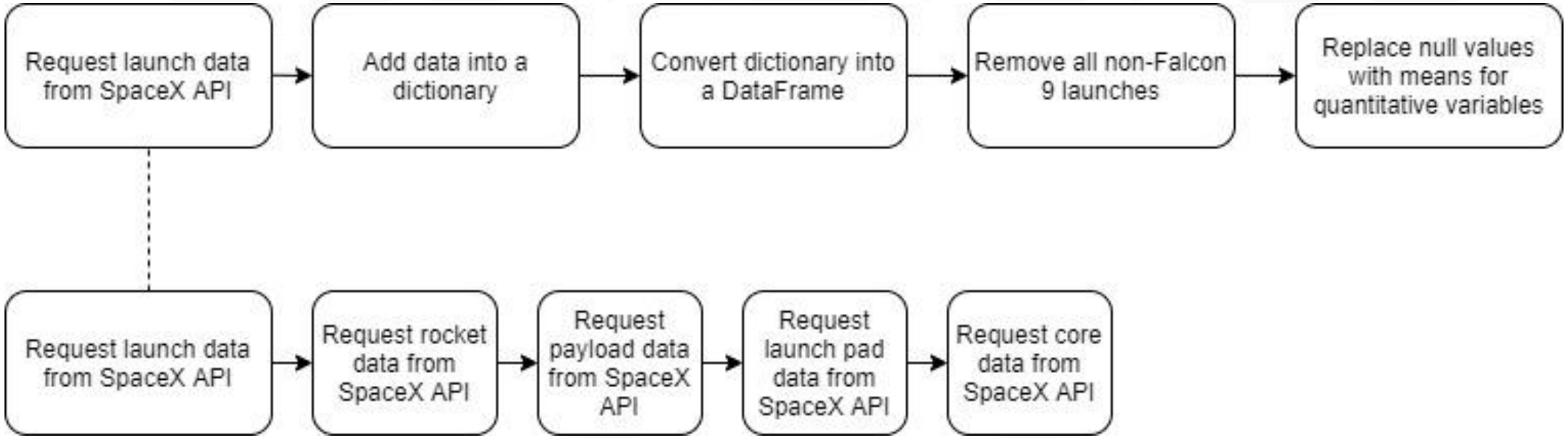
METHODOLOGY

Data collection - SpaceX API and Web scrapping

*Download a .json file containing rocket launch data from API

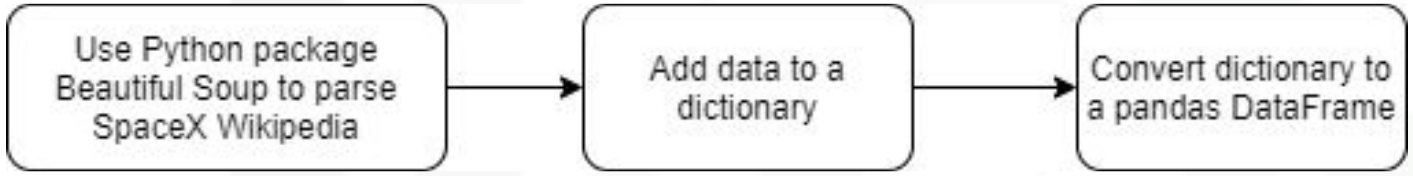
*Parse the information into a dataframe

*Convert null values (numerical) into the mean of the column

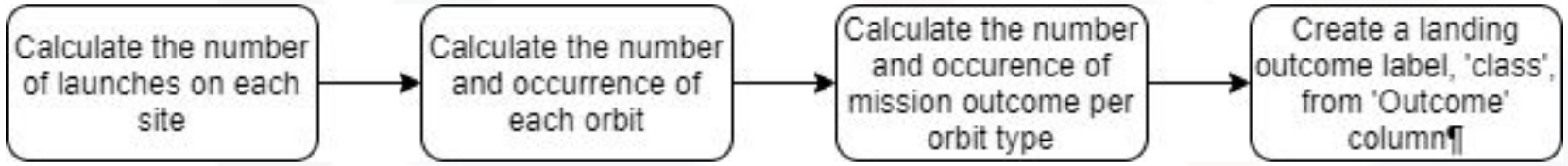


METHODOLOGY

Data collection - Web scrapping



Data wrangling



METHODOLOGY

Successful or Failure?

On how our data is organized, we can observe 8 different outcomes for this:

- True ASDS → Successful landing to drone ship
- True RTLS → Successful landing on a ground pad
- True Ocean → Successful landing in ocean
- None None → Failed to land
- None ASDS → Failed to land
- False ASDS → Failed to land on drone ship
- False RTLS → Failed to land on ground pad
- False Ocean → Failed to land in ocean

We created the column ‘class’ to distinguish successful than failure recovery:

0 → successful

1 → Failure

METHODOLOGY

EDA with data visualization:

*Explore the relationship between variables: histograms, scatterplots etc

*Matplotlib, plotly

*One-hot encoding

*See results for details

EDA with SQL:

*Get a better sense for the relationships between variables like:

Launch Site

Payload Mass (kg)

Mission Outcome

Booster Version

Date

METHODOLOGY

Interactive Map with Folium:

*View the location of each Falcon 9 launch site

*How many successful launches occurred at each location

*Determine distances to the closest coastline, city, railway, and highway.

Dashboard with Plotly Dash:

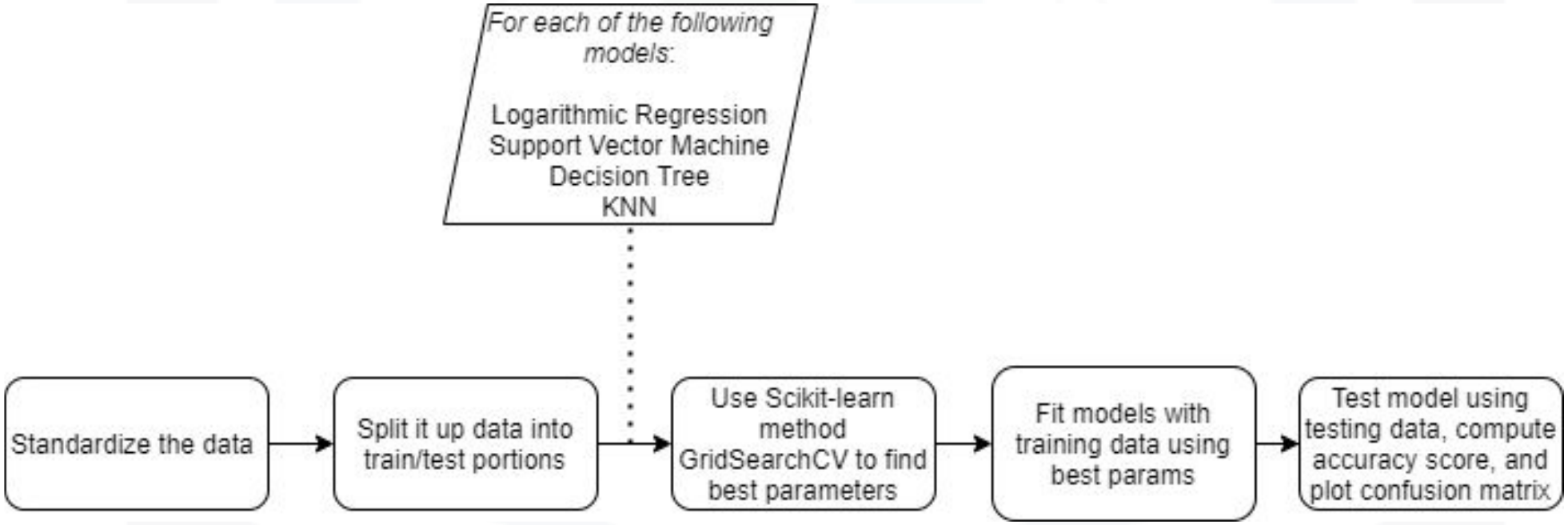
*A menu showing pie charts about the successful/failure recoveries on each site.

*A Recovery Outcome vs. Payload Mass scatter plot with a range (0-10000kg).

* It provides insights about the launch sites and payload masses relationship wrt. the recovery outcomes.

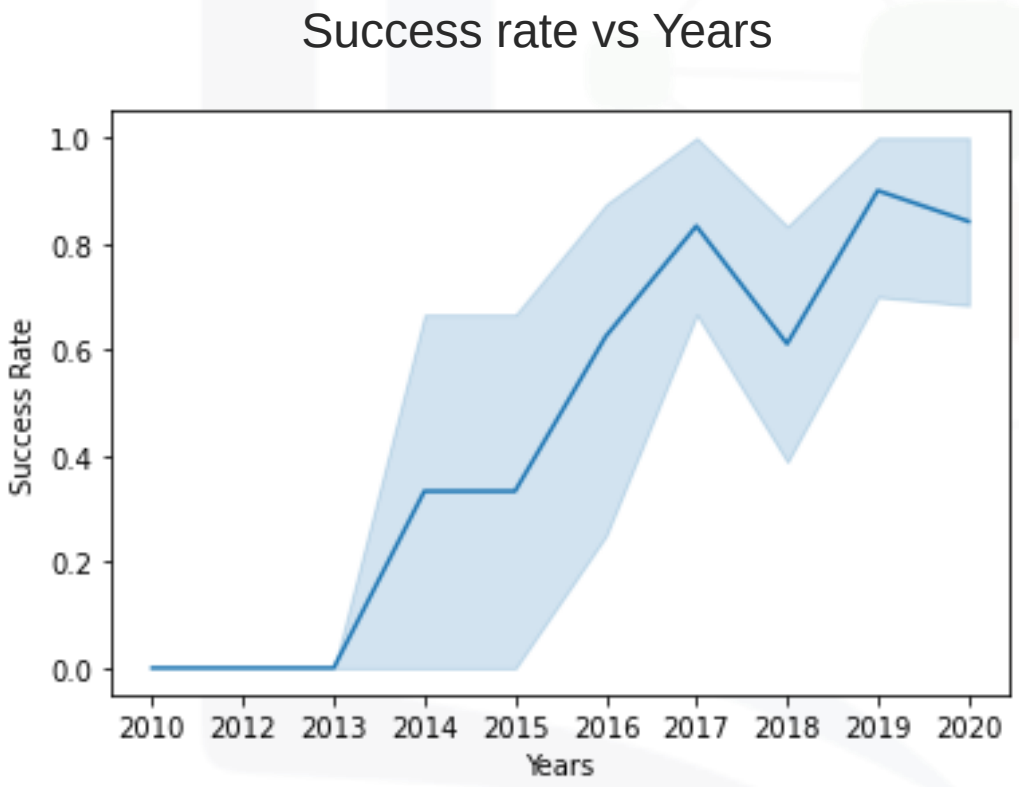
METHODOLOGY

Machine Learning Analysis (Classification)



RESULTS AND DISCUSSION - EDA

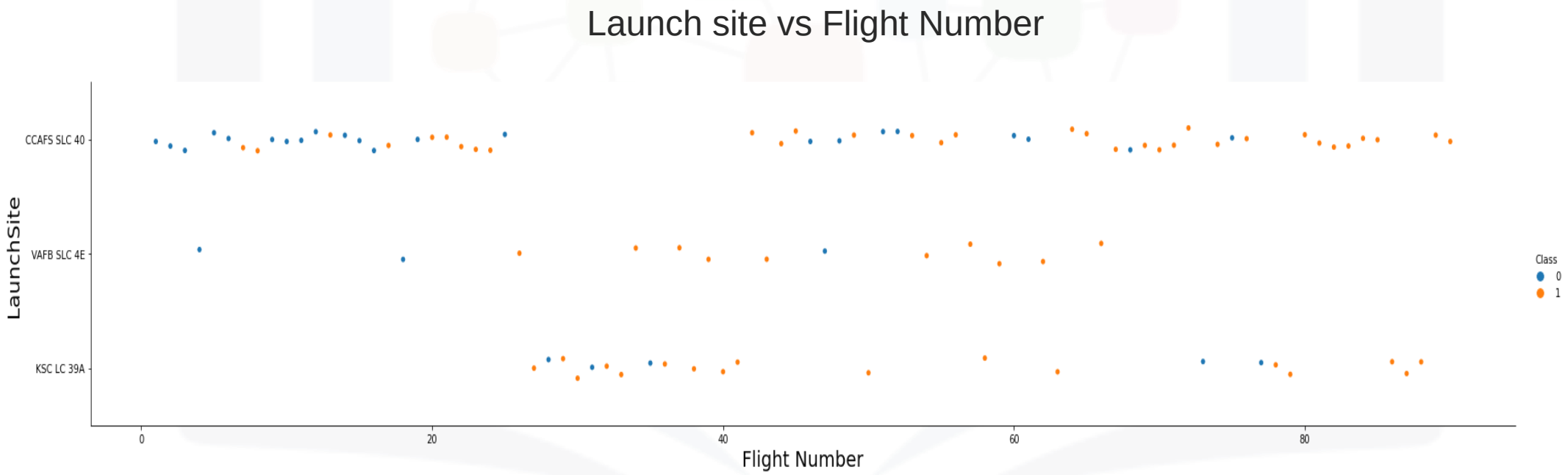
RESULTS - EDA



Launches are more successful from 2013 and keep increasing until 2020.

RESULTS - EDA

Site KSC LC-39A appears to be ideal as it has a success rate of over 75%

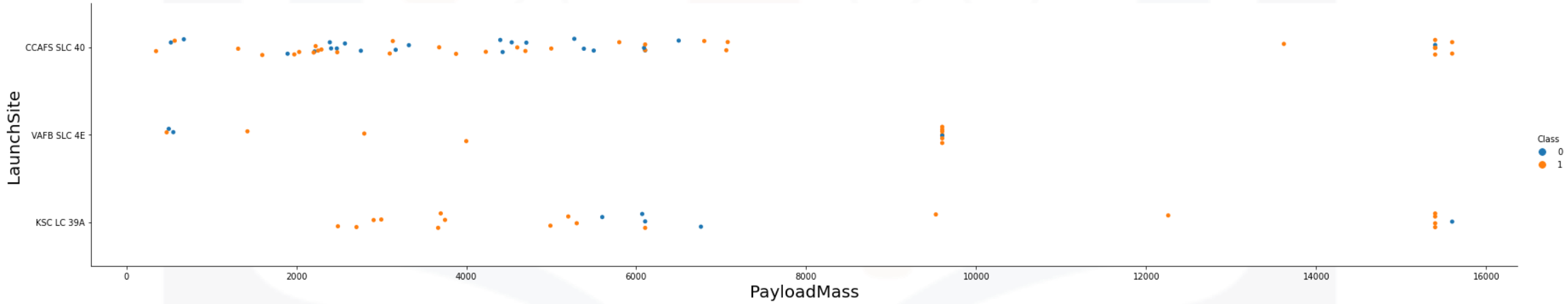


RESULTS - EDA

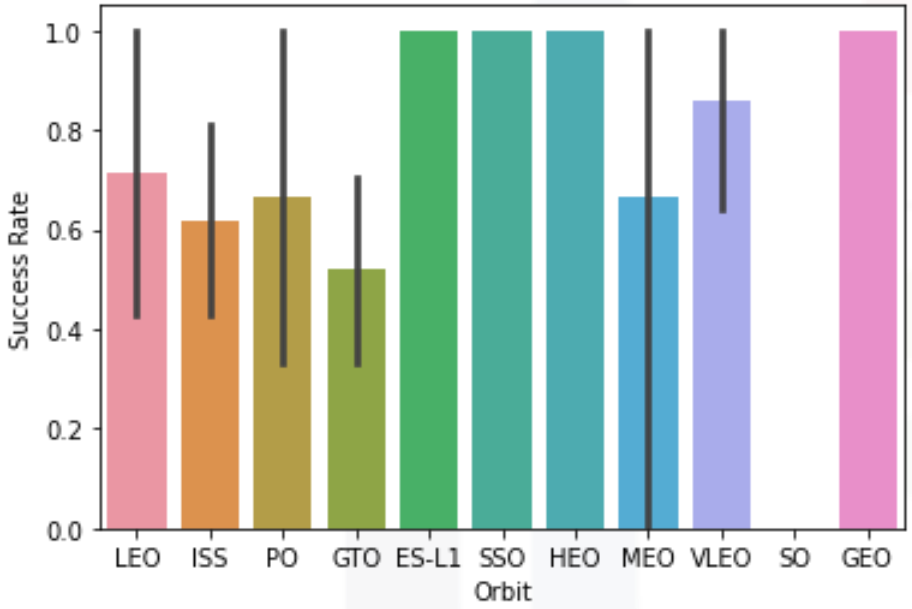
Light payloads are evidently easier to recover:

Most successful recoveries occur when the payload is between 2000kg - 4000kg

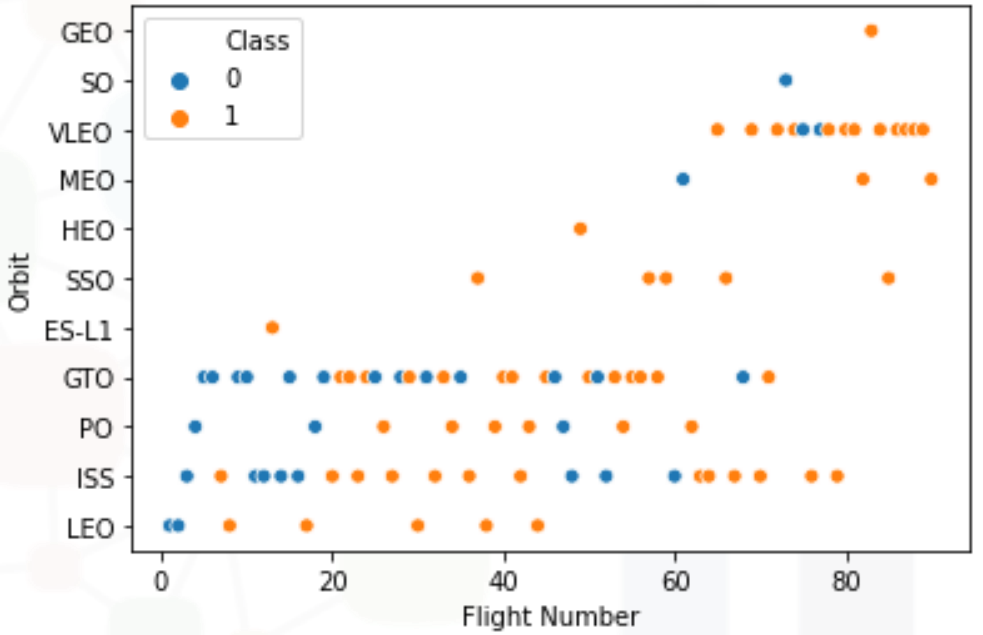
Launch site vs Payload mass



RESULTS - EDA



Success Rate vs. Orbit Type



Orbit Type vs Flight number

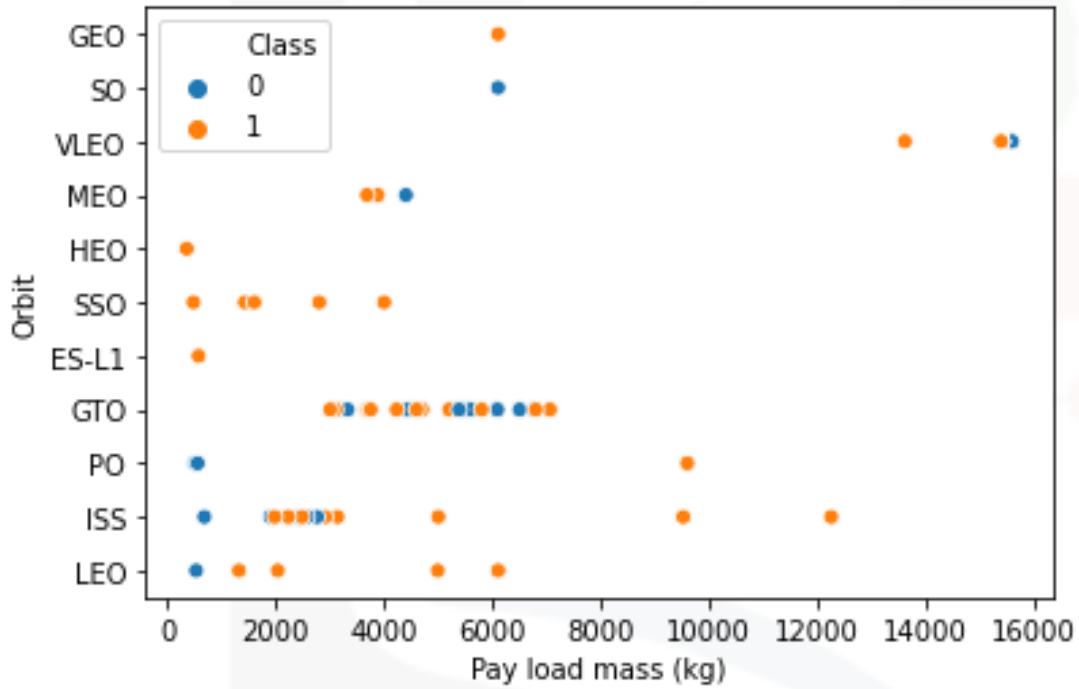
*ES-L1, SSO, HEO, and GEO orbits are very reliable

*ES-L1, HEO, and GEO have only 1 launch

*LEO, SSO, and VLEO all have high success rates and high sample sizes

RESULTS - EDA

Orbit Type vs Payload mass



The success of the LEO and SSO orbits is due to lighter payloads

RESULTS – EDA - SQL

According to SQL, we have 4 unique launch site locations

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

First 5 records for launch sites that 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	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

RESULTS – EDA - SQL

With QSL querying we also learned that:

Rockets with by booster version F9 v1.1 carry an average mass of

1
2534

Total payload mass (in kg) carried by boosters from NASA

1
111268

The first successful Stage One recovery landing occurred on

1
2010-06-04

Names of boosters which have successfully landed on a drone ship and had payload mass greater than 4000kg but less than 6000kg

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

101 missions recorded in this database.

Successful

1
100

Failure

1
1

RESULTS – EDA - SQL

With QSL querying we also learned that:

Names of boosters that have carried the maximum payload mass

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

In 2015, there were two launches which resulted in a failed Stage One recovery

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

RESULTS – EDA - SQL

With QSL querying we also learned that:

Rank of the types and number of landing outcomes (in descending order) between dates 2010-06-04 and 2017-03-20

landing__outcome	counts
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

RESULTS AND DISCUSSION - LAUNCH SITE ANALYSIS

RESULTS - LAUNCH SITE ANALYSIS

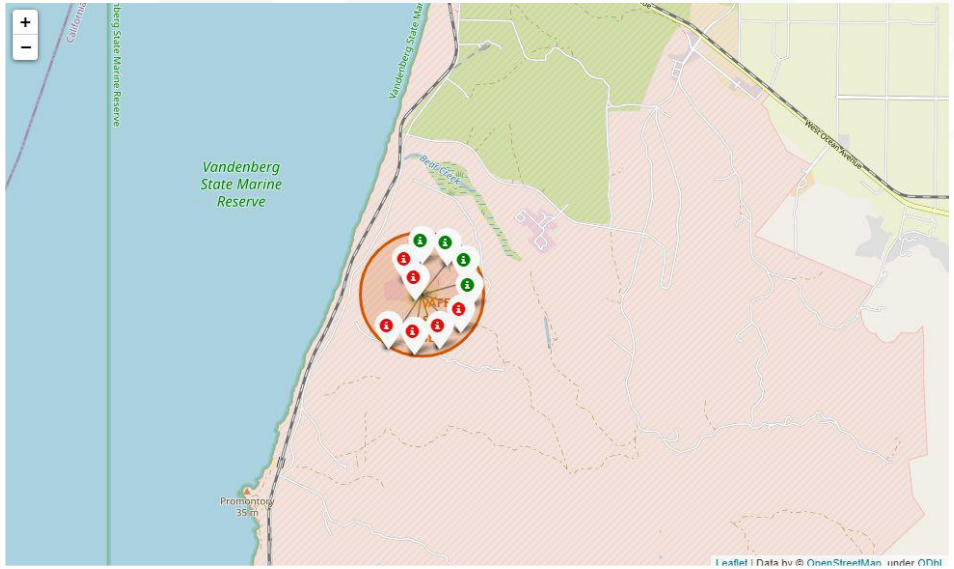
The rest of the launching sites are located on the Atlantic coast (Florida)

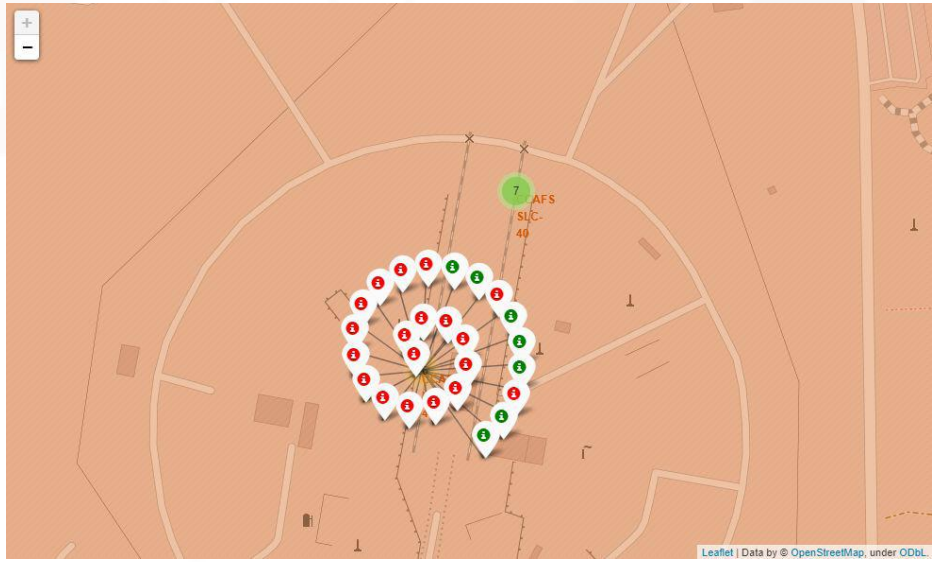
SpaceX has one launch site on the Pacific coast

RESULTS - LAUNCH SITE ANALYSIS

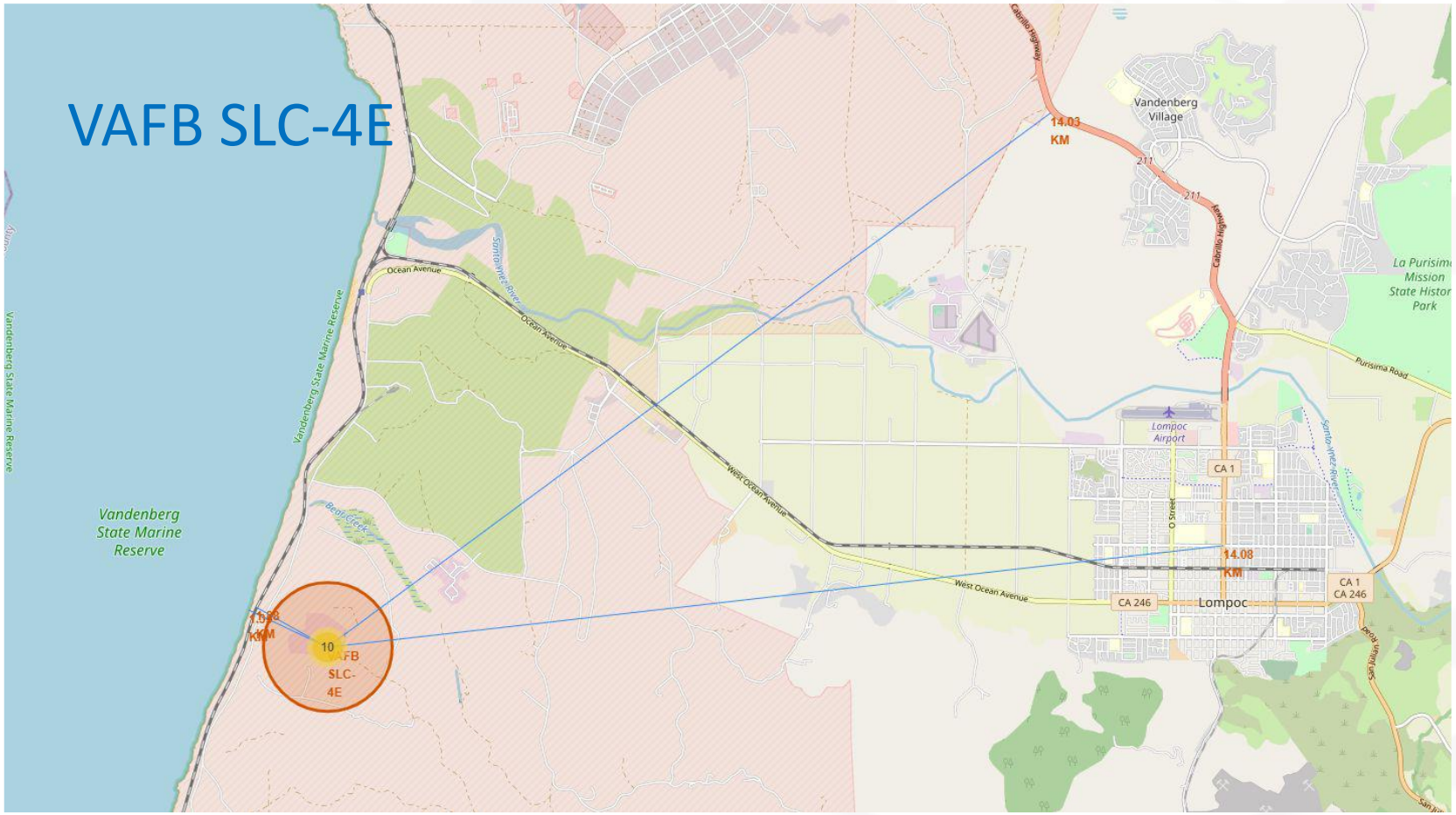
Each marker represents a Falcon 9 Stage One recovery

Green – successful recovery
Red – Failure recovery





RESULTS - LAUNCH SITE ANALYSIS



Each blue line represents the distance to the nearest

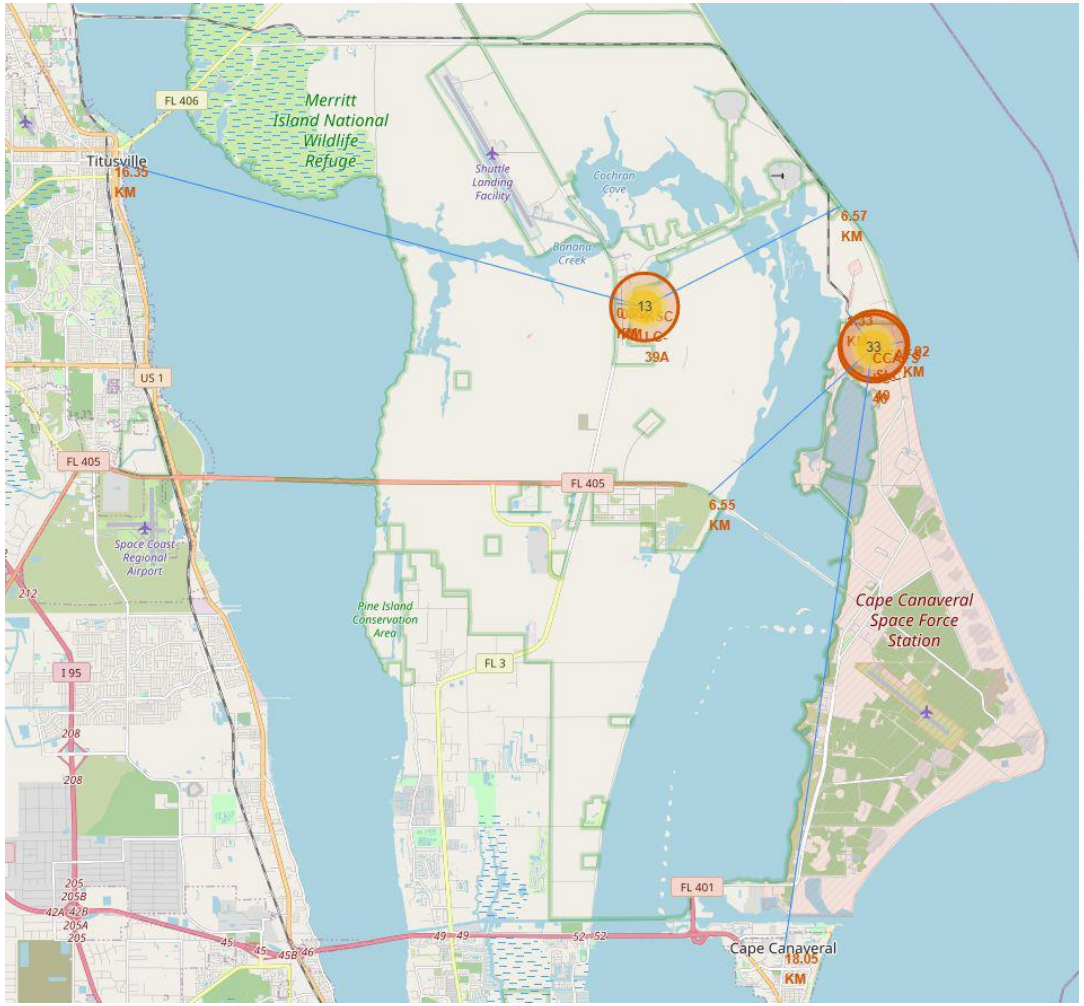
- coastline
- city/town
- railway
- highway

- Reasonable distance from the closest city and highway (~14km)
- Very close to the coast and the nearest railway (<2km)

IBM Developer

SKILLS NETWORK

RESULTS - LAUNCH SITE ANALYSIS



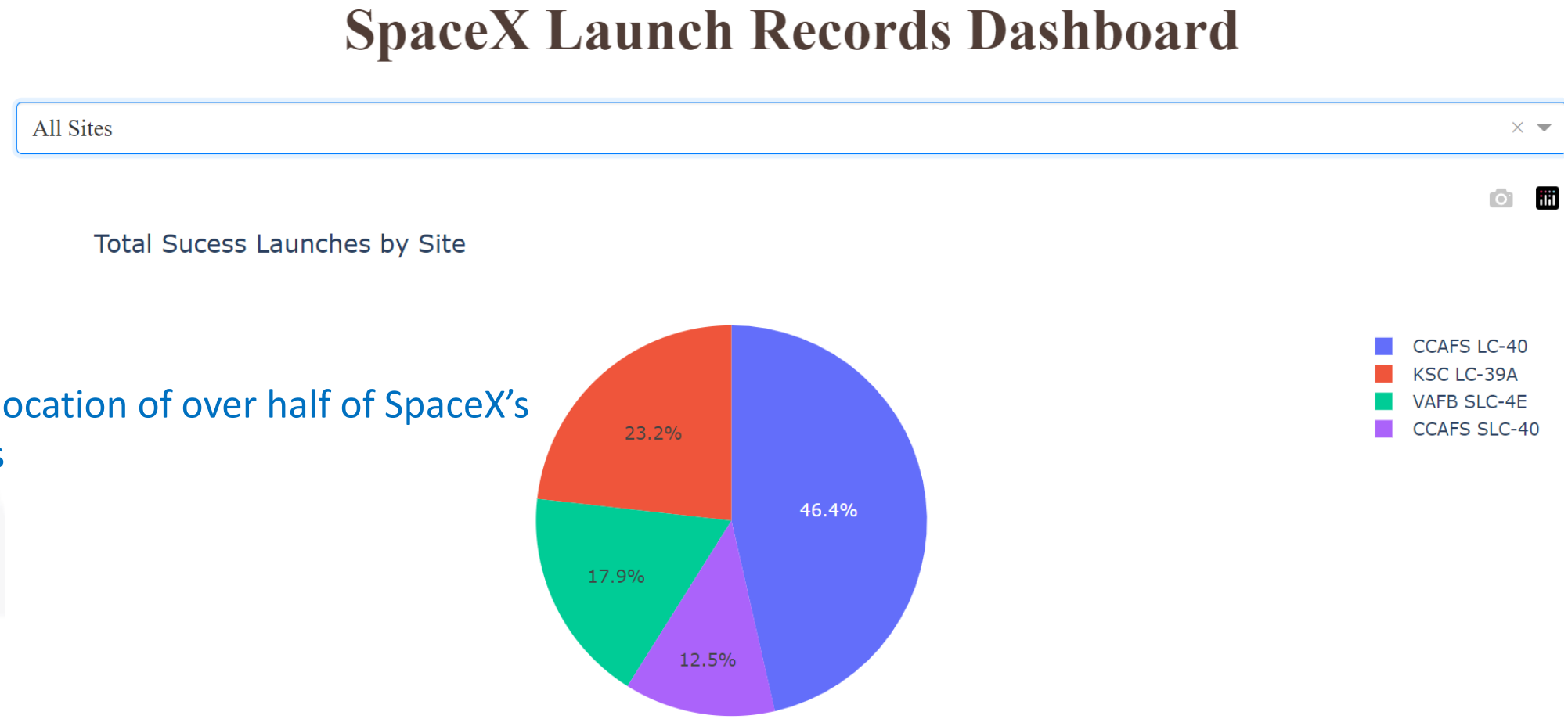
KSC LC-39A
CCAFS LC-40,
CCAFS SLC-40

- Reasonable distance from the nearest towns (>15km)
- Very close to the coast and the nearest railway (<7km)

RESULTS AND DISCUSSION - DASHBOARD

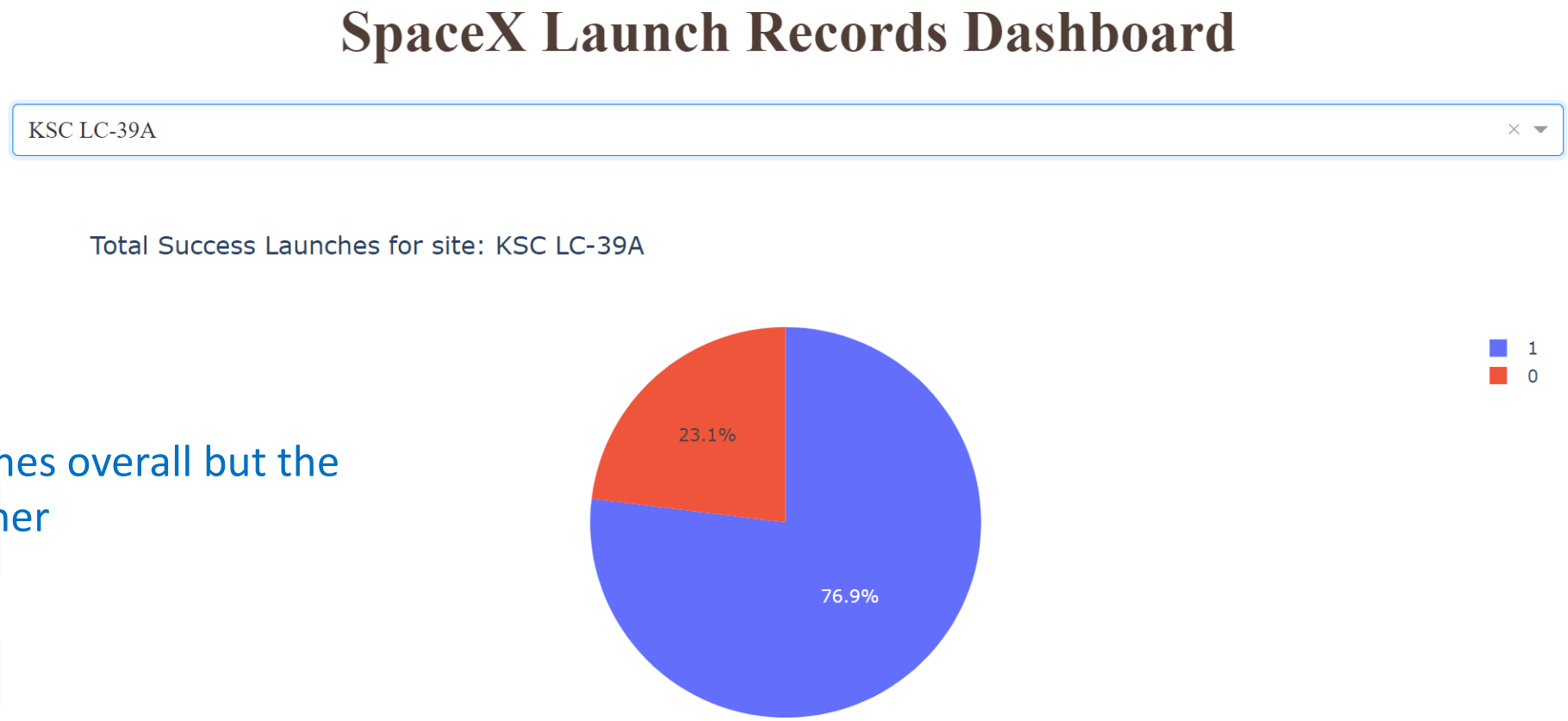
RESULTS - DASHBOARD

Here is a breakdown of the total amount of successful launches by site:



CCAFS LC-40 is the location of over half of SpaceX's Successful launches

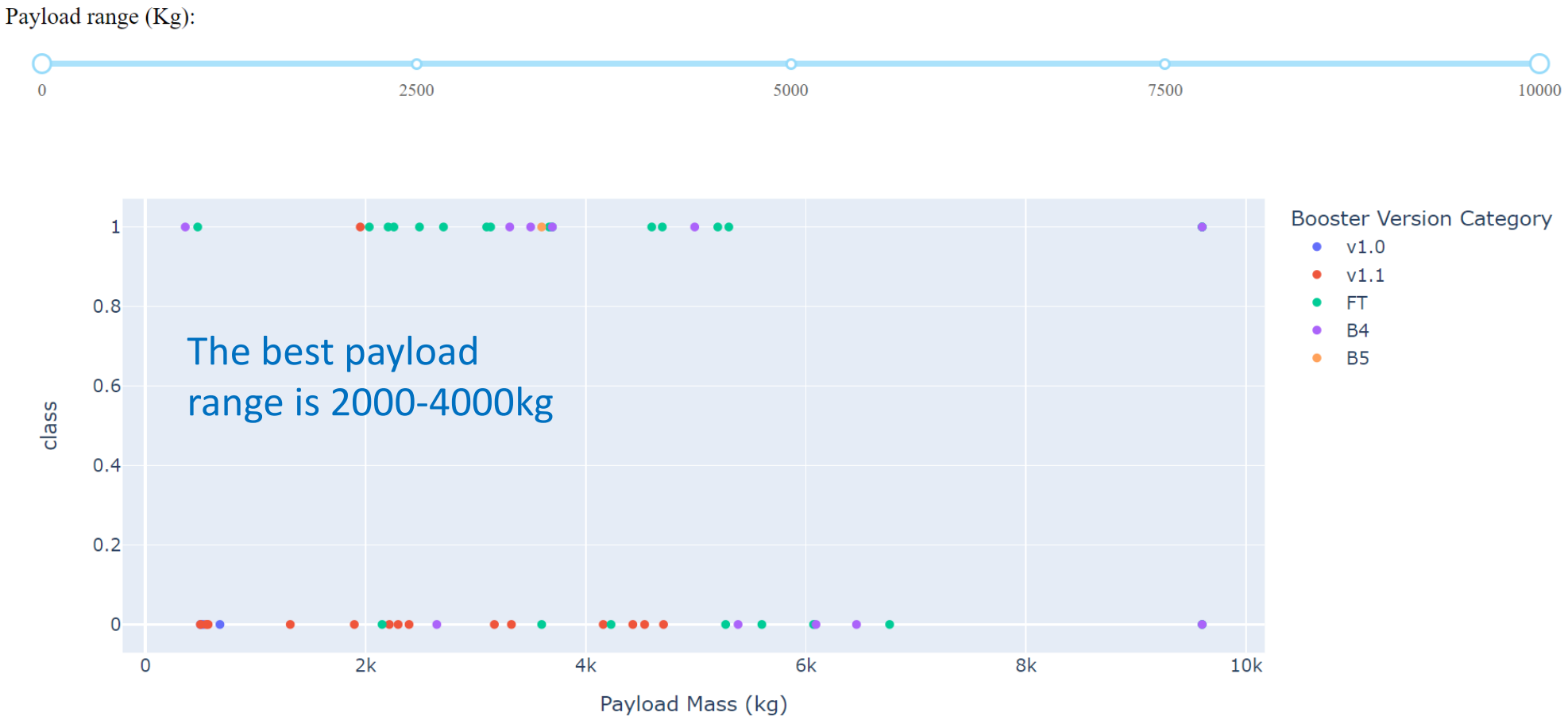
RESULTS - DASHBOARD



KSC LC-39A has fewer launches overall but the proportion of success is higher

RESULTS - DASHBOARD

Here is a breakdown of the total amount of successful launches by payload range:



The v1.1 booster Have failed the most while the FT booster is very successful

The best payload range is 2000-4000kg

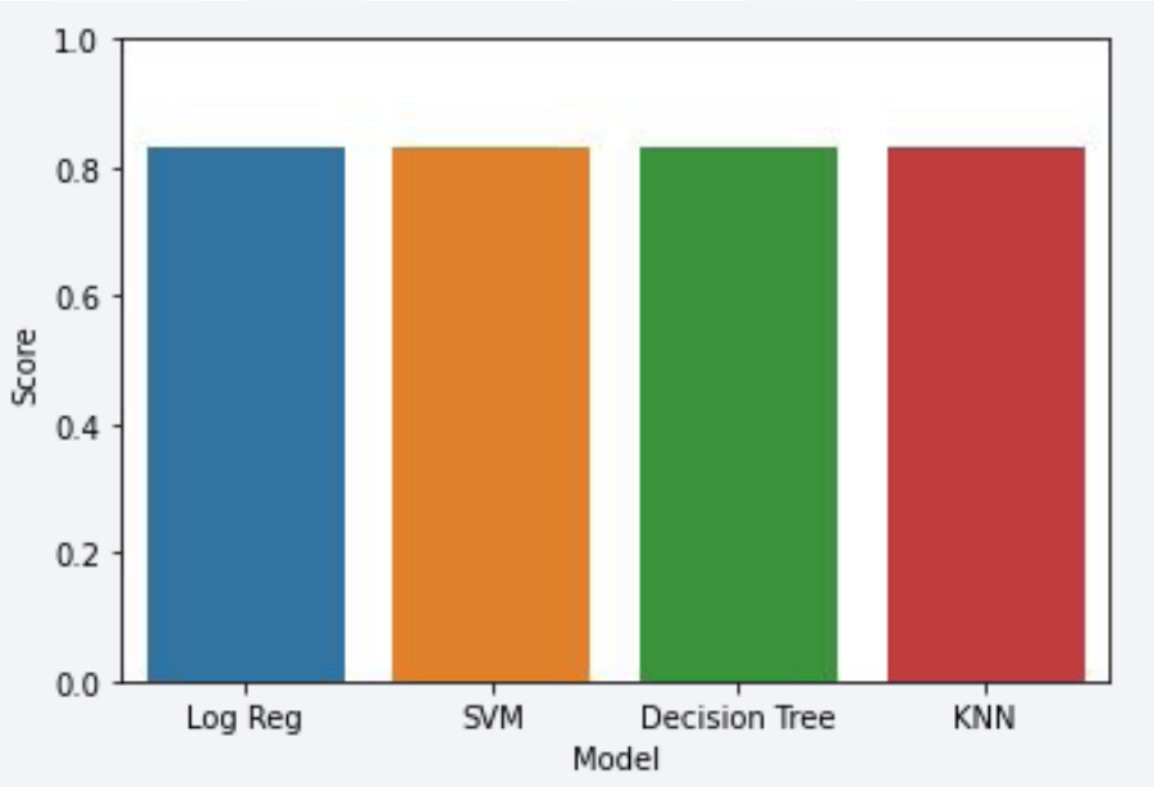
RESULTS AND DISCUSSION – MODEL DEVELOPMENT

RESULTS – MODEL DEVELOPMENT

We evaluated the different models using Grid Search to iterate and find the best hyperparameters for each one.

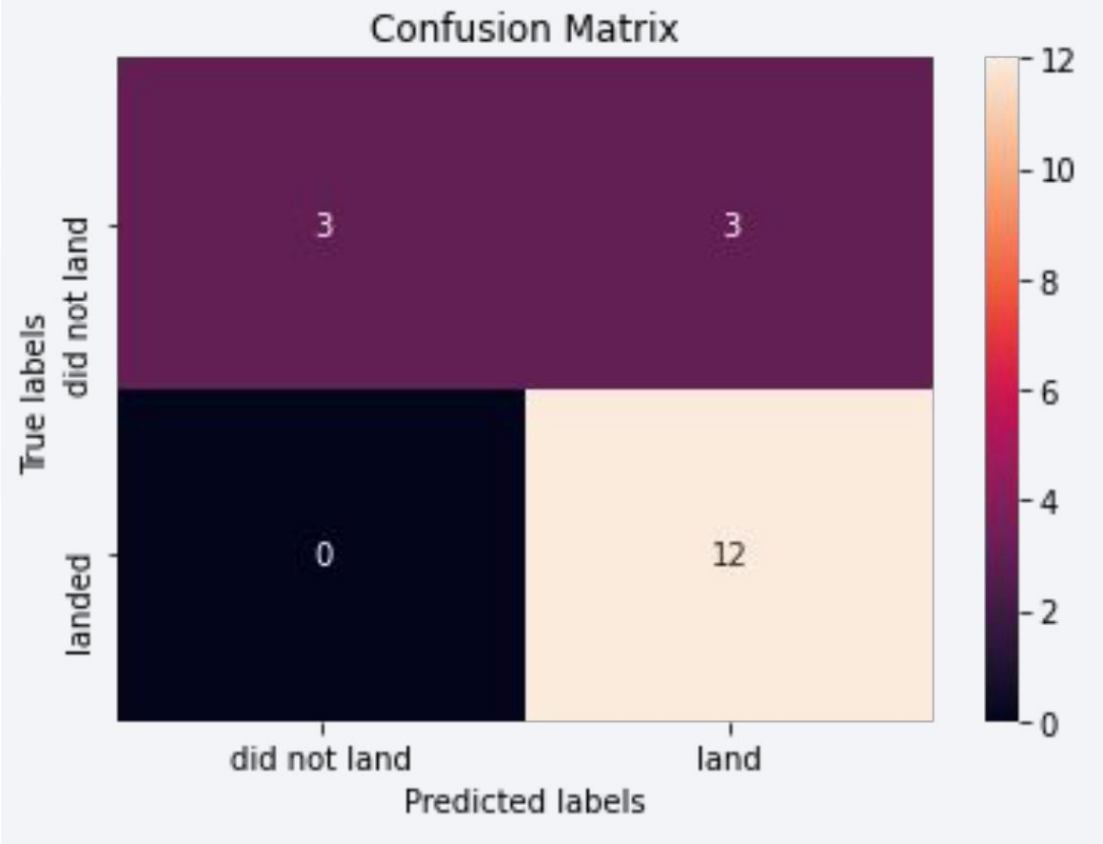
This bar plot displays the score for each model.

When evaluating the models in the test data, all the models exhibit similar scores of about 83%.



RESULTS – MODEL DEVELOPMENT

However, the confusion matrix shows that the models wrongly predicted three landings as being successful where they were not: False positives.



CONCLUSION

SpaceX's successful recovery data starts in 2013 and generally display the following properties:

- Light payload (in the range 2000-4000kg)
- Mostly launched from site KSC LC-39A
- Successfully recovered via drone ship

Our model can predict the outcome of a given recovery with a reasonable degree of accuracy, 83.33%

Check my github: https://github.com/angelina5285/Data-Science-Lab