IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Simon Nachevski
19.12.2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Methodology

- Data Collection Methodology:
  - Collect SpaceX API rocket launch data
- Data Wrangling
  - Create and modify columns for further use
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Split dataset into training and testing data
  - Develop machine learning models such as KNN, Decision Trees, Logistic Regression and SVM
  - Train and evaluate the models

Introduction

- Space Exploration companies can spend up to $165 million per launch
- SpaceX have cut their costs up to $62 million per launch
- With the recovery of the Stage One part of the rocket, which is a very expensive part
- But not always…

Section 1

Methodology

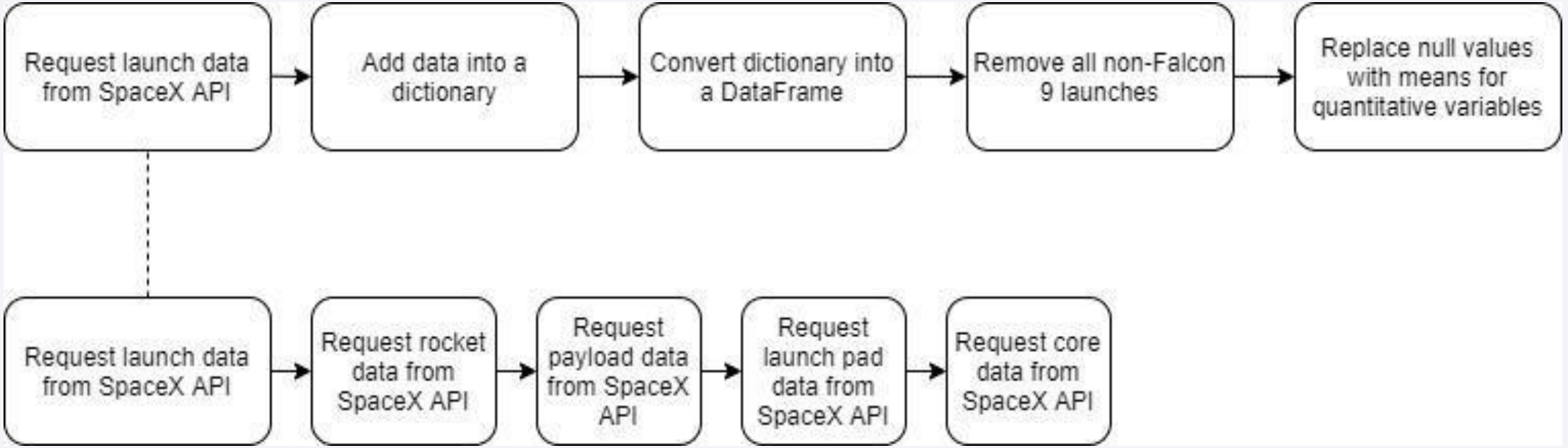
Methodology

- Data Collection Methodology:
  - Collect SpaceX API rocket launch data
- Data Wrangling
  - Create and modify columns for further use
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Split dataset into training and testing data
  - Develop machine learning models such as KNN, Decision Trees, Logistic Regression and SVM
  - Train and evaluate the models

Data Collection

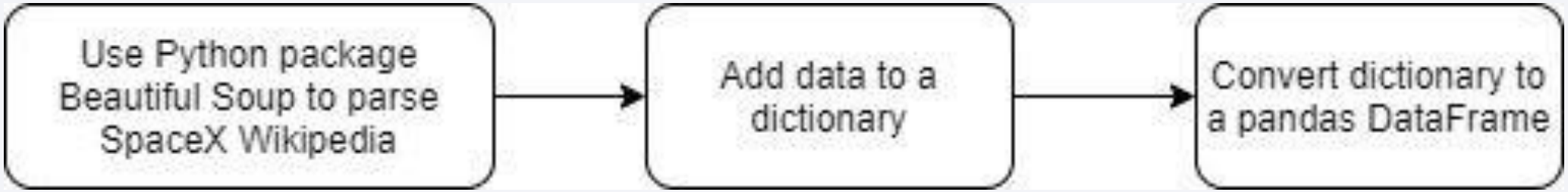
- Data Collection Methodology
  - Download SpaceX data using the SpaceX API
  - Create a dataframe for further use
  - Clean the data

Data Collection – SpaceX API



GitHub

Data Collection - Scraping



GitHub

Data Wrangling

Determine whether the launch was successful or not

From the dataframe, 8 possible outcomes can be defined:

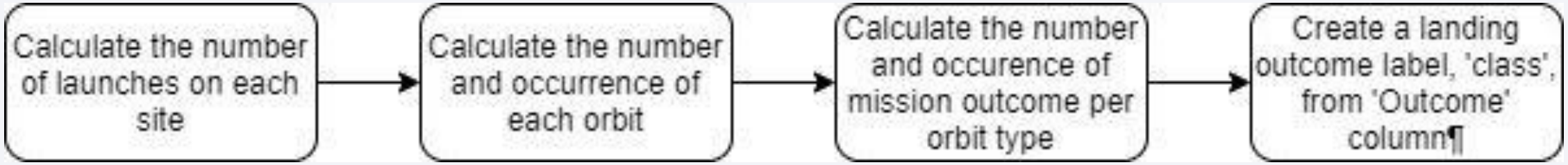
- True ASDS: Successful landing to a drone ship
- True RTLS: Successful landing on a ground pad
- True Ocean: Successful landing in the ocean
- None None: Failed to land
- None ASDS: Failed to land
- False ASDS: Failed to land on a drone ship
- False RTLS: Failed to land on a ground pad
- False Ocean: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

Data Wrangling

- The Stage One recovery data is split in new columns depending on the recovery:

  1 – Successful Recovery
  0 - Failed

Data Wrangling



GitHub

EDA with Data Visualization

- Plots of relationships within variables:
  - Cat plot of Flight Number and Payload
  - Cat plot of Flight Number and Launch Site
  - Scatter plot of Launch Site and Payload
  - Bar plot of Success Rate and Orbit type
  - Scatter plot of Orbit type and Flight Number
  - Scatter plot of Orbit type and Payload
  - Line plot of Success rate and Time (Years)

EDA with SQL

With the help of SQL, we make some basic queries on our data to get a better sense for the relationships between variables, particularly the following:

- Launch Site
- Payload Mass in kg
- Mission Outcome
- Booster Version
- Date

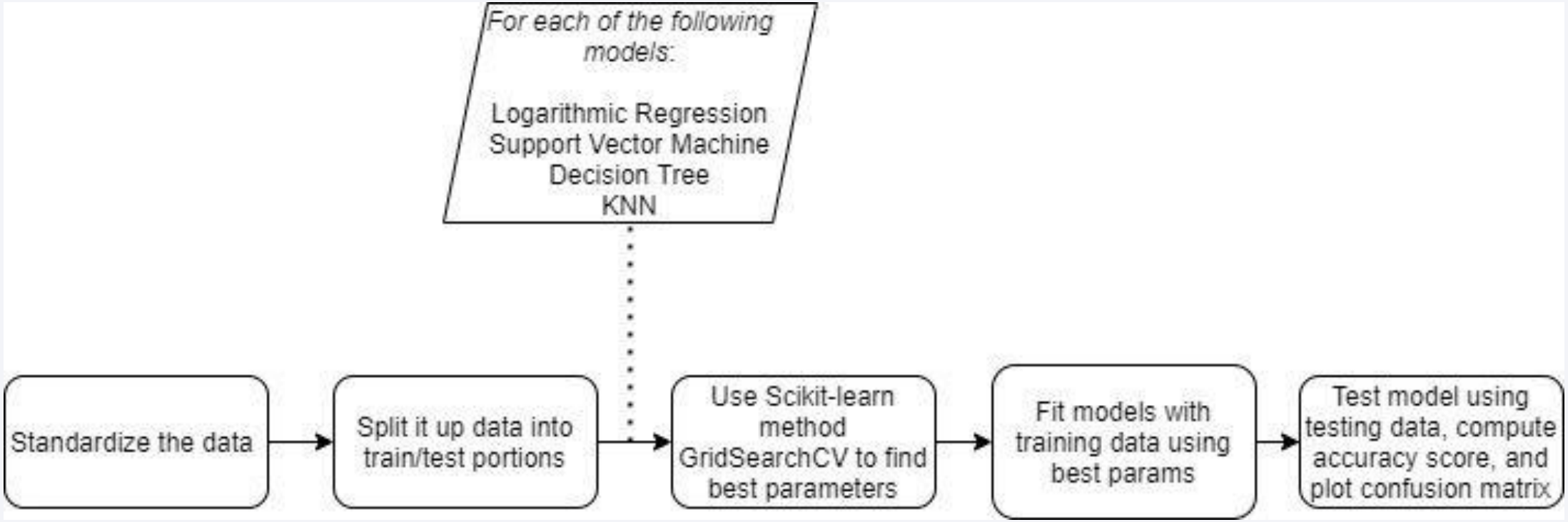
GitHub

Build an Interactive Map with Folium

- With Folium, an interactive map has been created in which you can:
  - Falcon 9 launch sites as circles
  - Launches by location, represented by markers. Green markers are successful recoveries. Red markers are unsuccessful recoveries
  - Distances to closest coastlines , cities, railways and highways, shown as a blue line

GitHub

Predictive Analysis

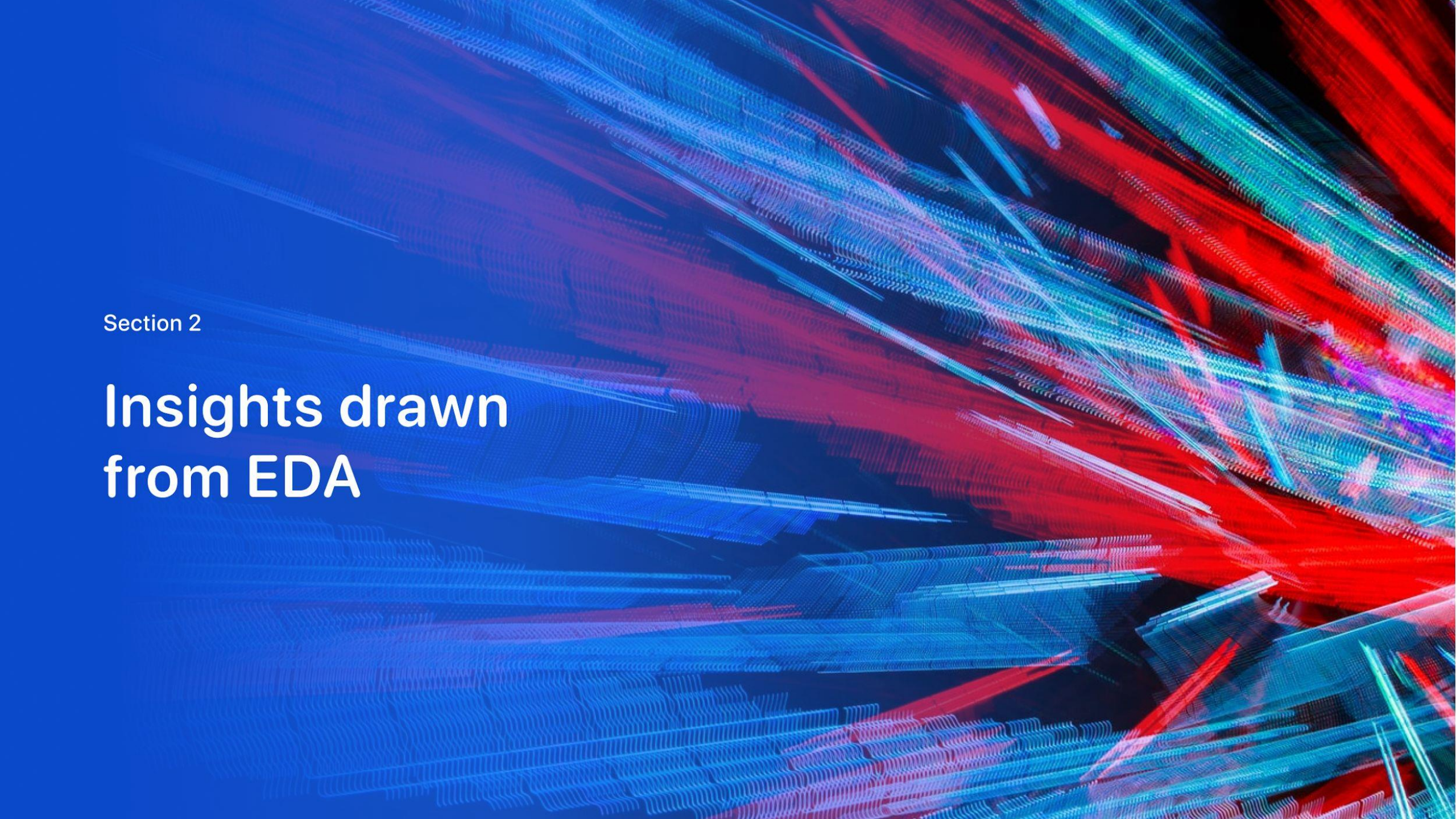


GitHub

Results

EDA Results:

- SpaceX has gotten better at launching rockets overtime. Launches are most successful when launched in 2017 or later
- Light payloads are evidently easier to recover, as most successful recoveries occur when the payload has a mass between 2000kg and 4000kg)
- Launch site KSC LC-39A appears to be ideal as it has a success rate of over 75%
- The best recovery method is through drone ships



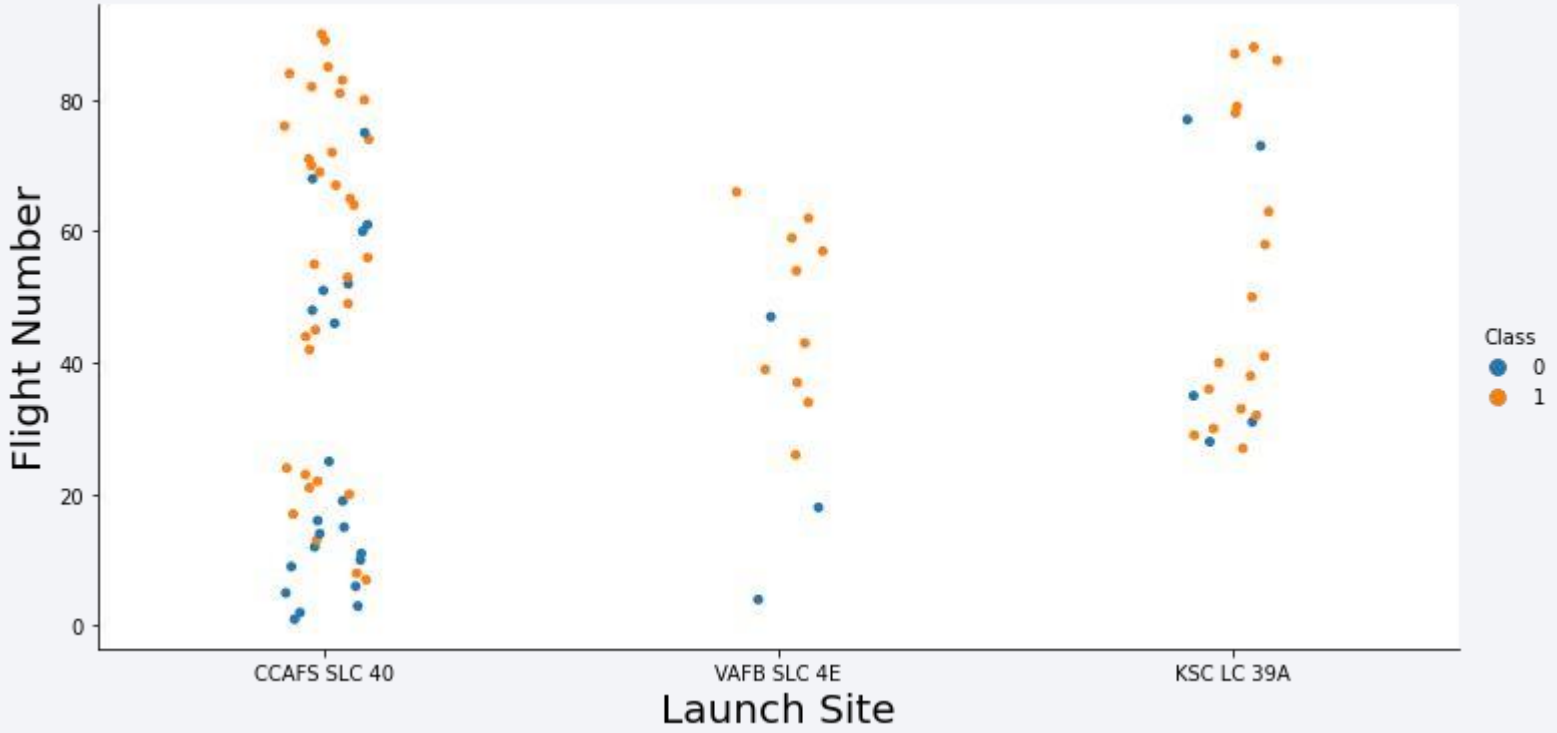
Section 2

Insights drawn from EDA

EDA with Data Visualization

KSC LC-39A seems to be the most consistent launch site

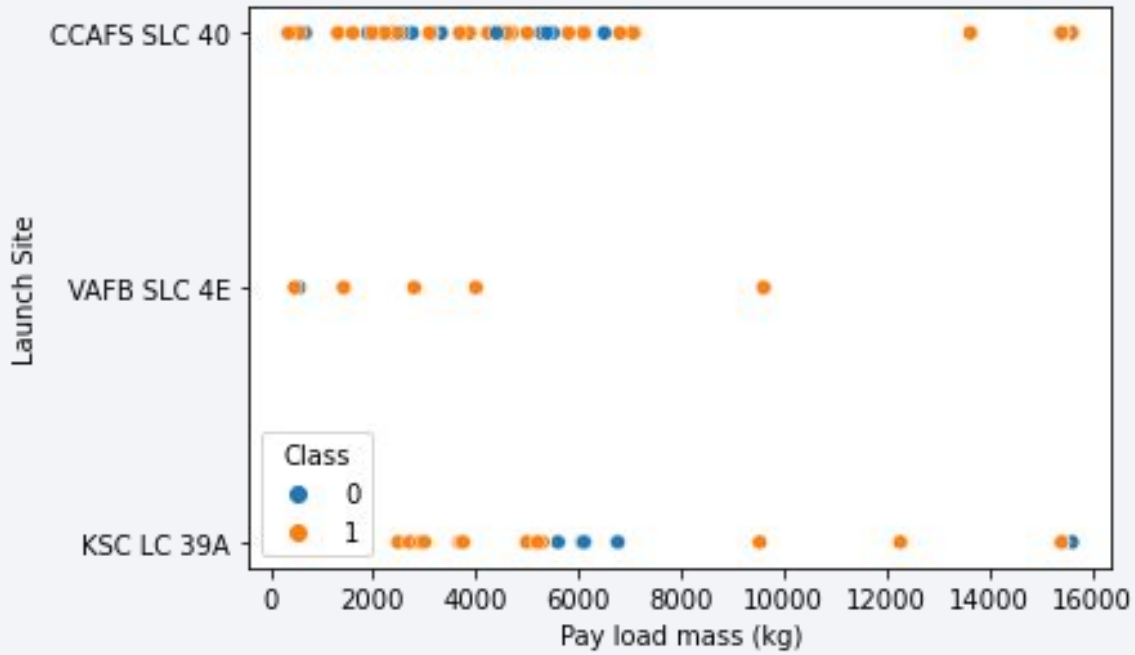
Flight Number vs Launch Site (Cat Plot)



EDA with Data Visualization

The smaller the payload, the higher the success rate

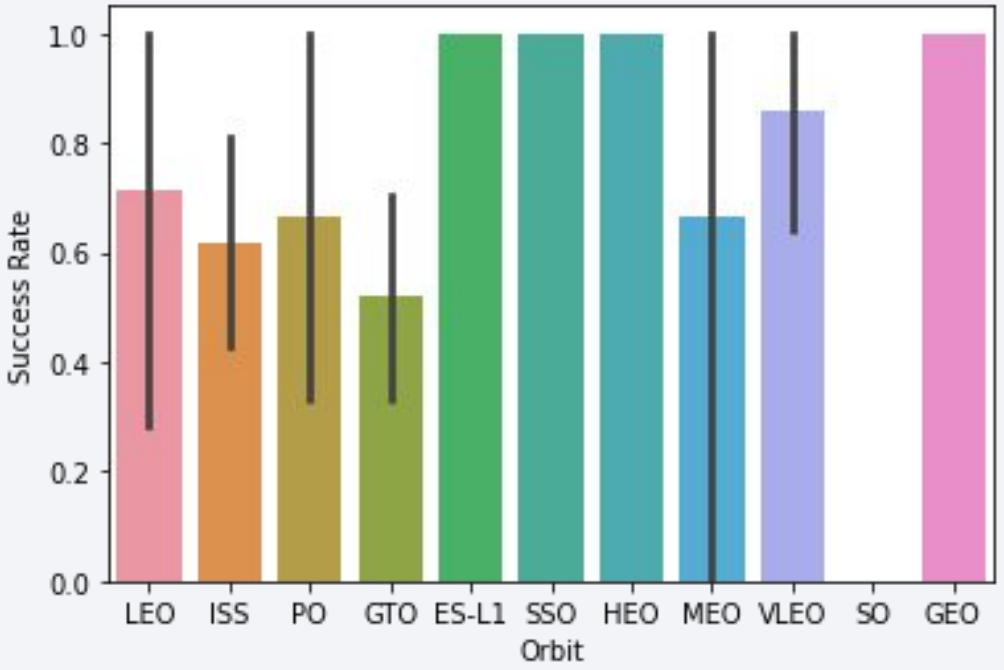
Launch Site vs Payload Mass (Scatter Plot)



EDA with Data Visualization

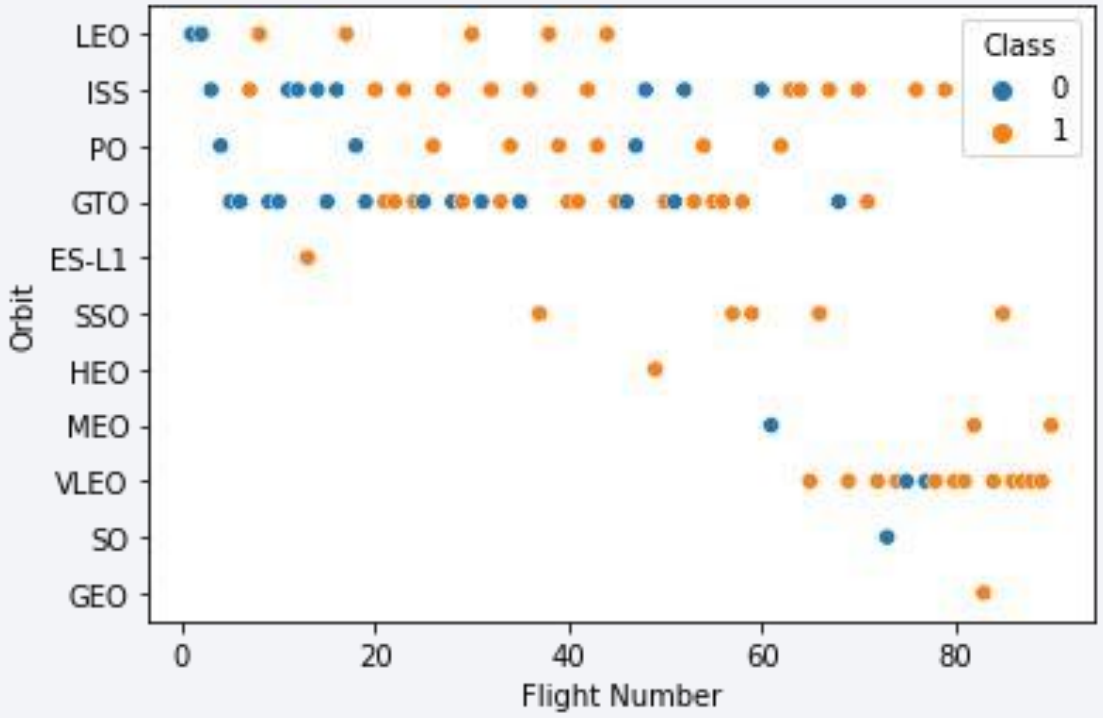
The ES-L1, SSO, HEO, and GEO orbits are very reliable

Success Rate vs. Orbit Type (Bar plot)



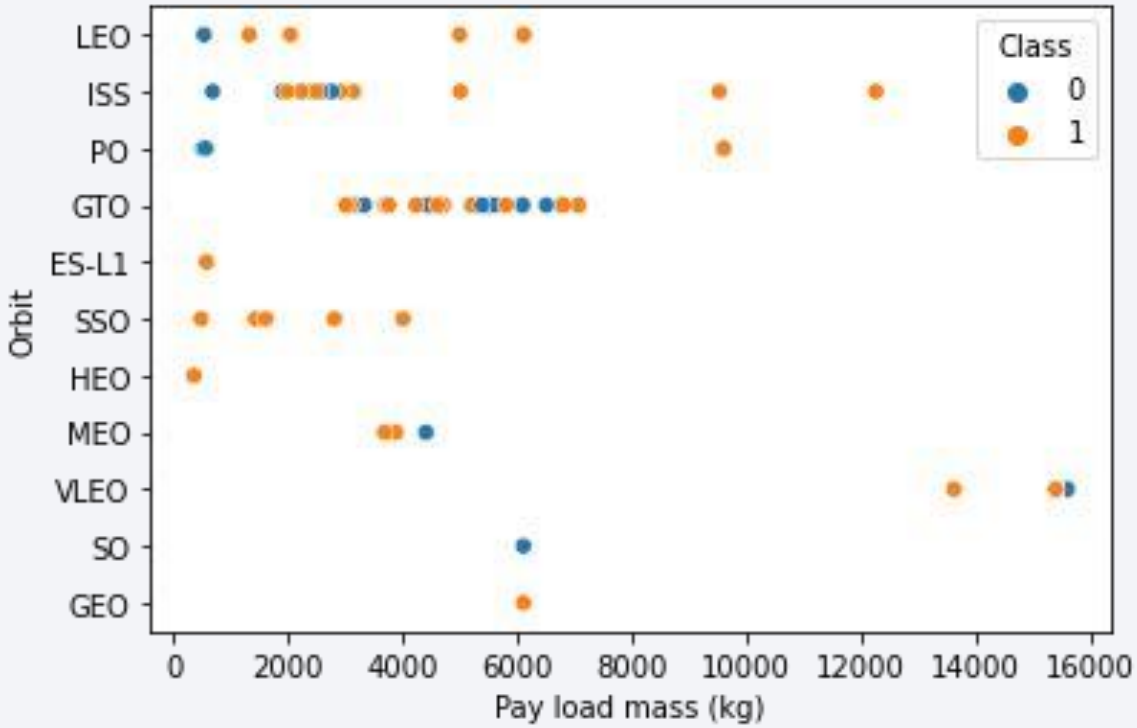
Flight Number vs. Orbit Type

- ES-L1, HEO, and GEO success rates are skewed due to each having only 1 launch
- LEO, SSO, and VLEO have higher success rates and higher number of samples



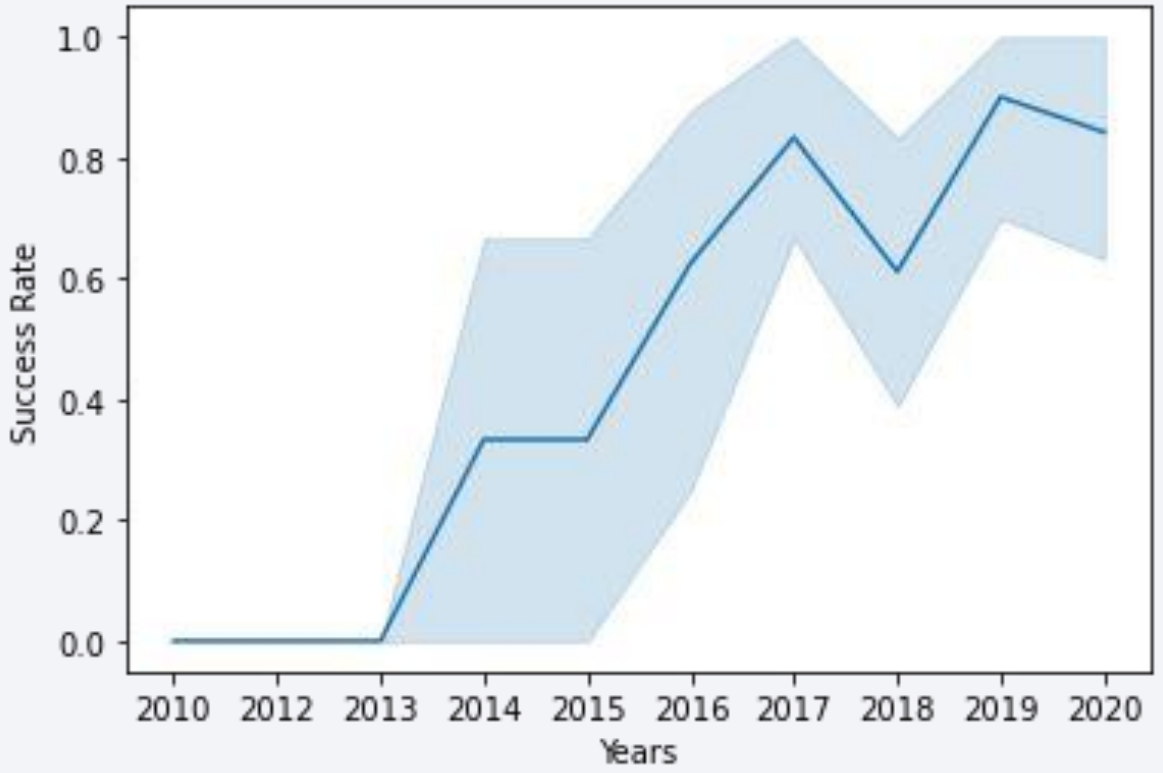
Payload vs. Orbit Type

- LEO and SSO orbits' success may be due to light payloads



Launch Success Yearly Trend

- Overall success rate increased the most from 2013 to 2017
- Since 2017, the success rate has lowered



Launch Sites

There are four launch sites:

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

Launch Site Names Begin with 'CCA'

- 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	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

Total Payload Mass

- Here is the total payload mass (in kg) carried by boosters from NASA:

SUM(payload_mass__kg_)
111268

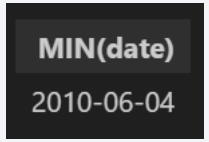
Average Payload Mass by F9 v1.1

- On average, rockets with by booster version F9 v1.1 carry a mass of 2534,66

AVG(payload_mass__kg_)
2534.6666666666665

First Successful Ground Landing Date

- The first successful Stage One recovery landing occurred on:



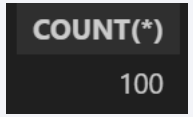
Successful Drone Ship Landing with Payload between 4000 and 6000

- This is a list of the names of boosters which have successfully landed on a drone ship and had payload mass greater than 4000kg but less than 6000kg

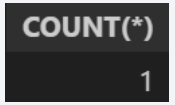
Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

- In total, there were 101 missions recorded in this database, which were either:
  - Success



  - Failure



Boosters Carried Maximum Payload

- List of the names of boosters which 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

2015 Launch Records

- 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

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

- Rank of the types and number of landing outcomes between dates 2010-06-04 and 2017-03-20:

Landing_Outcome	counts
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 4

Launch Sites Proximities Analysis

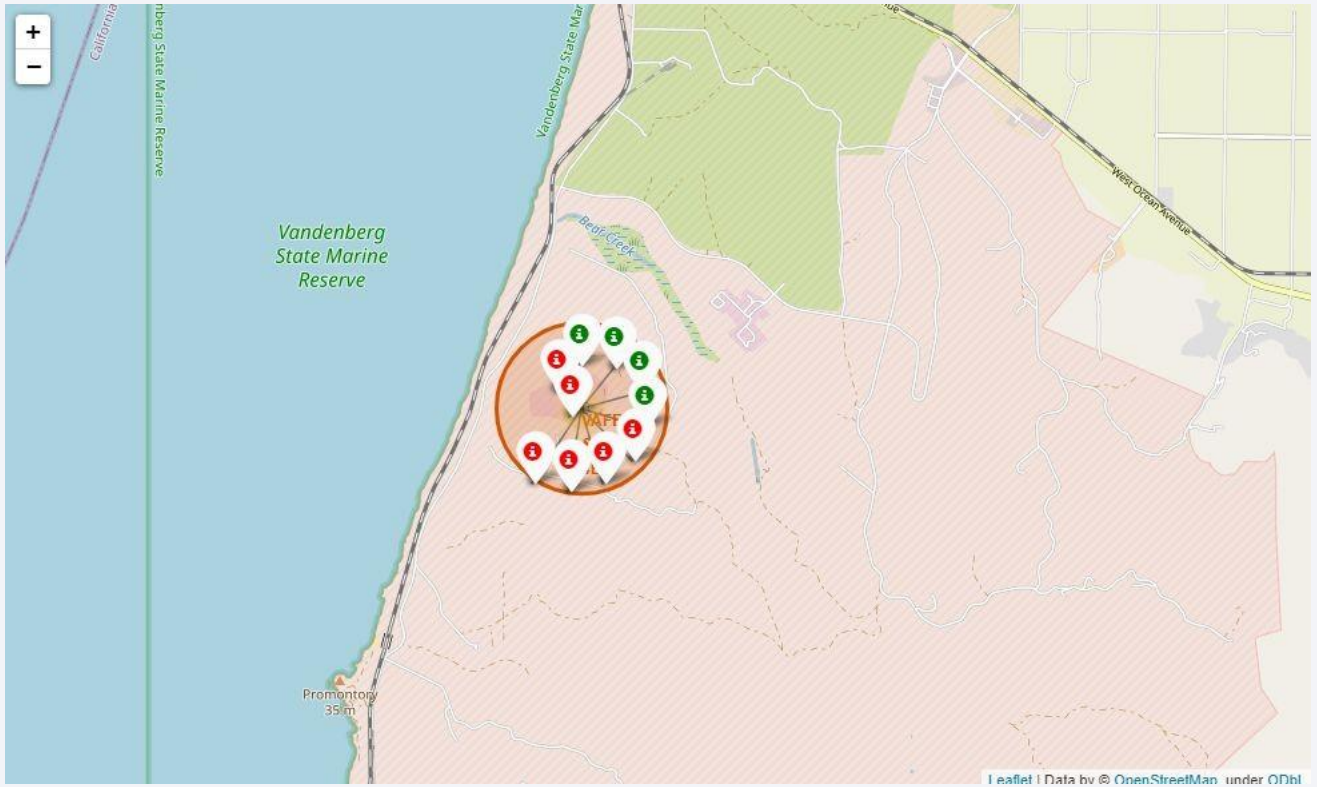
Map of Launch Site Locations

- As can be seen, SpaceX has one launch site on the Pacific coast of Southern California
- The rest of the launch sites are located on the Atlantic coast of Florida

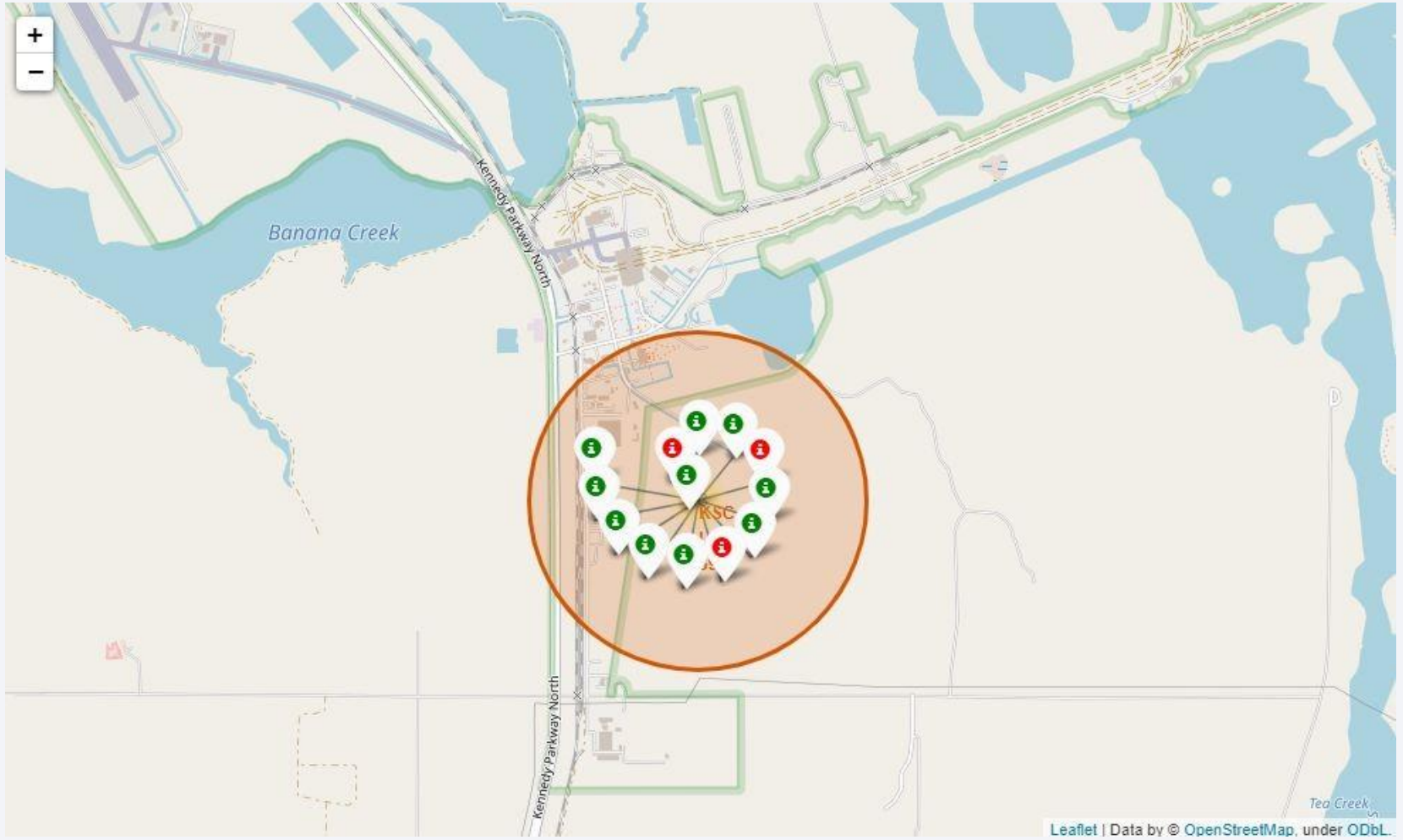
VAFB SLC-4E Recovery Outcomes

- The markers represent a Falcon 9 Stage One recovery

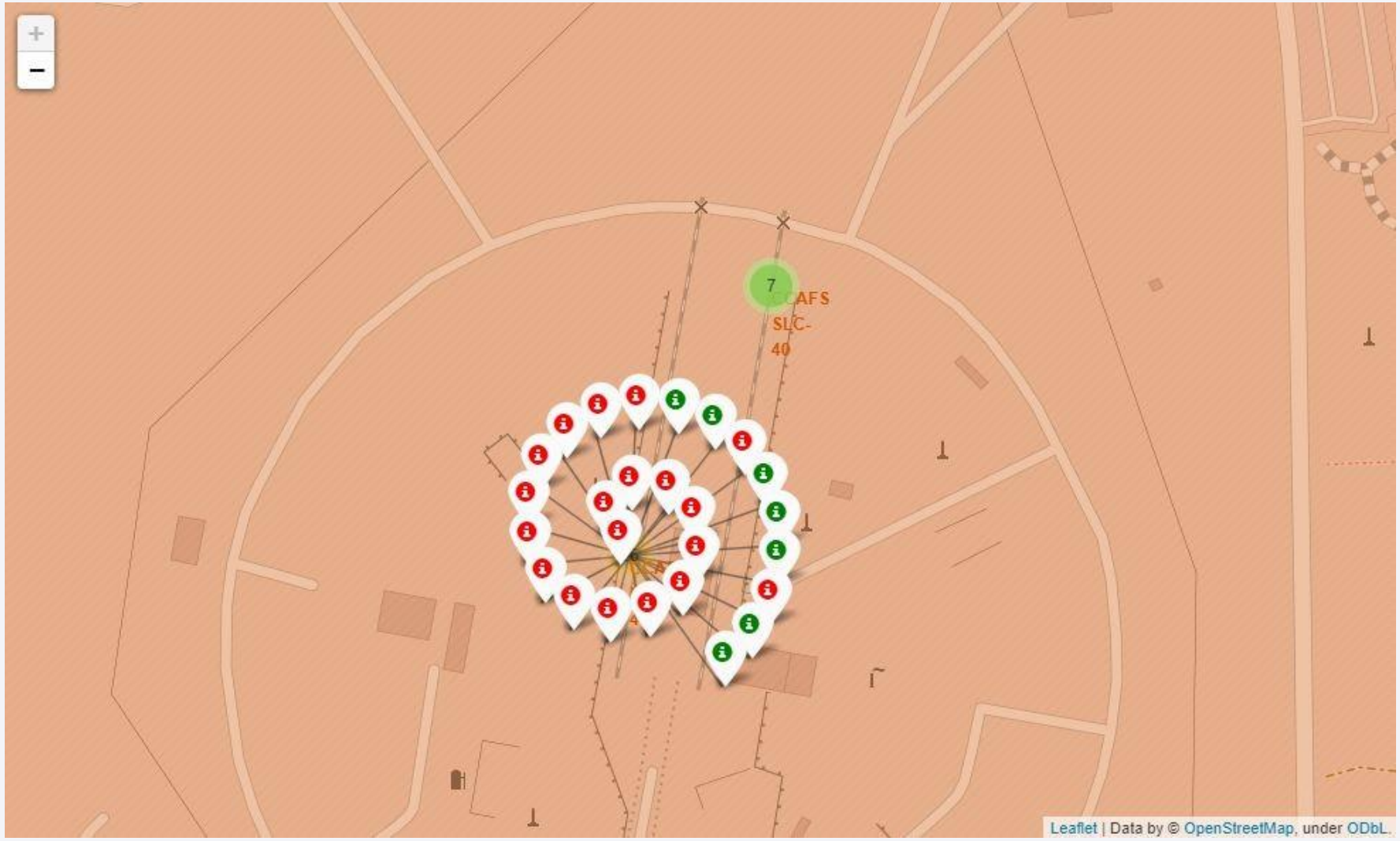
- Green is a successful recovery
- Red is an unsuccessful recovery



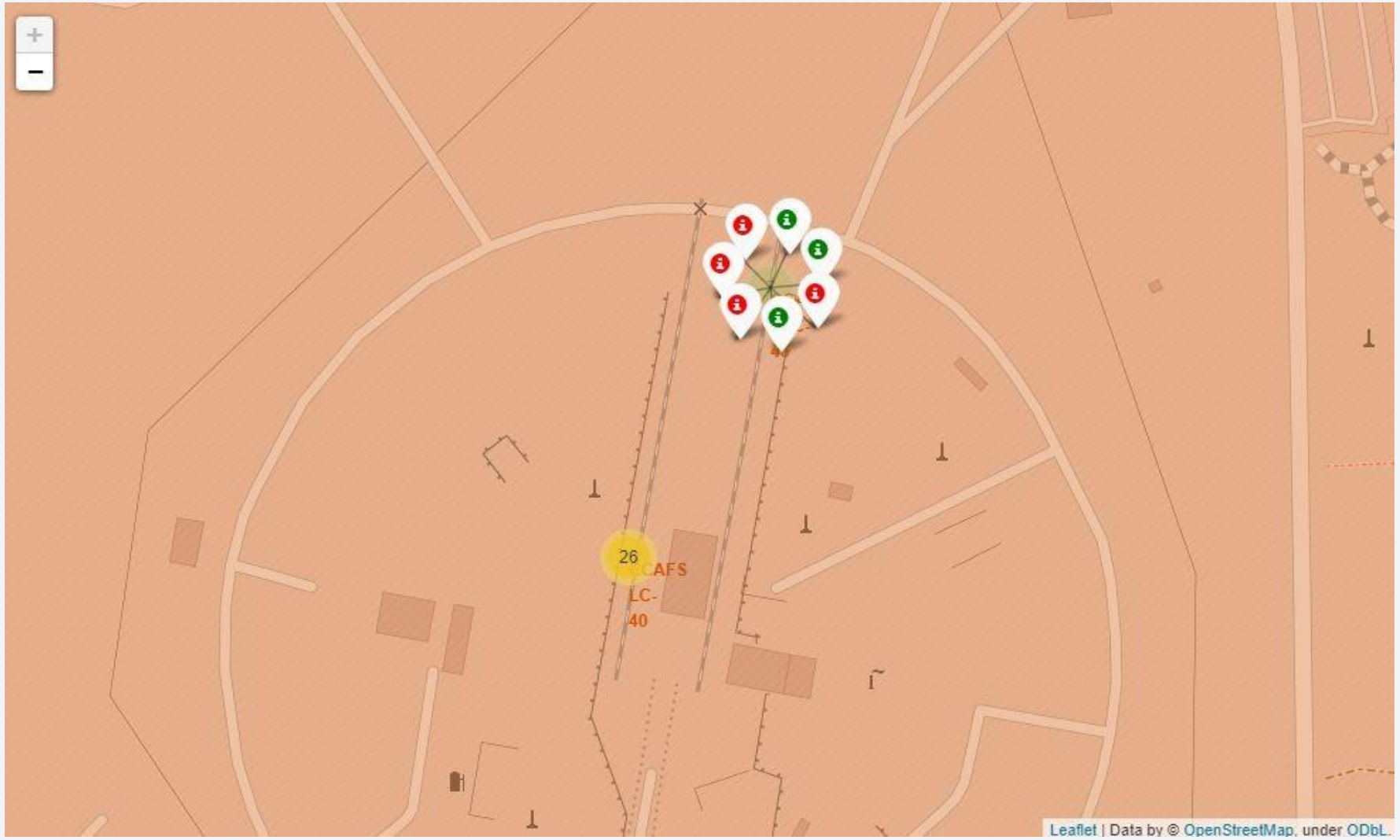
KSC LC-39A Recovery Outcomes



CCAFS LC-40 Recovery Outcomes

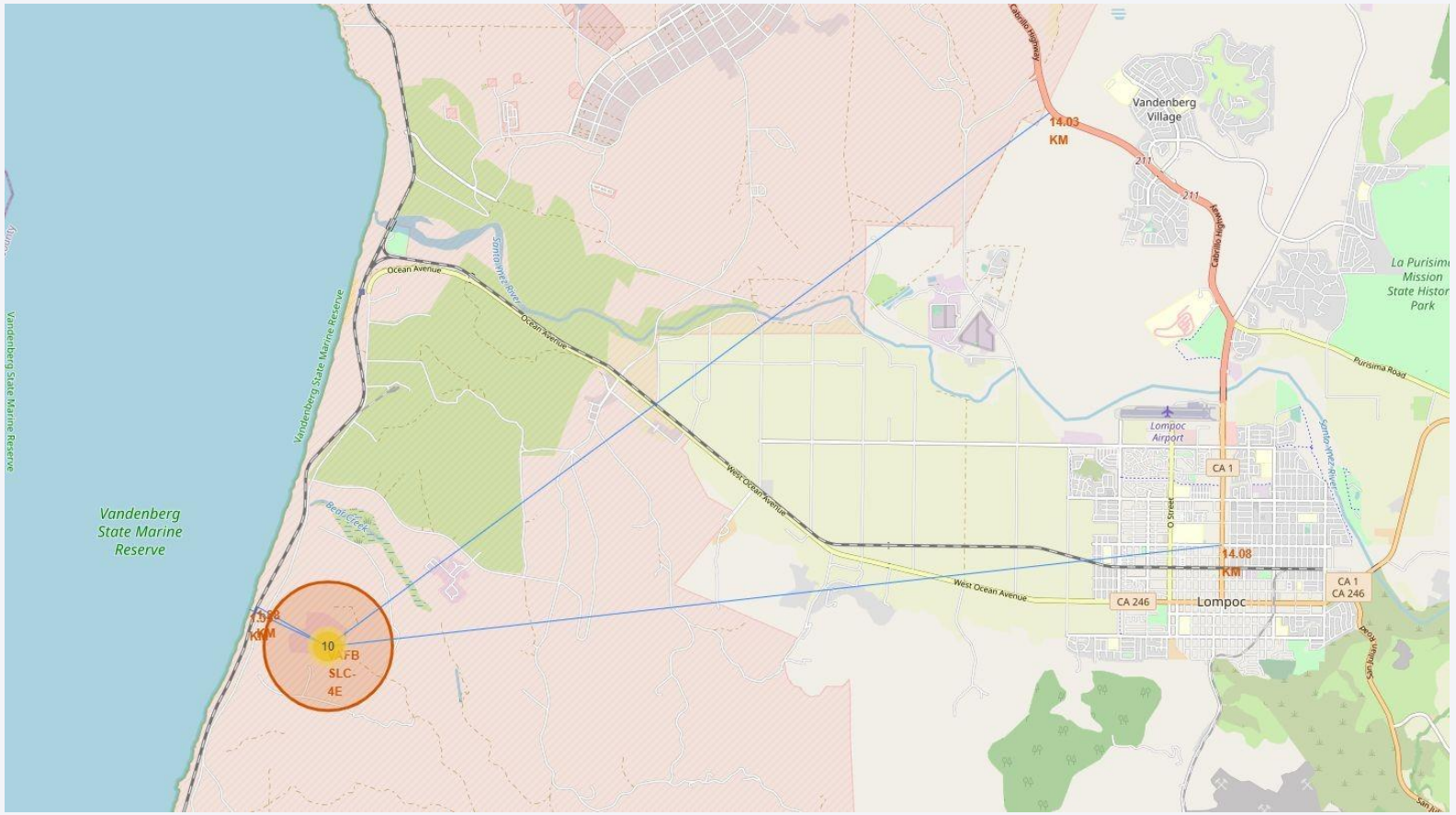


CCAFS SLC-40 Recovery Outcomes



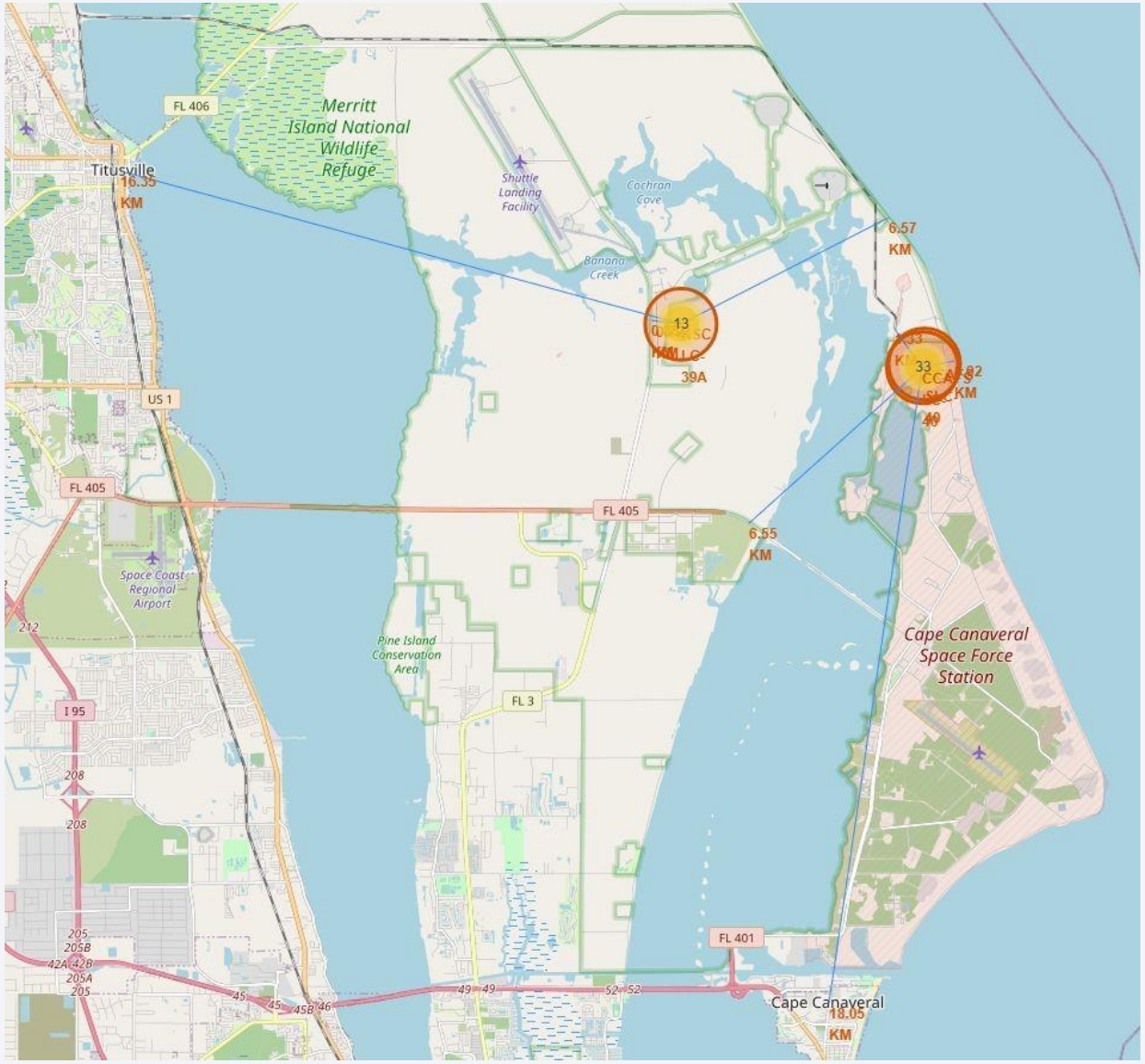
VAFB SLC-4E Nearby Locations

- Each blue line represents the distance to the nearest
  - Coastline
  - City
  - Railway
  - Highway



KSC LC-39A, CCAFS LC-40, and CCAFS SLC-40 Nearby Locations

- Less than 15km distance to the nearest towns
- Launch sites located less than 7km to the coast and surprisingly close to railways and highways

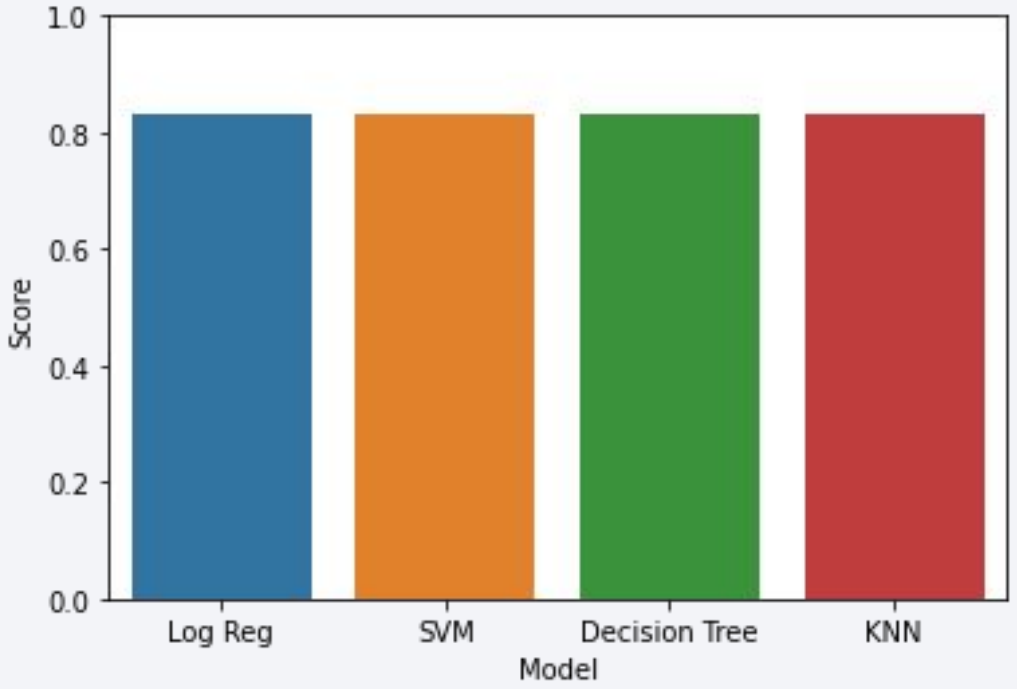


Section 6

Predictive Analysis (Classification)

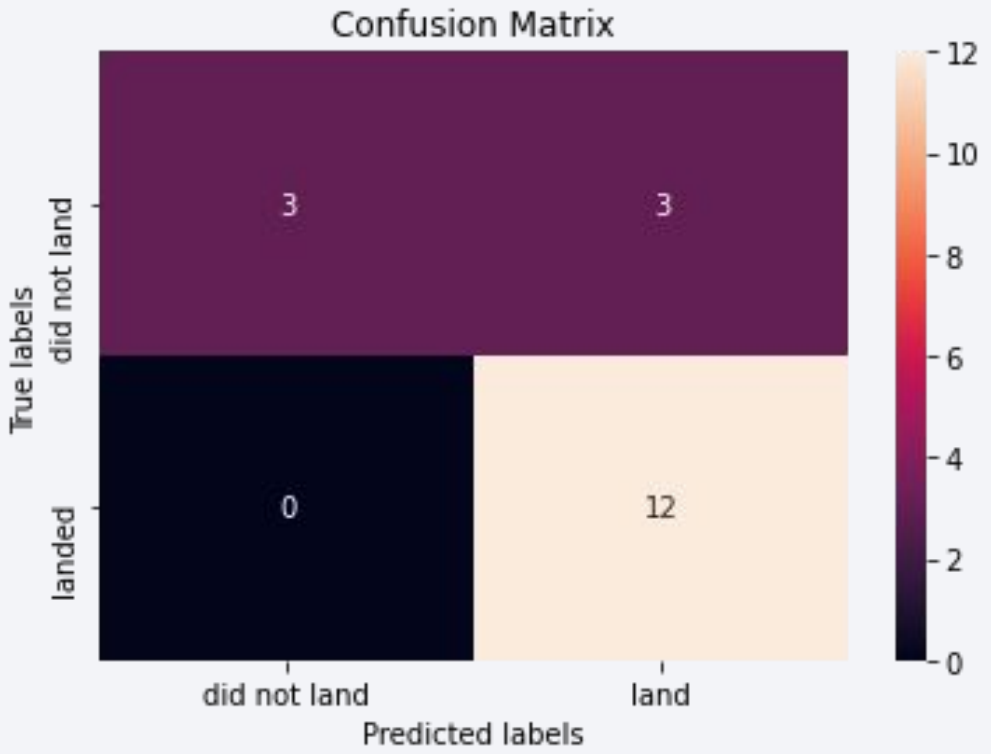
Classification Accuracy

- The bar plot shows the accuracy of the different classification models
- All models have an accuracy rate of 83,33%



Confusion Matrix

- The model had good predictions
- Although it had good predictions, there were some cases in which it could not predict correctly



Conclusions

- From all of the successful recoveries, these properties were the most common ones:
  - A launch date in the year 2017 or later
  - A payload in the range from 2000 up to 4000kg
  - KSC LC-39A launch site
  - Drone ship successful recovery

- The models predict the outcome with a success rate of 83.33%

Thank you!