Falcon 9 First Stage Landings

A Data Science Approach

OUTLINE

Executive Summary

Summary of methodologies

This research aims at exploring past launch data to gain insights for predicting the success rate of first stage landings. The following methodologies were adopted to make this determination:

Data Collection & Wrangling

- Data collection: SpaceX REST API and web scraping from Wikipedia
- Data wrangling: Prepare data for analysis & modeling

Exploratory Data Analysis (EDA)

- EDA with SQL
- EDA with Visualization

Data Visualization

- Interactive Visual Analytics with Folium & Plotly Dash

Modeling

- Predictive Analytics via classification models

Summary of all results

Exploratory Data Analysis:

- The KSC LC 39A was the landing site with the highest success rate
- The following 4 orbits had 100% success rate: ES-l1, GEO, HEO & SSO
- The Launch success rate has been trending upwards since 2013

Visual Analytics:

- All launch sites are close to the coastline as well as the equator

Predictive Analytics:

- The best predictive model for the dataset was the Decision Tree model

INTRODUCTION

Background

SpaceX is a leading developer of reusable rockets for space travel missions with an advertised cost of $62 millions as compared to their competitors' costs of above $165 million. The Falcon 9 first stage which is equipped with landing legs and grid fins, allows it to return to Earth after delivering its payload into space. Hence, our ability to determine whether the first stage will land can be a factor in determining the cost of a launch because reuse provides significant cost savings to SpaceX.

The Goal

The purpose of this project is to analyze data from previous launches to predict the success rate of first stage landings.

The Objectives

- Determine factors that influence successful first stage landing of Falcon 9
- Evaluate lading outcomes over time

METHODOLOGY

- **Data Collection & Wrangling**
- Data collection: SpaceX REST API and web scraping from Wikipedia
- Data wrangling: Prepare data for analysis & modeling
- **Exploratory Data Analysis (EDA)**
- EDA with SQL
- EDA with Visualization
- **Data Visualization**
- Interactive Visual Analytics with Folium & Plotly Dash
- **Modeling**
- Predictive Analytics via classification models

Data Collection: SpaceX API

- The request.get() was deployed via the SpaceX API to obtain the launch data as response
- Response content was decoded via .json() and converted into a Pandas dataframe using .json_normalize()
- Applied SpaceX API to obtain specific launch data stored in a list for creating a new dataframe
- Filtered the dataframe to include only Falcon 9 launches
- Replaced missing Payload Mass values with calculated .mean()
- Export data to csv file

Data Collection: Web Scrapi

- Requested Falcon 9 launch records from Wikipedia
- Created a BeautifulSoup object from the HTML response
- Extracted column names from HTML table header
- Passed the HTML tables and created a dataframe from the dictionary
- Export data to csv file

Data Wrangling

- Performed Exploratory Data Analysis (EDA) and determine the training labels
- Calculated the following:

  No. of launches per site

  No. & occurrence of each orbit

  No. & occurrence of mission outcome of the orbits
- Created a binary landing outcome label from Outcome column
- Outcome column labels:

  successful landing = 1

  unsuccessful landing = 0
- Export results to csv file

Exploratory Data Analysis: Visualization

The following charts were plotted to analyze relationships between some variables:

- Flight No. vs Launch Site
- Payload (kg) vs Launch Site
- Success rate vs Orbit type
- Payload Mass (kg) vs Orbit type

Exploratory Data Analysis: SQL

SQL queries were executed to gain further insights into the SpaceX dataset through the following sets of information displays and lists.

Displays:

- Names of the unique launch sites in the space mission
- Records where launch sites begin with CCA
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1

Lists:

- Date of the first successful landing outcome on the ground pad
- Names of the boosters which successfully landed on drone ship and had payload mass between 4000 and 6000
- Total number of successful and failed mission outcomes
- Failed landing outcomes on drone ship, booster versions & launch site for the months in the year 2015.

Interactive Visual Analytics: Folium

- Marked all launch sites on a map

  Added circle objects based on its coordinate (Latitudes, Longitudes) values.

  Added Launch site name as a popup label
- Marked the outcome of launches for each site on the map

  Added markers to indicate success([green], class=1) and failure([red], class=0)
- Calculated the distances between a launch site(CCAFS SLC-40) to various proximities such as:

  the closest coastline
  the closest railway line
  the closest highway

Interactive Visual Analytics: Plotly Dash

- Created a dropdown list with all launch sites
- Plotted a pie chat to visualize launch success counts as a percent of total
- Configured a Payload Range Slider

  To enable users analyze if variable payloads are correlated with mission outcomes
- Plotted a scatter graph of Payload Mass vs Launch outcome(class):

  Enables users to visualize the correlation between Payload Mass vs Launch outcomes

Predictive Analytics:

Classification

- Created a NumPy array from the column class
- Standardize the data with StandardScaler, fit and then transform it
- Applied the function train_test_split to split the data
- Created a GridSearchCV object with cv=10 to obtain the best parameters
- Applied the GridSearchCV to the following algorithms:

  Logistic regression, Support Vector Machines, Decision Trees, and K-Nearest Neighbor

- Used accuracy as a metric on the test data with .score() for all the models
- Determined the best classification model

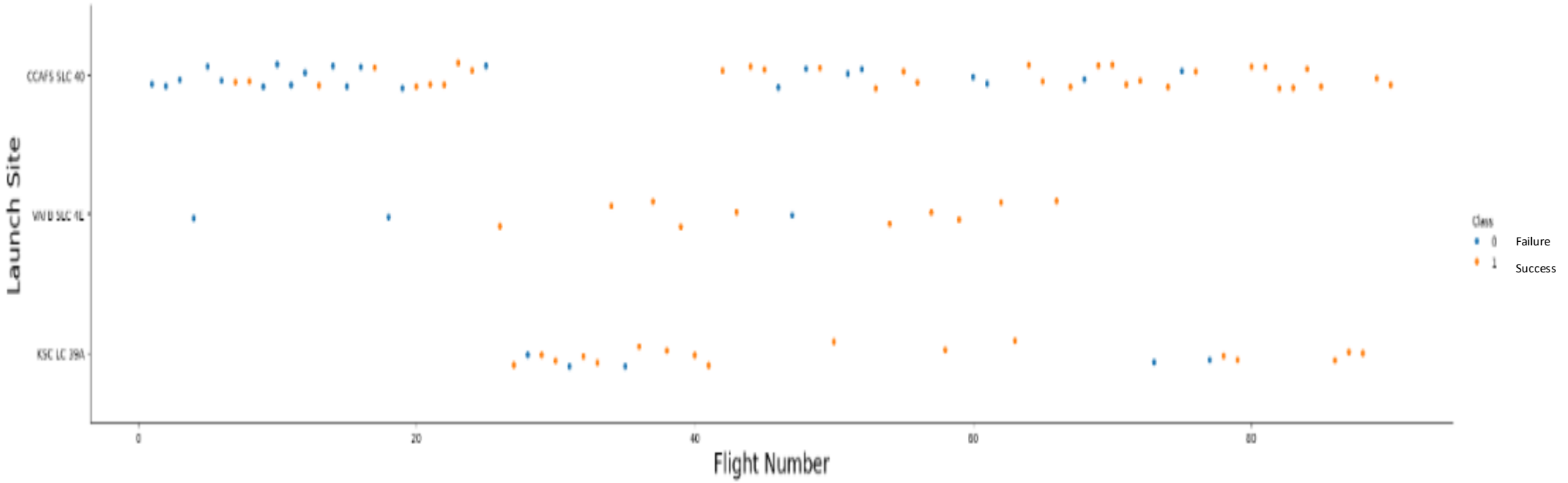
Results

- Exploratory Data Analysis (EDA)
- Interactive Visual Analytics
- Predictive Analytics

Exploratory Data Analysis (EDA)

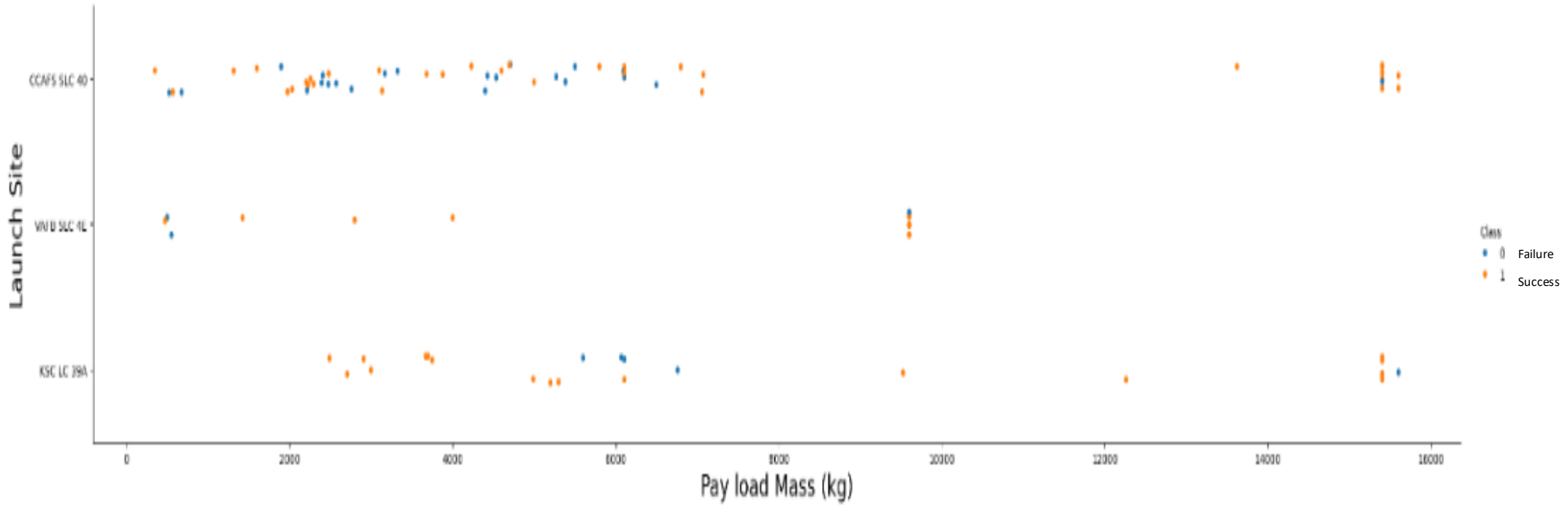
Flight Number vs Launch Site(EDA)

- The highest success rates was recorded by KSC LC 39A and VAFB SLC 4E
- CCAFS SLC 40 recorded the highest number of launches



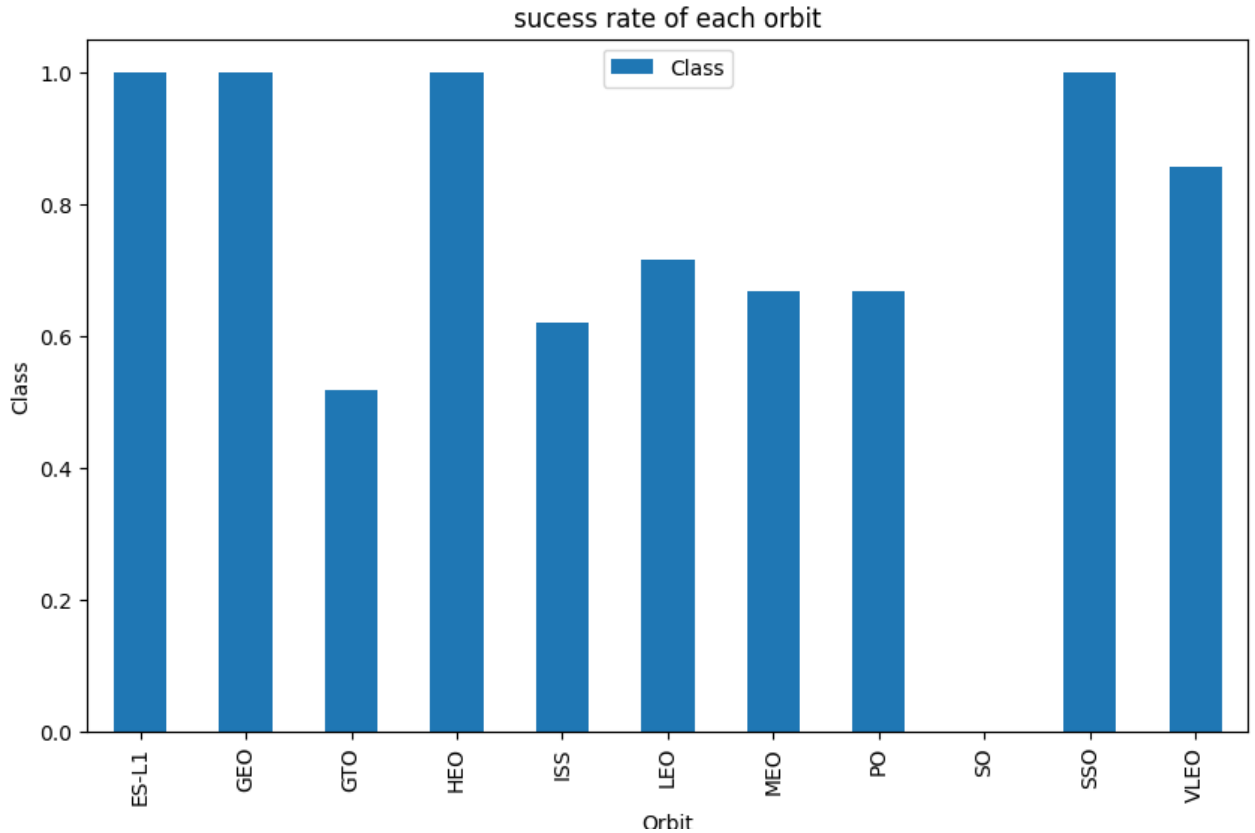
Payload vs Launch Site(EDA)

- Higher success rates have been associated with Payload Masses above 8,000kg
- All Pay load masses below 5,000 kg were successful launches from KSC LC 39A
- No launches were recorded for Pay load Mass rages of between 7,000 kg and 9,000kg



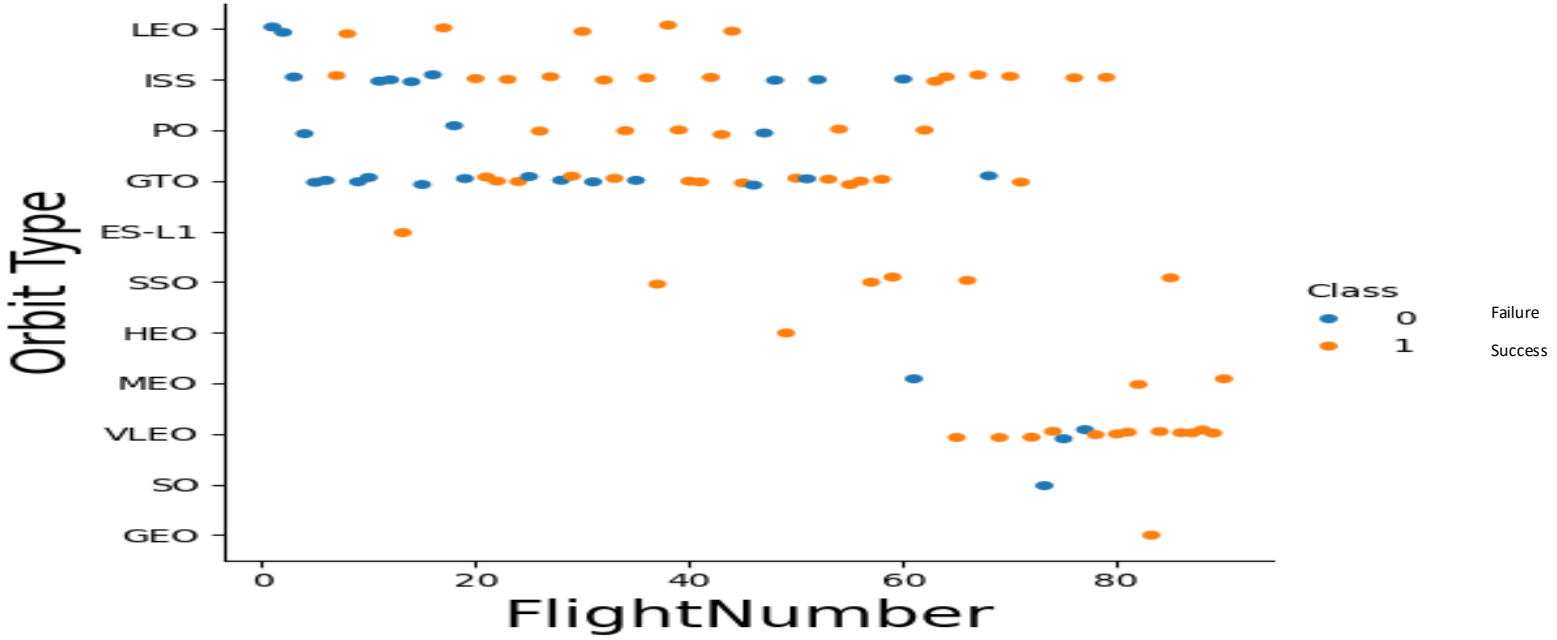
Success Rate vs Orbit Type (EDA)

- The following Orbits had 100% success rates: ES-L1, GEO, HEO and SSO
- The SO Orbits had zero success rate



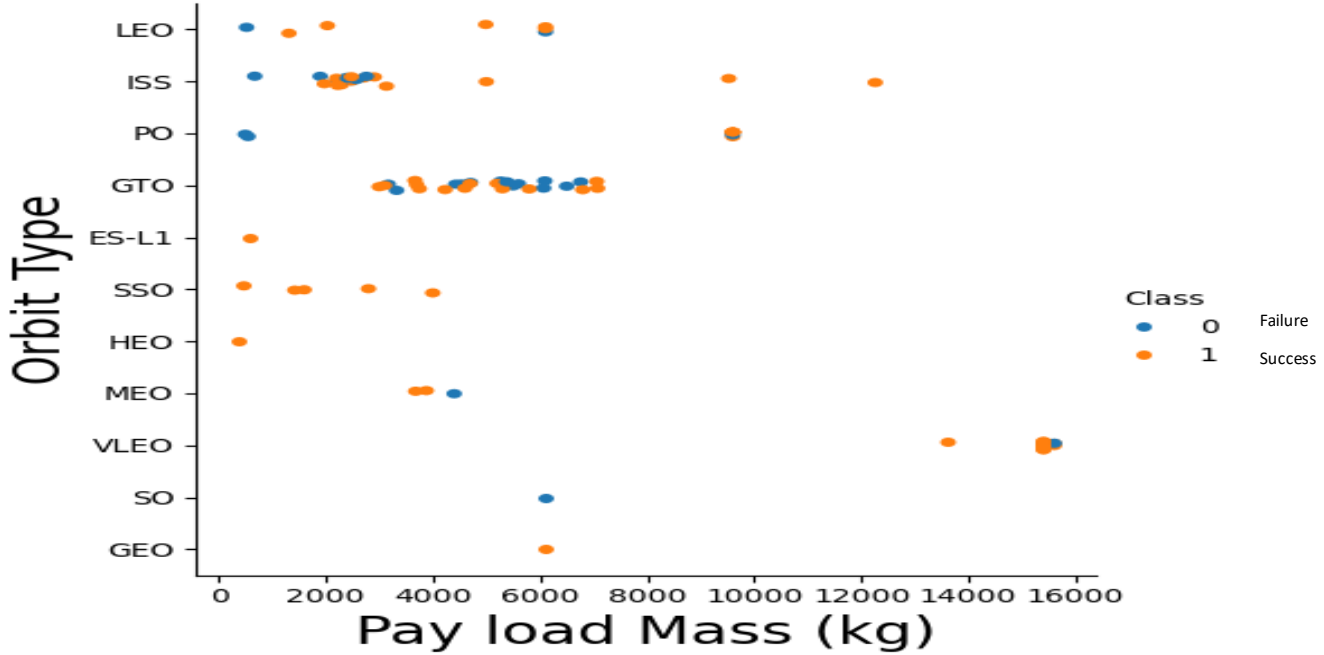
Flight Number vs Orbit Type(EDA)

- The success rate of Orbits appear to increase for higher flight numbers as shown by the LEO orbit
- The GTO orbit is however an exception to the trend and thus shows no increasing relationship with the flight number.



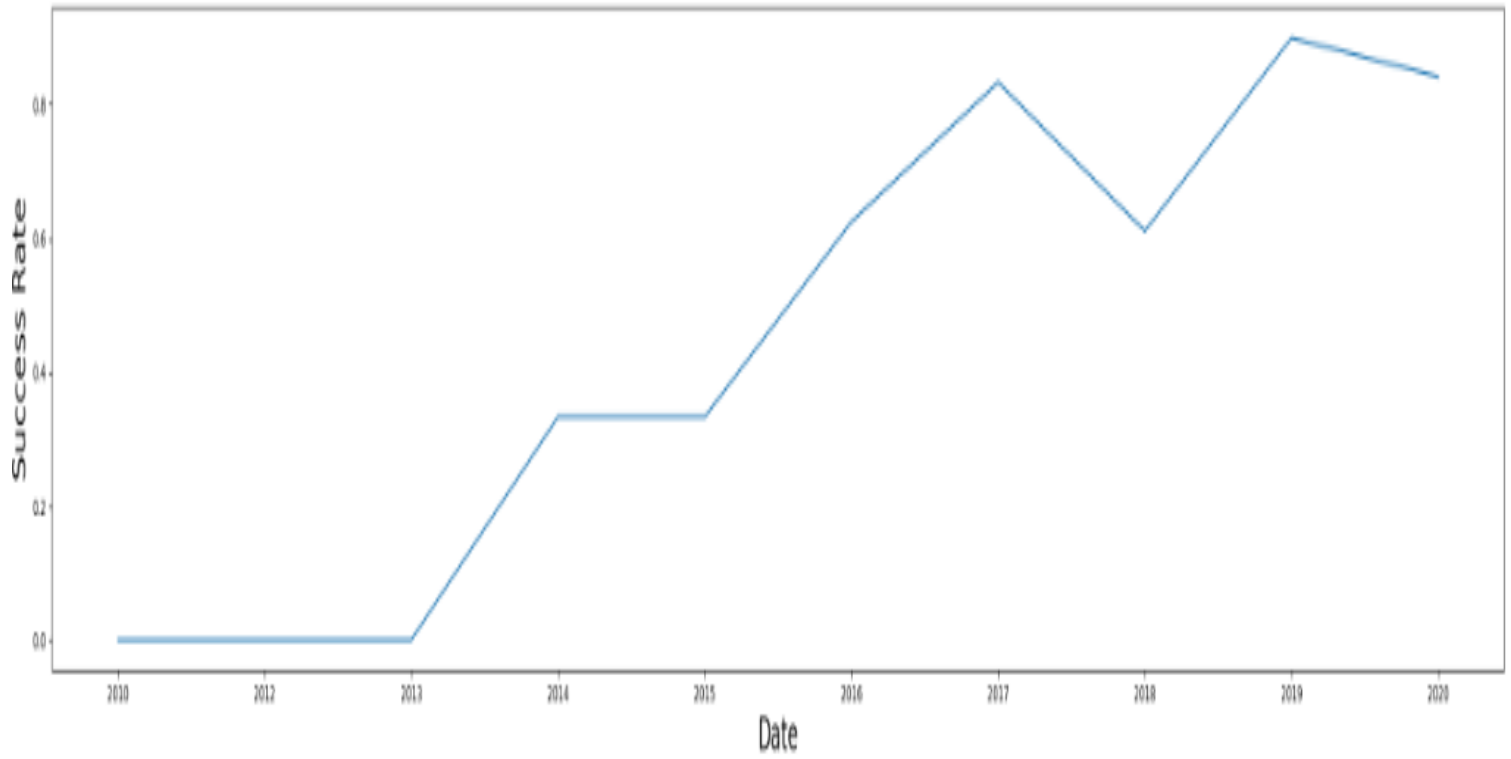
Payload vs Orbit Type(EDA)

The orbits LEO, ISS and PO have had successful landing outcomes with higher Payloads



Yearly Success Trend (EDA)

The success rate has seen upward trend since 2013



Launch Site Information

Launch Site Names

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

Launch Site Names Starting with CCA

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO
2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)

Payload Mass

Total Payload carried by NASA boosters

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass carried by booster F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

Landing & Mission Outcome Details

First successful Ground Landing Date was December 22nd 2015

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with payload mass between 4000 and 6000

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

Total Number of Successful and Failed Mission Outcomes

Mission_Outcome	total_number
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Booster versions with Maximum Payload

Names of the booster versions 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

The list of failed landing outcomes in drone ship with their booster versions and launch site names for 2015

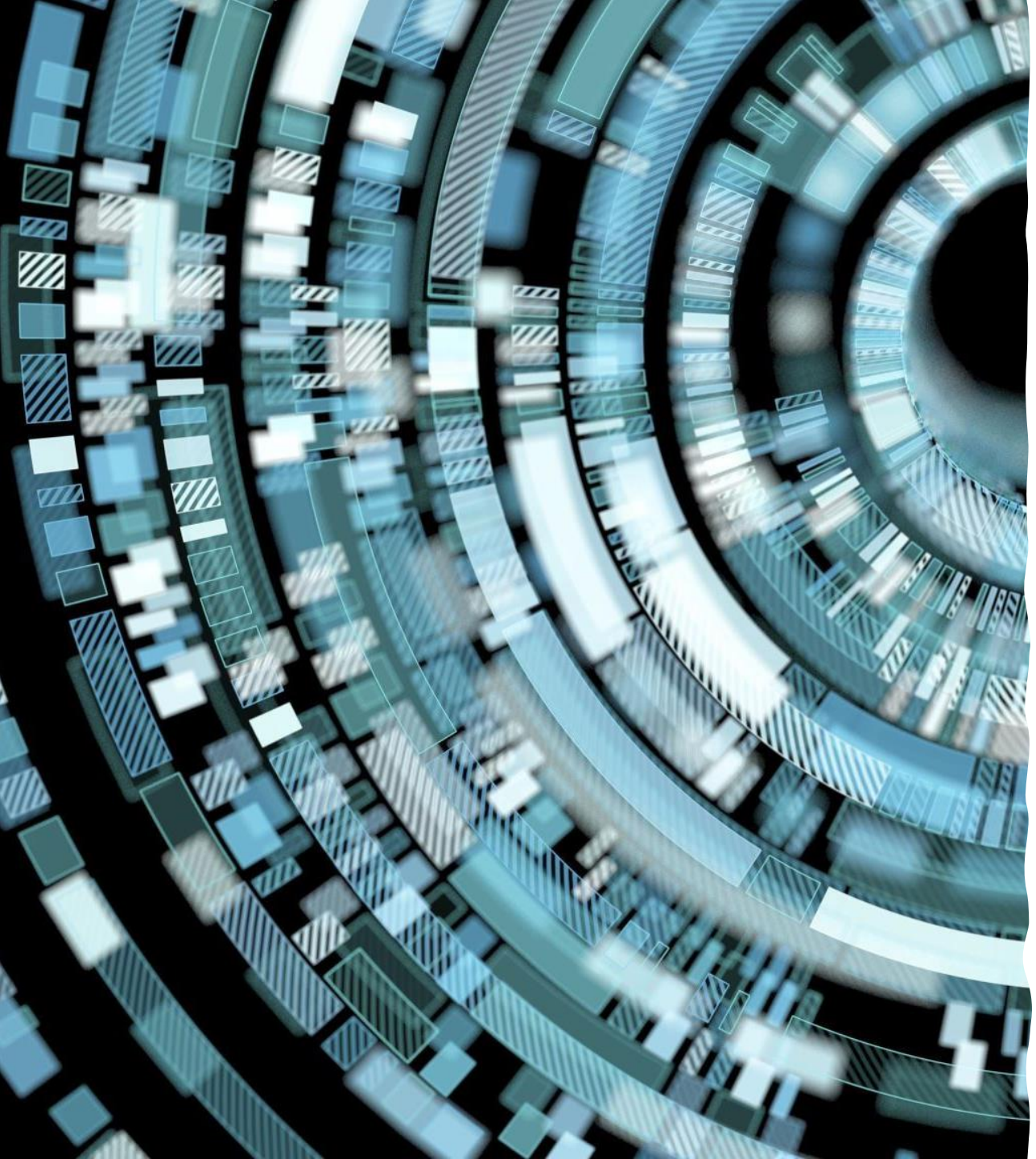
month	Date	Booster_Version	Launch_Site	Landing_Outcome
10	2015-10-01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Rank of Landing Outcomes between 2010-2017

The Rank of landing outcomes (failed drone ship or Success ground pad) between June 2010 and March 2017

Landing_Outcome	COUNT_LAUNCHES
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

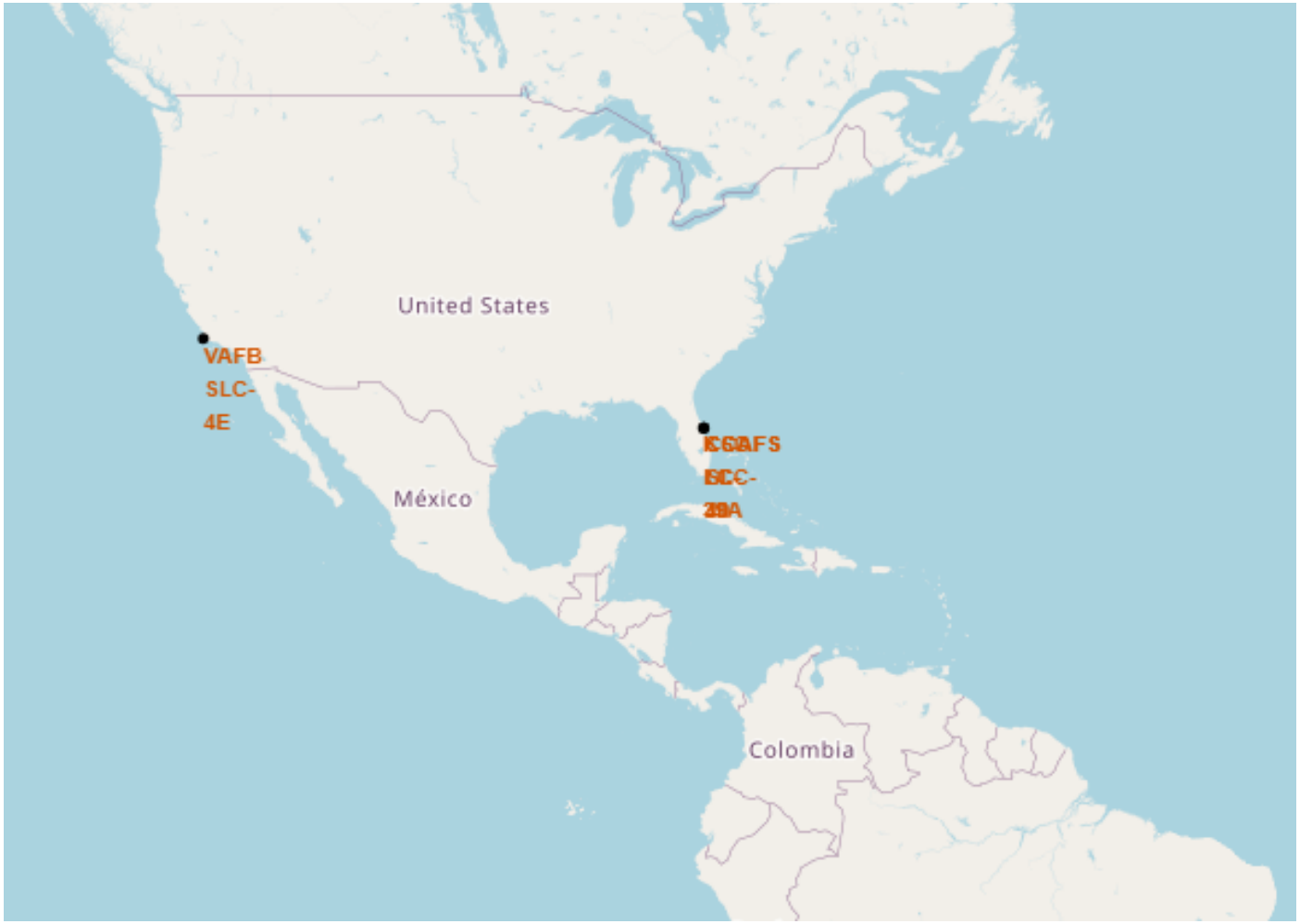
- The landing outcome for ground pad was the most successful (100%)
- Landing outcome for drone ship was mixed(50%)



Interactive Visual Analytics

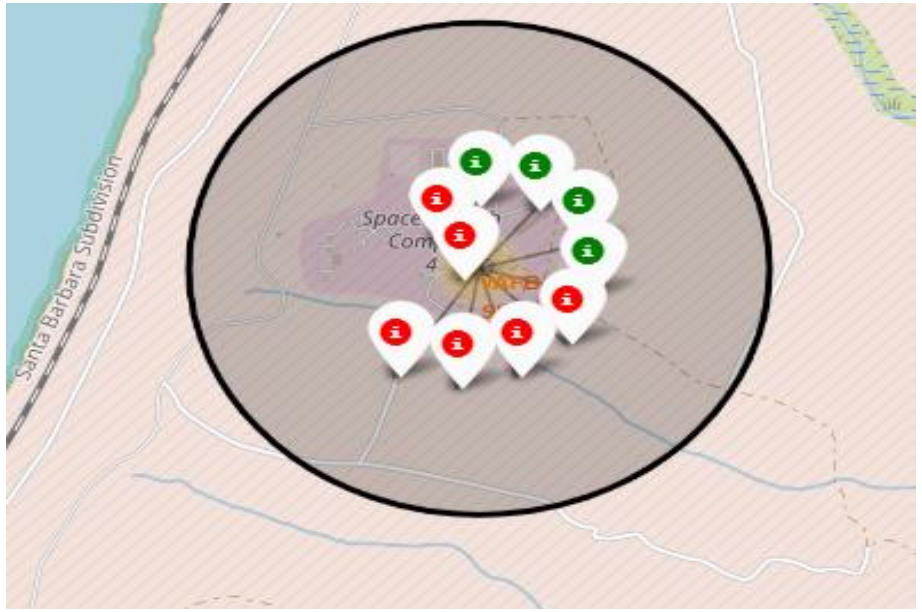
Map of all SpaceX Launch Sites

All SpaceX launch sites are located in the United States of America at the coasts of California and Florida.



Markers indicating Launch sites

Florida Launch Sites

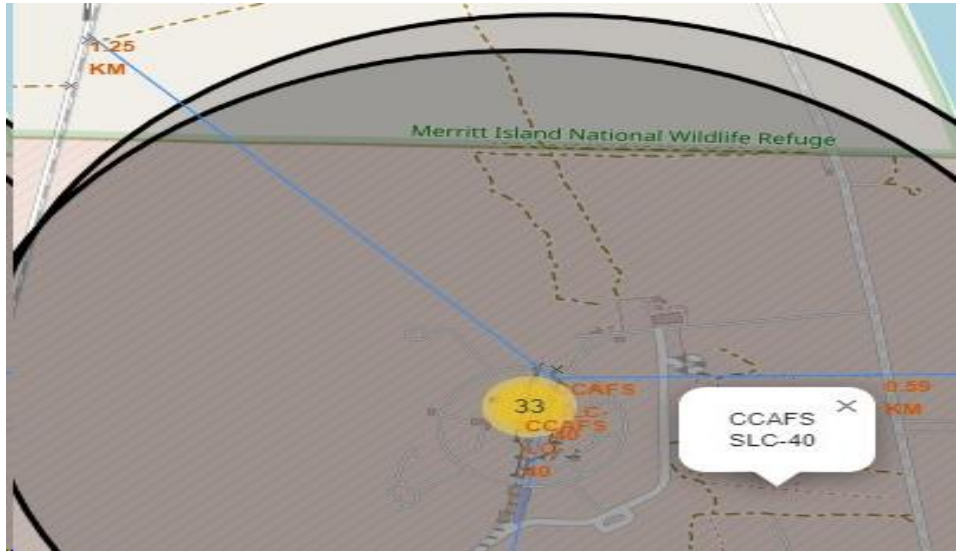


California Launch Site

Green Markers: Successful Launches

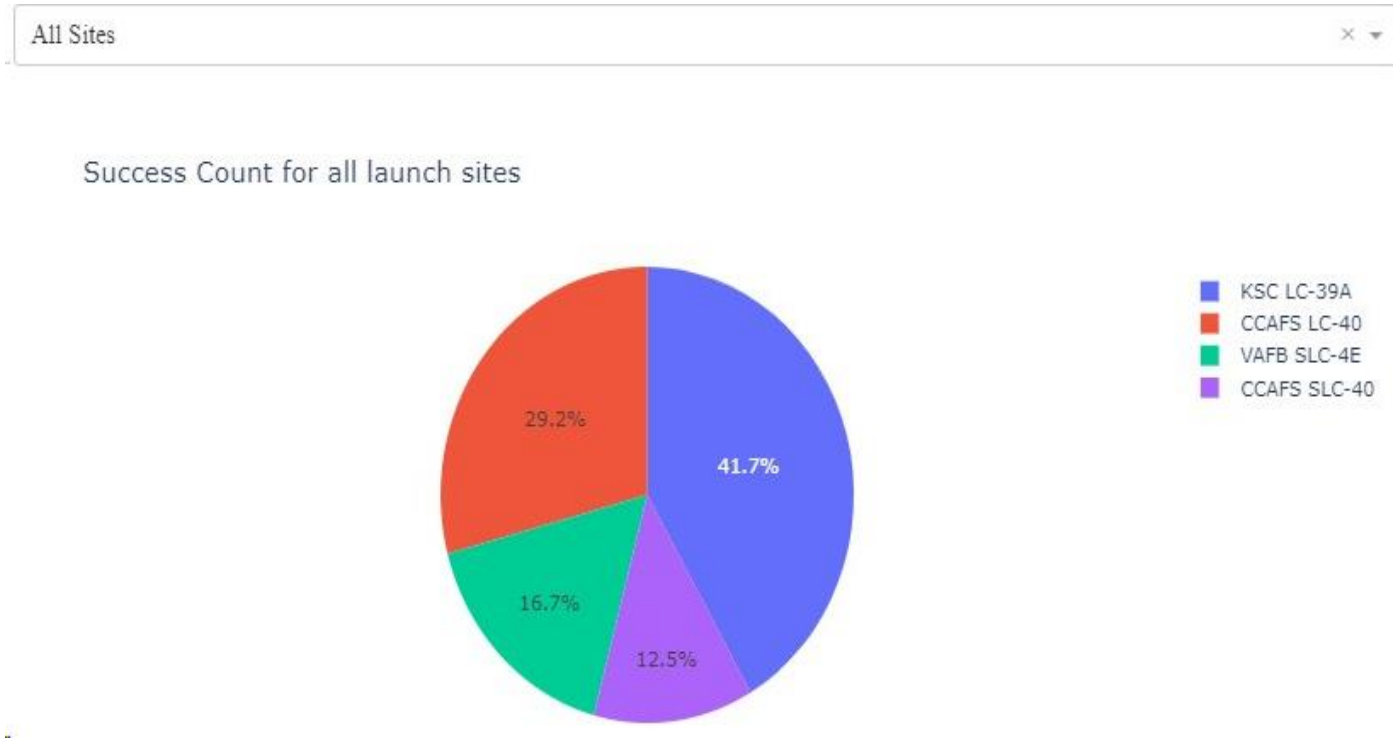
Red Markers: Failed Launches

Launch site Proximities

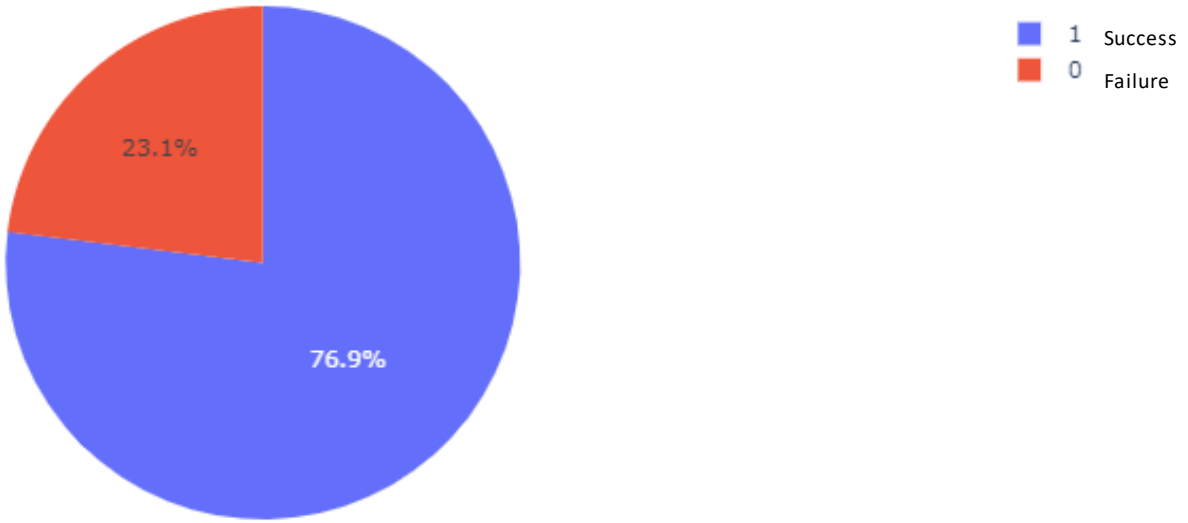


- Launch sites are close to the coast and near the equator as well
- Launch sites are also far from cities as well as major infrastructure such highways and railway lines

Pie chart of Success for all launch sites



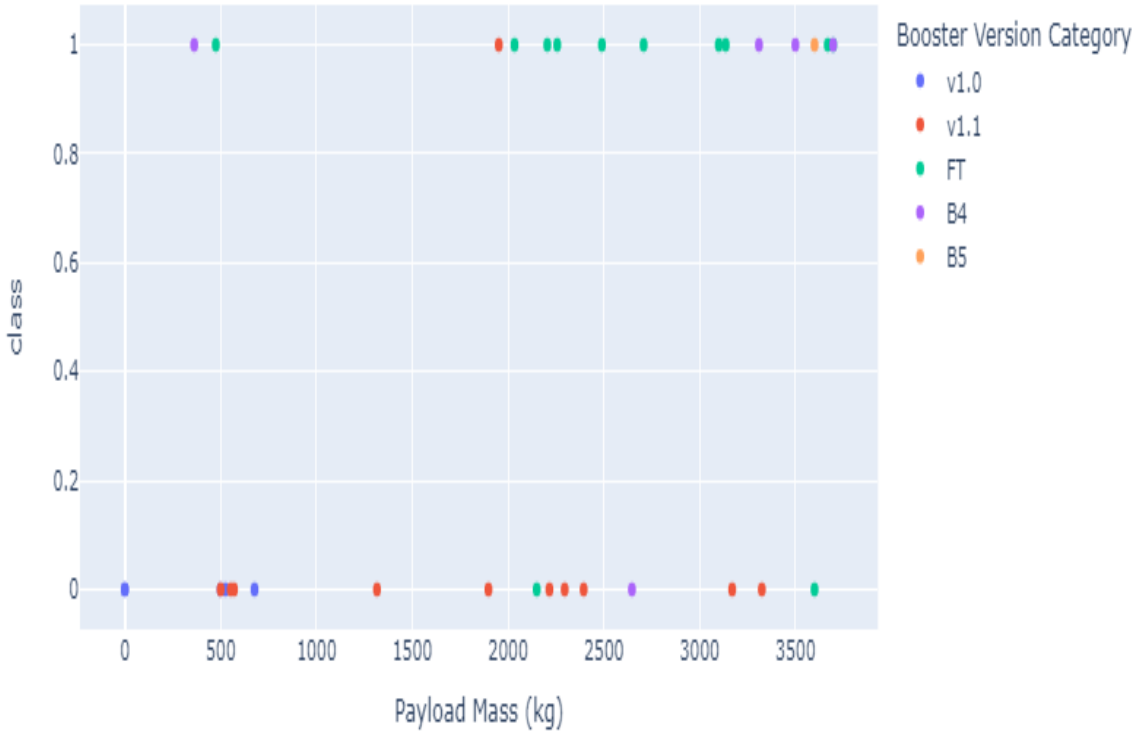
Pie Chart for Launch Site with highest Success rate

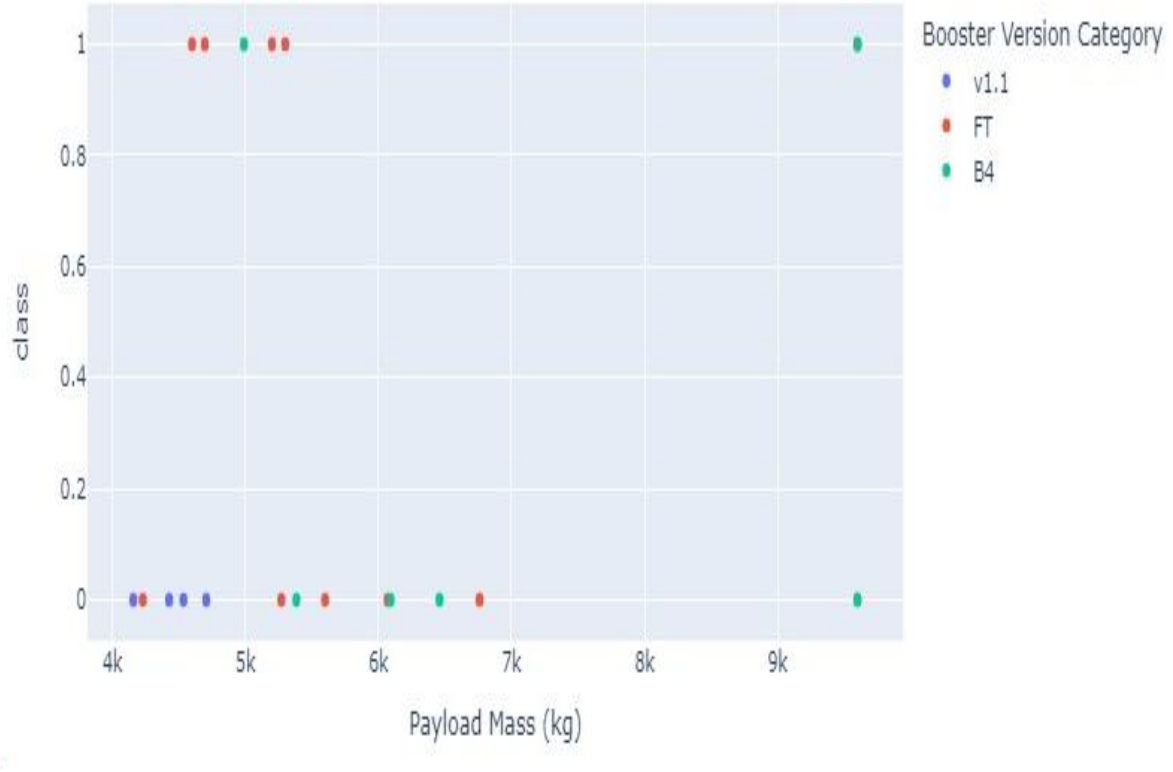


KSC-LC-39A has the highest success rate 76.9%

Lowest & highest success rate per payload ranges

- Success rates are higher for payload masses below 4,000kg
- Booster Version FT appears to have been successful in lower payloads ranges

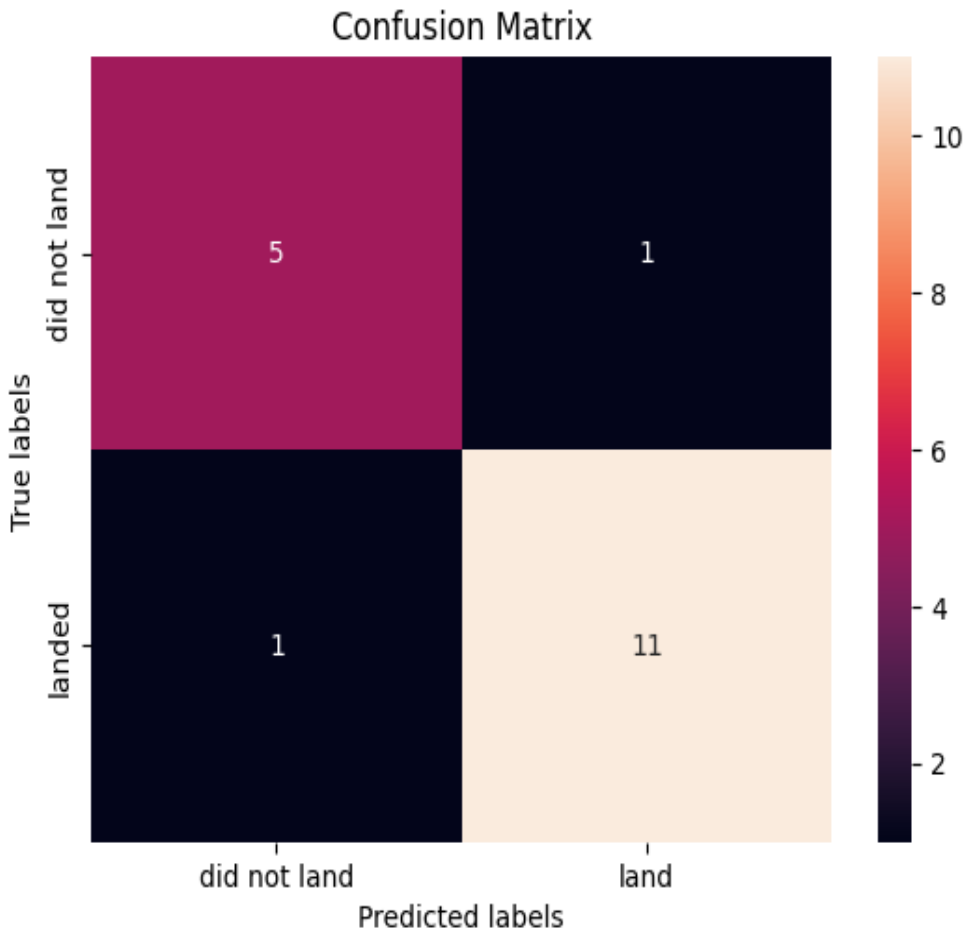




Predictive Analytics

Classifications

Confusion Matrix



Decision Tree

The confusion matrix here indicates that this model can distinguish between the two classes.

The Accuracy score is 88%

Classification Accuracy

The results from the four models has confirmed the Decision Tree as the model with the highest classification accuracy

```
print('Accuracy for Logistics Regression method:', logreg_cv.score(X_test, Y_test))
print( 'Accuracy for Support Vector Machine method:', svm_cv.score(X_test, Y_test))
print('Accuracy for Decision tree method:', tree_cv.score(X_test, Y_test))
print('Accuracy for K nearsdt neighbors method:', knn_cv.score(X_test, Y_test))
```

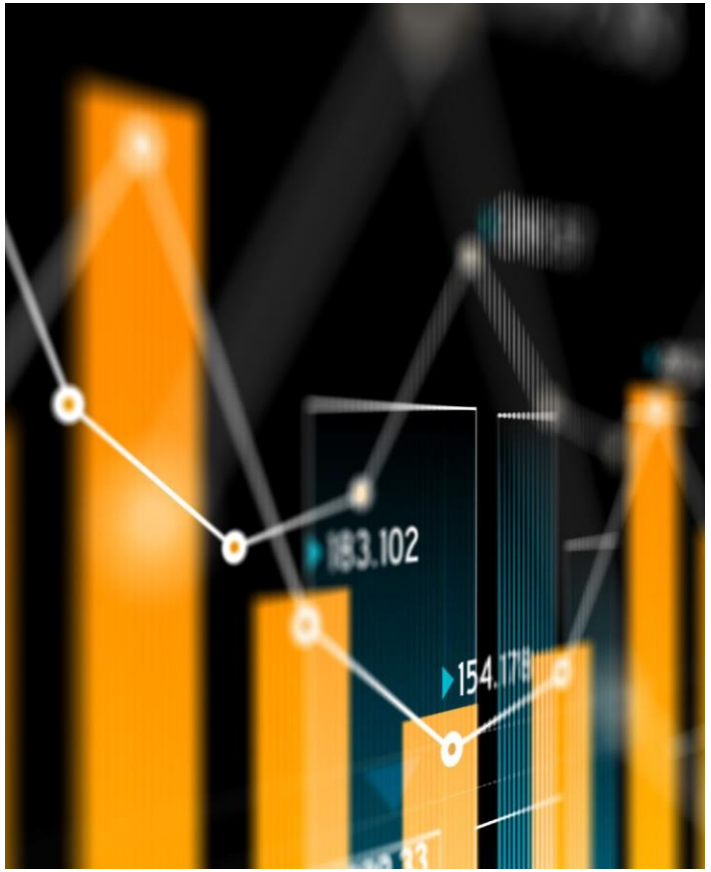
```
Accuracy for Logistics Regression method: 0.8333333333333334
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.8888888888888888
Accuracy for K nearsdt neighbors method: 0.8333333333333334
```

Conclusions

- The following insights were obtained from the dataset:
- The KSC LC 39A was the landing site with the highest success rate
- The Ground pad had the highest landing outcome (100%) between 2010 and 2017
- The following four orbits had 100% success rate: ES-l1, GEO, HEO & SSO
- The Launch success rate has been trending upwards since 2013
- Most launch sites are close to the coastline as well as the equator
- The best predictive model for the dataset was the Decision Tree model

The research has identified some of the factors which can influence the successful landing/reuse of rockets as:

- The launch facility (KSC LC-39A)
- The location of launch sites (proximity to coast and the equator)
- The type of Orbit
- Launch success has improved over time which could be due to either experience or economies of scale



APPENDIX