SpaceX Falcon 9 First Stage Landing Prediction

Tan Jie Qi

25 June 2023

Table of Contents

Executive Summary

This report aims to predict the likelihood of a successful rocket landing, which determines SpaceX's Falcon 9 launch price. To make this prediction, the following methodologies were applied:

- **Collect** data using SpaceX API and web scraping techniques
- **Data wrangling** to clean up data
- Analyze and explore data using **SQL queries** as well as **data visualization** to identify potential relationships between various variables vis-à-vis landing outcome
- Visualize launch site success rates and geographical location and proximities using **Folium map**
- Build **machine learning models** to predict landing outcomes

Results

- Overall, success rate of Falcon 9 first stage landing has improved over the years.
- Success rate is typically higher with higher payload mass across the launch sites.
- Launch sites are generally closer to coastline, railway and highways, but far away from cities to minimize potential damage.
- Launch Site KSC LC-39A has the highest likelihood of first stage recovery, with 100% success rate for launches with payload mass below 5,600kg.

Introduction

Background

SpaceX, a leader in the space industry, advertises Falcon 9 rocket launches with a cost of $62mil, approximately ***1/3*** of other providers with cost upward of $165mil. This is mainly due to SpaceX's reusability of the first stage landing.

Therefore, by determining if the first stage will land, we can determine the cost of a launch.

Objective

In this report, we use public data and machine learning models to **predict whether Falcon 9 first stage will land successfully**.

Methodology

1. Data Collection – API

- Request rocket launch data from SpaceX API with following URL: *https://api.spacexdata.com/v4/launches/past*
- Decode the response content as Json using *.json()* and turn it into a Pandas dataframe using *.json_normalize()*
- Filter the dataframe using the "*BoosterVersion*" column to only keep the Falcon 9 launches
- Replace missing values of Payload Mass with the calculated mean value
- Export data to CSV file

Methodology

2. Data Collection – Web Scraping

- Perform web scraping to collect Falcon 9 historical launch records from Wikipedia using *BeautifulSoup*
- Extract column names from HTML table header
- Create Pandas dataframe by parsing the HTML tables
- Export data to CSV file

Methodology

3. Data Wrangling

- Identify missing values in each attribute – decision to leave missing values in Landing Pad column as-is
- Perform EDA – Calculate:
  - Number of launches on each Launch Site
  - Number and occurrence of each orbit
  - Number and occurrence of mission outcome per orbit type
- Create a binary "Class" variable based on "Outcome" column
- Export data to CSV file

Methodology

4. EDA with SQL

Queries run include:

- Names of unique Launch Sites
- Date where successful landing outcome in drone ship was first achieved
- Names of boosters which have success in ground pad and payload mass between 4,000 and 6,000 kg
- Total number of successful and failure mission outcomes
- Names of Booster Version which have carried the maximum payload mass

Methodology

5. EDA with Data Visualization

Plot **scatter plots** to visualize relationship between different variables and the launch outcome

Plot **bar chart** to visualize and identify Orbit Types with high launch success rate

Plot **line chart** to observe trend of launch success rate over the years

Methodology

6. Build Interative Map with Folium

- Create Folium map
- Mark all Launch Sites
- Add coloured markers of Launch Outcomes – **green** markers for successful and **red** for unsuccessful launches at each launch site to show which launch sites have high success rates
- Add coloured lines to indicate distance of launch site from the nearest coastline, railway, highway and city

Methodology

7. Dashboard with Plotly Dash

Build an interactive dashboard with Plotly Dash with following features:

- Dropdown list to allow users to select all or specific Launch Sites
- Pie chart showing success rate of launches based on selected Launch Sites
- Range slider to select Payload Mass range
- Scatter chart showing payload mass vs success rate

Methodology

8. Predictive Analysis

- Create NumPy array from Class column
- **Standardise data** with *.StandardScaler()*
- **Split data** using *train_test_split* function (parameter test_size set to 0.2 and random_state to 2)
- Create a GridSearchCV object with cv = 10 and apply on each of the following algorithm method/method:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree Classifier
  - K-Nearest Neighbour
- Evaluate the **accuracy score** and plot **confusion matrix** for each model to **identify the best model**

Results – Exploratory Data Analysis

Launch Site Names:

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

5 records for Launch Site 'KSC%'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
19/02/2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
16/03/2017	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600.0	GTO	EchoStar	Success	No attempt
30/03/2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300.0	GTO	SES	Success	Success (drone ship)
05/01/2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300.0	LEO	NRO	Success	Success (ground pad)
15/05/2017	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070.0	GTO	Inmarsat	Success	No attempt

Total Payload Mass carried by Boosters launched by NASA (CRS): 45,569 kg

Average Payload mass carried by Booster version F9 v1.1: 2,928.4 kg

1st Successful landing in Drone Ship: 4 Aug 2016

Results – Exploratory Data Analysis

Names of Boosters with success in Ground Pad and Payload mass between 4,000kg and 6,000kg:

Booster_Version
F9 FT B1032.1
F9 B4 B1040.1
F9 B4 B1043.1

Total number of successful and failure mission outcomes:

Mission_Outcome	COUNT(MISSION_OUTCOME)
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Results – Exploratory Data Analysis

Names of 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

Results – Exploratory Data Analysis

Records of Booster Version, Launch Site and month for successful landing outcome in Ground Pad in 2017:

Booster_Version	Launch_Site	substr(Date, 4, 2)
F9 FT B1031.1	KSC LC-39A	02
F9 FT B1032.1	KSC LC-39A	01
F9 FT B1035.1	KSC LC-39A	03
F9 B4 B1039.1	KSC LC-39A	08
F9 B4 B1040.1	KSC LC-39A	07
F9 FT B1035.2	CCAFS SLC-40	12

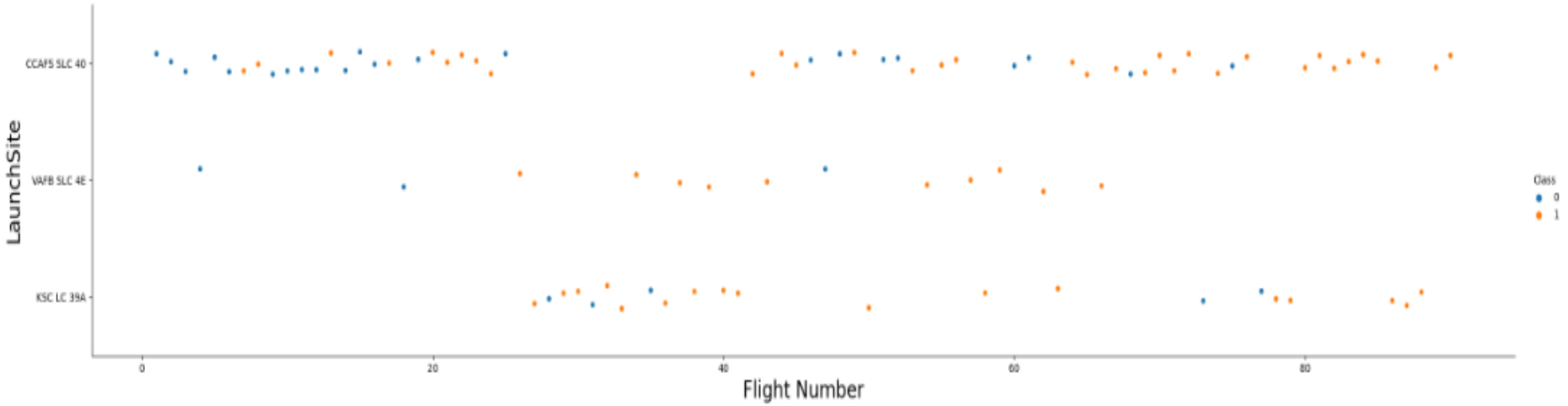
Results – Exploratory Data Analysis

Count of landing outcomes between 2010-06-04 and 2017-3-20:

Landing _Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

Results – Exploratory Data Analysis

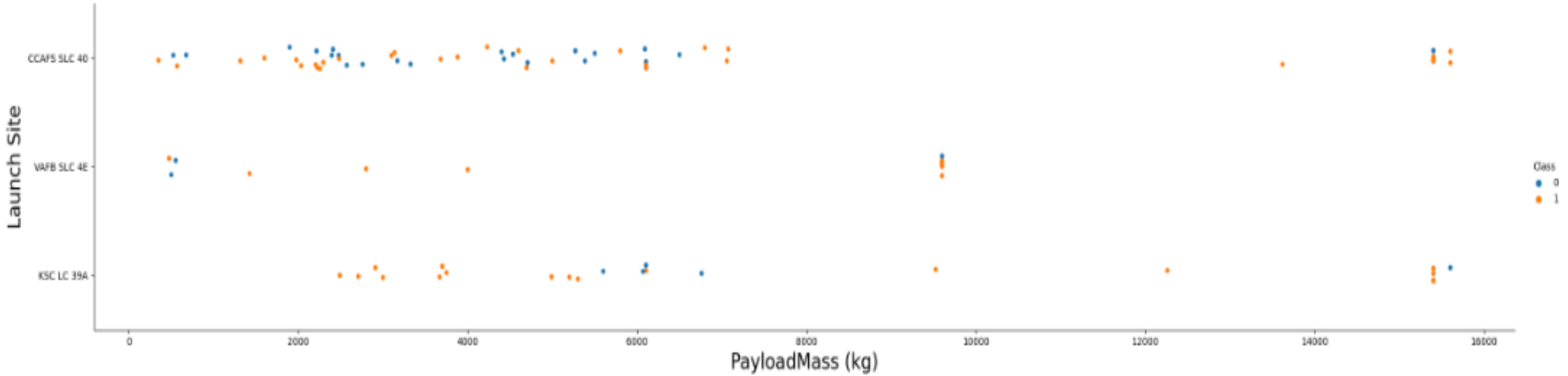
Flight Number vs Launch Site



- CCAFS SLC 40 has the highest number of flights. Nonetheless, it has the lowest success rate.
- As the flight number increases in each of the 3 Launch Sites, so does the success rate.

Results – Exploratory Data Analysis

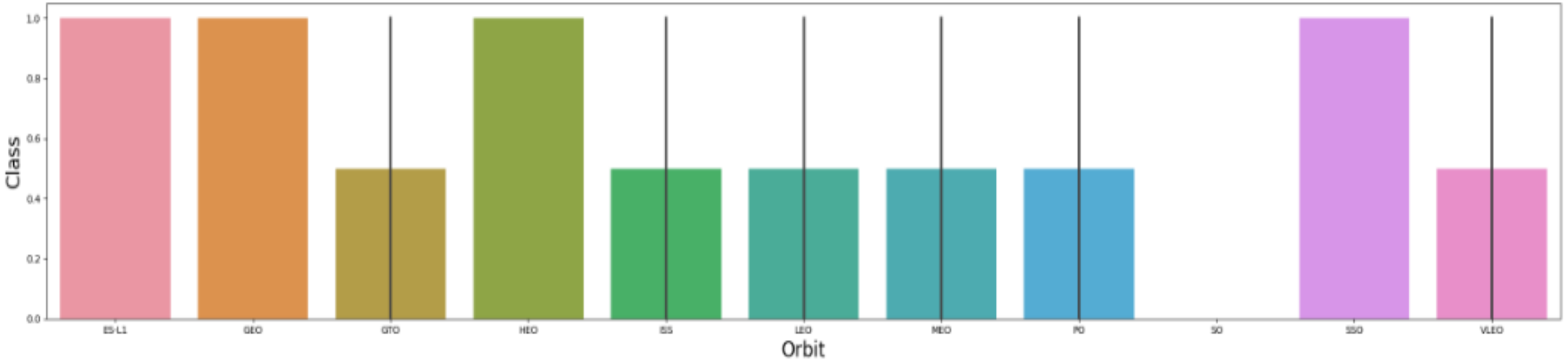
Payload vs Launch Site



- Typically, the higher the Payload Mass, the higher the success rate.
- Most launches with a Payload Mass greater than 8,000 kg were successful.
- KSC LC 39A has a 100% success rate for launches with Payload Mass below 5,600 kg.

Results – Exploratory Data Analysis

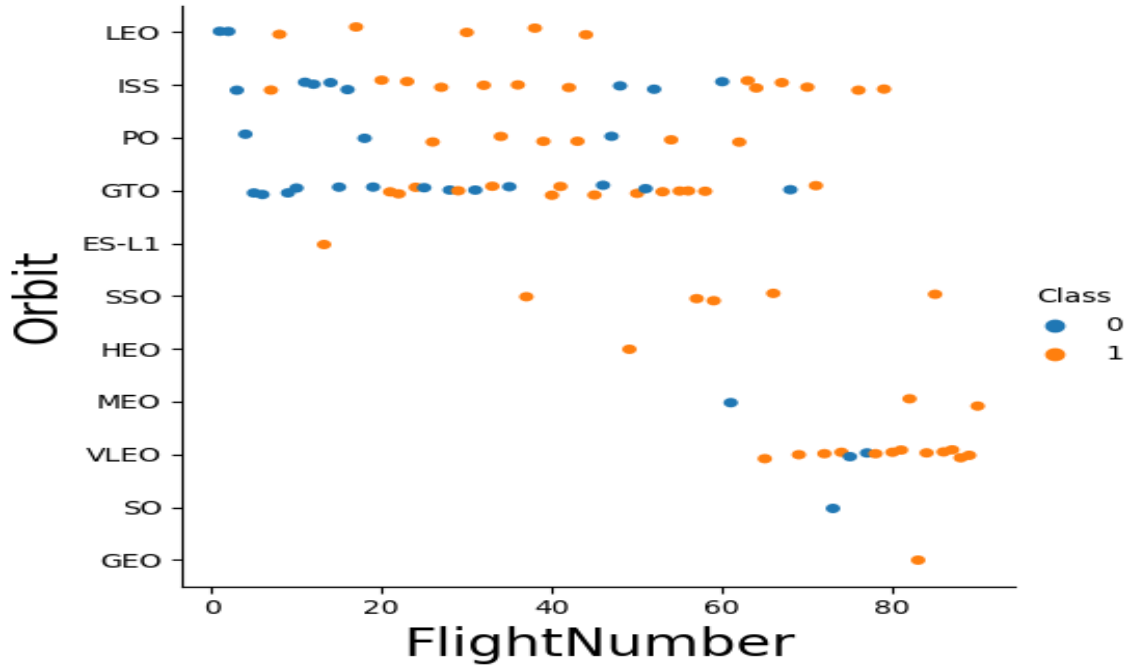
Success Rate by Orbit



- Success rate is 100% for the following orbits: **ES-L1**, **GEO**, **HEO** and **SSO**
- **SO** has lowest success rate of 0%
- The rest of the orbits have success rates between 40% to 50%

Results – Exploratory Data Analysis

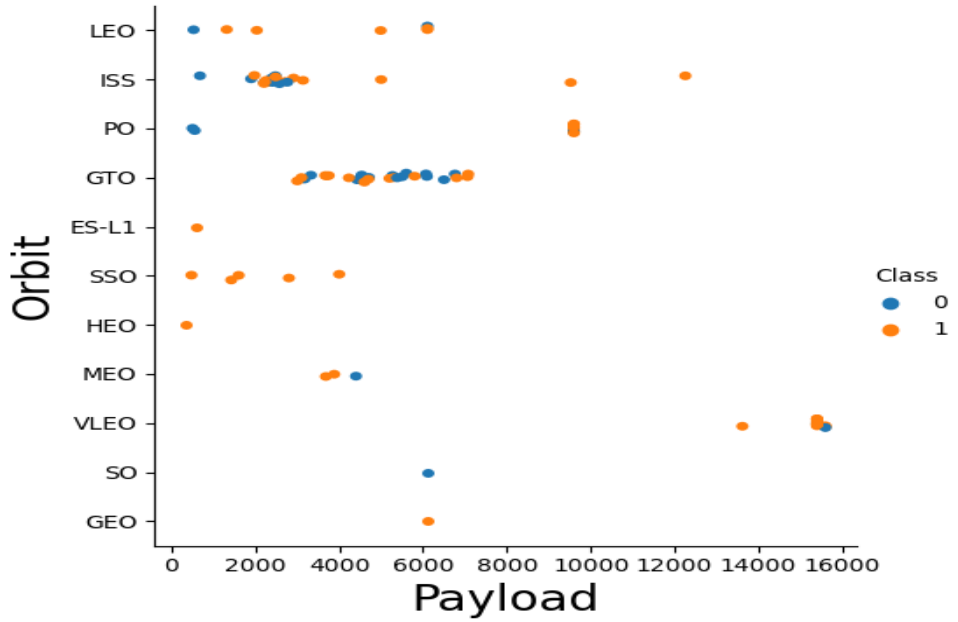
Flight Number vs Orbit



- Success rate typically increases with the number of flights for each orbit.
- This is especially the case for **LEO** orbit, but not so much the case for **GTO** orbit.

Results – Exploratory Data Analysis

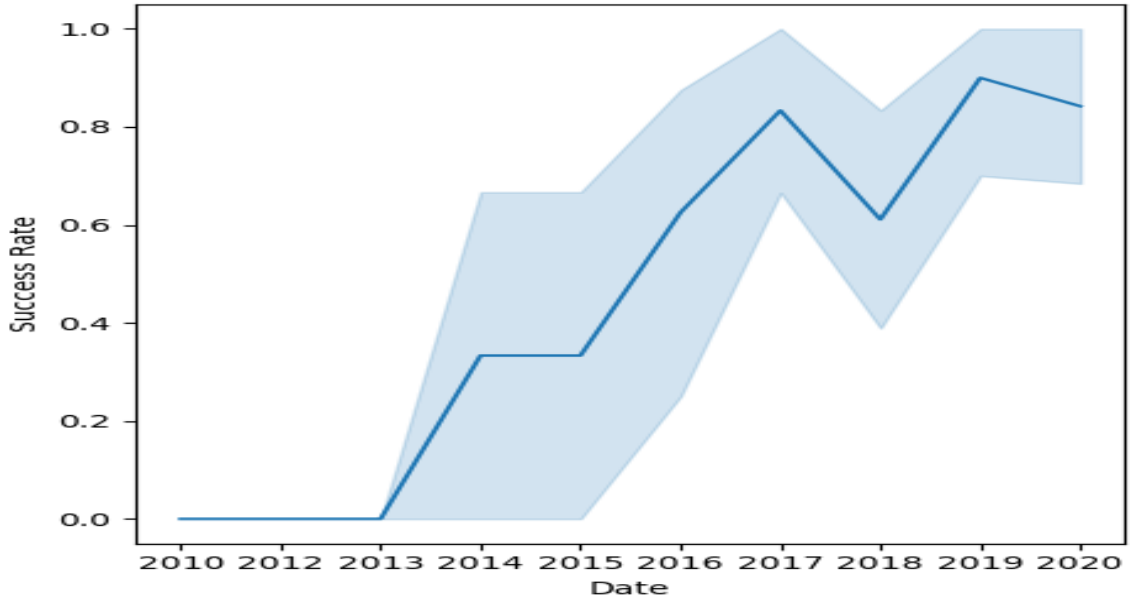
Payload vs Orbit



- Heavier payloads lead to higher success rates for **LEO**, **ISS** and **PO** orbits.
- This relationship does not seem to apply for **GTO** orbit.

Results – Exploratory Data Analysis

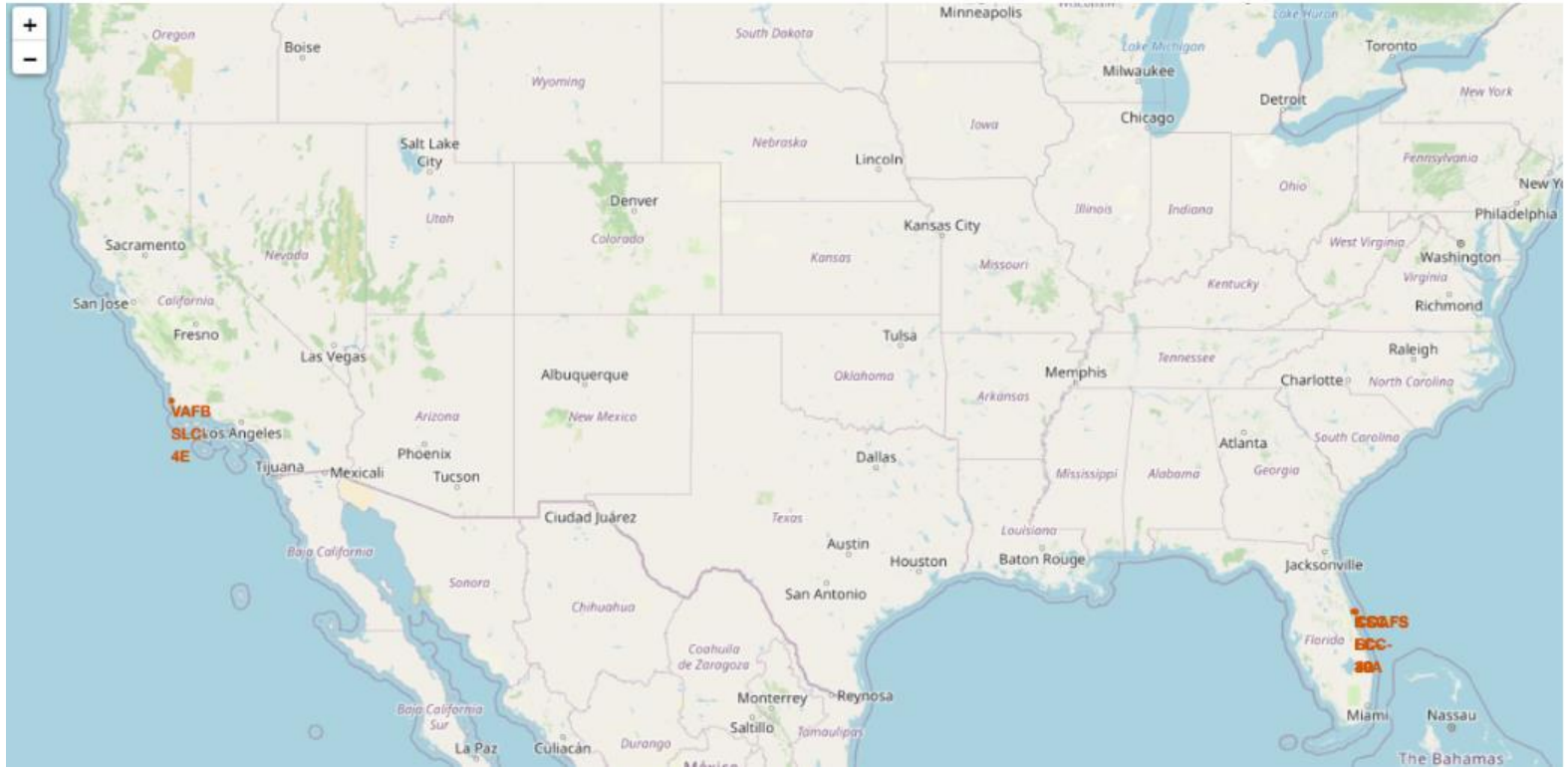
Launch Success Rate over Time



- Success rate generally improved from 2013 to 2020, with a slight drop between 2017 to 2018 and 2019 to 2020.

Results – Launch Site Analysis

Falcon 9 Launch Sites



We can see that the Launch Sites are all near to coastlines due to safety reasons to minimize damage in the event of unsuccessful launches.

Results – Launch Site Analysis

Launch Outcome at each Launch Site

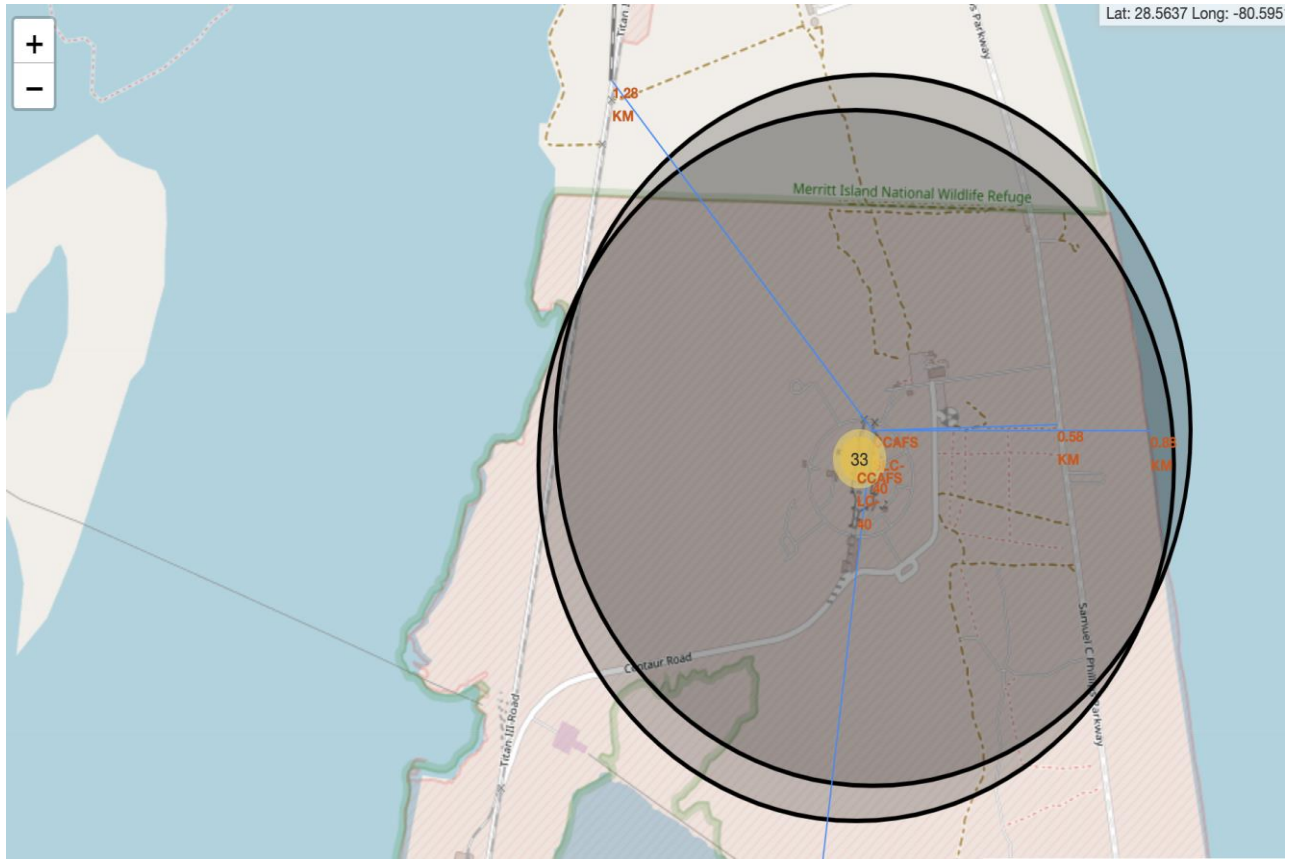
Green markers – Successful launch **Red** markers – Unsuccessful launch

KSC LC-39A has comparatively higher success rates compared to the other 2 Launch Sites.

Results – Launch Site Analysis

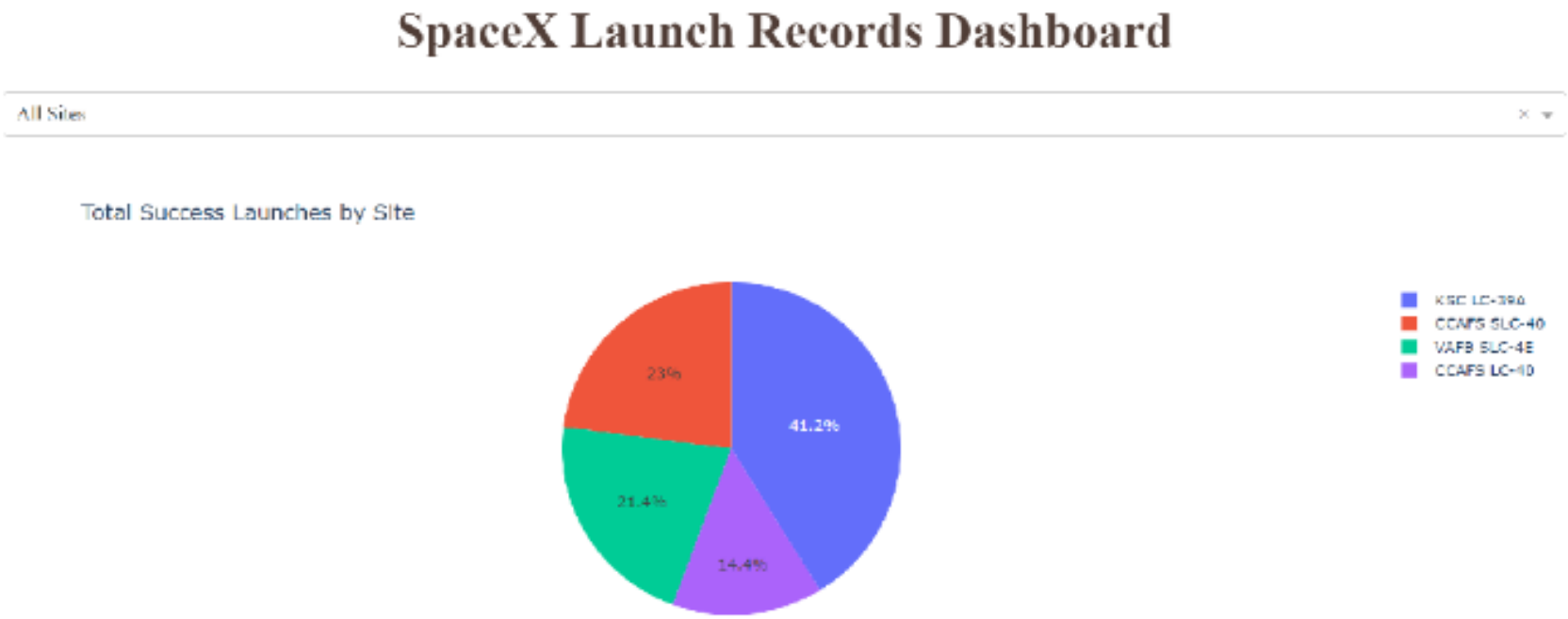
Launch Sites Proximities (CCAFS SLC-40)

- 0.88km from nearest coastline – near to coastline for safety reasons e.g. landing in water to minimize damage from failed launches
- 51.43km from nearest city – far from cities to minimize damage to human population and property from failed launches
- 0.58km from nearest highway and 1.28km from nearest railway – near to highway and railways to facilitate transport of people and material for launch activities



Results – Dashboard with Plotly

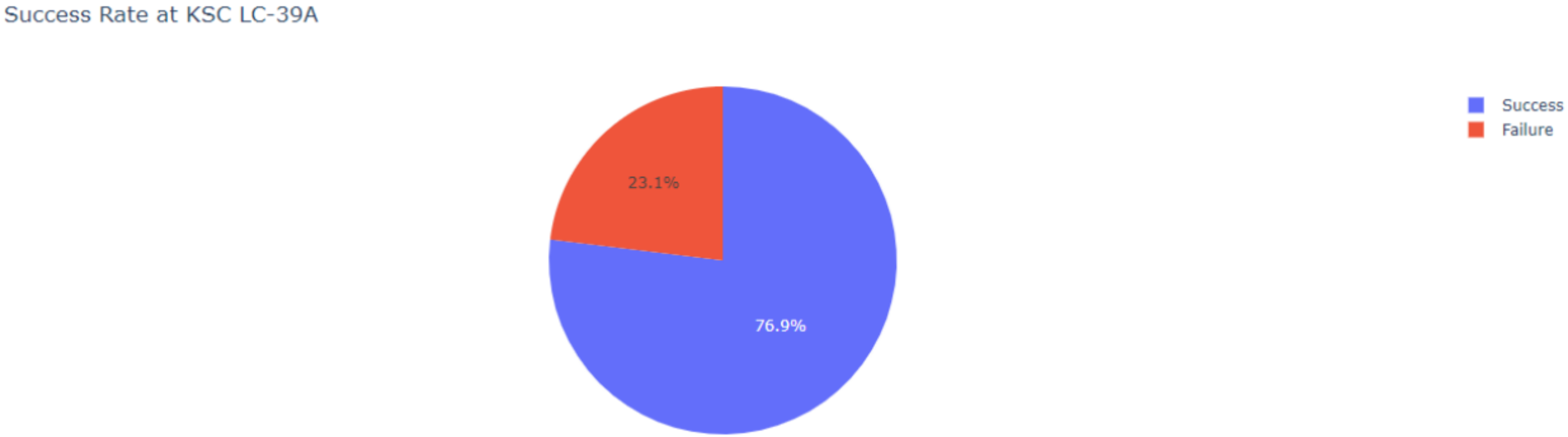
Success Rate by Launch Sites



KSC LC-39A has the highest number of successful launches (41.2% of total successful launches).

Results – Dashboard with Plotly

Launch Success (KSC LC-29A)



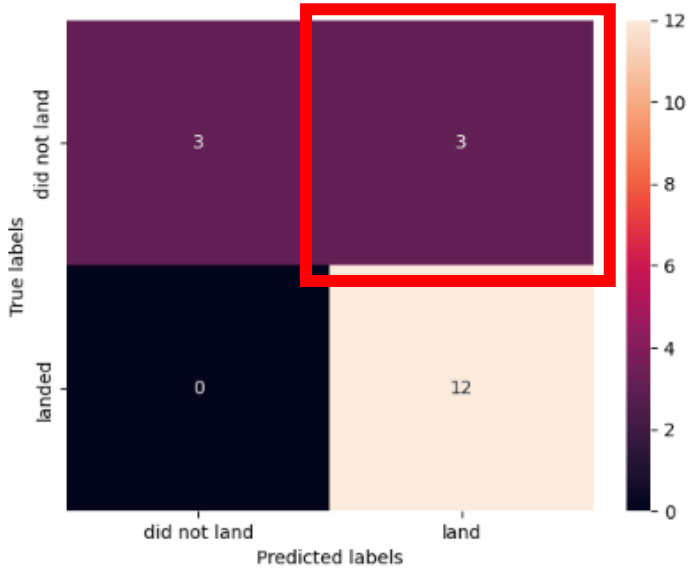
KSC LC-39A has high success rate of 76.9%.

Results – Predictive Analysis

Logistic Regression	Support Vector Machine	Decision Tree Classifier	K-Nearest Neighbour

Classification Models Accuracy: All 4 models have the **same accuracy score of *0.8333333***.

Confusion Matrix: All 4 models have **identical Confusion Matrix** as follow. Concern is with the false positives i.e. predicted "landed" but actual result is "did not land".



Conclusion

- Overall, the success rate of Falcon 9 first stage landing has improved over the years.
- Success rate is typically higher with higher payload mass across the launch sites.
- Orbits **ES-L1**, **GEO**, **HEO** and **SSO** have a 100% success rate.
- Rocket launch sites are generally closer to coastline, railway and highways, but far away from cities.
- Launch Site **KSC LC-39A** has the highest likelihood of first stage recovery; has a 100% success rate for launches with Payload mass less than 5,600 kg.
- It is possible to predict the landing outcome of the first stage of Falcon 9 with a reasonably high degree of accuracy based on the data set, with similar performance across the 4 models applied.