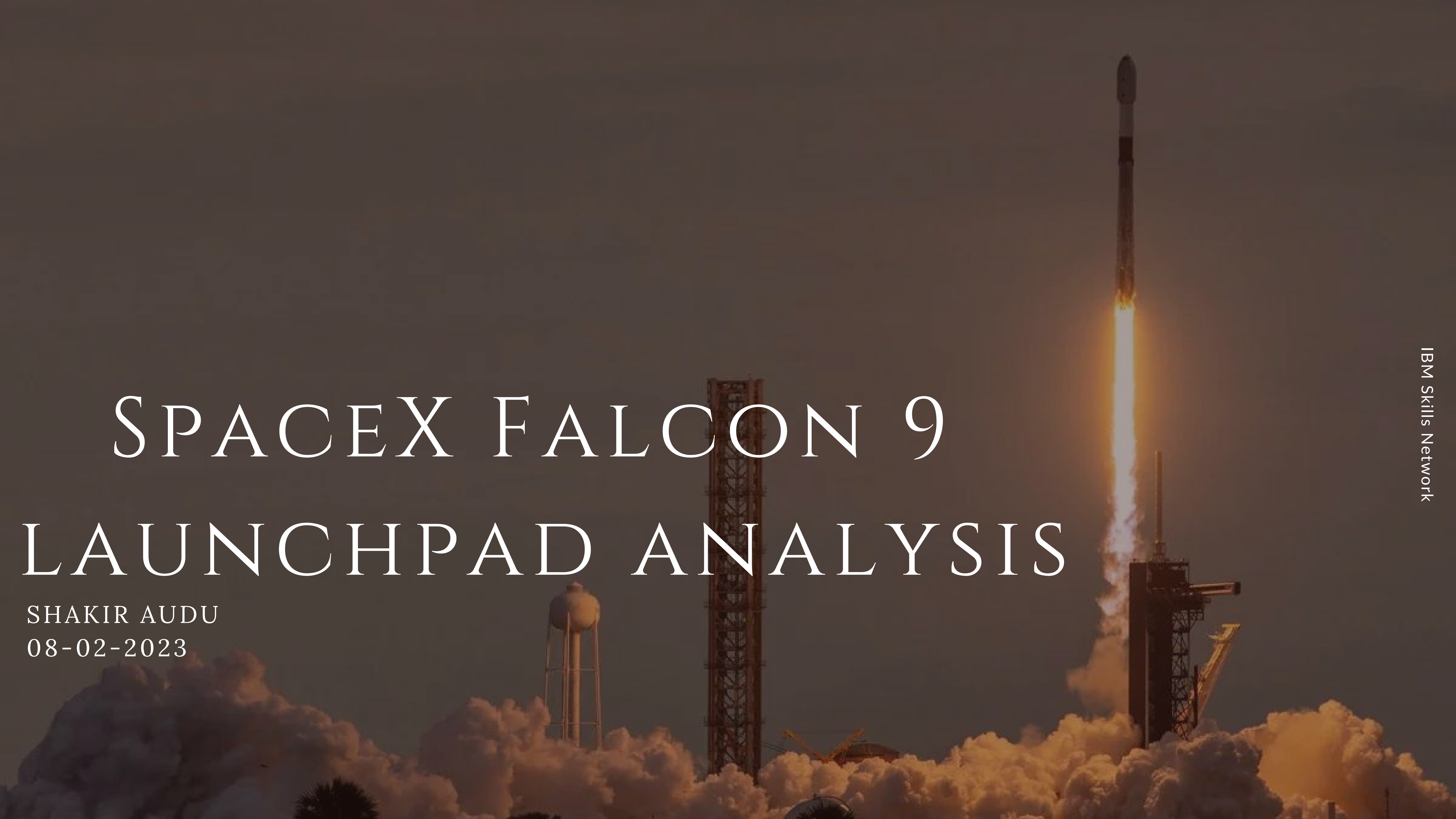


SpaceX Falcon 9 launchpad analysis

SHAKIR AUDU
08-02-2023

IBM Skills Network

Table of Contents

Executive Summary

Summary of the methodologies used:

- Data collection
- Data wrangling
- Exploratory Data Analysis with SQL and Data Visualization
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Summary of the results gotten:

- Exploratory Data Analysis results
- Interactive Analytics demo in screenshots
- Predictive Analysis results

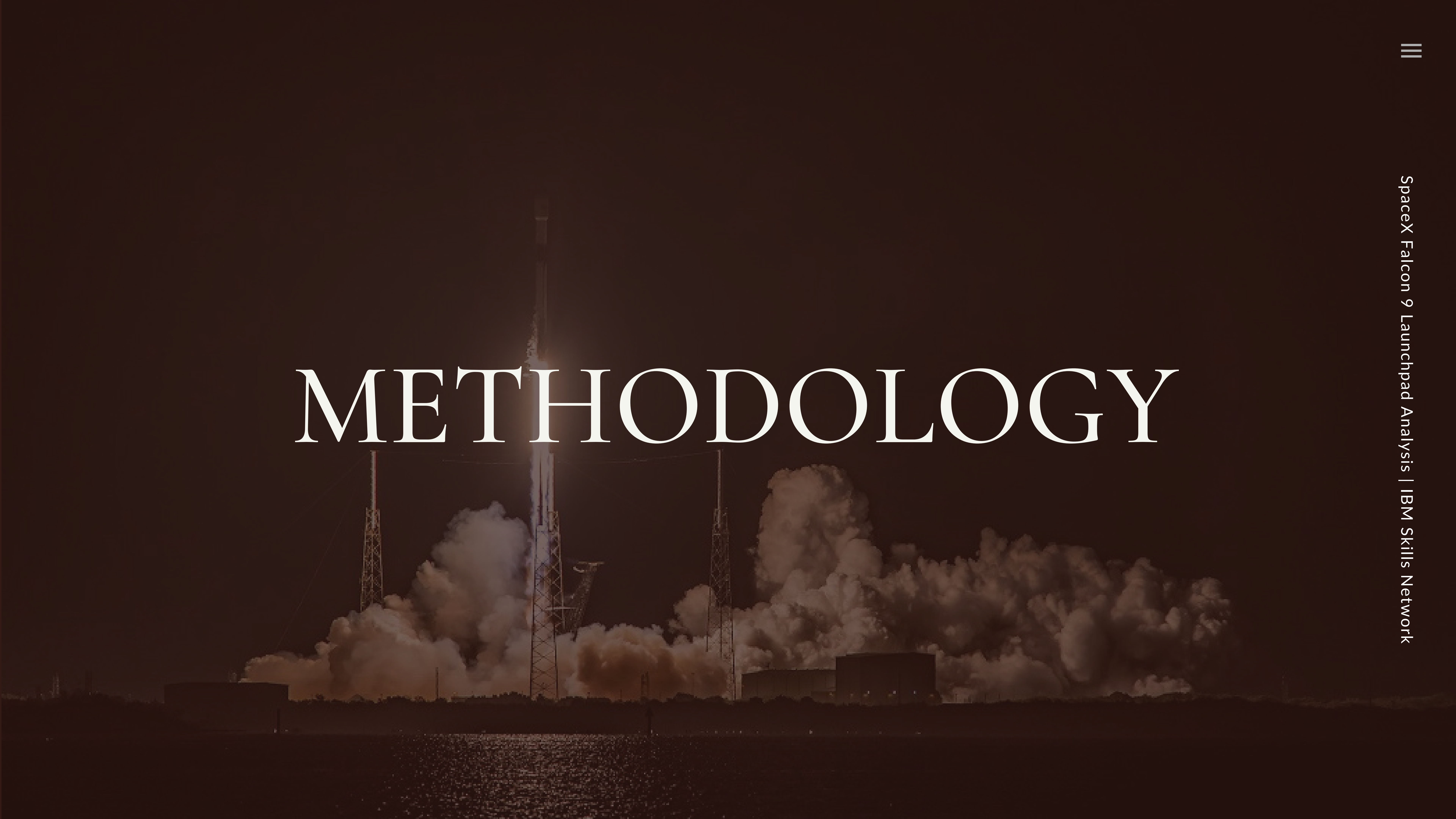
Introduction

SpaceX is the most successful company of the commercial space age, making space travel affordable. The company advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. Using public information and machine learning models, we are going to predict if SpaceX will reuse it's first stage.

Desired Results:

- Discover what variables affect the success rate of Space X's Falcon 9 rocket
- Predict whether future Falcon 9 rockets will land successfully

METHODOLOGY



Methodology

Data collection methodology:

- Using SpaceX Rest API
- Using Web Scrapping from Wikipedia

Data cleaning methodology:

- Filtering the data
- Dealing with missing values
- Using One Hot Encoding to prepare the data to a binary classification

Performed Exploratory Data Analysis (EDA) using SQL and visualization techniques

Performed Interactive Visual Analytics using Folium and Plotly Dash

Performed Predictive Analysis using Machine Learning classification models

Data Collection

The data used for the analysis was gotten from SpaceX API and SpaceX Wikipedia entry. A combination of both datasets was used for the purpose of having a richer dataset which will ultimately lead to more detailed analysis.

Data Collection-SpaceX API

Requested rocket launch data from SpaceX API

Created a dataframe from the data gotten

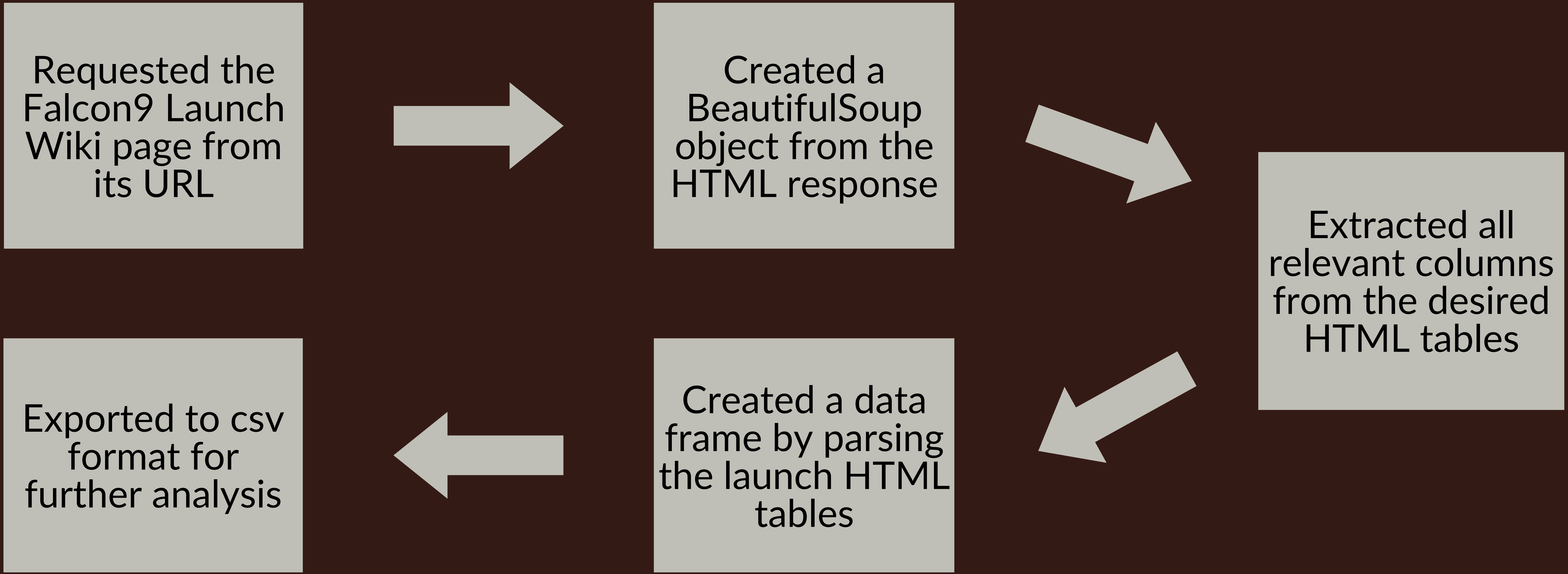
Filtered the dataframe to only include Falcon 9 launches

Dealt with incomplete values for 'PayLoad Mass ' data

Exported to csv format for further analysis

Github URL - Data Collection

Data Collection- Web scrapping



Github URL - Data Collection Webscrapping

Data Cleaning/Wrangling

Exploratory Data Analysis was performed to find some patterns in the data and determine what would be the label for training supervised models.

The following procedures were followed:

- Calculated the number of launches on each site
- Calculated the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome per orbit type
- Created a landing outcome label (1 means the booster successfully landed 0 means it was unsuccessful).

Github URL - Data Cleaning

Exploratory Data Analysis (EDA)

Further analysis was done using SQL queries to select and sort the data, and using visualization tools (Plotly Dash and Folium) to visualize the data and extract meaningful patterns to guide the modelling process. Finally feature engineering was done using OneHotEncoding to convert categorical variables to numeric ones to better fit machine learning algorithms to improve modelling.

Some of the patterns targeted:

SQL

- Unique Launch Sites
- Average Payload Mass
- Total number of successful and unsuccessful mission outcomes
- Count of successful landing outcomes between specific dates
- Successful boosters with payload mass greater than 4000 but less than 6000

Github URL - SQL

Data Visualization

- Relationship between several variable such as:

  Flight Number vs. Payload Mass

  Payload Mass vs Launch Site

  Success Rate vs Orbit Type

  Flight Number and Orbit type

  Payload Mass vs Orbit Type
- Visualize the yearly success rate.

Github URL - Visualization

Location Analysis

A Rocket's launch success rate may also depend on the location and proximities of the launch site. Finding an optimal location for building a launch site certainly involves many factors so location analysis was done. The tool used was **Folium** (A Python library used for visualizing geospatial data)

The following procedures were followed:

- All launch sites locations were marked and added to a map using their latitudes and longitudes
- The successful and failed launches for each site were marked on the map
- Discovered launch site proximity patters such as:

  Distance between launch sites

  How far the launch sites are from cities

  Distance between launch site to coast lines

  How relative the launch site is to the Equator

Github URL - Geospatial Analysis

Machine Learning Modelling

A Machine Learning pipeline was built to predict if the first stage will land given the data. To get the best model that fits the data.

Firstly the data was split into 2 sets (the training set and testing set that took 80% and 20% of the data respectively)

GridSearchCV (A cross-validation technique for finding the optimal hyperparameters to increase the model performance) was used to tune the best hyperparameter for the chosen classification models (SVM, Classification Trees and Logistic Regression)

Finally, the models were tested for accuracy using the following methods:

- Jaccard Score
- F1-Score
- Confusion matrix

Github URL - ML Modelling

RESULTS

EXPLORATORY ANALYSIS

A SUMMARY OF THE DATAFRAME

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
1	2010-06-04	Falcon 9	6104.959411764710	LEO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1.0	0	B0003	-80.577366	28.5618571	0
2	2012-05-22	Falcon 9	525.0	LEO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1.0	0	B0005	-80.577366	28.5618571	0
3	2013-03-01	Falcon 9	677.0	ISS	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1.0	0	B0007	-80.577366	28.5618571	0
4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	FALSE	FALSE	FALSE		1.0	0	B1003	-120.61082900000000	34.632093	0
5	2013-12-03	Falcon 9	3170.0	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1.0	0	B1004	-80.577366	28.5618571	0
6	2014-01-06	Falcon 9	3325.0	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1.0	0	B1005	-80.577366	28.5618571	0
7	2014-04-18	Falcon 9	2296.0	ISS	CCAFS SLC 40	True Ocean	1	FALSE	FALSE	TRUE		1.0	0	B1006	-80.577366	28.5618571	1
8	2014-07-14	Falcon 9	1316.0	LEO	CCAFS SLC 40	True Ocean	1	FALSE	FALSE	TRUE		1.0	0	B1007	-80.577366	28.5618571	1
9	2014-08-05	Falcon 9	4535.0	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1.0	0	B1008	-80.577366	28.5618571	0
10	2014-09-07	Falcon 9	4428.0	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1.0	0	B1011	-80.577366	28.5618571	0
11	2014-09-21	Falcon 9	2216.0	ISS	CCAFS SLC 40	False Ocean	1	FALSE	FALSE	FALSE		1.0	0	B1010	-80.577366	28.5618571	0
12	2015-01-10	Falcon 9	2395.0	ISS	CCAFS SLC 40	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb761634e7cb	1.0	0	B1012	-80.577366	28.5618571	0
13	2015-02-11	Falcon 9	570.0	ES-L1	CCAFS SLC 40	True Ocean	1	TRUE	FALSE	TRUE		1.0	0	B1013	-80.577366	28.5618571	1
14	2015-04-14	Falcon 9	1898.0	ISS	CCAFS SLC 40	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb761634e7cb	1.0	0	B1015	-80.577366	28.5618571	0
15	2015-04-27	Falcon 9	4707.0	GTO	CCAFS SLC 40	None None	1	FALSE	FALSE	FALSE		1.0	0	B1016	-80.577366	28.5618571	0
16	2015-06-28	Falcon 9	2477.0	ISS	CCAFS SLC 40	None ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb6bb234e7ca	1.0	0	B1018	-80.577366	28.5618571	0
17	2015-12-22	Falcon 9	2034.0	LEO	CCAFS SLC 40	True RTLS	1	TRUE	FALSE	TRUE	5e9e3032383ecb267a34e7c7	1.0	0	B1019	-80.577366	28.5618571	1
18	2016-01-17	Falcon 9	553.0	PO	VAFB SLC 4E	False ASDS	1	TRUE	FALSE	TRUE	5e9e3033383ecbb9e534e7cc	1.0	0	B1017	-120.61082900000000	34.632093	0
19	2016-03-04	Falcon 9	5271.0	GTO	CCAFS SLC 40	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032383ecb6bb234e7ca	1.0	0	B1020	-80.577366	28.5618571	0

The dataset is relatively small. It has 90 samples and 18 attributes namely;

FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude and Class (the target variable).

ANALYSIS RESULTS

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite,
```

Figure 1a

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
HEO       1
SO        1
GEO       1
Name: Orbit,
```

Figure 1b

Figure 1a and 1b show the number of launches by launch site and orbit type respectively.

The different Site Names are:

- Cape Canaveral Space Launch Complex 40 (CCAFS SLC 40)
- Kennedy Space Center Launch Complex 39A (KSC LC 39A)
- Vandenberg Air Force Base Space Launch Complex 4E (VAFB SLC-4E)

The 3 most visited orbits are:

- The Geosynchronous orbit (GTO)
- The modular space station (a habitable artificial satellite) in low Earth orbit
- The Very Low Earth Orbit (VLEO)

SQL ANALYSIS RESULTS

total_payload_mass
45596

Total payload mass carried for customer - NASA (CRS)

first_successful_landing
2015-12-22

Date of the first successful landing outcome (ground pad)

average_payload_mass
2534.6666666666665

Average payload mass carried by the F9 v1.1 booster version

SQL ANALYSIS RESULTS

mission_outcome	total_number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Figure showing the number of successful and failed missions.

The above figure shows a high rate of successful missions (99%), having one mission with an uncertain payload status.
Only 1 mission was unsuccessful.

SQL ANALYSIS RESULTS

landing__outcome	count_outcomes
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

Figure showing the number of successful and failed missions by landing type between 2010-06-04 and 2017-03-20

Most of the Falcon9 rockets within the specified time period did not land.

Majority of landing attempts were on drone ships.

Ground pad landing attempts had a 100% success rate.

SQL ANALYSIS RESULTS

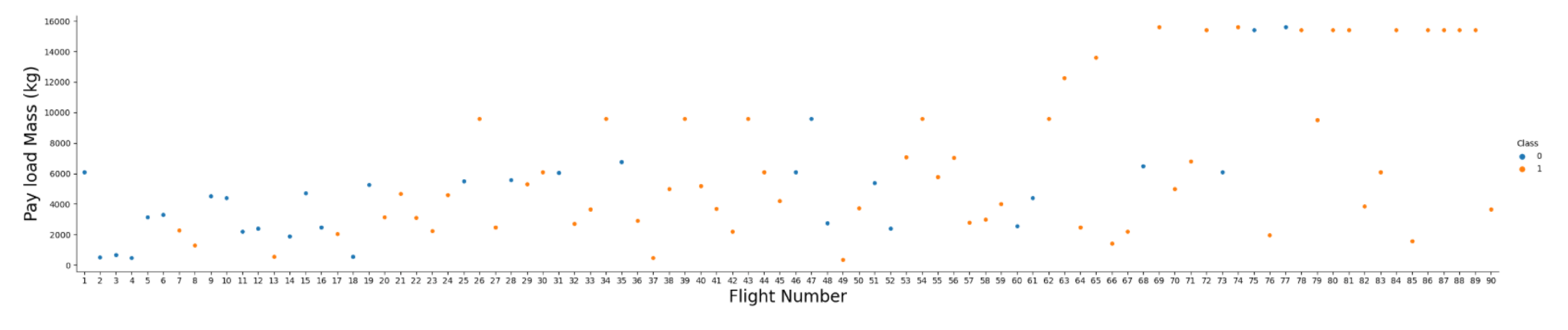
booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

The figure shows booster versions which have carried the maximum payload mass.

Different booster versions carried the maximum payload mass

All boosters were Falcon 9 Booster 5 B10 versions

VISUALIZATION RESULTS



A scatterplot showing the relationship FlightNumber vs. PayloadMass

Blue dot – failed landing

Orange dot – successful landing

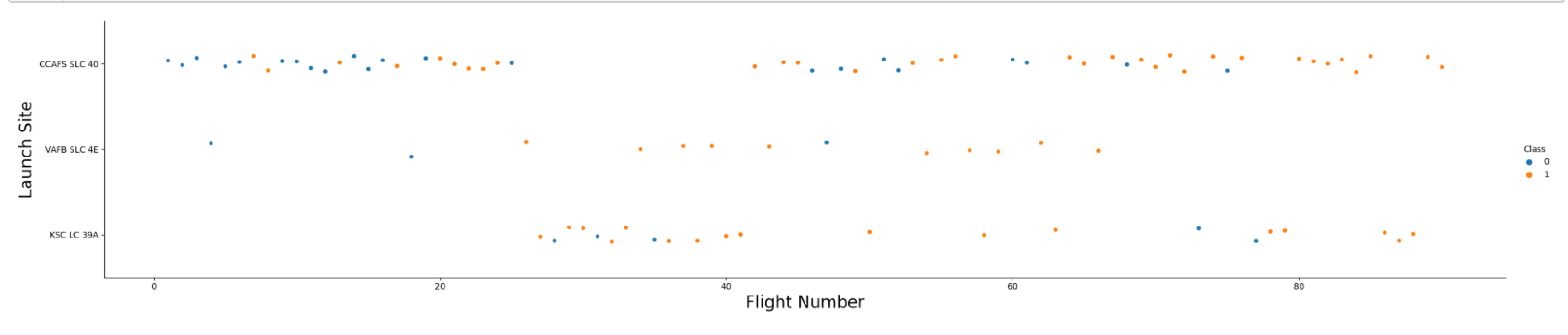
Most flights carried payload of less than 8,000kg

Rockets carrying a payload mass of higher than 10,000kg were relatively successful

The first time that rockets carried a maximum payload of 16,000kg was on the 69th flight

---

VISUALIZATION RESULTS



A scatterplot showing the relationship FlightNumber vs. Launch Site

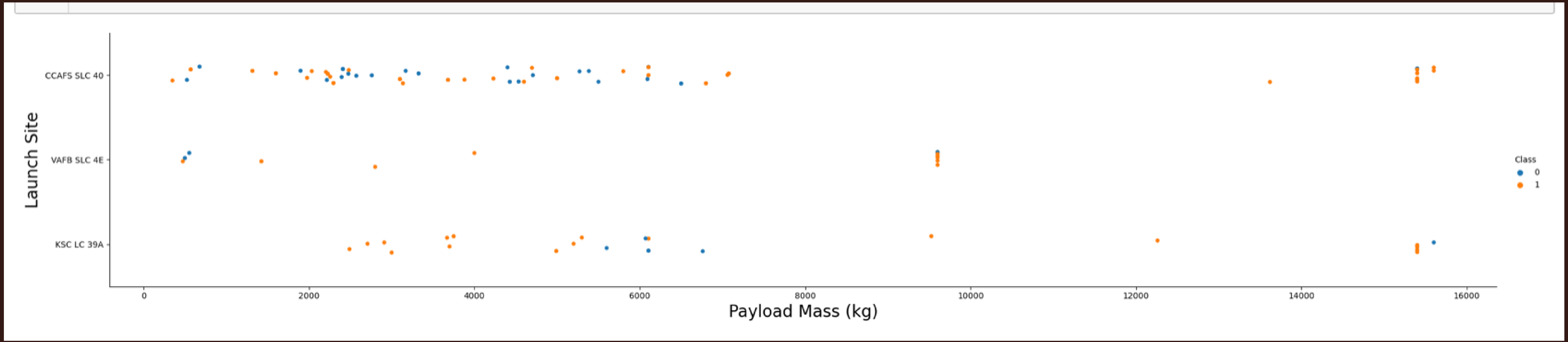
Blue dot – failed landing

Orange dot – successful landing

The Cape Canaveral Space Launch Complex 40 was used first and for majority of the flights

The Vandenberg Air Force Base Space Launch Complex has a similar landing success rate to the Kennedy Space Center Launch Complex 39A

VISUALIZATION RESULTS



A scatterplot showing the relationship Launch Site vs. PayloadMass

Blue dot – failed landing

Orange dot – successful landing

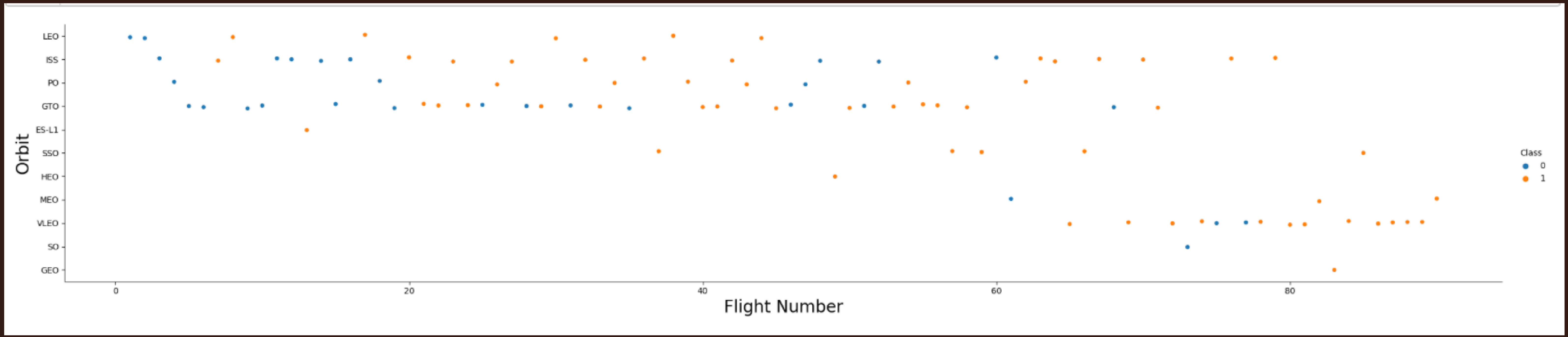
The majority of rockets with payload masses of 8,000 kg and lower were launched in Cape Canaveral Space Launch Complex 40

The maximum payload mass launched on Vandenberg Air Force Base Space Launch Complex was less than 10,000 kg

Flights carrying heavier payload masses (8000kg and higher) were more successful than flights with lesser payload masses

---

VISUALIZATION RESULTS



A scatterplot showing the relationship FlightNumber vs. Orbit Type

Blue dot – failed landing
Orange dot – successful landing

Most early flights were lauched to the following orbits: LEO, ISS, PO and GTO
The later flights were launched to VLEO

Very little flights were launched to higher orbits.

VISUALIZATION RESULTS

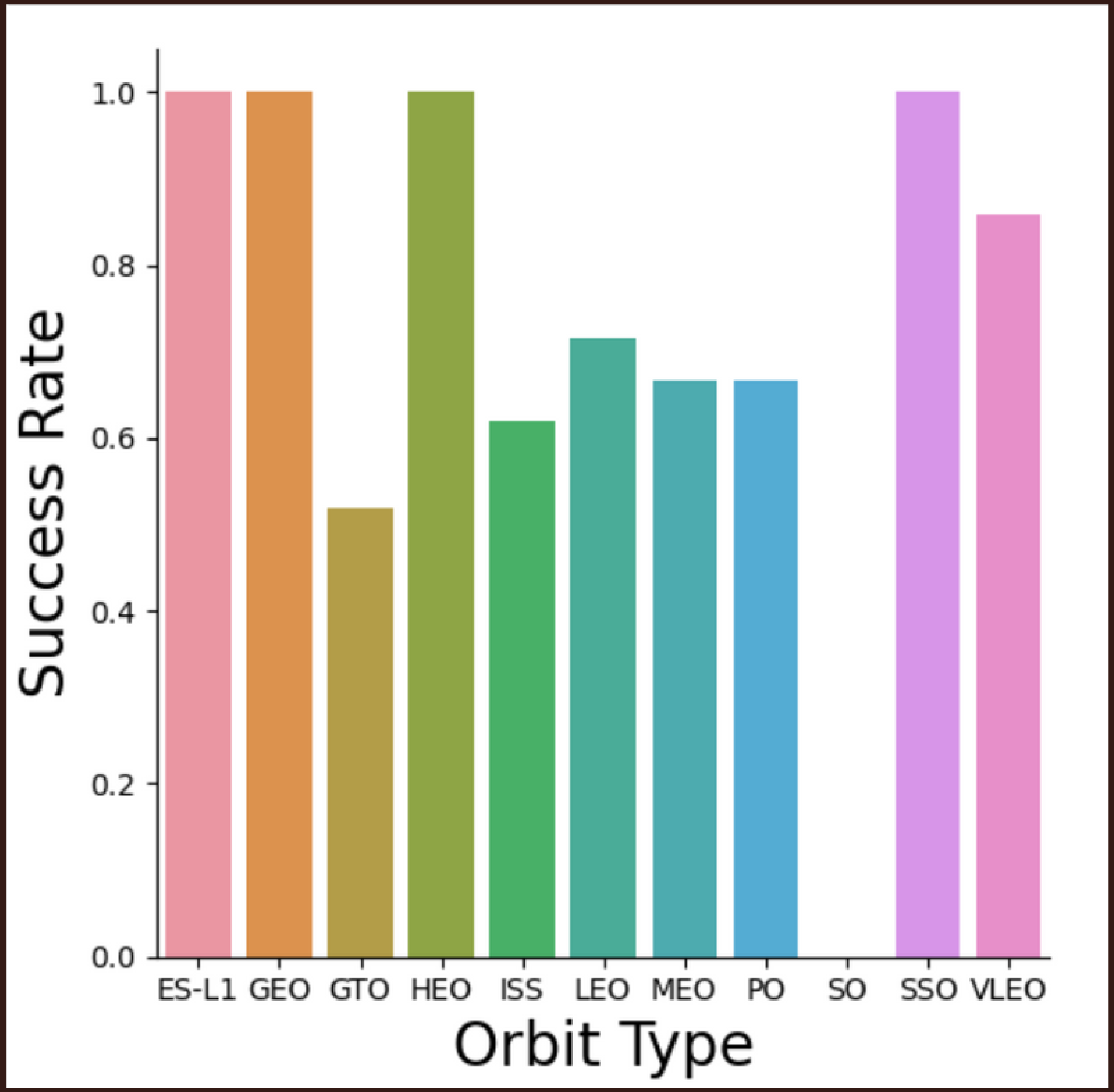


Chart showing the success rate of rockets launched to different orbits

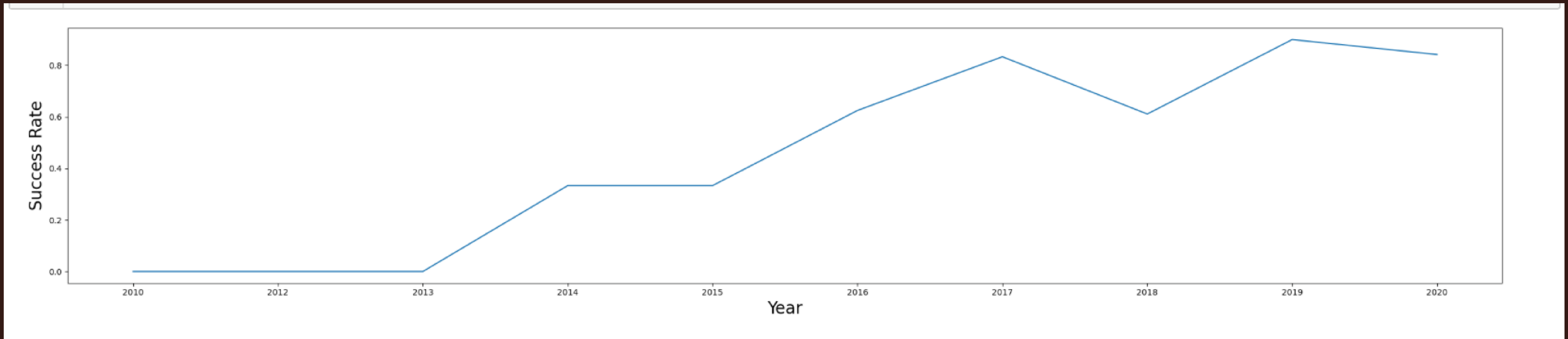
Rockets launched to the following orbits had a 100% success rate:

- GEO
- ES-L1
- HEO
- SSO

Rockets launched to most visited orbits had lower success rates

Half the rockets launched to the geosynchronous orbit were successful

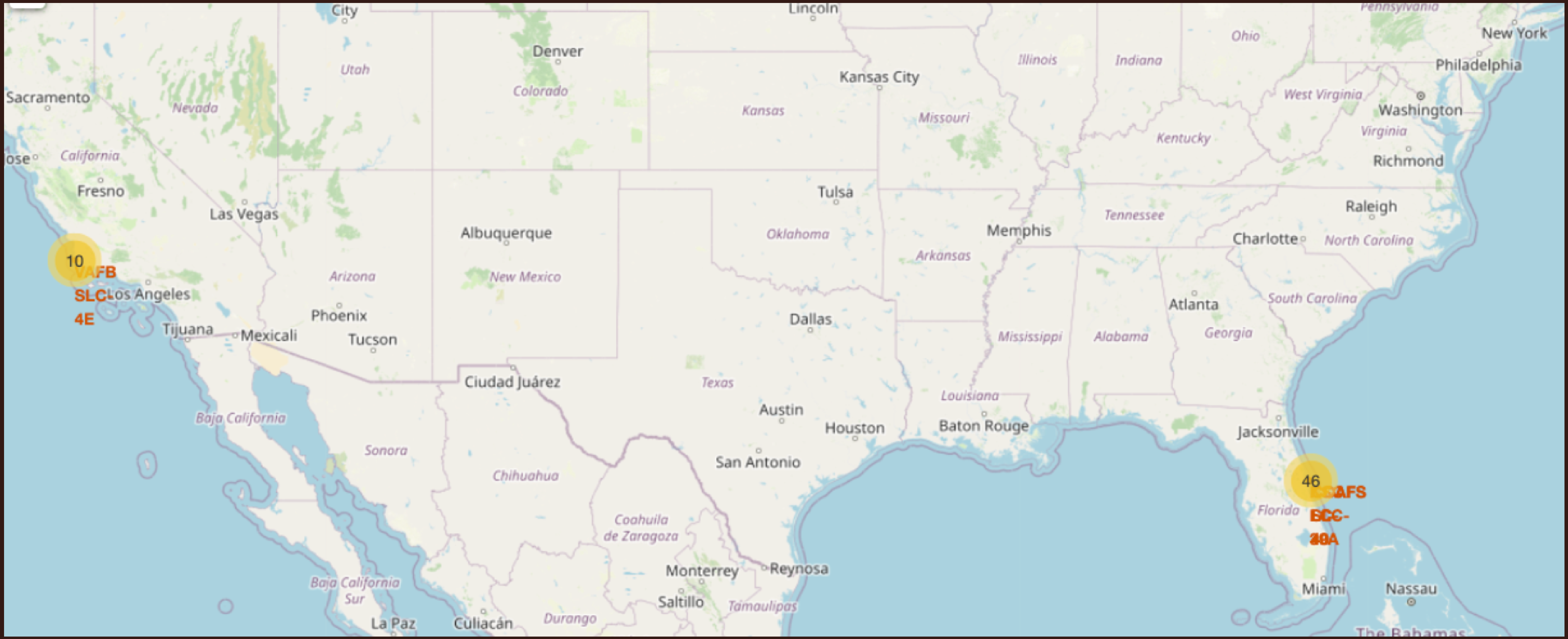
VISUALIZATION RESULTS



The figure shows the yearly success rate of Falcon9 rockets

- The natural trend is a yearly increase in the success rates of rocket launches
- For the first time, in 2017 the success rate dropped by 20%
- The year with the highest increase was between 2013 and 2014, were the success rate increased by 30%

VISUALIZATION RESULTS

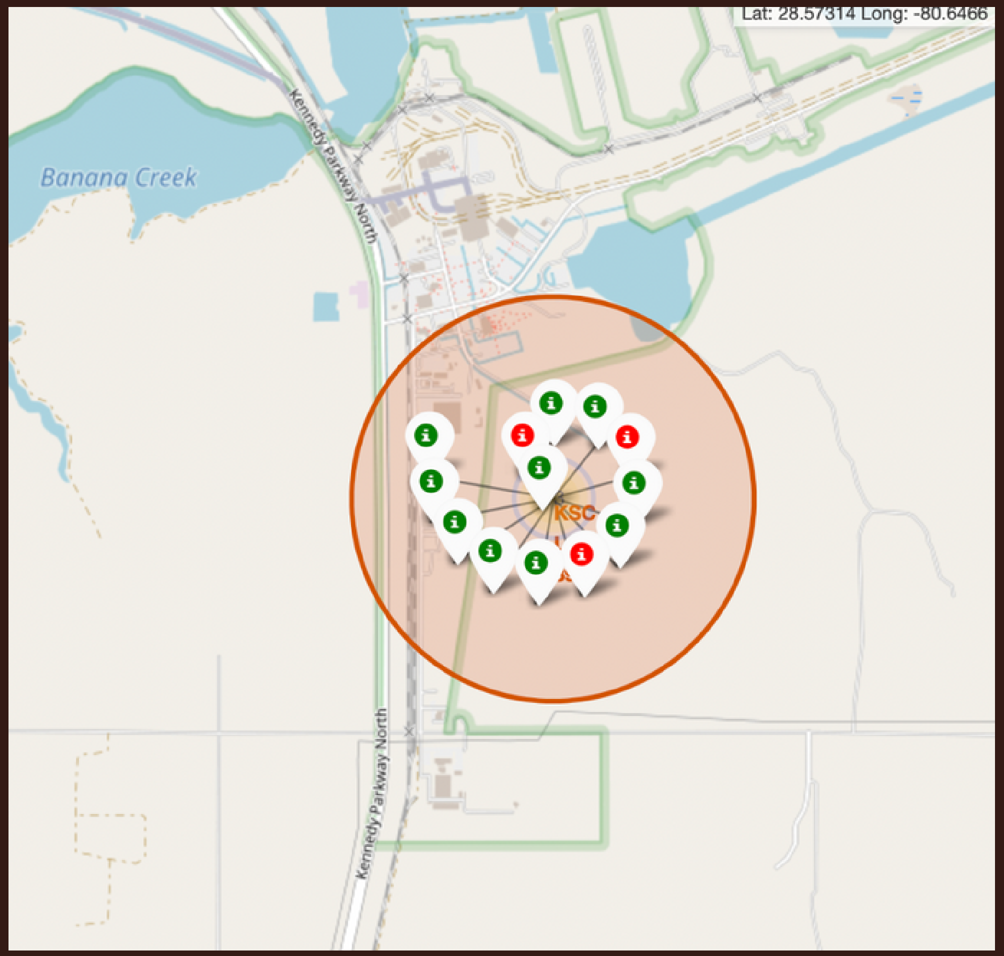


The figure shows the location of launch sites on a map

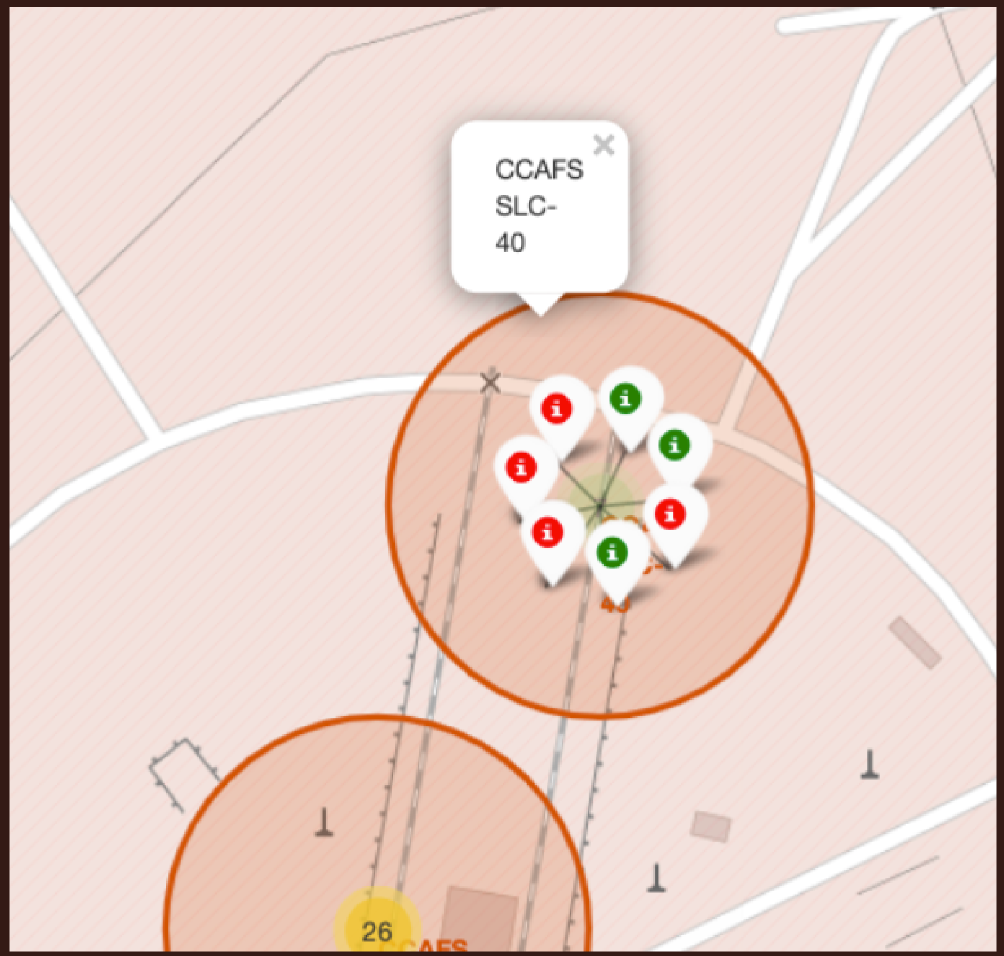
Notice that both launch sites displayed are close to the ocean, this can be taken as a safety requirement.

The launch sites are in line with the Earth's equator so they can take optimum advantage of the Earth's substantial rotational speed

VISUALIZATION RESULTS



The map shows the success rate of launch site KSC LC-39A

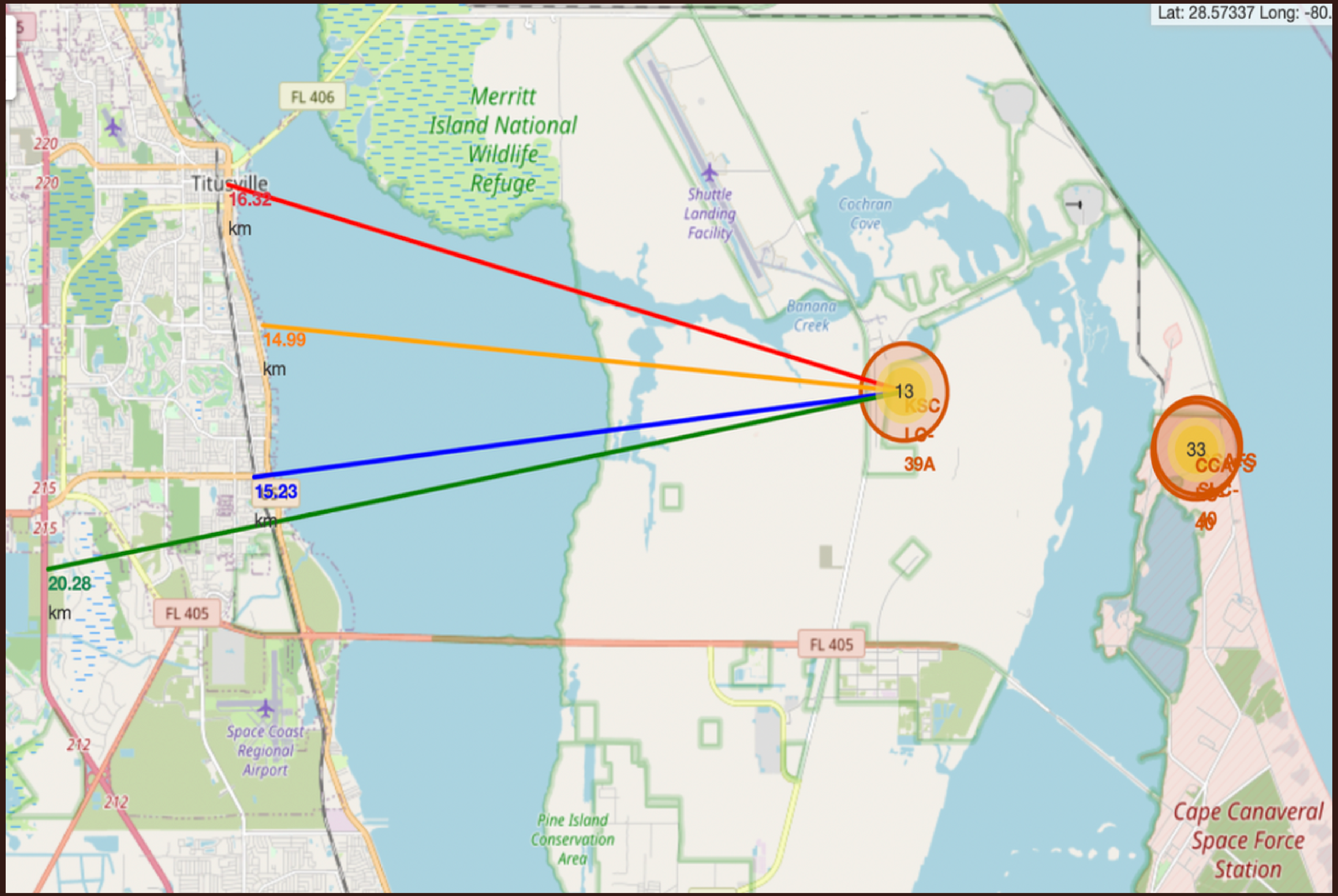


The map shows the success rate of launch site CCAFS SLC-40

Launch Site KSC LC-39A has a hifher success rate to laugh site CCAFS SLC-40

The green labels represent successful landings and the red labels represent failed landings

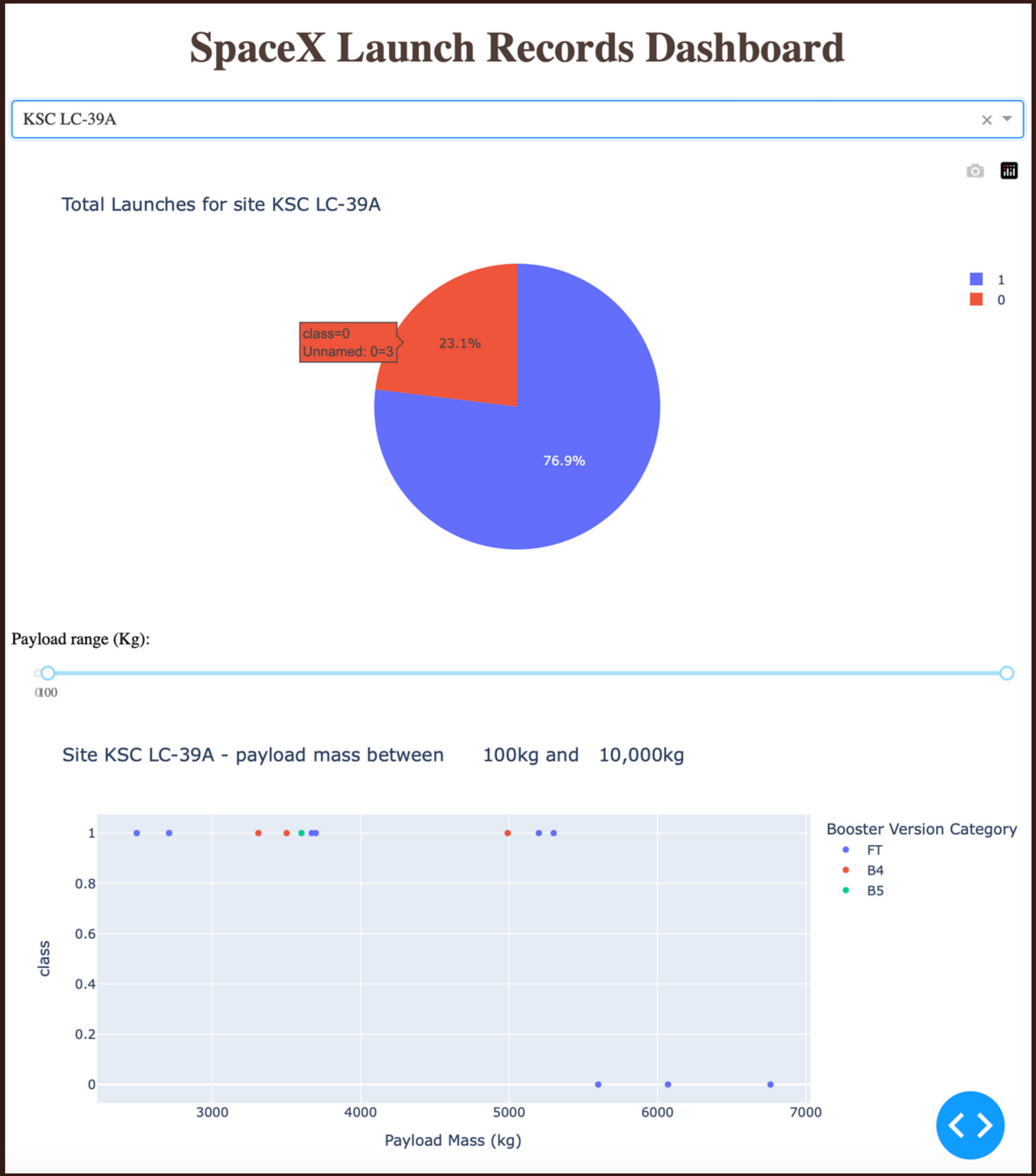
VISUALIZATION RESULTS



The map shows Cape Canaveral Space Launch Complex 40 and Kennedy Space Center Launch Complex 39A and their proximities

CITY	16.32km
RAILWAY	15.23km
HIGHWAY	20.28km
COASTLINE	14.99km

INTERACTIVE DASHBOARD RESULTS

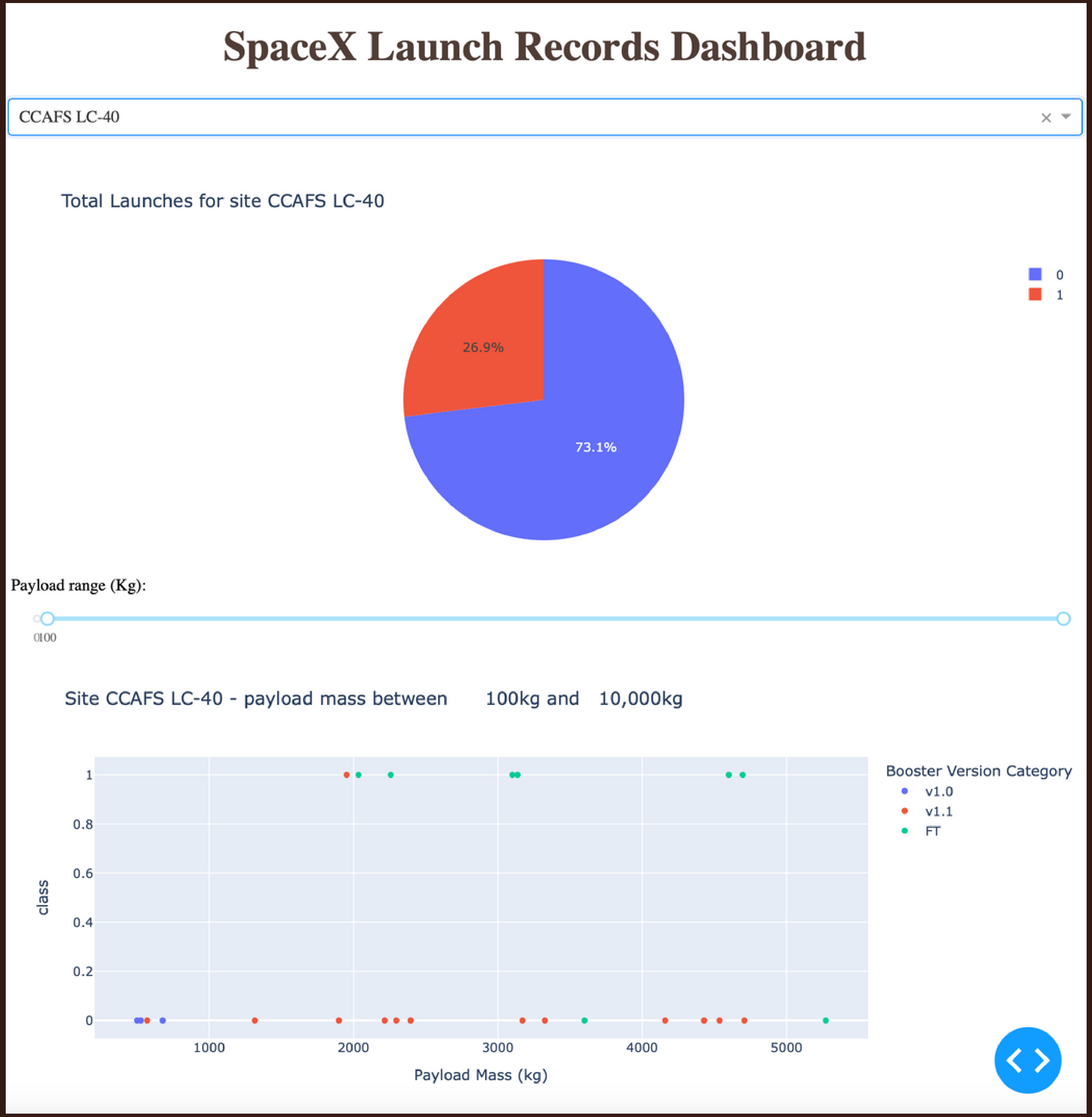


Kennedy Space Center Launch Complex 39A (KSC LC 39A)

Dashboard showing;

- The rate of successful launches
- The payload ranges for each booster version category (FT, B4, B5)

INTERACTIVE DASHBOARD RESULTS

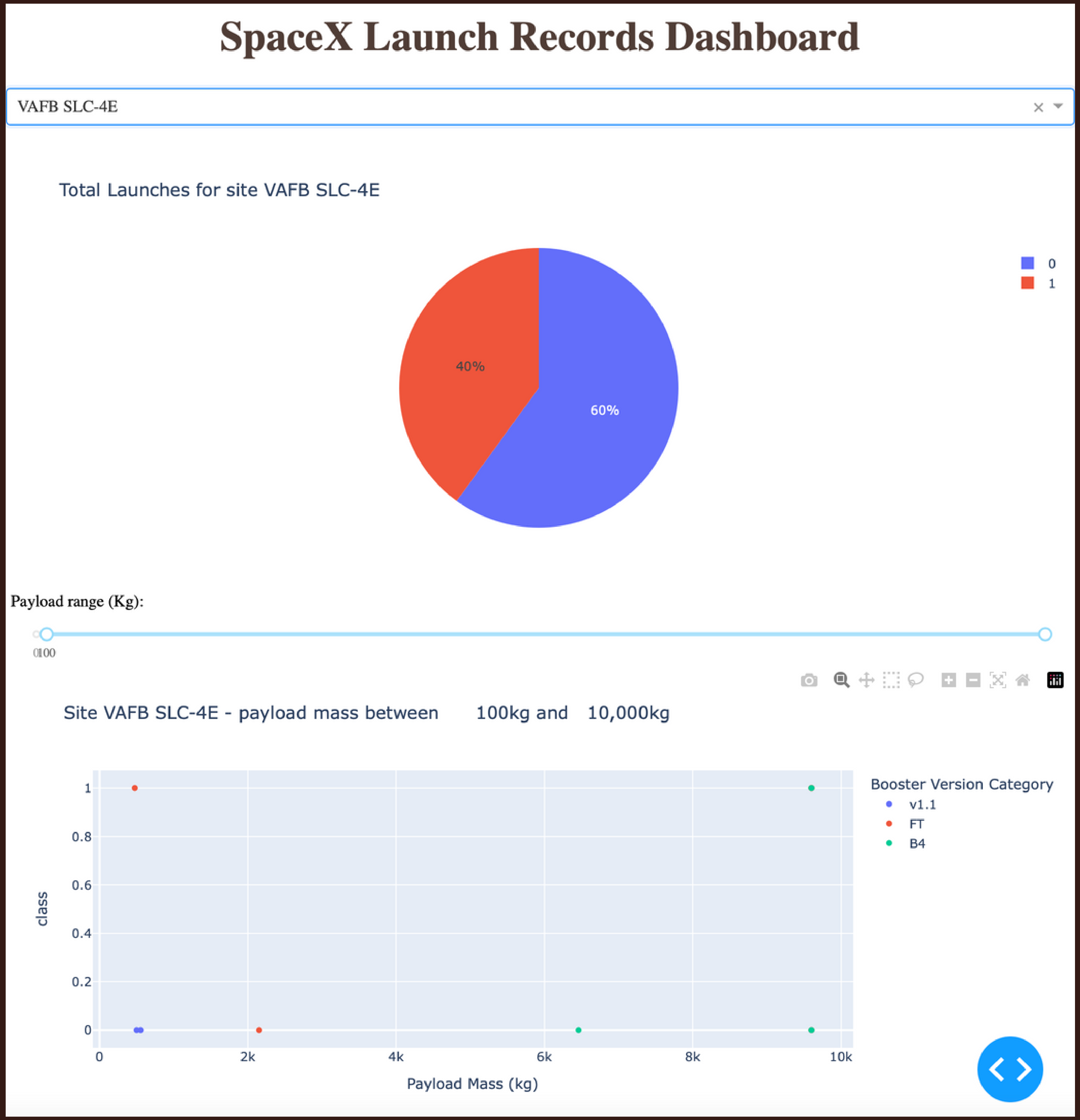


Cape Canaveral Space Launch Complex 40 (CCAFS LC 40)

Dashboard showing;

- The rate of successful launches
- The payload ranges for each booster version category (FT, B4, B5)

INTERACTIVE DASHBOARD RESULTS



Vandenberg Air Force Base Space Launch Complex 4E (VAFB SLC-4E)

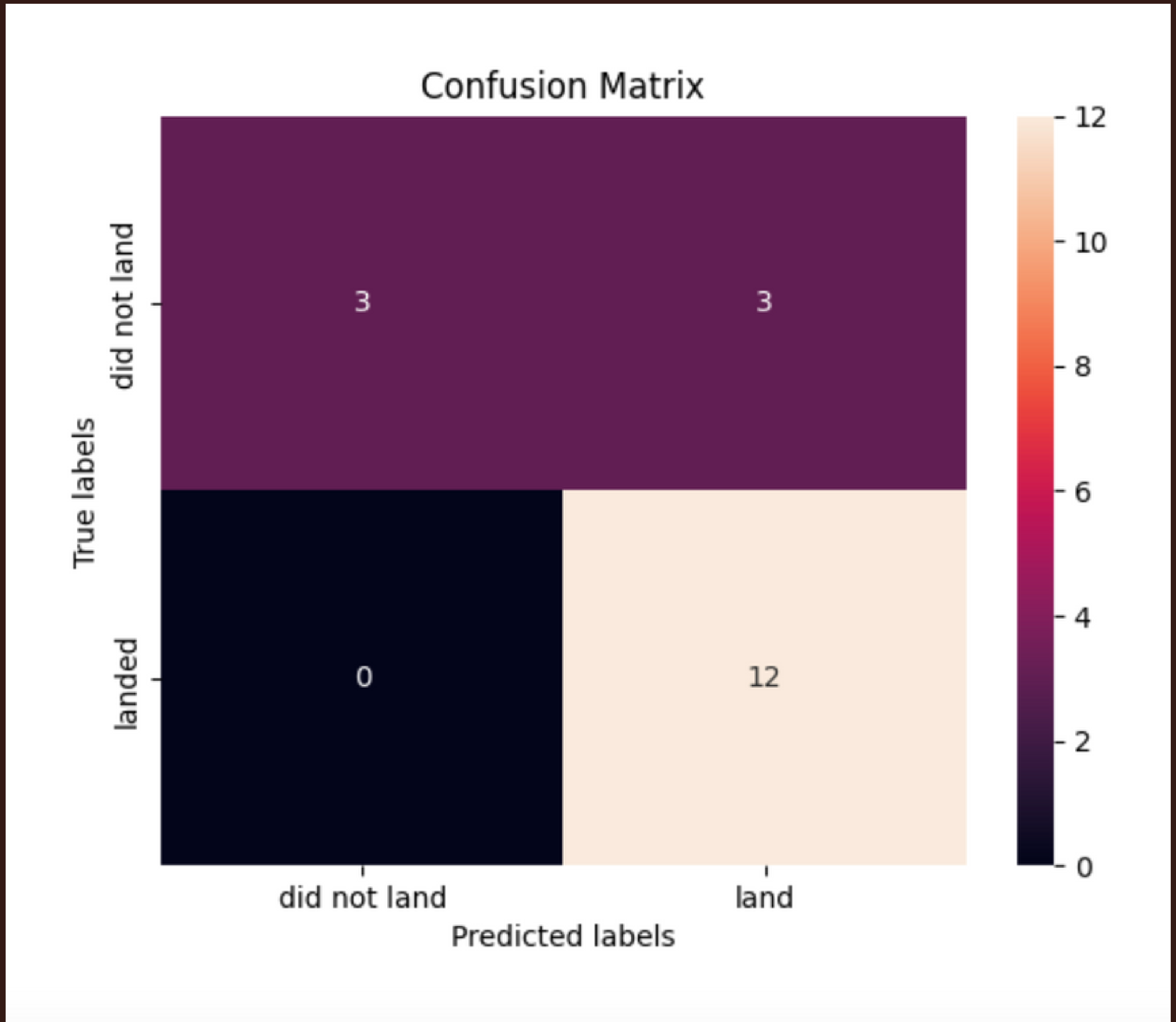
Dashboard showing;

- The rate of successful launches
- The payload ranges for each booster version category (FT, B4, B5)

RESULTS

PREDICTIVE ANALYSIS

MACHINE LEARNING MODELLING



Precision = 1

Recall = 0.8

We want our model to predict if flights will land so that we can make business decision.

So the precision of the model is more valuable.

From the Confusion Matrix shown, we can tell the following:

- The model rightly predicted all flights that landed.
- The model predicted that 3 flights did not land while they landed

MACHINE LEARNING MODELLING

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.819444	0.819444
F1_Score	0.909091	0.916031	0.900763	0.900763
Accuracy	0.866667	0.877778	0.855556	0.855556

- All algorithms did relatively well and had similar results. This can be as a result of using a small dataset.

- The algorithm that performed best was the Support Vector Machine (SVM). It had both higher scores and higher accuracy as well based on the scores and accuracy of the test set.

CONCLUSION

The goal was to use data to predict rocket landing outcome and with the result determine how to drive down the cost of space access.

Based on results gotten, Our model did really well predicting the test set

With the best algorithm having an overall accuracy of 87% whilst predicting 100% correctly on all rockets that landed

The analysis done can be used to determine the best launch conditions for better landing outcomes.

-

The model would perform better with a larger dataset.

APPENDIX

Special Thanks!

IBM Skills Network
Coursera

— Thank you