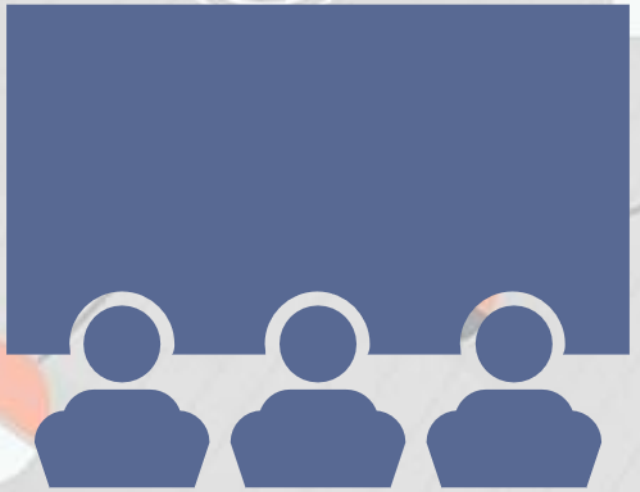
Space X Falcon 9 First Stage Landing Prediction

I Himika

GitHub Repository Link

OUTLINE



- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

EXECUTIVE SUMMARY

- SpaceY, a rocket company would like to bid against SpaceX, another rocket company.
- Spacex publicize Falcon 9 rocket launches cost on its website 2.6times less than the other providers primarily because Space X can reuse the first stage.
- Predicting the first stage rocket landing of SpaceX will also ascertain the cost of a launch.
- This report presents the model which can determine if the first stage will land, that will help to ascertain the price of each launch.

INTRODUCTION

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars;
- Whereas other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.
- Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

METHODOLOGY

- Data Collection-API
- Data Collection-Web Scraping
- Data wrangling
- EDA with SQL
- EDA with Visualization
- Interactive map with Folium (Launch Site Location)
- EDA and interactive visual analytics- (Launch Record Dashboard with Plotly Dash)
- Predictive Analysis

Data Collection-API

Requested rocket launch data from SpaceX API

Requested and parsed the SpaceX launch data using the GET request and store

Normalized data and stored in dataframe

Filtered the dataframe to only include Falcon 9 launches

Dealt with the missing values

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Le
0	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	Fa
1	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	Fa
2	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	Fa
3	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	Fa
4	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	Fa
...	...	...	...	...	...	...	...	...	...	...	
85	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	Tr
86	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	Tr

Fig1-Dataframe without any missing values

Data Collection-Web Scraping

Extracted a Falcon 9 launch records HTML table from Wikipedia

Extracetd all column/variable names from the HTML table header and storing in a dataframe

Parsed the table and converted it into a Pandas dataframe

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
104	8 January 2021 02:15[604]	F9 B5 △ B1060.4	CCSFS, SLC-40	Türksat 5A[605]	3,500 kg (7,700 lb)	GTO	Türksat	Success	Success (drone ship)
	A 3,500 kg (7,700 lb) satellite intended to be stationed at 31.0° east.[605] This is the most powerful satellite in Türksat's fleet[606] and will provide Ku-band television broadcast services over Turkey, the Middle East, Europe and Africa. The satellite was injected in to a Super-synchronous transfer orbit of 280 km × 55,000 km (170 mi × 34,180 mi) with 17.6° inclination.[607]								
105	20 January 2021 13:02[608]	F9 B5 △ B1051.8[609]	KSC, LC-39A	Starlink 16 v1.0 (60 satellites)	15,600 kg (34,400 lb)	LEO	SpaceX	Success	Success (drone ship)
	The first booster to successfully launch and land eight times. Achieved a record turnaround time between two launches of the same booster of only 38 days and brought the total of launched Starlink satellites to over 1000.[610] SpaceX stated that the landing would occur during higher winds than usual; this test to expand the landing envelope was successfully passed by the booster.[611]								
106	24 January 2021 15:00[612]	F9 B5 △ B1058.5[613]	CCSFS, SLC-40	*Transporter-1* (143 smallsat rideshare)	~5,000 kg (11,000 lb)	SSO	Various	Success	Success (drone ship)
	First dedicated smallsat rideshare launch, targeting a 525 km (326 mi) altitude orbit.[614] The launch deployed a record 143 satellites, consisting of 120 CubeSats, 11 microsatellites, 10 Starlinks, and 2 transfer stages. In addition, 2 hosted payloads and 1 non-separating dummy satellite[615] were[*failed verification*] launched.[616] These include SpaceBEE (x 36), Lemur-2 (x 8), ICEYE (x 3), UVSQ-SAT,[617] ELaNa 35 (PTD-1),[381] and multiple Kepler nanosats.[618][619] D-Orbit flew their ION SCV								

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	F9 v1.0B0003.1	Failure	4 June 2010	18:45
2	1	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0003.1	No attempt\n	4 June 2010	18:45
3	1	CCAFS	Dragon	0	LEO	NASA	Success\n	F9 v1.0B0003.1	No attempt	4 June 2010	18:45

Fig-2&3 Dtata from a web page to Dataframe

Data wrangling

Calculating the number of launches on each site

Calculating the number and occurence of mission outcome per orbit type

Calculating the number and occurrence of each orbit

Creating a landing outcome label from Outcome column

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

No. of launch sites

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

No. of occurence of mission outcome per orbit type

```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
```

No. of occurrence of each orbit

	FlightNumber	Orbit	LaunchSite	Outcome
0	1	LEO	CCAFS SLC 40	None None
1	2	LEO	CCAFS SLC 40	None None
2	3	ISS	CCAFS SLC 40	None None
3	4	PO	VAFB SLC 4E	False Ocean
4	5	GTO	CCAFS SLC 40	None None
5	6	GTO	CCAFS SLC 40	None None
6	7	ISS	CCAFS SLC 40	True Ocean
7	8	LEO	CCAFS SLC 40	True Ocean

	FlightNumber	Orbit	LaunchSite	Outcome	Class
0	1	LEO	CCAFS SLC 40	None None	0
1	2	LEO	CCAFS SLC 40	None None	0
2	3	ISS	CCAFS SLC 40	None None	0
3	4	PO	VAFB SLC 4E	False Ocean	0
4	5	GTO	CCAFS SLC 40	None None	0
5	6	GTO	CCAFS SLC 40	None None	0
6	7	ISS	CCAFS SLC 40	True Ocean	1
7	8	LEO	CCAFS SLC 40	True Ocean	1

EDA with SQL

Loaded the dataset into the corresponding table in a Db2 database

Executed SQL queries to get required information from the data corresponding to-

- Launch Site
- Payload Mass
- Booster Version
- Mission Outcome
- Booster Landing

(Results are shown in Result section)

EDA with Visualization

Performed Exploratory Data Analysis by Visualizing the relationship between-

- Flight Number and Launch Site
- Payload and Launch Site
- Success rate of each orbit type
- Flight Number and Orbit type
- Payload and Orbit type
- Visualize the launch success yearly trend
- Selecting the features that will be used in success prediction

(Results are shown in Result section)

EDA and interactive visual analytics- (Launch Site location Analysis)

Marked all launch sites on a map using Folium

Marked the success/failed launches for each site on the map

Calculated the distances between a launch site to its proximities such as-

Railway	Highway	Coastline	Cities

(Results are shown in Result section)

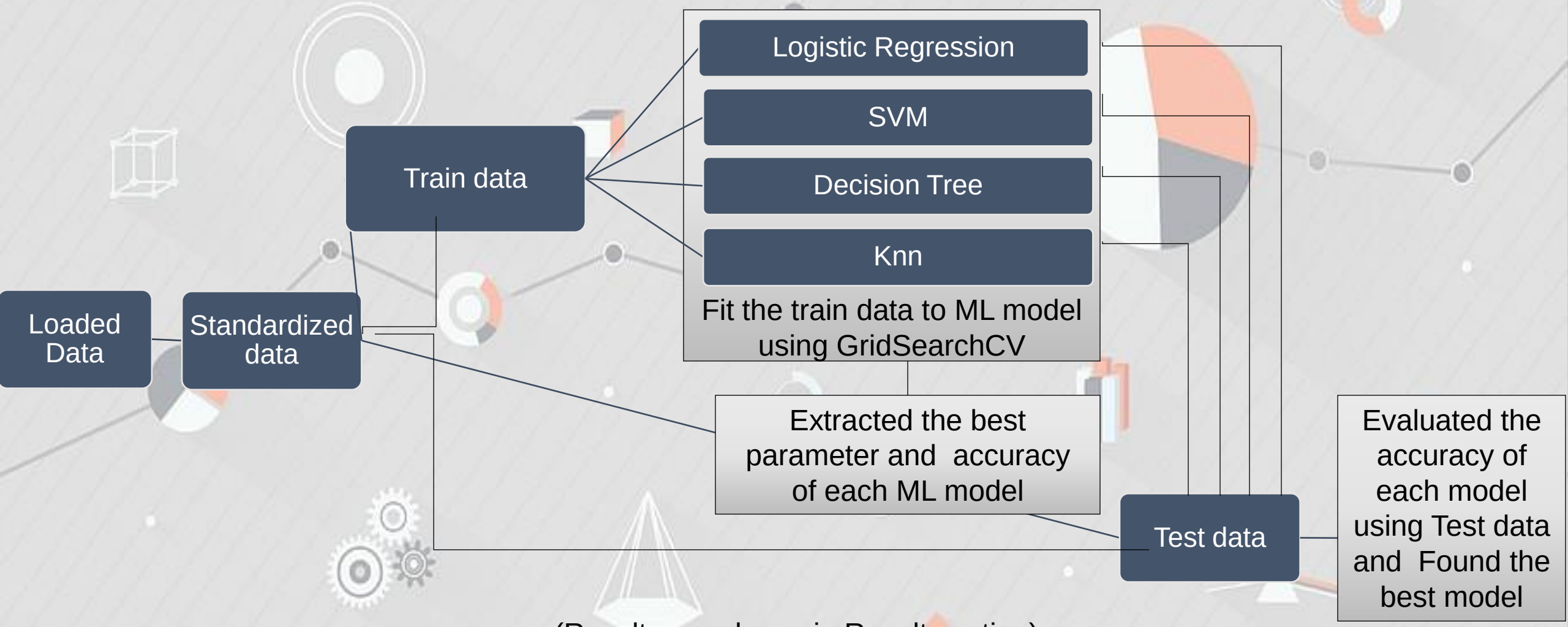
EDA and interactive visual analytics- (Launch Record Dashboard)

Built a Plotly Dash application for users to perform interactive visual analytics on SpaceX launch data in real-time by-

Using Pie chart which shows total Success launch by launch site	Using Range slider and Scatter plot to show correlation between Payload and Success for all launch sites

(Results are shown in Result section)

Predictive Analysis



(Results are shown in Result section)

RESULTS and INSIGHTS

- EDA with SQL results
- EDA with visualization insights
- Interactive map with Folium results insights
- Plotly Dash dashboard results insights

EDA with SQL results (1)

- Unique launch sites in the space mission

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

EDA with SQL results (2)

- 5 records where launch sites begin with the string '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

EDA with SQL results (3)

- Total number of successful and failure mission outcomes

sucessful	faliure
100	1

- Names of the booster_versions which have carried the maximum payload mass

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

EDA with SQL results (4)

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

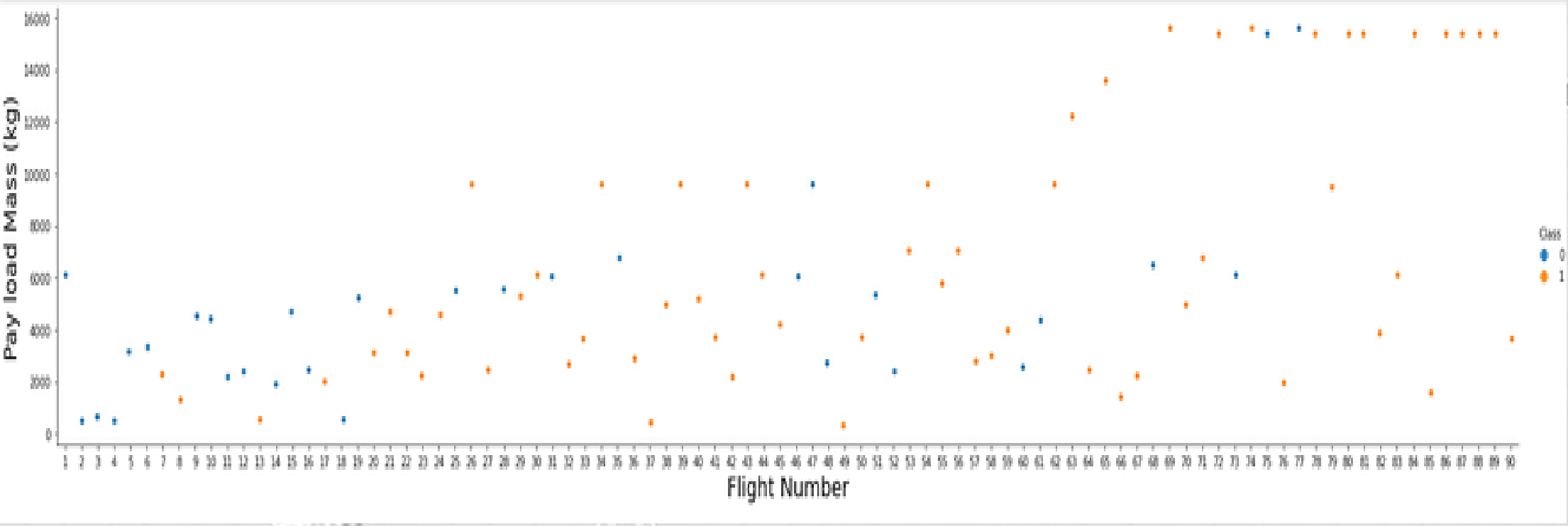
Landing _Outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing _Outcome	COUNT
Failure (drone ship)	5
Success (ground pad)	3

EDA with visualization Result and insights (1)

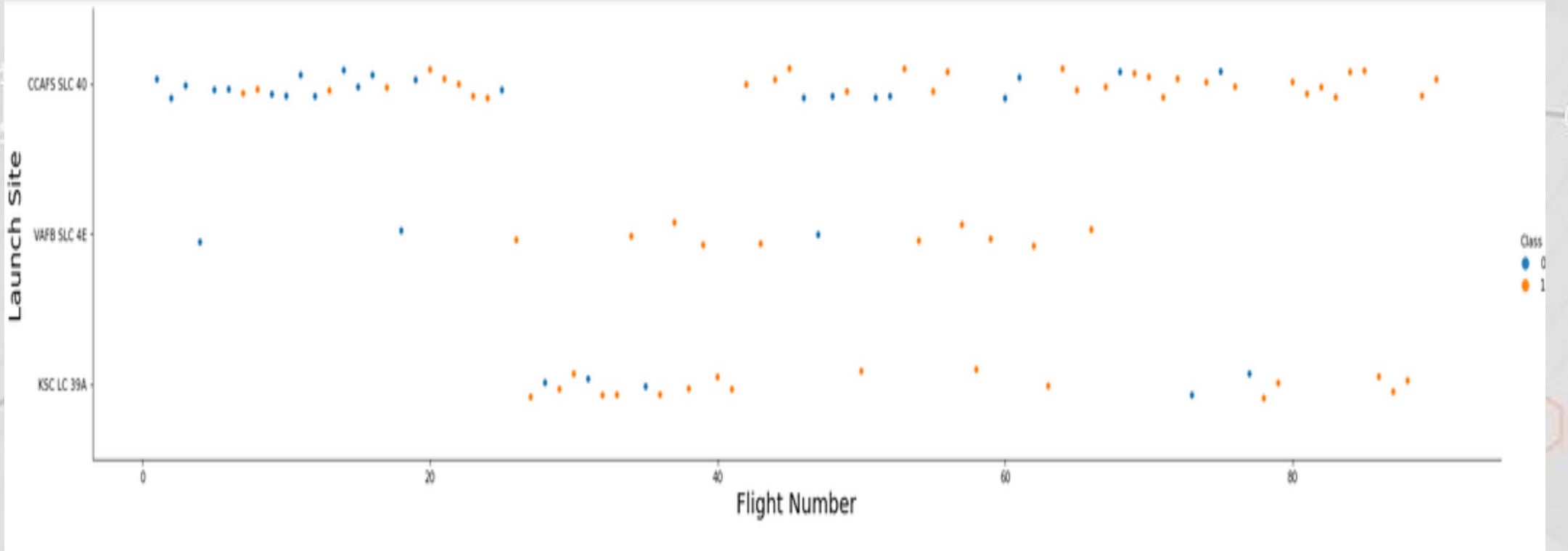
- Relationship between FlightNumber vs. PayloadMass



As the flight number increases, the first stage is more likely to land successfully and more massive the payload, less likely the first stage will return

EDA with visualization Result and insights (2)

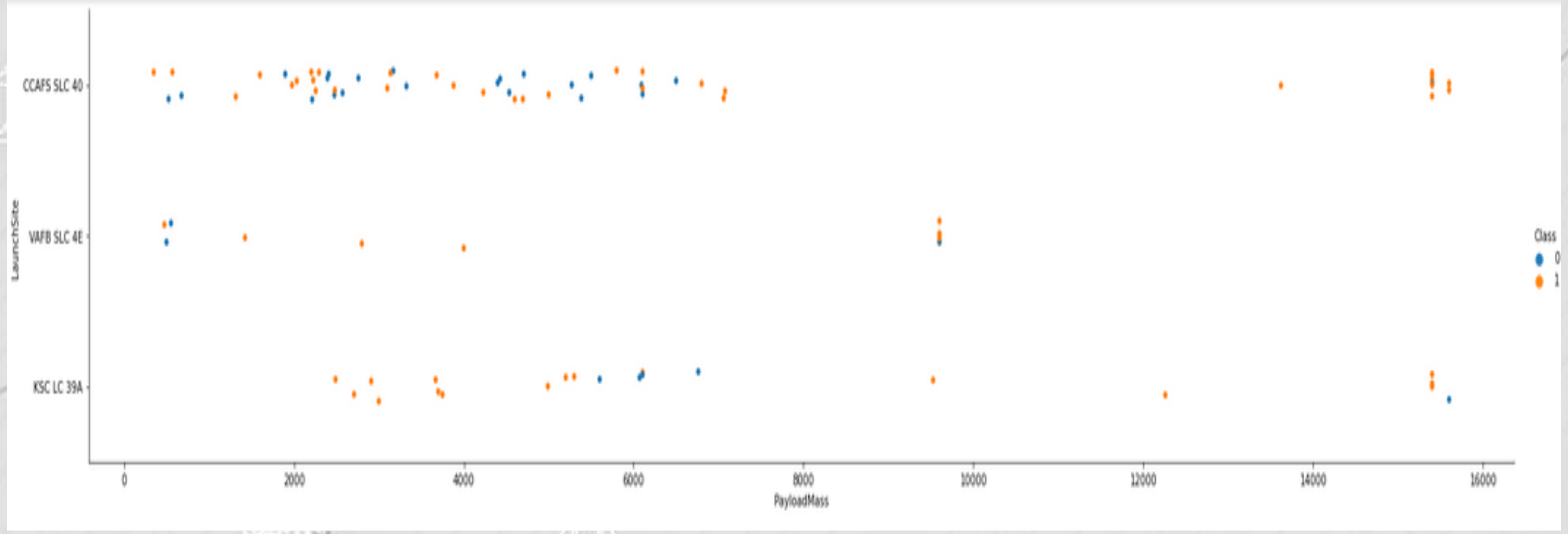
- Relationship between Flight Number and Launch Site



As the flight number increases, the first stage is more likely to land successfully, Launch site CASF SLC 40 has most failure landings than other sites.

EDA with visualization Result and insights (3)

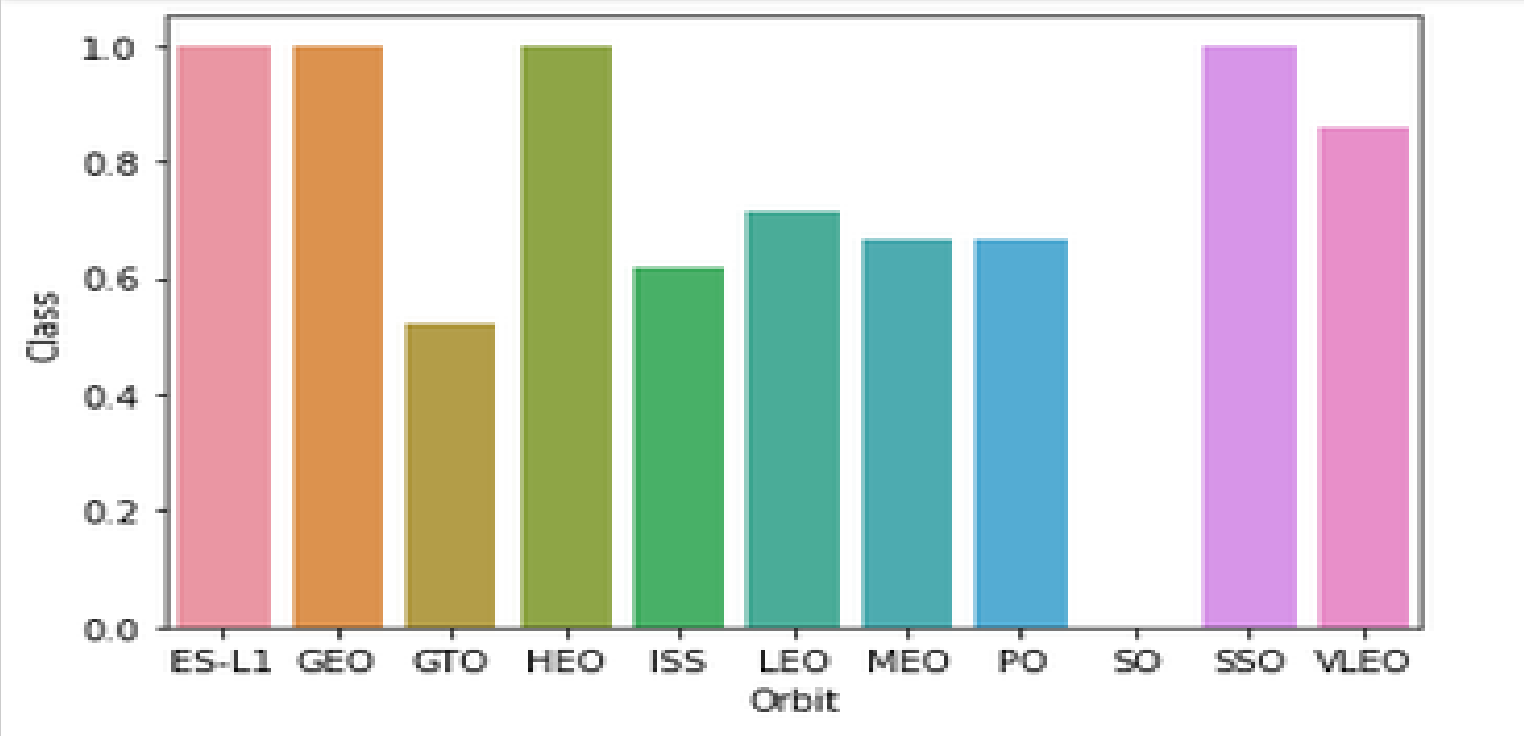
- Relationship between Payload and Launch Site



In VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

EDA with visualization Result and insights (4)

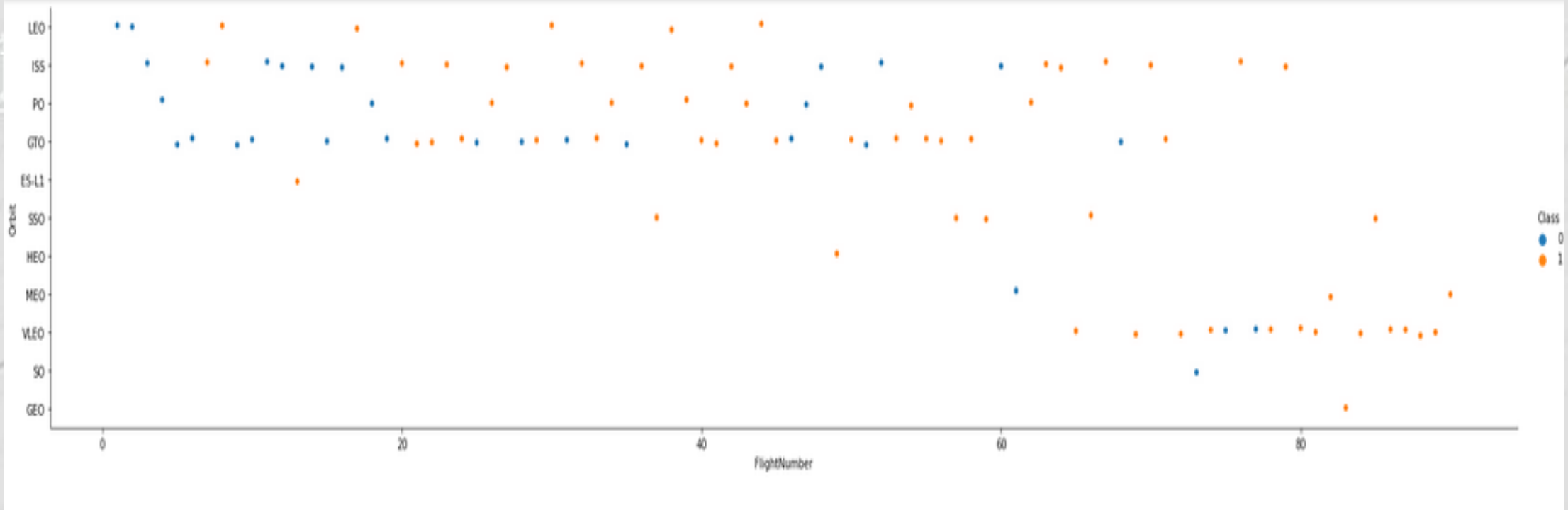
- Relationship between success rate of each orbit type



ES-L1, GEO, HEO and SSO orbits have maximum success rate where as GTO has lowest success rate.

EDA with visualization Result and insights (5)

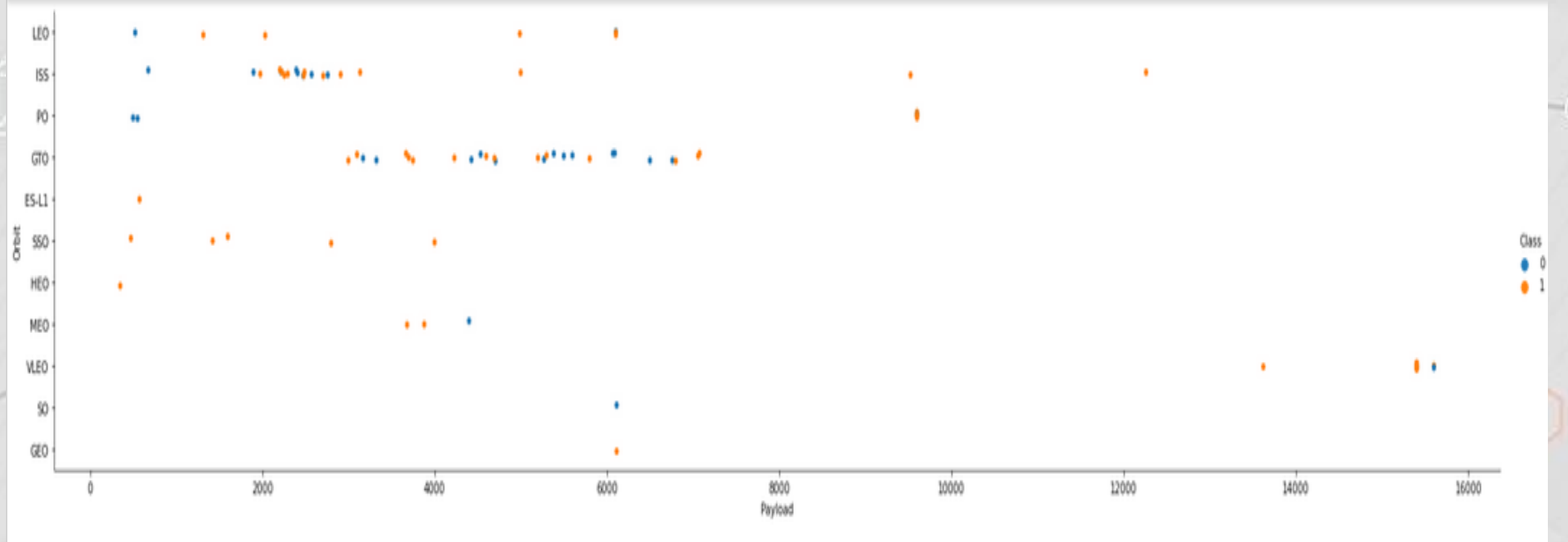
- Relationship between FlightNumber and Orbit type



The LEO orbit's success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

EDA with visualization Result and insights (6)

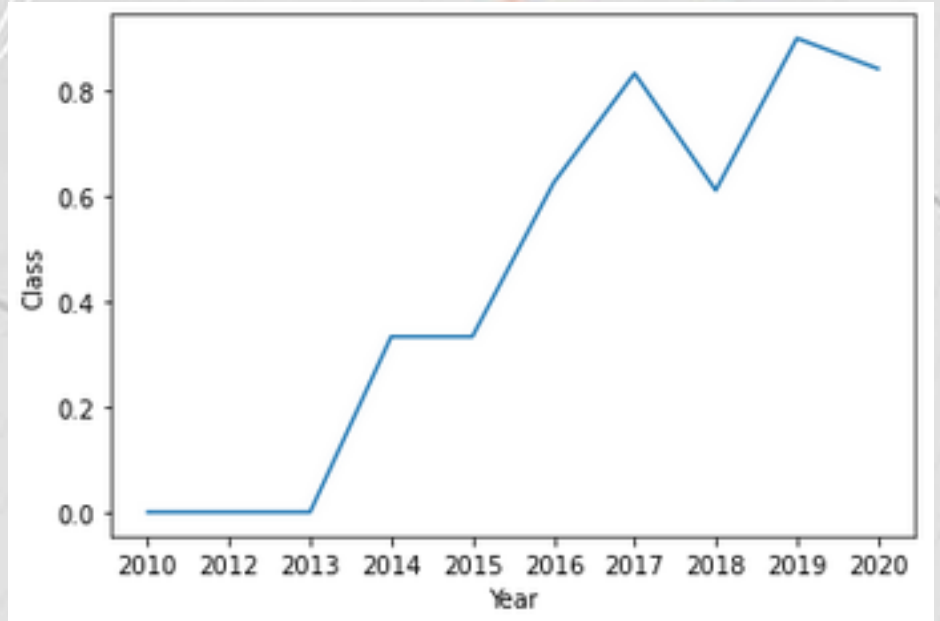
- Relationship between Payload and Orbit type and Orbit type



With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS orbits.

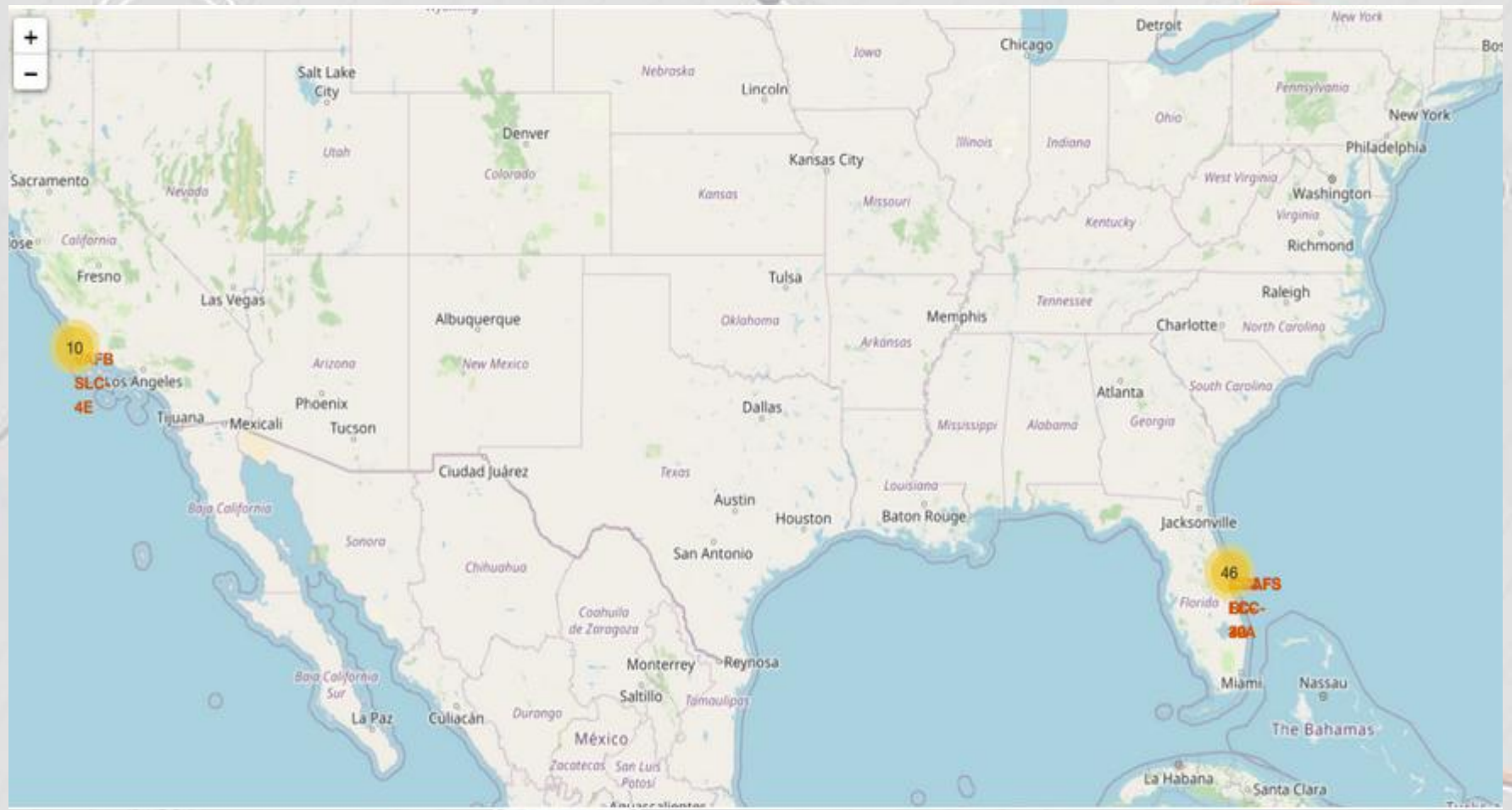
EDA with visualization Result and insights (7)

- Launch success yearly trend



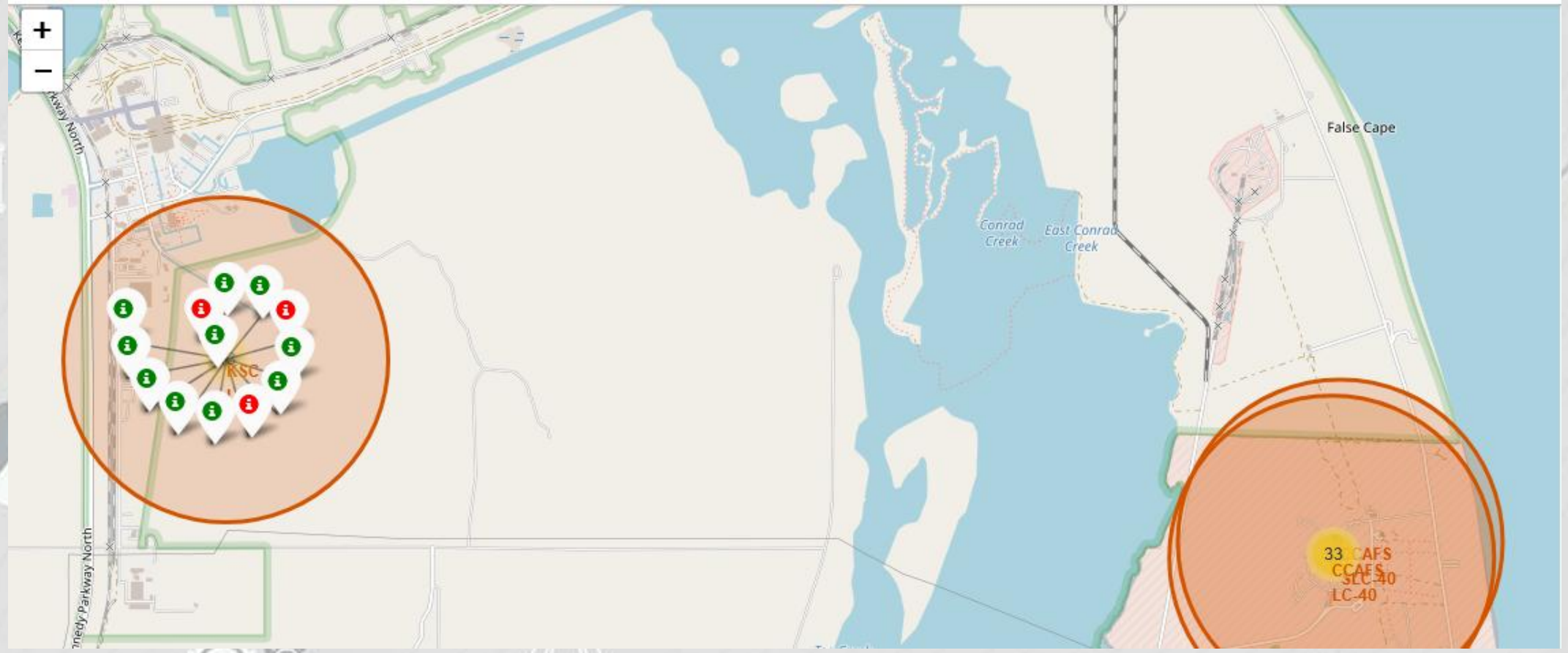
The launch success rate since 2013 kept increasing till 2020

Interactive map with Folium (Launch Site Location) (1)



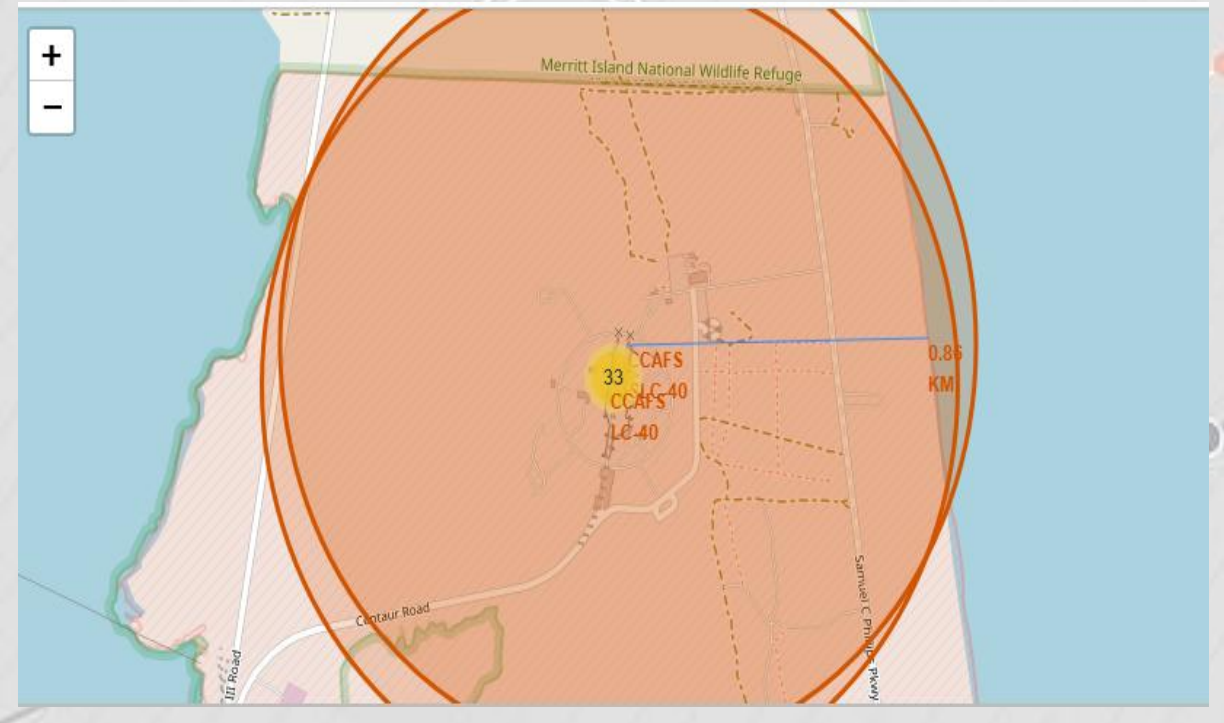
All launch sites are in proximity to both the Equator line and the coast

Interactive map with Folium (Launch Site Location) (3)



Launch site KSC LC-39A have relatively high success rates

Interactive map with Folium (Launch Site Location) (4)

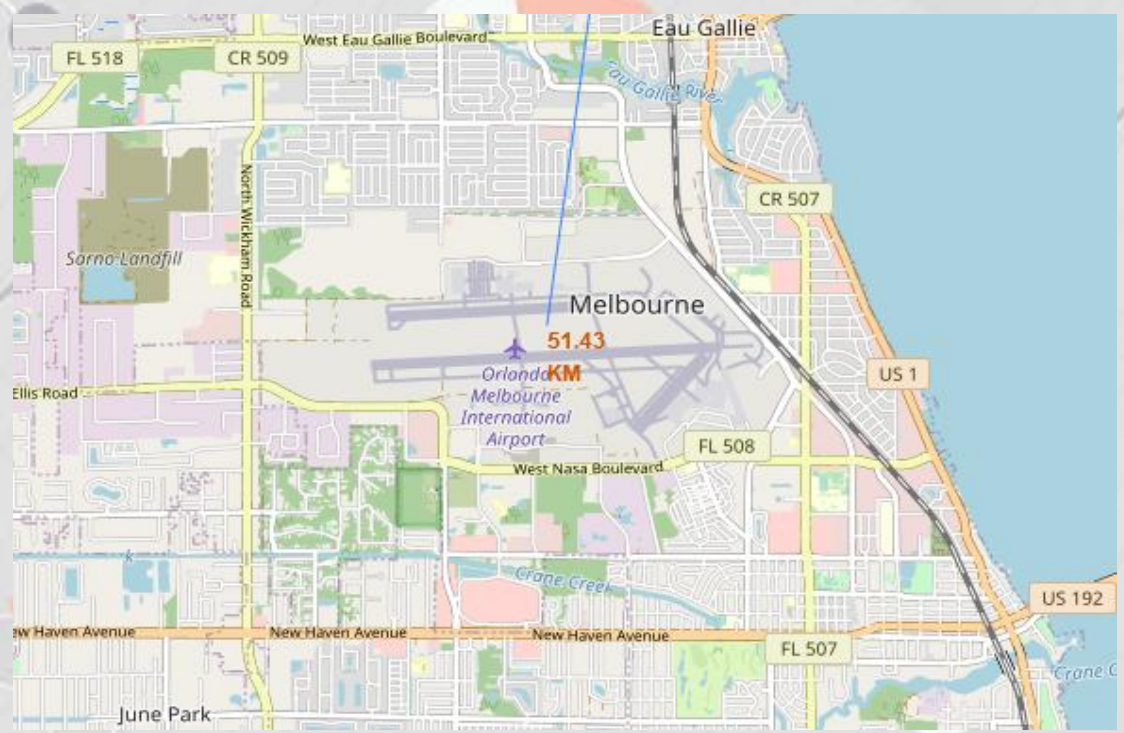


Launch sites are in close proximity to highways. For e.g. launch site CCAFS SLC-40 is 0.86km away from coastline

Launch sites are in close proximity to highways. Launch site CCAFS SLC-40 is 0.58 km away from coastline

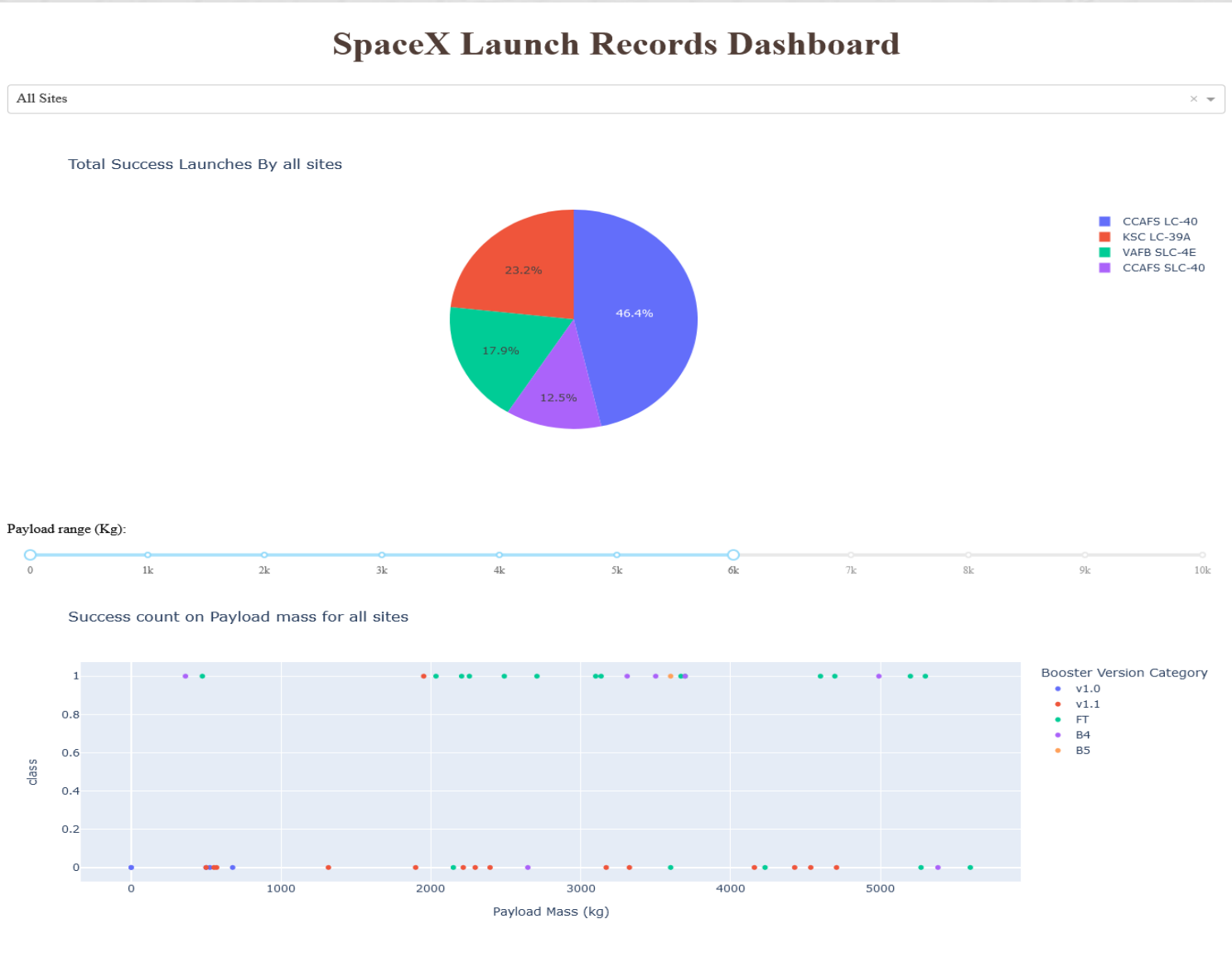
Interactive map with Folium (Launch Site Location) (5)

Launch sites are in close proximity to railways. e.g. launch site CCAFS SLC-40 is 1.28km away from railways



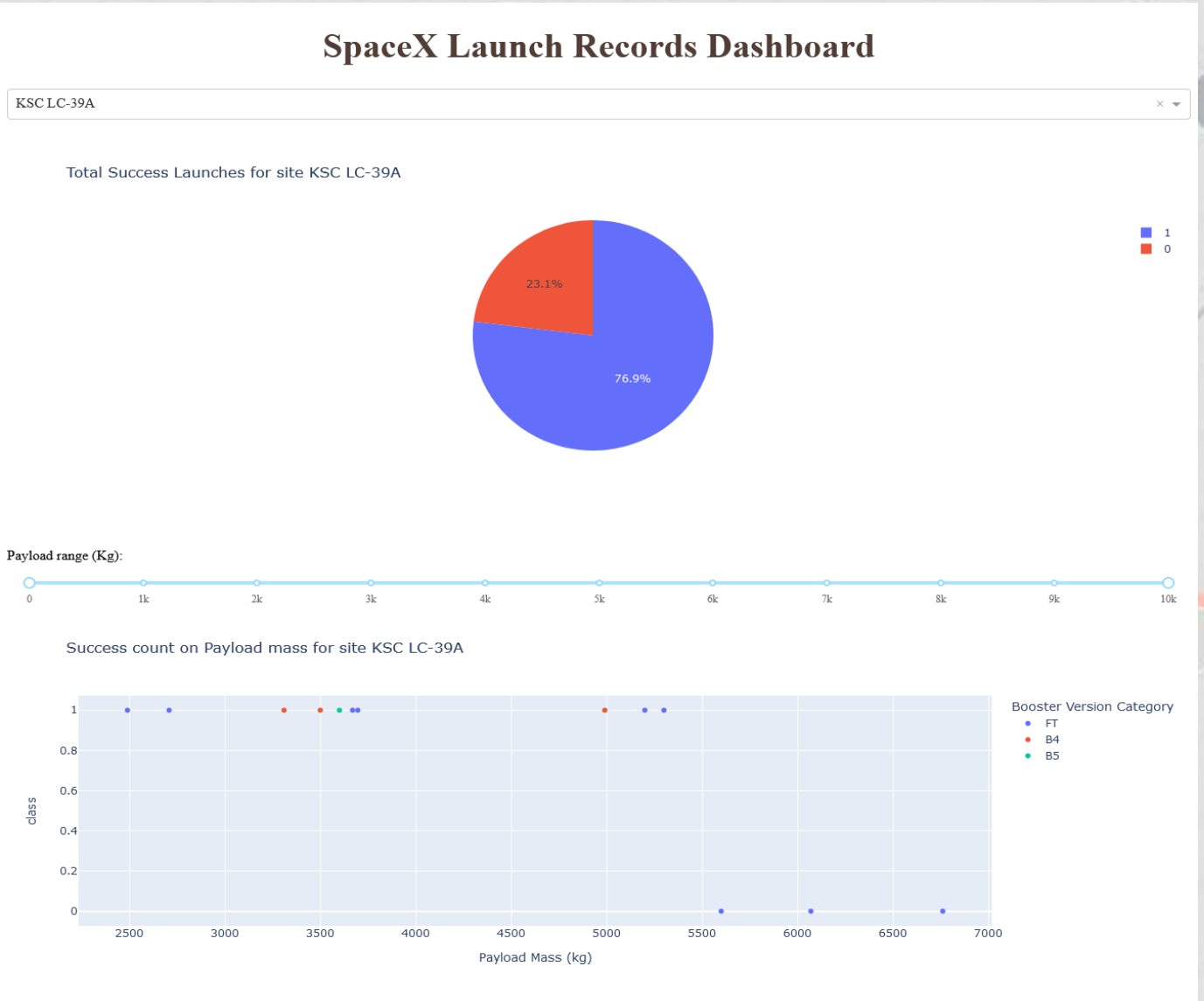
Launch sites keep certain distance away from cities for e.g. site CCAFS SLC-40 launch site is 51.43 km away from city Melbourn

Plotly Dash dashboard (1)



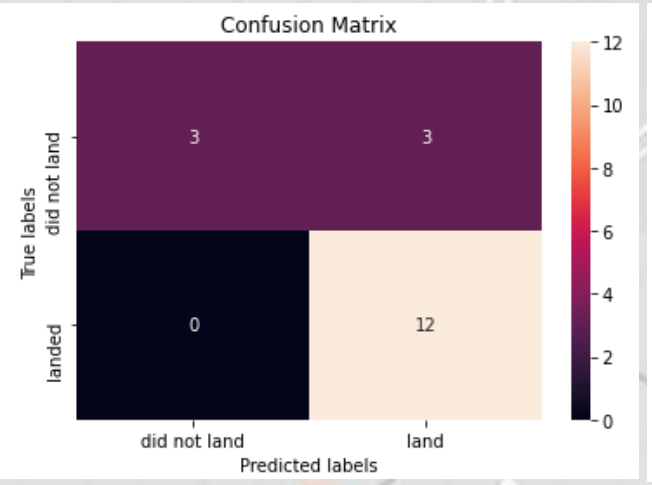
Dashboard showing success launch and success count on all launch site with payload range 0 to 6k

Plotly Dash dashboard (2)

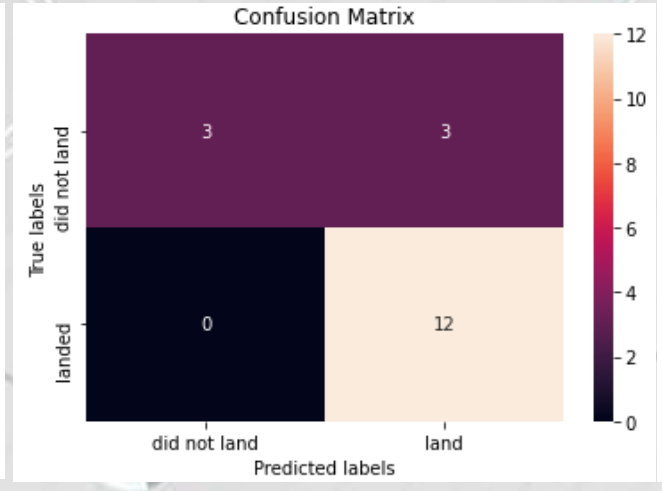


Dashboard shows launch site KSC LC-3A has highest success launch rate as 76.9% with success count on payload mass as 6750k.

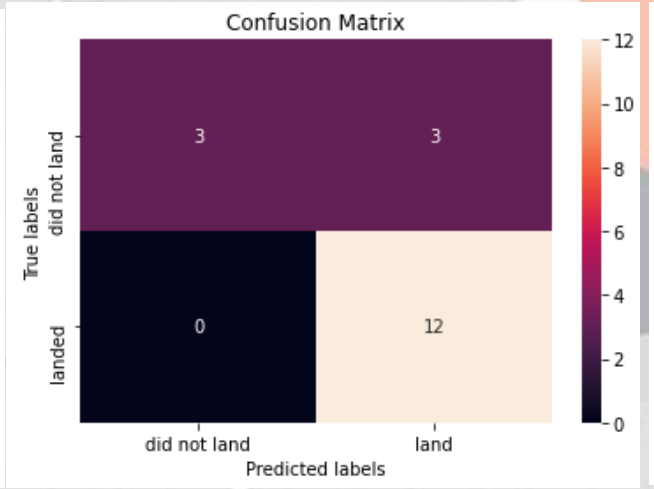
Predictive Analysis Result and Insights (1)



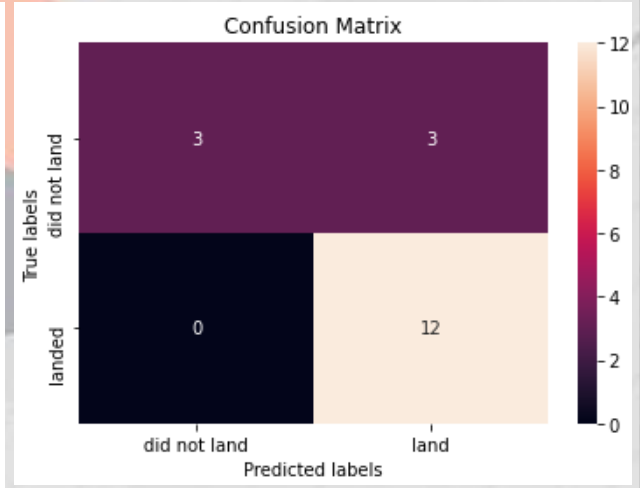
Logistic Regression



SVM



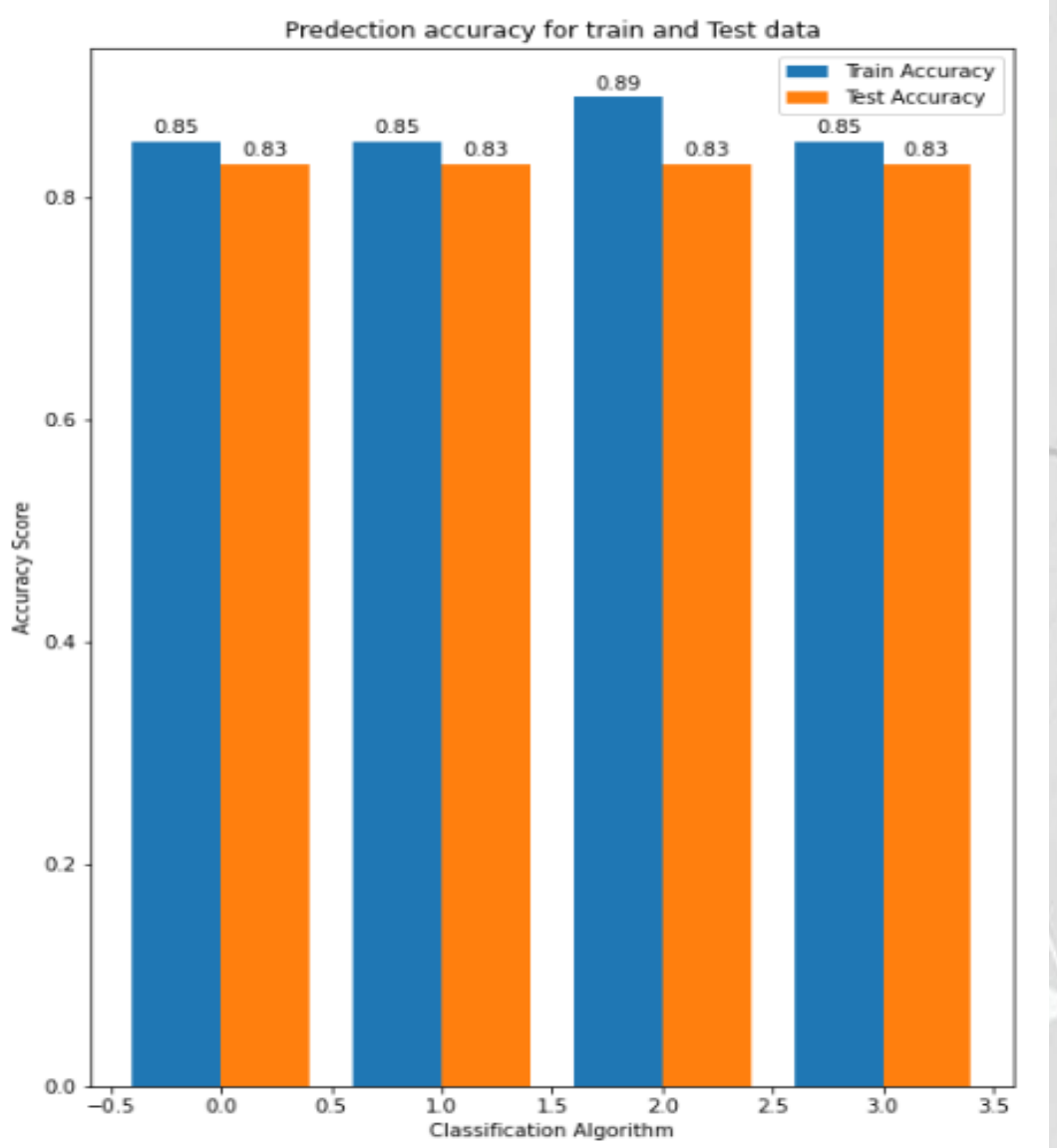
Decision Tree



KNN

Above are the confusion matrix for 4 classification algorithm used which shows all 4 can distinguish between the different classes but have detected false positive values. i.e. predicted 3 rocket landing which actually did not land.

Predictive Analysis Result and Insights (2)



- Graph show Logistic Regression, SVM and KNN classification algorithm preformed same with train and test data with 83% accuracy.
- Decision Tree, which fit train data slightly better i.e. 89% accuracy but accuracy for test data is same as other methods.

CONCLUSION

- We have determined if the first stage will land successfully using model created with Decision Tree algorithm giving 83% accuracy by using data insights like:
  - The first stage is more likely to land successfully in all launch site except for Launch site CASF SLC 40.
  - The launch success rate since 2013 kept increasing till 2020
  - ES-L1, GEO, HEO and SSO orbits have maximum success rate and
  - Launch site KSC LC-3A has highest success launch rate

- This information can be used to bid against SpaceX for a rocket launch by determine the launch cost.

THANK YOU