IBM DATA SCIENCE PROJECT

SpaceX Reusable Rocket

01

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Data Collection & Data Wrangling	Exploratory Analysis	Predictive Models
Collect data with SpaceX REST API and Web Scraping using BeautifulSoup. Data cleansing using python libraries.	Using SQL commands and visualization like bar chart, scatter plot. Using interactive map and dashboard to visualize more clearly.	Machine Learning Models used: • Logistic Regression • Support Vector Machine • K-Nearest Neighbors • Decision Trees All have 83.33% accuracy.

Introduction

- SpaceX

  an American aerospace manufacturer, space transportation services and communications company

- Cost per launch

  -Normal: $165 million
  -Falcon9: $62 million

Goal

To maximize success rate
and the best model to predict

SpaceX Falcon 9

Reusable Rocket Model

$62 million

per launch

Methodology

Data Collection

-Web Scraping
-SpaceX REST API

Data Wranging

-Removing irrelavant data
-Handling missing values
-Turning categorical variables into numerical variables

Exploratory Data Analysis using visualization

-Using visualization (scatter plot, bar chart) to present data

Methodology (continue)

Exploratory Data Analysis using SQL

-Using SQL commands to obtain the data insights

Interactive visual analytics

-Using Ploty Dash to create interactive dashboard and visualize to data
-Using Folium to create interactive maps

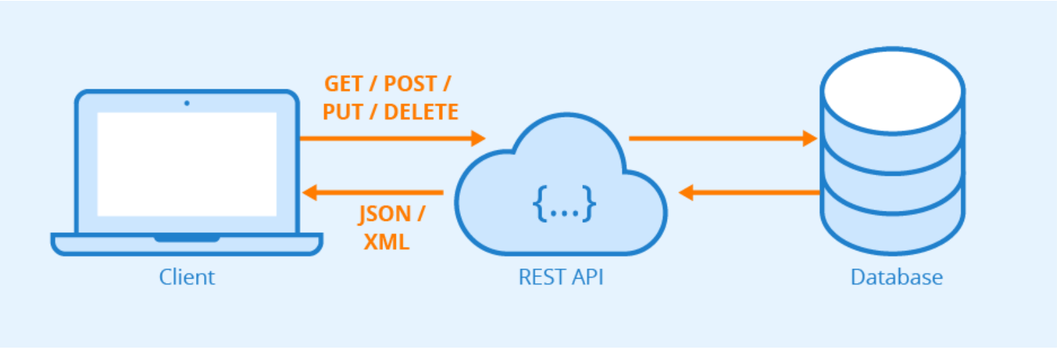
Prediction using classification models

-Using machine learning models such as KNN and SVM to predict the outcome
-Model evaluation

Data Collection

Using SpaceX REST API

:use get request to obtain the response and then convert it to data frame



Using Beautiful Soup

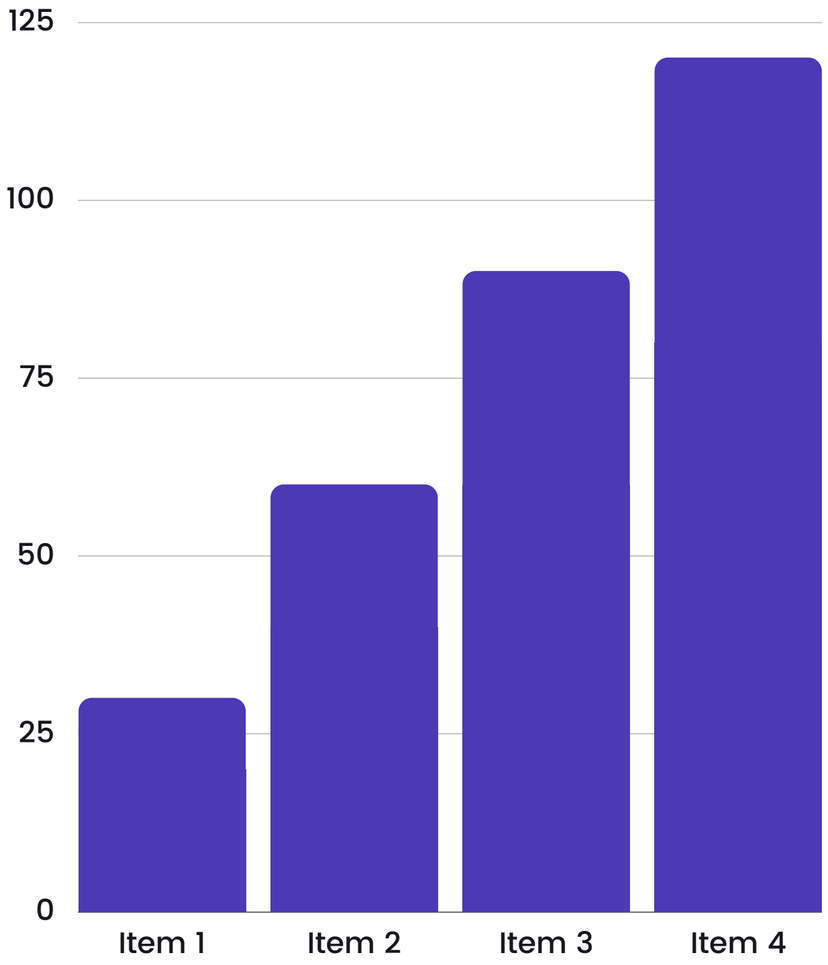
: use beautiful soup for web scraping from webpages to obtain the data

Data Wrangling

Filter the data frame
:only include Falcon 9 launches

Dealing with Missing Values
: replace them by the mean

Converting caterical variable into numerical
successful: 1
unsuccessful: 0



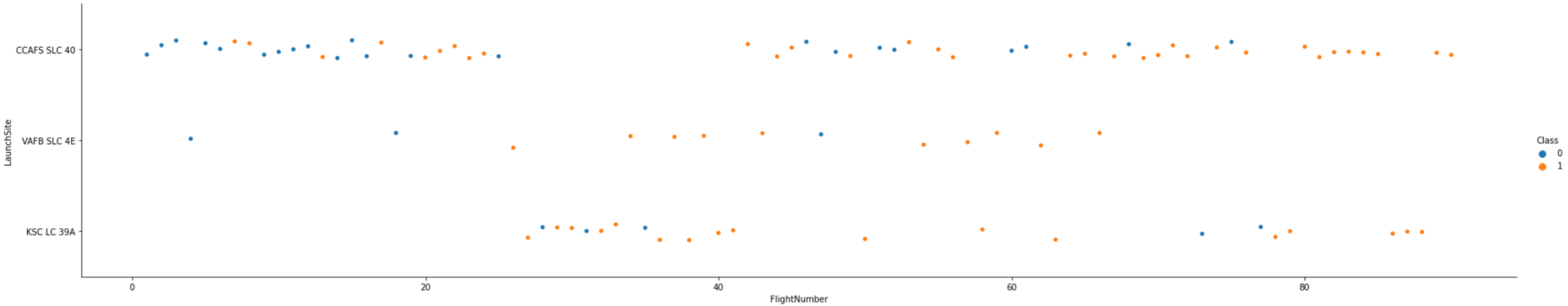
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

EDA with visualization

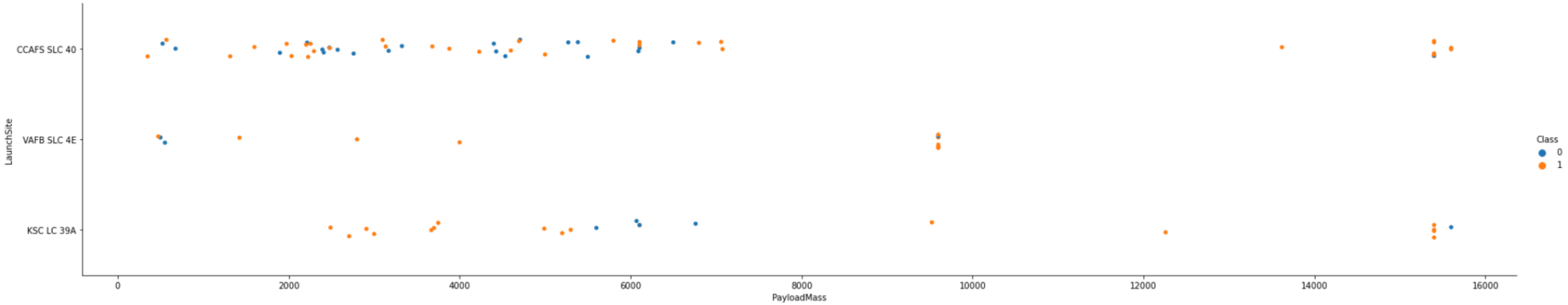
- Scatter plot between Flight Number and Launch Site
- Scatter plot between Payload and Launch Site
- Bar chart between success rate of each orbit type
- Scatter plot between FlightNumber and Orbit type
- Scatter plot between Payload and Orbit type
- Line chart of the launch success yearly trend

Flight Number vs. Launch Site



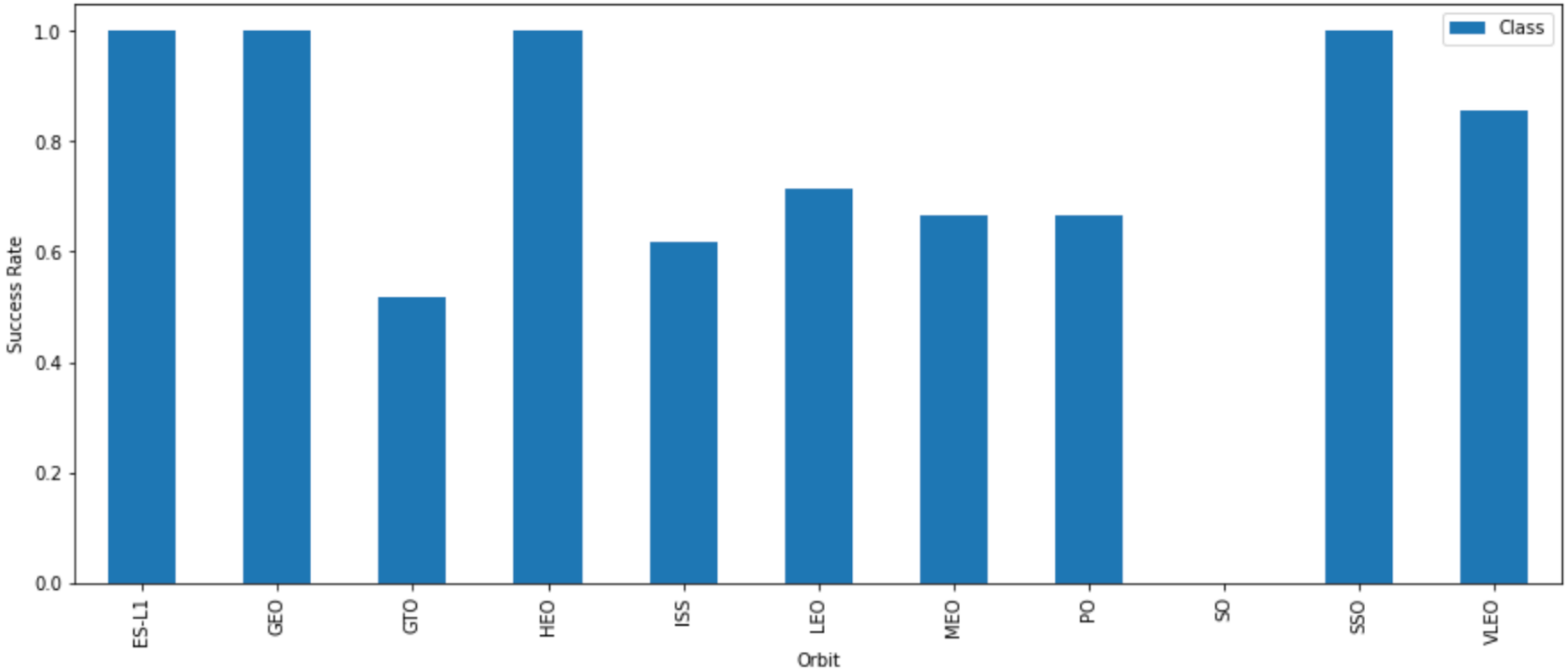
Higher flight number seems to have more success rate on the same launch site.

Payload vs. Launch Site



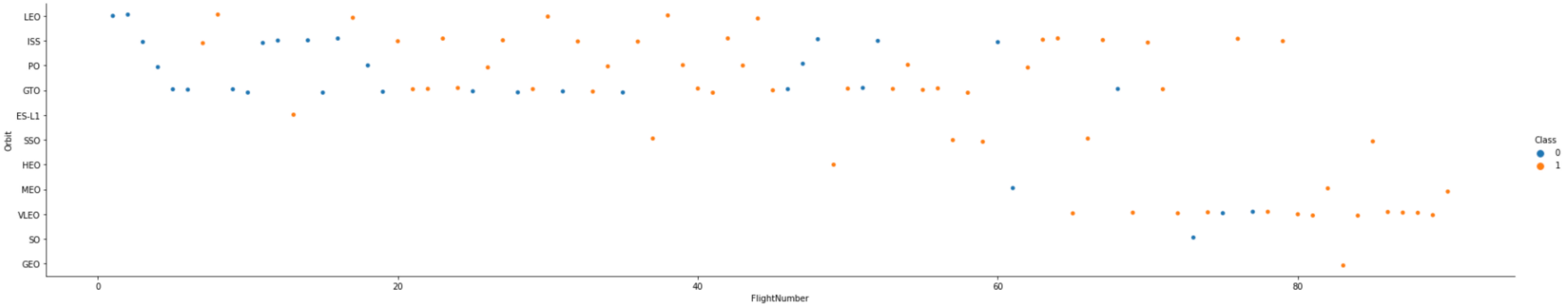
For CCAFS SLC 40 and VAFB SLC 4E, higher payload mass seems to be more successful. For KSC LC 39A, payload mass between 6,000-7,000 seems to be unsuccessful.

Success rate vs. Orbit type



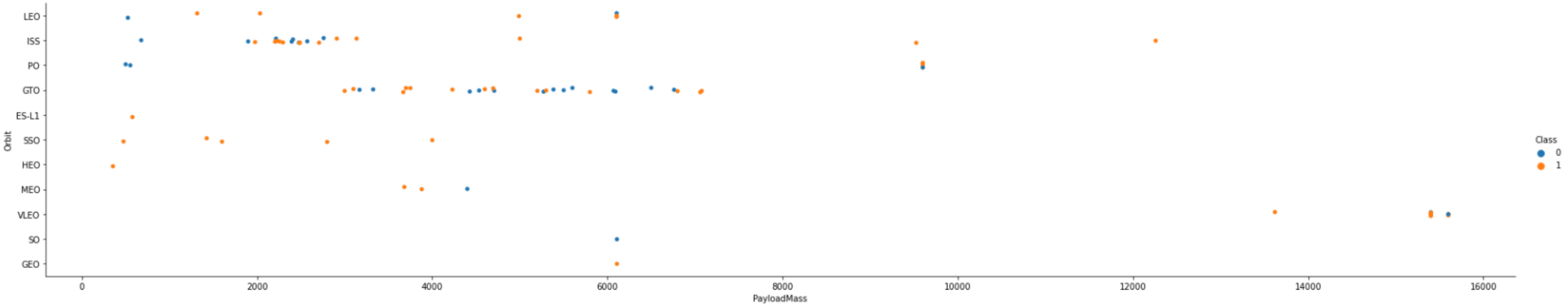
There are 4 orbit types that have the highest success rate, which are ES-L1, GEO, HEO, SSO.

Flight Number vs. Orbit type



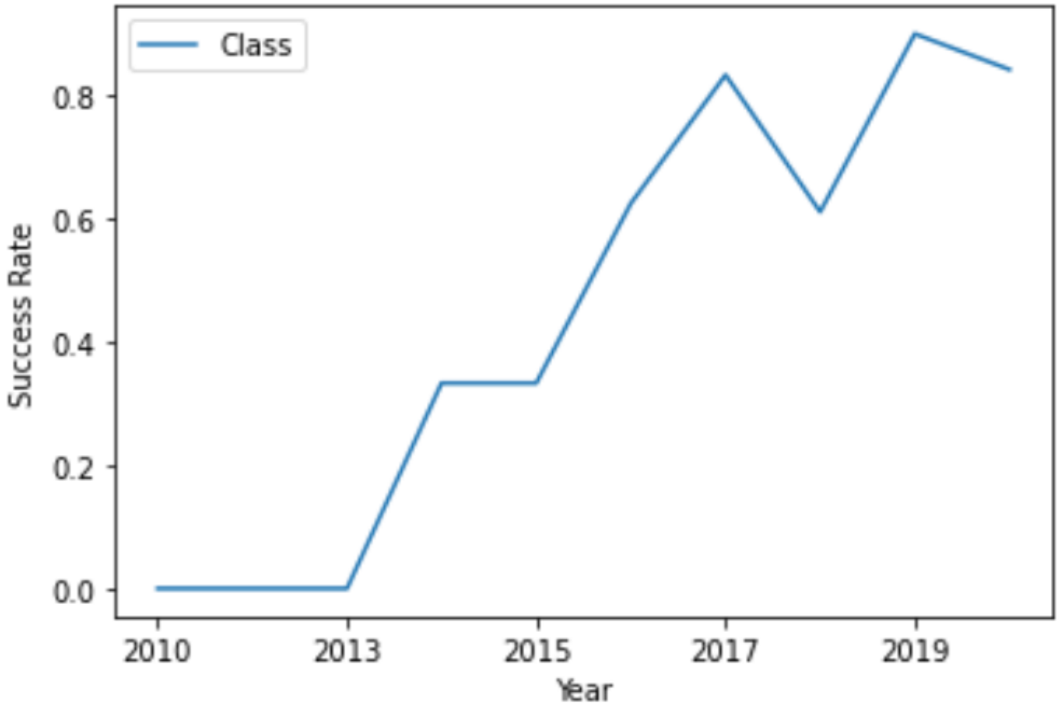
For GTO, less flight number seems to be unsuccessful on the same orbit type.

Payload vs. Orbit type



Higher payload mass seems to have negative influence with GTO orbit type.

Launch success yearly trend



Success rate's trend keeps increasing until 2020, which corresponds with the scatter plot that indicates that the higher flight number is, the higher success rate is.

EDA with SQL

Using SQL commands to gain more data insights

Task 1

Display the names of the unique launch sites in the space mission

```
%sql select DISTINCT launch_site from spacextbl
```

```
 * ibm_db_sa://xwx66212:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

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

EDA with SQL

Task 2

Display 5 records where launch sites begin with the string 'CCA'

```
%%sql select * from spacextbl
where launch_site like 'CCA%' limit 5
```

```
 * ibm_db_sa://xwx66212:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

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

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%%sql select sum(payload_mass__kg_) as sum from spacextbl
where customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://xwx66212:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

SUM
45596

Task 4

Display average payload mass carried by booster version F9 v1.1

```
%%sql select avg(payload_mass__kg_) as average from spacextbl
where booster_version like 'F9 v1.1%'
```

```
 * ibm_db_sa://xwx66212:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

average
2534

EDA with SQL

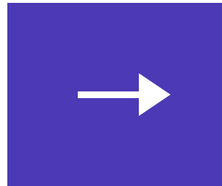
Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%%sql select min(date) from (select date from spacextbl
                             where landing_outcome = 'Success (ground pad)')
```

1
2015-12-22



EDA with SQL

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%%sql select booster_version from spacextbl
where payload_mass__kg_ between 4000 and 6000
```

```
 * ibm_db_sa://xwx66212:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

booster_version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1046.3
F9 B5B1054
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

EDA with SQL

Task 7

List the total number of successful and failure mission outcomes

```
%%sql select count(mission_outcome) as SUCESS_MISSION from spacextbl
where mission_outcome like 'Success%'
```

```
 * ibm_db_sa://xwx66212:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

sucess_mission
100

```
%%sql select count(mission_outcome) as FAILED_MISSION from spacextbl
where mission_outcome like 'Failure%'
```

```
 * ibm_db_sa://xwx66212:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

failed_mission
1

EDA with SQL

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%%sql select booster_version from spacextbl
where payload_mass__kg_ = (select max(payload_mass__kg_) from spacextbl)
```

```
 * ibm_db_sa://xwx66212:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:3119
8/bludb
Done.
```

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

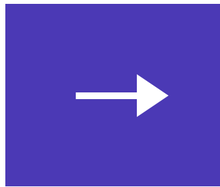
EDA with SQL

Task 9

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for the in year 2015

```
%%sql
Select SPACEXTBL.landing__outcome, SPACEXTBL.booster_version, SPACEXTBL.launch_site from SPACEXTBL
where SPACEXTBL.landing__outcome in ('Failure (drone ship)') and year(SPACEXTBL.Date) = 2015;
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40



EDA with SQL

Task 10

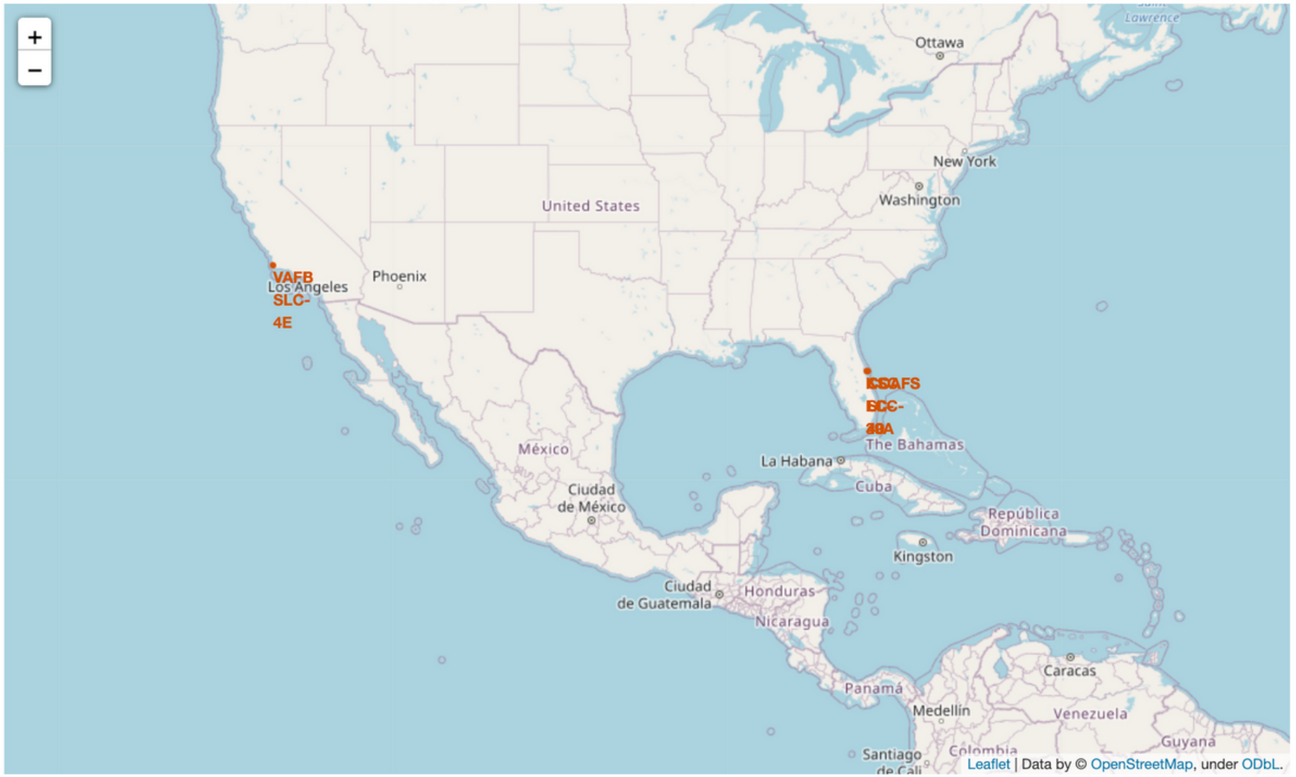
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%%sql
SELECT landing__outcome,COUNT(landing__outcome) as OccurenceValue FROM SPACEXTBL
where SPACEXTBL.Date between '2010-06-04' and '2017-03-20'
GROUP BY landing__outcome ORDER BY OccurenceValue desc;
```

landing__outcome	occurencevalue
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

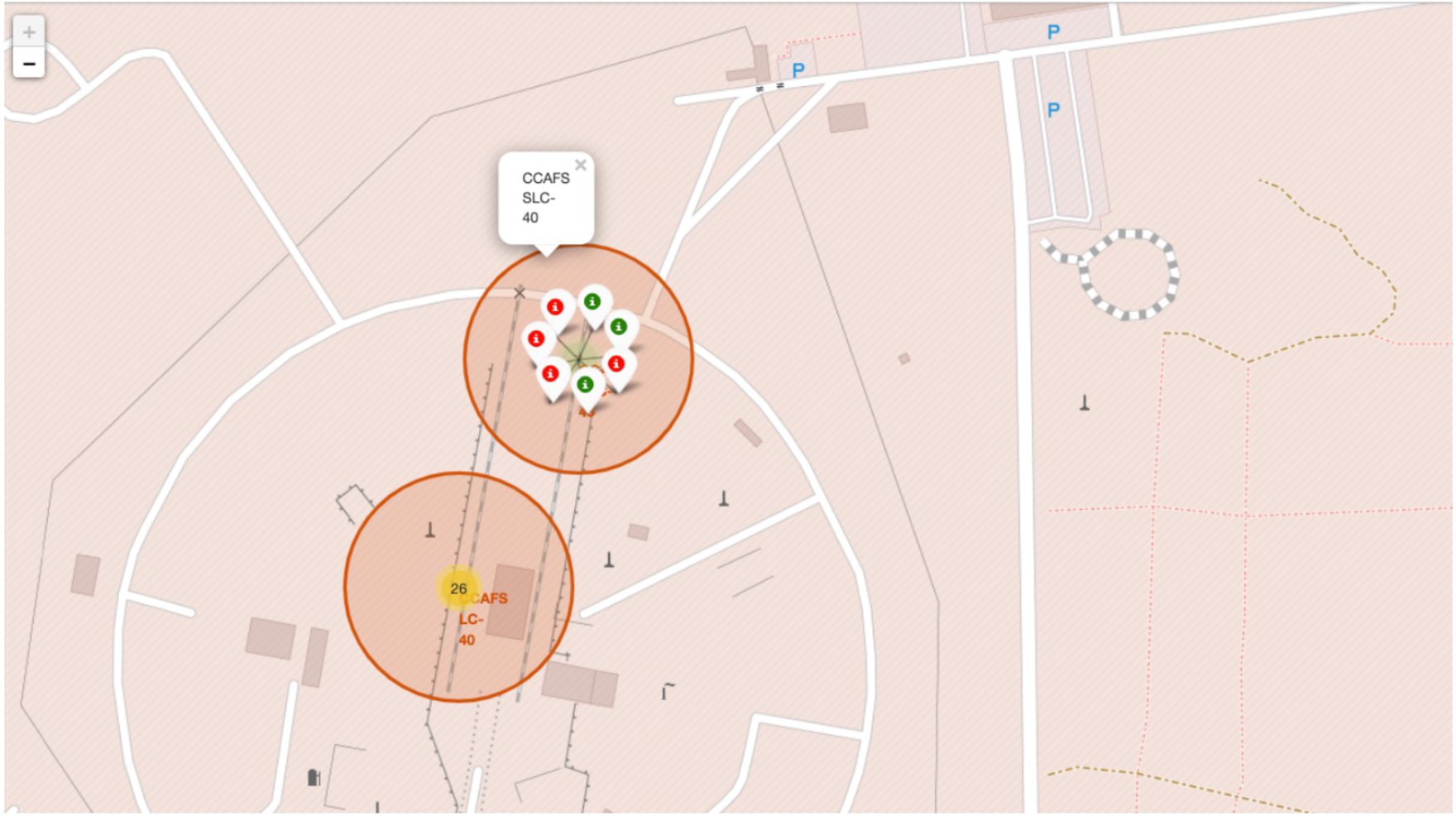
Build an interactive map with Folium

Locations of all launch sites

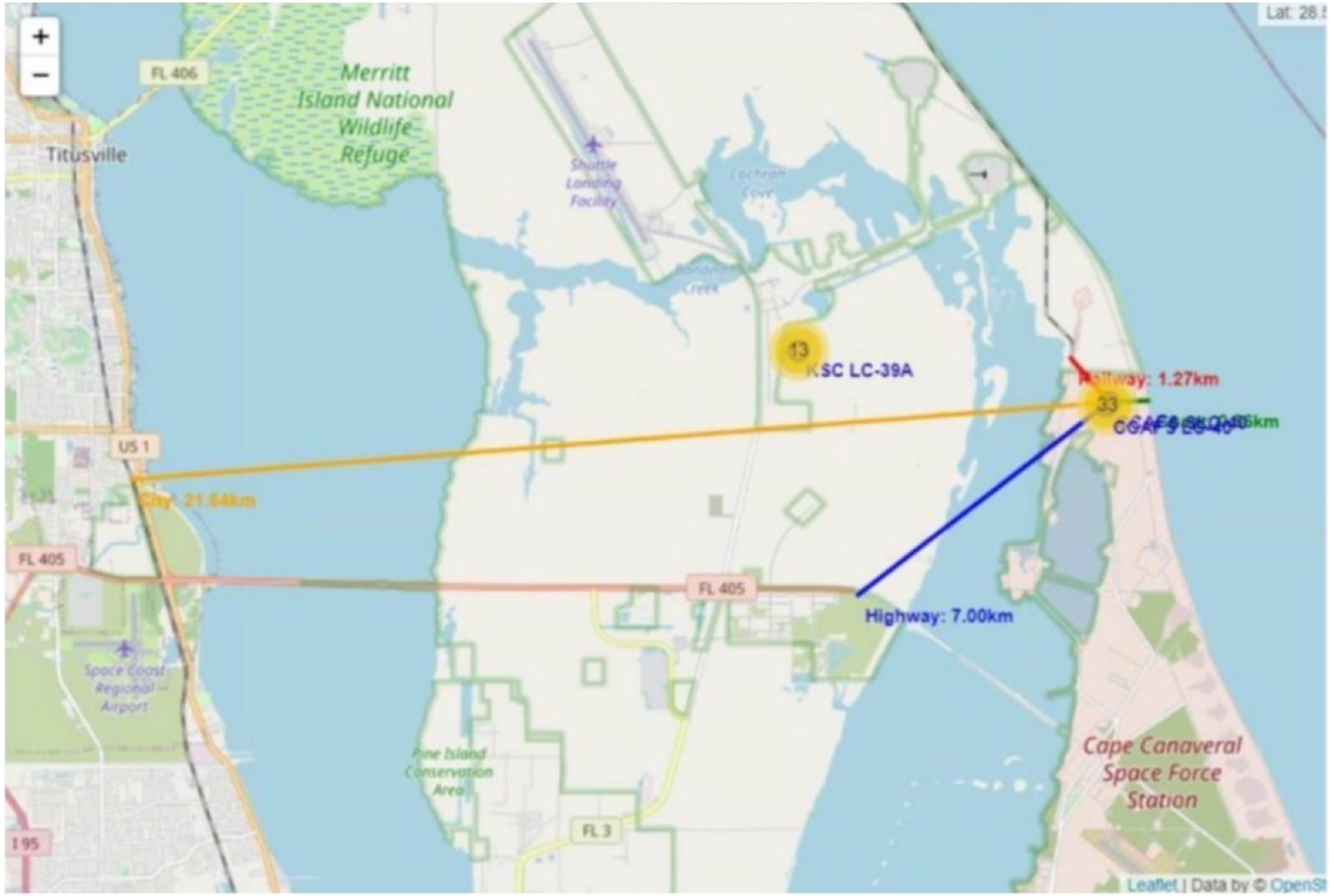


All locations of launch sites are near the ocean.

Folium Map

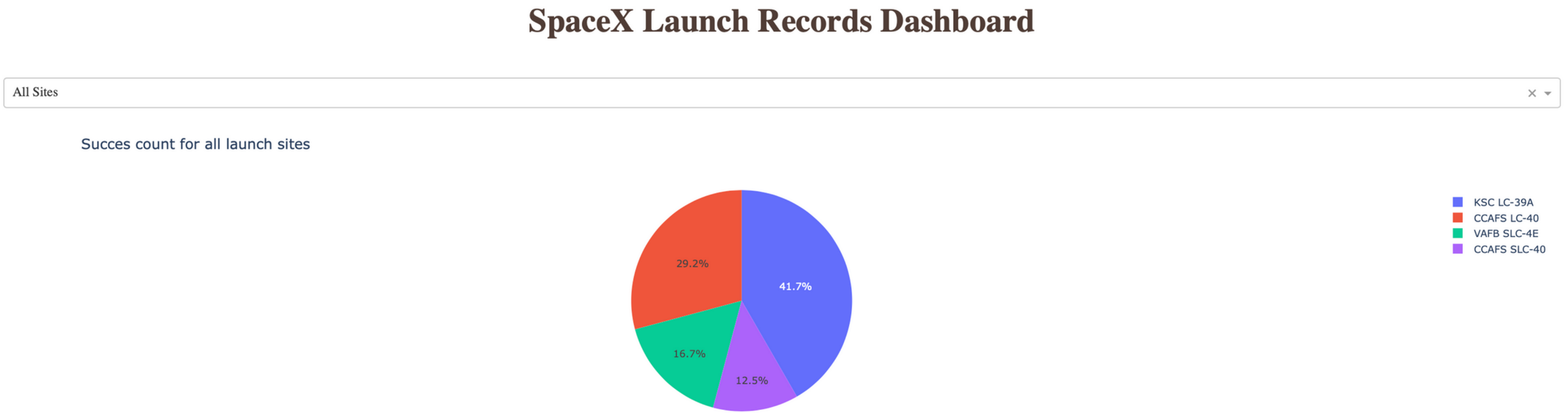


Folium Map


Merritt Island National Wildlife Refuge
Titusville
Shuttle Landing Facility
Cochran Cove
Banana Creek
13
KSC LC-39A
33
Highway: 1.27km
Highway: 7.00km
US 1
FL 405
FL 405
FL 406
Space Coast Regional Airport
Pine Island Conservation Area
FL 3
I 95
Cape Canaveral Space Force Station
Leaflet | Data by © OpenSt

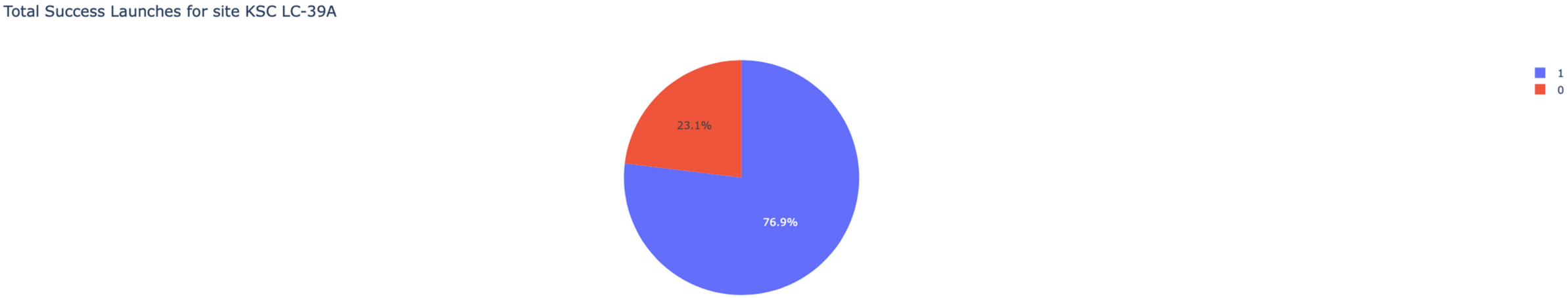
Build a dashboard

Using Plotly Dash



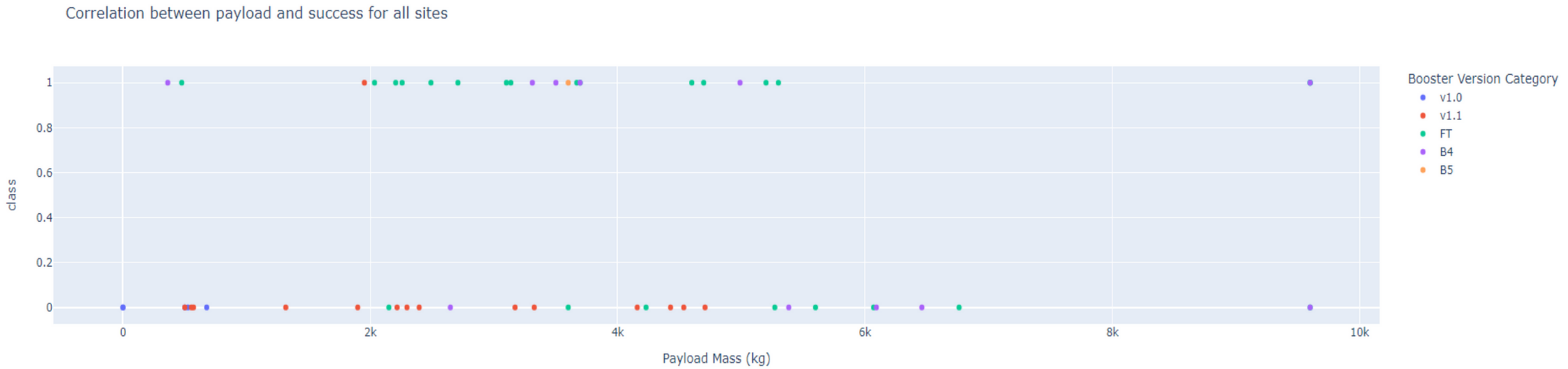
KSC LC-39A has the most successes with 41.7% of all combined.

Dashboard



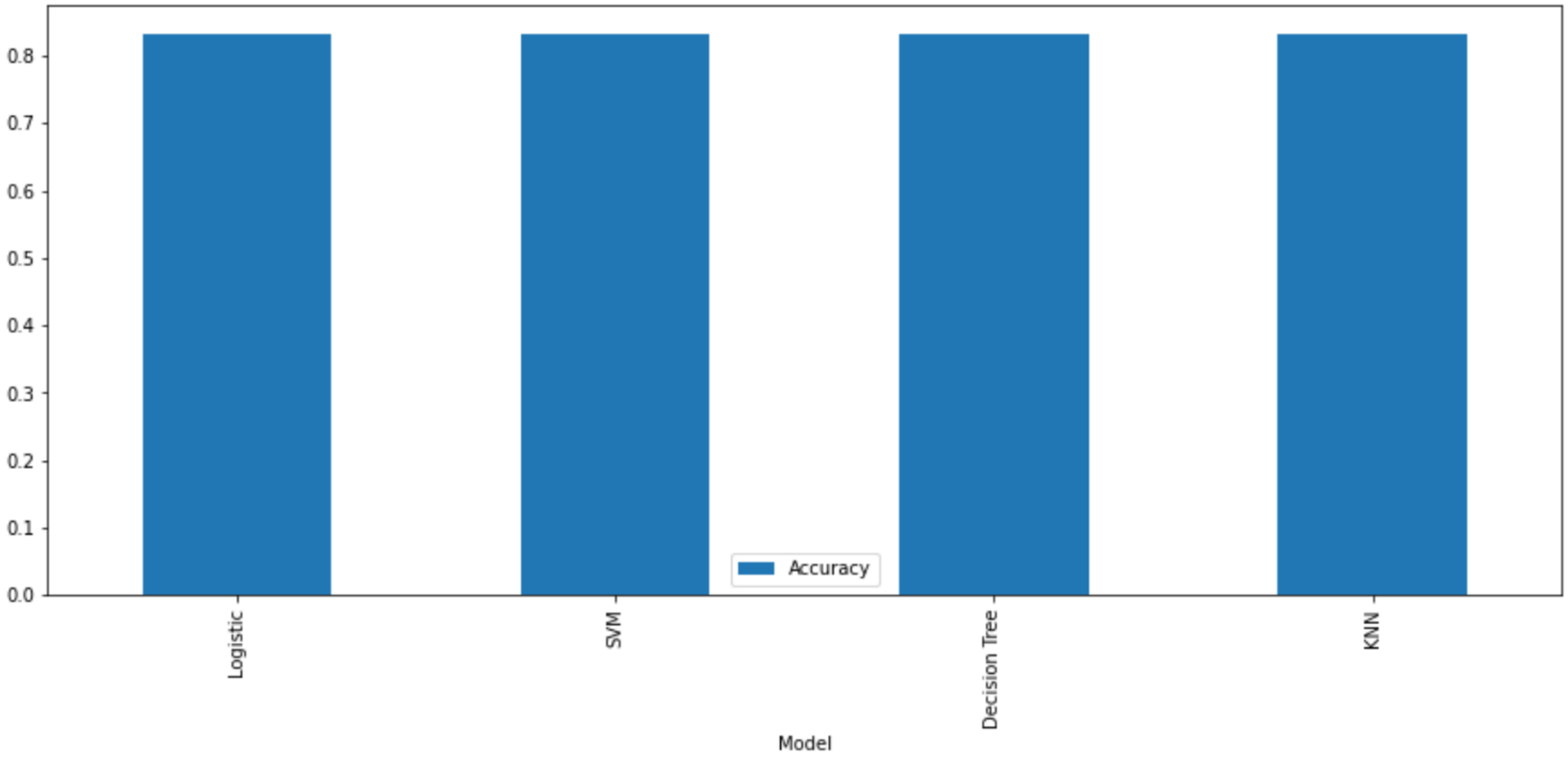
KSC LC-39A also has the most success rate of 76.9% of all operations launched at site KSC LC-39A.

Dashboard



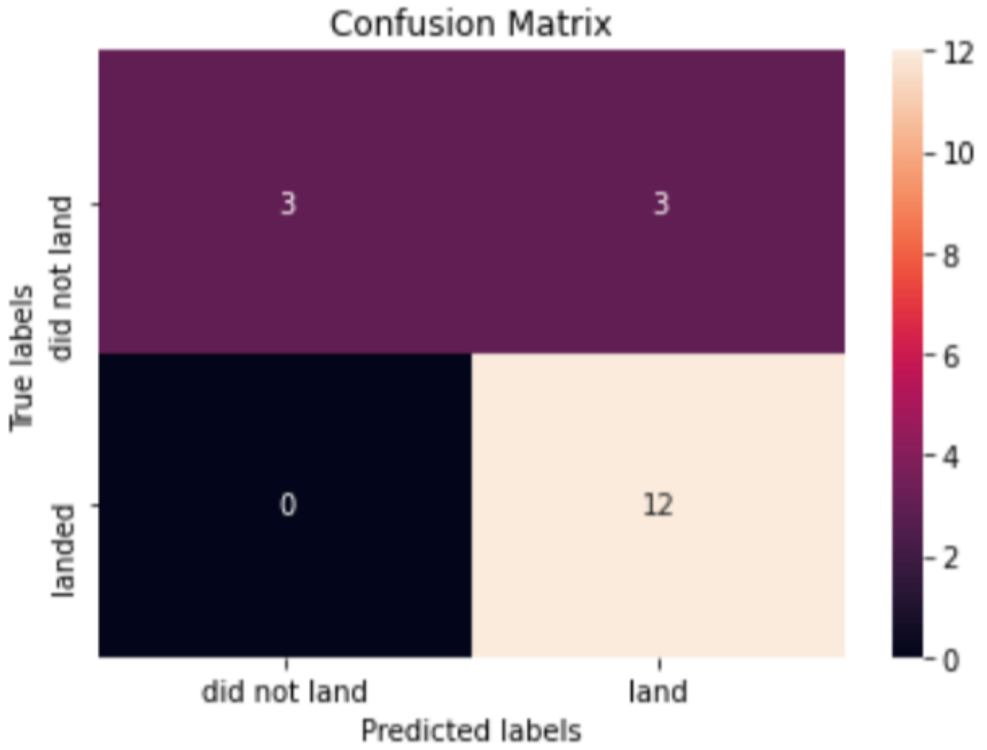
v1.1 booster version has a low success rate.

Predictive Analysis (Classification)



All models have the same accuracy of 0.83 .

Predictive Analysis (Classification)



This is the confusion matrix of all 4 models.

They have same confusion matrix with 18 sample tests.

Conclusions

Orbit type and payload mass

have influences on the success rate.

The success rate

keeps increasing yearly.

Launch Site locations

should be near the ocean.

Most success rate site

is KSC LC-39A

Machine Learning Models

All 4 models' ,namely LogReg, SVM, KNN, Tree, performances are similar for the data.