IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Francisco Machado e Costa
October 2023

Outline

Executive Summary

Data was extracted and explored. Several prediction models were built using python to predict launch success and first stage recovery

Study results

- Launch success rate has increased with time.
- Orbits **ES-L1**, **SSO**, **HEO** and **GEO** had the highest success rate.
- All machine learning algorithms we adequate for predicting launch outcome.
- A prediction accuracy of **83%** was achieved

Introduction

Background

- SpaceX 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. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

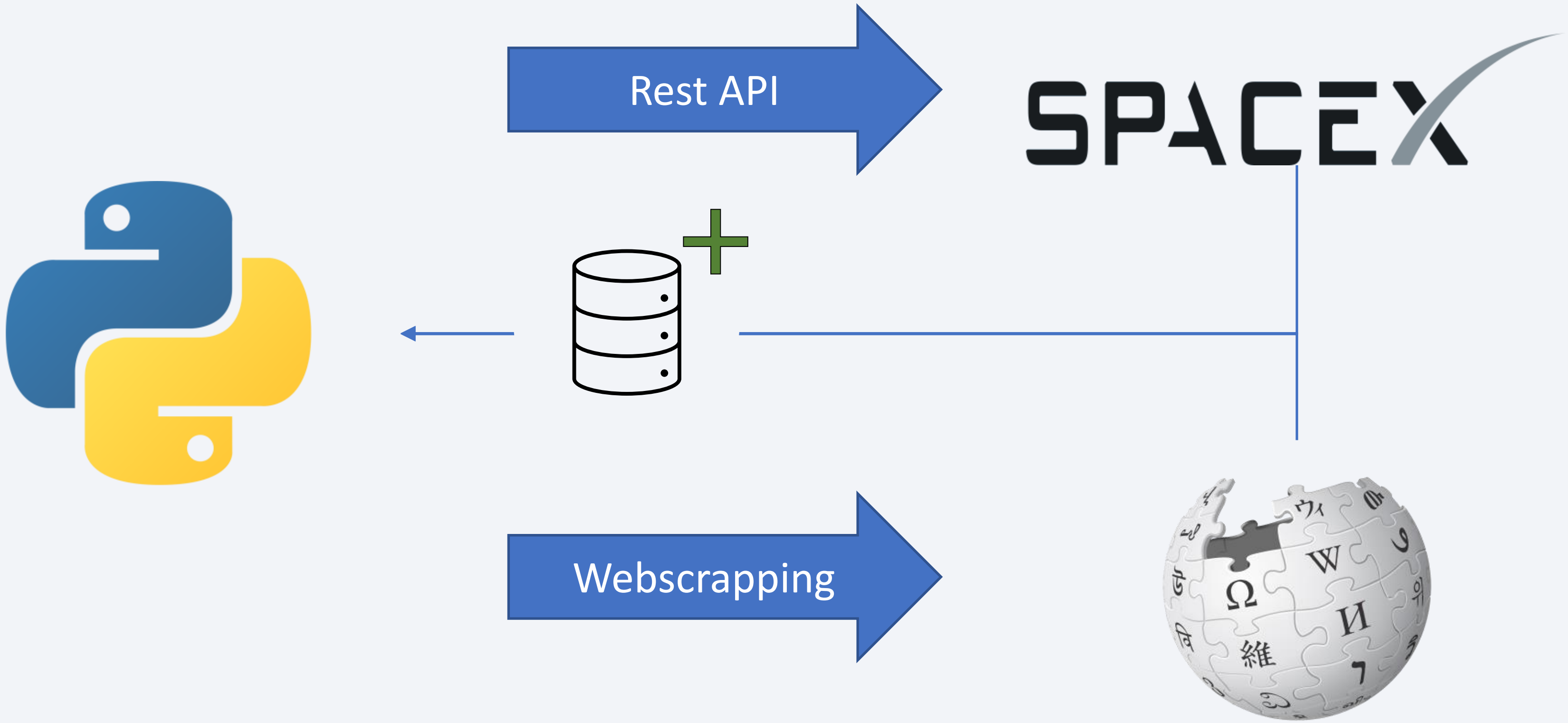
Objectives

- Determine the price of each launch.
- Gather information about Space X and create dashboards for your team.
- Determine if SpaceX will reuse the first stage
- Train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

Section 1

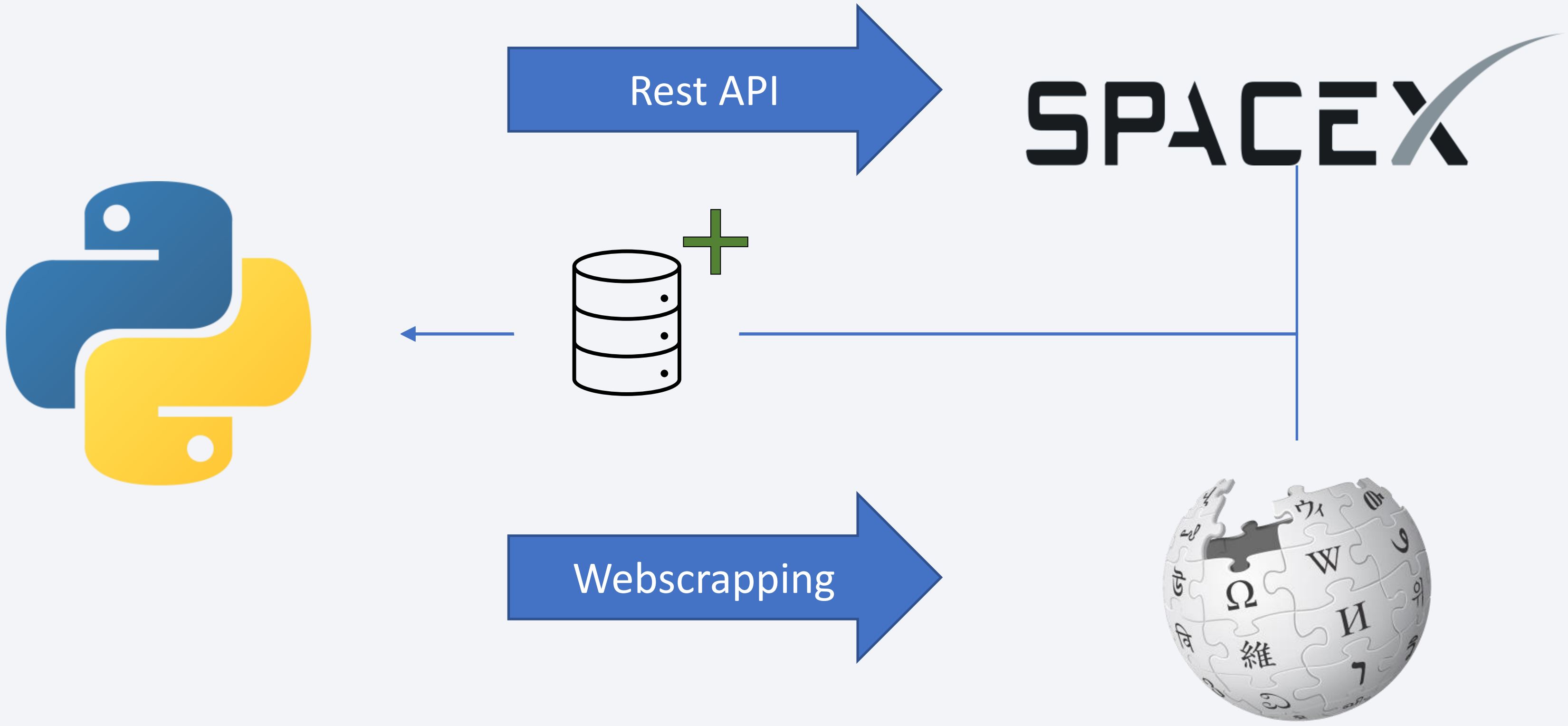
Methodology

Data Collection



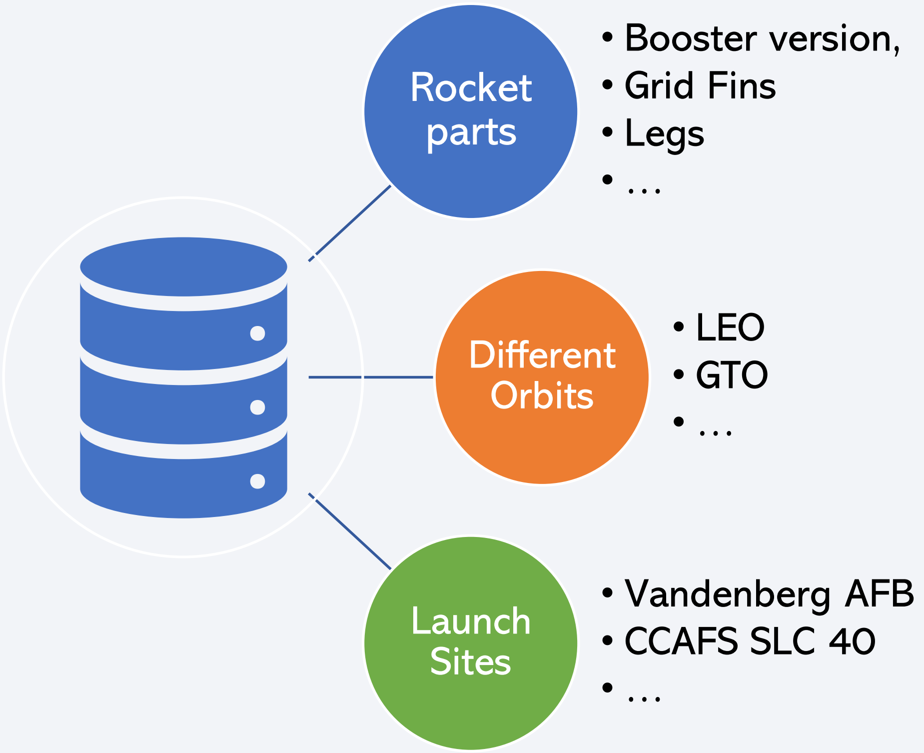
- Collect Data using SpaceX REST API (api.spacexdata.com/v4/launches/past) Using the GET request to extract and .json methods to parse :
- Web Scrape Falcon 9 Launch Data: using Python BeautifulSoup package to scrape HTML tables containing Falcon 9 launch records, parsing the data and converting it into a Pandas DataFrame for further analysis

Data Collection – Code



- Data requests and Webscrapping can be found as tasks 1 & 2 in the url https://github.com/frmachadoecosta/IBMDatascienceCapstone-SpaceX

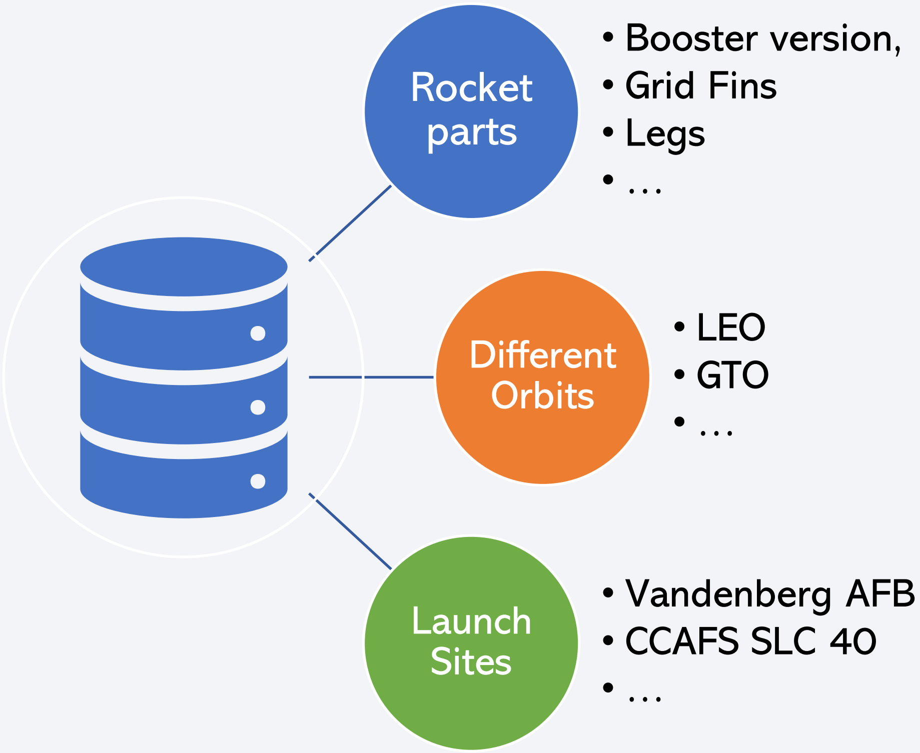
Data Wrangling



- Filter to exclude Falcon 1 launches
- Convert landing outcomes to Class Y (either 0 or 1), where 0 represents a bad outcome (booster did not land) and 1 represents a good outcome (booster did land).

Data Wrangling - Code

GitHub

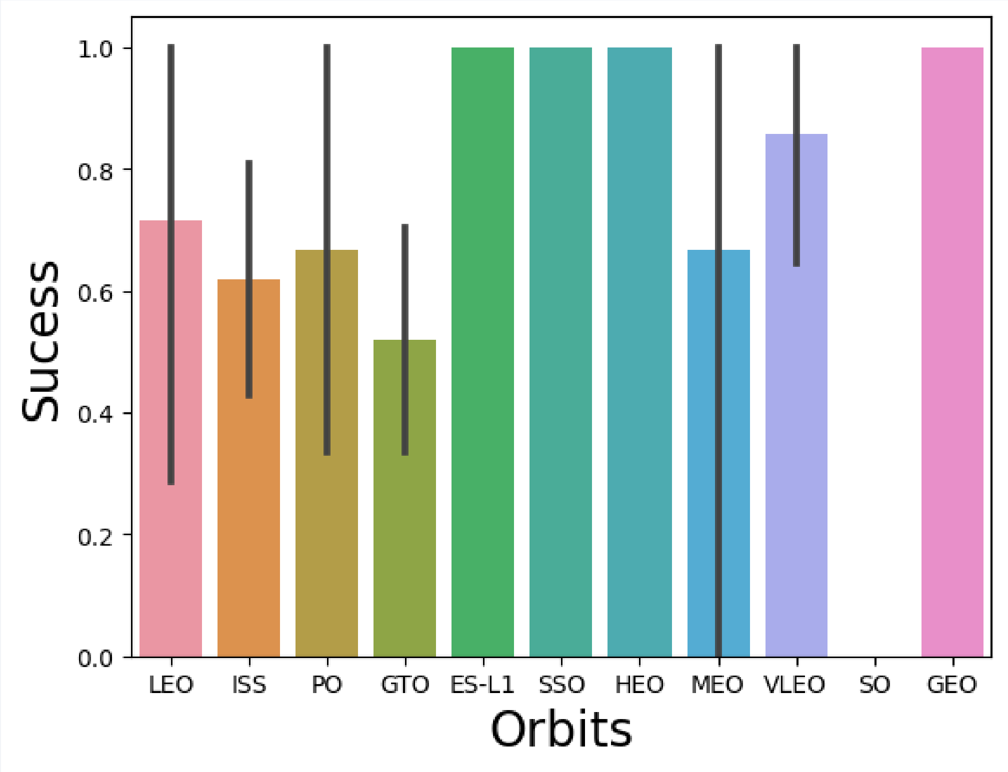


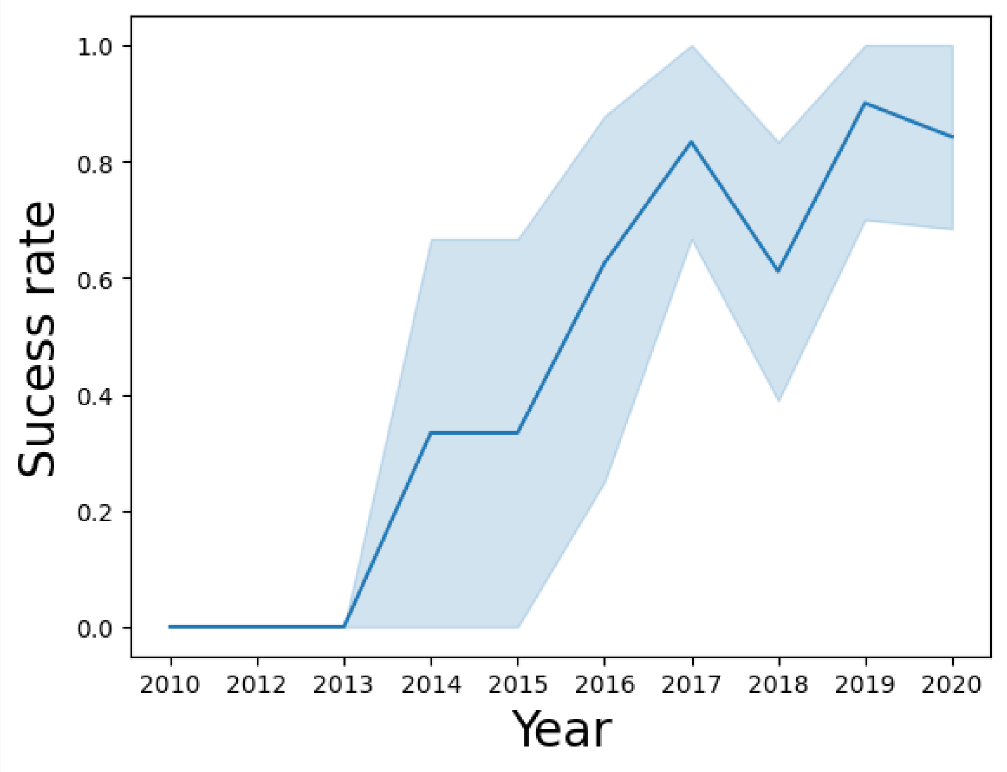
- All data refinements can be seen in Task3 of the repository https://github.com/frmachadoecosta/IBMDatascienceCapstone-SpaceX

EDA with Data Visualization and SQL

- Provide a better understanding of SpaceX's launch history, booster performance, and mission outcomes using SQL and data visualization libraries in python
- Tasks included displaying unique launch sites, filtering records based on specific criteria, calculating payload mass statistics, identifying successful and failed mission outcomes, and ranking landing outcomes within a specific date range.

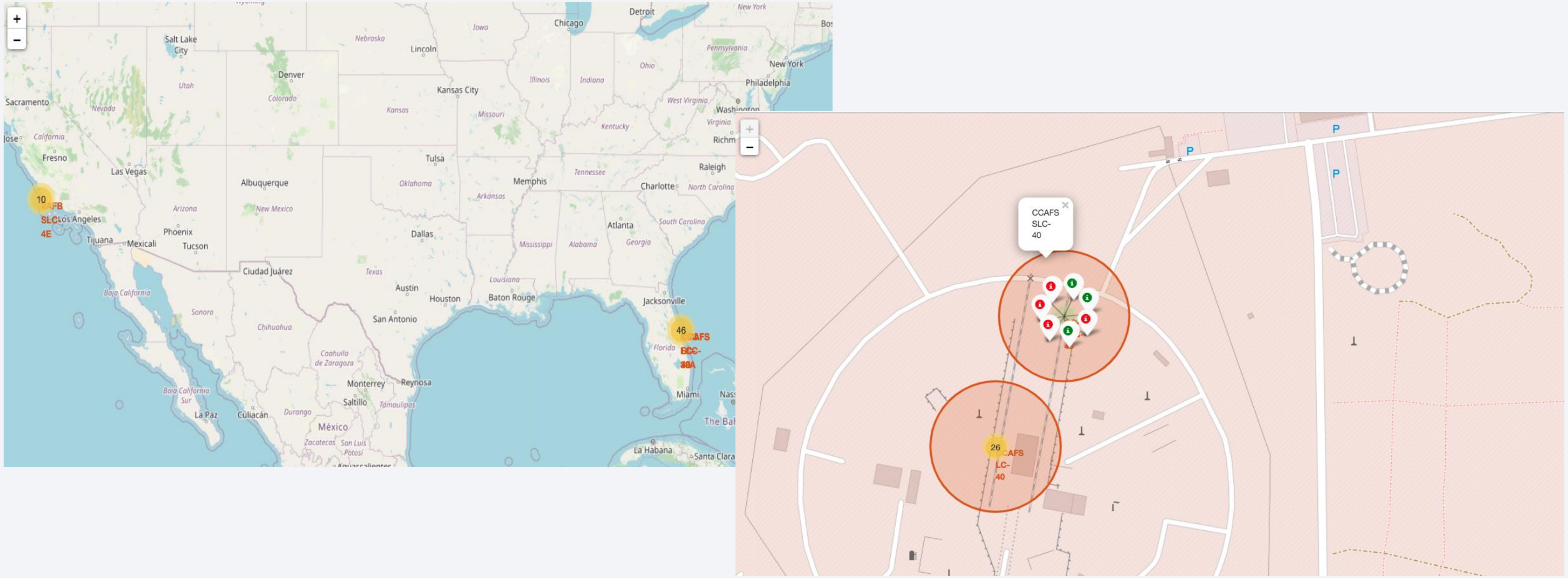
EDA with Data Visualization and SQL





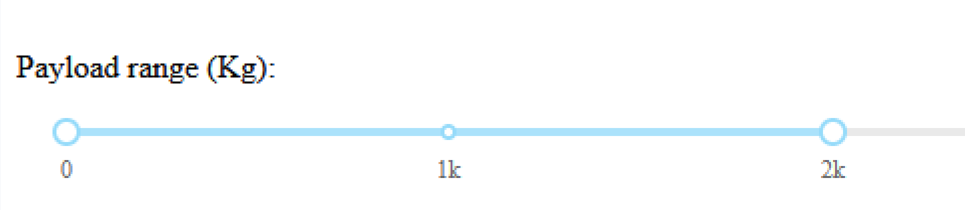
- All data refinements can be seen in Tasks 4 & 5 of the repository https://github.com/frmachadoecosta/IBMDatascienceCapstone-SpaceX

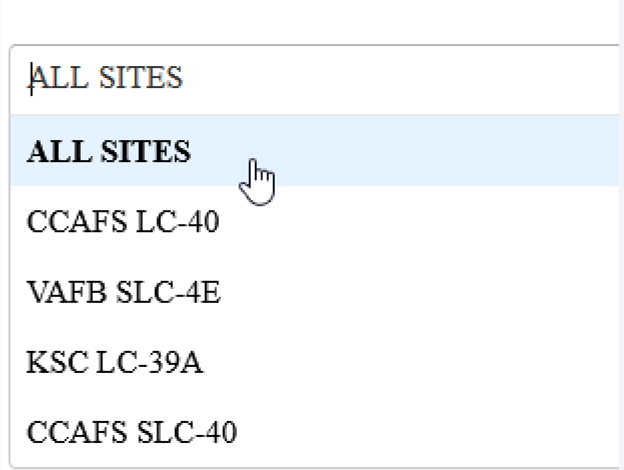
Build an Interactive Map with Folium

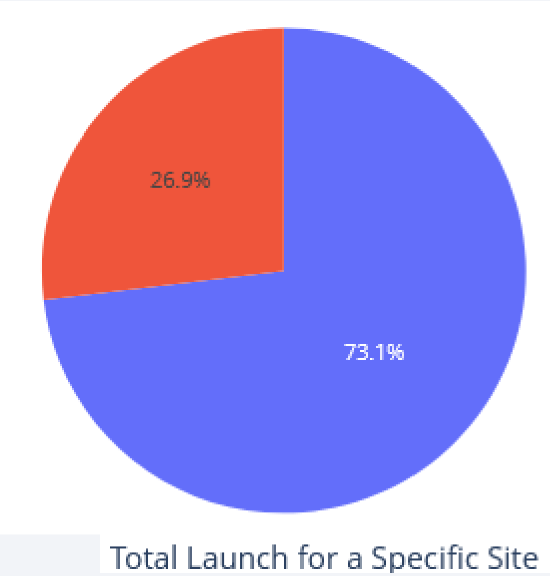


- Visualize launch site location as well as surrounding landmarks and plot a correlation between site and successful launches

Build a Dashboard with Plotly Dash

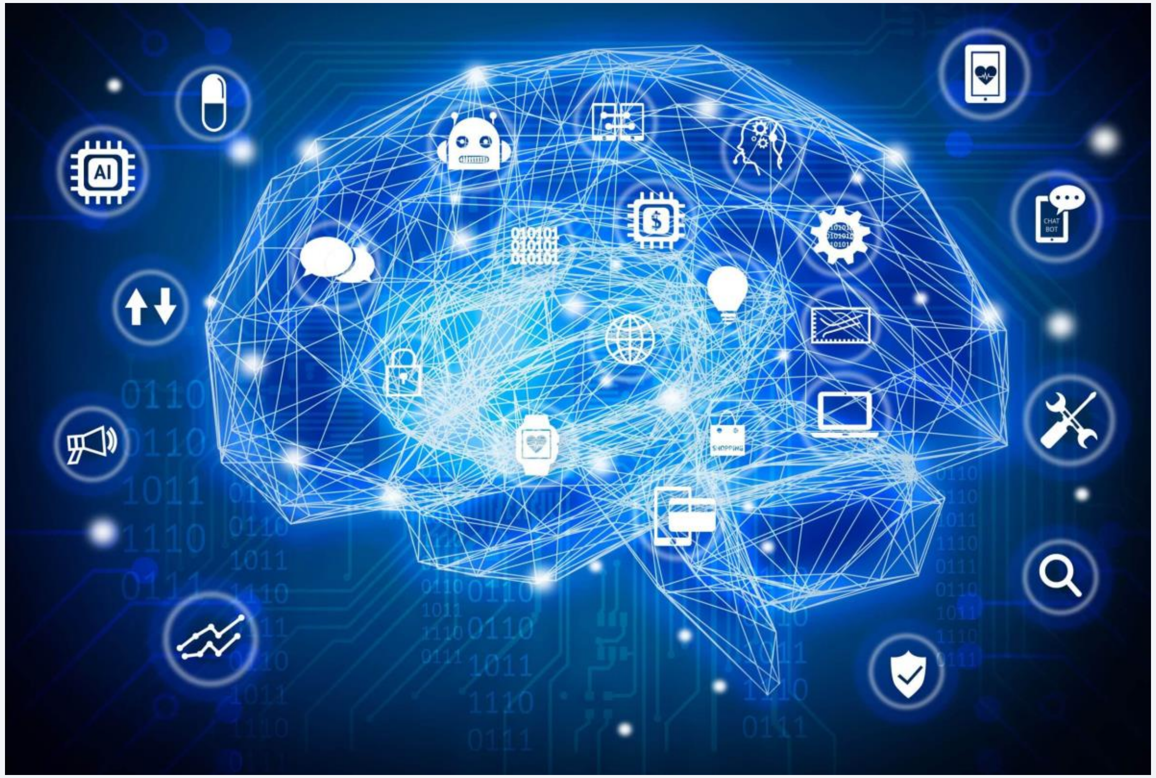






- Interactive data exploration with plotly
- Resulting app can be found in Task7 of the [repository]

Predictive Analysis (Classification)

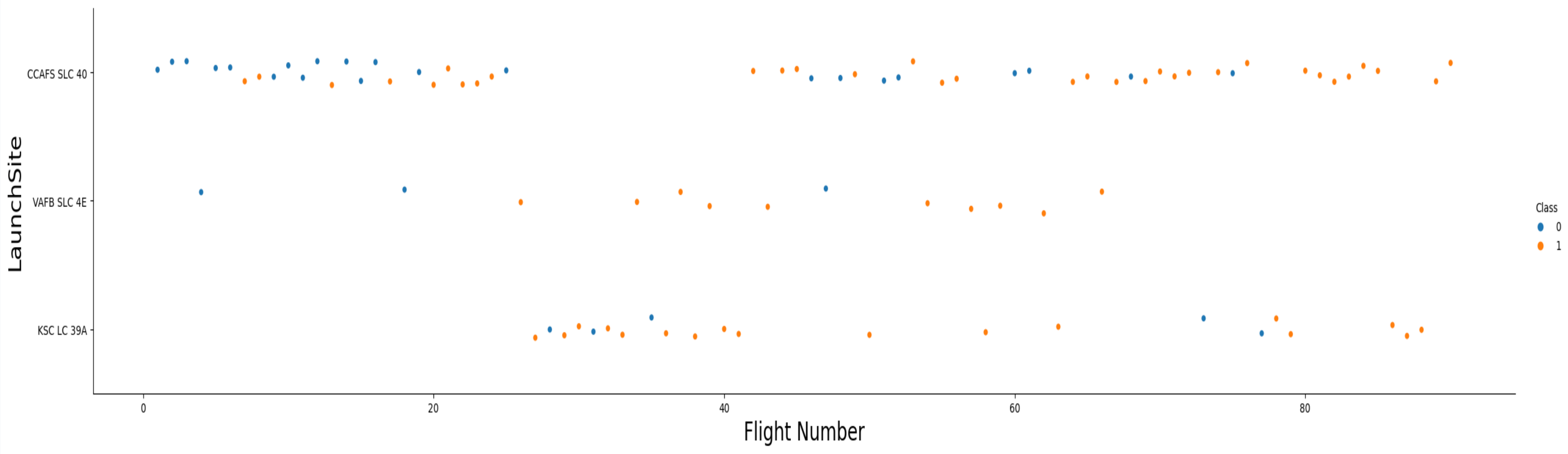


- Split data into training data and test data to find the best Hyperparameter for SVM, Classification Trees, and Logistic Regression.
- Find the method that performs best using test data.

Section 2

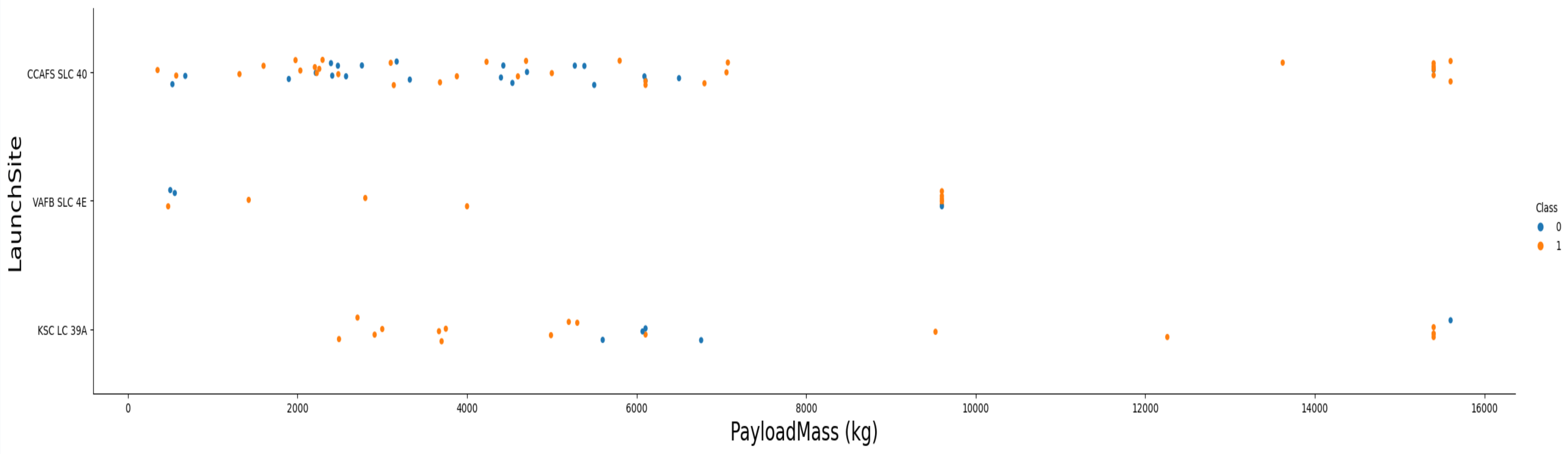
Insights drawn from EDA

Flight Number vs. Launch Site



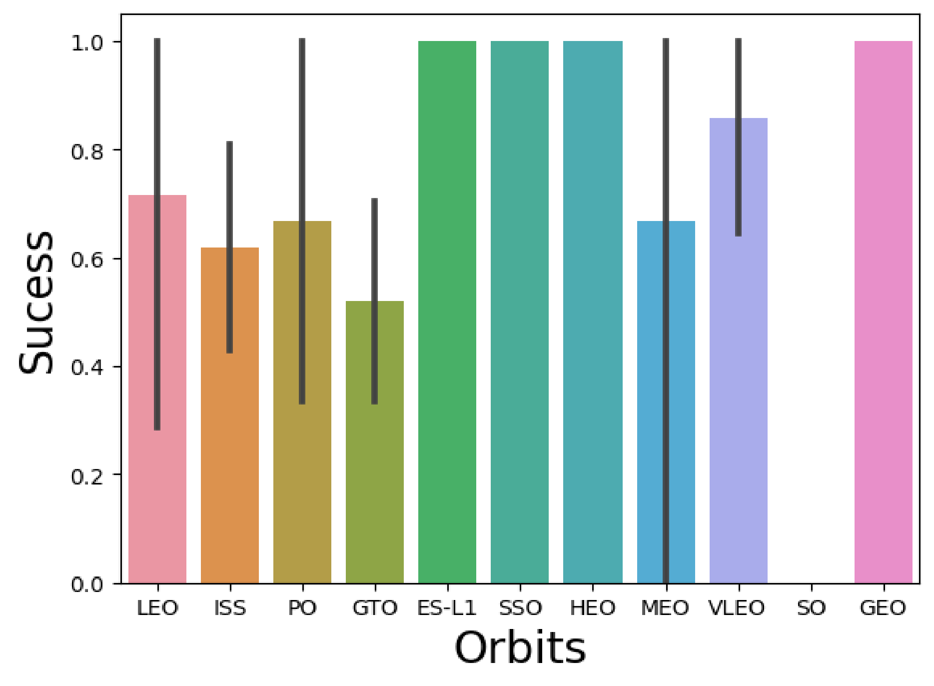
- We see a clear preference for CCAFS and KSC sites

Payload vs. Launch Site



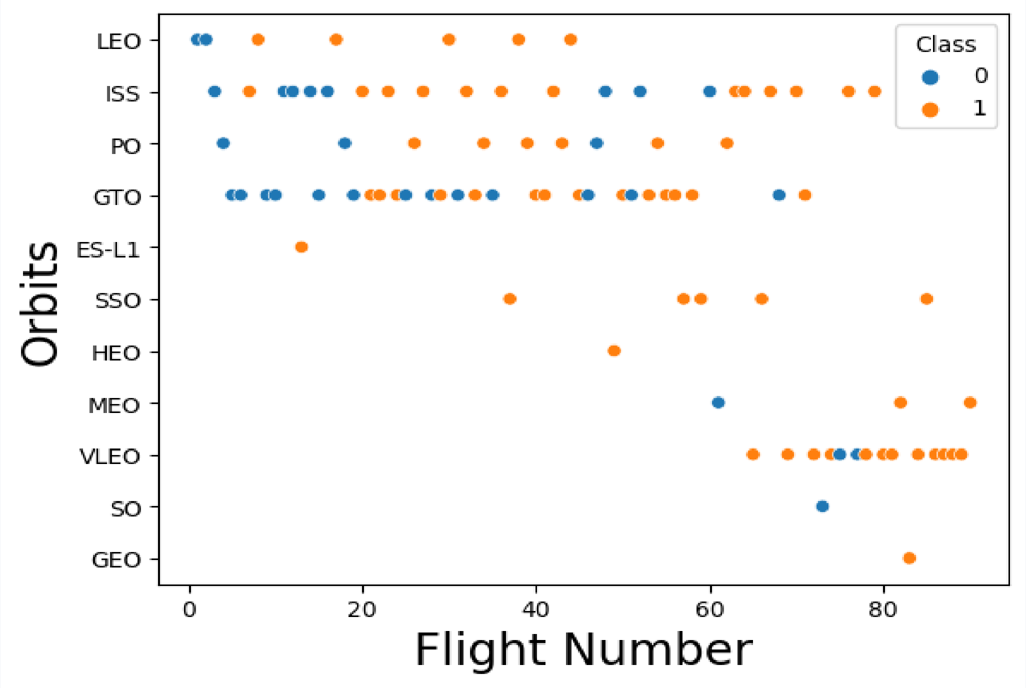
- Payloads larger than 10,000kg yield a higher success rate
- KSC sees some success with small payloads

Success Rate vs. Orbit Type



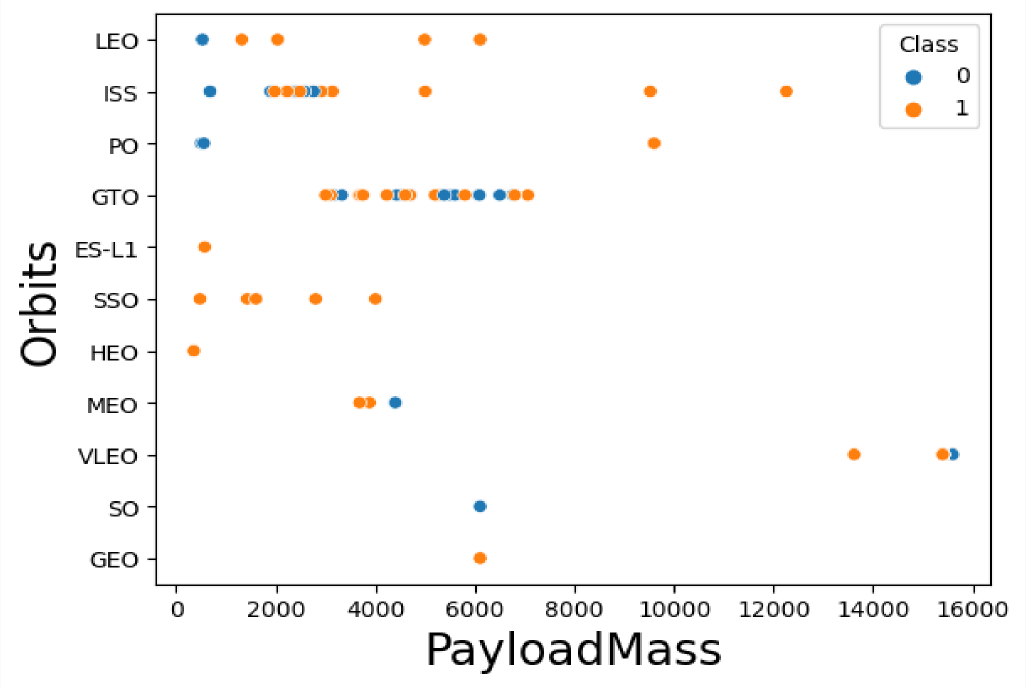
- Orbits ES-L1, SSO, HEO and GEO have the highest success rate

Flight Number vs. Orbit Type



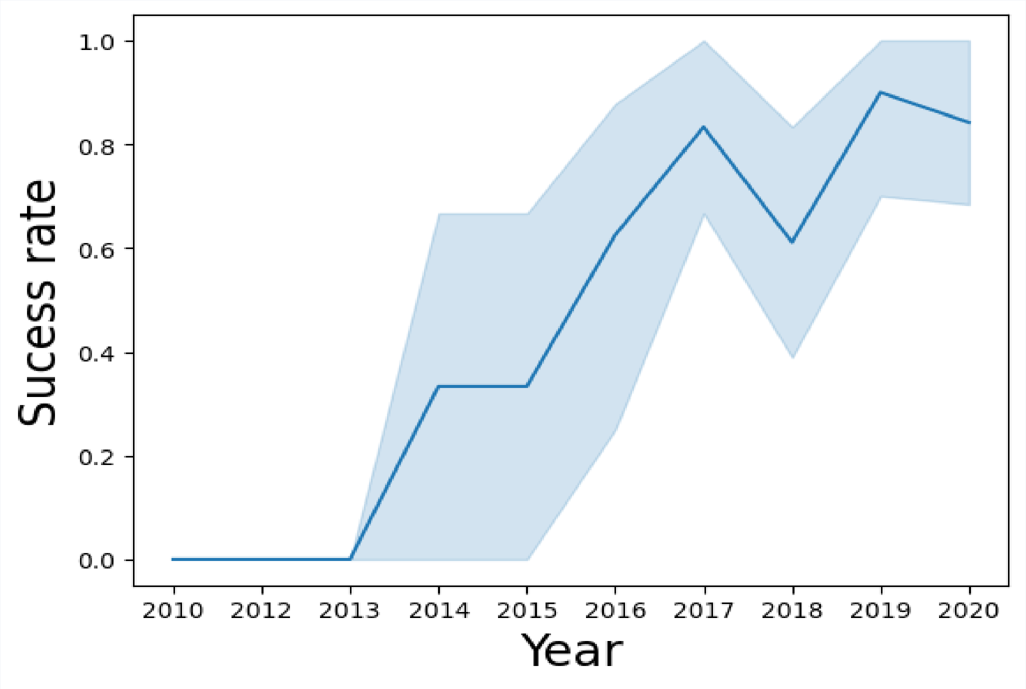
- Orbits VLEO have had some recent success along with the highest number of launches
- There seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



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

Launch Success Yearly Trend



- Success rate has increased since 2013

All Launch Site Names

```
%sql SELECT DISTINCT Launch_Site from SPACEXTBL
```

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

Launch Site Names Begin with 'CCA'

```
%sql SELECT DISTINCT * from SPACEXTBL where Launch_Site LIKE "CCA%" LIMIT  5
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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)	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") from SPACEXTBL where Customer = "NASA (CRS)"
```

```
 * sqlite:///my_data1.db
Done.
```

SUM("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") from SPACEXTBL where Booster_Version = "F9 v1.1"
```

```
 * sqlite:///my_data1.db
Done.
```

AVG("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Landing Date

```
%sql SELECT DISTINCT "Landing_Outcome" from SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome
Failure (parachute)
No attempt
Uncontrolled (ocean)
Controlled (ocean)
Failure (drone ship)
Precluded (drone ship)
Success (ground pad)
Success (drone ship)
Success
Failure
No attempt

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT Booster_Version from SPACEXTBL where Landing_Outcome = "Success (drone ship)"  AND PAYLOAD_MASS__KG_ >
```

```
 * sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT COUNT(*) as Sucess_count from SPACEXTBL WHERE Landing_Outcome LIKE "Success%"
```

```
 * sqlite:///my_data1.db
Done.
```

Sucess_count
61

```
%sql SELECT COUNT(*) as Failure_count from SPACEXTBL WHERE Landing_Outcome LIKE "Fail%"
```

```
 * sqlite:///my_data1.db
Done.
```

Failure_count
10

Boosters Carried Maximum Payload

```
%sql SELECT DISTINCT Booster_Version as  Carriedmaxpayload FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MAS
```

```
 * sqlite:///my_data1.db
Done.
```

Carriedmaxpayload
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

```
%sql SELECT substr(Date, 6, 2) as month,Landing_Outcome,Booster_Version,Launch_Site  FROM SPACEXTBL WHERE Landing_Outcome =
```

```
 * sqlite:///my_data1.db
Done.
```

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

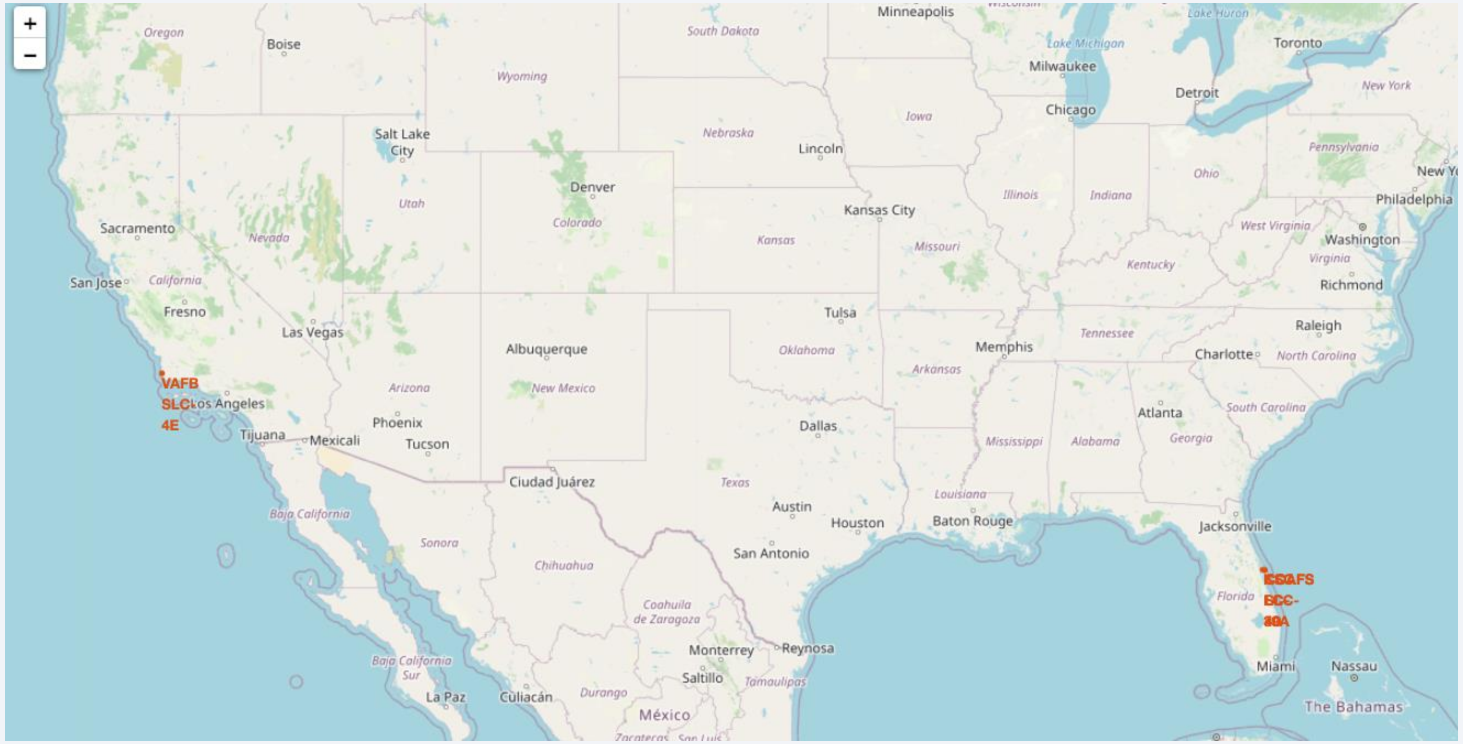
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing_Outcome	number
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

Section 3

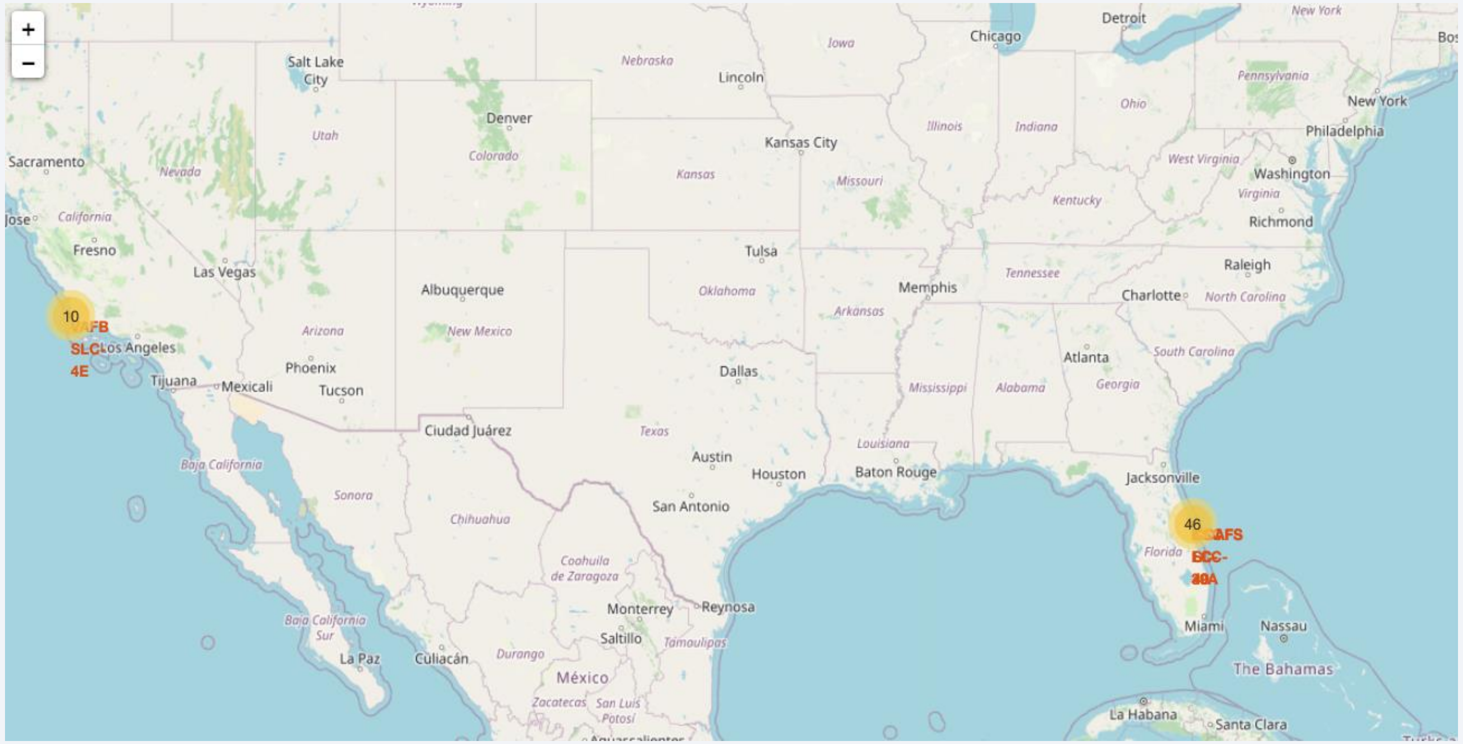
Launch Sites Proximities Analysis

Follium map of Launch site locations



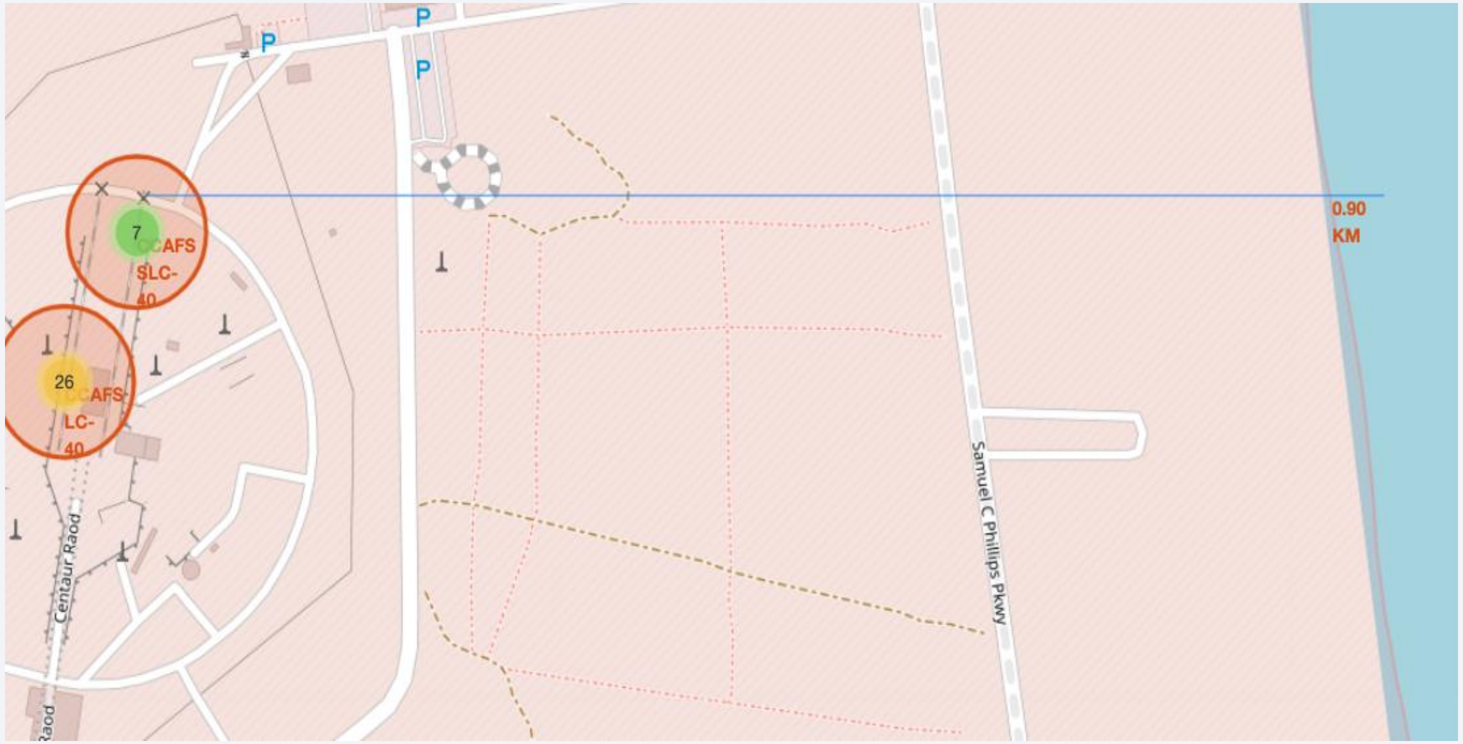
- Easter sites are closer to the equator

Success and Failures per Site



- Easter sites are preferred in # of launches

Site proximity markers



- All launch sites share a proximity to shore as well as railway stations

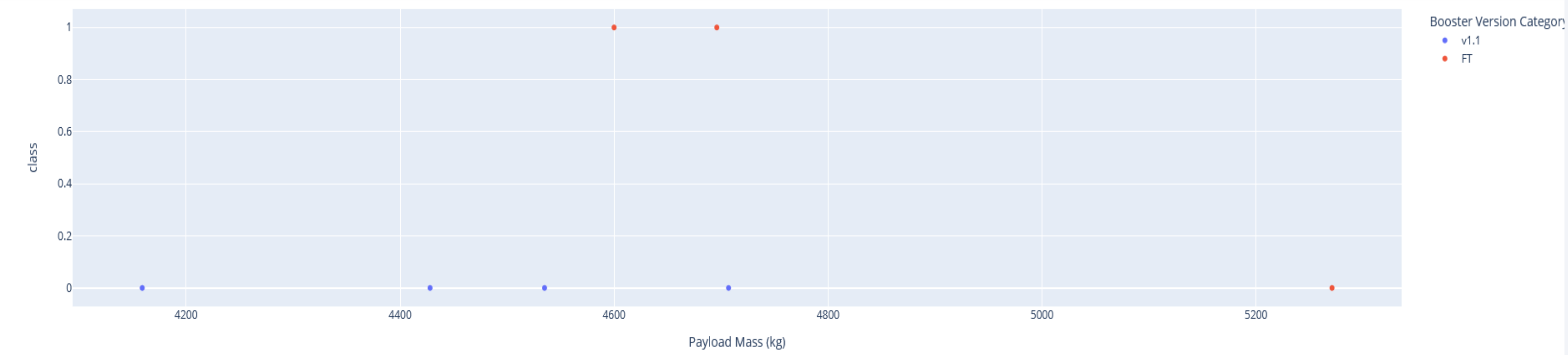
Section 4

Build a Dashboard with Plotly Dash

Success launches by Site

Launch site with highest launch success ratio

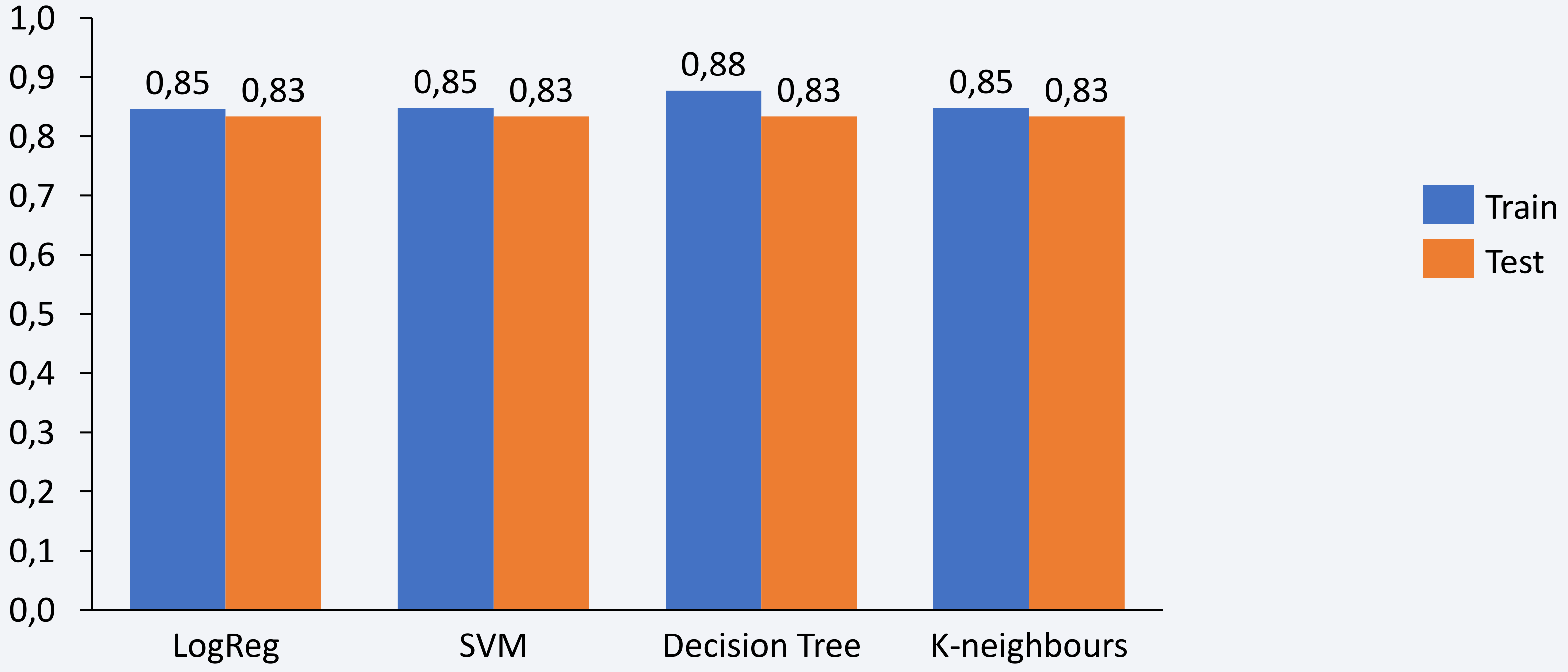
Payload Mass vs. Launch Outcome for all sites



Section 5

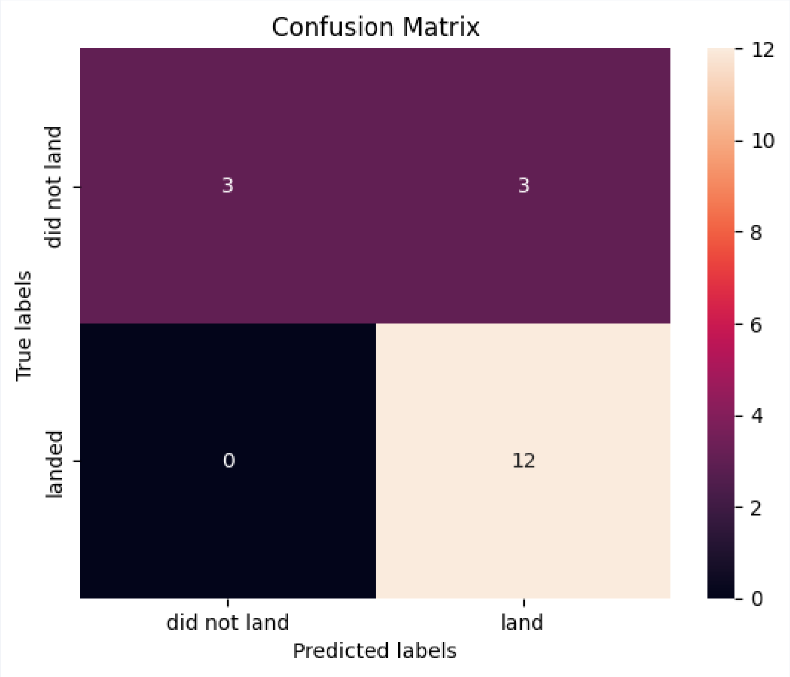
Predictive Analysis (Classification)

Classification Accuracy



- All models had a prediction accuracy of 83% with Decision Tree having a slightly higher result with the train data

Confusion Matrix



- Model has a high True and False positive rate

Conclusions

- Launch success rate has increased with time.
- Orbits ES-L1, SSO, HEO and GEO had the highest success rate.
- All machine learning algorithms we adequate for predicting launch outcome.

Thank you!