IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Modupe Adeoti
July 16th, 2025

Outline

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

Executive Summary

As Chief Data Scientist at Space Y, I spearheaded an ambitious, data-driven initiative to support the company's mission: to become the first commercial entity to make space travel a familiar and accessible reality for households around the globe. We utilized thorough rigorous research, analytical methodologies, and grounded robust dataset

Methodologies

- Data Collection from API and Web scraping.
- Data Wrangling.
- Exploratory Data Analysis (EDA)using SQL, Pandas and Matplotlib.
- Interactive Visual Analytics and Dashboard with Follium

Results

- The best Hyperparameters for logistics Regression, SVM, KNN Neighbors and Decision Tree.
- The best performing model using test/train split training.

Introduction

Space Y is entering the commercial space race and its biggest competitor is Space X.

We aim to make rocket launches relatively inexpensive for everyone. By leveraging data from our competitor we want to see how we can save millions in each launch of our rocket if we can reuse the first stage, find out the price of each launch and use machine learning models to predict if our competitor reuse the first stage

Section 1

Methodology

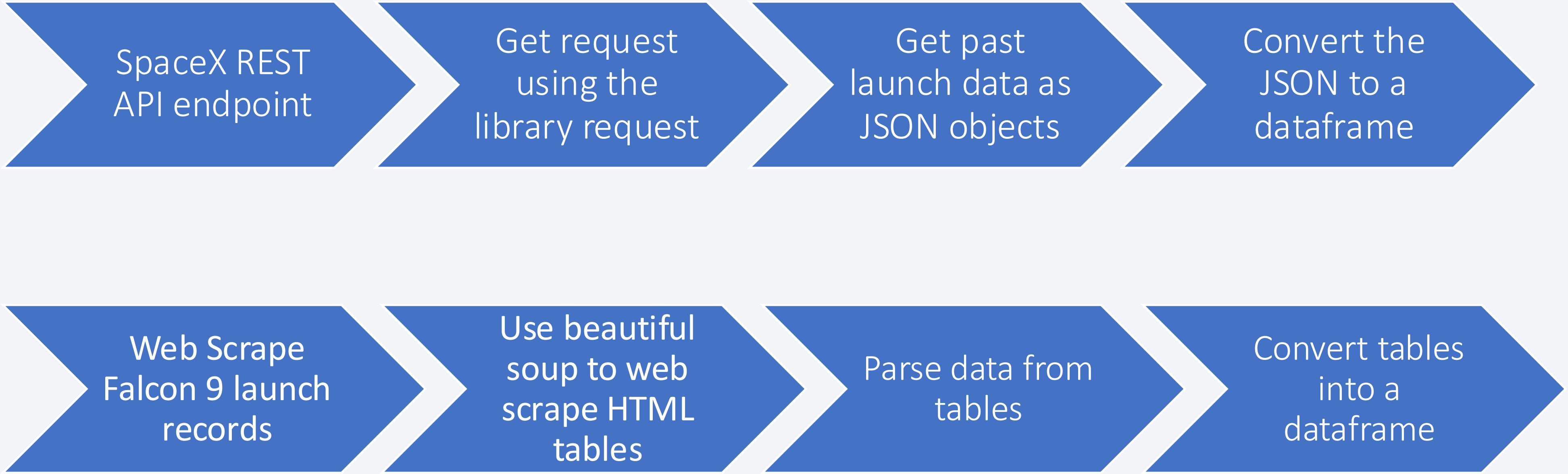
Methodology

Executive Summary

- Data collection methodology:
  - The data was gathered from SpaceX REST API and web scraping from wiki pages
- Perform data wrangling
  - The data collected is in form of a JSON object and HTML tables, after that the data is converted into a Pandas dataframe for visualization and analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use of machine learning to determine if the first stage of Falcon 9 will land succesfully

Data Collection

Data was gathered from SpaceX REST API and web scarped from wiki pages



Data Collection – SpaceX API

SpaceX REST API endpoint

Get request using the request library

Extract inormation about booster, name, launchsite, payload mass and landing site

Get the past launch data as a JSON object

Data Wrangling

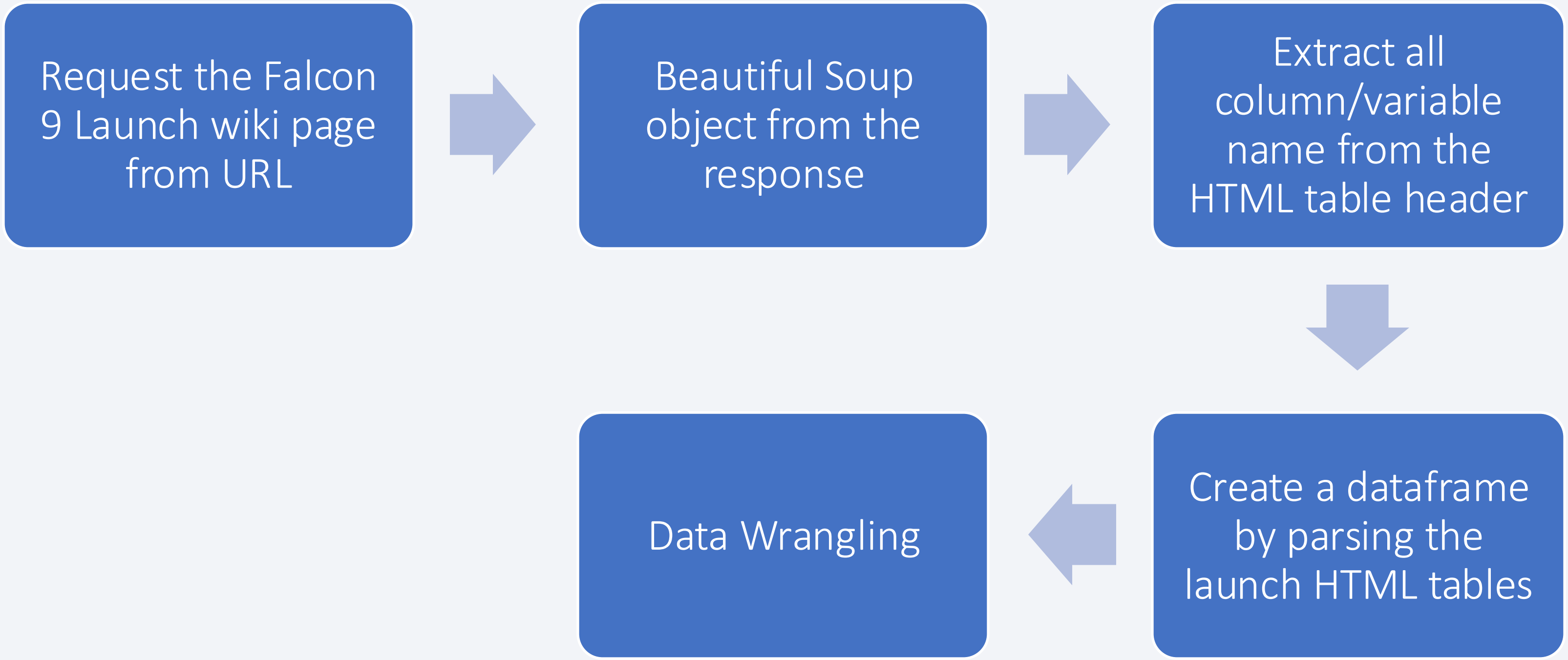
Deal with Missing Values

Filter the dataframe to only include Falcon 9 launches

Convert to JSON to a dataframe

- Data Collection API Notebook

Data Collection - Scraping

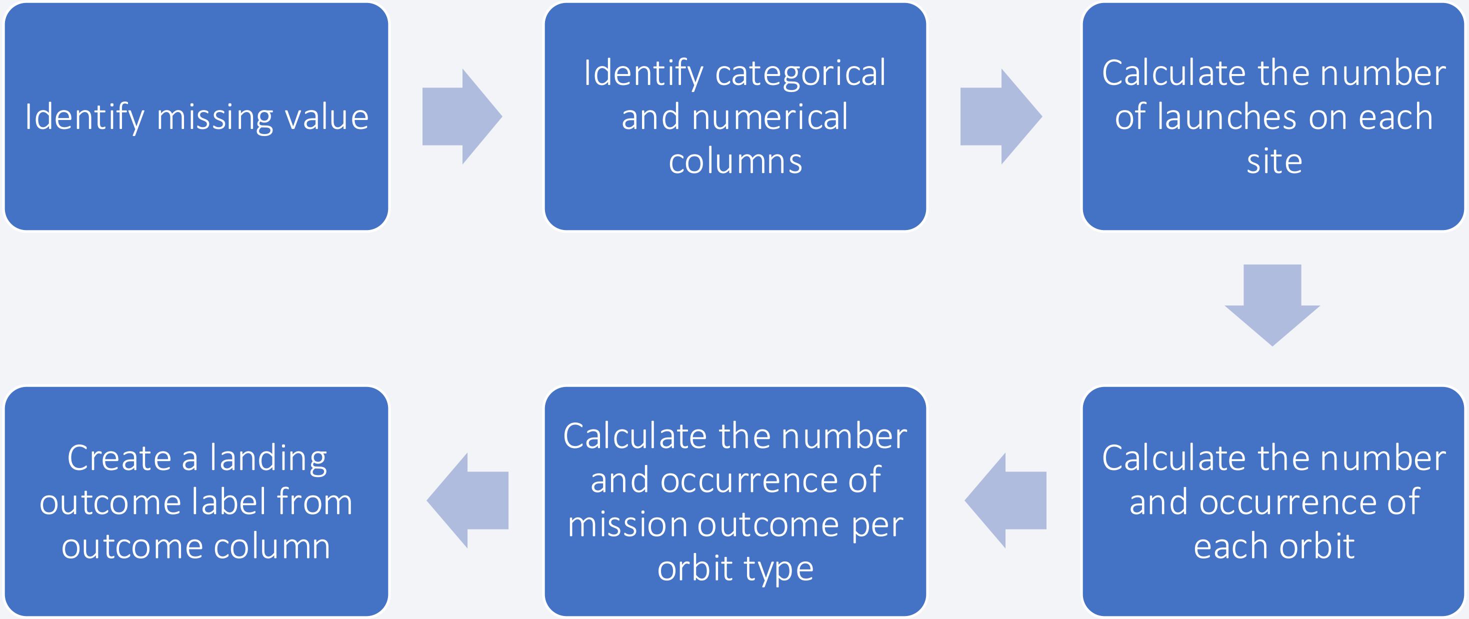


Perform web scraping to collect Falcon 9 historical launch records from Wikipedia page

[Web scraping Notebook]

Data Wrangling

Perform Exploratory Data Analysis to find patterns in the data and determine what would the label for train supervised models



The variable represents the classification outcome of each launch. Zero means the firststage did not land successfully while One means it landed successfully.

[Data Wrangling notebook](#)

EDA with Data Visualization

- Summary of charts that were plotted:
- Catplot to visualize the relationship between flight number and payload
- Catplot to visualize the relationship between flight number and launch site
- Catplot to visualize the relationship between payload and lunchsite
- Barchart to visualize the relationship between success rate of each orbit type
- Catplot to visualize the relationship between flight number and orbit type
- Catplot to visualize the relationship between payload and orbit type
- Line chart to visualize the launch success yearly trend

EDA with visualization notebook

EDA with SQL

- Used EDA with SQL to :
  - Display the names and unique launch sites in the space mission
  - Display 5 records where the launch site beings with CCA
  - Display the total payload mass carried by boosters launches by NASA (CRS)
  - Displa the average payload mass carried by booster version F9 v1:1
  - Listed the date when the first successful landing outcome in ground pad was achieved
  - Listed the names of the boosters which have success in drone ship and have payload mass >4000 and <6000
  - Listed the total numbers of successful and failed mission outcomes
  - List the names of booster version which have carried the maximum payload mass using a subquery

EDA with SQL notebook

Build an Interactive Map with Folium

- Summarize of the map objects:
- Folium.Circle and folium.Marker- use this to add a highlighted circle area with a text lable on a specific coordinate for each launch site on the site map
- MarkerCluster – use for object for simplifying a map containing many markers having the same coordinate
- MousePosition – use on the map to get coordinate for a mouse over a point on the map
- Folium.Polyline- use to draw a line between a launch site to its closest city, railway and highway

Launch site interactive map with folium

Build a Dashboard with Plotly Dash

- The dashboard applications contains input component such as dropdown lista and a range sliders to interact with a pie chart and scatter chart
- A launch site dropdown input component
- A callback function to render success pie chart based on selected site dropdown
- A range slider to select payload

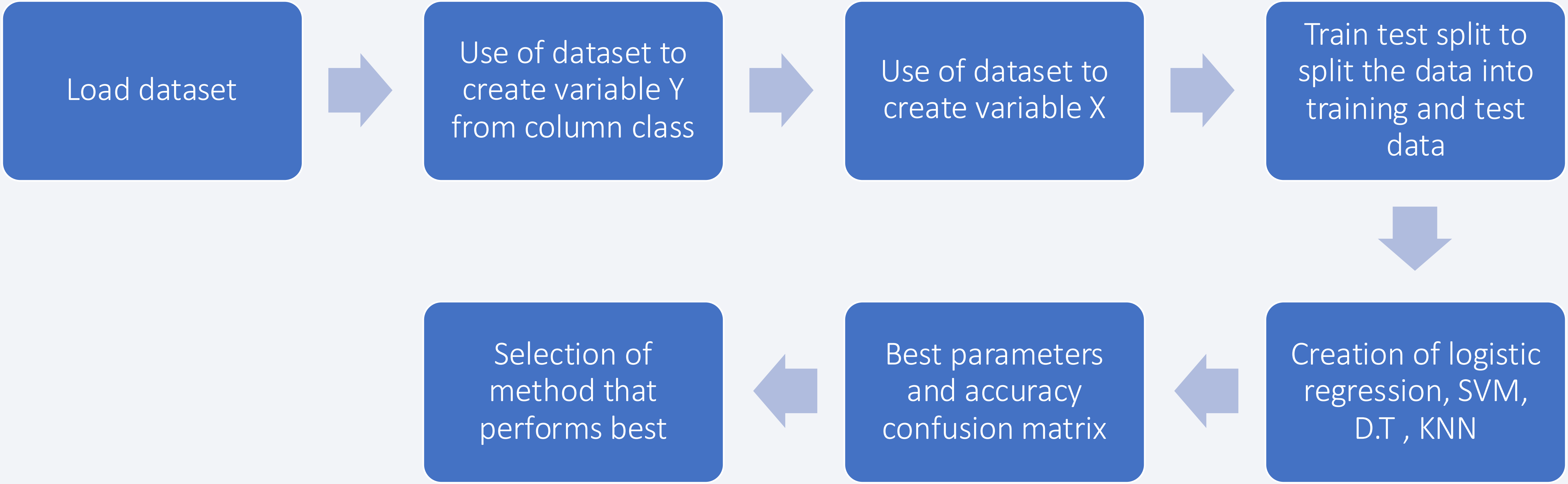
Interactive Dashboard with Plotly Dash

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model.
  - Creation of NumPy array from the column class
  - Data Standardization
  - Use of the function train-test-split to slpit
  - identify the best hyperparameters for logistics regression, SCM, decision tree and KNN classifiers
  - Identify the method that performs best using test data

[Machine learning predictive analysis]

Predictive Analysis (Classification)



[Machine learning predictive analysis]

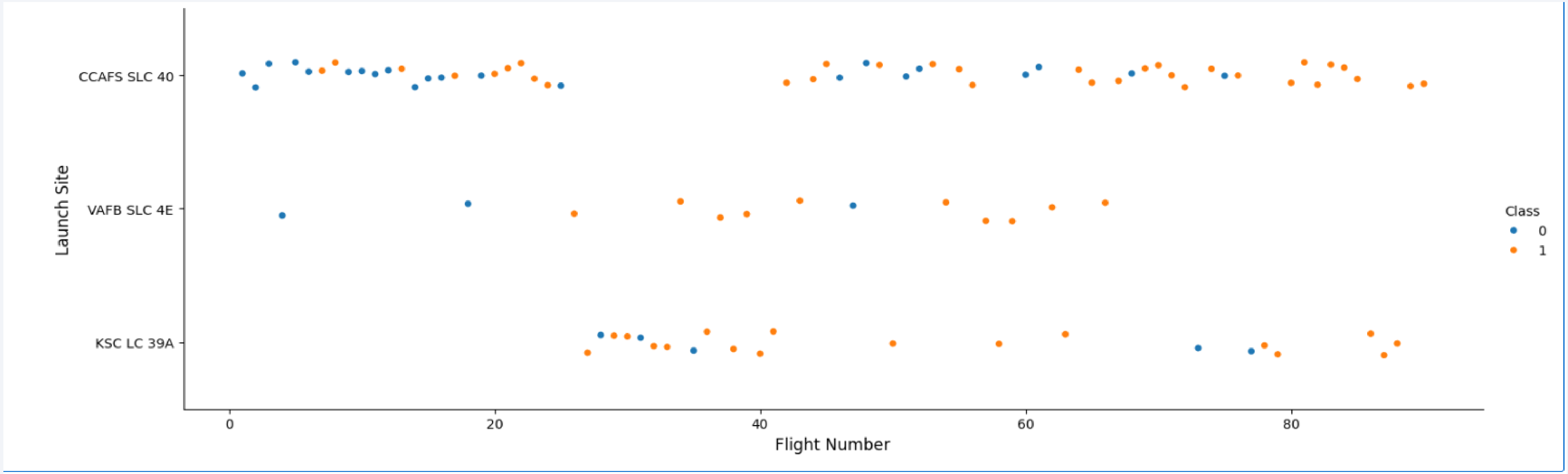
Results

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

Section 2

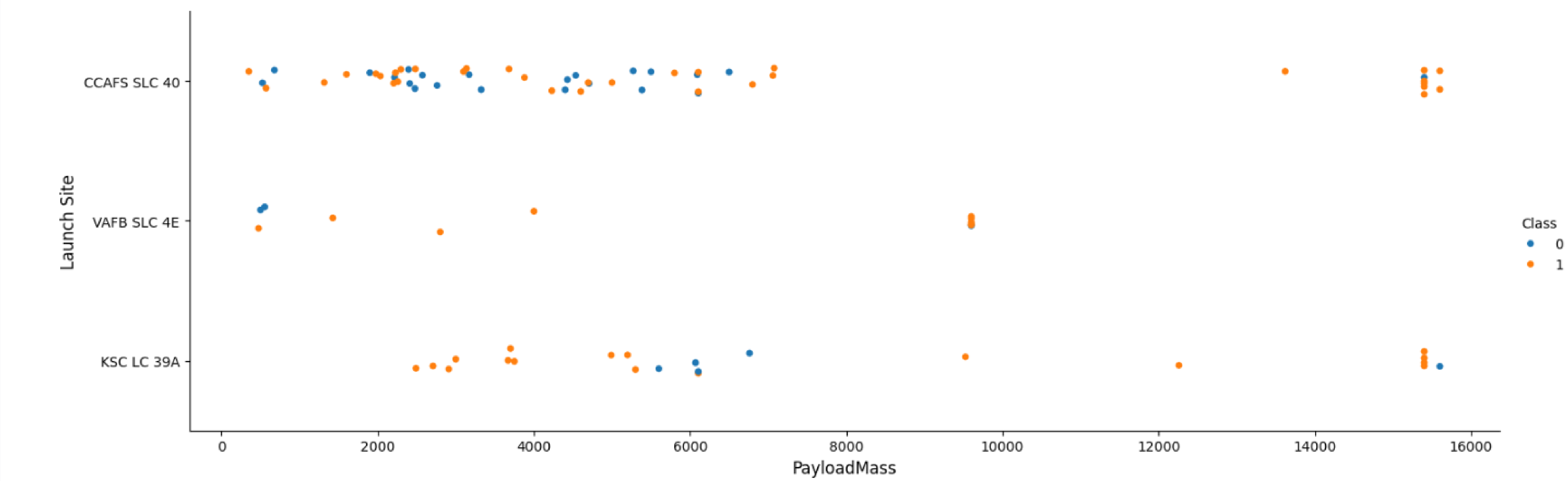
Insights drawn from EDA

Flight Number vs. Launch Site



With time the successful rate has increased for every launch site,
VAFB SLC 4E and KSC LC 39A has a higher successful rate but represents one third of the total launches

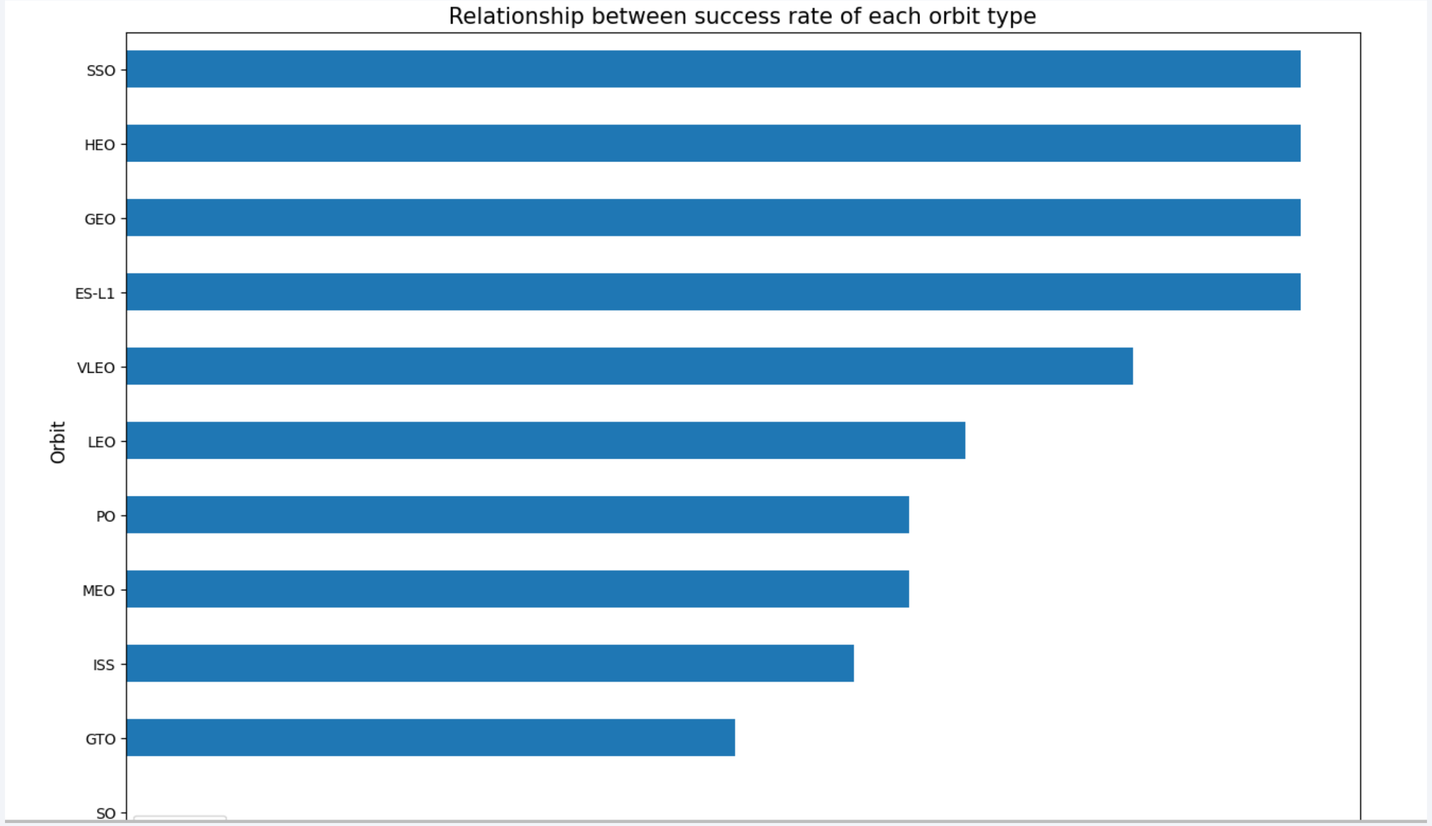
Payload vs. Launch Site



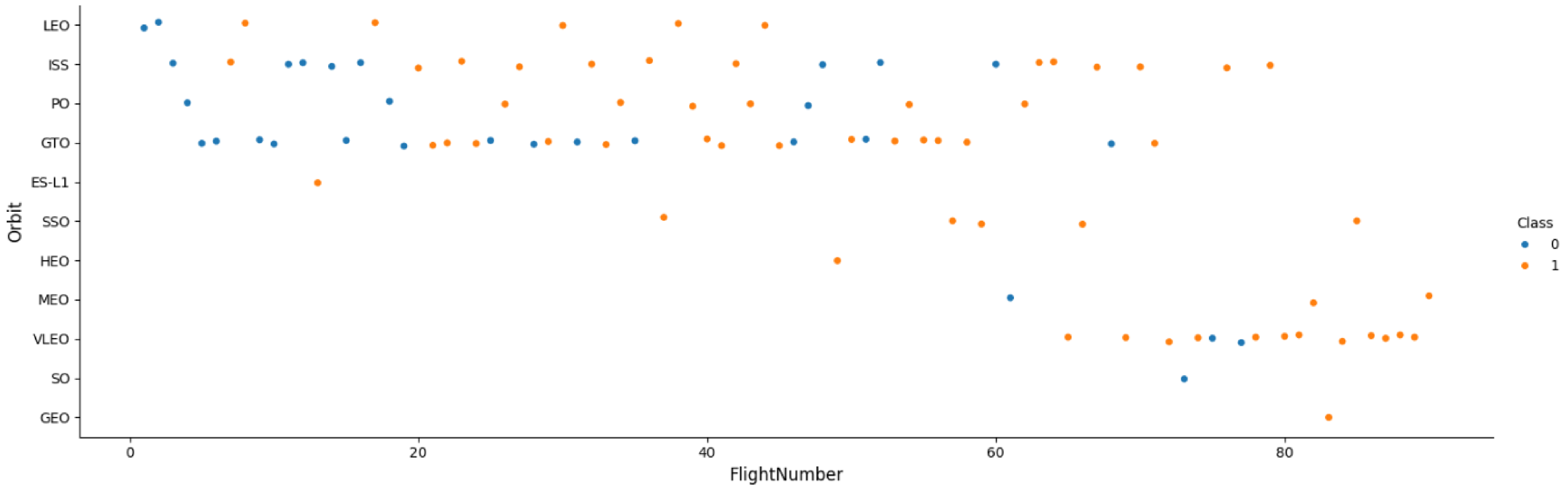
In VAFB-SLC launch site there are no rockets launched for heavy payloadmass greater than 10000kg
In KSC LC Launch site there are no rocket launched lower than payloadmass less than 2500kg
CCAFS SLC has launched rocket less than 7500kg and more than13000kg but not in between

Success Rate vs. Orbit Type

- The first 4 Orbit types has the best successful rate.



Flight Number vs. Orbit Type

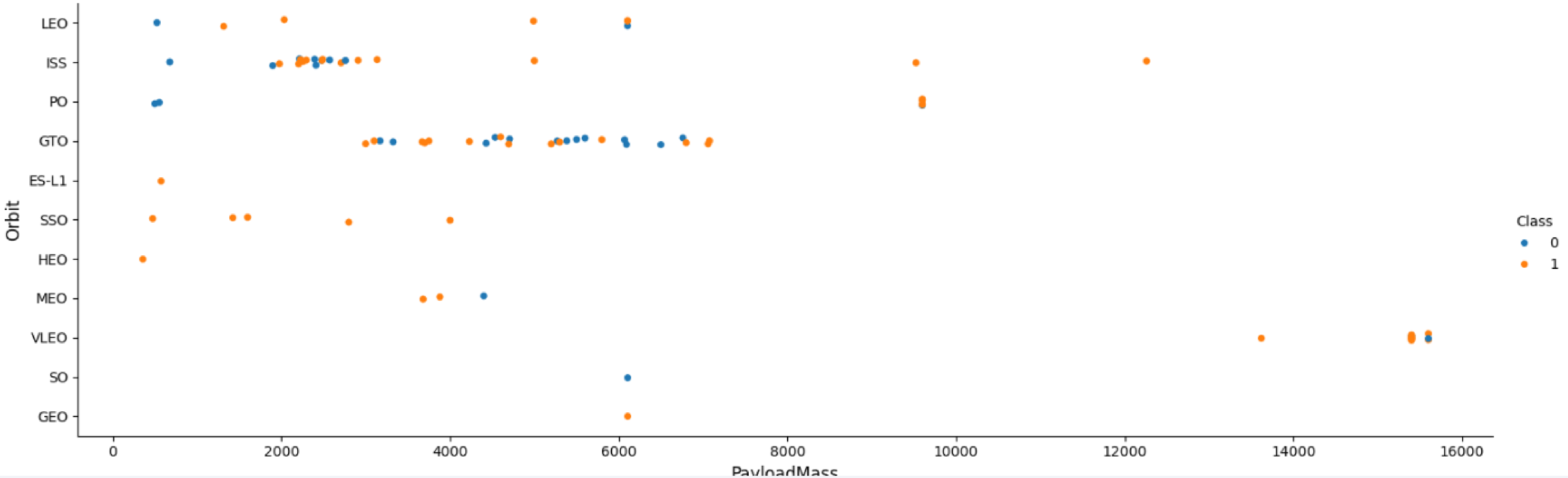


There were more failure at the beginning of the series of launches but after the first 40 attempts, the raito improves by reducing the 50 percent of unsuccessful landings

GTO and ISS orbit has the higher concentration of launches with the lowest ratio of successful landings

The orbit with higher successful rate, has one or just few numbers of launches

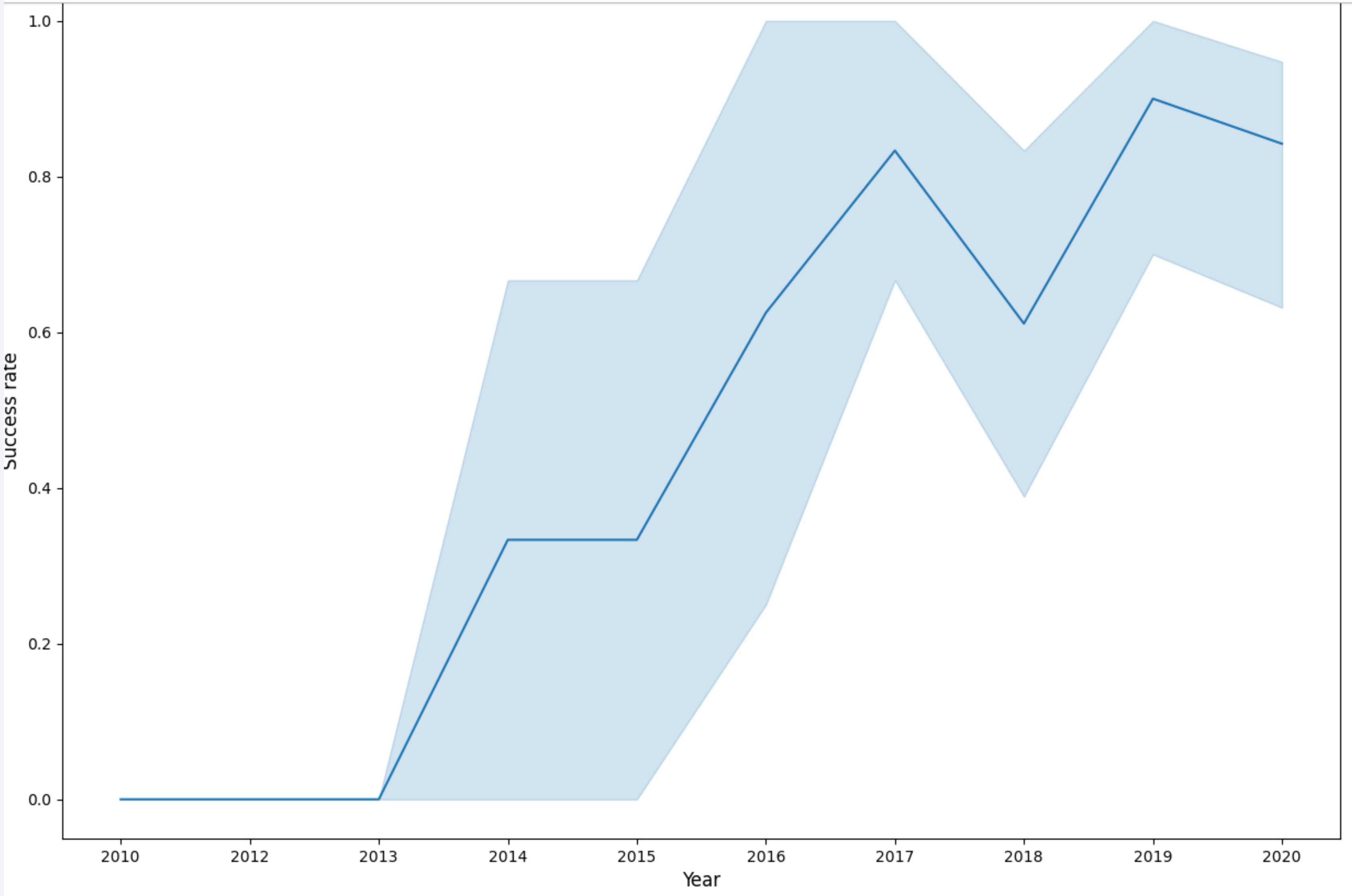
Payload vs. Orbit Type



- Exists a visible limit of Payload around 7600 kg. Less than 10 launches exceed that limit.
- With heavy payloads the successful landing rate are more for Polar, LEO and ISS.
- However for GTO, we cannot distinguish this well as both, positive landing rate and negative landing are both there here.

Launch Success Yearly Trend

The success rate since 2013 kept increasing until 2020



All Launch Site Names

Task 1

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

[11]:
```
%sql select DISTINCT "Launch_Site" from SPACEXTBL
```

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

[11]:

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

I used distinct function to drop all repeated names

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Using WHERE, LIKE, and LIMIT function

```
15]: %sql select * from SPACEXTBL WHERE "Launch_Site" like 'CCA%' limit 5
 * sqlite:///my_data1.db
Done.
```

15]:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Using SUM and Where functions

Task 3

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

```
[4]: %sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer='NASA (CRS)'
```

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

4]:

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Used same logic as above

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

```
[25]: %sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where "Booster_Version" ='F9 v1.1'
```

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

[25]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Using MIN, WHERE and AS

Hint:Use min function

```
[26]: %sql SELECT MIN(DATE) AS first_successful_landing FROM SPACEXTBL WHERE (landing_outcome)='Success (ground pad)';
```

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

[26]:

first_successful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Used WHERE and AND functions to query the data

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, payload_mass__kg_, landing_outcome FROM SPACEXTBL \
        WHERE landing_outcome='Success (drone ship)' AND (payload_mass__kg_ BETWEEN 4000 AND 6000) ;
```

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

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Used Count, AS, and Group by functions

```
%sql SELECT mission_outcome, COUNT(mission_outcome) AS TOTAL FROM SPACEXTBL GROUP BY mission_outcome;
```

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

Mission_Outcome	TOTAL
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Used DISTINCT, MAX, AS in a subquery form also applied LIMIT 5 to see the first 5 data

```
%sql SELECT DISTINCT(booster_version), (SELECT MAX(payload_mass__kg_) AS "maximum_payload_mass" FROM SPACEXTBL) FROM SPACEXTBL limit 5
```

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

Booster_Version	(SELECT MAX(payload_mass__kg_) AS "maximum_payload_mass" FROM SPACEXTBL)
F9 v1.0 B0003	15600
F9 v1.0 B0004	15600
F9 v1.0 B0005	15600
F9 v1.0 B0006	15600
F9 v1.0 B0007	15600

2015 Launch Records

- The failed landing outcomes in drone ship , their booster version and launch site names for the year 2015

```
%sql SELECT landing_outcome, booster_version, launch_site, DATE FROM SPACEXTBL WHERE landing_outcome LIKE '%Failure (drone ship)%' AND (DATE LIKE '2015%') ;
```

 * sqlite:///my_data1.db
Done.

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

- 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 "total" FROM SPACEXTBL WHERE (DATE BETWEEN '2010-06-04' AND '2017-03-20')
```

```
 * ibm_db_sa://ycy00214:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

landing_outcome	total
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

Section 3

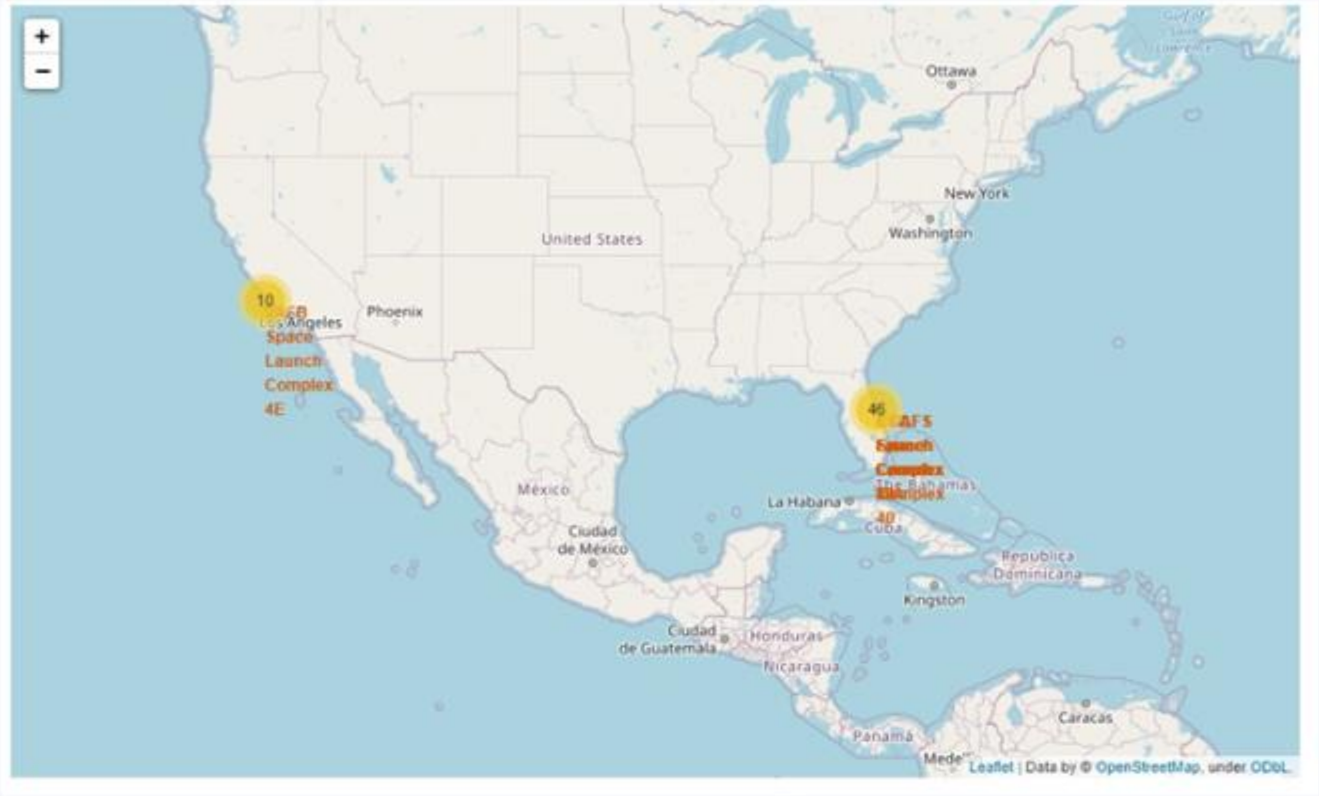
Launch Sites Proximities Analysis

All Launch Sites

- All launch sites are in very close proximity to the coast and into restricted areas

Success/Failed Launches for Each Sites

- The frist map shows clusters for every launch sites, the second shows a green marker if launch was successful and a red marker if a launch failed.



Launch site and its proximities

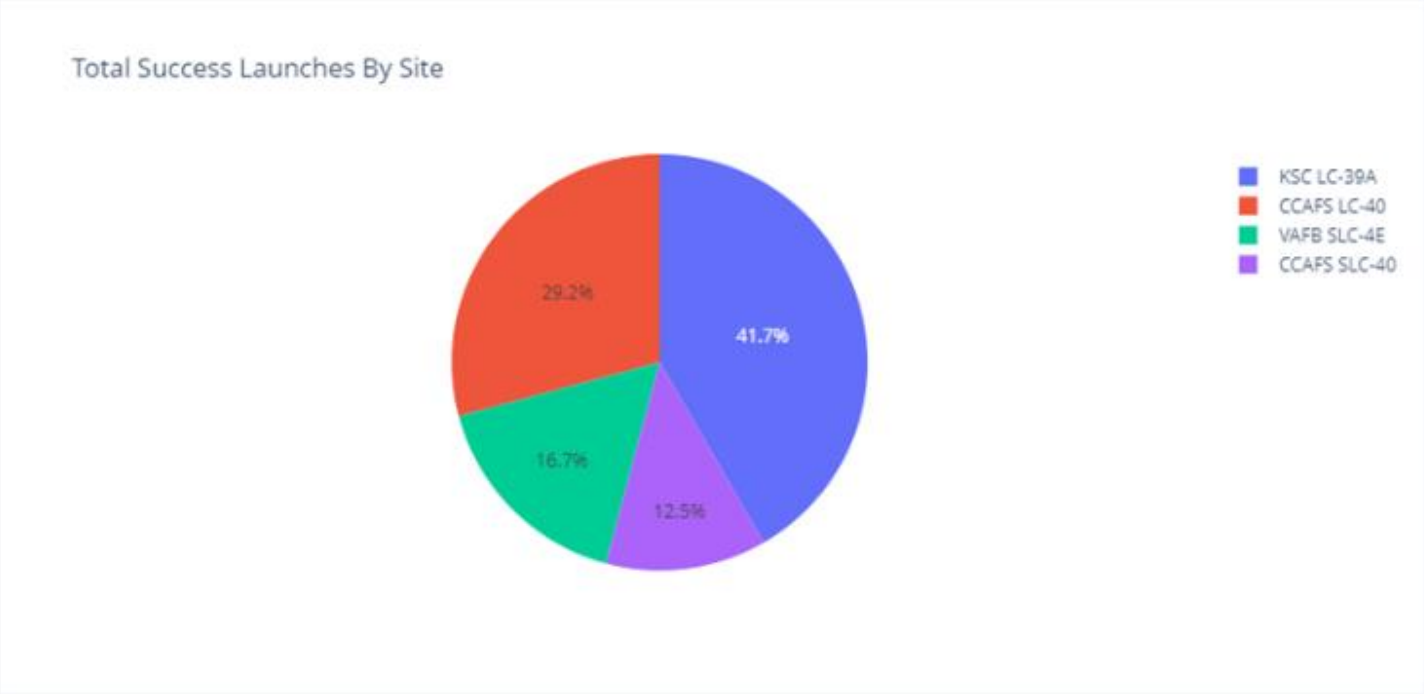
- Launch site are near railways, roads, highways and coastlines.

Section 4

Build a Dashboard with Plotly Dash

Total Success Launches by site

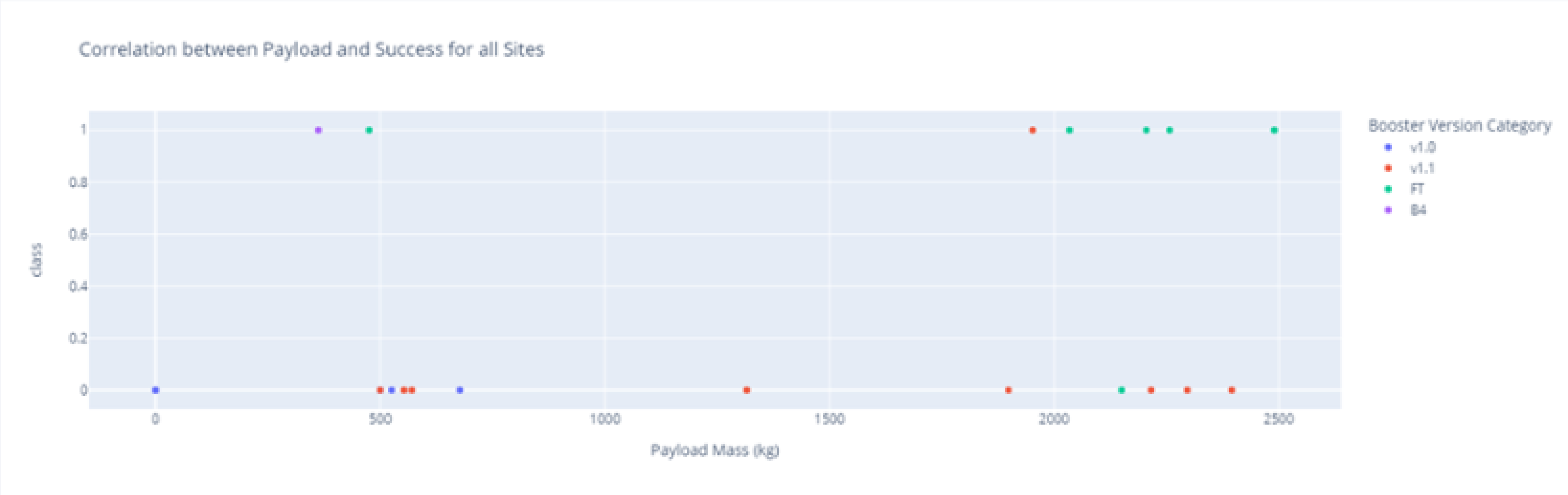
KSC LC-39A is the site with the higher success launches followed by CCAFS LC-40.



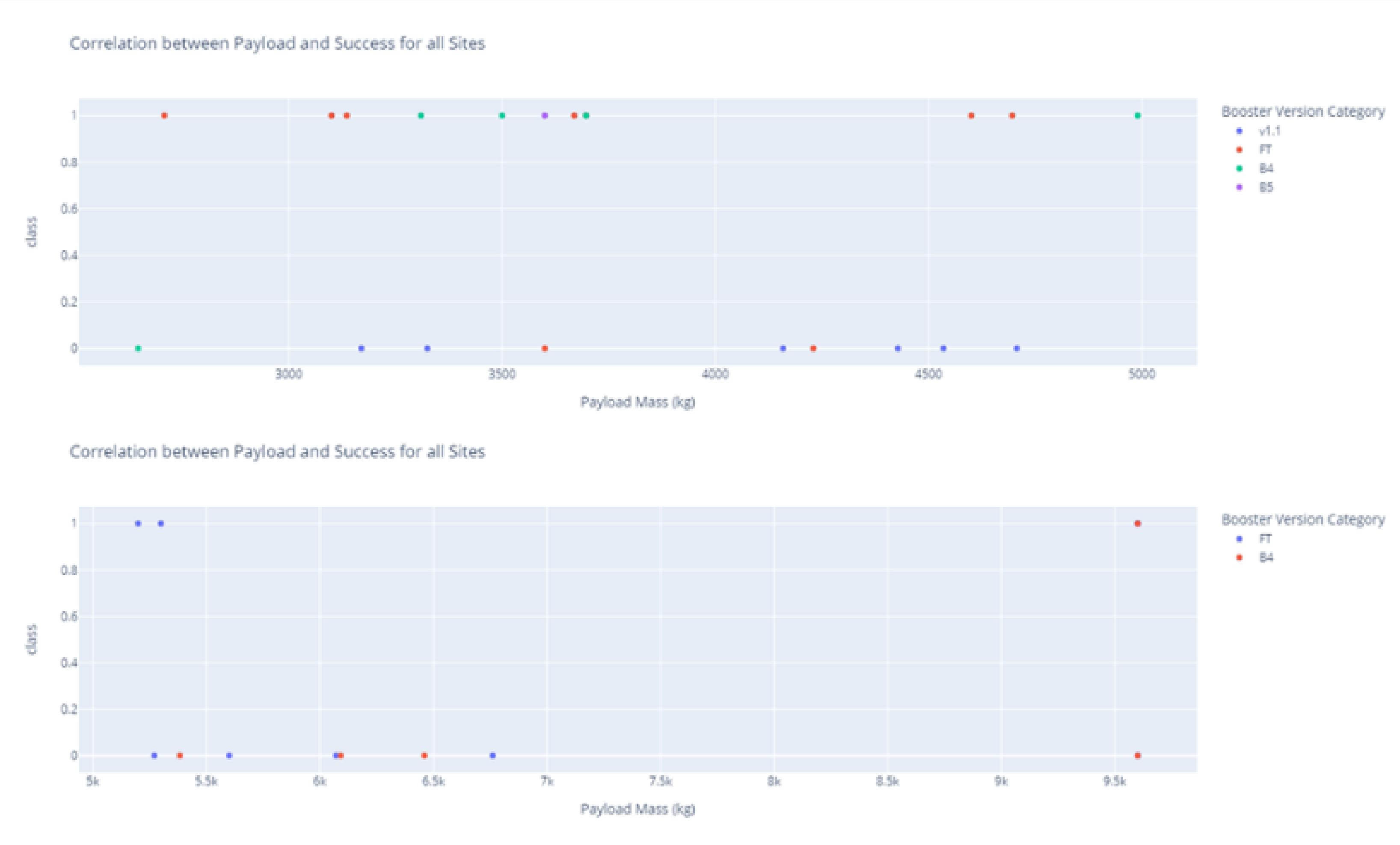
Payload vs Launch outcome

Scatter plot for all sites with 2500(kg), 5000(kg) and 10000(kg) payload ranges.

The 2500-5000(kg) range concentrate the majority of the successfully launches, the 0-2500(kg) range has most failed launches but all three are similar.



Payload vs Launch Outcome

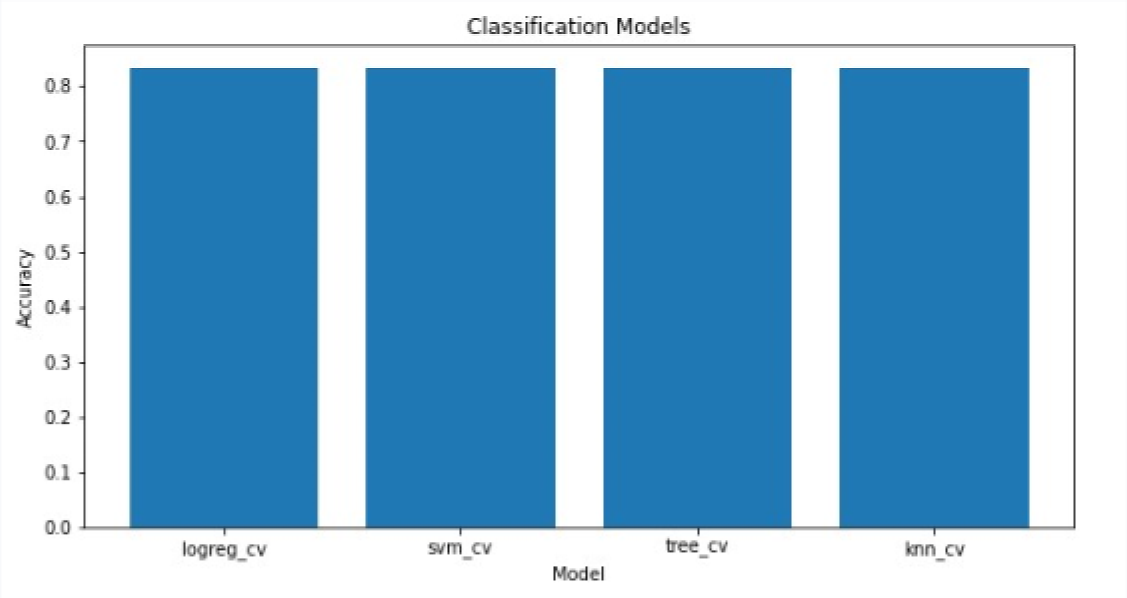


Section 5

Predictive Analysis (Classification)

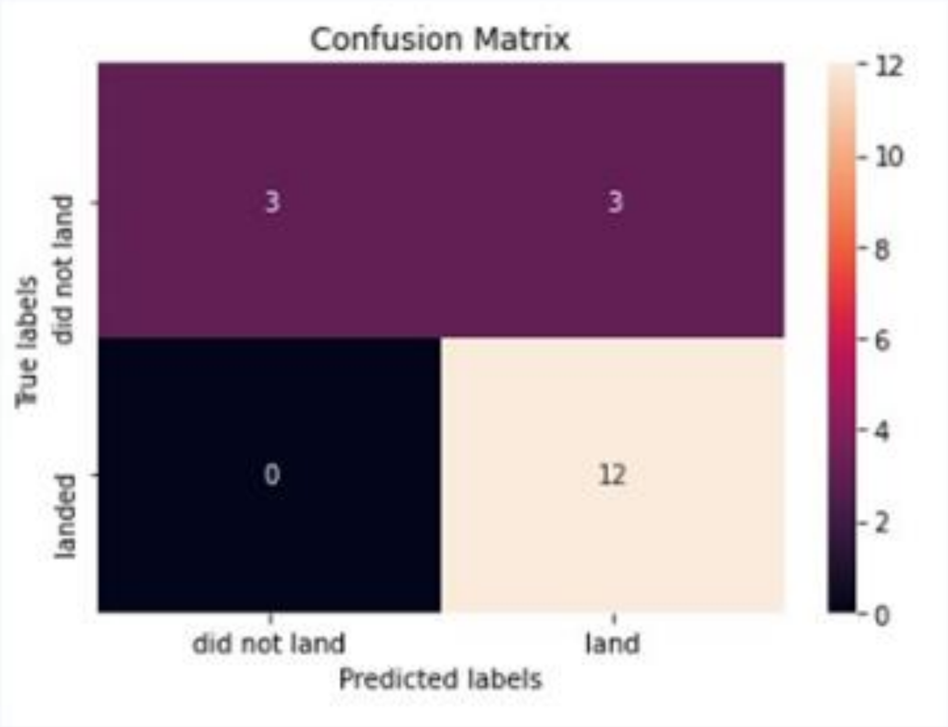
Classification Accuracy

- The Accuracy is the same for all models



Confusion Matrix

- The confusion Matrix is the same for all models



Conclusions

- As all algorithms are giving the same accuracy, they perform same
- By using our machine learning model , we can predict if the first stage of our competitor will land and determine the cost of a launch

Appendix

- For the notebooks , check the link below to the GitHub repository

Data Science Captsone

Thank you!