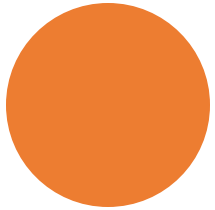
IBM Developer SKILLS NETWORK

Winning Space Race with Data Science

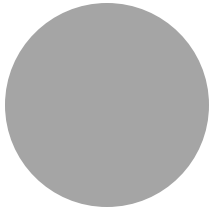
Hasan DASDEMIR

October 1, 2021

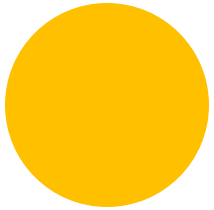
Outline



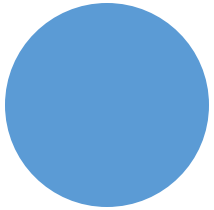
EXECUTIVE SUMMARY



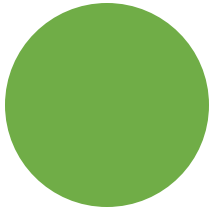
INTRODUCTION



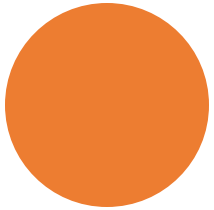
METHODOLOGY



RESULTS



CONCLUSION



APPENDIX

Executive Summary

Summary of Methodologies

- Data Collection & Wrangling
- EDA (Python & SQL)
- Interactive Analytics
  - Map (Folium)
  - Dashboard (Ploty)
- Clasification

Summary of Results

- EDA Result
- Analytics - Dashboard
- Predictive Analysis

Introduction

Project & Context

- Sending spacecraft to the International Space Station.
- If the first stage will land successfully, we can determine the cost of a launch.
- The second stage, helps bring the payload to orbit,

Problems

- Are we going to use first stage for the save money?
- We are going to try to predict whether SpaceY will attempt to land rocket or not.
- If SpaceY will reuse the first stage then we can save some money

Methodology

Methodology

Data collection	Data Wrangling	EDA	Visual Analystics	Predictive Analysis
• SpaceX API • Launches, rocket info, payload delivered, launch & landing specifications, and landing outcome. • Web Scraping Wikipedia	• Dropping unnecessary columns • Finding features	• Matplotlib • SQL	• Folium • Ploty Dash	• Classification

Data Collection

SpaceX API
Githup URL

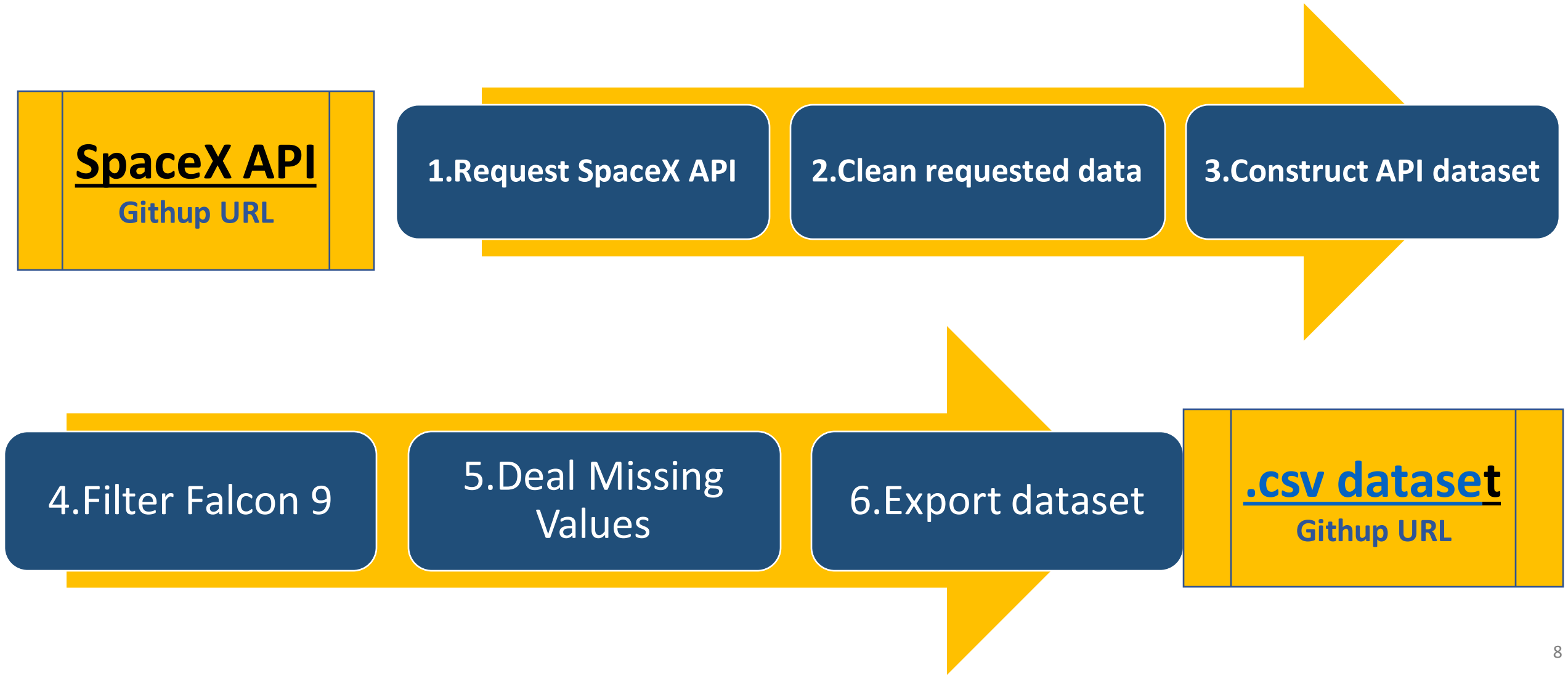
Web Scraping
Githup URL

EDA
Githup URL

EDA Dataviz
Githup URL

EDA SQL
Githup URL

Data Collection - SpaceX API



Data Collection - Web Scraping

Web Scraping
Githup URL

1.Request Falcon 9 Wiki Page

2.Create BeautifulSoup

3.Extract all columns

4.Parsing HTML table

5.Creat & fill launch dict.

6.Export dataset

.csv dataset
Githup URL

Data Wrangling

Perform exploratory Data Analysis and determine Training Labels

EDA
Githup URL

1.Load
Space X dataset

2.Calculate
Sites Launches

3.Calculate
Orbit occurrence

4.Calculate
outcome orbit type

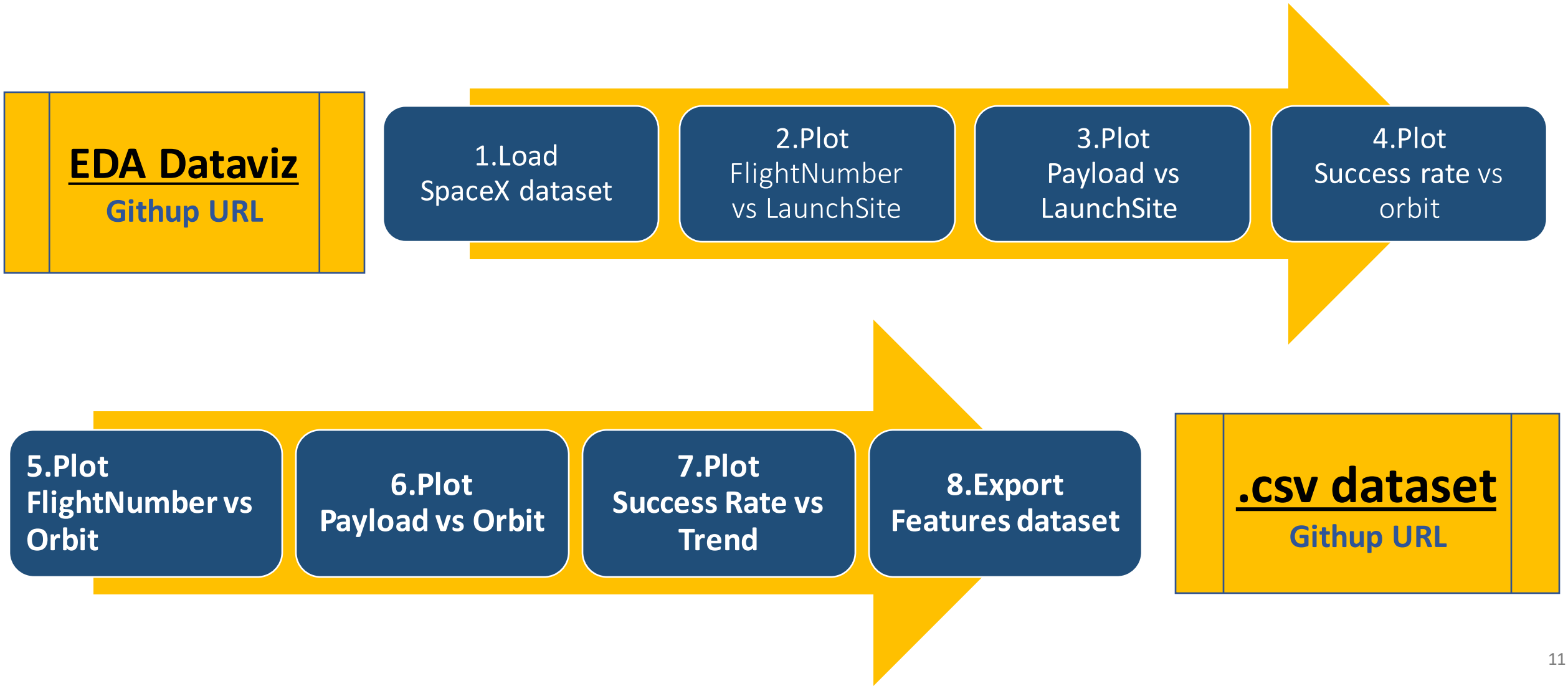
5.Create
landing outcome label

6.Export
dataset

.csv dataset
Githup URL

EDA with Data Visualization

Perform Wxploratory Data Analysis and determine Training Labels or Features



EDA with SQL

Getting more information to understan the SpaceX Dataset

EDA SQL

Githup URL

1. Display the names of the unique launch sites in the space mission
2. Display 5 records where launch sites begin with the string 'CCA'
3. Display the total payload mass carried by boosters launched by NASA (CRS)
4. Display average payload mass carried by booster version F9 v1.1
5. List the date when the first successful landing outcome in ground pad was acheived.
6. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. List the total number of successful and failure mission outcomes
8. List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
9. List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
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

Build an Interactive Map with Folium

Launch Sites Location Analysis / Interactive Visual Analytics with Folium

Map Folium

Githup URL

1.Mark
Launch sites on a map

- To see
  Exact Launch locations

2.Mark
the success/failed launches

- To see
  which sites have high success rates

3.Calculate
the distances between a launch site to its proximities

- To see
  Points of interests
- Railways
- Highways
- Costline
- Cities

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

interactive visual analytics on SpaceX launch data in real-time

Dashboard with Ploty

Githup URL

1. Add Drop-down

- To show and select a Launch Sites

2.Add Callback **function**

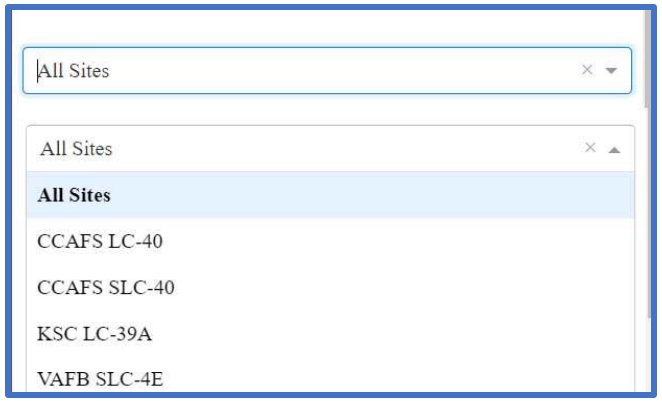
- To render `success-pie-chart` based on selected site dropdown

3.Add Range **Slider**

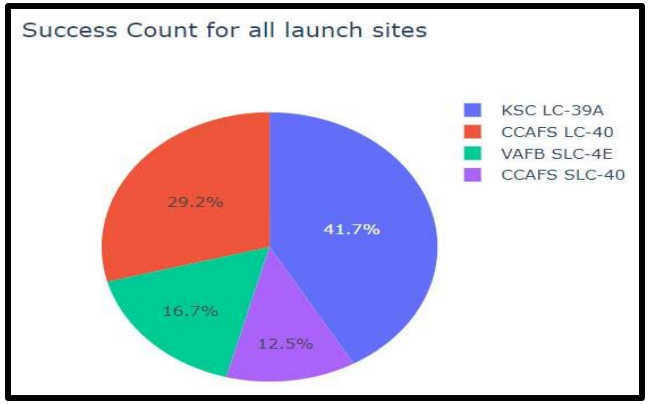
- To Select Payload

4.Add Callback **function**

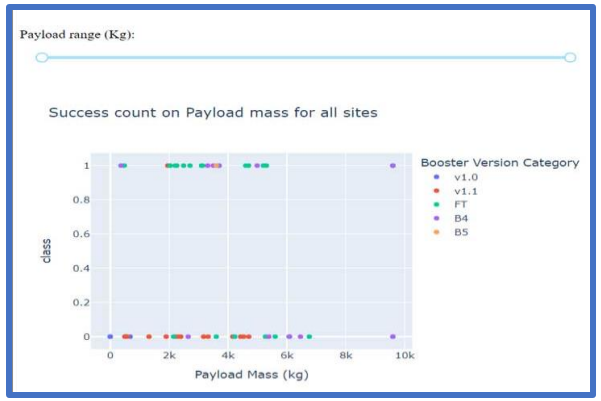
- To render `success-payload-scatter-chart` scatter plot



DropDown : Show and select all or single site to interact pieChart



PieChart : Shows Success Count for all or single site



Scatter : Shows the relationship Outcome and Payload Mass.

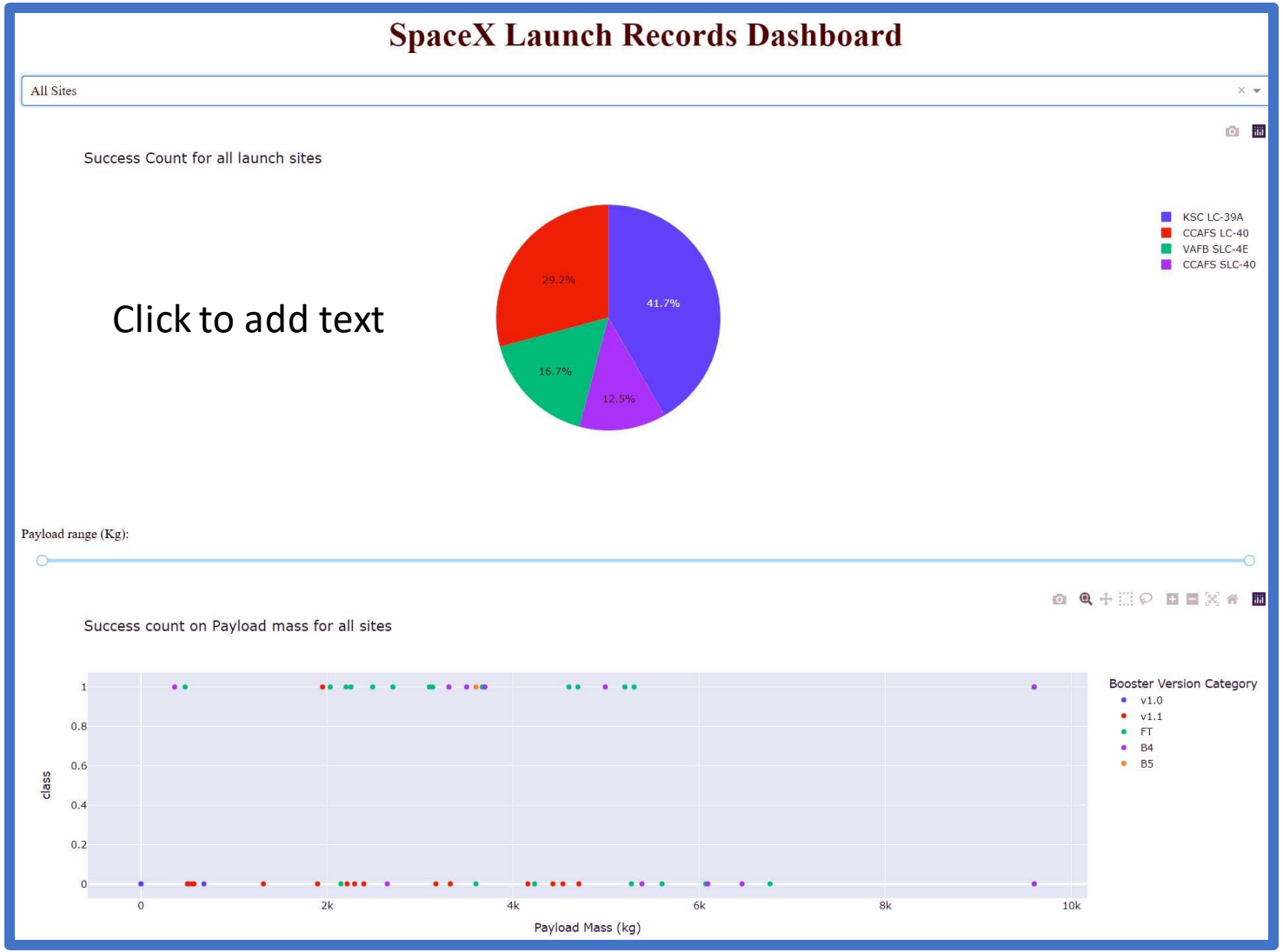
See FullScreen Dash on next slide

Build a Dashboard with Plotly Dash

interactive visual analytics on SpaceX launch data in real-time

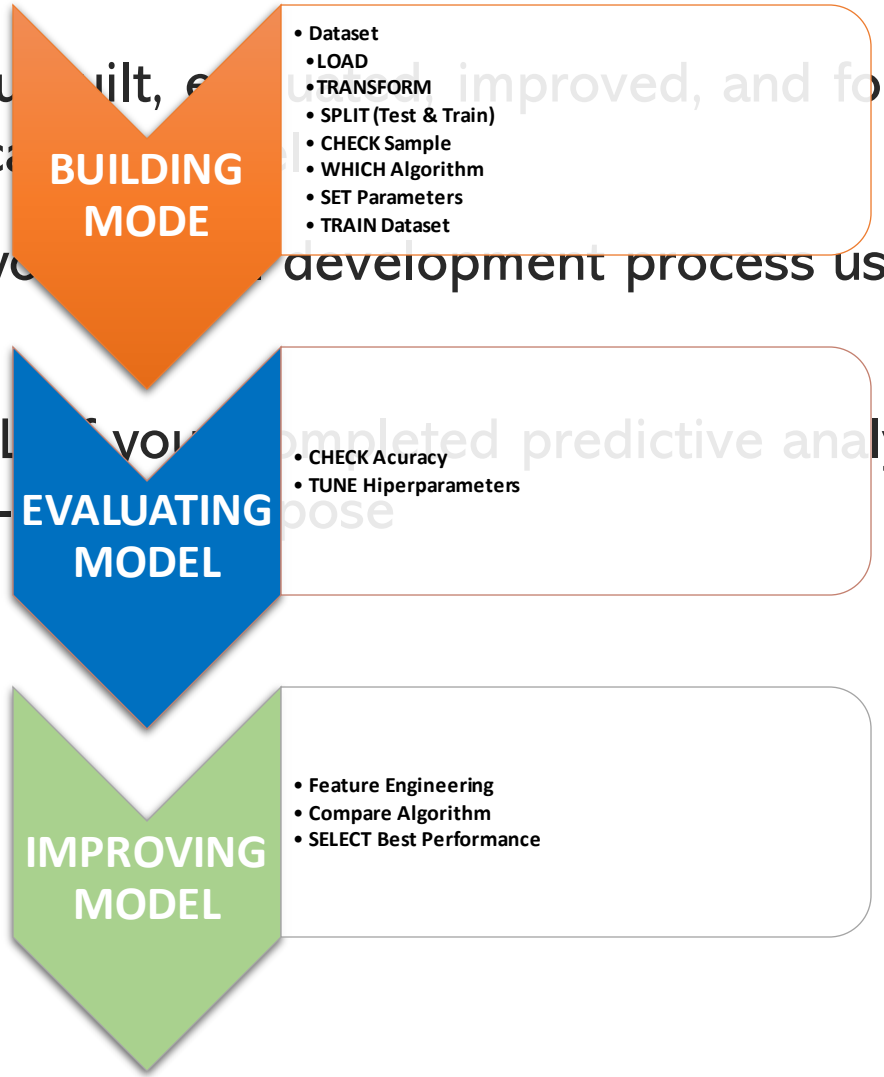
Dashboard with Ploty

Githup URL



Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose



Predictive Analysis (Classification)

Machine Learning Prediction

Githup URL

BUILDING MODE

- **Dataset**
- LOAD
- TRANSFORM
- SPLIT (Test & Train)
- CHECK Sample
- WHICH Algorithm
- SET Parameters
- TRAIN Dataset

EVALUATING MODEL

- CHECK Acuracy
- TUNE Hiperparameters

IMPROVING MODEL

- FEATURE Engineering
- COMPARE Algorithm
- SELECT Best Performance

Result

Exploratory data analysis / Interactive Analytics / Predictive Analysis

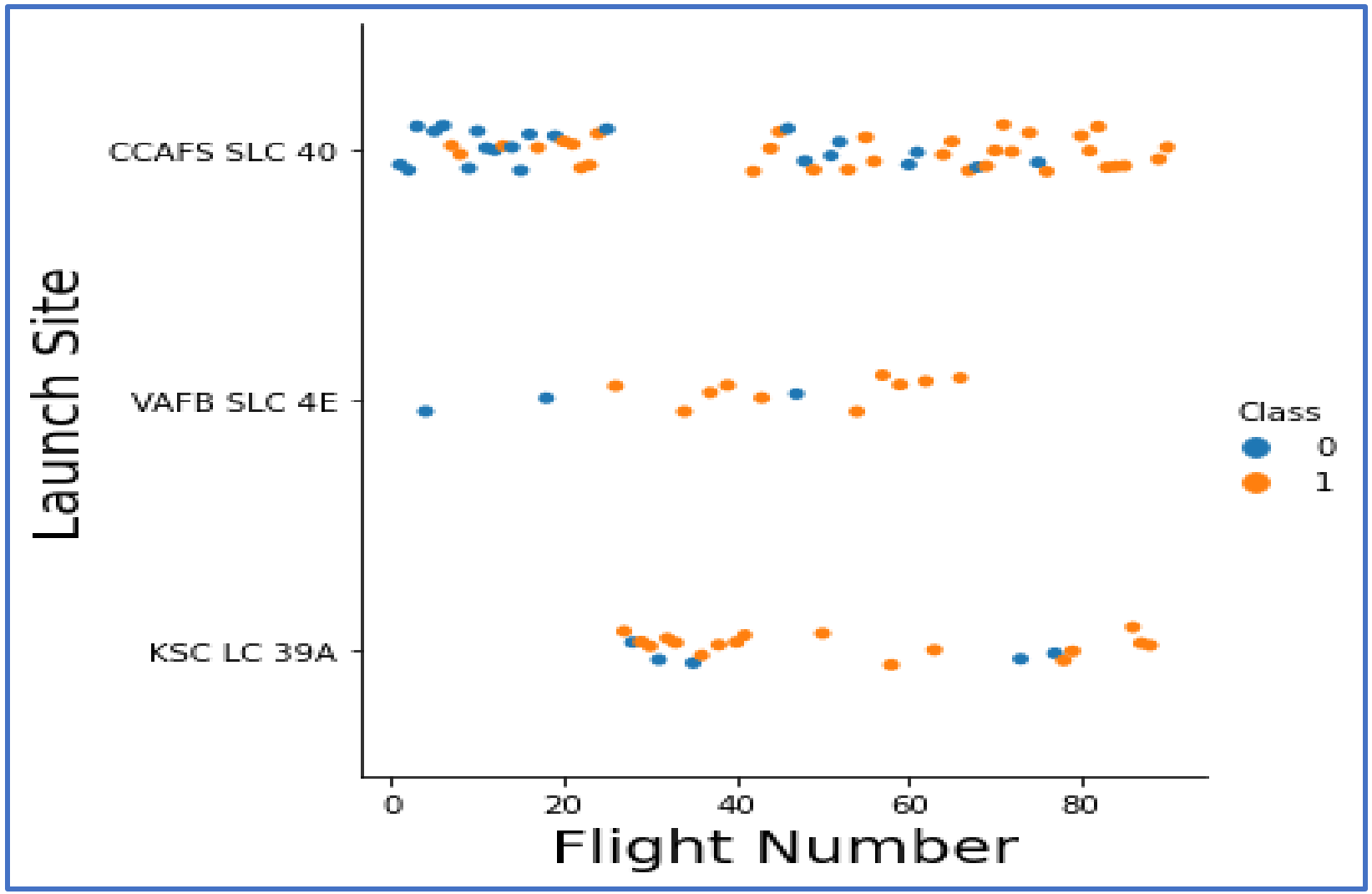
Exploratory data analysis results

Interactive analytics demo in screenshots

Predictive analysis results

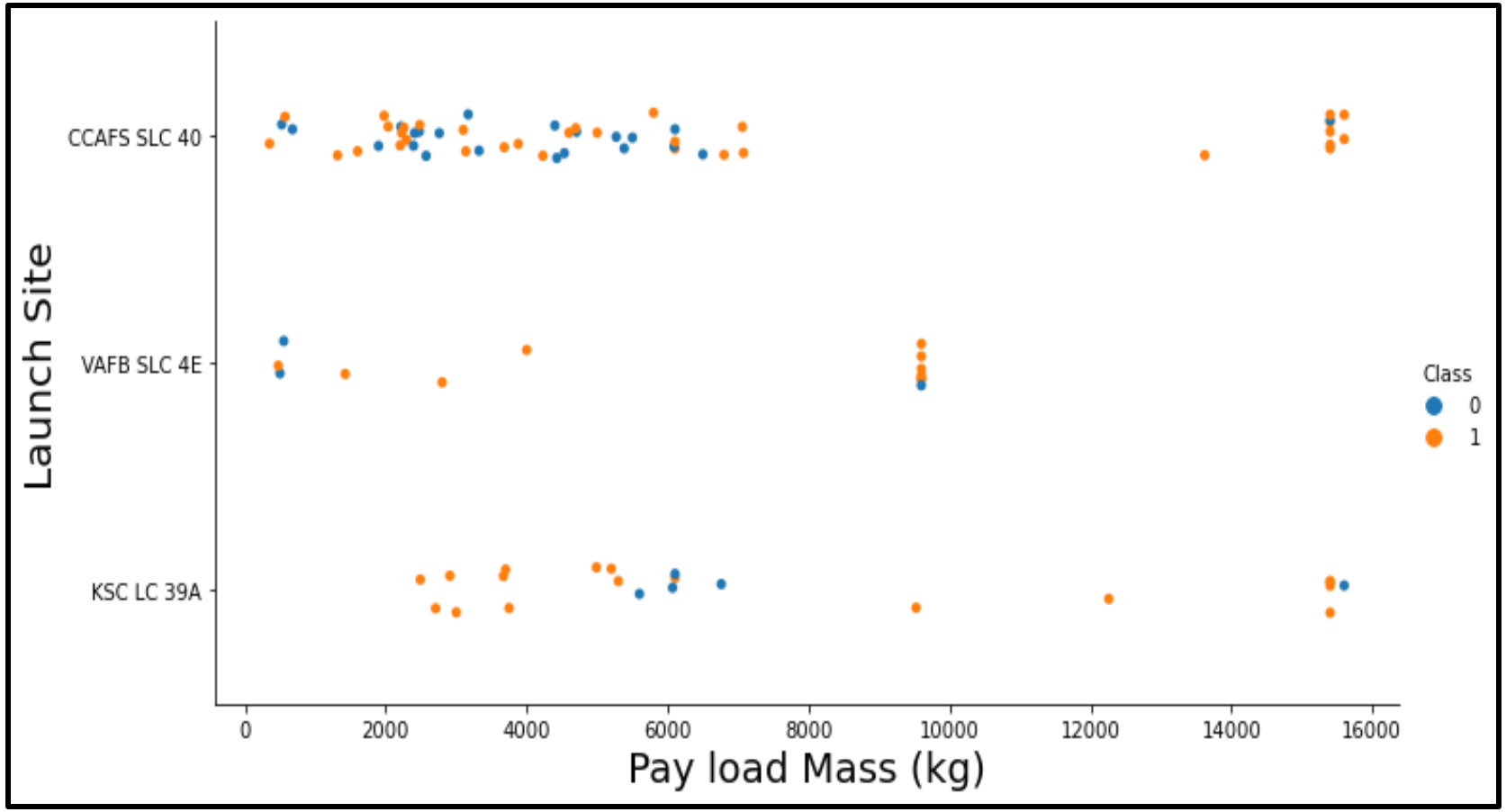
Insights drawn from EDA

Flight Number vs. Launch Site



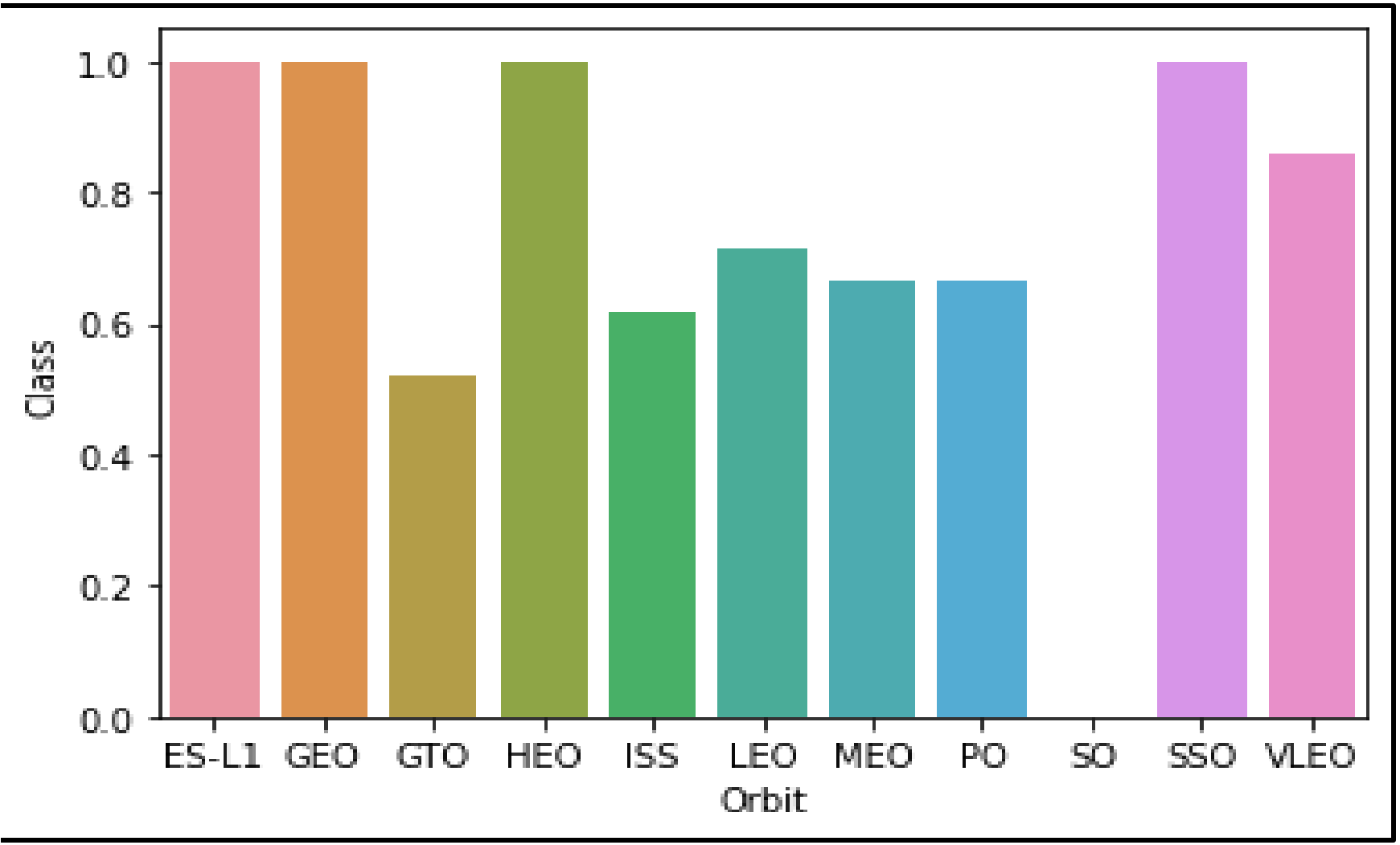
Especially, CCAFS-SLC 40 site is
the most popular site than the other sites.

Payload vs. Launch Site



CCAFS-SLC 40 and KSC LC 39A are the most popular site than the other But they have low Pay Load Mass rate.

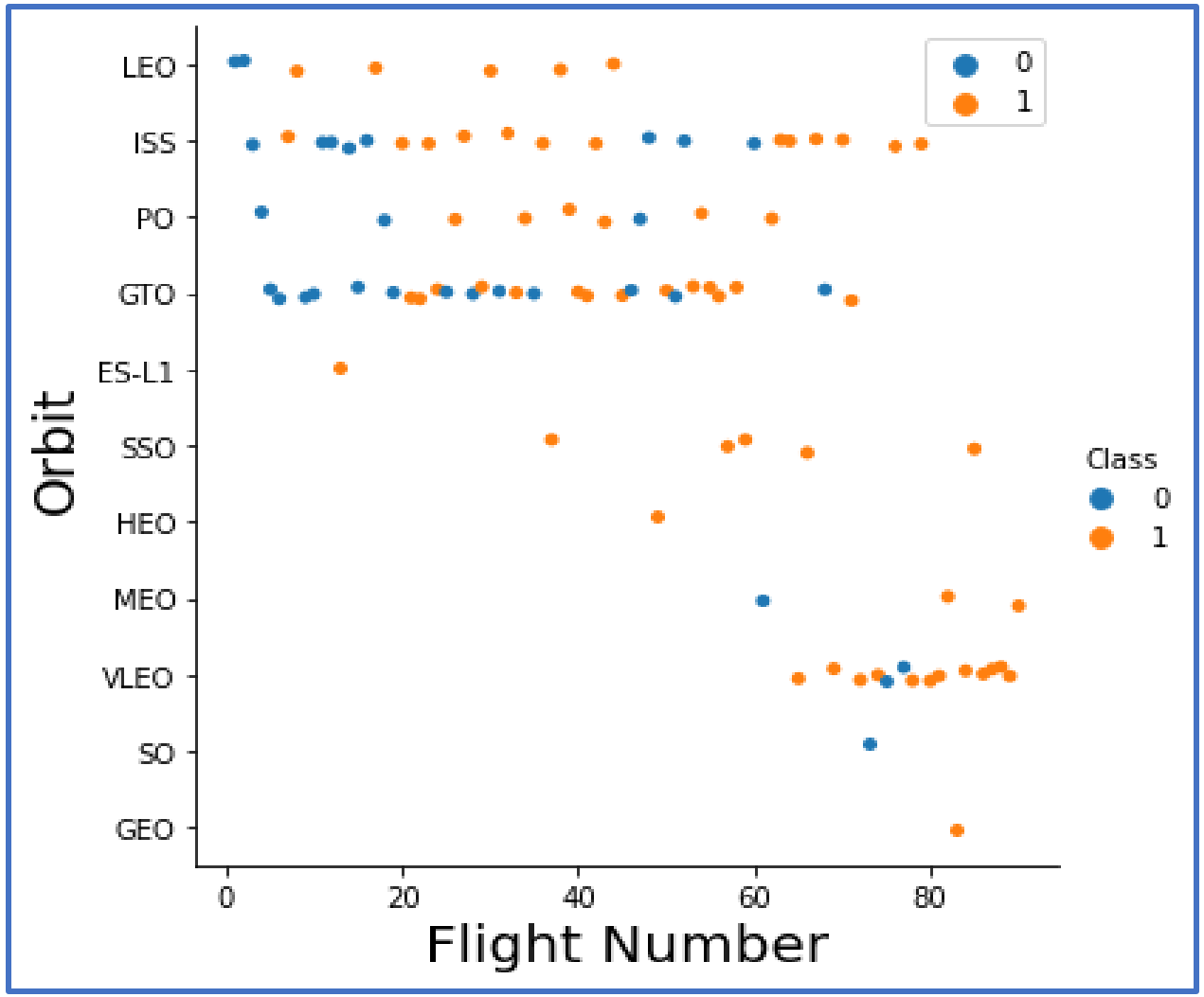
Success Rate vs. Orbit Type



ES-LI, GEO, HEO and SSO have the best success rate than the others

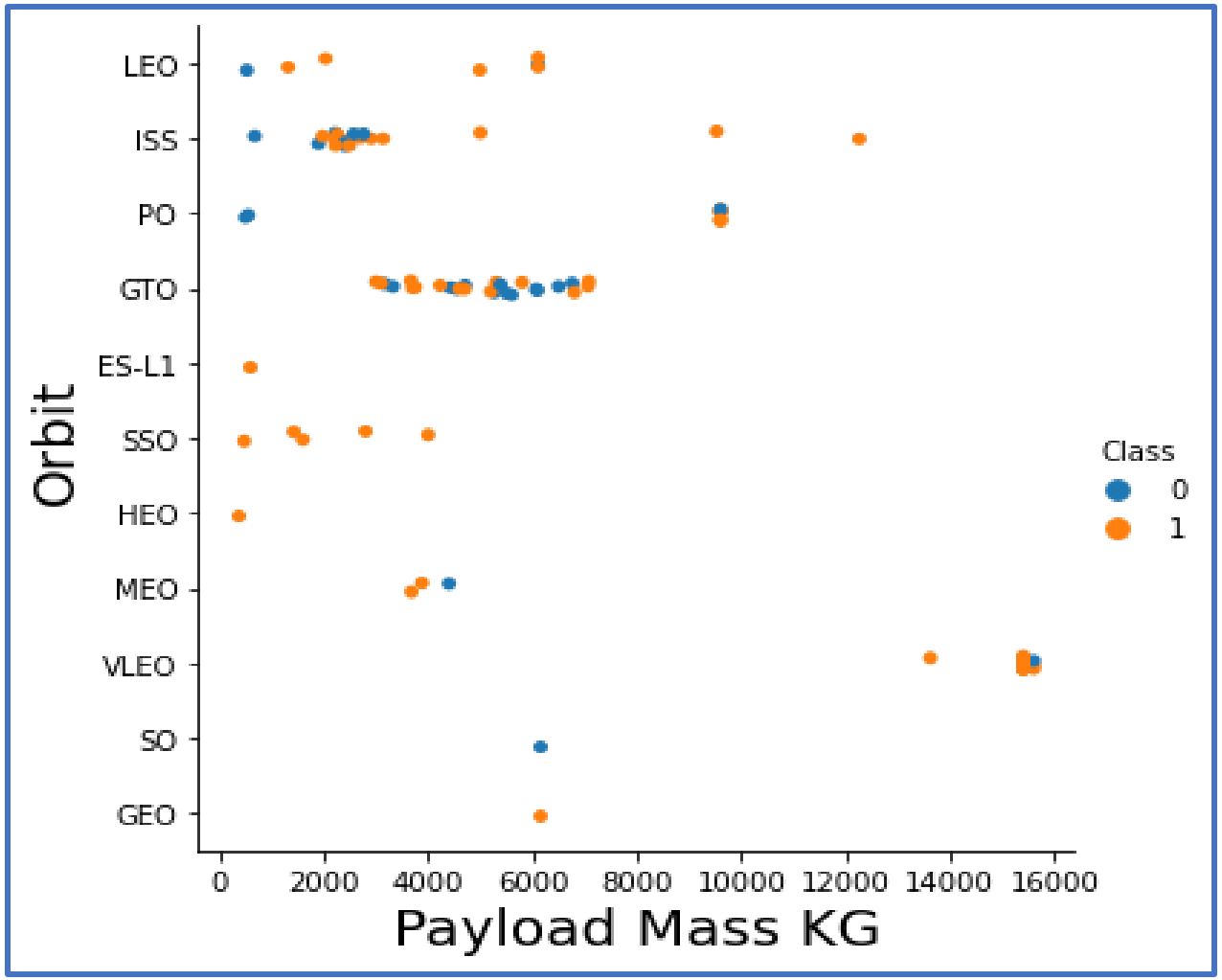
GTO has lowest success rate

Flight Number vs. Orbit Type



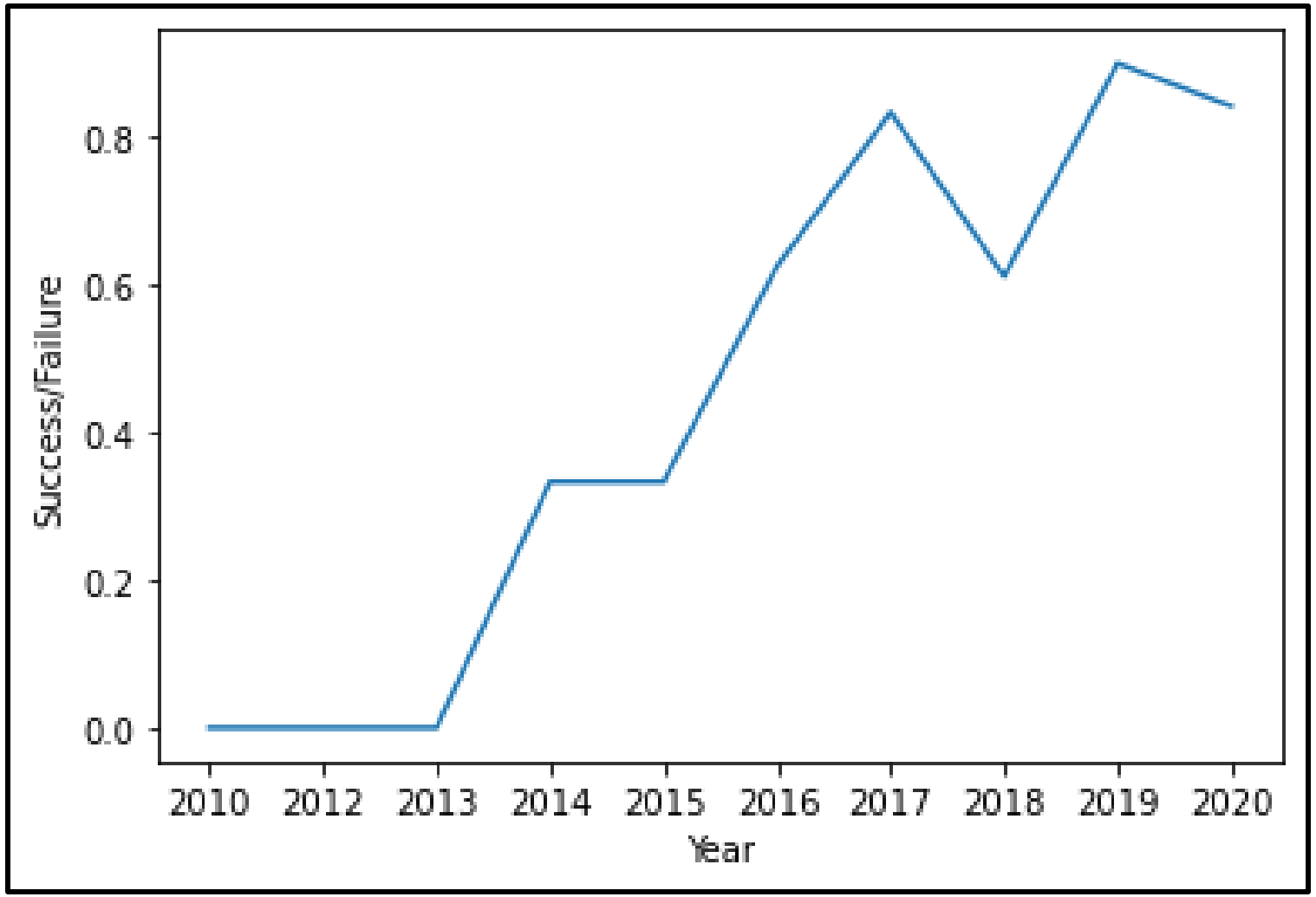
Especially, LEO orbit has most affordable orbit. Also, we cannot seen any relationship between Flight number and Orbit. Do you ?

Payload vs. Orbit Type



GTO and ISS have heavy pay loads then the other orbit!

Launch Success Yearly Trend



Success rate started to increase in 2013 and kept skyrocketing till 2020. 2018 and 2019 are also not bad!

EDA with SQL

All Launch Site Names

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

In [5]:
```
%sql SELECT DISTINCT launch_site FROM SPACEXTBL
```

Out[5]:

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

If we use DISTINCT keyword in SQL command to get unique site names.

Launch Site Names Begin with 'CCA'

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

In [6]:
```
%sql SELECT LAUNCH_SITE \
        FROM SPACEXTBL \
        WHERE LAUNCH_SITE LIKE 'CCA%' \
        LIMIT 5
```

Out[6]:

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

If we use LIKE keyword with 'CCA%' in SQL command to get site name starts with "CCA". Using LIMIT 5 keywords to get top five records

Total Payload Mass

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

```
In [9]: %sql SELECT sum(PAYLOAD_MASS__KG_) AS "Total Payload Mass" FROM SPACEXTBL WHERE customer='NASA (CRS)'
```

Out[9]:

Total Payload Mass
45596

Selected only "NASA (CRS)" records to get Total Payload Mass with SUM() function; Total Payload Mass is 45596

Average Payload Mass by F9 v1.1

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

In [10]:
```
%sql SELECT avg(PAYLOAD_MASS__KG_) AS "Average Payload Mass" FROM SPACEXTBL WHERE booster_version='F9 v1.1'
```

Out[10]:

Average Payload Mass
2928.400000

Selected only "F9 v1.1" records to get Avarage Payload Mass with AVG() function; Total Payload Mass is 2928.4

First Successful Ground Landing Date

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

In [11]:
```
%sql SELECT MIN(DATE) FROM SPACEXTBL WHERE mission_outcome='Success'
```

Out[11]:

1
2010-06-04

Selected only "Success" records to get first successful landing with MIN() function;
The date is 2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

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

In [12]:
```
%sql SELECT * FROM SPACEXTBL \
            WHERE (payload_mass__kg_ > 4000 and payload_mass__kg_ < 6000) AND \
            landing__outcome='Success (drone ship)'
```

Out[12]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Selected only "Success (drone ship)" records to get *payload mass greater than 4000 but less than 6000*;

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

In [13]:
```
%sql SELECT COUNT(mission_outcome) AS Total_Success FROM SPACEXTBL WHERE mission_outcome LIKE '%Success%'
```

Out[13]:

total_success
100

Selected only "Success" records to get *Total Success mission with count() function; Total Success is 100.*

In [14]:
```
%sql SELECT COUNT(mission_outcome) AS Total_Failure FROM SPACEXTBL WHERE mission_outcome LIKE '%Failure%'
```

Out[14]:

total_failure
1

Selected only "Failure" records with LIKE "%Failure" to get *Total Success mission with count(); Total Failure is 1.*

Boosters Carried Maximum Payload

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

In [16]:
```
%sql SELECT BOOSTER_VERSION as boosterversion, PAYLOAD_MASS__KG_ from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
```

Out[16]:

boosterversion	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

Selected Boosters Carried with Maximum Payloads with sub query

2015 Launch Records

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

In [17]:
```
%sql SELECT booster_version, launch_site, year(DATE), mission_outcome, landing__outcome \
        FROM SPACEXTBL WHERE year(DATE) = 2015 AND \
        landing__outcome='Failure (drone ship)'
```

Out[17]:

booster_version	launch_site	3	mission_outcome	landing__outcome
F9 v1.1 B1012	CCAFS LC-40	2015	Success	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	2015	Success	Failure (drone ship)

Selected only "Failure (drone ship)" in 2015 records to get booster version, launch site name

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

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

In [18]:
```
%sql SELECT landing__outcome, DATE FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' \
    ORDER BY DATE DESC
```

Out[18]:

landing__outcome	DATE
No attempt	2017-03-16
Success (ground pad)	2017-02-19
Success (drone ship)	2017-01-14
Success (drone ship)	2016-08-14
Success (ground pad)	2016-07-18
Failure (drone ship)	2016-06-15
Success (drone ship)	2016-05-27
Success (drone ship)	2016-05-06
Success (drone ship)	2016-04-08
Failure (drone ship)	2016-03-04
Failure (drone ship)	2016-01-17
Success (ground pad)	2015-12-22
Precluded (drone ship)	2015-06-28

Selected Landing Outcomes between 2010-06-04 & 2017-03-20 to get from SpaceX dataset!

Launch Sites Proximities Analysis

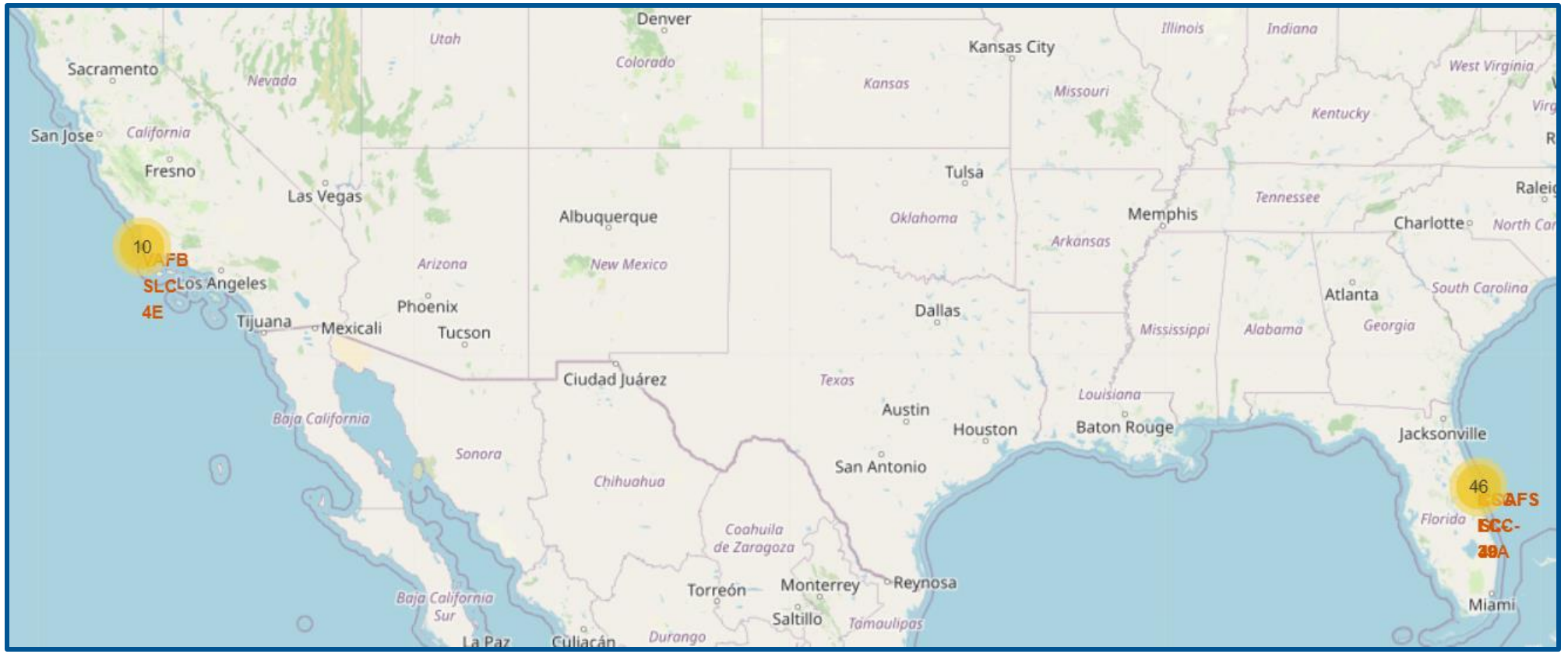
Launch Sites Proximities Analysis

Folium Map Screenshot 1

SpaceX Launch site is located in East and west cost of United States Florida, California

Launch Sites Proximities Analysis

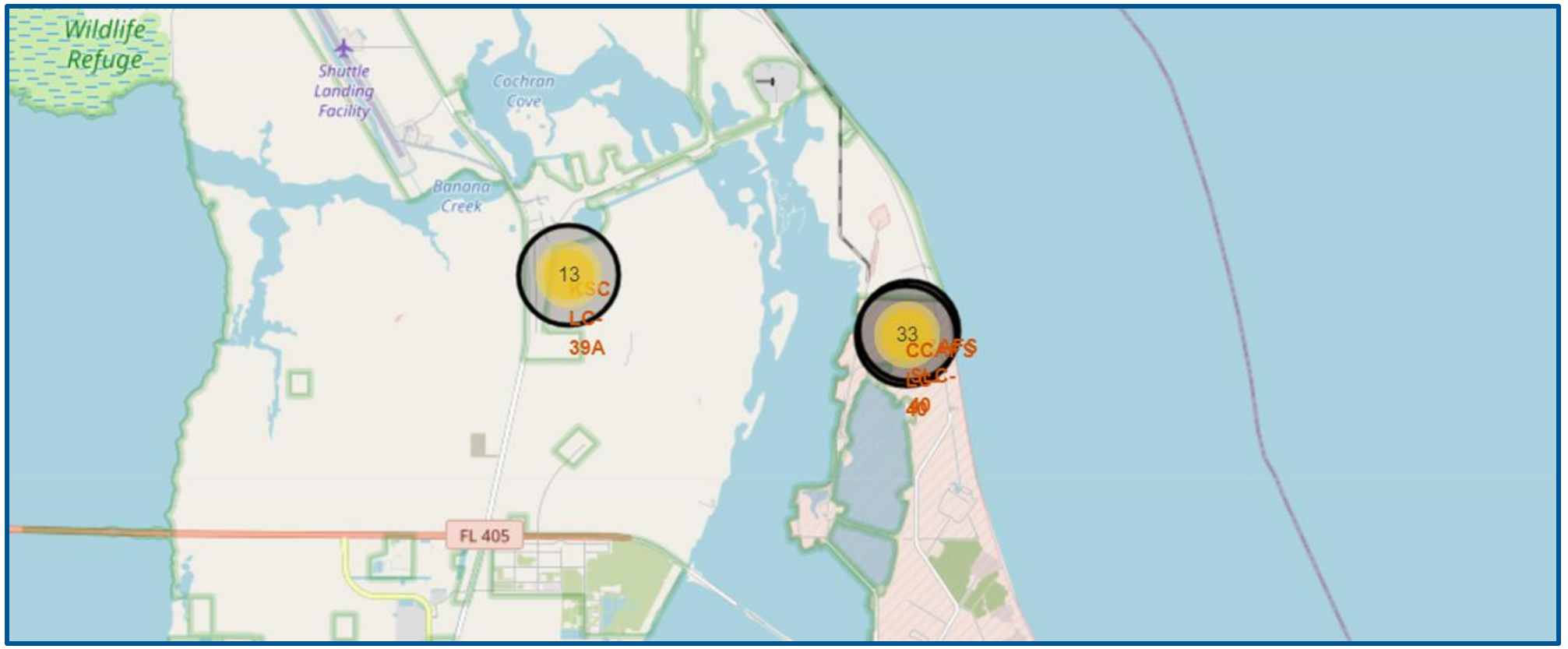
Folium Map Screenshot 2



Map shows number of the success launches per each location.

Launch Sites Proximities Analysis

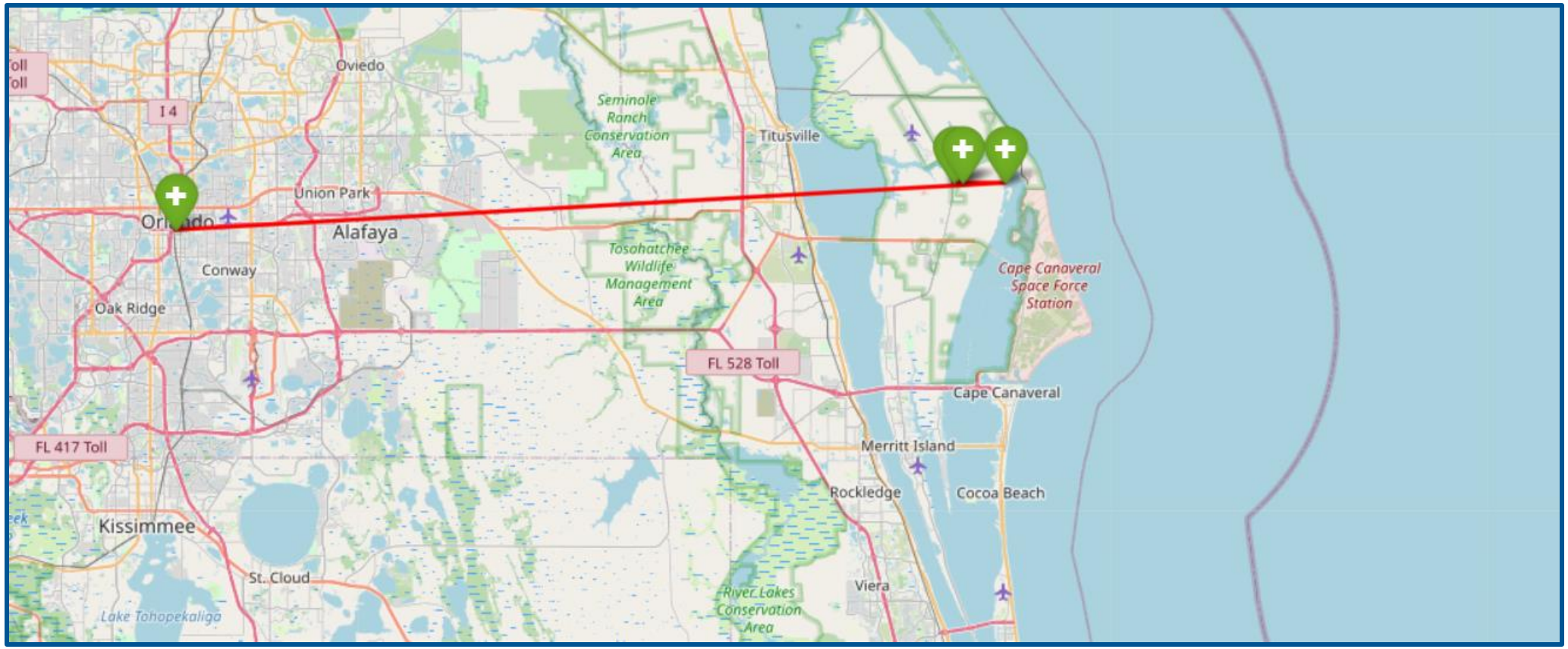
Folium Map Screenshot 3



Map shows number of the success launches per each location for Florida

Launch Sites Proximities Analysis

Folium Map Screenshot 4

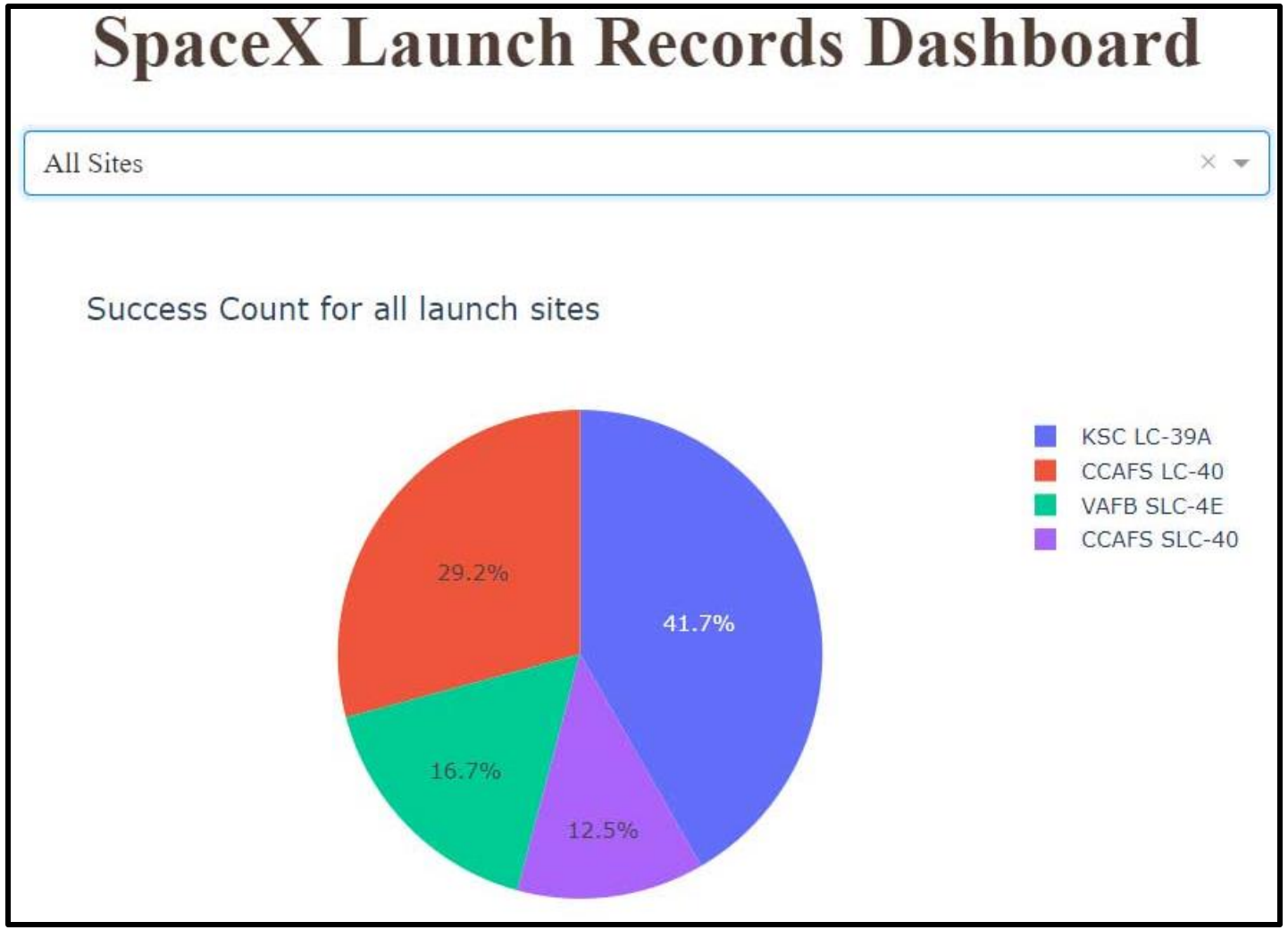


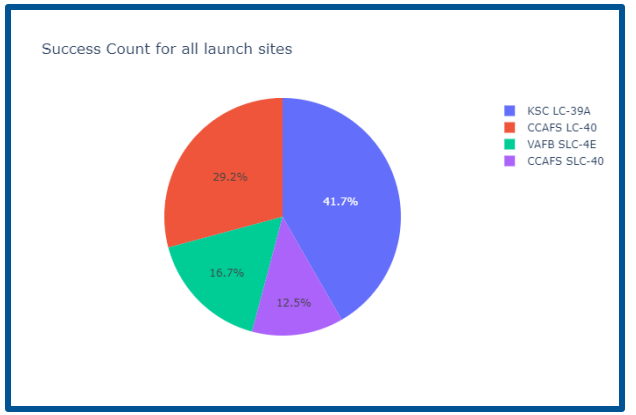
The map shows with red lines the proximity points of interest in Florida City.

Dashboard with Ploty Dash

SpaceX Launch Record Dashboard

Screenshot 1.

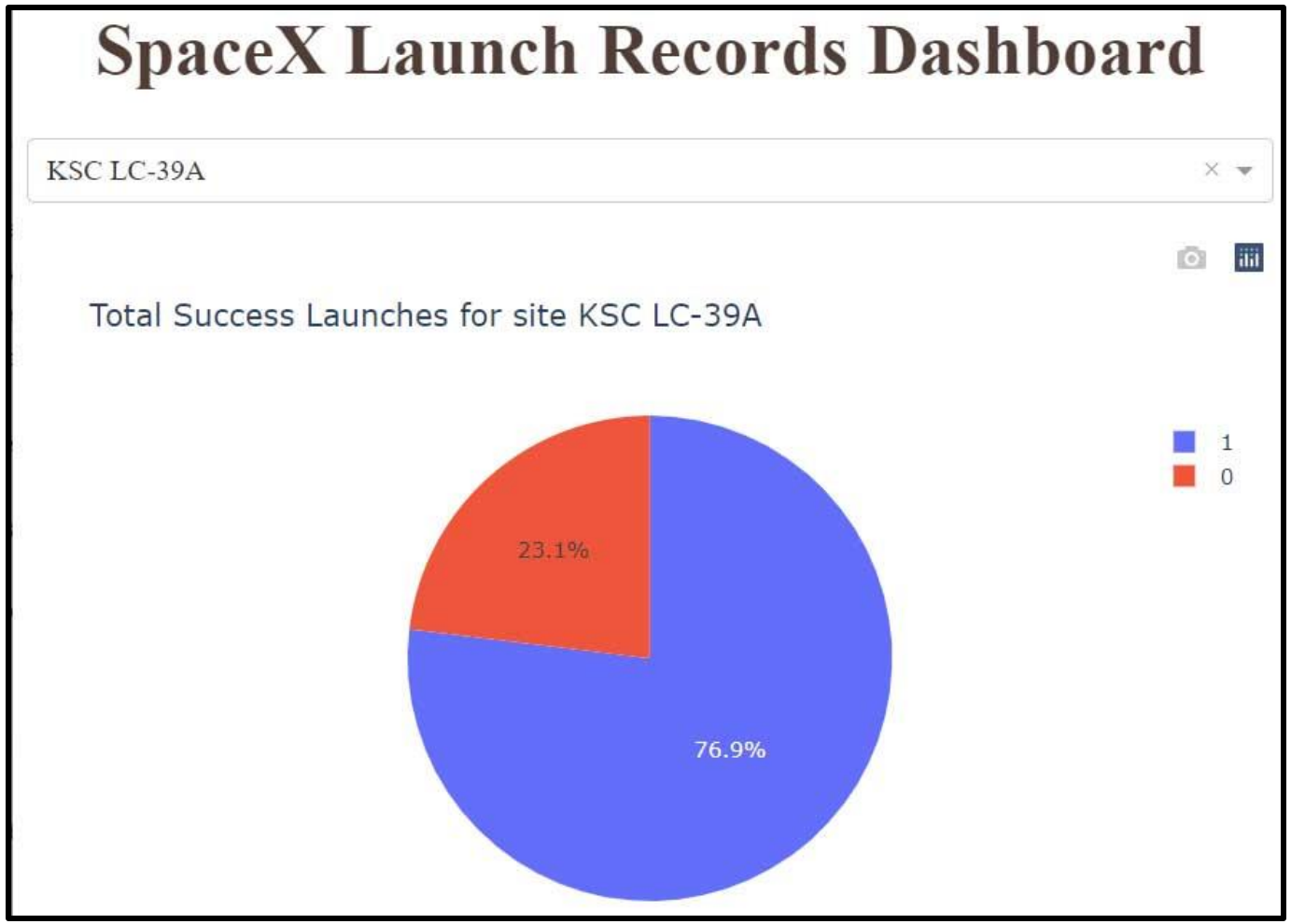


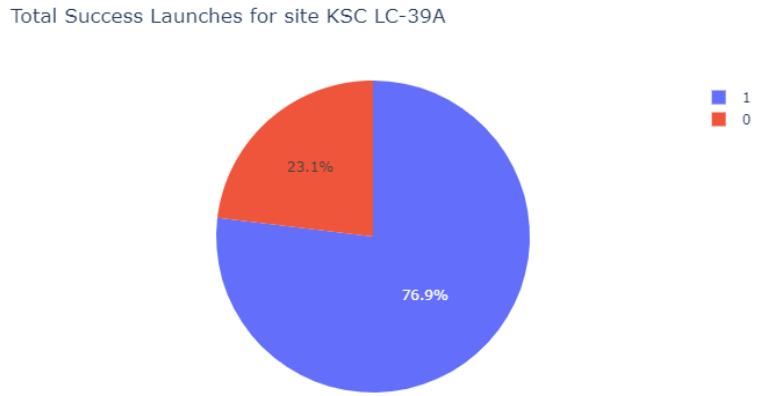


KSC LC-39A has the most successful lanunches than the other site than CCFS LC-40, VAF8 SLC-4E and CCAFS SLC-40

SpaceX Launch Record Dashboard

Screenshot 2.

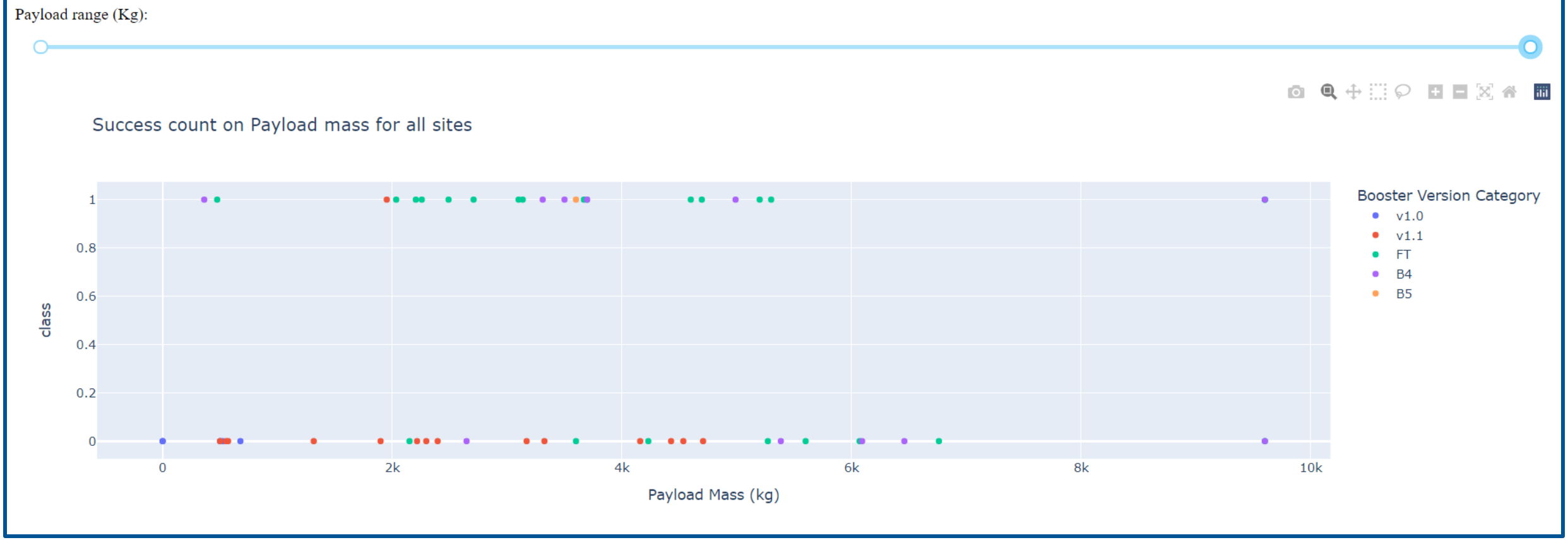




KSC LC-39A has the highest successful rate

SpaceX Launch Record Dashboard

Screenshot 3.1 same payloads with Full range slider



Booster Version Category v1.0 and v1.1 have highest success rate than the others

SpaceX Launch Record Dashboard

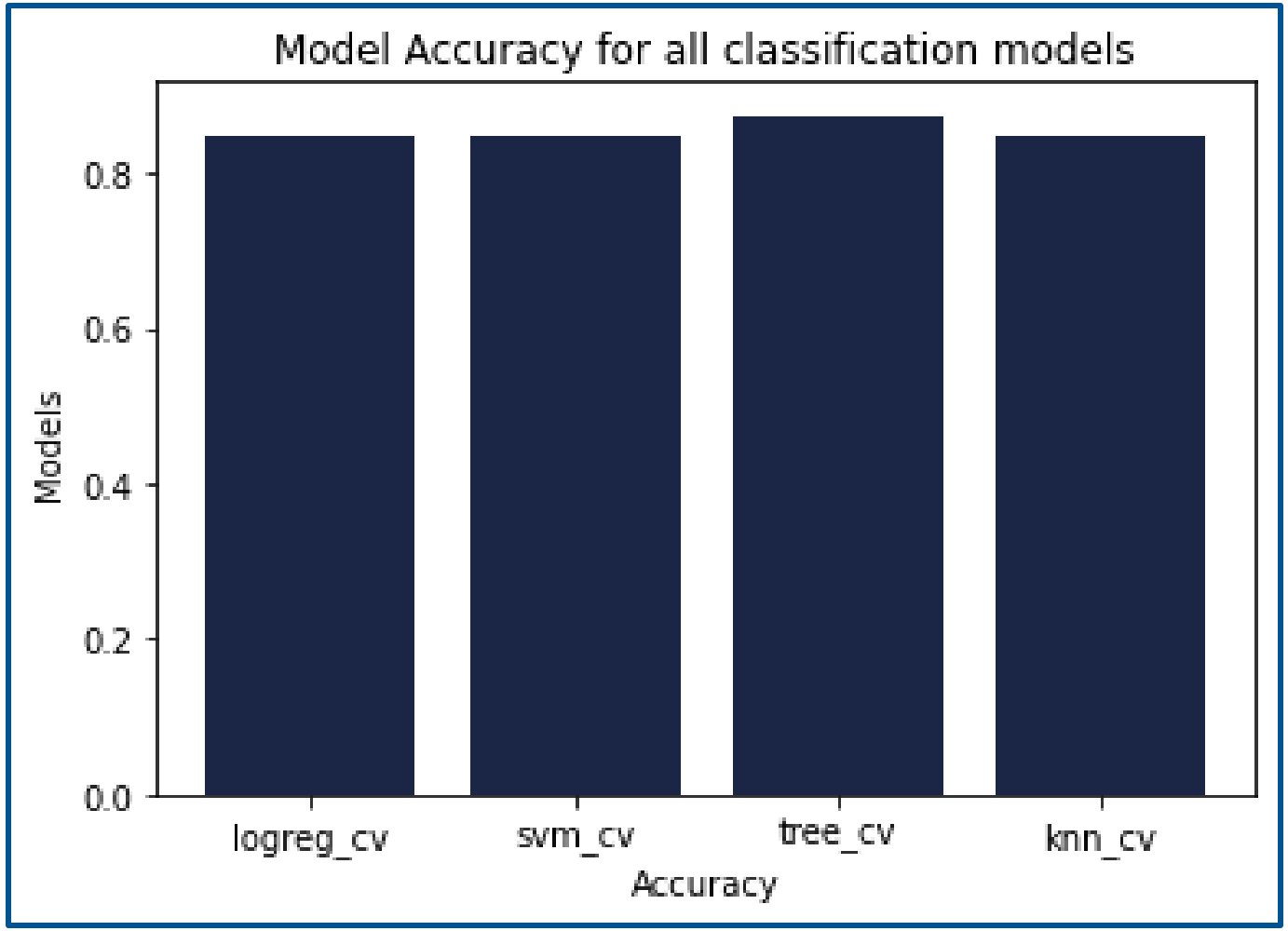
Screenshot 3.2 with different payload selected in the range slider

Booster Version Category v1.0 and v1.1 have highest success rate than the others

Predictive Analysis (Classification)

Section 5

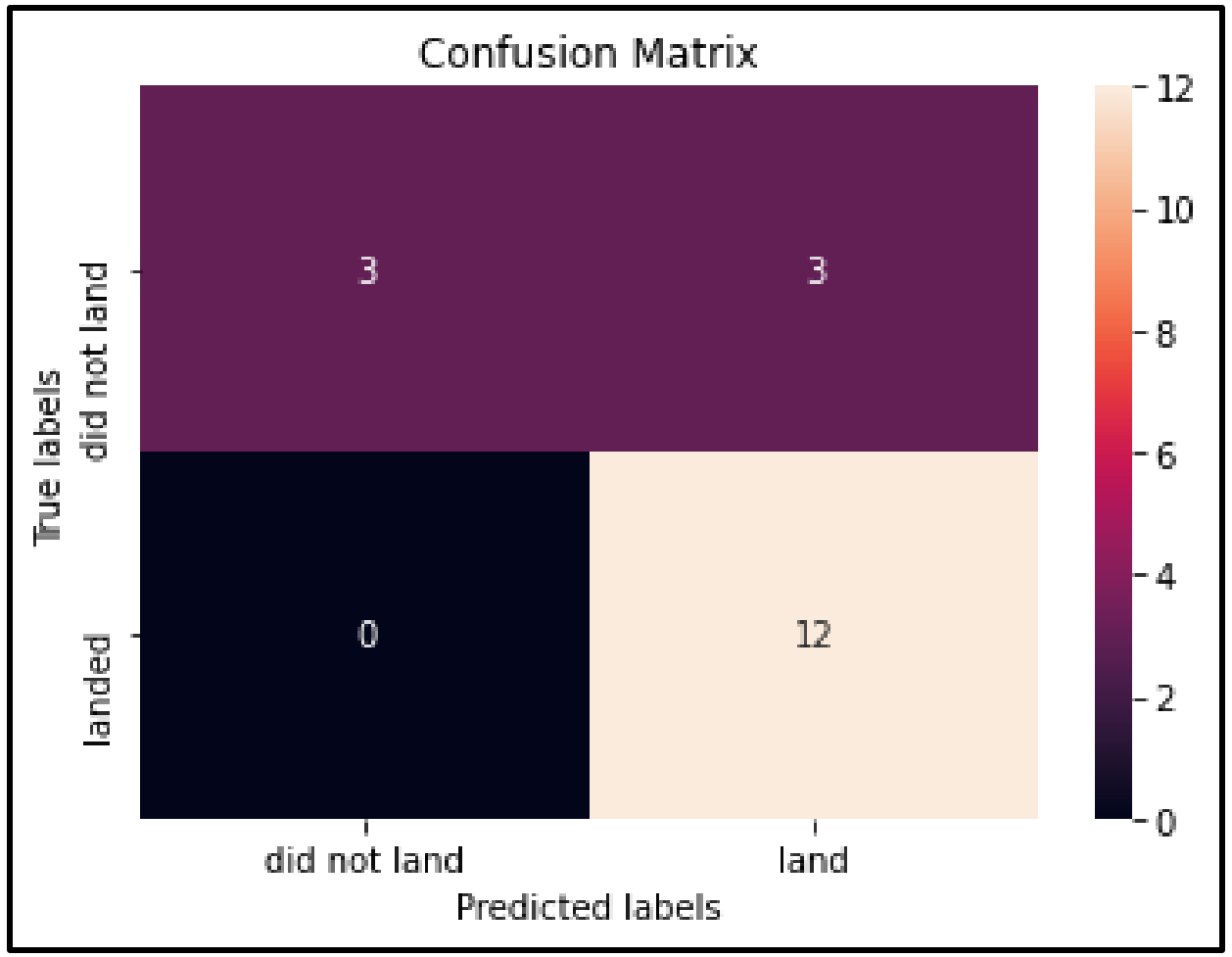
Classification Accuracy




```
scores
[104]  ✓ 0.2s
...  {'logreg_cv': 0.8464285714285713,
      'svm_cv': 0.8482142857142856,
      'tree_cv': 0.8732142857142857,
      'knn_cv': 0.8482142857142858}
```

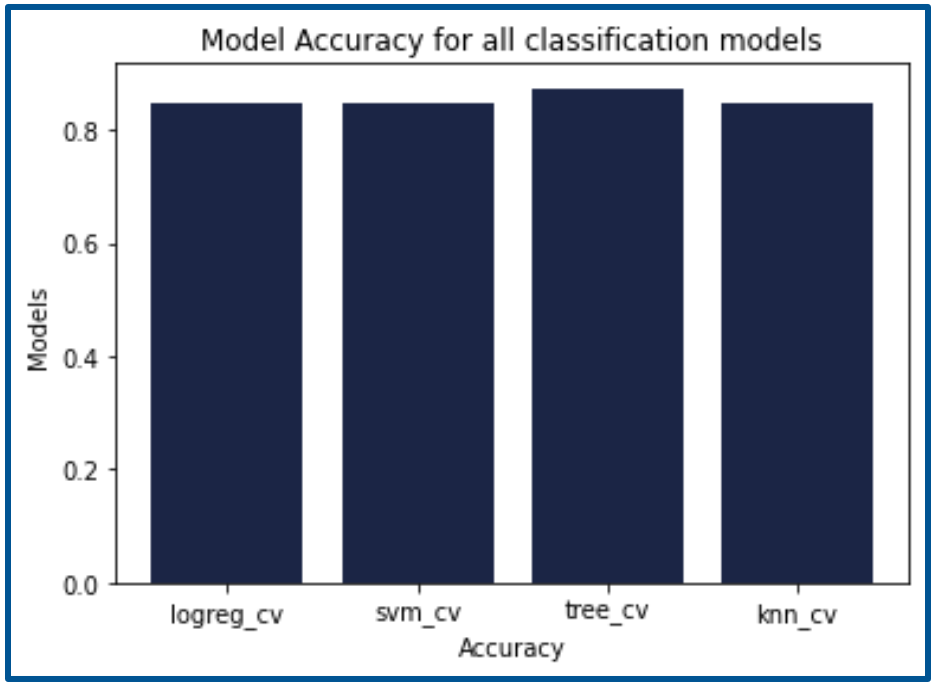
Decision Tree Clasifieris the highest score 0.87 accuracy

Confusion Matrix



The Confusion Matrix for the Decision Tree model

Conclusions



Tree Classifier Algorithm the best score 0.87 for the Machine Learning

Orbit ES-LI, GEO, HEO and SSO have the best success rate than the others GTO has lowest success rate

CCAFS-SLC 40 and KSC LC 39A are the most popular site than the other But they have low Pay Load Mass rate.

Success rate started to increase in 2013 and kept skyrocketing till 2020. 2018 and 2019 are also not bad!

Thank you