IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tamir Rozenfeld

Outline

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

Executive Summary

- Methodologies -
    - Data Collection via API
    - Web scrapping
    - Data Wrangling
    - Exploratory Data Analysis (EDA) with Data Visualization
    - EDA with SQL
    - Interactive map with Folium
    - Dashboards with Plotly Dash
    - Predictive Analysis using Machine Learning
    - A dashboard for Space X as a client
- Result -
    - SpaceX will launch on first stage successfully.

Introduction

- Project background and context
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars and other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

- Problems you want to find answers
- We need to determine the Falcon 9 rocket will land on first stage or not.
- If we could solve above problem then we can determine the cost of a launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data Collection using SpaceX API and using json normalize method
- Perform data wrangling
  - Deal with missing values and get the columns unique values and its count.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Import CSV data using IBM DB2 resource and extract the needful data for analysis.
- Perform interactive visual analytics using Folium and Plotly Dash
  - Display the data using Folium for Maps and Plotly for Dashboard.
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- we will define a series of helper functions that will help us use the API to extract information using identification numbers in the launch data.
- Requesting rocket launch data from SpaceX API with the following URL
- https://api.spacexdata.com/v4/launches/past
- we then decode the response content as a Json using .json() and turn it into a Pandas dataframe using .json_normalize()

Data Wrangling

- The number of launches on each site

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

- The occurrence of each orbit

```
GTO       27
ISS       21
VLEO      14
PO         9
LEO        7
SSO        5
MEO        3
ES-L1      1
HEO        1
SO         1
GEO        1
Name: Orbit, dtype: int64
```

- The occurence of mission outcome per orbit type

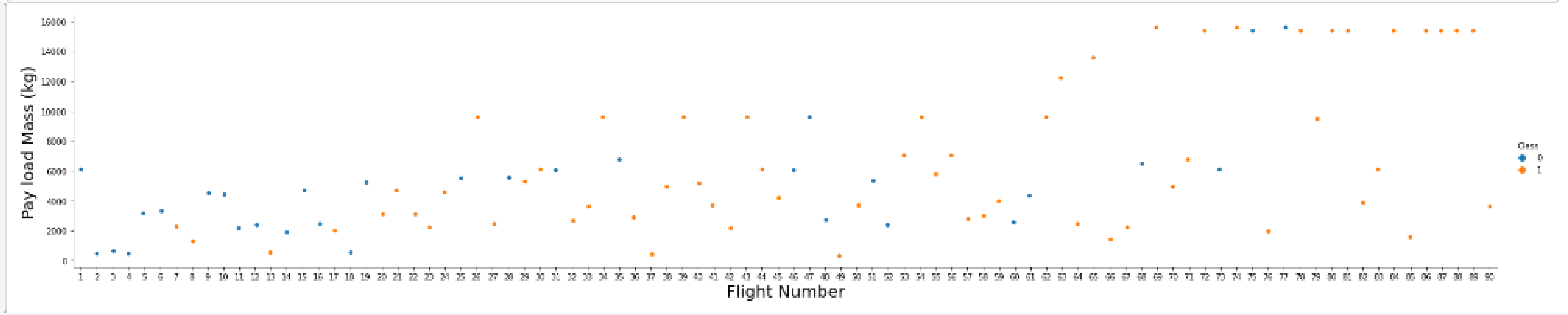
```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: Outcome, dtype: int64
```

- A landing outcomes, 1 means landed successfully.

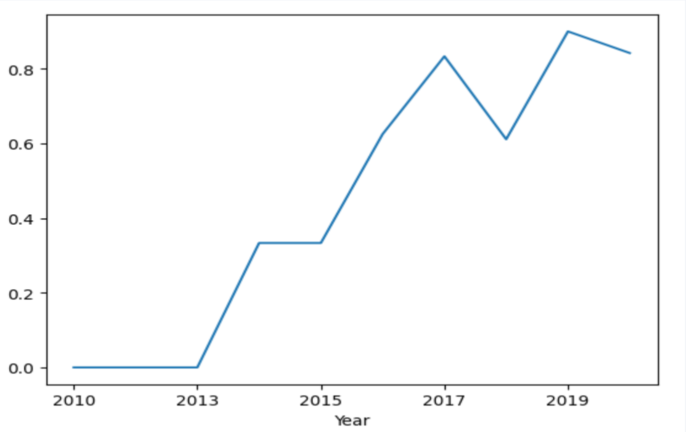
	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

- https://github.com/herambgaidhani/Winning-Space-Race-with-Data-Science/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization



We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E ha s a success rate of 77%.



The success rate since 2013 kept increasing till 2020.

https://github.com/herambgaidhani/Winning-Space-Race-with-Data-Science/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- The SQL queries performed -
- The unique launch sites names in the space mission
- The launch sites begin with the string 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- the date when the first successful landing outcome in ground pad was achieved.
- the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- the total number of successful mission outcomes is 99 and failure is 1

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

total_payload_mass_nasa_crs
45596

date_first_groundpad_success
2015-12-22

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

https://github.com/herambgaidhani/Winning-Space-Race-with-Data-Science/blob/main/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- Goal: find insights on best Launch Sites Locations using interactive maps;
- Launch sites are indicated by markers on a map;
- Mark success/failed launches for each site with marker clusters;
- Calculate distances (denoted by lines in the map) between a launch site to its proximities.

Build a Dashboard with Plotly Dash

- We created a Plotly Dash application to perform interactive visual analytics on SpaceX launch data in real-time.
- We used pie charts, rangeslider and scatter plots to visualize data.
- Pie charts for the percentage of successful launches by site, in order to determine the best launch site.
- Rangeslider allows to select a payload mass in a range.
- Scatter plots to study the relation between payloads and launch sites, in order to better understand the best launch sites according to payloads.

Predictive Analysis (Classification)

- We compared four different classification models in order to find the best model to predict if a launch is successful or not:
  - Logistic regression;
  - Support Vector Machine (SVM);
  - Decision tree;
  - K Nearest Neighbors (KNN).

Results

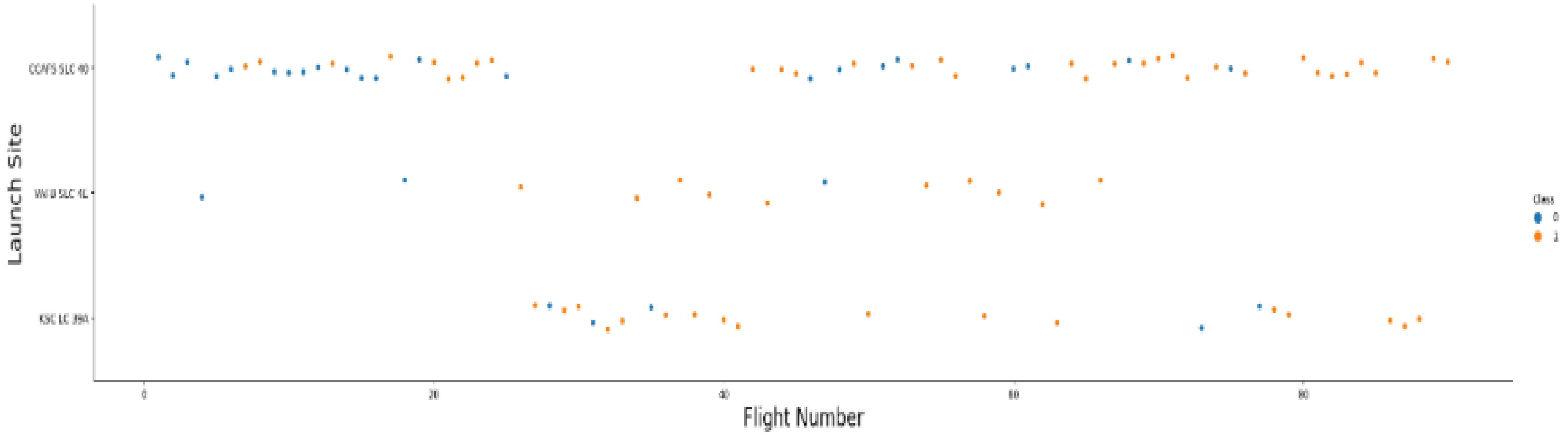
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

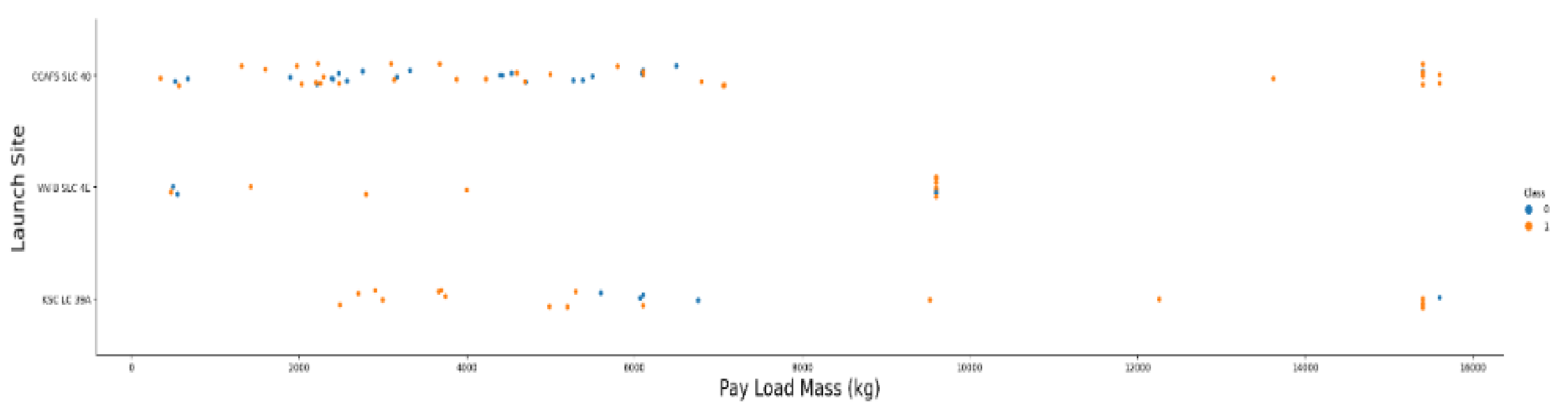
Flight Number vs. Launch Site

- Most of first launches were performed in CCAFS SLC 40.
- The success rate improves over time for every site.



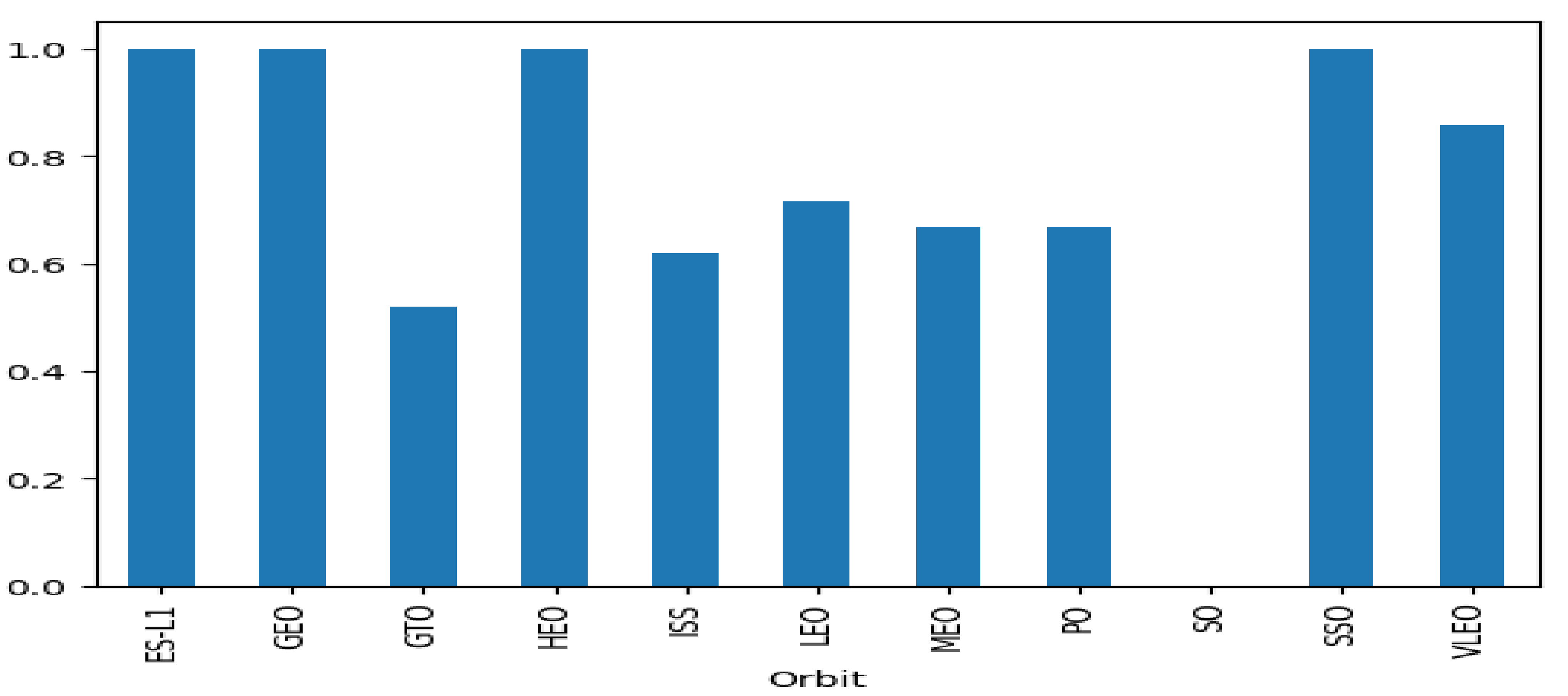
Payload vs. Launch Site

- KSC LC 39A has best success rate.
- Payloads with more than 8000kg have high success rate.



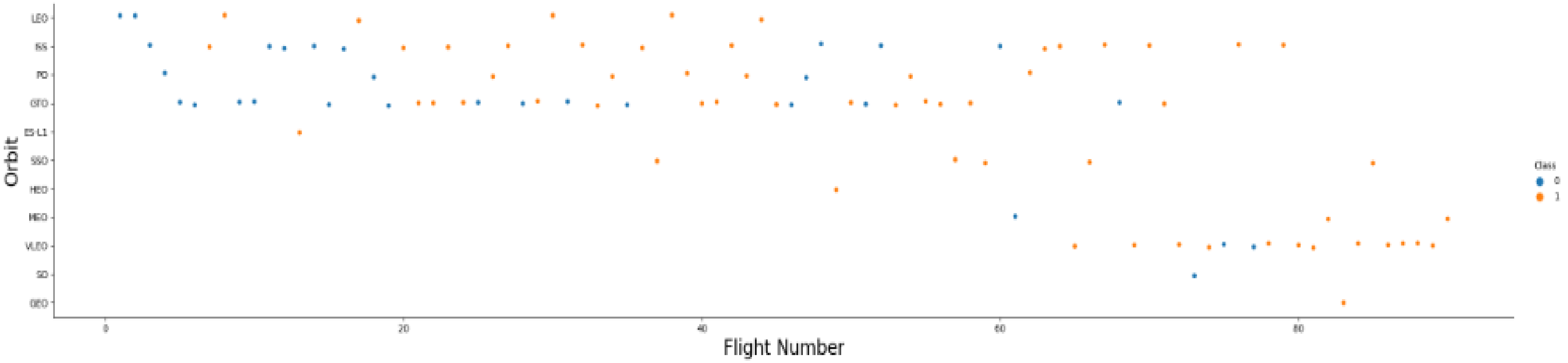
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have the highest success rate
- SO and GTO have the worst success rate



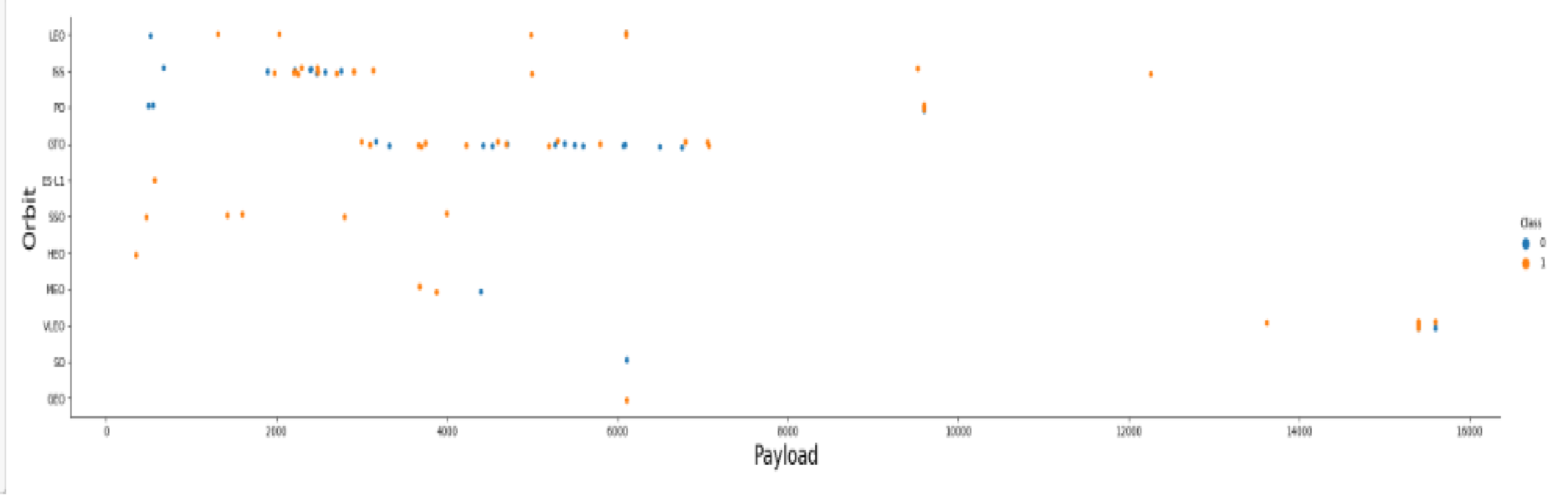
Flight Number vs. Orbit Type

- LEO: the success related to the number of flights
- GTO: No relation between flight number and success
- VLEO: Most of successful launches in the last period



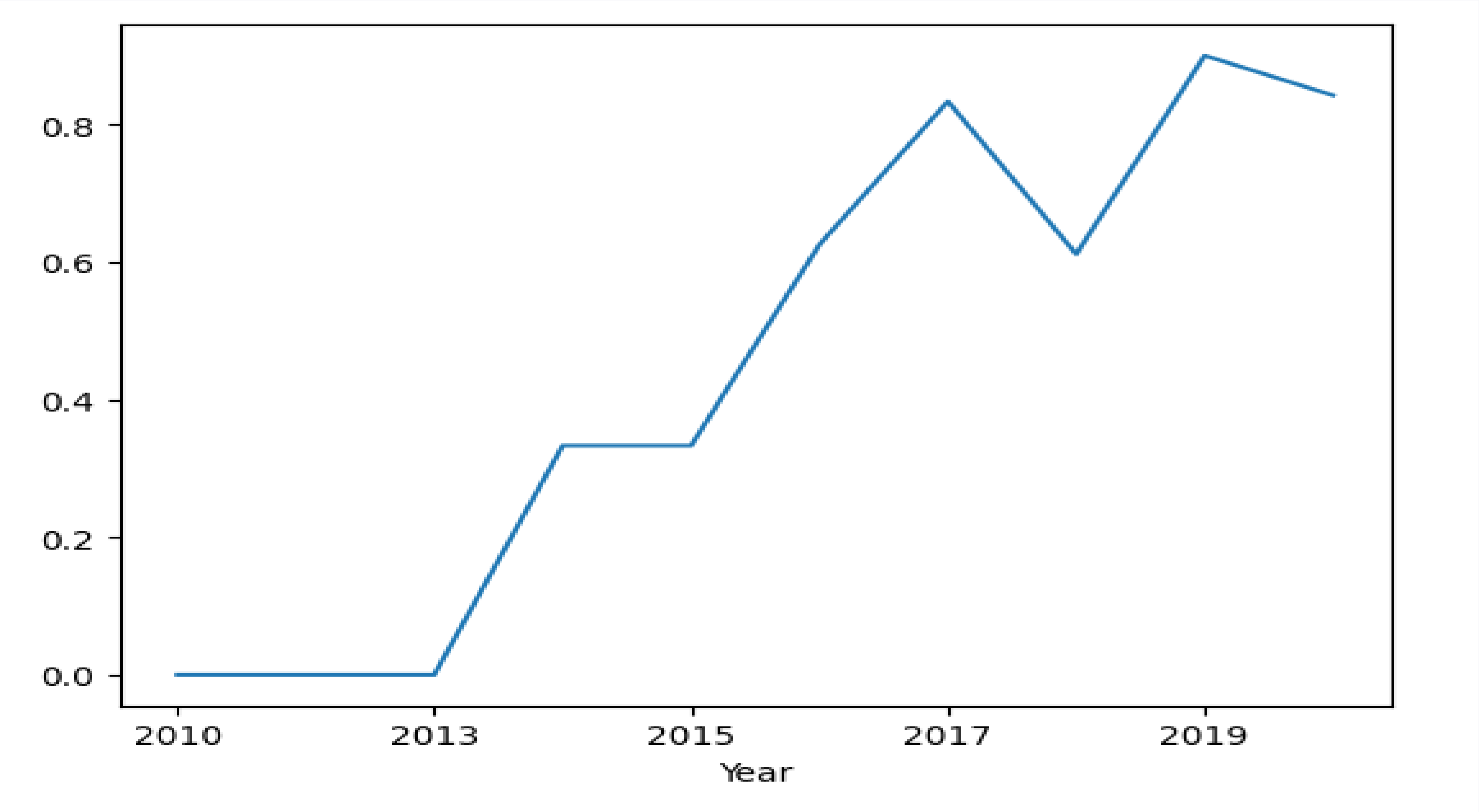
Payload vs. Orbit Type

- Launches over 8000kg are in VLEO, PO and ISSS



Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020
- Success rate Zero between 2010 and 2013



All Launch Site Names

- The names of the unique launch sites i.e. there are only 4 launch sites.

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

Launch Site Names Begin with 'CCA'

- First 5 records where launch sites begin with `CCA`

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

- Total payload mass (in kg) carried by boosters launched by NASA (CRS):

total_payload_mass_nasa_crs
45596

Average Payload Mass

- The average payload mass carried by booster version F9 v1.1

avg_payload_mass_f9v11
2928

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad -

date_first_groundpad_success
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful mission outcomes are 99 and failure mission outcomes are only 1

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

2015 Launch Records

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

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

landing__outcome	total
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Success (ground pad)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	1
Precluded (drone ship)	1

Section 3

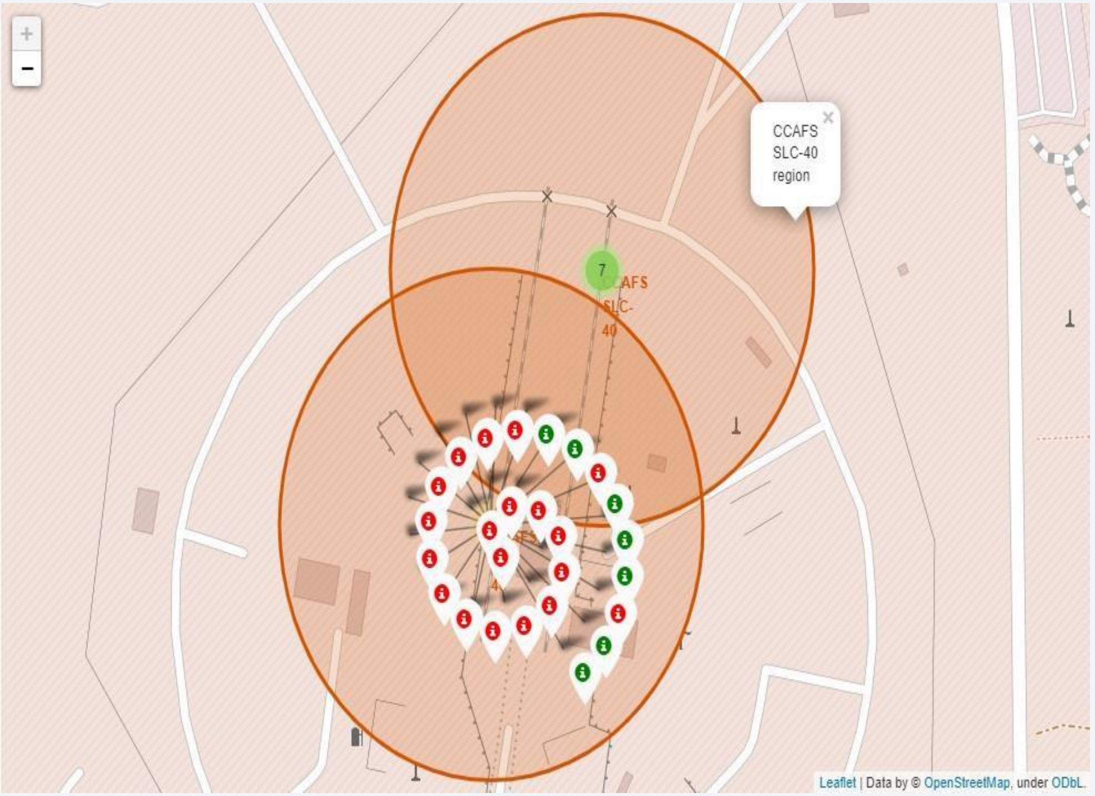
Launch Sites Proximities Analysis

Folium Map all launch sites

- It Includes all launch sites' location markers -

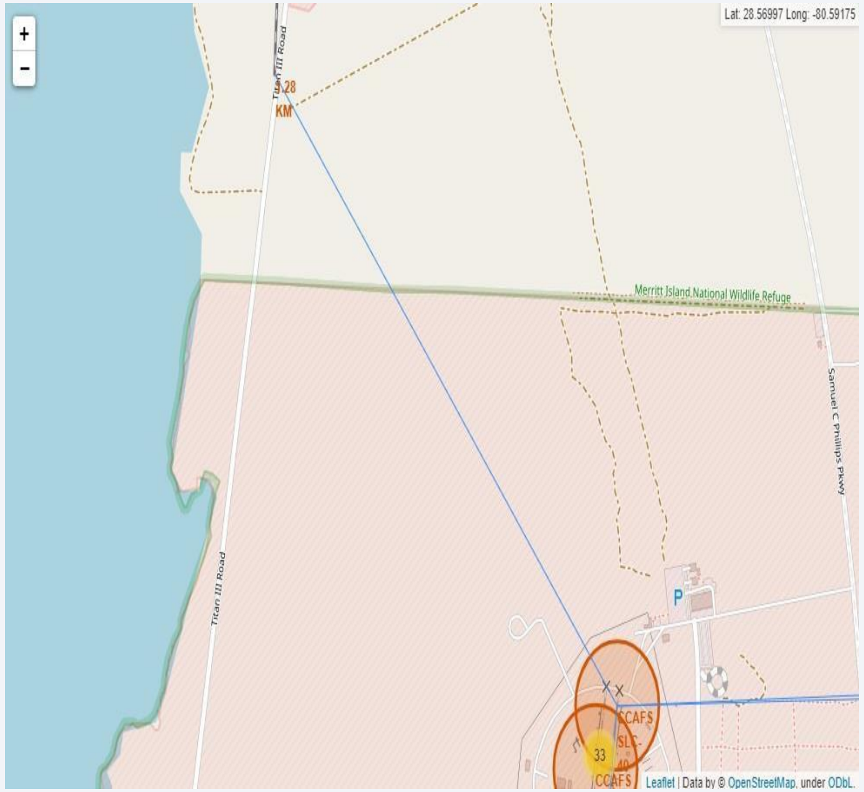
Launch Outcomes on the Map using Folium

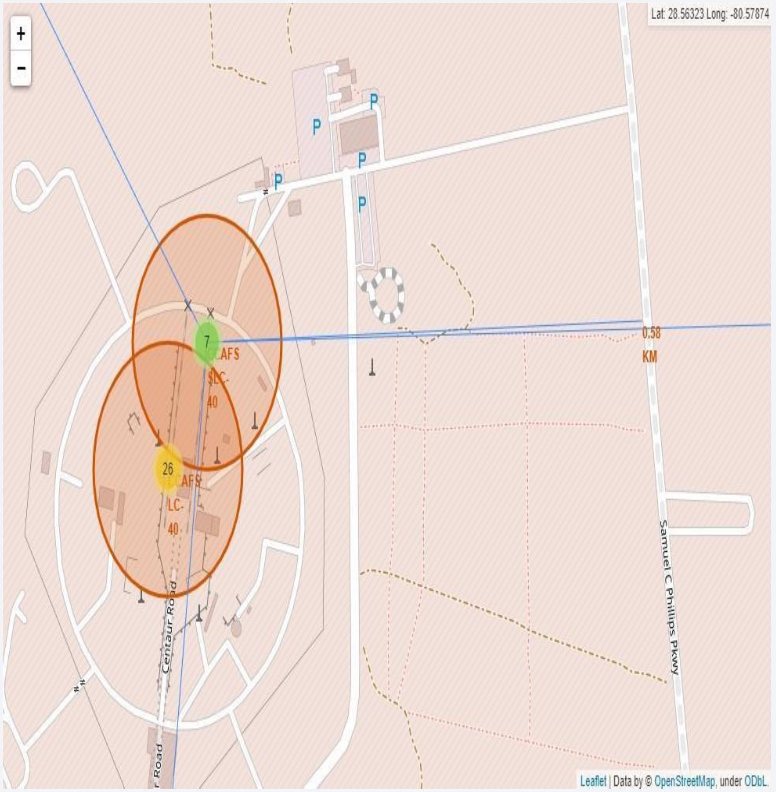
- Green markers indicate successful launches, red markers indicate failed launches

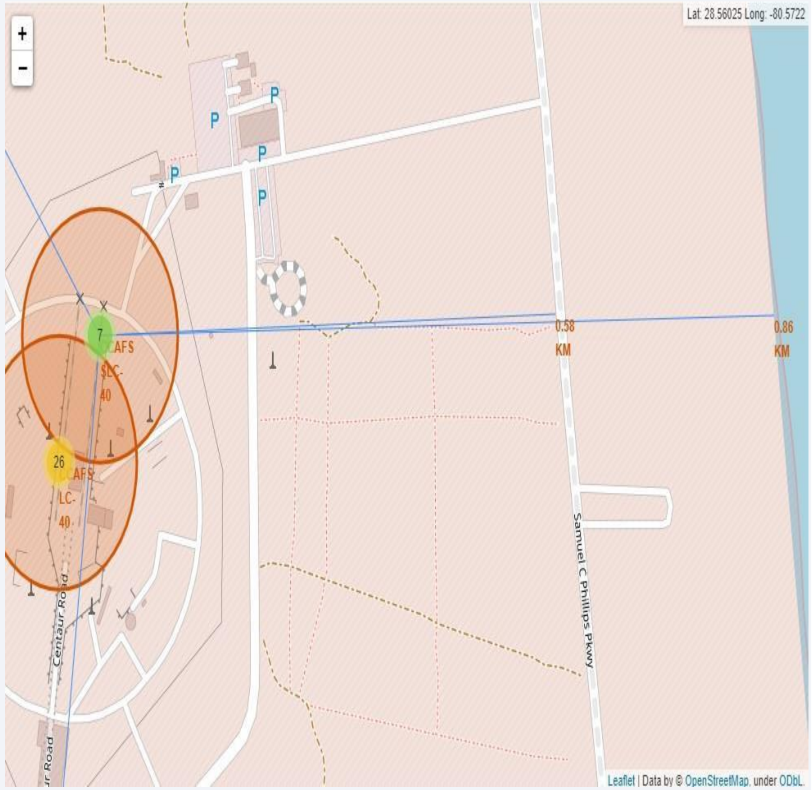


Launch Sites

- Railway, highway, coastline respectively





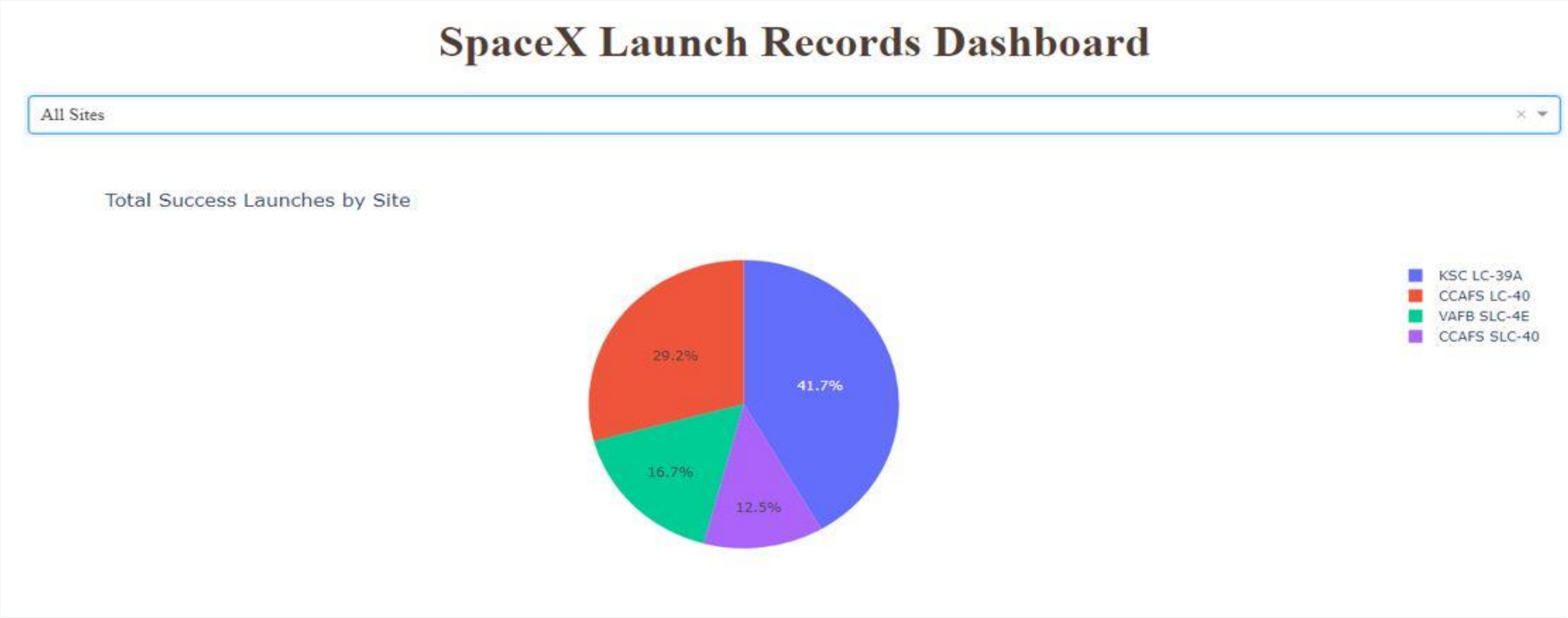


Section 4

Build a Dashboard with Plotly Dash

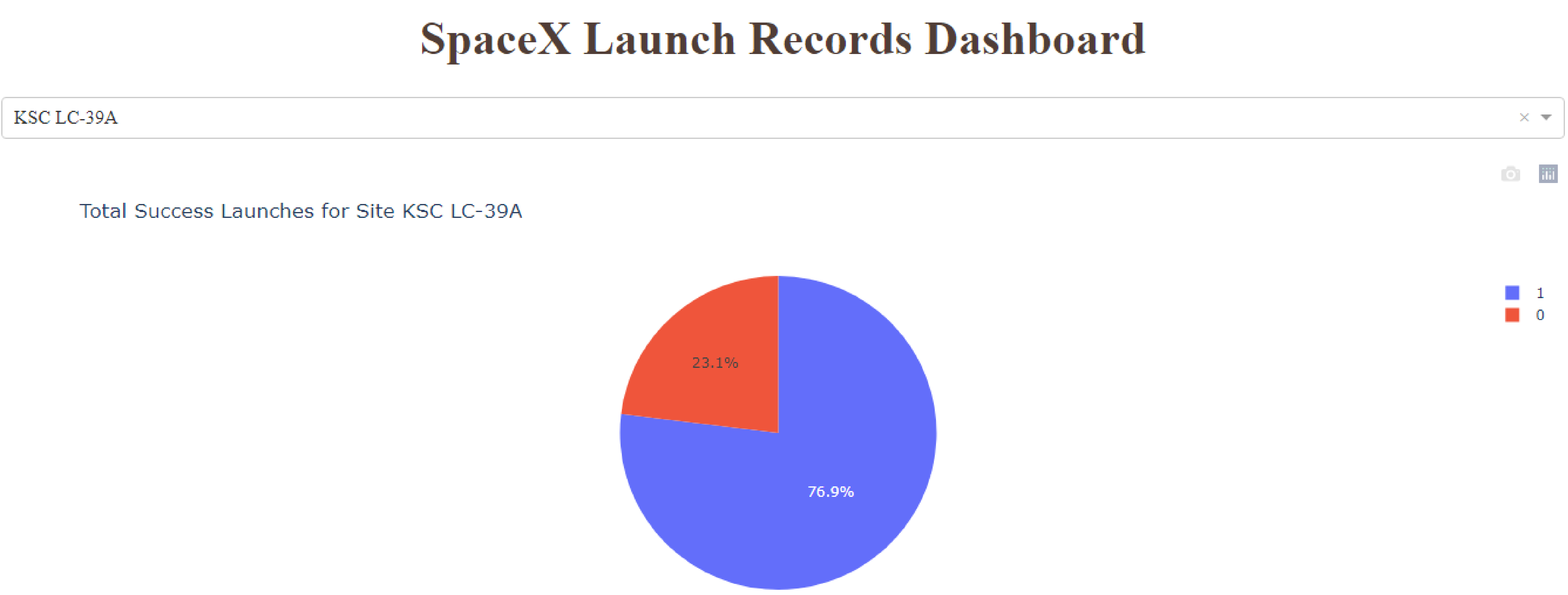
Dashboard of launch success count for all sites

- This pie-chart shows us the success rate for launches by site.



Total Success launches for site KSC LC–39A

- Site KSC LC–39A shows the most success rate.



Correlation between payload and success for all sites

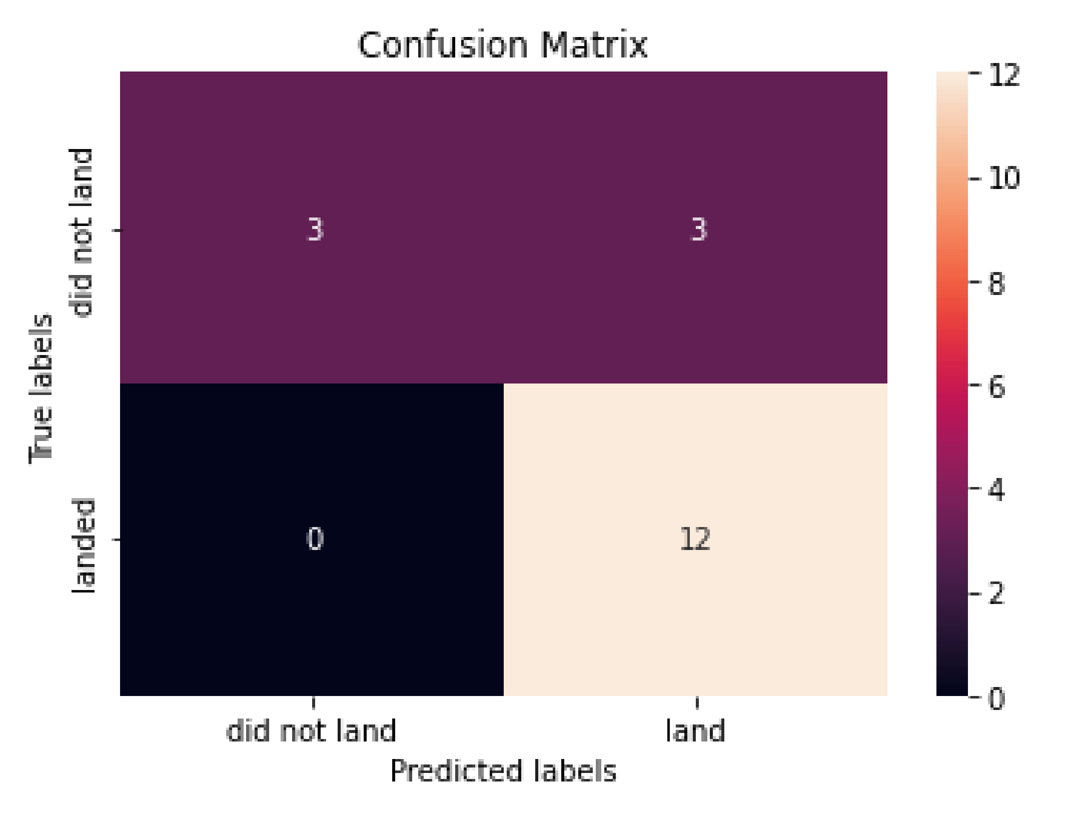
- Payload and success correlation shows that most are within 7000 KG payload range.

Section 5

Predictive Analysis (Classification)

Confusion Matrix

- The Decision Tree model performed the best.



Conclusions

- The best launch site is KSC LC-39A: it seems to perform with success both with light and heavy Payload Mass;
- The best orbits are GEO, HEO, SSO and ES-L1. It is worth taking into account VLEO, which has a lot of flights in the last period;
- Depending on the orbit, the Payload Mass can be a feature which influences
- the success of a mission;
- In general success rate increases with the number of flights and it started becoming higher from 2013, probably due to gain in knowledge and improvements in technologies;
- We decided to take the Decision Tree Model to predict the success of a
- mission: it has the best Accuracy on train data and the best test Accuracy.

Appendix

Some useful links that I have referred –

- https://docs.python.org/3/library/
- https://seaborn.pydata.org/
- https://pandas.pydata.org/docs/
- https://scikit-learn.org/stable/tutorial/index.html

Thank you!