IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Raul Gomez Rodriguez

2024-10-14

Outline

- Executive Summary
  - Summary of methodology
  - Summary of results
- Introduction
  - Project context
  - Exploration
- Methodology
  - Data collection method
  - Perform data Wrangling
  - Perform exploration of data analysis(EDA)
- Results
- Conclusion
- Appendix

Executive Summary

- The following methodologies were used to analyze data:
  - web scraping and SpaceX AP for the collection data
  - Exploratory Data Analysis (EDA)
  - public sources were used to collect data
  - The EDA serve us to identify the best to predict success of launchings;
  - Also Machine learning was used to predict the best model to use and the site landing with the best options,

Introduction

- We need to evaluate wich if a new company could get best results as SpaceX.
- Wich could be the best place
- The total cost of launches with a successful kanding result

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
  - Data from Space X was obtained from:
    - Space X API (https://api.spacexdata.com/v4/rockets/).
    - WebScarping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling

  new outcomes were created before analize to créate a new outcomes
- exploratory data analysis (EDA) using SQL and visualization
- Folium and Plotly Dash were used for visualization propouses
- Different classification models were used for predictive analysis

Data Collection

All the data were collected from:

Space X API:

https://api.spacexdata.com/v4/rockets/

and from Wikipedia:

https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches

And web scraping technics were used.

Data Collection – SpaceX API

1 - Request API from SpaceX

2 - Data was filtered to include the launches of falcon 9

3 - Missing values are treated as well

The results are disponible at:

https://github.com/ra13ul/Applied_Data_Science_Capstone/blob/main/05%20-%20jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

Request data from wikipedia page

Extract all the columns and values

Create a new data frame with the obtained values

The results are disponible at:

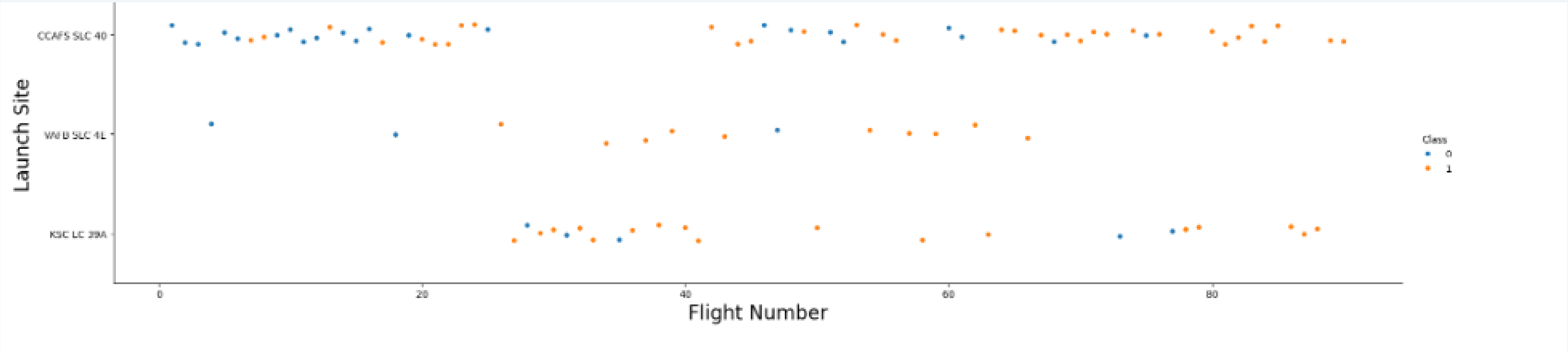
https://github.com/ra13ul/Applied_Data_Science_Capstone/blob/main/06%20-%20jupyter-labs-webscraping.ipynb

Data Wrangling

- Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.

EDA with Data Visualization

- Visualize the relationship between Flight Number and Launch Site
- Visualize the relationship between Payload Mass and Launch Site
- Visualize the relationship between success rate of each orbit type
- Visualize the relationship between FlightNumber and Orbit type
- Visualize the relationship between Payload Mass and Orbit type ….



For full datails: https://github.com/ra13ul/Applied_Data_Science_Capstone/blob/main/14%20-%20jupyter-labs-eda-dataviz.ipynb

EDA with SQL

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

For all datails:
https://github.com/ra13ul/Applied_Data_Science_Capstone/blob/main/12%20-%20jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities
- Draw a line between the marker to the launch site

Here’s the information:
https://github.com/ra13ul/Applied_Data_Science_Capstone/blob/main/17%20-%20lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- We use graphs and plots to visualize data:
  - Percentage of launches by site
  - Payload range

The app:
https://github.com/ra13ul/Applied_Data_Science_Capstone/blob/main/18%20-%20spacex_dash_app.py

Predictive Analysis (Classification)

- logistic regression, support vector machine, decision tree and k nearest neighbors were the models ued.

- The steps:
  - Data preparation
  - Combinations of Hyperparameters for the models
  - Results compared.

Follow this link for all the proccess
https://github.com/ra13ul/Applied_Data_Science_Capstone/blob/main/20%20-%20SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- Results for Exploratory data analysis
  - We detect four different launch sites
  - F9 v1.1 booster has and average payload = 2,928 kg
  - The first success landing outcome happened in 2015 was the first success landing outcome
  - Falcon 9 booster in several versions were successful landing in drone ships with a payload lower,
  - Almost 100% of mission outcomes were successful
  - As years passed the number of landing outcomes is higher.

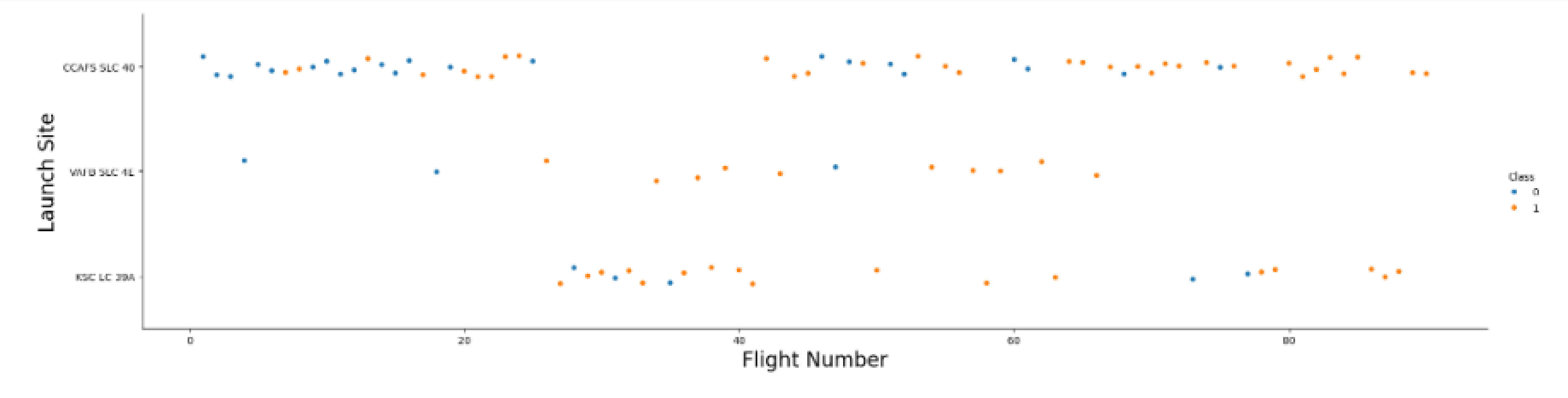
Results

- Decision Tree Classifier is the best model to predict successful landings,
- accuracy of 87%
- test data accuracy 94%.

Section 2

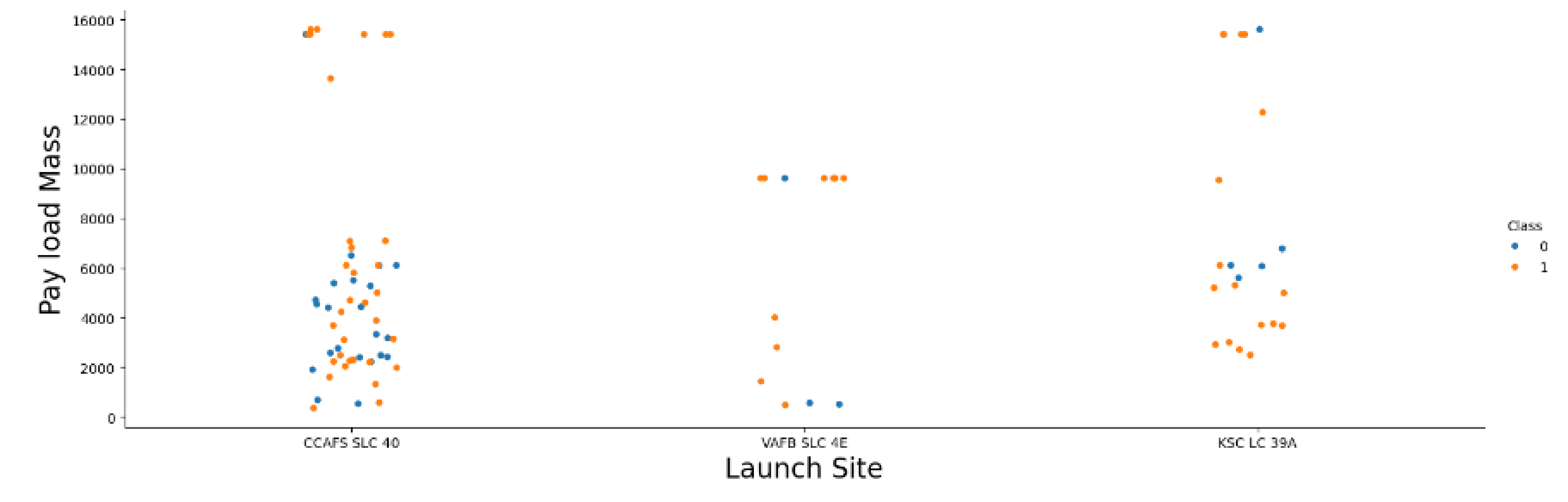
Insights drawn from EDA

Flight Number vs. Launch Site



- best launch site is CCAF5 SLC 40, followed by VAFB SLC 4E and KSC LC 39A
- success rate increase with the time.

Payload vs. Launch Site

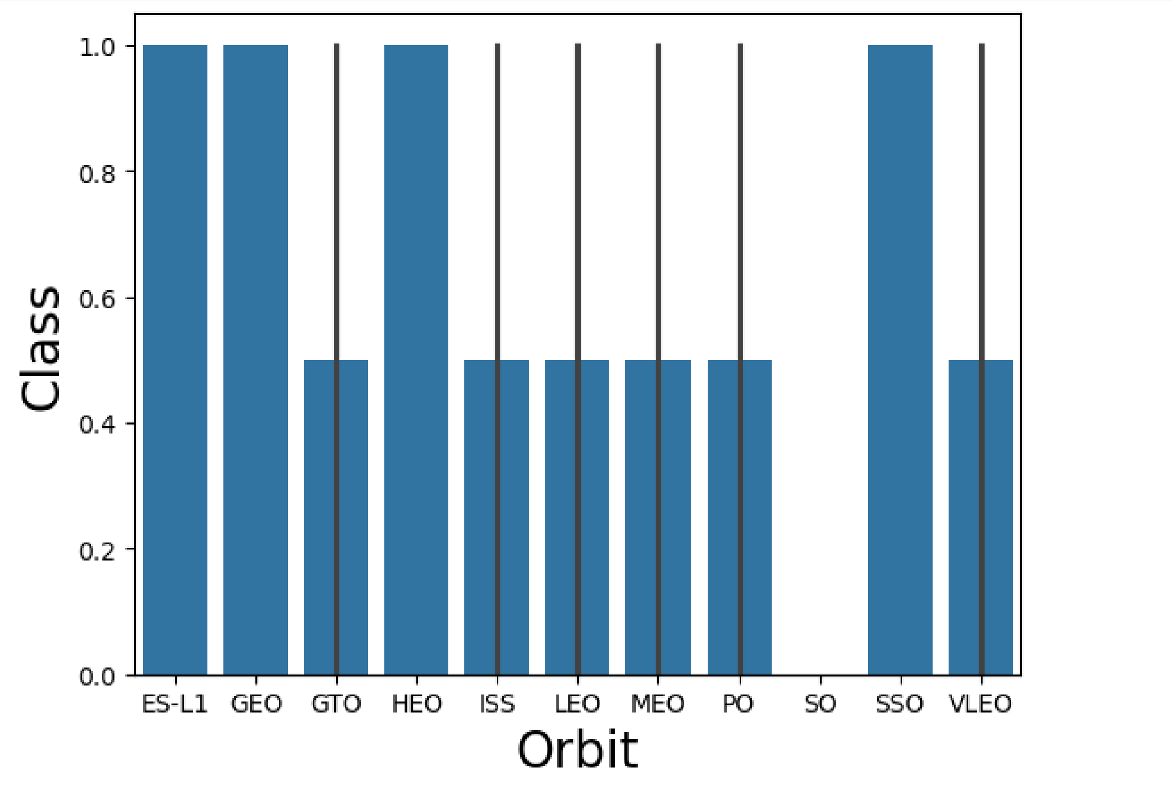


- Payloads over 9,000kg have very good success rate in mostly places;
- Payloads over 12,000kg only take place in CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

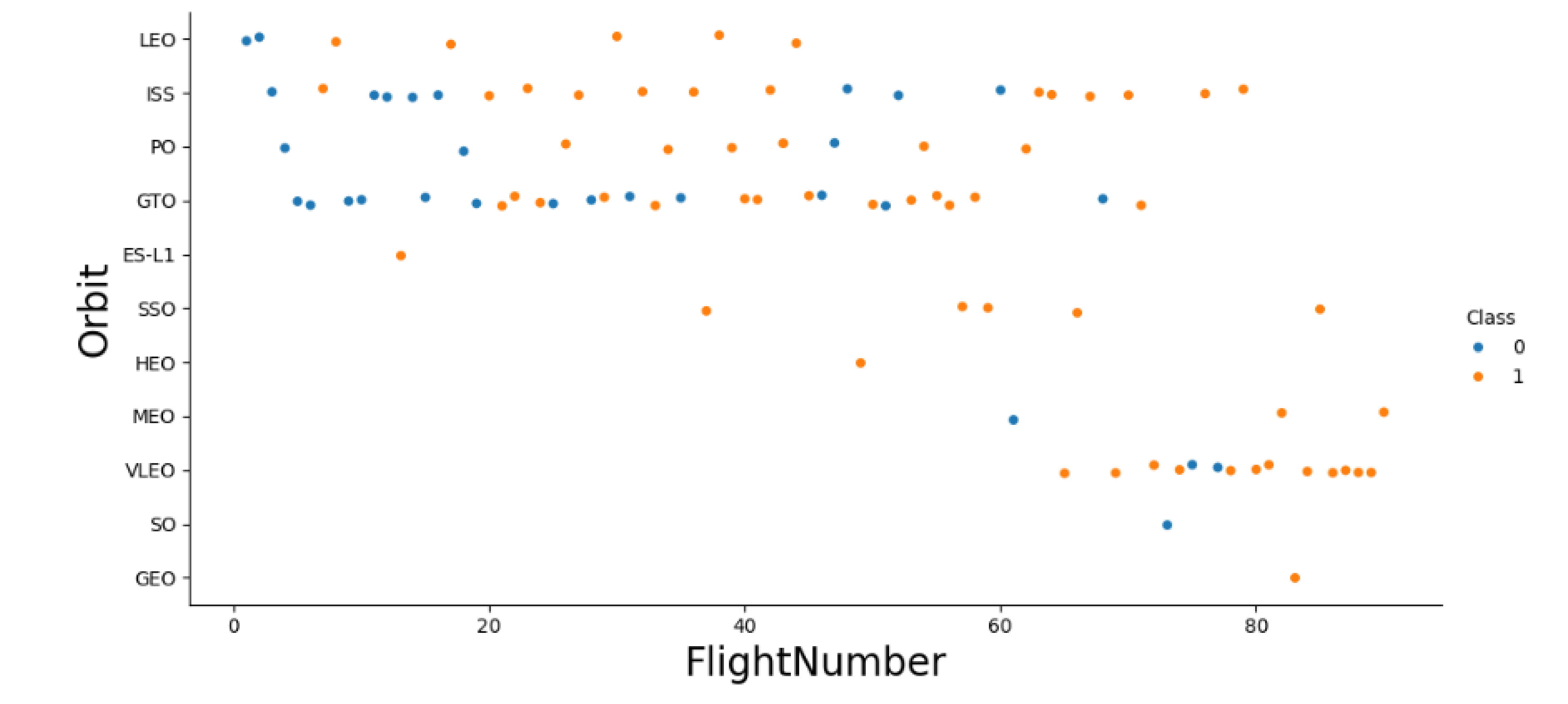
- The better results are in the orbits:

  ES-L1
  GEO
  HEO
  SSO.



Flight Number vs. Orbit Type

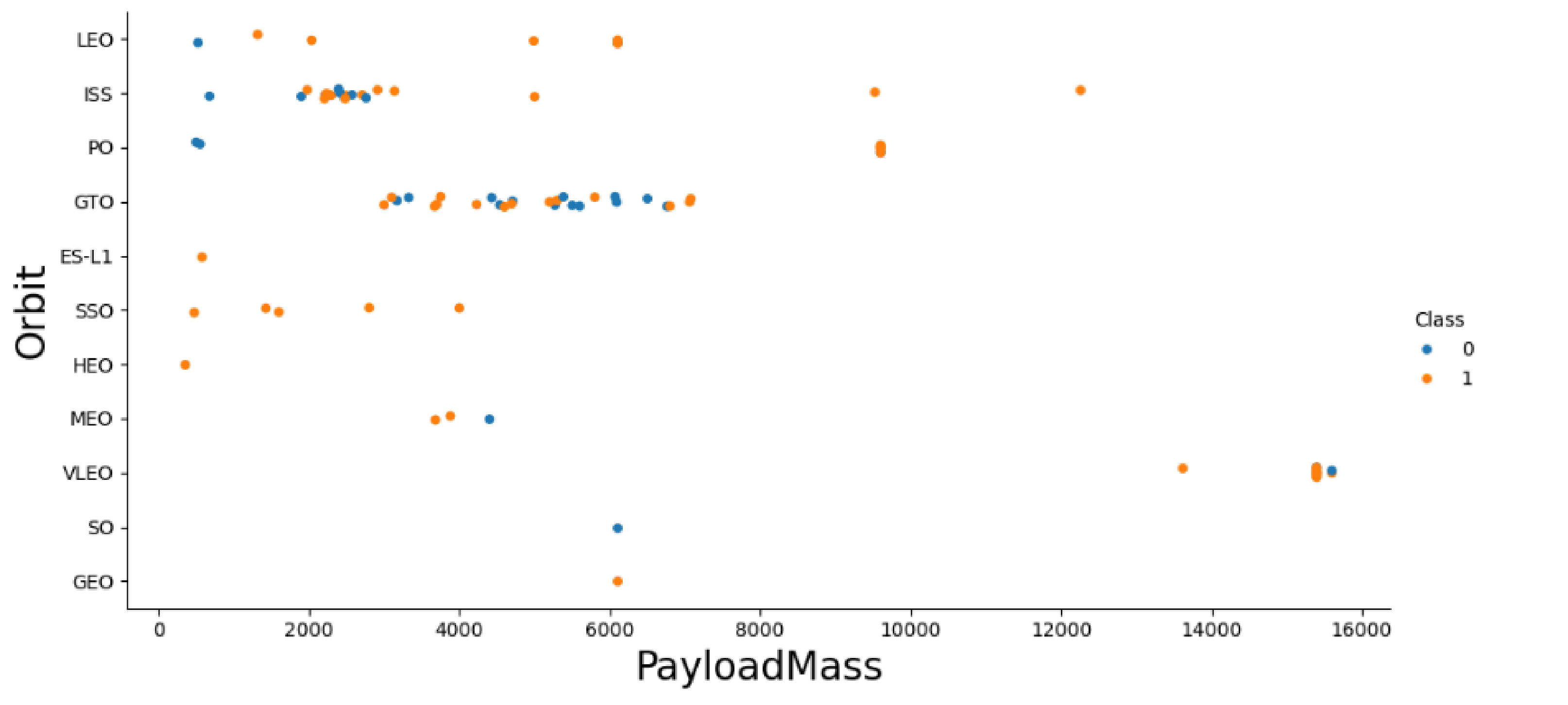
Generally all orbits has betters results as the flights numbers increase



Payload vs. Orbit Type

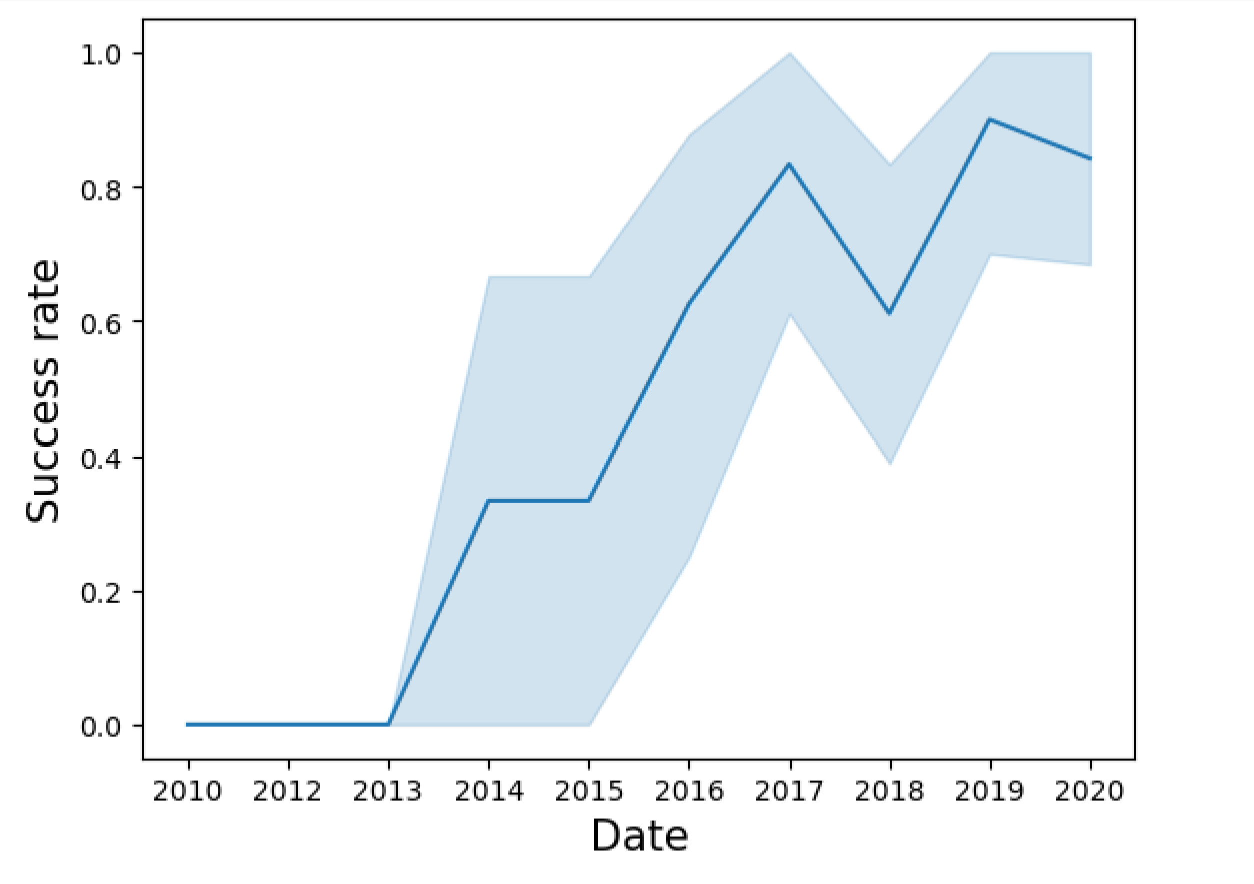
GTO could be most used with payloads between 3500 and 6500

SO and GEO orbits are the fewer used



Launch Success Yearly Trend

Success rate increasing in 2013 until 2020 only in 2018 and 2019 have been descent,



All Launch Site Names

There 4 launch sites:

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

Launch Site Names Begin with 'CCA'

Task 2

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

In [14]:
```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

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

Out[14]:

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

Finding where launch sites begin with \`CCA\`, 5 sites

Total Payload Mass

- Total payload boosters from NASA (CRS):

Task 3

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

In [15]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS "total Payload (Kg)" FROM SPACEXTBL WHERE Customer LIKE 'NASA (CRS)';
```

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

Out[15]:

total Payload (Kg)
45596

Average Payload Mass by F9 v1.1

Task 4

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

```
In [16]: %sql SELECT AVG(PAYLOAD_MASS__KG_) AS "total Payload (Kg)", Booster_Version FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1
```

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

Out[16]:

total Payload (Kg)	Booster_Version
2928.4	F9 v1.1

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

In [17]:

```
%sql SELECT MIN(DATE) as " first succesful landing outcome in ground pad"  FROM SPACEXTBL WHERE Landing_Outcome = 'Success
```

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

Out[17]:

first succesful landing outcome in ground pad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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, Customer, PAYLOAD_MASS__KG_,  Landing_Outcome \
    FROM SPACEXTBL \
    WHERE (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000) AND Landing_Outcome = 'Success (drone ship)';
```

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

Booster_Version	Customer	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	SKY Perfect JSAT Group	4696	Success (drone ship)
F9 FT B1026	SKY Perfect JSAT Group	4600	Success (drone ship)
F9 FT B1021.2	SES	5300	Success (drone ship)
F9 FT B1031.2	SES EchoStar	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

In [19]:

```
%sql SELECT Mission_Outcome, COUNT(Mission_Outcome) AS 'TOTAL MISSION OUTCOMES' FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

Out[19]:

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

Boosters Carried Maximum Payload

Customer	Date	Booster_Version	PAYLOAD_MASS__KG_
SpaceX	2019-11-11	F9 B5 B1048.4	15600
SpaceX	2020-01-07	F9 B5 B1049.4	15600
SpaceX	2020-01-29	F9 B5 B1051.3	15600
SpaceX	2020-02-17	F9 B5 B1056.4	15600
SpaceX	2020-03-18	F9 B5 B1048.5	15600
SpaceX	2020-04-22	F9 B5 B1051.4	15600
SpaceX, Planet Labs	2020-06-04	F9 B5 B1049.5	15600
SpaceX	2020-09-03	F9 B5 B1060.2	15600
SpaceX	2020-10-06	F9 B5 B1058.3	15600
SpaceX	2020-10-18	F9 B5 B1051.6	15600
SpaceX	2020-10-24	F9 B5 B1060.3	15600
SpaceX	2020-11-25	F9 B5 B1049.7	15600

2015 Launch Records

Date(Month)	Year	Landing_Outcome	Booster_Version	Launch_Site
01	1	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	1	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

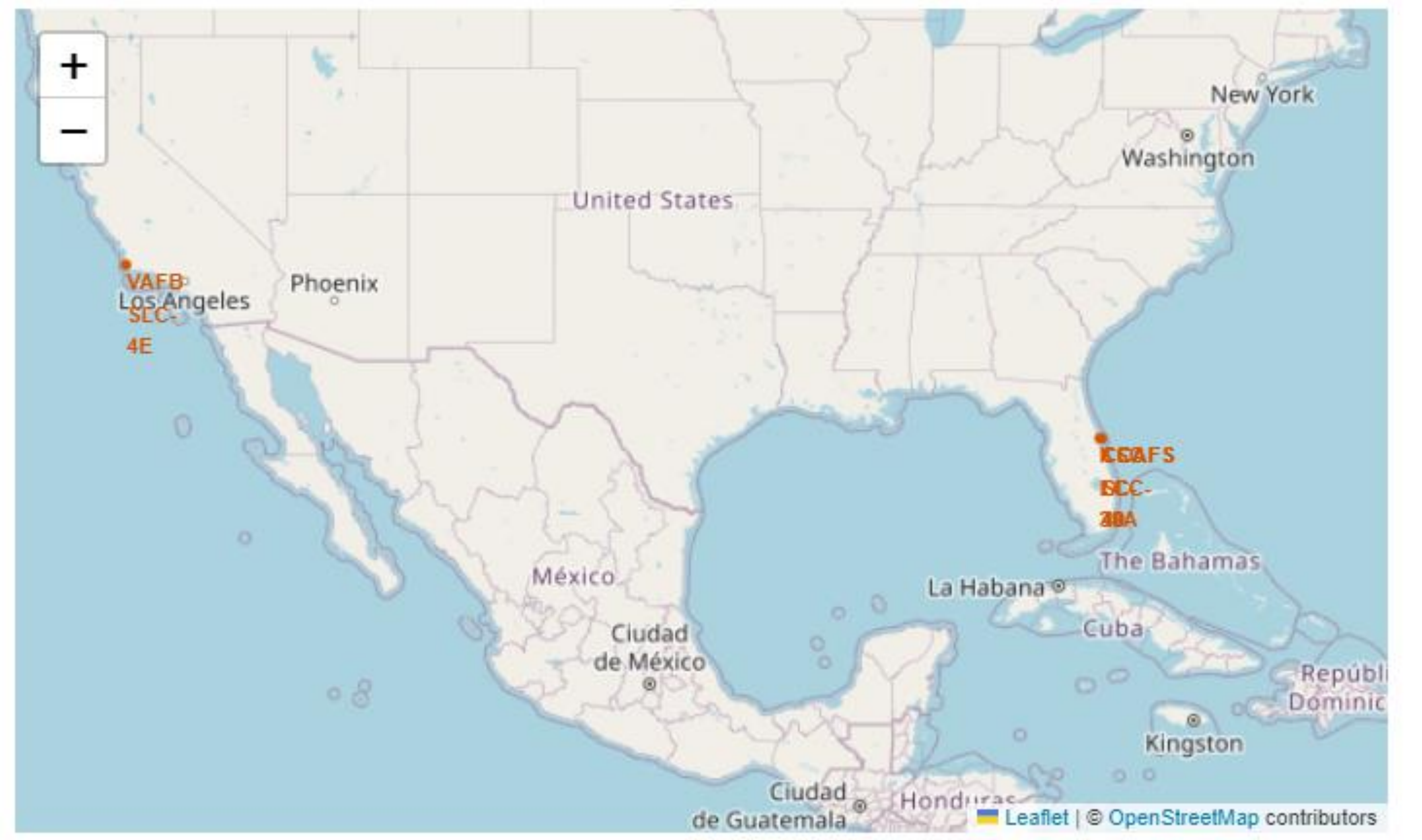
Total outcome	Landing_Outcome	Booster_Version	Launch_Site
3	Success (ground pad)	F9 FT B1019	CCAFS LC-40
5	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40

Section 3

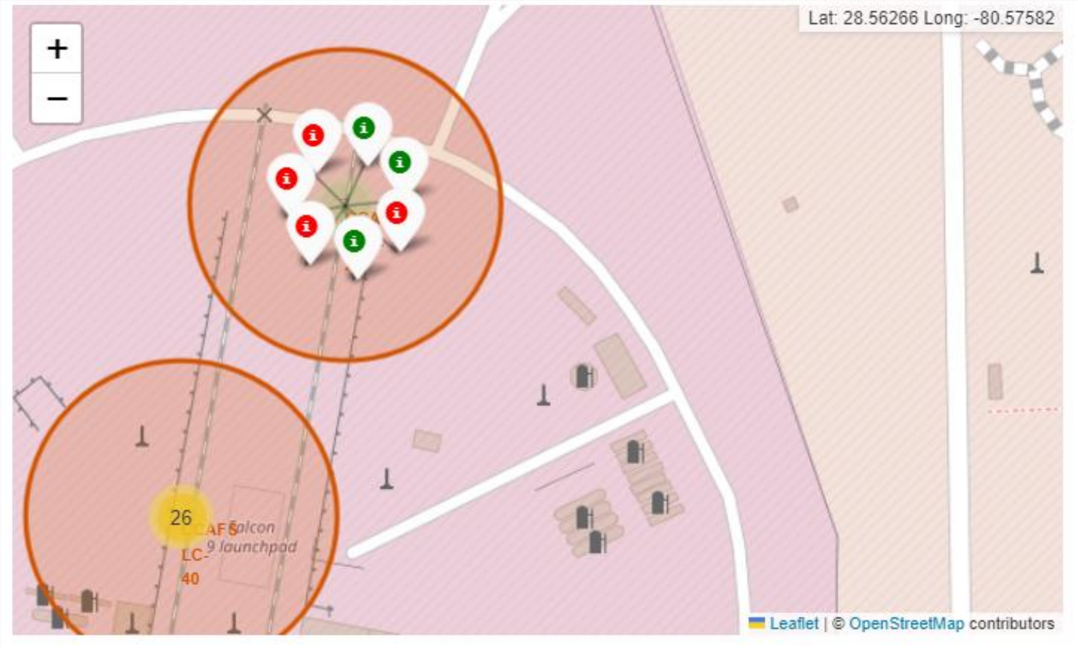
Launch Sites Proximities Analysis

Launches sites

Launch sites are close the sea, railways and motorways



Launch outcomes by site

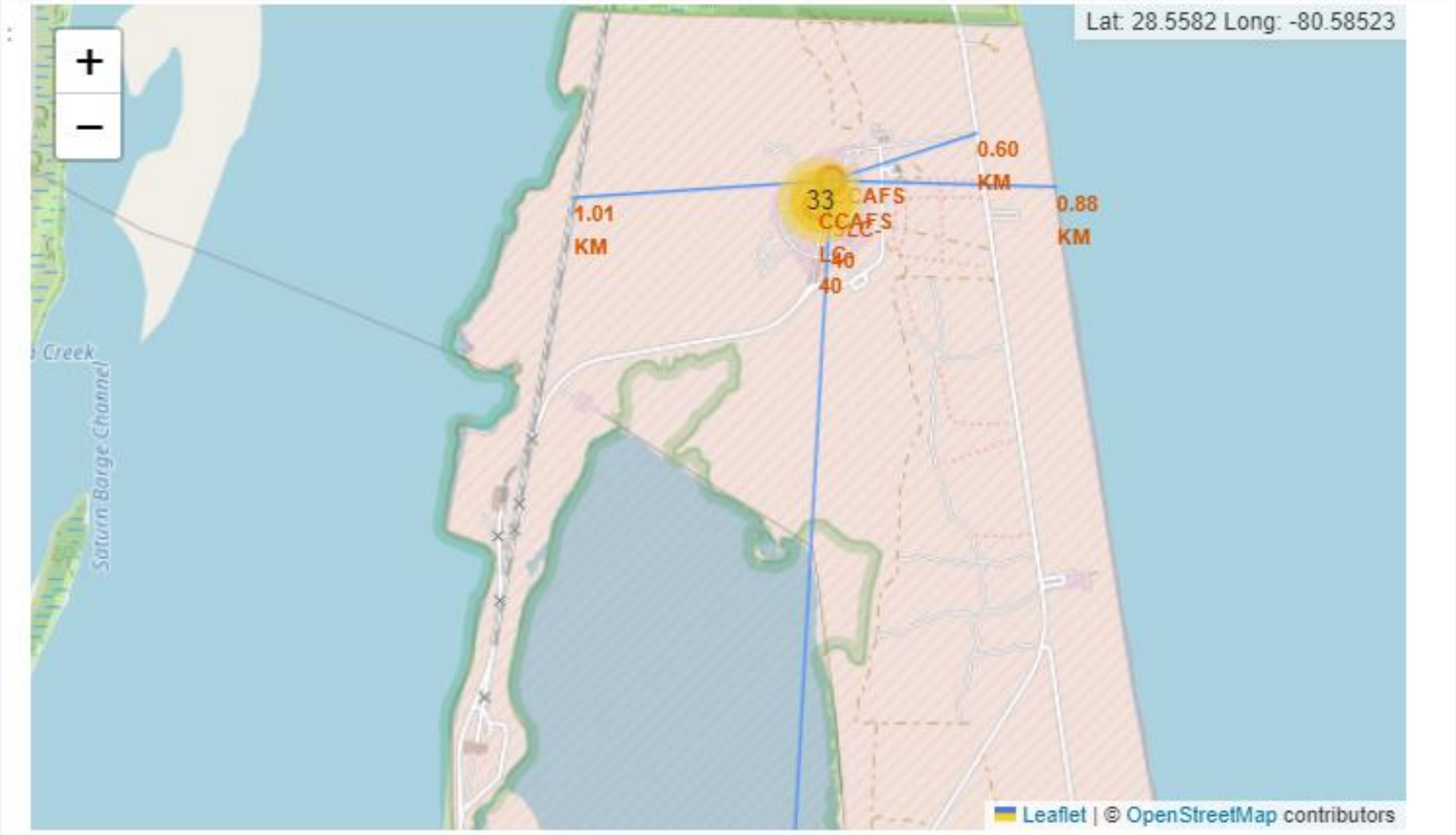


Launch site outcome CCAFS SLC-40 (top)

The green markers are successful and red one failure.

Logistics and safety

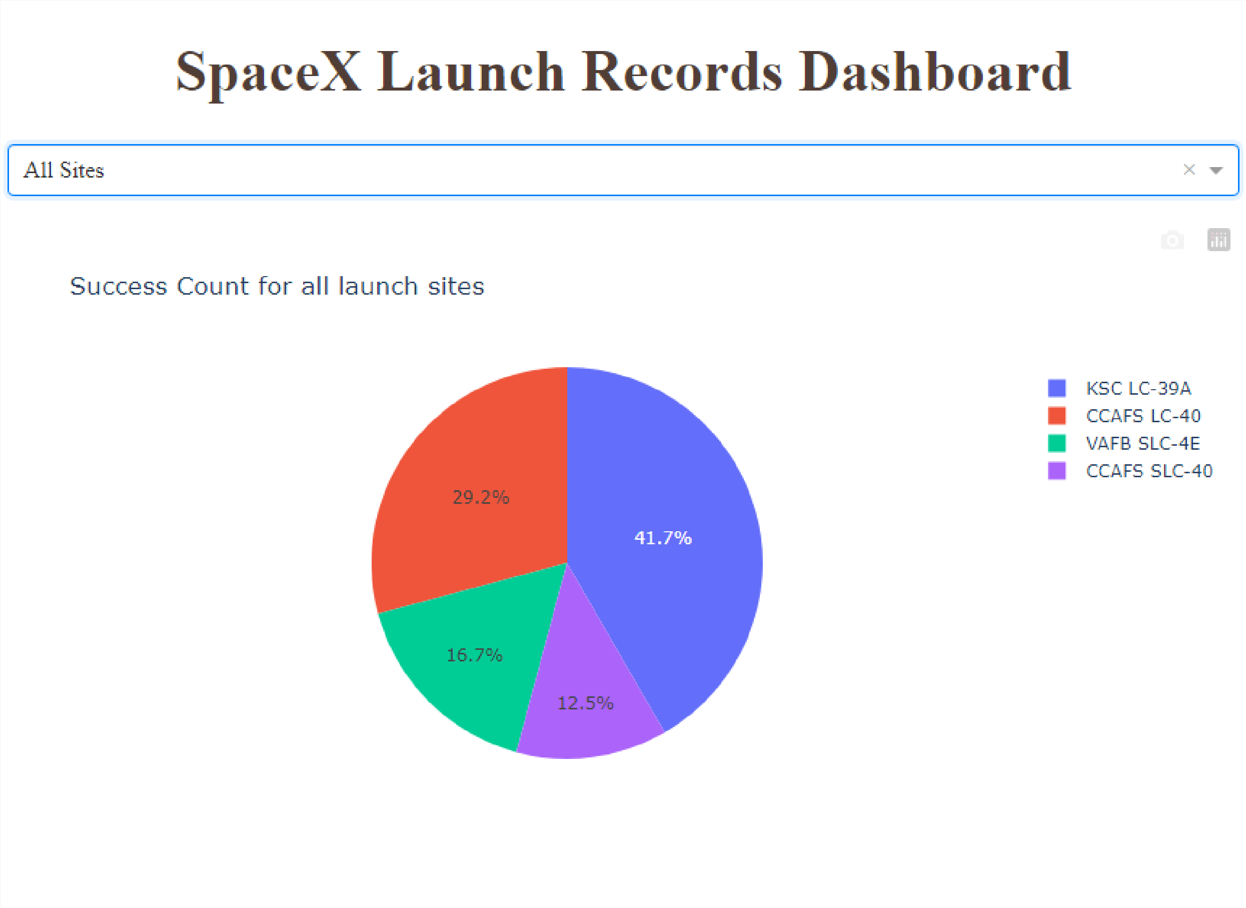
CCAFS SLC-40 has good logistics aspects, being near railroad and road and not so far from a city.



Section 4

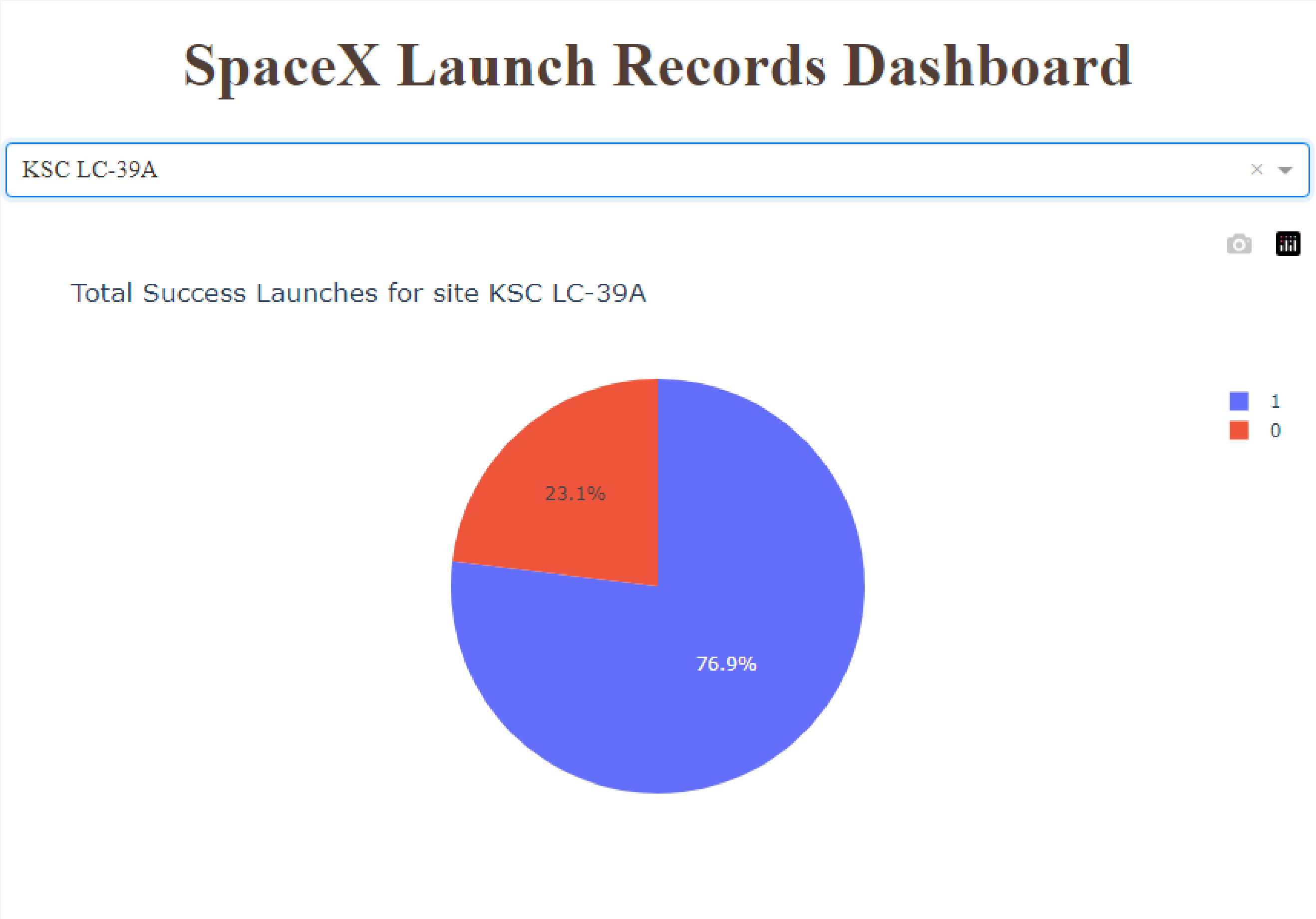
Build a Dashboard with Plotly Dash

Successful Launches Sites

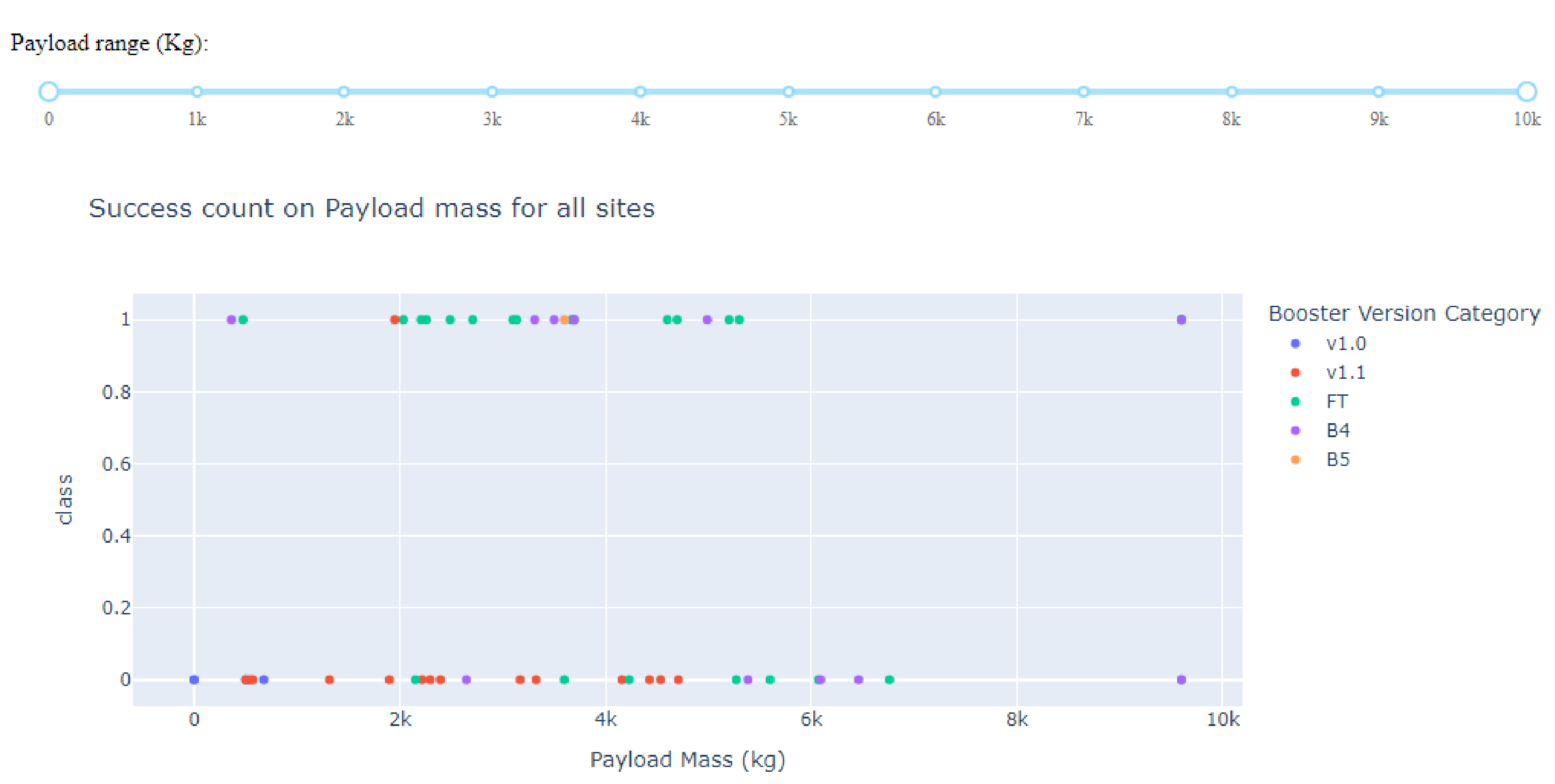


Lauch success ratio for KSC LC-39

76.9% SUCCESS (BLUE) of launches are successful in this site.



Payload versus Launch Outcome



Section 5

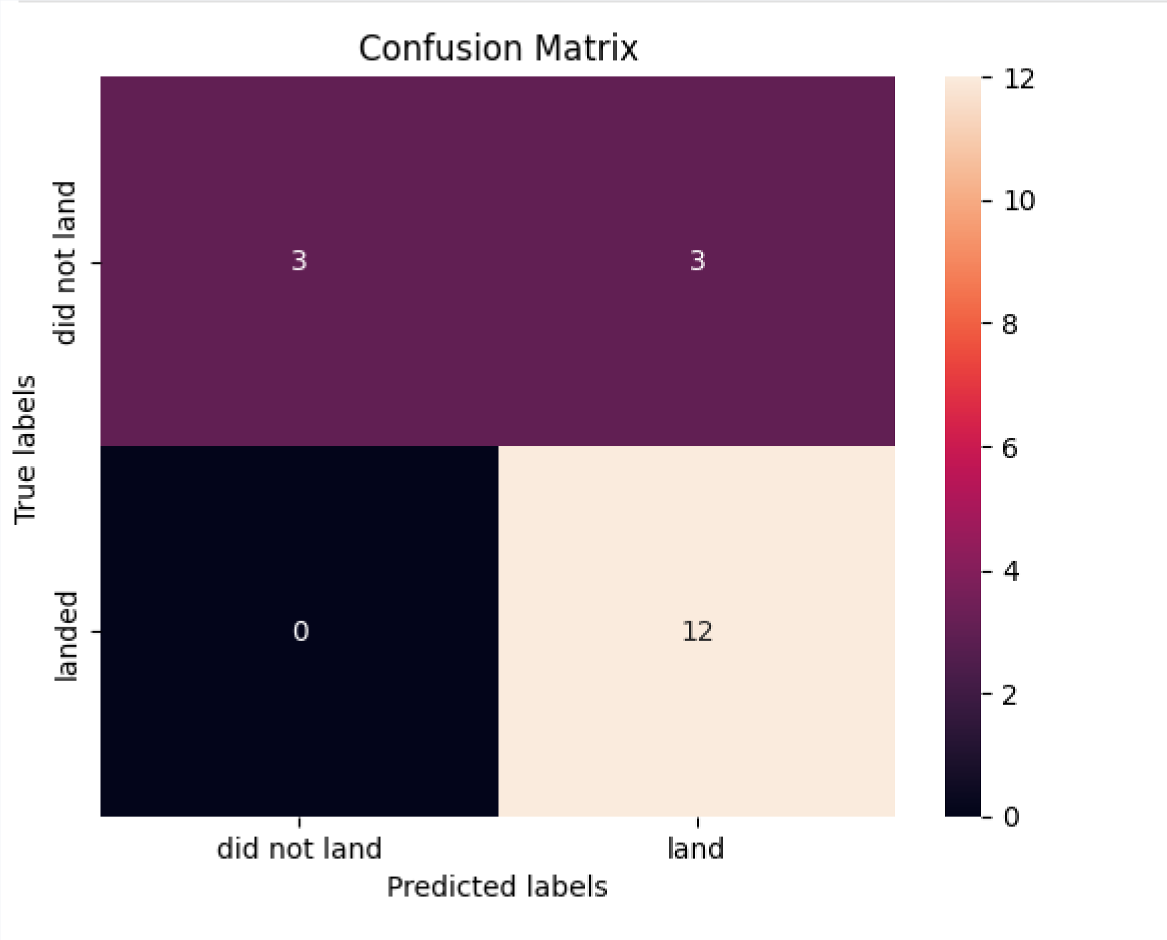
Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested, logReg, SVM, decision tree and knn
- The model with the highest classification accuracy is Decision Tree Classifier has accuracies of the 87%. Is the best model,

Confusion Matrix

Confusion matrix of Decision Tree Classifier has best true positive and true negatives.



Conclusions

The best launch site seem tobe KSC LC-39A

Launches above 7000 kg have the best results

To predict successful landing, best model could be Decision Tree Classifier

Appendix

- All the notebook, dashboard … and relevant datasets are available in:
- https://github.com/ra13ul/Applied_Data_Science_Capstone/tree/main

Thank you!