IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Zhicheng Wang>
<12.16.2021>

Outline

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

Introduction

- Project background and context
- 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.
- Problems you want to find answers
- So if we can determine if the first stage will land, we can determine the cost of a launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - By making a request to the SpaceX API
- Perform data wrangling
  - By performing some (EDA) (more details in later stage) to find the patterns of the date
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

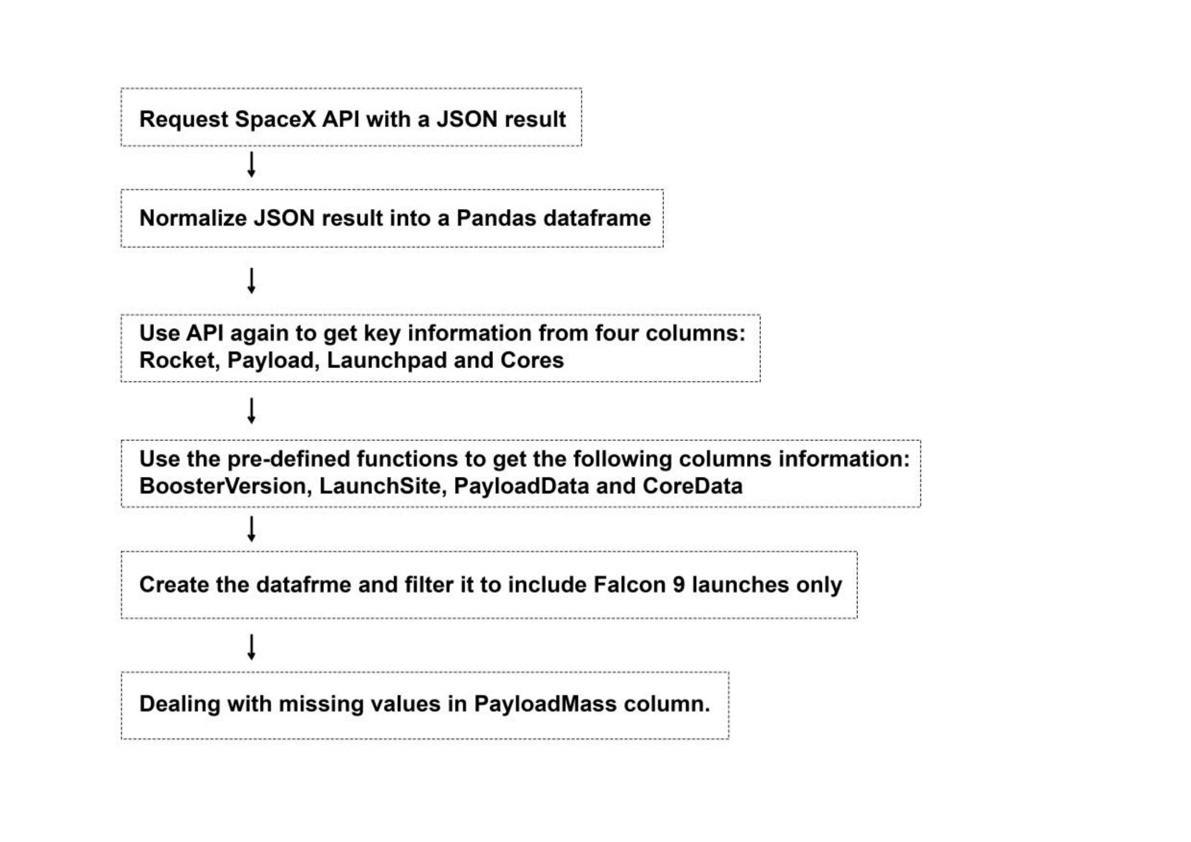
Data Collection

- The data was collected by using API request from SpaceX public API and by web scraping from SpaceX Wikipedia page.
- The following pages will show the data collection process.

Data Collection – SpaceX API

Github:

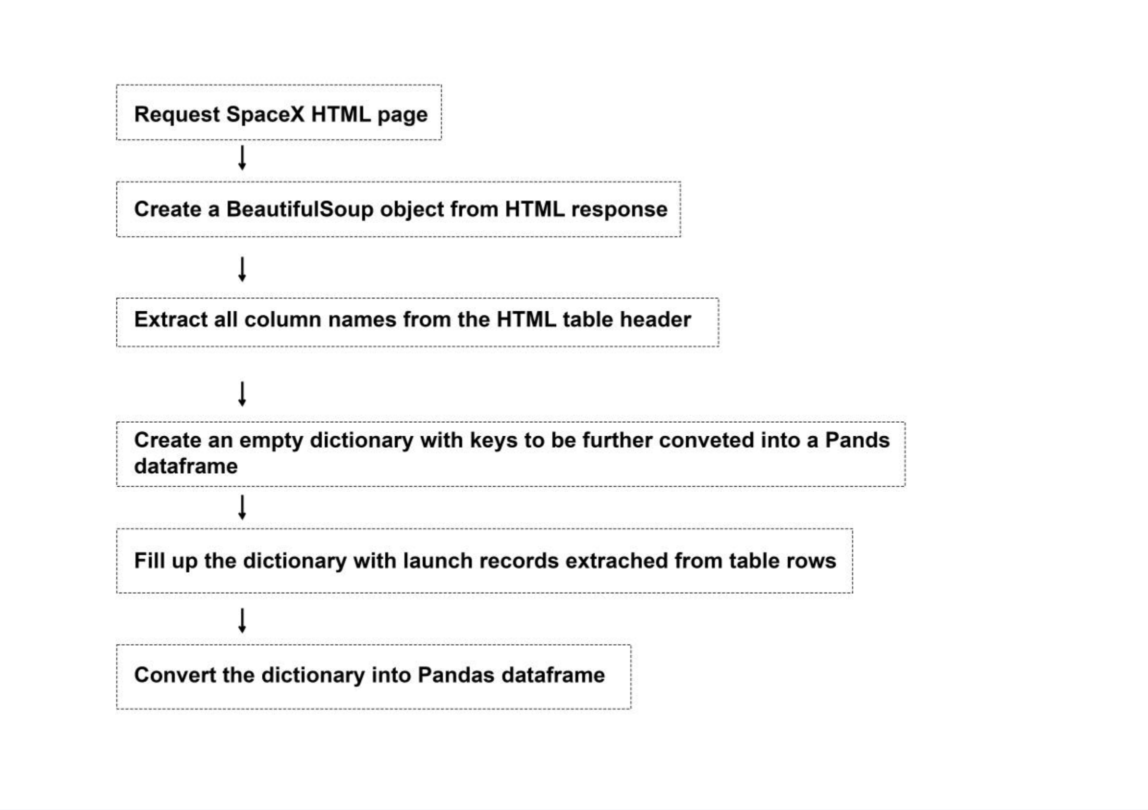
https://github.com/Dogecat0/IBM_DataScience_Captone_Porject/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

Github:

https://github.com/Dogecat0/IBM_DataScience_Captone_Porject/blob/main/jupyter-labs-webscraping.ipynb

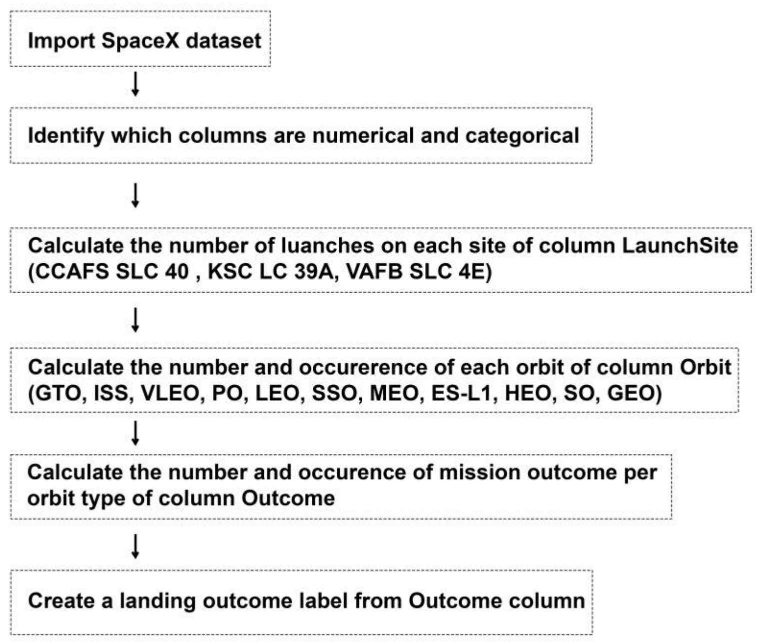


Data Wrangling

- First, understand the success or failure of all the landings through the different terms like 'True Ocean', 'False RTLS', etc. Then process the data by converting those landing outcomes into Training Labels with 1 means the booster successfully landed and 0 means it was unsuccessful.

GitHub:

https://github.com/Dogecat0/IBM_DataScience_Captone_Porject/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Scatter Point Plot, Bar Plot, Line Chart Plot were created to compare relationships between variables. Then we can decide if there is any relationship between them so that they could be used in training the machine learning model.

Github:

https://github.com/Dogecat0/IBM_DataScience_Captone_Porject/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- The performed SLQ queries are:

Names of unique launch sites, Records where launch sites begin with 'CCA', Total payload mass carried by boosters launched by NASA(CRS), Average payload mass carried by booster version F9 v1.1, The first successful landing outcome in ground pad was achieved, Boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000, Total number of successful and failure mission outcomes, Booster versions which have carried max payload mass, The failed landing outcomes in drone ship, Rank the count of landing outcomes between 2010-06-04 and 2017-03-20

GitHub:

https://github.com/Dogecat0/IBM_DataScience_Captone_Porject/blob/main/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- Folium maps mark Launch Sites, successful and unsuccessful landings, and a proximity example to key locations: Railway, Highway, Coast, and City.
- It allows us to understand the reason behind the chosen site. Also, visualizes the relationship between successful landing and sites.

GitHub:

https://github.com/Dogecat0/IBM_DataScience_Captone_Porject/blob/main/lab_jupyter_launch_site_location%20(1).ipynb

Build a Dashboard with Plotly Dash

- Dashboard includes a pie chart and a scatter plot. Pie chart can show distribution of successful landings across all launch sites and also show the success rate of different individual site. Scatter plot has two inputs: The site and payload mass on a slider between 0 and 10000 kg. The pie chart is used to visualize launch site success rate. The scatter plot can help us to check how successful landings distributed across different launch sites, payload mass, and booster version category.

Github:

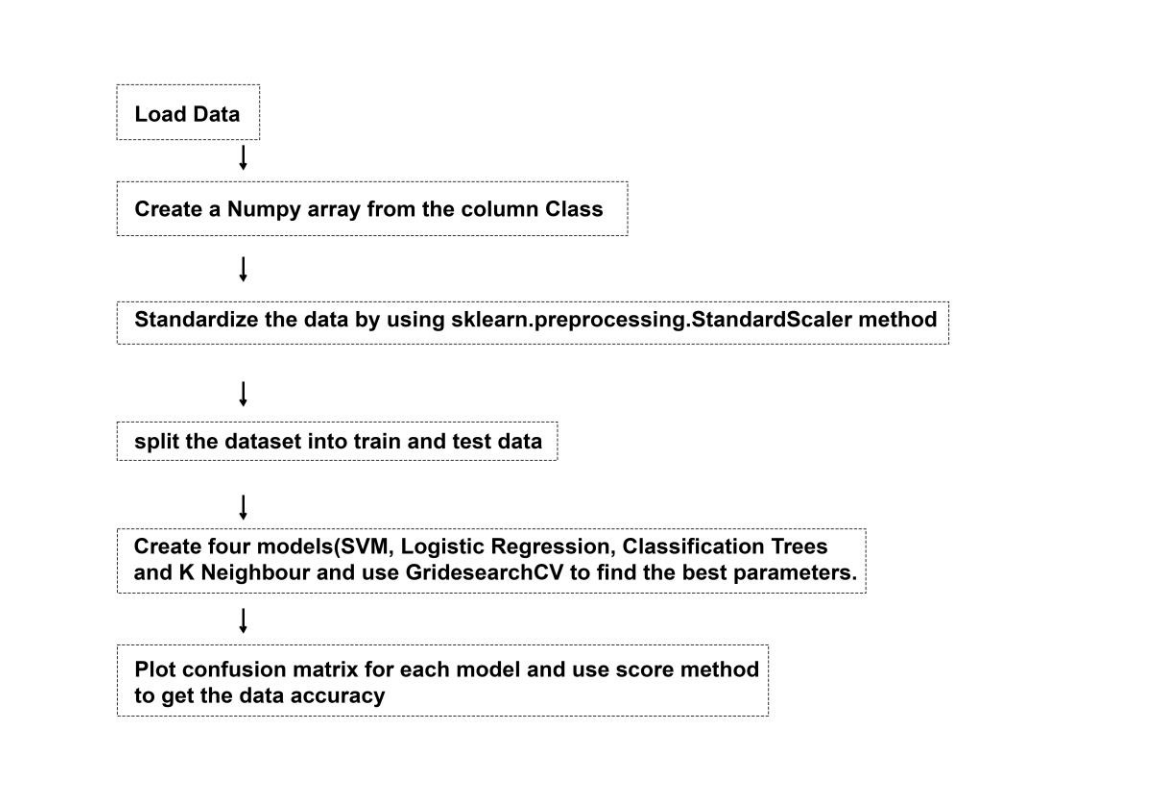
https://github.com/Dogecat0/IBM_DataScience_Captone_Porject/blob/main/dash_lab.ipynb

Predictive Analysis (Classification)

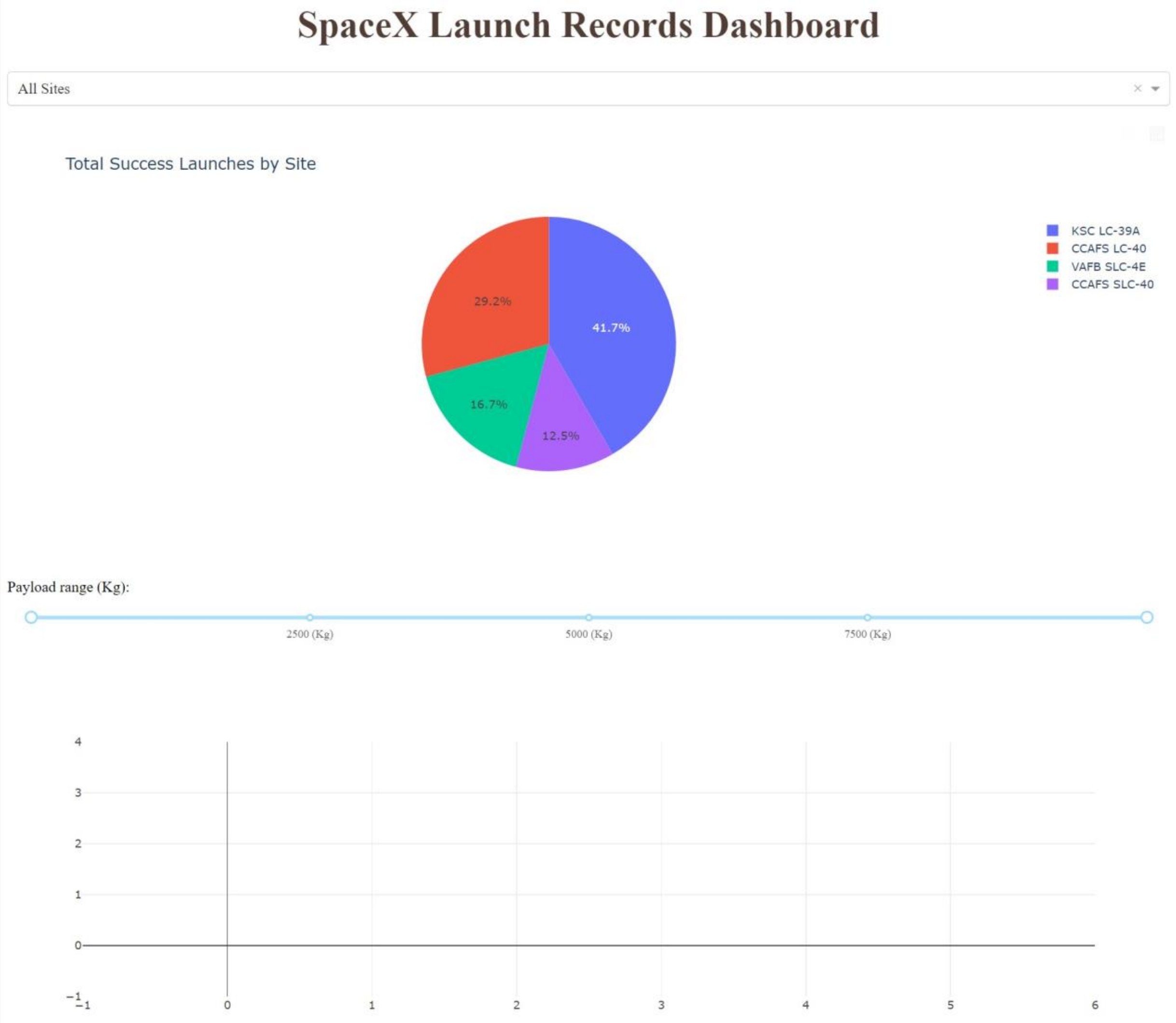
- Create a column for the class, then standardize the data and split into training data and test data for four models (SVM, Classification Trees, Logistic Regression and K Neighbours). Next is to find the best Hyperparameter for four models to achieve the best model accuracy.

Github:

https://github.com/Dogecat0/IBM_DataScience_Captone_Porject/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

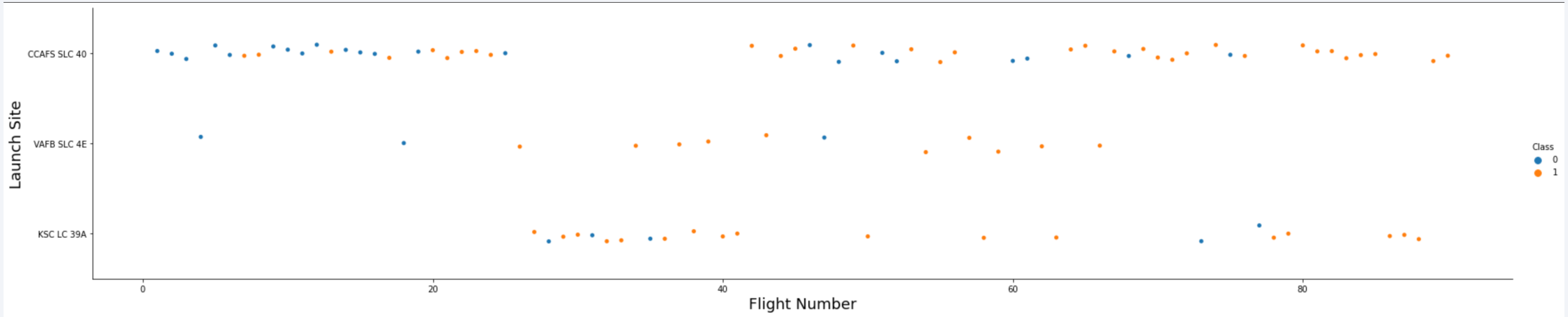


The interactive dashboard is shown on the left page and the results from the machine learning model prediction of all four models is 83.33%

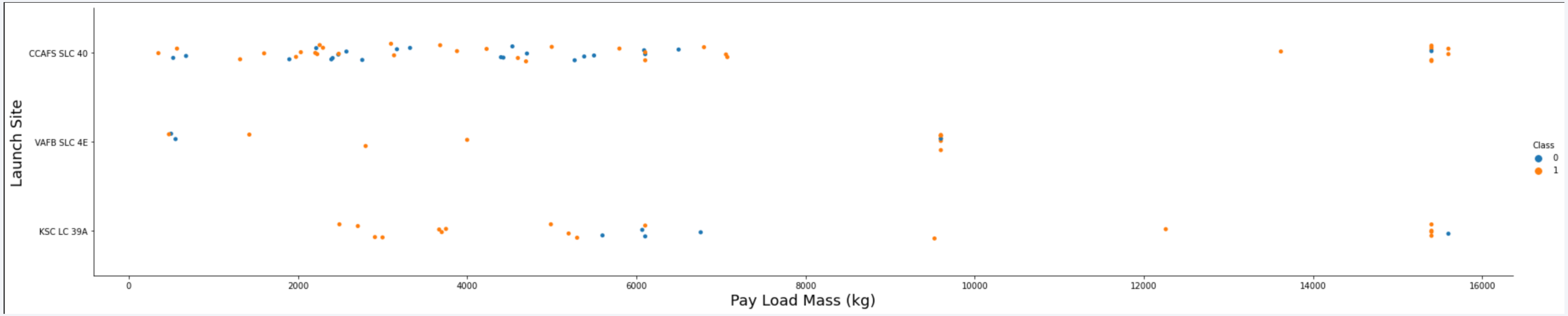
Section 2

Insights drawn from EDA

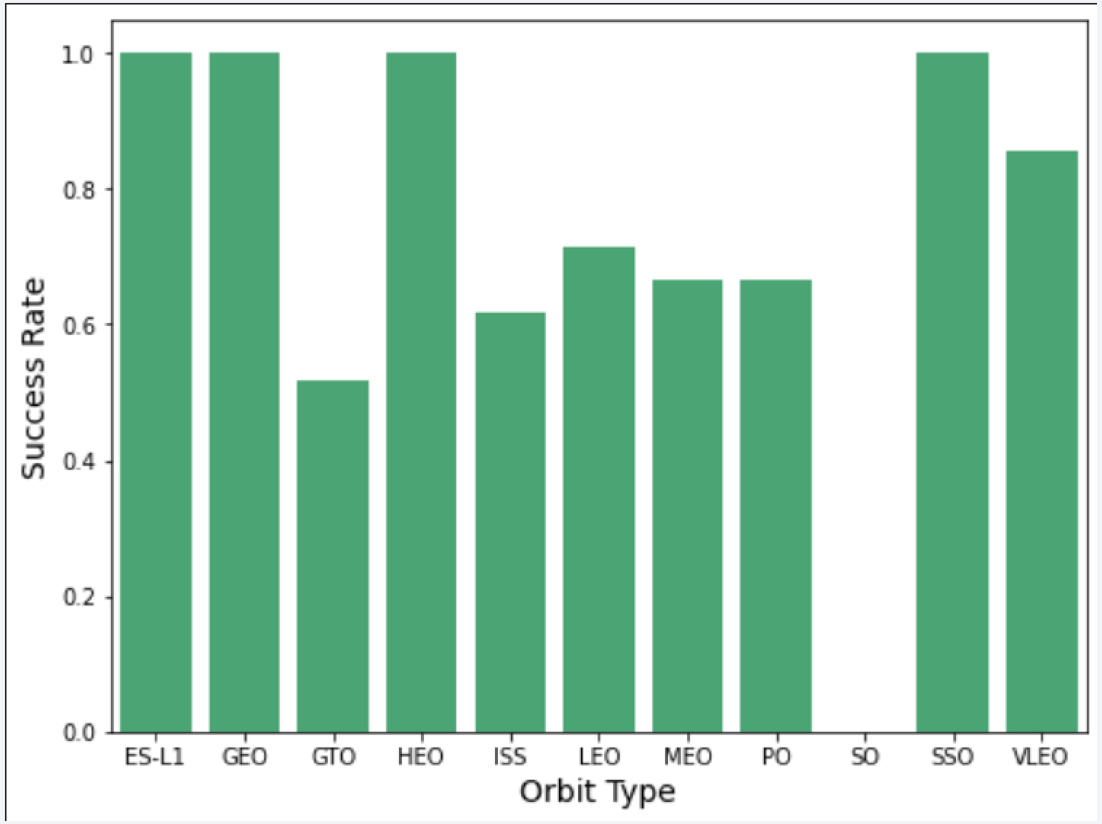
Flight Number vs. Launch Site



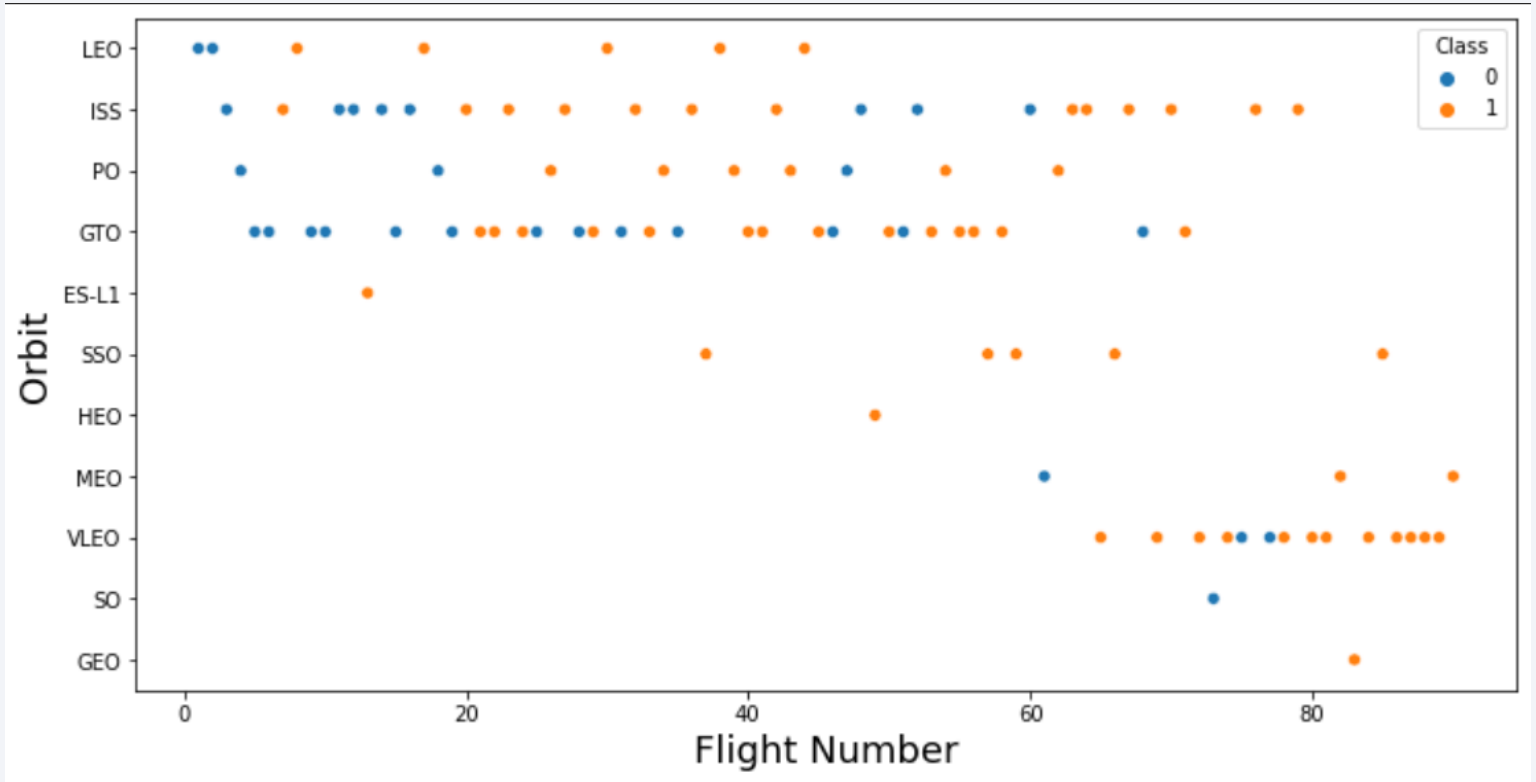
Payload vs. Launch Site



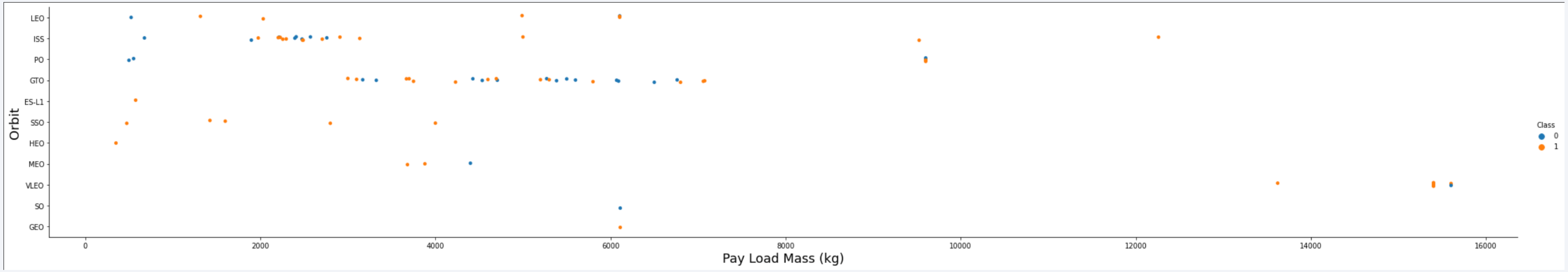
Success Rate vs. Orbit Type



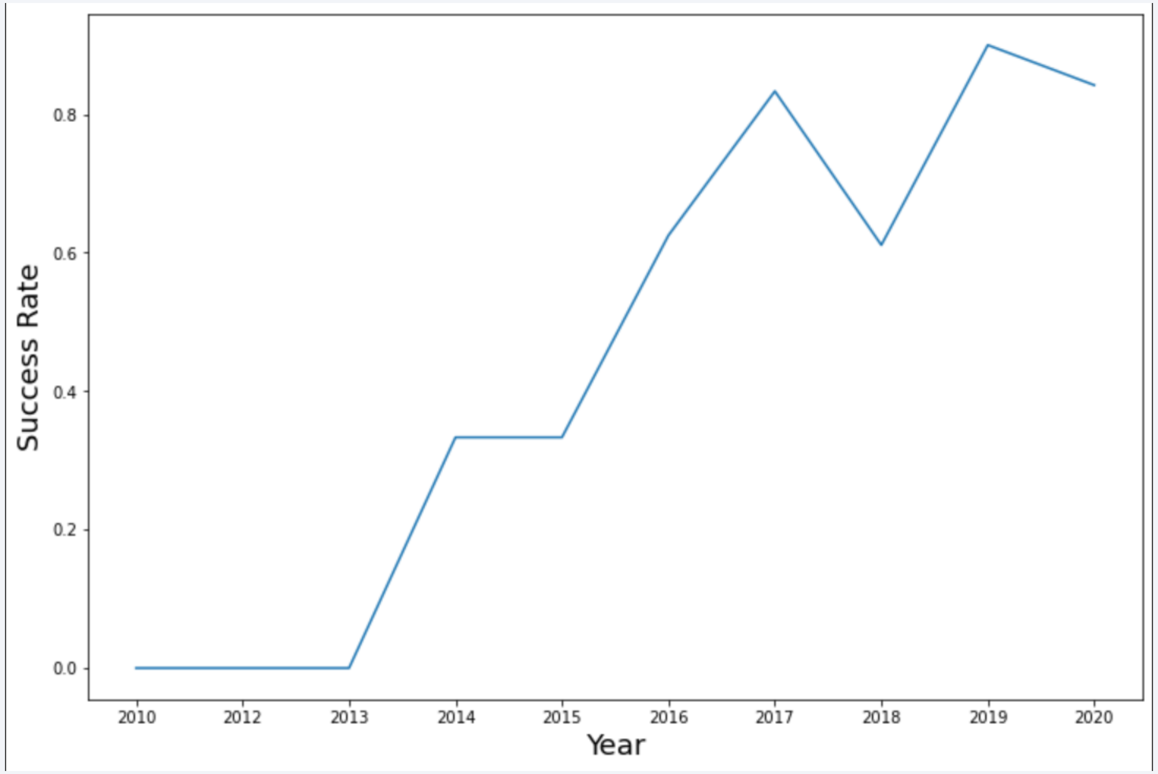
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

[9]
```
%sql
```
```
'Connected: fqr34907@bludb_1'
```

[14]
```
%%sql
select distinct launch_site
from spacextbl
```
```
Done.
```

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

Launch Site Names Begin with 'CCA'

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

```
%%sql
select *
from spacextbl
where launch_site like 'CCA%'
limit 5
```

Done.

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

```
%%sql
select sum(payload_mass__kg_)
from spacextbl
where customer = 'NASA (CRS)'
```

Done.

1
45596

Average Payload Mass by F9 v1.1

```
%%sql
select avg(payload_mass__kg_)
from spacextbl
where booster_version = 'F9 v1.1'
```

Done.

1
2928

First Successful Ground Landing Date

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

Hint:Use min function

```
%sql select distinct landing__outcome from spacextbl
```

Done.

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

```
%%sql
select min(date)
from spacextbl
where landing__outcome = 'Success (ground pad)'
```

Done.

1
2015-12-22

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

```
%%sql
select booster_version
from spacextbl
where landing__outcome = 'Success (drone ship)' and 4000 < payload_mass__kg_ < 6000
```

Done.

booster_version
F9 FT B1021.1
F9 FT B1023.1
F9 FT B1029.2
F9 FT B1038.1
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql select distinct mission_outcome from spacextbl
```

Done.

mission_outcome
Failure (in flight)
Success
Success (payload status unclear)

```
%%sql
select count(mission_outcome)
from spacextbl
where mission_outcome like 'Success%'
```

Done.

1
100

Boosters Carried Maximum Payload

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

```
%%sql
select booster_version
from spacextbl
where payload_mass__kg_ = (select max(payload_mass__kg_) from spacextbl)
```

Done.

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

```
%sql select distinct landing__outcome from spacextbl
```

Done.

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

```
%%sql
select landing__outcome, booster_version, launch_site, date
from spacextbl
where landing__outcome = 'Failure (drone ship)' and date like '2015%'
```

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

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

```
%sql select date from spacextbl limit 5
```

Done.

DATE
2010-06-04
2010-12-08
2012-05-22
2012-10-08
2013-03-01

```
%%sql
select count(landing__outcome),landing__outcome
from spacextbl
where date between '2010-06-04' and '2017-03-20'
group by landing__outcome
order by count(landing__outcome) desc
```

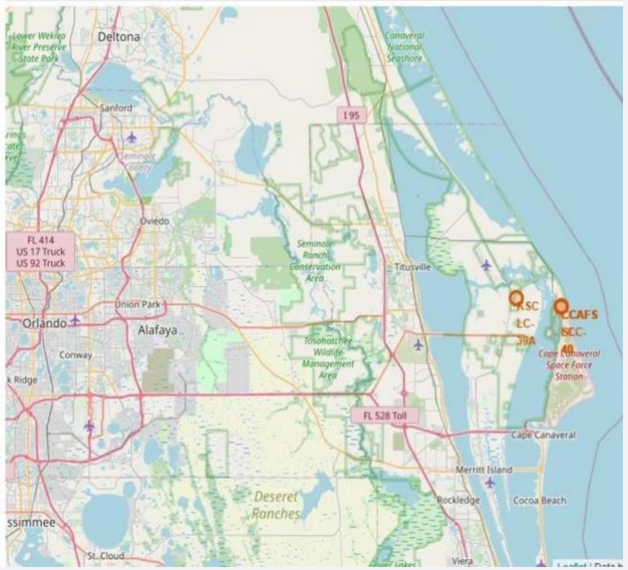
Done.

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

Section 4

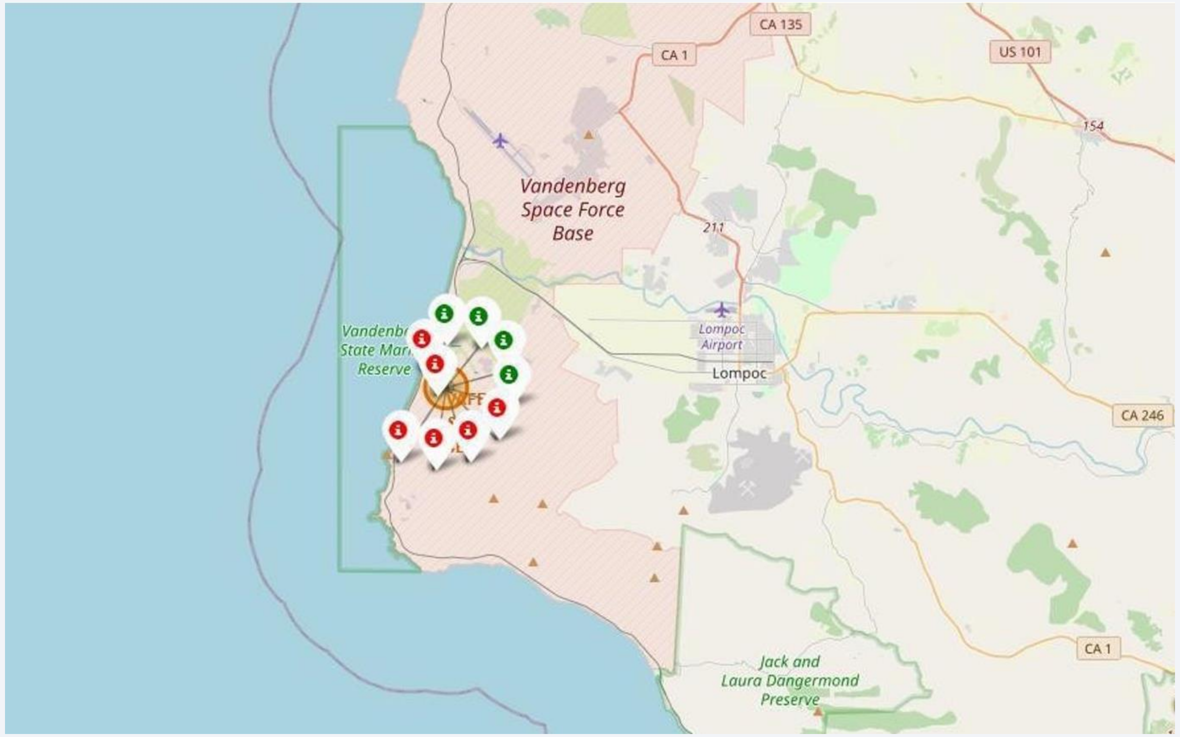
Launch Sites Proximities Analysis

<Location of Launch Site>



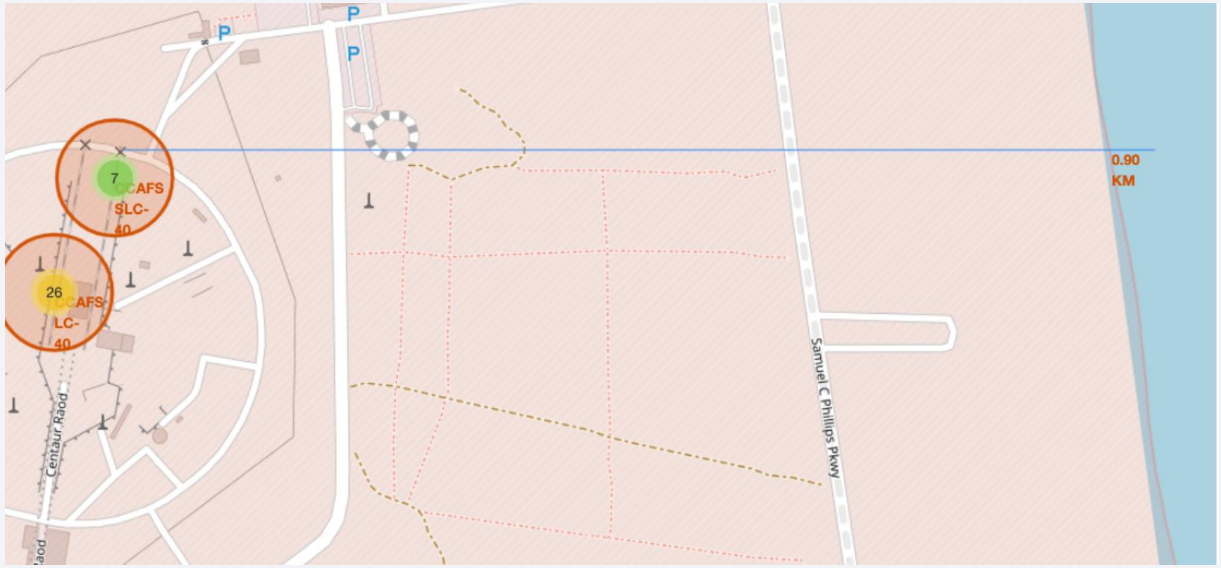
Map on the left shows all launch sites in US map. The right map shows the two Florida launch sites since they are very close to each other. All site and close to sea and equator.

<Launch Markers>



Clusters can be clicked on to
display each successful landing with green
icon and failed
landing with red icon. In this example VAFB SLC-
4E shows 4 successful landings and 6 failed landings.

<Key Location Distance Proximities>



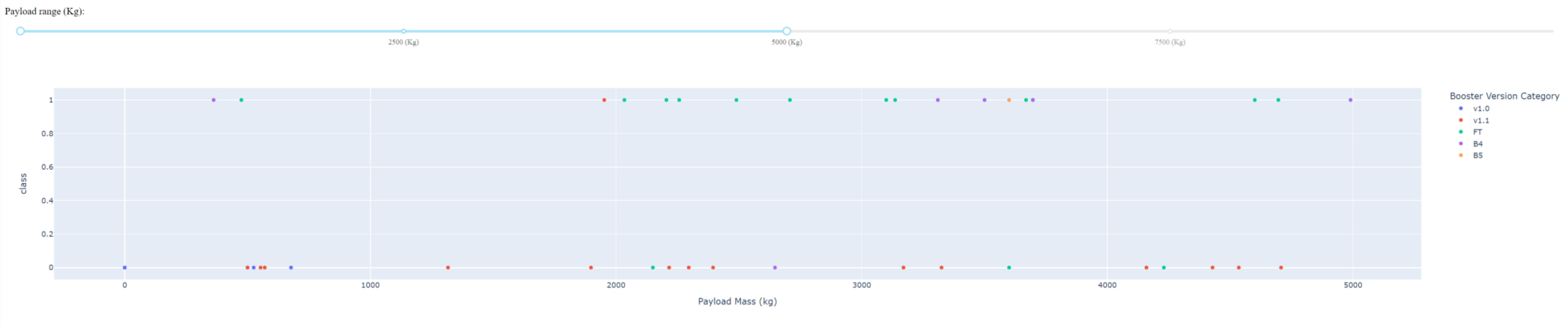
Section 5

Build a Dashboard with Plotly Dash

<Successful Launches of All Sites>

<Launch Site with Highest Succes Ratio>

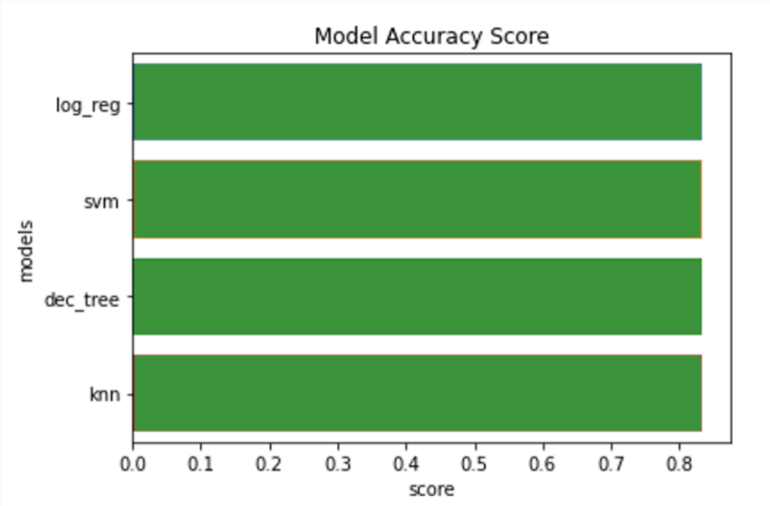
<Payload vs Launch Outcome>



Section 6

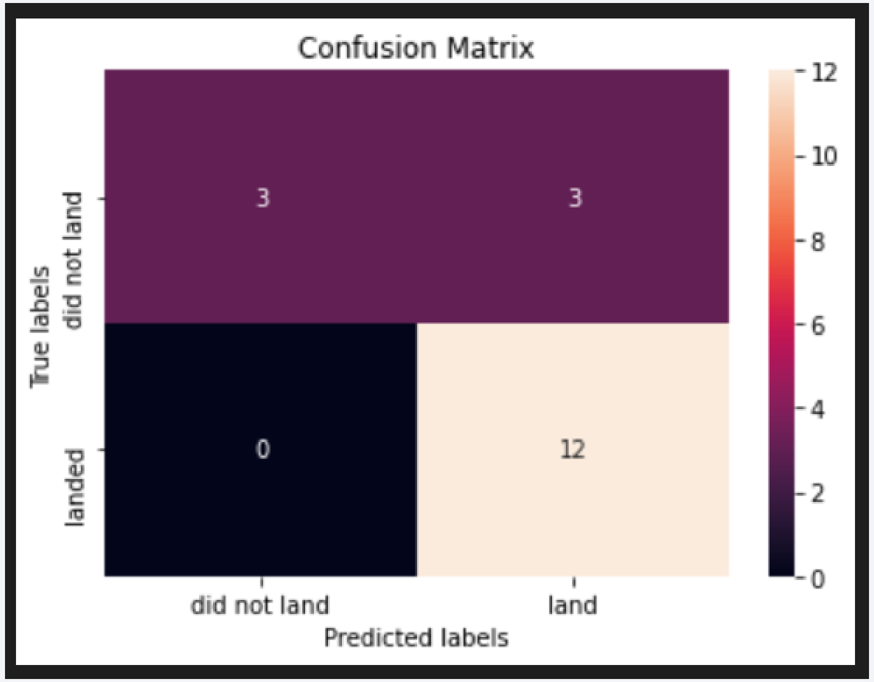
Predictive Analysis (Classification)

Classification Accuracy



All four models have the same accuracy score with 83.33%

Confusion Matrix



Since the all four models have the same accuracy so the confusion matrix is also the same and from the screenshot that we can see that all four models can distinguish between the different classes. We see that the major problem is false positives. Our models over predict successful landings with 3 predictions of success which are unsuccess in reality.

Conclusions

Back to the starting point of this project, Space X 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 Space X 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 space X for a rocket launch. So with the work we carried out, we can predict that the success rate of the first stage of Falcon 9 is 83.33% which has a relatively high rate of success which should be used as an evidence to help the stakeholders to decide whether future work should be continued or not. But also if more data can be collected in the future, we will be able to carry out more precise model with better prediction.

Thank you!