IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Bora Jeong

Mar 26, 2023

Outline

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

Executive Summary

- Summary of methodologies
- This project has been done with below sequence:
  - Collect the Space X's historical launch data from Space X API or webscrapping from wiki pages using request library.
  - Proceed the data wrangling for preprocessing the data. This includes extracting only necessary data, removing duplicates, all the datatype correction and dealing with missing values.
  - Go through EDA process with data visualization. Figure out what might be the variable that affects the success rate.
  - Build the interactive map visualization and dashboard for stakeholders.
  - Build, train the model and predict the success rate.
- Summary of all results
  - From our model, the first stage landing success predicted to be 8 out of 10.
  - The success rate has been significantly increased over the years.
  - Heavier payload mass, orbit type and flight number turn out to affect the success rate.

Introduction

- Project background and context
  - SpaceX advertises Falcon9 rocket launches and its cost savings from reusing the first stage. Reusing the first stage is not common for other companies, which this is Space X's competitive advantage.
  - However, the first stage doesn't land successfully all the time. So, impossible to reuse the first stage all the time.
  - The competitor Space Y want to predict the price of each launch of Space X by predicting if Space X will reuse the first stage or not

- Problems you want to find answers
  - The problem to solve is **determining the cost of each launch of Space X**
  - This can be achieved by **predicting that the first stage will land successfully or not**
  - The data scientists in Space Y will build and train the machine learning model to predict the first stage's success

Section 1

Methodology

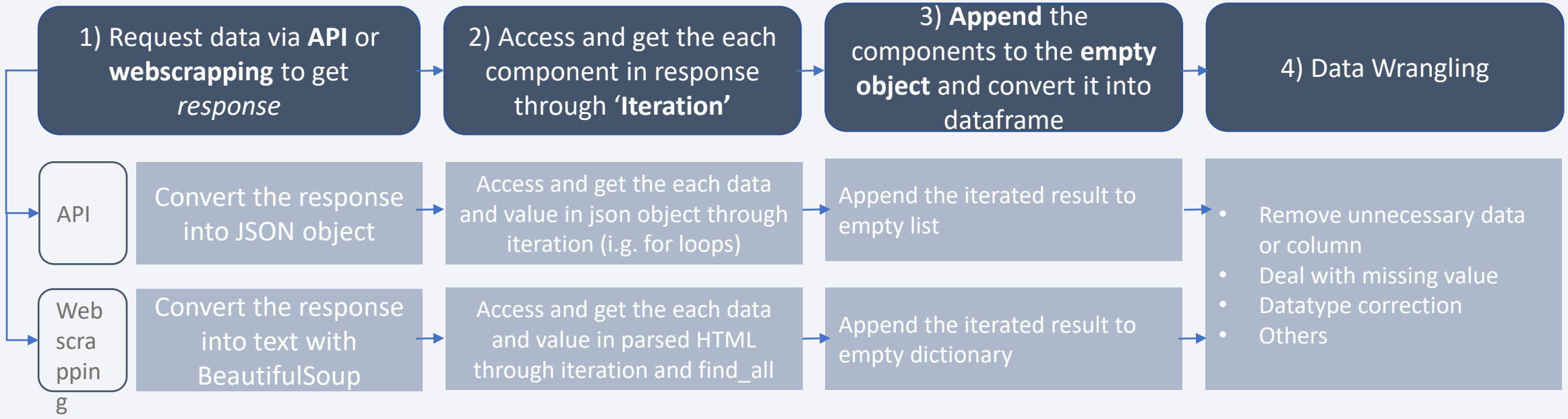
Methodology

Executive Summary

- Data collection methodology:
  - Collected raw data from Space X API or Webscrapping using request library and convert into dataframe
- Perform data wrangling
  - Filtered unnecessary data (i.e. Falcon 1), dealt with missing value and reformed landing success with 1 or 0 for model training purpose
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- After going through feature engineering and training test data split process, built and train the four classification models with hyperparameter tuning and accessed each model's performance based on accuracy score

Data Collection

- Able to collect the historical rocket launch data from API or Web scrapping
- Even though there are some minor differences between both ways in process but in overall, the stage is similar as below.



Data Collection – SpaceX API

1) Collect the data from several API and convert it into dataframe

Request rocket launch related data from SpaceX API with following URL, using Get request (Requests.get()) and convert it into json object with json ()

↓

URLs:
https://api.spacexdata.com/v4/rockets/ for rocket name
https://api.spacexdata.com/v4/launchpads/ for launch site name
https://api.spacexdata.com/v4/payloads/ for payload
https://api.spacexdata.com/v4/cores/ for other informations

↓

Append the data to empty list and make all the list into dataframe

↓

2) Data Wrangling

In the dataframe, exclude Falcon 1 data

↓

Deal with missing value, replacing them with mean value for payload mass

- Please find below Github URL:

https://github.com/borajeong88/Capstone_project/blob/a35b9f00083e9b78a792699d19486614a902c995/%5BCapstone_week1%5D%20Finished_Complete%20the%20Data%20Collection%20from%20API%20Lab.ipynb

Data Collection - Scraping

1) Request the Falcon 9 Launch records from wiki page via BeautifulSoup

Web scrape html table with BeautifulSoup package that contains Falcon 9 historical launch records from a Wikipedia via Requests.get() and parse the response

2) Get the column name

Extract the column name from HTML table header by using 'Find_all('table') & Find_all('th')

3) Parse the data and convert it into dataframe

Make the empty dictionary with key name from the stage 2

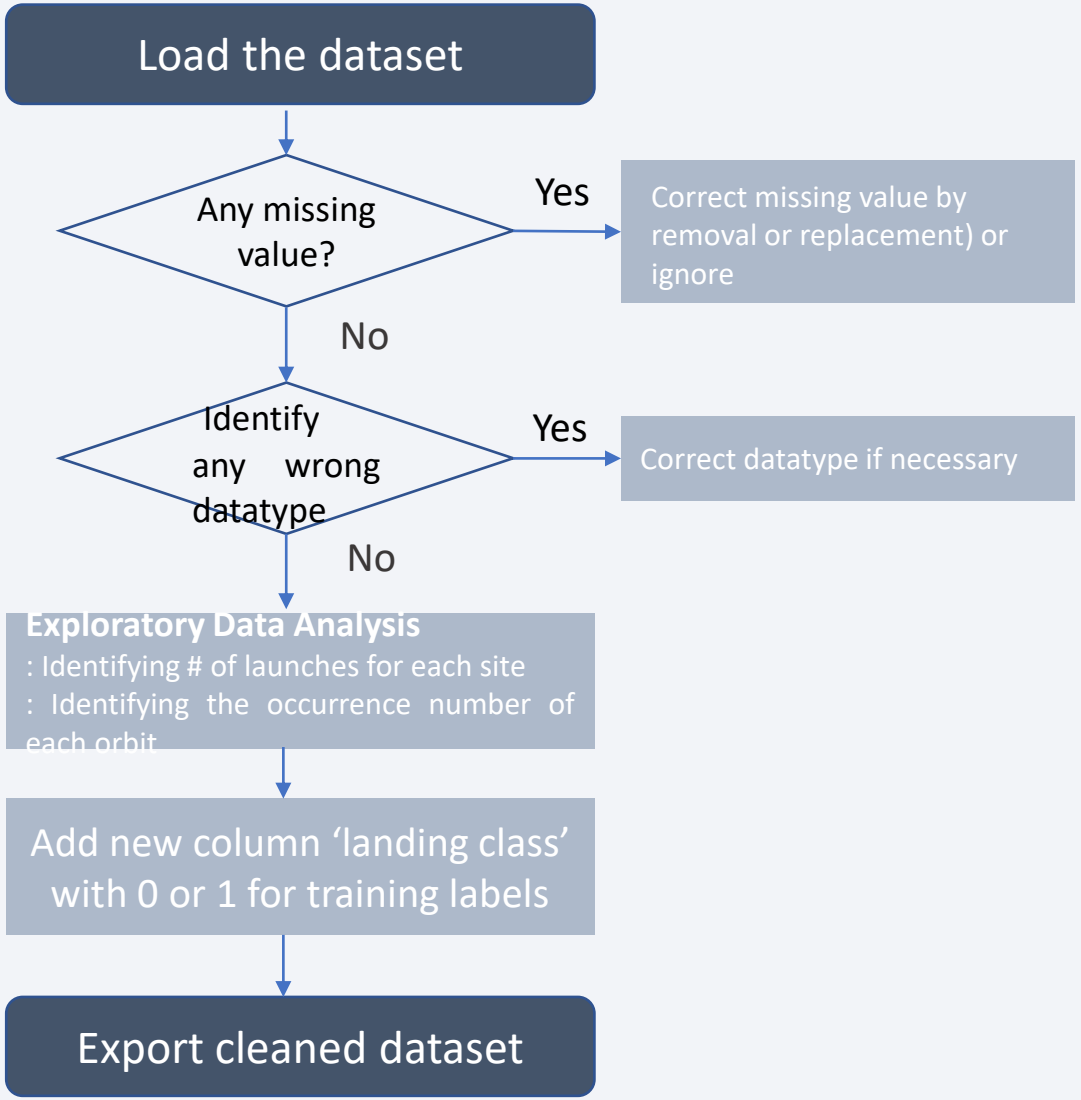
Parse the launch HTML table and get each component such as *flight number, date, booster version, launch site and payload, etc.* with Find_all('tr') & find_all ('td')

Fill the dictionary made earlier with parsing data from table and convert the dictionary into dataframe

- Please find below Github URL:

https://github.com/borajeong88/Capstone_project/blob/8ae7fca5fb79c7771d98ba5da838aa9ce626ffb2/%5BCapstone_week1%5D%20Finished_Complete%20the%20Data%20Collection%20with%20Web%20Scrapping%20Lab.ipynb

Data Wrangling



- Please find data wrangling processed in the left flowchart.
  - After loading data from collection stage, figured out any missing value or incorrect datatype.
  - Went through simple EDA process which includes identifying # of launches for each site and occurrence number from each orbit by using 'value_counts' method.
  - To label the training data (Y variable), added the new grouped data named 'landing class' which only includes 0 for landing failure and 1 for landing success

- Please find below Github URL:

https://github.com/borajeong88/Capstone_project/blob/3cdda348a83fe429c18e3fbab3c00ac1ee016bc5/%5BCapstone_week1%5D%20Finished_Data%20Wrangling.ipynb

EDA with Data Visualization

- Mostly, **scatterplot** was plotted since scatterplot is intensive to figure out the relationship between two numerical variables. Also, adding 'hue' parameters for class (=success or not) allows me to find pattern more easily. In my exercise, scatterplots were plotted for:
  - Flight number vs. payload mass
  - Flight number vs. launchsite
  - Payload mass vs. launchsite
  - Flight number vs. orbit type
  - Payload mass vs. orbit type
- However, other chart such as **bar chart** and **line chart** also has been used for launch success rate by orbit type and average success rate by year trend respectively.
- For notebook as a reference, please find below Github URL:

https://github.com/borajeong88/Capstone_project/blob/56a12c792e1ab4fad99ebffc635f3f9021a93c4b/%5BCapstone_week2%5D%20Finished_Complete%20EDA%20with%20Visualization%20Lab.ipynb

EDA with SQL

Exploratory Data Analysis	SQL Queries used
• What is the unique launch sites names?	• select distinct Launch_Site from SPACEXTBL
• 5 records where launch sites begin with CCA	• select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5
• What is the toal payload mass carried by boosters launched by NASA (CRS)	• select sum(PAYLOAD_MASS__KG_) as total_payload_mass from SPACEXTBL where Customer = 'NASA (CRS)';
• What is the average payload mass carried by booster version F9 v1.1?	• select avg(PAYLOAD_MASS__KG_) as average from SPACEXTBL where Booster_Version like 'F9 v1.1%';
• When was the first succesful landing outcome in ground pad?	• select min(Date) as first_date from SPACEXTBL where [Landing _Outcome] = 'Success (ground pad)';
• What are the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000?	• select distinct Booster_Version from SPACEXTBL where (PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000) and [Landing _Outcome] = 'Success (drone ship)'

- For notebook as a reference, please find below Github URL:

https://github.com/borajeong88/Capstone_project/blob/2bad29b793a9754a5c1947acd3d54786fbfe449b/%5BCapstone_week2%5D%20Finished_Complete%20EDA%20with%20SQL.ipynb

EDA with SQL

Exploratory Data Analysis

- What is the total number of successful and failure mission outcomes?
- What is the booster_versions which have carried the maximum payload mass?
- The month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017

SQL Queries used

- select Mission_Outcome, count(Mission_Outcome) as total_num_mission_outcomes from SPACEXTBL group by Mission_Outcome;
- select distinct Booster_Version from SPACEXTBL
- where (PAYLOAD_MASS__KG_) in (select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
- select substr(Date, 4, 2) as month, substr(Date, 7, 4) as year, Booster_Version, Launch_Site from SPACEXTBL where substr(Date, 7, 4) = '2015' and [Landing _Outcome] = 'Failure (drone ship)'
- select *,
  substr(Date, 7, 4) || '-' || substr(Date, 4, 2) || '-' || substr(Date, 1, 2) as Date2,
  rank () over (order by substr(Date, 7, 4) || '-' || substr(Date, 4, 2) || '-' || substr(Date, 1, 2) desc) Rank from SPACEXTBL
  where
  Date2 between '2010-06-04' and '2017-03-20' and
  [Landing _Outcome] like 'Success%'
  order by Date2 desc

Build an Interactive Map with Folium

- First, created the initial map object
- And, added marker (*folium.marker*) , icon marker *(folium.Icon)*, lines *(folium.PolyLine)*, circles (*folium.Circle*) and mark cluster, to figure out:

I. The exact location of each site

II. # of launch attempts and compare how many success or failure per each site

III. The near proximities

- For notebook as a reference, please find below Github URL:

https://github.com/borajeong88/Capstone_project/blob/80599f0aa1af6d9eb2ada4080046402b9832c9e2/%5BCapstone_week3%5D%20Finished_Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- In a dashboard, two charts has been created and added:

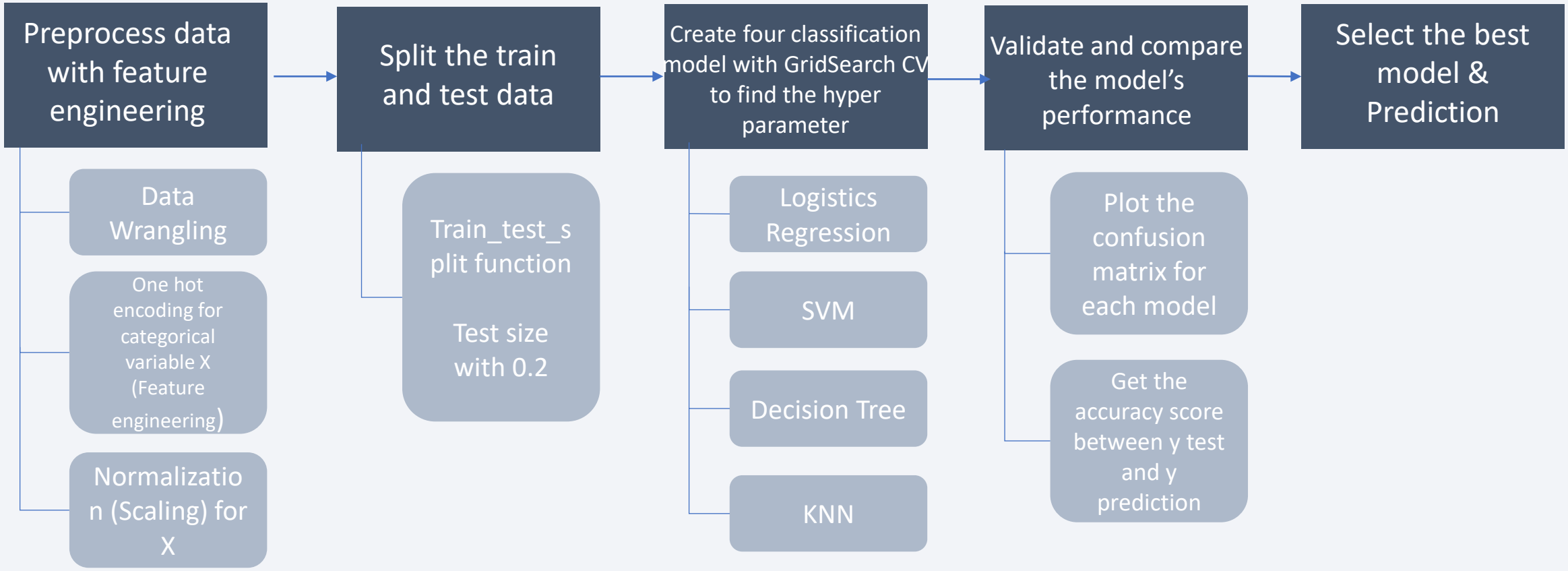
  - First, the **pie chart** shows the **successful launches count by all sites**, This will help audience to identify which site has the largest success launches.
  - For more details, audience can select the site from the drop down. if the specific site selected, audiences also check the exact success count and failure count for the selected site. This will allow audience to compare success rate by each site.
  - In addition, **scatterplot** between **payload mass** and **launch success rate** was plotted as well. Audience also can check the success rate by F9 booster version. They can also control the payload range with range slider. This will help them to figure out easily which payload range has the highest success rate.

- For notebook as a reference, please find below Github URL:

https://github.com/borajeong88/Capstone_project/blob/eb4e283279a6b7d21d3ba89df3cd6d174537096c/%5BCapstone_week3%5D%20Finished_Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash.ipynb

Predictive Analysis (Classification)

- After preprocessing the X and Y data, split the train and test data
- Create several classification models with hyperparameter tuning and access each model's performance
- Predict y value with the best model. Please find below flow chart for more details.



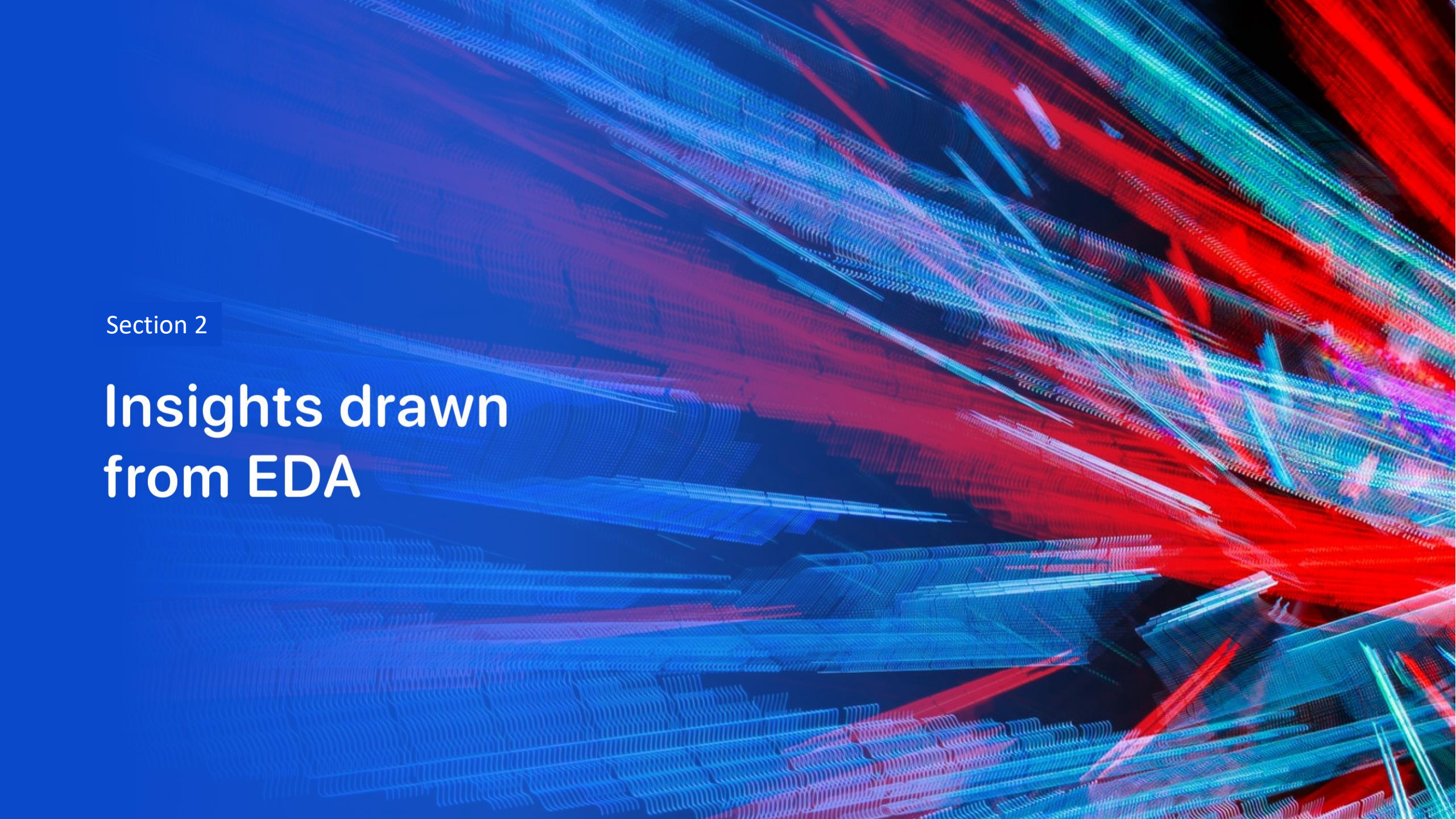
Predictive Analysis (Classification)

- For notebook as a reference, please find below Github URL:

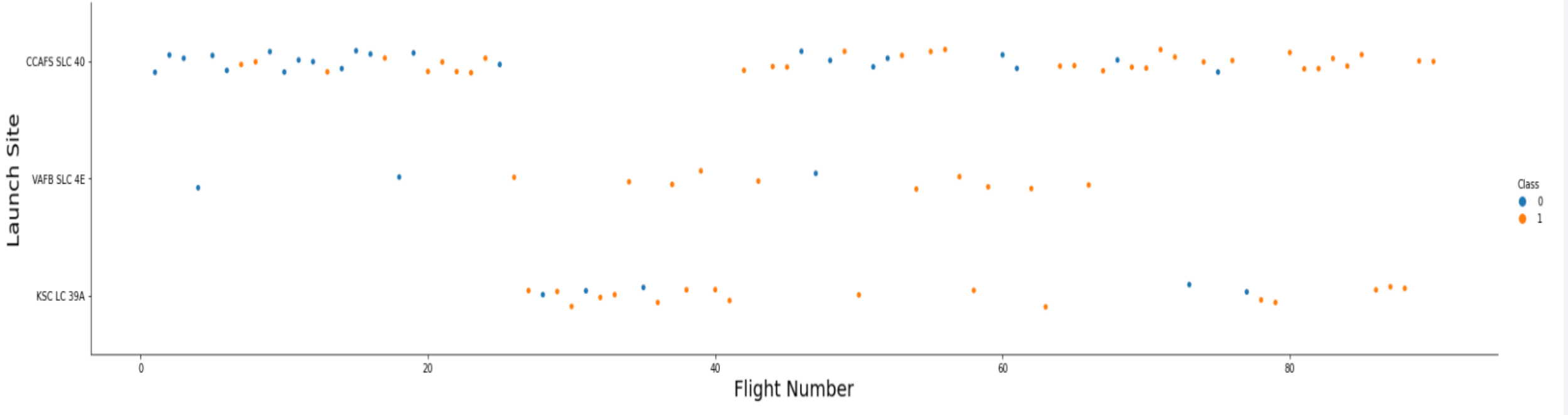
https://github.com/borajeong88/Capstone_project/blob/2bad29b793a9754a5c1947acd3d54786fbfe449b/%5BCapstone_week4%5D%20Finished_Complete%20the%20Machine%20Learning%20Prediction%20Lab.ipynb

Results

- In the early stage, especially before 2013, Space X's the first stage landing success rate was very low but **it significantly increase over years**. After 2019, its rate reaches over 80%
- There are **four different launch site**. **One in CA**, and **the other threes are in FL**. Each site has different number of trial and different success rate.
- Even though launches tried in **CCAFS LC-40 the most**, but **KSC LC-39A** has the **highest success rate**.
- The each site aims to dedicate the different types of orbit.
- The heavier payload and orbit type has some connection with the launch's success rate.
- From the trained model, out of 18 cases, 15 landings were predicted to be a success with 3 failure **(predicted success rate: 83%)**
- The four different models have been built to compare its performance and the result was same. our model's accuracy score is 83%

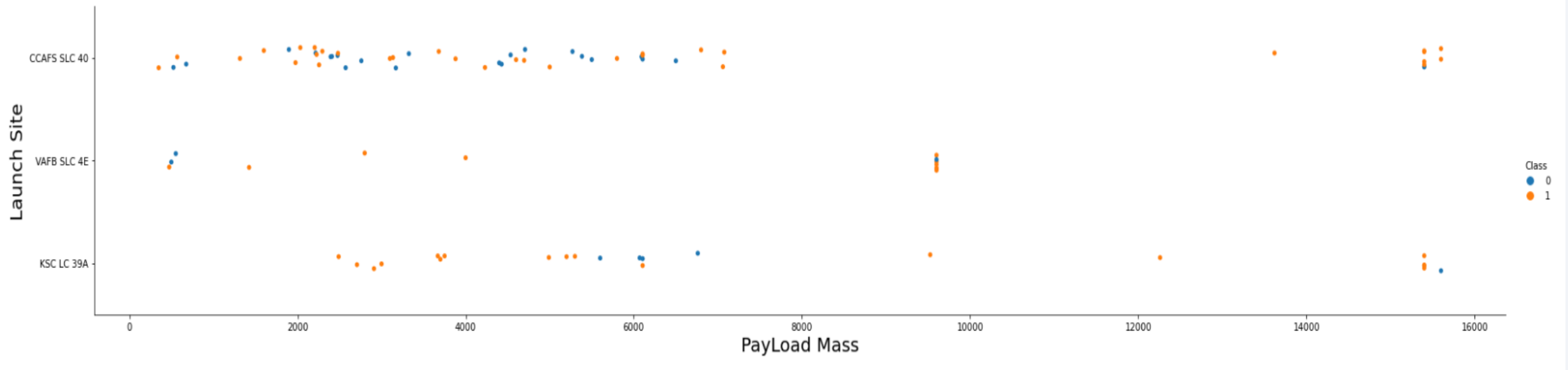

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



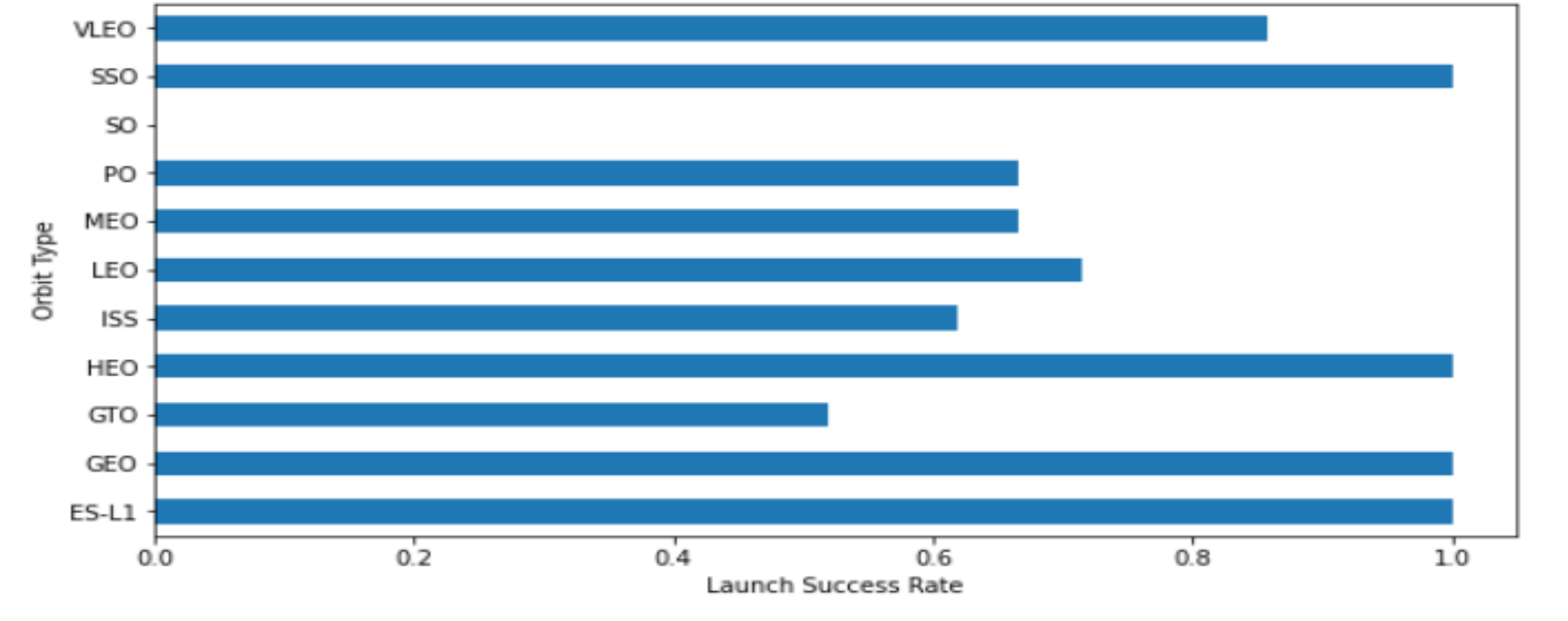
- Plotted scatterchart with catplot() in Seaborn library.
- Even though 'KSC LC 39A' and 'VAFB SLC 4E' site has lower number of trial, its success rate looks higher.
- CCAFS SLC 40 sites shows the higher success rate as launch attempts increases (with higher flightnumber)

Payload vs. Launch Site



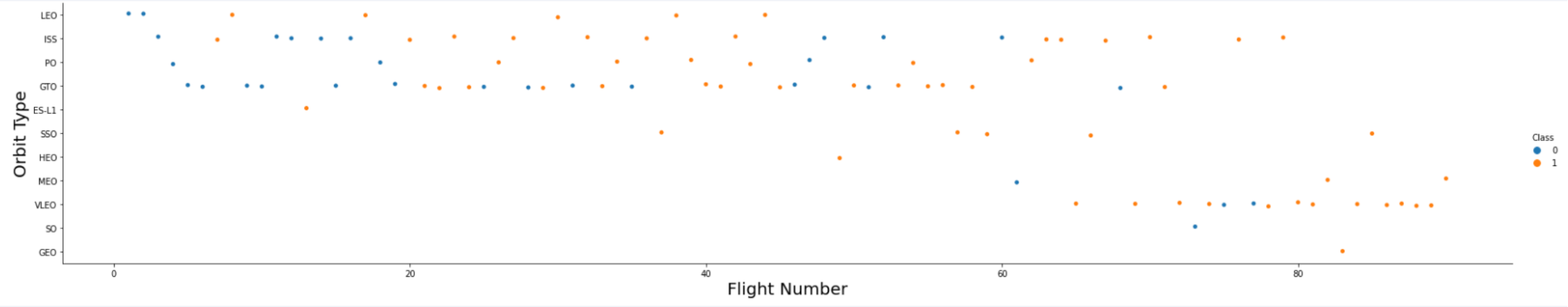
- Plotted scatterchart with catplot() in Seaborn library.
- 'VAFB SLC 4E' site has no rocket launched for heavy payload mass that is greater than 10000
- It looks like that there is relevant pattern between payload mass and success rate. The rockets with lighter payload mass shows the less success rate.

Success Rate vs. Orbit Type



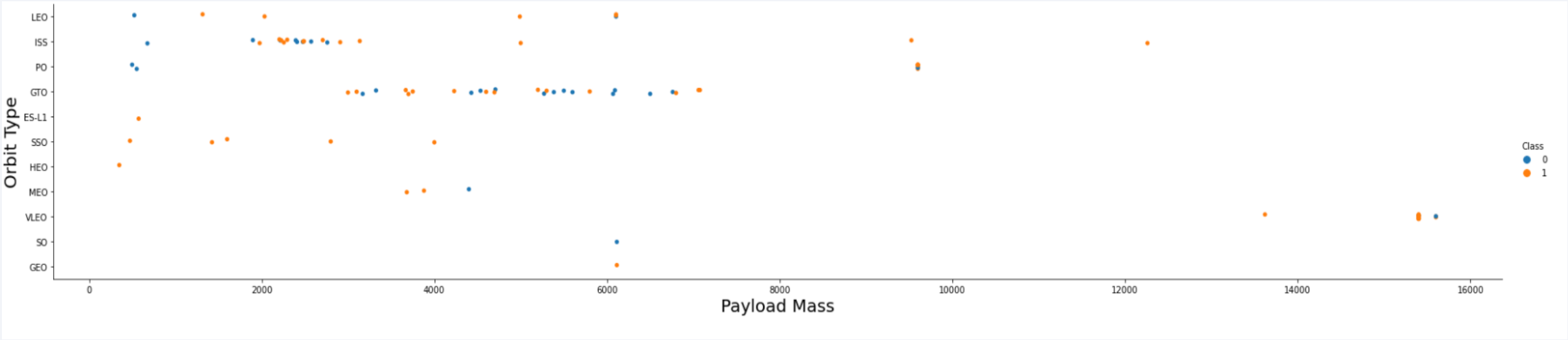
- Plotted bar chart between ‘success rate’ and ‘orbit type’
- Orbit type ‘GTO’ has the lowest success rate, while orbit like ‘SSO’ shows the 100% launch success rate.

Flight Number vs. Orbit Type



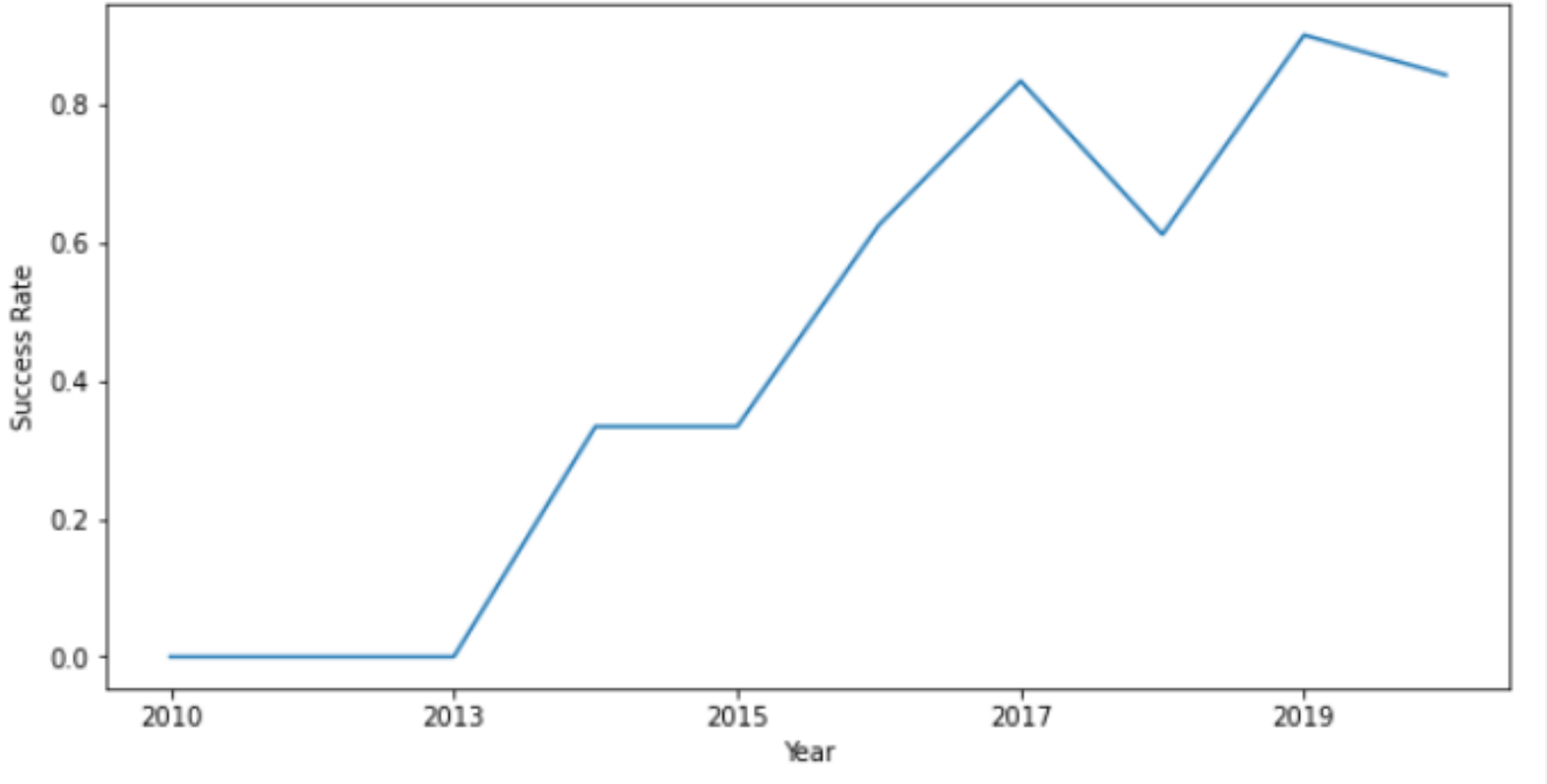
- Plotted scatterchart with catplot() in Seaborn library.
- In LEO orbit, the success appears to increase with higher flight number; on the other hand, there seems to be no relationship between flight number when in GTO orbit.
- There are some orbits types such as SSO and VLEO related to many launch attempts (= higher flight number)

Payload vs. Orbit Type



- Plotted scatterchart with catplot() in Seaborn library.
- The correlation between payload mass and success rate differs depends on orbit type.
- For example, the more positive landing success with heavier payload mass for LEO and ISS; while there are no clear tendency for orbit type, GTO.

Launch Success Yearly Trend



- Plotted line chart with x axis of ‘year’ and y axis of ‘success rate’
- Until 2013, the success rate is 0% but it start to increase after that. It continues to increase over years.

All Launch Site Names

- The unique launch sites are:

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

- '**Distinct**' syntax was used to get the unique launch site name in select statement

Task 1

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

[7]:
```
%%sql

select distinct Launch_Site from SPACEXTBL
```

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

[7]:

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

Launch Site Names Begin with 'CCA'

- Please find below screenshot for query and its result
- Used **'like' with '%'** in '**where' clauses** to find launch sites that begins with 'CCA'

```
8]: %%sql

select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5
```

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

8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- The total payload by 'NASA CRS' is ***45,596 KG***
- Please find below query and its result
- Used **'Sum' function** to get total value in select statement and filtered 'NASA' from where clauses.

Task 3

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

[9]:
```
%%sql

select sum(PAYLOAD_MASS__KG_) as total_payload_mass from SPACEXTBL where Customer = 'NASA (CRS)';
```

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

[9]:

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- The average payload mass by booster version F9 v1.1 is ***2534.67KG***
- Please find below query and its result
- Used '**Avg**' function in select statement and filtered '**F9 v1.1' from where clauses**

Task 4 ¶

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

```
]: %%sql

   select avg(PAYLOAD_MASS__KG_) as average from SPACEXTBL where Booster_Version like 'F9 v1.1%';
```

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

]:

average
2534.6666666666665

First Successful Ground Landing Date

- The first dates of the successful landing outcome on ground pad was *May 1st, 2017*
- Used 'Min' function to get the first date and filtered 'ground pad success' from 'Where' clauses.

Task 5

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

Hint:Use min function

```
[1]: %%sql

     select min(Date) as first_date from SPACEXTBL where [Landing _Outcome] = 'Success (ground pad)';
```

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

[1]:

first_date
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which landed on drone ship with payload mass between 4000 and 6000s are:

  F9 FT B1022

  F9 FT B1026

  F9 FT B1021.2

  F9 FT B1031.2

- Used 'Distinct' function in select statement to get the unique boosters name that meet the condition

- Conditions are described in where clauses

Task 6

List the names of the boosters which have success in drone ship and have payl

```
%%sql

select distinct Booster_Version
from SPACEXTBL
where (PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000) and
[Landing _Outcome] = 'Success (drone ship)'
```

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

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

Total Number of Successful and Failure Mission Outcomes

- Please find below query and its result
- Used ‘Group by’ function to aggregate the outcome number by mission success or failure

Task 7 ¶

List the total number of successful and failure mission outcomes

```
%%sql

select Mission_Outcome, count(Mission_Outcome) as total_num_mission_outcomes from SPACEXTBL group by Mission_Outcome;
```

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

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

Boosters Carried Maximum Payload

- Below is the boosters names with maximum payload mass

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

- Used 'subquery' to find maximum payload mass in where clauses

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Us

```
%%sql

select distinct Booster_Version
from SPACEXTBL
where (PAYLOAD_MASS__KG_) in (select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
```

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

2015 Launch Records

- Please find below query and its result
- Used ‘substr’ function in select statement to get the ‘Year’ from Date column
- Also, indicated two conditions (Year 2015 and failure drone ship) in ‘where’ clauses

[19]:
```
%%sql

select substr(Date, 4, 2) as month, substr(Date, 7, 4) as year, Booster_Version, Launch_Site
from SPACEXTBL
where substr(Date, 7, 4) = '2015' and
[Landing _Outcome] = 'Failure (drone ship)'
```

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

[19]:

month	year	Booster_Version	Launch_Site
01	2015	F9 v1.1 B1012	CCAFS LC-40
04	2015	F9 v1.1 B1015	CCAFS LC-40

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

- Please find the query and its result in next slide
- Used ‘substr’ function in select statement to get the new ‘Date’ column in ‘yyyy-mm-dd’ format
- Filtered date (2010-06-04, 2017-03-20) in where clauses and landing outcome that begins with ‘Success’
- Added ‘Rank’ column with ‘Rank over()’ function

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

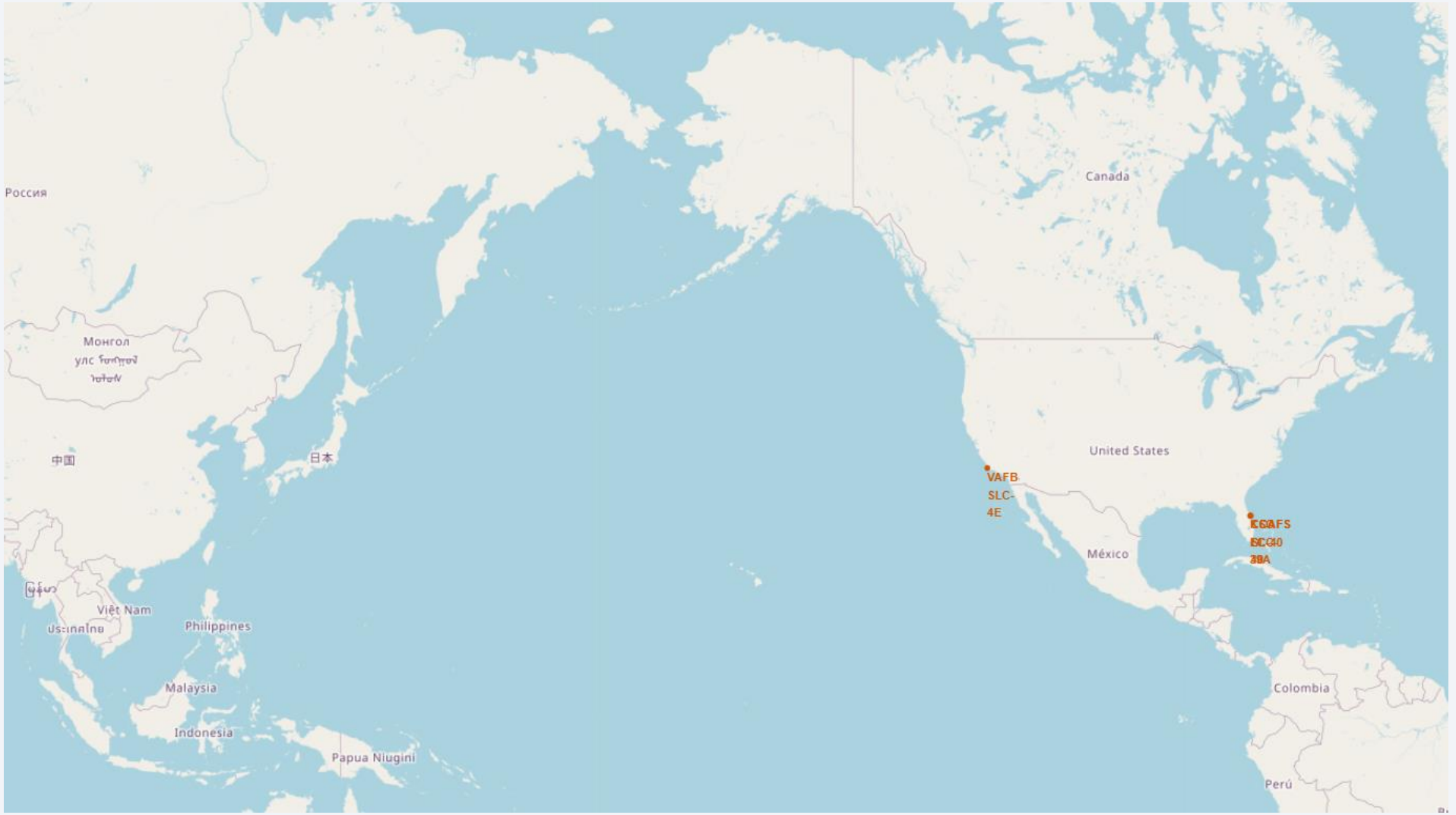
```
%%sql

select *,
    substr(Date, 7, 4) || '-' || substr(Date, 4, 2) || '-' || substr(Date, 1, 2) as Date2,
    rank () over (order by substr(Date, 7, 4) || '-' || substr(Date, 4, 2) || '-' || substr(Date, 1, 2) desc) Rank
from SPACEXTBL
where
        Date2 between '2010-06-04' and '2017-03-20' and
        [Landing _Outcome] like 'Success%'
order by Date2 desc
```

e	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome	Date2	Rank
- - 7	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)	2017-02-19	1
- - 7	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)	2017-01-14	2
- - 6	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)	2016-08-14	3
- - 6	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)	2016-07-18	4
- - 6	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)	2016-05-27	5
- - 6	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)	2016-05-06	6
- -	20:43:00	F9 FT B1021.1	CCAFS LC-	SpaceX	3136	LEO	NASA (CRS)	Success	Success (drone	2016-	7

Section 3

Launch Sites Proximities Analysis

Space X's Falcon 9 rocket launches sites location



- There are four launch sites near coastal line
- One near Santa Maria City, CA and the other threes are located near Orlando City, Miami.

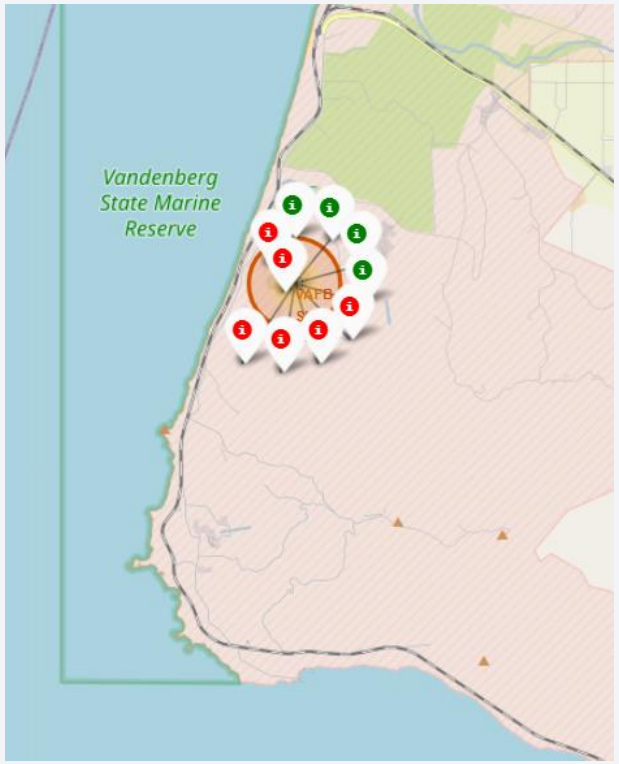
The landing success rate by each site

The landing success rate by each site

- All four sites shows different success rate. KSC LC-39A has the highest rate 77%, while CCAFS LC-40 has the lowest rate 27%.
- Among the four sites, CCAFS LC-40 has the most launch attempts (26 trial) including earlier trial (based on the lower flight number), which can affect the lower average success rate.

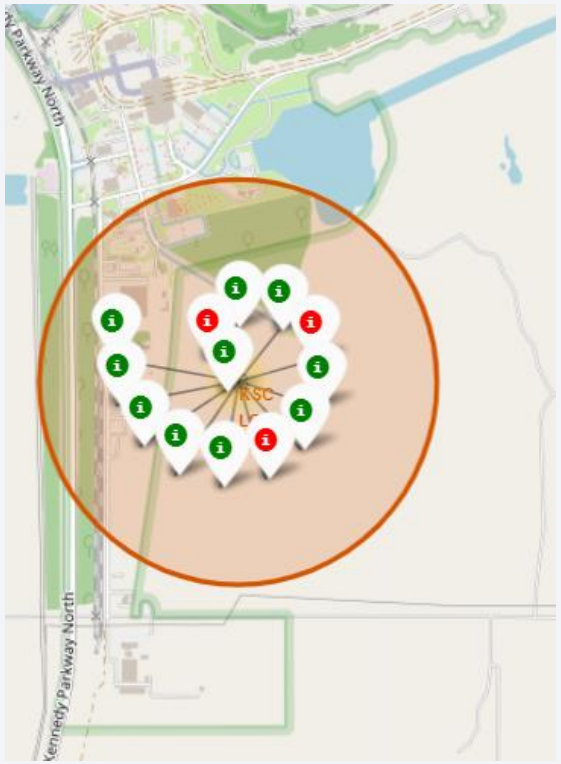
- VAFB SLC-4E

Success rate: **40.0%**



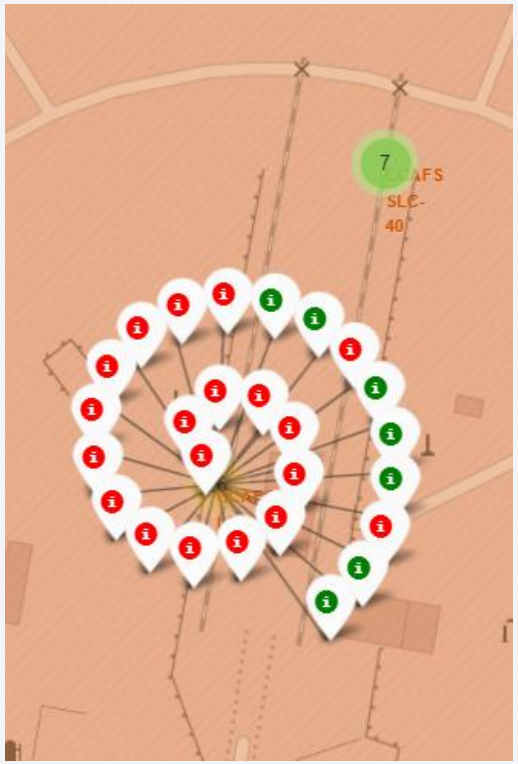
- KSC LC-39A

Success rate: **76.9%**



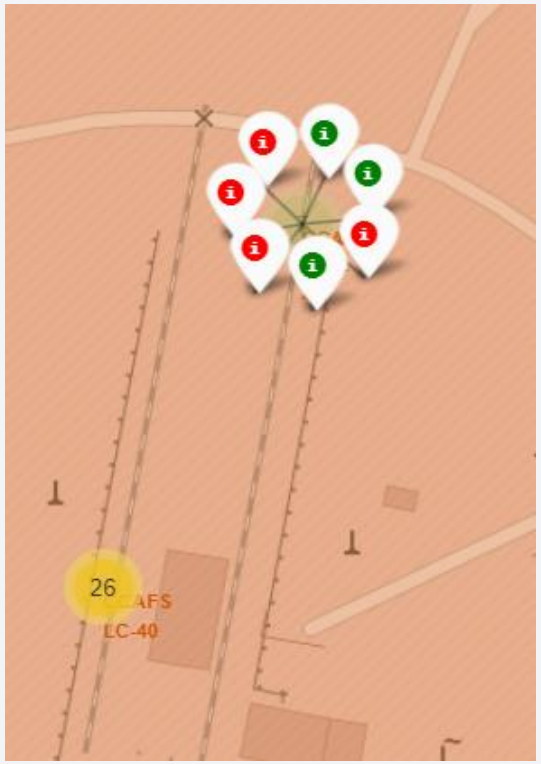
- CCAFS LC-40

Success rate: **26.9%**

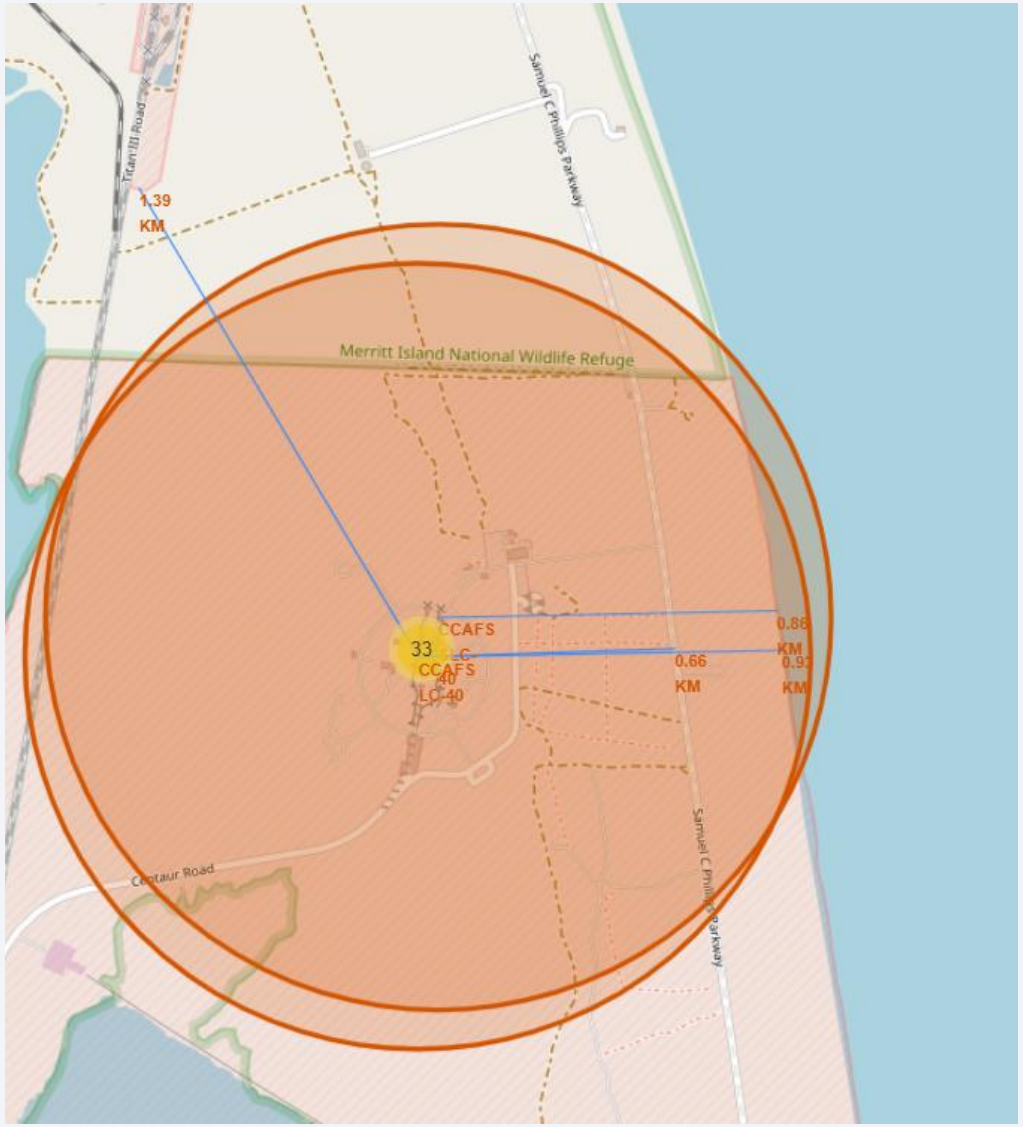


- CCAFS SLC-40

Success rate: **42.9%**

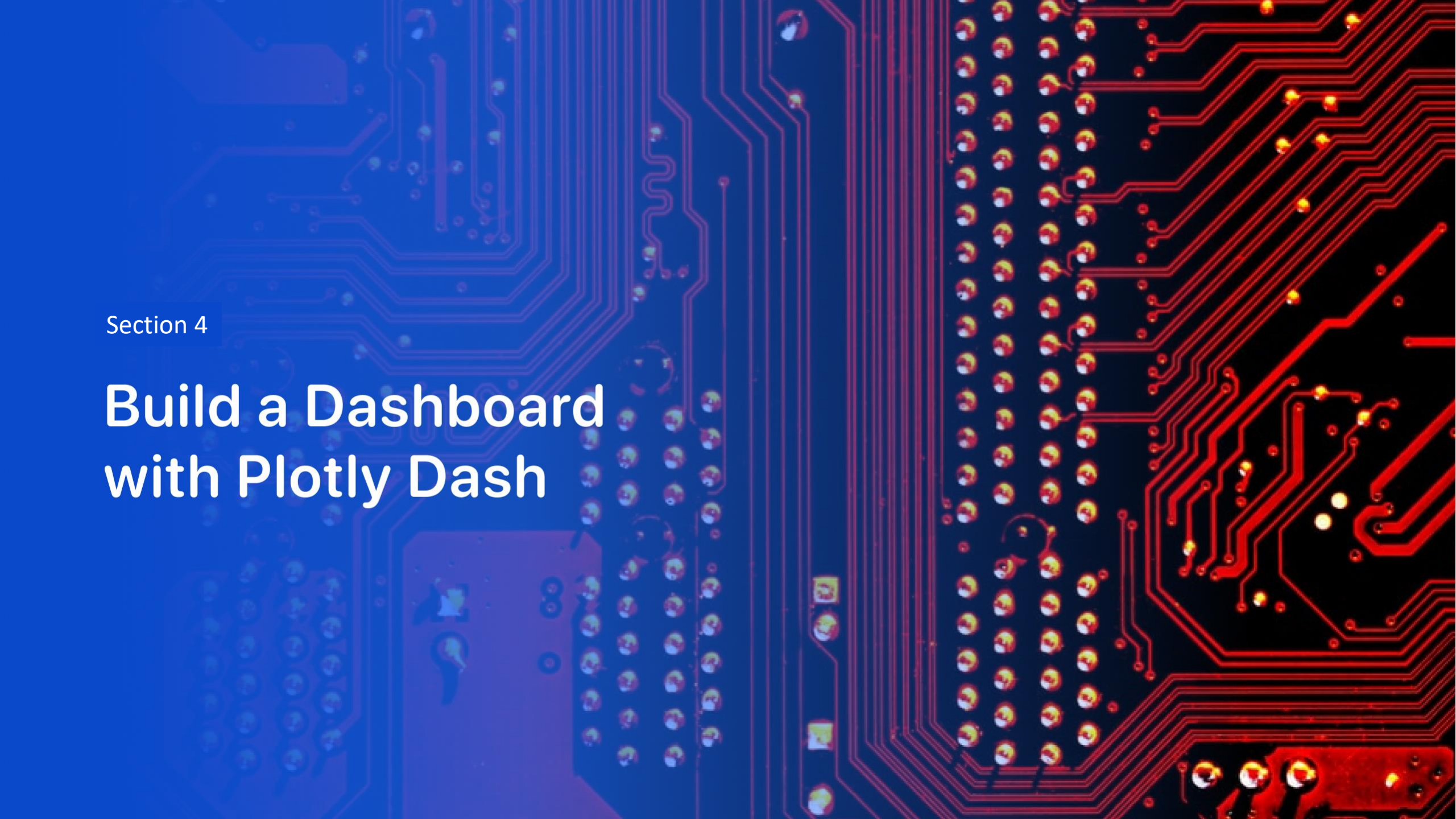


CCAFS LC-40 and its proximities to the nearest highway, coastline and railway

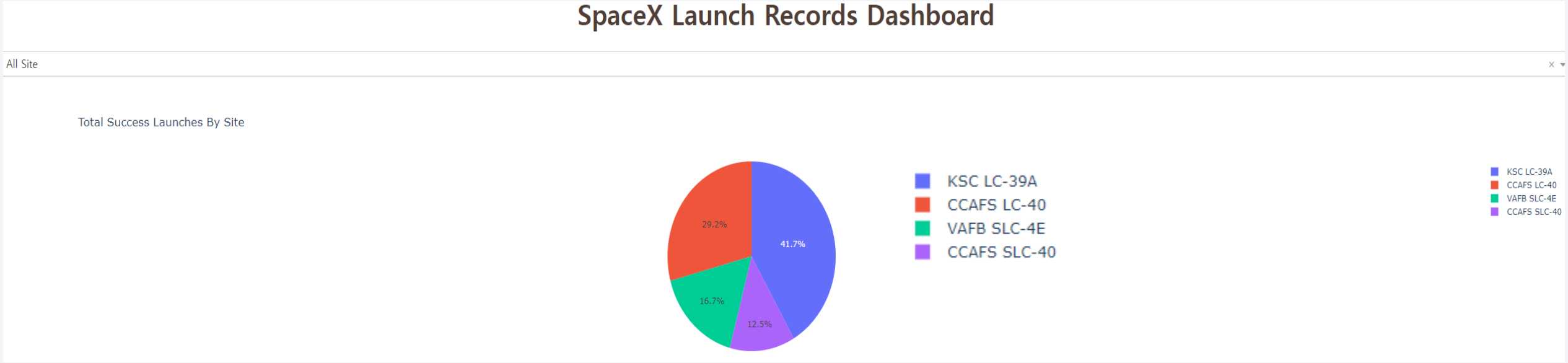


- In overall, it looks that CCAFS LC-40 has the very good proximities of transportation and coastline.

  I. The distance between CCAFS LC-40 and the nearest highway, Samuel C Phillips Pkwy is 0.66km
  II. The distance between CCAFS LC-40 and the nearest coastline is 0.93km
  III. The distance between CCAFS LC-40 and the nearest railway NASA railroad is ~1.4km


Section 4
Build a Dashboard with Plotly Dash

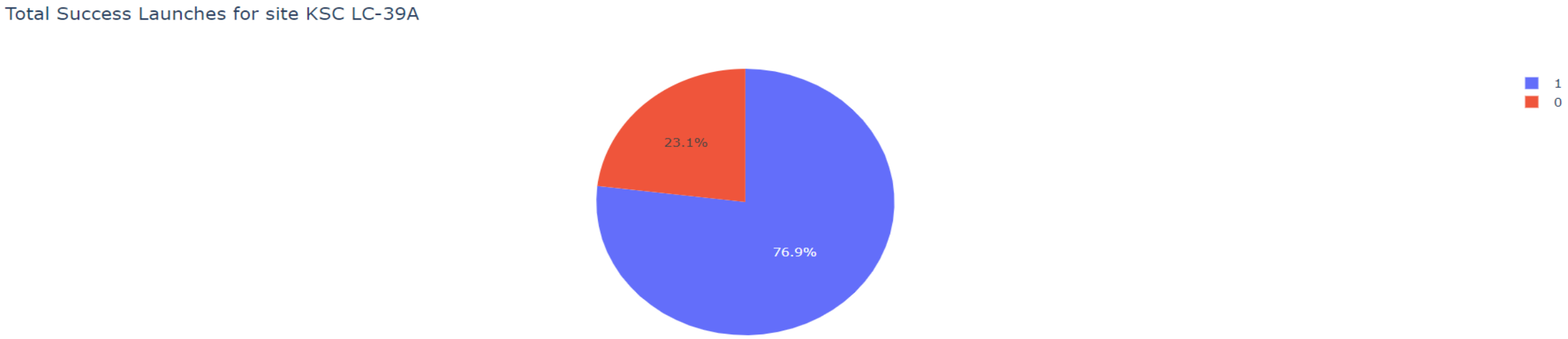
The launch success count by all sites



- **KSC LC-39A** has **the most success case** followed by CCAFS LC-40
- KSC LC-39A has ten success launches, while CCAFS SLC-40 has three success launches.

The launch success rate for KSC LC-39A

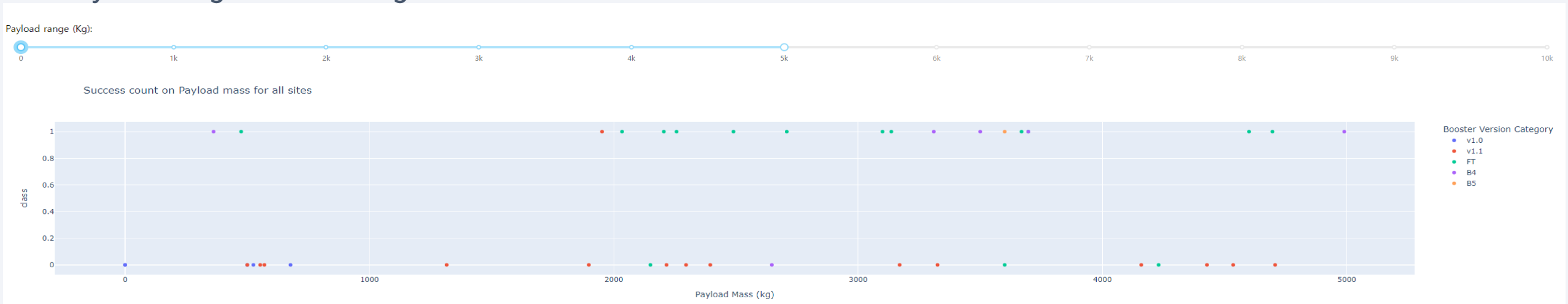
KSC LC-39A



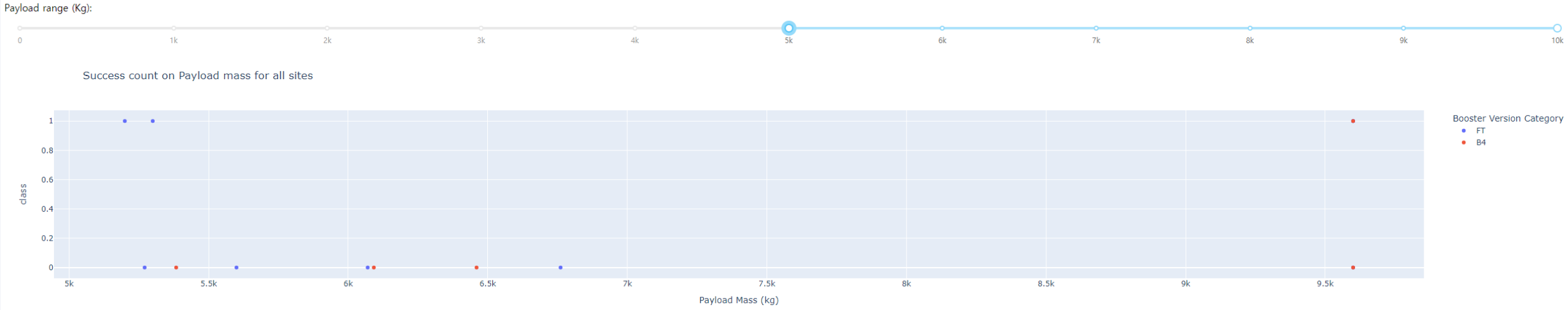
- **KSC LC-39A** site has **the highest** launch success rate with **76.9%**
- There has been 13 launch attempts in KSC LC-39A and 10 attempts were succeeded, leading to 76.9% success rate.

Payload vs. Launch Outcome

1. Payload range **below** 5K Kg selected:

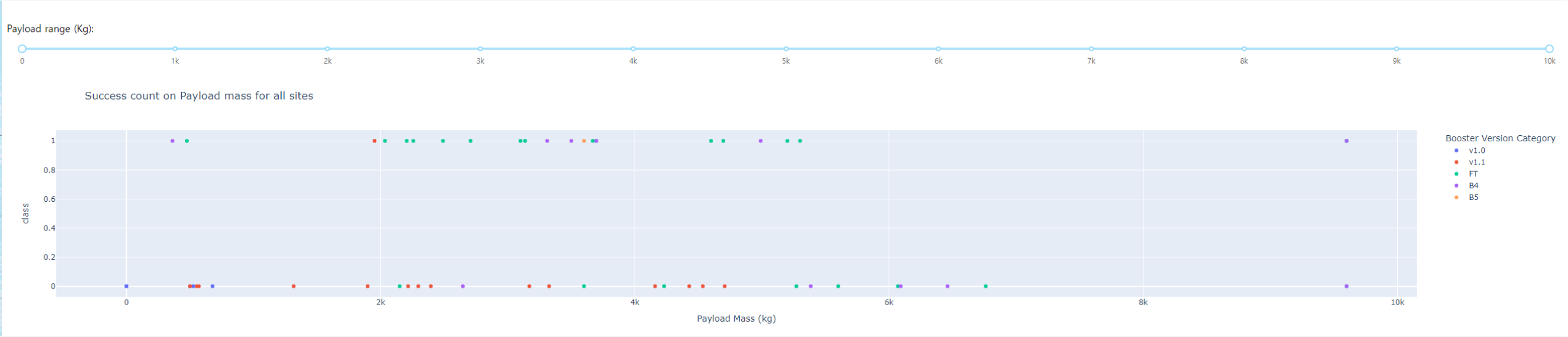


2. Payload range **above** 5K Kg selected:



Payload vs. Launch Outcome

3. Payload range (all)



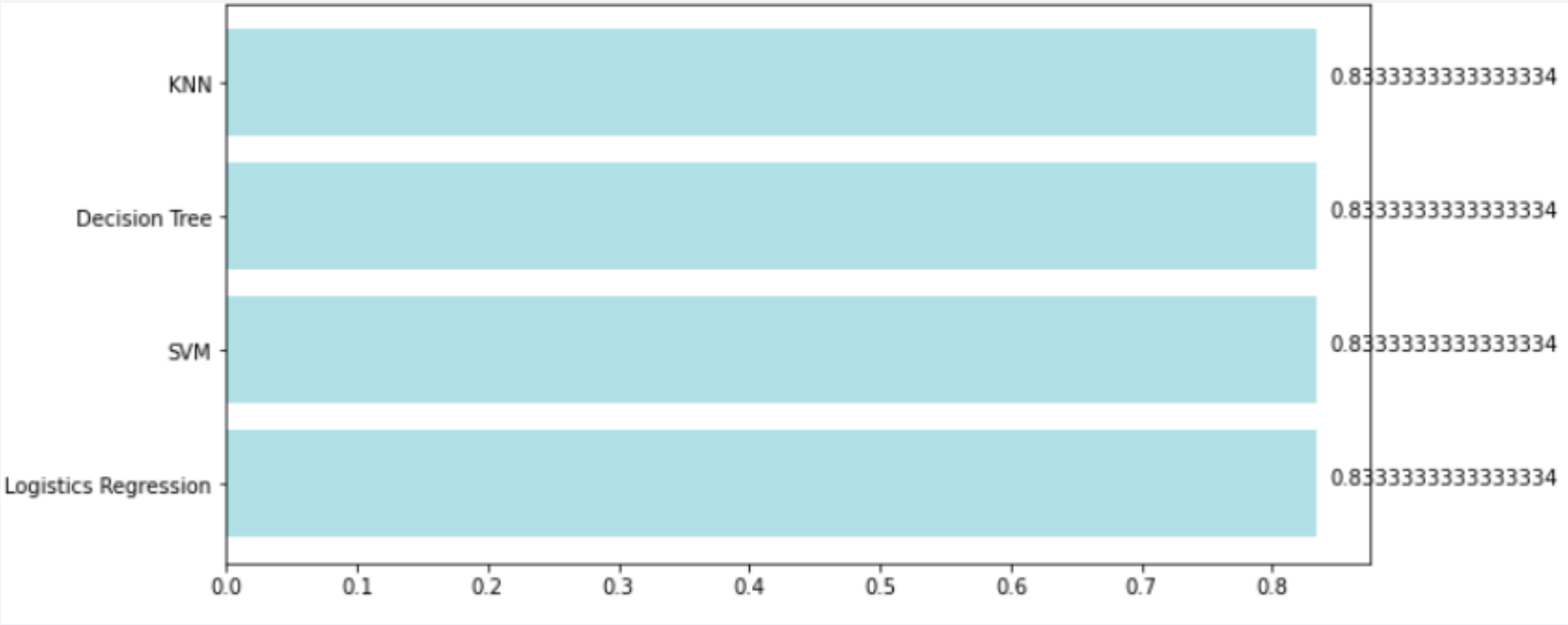
- Booster Version **FT** has the **largest success rate**, while v1.1 has the lowest success rate.
- There has been the most attempts in **payload mass range 2K – 4K**; This range also has the largest success count.
- The middle level of payload mass looks related to high success rate.

Section 5

Predictive Analysis (Classification)

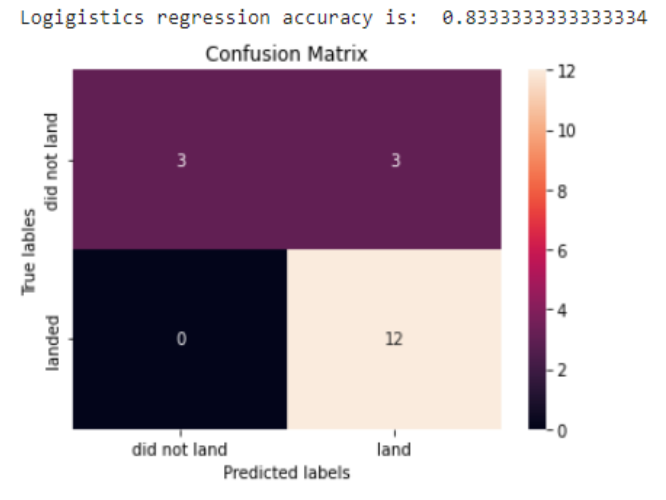
Classification Accuracy

- When it comes to accuracy score between y test and y prediction, it turns out that all four models (Logistics Regression, SVM, Decision Tree, KNN) has **the same accuracy** with **83.3%**
- Since our **test data is small set data**, that lead to the same model assessment result.
- Please find below bar chart.

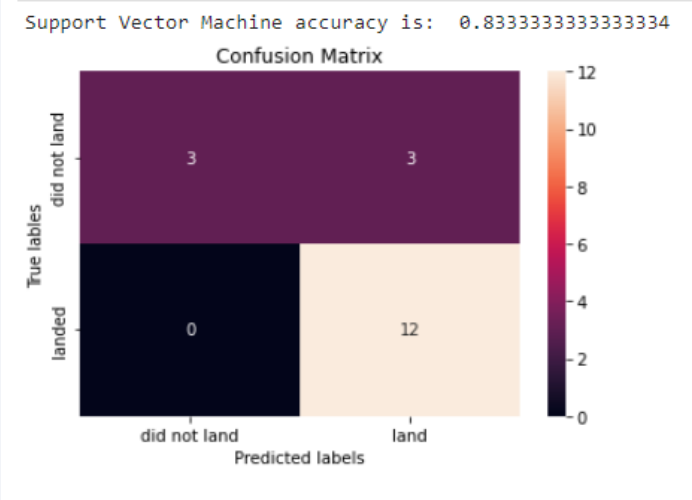


Confusion Matrix

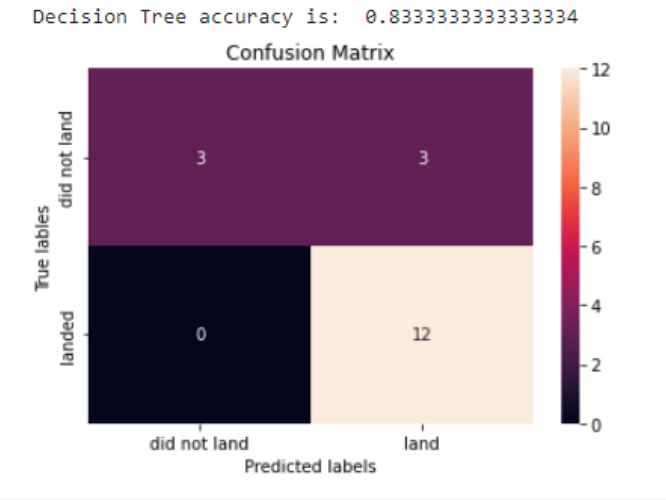
Logistics Regression



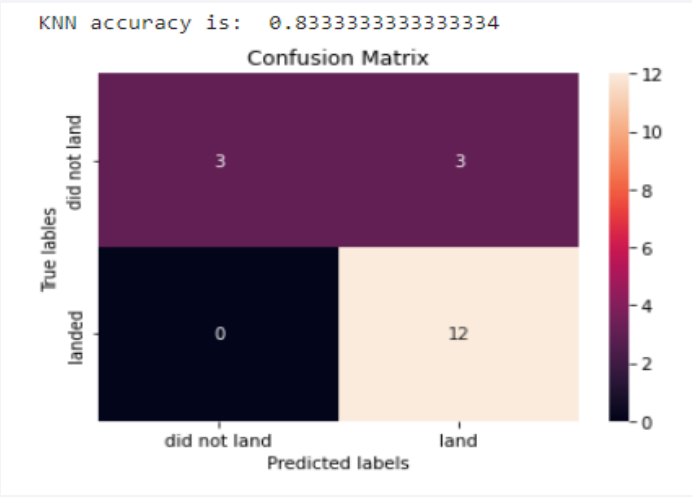
SVM



Decision Tree



KNN



- All four models showed **the same confusion matrix result** when it comes to scoring for y test and y prediction
- The accuracy score is the same for all models, **83.3%**
- With 18 test set, 15 prediction was correct, leading to 83.3%
  : The true positive #: 12
  : The true negative #: 3)

Conclusions

- Our model predicted the overall landing success rate as 83%, which means that when we had a 18 cases, it predicted 15 cases as a success and 3 cases as a failure.
- From the perspective of accuracy score, it is hard to select which models is the best model for prediction since all four classification models showed the same score.
- Out of 18 test set, 15 prediction was correct, leading to 83.3% accuracy for all models, which is good level of prediction accuracy.
- The small set of test data might affect this same result among all models. Once we secure more launch related data in the future, we can go through the same process and access each model with more details.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!