IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Do Minh Quoc
Sep 05 2023

Outline

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

Executive Summary

- The ultimate purpose of this project is creating a model which can predict the outcomes of launching a rocket : in the process of finding the best model, data would be collected throught API and website. After that, the most relevant data would be used to reveal the relation between out outcome and contributiong factors. There will be some visualization to see this more clearly

- Result: The contributor to the success rate of launch is : Launch site, the Orbit that rocket will launch to, the payload and the year of launching

Introduction

- In the modern world, space age is for everyone. SpaceY is a startup to help space travel is affordable. While other company could take 165 milions dollar for every launch of a rocket. SpaceY's mission is to reduce half of the cost to less then 65 milion dollars by reusing the first phase of the rocket just like SpaceX

- By using the collected data, SpaceY will offer the best cost and the rate of reuse parts of rocket to save cost and reduce the price the most

Section 1

Methodology

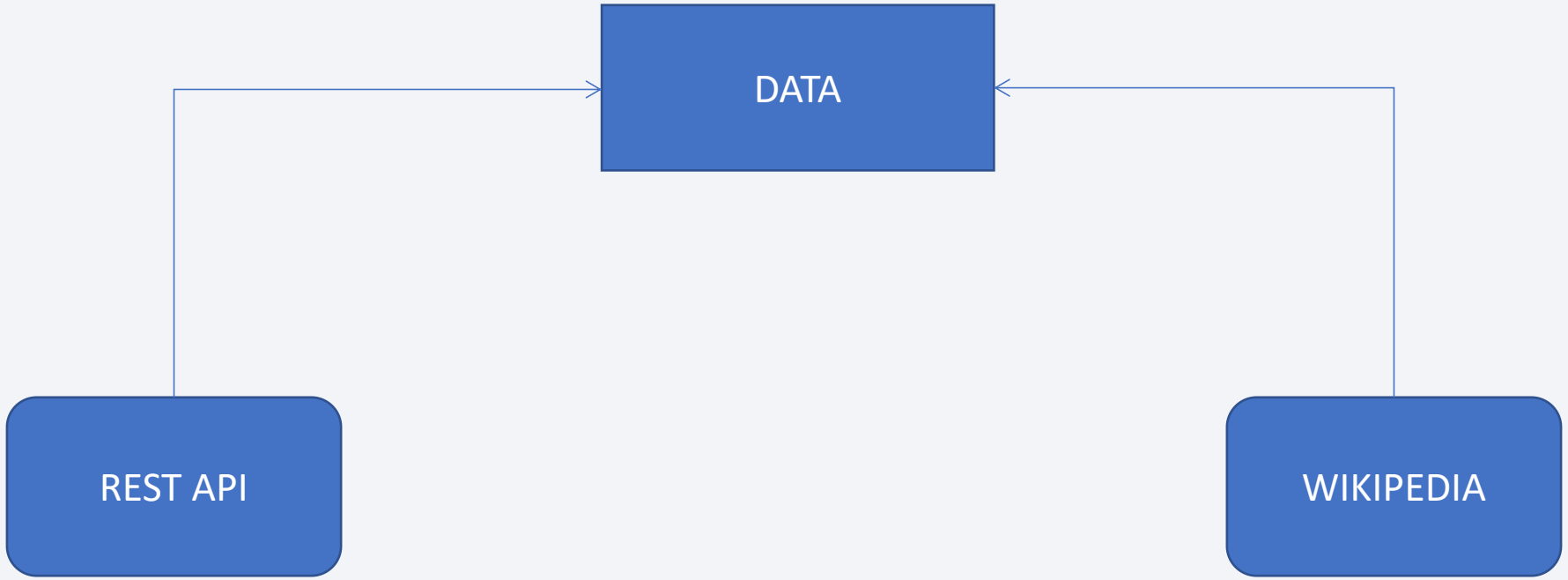
Methodology

Executive Summary

- Data were collected thourgh SPACEX REST API and WIKIPEDIA website
- Data is processed to filter on Falcon 9 data, replace the missing values, replace ID code of values in data by name
- Exploratory data analysis (EDA) to see the factors contribute to the launching rate success rate
- Visial data on worldmap to see insight of data and more factors affects the models
- Data will used to train some model classification in order to choose the best model for the future predictions.

Data Collection

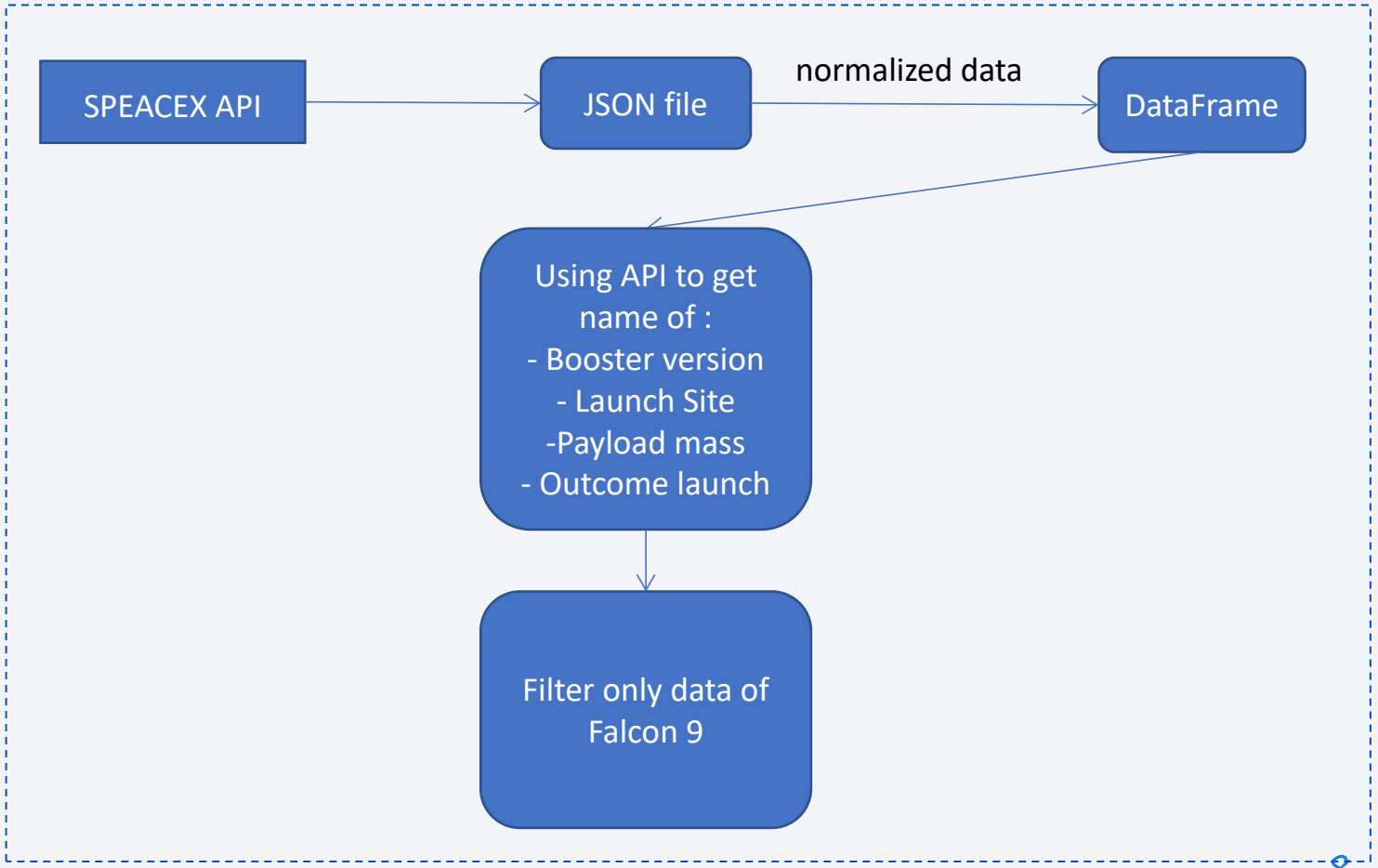
- Data collected by SpaceX API and WIKIPEDIA



Data Collection – SpaceX API

- GitHub URL :

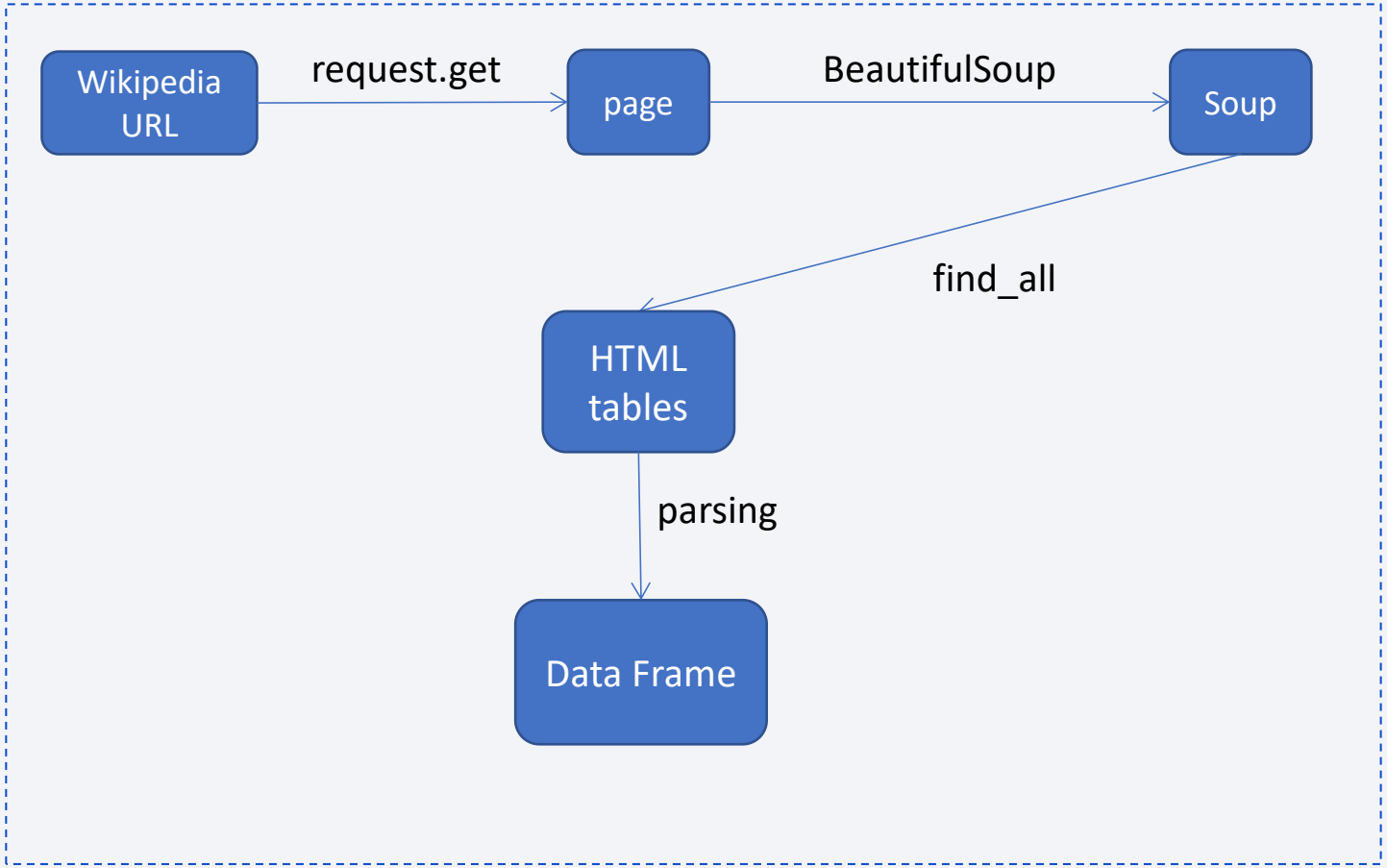
https://github.com/doquuoc/SPACEY---DATA-CAPSTONE-PROJECT/blob/1c6684af91dedbbf6f9e9f68472b16f47785d502/1-%20Data%20collecting.ipynb



Data Collection - Web Scraping

GitHub URL :

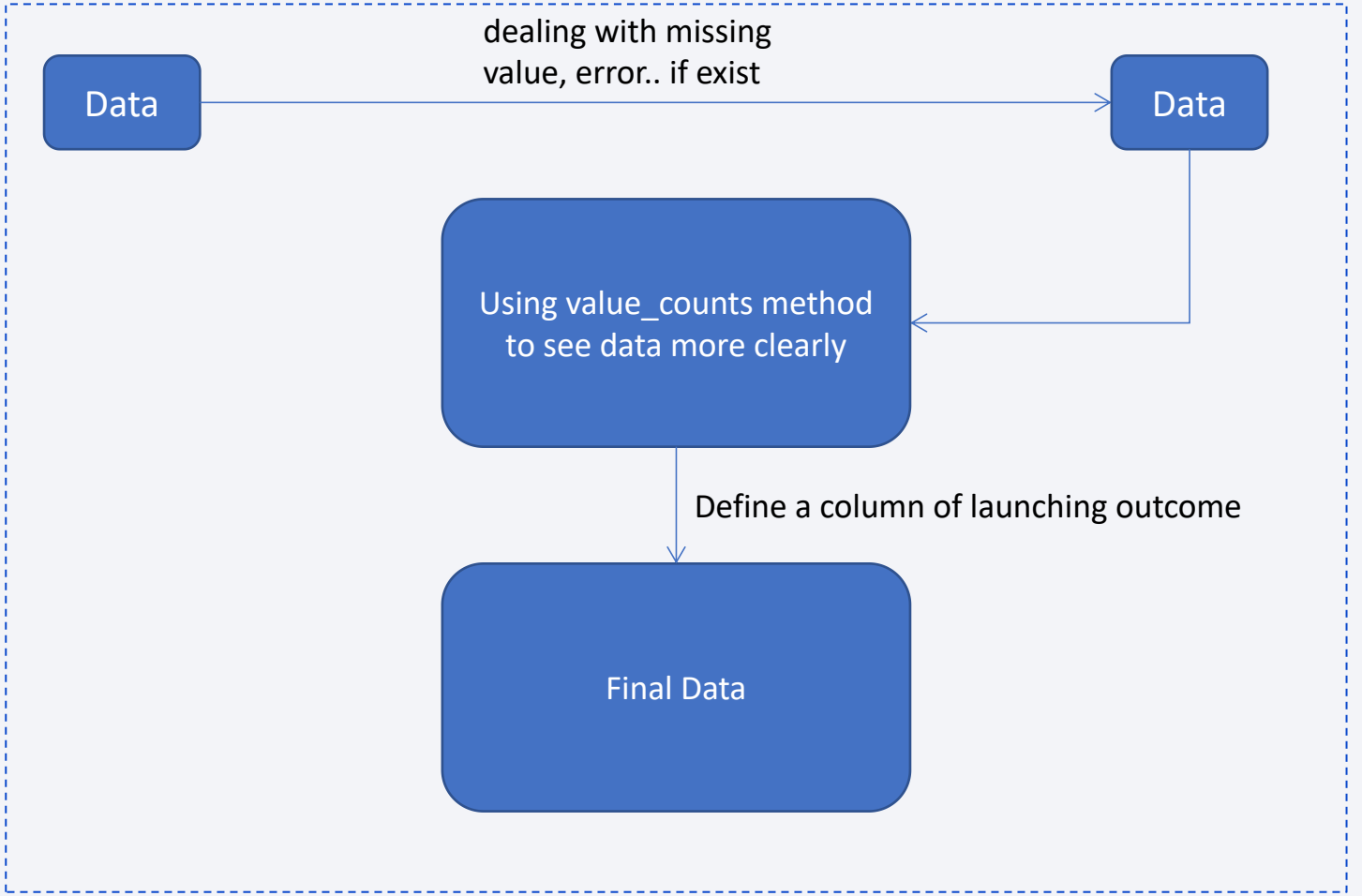
https://github.com/doquuoc/SPACEY---DATA-CAPSTONE-PROJECT/blob/1c6684af91ddbbf6f9e9f68472b16f47785d502/2-%20Data%20web%20scraping.ipynb



Data Wrangling

GitHub URL :

https://github.com/doquuoc/SPACEY---DATA-CAPSTONE-PROJECT/blob/1c6684af91dedbbf6f9e9f68472b16f47785d502/3-%20Data%20Wrangling.ipynb



EDA with Data Visualization

- Chart used in this project:

1.Scatter chart to represent the relation between Flight Number and Payload mass and Launch Site

2. Bar chart to represent the successful rate of every orbit type

3. Line chart to represent the successful rate over years

- GitHub URL :

https://github.com/doquuoc/SPACEY---DATA-CAPSTONE-PROJECT/blob/1c6684af91dedbbf6f9e9f68472b16f47785d502/5%20-%20Data%20visualization%20and%20One%20hot%20conder.ipynb

EDA with SQL

Some of bullet point data using SQL:

- Select distinct launch site
- Calculate the average payload mass
- Show the first day of successful launch on a groud pad
- Show the the successful launch drone ship which have the payload mass form 4000 - 6000
- Count the total number of failure and successful launch
- Show the booster version with maximum payload
- List the records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 201

GitHub URL :

- https://github.com/doquuoc/SPACEY---DATA-CAPSTONE-PROJECT/blob/1c6684af91dedbbf6f9e9f68472b16f47785d502/4%20-%20Query%20data%20with%20SQL.ipynb

Build an Interactive Map with Folium

A folium map with:

- Map show the position of 3 Launch site with mark of every Launch site for a clear visibility.
- Marker cluster to so which launch is fail or success in every launch site to have an overview of successful rate of every launch site
- Arrow to show the distance between Launch site and Central city or Railway or Coast to discover the successful rate, number of launch of the is affected by those distances

GitHub URL :

https://github.com/doquuoc/SPACEY---DATA-CAPSTONE-PROJECT/blob/1c6684af91dedbbf6f9e9f68472b16f47785d502/6%20-%20Folium%20visualization.ipynb

Build a Dashboard with Plotly Dash

Dashboard summary:

- A pie chart to show the successful rate or failure rate of every launch site or percentage of successful rate of each launch site in total successful launch
- A range slider to select the payload mass range
- A scatter chart show the relation of playload mass and launch outcome and booster version

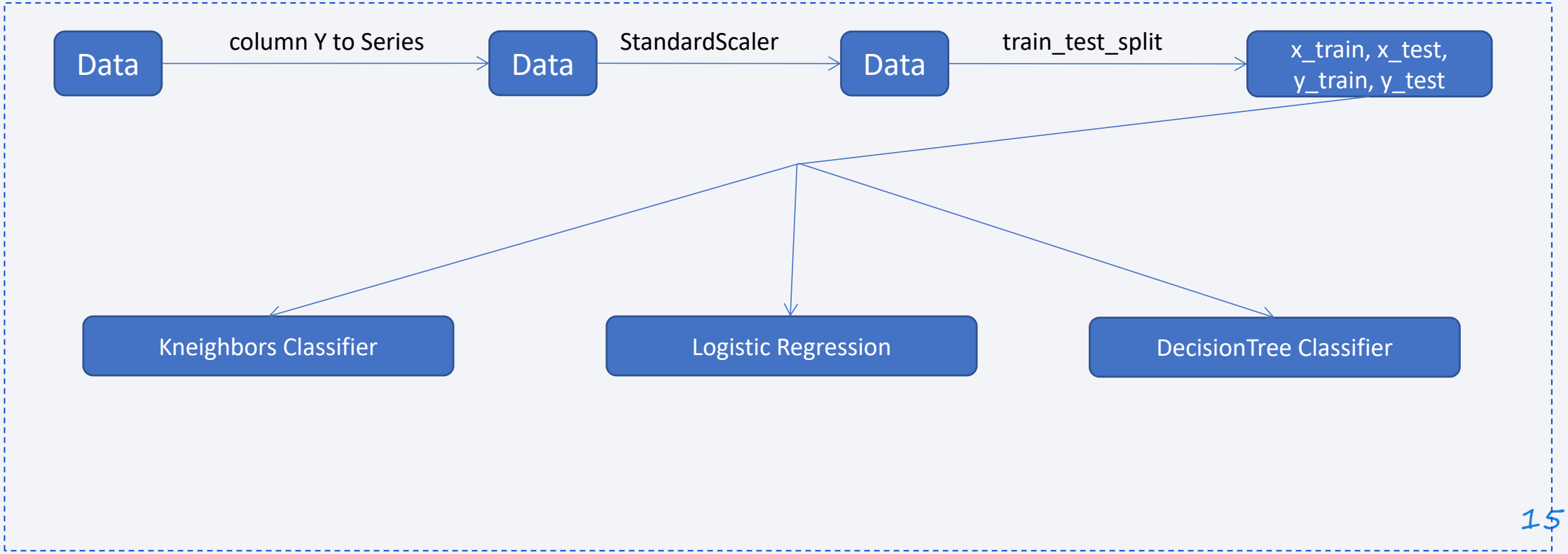
GitHub URL:

https://github.com/doquuoc/SPACEY---DATA-CAPSTONE-PROJECT/blob/1c6684af91dedbbf6f9e9f68472b16f47785d502/7%20-%20Dash%20app%20-%20SpaceX.py

Predictive Analysis (Classification)

GitHub URL:

https://github.com/doquuoc/SPACEY---DATA-CAPSTONE-PROJECT/blob/1c6684af91dedbbf6f9e9f68472b16f47785d502/8%20-%20Mechine%20Learning%20.ipynb



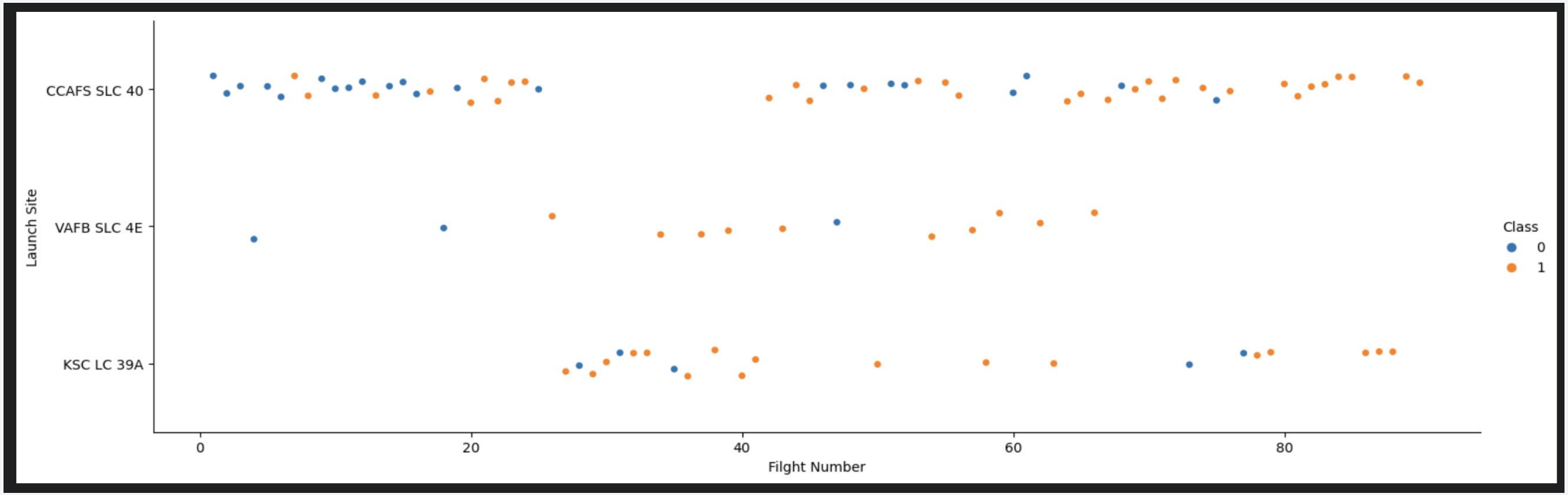
Results

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

Section 2

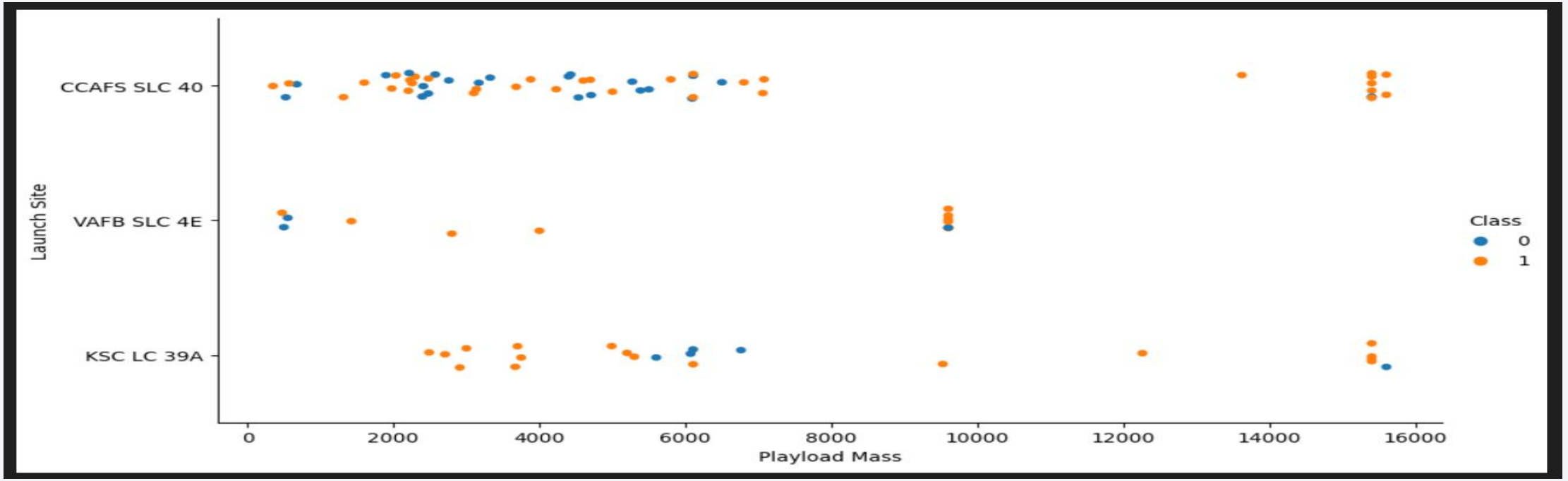
Insights drawn from EDA

Flight Number vs. Launch Site



- As you can see, as Flight Number increase, all of 3 Launch site increase the success rate
- But overall, the VAFB SLC 4E and KSC LC39A have more success rate of launching than the CAFS SLC 40

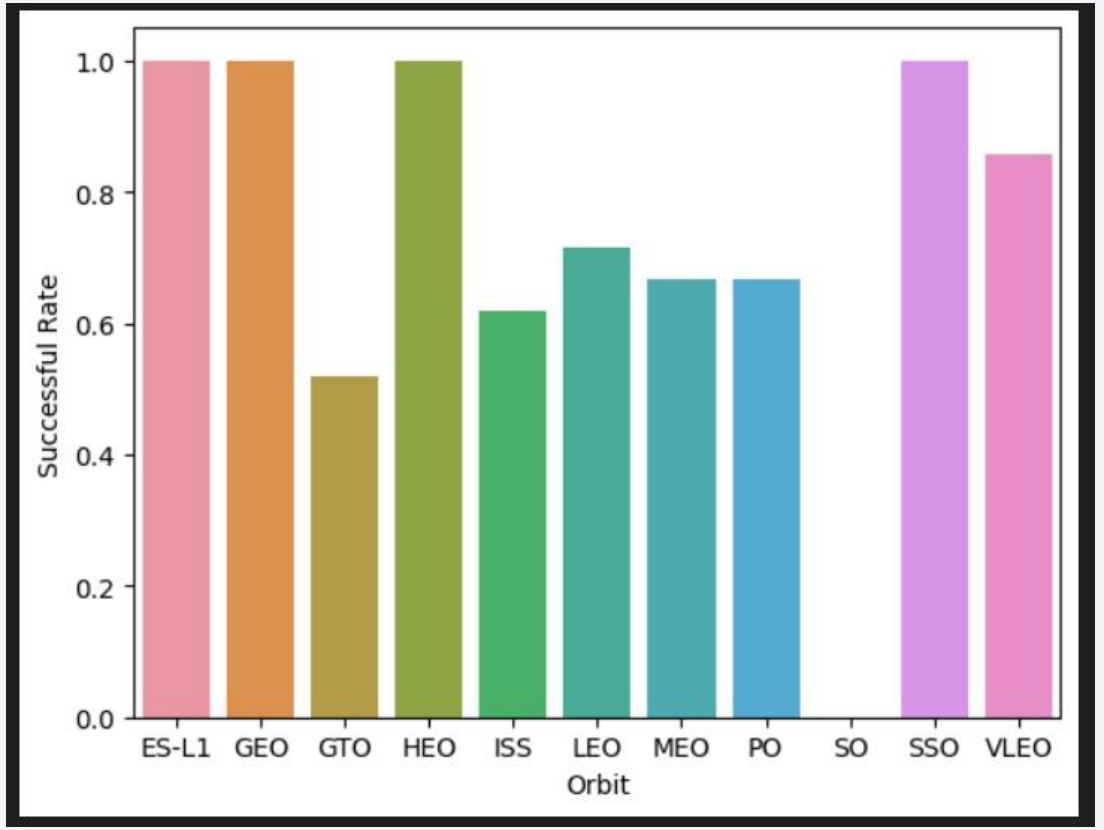
Payload vs. Launch Site



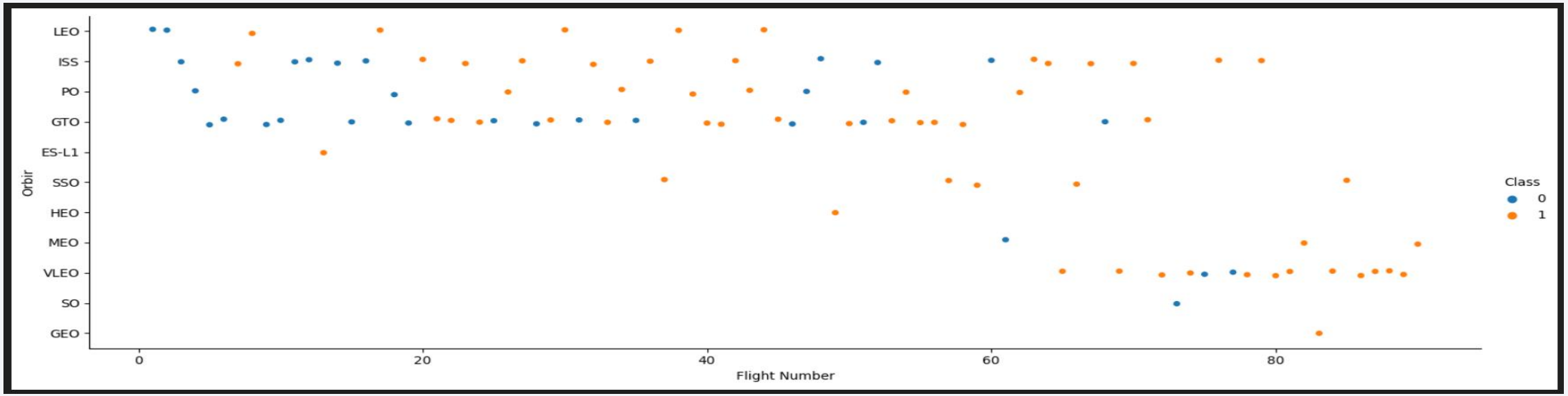
- As you can see, the **VAFE SLC 4E** launch site only for launching rocket under 10,000kg payload while the others could possibly carry payload about 16,000kg
- Under 8,000 kg payload. **KSC LC 39A** have more successful rate than **VAFB SLC 4E** and **CCAFS SLC 40** launch site

Success Rate vs. Orbit Type

- Launching rocket to ES-L1 , GEO, HEO and SSO get 100% of successful rate while the GTP get the lowest however SSO get 5 launch times but the others only 1 times.
- GTO, ISS and VLEO get the most launch times : 27, 21 ,14 respectively so they get the lowest successul rate

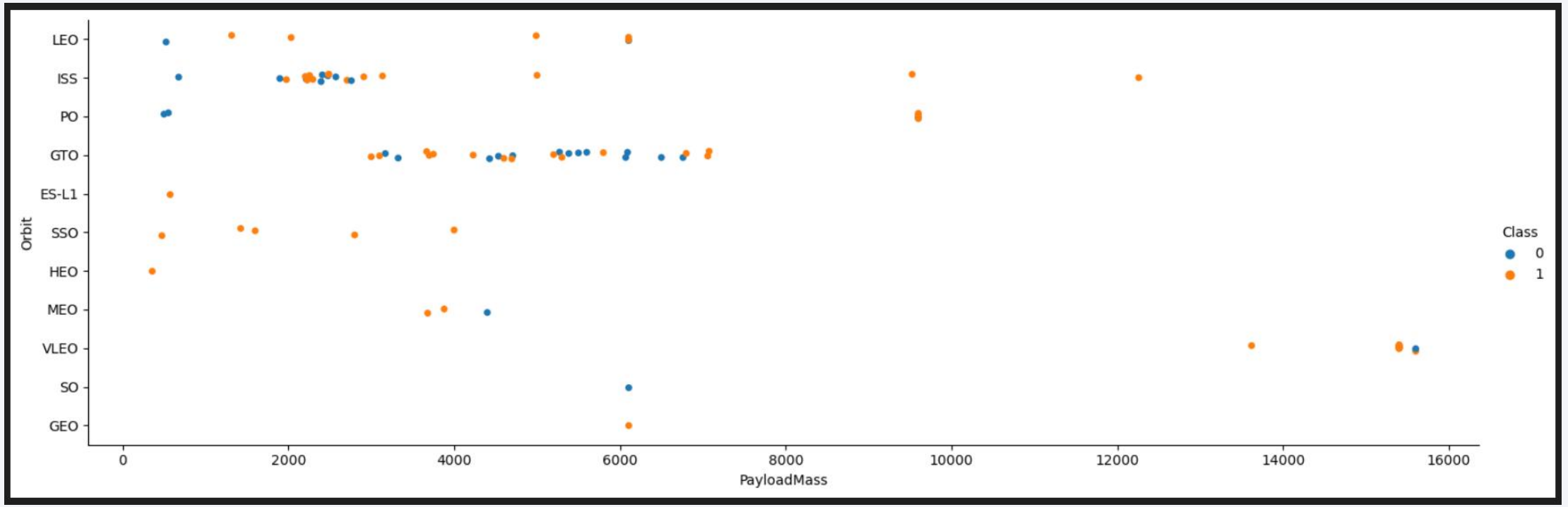


Flight Number vs. Orbit Type



- The chart show that the the higher the orbit, the less lauching times, as the Filght number increase, the company have a tendency to launch rocket into the higher orbit than before
- The higher orbit, the more higher rate of success.

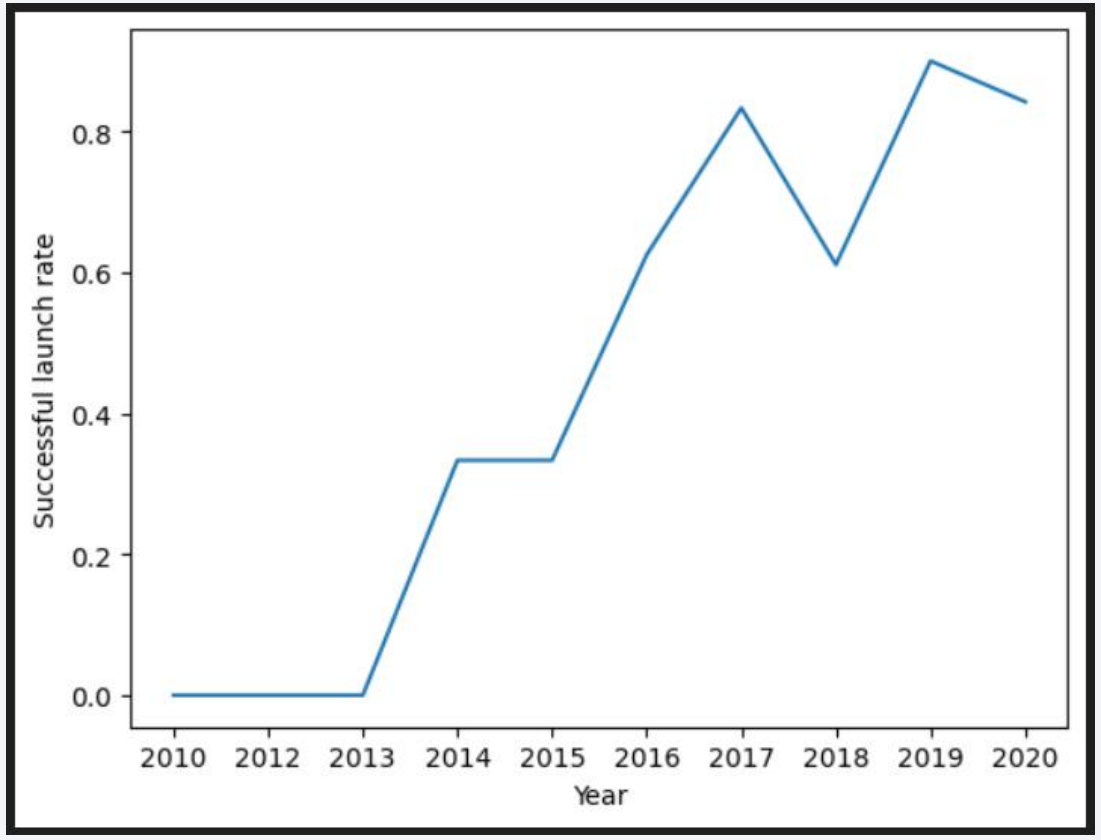
Payload vs. Orbit Type



- LEO , ISS and PO are orbit target of launching wide range of Payload rocket 500 – 12,000
- VLEO usually carry a heavy Payload about 16,000kg

Launch Success Yearly Trend

- As you can see, overtime the rate of success increaser because of the more advance of technology



All Launch Site Names

There are total 4 launch site:

- CCAFS LC-40
- VAFB SLC-40
- KSC LC-39A
- CCAFS SLC-40

```
%sql SELECT DISTINCT("Launch_Site") FROM SPACEXTABLE;
 ✓ 0.0s

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

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

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE "CCA%" LIMIT 5;
```

Python

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

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

- Example of 5 record where launch site begin with CCA
- As you can see SpaceX and Nasa attempt to land was fail, after that, NASA have no attempt to land back the ground

Total Payload Mass

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

```
%sql SELECT "Customer",SUM("PAYLOAD_MASS__KG_") AS "Total payload mass(kg)" FROM SPACEXTABLE WHERE "Customer" = "NASA (CRS)";
```

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

Customer	Total payload mass(kg)
NASA (CRS)	45596

- Payload max carried by boosters launch by NASA(CRS) was 45,596kg

Average Payload Mass by F9 v1.1

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

```
%sql SELECT "Booster_Version",AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = "F9 v1.1";
```

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

Booster_Version	AVG("PAYLOAD_MASS__KG_")
F9 v1.1	2928.4

Average payload mass carried by booster version F9 1.1 was 2,928 kg

First Successful Ground Landing Date

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

Hint:Use min function

```
%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (ground pad)";
```

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

MIN(Date)
2015-12-22

As record, the first succesful landing outcome in ground pad was 22/12/2015

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", "PAYLOAD_MASS__KG_", "Landing_Outcome" FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (drone ship)" AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000;
```

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

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

This is the list of the booster which have success in landing drone ship and having payload mass from 4000 to 6000kg:

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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") FROM SPACEXTABLE GROUP BY "Mission_Outcome";
```

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

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

In total, there are 100 missions were succesful and 1 mission was fail.

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 SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
```

```
 * 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
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

These are list of booster versions which carried the maximum payload

2015 Launch Records

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

```
%sql SELECT SUBSTR("Date",6,2) AS Month, "Booster_Version", "Launch_Site", "Landing_Outcome" FROM SPACEXTABLE WHERE "Landing_Outcome" = "Failure (drone ship)" AND substring("Date",1,4) = "2015";
```

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

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

There were 2 failure landing in drone ship in 2015 which booster version is F9 v1.1 B1012 and F9 v1.1 B1015

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 "Landing_Outcome", COUNT("Landing_Outcome") \
FROM SPACEXTABLE \
WHERE substring("Date",1,4) <= "2017" AND substring("Date",1,4) > "2010" \
GROUP BY "Landing_Outcome" ORDER BY COUNT("Landing_Outcome") DESC ;
```

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

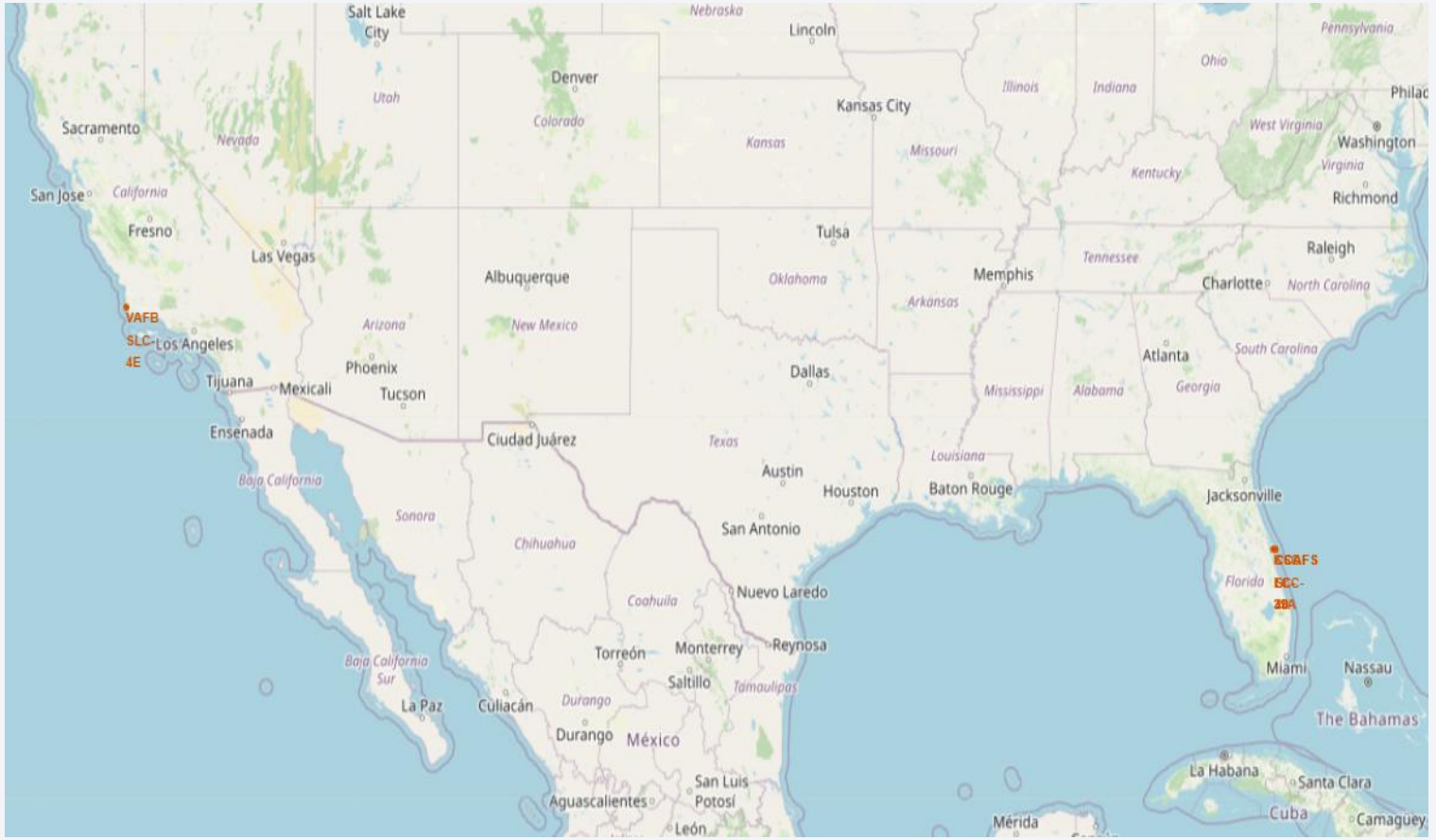
Landing_Outcome	COUNT("Landing_Outcome")
Success (drone ship)	12
No attempt	12
Success (ground pad)	8
Failure (drone ship)	5
Controlled (ocean)	4
Uncontrolled (ocean)	2
Precluded (drone ship)	1

As you can see the most success is landing on drone ship, next is no landing, and than ground pand.
The failure rate of landing drone ship also relatively hight: 5 failure
Next is landing on ocean with 4 landing were controlled and 2 landing were uncontrolled

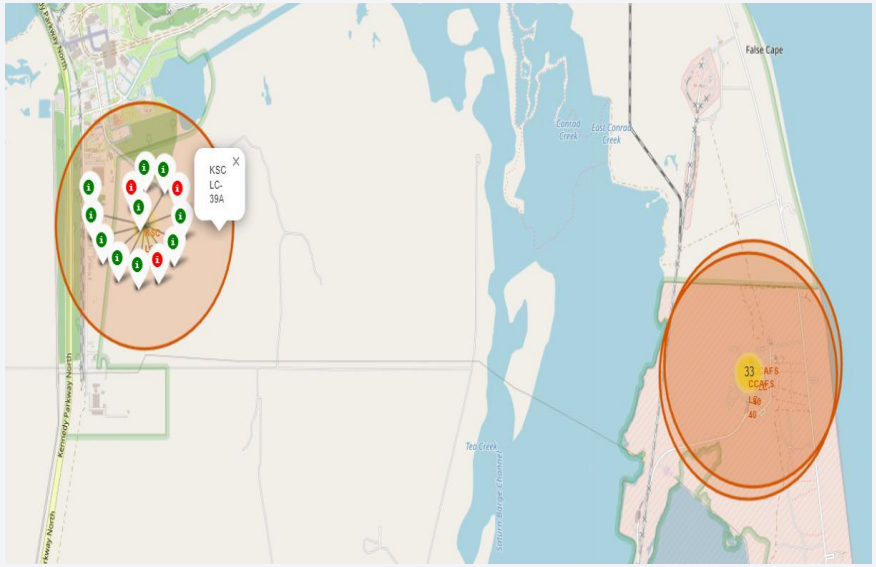

Section 3
Launch Sites
Proximities Analysis

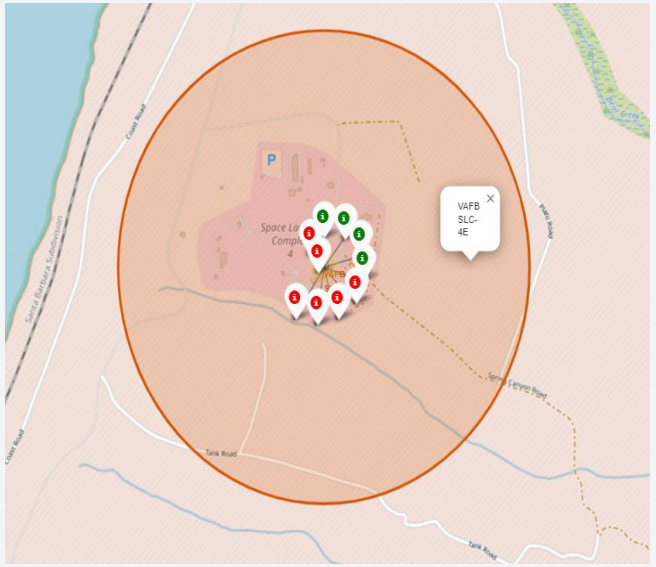
Launch Site Map

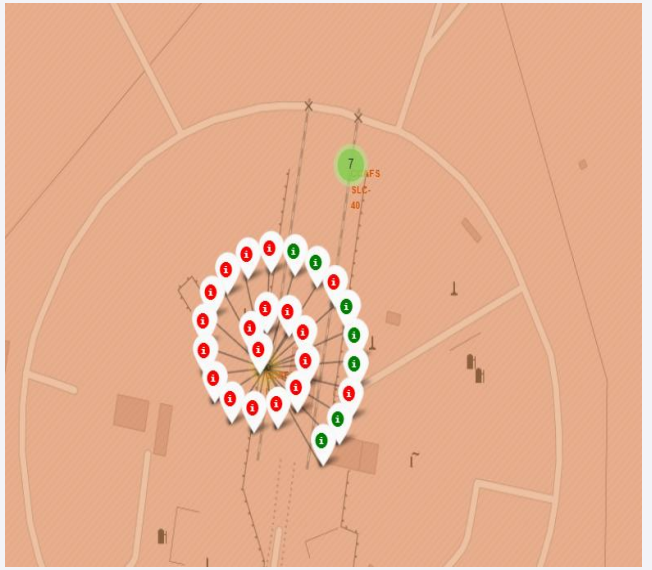
- As you can see on the map, there are 4 total Launch site: 1 on the West and 3 on the East of America
- All of the launch sites also locate near by the coast for better control of landing in ocean and ship



Failure and Success landing outcome of each Launch site



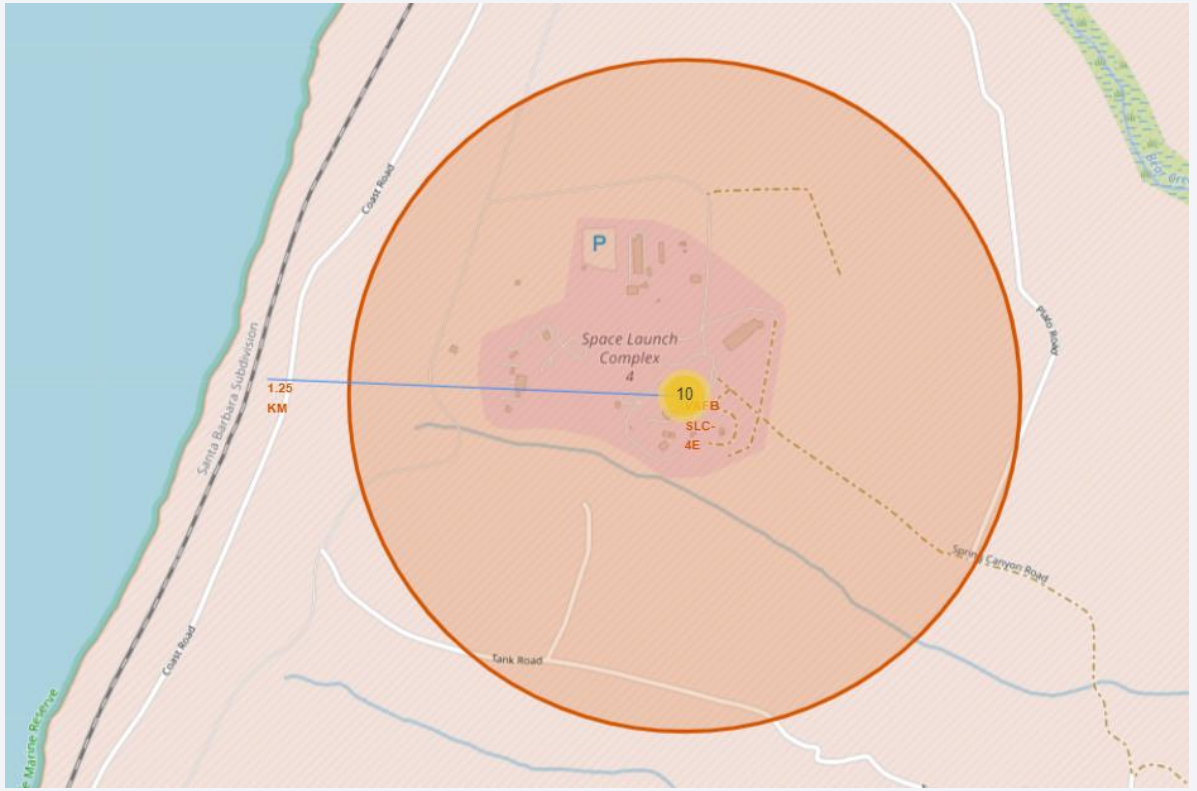


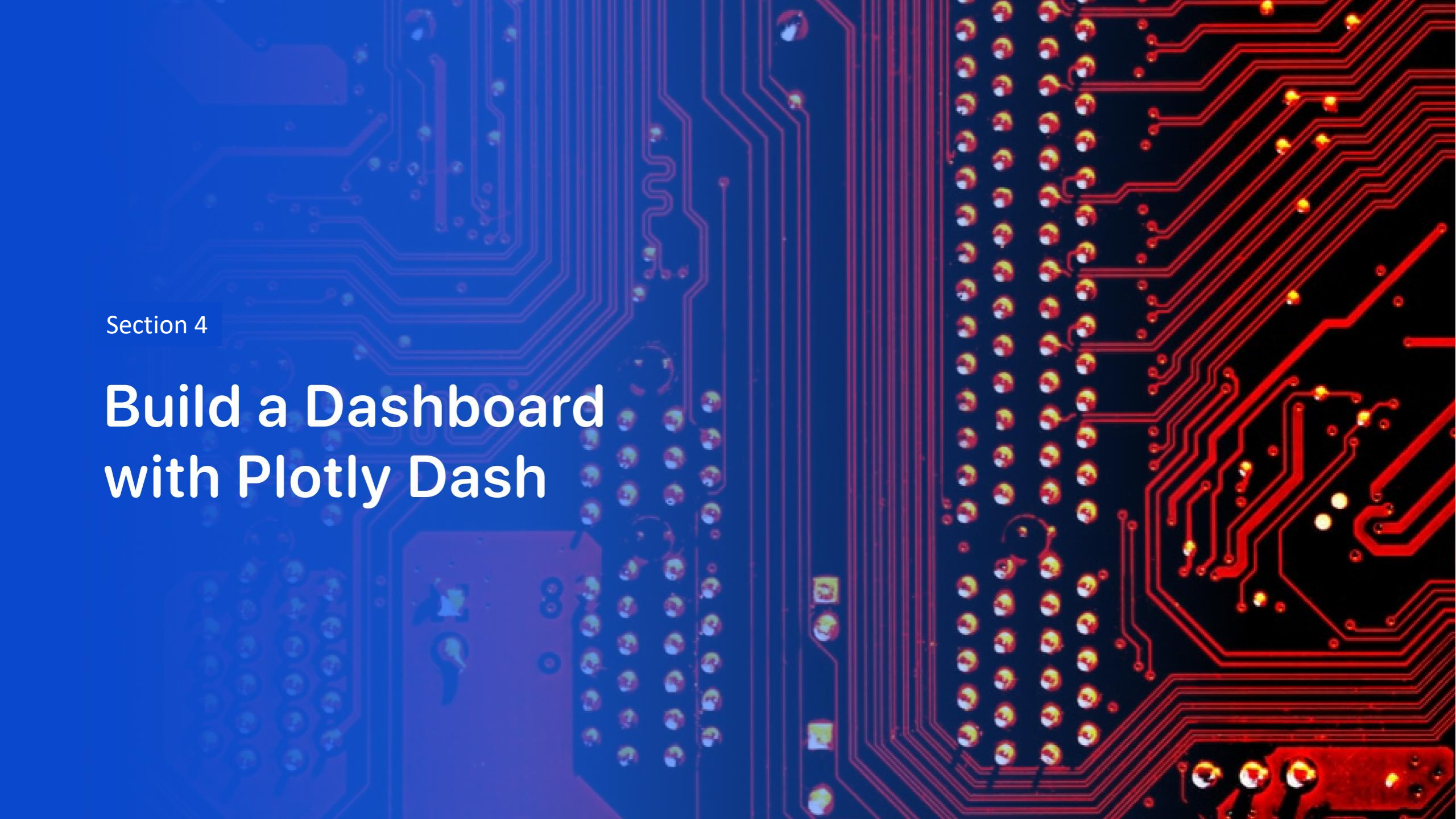


The map show the outcome launching of every launching site:
- The VAFB-SLC-4E have 10 launch times but only have 4 success
- CCAFS LC-40 launch 26 times but got 7 times success
- CCAFS SLC-40 have 3 successful launch out of 7 launch times
- KSC LC-39A got the best succes rate, 10 successful launch out of 13 launch time.

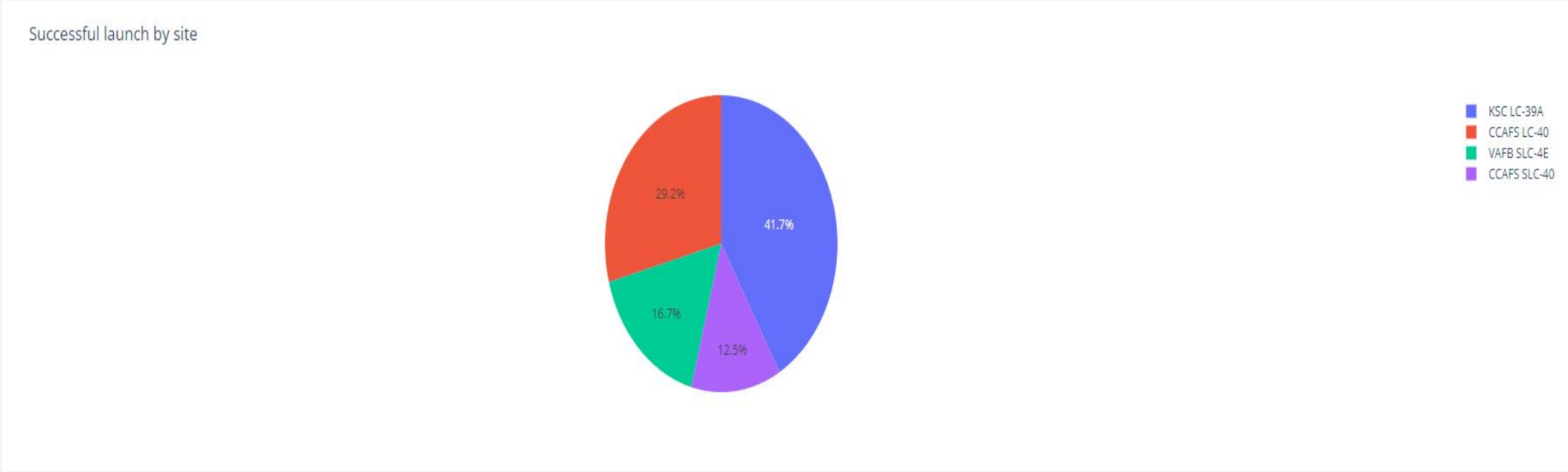
Launch site transportations

All launch site locate nearby the coast, all of them also have a good transportation : near the railway, highway but stay way from the central city.




Section 4
Build a Dashboard with Plotly Dash

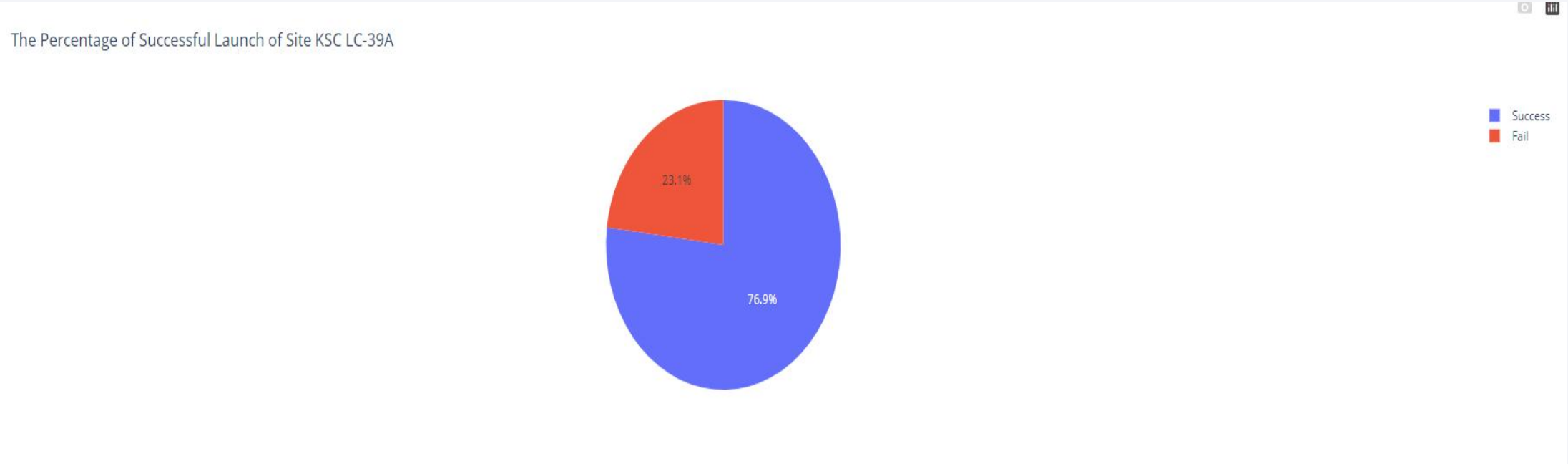
Successful rate contribution of each launch site



The pie chart show the contribution of each launch site to the total successful launch. As you can see, KSC LC-39A count for 41,7% in total. It meam that launching in this launch site get more rate of success than others.

The rest is CCAFS LC-40A, VAFB SLC-4E, CCAFS SLC-40 count for 29.2% , 16.7&, 12.5% respectively

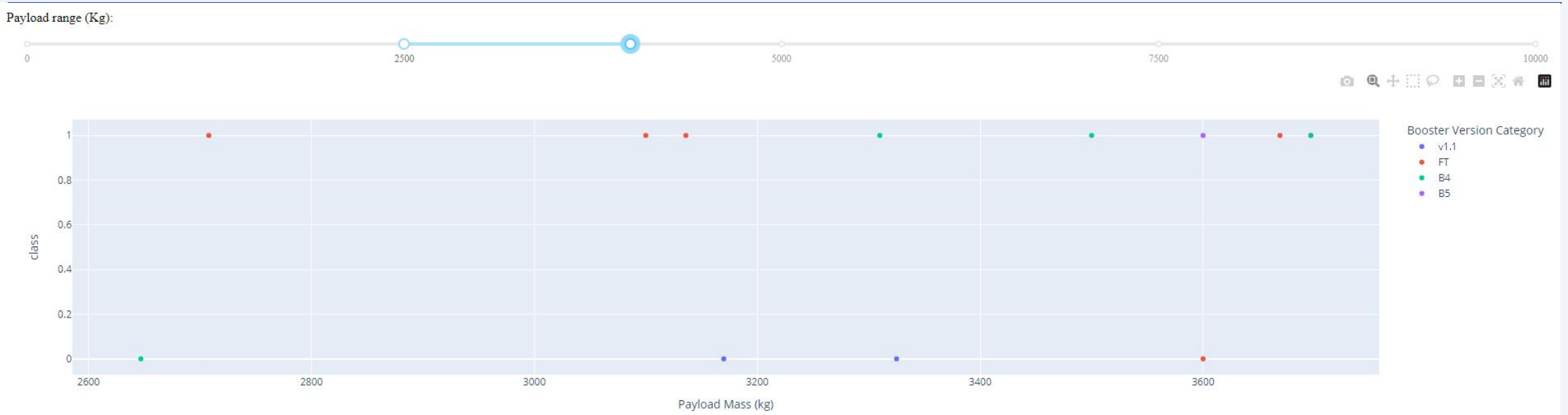
Success and failure rate of KSC LC-39A Launch site



Go further into the KSC LC-39A launch site which count for the most successful launch.

As the pie chart shows that 76.9% of launching was successful. It shows that launch rocket in this site would have a greater success rate

<Dashboard Screenshot 3>

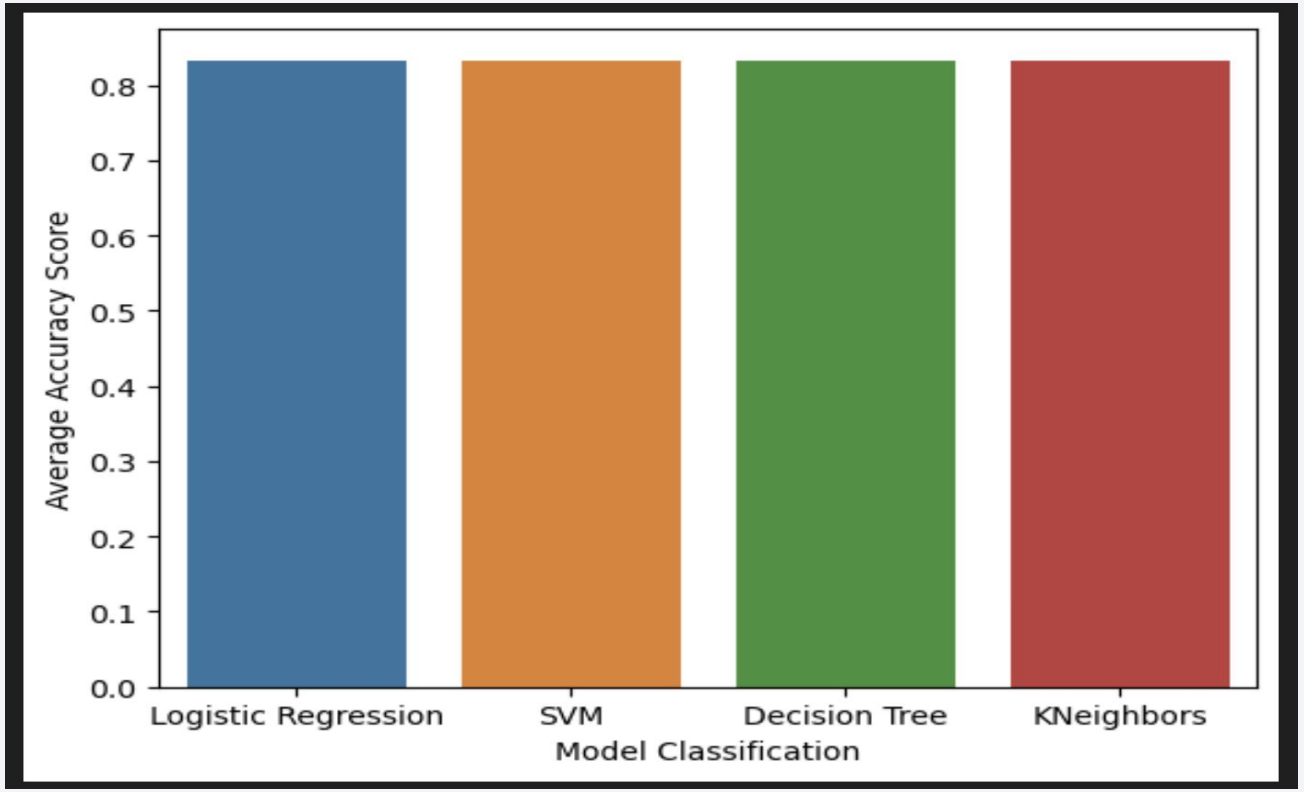


- As you can see, from the payload range between 2500kg and 3750kg. The success rate much higher than the failure rate, about 100%, version FT acount for the most times of successful launch, and the next second place is version B4.
- v1.1 have no successful launch in this range.

Section 5

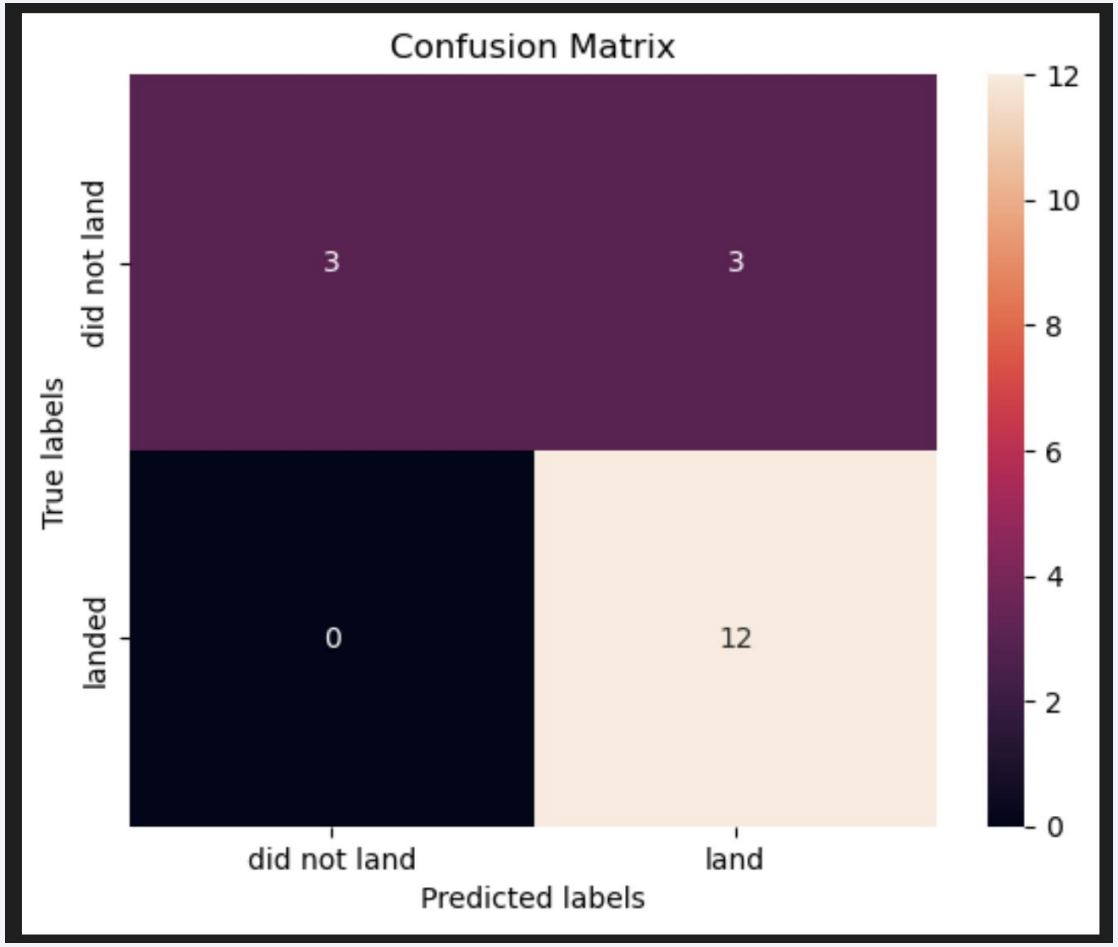
Predictive Analysis (Classification)

Classification Accuracy



The bar chart shows the accuracy score of 4 models was built.
All of 4 models got 0.833 accurcy score, so 4 models were equally work.

Confusion Matrix



- As you can see in the Confusion matrix, all models's predictions are very good at predict true positive (12) and false negative (0)
- But all model not good at predict true negative (3) and false positive (3)

Conclusions

- Conclusion 1: We can use a model to predict the outcome of a launching, the model is very good at predicting the true outcome of a successful launch.
- Success rate of a launch are affected by various factors: launch-site, payload and version of booster

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!