IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Fadhil Reza Maulana
December, 12 2022

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

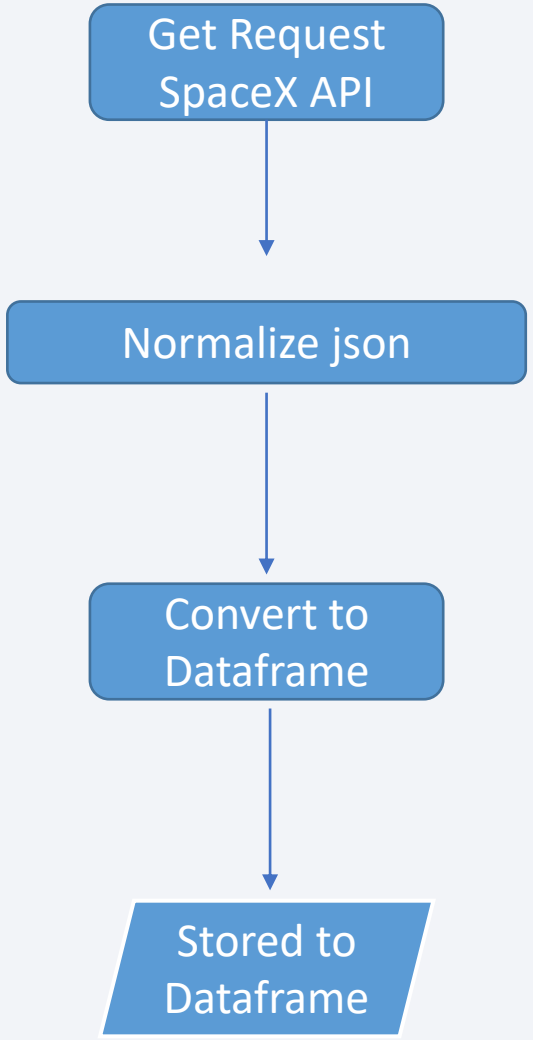
- Data collection methodology:
  - Data collected use SpaceX API and scrapping from Wikipedia
- Perform data wrangling
  - Data was encoded by one hot encoding and was sclaed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

SpaceX API Data Collection Github URL

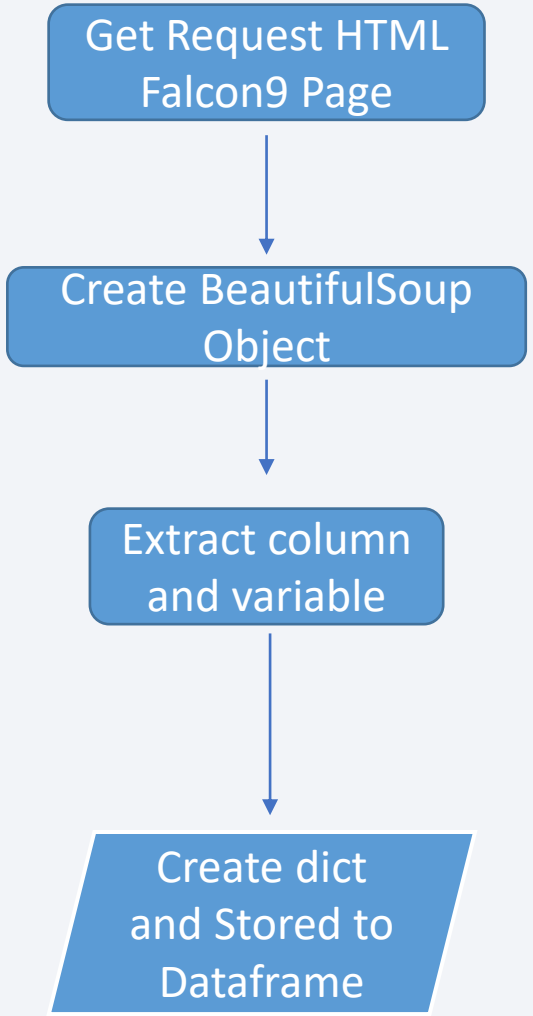


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

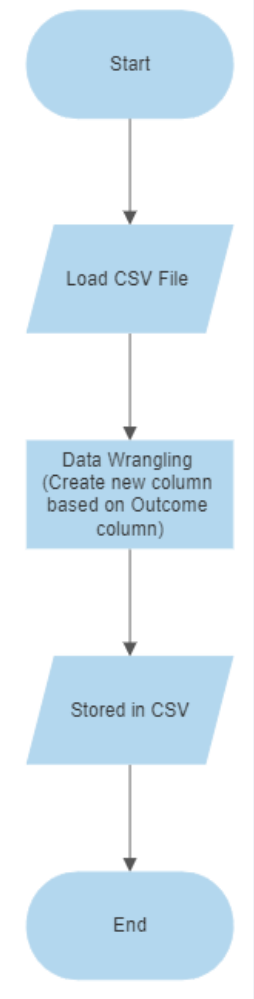
Falcon9 HTML Web Scrapping GitHub URL



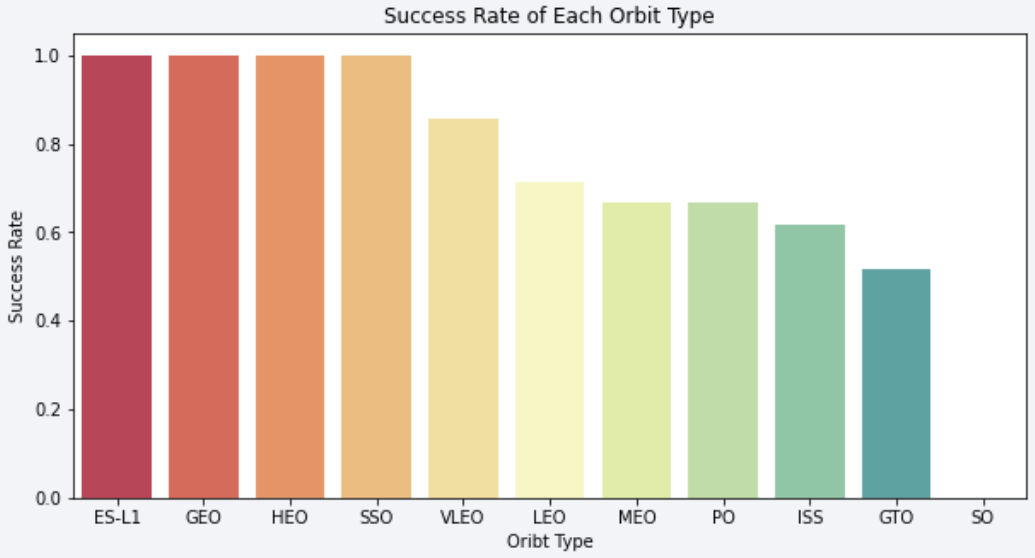
Data Wrangling

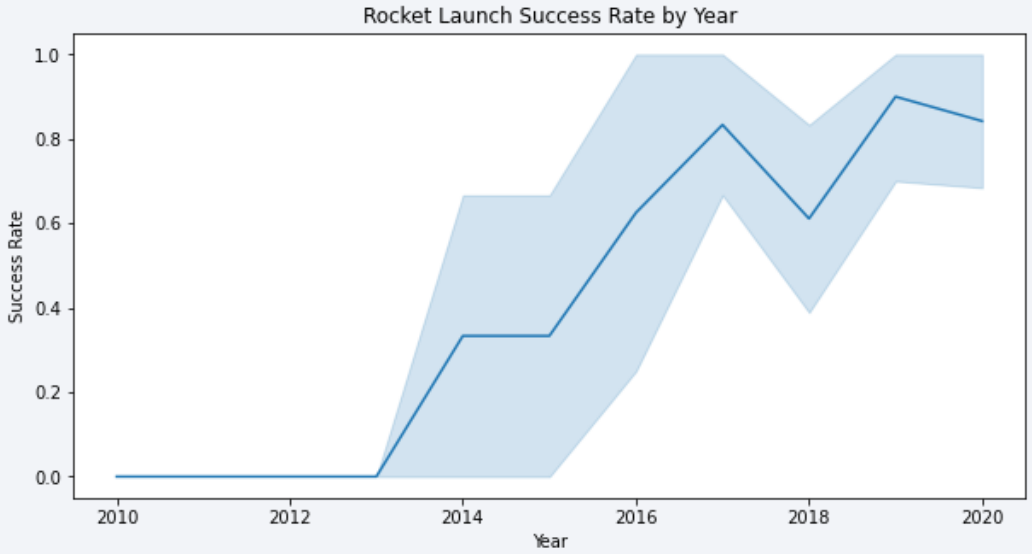
- Describe how data were processed

[Data Wrangling GitHub URL]



EDA with Data Visualization





Both chart show us which Orbit Type and Year with the highest success rate

[EDA Data Viz GitHub URL]

EDA with SQL

EDA with SQL GitHub Link

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Folium EDA GitHub URL

Build a Dashboard with Plotly Dash

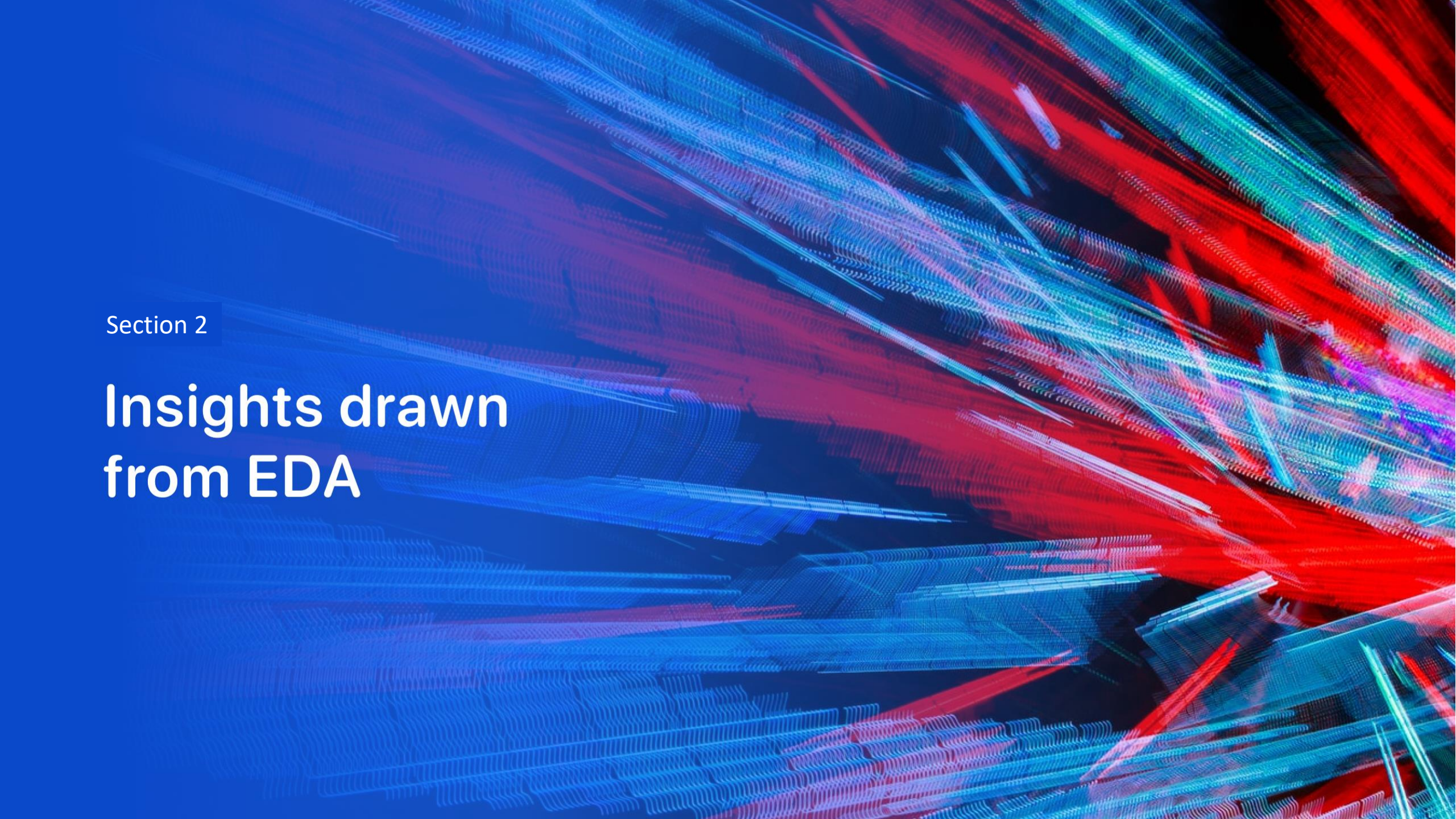
Dashboard with Plotly Dash GitHub Link

Predictive Analysis (Classification)

SpaceX Falcon 9 Predictive Classification Model GitHub Link

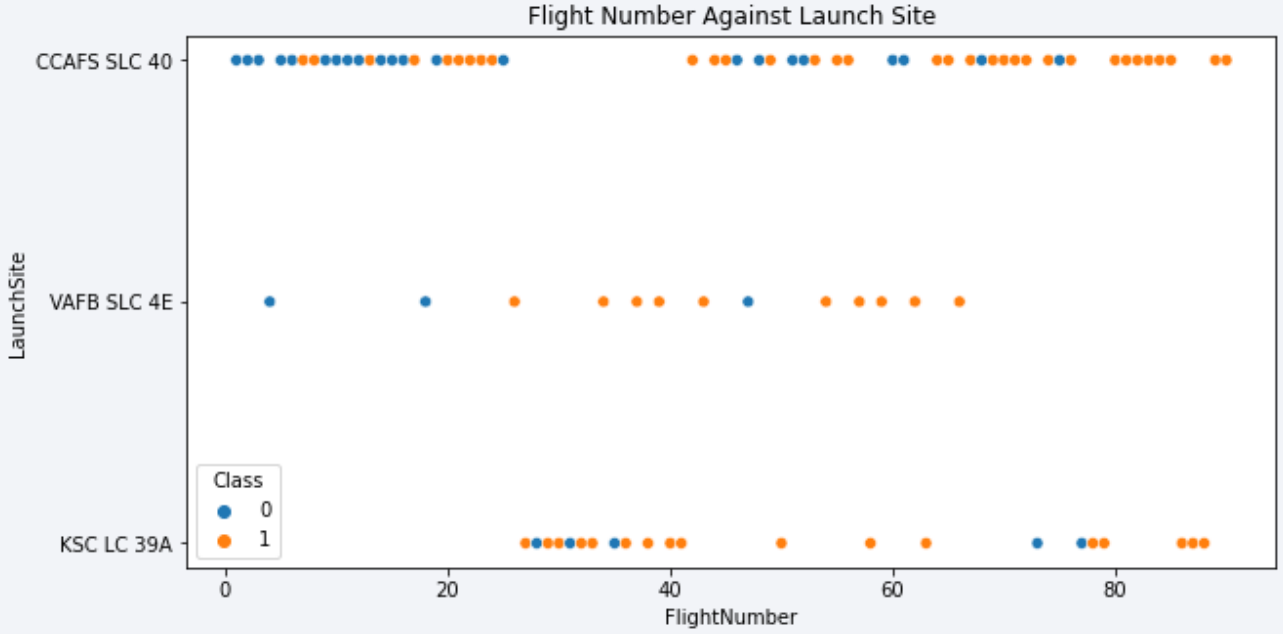
Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

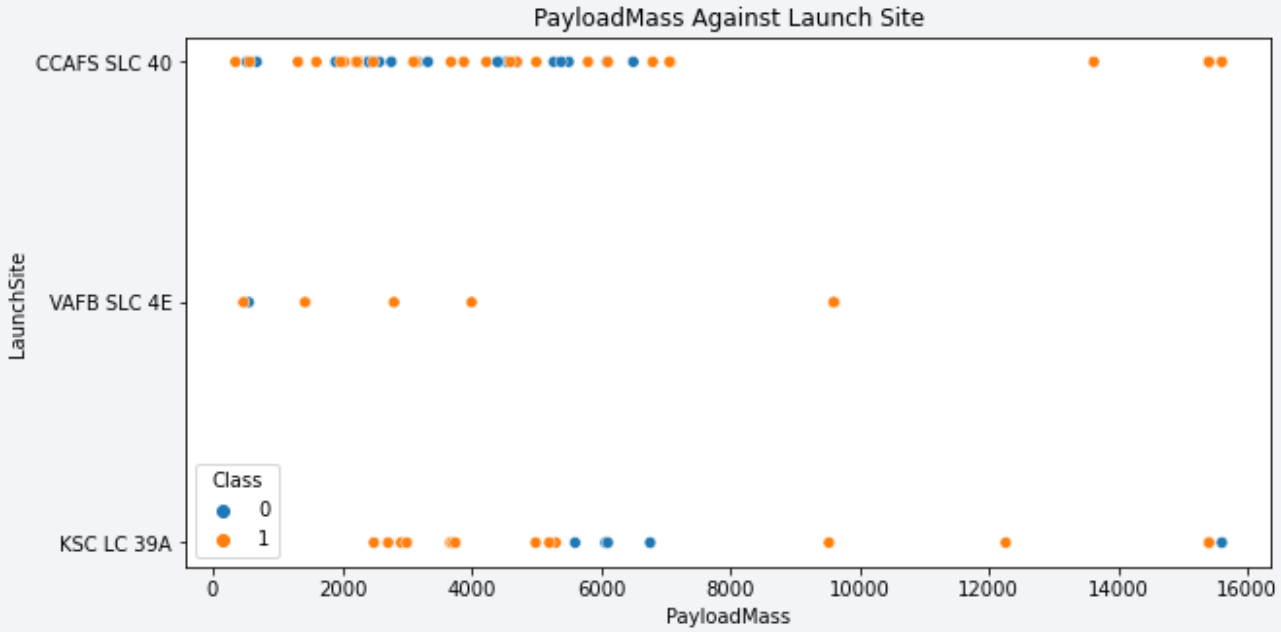
- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations



From the plot we can see that most Flight Number from all LaunchSite was successfully landed

Payload vs. Launch Site

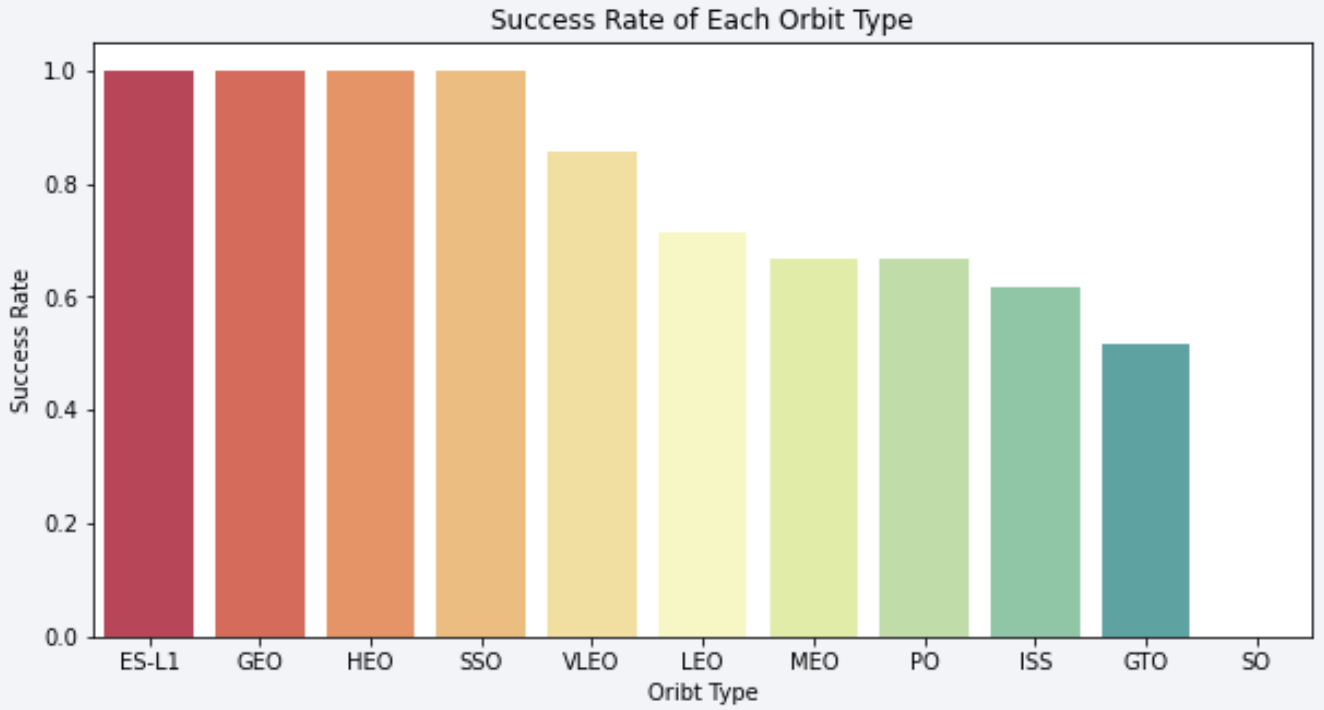
- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations



From plot we can see that Launch Site 'CCAFS SLC 40' is launchsite with the highest landing failure for Falcon9 with Payload under 6000 Kg

Success Rate vs. Orbit Type

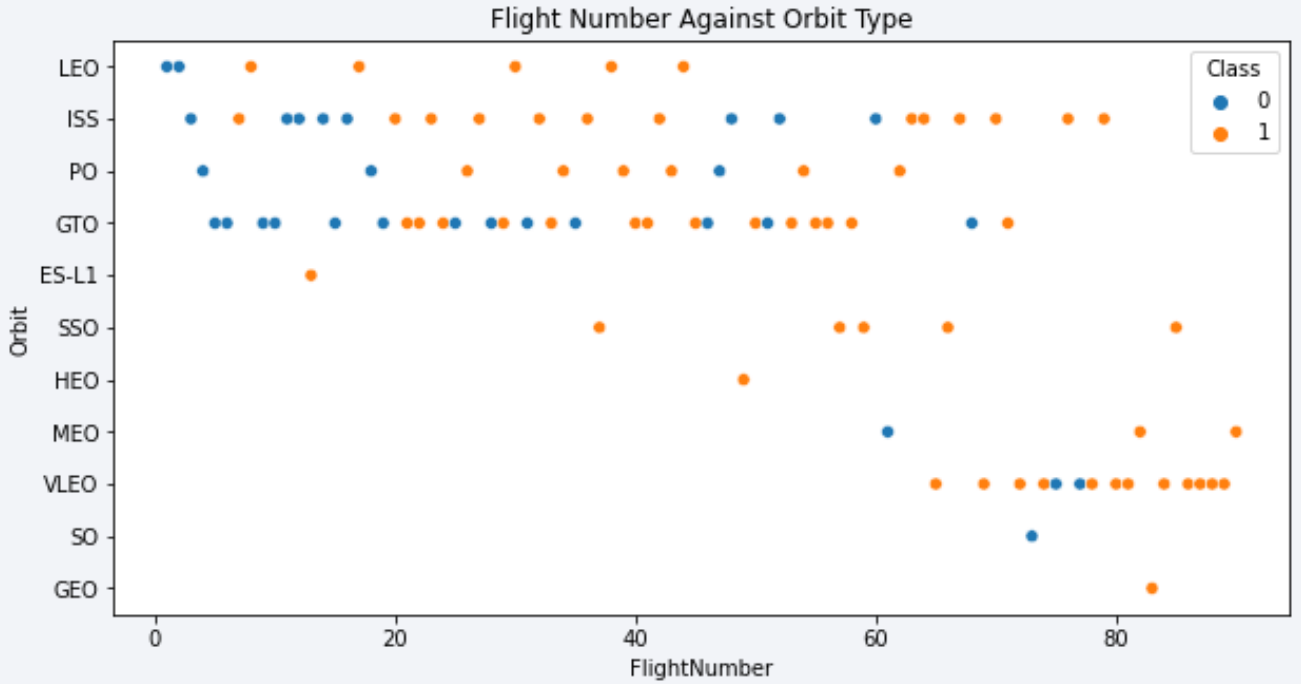
- Show a bar chart for the success rate of each orbit type
- Show the screenshot of the scatter plot with explanations



Graph above show us that ES-L1 orbit type has the highest success rate than other orbit type

Flight Number vs. Orbit Type

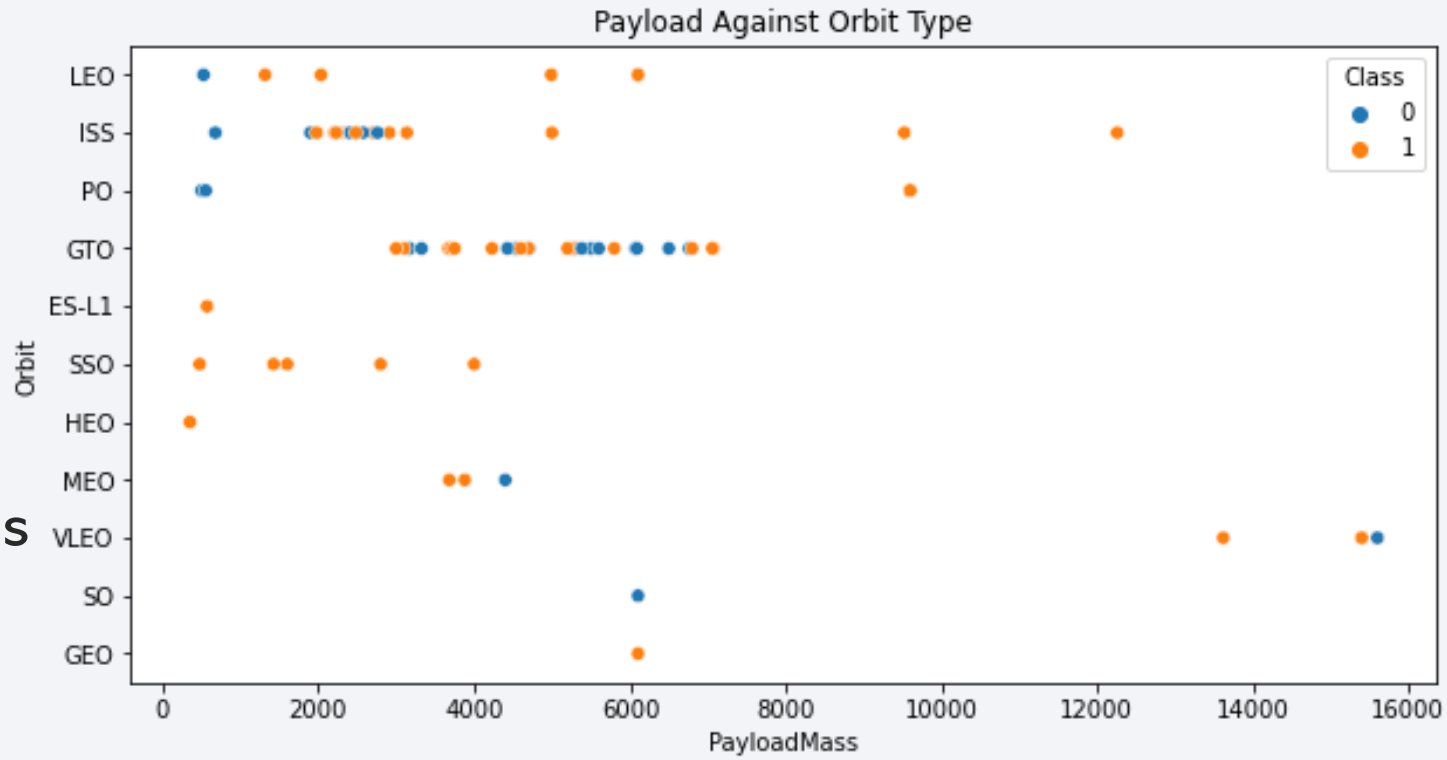
- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations



VLEO orbit type become the most successful orbit with the highest rate of flight number

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations

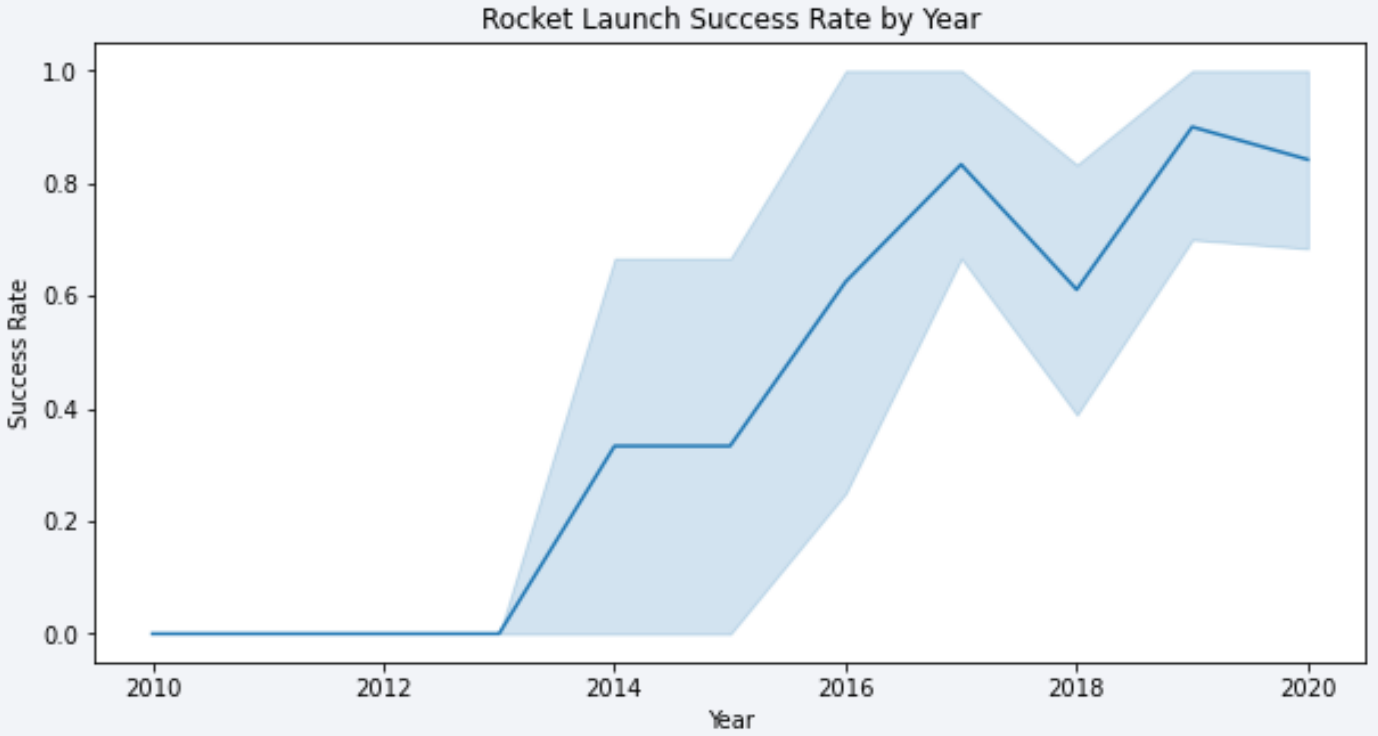


- VLEO become an Orbit with flight payloadmass over 12000 Kg

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations



From the plot above we can see that The success rate from rocket launch each year is mostly increase, exceptional form year 2017 to 2018 there's slightly decreseasing.

All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

```
task_1 = '''
        SELECT DISTINCT LaunchSite
        FROM SpaceX
'''
create_pandas_df(task_1, database=conn)
```

	launchsite
0	KSC LC-39A
1	CCAFS LC-40
2	CCAFS SLC-40
3	VAFB SLC-4E

With Distinct from LaunchSite column we can get unique value from that column. And now we get there are 4 unqieu Launch Site Name

Launch Site Names Begin with 'CCA'

```
task_2 = '''
        SELECT *
        FROM SpaceX
        WHERE LaunchSite LIKE 'CCA%'
        LIMIT 5
        '''
create_pandas_df(task_2, database=conn)
```

Python

	date	time	boosterversion	launchsite	payload	payloadmasskg	orbit	customer	missionoutcome	landingoutcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

To get top 5 Launch Site Name that begin with CCA we use LIKE and limit to only 5 query result

Total Payload Mass

```
task_3 = '''
        SELECT SUM(PayloadMassKG) AS Total_PayloadMass
        FROM SpaceX
        WHERE Customer LIKE 'NASA (CRS)'
        '''
create_pandas_df(task_3, database=conn)
```

total_payloadmass
0 45596

To get Total Payload carried by NASA we use SUM on PayloadMassKg and use like filter on Customer column

Average Payload Mass by F9 v1.1

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

```
task_4 = '''
        SELECT AVG(PayloadMassKG) AS Avg_PayloadMass
        FROM SpaceX
        WHERE BoosterVersion = 'F9 v1.1'
        '''
create_pandas_df(task_4, database=conn)
```

	avg_payloadmass
0	2928.4

Query above show us how to get avg payload mass that carried by booseter version F9 v1.1

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
task_5 = '''
        SELECT MIN(Date) AS FirstSuccessfull_landing_date
        FROM SpaceX
        WHERE LandingOutcome LIKE 'Success (ground pad)'
        '''
create_pandas_df(task_5, database=conn)
```

	firstsuccessfull_landing_date
0	2015-12-22

To find the first successful landing that means we have to query the minum date that exist in this database. So we have to use MIN on Date column and use like filter on landingoutcome column to have value Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
task_6 = '''
        SELECT BoosterVersion
        FROM SpaceX
        WHERE LandingOutcome = 'Success (drone ship)'
            AND PayloadMassKG > 4000
            AND PayloadMassKG < 6000
        '''
create_pandas_df(task_6, database=conn)
```

	boosterversion
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Use filter on landing outcome and payloadmasskg with value over 4000 and under 6000

Total Number of Successful and Failure Mission Outcomes

```
task_7a = '''
        SELECT COUNT(MissionOutcome) AS SuccessOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Success%'
        '''

task_7b = '''
        SELECT COUNT(MissionOutcome) AS FailureOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Failure%'
        '''
print('The total number of successful mission outcome is:')
display(create_pandas_df(task_7a, database=conn))
print()
print('The total number of failed mission outcome is:')
create_pandas_df(task_7b, database=conn)
```

```
The total number of successful mission outcome is:
```

	successoutcome
0	100

```
The total number of failed mission outcome is:
```

	failureoutcome
0	1

Boosters Carried Maximum Payload

```
task_8 = '''
        SELECT BoosterVersion, PayloadMassKG
        FROM SpaceX
        WHERE PayloadMassKG = (
                              SELECT MAX(PayloadMassKG)
                              FROM SpaceX
                              )
        ORDER BY BoosterVersion
        '''
create_pandas_df(task_8, database=conn)
```

	boosterversion	payloadmasskg
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
5	F9 B5 B1051.3	15600
6	F9 B5 B1051.4	15600
7	F9 B5 B1051.6	15600
8	F9 B5 B1056.4	15600
9	F9 B5 B1058.3	15600
10	F9 B5 B1060.2	15600
11	F9 B5 B1060.3	15600

2015 Launch Records

```
task_9 = '''
        SELECT BoosterVersion, LaunchSite, LandingOutcome
        FROM SpaceX
        WHERE LandingOutcome LIKE 'Failure (drone ship)'
            AND Date BETWEEN '2015-01-01' AND '2015-12-31'
        '''
create_pandas_df(task_9, database=conn)
```

	boosterversion	launchsite	landingoutcome
0	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

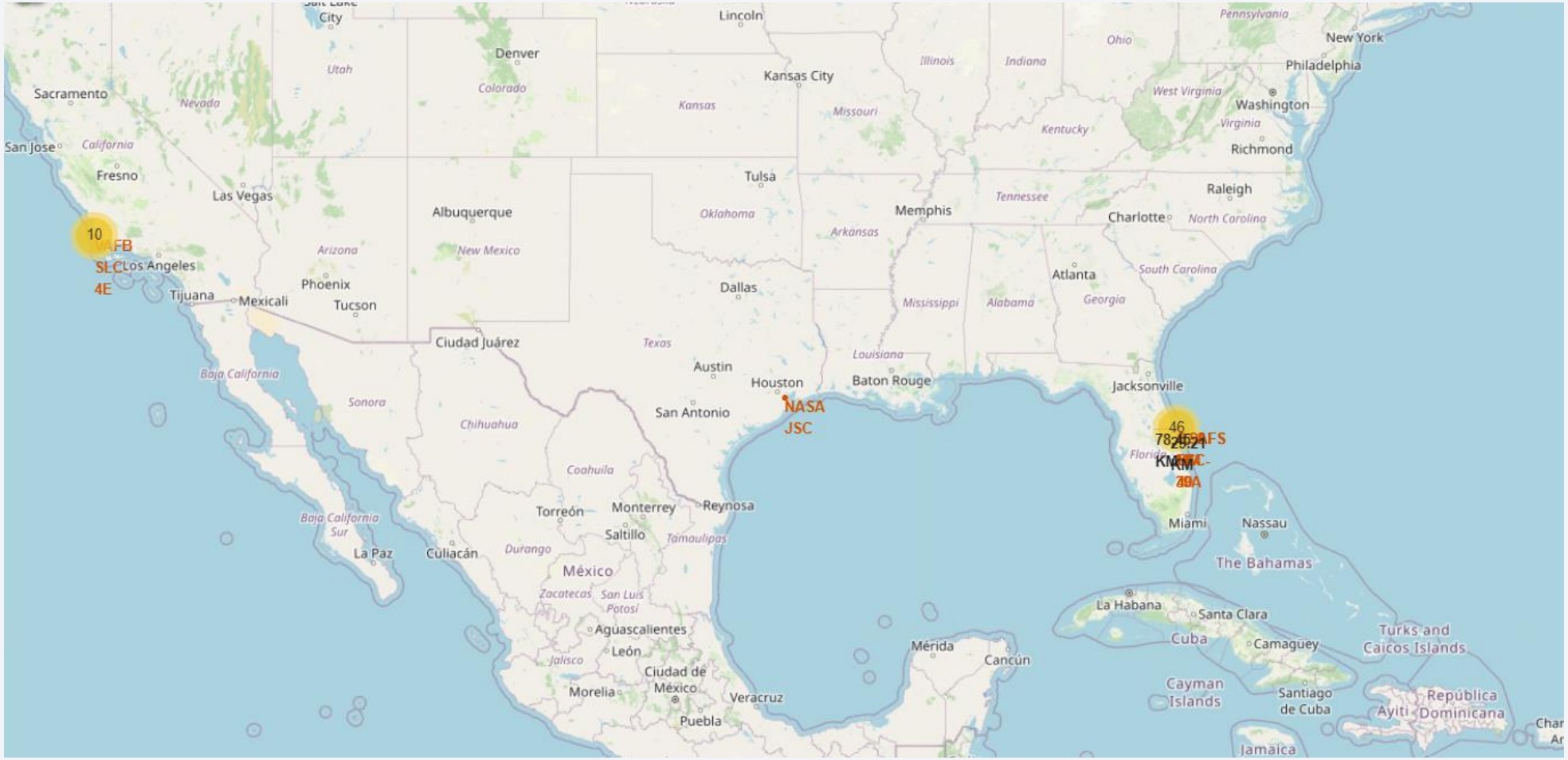
```
task_10 = '''
        SELECT LandingOutcome, COUNT(LandingOutcome)
        FROM SpaceX
        WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
        GROUP BY LandingOutcome
        ORDER BY COUNT(LandingOutcome) DESC
        '''
create_pandas_df(task_10, database=conn)
```

	landingoutcome	count
0	No attempt	10
1	Success (drone ship)	6
2	Failure (drone ship)	5
3	Success (ground pad)	5
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Precluded (drone ship)	1
7	Failure (parachute)	1

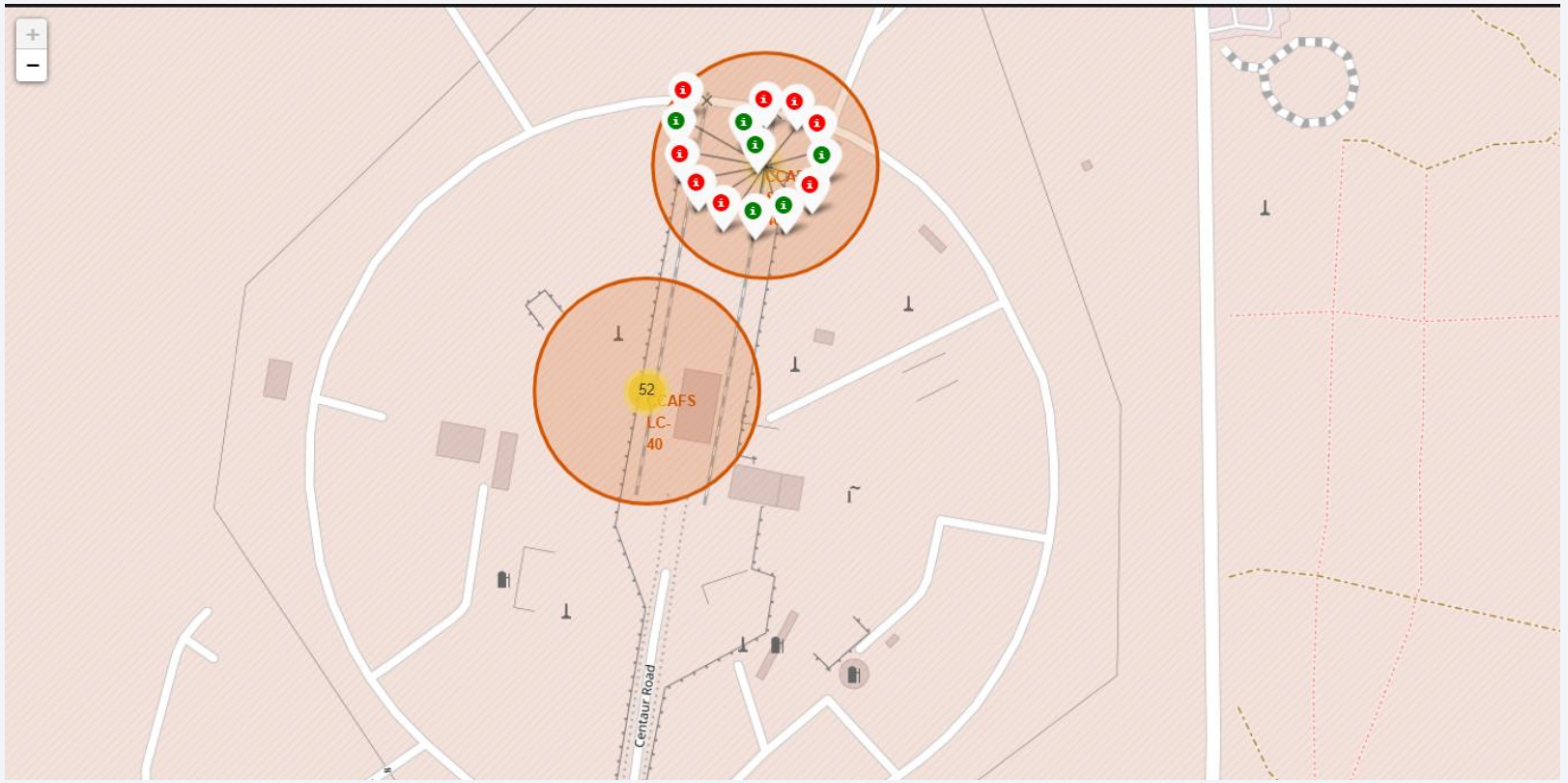
Section 3

Launch Sites Proximities Analysis

Rocket Launch Site Location

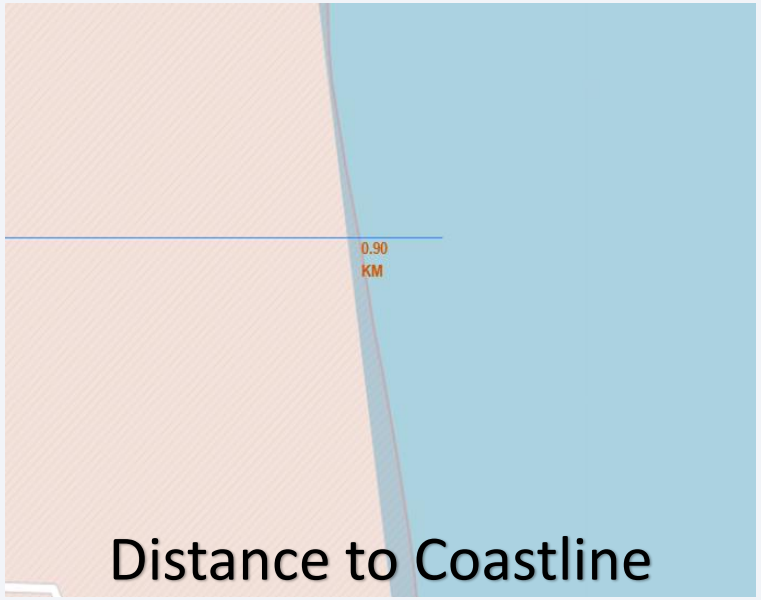


CCAFS ALC-40 Launch Site Marker Color

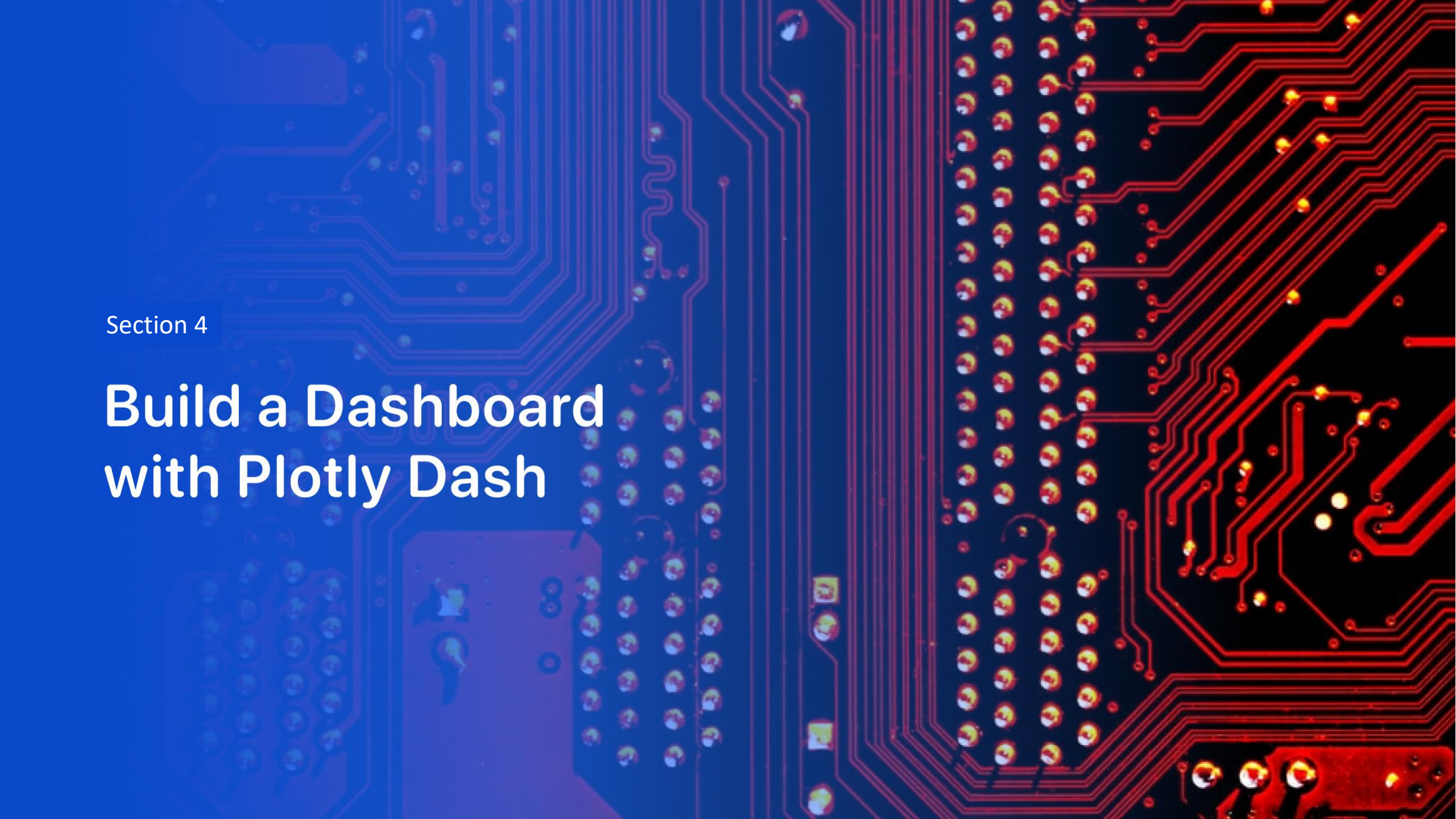


Maps above show us the coordinate of CCAFS ALC-40 Launch Site. Green markers indicates a successful landing and the red markers indicates an unsuccessful landing

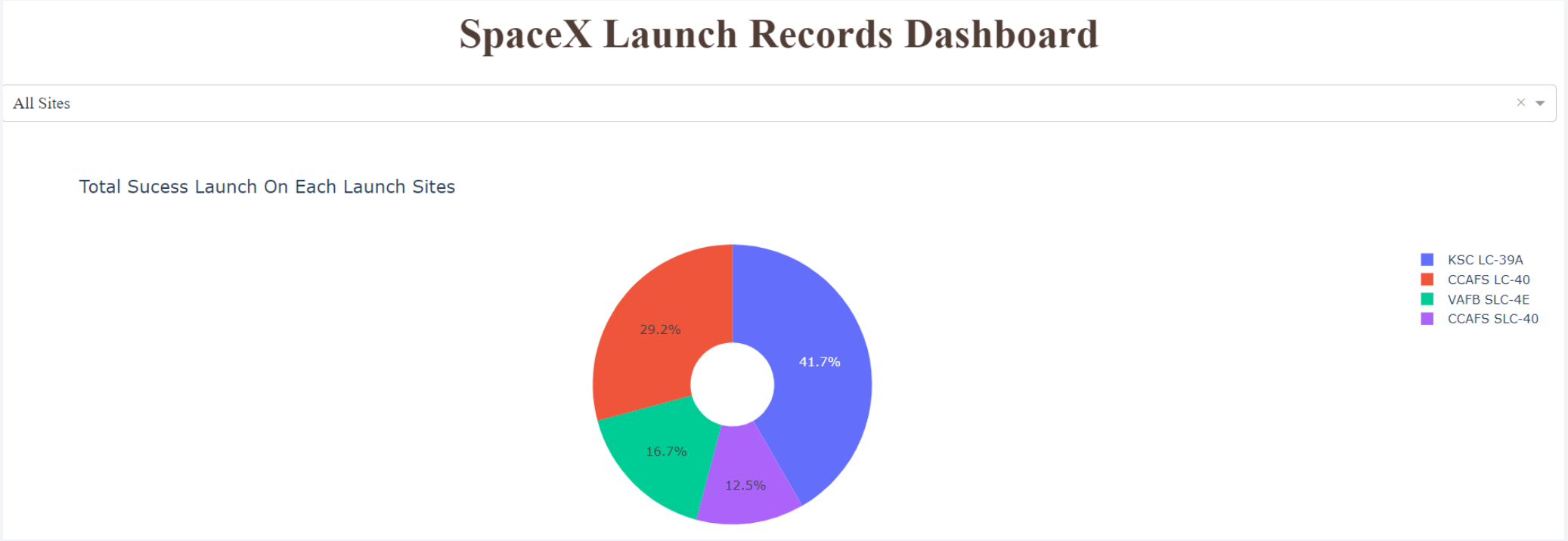
Launch Site Location Distance to Public Places



Are launch sites in close proximity to railways? No
Are launch sites in close proximity to highways? No
Are launch sites in close proximity to coastline? Yes
Do launch sites keep certain distance away from cities? Yes

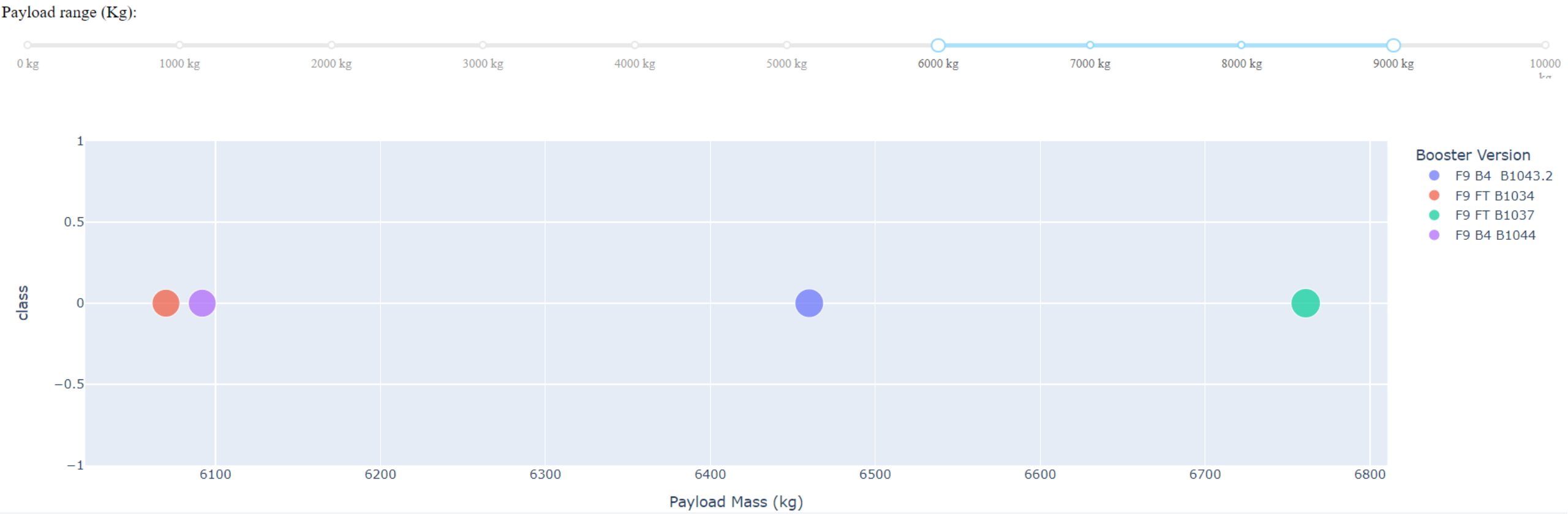

Section 4
Build a Dashboard with Plotly Dash

Total Success Launch on Each Launch Site



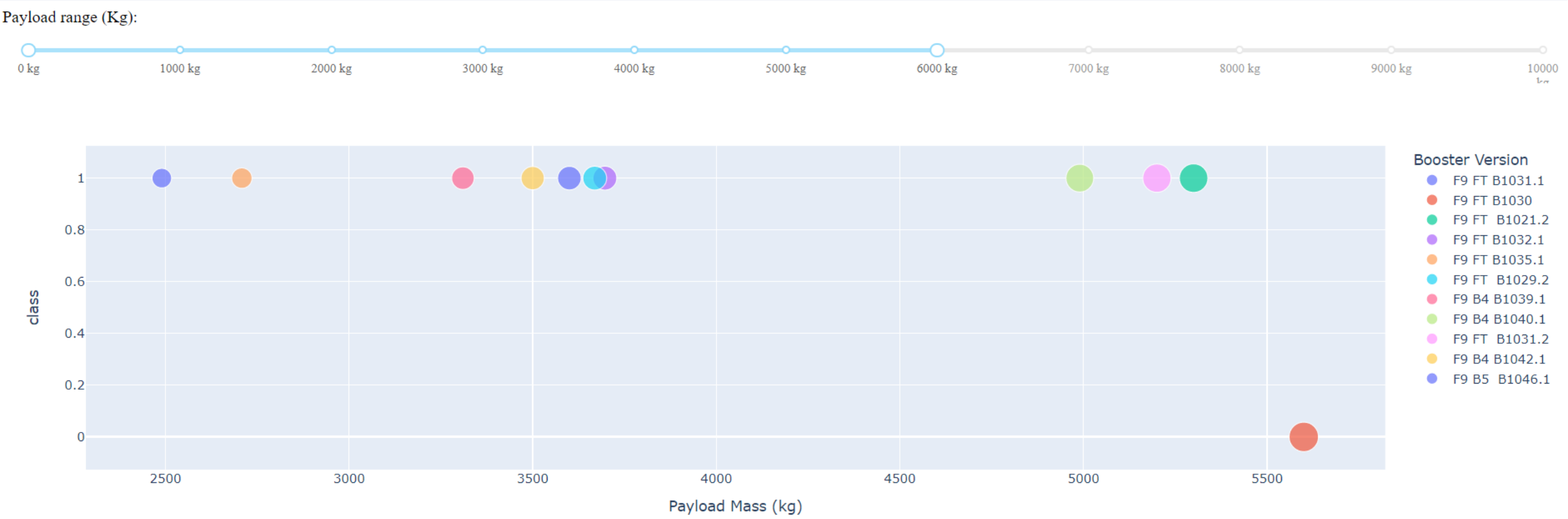
From chart above,KSC LC-39A is the most successful launch site, with success rate around 41 %, and CCAFS SLC-40 launch site with the lowest success rate with 12, 5 %

Booster Version Success Rate Based on Payload Mass (kg)



For rocket launch with Payload mass around 60.000 Kg – 90.000 Kg. we can see that all the Booster Version had failed all the launches.

Booster Version Success Rate Based on Payload Mass (kg)

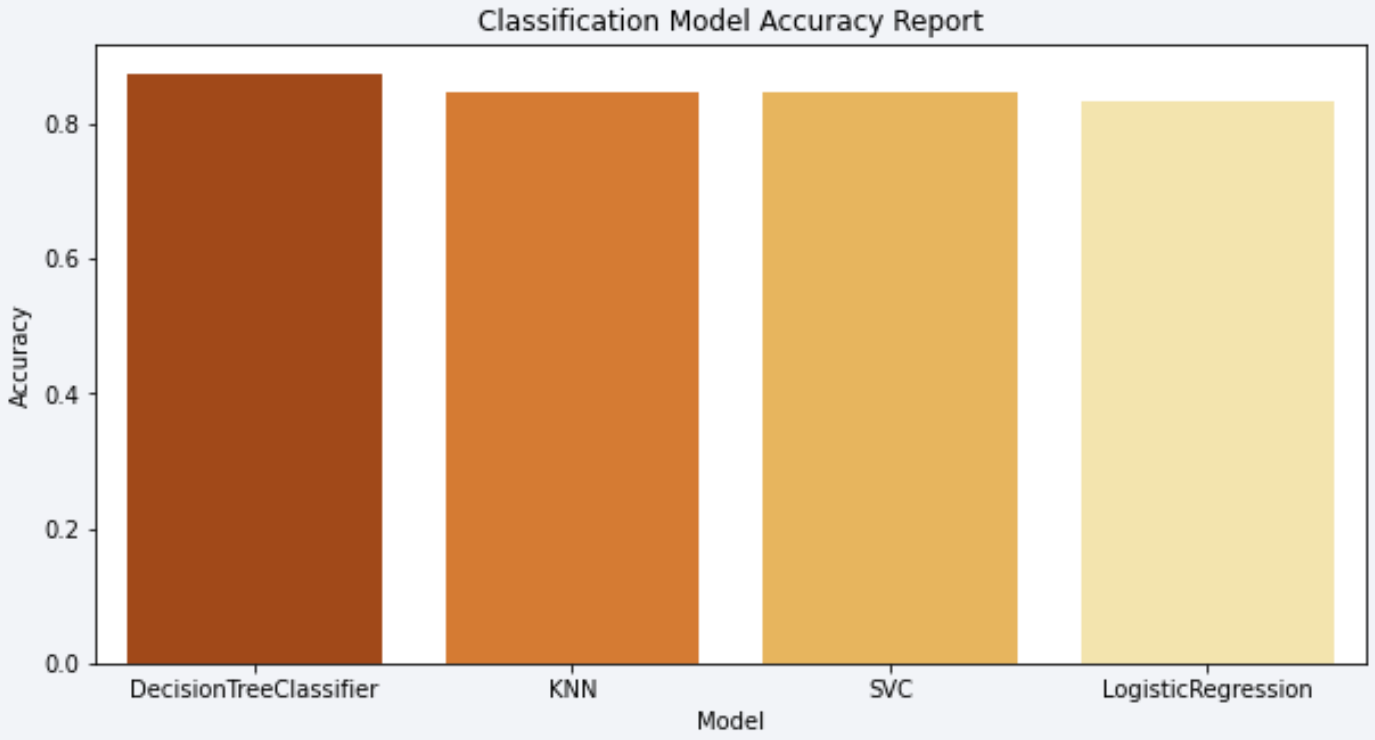


For rocket launch with Payload mass under 5500 Kg. we can see that all the Booster Version had succeed all the launches

Section 5

Predictive Analysis (Classification)

Classification Accuracy



From the plot beside, we can see that in this dataset, ***DecisionTreeClassifier*** is the best model with the best accuracy result

Confusion Matrix

This model has the highest accuracy among other algorithm, and it can be proved by this confusion matrix beside. the model only made 3 prediction errors, which predicting the rockets that didn't land that did actually it was landed.

Conclusions

- Launch Site ‘CCAFS SLC 40’ is launchsite with the highest landing failure for Falcon9 with Payload under 6000 Kg
- that ES-L1 orbit type has the highest success rate than other orbit type
- VLEO orbit type become the most successful orbit with the highest rate of flight number
- The success rate from rocket launch each year is mostly increase

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!