IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yaser Alburshiad
June 2023

Outline

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

Executive Summary

- In this capstone project, we will predict whether the SpaceX Falcon 9 first stage will land successfully. If we can determine if the first stage will land, we can determine the cost of a launch. This will be achieved using classfiers machine learning algorithm such as logistics regression, KNN, etc.
- The data is collected through REST API and web scrapping, then some data wrngling techniques are applied to do Exploratory Data Analysis (EDA) and visualization. Last, we use machine learning to predict the success of a launch.
- I found out that there is a correlation between some features of used rockets and the success of a mission.

Introduction

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. 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. 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 SpaceX for a rocket launch.
- The problem to solve is to determine if the first stage will land successfully

Section 1

Methodology

Methodology

Executive Summary

- To build our dataset we collected data from SpaceX API & Web scrapping from Wikipedia, then we clean the data by removing null values and not needed columns using panda library from python.
- After we get the needed dataset, we did exploratory data analysis (EDA) using visualization tools such as matplotlib and seaborn libraries, as well as answering questions using SQL queries. It make us understand the data deeply and indicate a relations between the rocket features and the success of a mission.
- Then, we used Folium to create an interactive map with the location of each launch site and mapped successful/failed launched for each site. Also, using Plotly Dash we built a dashboard to help us identify the most successful launch site.
- Finally, we prepare the dataset to be used in machine learning algorithm by changing categorical values to numerical values (One Hot Encoding), then perform Logistic Regression, SVM, Decision Tree, & KNN to get the best classifier that can predict successful launches.

Data Collection – SpaceX API

```python
# Takes the dataset and uses the cores column to call the API and append the data to the lists
def getCoreData(data):
    for core in data['cores']:
            if core['core'] != None:
                response = requests.get("https://api.spacexdata.com/v4/cores/"+core['core']).json()
                Block.append(response['block'])
                ReusedCount.append(response['reuse_count'])
                Serial.append(response['serial'])
            else:
                Block.append(None)
                ReusedCount.append(None)
                Serial.append(None)
            Outcome.append(str(core['landing_success'])+' '+str(core['landing_type']))
            Flights.append(core['flight'])
            GridFins.append(core['gridfins'])
            Reused.append(core['reused'])
            Legs.append(core['legs'])
            LandingPad.append(core['landpad'])
```

Now let's start requesting rocket launch data from SpaceX API with the following URL:

```python
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```python
response = requests.get(spacex_url)
```

Check the content of the response

```python
print(response.content)
```

- Github link: https://github.com/YABSHD/-spaceX-landing-success-prediction/blob/main/1_jupyter-labs-spacex-data-collection-api.ipynb

1. Request and parse the SpaceX launch data using the GET request
2. Normalize JSON response into a dataframe
3. Create new pandas dataframe from the needed columns
4. Get data from Rocket9 launcher
5. Handle missing values
6. Export to CSV file

Data Collection - Scraping

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon
```

Next, request the HTML page from the above URL and get a `response` object

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP res

```
# use requests.get() method with the provided static_url
# assign the response to a object
response = requests.get(static_url)
```

Create a `BeautifulSoup` object from the HTML `response`

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response.content, 'html.parser')
```

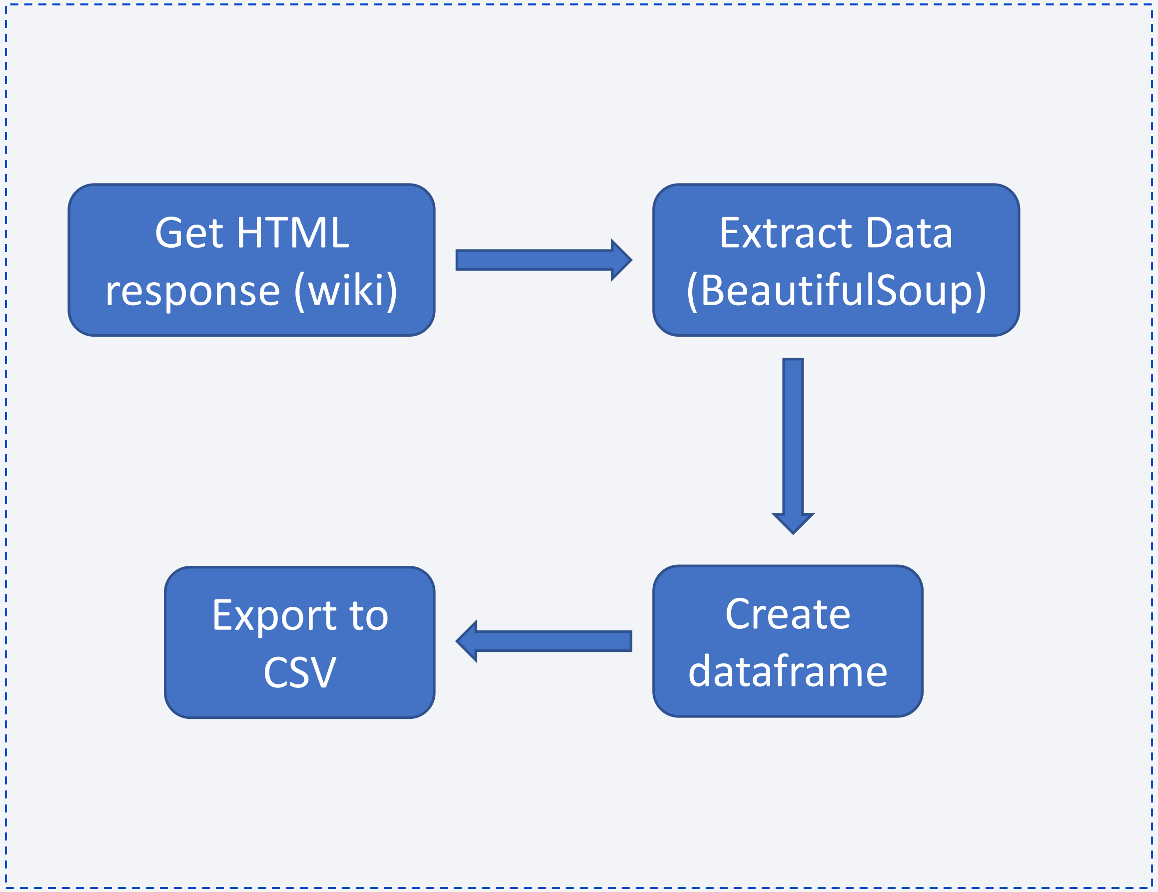
Print the page title to verify if the `Beautifu`

```
# Use soup.title attribute
soup.title
```

```
# Use the find_all function in the BeautifulSoup object, with element ty
# Assign the result to a list called `html_tables`
html_tables = soup.find_all('table')
```

Starting from the third table is our target table contains the actual launch records.

```
# Let's print the third table and check its content
first_launch_table = html_tables[2]
print(first_launch_table)
```



- Github link: https://github.com/YABSHD/-spaceX-landing-success-prediction/blob/main/2_jupyter-labs-webscraping.ipynb

Data Wrangling

1. Calculate the number of launches on each site

2. Calculate the number and occurrence of each orbit

3. Calculate the number and occurence of mission outcome per orbit type

4. Create a landing outcome label from Outcome column

5. Export to CSV file

- Github link:

https://github.com/YABSHD/-spaceX-landing-success-prediction/blob/main/3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

1. Scatter Plots, which used to show relations between two attributes. It is used to test the relation between:
   - Flight Number vs. Launch Site
   - Payload vs. Launch Site
   - Flight Number vs. Orbit Type
   - Payload vs. Orbit Type

2. Bar chart, to compare the Success Rate for different Orbit Types

3. Line chart to draw the success rate over years

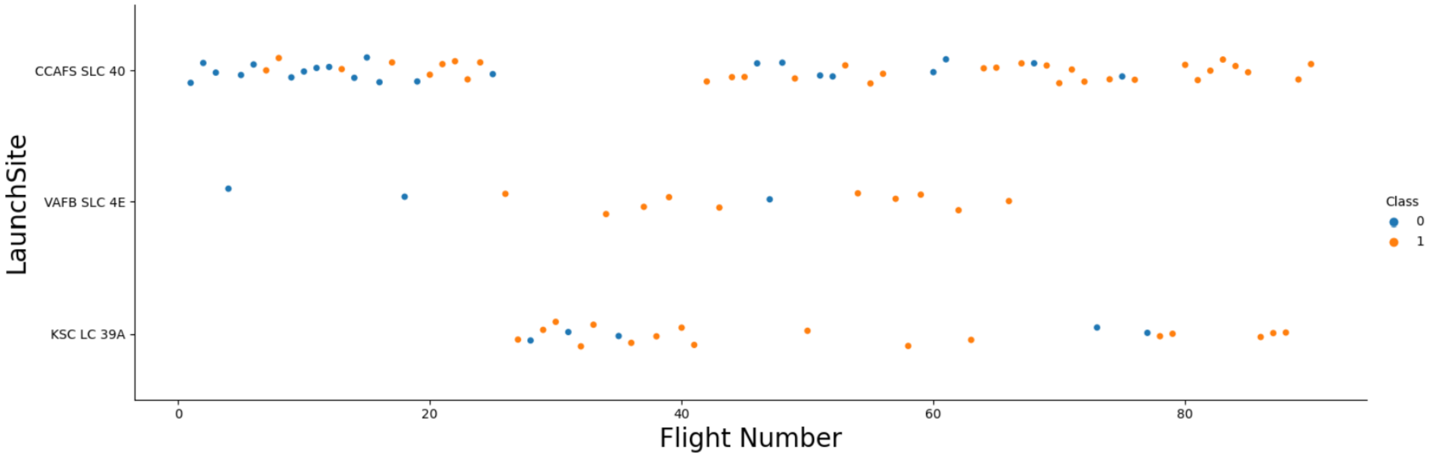
EDA with SQL

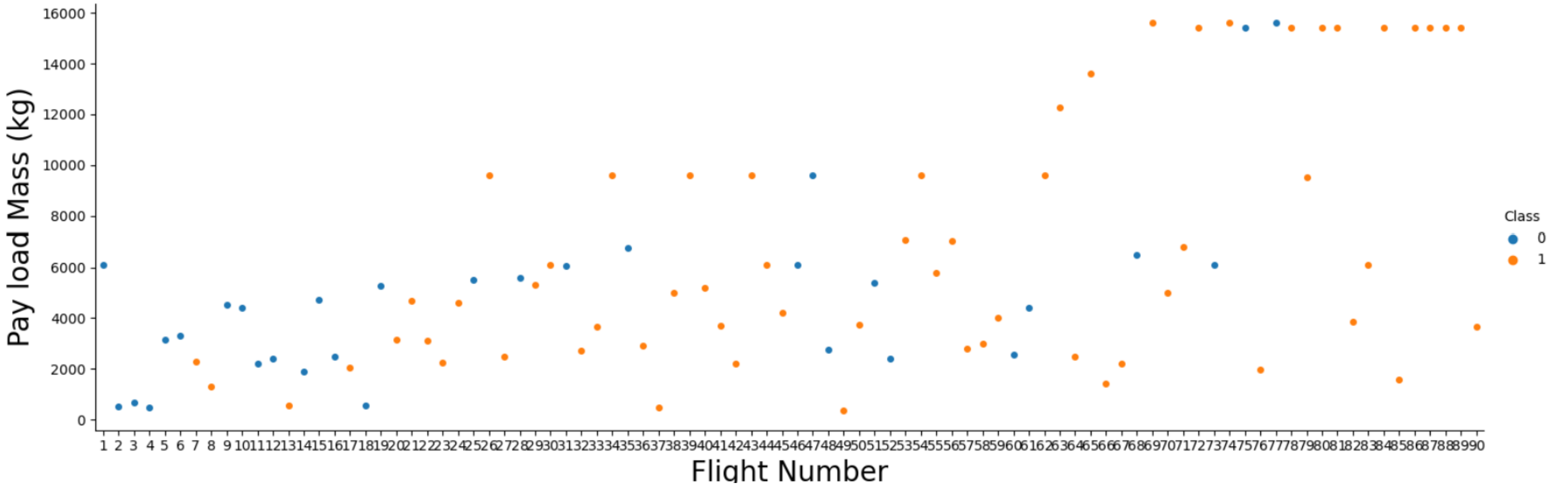
- The following queries are performed to EDA
  - **Display the names of the unique launch sites in the space mission**
  - **Display 5 records where launch sites begin with the string 'CCA'**
  - **Display the total payload mass carried by boosters launched by NASA (CRS)**
  - **Display average payload 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.**
- Github link:

https://github.com/YABSHD/-spaceX-landing-success-prediction/blob/main/4_jupyter-labs-eda-sql-coursera_sqllite.ipynb

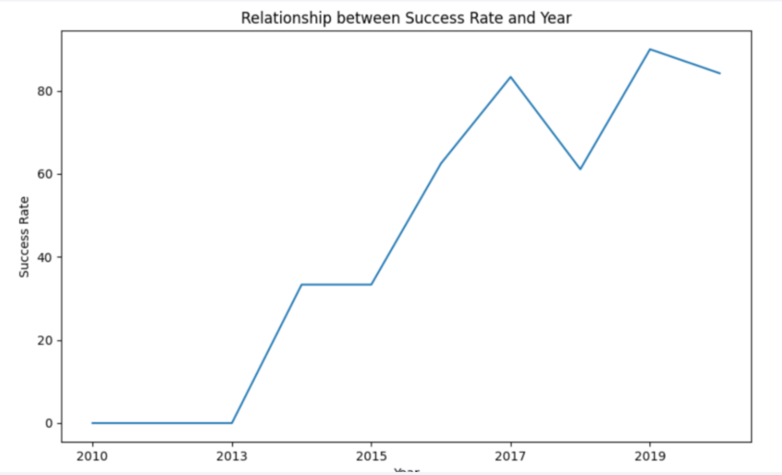
EDA with Data Visualization

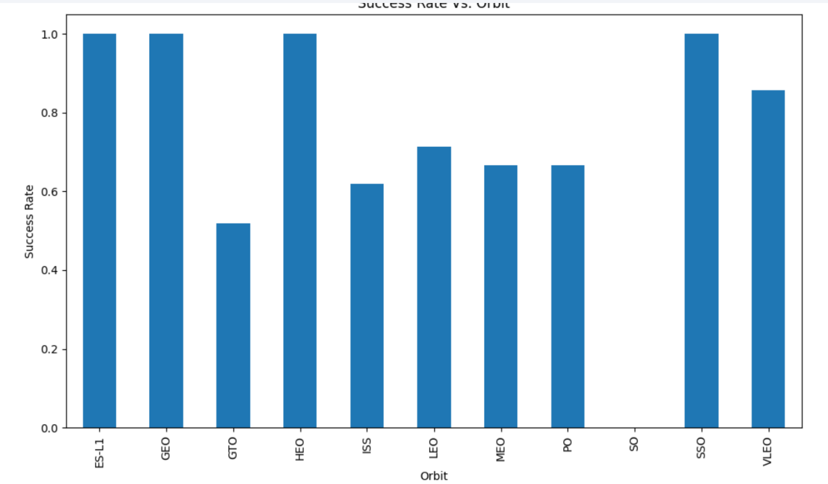
visualized the SpaceX launch dataset using matplotlib and seaborn and discovered some preliminary correlations between the launch site and success rates





EDA with Data Visualization





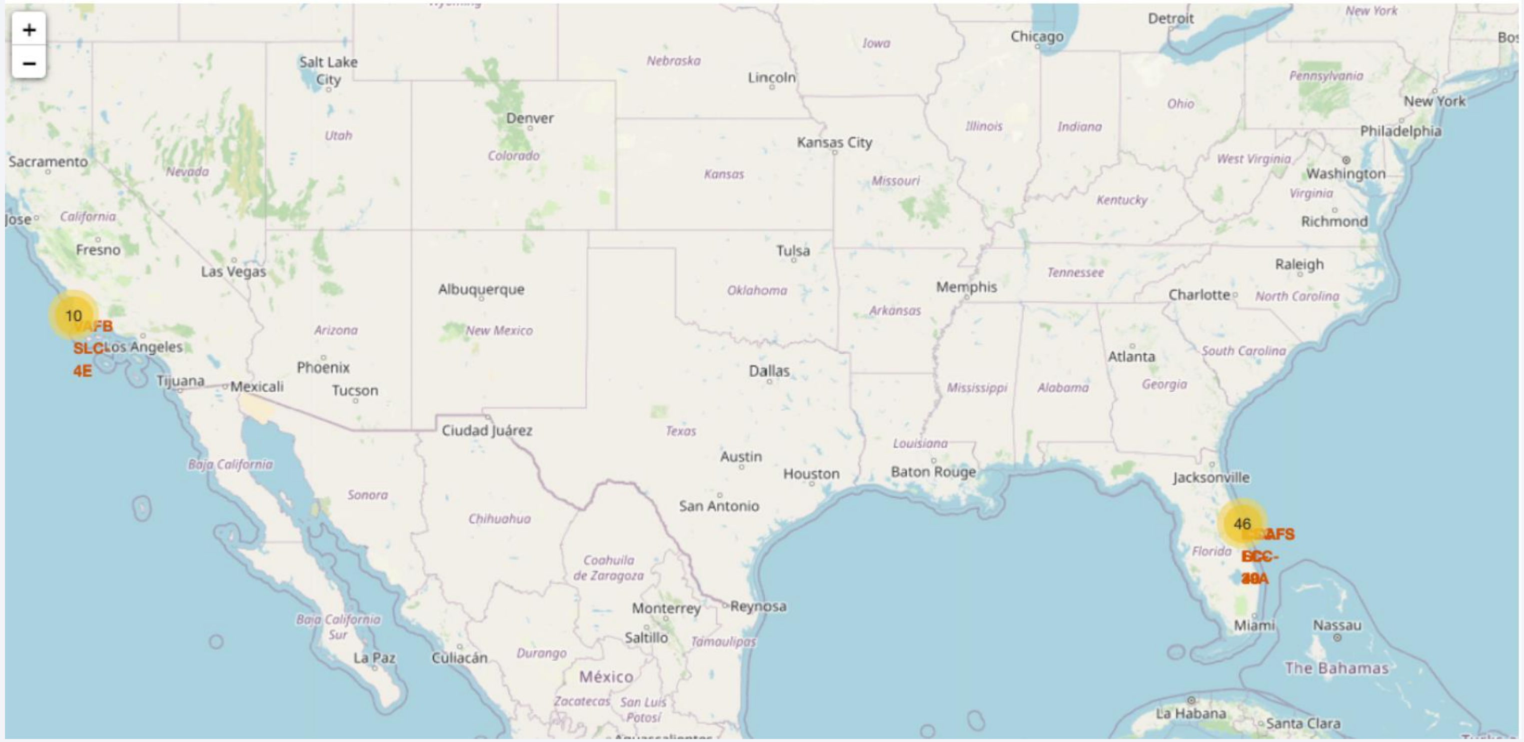
Pay Load vs. Flight No → Site vs. Flight No → Site vs. Pay Load → Success Rate → Feature Engineering

- Github link:

https://github.com/YABSHD/-spaceX-landing-success-prediction/blob/main/5_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

Build an Interactive Map with Folium

- The following data is mapped using Folium:
  - All launch sites on a map
  - the success/failed launches for each site
  - Calculate the distances between a launch site to its proximities



Build an Interactive Map with Folium

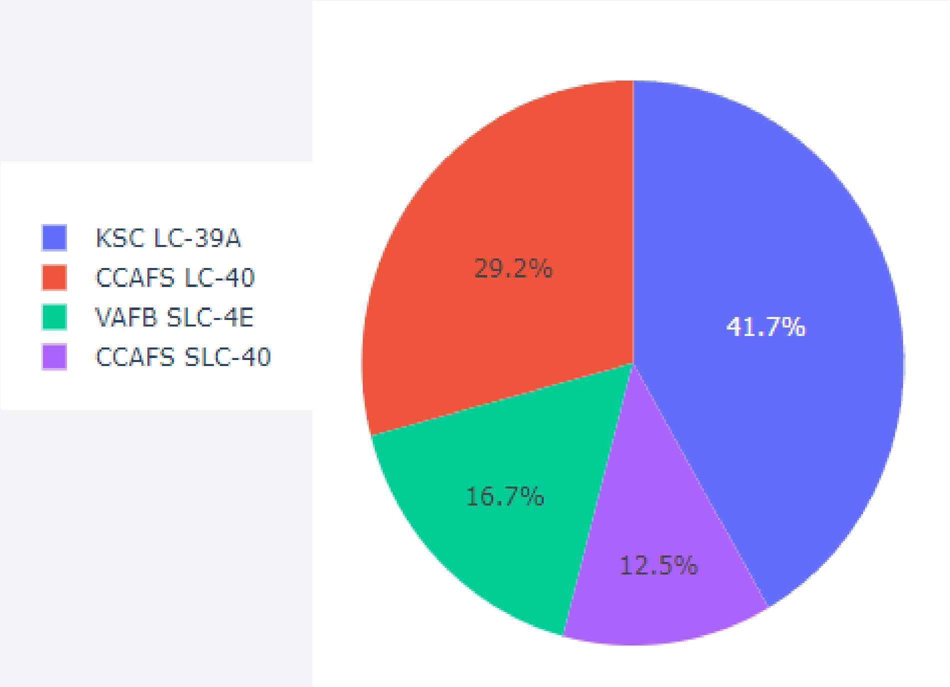
- By adding these objects, following geographical patterns about launch sites are found:
    - Are launch sites in close proximity to railways? Yes
    - Are launch sites in close proximity to highways? Yes
    - Are launch sites in close proximity to coastline? Yes
    - Do launch sites keep certain distance away from cities? Yes

- Github link:

https://github.com/YABSHD/-spaceX-landing-success-prediction/blob/main/6_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- The dashboard is divided into two parts:
  - Pie chart for the successful launch by each site
  - Scatter point for the relationship between landing outcomes and the payload



Predictive Analysis (Classification)

- Fix the data to be utilized to machine learning algorithms:
  - create a column for the class
  - Standardize the data
  - Split into training data and test data
  - Find best Hyperparameter for SVM, Decision Trees, K-Nearest Neighbours and Logistic Regression.
  - Find the best algorithm to classify by calculating Test Data Accuracy

- Github link:

https://github.com/YABSHD/-spaceX-landing-success-prediction/blob/main/7_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

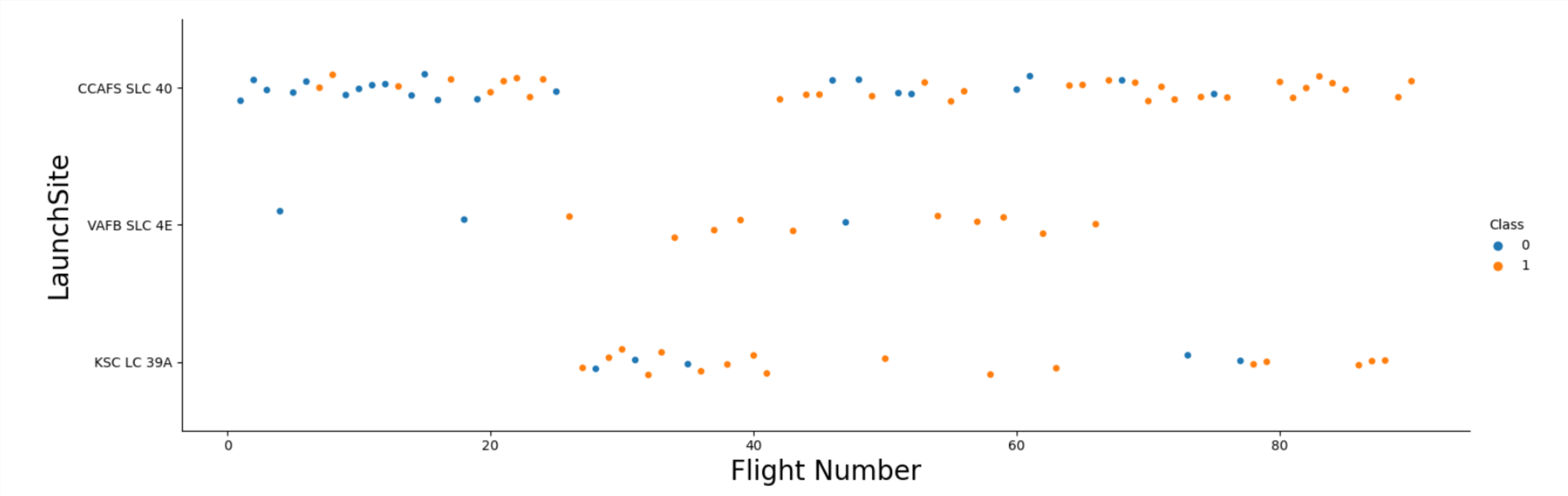
- The results of the exploratory data analysis revealed that the success rate of the Falcon 9 landings was 66.66%
- The Machine Learning predictive analysis showed that the highest accuracy classification model is Decision Tree with accuracy 94.4%

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

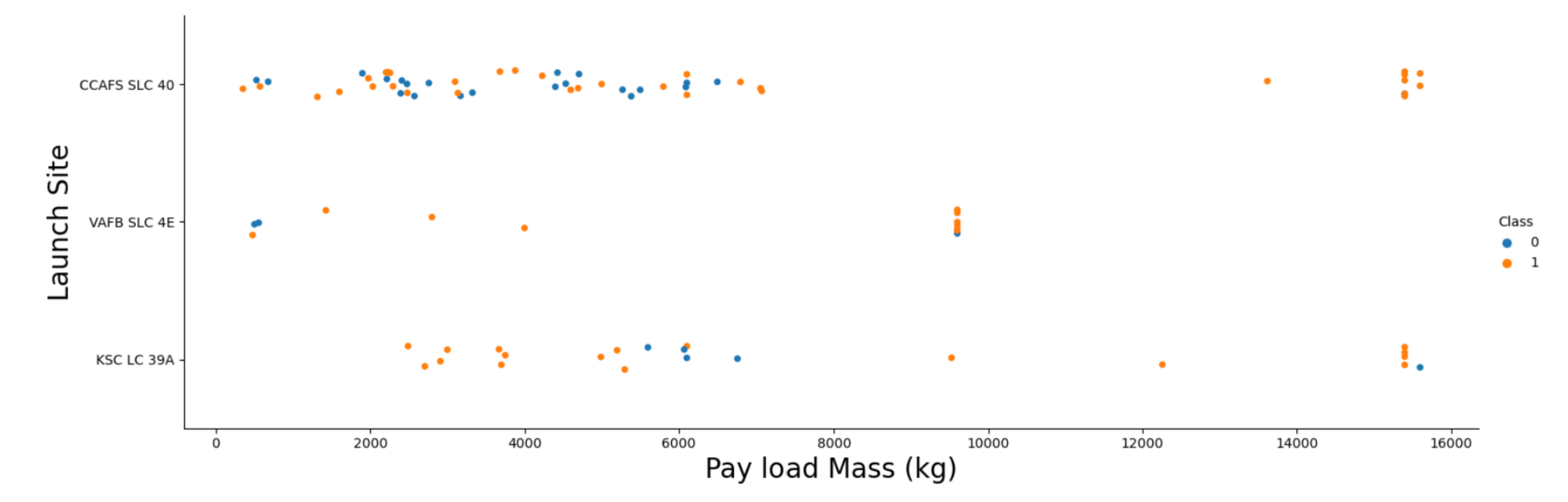
- The figure represent failed mission with blue dots and successful mission with orange dots.
- It also shows that the success rate increased as the number of flights increased.



scatter plot of Flight Number vs. Launch Site

Payload vs. Launch Site

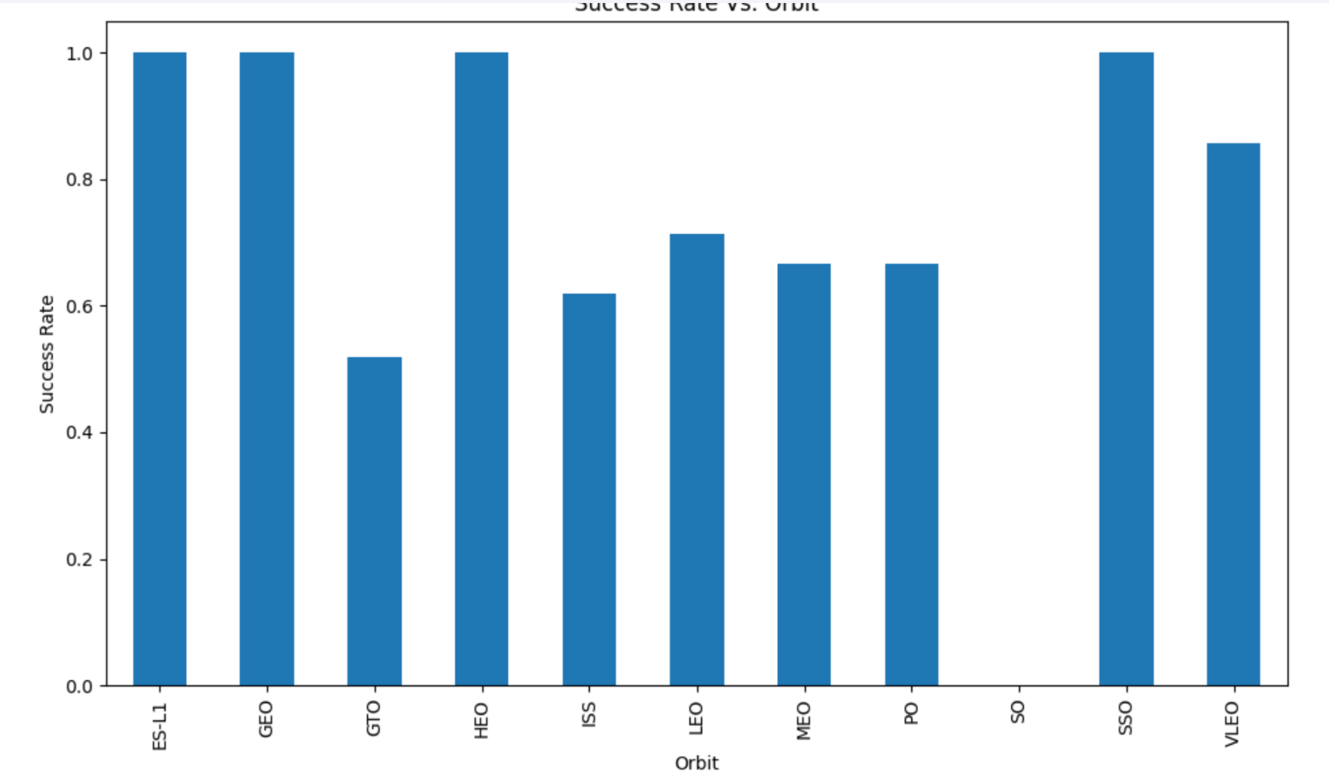
- The figure represent failed mission with blue dots and successful mission with orange dots.
- It shows no clear relationship between launch site and payload.



scatter plot of Payload vs. Launch Site

Success Rate vs. Orbit Type

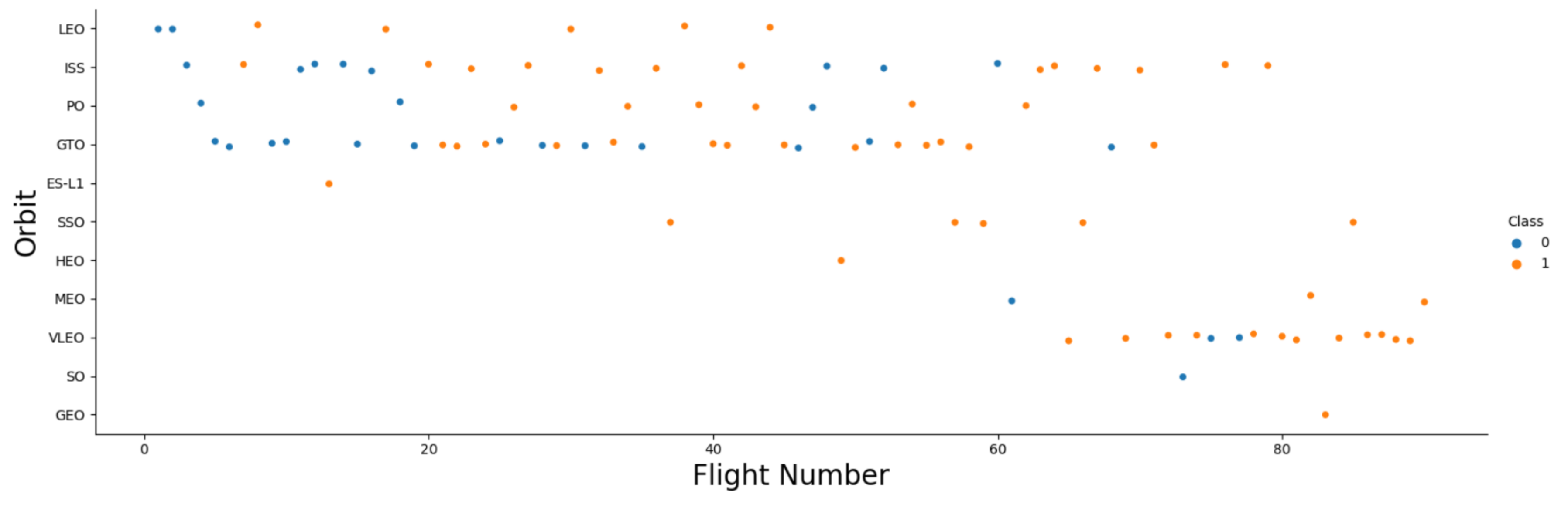
- Four orbits have 100% success rate (ES-11, GEO, HEO, SSO)



success rate of each orbit type

Flight Number vs. Orbit Type

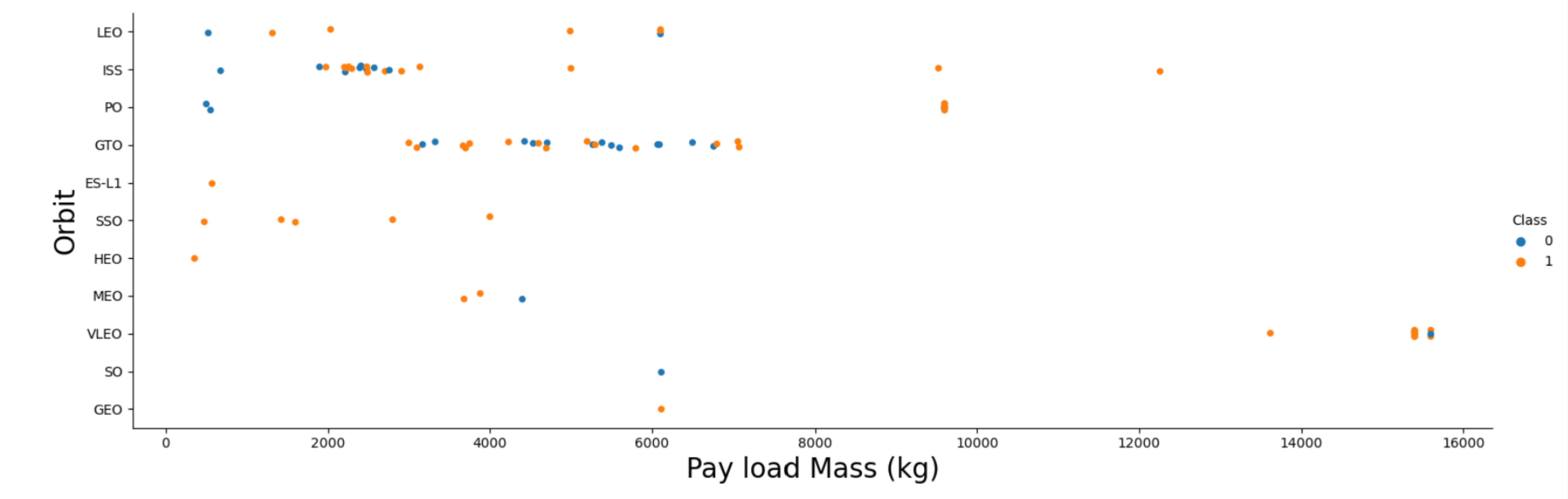
- The figure represent failed missions with blue dots and successful missions with orange dots.
- The success rate increase after 20 flights.
- Increase in flight number lead to increase in success rate for most orbits.



Scatter point between Flight number vs. Orbit

Payload vs. Orbit Type

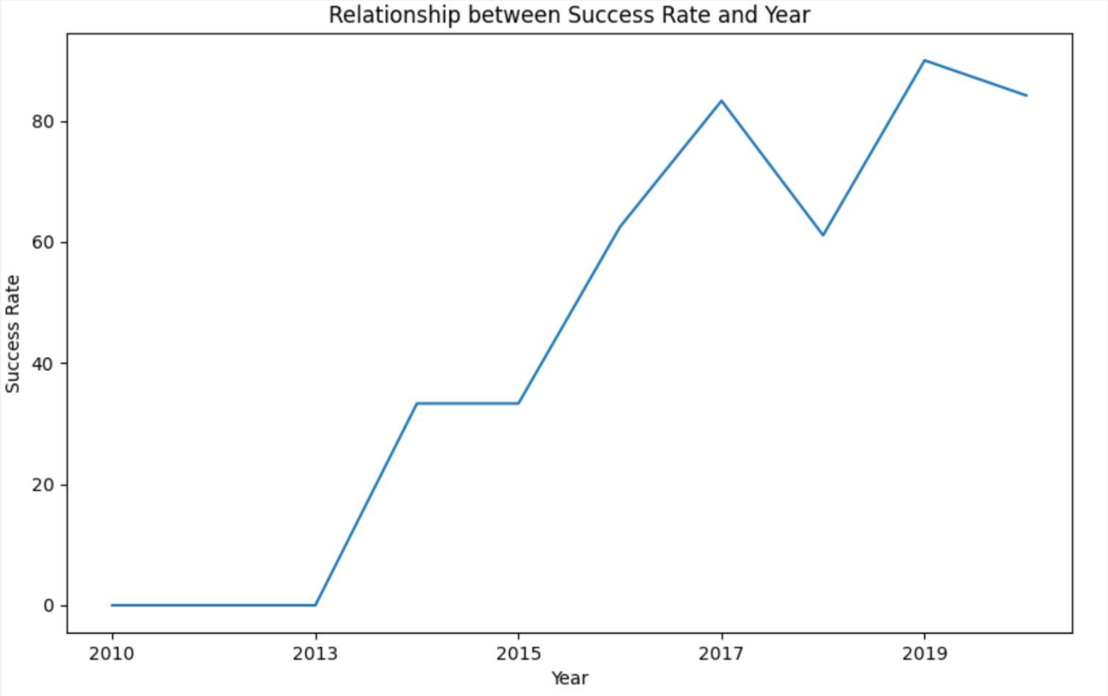
- The figure represent failed missions with blue dots and successful missions with orange dots.
- There is relationship between the payloads and orbits. As the heavier pay load, the success rate increases in most orbits.



scatter point of payload vs. orbit type

Launch Success Yearly Trend

- There is a correlation between success rate and year pass. the success rate increase with years except 2018 and 2020.



line chart of yearly average success rate

All Launch Site Names

- The names of the unique launch sites:

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

```
%sql select DISTINCT(Launch_Site) from SPACEXTBL
```

```
 * 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'

- Find 5 records where launch sites 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
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outco
)6/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachu
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachu
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No atter
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No atter
)3/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No atter

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

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

```
%sql select TOTAL(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer = 'NASA (CRS)'
```

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

TOTAL(PAYLOAD_MASS__KG_)
45596.0

Average Payload Mass by F9 v1.1

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

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

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version = 'F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

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

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

Hint:Use min function

```
%sql select min(Date) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

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

min(Date)
01/08/2018

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

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 distinct(Booster_Version)
from SPACEXTBL
where PAYLOAD_MASS__KG_ between 4000 and 6000
AND Landing_Outcome LIKE '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

- Calculate the total number of successful and failure mission outcomes

List the total number of successful and failure mission outcomes

```
%%sql select  Mission_Outcome, count(*) number_of_mission
from SPACEXTBL
group by Mission_Outcome
```

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

Mission_Outcome	number_of_mission
None	898
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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)
```

```
 * 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

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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,4,2) as month, BOOSTER_VERSION, LAUNCH_SITE, Landing_Outcome
FROM SPACEXTBL where Landing_Outcome = 'Failure (drone ship)' and substr(Date,7,4)='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)

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

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

```
%%sql select LANDING_OUTCOME, count(LANDING_OUTCOME) from spacextbl
where Mission_Outcome = 'Success'   and  "Date" between '04-06-2010' and '20-03-2017'
GROUP BY  LANDING_OUTCOME ORDER BY COUNT(LANDING_OUTCOME) DESC;
```

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

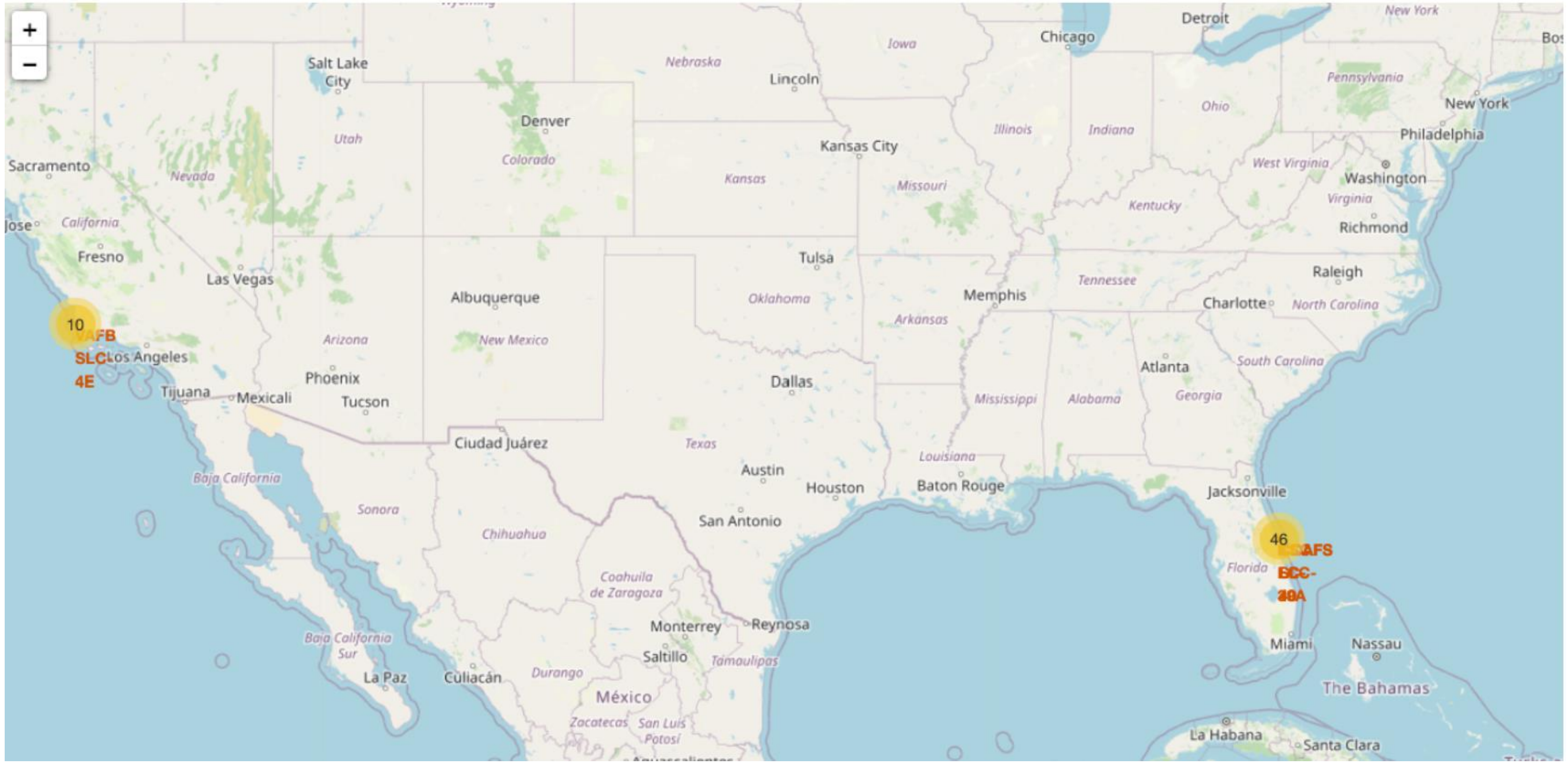
Landing_Outcome	count(LANDING_OUTCOME)
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

Section 3

Launch Sites Proximities Analysis

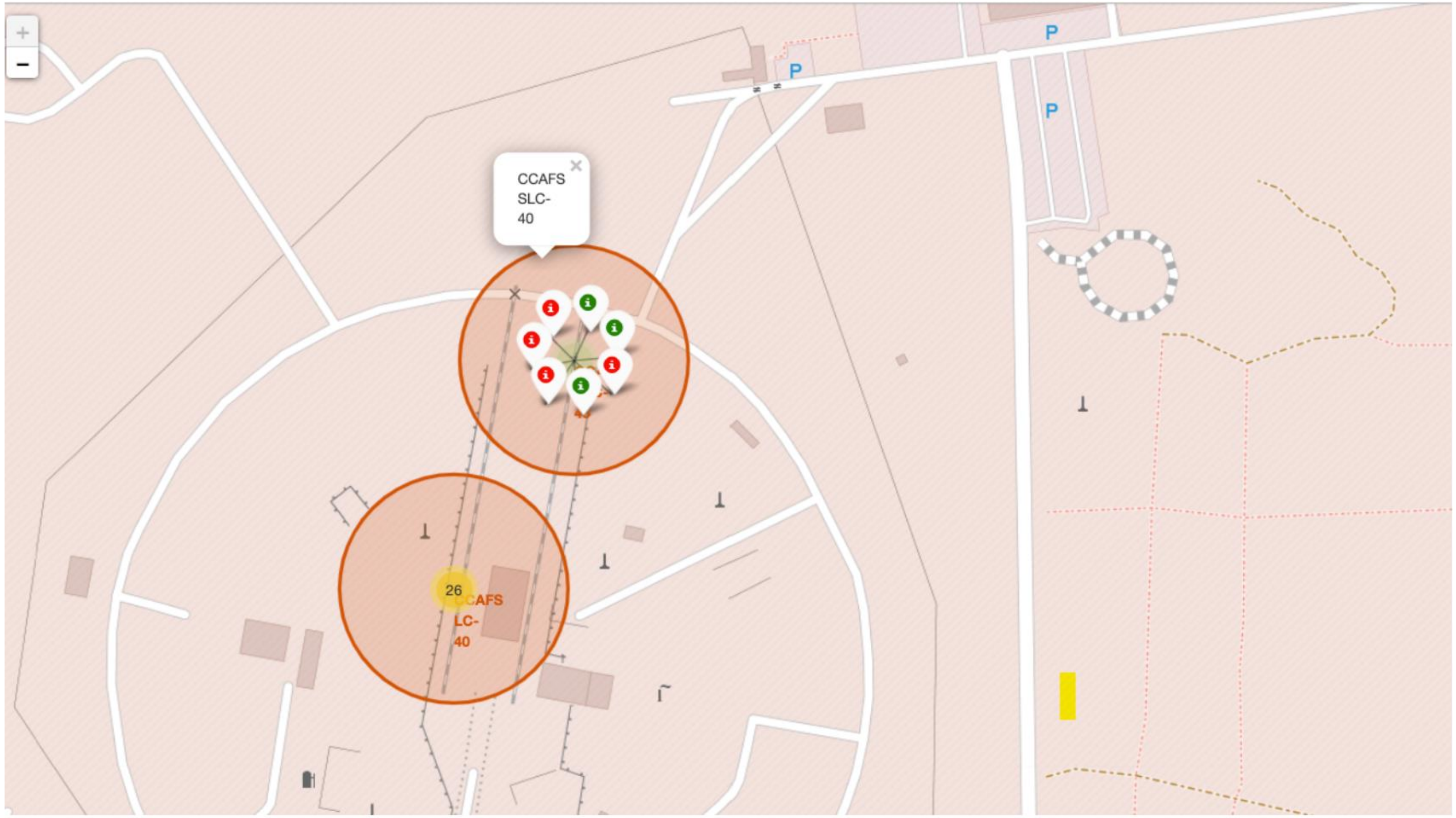
Build an Interactive Map with Folium

- The yellow markers are indicators of where the locations of all the SpaceX launch sites. The location of the sites are near coastlines and far from cities.



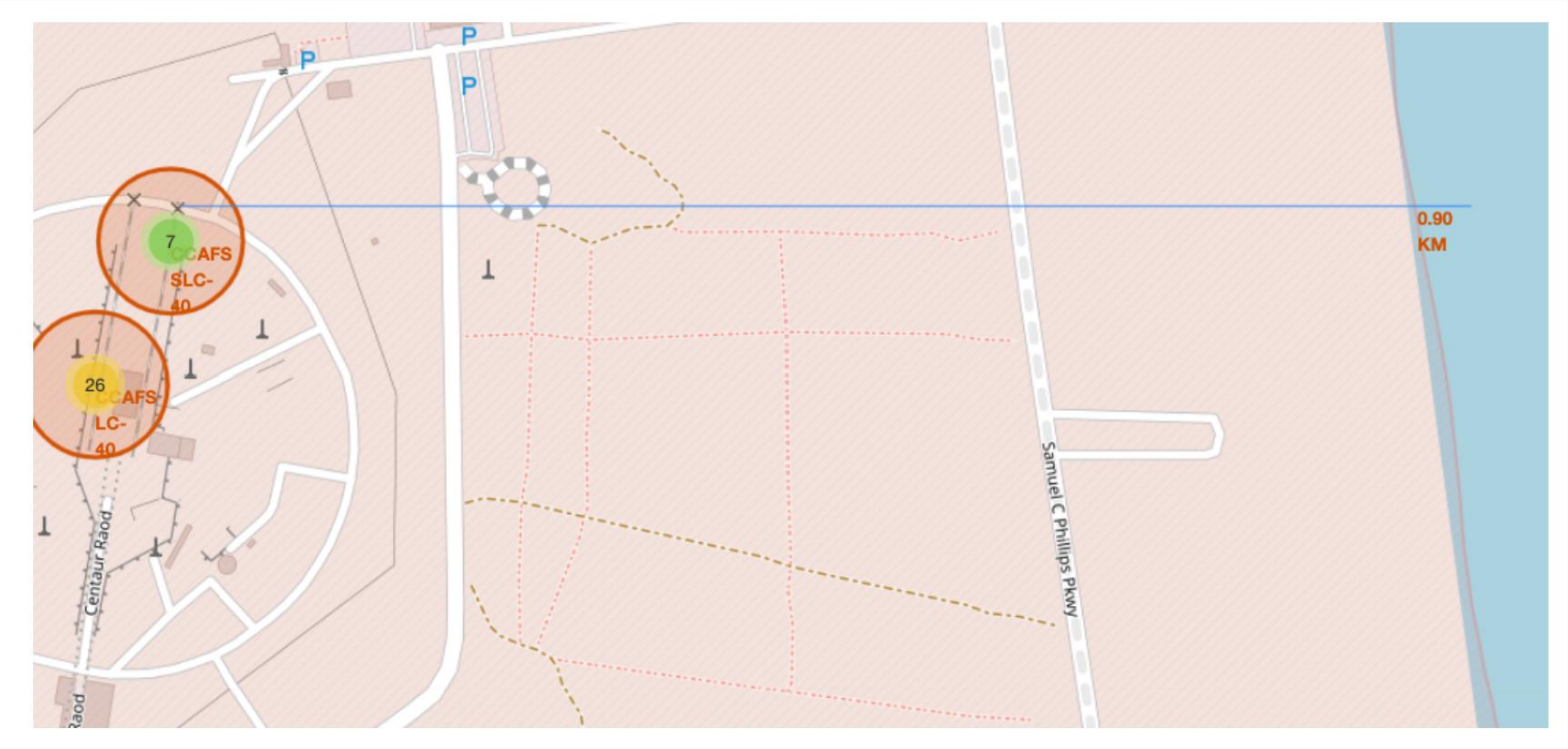
Build an Interactive Map with Folium

- When click on a launch site a clusters of successful landings (green) or failed landing (red) showing on the map.



<Folium Map Screenshot 3>

- All sites are close to coastlines, and transportations.

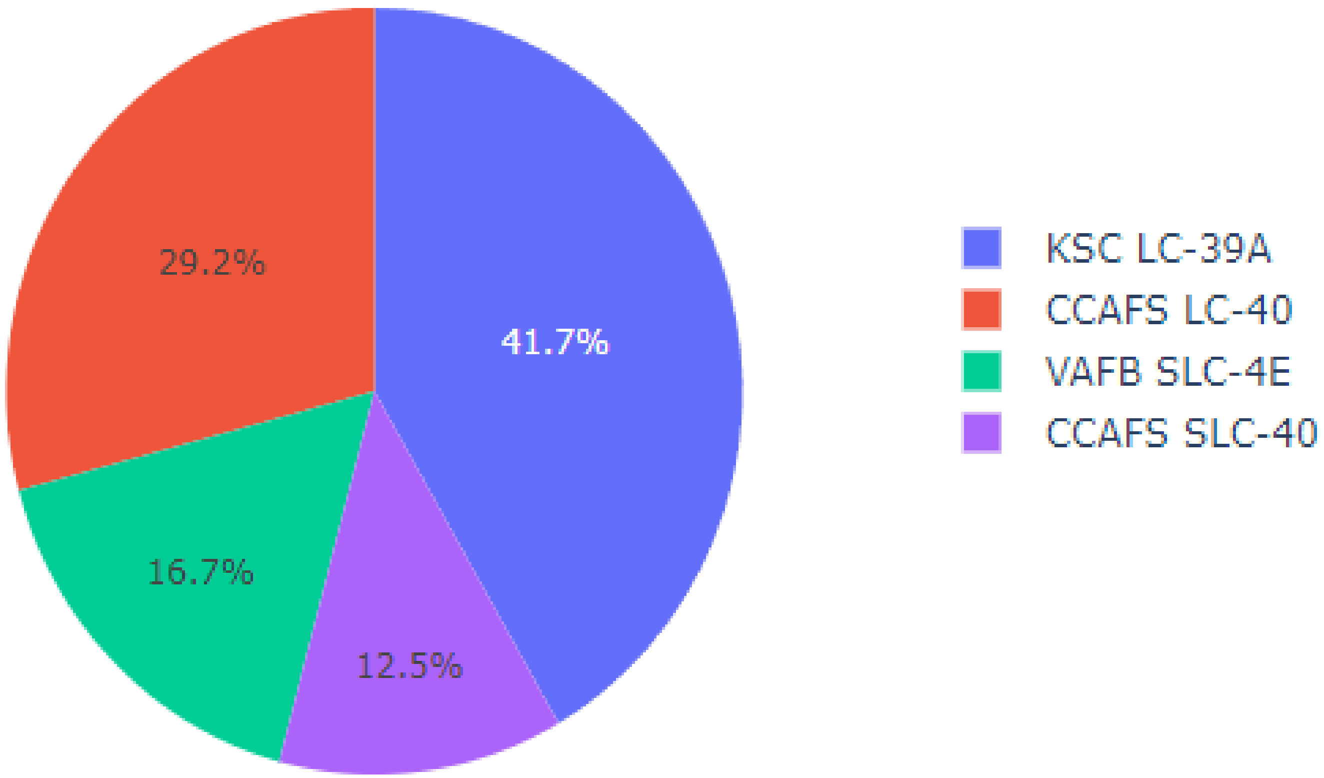


Section 4

Build a Dashboard with Plotly Dash

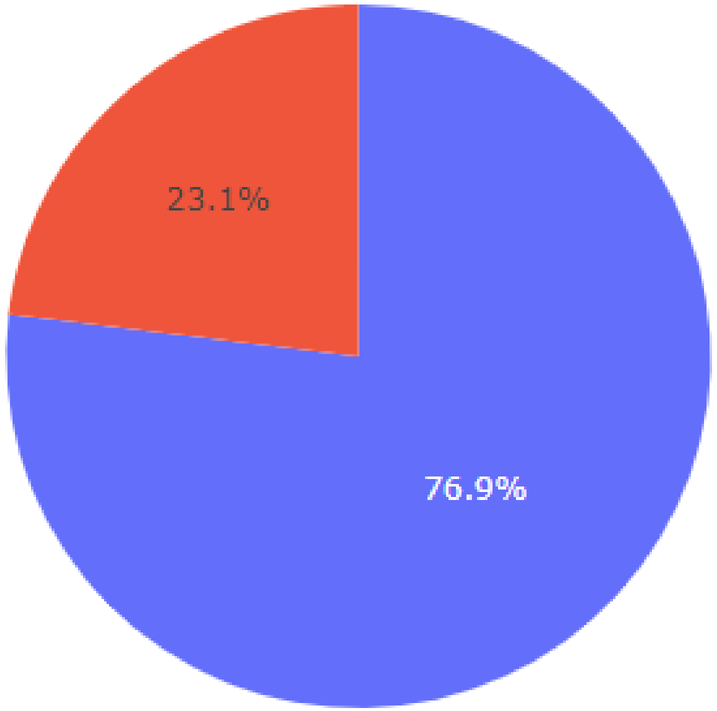
Most successful launches

The most successful launches is **KSC LC-39A**.

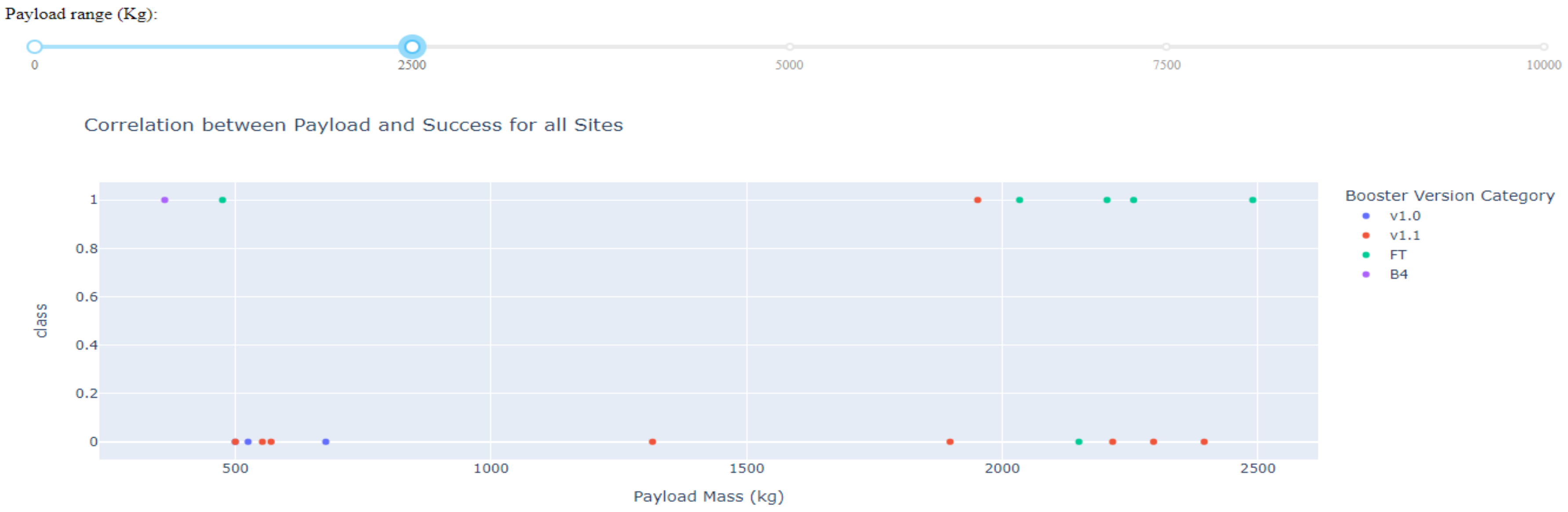


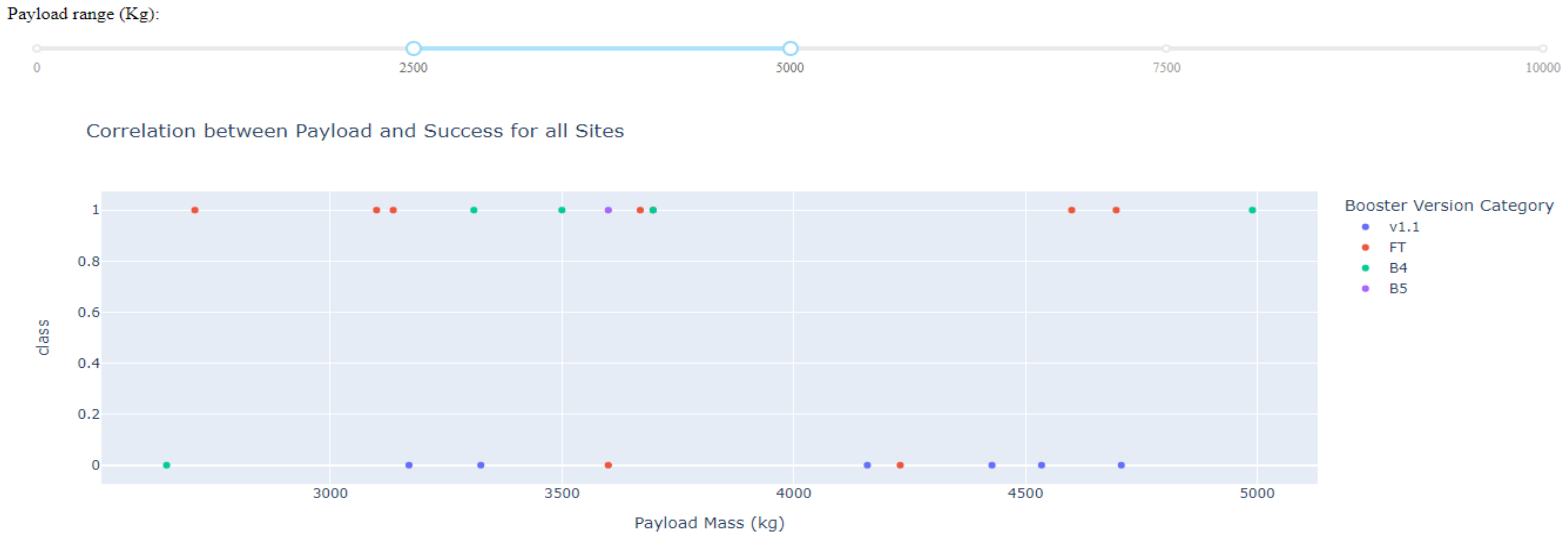
Highest success rate

The **KSLC-39A** has the highest success rate with 76.9%.



Payloads vs Launch Outcome



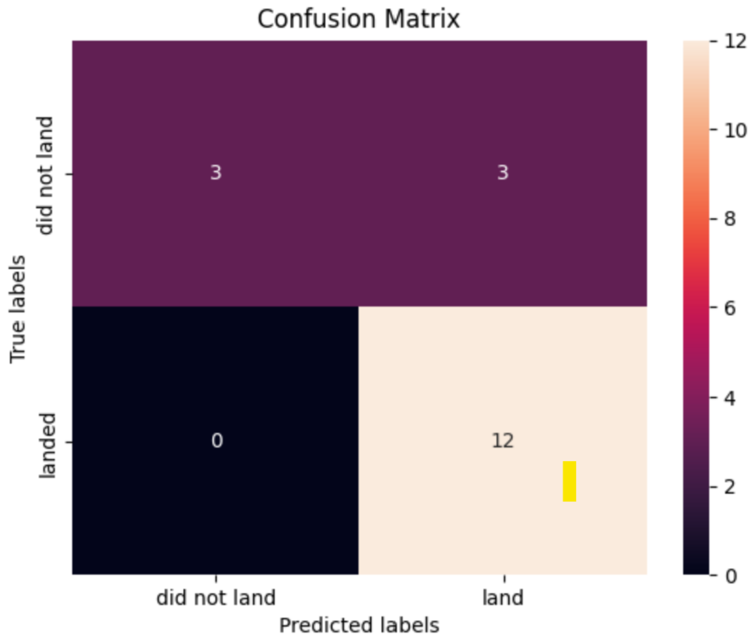


Section 5

Predictive Analysis (Classification)

Predictive Analysis (Classification)

- Fix the data to be utilized to machine learning algorithms:
  - create a column for the class
  - Standardize the data
  - Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression



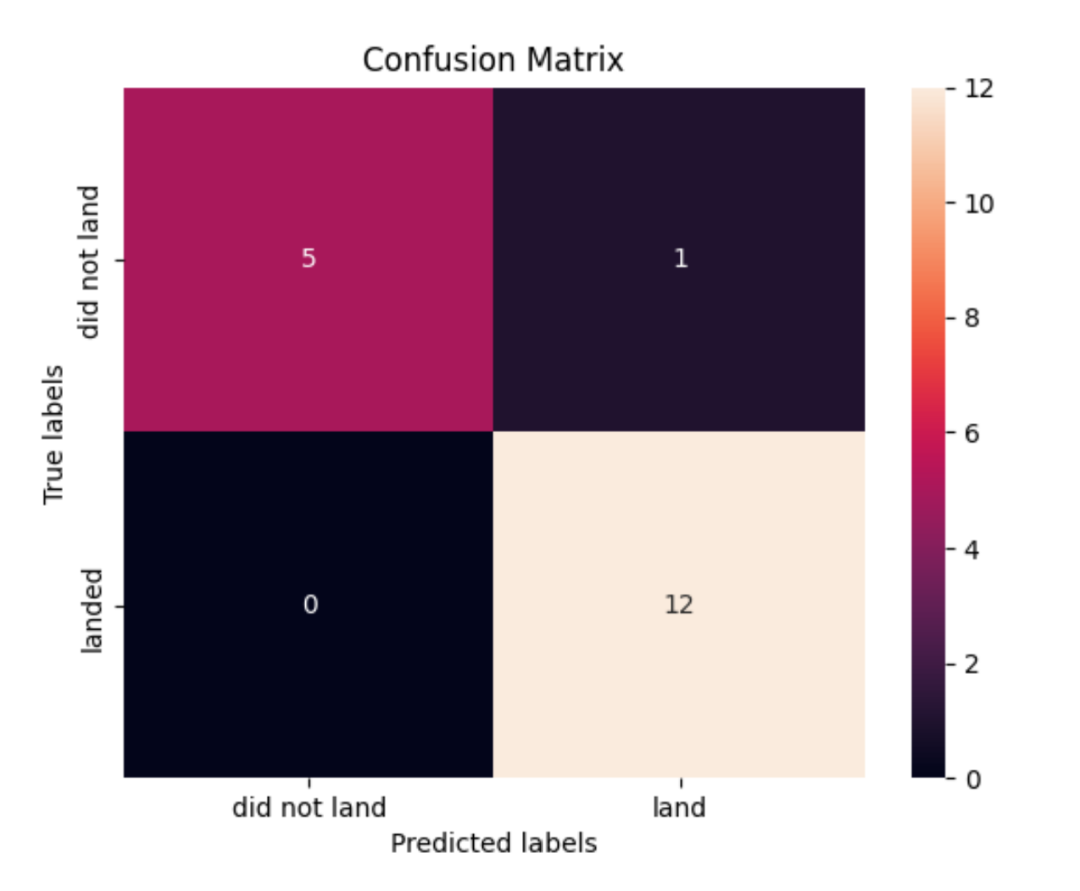
Classification Accuracy

- The highest accuracy classification model is Decision Tree with accuracy 94.4%

Method	Logistic_Reg	SVM	Decision Tree	KNN
Test Data Accuracy	0.833333	0.833333	0.944444	0.833333

Confusion Matrix

- The confusion matrix of the decision tree model



- The model predicted 12 successful landings when the True label was successful.

- The model predicted 1 successful landings when the True label was unsuccessful landing.

Conclusions

- The results of the exploratory data analysis revealed that the success rate of the Falcon 9 landings was 66.66%
- The EDA showed that the success rate increase over the years and there is a positive correlation between number of flights and success rate.
- Most successful lunches are to orbits SSO, GEO, HEO, & ES-L1.
- The Machine Learning predictive analysis showed that the highest accuracy classification model is Decision Tree with accuracy 94.4%

Appendix

- The project GitHub: https://github.com/YABSHD/-spaceX-landing-success-prediction/tree/main

- SQL Help: https://www.w3schools.com/sql/

Thank you!