Data Science Capstone project

<SALAMI, Olatunde Stephen>

<August 10, 2021>

Outline

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

Executive Summary

- Summary of methodologies:

SpaceX's Falcon9 rocket launch trip is advertised as 165 million on their website, this is at 3x lesser than there contemporaries. The low pricing from SpaceX is because they can reuse the first stage, which often depends on a successful falcon9 landing.

In this report, I collected, processed, and analysed historic Falcon9 rocket launches, then predicted using Machine Learning approach, the possible outcome from the first launch.

- Summary of all results:

- A successful rocket launch is depended on many factors such as payload mass, orbit type, launch sites, etc.
- Advancement in the technology is aiding improved research, thus contributing to closing the knowledge gap, thus improving the success rate year on year.

Introduction

- Project background and context

❖ Falcon9 is part of the falcon rocket family. It is designed and operated by private manufacturer SpaceX. They are mainly used for space exploration such as; cargo delivery, crewed flights to the International Space Station (ISS), etc. (Wikipedia).

❖ Since June 2010, rockets from the falcon9 family have been launched 126 times, with 124 full mission successes, one partial failure and one total loss of spacecraft (Wikipedia), hence they have a good record of launches.

❖ The landing performed on Falcon9 after there launches is usually controlled, hence there high success landing rate. This allows for of reusability, and thus low fare pricing compared to there competitors.

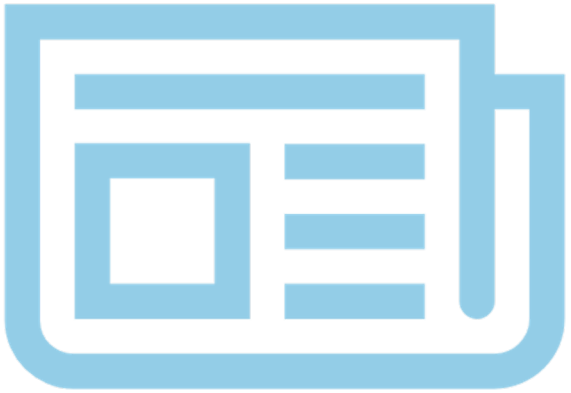
- Problems to be answered

For other manufacturers to stand a chance against SpaceX, historical falcon9 launch data was analyzed for key questions such as:

- What are the principal factors for a successful landing?
- What factors are associated with failure landing?

If known, Information on a possible successful rocket landing can be used by other rocket companies to bid against SpaceX's rocket launch pricing.

Methodology



- Data collection methodology:
  - Describes how data were collected
- Performed data wrangling
  - Describe how data were processed
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - How to build, tune, evaluate classification models

Methodology

Data collection

- How data sets was collected.

Data used for this analysis was collected using two different approaches:

- Through REST API request made directly to the SpaceX API
- Web scraping the Wikipedia page for relevant Falcon9 historical launch table

Here’s a Github link to the complete notebook of the data collection

Data collection – REST API

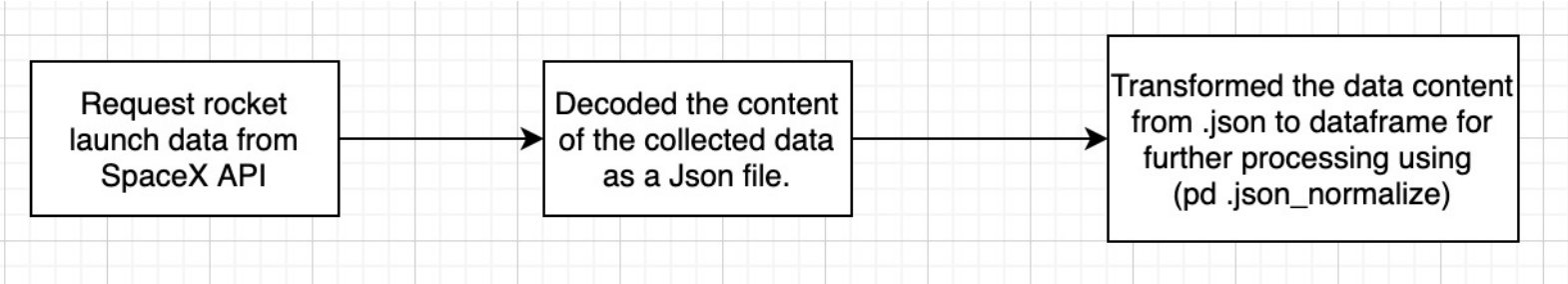
- Initiated a request to SpaceX API

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

- Then converted the response content into a pandas dataframe for Cleaning, and further processes

```
response.json()
data = pd.json_normalize(response.json())
```

- Below is a flowchart of the collected process.



Here's a Github link to the complete notebook of the data collection

Data collection – Web scraping

- The webscraping was done using BeautifulSoup (a python package)

```
!pip3 install beautifulsoup4
!pip3 install requests
```

- Extracted a falcon9 launch record HTML table from Wikipedia

```
html_data = requests.get(static_url).text
```

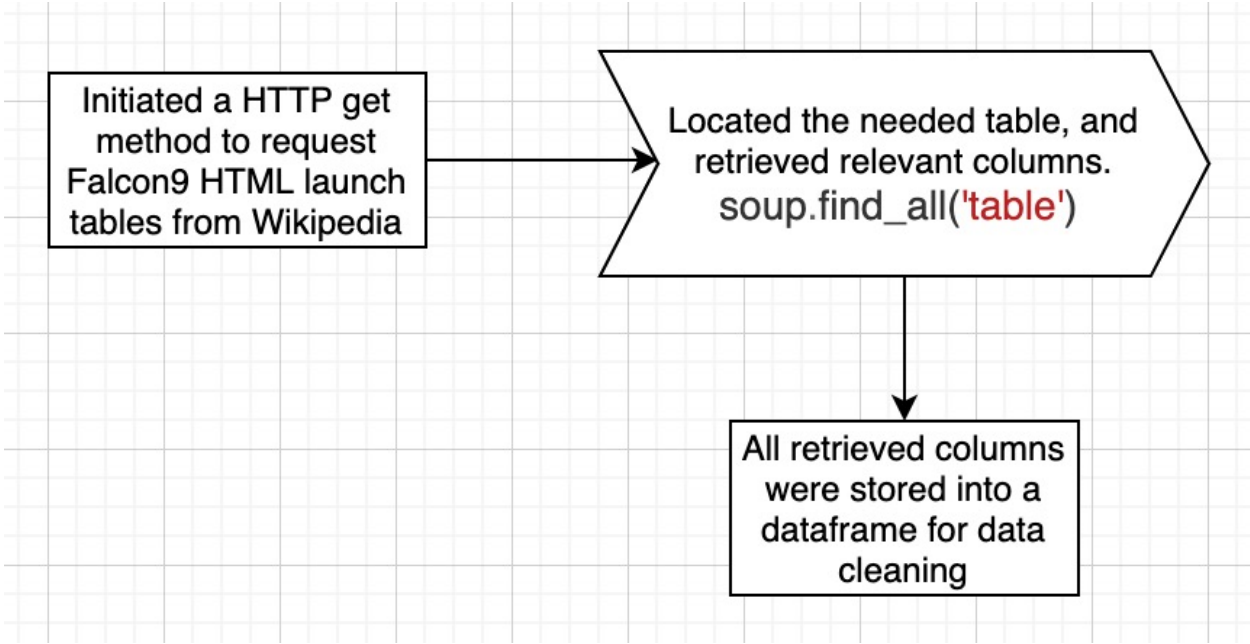
```
html_tables = soup.find_all('table')
```

- Parsed the table and converted it to a dataframe.

```
soup = BeautifulSoup(html_data, 'html5lib')
```

```
df=pd.DataFrame(launch_dict)
```

A flowchart of web scraping process.



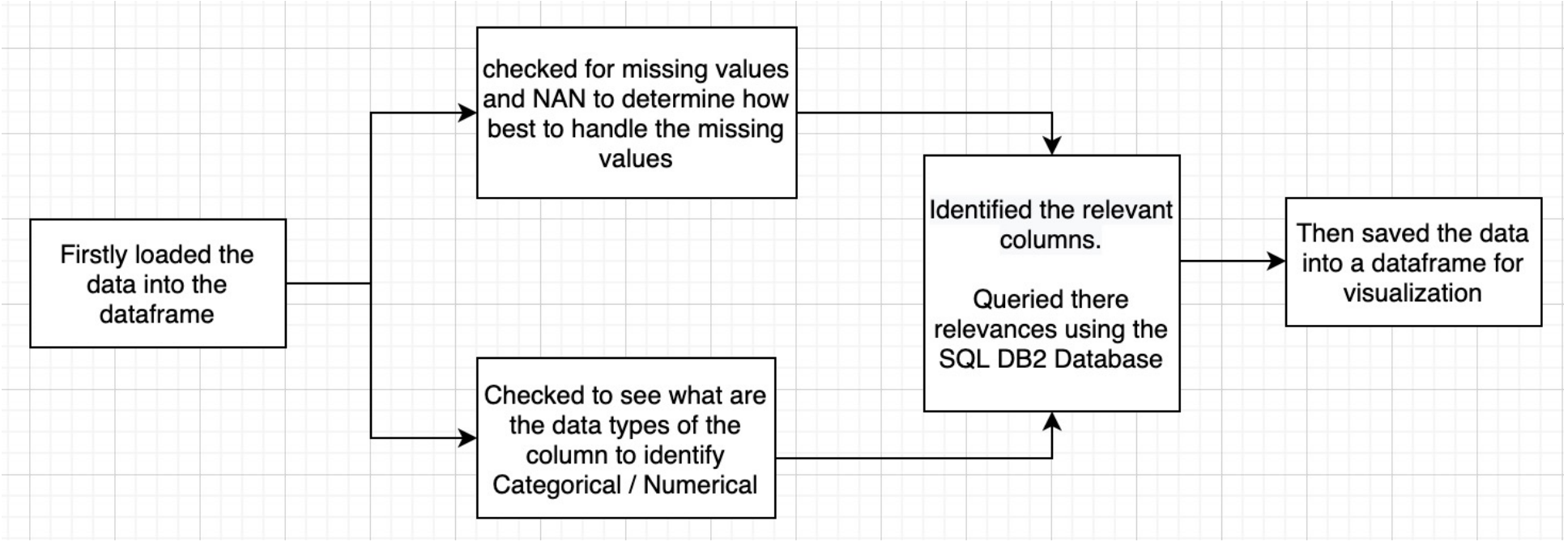
Here's a Github link to the complete notebook for the webscraping

Data wrangling

The objective of the data wrangling was to:-

* Understand the dataset
* Data cleaning, and identifying patterns in the dataset
* Determine the labels for training the machine learning model.

Here's a flowchart of the processes



Here's a Github link to the complete notebook for the data wrangling

EDA with SQL

- Summary of some performed SQL queries using bullet points
- Displayed the number of unique lunch sites on the space mission

```
%sql SELECT DISTINCT (launch_site) FROM SPACEXDATASET;
```

- Queried the average payload mass carried by Booster version F9 V1.1

```
%sql SELECT AVG(payload_mass__kg_) AS AVG_PAYLOAD_MASS FROM SPACEXDATASET WHERE booster_version = 'F9 v1.1'
```

- Queried for the first successful landing outcome

```
%sql SELECT MIN(Date) AS FIRST_SUCCESS_LANDING FROM SPACEXDATASET WHERE LANDING__OUTCOME LIKE '%Success (ground pa
d)%';
```

- Also queried for the total number of successful outcomes

```
%sql SELECT COUNT(MISSION_OUTCOME) AS SUCCESSFUL_OUTCOME FROM SPACEXDATASET WHERE MISSION_OUTCOME LIKE '%Success%';
```

Here's a Github link to the complete notebook for the data wrangling

EDA with data visualization

- Summary of what charts were plotted and why used those charts

❖ Plotted a catplot to discover the relationship between Flight-Number and Payload Mass, the essence to to observe how these variables would affect launch outcome.

❖ Plotted a chart to relationship between Flight Number and Launch Site, and to establish how launch sites vary with flight numbers.

❖ Also tried to observe if there's a relationship between payload mass and launch sites.

❖ Then I also plotted a chart to visualize relationship between success rate of each orbit type

❖ Also the relationship between payload mass and orbit type.

Here's a Github link to the complete notebook for the data visualization exercise

Build an interactive map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map

❖Created and added folium.Circle and folium.Marker for each launch site on the site map, to visualize the relative positions of the key locations

❖Created a new column in launch_sites dataframe called marker_color to store the marker colors based on the class value (successful/failed rocket landing).

❖Drew a PolyLine between launch sites and some neighbouring facilities to check for there proximities to public paces.

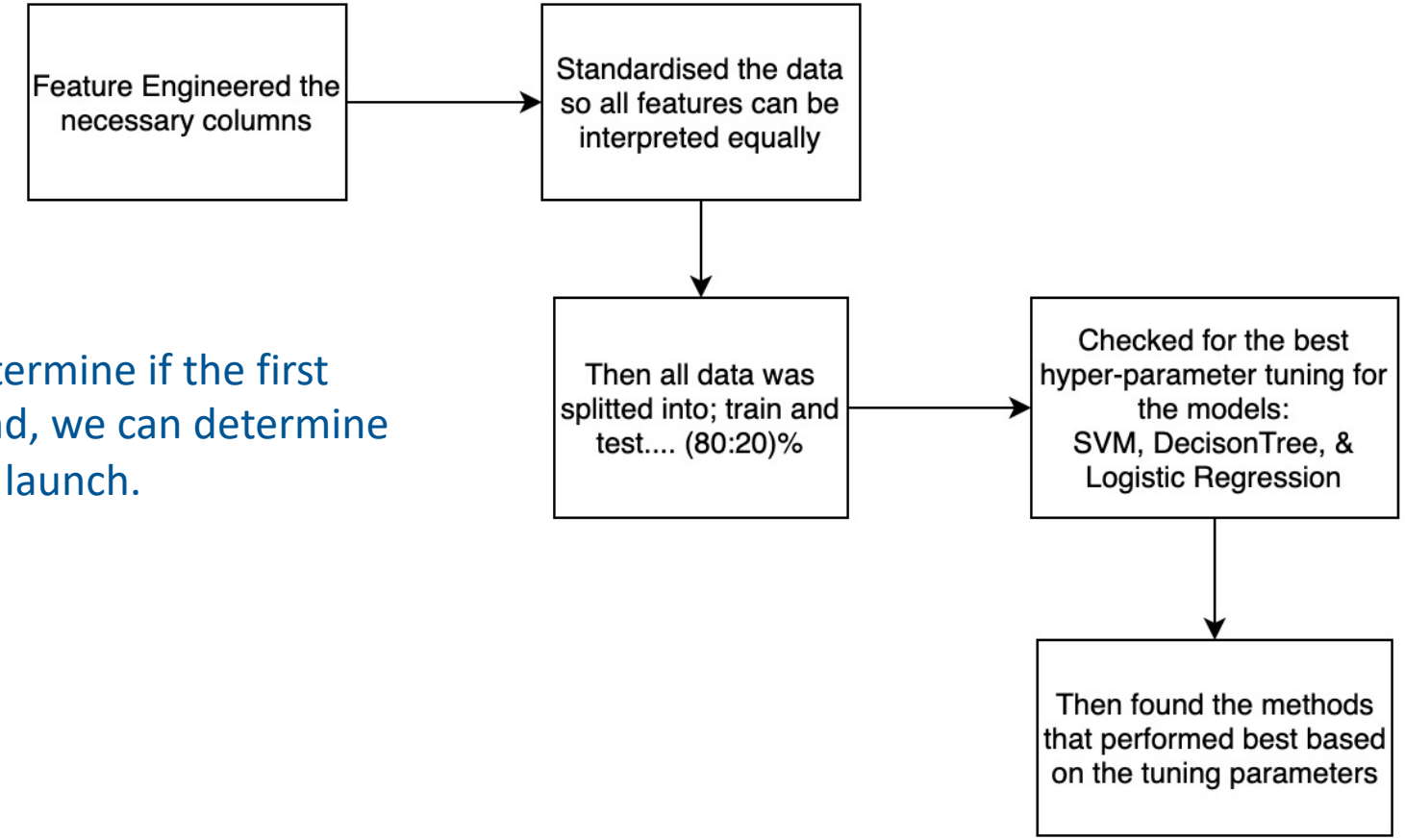
Here's a Github link to the complete notebook for these tasks.

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive analysis (Classification)

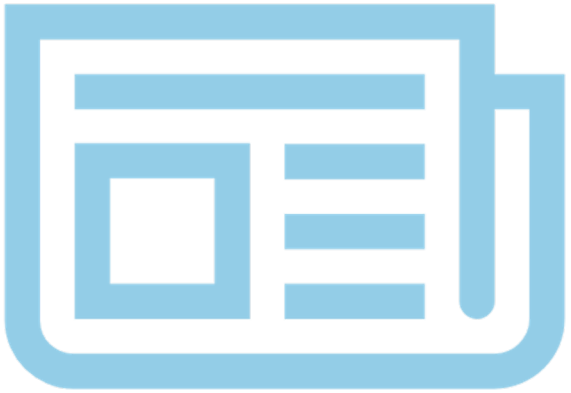
A machine learning pipeline to predict the outcome of the first stage landing given the data from the preceding labs.



If we can determine if the first stage will land, we can determine the cost of a launch.

Here's a Github link to the complete notebook for the model prediction.

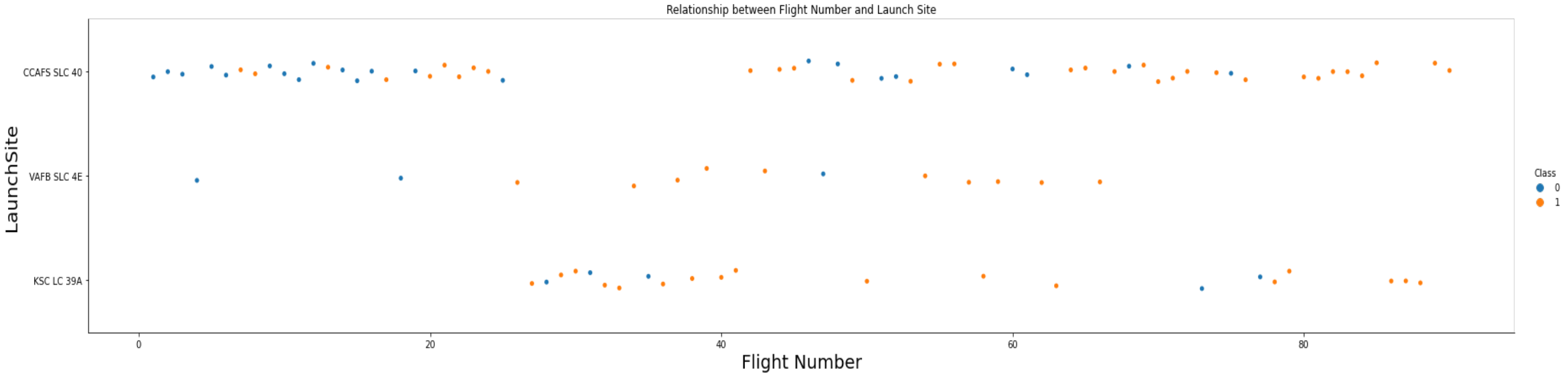
Results



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

EDA with Visualization

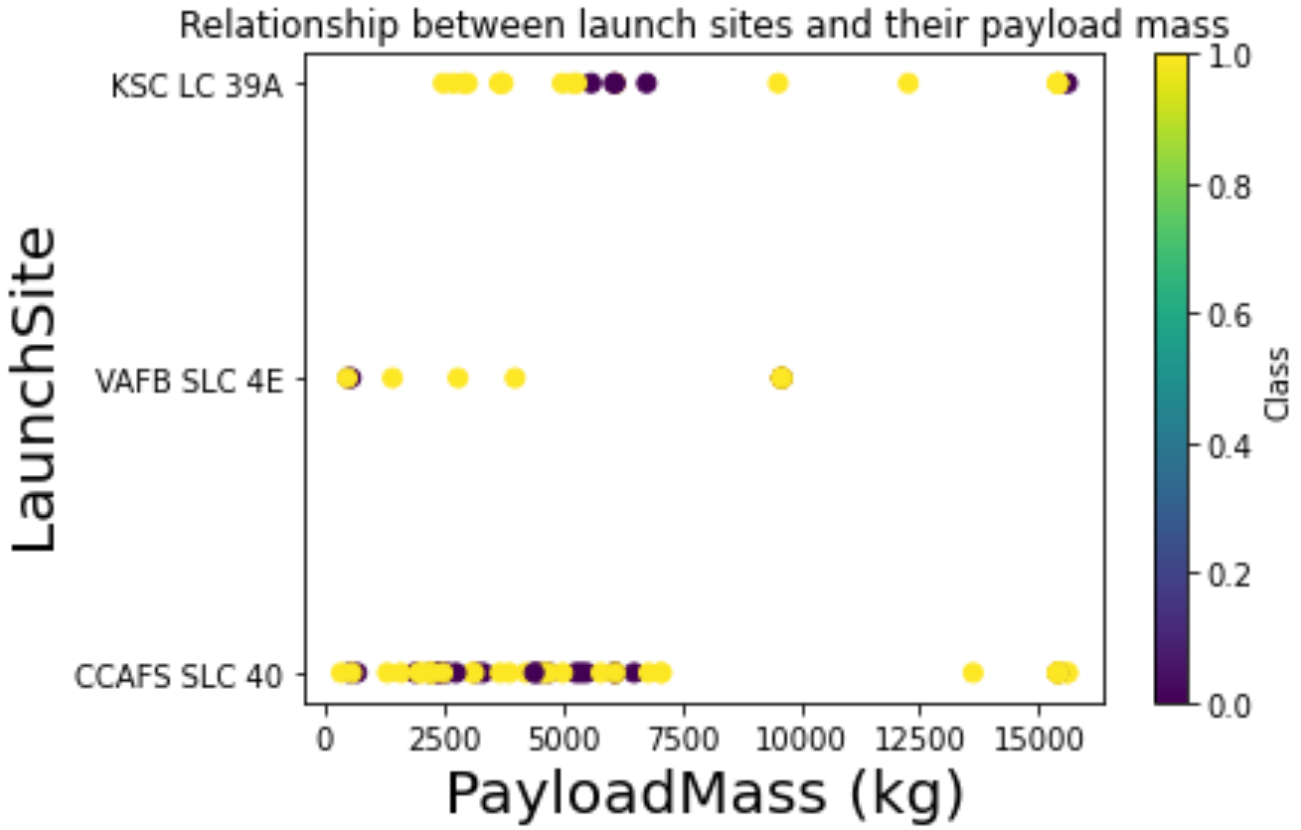
Flight Number vs. Launch Site



From the plot above, we can deduce that :

- majority of the flights took place at the CCAFS SLC40 launch site.
- That a huge number of flights from the VAFB SLC4E launch site were successful.

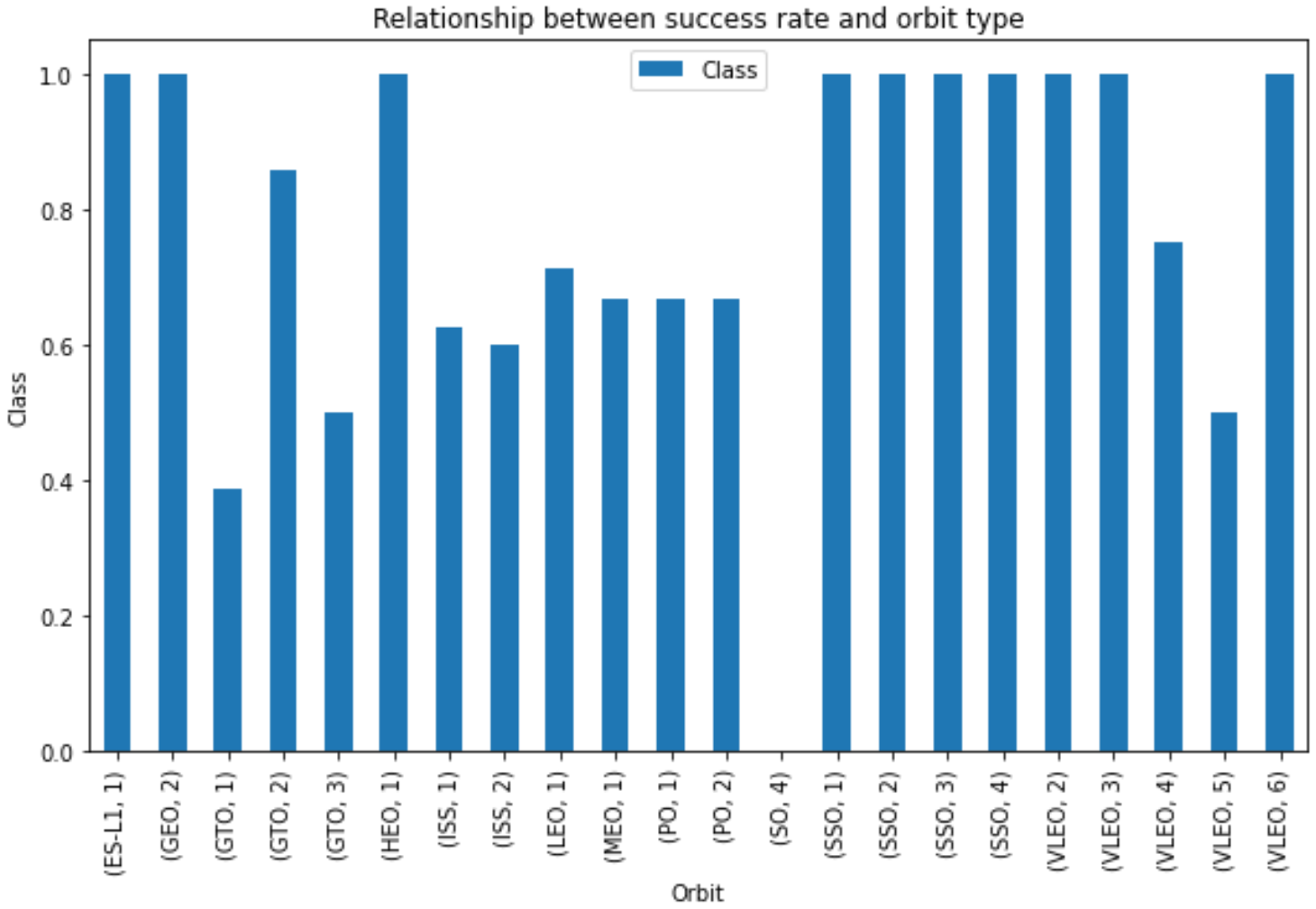
Payload vs. Launch Site



From the plot, we can deduce that :

- Flight from VAFB SLC4E: Have a max payload of 10,000kg.
- Most of the flights from the CCAFS SLC40 & KSC LC39A are concentrated within 0-7,500kg payload mass range.

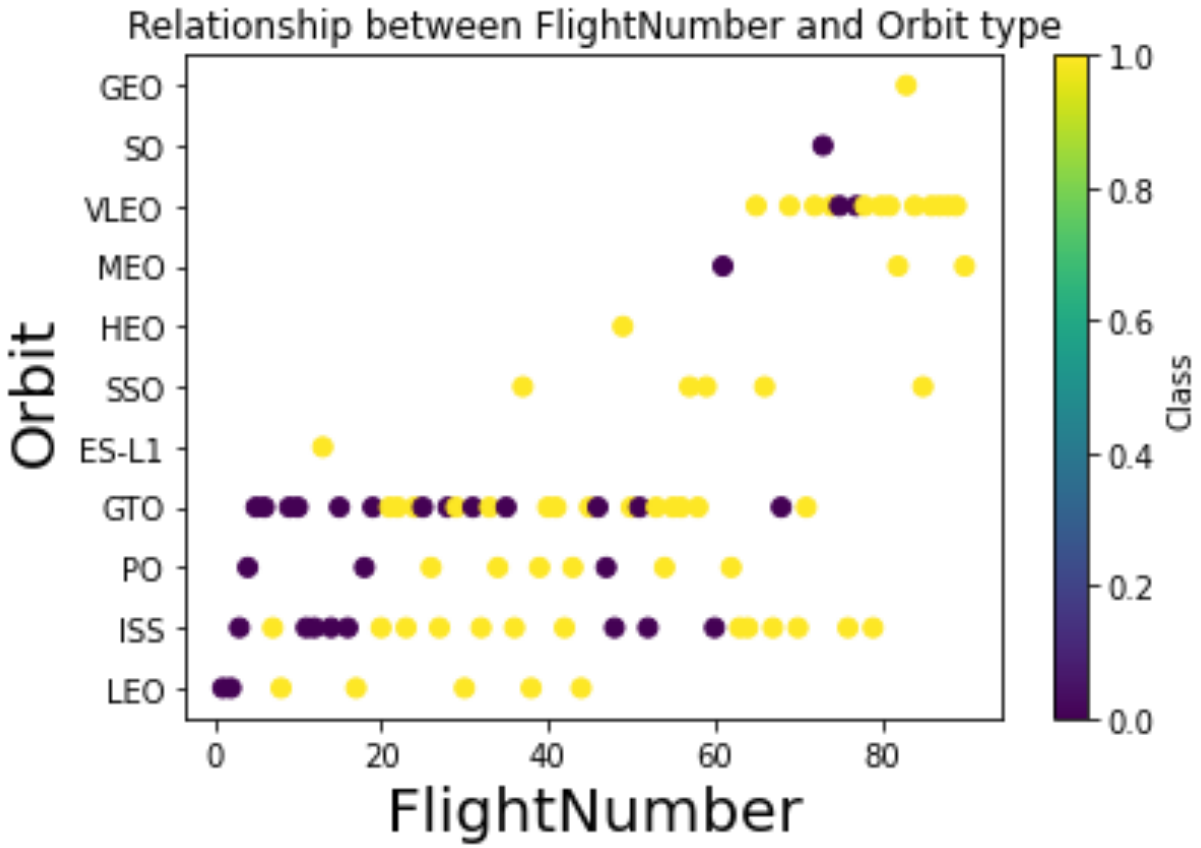
Success rate vs. Orbit type



From the plot, we can deduce that :

- Orbits with high success rate includes: ESL, GEO, HEO, SSO, & VLEO
- Also, one of the orbits with the lowest success rate is GTO

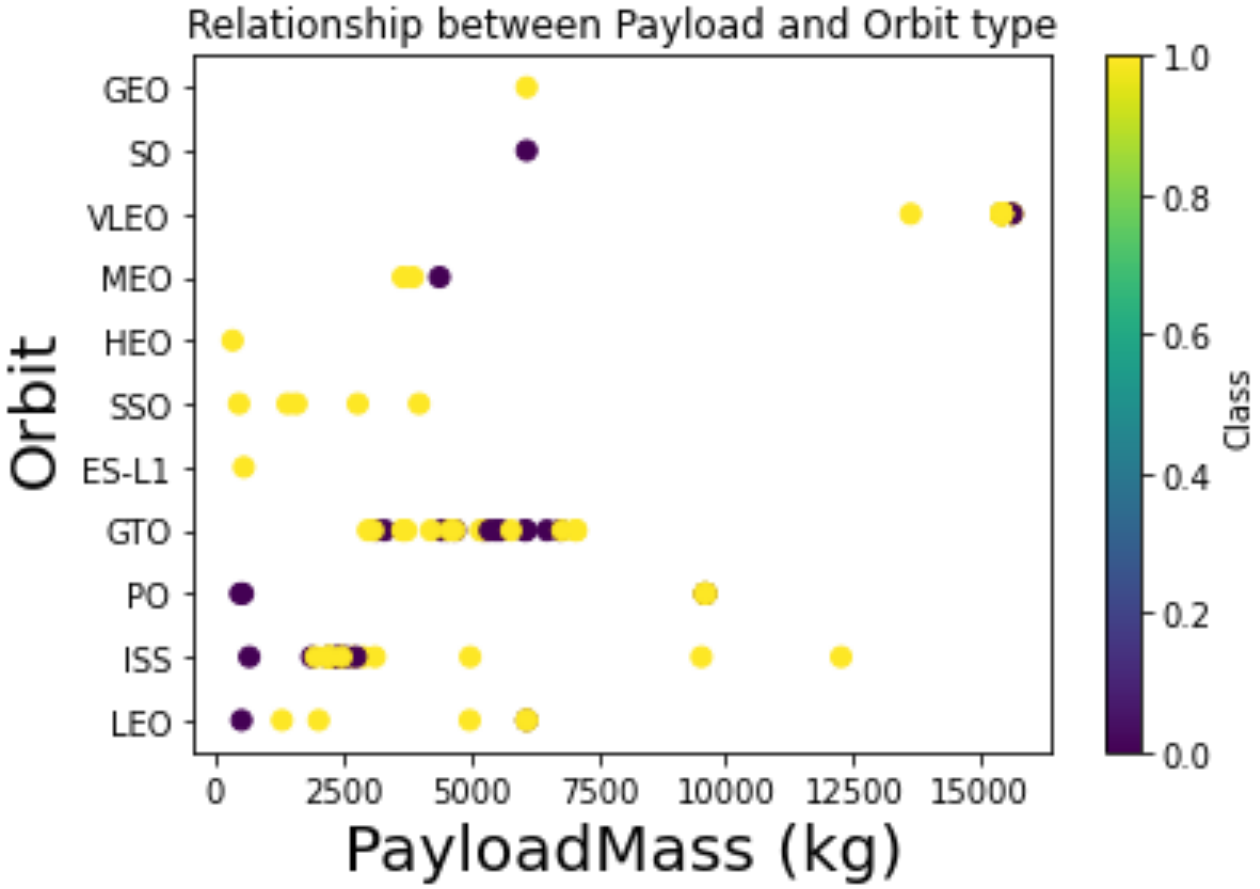
Flight Number vs. Orbit type



From the plot, we can deduce that :

- LEO orbit the Success appears related to the number of flights;
- on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit type



From the plot, we can observe that :

- Heavy payloads have a negative influence on GTO orbits and positive on PO and Polar LEO (ISS) orbits.

Launch success yearly trend

Show a line chart of yearly average success rate

Show the screenshot of the scatter plot with explanations

EDA with SQL

All launch site names

- Find the names of the unique launch sites

```
%sql SELECT DISTINCT (launch_site) FROM SPACEXDATASET;
```

- Query result with a short explanation here

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

The above query return with 5 unique launch sites from the SpaceX Dataset

Launch site names begin with `CCA`

- launch sites beginning with `CCA`

```
%%sql
SELECT BOOSTER_VERSION, CUSTOMER, launch_site, LANDING__OUTCOME, PAYLOAD_MASS__KG_
    FROM SPACEXDATASET
    WHERE launch_site LIKE '%CCA%'  LIMIT 5;
```

- Query result with a short explanation here

booster_version	customer	launch_site	landing__outcome	payload_mass__kg_
F9 v1.0 B0003	SpaceX	CCAFS LC-40	Failure (parachute)	0
F9 v1.0 B0004	NASA (COTS) NRO	CCAFS LC-40	Failure (parachute)	0
F9 v1.0 B0005	NASA (COTS)	CCAFS LC-40	No attempt	525
F9 v1.0 B0006	NASA (CRS)	CCAFS LC-40	No attempt	500
F9 v1.0 B0007	NASA (CRS)	CCAFS LC-40	No attempt	677

Total payload mass

- Calculate the total payload carried by boosters from NASA

```
%sql SELECT SUM(payload_mass__kg_) AS Total_payload_mass, CUSTOMER FROM SPACEXDATASET GROUP BY CUSTOMER
```

- Present your query result with a short explanation here

total_payload_mass	customer
7759	ABS Eutelsat
8963	AsiaSat
3669	Bulsatcom
3000	CONAE
3130	CONAE, PlanetIQ, SpaceX
4200	Canadian Space Agency (CSA)
5600	EchoStar
5300	Es hailSat
2150	Hisdesat exactEarth SpaceX
6092	Hispasat NovaWurks
6070	Inmarsat
6761	Intelsat

The above query displays the total payload mass carried by boosters launched by NASA (CRS)

Average payload mass by F9 v1.1

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

```
%sql SELECT AVG(payload_mass__kg_) AS AVG_PAYLOAD_MASS FROM SPACEXDATASET WHERE booster_version = 'F9 v1.1'
```

- Present your query result with a short explanation here

avg_payload_mass
2928.400000

The above query displays average payload mass carried by booster version F9 v1.1, with an output of 2928,4kg

First successful ground landing date

- Find the date when the first successful landing outcome in ground pad

```
%sql SELECT MIN(Date) AS FIRST_SUCCESS_LANDING FROM SPACEXDATASET WHERE LANDING__OUTCOME LIKE '%Success (ground pa
d)%';
```

- Present your query result with a short explanation here

first_success_landing
2015-12-22

The above query the date when the first successful landing outcome in ground pad was achieved. (12-22-2015)

Successful drone ship landing with payload between 4000 and 6000

- List the names of boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%%sql
SELECT BOOSTER_VERSION, CUSTOMER, payload_mass__kg_, LANDING__OUTCOME
    FROM SPACEXDATASET
    WHERE LANDING__OUTCOME LIKE '%Success (ground pad)%' and (payload_mass__kg_ BETWEEN 4000 AND 6000);
```

- Present your query result with a short explanation here

booster_version	customer	payload_mass__kg_	landing__outcome
F9 FT B1032.1	NRO	5300	Success (ground pad)
F9 B4 B1040.1	U.S. Air Force	4990	Success (ground pad)
F9 B4 B1043.1	Northrop Grumman	5000	Success (ground pad)

The output of the above query shows that there are 3 booster versions within this category.

Total number of successful and failure mission outcomes

- Calculate the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(MISSION_OUTCOME) AS SUCCESSFUL_OUTCOME FROM SPACEXDATASET WHERE MISSION_OUTCOME LIKE '%Success%';
```

```
%sql SELECT COUNT(MISSION_OUTCOME) AS FAILURE_OUTCOME FROM SPACEXDATASET WHERE MISSION_OUTCOME LIKE '%Failure%';
```

- Present your query result with a short explanation here

successful_outcome	failure_outcome
100	1

The above query checks for the number of successful and failed missions, which returns with an output for 100 success missions and 1 failed mission from the spaceX dataset.

Boosters carried maximum payload

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

```
%%sql
SELECT BOOSTER_VERSION, (SELECT MAX(payload_mass__kg_) FROM SPACEXDATASET) AS MAX_PAYLOAD
    FROM SPACEXDATASET
```

- Present your query result with a short explanation here

booster_version	max_payload
F9 v1.0 B0003	15600
F9 v1.0 B0004	15600
F9 v1.0 B0005	15600
F9 v1.0 B0006	15600
F9 v1.0 B0007	15600
F9 v1.1 B1003	15600
F9 v1.1	15600
F9 v1.1	15600
F9 v1.1	15600
F9 v1.1	15600
F9 v1.1	15600

The above query returns the names of the booster versions which have carried the maximum payload mass. With the maximum payload mass of 15,500kg, the list is quite long and exhaustive too.

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

```
%%sql
SELECT *
    FROM SPACEXDATASET
    WHERE LANDING__OUTCOME LIKE '%Failure%'
    AND YEAR(DATE) = '2015'
```

- Present your query result with a short explanation here

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2015-10-01	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

The search return only two rows as the output for this category.

Rank success count between 2010-06-04 and 2017-03-20

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

```
%%sql
SELECT *
    FROM SPACEXDATASET
    WHERE LANDING__OUTCOME LIKE '%Succes%' AND
    YEAR (DATE) < 2017;
```

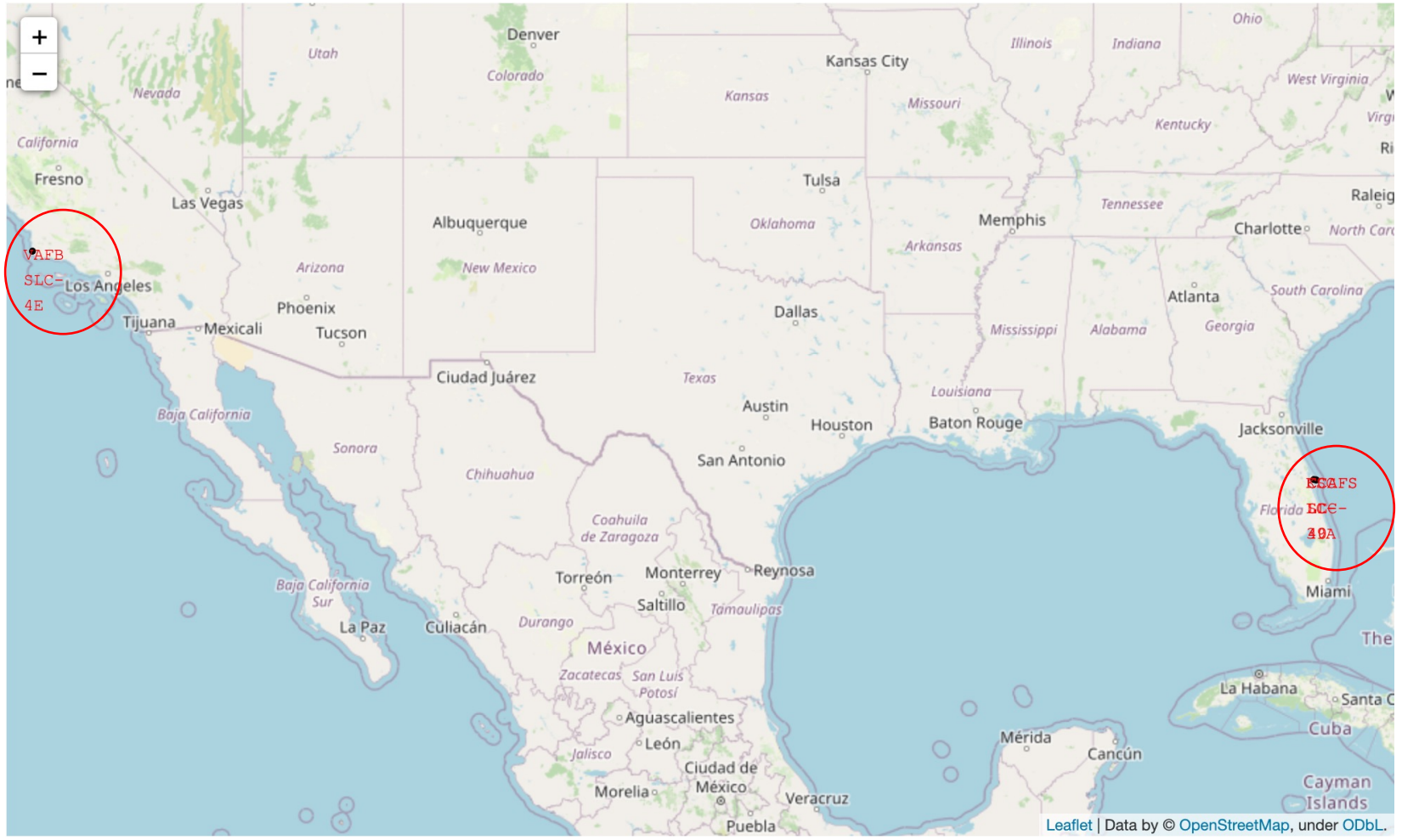
The query returns the count of successful landing outcomes between 2010 and 2017

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)
2016-08-04	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2016-06-05	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)

Interactive map with Folium

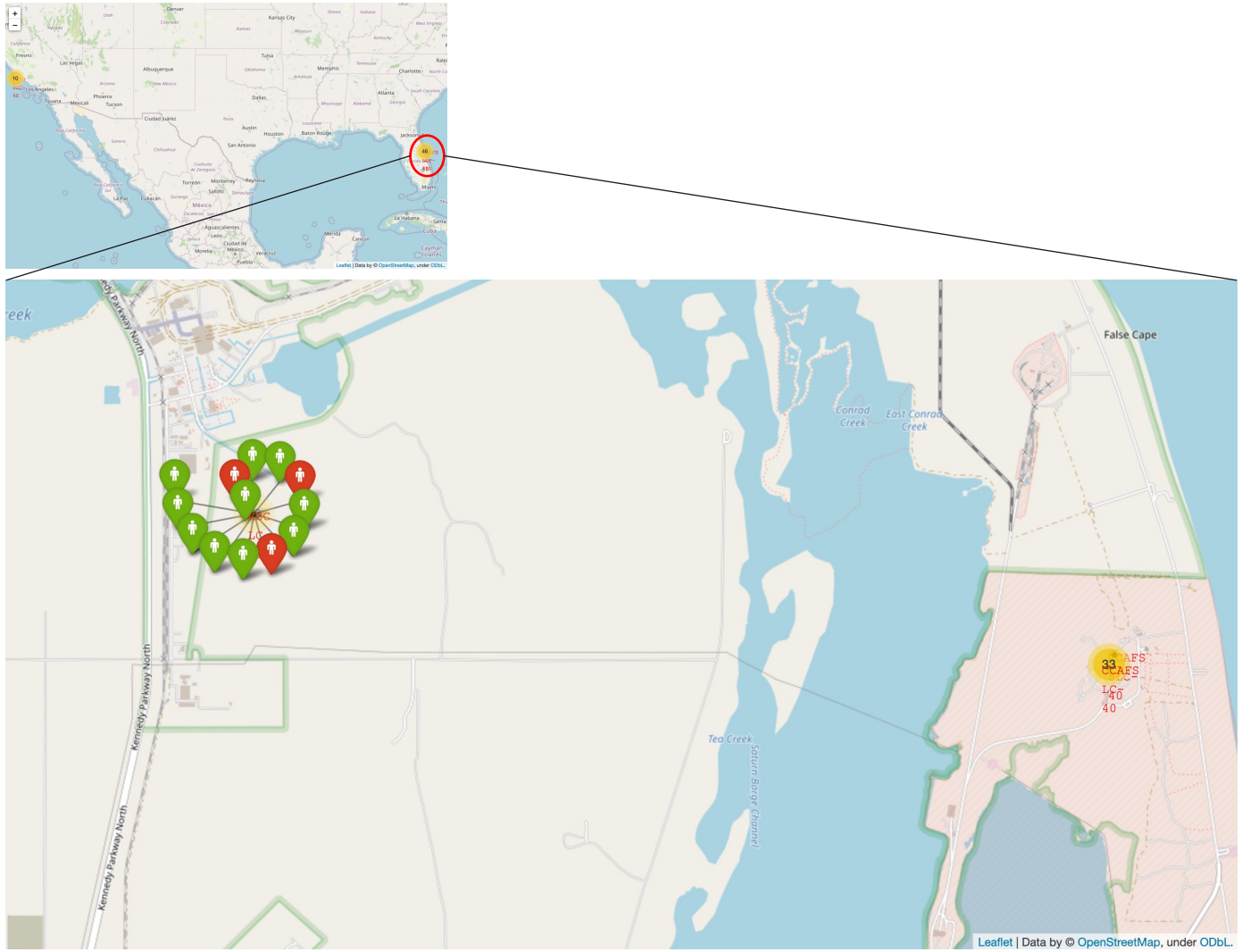
< launch sites' location markers on a global map >

The launch site locations are located on the East and West coast of the US.



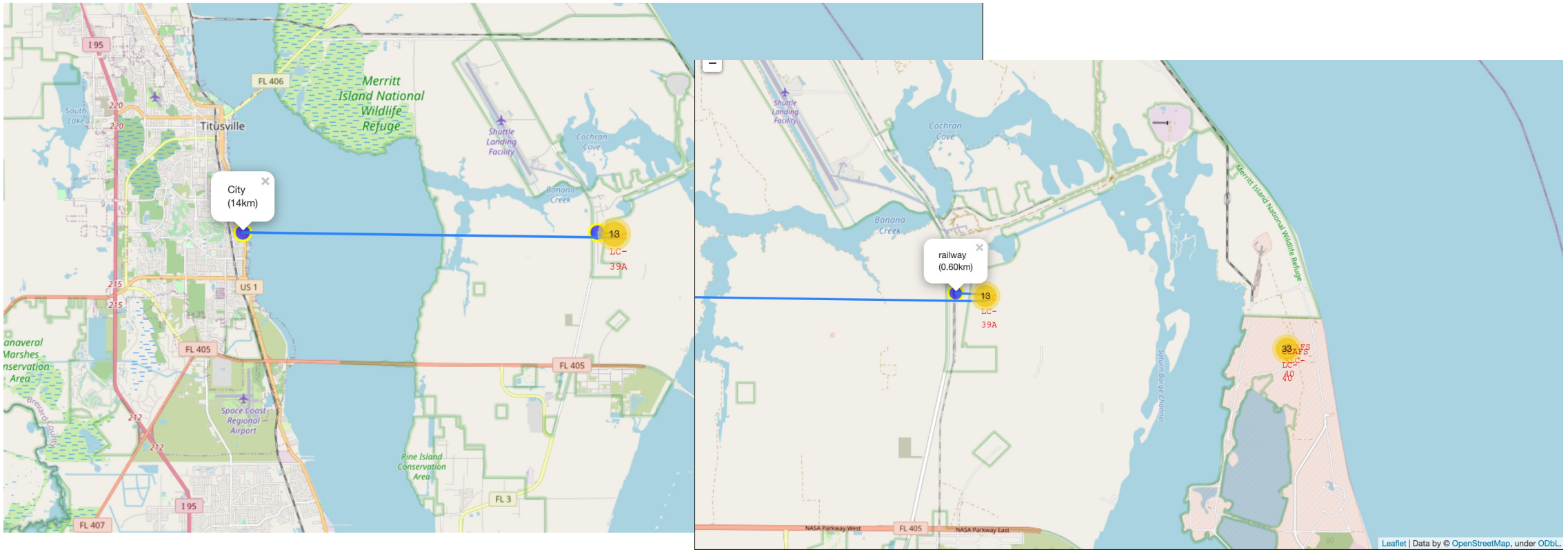
< color-labeled launch records on the map >

- The map shows the location of some of the launch sites on the Eastern coast of the US.
- The green points indicates successful launch and red for failed launch



< A selected launch site to its proximities railway, coastline with distance >

- The launch sites is in close proximty to a railway, about 0.6km away, but are somewhat far from the nearest city at about 14km away.



Build a Dashboard with Plotly Dash

<Dashboard screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard screenshot 3>

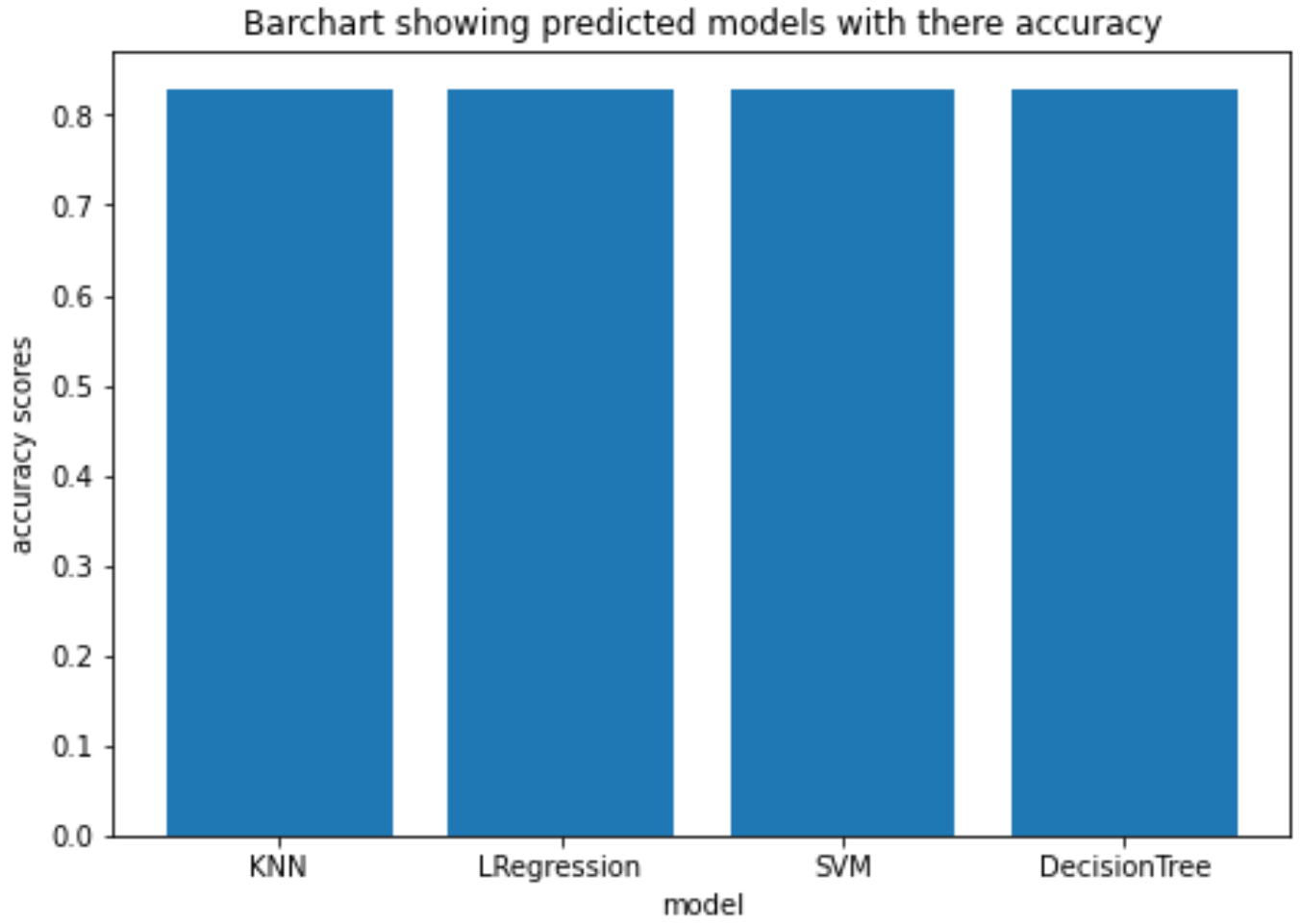
- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot

Predictive analysis (Classification)

Classification Accuracy

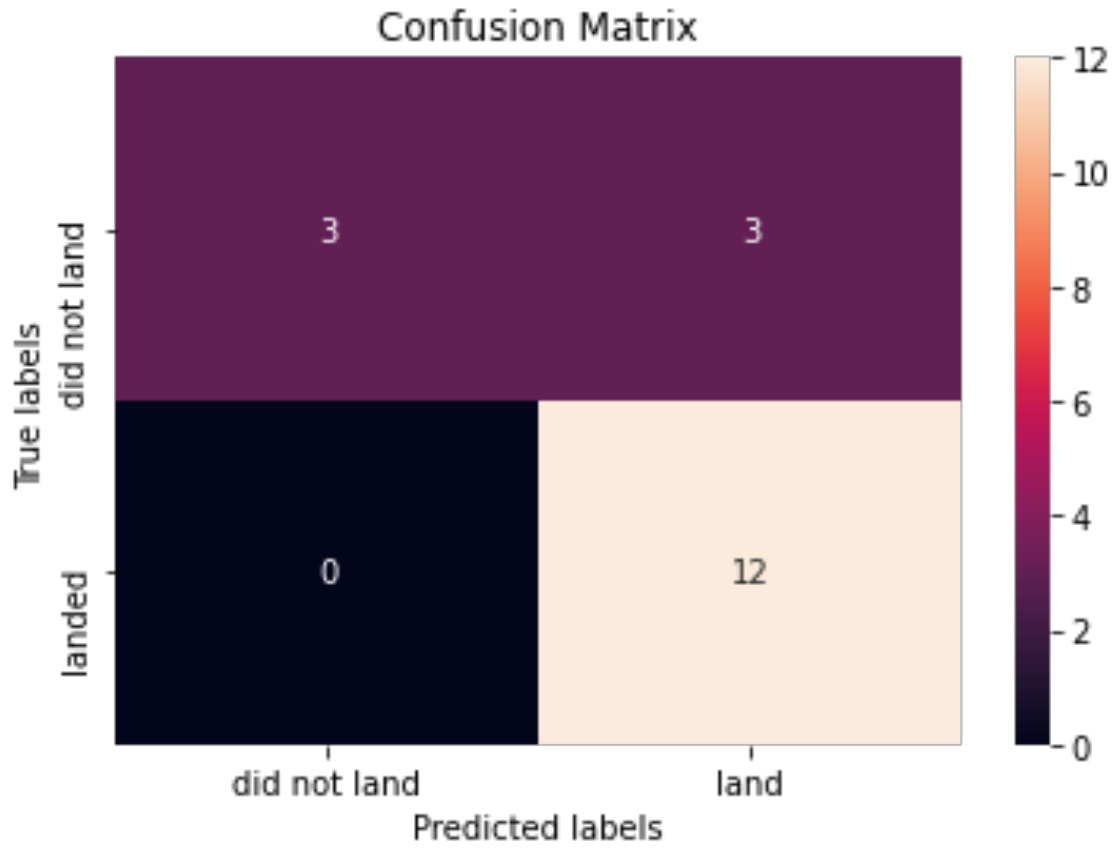
From the chart, we can observe that:

All models: ('Logistic Regression,', 'Decision Tree', 'KNN', '&', 'SVM') provided similar results with 0.83 as the final values



Confusion Matrix

Here’s the confusion matrix of the best performing model



CONCLUSION

- **Point1**: Most of the Falcon9 rocket launching took place at the CCAFS LC-40, with over 60% success rate.
- **Point 2**: Falcon9 rocket launches with relatively heavy payload masses are observed to have high success rate
- **Point 3**: Orbits with high success rate includes: ESL, GEO, HEO, SSO, & VLEO
- **Point 4**: All models: ('Logistic Regression,', 'Decision Tree', 'KNN', '&', 'SVM') provided similar results with 0.83 as the final values

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