IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Taufiq Nurrohman
01-Feb-2025

Outline

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

Executive Summary

Summary of methodologies

- Data Collection
- Data Wrangling
- EDA with Data Visualization
- EDA with SQL
- Building an interactive map with folium
- Building a dashboard with Plotlydash
- Predictive Analysis (Classification)

Summary of all results

- EDA results
- Interactive Analysis
- Predictive Analysis

Introduction

- Project background and context
  - Analyzing factors that influence rocket missions and result in successful outcomes
- Problems you want to find answers
  - Determining what factors predict successful mission outcomes

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Using REST Apis
    - Using Beautiful Soup package to web scrape HTML tables
- Perform data wrangling
    - Cleaning, transforming & structuring raw data into a useable format by handling missing values, removing duplicates and correcting inconsistencies.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and PlotlyDash
- Perform predictive analysis using classification models
    - Build a machine learning pipeline, determine model with best accuracy, output the confusion matrix

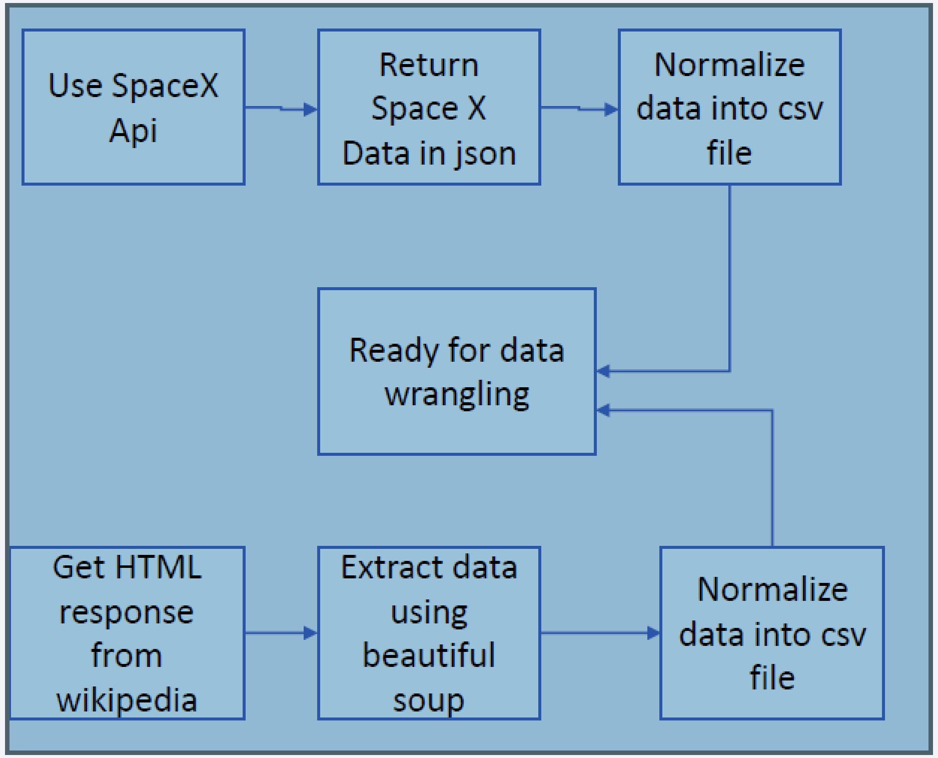
Data Collection

Data Collection steps -

•Gather data from Space X REST Api

•API will give data about launches, rockets used, payload delivered, launch specification, landing specification and landing outcome.

•Another popular data source is web scrapping using Beautiful Soup



Data Collection – SpaceX API

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

- Add the GitHub URL

https://github.com/smoha-git/Firstrepo/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

1. Getting Response from API

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

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

2.

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

3.

```
# Call getLaunchSite
getLaunchSite(data)
```

```
# Call getPayloadData
getPayloadData(data)
```

```
# Call getCoreData
getCoreData(data)
```

4.

```
# Create a data from launch_dict
data=pd.DataFrame(launch_dict)
```

Show the summary of the dataframe

```
# Show the head of the dataframe
data.info()
```

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

https://github.com/smoha-git/Firstrepo/blob/main/Webscraping_Subhasree.ipynb

```
import plotly.graph_objects as go
from plotly.subplots import make_subplots
```

```
import pandas as pd
```

```
import requests
from bs4 import BeautifulSoup
```

```
import plotly.graph_objects as go
from plotly.subplots import make_subplots
```

```
tesla_data = tesla.history(period="max")
```

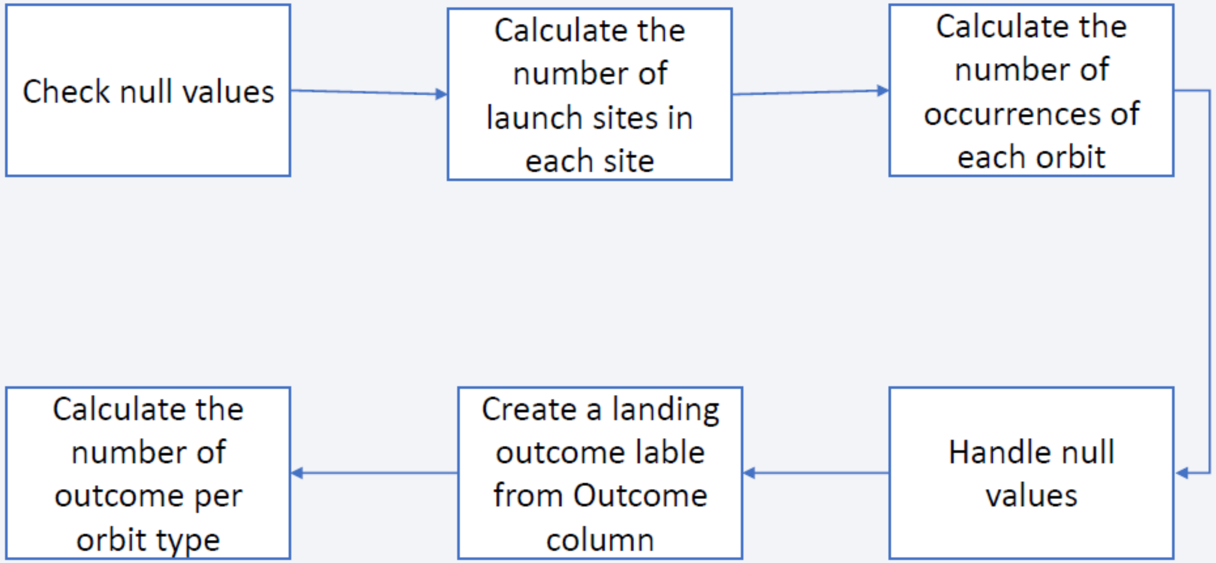
```
tesla_data.head(5)
```

	Open	High	Low	Close	Volume	Dividends	Stock Split
Date							
2010-06-29 00:00:00-04:00	1.266667	1.666667	1.169333	1.592667	281494500	0.0	0.

Data Wrangling

- Describe how data were processed

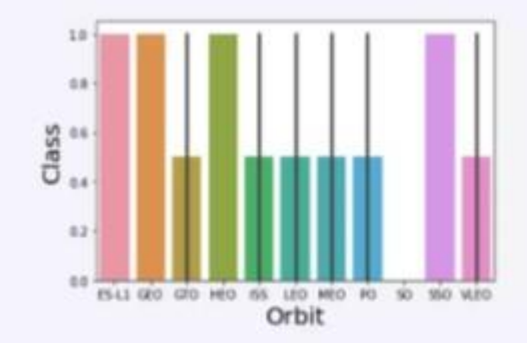
- Add the GitHub URL

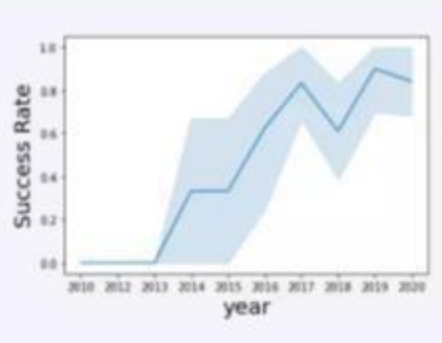


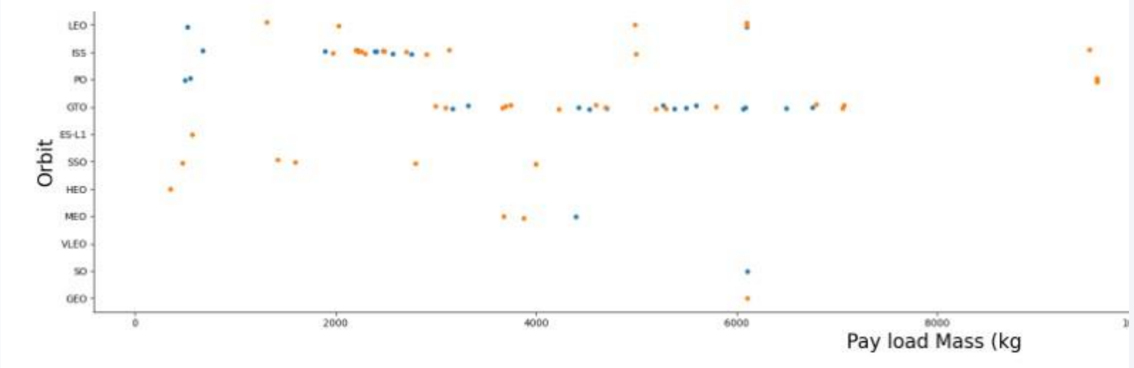
https://github.com/smoha-git/Firstrepo/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

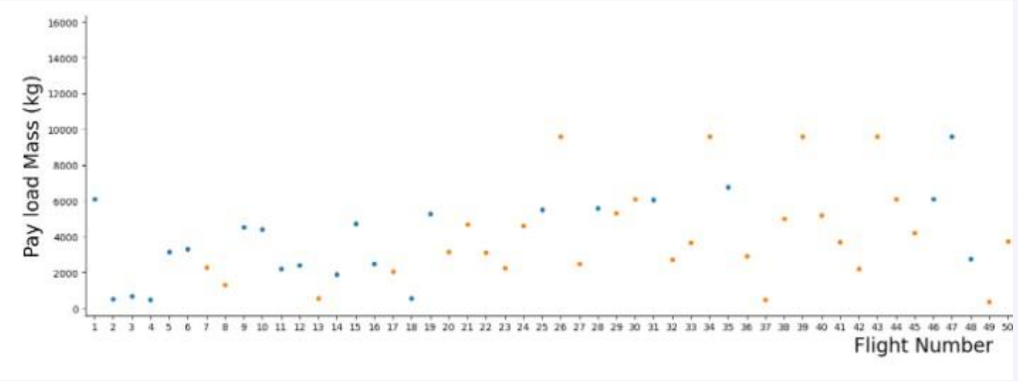
EDA with Data Visualization

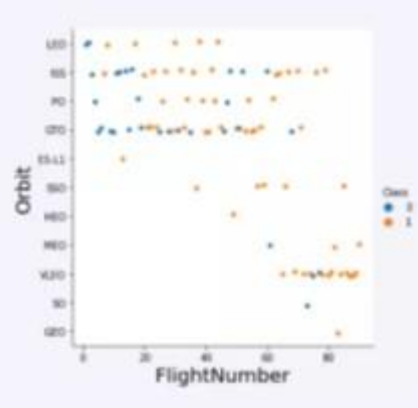
https://github.com/smoha-git/Firstrepo/blob/main/edadataviz.ipynb

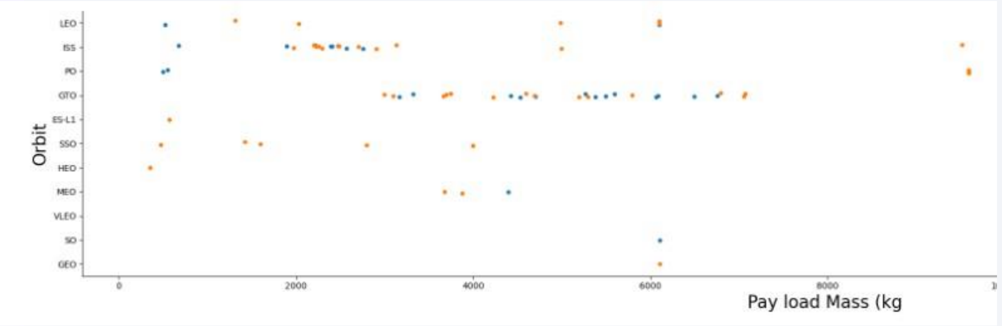












EDA with SQL

SQL queries performed:

- Create table from another table
- Select Distinct
- Applied Where, Limit, and Between conditional clauses
- Aggregation functions like Sum, Count and Average

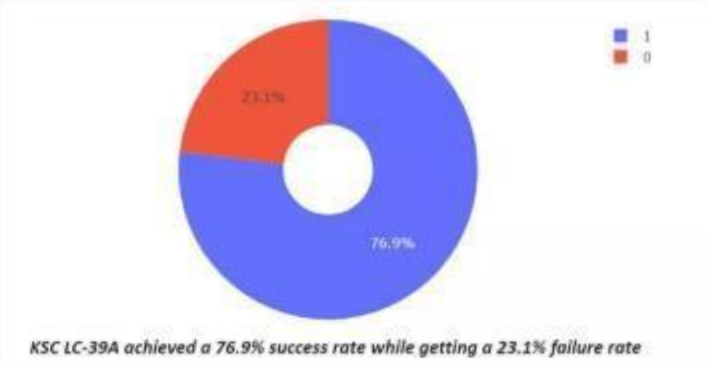
https://github.com/smoha-git/Firstrepo/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

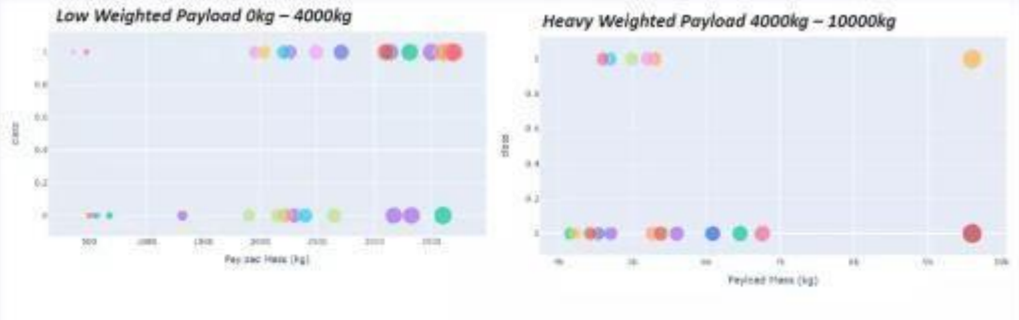
Map markers have been added to the map with aim to finding an optimal location for building a launch site

Build a Dashboard with Plotly Dash

We can see that KSC LC-39A had the most successful launches from all the sites



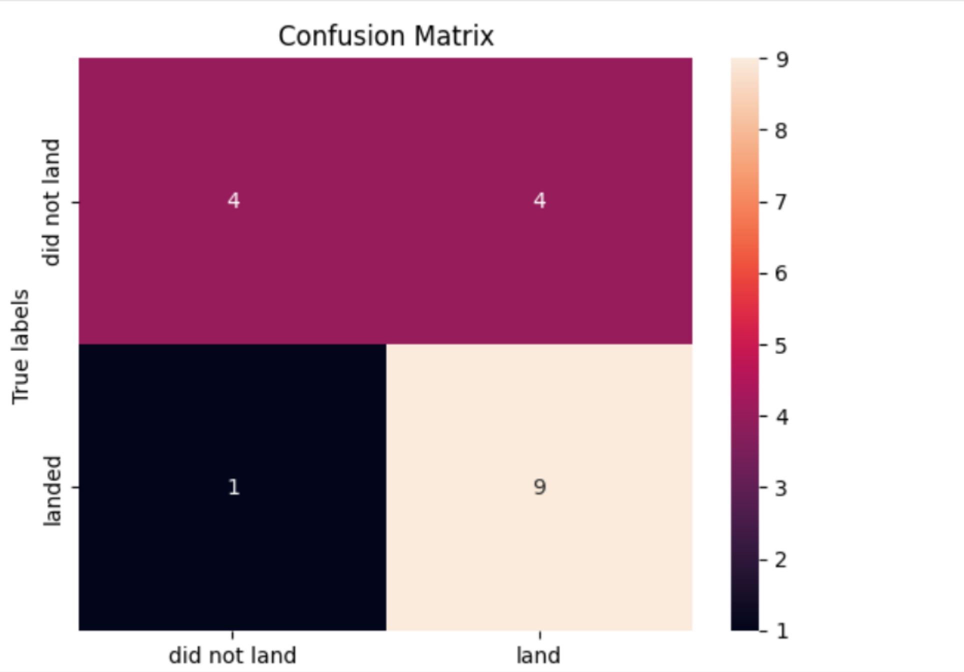
KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate



We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Predictive Analysis (Classification)

- Built, evaluated, improved, and found the best performing classification model
- Applied a standardScaler() function to preprocess data
- Split the data into training and testing samples
- Applied GridSearch to find the best hyperparameters for different classification
- Evaluated the various models' accuracy and analysed the confusion matrix



https://github.com/smoha-git/Firstrepo/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

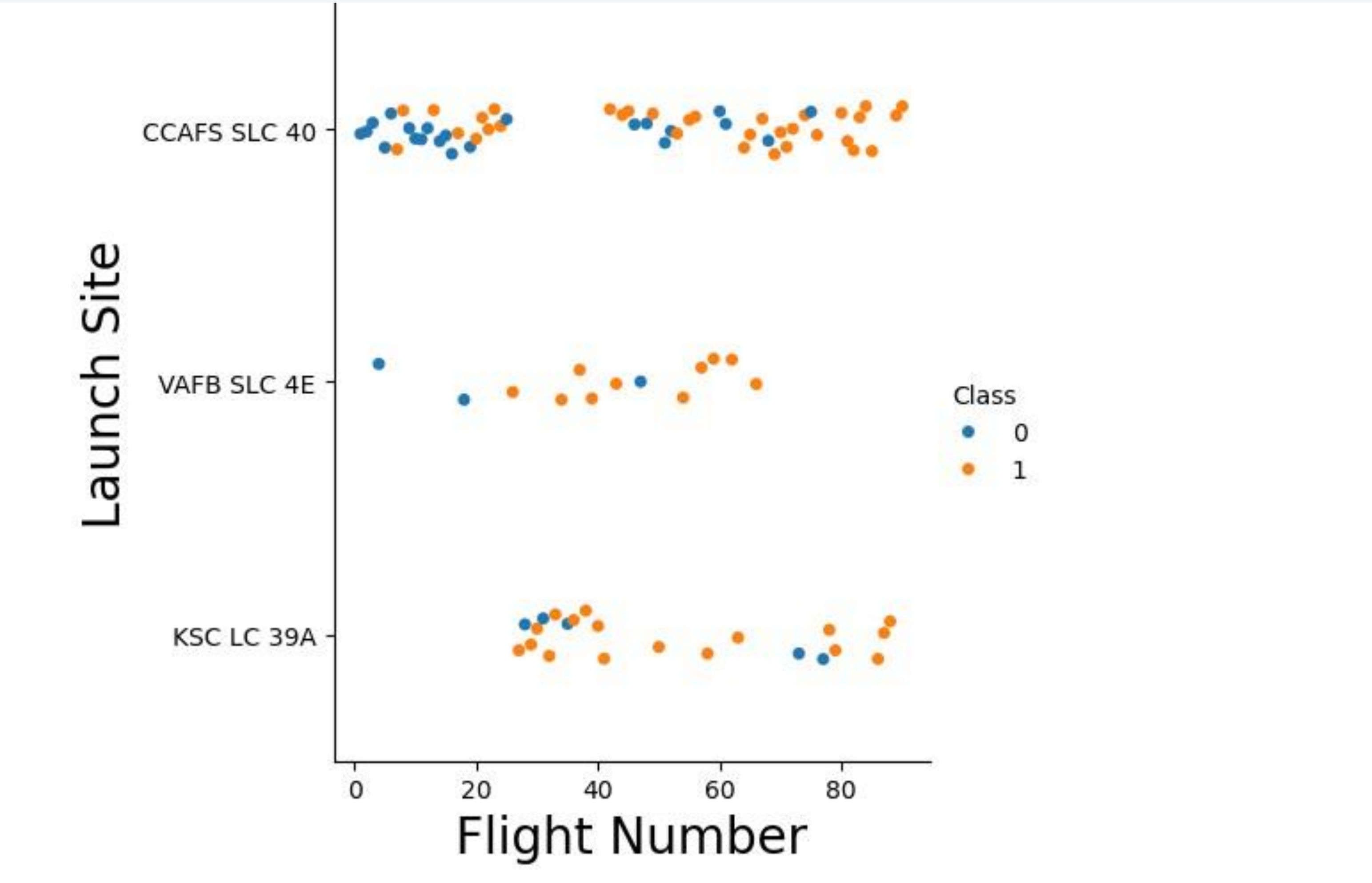
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

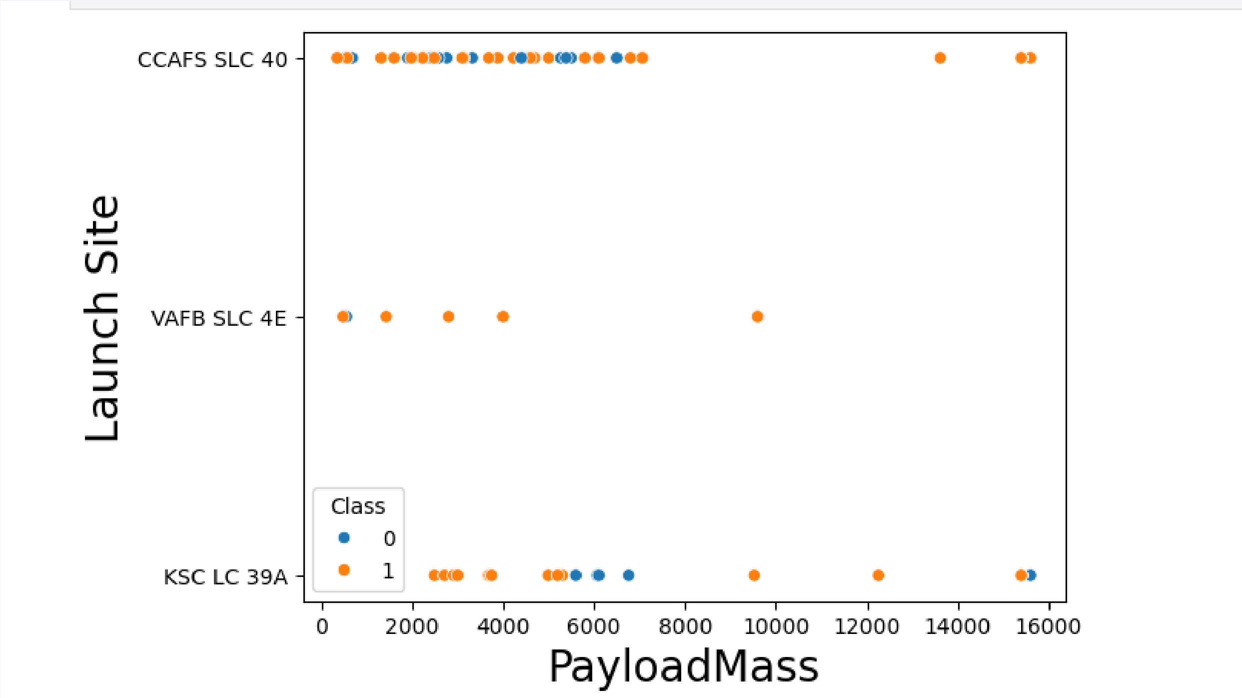
Flight Number vs. Launch Site

Launches from CCAFS SLC 40 are higher in number and higher chances of success



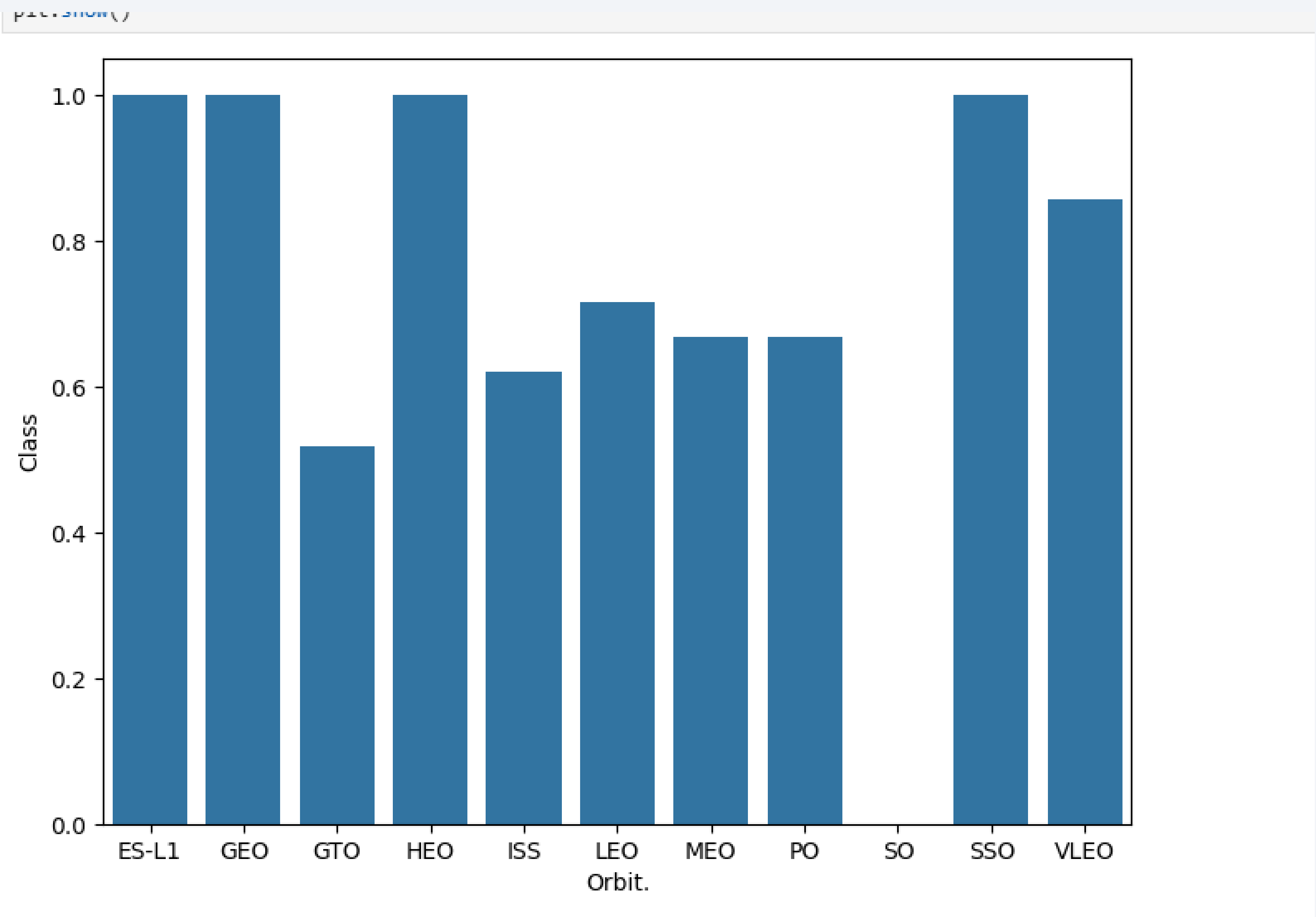
Payload vs. Launch Site

CCAFS SLC 40 has launched majority with payload mass



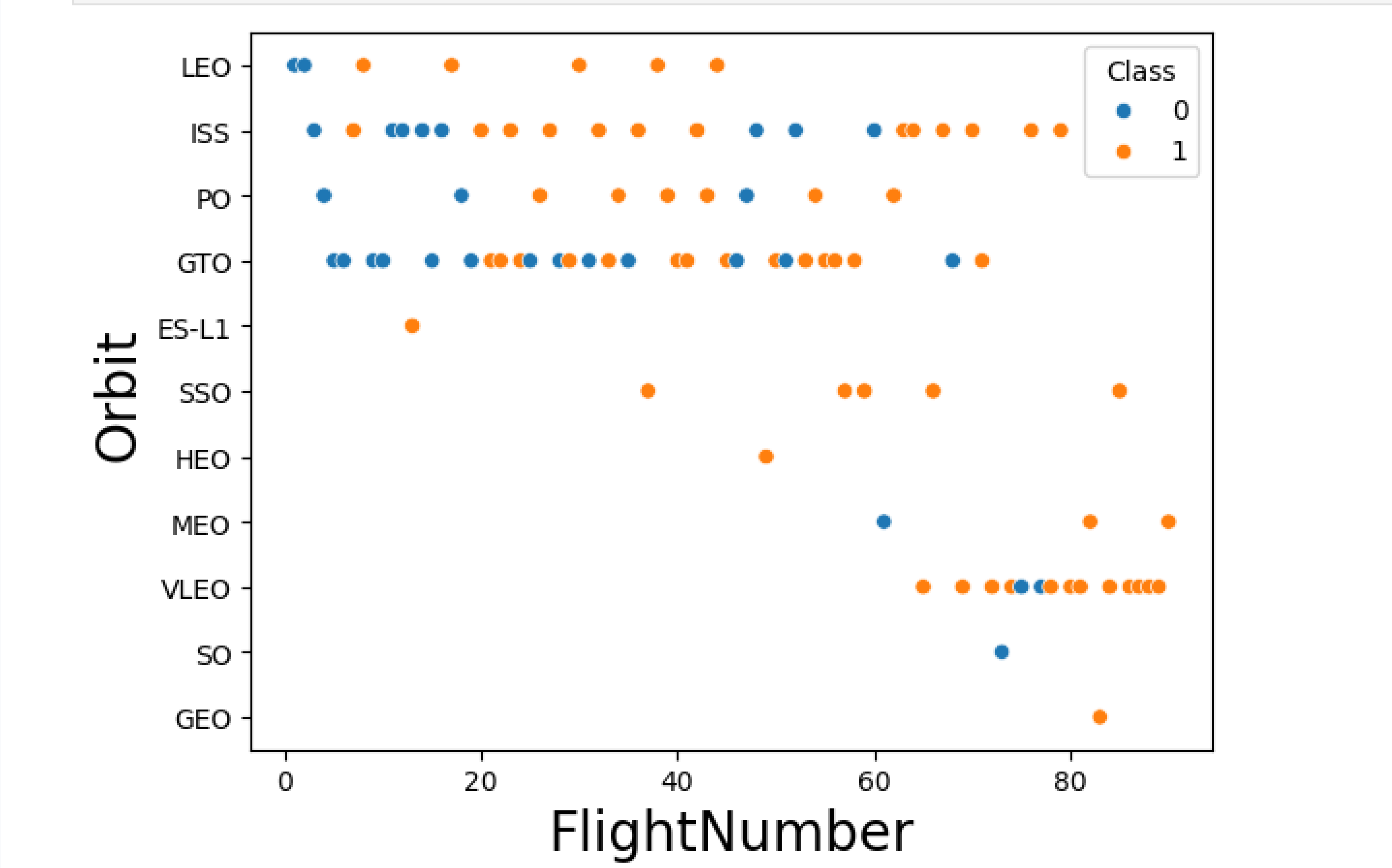
Success Rate vs. Orbit Type

ES-L1, GEO, HEO & SSO are among the highest success rate



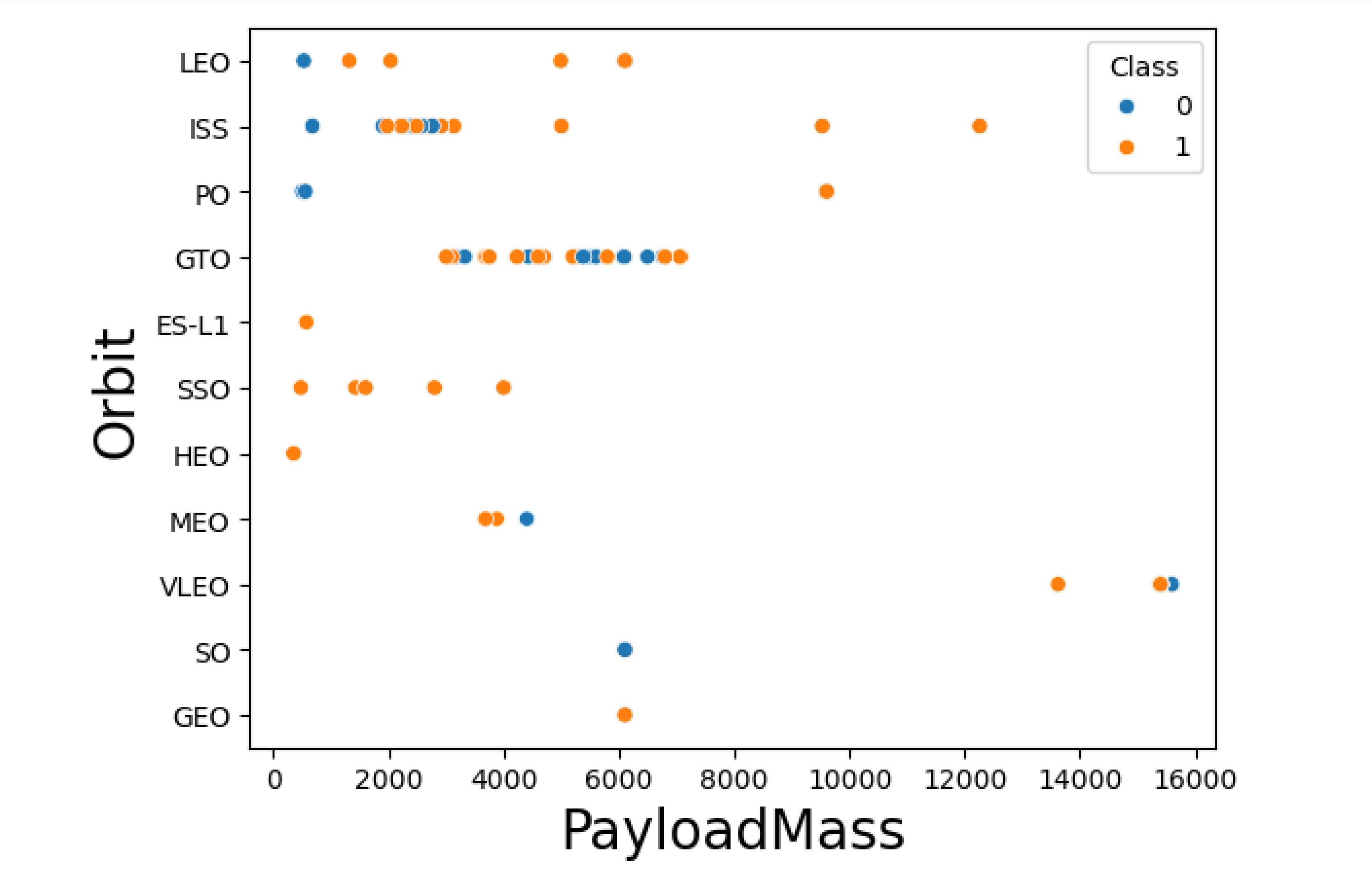
Flight Number vs. Orbit Type

Towards the end, all orbit types showed successful outcomes as the flight numbers increased, VLEO has a different trend



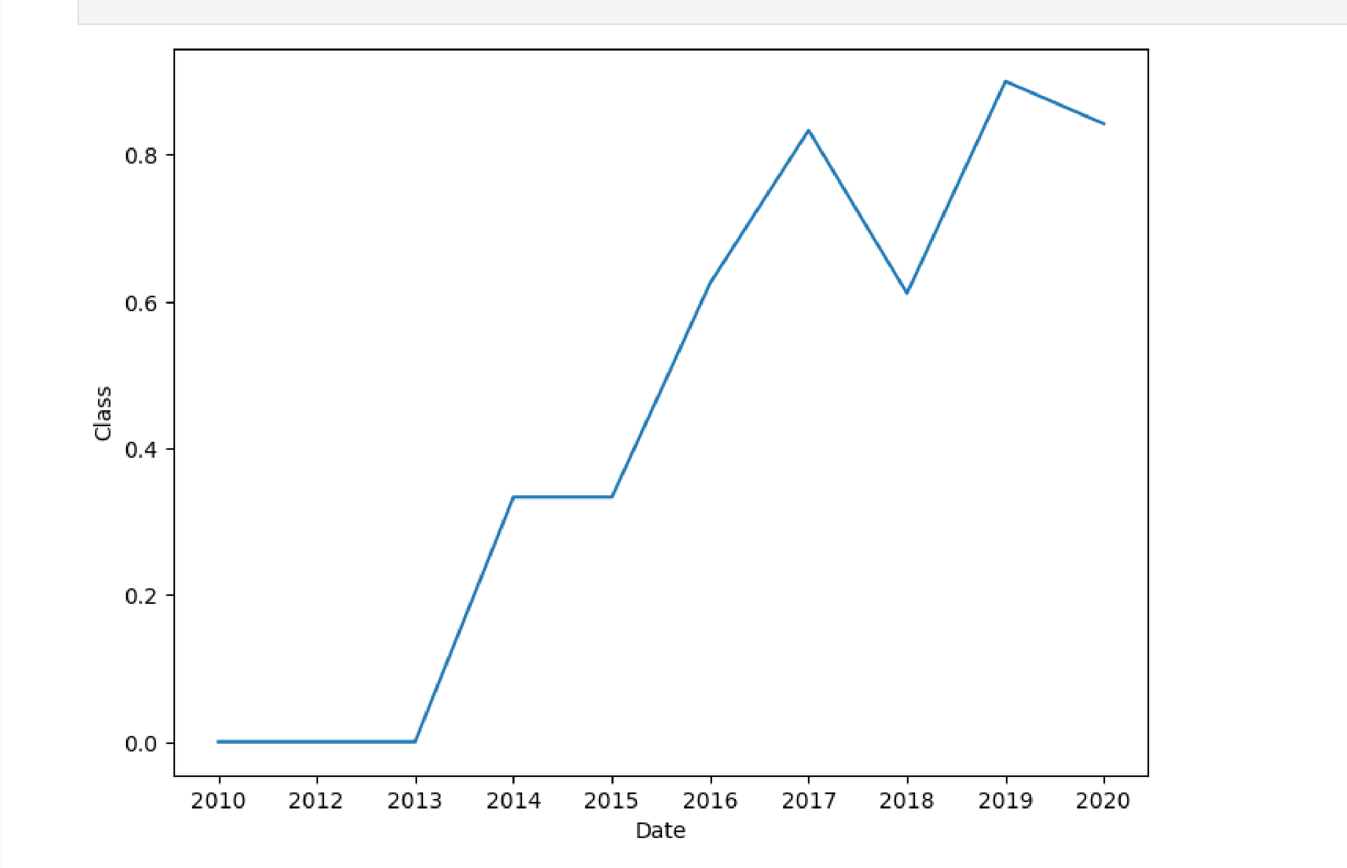
Payload vs. Orbit Type

Higher payloads showed better results across most orbit types for mission success



Launch Success Yearly Trend

The success rate since 2013 kept increasing till 2020



All Launch Site Names

%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE

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
```

Date	Time (UTC)	Booster_Version	Launch_Site
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_)  FROM  SPACEXTABLE  WHERE  Customer="NASA  (CRS)"
```

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

%sql SELECT Date FROM SPACEXTABLE WHERE Landing_Outcome="Success (ground pad)" LIMIT 1

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select BOOSTER_VERSION from SPACEXTBL where Landing__Outcome
= 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and
PAYLOAD_MASS__KG_ < 6000
```

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

Total Number of Successful and Failure Mission Outcomes

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

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

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE
PAYLOAD_MASS__KG_=(SELECT MAX(PAYLOAD_MASS__KG_)
FROM SPACEXTABLE)
```

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

%sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2015-01-01' and '2015-12-31') order by date desc

time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
		CCAFS LC-				SKY Perfect JSAT		

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

```
%sql SELECT COUNT(Landing_Outcome), Landing_Outcome
FROM SPACEXTABLE WHERE Date BETWEEN "2010-06-04" AND
"2017-03-20" GROUP BY Landing_Outcome ORDER BY
COUNT(Landing_Outcome) DESC
```

COUNT(Landing_Outcome)	Landing_Outcome
10	No attempt
5	Success (drone ship)
5	Failure (drone ship)
3	Success (ground pad)
3	Controlled (ocean)
2	Uncontrolled (ocean)
2	Failure (parachute)
1	Precluded (drone ship)

Section 3

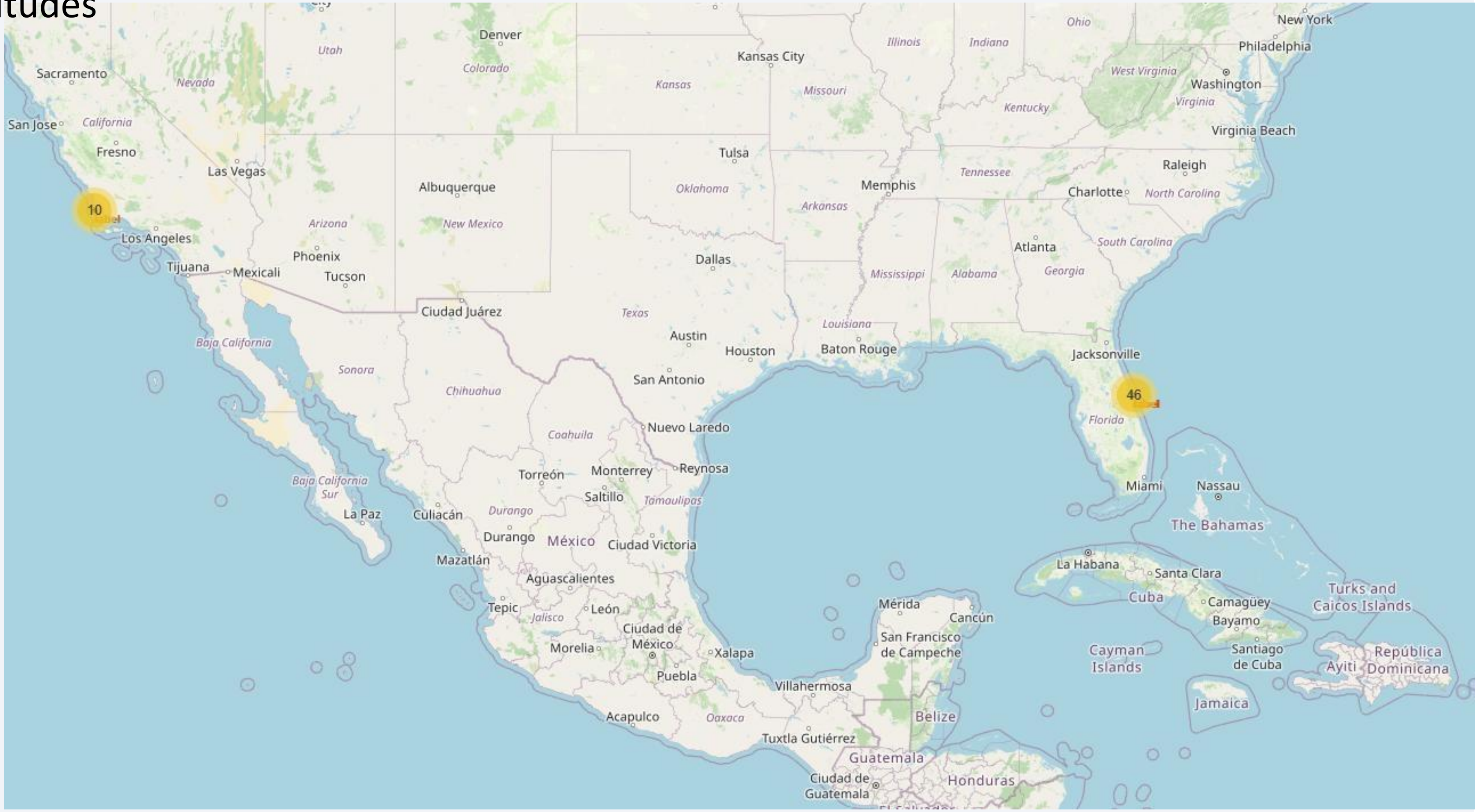
Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

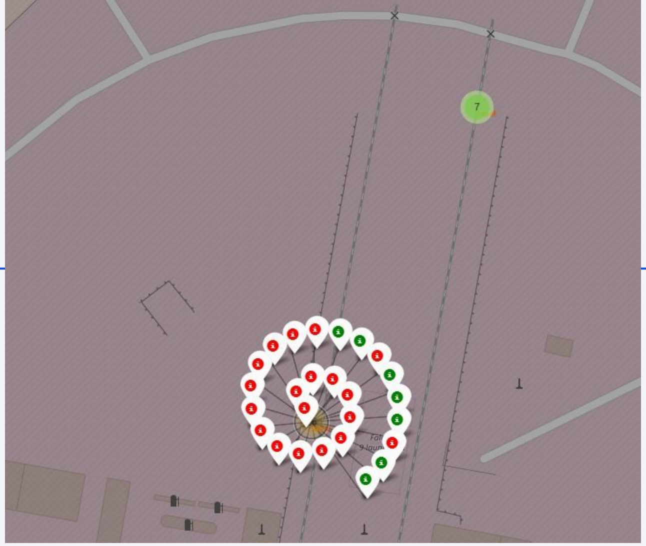
Insights

The launch sites are close to the coastal regions.

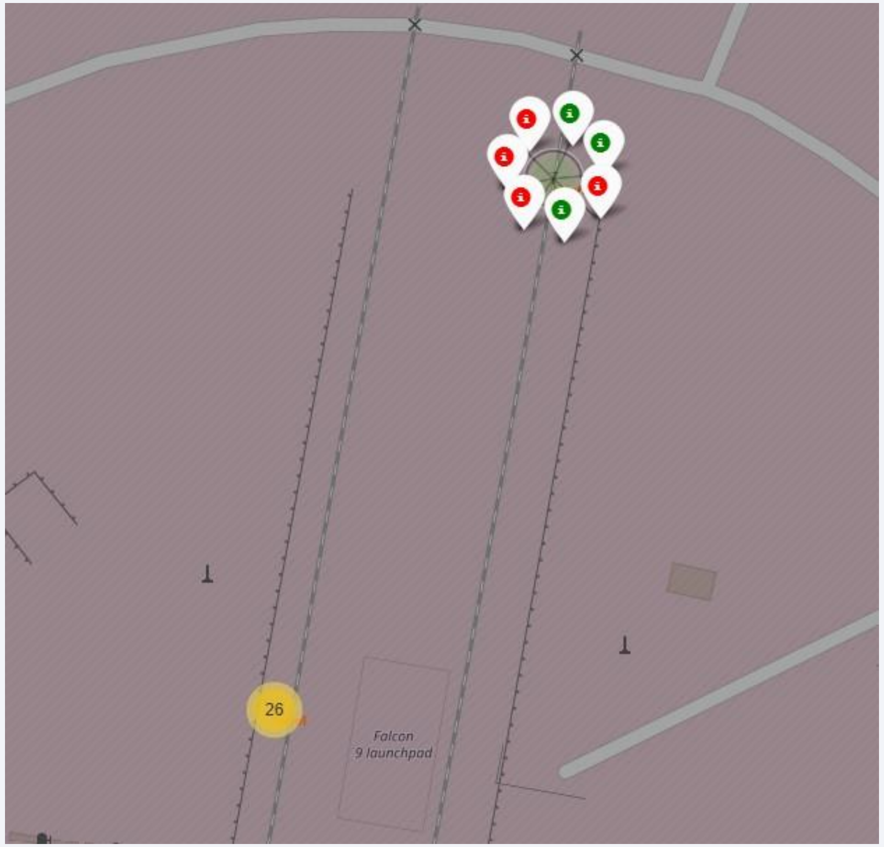
The launch sites are on similar latitudes



<Folium Map Screenshot 2>



Insights
The two launch sites on the east coast had more failed launches than successful launches.
This is possibly due to the fact that initial launches failed and over time they improved



<Folium Map Screenshot 3>

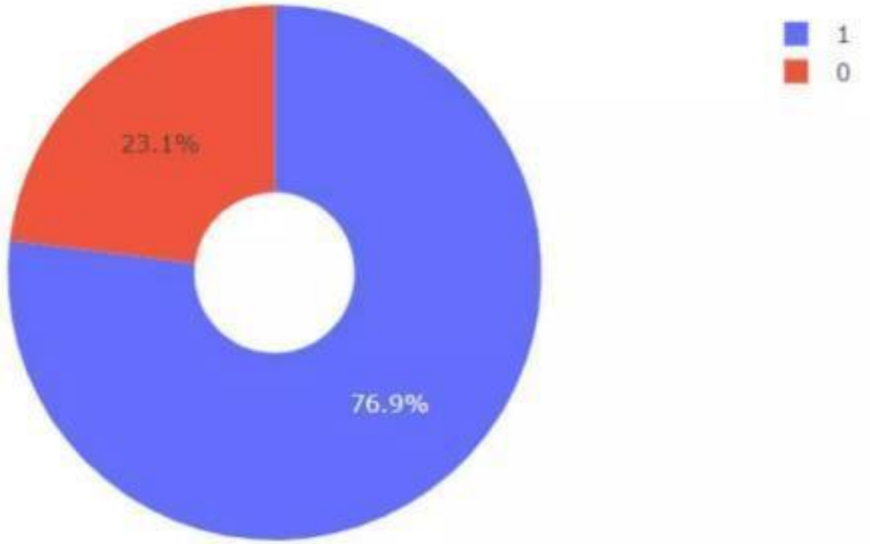
Launch site CCAFS SLC-40 is 0.87km from the coastline

Section 4

Build a Dashboard with Plotly Dash

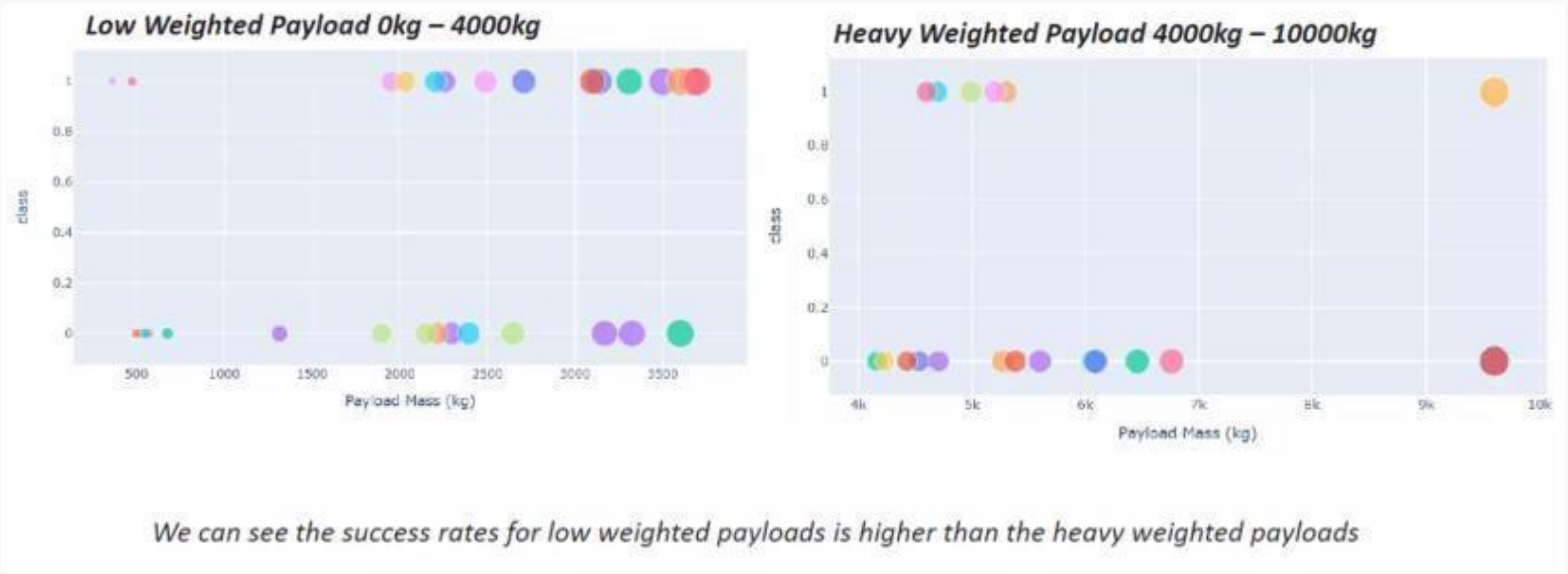
<Dashboard Screenshot 1>

<Dashboard Screenshot 2>



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

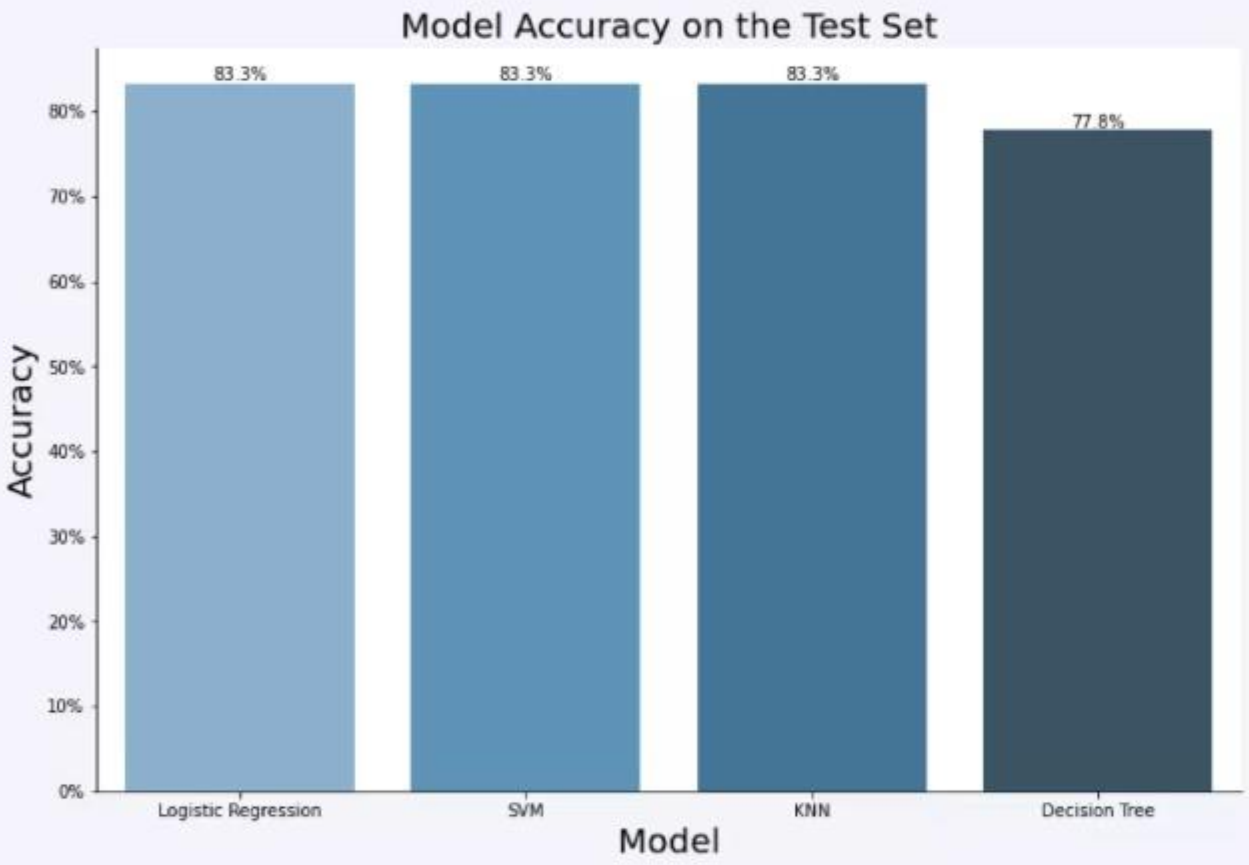
<Dashboard Screenshot 3>



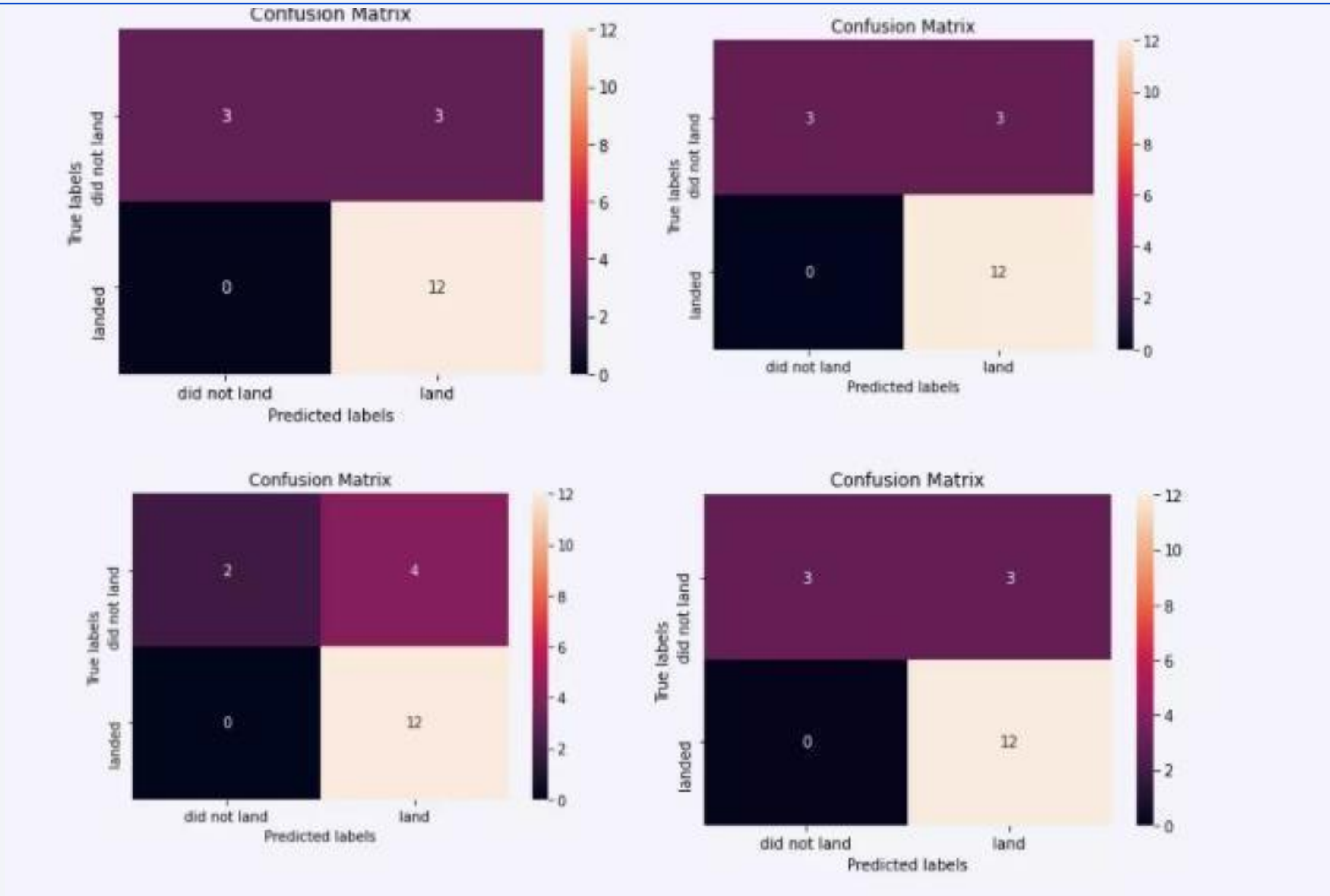
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

There is plenty of SpaceX launch data available online

- The two launch sites on the east coast had more failed launches than successful launches.
- The KSC LC -39A launch site had the highest success rate
- The decision tree was the best model for classification purposes

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!