FALCON 9 BOOSTER LANDING SUCCESS PREDICTION

Capstone

Shahadat Hussain

Researcher, Mechanical Engineering

April, 2024

https://sites.google.com/view/shahadathussain/

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

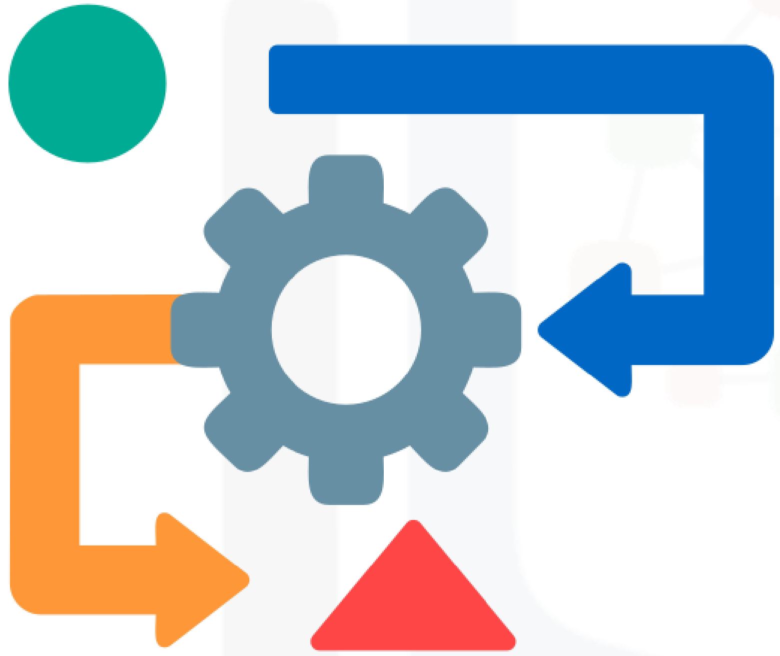
EXECUTIVE SUMMARY

- Analysis of space mission data reveals a positive trend of improving success rates over time.
- Key findings highlight:
  - Effectiveness of KSC LC-39A as a launch site.
  - Potential of CCAFS SLC-40 for handling heavier payloads.
  - Importance of strategic mission planning, evident in consistent success rates in specific orbits.
  - Observed increase in success rates beyond 2013 reflects advancements in technology and operational practices.
- Insights underscore the importance of continuous innovation and collaboration in achieving successful space missions.

INTRODUCTION

- The capstone project aims to predict the successful landing of the Falcon 9 first stage.
- SpaceX promotes Falcon 9 rocket launches on their website, priced at $62 million, significantly lower than competitors' costs of over $165 million, largely due to the reusable first stage.
- Predicting the first stage's landing success enables estimation of launch costs, crucial for competing bids against SpaceX.
- Using data science methods, variables like Payload Mass, Flight Number, Orbit Type, Launch Site location, and other factors were examined to gain insights into their behaviors and dependencies, particularly regarding landing success and mission success rates.

METHODOLOGY



- Data Collection
    - APIs
    - Web scraping using BeautifulSoup library
- Data Wrangling
- Exploratory Data Analysis
    - Pandas, matplotlib, numpy, and scipy
    - SQL
- Interactive Visual Analytics
    - Using Folium map library
    - Dashboard using Plotly Dash
- Prediction Models
    - SVM
    - Classification Trees
    - Regression

RESULTS

DATA COLLECTION THROUGH API

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

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

In [19]:
```
# Use json_normalize meethod to convert the json result into a dataframe
response_json = response.json()
df = pd.json_normalize(response_json)
```

Using the dataframe `data` print the first 5 rows

In [20]:
```
# Get the head of the dataframe
df.head()
```

Out[20]:

	static_fire_date_utc	static_fire_date_unix	net	window	rocket	success	failures	details	crew	ships	capsules	
		1.142554e+09	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}]	Engine failure at 33 seconds and loss of vehicle	[]	[]	[]	[5eb0e4t
1	None	NaN	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 301, 'altitude': 289, 'reason': 'harmonic oscillation	Successful first stage burn and transition to second stage, maximum altitude 289 km, Premature engine	[]	[]	[]	[5eb0e4t

DATA COLLECTION WITH WEB SCRAPING

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

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

```python
# Use BeautifulSoup() to create a Beautiful
import html5lib
soup=BeautifulSoup(response_text, 'lxml')
```

Print the page title to verify if the `BeautifulSoup` object

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

```
[6]: <title>List of Falcon 9 and Falcon Heavy la
```

```python
# Use the find_all function in the BeautifulSoup object, with element type `table`
# 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.

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

```
<table class="wikitable plainrowheaders collapsible" style="width: 100%;">
<tbody><tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date and<br/>time (<a href="/wiki/Coordinated_Universal_Time" title="Coordinated Universal Time">UTC</a>)
</th>
<th scope="col"><a href="/wiki/List_of_Falcon_9_first-stage_boosters" title="List of Falcon 9 first-stage boosters">Versio
n,<br/>Booster</a> <sup class="reference" id="cite_ref-booster_11-0"><a href="#cite_note-booster-11">[b]</a></sup>
</th>
```

EDA WITH SQL

[10]:
```
%sql Select distinct Launch_Site from SPACEXTABLE
```

* sqlite:///my_data1.db
Done.

Out[10]:

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

Display 5 records where launch sites begin with the string 'CCA'

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

* sqlite:///my_data1.db
Done.

5]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

DATA WRANGLING

```
In [3]: df.isnull().sum()/len(df)*100
```

```
Out[3]: FlightNumber      0.000000
        Date              0.000000
        BoosterVersion    0.000000
        PayloadMass       0.000000
        Orbit             0.000000
        LaunchSite        0.000000
        Outcome           0.000000
        Flights           0.000000
        GridFins          0.000000
        Reused            0.000000
        Legs              0.000000
        LandingPad       28.888889
        Block             0.000000
        ReusedCount       0.000000
        Serial            0.000000
        Longitude         0.000000
        Latitude          0.000000
        dtype: float64
```

Identify which columns are numerical and

Use the method `value_counts()` on the column `LaunchSite` to determine the number of launches on each site:

```
# Apply value_counts() on column LaunchSite
launch_counts = df['LaunchSite'].value_counts()
launch_counts
```

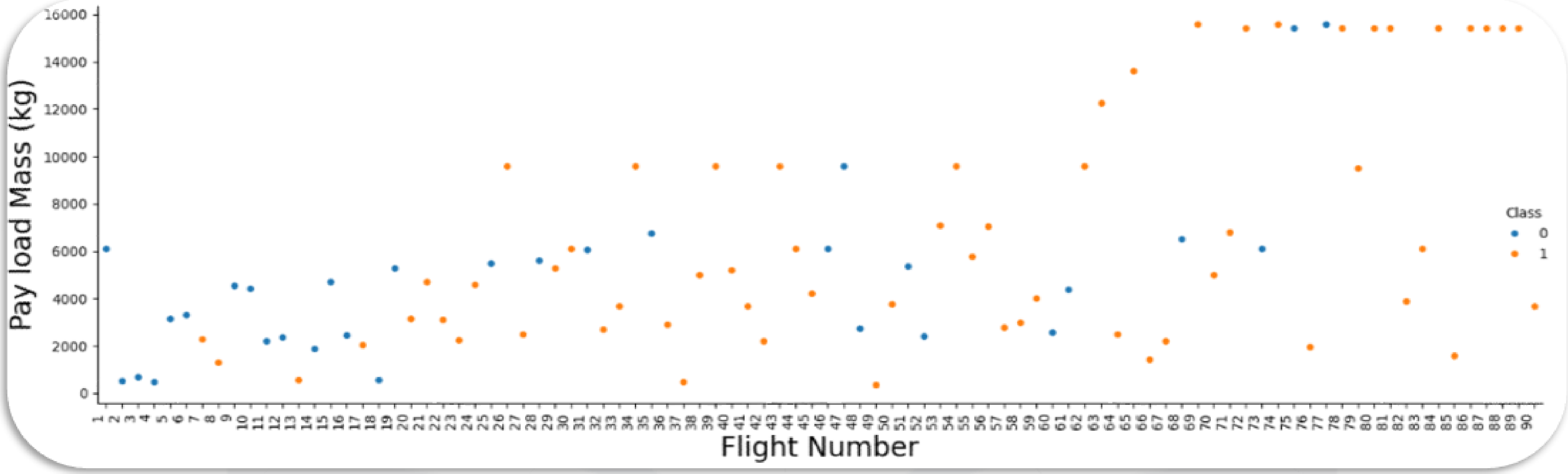
```
LaunchSite
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: count, dtype: int64
```

Using the `Outcome`, create a list where the element is zero if the corresponding row in `Outcome` is in the set `bad_outcome`; otherwise, it's one. Then assign it to the variable `landing_class`:

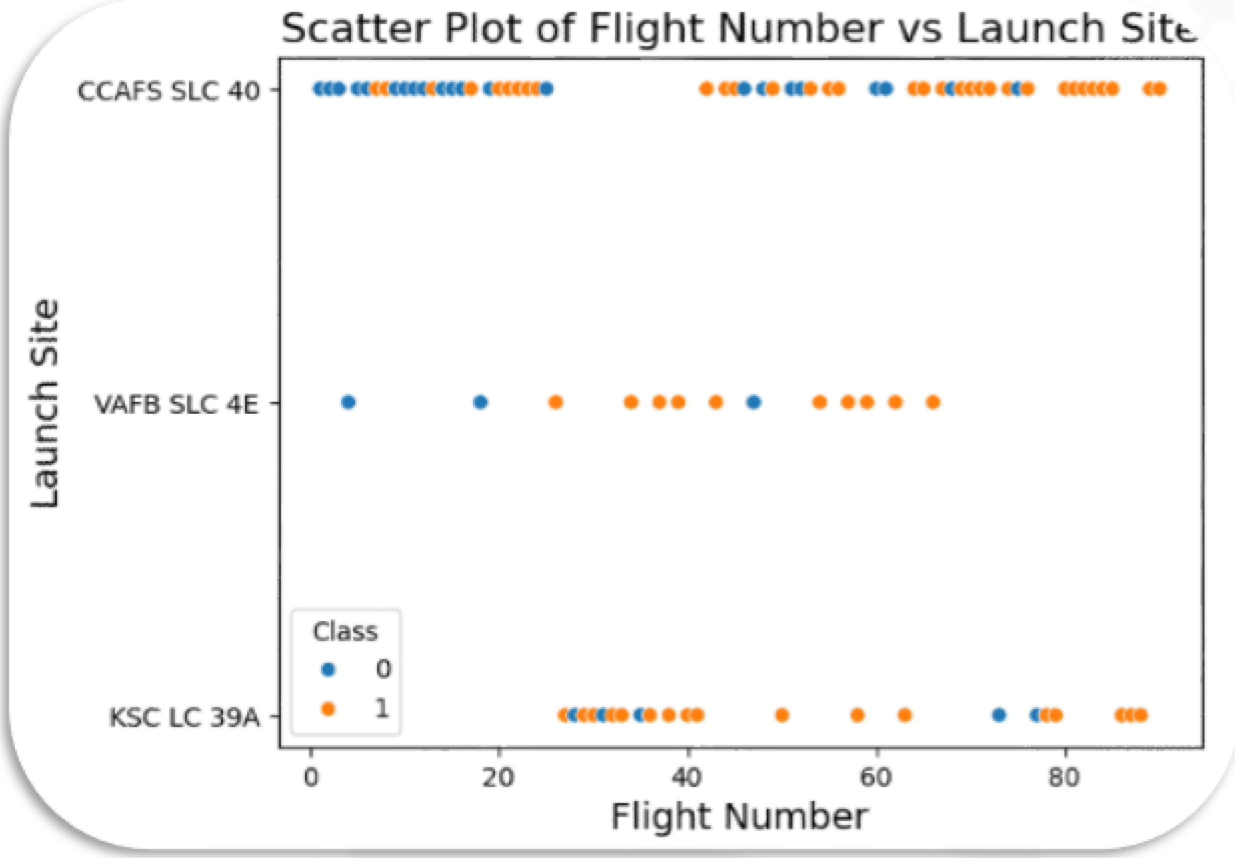
```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise


landing_class = [0 if outcome in bad_outcomes else 1 for outcome in df['Outcome']]


print("Landing class:")
print(landing_class)
```

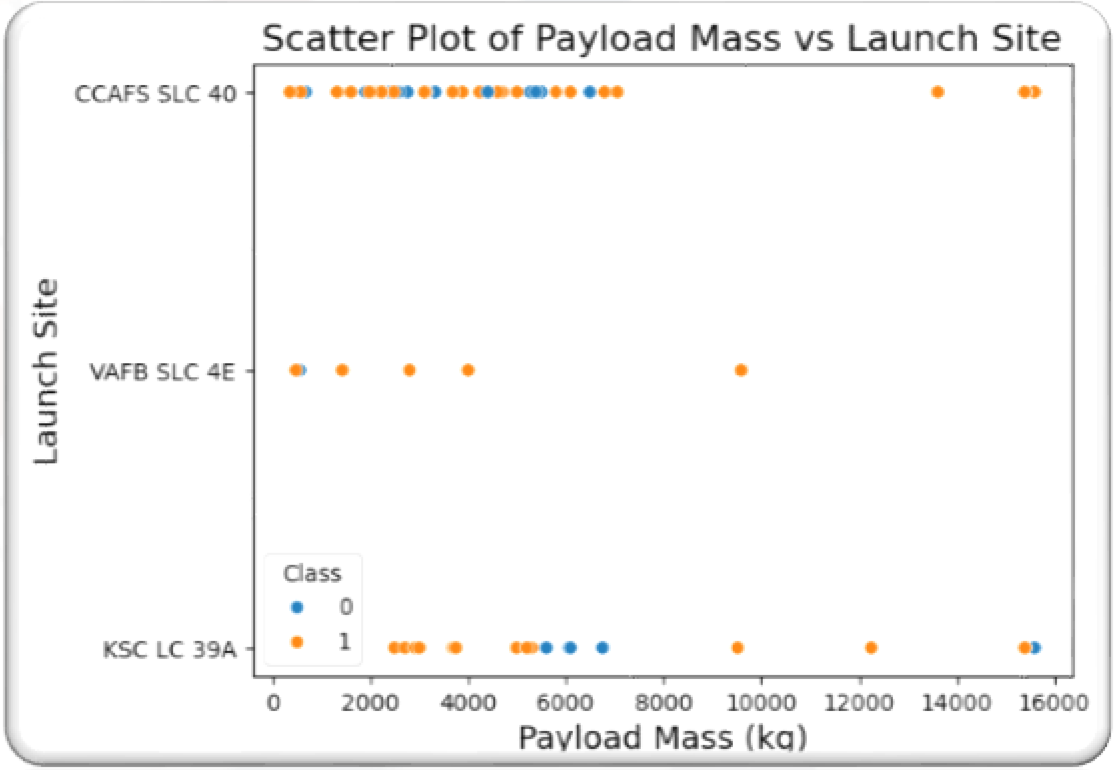
```
Landing class:
[0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1]
```

PAYLOAD MASS VS FLIGHT NUMBER

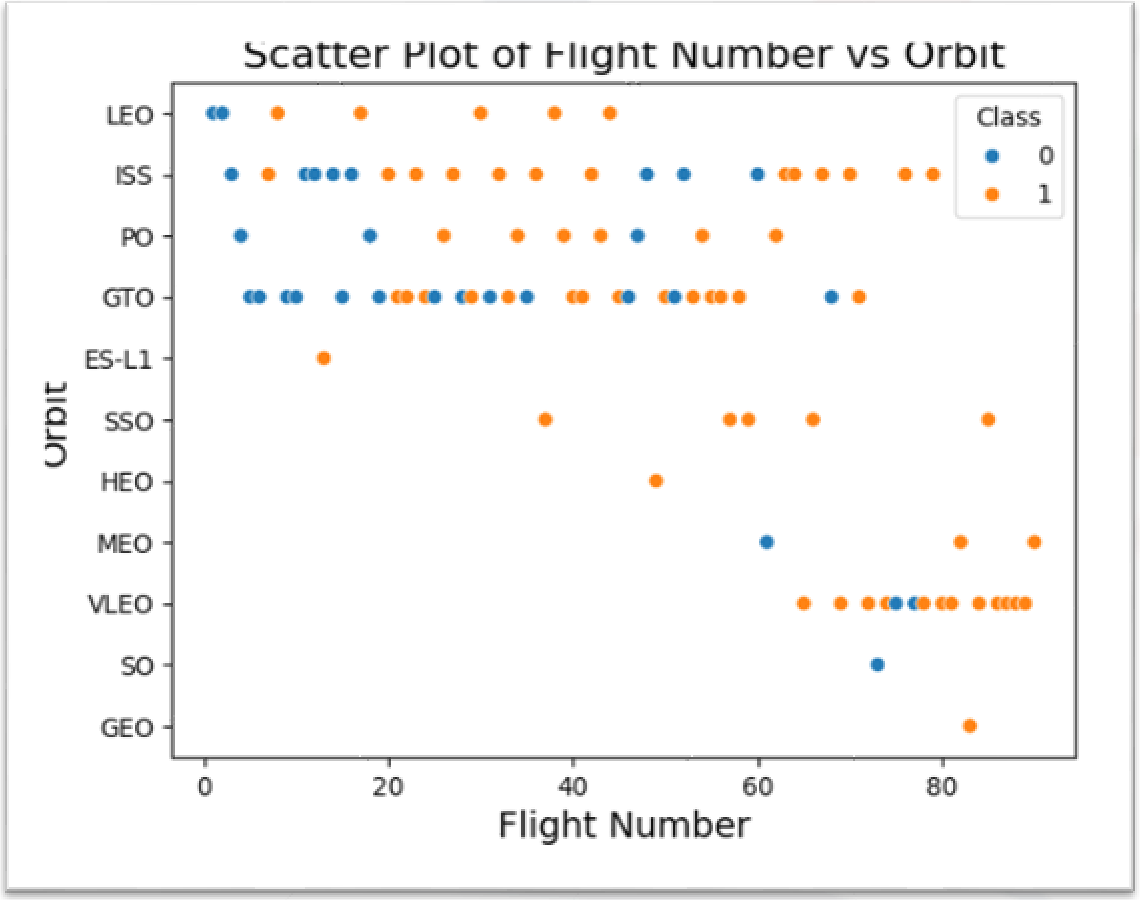


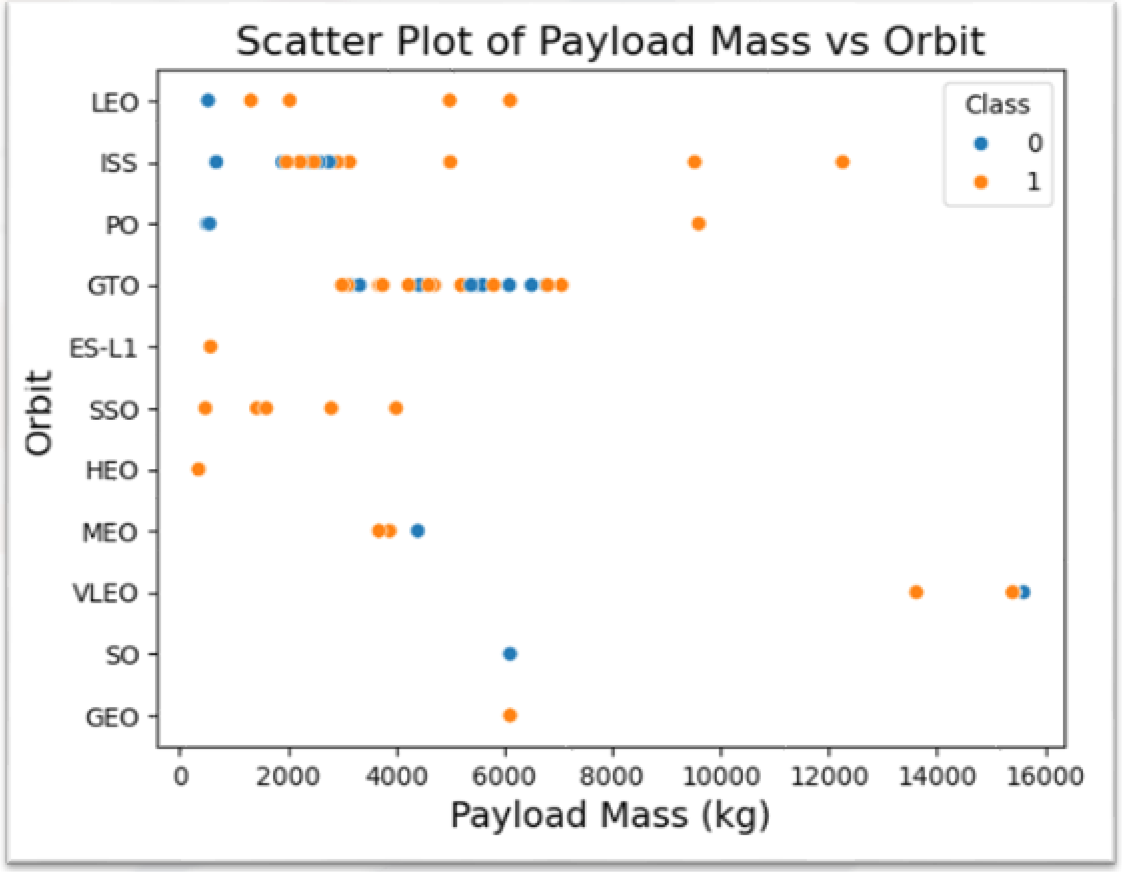
LAUNCH SITE VS FLIGHT NUMBER AND PAYLOAD MASS



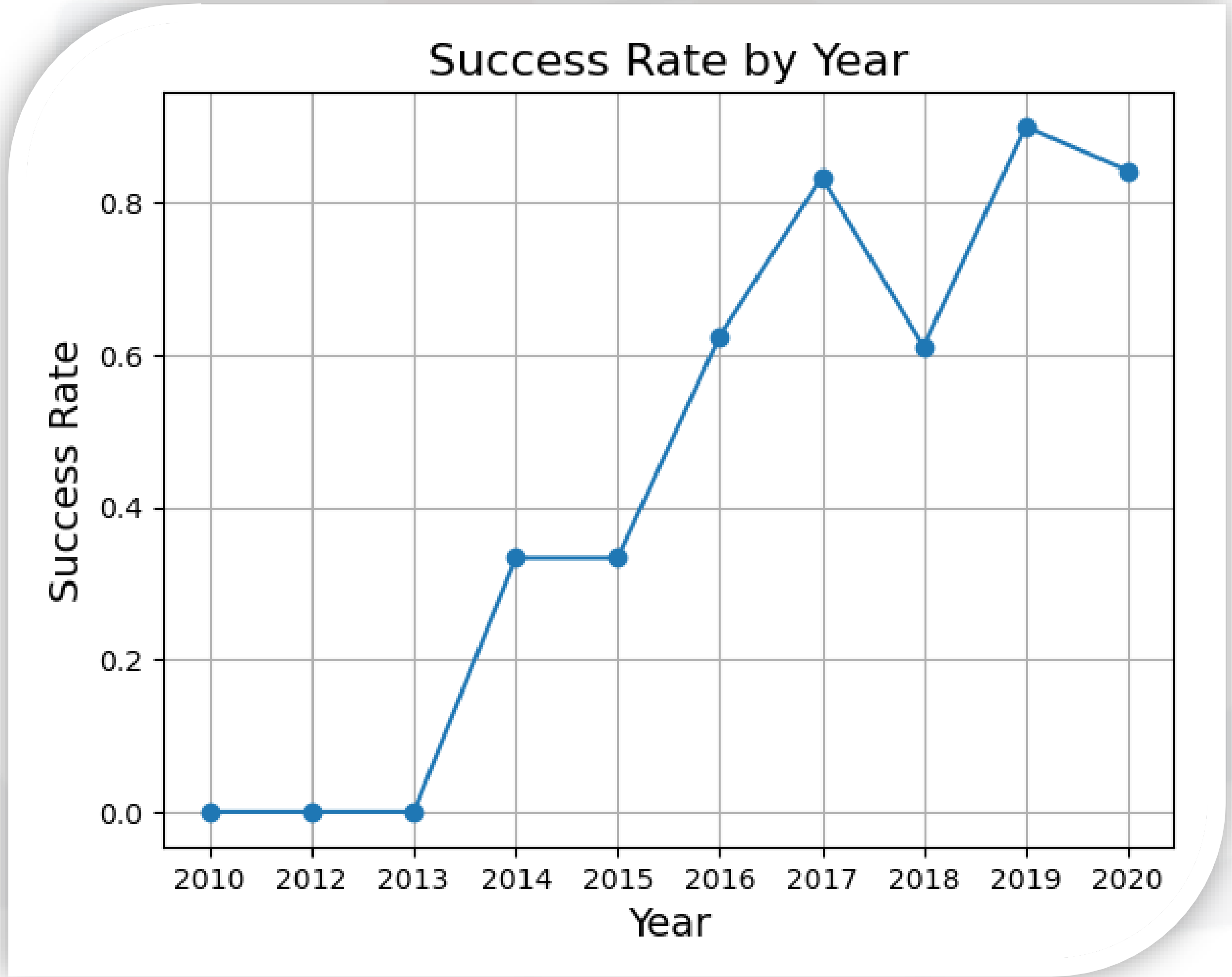


ORBIT TYPE VS FLIGHT NUMBER AND PAYLOAD MASS





YEARLY SUCCESS RATE



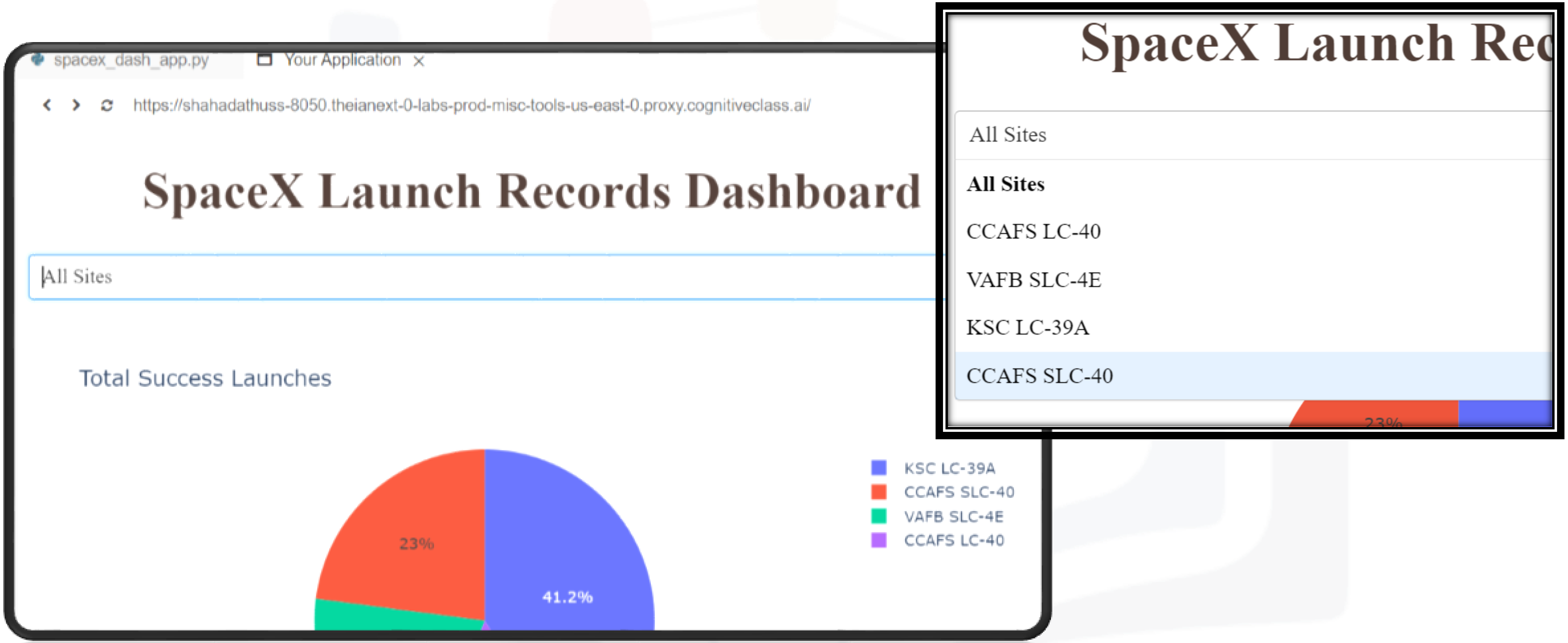
GEO-LOCATION MAP DATA VISUALIZATION

	Launch Site	Total No. of Lau nches	Successful L anding of First Stage
1	CCAFS LC-40	26	7
2	CCAFS SLC-40	7	3
3	KSC LC-39A	13	10
4	VAFB SLC-4E	10	4

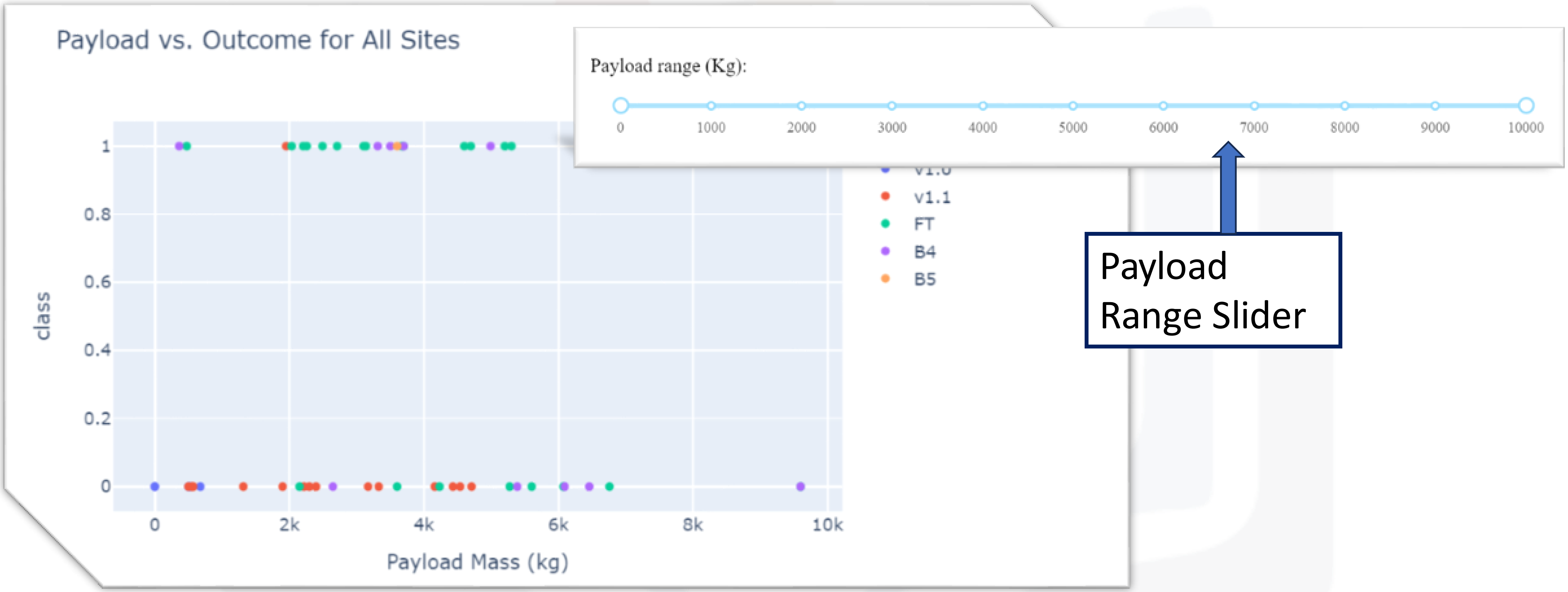
DASHBOARD

https://github.com/drdataengg/Capstone_Applied_Data_Science_IBM/blob/082882e665b1711d0e0ac46fcc36b2a4bf52a37c/spacex_dash_app.py

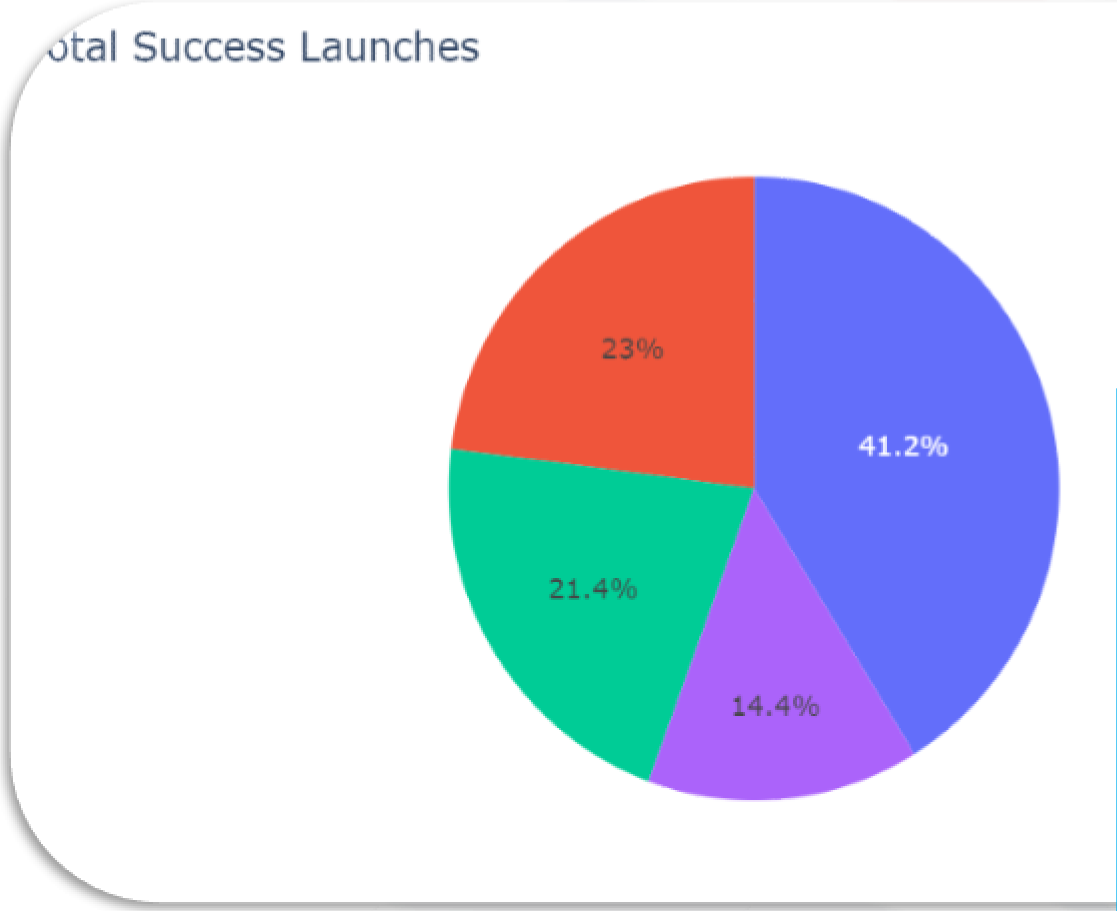
DASHBOARD TAB 1



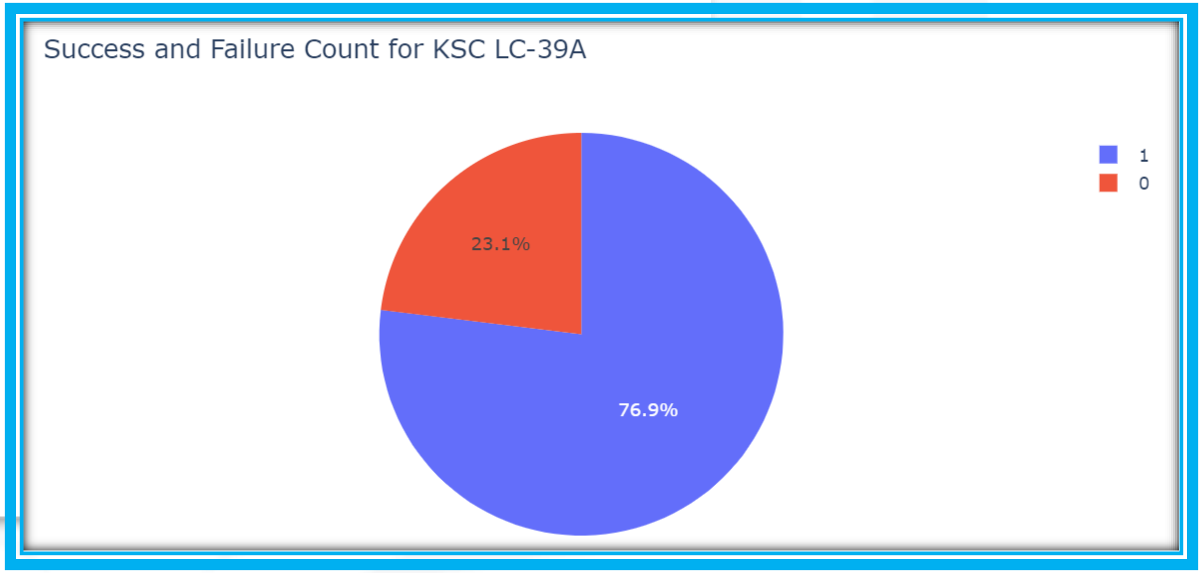
DASHBOARD TAB 2



DASHBOARD TAB 3



Launch Site with highest landing success



MACHINE LEARNING PREDICTION MODELS

TASK 12

Find the method performs best:

```
max_accuracy = max(accuracy_scores.values())
best_methods = [method for method, accuracy in accuracy_scores.items() if accuracy == max_accuracy]
print("Best performing method(s):", ', '.join(best_methods))
print("Accuracy of the best performing method(s):", max_accuracy)
```

```
Best performing method(s): Logistic Regression, Support Vector Machine, Decision Tree, K-Nearest Neighbors
Accuracy of the best performing method(s): 0.8333333333333334
```

- The best method for the prediction models was found to be Logistic Regression, Decision Tree, Support Vector Machine and K-Nearest Neighbors.
- They all have an accuracy of 83.33% on the test data modeled by the machine learning algorithms.

DISCUSSION

- Increasing flight numbers tend to correlate with a higher rate of successful landings, indicating a potential continuous learning process among engineers involved in the project, building on previous results.
- The success rate at each launch site tends to increase as the flight number increases, suggesting a learning curve and improvement in processes over time.
- KSC LC 39A exhibits the highest success rate among all launch sites.
- The majority of payload masses sent are below 7000 kg.
- For payloads exceeding 12000 kg, CCAFS SLC-40 demonstrates a better success rate compared to other launch sites.
- SSO orbit shows a 100% success rate, with LEO orbit also demonstrating higher success rates.
- Beyond 2013, there is an observed increase in success rate with each passing year.
- All four launch sites are situated closer to the coastline, railway, and highway, while being farther from urban areas. This placement ensures safety in case of failure, as launch sites are kept away from densely populated areas. Additionally, proximity to the coast, railway, and highway provides logistical advantages for transportation of heavy rocket or payload objects.

OVERALL FINDINGS & IMPLICATIONS

Findings

- The success rates at each launch site tend to improve with increasing flight numbers, reflecting a learning curve and refinement of processes.
- Beyond 2013, there is a notable increase in success rates, indicating advancements in technology and operational practices.
- The strategic positioning of launch sites away from urban areas, closer to the coastline, railway, and highway, ensures public safety in the event of failure and facilitates efficient logistics for transporting heavy payloads.

Implications

- The findings suggest that the aerospace industry is evolving, with a focus on continuous improvement and learning from past experiences.
- The observed increase in success rates over time underscores the importance of ongoing research, development, and collaboration within the space exploration community.
- The strategic placement of launch sites highlights the significance of safety considerations and logistical efficiency in mission planning and execution.

CONCLUSION

- The analysis indicates a positive trend of improving success rates in space missions, driven by continuous learning and refinement of engineering practices.
- KSC LC 39A emerges as an effective launch site, while CCAFS SLC-40 shows promise for handling heavier payloads.
- Consistent success rates in specific orbits highlight the significance of strategic mission planning.
- The observed increase in success rates beyond 2013 reflects advancements in technology and operational practices, emphasizing the dynamic nature of the aerospace industry.

APPENDIX

Calculation of distance between launch sites and other points of interests

	Launch Site	Coordinates	Total Launches	Success	Coastline	Railway	Highway	City	Distance to Coastline	Distance to Railway	Distance to Highway	Distance to City
0	CCAFS LC-40	[28.562302, -80.577356]	26	7	[28.56288, -80.56789]	[28.56178, -80.58718]	[28.56235, -80.57063]	[28.08773, -80.65567]	0.926991	0.961492	0.657104	53.34058
1	CCAFS SLC-40	[28.563197, -80.57682]	7	3	[28.56288, -80.58794]	[28.56178, -80.58718]	[28.56235, -80.57063]	[28.08773, -80.65567]	1.086910	1.024296	0.612011	53.44665
2	KSC LC-39A	[28.573255, -80.646895]	13	10	[28.58015, -80.64205]	[28.57317, -80.65404]	[28.57303, -80.6537]	[28.51456, -81.35834]	0.901196	0.698008	0.665203	69.82107
3	VAFB SLC-4E	[34.632834, -120.610745]	10	4	[34.63487, -120.62503]	[34.63519, -120.62373]	[34.63225, -120.59916]	[34.64874, -120.45742]	1.326832	1.216936	1.062264	14.14223