IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Snehar Singh Gujral
27/12/2024

Outline

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

Executive Summary

Objective:

Predict the successful landing of SpaceX Falcon 9's first stage using historical launch data.

Key Findings:

High accuracy models achieved over 90% accuracy.

ROC-AUC scores indicate excellent classification performance.

Tools and Techniques Used:

Python, Pandas, NumPy, Requests, SQL, Folium, Plotly Dash, Scikit-learn.

Introduction

Background	Space X 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 Space X 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 space X for a rocket launch.
Problem Statement	Predicting the success of first-stage landings to optimize launch operations and reduce costs.
Goals:	Analyze historical launch data. Develop predictive models to forecast landing success. Create interactive visualizations for data-driven insights

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

- https://github.com/irhen-chan/-IBM-Data-Science-Professional-Certificate/blob/8ab8b2034da8f3e3a414c3bab608b7fc0944fbd7/applied_data_sci_project/jupyter-labs-spacex-data-collection-api.ipynb

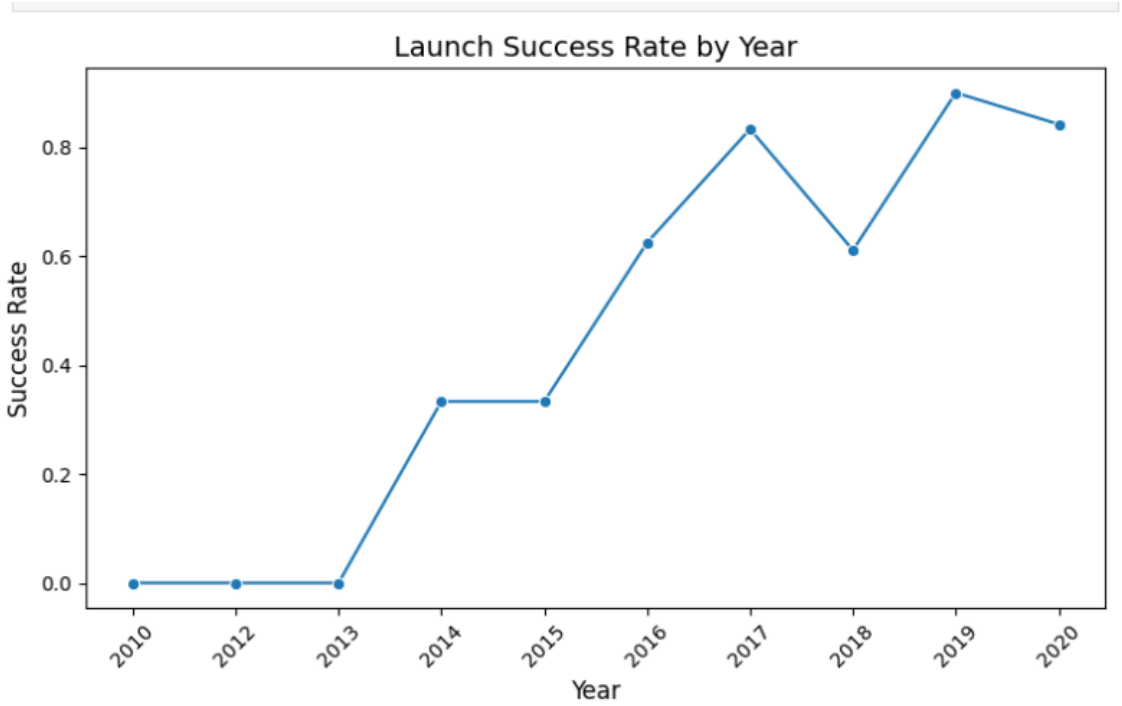
```
import requests
import pandas as pd

spacex_url = "https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)
launches_data = response.json()
df = pd.DataFrame(launches_data)
```

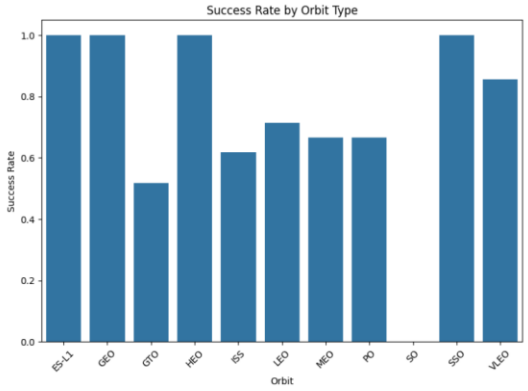
Data Wrangling

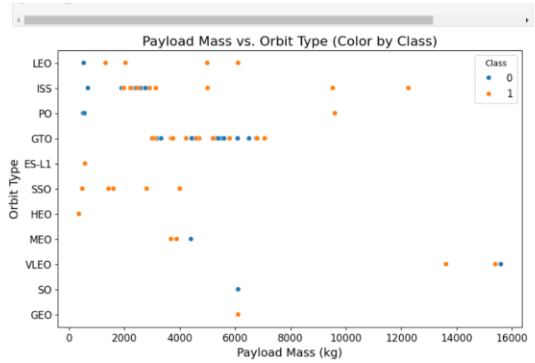
- Handling missing values.
- Encoding categorical variables using One-Hot Encoding.
- Feature engineering (e.g., extracting launch dates, booster versions)
- https://github.com/irhen-chan/-IBM-Data-Science-Professional-Certificate/blob/8ab8b2034da8f3e3a414c3bab608b7fc0944fbd7/applied_data_sci_project/labs-jupyter-spacex-Data%20wrangling.ipynb.

EDA with Data Visualization

https://github.com/irhen-chan/-IBM-Data-Science-Professional-Certificate/blob/8ab8b2034da8f3e3a414c3bab608b7fc0944fbd7/applied_data_sci_project/edadataviz.ipynb







EDA with SQL

Done.

[57]:

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

Done.

[52]:

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

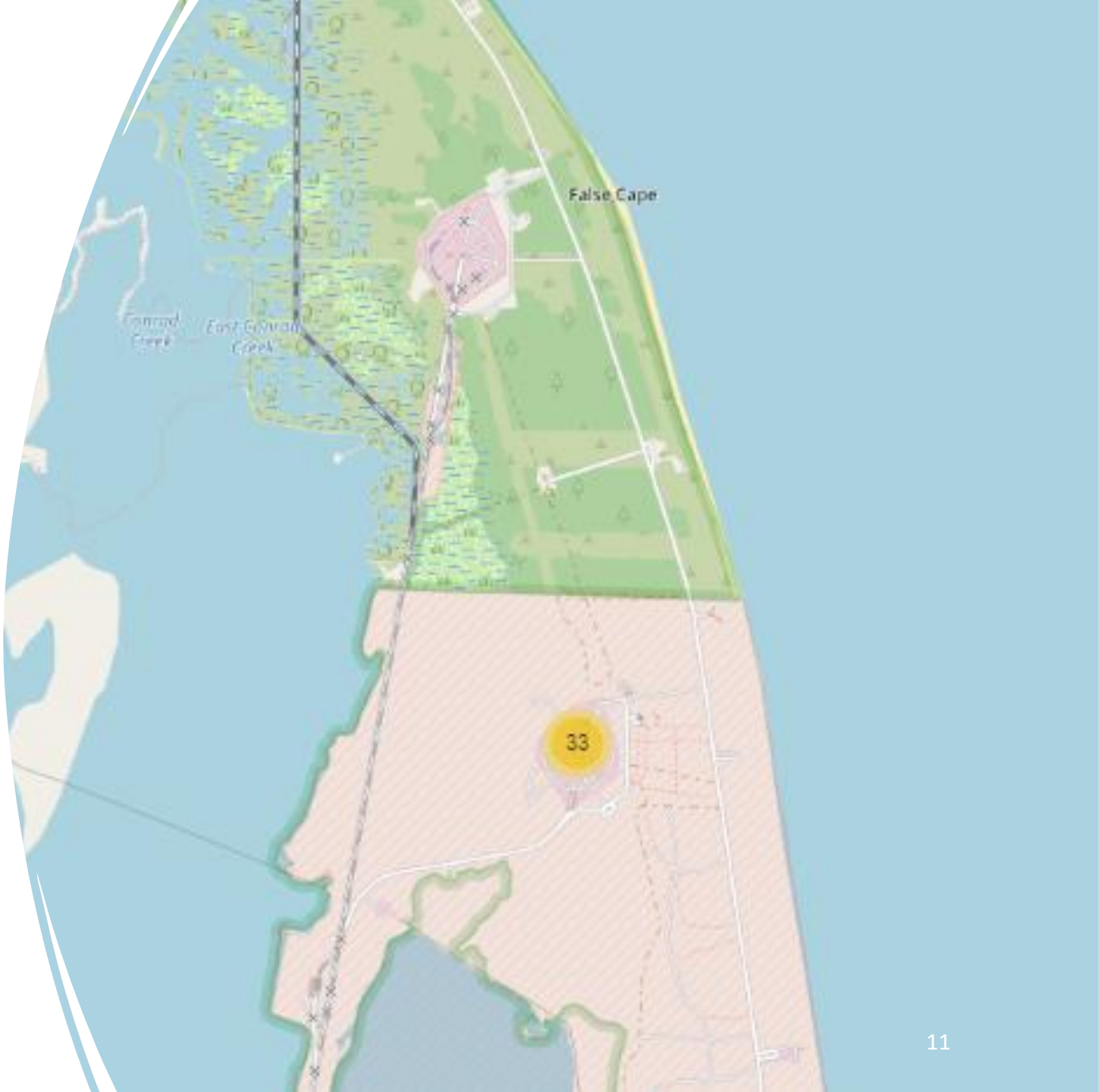
]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	
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	
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	

https://github.com/irhen-chan/-IBM-Data-Science-Professional-Certificate/blob/8ab8b2034da8f3e3a414c3bab608b7fc0944fbd7/applied_data_sci_project/jupyter-labs-eda-sql-coursera_sqllite.ipynb

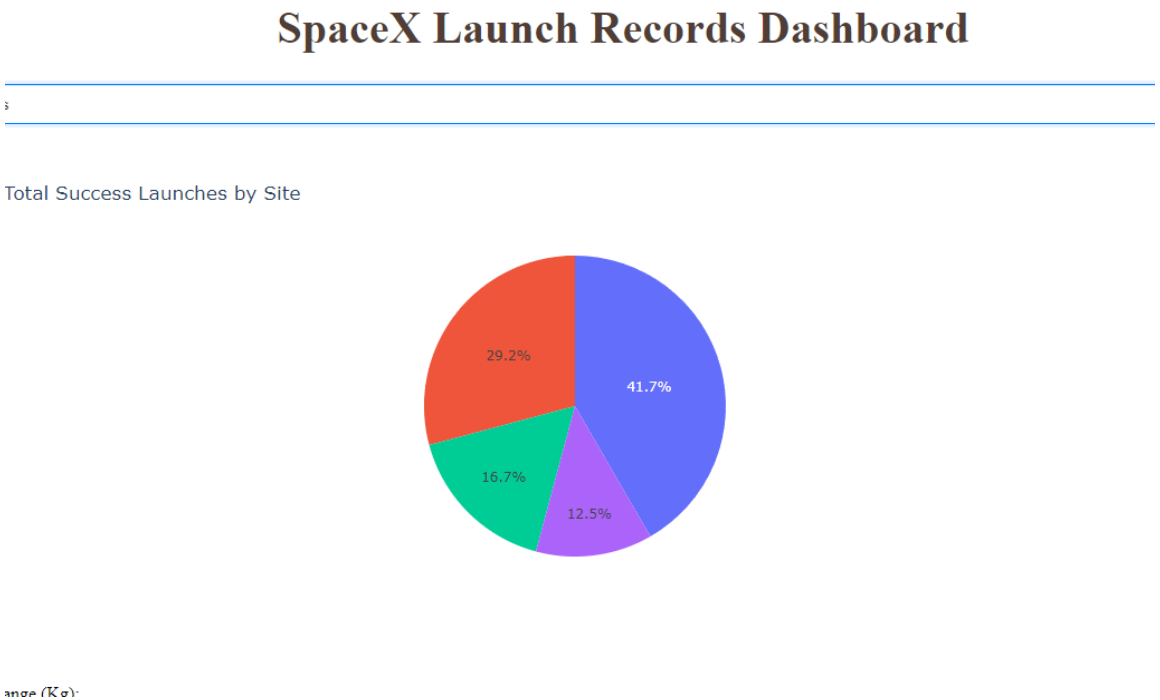
Build an Interactive Map with Folium

- https://github.com/irhen-chan/-IBM-Data-Science-Professional-Certificate/blob/8ab8b2034da8f3e3a414c3bab608b7fc0944fbd7/applied_data_sci_project/lab_jupyter_launch_site_location.ipynb



Build a Dashboard with Plotly Dash

- https://github.com/irhen-chan/-IBM-Data-Science-Professional-Certificate/blob/8ab8b2034da8f3e3a414c3bab608b7fc0944fbd7/applied_data_sci_project/spacex_dash_app.py

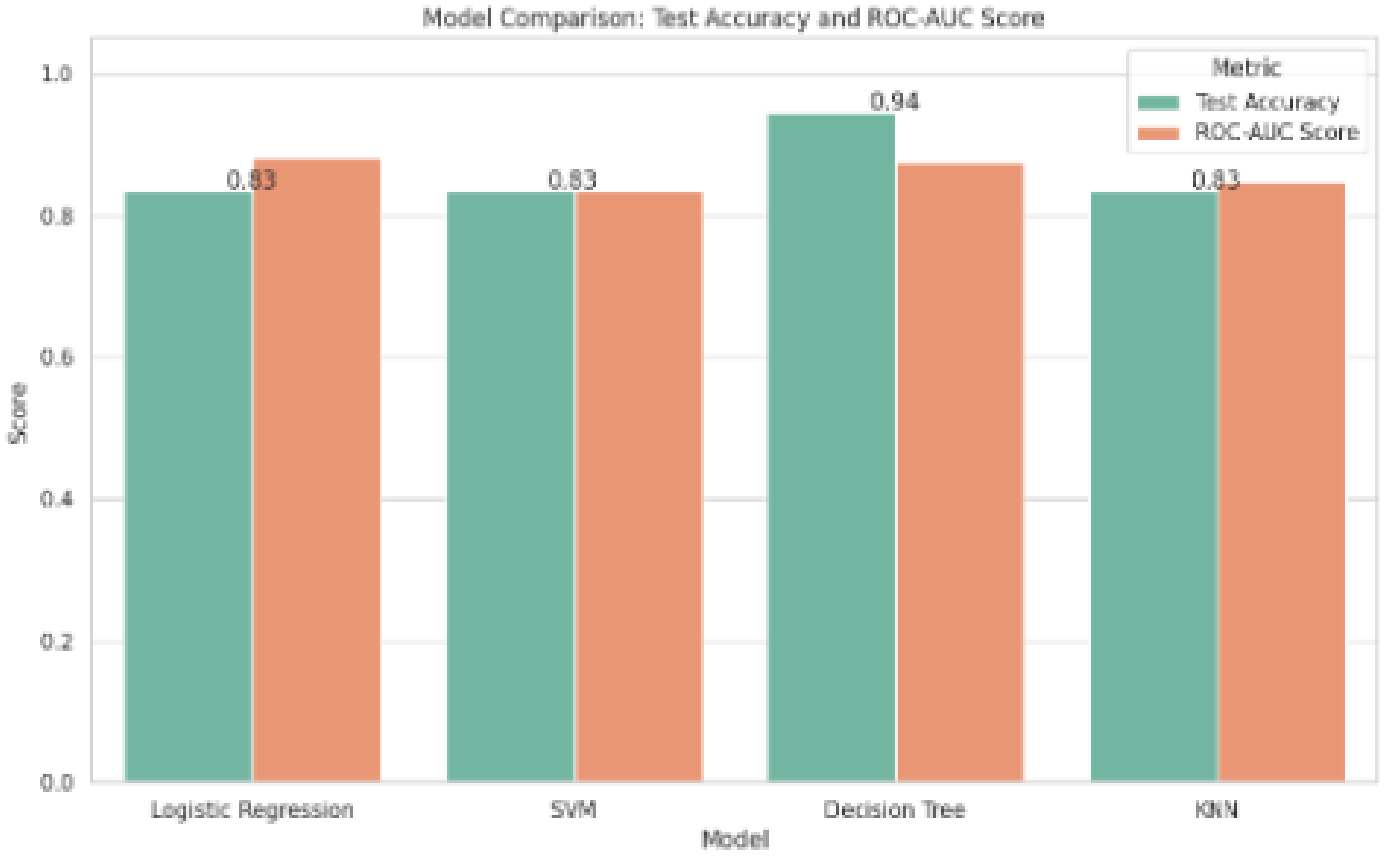


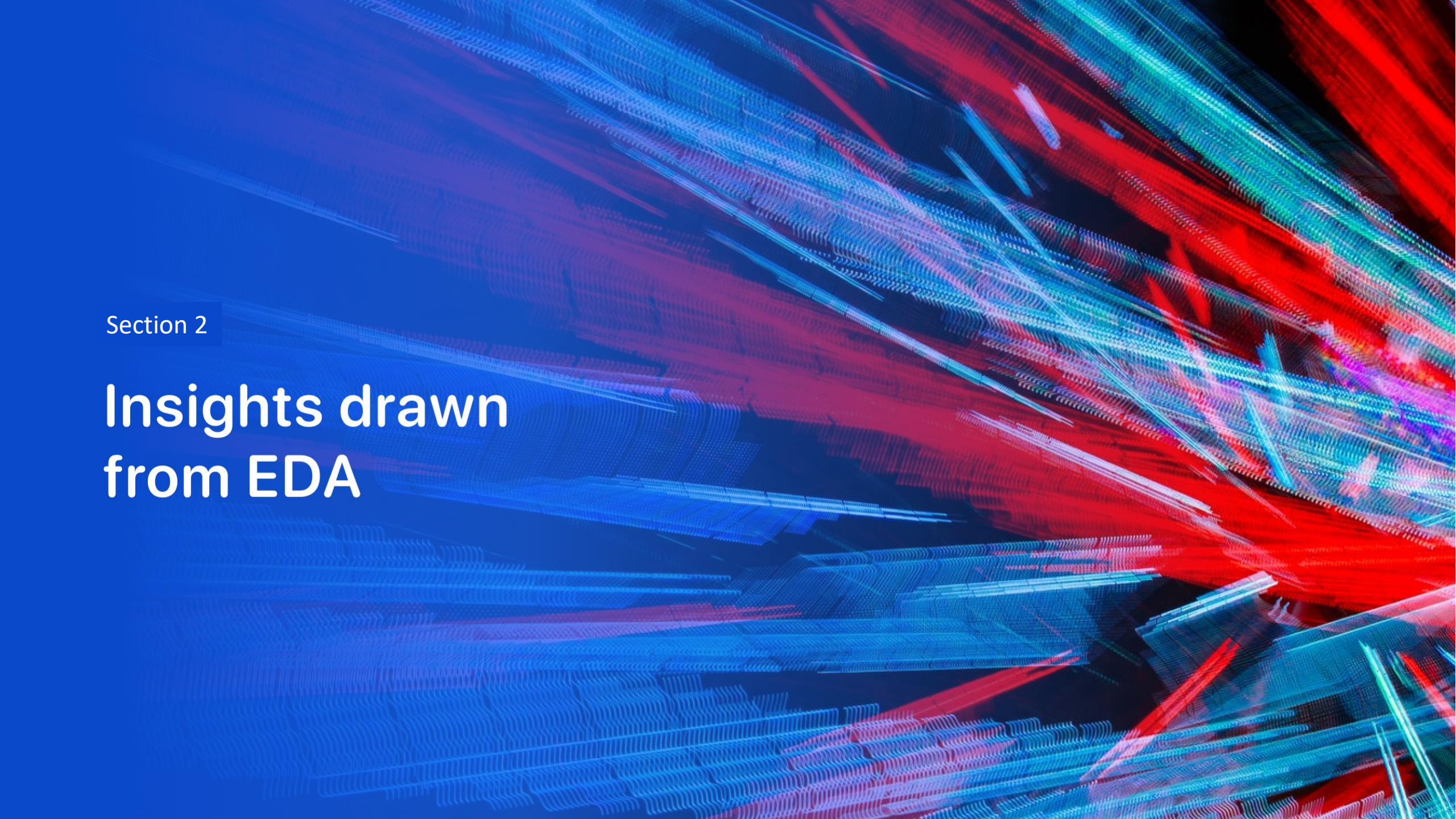
Predictive Analysis (Classification)

- https://github.com/irhen-chan/-IBM-Data-Science-Professional-Certificate/blob/8ab8b2034da8f3e3a414c3bab608b7fc0944fbd7/applied_data_sci_project/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

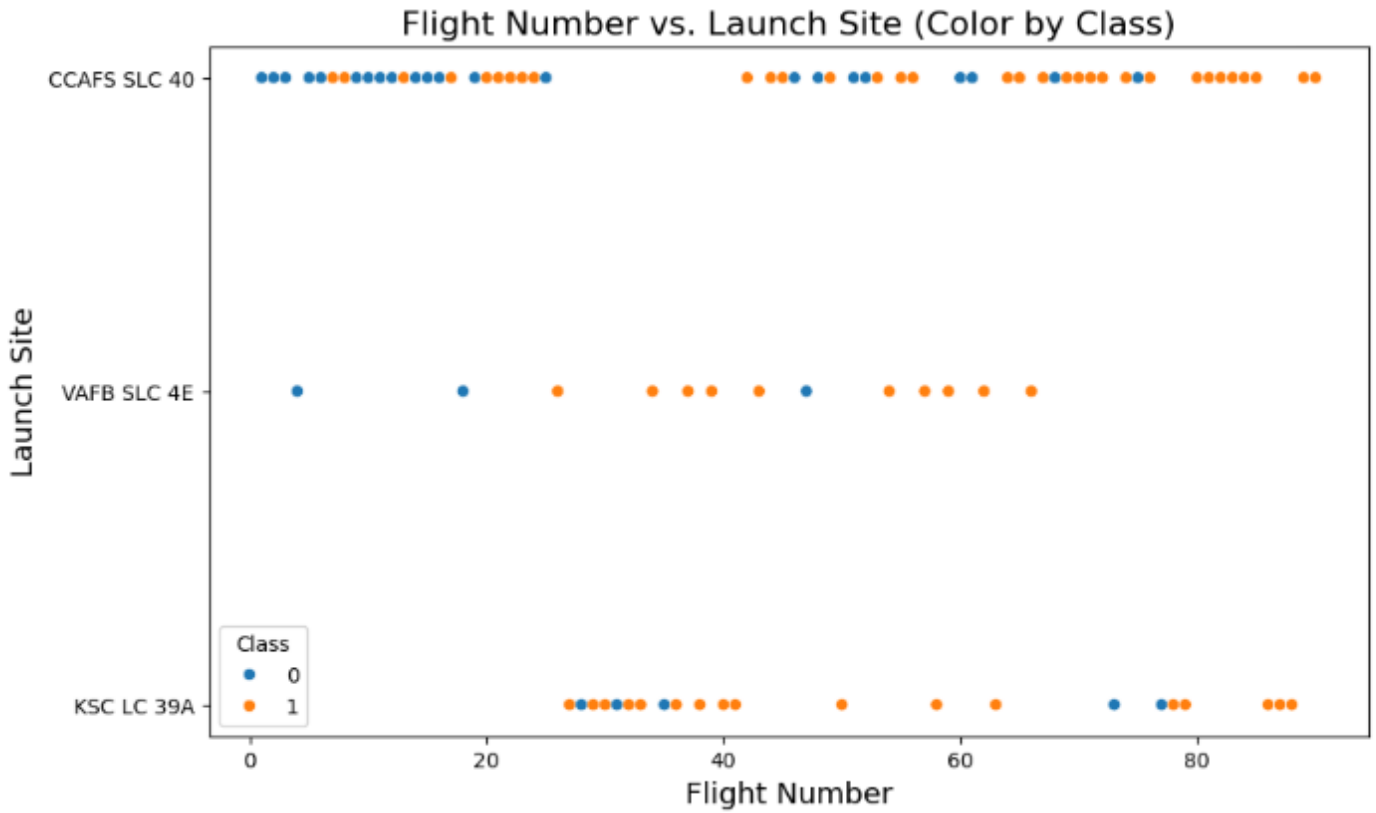
- **Models Evaluated:**
  - Logistic Regression
  - Support Vector Machines (SVM)
  - Decision Trees
  - K-Nearest Neighbors (KNN)

Results

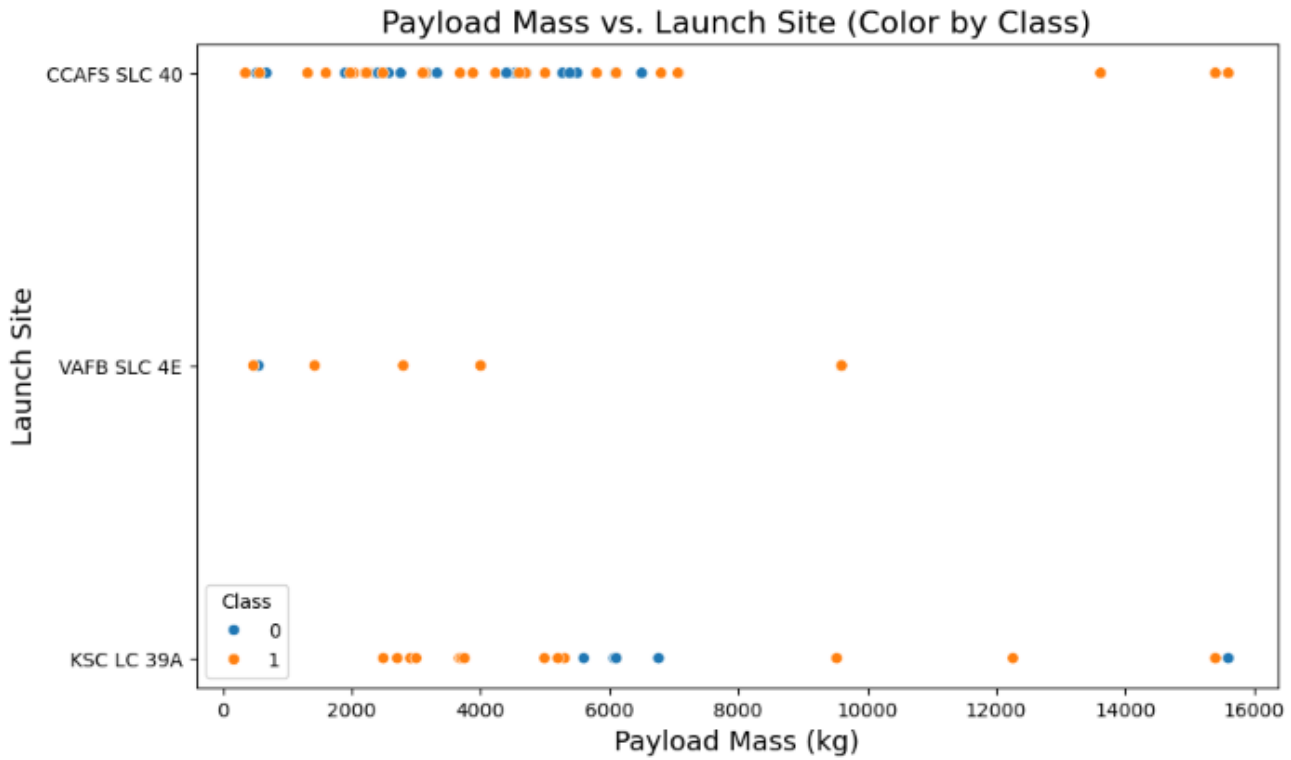

Model Comparison: Test Accuracy and ROC-AUC Score
Metric
Test Accuracy
ROC-AUC Score
1.0
0.8
0.6
0.4
0.2
0.0
Score
0.83
0.83
0.94
0.83
Logistic Regression
SVM
Decision Tree
KNN
Model


Section 2
Insights drawn from EDA

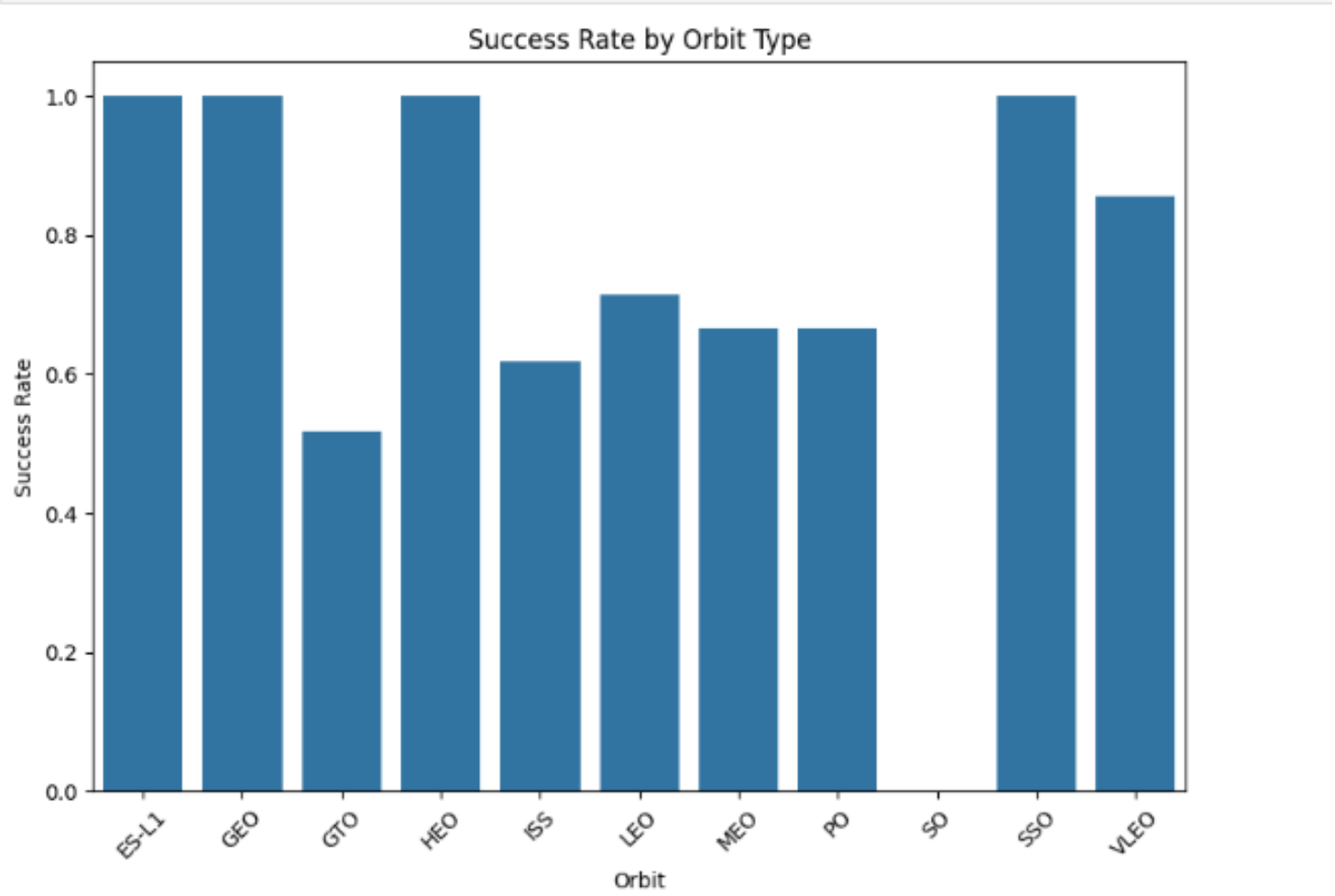
Flight Number vs. Launch Site



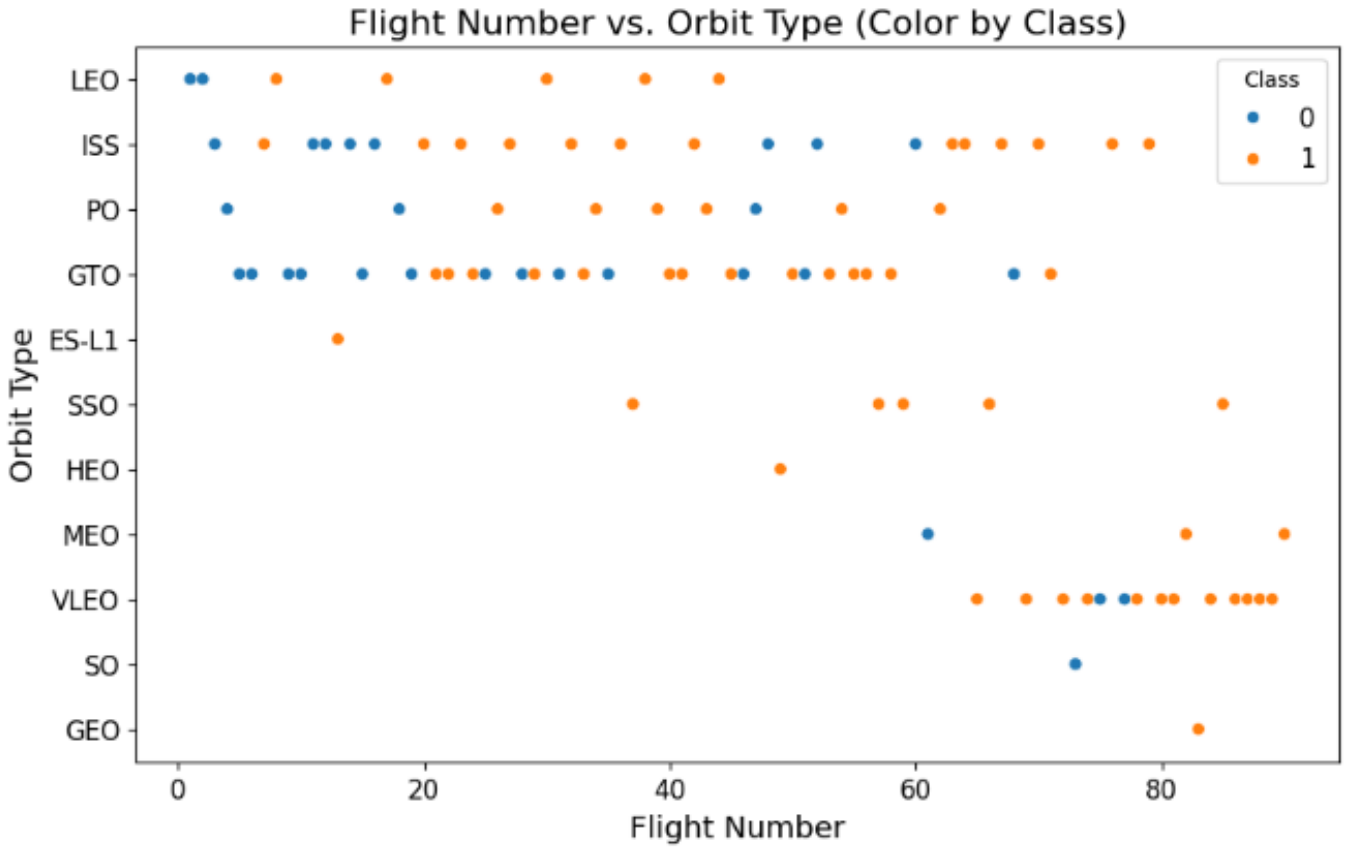
Payload vs. Launch Site



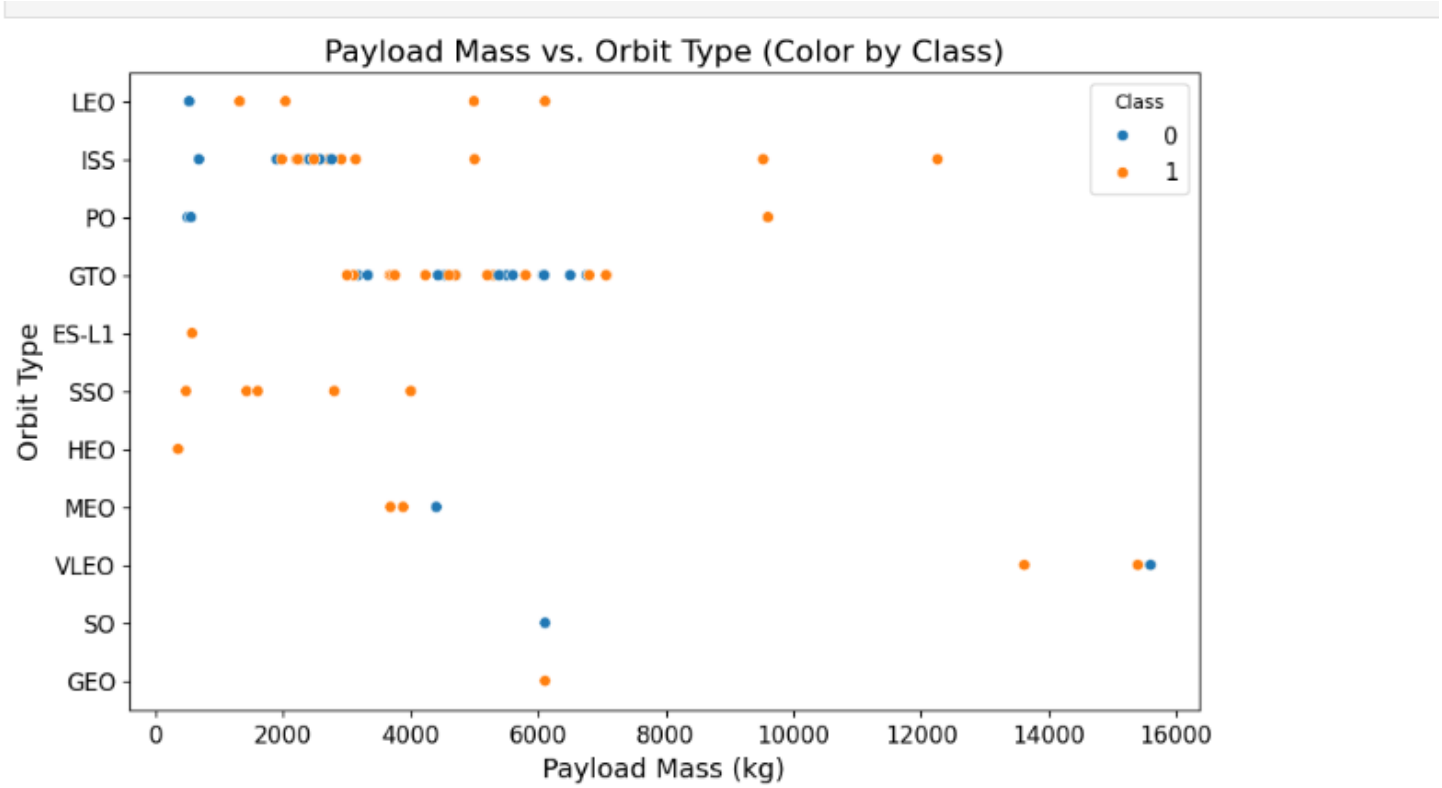
Success Rate vs. Orbit Type



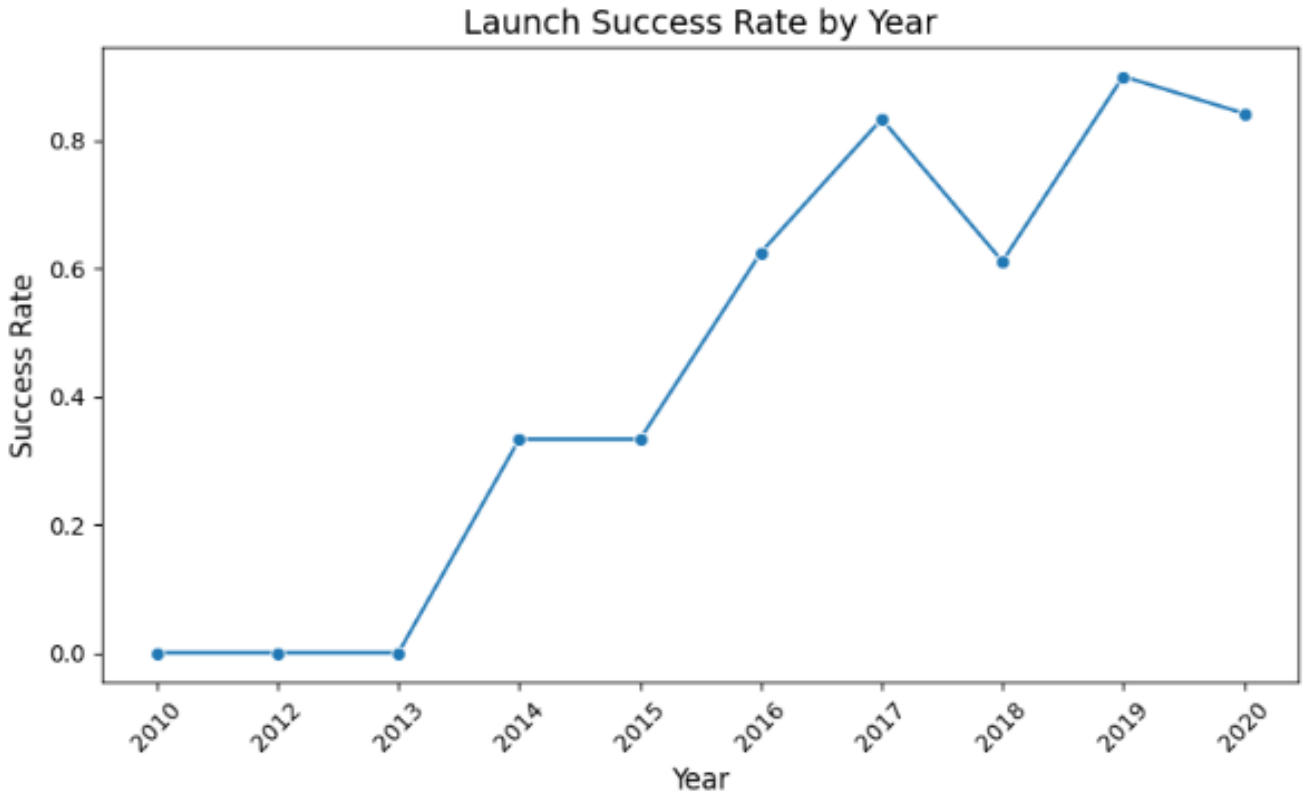
Flight Number vs. Orbit Type

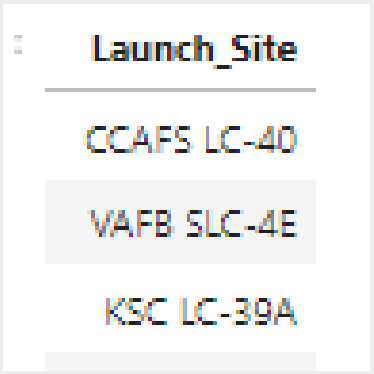


Payload vs. Orbit Type



Launch Success Yearly Trend





All Launch Site Names

Launch Site Names Begin with 'CCA'

```
]: %sql SELECT * FROM SPACEXTABLE 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_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

Task 3

Total Payload Mass

Total_Payload
48213

Average Payload Mass by F9 v1.1

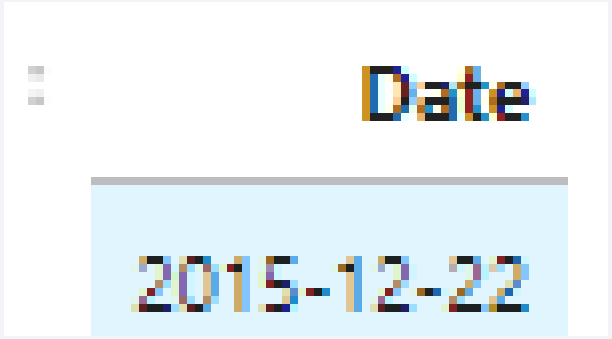
[41]:
```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS "Avg_Payload" FROM SPACEXTABLE WHERE "Booster_Version" like 'F9 v1.1%';
```

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

[41]:

Avg_Payload
2534.6666666666665

First Successful Ground Landing Date



Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT "Booster_Version"FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)'  AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000;
```

```
 * 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

List the total number of successful and failure mission outcomes ¶

```
]: %sql SELECT "Mission_Outcome", COUNT(*) AS "Outcome_Count" FROM SPACEXTABLE GROUP BY "Mission_Outcome";
```

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

]:

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

Boosters Carried Maximum Payload

```
55]: %sql select "Booster_Version" from SPACEXTABLE where "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") From SPACEXTABLE)
```

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

55]:

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

56]:

Month	Landing_Outcome	"BoosterVersion"	Launch_Site
01	Failure (drone ship)	BoosterVersion	CCAFS LC-40
04	Failure (drone ship)	BoosterVersion	CCAFS LC-40

2015 Launch Records

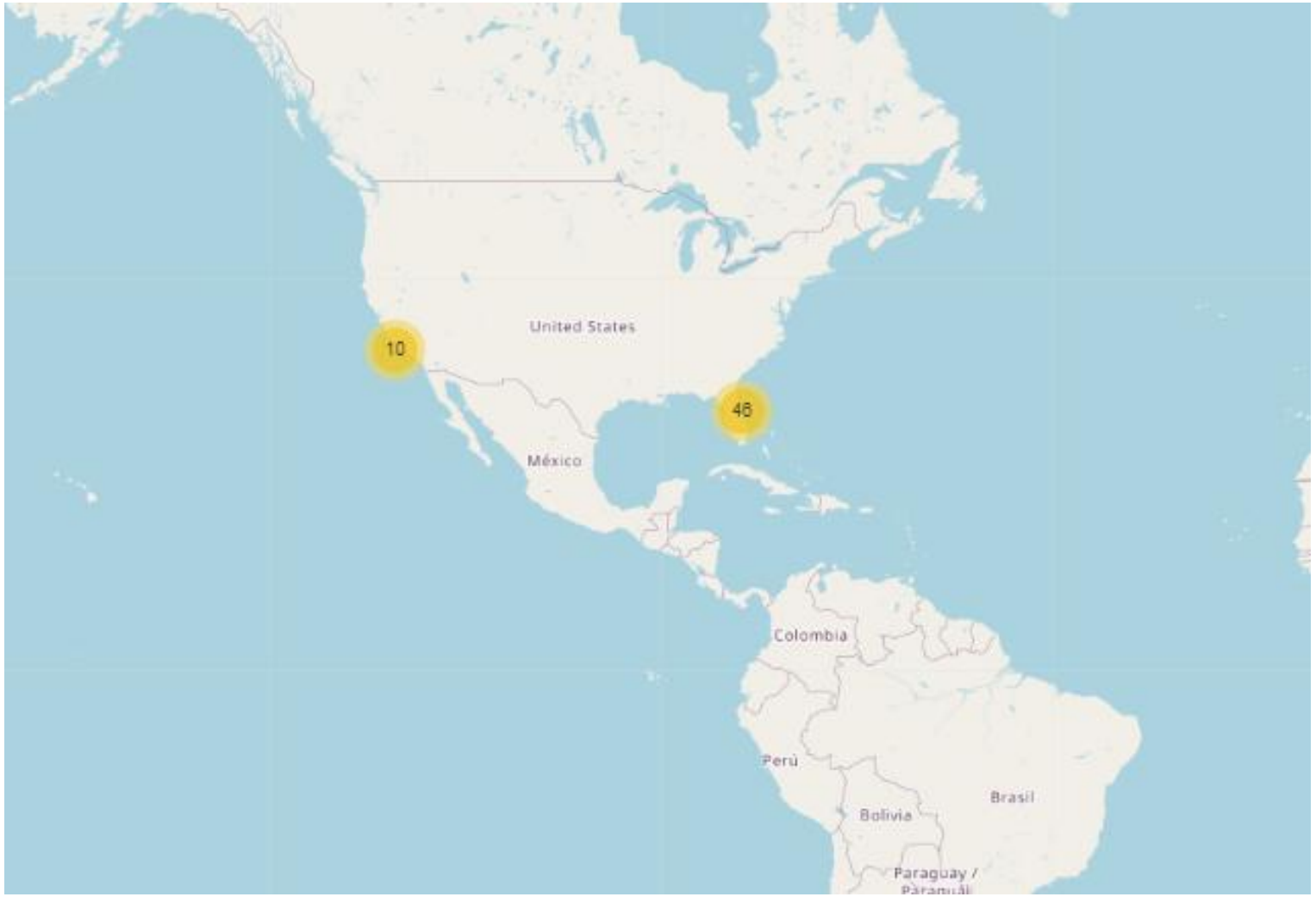
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

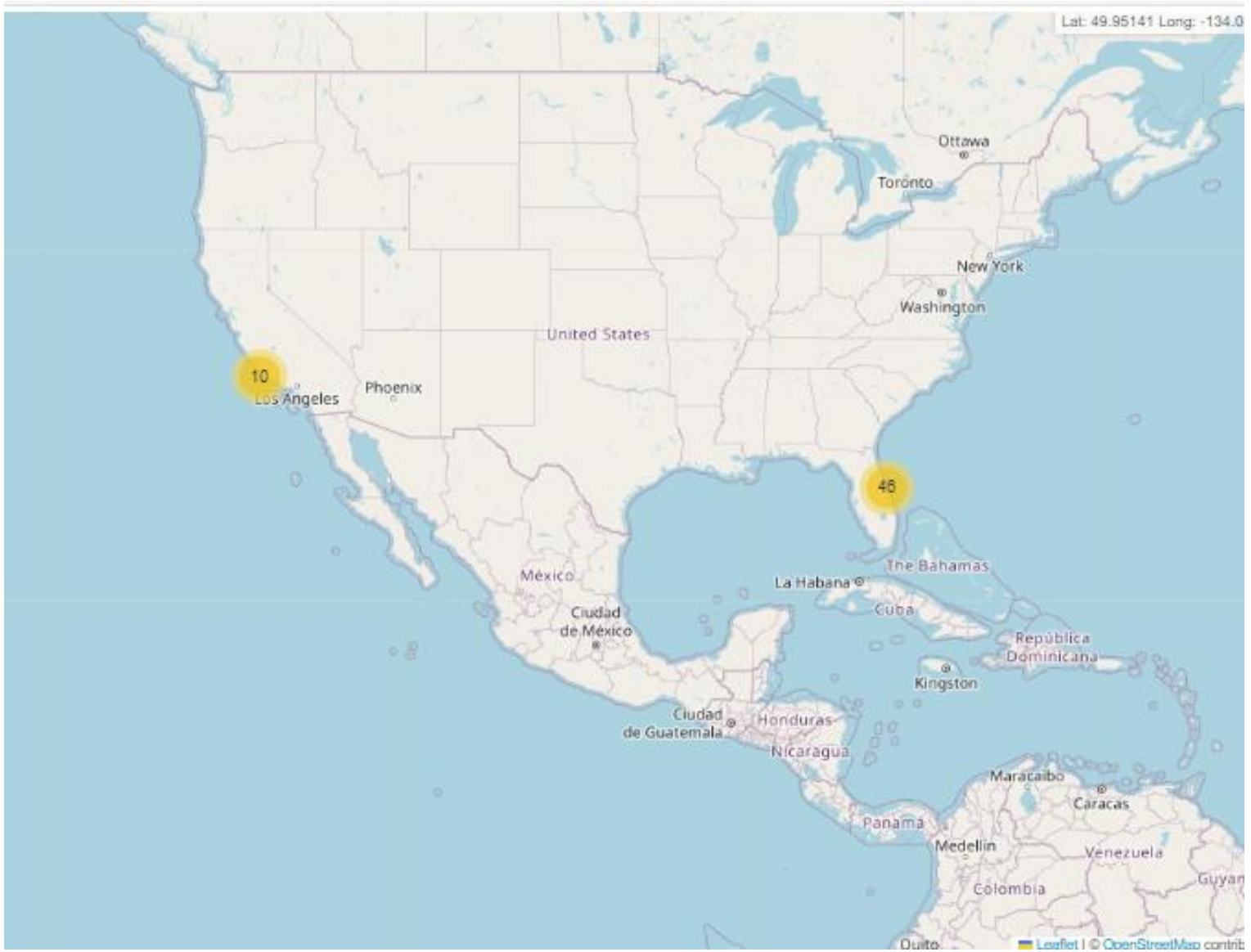
Section 3

Launch Sites Proximities Analysis

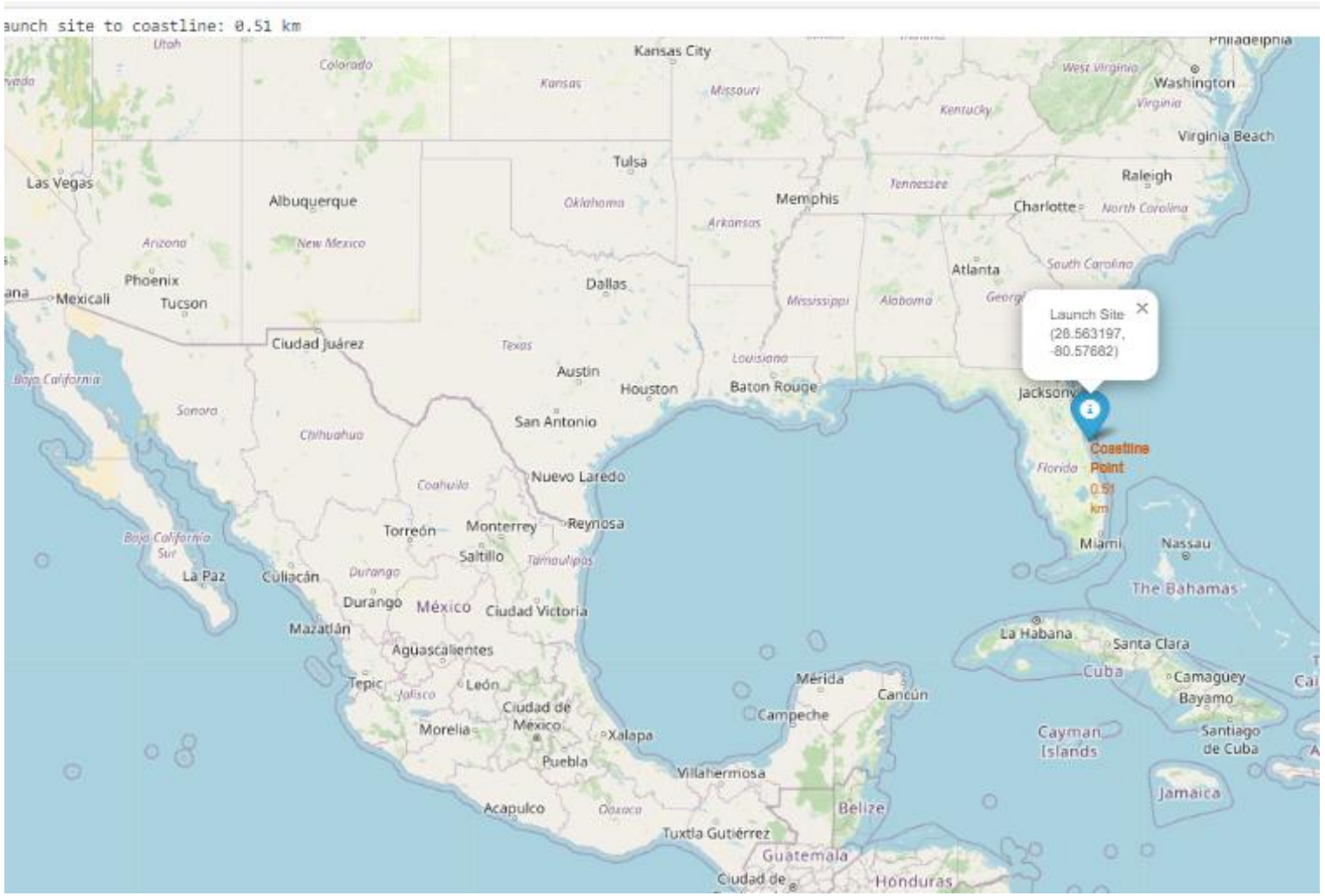
<Folium Map Screenshot 1>

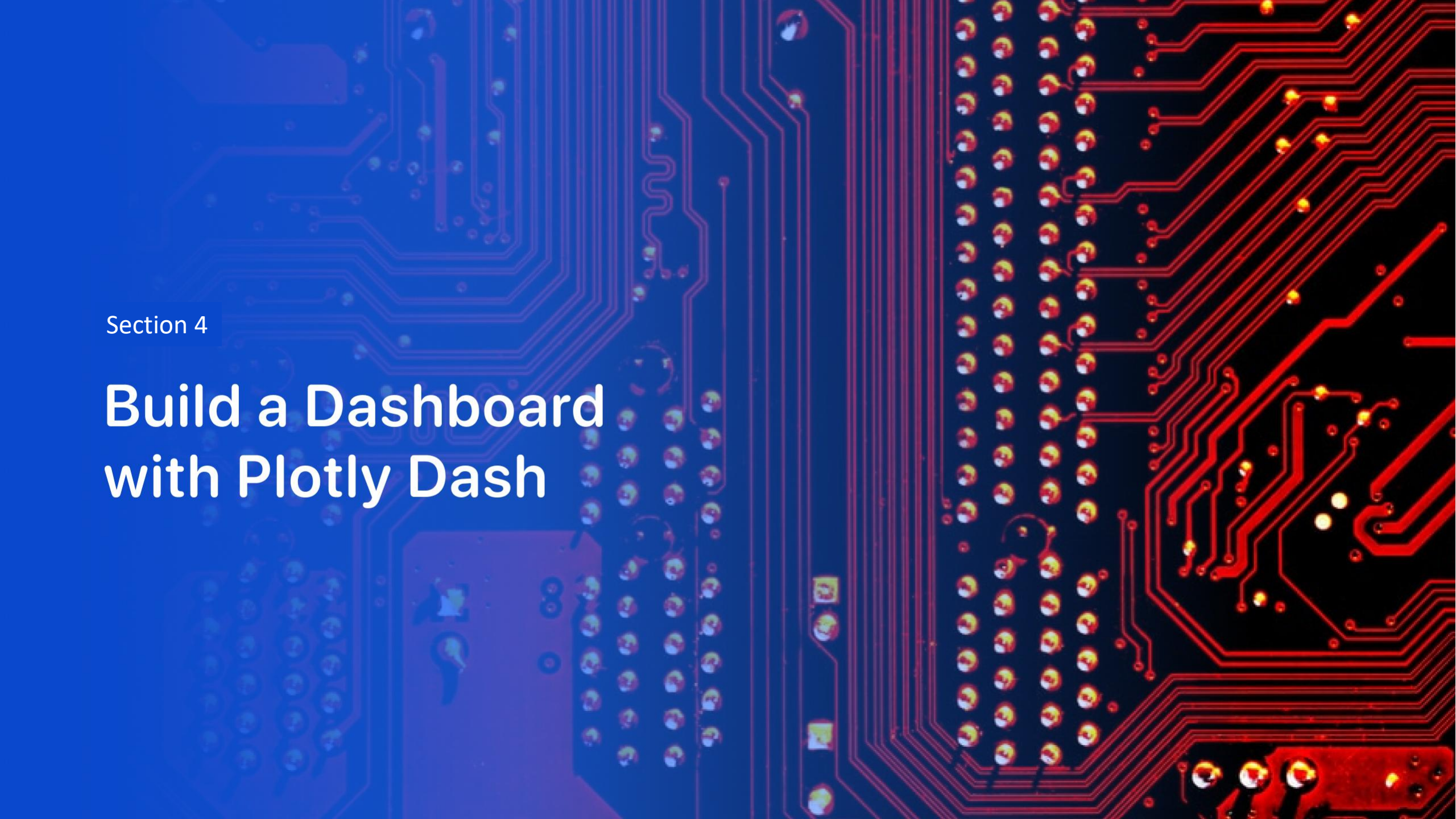


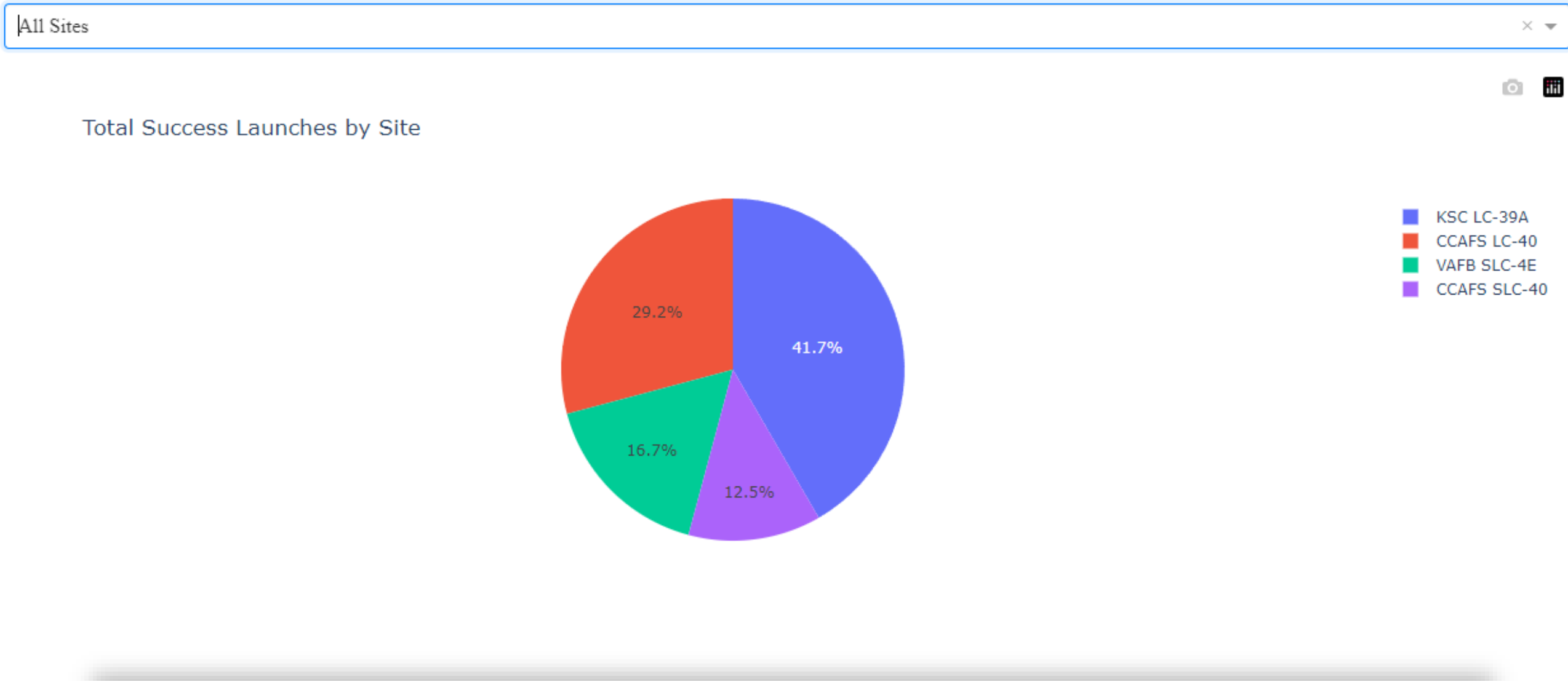
<Folium Map Screenshot 2>



<Folium Map Screenshot 3>




Section 4
Build a Dashboard with Plotly Dash



<Dashboard Screenshot 1>

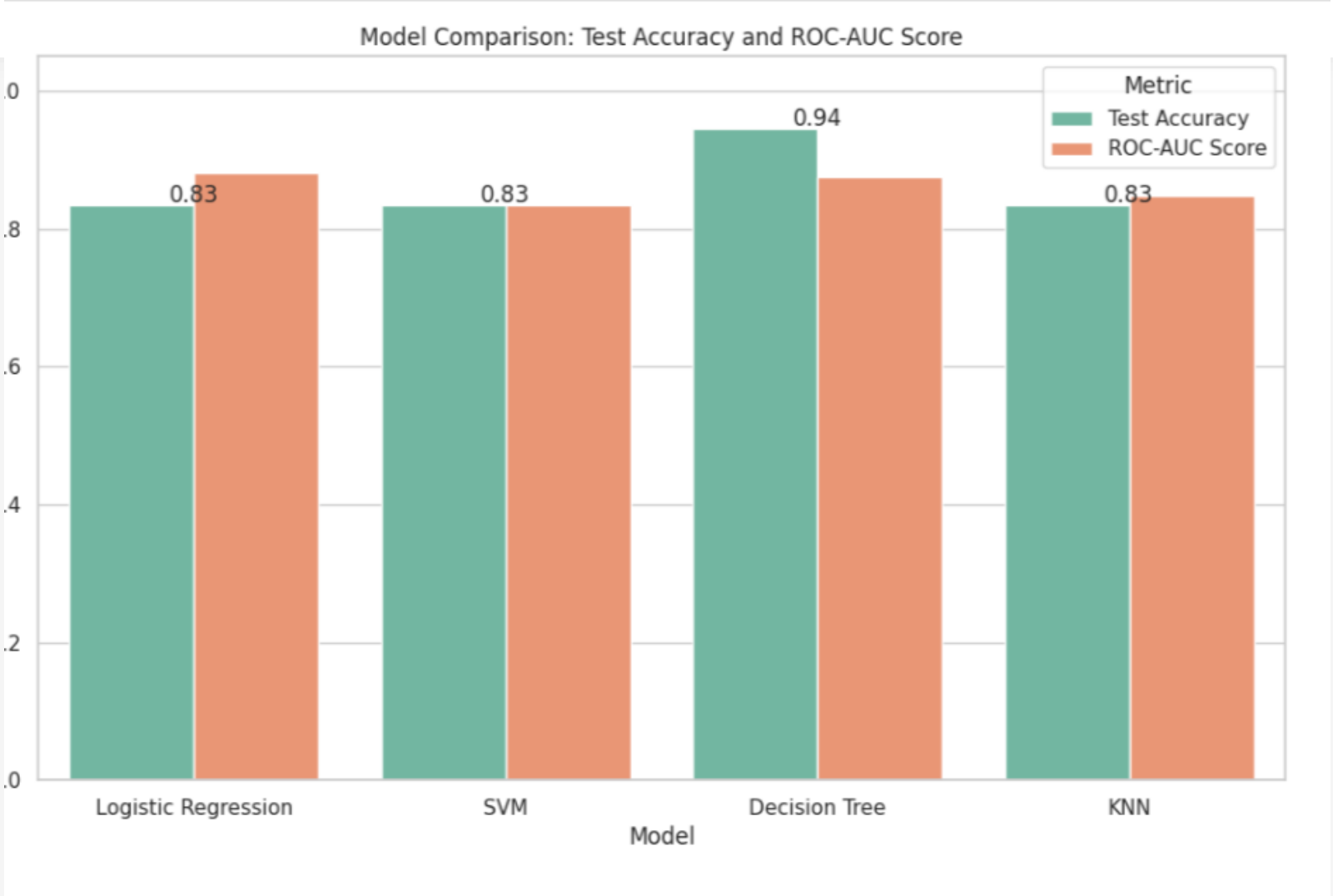
<Dashboard Screenshot 2>

<Dashboard Screenshot 3>

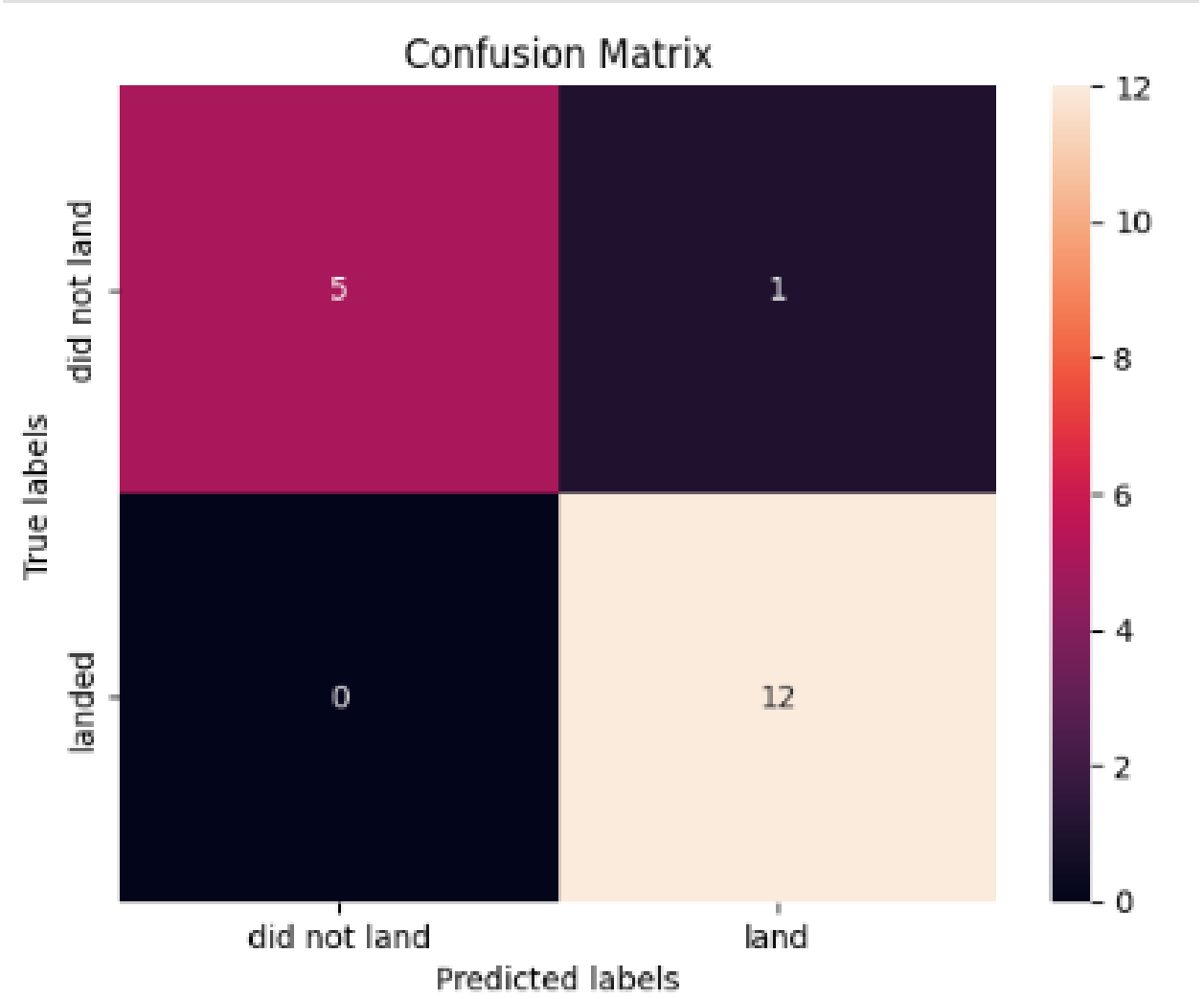
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- **Summary of Findings:**
  - Successful prediction of first-stage landing with high accuracy.
  - Identification of key factors influencing landing success.
- **Implications:**
  - Enhancing operational efficiency for SpaceX.
  - Reducing costs through better prediction models.
- **Future Work:**
  - Incorporating real-time data for dynamic predictions.
  - Exploring ensemble models for improved performance.
  - Deploying the dashboard for broader accessibility.

Thank you!