IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- The analysis begins with data collection from the SpaceX API, followed by a rigorous cleaning and preprocessing phase to ensure data quality. Exploratory Data Analysis (EDA) uncovers key patterns using visualizations and summary statistics. Feature engineering is applied to enhance predictive modeling, and machine learning techniques, such as logistic regression and Support Vector Machines, are used to identify factors influencing launch success. Cross-validation and hyperparameter tuning optimize model performance. Data visualization highlights key insights, ensuring clarity and actionable recommendations.
- The analysis is expected to reveal trends in SpaceX launch success rates, highlighting influential factors like rocket type, payload weight, and launch location. We anticipate finding correlations between payload characteristics and mission outcomes, alongside a clear timeline of improvements in launch success. Additionally, geographical analysis should pinpoint the most reliable launch sites. The final insights aim to offer recommendations for optimizing mission planning and identify data-driven factors contributing to SpaceX's increasing success over time.

Introduction

- SpaceX, a leading aerospace manufacturer and space transportation company, has revolutionized the industry with reusable rockets and ambitious goals of making space travel more accessible. As the company continues to innovate and expand, understanding the factors that drive the success of its missions is crucial. This project leverages data from the SpaceX API, which provides comprehensive information on launches, payloads, and mission outcomes. By analyzing historical launch data, the project aims to explore patterns and trends that have influenced SpaceX's impressive track record, using a data-driven approach to extract meaningful insights from a decade of space exploration.
- The core question driving this analysis is: What are the key factors that influence the success or failure of SpaceX missions? We aim to identify variables—such as rocket type, payload characteristics, launch location, and timing—that correlate with mission outcomes. By understanding these factors, we seek to provide data-backed insights into how SpaceX can further optimize its launches, improve reliability, and minimize risks. This analysis intends to highlight the variables that matter most, guiding future strategic decisions and supporting SpaceX's overarching mission of efficient and sustainable space exploration.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was retrieved using the SpaceX API, providing detailed information on rocket launches, payloads, and mission outcomes.
- Perform data wrangling
  - The data underwent cleaning and preprocessing to remove inconsistencies, handle missing values, and transform variables for accurate analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Predictive models were developed using machine learning techniques, with model performance optimized through hyperparameter tuning and evaluated using accuracy metrics.

Data Collection

- How data sets were collected.
  - Request and parse the SpaceX launch data using the GET request
  - Filter the dataframe to only include Falcon 9 launches
  - Dealing with Missing Values

Data Collection – SpaceX API

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

- GitHub URL: [applied-data-science-capstone-assignments/jupyter-labs-spacex-data-collection-api.ipynb at main · kaechen413/applied-data-science-capstone-assignments]

- Request and parse the SpaceX launch data using the GET request
- Filter the dataframe to only include Falcon 9 launches
- Dealing with Missing Values

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- GitHub URL: [applied-data-science-capstone-assignments/jupyter-labs-webscraping.ipynb at main · kaechen413/applied-data-science-capstone-assignments]

1. Request the Falcon9 Launch Wiki page from its URL

2. Extract all column/variable names from the HTML table header

3. Create a data frame by parsing the launch HTML tables

Data Wrangling

- Describe how data were processed
  - Load Space X dataset
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurence of mission outcome of the orbits
  - Create a landing outcome label from Outcome column

- GitHub URL:applied-data-science-capstone-assignments/labs-jupyter-spacex-Data wrangling.ipynb at main · kaechen413/applied-data-science-capstone-assignments

EDA with Data Visualization

- Scatter point charts and bar charts were used to visualize the replationship among Flight number and Launch Site, Payload Mass, Orbit type and Launch Site, through which we could see the relationship very clearly
- GitHub URL: [applied-data-science-capstone-assignments/edadataviz.ipynb at main · kaechen413/applied-data-science-capstone-assignments]()

EDA with SQL

- the SQL queries I performed
  - **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 achieved**
  - **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 landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order**
- GitHub URL: [applied-data-science-capstone-assignments/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · kaechen413/applied-data-science-capstone-assignments]()

Build an Interactive Map with Folium

- The map objects such as markers, circles, lines, etc. I created and added to a folium map
  - Circles
  - Markers
  - Popups
- GitHub URL: applied-data-science-capstone-assignments/lab_jupyter_launch_site_location.ipynb at main · kaechen413/applied-data-science-capstone-assignments

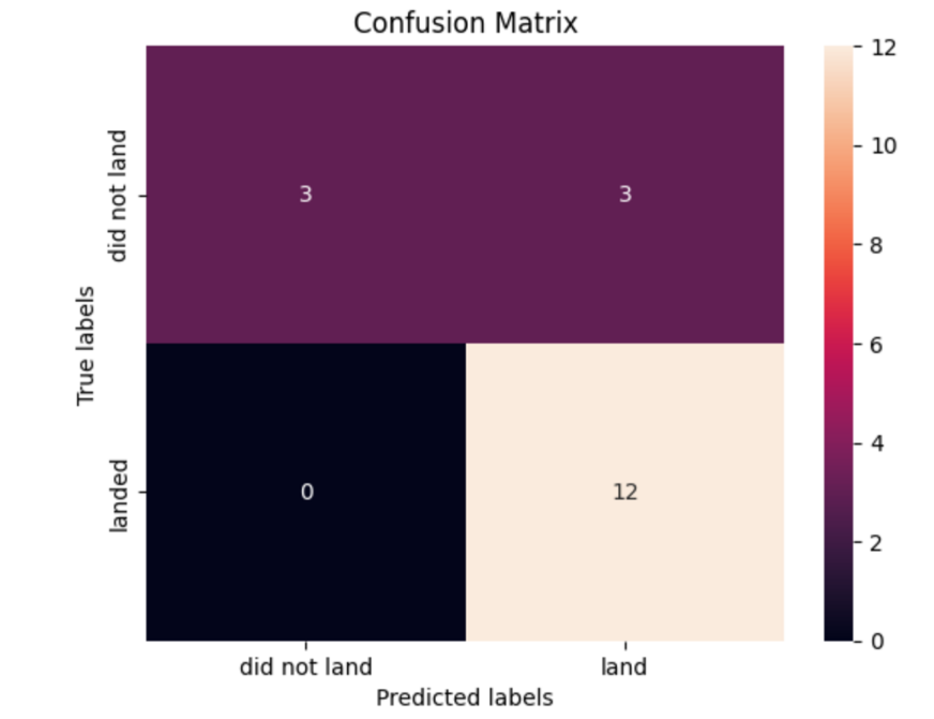
Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- GitHub URL: applied-data-science-capstone-assignments/lab_jupyter_launch_site_location.ipynb at main · kaechen413/applied-data-science-capstone-assignments

Predictive Analysis (Classification)

- How I built, evaluated, improved, and found the best performing classification model
  - Create a NumPy array from the column Class in data
  - Standardize the data
  - Use the function train_test_split to split the data X and Y into training and test data.
  - Create a logistic regression
  - Calculate the accuracy on the test data
  - Check the confusion matrix

- GitHub URL: [applied-data-science-capstone-assignments/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · kaechen413/applied-data-science-capstone-assignments]()



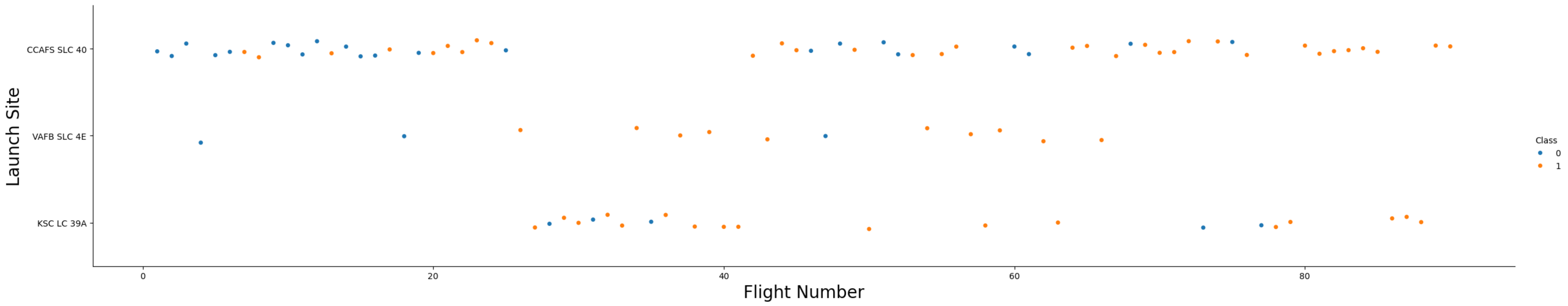
Results

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

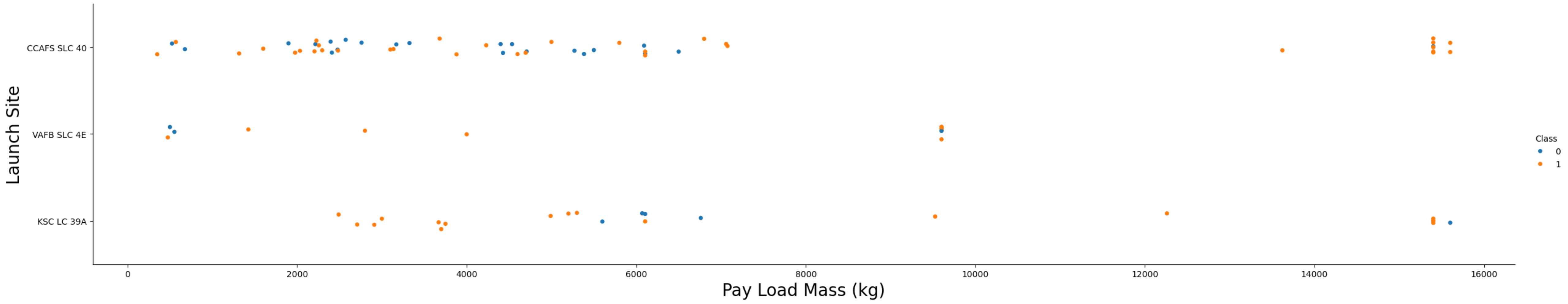
Section 2

Insights drawn from EDA

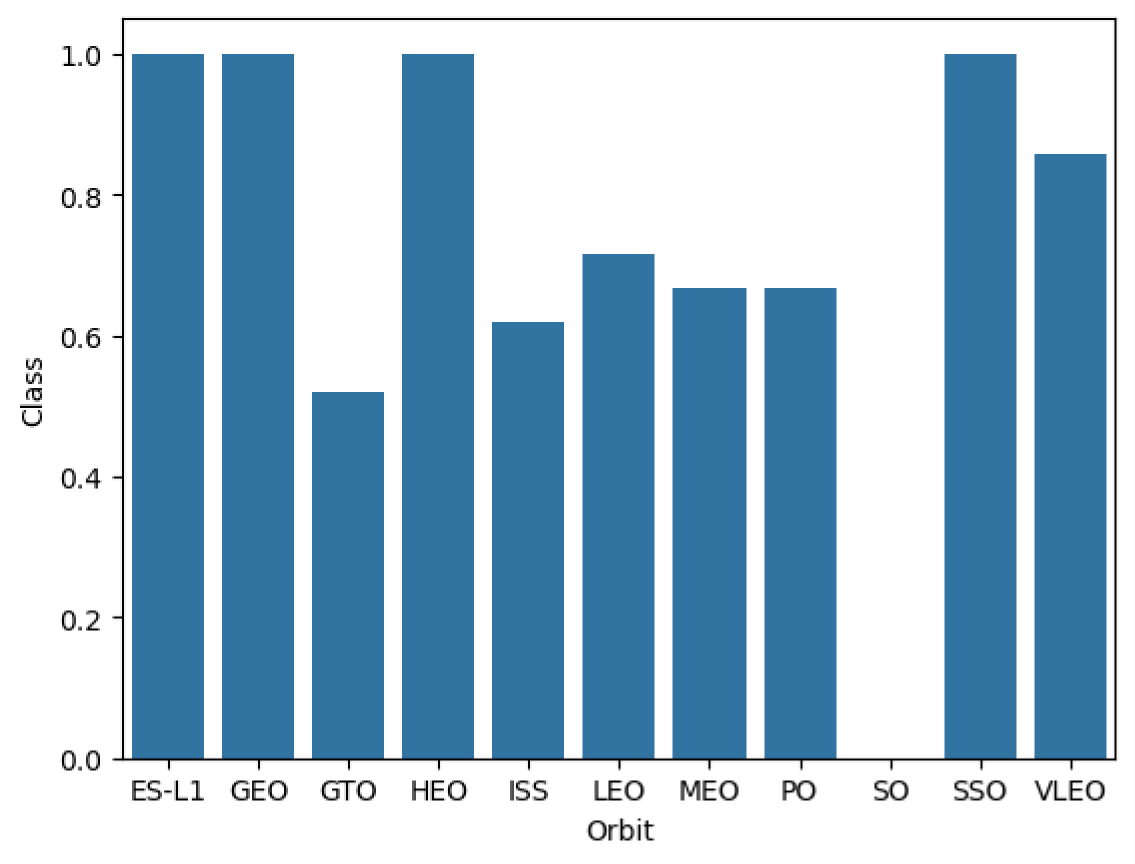
Flight Number vs. Launch Site



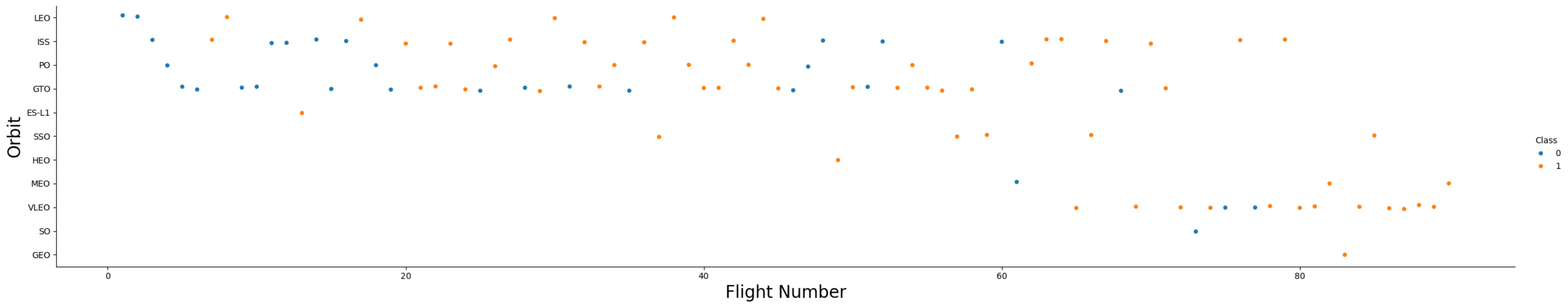
Payload vs. Launch Site



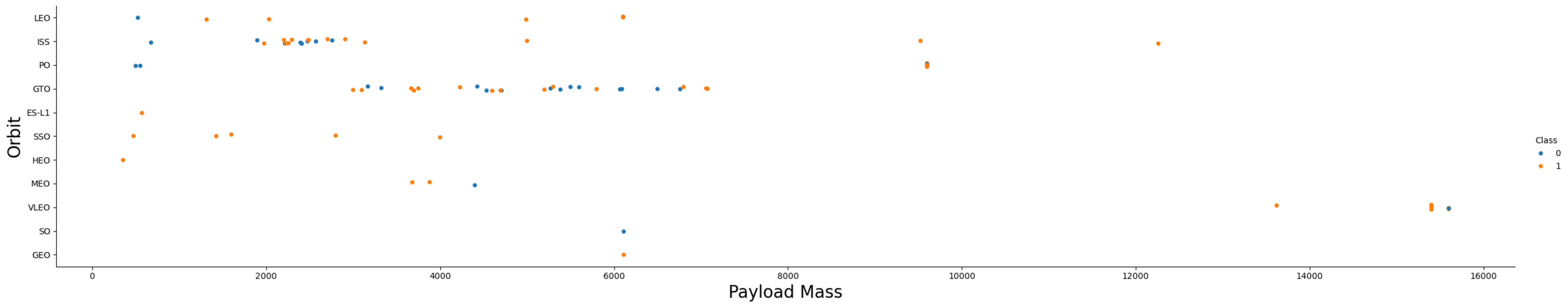
Success Rate vs. Orbit Type



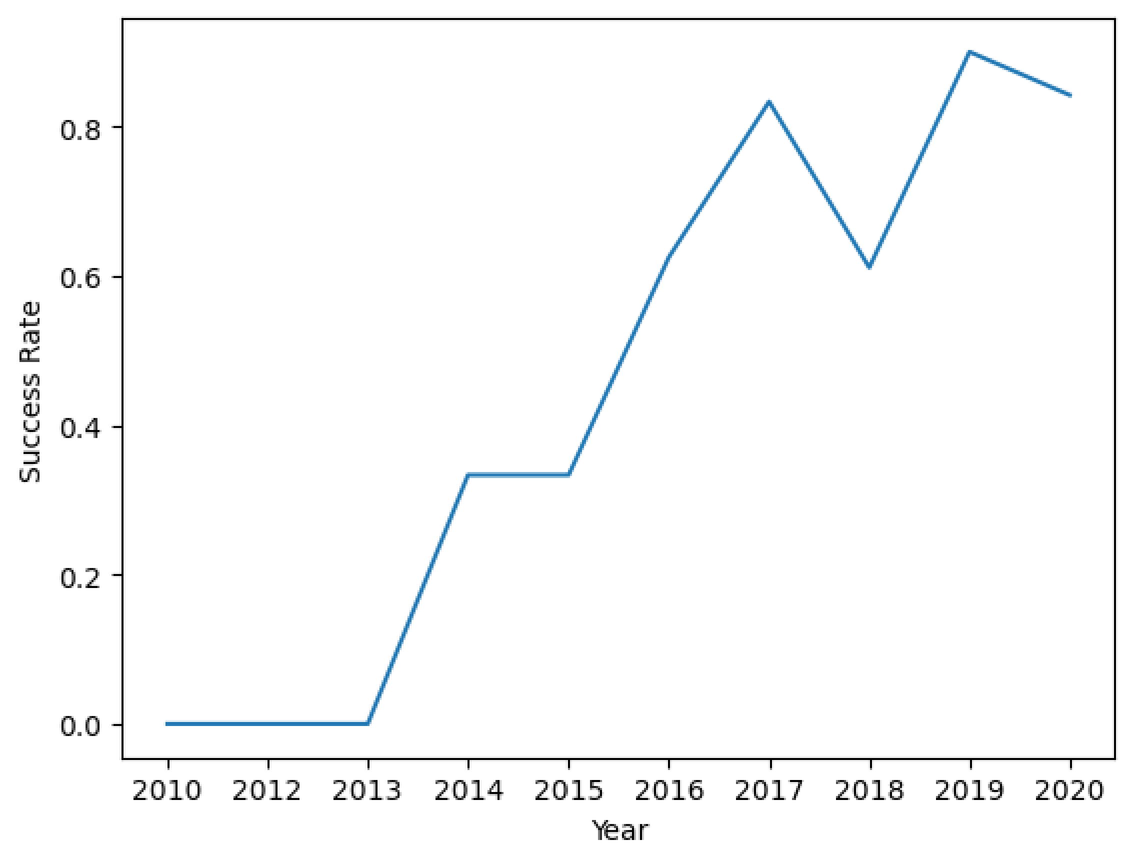
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

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

Total Payload Mass

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 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

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

Boosters Carried Maximum Payload

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

substr(Date, 6,2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Landing_Outcome	count(Landing_Outcome)
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

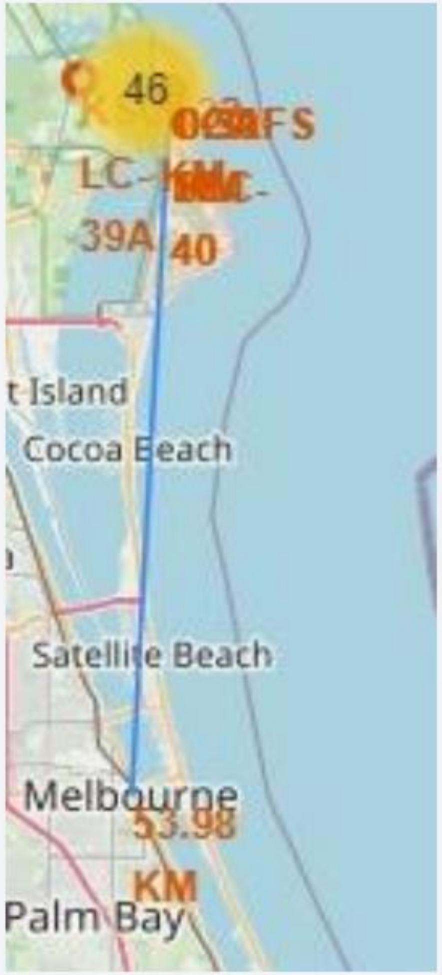
Launch Sites Proximities Analysis

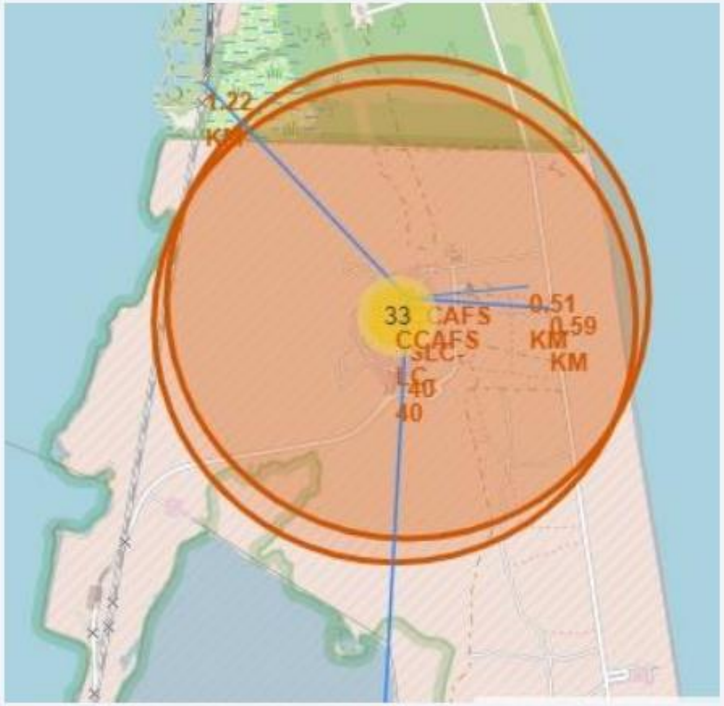
<Folium Map Screenshot 1>

<Folium Map Screenshot 2>

- Green show successful launches and Red show Failed launches
- From the color-labeled markers we can easily identify which sites have relatively high success rates.

<Folium Map Screenshot 3>





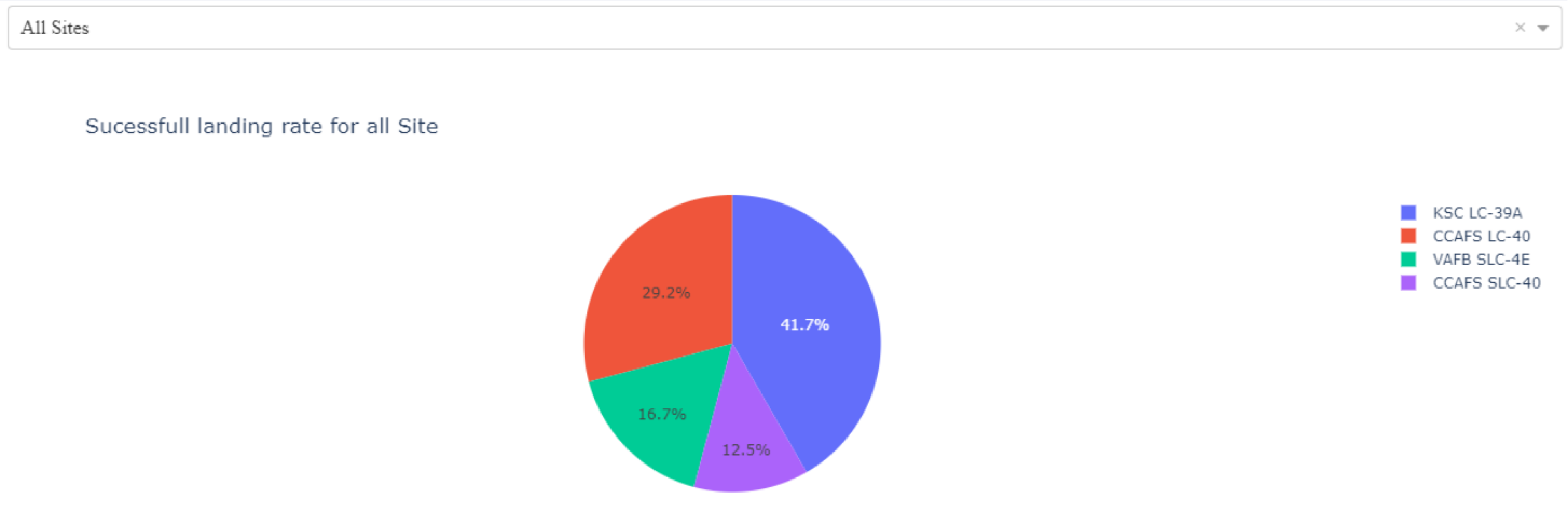
- We choose the SLC40 site to calculate the distances
- From the resulted line distances on the map, we could observe that:

- Launch sites are in close proximity to railways, highways and coastline to facilitate in transportation and logistics
- Launch sites keep a certain distance away from cities to ensure public safety.

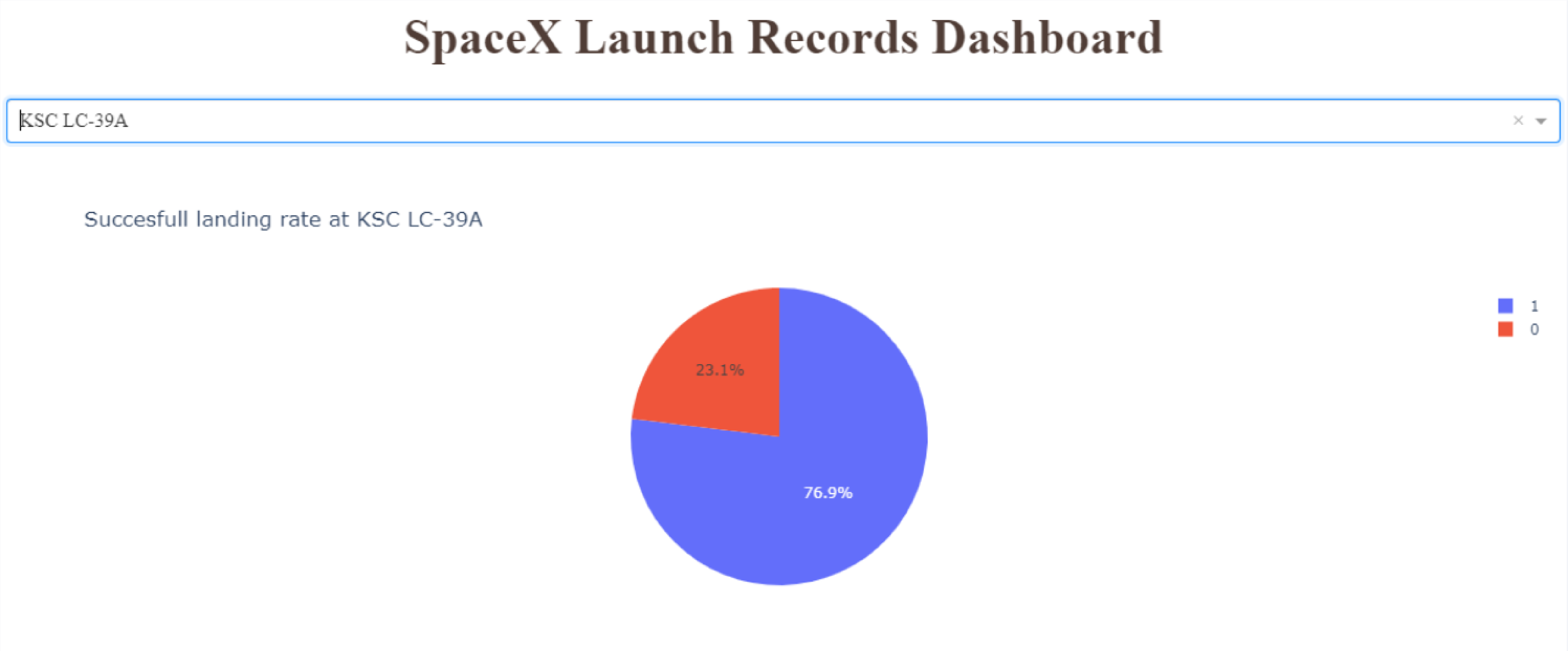
Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>



<Dashboard Screenshot 2>



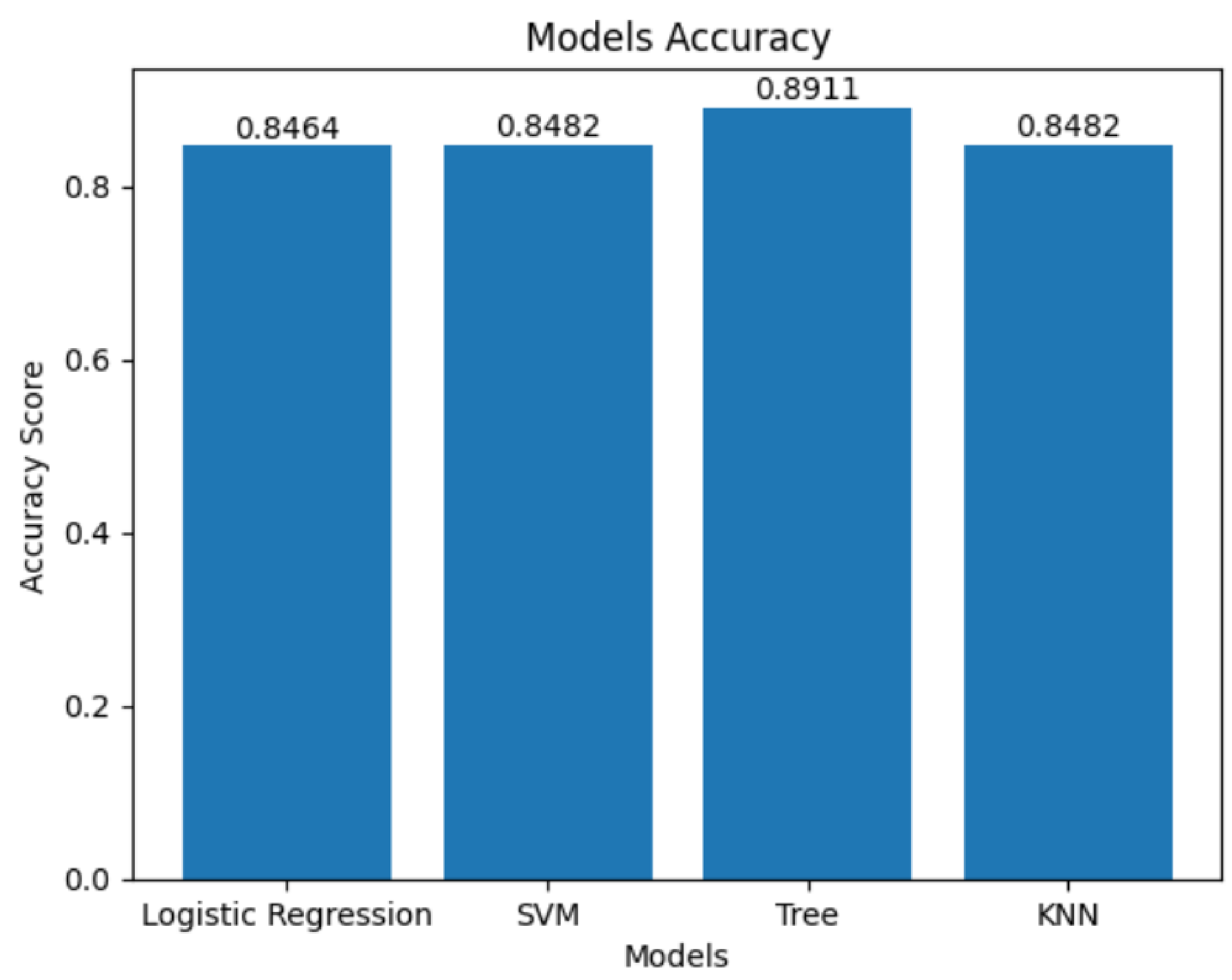
<Dashboard Screenshot 3>

- 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, such as which payload range or booster version have the largest success rate, etc.

Section 5

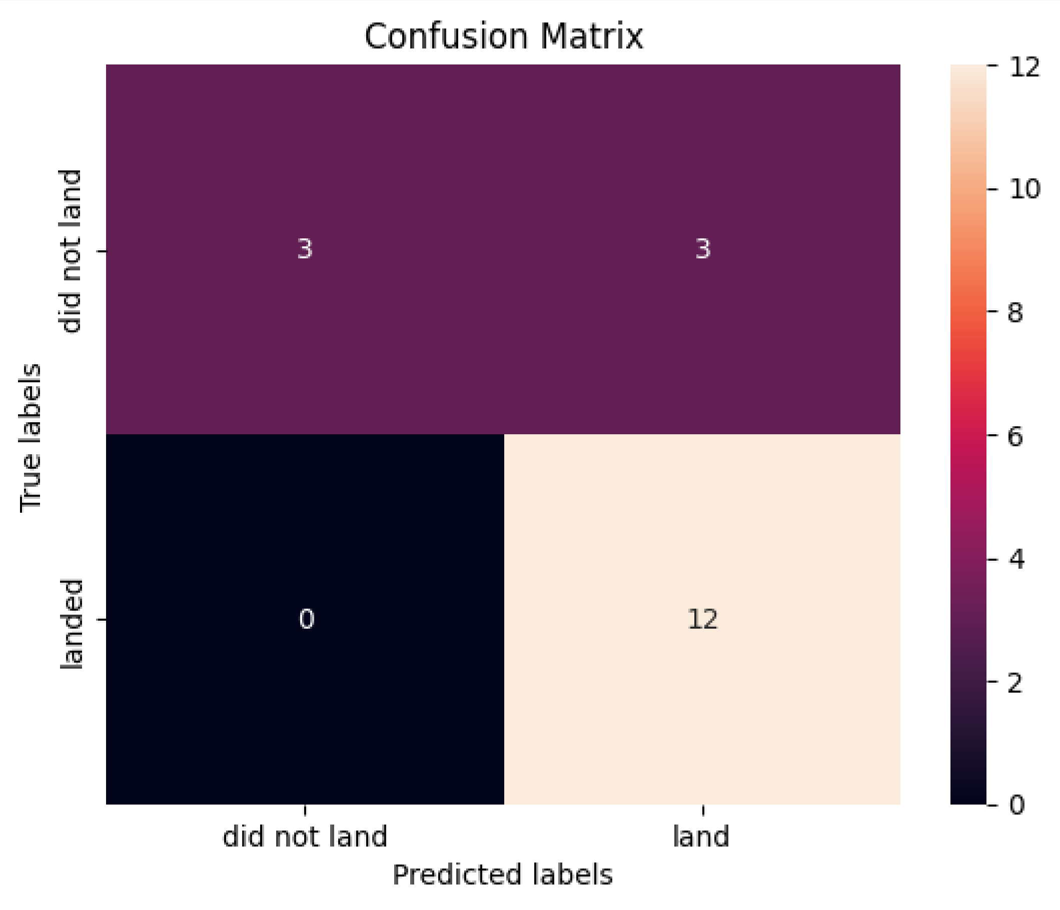
Predictive Analysis (Classification)

Classification Accuracy



As observed from the bar chart, the highest accuracy score is belongs to the **Tree model**.

Confusion Matrix



Conclusions

From the above analyses, we can conclude that:

• The success rate of launches has shown a steady increase over time as the number of flight grows. Notably, it has consistently risen from 2013 to 2020.

• The orbit that exhibit the highest success rates are ES-L1, GWO, HEO and SS0

• For heavy payload masses, the success rates are higher in PO, ISS and LEO orbits. In contrast, ES-L1, SSO, HEO and MEO orbits demonstrate higher success rates for low payload masses.

• Launches with low payload masses (between 0 and 4,000 kg) have a significantly higher success rate compared to those with high payload masses (between 4,000 and 10,000 kg) • All launch sites are located above the equator line and in close proximity to coastlines. This position facilitates transportation and logistics through near by railways, highways, and coastal access. Additionally, these sites are strategically distanced from urban areas to ensure public safety.

• KSC LC-39 stands out as the site with the highest number of successful launches.

• The Tree classifier model achieves the highest accuracy score among the model evaluated.

Thank you!