IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Madhava Krishnan N S
18/02/2024

Outline

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

Executive Summary

- Utilized SpaceX API and web scraping for data collection, followed by extensive wrangling and outcome labeling. Explored data with visualization and SQL, and performed predictive analysis using classification models, building, tuning, and evaluating each.

Key Results:

- Launch Site Performance: 'CCAFS SLC 40' struggled initially; 'VAFB SLC 4E' and 'KSC LC 39A' consistently outperformed.
- Orbit Success Rates: 'ES-L1', 'GEO', 'HEO', and 'SSO' achieved 100% success; 'GTO', 'ISS', 'LEO', 'MEO', 'PO', 'VLEO' >50%; 'SO' none.
- Payload and Landing Insights: Successful landings in Polar, LEO, ISS; GTO struggles distinguishing positive/negative landings.

Executive Summary

- Temporal Trends: Success rate steadily increased from 2013 to 2020.
- Payload Statistics: Total payload mass: 45,596 kg; Average: 2,534.67 kg.
- Geographical & Strategic Considerations: Florida sites near Equator maximize efficiency; coastal launches enhance safety. Launch sites strategically located for optimal logistics.
- Launch Site Comparison: KSC LC-39A: highest launches and success rate; CCAFS LC-40: lowest successful launches.
- Payload and Booster Analysis: Payloads 2k-4k kg have highest success; Booster FT excels, v1.1 lags.
- Model Performance: Logistic Regression, SVM, Decision Tree, and KNN achieved 0.83 accuracy.
- Model Evaluation: Consistent confusion matrix; primary challenge: false positives.

Introduction

- In this project, we delve into SpaceX's Falcon 9 rocket launches, celebrated for their cost efficiency at 62 million dollars—markedly below competitors charging upwards of 165 million dollars.
- The primary challenge addressed is predicting the landing success of the first stage, a critical factor influencing launch costs.
- This predictive capability is not only instrumental for optimizing SpaceX's operations but also provides potential competitors with valuable insights for informed decision-making when bidding against SpaceX for rocket launches.

Section 1

Methodology

Methodology

Executive Summary

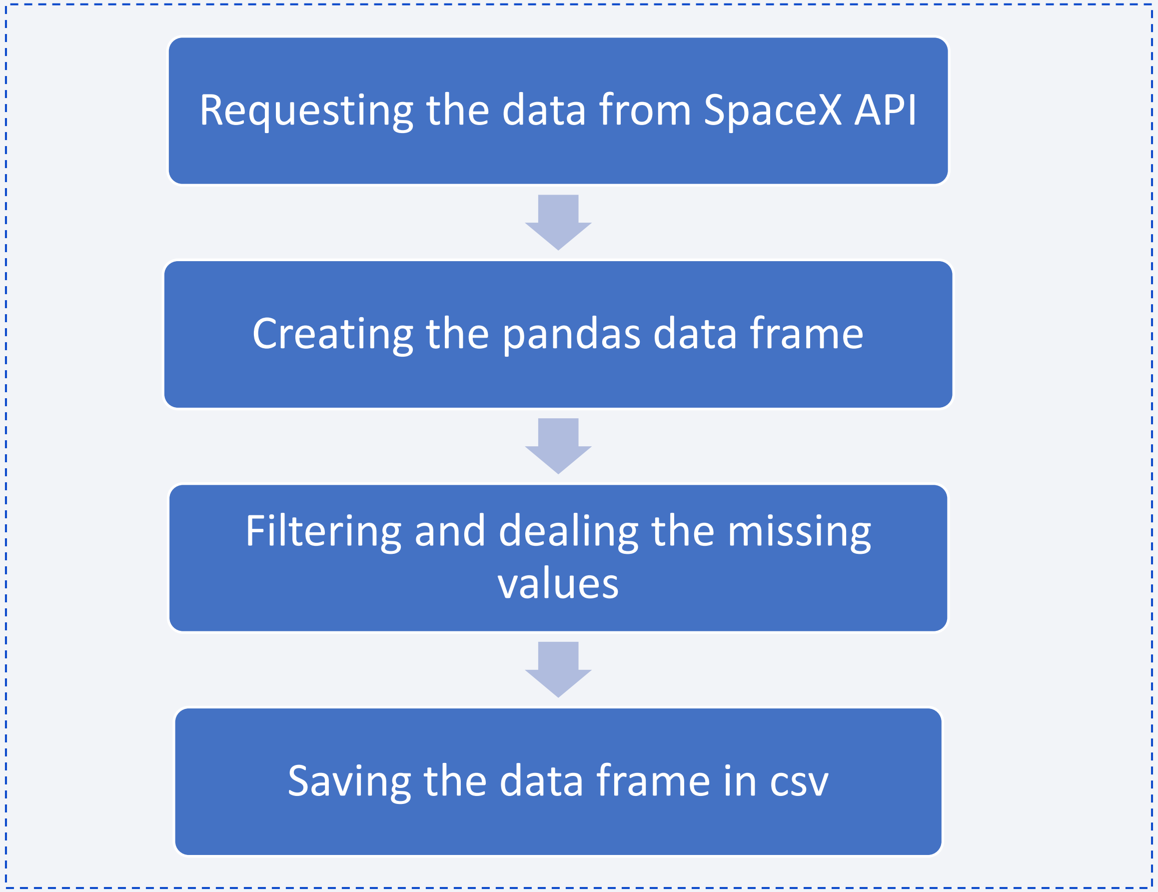
- Data collection methodology:
  - SpaceX API and web scrapping the Wikipedia.
- Perform data wrangling
  - Labelling the outcome
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building, tuning and evaluation of models.

Data Collection

- The data were collected using the SpaceX API and by scraping the Wikipedia page through the use of the 'requests' and 'BeautifulSoup' modules in Python.
- The data collection process is illustrated using key phrases and flowcharts in the following slide.

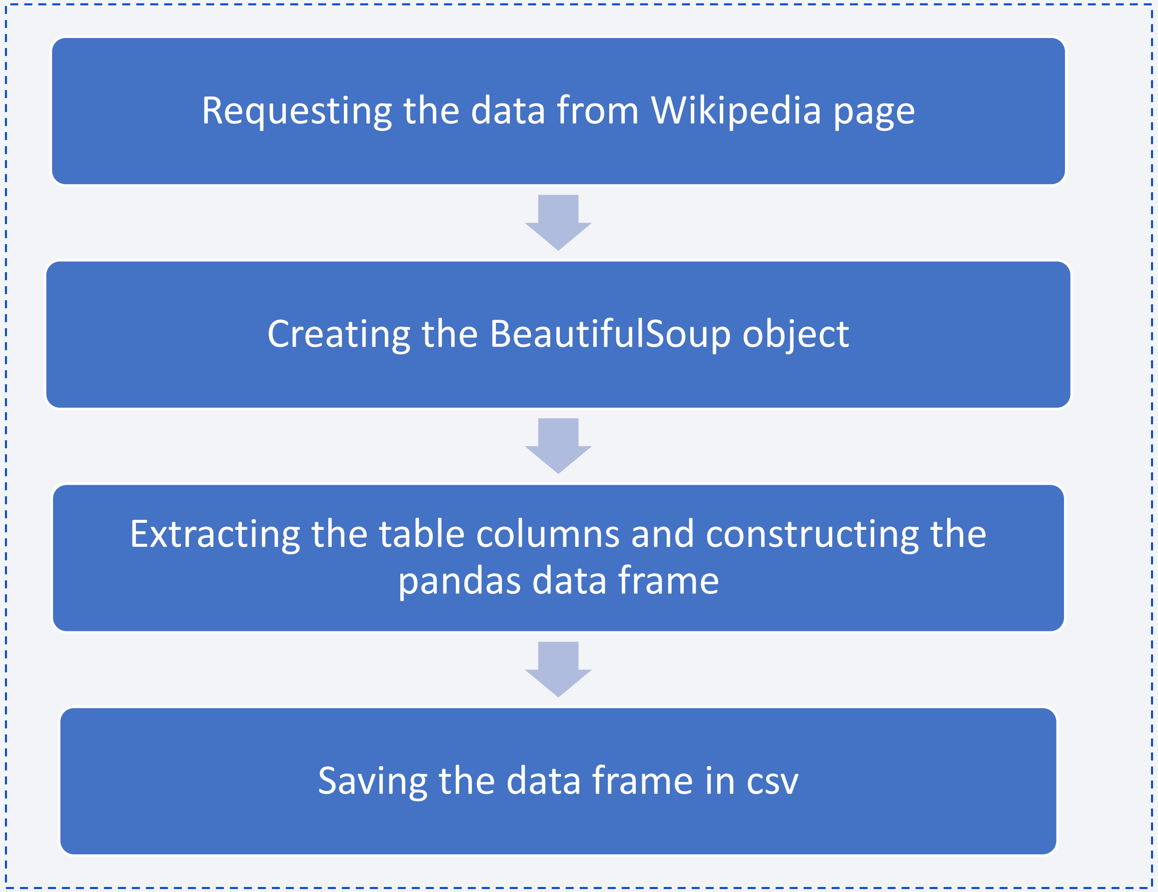
Data Collection – SpaceX API

- The necessary Falcon 9 rocket launch data was obtained using the SpaceX API and subsequently saved in cleaned form in CSV file format. This process is illustrated in the accompanying flow chart.
- The GitHub URL of the completed SpaceX API calls notebook: https://github.com/Madhavananalyst/Applied-Data-Science-Capstone-project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



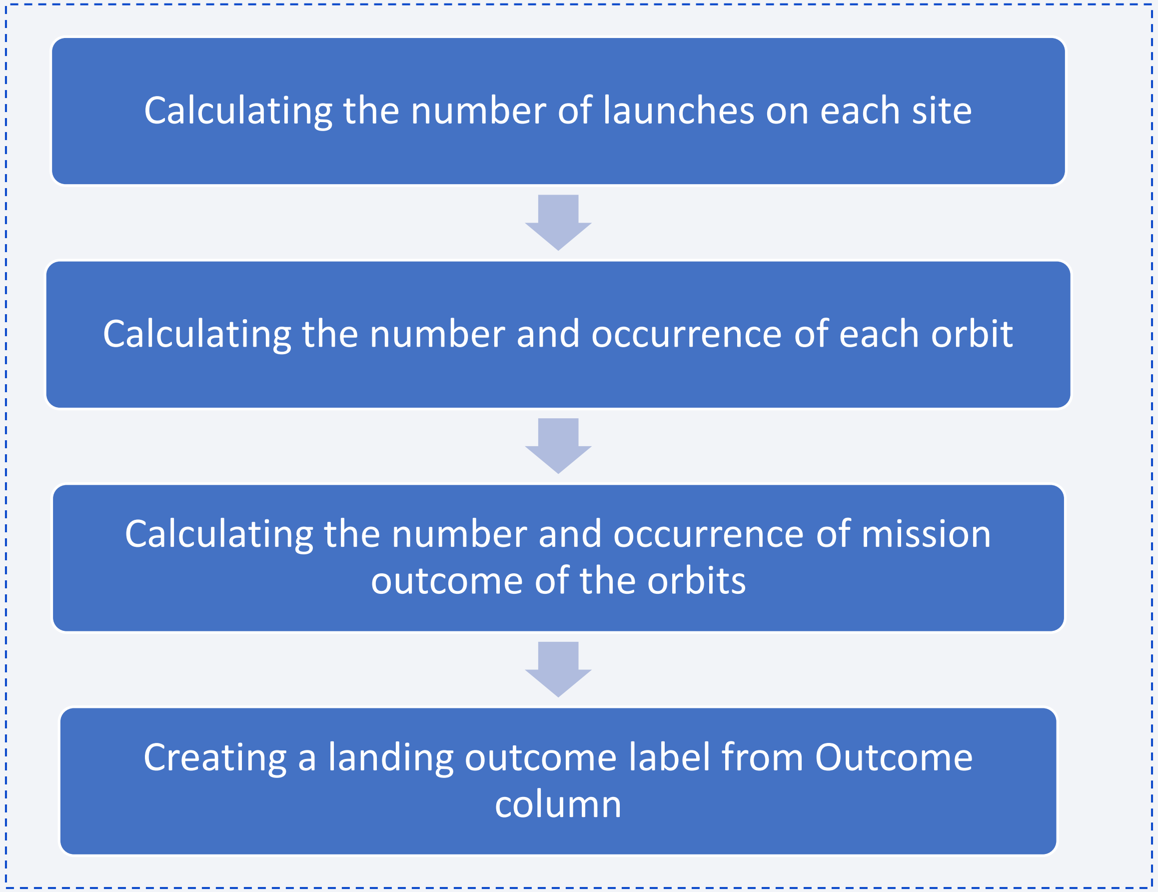
Data Collection - Scraping

- The necessary Falcon 9 rocket launch data was acquired by web scraping the Wikipedia page, and the cleaned data was then saved in CSV file format. This process is illustrated in the accompanying flowchart.
- The GitHub URL of the completed Scraping the web page notebook: https://github.com/Madhavananalyst/Applied-Data-Science-Capstone-project/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Performed some Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models. This process in illustrated in the flowchart.
- The GitHub URL of the completed Data wrangling notebook: https://github.com/Madhavananalyst/Applied-Data-Science-Capstone-project/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- The following charts were plotted:
  - Catplot between Flight Number and Launch Site, Payload and Launch Site, Flight Number and Orbit type and Payload and Orbit type.
  - Bar chart of success rate of each orbit type.
  - Trend line of launch success yearly trend.
- The charts were used to understand the relationship between different variables, success rate of each orbit and the yearly trend.
- The GitHub URL of completed EDA with data visualization notebook: https://github.com/Madhavananalyst/Applied-Data-Science-Capstone-project/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- The following SQL queries were 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 successful 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.
- The GitHub URL of completed EDA with SQL notebook: https://github.com/Madhavananalyst/Applied-Data-Science-Capstone-project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

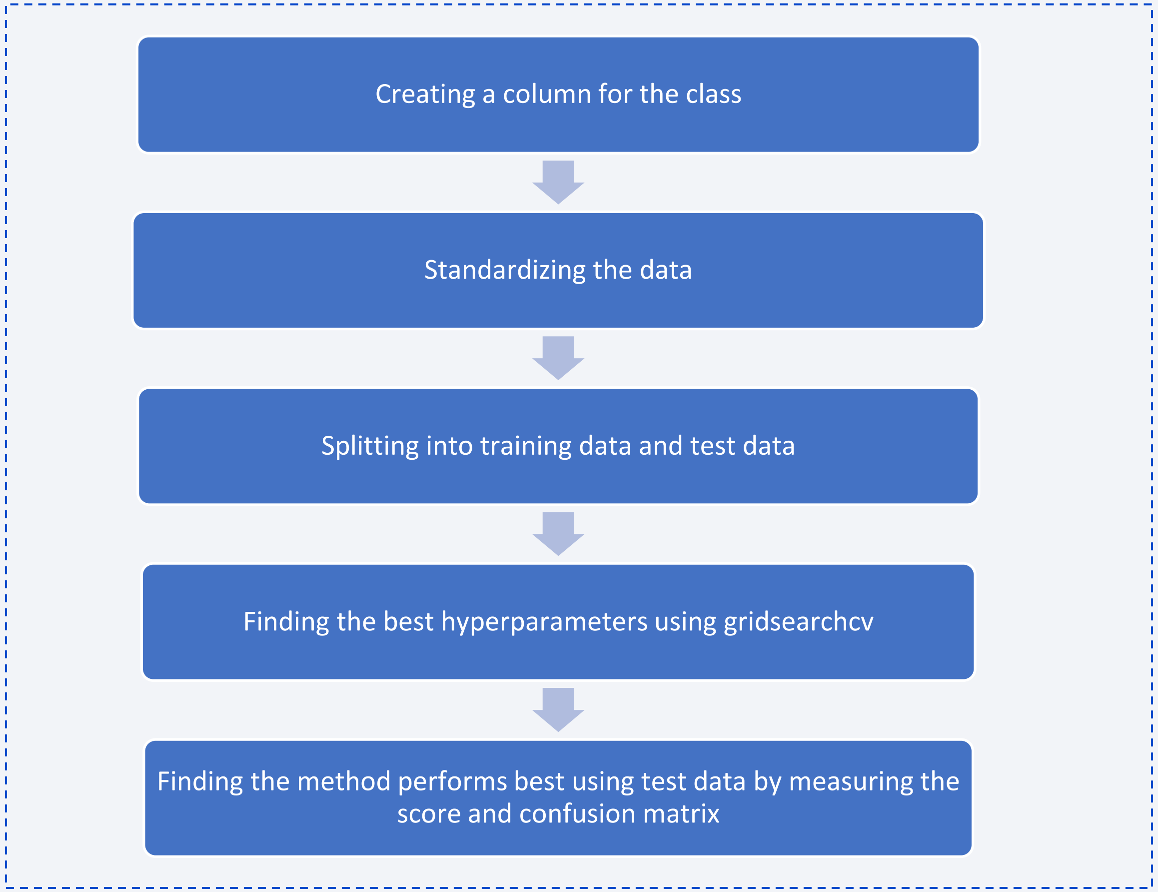
- Added the circle marker with label for NASA Johnson Space Center and launch sites. Added marker cluster for landing outcomes such as green marker represents the success and red represents the failure. Then, added mouse position and polylines that show the distance between launch sites to closest coastline, railway, highway and city.
- These were added to understand how the location of launch site affects the launching outcome.
- The GitHub URL of completed interactive map with Folium map: https://github.com/Madhavananalyst/Applied-Data-Science-Capstone-project/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Added the dropdown menu to select the launch sites and slider to select the payload range.
- Added the pie chart that shows the percentage of success rate of launch sites or the success and failure percentage of selected launch site. And added the scatter plot that shows the correlation between payload and launch outcomes.
- These were added to understand how the launch site and payload affects the launch outcomes.
- The GitHub URL of completed Plotly Dash lab: https://github.com/Madhavananalyst/Applied-Data-Science-Capstone-project/blob/main/dash.py

Predictive Analysis (Classification)

- The Logistic regression, SVM, Decision tree and KNN models were build. This process in illustrated in the flowchart.
- The GitHub URL of the completed predictive analysis (Classification) notebook: https://github.com/Madhavananalyst/Applied-Data-Science-Capstone-project/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb



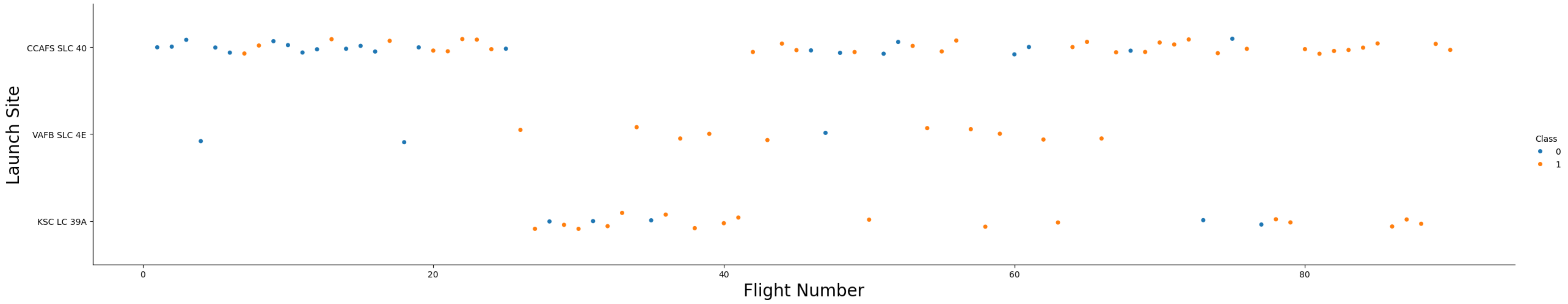
Results

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

Section 2

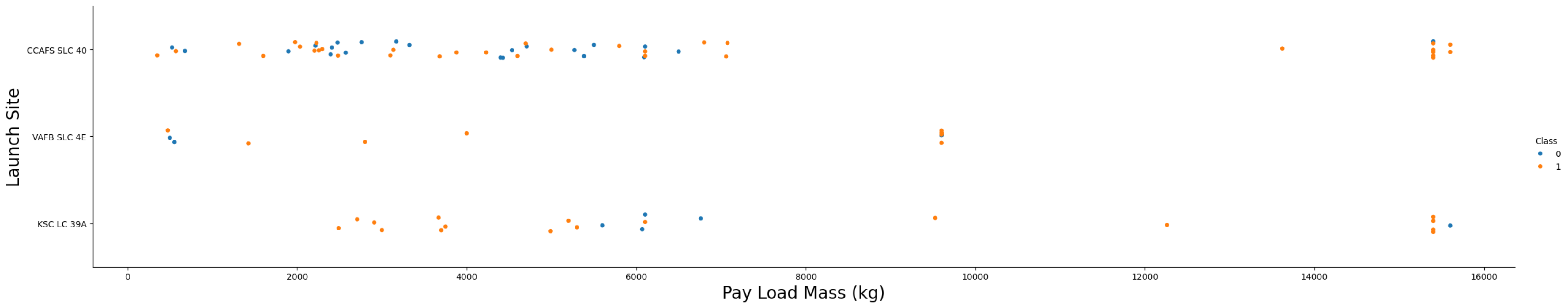
Insights drawn from EDA

Flight Number vs. Launch Site



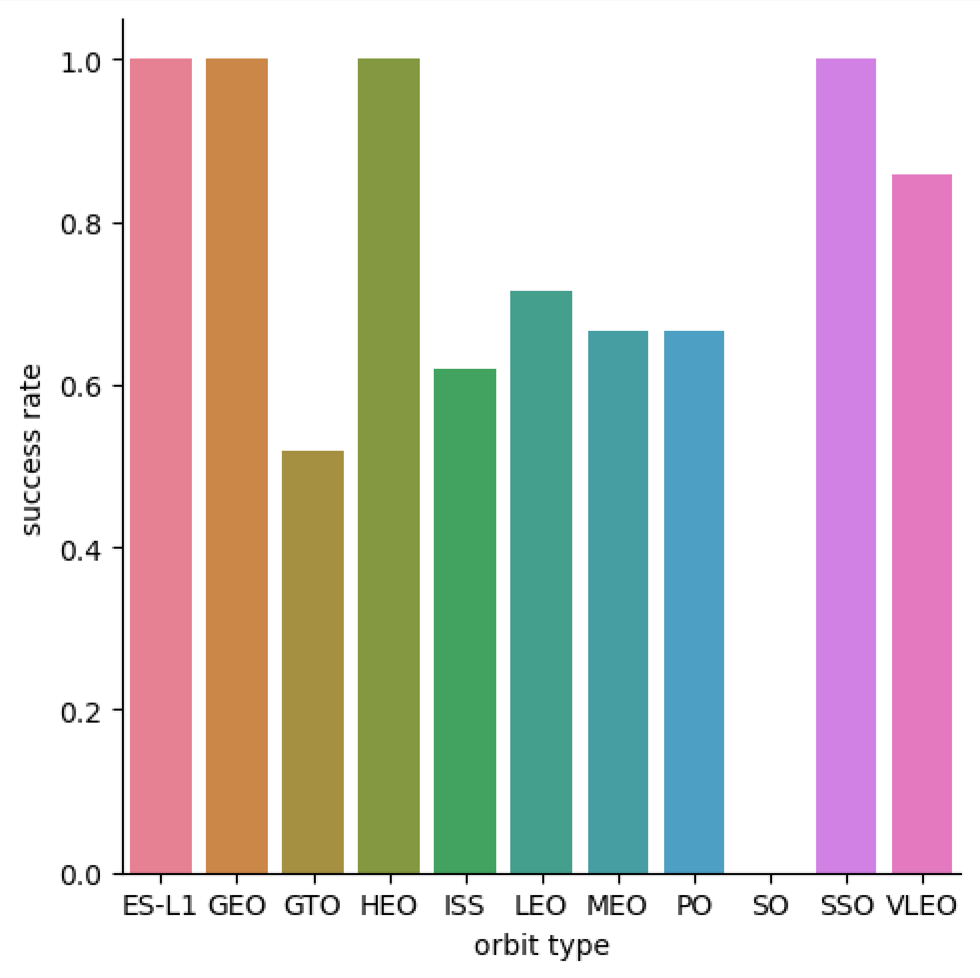
- The launch site 'CCAFS SLC 40' was frequently used, and its success rate was initially very low. In contrast, the success rates of the launch sites 'VAFB SLC 4E' and 'KSC LC 39A' were higher than that of 'CCAFS SLC 40'.
- However, in recent years, the success rates for all launch sites have significantly improved.

Payload vs. Launch Site



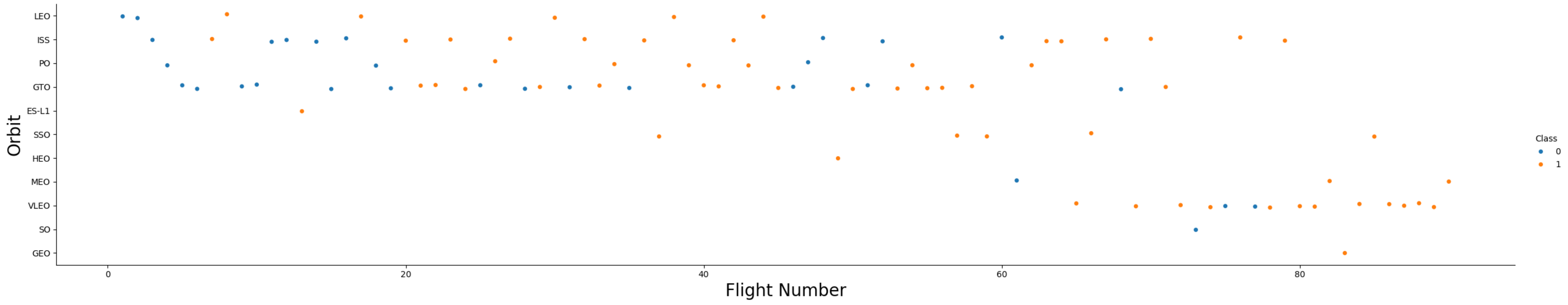
- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type



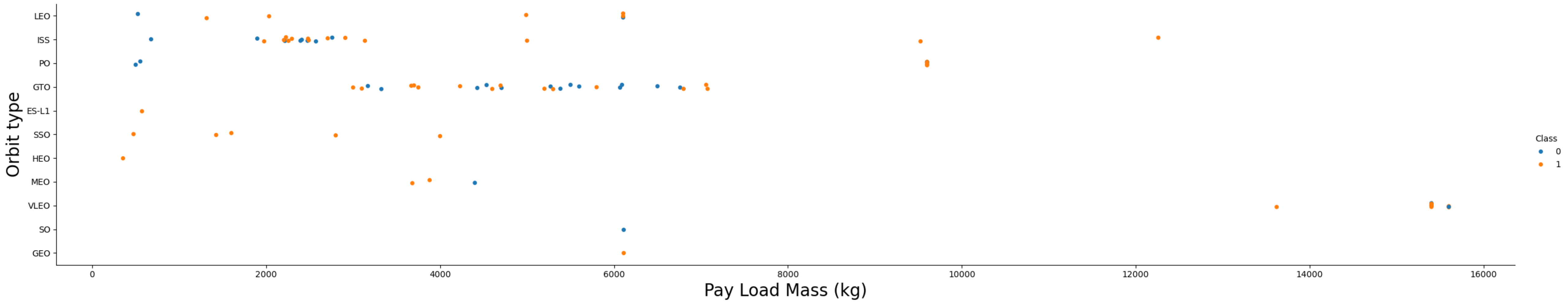
- The orbits 'ES-L1', 'GEO', 'HEO', and 'SSO' have a success rate of 100 percentage.
- 'GTO', 'ISS', 'LEO', 'MEO','PO' and 'VLEO' have a success rate of more than 50 percent.
- On the other hand, 'SO' has no success rate.

Flight Number vs. Orbit Type



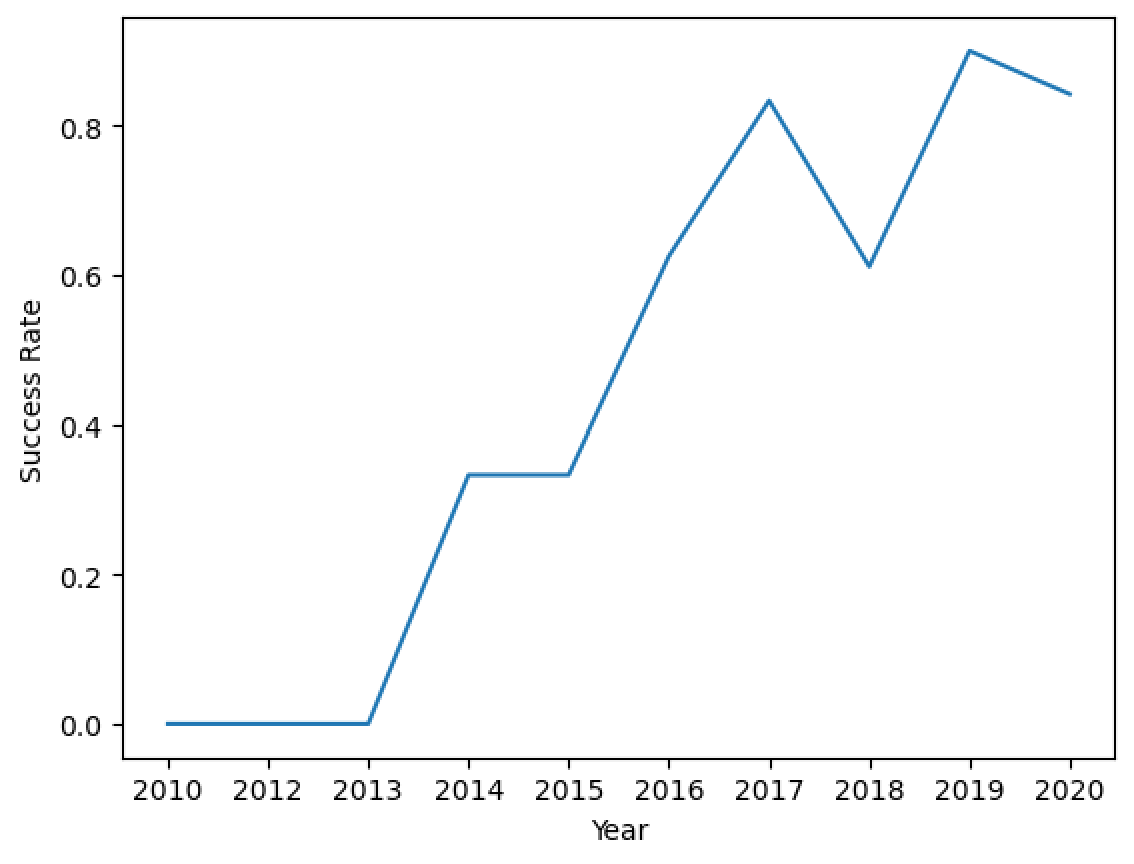
- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



- We can observe that the success rate since 2013 kept increasing till 2020.

All Launch Site Names

```
%sql select distinct(Launch_site) from SPACEXTABLE
```

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

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
```

```
 * 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

Total Payload Mass

```
%sql select sum(PAYLOAD_MASS__KG_) as total_payload_mass from spacextable where customer = "NASA (CRS)"
```

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

total_payload_mass
45596

in kg

Average Payload Mass by F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) as average_payload_mass from spacextable where Booster_Version = "F9 v1.1"
```

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

average_payload_mass in kg
2928.4

```
%sql select avg(PAYLOAD_MASS__KG_) as average_payload_mass from spacextable where Booster_Version like "%F9 v1.1%"
```

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

average_payload_mass in kg
2534.6666666666665

First Successful Ground Landing Date

```
%sql select min(Date) from spacextable where Landing_outcome = "Success (ground pad)"
```

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql select booster_version from spacextable where Landing_outcome = "Success (drone ship)" and PAYLOAD_MASS__KG_ between 4000 ar
```

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

```
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

```
%sql select mission_outcome, count(*) from spacextable group by mission_outcome
```

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

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

Boosters Carried Maximum Payload

```
%sql select booster_version from spacextable where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spacextable)
```

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

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 substr(Date,6,2) as month, landing_outcome, booster_version, launch_site
from spacextable
where substr(Date,0,5)='2015'
and Landing_Outcome = 'Failure (drone ship)'
```

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

month	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

```
%%sql
select landing_outcome, count(landing_outcome) as number_of_times from spacextable
where date between "2010-06-04" and "2017-03-20"
group by landing_outcome
order by number_of_times desc
```

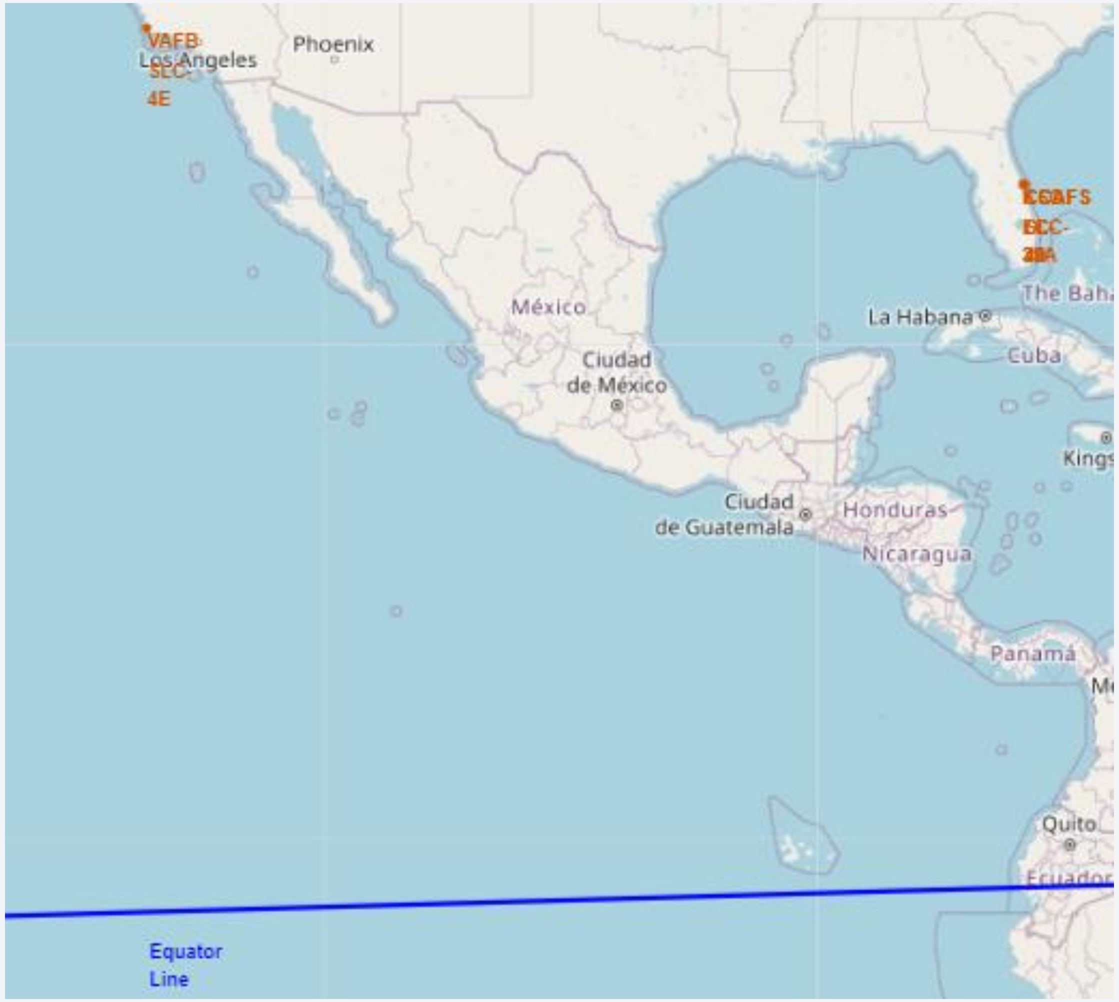
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	number_of_times
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

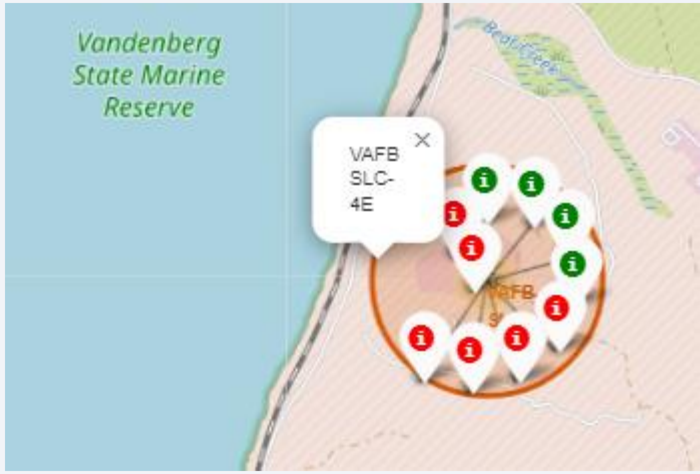
Launch sites' location

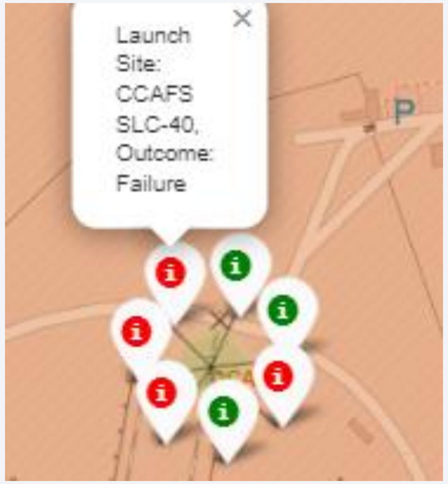


Launch sites are marked in red circles.

- The Florida launch sites (CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A) are closer to the Equator than VAFB SLC-4E, which is located in California and farther away.
- All launch sites are in proximity to a coast.
- Launching near the Equator maximizes the Earth's rotational speed, providing a natural boost for rockets. This efficiency is crucial for payload capacity, cost-effectiveness, and mission success, making it an important consideration for space agencies, satellite operators, and commercial client.
- Coastal launch sites offer overwater trajectories, enhancing safety by minimizing populated areas overflown during launches.

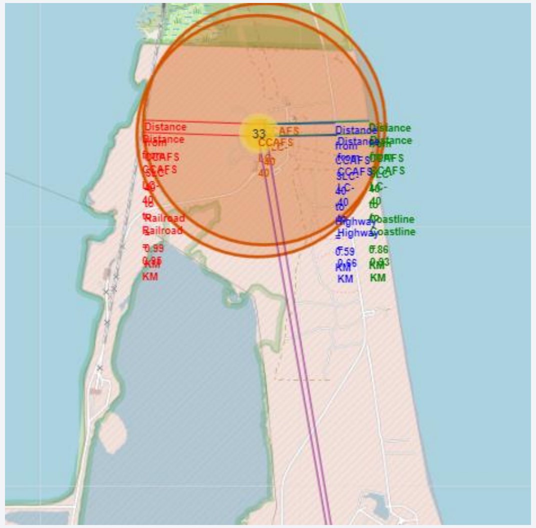
Launching outcomes of launch sites

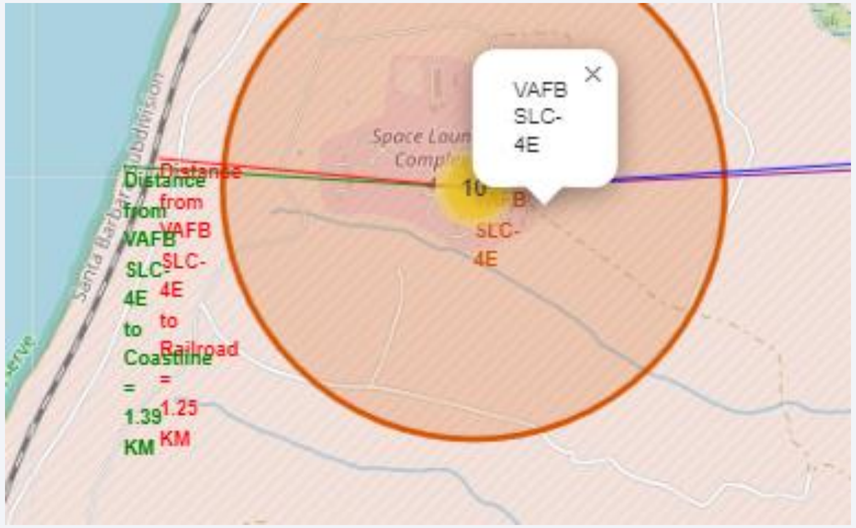




- Green markers represents the success and red markers represents failure.
- Launch site KSC LC-39A has very high success rate whereas Launch site CCAFS LC-40 has very low success rate.

Distance from launch sites to its proximities





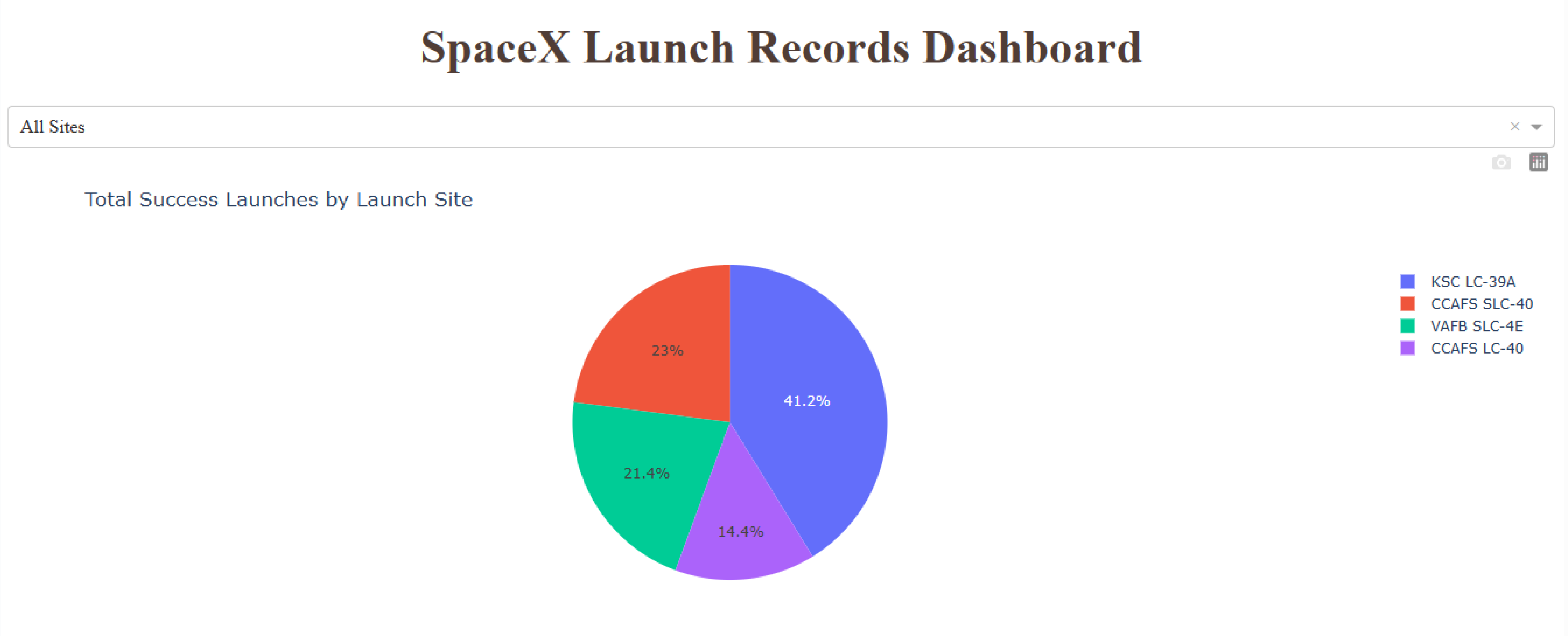
Launch Site	Coastline Distance in km	Highway Distance in km	Railroad Distance in km	City Distance in km
CCAFS LC-40	0.928084	0.655233	0.952528	8.254539
CCAFS SLC-40	0.862105	0.588128	0.986247	8.343822
KSC LC-39A	3.283125	0.838578	0.690150	12.450739
VAFB SLC-4E	1.392655	14.015603	1.254472	14.009983

- The location of launch sites appears to be strategically chosen to be near essential transportation routes (railways and highways), close to the coastline for safe launch trajectories, and at a safe distance from populated areas.
- These factors are likely considered to optimize logistics, safety, and operational efficiency.

Section 4

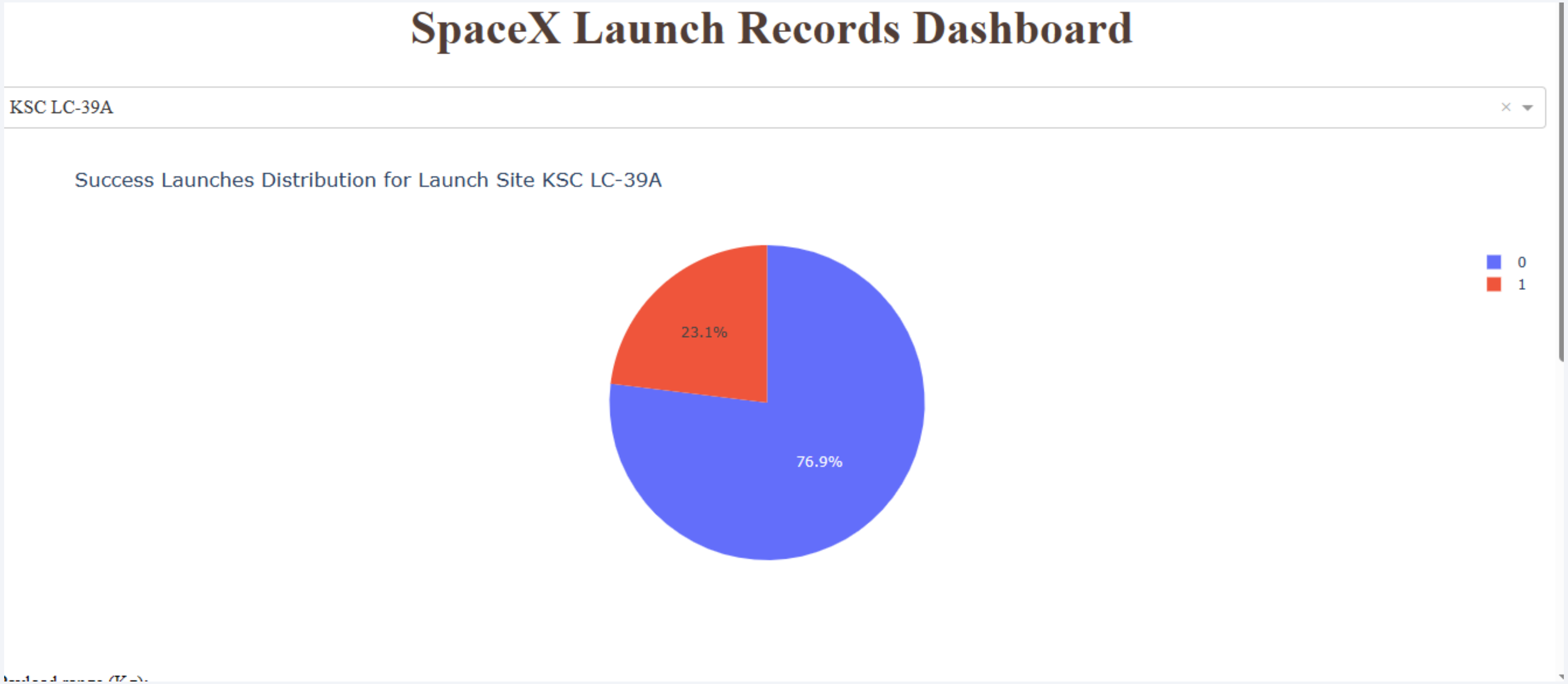
Build a Dashboard with Plotly Dash

Total success launches by launch site



- The launch site KSC LC-39A has the highest number of successful launches.
- The launch site CCAFS LC-40 has the lowest number of successful launches.

KSC LC – 39A : The launch site with the highest success rate



- The launch site KSC LC-39A has the highest success rate of 76.9 percentage.

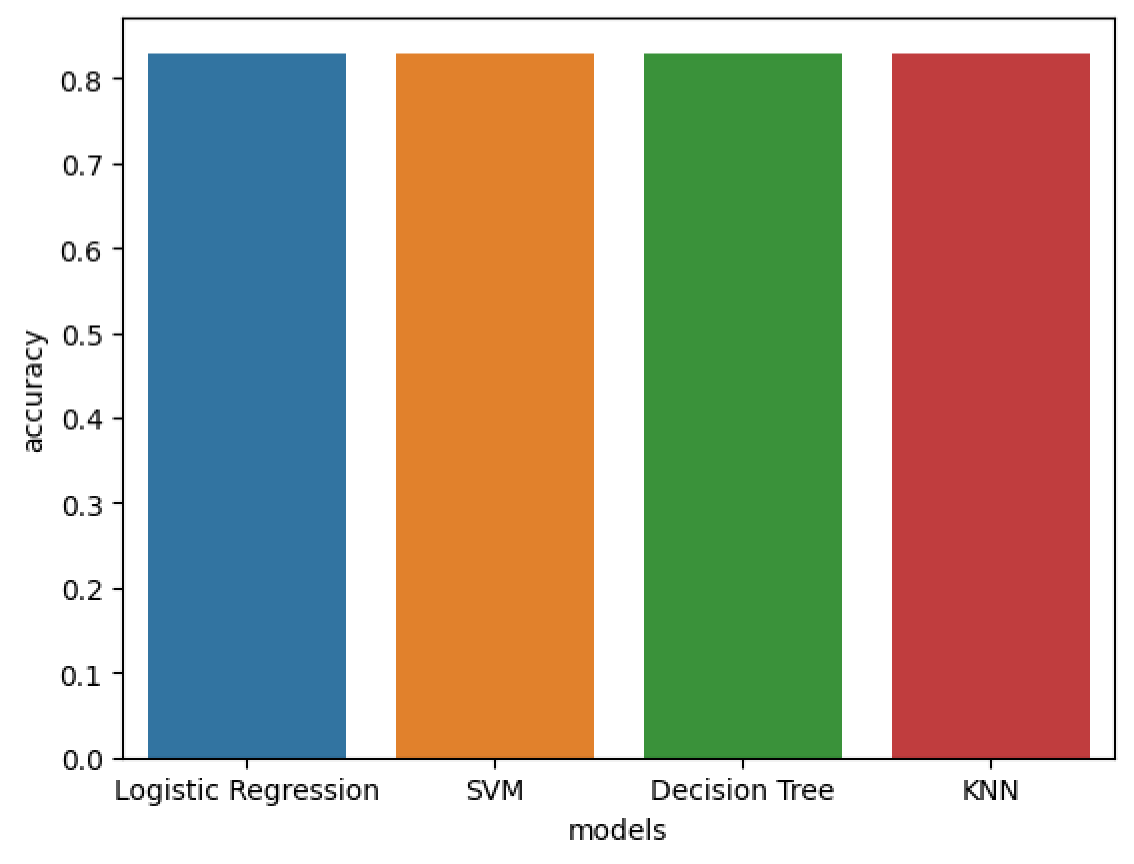
Payload vs. Launch Outcomes for all sites

- The payload range between 2k and 4k kg has the highest success rate, whereas payload masses greater than 6k have almost no success rate at all.
- The booster version FT has the highest success rate, whereas v1.1 has the lowest success rate.

Section 5

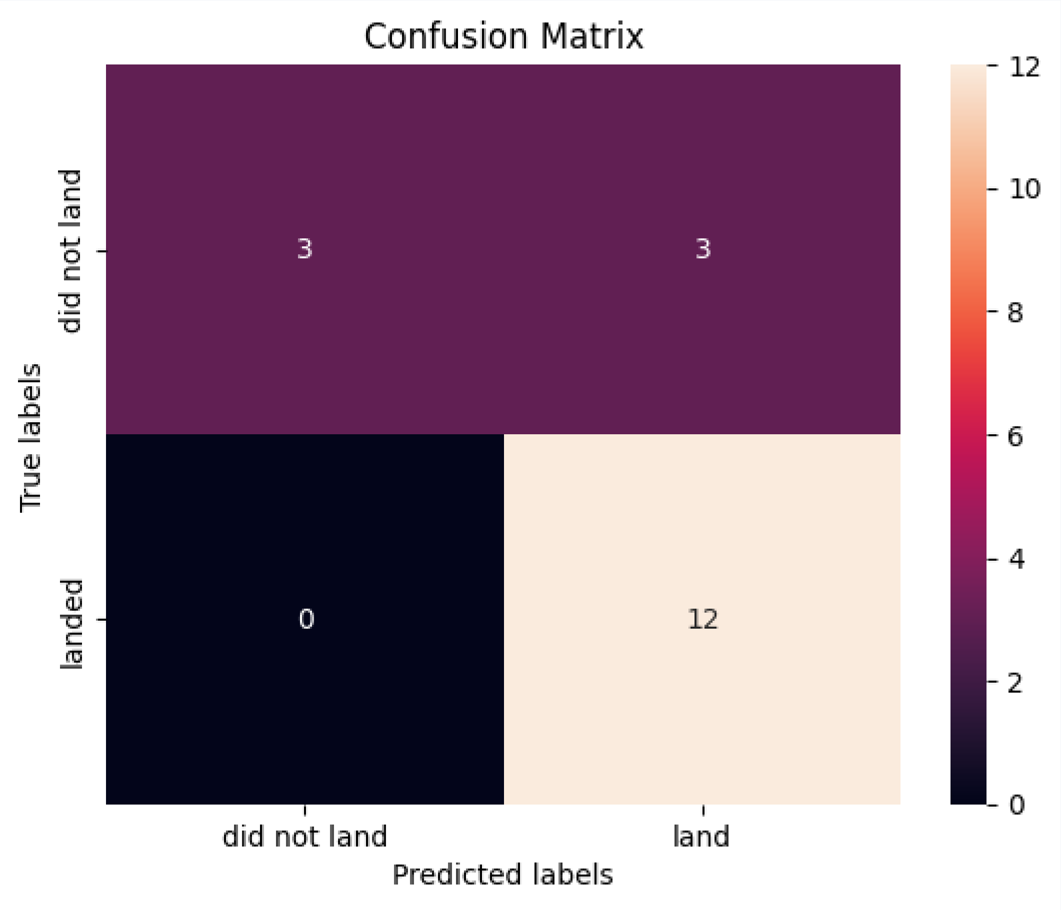
Predictive Analysis (Classification)

Classification Accuracy



- All four models, namely Logistic Regression, SVM, Decision Tree, and KNN, performed equally well with an accuracy of 0.83.

Confusion Matrix



- The confusion matrix is consistent across all four models.
- Each model effectively distinguishes between different classes.
- Notably, the primary challenge observed is the occurrence of false positives.

Conclusions

In closing, our exploration of SpaceX's Falcon 9 rocket launches has yielded invaluable insights into the dynamic realm of space exploration and cost-efficiency strategies. The predictive analysis of landing success, achieved through a meticulous combination of data collection methodologies and advanced classification models, has illuminated critical patterns and trends.

Key Reflections:

- Operational Enhancements: The success rates of various launch sites, orbital trajectories, and payload configurations have showcased a trajectory of improvement. This evolution underscores the efficiency gains achieved by SpaceX over the years.
- Geographical and Strategic Significance: The strategic selection of launch sites near essential routes, coastlines, and safe distances from populated areas highlights the careful considerations made to optimize logistical efficiency and ensure the safety of space missions.
- Predictive Power: The accuracy achieved by classification models—Logistic Regression, SVM, Decision Tree, and KNN—at 0.83 provides a robust foundation for predicting landing success, empowering decision-makers with actionable insights.

Conclusions

Future Implications:

- Cost Projections: The predictive capabilities developed in this study offer a pathway to revolutionize cost projections for rocket launches. The direct correlation between landing success and cost implications lays the groundwork for informed decision-making in the aerospace industry.
- Strategic Positioning: Beyond SpaceX, these findings empower potential competitors with critical insights for strategic positioning when bidding against SpaceX for rocket launches. Understanding success rates and their impact on costs is pivotal for navigating the competitive landscape.

Final Considerations:

- As we conclude, the fusion of data science methodologies and comprehensive results provides a lens through which we can not only understand the intricacies of Falcon 9 launches but also influence future strategies in the dynamic and ever-evolving field of space exploration. The journey undertaken in this project serves as a testament to the power of predictive analytics in shaping the trajectory of space missions and economic efficiency in the aerospace industry.

Appendix

- The SpaceX API: https://api.spacexdata.com/v4/launches/past
- The Falcon9 Launch Wiki page: https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Thank you!