IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Michael Diop Rogandji
07/15/2024

Outline

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

Executive Summary

- Data Collection
- Data Wrangling
- Exploratory Data Analysis (EDA)
- Interactive Visual Analytics and Dashboard
- Predictive Analysis

Introduction

- This project aims to predict the success of the Falcon 9 rocket's first stage landing, which is crucial for reducing launch costs and enhancing competitive strategies in the commercial rocket market. SpaceX’s ability to reuse the first stage contributes significantly to its lower launch price of $62 million compared to competitors.
- Key Factors Influencing Success:
  - Rocket Variables: Analyzing technical specifications, flight dynamics, and landing systems.
  - Launch Conditions: Evaluating weather, sea state, and payload characteristics.
  - Mission Profiles: Considering mission type, orbit requirements, and flight duration.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API: Gathered data directly from SpaceX's API.
  - Web Scraping: Collected additional data from Wikipedia to supplement the dataset.
- Perform data wrangling
  - Data Wrangling: Cleaned raw data for machine learning applications.
  - One-Hot Encoding: Converted categorical data into numerical format and dropped irrelevant columns.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Applied classification models to predict landing success.
  - Focused on building, tuning, and evaluating various classification models to improve prediction accuracy.

Data Collection

- SpaceX API: Used GET requests to retrieve data, decoded the JSON response, and converted it into a pandas DataFrame using .json_normalize().
- Data Cleaning: Checked for and addressed missing values to ensure data quality.
- Web Scraping: Wikipedia: Used BeautifulSoup to scrape Falcon 9 launch records, extracting HTML tables and converting them into pandas DataFrames for further analysis.
- Objective: To gather and prepare comprehensive data for analysis, focusing on cleaning, transforming, and organizing data from multiple sources for predictive modeling and insights.

Data Collection – SpaceX API

- We utilized GET requests to the SpaceX API to collect data, followed by cleaning, basic data wrangling, and formatting to prepare the dataset for analysis.

- Github link : https://github.com/kur0ii/SpaceX-Falcon9-Landing-Predictions

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
[81]: static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/dat
```

We should see that the request was successfull with the 200 status response code

```
[82]: response.status_code
[82]: 200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_norma`

```
[83]: # Use json_normalize meethod to convert the json result into a dataframe
      json_result = requests.get(static_json_url).json()
```

Using the dataframe `data` print the first 5 rows

```
[84]: data = pd.json_normalize(json_result)
      data
```

[84]:

	static_fire_date_utc	static_fire_date_unix	tbd	net	window	rocket	success	details
0	2006-03-	1.142554e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Engine failure at 33 seconds

Data Collection - Scraping

- We employed web scraping with BeautifulSoup to extract Falcon 9 launch records, then parsed the data and converted it into a pandas DataFrame.

- Github link : https://github.com/kur0ii/SpaceX-Falcon9-Landing-Predictions

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

```
soup = BeautifulSoup(response.text,"html.parser")
```

```
soup.title
```

```
<title>List of Falcon 9 and Falcon Heavy launches
```

```
html_tables = soup.find_all("table")
```

Starting from the third table is our target table contai

You should able to see the columns names embedde

Next, we just need to iterate through the `<th>` elem

```
column_names = []

th_elements = first_launch_table.find_all("th")
for element in th_elements :
    name = extract_column_from_header(element)
    if name is not None and len(name) > 0 :
        column_names.append(name)
```

Data Wrangling

- We conducted exploratory data analysis to identify training labels.
- We also created a landing outcome label based on the outcome column and exported the results to a CSV file.

- Github link : https://github.com/kur0ii/SpaceX-Falcon9-Landing-Predictions

```python
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
# Assign landing_class based on the Outcome column
landing_class = df['Outcome'].apply(lambda x: 1 if "True" in x else 0)
```

This variable will represent the classification variable that represents the outcom
landed Successfully

```python
df['Class']=landing_class
df[['Class']].head(8)
```

EDA with Data Visualization

- Scatter plots: Plotted for various variable pairs like Flight Number vs. Payload Mass and Payload Mass vs. Orbit Type, to examine relationships between them, which can inform machine learning models.
- Bar charts: Used to compare discrete categories, showing the relationship between specific categories and a measured value.
- Line charts: Used to display trends over time, particularly for showing the yearly trend of success rates.
- Github link : https://github.com/kur0ii/SpaceX-Falcon9-Landing-Predictions

EDA with SQL

- Unique Launch Sites: Displayed the names of unique launch sites.
- Launch Sites Starting with 'CCA': Retrieved 5 records with launch sites beginning with 'CCA'.
- Total Payload by NASA (CRS): Calculated the total payload mass for NASA (CRS) launched boosters.
- Average Payload for F9 v1.1: Displayed the average payload mass for the F9 v1.1 booster version.
- First Successful Ground Pad Landing: Listed the date of the first successful landing on a ground pad.
- Boosters with Specific Payload and Success: Listed boosters with successful drone ship landings and payloads between 4000 and 6000.
- Mission Outcomes: Counted the total number of successful and failed mission outcomes.
- Boosters with Maximum Payload: Identified the booster versions that carried the maximum payload mass.
- Failed Landings in 2015: Listed failed drone ship landings in 2015, along with their booster versions and launch sites.
- Ranking Landing Outcomes: Ranked the count of landing outcomes between 2010-06-04 and 2017-03-20 in descending order.
- Github link : https://github.com/kur0ii/SpaceX-Falcon9-Landing-Predictions

Build an Interactive Map with Folium

- Markers for Launch Sites: Added markers with circles, popup labels, and text labels to represent the NASA Johnson Space Center and other launch sites, indicating their geographical locations and proximity to the Equator and coasts.
- Coloured Markers for Launch Outcomes: Used colored markers (green for success and red for failure) within a marker cluster to visualize the success rates of different launch sites.
- Distance Visualization: Added colored lines to illustrate the distances between the KSC LC-39A launch site and nearby features such as a railway, highway, coastline, and closest city.
- Github link : https://github.com/kur0ii/SpaceX-Falcon9-Landing-Predictions

Build a Dashboard with Plotly Dash

- Launch Sites Dropdown List: Implemented a dropdown menu to allow users to select specific launch sites.
- Pie Chart for Launch Success: Added a pie chart displaying the total successful launches across all sites, and the success vs. failure counts for a selected site.
- Payload Mass Range Slider: Introduced a slider for users to select and filter by payload mass range.
- Scatter Chart of Payload vs. Success Rate: Created a scatter chart to visualize the correlation between payload mass and launch success across different booster versions.
- Github link : https://github.com/kur0ii/SpaceX-Falcon9-Landing-Predictions

Predictive Analysis (Classification)

- Data Loading and Preparation: Data was loaded using NumPy and Pandas, then transformed and split into training and testing sets.
- Model Building and Hyperparameter Tuning: Multiple machine learning models were built, and hyperparameters were optimized using GridSearchCV.
- Model Evaluation and Improvement: Accuracy was used as the evaluation metric. The model was improved through feature engineering and algorithm tuning.
- Best Model Identification: The best-performing classification model was identified based on its performance.
- Github link : https://github.com/kur0ii/SpaceX-Falcon9-Landing-Predictions

Results

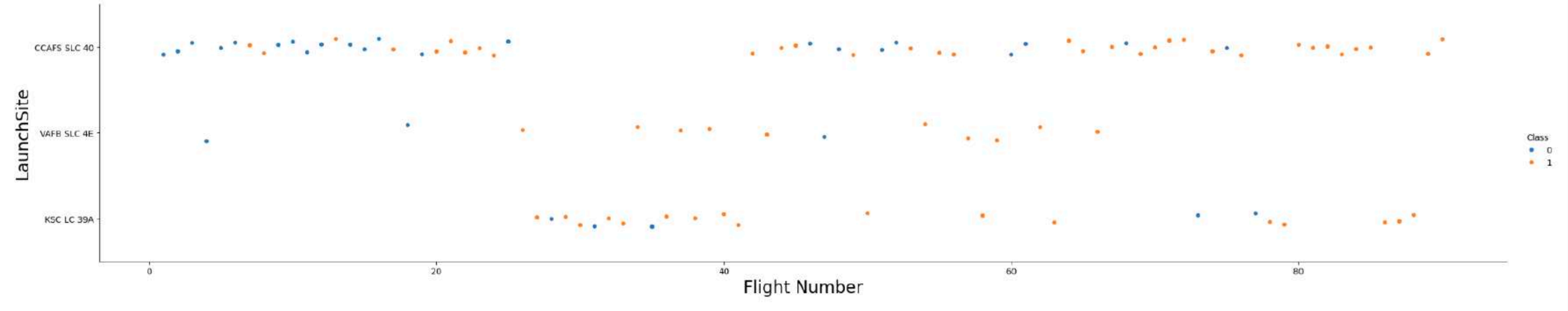
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

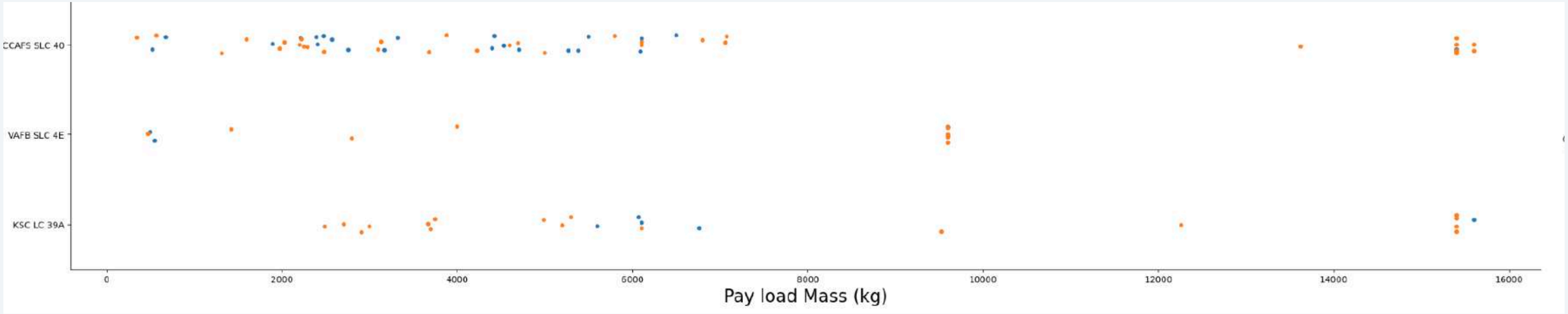
Flight Number vs. Launch Site

- Flight Success Over Time: The earliest flights failed, while the latest flights were all successful, suggesting an improvement in success rates over time.
- CCAFS SLC 40 Launch Site: This site accounts for about half of all launches.
- Higher Success Rates at Specific Sites: VAFB SLC 4E and KSC LC 39A have notably higher success rates.
- Increasing Success with New Launches: It's assumed that each new launch has a higher likelihood of success, indicating a trend of improving reliability.



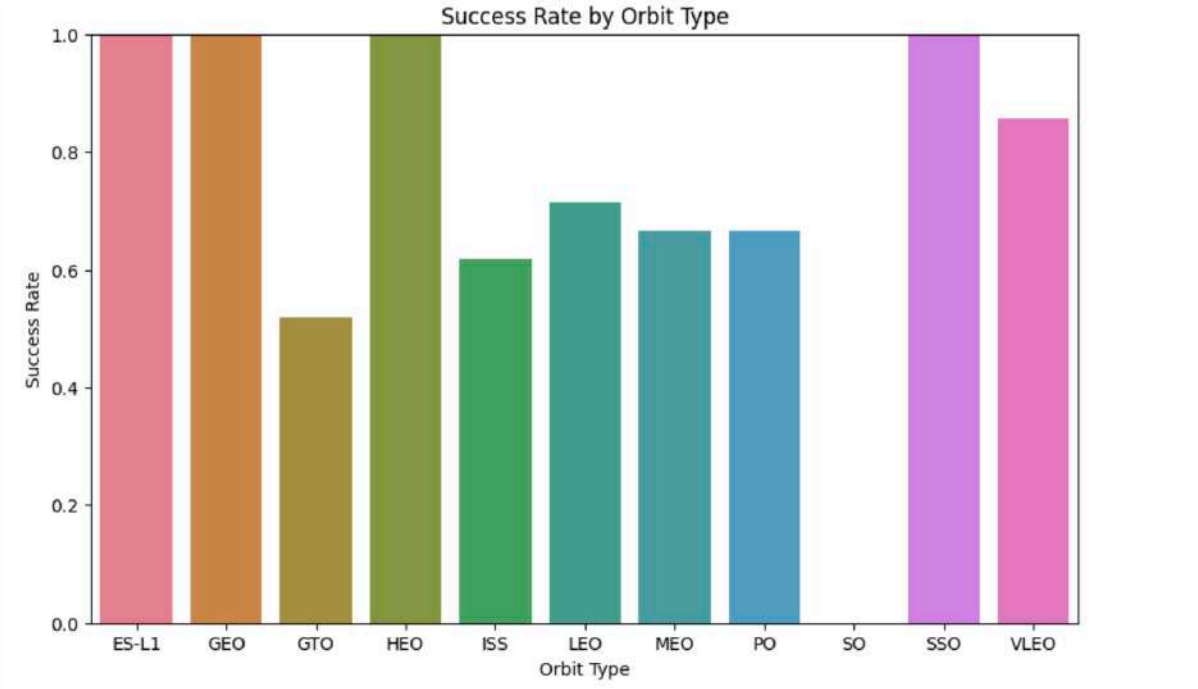
Payload vs. Launch Site

- Payload Mass and Success Rate: Across all launch sites, higher payload mass is associated with a higher success rate.
- Success with High Payloads: Most launches carrying payloads over 7,000 kg were successful.
- KSC LC 39A Performance: The KSC LC 39A launch site has a 100% success rate for payloads under 5,500 kg as well.



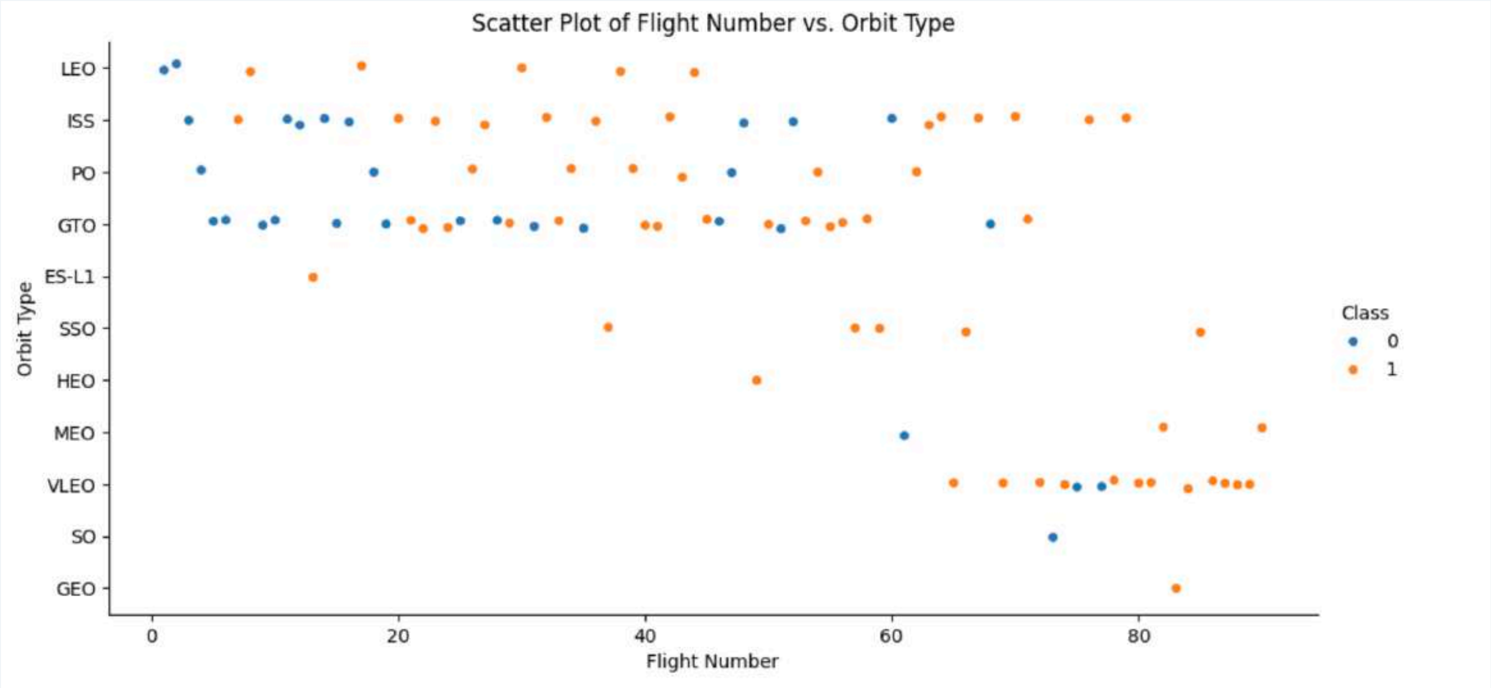
Success Rate vs. Orbit Type

- Orbits with 100% Success Rate: ES-L1, GEO, HEO, and SSO orbits have achieved a 100% success rate.
- Orbit with 0% Success Rate: SO orbit has a 0% success rate.s



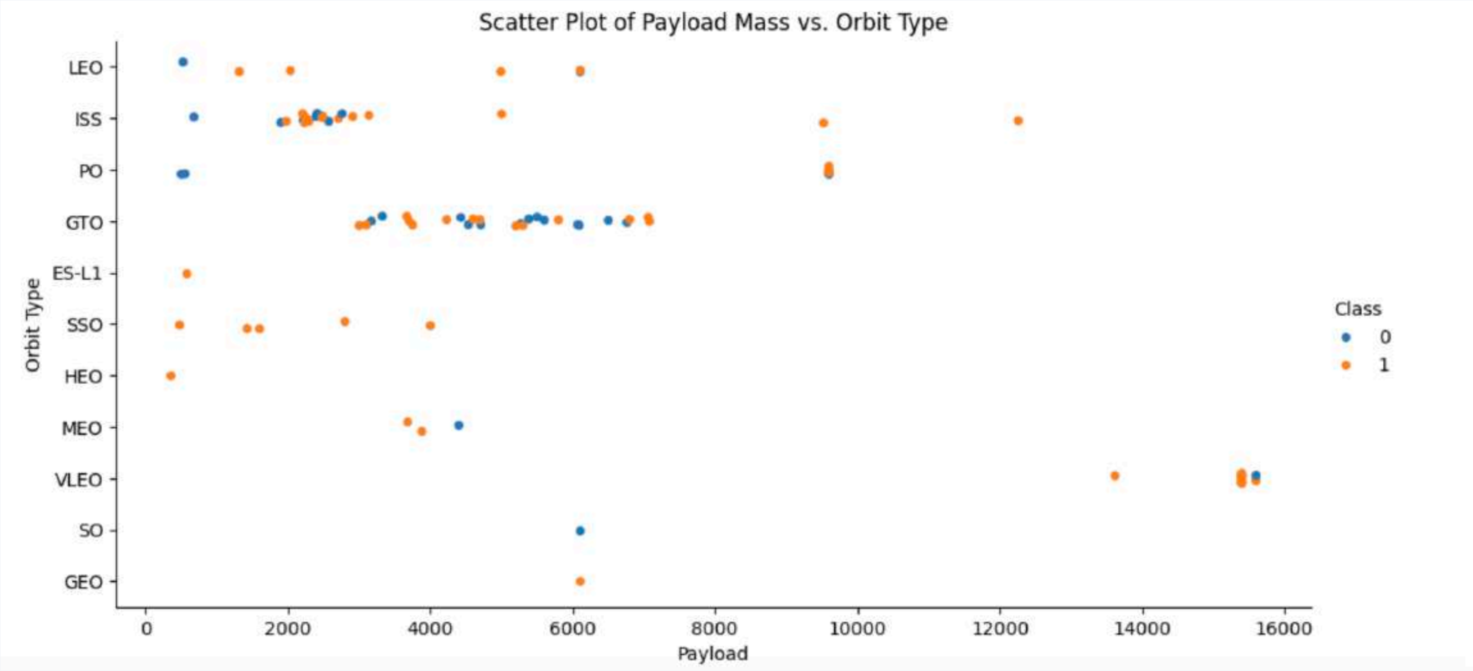
Flight Number vs. Orbit Type

- LEO Orbit: Success rates in the LEO orbit appear to be related to the number of flights, suggesting that more flights might improve success rates.
- GTO Orbit: There is no apparent relationship between the number of flights and success rates in the GTO orbit.



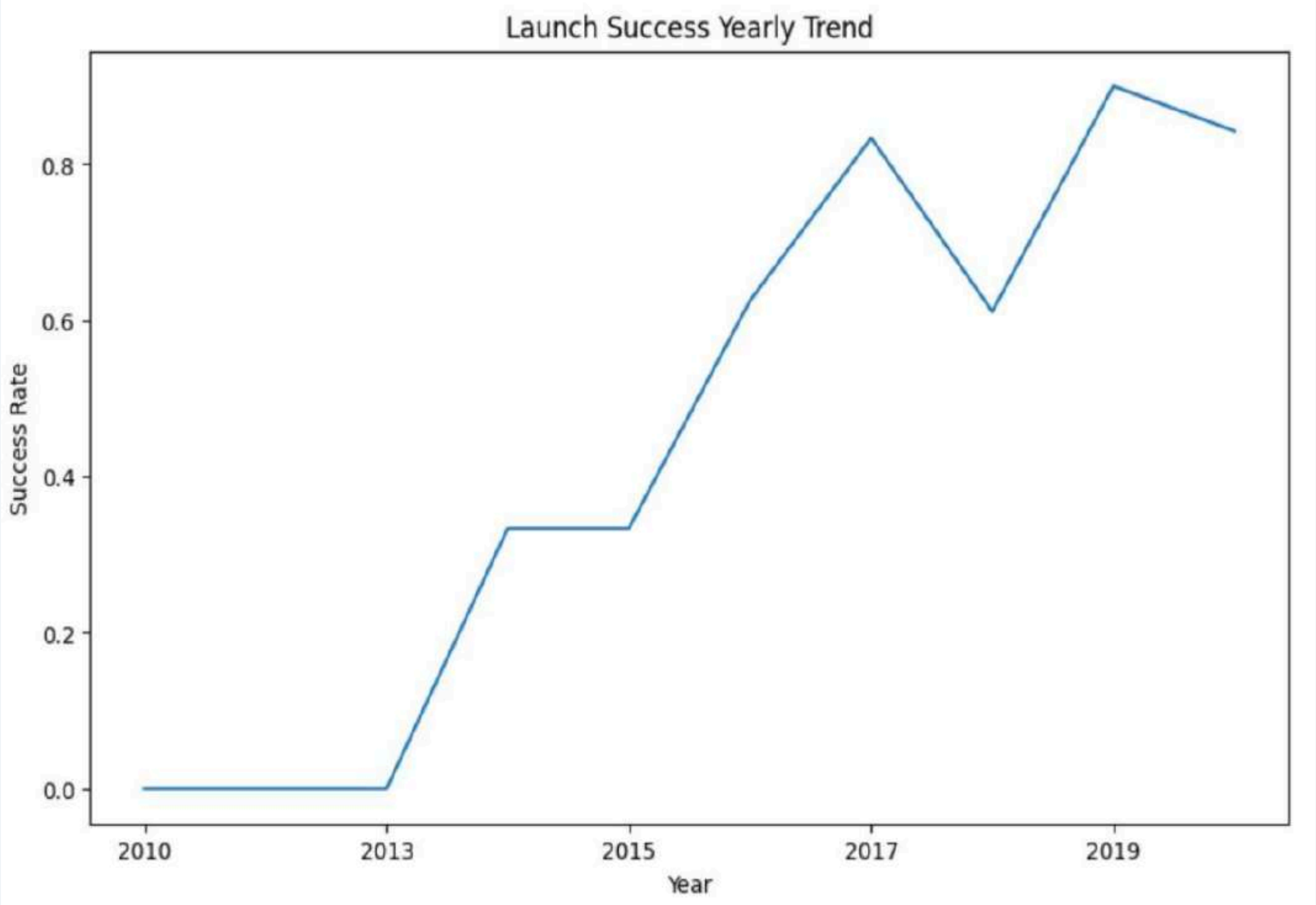
Payload vs. Orbit Type

- Payloads Above 10,000 kg: These were placed in PO (Polar Orbit), ISS (International Space Station), and LEO (Low Earth Orbit) orbits.
- Payloads Between 4,000 kg and 8,000 kg: These were placed in the GTO (Geostationary Transfer Orbit).



Launch Success Yearly Trend

- Success Rate Trend: The plot shows that the success rate has been steadily increasing from 2013 through 2020.



All Launch Site Names

- The query below display the names of the unique launch sites in the space mission :

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTBL ;
```

```
 * 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'

- The query below display 5 records where launch sites begin with the string ‘CCA'

```
%sql SELECT * FROM SPACEXTBL 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

- The query below display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The query below calculate the average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version like 'F9 v1.1%';
```

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

- The query below displays the date of the first successful landing outcome on ground pad

```
%sql SELECT MIN(DATE) FROM SPACEXTBL WHERE Landing_Outcome LIKE 'SUCCESS%' ;
```

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

MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The query below displays the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%sql SELECT Booster_Version FROM SPACEXTBL WHERE Mission_Outcome LIKE 'SUCCESS' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
```

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

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022

Total Number of Successful and Failure Mission Outcomes

- The query below displays the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(*) FILTER (WHERE Mission_Outcome LIKE 'SUCCESS') as successful,COUNT(*) FILTER \
    (WHERE Mission_Outcome NOT LIKE 'SUCCESS') as failure FROM SPACEXTBL;
```

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

successful	failure
98	3

Boosters Carried Maximum Payload

- The query below displays the names of the booster which have carried the maximum payload mass

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

```
 * 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

2015 Launch Records

- The query returns records from 2015 where the landing outcome involved a failure on a drone ship, along with the month, landing outcome details, booster version, and launch site.

```
%sql SELECT substr(Date, 6, 2) as monthname, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL  WHERE substr(Date, 0, 5) = '2015'  ANI
```

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

monthname	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

- This query counts how many times each landing outcome occurred between June 4, 2010, and March 20, 2017. It groups the results by Landing_Outcome, orders them by the count in descending order, and returns the landing outcome and its corresponding count.

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.

```
%sql SELECT Landing_Outcome, COUNT(*) as outcome_count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outco
```

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

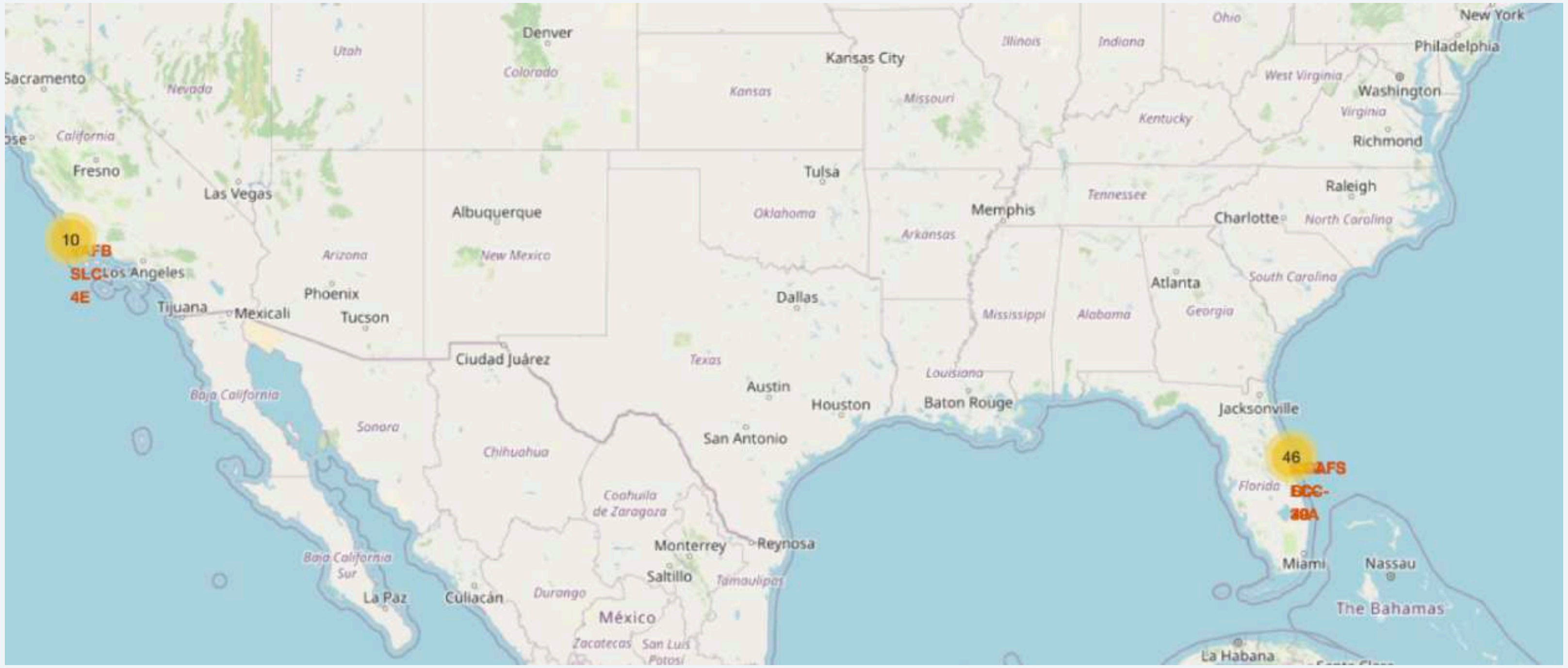
Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

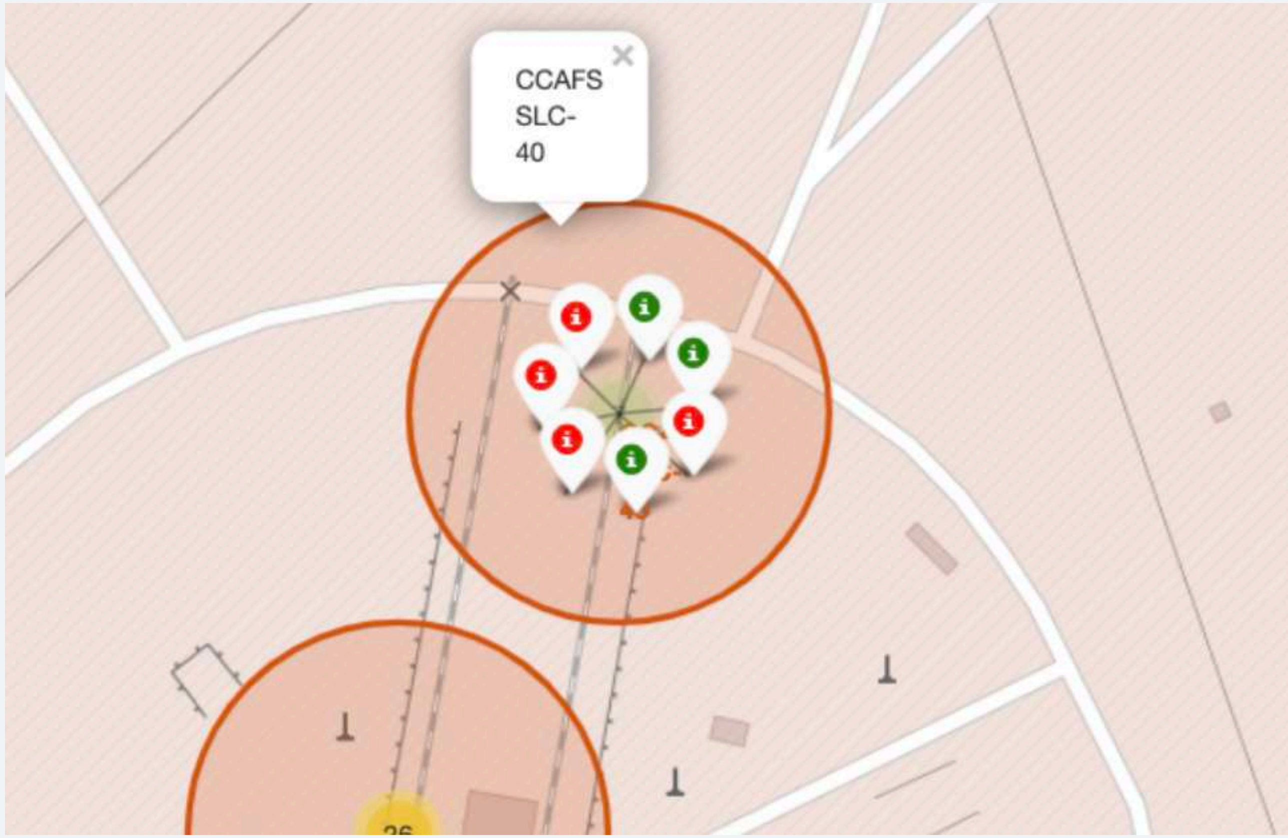
Global Launch Sites: Key Locations in California and Florida, USA

- Most launch sites are near the equator because the Earth rotates faster there, giving rockets an extra speed boost of 1670 km/hour due to inertia, which helps them reach orbit. Additionally, launch sites are close to the coast to minimize the risk of debris falling on populated areas by launching rockets over the ocean.



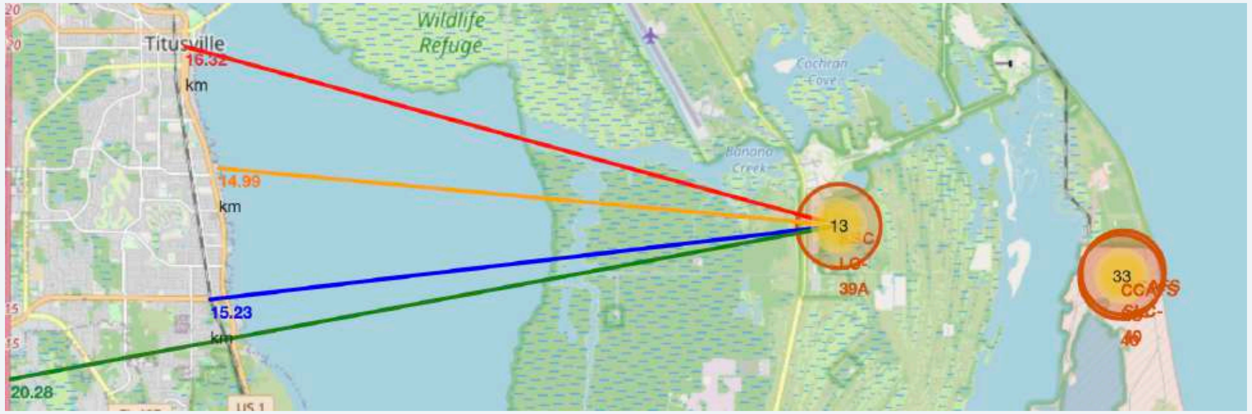
Color-Coded Global Launch Sites: Visualizing Launch Records on the Map

- Green markers shows a successful launch whereas red markers shows a failed launch.



Proximity Analysis: Distances from KSC LC-39A to Nearby Landmarks

- It is relatively close to a railway (15.23 km), highway (20.28 km), coastline (14.99 km), and the nearest city, Titusville (16.32 km).
- Given that a failed rocket could travel 15-20 km in seconds, this proximity to populated areas could pose a safety risk.

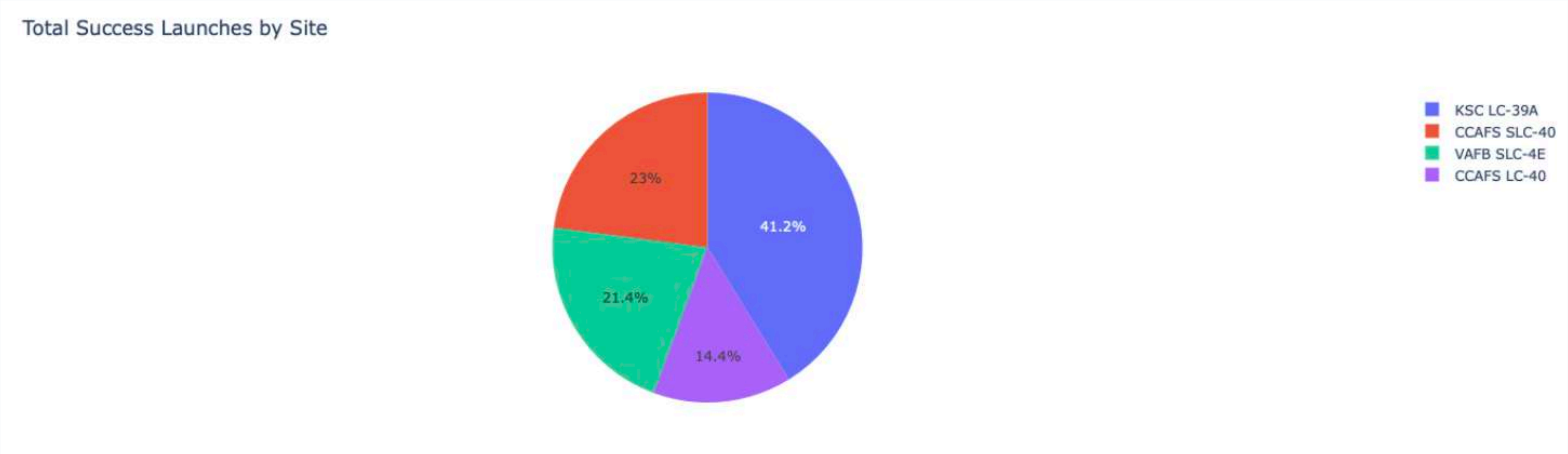


Section 4

Build a Dashboard with Plotly Dash

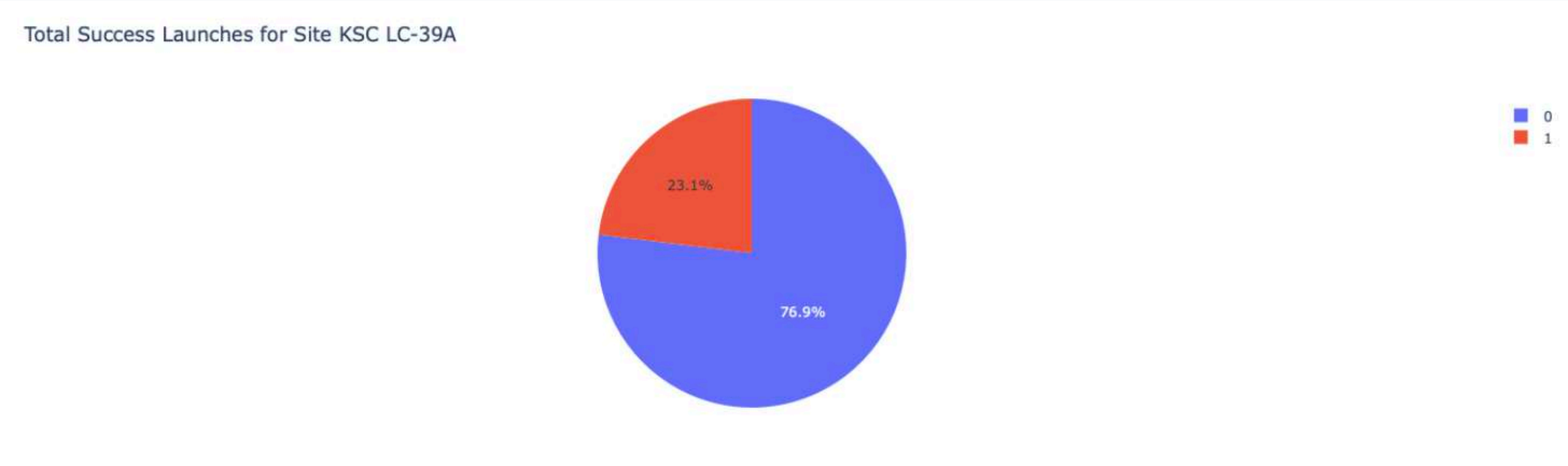
Total Success Launches by Site

- KSC LC-39A has the highest number of successful launches among all the sites.



Total Success Launches for Site KSC LC-39A

- KSC LC-39A has the highest launch success rate at 76.9%.



Payload Mass and Launch Outcomes Across All Sites

- The charts reveal that payloads ranging from 2,000 to 5,500 kg have the highest success rate.

Section 5

Predictive Analysis (Classification)

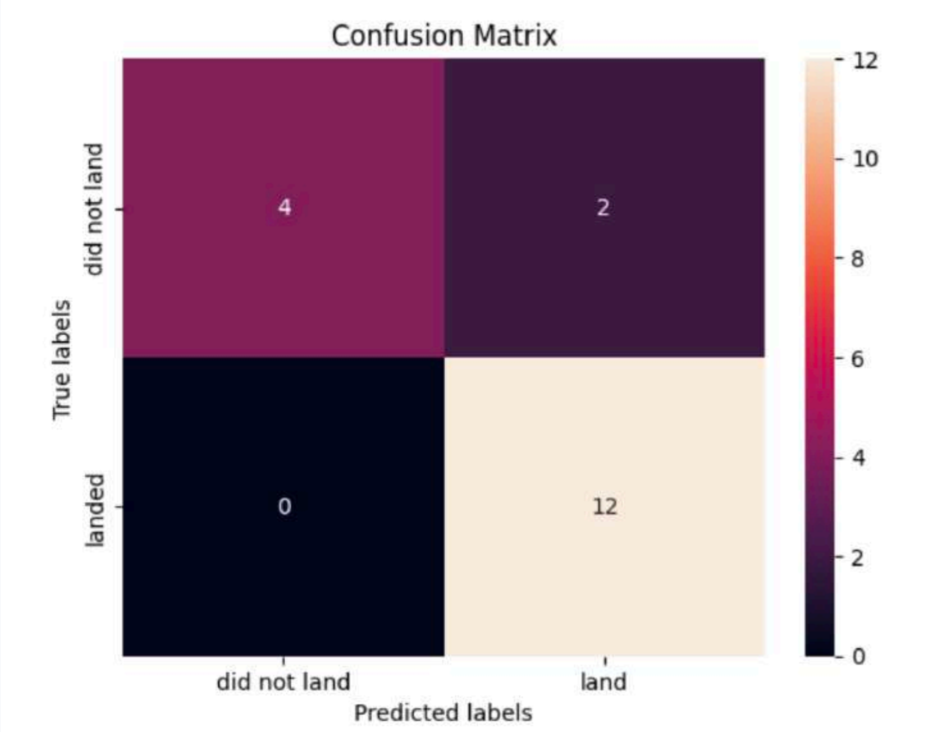
Classification Accuracy

- Tree Model Performance: The Decision Tree model achieved the highest accuracy across the entire dataset.

	K-Nearest Neighbors	Support Vector Machine	Decision Tree	Logistic Regression
accuracy	0.833333	0.833333	0.888889	0.833333
precision	0.866667	0.866667	0.904762	0.866667
recall	0.833333	0.833333	0.888889	0.833333
f1_score	0.814815	0.814815	0.882051	0.814815

Confusion Matrix

- The confusion matrix indicates:
  - 12 True Positives
  - 3 False Positives
  - 3 True Negatives
  - 0 False Negatives



Conclusions

- Success Rate Improvement: Rocket launch success rates have increased since 2013.
- Best Orbits: ES-L1, GEO, HEO, and SSO orbits have a 100% success rate.
- Top Launch Site: KSC LC-39A has the highest success rate among launch sites.
- Best Model: The Decision Tree model provides the best accuracy for the SpaceX dataset.
- Payload Mass: Lower payload masses generally result in better launch success.

Appendix

- Github folder link : https://github.com/kur0ii/SpaceX-Falcon9-Landing-Predictions

Thank you!