IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohamad Sabha
11-May--2025

Outline

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

Executive Summary

- **Summary of methodologies**
    - Data Collection using SpaceX API
    - Data Collection with Web Scraping
    - Data Wrangling
    - Exploratory Data Analysis using SQL
    - EDA DataViz Using Python Pandas and Matplotlib
    - Launch Sites Analysis with Folium-Interactive Visual Analytics and Ploty Dash
    - Machine Learning Landing Prediction
- **Summary of all results**
    - Exploratory Data Analysis (EDA) Results Using Visualization and SQL
    - Interactive Visualization Analytics and Dashboards using Folium and Plotly
    - Predictive Analysis Using Classification Models

Introduction

- Project background and context

SpaceX 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 SpaceX 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 SpaceX for a rocket launch.

Falcon 9 is a two-stage reusable rocket developed and manufactured by SpaceX, a private aerospace company founded by Elon Musk. Here are some key features and missions of Falcon 9 rockets

Reusability: One of the notable features of Falcon 9 is its reusability. The first stage of the rocket is designed to return to Earth after launch, landing vertically either on land (at SpaceX's landing zones) or on an autonomous drone ship in the ocean. This reusability significantly reduces the cost of space launches.

Payload Capacity: Falcon 9 is capable of delivering a variety of payloads to orbit, including satellites, cargo resupply missions to the International Space Station (ISS), and even crewed missions. It has a payload capacity of up to 22,800 kilograms (50,300 pounds) to low Earth orbit (LEO) and up to 8,300 kilograms (18,300 pounds) to geostationary transfer orbit (GTO).

- Problems you want to find answers

In this capstone, we will predict if the Falcon 9 first stage will land successfully using data from Falcon 9 rocket launches advertised on its website.

Section 1

Methodology

Methodology

- 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

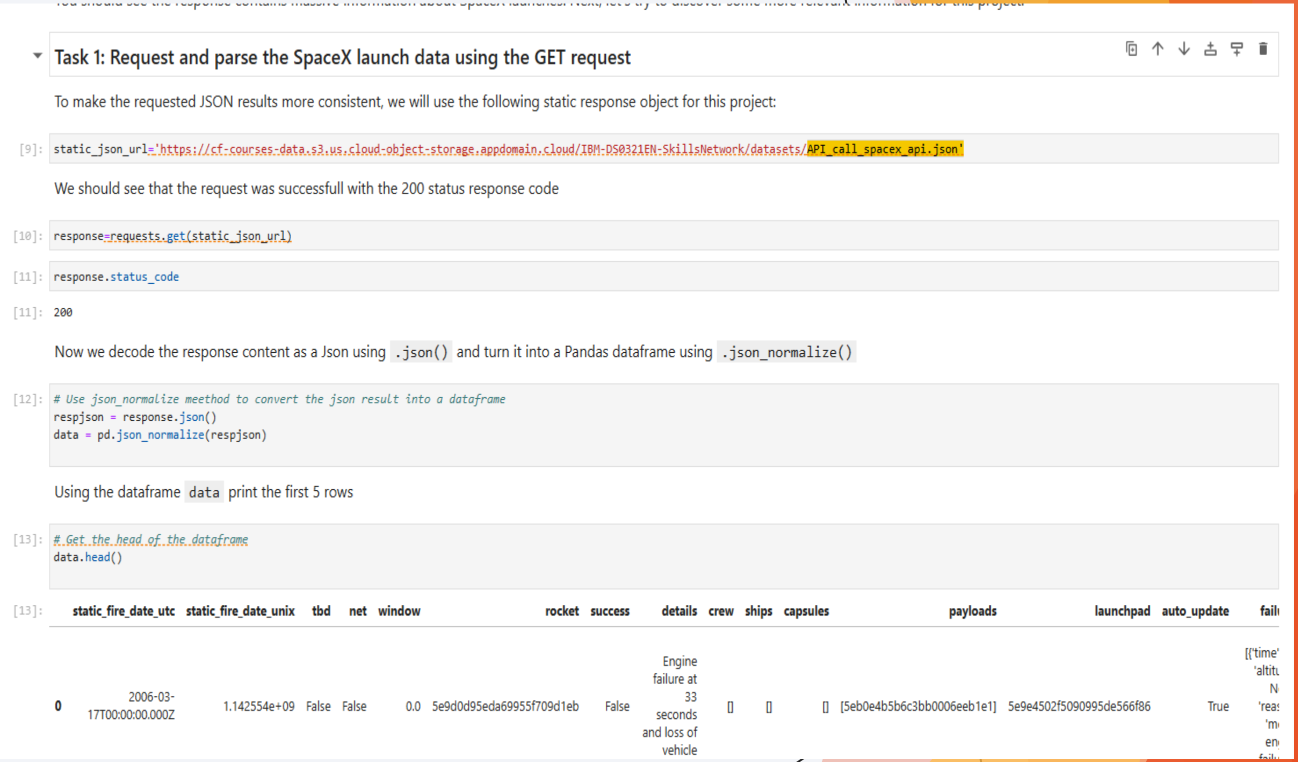
Data Collection

- Data collection began by using the SpaceX RESTful API through GET requests. A set of helper functions was created to streamline interaction with the API, allowing data extraction using launch identification numbers. Rocket launch information was then retrieved from the SpaceX API endpoint.
- To ensure uniformity in the JSON responses, the retrieved launch data was parsed from the GET request, decoded, and transformed into a Pandas DataFrame for easier analysis.
- Additionally, historical Falcon 9 launch records were gathered through web scraping from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches. Using the requests and BeautifulSoup libraries, the HTML table containing the launch data was extracted, parsed, and also converted into a Pandas DataFrame.

Data Collection – SpaceX API

- The data was obtained from the SpaceX RESTful API by sending a GET request. The launch data retrieved from the API was then parsed, the response content was decoded as JSON, and subsequently converted into a Pandas DataFrame for further analysis.
- You can find the completed notebook for the SpaceX API calls at the following GitHub URL.

https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/1-%20jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Conducted web scraping to gather historical Falcon 9 launch data from a Wikipedia page. Utilizing the requests and BeautifulSoup libraries, the HTML table containing the launch records was extracted and parsed. The structured data was then converted into a Pandas DataFrame for analysis.
- The completed web scraping notebook is available at the following GitHub URL.
- https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/2-%20jupyter-labs-webscraping.ipynb

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

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

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

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response.content, 'html.parser')
```

Print the page title to verify if the `BeautifulSoup` object was created properly

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

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

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

Next, we want to collect all relevant column names from the HTML table header

Let's try to find all tables on the wiki page first. If you need to refresh your memory about `BeautifulSoup`, please check the external reference link towards the end of this lab

```
# 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.

Data Wrangling

- We performed Some Data wrangling techniques to identify patterns in the dataset and define suitable labels for training supervised learning models. The dataset included various mission outcomes indicating whether booster landings were successful.
- For example:
  - True Ocean indicated a successful landing in a designated ocean region, while False Ocean denoted a failed landing attempt.
  - True RTLS signified a successful landing on a ground pad, whereas False RTLS indicated failure.
  - True ASDS referred to a successful landing on a drone ship, and False ASDS represented an unsuccessful attempt.
- These outcomes were converted into binary training labels, where 1 represented a successful landing and 0 indicated failure.
- The analysis included the following tasks:
  - Calculated the number of launches at each launch site
  - Determined the count and frequency of each orbit type
  - Analyzed mission outcomes by orbit type
  - Created a binary landing outcome label from the Outcome column
- Here is the GitHub URL of the completed data wrangling related notebooks

https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/3-%20labs-jupyter-spacex-Data%20wrangling.ipynb

TASK 2: Calculate the number and occurrence of each orbit

Use the method `.value_counts()` to determine the number and occurrence of each orbit in the column `Orbit`

```
# Apply value_counts on Orbit column
occurence = df['Orbit'].value_counts()
occurence
```

```
Orbit
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
HEO       1
ES-L1     1
SO        1
GEO       1
Name: count, dtype: int64
```

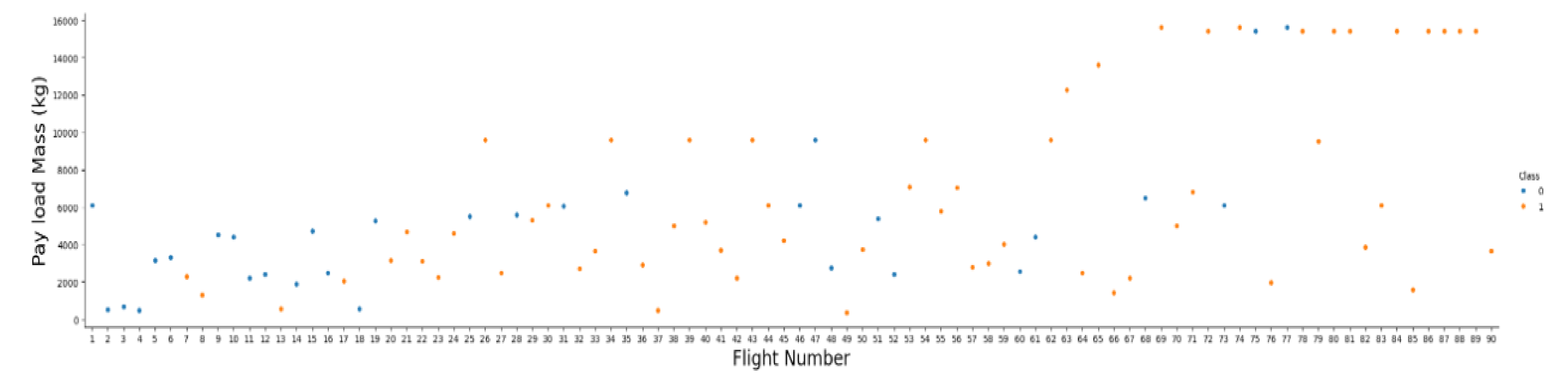
TASK 3: Calculate the number and occurence of mission outcome of the orbits

Use the method `.value_counts()` on the column `Outcome` to determine the number of `landing_outcomes`.Then assign it to a variable landing_outcomes.

```
# landing_outcomes = values on Outcome column
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes
```

```
Outcome
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: count, dtype: int64
```

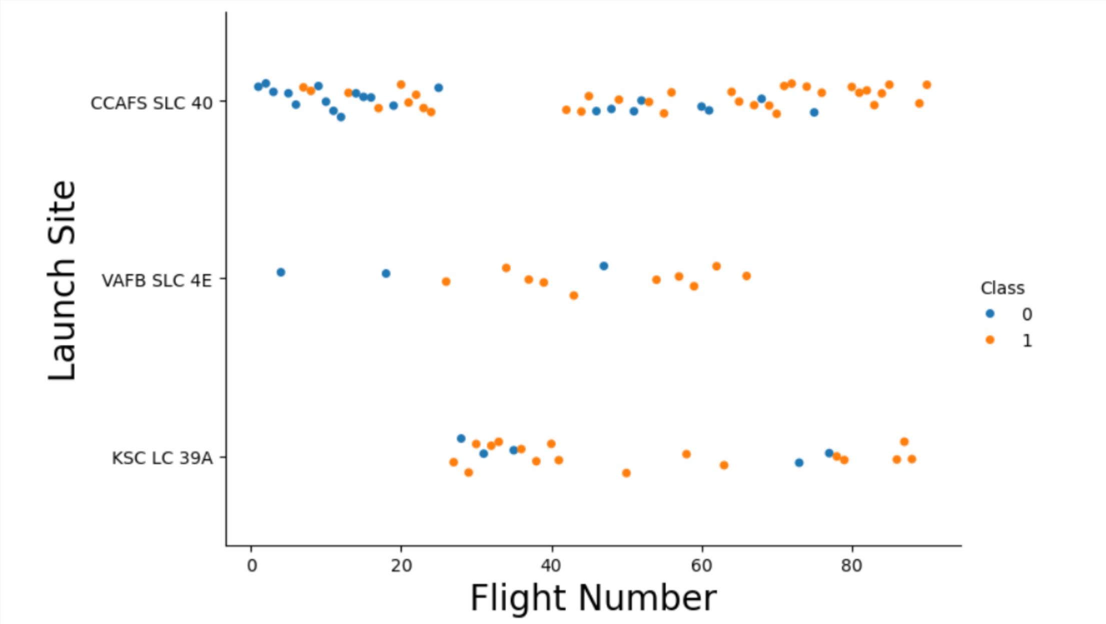
EDA with Data Visualization 1



After plotting out the FlightNumber vs. PayloadMassand overlay the outcome of the launch. We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass also appears to be a factor; even with more massive payloads, the first stage often returns successfully.

Completed EDA with data visualization notebook can be found in the following link : https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/5-%20EDA%20DataViz.ipynb

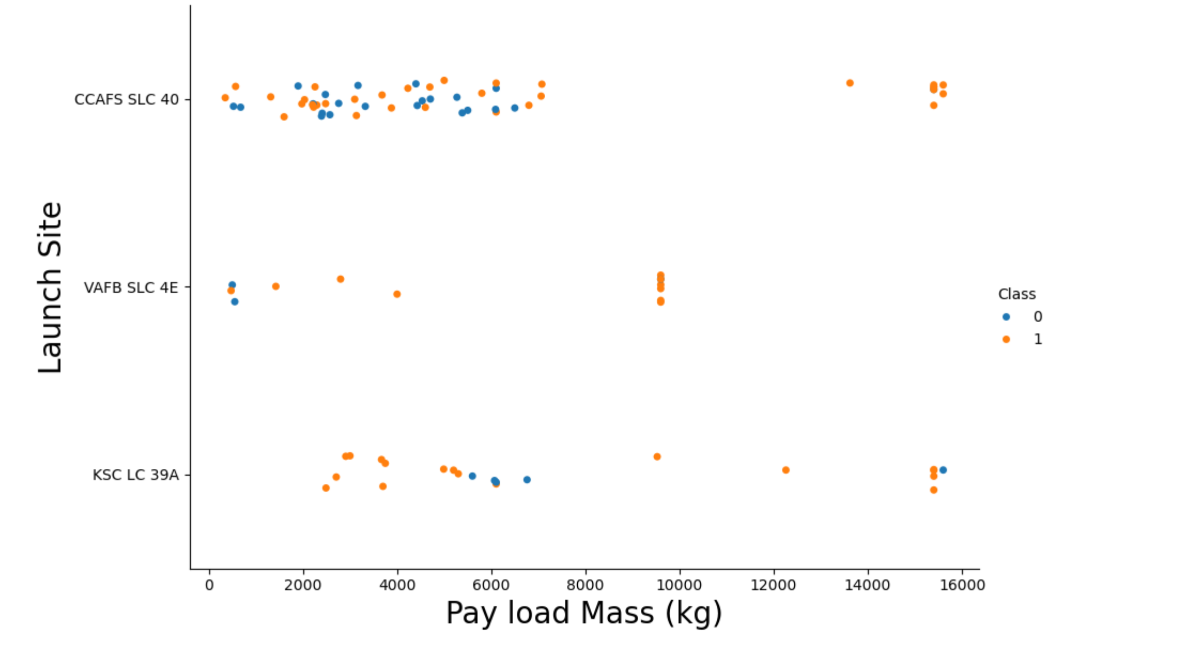
EDA with Data Visualization 2



Using the function catplot to plot FlightNumber vs LaunchSite, set the parameter x parameter to FlightNumber,set the y to Launch Site and set the parameter hue to 'class'. We can see that the CCAFS SLC 40 has the highest number of successful lunches as the flight number increases.

Completed EDA with data visualization notebook can be found in the following link : https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/5-%20EDA%20DataViz.ipynb

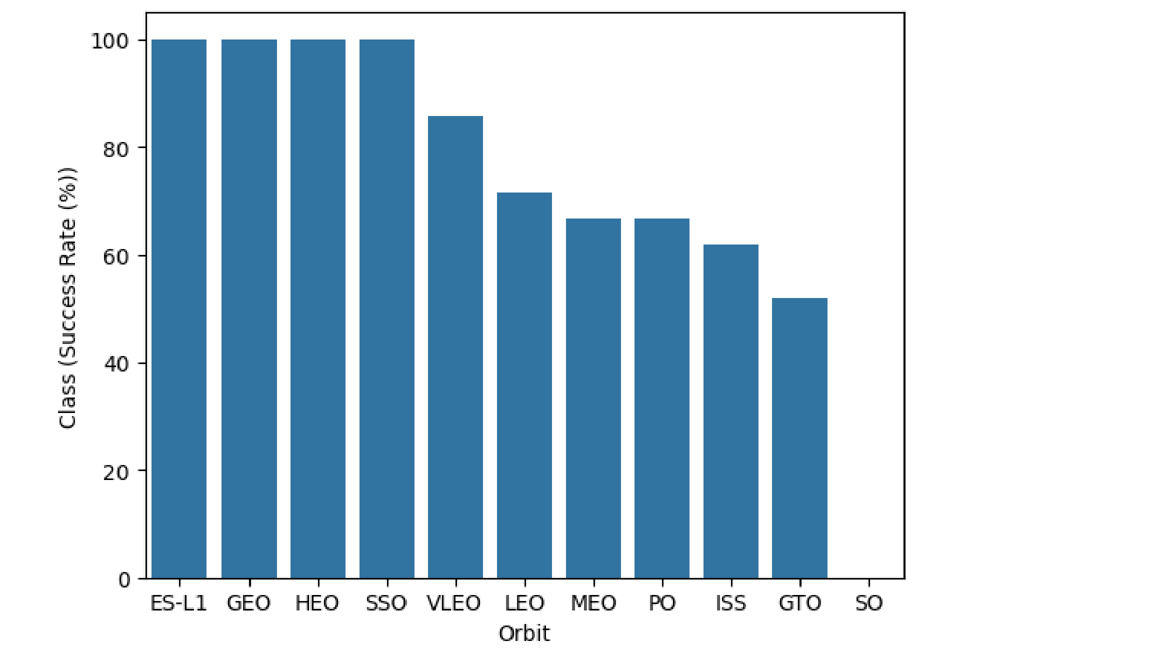
EDA with Data Visualization 3



We also want to observe if there is any relationship between launch sites and their payload mass., if we observe Payload Mass Vs. Launch Site scatter point chart we will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)

Completed EDA with data visualization notebook can be found in the following link : https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/5-%20EDA%20DataViz.ipynb

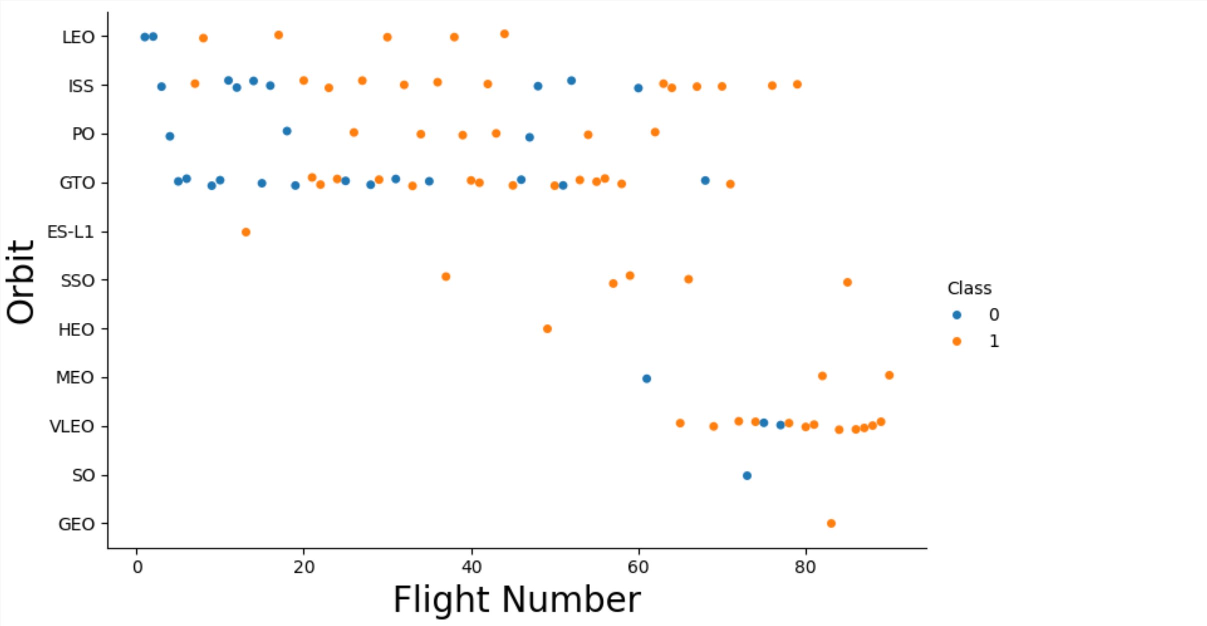
EDA with Data Visualization 4



Next, we want to visually check if there are any relationship between success rate and orbit type. we can see that ESL1 ,GEO and HEO orbits have the highest success rates and the SO has the Lowest 0 %.

Completed EDA with data visualization notebook can be found in the following link : https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/5-%20EDA%20DataViz.ipynb

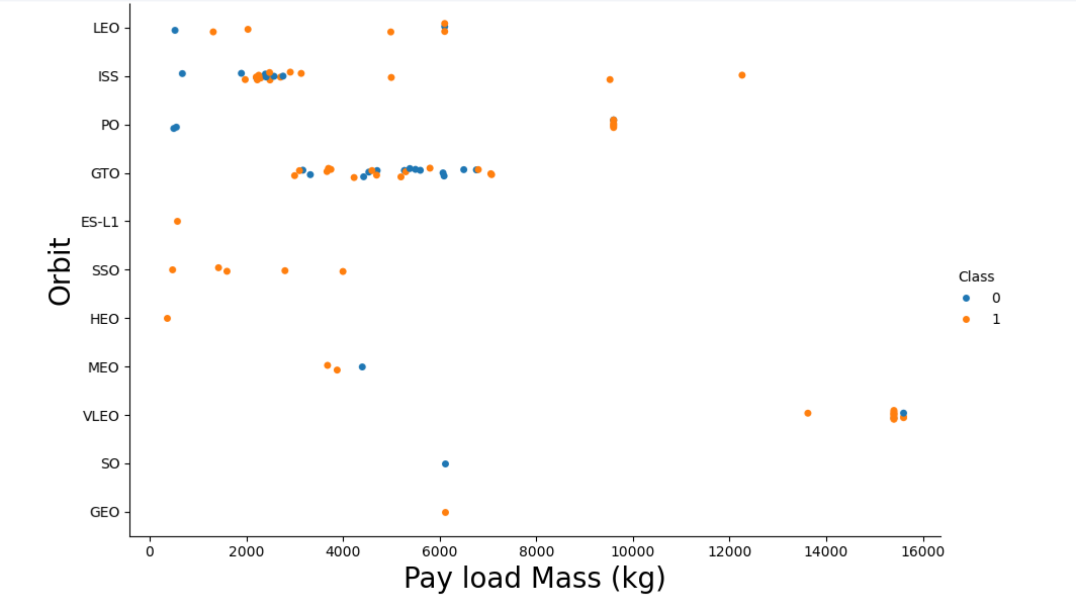
EDA with Data Visualization 5



For each orbit, we want to see if there is any relationship between FlightNumber and Orbit type. we can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Completed EDA with data visualization notebook can be found in the following link : https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/5-%20EDA%20DataViz.ipynb

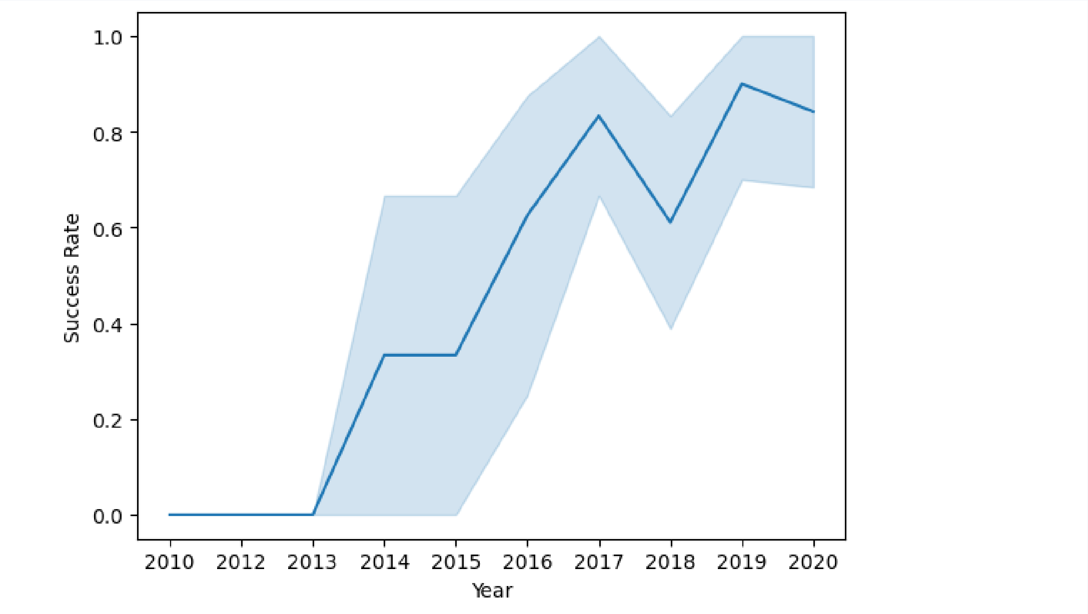
EDA with Data Visualization 6



Similarly, we can plot the Payload Mass vs. Orbit scatter point charts to reveal the relationship between Payload Mass and Orbit type. With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Completed EDA with data visualization notebook can be found in the following link : https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/5-%20EDA%20DataViz.ipynb

EDA with Data Visualization 7



we can plot a line chart with x axis to be Year and y axis to be average success rate, to get the average launch success trend.. we can observe that the sucess rate since 2013 kept increasing till 2020

Completed EDA with data visualization notebook can be found in the following link : https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/5-%20EDA%20DataViz.ipynb

EDA with SQL 1

- The following SQL queries were performed for EDA

  - 1- Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
```

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

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

  - 3- Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)';
```

  - 4- Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version LIKE 'F9 v1.1%';
```

  - 5- List the date when the first succesful landing outcome in ground pad was acheived.

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)';
```

  - 6- 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 DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE "Landing_Outcome" = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

EDA with SQL 2

- The following SQL queries were performed for EDA

  - 7- List the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

  - 8- List all the booster_versions that have carried the maximum payload mass. Use a subquery.

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

  - 9- 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.

```
%sql SELECT SUBSTR(Date,6,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Failure%drone%' AND SUBSTR(Date,0,5) = '2015';
```

  - 10- 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 Numbers FROM SPACEXTBL WHERE (Landing_Outcome LIKE 'Success%' Or Landing_Outcome LIKE 'Failure%') AND (Date BETWEEN '2010-06-04' AND '2017-03-20') GROUP BY Landing_Outcome ORDER BY Numbers DESC;
```

Here is the GitHub URL of your completed EDA with SQL notebook.

https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/4-%20jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

We performed more interactive visual analytics using Folium.in order to marked all the launch sites, and created map objects such as markers, circles, lines to mark the success or failure of launches for each launch site. The following points was achived

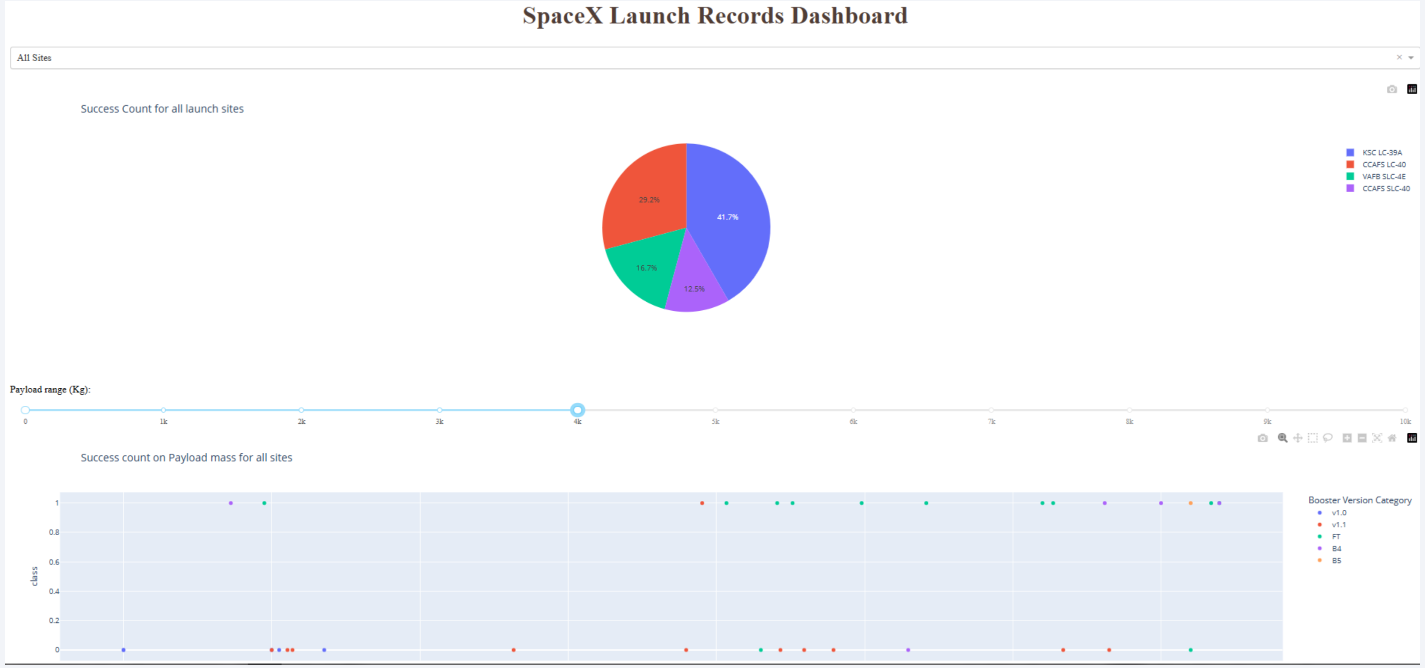
- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities . We were able to find some geographical patterns about launch sites

The completed interactive map with Folium map can be found in the following link: https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/6-%20Folium_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

We Developed an interactive dashboard application using Plotly Dash, which included the following features:

- Implemented a dropdown component to select the launch site
- Created a callback function to dynamically generate a success pie chart based on the selected site
- Added a range slider to enable selection of payload mass range
- Developed a callback function to display a success vs. payload scatter plot based on the selected inputs



Completed data visualization notebook of Dashboard with Plotly Dash can be found in the following link : https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/7-%20spacex-dash-app.py

Predictive Analysis (Classification)

Model Development Process

1. Data Preprocessing

We began by preparing the dataset:

- Cleaned and normalized the features.
- Encoded categorical variables to numerical format.
- Split the data into training and testing sets to evaluate generalization performance.

2. Model Selection & Initial Evaluation

We selected four popular classification algorithms:

- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree
- K-Nearest Neighbors (KNN)

Each model was trained on the training dataset and evaluated using cross-validation accuracy to assess its baseline performance.

3. Hyperparameter Tuning

To improve model performance, we performed Grid Search with Cross-Validation (GridSearchCV) to fine-tune each model's hyperparameters:

- For example, the number of neighbors for KNN, or the penalty parameter for Logistic Regression.
- This helped minimize overfitting and optimize for generalization.

4. Final Evaluation on Test Set

After tuning, we evaluated all models on the **unseen test dataset**. we saw that , **Logistic Regression**, **SVM**, and **KNN** all achieved the highest accuracy on the test data, tying at **83.3%**. The **Decision Tree model** lagged behind at **72.2%**, likely due to overfitting on the training set or insufficient generalization capacity.

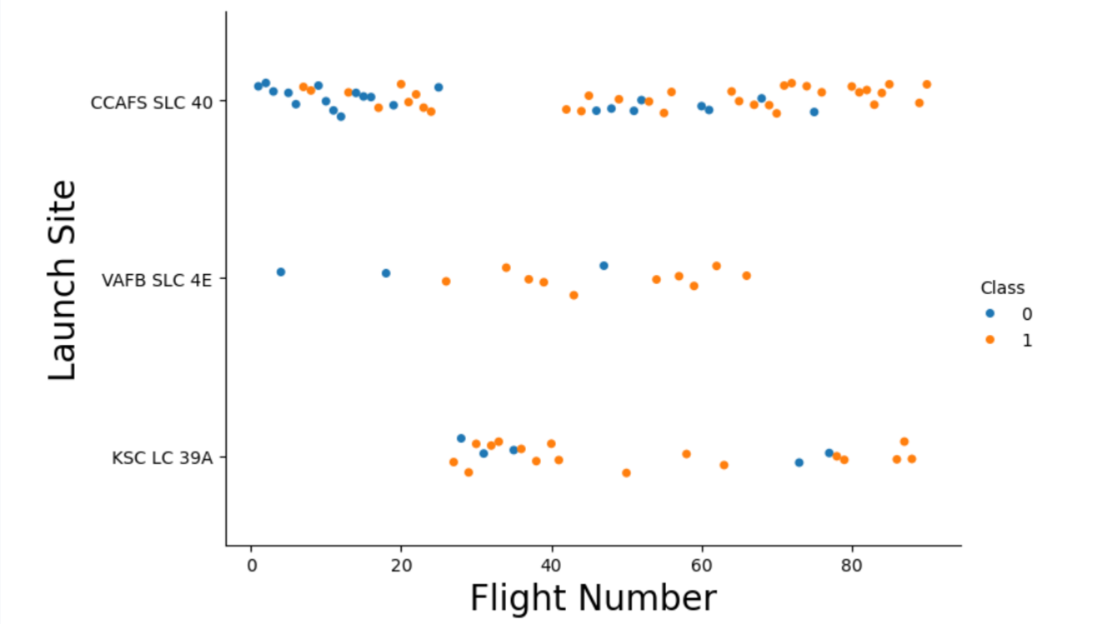
Results

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

Section 2

Insights drawn from EDA

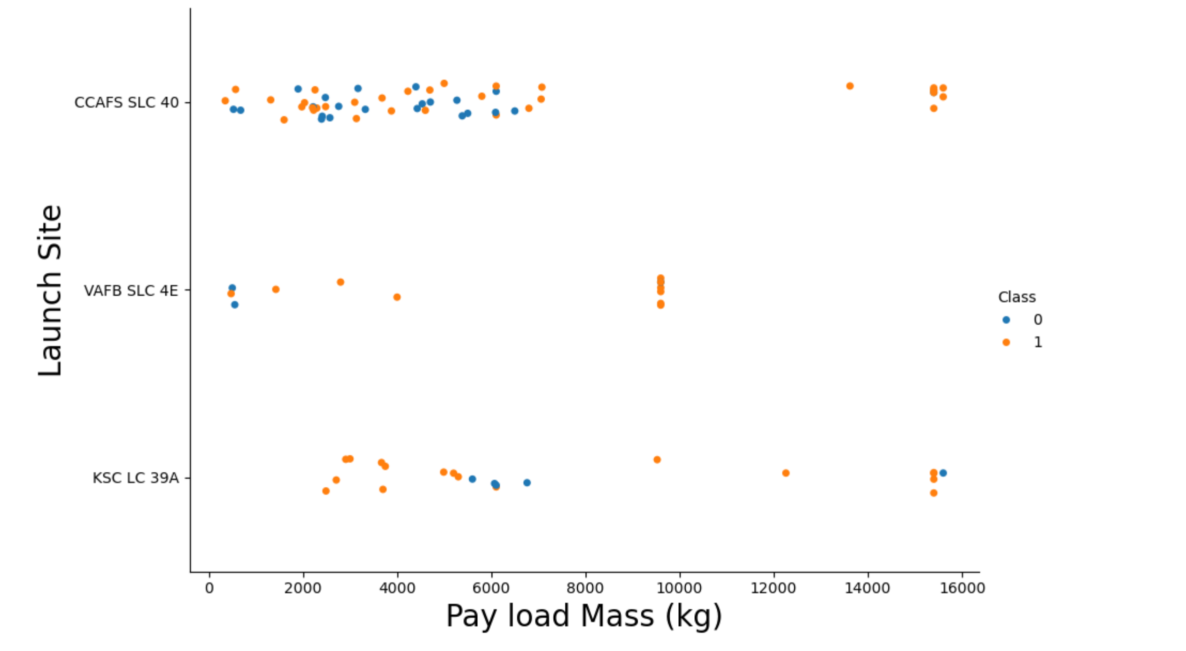
Flight Number vs. Launch Site



A scatter plot with explanations of Flight Number vs. Launch Site

Using the function catplot to plot FlightNumber vs LaunchSite, set the parameter x parameter to FlightNumber,set the y to Launch Site and set the parameter hue to 'class’. We can see that the CCAFS SLC 40 has the highest number of successful lunches as the flight number increases.

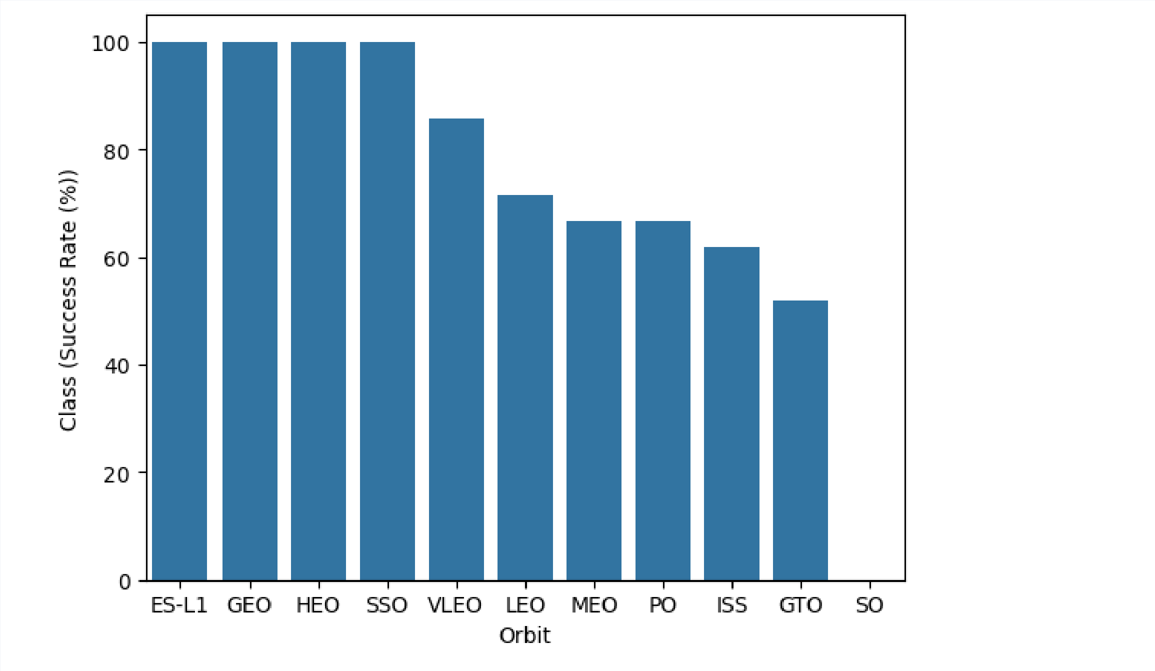
Payload vs. Launch Site



A scatter plot with explanation of Payload vs. Launch Site

if we observe Payload Mass Vs. Launch Site scatter point chart we will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)

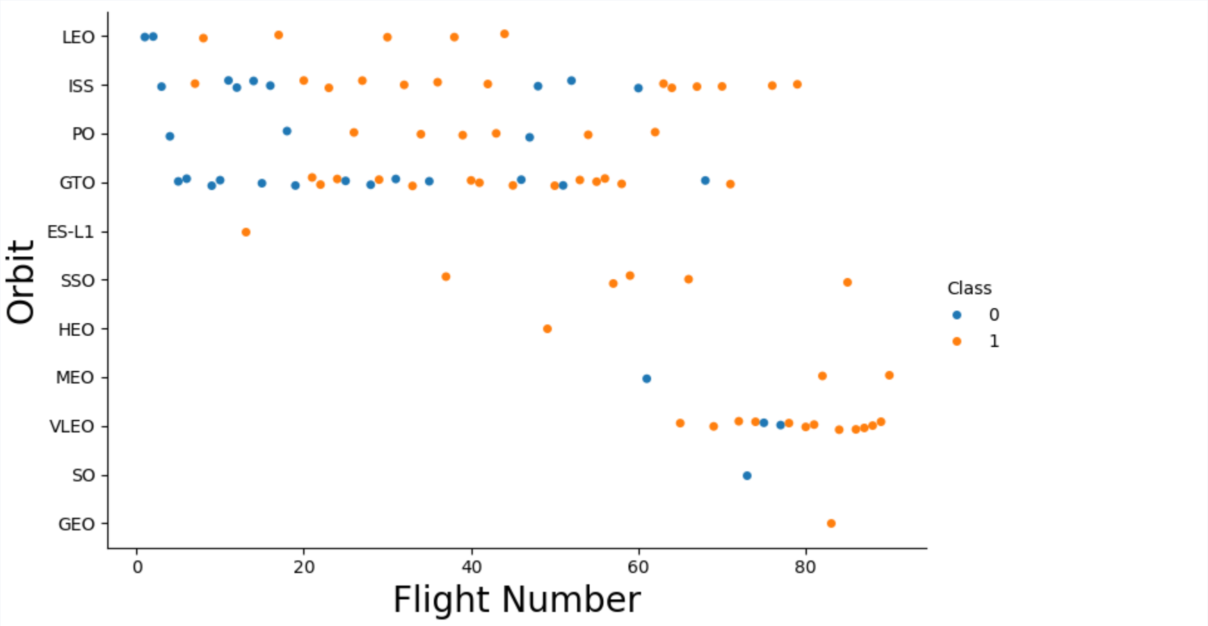
Success Rate vs. Orbit Type



Bar chart with explanations of Success Rate vs. Orbit Type

we can see that ESL1 ,GEO and HEO orbits have the highest success rates and the SO has the Lowest 0 %.

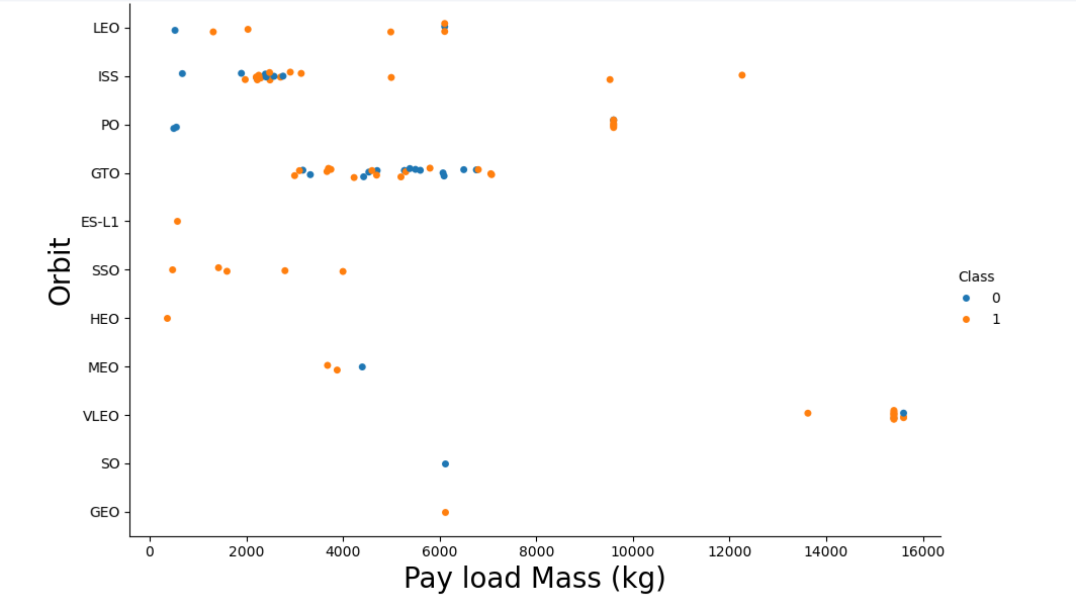
Flight Number vs. Orbit Type



A scatter plot with explanation of Flight Number vs. Orbit Type

For each orbit. we can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

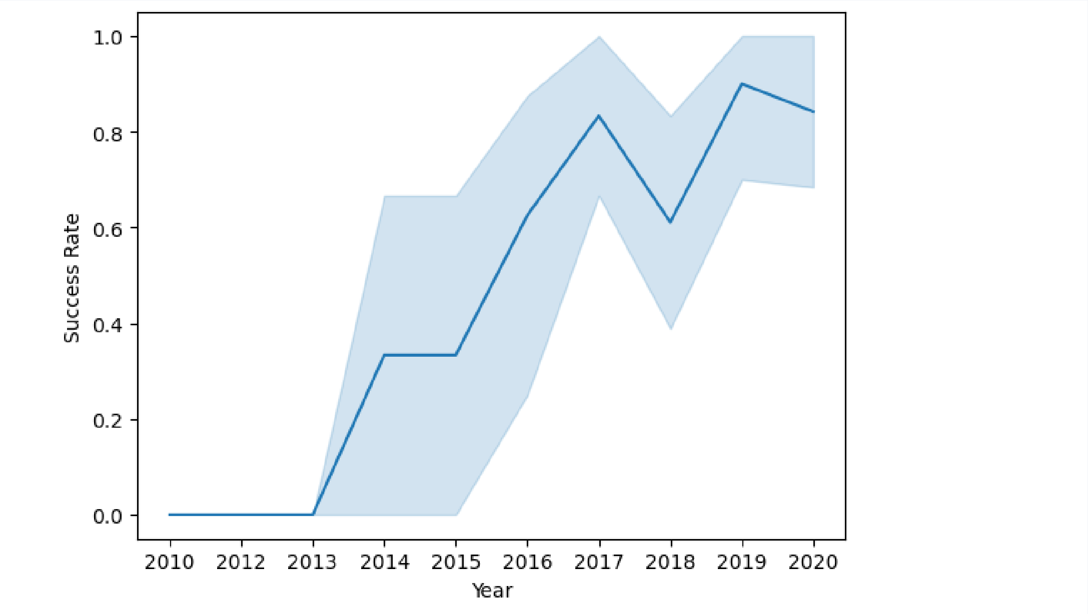
Payload vs. Orbit Type



A scatter plot with explanation of Payload vs. Orbit Type

With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



A line chart Explanation of yearly average success rate

we can plot a line chart with x axis to be Year and y axis to be average success rate, to get the average launch success trend. we can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- Find the names of the unique launch sites

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
```

- Results :

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

we selected all lunch site with applying distinct to escape redundancy

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

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

- Results :

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

We utilized the LIKE operator with the % wildcard in the WHERE clause to retrieve and display all records where the launch site names start with the prefix 'CCA'.

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)';
```

- Results :

Total Payload Mass(Kgs)	Customer
45596	NASA (CRS)

We Applied the SUM() function to calculate and display the total sum of the PAYLOAD_MASS_KG column for the customer 'NASA(CRS)'.

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version LIKE 'F9 v1.1%';
```

- Results :

Payload Mass Kgs	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

We applied the ‘AVG()’ function to return and dispaly the average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)';
```

- Results :

min(DATE)
2015-12-22

Using the minimum function , ‘MIN()’ to return and dispaly the first (oldest) date when first successful landing outcome on ground pad ‘Success (ground pad)occurred .

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql SELECT DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE "Landing_Outcome" = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

- Results :

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

We used the SELECT DISTINCT statement to retrieve and list the unique names of boosters with payloads between 4000 and 6000 kilograms, and a landing outcome of 'Success (drone ship)'.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

- Results :

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

Utilized the COUNT() function in conjunction with the GROUP BY statement to calculate and display the total number of mission outcomes.

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

- Results :

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

We applied a subquery inside the main query to retrieve the maximum payload and applied it to list all boosters that have carried the maximum payload which is 15,600 kg.

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql SELECT SUBSTR(Date,6,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Failure%drone%' AND SUBSTR(Date,0,5) = '2015';
```

- Results :

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

Applied the SUBSTR() function in the SELECT statement to extract the month and year from the Date column as instructed (date,6,2)and (date ,0,5), filtering for records where the year is 2015 and the landing outcome was 'Failure (drone ship)', and returned the matching entries.

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

- 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 Numbers FROM SPACEXTBL WHERE (Landing_Outcome LIKE 'Success%' Or Landing_Outcome LIKE 'Failure%') AND (Date BETWEEN '2010-06-04' AND '2017-03-20') GROUP BY Landing_Outcome ORDER BY Numbers DESC;
```

- Results :

Landing_Outcome	Numbers
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Failure (parachute)	2

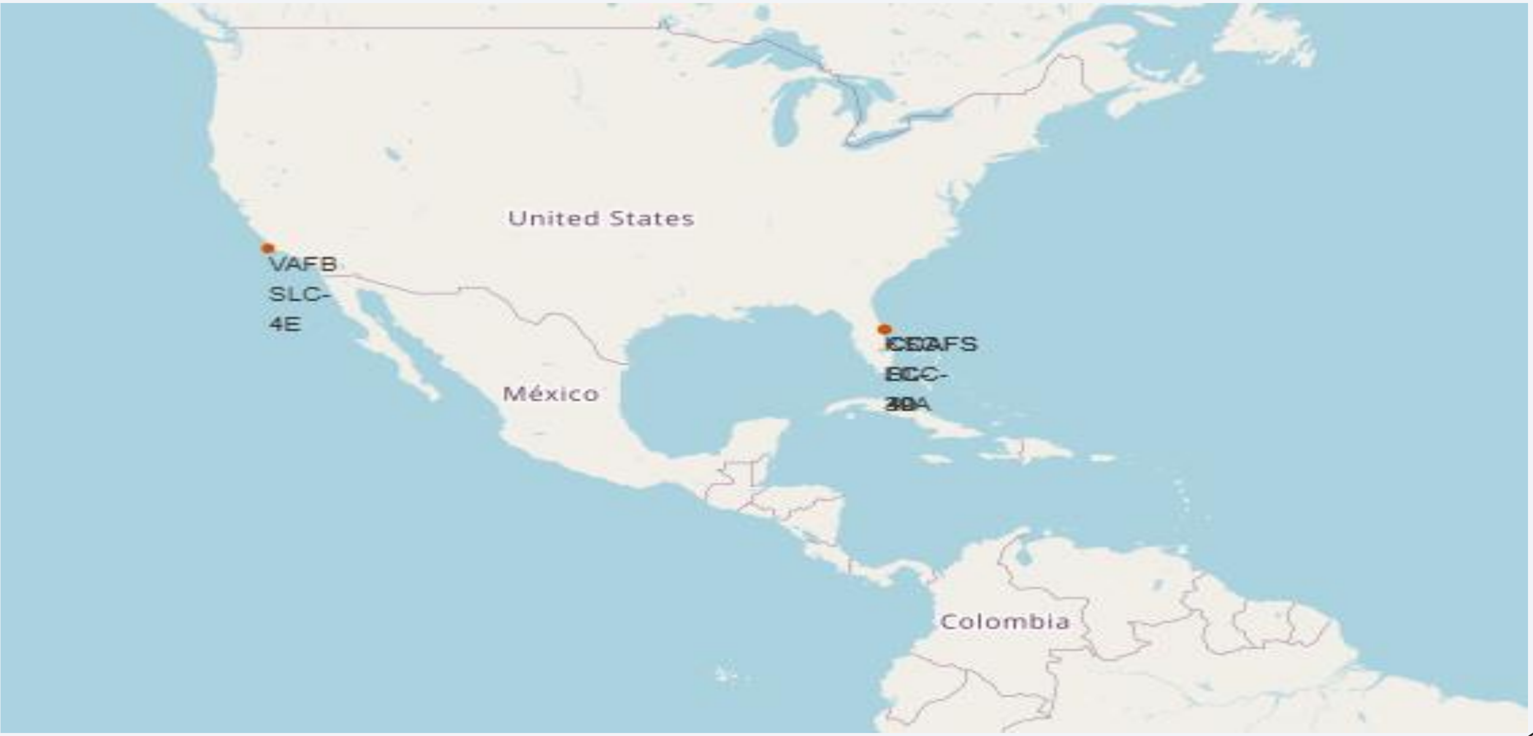
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

Section 3

Launch Sites Proximities Analysis

Plotted markers for all launch sites on a global map

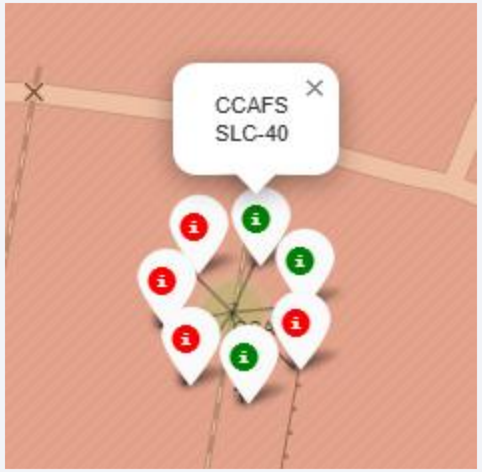
All launch sites are situated near the Equator, positioned toward the southern region of the U.S., and are located in close proximity to coastal areas.

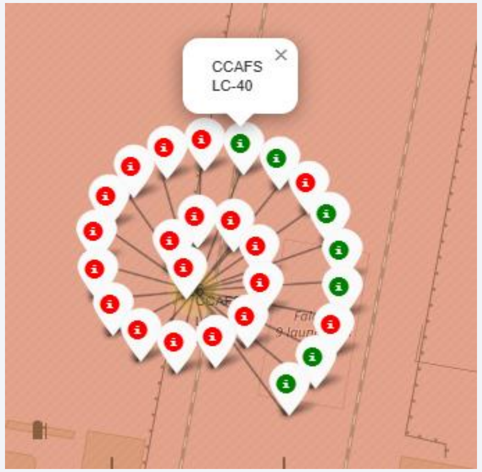


Displayed launch outcomes for each site on the map using color-coded markers

As discussed in the earlier slides, KSC LC-39A has consistently shown the highest launch success rate among all SpaceX launch sites, with other Florida-based sites such as CCAFS LC-40 and CCAFS SLC-40 demonstrating comparatively lower performance including the lunch site (Californai) VAFB SLC-4E that is located in the West Coast . The Folium map presented here visually reinforces this observation, as the launch site with the highest success rate—located in Florida—stands out with a greater number of successful launches compared to the other sites.

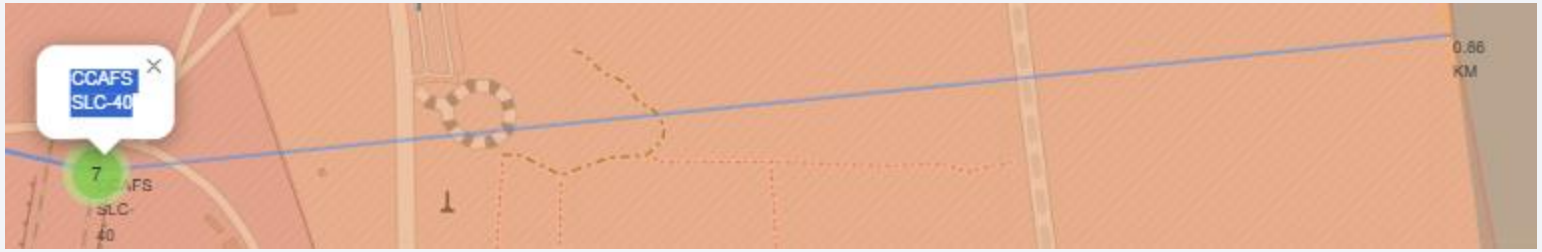
Florida-based sites



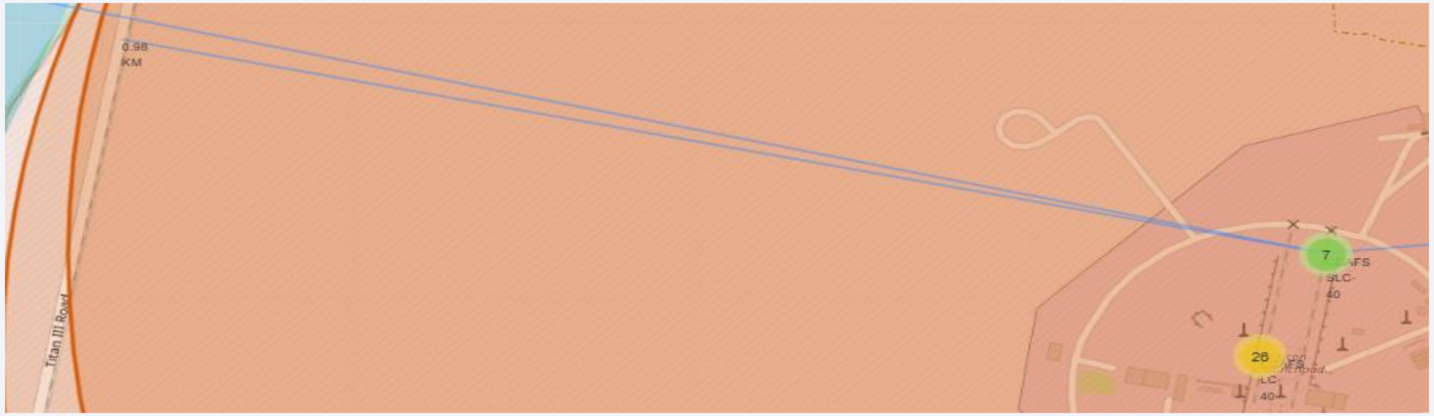


West Coast –based sites

Measured Distances between CCAFS SLC-40 launch site to its proximities



As mentioned previously, the CCAFS SLC-40 launch site is situated very close to the coastline, with a proximity of just under one kilometer(0.86km). This coastal location, like other launch sites, is ideal for safety and logistical reasons, and is clearly visible on the Folium map visualization.

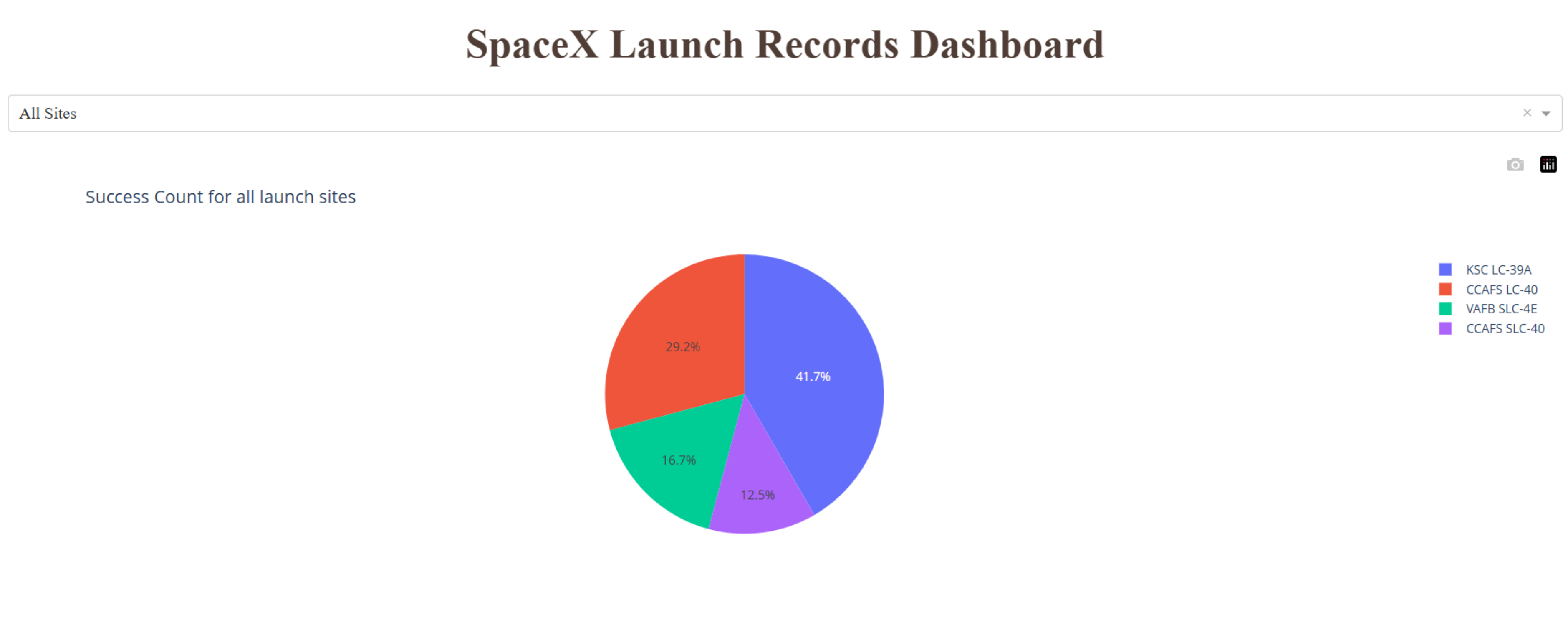


And it’s also located approximately 0.98 km from Titan III Road, highlighting its close accessibility to key infrastructure.

Section 4

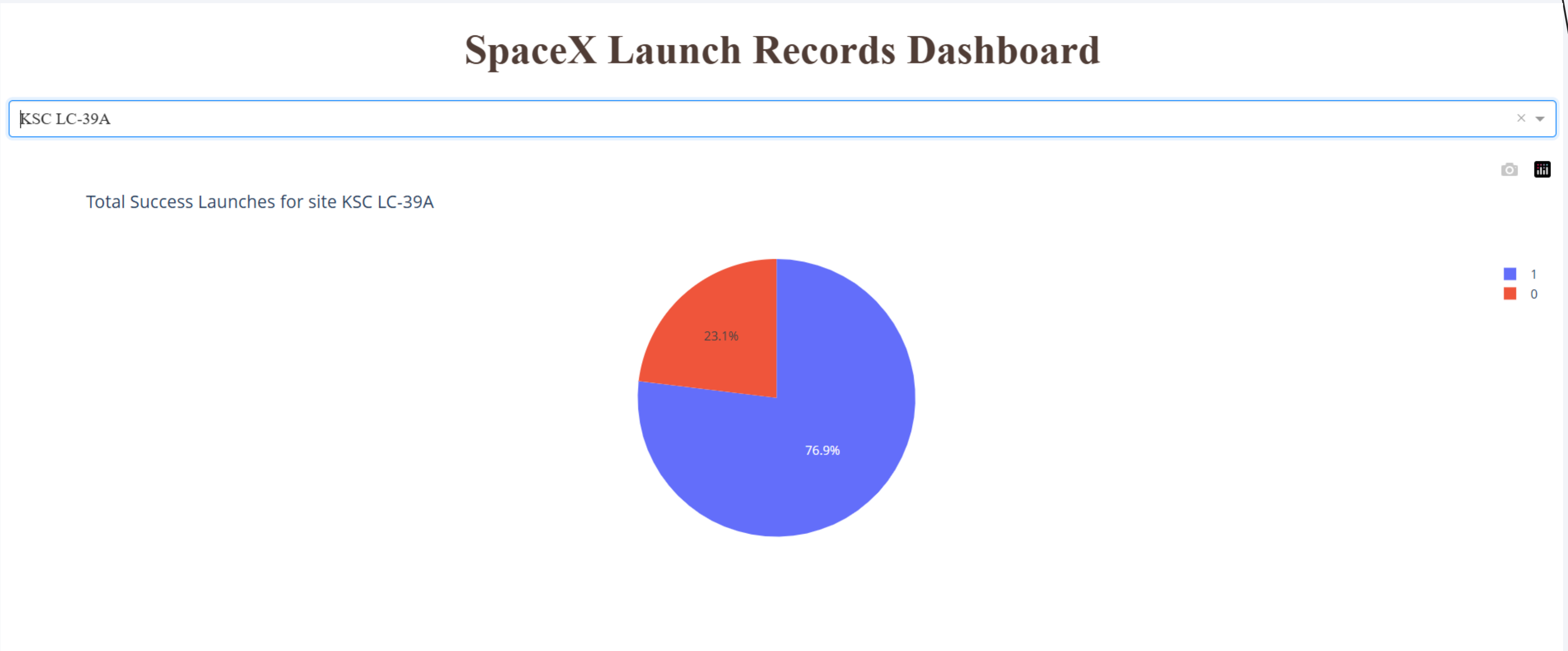
Build a Dashboard with Plotly Dash

A pie chart visualizing the launch success counts across all launch sites.



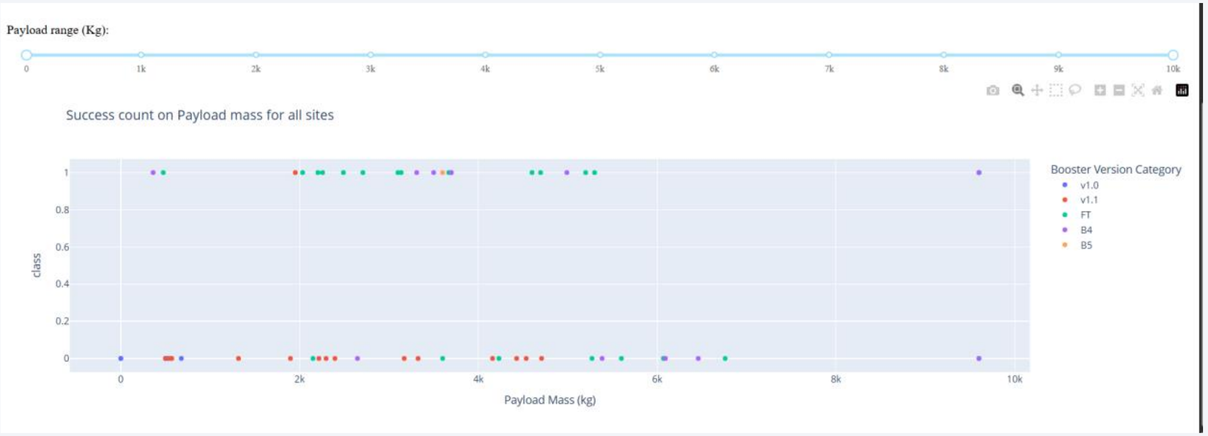
The KSC LC-39A launch site recorded the highest success rate at nearly 42% of all lunches of all sites, followed by CCAFS LC-40 at approximately 29%, VAFB SLC-4E at around 17%, and CCAFS SLC-40 with the lowest success rate of about 12.5%.

A Pie chart for the launch site with highest launch success ratio



Launch site KSC LC-39A had the highest success ratio of almost 77% success against 23% failed launches

Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

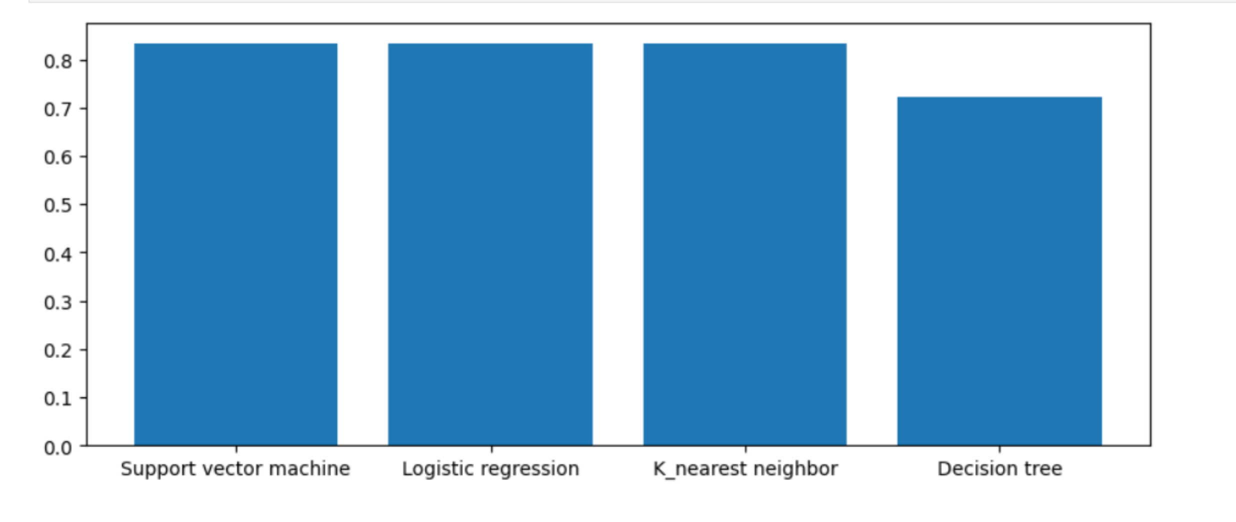


We can notice that the as the payload mass goes lower, the success rate is going higher

Section 5

Predictive Analysis (Classification)

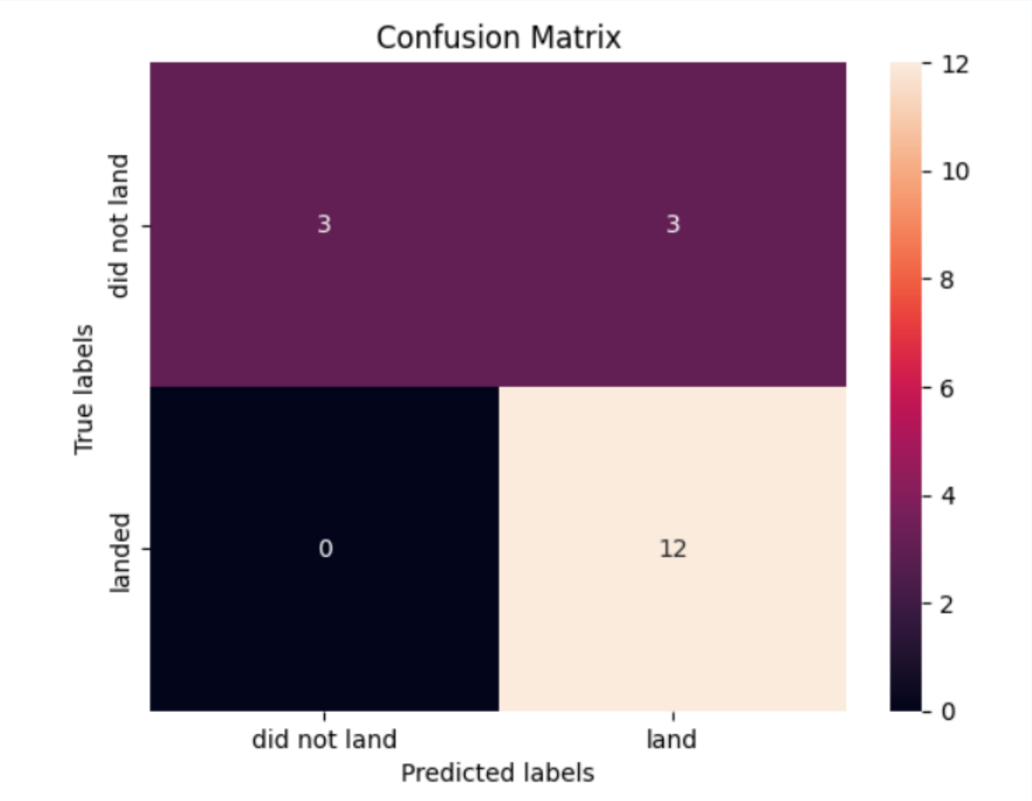
Classification Accuracy



Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.722222
KNN	0.833333

- As we can see, **Logistic Regression**, **SVM**, and **KNN** all achieved the highest accuracy on the test data, tying at **83.3%**. The **Decision Tree model** lagged behind at **72.2%**, likely due to overfitting on the training set or insufficient generalization capacity.

Confusion Matrix



- Examining the confusion matrix which is the same for all three models **Logistic Regression**, **SVM**, and **KNN** where all of them achieved the highest accuracy on the test data. we can can distinguish between the different classes. We see that the problem is false positives.

Conclusions

- As the flight number increases, the first stage is more likely to land successfully. The payload mass also appears to be a factor; even with more massive payloads, the first stage often returns successfully.
- The KSC LC-39A launch site recorded the highest success rate at nearly 42% of all lunches of all sites, followed by CCAFS LC-40 at approximately 29%, VAFB SLC-4E at around 17%, and CCAFS SLC-40 with the lowest success rate of about 12.5%.
- There has been an increase in the success rate since 201 3 kept increasing till 2020
- If we observe Payload Mass Vs. Launch Site scatter point chart we will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)
- For orbit type, SO has the least success rate while ES-L1, GEO, HEO and SSO have the highest success rate According to the yearly trend

Conclusions cont..

- All launch sites are situated near the Equator, positioned toward the southern region of the U.S., and are located in close proximity to coastal areas.
- The CCAFS SLC-40 launch site is situated very close to the coastline, with a proximity of just under one kilometer(0.86km). This coastal location, like other launch sites, is ideal for safety and logistical reasons, and is clearly visible on the Folium map visualization.
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.
- After training all models, we evaluated their performance on the test set and found that Logistic Regression, SVM, and KNN each achieved a test accuracy of 83.3%, outperforming the Decision Tree model, which achieved 72.2%. This process allowed us to identify the most reliable classifiers for this prediction task.

Appendix

All relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, can be found in the following GitHub Repo

https://github.com/MohamadSabha/SpaceX-Falcon-9-first-stage-Landing-Prediction/tree/main

Thank you!