IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nelson Ortega

November 24, 2023

Outline

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

Executive Summary

- The intent of this project is to dissect and understand the cost dynamics of SpaceX's operations.
- By studying the Falcon 9 rocket, cost-saving strategies, particularly its first stage launches through Data Analysis methodologies such as Data collection, Exploratory analysis, Visual analytics and predictive Analysis.
- These observations were vital, as the Falcon 9's first stage is not only the pillar of the rocket but also the key to SpaceX's cost-saving strategy.
- However, it was noticed that the first stage's recovery was not always certain, with occasional crashes or intentional sacrifices, depending on mission-specific factors like payload and orbital requirements.

Introduction

Project Background

There's a significant shift happening in the realm of space travel with several companies aiming to reduce costs and make it more accessible for everyone. Among these, SpaceX has gained particular attention for its noteworthy achievements, which include pioneering reusable rocket technology to cut down launch costs. As the space industry continues to evolve, it's crucial to understand and emulate such cost-saving measures to stay competitive.

Problem Statement

This project is centered around a key challenge: determining whether the first stage of a rocket, which is a major cost factor, can be landed and reused, and how this impacts the overall cost of space missions. The role within this project involves analyzing data from previous SpaceX launches to identify patterns and factors that contribute to successful first-stage recovery. By applying machine learning techniques to publicly available data, the aim is to predict the reusability of rocket components, which is instrumental in forecasting launch costs and ensuring market competitiveness.

Methodology

Methodology

- Executive Summary
- successful landing of Falcon 9's first stage.
- Data collection methodology:
  - Data in the project was collected primarily through the SpaceX REST API, which provided extensive details about SpaceX launches, such as rocket and payload specifics, launch and landing details, and outcomes.
- Perform data wrangling
  - Data wrangling was performed, focusing on various attributes of space launches like Flight Number, Launch Site, and Outcome, with explanations on specific columns such as LaunchSite, Orbits, and Outcome, which were used to classify the success of booster landings.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Construction of a machine learning pipeline, encompassing data preprocessing, splitting, model training with grid search for hyperparameter optimization, testing various algorithms, and evaluating performance using a confusion matrix to predict the

Data Collection

- **Here are the most important steps, enumerated from 1 to 6, based on the SpaceX launch data collection using an API:**

- **Accessing the SpaceX REST API**: Utilize the endpoint api.spacexdata.com/v4/launches/past to target specific past launch data.

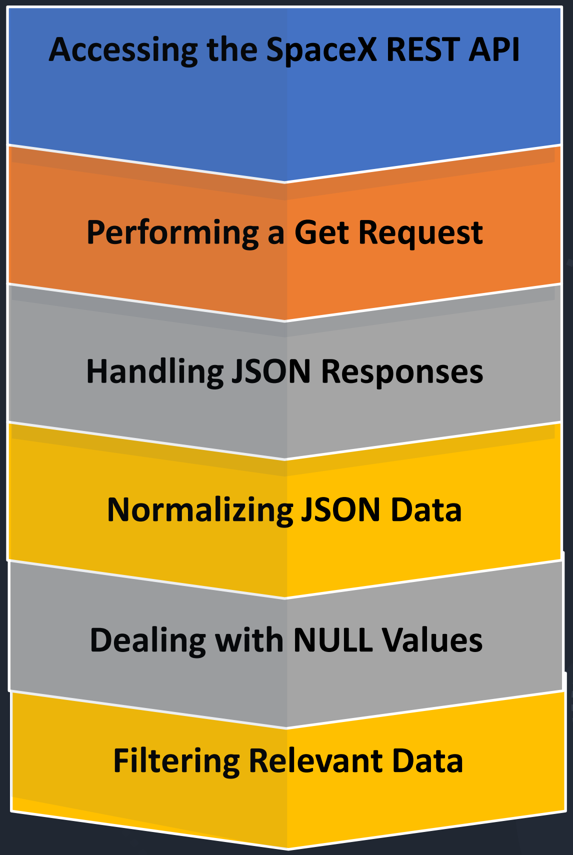
- **Performing a Get Request:** Use the requests library to perform a get request to obtain the launch data.

- **Handling JSON Responses:** Process the response from the API, which will be in the form of a list of JSON objects, each representing a launch.

- **Normalizing JSON Data:** Use the json_normalize function to convert the structured JSON data into a flat table (dataframe).

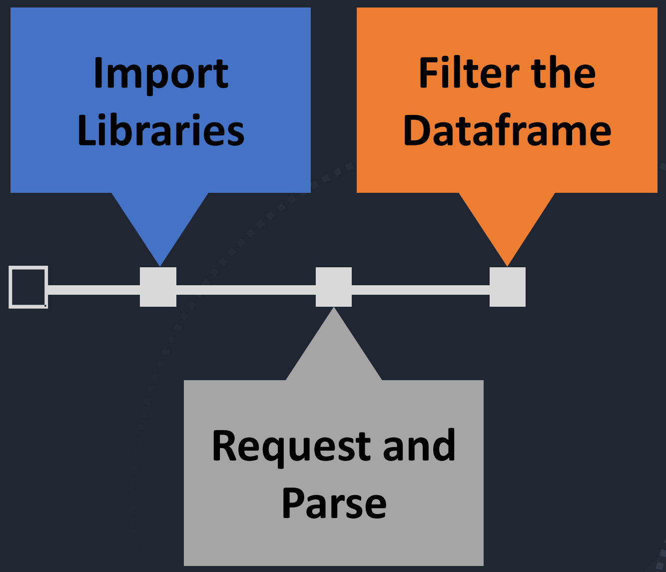
- **Dealing with NULL Values:** Develop a method to handle NULL values, such as calculating the mean for missing PayloadMass data and replacing NULL values with this mean.

- **Filtering Relevant Data:** Filter out irrelevant data, like removing Falcon 1 launch data to focus only on Falcon 9 launches and prepare the dataset for analysis.



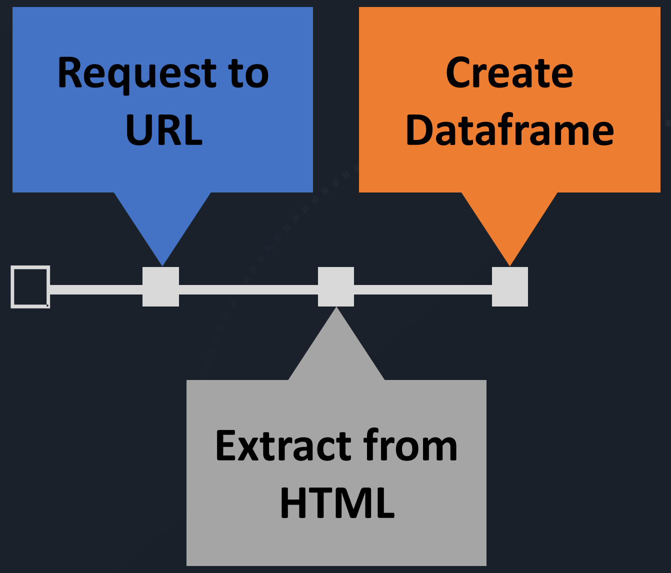
Data Collection – SpaceX API

- **Import Libraries and Define Auxiliary Functions**
  - Use several columns from the dataset to call the API and append the data to the list
  - Request rocket launch data from SpaceX API with the URL
- **Request and parse the SpaceX launch data using the GET request**
  - Decode the response content as a Json using .json() and turned to Pandas dataframe using .json_normalize()
  - Data from these requests was stored in lists and used to create a new dataframe.
- **Filter the dataframe to only include Falcon**
  - Filter the data dataframe using the BoosterVersion column.
  - Reset the FlgihtNumber column
- GitHub URL: https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- **Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia**
  - **Request the Falcon9 Launch Wiki page from its URL**
  - **Extract all column/variable names from the HTML table header**
  - **Create a data frame by parsing the launch HTML tables**
- **GitHub URL: https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb**



Data Wrangling

In this stage we will be transforming data from a "raw" form into another format to making it more appropriate for downstream purposes:

- **Identify and calculate the percentage of the missing values in each attribute**
- **Calculate the number of launches on each site**
- **Calculate the number and occurrence of each orbit**
- **Calculate the number and occurrence of mission outcome of the orbits**
- **Create a landing outcome label from Outcome column**
- **Determine the success rate: 66%**

GitHub [URL:https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb](https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb)

1

```
FlightNumber      0.000000
Date              0.000000
BoosterVersion    0.000000
PayloadMass       0.000000
Orbit             0.000000
LaunchSite        0.000000
Outcome           0.000000
Flights           0.000000
GridFins          0.000000
Reused            0.000000
Legs              0.000000
LandingPad       28.888889
Block             0.000000
ReusedCount       0.000000
Serial            0.000000
Longitude         0.000000
Latitude          0.000000
```

2

```
LaunchSite
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
```

3

```
Orbit
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
GEO       1
HEO       1
SO        1
```

4

0	0
1	0
2	0
3	0
4	0
...	...
85	1
86	1
87	1
88	1
89	1

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

5

EDA with Data Visualization

- Data Visualization was performed by generating Scatter plots, bar charts, and line charts given that to perform exploratory Data Analysis and Feature Engineering using Pandas and Matplotlib.
  - Scatter Plots: These are ideal for visualizing the relationship between two numerical variables.
  - Bar Charts: Bar charts are used to compare different groups or to track changes over time. In this case we used the x-axis represents orbits vs success rate in the y-axis.
  - Line Charts: Line charts are excellent for illustrating trends over time. In the following Line Chart the success rate since 2013 kept increasing till 2020 you can observe that the success rate since 2013 kept increasing till 2020
- GitHub: [URL:https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(2).ipynb](https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(2).ipynb)

EDA with SQL

Executed SQL Queries:

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

%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE¶

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

%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 20

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

%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'

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

%sql SELECT AVG(PAYLOAD_MASS__KG_)FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1’

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

%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)’

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 FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000

List the total number of successful and failure mission outcomes

%sql SELECT Mission_Outcome, COUNT(*) FROM SPACEXTABLE GROUP BY Mission_Outcome

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

%sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)

EDA with SQL (Continued)

- Executed SQL Queries:

- **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 **
  - **strftime('%m', Date) AS Month_Number,**
  - **CASE strftime('%m', Date)**
  - **WHEN '01' THEN 'January'**
  - **WHEN '02' THEN 'February'**
  - **WHEN '03' THEN 'March'**
  - **WHEN '04' THEN 'April'**
  - **WHEN '05' THEN 'May'**
  - **WHEN '06' THEN 'June'**
  - **WHEN '07' THEN 'July'**
  - **WHEN '08' THEN 'August'**
  - **WHEN '09' THEN 'September'**
  - **WHEN '10' THEN 'October'**
  - **WHEN '11' THEN 'November'**
  - **WHEN '12' THEN 'December'**
  - **END AS Month_Name,**
  - **Landing_Outcome,**
  - **Booster_Version,**
  - **Launch_Site**
  - **FROM SPACEXTABLE**
  - **WHERE substr(Date, 0, 5) = '2015' AND Landing_Outcome = 'Failure (drone ship)';**

EDA with SQL (Continued)

Executed SQL Queries:

- **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 Count_Landing_Outcome FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Count_Landing_Outcome DESC**
- GitHub: [URL:https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(2).ipynb](https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(2).ipynb)

Build an Interactive Map with Folium

- Folium was used to build an interactive map, allowing for dynamic data exploration and visualization through features like zooming, panning, filtering, and linking. The study included analyzing launch site geographies with Folium to gain deeper insights from the SpaceX dataset.
- The analysis included the launch site geo and proximities. The launch sites locations were marked with their close proximities on an interactive map. Then, we explored the map with those markers and try to discover any patterns from them. Finally, we were able to explain how to choose an optimal launch site.
- GitHub URL: https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- A dashboard application was built with the Python Plotly Dash package. This dashboard application contained input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart.
- The dashboard was built to find more insights from the SpaceX dataset more easily than with static graphs.
- GitHub URL: https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/spacex_dash_app%20(1).py

Predictive Analysis (Classification)

- In this Predictive Analysis, we built a machine learning pipeline to predict if the first stage of the Falcon 9 lands successfully.
- This included:
  - Preprocessing, allowing us to standardize our data, and Train_test_split, allowing us to split our data into training and testing data,
  - Train the model and perform Grid Search, allowing us to find the hyperparameters that allow a given algorithm to perform best.
  - Use the best hyperparameter values, we determine the model with the best accuracy using the training data.
  - The following algorithms were used: Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors.
  - Finally, we generate a confusion matrix visualize the performance of an algorithm
  - GitHub URL: https://github.com/kuriux1/IBM-Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite%20(1).ipynb

Results

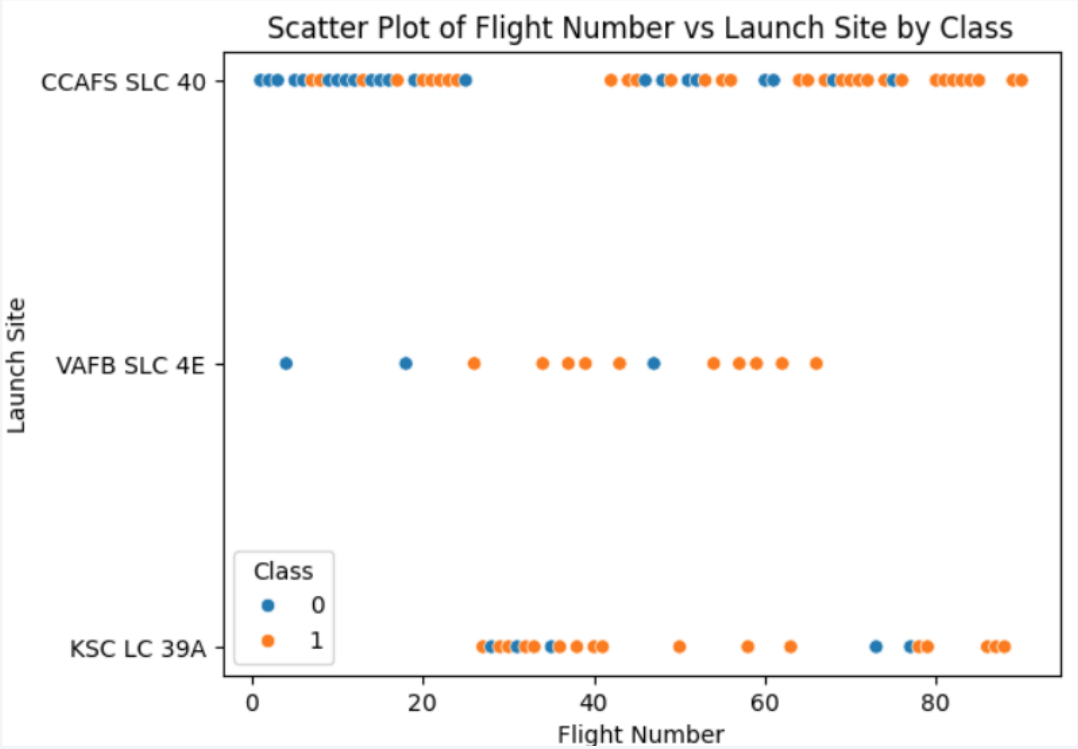
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

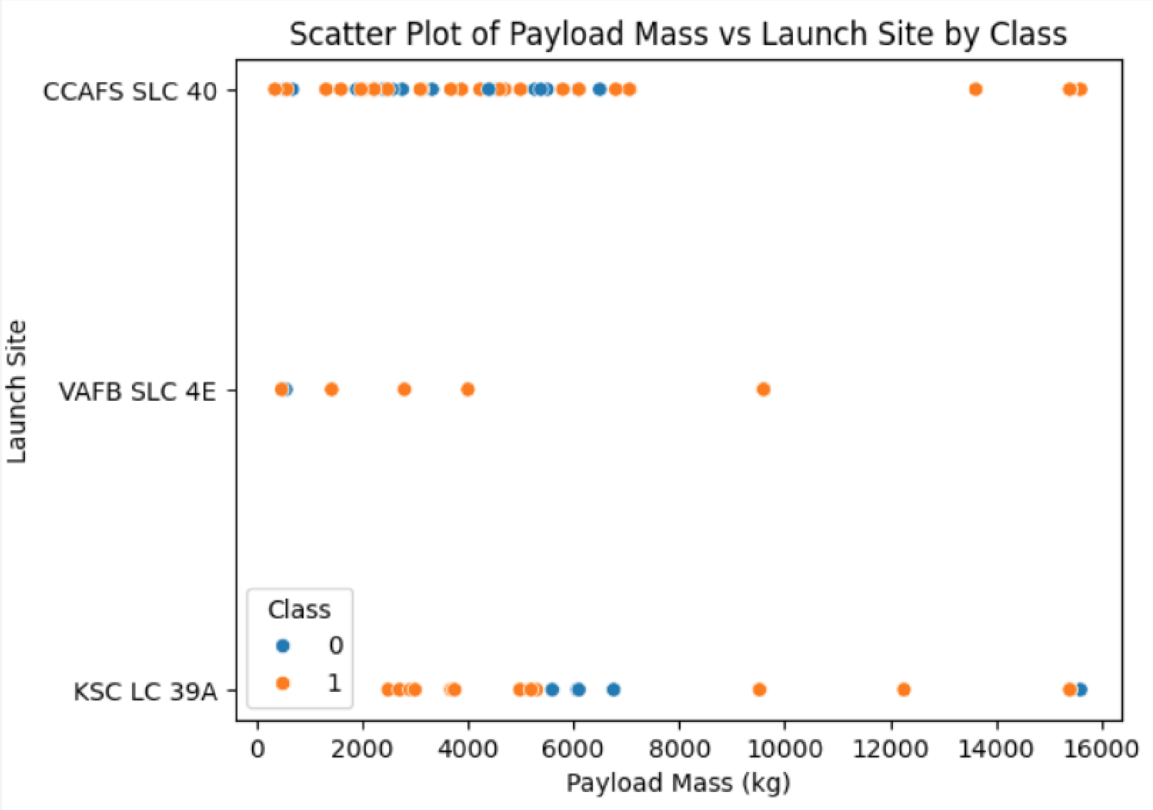
1. **Launch Outcomes by Site:**
   1. **CCAFS SLC 40** shows a large number of both successful and unsuccessful launches.
   2. **VAFB SLC 4E** has a smaller number of launches overall, with a mix of successes and failures.
   3. **KSC LC 39A** has many successful launches, especially in higher flight numbers, and a few failures.
2. **Trends**:
   1. There seems to be a trend of increasing success with higher flight numbers, particularly at KSC LC 39A.
   2. The distribution of successes and failures at VAFB SLC 4E appears more sporadic due to fewer launches.
3. **Success Rate Over Time**:
   1. As the flight number increases, the proportion of successful launches (Class 1) appears to increase, which could indicate improved reliability and processes over time.



Payload vs. Launch Site

Key observations :

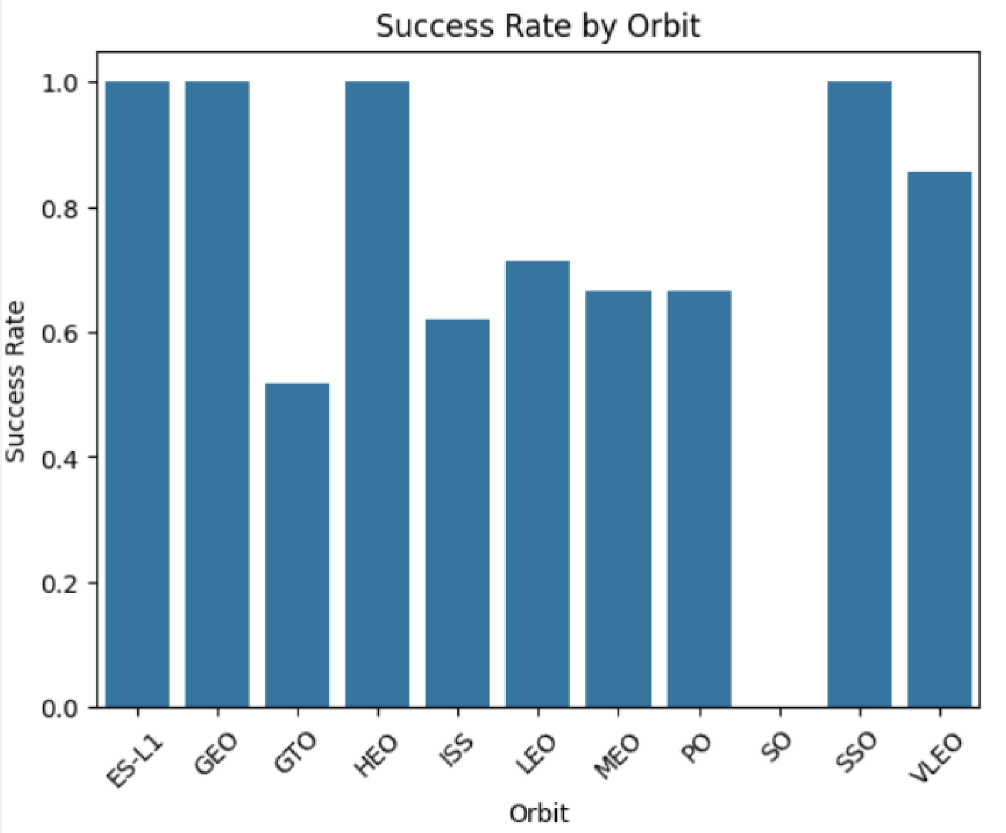
- Launches are grouped by the site, with three distinct launch sites displayed: CCAFS SLC 40, VAFB SLC 4E, and KSC LC 39A.
- A higher density of successful launches (Class 1) is apparent across varying payload masses, suggesting a general trend of success regardless of the mass of the payload for these sites.
- VAFB SLC 4E has fewer data points, indicating fewer launches from this site in the dataset, but maintains a high success rate across the range of payload masses represented.
- The payload mass for launches spans from light payloads close to 0 kg to heavy payloads over 15,000 kg.
- The KSC LC 39A site shows successful launches across the widest range of payload masses, including the heaviest payloads.
- There's a visible trend that as payload mass increases, the number of launches (data points) decreases, which might suggest heavier payloads are less common.
- Unsuccessful launches (Class 0) are fewer overall and occur across a range of payload masses, with no clear pattern indicating a higher likelihood of failure with increasing payload mass.
- Overall, the plot suggests that all three launch sites have a generally high success rate across a wide range of payload masses, with no immediate indication that payload mass significantly affects launch outcome based on the available data.



Success Rate vs. Orbit Type

Key observations:

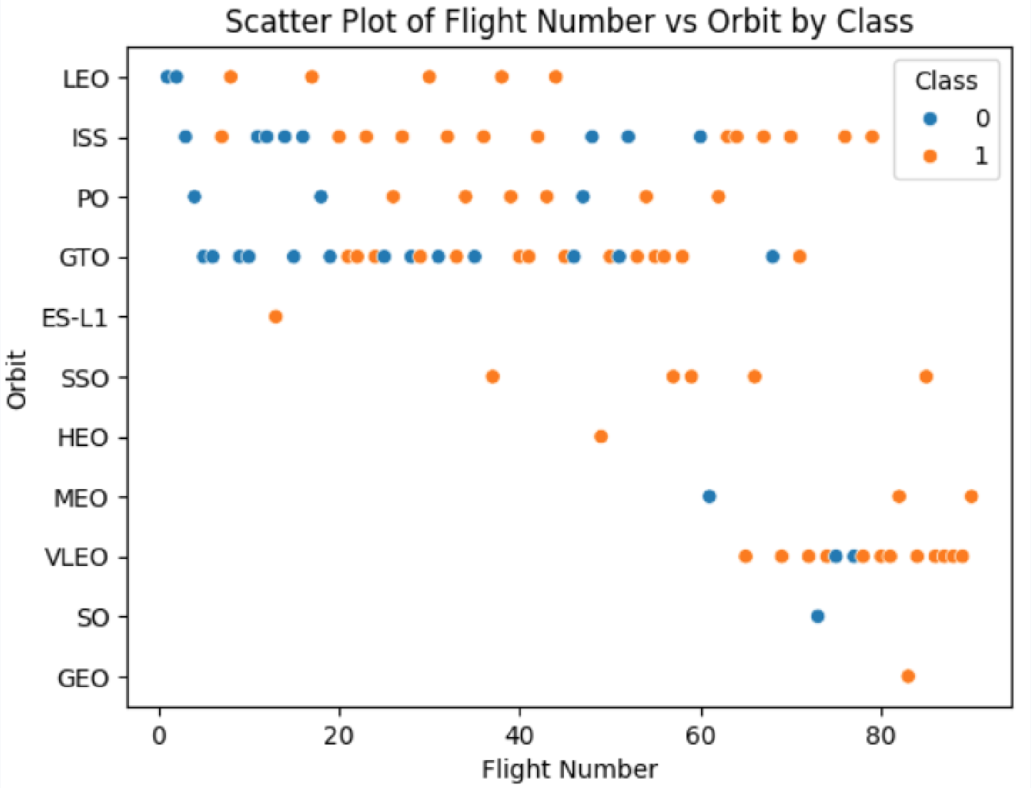
- Orbits like ES-L1, GEO, and VLEO show the highest success rates, close to 100%. In contrast, GTO (Geostationary Transfer Orbit) demonstrates a notably lower success rate compared to the others.
- The intermediate success rates are seen for orbits such as HEO (Highly Elliptical Orbit), ISS (International Space Station orbit), LEO (Low Earth Orbit), MEO (Medium Earth Orbit), PO (Polar Orbit), SO (Sun-synchronous Orbit), and SSO (Sun-Synchronous Orbit).
- Overall, the chart suggests that some orbits have historically been more challenging with respect to successful rocket launches than others.



Flight Number vs. Orbit Type

Key observations:

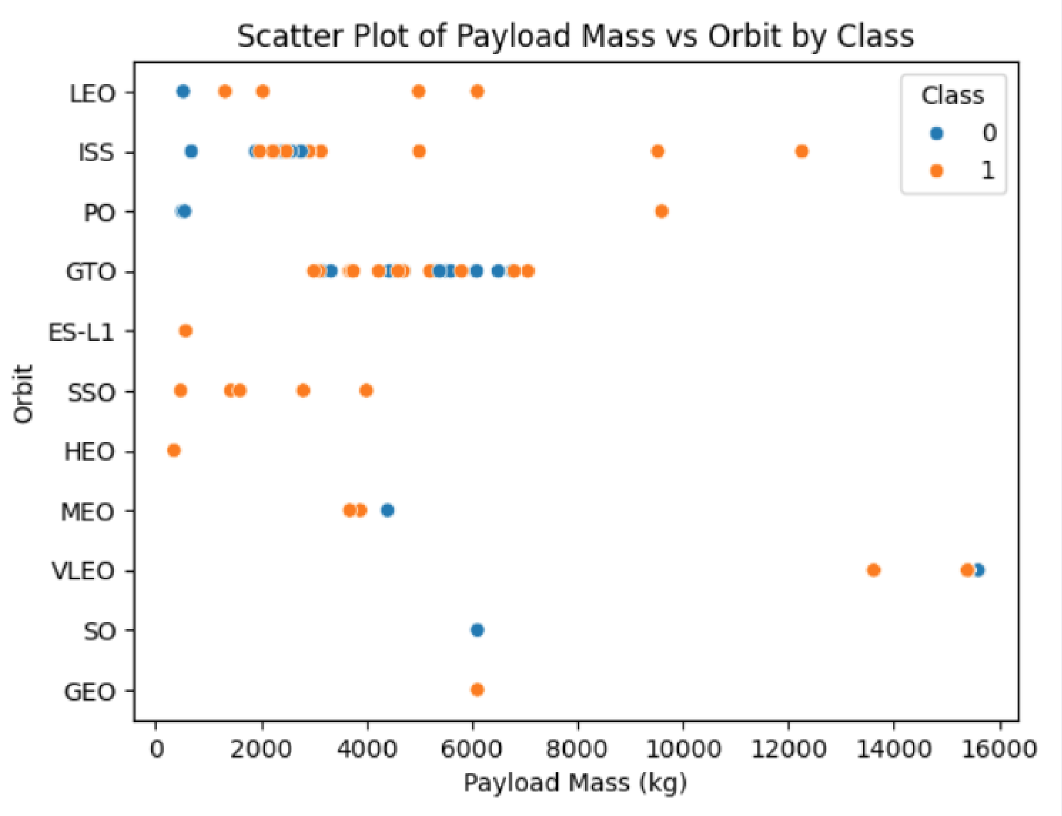
1. **Flight Distribution**: There's a spread of flight numbers across various orbits, suggesting a range of missions over time.
2. **Success and Failure Trends**: The plot shows both successful and unsuccessful flights across different orbits. A visual trend might suggest whether success rates improve with higher flight numbers.
3. **Orbit Specifics**:
    1. **LEO (Low Earth Orbit)** and **ISS (International Space Station)** trajectories seem to have a high number of flights with a mix of successes and failures.
    2. **GTO (Geostationary Transfer Orbit)** flights are also quite frequent but show a significant number of unsuccessful attempts, especially in earlier flights.
    3. **ES-L1**, **SSO (Sun-Synchronous Orbit)**, and **HEO (Highly Elliptical Orbit)** have fewer data points, indicating fewer flights to these orbits.
4. **Learning Over Time**: If there's a higher concentration of orange dots (successes) in flights with higher numbers, it may suggest that the success rate has improved as the flight number increases, potentially indicating learning and improvements in technology or processes.



Payload vs. Orbit Type

Key observations :

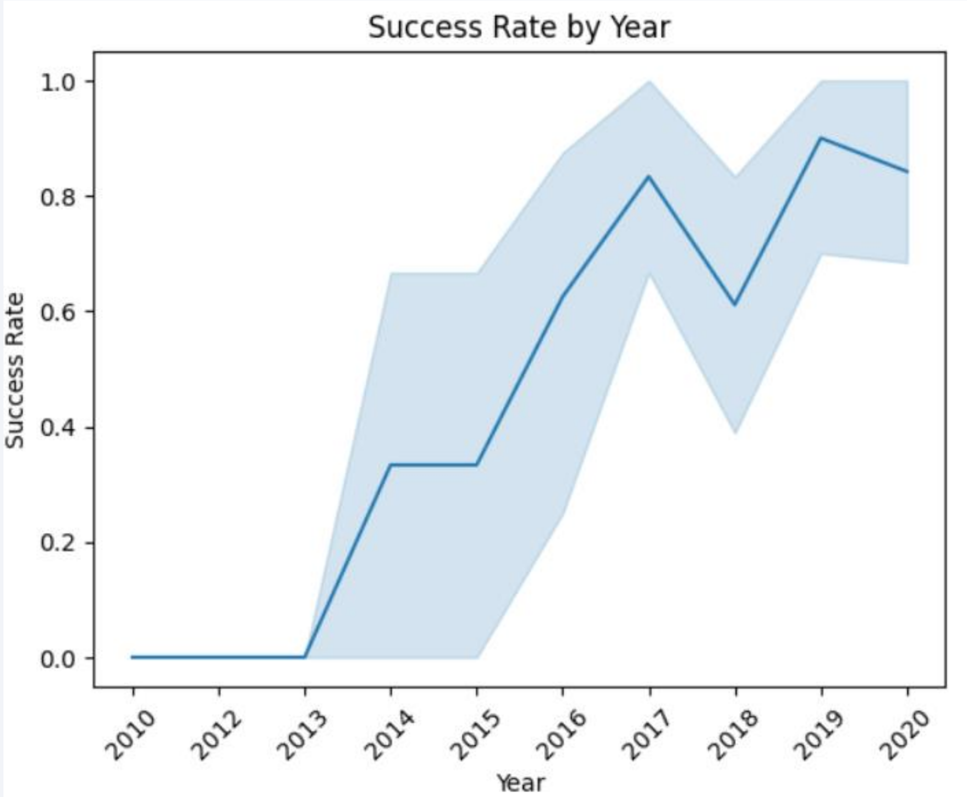
- The majority of launches, particularly to the ISS and LEO, are successful regardless of payload mass.
- GTO orbits have a mixture of success and failures, with no clear correlation between payload mass and launch outcome.
- Higher payload masses (above 10,000 kg) have been successfully placed into GTO and VLEO orbits.
- There are relatively few launches to ES-L1, SSO, HEO, MEO, VLEO, and GEO orbits, but these show high success rates, except for a single failure to MEO.



Launch Success Yearly Trend

Key observations :

- The success rate starts below 0.2 in 2010 and demonstrates a generally upward trend over the years, with some fluctuations.
- Notably, the success rate improves significantly from 2010 to 2013, stabilizes with some variance until 2016, then peaks and dips slightly before reaching its highest point in 2019, followed by a small decline in 2020.
- The overall trend suggests increased efficiency or effectiveness of the processes or technologies involved, with the best performance near the end of the decade.



All Launch Site Names (SQL Queries)

- Find the names of the unique launch sites
- Present your query result with a short explanation here

```
#%sql SELECT * FROM SPACEXTABLE LIMIT 5
%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

- Total payload carried by boosters from NASA

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.

```
%sql SELECT AVG(PAYLOAD_MASS__KG_)FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

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

```
%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

- 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 FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000
```

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- 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

- Booster which have carried the maximum payload mas

```
%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

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql SELECT \
 strftime('%m', Date) AS Month_Number,\
 CASE strftime('%m', Date)\
   WHEN '01' THEN 'January'\
   WHEN '02' THEN 'February'\
   WHEN '03' THEN 'March'\
   WHEN '04' THEN 'April'\
   WHEN '05' THEN 'May'\
   WHEN '06' THEN 'June'\
   WHEN '07' THEN 'July'\
   WHEN '08' THEN 'August'\
   WHEN '09' THEN 'September'\
   WHEN '10' THEN 'October'\
   WHEN '11' THEN 'November'\
   WHEN '12' THEN 'December'\
 END AS Month_Name,\
 Landing_Outcome,\
 Booster_Version,\
 Launch_Site\
FROM SPACEXTABLE\
WHERE substr(Date, 0, 5) = '2015' AND Landing_Outcome = 'Failure (drone shi
```

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

Month_Number	Month_Name	Landing_Outcome	Booster_Version	Launch_Site
01	January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

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

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 Count_Landing_Outcome FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Count_Landing_Outcome DESC
```

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

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

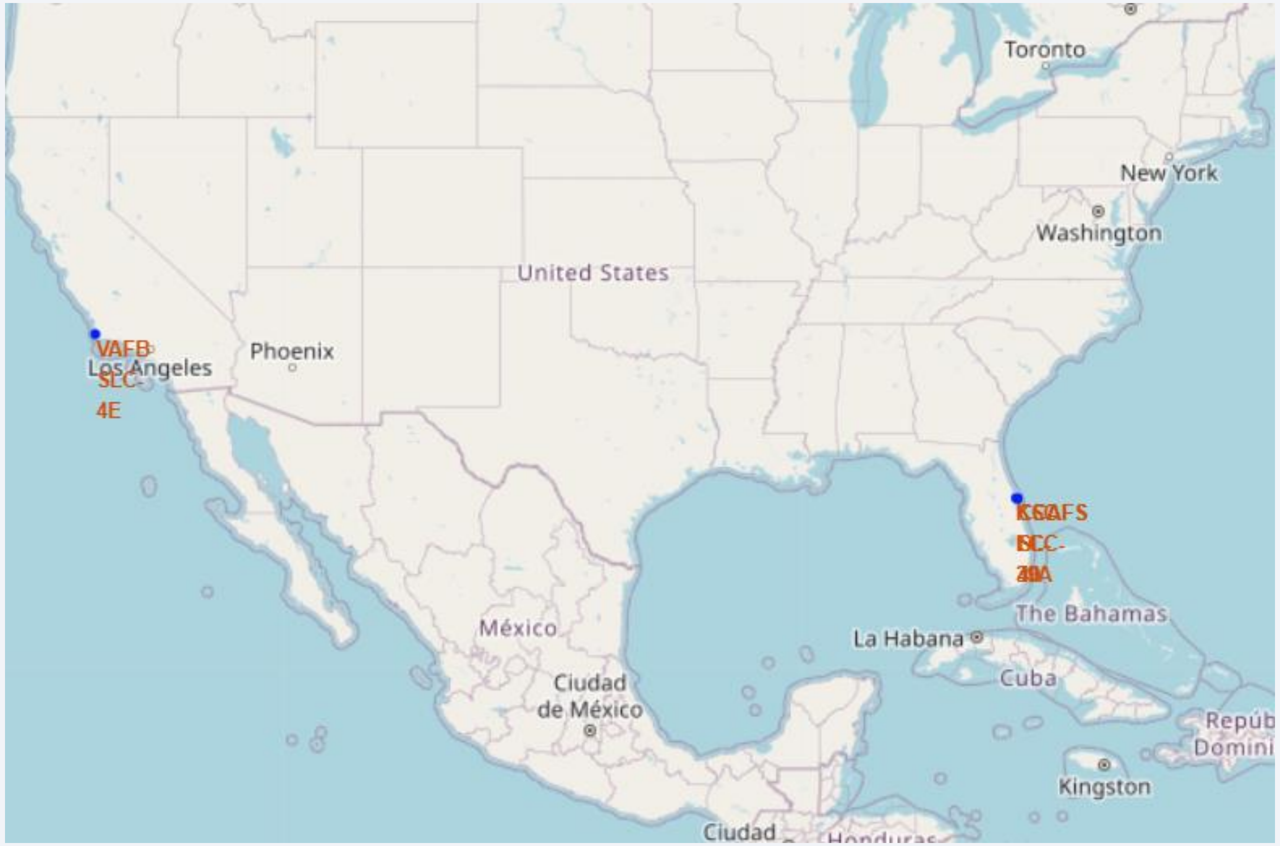
SpaceX Launching Sites

SpaceX Launching Sites are located near coastal areas. Some of the reasons are:

Safety: Coastal locations allow for launch trajectories over the ocean, which minimizes the risk to populated areas in the event of a launch failure.

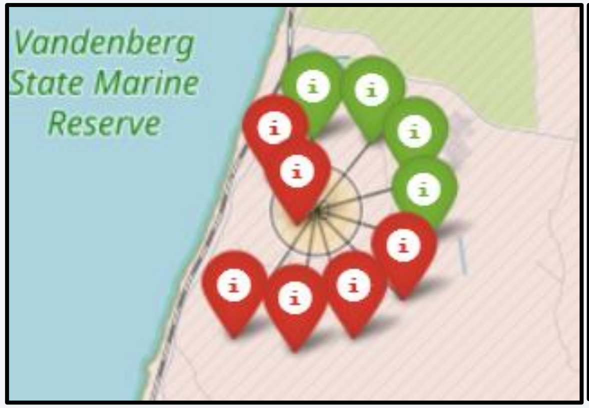
Security: Coastal sites can be more easily secured against unauthorized access, and provide clear zones that can be monitored for ship and air traffic during launches.

Environmental and Economic Impact: The environmental impact of launches can be better managed by directing potential hazards away from land, and the economic impact of a failed launch is less on water than on populated land areas.



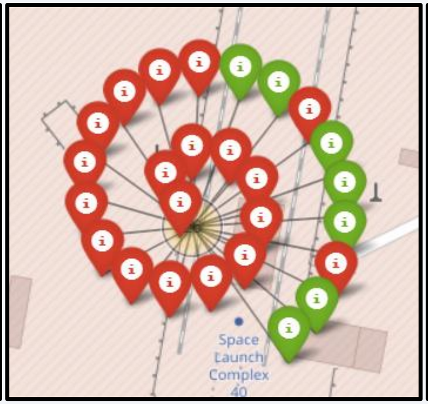
Launch Sites Clusters Successes or Failed

- In the diagrams below are shown the different sites use by space X rocket launches. Green icons represents successful launches and red icons represents failed launches.
- SpaceX's launch success rate varies by site, with some locations potentially offering more favorable conditions or having more advanced infrastructure, but overall, their demonstrated success across different sites reflects the company's ability to adapt to the unique challenges each launch site presents.

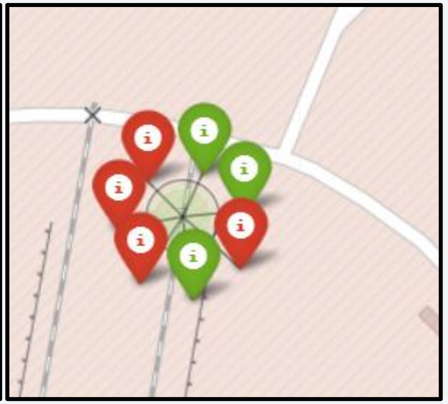


VAFB SLC-4E

KSC LC-39A



CCAFS LC-40



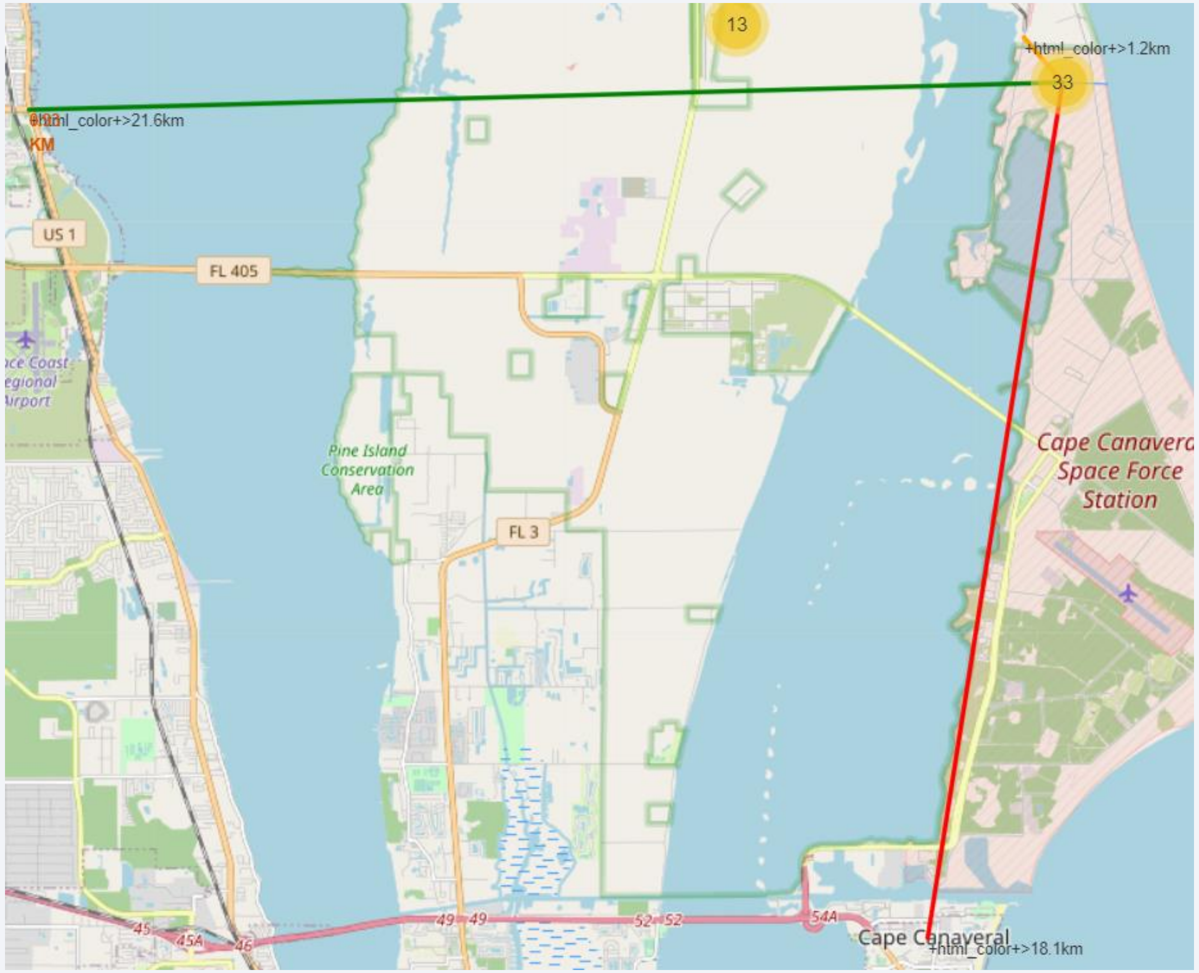
CCAFS SLC-40

Launch Site Proximities

Distances:

- City Distance (Cape Canaveral): 18.0 km
- Railway Distance (Northwest): 1.23 km
- Highway Distance (West): 21.5 km
- Coastline Distance (East): 0.93 km
- Explain the important elements and findings on the screenshot

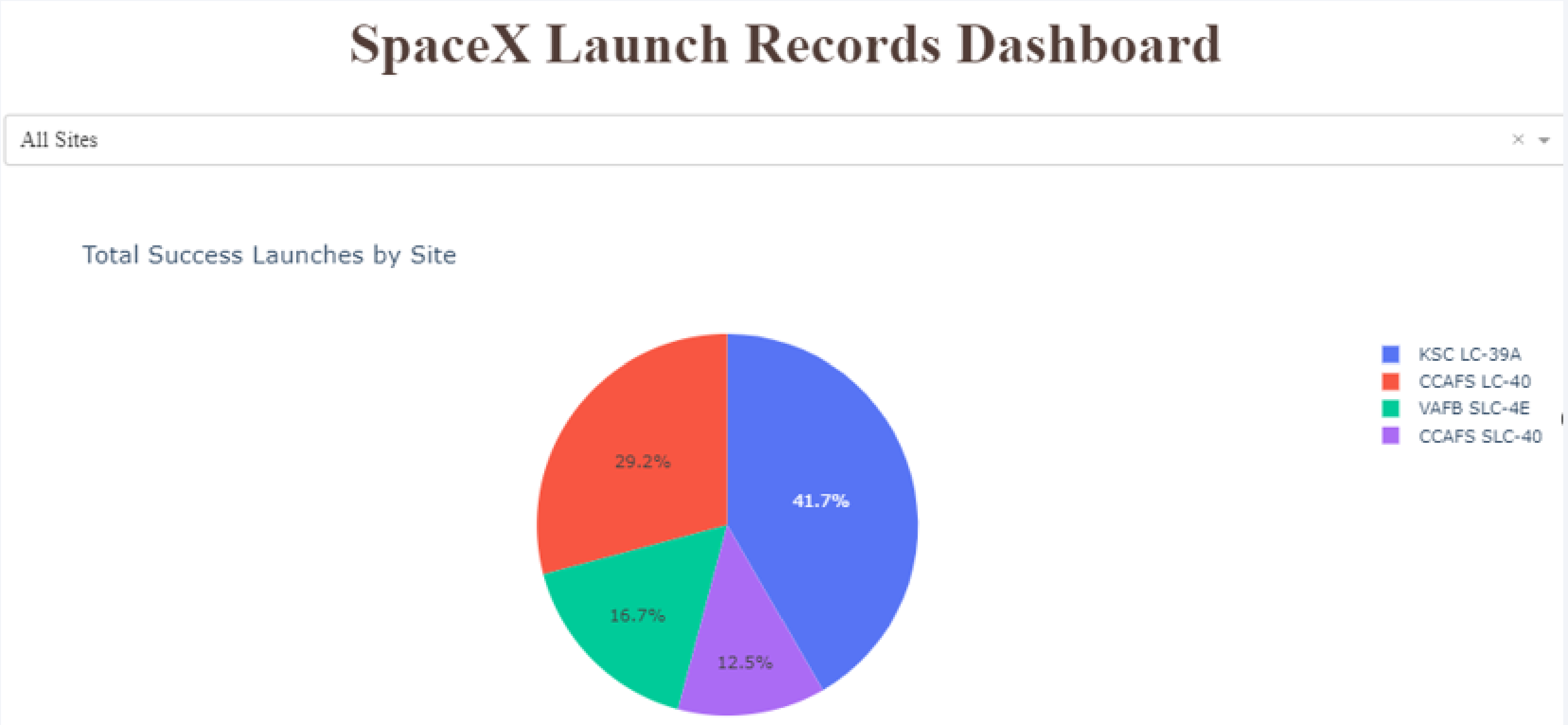
A rocket launch site is typically far from a city to ensure public safety and minimize the impact of noise and potential hazards, while proximity to railroads facilitates the efficient transportation of heavy and oversized launch equipment and materials to the site.



Section 4

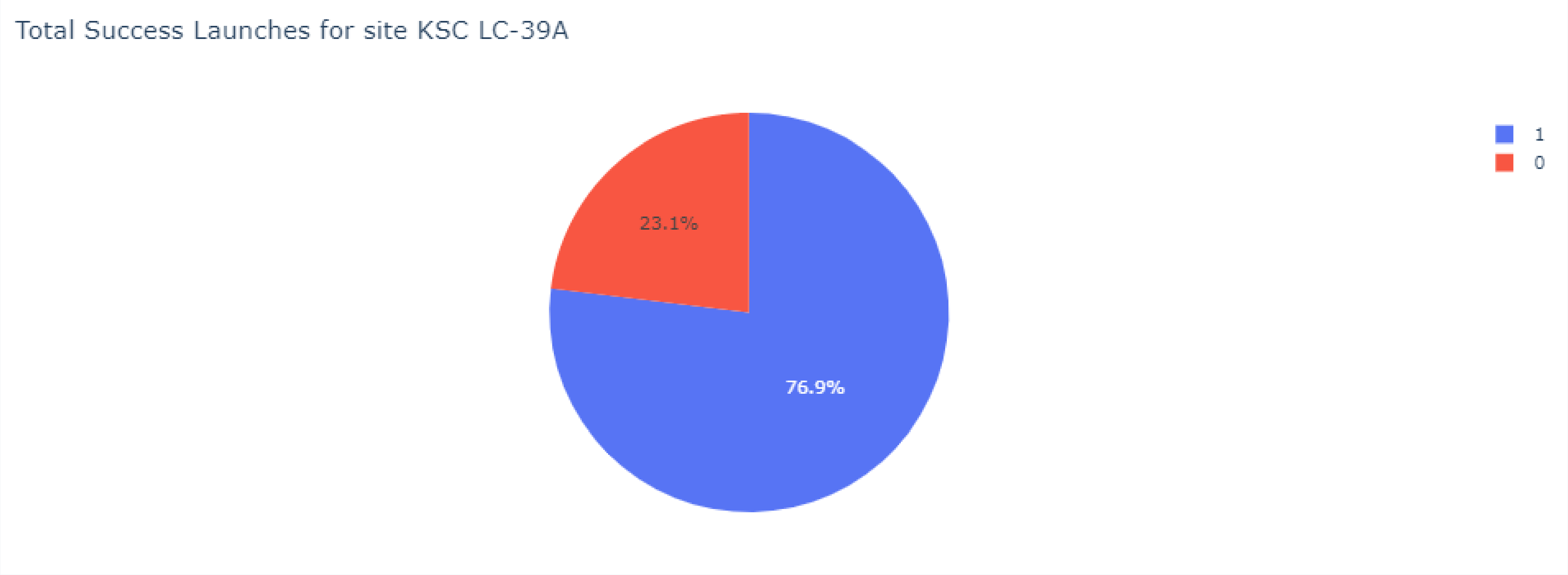
Build a Dashboard with Plotly Dash

Total Success Launces by Site



- As shown in the screenshot above the site KSC LC-39A has a 41.7% of all the successful launches by Space X.

Launch Site with Highest Launch Success Ratio



- Launch site with highest launch success ratio
- As shown in this Pie chart graph the site KSC LC-39A located on Florida US, has a 76.9% Launch success rate.

Correlation between Payload vs. Launch Outcome

- Show screenshot of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- The booster version Category FT 14 and 7 Failures has the overall highest successful rate with over 50% of success rate in the payload range of 0-7500 kg.

Section 5

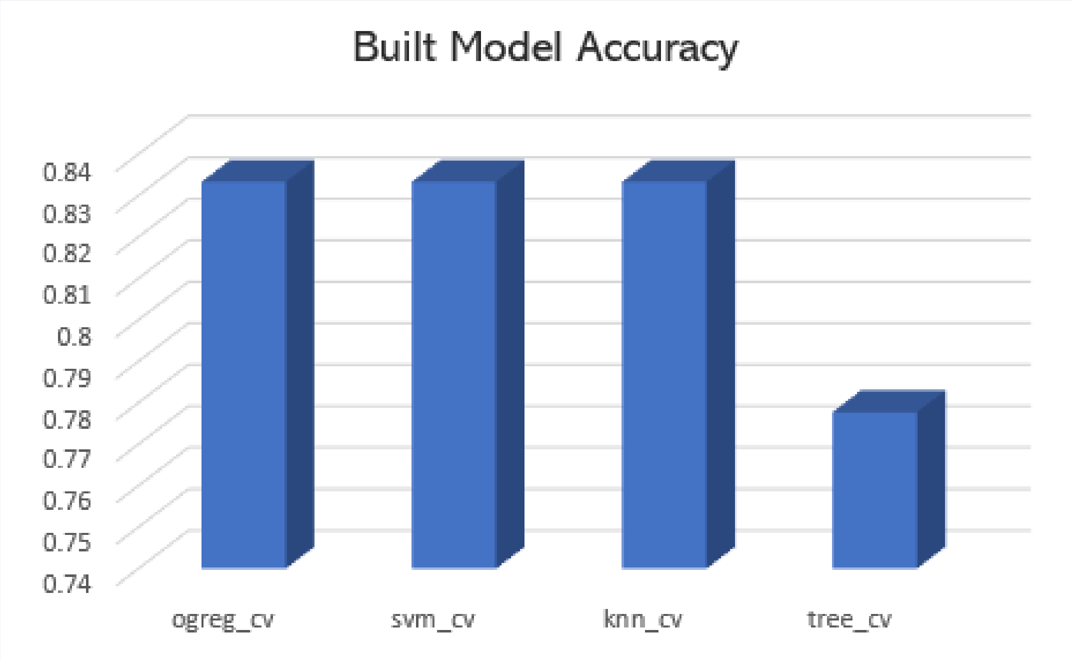
Predictive Analysis (Classification)

Classification Accuracy

Key observations :

The data results from machine learning models indicate that the logistic regression (ogreg_cv), support vector machine (svm_cv), and k-nearest neighbors (knn_cv) models all achieved the same accuracy score of approximately 83.33%. The decision tree model (tree_cv), however, scored slightly lower with an accuracy of approximately 77.78%.

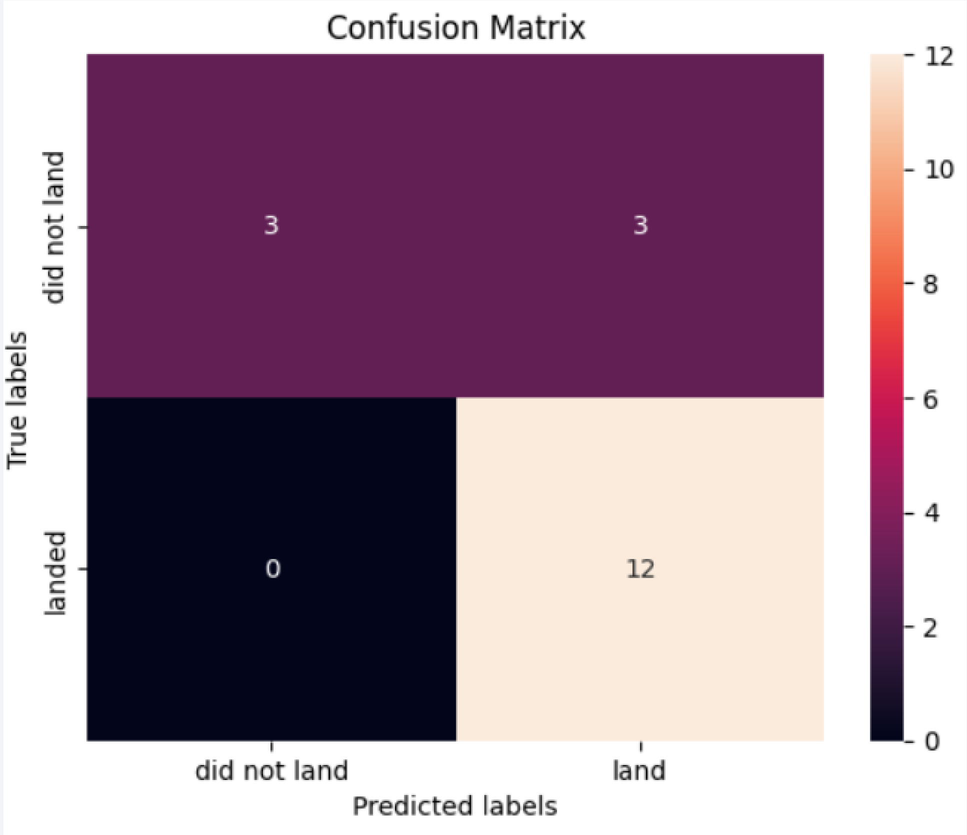
This suggests that the first three models are equally effective for the task at hand, while the decision tree model may be less reliable or overfitting to the training data in comparison.



Confusion Matrix

Key observations :

- There are 12 true positive cases where the model correctly predicted the rocket would land.
- There are 3 true negative cases where the model correctly predicted the rocket would not land.
- There are no false positive cases; the model did not incorrectly predict that the rocket would land when it didn't.
- There are 3 false negative cases where the model incorrectly predicted the rocket would not land when it actually did.
- From this matrix, we can infer that the model is better at predicting successful landings than unsuccessful ones, given that there are no false positives and the true positives significantly outnumber the true negatives. However, the presence of false negatives indicates that there's room for improvement in the model's ability to predict unsuccessful landings.



Conclusions

Based on the data and plots provided earlier, here are four key points that can be concluded regarding the prediction of successful Falcon 9 first stage landings:

1. **Increasing Experience Over Time**: The success rate by year plot shows a general upward trend in successful landings over time, suggesting that with each launch, SpaceX accumulates more data and experience, which likely contributes to refining their landing techniques and increasing the chances of success.

2. **Payload Mass Appears Non-Decisive**: The scatter plot of payload mass versus orbit indicates that successful landings are not exclusively dependent on the payload mass, as there are successful landings across a wide range of payload masses. This suggests that while payload mass is a factor in launch dynamics, it may not be a primary determinant of landing success.

3. **Influence of Orbital Destination**: Different orbits show varying levels of success, with some, like GTO, presenting more challenges as indicated by a mix of successful and unsuccessful landings. This implies that the destination orbit might play a role in the complexity of the landing maneuver, affecting the predictability of a successful landing.

4. **Learning from Past Launches**:The scatter plot of flight number versus orbit by class demonstrates an increase in successful landings as the flight number grows, hinting at SpaceX's learning curve and continuous improvement in launch operations. This increase in success rate with higher flight numbers could be leveraged as a predictive indicator for future successful landings.

In conclusion, while there is no single factor that guarantees the success of Falcon 9 first stage landings, the cumulative learning from past flights, considerations of payload mass, and the specific challenges posed by different orbital destinations can be integrated into a predictive model to estimate the likelihood of successful landings in future missions.

Appendix

- Relevant Websites:
  - https://www.spacex.com/
  - https://en.wikipedia.org/wiki/Falcon_9

Thank You!