Space-X - Mission Statement

To revolutionize space technology, with an ultimate goal of people to live on other planets.

Prepared by: Athar Iqbal

January 8th, 2025

Agenda

Executive Summary

Introduction

Methodology

Results

Conclusion

Executive Summary

Summary of Methodologies

The research aims to identify the key factors contributing to a successful rocket landing. To determine these factors, the following methodologies were employed:

Data Collection: Utilized the SpaceX REST API and web scraping techniques.

Data Wrangling: Created a success/fail outcome variable.

Data Exploration: Used data visualization techniques to examine factors such as payload, launch site, flight number, and yearly trends.

Data Analysis: Employed SQL to calculate statistics such as total payload, payload range for successful launches, and the number of successful and failed outcomes.

Launch Site Analysis: Explored success rates and proximity to geographical markers.

Visualization: Highlighted launch sites with the highest success rates and successful payload ranges.

Model Building: Predicted landing outcomes using logistic regression, support vector machines (SVM), decision tree, and K-nearest neighbors (KNN).

• Results

- Exploratory Data Analysis:
  - ❖ Launch success has shown improvement over time.
  - ❖ VFAB SLC 4E and KSC LC 39A has the highest success rate among all landing sites.
  - ❖ Orbits GTO, LEO, and Polar LEO has a higher success rate.
- Executive Summary Visualization/Analytics:
  - ❖ The majority of launch sites are situated near the equator and in close proximity to coastlines.

- Visualization/Analytics:
- Most launch sites are near the equator, and all are close to the coast
- Predictive Analytics:
- All models performed similarly on the test set. The decision tree model slightly outperformed

Introduction

Background

The commercial space age is here, making space travel more affordable. SpaceX has achieved significant milestones, such as launching spacecraft to the International Space. A key factor in SpaceX's success is the cost-effectiveness of its launches, with reusable first stages of the Falcon 9 rocket.

Problem Statement

SpaceX is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars whereas other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

Methodology

Methodology

Data Collection was performed using SpaceX REST APIs and web scraping techniques

Data Wrangling by filtering, augmenting missing values and applying one hot encoding to prepare the data for analysis and modeling

Exploration Data Analysis (EDA) was performed using SQL and data visualization techniques

Visualization was performed using Folium and Plotly Dash tools

Built Models to predict landing outcomes using classification models. Fine tuned models and evaluated to find the best model and parameters

Data Collection – REST APIs

Request Data – from SpaceX API (rocket launch data)

Decode Response – using .json() and converted t pandas data frames using .json_normalize() function

Request Information – iterated through the frames and requested additional information using custom functions

Create Dictionary – from the data collected

Create Data Frames – from the dictionary

Filter Data Frames – to ensure that data frames hold Falcon 9 launches data only

Replace Missing Values – Payload Mass attributes using .mean() function

Export the result sets into CSV files – use in later analysis

[Source Code: SpaceX API based Data Collection]

Data Collection – Web Scraping

Data Source – Wikipedia website to collect Falcon 9 Launch data (Wikipedia Flacon 9 Launches)

Decode Response – used BeautifulSoup object to parse the HTML response data

Column Name Extraction – used HTML table header to extract the column names

Data Extraction – used HTML table rows to collect the relevant data along with validation

Create Data Dictionary – from collected data

Create Data Frames – from the dictionary

Export the result sets into CSV files – use in later analysis

Source Code: Data Collection using Site Scrapping

Data Wrangling

Identified bad outcome for landings

Augmented new classification feature to the dataset for landing outcome

Identified Successful and Failed landing attempts for different types

Identified the percentage of missing information or NaN

Augmented missing data attributes using .mean() functions.

Source Code: Data Wrangling

EDA with Data Visualization

Charts to visualize the relationship

- Flight Number and Launch Site
- Payload and Launch Site
- Success rate of each orbit type
- Flight Number and Orbit type
- Payload and Orbit type
- Success yearly trend

Analysis

- Viewed relationship by using scatter plots . The variables could be useful for machine learning if a relationship exists
- Show comparisons among discrete categories with bar charts . Bar charts show the relationships among the categories and a measured value.

EDA with SQL

Queries performed to:

Display:

- Names of unique launch sites
- 5 records where launch site begins with 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1.

Result set:

- Date of first successful landing on ground pad
- Names of boosters which had success landing on drone ship and have payload mass greater than 4,000 but less than 6,000
- Total number of successful and failed missions
- Names of booster versions which have carried the max payload
- Failed landing outcomes on drone ship, their booster version and launch site for the months in the year 2015
- Count of landing outcomes between 2010-06-04 and 2017-03-20 (desc)

Folium Maps

Markers identifying Launch Sites

Added circle at NASA Johnson Space Center's coordinate with a popup label showing its name using its latitude and longitude coordinates

Added red circles at all launch sites coordinates with a popup label showing its name using its name using its latitude and longitude coordinates

Colored Markers of Launch Outcomes

Added colored markers of successful (green) and unsuccessful (red) launches at each launch site to show which launch sites have high success rates

Distances Between a Launch Site to Proximities

Added colored lines to show distance between launch sites and its proximity to the nearest coastline, railway, highway, and city

Dashboard with Plotly Dash

Developed dashboard application containing input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart. These application features were developed to answer the insights such as:

- Site that has the largest successful launches?
- Site that has the highest launch success rate?
- Payload range(s) with highest launch success rate?
- Payload range(s) with lowest launch success rate?
- F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) Payload vs Success rate

Predictive Analytics (Classification)

Creating a high-performing classification model involves several steps, starting from data preparation and ending with model evaluation and tuning. Here is the summary of the process:

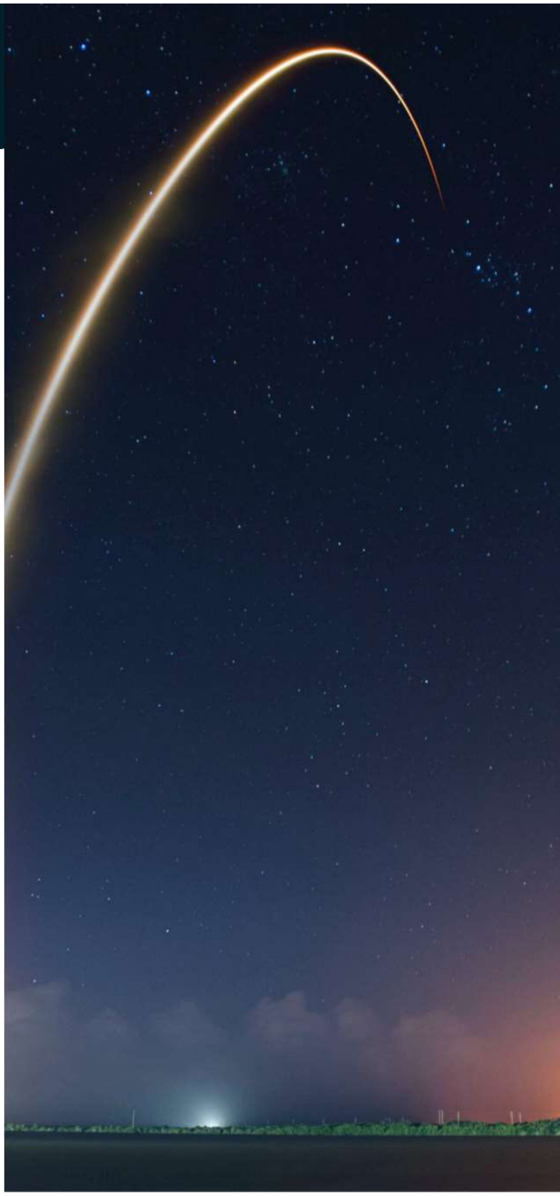
Prepare Data - Cleansed, preprocessed, and split the dataset into training and testing datasets

Build Model - Train multiple classification models (Logistic Regression, SVM, Decision Tree, KNN)

Evaluate - Used accuracy, score, and confusion matrix to evaluate each model

Improve - Optimized models via hyperparameter tuning and cross-validation

Select Best - Identified the model with the highest performance



Results

Results Summary

Exploratory Data Analysis:

- ❖ Launch success has shown improvement over time.
- ❖ VAFB SLC 4E and KSC LC 391 has the highest success rate among all landing sites.
- ❖ Orbits GTO, LEO, and Polar LEO has a higher success rate.

Executive Summary Visualization/Analytics:

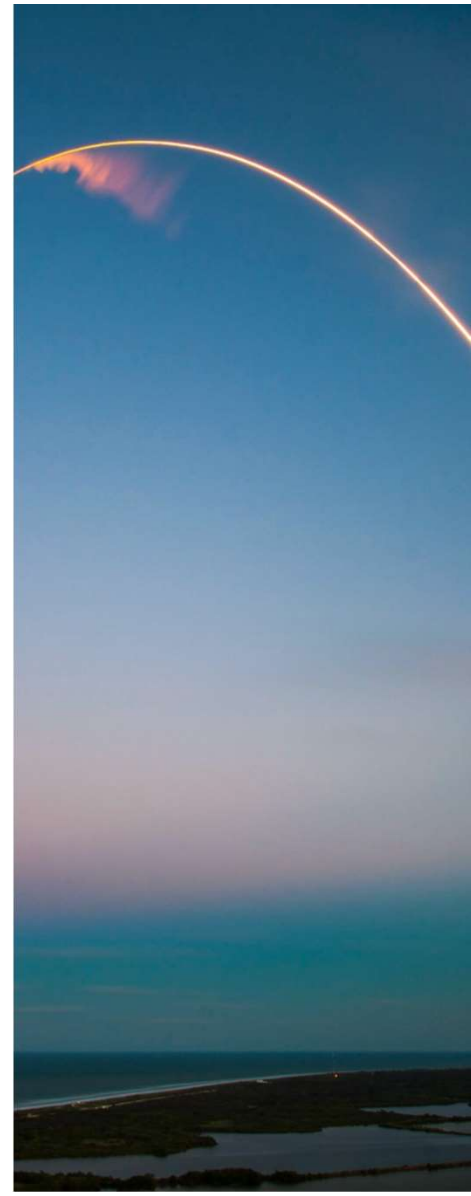
The majority of launch sites are situated near the equator and in close proximity to coastlines.

Visualization/Analytics:

Most launch sites are near the equator, and all are close to the coast

Predictive Analytics:

All models performed similarly on the test set. The decision tree model slightly outperformed

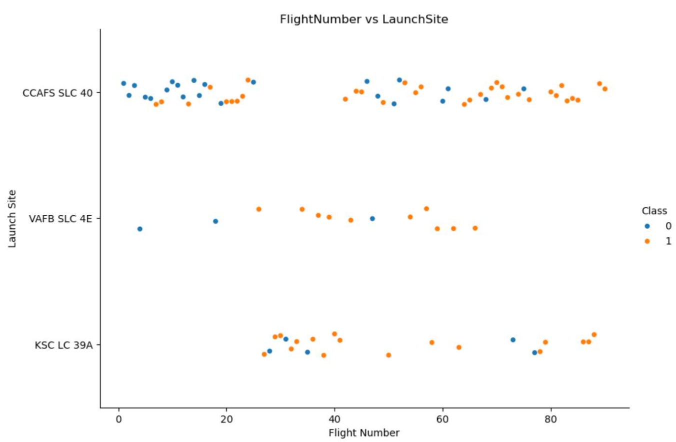


Source Code: EDA with Data Visualization Tools
Source Code: EDA with SQL

Flight Number vs Launch Site

Data Analysis

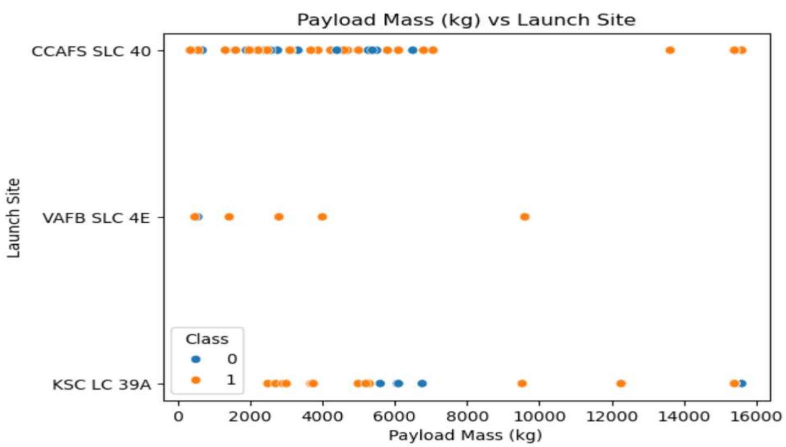
- **Initial flights** had a **lower** the success rate (blue = failure)
- **Later flights** had a **higher** the success rate (yellow = success)
- VAFB SLC 4E and KSC LC 39A have higher success rate
- Most of the launches took place from CCAFS SLC 40



Payload Mass vs Launch Site

Data Analysis

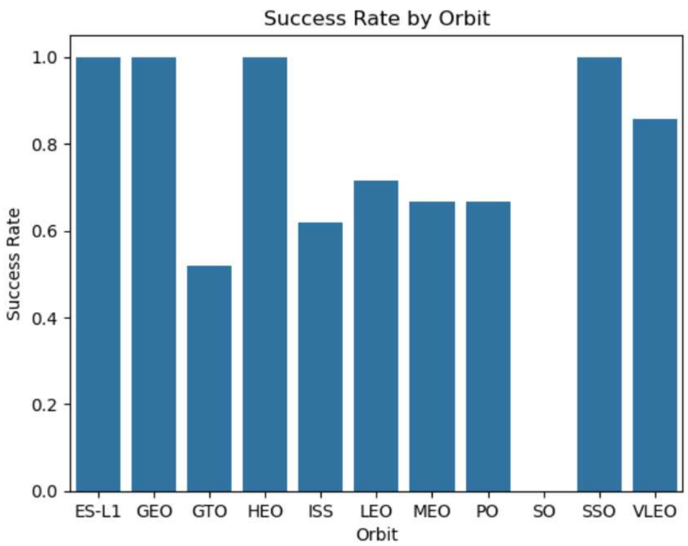
- **Higher** the Payload Mass (kg), **higher** the success rate
- Most launches with a payload of 7000 kg or higher had a successful mission outcome
- VAFB SLC 4E has higher success rate between 1,500-10,000 kg payload, and has not launched anything higher 10,000 kg
- KSC LC 39A has a 100% for launches carrying payload mass up to 5,500, it also has higher success rate between 9,000 and 15,000 kg
- Most of the launches took place from CCAFS SLC 40

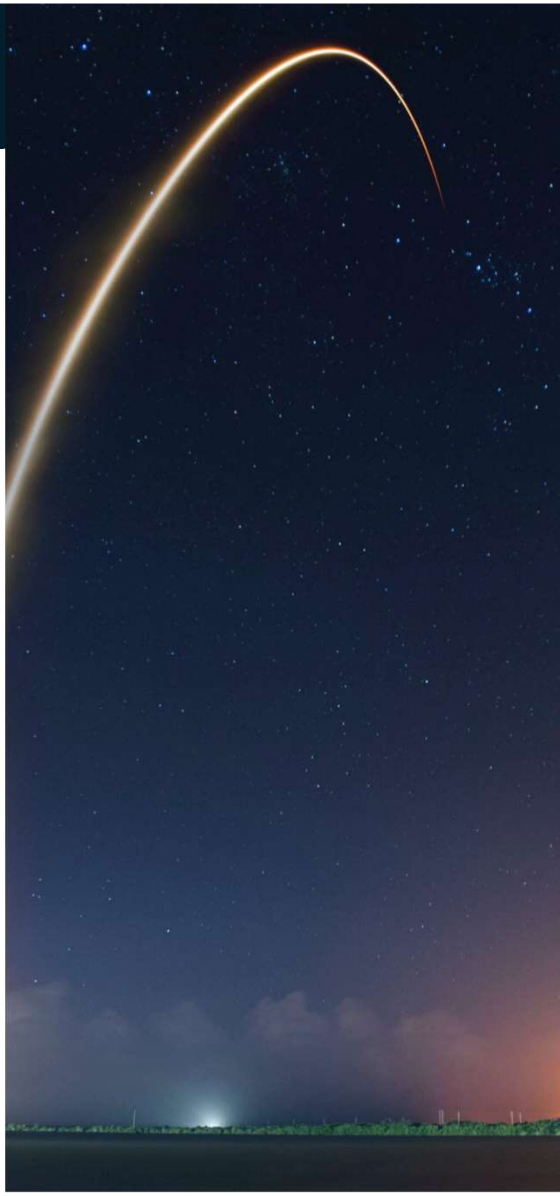


Success Rate by Orbit

Data Analysis

- 100% Success Rate – ES-L1, GEO, HEO and SSO
- 50%-80% Success Rate – GTO, ISS, LEO, MEO, PO
- 0% Success Rate - SO

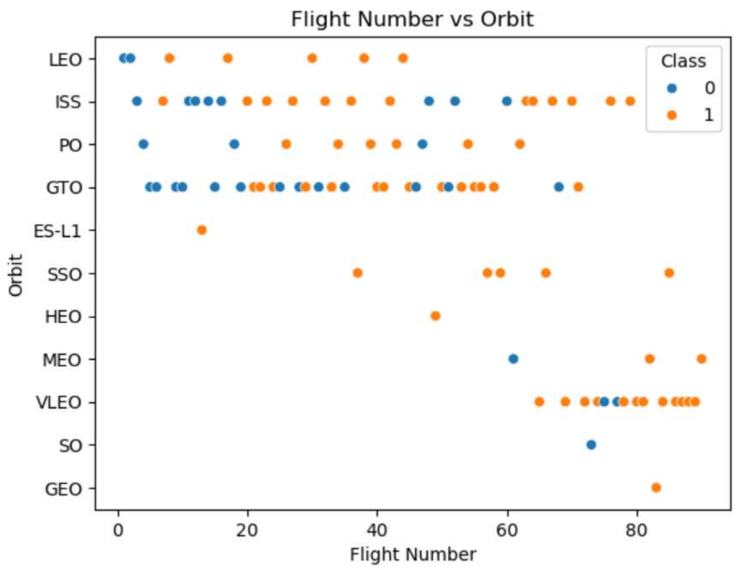




Flight Number vs Orbit

Data Analysis

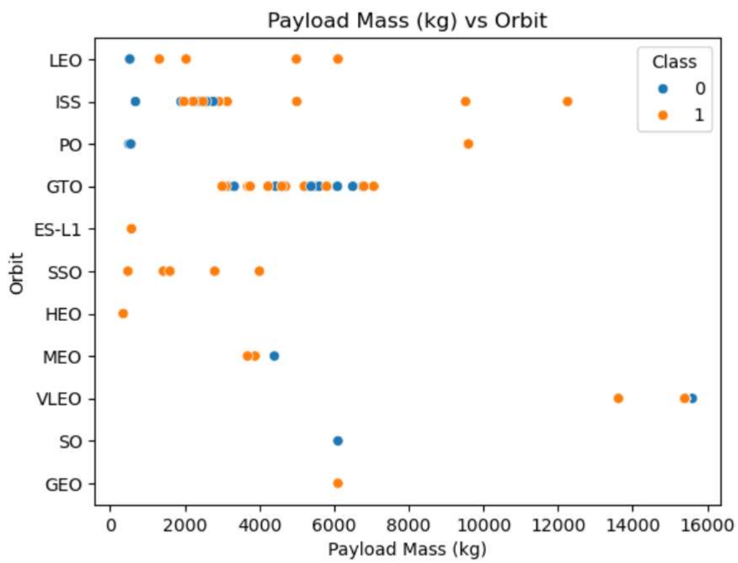
- **Higher** success rate is achieved with each attempt in the orbit
- **Higher Success relation** seems to appear for the VLEO orbit
- GTO orbit does not present this trend

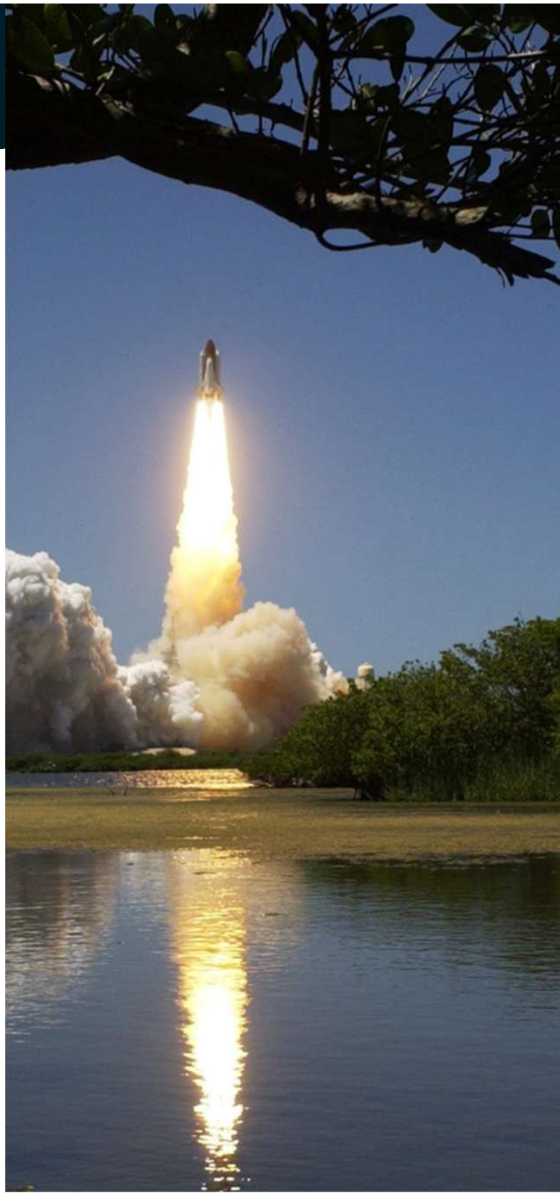


Payload Mass vs Orbit

Data Analysis

- Heavy payload mass has higher success rate for LEO, ISS and PO orbits
- SSO seems to Nan have success rate for up to 4,000 kg payload
- GTO seems to have mixed outcome for heavier payloads

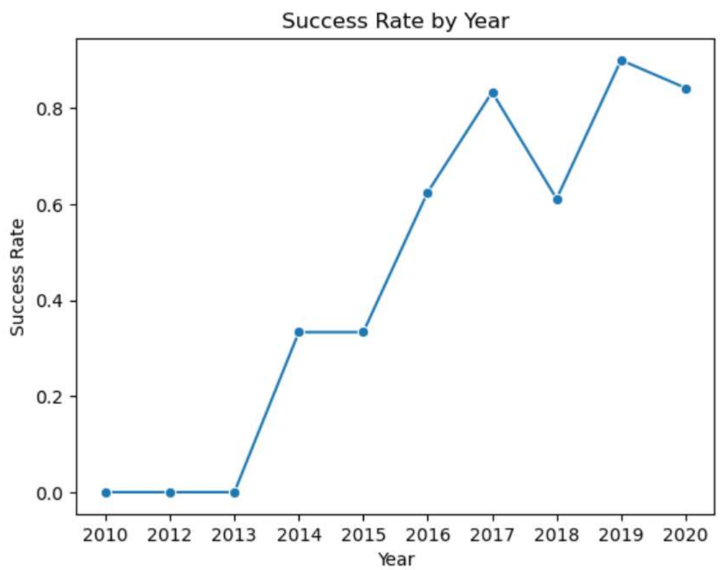




Launch Success over Time

Data Analysis

- Success rate improved significantly between 2013 thru 2017 and 2018 thru 2019
- Success rate dipped from 2017 thru 2018, and 2019 thru 2020
- Overall, success rate has improved starting 2013



Launch Sites Detail

Launch Sites Name

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40

%sql select distinct(Launch_Site) from SPACEXTABLE

Total Successful Launches Ranking by Site

CCAFS SLC-40	32
KSC LC-39A	25
CCAFS LC-40	25
VAFB SLC-4E	16

%sql select Launch_Site, count(*) from SPACEXTABLE where Mission_Outcome = 'Success' Group by Launch_Site order by count(*) desc

Record with Launch Sites Starting with CCA (first 5 record only)

%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5

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

Landing and Mission Details

First Successful Landing on Ground Pad

12/22/2015

%sql select Date from SPACEXTABLE where Landing_Outcome = 'Success (ground pad)' order by Date asc limit 1

Date
2015-12-22

Drone Ship Landing

Boosters versions payload mass between 4,000 and 6,000 kg.

F9 FT B1022, F9 FT B1026,

F9 FT B1021.2, F9 FT B1031.2

%sql select distinct(Booster_Version) from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000

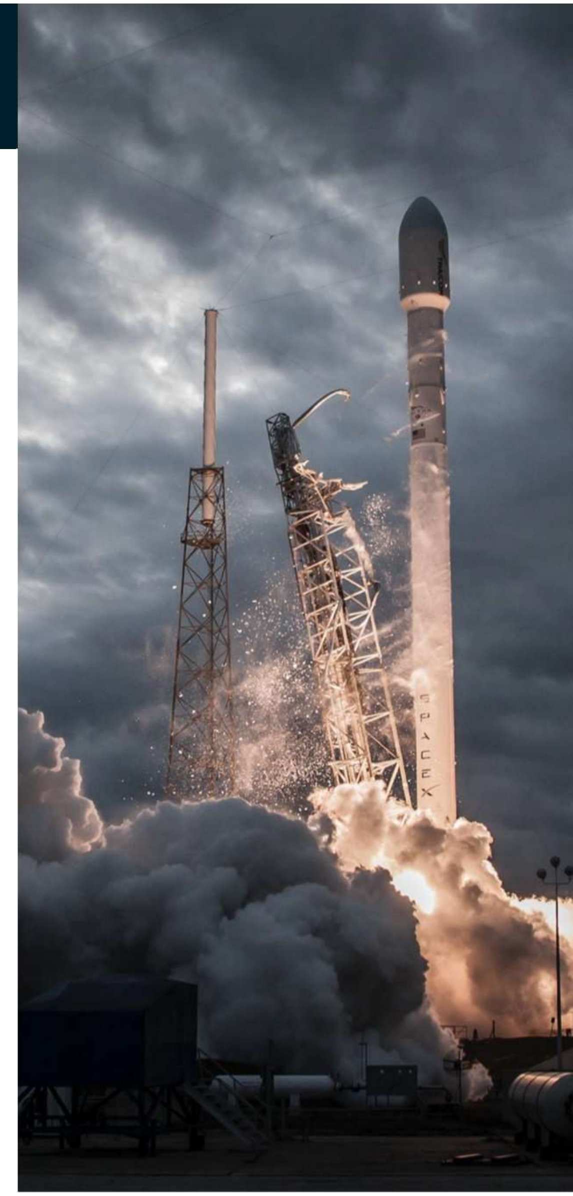
Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Payload Mass between 4,000 and 6,000 kg.

JCSAT-14, JCSAT-16, SES-10, SES-11 / EchoStar 105

%sql select distinct(Payload) from SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000

Payload
JCSAT-14
JCSAT-16
SES-10



Landing and Mission Details cont.

Total Number of Successful and Failed Mission Outcomes

1 Failure in Flight

99 Success

1 Success (Payload status unclear)

%sql select Mission_Outcome, count(Mission_Outcome) from SPACEXTABLE group by Mission_Outcome order by Mission_Outcome

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

Booster carrying maximum payload

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

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

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

Failed Landings on Drone Ship

%sql SELECT substr(Date, 6,2) AS Month, Booster_Version, Launch_Site, Landing_Outcome FROM SPACEXTABLE WHERE substr(Date, 1, 4) = '2015' AND Landing_Outcome = 'Failure (drone ship)'

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

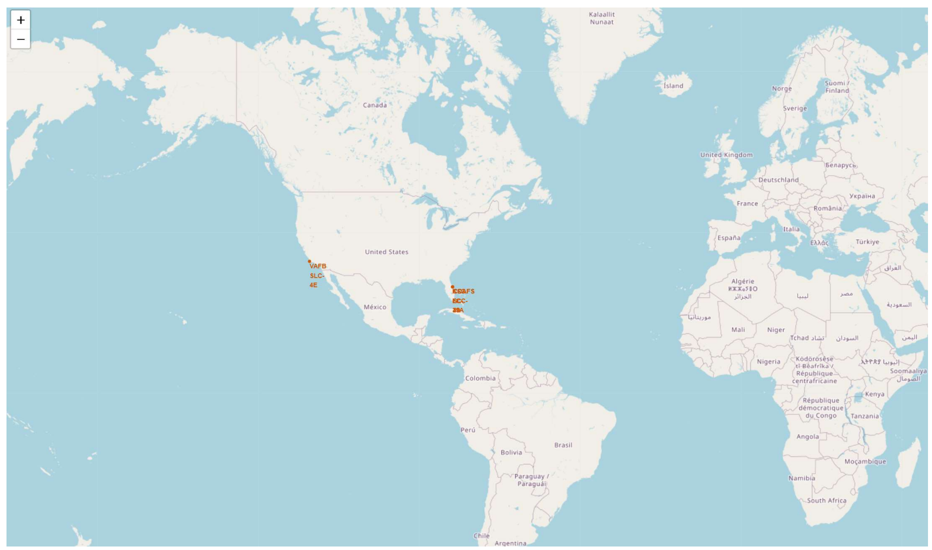
Ranking Landing Outcomes

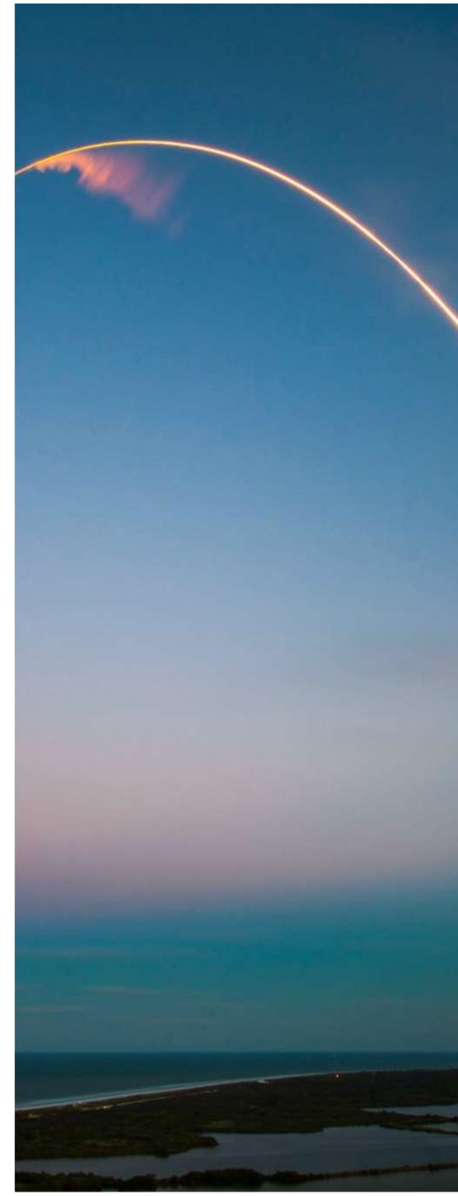
```
%sql SELECT Landing_Outcome, COUNT(*) AS Count FROM SPACEXTABLE WHERE Date BETWEEN
'2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Count DESC
```

Landing_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

Launch Site Analysis

Close to the Equator: Launch sites near the equator make it easier to reach equatorial orbit. This is because rockets benefit from Earth's rotational speed, particularly for prograde orbits. The rotational boost near the equator provides a natural advantage, reducing the need for extra fuel and boosters, and thus cutting costs.





[Source Code: Folium Maps]

Launch Site with Outcomes

Outcomes

- Successful launches are marked Green
- Failed Launches are marked with Red
- Launch Site CCASFS SLC-40 has 43% success rate

Launch Site with proximity to Landmarks

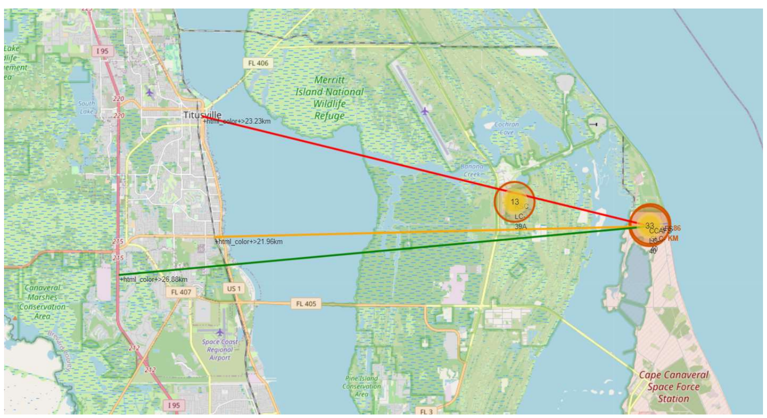
CCAFS SLC-40

.86 km from nearest coastline

21.96 km from nearest railway

23.23 km from nearest city

26.88 km from nearest highway



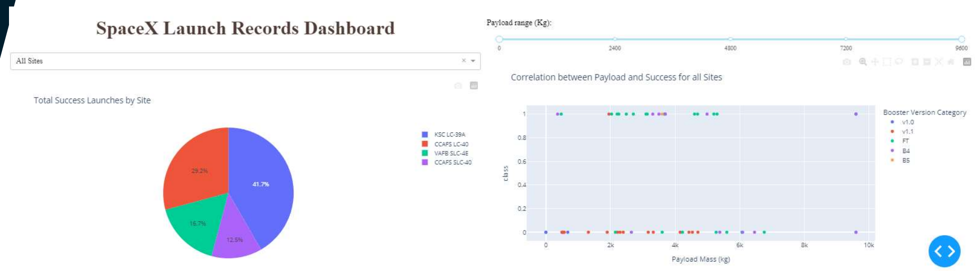
Coasts: Launch sites near coasts ensure that any spent rocket stages or failed launches fall into the ocean, avoiding people and property.

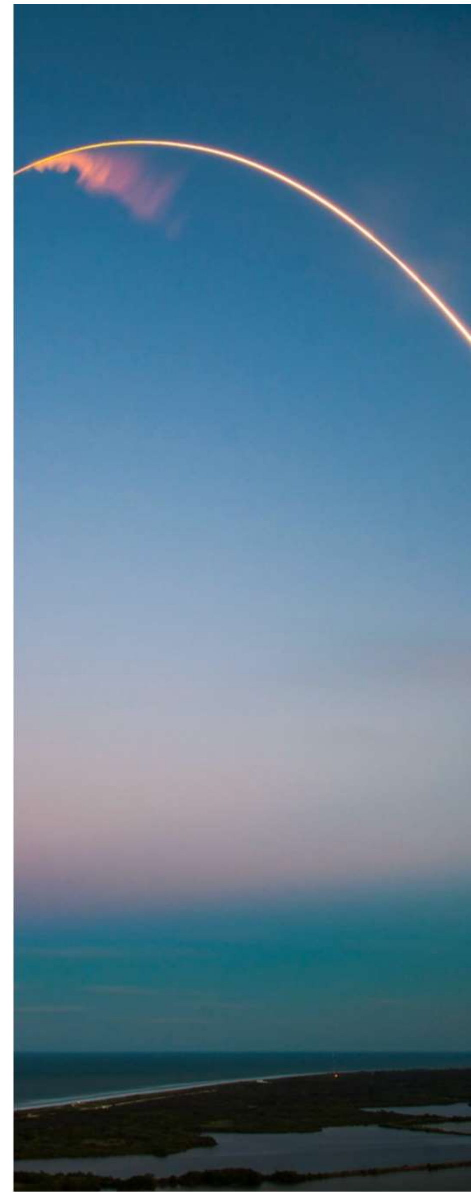
Safety and Security: An exclusion zone around the launch site keeps unauthorized individuals away, ensuring their safety and the security of the launch.

Transportation, Infrastructure, and Cities: Launch sites should be distant enough from cities and infrastructure to minimize potential damage from failed launches. However, they need to be close enough to roads, railways, and docks to efficiently transport people and materials for launch support activities.

Dashboard with Plotly

Dashboard with Plotly: Application was built with Plotly to analyze the success rate for all launch sites along with eh correlation between payload and success rate.



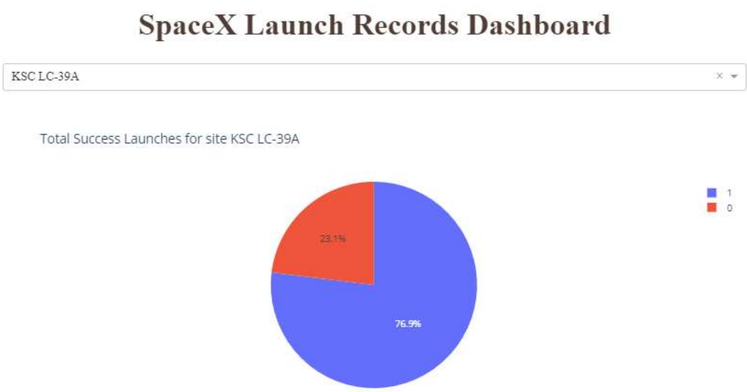


[Source Code: SpaceX Dashboard Application]()

Launch Site and Success Rate with Plotly

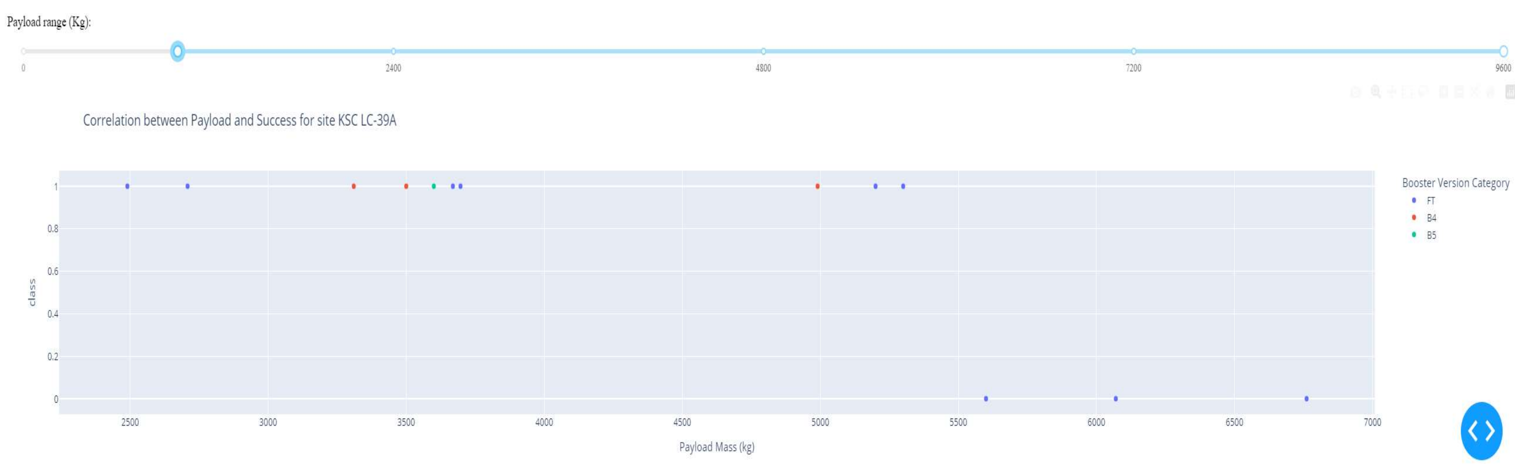
KSC LC-39A has the highest success rate among the launch sites.

79.6% Success Rate



Payload Mass and Success Rate with Plotly

Payloads between 2,000 kg and 5,000 kg have the highest success rate



Predictive Analytics

Summary

Performed exploratory Data Analysis and determined Training Labels

Created column for the class

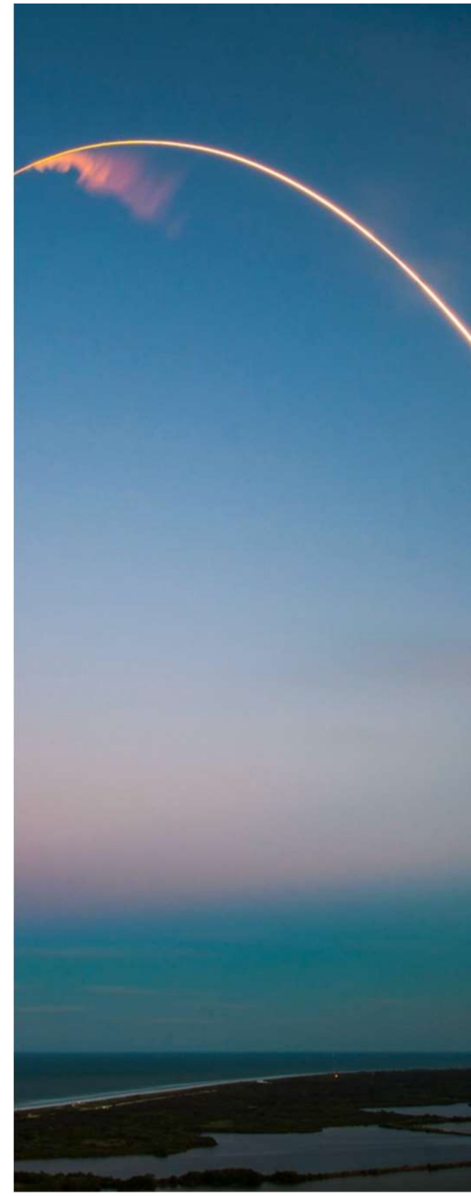
Standardized the data

Split the data into training data and test data

Identified best Hyperparameter for SVM, Classification Trees and Logistic Regression

Identified the method performed best using test data

[Source Code: Machine Learning Prediction]



Classification Accuracy

All the models performed similarly, achieving comparable scores and accuracy.

This outcome is likely due to the small dataset.

However, the **Decision Tree** model slightly outperformed the others when evaluating the .best_score_, which represents the average of all cross-validation folds for a single combination of parameters.

ll the models

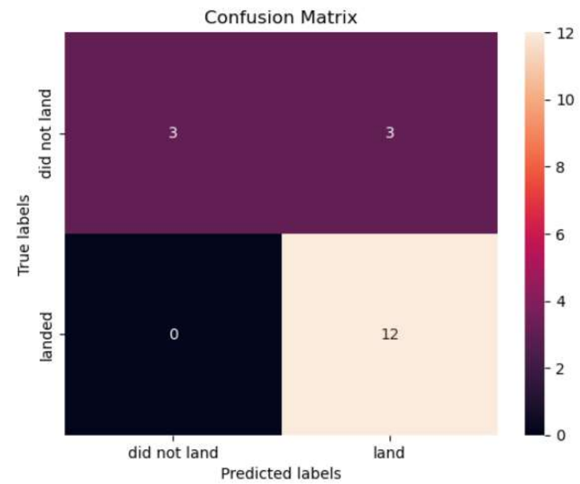
```
print("Test accuracy for Logistic Regression:", test_accuracy_logreg)
print("Test accuracy for SVM:", test_accuracy_svm)
print("Test accuracy for Decision Tree:", test_accuracy_tree)
print("Test accuracy for KNN:", test_accuracy_knn)

# Determine which method performs the best
best_model = max(test_accuracy_logreg, test_accuracy_svm, test_accuracy_tree, test_accuracy_knn)
print("The best performing model is:", best_model)
```

```
Test accuracy for Logistic Regression: 0.8333333333333334
Test accuracy for SVM: 0.8333333333333334
Test accuracy for Decision Tree: 0.9444444444444444
Test accuracy for KNN: 0.8333333333333334
The best performing model is: 0.9444444444444444
```

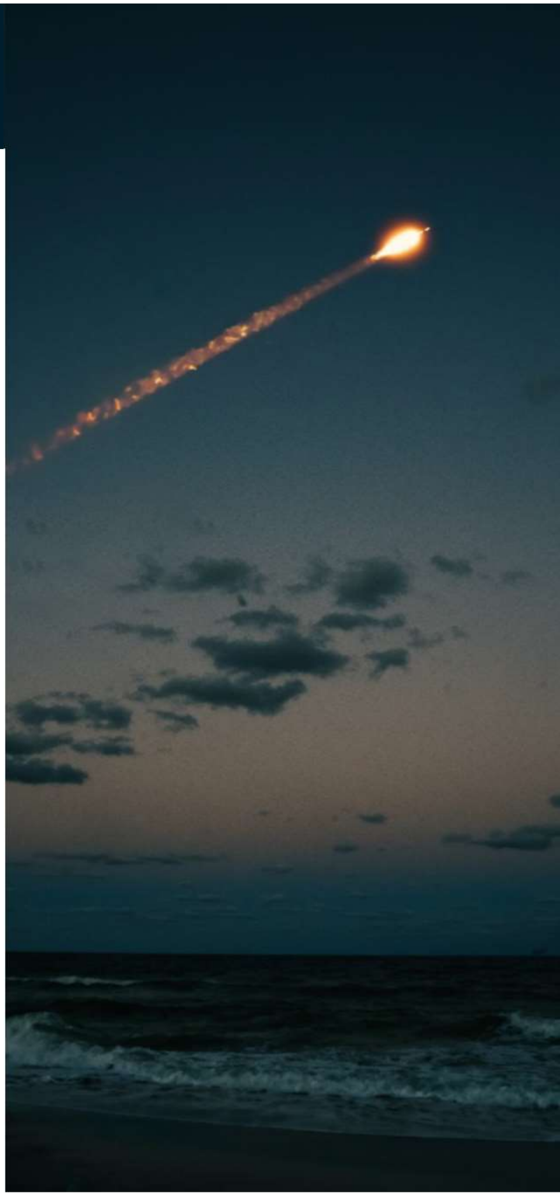
Confusion Matrix

Performance Summary

- All the confusion matrices were identical
- Confusion Matrix Outputs:
- 12 True positive
- 3 True negative
- 3 False positive
- 0 False Negative



Accuracy= (TP + TN) / (TP + TN + FP + FN) = .833



Conclusion

Conclusion

Model Performance: The models performed similarly on the test set with the decision tree model slightly outperforming

Equator: Most of the launch sites are near the equator for an additional natural boost -due to the rotational speed of earth – which helps save the cost of putting in extra fuel and boosters

Coast: All the launch sites are close to the coast

Launch Success: Increases over time

KSC LC-39A: Has the highest success rate among launch sites. Has a 100% success rate for launches less than 5,500 kg

Orbits: ES-L1, GEO, HEO, and SSO have a 100% success rate

Payload Mass: Across all launch sites, the higher the payload mass (kg), the higher the success rate

Appendix

Appendix

GitHub repository endpoint: Code Repository

Useful links to various source code:

Machine Learning Prediction

SpaceX API based Data Collection

Data Collection using Site Scrapping

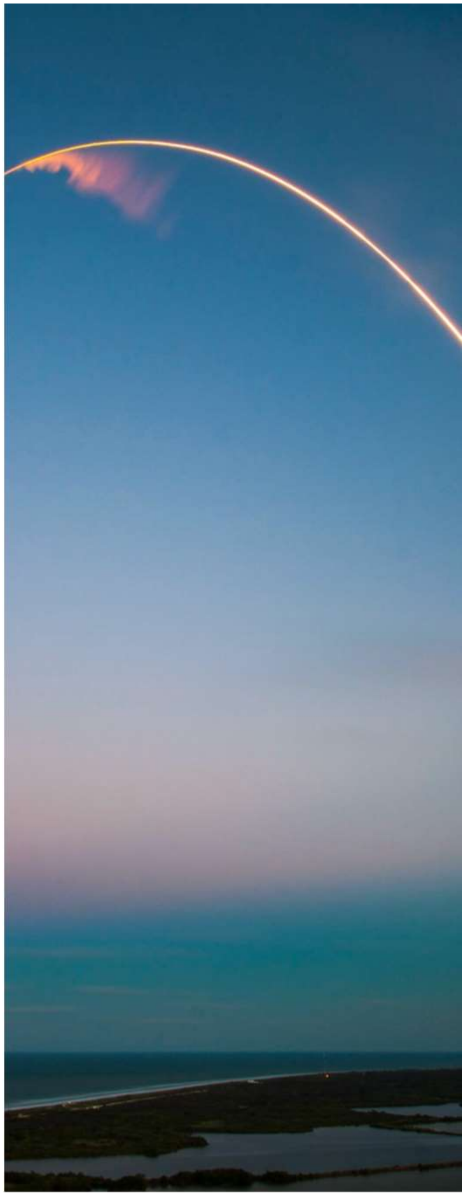
Data Wrangling

EDA with SQL

EDA with Data Visualization Tools

Folium Maps

Space-X Dashboard Application



THANK YOU

Athar Iqbal