# SpaceX Rocket Analytics project

A guide by Jake Best

# Executive Summary

SpaceX's rocket reuse strategy is yielding remarkable results, with key insights emerging from data analysis.

Proximity to the equator significantly boosts first stage landing success rates, leveraging Earth's rotation for efficient prograde orbits.

Heavier payloads, particularly those bound for Polar, Low Earth Orbit (LEO), and the International Space Station (ISS), exhibit higher rates of successful first stage landings.

However, in Geostationary Transfer Orbit (GTO) missions, the relationship between payload weight and landing success is less clear, with both positive and negative outcomes observed.

Understanding these dynamics is crucial for optimizing future missions and advancing the efficiency of rocket reusability in space exploration.

# Intro

SpaceX has redefined the landscape of space exploration through its pioneering approach to rocket reusability. By successfully recovering and refurbishing first stage boosters, SpaceX has not only reduced the cost of space travel but also unlocked new possibilities for rapid and sustainable access to space.

In this analysis, we delve into the statistics and data surrounding SpaceX's rocket reuse program in 4 stages:

1. Data Collection Stage
2. EDA
3. Data Visualization
4. Predictive Analysis

# Data Collections

## Data Collection Methodology

- Primarily gathered data through:
  - Web scraping Wikipedia pages on SpaceX using BeautifulSoup.
  - API requests to SpaceX website gathering the data provided from the company directly.

# EDA and Visual Analysis

## EDA Methodology

- Primary EDA methodologies:
    - Initially preformed data wrangling stage, which mostly comprised of handling missing data.
    - Moved on to SQL querying of data, which yielded insights into the payload vs boost version of the rockets success rate.
    - Finally moving on to EDA in python primarily to preform initial analysis in the factors:
        - Flight Number
        - Launch Site
        - Payload Mass
        - Orbit
        - Success rate (Class)

## Visualization Methodology

- Primary visualization methodologies:
    - During EDA of the SpaceX data we used the following visuals:
        - Scatterplots
        - Bar Graphs
        - Line Chart
    - After EDA, we preformed visualization using folium in Python. This allowed us to gather geographic details on the launches.
    - Lastly, we created interactive dashboards using Ploty Dash, which included:
        - Pie Chart
        - Scatterplot

# Predictive Analysis

## Predictive Analysis Methodology

- During the final stage, Predictive Analysis, we:
  - Split our data set into train, test sets for modeling via GridSearchCV. We used:
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree Classifier
    - K Nearest Neighbors

# EDA: SQL Results

## Task 3

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

```
[10]: %sql SELECT SUM(PAYLOAD_MASS__KG_) \
        FROM SPACEXTBL \
        WHERE CUSTOMER = 'NASA (CRS)';
```

* sqlite:///my_data1.db
Done.

[10]:

| SUM(PAYLOAD_MASS__KG_) |
|---|
| 45596 |

## Task 4

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

```
[11]: %sql SELECT AVG(PAYLOAD_MASS__KG_) \
        FROM SPACEXTBL \
        WHERE BOOSTER_VERSION = 'F9 v1.1';
```

* sqlite:///my_data1.db
Done.

[11]:

| AVG(PAYLOAD_MASS__KG_) |
|---|
| 2928.4 |

**Insight:**

The EDA analysis of the data provided for SpaceX's Falcon 9 rocket resulted in less of a payload and a high number of successful missions.

## Task 7

List the total number of successful and failure mission outcomes

```
19]: %sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
     FROM SPACEXTBL \
     GROUP BY MISSION_OUTCOME;
```

* sqlite:///my_data1.db
Done.

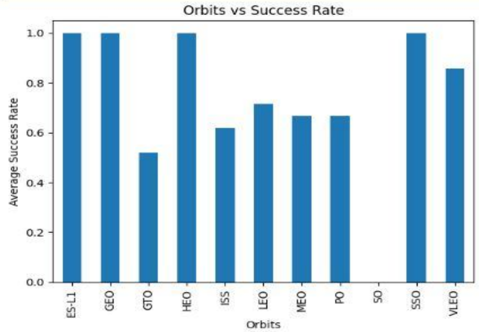
19]:

| Mission_Outcome | total_number |
|---|---|
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

# EDA: Python Visualization Results

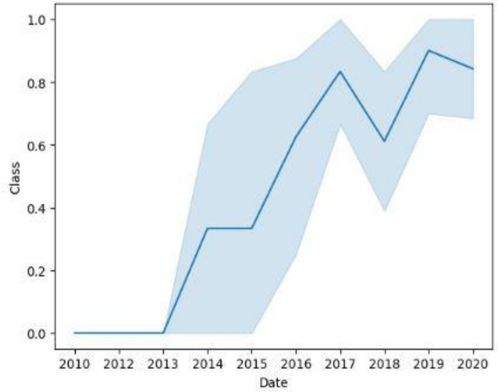
```python
# HINT use groupby method on Orbit column and get the mean of Class column
Orbit = df.groupby('Orbit')
Orbit_Class_Mean = Orbit['Class'].mean()
Orbit_Class_Mean

Orbit_Class_Mean.plot(kind = 'bar')
plt.xlabel('Orbit Type') # add to x-label to the plot
plt.ylabel('Average Success Rate') # add y-label to the plot
plt.title('Success Rate Of Each Orbit Type') # add title to the plot
plt.show()
```




**Insight:** From this graph we can observe the success rates of the different orbits

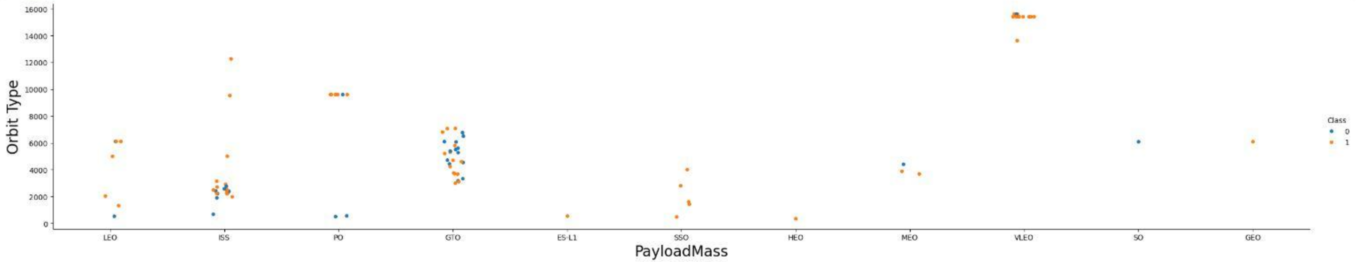
```python
# Plot a line chart with x axis to be the extracted year and y axis to be the success rate
sns.lineplot( x = "Date", y = "Class", data = df);
```




you can observe that the sucess rate since 2013 kept increasing till 2020

**Insight:** From this graph we can observe that the success rate of launches increased since 2013.

```python
# Plot a scatter point chart with x axis to be Payload and y axis to be the Orbit, and hue to be the class value
sns.catplot(y="PayloadMass", x="Orbit", hue="Class", data=df, aspect = 5)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("Orbit Type",fontsize=20)
plt.show()
```

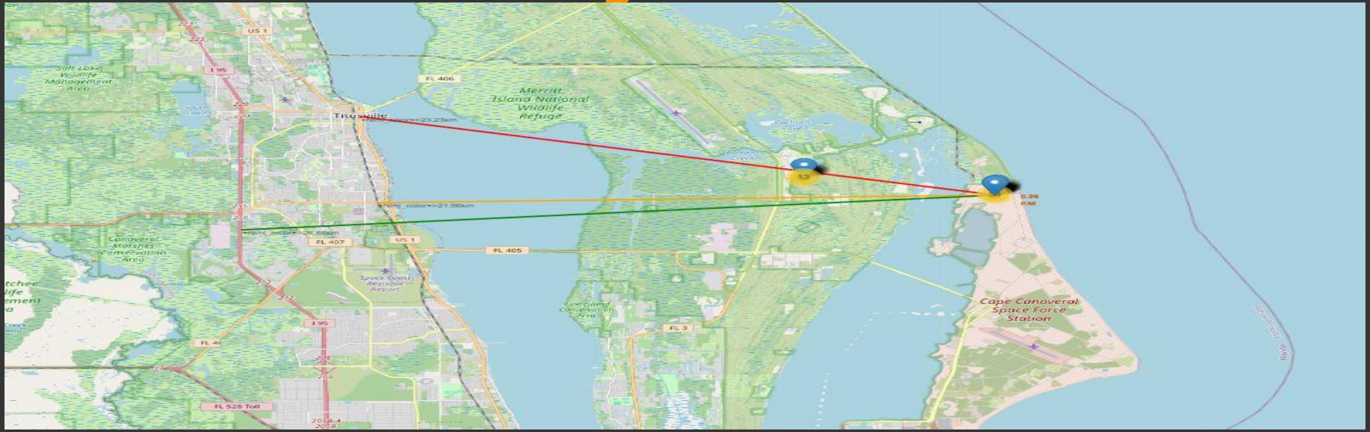


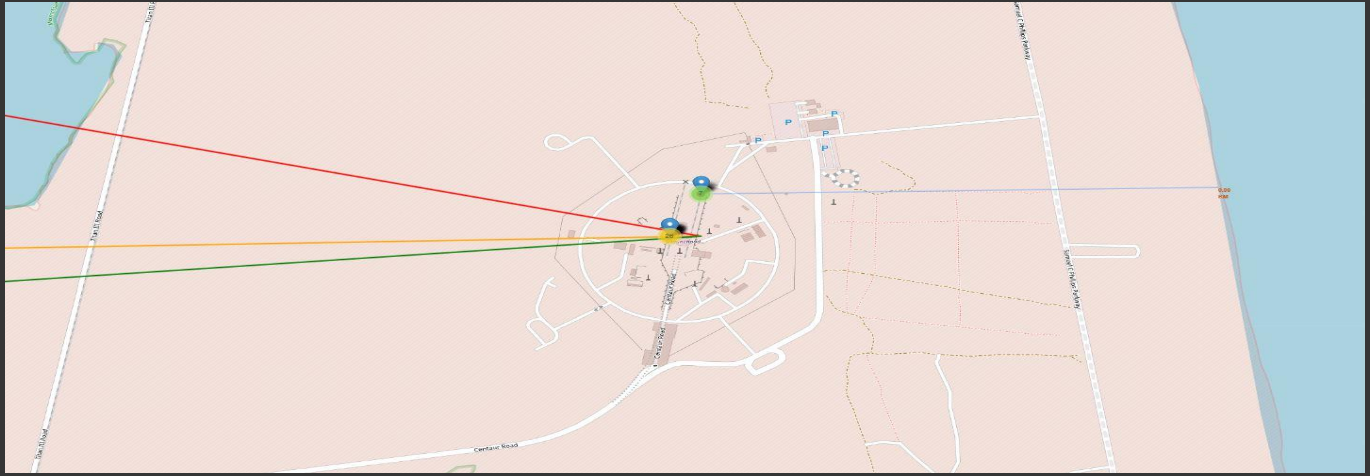

With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

**Insight:** From this graph we can see the payload mass for each orbit type.

# Visual Analysis: Folium Map




## Insight:

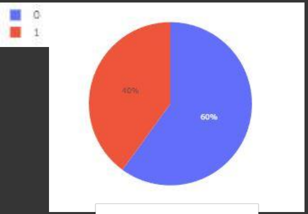
- Proximity to the Equator:

The closer the launch site to the equator, the easier it is to launch to equatorial orbit, and the more help you get from Earth's rotation for a prograde orbit.

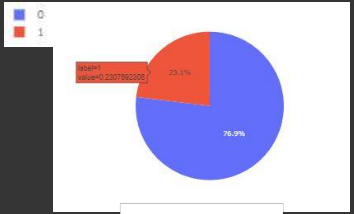
- Safety:

Coasts help ensure that spent stages dropped along the launch path or failed launches don't fall on people or property.

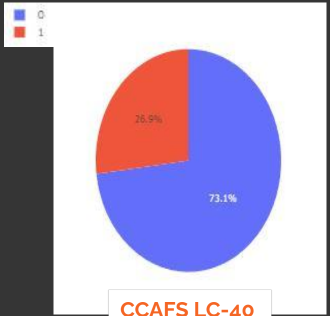
# Visual Analysis: Dash Visuals



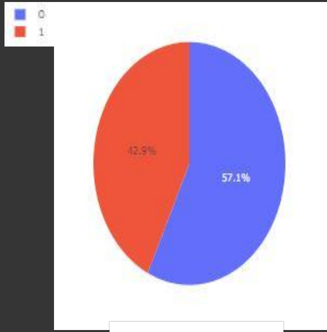

VAFB SLC-4E



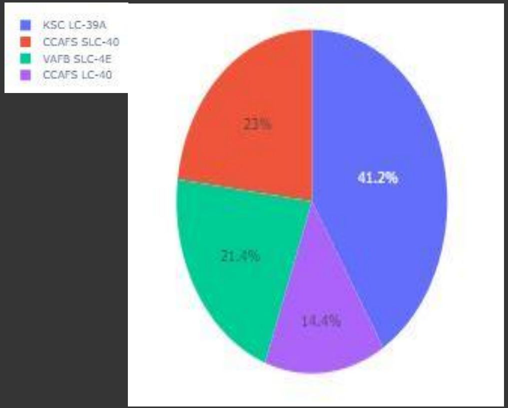

KSC LC-39-A




CCAFS LC-40



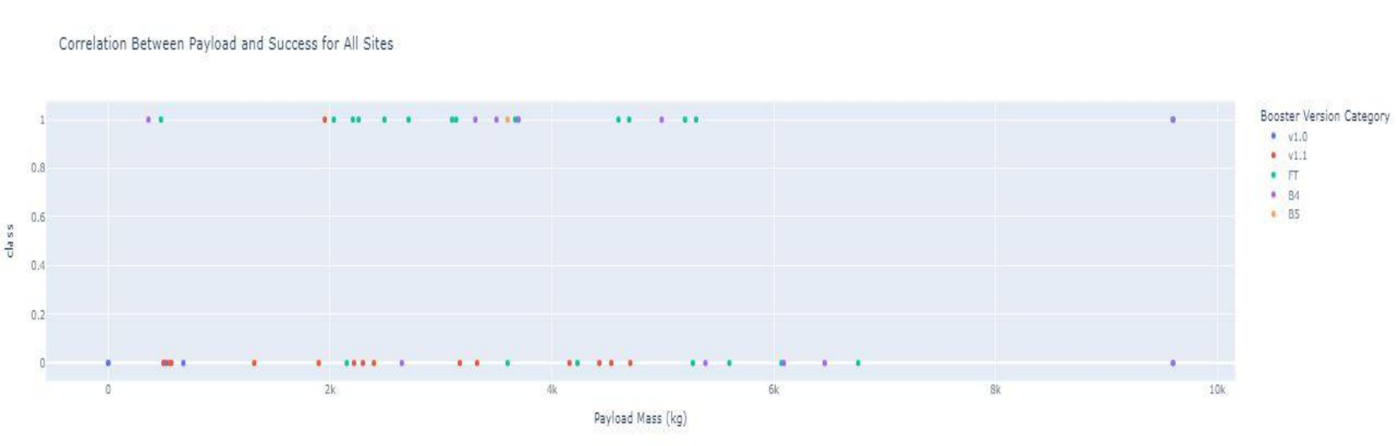

CCAFS SLC-40



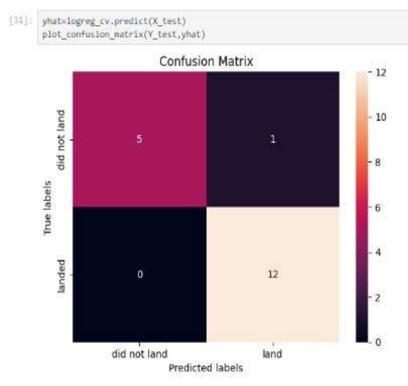

## Insight:

Each of these pie charts how the success vs failure rate of each launch site.

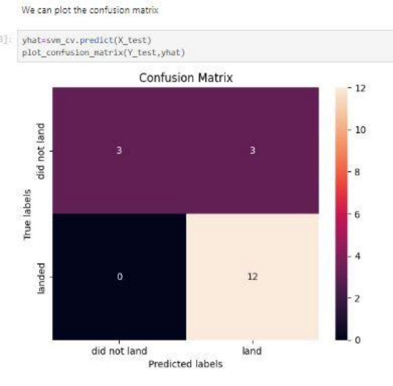
The Scatterplot shows the payload mass vs success rate of each boost version.



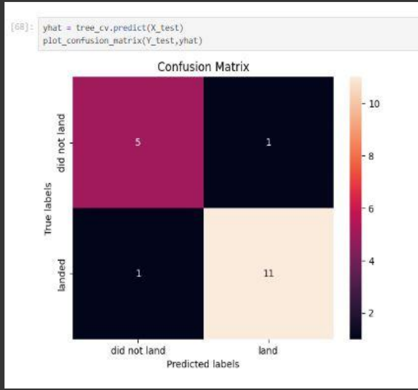
# Predictive Analysis: Machine Learning Model Results



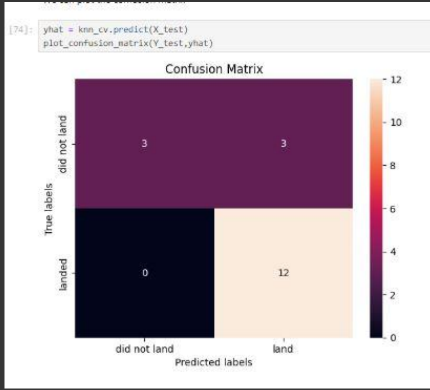

Logistic Regression




Support Vector Machine




Decision Tree Classifier




K Nearest Neighbors

## TASK 12

Find the method performs best:

```
[75]: print('Accuracy for Logistics Regression method:', logreg_cv.score(X_test, Y_test))
print( 'Accuracy for Support Vector Machine method:', svm_cv.score(X_test, Y_test))
print('Accuracy for Decision tree method:', tree_cv.score(X_test, Y_test))
print('Accuracy for K nearsdt neighbors method:', knn_cv.score(X_test, Y_test))

Accuracy for Logistics Regression method: 0.9444444444444444
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.8888888888888888
Accuracy for K nearsdt neighbors method: 0.8333333333333334
```

## Insight:

**Examining the confusion matrix, we see that the major problem is false positives for each of the models.**

**Checking the accuracy for the models, logistic regression had the highest accuracy score, thus making it the best method for predictive analysis in this case.**

# Conclusion

In conclusion, SpaceX's pursuit of rocket reusability stands as a testament to innovation and progress in space exploration. Drawing from the analysis, several key takeaways emerge:Proximity to the equator significantly boosts first stage landing success rates, leveraging Earth's rotation for efficient prograde orbits.

- Heavier payloads, particularly those bound for Polar, Low Earth Orbit (LEO), and the International Space Station (ISS), exhibit higher rates of successful first stage landings.
- Geostationary Transfer Orbit (GTO) missions present a nuanced challenge, with both positive and negative outcomes observed.
- These insights not only inform current operations but also pave the way for future advancements. As SpaceX continues to drive forward with its mission to revolutionize space travel, these findings serve as crucial guideposts, propelling us closer to a future where access to space is not only feasible but sustainable.

These insights not only inform current operations but also pave the way for future advancements. As SpaceX continues to drive forward with its mission to revolutionize space travel, these findings serve as crucial guideposts, propelling us closer to a future where access to space is not only feasible but sustainable.