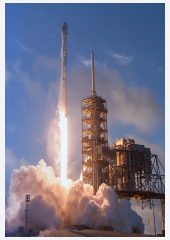
Winning Space Race with Data Science

Lorena Flores
October 2024

Outline

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

Executive Summary



- Summary of Methodologies

  - Data collection through Spacex rest API and web scraping.

  - Data wrangling.

  - Exploratory data analysis (EDA) using visualization and SQL.

  - Interative visual analytics using Folium and Plotly Dash.

  - Predictive analysis.

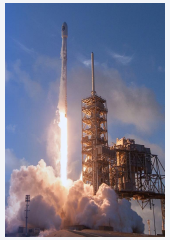
- Summary of Results

  - Collected valuable data from public sources.

  - Identified key features through EDA to predict the success of launches.

  - Machine Learning identified the optimal model to predict which characteristics are crucial for maximizing this opportunity, using all collected data.

Introduction



- Project background and context

SpaceX advertises Falcon 9 rocket launches on its website at a cost of 62 million dollars. In contrast, other providers charge upwards of 165 million dollars per launch. Much of the savings come from SpaceX's ability to reuse the first stage of the rocket. Therefore, if we can predict whether the first stage will land successfully, we can estimate the cost of a launch. This information can be valuable for alternative companies looking to bid against SpaceX for rocket launches.

- Problems to find answers

- What are the conditions for a successful landing?

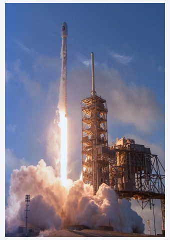
- How do different variables affect the success rate?

- What conditions must SpaceX achieve to maximize the rocket's success landing rate?

Section 1

Methodology

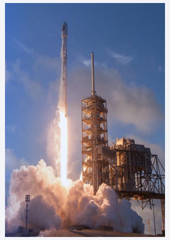
Methodology



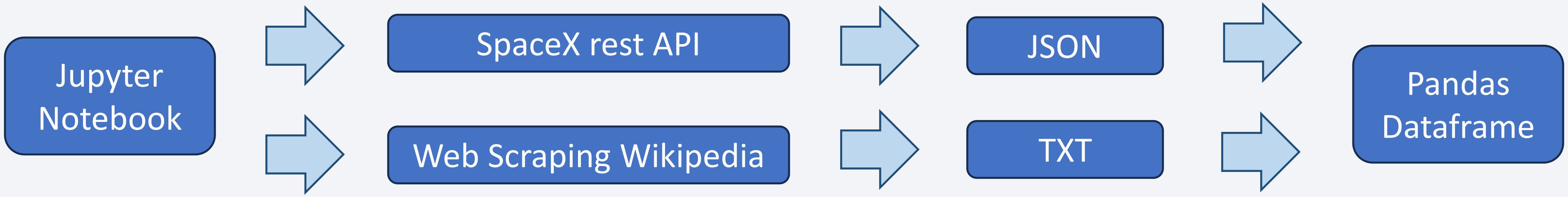
Executive Summary

- Data collection methodology
  - Using SpaceX rest API and web scraping from Wikipedia.
- Perform data wrangling
  - Data was cleaned by removing irrelevant columns and transformed using one-hot encoding for Machine Learning.
- 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.

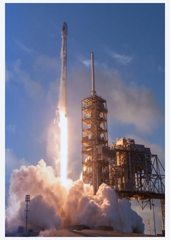
Data Collection



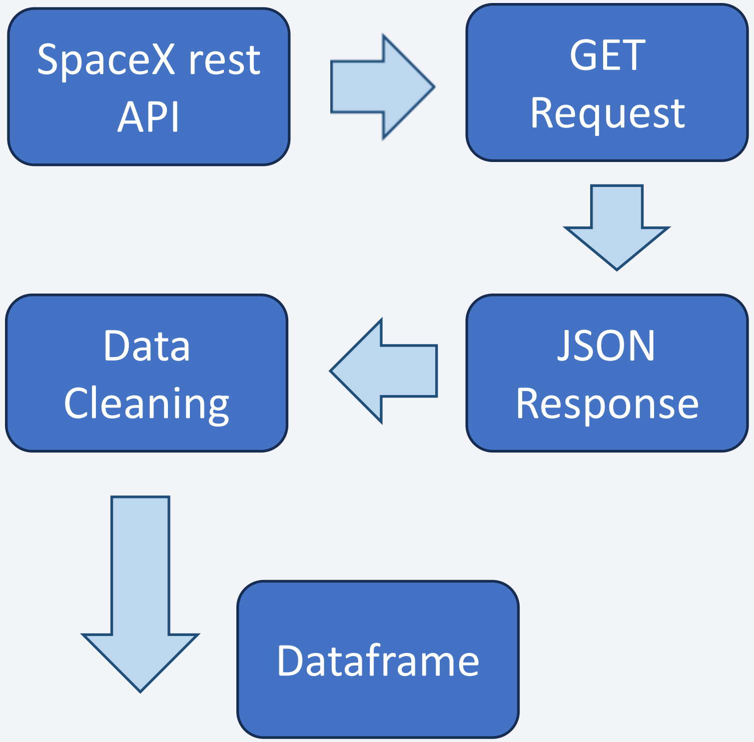
- Data was gathered using the SpaceX REST API, which provides information about launches, including rocket details, payload, launch and landing specifications, and outcomes. The goal is to predict whether SpaceX will attempt to land a rocket. The API endpoints start with (api.spacexdata.com/v4/rockets/). Additionally, Falcon 9 launch data was obtained by web scraping Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches).
- Data collection process



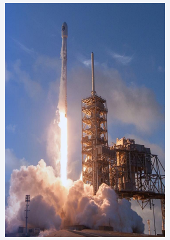
Data Collection – SpaceX API



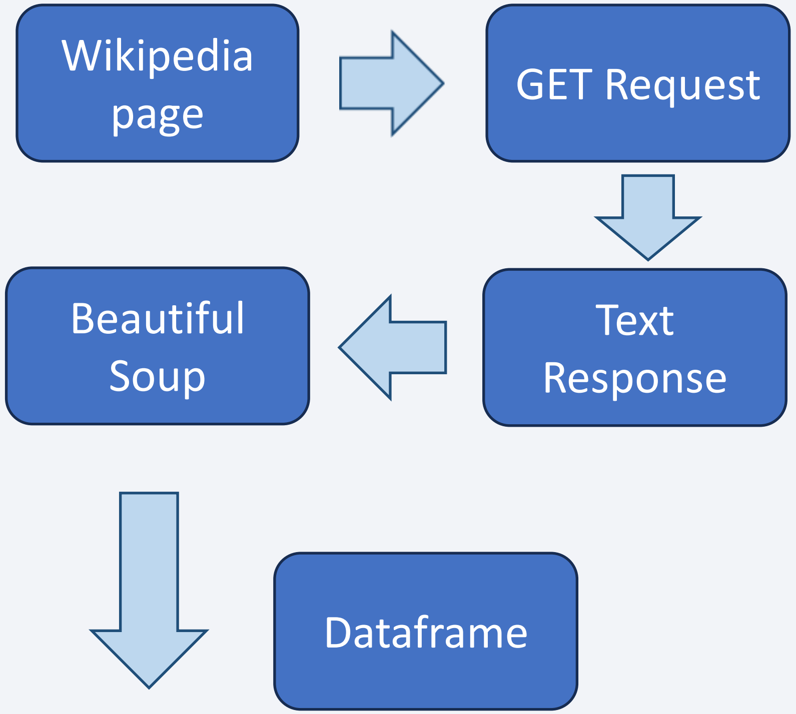
- Data collection from the SpaceX rest API begins by importing essential libraries such as pandas, NumPy and Requests. A GET request is then made to the API URL, which returns data in JSON format. This JSON data is converted into a DataFrame by extracting relevant information, including geospatial data, rocket type, orbit, flight number, and more.
- GitHub Link



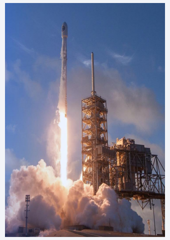
Data Collection - Scraping



- The required Python libraries were imported. The data source was a the Wikipedia page ‘SpaceX Falcon 9 First Stage Landing Prediction’. An HTTP GET request was initialized and the response was received in text format. The Beautiful Soup library was then used to extract the tables and columns from the text response, which were later converted into a Pandas Dataframe.
- GitHub Link



Data Wrangling



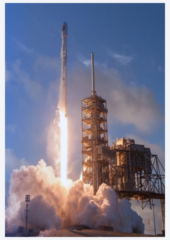
- Process

- Import Libraries and Load Dataset
- Identify and Handle Missing Values
- Identify Data Types
- Calculate Number of Launches at Each Site
- Calculate Number and Occurrence of Each Orbit
- Calculate Number and Occurrence of Mission Outcomes
- Create a Set of Unsuccessful Outcomes
- Create Landing Outcome Labels
- Add a 'Class' Column
- Calculate Success Rate
- Export Data

- GitHub Link

1. Data Discovery
2. Data Structuring
3. Data Cleaning
4. Data Enriching
5. Data Validating
6. Data Publishing

EDA with Data Visualization



* Scatter plots to show correlations:

- Flight Number vs Payload Mass

- Flight Number vs Launch Site

- Payload Mass vs Launch Site

- Flight Number vs Orbit Type

- Payload Mass vs Orbit Type.

* Bar Graph to compare:

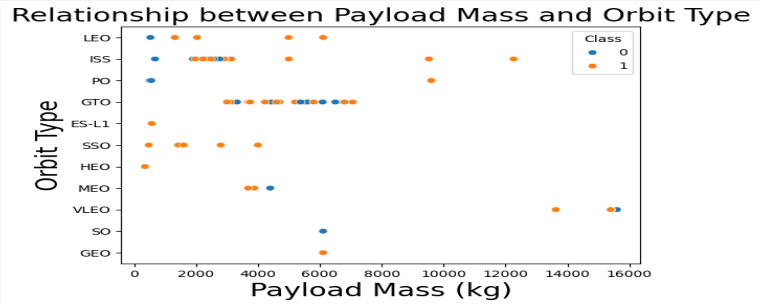
Success Rate by

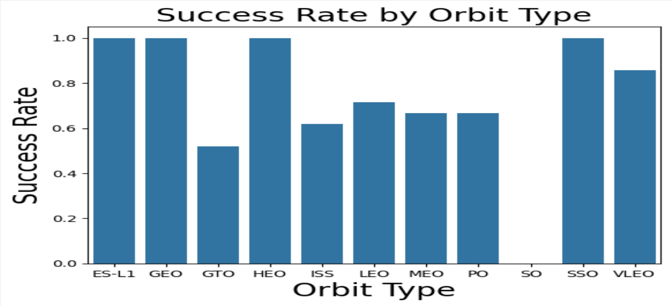
Orbit Type.

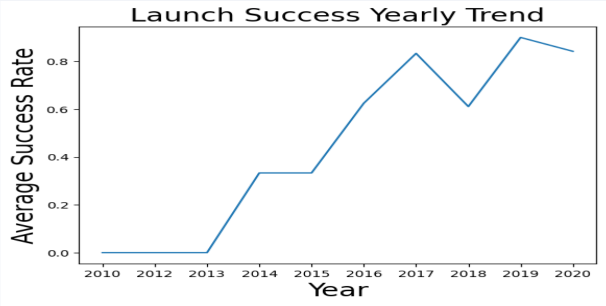
* Line Graph to show:

Launch Success

YearlyTrend.

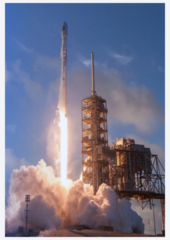






• GitHub Link

EDA with SQL



- SQL queries performed:

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

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

- Display the total payload mass carried by boosters launched by NASA (CRS) EDA WITHSQL 12.

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

- List the date when the first successful landing outcome in ground pad was achieved.

- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 - Listing the total number of successful and failure mission outcomes.

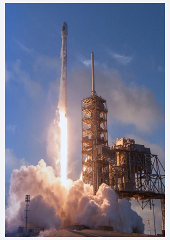
- List the names of the booster versions which have carried the maximum payload mass.

- List the records which will display the month names, failure landing_outcomes in drone ship, booster versions, and launch_site for the months in year 2015.

- 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.

- GitHub Link

Build an Interactive Map with Folium



- Map objects created and added to a folium map

- Marked all launch sites and added map objects such as markers, circles, and lines to indicate the success or failure of launches at each site.

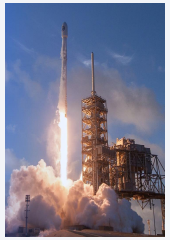
- Assigned launch outcomes to classes: 0 for failure and 1 for success.

- Used color-labeled marker clusters to identify launch sites with relatively high success rates.

- Calculated the distances between each launch site and its proximities.

- GitHub Link

Build a Dashboard with Plotly Dash



- Plots/graphs and interactions added to a dashboard:

- Added a dropdown list to enable Launch Site selection, including the following options: All Sites, CCAFS LC-40, CCAFS SLC-40, VAFB SLC-4E, KSC LC-39A.

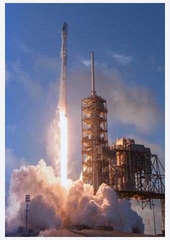
- Added a pie chart to show the total successful launches count for all sites.

- Added a slider to select payload, ranging from 0 to 10,000.

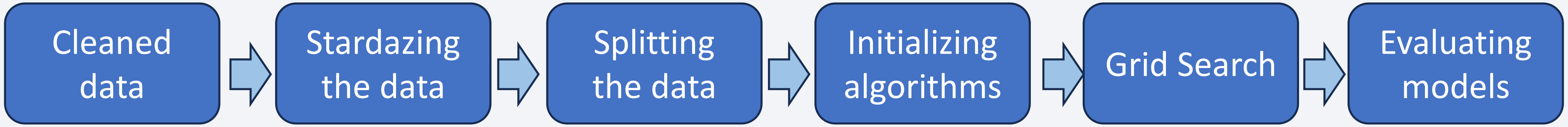
- Added a scatter chart to show the correlation between payload and launch success.

- [GitHub Link]

Predictive Analysis (Classification)

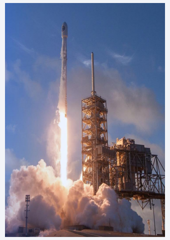


- Loading the cleaned data.
- Standardizing the data to prevent bias.
- Splitting the data: 80% for training and 20% for testing.
- Initializing different classification algorithms: Logistic Regression, Support Vector Machine, Decision Tree and K-Nearest Neighbors.
- Using Grid Search to find the best parameters.
- Evaluating the models using: Confusion Matrix, F1 Score and Jaccard Score



- GitHub Link

Results



- Exploratory Data Analysis (EDA) results

- SpaceX uses four different launch sites.

- The first launches were conducted for SpaceX itself and NASA.

- The average payload of the F9 v1.1 booster is 2,928 kg.

- The first successful landing outcome occurred in 2015, five years after the first launch.

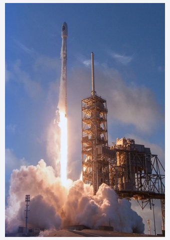
- Many Falcon 9 booster versions successfully landed on drone ships with payloads above the average.

- Almost 100% of mission outcomes were successful.

- Two booster versions failed to land on drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015.

- The number of successful landing outcomes improved over the years.

Results



- Interactive analytics demo in screenshots

- It was possible to identify that launch sites are typically located in safe areas, near the sea, and have good logistic infrastructure around them.

- Most launches happen at East Coast launch sites.

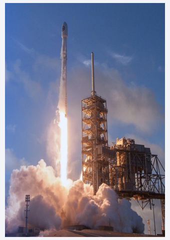
- Predictive analysis results

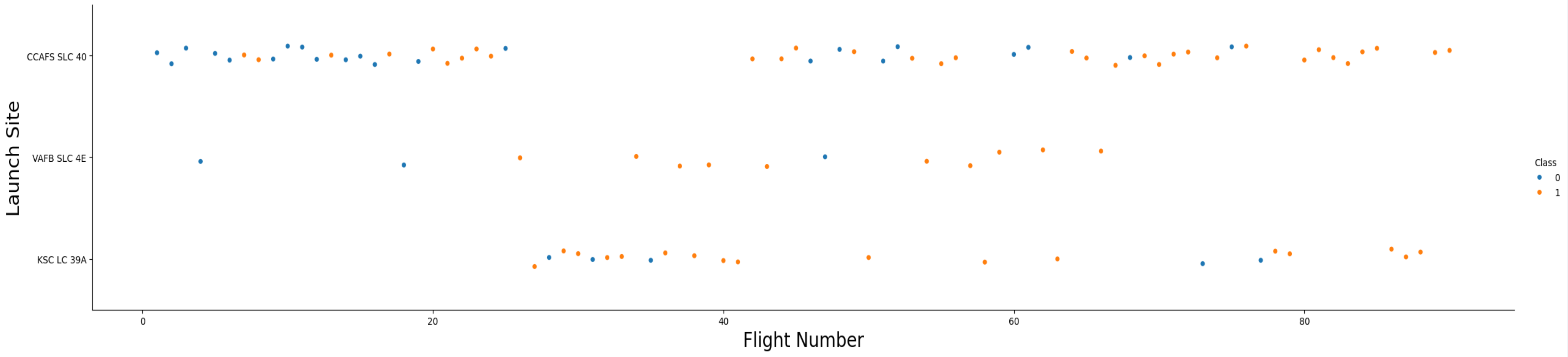
- Predictive analysis showed that the Decision Tree Classifier is the best model to predict successful landings, with an accuracy of over 87% and an accuracy for test data of over 94%.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site





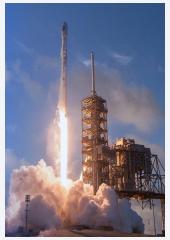
- **Analysis of Launch Sites:**

- CCAFS SLC 40: This is the most frequently used site for launching SpaceX's rockets, with 55 trials. Out of these, 33 were successful, and 22 failed, resulting in a 60% success rate.

- VAFB SLC 4E: This is the least frequently used site for launching SpaceX's rockets, with 13 trials. Out of these, 10 were successful, and 3 failed, resulting in a 77% success rate.

- KSC LC 39A: This is a moderately used site for launching SpaceX's rockets, with 22 trials. Out of these, 17 were successful, and 5 failed, resulting in a 77% success rate.

Payload vs. Launch Site



- **Analysis of Payload Mass and Launch Sites:**

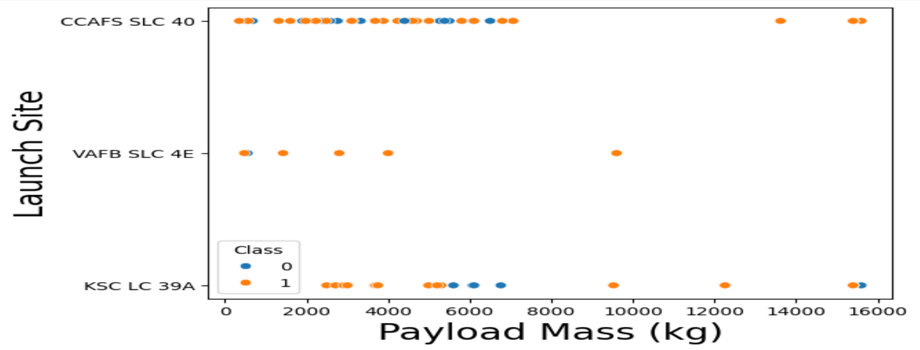
\- CCAFS SLC 40 and Payload Mass:

1. As the payload mass increases, the likelihood of the launch site being CCAFS SLC 40 also increases.
2. The heaviest payload ever launched from this site was 15,000 kilograms.
3. CCAFS SLC 40 is the preferred site for launching heavy payloads.

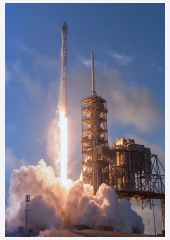
\- VAFB SLC 4E and Heavy Payloads:

1. At the VAFB SLC 4E launch site, there have been no rockets launched with a payload mass greater than 10,000 kilograms.
2. VAFB SLC 4E is not used for launching heavy payloads.

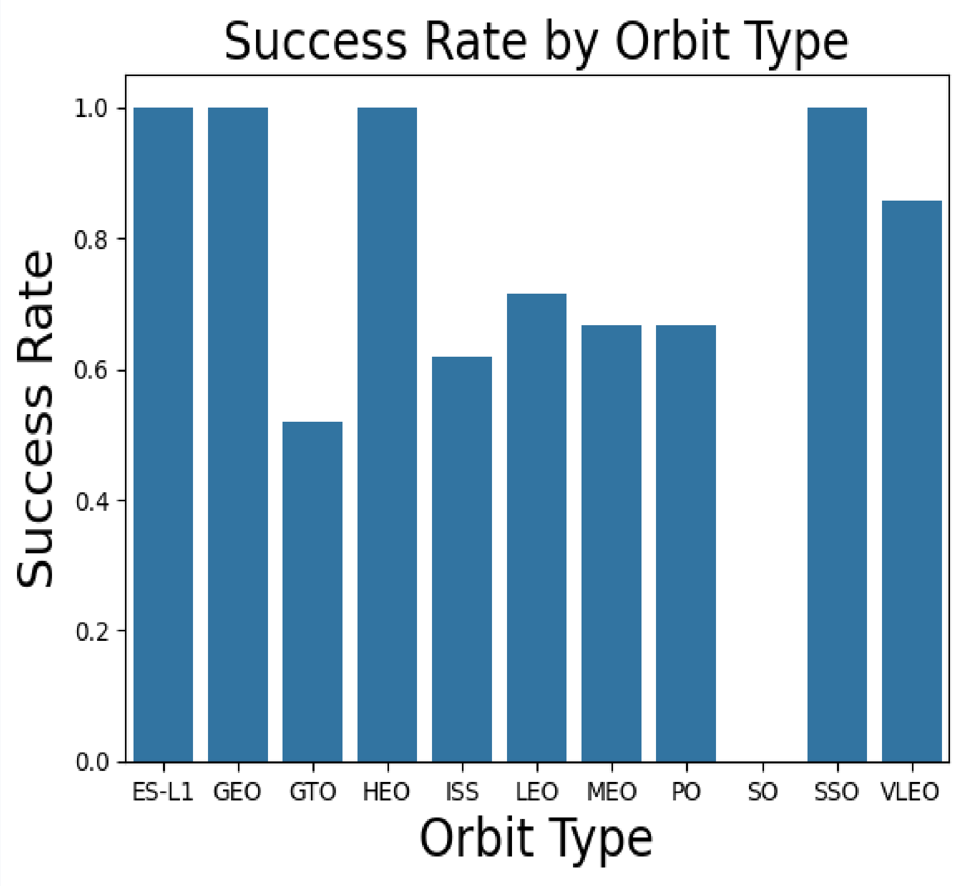
\- Overall Preference: CCAFS SLC 40 is the go-to launch site for SpaceX when it comes to heavy payloads, due to its capability to handle larger masses.



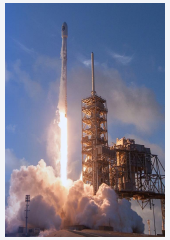
Success Rate vs. Orbit Type



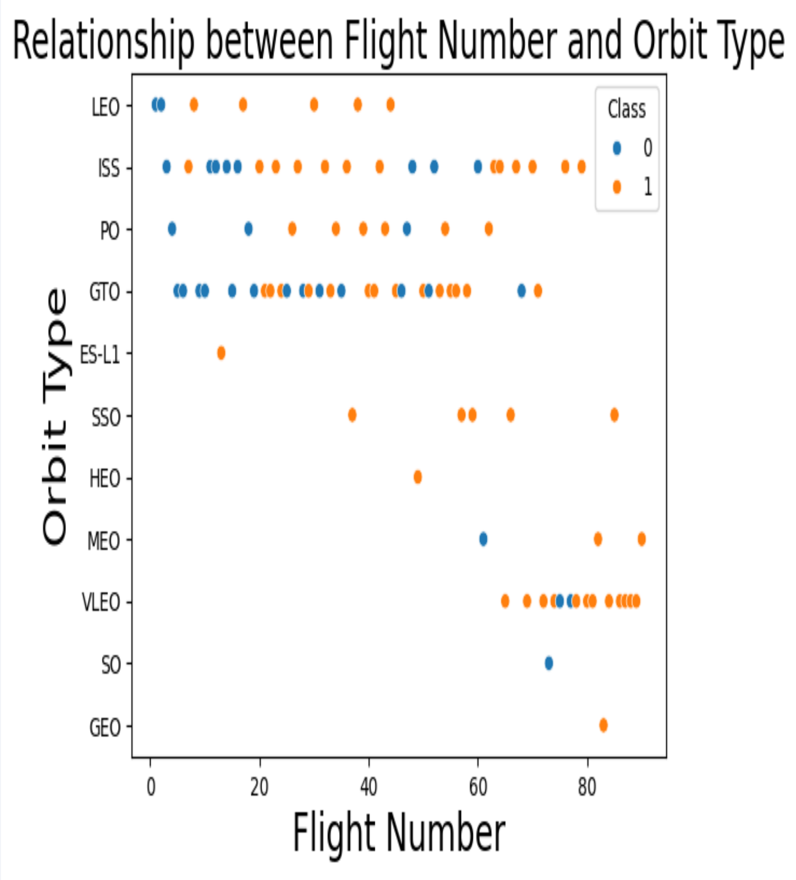
- ES-L1, GEO, HEO, and SSO Orbits: All first stages launched into these orbits have successfully landed, demonstrating a perfect landing record.
- SO Orbit: There has never been a successful landing from this orbit, indicating challenges or limitations specific to it.
- Other Orbits: The landing success rates for first stages launched into other orbits range between 50% and 90%, showing variability in landing outcomes depending on the orbit type.



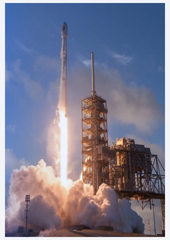
Flight Number vs. Orbit Type



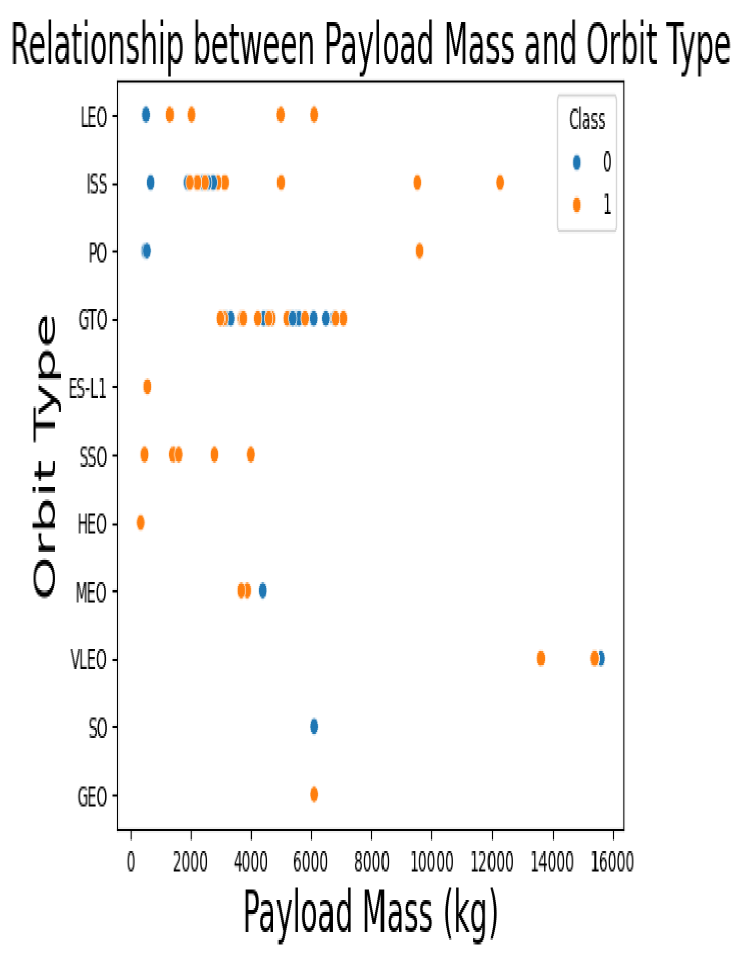
- **LEO Orbit**: Clear relationship between the number of flights and success rate. As the number of flights increases, the success rate also tends to be higher.
- **GTO Orbit**: No correlation between the number of flights and success rate. The success rate seems to have improved over time across all orbits.
- **VLEO Orbit**: shows potential as a new business opportunity due to the recent increase in its usage frequency. It also has a high success rate.
- **High Flight Numbers**: GEO, SO, VLEO, MEO, HEO, and SOO tend to have higher flight numbers.
- **Utilization of Orbits**: VLEO orbit has a high success rate. LEO and GTO orbits are heavily utilized, with numerous flights ranging from 0 to over 80.
- **SO and GEO Orbits**: have very limited data points, suggesting that missions targeting these orbits are less common or have specific purposes.
- **PO Orbit**: has a moderate number of flights, with nine data points ranging from 0 to 60, indicating that missions requiring polar orbits are conducted relatively frequently.



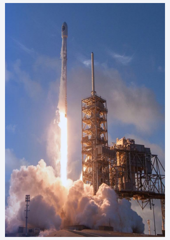
Payload vs. Orbit Type



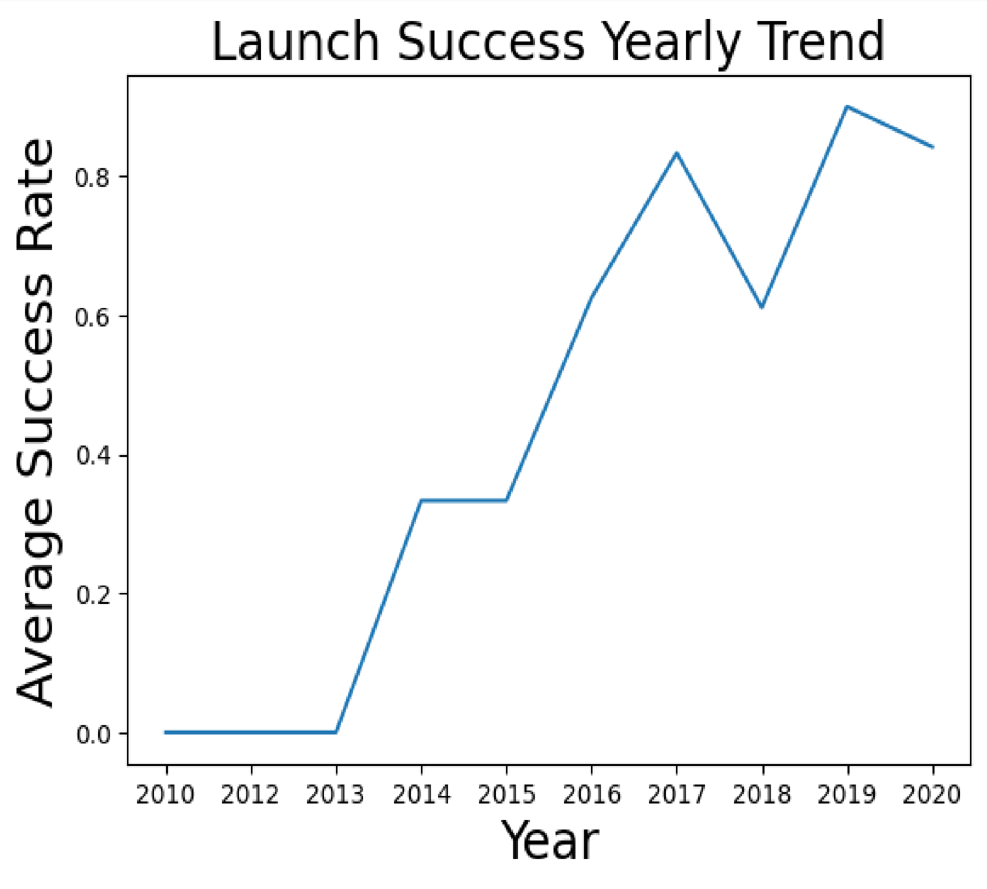
- **Heavy Payloads**: They have a negative impact on GTO orbits but positively influence Polar LEO (ISS) orbits. There is no clear relationship between payload mass and success rate for GTO orbits.
- **ISS Orbit**: supports the widest range of payloads and boasts a high success rate.
- **SO and GEO Orbits**: see fewer launches. For heavy payloads, successful landings are more frequent in PO, LEO, and ISS orbits. Both ISS and LEO orbits show higher success rates as payload mass increases.
- **SSO Orbit**: has a higher success rate with lower payload masses. GTO orbits accommodate a wide range of payload masses.
- **LEO Orbit**: has a relatively lower payload capacity.
- **ES-L1, HEO, SSO Orbits**: have specific mission requirements, reflected in their limited payload range.



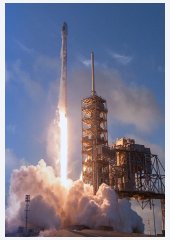
Launch Success Yearly Trend



- The success rate remains flat between 2010-2013.
- In 2014, there is a noticeable increase in the success rate.
- In 2015, there is a substantial upward trend in the success rate.
- In 2018, there is a slight decline in the success rate from the high point in 2017.
- In 2019, there is a notable recovery.
- In 2020, the success rate remains close to 0.8, suggesting that SpaceX has reached a stable state of high mission success.



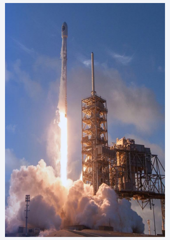
All Launch Site Names



- The keyword “DISTINCT” was used to display only unique launch sites from the SpaceX data.

```
# Ejecutar la consulta SQL
cur.execute('''
    SELECT DISTINCT "Launch_Site"
    FROM SPACEXTABLE;
''')

# Obtener los resultados
launch_sites = cur.fetchall()

# Mostrar los resultados
for site in launch_sites:
    print(site)

# Cerrar la conexión
con.close()
```

```
('CCAFS LC-40',)
('VAFB SLC-4E',)
('KSC LC-39A',)
('CCAFS SLC-40',)
```



Launch Site Names Begin with 'CCA'

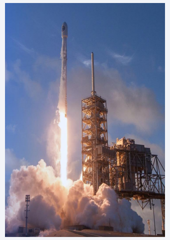
- The query below was used to display 5 records where launch sites begin with ‘CCA’.

```python
# Ejecutar la consulta SQL
cur.execute('''
    SELECT *
    FROM SPACEXTABLE
    WHERE "Launch_Site" LIKE 'CCA%'
    LIMIT 5;
''')

# Obtener los resultados
records = cur.fetchall()

# Mostrar los resultados
for record in records:
    print(record)

# Cerrar la conexión
con.close()
```

Python

```
('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',
('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

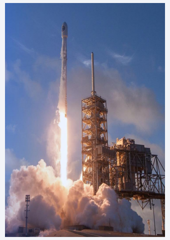


- SELECT SUM(PAYLOAD_MASS__KG_) AS TotalPayloadMass: calculates the total payload mass (in kilograms) from the SPACEXTABLE and assigns it the alias TotalPayloadMass.
- FROM SPACEXTABLE: Specifies that the data is being retrieved from the table named SPACEXTABLE.
- WHERE “Customer” = ‘NASA (CRS)’: Filters the results to include only records where the customer is ‘NASA (CRS)’, referring to missions associated with NASA’s Commercial Resupply Services.

```
# Ejecutar la consulta SQL
cur.execute('''
    SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass
    FROM SPACEXTABLE
    WHERE "Customer" = 'NASA (CRS)';
''')

# Obtener el resultado
total_payload_mass = cur.fetchone()[0]

# Mostrar el resultado
print("Total Payload Mass carried by NASA (CRS) boosters:", total_payload_mass, "kg")

# Cerrar la conexión
con.close()
```

```
Total Payload Mass carried by NASA (CRS) boosters: 45596 kg
```

Average Payload Mass by F9 v1.1

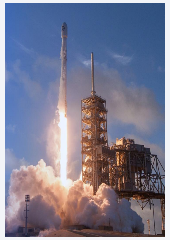


- To find the average Payload Mass carried by Booster Version F9 v1.1, the avg() function is used to return the Average Payload Mass, and the condition “Booster_Version” = ‘F9 v1.1’ is applied to filter the records by booster version.

```
# Ejecutar la consulta SQL
cur.execute('''
    SELECT AVG("Payload_Mass__kg_") AS Average_Payload_Mass
    FROM SPACEXTABLE
    WHERE "Booster_Version" = 'F9 v1.1';
''')

# Obtener el resultado
average_payload_mass = cur.fetchone()[0]

# Mostrar el resultado
print("Average Payload Mass carried by booster version F9 v1.1:", average_payload_mass, "kg")

# Cerrar la conexión
con.close()
```

```
Average Payload Mass carried by booster version F9 v1.1: 2928.4 kg
```

First Successful Ground Landing Date

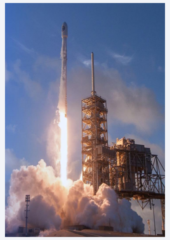


- To find the earliest successful ground pad landing date, the MIN() function is used to select the minimum date from the Date column, assigning it the alias First_Successful_Landing. Data is retrieved from the SPACEXTABLE table, and the results are filtered to include only records where the landing outcome is 'Success (ground pad)'.

```
# Ejecutar la consulta SQL
cur.execute('''
    SELECT MIN("Date") AS First_Successful_Landing
    FROM SPACEXTABLE
    WHERE "Landing_Outcome" = 'Success (ground pad)';
''')

# Obtener el resultado
first_successful_landing = cur.fetchone()[0]

# Mostrar el resultado
print("Date of the first successful landing on ground pad:", first_successful_landing)

# Cerrar la conexión
con.close()
```

```
Date of the first successful landing on ground pad: 2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

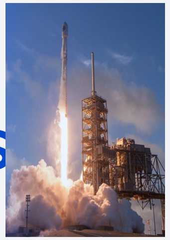


- The WHERE clause is used to filter for boosters that have successfully landed on a drone ship. The AND condition is applied to determine successful landings with a payload mass greater than 4000 but less than 6000.

```
# Ejecutar la consulta SQL
cur.execute('''
    SELECT "Booster_Version"
    FROM SPACEXTABLE
    WHERE "Landing_Outcome" = 'Success (drone ship)'
      AND "Payload_Mass__kg_" > 4000
      AND "Payload_Mass__kg_" < 6000;
''')

# Obtener los resultados
boosters = cur.fetchall()

# Mostrar los resultados
for booster in boosters:
    print(booster)

# Cerrar la conexión
con.close()
```

```
('F9 FT B1022',)
('F9 FT B1026',)
('F9 FT  B1021.2',)
('F9 FT  B1031.2',)
```

Total Number of Successful and Failure Mission Outcomes

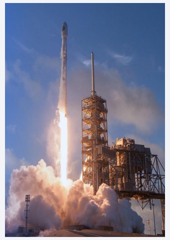


- To find the total number of successful and failed mission outcomes, the count(*) statement is used to retrieve the number of records. The group by statement is applied to group records by mission outcomes.

```
# Ejecutar la consulta SQL
cur.execute('''
    SELECT "Booster_Version"
    FROM SPACEXTABLE
    WHERE "Landing_Outcome" = 'Success (drone ship)'
      AND "Payload_Mass__kg_" > 4000
      AND "Payload_Mass__kg_" < 6000;
''')

# Obtener los resultados
boosters = cur.fetchall()

# Mostrar los resultados
for booster in boosters:
    print(booster)

# Cerrar la conexión
con.close()
```

```
('F9 FT B1022',)
('F9 FT B1026',)
('F9 FT  B1021.2',)
('F9 FT  B1031.2',)
```

Boosters Carried Maximum Payload

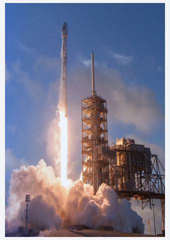


- To find the names of Booster Versions that have carried the maximum Payload Mass, a subquery is used to get the maximum Payload Mass. The max() function is applied to return the largest value of Payload Mass.

```
# Ejecutar la consulta SQL con subconsulta
cur.execute('''
    SELECT "Booster_Version"
    FROM SPACEXTABLE
    WHERE "Payload_Mass__kg_" = (
        SELECT MAX("Payload_Mass__kg_")
        FROM SPACEXTABLE
    );
''')

# Obtener los resultados
booster_versions = cur.fetchall()

# Mostrar los resultados
for version in booster_versions:
    print(version)

# Cerrar la conexión
con.close()
```

```
('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

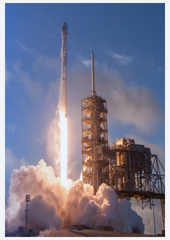


- To find the failed landing outcomes on a drone ship, along with their booster versions and launch site names for the year 2015, the year() function is used to extract the year part of a date.

```
cur.execute('''
    SELECT
        CASE substr("Date", 6, 2)
            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 "Landing_Outcome" = 'Failure (drone ship)'
      AND substr("Date", 0, 5) = '2015';
''')

# Obtener los resultados
records = cur.fetchall()

# Mostrar los resultados
for record in records:
    print(record)

# Cerrar la conexión
con.close()
```

```
('January', 'Failure (drone ship)', 'F9 v1.1 B1012', 'CCAFS LC-40')
('April', 'Failure (drone ship)', 'F9 v1.1 B1015', 'CCAFS LC-40')
```

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



- We selected Landing Outcomes and the COUNT of Landing Outcomes from the data and used the WHERE clause to filter for Landing Outcomes BETWEEN 2010-06-04 to 2010-03-20.
- We applied the GROUP BY clause to group the Landing Outcomes and the ORDER BY clause to order the grouped Landing Outcome in descending order.

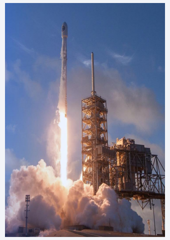
```
# Ejecutar la consulta SQL
cur.execute('''
    SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count
    FROM SPACEXTABLE
    WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
    GROUP BY "Landing_Outcome"
    ORDER BY Outcome_Count DESC;
''')

# Obtener los resultados
landing_outcomes = cur.fetchall()

# Mostrar los resultados
for outcome in landing_outcomes:
    print(outcome)

# Cerrar la conexión
con.close()
```

```
('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

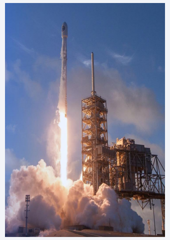
Launch Sites Proximities Analysis

Folium Map: All Launch Sites

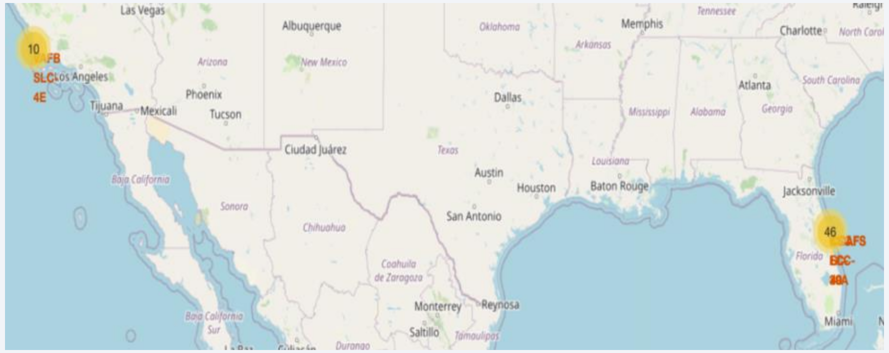


- SpaceX strategically selects locations for its launch sites near water to avoid accidents. The launch sites are located in California and Florida.

Folium Map: Success/failed launches



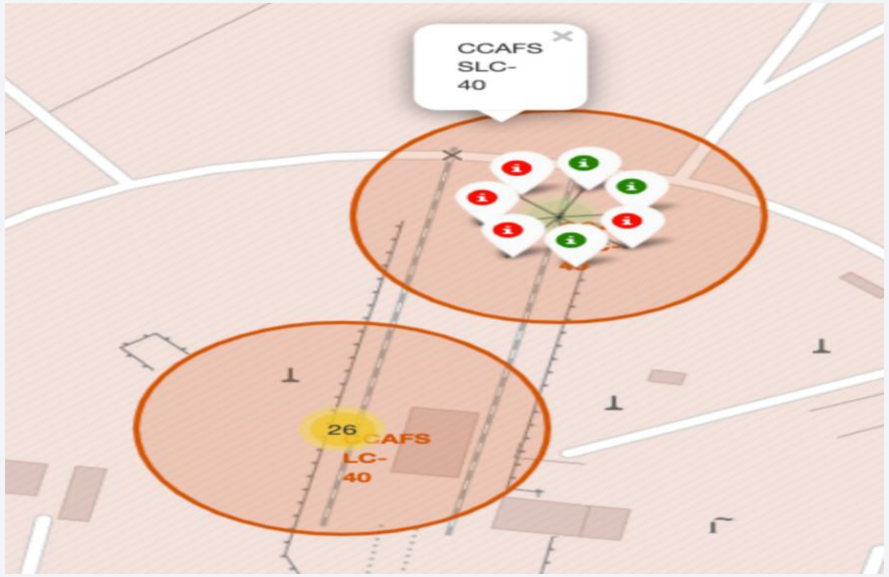
- At the top of the map, yellow is visible, which is a mix of green and red colors, indicating both successful and failed landings in that area.



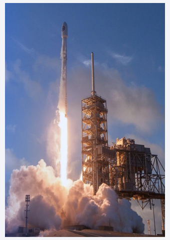
- From the color-labeled markers in the marker clusters, it is easy to identify which launch sites have relatively high success rates:

  - Green Marker: Successful Return.

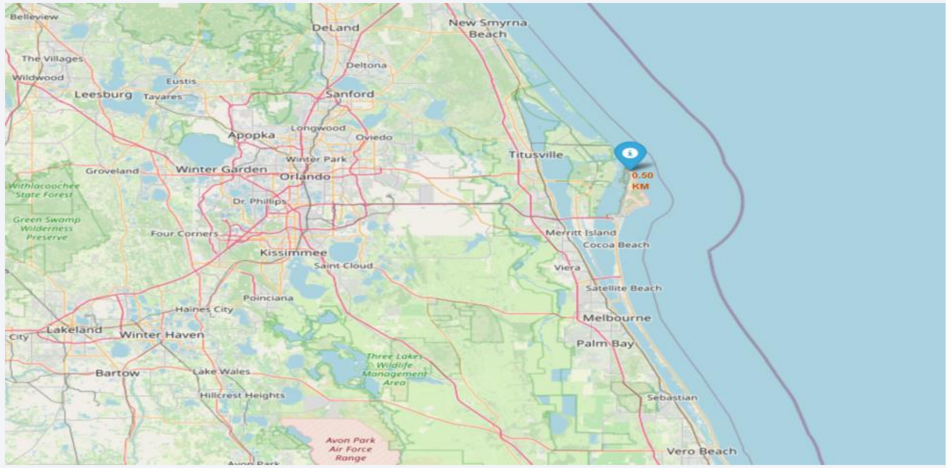
  - Red Marker: Failed Return.

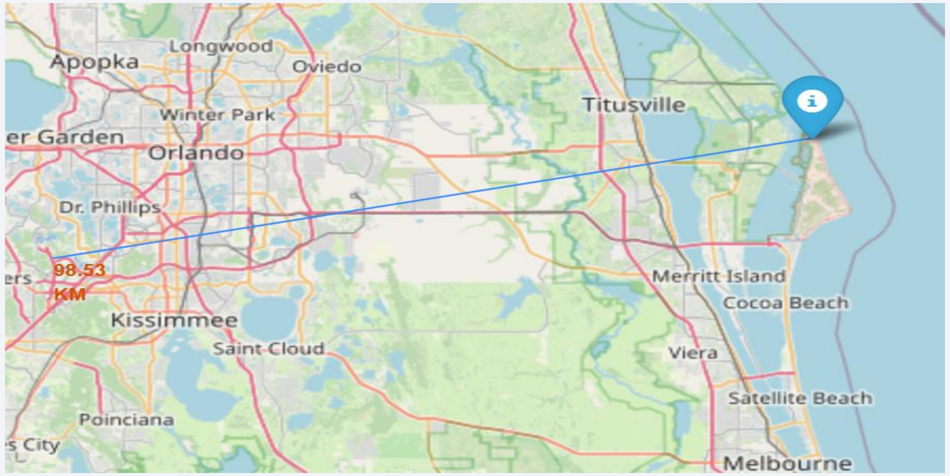


Folium Map: Launch site and its proximities



- The generated Folium map shows a selected launch site near the coastline. The map includes surrounding cities like Orlando and Kissimmee, and major highways connecting them. The coastline is clearly visible to the east.
- The Launch Site Location is strategically placed near the coast to minimize risks during launches.
- The site is close to major highways, facilitating transportation and logistics.
- The coastal location helps in managing safety zones during launches.
- Distance to closest coastline: 0.50 km.
- Distance to closest city: 98.53 km.

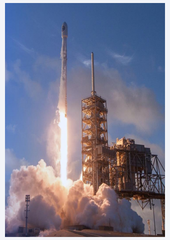




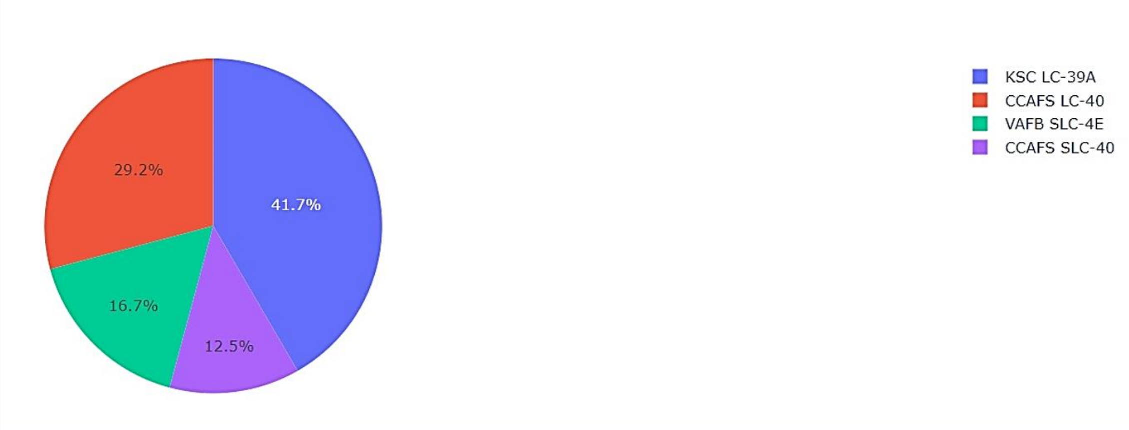
Section 4

Build a Dashboard with Plotly Dash

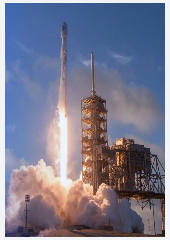
Dashboard: Launch Success count for all sites



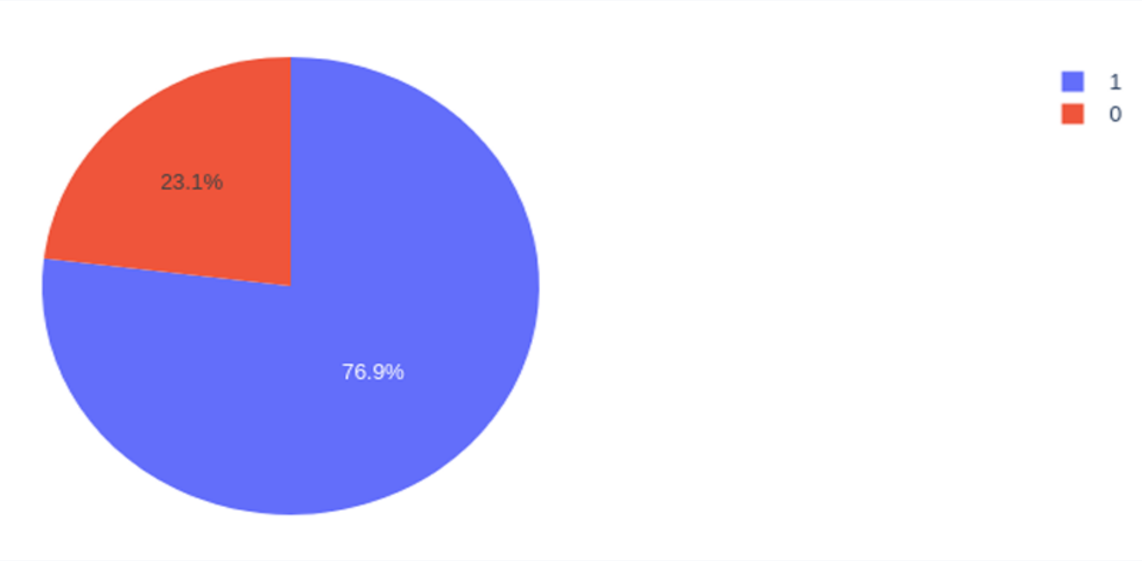
- The graph displays the success percentages for each launch site in terms of first stage returns:
- KSC LC-39A is the top-performing site, with 41.7% of all successful rocket launches.
- CCAFS SLC-40 has the lowest success rate, with only 12.5% of successful launches.
- The pie chart highlights that KSC LC-39A is the most successful launch site, occupying nearly half of the chart. In contrast, CCAFS SLC-40 is the least successful, with just 12.5% success.



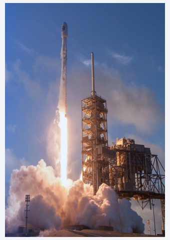
Dashboard: Launch Site with highest Launch Success Ratio



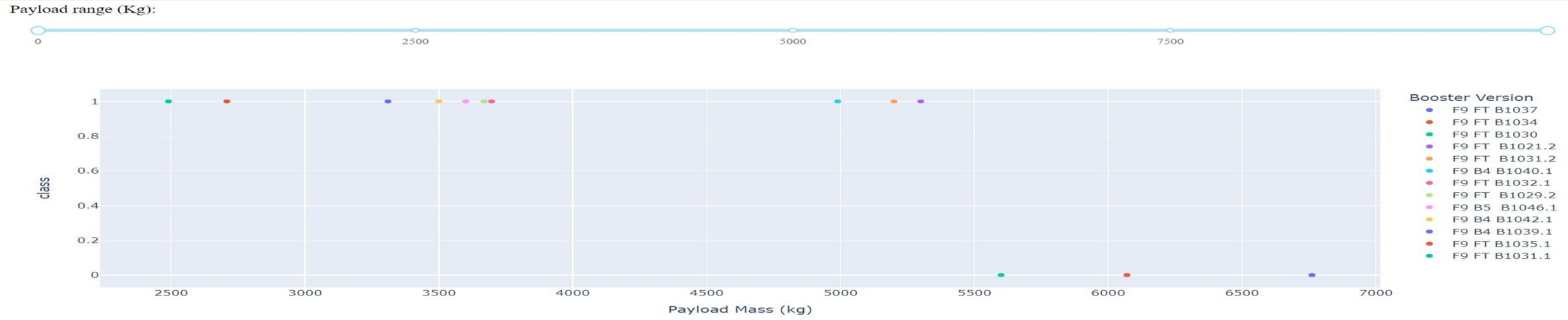
- The pie chart clearly shows that KSC LC-39A has the highest success ratio. The success rate is represented in blue at 76.9%, while the failure rate is shown in red at 23.1%.



Dashboard: Payload vs Launch Outcome for all sites



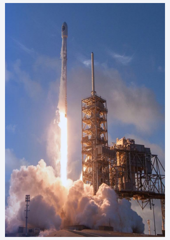
- This interactive scatter plot shows that payloads under 4000 kg are more likely to result in successful launches. The booster version with the highest success rate is the Falcon 9 Full Thrust (F9 FT).
- The payload range with the highest success rate is 2500-5000 kg.
- The booster version with the lowest success rate is the Falcon 9 Block 4 (F9 B4).
- The payload range with the lowest success rate is 0-2500 kg.



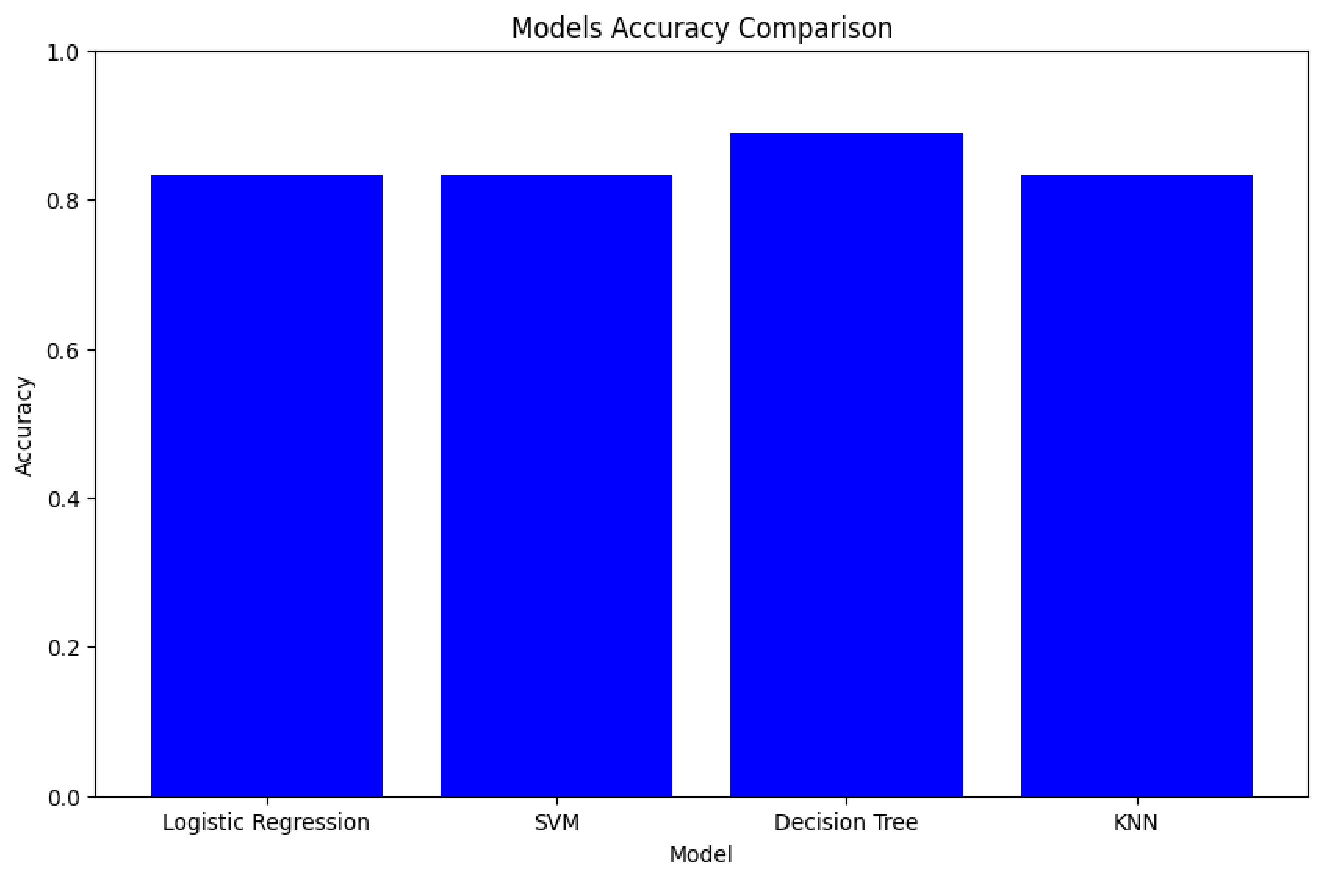
Section 5

Predictive Analysis (Classification)

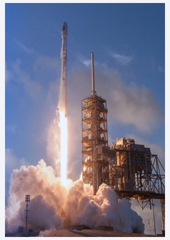
Classification Accuracy



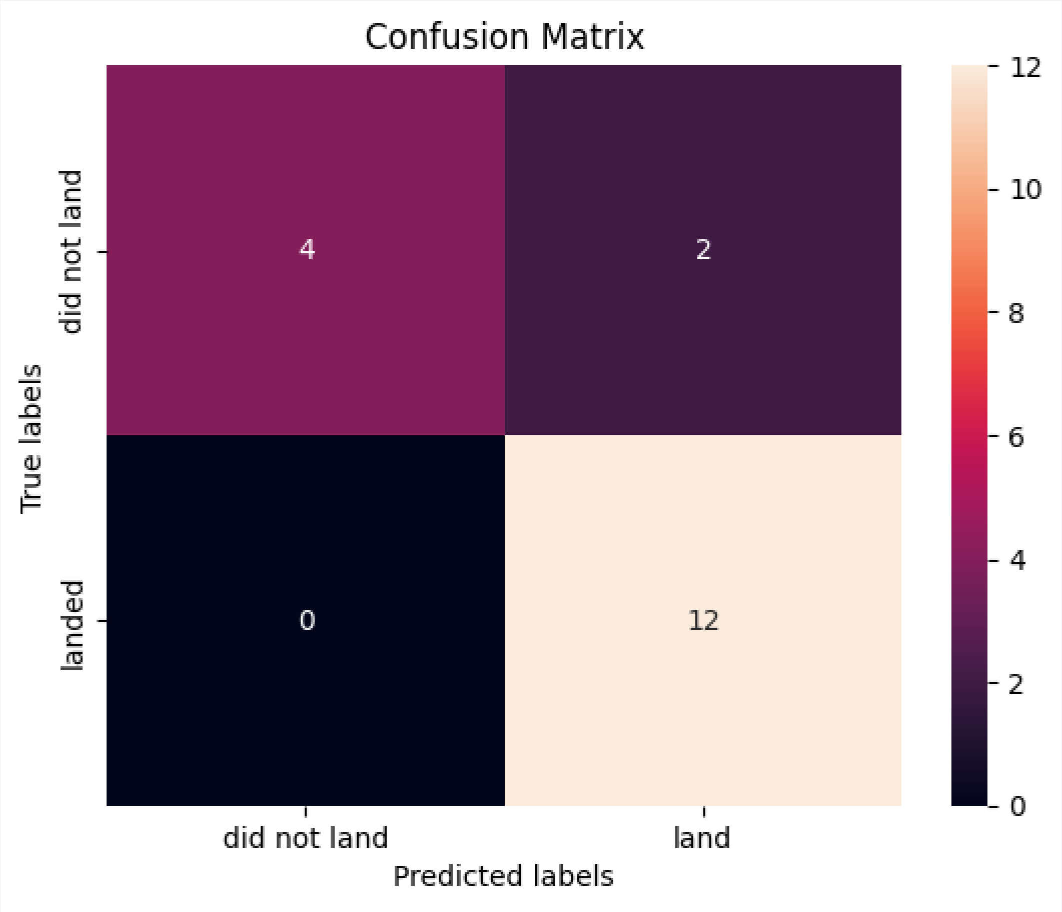
- The decision tree classifier is the model with the highest classification accuracy



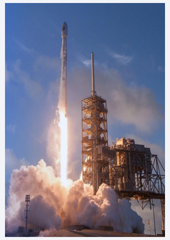
Confusion Matrix



- The best performing method is the Decision Tree Classifier. The Confusion Matrix for the Decision Tree Classifier indicates that the model can effectively distinguish between different classes. However, a significant issue is the occurrence of false positives, where unsuccessful landings are incorrectly classified as successful landings by the model.



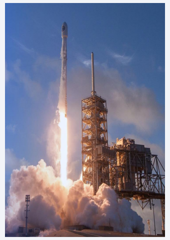
Conclusions I



1. **Key Findings**

- Top Launch Site: KSC LC-39A.
- Lower Risk: Launches above 7,000 kg.
- Improvement: Successful landings have improved over time.
- Prediction: Decision Tree Classifier boosts profits.
- Success Correlation: More flights at a site lead to higher success rates.
- Orbit Efficiency: ES-L1, GEO, HEO, SSO, and VLEO orbits have the highest success rates.

Conclusions II



2. Implications

- Cost Efficiency: Strategic site selection and payload planning.
- Operational Optimization: Enhanced operations and equipment.
- Risk Mitigation: Reduced launch failures.

3. Future Research

- Advanced Models: More accurate machine learning algorithms.
- Real-time Analysis: Implement real-time data processing.
- Environmental Factors: Investigate weather and geographic impacts.
- Market Expansion: New orbits and clients.

Appendix

- GitHub Link

Thank you!