IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Amber Henry
01/20/2025

Outline

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

Executive Summary

- Summary of methodologies
  - Collecting the Data
  - Data Wrangling
  - Exploratory Data Analysis (EDA)
  - Interactive Visual Analytics and Dashboard
  - Predictive Analytics
- Summary of all results

Introduction

- We aim to predict the success of Falcon 9’s first-stage landing. SpaceX promotes its Falcon 9 launches at a price of $62 million, while competitors charge upwards of $165 million per launch. A significant portion of SpaceX's cost savings comes from its ability to reuse the first stage of the rocket. By accurately forecasting whether the first stage will land successfully, we can estimate the overall cost of a launch. This data could be valuable for other companies looking to compete with SpaceX in the rocket launch market.
- In this presentation, we will answer the following:
  1. How can we predict the success of Falcon 9’s first-stage landing?
  2. What role does the reusability of the first stage play in reducing the cost of SpaceX launches?
  3. How can this prediction model help other companies compete with SpaceX for rocket launches?
  4. What data sources and methodologies were used to build the predictive model?
  5. What insights were gained from exploratory data analysis (EDA) and visualizations?

Section 1

Methodology

Methodology

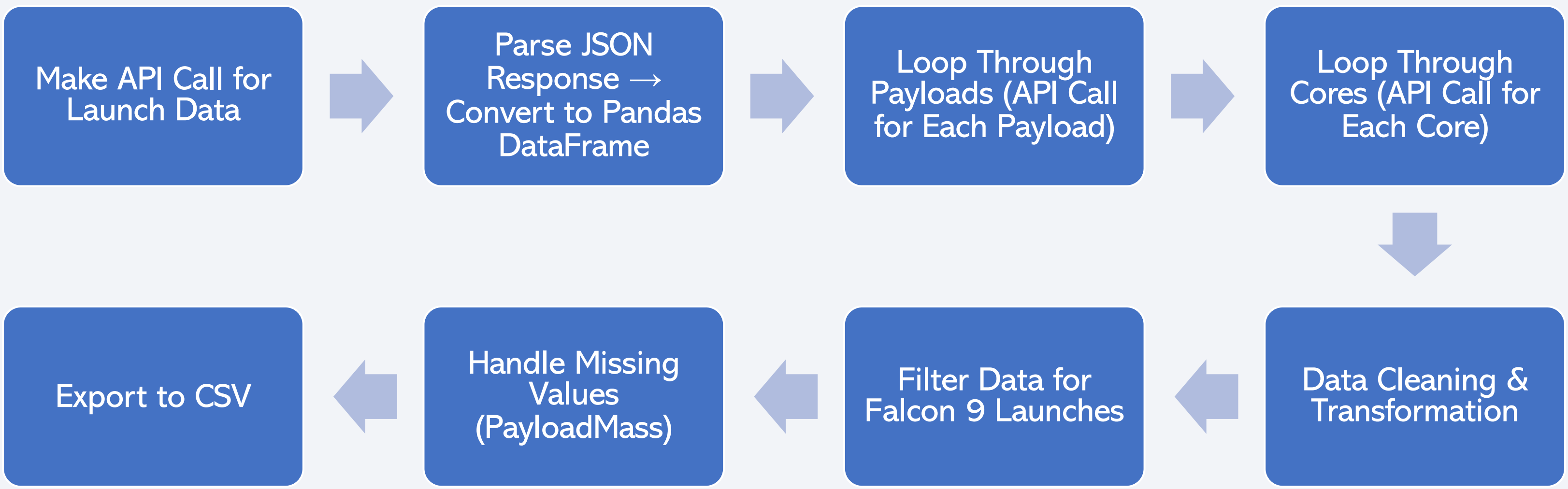
Executive Summary

- Data collection methodology:
  - SpaceX REST API, Web scraping
- Perform data wrangling
  - Find patterns in data and determine labels for training models
- 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

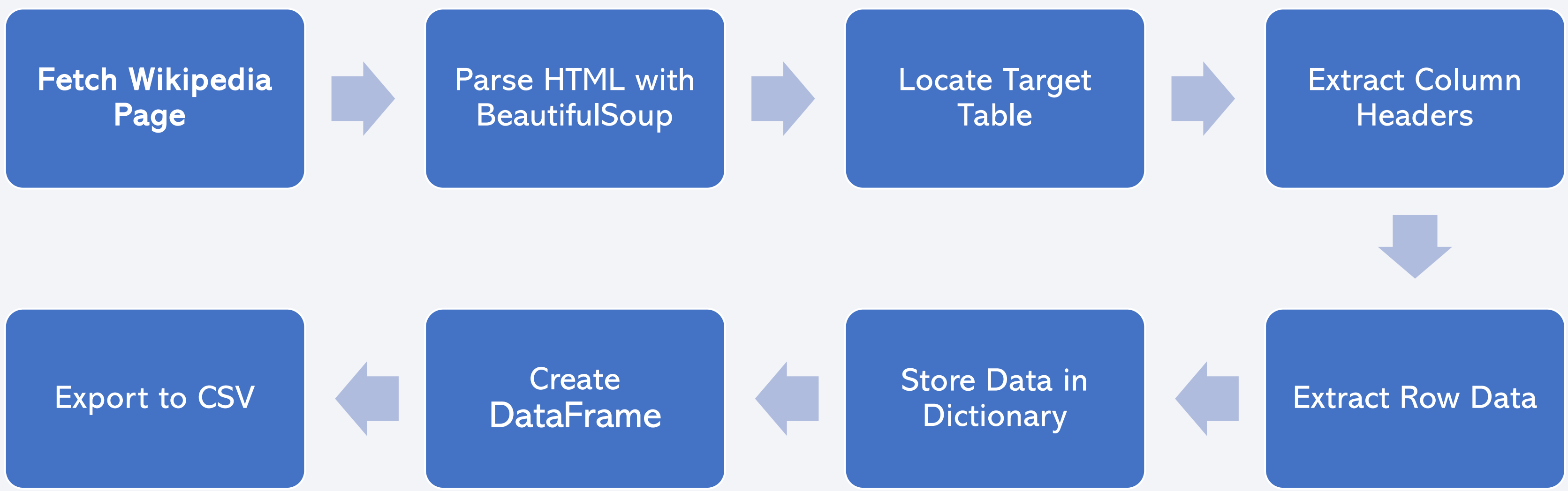
- **GET Request**: Used to pull data from SpaceX API.
- **Payload Data**: Collected payload mass and orbit info.
- **Core Data**: Information about rocket cores, such as landing success and reuse count.
- **Static Data**: Used to ensure consistency in the dataset (via a provided JSON file).
- **Filtering**: Isolated Falcon 9 data by **BoosterVersion**.
- **Missing Values**: Handled by replacing np.nan values with the **mean**.
- **Data Wrangling**: Reassigned **FlightNumber** to ensure sequential numbering.

Data Collection – SpaceX API



https://github.com/amberlhenry/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



https://github.com/amberlhenry/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Libraries Installed: Used Pandas and NumPy for data manipulation and analysis.
- Data Cleaning: Identified and calculated the percentage of missing values.
- Feature Engineering:
  - Determined training labels (Class variable: 0 = unsuccessful landing, 1 = successful landing).
  - Calculated number and occurrences of orbits using .value_counts().
  - Analyzed mission outcomes per orbit to classify success rates.
- Dataset Exploration: Reviewed key attributes like FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, and Outcome.
- Data Export: Saved cleaned and processed dataset for further analysis

https://github.com/amberlhenry/Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

1. **Flight Number vs. Launch Site**
   - Scatter Plot Purpose: To examine the relationship between the number of launches and their success across different launch sites.
   - Higher flight numbers correlate with more successful landings.
2. **Payload Mass vs. Launch Site**
   - Scatter Plot Purpose: To understand how payload mass impacts launch success at various launch sites.
   - No heavy payloads (>10,000 kg) were launched from VAFB-SLC.
3. **Orbit Type vs. Success Rate**
   - Bar Chart Purpose: To analyze success rates for each orbit type.
   - Polar, LEO, and ISS orbits have higher success rates compared to GTO.
4. **Flight Number vs. Orbit Type**
   - Scatter Plot Purpose: To identify the relationship between flight numbers and success rates for each orbit type.
   - LEO orbits show a positive correlation, while GTO lacks a clear pattern.
5. **Payload Mass vs. Orbit Type**
   - Scatter Plot Purpose: To observe how payload mass affects success rates for different orbits.
   - Polar, LEO, and ISS orbits show higher success rates for heavy payloads.
6. **Yearly Launch Success Trend**
   - Line Chart Purpose: To track the trend of launch success rates over time.
   - Success rates have steadily increased from 2013 to 2020.

https://github.com/amberlhenry/Applied-Data-Science-Capstone/blob/main/edadataviz.ipynb

EDA with SQL- Queries

- **Unique Launch Sites**
  - Displayed distinct launch site names.
  - Identified available launch sites for analysis.
- **Launch Sites Starting with 'CCA'**
  - Filtered for launch sites beginning with "CCA."
  - Narrowed focus to specific launch sites for further study.
- **Total Payload Mass by NASA (CRS) Boosters**
  - Calculated average payload mass for NASA missions.
  - NASA missions averaged 2,928.4 kg payload mass.
- **First Ground Pad Success**
  - Identified the earliest date of a successful ground pad landing.
  - Achieved on 2015-12-22.
- **Boosters with Drone Ship Success (Payload 4000–6000 kg)**
  - Listed boosters with successful landings and mid-range payloads.
  - Boosters like F9 FT B1022 and B1026 were identified.
- **Mission Outcome Counts**
  - Summarized total counts for successful and failed missions.
  - 100 successes vs. 1 in-flight failure.
- **Boosters with Maximum Payload**
  - Used a subquery to find boosters carrying maximum payloads.
  - Multiple F9 B5 boosters were identified.
- **Failures on Drone Ship in 2015**
  - Filtered for failed drone ship landings in 2015.
  - Failures occurred in January and April at CCAFS LC-40.
- **Ranked Landing Outcomes (2010–2017)**
  - Counted and ranked landing outcomes within a specific timeframe.
  - "No attempt" was the most common outcome, followed by "Success (drone ship)."

https://github.com/amberlhenry/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Map Features:

- **Launch Site Markers:** Placed at SpaceX launch sites (CCAFS LC-40, SLC-40; KSC LC-39A; VAFB SLC-4E) to identify key locations.
- **Highlighted Circles:** Colored circles around sites to visually highlight the area.
- **Mouse Position Plugin:** Displays cursor's latitude and longitude for precise navigation.
- **Marker Clusters:** Groups markers for better readability and navigation at different zoom levels.
- **Interactive Pop-ups:** Provide additional site details when clicked.
- **Map Centering & Zoom:** Initial view centered on NASA Johnson Space Center for a familiar starting point.

https://github.com/amberlhenry/Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- **Dropdown Menu**: Filters launch data by site or shows all sites.
  Purpose: Tailored analysis of success rates.
- **Pie Chart**: Displays total launches by site or success vs. failure for a selected site.
  Purpose: Visualize launch performance metrics.
- **Payload Slider**: Filters data by payload mass range.
  Purpose: Analyze payload impact on success.
- **Scatter Plot**: Correlates payload mass with success, colored by booster type.
  Purpose: Explore payload and booster effects on outcomes.
- **Interactive Features**

1. Pie chart updates dynamically based on site selection.
2. Scatter plot filters based on site and payload range.

- *Why*: Simplifies data exploration for SpaceX launches with clear, interactive visuals.

https://github.com/amberlhenry/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

Model Development Process:

1. Exploratory Data Analysis:
- Investigated dataset, created class labels for prediction.

2. Data Preprocessing:
- Standardized data and split into training and test sets.

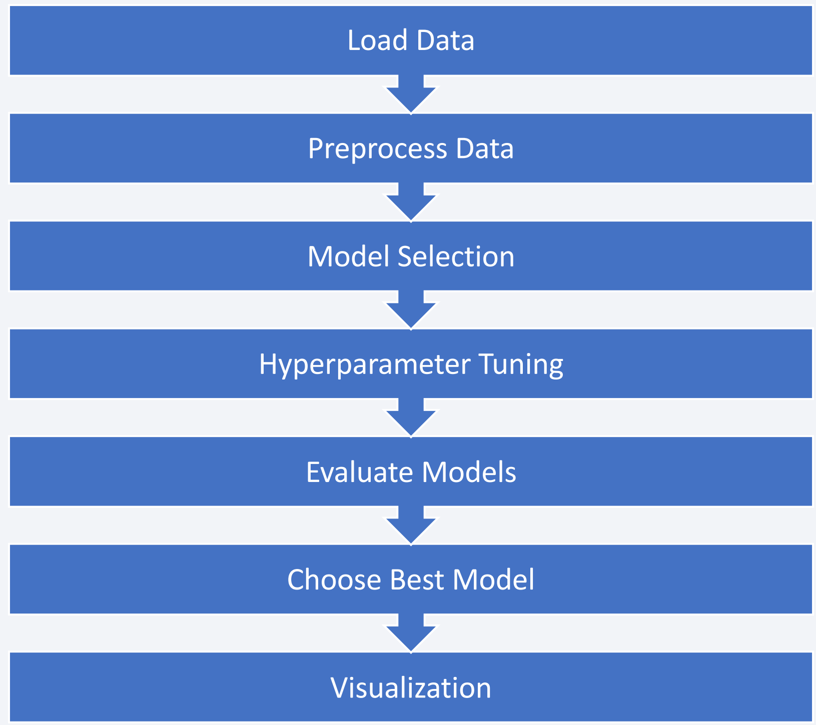
3. Model Selection:
- Evaluated models: SVM, Classification Trees, Logistic Regression.

4. Hyperparameter Tuning:
- Used GridSearchCV for best hyperparameters (e.g., for SVM and Decision Trees).

5. Model Evaluation:
- Compared models using test data and confusion matrix for performance assessment.

6. Best Model:
- Decision Tree performed best with a score of 0.889.

7. Visualization:
- Displayed model performance with bar chart highlighting best model.



https://github.com/amberlhenry/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

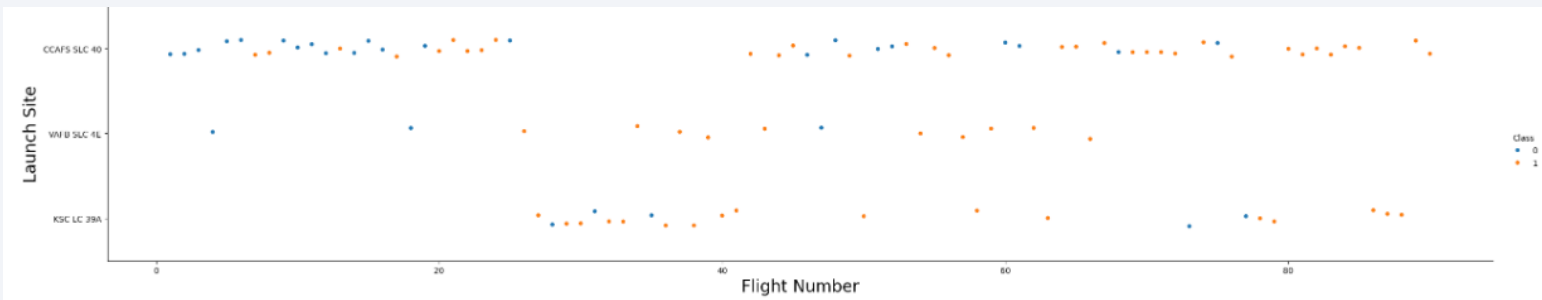
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

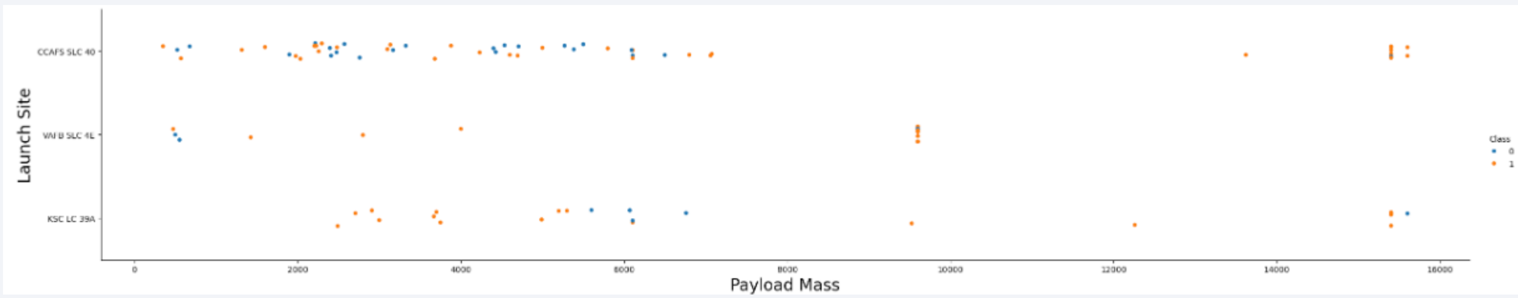
HENRY

Flight Number vs. Launch Site



- **Flight Success Trends:**
  - Earlier flights (blue) had a lower success rate; later flights (orange) had a higher success rate.
- **Launch Site Insights:**
  - Around half of the launches were from CCAFS SLC 40.
  - VAFB SLC 4E and KSC LC 39A show higher success rates.
- **Inference:**
  - Newer launches generally have a higher success rate, suggesting improvement over time.

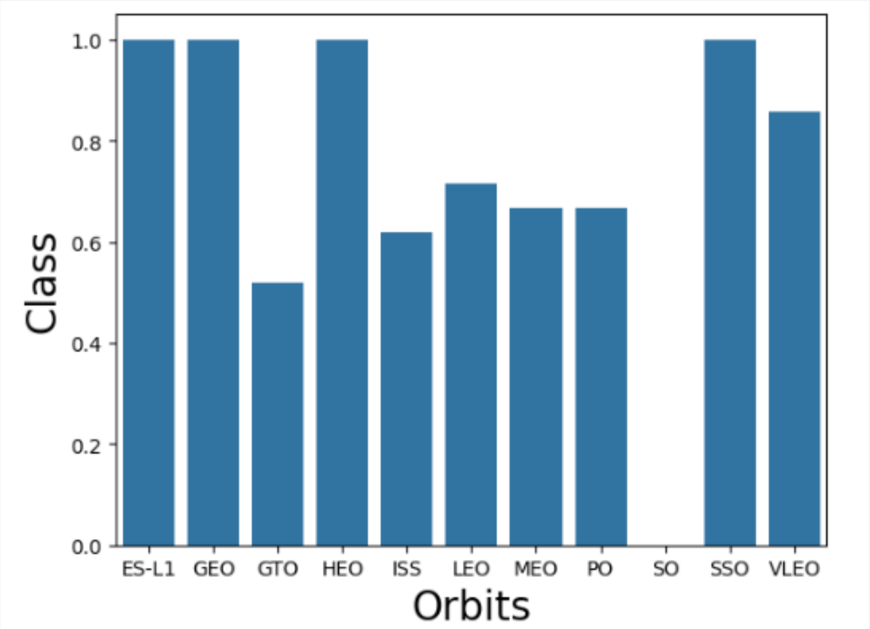
Payload vs. Launch Site



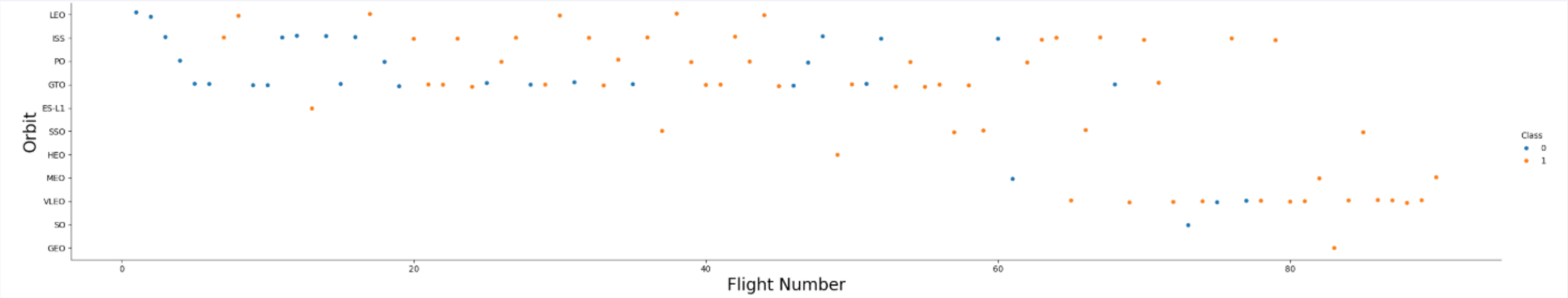
- **Payload Mass and Success:** Higher payload mass generally correlates with higher success rates.
- **Launch Success by Mass:** Most launches with payloads over 7,000 kg were successful.
- **Launch Site Insights:**
  - KSC LC 39A has a 100% success rate for launches under 5,500 kg.
  - AFB SLC 4E has not launched payloads over 10,000 kg.

Success Rate vs. Orbit Type

- 100% Success rates:
  - ES-L1
  - GEO
  - HEO
  - SSO
  - VLEO



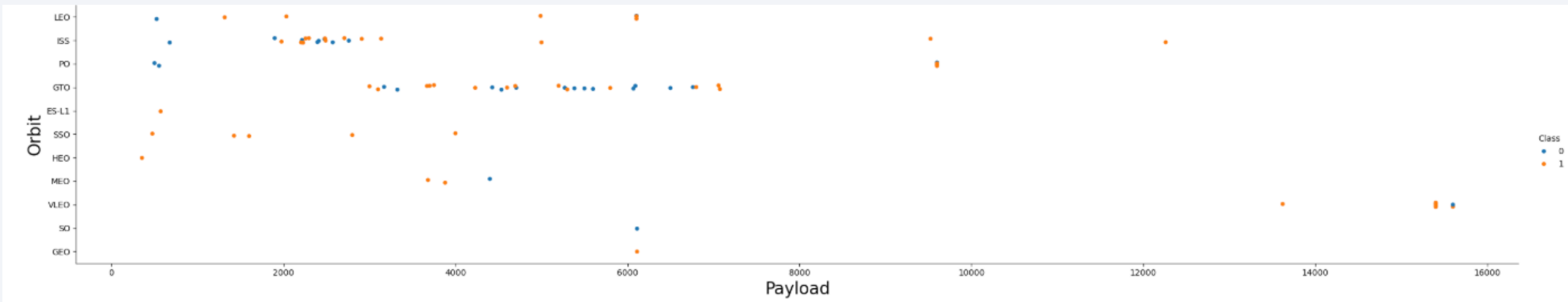
Flight Number vs. Orbit Type



- **LEO Orbit**: Success increases with higher flight numbers.
- **GTO Orbit**: No clear relationship between flight number and success.

HENRY

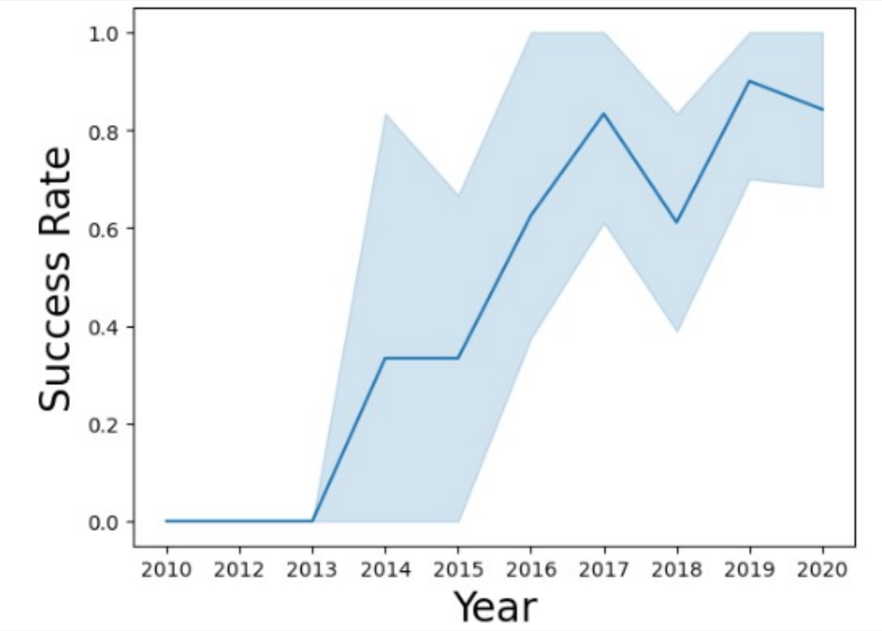
Payload vs. Orbit Type



- **Heavy Payloads**: Higher landing success rates for Polar, LEO, and ISS orbits, especially pronounced for LEO.
- **GTO Orbits**: Success and failure rates are mixed, making outcomes harder to distinguish.

Launch Success Yearly Trend

- **Success Rate (2013–2020):** Steady increase observed over the years.



All Launch Site Names

- **Launch Site Names**
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

```
%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

- The filtered dataset shows 5 records where the launch site begins with ‘CCA’. These records represent SpaceX launches from sites with names starting with “CCA”, which might indicate specific launch locations used for particular mission types. By focusing on these sites, we can analyze patterns related to cost efficiency and success rates for missions launched from these locations.

Total Payload Mass

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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer like 'NASA (CRS)'
```

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

sum(PAYLOAD_MASS__KG_)
45596

- Total payload carried by boosters from NASA: **45,596 kg**
- This indicates the combined mass of the payloads launched in these specific NASA missions, providing insight into the scale and capacity of SpaceX's missions in collaboration with NASA.

Average Payload Mass by F9 v1.1

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

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like 'F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

- Average payload mass carried by booster version F9 v1.1: **2,928.4 kg**
- This result provides insight into the typical payload capacity of the F9 v1.1 booster, helping to assess its performance and efficiency for SpaceX missions.

First Successful Ground Landing Date

```
%sql select min(Date) from SPACEXTABLE where Landing_Outcome like 'Success (ground pad)'
```

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

min(Date)
2015-12-22

- The first successful landing on a ground pad by SpaceX was achieved on **December 22, 2015**.
- This historic milestone marked a significant advancement in rocket reusability, demonstrating SpaceX's ability to land and recover the first stage of its Falcon 9 rocket, paving the way for future cost-effective space missions.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select Booster_Version from SPACEXTABLE where Landing_Outcome like '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

- These boosters were part of SpaceX missions that not only achieved successful landings on drone ships but also carried a payload mass within the specified range, showcasing the efficiency and versatility of the Falcon 9 rocket series.

Total Number of Successful and Failure Mission Outcomes

```
%sql select Mission_Outcome, count(*) as count from SPACEXTABLE where Mission_Outcome like 'Success%' union select Mission_Outcome, count(*) from SPACEXTABLE where Mission_Outcome like 'failure%'
```

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

Mission_Outcome	count
Failure (in flight)	1
Success	100

- The total number of mission outcomes are as follows:
  - **Success**: 100 missions
  - **Failure (in flight)**: 1 mission
- This indicates that SpaceX has had a significantly higher success rate in its missions, with a very small number of failures, showcasing the reliability and effectiveness of its launch system.

Boosters Carried Maximum Payload

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

```
%sql select distinct 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

The records for failed drone ship landings in 2015 show the following:

- **January**:
  - **Booster Version**: F9 v1.1 B1012
  - **Launch Site**: CCAFS LC-40
- **April**:
  - **Booster Version**: F9 v1.1 B1015
  - **Launch Site**: CCAFS LC-40

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

These are the missions where the landing on the drone ship failed during 2015, categorized by the respective months and launch details. This highlights the timing and specifics of the missions that faced landing challenges.

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

The ranking of landing outcomes between **2010-06-04** and **2017-03-20** is as follows:

1. **No attempt:** 10 occurrences
2. **Success (drone ship):** 5 occurrences
3. **Failure (drone ship):** 5 occurrences
4. **Success (ground pad):** 3 occurrences
5. **Controlled (ocean):** 3 occurrences
6. **Uncontrolled (ocean):** 2 occurrences
7. **Failure (parachute):** 2 occurrences
8. **Precluded (drone ship):** 1 occurrence

- This ranking shows the frequency of various landing outcomes, with the most common being “No attempt” followed by successes and failures on drone ships, providing a clear picture of SpaceX’s landing attempts during this period.

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

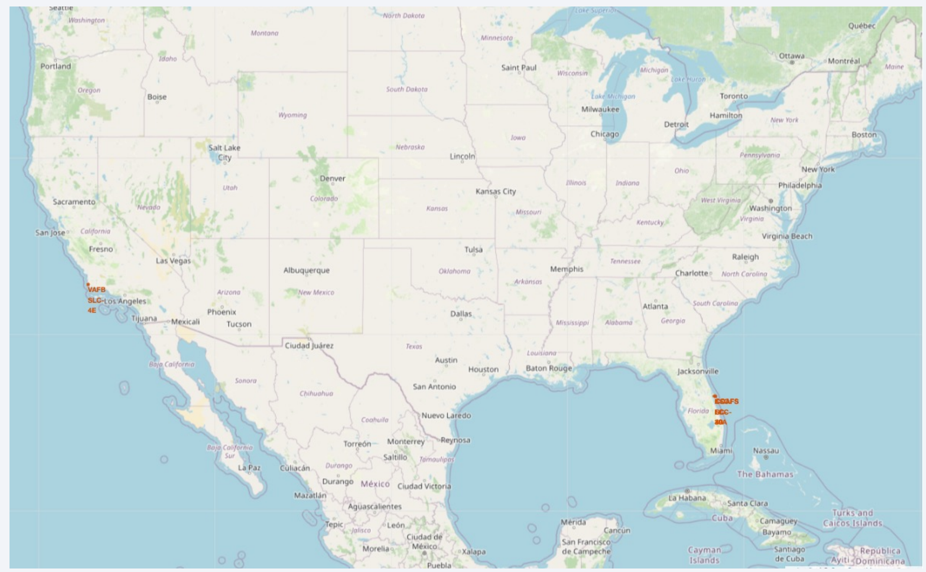
Section 3

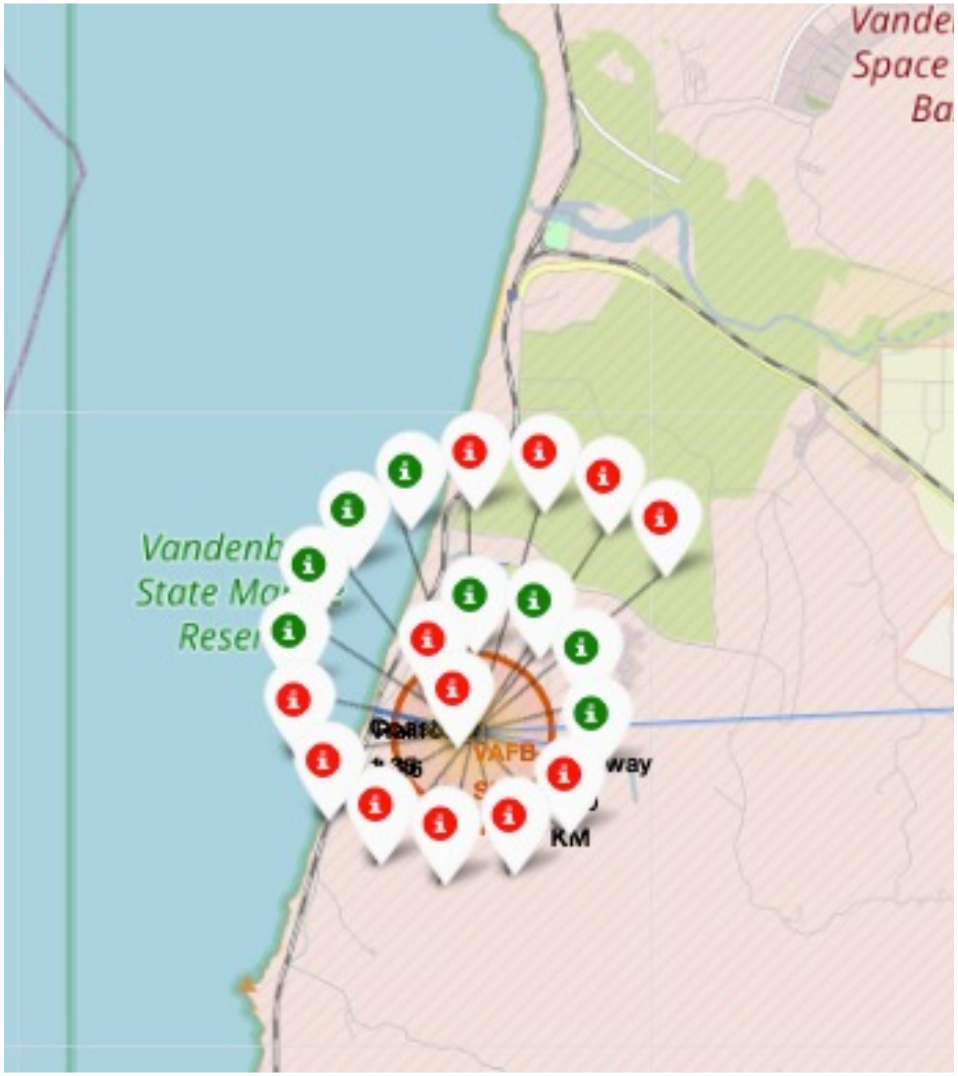
Launch Sites Proximities Analysis

HENRY

Global Launch Site Map

- **Map Visualization**: Displays SpaceX launch sites using Folium, marked by circles and markers.
- **Equator Proximity**: Most sites are near the Equator for added rotational speed benefits.
- **Coastline Locations**: All sites are near coastlines to minimize risk during launches over the ocean.





Launch Outcomes by Site

- **Launch Outcomes Map**: Displays color-coded markers for successful and failed launches.
- **Visual Performance**: Different colors help identify sites with high success rates versus those with failures.
- **Insights**: Offers a quick overview of SpaceX's launch site performance across the globe.

Launch Site and Proximities Map

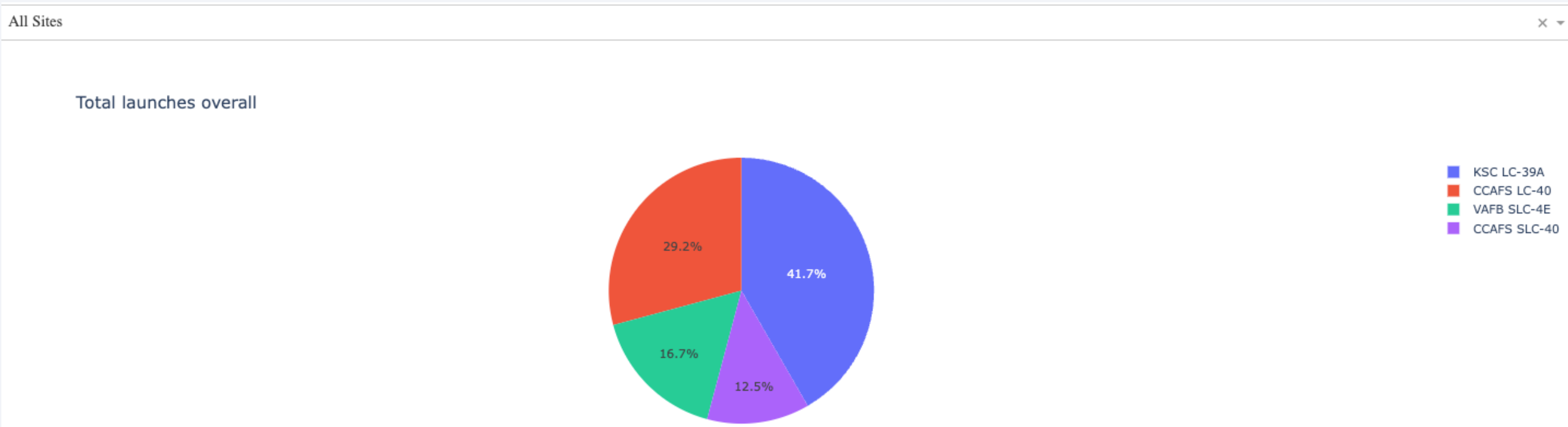
- **Launch Site Proximity:** Shows distances to nearby railways, highways, and coastlines.
- **Geographical Context:** Helps assess the site’s accessibility and safety.
- **Distance Info:** Displays distances to infrastructure, aiding logistics and planning.

Section 4

Build a Dashboard with Plotly Dash

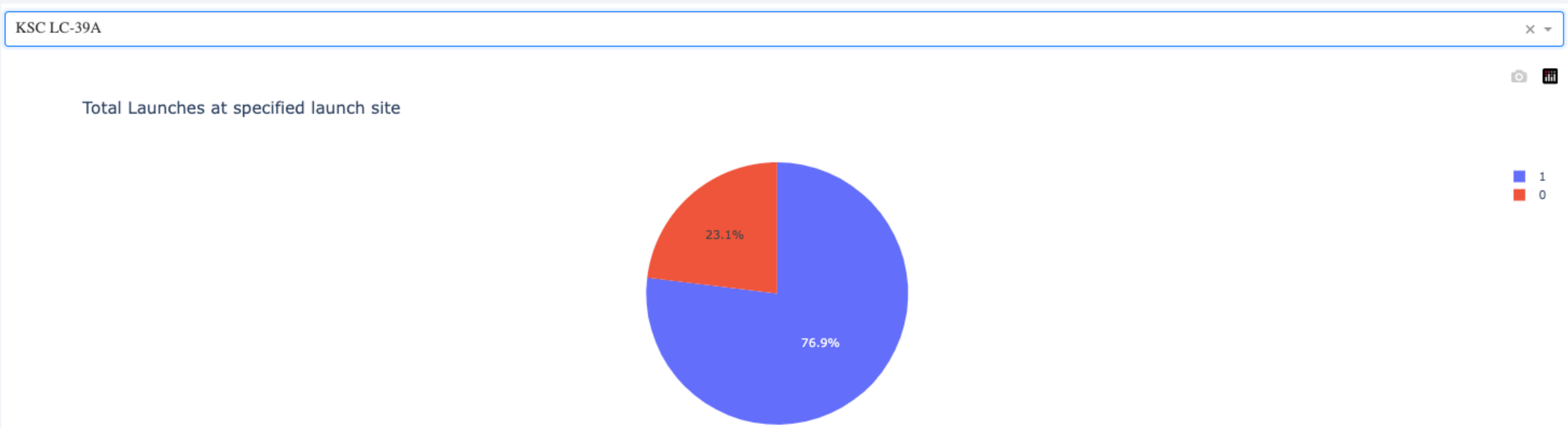
HENRY

Launch Success by Site



- **Launch Success Distribution**: Pie chart showing the success rate across all launch sites.
- **KSC LC-39A** has the highest success rate at **41.7%**.
- **Insights**: Visualizes success rates, highlighting the most successful launch sites.

Launch Success KSC LC-29A



- **KSC LC-39A Launch Success Breakdown:** Pie chart showing launch success ratio.
- **KSC LC-39A** has **76.9% success** and **23.1% failure**.
- **Insights:** Visualizes the high success rate for this site compared to failures.

Payload Mass Success

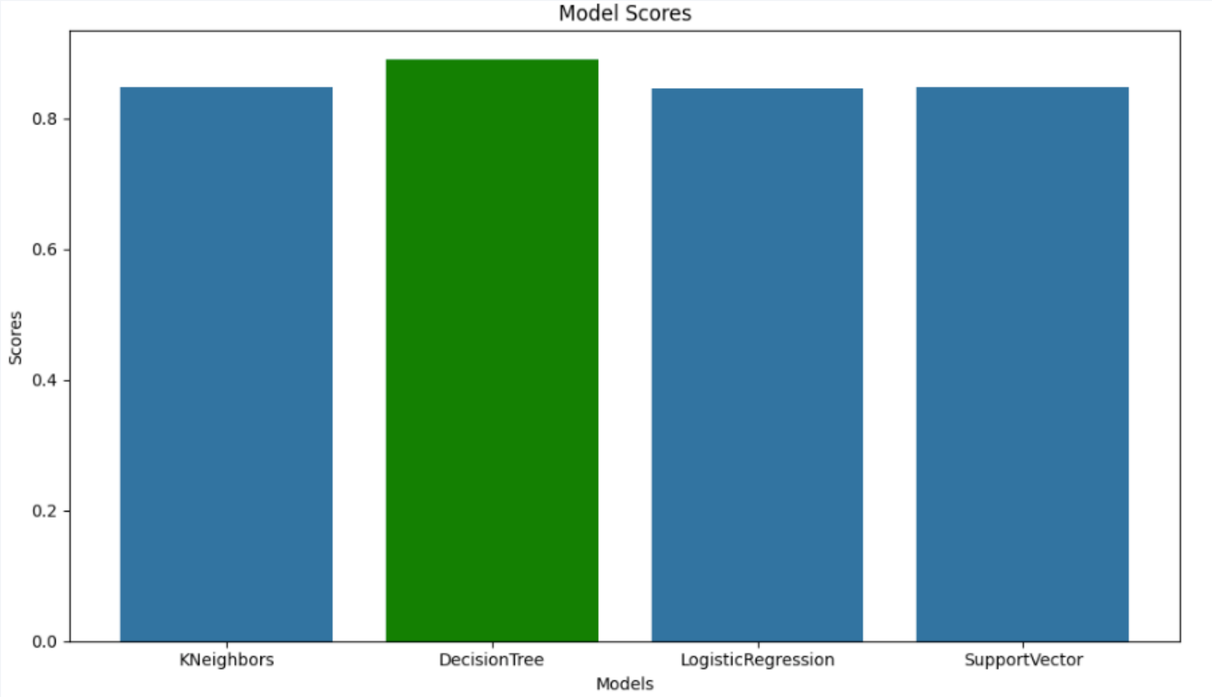
- **Payload vs. Launch Outcome:** Scatter plot showing launch outcomes based on payload weight.
- Payloads between **2000 kg and 5000 kg** have the highest success rate.
- **Insights:** Success rate is higher for mid-range payloads, with **1** indicating success and **0** indicating failure.

Section 5

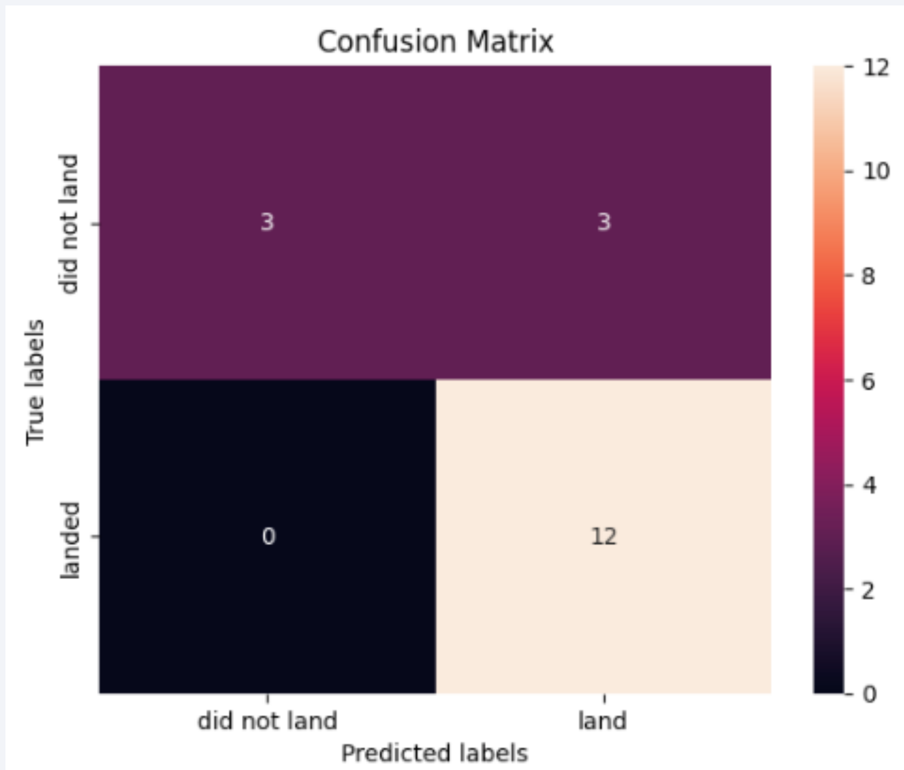
Predictive Analysis (Classification)

HENRY

Classification Accuracy



Confusion Matrix



Conclusions

1. **Model Performance:** The decision tree model slightly outperforms others on the test set.
2. **Equator:** Many launch sites are positioned near the equator, benefiting from the Earth's rotational speed, which reduces fuel and booster costs.
3. **Coastal Proximity:** All launch sites are located near coastlines for safer launches.
4. **Launch Success Over Time:** The success rate of launches has been increasing over time.
5. **KSC LC-39A Performance:** This site has the highest success rate, achieving 100% success for launches with payloads under 5,500 kg.
6. **Orbit Performance: ES-L1**, **GEO**, **HEO**, and **SSO** orbits consistently show a 100% success rate.
7. **Payload Mass and Success:** Launches with higher payload mass generally have higher success rates across all sites.

Appendix

- SpaceX Falcon 9 Data Sheet
- SpaceXNow Statistics
- SpaceX Falcon 9 Overview
- SpaceX Falcon 9 v1.2 Data Sheet
- Analysis of Flight Data in SpaceX's Falcon 9
- SpaceX Falcon 9 Rocket Facts
- SpaceX Falcon 9 First Stage Landing Prediction
- SpaceX Falcon 9 1st Stage Success Landing Prediction
- SpaceX Statistics
- SpaceX Falcon 9 Flight Analysis


Thank you!
HENRY