IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Homero Garza
August 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data Wrangling
  - EDA with data visualization
  - Eda with SQL
  - Building interactive map with Folium
  - Building dashboard with Plotly Dash
  - Predictive analysis
- Summary of all results
  - EDA results
  - Interactive analytics
  - Predictive analysis

Section 1

Methodology

Introduction

- Project background and context
  - SpaceX's cost-effective launches are priced at $62 million compared to competitors $165 million largely stem from their first stage reusability. Projecting successful landings not only informs launch outcomes but also aids cost estimation which is valuable for competitors challenging SpaceX bids.
- Problems you want to find answers
  - This capstone project aims to predict the successful landings of the Falcon 9 first stage rocket booster.

Methodology

- Data collection methodology:
  - SpaceX API
  - Web scraping from Wikipedia
- Perform data wrangling:
  - Cleaning of null values and useless information, encoding features 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:
  - Logistic Regression, KNN, SVM, Decision tree all tuned using GridSearchCV and evaluated to find champion model.

Data Collection

- How the data was collected:
  - SpaceX launch data was gathered from SpaceX REST API.
  - This API provided data on type of rockets used, payloads delivered, launch specifications, landing specifications, and landing outcomes.
  - The second half of the data was sourced by web scraping Falcon 9 launch data on Wikipedia using BeautifulSoup.

Data Collection – SpaceX API

1. Requesting and parsing the SpaceX launch data.

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

2. Filter the data frame to include only Falcon 9 launches.

```
data_falcon9 = data[data['BoosterVersion'] != 'Falcon 1']
```

3. Dealing with missing values.

```
# Calculate the mean value of PayloadMass column
data_mean = data_falcon9['PayloadMass'].mean()
# Replace the np.nan values with its mean value
data_falcon9['PayloadMass'] = data_falcon9['PayloadMass'].replace(np.nan, data_mean)
```

GitHub Link: API Data Collection

Data Collection - Scraping

1. Requesting Falcon 9 launch wiki page from URL.

2. Extract all column names from HTML Table header.

3. Create a data frame by parsing the launch HTML tables.

```
# use requests.get() method with the provided static_url
# assign the response to a object
response = requests.get(static_url)
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response.content, 'html.parser')
```

```
# Use the find_all function in the BeautifulSoup object, with element type `table`
# Assign the result to a list called `html_tables`
html_tables = soup.find_all('table')
```

```
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all('table',"wikitable plainrowheaders collapsible")):
    # get table row
    for rows in table.find_all("tr"):
        #check to see if first table heading is as number corresponding to launch a number
        if rows.th:
            if rows.th.string:
                flight_number=rows.th.string.strip()
                flag=flight_number.isdigit()
        else:
            flag=False
        #get table element
        row=rows.find_all('td')
        #if it is number save cells in a dictonary
        if flag:
            extracted_row += 1
            # Flight Number value
            # TODO: Append the flight_number into launch_dict with key `Flight No.`
            launch_dict['Flight No.'].append(flight_number)
            #print(flight_number)
            datatimelist=date_time(row[0])
```

GitHub Link: [Web Scraping Data Collection]

Data Wrangling

1. Check for null values.
2. Calculate the number of launches at each site.
3. Calculate the number of orbit occurrences.
4. Calculate the number of occurrences of mission outcome per orbit type.
5. Create a label column for the outcome column.
6. Handle null values.

GitHub Link: Data Wrangling

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts

1. Scatter plot looking at relationship between flight number and launch site.
2. Scatter plot looking at relationship between payload and launch site.
3. Bar chart looking at the success rate of each orbit.
4. Scatter plot looking at relationship between flight number and orbit type.
5. Scatter plot looking at relationship between payload mass and orbit type.
6. Line chart looking at trend of success by year.

GitHub Link: EDA with Data Visualization

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - Displaying the names of the unique launch sites.
  - Displaying 5 records where launch sites begin with 'KSC'.
  - Displaying the total payload mass of launched boosters from NASA.
  - Displaying the average payload mass carried by Falcon 9 boosters.
  - Listing the dates of the first successful ground pad landing.
  - Listing the names of the boosters with successful drone ship landings that also had a payload mass greater than 4,000 kg but less than 6,000 kg.
  - Listing the total number of successful and failed mission outcomes.
  - Listing the names of the booster versions that carried the maximum payload.
  - Listing the records of failed landings on drone ship that include the month, booster version and launch site in 2015.
  - Ranking the count of successful landing outcomes between June, 4 2010 and March, 20 2017 in descending order.

GitHub Link: EDA with SQL

Build an Interactive Map with Folium

- Added map markers to launch site locations, and colored markers at site locations to distinguish successful landings (green) and failed landings (red).
- Added lines to map to show launch site proximities to closest coastline, railway and freeway labeled with the distance in kilometers.

GitHub Link: Interactive Map with Folium

Build a Dashboard with Plotly Dash

- Dashboard contains a drop down list of desired launch site locations to explore.
- The first visual is a pie chart that shows the ratio of successful landings compared to failed landings.
- The second visual contains a slider that can specify which boosters to look at based on payload mass and their landing outcomes.
- These two visuals show how successful outcomes are based on launch site location and by payload mass.

GitHub Link: Dashboard with Plotly Dash

Predictive Analysis (Classification)

- Built 4 different classification models to look at: Logistic Regression, SVM, KNN and Decision Tree.
- Each model was tuned to find the best performance using GridSearchCV.
- The Logistic Regression and Decision Tree model performed the best both with an 83% accuracy score.
- The Decision Tree beat out the Logisitic Regression model and was determined the champion model based on the confusion matrix metrics and the recall score

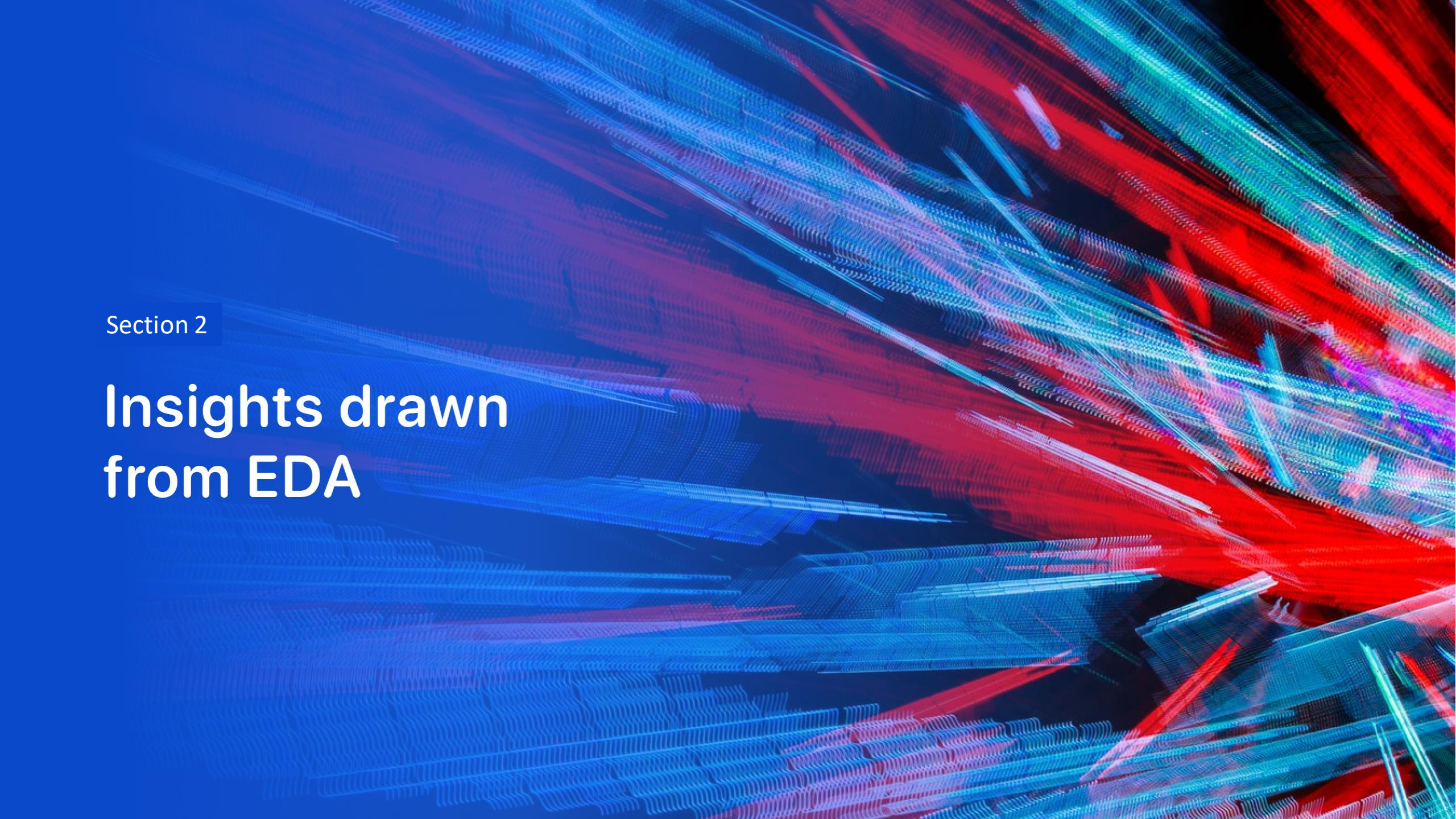
GitHub Link: Predictive Analysis

Results

- Exploratory data analysis results
  - Lower weighted payloads perform better than heavier payloads
  - The success rates for SpaceX launches has a positive relationship with time in years.
  - KSC LC 39A had the most successful launches out of all launch sites.
  - Orbits GEO, HEO, SSO, ES L1 had the best success rates
- Predictive Analysis results
  - Out of the 4 models built and tested the Decision Tree model performed the best.
  - The Decision Tree model scored an 83% on the test set.

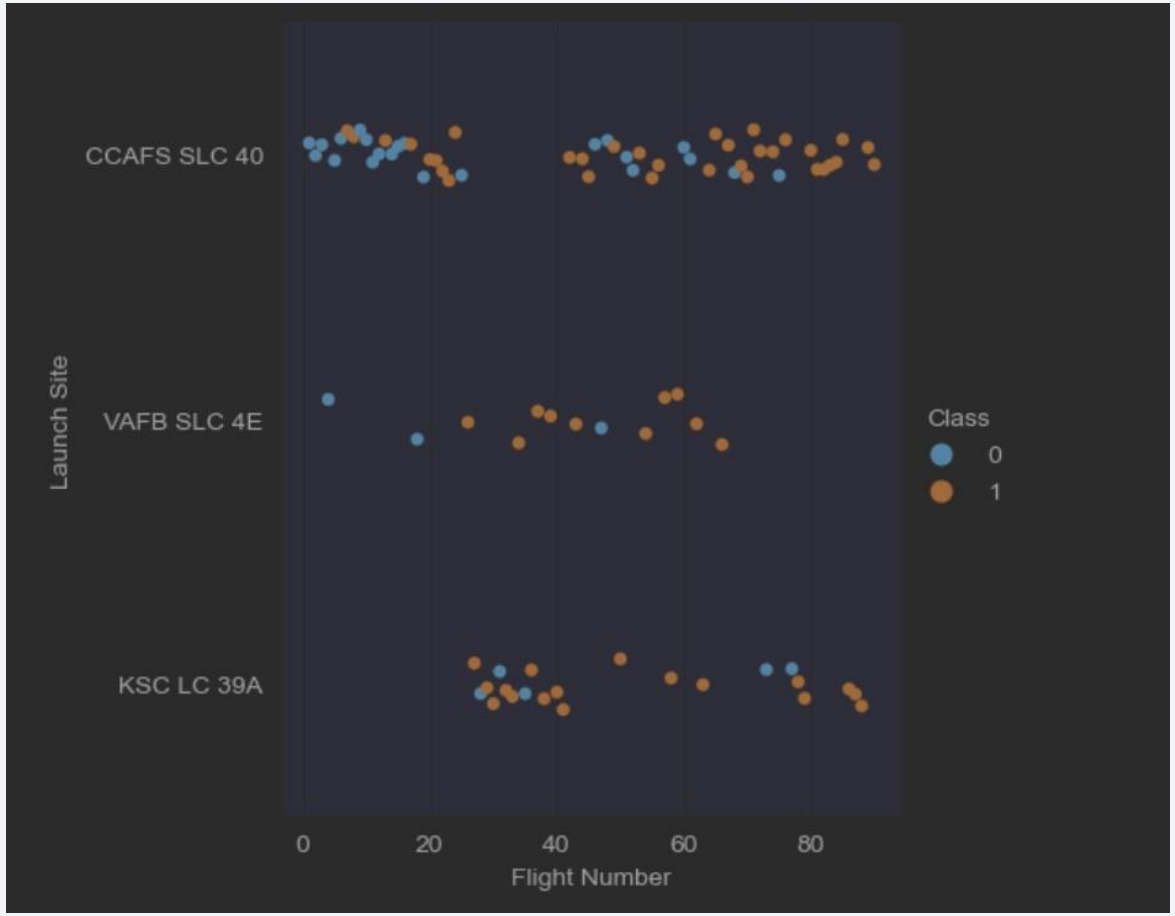
Section 2

Insights drawn from EDA



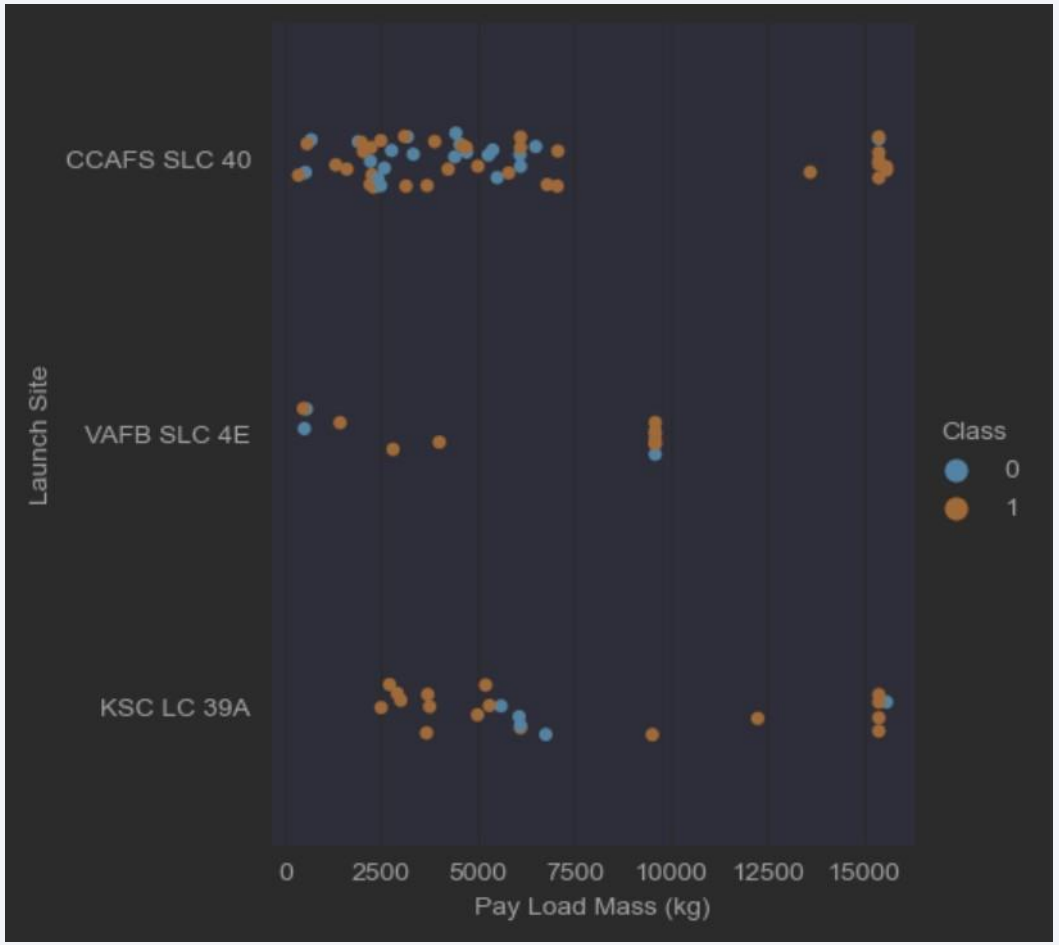
Flight Number vs. Launch Site

- There is a positive relationship between the number of launches and successful landings.
- CCAFS SLC 40 has the most launches.
- KSC LC 39A has the highest ratio of successful launches to failed launches.
- VAFB SLC 4E has the least amount of launches.



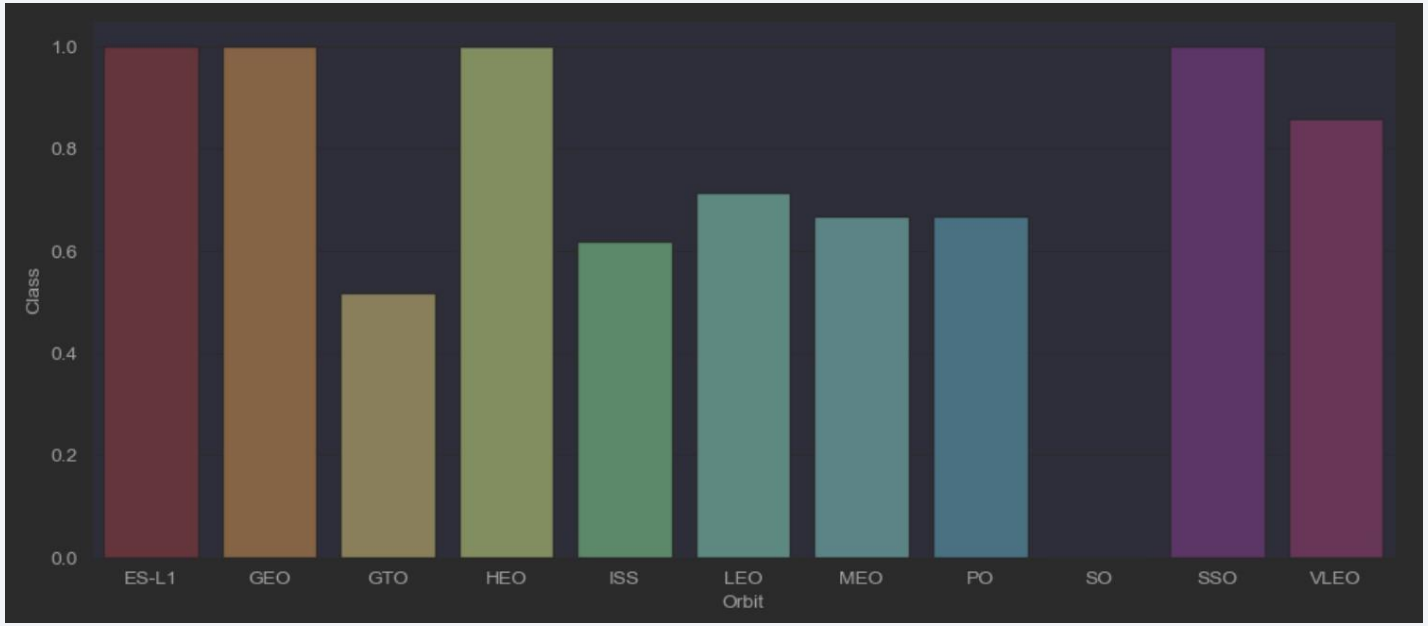
Payload vs. Launch Site

- At VAFB SLC 4E there are no launches with a payload greater than 10,000 kg.
- Most launches at CCAFS SLC 40 have a payload less than 7,500 kg.



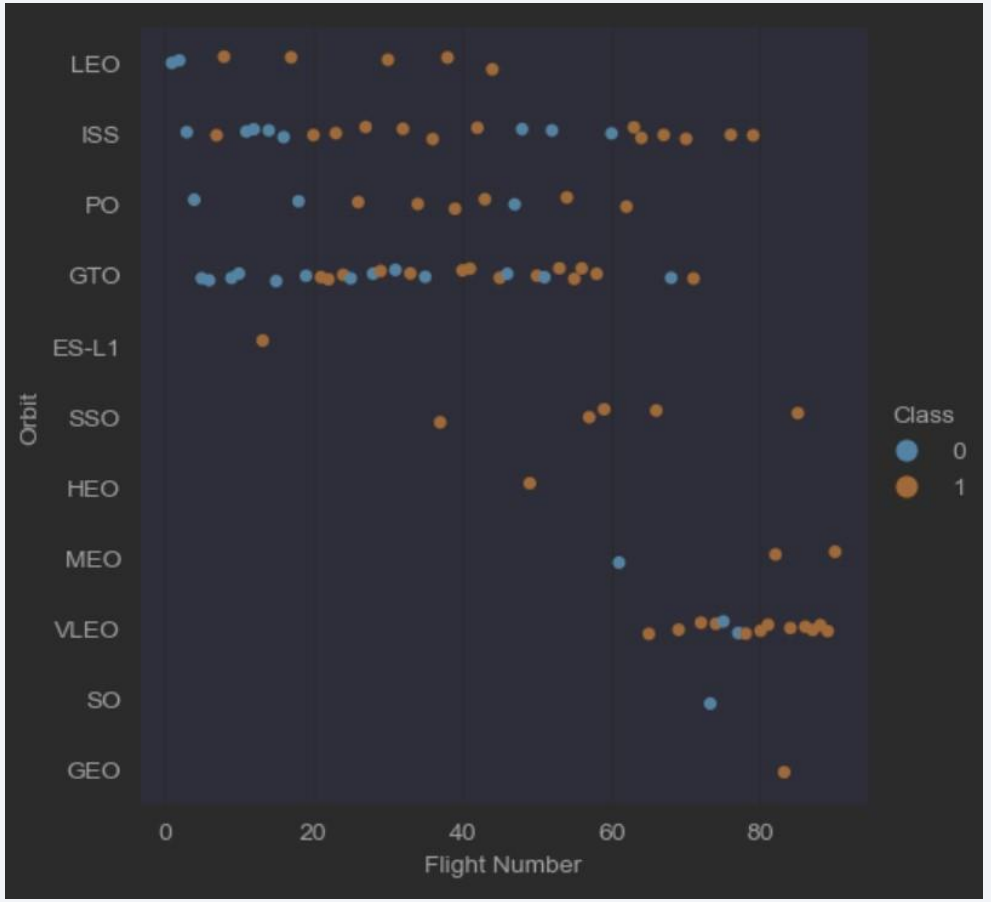
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO have the highest orbit success rates.
- GTO has the lowest success rate followed by ISS.



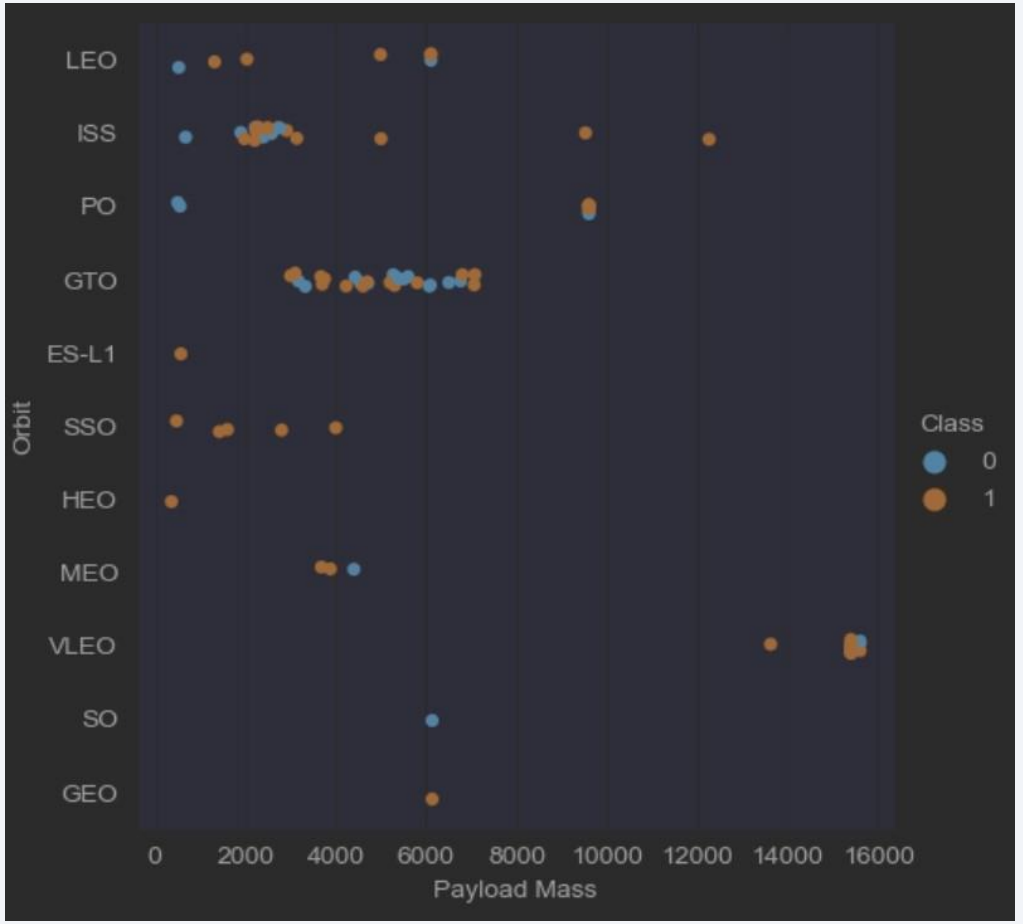
Flight Number vs. Orbit Type

- The most launches are into ISS and GTO orbit.
- There is a positive relationship with the number of launches and success for the LEO orbit.
- There is no distinct relationship between flight number and success for the GTO orbit.



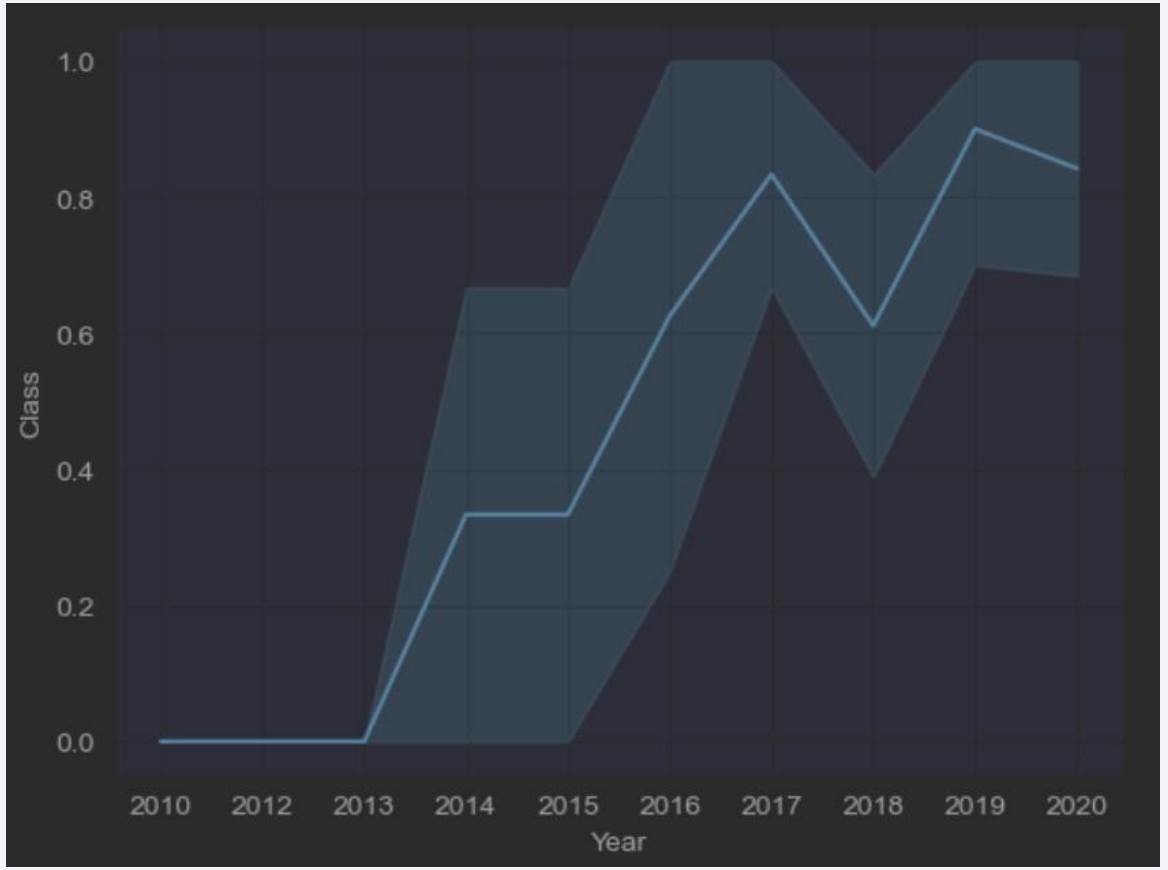
Payload vs. Orbit Type

- For heavier payloads the success rates are higher for ISS, PO and VLEO orbits.
- For the GTO orbit it is hard to distinguish if there are more successful launches or failed launches.



Launch Success Yearly Trend

There is a clear positive linear relationship between launch success as time goes on.



All Launch Site Names

Using SQLite this query found the unique launch sites for the Falcon 9 booster launches.

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

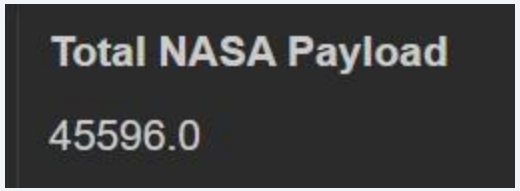
Launch Site Names Begin with 'CCA'

This query retrieved 5 launch records with launch sites that began with "CCA".

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

This query calculated the total payload mass (kg) of launches from NASA. As you can see the total payload mass from NASA launches is 45,596 kg.



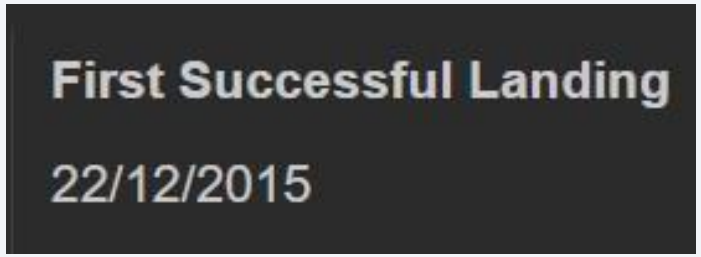
Average Payload Mass by F9 v1.1

The average payload mass of all Falcon 9 launches is 2534.6 kg.

Average Payload Mass
2534.6666666666665

First Successful Ground Landing Date

This query found the date of the first successful Falcon 9 booster landing on the ground pad. The date was December 22, 2015.



Successful Drone Ship Landing with Payload between 4000 and 6000

This query presented a list booster names that landed successfully on the drone ship and also had a payload mass greater than 4,000 kg but less than 6,000 kg.

Booster Name
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

This query calculated the total number of successful missions and the total number of failed missions. There is 1 failed mission and 100 successful missions.

Successful Mission Counts	Failed Mission Counts
100	1

Boosters Carried Maximum Payload

Here is a list of Falcon 9 boosters that have carried the maximum payload mass at 15,600 kg.

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

This query presents a list of failed landings in 2015. Both of the failed landings were meant to land on the drone ship and were launched from site CCAFS LC-40. The failed landings were in April and October of 2015.

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

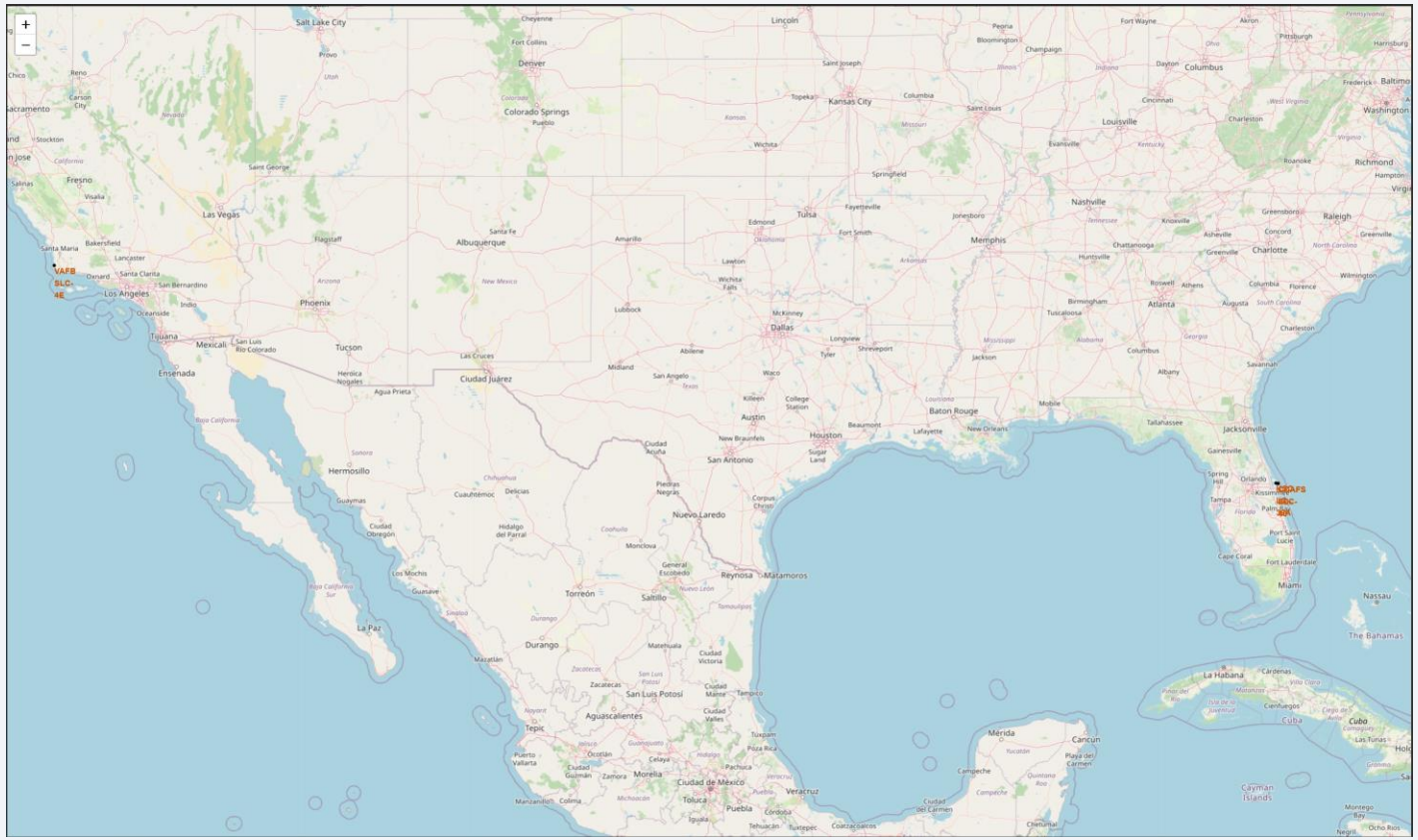
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Here is a list of landing outcomes ranked in descending order by their count between June 4, 2010 and March 20, 2017.
- The landing outcome Success has the most occurrences at 38.

Landing_Outcome	Landing Count
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1
None	0

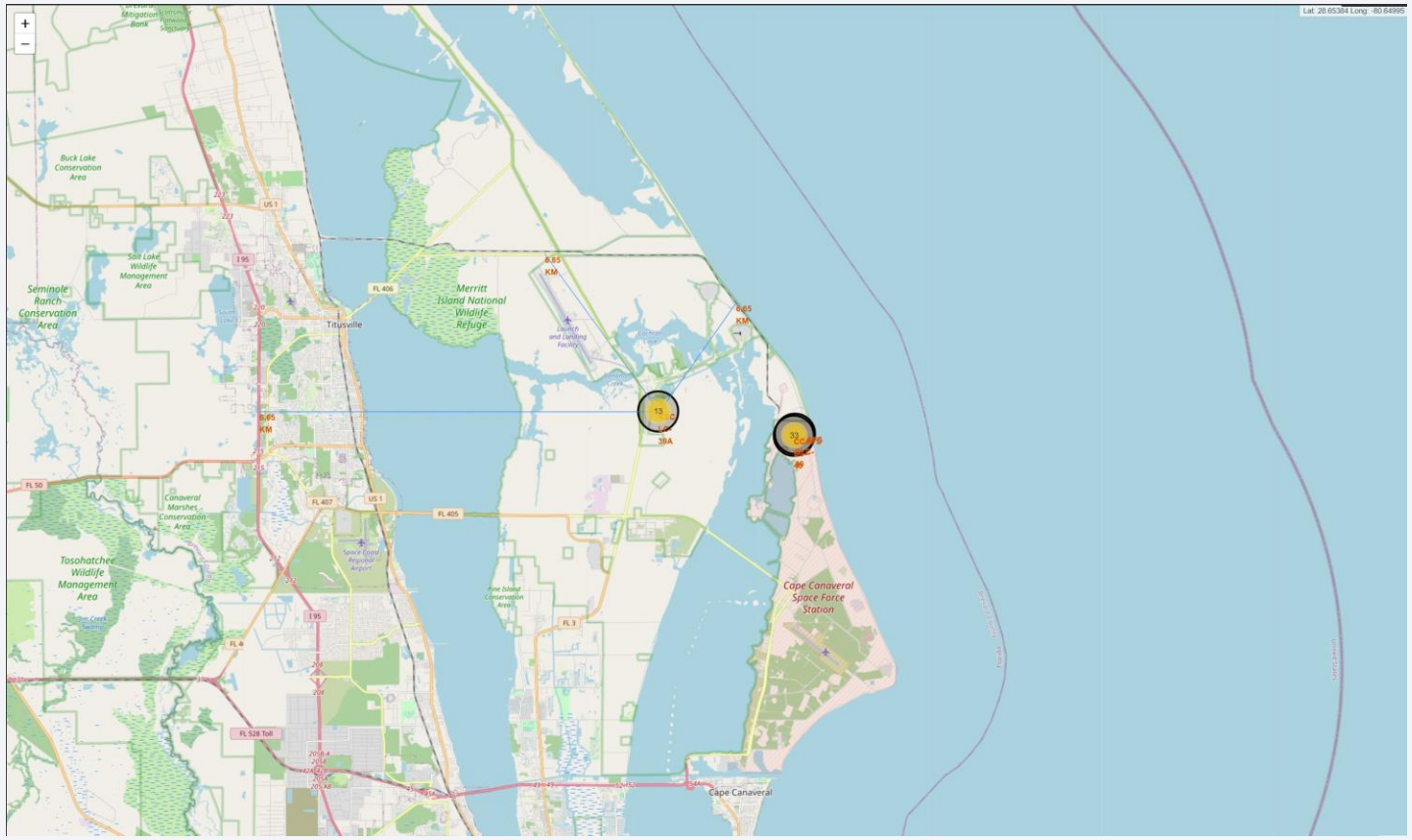

Section 3
Launch Sites
Proximities Analysis

Launch Site Locations Across the U.S



The locations of the sites used to launch Falcon 9 boosters are situated on both the west coast in California and the East Coast in Florida.

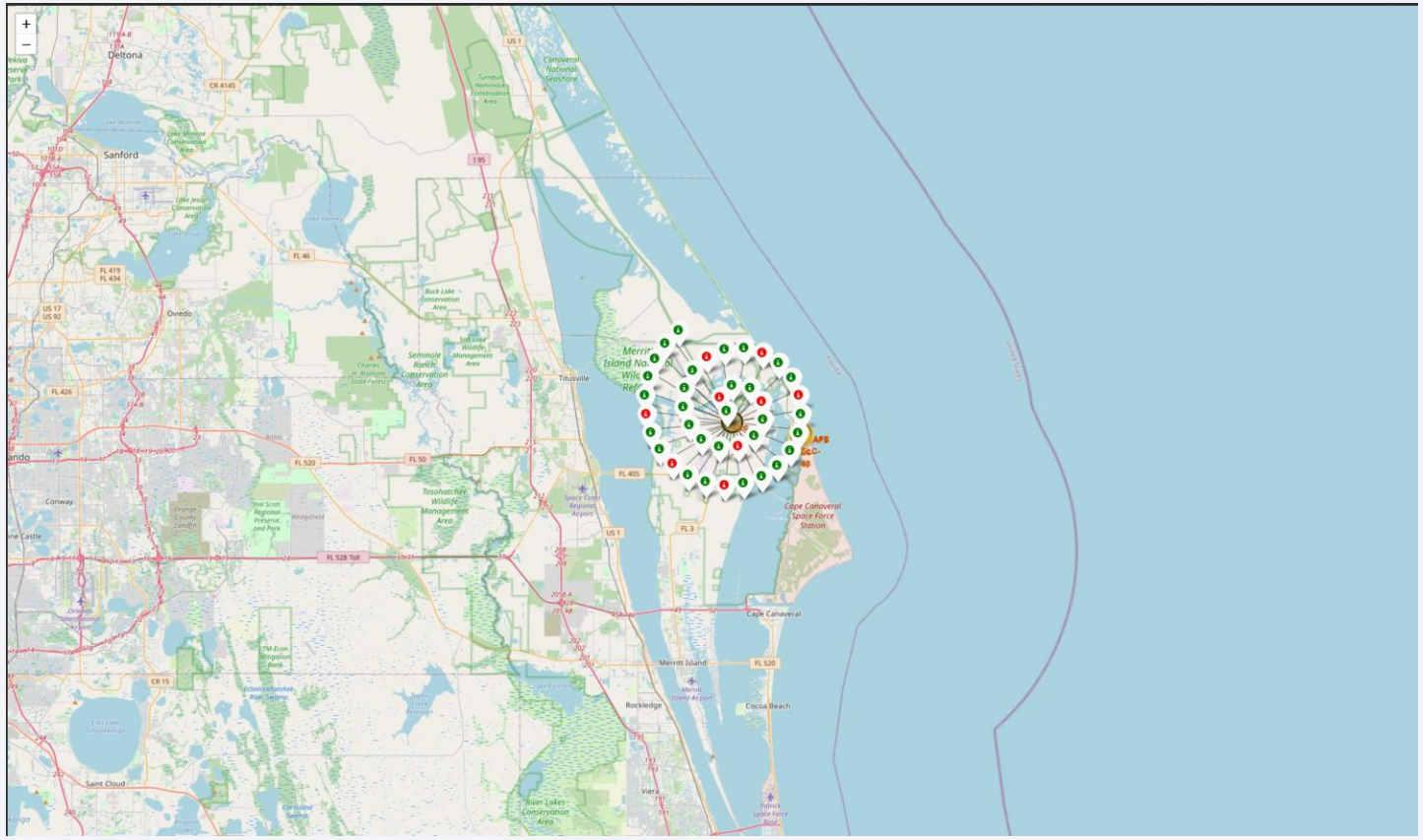
Launch Site Proximities



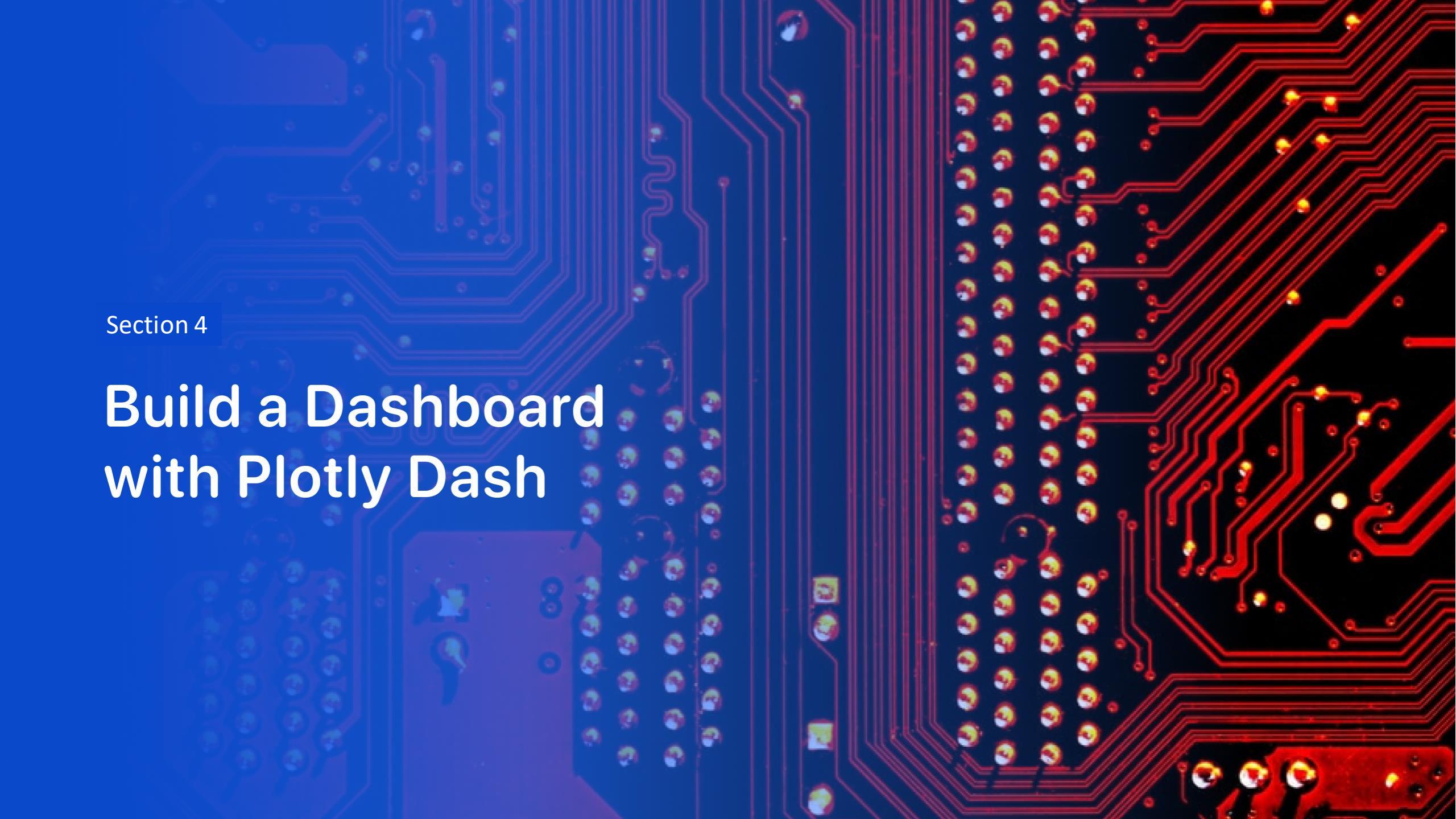
This example shows KSC LC-39A and a few close proximity landmarks including the nearest coastline, railway and freeway.

These landmarks are shown with a line that displays the distance to the launch site in kilometers.

Launch Outcomes Using Colored Labels

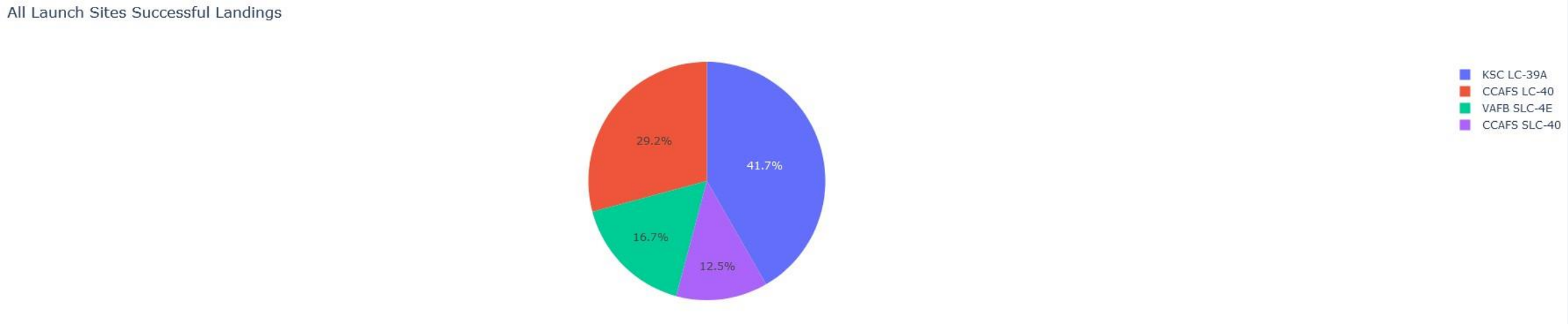


Each launch site contains colored labels of mission outcomes. Green labels represent successful missions and red labels represent failed missions.


Section 4
Build a Dashboard with Plotly Dash

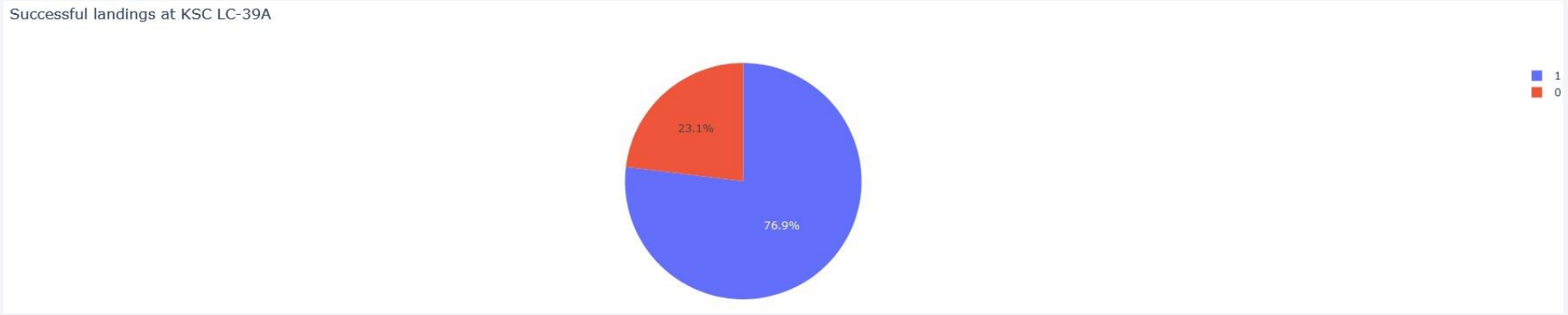
Successful Landings Across All Launch Sites

- This pie chart shows the launch success of each launch site and the percentage of successful landings compared to the other sites.
- You can see site KSC LC-39A had the most successful landings compared to the other sites and makes up 41.7% of the total successful landings.
- Site CCAFS SLC-40 had the least amount of successful landings.



Launch Site With The Most Successful Landings

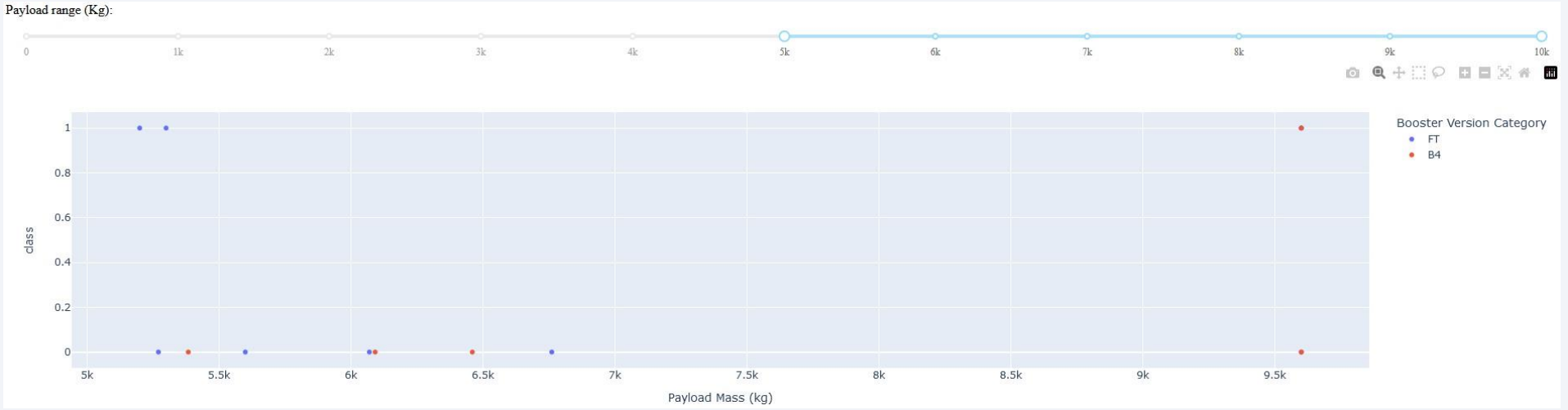
As mentioned in the previous slide, Site KSC LC-39A had the most successful landings with a total of about 77% success rate.



Payload Vs. Launch Outcome

Launch Outcomes for Payloads <= 5,000 kg

Launch Outcomes for Payloads >= 5,000 kg

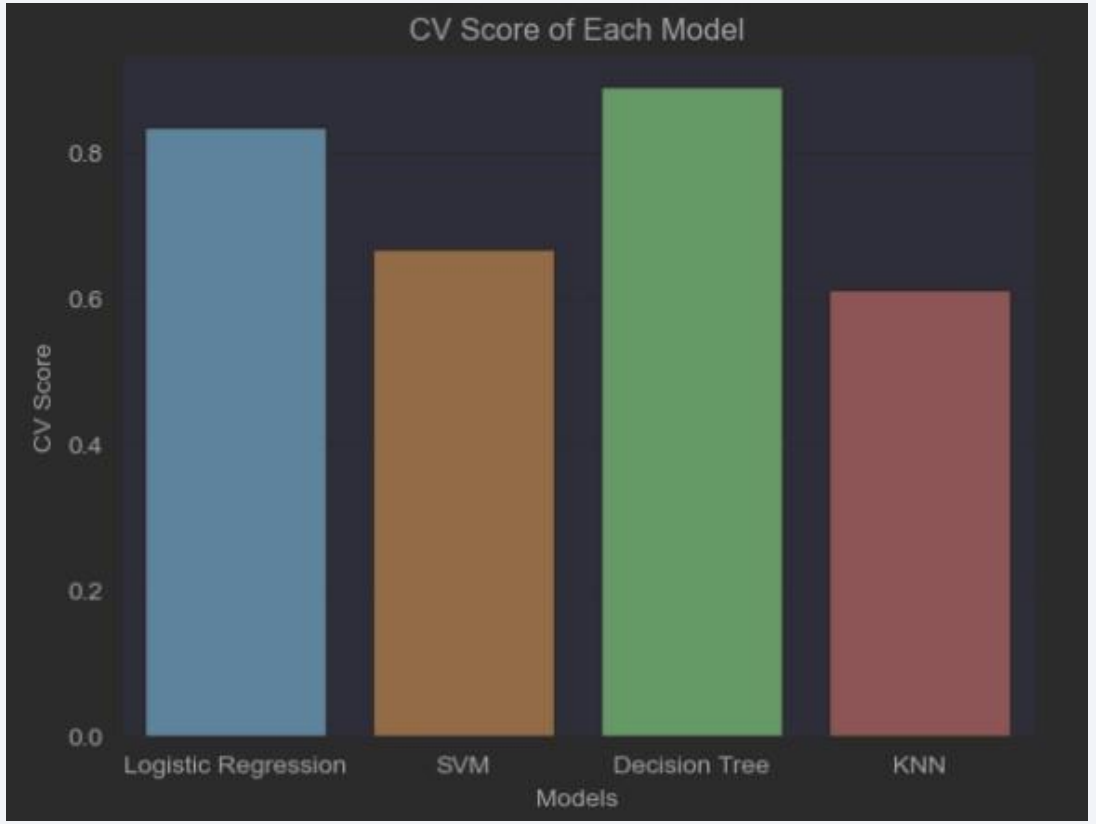


Section 5

Predictive Analysis (Classification)

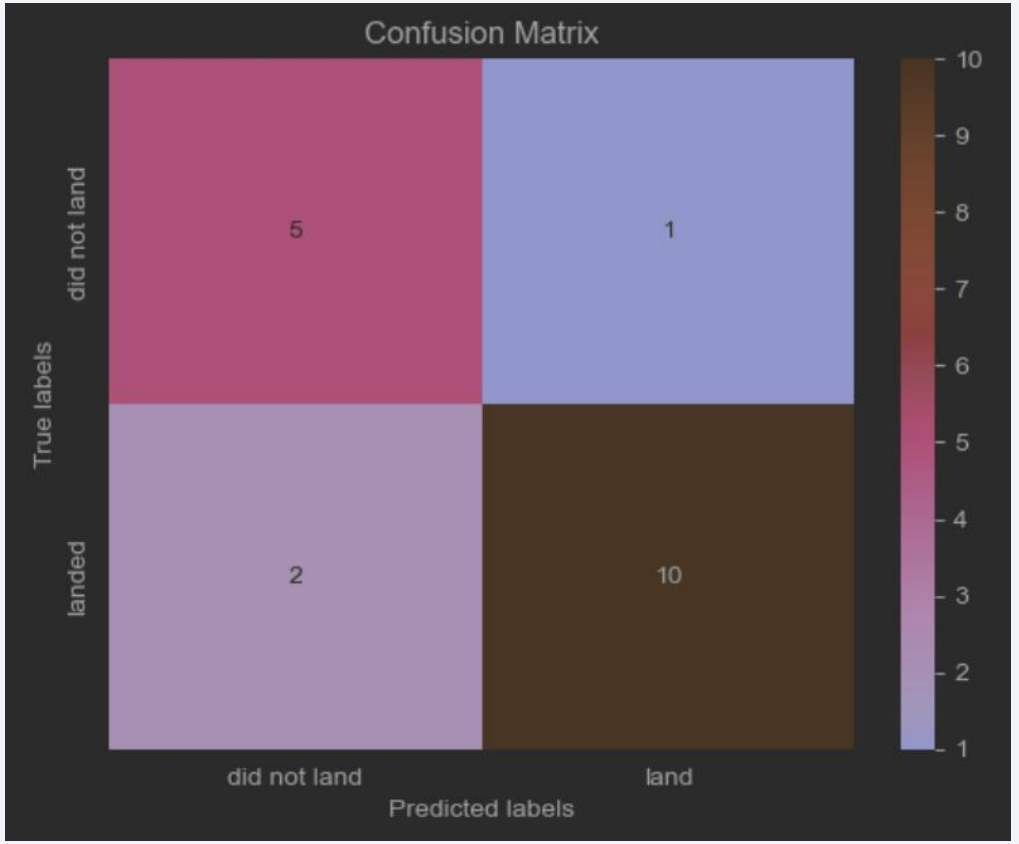
Classification Accuracy

- Creating a bar chart to compare the scores of each model implemented we can see the logistic regression model and decision tree scored the highest.
- The decision tree performed better than the logistic regression model and is the champion model with a score of 85%.



Confusion Matrix – Decision Tree

- Looking at the confusion matrix of the decision tree's performance we can see how the model correctly predicted true and false labels.
- The Decision Tree model predicted 5 True Positive labels out of 6 or 83%.
- The Decision Tree model predicted 10 True Negative labels out of 12 or 83%.



Conclusions

- The success rates for SpaceX launches are positively correlated with time.
- KSC LC 39A had the most successful launches.
- The Decision Tree and Logistic Regression models performed the best with scores of 83%.
- The Decision Tree model was championed based on better recall performance and true predictions.

Thank you!