IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Sahngyoon Rhee
June 28, 2023

# Outline

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

# Executive Summary

- Accessed data of SpaceX's Falcon 9 launches through SpaceX REST API
- Analyzed data such as launch date, success or failure of launch, success or failure of landing, reason for failures of each, payload weight, launchpad location, and rocket orbit in space.
- After analysis, I have found that the features of a falcon 9 launch that would most likely lead to a successful launch is one with FT booster version, has payload between 300 kg and 5330 kg, aims for SSO orbit in space, and is launched from KSC LC 39-A launch site in California.

# Introduction

- SpaceX has a leading competitive edge on space exploration and satellite launch due to its economical reuse of parts of a space rocket
- According to its website, SpaceX can launch one of its rockets, Falcon 9, with 62 million dollars; other providers may spend more than 165 million dollars to launch a rocket into space
- Analysis of SpaceX's launch data would give us significant insights into what are some features of a successful launch
- The goal of this data analysis is to find the features of Falcon 9 launch that would most likely lead to a success

Section 1

# Methodology

# Methodology

## Executive Summary

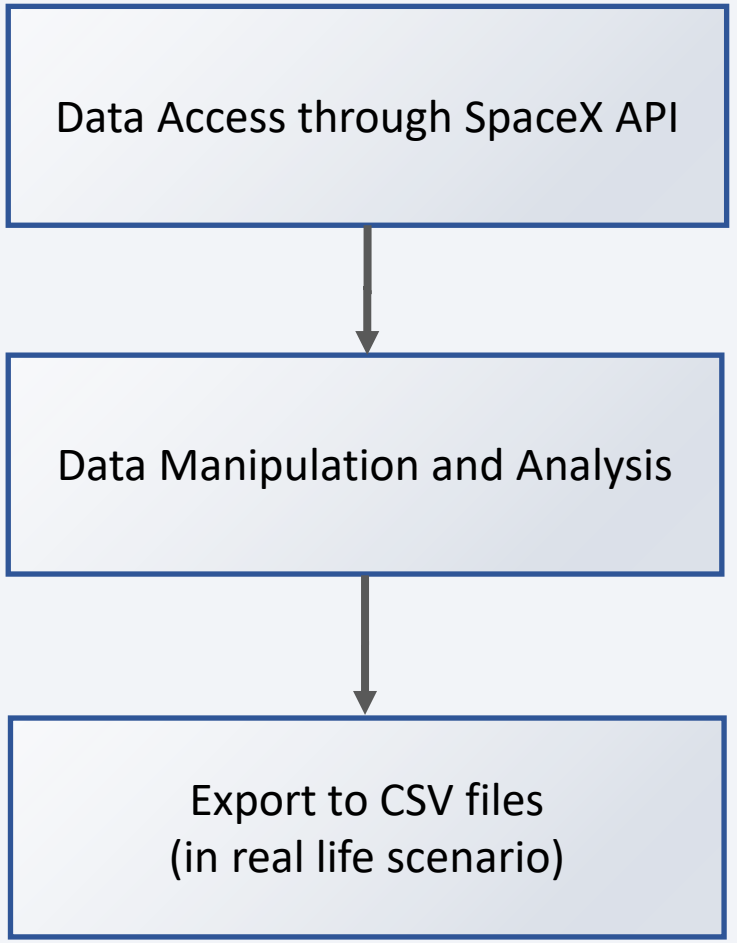
- Data collection methodology:
  - Requested data through SpaceX REST API
- Perform data wrangling
  - Replaced missing payload mass value of five launches with average of all other
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

# Data Collection

- Data collected through SpaceX API
- Collected information such as:
  - Launch date
  - Success/failure of launch + reason for failure
  - Success/failure of landing + reason for failure
  - Type of payloads and their weight
  - Launchpad location
  - Rocket orbit in space

# Data Collection – SpaceX API

- Used REST calls to SpaceX API to secure data
- Cleaned and analyzed data stored in local Jupyter (as this is a practice project)
- Github: [Practice-for-Data-Science/Step_1 Data Collection API.ipynb at main · srheegit/Practice-for-Data-Science (github.com)]



# Data Collection - Scraping

- Made URL request on Wikipedia page about Falcon 9 Launches
- Extracted the launch information table on the website and created a dataframe with the information
- [Practice-for-Data-Science/Step 2 Webscraping.ipynb at main · srheegit/Practice-for-Data-Science (github.com)]

# Data Wrangling

- Calculated the number of:
  - launches on each site
  - launches of each orbit in space
  - Landing successes of each landing type and location
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose
- Practice-for-Data-Science/Step_3_Data_Wrangling.ipynb at main · srheegit/Practice-for-Data-Science (github.com)

# EDA with Data Visualization

| Made visualizations of… | …to see this trend: |
|---|---|
| Scatterplot of Flight number and launch site colored red or blue (success or failure) | as the flight number increased (i.e. as time progressed), more frequent successes. The launch site CCAPS SLC 40 showed greatest improvement and most frequently used in latest launcheSs |
| Scatterplot of Payload Mass. Vs. lauch site color red or blue | for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000). |
| Success rate of launches for each orbit | ES-L1, GEO, HEO, and SSO had perfect launch success rate, while GTO orbit had the lowest rate |
| Scatterplot of Flight number and Orbit type colored red or blue | in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit. |

Practice-for-Data-Science/Step 5 EDA Data Visualization.ipynb at main · srheegit/Practice-for-Data-Science (github.com)

# EDA with SQL

- Performed SQL queries to find out data such as:
  - Launch sites
  - The total and average payload mass in kg
  - Boosters that have success in drone ship and have payload in a certain range
  - Total number of successful and failure mission outcomes
  - Names of booster versions with maximum payload
- Practice-for-Data-Science/Step 4 Eda SQL.ipynb at main · srheegit/Practice-for-Data-Science (github.com)

# Build an Interactive Map with Folium

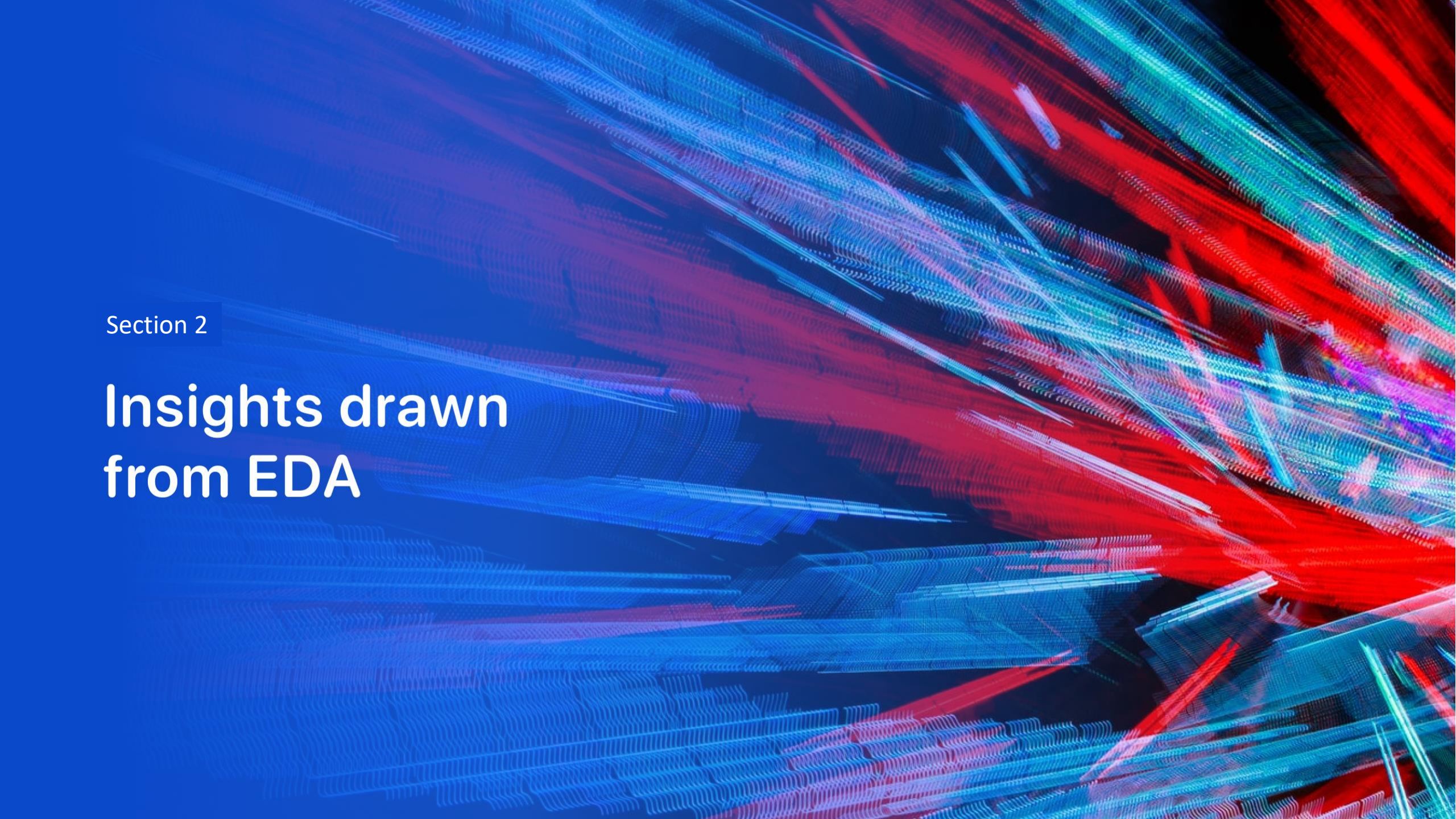
- Using Folium, created a map of marked launch sites and their respective number of successful and unsuccessful launches
- Through the visuals, we can clearly see that all launches are mostly done in Florida, where three launch sites are in very close proximity
- Most launch sites are close to:
  - Railroads
  - Shorelines
  - Highways
- [Practice-for-Data-Science/Step 6 Launch Site Locations.ipynb at main · srheegit/Practice-for-Data-Science (github.com)]

# Build a Dashboard with Plotly Dash

- Dashboard contains:
  - One drop-down to select all launch sites or a specific launch site
  - One range slider to select payload mass in kilograms
- Displays:
  - Pie chart of successes vs. failures on a specific launch site if a site is selected
  - A scatterplot of payload mass and lauch sites colored red or blue (success or failure)
  - Data adjusted accordingly if all sites are selected (shows total share of successes of different launch sites
- [Practice-for-Data-Science/Step 7 Dash Application Code.ipynb at main · srheegit/Practice-for-Data-Science (github.com)]

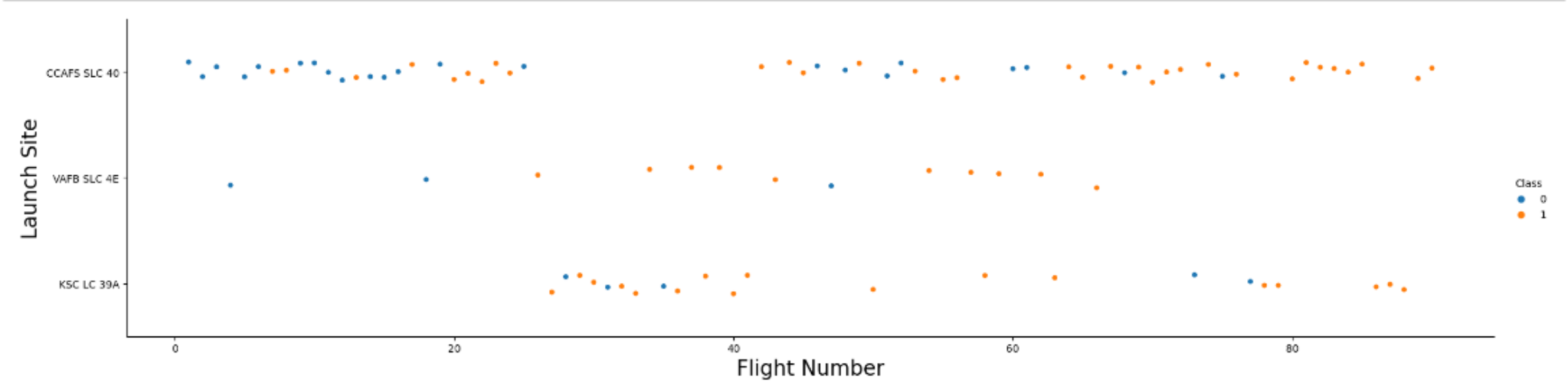
# Predictive Analysis (Classification)

- Used one-hot encoding, standardization, train_test_split, and various metrics and machine learning models to see that Logistic Regression was the best model to use in predicting the success or failure given a set of features
- Practice-for-Data-Science/Step 8 Machine Learning Prediction.ipynb at main · srheegit/Practice-for-Data-Science (github.com)


Section 2
Insights drawn from EDA

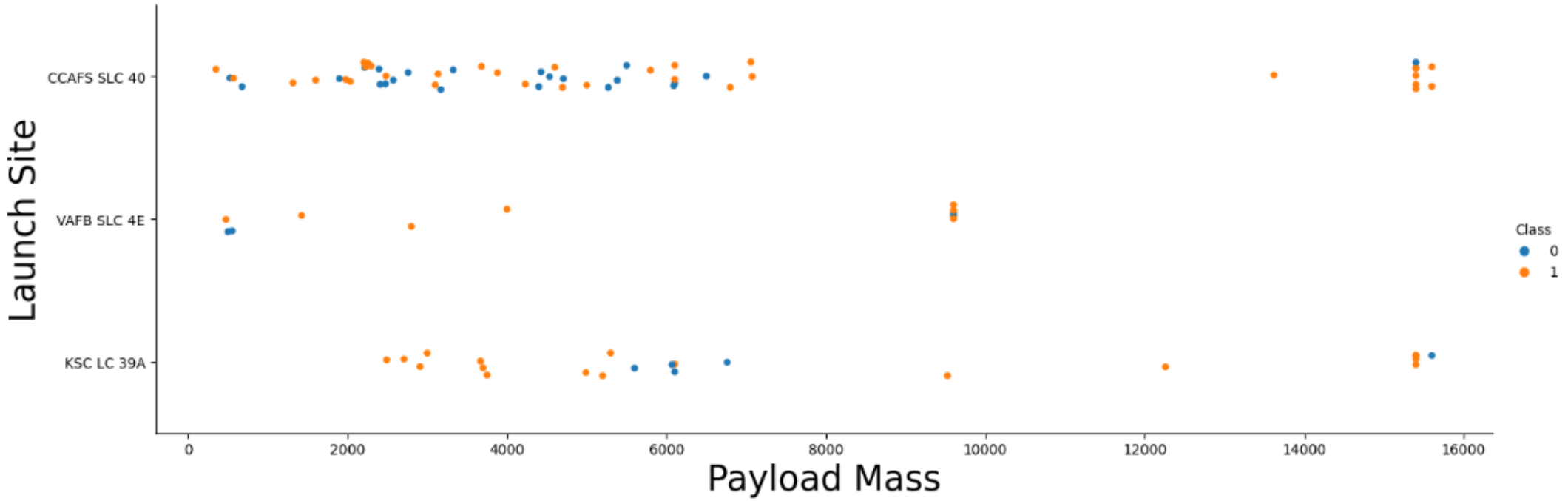
# Flight Number vs. Launch Site

- Class 0 is failure and 1 is success
- As the flight number increases, we see that success rate increases for all three launch sites
- This could be because there are insights gathered from earlier launch failures to improve on future launches



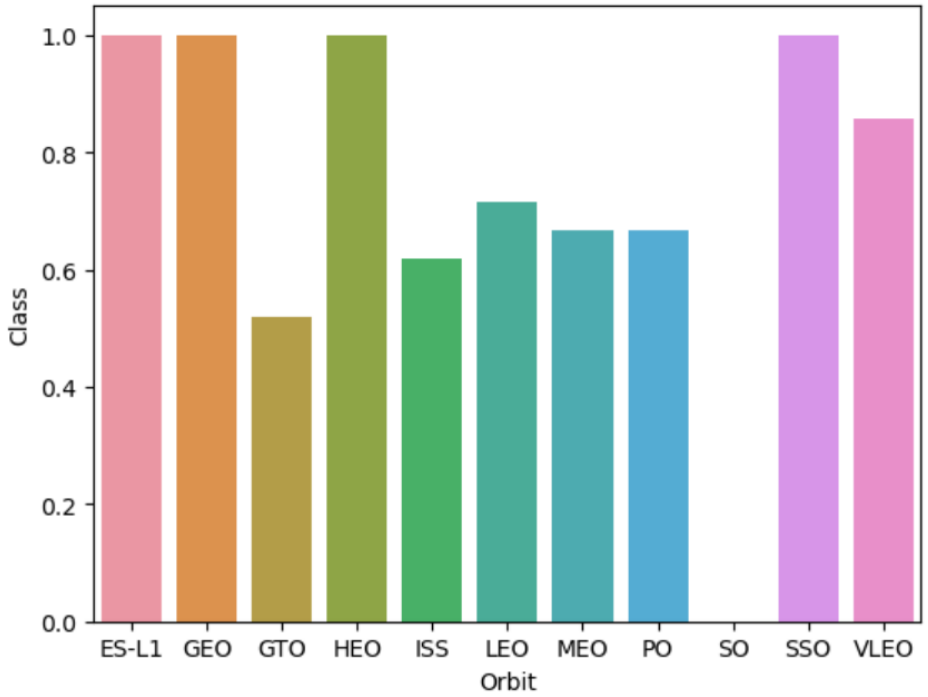
# Payload vs. Launch Site

- Blue (0) is Failure, Orange (1) is Success
- The launch site CCAFS SLC-40 is best suited for heavier payload mass
- The launch site KSC LC 39A is best suited for light payload mass
- For the site FVAFB-SLC, there are no rockets launched for heavy payload



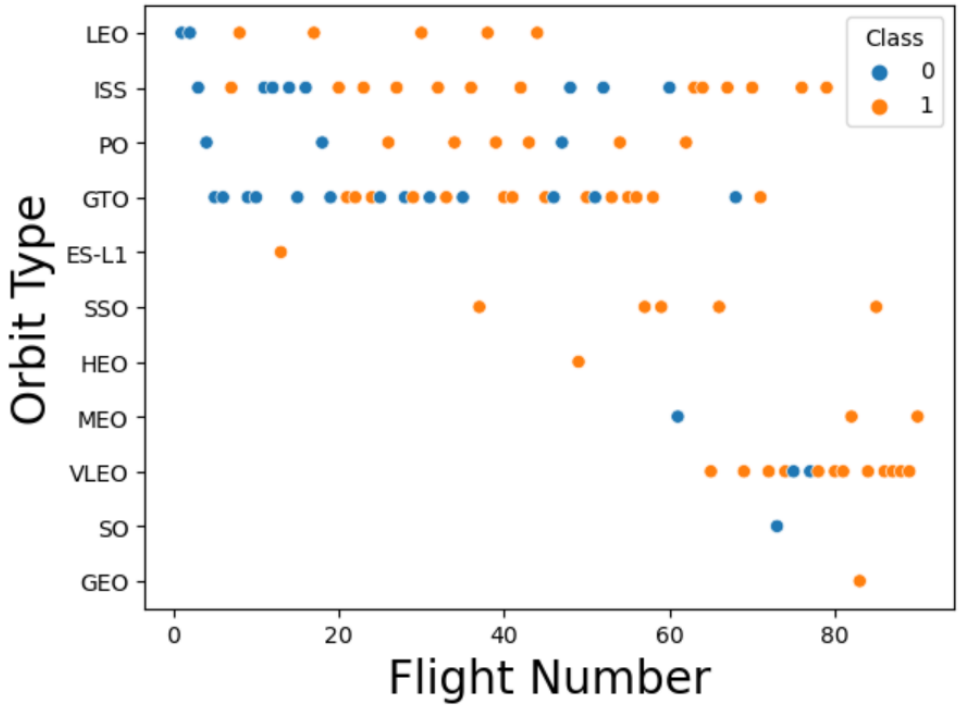
# Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, SSO has perfect success rate
- GTO and ISS are among the lowest success rate



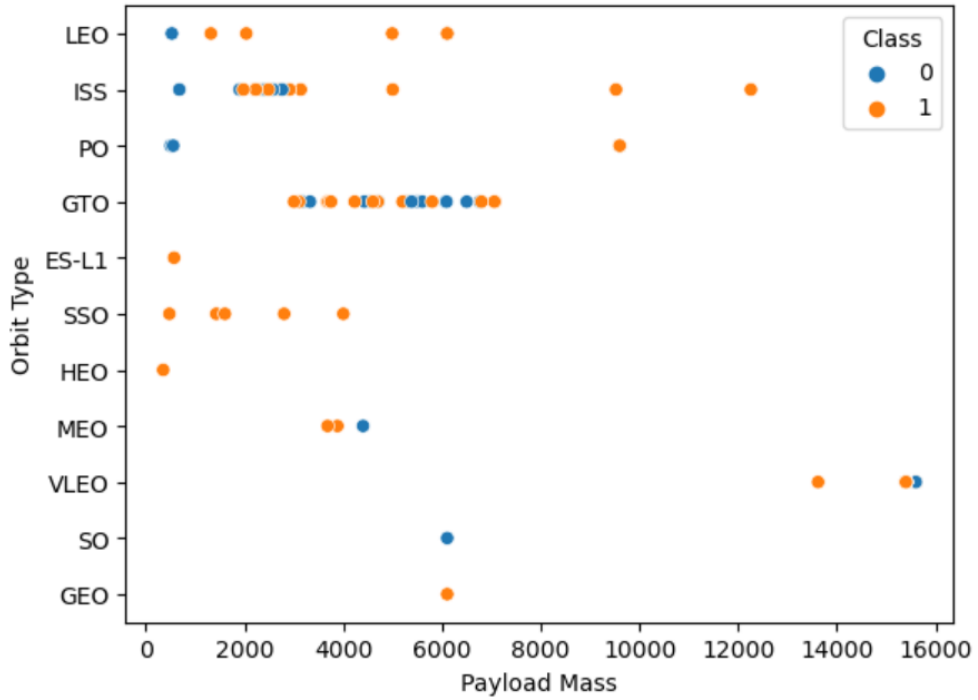
# Flight Number vs. Orbit Type

- LEO shows clear signs of improvement in performance
- ISS, PO, and VLEO also shows signs of improvement
- In GTO the result seems to be mixed, indicating the difficulty of launching to the GTO orbit



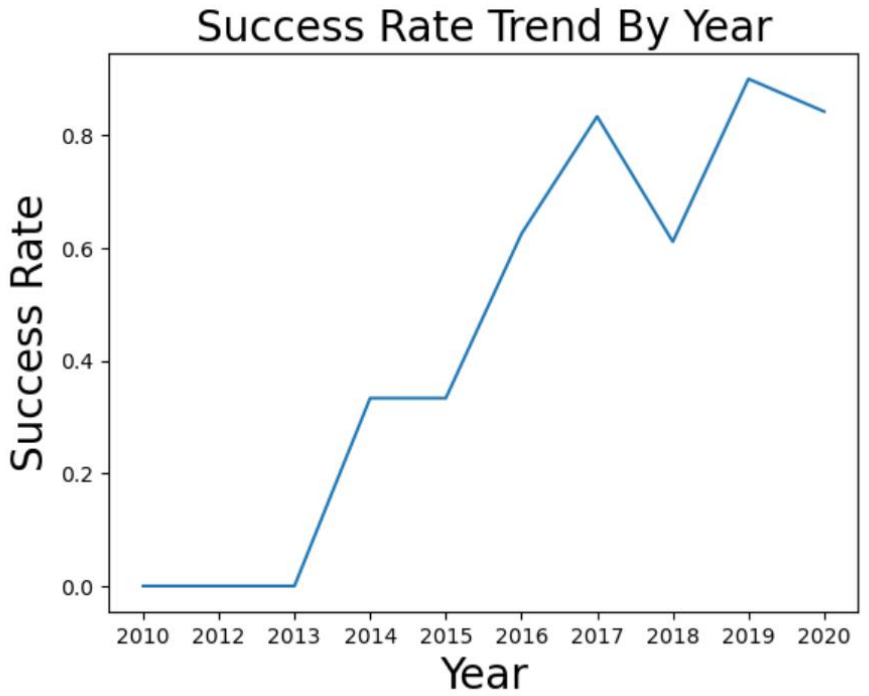
# Payload vs. Orbit Type

- For heavy payloads, Polar, LEO, and ISS have high positive landing rate
- For GTO, there is no immediate impact of payload mass to positive landing rate



# Launch Success Yearly Trend

- Success rate has increased in general as years progressed



# All Launch Site Names

- Names of Unique Launch Sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E
- First three sites are located close to one another in about 50 miles east of Orlando, Florida
- VAFB SLC-4E is about 140 miles northwest of Los Angeles

# Launch Site Names Begin with 'CCA'

- First five launches in launch sites beginning ‘CCA’ – all in Florida
- All five had small payload launches, successful mission outcome, and no landings

```
In [40]: %sql select * from spacextbl where launch_site like 'CCA%' limit 5;
 * sqlite:///my_data1.db
Done.
```

Out[40]:

| 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

- Total payload carried by boosters from NASA: **45596 KG**

# Average Payload Mass by F9 v1.1

- Average payload carried by booster F9 v1.1: **2928.4 KG**

# First Successful Ground Landing Date

- First successful landing on ground pad: **January 8, 2018**

## Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

- F9 FT: B1022, B1026, B1021.2, B1032.1, B1031.2
- F9 B4: B1040.1, B1043.1
- F9 B5: B1046.2, B1047.2, B1046.3, B1048.3, B1051.2, B1060.1, B1058.2, B1062.1

# Total Number of Successful and Failure Mission Outcomes

- Total number of failure mission outcomes: 1
- Total number of successful mission outcomes: 100
- Most of Falcon 9 launches' landing failures are done on purpose

# Boosters Carried Maximum Payload

- Names of boosters which have carried the maximum payload mass:

- F9 B5: B1048.4, B1049.4, B1051.3, B1056.4, B1048.5, B1051.4, B1049.5, B1060.2, B1058.3, B1051.6, B1060.3, B1049.7

# 2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

| Month_names | Booster_Version | Launch_Site |
|---|---|---|
| 10 | F9 v1.1 B1012 | CCAFS LC-40 |
| 04 | F9 v1.1 B1015 | CCAFS LC-40 |

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

| Landing_Outcome | count |
|---|---|
| Success | 20 |
| No attempt | 9 |
| Success (drone ship) | 8 |
| Success (ground pad) | 7 |
| Failure (drone ship) | 3 |
| Failure | 3 |
| Failure (parachute) | 2 |
| Controlled (ocean) | 2 |
| No attempt | 1 |

- Rank of 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

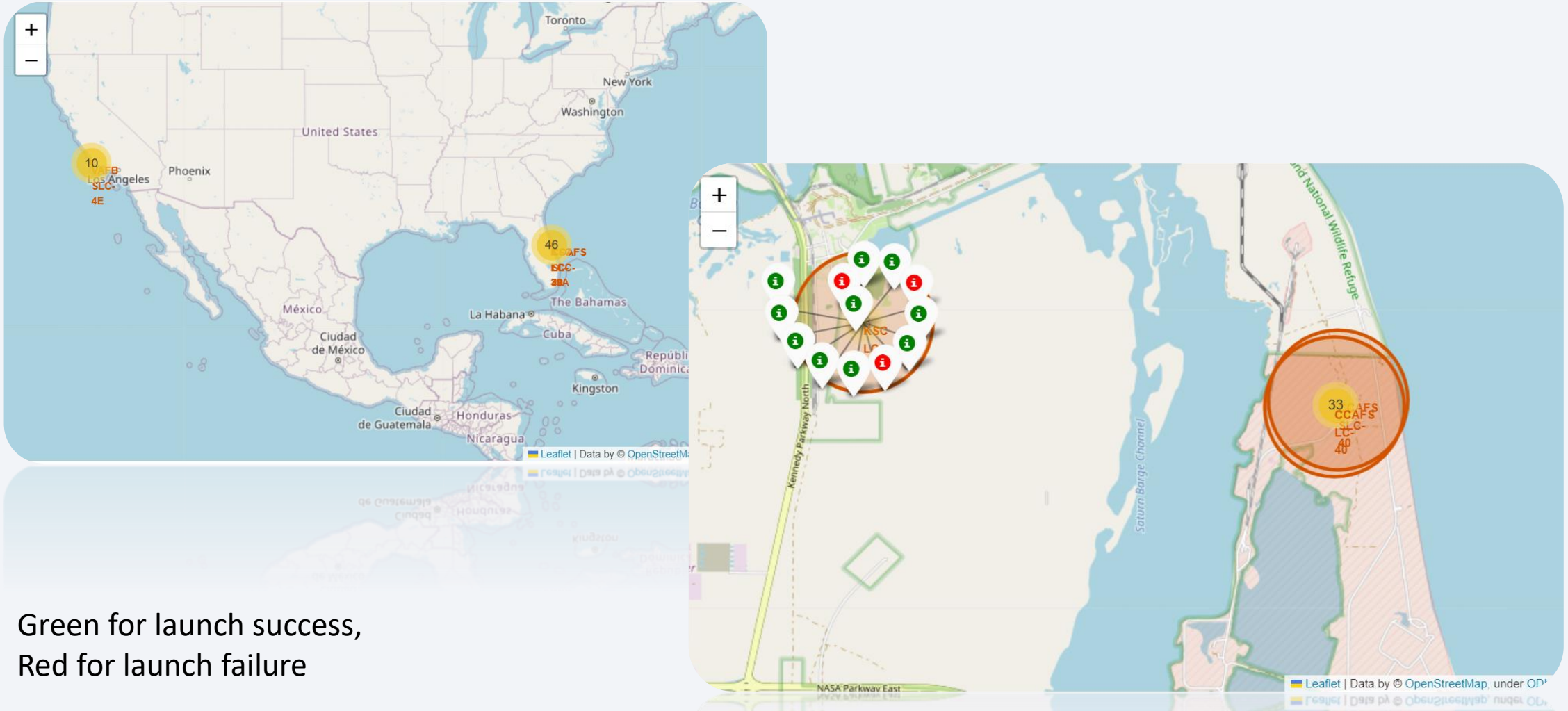
Section 3

# Launch Sites Proximities Analysis

# Map of all launch locations

- Three out of four launch locations are in Florida

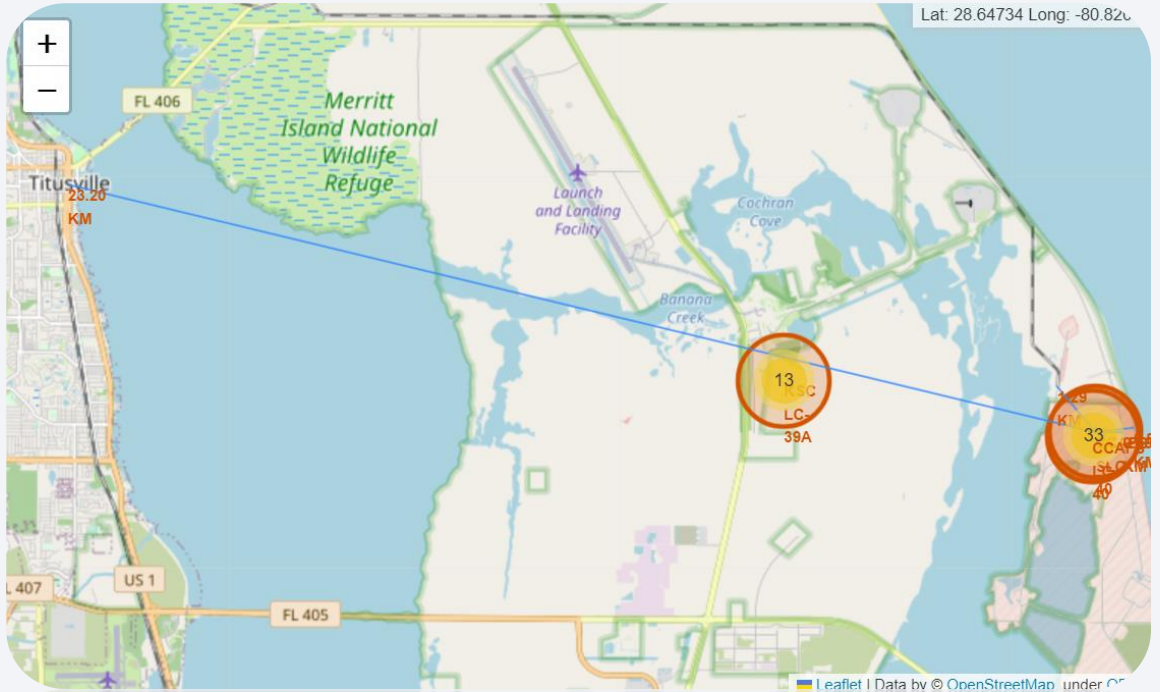
# Color-labeled launch outcomes

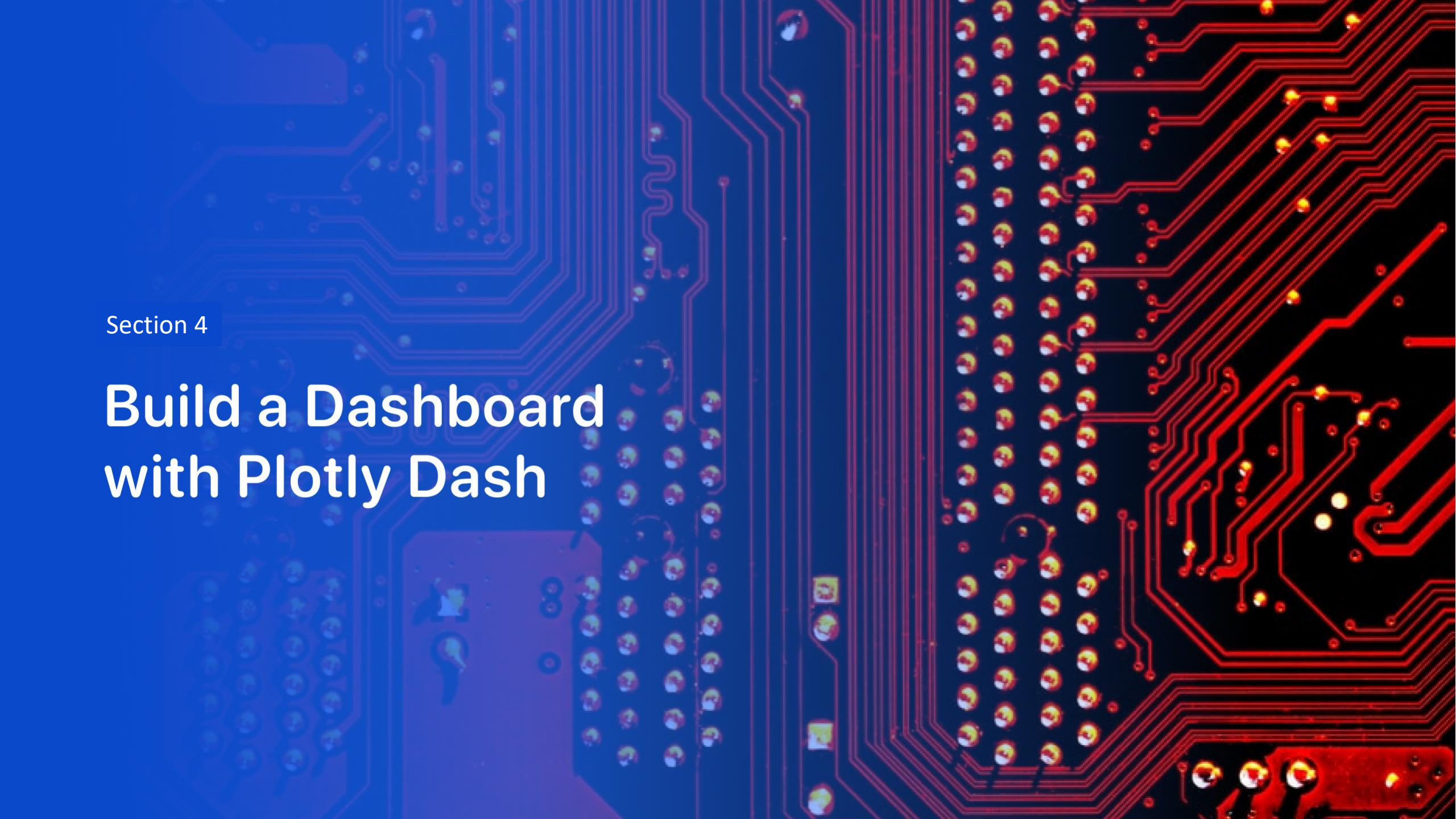



Green for launch success,
Red for launch failure

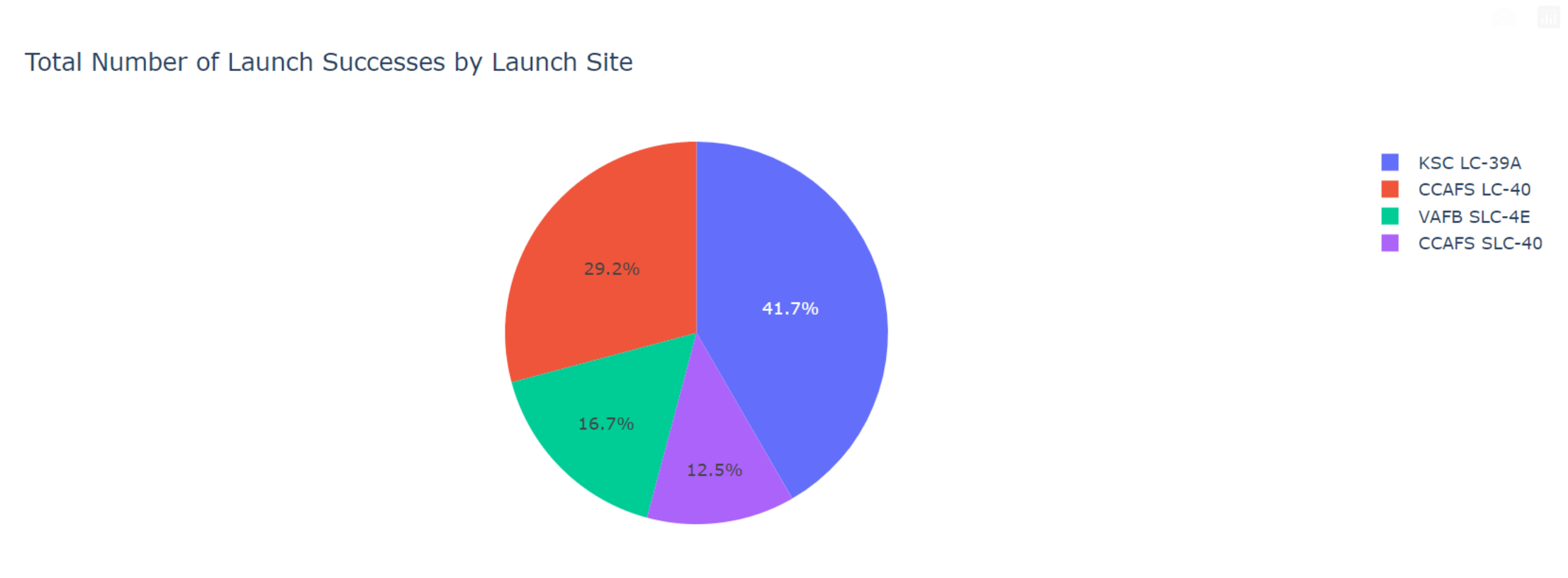
# Places surrounding launch sites

- Launch sites are close to railway, highway, coastline, as shown in the map for one launch location in Florida

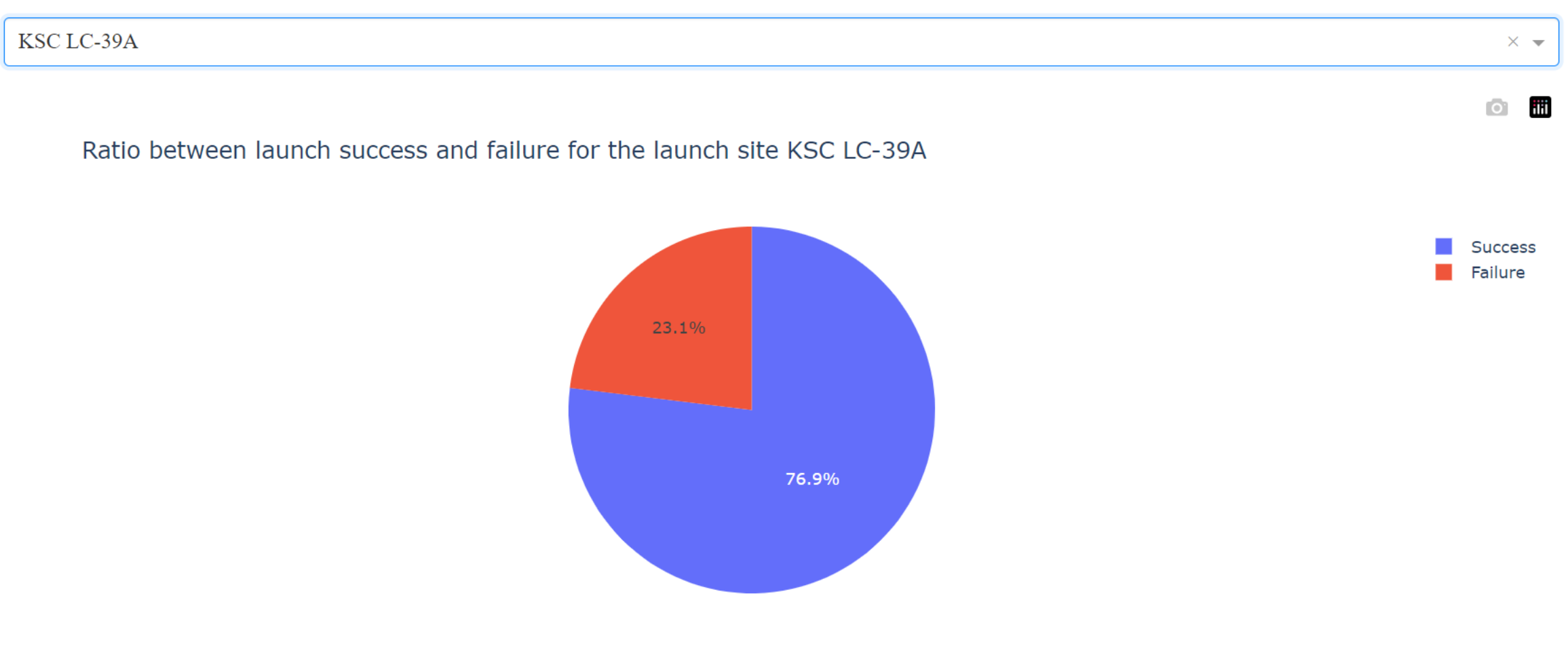




Section 4
Build a Dashboard with Plotly Dash

# Launch Success Count for All Sites



# Launch site with the highest launch success ratio



# Payload vs. Launch Outcome for All Sites

- Scatterplot on next slide
- Payload vs. Launch Outcome scatter plot for all sites, with payload approximately between 300kg and 5330kg
- The selected payload range has the largest success rate
- FT booster version has the largest success rate

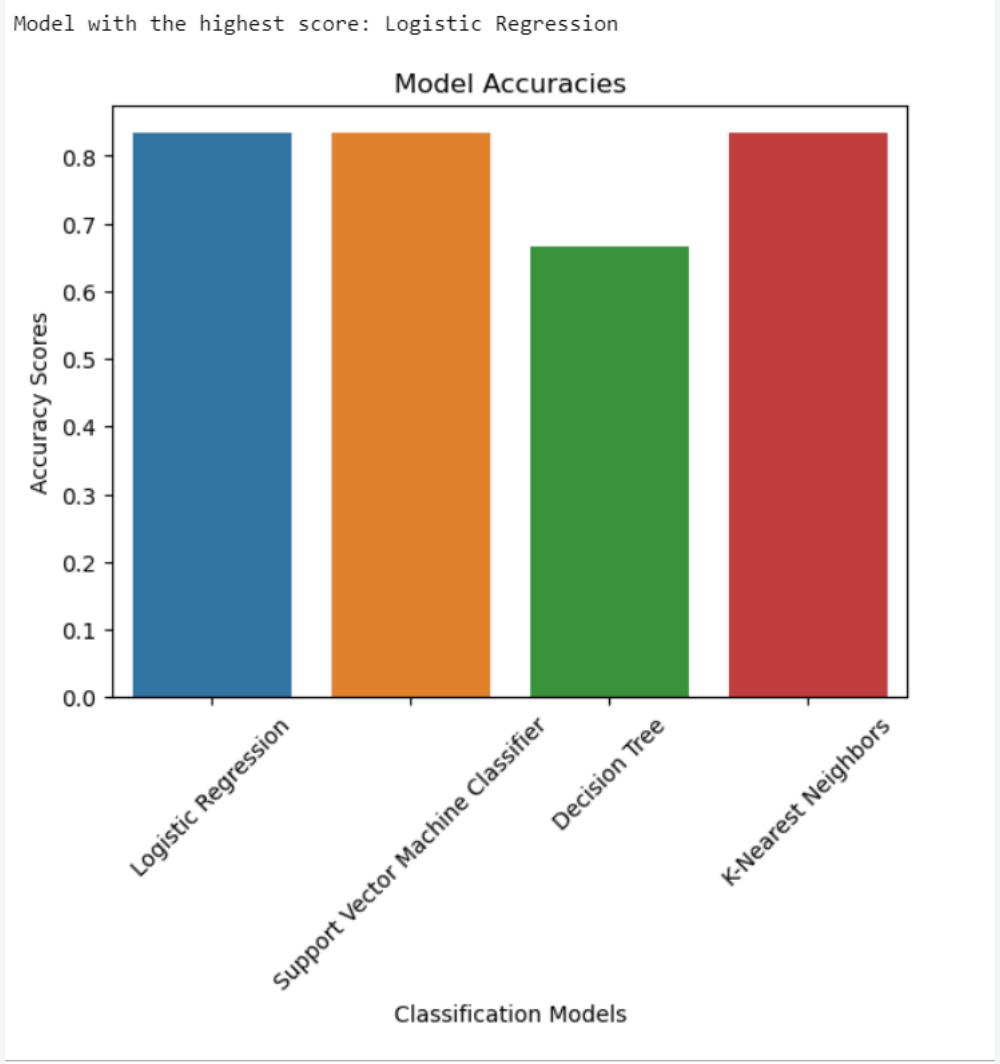
# Payload vs. Launch Outcome for All Sites

Section 5

# Predictive Analysis (Classification)

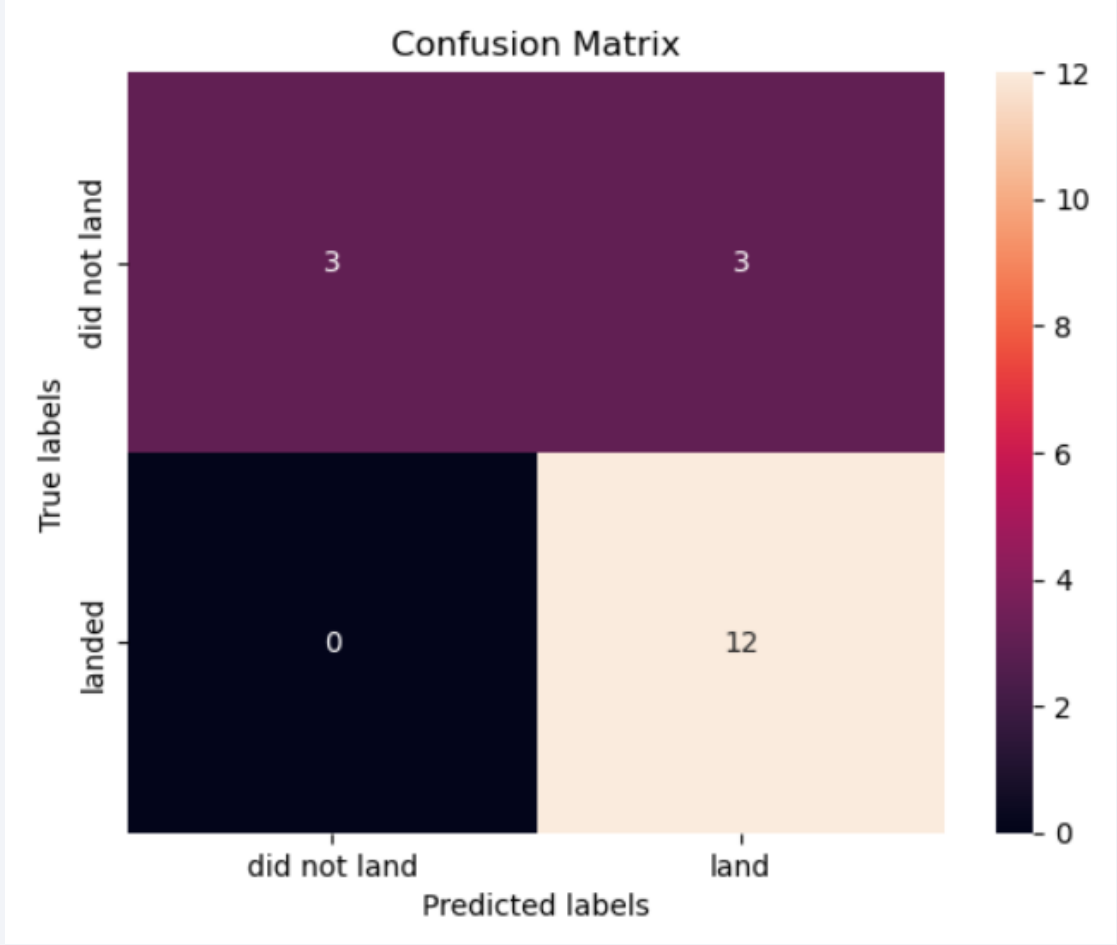
# Classification Accuracy By Classification Models

- Logistic Regression is the most accurate



# Confusion Matrix

- Confusion matrix of the best performing model, Logistic Regression
- We see that false positives are the only error of the logistic regression



# Conclusions

- Among all orbits with at least five launches, SSO is the orbit with the greatest success rate
- As shown by success rate by year and success rate by flight number, more trials leads to higher success rate
- Launch site KSC LC 39-A has the highest success rate, both overall and for light payload mass
- Payloads approximately between 300kg and 5330kg – which are relatively light - have the largest success rate
- FT booster version has the largest success rate
- **Therefore, the features of a falcon 9 launch that would most likely lead to a successful launch is one with FT BOOSTER VERSION, has PAYLOAD BETWEEN 300KG AND 5330KG, AIMS FOR SSO ORBIT IN SPACE, and IS LAUNCHED FROM KSC LC 39-A Launch site in California.**

# Appendix

- A portion of my Dash application code is shown on the right

```
# Read the airline data into pandas dataframe
spacex_df = pd.read_csv("spacex_launch_dash.csv")
max_payload = spacex_df['Payload Mass (kg)'].max()
min_payload = spacex_df['Payload Mass (kg)'].min()

# Create a dash application
app = dash.Dash(__name__)

# Create an app layout
app.layout = html.Div(children=[html.H1('SpaceX Launch Records Dashboard',
                                        style={'textAlign': 'center', 'color': '#503D36',
                                               'font-size': 40}),
                                # A dropdown list to enable Launch Site selection
                                # The default select value is for ALL sites
                                dcc.Dropdown(id='site-dropdown',
                                options=[{'label': 'All Sites', 'value': 'ALL'},
                                {'label': 'CCAFS LC-40', 'value': 'CCAFS LC-40'},
                                {'label': 'CCAFS SLC-40', 'value': 'CCAFS SLC-40'},
                                {'label': 'KSC LC-39A', 'value': 'KSC LC-39A'},
                                {'label': 'VAFB SLC-4E', 'value': 'VAFB SLC-4E'}
                                ],
                                value='ALL',
                                placeholder='All Sites',
                                searchable = True
                                ),
                                html.Br(),

                                # A pie chart to show the total successful launches count for all sites
                                # If a specific launch site was selected, show the Success vs. Failed counts for the site
                                html.Div(dcc.Graph(id='success-pie-chart')),
                                html.Br(),

                                html.P("Payload range (Kg):"),
                                # A slider to select payload range
                                dcc.RangeSlider(id='payload-slider',
                                                min=0, max=10000, step=10,
                                                marks={0: '0',
                                                       2500: '2500',
                                                       5000: '5000',
                                                       7500: '7500',
                                                       10000: '10000'},
                                                value=[0,10000]),

                                # A scatter chart to show the correlation between payload and launch success
                                html.Div(dcc.Graph(id='success-payload-scatter-chart')),
                                ])

# A callback function for `site-dropdown` as input, `success-pie-chart` as output
@app.callback(Output(component_id='success-pie-chart', component_property='figure'),
                    Input(component_id='site-dropdown', component_property='value'))
def get_pie_chart(entered_site):
    filtered_df = spacex_df
    if entered_site == 'ALL':
        filtered_df = spacex_df.groupby('Launch Site')['class'].sum().reset_index()
```

# Thank you!