# Space-X Launch History: Analysis and Predictions

Prepared by: Garrett Meek

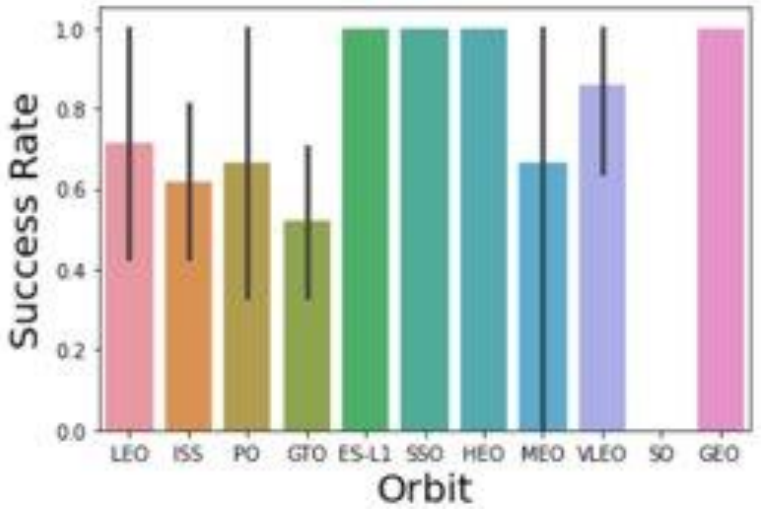
Date: 8-23-21

# Outline


Image credit:
https://phys.org/news/2021-02-spacex-starship-prototype-rocket-fireball.html

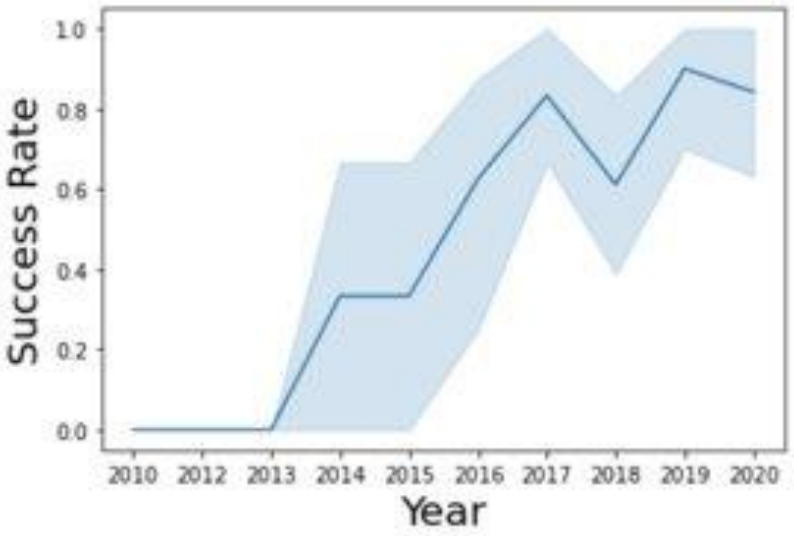
- Executive Summary
- Introduction
- Methodology
- Results
- Discussion
- Conclusion
- Appendix

# Executive Summary



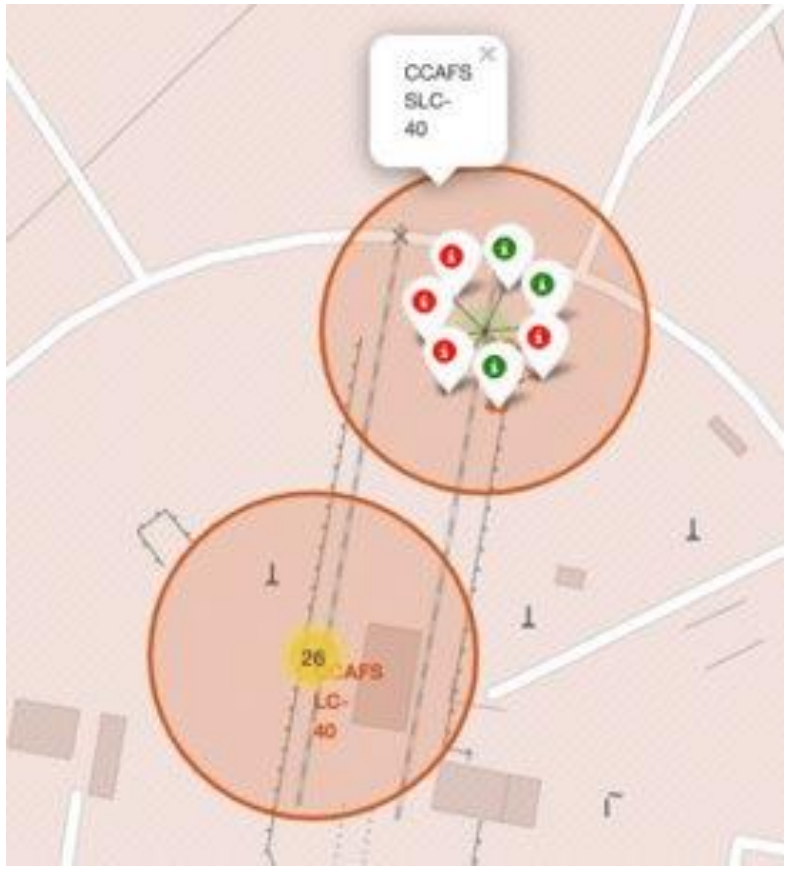

- Space-X rockets are unique
  - Rockets typically cost tens of millions of dollars each
  - Space-X rockets can be reused after a successful launch
- Predicting launch success is thus vitally important to predicting Space-X's performance as a business
- Space-X publishes data about each launch
- Analysis shows that launch success is:
  - Most likely for rockets that have payload masses > 9,000 kg
  - Steadily increasing over time
  - Higher for specific orbit types (see figure at left)
- A decision tree classifier, trained on these data, is able to predict launch success with ~94% accuracy

# Introduction




- Space-X rockets initially used 'Falcon 1' boosters, but eventually transitioned to 'Falcon 9' boosters
  - Launch success rates changed significantly with this shift
- Launches can occur at three different sites
- Rockets can follow one of ten different orbital trajectories
- The "payload mass" is an indicator of rocket performance, where a higher payload is desirable for Space-X's long-term business goals
  - The "payload mass" of Space-X rockets increased with the transition from 'Falcon 1' to 'Falcon 9' boosters

# Methodology




- Data collection methodology
  - Launch data was collected using the SpaceX API
- Data wrangling details
  - Launches with 'Falcon 1' boosters were removed due to low representation in the dataset
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis with the following models:
  - Logistic regression, K-nearest neighbors, support vector machine, decision tree

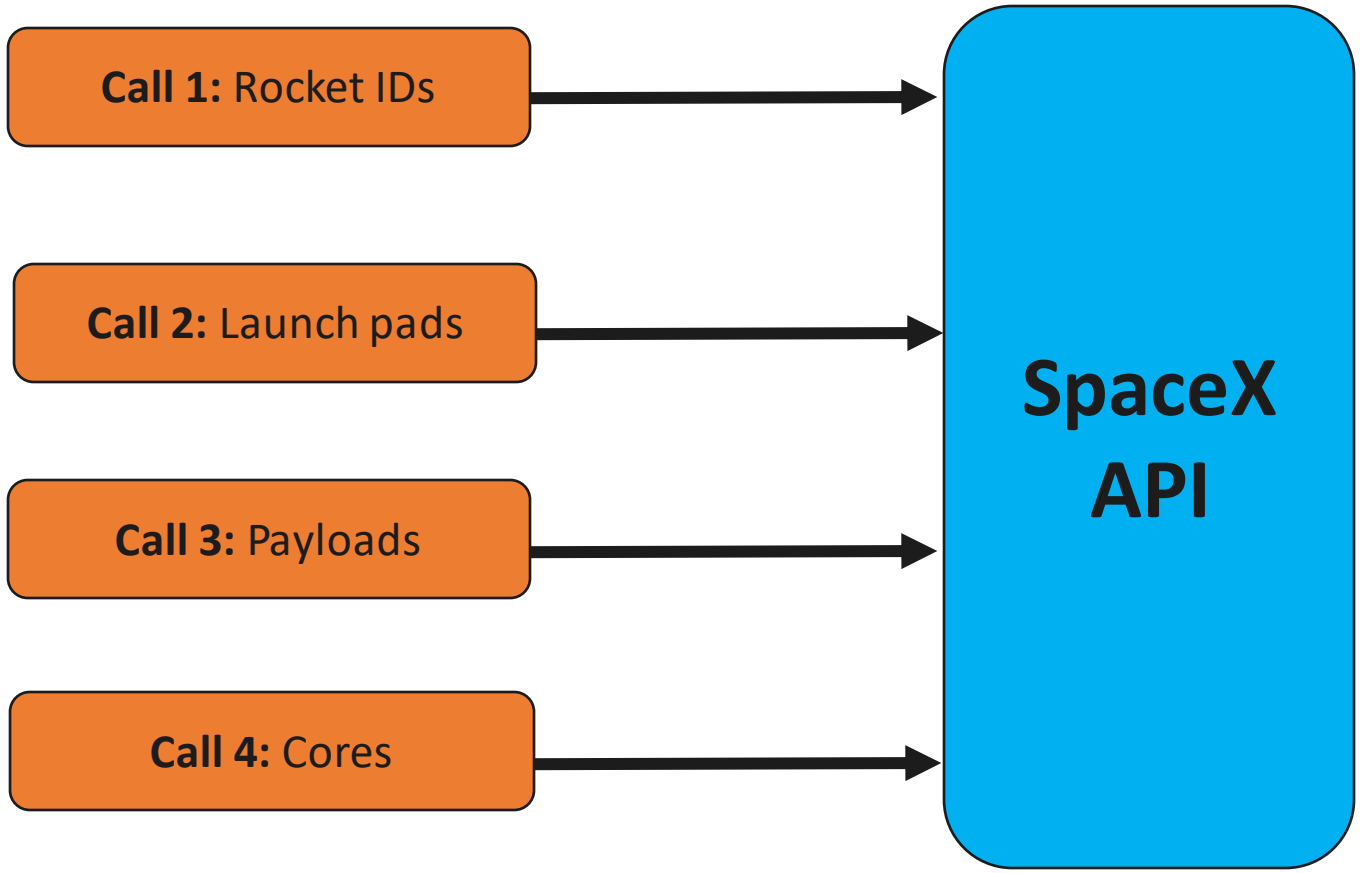
# Data collection

- Data was collected using the SpaceX API
- Requested JSON results were turned into a Pandas dataframe
- Kept only the features that we are interested in:
  - Rocket ID, payload mass, launchpad, date, cores
- Dropped the launch data for rockets with a "Falcon 1" booster
- The average payload mass was used to fill in missing data points

# Data collection – SpaceX API

1. Call SpaceX API to retrieve data using the list of rocket IDs
2. Call SpaceX API to retrieve data using the list of launch pads
3. Call SpaceX API to retrieve data using the list of payloads
4. Call SpaceX API to retrieve data using the list of cores
5. Keep only the features that we are interested in
6. Dropped the launch data for rockets with a "Falcon 1" booster
7. The average payload mass was used to fill in missing data points

# Added a flowchart of SpaceX API calls here

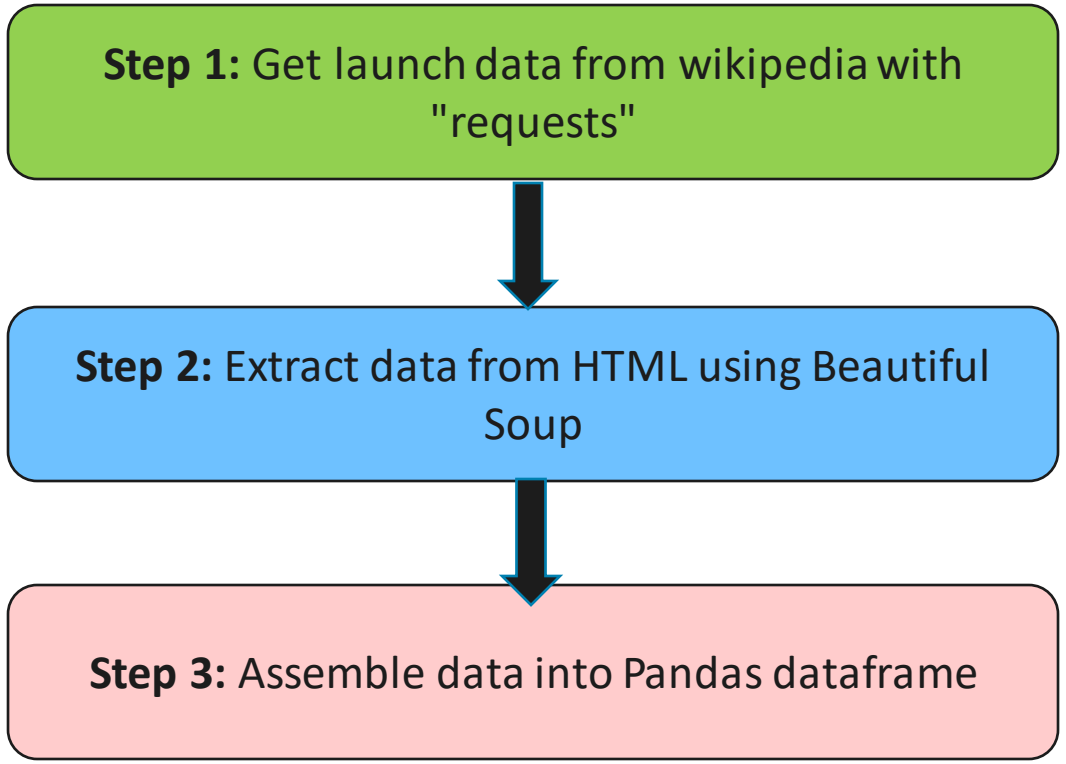



Github Link to Jupyter Notebook with Data Collection

# Data collection – Web scraping

1. Launch data was also "scraped" from a Wikipedia site that provides a snapshot of data for each launch, using "requests"
2. Beautiful Soup was used to extract the data as text from the "scraped" HTML
3. HTML data was assembled into a Pandas dataframe for further analysis

# Add a flowchart of web scraping here




Github Link to Jupyter Notebook with Web Scraping

# Data wrangling

- Landing outcomes were assigned to "good" and "bad" categories to simplify classification
- The launch success was one-hot coded to make a "Class" variable in which a value of 0 means the launch failed, and a value of 1 means the launch succeeded

Github Link to Jupyter Notebook with Web Wangling

# EDA with data visualization

- The following scatter plots were created during exploration of the data:
  - Payload mass vs. Flight Number
  - Launch Site vs. Flight Number
  - Launch Site vs. Payload mass
  - Orbit type vs. Flight Number
  - Orbit type vs. Payload mass
- A bar chart of launch success was created for each launch site
- A line chart of launch success as a function of payload mass was created

Github Link to Jupyter Notebook with EDA and Data Visualization

# EDA with SQL

- Loaded Coursera-provided CSV dataset into IBM DB2
- Used the following SQL queries:
  - `select distinct(launch_site) from SPACEXDATASET`
  - `select * from SPACEXDATASET where launch_site like '%CCA%' limit 5`
  - `select SUM(payload_mass__kg_) from SPACEXDATASET where customer like '%NASA%'`
  - `select AVG(payload_mass__kg_) from SPACEXDATASET where booster_version = 'F9 v1.1'`
  - `select MIN(DATE) from SPACEXDATASET where landing__outcome like '%ground pad%'`
  - `select booster_version from (select * from SPACEXDATASET where landing__outcome like '%drone ship%') where payload_mass__kg_ between 4000 and 6000select COUNT(mission_outcome) from SPACEXDATASET where mission_outcome like 'Success'`

Github Link to Jupyter Notebook EDA with SQL

# Build an interactive map with Folium

- Launch locations and success rate were analyzed/plotted using the Folium Python package
- Folium Circle objects were added to a map in order to plot the coordinates of each launch location
- A marker cluster object was created to enable better visualization of launch success rate, since circles and markers are clustered at the launch locations
- The distance was calculated to the nearest railway by creating and applying a distance function to the launch site coordinates

Github Link to Jupyter Notebook Visual Analytics with Folium

# Build a Dashboard with Plotly Dash

- Pie charts of the launch success for all launch sites, and by individual launch sites, were created on a dashboard where launch site can be selected from a menu
- A slider was created to select the payload mass range for displayed data
- A scatter plot was created to display the launch success for all missions, color-labeled by the booster version that was used

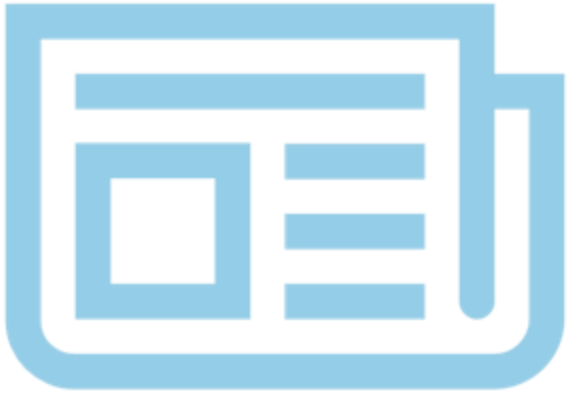
Github Link to Python File with Plotly Dashboard

# Predictive analysis (Classification)

- Data was standardized and separated into training (80% of data) and testing (20%) groups
- Classification models were trained in order to predict if a launch was successful
- Model parameters with a good fit (high prediction accuracy) were identified using the "GridSearchCV" function
- The following models were used:
  - Logistic regression, K-nearest neighbors, support vector machine, and a decision tree
  - The "score()" method was used to identify the model with the highest accuracy

Github Link to Jupyter Notebook Predictive Analysis Lab

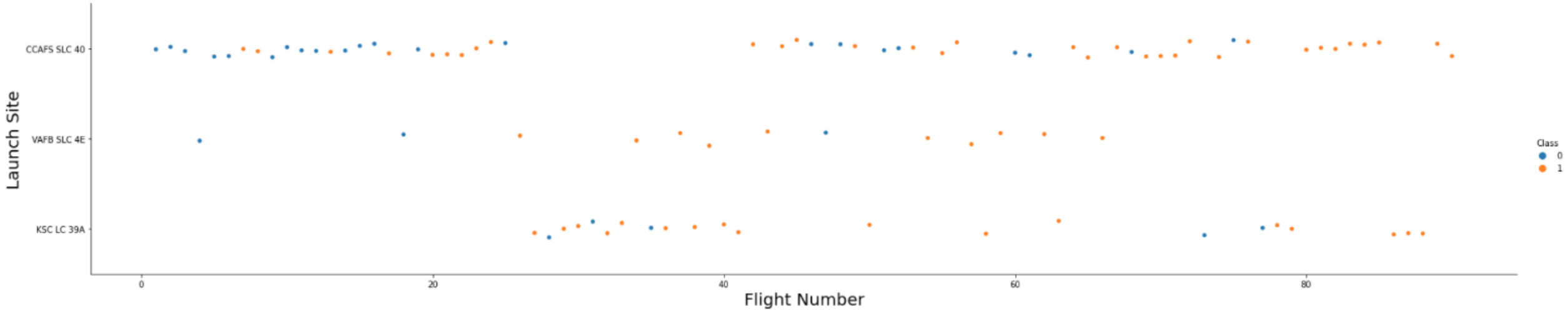
# Results



- Exploratory data analysis results
- Interactive analytics dashboard screenshots
- Predictive analysis results

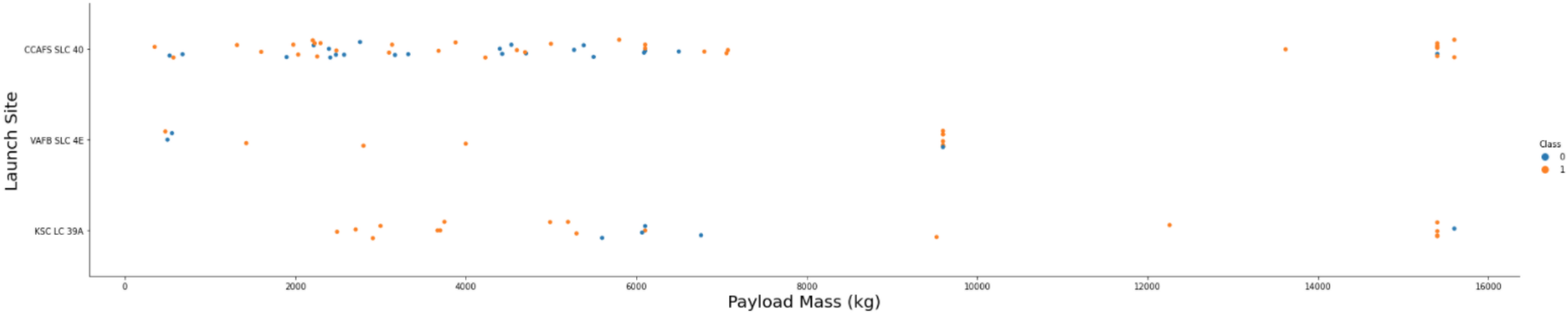
# EDA with Visualization

# Flight Number vs. Launch Site



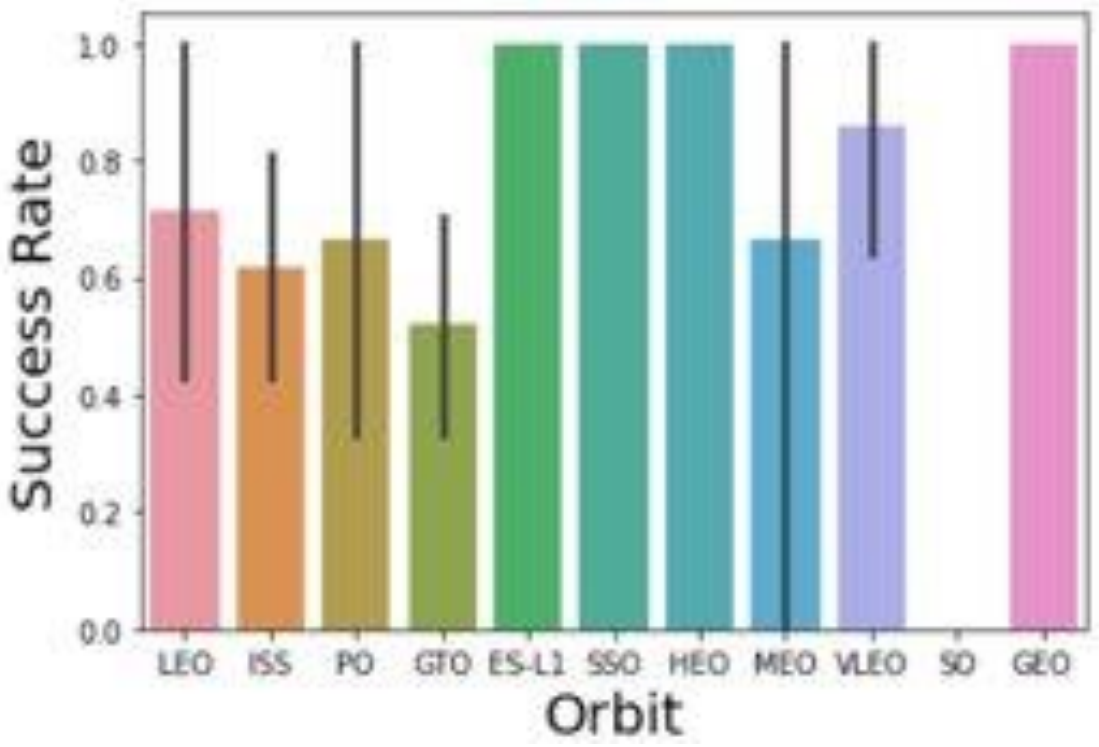

- The figure above shows how the launch sites that are used have changed over time, as the flight number increases with time
- The marker color shows the success of the launch, and demonstrates that the launch success rate has also improved over time, across all launch sites

# Payload vs. Launch Site



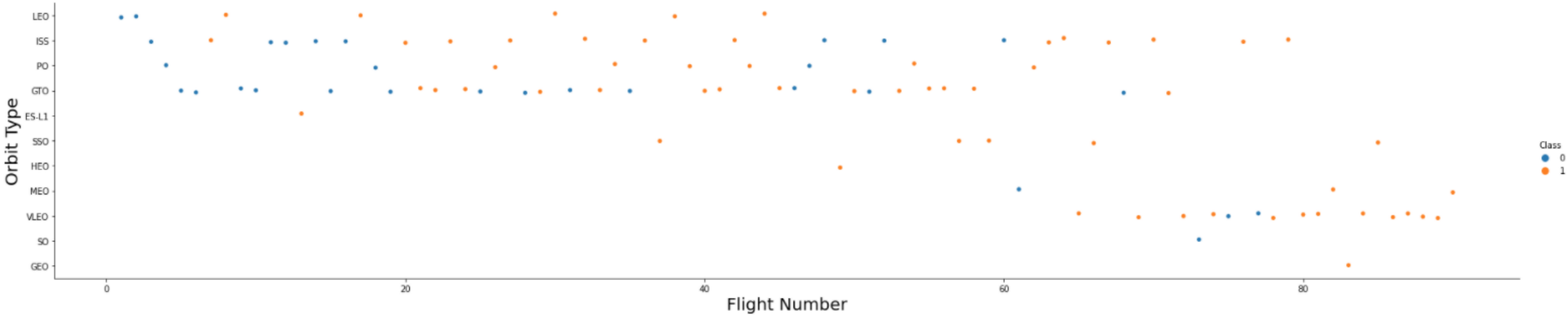

- The figure above shows how the payload mass has changed with the launch site used
- The distribution of data shows that the rockets with the highest payload masses were launched from the 'CCAFS' and 'KSC' sites
- Rockets with payload masses of ~9000 kg were primarily launched from the 'VAFB' site
- The 'CCAFS' site has been used to launch the largest number of rockets

# Success rate vs. Orbit type



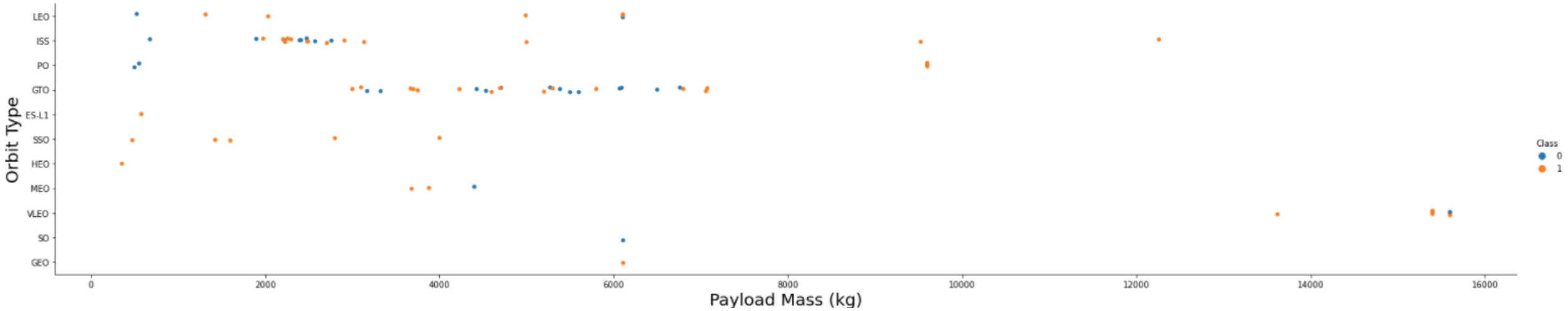

- The figure shows how the launch success rate varies with the type of orbital trajectory the rocket follows
- The data show that all attempts at 'ES-L1', 'SSO', 'GEO', and 'HEO' orbit have been successful
- The 'MEO' orbital trajectory exhibits the highest variation in launch success

# Flight Number vs. Orbit type



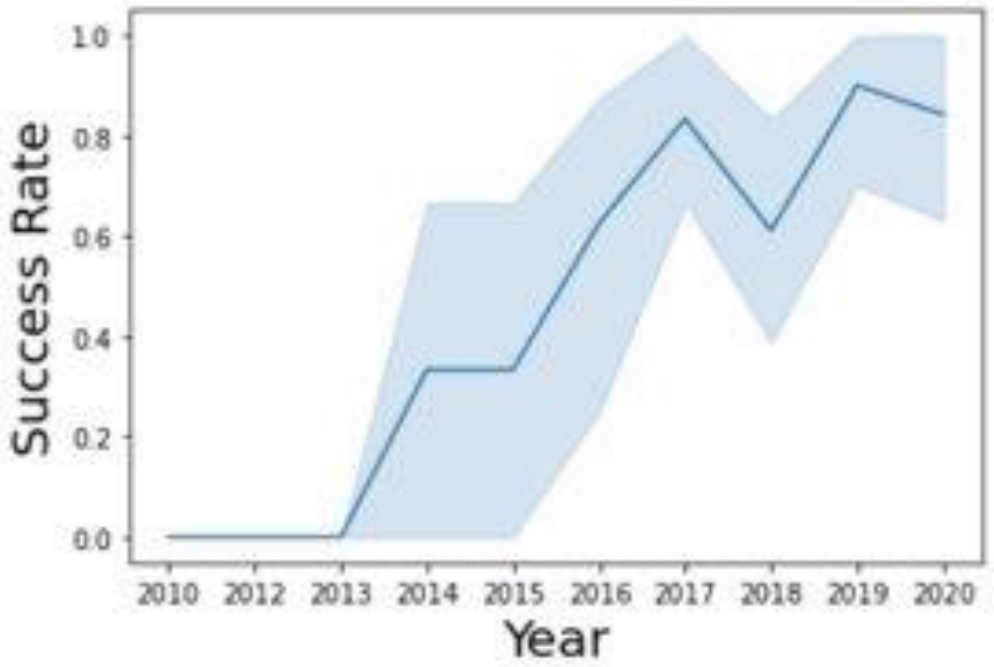

- The figure above shows how the orbital trajectory of the rocket has changed with the flight number (and time)
- The majority of recent launches (higher flight numbers) have followed the 'VLEO' launch trajectory
- The majority of early launches followed the 'LEO', 'ISS', and 'GEO' launch trajectories

# Payload vs. Orbit type




- The figure above shows how the orbital trajectory of the rocket is related to the payload mass
- The highest payload masses were attempted with a 'VLEO' orbital trajectory
- A high number of broadly-distributed payload masses has been attempted with the 'GEO' orbital trajectory

# Launch success yearly trend




- The figure shows how the launch success rate varies with time
- The trendline clearly shows that the launch success rate has improved over time
- In the year 2020 the launch success rate was ~84%

# EDA with SQL

# All launch site names

| launch_site |
| --- |
| CCAFS LC-40 |
| CCAFS SLC-40 |
| KSC LC-39A |
| VAFB SLC-4E |

CCAFS = Cape Canaveral

LC = Launch Complex

SLC = Space Launch Complex

KSC = Kennedy Space Center

VAFB = Vanderberg Airforce Base

# Launch site names begin with `CCA`

| column_0 | DATE | time__utc_ | booster_version | launch_site | payload |
|---|---|---|---|---|---|
| 0 | 2010-04-06 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit |
| 1 | 2010-08-12 | 15:43:00 | F9 v1.0 B0004 | CCAFS LC-40 | Dragon demo flight C1, two CubeSats, barrel of Brouere cheese |
| 2 | 2012-05-22 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 |
| 3 | 2012-08-10 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 |
| 4 | 2013-01-03 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 |

- Shown at left are some of the launch data for launch sites beginning with the name 'CCA'

# Total payload mass

```
%sql select SUM(payload_mass__kg_)
```

```
 * ibm_db_sa://wfy19102:***@ba99a9e
Done.
```

| 1 |
| --- |
| 107010 |

The total payload mass for all launches performed by NASA was obtained with an SQL query

The total payload mass for these launches was:

107,010 kg

# Average payload mass by F9 v1.1

```
%sql select AVG(payload_mass__kg_)
```

```
 * ibm_db_sa://wfy19102:***@ba99a9e
Done.
```

| 1 |
| --- |
| 2928 |

- The average payload mass carried by a booster version F9 v1.1 rocket was obtained with an SQL query
- The average payload mass was:

2,928 kg

# First successful ground landing date

```
%sql select MIN(DATE)
```

```
 * ibm_db_sa://wfy1916
Done.
```

| 1 |
| --- |
| 2015-12-22 |

- The first successful ground landing was obtained with an SQL query
- The first successful ground landing took place on:

December 22, 2015

# Successful drone ship landing with payload between 4000 and 6000

```
%sql select booster_version
between 4000 and 6000;
```

```
 * ibm_db_sa://wfy19102:***@
Done.
```

| booster_version |
|---|
| F9 FT B1020 |
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |

- The names of boosters which have success in drone ship, and have payload mass greater than 4000 but less than 6000, were obtained with an SQL query
- The booster names are shown at left

# Total number of successful and failure mission outcomes

```
%sql select COUNT(mission_outcome) from SPACEXDATASET where mission_outcome like 'Success';
```

```
 * ibm_db_sa://wfy19102:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.dat
Done.
```

]:

| 1 |
|---|
| 99 |

```
%sql select COUNT(mission_outcome) from SPACEXDATASET where mission_outcome like '%Fail%';
```

```
 * ibm_db_sa://wfy19102:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.dat
Done.
```

]:

| 1 |
|---|
| 1 |

- The total number of successful (99) and failed (1) missions were obtained with an SQL query (above)

# Boosters that carried maximum payload

```
%sql select booster_version from SPACEXDATASET
```

```
 * ibm_db_sa://wfy19102:***@ba99a9e6-d59e-48
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 |

- The names of the boosters that carried the maximum payload were obtained with an SQL query
- They are shown at left

# 2015 launch records

```
%sql select booster_version, launch_site
```

```
 * ibm_db_sa://wfy19102:***@ba99a9e6-d
Done.
```

]:

| booster_version | launch_site |
|---|---|
| F9 v1.1 B1012 | CCAFS LC-40 |
| F9 v1.1 B1015 | CCAFS LC-40 |

- The booster versions and launch site for records with a failed landing outcome in drone ship during the year 2015 were obtained with an SQL query

# Rank success count between 2010-06-04 and 2017-03-20

```
ect landing__outcome, count(landing__outcome)
```

```
 * ibm_db_sa://wfy19102:***@ba99a9e6-d59e-
Done.
```

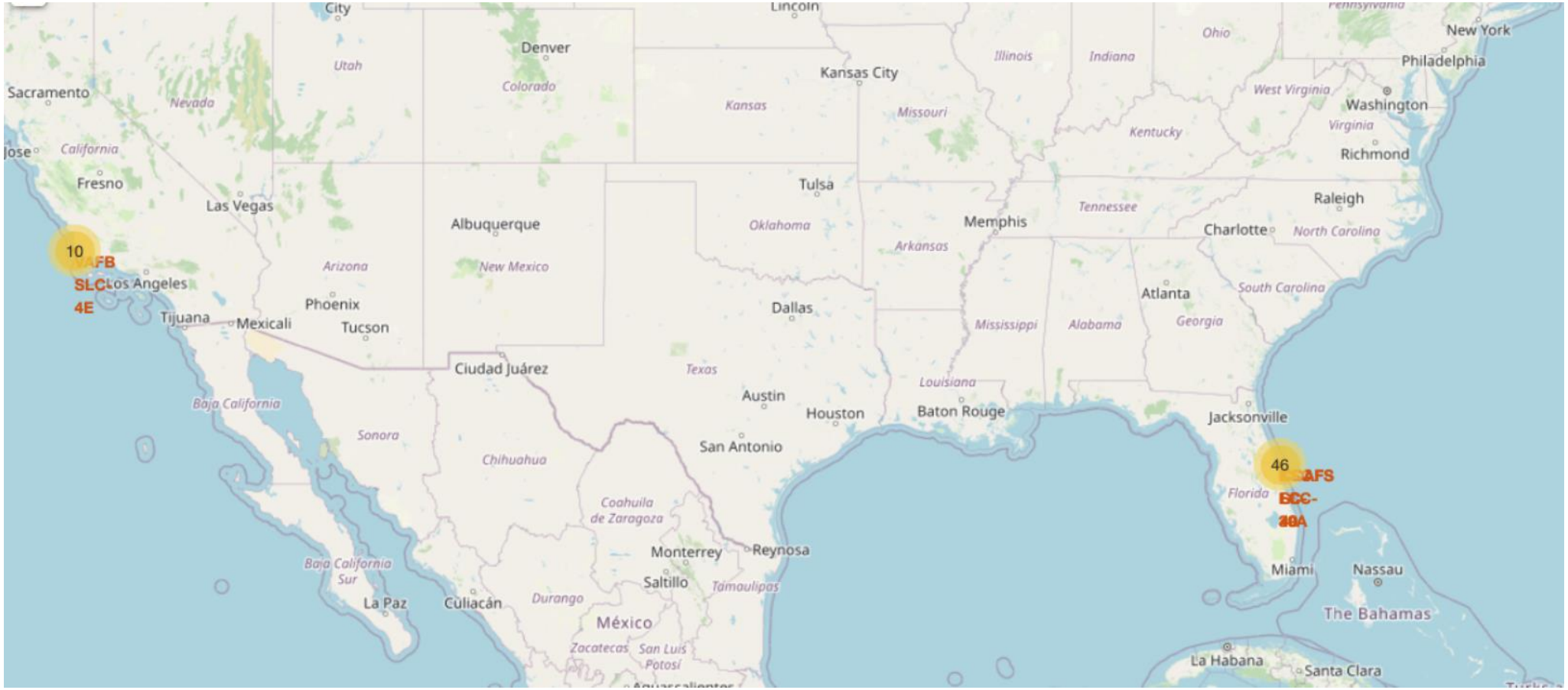
]:

| landing__outcome | 2 |
|---|---|
| No attempt | 10 |
| Failure (drone ship) | 5 |
| Success (drone ship) | 5 |
| Success (ground pad) | 5 |
| Controlled (ocean) | 3 |
| Uncontrolled (ocean) | 2 |
| Failure (parachute) | 1 |
| Precluded (drone ship) | 1 |

- The count of successful landing outcomes between 2010-06-04 and 2017-03-20 were listed in descending order using an SQL query

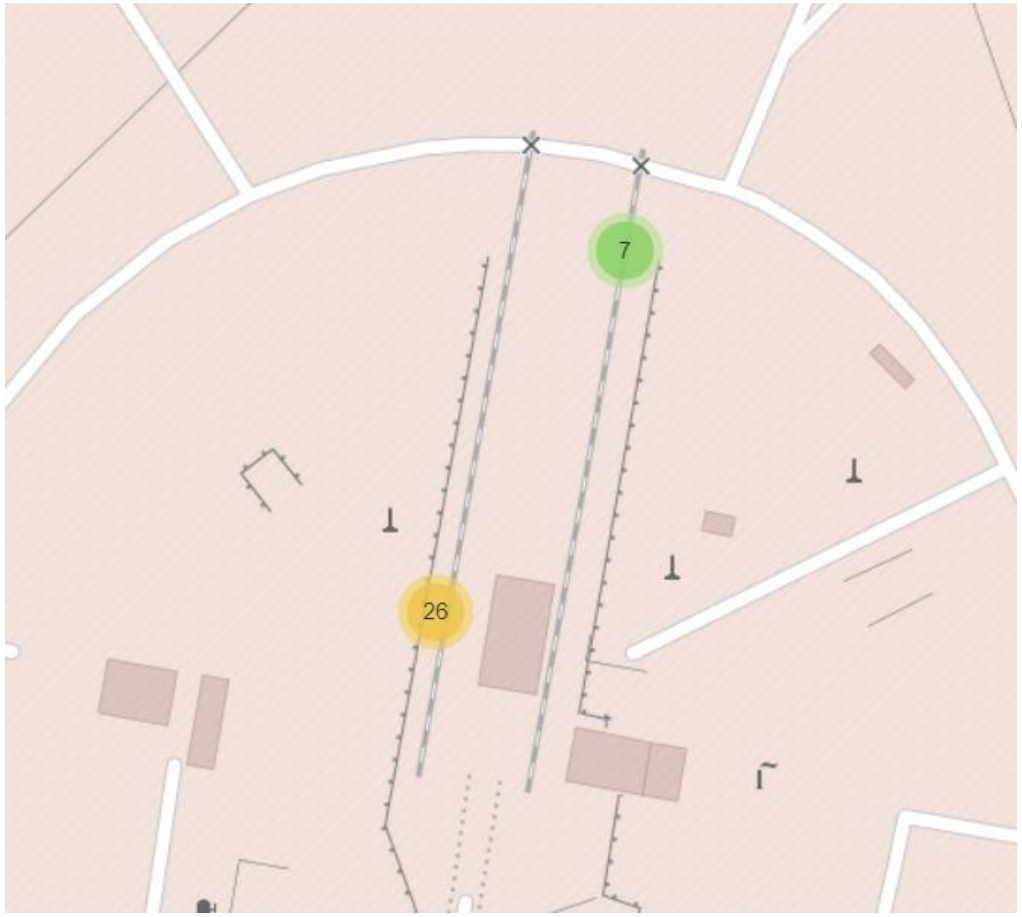
# Interactive map with Folium

# Folium map of launch site locations



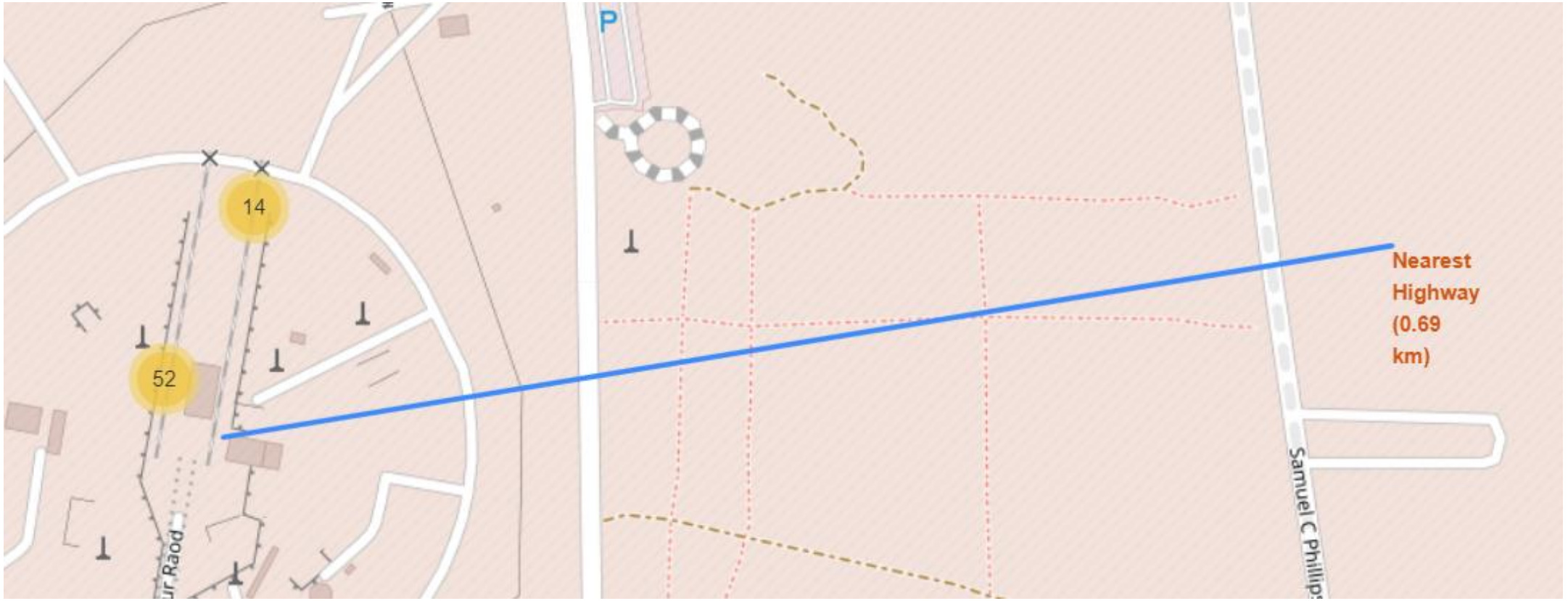

- All launch site locations are shown in the map above

# Folium map with color-labeled launch results




- The Cape Canaveral (both launch site) launch results are shown as a Folium "Marker cluster"
- The color of the markers is determined by the success of the launch result

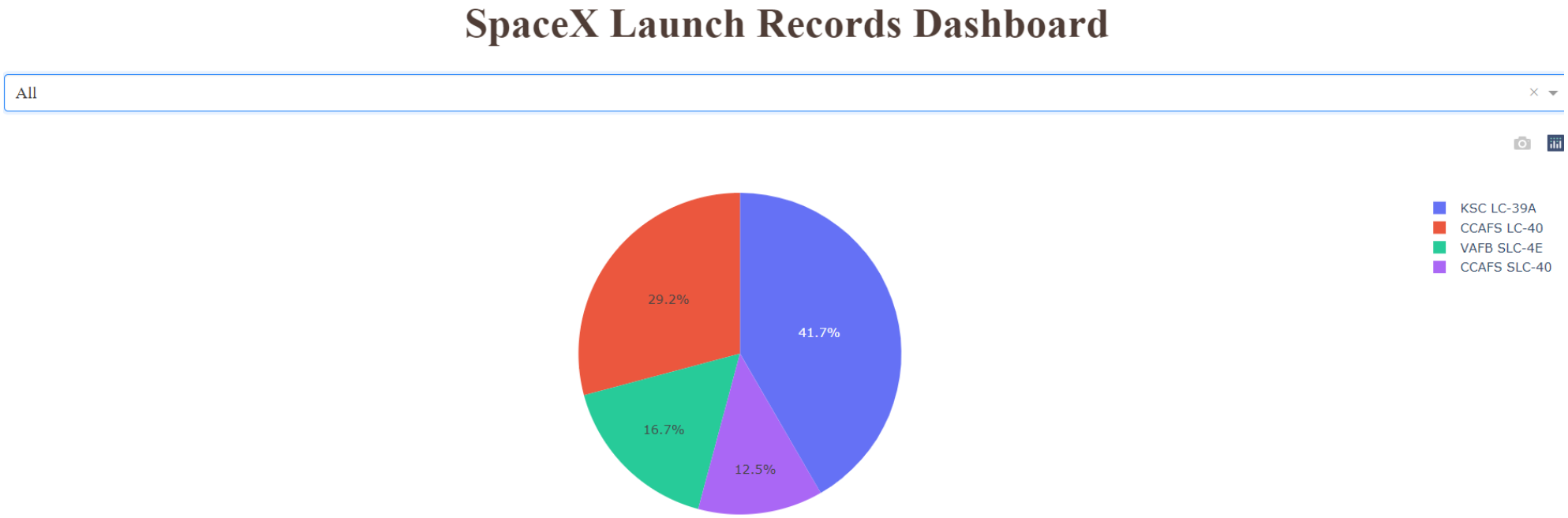
# Folium map with a nearby highway




- Shown is a Folium map with a line between the launch sites and the nearest highway (0.69 km)

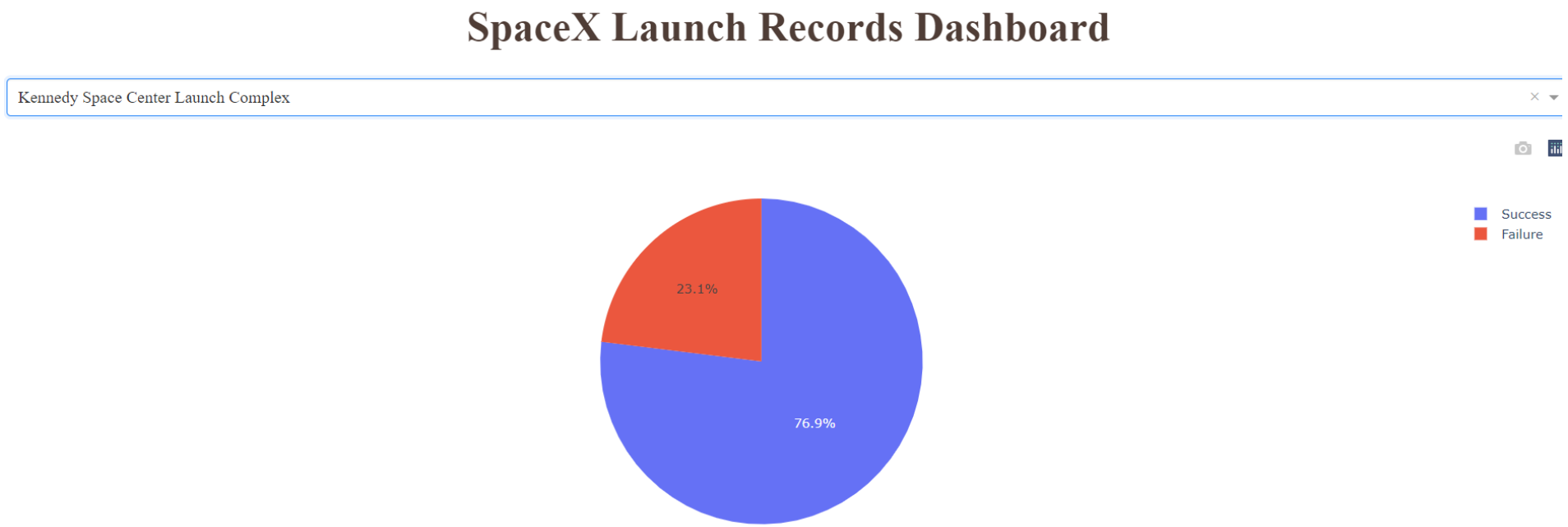
# Build a Dashboard with Plotly Dash

# Launch Success Rate as Pie Chart



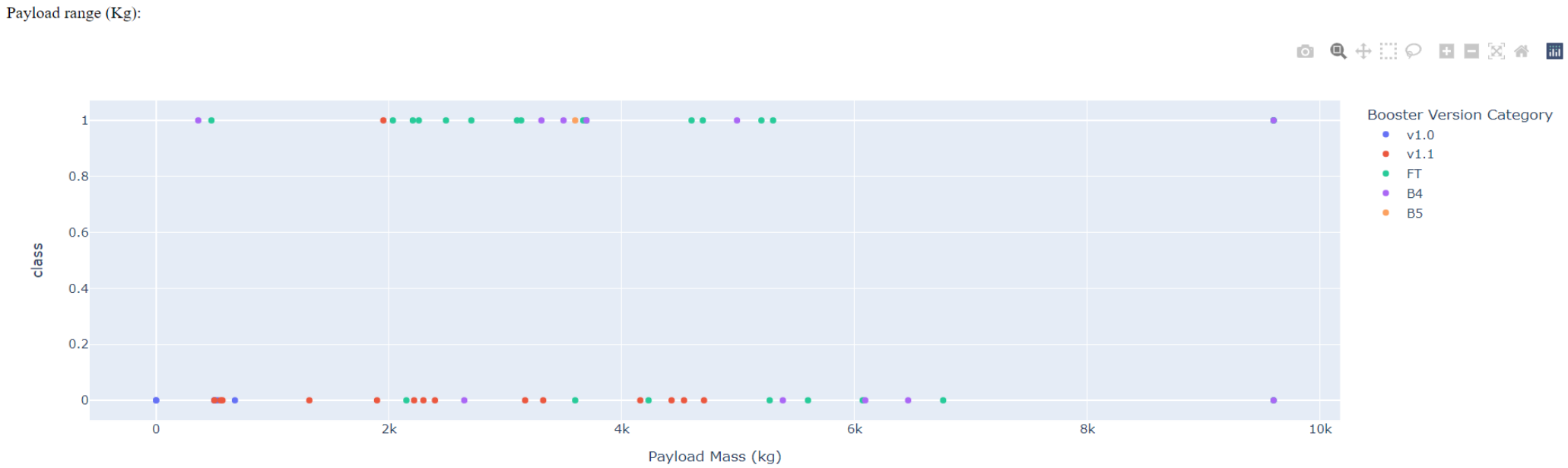

- A dashboard was created to display the number of successful launches at each site in a pie chart

# Kennedy Space Center Launch Success Rate




- A pie chart was also generated for the Kennedy Space Center site, individually, since it had the highest number of successful launches

# Payload vs. Launch Outcome Scatter Plot



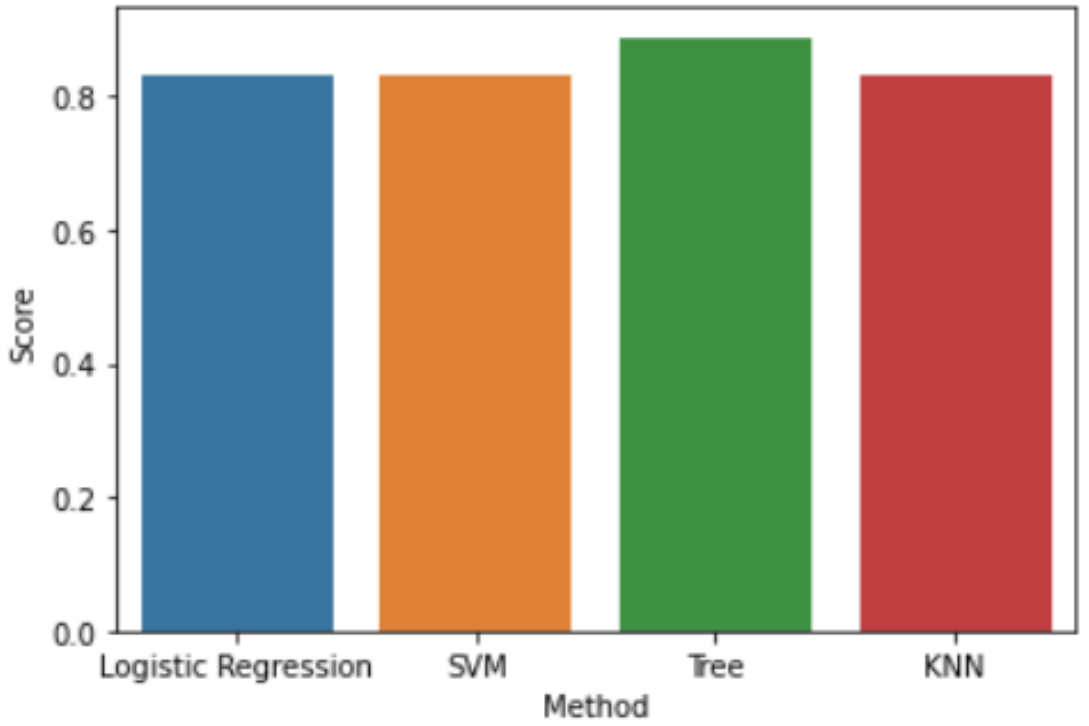

- The launch outcome was plotted against the payload mass, with data color-labeled according to Booster version

# Predictive analysis (Classification)

# Classification Accuracy

The bar chart at right shows the classification accuracy (score) obtained for all models applied to the Space-X dataset.

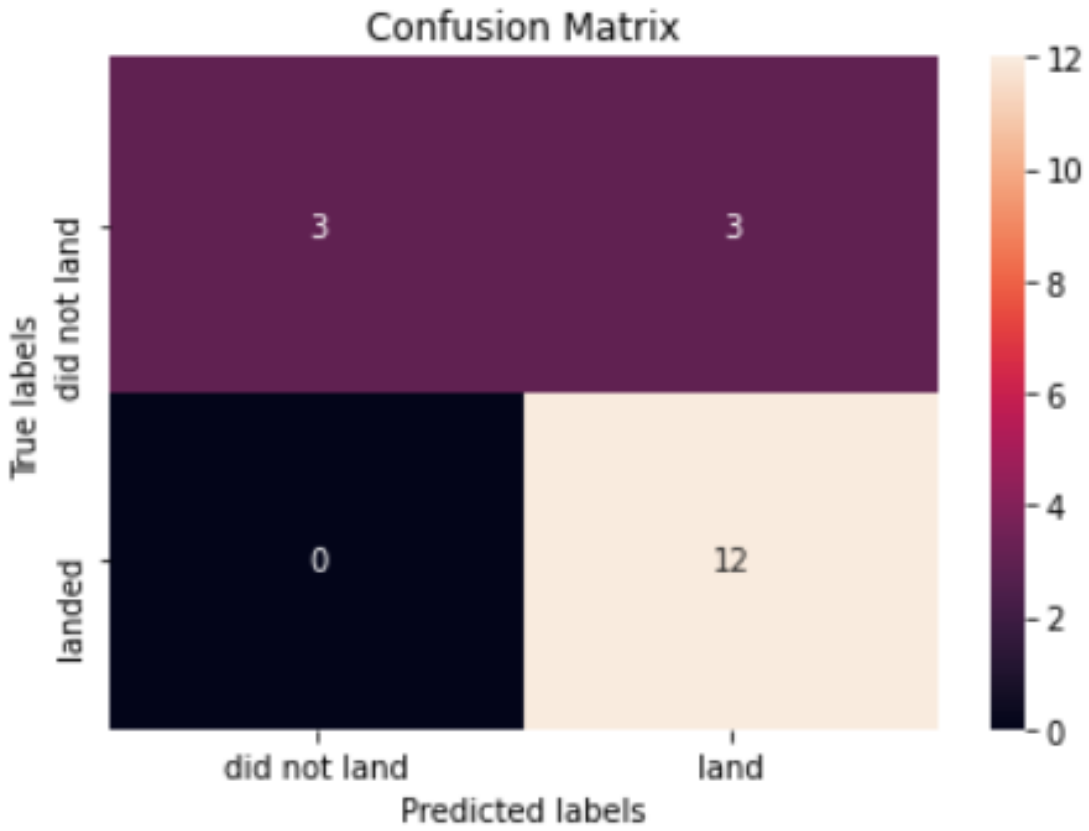
The chart shows that a decision tree classifier had the highest classification accuracy when determining if a launch would be successful.



# Confusion Matrix

Shown at right is the confusion matrix for the Decision Tree classifier.

The confusion matrix shows that there are a significant number of false positives (3), where the predicted outcome of a launch is that it would be successful when in actuality it wouldn't land.



# CONCLUSION

- Webscraping can be used to obtain real-time data that provides insight about companies such as Space-X
- A dashboard can be used to visualize data given user-defined input parameters
- When analyzing Space-X launch data using a variety of classification models, a decision tree had the highest accuracy for predicting the success of a launch (83%)