IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Malcomb C. Brown
Jan. 8, 2022

Outline

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

Executive Summary

- Ability to develop a cost analysis for our rockets based on expected successful landing
- Collecting, wrangling, and analyzing SpaceX Falcon 9 launch data
- Able to predict successful landings with over 80% accuracy

Introduction

- Reducing the cost of rockets
  - Being able to reuse the rockets
- What are the factors contributing to a successful landing?
  - Launch Site
  - Payload Mass
  - Orbit
- Able to correctly predict successful landings with over 80% accuracy

Section 1

Methodology

Methodology

Executive Summary

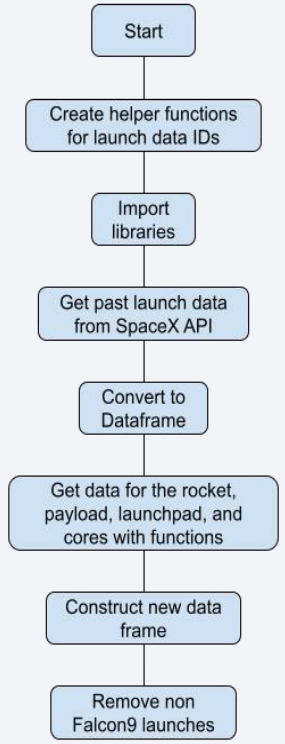
- Data collection methodology:
  - Gathering past launch data using the SpaceX REST API
  - Web Scraping relevant Wiki pages with the Python library Beautiful Soup
- Perform data wrangling
  - Deal with any missing data and null values
  - Convert landing outcomes to allow for classification models
- 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

- How data sets were collected.
  - Acquiring past launch data using the SpaceX REST API
  - Use of the requests library to obtain the information from the API as a JSON file
  - Convert the JSON data into a data frame with the Pandas library
  - Web Scraping Falcon 9 launch data from Wiki pages using the Python Beautiful Soup library
  - Parse and Convert the HTML tables to a Pandas data frame

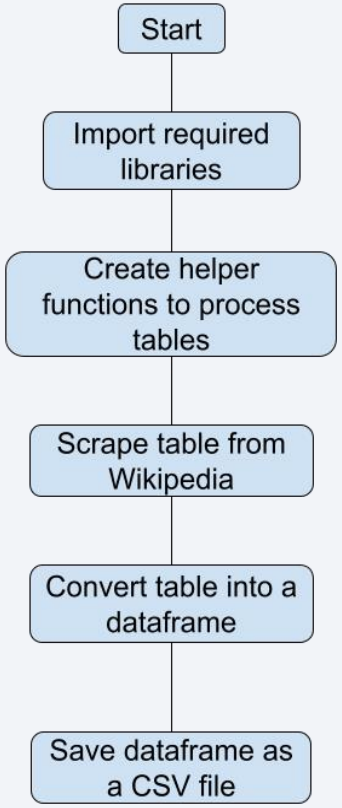
Data Collection – SpaceX API

- Request and Parse launch data with a GET request
- Retrieve specific data on the rocket, payload, launch pad, and cores
- Filter the resulting data frame to only include Falcon9 launches
- https://github.com/Ash021980/spaceyibmcap/blob/main/Data%20Collection%20API%20Lab.ipynb



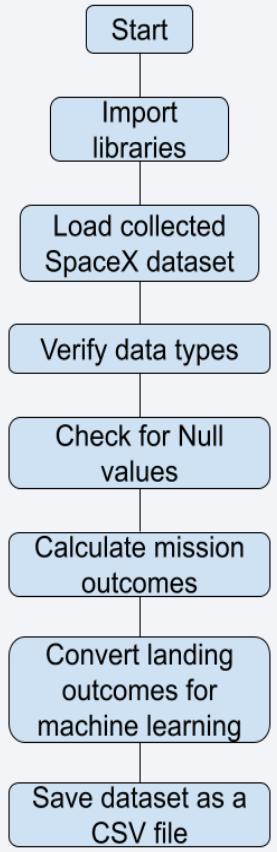
Data Collection - Scraping

- Extract Falcon9 launch records HTML table from Wikipedia with Beautiful Soup
- Parse the table and convert it to a Pandas data frame
- https://github.com/Ash021980/spaceyibmcap/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Performed EDA on collected SpaceX data
- Verify and clean data set
- Handle Null values in data
- Handle any data type inconsistencies
- Determine relevant training labels
- Transformed landing outcomes to success, 1, and failure, 2, class
- https://github.com/Ash021980/spaceyibmcap/blob/main/Data%20Wrangling%20Lab.ipynb



EDA with Data Visualization

- Categorical plots were used to see the effects different categories have on the success rate of the landing outcome.
- A bar chart was used to view the relationship between the landing outcome success rate and the type of orbit.
- A line plot was used to visualize the success rate trend over time.
- https://github.com/Ash021980/spaceyibmcap/blob/main/SpaceY%20Visualizations%20.ipynb

EDA with SQL

- What unique launch sites have been used?
- What the average pay load mass carried by a Falcon9 v1.1 booster?
- When was the first successful landing outcome achieved on a ground pad?
- What is the total number of successful and failed missions?
- What was the total number of landing outcomes from 2010-06-04 and 2017-03-20 and rank them?
- https://github.com/Ash021980/spaceyibmcap/blob/main/EDA%20with%20SQL%20.ipynb

Build an Interactive Map with Folium

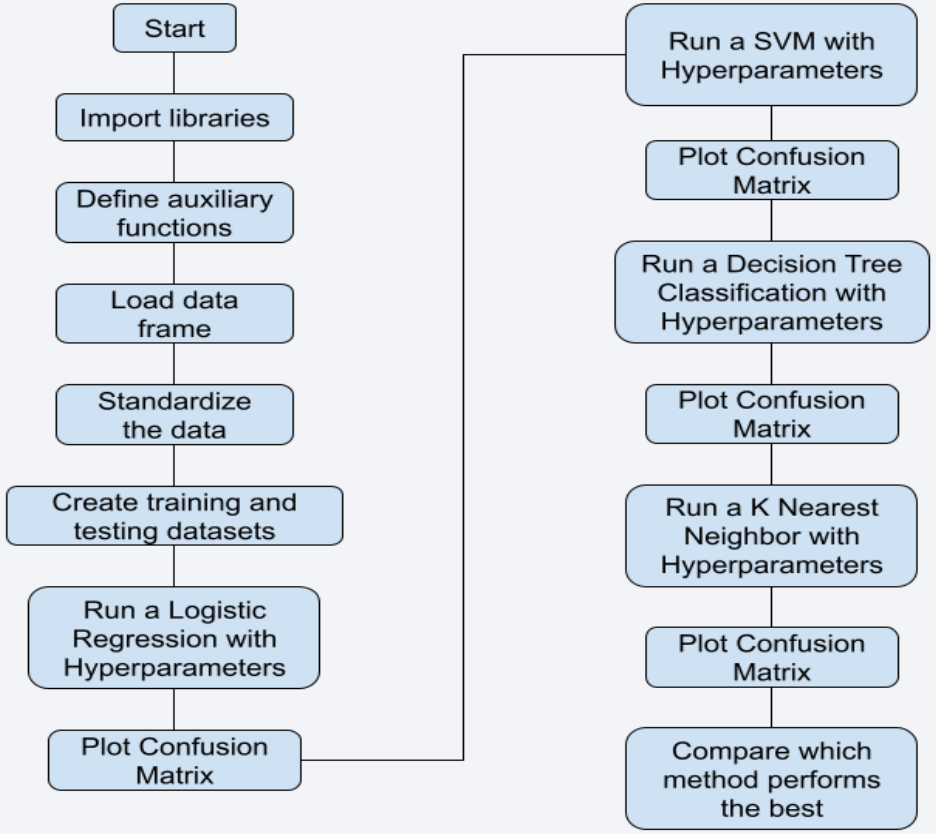
- Initialize a Folium Map object to get an interactive map
- Marker objects identify launch site locations.
- Marker Clusters visualize successful and failed launch outcomes at each launch site.
- Circles allowed for the highlighting specific coordinates and overlaying text labels.
- Mouse position provided for the viewing latitude and longitude for a mouse over on the map.
- Polylines provided lines from launch sites and specified map coordinates.
- https://github.com/Ash021980/spaceyibmcap/blob/main/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Display the success rates for each launch site in a Pie chart.
  - Controlled with a dropdown menu to drilldown by specific site
- Categorical plot to visualize the success rate for each booster version and by its pay load mass.
  - Pay load mass can be adjusted by a slider
- https://github.com/Ash021980/spaceyibmcap/blob/main/spacex_dash_app.py.ipynb

Predictive Analysis (Classification)

- Feature Engineering
  - Create dummy variables for categorical columns
  - Cast all numeric columns as float64
- Perform exploratory data analysis and determine training labels
  - Create a column for the class
  - Standardize the data
  - Split the data into training and testing datasets
- Find the best hyperparameters for SVM, Classification Trees, and Logistic Regression
  - Find the method that performs best using test data
- https://github.com/Ash021980/spaceyibmcap/blob/main/Machine%20Learning%20Prediction%20Lab.ipynb

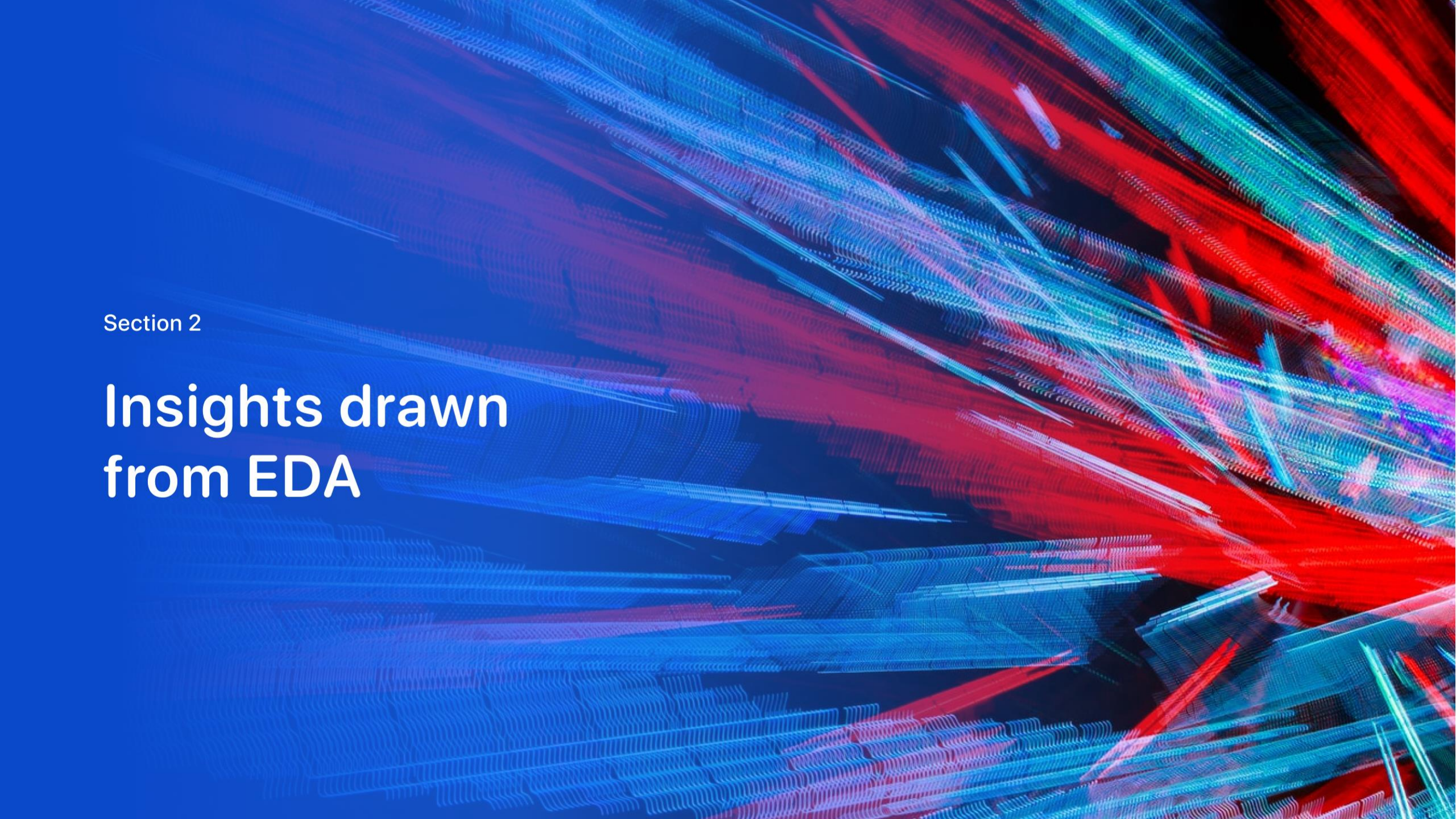


Results

- Exploratory data analysis results
  - Located the 4 launch sites used
  - Total number of successful and failed mission outcomes
  - Total for each type of landing outcome
- Interactive analytics
  - Launch site proximity to railways, highways, and coastline
  - Distance from launch site to nearest city
- Predictive analysis results
  - No significant difference in the accuracy among the models
  - Decision Tree Classification showed the only improvement over the training and testing sets

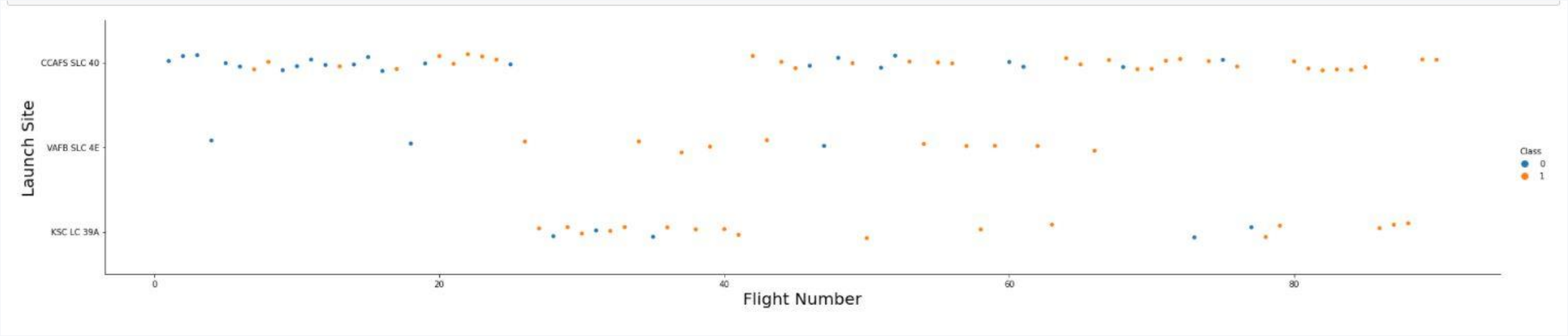
Section 2

Insights drawn from EDA



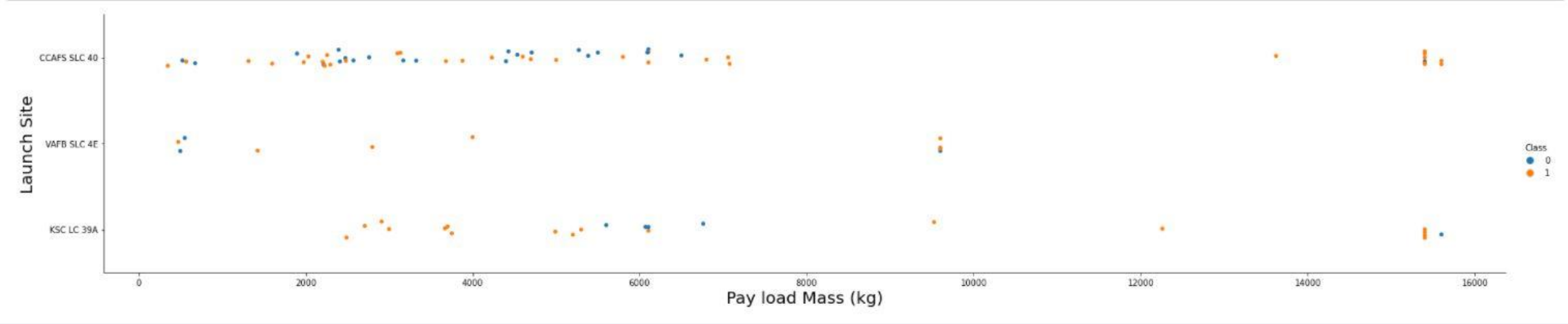
Flight Number vs. Launch Site

- Visualize the landing outcome for each flight.
- Became more successful as more flights were done. (Experience)
- Most flights originate from CCAFS SLC-40.
- KSC LC-39A and VAFB SLC-4E have high success rates but low usage.



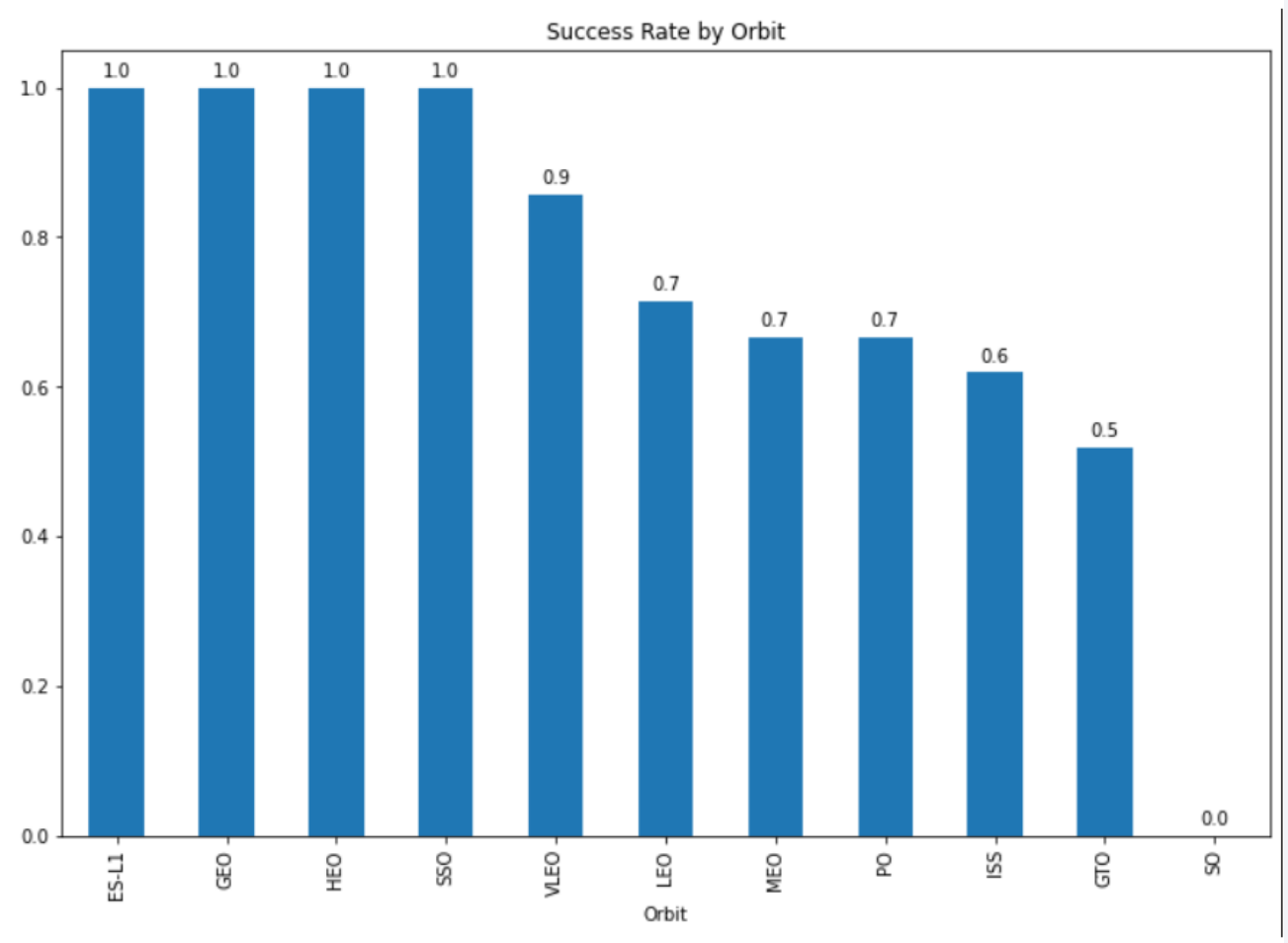
Payload vs. Launch Site

- Plot the landing outcome by launch site and pay load mass (kg).
- Success rate increases for payloads over 10,000 kg.
- VAFB SLC-4E has no launches with a payload over 10,000 kg.
- KSC LC-39A has only successful landing outcomes with a payload mass less than 5,000 kg.



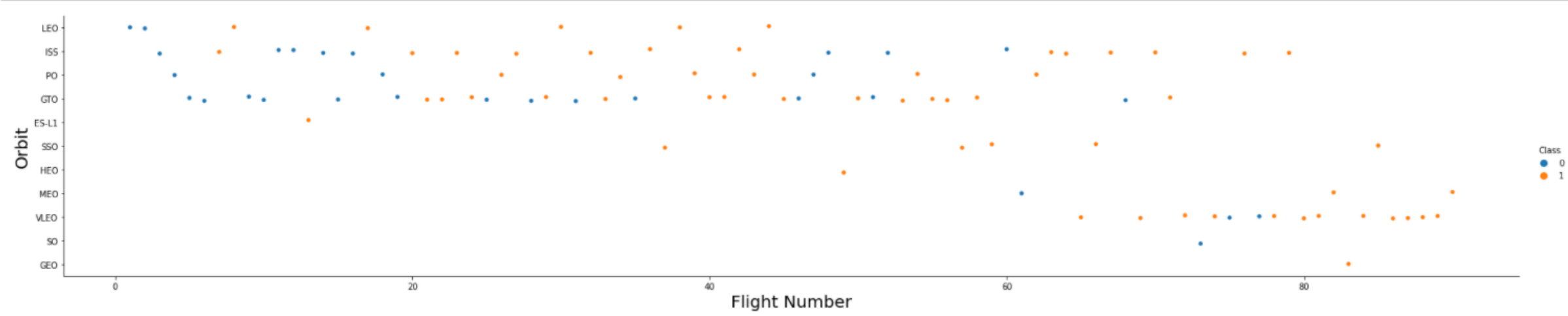
Success Rate vs. Orbit Type

- Orbits with highest success rate:
  - ES-L1
  - GEO
  - HEO
  - SSO
  - VLEO



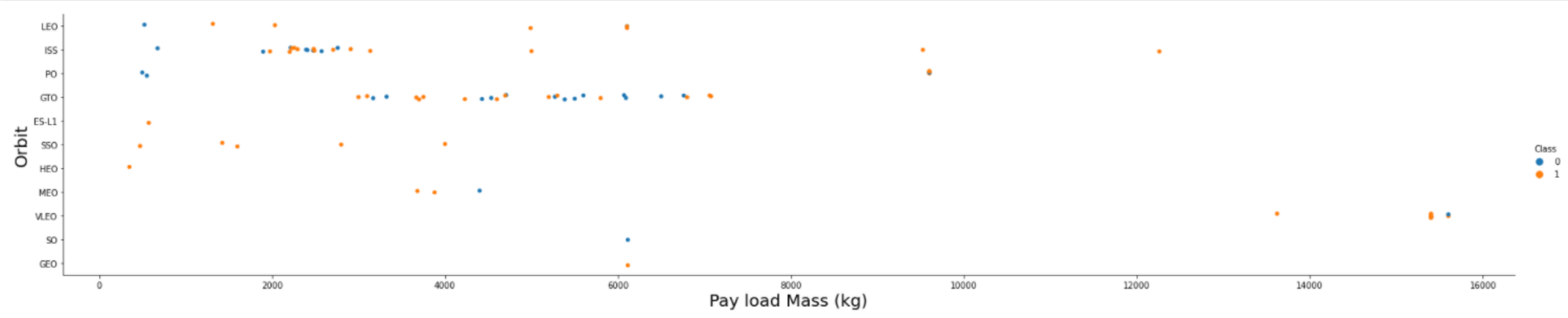
Flight Number vs. Orbit Type

- Orbits with 100% success rate have only completed 5 or fewer flights.
  - SSO is the only one with more than 1 flight
- VLEO, with a 90% success rate, is the most frequented orbit after the 60th flight.
- Success rate increases with the number of flights to a specific orbit.



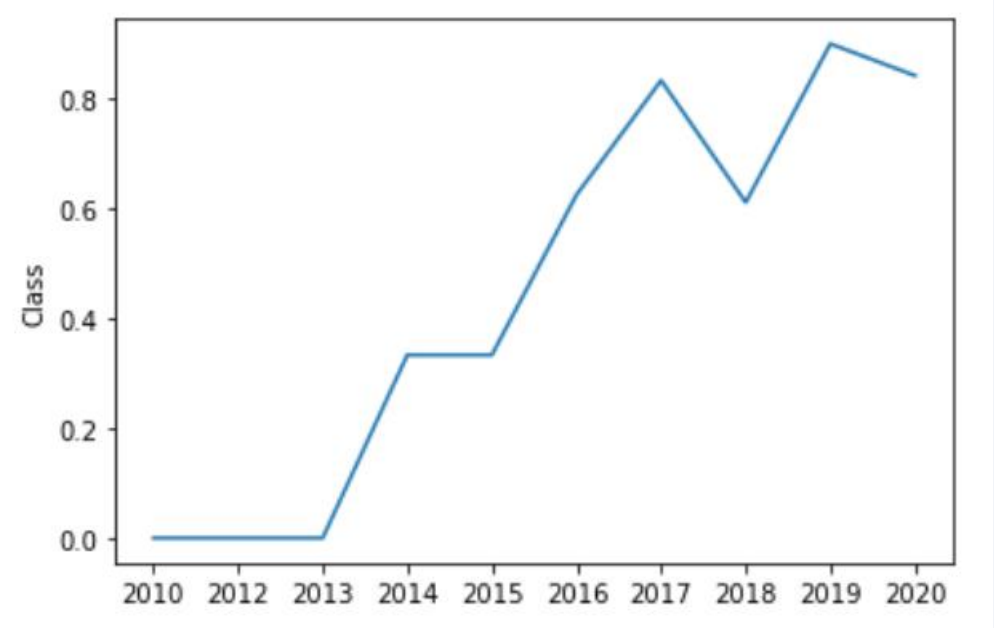
Payload vs. Orbit Type

- Heavy pay loads, over 10,000 kg, have positive success rates to the Polar, LEO, and ISS orbits.
- Pay loads under 5,000 kg have positive success rates to the SSO, ES-L1, and HEO orbits.
- Cannot distinguish between positive or negative landing for the GTO orbit.



Launch Success Yearly Trend

- Since 2013, there has been a steady increase in the success rate.
- From 2017 to 2018, there was a significant reduction in the success rate.



All Launch Site Names

- Explored the SpaceX historical launch data
- Found 4 launch facilities used for launch missions
  - **VAFB SLC 4E** , Vandenberg Air Force Base Space Launch Complex 4E
  - **KSC LC 39A**, Kennedy Space Center Launch Complex 39A
  - **CCAFS LC-40**, Cape Canaveral Space Launch Complex 40
  - **CCAFS SLC-40**, Cape Canaveral Space Launch Complex 40
- SQL query in Appendix

Launch Site Names Begin with 'CCA'

- Looked at 5 missions to launch from Cape Canaveral and Merritt Island launch sites
- First 5 missions were to LEO
- 4 of the 5 had NASA as the customer
- All missions were successful
- No successful attempts to retrieve the first stage

```
%%sql
SELECT * FROM SPACEXDATASET
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5
```

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- Calculated the total payload carried by boosters from NASA Commercial Resupply Services
- Over 45,000 kg, @100,521 lbs., were carried

```
%%sql
SELECT sum(payload_mass__kg_) as total_payload FROM SPACEXDATASET
WHERE customer like 'NASA (CRS)'
```

total_payload
45596

Average Payload Mass by F9 v1.1

- Calculated the average payload mass carried by booster version F9 v1.1
- On average 2,534 kg, @5,586 lbs., were carried

```
%%sql
SELECT avg(payload_mass__kg_) as average_payload FROM SPACEXDATASET
WHERE booster_version like 'F9 v1.1%'
```

average_payload
2534

First Successful Ground Landing Date

- When was the first successful landing outcome on ground pad?
- December 22, 2015
- That's 5 years after the first launch

```
%%sql
SELECT min(DATE) as first_successful_landing FROM SPACEXDATASET
WHERE landing__outcome like '%Success%'
```

first_successful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Which boosters successfully landed on drone ship that had payload mass greater than 4000 but less than 6000 kg?
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

- SQL query and data frame in Appendix

Total Number of Successful and Failure Mission Outcomes

- Calculate the total of each mission outcome
- There was only 1 failure (in flight)
- There were 100 successful missions

mission_outcome	total
Success	99
Failure (in flight)	1
Success (payload status unclear)	1

```
%%sql
SELECT mission_outcome, count(mission_outcome) as total FROM SPACEXDATASET
GROUP BY mission_outcome
ORDER BY total desc
```

Boosters Carried Maximum Payload

- Which booster carried the maximum payload mass?
- The F9 B5 booster versions carried the max payload mass, which was 15,600 kg

```
%%sql
SELECT booster_version, payload_mass__kg_ FROM SPACEXDATASET
WHERE payload_mass__kg_ LIKE
    (SELECT max(payload_mass__kg_)
    FROM SPACEXDATASET)
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- There were 2 failed landing outcomes on drone ships
- Both were booster version F9 v1.1
- Launched from CCAFS LC-40

```
%%sql
SELECT landing__outcome, booster_version, launch_site, DATE FROM SPACEXDATASET
WHERE landing__outcome LIKE 'Failure (drone ship)'
AND DATE LIKE '2015%'
```

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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

- Rank of landing outcomes between 2010-06-04 and 2017-03-20
- No failed landing outcomes to a ground pad
- 33% of the time no attempt was made at booster recovery
- Landing to a drone ship was successful at rate of 50%
- Query and data frame presented in the Appendix

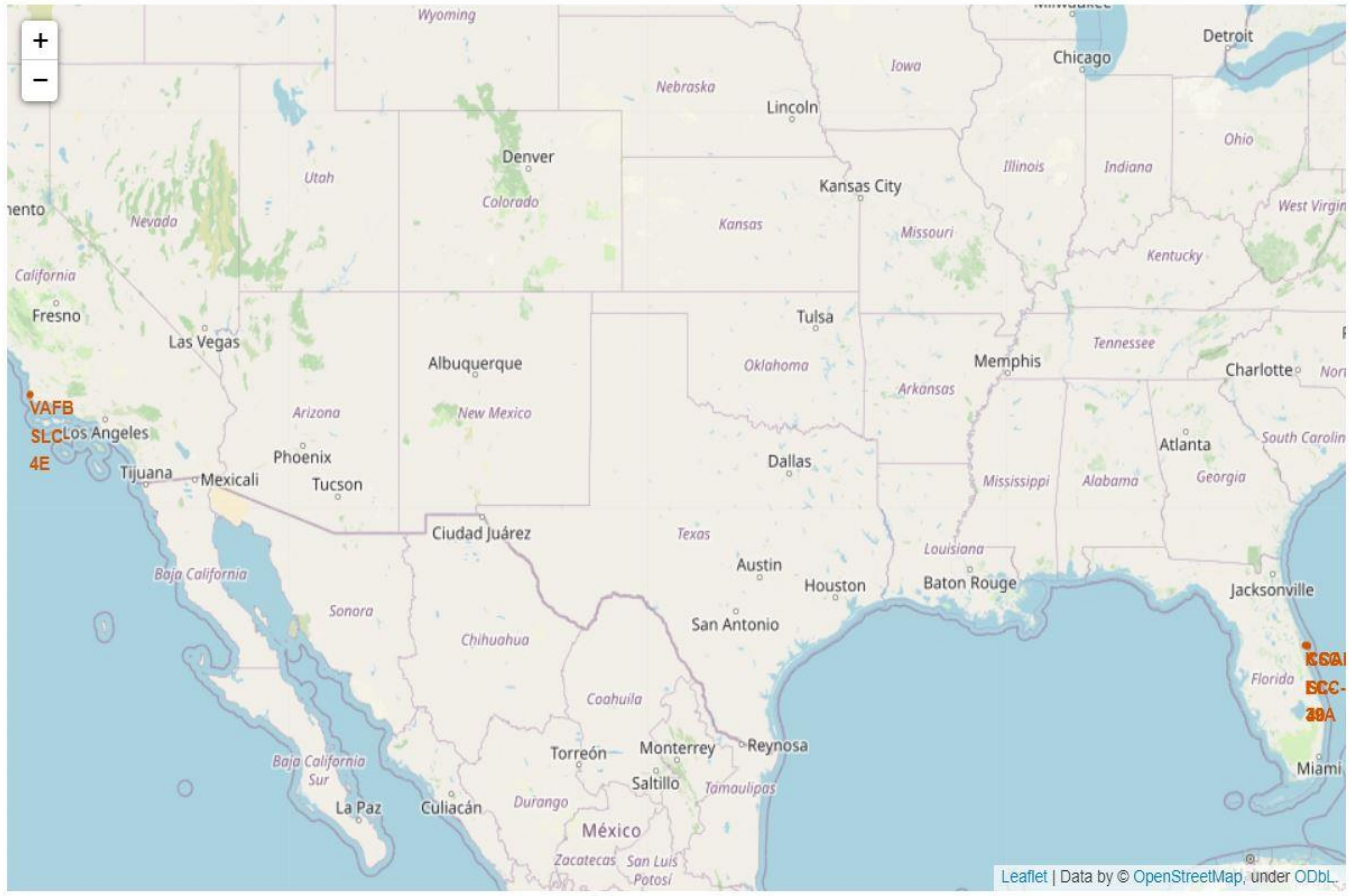
Section 4

Launch Sites Proximities Analysis

Launch Site Locations

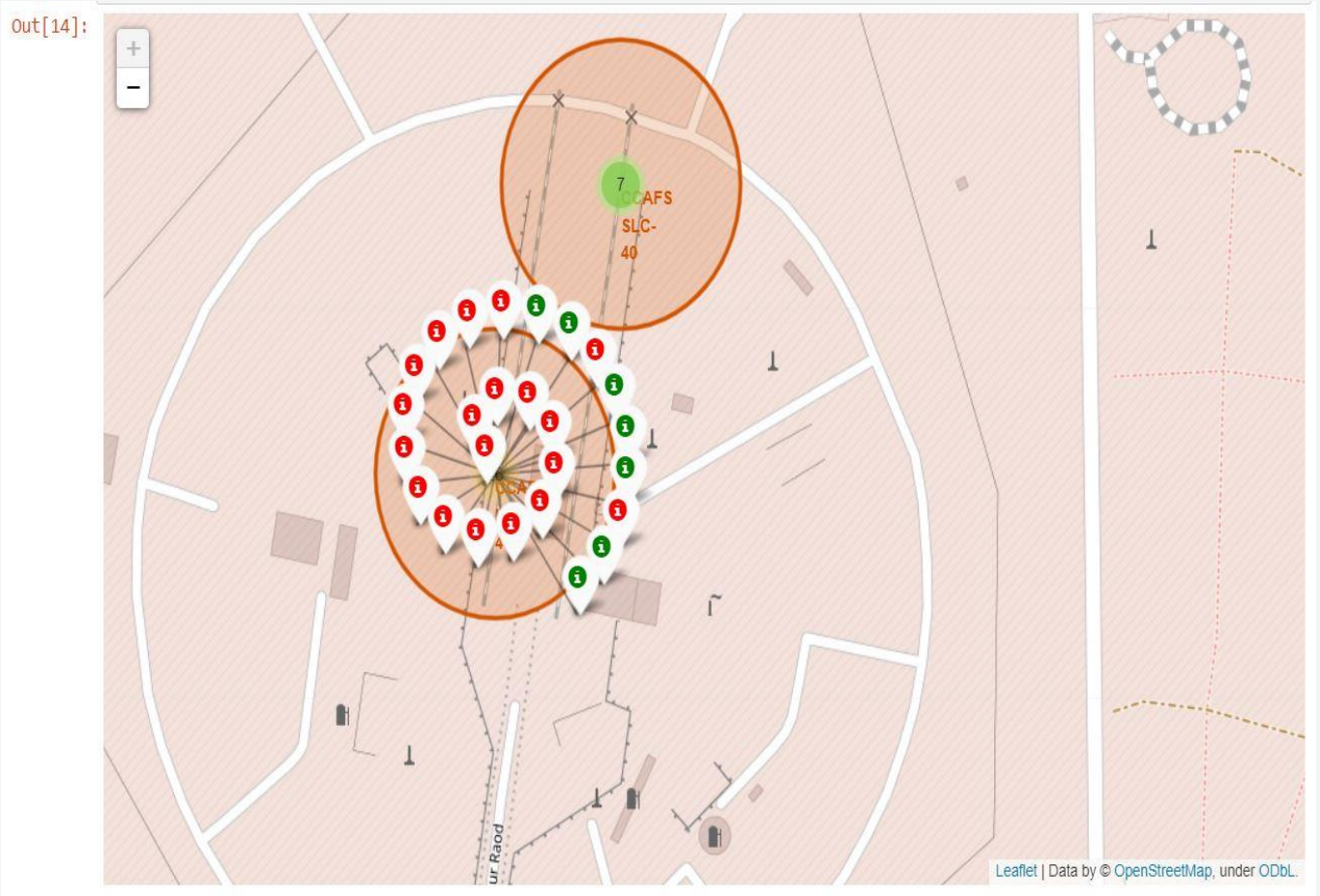
- Launch sites located near the coast.
- 75% of launch sites are in Florida

Out[9]:



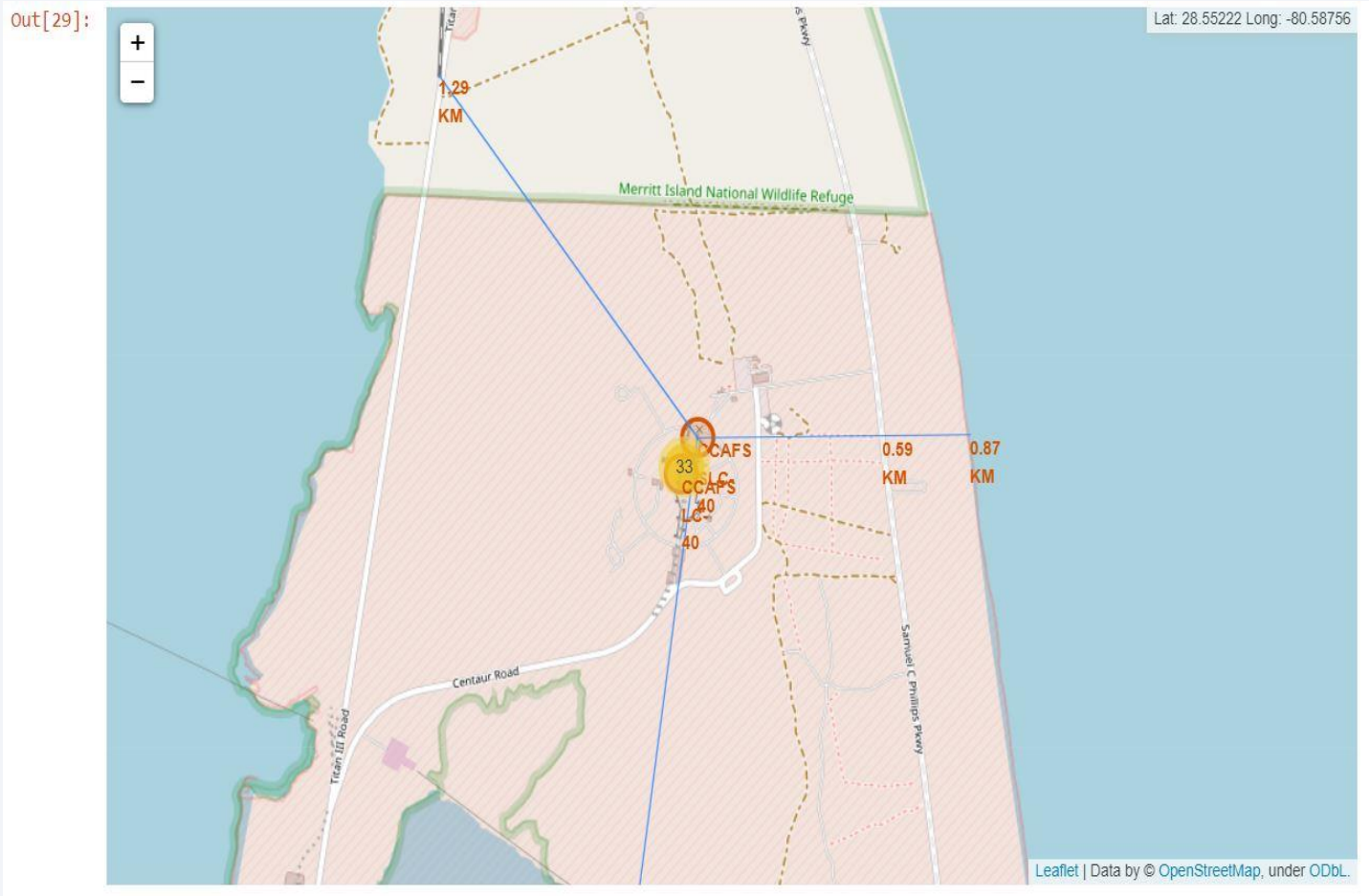
Launch Site Landing Outcomes

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot



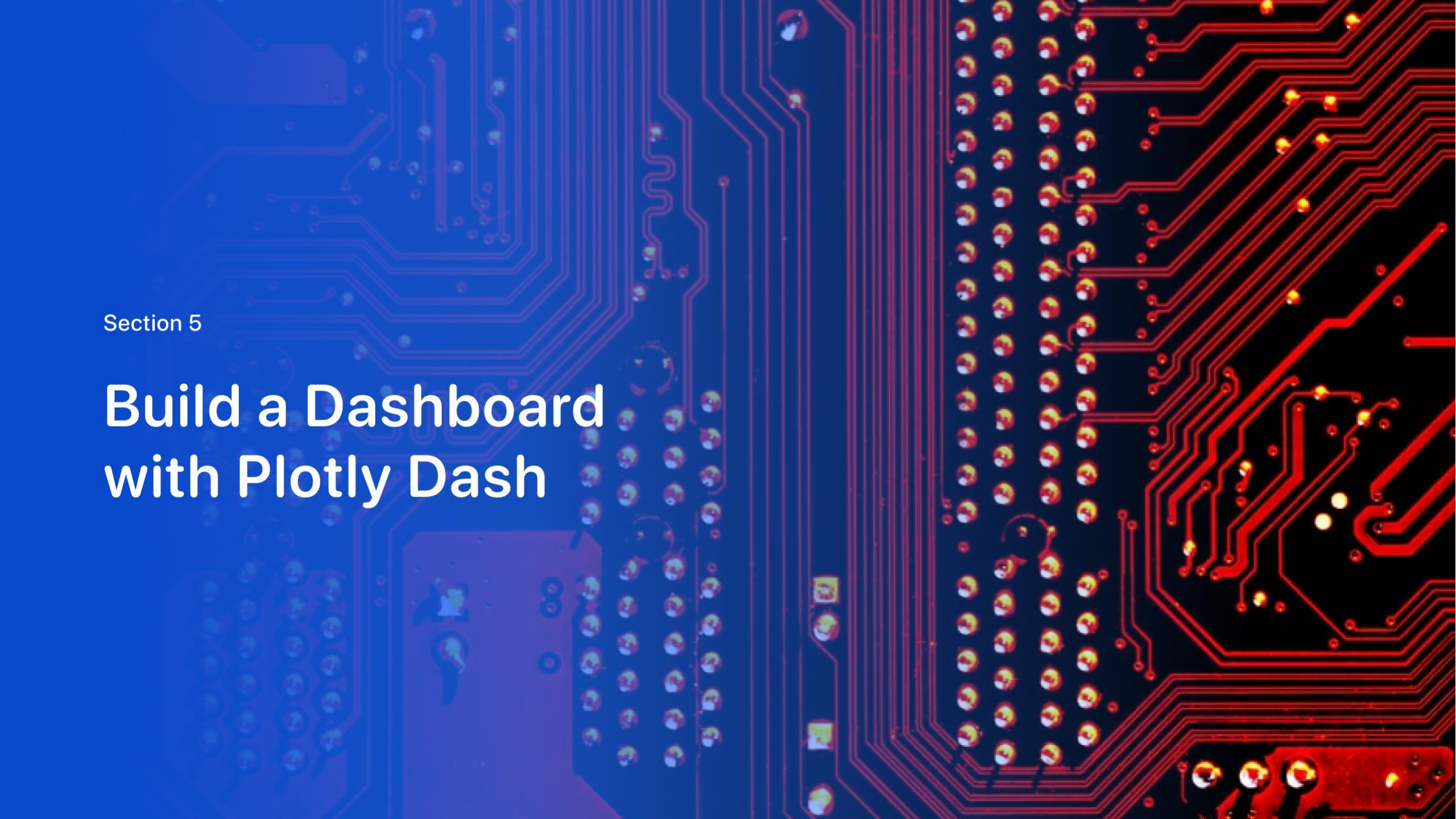
Distances between Launch Site and its Proximities

- Launch sites are nearby:
  - Railways and highways for logistics
  - Coastlines for safety
- Launch sites are not nearby:
  - Cities for safety



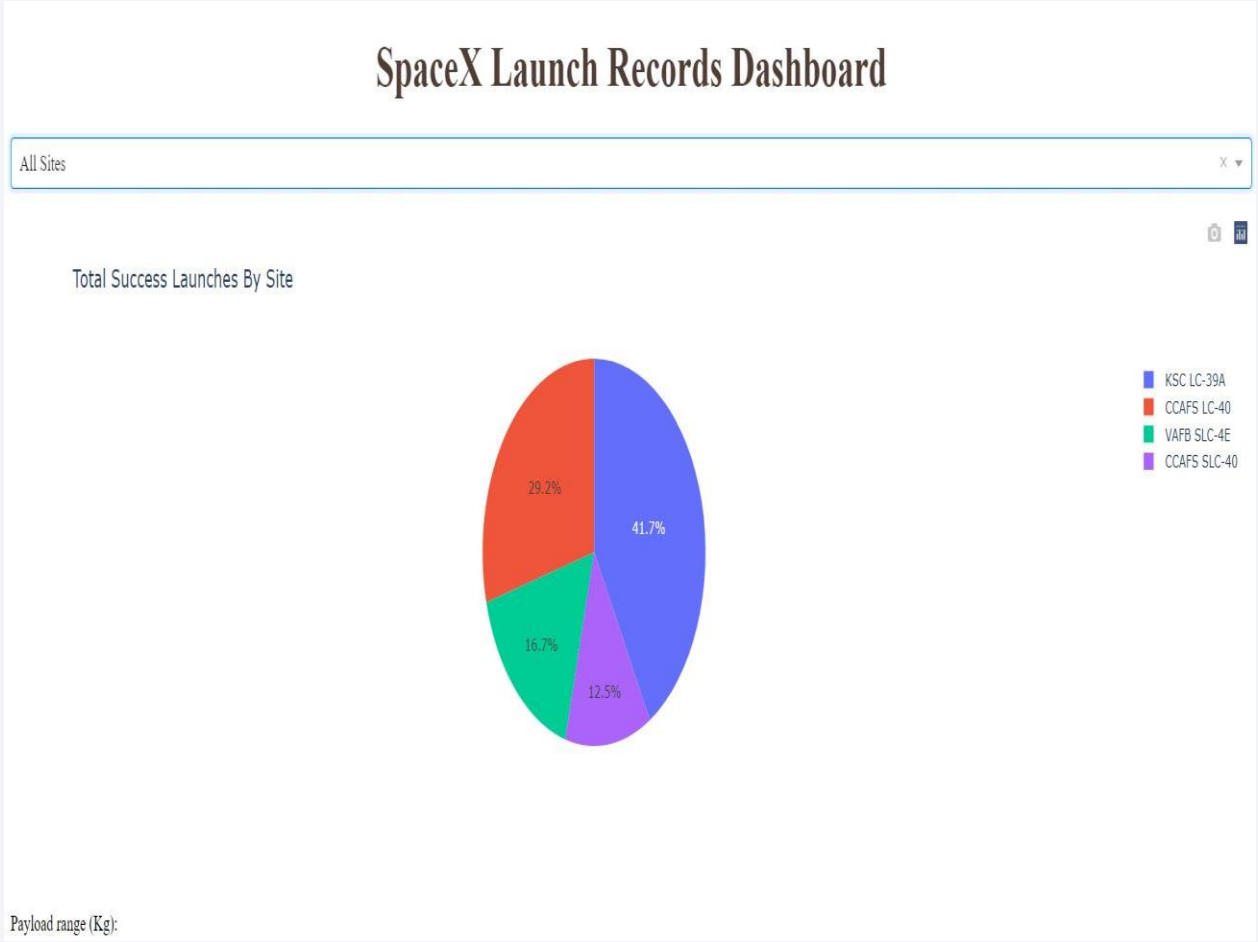
Section 5

Build a Dashboard with Plotly Dash



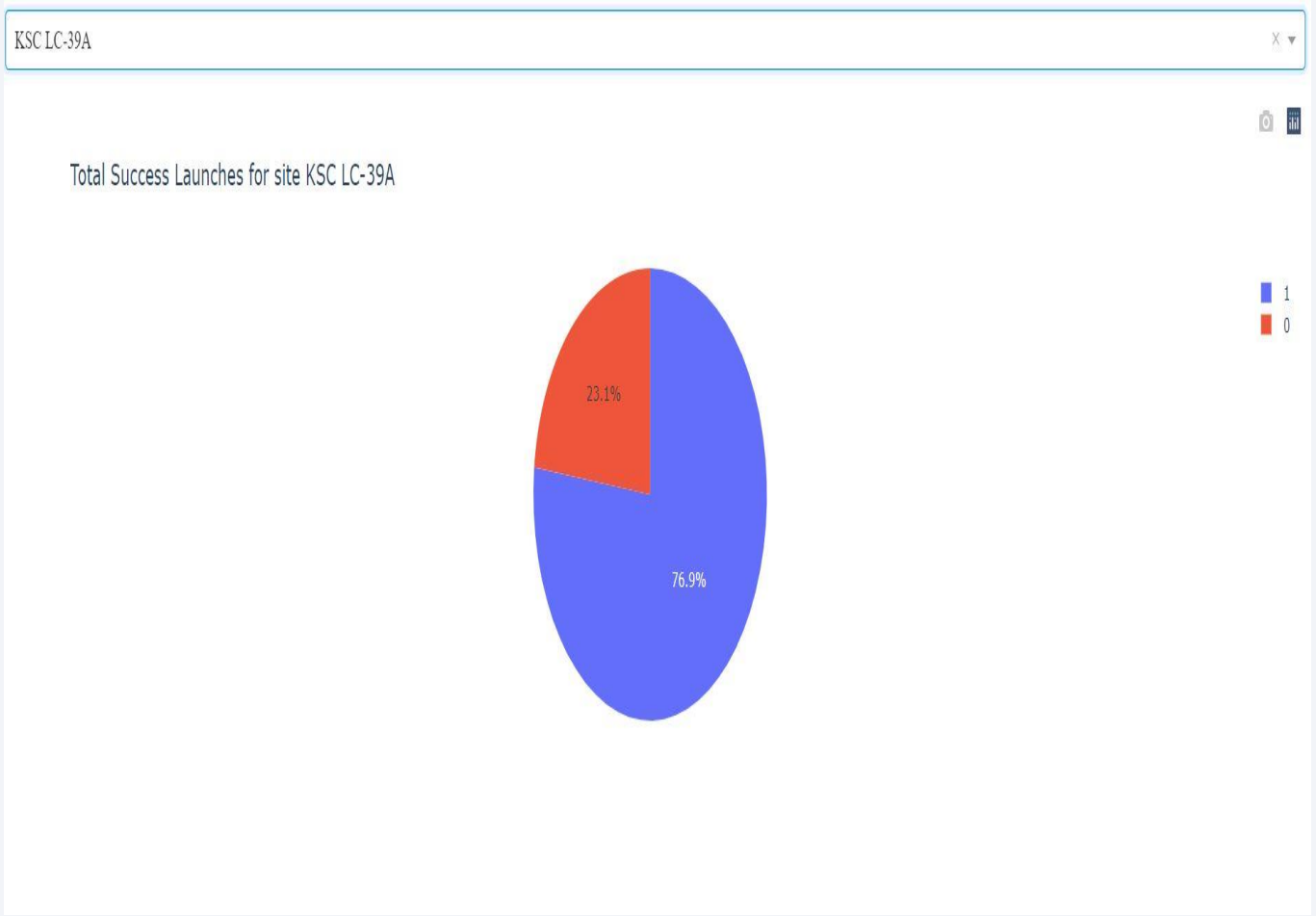
Total Successful Launches

- KSC LC-39A has the highest number of successful landing outcomes
- VAFB SLC-4E and CCAFS SLC-40 have the least number of successful landings
- KSC LC-39A and CCAFS LC-40 have over 70% of all successful landings



KSC LC-39A Successful vs Failed launches

- Over two-thirds of all launches ended successfully
- Conducted the second greatest number of total launches
- Only 13 launches have been conducted

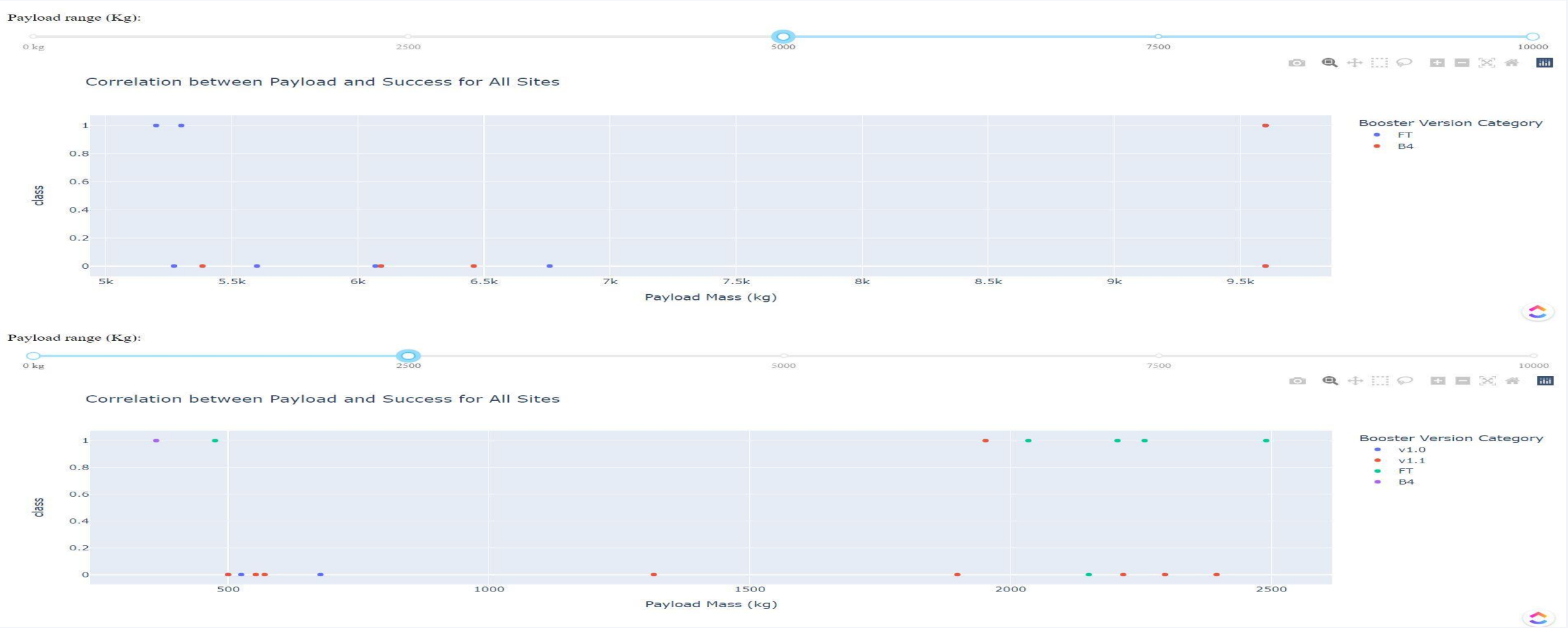


Payload vs Landing Outcome Success

- Falcon9 booster version vFT have the greatest number of successful landing outcomes overall.
- Booster vFT also has the highest number of successful landings with a payload mass between 2000kg and 4000kg.
- Booster version v1.1 was only successful once and with a payload less than 2000kg.
- Booster vB4 has the only successful landing over 6000kg.

Payload vs Landing Outcome Success

Payload vs Landing Outcome Success

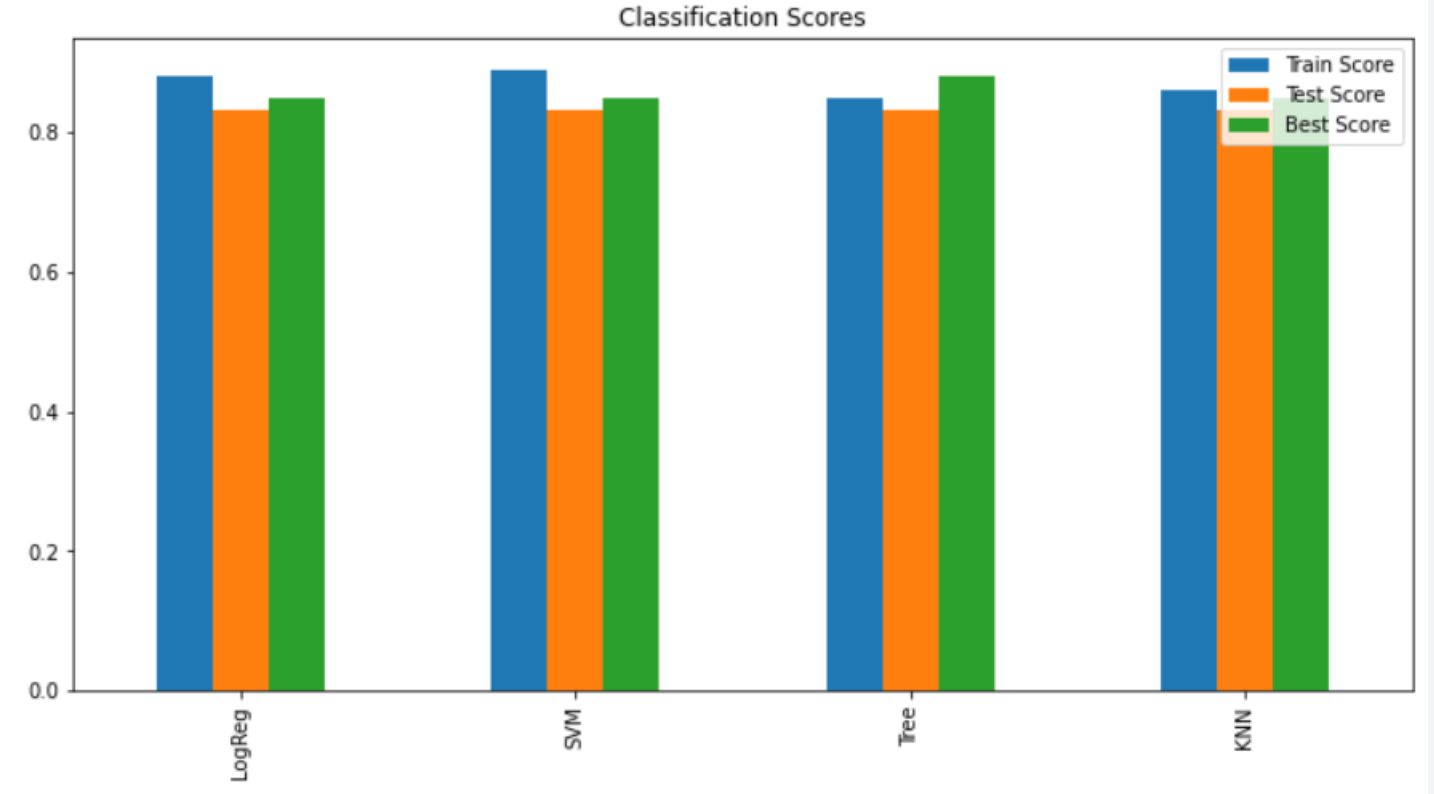


Section 6

Predictive Analysis (Classification)

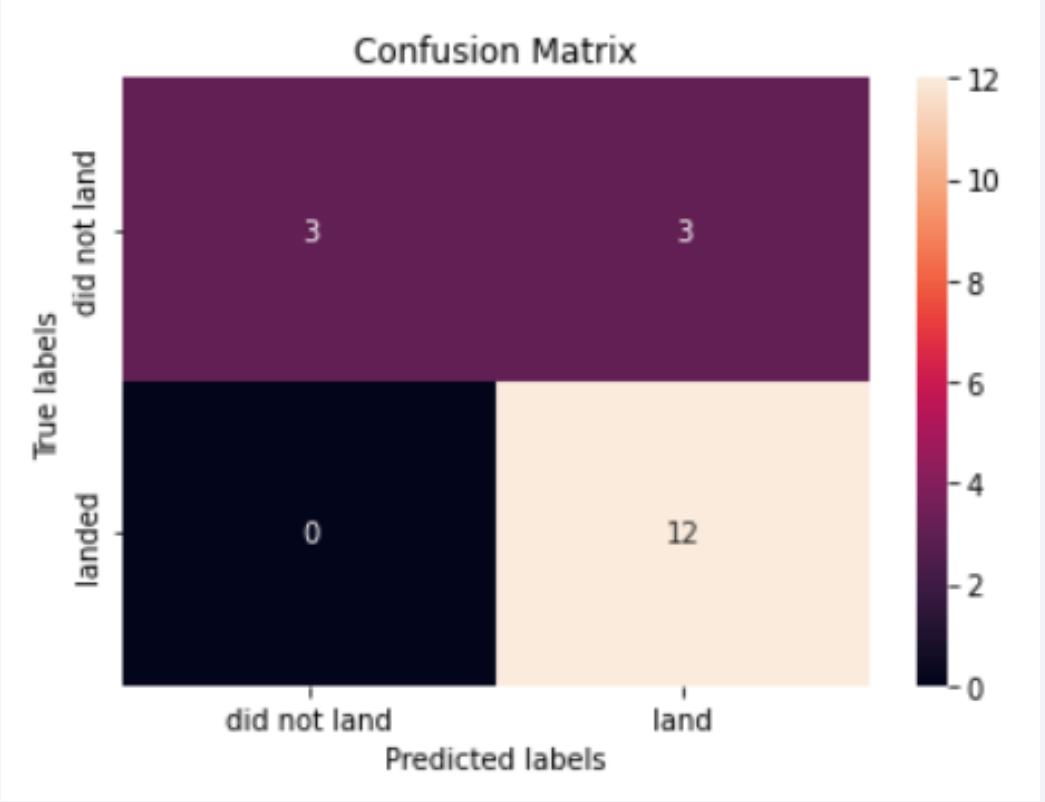
Classification Accuracy

- All models have an 83% classification accuracy
- When applying hyperparameters, the Decision Tree Classification model is 88% accurate
- All other models only improve to an 85% classification accuracy



Confusion Matrix

- Decision Tree Confusion Matrix
- 15 landings successfully predicted
- 3 landing outcomes were predicted unsuccessfully



Conclusions

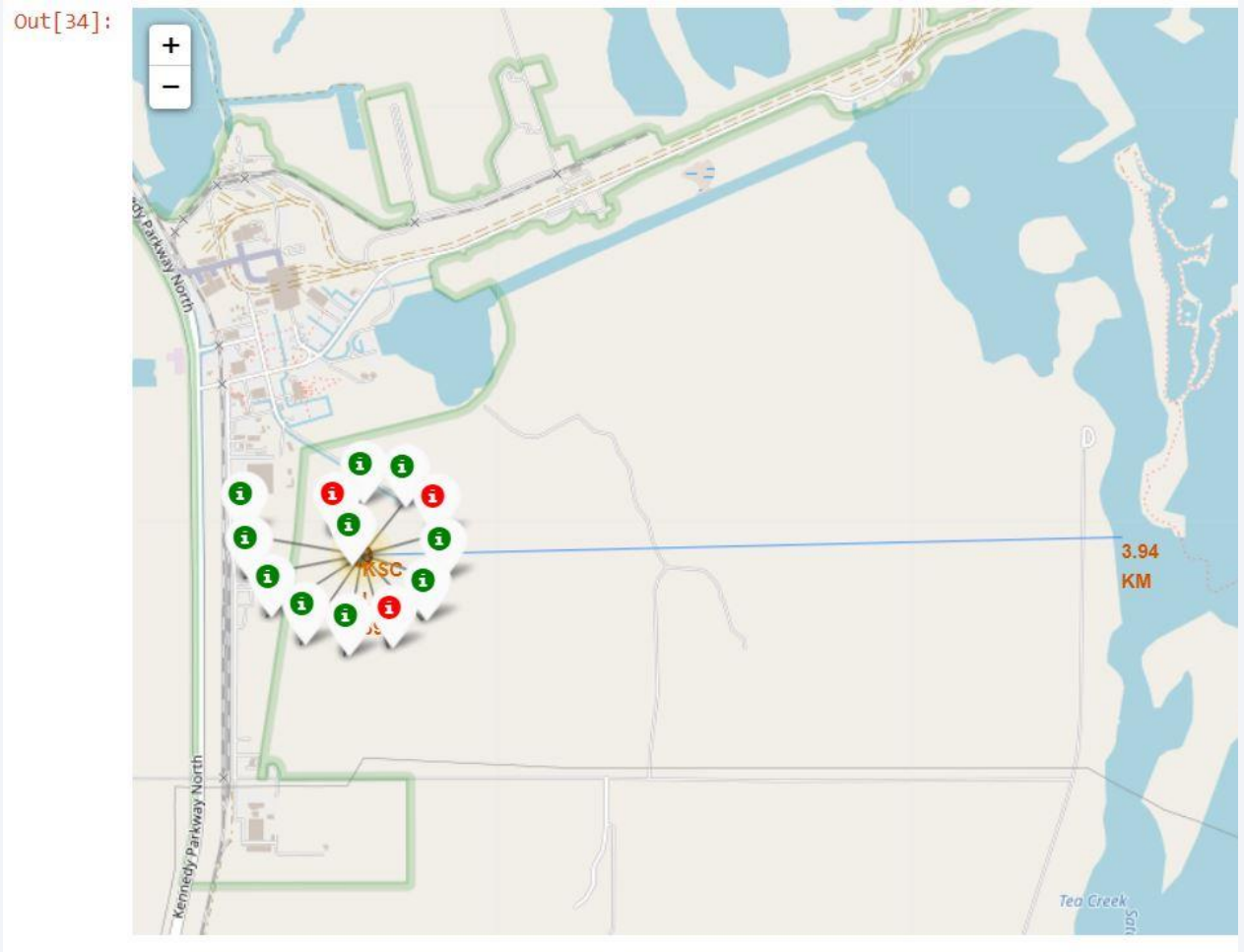
- Timeframe for successful booster recovery
  - 4 to 5 years after first flight
- Location of the launch site
  - KSC LC-39A
  - CCAFS site complex 40
- Pay load mass (in kg)
  - 2,000 kg or less
- Booster versions
  - F9 FT
  - F9 v1.1
- Orbit
  - SSO
  - ES-L1
  - HEO
- Try to acquire NASA as a vendor
  - NASA CRS or Commercial Resupply Services

Appendix

- Complete GitHub repo:
  - https://github.com/Ash021980/spaceyibmcap.git

Appendix

- KSC LC-39A at 3.94 km from the nearest coastline has the highest success rate and count
- Only launch site more than 1.5 km away from a coastline
- Has the second highest number of launches



Appendix

- Unique Launch Site SQL query

```
%%sql
Select distinct(Launch_Site) from SPACEXDATASET
```

- Resulting data frame

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

Appendix

- Successful drone ship landing outcome with a payload mass between 4000 and 6000 kg SQL query

```
%%sql
SELECT booster_version, payload_mass__kg_, landing__outcome FROM SPACEXDATASET
WHERE (landing__outcome like 'Success (drone ship)') and payload_mass__kg_ between 4000 and 6000
```

- Resulting data frame

booster_version	payload_mass__kg_	landing__outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Appendix

- Rank of landing outcomes between 2010-06-04 and 2017-03-20

```
%%sql
SELECT landing__outcome, count(landing__outcome) as total FROM SPACEXDATASET
WHERE (DATE BETWEEN '2010-06-04' and '2017-03-20')
GROUP BY landing__outcome
ORDER BY total desc
```

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

Appendix

- Classification data frame

	Train Score	Test Score	Best Score
LogReg	0.88	0.83	0.85
SVM	0.89	0.83	0.85
Tree	0.85	0.83	0.88
KNN	0.86	0.83	0.85

Thank you!