IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jia Shen Ang

18 July 2022

Outline

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

Executive Summary

- Summary of methodologies – The following methodologies were used to analyze data:
  - Data Collection using web scraping and the SpaceX API
  - Exploratory Data Analysis (EDA) – data wrangling, data visualization
  - Machine learning predictions
- Summary of all results
  - Data was successfully collected using web scrapping and API
  - EDA allowed us to identify key features of successful landings.
  - Machine Learning Prediction showed the best model to use when making predictions on the collected data.

Introduction

- This project is to evaluate the viability of our new company SpaceY and its ability to compete with SpaceX
- Problems we want to find answers to:
  - To find the best way to estimate the total cost of launches by predicting successfully landing of rocket first stages
  - Identify ideal locations for where we can launch rockets.

Section 1

Methodology

Methodology

Executive Summary

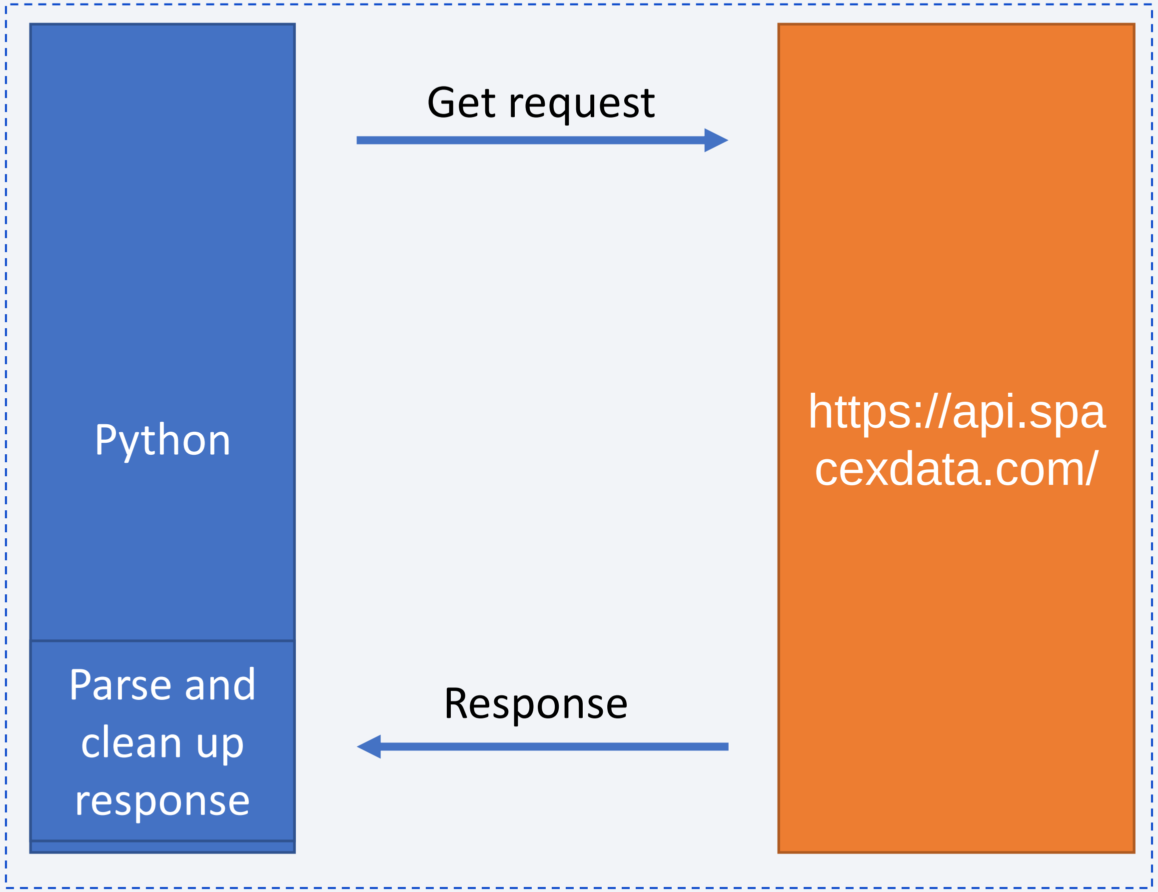
- Data collection methodology:
  - SpaceX data was acquired from two sources, the first from Wikipedia. The second from the SpaceX API
- Perform data wrangling
  - Collected data was processed to create labels based on the landing outcome
- Perform exploratory data analysis (EDA) using visualization 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 – SpaceX API

- Data was collected by sending a **get request** to the SpaceX API
- The API responded to the **get request** in a **JSON format**
- The data was parsed into a **Pandas Data Frame**
- The data frame was then cleaned up
    - Filtering only Falcon 9 rockets
    - Dealing with missing values
- Data then exported to CSV

Github URL:

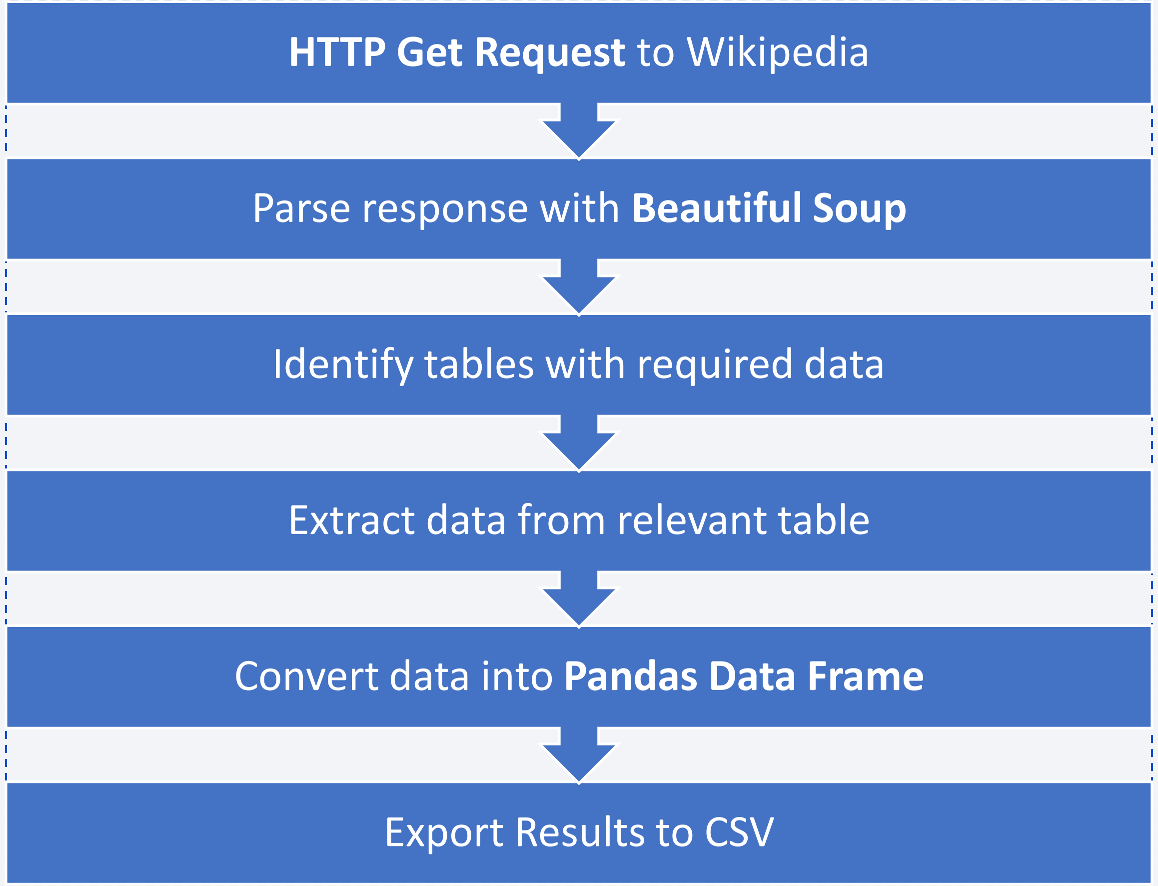
https://github.com/jiashenang/IBM-AppliedDataScienceCapstone/blob/master/1-1%20-%20Data%20Collection%20API.ipynb



Data Collection - Scraping

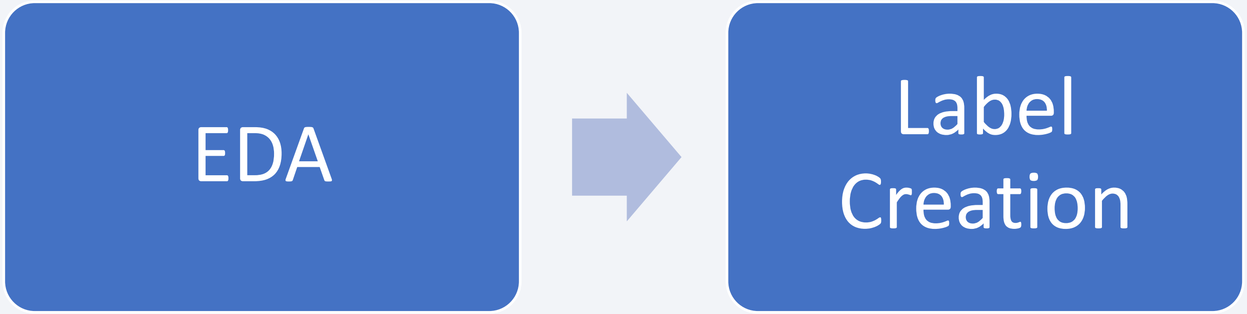
- Data was scraped from tables on Wikipedia
- Scraping was made possible with the following libraries
  - Request
  - Beautiful soup
  - Pandas

Github URL: https://github.com/jiashenang/IBM-AppliedDataScienceCapstone/blob/master/1-2%20-%20Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Initial Exploratory Data Analysis was done to see:
  - Number of launch sites
  - Number of each orbit type
  - Outcomes of landing attempts
- Finally, labels were created for each landing outcome
  - 1 for a successful landing attempt
  - 0 for an unsuccessfully landing attempt



Github URL: https://github.com/jiashenang/IBM-AppliedDataScienceCapstone/blob/master/1-3%20-%20Data%20Wrangling.ipynb

EDA with Data Visualization

- In general Scatter Charts were plotted to see the relationship between pairs of features
- A bar chart was plotted to visualize success of each type of orbit
- And a line chart was used to visualize the overall success rate over time.

Github URL: https://github.com/jiashenang/IBM-AppliedDataScienceCapstone/blob/master/2-2%20-%20EDA%20with%20Visualization.ipynb

EDA with SQL

- Query the name of the unique launch sites
- Query 5 records where the launch site's name begins with the string 'CCA'
- Query the total payload mass carried by boosters launched by NASA (CRS)
- Query the average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Rank 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

Github URL: https://github.com/jiashenang/IBM-AppliedDataScienceCapstone/blob/master/2-1%20-%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and map cluster were use to a folium map
  - Marker were used to mark key launch sites
  - Circles were used to highlight specific coordinates
  - Lines were used to show distance between two points
  - Clusters were used to group together events – in this case launch attempts

- Github URL: https://github.com/jiashenang/IBM-AppliedDataScienceCapstone/blob/master/3-1%20-%20Interactive%20Visual%20Analytics%20with%20Folium.ipynb

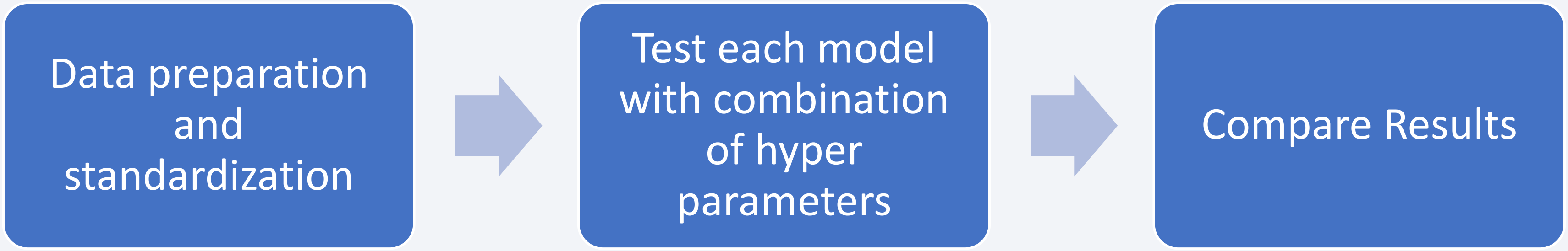
Build a Dashboard with Plotly Dash

- The following plots and charts were plotted
  - A pie of successful launch by site, or from the site depending on the value selected
  - A scatter to plot payload mass and successful landing, payload mass can be selected using a slider
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Github URL: https://github.com/jiashenang/IBM-AppliedDataScienceCapstone/tree/master/3-2%20-%20Dash

Predictive Analysis (Classification)

- 4 Models were built using: Logistic Regression, SVM, Classification Trees and K Nearest Neighbors
- The models were then compared



Github URL: https://github.com/jiashenang/IBM-AppliedDataScienceCapstone/blob/master/4-1%20-%20Machine%20Learning%20Prediction.ipynb

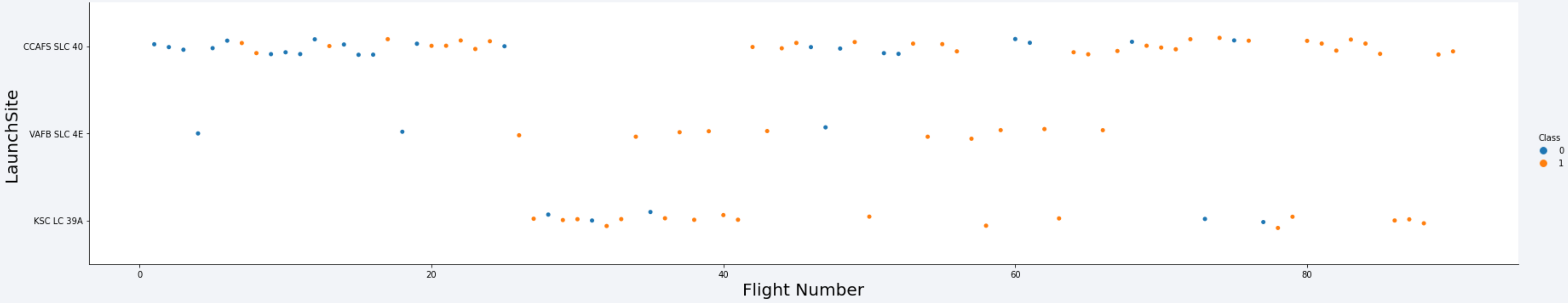
Results

- Exploratory data analysis results
  - SpaceX uses 4 launch sites
  - SpaceX has been able to successfully launch payloads into orbit very reliably around 100% of the time on some orbits. They are also now able to land about 80% of their first stage boosters giving them enormous cost savings in reuse.
- Interactive analytics demo in screenshots
- Predictive analysis results
  - Predictive analysis did not yield any models which stood out from the rest they preformed equally when using the test data.

Section 2

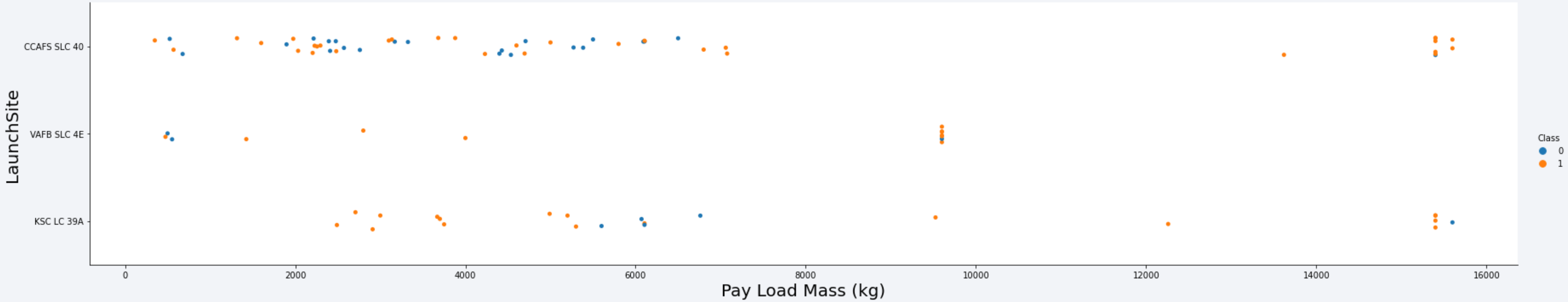
Insights drawn from EDA

Flight Number vs. Launch Site



- Earlier launches took place from CCAFS SLC, and many of those ended in failure to land.
- At around 20 flights, the success rate of missions went up dramatically
- Of the three sites, VAFB SLC which is on the west coast is used the least.
- Large gap in launches around flight 29 from CCAFS SLC caused by a launch pad explosion of a Falcon 9 rocket during testing.

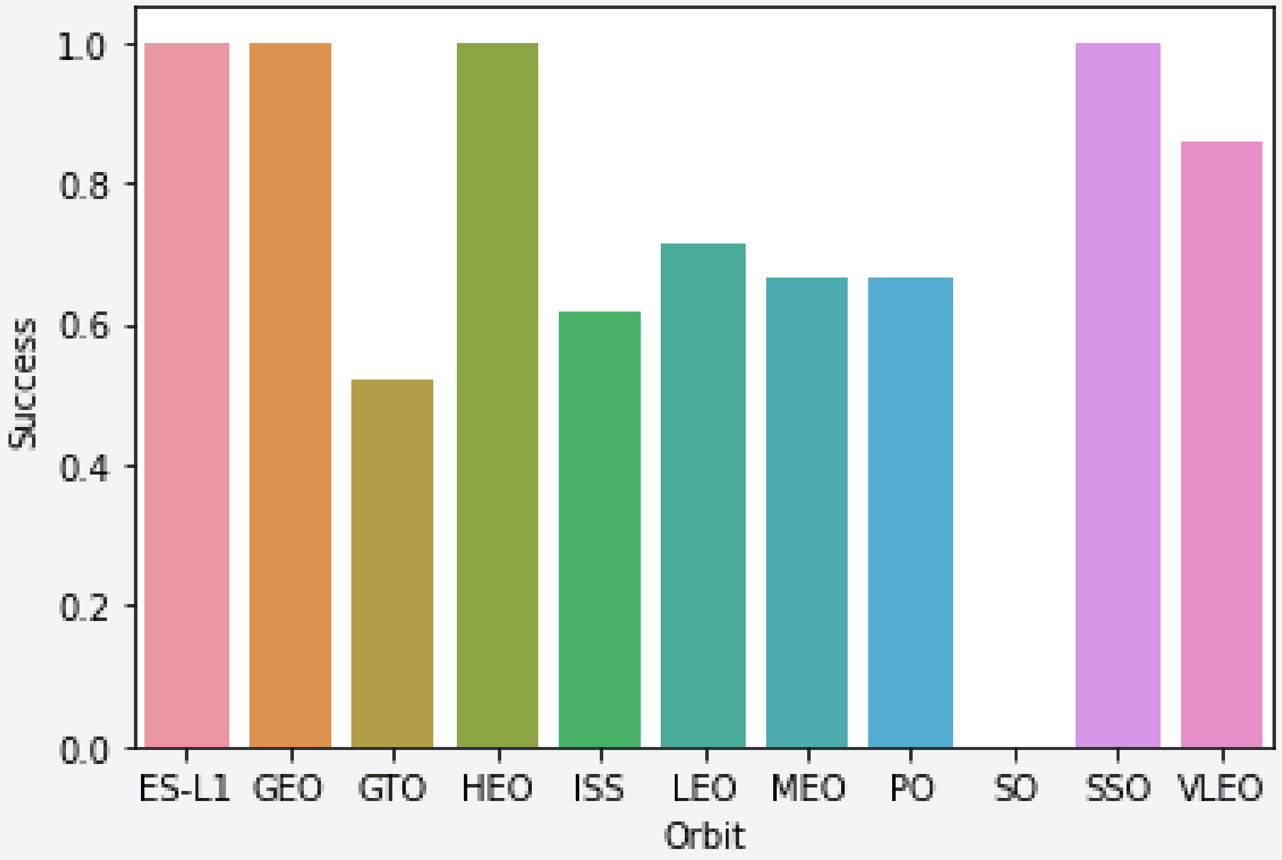
Payload vs. Launch Site



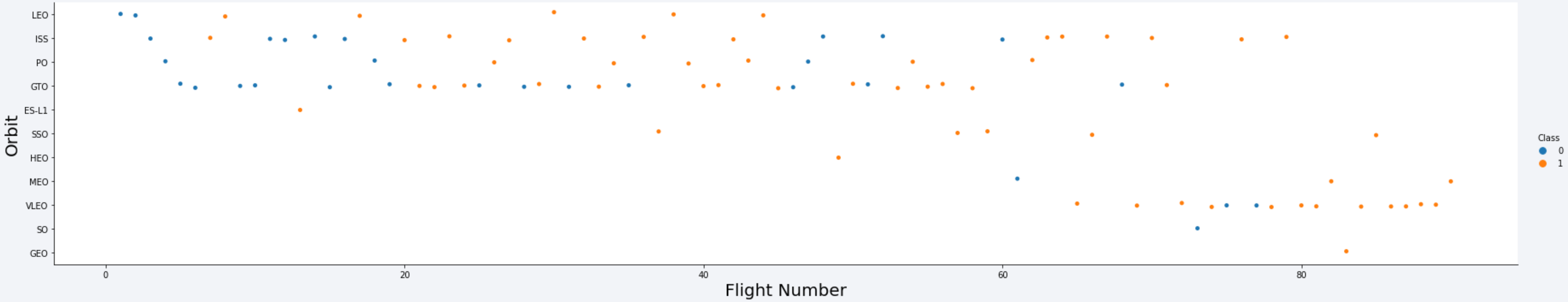
- VAFB-SLC has no payloads with mass greater than 10,000 kg.
- Majority of launch have a payload mass of under 8,000 kg.

Success Rate vs. Orbit Type

- ES-L1, GRO, HEO, SSO orbit types have the highest success rate. Succeeding 100% of the time.
- While GTO has the lowest success rate.

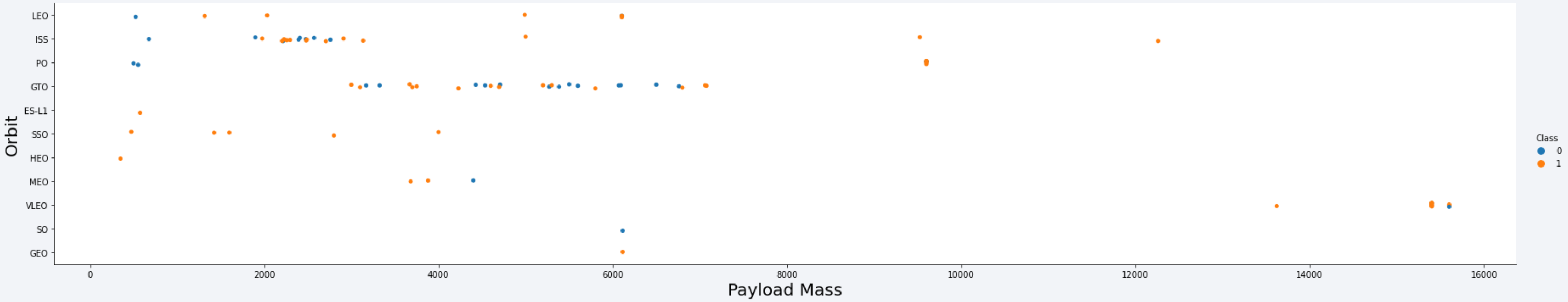


Flight Number vs. Orbit Type



- Earlier flights focused on sending payload to LEO, ISS, PO, and GTO orbits.
- Later flights started to send payload to VLEO – many of those are Starlink launches.

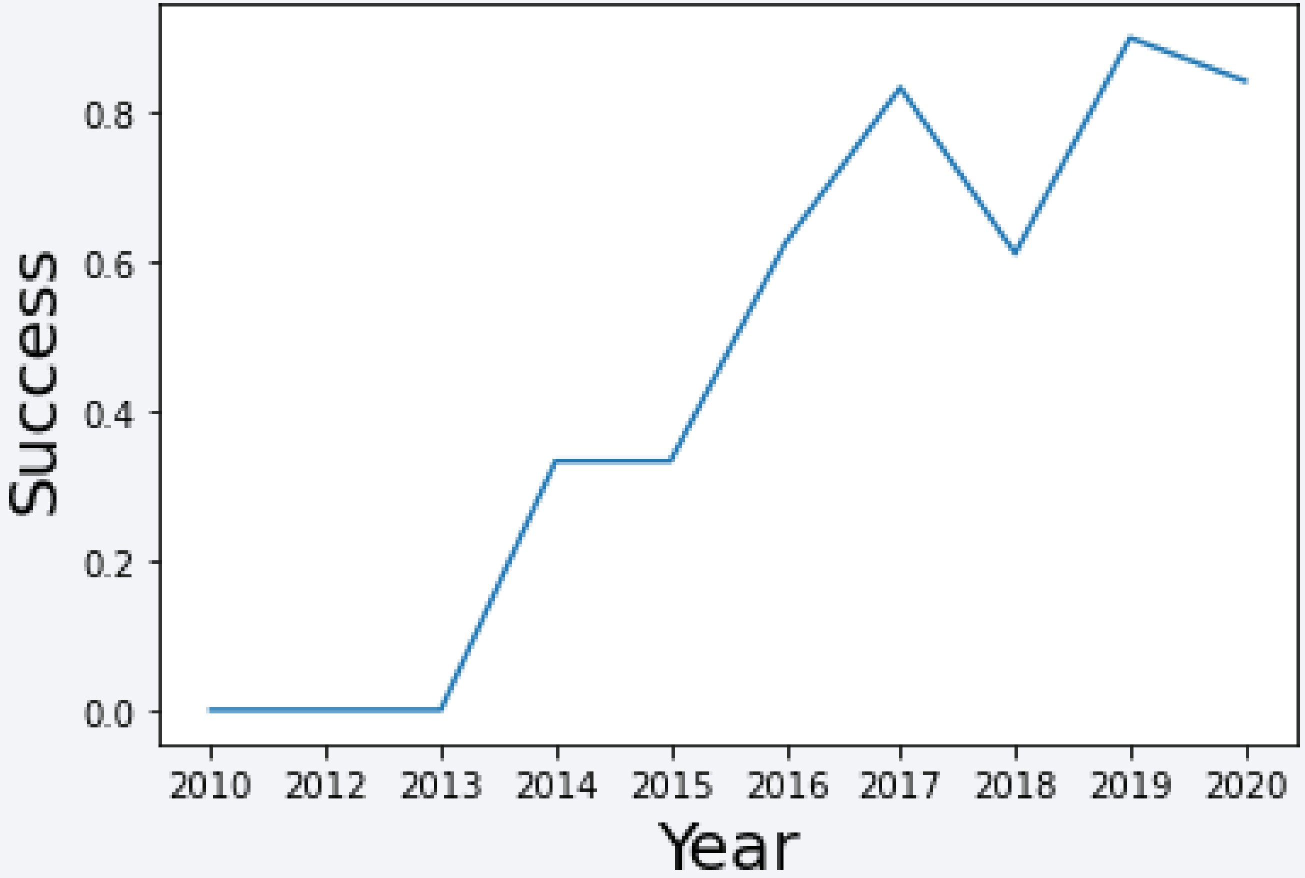
Payload vs. Orbit Type



- Certain orbits have certain payload ranges
- When launching to VLEO the Falcon 9 can carry close to 16,000 kg
- When flying to the ISS, a typical launch would carry payload in 2,000 – 3,000 kg
- While to GTO the payload range of 3,000 – 7,000 kg

Launch Success Yearly Trend

- SpaceX has made dramatic improvements to successfully landing their boosters.



All Launch Site Names

```
%%sql
SELECT DISTINCT launch_site FROM SPACEXTBL
```

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

Launch Site Names Begin with 'CCA'

```
%%sql
SELECT *
FROM SPACEXTBL
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

```
%%sql
SELECT SUM(payload_mass__kg_) AS "Total Payload Mass Carried by Booster Launched by NASA (CRS)"
FROM SPACEXTBL
WHERE customer = 'NASA (CRS)'
```

Total Payload Mass Carried by Booster Launched by NASA (CRS)
45596

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(payload_mass__kg_) AS "Average Mass"
FROM SPACEXTBL
WHERE booster_version LIKE 'F9 v1.1%'
```

Average Mass
2534

First Successful Ground Landing Date

```
%%sql
SELECT MIN(DATE)
FROM SPACEXTBL
WHERE landing__outcome = 'Success (ground pad)'
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT *
FROM SPACEXTBL
WHERE landing__outcome = 'Success (drone ship)'
    AND payload_mass__kg_ BETWEEN 4000 AND 6000
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

```
%%sql
SELECT mission_outcome AS "Mission Outcome", COUNT(mission_outcome) AS "Count"
FROM SPACEXTBL
GROUP BY mission_outcome
```

Mission Outcome	Count
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

```
%%sql
SELECT booster_version
FROM SPACEXTBL
WHERE payload_mass__kg_=(SELECT MAX(payload_mass__kg_)
                         FROM SPACEXTBL)
```

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

```
%%sql
SELECT *
FROM SPACEXTBL
WHERE landing__outcome = 'Failure (drone ship)'
    AND DATE LIKE '2015%'
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2015-01-10	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

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

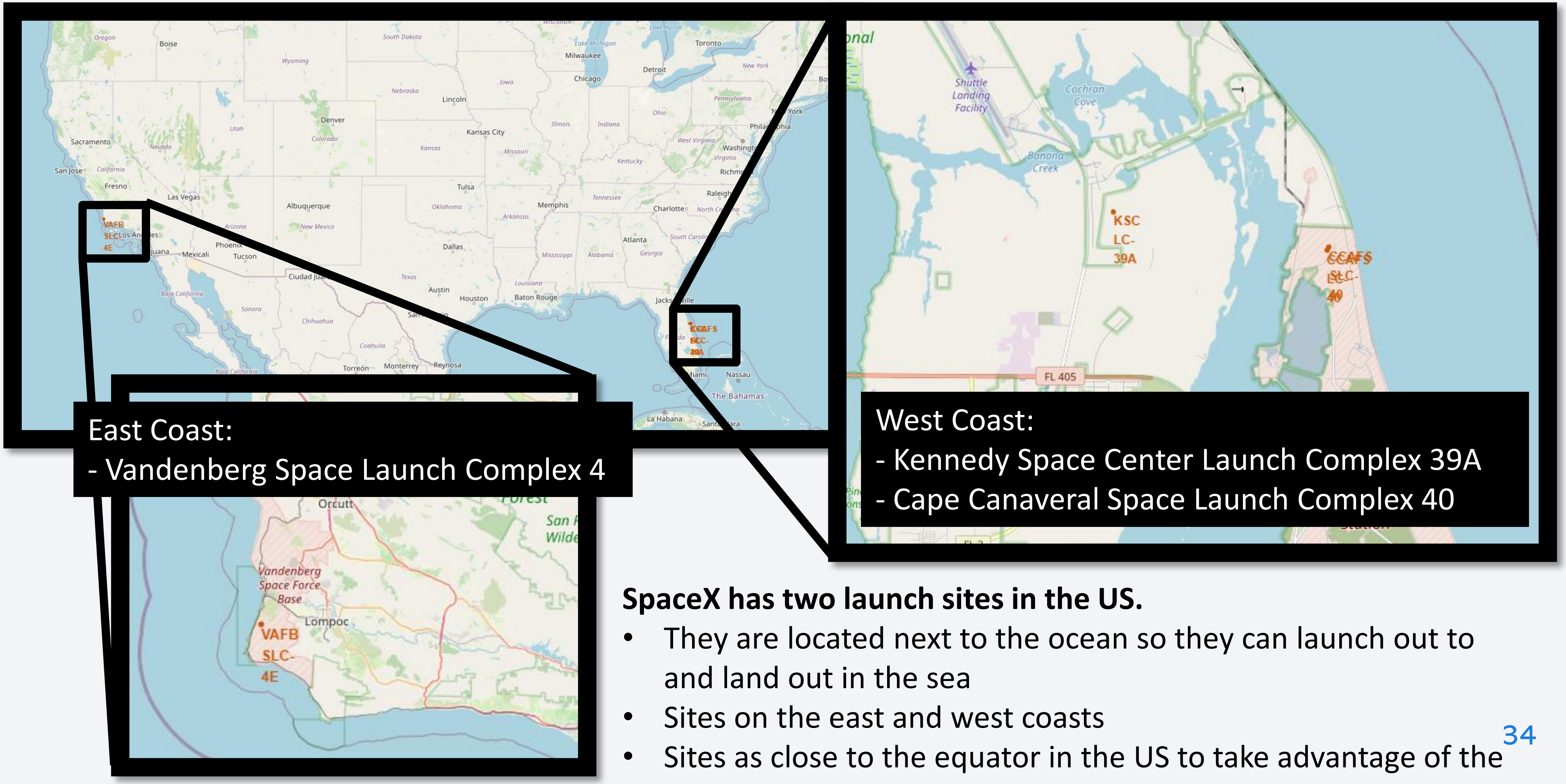
```
%%sql
SELECT landing__outcome AS "Landing Outcomes", COUNT(landing__outcome) AS "Count"
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing__outcome
ORDER BY Count desc
```

Landing Outcomes	Count
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

Section 3

Launch Sites Proximities Analysis

Locations of SpaceX Launch Sites

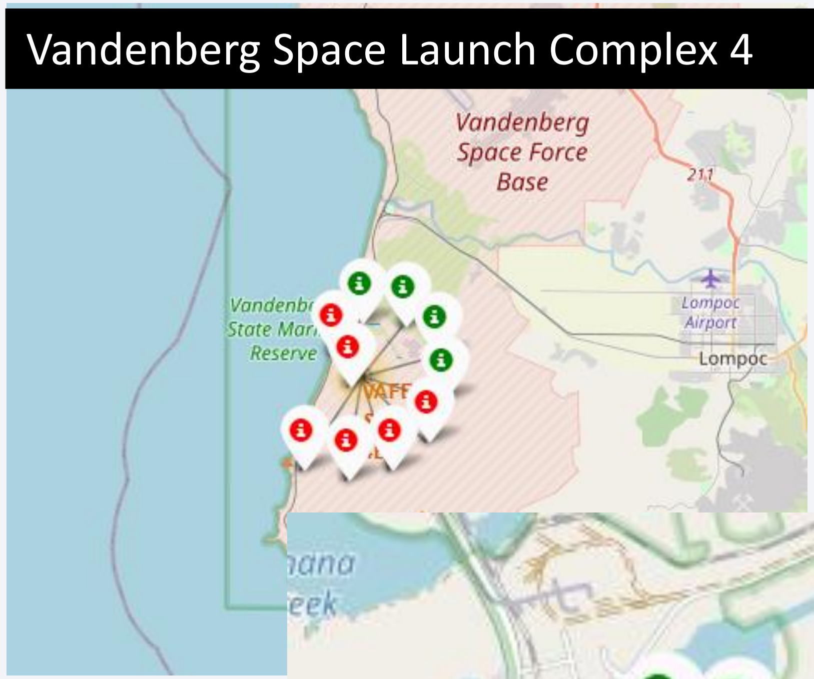


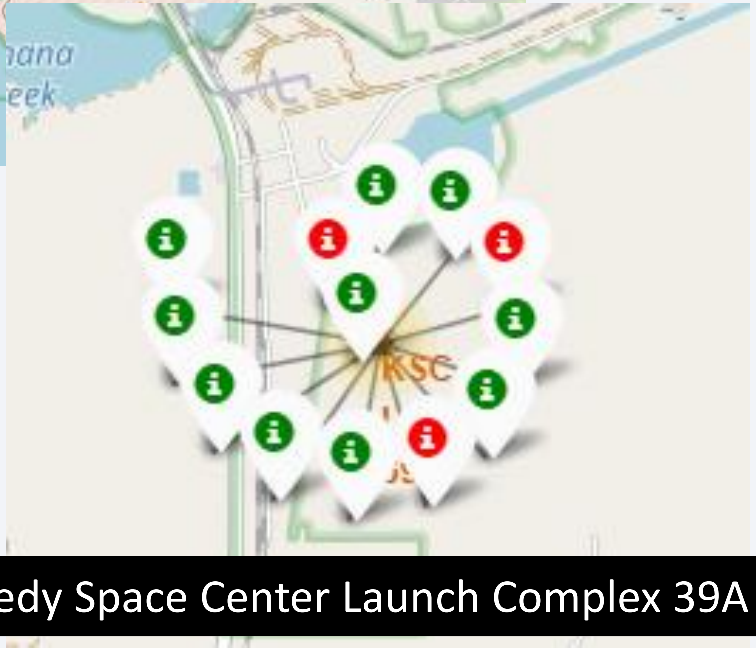
SpaceX has two launch sites in the US.

- They are located next to the ocean so they can launch out to and land out in the sea
- Sites on the east and west coasts
- Sites as close to the equator in the US to take advantage of the earth's spin.

Launch outcomes from each site visualized

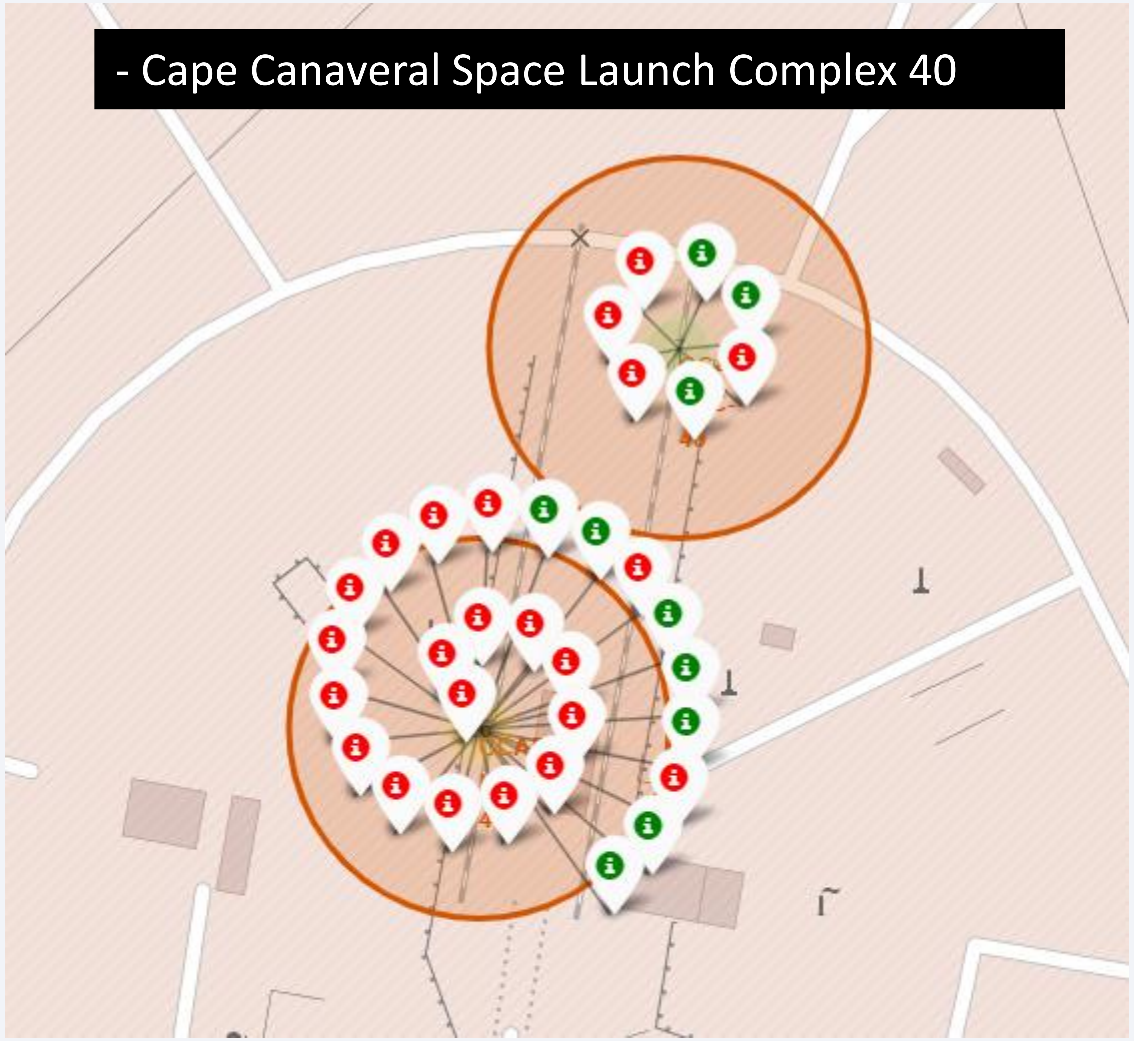
Vandenberg Space Launch Complex 4



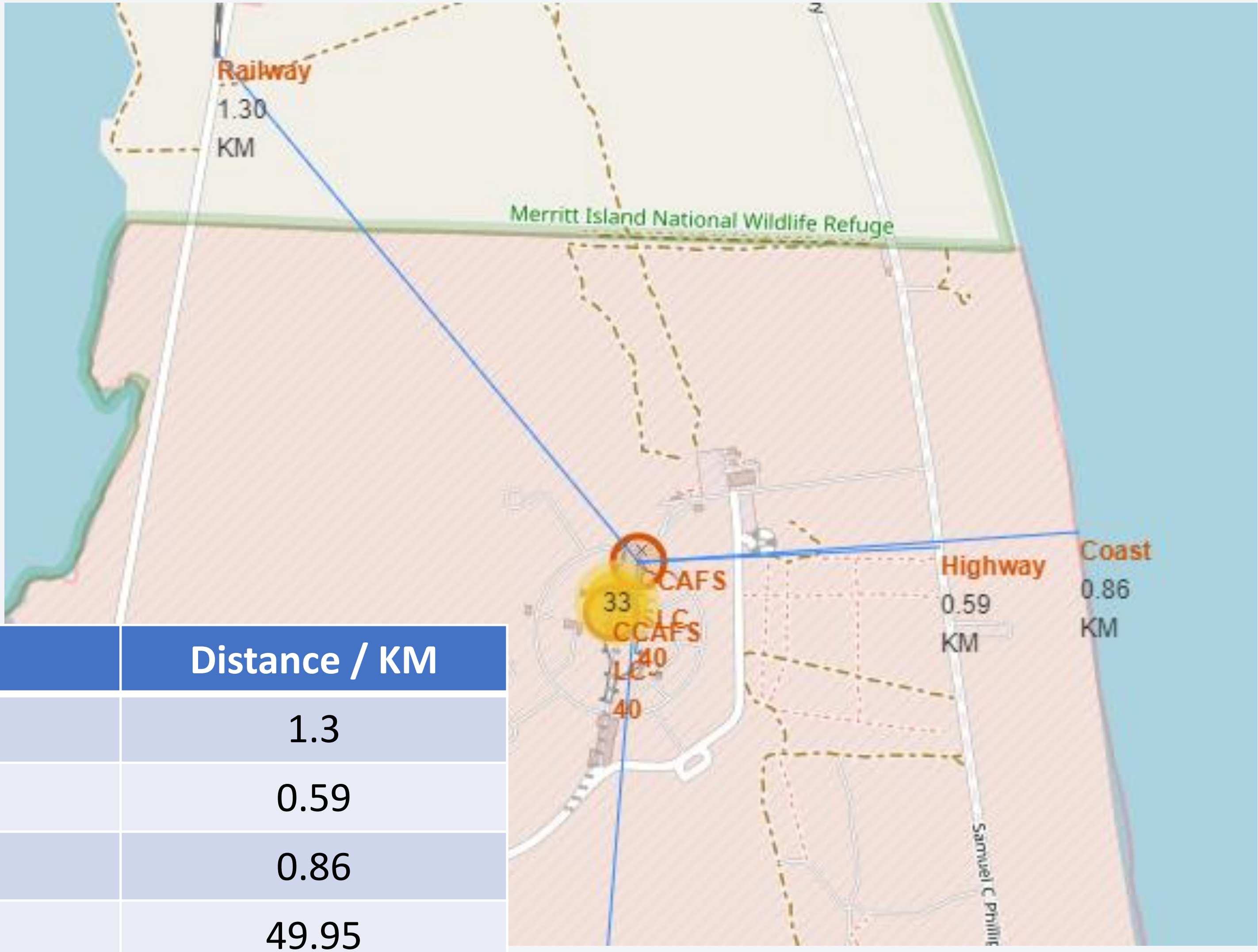


Kennedy Space Center Launch Complex 39A

- Cape Canaveral Space Launch Complex 40



Proximity of the CCAFS SLC-40 launch site to key sites



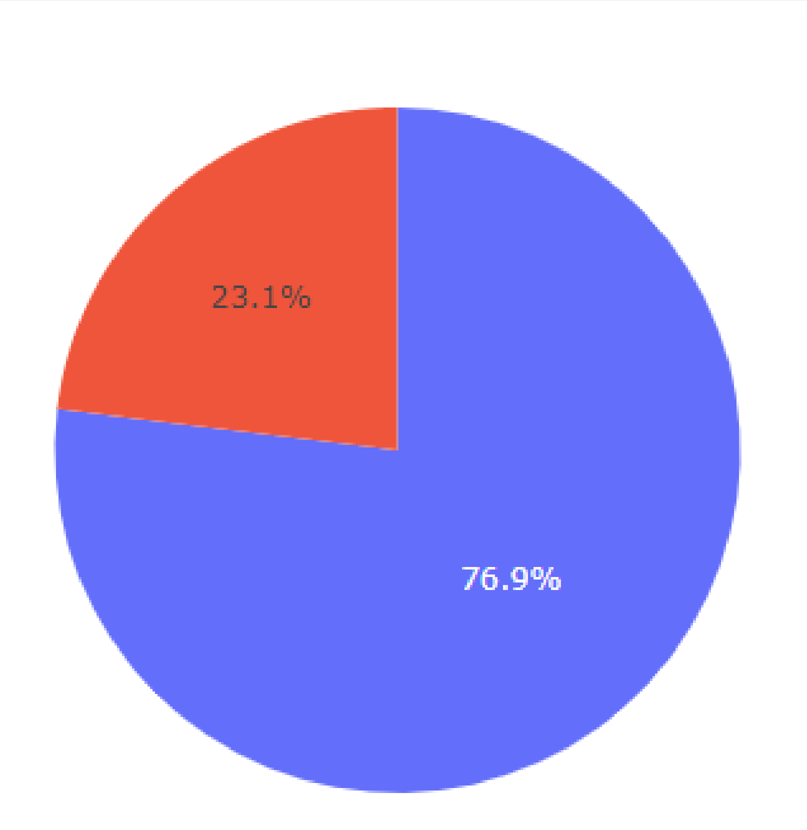
Site	Distance / KM
Railway	1.3
Highway	0.59
Coast	0.86
City	49.95

Section 4

Build a Dashboard with Plotly Dash

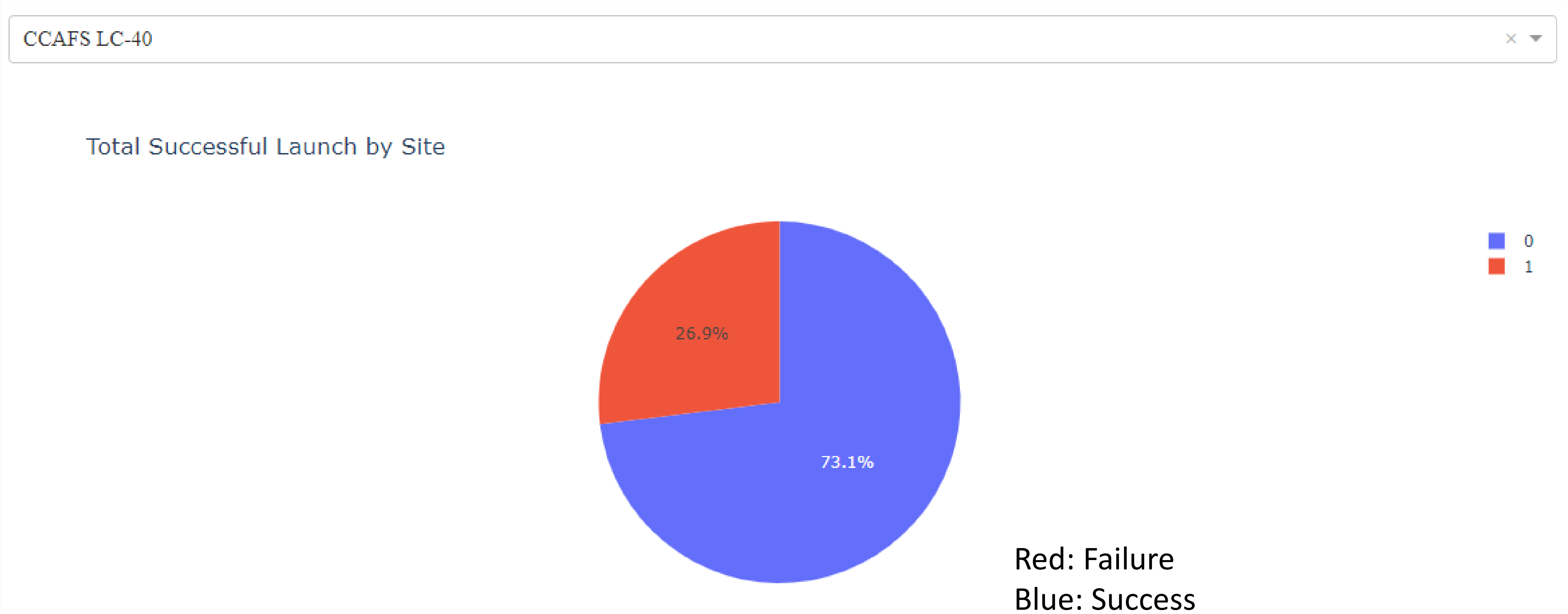
Total Success and Failures for SpaceX Launches

- SpaceX are successfully launching their payloads over 75% of the time



Red: Failure
Blue: Success

CCAFS LC-40 has launched the most successful number of missions



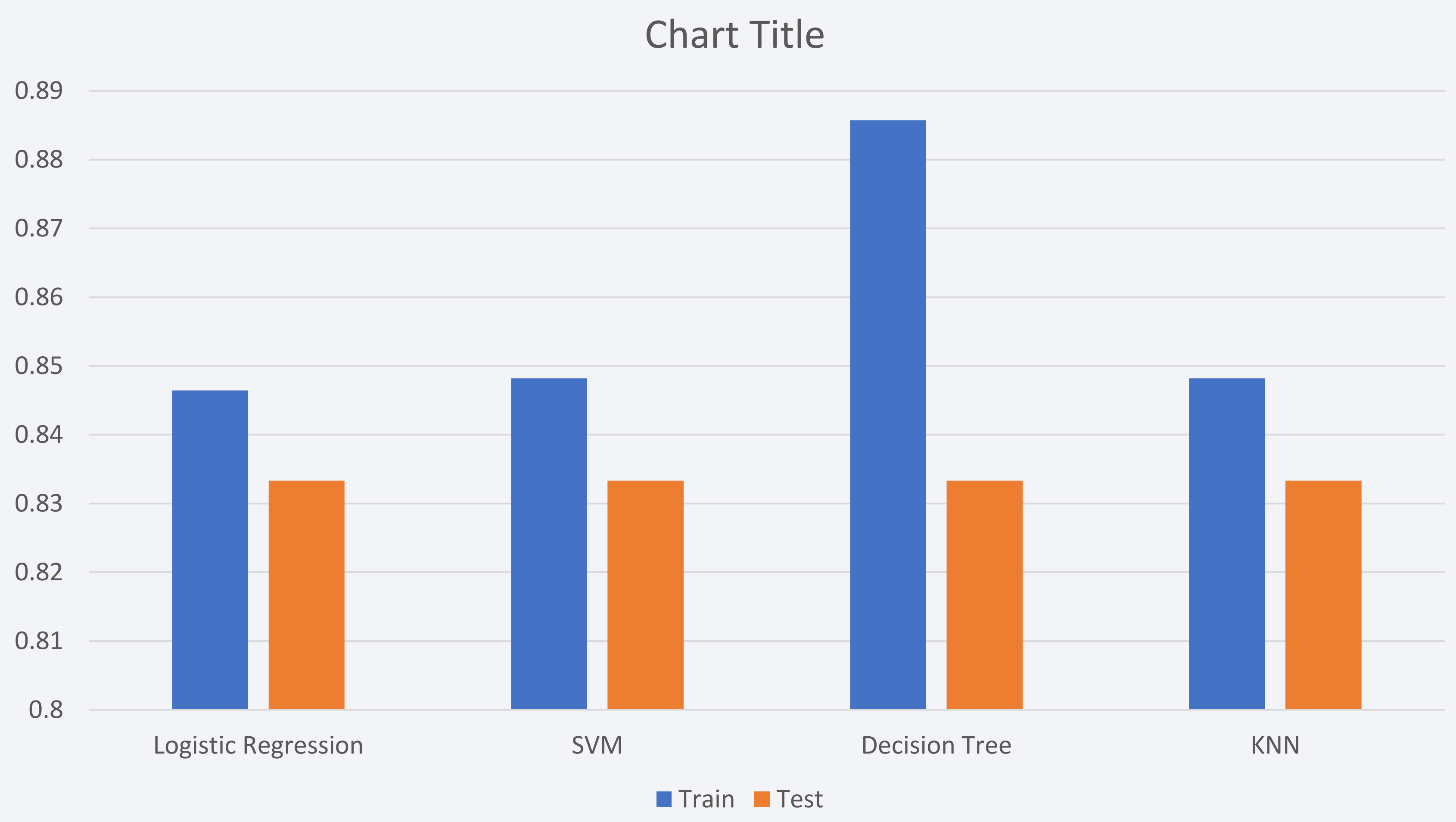
Red: Failure
Blue: Success

Payload vs. Launch Outcome

Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

Conclusions

- Model preformed the same when used test data
- Difference was observed with training data – there the decision tress preformed better.

Thank you!