IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rishi Ramaratnam
June 11-2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data is collected from Web Scraping and Space X API Data
  - Data Wrangling
  - Data Visualization via Python and SQL
  - Use of Interactive Dashboards and Maps for Data Analytics
  - Use of Machine Leaning Classification Algorithms for predictive analysis

- Summary of all results
  - Results from Data Analysis
  - Interactive Dashboard and Maps analysis for easy visual analysis
  - Model evaluation
  - Use of Classification models to determine successful predictions

Introduction

- Project background and context
  - Working for Space Y to compete with Space X
  - Want to Determine the price of each launch of Space X
  - Falcon 9, space X's Rocket, is known for the recuse of its stage 1 rockets while delivering payloads
- Problems you want to find answers
  - The project aims to predict the success rate of stage 1 Falcon 9 rockets using classification models
  - Want to know the chances the stage 1 is used.
  - How much does size of payload , location of launch, or orbit effect outcome of rocket launch?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Got Data from SpaceX API
  - Wikipedia collection of data
- Perform data wrangling
  - Used EDA for pattern recognition
  - Fixed missing values and deleted un-needed values
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Used KNN, Logistic regression and support Vector Machine Tree decision classifier to find best model

Data Collection

- Describe how data sets were collected.
  - Used an API from SpaceX
  - And a Wikipedia Webpage to get Falcon 9 launch records

Data Collection – SpaceX API

- On the right, we can see the basic flow of events we took to collect the data from the API.
- https://github.com/puligoat11/SpaceX-Falcon-9-Final-Project-/blob/main/Part%201Data-collection-api.ipynb

Step 1: Request Data using Get

Step 2: Convert into Data Frame using json

Step 3: Data Wrangling for missing data

Step 4: Save to CSV

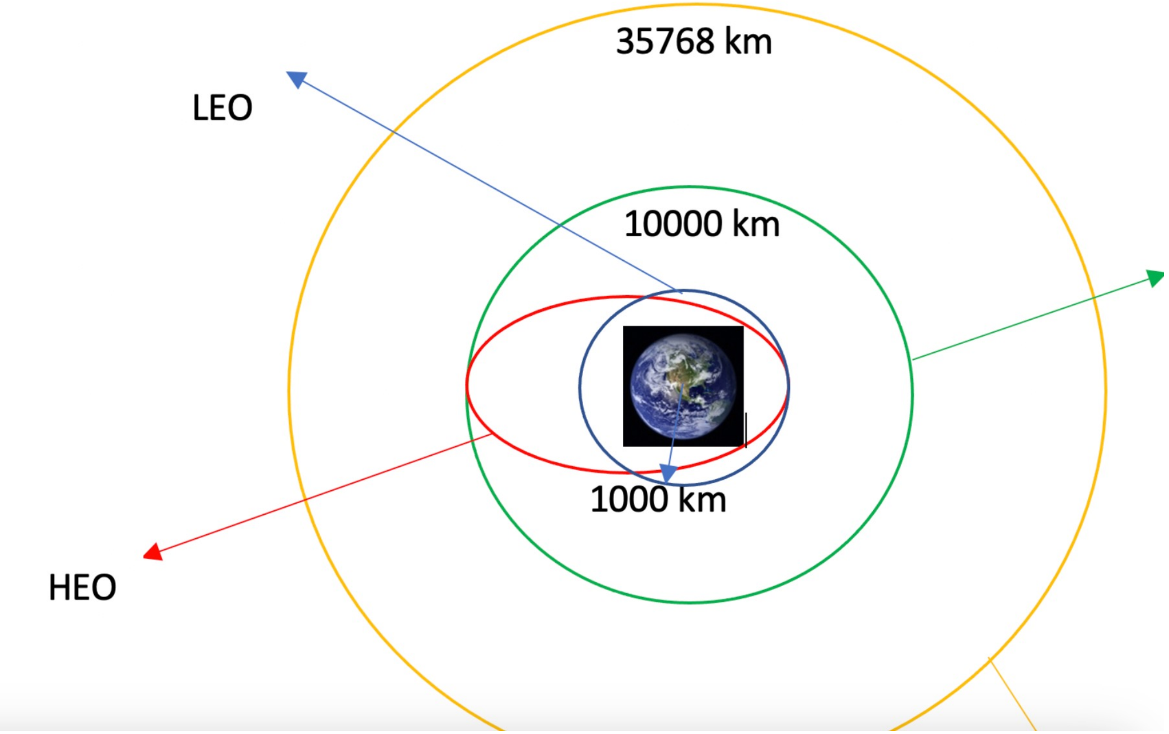
Data Collection - Scraping

- On the right, we can see the basic flow of events we took to scrape the data from the Wiki Page
- https://github.com/puligoat11/SpaceX-Falcon-9-Final-Project-/blob/main/Data%20Webscraping.ipynb

1. Request Data from Wiki page
2. Extract collum data from table header
3. Parse data to create data frames
4. Export to csv

Data Wrangling

- Found Missing Values
- Calculated number of launches for each site
- Calculated Number and occurrence of each orbit
- Mission outcome results
- Saved as a csv file
- https://github.com/puligoat11/SpaceX-Falcon-9-Final/blob/main/Data%20wrangling.ipynb



EDA with Data Visualization

- Scatterplot Relationship between Flight Number and Launch site
- Scatterplot Relationship between Payload and Launch site
- Bar chart for success rate by Orbit Type
- Scatter between Flight Number and Orbit Type
- Scatter between Payload and Orbit Type
- Features Engineering
  - We use scatter plots for two sets of categorical data, barchart for more than 2 sets of categorical data, and lineplot to see the time series
- https://github.com/puligoat11/SpaceX-Falcon-9-Final-Project-/blob/main/Expolring%20and%20Preping%20Data.ipynb

EDA with SQL

- Summarize SQL Queries
  - Names of Unique Launch Sites
  - Total Payload mass carried by boosters of certain type
  - Find date of a specific successful landing
  - Find names of boosters that were successful in drone ship and have a mass between 4000 and 6000
  - Rank Landing Outcomes
- https://github.com/puligoat11/SpaceX-Falcon-9-Final-Project-/blob/main/SQL_Lab.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Marked all launch sites on a map
  - Marked success/ failures for launches on a map
  - Calculated the distances between a launch site to its proximities
- Explain why you added those objects
  - Wanted to see if there was any trends in geographic location that lead to success's or failures for launches
- https://github.com/puligoat11/SpaceX-Falcon-9-Final-Project-/blob/main/Map%20Analytics.ipynb

Build a Dashboard with Plotly Dash

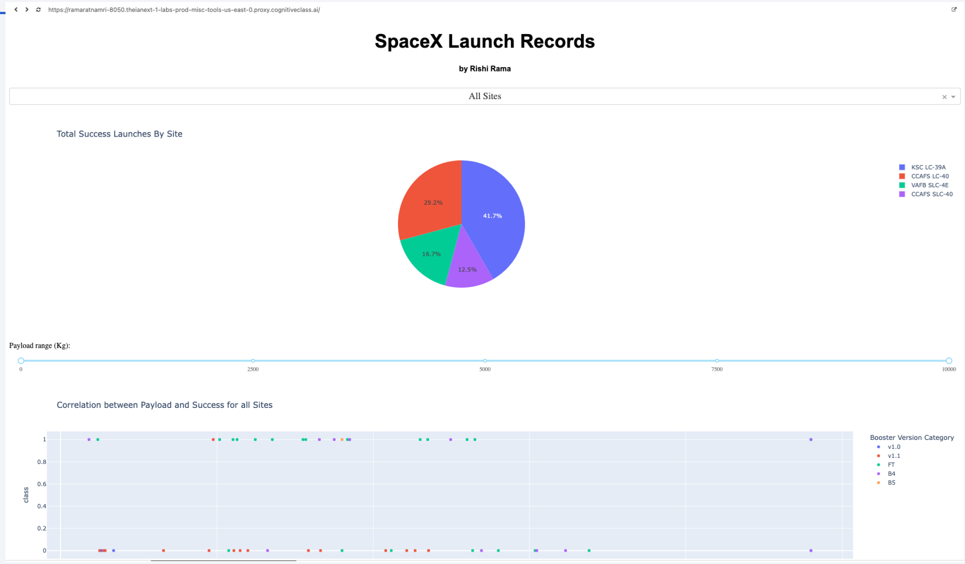
- Summarize what plots/graphs and interactions you have added to a dashboard
  - Launch site info (successes) through a pie chart
  - Payload select through a slicer range
  - A success payload scatter chart plot that is based on selected site
- We picked these to find the relationship for success between launch site and payload
- https://github.com/puligoat11/SpaceX-Falcon-9-Final-Project-/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Steps to find the best performing classification model
    1. Data Wrangling/ Standardization
    2. Splitting Data into Training and Test Data
    3. Ran Predictive Models
    4. Model evaluation and selection
- Created logistic regression, support vector machine, decision tree, and KNN models,
- https://github.com/puligoat11/SpaceX-Falcon-9-Final-Project-/blob/main/Machine_Learning_Prediction_Part_5.jupyterlite%20(1).ipynb

Results

- Exploratory data analysis results
  - We can see success rate has gone up since 2013-2017
- Interactive analytics demo in screenshots to right
- Predictive analysis results
  - All models are roughly similar with accuracy's of around 83%
  - However, the tree classifier model had an accuracy of around 89% considering all the parameters.

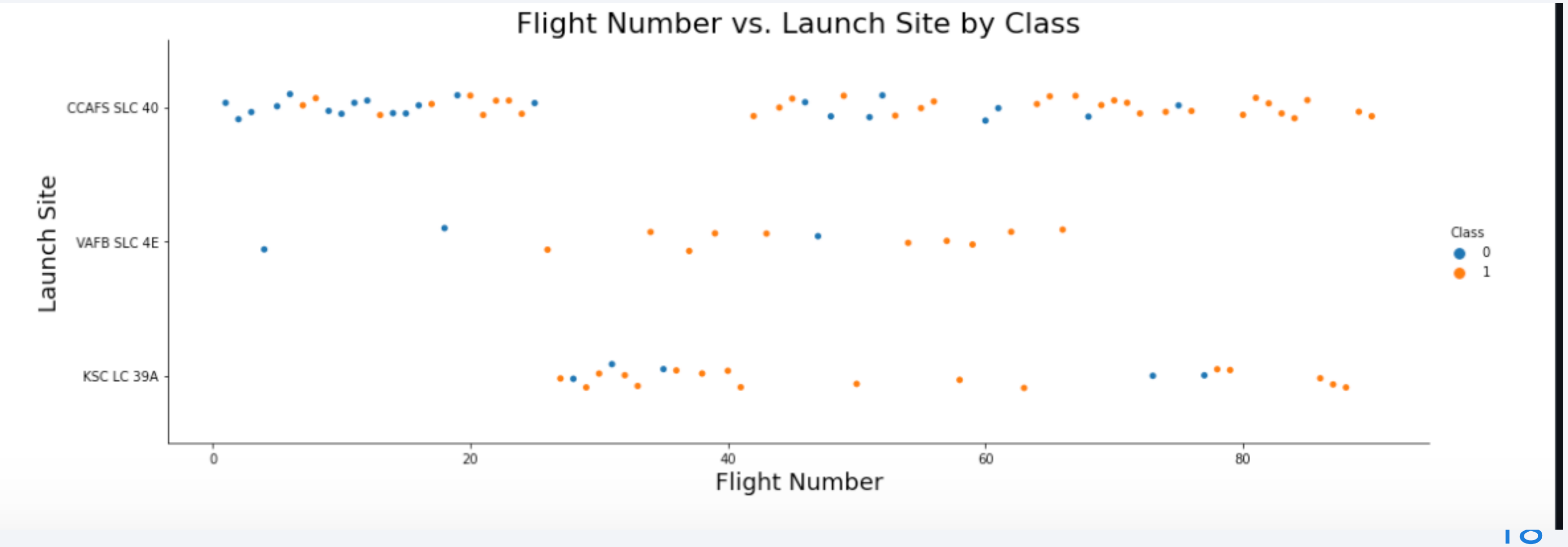


Section 2

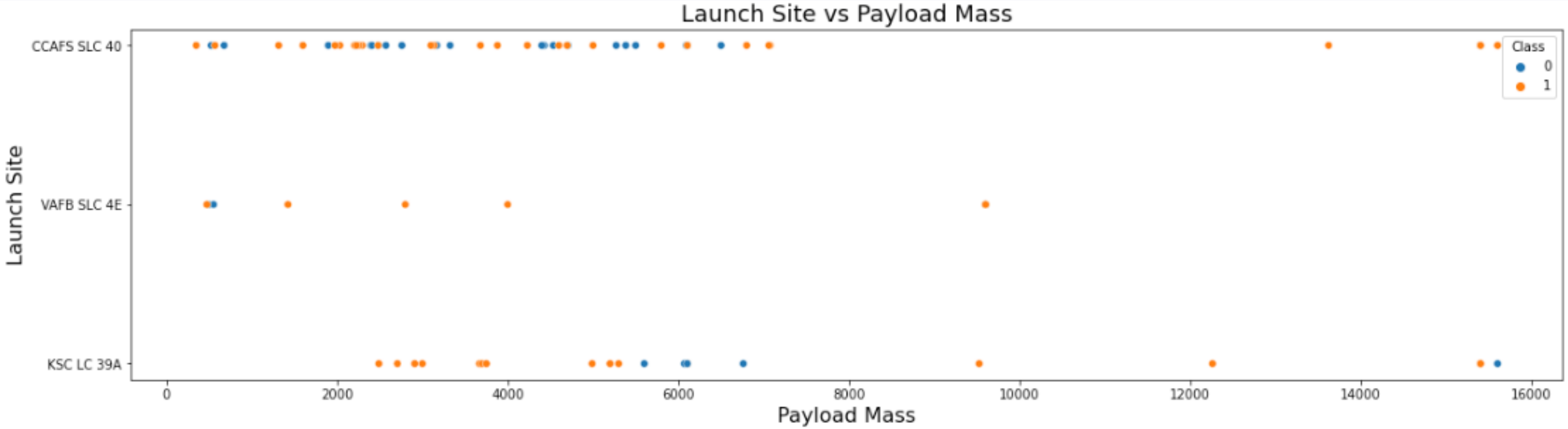
Insights drawn from EDA

Flight Number vs. Launch Site

- KSC LC 39A has a higher success rate
- CCAFS has launched the most rockets



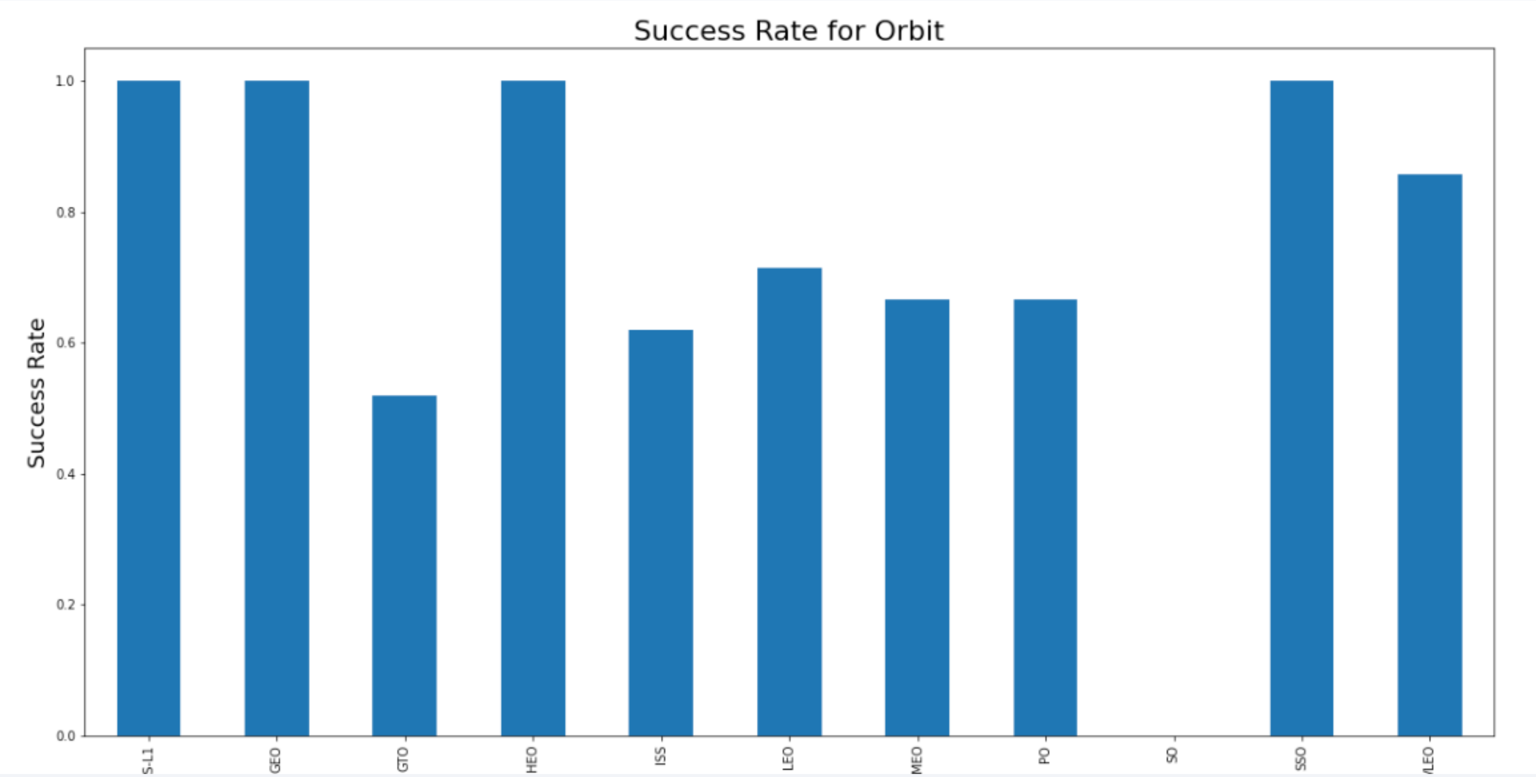
Payload vs. Launch Site



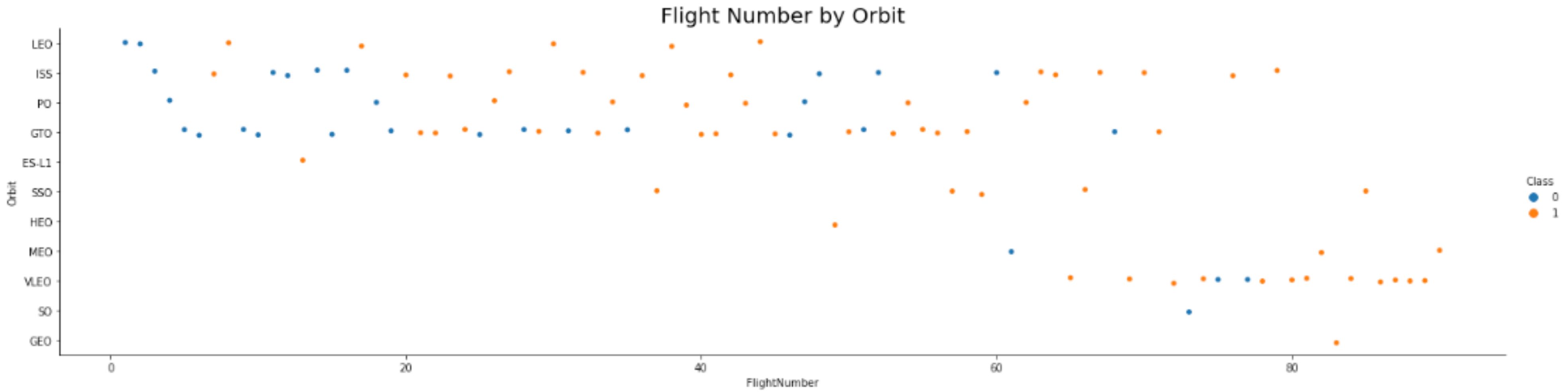
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

- GEO, HEO, and SSO have the highest orbit success rate
- SO has the lowest, with GTO second lowest

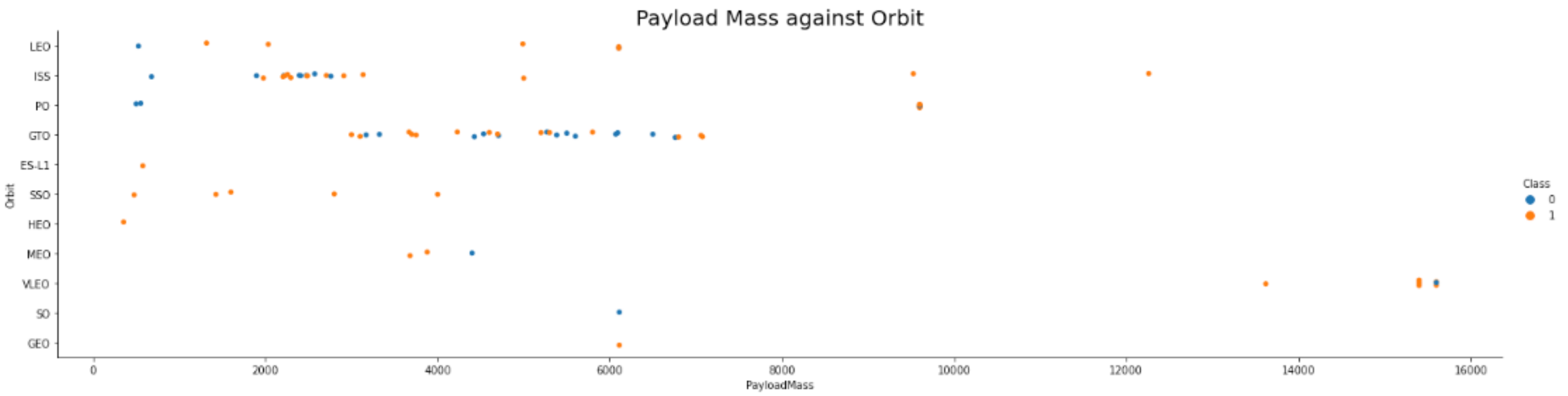


Flight Number vs. Orbit Type



You should see that 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.

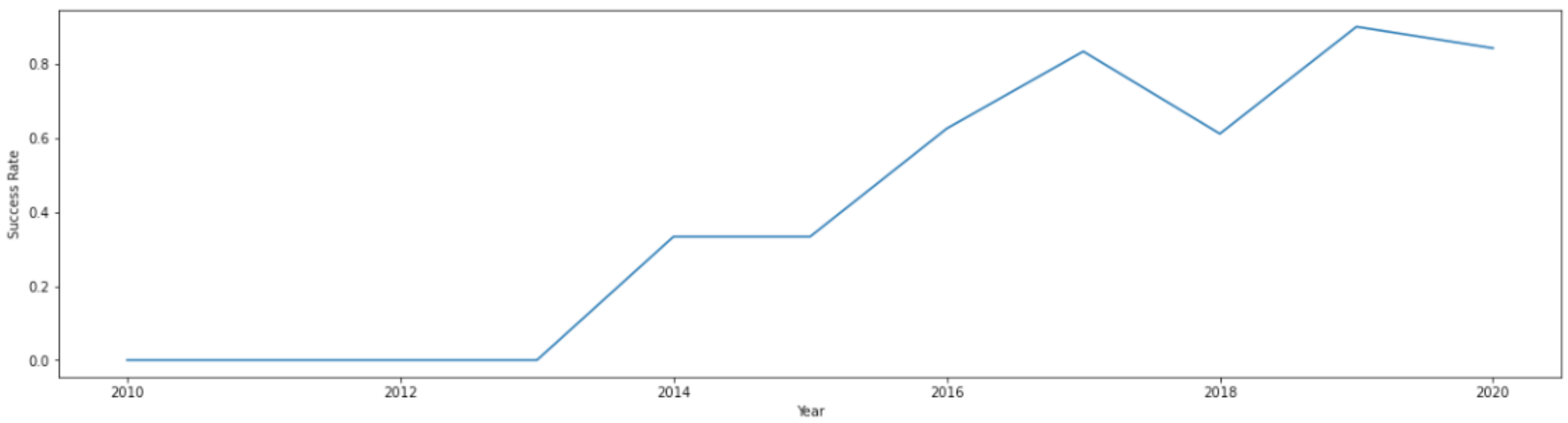
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



You can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

All Launch Site Names

- We see 4 different launch sites

Display the names of the unique launch sites in the space mission

In [9]:

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[9]:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- All the orbits are LEO

In [11]:

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

```
* sqlite:///my_data1.db
Done.
```

Out[11]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO
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)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)

Total Payload Mass

- Total Mass is 45,596 KG

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS total_payload_mass, Customer FROM 'SPACEXTBL'
```

```
* sqlite:///my_data1.db
Done.
```

total_payload_mass	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

- Avg Payload mass here is 2535 KG

Task 4

Display average payload mass carried by booster version F9 v1.1

[13]:

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS average_payload_mass FROM 'SPACEXTBL' WHERE B
```

```
 * sqlite:///my_data1.db
Done.
```

[13]:

average_payload_mass
2534.6666666666665

First Successful Ground Landing Date

- First ground landing is on December 22, 2015

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
]: #%sql SELECT MIN(DATE) FROM 'SPACEXTBL' WHERE
   %sql select min(date) as DATE from 'SPACEXTBL' where landing_outcome like '%ground
```

```
 * sqlite:///my_data1.db
Done.
```

]:

DATE
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql SELECT DISTINCT Booster_Version FROM SPACEXTBL WHERE Mission_Outcome = 'Success' AND PAYLOAD_MASS__KG_ > 4
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

- There are 98 Success and 1 Failure

List the total number of successful and failure mission outcomes

34]:
```
%sql SELECT Mission_Outcome, COUNT(Mission_Outcome) as Total FROM SPACEXTBL GROUP BY Mission_Outcome;
```

```
 * sqlite:///my_data1.db
Done.
```

34]:

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
]: %sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXT
```

```
* sqlite:///my_data1.db
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

2015 Launch Records

```
In [38]: %sql SELECT DATE, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE substr(DATE, 1, 4) = '2015' AND LANDING__OUT

 * sqlite:///my_data1.db
(sqlite3.OperationalError) no such column: LANDING__OUTCOME
[SQL: SELECT DATE, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE substr(DATE, 1, 4) = '2015' AND LANDING__OUT
COME = 'Failure (drone ship)']
(Background on this error at: https://sqlalche.me/e/20/e3q8)

In [ ]:
```

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

```
%sql SELECT LANDING__OUTCOME, COUNT(LANDING__OUTCOME) AS Landing_Count FROM SPACEXTBL WHERE substr(DATE, 0, 5) >
```

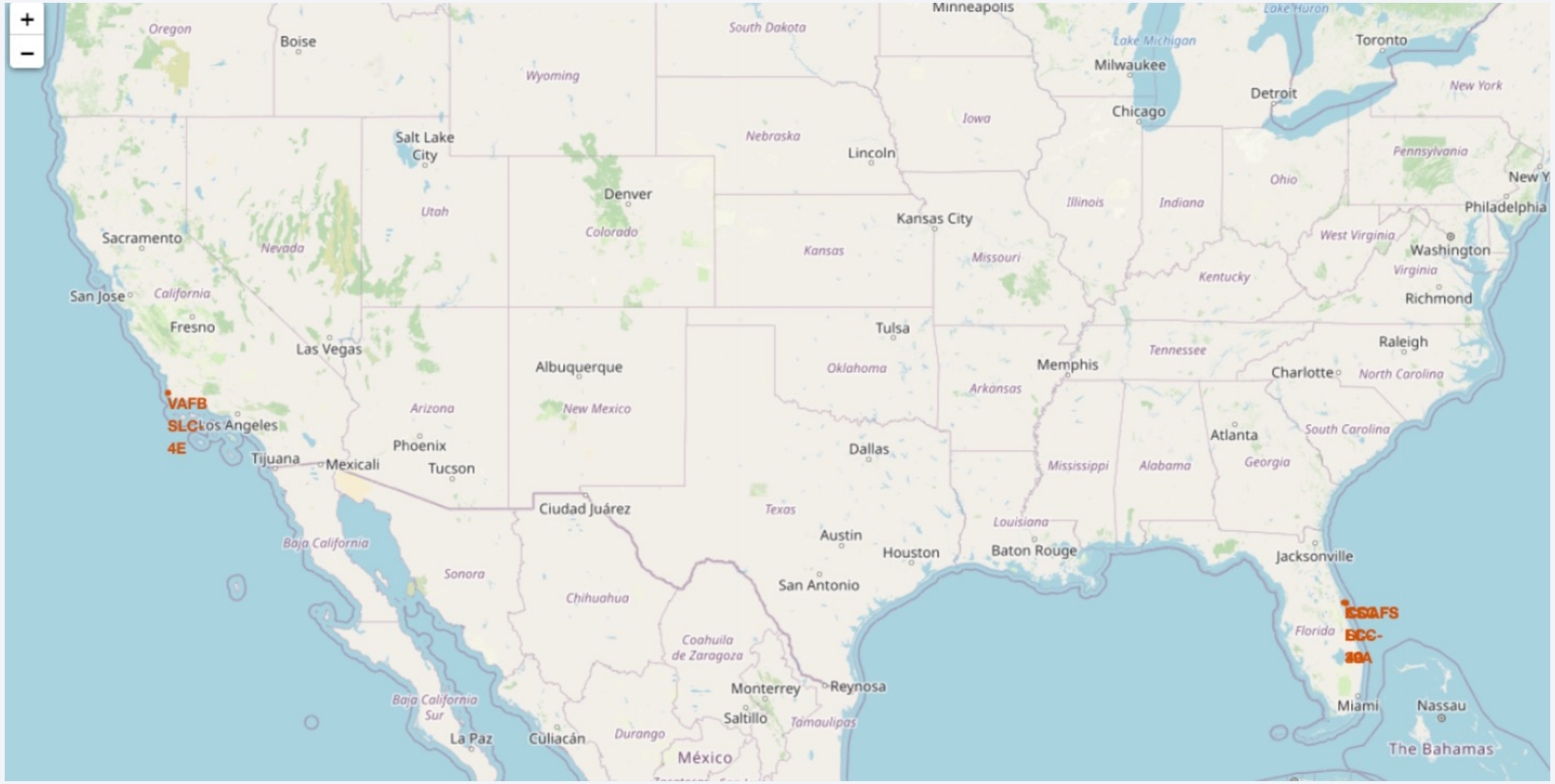
```
 * sqlite:///my_data1.db
(sqlite3.OperationalError) no such column: LANDING__OUTCOME
[SQL: SELECT LANDING__OUTCOME, COUNT(LANDING__OUTCOME) AS Landing_Count FROM SPACEXTBL WHERE substr(DATE, 0, 5) >
= '2010' AND substr(DATE, 0, 5) <= '2017' AND substr(DATE, 6, 2) >= '06' AND substr(DATE, 6, 2) <= '03' GROUP BY
LANDING__OUTCOME ORDER BY Landing_Count DESC]
(Background on this error at: https://sqlalche.me/e/20/e3q8)
```

Section 3

Launch Sites Proximities Analysis

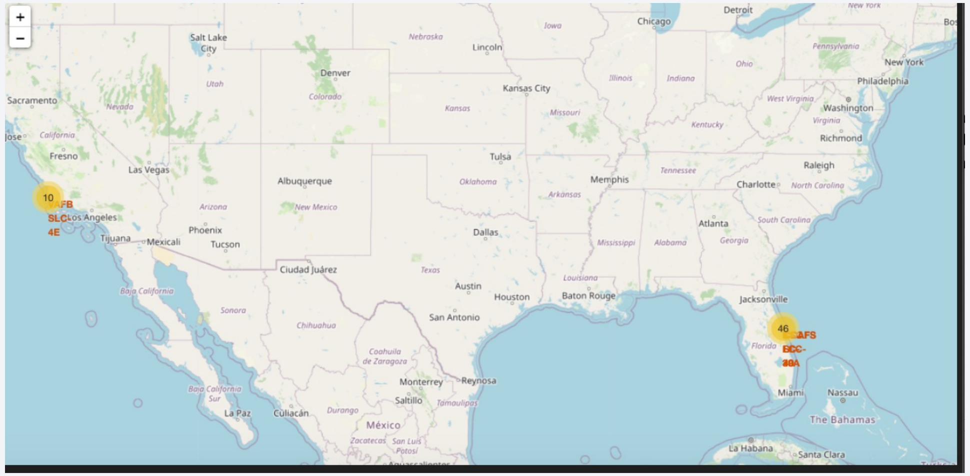
USA Launch Sites

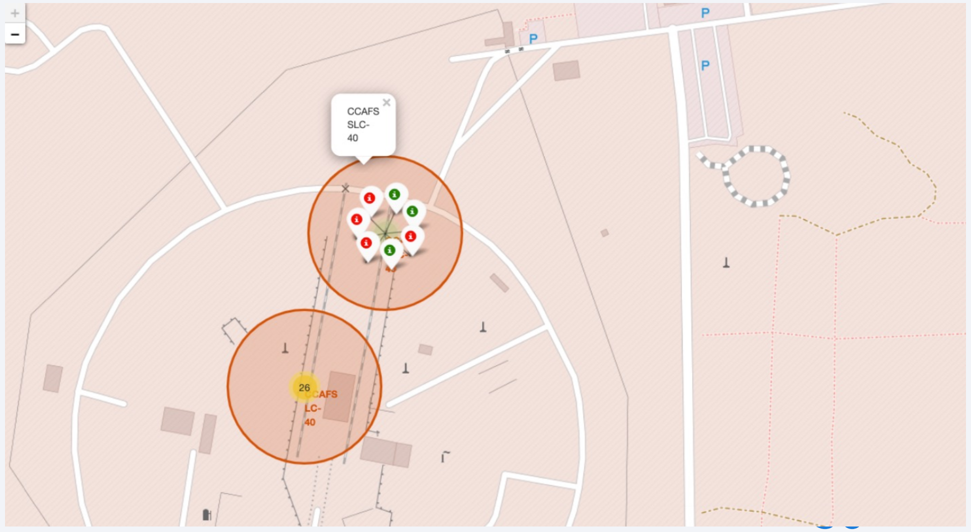
- We find rocket launch bases on the coast
- And near equator



Other Launch Sites

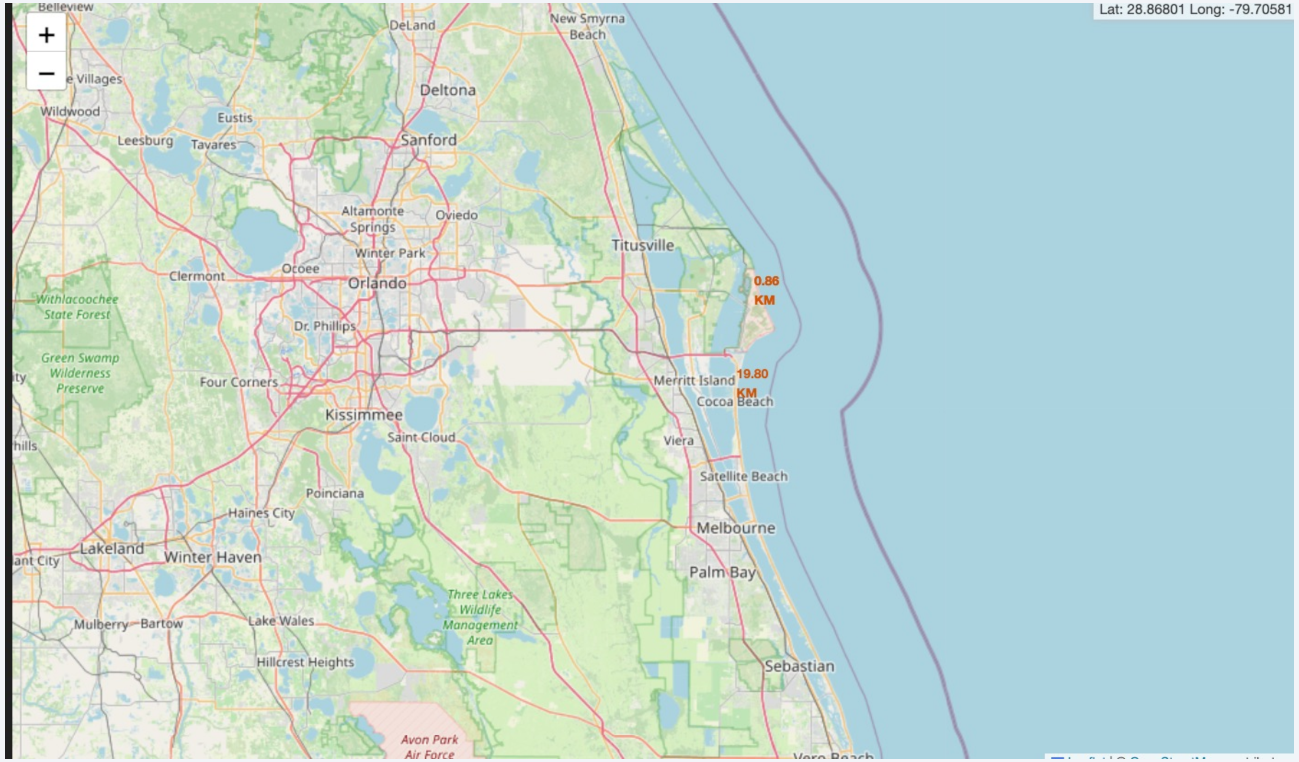
- These clusters can help us easily identify successes





Launch Site Proximity

- For example we can see that cape Canaveral is 19.8 meters from coast.
- Useful tool to see how close launch site is from certain areas such as highways/ coast

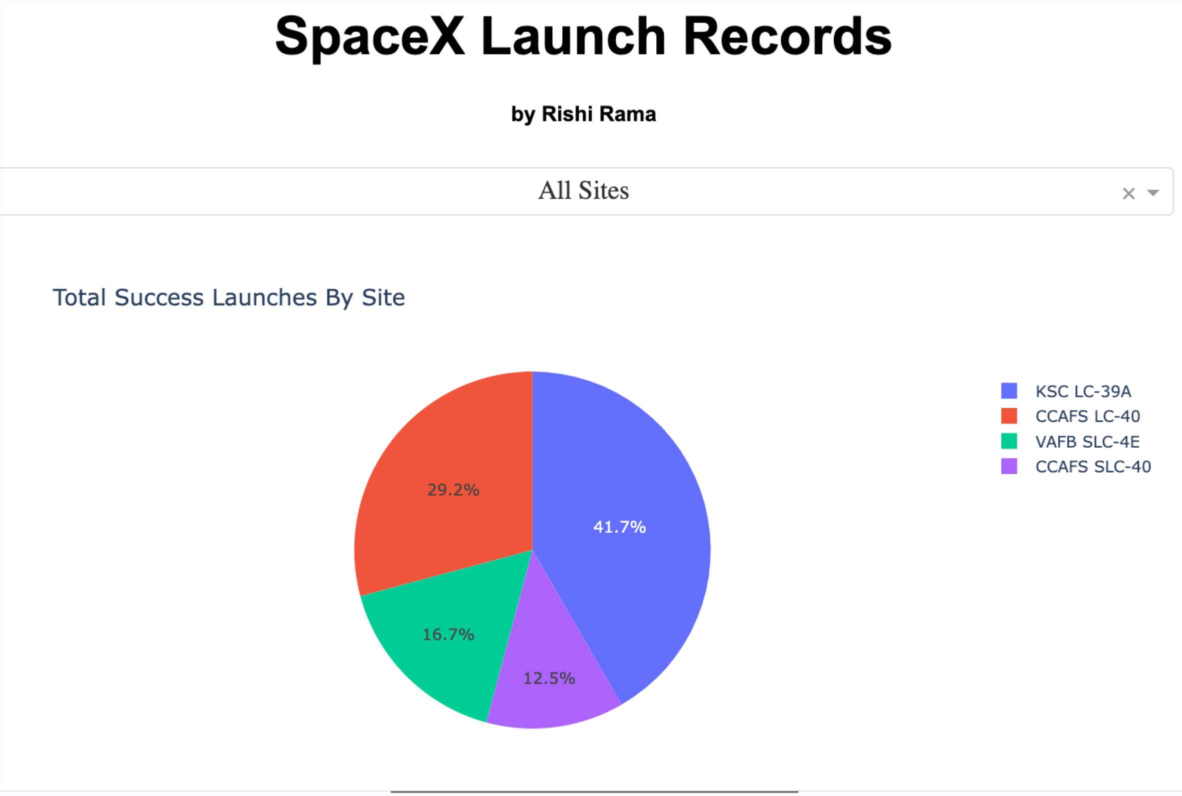


Section 4

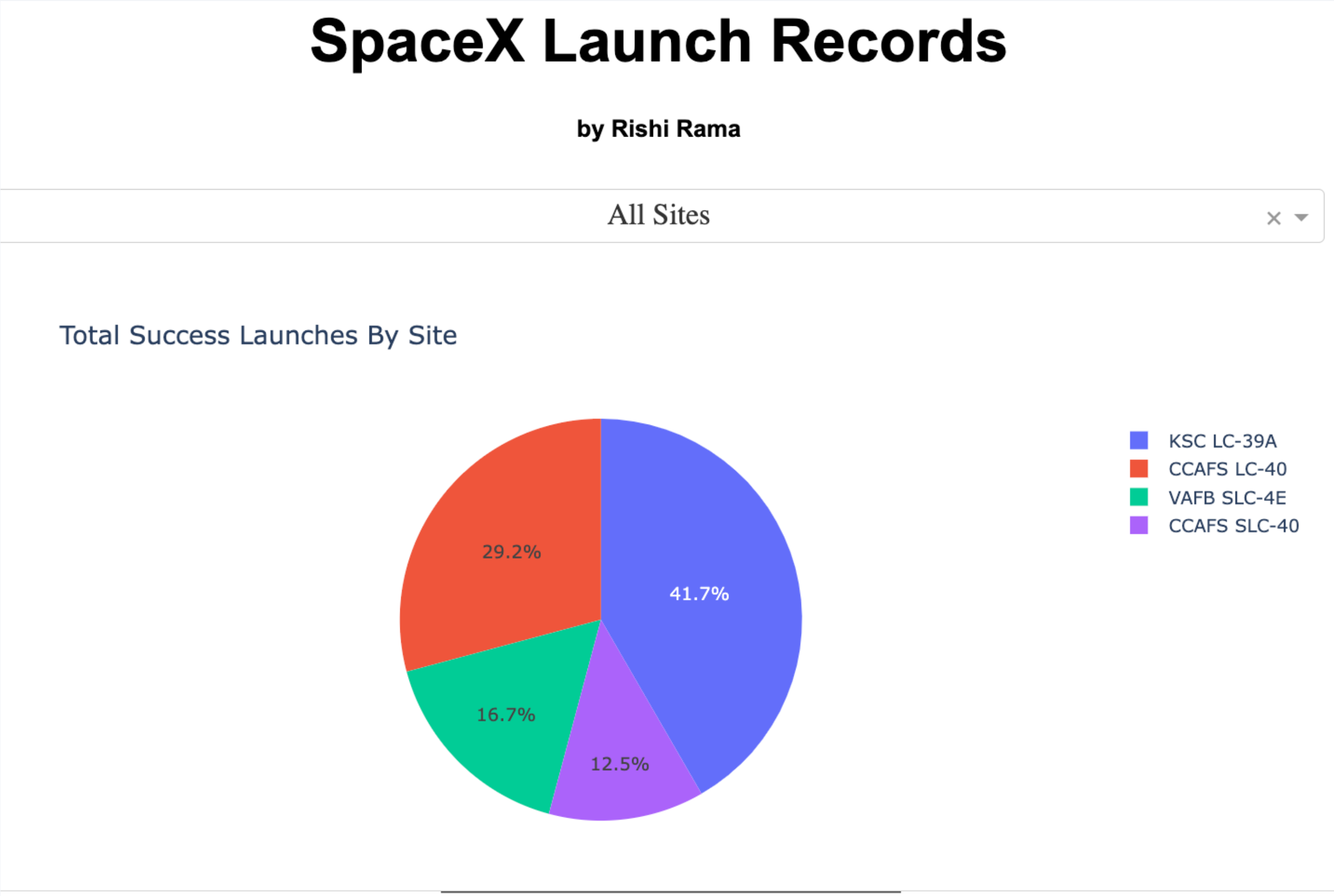
Build a Dashboard with Plotly Dash

All Sites Launch

- KSC leads the way with 42% while VAFB is last with 17%



Highest Launch Success



Payload Slider Graph

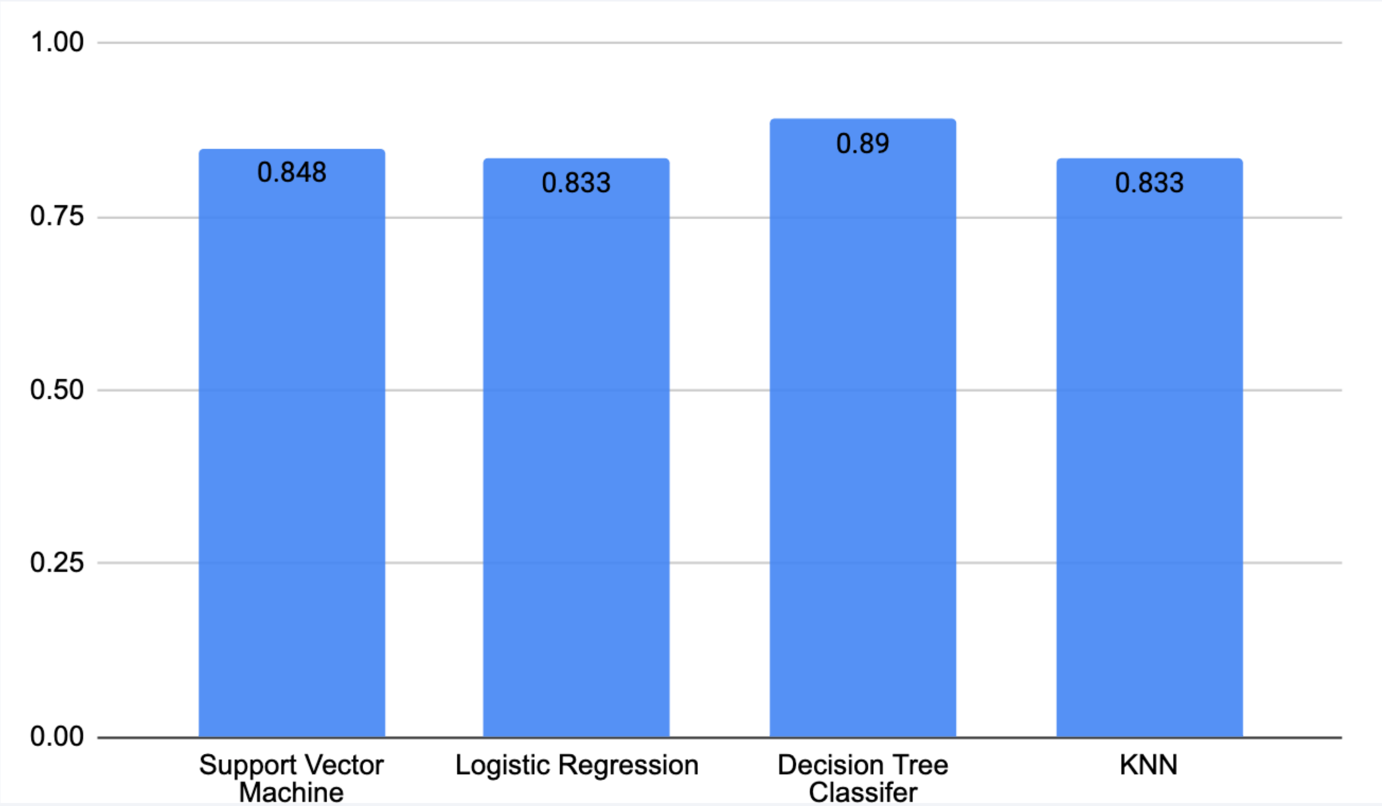
- 2500 to 5000 has good payload success
- FT has the highest rate of success

Section 5

Predictive Analysis (Classification)

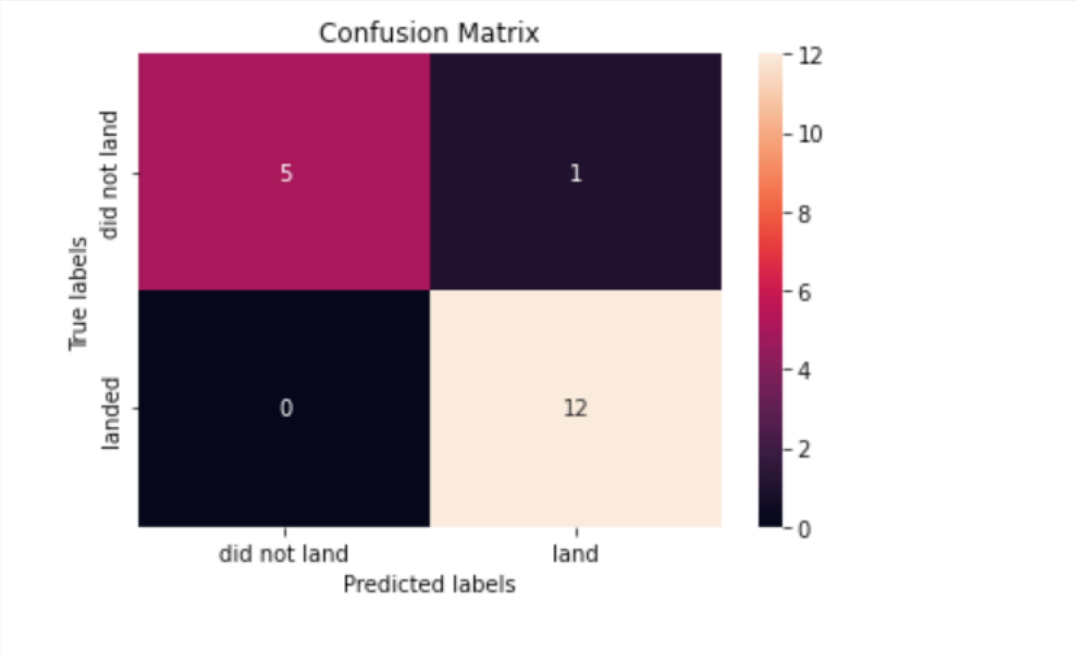
Classification Accuracy

- Decision Tree Classifier has the highest model Accuracy of around 0.89



Confusion Matrix

- The matrix shows us that we will have 5 TP, 1 False Negatives, 0 False Positives, and 12 true Negatives



Conclusions

- There has clearly been a positive trend of success since 2013 till 2020
- For all site's, KSLA 39A has the highest launch success rate.
- GEO, HEO, and SSO have the highest orbit success rate
- The size of the payload clearly matters. The heavier the initial payload the lower chance of success that stage one returns.
- With an accuracy of 89%, the decision tree classifier was the best prediction model given the parameters.

Appendix

- https://github.com/puligoat11/SpaceX-Falcon-9-Final-Project-

Thank you!