IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sharlmagne Henry
May 4, 2022

Outline

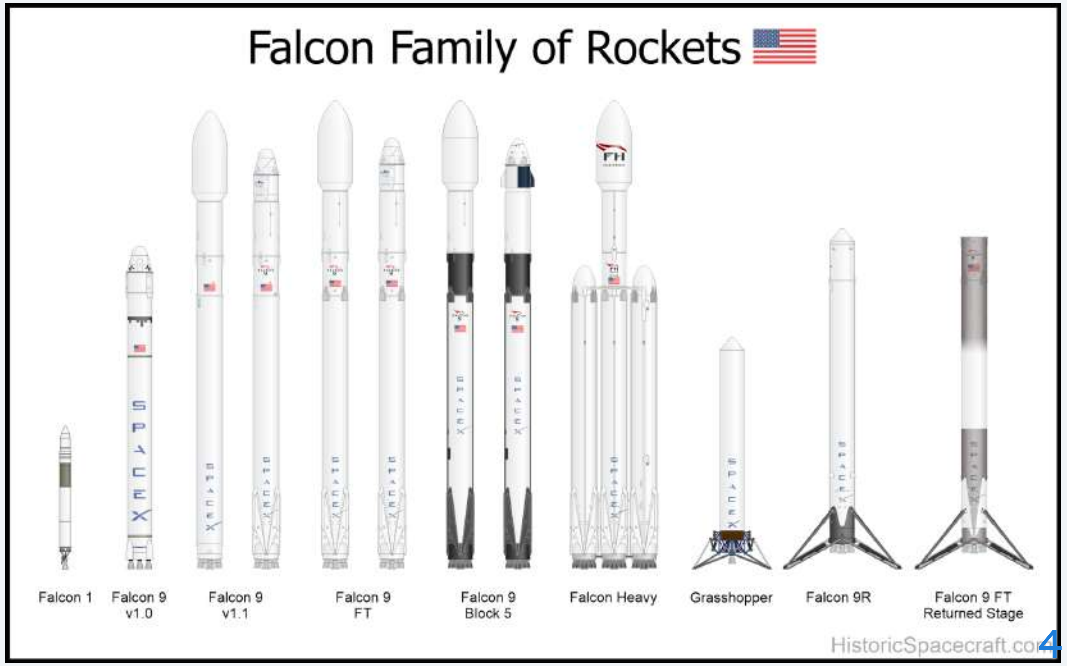
- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies
  - Data was collected through the SpaceX open source REST API
  - Falcon 9 launches were used as the premise for this exercise
  - Historical information was scraped from Wikipedia for Falcon 9 launches
  - Data wrangling, Explanatory Data Analysis (EDA) and data visualization was done to the data
  - Machine Learning predictions done using Logistic Regression, Support Vector Machine (SVM), Decision Tree Classifier and k Nearest Neighbors.
- Summary of all results
  - Success is dependent on the number of flight, launch site and the mass of the booster in some cases.
  - Success rate improved over the period of 2013 to 2020

Introduction

- SpaceY would like to position itself as an alterative to SpaceX by effectively and purposefully competing against their 62 million dollar rocket launches.
- The objective is to predict if the Falcon 9 first stage will land successfully and ultimately use that information to lower the cost of rocket launches.
  - Where are the best locations to launch from
  - How size affect launches



Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from
    - SpaceX REST API
    - Wikipedia (Historical records for Falcon 9 Launches)
- Perform data wrangling
  - A outcome field was constructed to determine a definite failure or success situation.

Methodology

Executive Summary - continued

- 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 was standardized
  - Split into training and test
  - Use KNN, Logistic Reg, SVM and Decision Tree Classifier
  - Determine the accuracy of the models

Data Collection

- Data was collected from
  - SpaceX REST API
    - https://api.spacexdata.com/v4/launches/past
  - Historical records for Falcon 9 Launches
    - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

SpaceX API
- Get request

Create Database
- Transform into Pandas DB
- Filter for Falcon 9 Launches

Handle Missing Values
- Replace missing values with mean

Source – GitHub

https://github.com/Sharlmagne/ibm-data-science/blob/f80040190cedb7f231d66aa3fecdaef7735e89cf/Data%20Collection%20API.ipynb

Data Collection – Scraping

Request HTML Table	• Get request Wiki page with falcon 9 launch history
Create Database	• Use HTML table headers to create columns in new database
Populate dataframe	• Add to dataframe by parsing HTML tables

Source – GitHub

https://github.com/Sharlmagne/ibm-data-science/blob/master/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

Launches per Launch Site	• Calculate the number of launches on each site
Launches per orbit	• Number of launches for every orbit types
Outcome Types	• Determine the number of outcomes per outcome type
Classify Outcomes	• Organize each of the 8 outcome types into a class of either a failure or success

Source – GitHub

https://github.com/Sharlmagne/ibm-data-science/blob/master/Data%20Wrangling.ipynb

EDA with Data Visualization

- A combination of Line, Bar and Scatter plot was done to analyze the data
  - Flights were plotted against Pay Load Mass, Launch Site and Orbit (Scatter Plot)
  - Year vs Success Rate (Line Graph)
  - Orbit vs Success Rate (Bar Graph)
  - Pay Load Mass vs Orbit (Scatter Plot)

Source – GitHub

https://github.com/Sharlmagne/ibm-data-science/blob/master/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

List of all the SQL queries performed on data

1. *Display the names of the unique launch sites in the space mission¶*
2. *Display 5 records where launch sites begin with the string 'CCA'*
3. *Display the total payload mass carried by boosters launched by NASA (CRS)*
4. *Display average payload mass carried by booster version F9 v1.1*
5. List the date when the first successful landing outcome in ground pad was achieved.
6. List the names if the boosters which have success in drone ship and payload mass greater than 4000 but less than 6000
7. List the total number of successful and failure mission outcomes
8. List the names of the booster versions which have carried the maximum payload mass
9. List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
10. Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order

Source – GitHub - https://github.com/Sharlmagne/ibm-data-science/blob/master/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Locations of all the launch sites were located on map using a circle
  - latitude and longitude coordinates were uses to determine exact locations
- Failed and successful launches were indicated by red and green markers respectively
- Distances between a launch site to its proximities were represented by lines
  - Proximities include highways, railways, coastlines and cities
  - This was done to determine the relative closeness among launch sites and their proximities.

Source – GitHub - https://github.com/SharImagne/ibm-data-science/blob/master/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- Plots done to visualize the data
  - Success count for all launch sites
  - Success count on payload mass for all sites

With the plot above one could determine the the most successful launch sites and how the weight affect it.

Source – GitHub - https://github.com/Sharlmagne/ibm-data-science/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- The classification models used for prediction
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree Classifier
  - k Nearest Neighbors

All models performed similarly with the test sample

Prep Data
- Load data in dataframe
- Standardize data to use with models

Train/Test split
- Train size – 80%
- Test size 20%

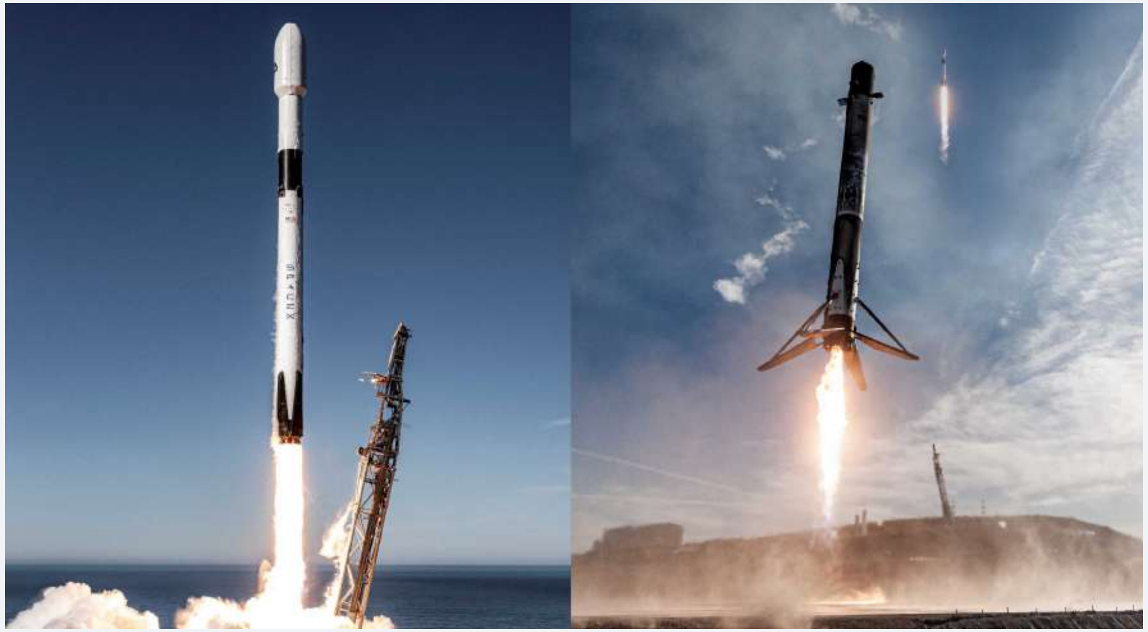
Run Model and determine best Hyperparameters

Determine the accuracy of the models

Source – GitHub - https://github.com/Sharlmagne/ibm-data-science/blob/master/Machine%20Learning%20Prediction.ipynb

Results

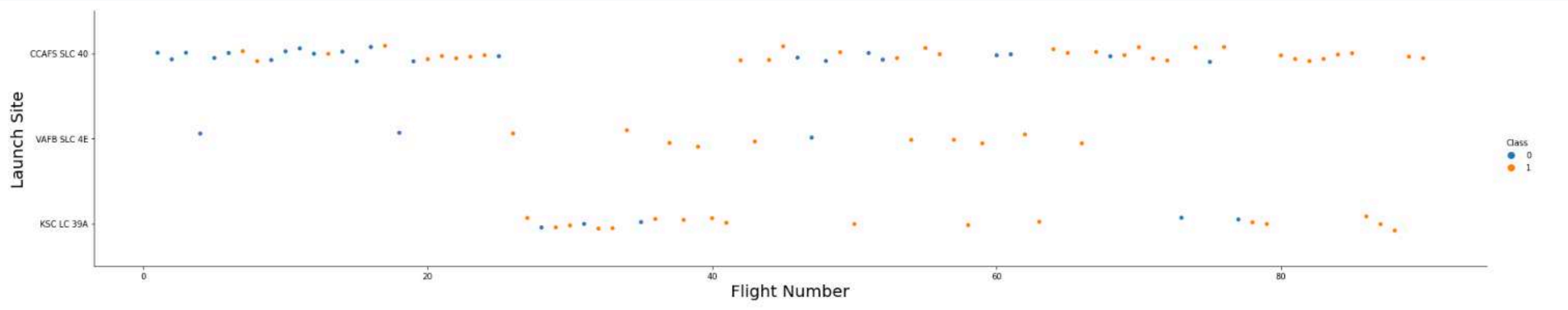
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



Section 2

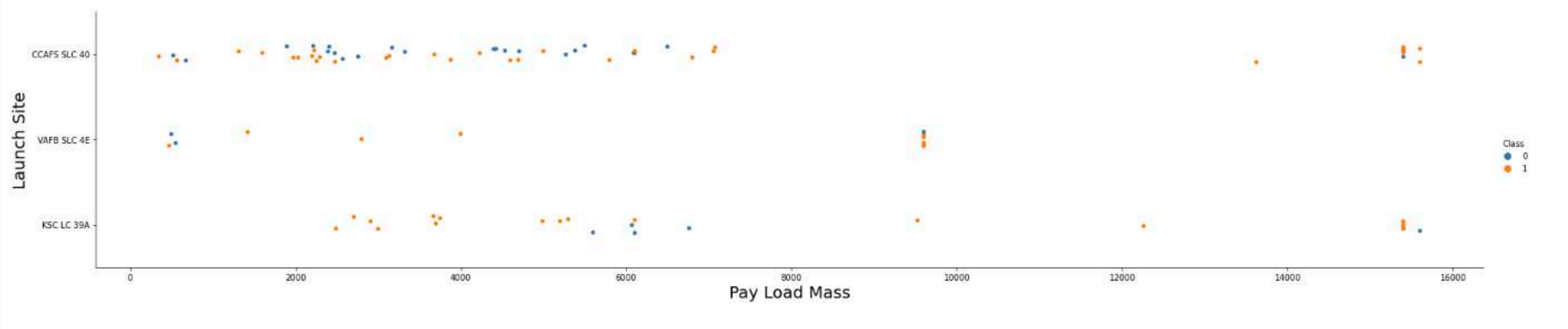
Insights drawn from EDA

Flight Number vs. Launch Site



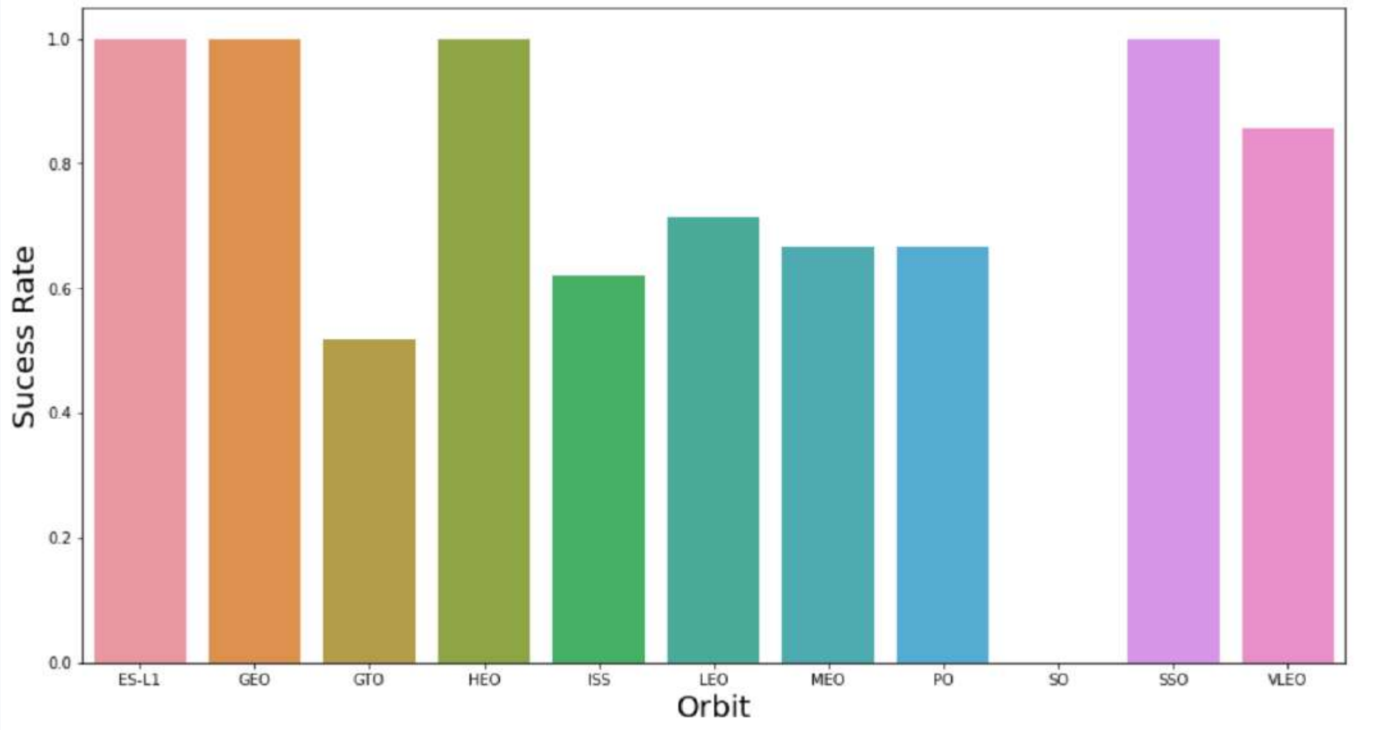
- As the number of flights increases the number of successes increase

Payload vs. Launch Site



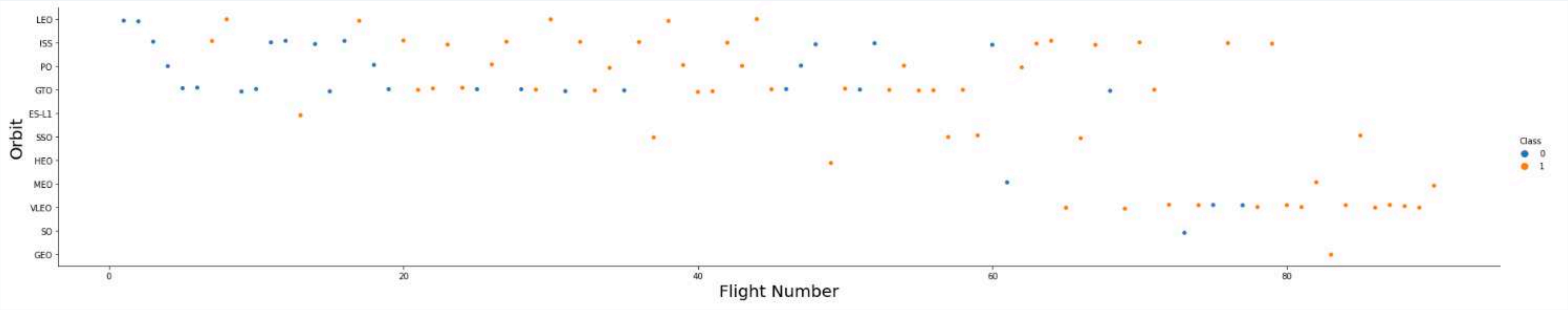
- We see that different launch sites have different success rates.
- CCAF LC-40 has a success rate of 60%
- KSC LC-39A and VAFB SLC 4E has a success rate of 77%

Success Rate vs. Orbit Type



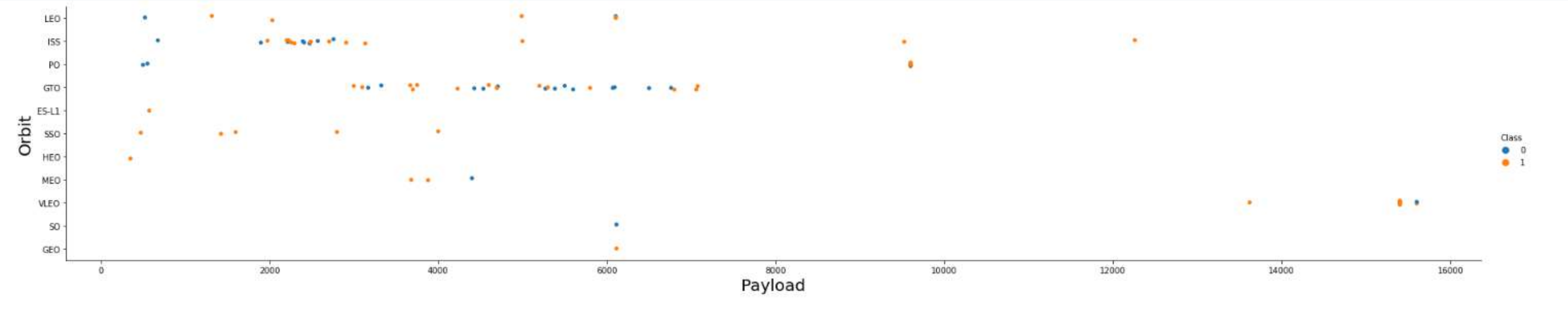
- ES-L1, GEO, HEO AND SSO all had successful rate of 100%
- GTO had the lowest just above 50% followed by ISS around 60%

Flight Number vs. Orbit Type



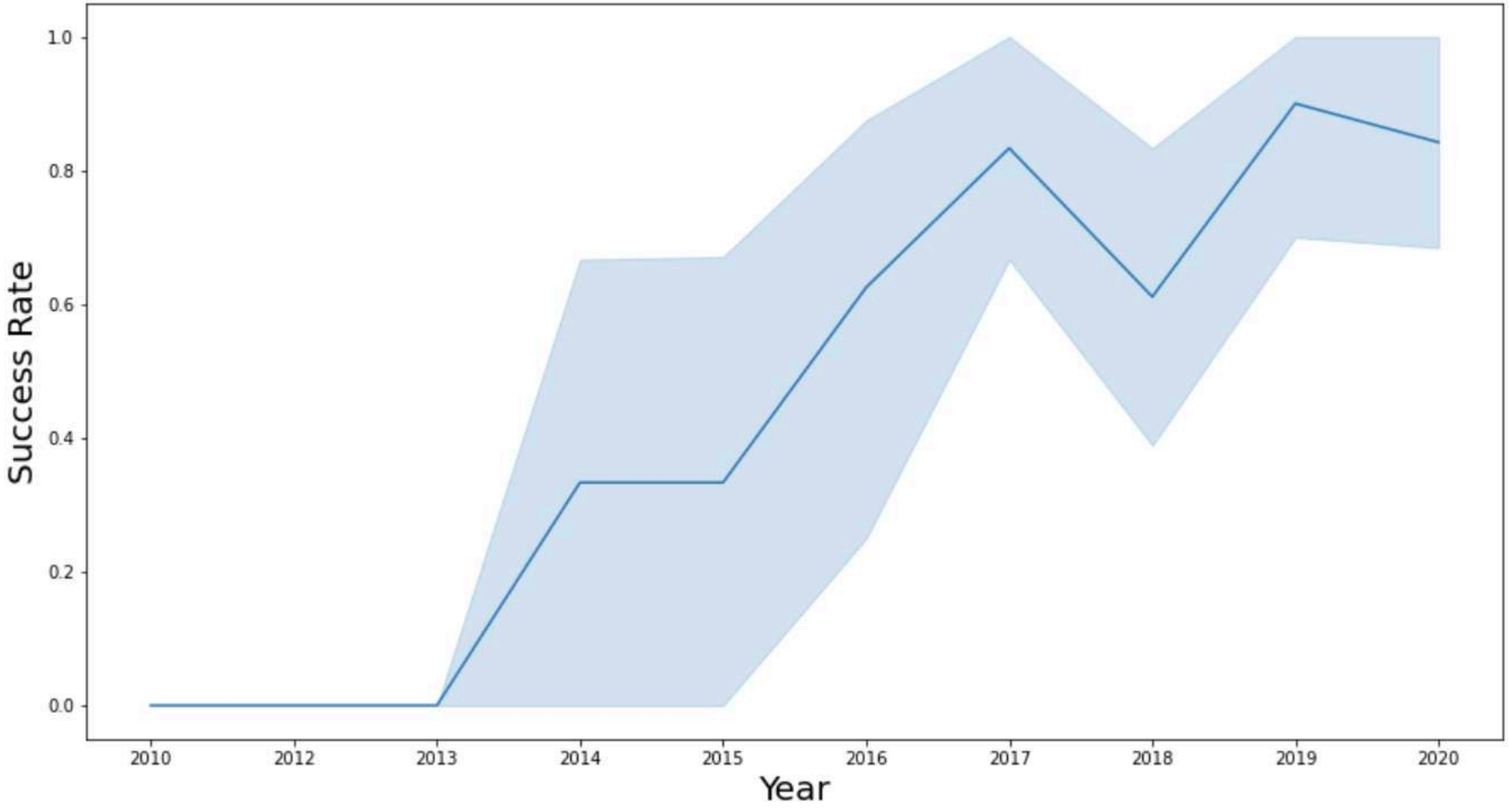
- LEO orbit successes appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit

Payload vs. Orbit Type



- With heavy payloads the successful landing are more for Polar, LEO and ISS. However for GTO we cannot distinguish the well as both positive landing rate and negative landing rate are both present among the different payload masses

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

```
%sql select distinct(launch_site) from SPACEXDATASET;
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

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

- Total of five distinct launch sites

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

```
%sql select * from SPACEXDATASET where launch_site like 'CCA%' limit 5;
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

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

- Using the like operator `CCA` was isolated to search for all data with it

Total Payload Mass

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

```
%sql select sum(payload_mass__kg_) as payload_sum_for_NASA_CRS from SPACEXDATASET where customer like 'NASA (CRS)';
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

payload_sum_for_nasa_crs
45596

- Total payload of 45,596kg for all the boosters launch by NASA.

Average Payload Mass by F9 v1.1

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

```
%sql select avg(payload_mass__kg_) as average_payload_mass from SPACEXDATASET where booster_version like 'F9 v1.1%';
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

average_payload_mass
2534

- The average payload mass carried by booster version F9 v1.1 was found to be 2534kg

First Successful Ground Landing Date

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

Hint:Use min function

```
%sql select date, launch_site, landing__outcome from SPACEXDATASET where landing__outcome in 'Success (ground pad)' or
der by date;
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

DATE	launch_site	landing__outcome
2015-12-22	CCAFS LC-40	Success (ground pad)
2016-07-18	CCAFS LC-40	Success (ground pad)
2017-02-19	KSC LC-39A	Success (ground pad)
2017-05-01	KSC LC-39A	Success (ground pad)
2017-06-03	KSC LC-39A	Success (ground pad)
2017-08-14	KSC LC-39A	Success (ground pad)
2017-09-07	KSC LC-39A	Success (ground pad)
2017-12-15	CCAFS SLC-40	Success (ground pad)
2018-01-08	CCAFS SLC-40	Success (ground pad)

- First successful landing was on December 22, 2015 from the CCAFS LC-40 Launch site

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 booster_version, payload_mass__kg_, landing__outcome \
from SPACEXDATASET \
where landing__outcome in 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000;
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

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)

- Only four boosters successfully landed on drone ship and had payload mass greater than 4000 but less than 6000.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql select count(mission_outcome) as total_outcomes, mission_outcome from SPACEXDATASET group by mission_outcome;
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0ngnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

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

- The results show a total of one failure with 100 successes. However the payload status was unclear for one of the missions.

Boosters Carried Maximum Payload

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

```
%sql select booster_version, payload_mass__kg_  from SPACEXDATASET where payload_mass__kg_ = (select MAX(payload_mass_
_kg_) from SPACEXDATASET) ;
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

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

- Total of 12 booster which have carried the maximum payload mass.

2015 Launch Records

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

```
%sql select landing__outcome, booster_version, launch_site, date from SPACEXDATASET where year(date) = 2015 and landin
g__outcome in 'Failure (drone ship)';
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

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

- Only two “landing outcomes” resulted in failure for 2015

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

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

```
%sql select count(landing__outcome) as total, landing__outcome \
from SPACEXDATASET \
where date between '2010-06-04' and '2017-03-20' \
group by landing__outcome \
order by total desc;
#having date between '2010-06-04' and '2017-03-20';
```

```
 * ibm_db_sa://yxk13191:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:3124
9/bludb
Done.
```

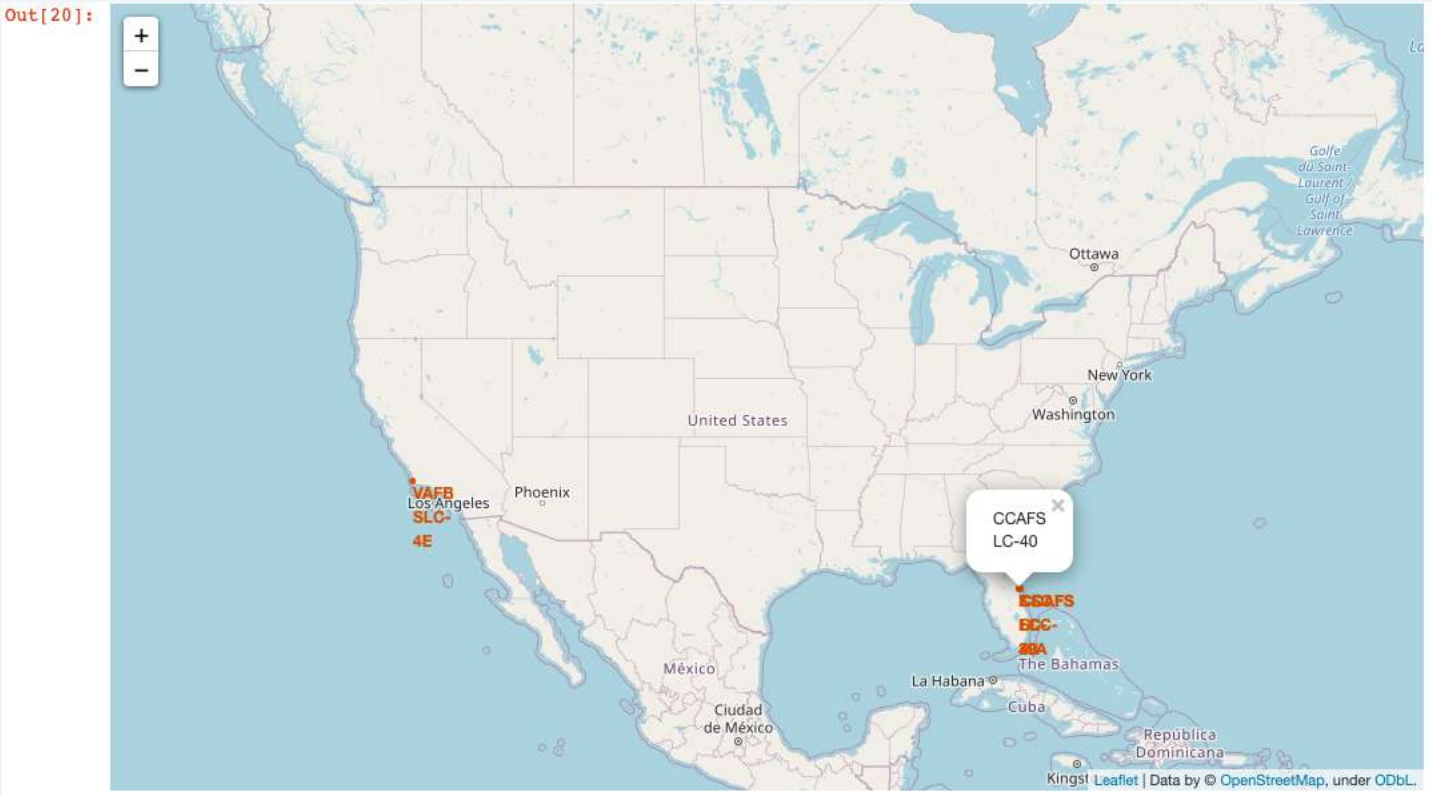
total	landing__outcome
10	No attempt
5	Failure (drone ship)
5	Success (drone ship)
3	Controlled (ocean)
3	Success (ground pad)
2	Failure (parachute)
2	Uncontrolled (ocean)
1	Precluded (drone ship)

- The counts of the different landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order.

Section 3

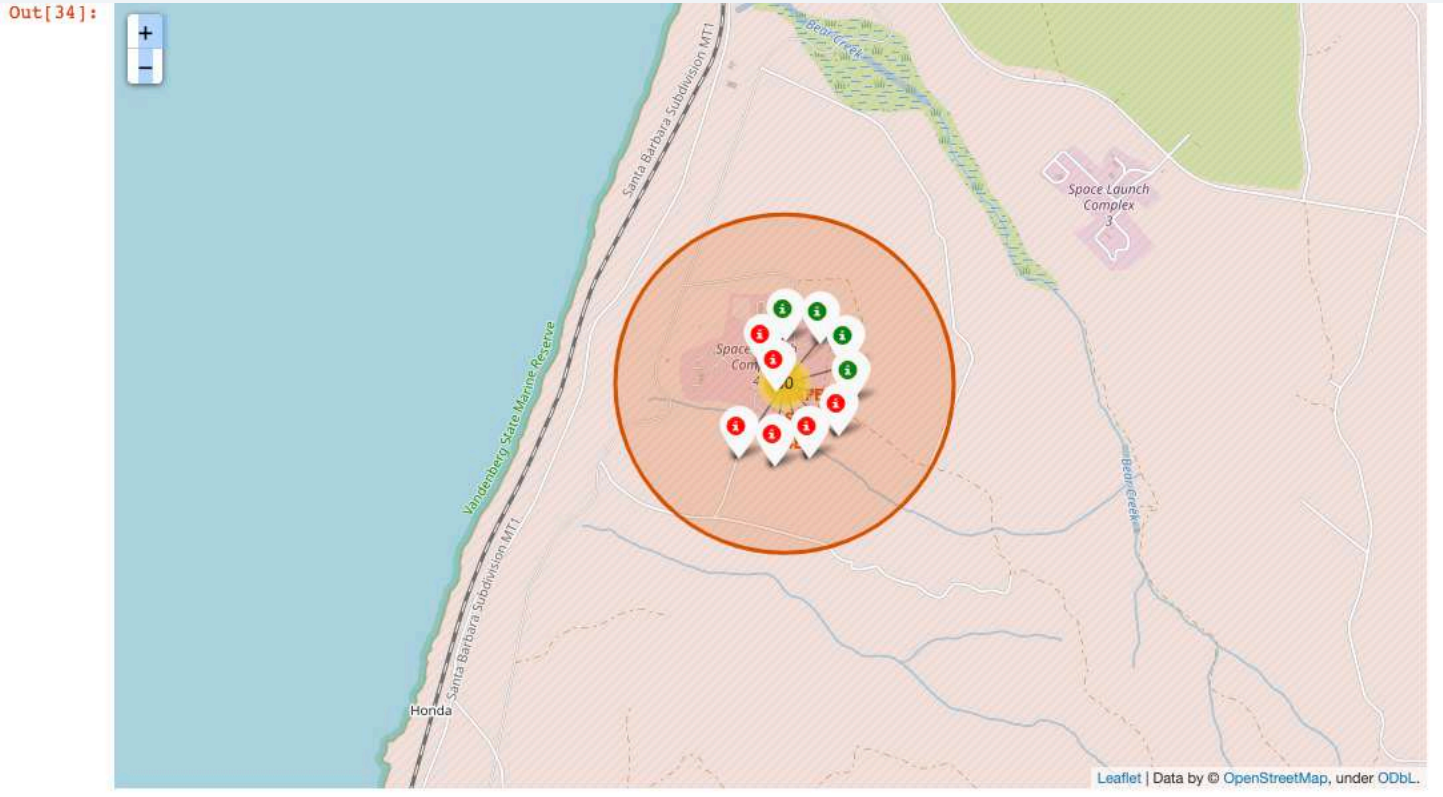
Launch Sites Proximities Analysis

All Launch Sites



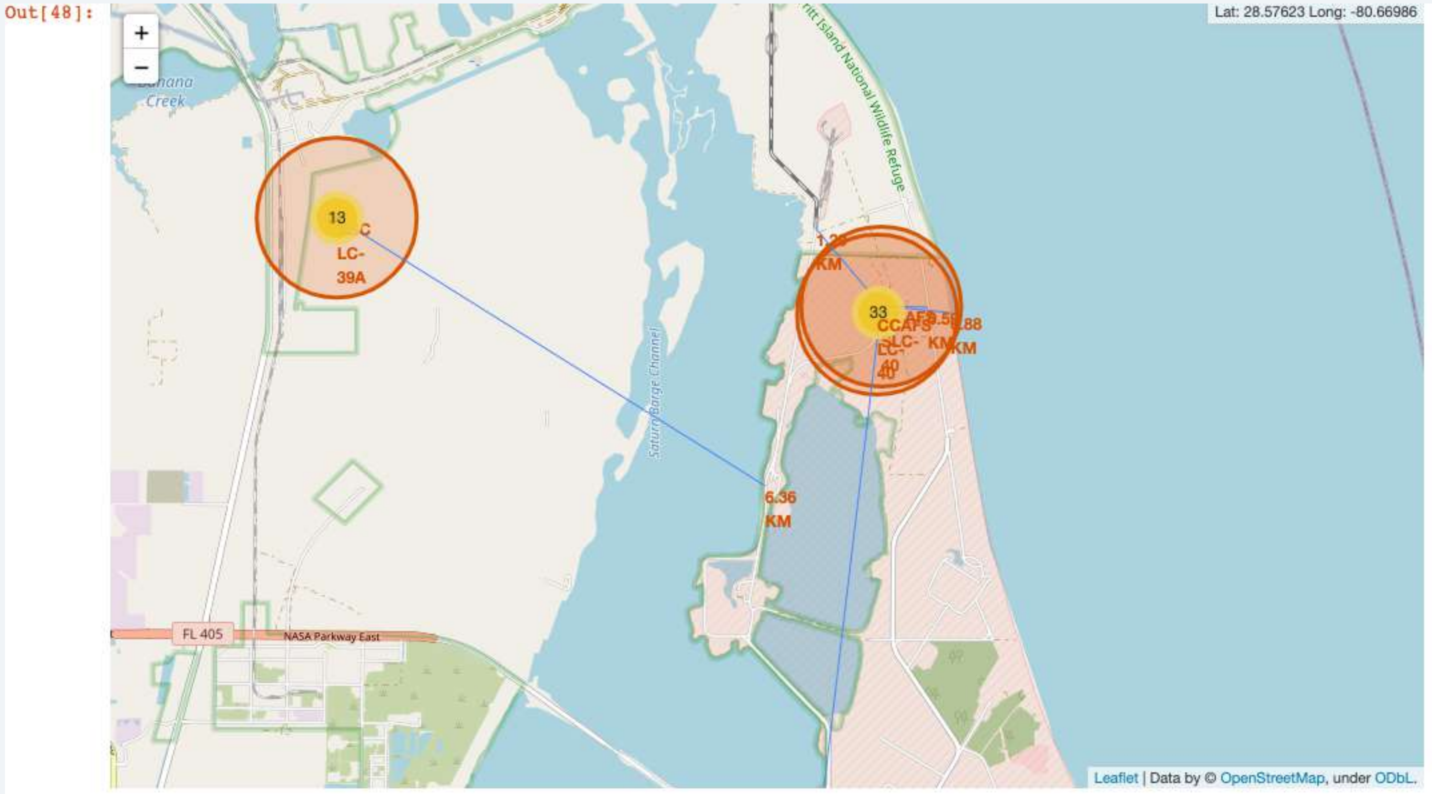
- It is apparent that launch sites are situated closer to the equator and coastlines.

Failure/Success Makers



- This is an example of the failed launches in red and the successful launches in green for the VAFB SLC-AE launch site

Launch site and its proximities

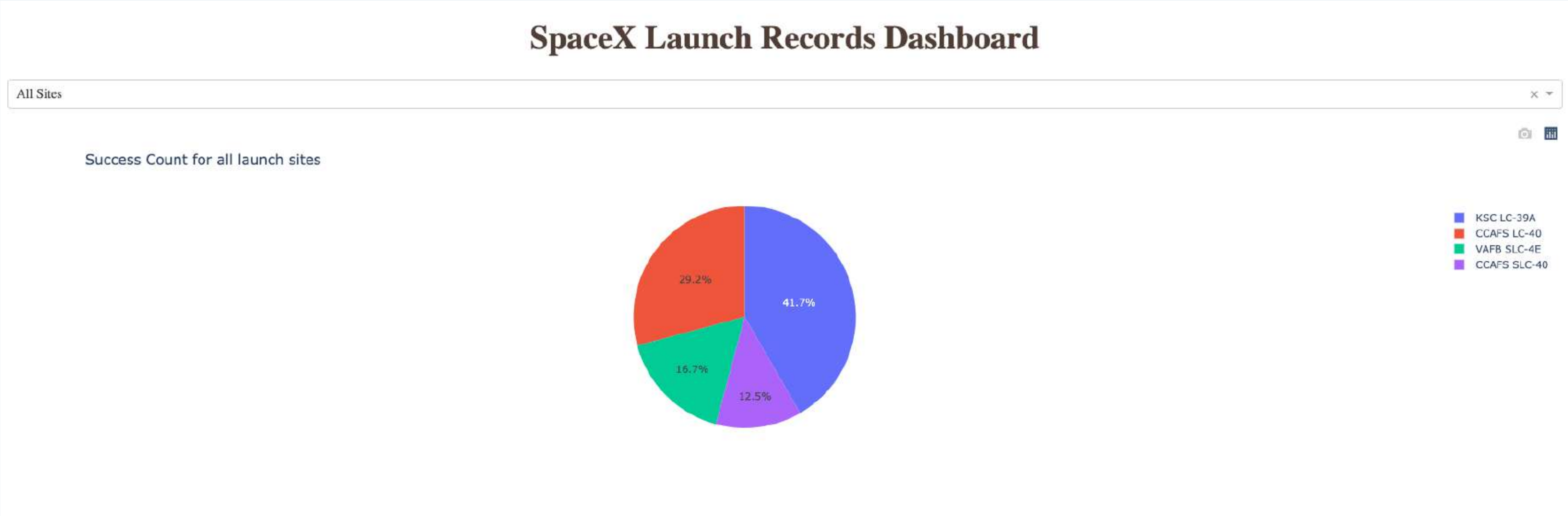


- Launch sites tend to be closer to railways, roads and coastlines for transportation and failures reasons (less than 1.5km). However, it is the opposite city/highly dense regions (over 50km).

Section 4

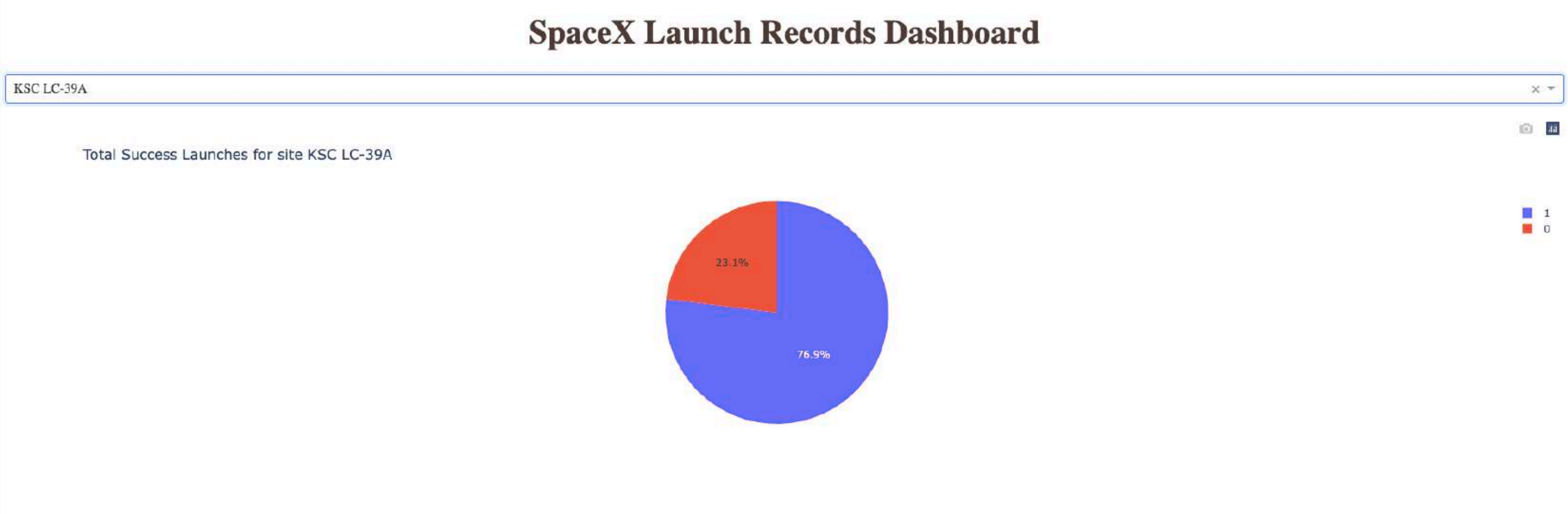
Build a Dashboard with Plotly Dash

Success count ratio among launch sites



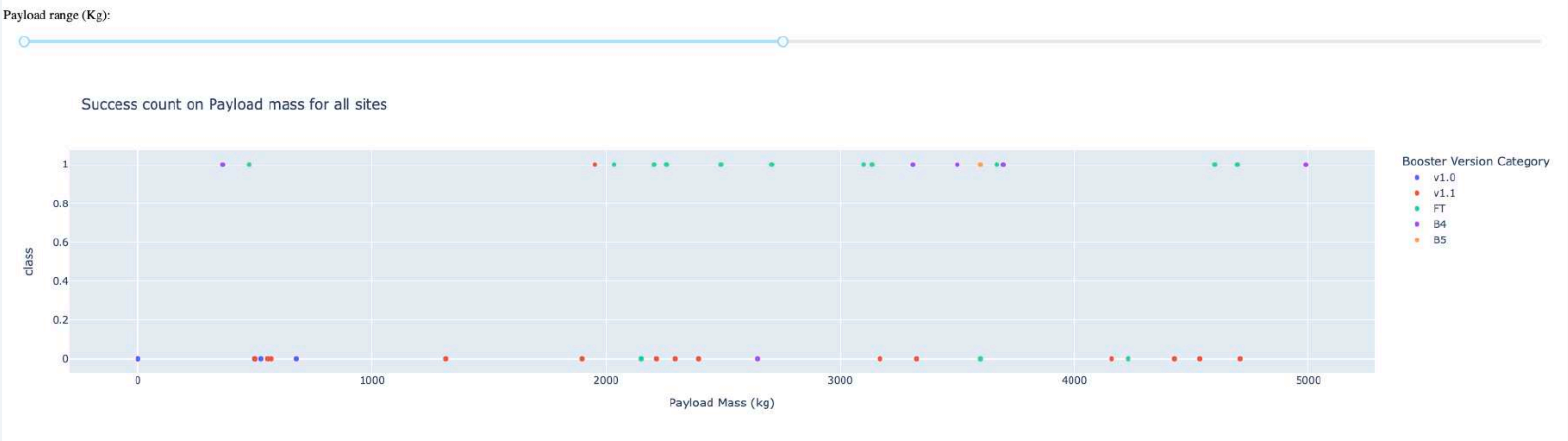
- KSC LC-39A has the most successful launches followed by CCAFS LC-40

Success Launches Rate for KSC LC 39A



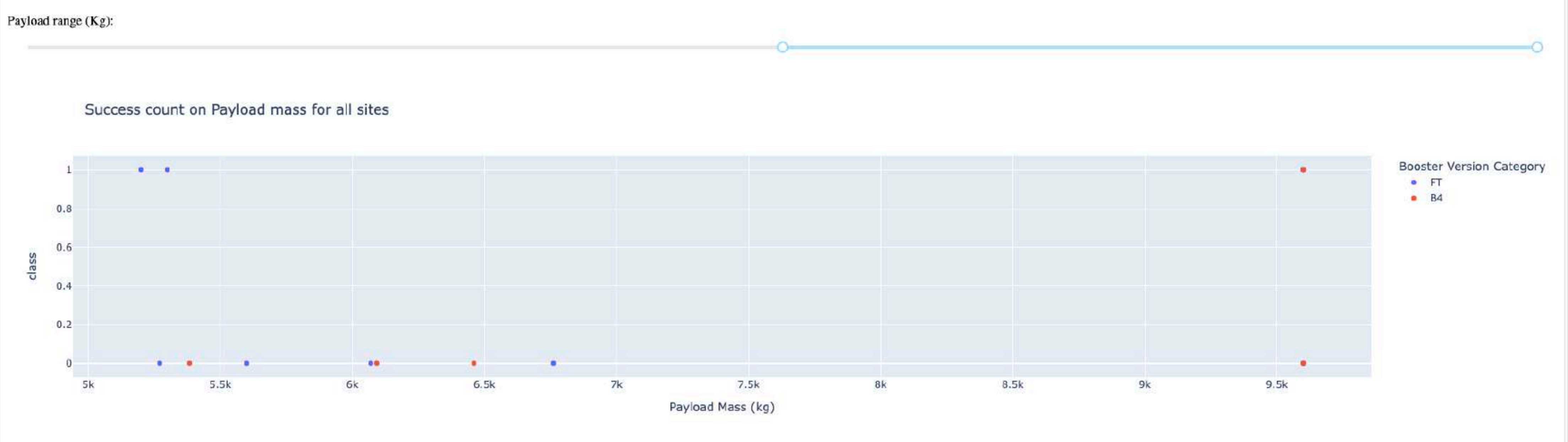
- KSC LC-39A had 76.9% success rate for launches.

Payload vs Launch Outcome



- Payload vs. Launch Outcome scatter plot for all sites, with different payload between 0kg and 5000kg
- According the the plot above “FT booster version” has the most successful launches with payload mass below 5000kg

Payload vs Launch Outcome

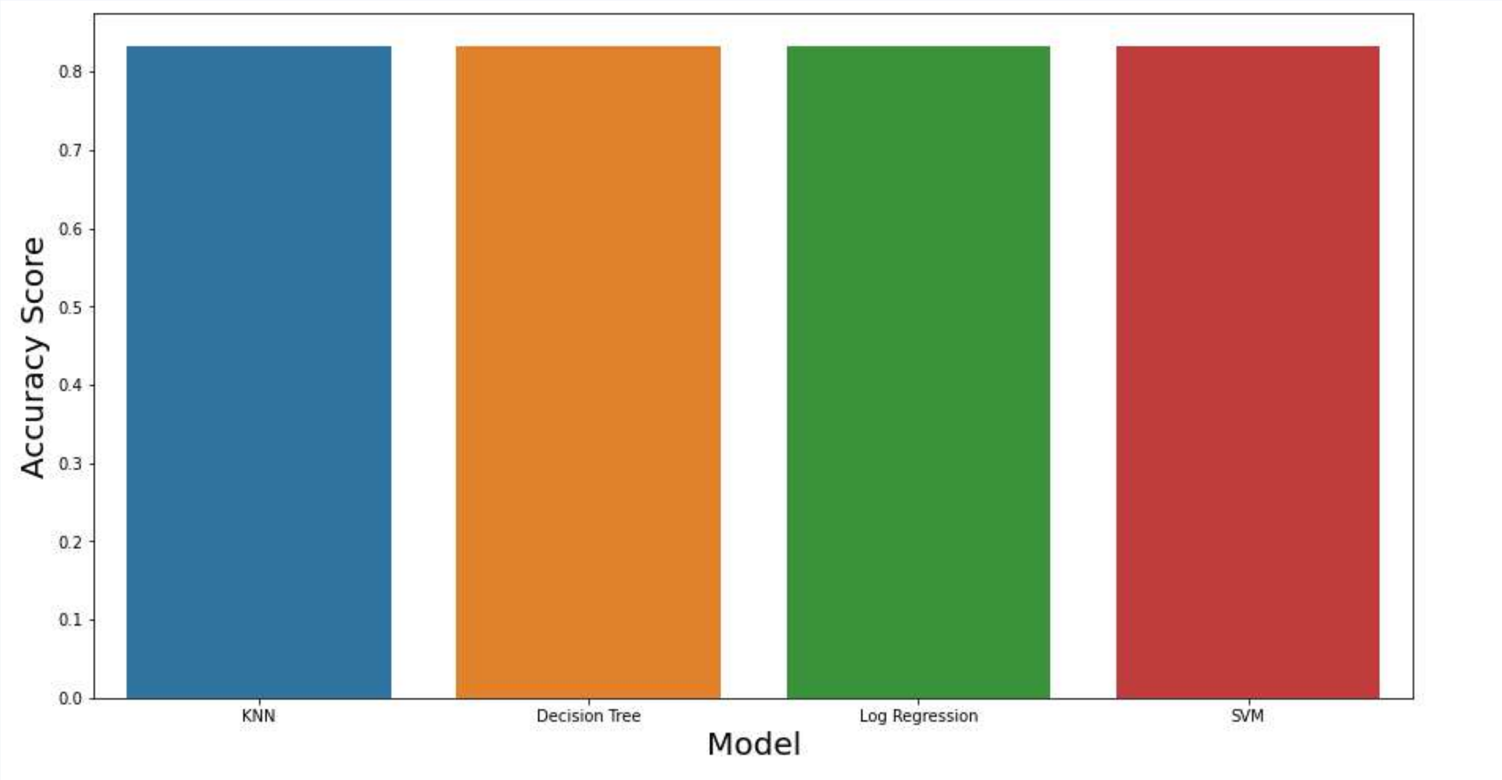


- Payload vs. Launch Outcome scatter plot for all sites, with different payload between 5000kg and 10000kg
- According the the plot above FT and B4 booster version were only two to have successful launches over 5000kg with B4 being the only successful booster version above 7000kg

Section 5

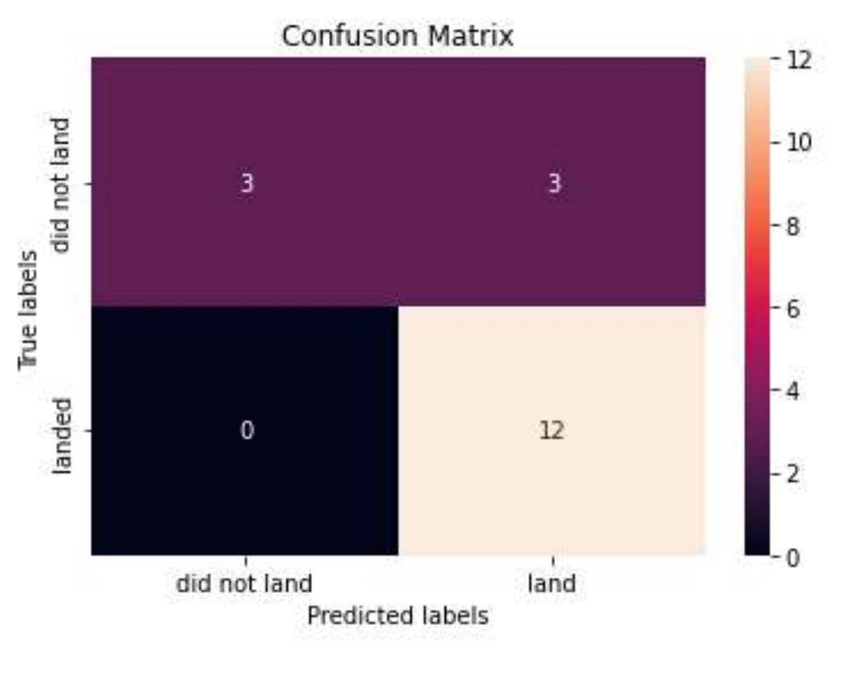
Predictive Analysis (Classification)

Classification Accuracy



- Model accuracy for all built classification models, in a bar chart
- All the models had the same test set accuracy of 83.33%

Confusion Matrix



- The confusion matrix was similar for all the models tested.
- According to the confusion matrix the model predicted that 3 launches did not land when actual stated that 6 did not land.
- The model predicted that 15 of the launches landed when in fact 12 landed.
- The model appears to be more optimistic than the actual with an overall accuracy of 83.33%

Conclusions

- Spaces launches are dependent of launch site location, payload mass, booster version & number of flight
- SpaceY has a 83.33% guarantee of successful flights
- The most successful launch site is KSC LC-39A
- Its necessary to select the right booster version for a certain payload mass.
  - FT booster version is more successful with loads under 6000kg
  - B4 booster version is responsible for success over 9000kg
- Successes increase overtime

Appendix

- References
  - Explanation about launch site proximities - https://solarsystem.nasa.gov/basics/chapter14-1/

Thank you!