Winning Space Race with Data Science

Gabriele Cerioli

GitHub Repo:

https://github.com/Gabmi73/coursera-capstone-project/tree/master

Outline

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

Executive Summary

Summary of methodologies

- Data collected by SpaceX API and Wikipedia Falcon 9 records
- Performed EDA
- Performed predictive analysis on four classification models
- Performed interactive map visualizations for impact on population neighborhood

Summary of all results

- Success rates of both launching (99%) and landing outcomes (66,66%) of Falcon 9 Rockets
- Best Launch Site performance (Cape Canaveral Space Launch Complex 40)
- Accuracy for predicting landing outcomes: 83,3%

Introduction

- SpaceX’ Falcon 9 rocket launches costs are sensible less than other providers
- The key success of its savings is because SpaceX reuses the rockets after the first stage boosters
- All available information and records about Falcon 9’ past launches and landings are provided by both the SpaceX API and the Wikipedia webpage
- This information can be also used whether an alternative company would like to bid against SpaceX
- In this project I will predict if the Falcon 9 first stage will land successfully

Section 1

Methodology

Methodology

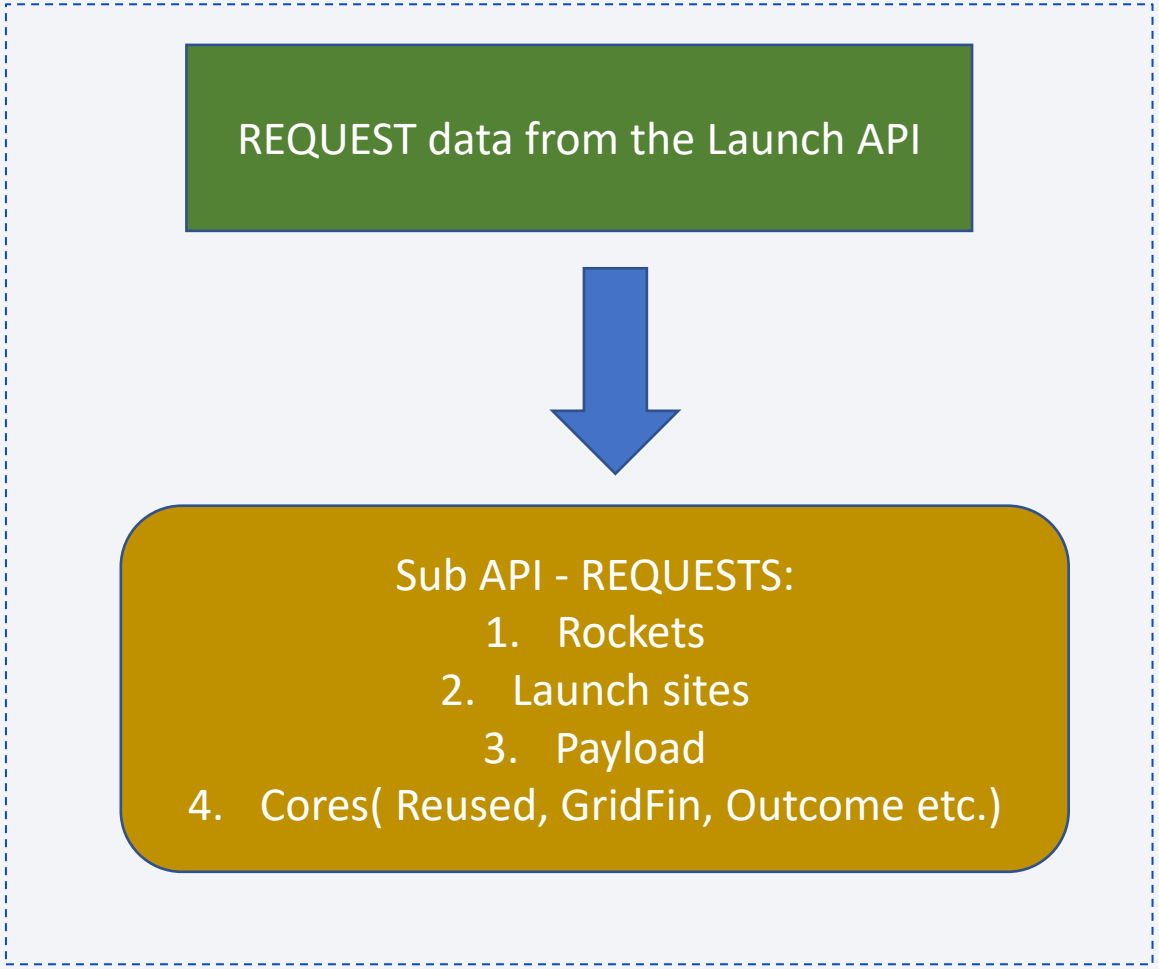
Executive Summary

- Data sets were collected both by the SpaceX data API and Wikipedia Falcon 9 and Falcon Heavy Launches Records
- Data was processed into a Data Frame, cleaned and analyzed with Python Pandas
- Performed exploratory data analysis (EDA) both with SQL and Viz
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models: splitting train-test set, choose best parameters, fitted models (Logistic Regression, SVM, Decision Tree, K-NN), evaluated models and test

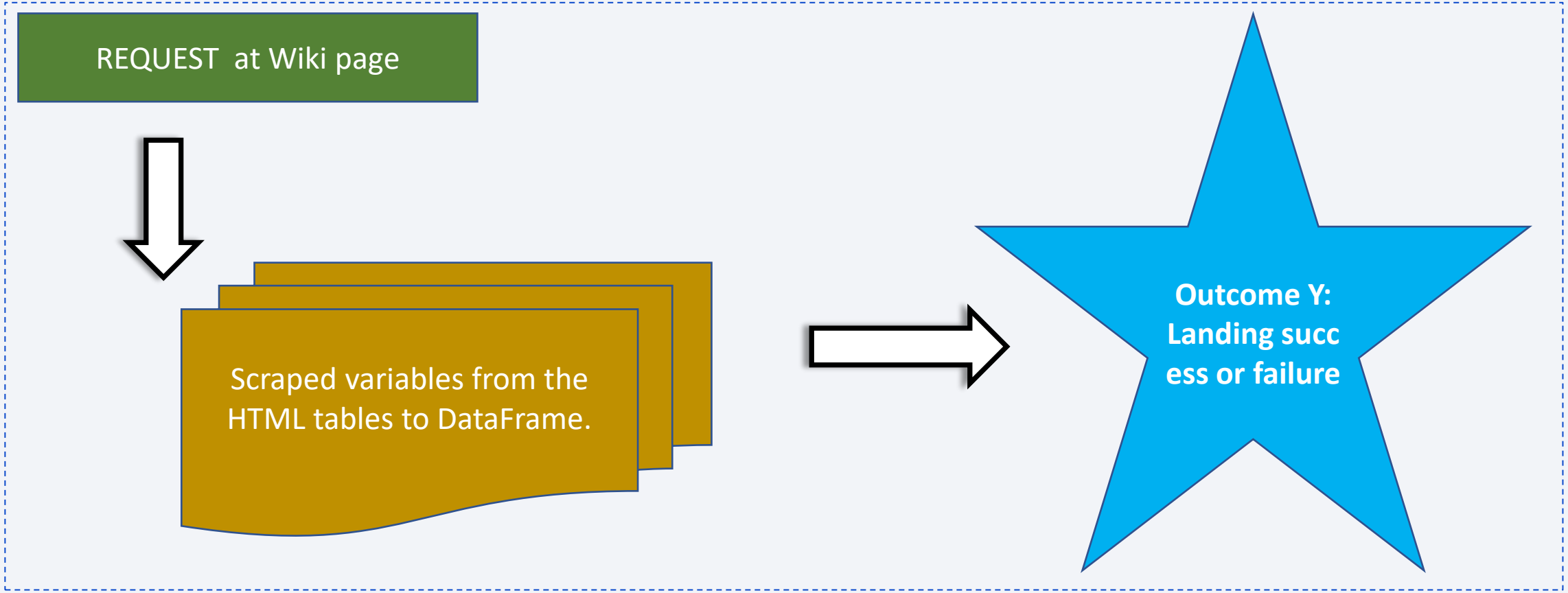
Data Collection – SpaceX API

SpaceX API calls notebook available at:

https://github.com/Gabmi73/coursera-capstone-project/blob/master/Data_Collection_API.ipynb



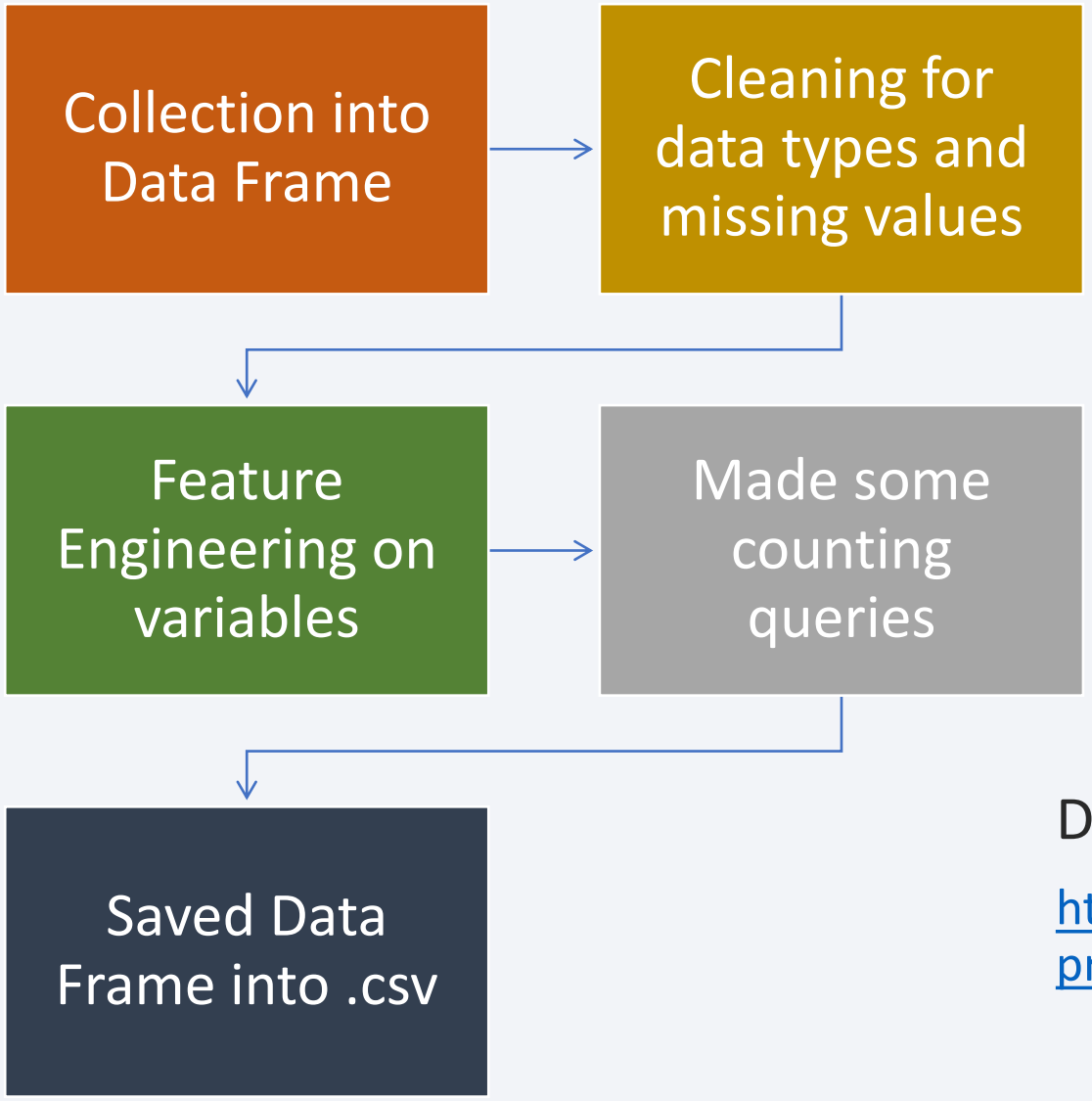
Data Collection - Scraping



Web-scraping notebook available at:

https://github.com/Gabmi73/coursera-capstone-project/blob/master/Web_ScrapingLab.ipynb

Data Wrangling



Data wrangling notebook available at:

https://github.com/Gabmi73/coursera-capstone-project/blob/master/Data_Wrangling.ipynb

EDA with Data Visualization

- Scatterplots were used to check the relationships between four variables: Flight Number, Payload Mass, Launch Sites and Orbit. Scatters better show linearities and homogeneous groups.
- A trend line was used to visualize the yearly trend line of the launch success
- A bar plot shows different Success Rates for each orbit

Data Vizualization notebook available at:

https://github.com/Gabmi73/coursera-capstone-project/blob/master/EDA_with_Viz.ipynb

EDA with SQL

SQL EDA available at: https://github.com/Gabmi73/coursera-capstone-project/blob/master/EDA_with_SQL.ipynb

- Distinct Launch Site names
- Displayed first 5 record starting with CCA
- Selected total Payload Mass carried by Boosters launched by NASA (CRS)
- Calculated average Payload Mass carried by Booster version F9 v1.1
- Extracted the date of the first successful landing
- Displayed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Counted the total number of successful and failure mission outcomes
- Displayed the names of the booster_versions which have carried the maximum payload mass
- Displayed the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Ranked the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20

Build an Interactive Map with Folium

Maps have been enriched by following Folium objects and for these reasons:

- circles, to visualize the number of launches of each site
- markers, to point each launch both successful and failure (green vs red)
- lines, to visualize the distance between sites and main public infrastructure as highways, railways and the closest city

Folium notebook available at:

https://github.com/Gabmi73/coursera-capstone-project/blob/master/Visual_Analytics_Folium.ipynb

Build a Dashboard with Plotly Dash

Dashbord created with following graphs and for related reasons:

- Pie Charts: to visualize in a glimpse the success of launches for all sites together and for each one individually
- Scatter Plots to visualize the relations between the most successful Booster Version per Payload Mass
- A scroll bar to select the range of Payload Mass and restrict/enlarge data visualization

Plotly Dash Python file available at:

https://github.com/Gabmi73/coursera-capstone-project/blob/master/DASHBOARD_CAPSTONE.py

Predictive Analysis (Classification)

For each one of the 4 models the following pipeline has been followed:

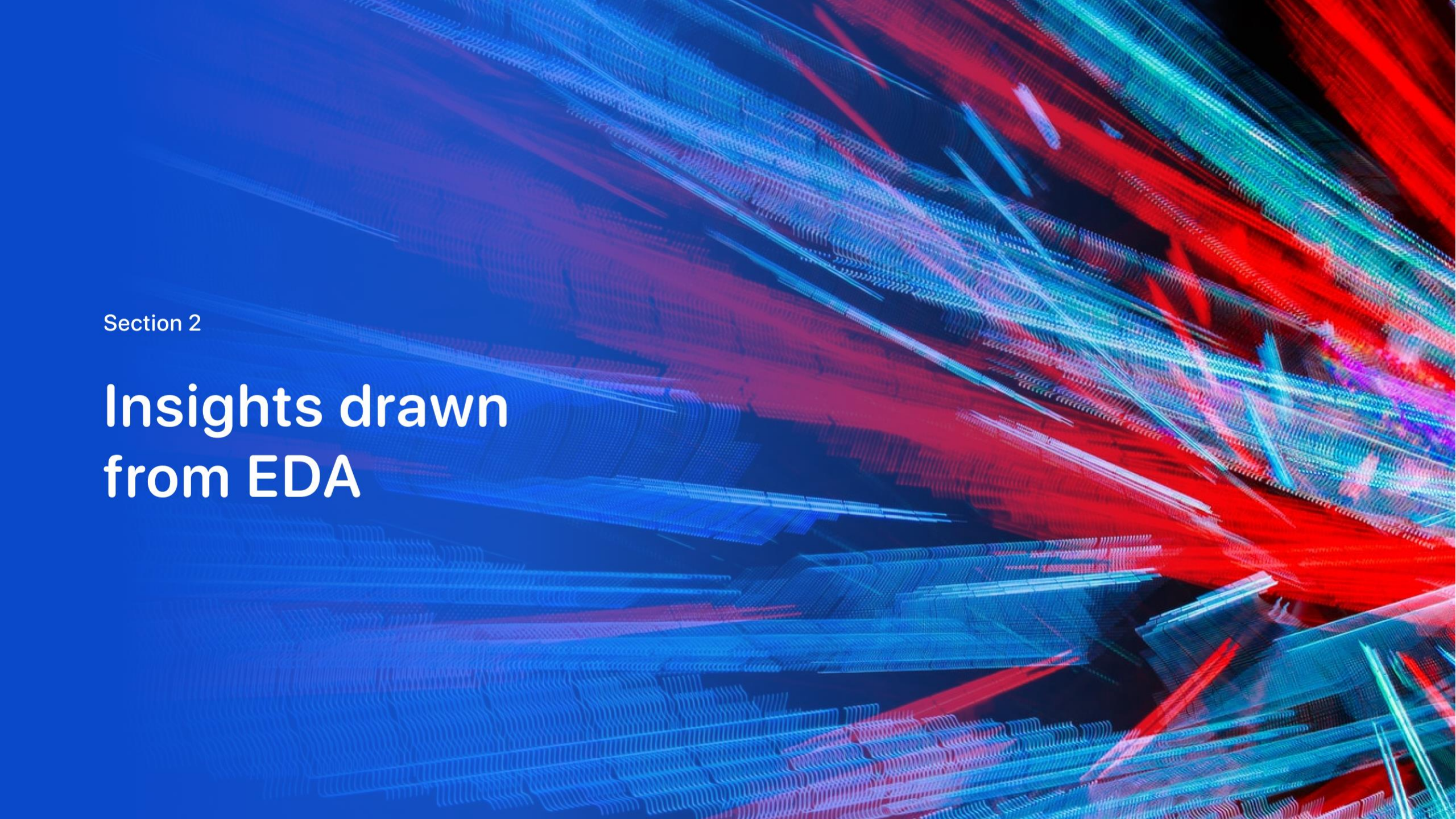
Notebook of predicting analysis available at:
https://github.com/Gabmi73/coursera-capstone-project/blob/master/Prediction.ipynb

Results

- First stage of Falcon 9 is successful
- Most failure SpaceX launches are located in Cape Canaveral Space Launch Complex 40
- Trend of last 5 years tend to increase in Success Rate for launching
- Interactive analytics shows the distance from the first inhabited area and other infrastructures
- Second stage of Landing Success Rate is 66,66 %
- Maybe the best classification model should be Decision Tree
- PayLoad Mass and Orbit have a role in predicting outcomes
- Best accuracy of prediction for landings is 83%

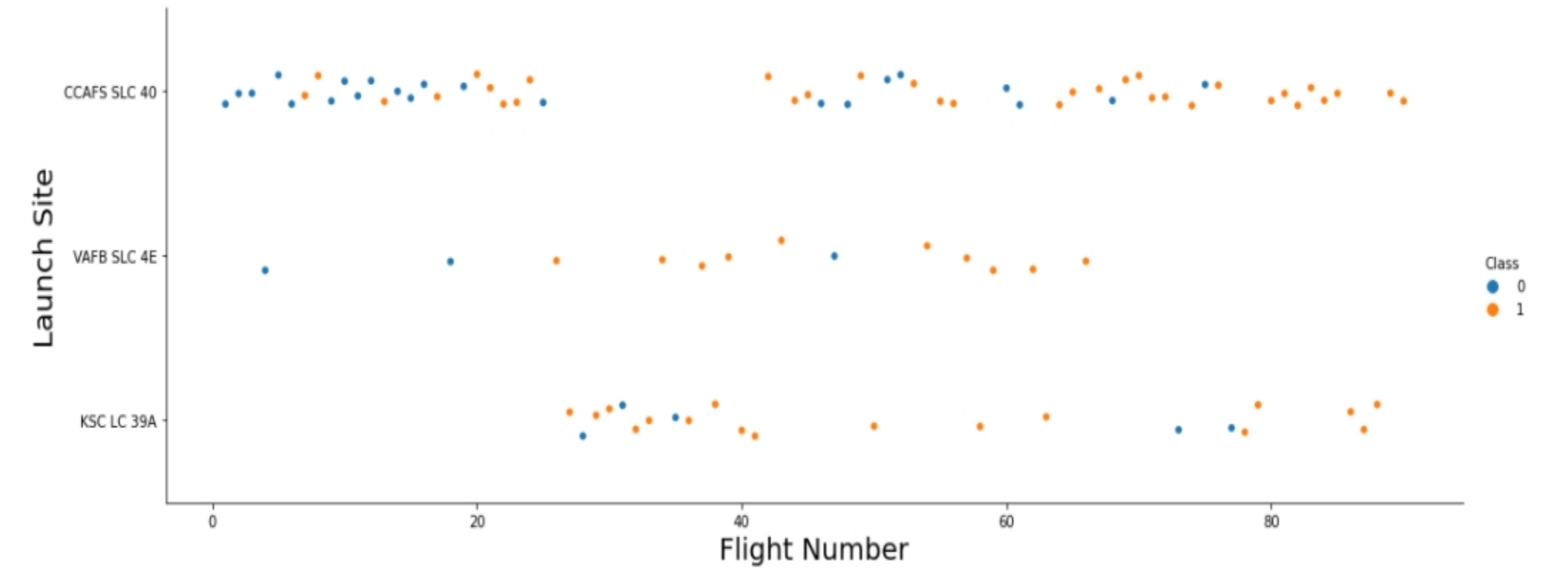
Section 2

Insights drawn from EDA



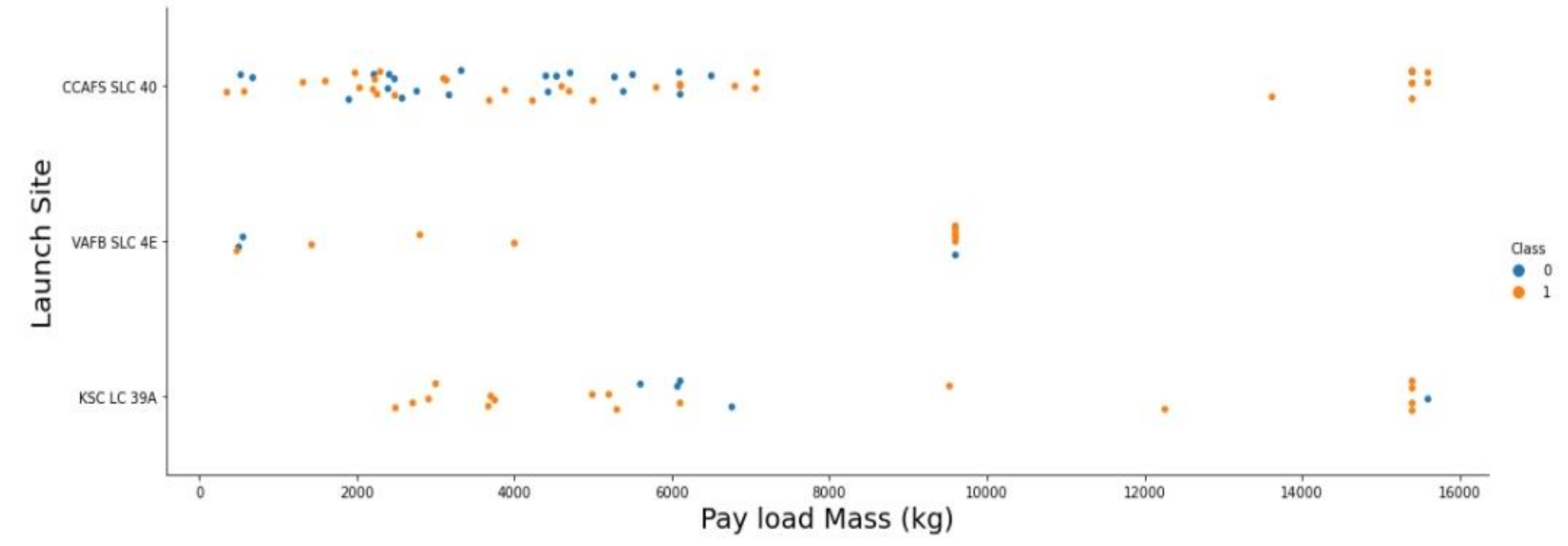
Flight Number vs. Launch Site

The following scatterplot shows the relationship between the Launch Site, in success (Class 1) or failure (Class 0) of Falcon 9’ landings, in function of the number of the flights



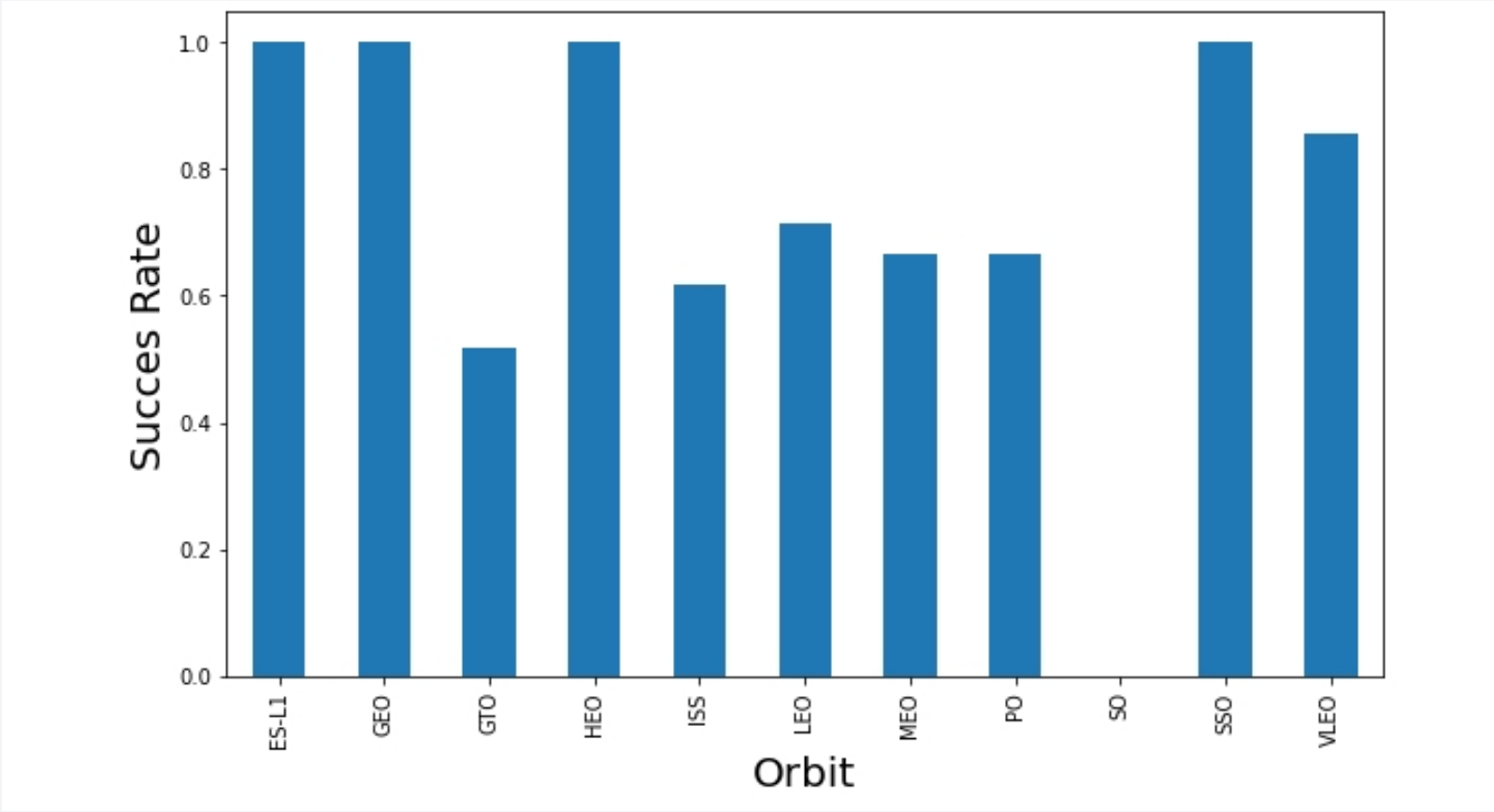
Payload vs. Launch Site

As follows the graph relationship between the Launch Site, in success (Class 1) or failure (Class 0) of Falcon 9' landings, in function of the Payload Mass (in kg)



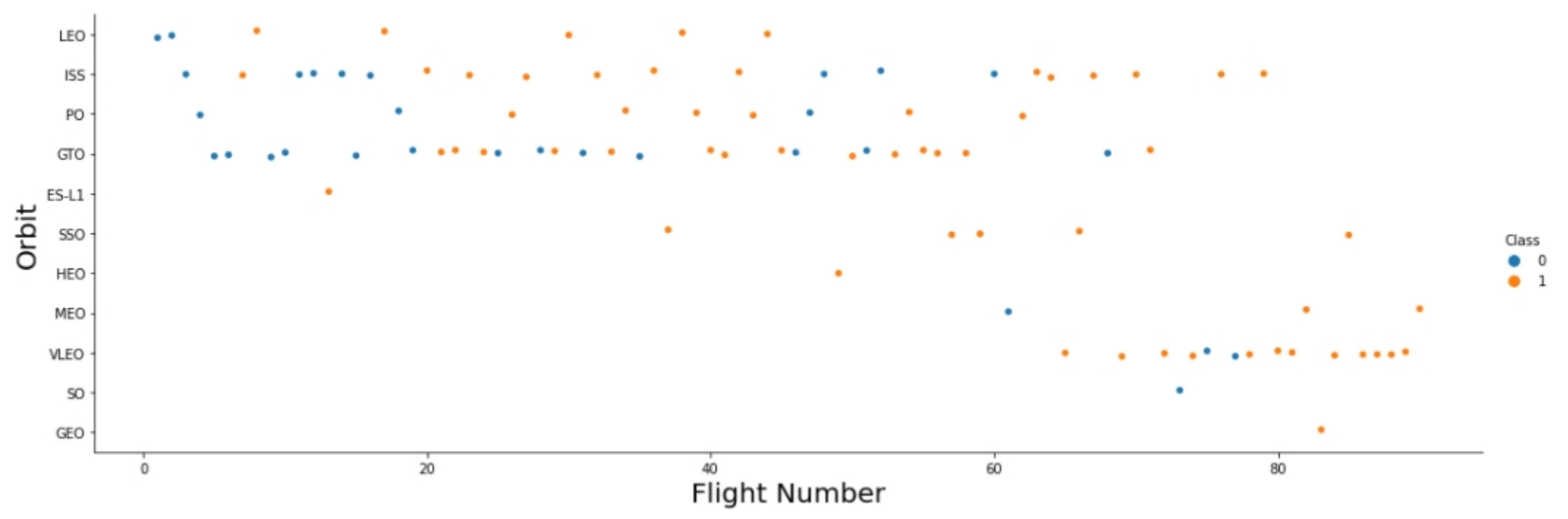
Success Rate vs. Orbit Type

For each Orbit Type a bar revelas the succes rate of landings



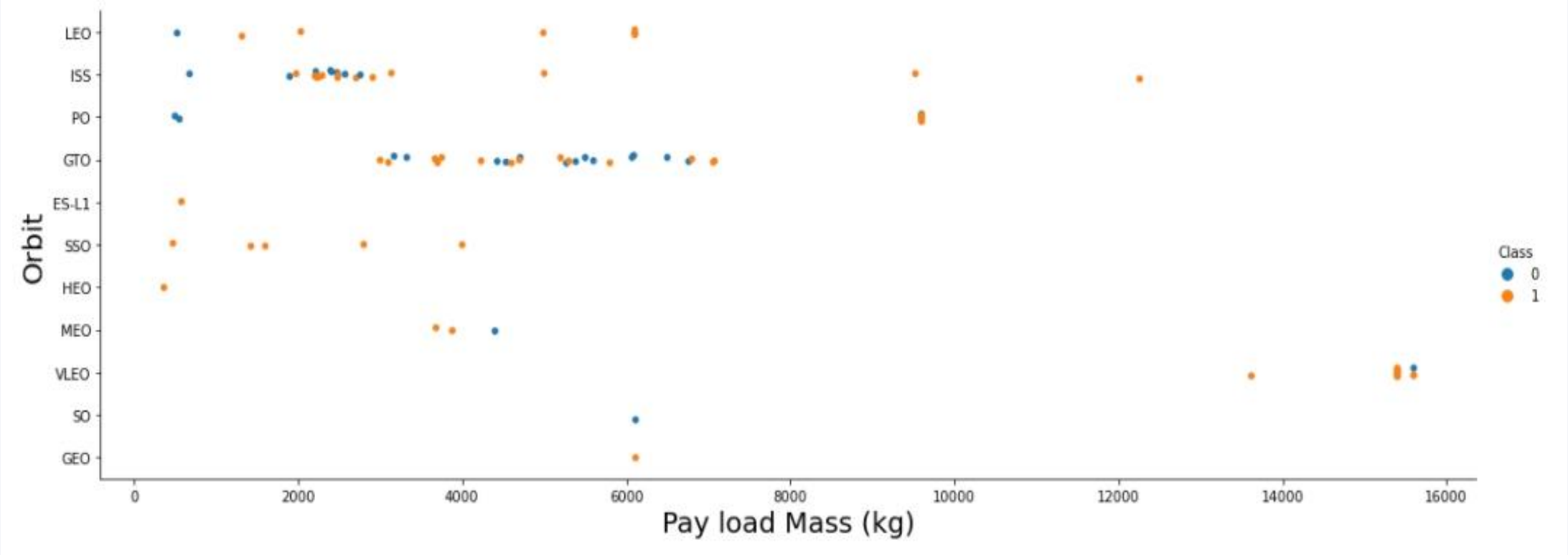
Flight Number vs. Orbit Type

This scatterplot shows the relationship between the Orbit Type, in success (Class 1) or failure (Class 0) of Falcon 9’ landings, in function of the number of the flights



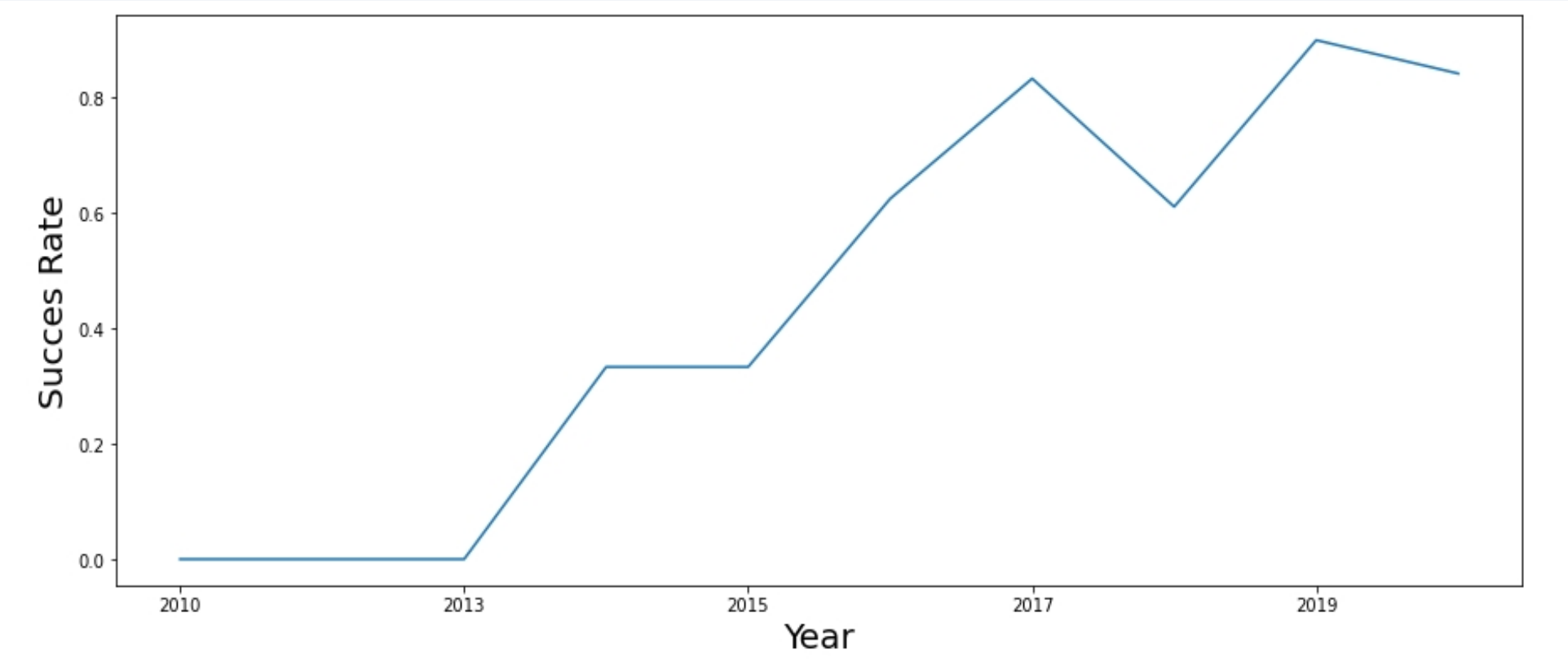
Payload vs. Orbit Type

In this slide: the relationship between the Orbit Type, in success (Class 1) or failure (Class 0) of Falcon 9’ landings, in function of the Payload Mass (in kg)



Launch Success Yearly Trend

The trend for each year of succesful landings



All Launch Site Names

Acronym	Site Name
CCAFS SLC 40	Cape Canaveral Space Launch Complex 40
CCAFS LC 40	Cape Canaveral Launch Complex 40
KSC LC 39A	Kennedy Space Center Launch Complex 39A
VAFB SLC 4E	Vandenberg Air Force Base Space Launch Complex 4E

Acronyms and related names of the three launch sites of Falcon 9 rockets

Launch Site Names Begin with 'CCA'

First five records of Launch Site starting with ‘CCA’

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

Total Payload Mass carried by boosters launched by NASA (CRS) is:

45596 KG

Average Payload Mass by F9 v1.1

The average Payload Mass carried by booster version F9 v1.1 is:

2928.40

First Successful Ground Landing Date

The date when the first successful landing outcome in ground pad was achieved:

22^{nd} - Dec - 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Payload	payload_mass__kg_
JCSAT-14	4696
JCSAT-16	4600
SES-10	5300
SES-11 / EchoStar 105	5200

Total Number of Successful and Failure Mission Outcomes

The following list of total nr of successful and failure mission outcomes (101)

All is clear, only one failure against one hundred successes

Mission_Outcome	n
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Total Number of Successful and Failure Landing Outcomes

The following is the list of all successful and failure landings. On a total of 101 launches 66 were successful

landing__outcome	n
Success	38
No attempt	22
Success (drone ship)	14
Success (ground pad)	9
Controlled (ocean)	5
Failure (drone ship)	5
Failure	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Boosters Carried Maximum Payload

Booster_Version	payload_mass__kg_	Payload
F9 B5 B1048.4	15600	Starlink 1 v1.0, SpaceX CRS-19
F9 B5 B1049.4	15600	Starlink 2 v1.0, Crew Dragon in-flight abort test
F9 B5 B1051.3	15600	Starlink 3 v1.0, Starlink 4 v1.0
F9 B5 B1056.4	15600	Starlink 4 v1.0, SpaceX CRS-20
F9 B5 B1048.5	15600	Starlink 5 v1.0, Starlink 6 v1.0
F9 B5 B1051.4	15600	Starlink 6 v1.0, Crew Dragon Demo-2
F9 B5 B1049.5	15600	Starlink 7 v1.0, Starlink 8 v1.0
F9 B5 B1060.2	15600	Starlink 11 v1.0, Starlink 12 v1.0
F9 B5 B1058.3	15600	Starlink 12 v1.0, Starlink 13 v1.0
F9 B5 B1051.6	15600	Starlink 13 v1.0, Starlink 14 v1.0
F9 B5 B1060.3	15600	Starlink 14 v1.0, GPS III-04
F9 B5 B1049.7	15600	Starlink 15 v1.0, SpaceX CRS-21

This is the list of the names of booster versions which have carried the maximum payload mass.

2015 Launch Records

Date	Landing__Outcome	Booster_Version	Launch_Site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Failed landing outcomes in drone ship and their booster versions in year 2015, both launched at Cape Canaveral Space Launch Complex 40

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

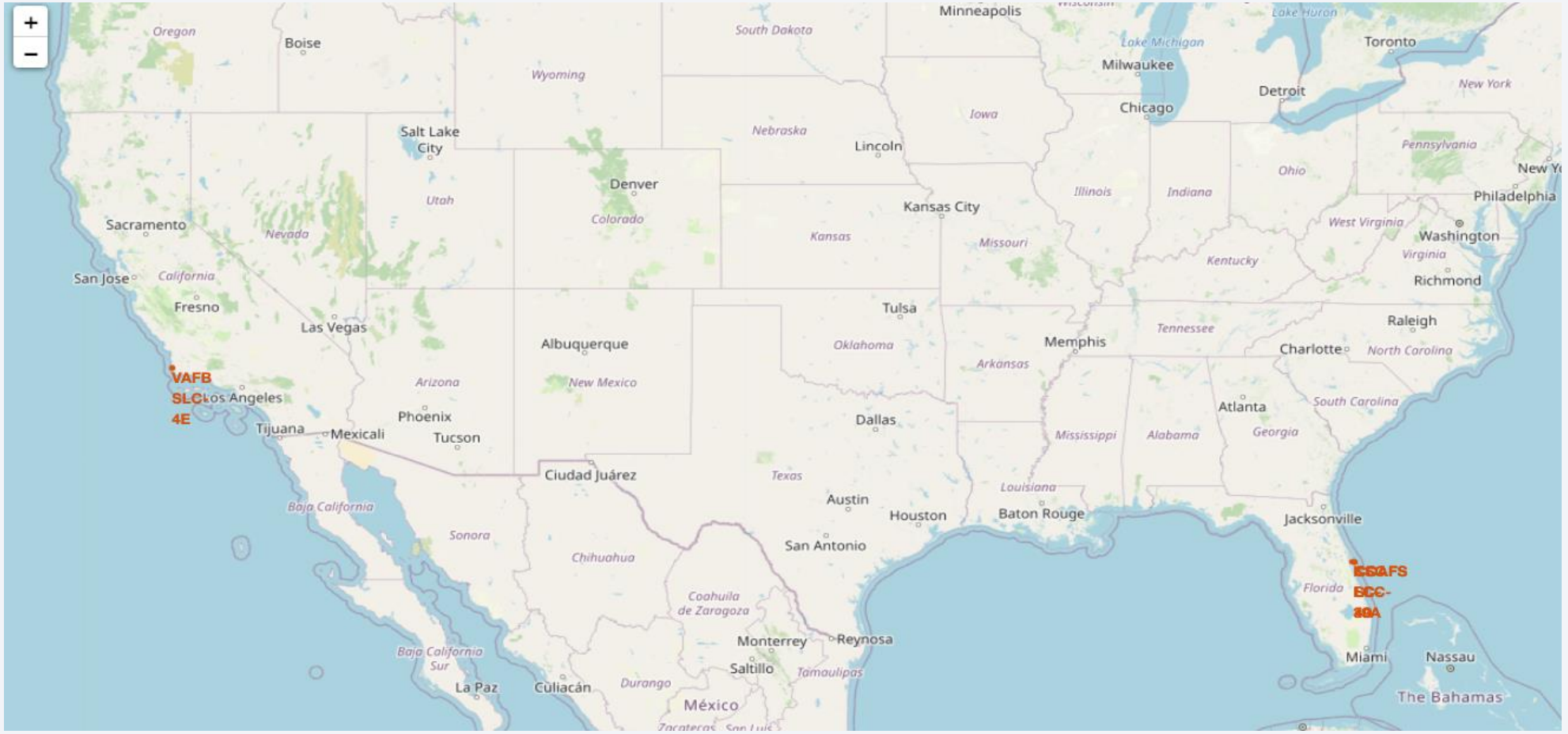
Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20

Date	Landing__Outcome
2015-12-22	Success (ground pad)
2016-07-18	Success (ground pad)
2017-02-19	Success (ground pad)
2015-01-10	Failure (drone ship)
2015-04-14	Failure (drone ship)
2016-01-17	Failure (drone ship)
2016-03-04	Failure (drone ship)
2016-06-15	Failure (drone ship)

Section 4

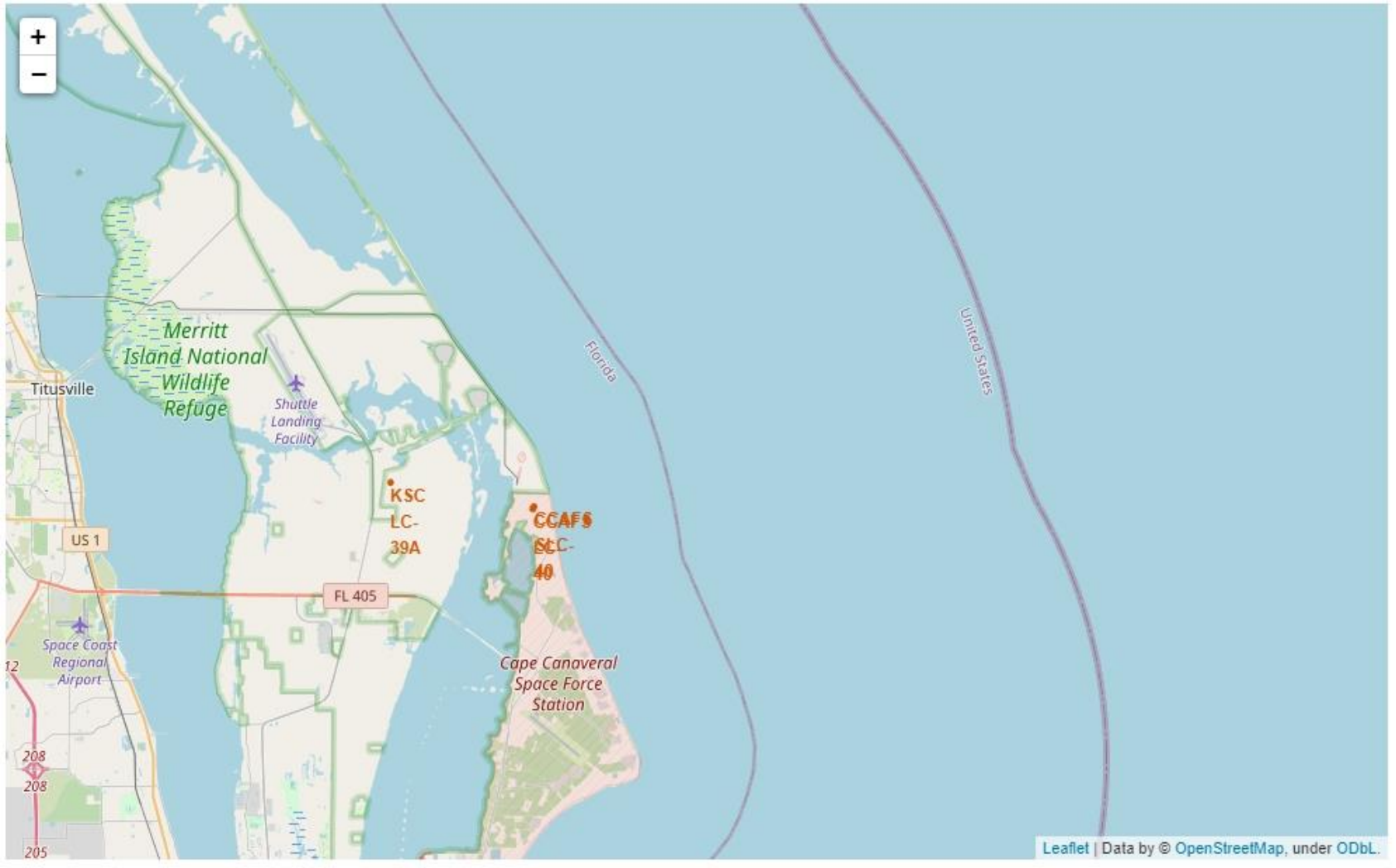
Launch Sites Proximities Analysis

All US Launch Sites - 1



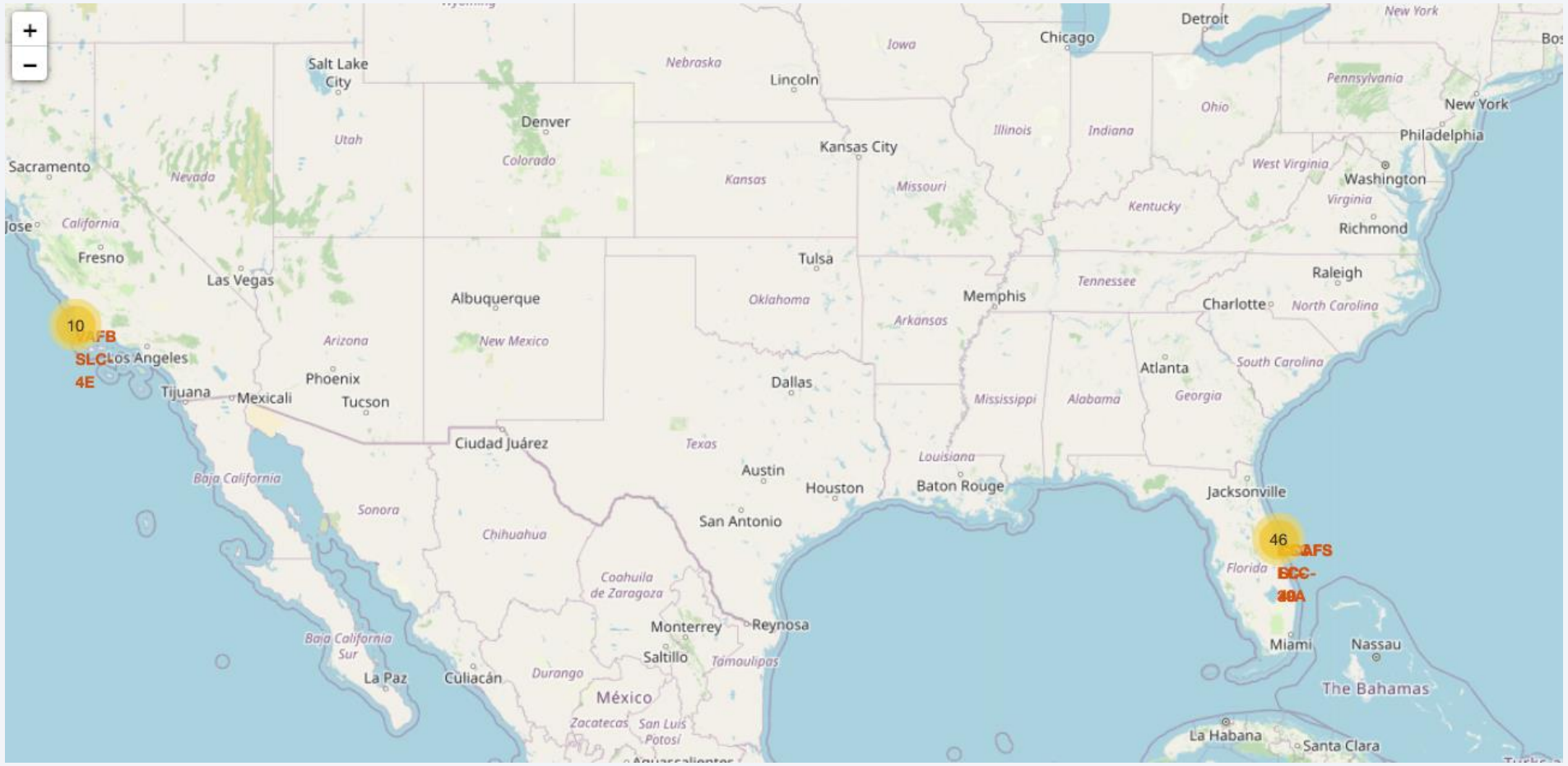
One site is in the east coast, CA, while the other two are in the west. The sites in Florida are very close, see next slide for better visualization.

All US Launch Sites - 2



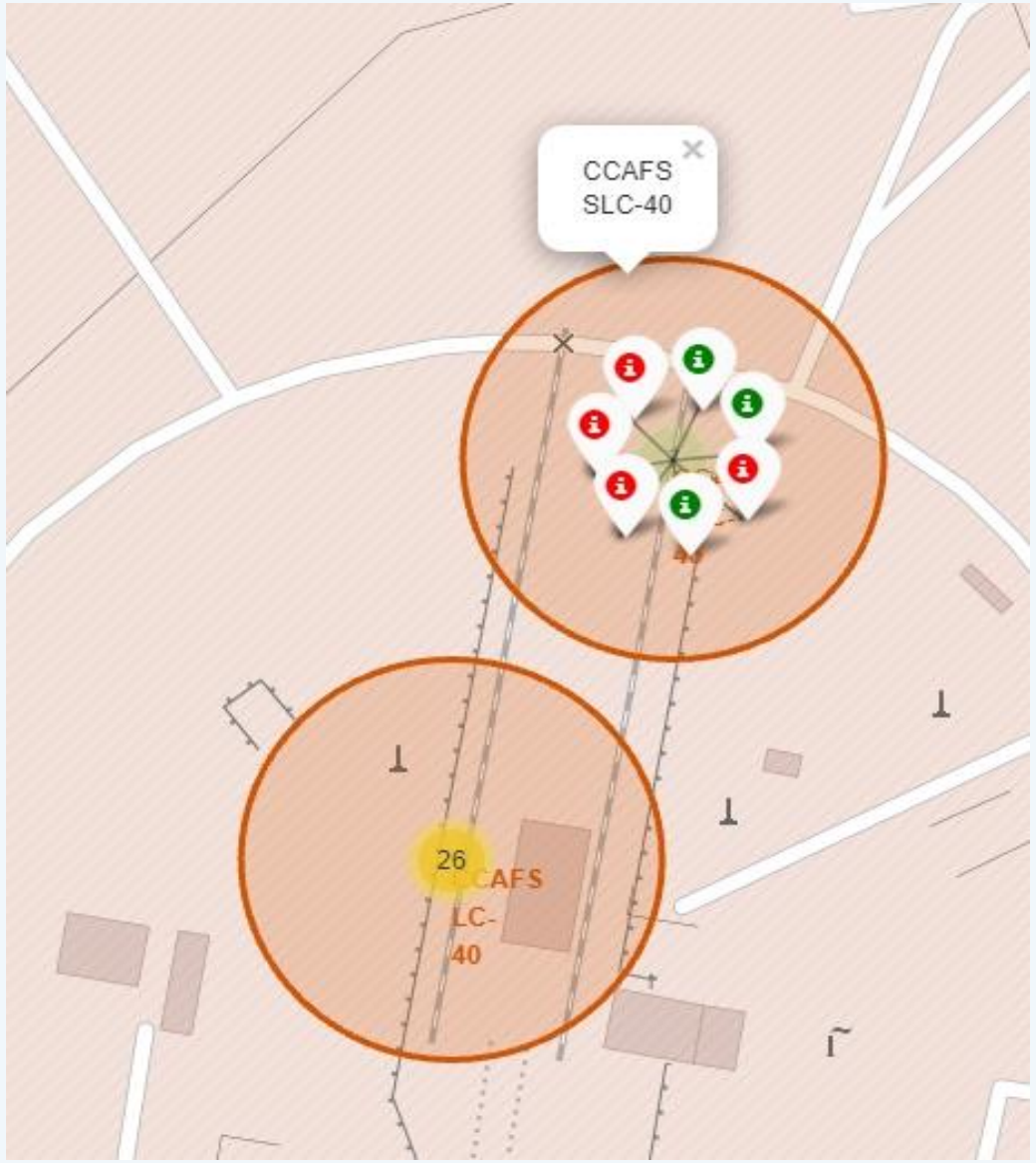
Sites in Florida

Outcomes for each Site - 1



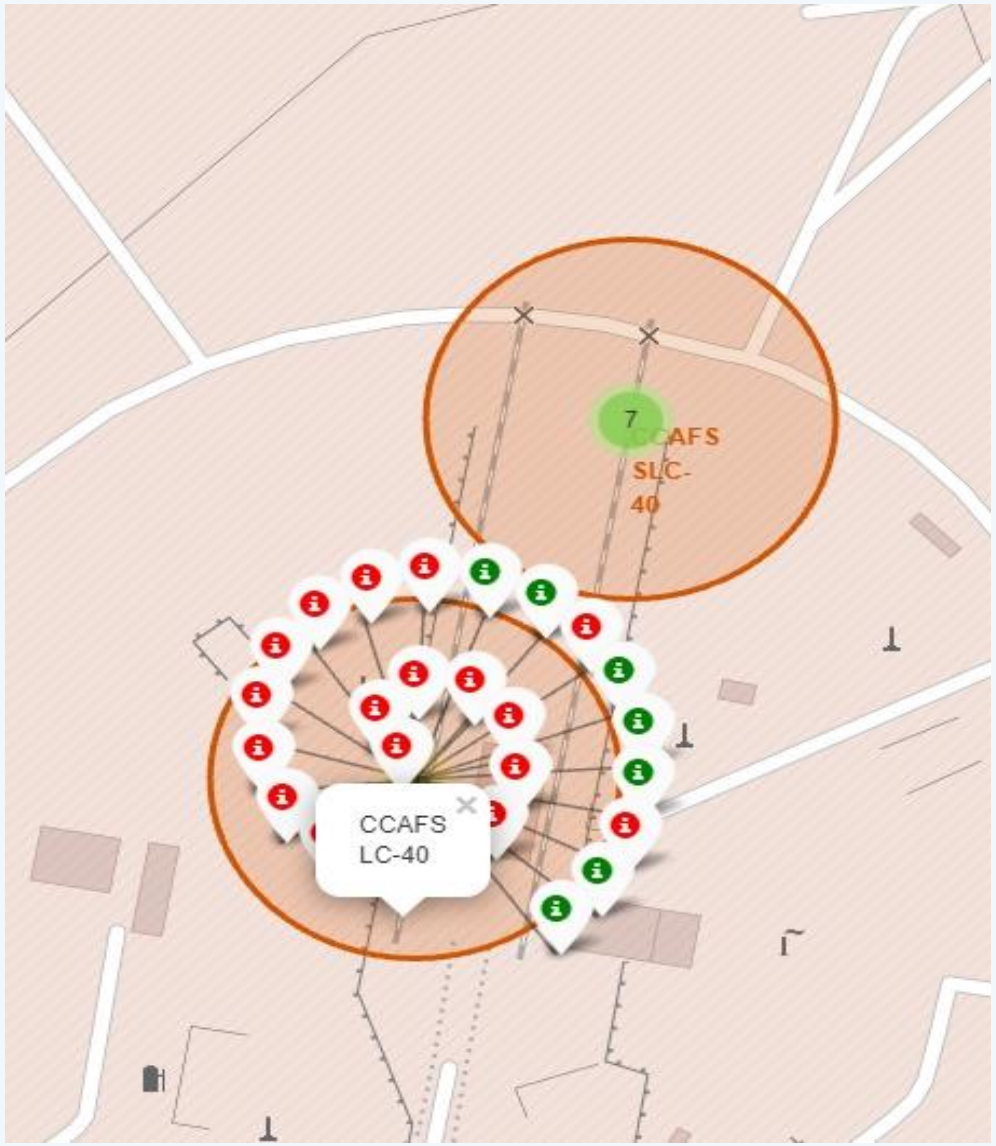
Global View shows that most of launches are in Florida’s sites

Outcomes for each Site – CCAFS SLC-40



On total of 7 launches, 3 were success and 4 failure

Outcomes for each Site – CCAFS LC-40

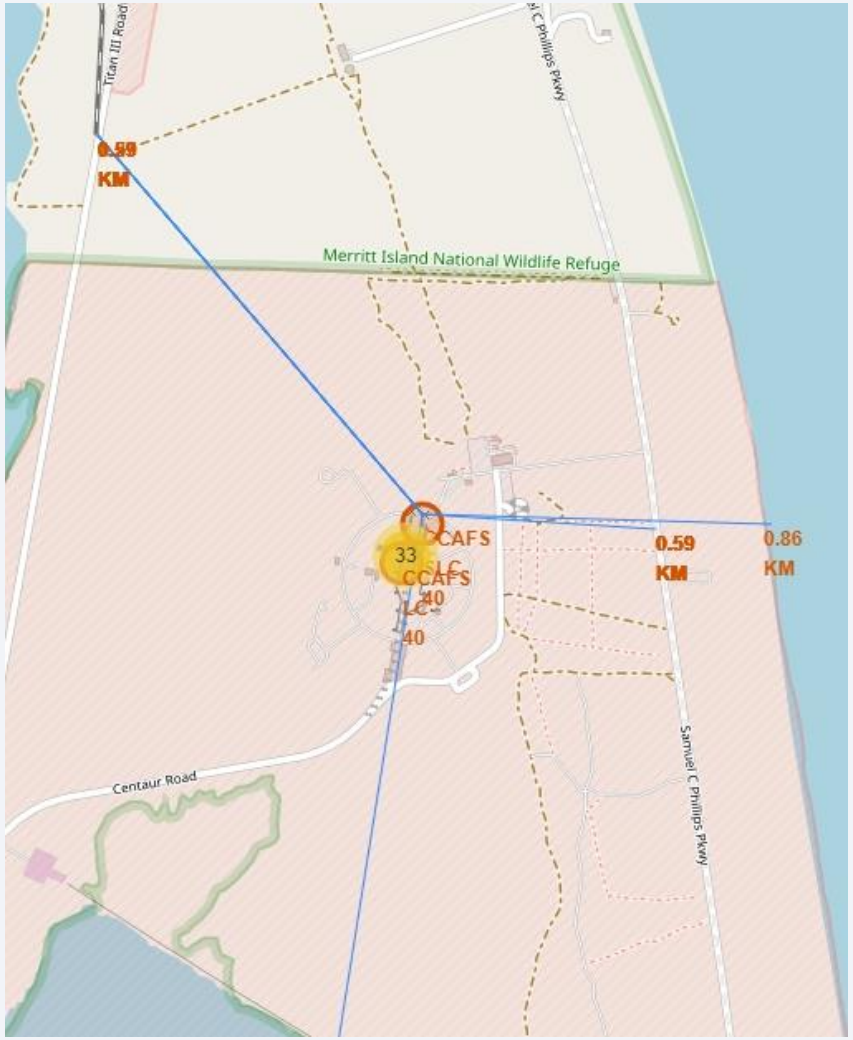


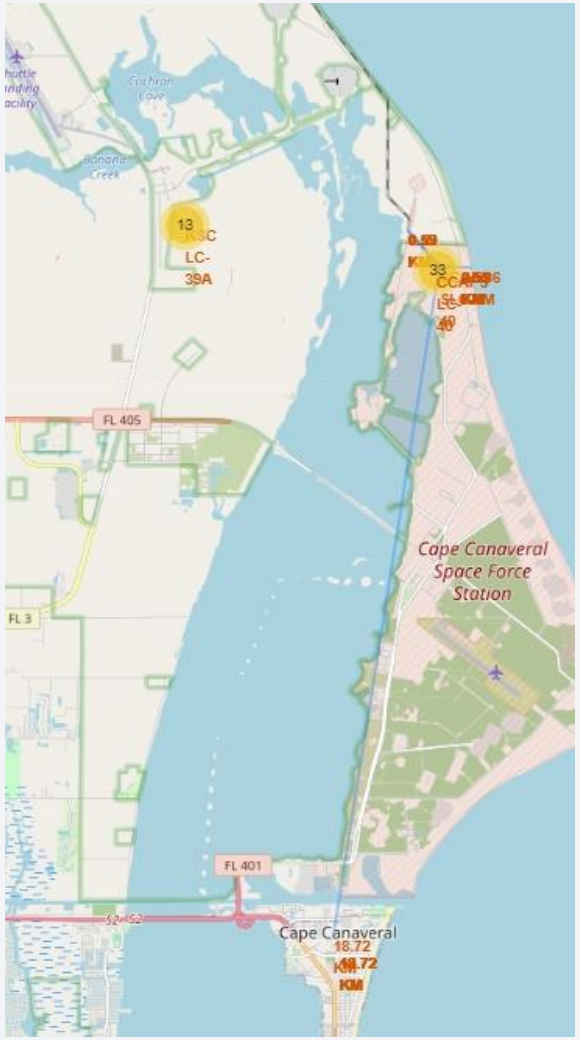
We can see that only 7 launces over 26 were succesful

Outcomes for each Site – VAFB SLC 4E

In California 4 launces over 10 were succesful

Distances from infrastructure and inhabitated areas



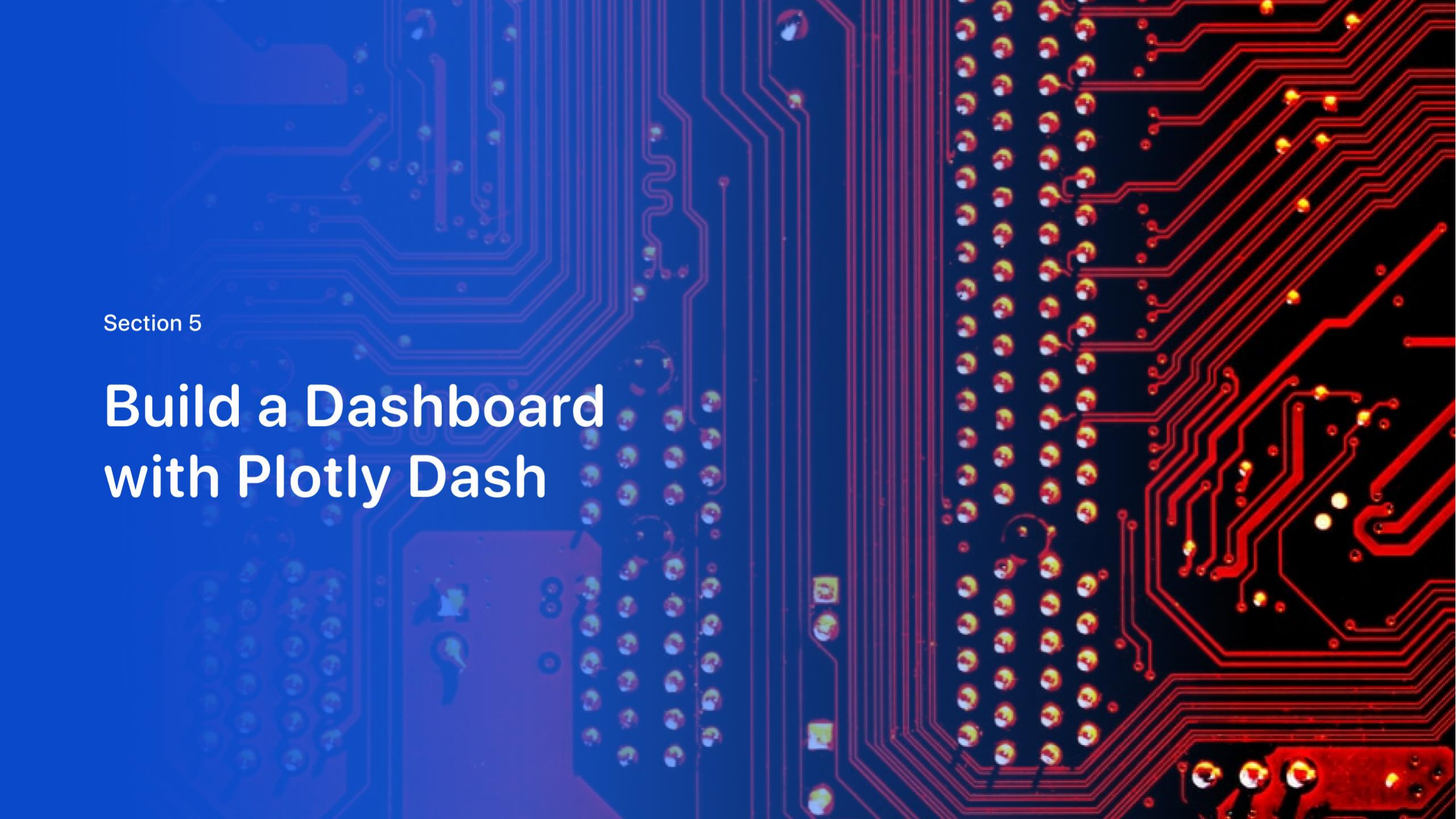


CCAFS SLC/LC 40 Sites are very close to the coastline and public infrastructure such as railway and highway (less than only 1 km).

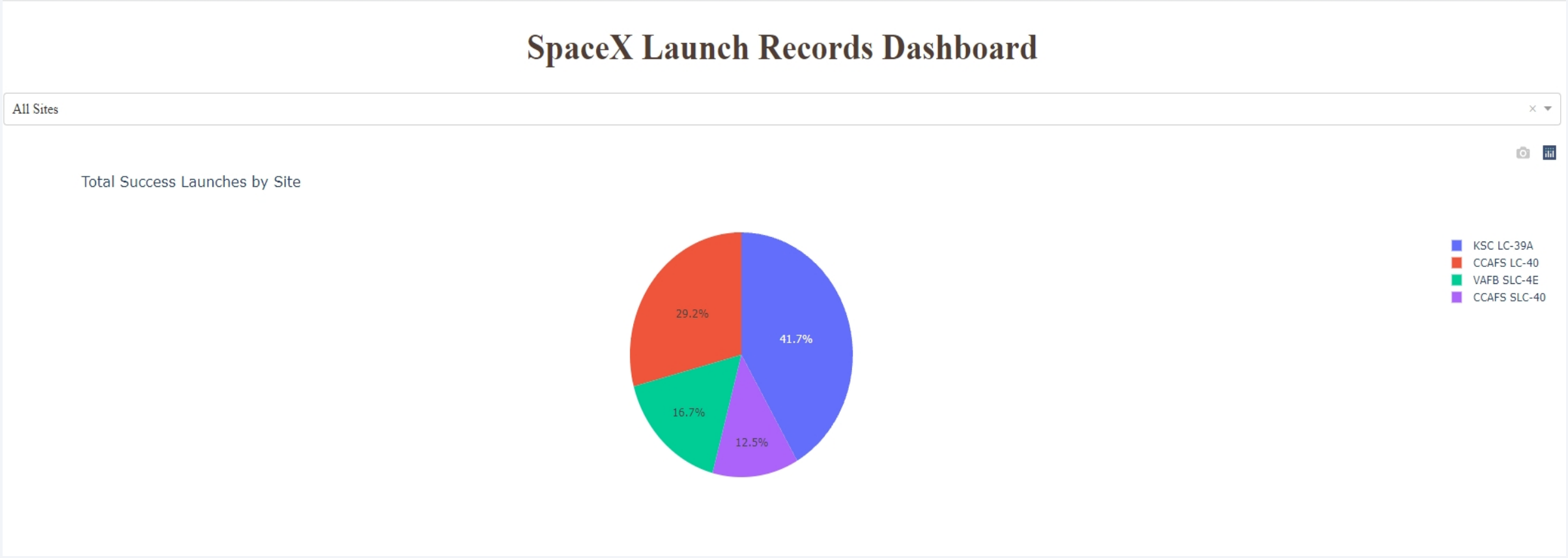
The first closer city, that is Cape Canaveral is quite distant: 18.72 KM

Section 5

Build a Dashboard with Plotly Dash

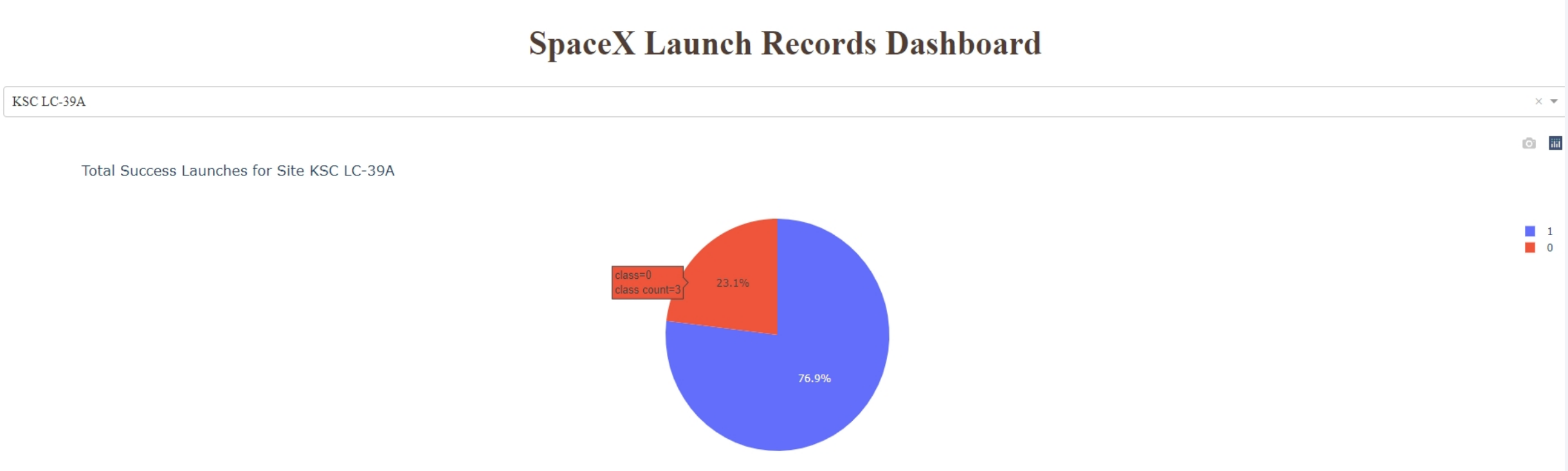


Total Success Launches by Site



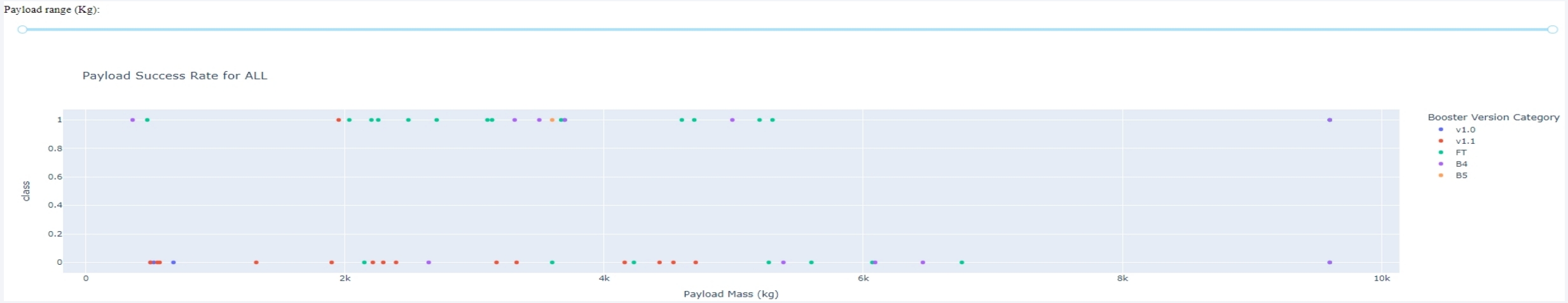
The two most successful places for launches are KSC LC-39A (41,7 % success rate) and CCAFS LC-40 (29,2 % s.r.)

Highest Launch Site Success Ratio



The Kennedy Space Center Launch Complex 39A recorded an s.r. of 76.9 % success launches (10 successful and 3 failed)

Booster Version vs Payload Mass



Most successful Booster Version per Payload Mass is FT. Also for lower Payload Mass (max 5,300 kg) the most successful Booster Version is FT

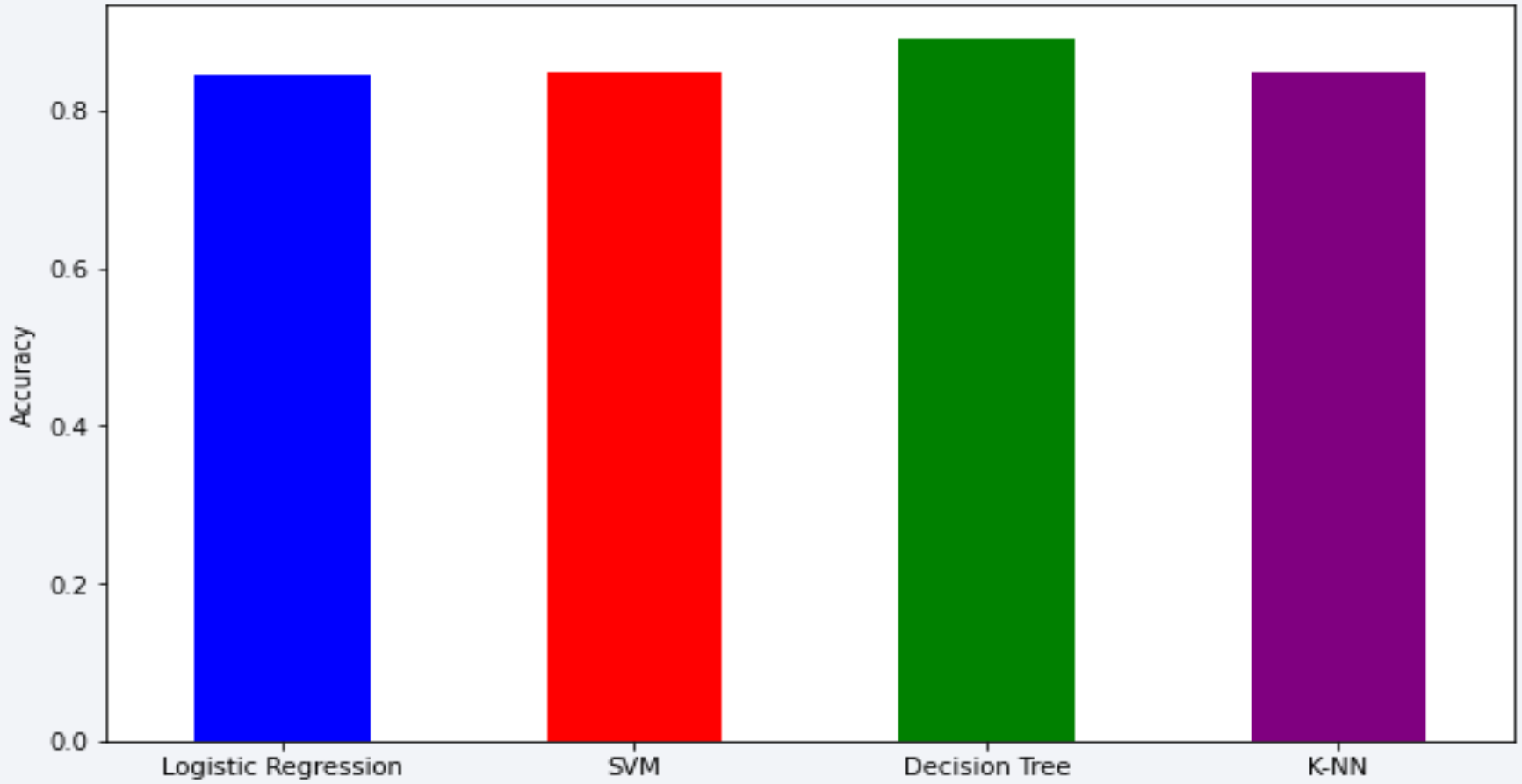
Section 6

Predictive Analysis (Classification)

Classification Accuracy

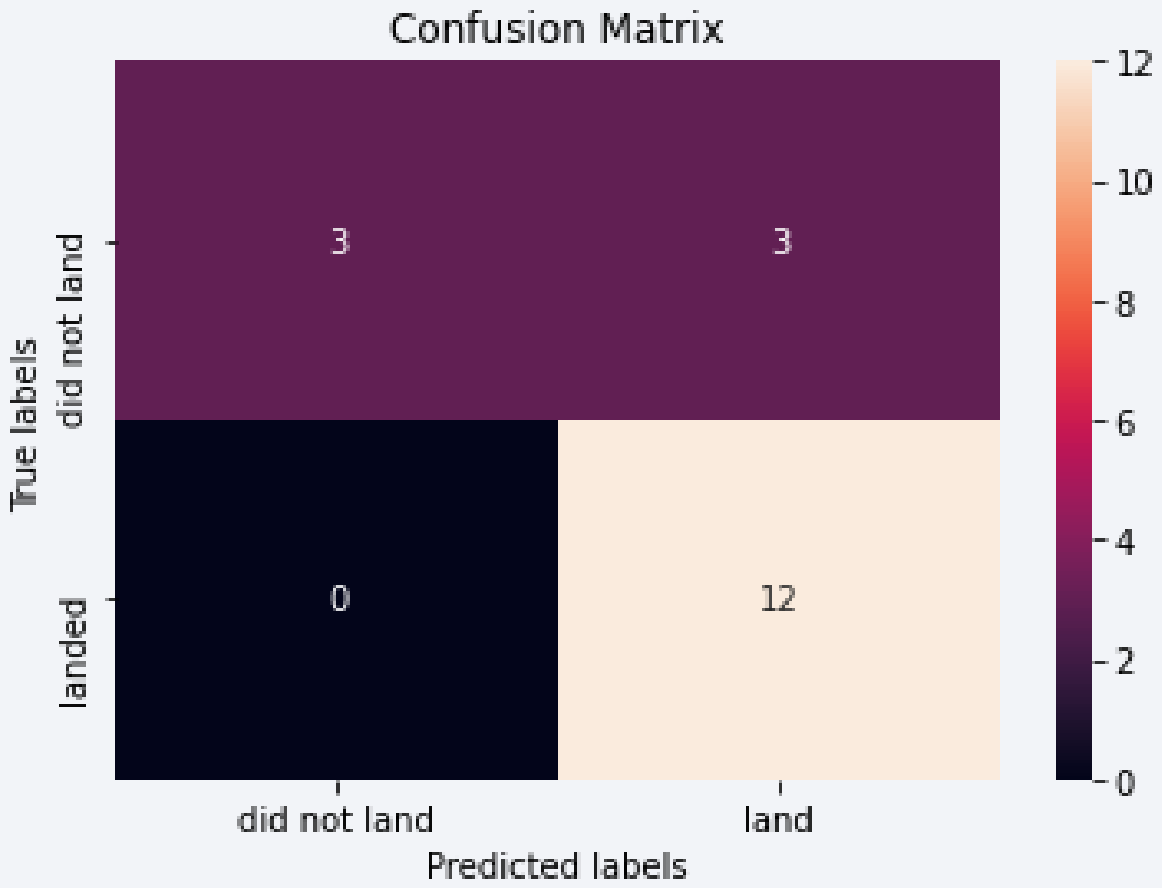
Decision Tree model has the highest Accuracy score: 0.891

All the models have same *Test Accuracy* which is 0.833



Confusion Matrix

Considering only the Accuracy on the Test Set, all the results are the same: there is no best model. However, the Decision Tree ML shows the highest Accuracy score in Training Set



Accuracy is good but in three times the model bad predicted the rocket landing, while it did not actually

Conclusions

- SpacesX is increasing its success in reusing Falcon 9 rockets after landings from first stage launches (Success Rate is about 67%)
- Site positions enough distanced from population neighborhoods tend to have more chance of a good outcome
- Predictions on landing’ outcomes have an accuracy of about 84%
- We can conclude that also different companies would save money by adopting Falcon 9 technology

Appendix

All Python notebooks and files are available at my GitHub

https://github.com/Gabmi73/coursera-capstone-project/tree/master

Thank you!