IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies

Data collected using SpaceX API and web scraping.

Exploratory Data Analysis used data wrangling, visualization and interactive visual analysis.

Machine learning prediction for different models.

- Summary of all results

Data collected from more than one source

Shows insights for different variables.

Shows the prediction and evaluation for different models.

Introduction

The objective of the project is to find if a new company will be able to compete with spaceX.

- Calculate the cost of the launch, the best place to launch and predict if the rocket is going to launch successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected from the SpaceX API and Wikipedia.
- Perform data wrangling
  - Perform exploratory Data Analysis and determine Training Labels
  - Perform exploratory data analysis (EDA) using visualization and SQL
- We Performed interactive visual analytics using Folium and Plotly Dash
- We Performed predictive analysis using classification models
- Split to train and test data and then evaluate the different models.

Data Collection

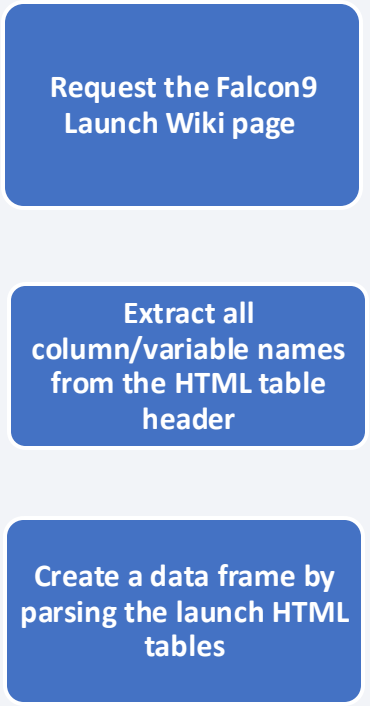
- Data collected from SpaceX API and Wikipedia which will be explained more in next slides.

Data Collection – SpaceX API

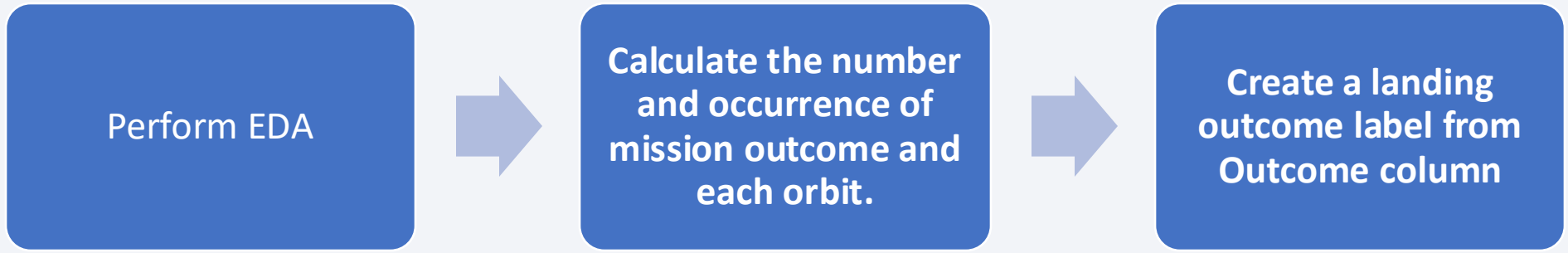
- GitHub URL : https://github.com/MohamedAlz66/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- GitHub URL : https://github.com/MohamedAlz66/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling



GitHub : https://github.com/MohamedAlz66/Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb

EDA with Data Visualization

- I used scatterplot to show the relationship between two variables such as PayloadMass and LaunchSite.
- Bar chart was also used to show the success rate of each orbit.
- Line chart was used to show that the success rate since 2013 kept increasing till 2020.
- GitHub URL : https://github.com/MohamedAlz66/Applied-Data-Science-Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- **Display the names of the unique launch sites in the space mission**
- **Display 5 records where launch sites begin with the string 'CCA'**
- **Display the total payload mass carried by boosters launched by NASA (CRS)**
- **Display average payload mass carried by booster version F9 v1.1**
- **List the date when the first succesful landing outcome in ground pad was acheived.**
- **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 records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months 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/MohamedAlz66/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- map objects such as markers, circles, lines and marker cluster was created and added to a folium map.
- Marker show specific points.
- Circle to show area around specific coordinates.
- Lines to show distance between two coordinates.
- Marker cluster for group events in a coordinate

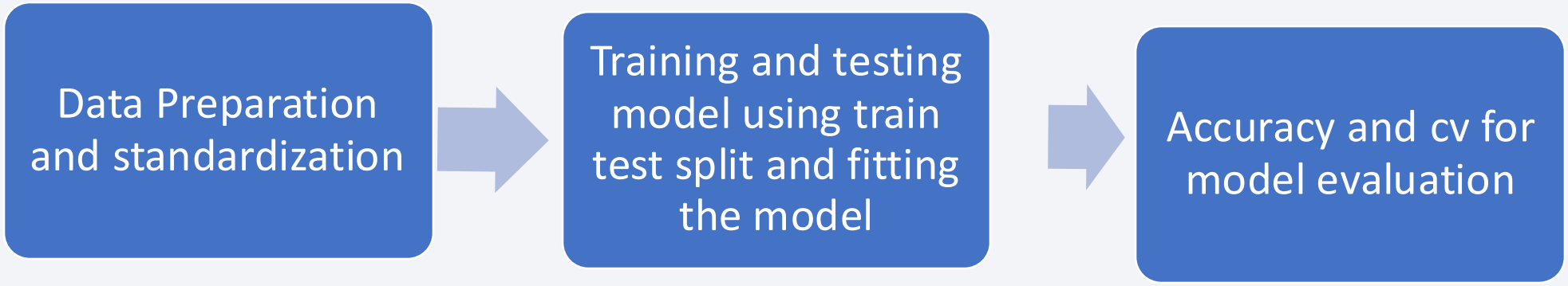
- GitHub URL : https://github.com/MohamedAlz66/Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Scatter plot and pie chart was used to visualize data
- It was used to show the relationship between payload mass and launch sites to show the best place to launch from.

- GitHub URL : https://github.com/MohamedAlz66/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)



GitHub URL : https://github.com/MohamedAlz66/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- Exploratory data analysis results

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Lati
1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.56
2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.56
3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.56
4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.63
5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.56

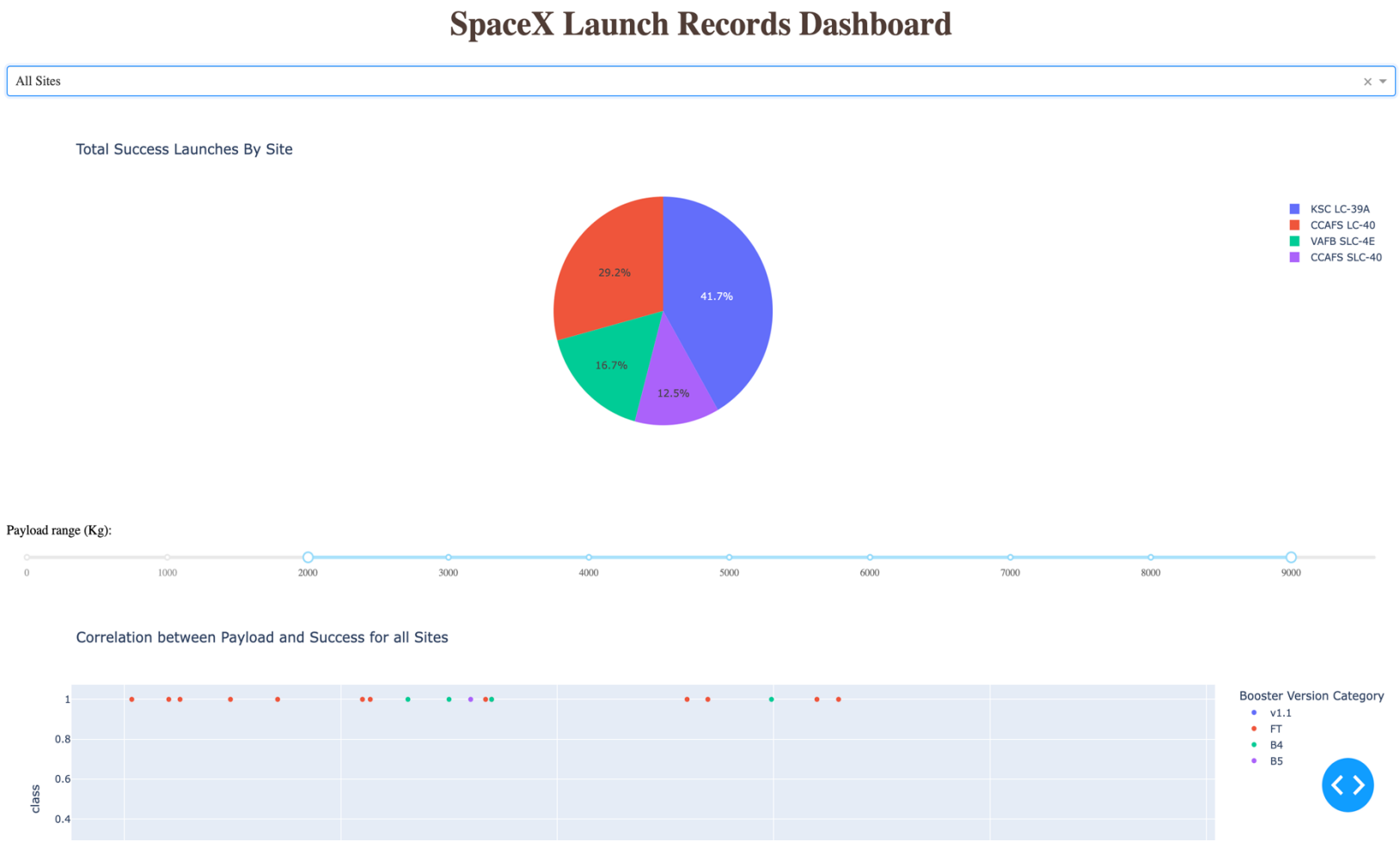
Results

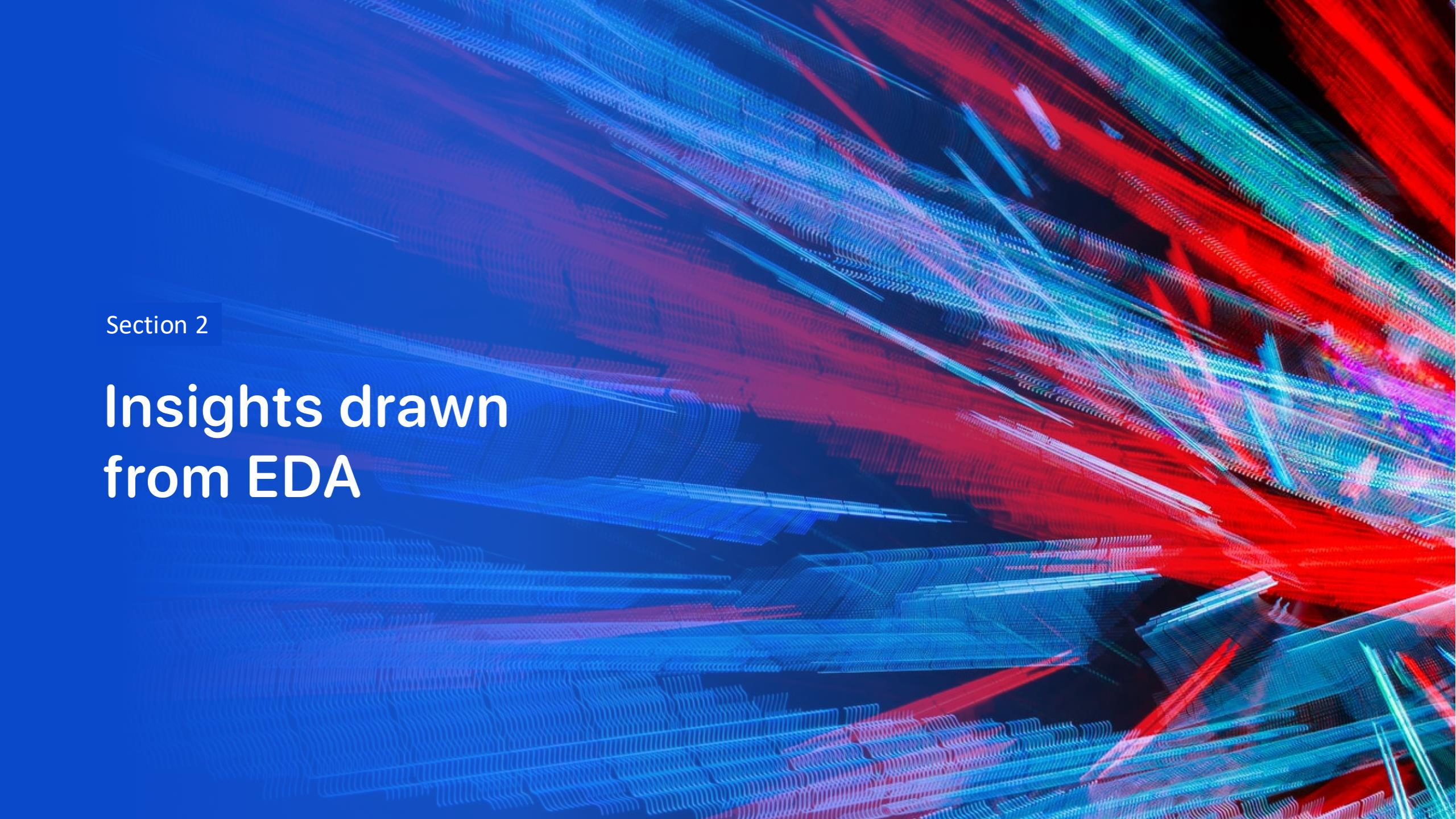
- Predictive analysis shows that the highest model accuracy was decision tree.

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.8333333333333334
SVM	0.84821	0.8333333333333334
Tree	0.8625	0.9444444444444444
KNN	0.84821	0.8333333333333334

Results

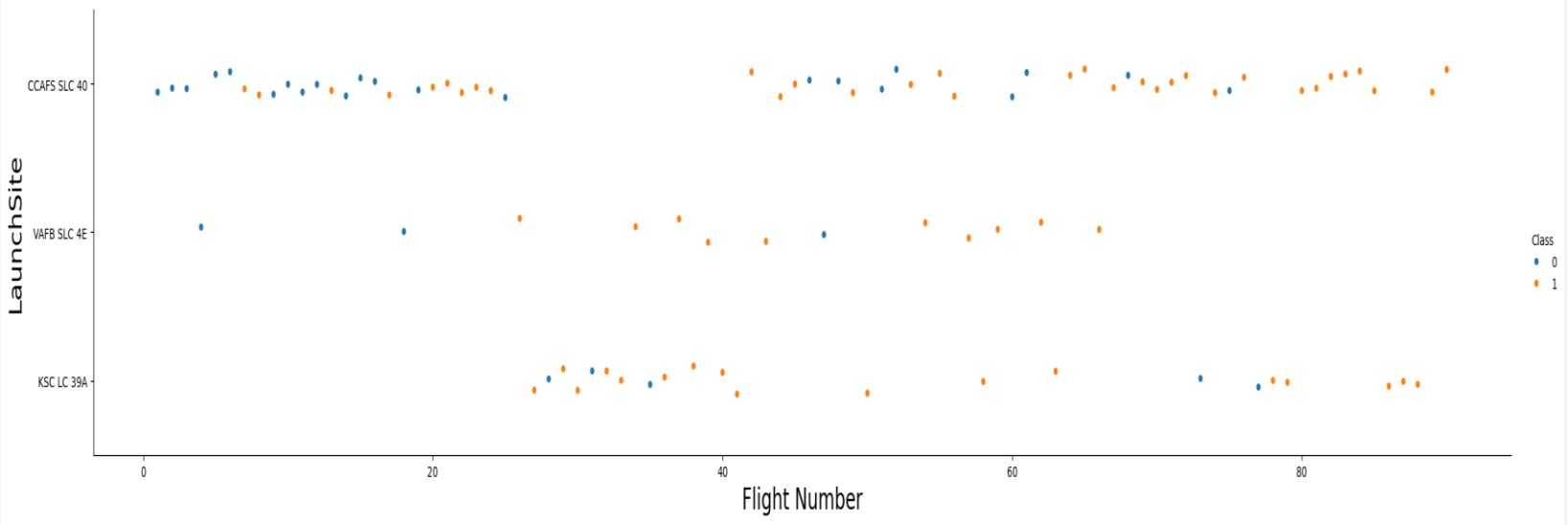
- Predictive analysis results



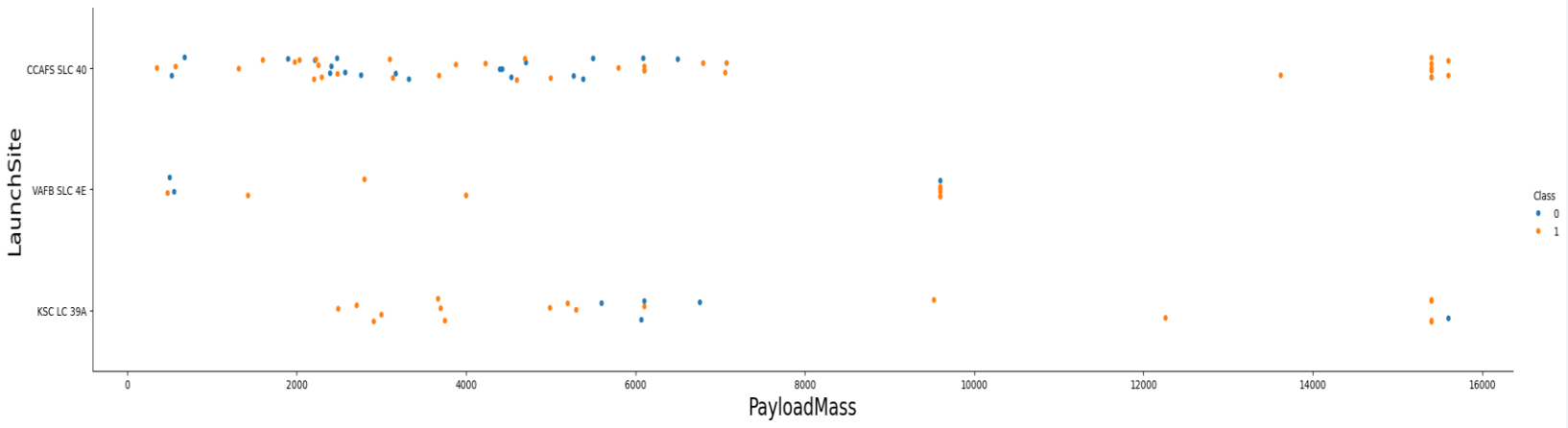

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- The scatterplot shows that CCAFS SLC 40 was the best launchsite after more than 80 flights and VAFB SLC 4E had the least number of launches.



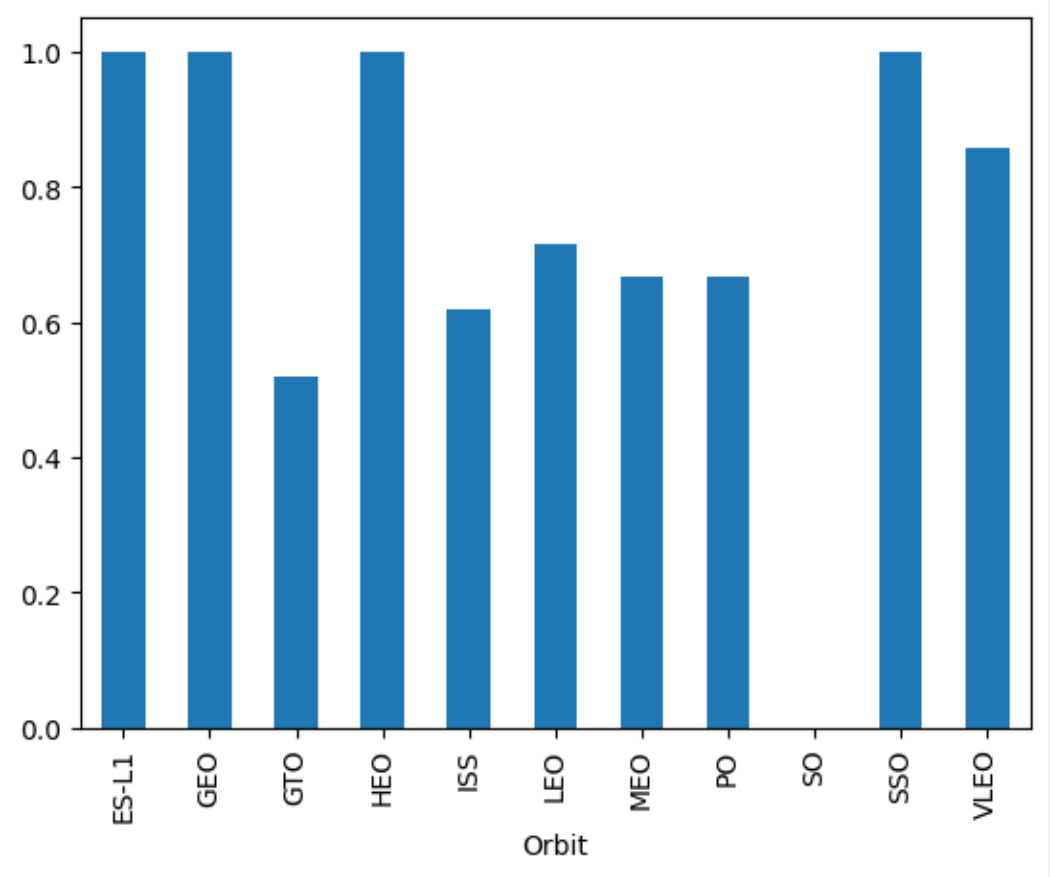
Payload vs. Launch Site



- The scatterplot shows that payload over 7000 have high success rate but less launches and KSC LC 39A have a great success rate under 6000.

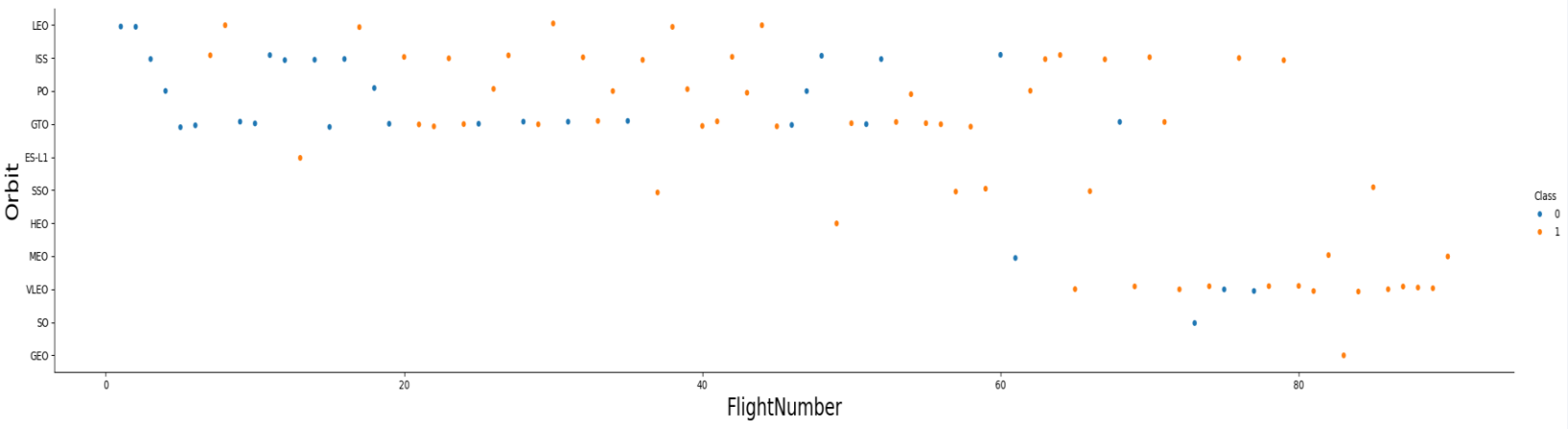
Success Rate vs. Orbit Type

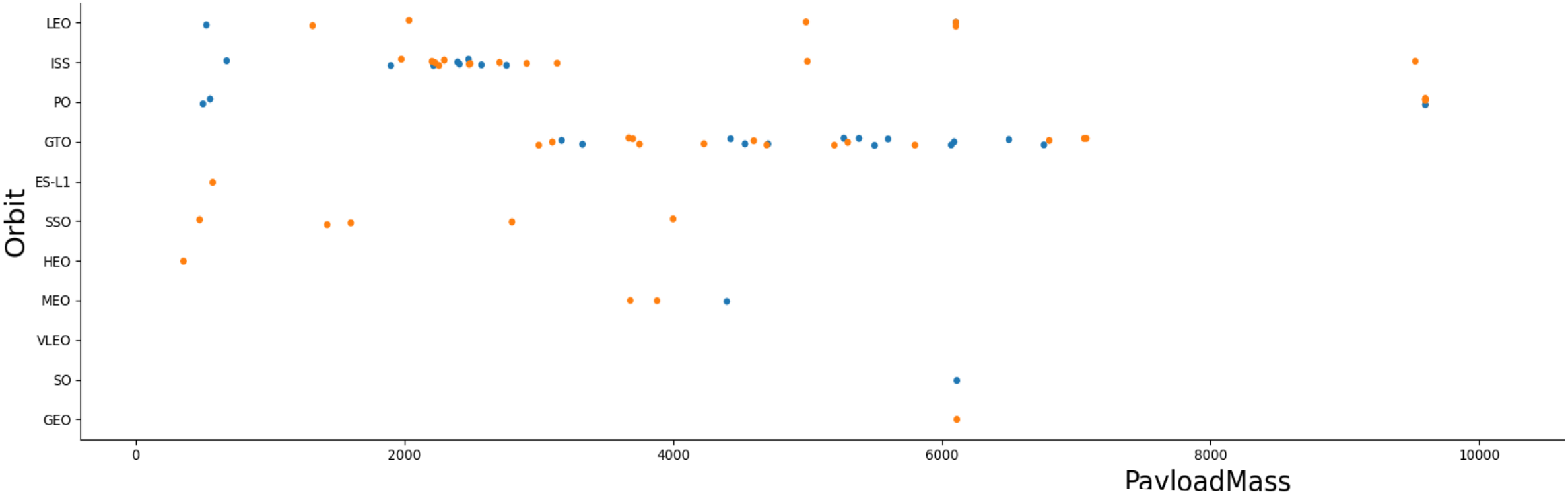
- The bar chart shows that ES-L1, GEO,HEO and SSO have a rate higher than 98%.
- SO have the least success rate.



Flight Number vs. Orbit Type

- VLEO and GTO had the highest success rate
- VLEO started launching after 60 flights.



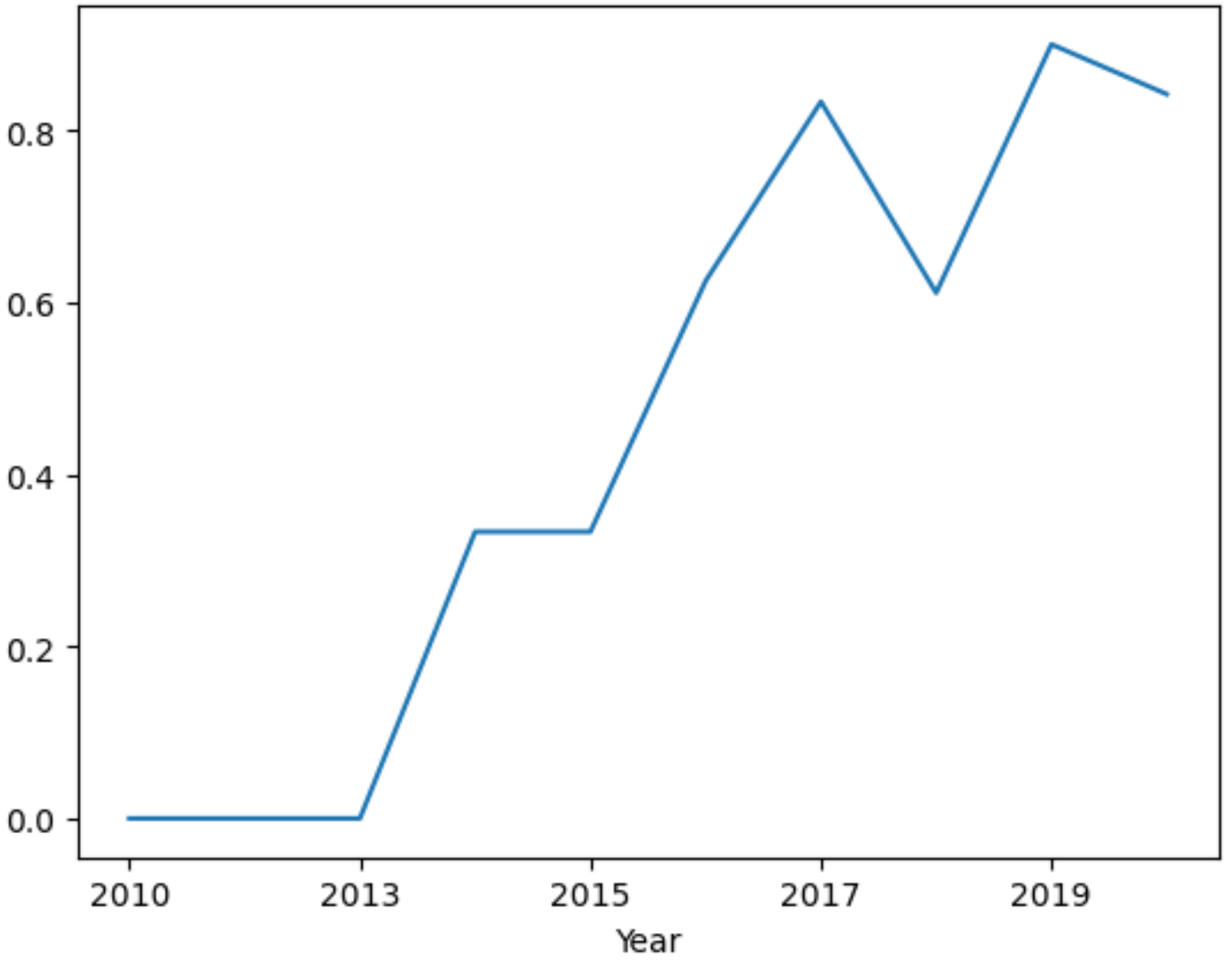


Payload vs. Orbit Type

- GEO and SO only had one launch
- GTO and ISS have the highest number of launch

Launch Success Yearly Trend

The line chart shows the average success rate over the years with the highest rate in 2019.



All Launch Site Names

- This the unique names of the 4 launch sites that was found using the following query
- sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;

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

Launch Site Names Begin with 'CCA'

- The following table shows 5 samples of launch sites that begins 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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- This is **total payload mass carried by boosters launched by NASA and was calculated by summing all the payloads that contains NASA as a customer.**

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

- The following output **Display average payload mass carried by booster version F9 v1.1 and was calculated using the specified booster version.**

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

- We Found the date of the first successful landing outcome on ground pad using the following query
- sql SELECT MIN(DATE) AS FIRST_SUCCESS_GP FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Success (ground pad)';

FIRST_SUCCESS_GP
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The output shows the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes is shown in the following output and it shows that there was 1 failure and 99 flight succussed.

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

Boosters Carried Maximum Payload

- The table shows the names of the booster which have carried the maximum payload mass.

Booster_Version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

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

Booster_Version	Launch_Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

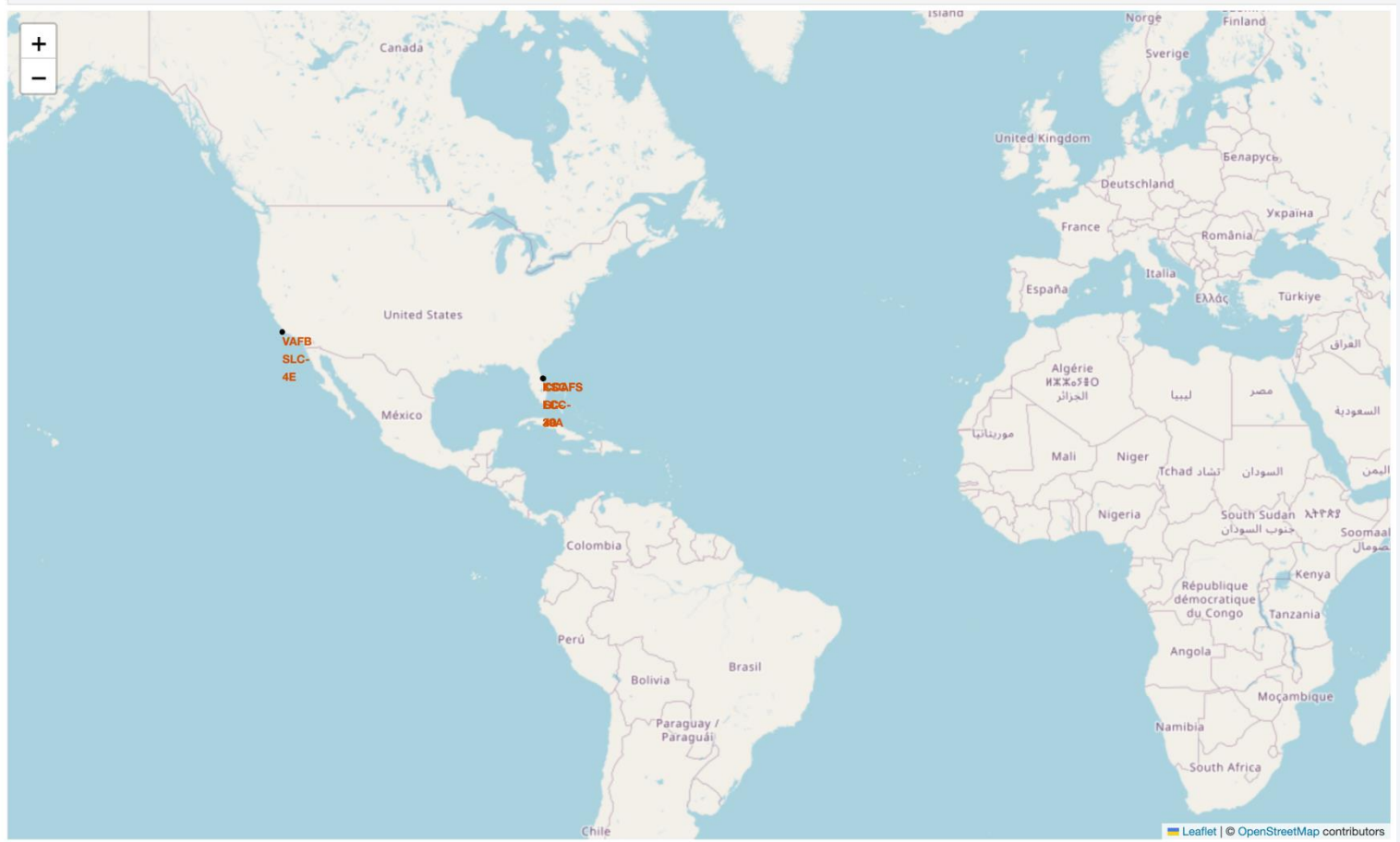
- The output shows the Ranking of the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

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

Section 3

Launch Sites Proximities Analysis

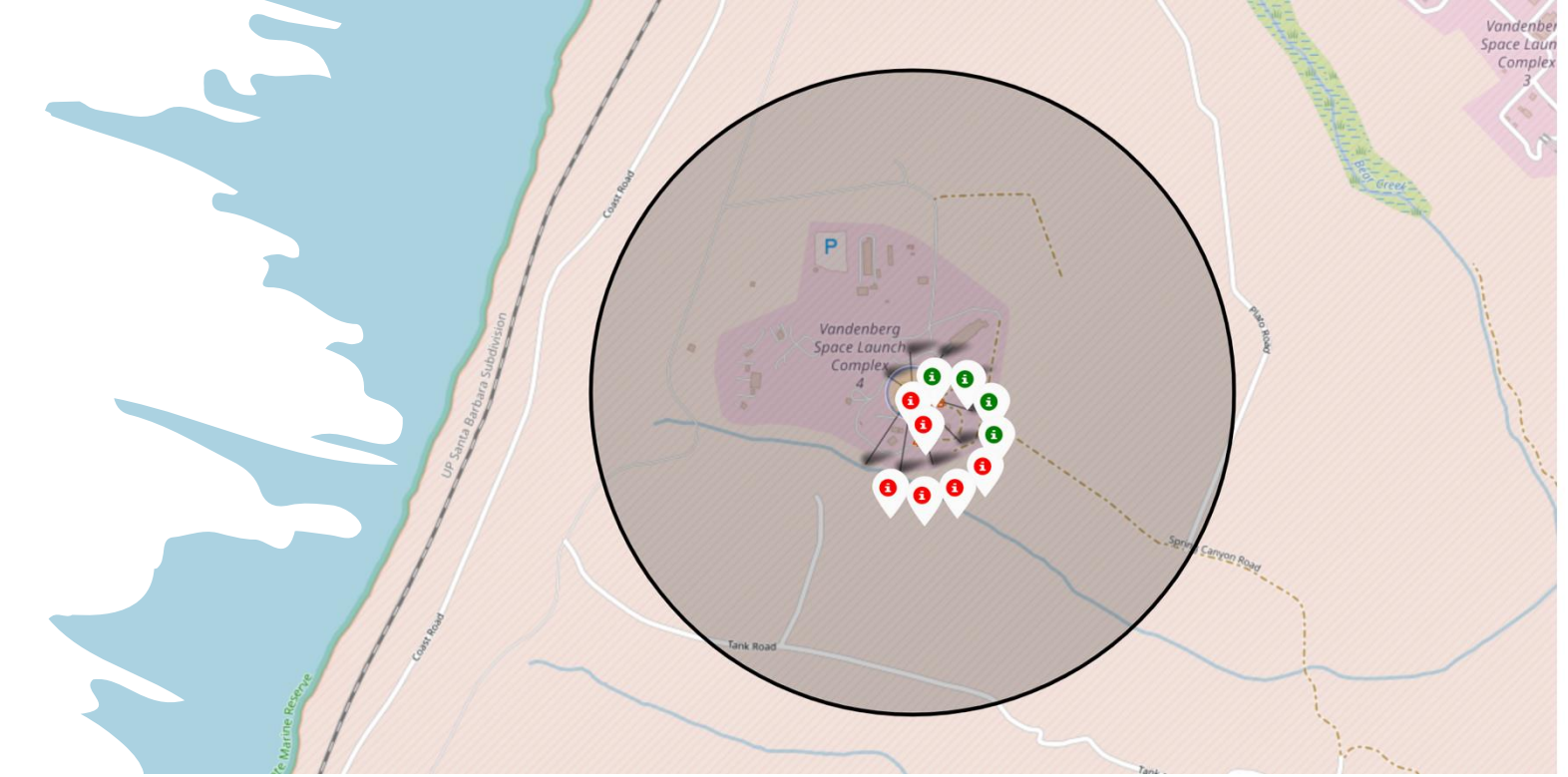
Launch Sites

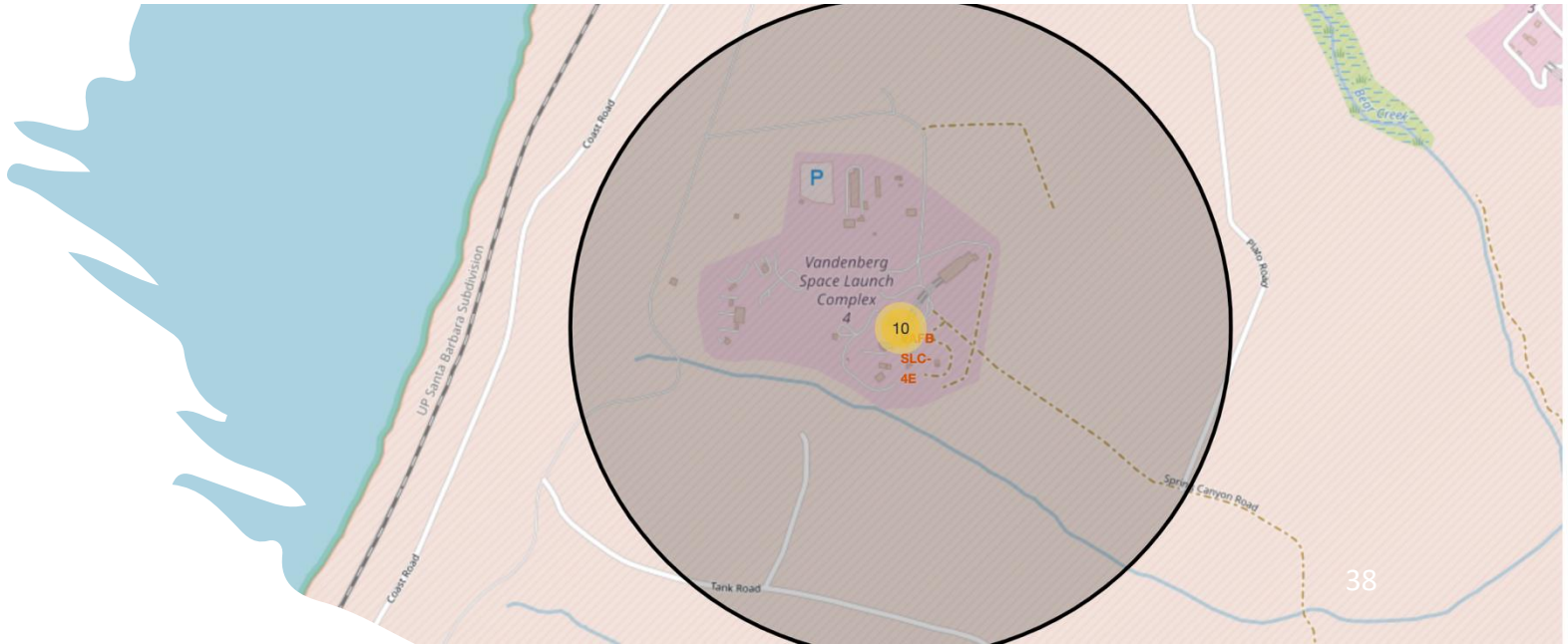


The map shows all the available launch sites.

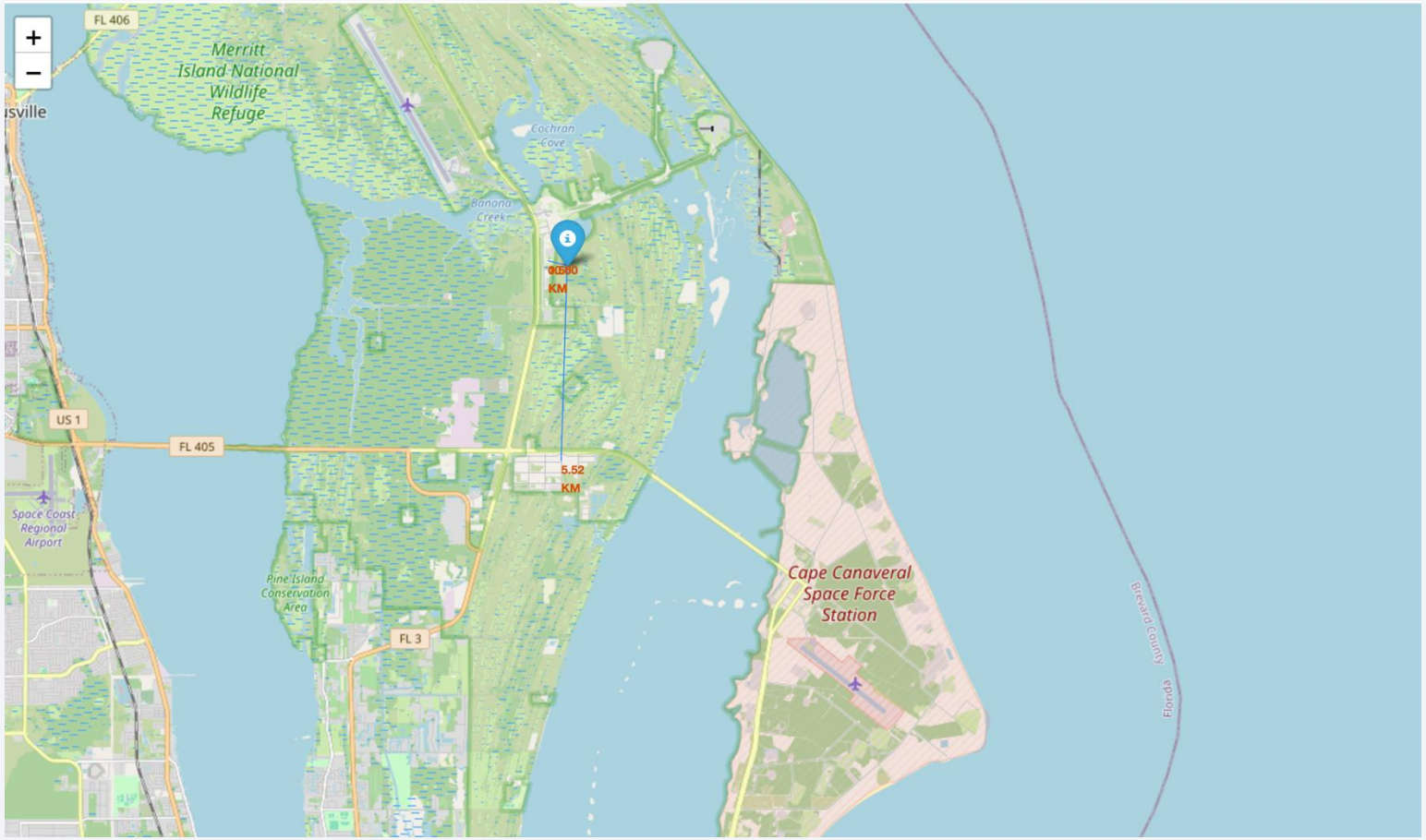
Launch site success rate

You can see that the following launch sites have higher failure rate than success while red is failure and green is success.

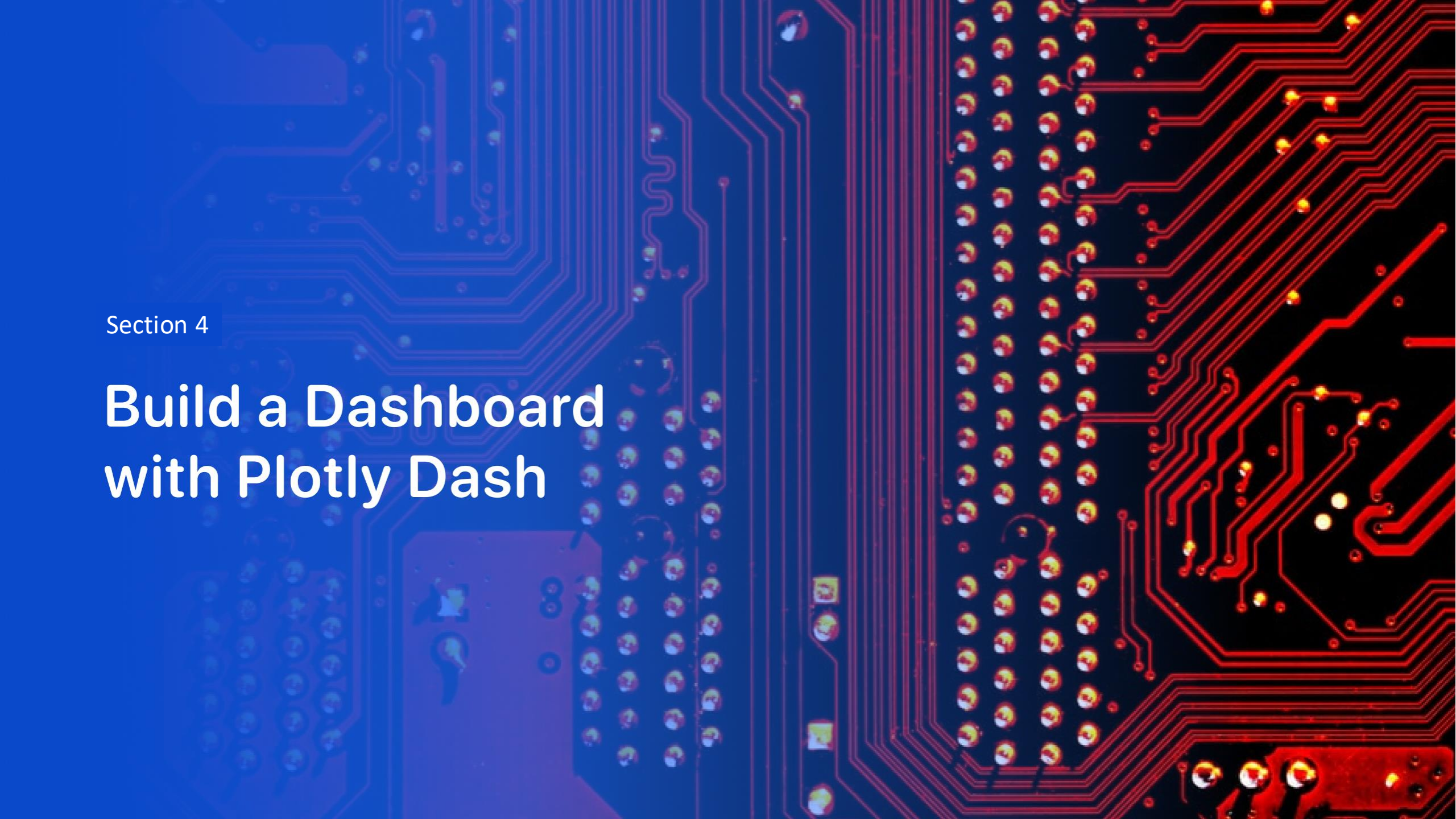


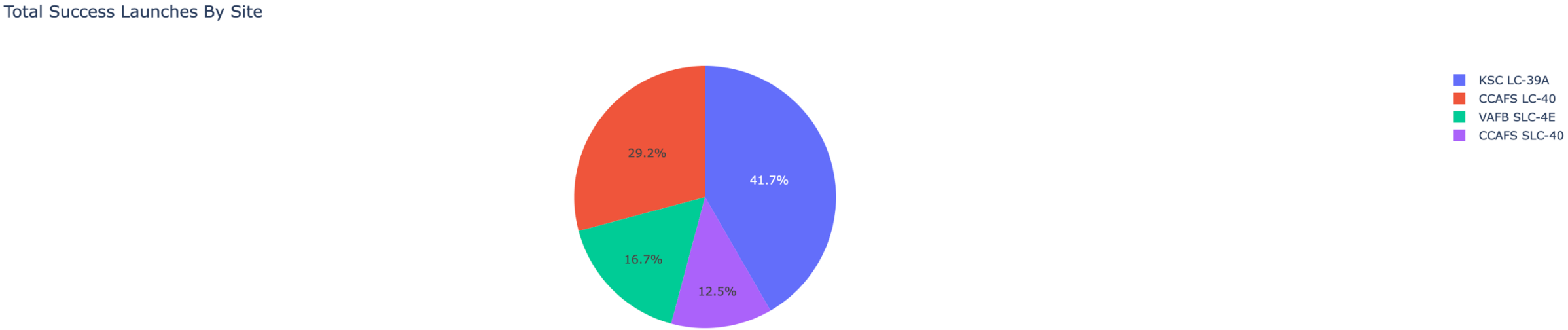


Launch site neighbourhood



The map show the selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

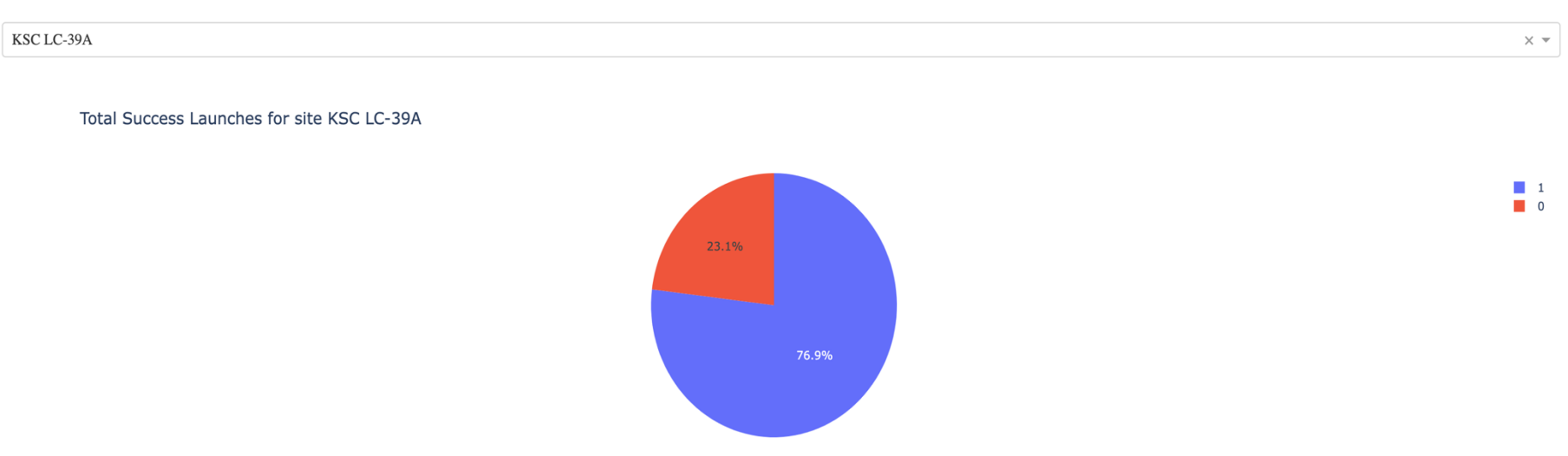

Section 4
Build a Dashboard with Plotly Dash



Launch success

- KSC LC-39A had the highest successful rate between all sites and CCAFS SLC-40 had the least successful launches

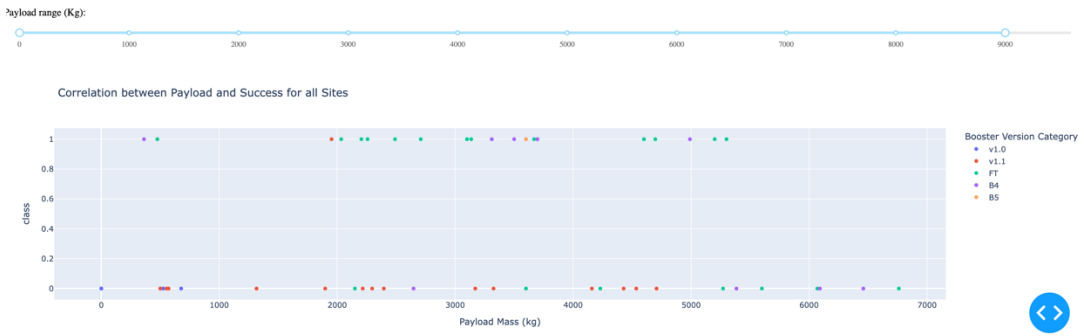
success launches rate



The KSC LC-39A had 76.9% of successful launches.

Payload and successful launches

The scatter plots show the different payload and booster versions and success rates



Section 5

Predictive Analysis (Classification)

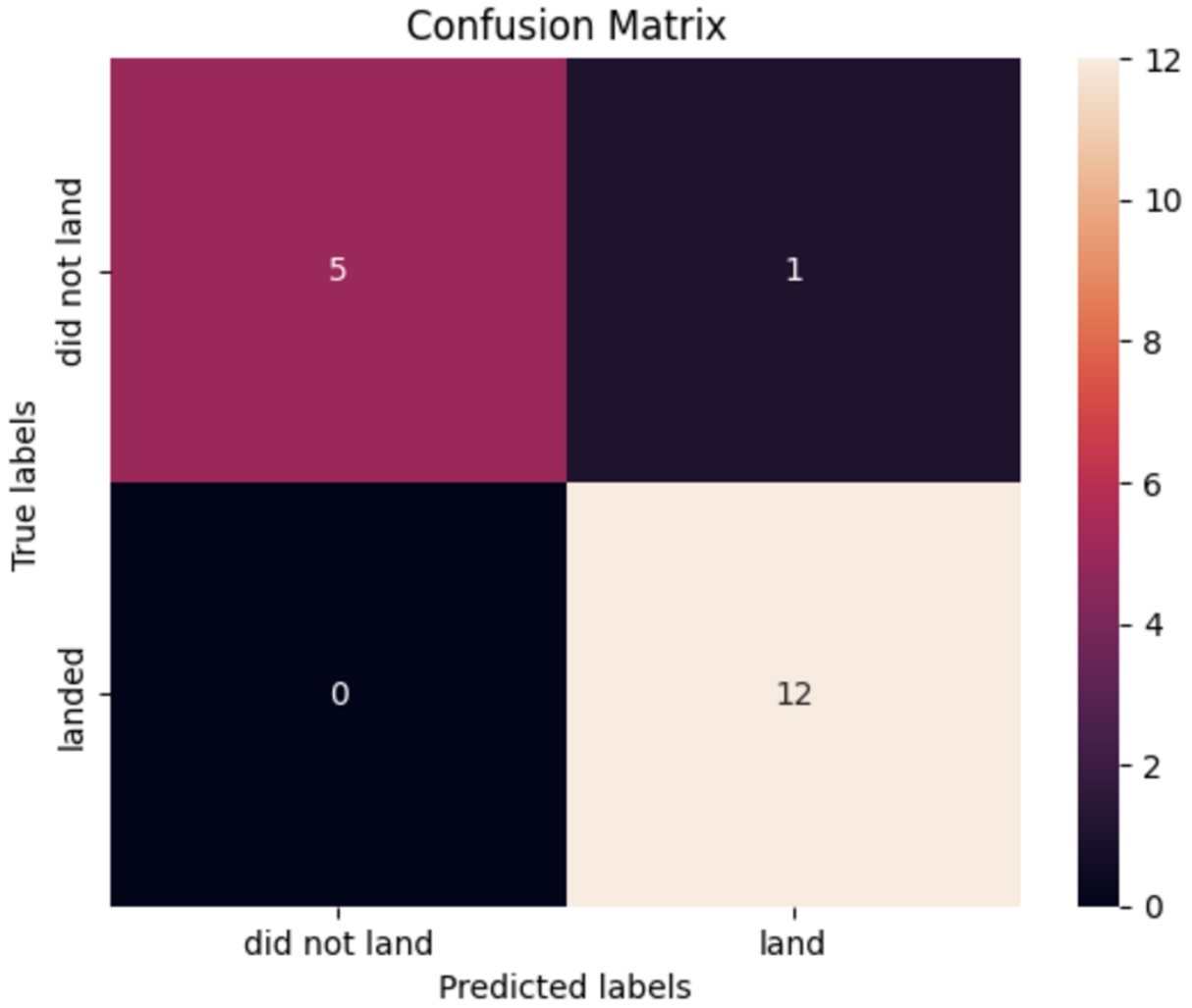
Classification Accuracy

- The model with the highest accuracy was decision tree.

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.8333333333333334
SVM	0.84821	0.8333333333333334
Tree	0.8625	0.9444444444444444
KNN	0.84821	0.8333333333333334

Confusion Matrix

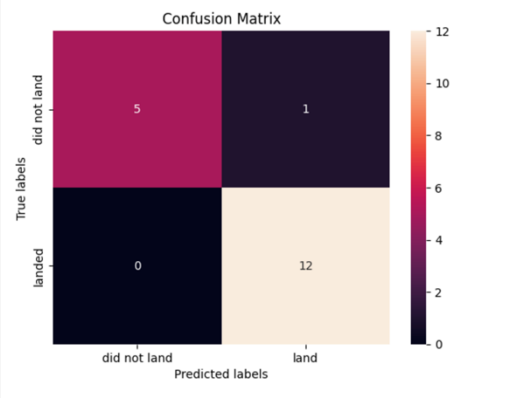
The confusion matrix of the decision tree shows that there was one wrong prediction only.

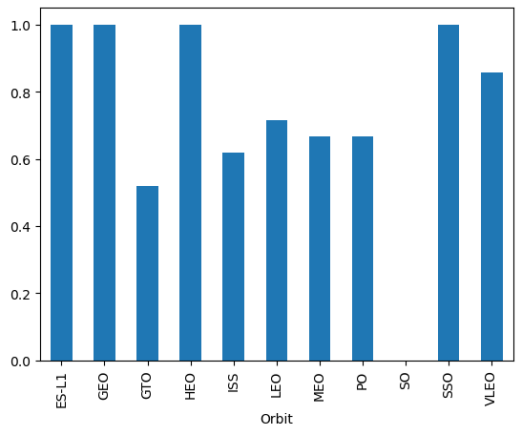


Conclusions

- When the weight increase the failure launch rate decrease.
- Decision tree had the highest accuracy but the other models wasn’t so far away.
- We discovered that there was one failure and 99 success launches.

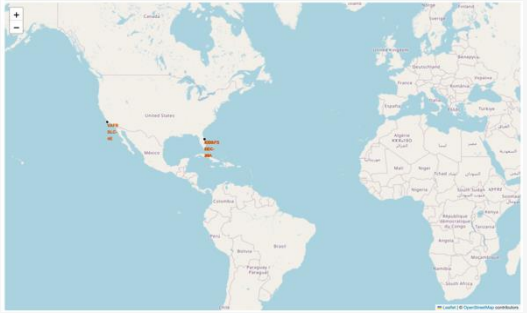
Appendix





FIRST_SUCCESS_GP

2015-12-22



Thank you!