IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Khairil Azmi Ashari
10 August 2022

Outline

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

Executive Summary

In this project we do data collection methodology using SpaceX API and Web scraping. Then, Perform data wrangling do data analysis, deal with missing value, and then create label for success/failure landing outcome. Perform exploratory data analysis (EDA) using visualization and SQL. Performing interactive visual analytics using Folium and Plotly Dash, then finally perform predictive analysis using classification models using Logistic Regression, Support Vector, Decision Tree, and K-Nearest to predict launch result and find which classification model give the best accuracy for result.

SpaceX launch sucess rate since 2013 kept increasing till 2020. ES-L1, GEO, HEO, and SSO orbit type had 100% success rate. KSC LC-39A had highest success rate with 76.9% success rate and it seen on the map data. Payload range 2k-5k is the optimize size payload for success land with FT Booster Version had the highest success count**.** Decision Tree had the highest best accuracy classification with 0.8889 or equal to 88.89% Accuracy.

Introduction

- Space X have project Starlink, a satellite internet constellation providing satellite Internet access. Sending manned missions to Space. One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars. other providers cost upwards of 165 million dollars each, much of the savings is because. SpaceX can reuse the first stage.
- Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

Methodology

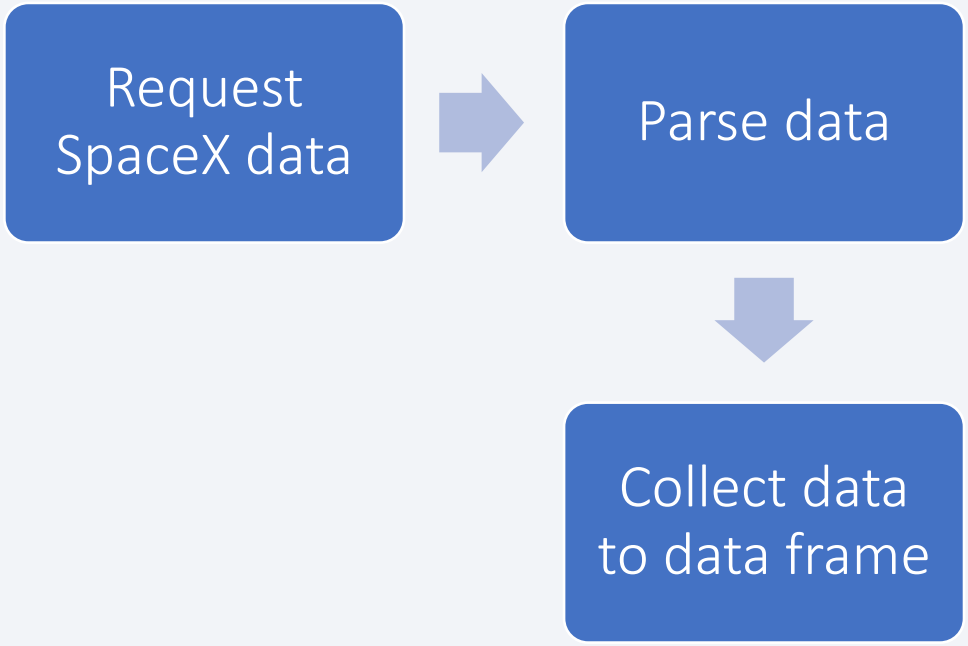
Methodology

- Executive Summary
- Data collection methodology:
  - Here we collect data using 2 method SpaceX API and Web scraping.
- Perform data wrangling
  - Here we do data analysis, deal with missing value, and then create label for success/failure 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
  - Here we use Logistic Regression, Support Vector, Decision Tree, and K-Nearest to predict launch result and find which classification model give the best accuracy for result.

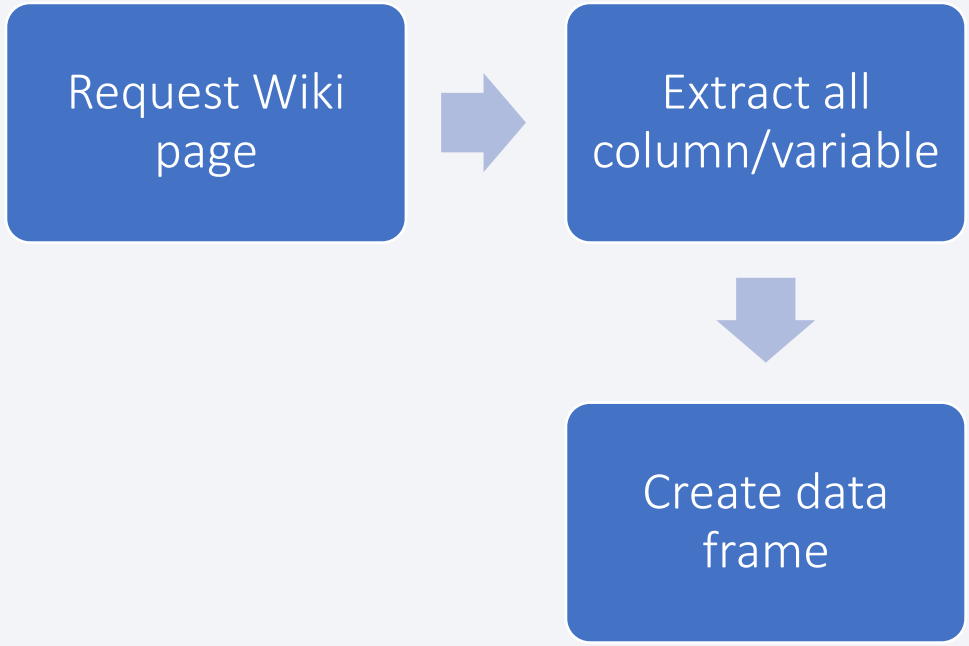
Data Collection

- Data sets were collected using SpaceX API and Web scraping from Wikipedia about List of Falcon 9 Falcon Heavy Launches.

SpaceX API process Flowchart



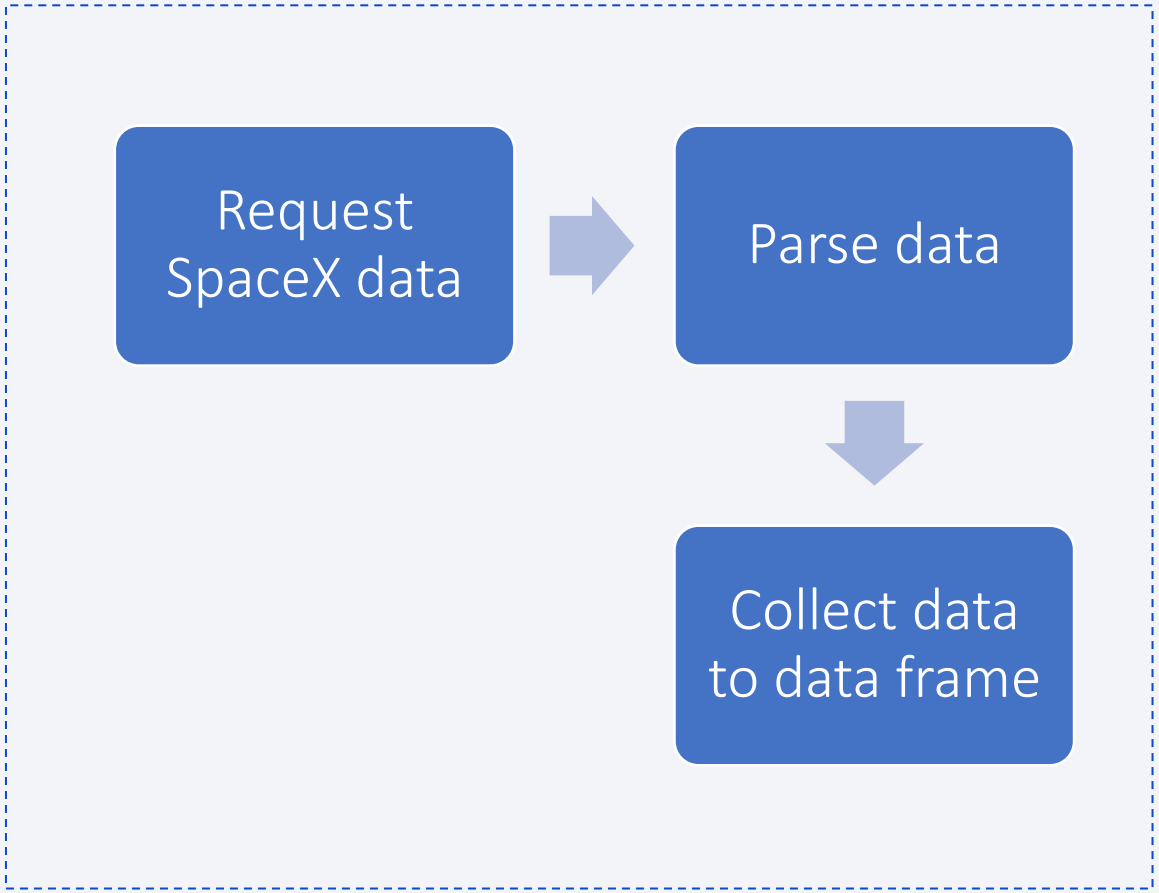
Web Scraping process Flowchart



Data Collection – SpaceX API

- Link Process

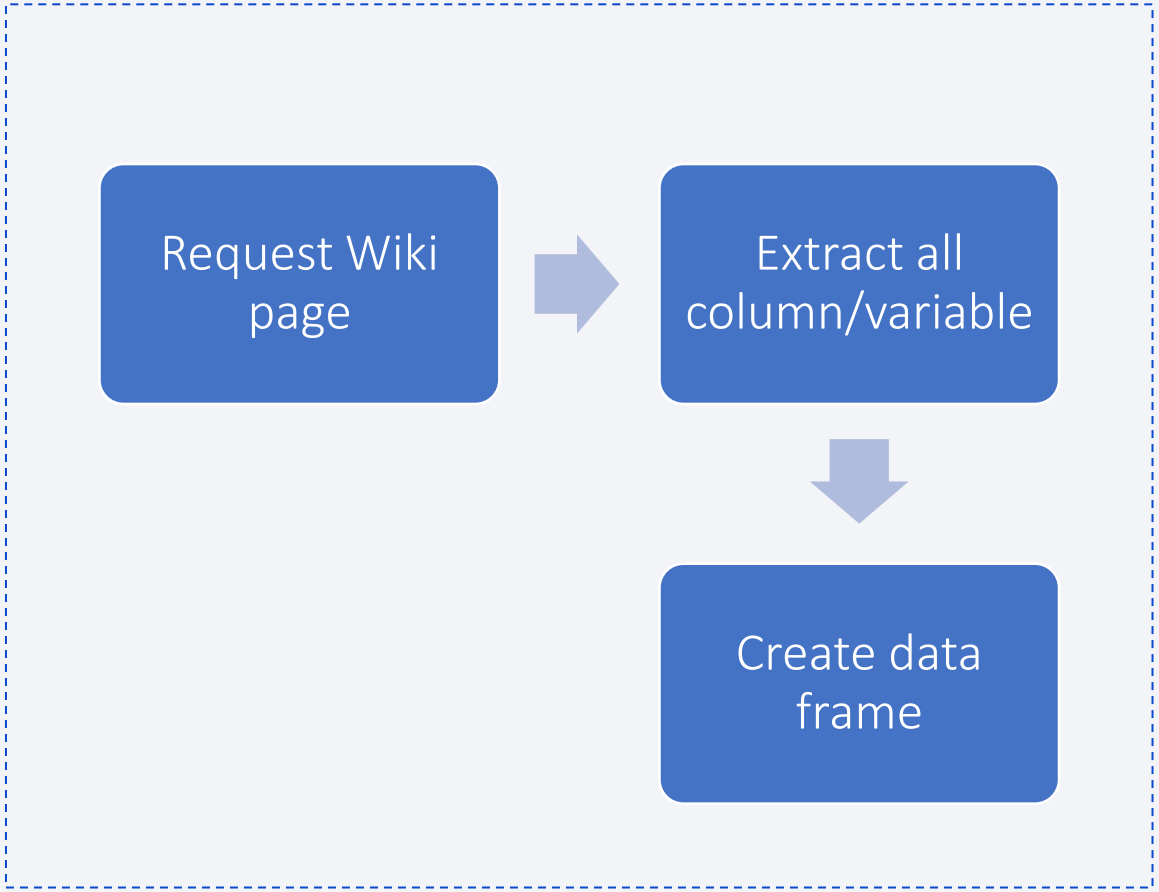
[Data Collection - SpaceX API]



Data Collection - Scraping

- Link process :

Data Collection - Web scraping process

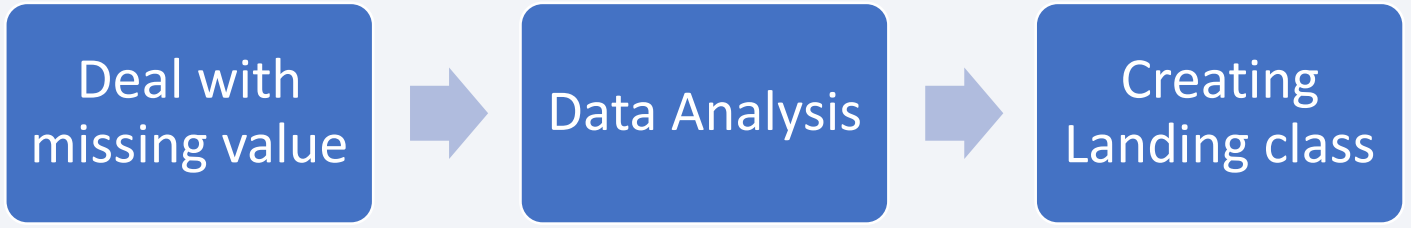


Data Wrangling

- In data wrangling, We do Data Analysis first we detect any missing value then dealing with it. We do 'mean' with missing Payload Mass value, We do nothing on Landing Pad because we don't need it, then we count every value to see unique Launch site, Orbit, Outcome . Finally Create a landing outcome (Class) label from Outcome column to inform success or failure landing.

- Data Wrangling process:

Data Wrangling Process



EDA with Data Visualization

- First we plot `Flight Number` vs. `Payload Mass` as the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.
- Launch site vs Flight Number, show as the flight number above 80 all launches identify as success launch. In Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
- ES-L1, GEO, HEO, and SSO orbit type had 100% success rate. On the other hand SO orbit had 0% success rate. LEO orbit the Success appears related to the number of flights.
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS. And finally from last plot you can observe that the sucess rate since 2013 kept increasing till 2020
- EDA with Data Visualization Proccess

EDA with SQL

- First we do Mark all launch sites on U.S. map to see position of launch sites, Then Mark the success/failed launches for each sites to see launch result from map, and last we calculate the distances between a launch site to its proximities.
- All the proximities created are closest coast, highway, railroad, and city, to see how launch sites created far enough from surrounding, to keep safety of people.
- Interactive Map with Folium Process

Build an Interactive Map with Folium

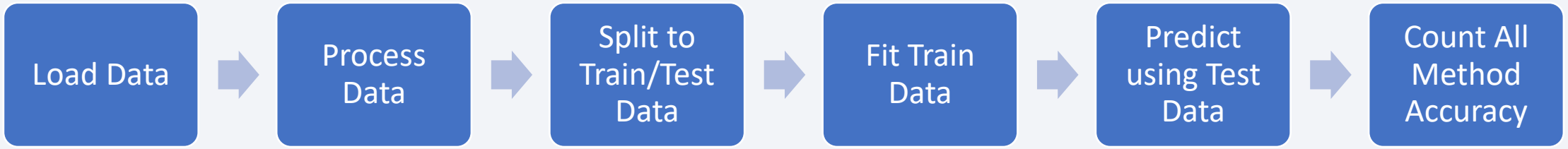
- First we do Mark all launch sites on U.S. map to see position of launch sites, Then Mark the success/failed launches for each sites to see launch result from map, and last we calculate the distances between a launch site to its proximities.
- All the proximities created are closest coast, highway, railroad, and city, to see how launch sites created far enough from surrounding, to keep safety of people.
- Interactive Map with Folium Process

Build a Dashboard with Plotly Dash

- In Dashboard we insert Pie Chart showing landing success rate for All or specific launching site according to user select on dropdown menu.
- And below it, we insert Scatter plot showing relation between Payload mass range to launch success based on slider range user input and launching site/all user selected before.
- Plotly Dash App Source Code

Predictive Analysis (Classification)

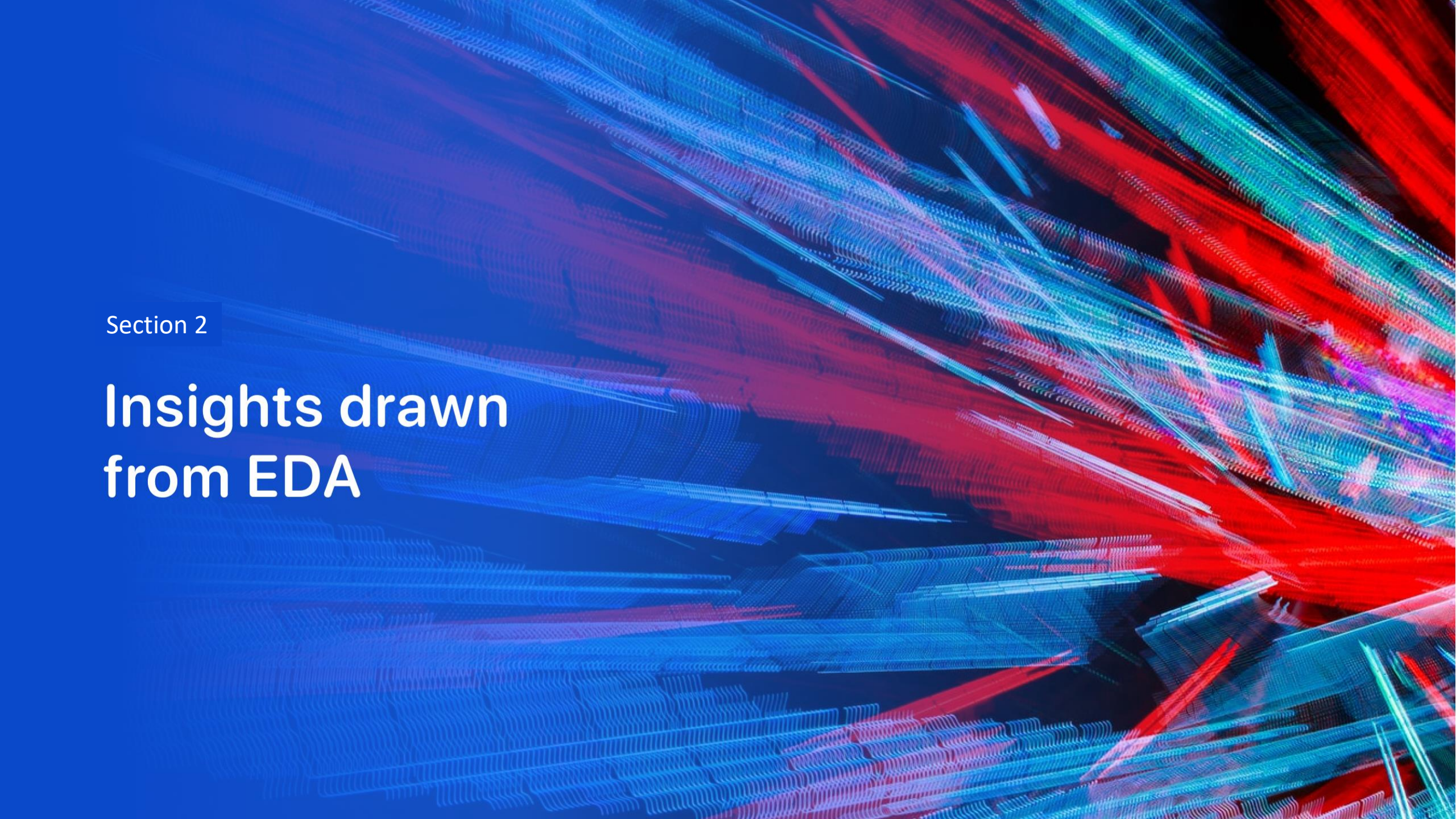
- First we load data that already collected and wrangled, than store Class data to Y and Features data to X, standardize X data. Split X and Y data to Train and Test data. We finish prepared data before do predictive analysis (classification).
- We Fit Train data using Logistic Regression, Support Vector Machine, Decision Tree, K-Nearest Neighbors classification method. Do predict on Test data. And Finally we count Predict Accuracy Score to see the best Classification method to do Classification.

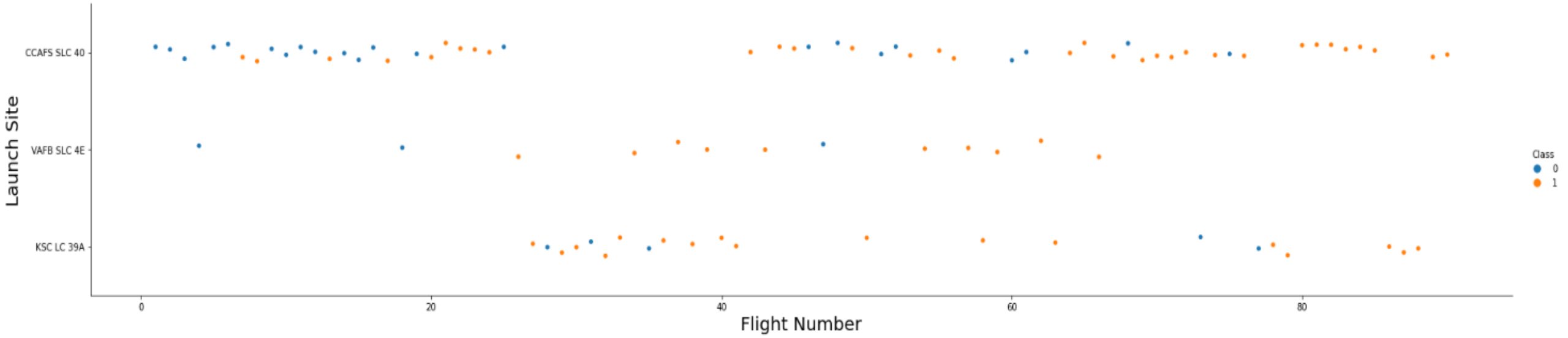


- Predictive Analysis/Classification process

Results

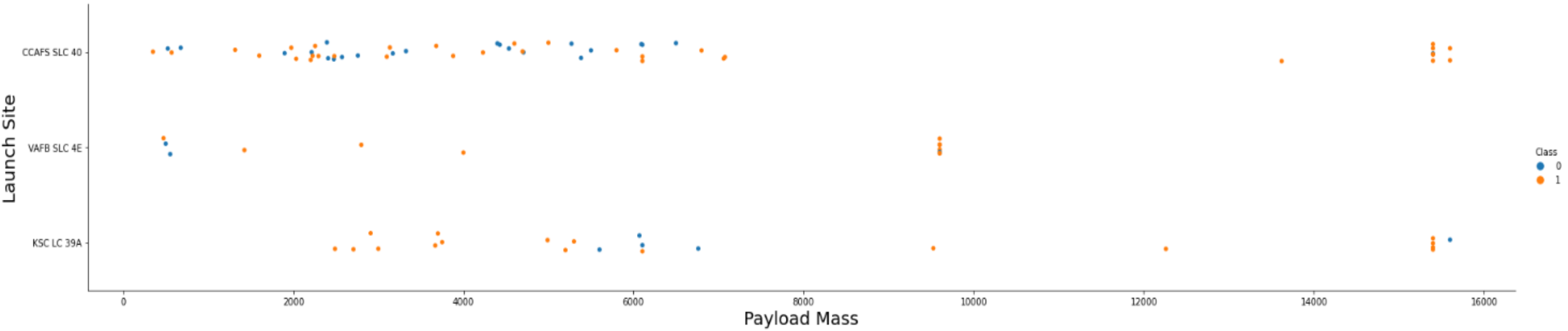
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA



Flight Number vs. Launch Site

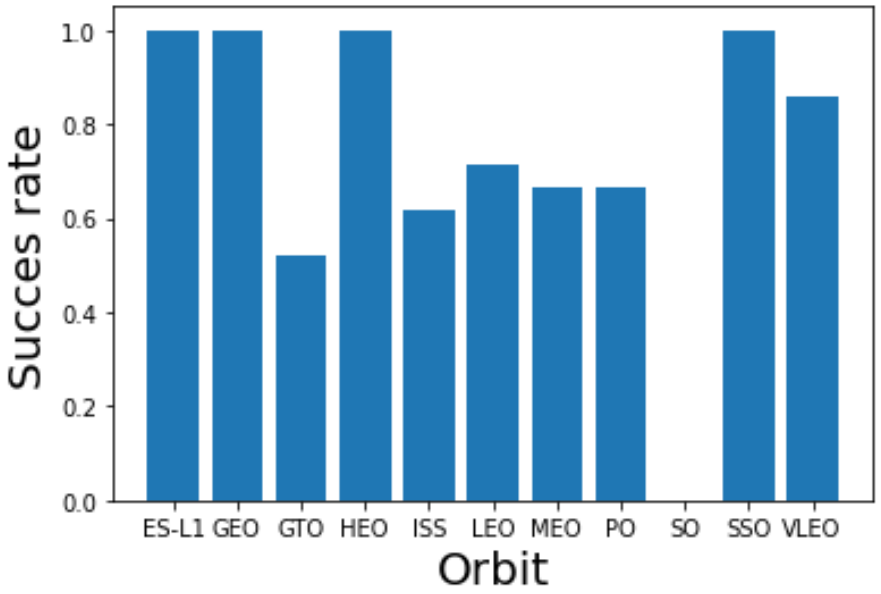
- As the flight number above 80 all launches identify as class 1 (yellow dot) which mean success launch.



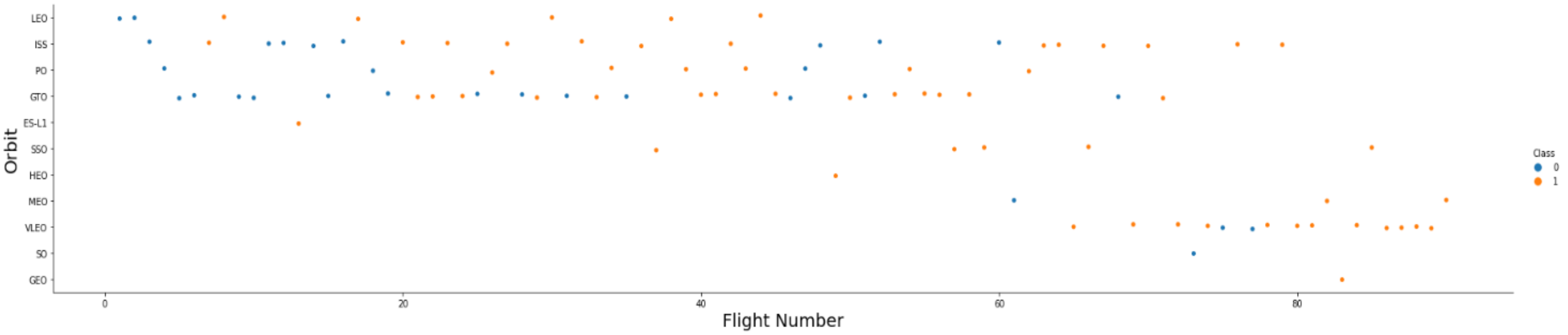
Payload vs. Launch Site

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

Success Rate vs. Orbit Type

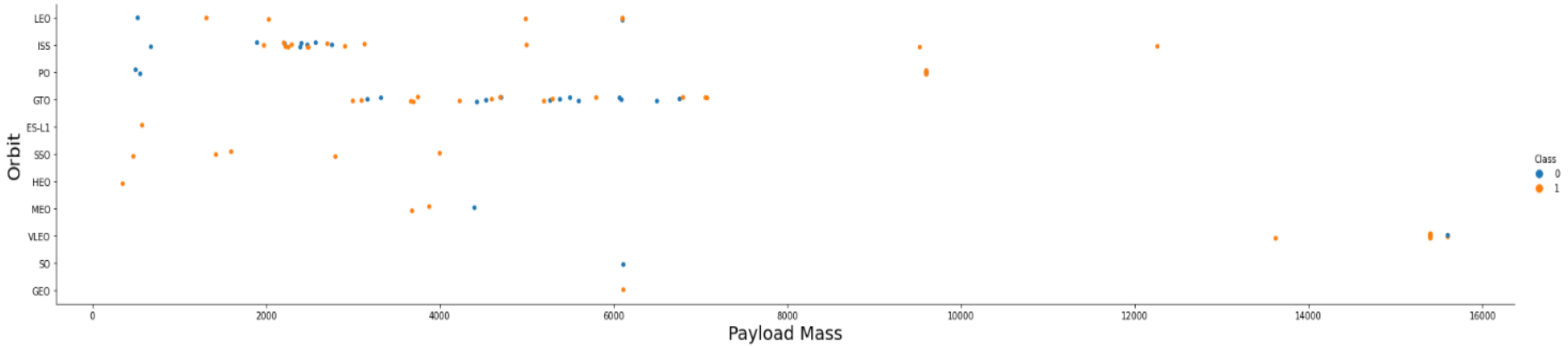


- We can see ES-L1, GEO, HEO, and SSO orbit type had 100% success rate.
- On the other hand SO orbit had 0% success rate.



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

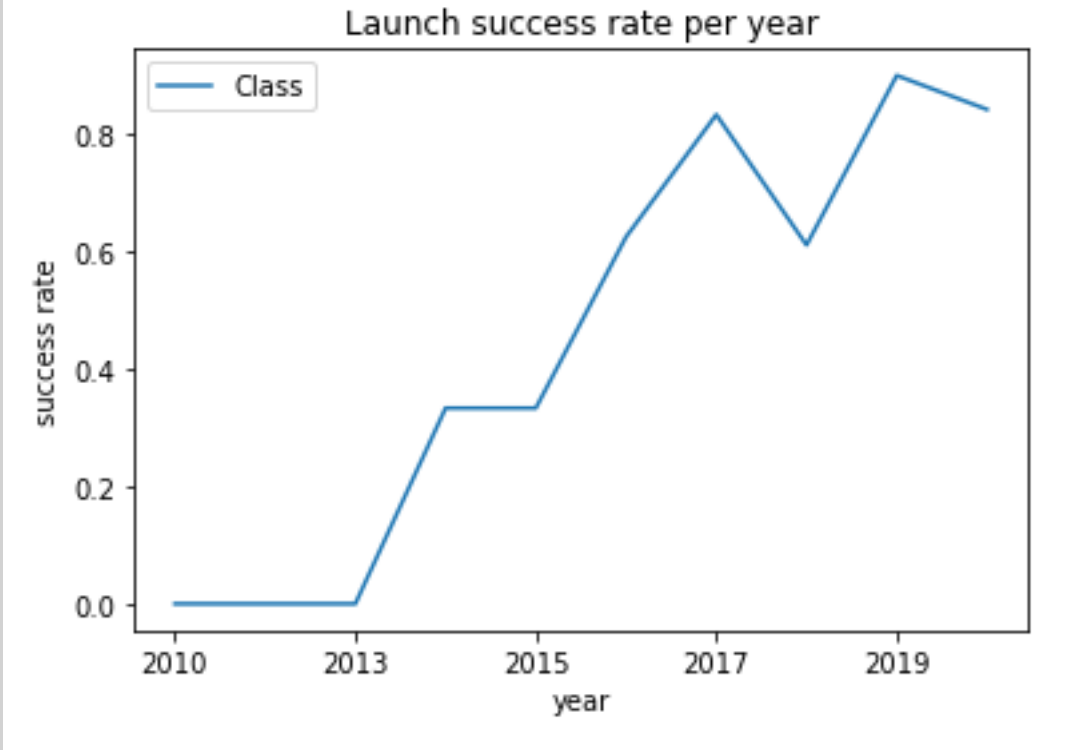


Payload vs. Orbit Type

- 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

- From Yearly success rate we got Space X success launch rate increase from 2013.
- And kept increasing until 2020.



All Launch Site Names

- Here is All launch sites name on SPACEX data

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Launch Site Names Begin with 'CCA'

- Showing list of 5 First Launch Site which Begin with 'CCA'

Total Payload Mass

- Total Payload Mass was carried by NASA (CRS) is 45596 kg.

Total Payload Mass(kg)
45596

Average Payload Mass by F9 v1.1

- Average Payload mass for F9 v1.1 is 2928.4 kg.

F9 v1.1 AVG Payload Mass(kg)
2928.4

First Successful Ground Landing Date

- From data gather we got, SpaceX First successful Ground Landing happen on 1 May 2017.

First Succesful Landing(Date)
01-05-2017

Booster Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Successful Drone Ship Landing with Payload between 4000 and 6000

- Booster Version F9 FT be the most success drone ship landing between 4000 and 6000 Payload mass.

Total Number of Successful and Failure Mission Outcomes

- Almost All Mission Outcome count as success, only 1 failure happen what fascinating.

Mission Outcome	Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Booster Version F9 B5 carried all maximum payload

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

2015 Failed Launch Records

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

On Failed Launch Records shown on 2015 shows:

- Booster Version F9 v1.1 and Launch site CCAFS LC-40 it could be the problem in Failed Launch

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

- From table on the right, it show Landing outcomes between 4 June 2010 and 20 march 2017.
- Between that date we can infer in a lot of success landing happen only a few fail happen between that 7 year .

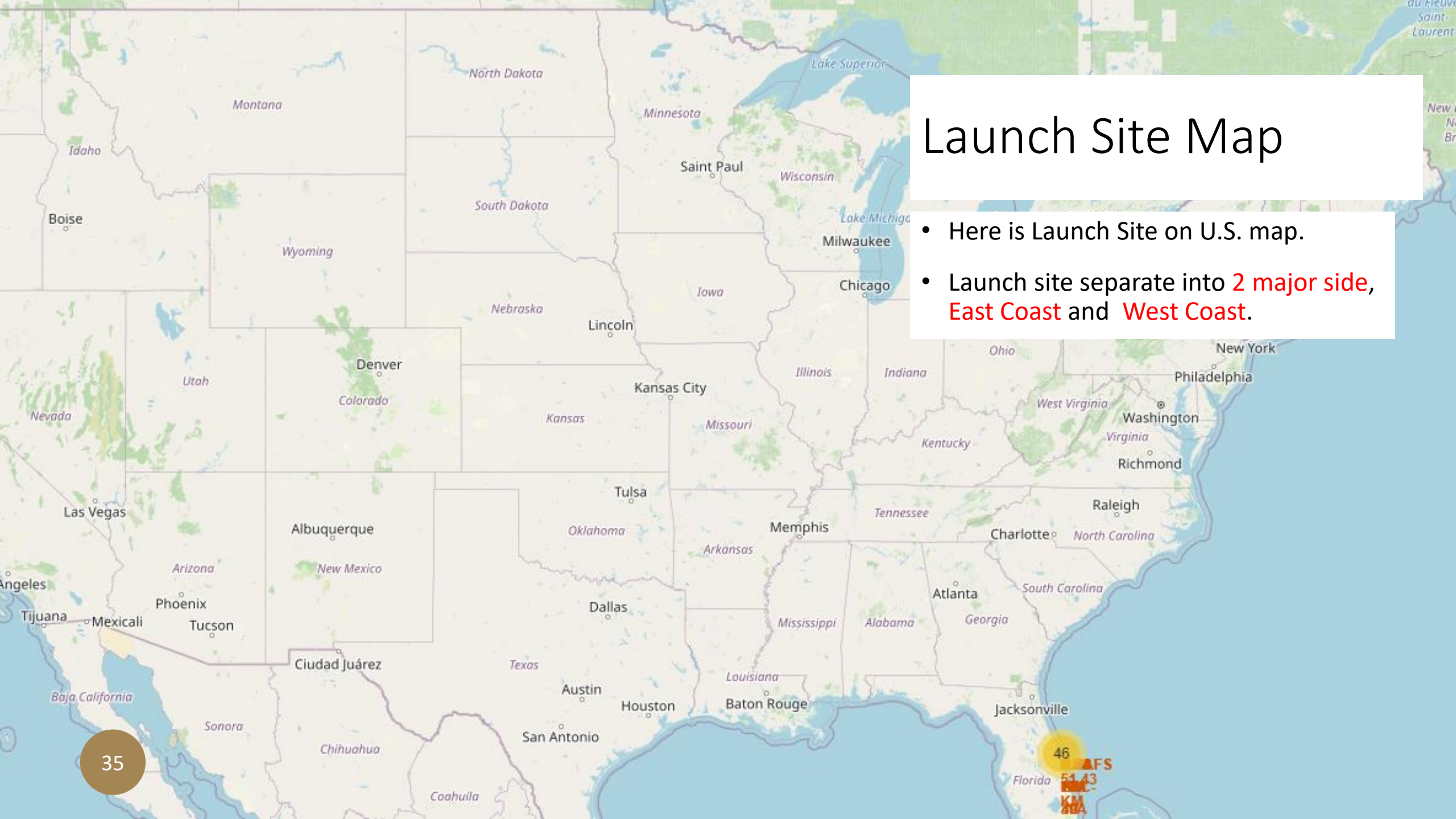
Landing Outcomes	Landing _Outcome
21	Success
10	No attempt
8	Success (drone ship)
6	Success (ground pad)
4	Failure (drone ship)
3	Failure
3	Controlled (ocean)
2	Failure (parachute)
1	No attempt

Section 3

Launch Sites Proximities Analysis

Launch Site Map

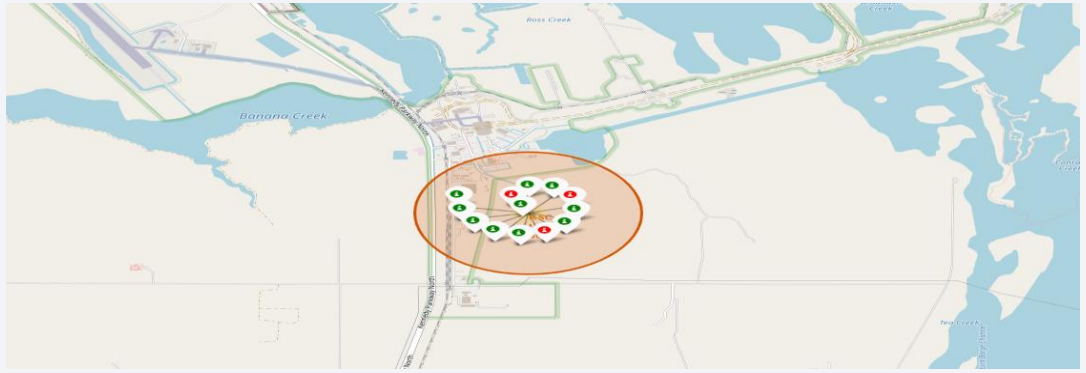
- Here is Launch Site on U.S. map.
- Launch site separate into 2 major side, East Coast and West Coast.



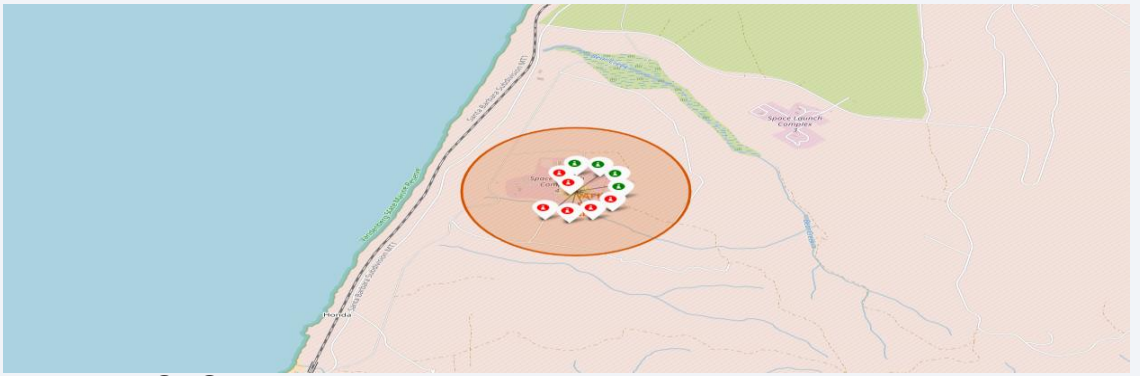
Launch Outcomes each Sites

CCA SLC-40

CCA LC-40

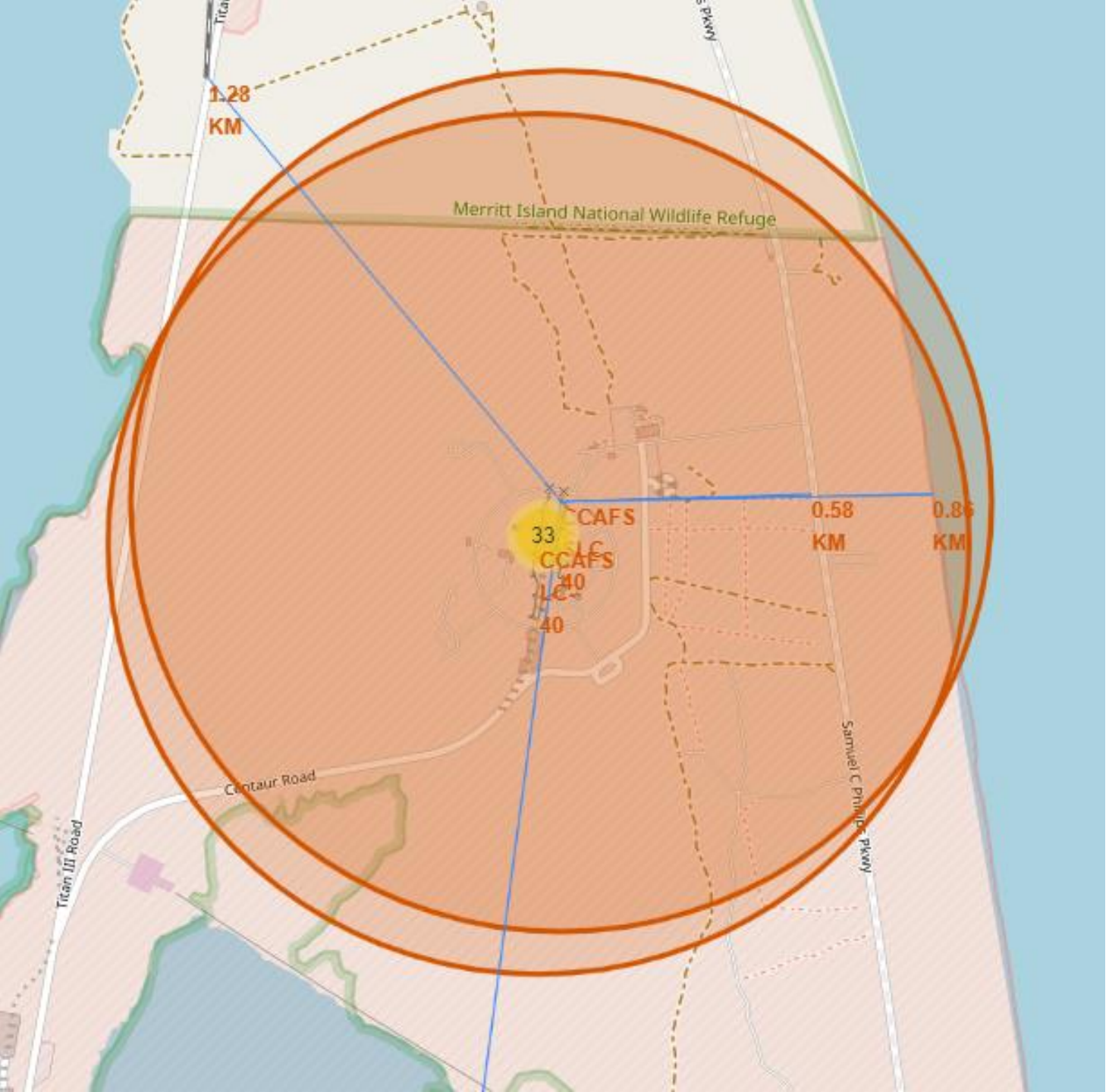


KSC LC-39A



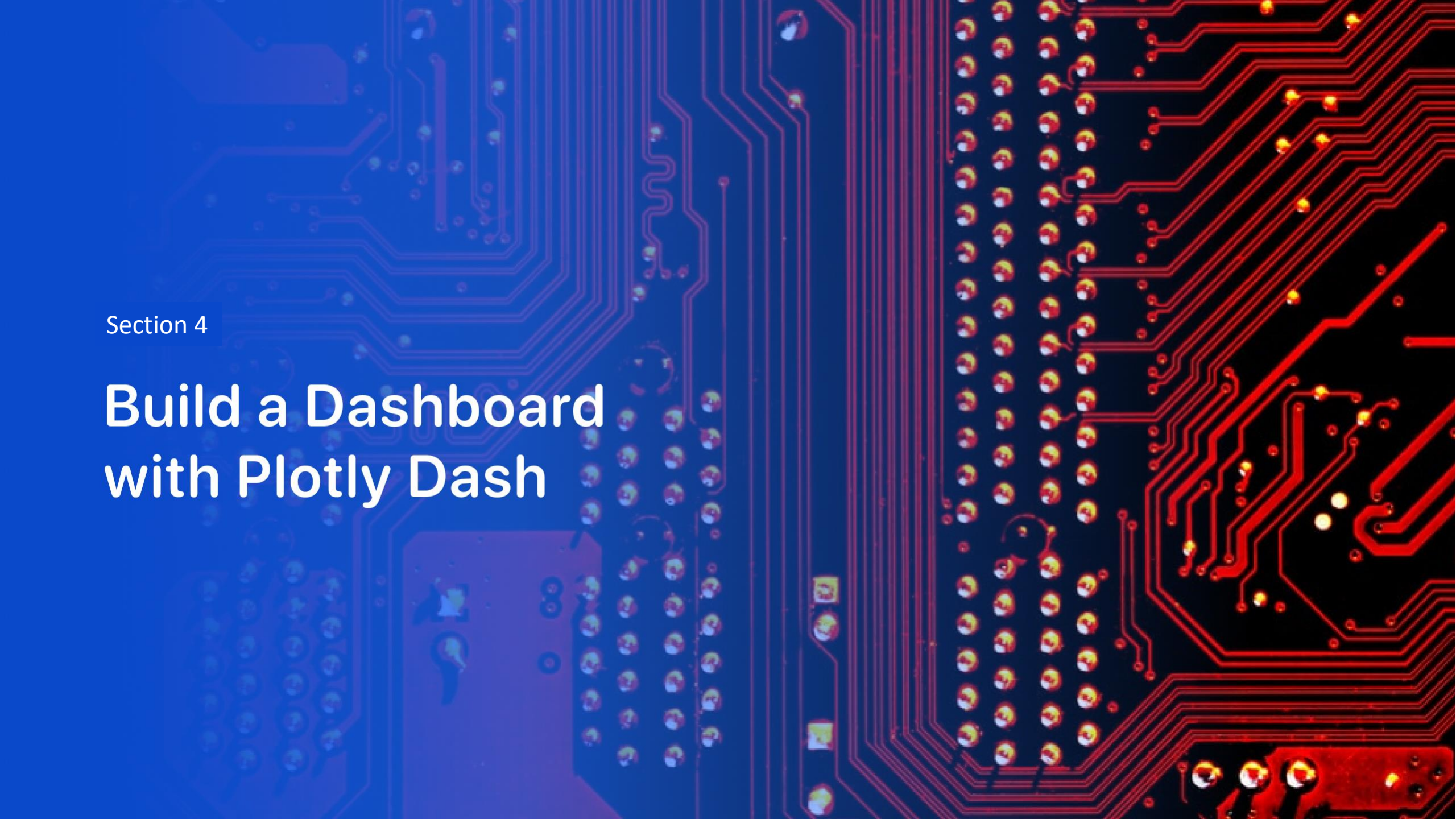
VAFB SLC-4E

- Here we got Launch Outcomes on every Launch Sites
- CCA SLC-40 had the Least Launch Success
- KSC LC-39A had the Most Launch Success



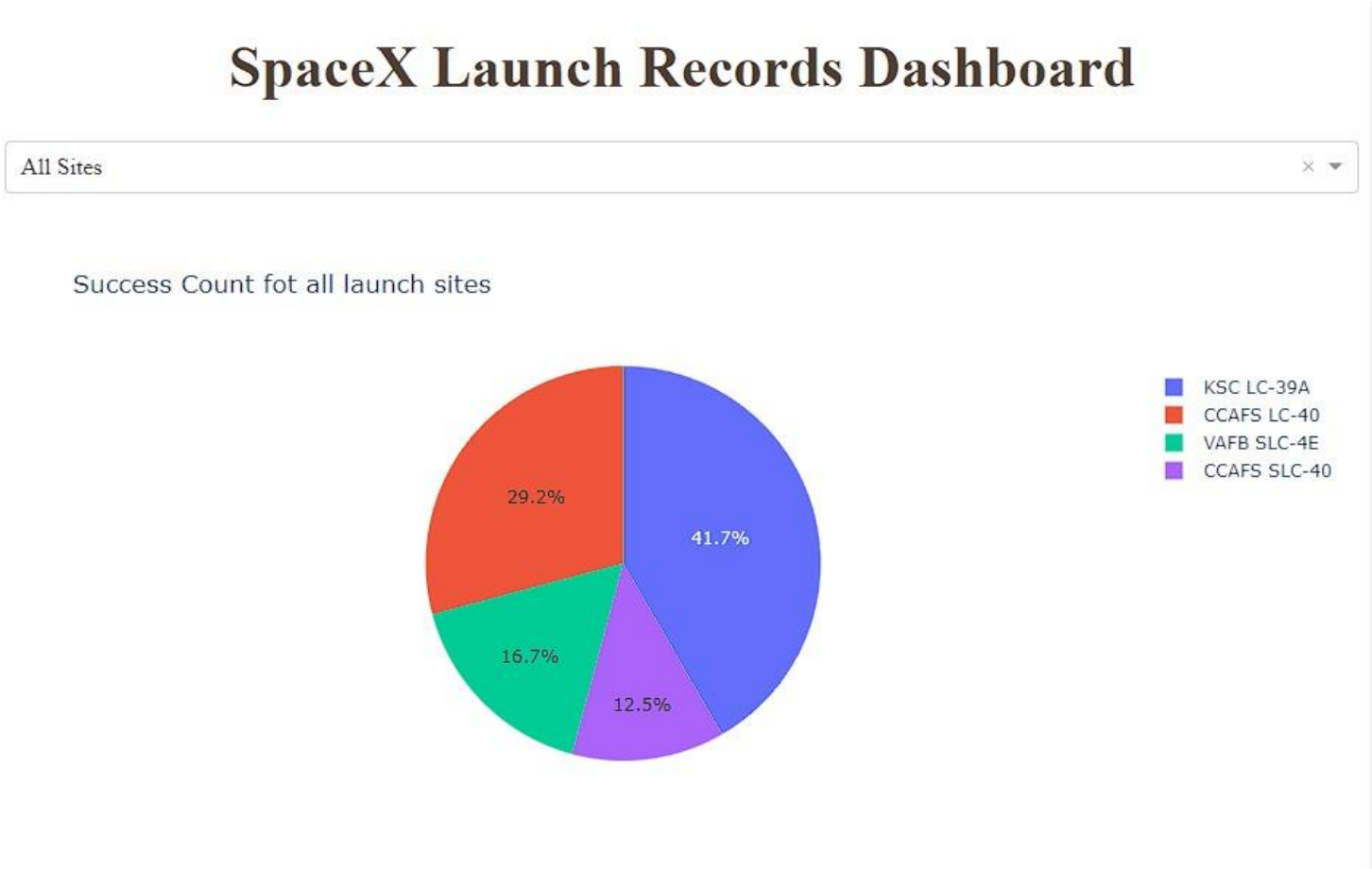
Launch Site Proximity

- We got from example of Launch site CCAFS SLC-40
- The Launch Site got proximity about 1.28 KM on nearby railroad
- 0.58 KM Proximity from nearby Highway
- And 0.86 KM form the nearby coast


Section 4
Build a Dashboard with Plotly Dash

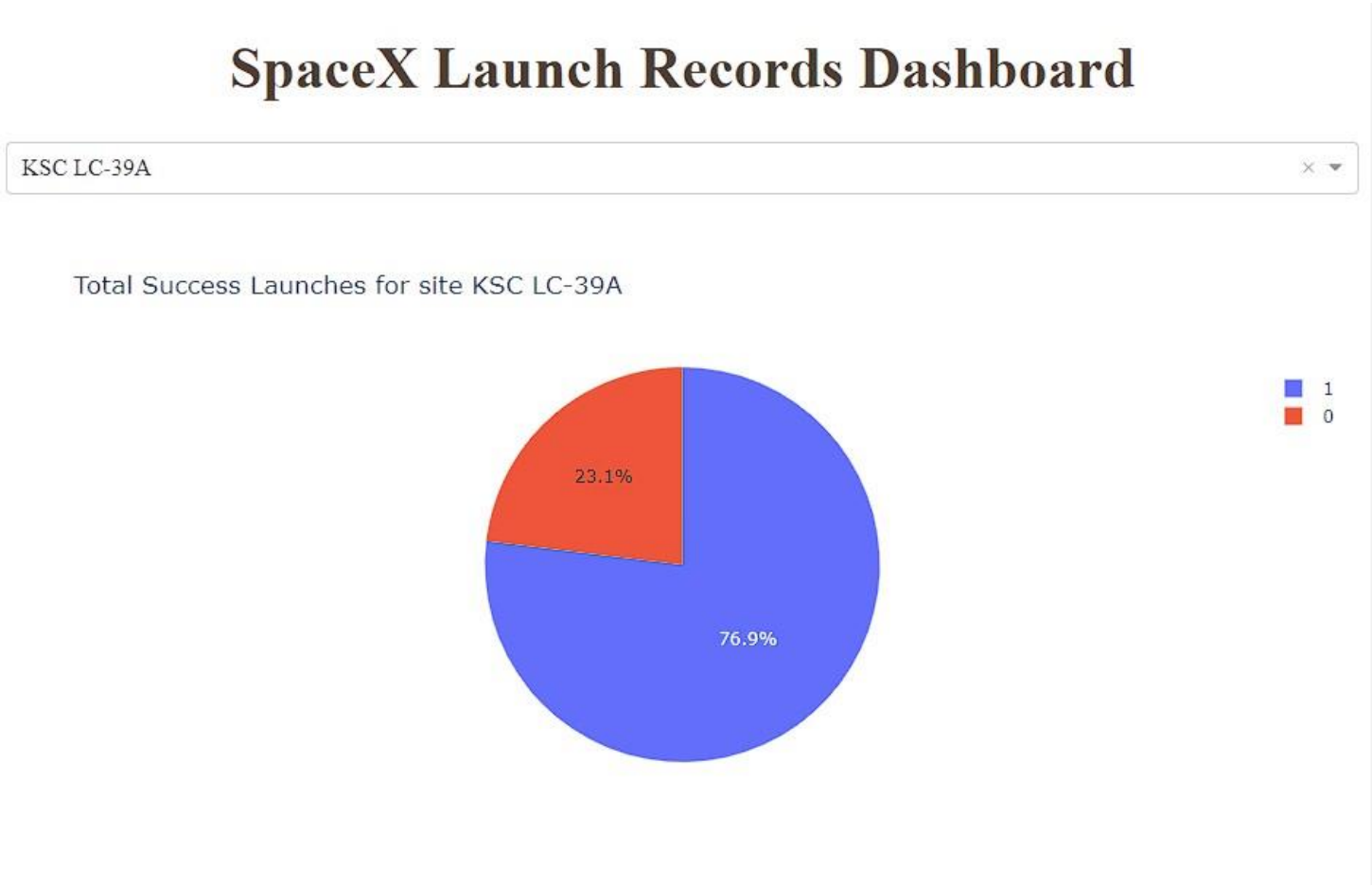
Launch Success

- From Pie chart Dashboard of All sites launch Success we got KSC LC-39A had highest launch success rate with 41.7%
- On contrary CCAFS SLC-40 had the lowest launch success rate with 12.5%



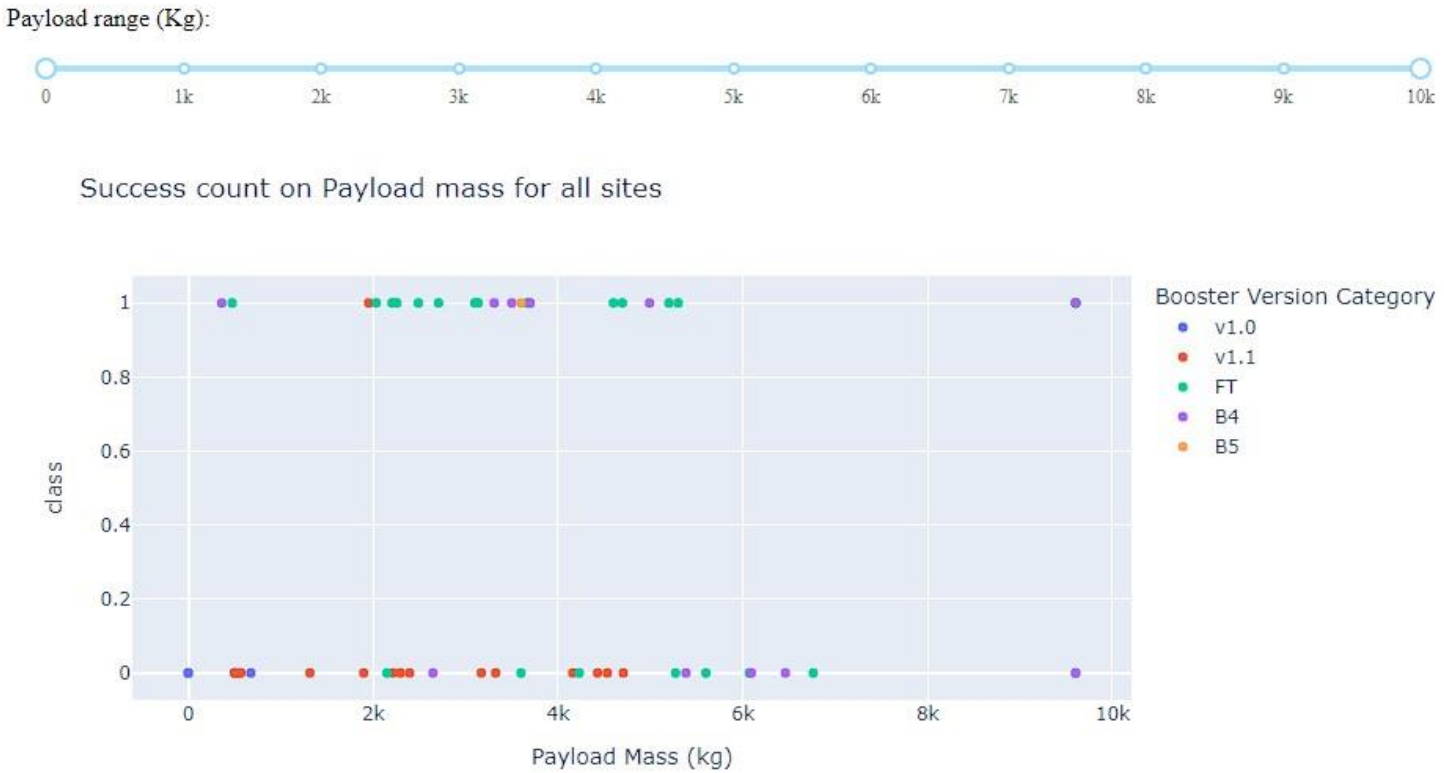
Launch Success (KSC LC-39A)

- When we focus on site KSC LC-39A as the highest success rate
- We got it had 76.9% success rate in total and 23.1% failure rate which fascinating

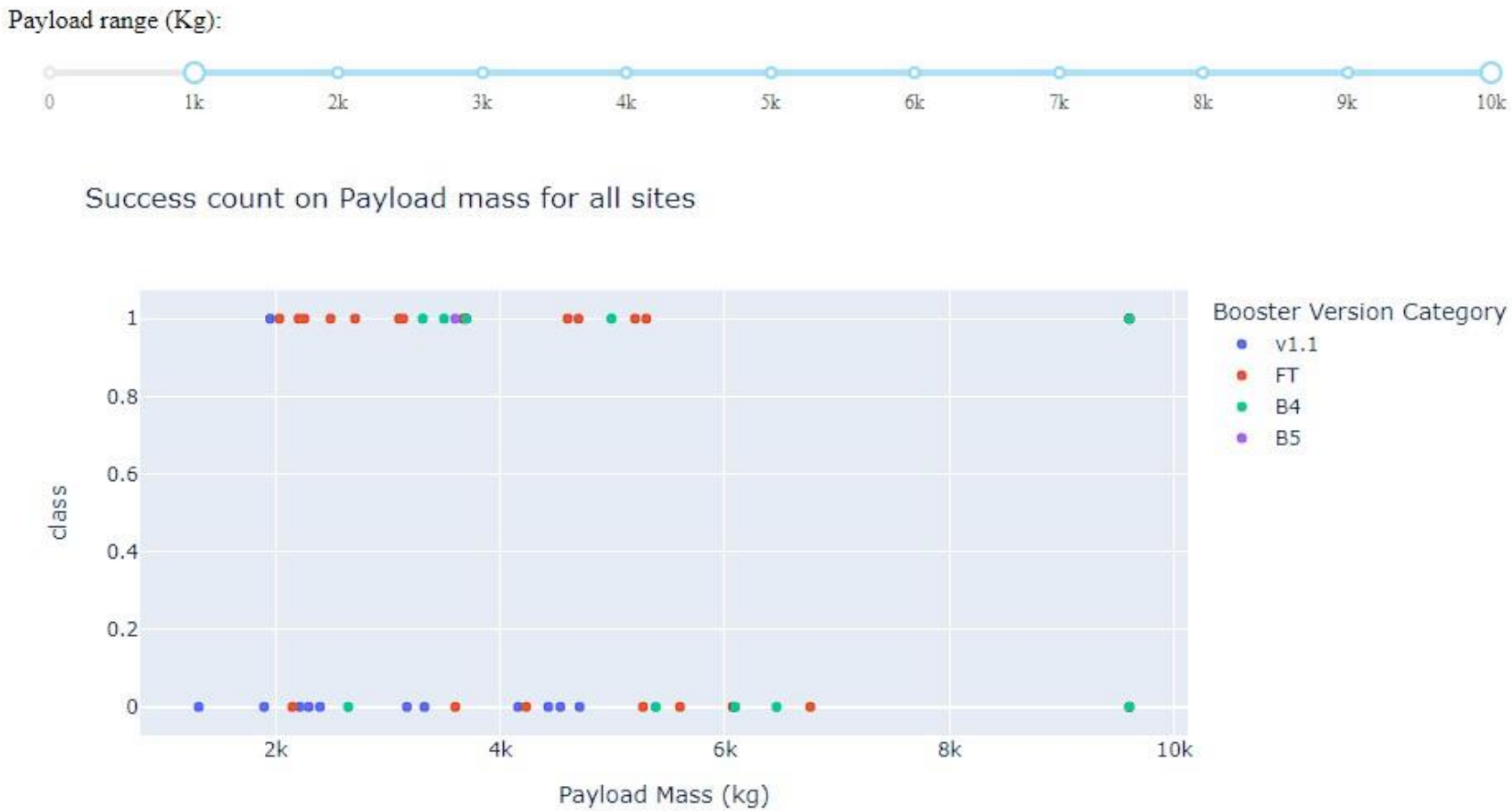


Payload vs Launch Outcome

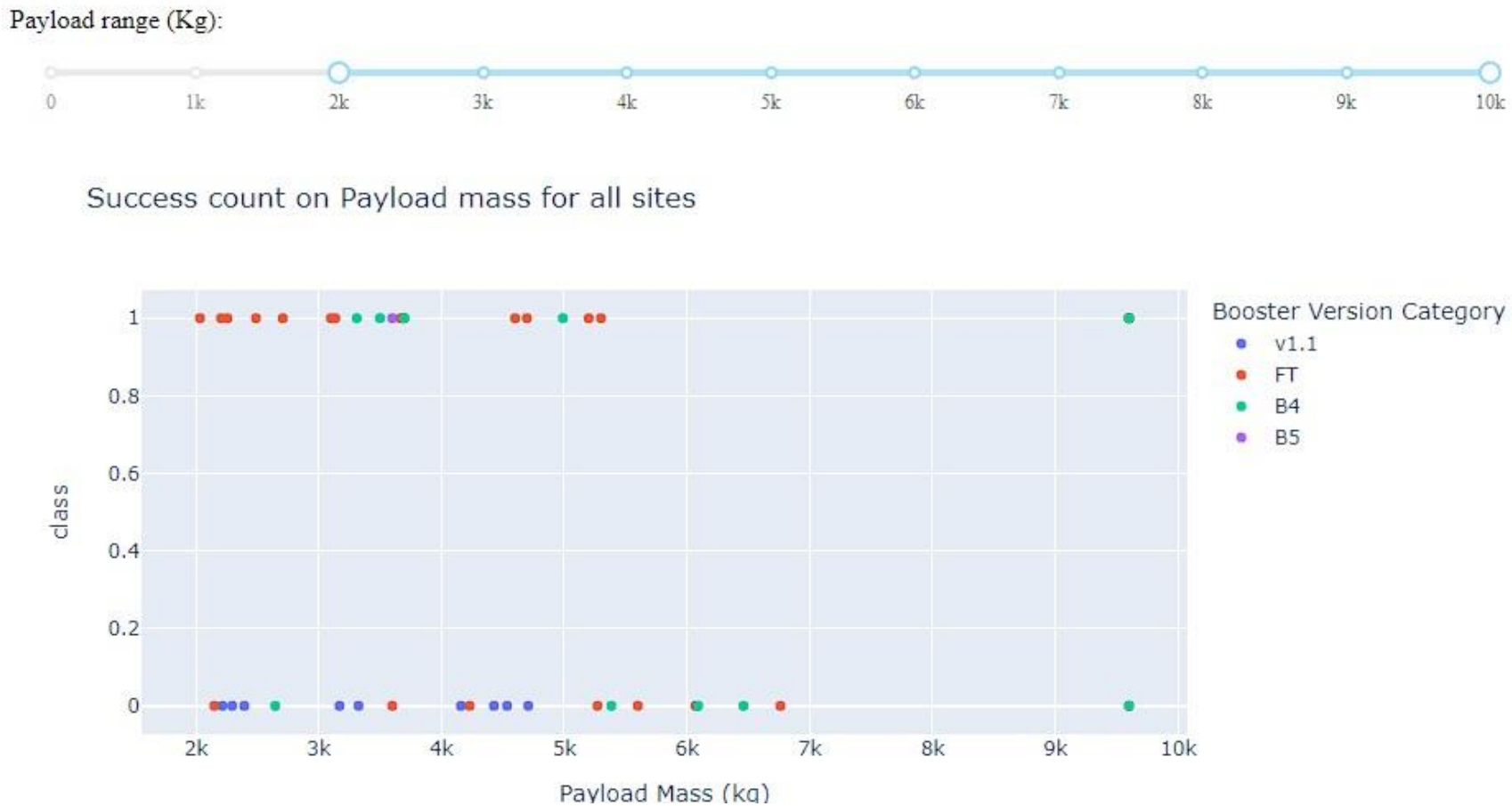
- This show initial condition when we want to show related between Payload range (kg) and Booster Version on Succes Rate (class = 1)
- We would try to slide the slider and see want the relation and tell the best success rate on which Payload range and Booster Version.



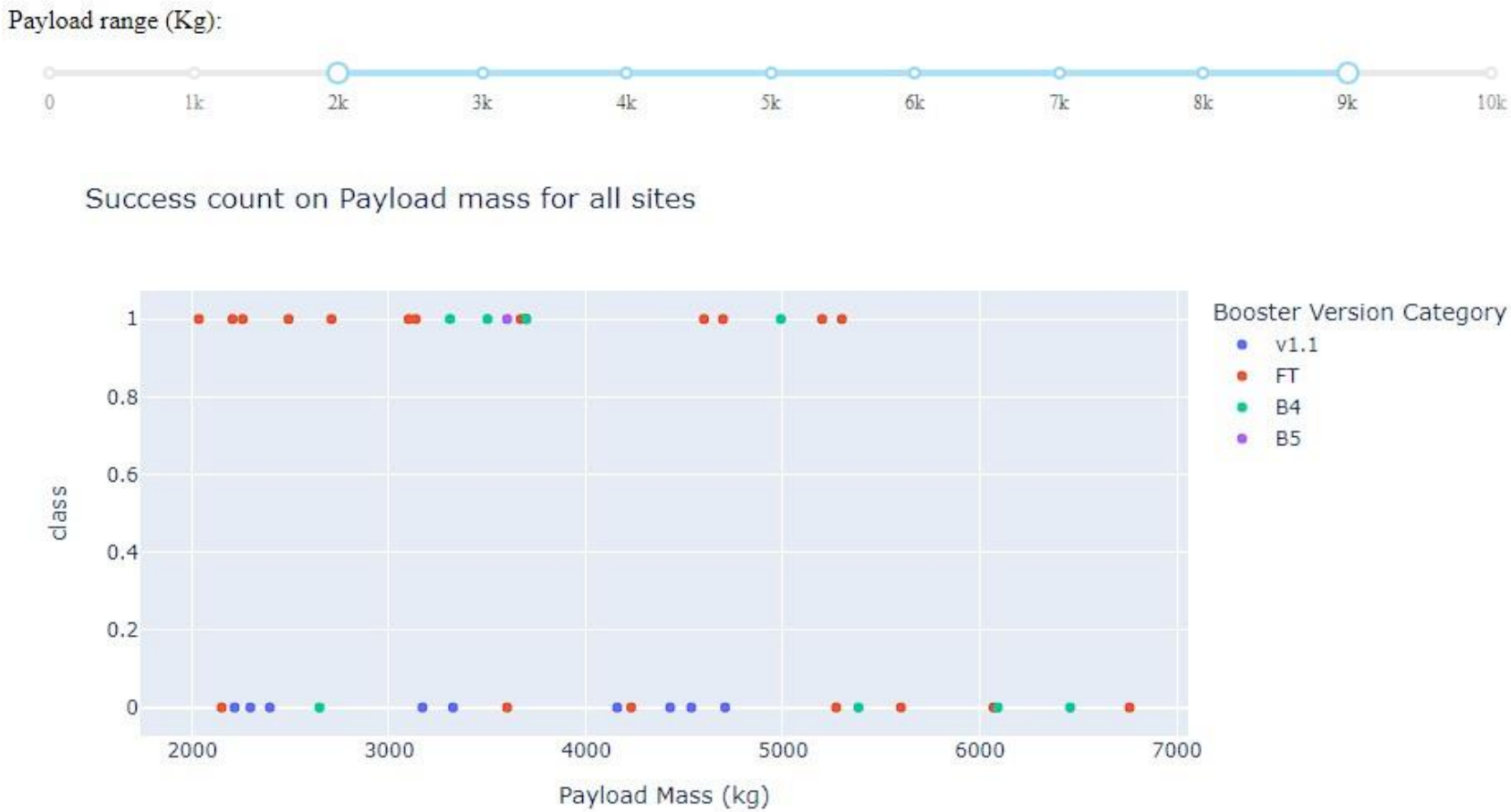
Payload range (1k – 10k)



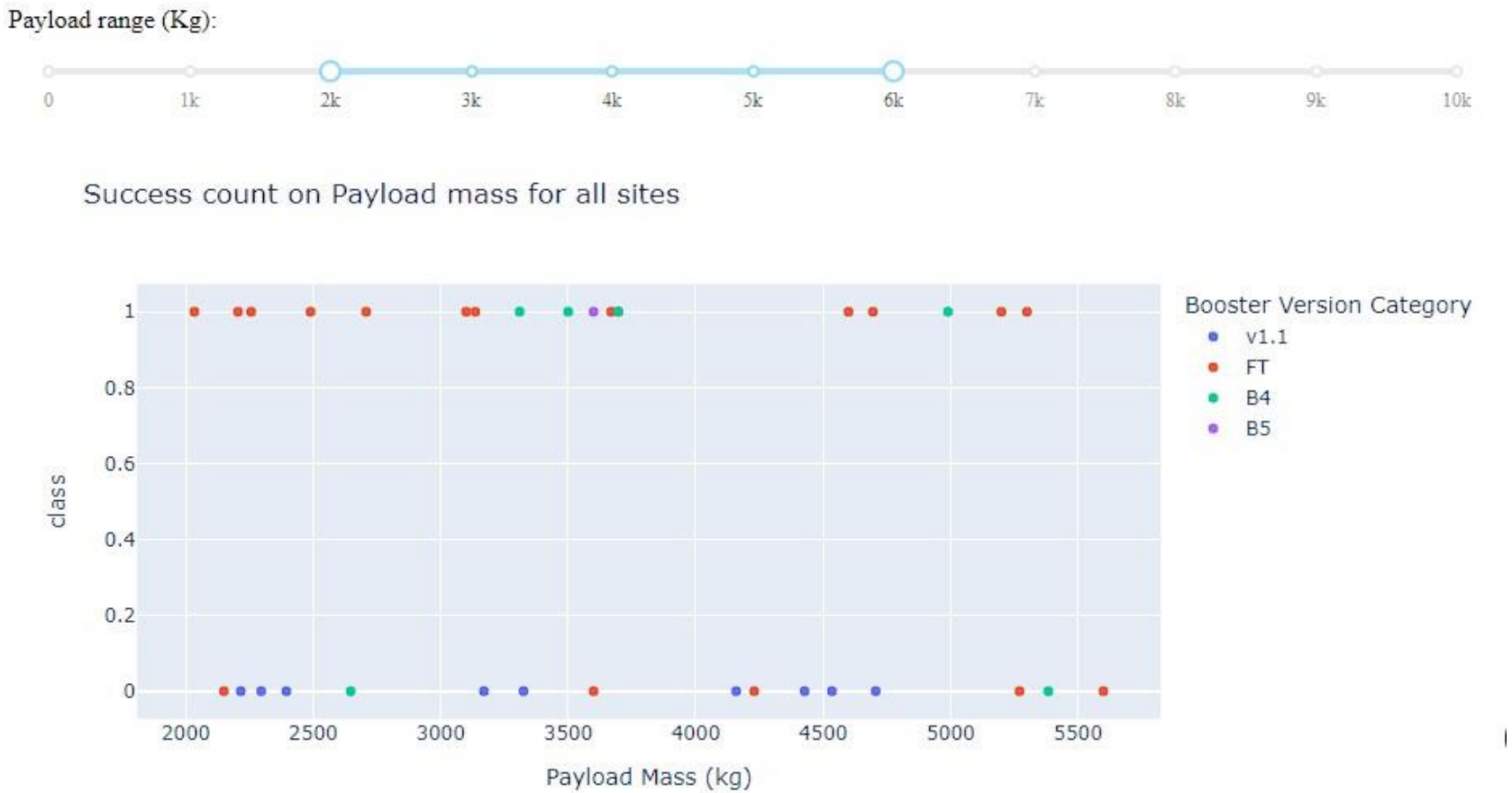
Payload range (2k – 10k)



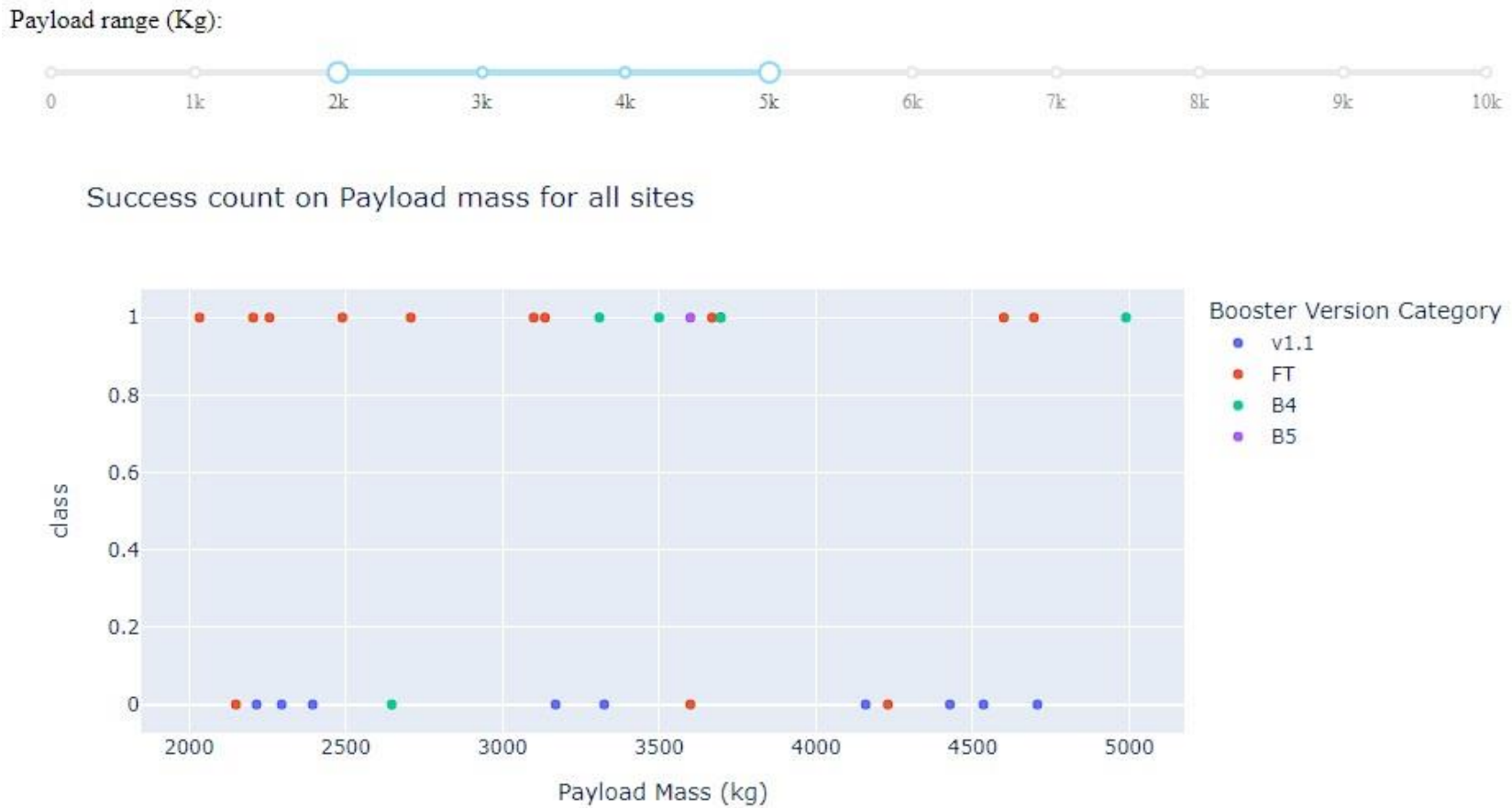
Payload range (2k – 9k)



Payload range (2k – 6k)

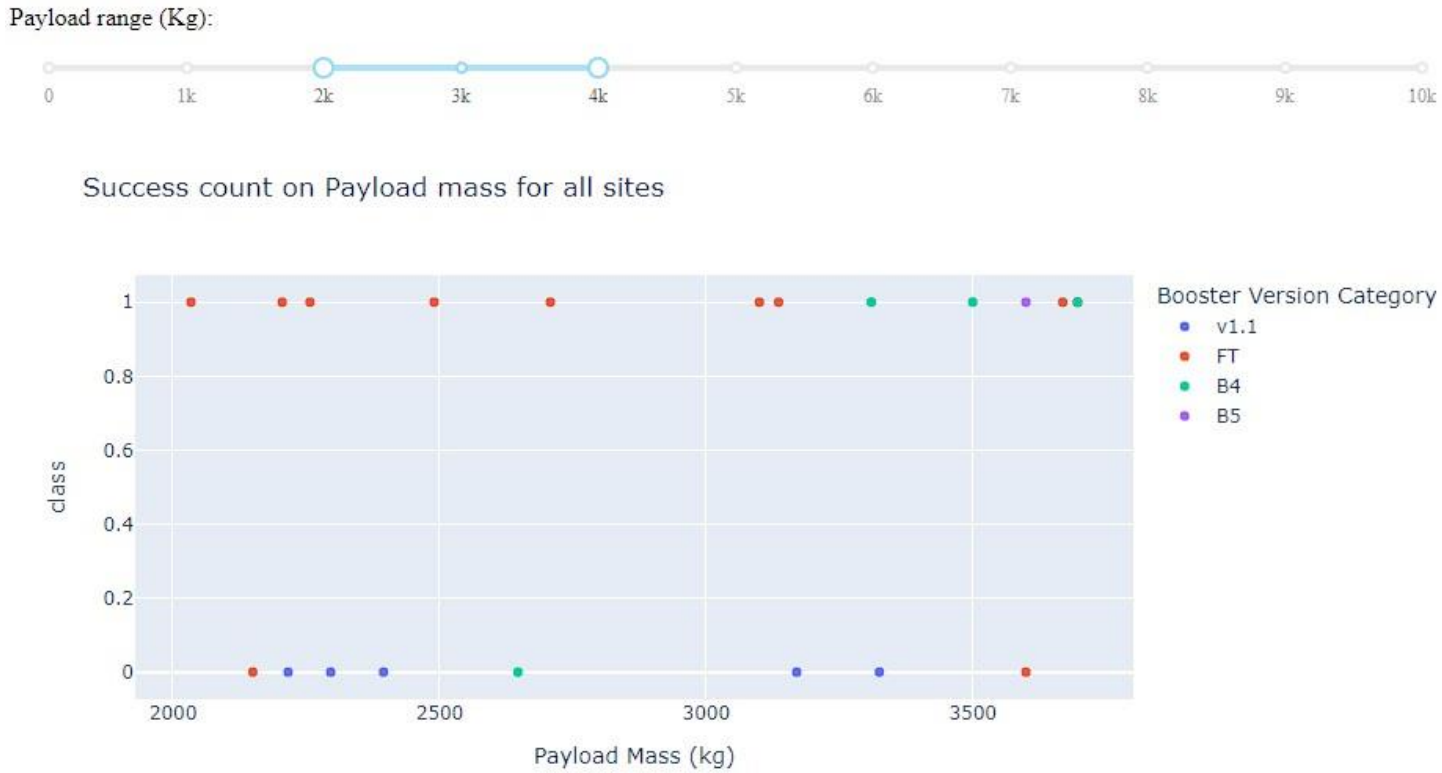


Payload range (2k – 5k)



Payload range (2k-5k)

- We can infer Payload range 2k-5k is the optimize size payload for success land
- As the payload range max lower the failure rate count reduce
- And can we infer on this optimize payload range FT Booster Version had the highest success count.

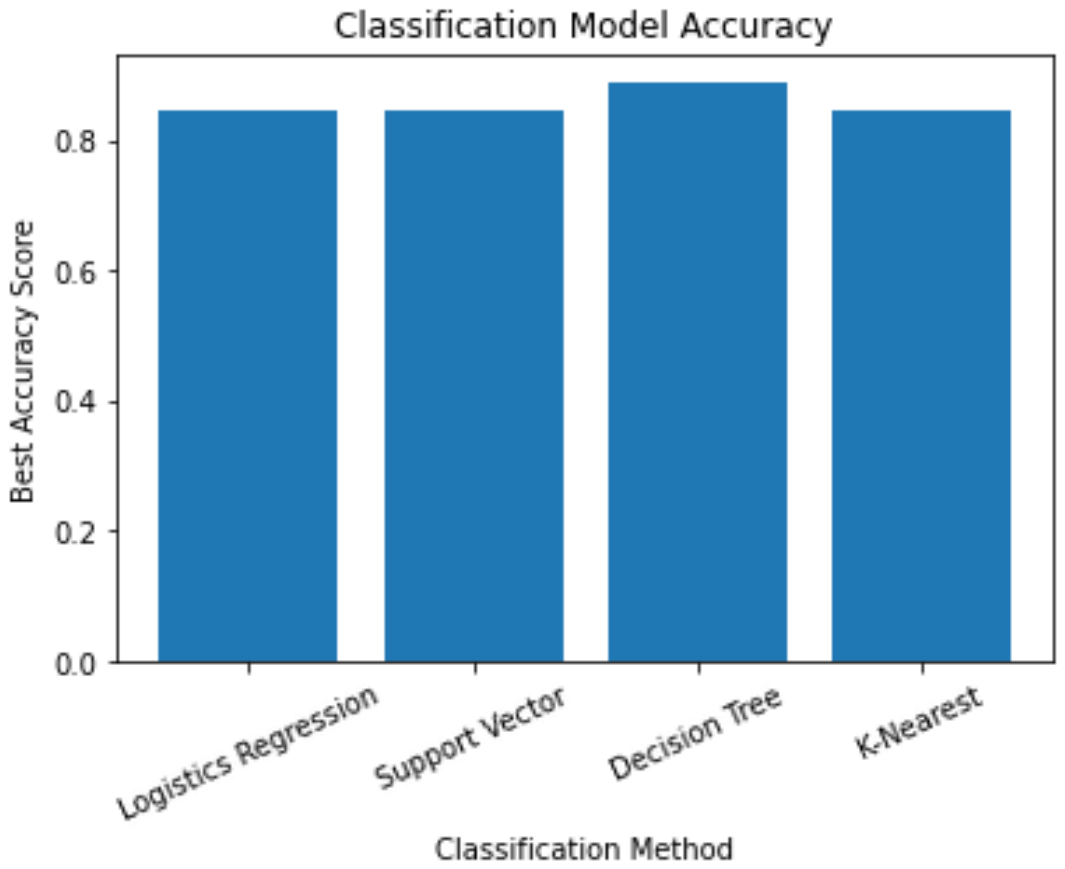


Section 5

Predictive Analysis (Classification)

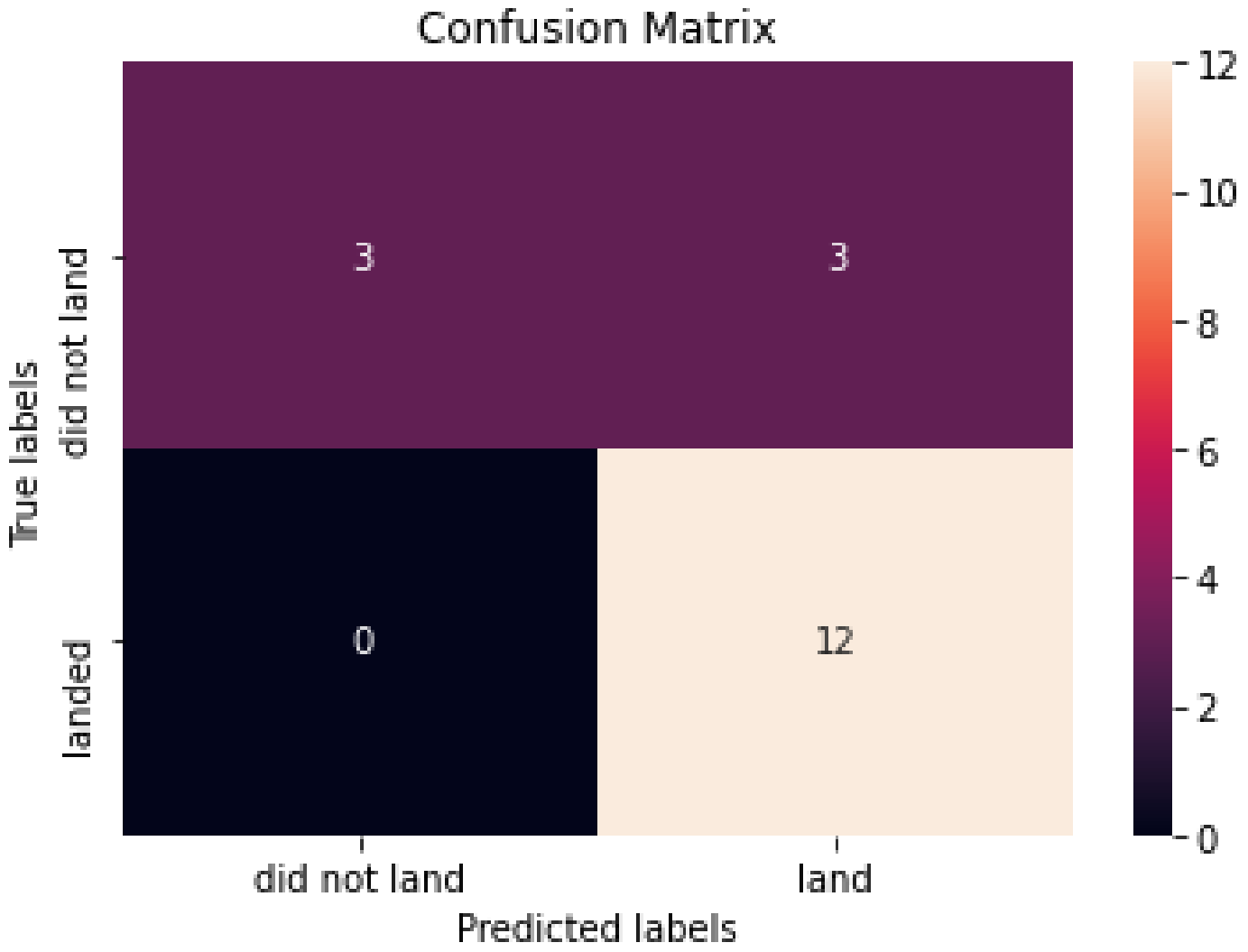
Classification Accuracy

- Even there is slightest different best score on every Classification method
- Decision Tree had the highest best accuracy classification with 0.8889 or equal to 88.89% Accuracy



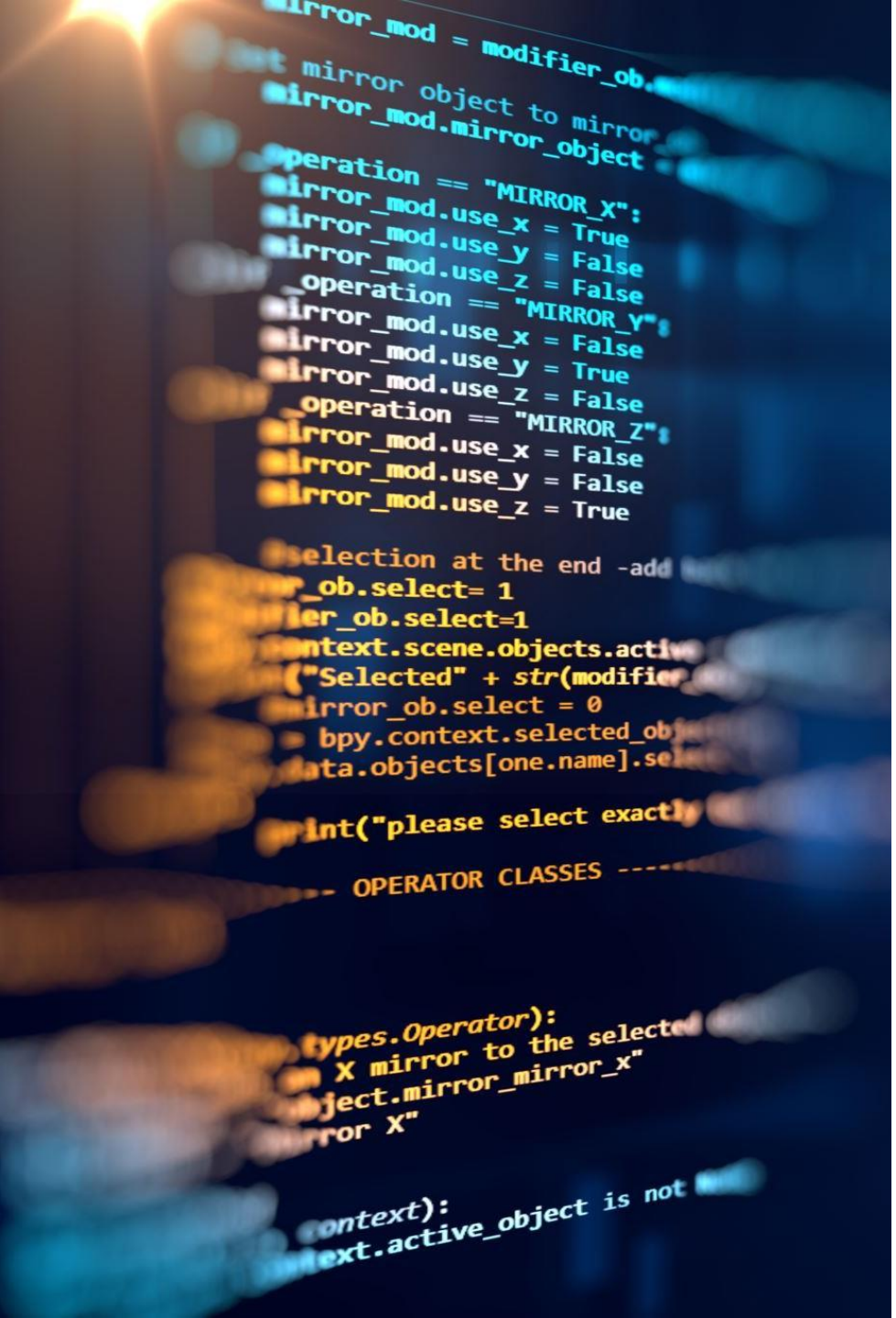
Confusion Matrix (Decision Tree)

- The result from Confusion matrix of Decision Tree show the Model Predict 3 'did not land' and True Labels so showing 100% accuracy on 'did not land'
- The model predict 12 'land' and so the True result but predict 3 on did not land which false positif, showing 80% accuracy on 'land'



Conclusions

- SpaceX launch sucess rate since 2013 kept increasing till 2020.
- ES-L1, GEO, HEO, and SSO orbit type had 100% success rate.
- KSC LC-39A had highest success rate with 76.9% success rate and it seen on the map too.
- Payload range 2k-5k is the optimize size payload for success land with FT Booster Version had the highest success count.
- Decision Tree had the highest best accuracy classification with 0.8889 or equal to 88.89% Accuracy.



Appendix

- All Python code snippets, SQL queries, charts, Notebook, or data sets used in this Project saved to this Github link.

Thank you!