IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

YU SHING LUI
17 JAN 2022

Outline

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

Executive Summary

- All the data is coming from public SpaceX API and SpaceX Wikipedia page.
- Exploring data using SQL, data visualization, folium map and dashboards.
- Finding the best parameters for ML model and visualize the accuracy score
- ML model: Logistic Regression, Support Vector Machine, Decision Tree, and KNN
- All the results are above 83%

Introduction

- Commercial Space is not a dream and it is started
- Space X has a good stock price nowadays
- Space X has a huge competitor is Space Y
- The problem we have to solve on this project is about using machine learning to predict successful stage 1 recovery

Section 1

Methodology

Methodology

Executive Summary

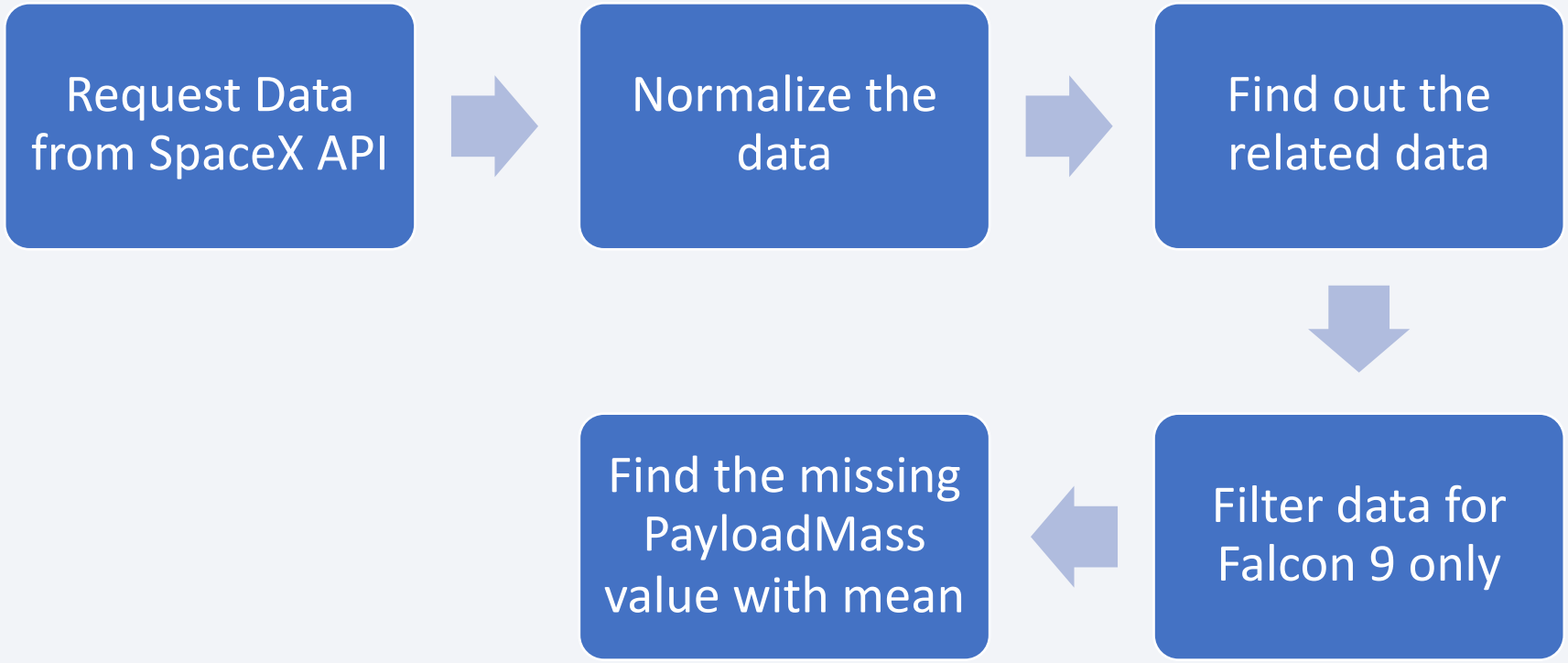
- Data collection methodology:
  - All the data is coming from public SpaceX API and SpaceX Wikipedia page
- Perform data wrangling
  - Classifying true landings as successful or not
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Fine tune all those models that using GridSearchCV

Data Collection

- All the data is coming from public SpaceX API and SpaceX Wikipedia page.
- The attributes of Space X are FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude
- The attributes of Wikipedia are Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

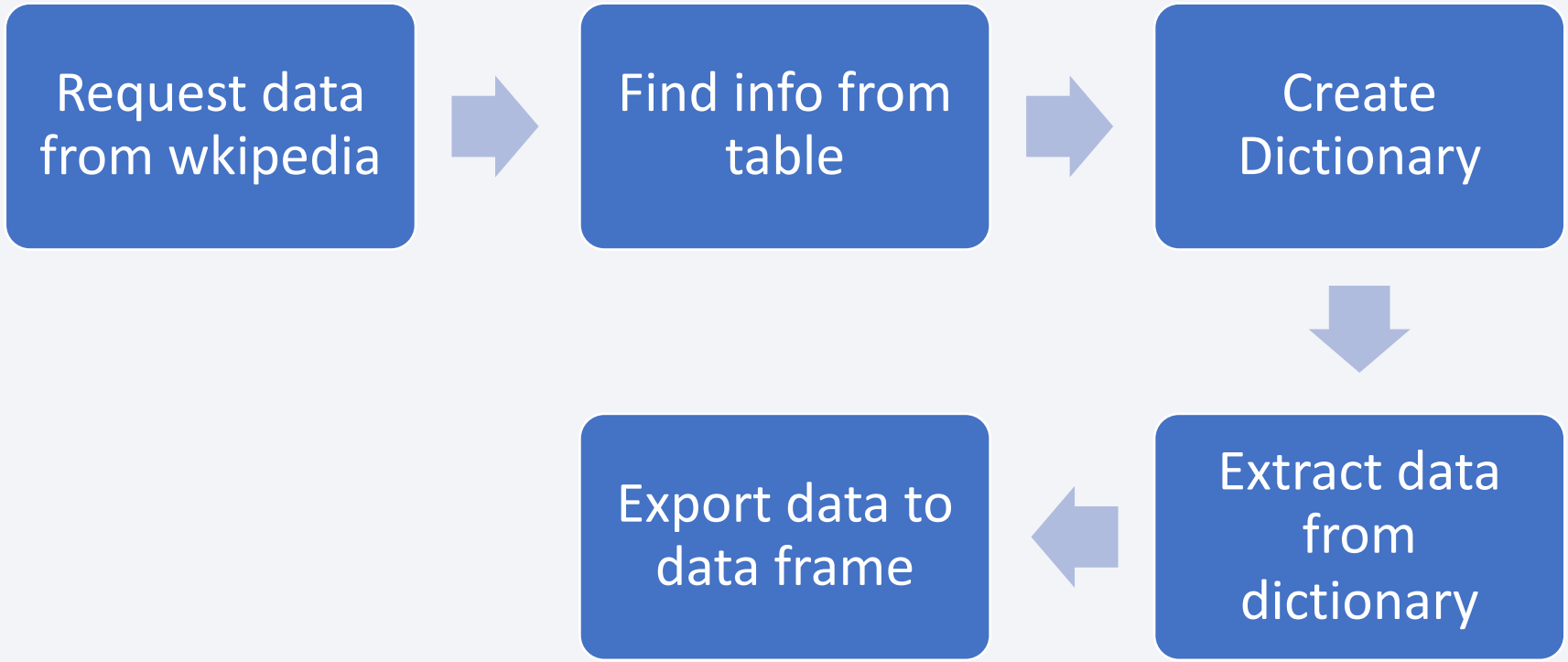
Data Collection – SpaceX API

- https://github.com/yushinglui/ibmAppliedDataScienceCapstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- https://github.com/yushinglui/ibmAppliedDataScienceCapstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Add the training attributes with landing outcome which successful=1 & failure=0
- “Mission Outcome” and ”Landing Location” are necessary for the result of column
- New attribute of “class” which contain a value of 1 if “Mission Outcome” is correct, and a value of 0 is opposite way
- Set to 1 if ASDS, RTLS & Ocean are true
- Set to 0 if ASDS, RTLS & Ocean are false

https://github.com/yushinglui/ibmAppliedDataScienceCapstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Finding the performance of the attributes of Flight Number, Payload Mass, Launch Site, Orbit, Class and Year.
- For those charts, it will present Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit vs. Success Rate, Flight Number vs. Orbit, Payload vs Orbit, and Success Yearly Trend.
- Using different charts like line charts, bar plots and scatter plots to find the relationship before we start to train the models

https://github.com/yushinglui/ibmAppliedDataScienceCapstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Dataset loaded into IBM DB2 database
- It integrate with python for queried
- Information is about launch site names, mission outcomes, various pay load sizes of customers and booster versions, and landing outcomes

https://github.com/yushinglui/ibmAppliedDataScienceCapstone/blob/main/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- Launch Sites, successful and unsuccessful landings, and a key locations of Railway, Highway, Coast, and City which all included in folium map
- The reason is launch sites may locate where they are
- Display successful landing relate to location

https://github.com/yushinglui/ibmAppliedDataScienceCapstone/blob/main/lab_jupyter_launch_site_location.ipynb

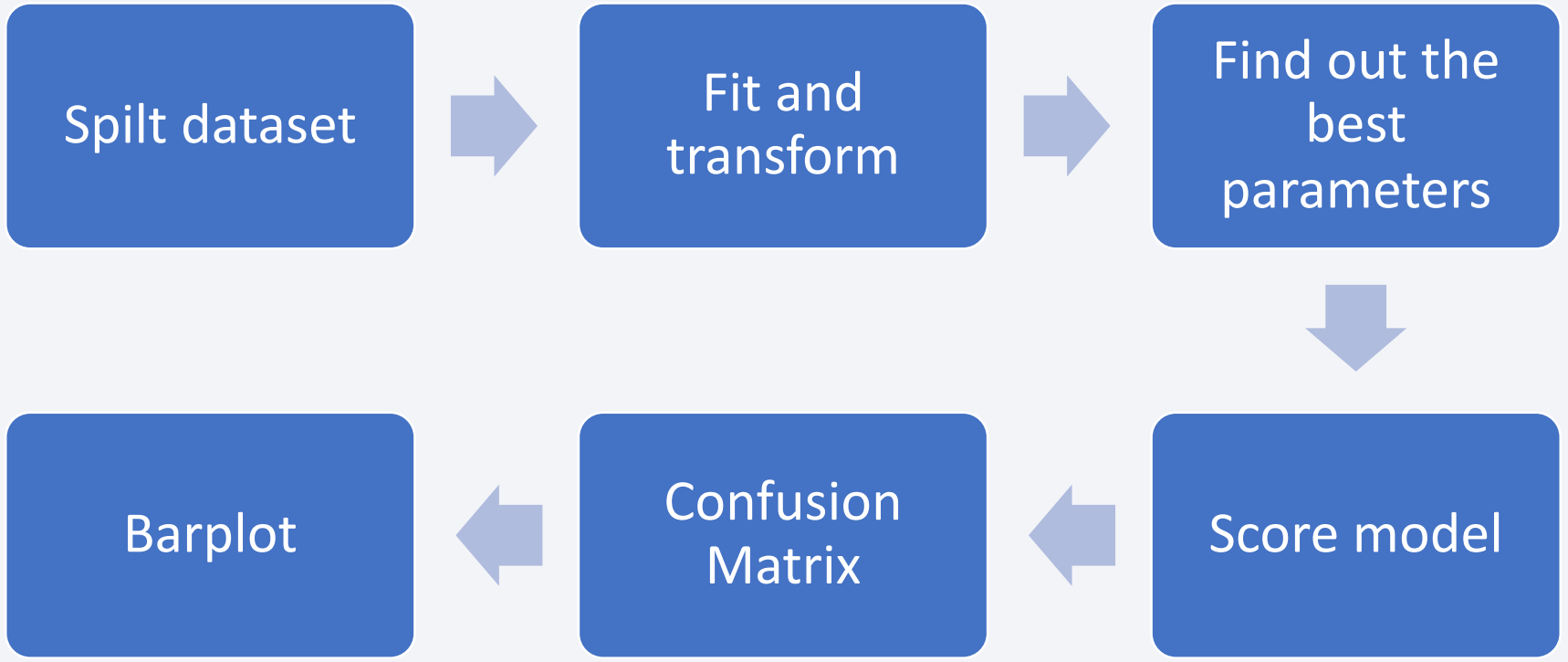
Build a Dashboard with Plotly Dash

- Pie chart and scatter chart is added into a dashboard.
- Pie chart can visualize launch site success rate.
- Scatter plot can find out how varies across launch sites, payload mass, and booster version category.

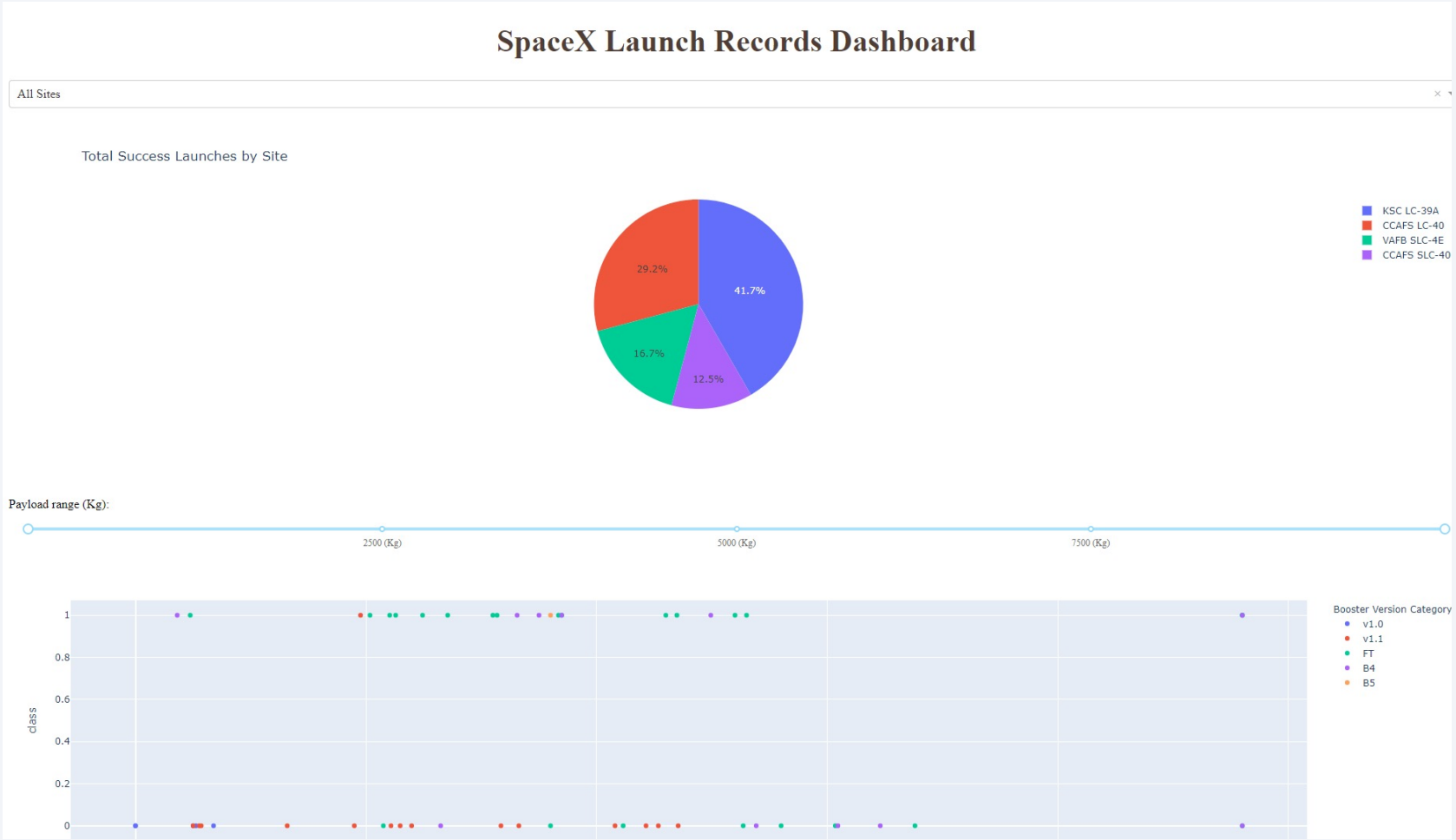
https://github.com/yushinglui/ibmAppliedDataScienceCapstone/blob/main/dashboard.ipynb

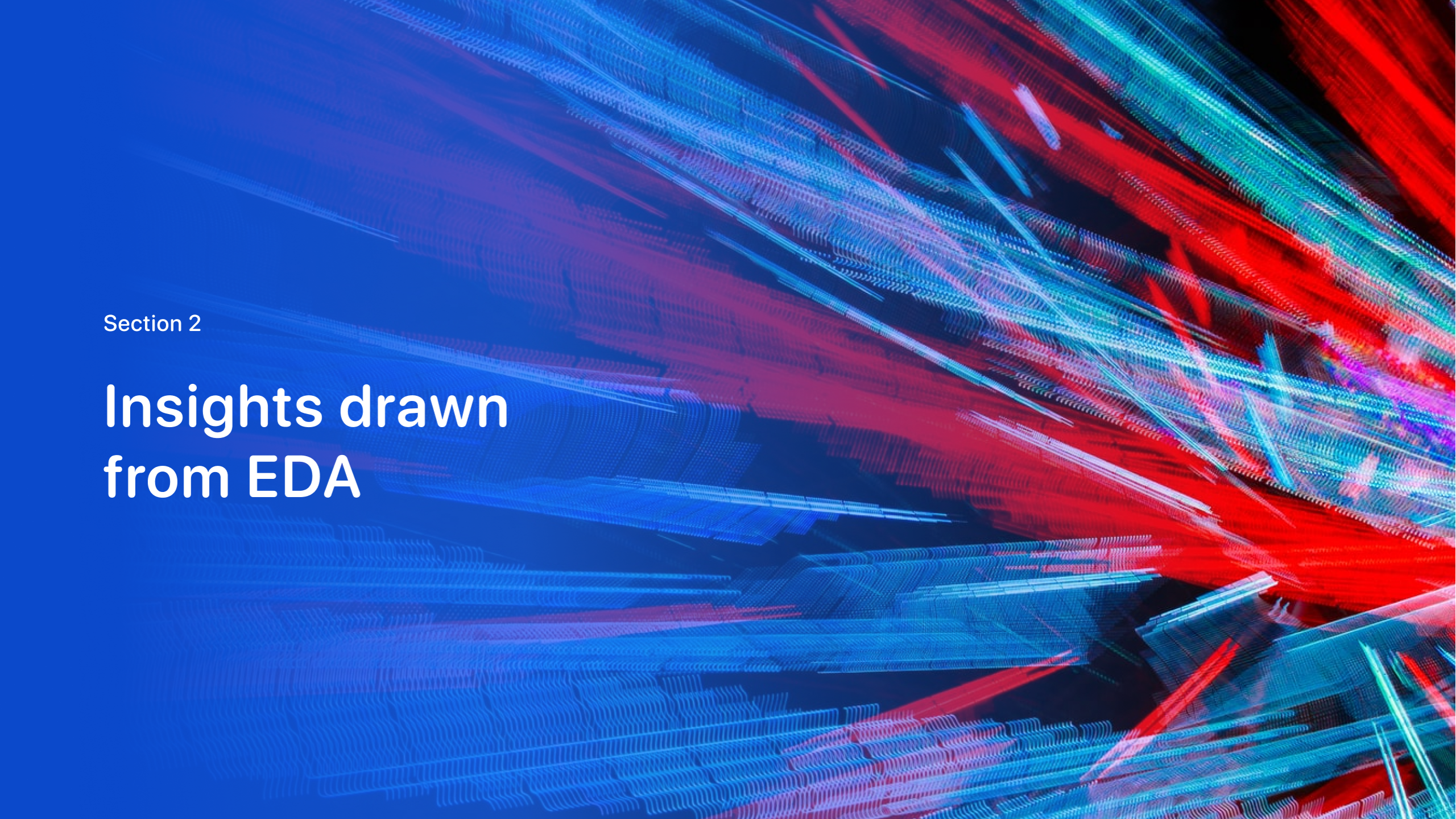
Predictive Analysis (Classification)

- http://localhost:8889/notebooks/Downloads/final/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

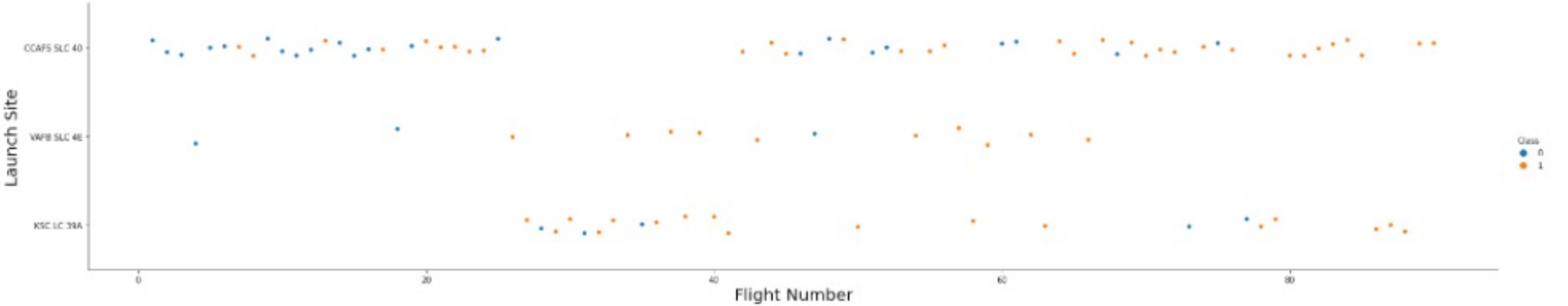

SpaceX Launch Records Dashboard
All Sites
Total Success Launches by Site
29.2%
41.7%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
Payload range (Kg):
2500 (Kg)
5000 (Kg)
7500 (Kg)
class
1
0.8
0.6
0.4
0.2
0
Booster Version Category
v1.0
v1.1
FT
B4
B5



Section 2

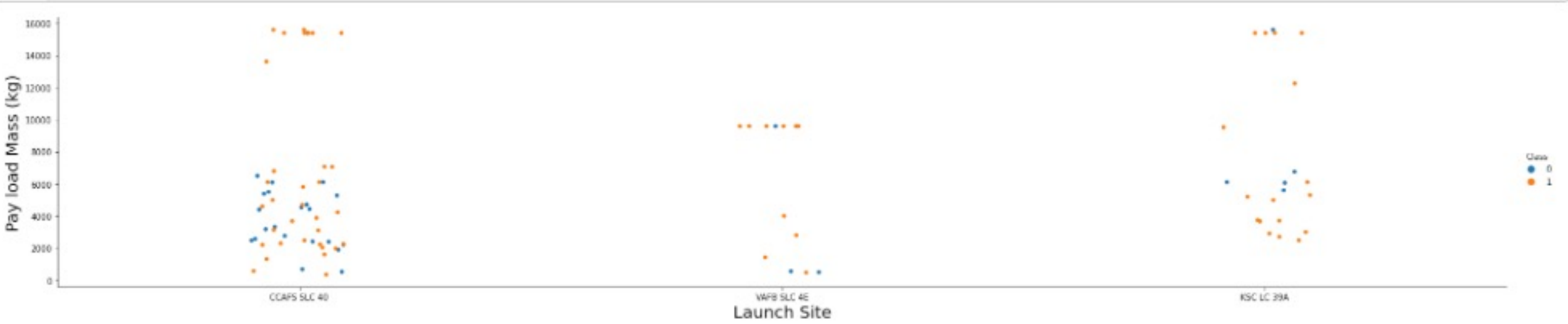
Insights drawn from EDA

Flight Number vs. Launch Site



- It shows the relationship between Flight Number and Launch Site
- Success rate is creased from flight 20
- CCAFS is the major launch site

Payload vs. Launch Site

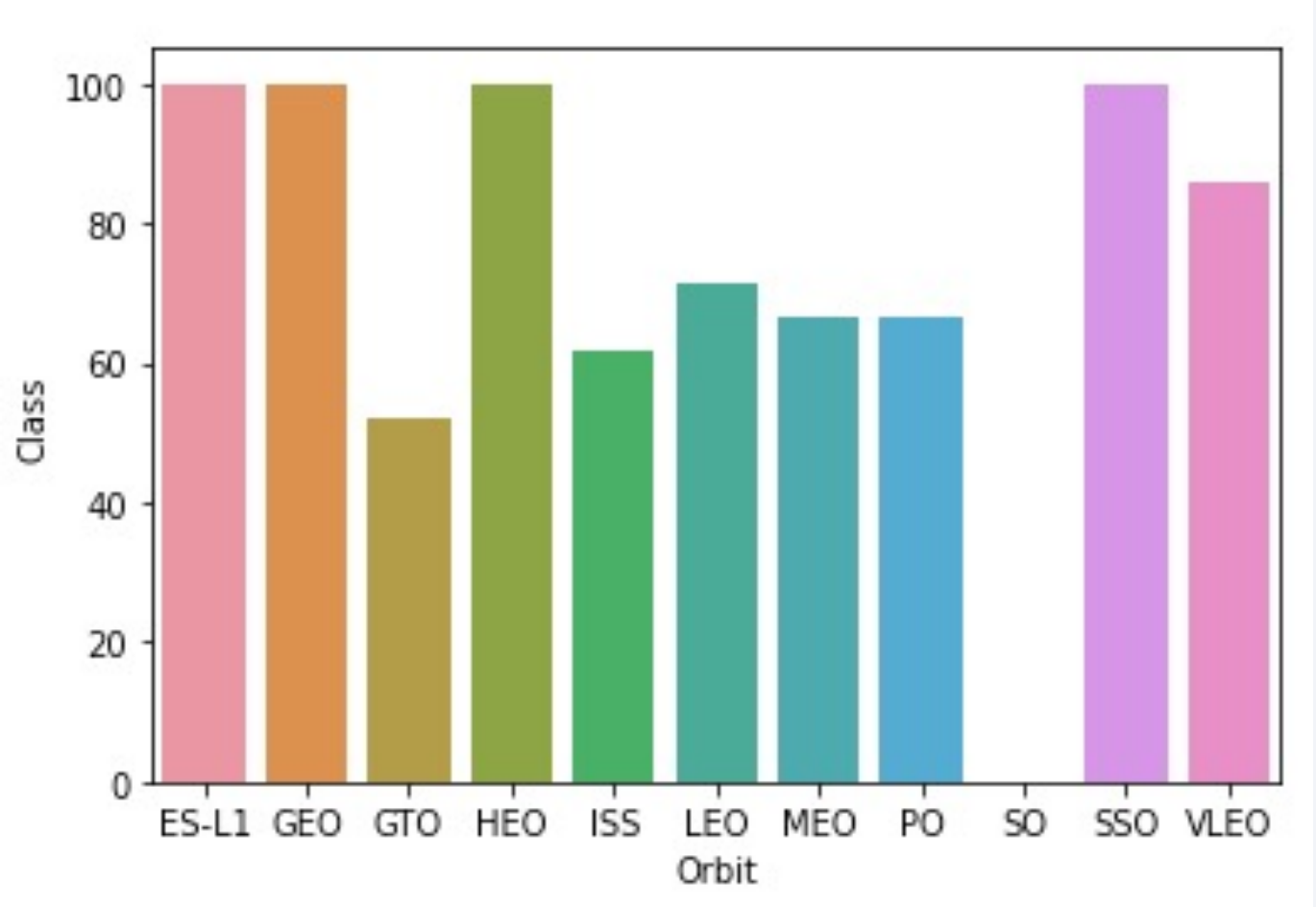


- It shows the relationship between Payload and Launch Site

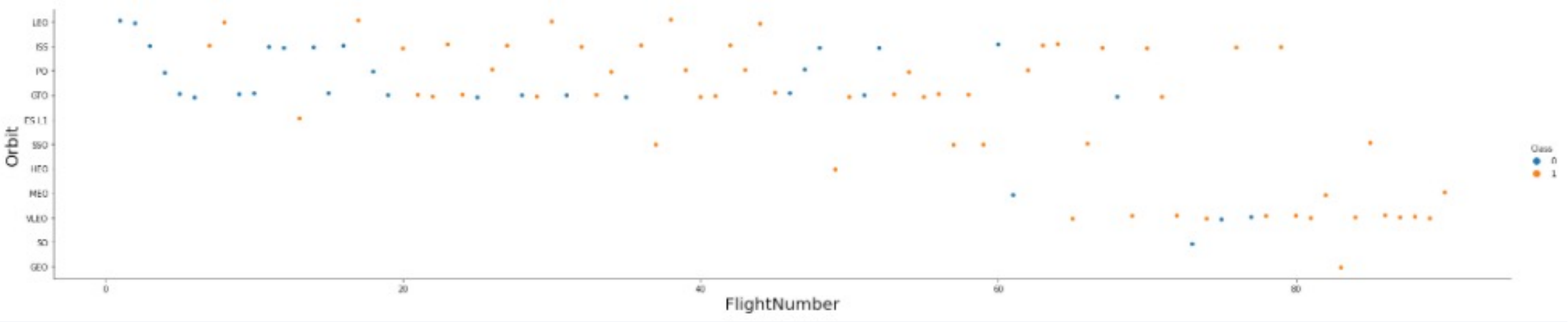
- VAFB-SLC launch site there are no rockets launched for heavy pay load mass (greater than 10000)

Success Rate vs. Orbit Type

- It shows the relationship between Success Rate and Orbit type
- ES-L1, GEO, HEO, and SSO are the highest success rate
- SO is the lowest success rate

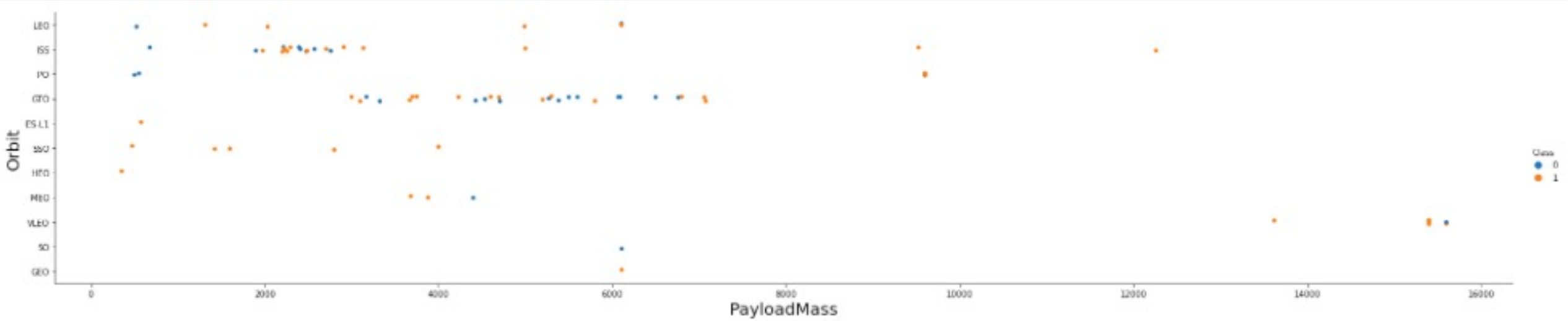


Flight Number vs. Orbit Type



- It shows the relationship between Flight Number and Orbit type
- The LEO orbit the Success appears related to the number of flights
- No relationship between flight number when in GTO orbit.

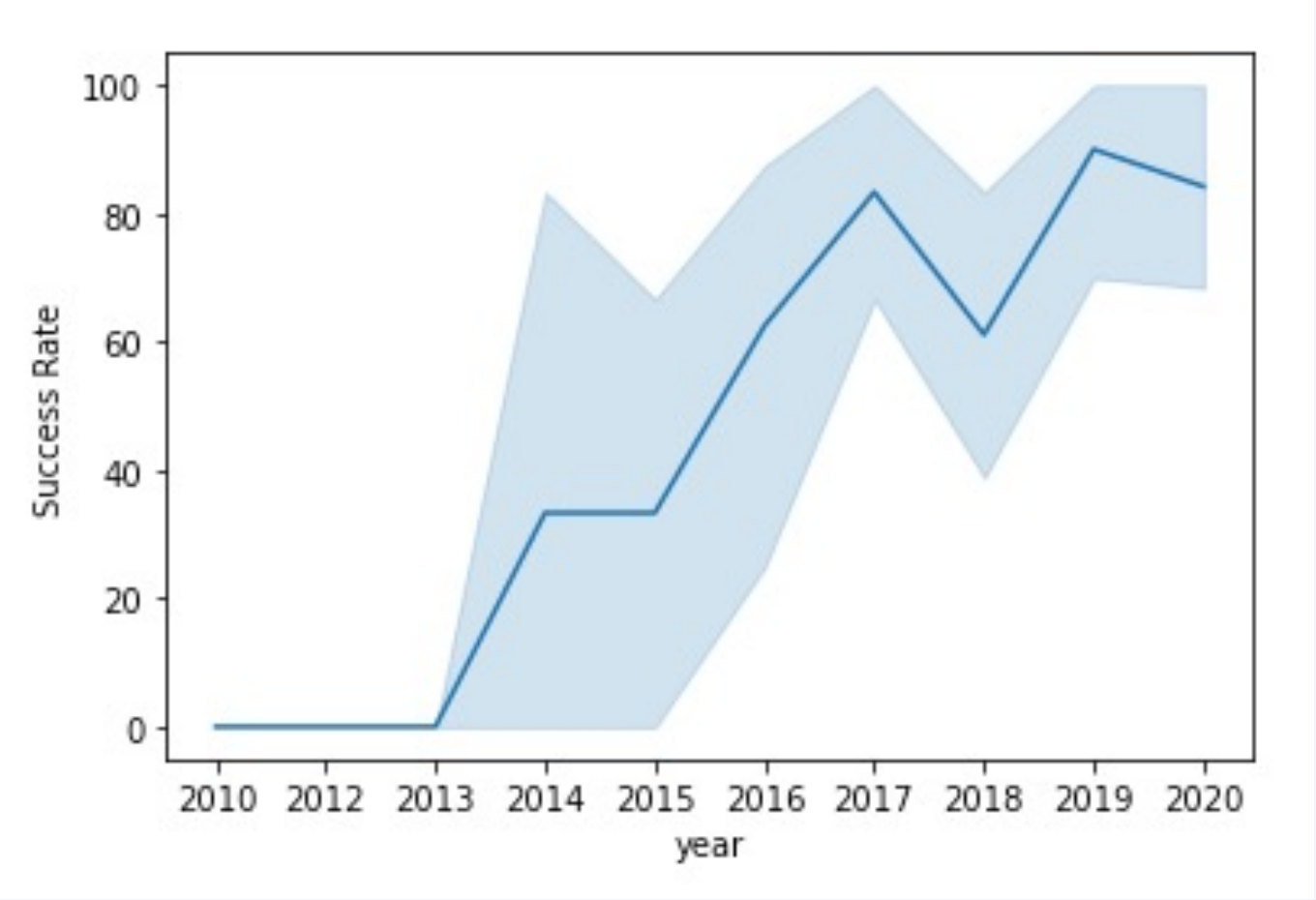
Payload vs. Orbit Type



- It shows the relationship between Payload and Orbit type
- Positive landing rate are more for Polar, LEO and ISS
- GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here

Launch Success Yearly Trend

- Show a line chart of the launch success yearly trend
- The success rate since 2013 kept increasing till 2020



All Launch Site Names

- CCAFS LC-40 was the previous name
- Actually, the site name of CCAFS SLC-40 and CCAFSSLC-40 are the same because data entry error
- Three launch sites from database only

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

Launch Site Names Begin with 'CCA'

- The total payload mass carried by boosters launched by NASA (CRS)

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

- The total payload mass carried by boosters launched by NASA (CRS)

sum_payload_mass_kg
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1

avg_payload_mass_kg
2928

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad

first_success
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Four 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

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

Boosters Carried Maximum Payload

- Those 12 boosters which have carried the maximum payload mass

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 Launch Records

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

MONTH	landing__outcome	booster_version	payload_mass__kg_	launch_site
January	Failure (drone ship)	F9 v1.1 B1012	2395	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	1898	CCAFS LC-40

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

- 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

landing__outcome	no_outcome
Success (drone ship)	5
Success (ground pad)	3

Section 4

Launch Sites Proximities Analysis

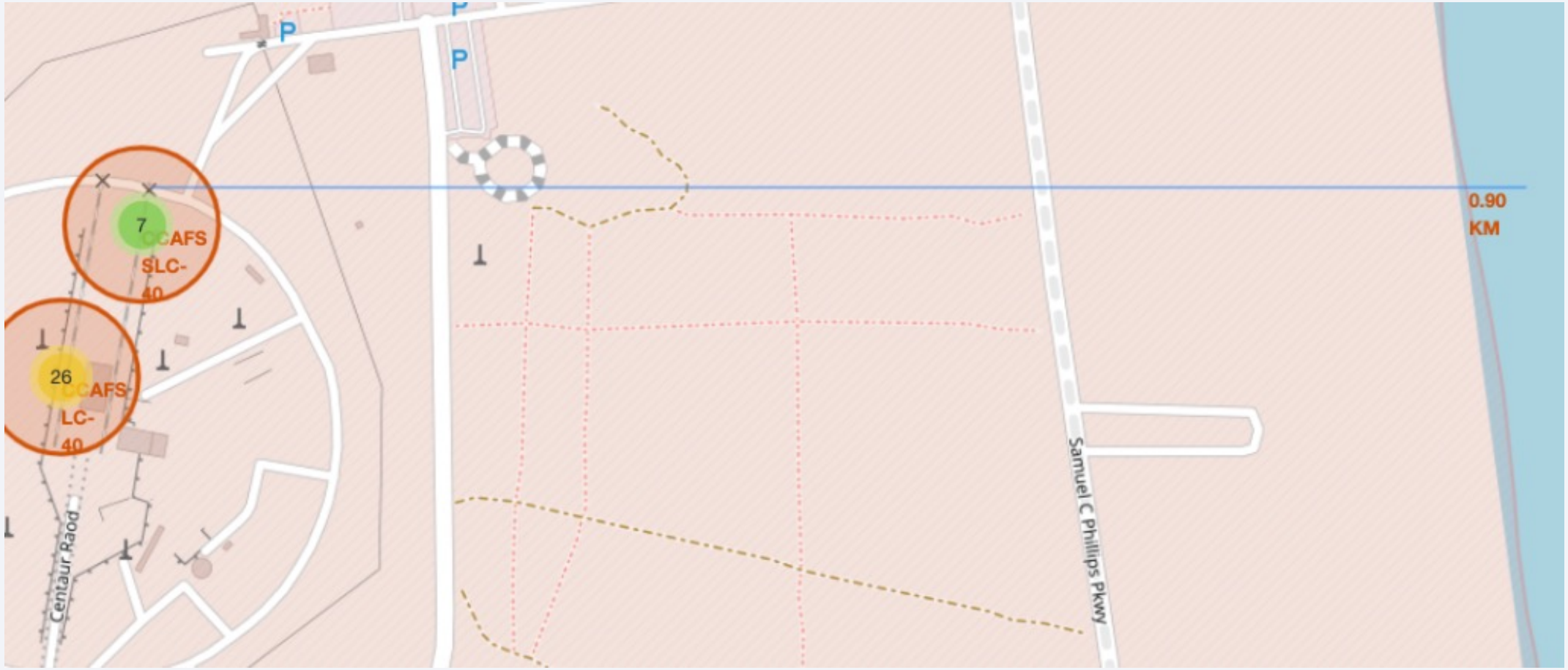
Launch Sites

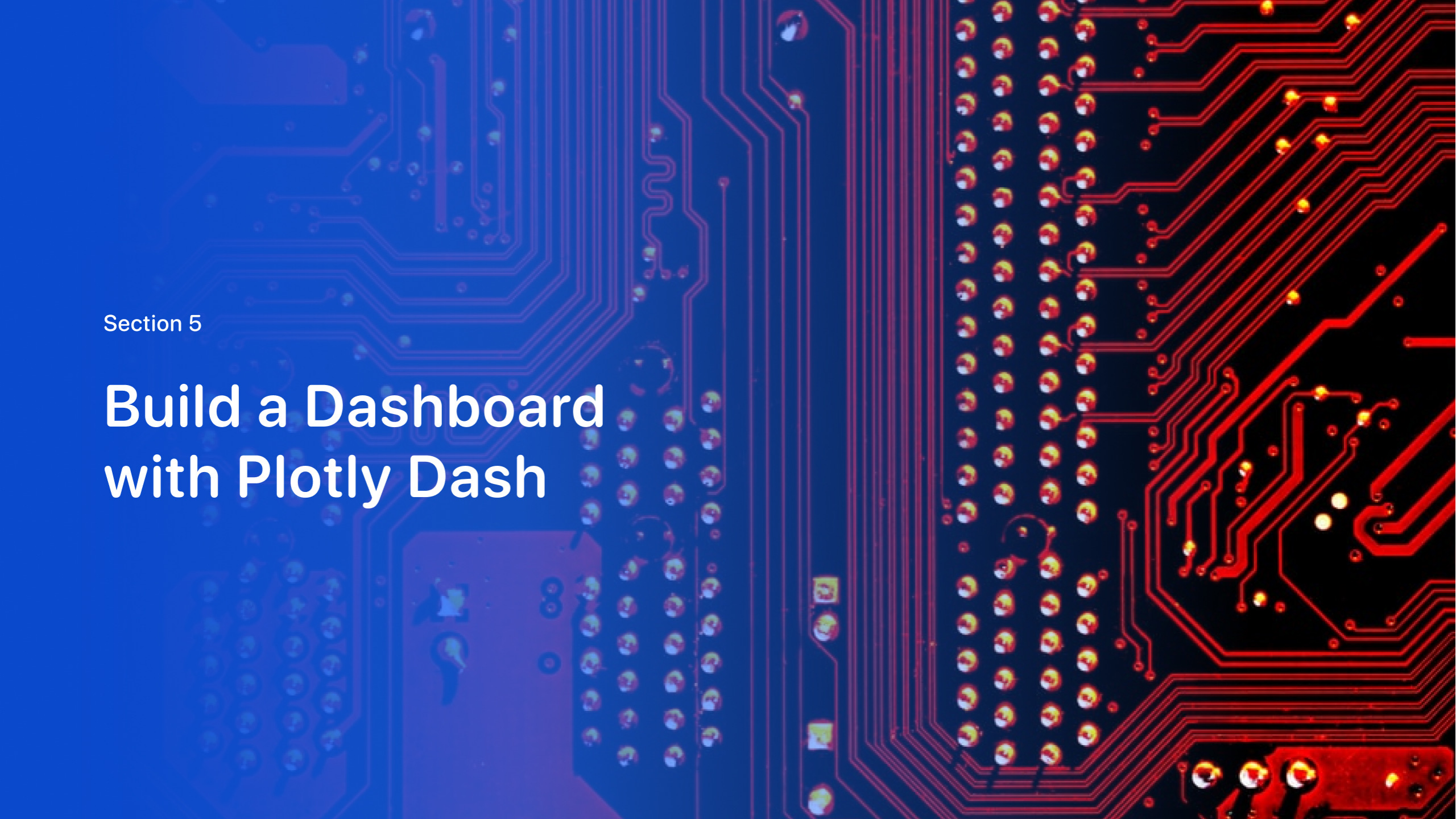
- Those launch sites are very close
- Launch sites in proximity to the Equator line
- They all nearby the ocean

The success/failed launches for each site

- Green one is success landing
- Red on is failed landing
- Able to easily identify which launch sites have relatively high success rates

The distances between a launch sites



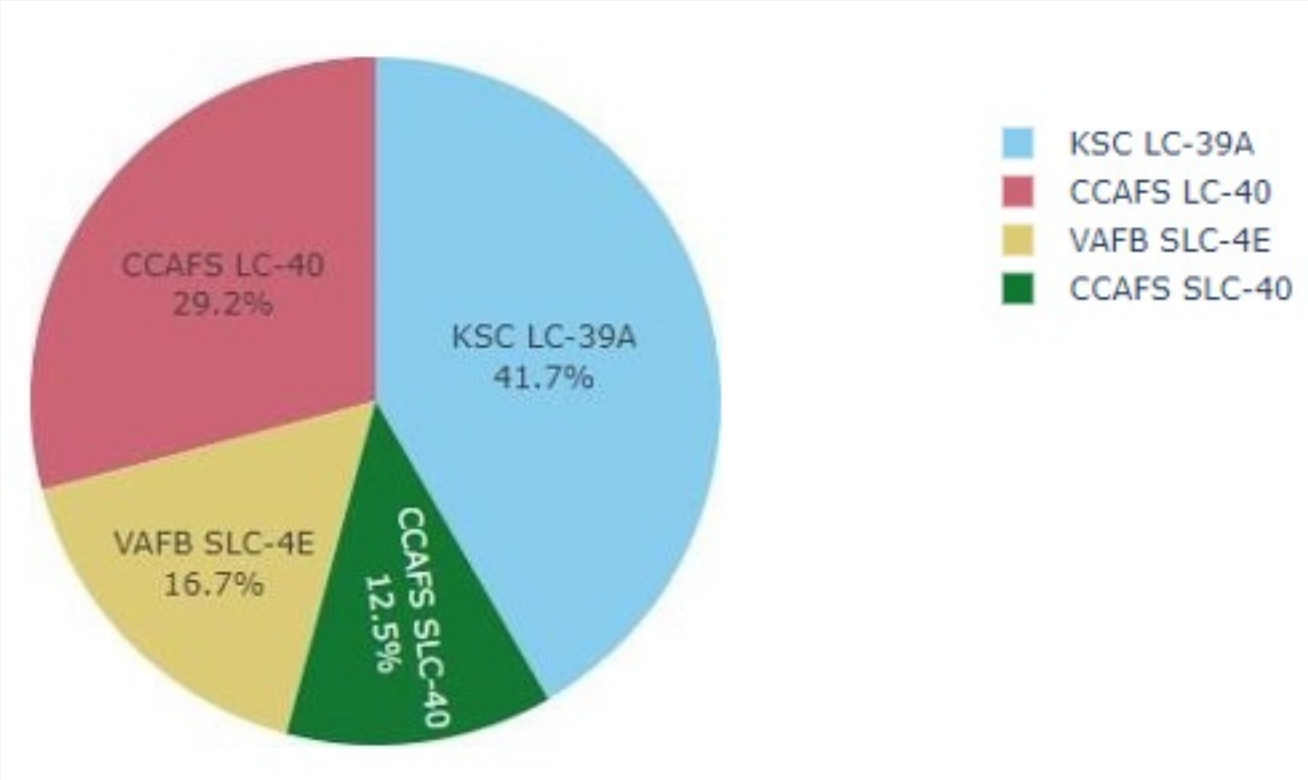


Section 5

Build a Dashboard with Plotly Dash

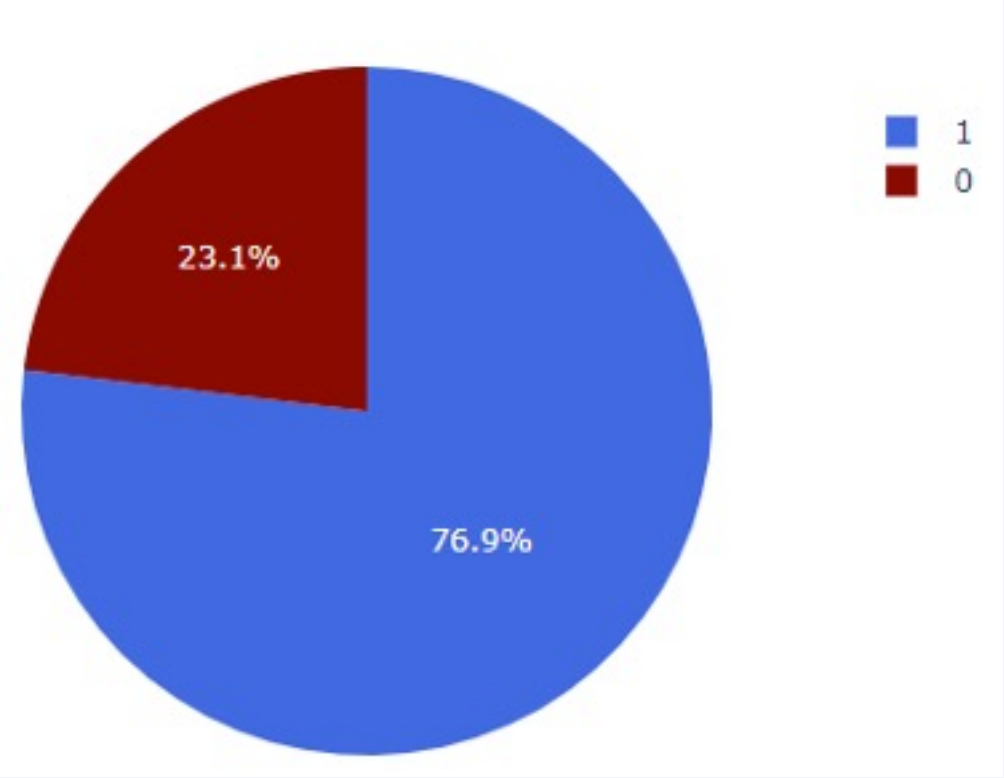
Launch success count for all sites

- KSC LC-39A is the highest success rate
- CCAFS SLC-40 is the lowest success rate
- However, CCAFS LC-40 and CCAFS SLC-40 are the same, so the result is the same success rate with KSC LC-39A

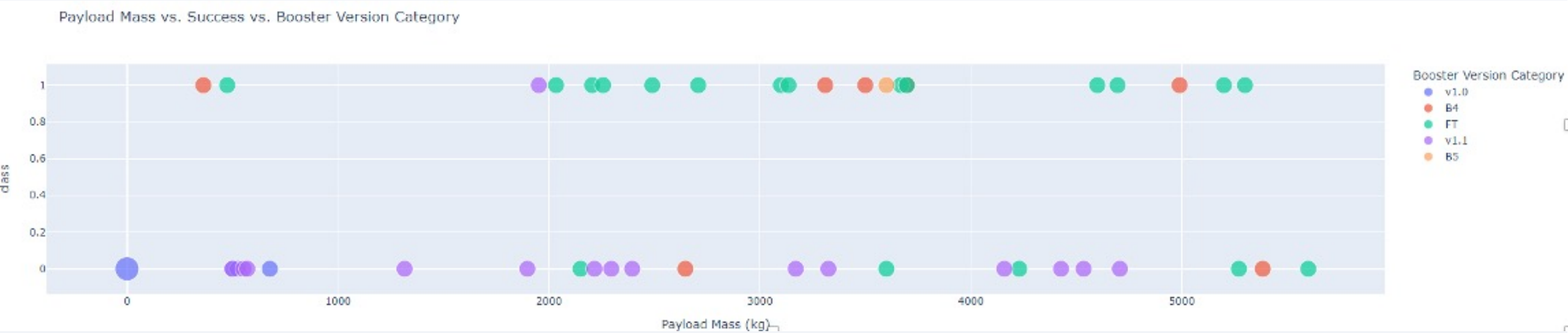


The launch site with highest launch success ratio

- KSC LC-39A is the highest success rate
- The color of blue means success landing
- Success rate is nearly 80%



Payload vs. Launch Outcome scatter plot for all sites

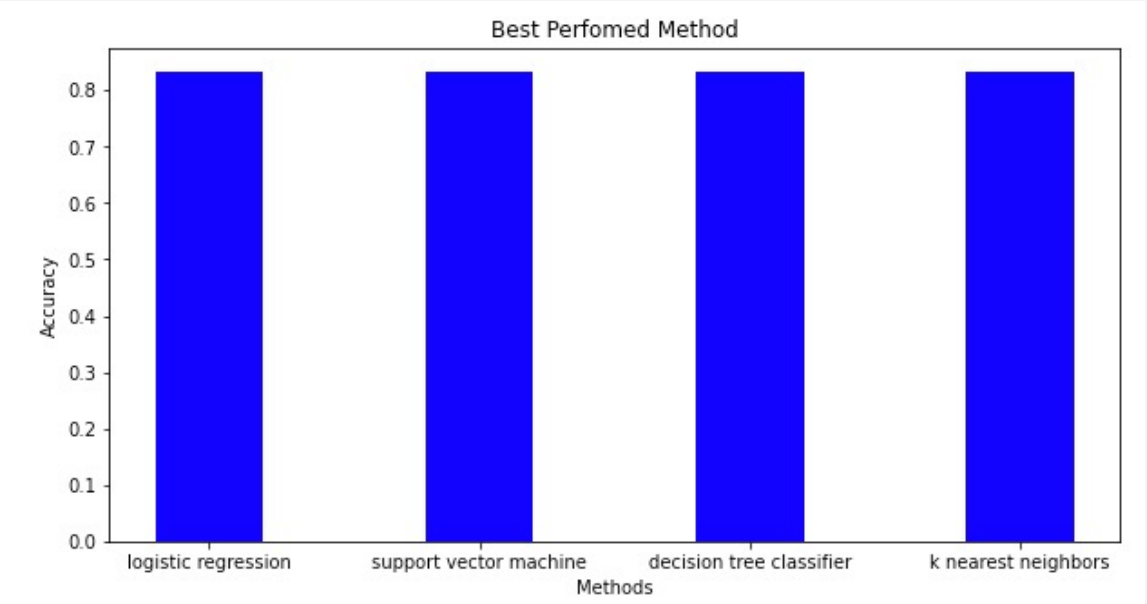


Section 6

Predictive Analysis (Classification)

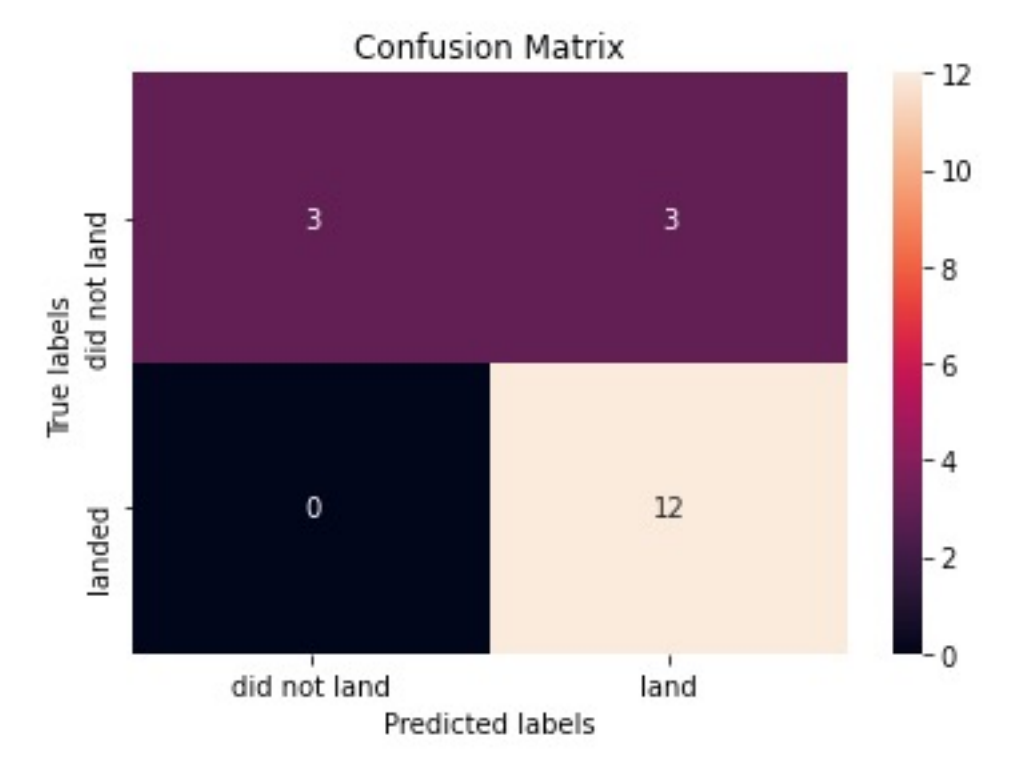
Classification Accuracy

- All models of accuracy rate are similar
- The accuracy rate is around 83%
- The size of data sample seems not enough to do analyzing



Confusion Matrix

- All models of confusion matrix result are the same
- The models are over predicting successful landing
- Need more data sample



Conclusions

- A machine learning models with an accuracy rate of 83%
- All models of confusion matrix result are the same
- The size of data sample seems not enough to do analyzing
- Need to improve accuracy rate and confusion matrix
- Find more data sample for further action

Appendix

- URL
- https://github.com/yushinglui/ibmAppliedDataScienceCapstone
- https://cocl.us/n40EW
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_5/submission_overview.md.html?origin=www.coursera.org

Thank you!