IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yusuke Ishida
October 19, 2024

Outline

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

Executive Summary

- Collected data
- Preprocess data
- Develop predictive model for launch outcome
- With over 83% accuracy. We can now predict whether a launch will be successful or not that would help in bidding contracts

Introduction

- Predicting the success of Falcon 9 first stage launch
- Successfully predicting the success allows us to determine the cost of a launch

Section 1

Methodology

Methodology

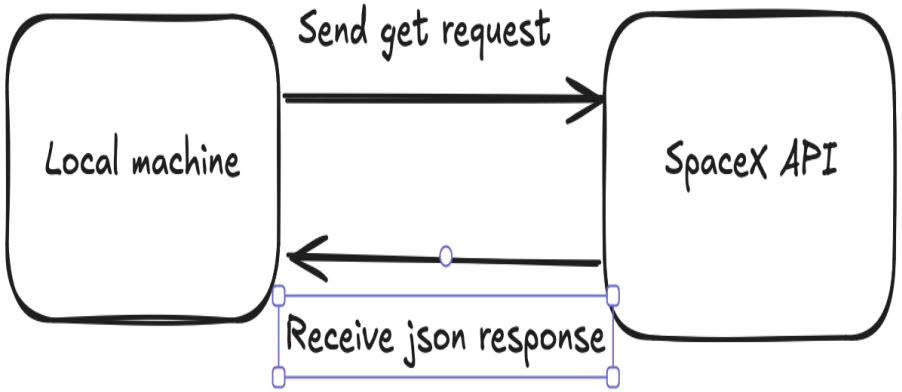
Executive Summary

- Data collection methodology:
  - SpaceX API
  - Scraping Wikipedia
- Perform data wrangling
  - Handling null values
  - Add labels based on Outcomes
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API

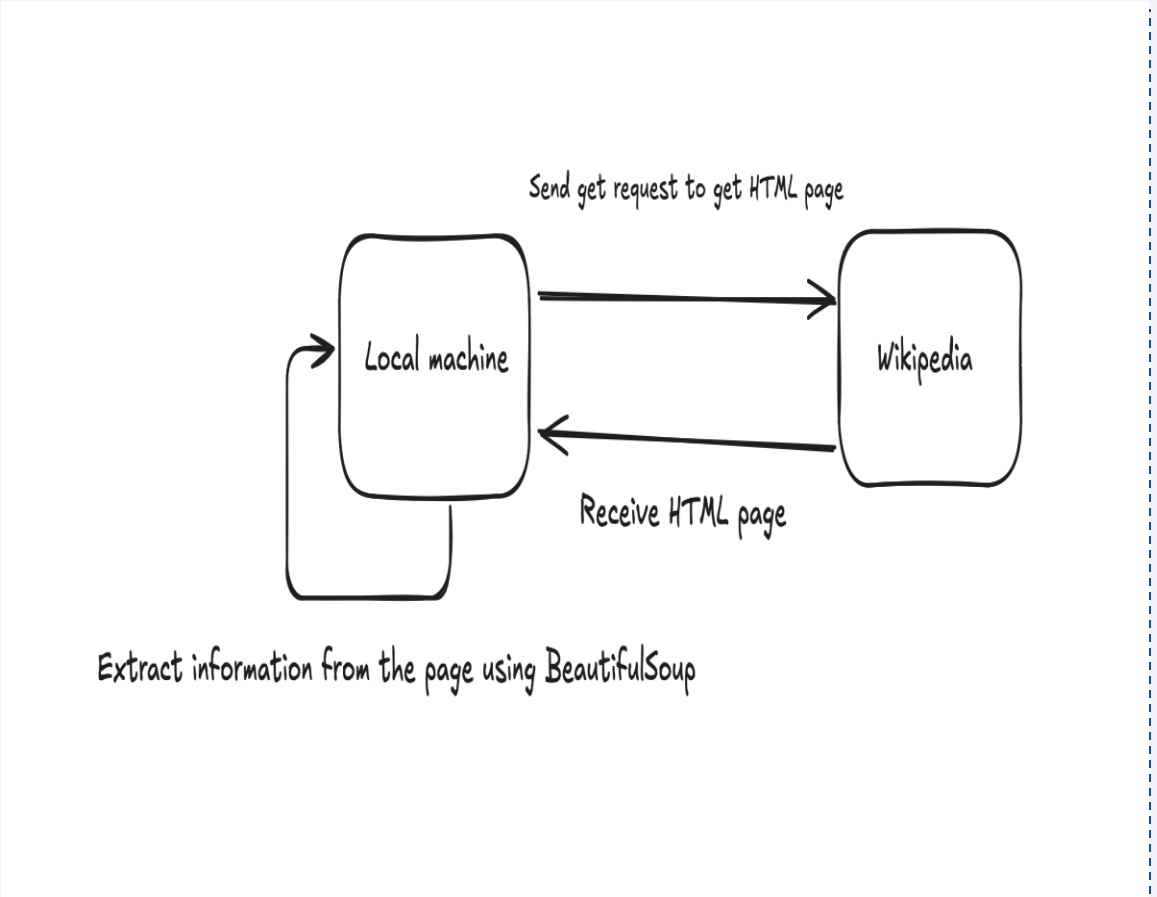
- Using requests library to send HTTP Get request.

- Notebook used for sending the request. https://github.com/yusuke0127/ibm-ds-capstone/blob/main/notebooks/jupyter-labs-spacex-data-collection-api.ipynb



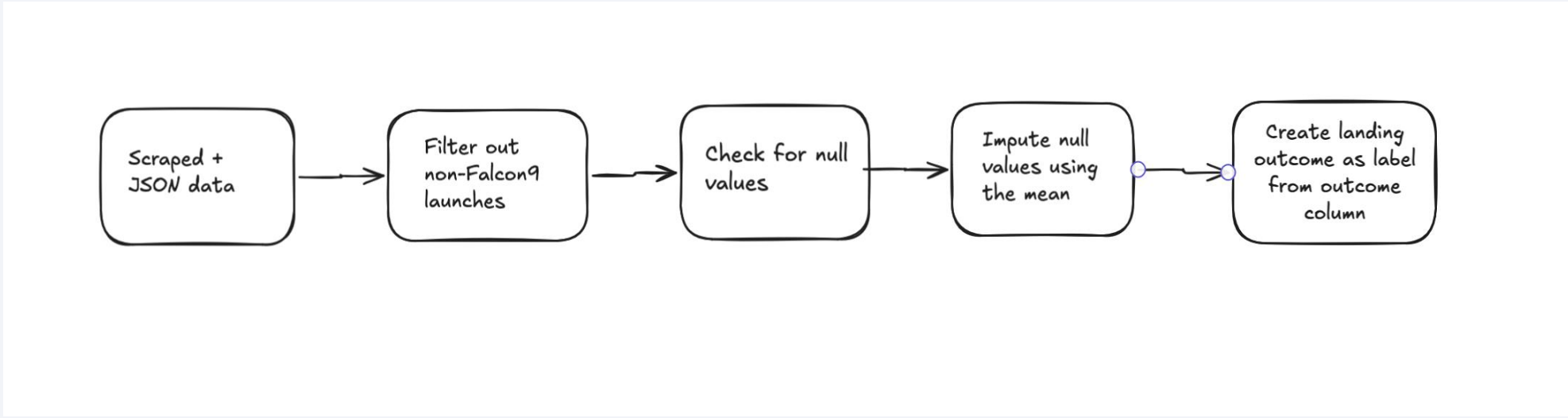
Data Collection - Scraping

- Use requests library to get HTML page and BeautifulSoup to extract information from the page.
- Notebook used for scraping. https://github.com/yusuke0127/ibm-ds-capstone/blob/main/notebooks/jupyter-labs-webscraping.ipynb



Data Wrangling

- Combined scraped data and JSON data into a single data frame
- Remove null values and fill it using the mean values of each columns.
- Added landing outcome labels from outcome columns
- https://github.com/yusuke0127/ibm-ds-capstone/blob/main/notebooks/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Visualized using scatter plot the launch site and number of flights to check the distribution of launches per launch site
- Visualized relationship between Payload mass and launch site to check correlation
- Visualized the relationship between orbits and success rate and found out that some orbits had higher success rate than others.
- Visualized yearly launch to see if there's a trend and it clearly shows that from 2013 that the success rate have been going up.
- https://github.com/yusuke0127/ibm-ds-capstone/blob/main/notebooks/edadataviz.ipynb

EDA with SQL

- Get the names of the launch sites
- Display 5 launch sites that starts with CCA
- Display total payload mass of launches boosted by NASA
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- https://github.com/yusuke0127/ibm-ds-capstone/blob/main/notebooks/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

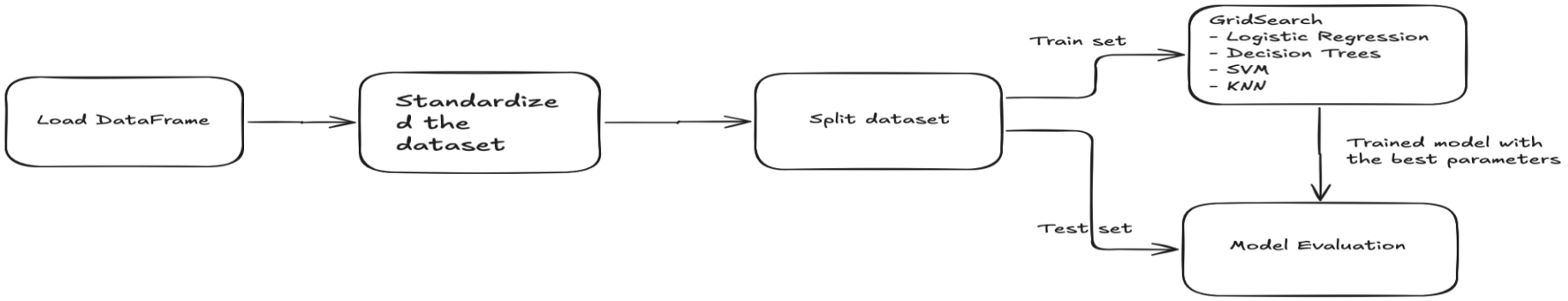
- Added a Circle and Marker to launch sites to identify where they are in the map.
- Also added Icon colors to markers to easily identify successful launches to failures.
- Added Lines to visualize distance of launch sites from the coast and railways
- Notebook used for map visualizations: https://github.com/yusuke0127/ibm-ds-capstone/blob/main/notebooks/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Added dropdown for selecting launch sites and pie graph to visualize number of launches between sites and a scatter plot with payload mass and success as the variables.
- Having a pie charts makes it easier to compare between different launch sites and a scatterplot helps in checking correlation between payload mass and successful launches.
- https://github.com/yusuke0127/ibm-ds-capstone/blob/main/spacex_dash_app.py

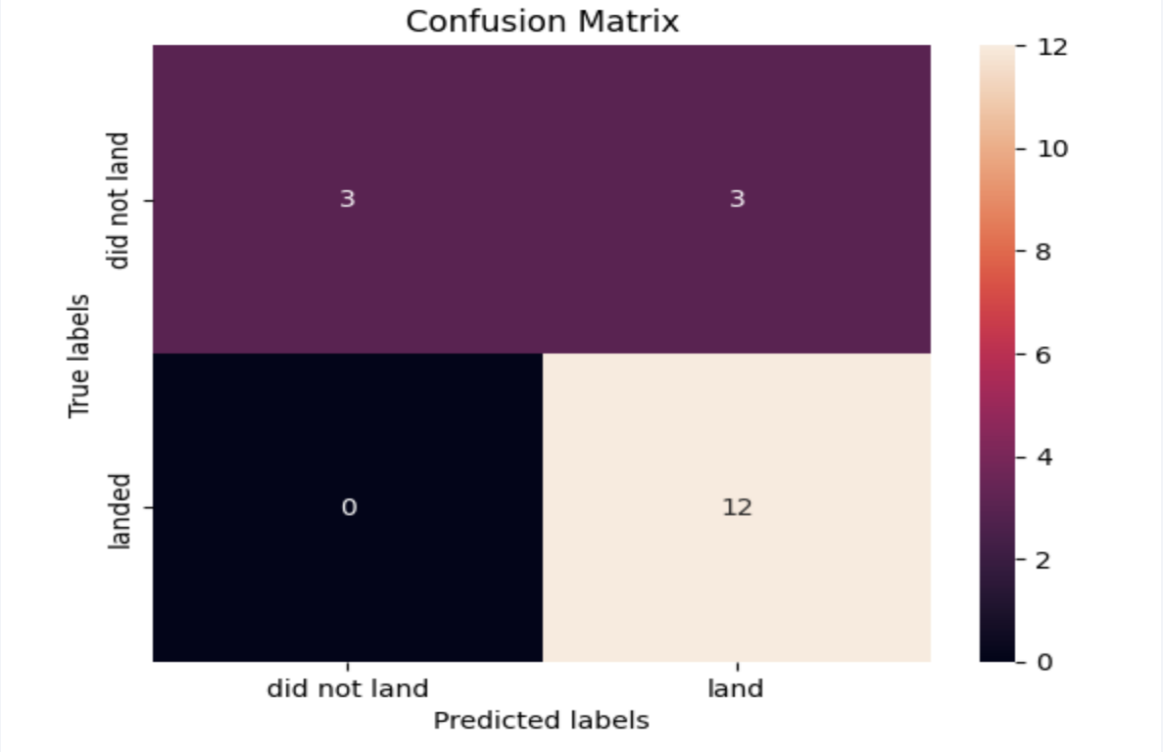
Predictive Analysis (Classification)

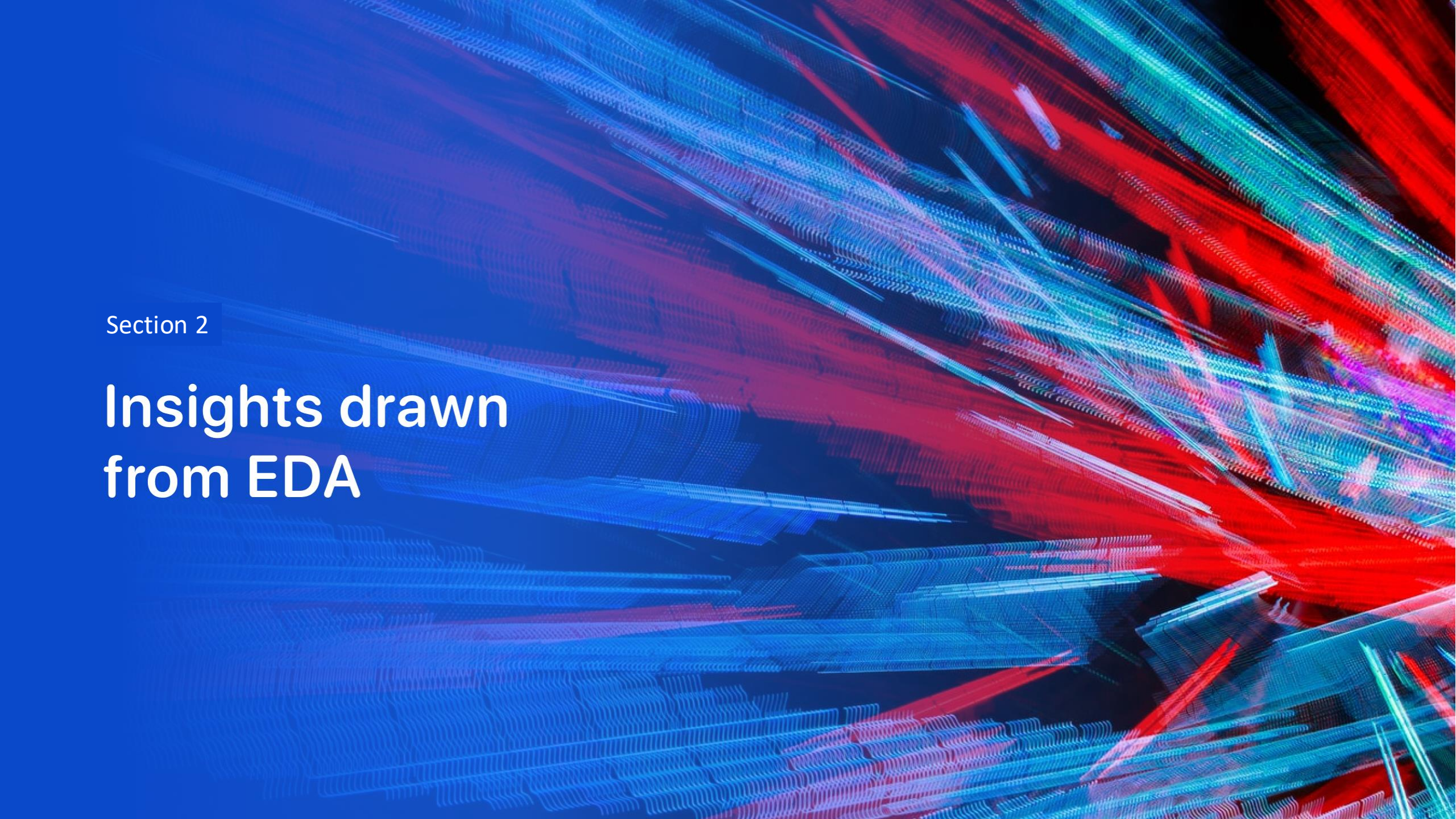
- Built the model by standardizing the dataset and splitting it into train and test set. Run Grid search for Logistic Regression, Decision Trees, SVM and KNN to get the best parameters for each model. All models except SVM scored the same 0.8333
- https://github.com/yusuke0127/ibm-ds-capstone/blob/main/notebooks/SpaceX_Machine%20Learning%20Prediction_Par



Results

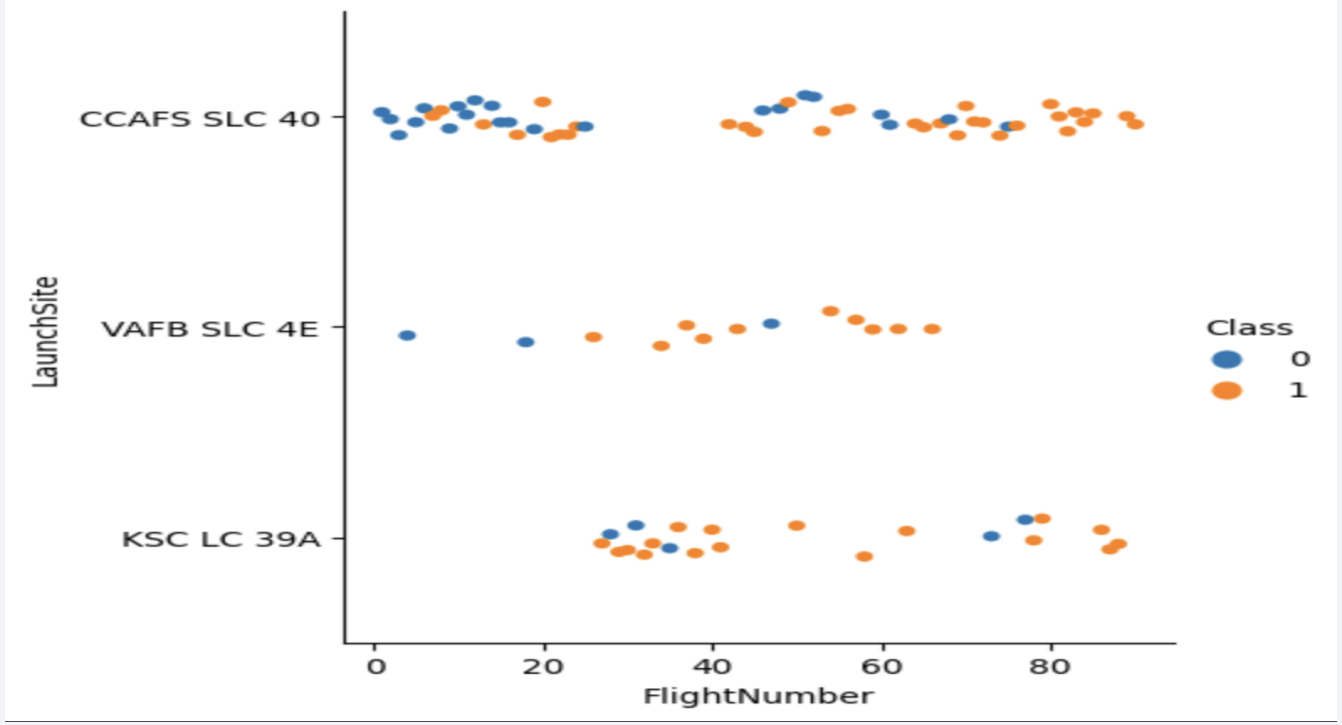
- A KNN model gives us 83% accuracy whether a launch will be successful or not




Section 2
Insights drawn from EDA

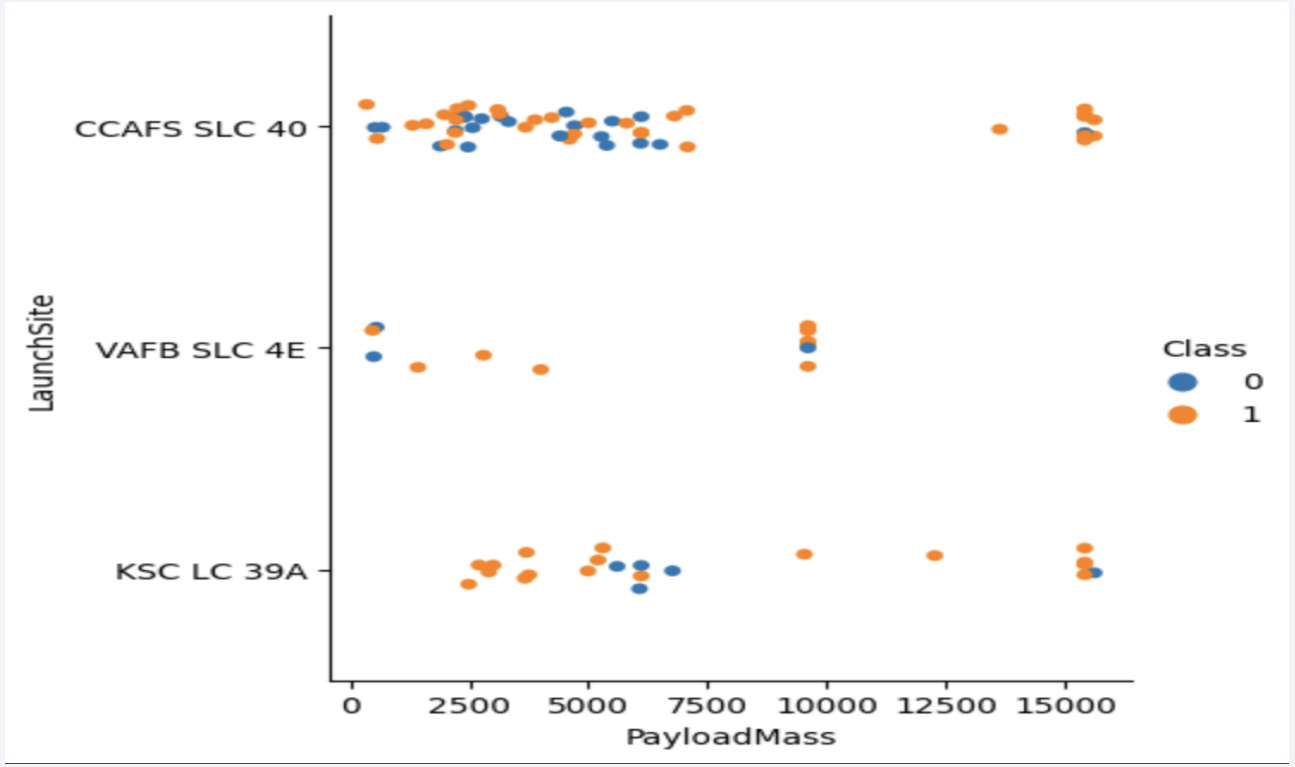
Flight Number vs. Launch Site

- The higher number of flights for both CCAFS SLC-40 and VAFB SLC4E tends to have higher success rates while it is not the case for KSC LC 39A.



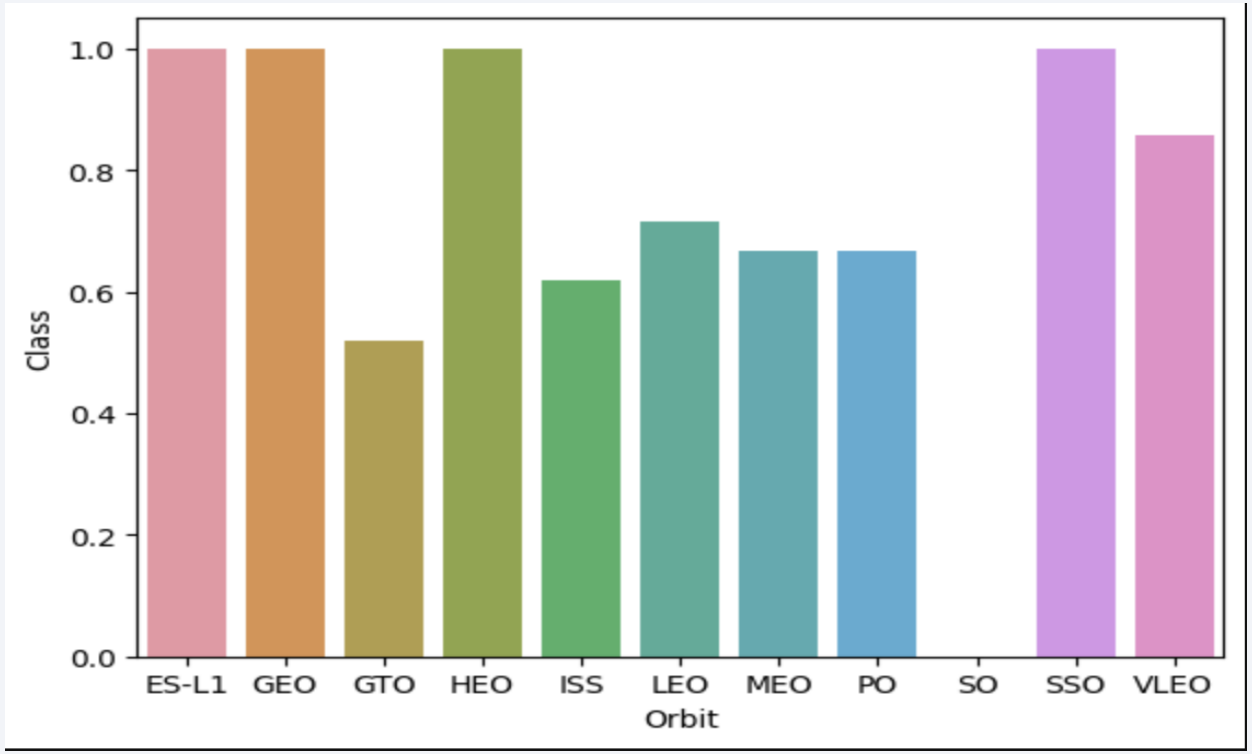
Payload vs. Launch Site

- The higher the heavier payload mass for CCAFS SLC 40 seems to have better success rate while that's not necessarily the case for both VAFB SLC 4E and KSC LC 39A



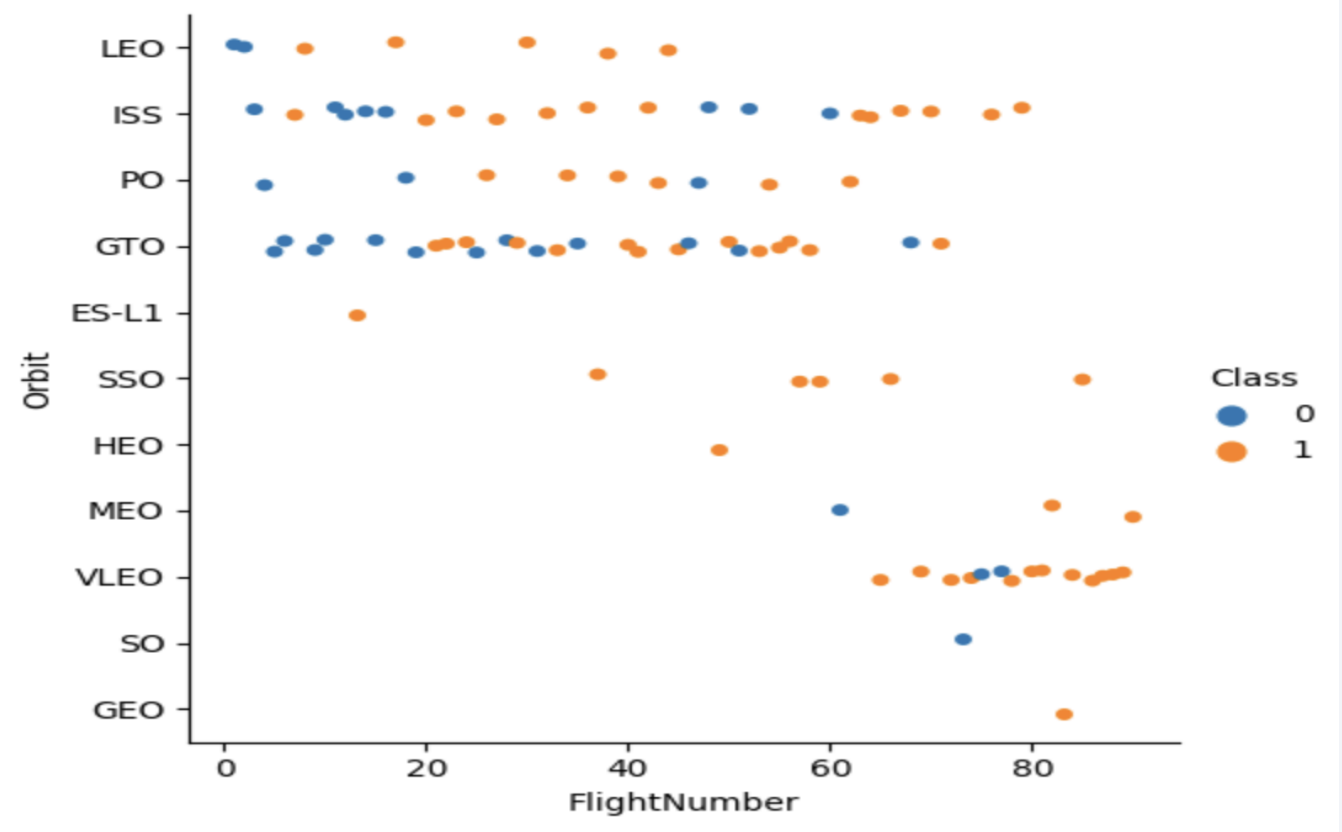
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have 100% success rate and SO have 0% success rate



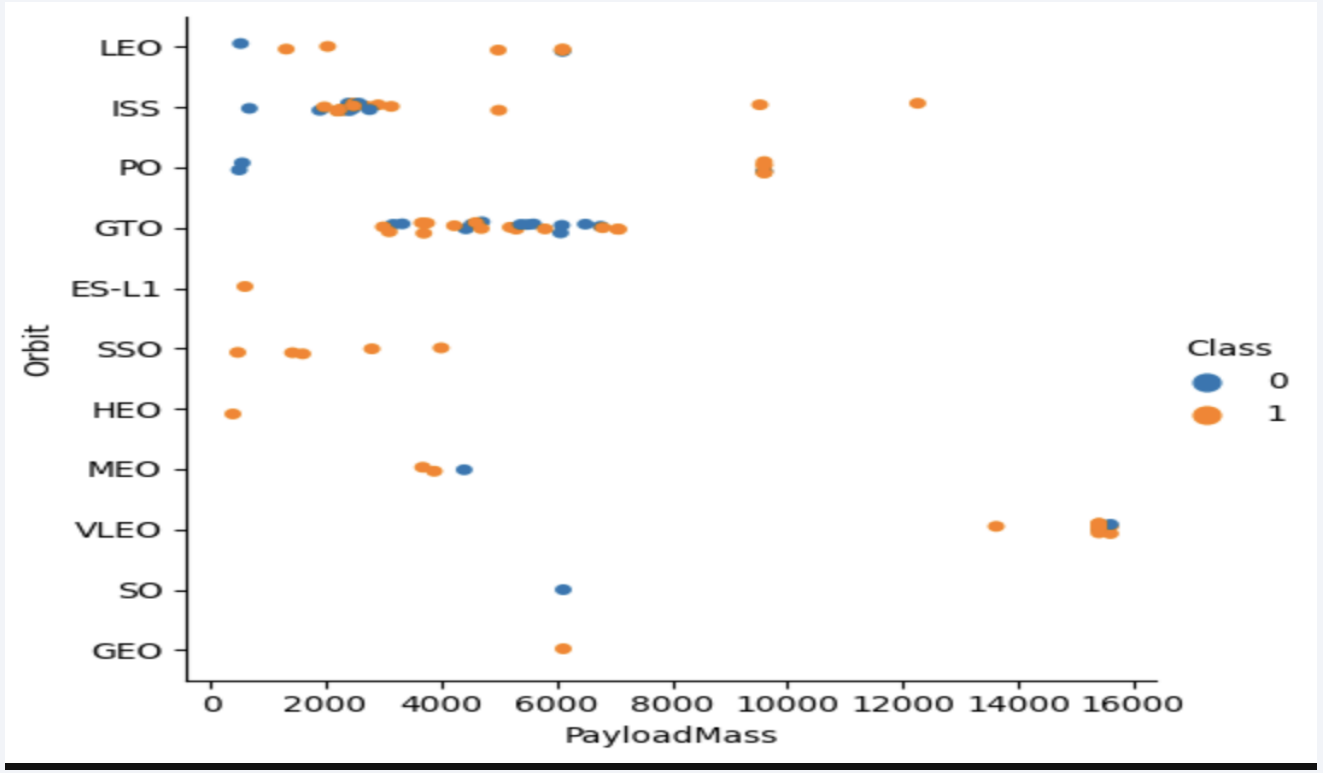
Flight Number vs. Orbit Type

- LEO and MEO have a higher success rate the more flights



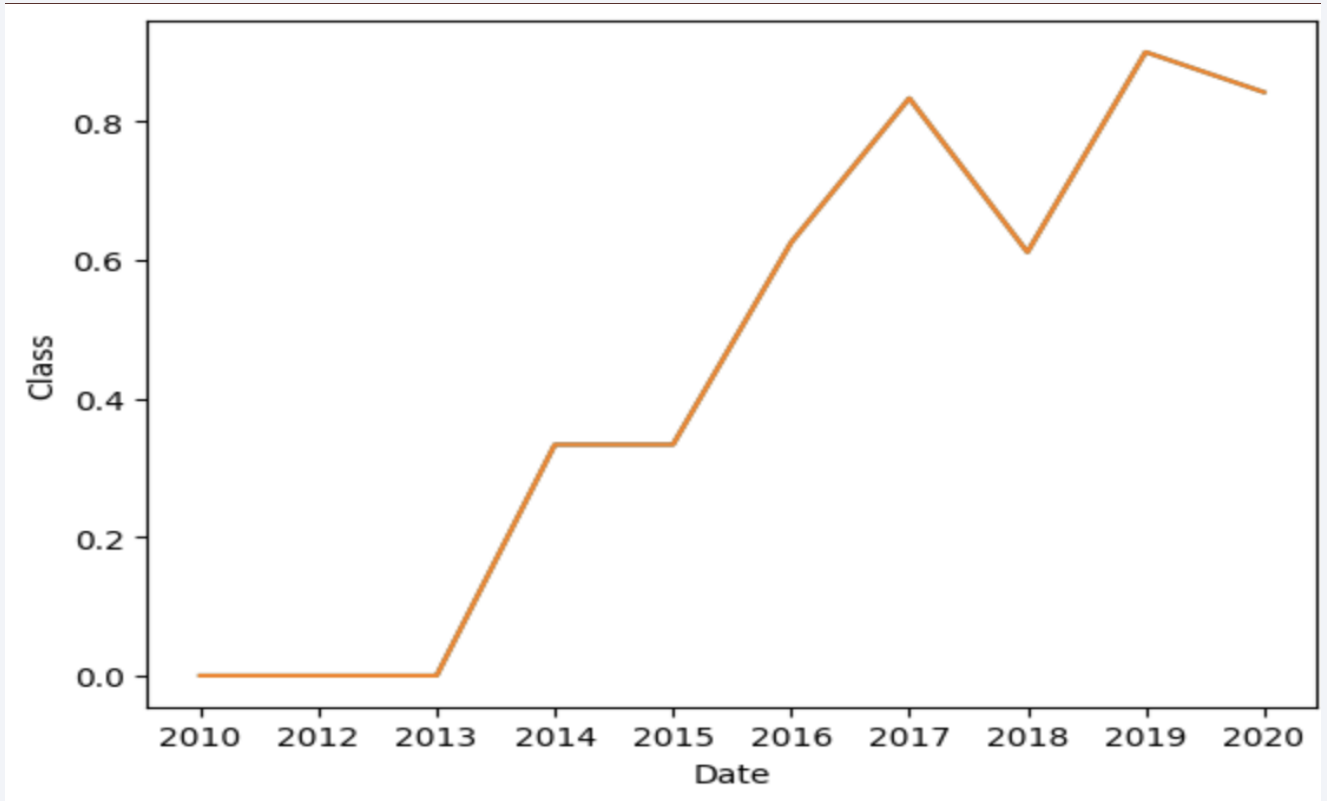
Payload vs. Orbit Type

- LEO, ISS AND PO seems to have better success rates the higher the payload mass while it's not the case for GTO



Launch Success Yearly Trend

- Success have been to upward trend since 2013



All Launch Site Names

- There are 4 launching sites

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Found the query using LIKE command

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Lan
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Fa
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	Fa
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	

Total Payload Mass

- Total payload carried by NASA (CRS)

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 using AVG function

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2013-12-03	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success

First Successful Ground Landing Date

- 2015 was the first year to have successful ground landing

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Land
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Su
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Su
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Su
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Su
2017-06-03	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708	LEO (ISS)	NASA (CRS)	Success	Su

Successful Drone Ship Landing with Payload between 4000 and 6000

- Only 4 successful drone ship landing with payload between 4000 and 6000

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landi
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	S
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	S
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	S
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	S

Total Number of Successful and Failure Mission Outcomes

- Theres more successful missions

Landing_Outcome	COUNT(*)
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

- F9 B5 B1060.3 has carried the highest payload

Booster_Version	MASS
F9 B5 B1060.3	15600

2015 Launch Records

- All 2015 launch records

SUBSTR(Date, 6, 2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
02	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
03	No attempt	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	No attempt	F9 v1.1 B1016	CCAFS LC-40
06	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

- Landing outcomes between the date 2010-06-04 and 2017-03-20 has the same number of success as it has failures

Landing_Outcome	outcomes	RANK() OVER(ORDER BY COUNT(*) DESC)
No attempt	10	1
Success (drone ship)	5	2
Failure (drone ship)	5	2
Success (ground pad)	3	4
Controlled (ocean)	3	4
Uncontrolled (ocean)	2	6
Failure (parachute)	2	6
Precluded (drone ship)	1	8

Section 3

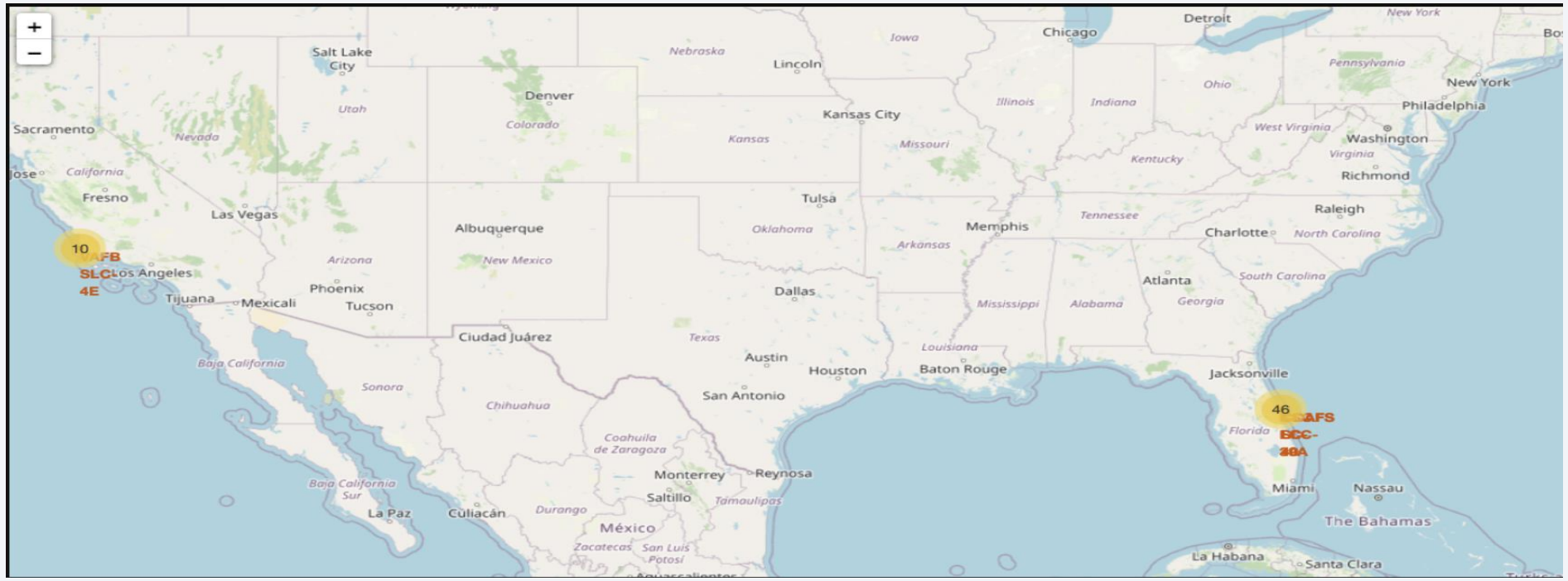
Launch Sites Proximities Analysis

Launch Site Locations

- There are two major launch sites in the US

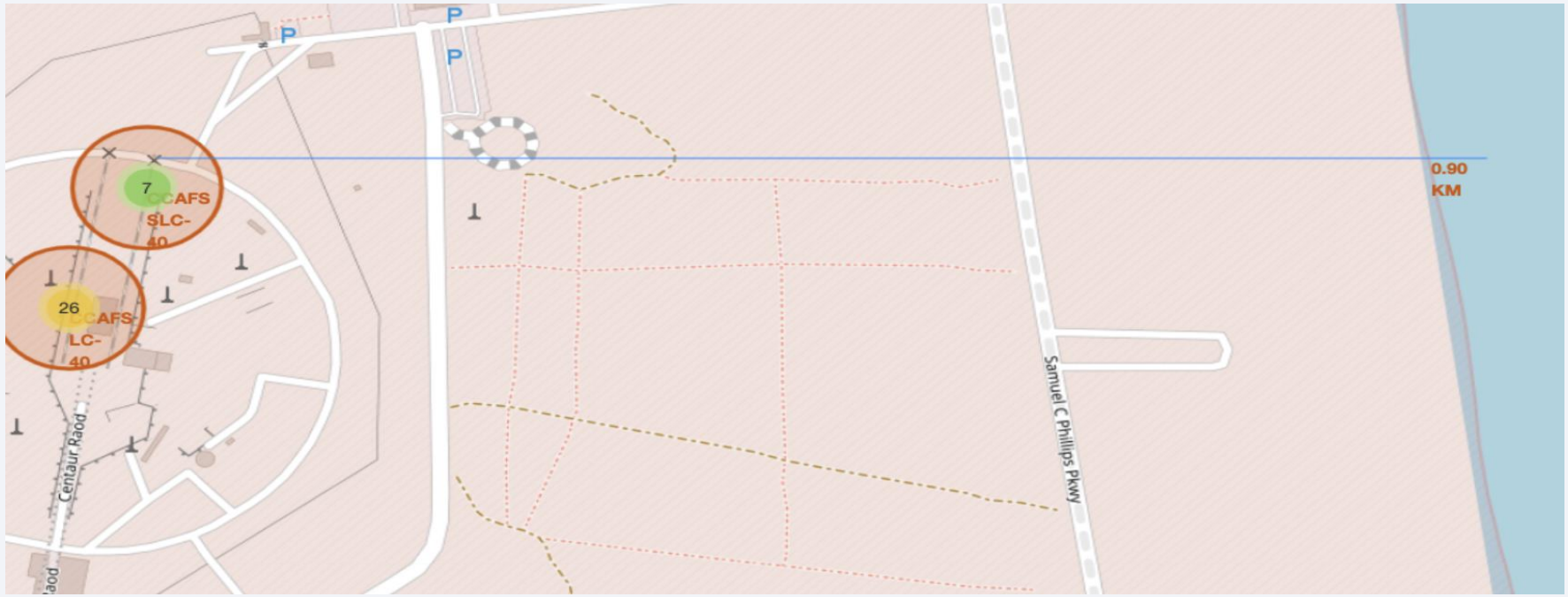
Number of launches per site

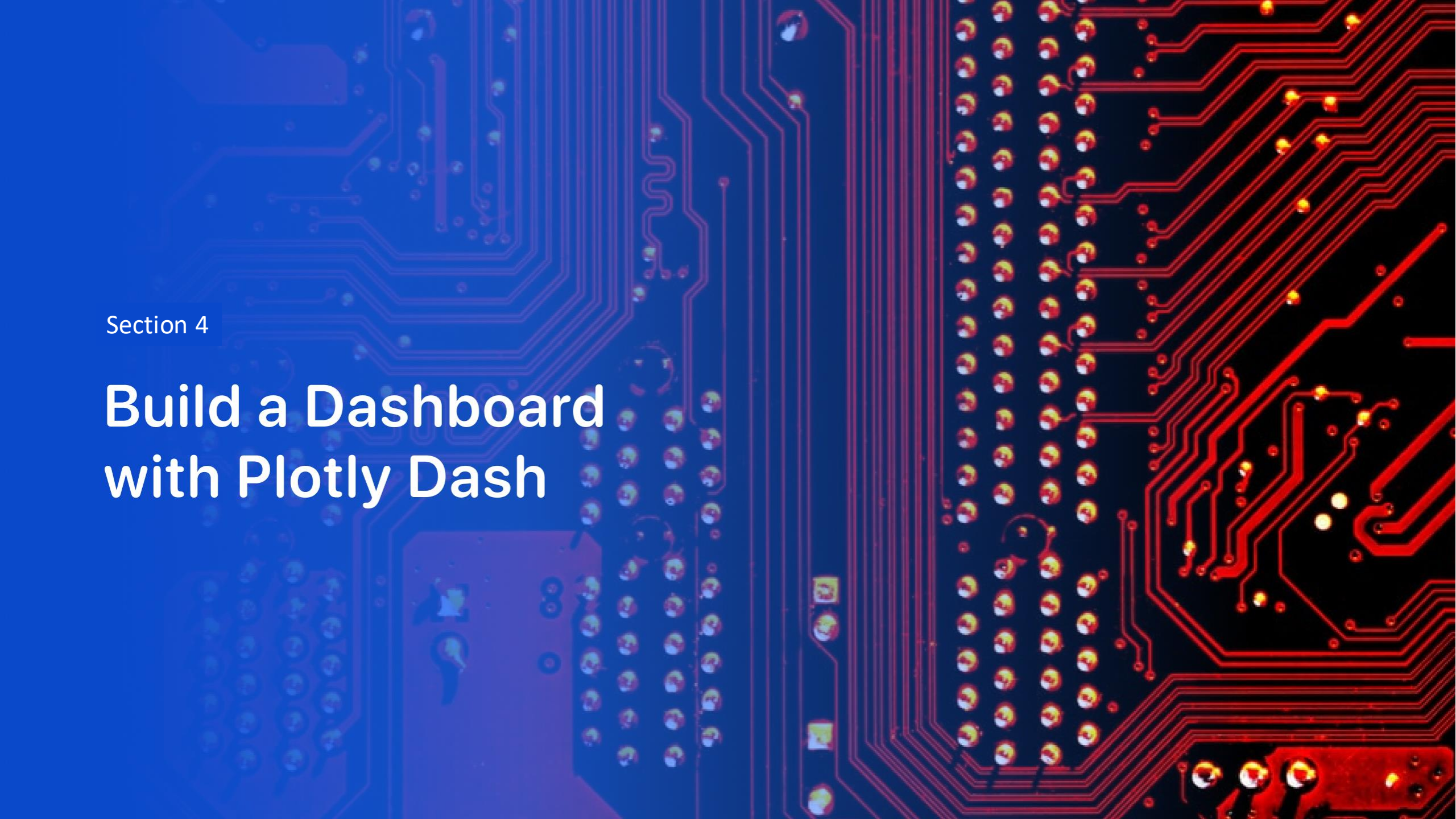
- Successful launches are marked green and failed ones red. We can see there's a fair number of failed and successful launches per site hence the yellow marker.



Common theme for launching sites

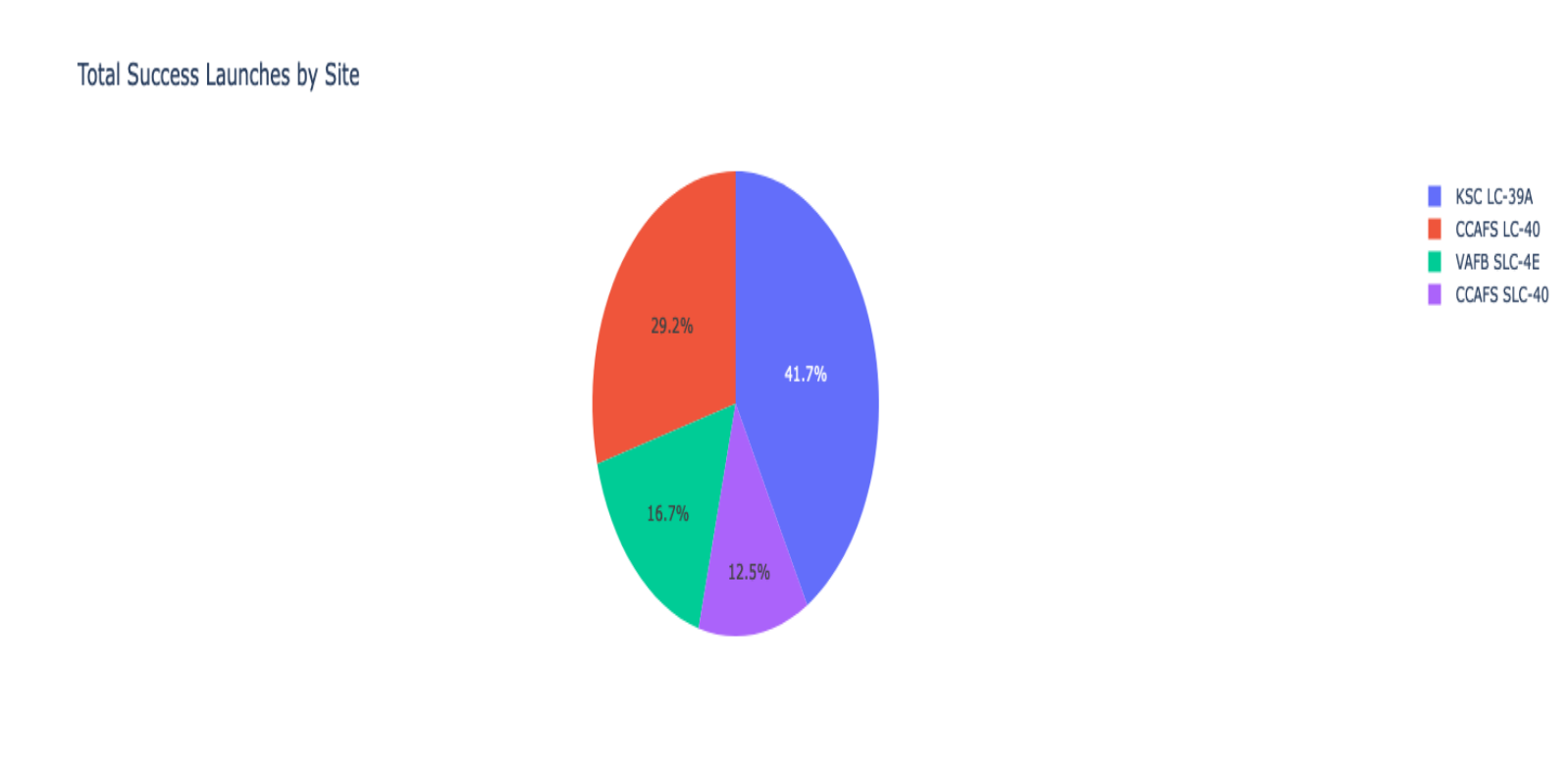
- All sites are very close to the port, railway tracks and highway roads.




Section 4
Build a Dashboard with Plotly Dash

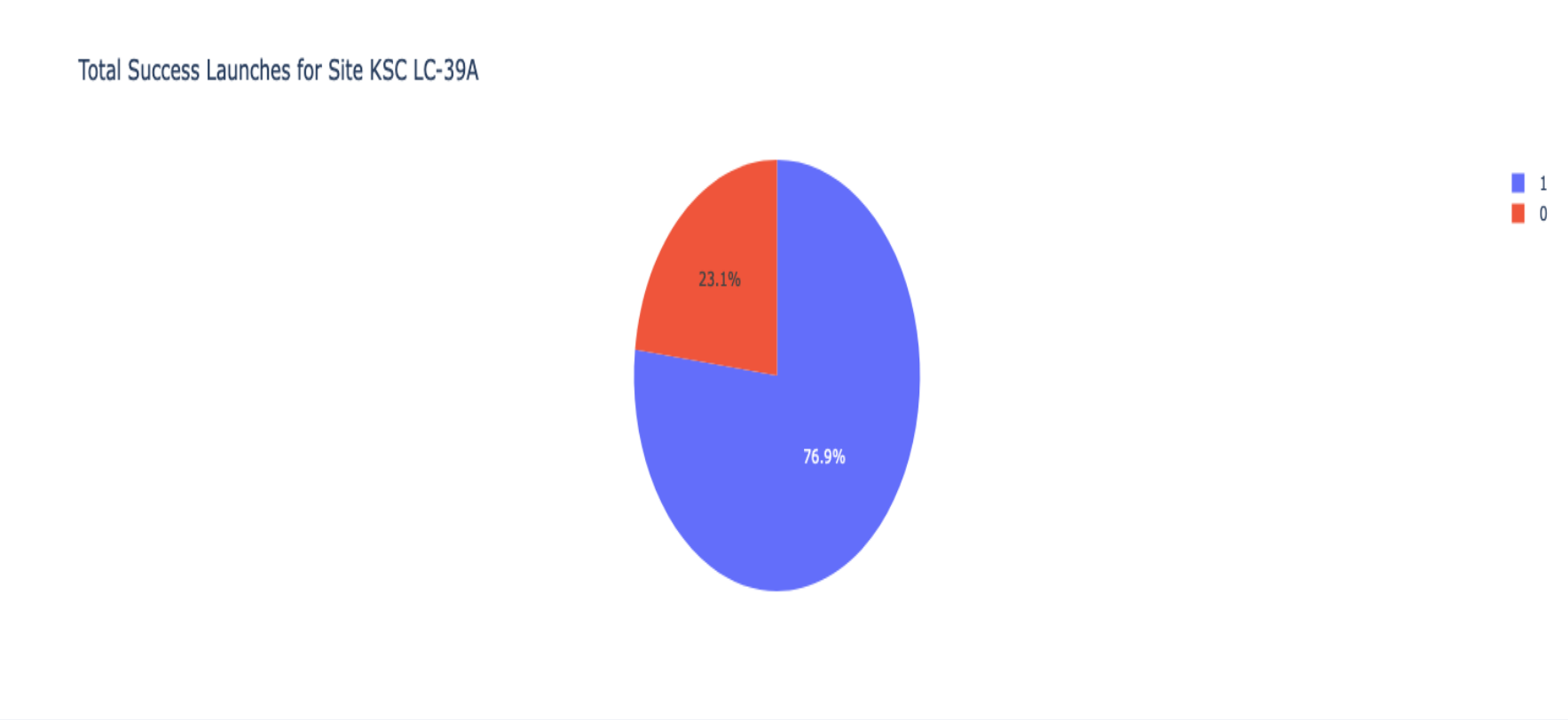
Total Success Launches

- Majority of successful launches was launched from from KSC LC-39A



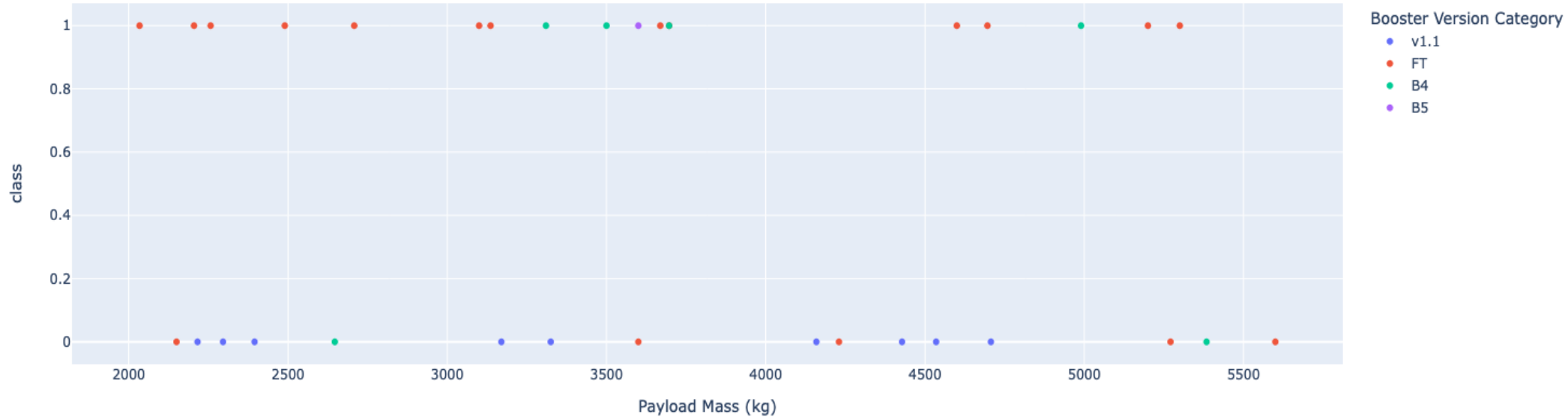
Launch success ratio for site KSC LC-39A

- KSC LC-39A also has the highest success launch rate



Most common payload range

- Most launches have a payload between 2,000 to 6,000 kilos. With FT booster being the most successful and v1.1 as the least successful one.

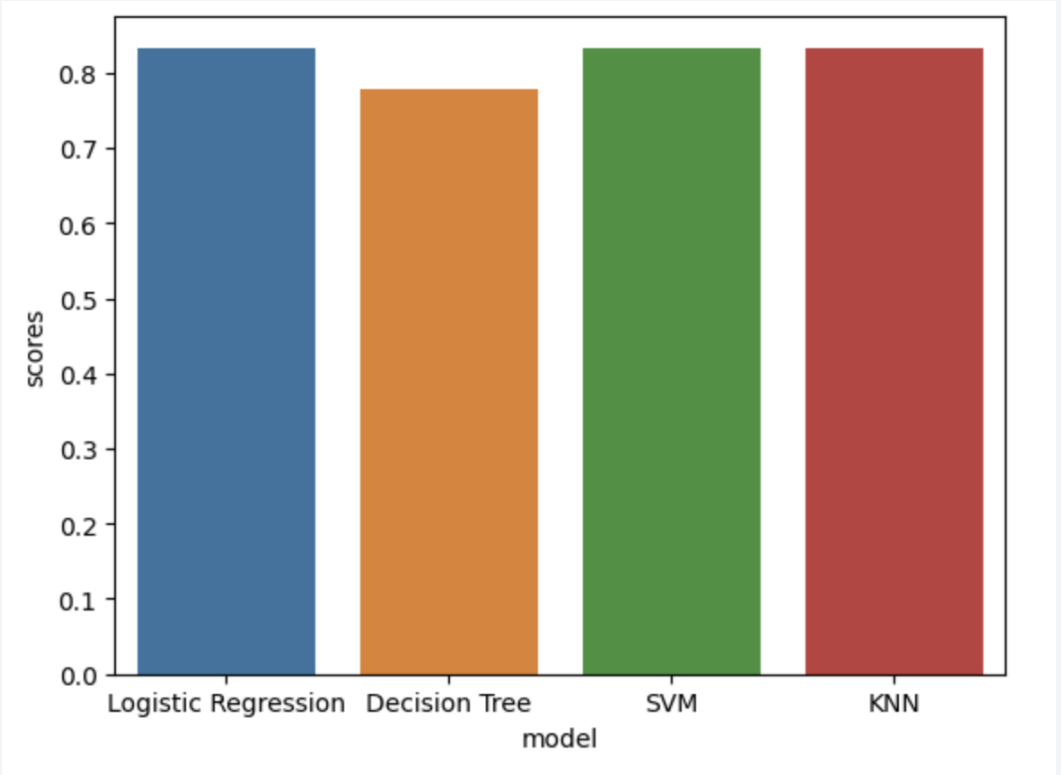


Section 5

Predictive Analysis (Classification)

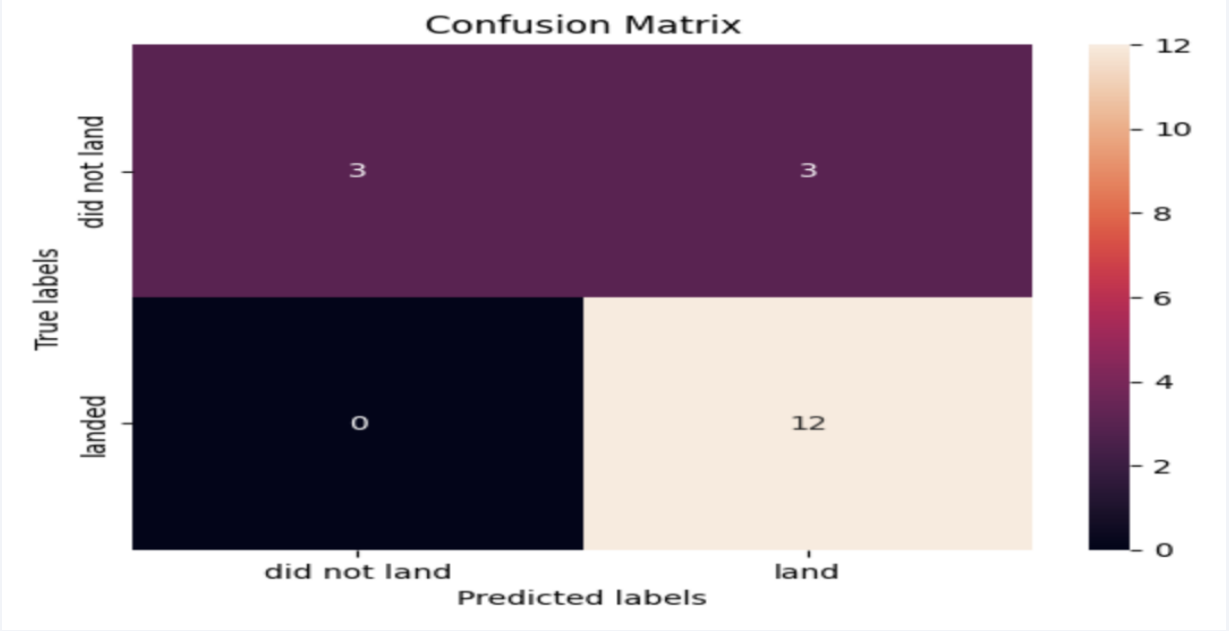
Classification Accuracy

- Most models performs the same except for Decision Tree.



Confusion Matrix

- Since SVM, Logistic Regression and KNN have the same scores which gives multiple options. Explainability is also important hence KNN seems to be the ideal one to use.



Conclusions

- All launch sites are close to the coast
- KSC LC-39A have the most successful launches and also have the most successful rate with CCAFS LC-40 a close second
- CCAFS SLC-40 is the site with the least successful launch rate
- Most launches have a payload between 2000 to 6000 kilos with FT being the most successful
- We can predict with high accuracy if a launch will be successful or not

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!