IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jainish Patel
1/5/2023

Outline

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

Executive Summary

❖Summary of Methodologies:

- Data Collection API
- Data Collection with Web Scrapping
- Data Wrangling
- EDA with SQL
- EDA with Visualization
- Visual Analytics with Folium
- Dashboard using Plotly
- Prediction using Machine Learning

❖Summary of Results:

- Exploratory Data Analysis Results
- Predictive Analysis Results

Introduction

- **Project background and context:**
  - SpaceX a rocket company launches satellites at low price like 70% less than their competitor since they land their satellites for reusing them to launch.
- **Problems you want to find answers:**
  - We use the previous data of launches of Falcon 9 rocket to predict the probability of the booster landing back to the pad influenced/correlated with the space launch site, the payload orbit, mass, landing pad location and the version of the booster.

Section 1

Methodology

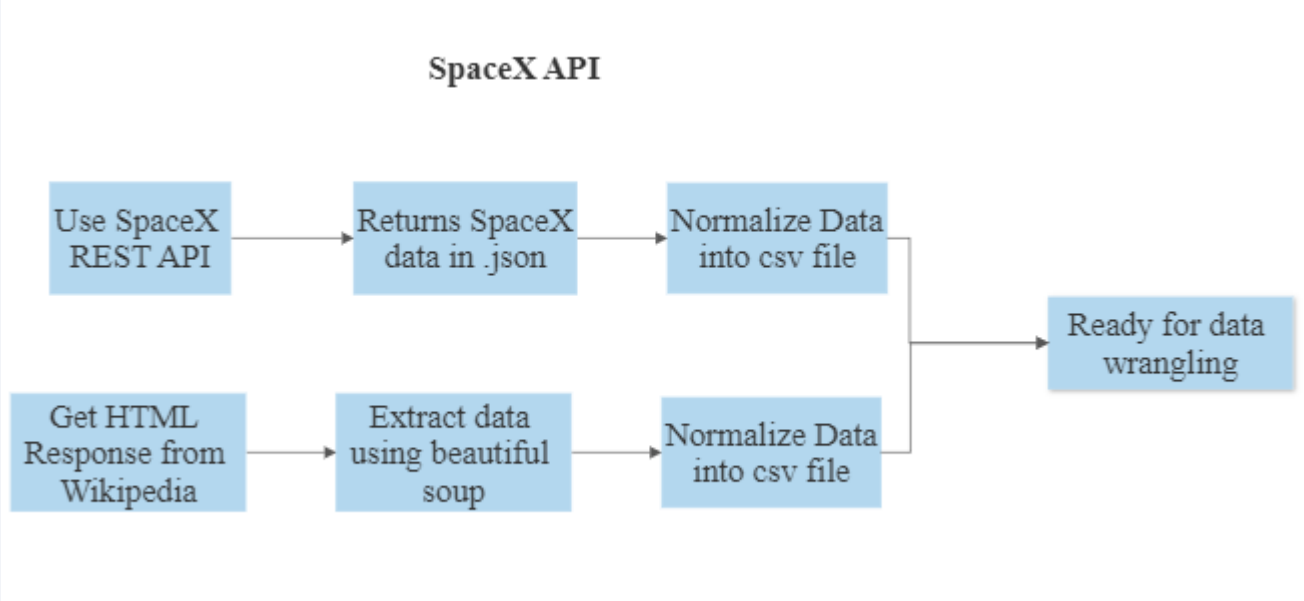
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
- Perform data wrangling:
  - Web Scrapping from Wikipedia
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LR, KNN, SVM models have been built and evaluated for best classifier

Data Collection

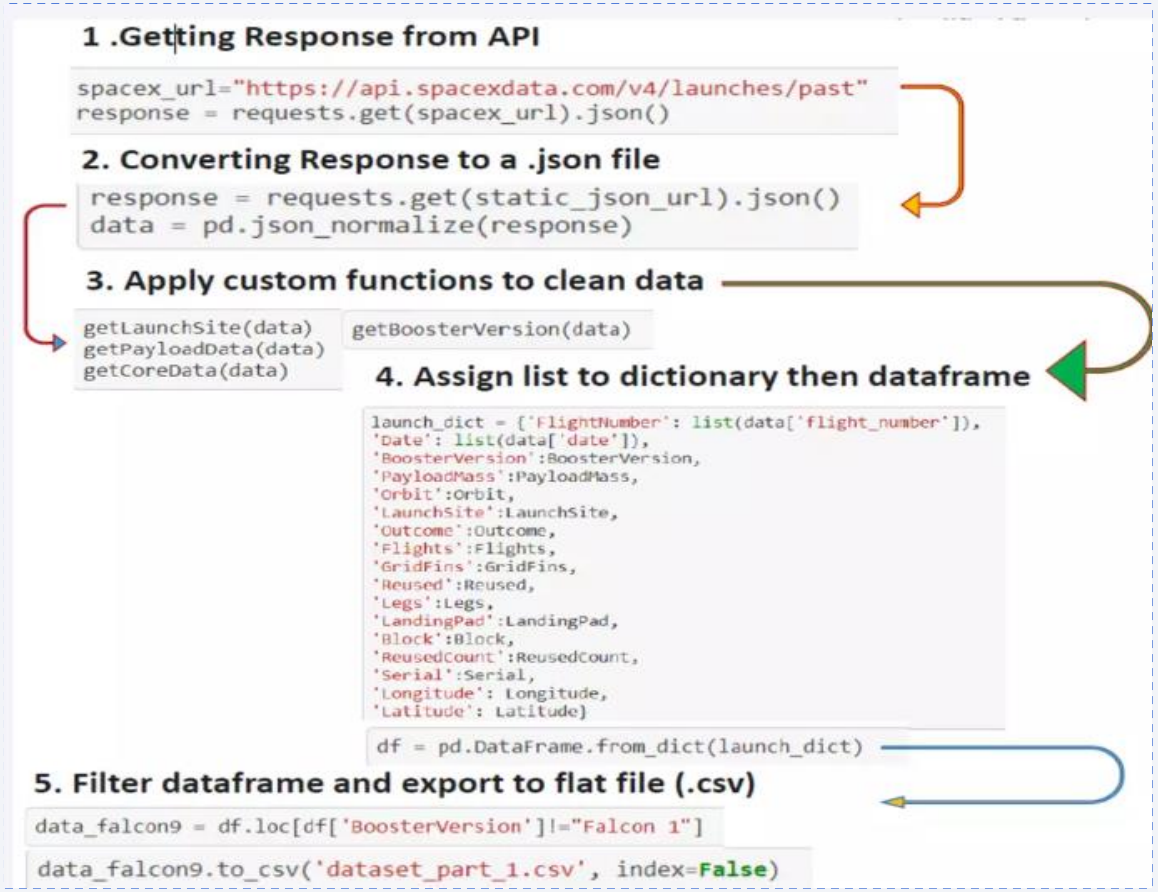
- REST API :
  - Using the REST API we extract the data in form of JSON and transform it to a dataframe using inbuilt python pandas method normalize .
- WEB SCRAPING :
  - Web scraping SpaceX launches from Wikipedia and converting it into a dataframe.



Data Collection – SpaceX API

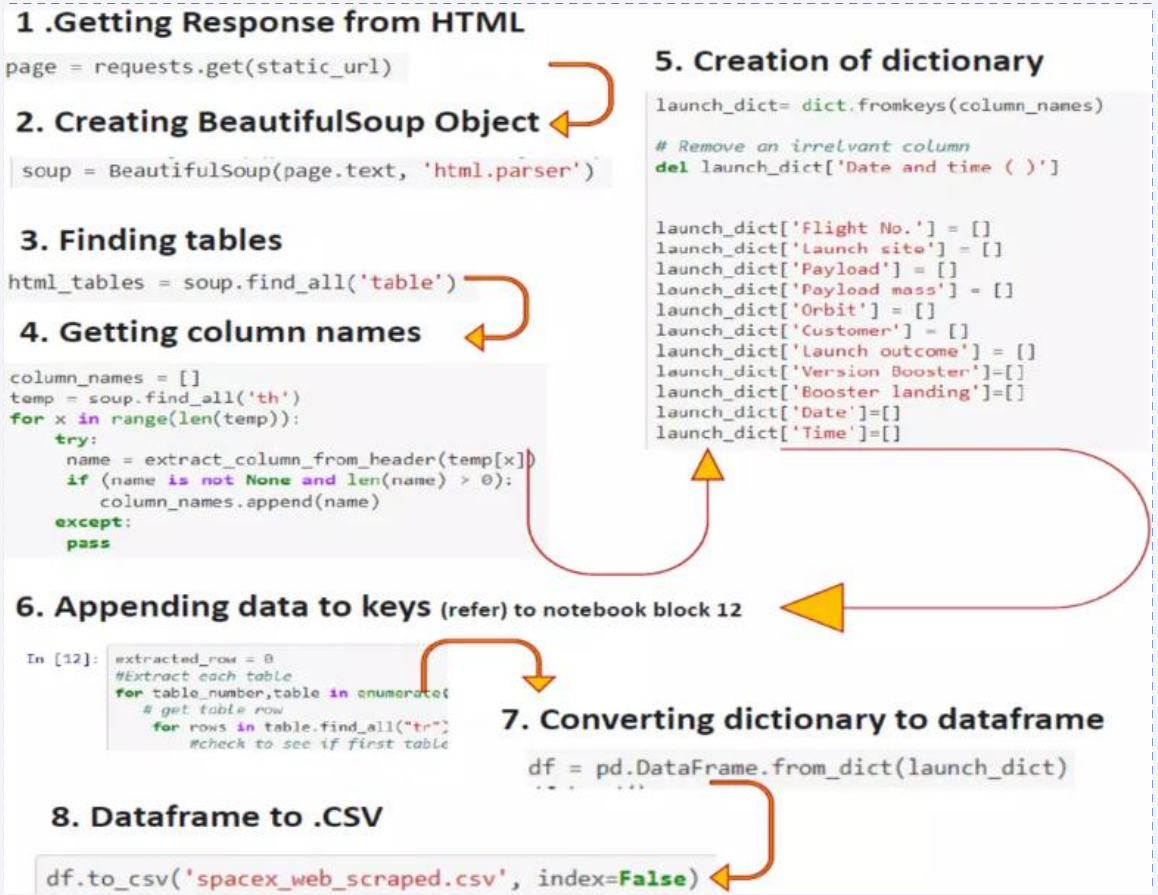
- SpaceX launch data that is gathered from the SpaceX REST API.
- This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications and landing outcomes.

- GitHub: https://github.com/Jainish1019/Applied_Data_Science/blob/master/Data%20Collection%20API-1.ipynb



Data Collection - Scraping

- In the data set, there are several different cases where the booster did not land successfully.
- GitHub: https://github.com/Jainish1019/Applied_Data_Science/blob/master/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- In the data set, there are several different cases where the booster did not land successfully.
- Sometimes a landing was attempted but failed due to an accident; for example, True Ocean means the mission outcome was successfully landed to a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean. True RTLS means the mission outcome was successfully landed to a ground pad False RTLS means the mission outcome was unsuccessfully landed to a ground pad.
- True ASDS means the mission outcome was successfully landed on a drone ship False ASDS means the mission outcome was unsuccessfully landed on a drone ship.
- GitHub: https://github.com/Jainish1019/Applied_Data_Science/blob/master/Data%20Wragling.ipynb

EDA with Data Visualization

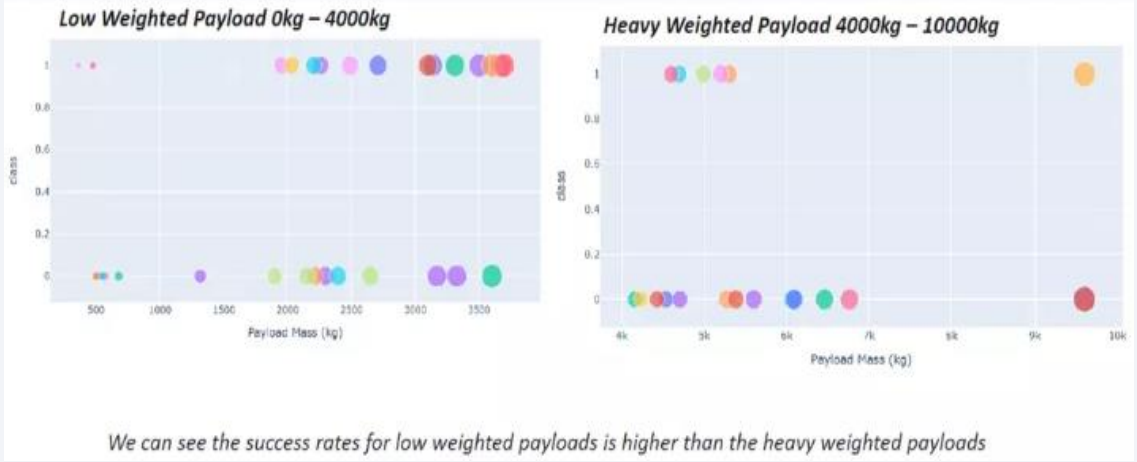
- Graphs used:
  - Scatter Plot
  - Bar Chart
  - Line Chart
- Because they are the once that best highlight the relationship between variables considered.
- GitHub: https://github.com/Jainish1019/Applied_Data_Science/blob/master/EDA%20with%20Visualization.ipynb

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 successful landing outcome in the 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.
- List of total number of successful and failure mission outcomes.
- List the names of the booster_versions which have carried the maximum payload mass.
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015.
- GitHub: https://github.com/Jainish1019/Applied_Data_Science/blob/master/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Summary of map objects:
  - Markers: Show a geo location from latitude and longitude data
  - Cluster: Show a group of markers
  - Circles: Show a single location
  - Lines: Show distance between two I have added object to find some geographical patterns about launch site
- GitHub: https://github.com/Jainish1019/Applied_Data_Science/blob/master/Interactive%20Visual%20Analytics%20with%20Folium.ipynb



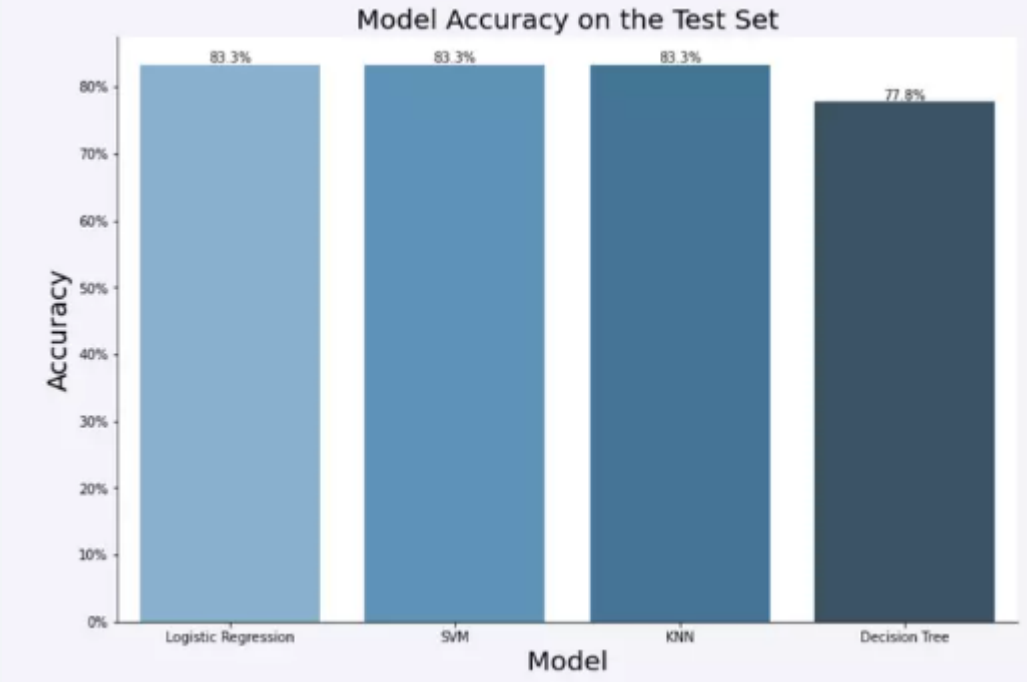
Build a Dashboard with Plotly Dash

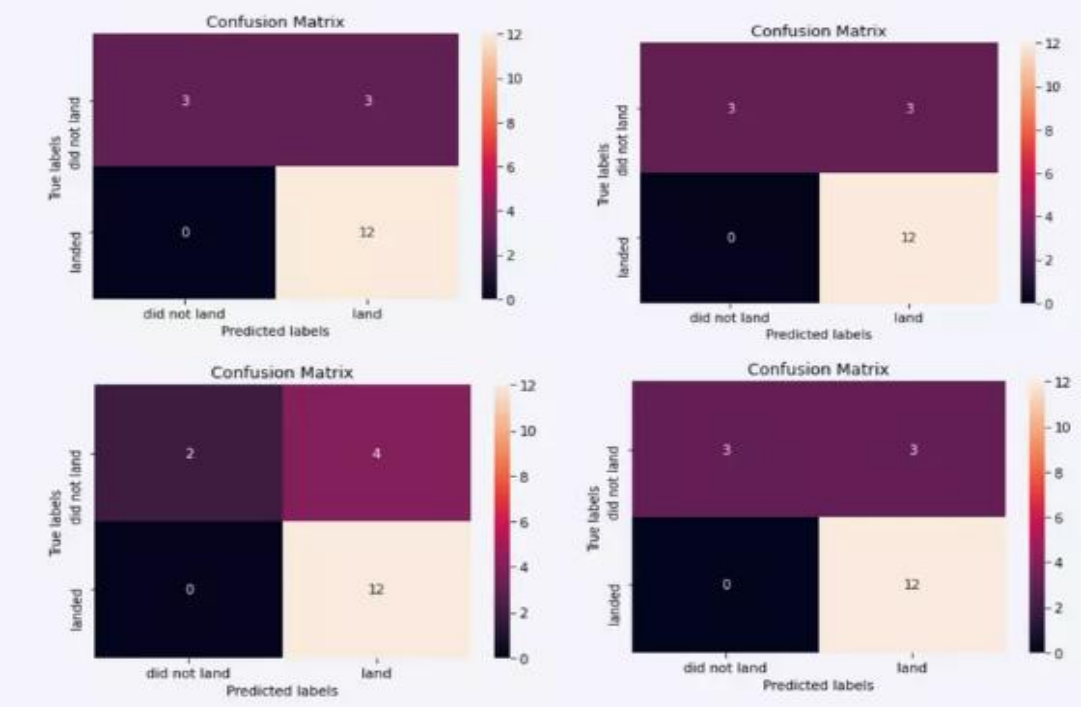
- Summary of plots:
  - Bar: Show categories differences
  - Line: Reports time series changes
  - Pie: Shows the percentage of events
  - Tree: Shows complex relationship of variables in interactive way
  - Map: Shows variables of states on a map

- GitHub: https://github.com/Jainish1019/Applied_Data_Science/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

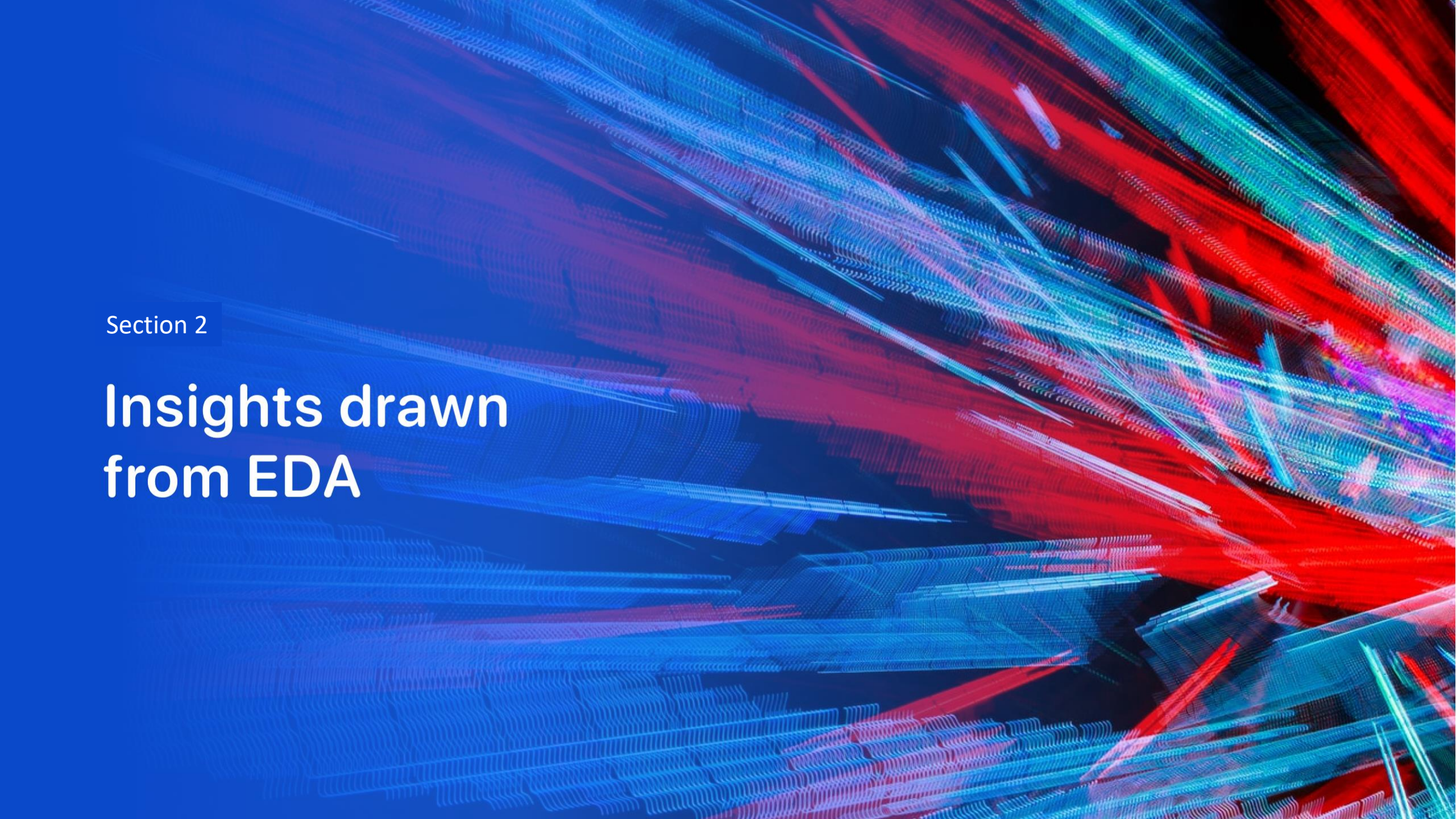
- The SVM, KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.



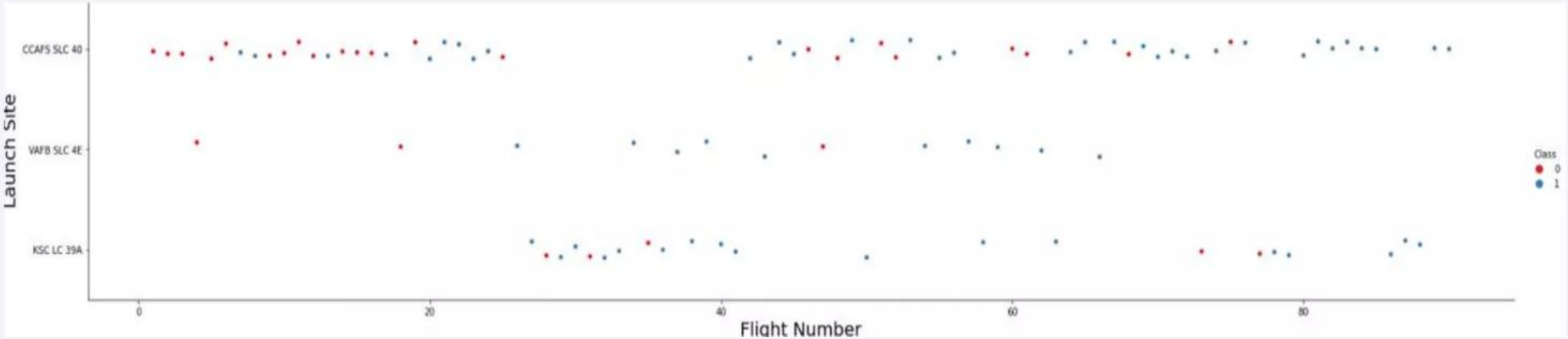


Results

- Exploratory data analysis results
  - Both API and web scraping are capable to collect SpaceX data.
- Interactive analytics demo in screenshots
  - EDA with SQL is effective for data filtering
  - EDA with interactive visualization provides informative information
  - Plotly Dash is powerful to show instant data change •
- Predictive analysis results
  - Decision Tree Classifier Algorithm has the best accuracy of predicting

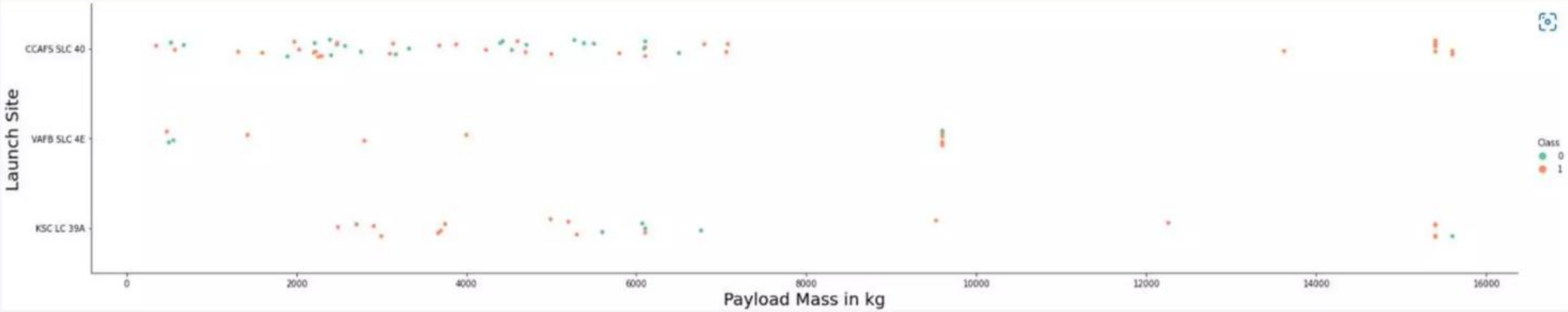

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



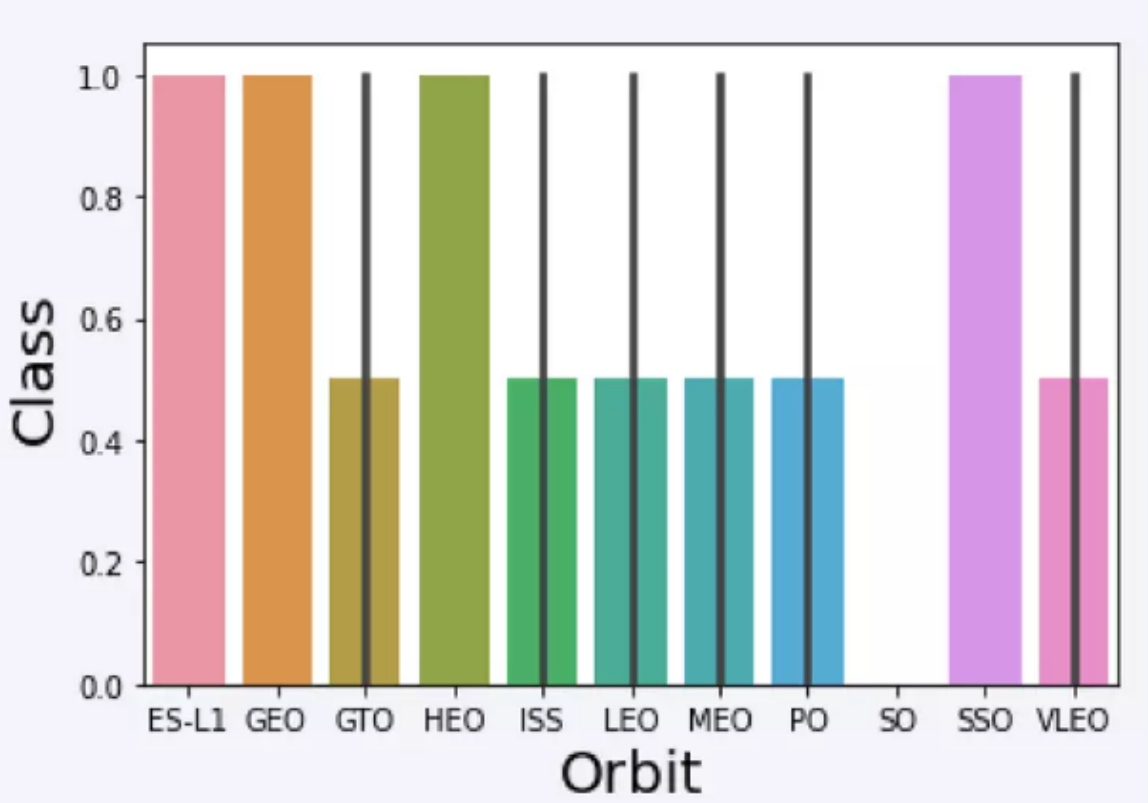
- Launches from the site of CCAFS SLC 40 are significantly higher than any other.

Payload vs. Launch Site



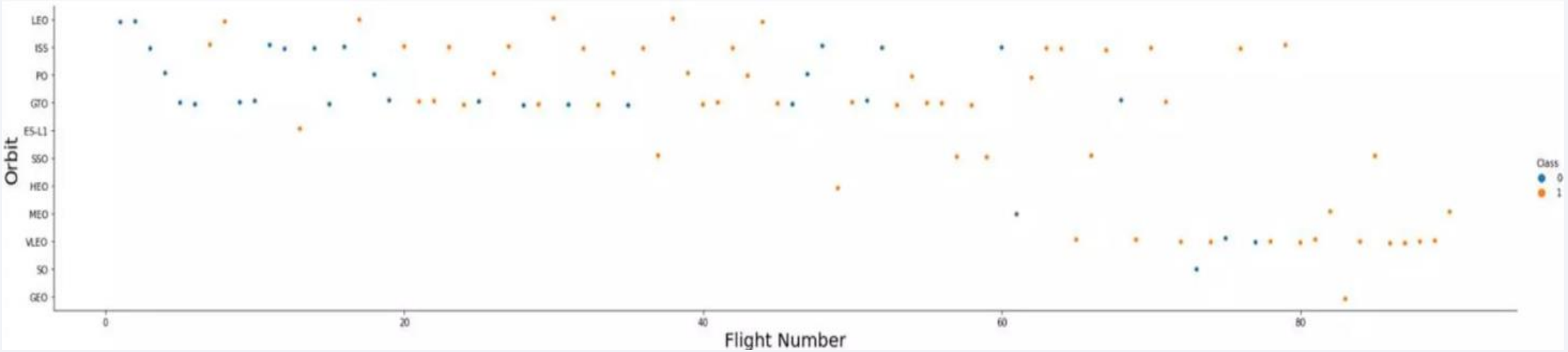
- The majority of Pay Loads with lower Mass have been launched from CCAFS SLC 40.

Success Rate vs. Orbit Type



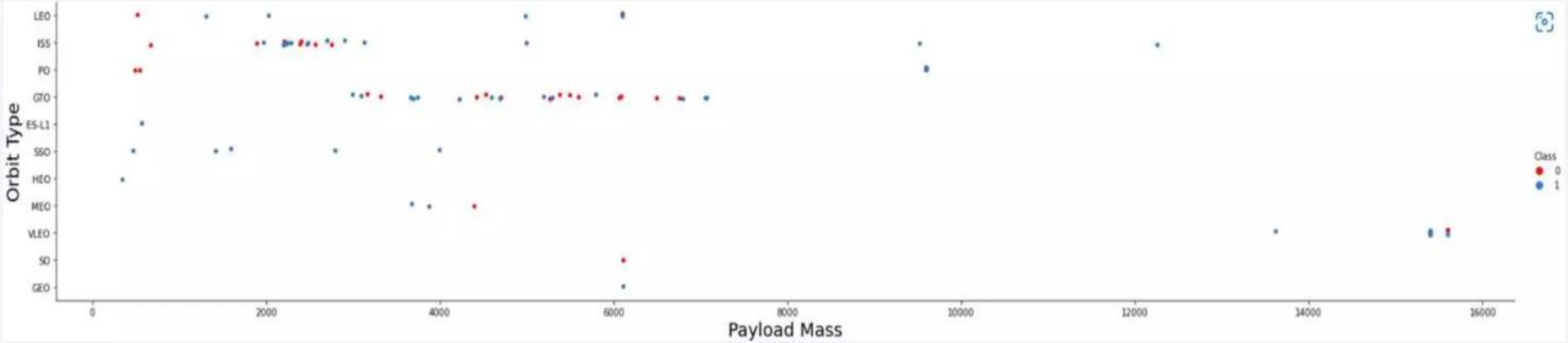
- We can see the orbits with the highest success rate are : SSO, HEO, GEO, ES-L1.
- While the GTO Orbit it is the one with lowest rate

Flight Number vs. Orbit Type



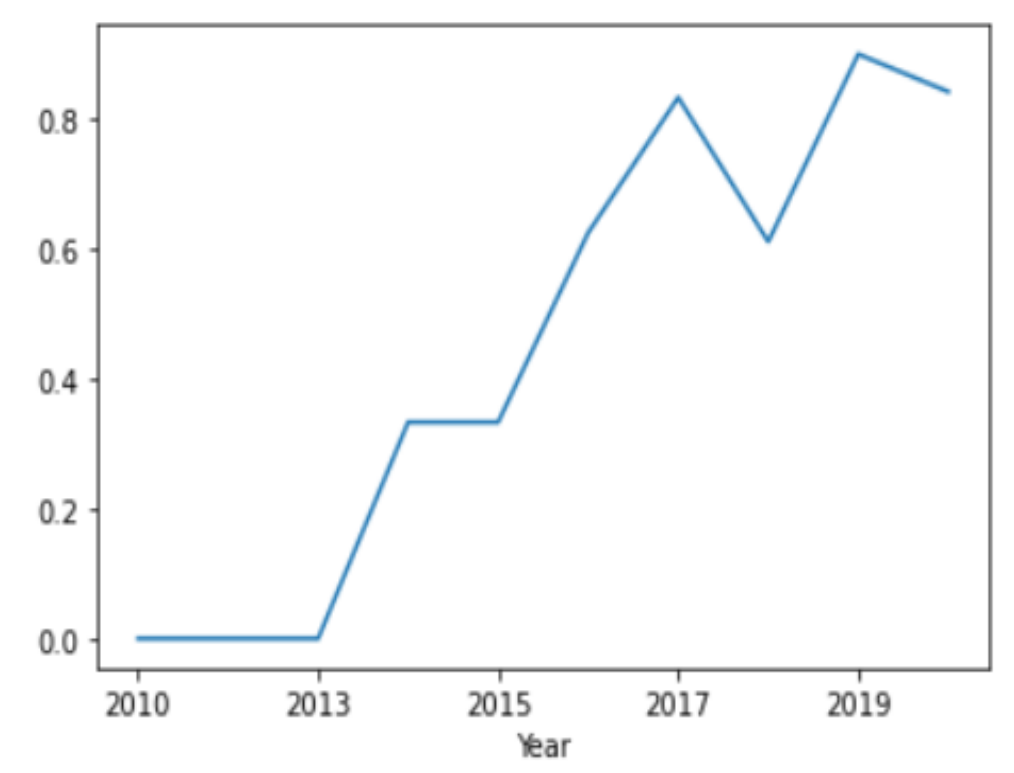
- You should see that in the LEO orbit the Success appears related to the number of flights.
- On the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- You should observe that Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend



- Launch success rate has increased significantly since 2013 and has establised since 2019, potentially due to advance in technology and lessons learned.

All Launch Site Names

- %sql select DISTINCT Launch_Site from SPACEX

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

Launch Site Names Begin with 'CCA'

- %sql select * from SPACEX where Launch_Site like 'CCA%' limit 5

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

- %sql select sum(PAYLOAD_MASS_KG_) as sum from SPACEX where customer like 'NASA (CRS)'

SUM
45596

Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS_KG_) as Average from SPACEX where Booster_Version like 'F9 v1.1%'

average
2534

First Successful Ground Landing Date

- %sql select min(date) as Date from SPACEX where Mission_Outcome like 'Success'

DATE
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql select Booster_Version from SPACEX where (Mission_Outcome like 'Success') AND (PAYLOAD_MASS_KG_ between 4000 AND 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

- %sql SELECT mission_outcome, count(*) as Count FROM SPACEX GROUP by mission_outcome ORDER BY mission_outcome

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

Boosters Carried Maximum Payload

- maxm = %sql select max(payload_mass_kg_) from SPACEX

  maxv = maxm[0][0]

  %sql select booster_version from SPACEX where payload_mass_kg_ =(select max(payload_mass_kg_) from SPACEX)

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

2015 Launch Records

- %sql select * from SPACEX where Landing_Outcome like 'Success%' and (DATE between '2015-01-01' and '2015-12-31') order by date desc

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

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

- %%sql

  select Landing_Outcome, count(*) as count from SPACEX

  where Date >= '2010-06-04' AND Date <= '2017-03-20’

  GROUP by Landing_Outcome ORDER BY count Desc

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

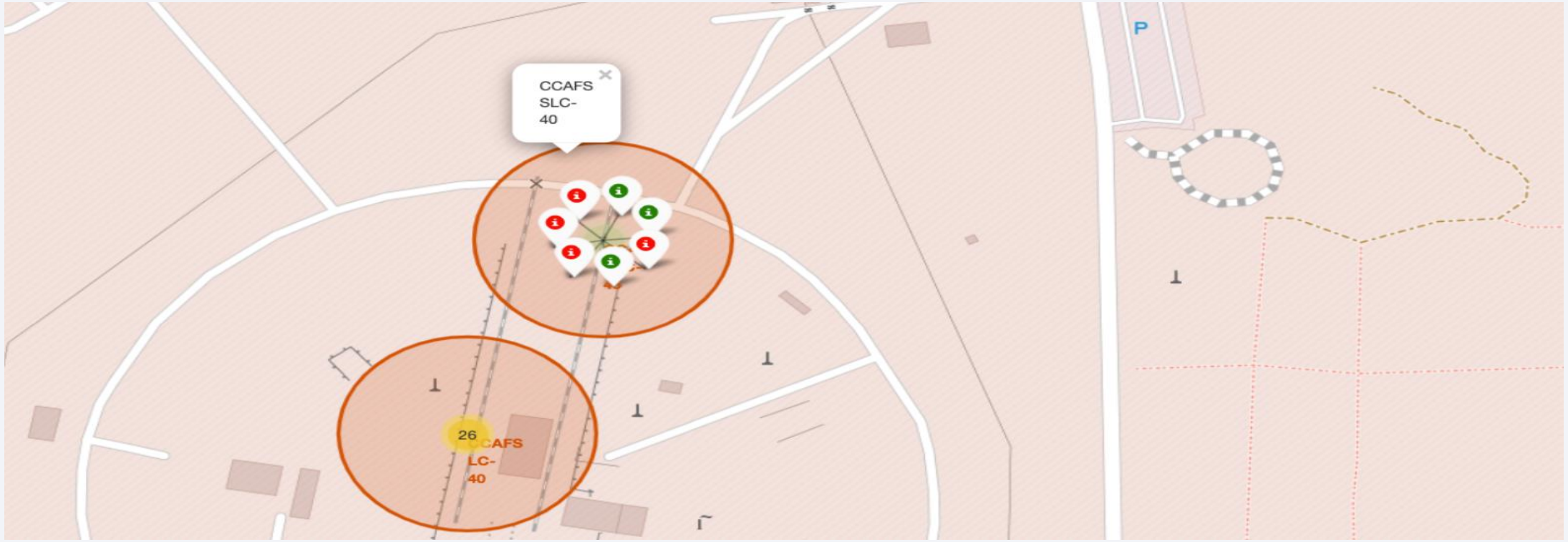
Section 3

Launch Sites Proximities Analysis

All launch sites on a map

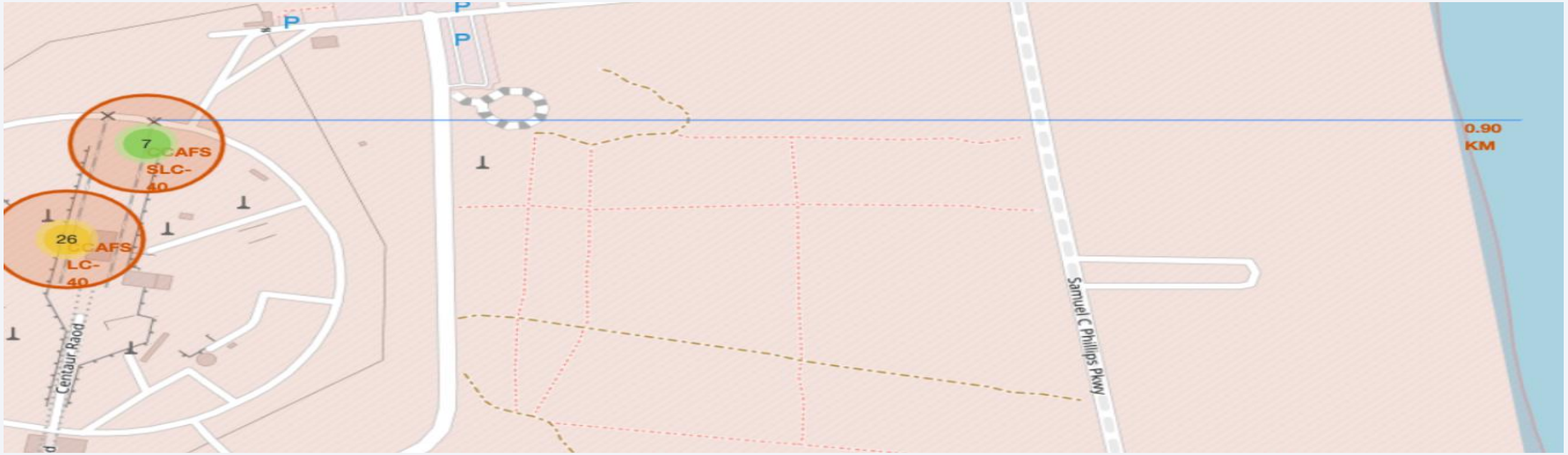
- Both ground and sea surface sites are necessary also south areas maybe a proper areas.
- The transportation base chosen maybe important.

Success/Failed launches for each site on the map

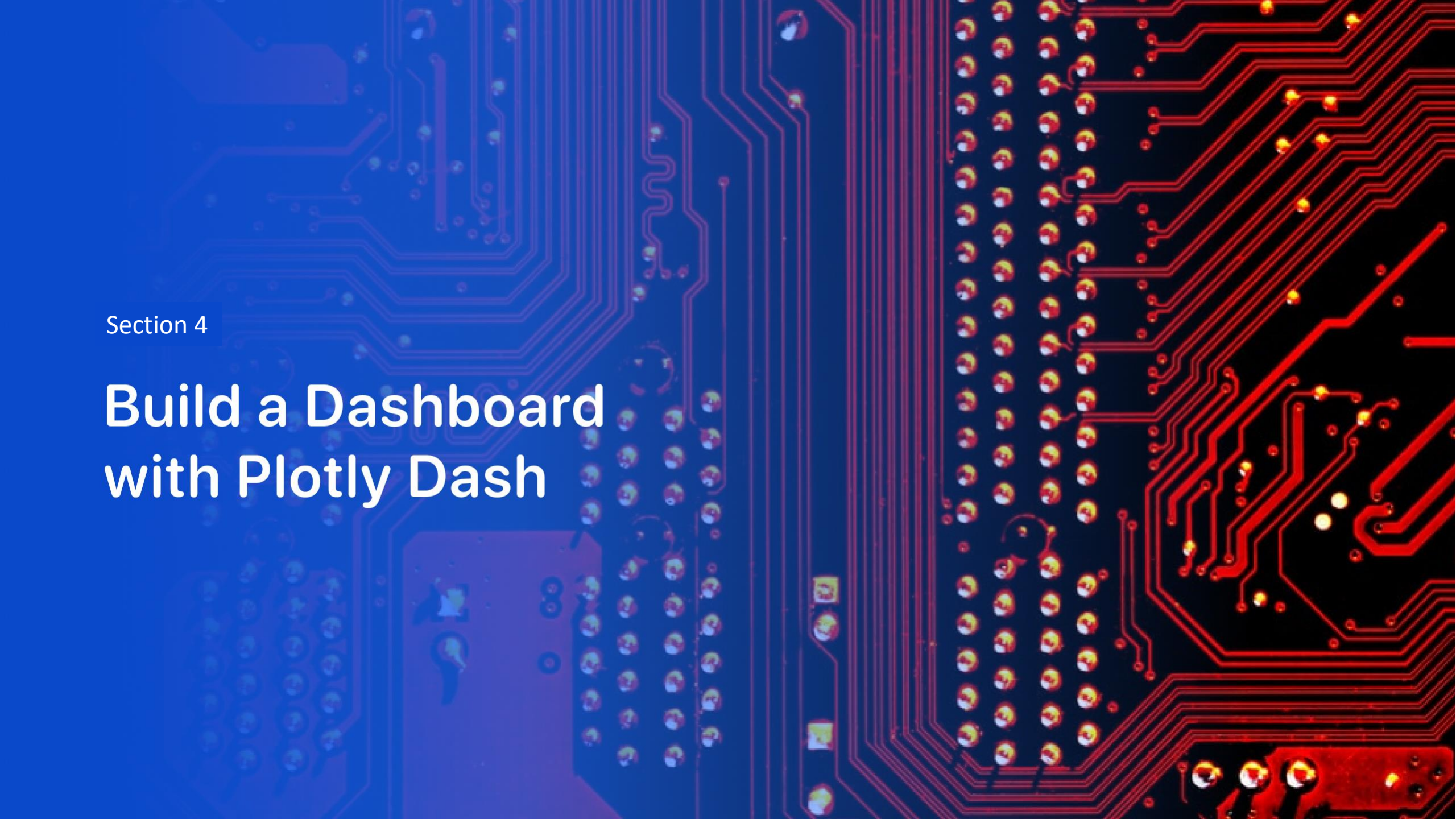


- Color icons are impressive way to show rate of success.
- Here CCAFS SLC-40 shows approx. 43% success rate.

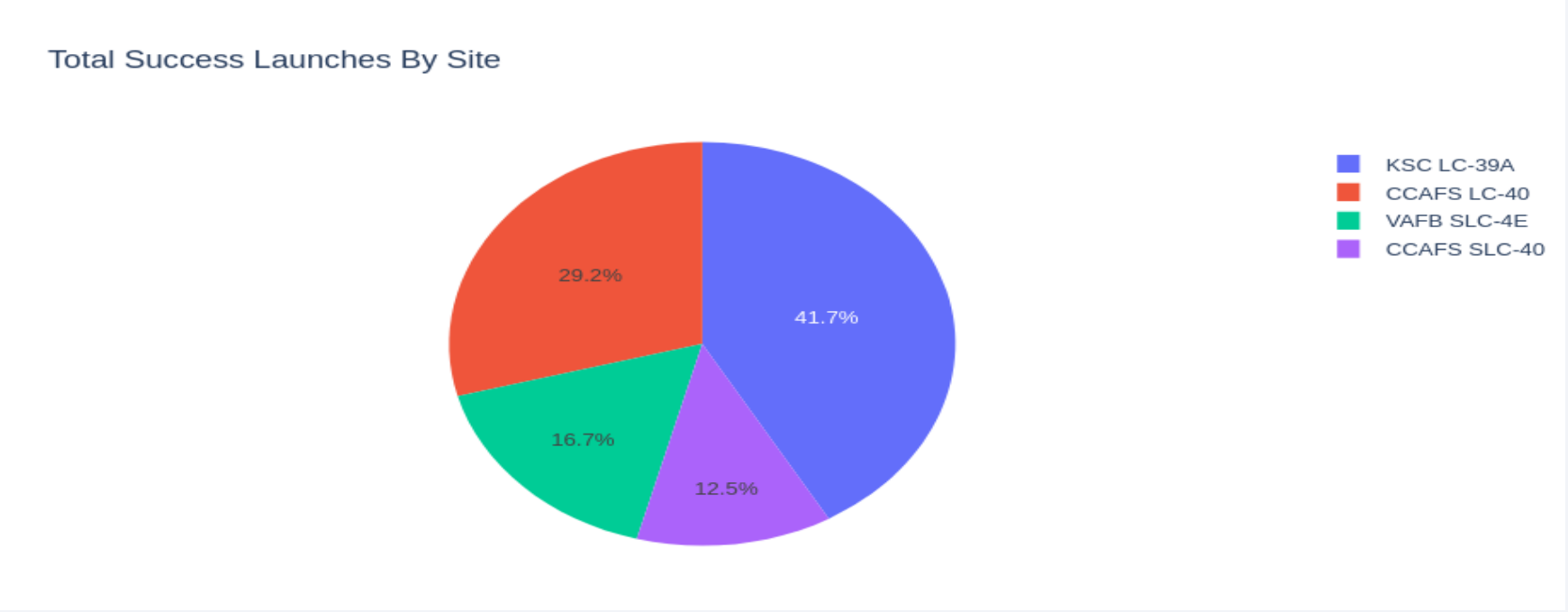
Distance between launch sites



- This map shows the distance from point of interest.
- In this case the distance from railway coast or port may be important.

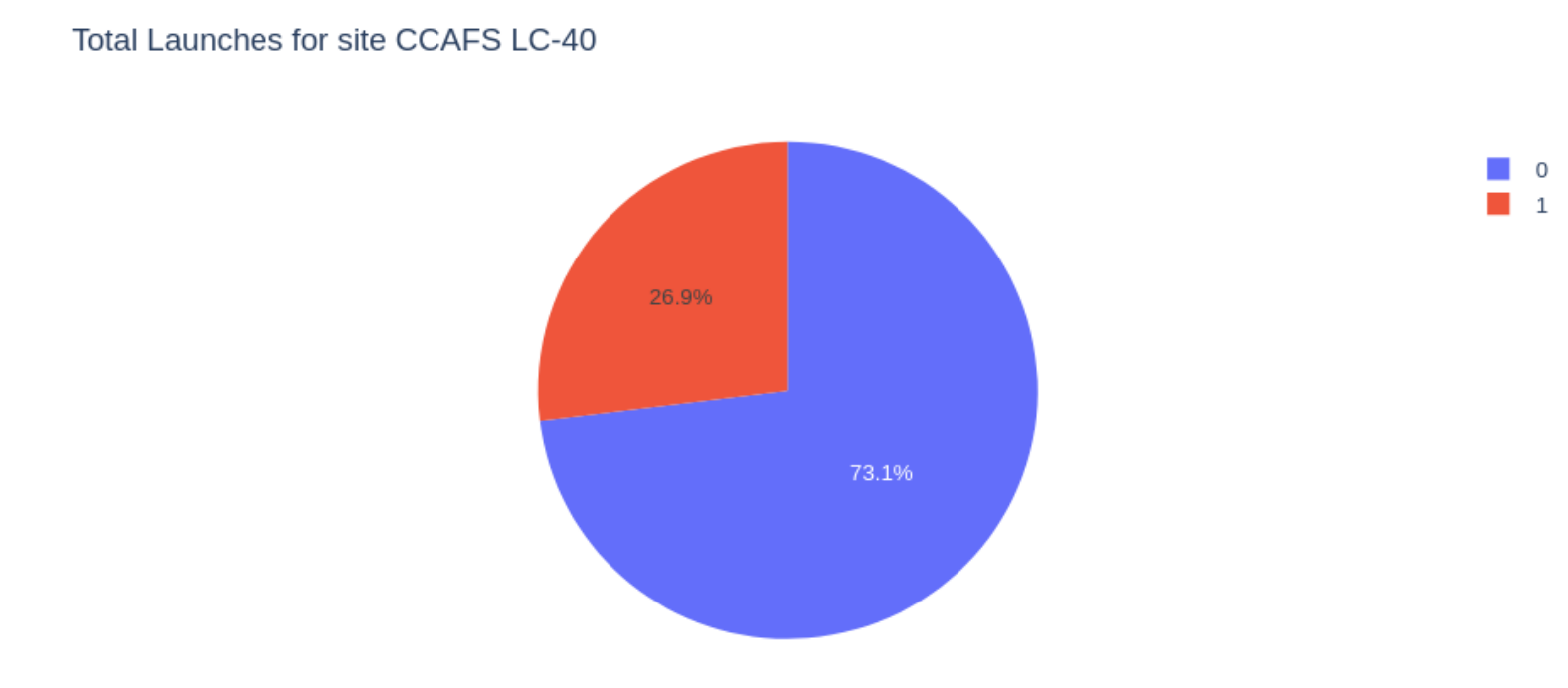

Section 4
Build a Dashboard with Plotly Dash

The most successful launch sites

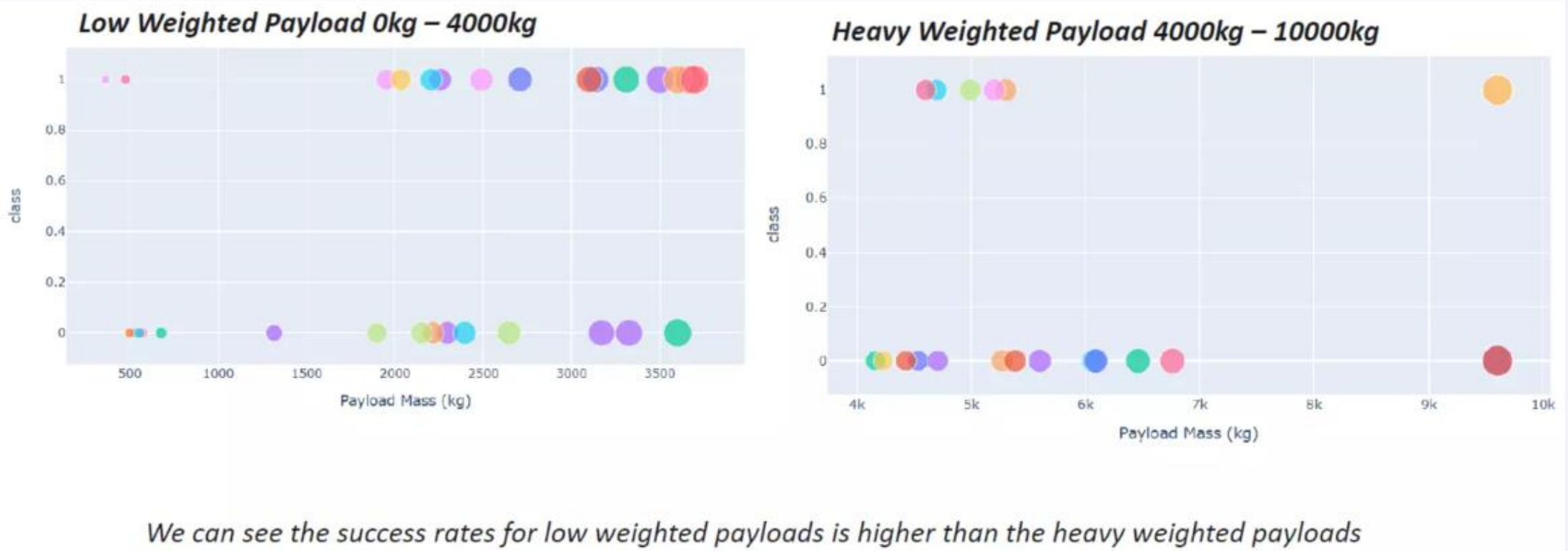


- We can see that CCAFS SLC-40 had the most successful launches from all the sites.

Success rate by site



Payload and Success launch

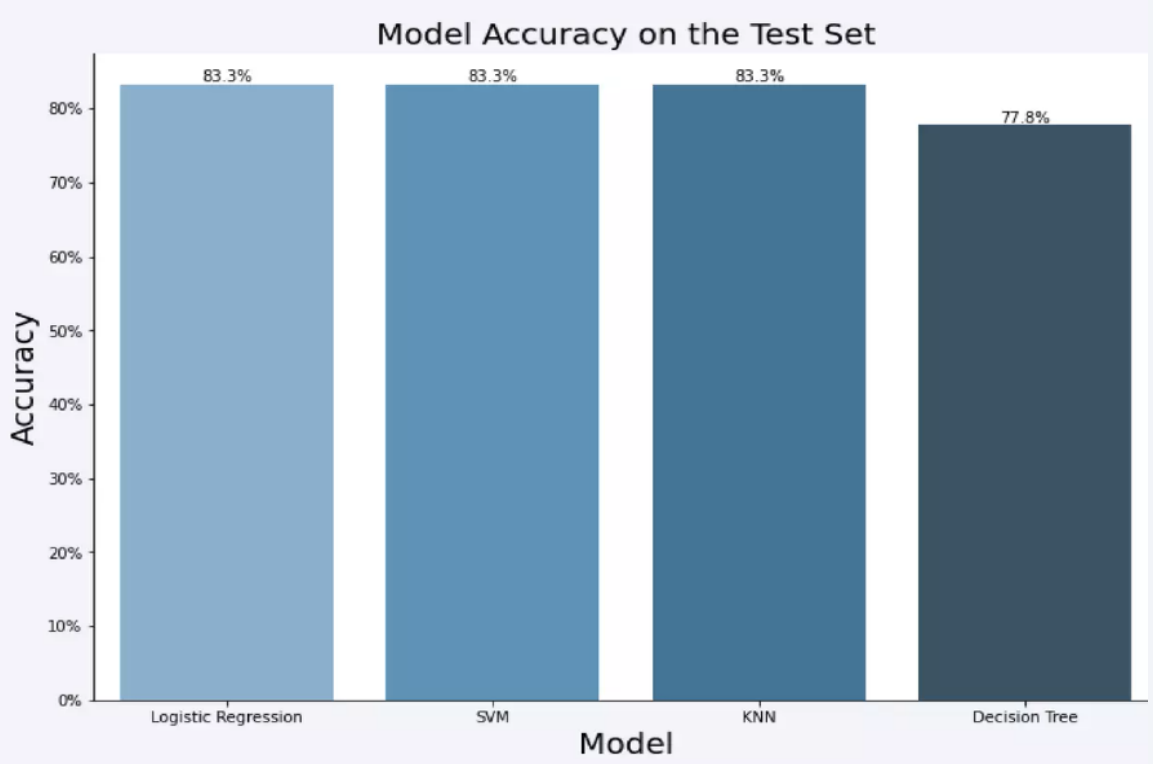


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 5

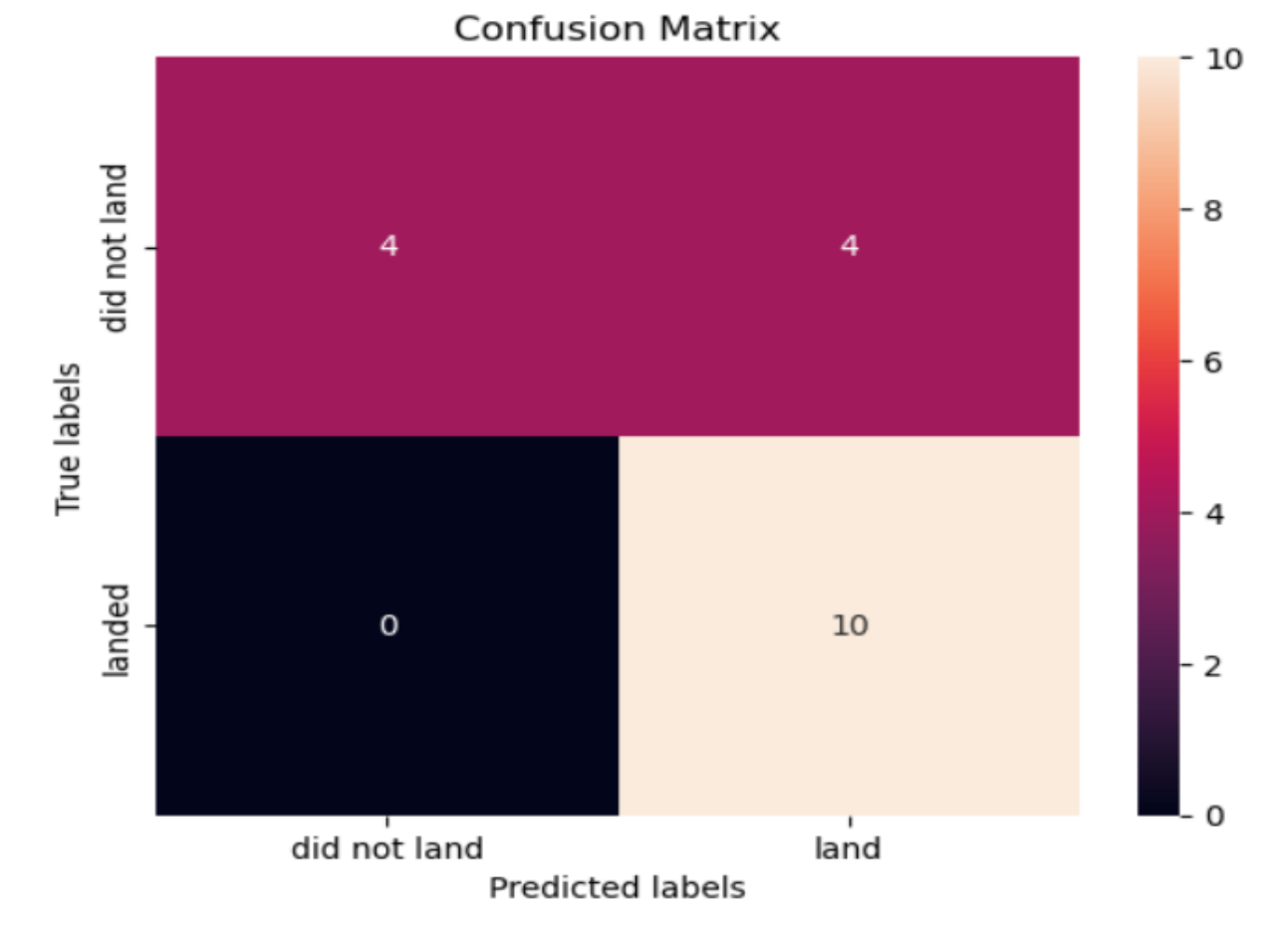
Predictive Analysis (Classification)

Classification Accuracy



- As you can see our accuracy is extremely close.

Confusion Matrix



Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO, ES L1 has the best Success Rate.

Thank you!