IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yassin Abda
30/04/2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data Wrangling
  - EDA with data visualization
  - EDA with SQL
  - Building an interactive map with Folium
  - Building a dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all results
  - EDA results
  - Interactive analytics
  - Predictive analysis

Introduction

- Project background and context
  - SpaceX can reuse the first stage of the Falcon 9 rockets; which in turn, saves a considerable amount of money leading it to be much cheaper than its competitors. With a cost of 62 million dollars per Falcon 9 rocket compared to a cost of upwards to 165 million dollars for its competitors.
- Problems you want to find answers
  - Predicting if the first stage of the Falcon 9 rockets will land successfully or not.

Section 1

Methodology

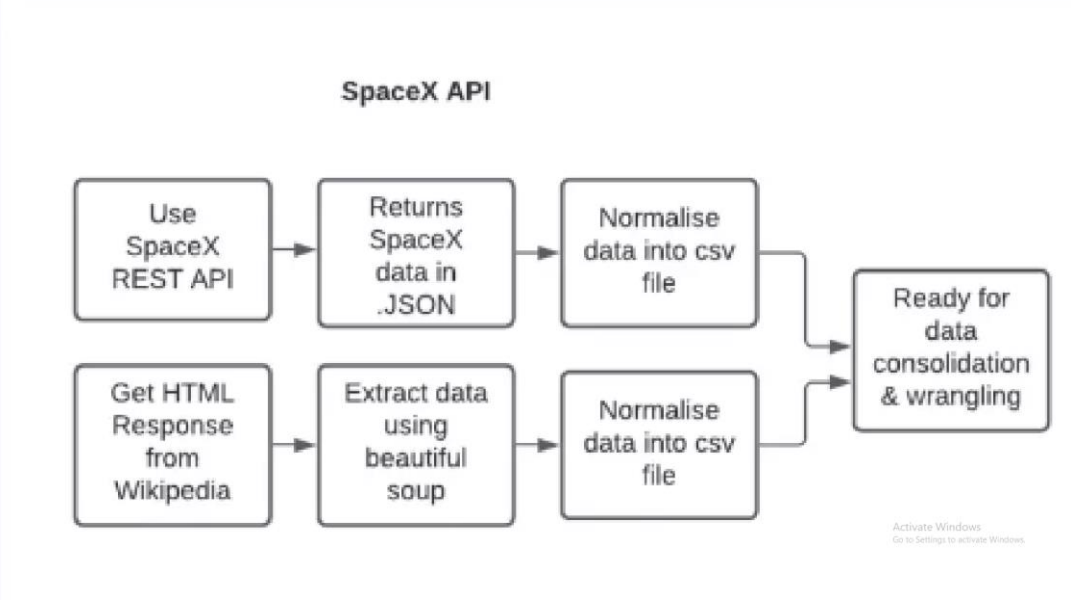
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - One hot Encoding data fields for Machine Learning and data cleaning of null values and irrelevant columns
- 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, DT models have been built and evaluated for the best classifier

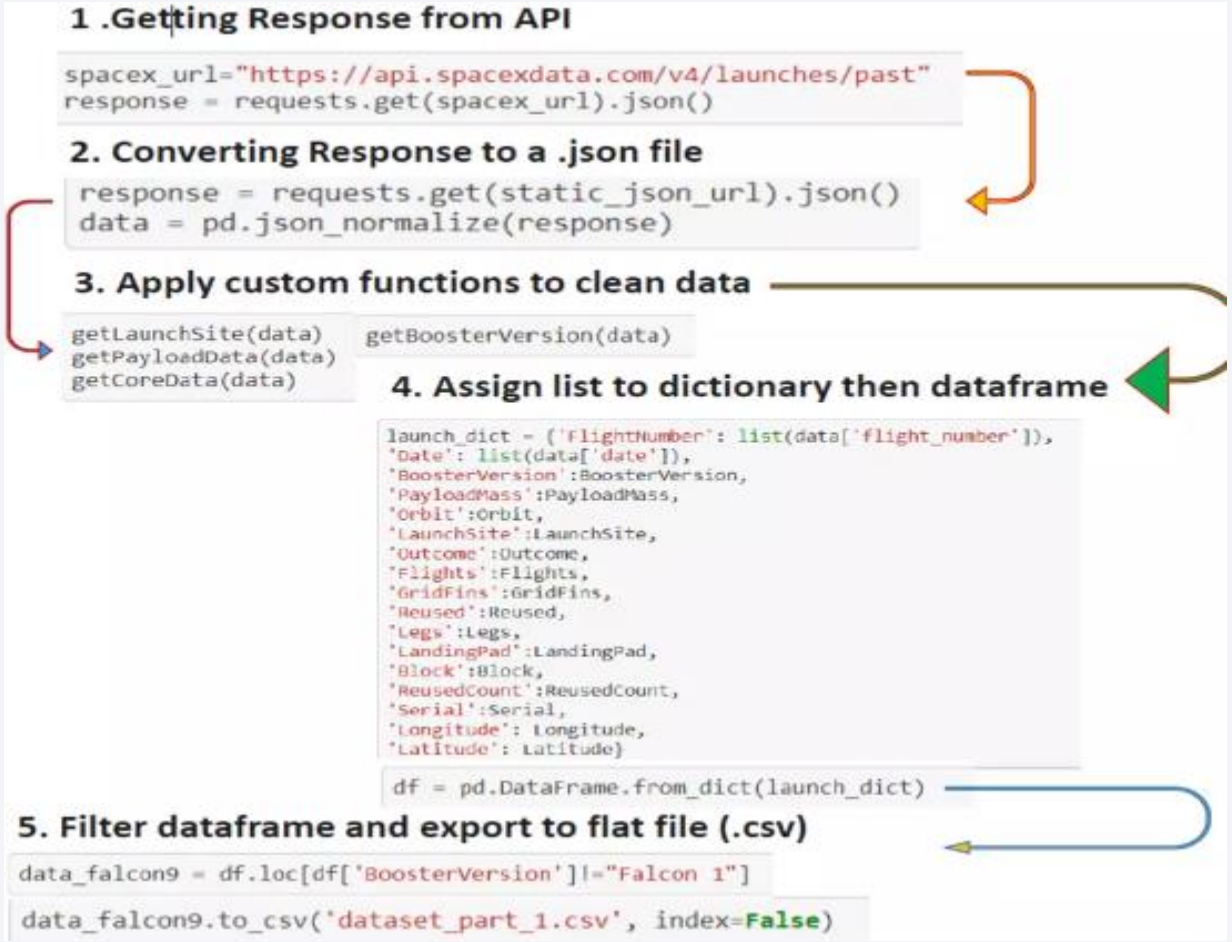
Data Collection

- The following datasets were collected:
  - 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 specification and landing outcome.
  - The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.
  - Another popular data source for obtaining Falcon 9 launch data is web scrapping from Wikipedia using BeautifulSoup.



Data Collection – SpaceX API

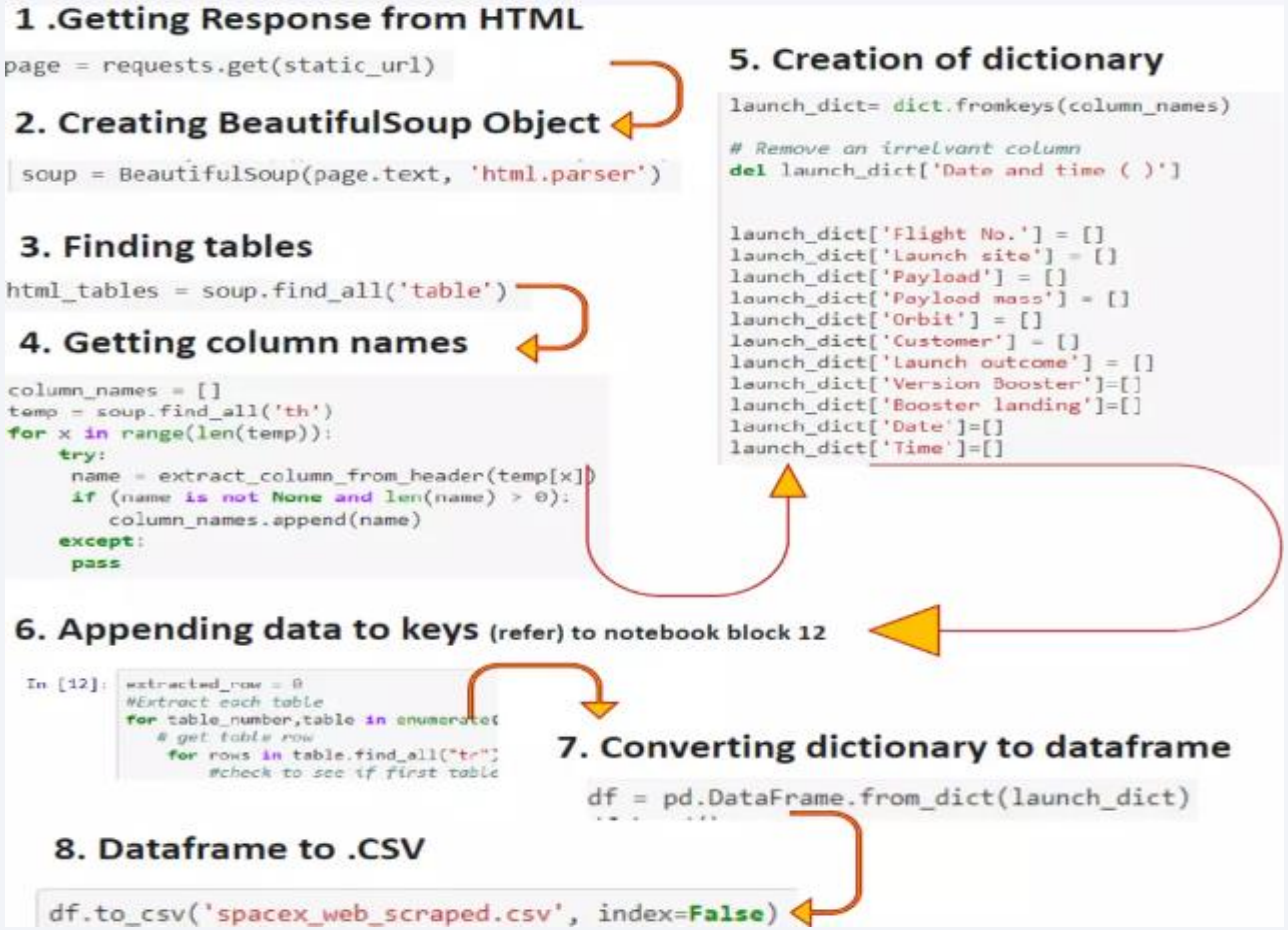
- Data collection with SpaceX REST calls
- https://github.com/yassin-asu/data-science-capstone-/blob/26caaf9ac9ad8d141cc6c34d3e4085601bce6101/data%20wrangling.ipynb



Data Collection - Scraping

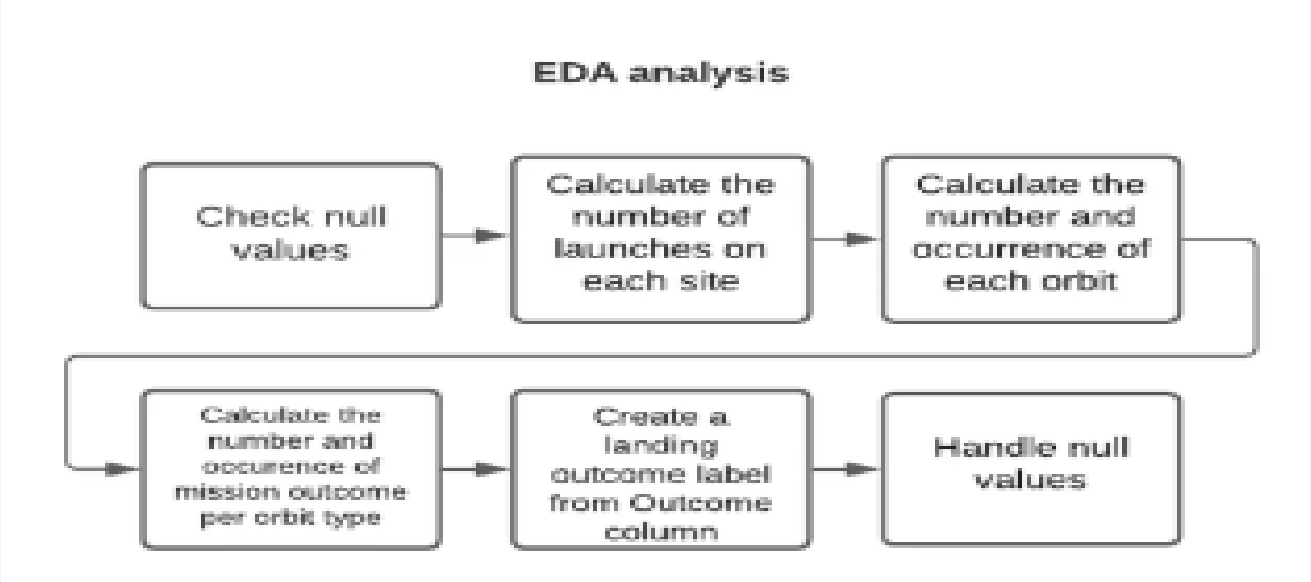
- Web Scrapping from Wikipedia

- https://github.com/yassin-asu/data-science-capstone-/blob/26caaf9ac9ad8d141cc6c34d3e4085601bce6101/data%20collection%20with%20web%20scraping.ipynb



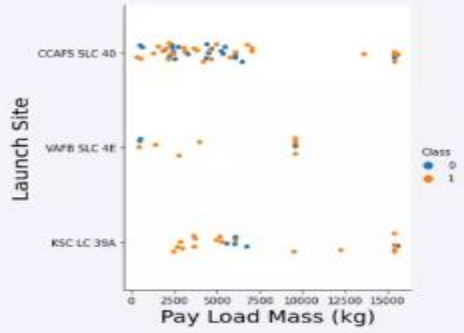
Data Wrangling

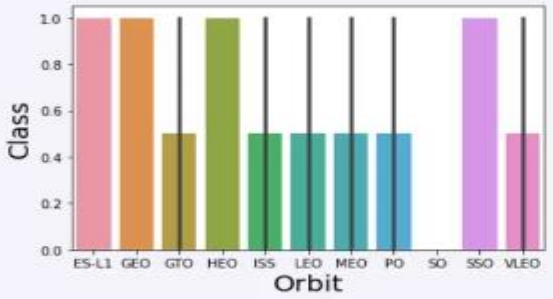
- EDA analysis
- https://github.com/yassin-asu/data-science-capstone-/blob/26caaf9ac9ad8d141cc6c34d3e4085601bce6101/eda.ipynb

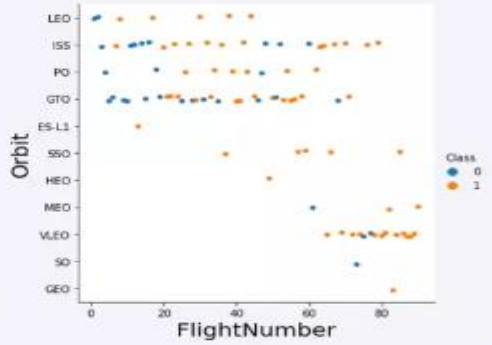


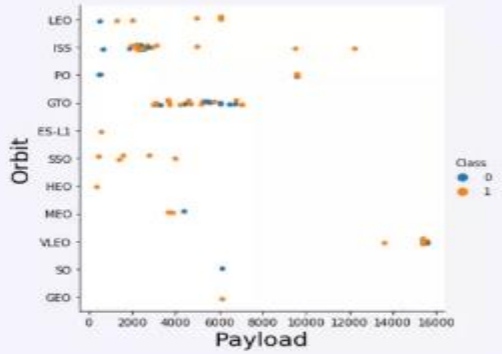
EDA with Data Visualization

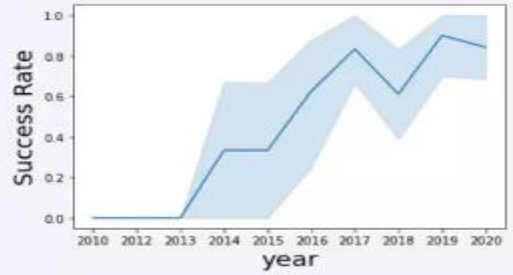
- Visualizing the data to further simplify and understand.
- https://github.com/yassin-asu/data-science-capstone-/blob/26caaf9ac9ad8d141cc6c34d3e4085601bce6101/eda%20dataviz.ipynb











EDA with SQL

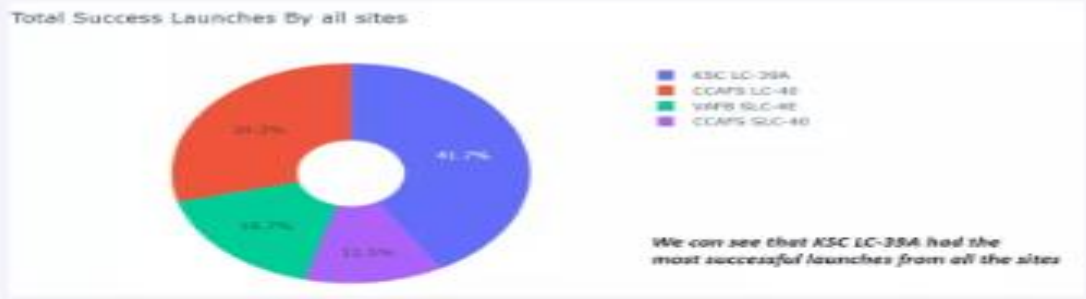
- Performing EDA using SQL:
  - Displaying the names of the unique launch sites in the space mission
  - Displaying 5 records where launch sites begin with the string 'KSC'
  - Displaying the total payload mass carried by boosters launched by NASA (CRS)
  - Displaying average payload mass carried by booster version F9 v1.1
  - Listing the date where the successful landing outcome in drone ship was achieved.
  - Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000
  - • Listing the total number of successful and failure mission outcomes
  - ⬤ Listing the names of the booster_versions which have carried the maximum payload mass.
  - • Listing the records which will display the month names, successful landing_outcomes in ground pad booster versions, launch_site for the months in year 2017
  - • Ranking the count of successful landing_outcomes between the date 2010 06 04 and 2017 03 20 in descendingorder.
- https://github.com/yassin-asu/data-science-capstone-/blob/26caaf9ac9ad8d141cc6c34d3e4085601bce6101/eda%20sql.ipynb

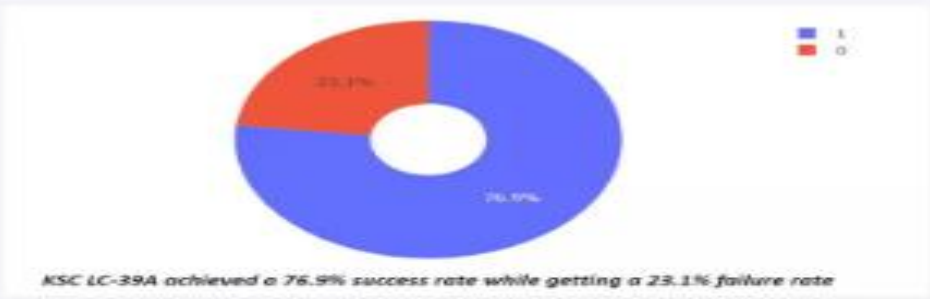
Build an Interactive Map with Folium

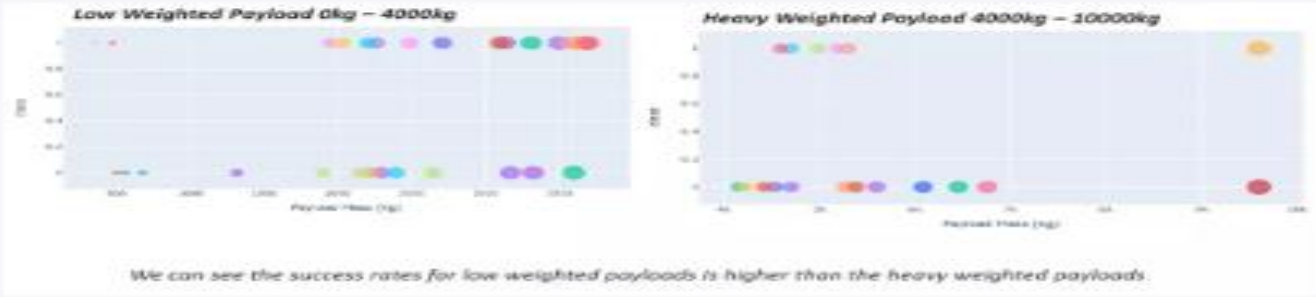
- The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors and hopefully we could discover some of the factors by analyzing the existing launch site locations.
- https://github.com/yassin-asu/data-science-capstone-/blob/26caaf9ac9ad8d141cc6c34d3e4085601bce6101/folium%20launch%20site%20location.ipynb

Build a Dashboard with Plotly Dash

- Plots and graphs using Plotly dash for a more interactive experience
- https://github.com/yassin-asu/data-science-capstone-/blob/26caaf9ac9ad8d141cc6c34d3e4085601bce6101/dashboard.ipynb

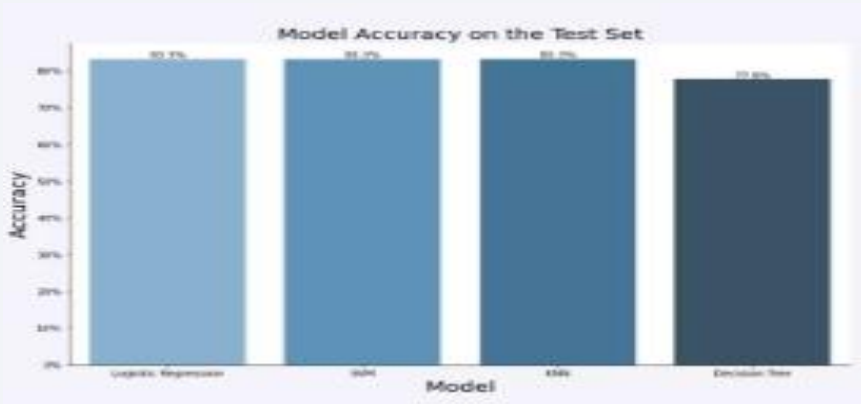


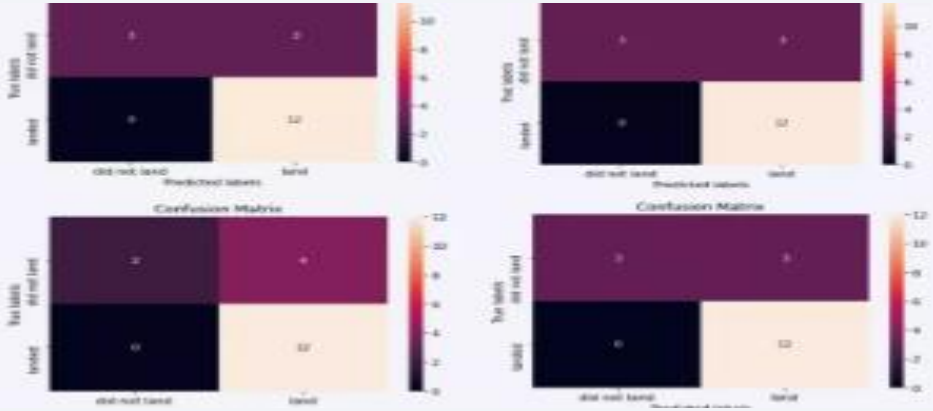




Predictive Analysis (Classification)

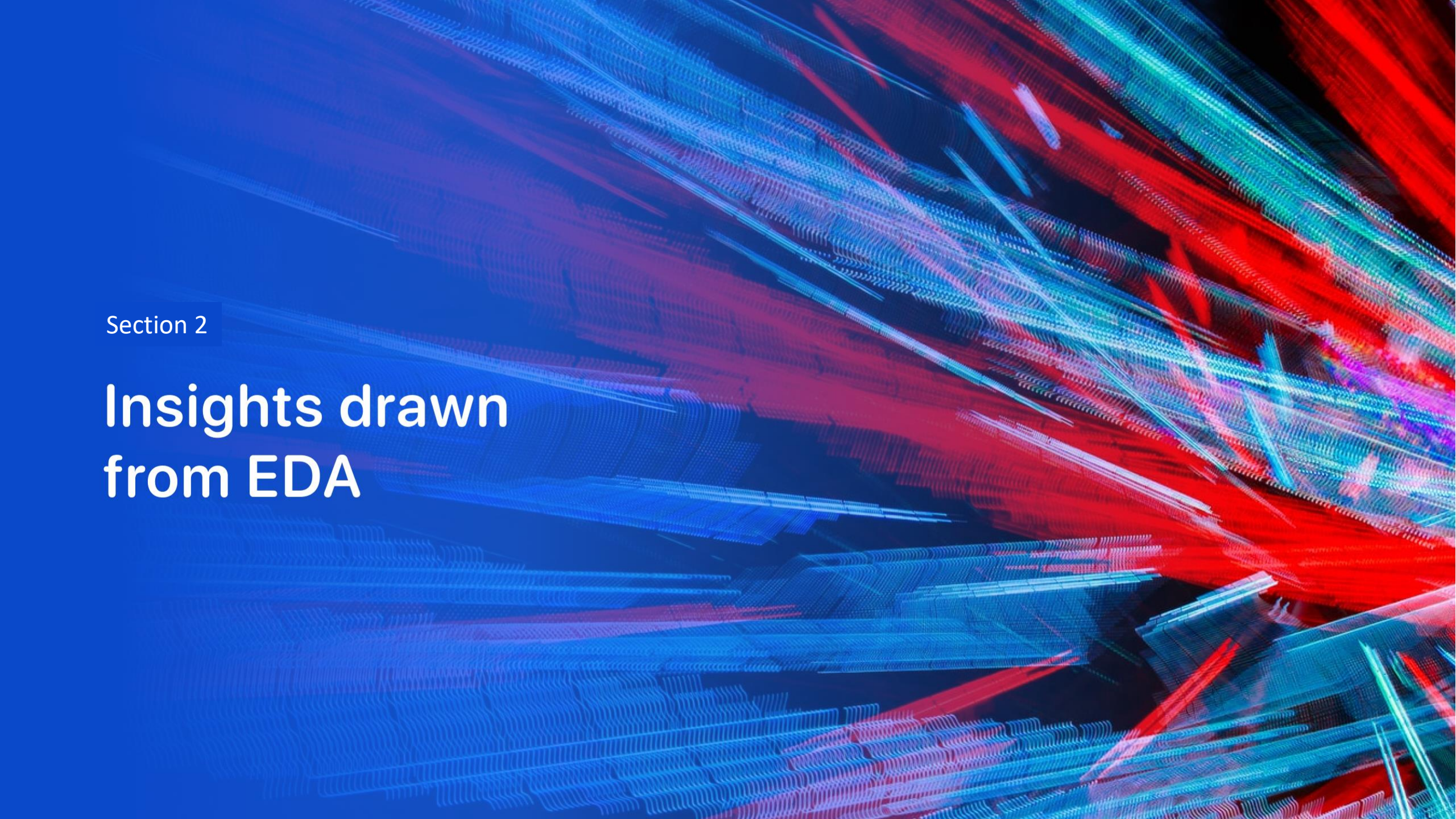
- The SVM, KNN and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.
- https://github.com/yassin-asu/data-science-capstone-/blob/26caaf9ac9ad8d141cc6c34d3e4085601bce6101/predicitive%20pipeline.ipynb



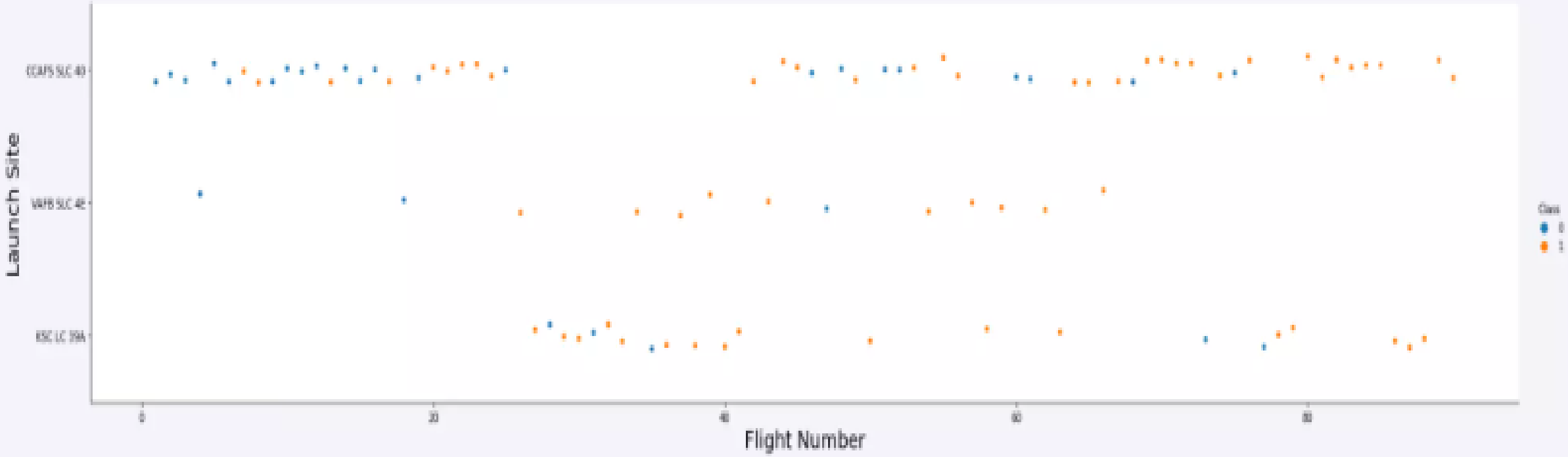


Results

- 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.


Section 2
Insights drawn from EDA

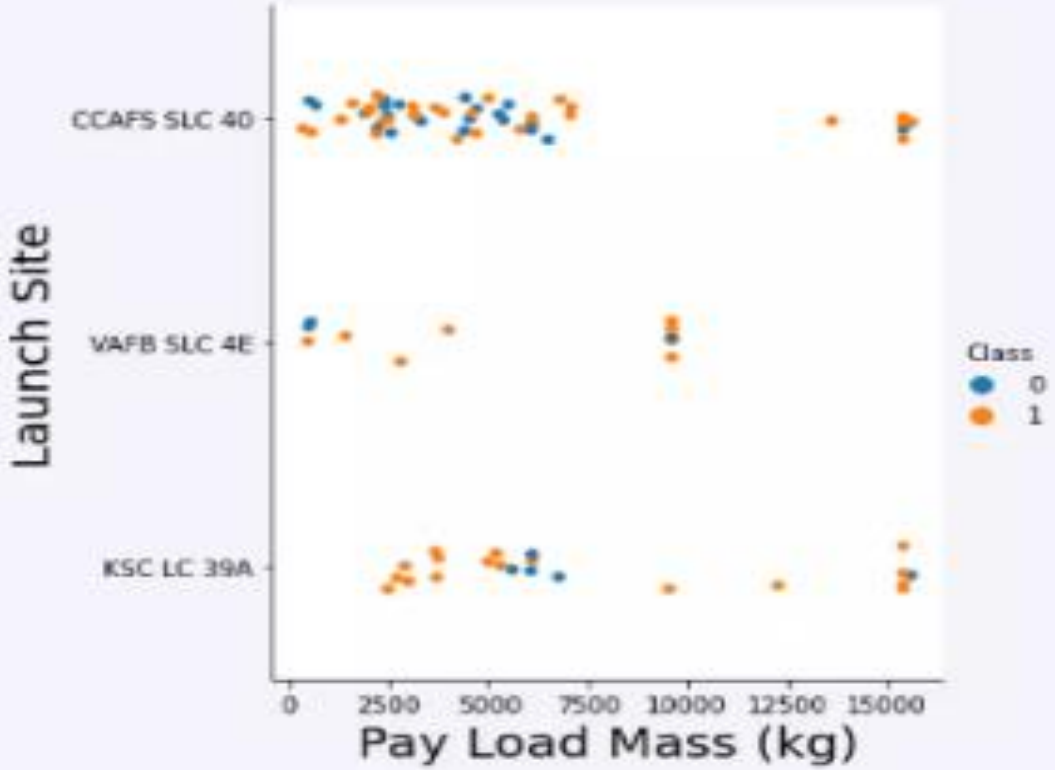
Flight Number vs. Launch Site



- Launches from the site of CCAFS SLC 40 ae significantly higher than launches from other sites

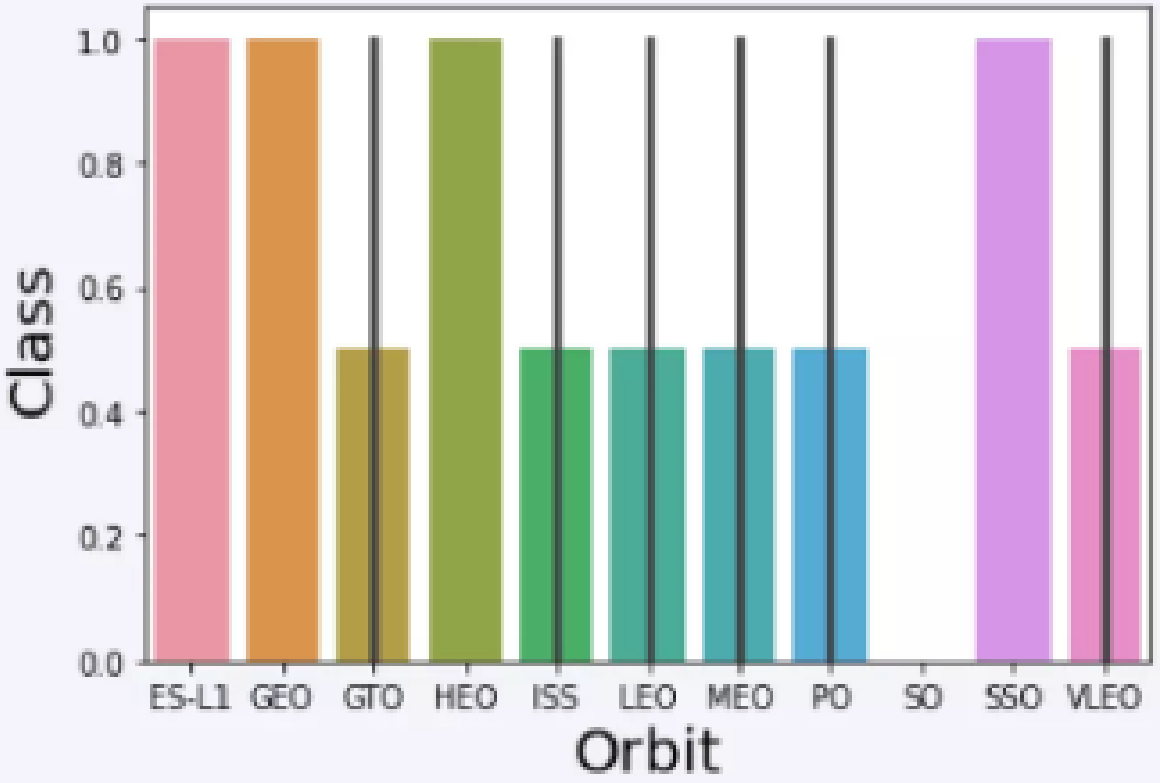
Payload vs. Launch Site

- The majority of IPay loads with lower Mass have been launched from CCAFS SLC 40



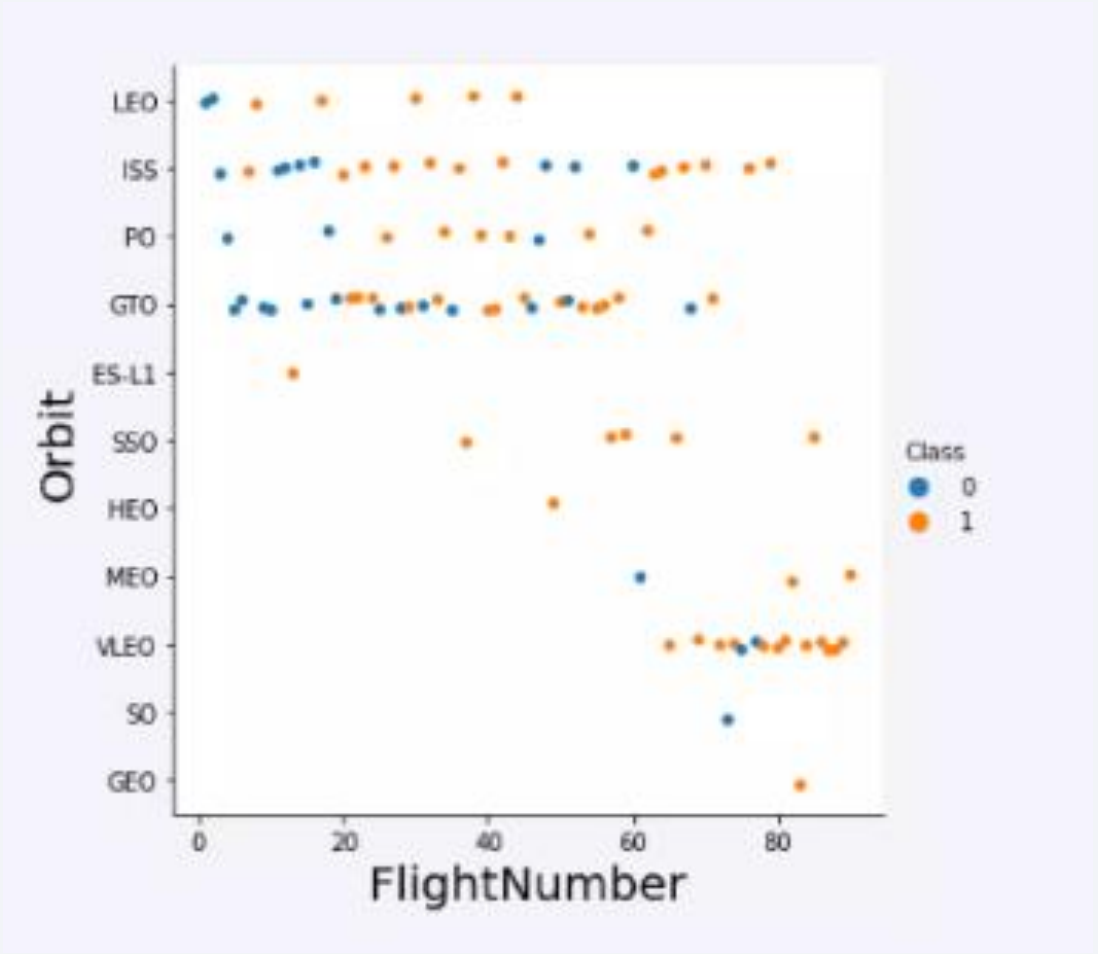
Success Rate vs. Orbit Type

- The orbit of ES-L1,GEO, HEO, SSO are among the highest success rate



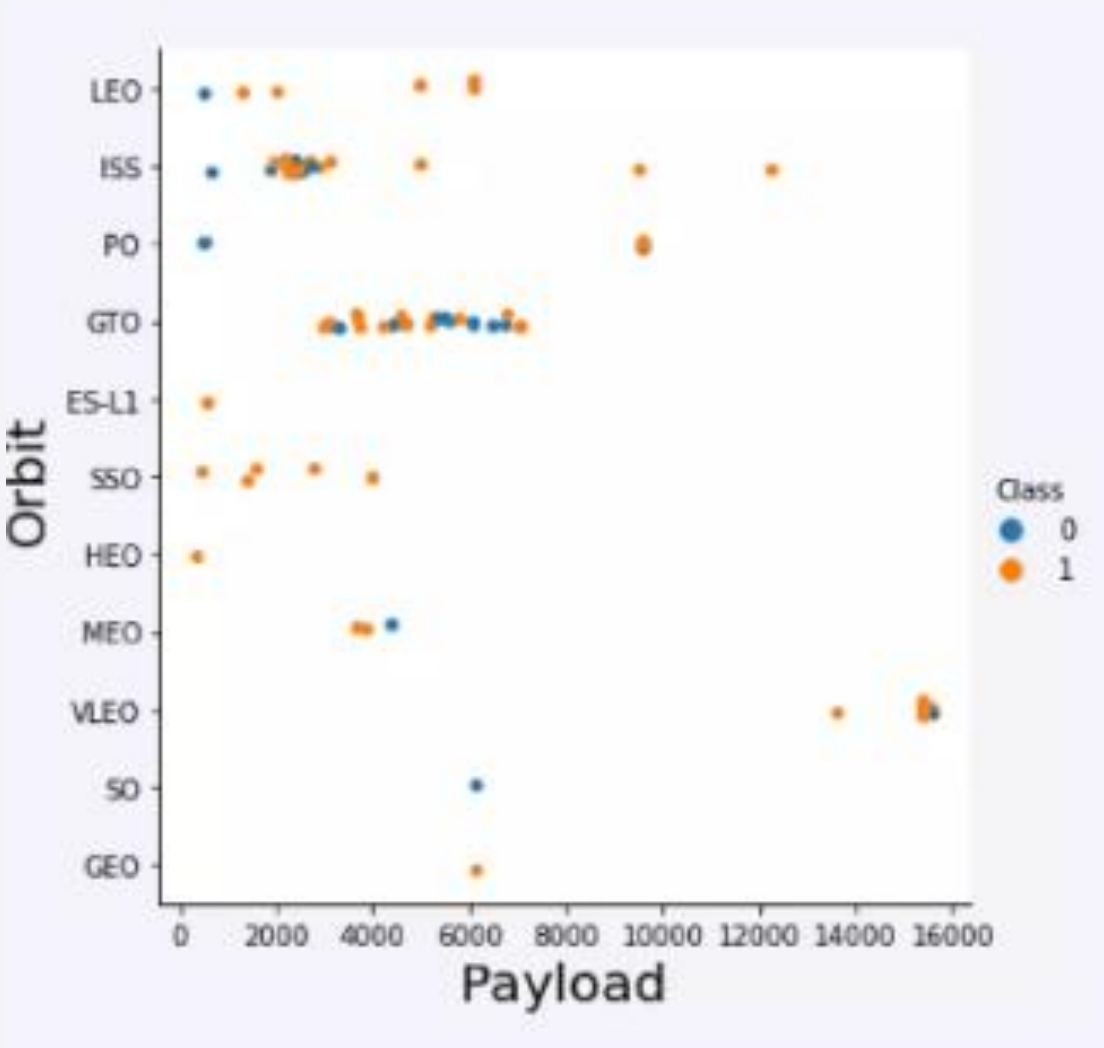
Flight Number vs. Orbit Type

- A trend can be observed of shifting to VLEO launches in recent years.



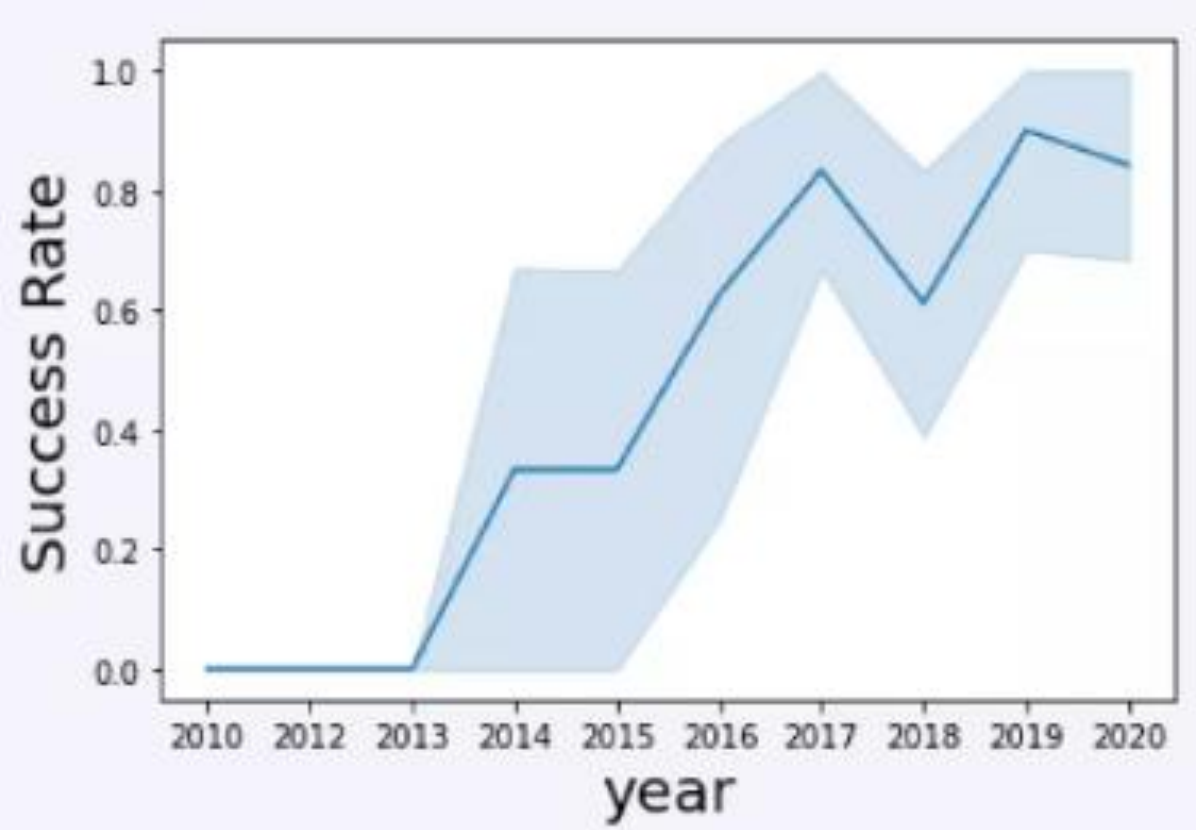
Payload vs. Orbit Type

- There are strong correlation between ISS and payload at the range around 2000, as well as between GTO and the range of 4000-8000.



Launch Success Yearly Trend

- Launch success rate increased significantly since 2013 and has stabilized since 2019, potentially due to advances in tech. and lessons learned.



All Launch Site Names

- %sql select distinct(LAUNCH_SITE) from SPACEXTBL

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

Launch Site Names Begin with 'CCA'

- %sql select * from SPACEXTBL 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	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

- %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
- Result is **45596**

Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
- Result is **2928.4**

First Successful Ground Landing Date

- %sql select min(DATE) from SPACEXTBL where LANDING _OUTCOME = 'Success (ground pad)'
- Result is 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql select BOOSTER_VERSION from SPACEXTBL where Landing __OUTCOME ='Success (drone ship)' and PAYLOAD_MASS___KG_> 4000 and PAYLOAD_MASS__KG_< 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 count9MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'
- Result is 100

Boosters Carried Maximum Payload

- %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ =(select mas(PAYLOAD_MASS__KG_) from SPACEXTBL)

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 SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2015-01-01' and '2015-12-31') order by date desc

time_utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing_outcome
14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
		CCAFS LC-				SKY Perfect JSAT		

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

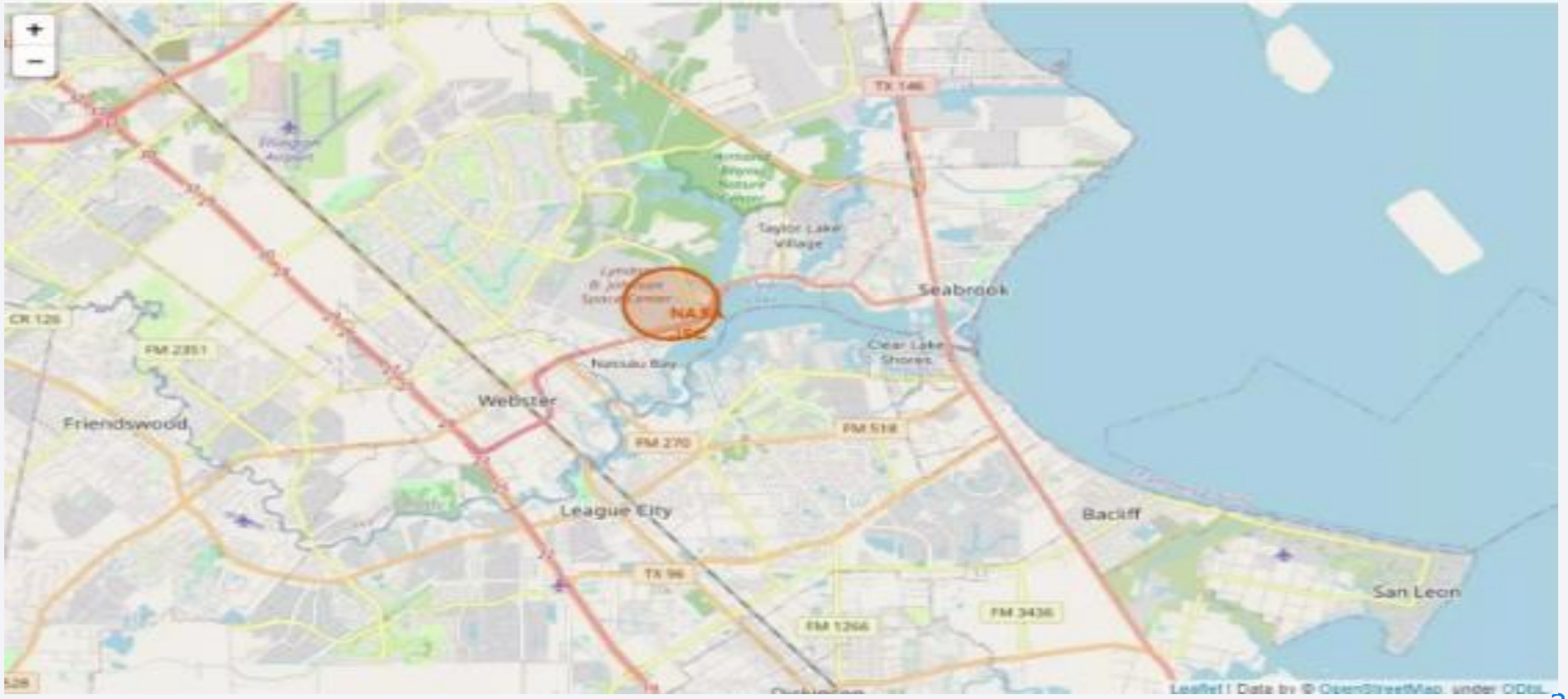
- %sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Section 3

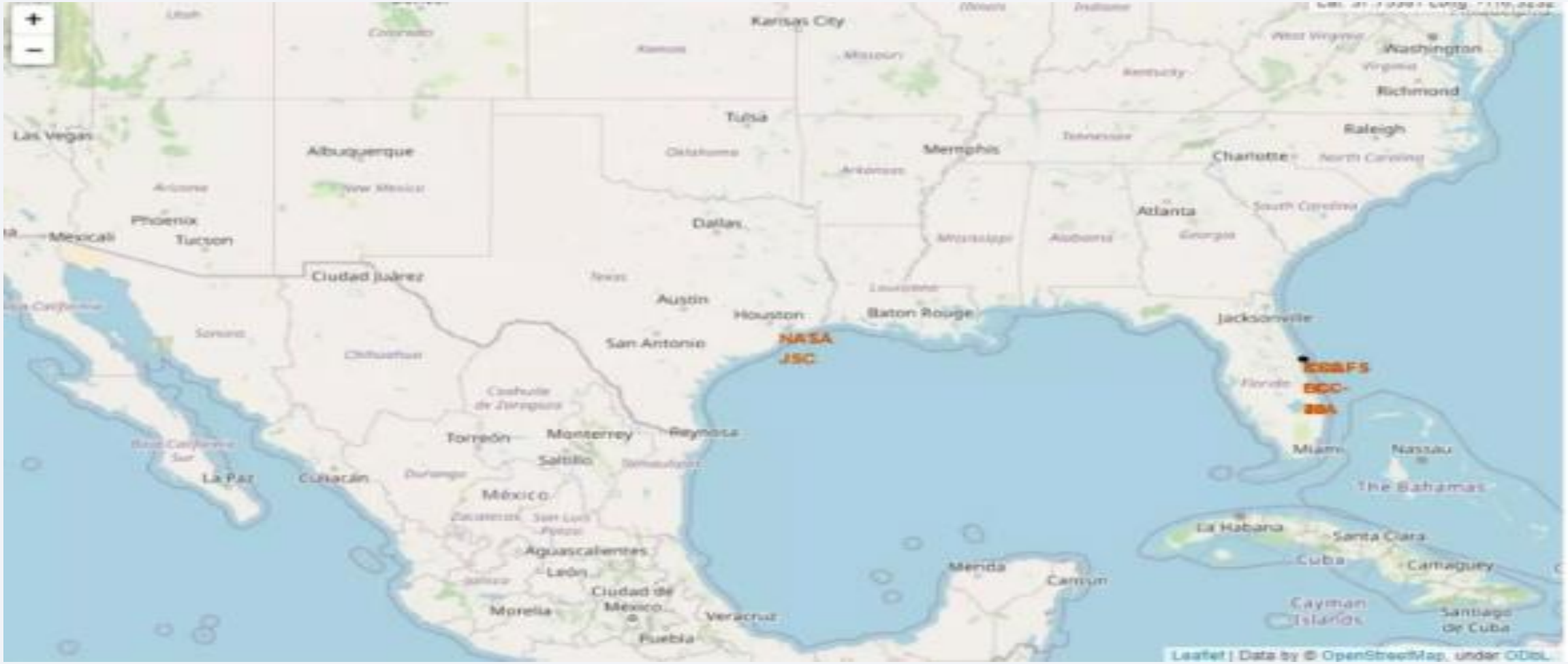
Launch Sites Proximities Analysis

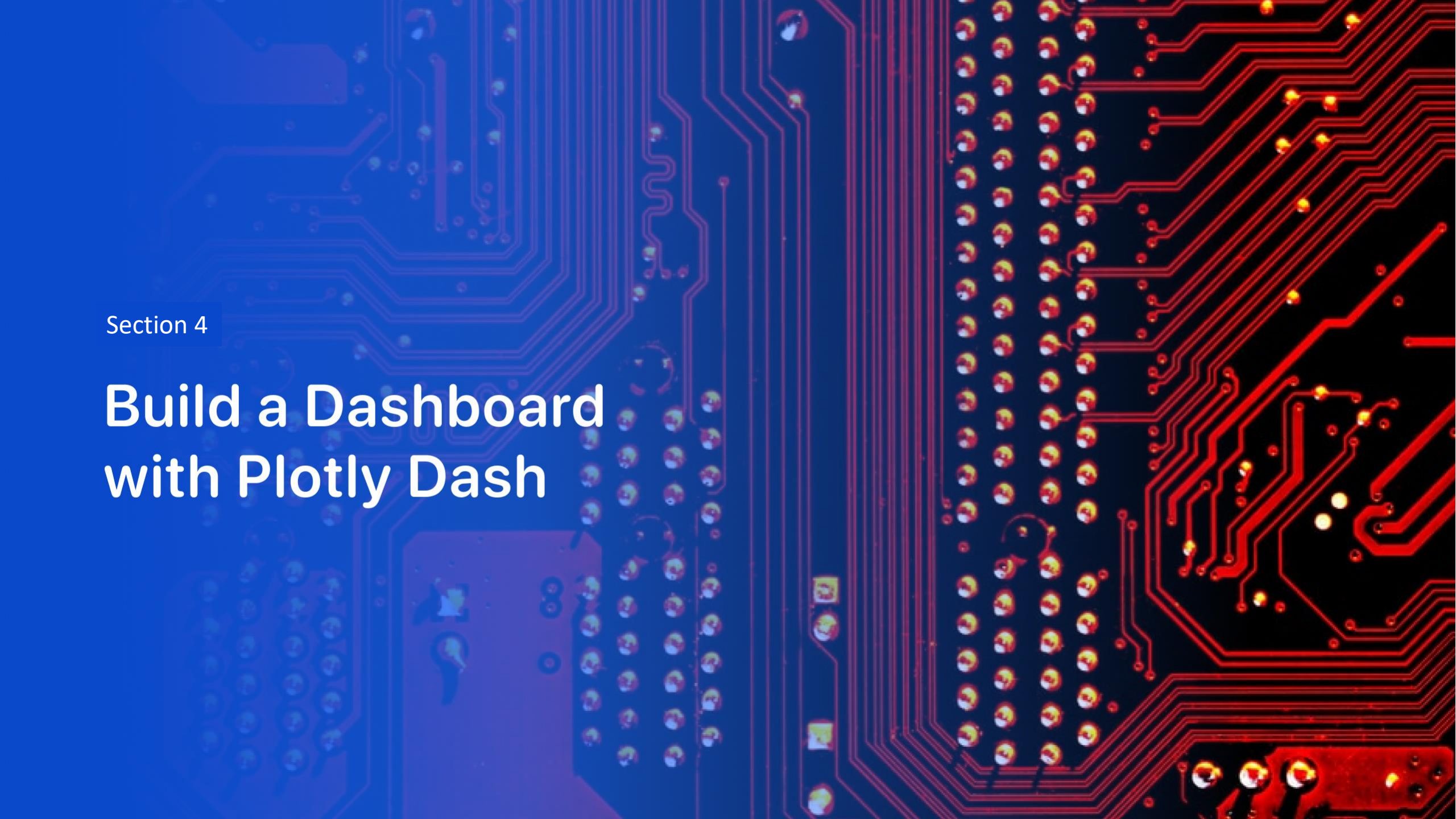
ALL launch sites marked on a map



Success/failed launches marked on map

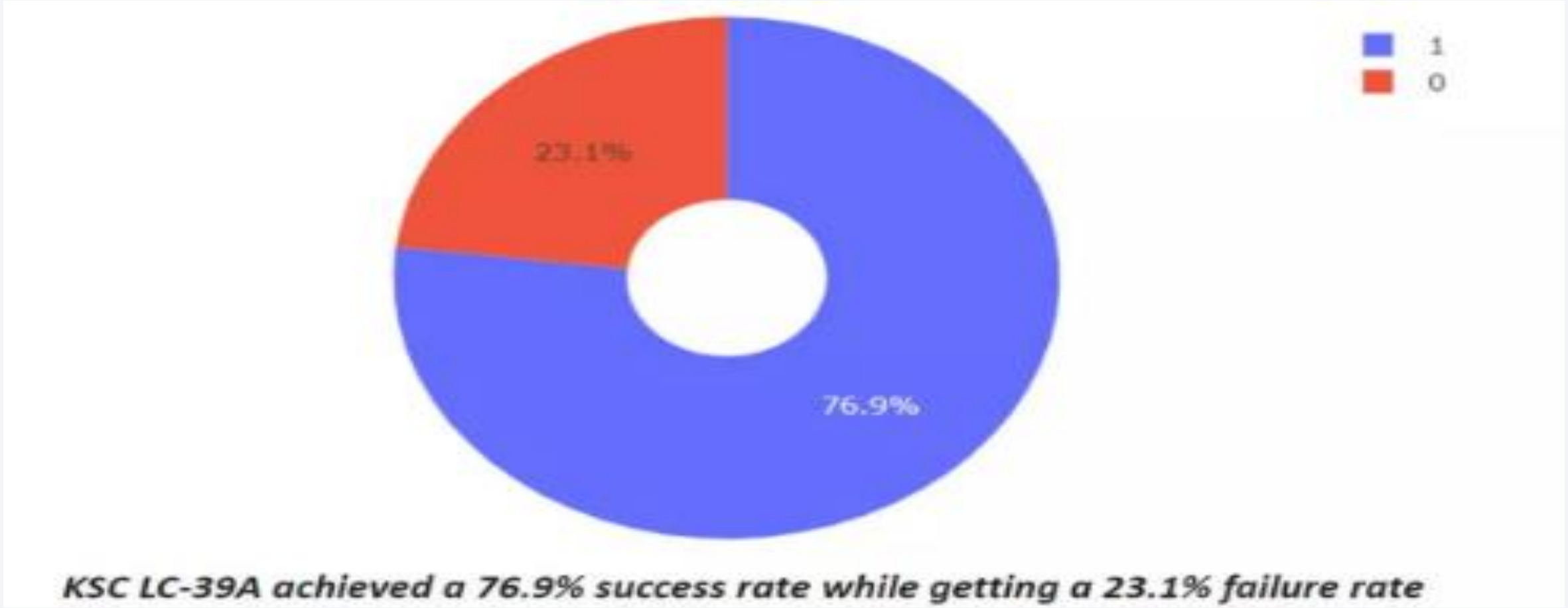
Distances between launch site to its proximities




Section 4
Build a Dashboard with Plotly Dash

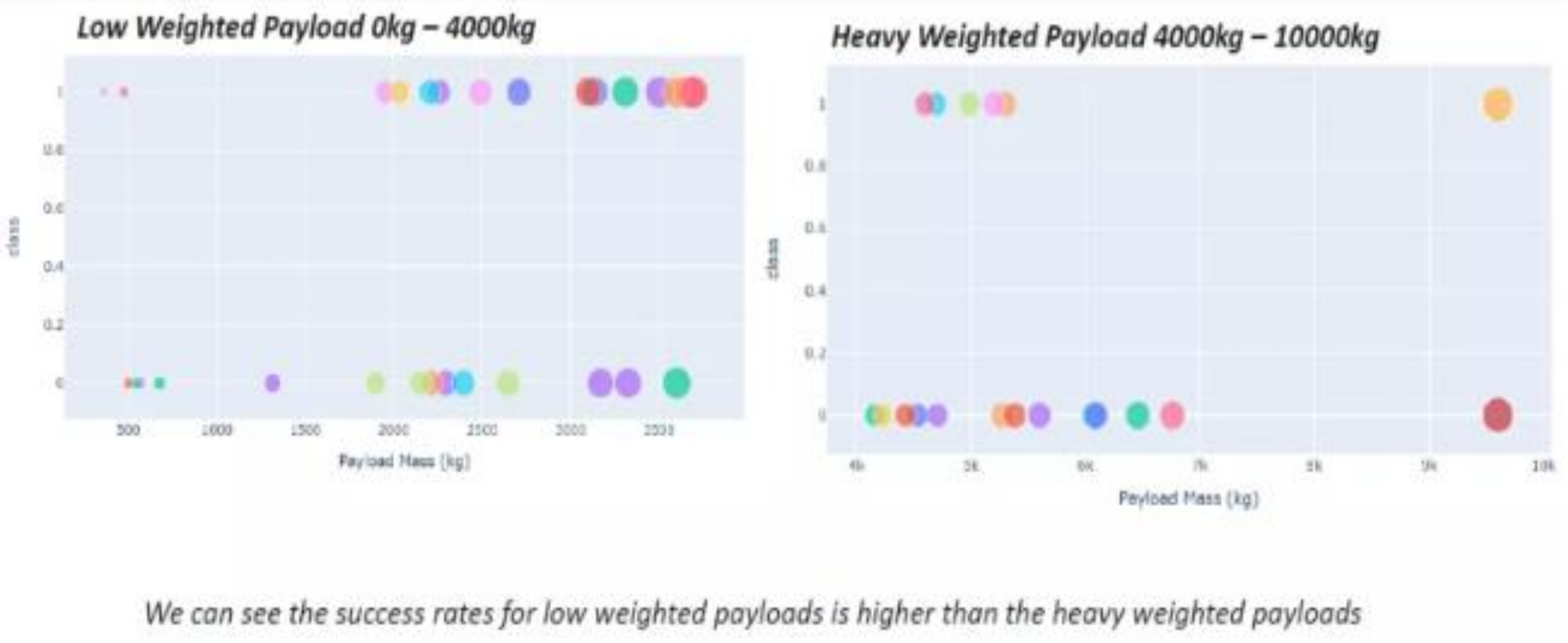
Total success launches by sites

Success rate by site



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs launch outcome

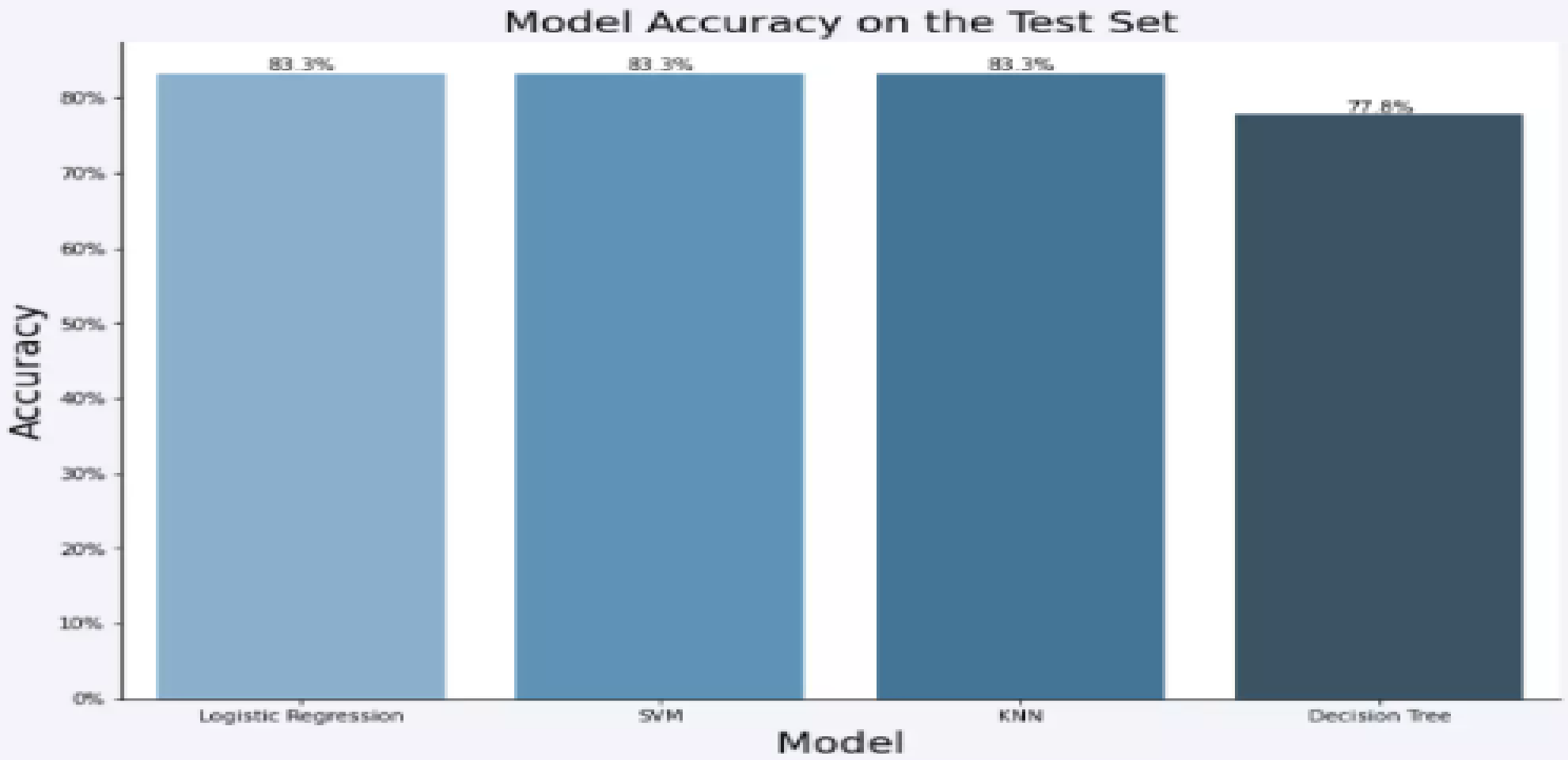


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

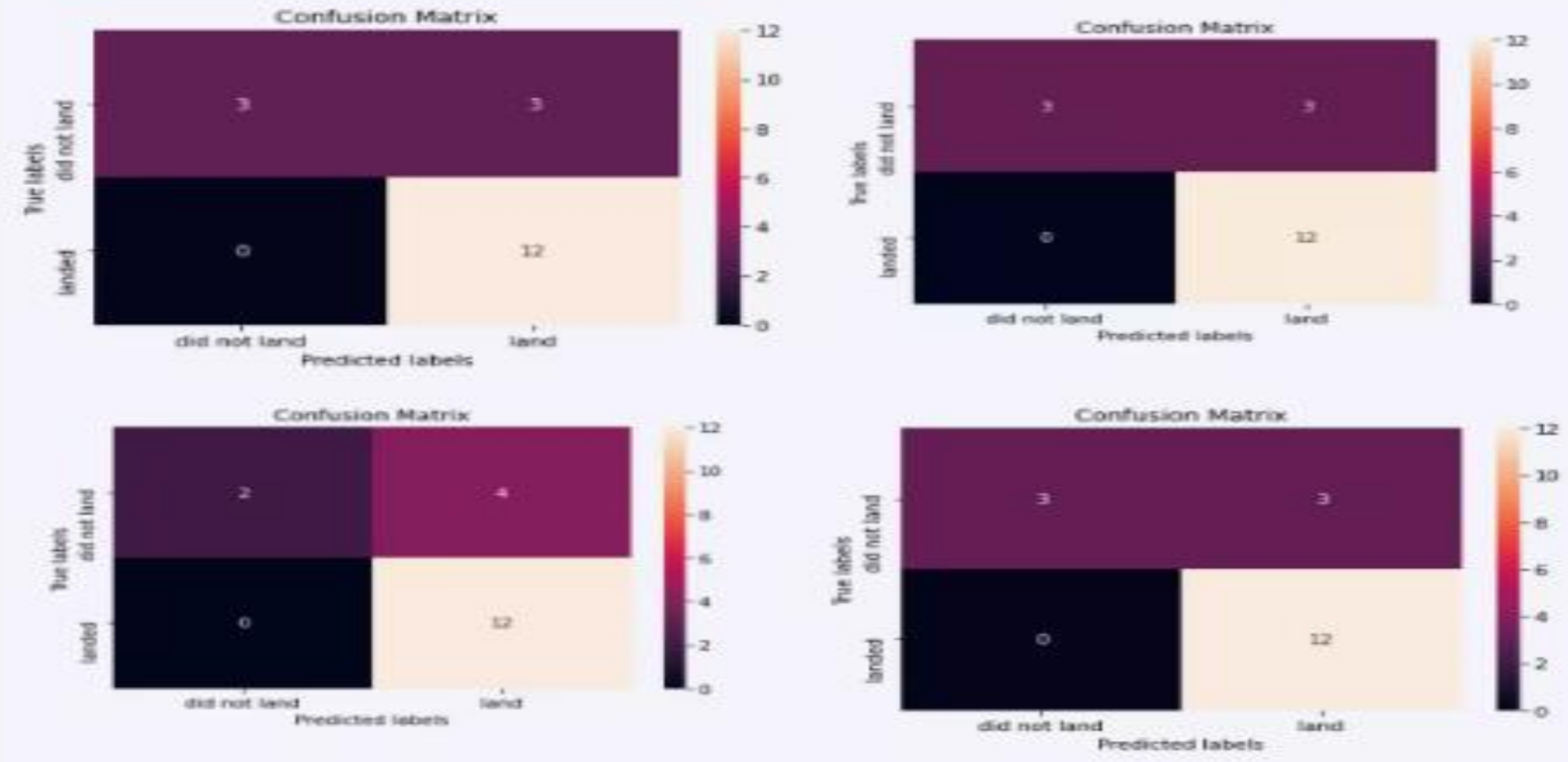
Section 5

Predictive Analysis (Classification)

Classification Accuracy



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.

Appendix

- https://github.com/yassin-asu/data-science-capstone-.git

Thank you!