Winning Space Race with Data Science

Leonardo Vinciprova

7-Mar-2022

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 is a big rocket manufacturer and perhaps the most renowned company in doing so. If not self-evident, due to its size and technological circumstances, SpaceX is full of useful data for data scientists to help improve decision making for stakeholders
- Problems you want to find answers
  - The project task is to predict with some degree of confidence if the first stage of the SpaceX Falcon 9 rocket will land successfully

Section 1

Methodology

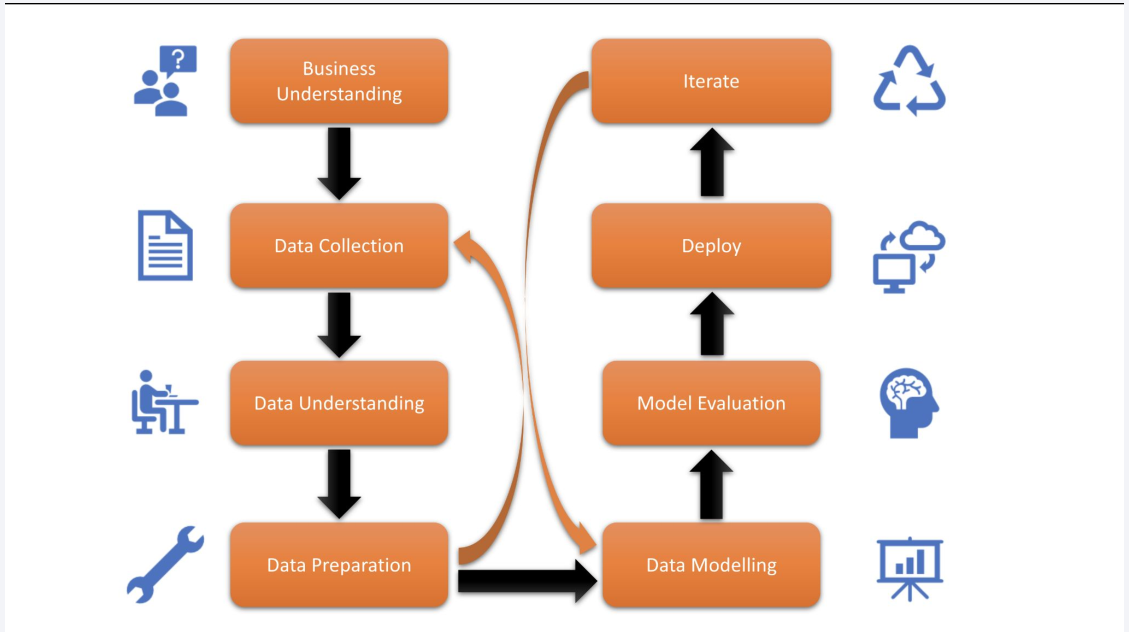
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping from Wikipedia data
- Perform data wrangling
  - One Hot Encoding data fields for Machine Learning and data cleaning of null values and, or, 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 was collected:
  - SpaceX launch data.
  - Through the APIs we will get data about launches such as: information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
  - It was also made use of methods like “Web Scrapping” and the “beatifulsoup”.



Data Collection – SpaceX API

- Data collection with SpaceX API REST
- Access codes at: https://github.com/GavetaDoPolvo/FINAL-ASSIGNMENT/blob/main/Data%20Collecting.ipynb

1 .Getting Response from API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url).json()
```

2. Converting Response to a .json file

```
response = requests.get(static_json_url).json()
data = pd.json_normalize(response)
```

3. Apply custom functions to clean data

```
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)
```

```
getBoosterVersion(data)
```

4. Assign list to dictionary then dataframe

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

```
df = pd.DataFrame.from_dict(launch_dict)
```

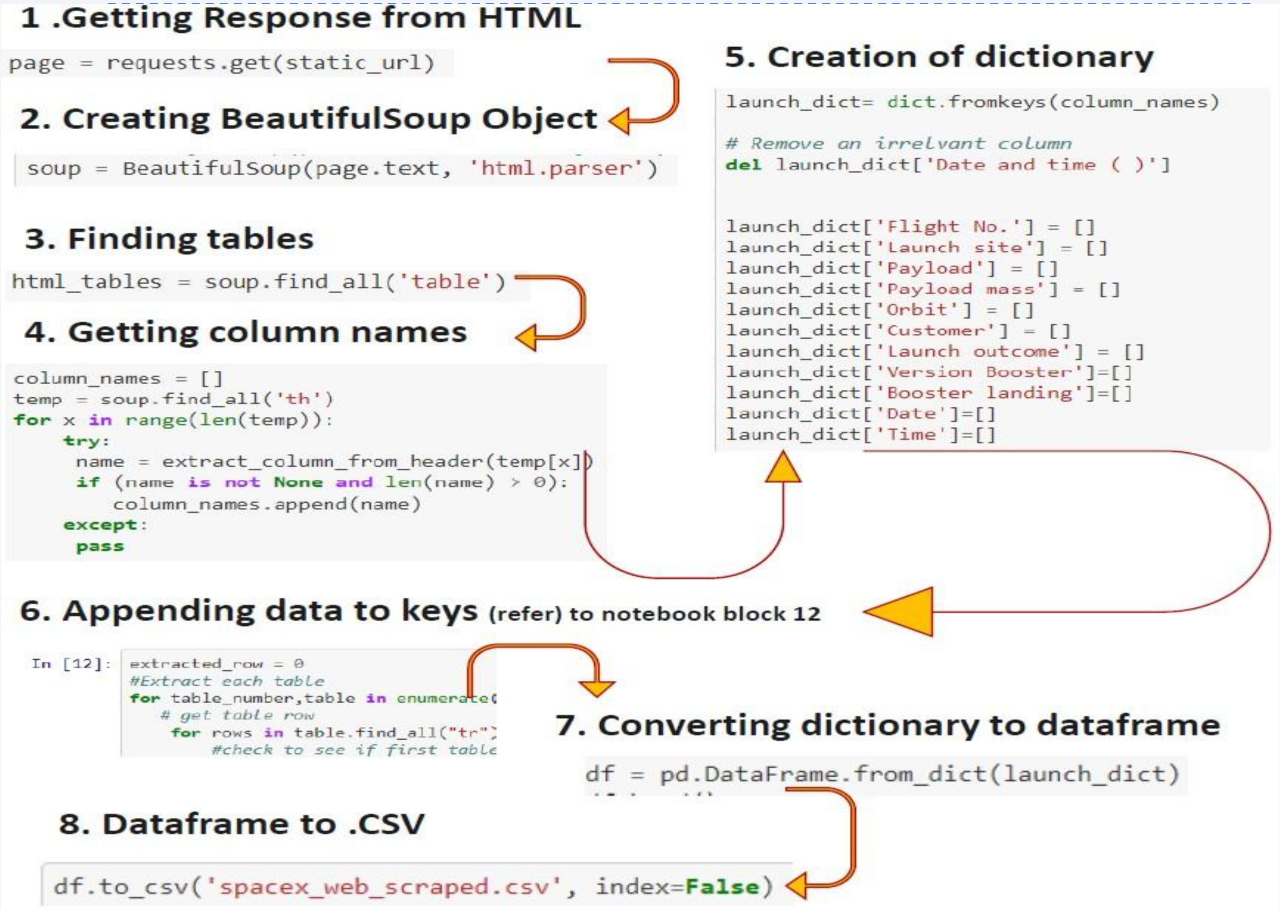
5. Filter dataframe and export to flat file (.csv)

```
data_falcon9 = df.loc[df['BoosterVersion']!="Falcon 1"]
```

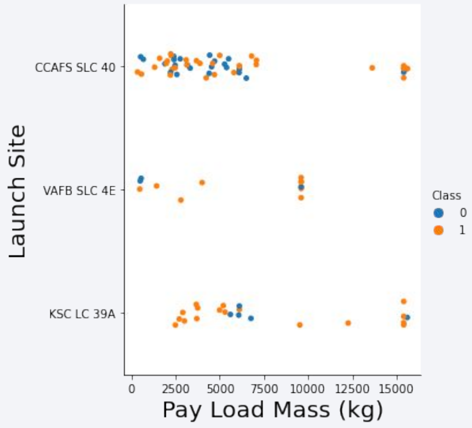
```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

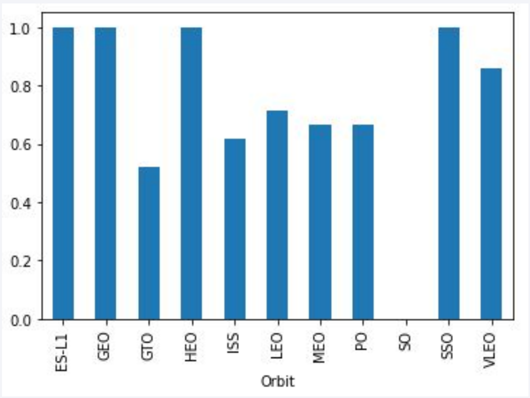
Data Wrangling

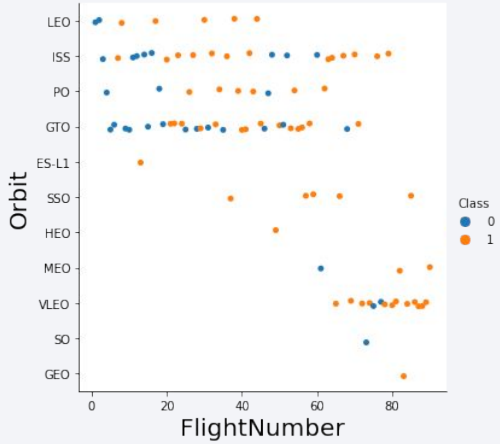
- Through Python in Jupyter platform,
- Coding and testing allowed for useful inferences on how to progress
- More information on the image on the right or at the notebook itself(https://github.com/GavetaDoPolvo/FINAL-ASSIGNMENT/blob/main/Web%20Scrapping.ipynb)

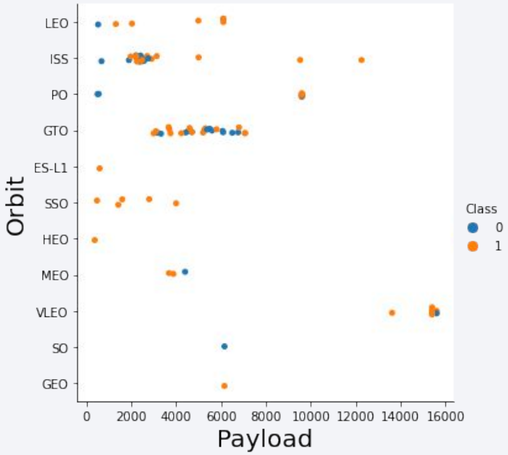


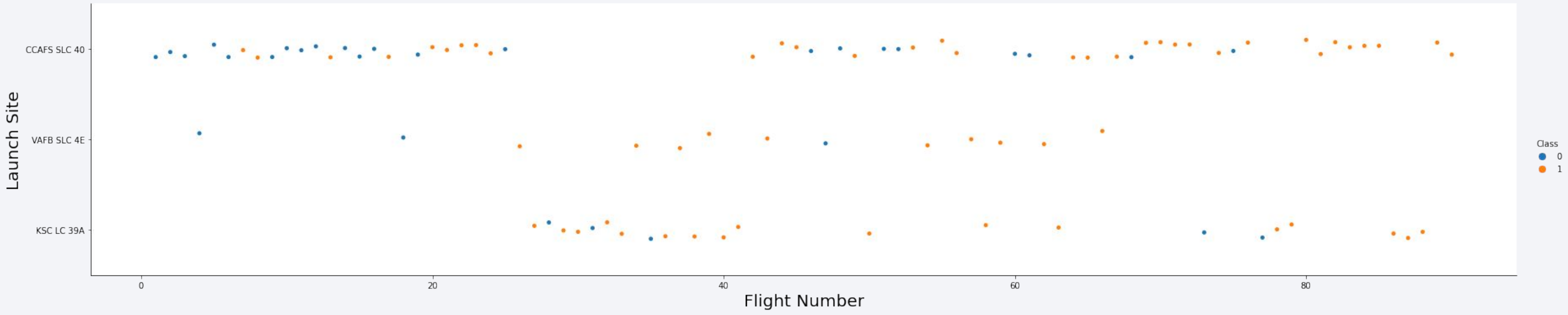
EDA with Data Visualisation

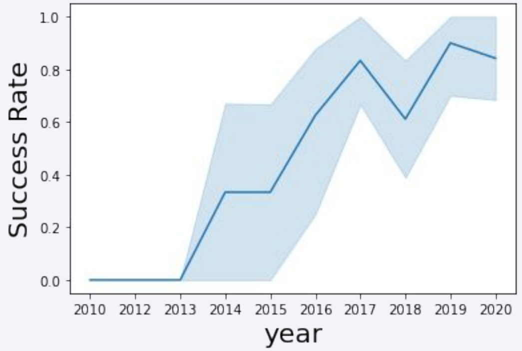






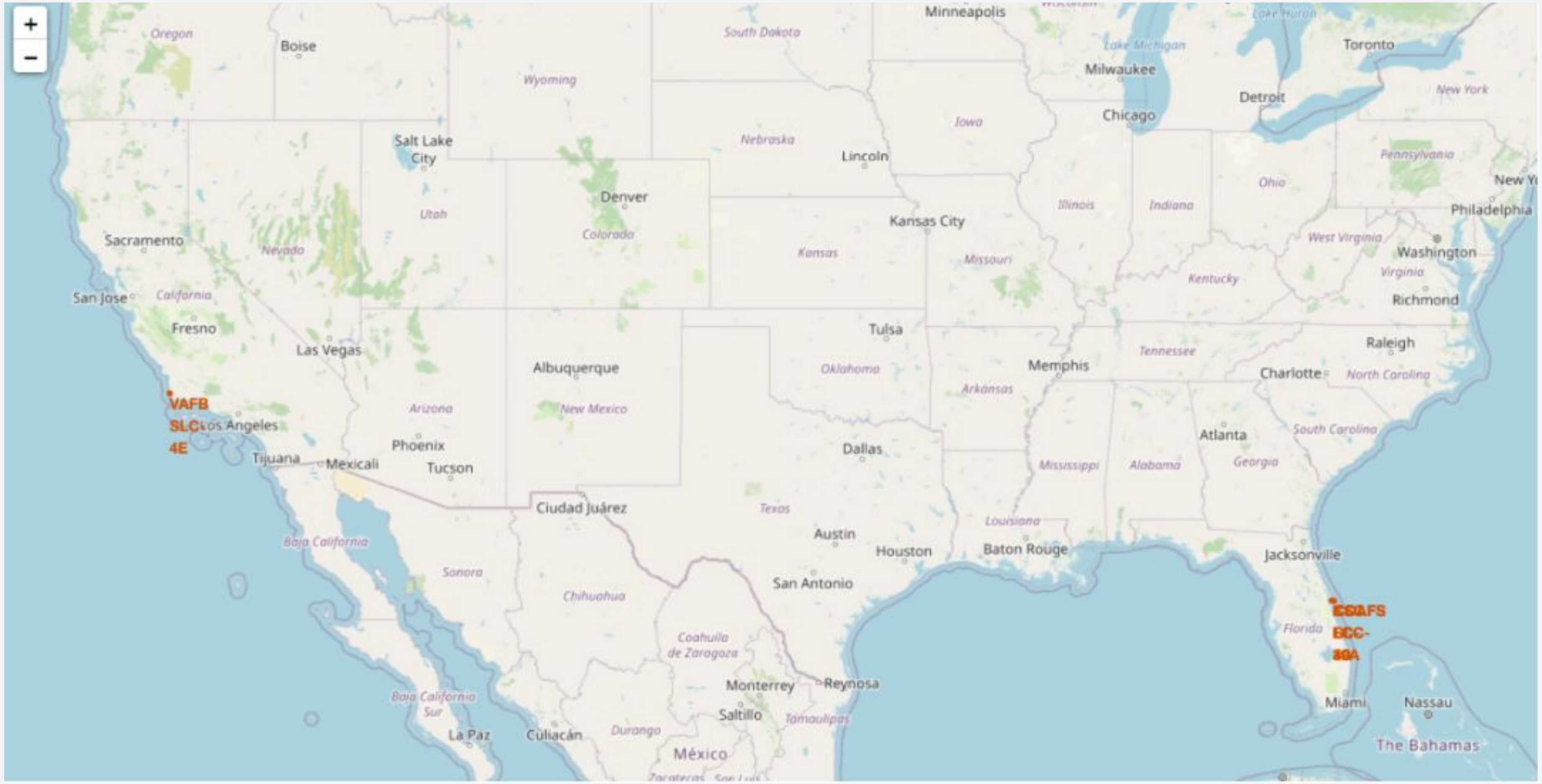






More on:
https://github.com/GavetaDoPolvo/FINAL-ASSIGNMENT/blob/main/Exploratory%20Analysis%20using%20pandas%20and%20matplotlib.ipynb

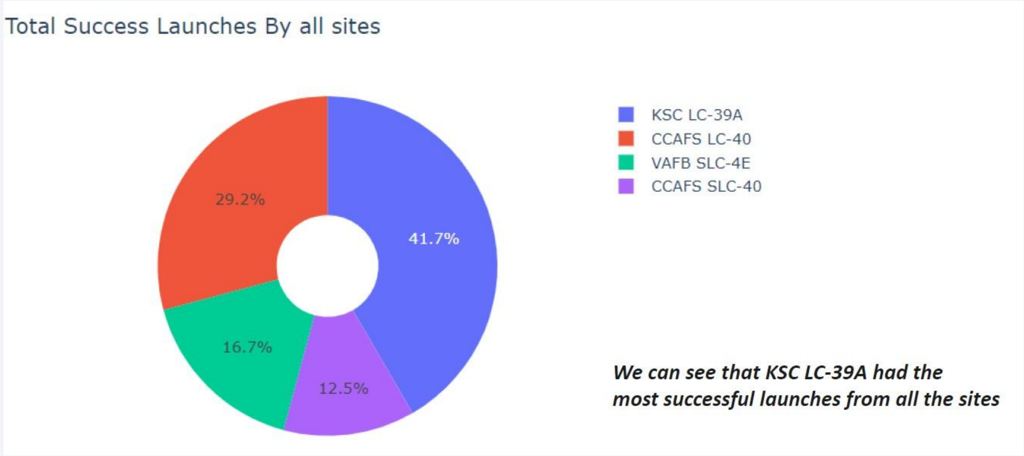
Build an Interactive Map with Folium

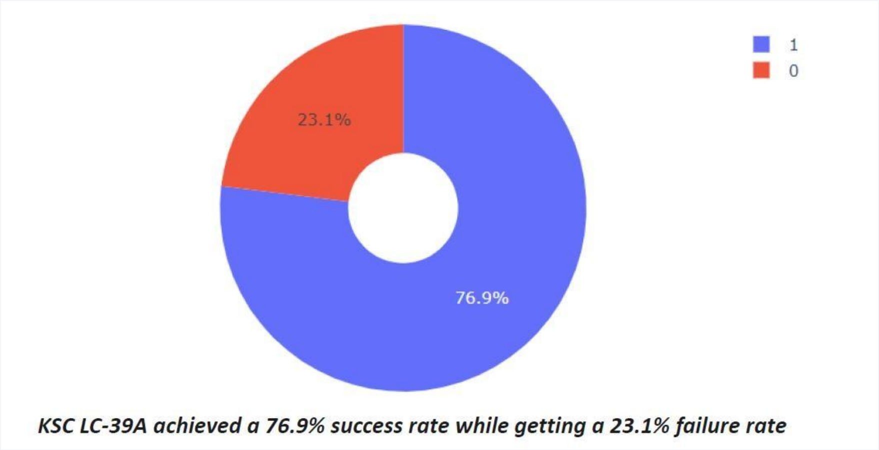


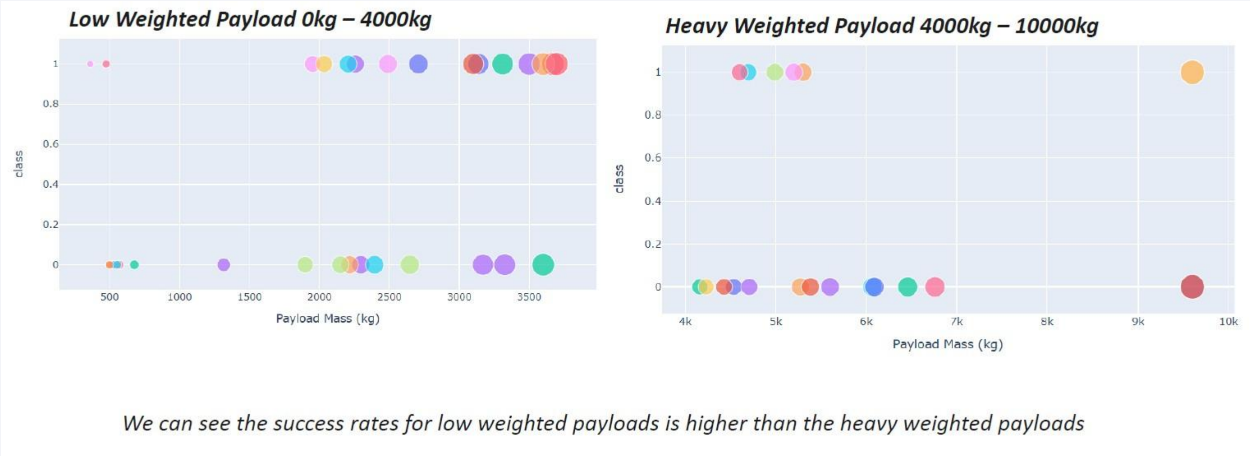
The markers were added to serve in finding an ideal location for building a launch site

(https://github.com/GavetaDoPolvo/FINAL-ASSIGNMENT/blob/main/Interactive%20visual%20analytics%20and%20dashboard.ipynb)

Build a Dashboard with Plotly Dash



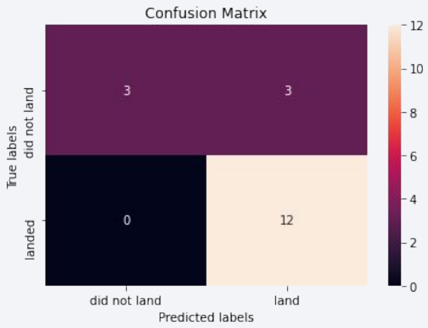


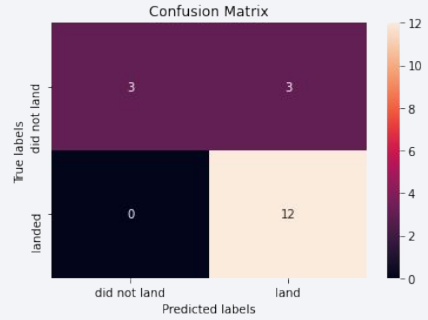


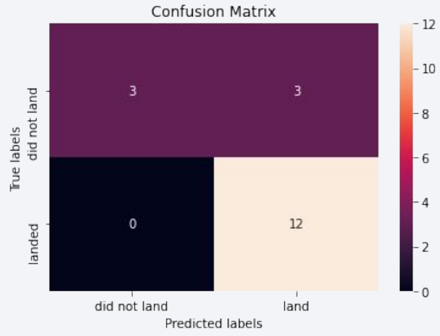
(https://github.com/GavetaDoPolvo/FINAL-ASSIGNMENT/blob/main/DASH)

Predictive Analysis (Classification)

- The SVM, KNN, and Logistic Regression model resulted in a prediction with 83.3% of confidence, while SVM performs better on “Area Under the Curve”, at 0.958.







(https://github.com/GavetaDoPolvo/FINAL-ASSIGNMENT/blob/main/Predictive%20Analysis.ipynb)

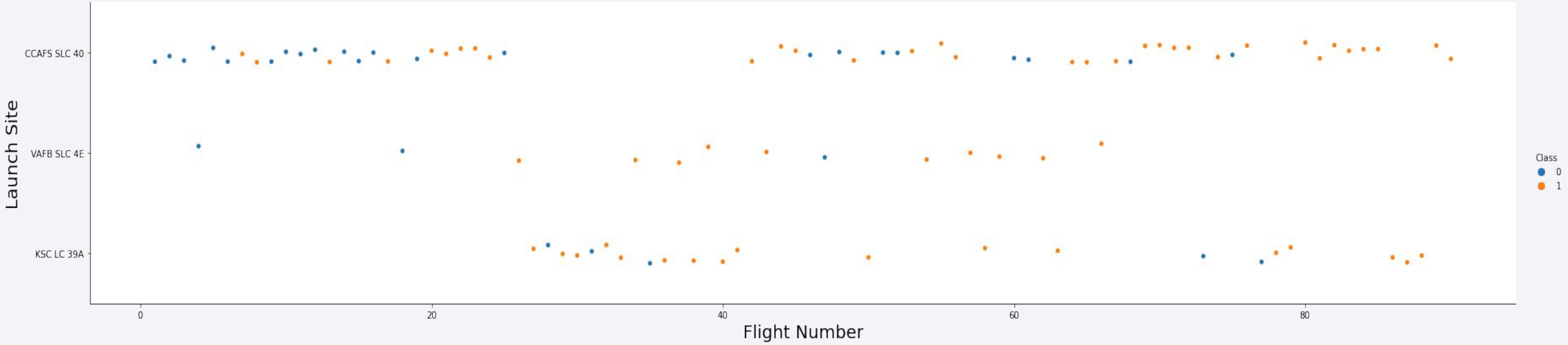
Results

- The SVM, KNN, and Logistic Regression models were found to be the best for prediction accuracy in this dataset.
- Low weighted payloads were more successful than the heavier payloads.
- There is a correlation between success rates and, presumably, development years.
- The launch site associated with the best outcomes was KSC LC 39A.
- Orbit GEO,HEO,SSO,ES L1 showed the most positive results.

Section 2

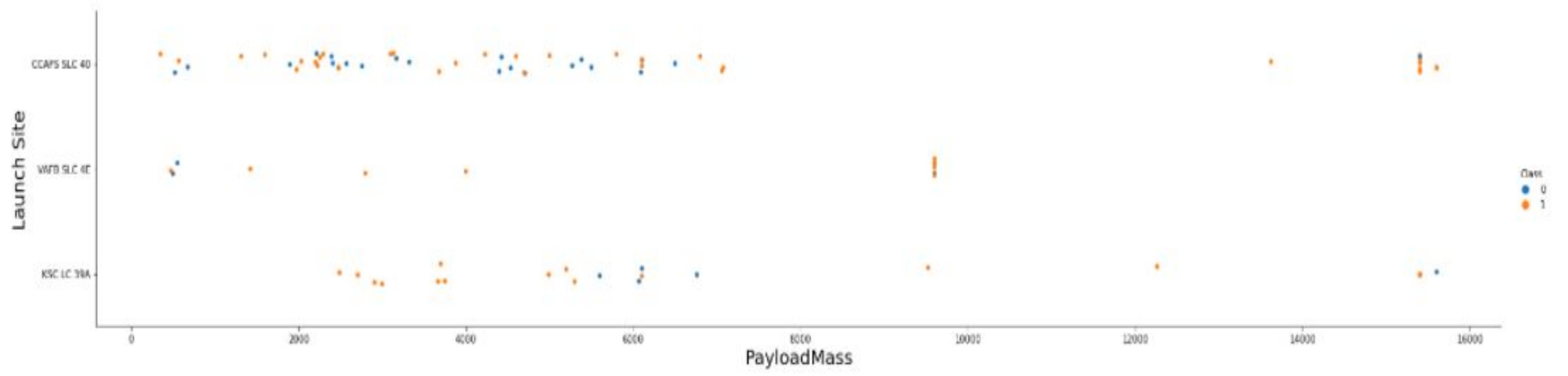
Insights drawn from EDA

Flight Number vs. Launch Site



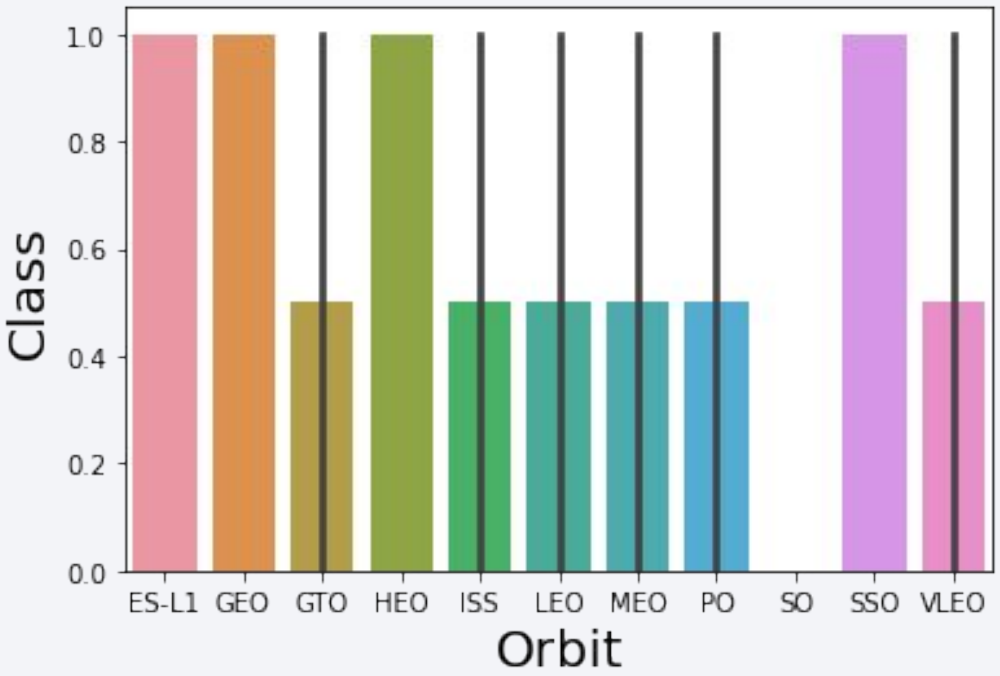
- Launches performed at the site CCAFS SLC 40 were significantly higher than launches from the other sites.

Payload vs. Launch Site



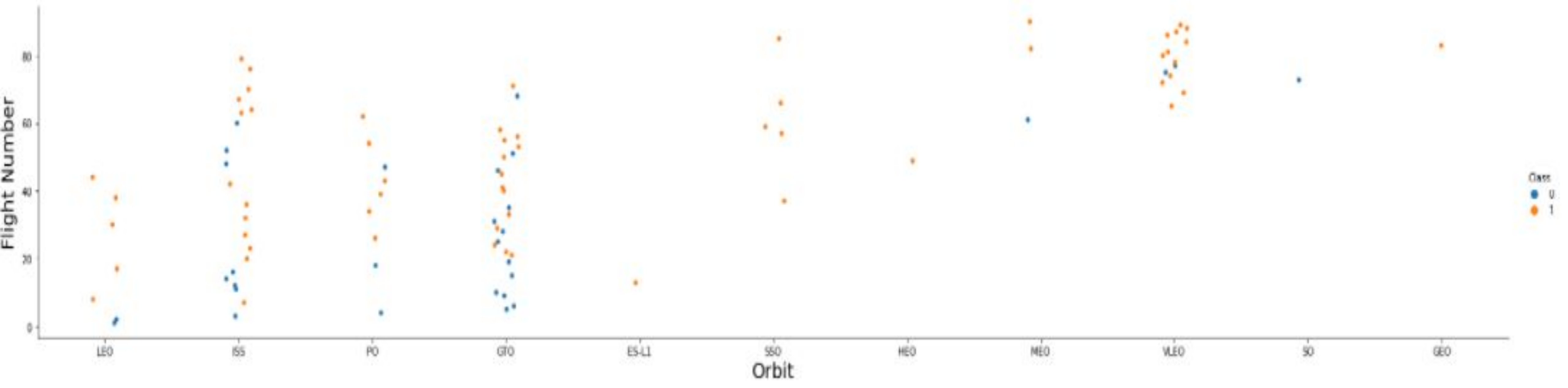
- Most of “Pay Loads” with lower “Mass” values were launched from CCAFS SLC 40

Success Rate vs. Orbit Type



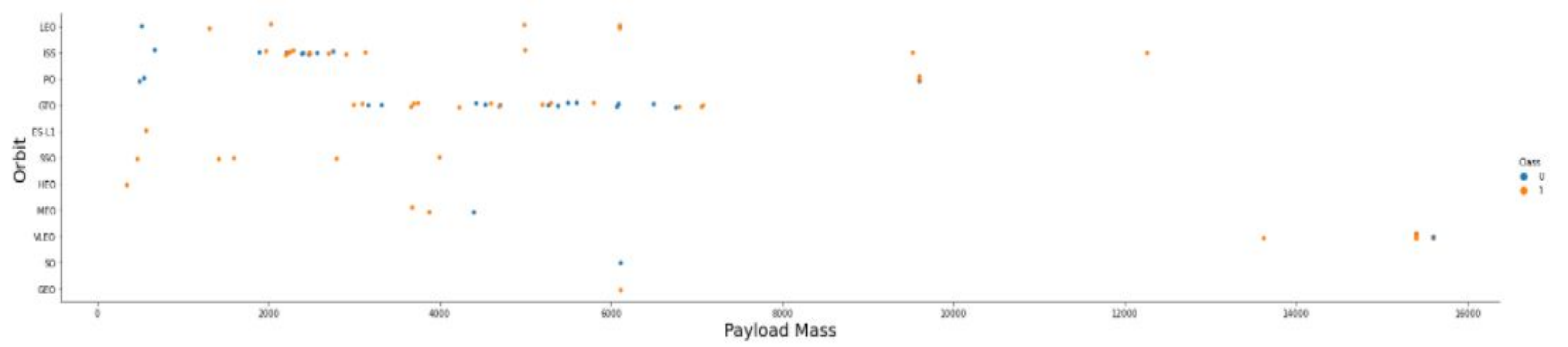
- Evidently, ES-L1, GEO, HEO, SSO had the highest success rate among the other orbit types.

Flight Number vs. Orbit Type



- There is an observable tendence of a change in distribution of launches in recent years.

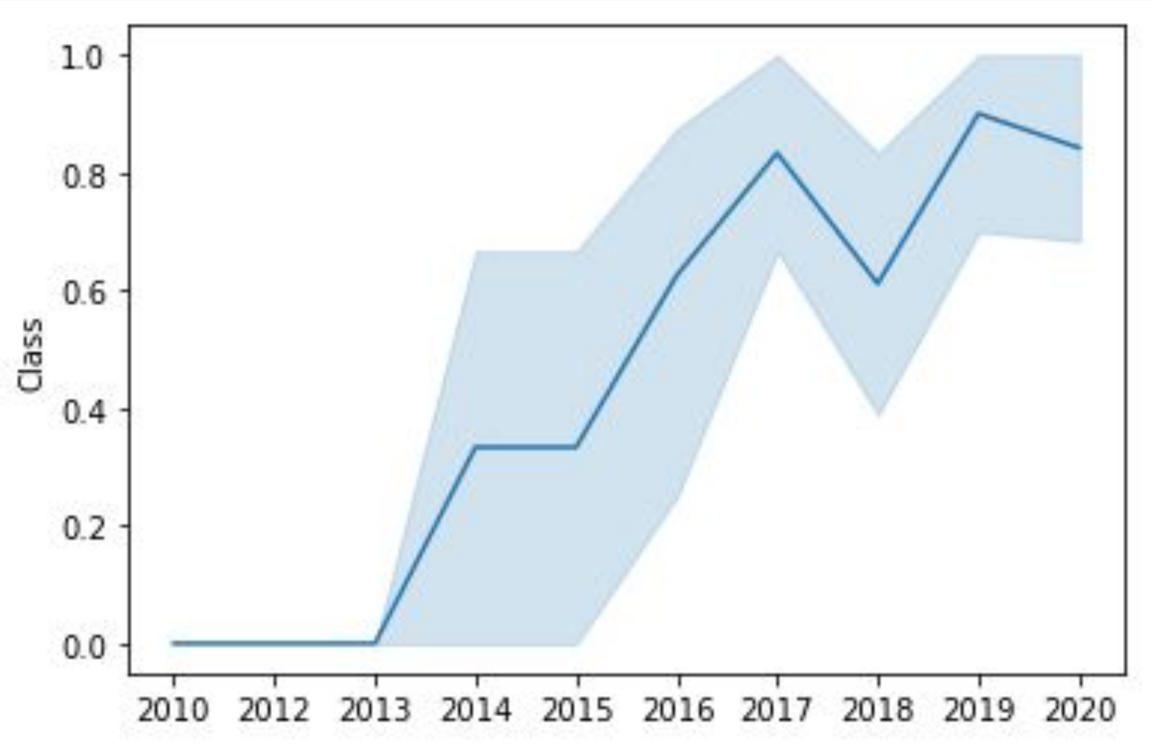
Payload vs. Orbit Type



- ISS and Payload at the range around 2000 are higl, as well as between GTO and the range of 4000-8000.

Launch Success Yearly Trend

- As shown in the graphic on the right, the success rates for launches increased greatly since 2013 but then it indicated either a normalization or a light fall after 2019



All Launch Site Names

- Using the query “%sql select distinct(LAUNCH_SITE) from SPACEXTBL”, we’ve gotten:

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

Launch Site Names Begin with 'CCA'

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

- Query: “%sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5”

Total Payload Mass

- Query: “%sql select sum(PAYLOAD_MASS   KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'”

- Output: 45596

Average Payload Mass by F9 v1.1

- Query: “%sql select avg(PAYLOAD_MASS_KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'”

- Output: 2928.400000

First Successful Ground Landing Date

- Query: “%sql select min(DATE) from SPACEXTBL where Landing Outcome = 'Success (ground pad)'”

- Output: 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query: “%sql select BOOSTER_VERSION from SPACEXTBL where Landing Outcome = 'Success (drone ship)' and PAYLOAD_MASS_KG_ > 4000 and PAYLOAD_MASS_KG_ < 6000”

- Output:

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Query: “%sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'”

- Output: 100

Boosters Carried Maximum Payload

- Query: “%sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS KG = (select max(PAYLOAD_MASS KG_) from SPACEXTBL)”

- Output on the right

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

- Query: “%sql select * from SPACEXTBL where Landing Outcome like 'Success%' and
(DATE between ‘2015-01-01' and ‘2015-12-31') order by date desc”
- Output:

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)

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

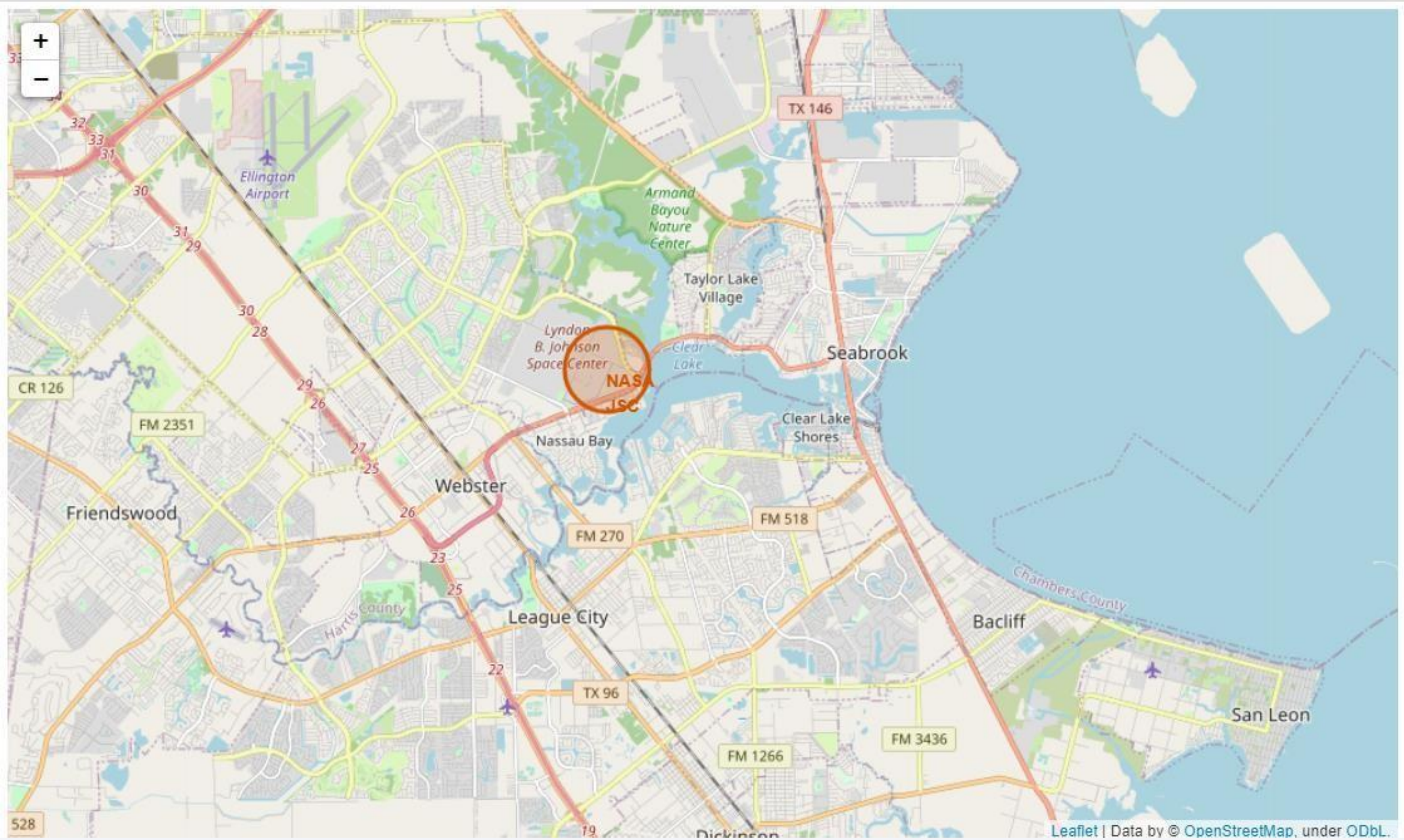
- Query: “%sql select * from SPACEXTBL where Landing Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc”
- Output:

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)

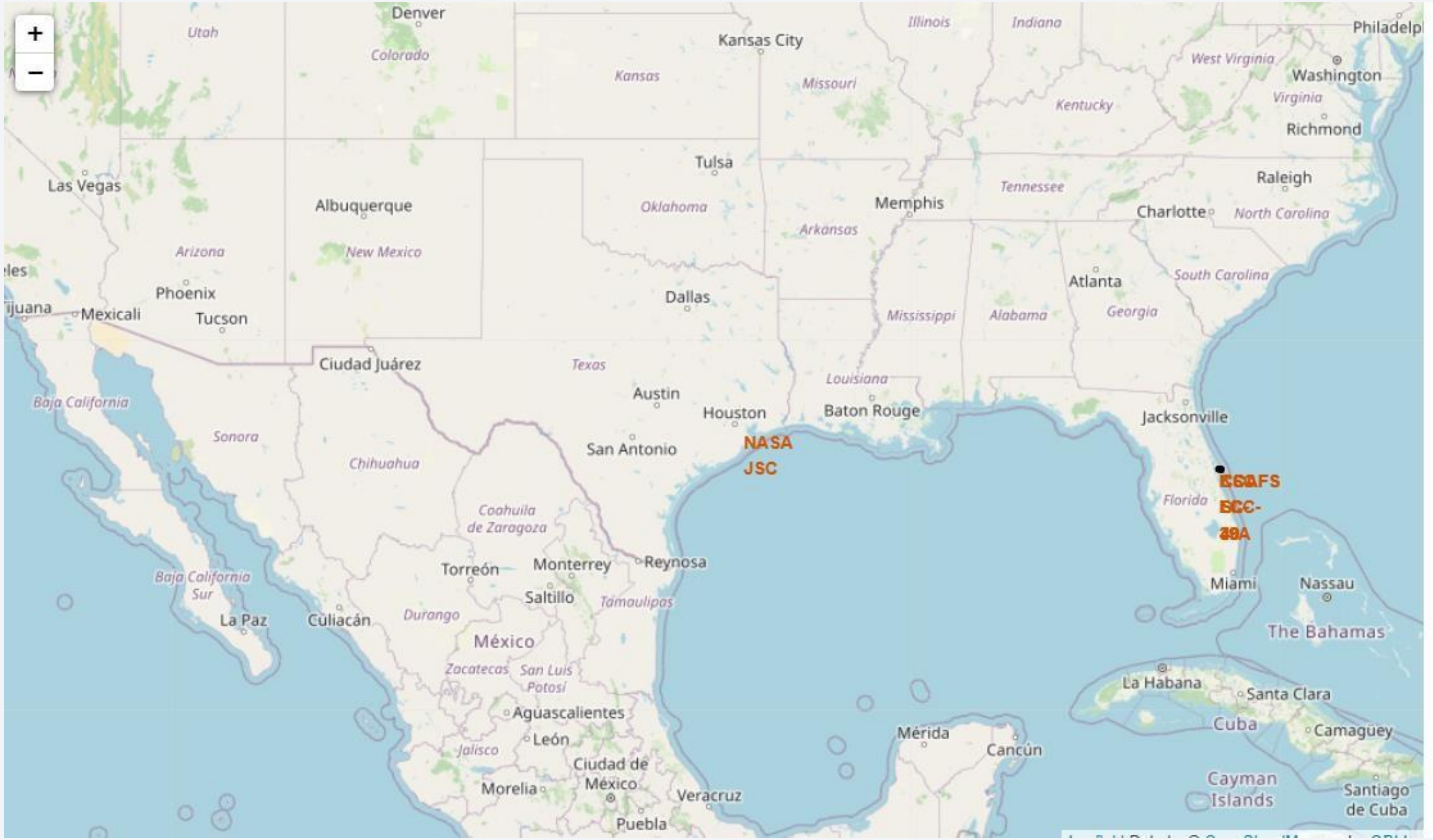
Section 4

Launch Sites Proximities Analysis

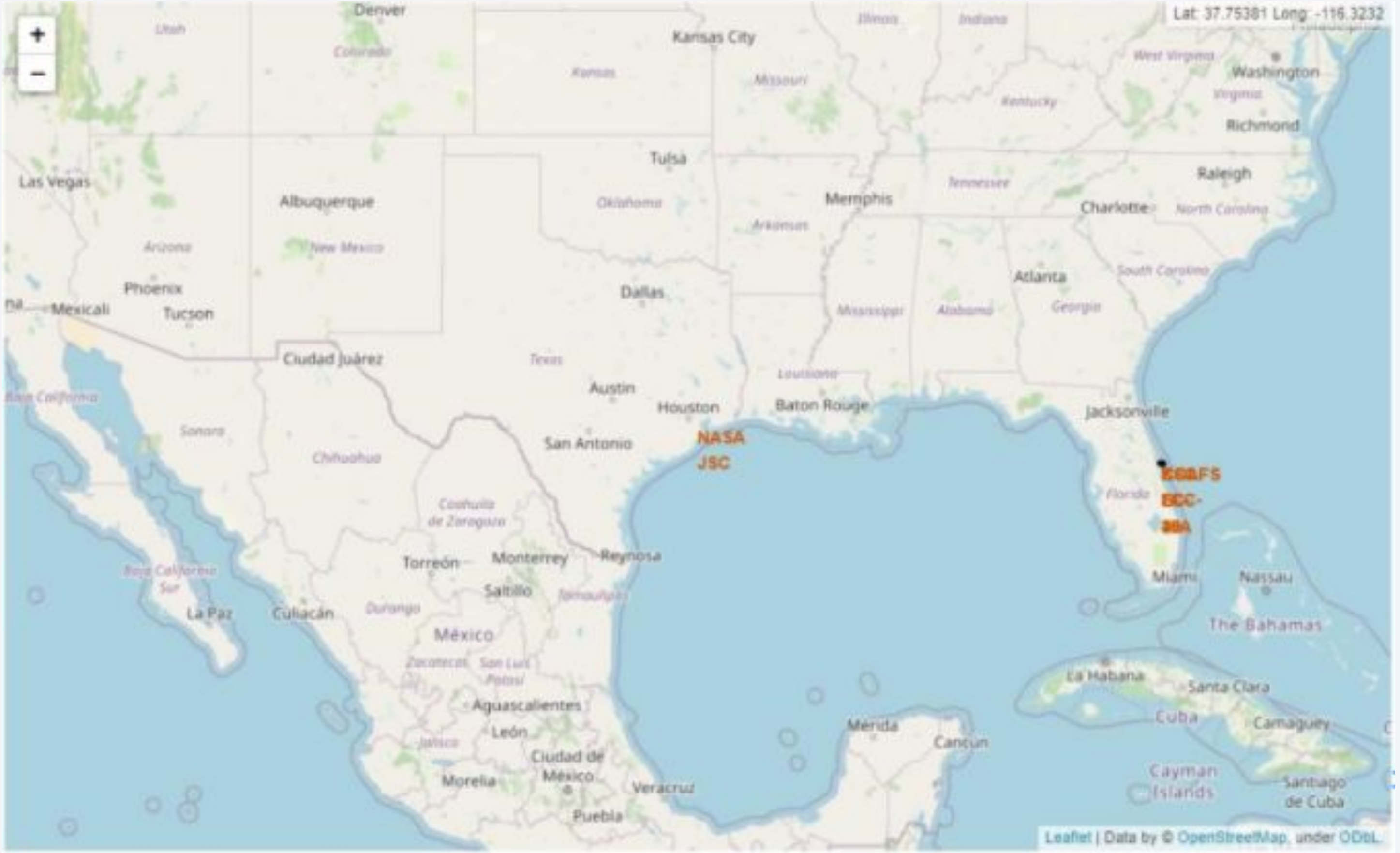
All launch sites marked on a map



Succeeded/failed launches

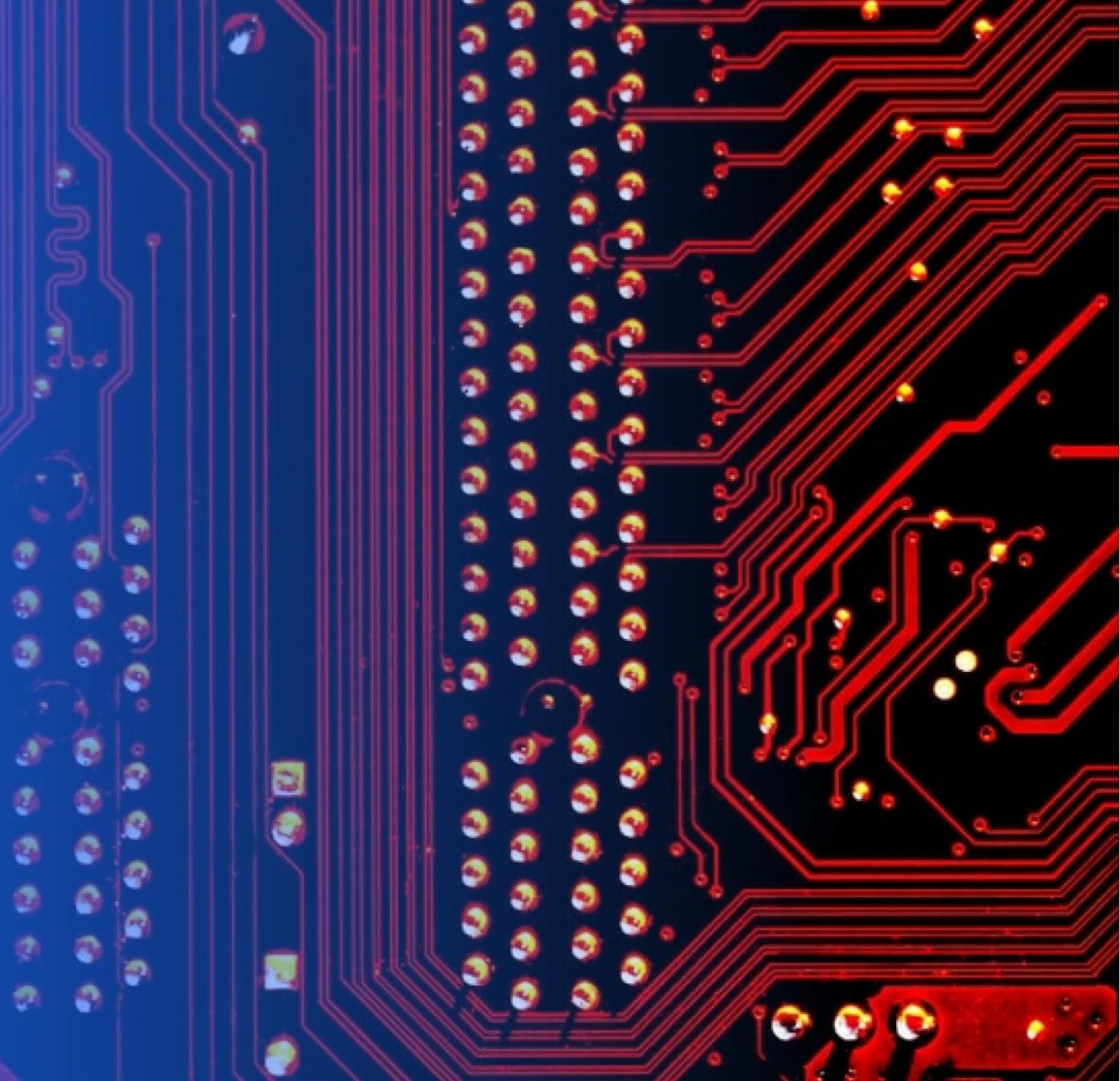


Distances between launch sites and its proximities

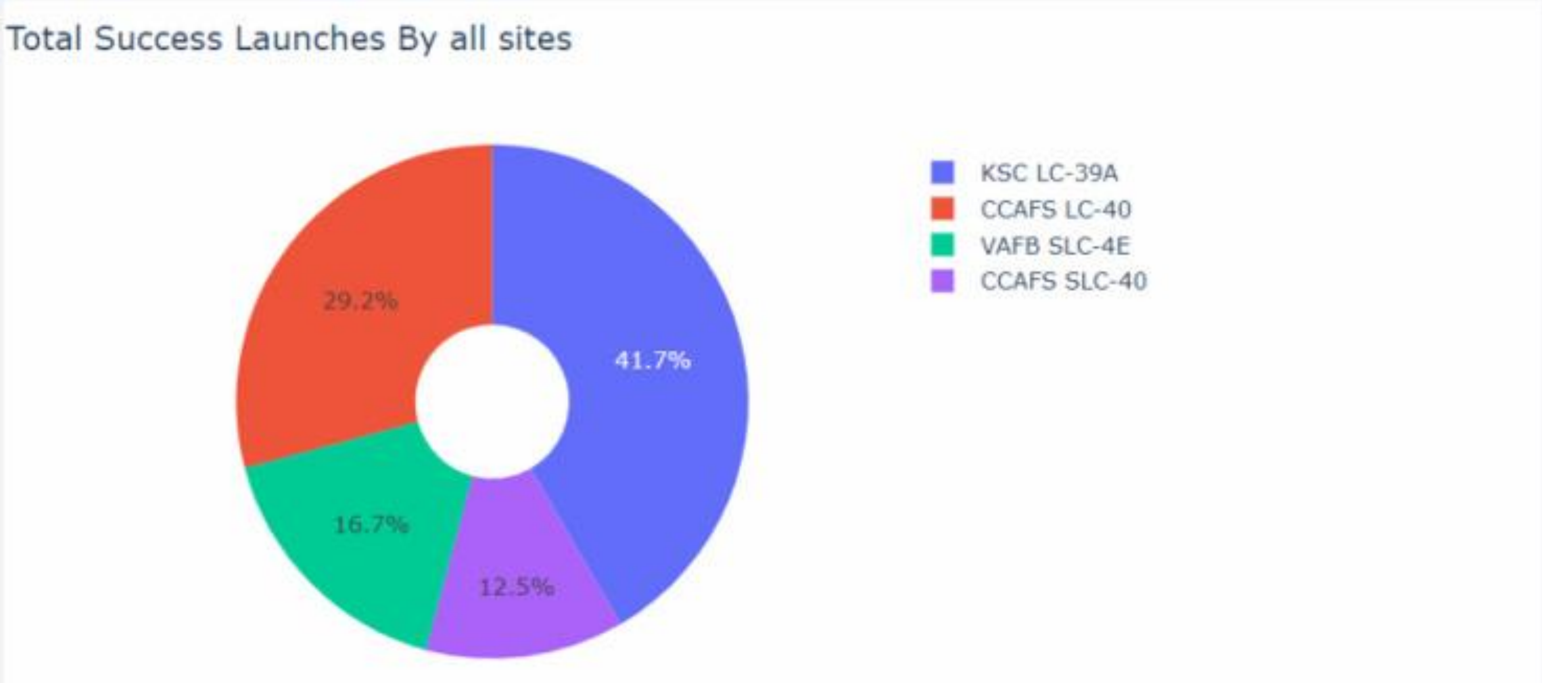


Section 5

Build a Dashboard with Plotly Dash

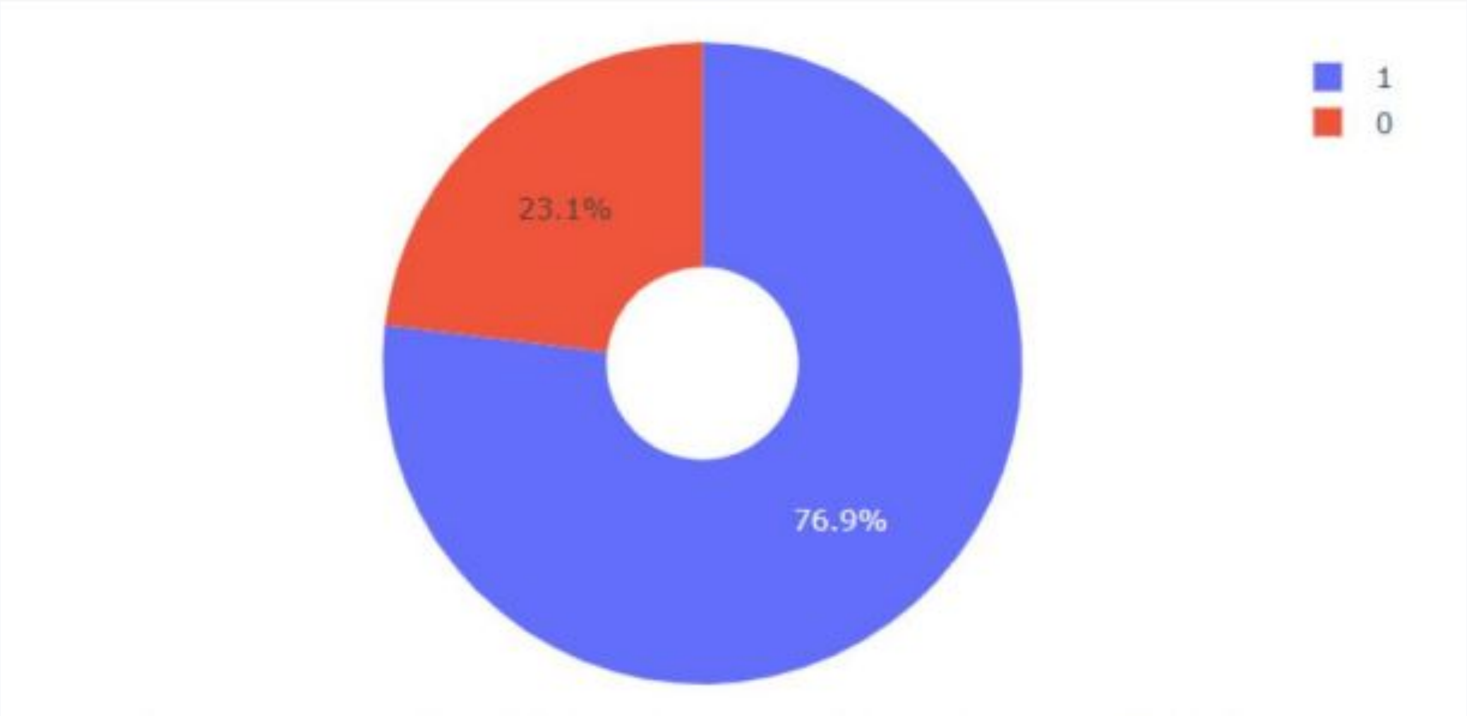


Total success launches by all sites



- So, as the pie chart demonstrates, the KSC LC-39A site has the best success rate, followed by CCAFS LC-40, and a close tie for VAFB SLC-4E and CCAFS SLC-40

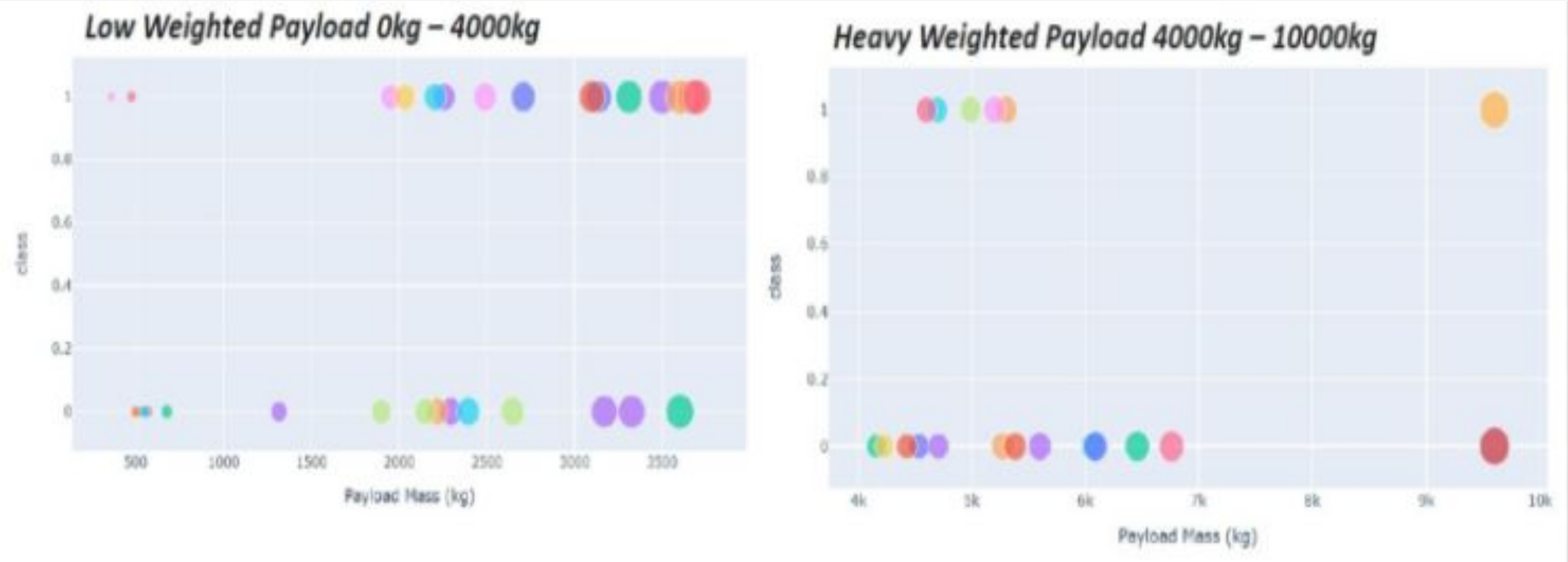
Success rate by site



- KSC LC-39A succeeded in 76.9% of the time

Payload vs launch outcome

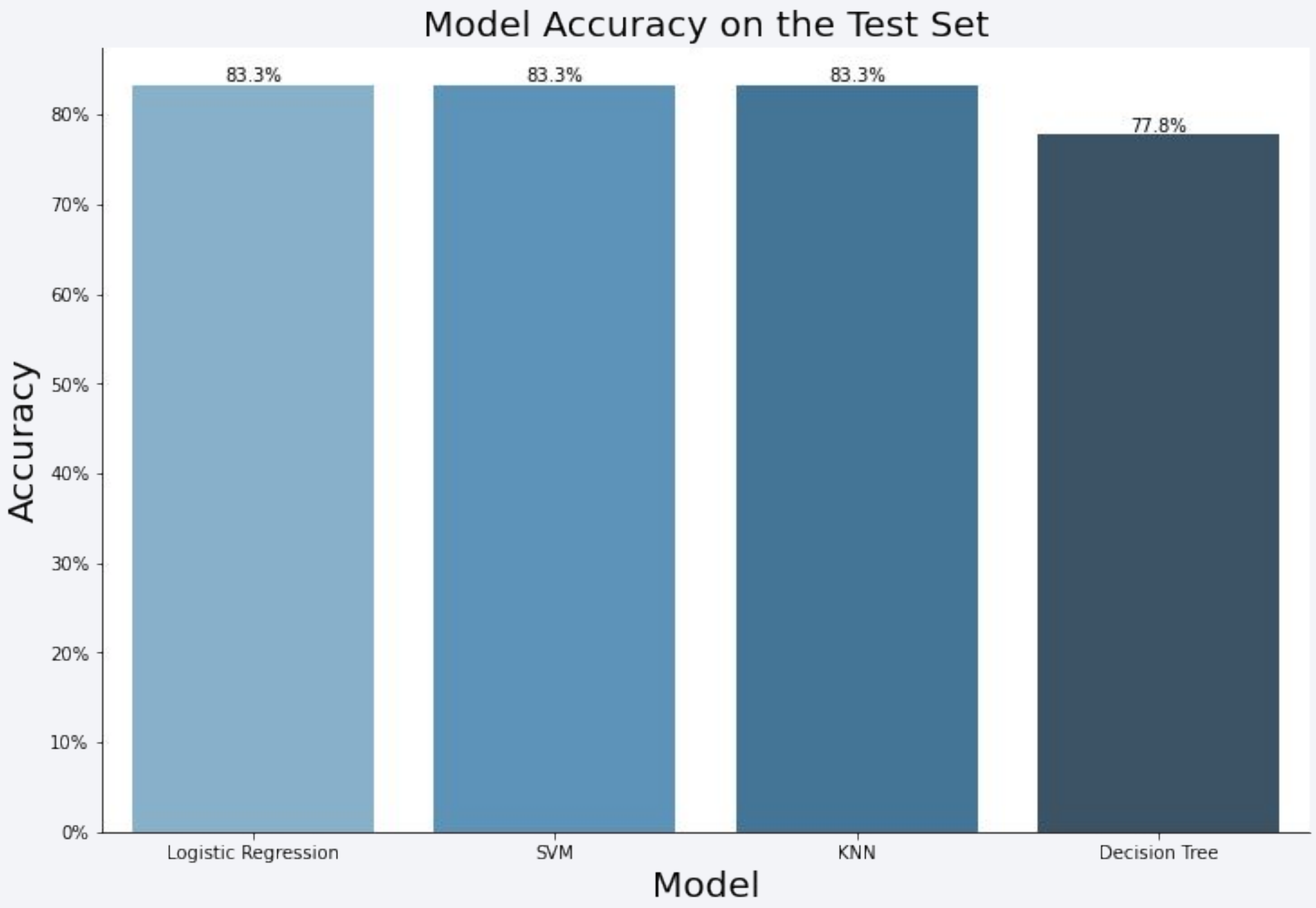
- The graphic below shows that low weighted rockets have higher odds of success compared to the heavy weighted



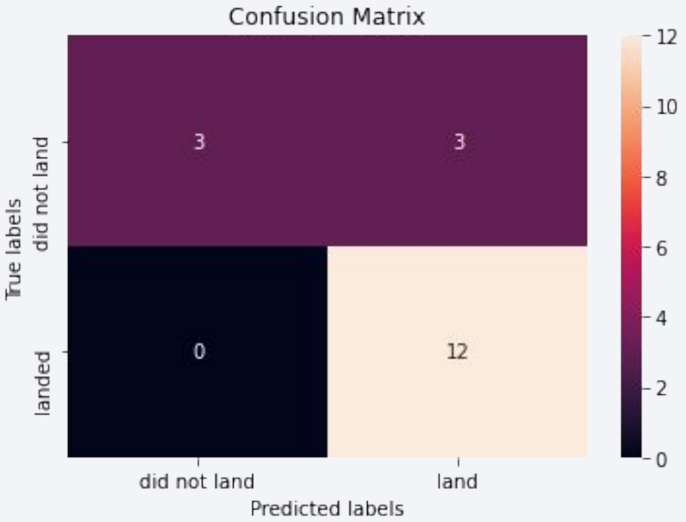
Section 6

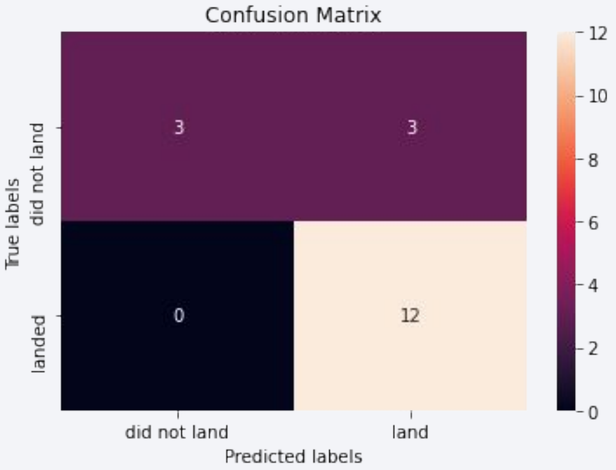
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix





Conclusions

Reiteration of the results:

- The SVM, KNN, and Logistic Regression models were found to be the best for prediction accuracy in this dataset.
- Low weighted payloads were more successful than the heavier payloads.
- There is a correlation between success rates and, presumably, development years.
- The launch site associated with the best outcomes was KSC LC 39A.
- Orbit GEO,HEO,SSO,ES L1 showed the most positive results.

And: the models were find to be highly accurate, so the inferences provided in this presentation are, in fact, reliable

Thank you!