Data Science And Space

Ayush ShANdIlyA

27-07-2024

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 (Classifications)

- Summary of all results
  - EDA results
  - Interactive analytics
  - Predictive analysis

Introduction

- Project background and context
  - SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars;
  - Other providers cost upward of 165 million dollars each, much of the saving is because SpaceX can reuse the first stage.

- Problems you want to find answers
  - The project task is to predicting if the first stage of the SpaceX Falcon 9 rocket will land successfully.

Section 1

Methodology

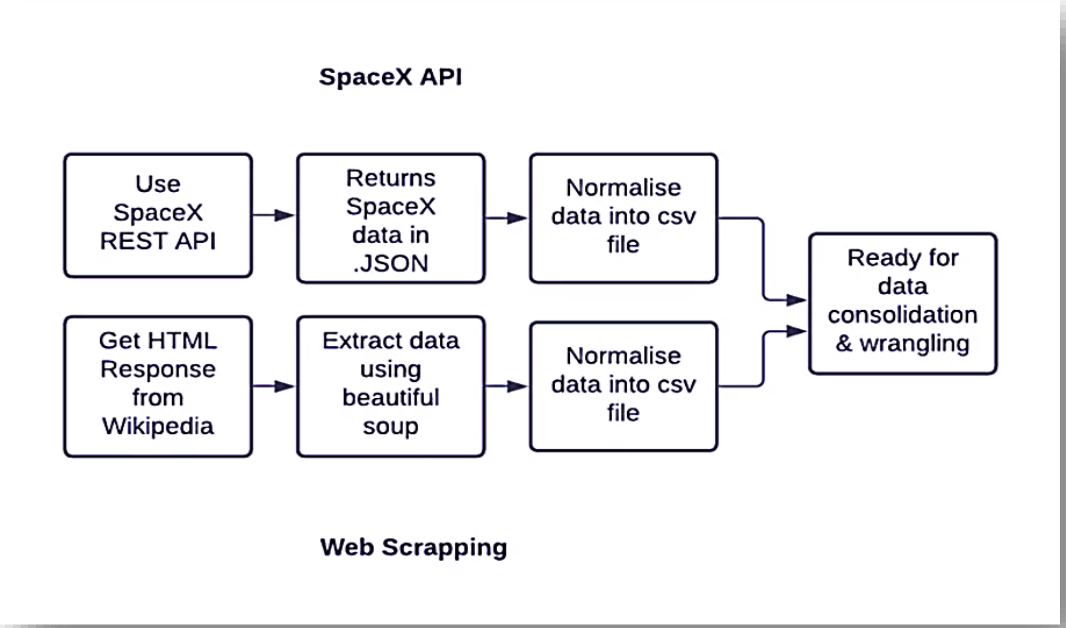
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - One Hot Encoding data fields for Machine Learning and 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
  - Linear Regression (LR), KNN, SVM, DT models have been built and evaluated for the best classifier

Data Collection

- Describe how data sets 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 specifications, 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 scraping Wikipedia using BeautifulSoup.



Data Collection – SpaceX API

- Data collection with SpaceX RESET calls.

https://github.com/ShanAyush/IBM-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

1. Getting Response from API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

2. Converting Response to a .json file

```
# Use json_normalize meethod to convert the json result into a dataframe
dataa = response.json()
data = pd.json_normalize(dataa)
```

3. Apply custom functions to clean data

```
# Call getLaunchSite
getLaunchSite(data)
```

```
# Call getPayloadData
getPayloadData(data)
```

```
# Call getCoreData
getCoreData(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}
```

```
launch_df = pd.DataFrame(launch_dict)
```

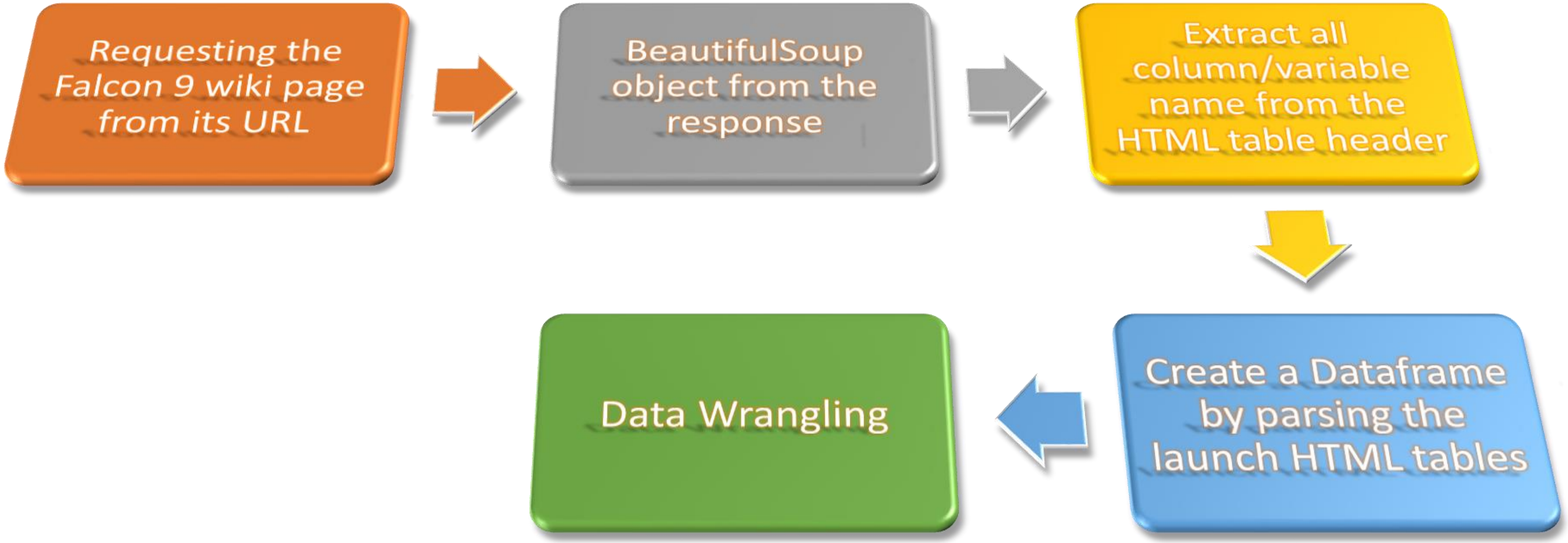
5. Filter dataframe and export to flat file (.csv)

```
data_falcon9 = launch_df[launch_df['BoosterVersion'] != 'Falcon 1']
```

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

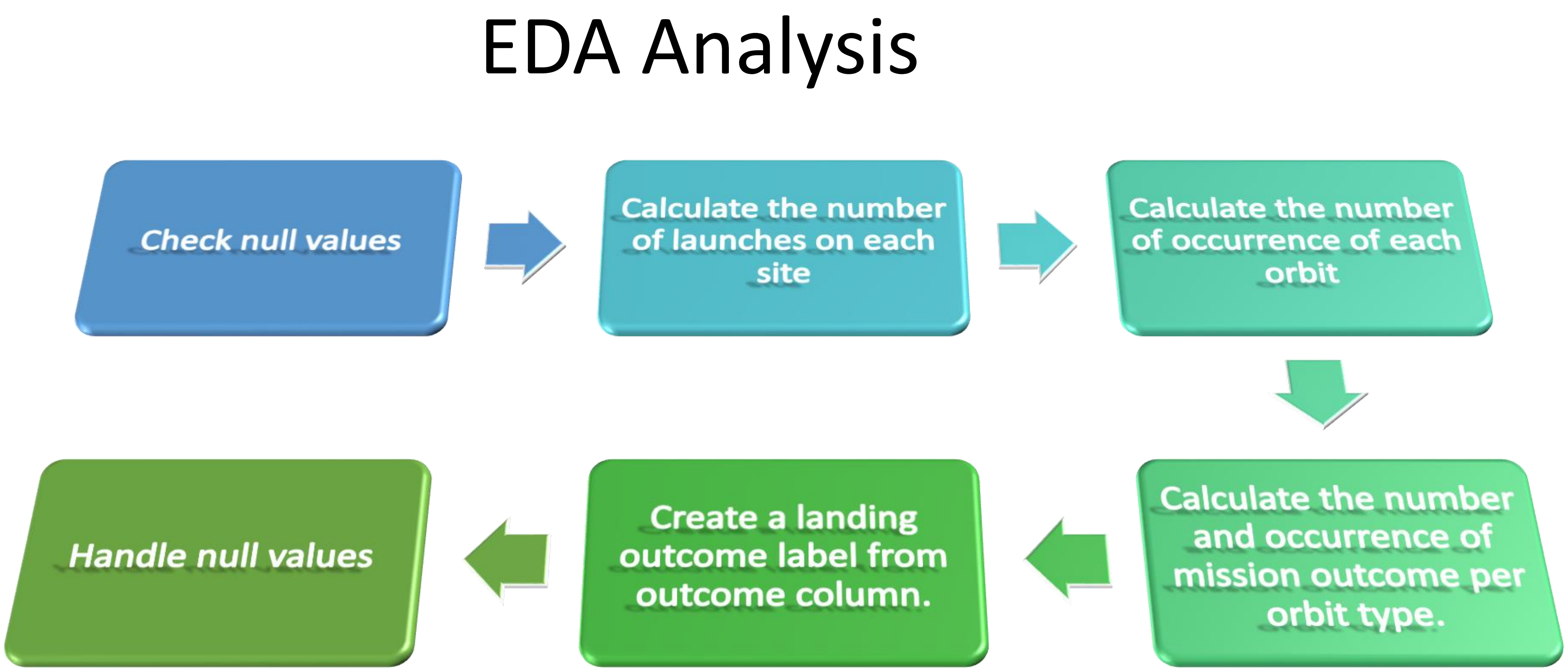
Data Collection - Scraping

- Perform web scraping to collect Falcon 9 historical launch records from Wikipedia page.



https://github.com/ShanAyush/IBM-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling



https://github.com/ShanAyush/IBM-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Summarize what charts were plotted and why you used those charts

- Catplot to visualize the relationship between flight number and payload.
- Catplot to visualize the relationship between flight number and launch site.
- Catplot to visualize the relational sip between payload and launch side.
- Bar chart to visualize the relationship between success rate of each orbit type
- Catplot to visualize the relationship between flight number and orbit type.
- Catplot to visualize the relationship between payload and orbit type.
- Line chart to visualize the launch success yearly trend.

https://github.com/ShanAyush/IBM-Data-Science-Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- SQL queries performed include:
  - Displaying the name of the unique launch site in the space mission.
  - Displaying five record 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 data 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_version which have carried the maximum payload mass.
  - Listing the records which will display the month names, successful landing_outcomes in ground pad, boosters_Version, 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 descending order.

https://github.com/ShanAyush/IBM-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

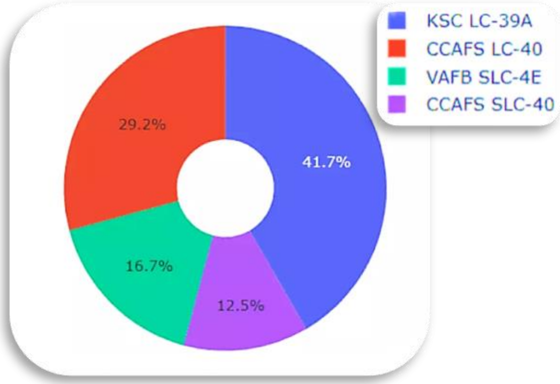
Build an Interactive Map with Folium

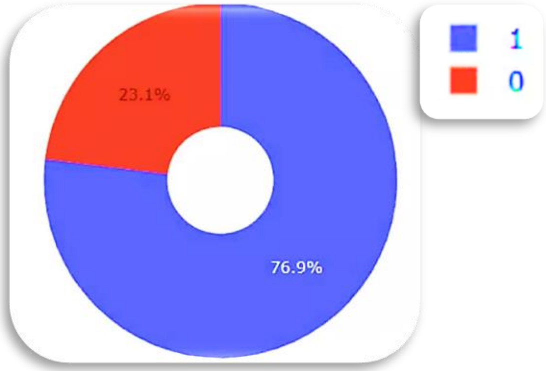
Map markers have been added to the map with aim to finding an optimal location for building a launch site.

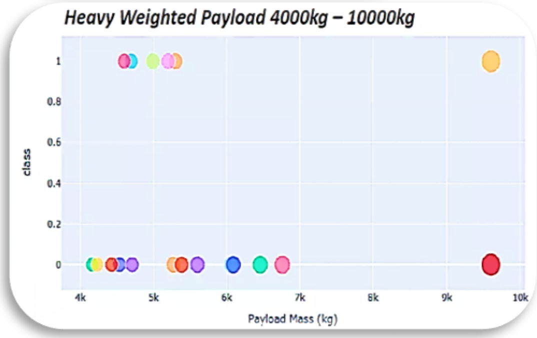
https://github.com/ShanAyush/IBM-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

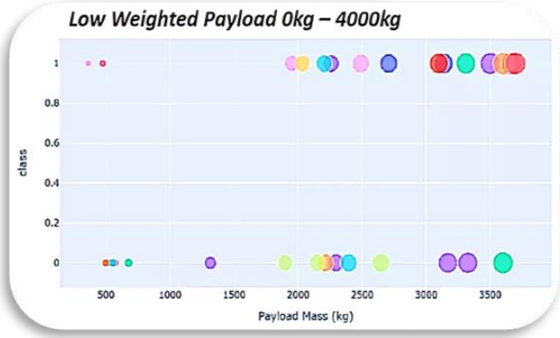
Build a Dashboard with Plotly Dash

- Summary of plot/graph and interactions on spaceX launch data in real-time
  - A callback function to render the success-payload-scatter-chart scatter plot.
  - To visually observe how payload may be correlated with mission outcome for selected site.





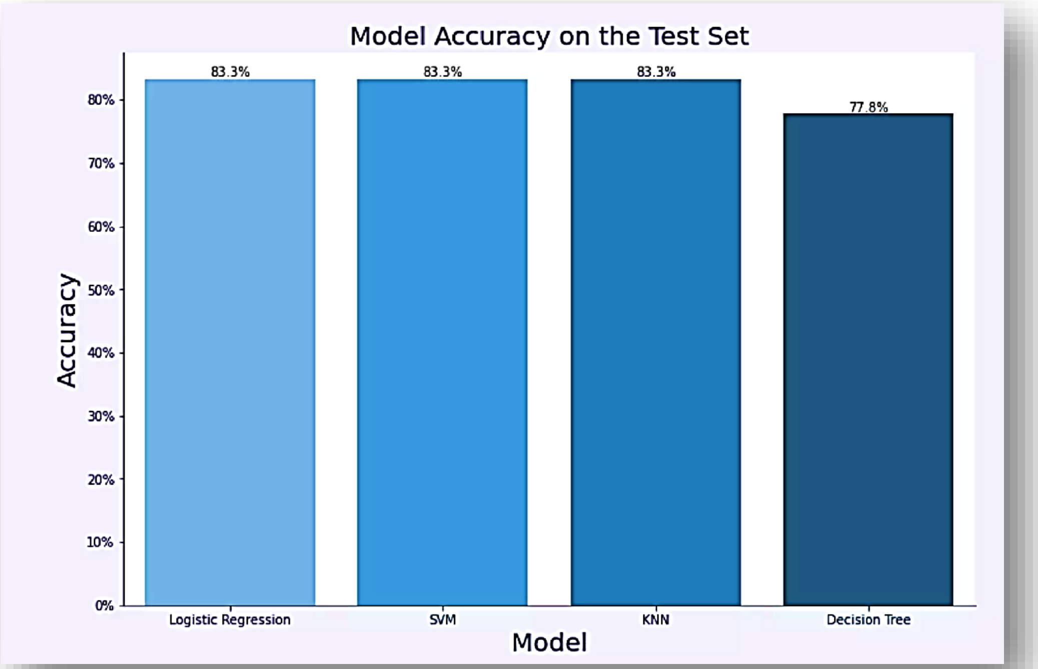


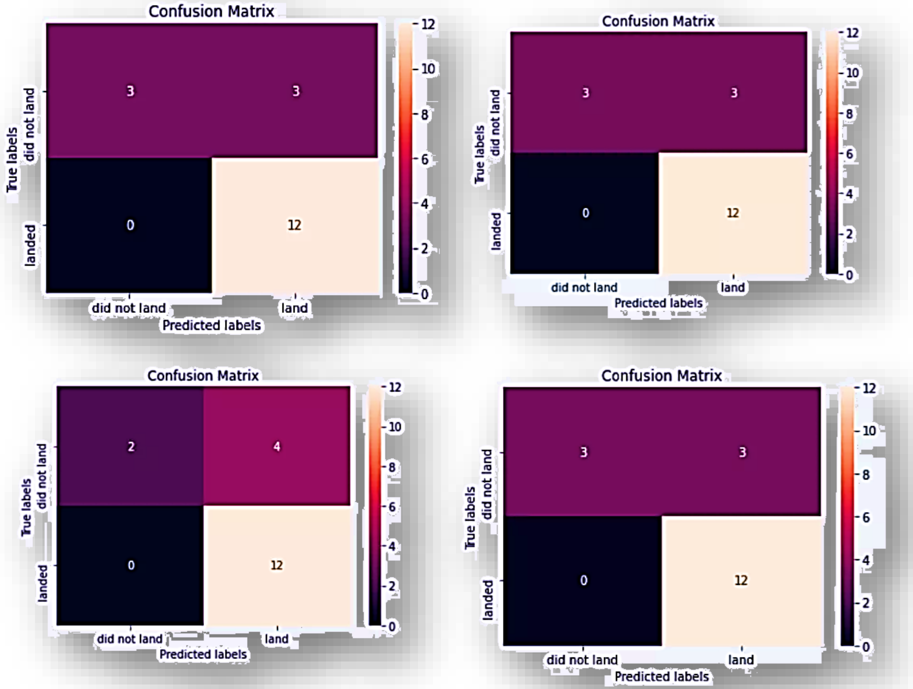


https://github.com/ShanAyush/IBM-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- The SVM, KNN and Logistic Regression model achieved 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/ShanAyush/IBM-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- The SVM, KNN and Logistic regression models are the best in term of prediction accuracy of the dataset.

- Low weighted payloads perform better than the heavier payloads.

- The success rate for SpaceX launch is directly proportional time in years they will eventually perfect the launches.

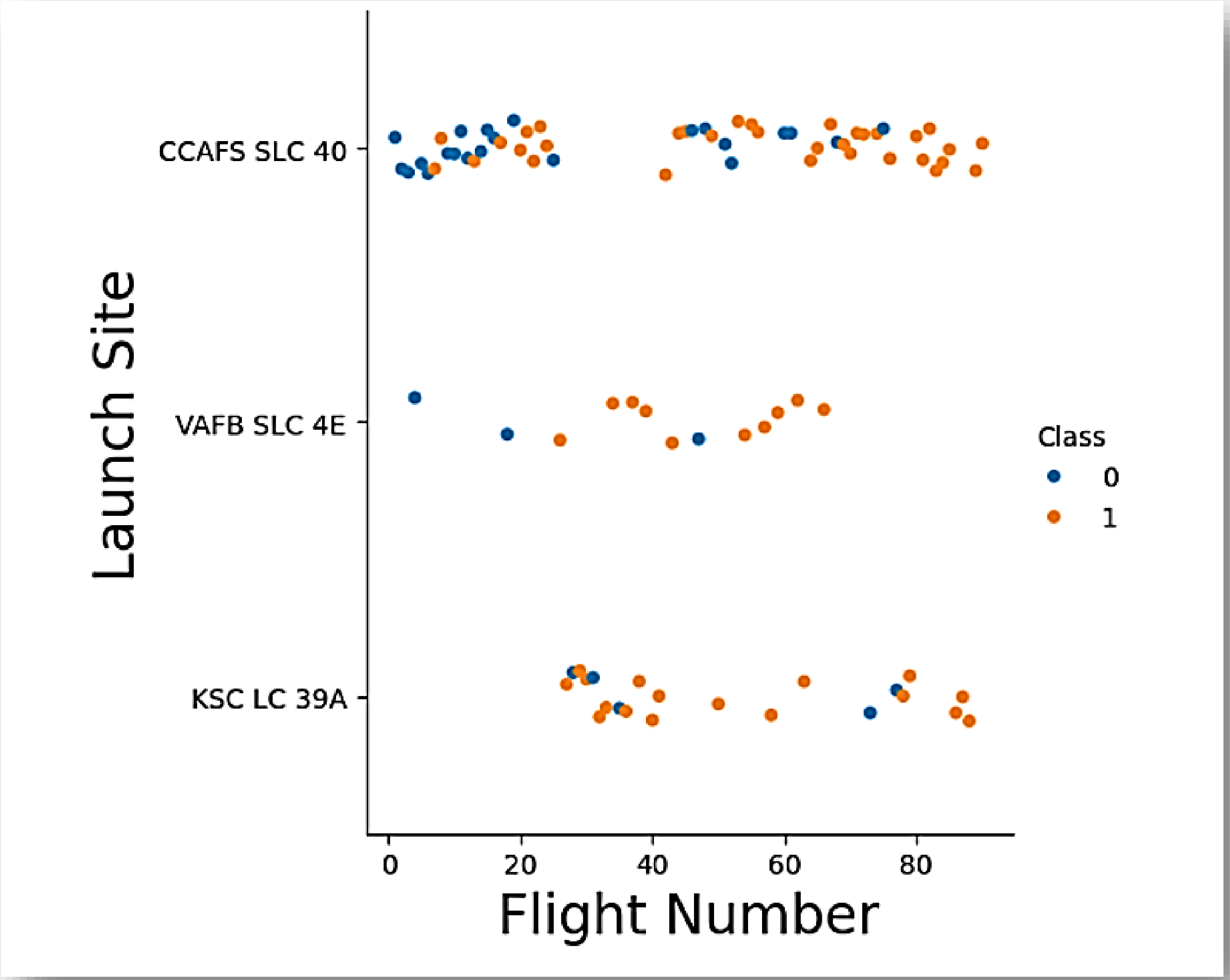
- KSC LC 39A had the most successful launches from all the sites.

- Orbit GEO, HEO, SSO, ES L1 has the best success Rate.

Section 2

Insights drawn from EDA

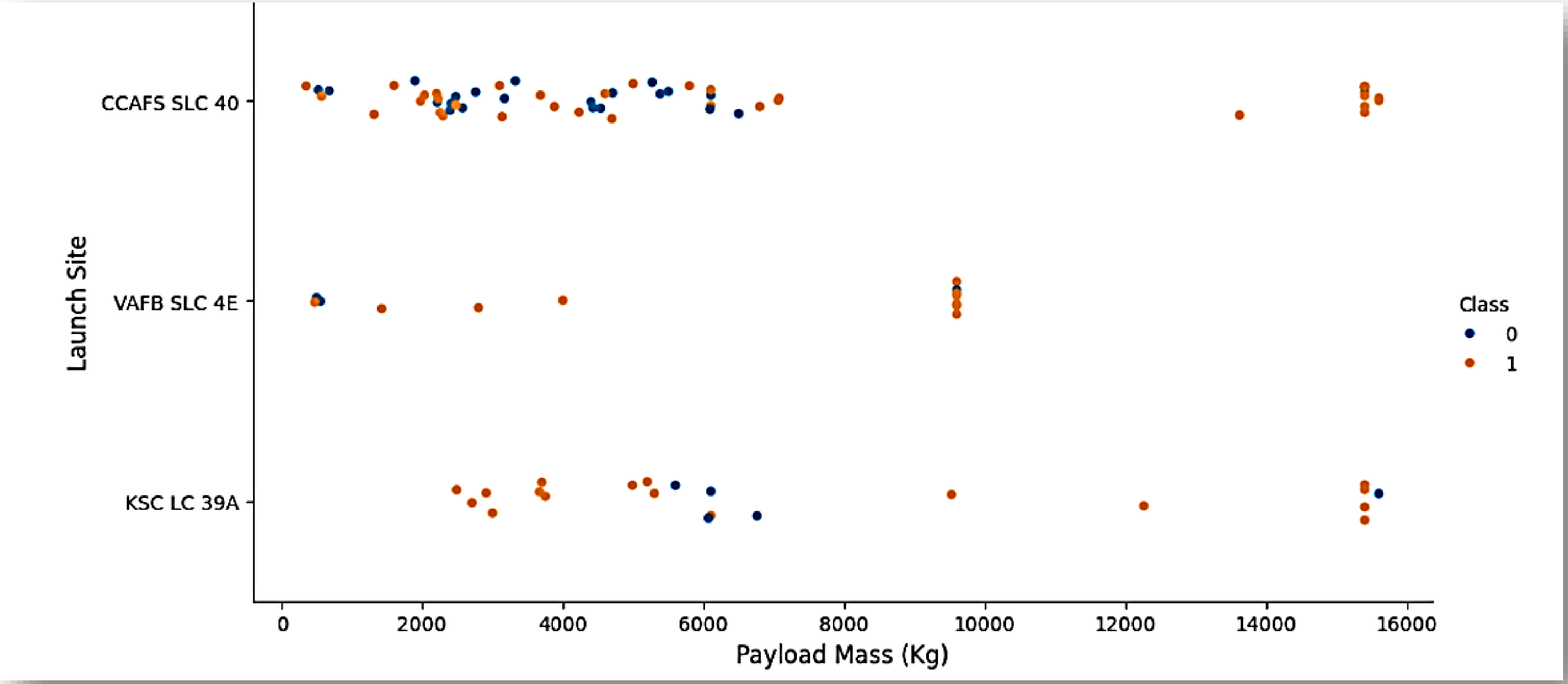
Flight Number vs. Launch Site



- Launches from the site of CCAFS 40 are significantly higher than launches from sites.

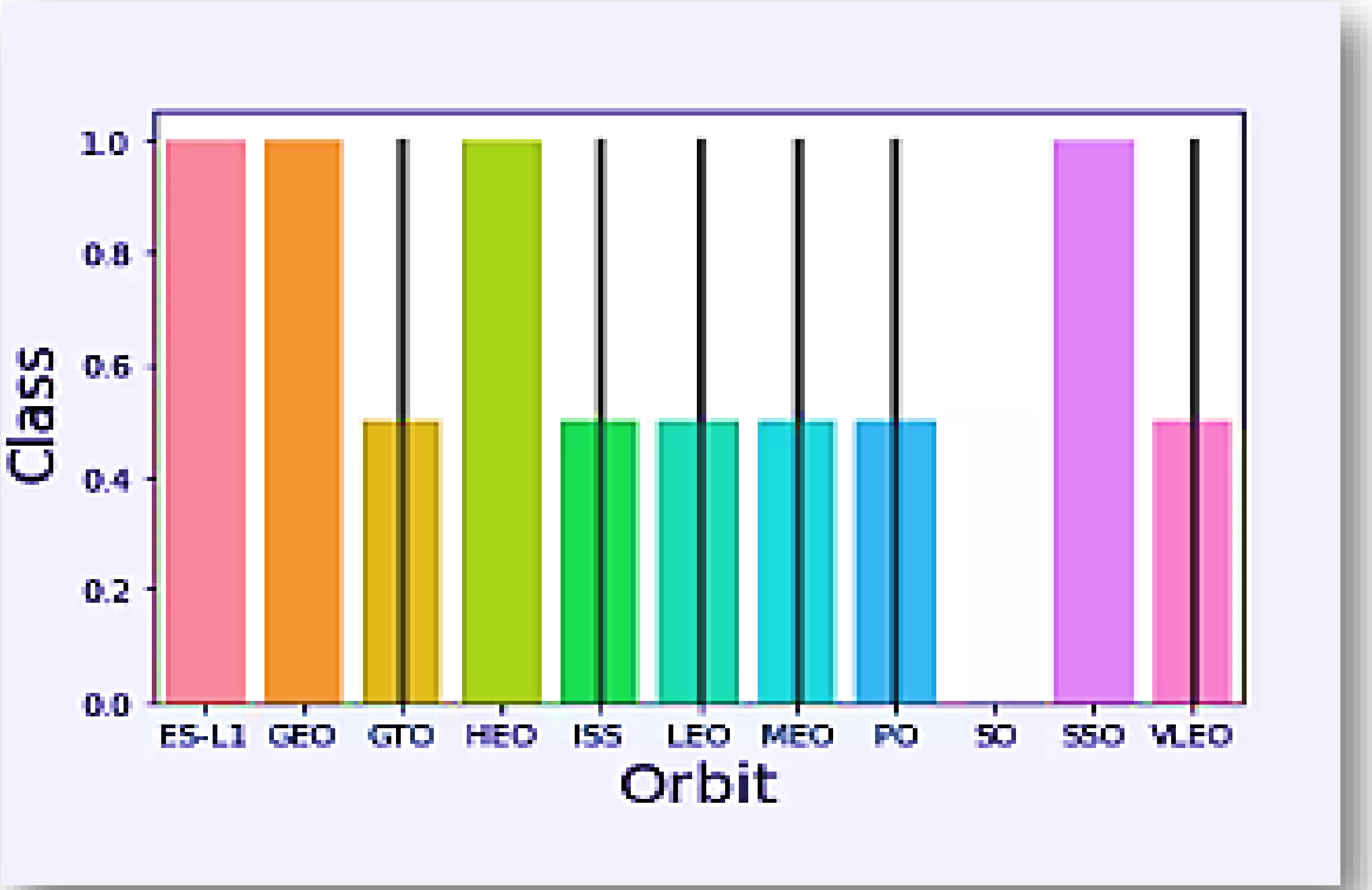
Payload vs. Launch Site

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



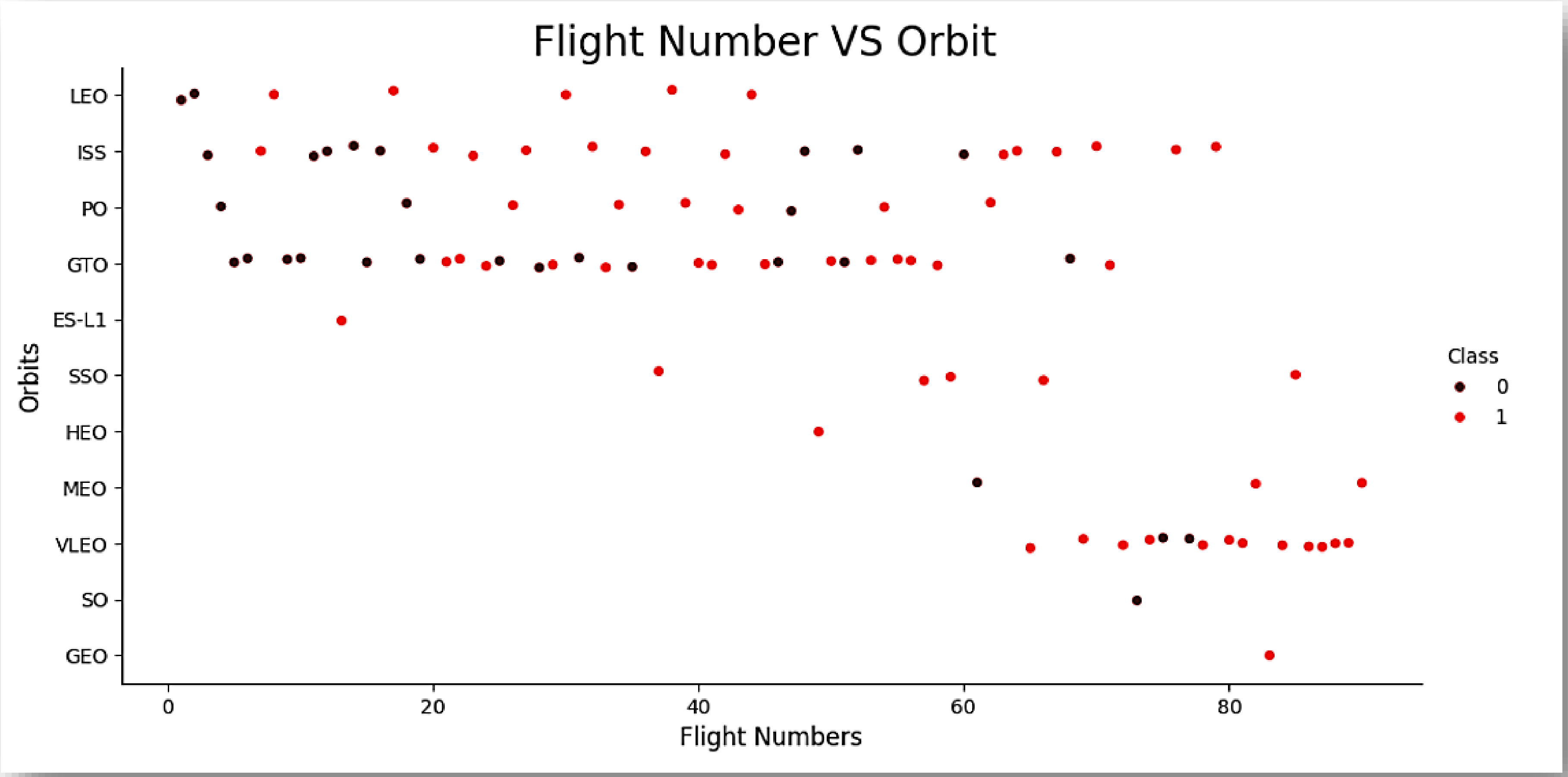
Success Rate vs. Orbit Type

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



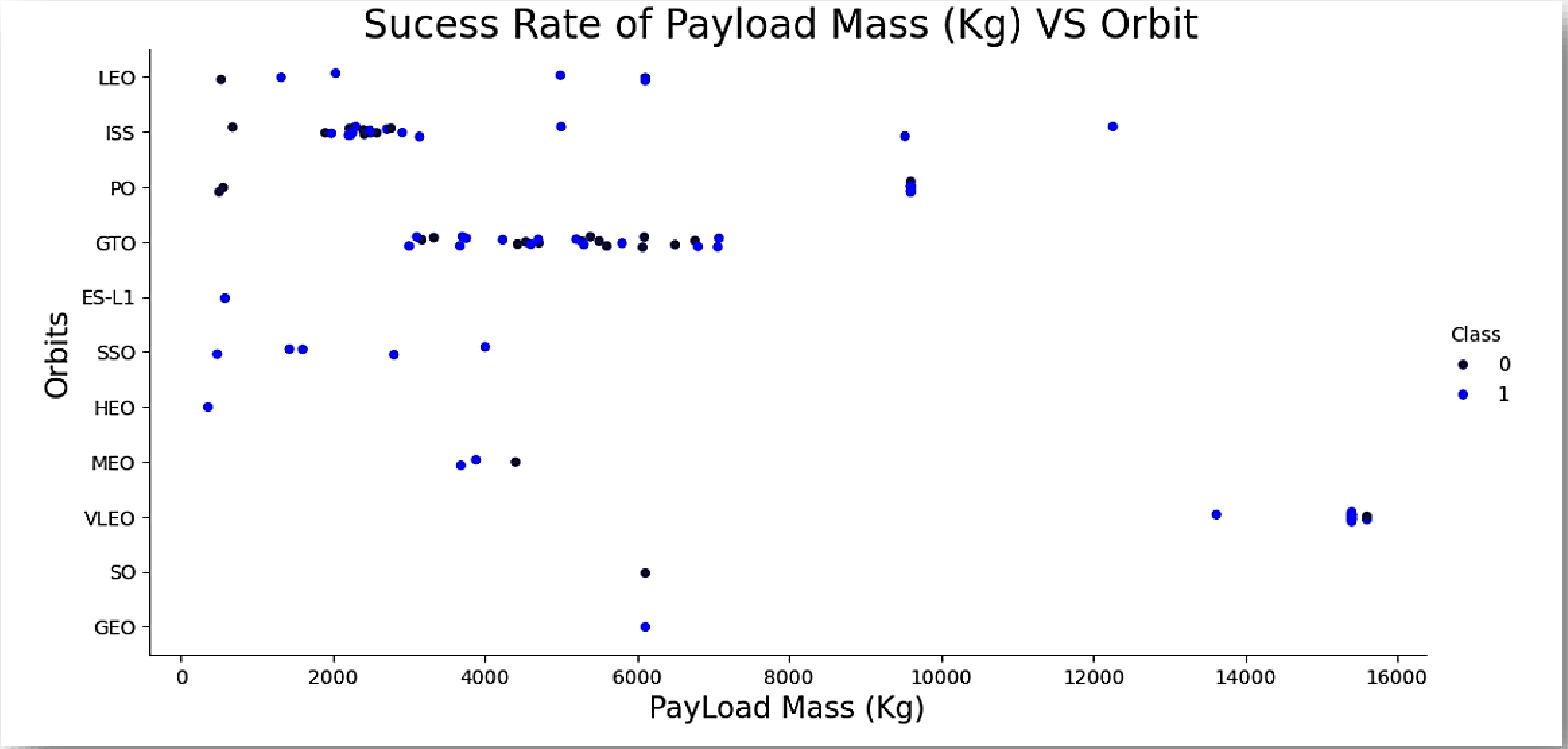
Flight Number vs. Orbit Type

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

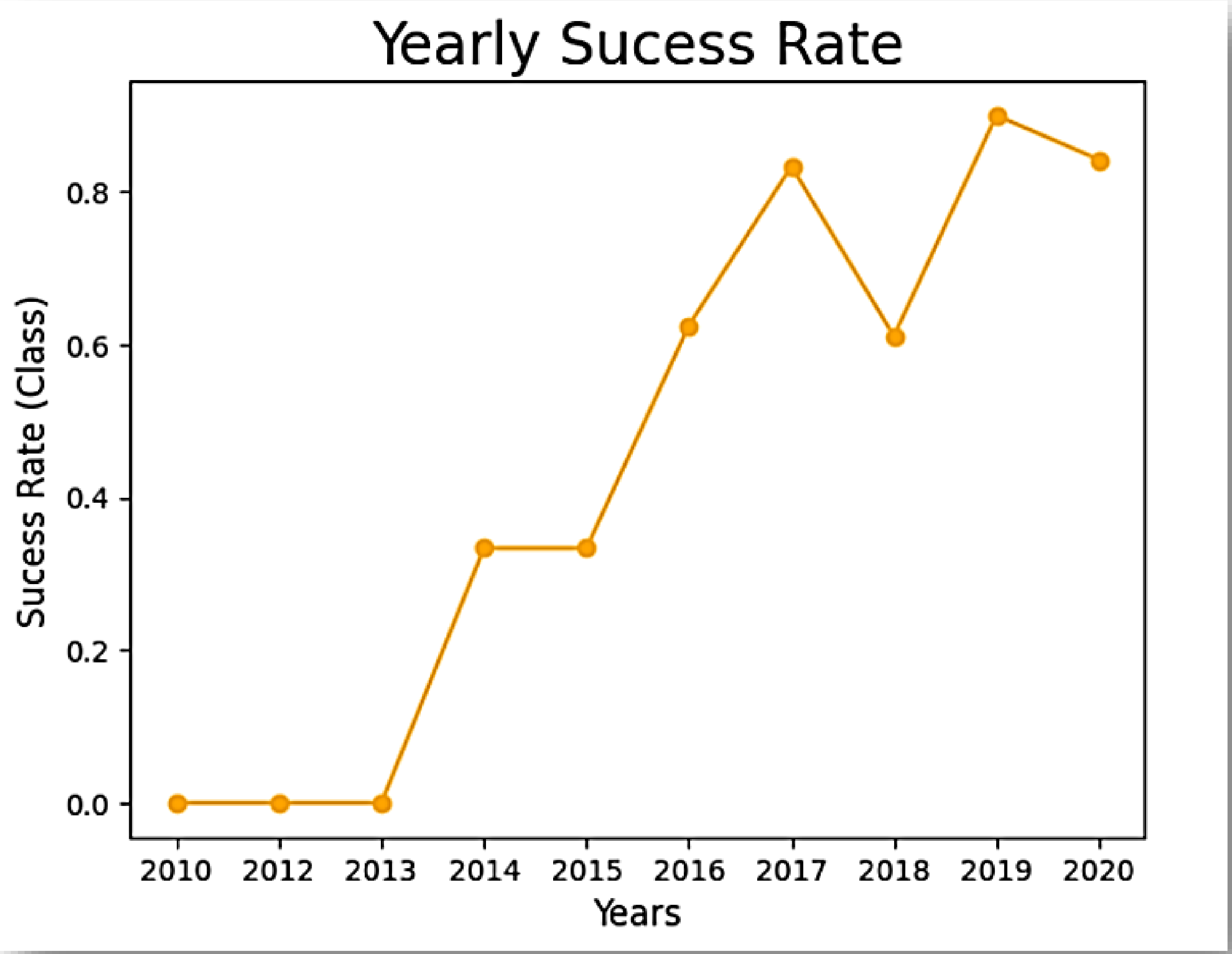


Payload vs. Orbit Type

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



Launch Success Yearly Trend



- Launch success rate has increased significantly since 2013 and has stabilized since 2019, potentially due to advance in technology 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	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) from SPACEXTBL where CUSTOMER = ‘NASA(CRS)’

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSSTER_VERSION = ‘F9 v1.1’

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

- %sql select min(DATE) from SPACEXTBL where Landing_Outcome = ‘Success(ground pad)’

First_Successful_Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql select BOOSTER_VERSION from SPACEXTBL Landinng_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 count (MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = ‘Success’ or MISSION_OUTCOME = ‘Failure(in flight)’

Mission_Outcome	Total_Number
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(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 ‘Failure%’ and (DATA between ‘2015-01-01’ and ‘2015-12-31’) order by data desc

Month_Name	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 * from SPACEXTBL where Landing_Outcome like ‘Success%’ and (DATA between ‘2010-06-04’ and ‘2017-03-20’) order by data desc

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

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

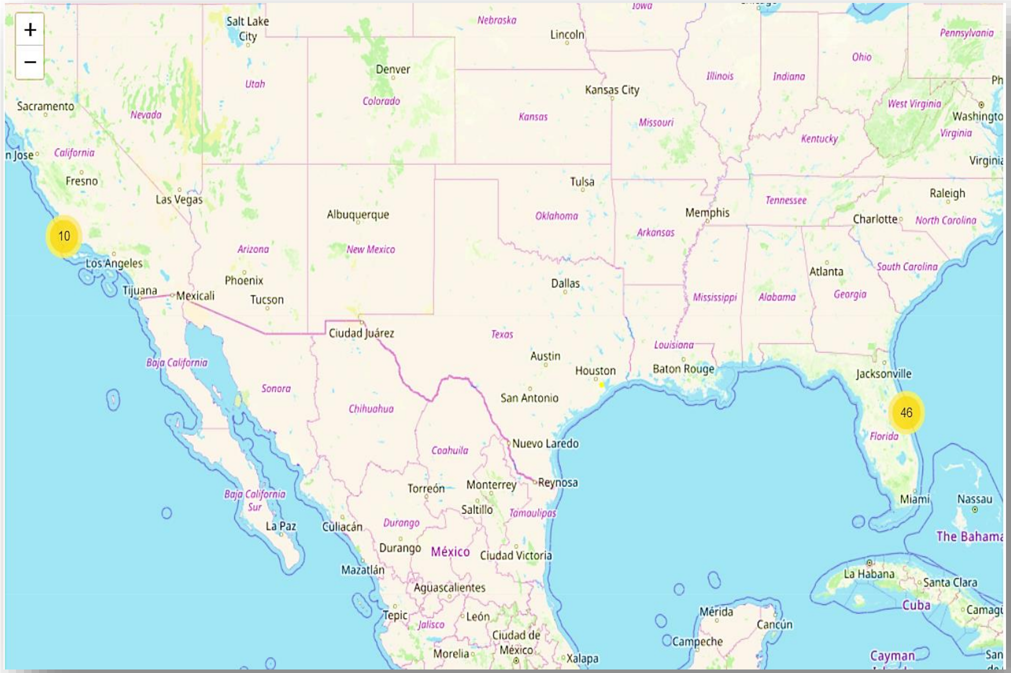
Launch Sites Proximities Analysis

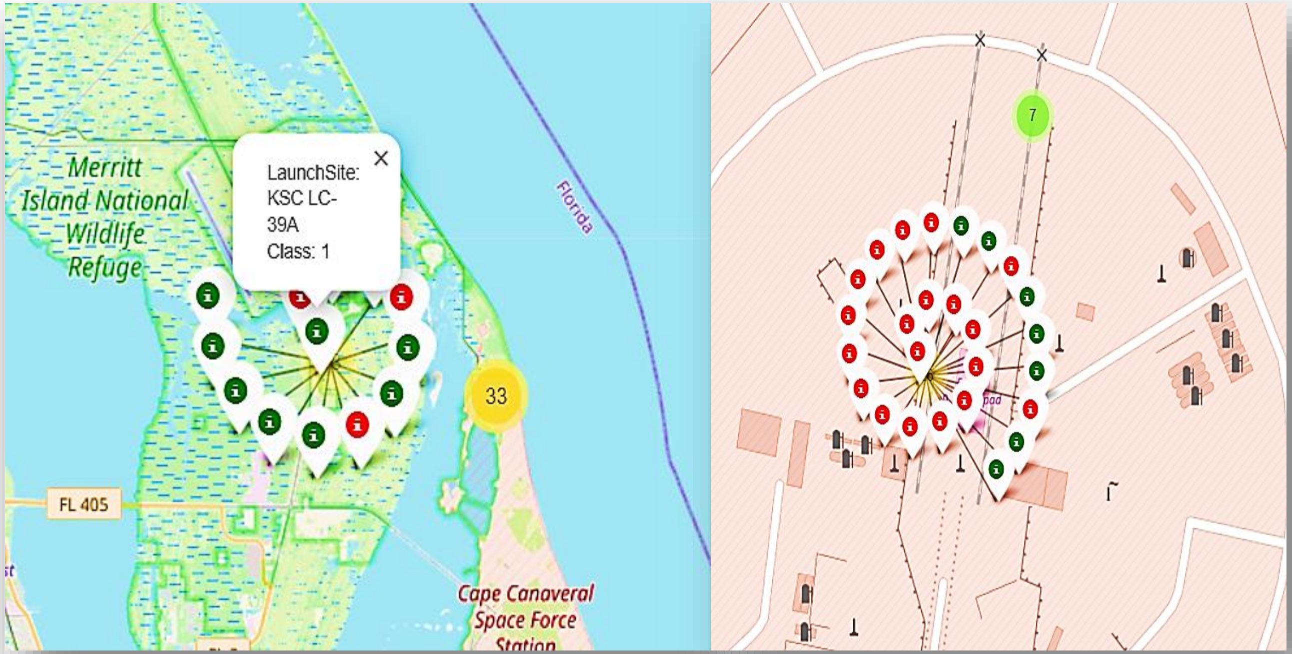
Global Map of SpaceX Launch Sites

- All launch Site marked on a map

Success/Failed Launches For Each Site

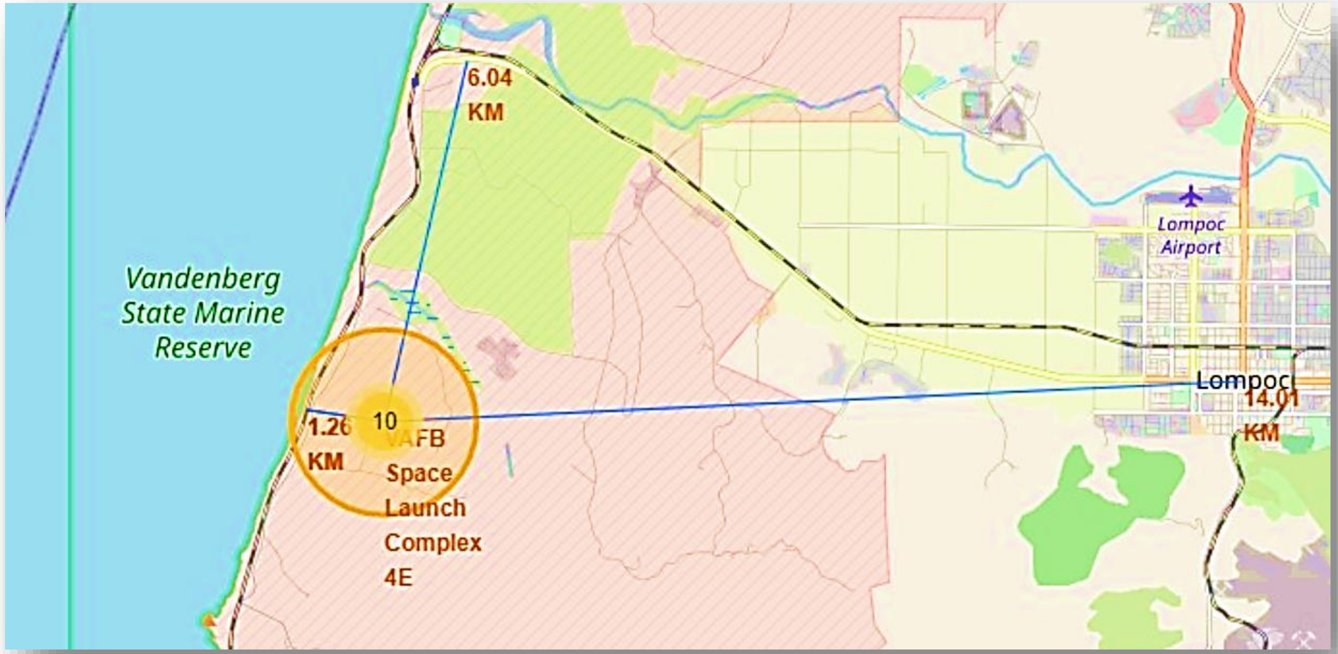
- The first map shows clusters of every launch site, the second show a green marker if a launch was successful and a red if a launch was failed.

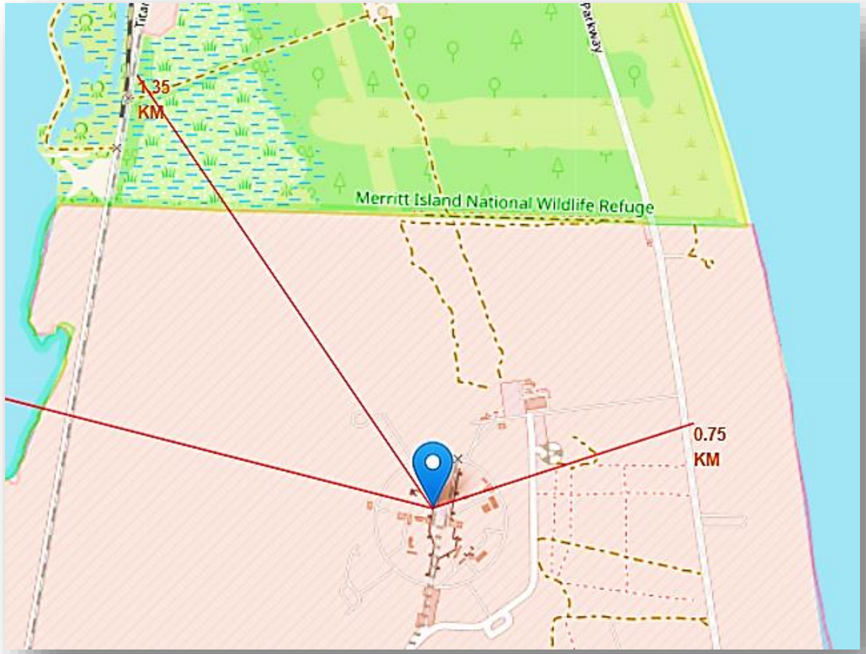




Distance between a launch site to its proximities

- Launch sites are near to railway, roads, highways and coastline. For maintain a safe distance with near cities.



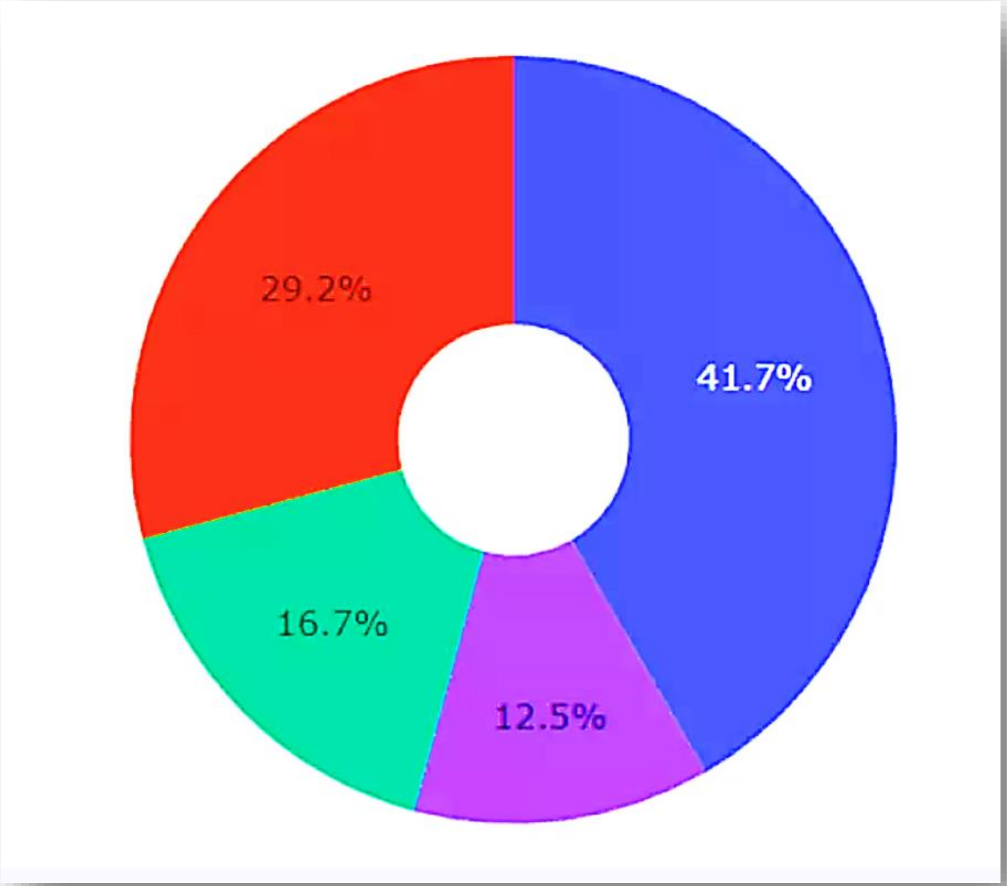


Section 4

Build a Dashboard with Plotly Dash

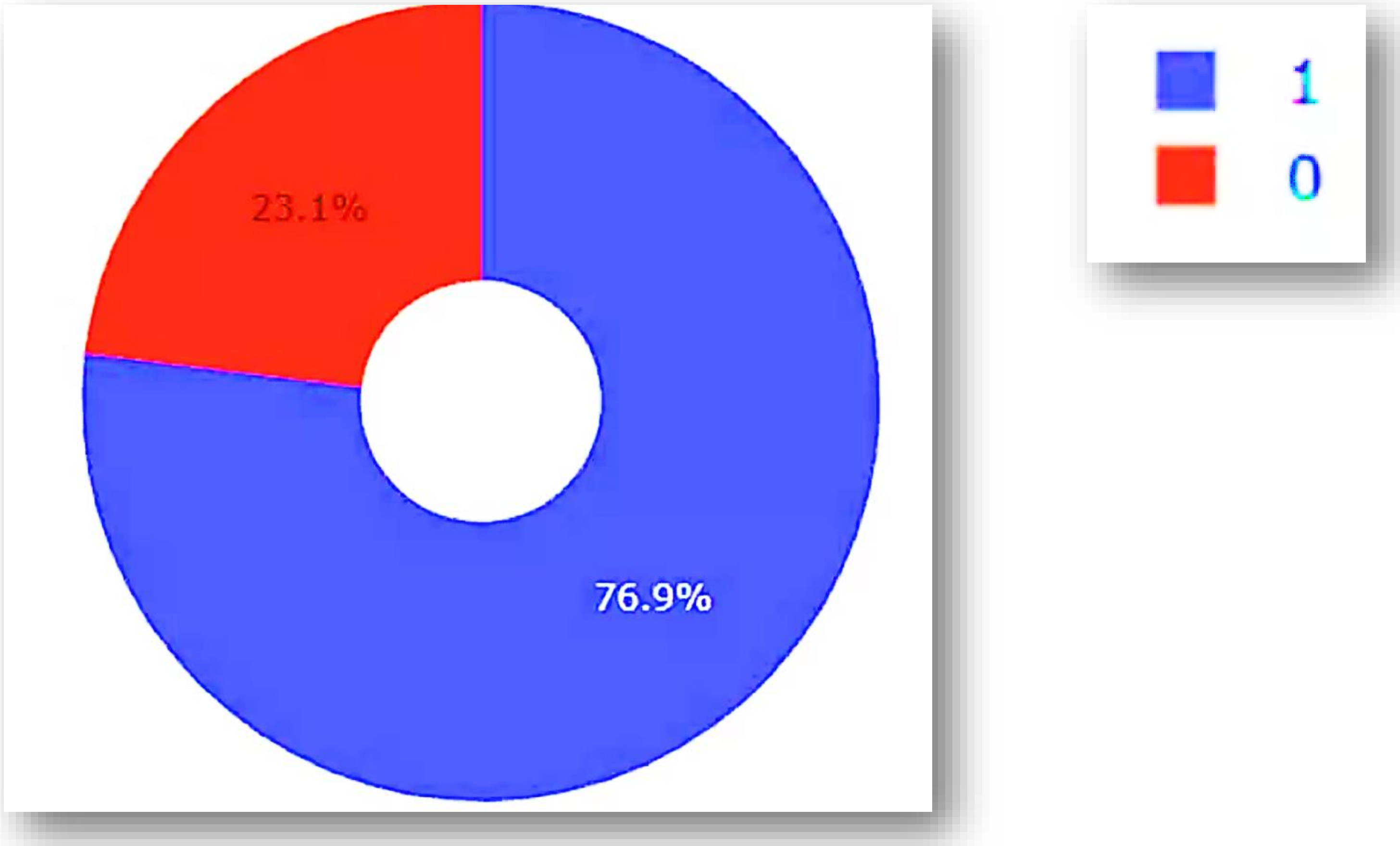
Total Success launches by all sites

- Total Success Launches



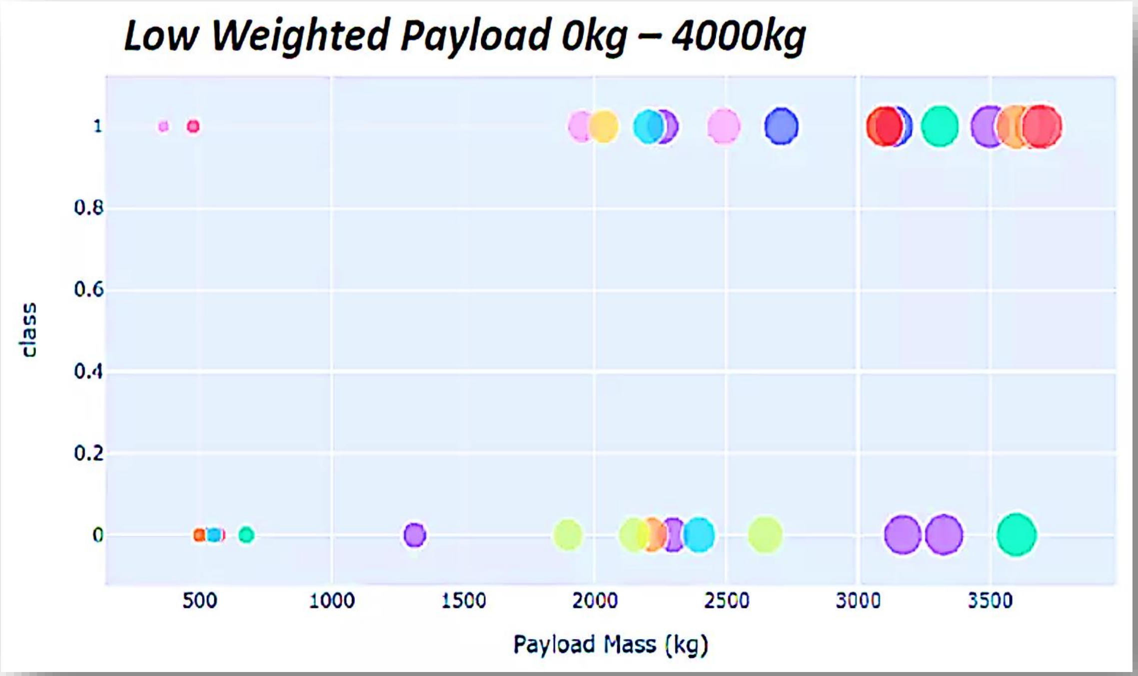
We can see that KSC LC-39A had the most successful launches from all the sites

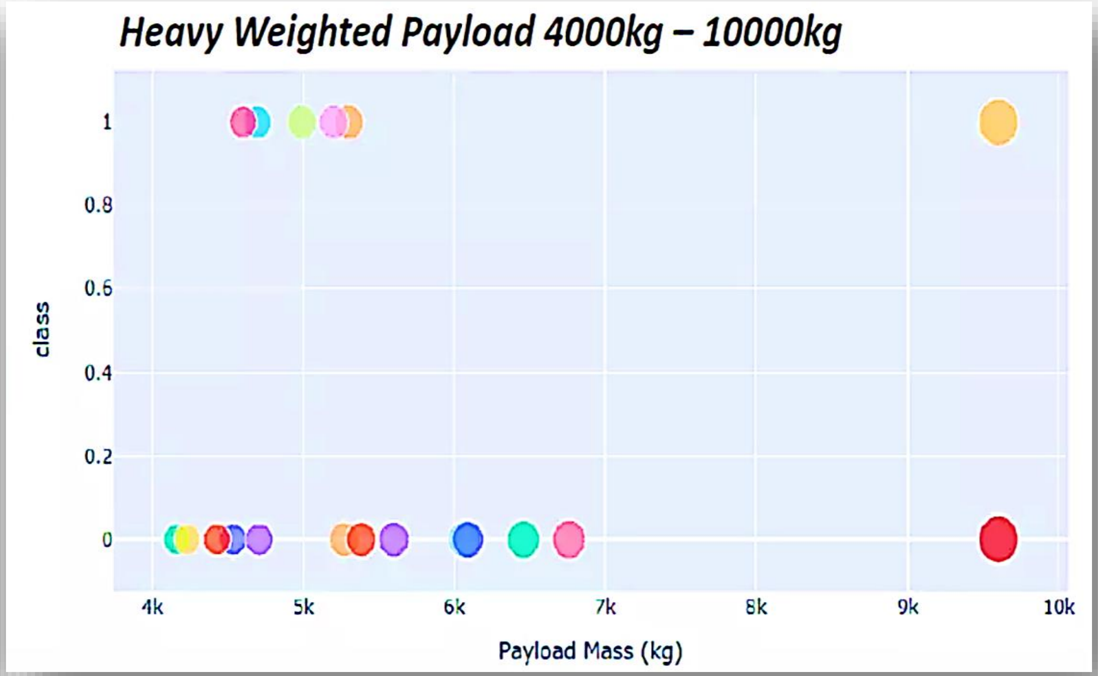
Success rate by site



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

Payload vs launch outcome



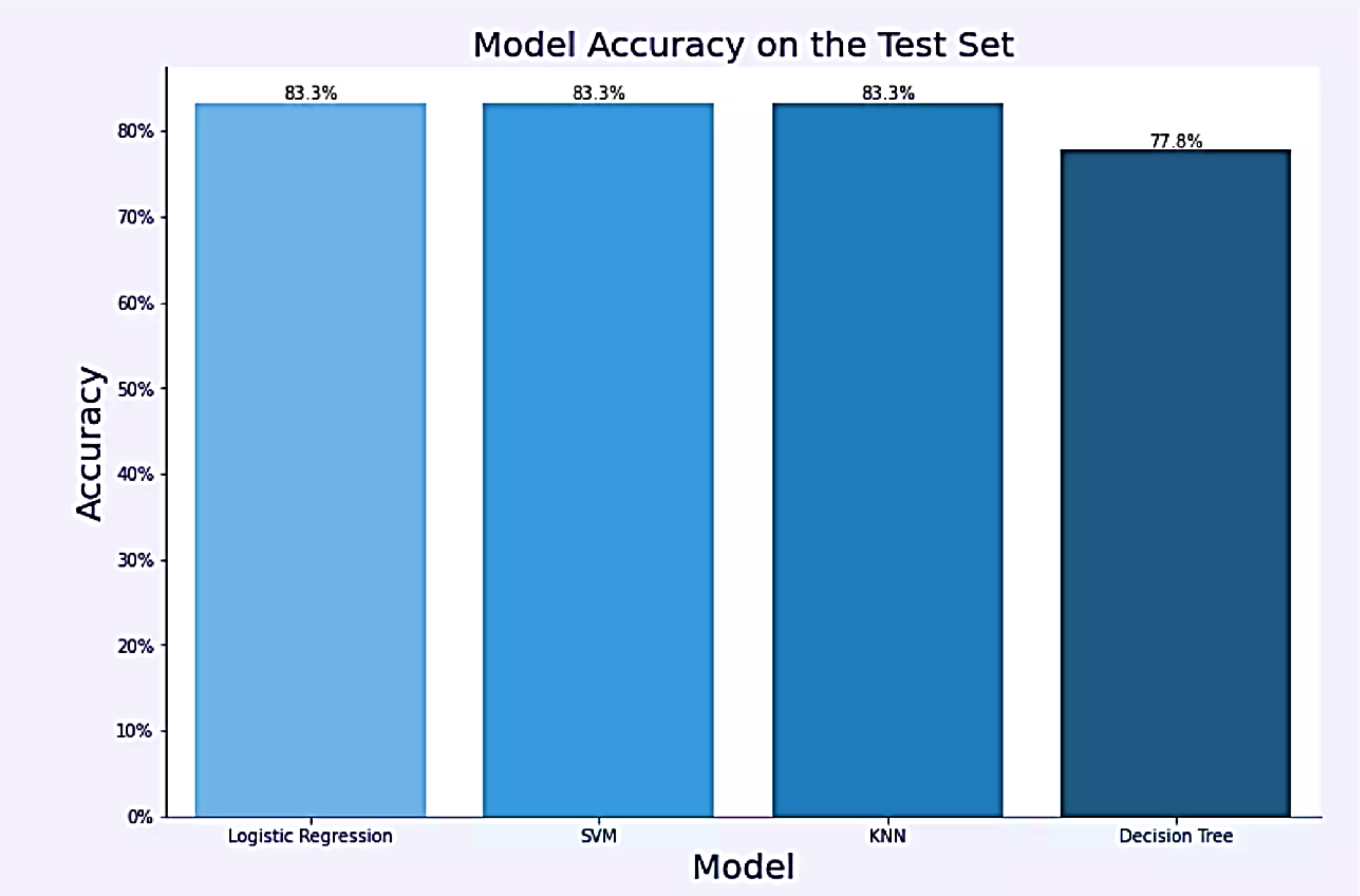


We can see the success rate for low weighted payloads is higher than the heavy weighted payloads.

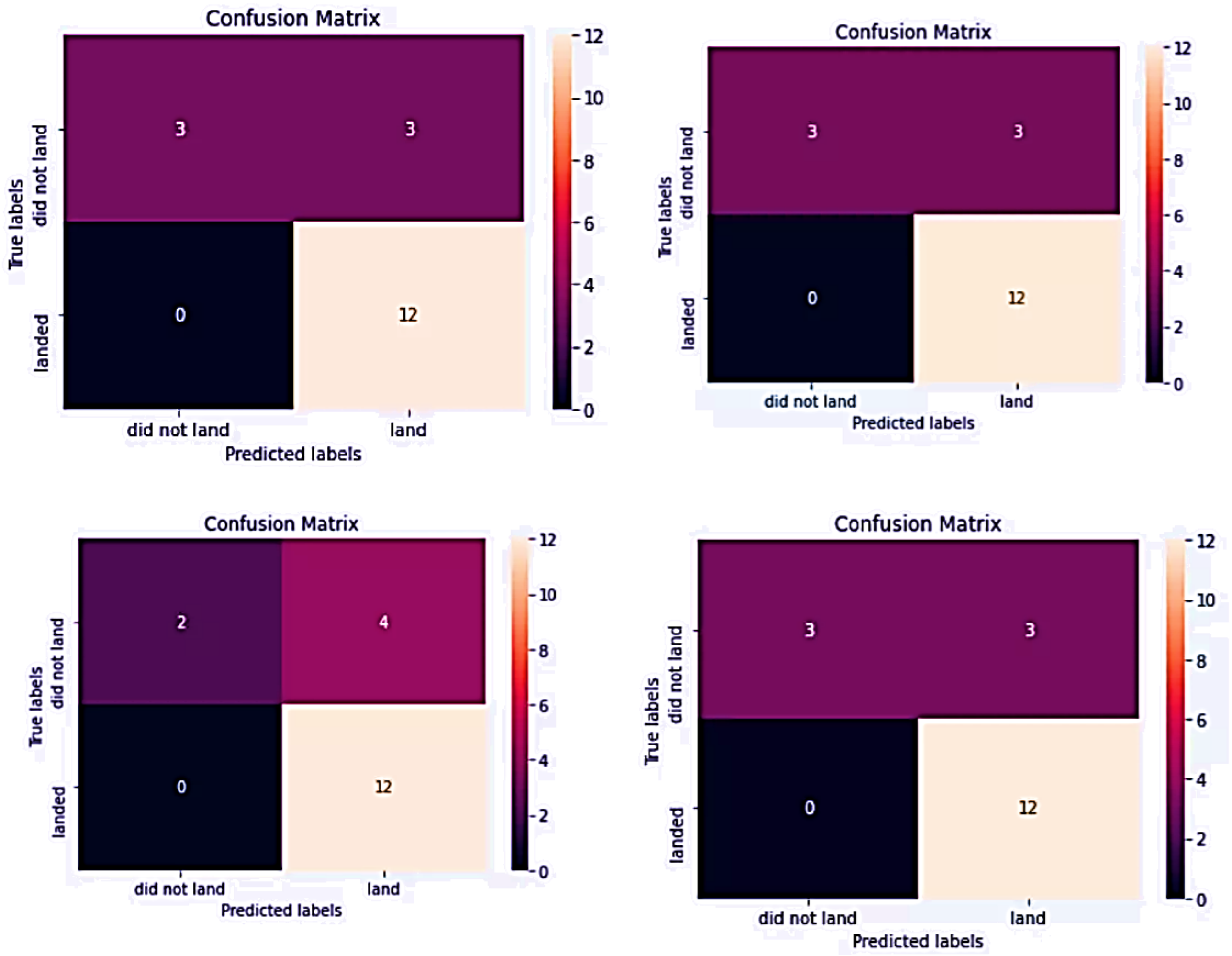
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

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

Appendix

- For notebooks, dataset and scripts, follow this GitHub repository link:

https://github.com/ShanAyush/IBM-Data-Science-Capstone