

Winning Space Race with Data Science

Nilesh Bharwad
January 24th 2025



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 advertise Falcon 9 rocket launches on the website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the saving is due to reusable rockets in the first stage of launch.
- Problems you want to find answers
 - The project task is to predict whether the first stage of each launch of Falcon 9 rocket will land successfully or not.

Section 1

Methodology

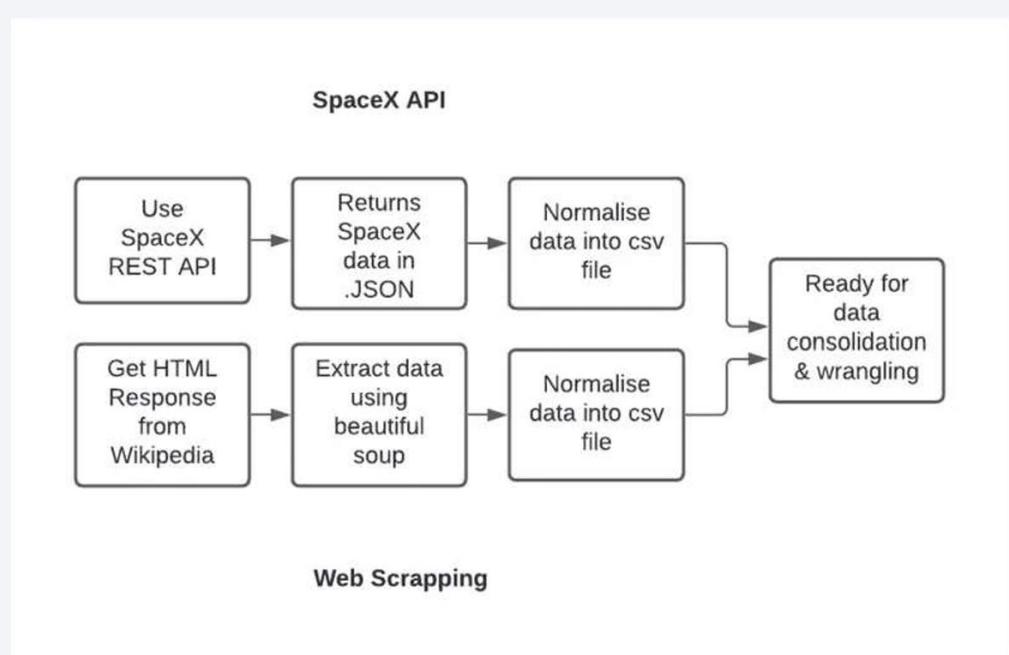
Methodology

Executive Summary

- Data collection methodology:
 - SpaceX Rest API
 - Web Scrapping from Wikipedia
- Perform data wrangling
 - One Hot Encoding data fields for Machine Learning data cleaning 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 and 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 specifications, landing specifications, and landing outcomes.
 - The SpaceX REST API endpoints, or URLs, starts with api.spacexdata.com/v4/.
 - Another popular data source for obtaining Falcon 9 launch data is web scrapping from www.Wikipedia.com using BeautifulSoup method.



Data Collection – SpaceX API

- Data Collection with SpaceX Rest API calls.

- [nbharwad/IBM Data Science Certificate: IBM Data Science Certificate.](#)

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)

4. Assign list to a variable

```
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 Collection - Scraping

- Web Scrapping from Wikipedia
- [IBM Data Science Certificate/jupyter-labs-webscraping.ipynb at main · nbharwad/IBM Data Science Certificate](#)

1 .Getting Response from HTML

```
page = requests.get(static_url)
```

2. Creating BeautifulSoup Object

```
soup = BeautifulSoup(page.text, 'html.parser')
```

3. Finding tables

```
html_tables = soup.find_all('table')
```

4. Getting column names

```
column_names = []
temp = soup.find_all('th')
for x in range(len(temp)):
    try:
        name = extract_column_from_header(temp[x])
        if (name is not None and len(name) > 0):
            column_names.append(name)
    except:
        pass
```

6. Appending data to keys (refer) to notebook block 12

```
In [12]: extracted_row = 0
#Extract each table
for table_number,table in enumerate(
    # get table row
    for rows in table.find_all("tr"):
        #check to see if first table
```

8. Dataframe to .CSV

```
df.to_csv('spacex_web_scraped.csv', index=False)
```

5. Creation of dictionary

```
launch_dict= dict.fromkeys(column_names)

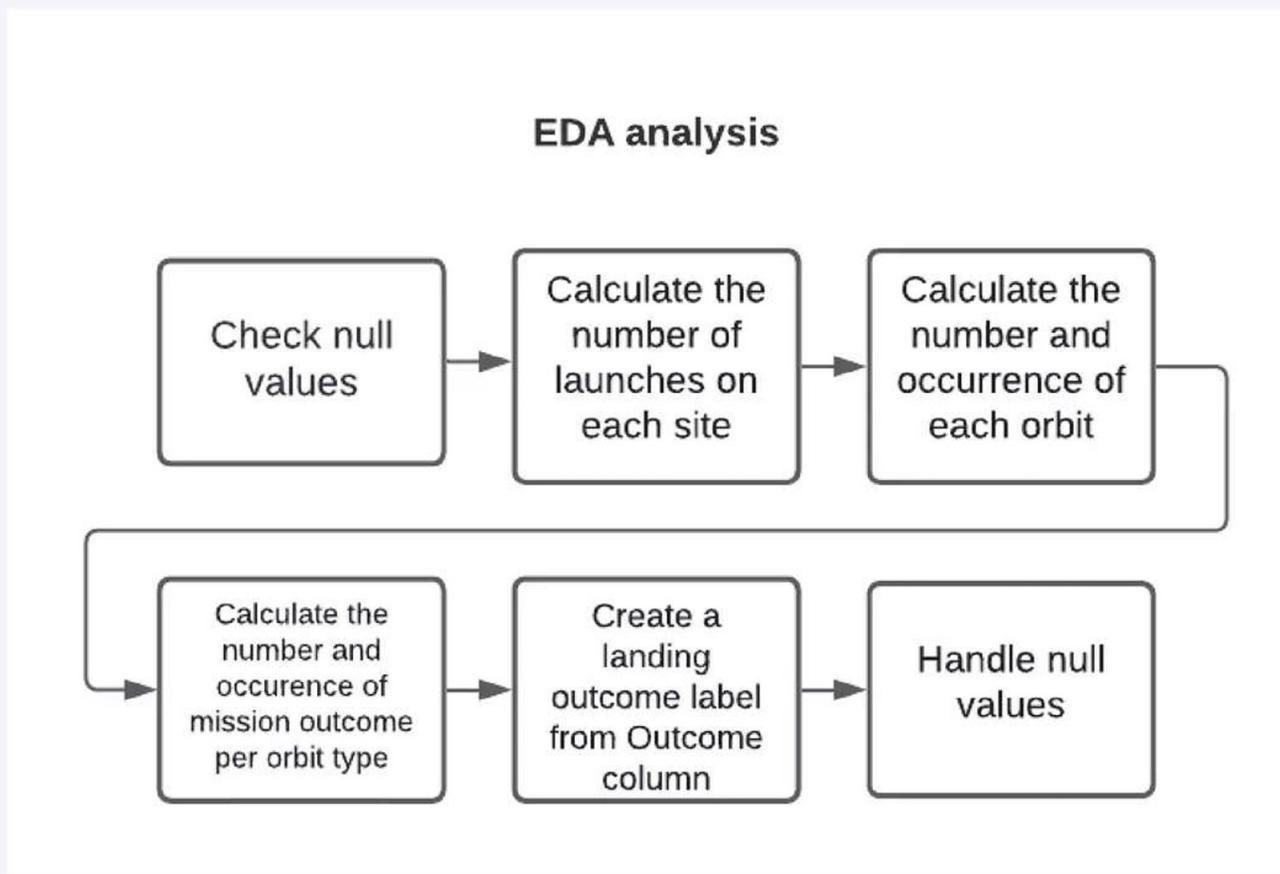
# Remove an irrelevant column
del launch_dict['Date and time ( )']

launch_dict['Flight No.']= []
launch_dict['Launch site']= []
launch_dict['Payload']= []
launch_dict['Payload mass']= []
launch_dict['Orbit']= []
launch_dict['Customer']= []
launch_dict['Launch outcome']= []
launch_dict['Version Booster']= []
launch_dict['Booster landing']= []
launch_dict['Date']= []
launch_dict['Time']= []
```

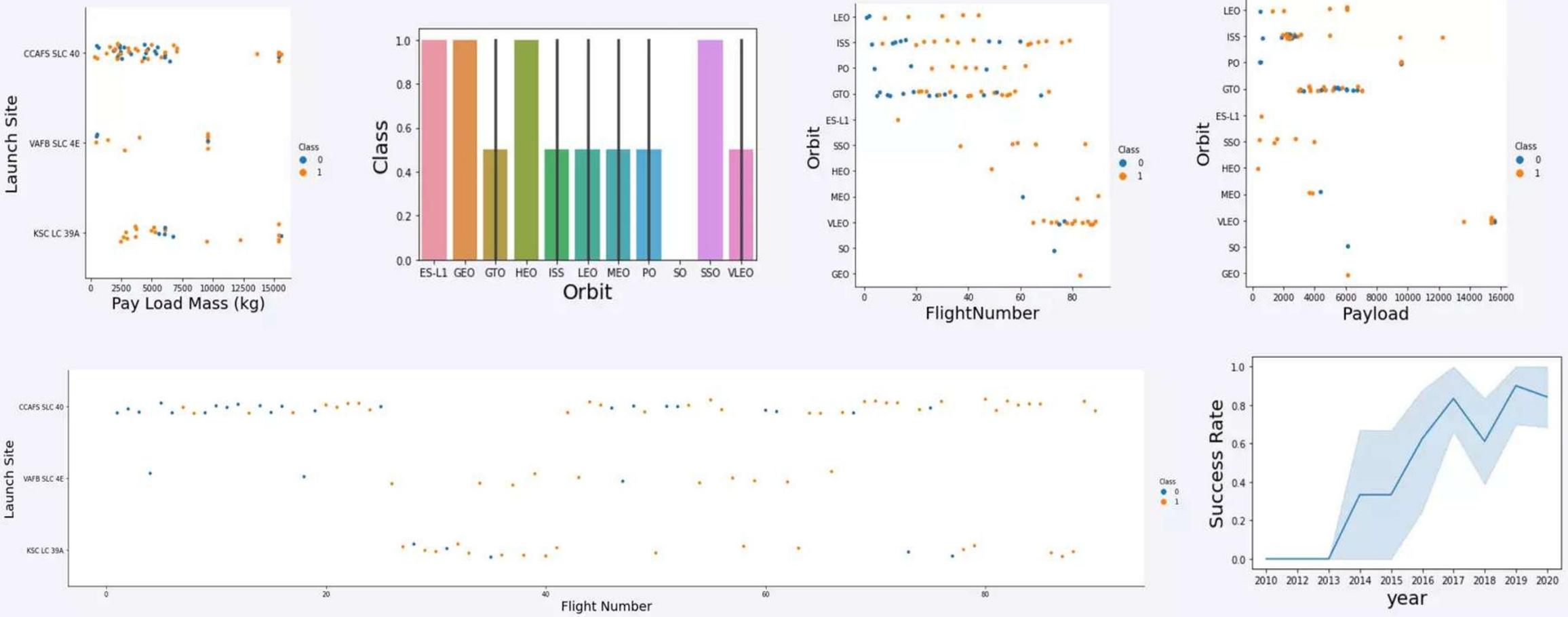
7. Converting dictionary to dataframe

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

Data Wrangling



EDA with Data Visualization



- [nbharwad/IBM Data Science Certificate: IBM Data Science Certificate.](#)

EDA with 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 between 4,000 and 6,000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster_versions which have carried the max. payload mass.
- Listing the records which will display the month names, successful landing outcomes in ground pad, booster versions, lannch site for the months in year 2017
- Ranking the count of successful landing outcomes between the dates 2010 and 2017.
- Link: [IBM Data Science Certificate/jupyter-labs-eda-sql-coursera_sqlite.ipynb at main · nbharwad/IBM Data Science Certificate](#)

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.

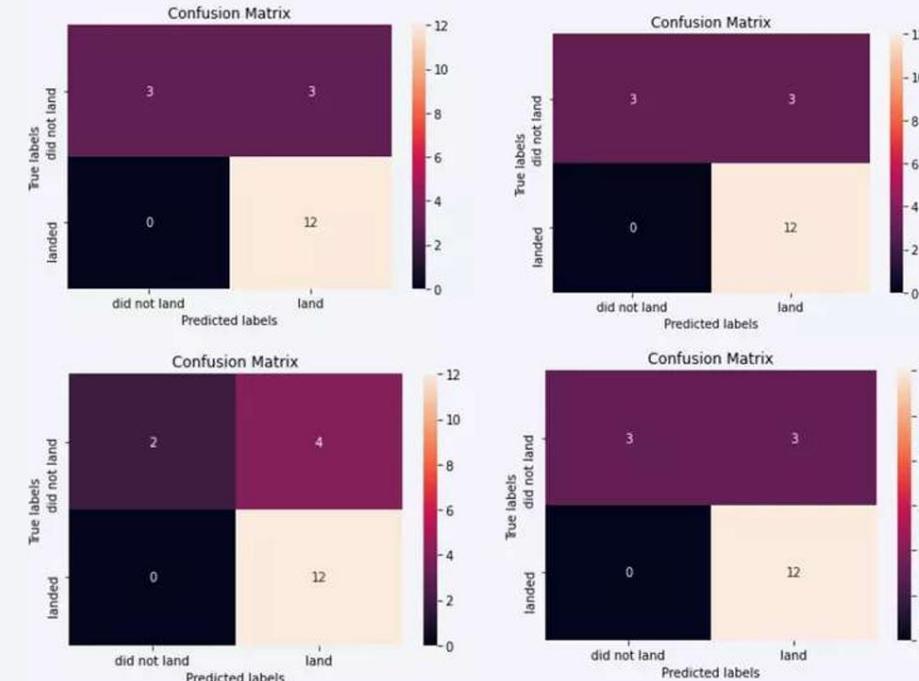
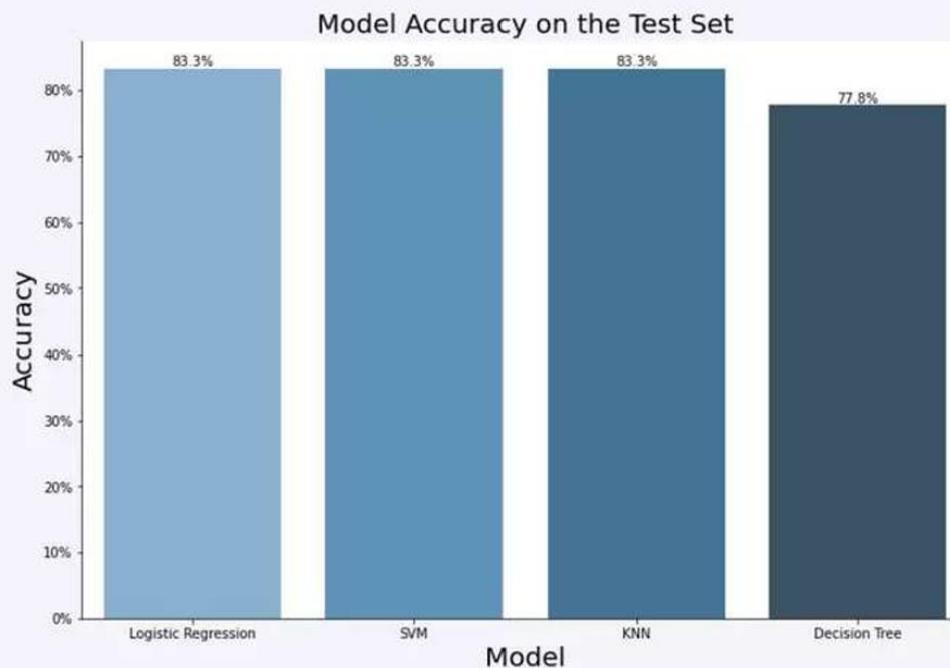
[IBM Data Science Certificate/Interactive Visual Analytics with Folium.ipynb at main · nbharwad/IBM Data Science Certificate](#)

Build a Dashboard with Plotly Dash



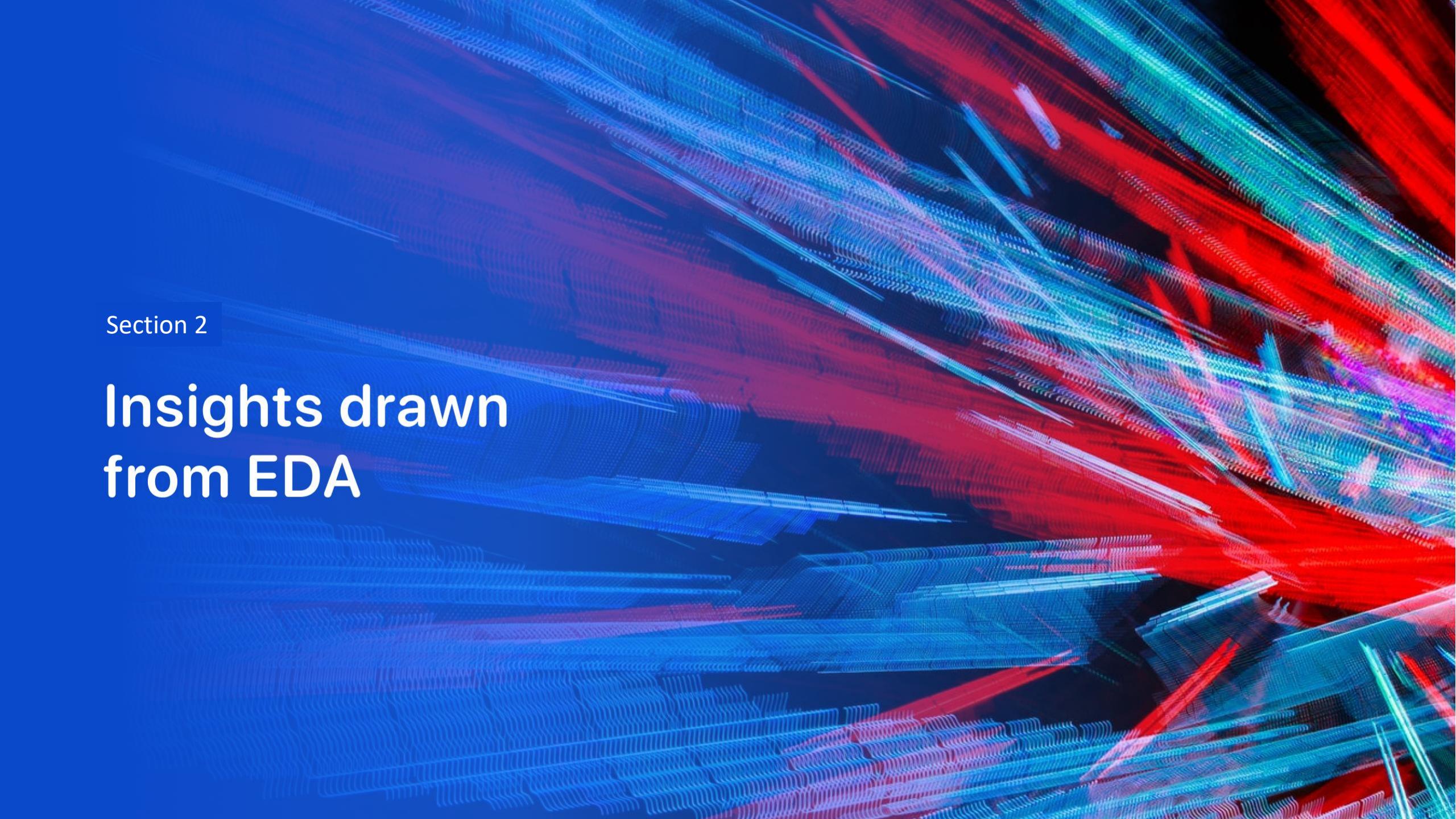
Predictive Analysis (Classification)

- The SVM, KNN, and LR model achieved the highest accuracy i.e. 83.3%, while the SVM performs the best in terms of Area Under the curve at 0.958



Results

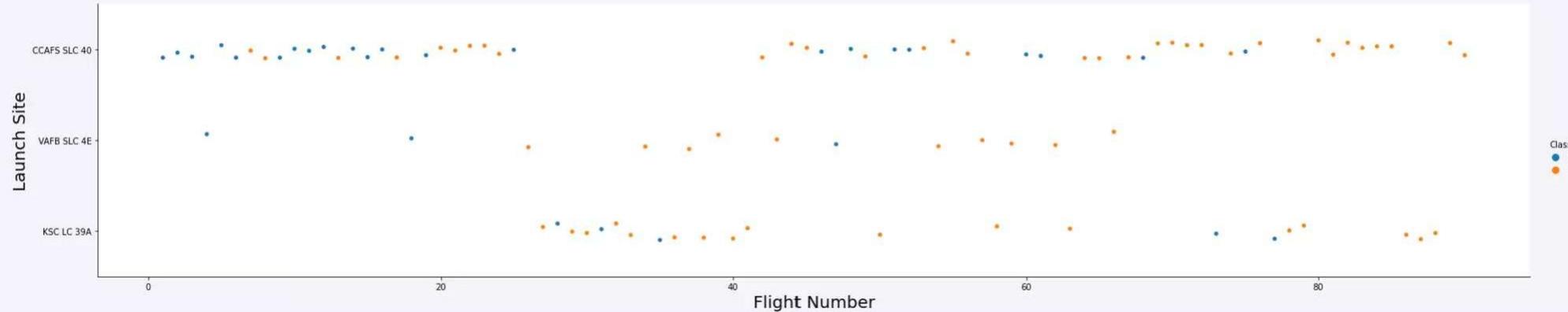
- The SVM, KNN and LR models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads performs better than the heavier payloads
- The success rates for SPACEX launches is directly proportional time in years they will eventually perform the launches. KSC LC 39A had the most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

The background of the slide features a complex, abstract digital visualization. It consists of numerous thin, glowing lines that create a sense of depth and motion. The lines are primarily blue and red, with some green and purple highlights. They form a grid-like structure that curves and twists across the frame, resembling a three-dimensional space or a network of data points. The overall effect is futuristic and dynamic.

Section 2

Insights drawn from EDA

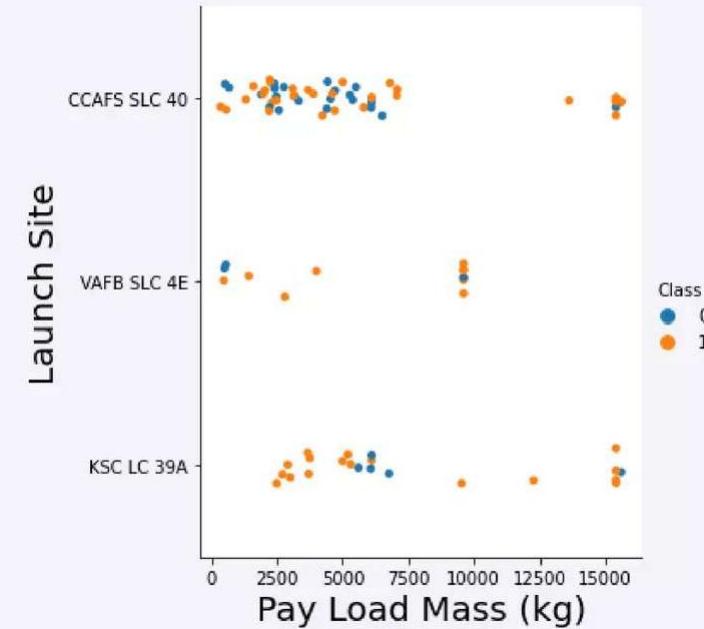
Flight Number vs. Launch Site



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

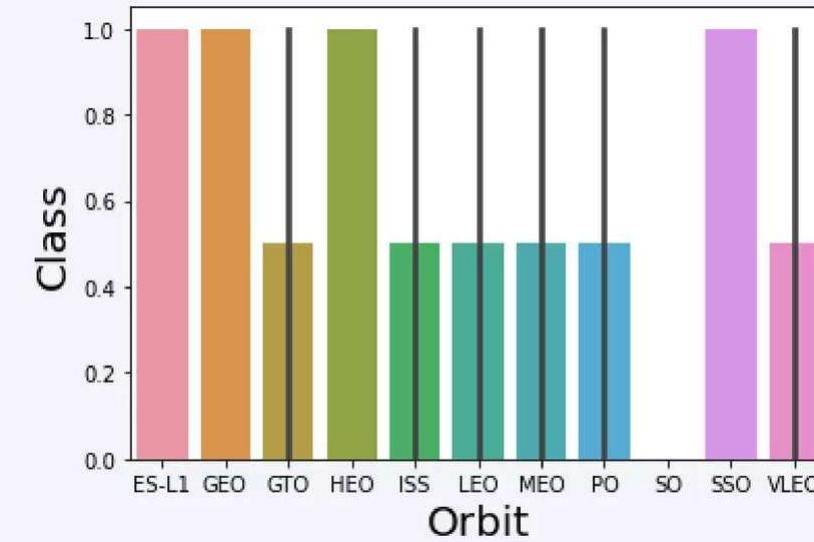
Payload vs. Launch Site

- The Majority of payloads with lower Mass have been launched from CCAFS SLC 40.



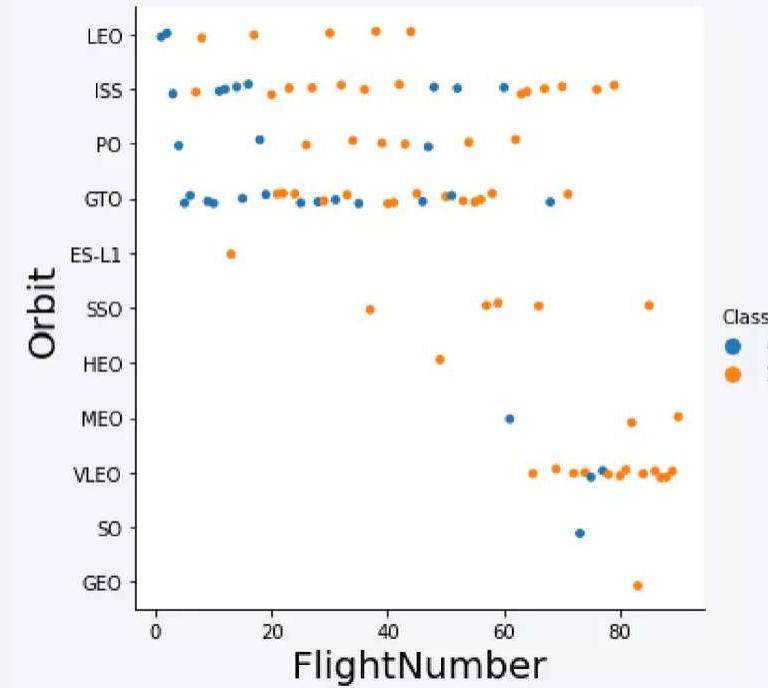
Success Rate vs. Orbit Type

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



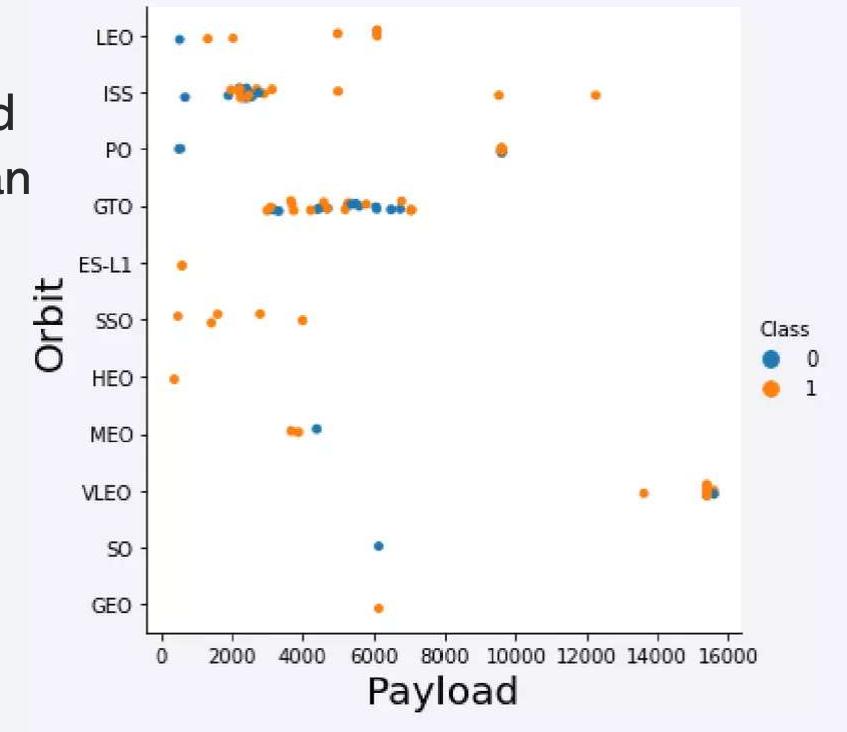
Flight Number vs. Orbit Type

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



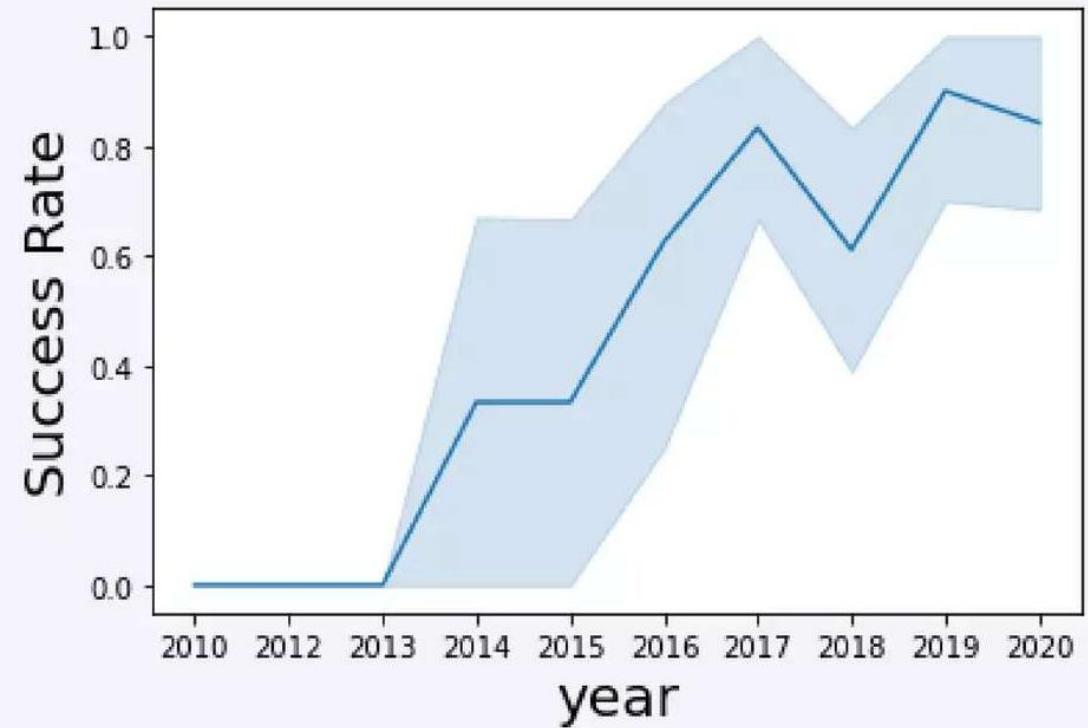
Payload vs. Orbit Type

- There are strong correlation between ISS and payload at the range around 2000, as well as between GTO and the range between 4,000 and 8,000.



Launch Success Yearly Trend

- Launch success rate has increased significantly since 2013 and has stabilised since 2010 and potentially due to advance in technology and lessons learned.



All Launch Site Names

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

- %sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5

Total Payload Mass

- %sql select sum(PAYLOAD_MASS_KG) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'



45596

A white rectangular box containing the number 45596. Each digit is rendered in a different color: 4 is blue, 5 is red, 5 is green, 9 is orange, and 6 is purple.

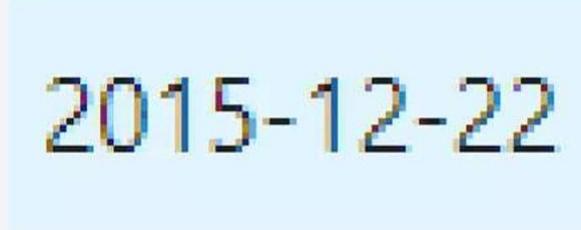
Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS_KG) from SPACEXTBL where BOOSTER_VERSION = 'F9V1,1'

```
2928.400000
```

First Successful Ground Landing Date

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



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 count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'



100

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 '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)
05:21:00	F9 FT B1022	CCAFS LC-	JCSAT-14	4000	GTO	SKY Perfect JSAT	Success	Success (drone ship)

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

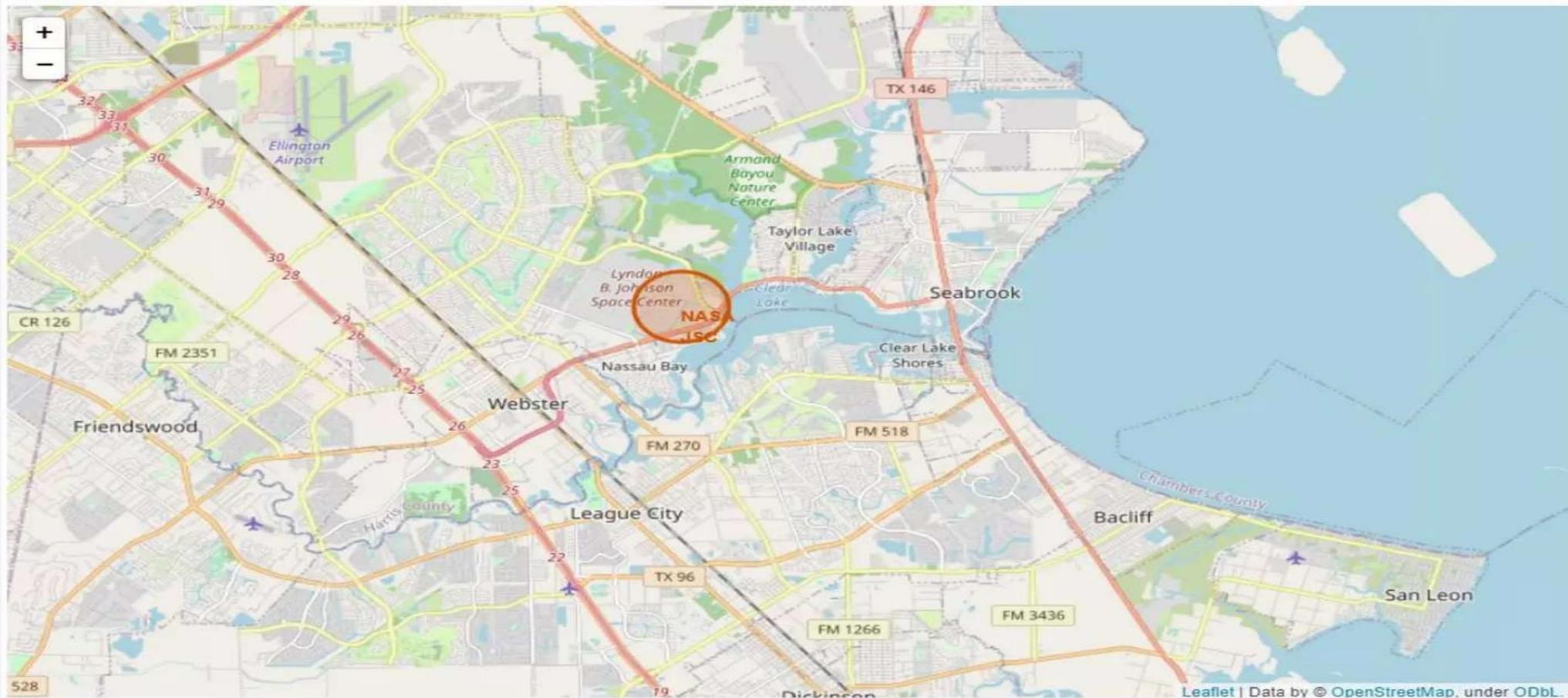
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	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)

The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth against the dark void of space. City lights are visible as numerous small white and yellow dots, primarily concentrated in the lower right quadrant where the United States appears. In the upper left quadrant, the green and blue glow of the aurora borealis is visible in the upper atmosphere.

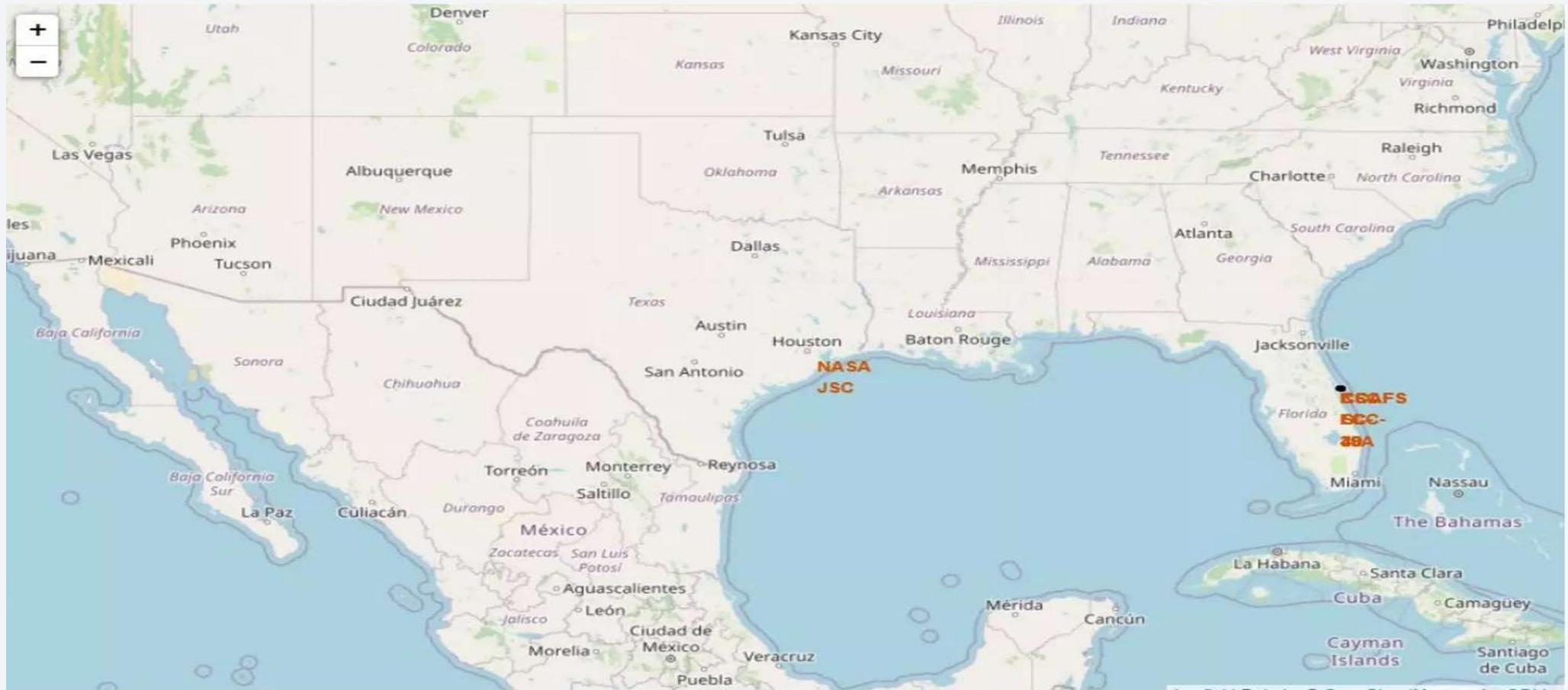
Section 3

Launch Sites Proximities Analysis

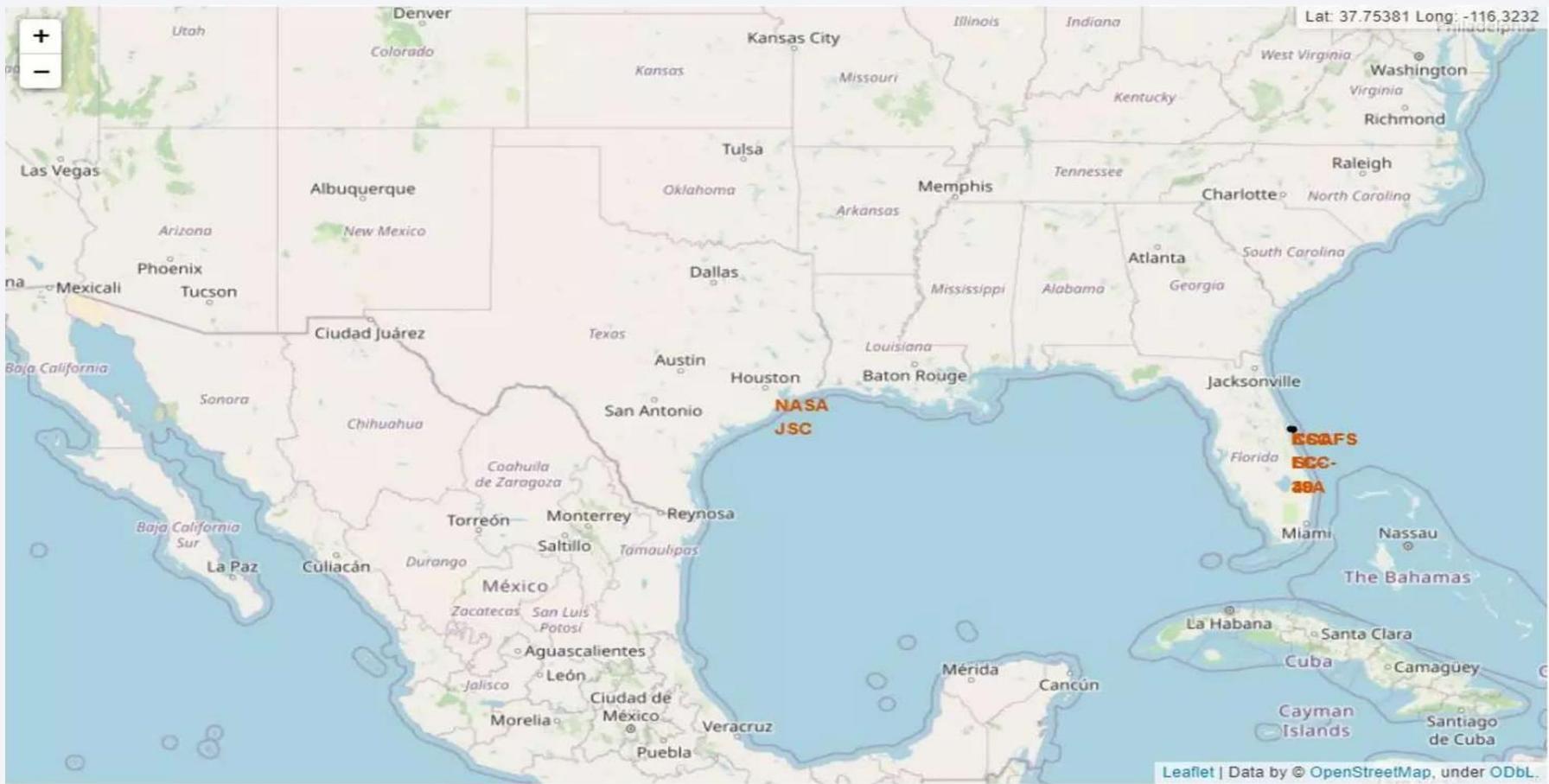
Marked launch sites on map!



Success/Failed Launch Sites Marked on the Map!

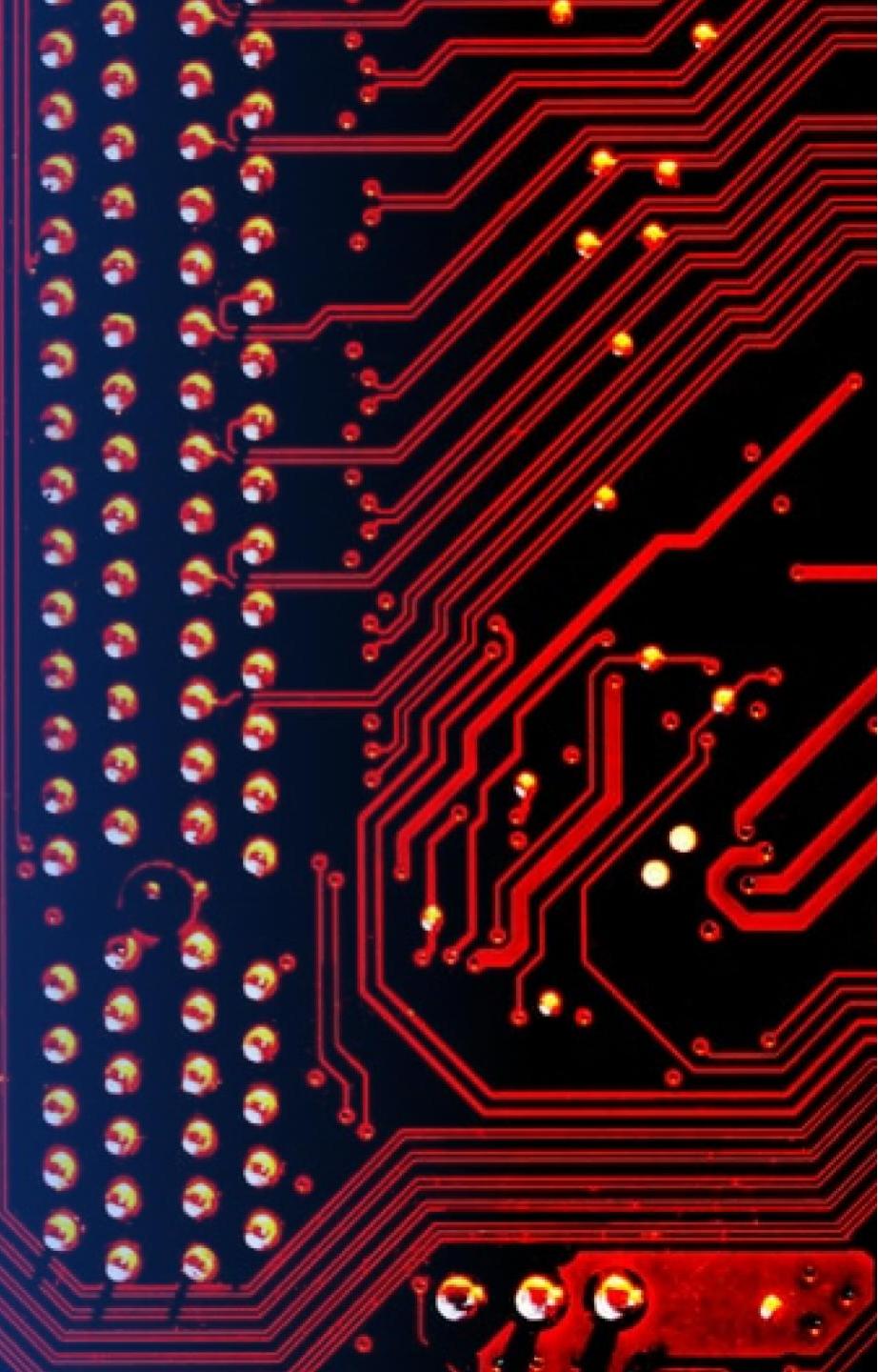


Distance between launch sites to it's proximity



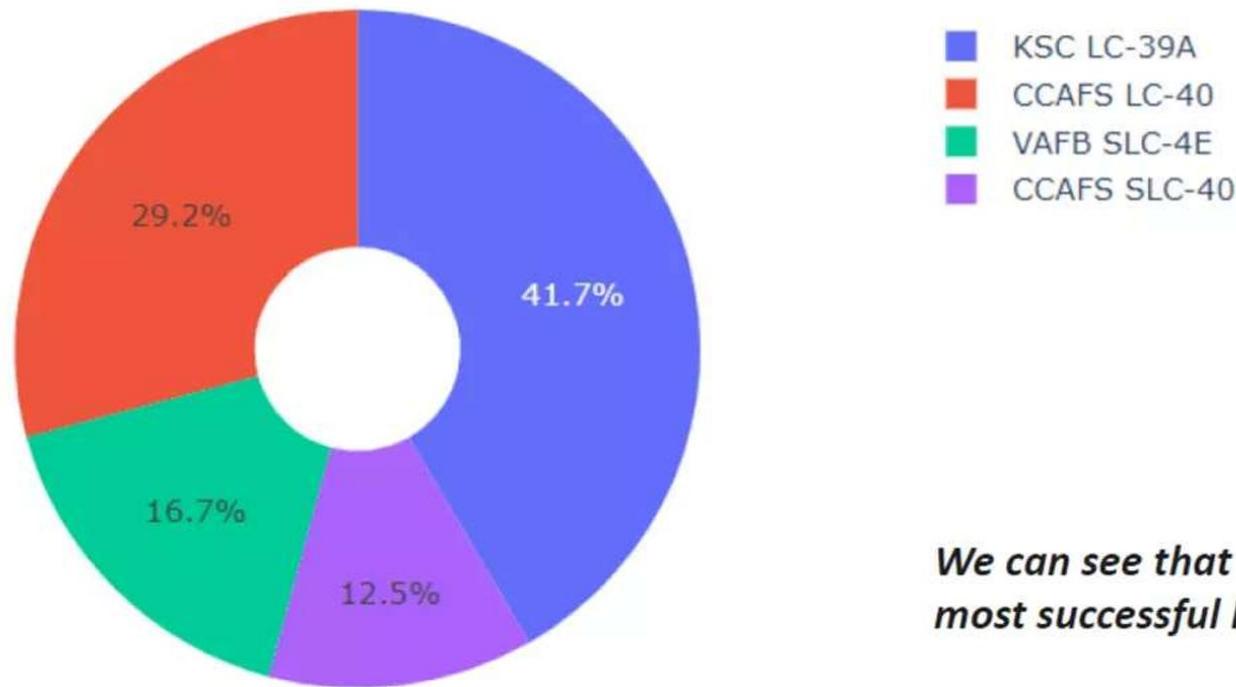
Section 4

Build a Dashboard with Plotly Dash



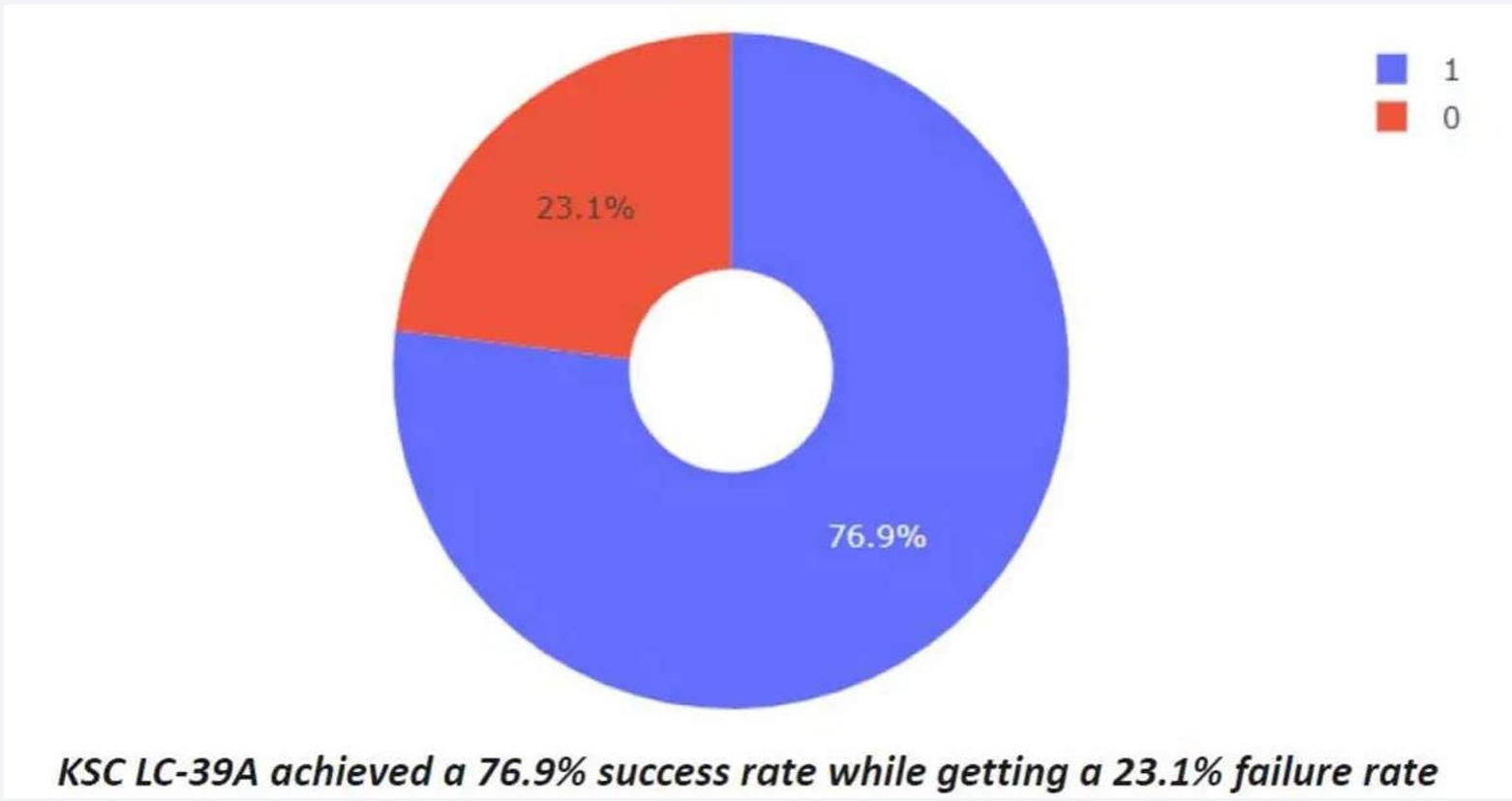
Total Success Launches by all sites!

Total Success Launches By all sites

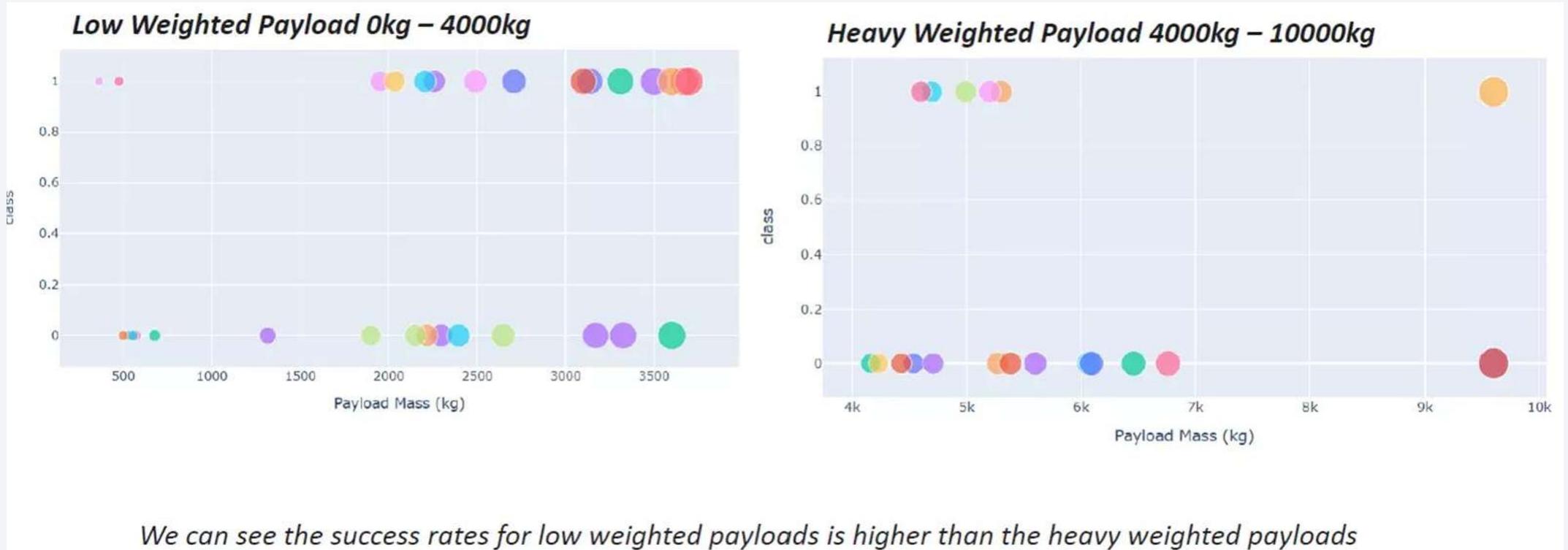


We can see that KSC LC-39A had the most successful launches from all the sites

Success rate by site!



Payload vs Launch outcome!

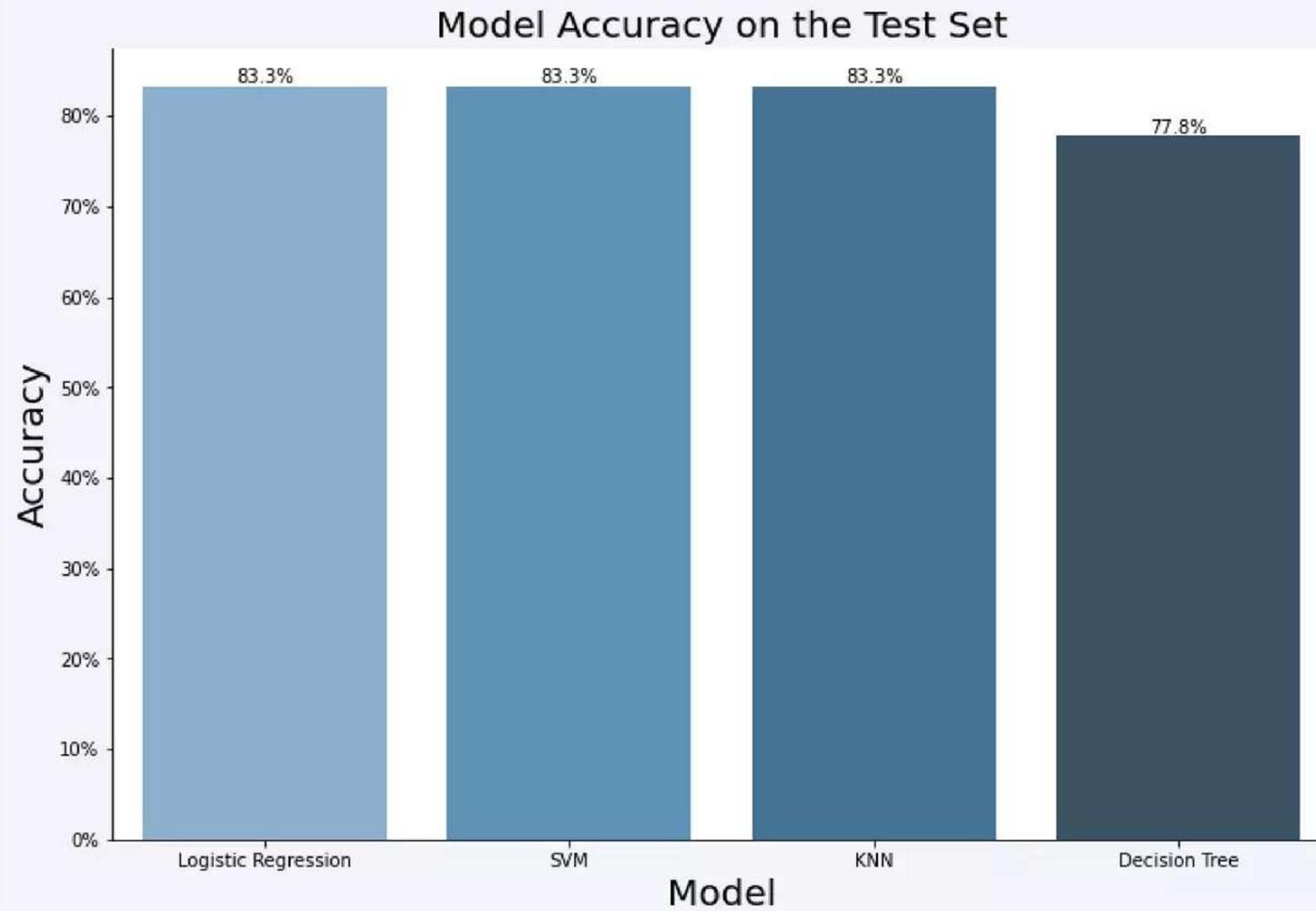


Section 5

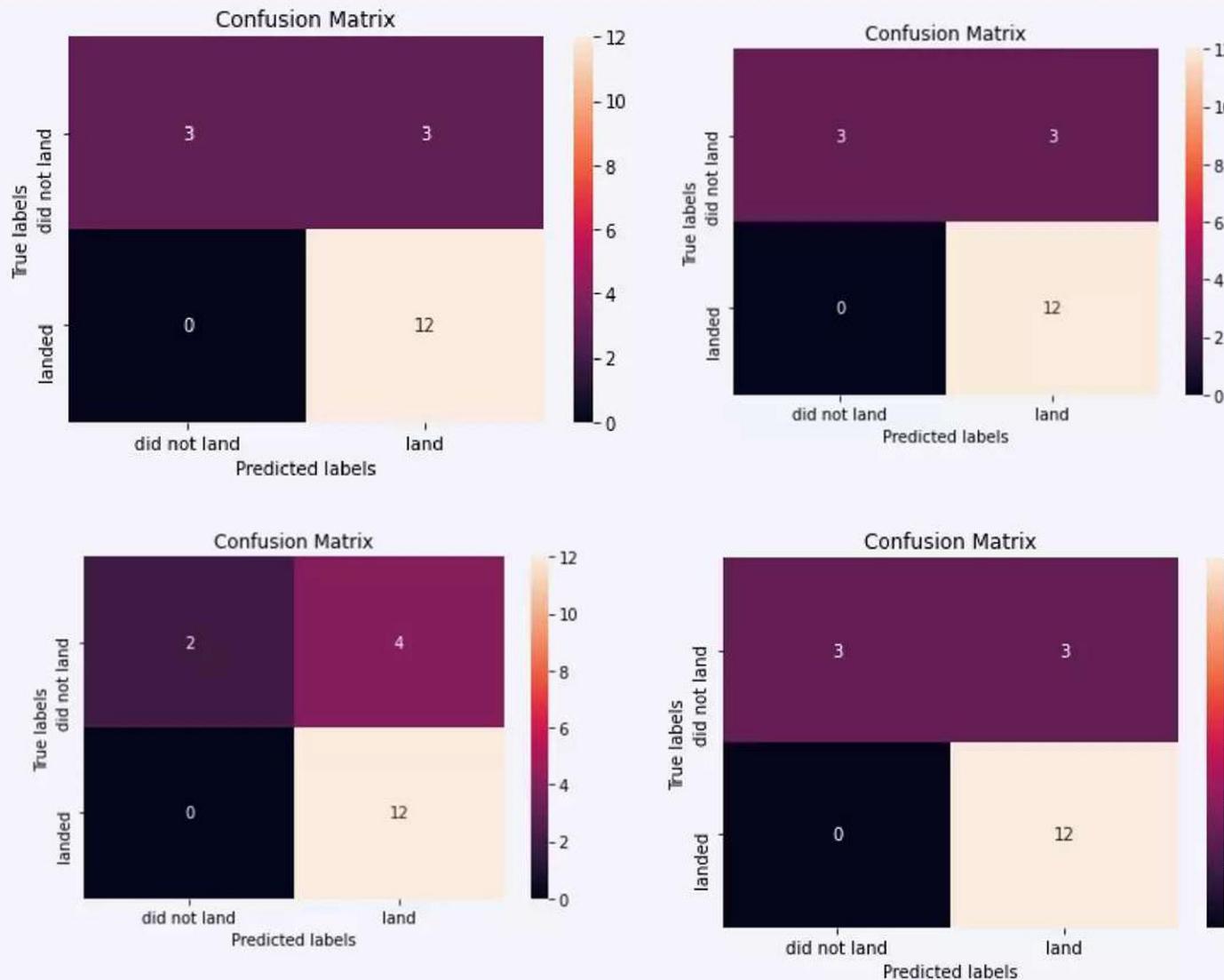
Predictive Analysis (Classification)

Classification Accuracy

Model 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

- [nbharwad/IBM Data Science Certificate: IBM Data Science Certificate.](#)

Thank you!

