

Winning Space Race with Data Science

Syed Muhammad Abdullah
28 Jun 2024



Outline

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

Executive Summary

- Falcon-9 Rocket first launch data was analyzed deeply
- Data was further used to generate a model to predict the future-success of rkt

Introduction

- SpaceX advertises falcon9 rocket launches on its website with a cost of 62 Mn \$, other providers cost upward of 165 Mn \$.
- The project task is to predict 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 scraping from the internet
- 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

- Data was used from the SPACEX launch data REST API which provides us with the information about:
 - Launches
 - Rocket type
 - Payload delivered
 - Landing specs
 - Landing outcome
 - Launch specs

Data Collection – SpaceX API

- Data was collected with SPACEX REST calls

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)    getBoosterVersion(data)  
getPayloadData(data)  getCoreData(data)
```

4. Assign list to dictionary then dataframe

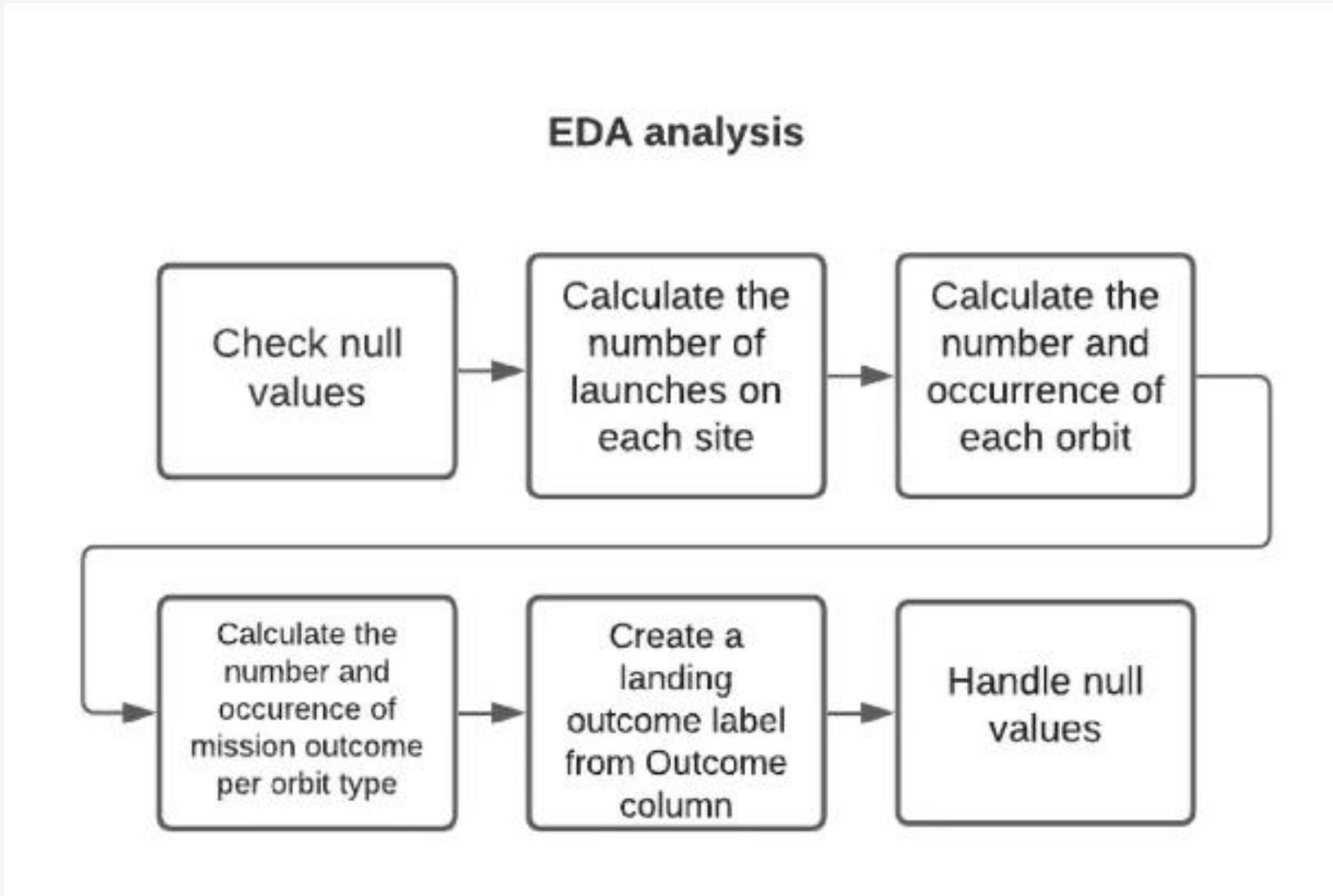
```
launch_dict = {'FlightNumber': list(data['flight_number']),  
              'Date': list(data['date']),  
              'BoosterVersion':BoosterVersion,  
              'PayloadMass':PayloadMass,  
              'Orbit':orbit,  
              'LaunchSite':LaunchSite,  
              'Outcome':Outcome,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

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

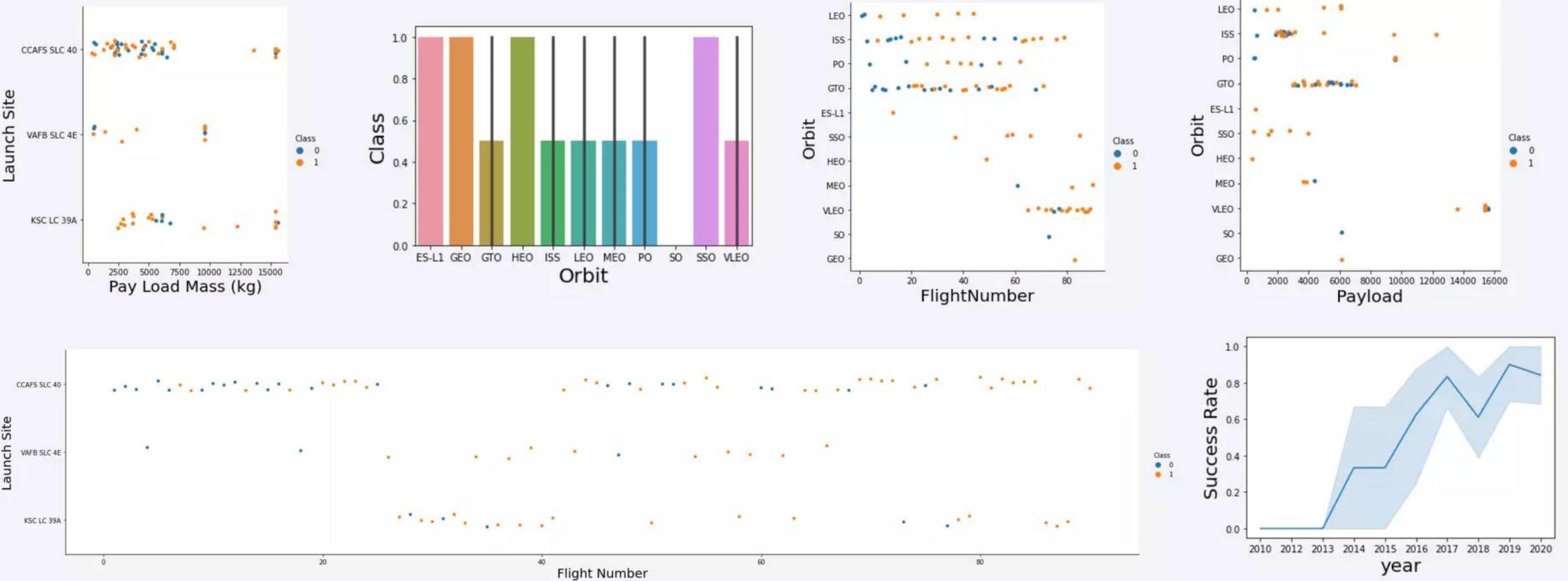
5. Filter dataframe and export to flat file (.csv)

```
data_falcon9 = df.loc[df['BoosterVersion']!='Falcon 1']  
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

Data Wrangling



EDA with Data Visualisation

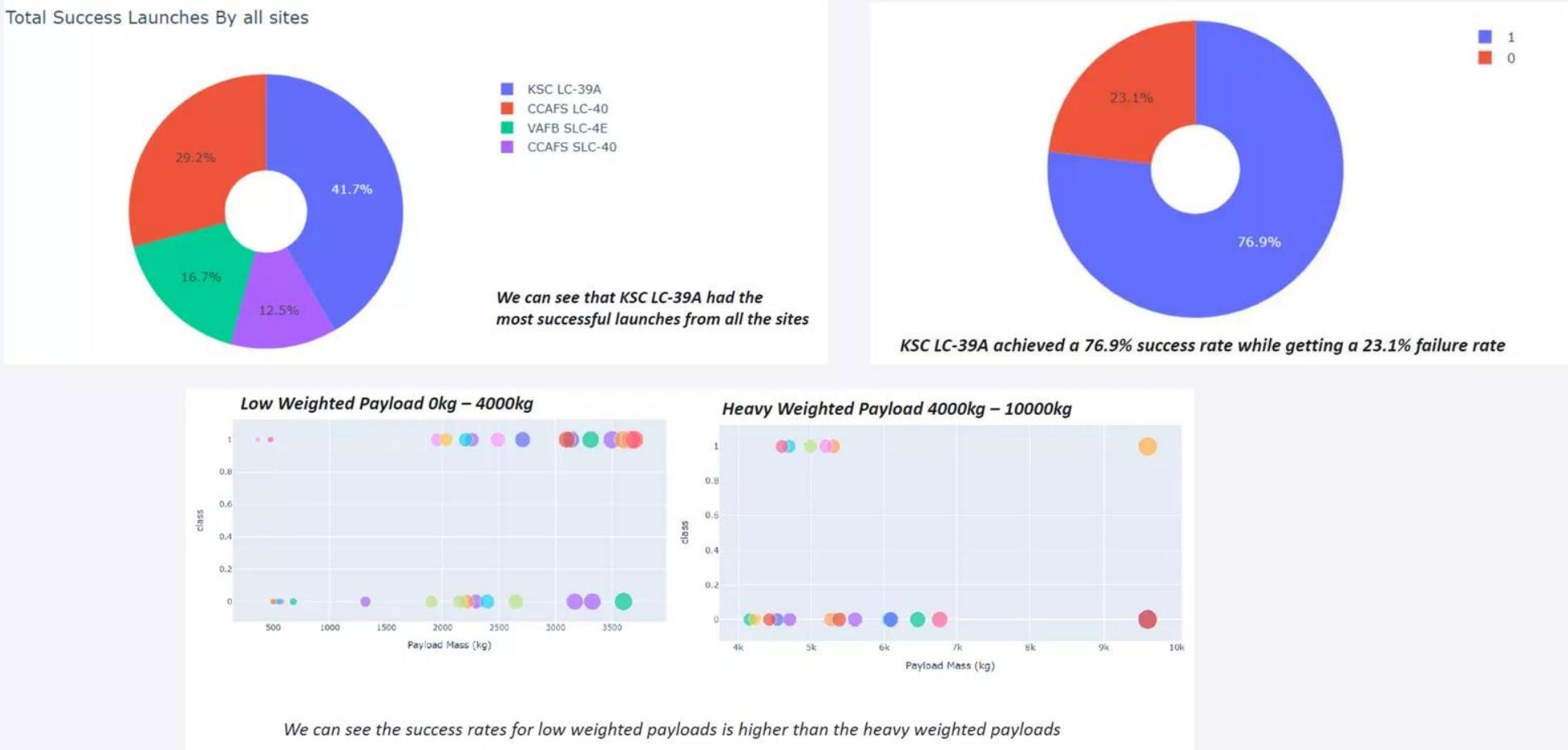


<https://github.com/initiative1972/data-science/blob/master/10-IBM%20DS%20Capstone-lab4-EDA%20viz.ipynb>

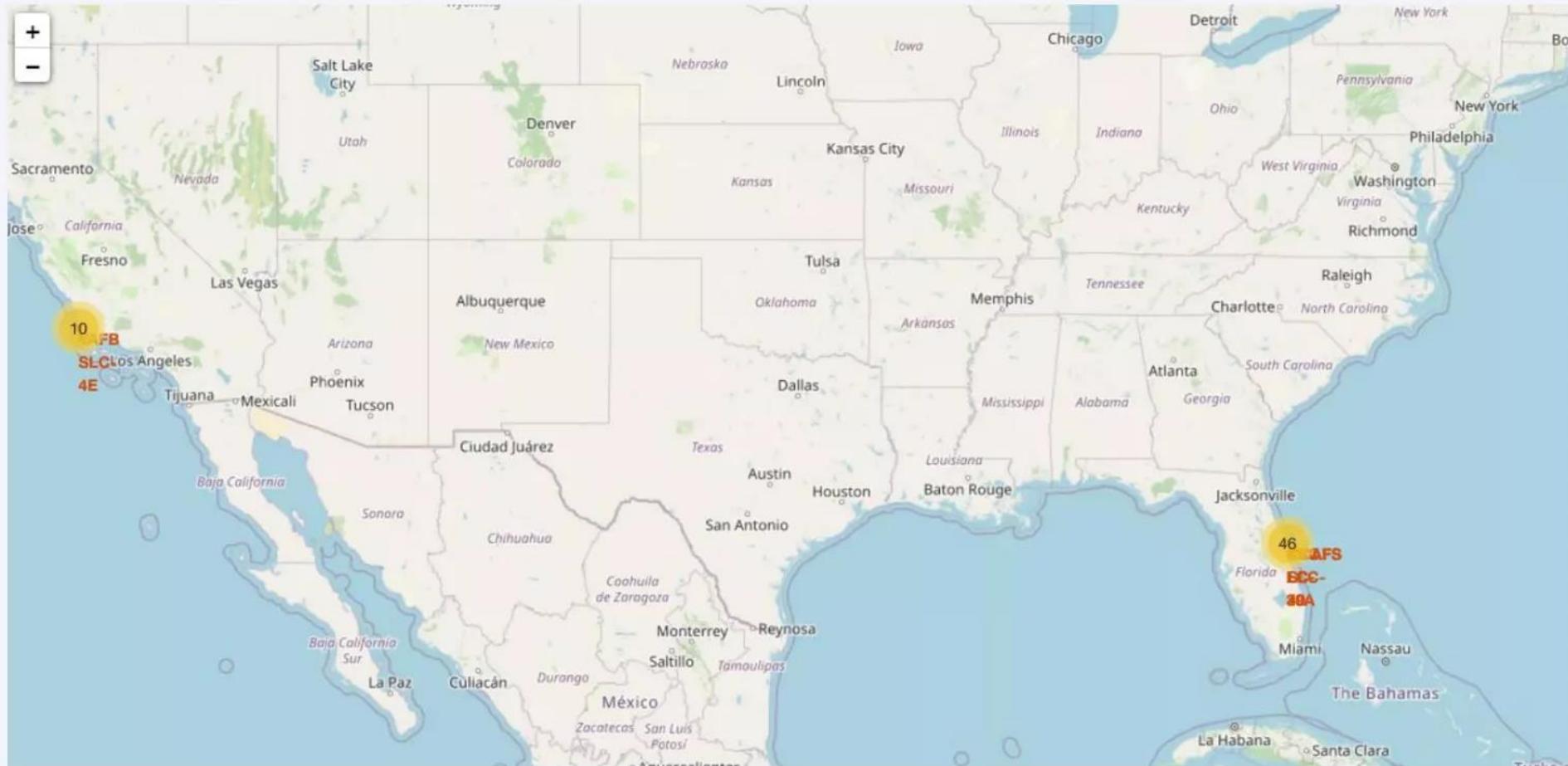
EDA with SQL

- SQL Queries performed include:
 - Displaying the names of the unique launch sites in the mission
 - Displaying records where launch sites begin with the string KSC
 - Displaying the total payload mass carried
 - Displaying the average payload mass
 - Listing the details of successes and failures of landings

Build a Dashboard with Plotly Dash

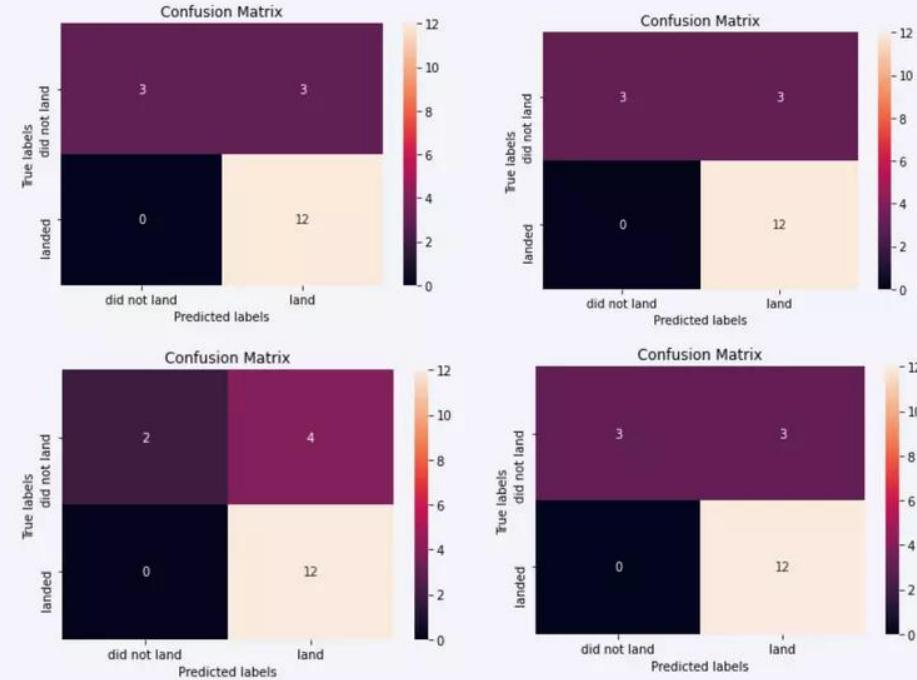
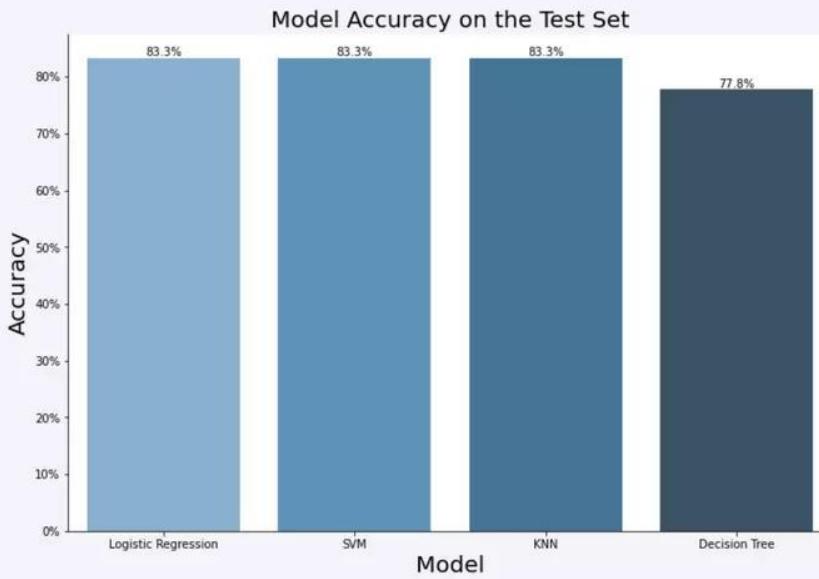


Build an Interactive Map with Folium



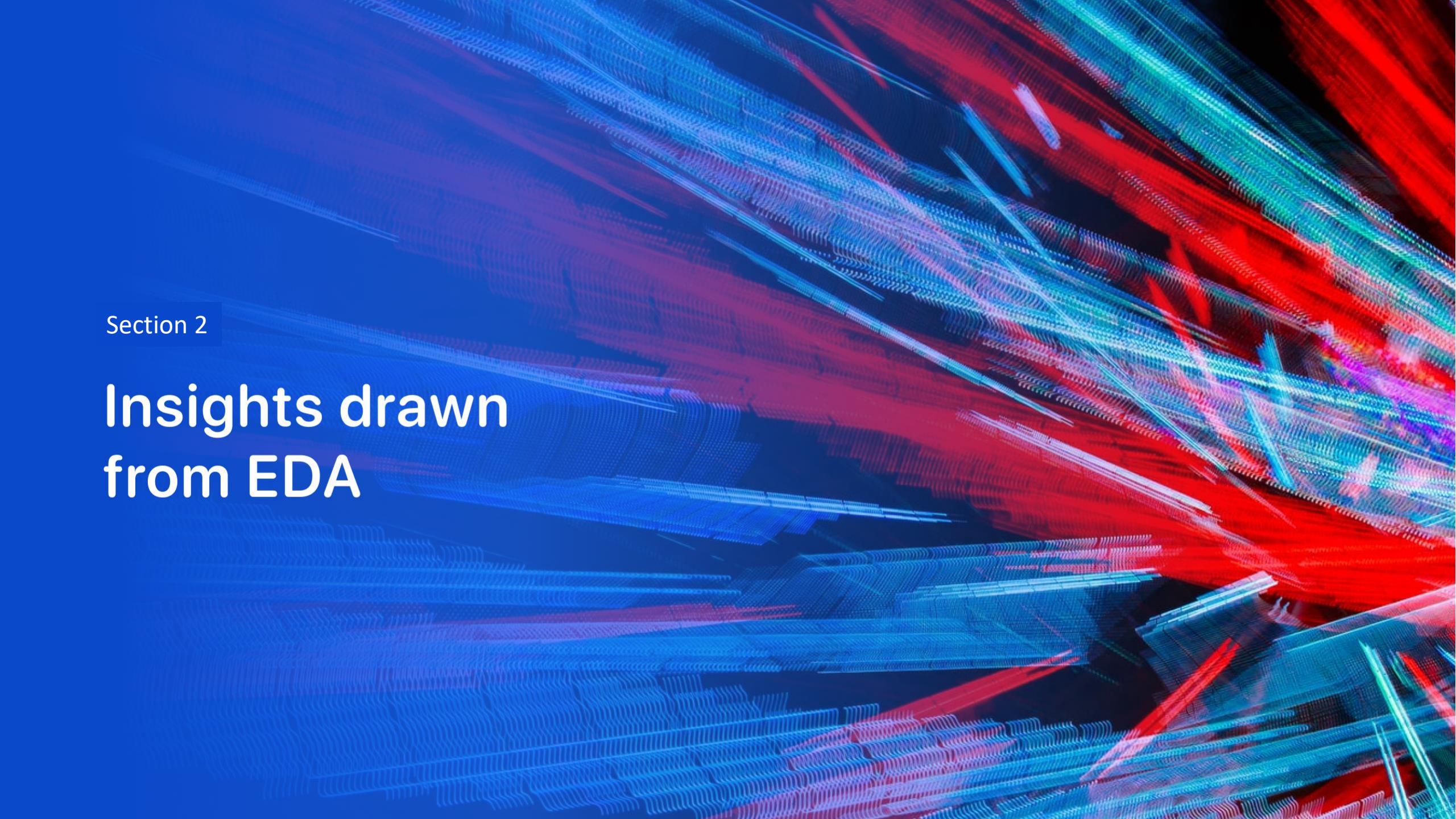
Predictive Analysis (Classification)

- 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

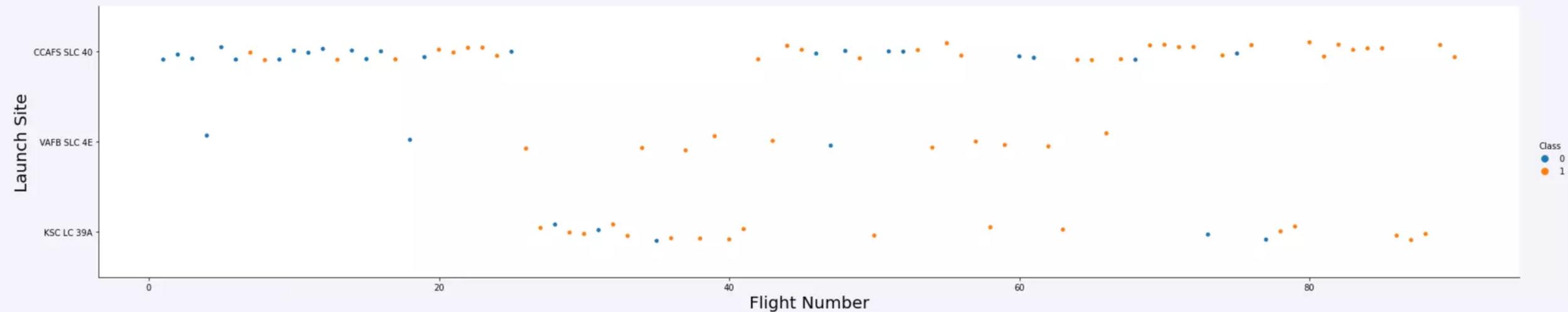
- The SVM, KNN and Logistic Regression models are the best in terms of prediction accuracy for this dataset
- Low weighted payloads perform better than the heavier payloads
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches
- KSC LC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

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

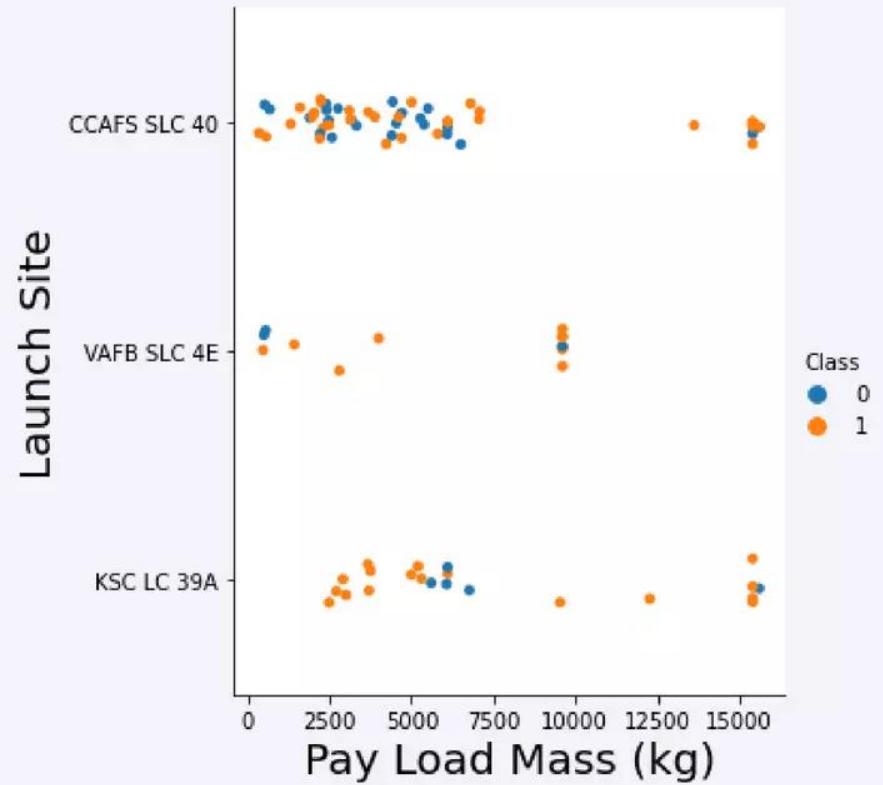
Insights drawn from EDA

Flight Number vs. Launch Site



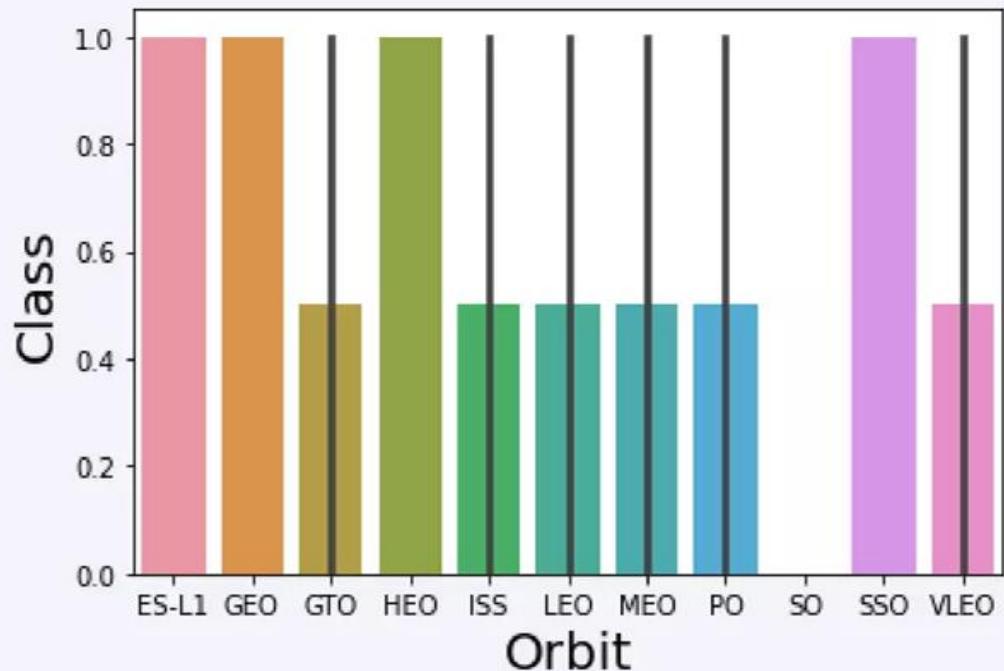
- Launches from the site of CCAFS SLC 40 are significantly higher than launches from other sites.

Payload vs. Launch Site



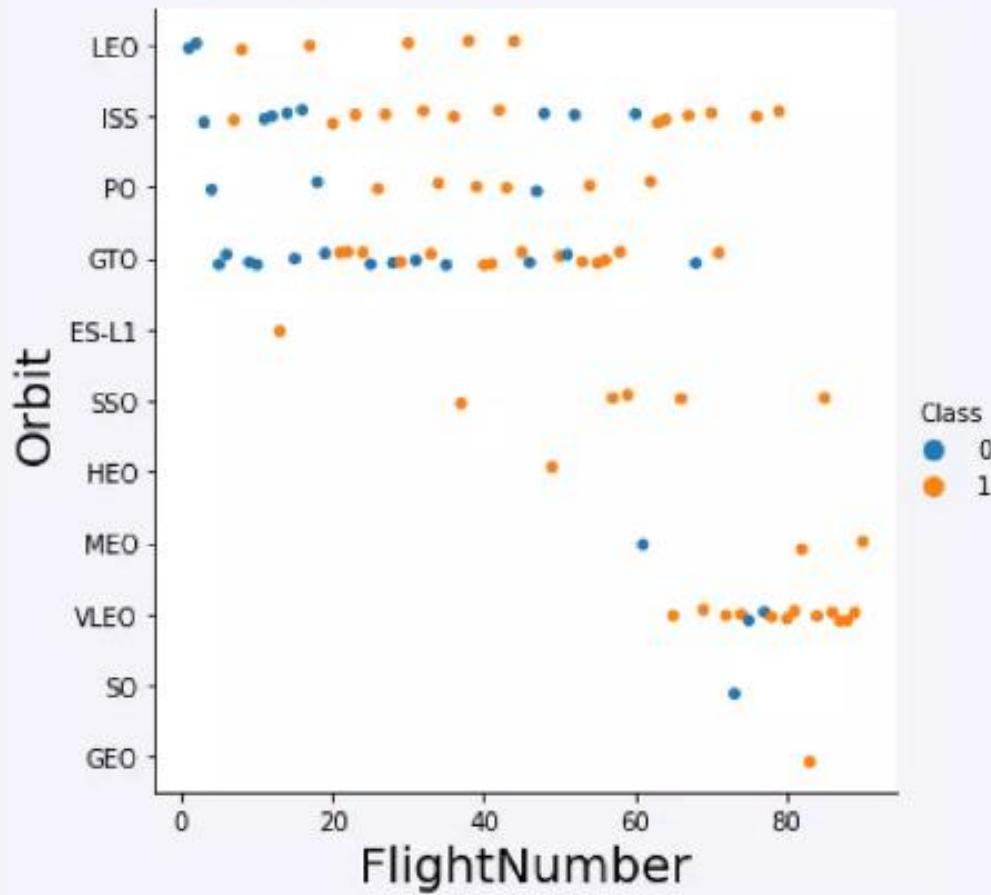
- The majority of IPay Loads with lower Mass have been launched from CCAFS SLC 40.

Success Rate vs. Orbit Type



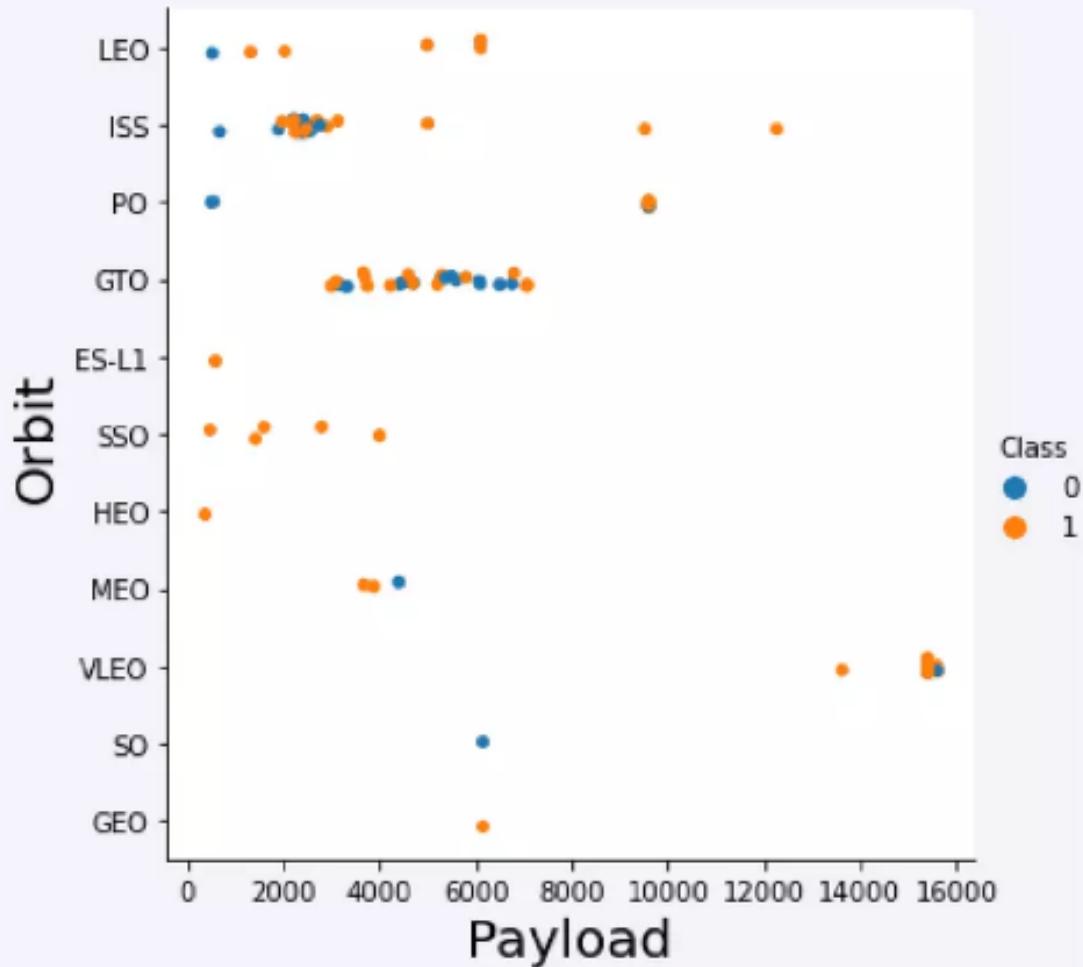
- The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.

Flight Number vs. Orbit Type



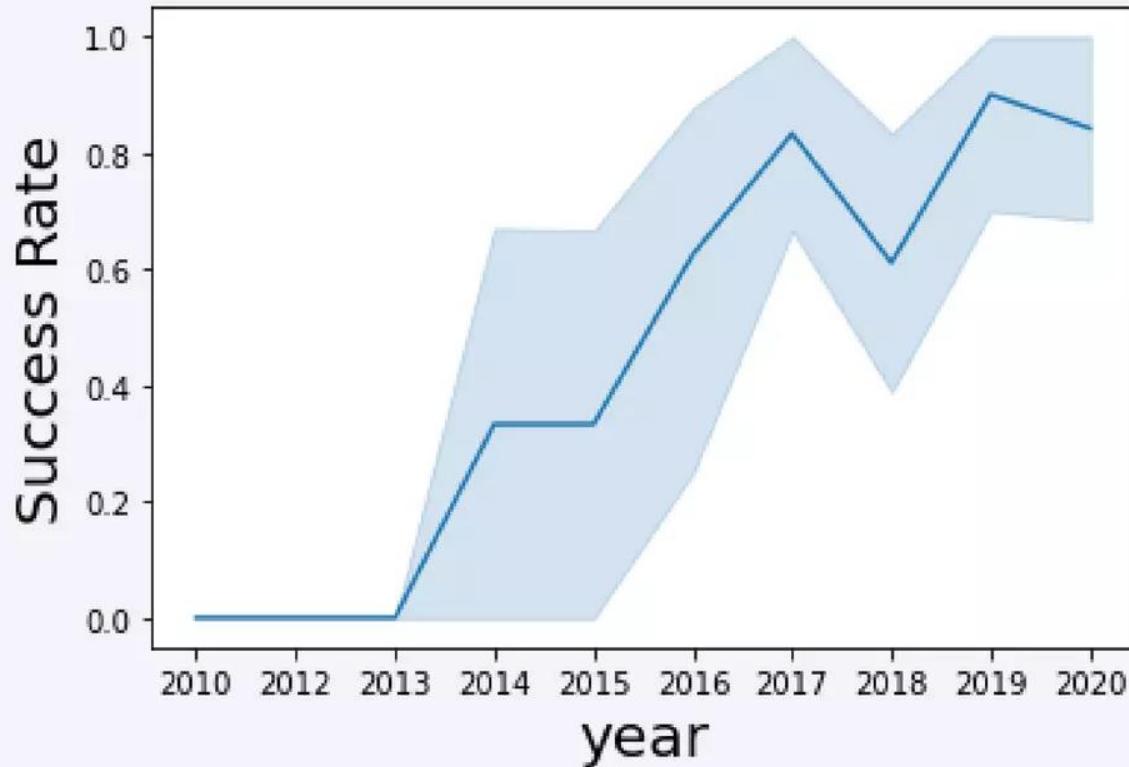
- A trend can be observed of shifting to VLEO launches in recent years.

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 of 4000-8000.

Launch Success Yearly Trend



- Launch success rate has increased significantly since 2013 and has stabilised 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	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)'

45596

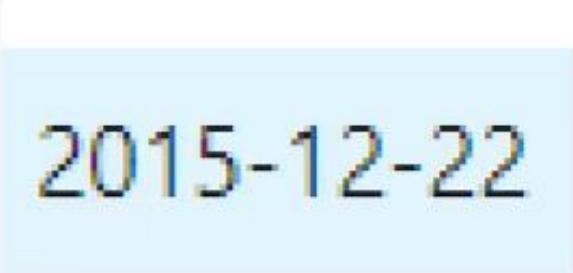
Average Payload Mass by F9 v1.1

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

```
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)
00:22:00	F9 FT B1022.1	CCAFS LC-40	Intelsat 35e	4000	LEO (ISS)	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

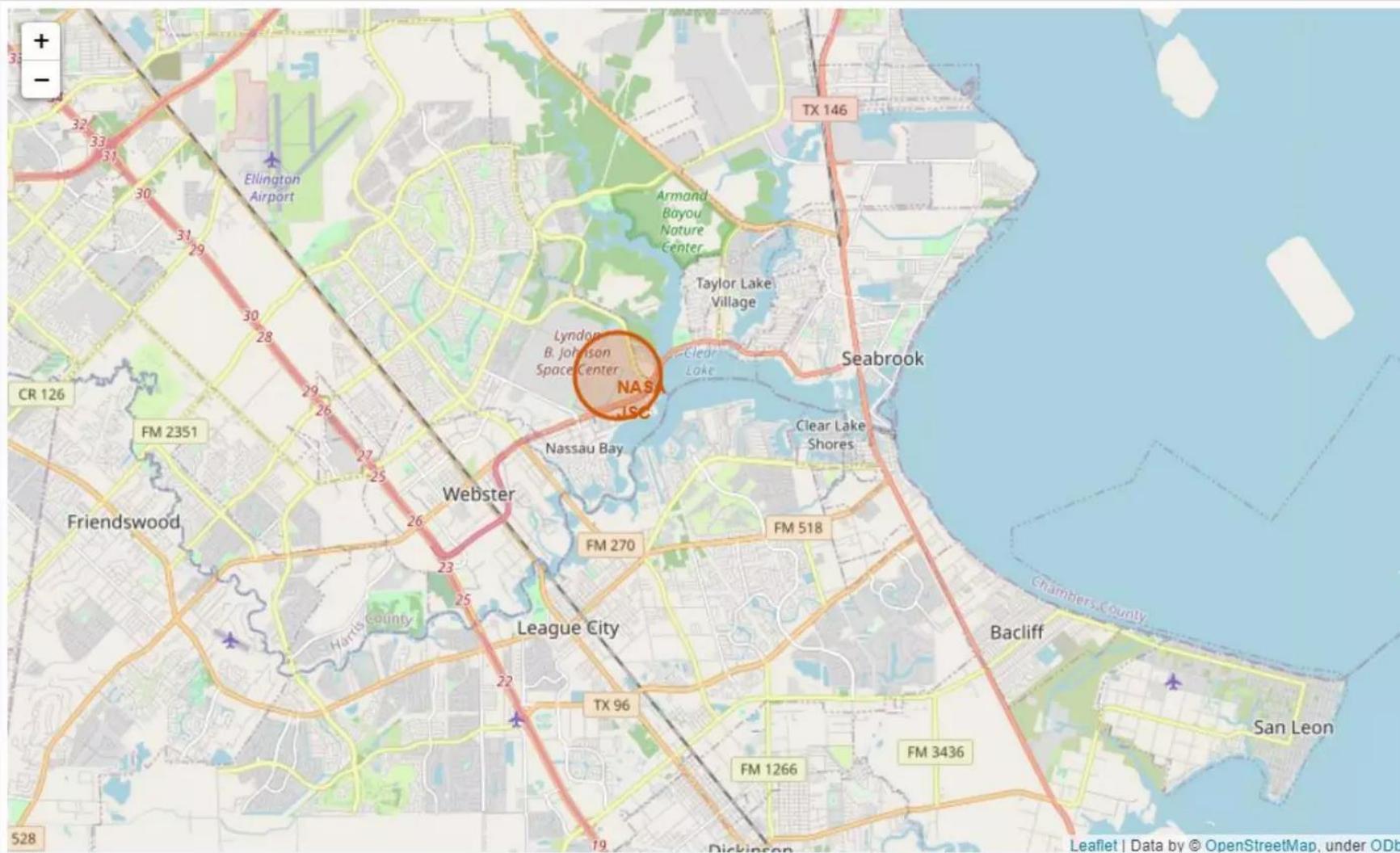
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)	
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)	
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	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's horizon against a dark blue sky. Numerous glowing yellow and white points represent city lights, concentrated in coastal and urban areas. In the upper right quadrant, there are bright green and yellow bands of light, likely the Aurora Borealis or Australis. The overall atmosphere is dark and mysterious.

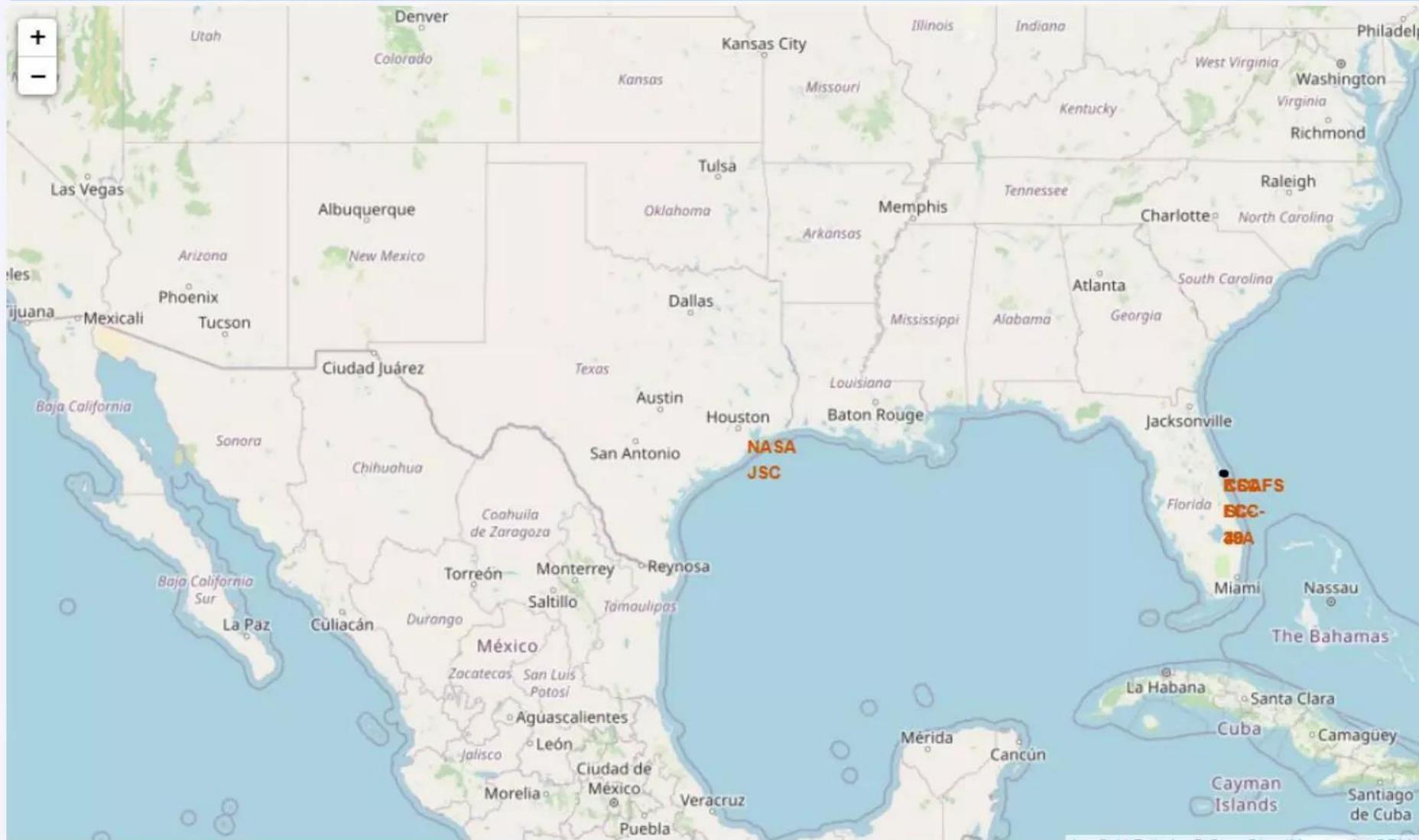
Section 3

Launch Sites Proximities Analysis

All launch sites marked on a map

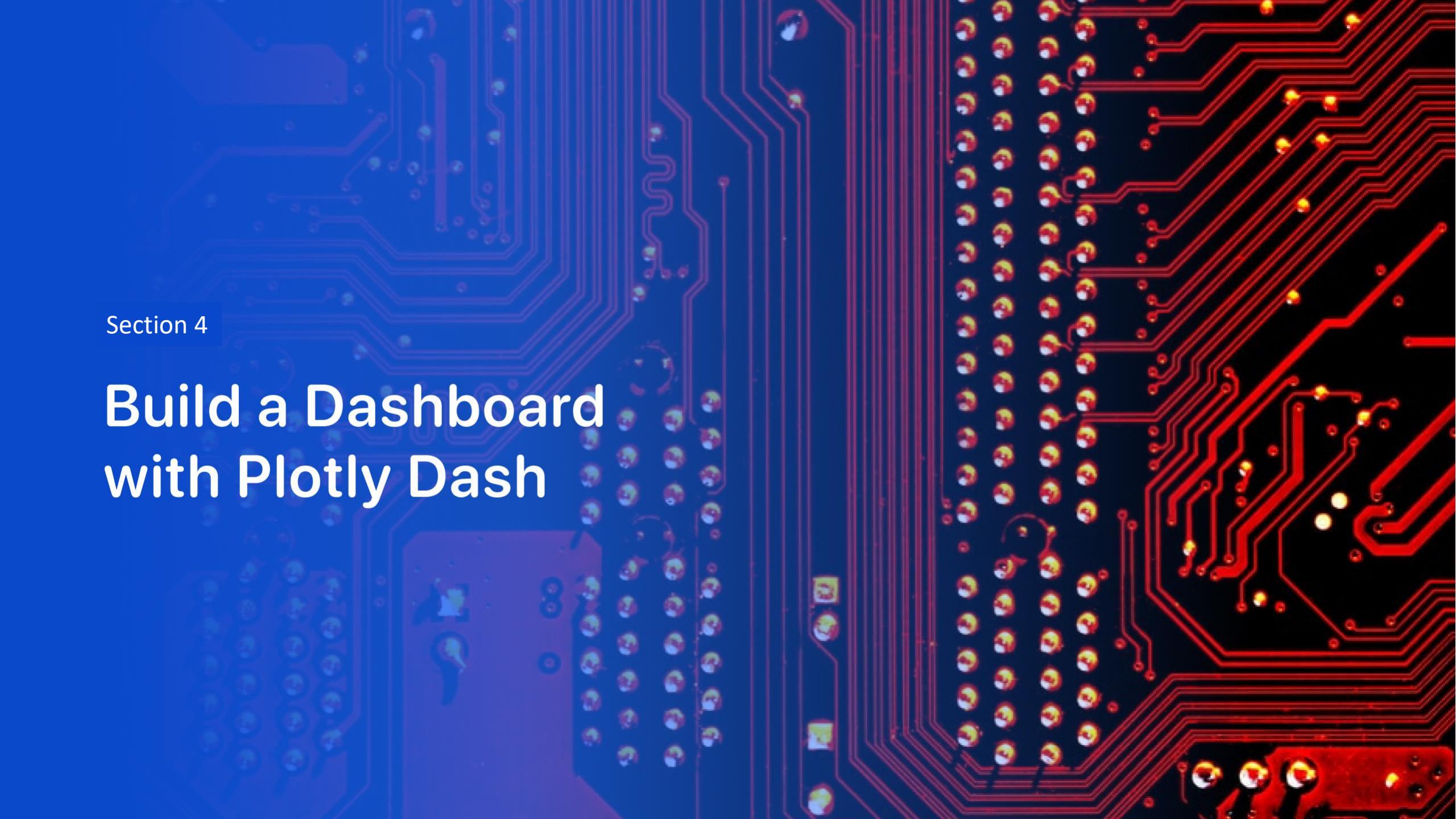


Success/failed launches marked on the map



<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot

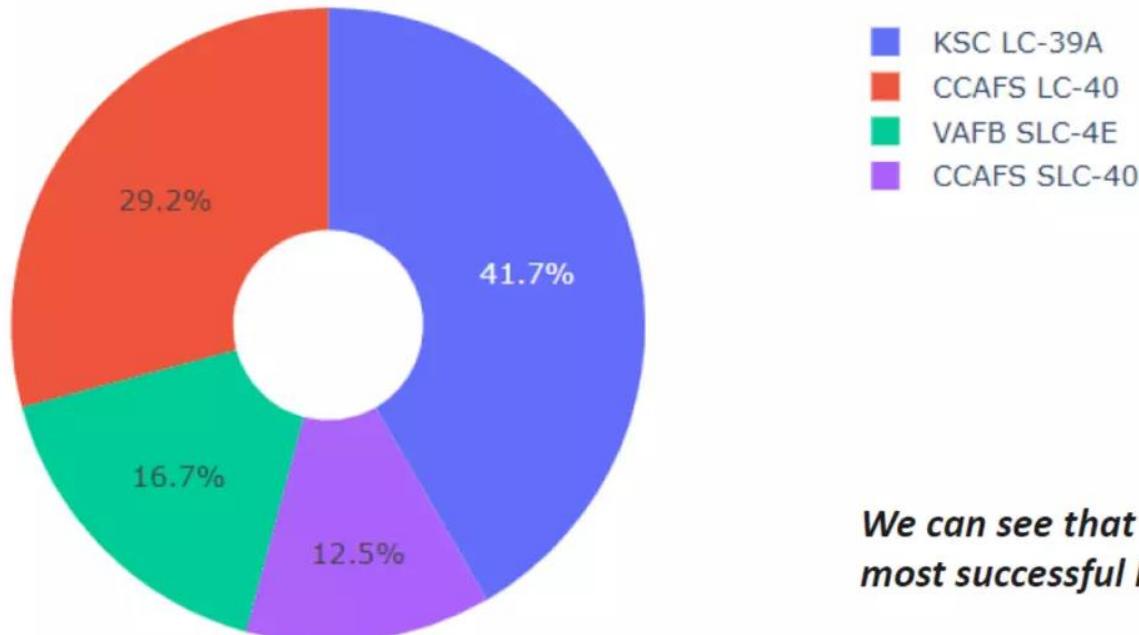


Section 4

Build a Dashboard with Plotly Dash

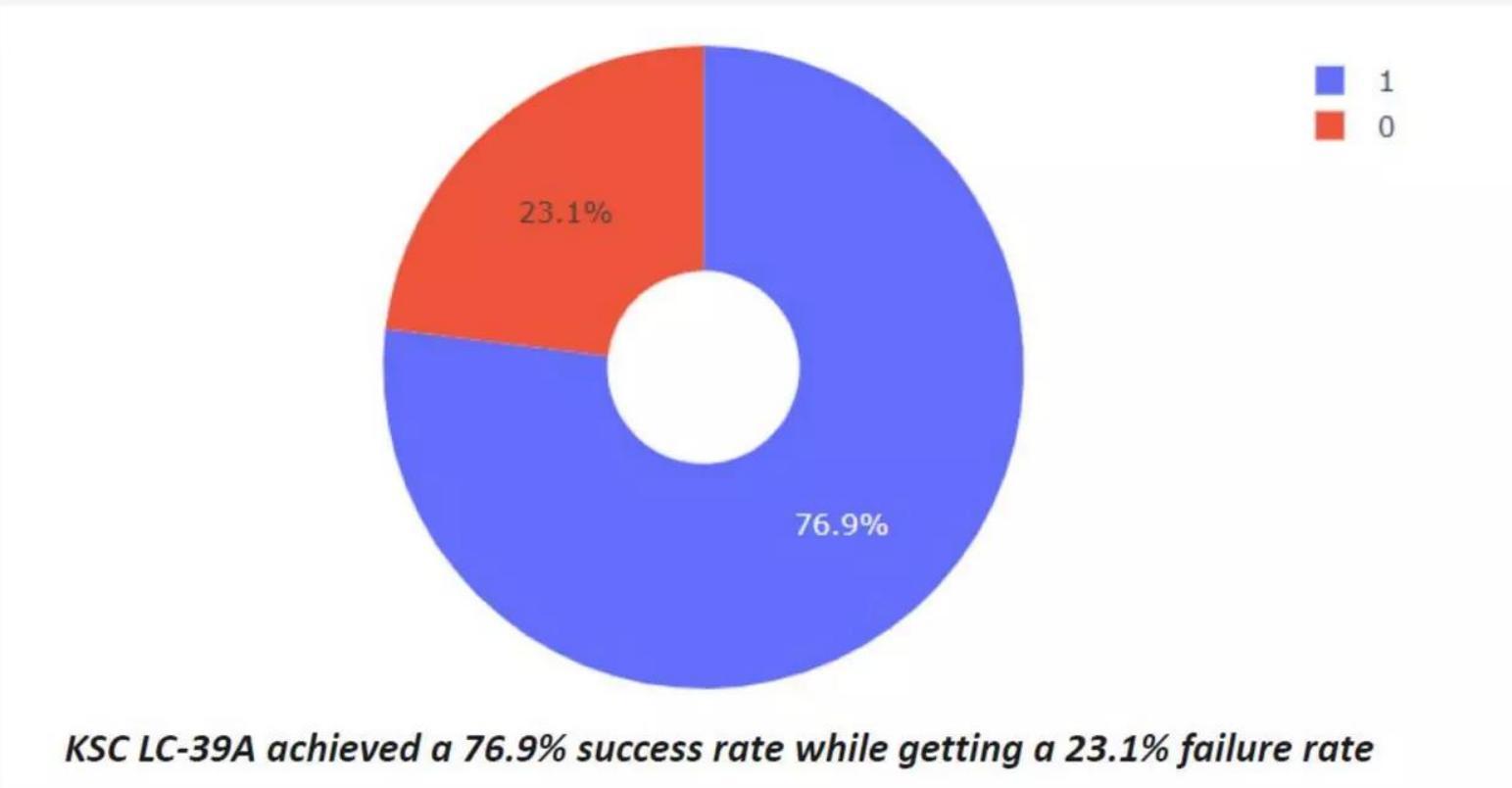
Total success launches by all sites

Total Success Launches By all sites

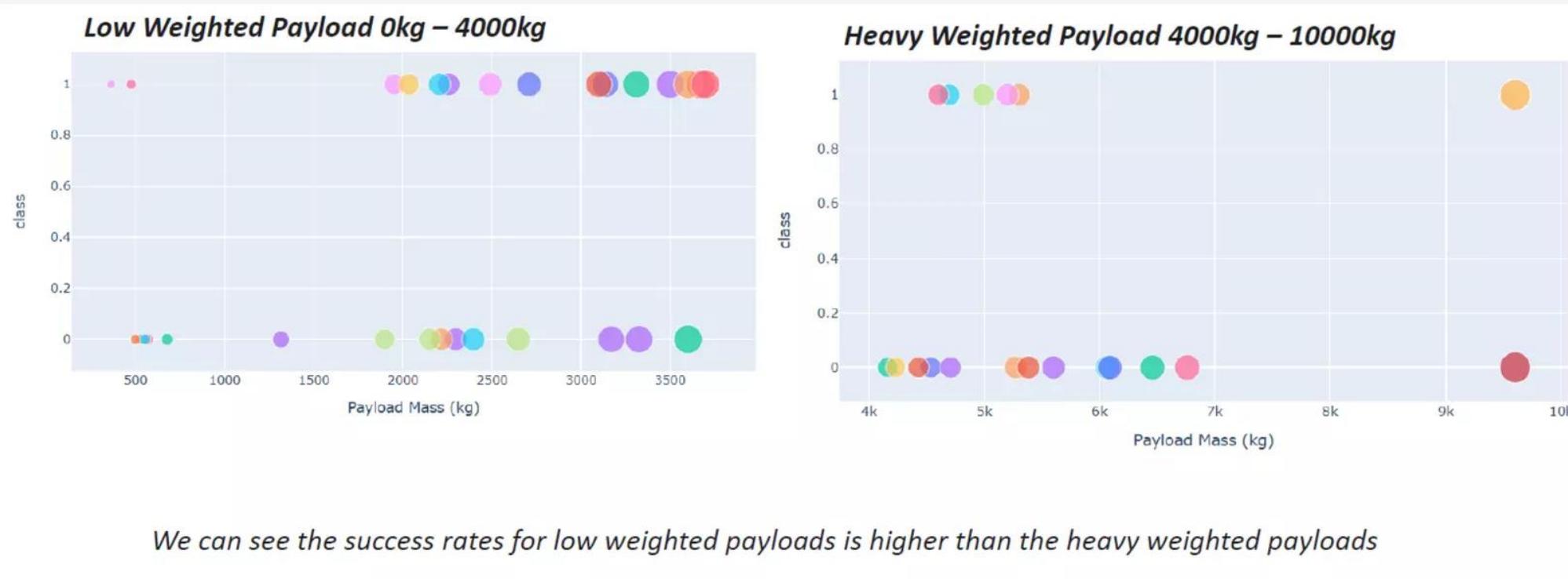


We can see that KSC LC-39A had the most successful launches from all the sites

Success rate by site



Payload vs launch outcome

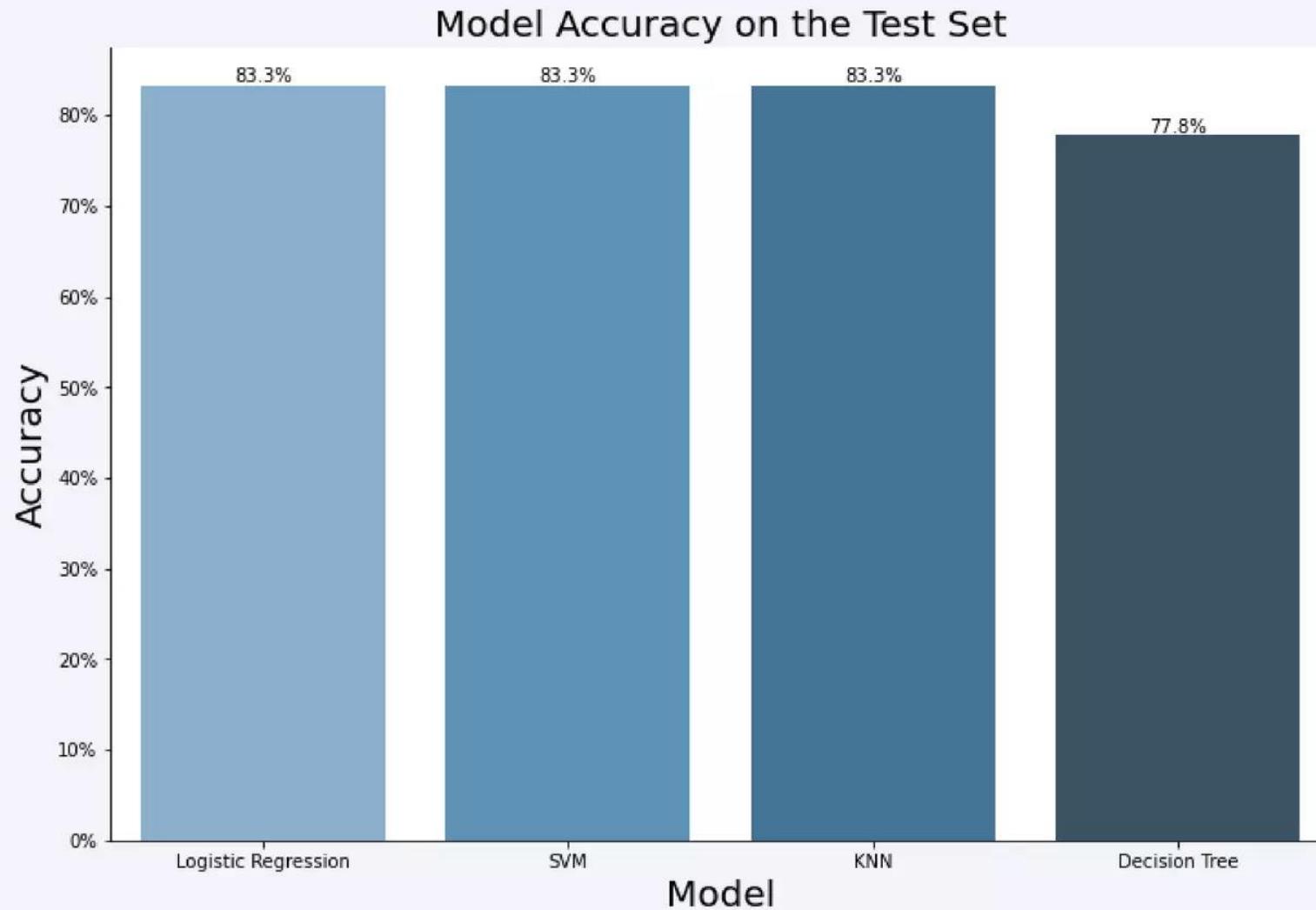


The background of the slide features a dynamic, abstract design. It consists of several curved, overlapping bands of color. A prominent band on the left is a bright blue, while another on the right is a warm yellow. These colors transition into lighter, more diffused tones towards the edges of the frame. The overall effect is one of motion and depth.

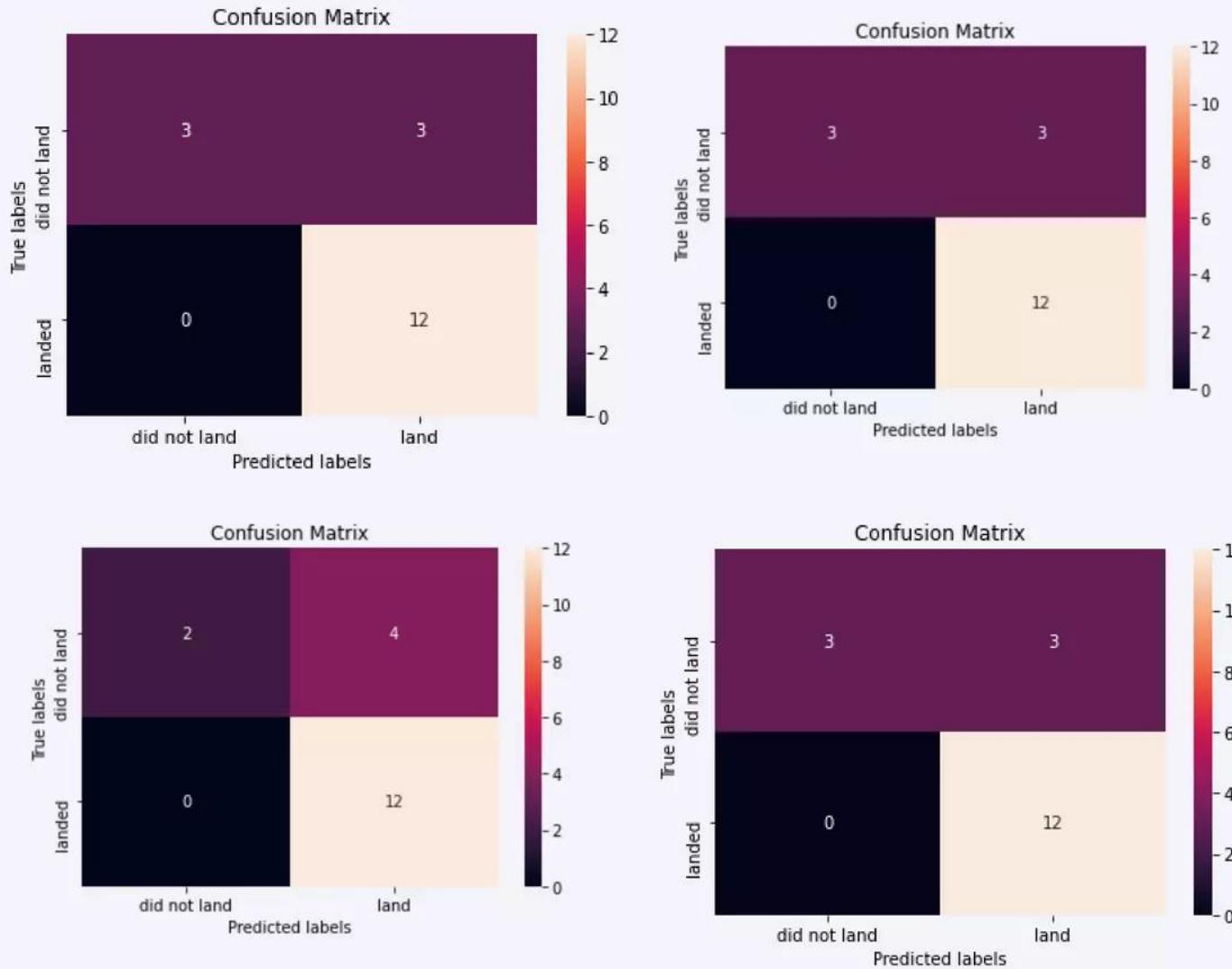
Section 5

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.
- Orbit GEO, HEO, SSO, ES L1 has the best success rates
- KSC LC 39A had the most successful launches from all the sites.

Thank you!

