IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Chandana Priya Gandikota

09/11/2023

Outline

Executive Summary

Introduction

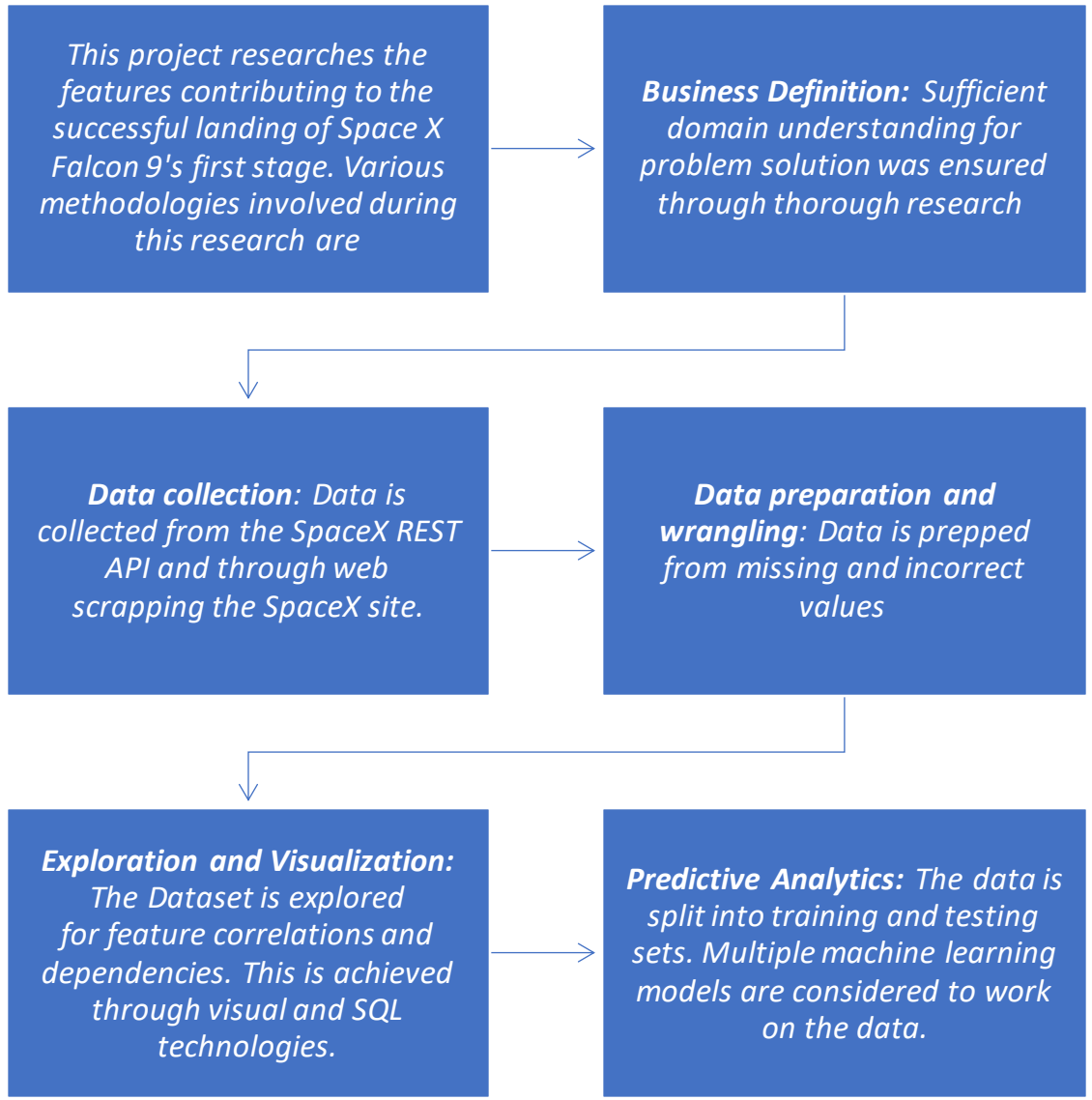
Methodology

Results

Conclusion

Appendix

Executive Summary - *Methodologies*



Executive Summary - Results

Data: The data is received from the SpaceX REST API and their website

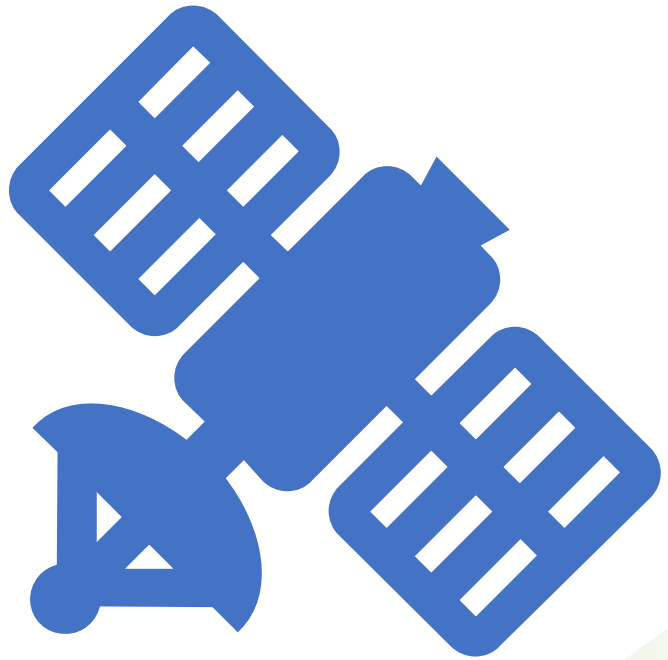
Data wrangling: All the Null values are removed

Exploration and Visualization: below are the major launching sites with the launching success rate increasing from 2013 till 2020.

Predictive Analytics : While multiple classification models have similar accuracy at 83.3%, Decision Tree performed slightly better with higher F1 Score

Introduction

Space X is a commercial space industry giant with accomplishments of sending spacecraft to the international space station, establishing Starlink, a satellite internet constellation and sending manned, unmanned missions to space. Space X boosts of offering space launches for $62 millions while the other providers cost at $165 million dollars. This is possible mainly because of the reusable first stage of Falcon 9 rocket.



Questions to Answer

- *Checking the successful landing of Falcon 9 stage 1*
- *Can the price of launch be determined accurately*

Section 1

Methodology

Methodology

Executive Summary

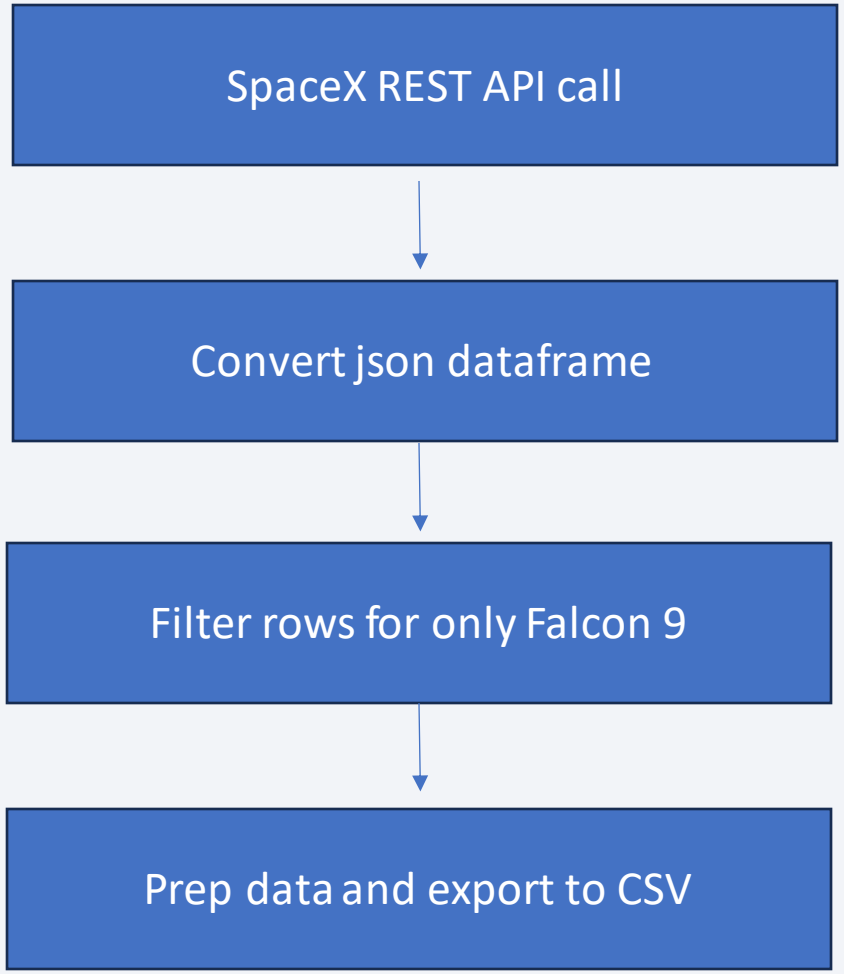
- Data collection methodology: The data was collected through Space X REST API and Space X website
- Perform data wrangling : Prep the data for null and inconsistent values
- Perform exploratory data analysis (EDA) using visualization and SQL: explore for relations between data features
- Perform interactive visual analytics using Folium and Plotly Dash: Using visual techniques to understand data
- Perform predictive analysis using classification models: Use machine learning models to fit the training data.

Git hub link

https://github.com/chandana-priya-g/DataScienceTools/blob/main/FinalNotebooks/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

https://github.com/chandana-priya-g/DataScienceTools/blob/main/FinalNotebooks/jupyter-labs-spacex-data-collection-api.ipynb

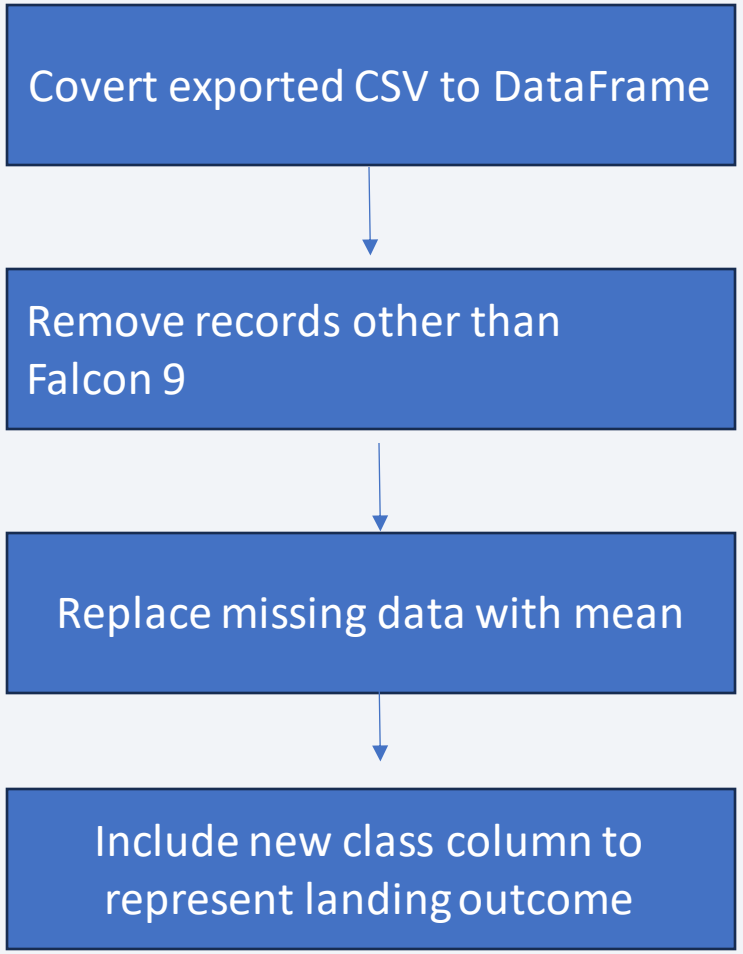
Data Collection



Data Collection – SpaceX API

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Data Wrangling



EDA with Data Visualization

Below are the various aspect part of the Exploratory Data Analysis

- *See how the FlightNumber (indicating the continuous launch attempts.) and Payload variables would affect the launch outcome.*
- *A drill down to each site visualize its detailed launch records.*
- *Patterns with Flight Number vs. Launch Site scatter point plots.*
- *Any relationship between launch sites and their payload mass.*
- *Visualize the relationship between success rate of each orbit type*
- *Visualize the relationship between FlightNumber and Orbit type*
- *Visualize the relationship between Payload and Orbit type*
- *Visualize the launch success yearly trend*
- *Features Engineering*

EDA with SQL

Below are the various aspect part of the Exploratory Data Analysis using SQL

- *Display the names of the unique launch sites in the space mission*
- *Display 5 records where launch sites begin with the string 'CCA'*
- *Display the total payload mass carried by boosters launched by NASA (CRS)*
- *Display average payload mass carried by booster version F9 v1.1*
- *List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000*
- *List the total number of successful and failure mission outcomes*
- *List the names of the booster_versions which have carried the maximum payload mass. Use a subquery*
- *List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.*
- *Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.*

Build an Interactive Map with Folium

The interactive data visualization using Folium was used to better understand the launch locations. A successful launch also depends on location and proximity of launch sites, including initial positions of rocket trajectories. Folium is a handy tool of such geographical visualization.

Various factors analyzed using Folium are:

- *Marking all launch sites on a map*
- *Marking the success/failed launches for each site on the map*
- *Calculating the distances between a launch site to its proximities*
- *Drawing a line between a launch site to a coastline point close by*
- *Drawing a line between a launch site to its closest city, railway, highway, etc*

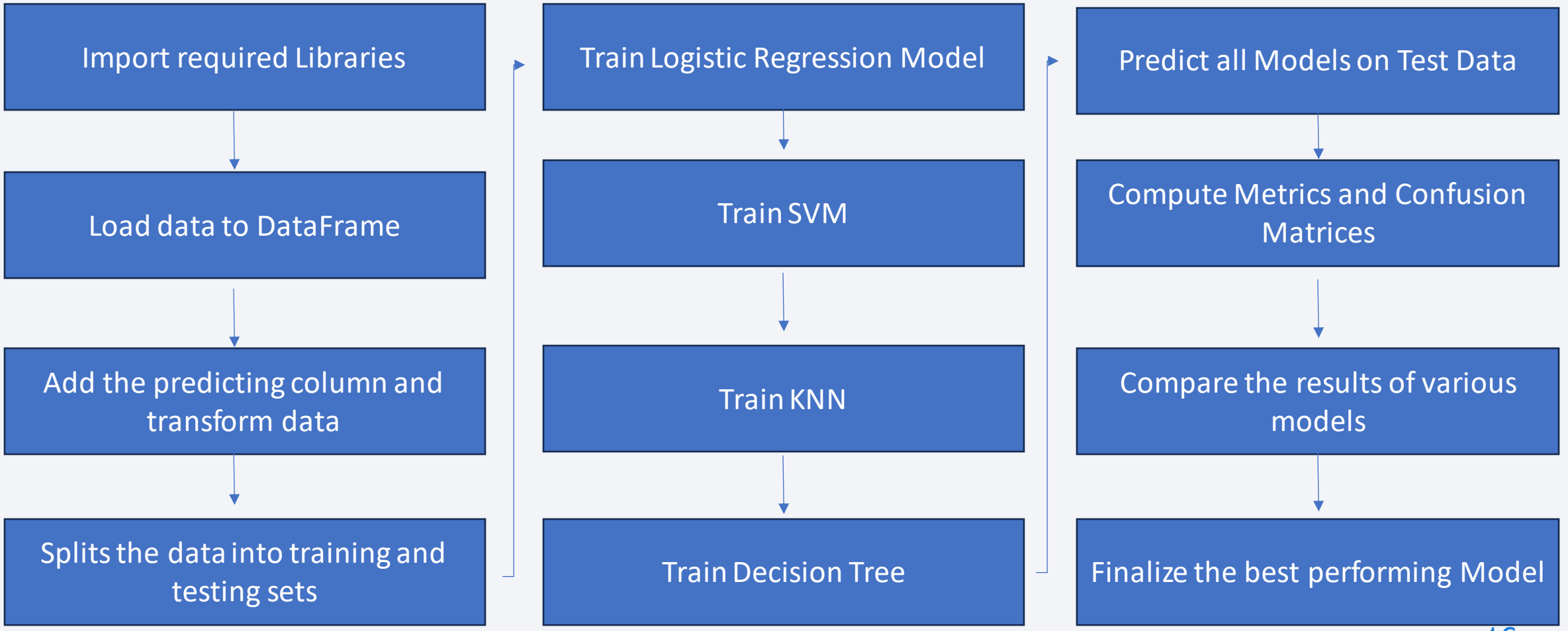
Build a Dashboard with Plotly Dash

The application allows the users to perform interactive visualization on SpaceX data using charts and other figures, helping in better exploration.

The attributes explored in this section are:

- *Add a Launch Site Drop-down Input Component*
- *Add a Range Slider to Select Payload*
- *Creating scatter plots based on selected payloads and launch sites*

Predictive Analysis (Classification)



Results Summary

Exploratory data analysis results

GEO, HEO, ES-L1, SSO have 100% success rate

The rate of success of landing increased after 2013 till 2020

Interactive analytics demo in screenshots

The Launch Sites are closer to coastline

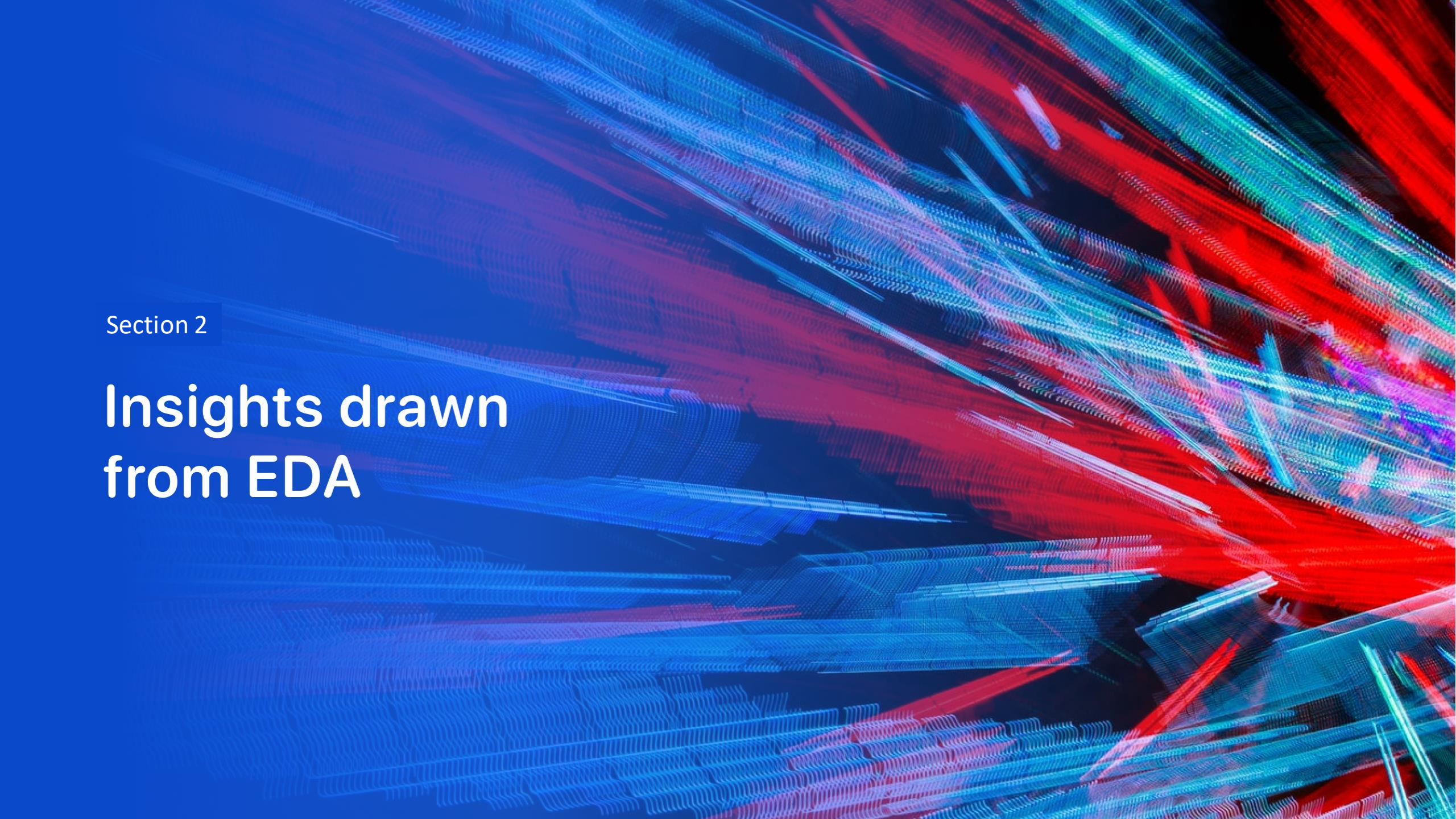
Launch Sites are accessible through roads and railways. However, they are far away from cities and other places of public activity

Predictive analysis results

Almost all the models test have similar accuracy. However, Decision Trees perfom better on the basis of F1 Score

Section 2

Insights drawn from EDA



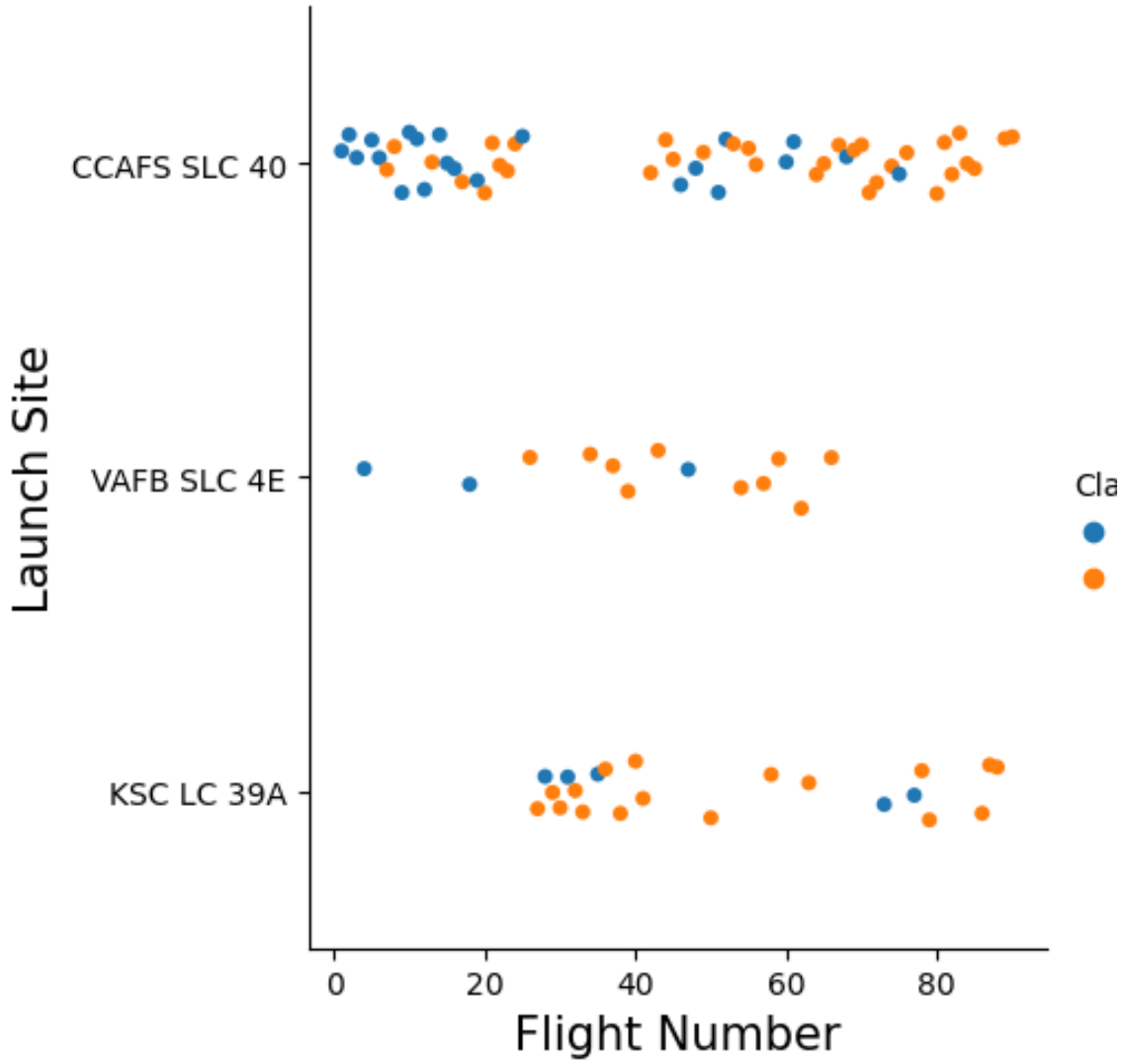
Git hub link

https://github.com/chandana-priya-g/DataScienceTools/blob/main/FinalNotebooks/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite%20(1).ipynb

Flight Number vs. Launch Site

Insights

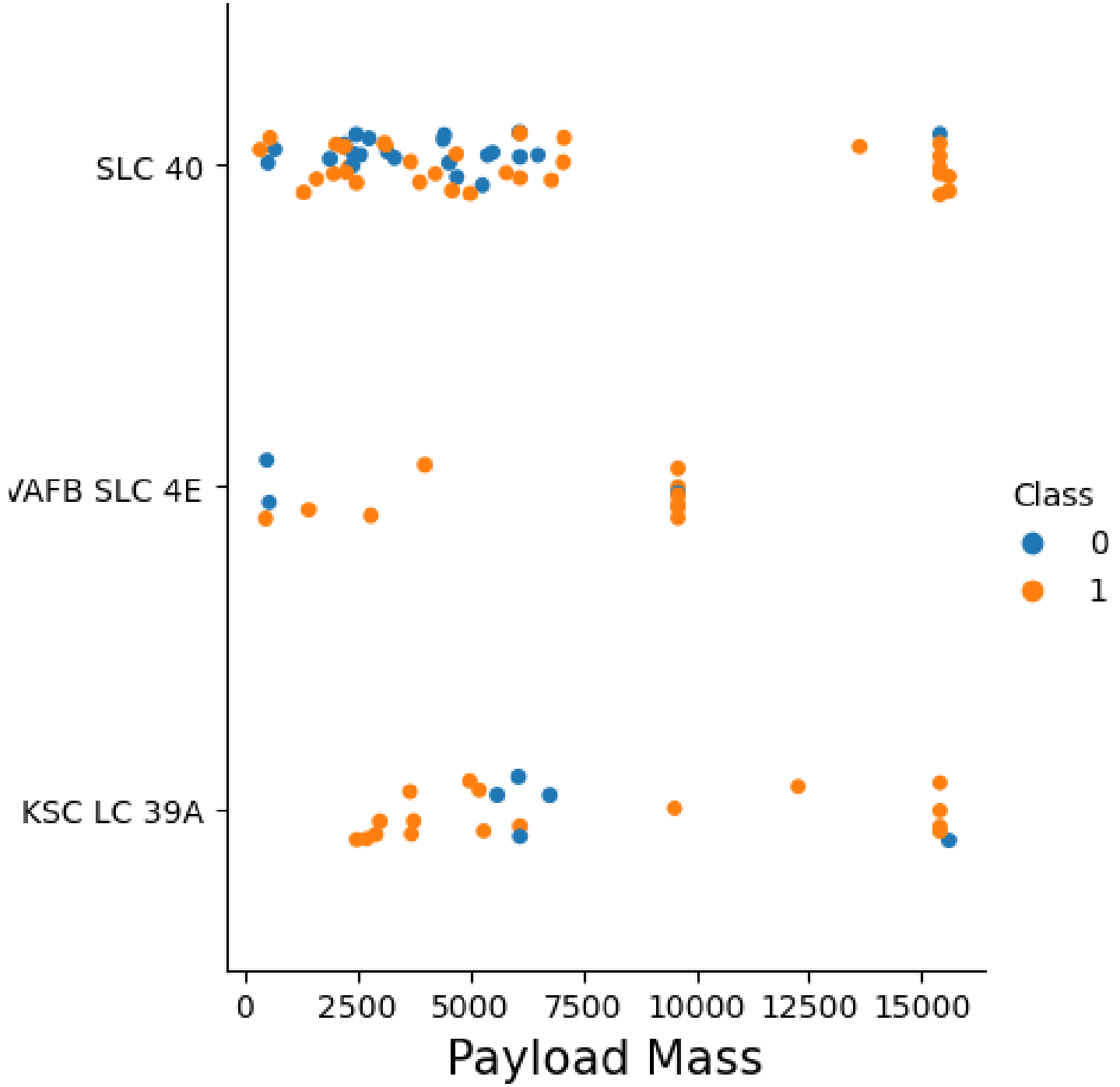
- CCAFS SLC 40 has high launches
- The success rate of later launches are higher
- Lower rate of success for earlier launches



Payload vs. Launch Site

Insights

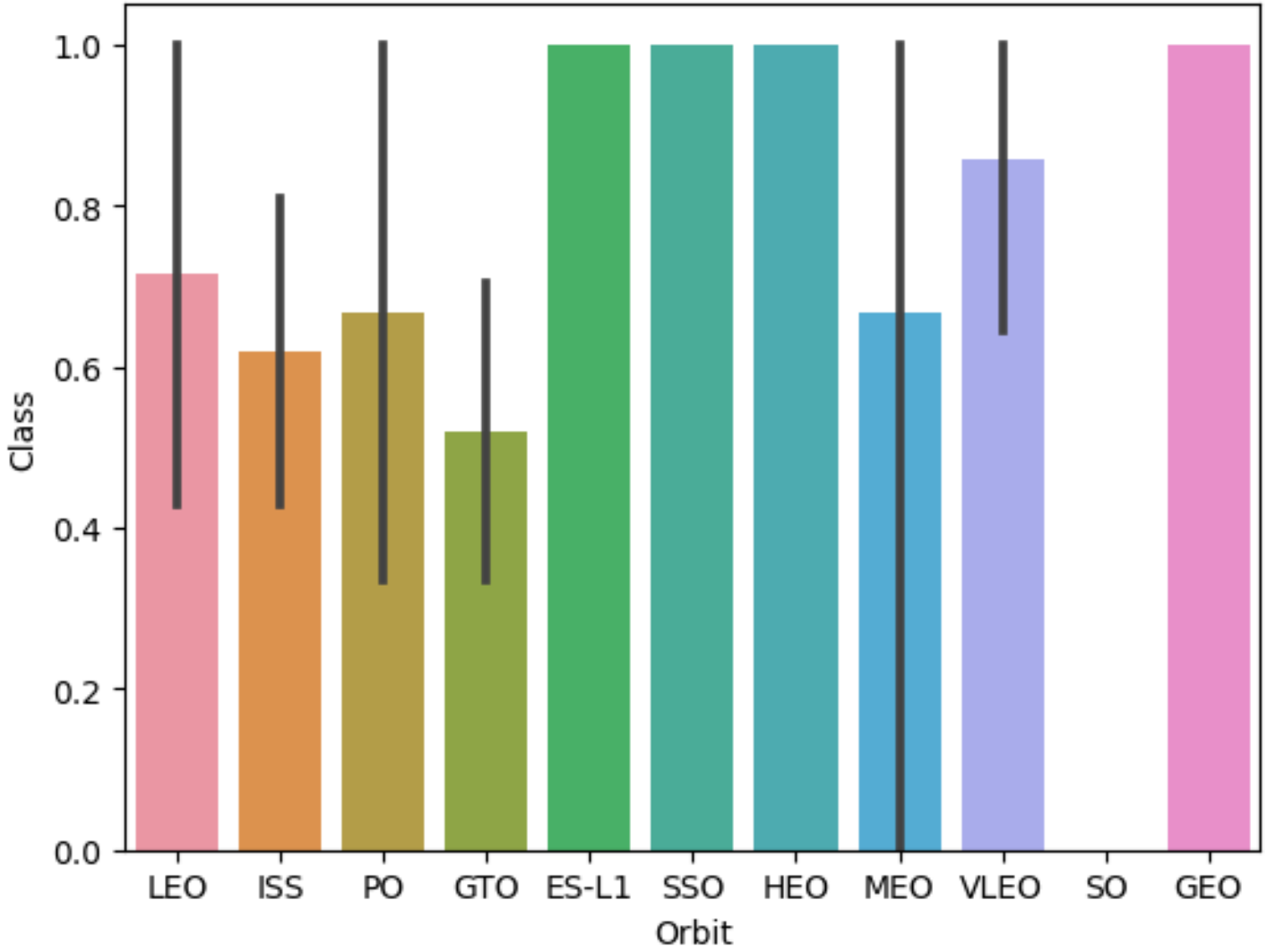
- KSC LC 39A has high success rate for all kinds of payloads similar to VAFB SLC 4E
- CCAFS SLC 40 payload is not a significant attribute in determining success of landing
- 15000 payload capacities have very few failures and 10000 payload capacities have not failures.



Success Rate vs. Orbit Type

Insights

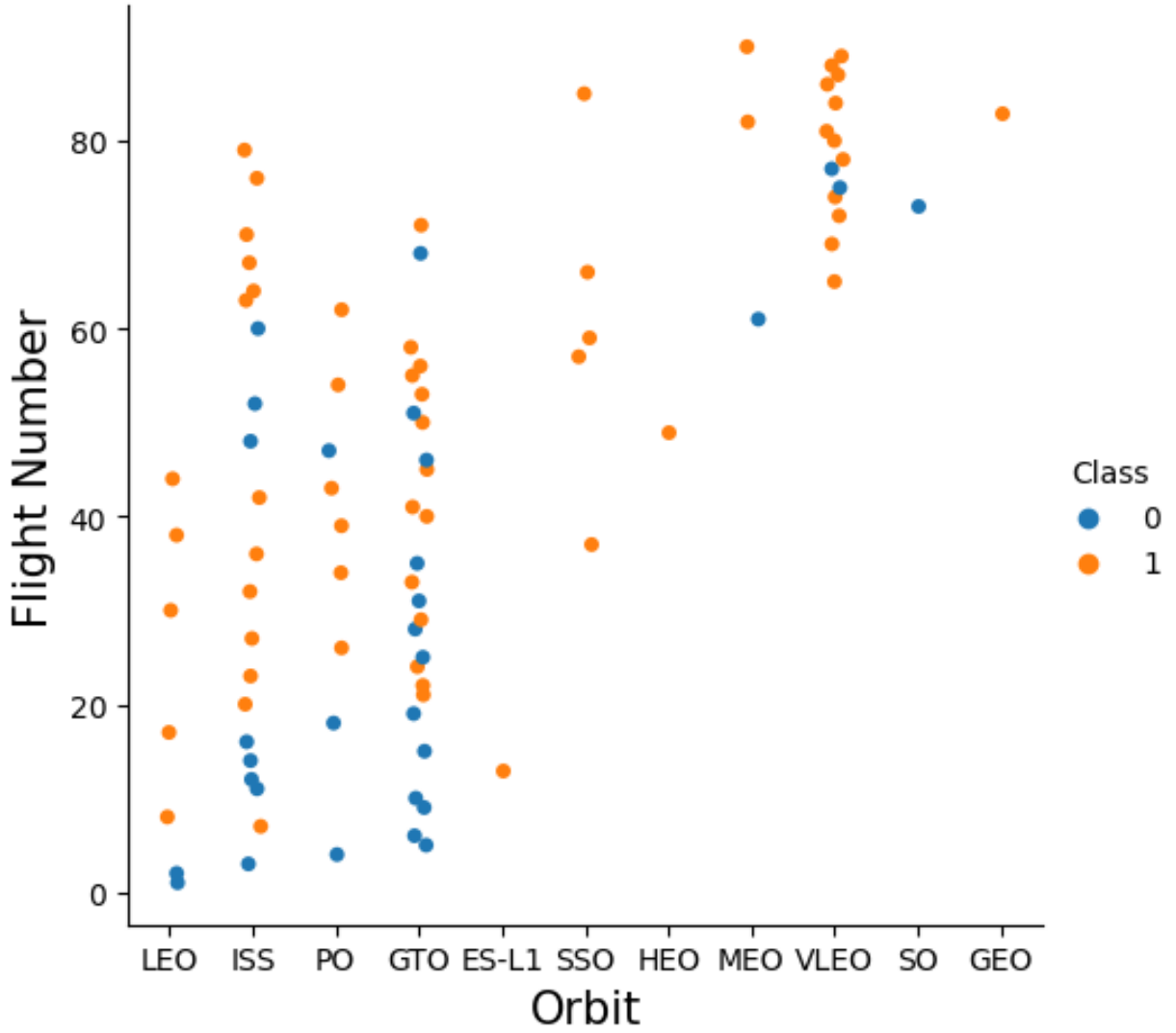
- ES-L1, SSO, HEO, GEO recorded 100% success rate
- SO show 0% success rate

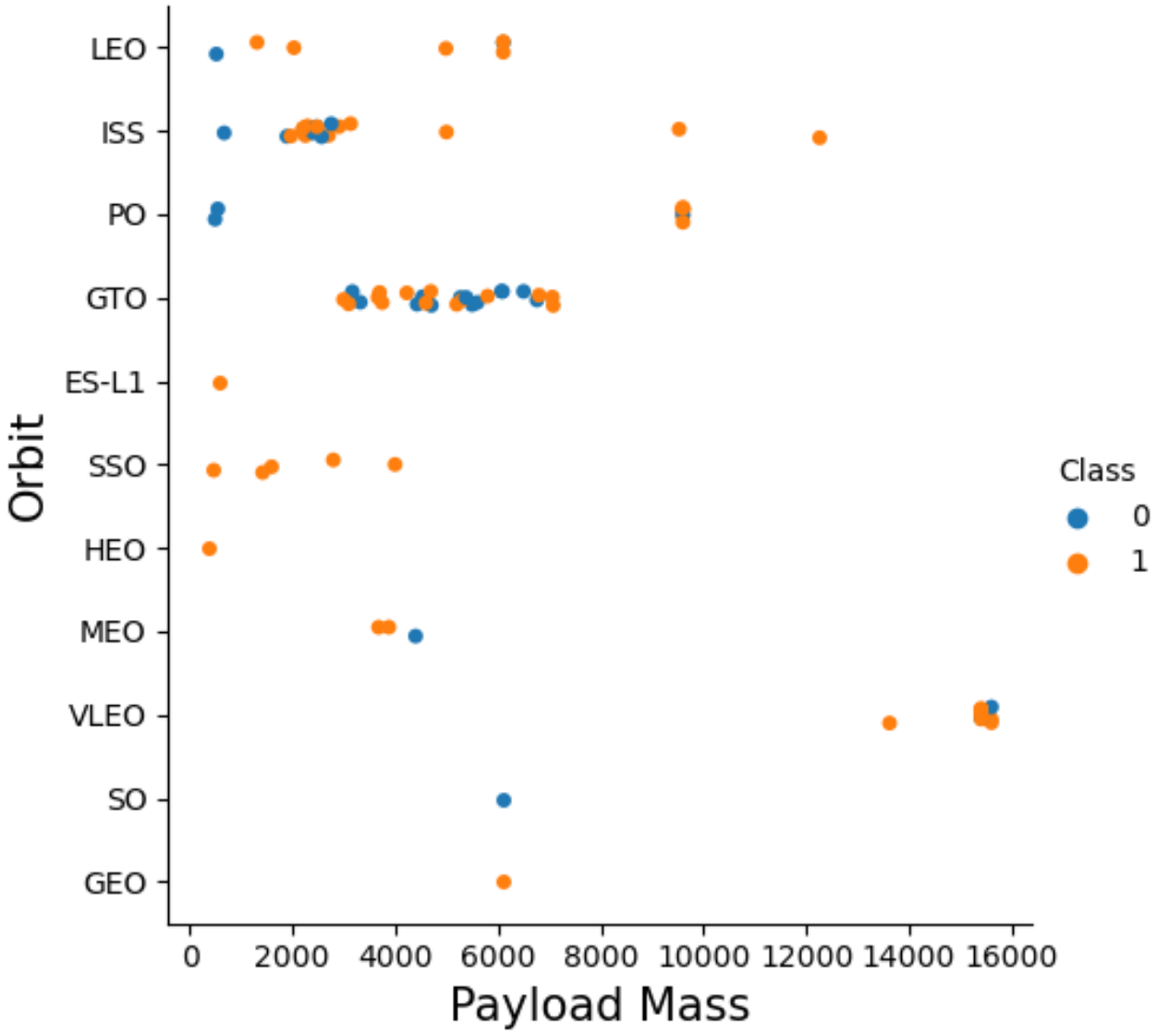


Flight Number vs. Orbit Type

Insights

- The LEO orbit the Success appears to be related to the number of flights
- here seems to be no relationship between flight number when in GTO orbit.





Payload vs. Orbit Type

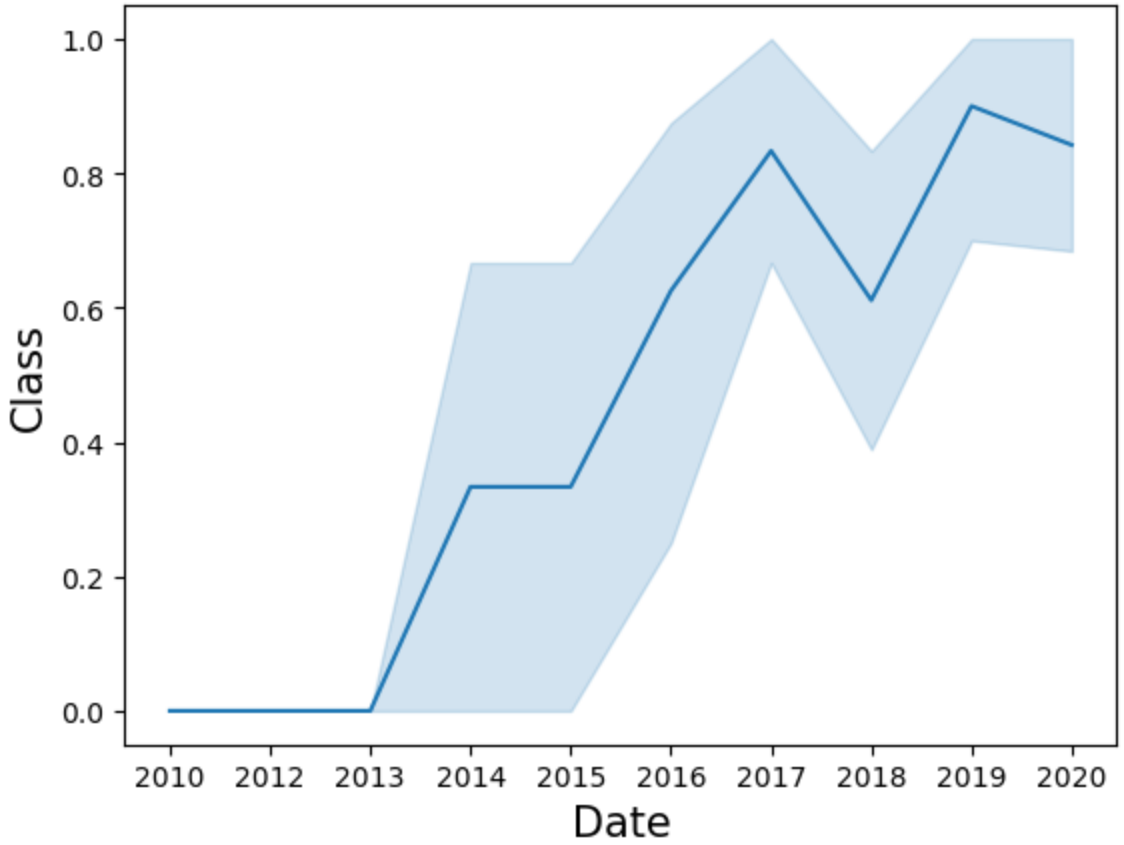
Insights

- Higher payloads witnessed near complete success rate
- Lower payloads show a mix of successful and failed landings
- ES L1, SSO and HEO have 100% success rates

Launch Success Yearly Trend

Insights

- The rate of success increased since2013 till 2020



Git hub link

https://github.com/chandana-priya-g/DataScienceTools/blob/main/FinalNotebooks/jupyter-labs-eda-sql-coursera_sqllite.ipynb

All Launch Site Names

```
%sql select distinct Launch_Site
from SPACEXTABLE
```

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

All the launches happened with in the four launch sites

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEXTABLE WHERE
Launch_Site LIKE 'CCA%' LIMIT 5
```

Done.

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Most launches seem to be a success with the launch site

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM
SPACEXTABLE WHERE Customer like '%NASA (CRS)%'
```

```
 * sqlite:///my_data1.db
Done.
SUM(PAYLOAD_MASS__KG_)
48213
```

NASA accounted for a major portion of Falcon 9 Payload in recent times

Average Payload Mass by F9 v1.1

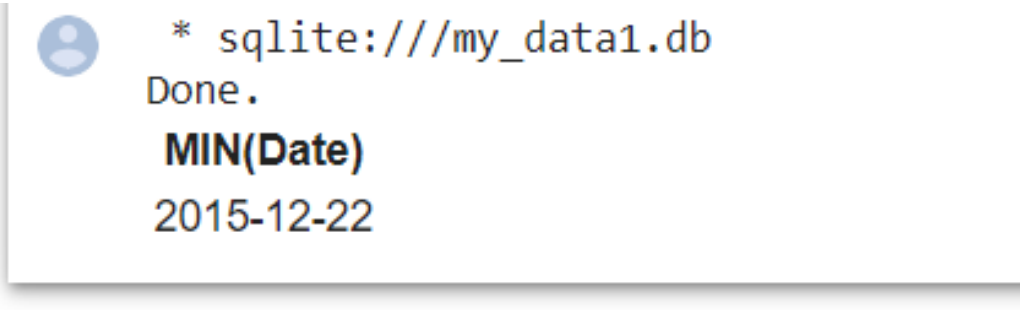
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM
SPACEXTABLE WHERE Booster_Version like '%F9
v1.1%'
```

```
 * sqlite:///my_data1.db
Done.
AVG(PAYLOAD_MASS__KG_)
2534.6666666666665
```

The average payload capacities from Falcon 9 can be termed a medium to low end.

First Successful Ground Landing Date

```
%sql SELECT MIN(Date) FROM SPACEXTABLE
WHERE Landing_Outcome like '%Success
(ground pad)%'
```



The first successful landing was achieved after 2012 where the rate of success started to rise.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT
Booster_Version,Landing_Outcome,PAYLOAD_MASS__KG_
FROM SPACEXTABLE WHERE Landing_Outcome LIKE
'%Success (drone ship)%' AND PAYLOAD_MASS__KG_
BETWEEN 4000 AND 6000
```

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

The number of such landings seem to be still very less. It can be inferred that Drop Ship landing is still not the most popular choice for landing.

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT COUNT(Booster_Version) FROM SPACEXTABLE
WHERE Landing_Outcome like '%Success%'
```

%sql SELECT COUNT(Booster_Version) FROM SPACEXTABLE WHERE Landing_Outcome like '%Failure%'

```
 * sqlite:///my_data1.db
Done.
COUNT(Booster_Version)
61
```

```
Done.
COUNT(Booster_Version)
10
```

The number of successful landing is significantly higher than that of the failed.

Boosters Carried Maximum Pay load

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE
PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_)
FROM SPACEXTABLE)
```

Done.

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

The Maximum payload carrying doesn't seem to be exclusive to fewer boosters.

2015 Launch Records

```
%sql SELECT
Booster_Version,Landing_Outcome,Launch_Site,substr
(Date, 4, 1) as Month FROM SPACEXTABLE WHERE
Landing_Outcome LIKE '%Failure (drone ship)%'
```

Booster_Version	Landing_Outcome	Launch_Site	Month
F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40	5
F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40	5
F9 v1.1 B1017	Failure (drone ship)	VAFB SLC-4E	6
F9 FT B1020	Failure (drone ship)	CCAFS LC-40	6
F9 FT B1024	Failure (drone ship)	CCAFS LC-40	6

Most of them are from CCAFS LC-40

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

```
%sql SELECT Landing_Outcome, COUNT() AS Frequency
FROM SPACEXTABLE WHERE Date BETWEEN "2010-06-04" AND
"2017-03-20" GROUP BY Landing_Outcome Order by
Frequency desc
```

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

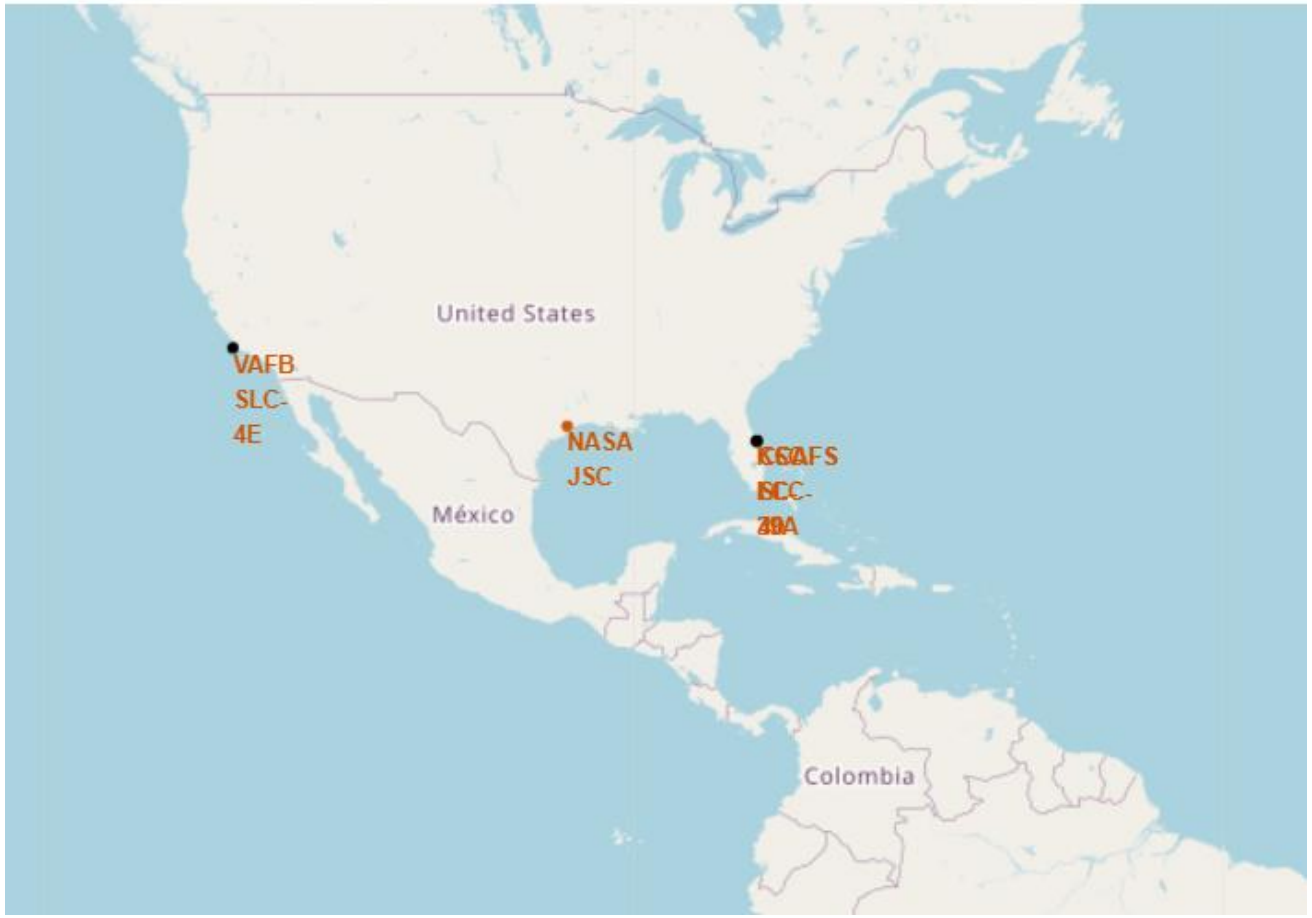
While the landing doesn't seem to be attempted most of the times, possibly for reasons such as fuel, system settings etc, most successful landings were on the ground pad. Though drone ship has similar success count, it has equal number of failures. Other options such as parachutes doesn't seem to be chosen a lot.


Section 3
Launch Sites
Proximities Analysis

Git hub link

https://github.com/chandana-priya-g/DataScienceTools/blob/main/FinalNotebooks/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Geographic representation of Launch Sites



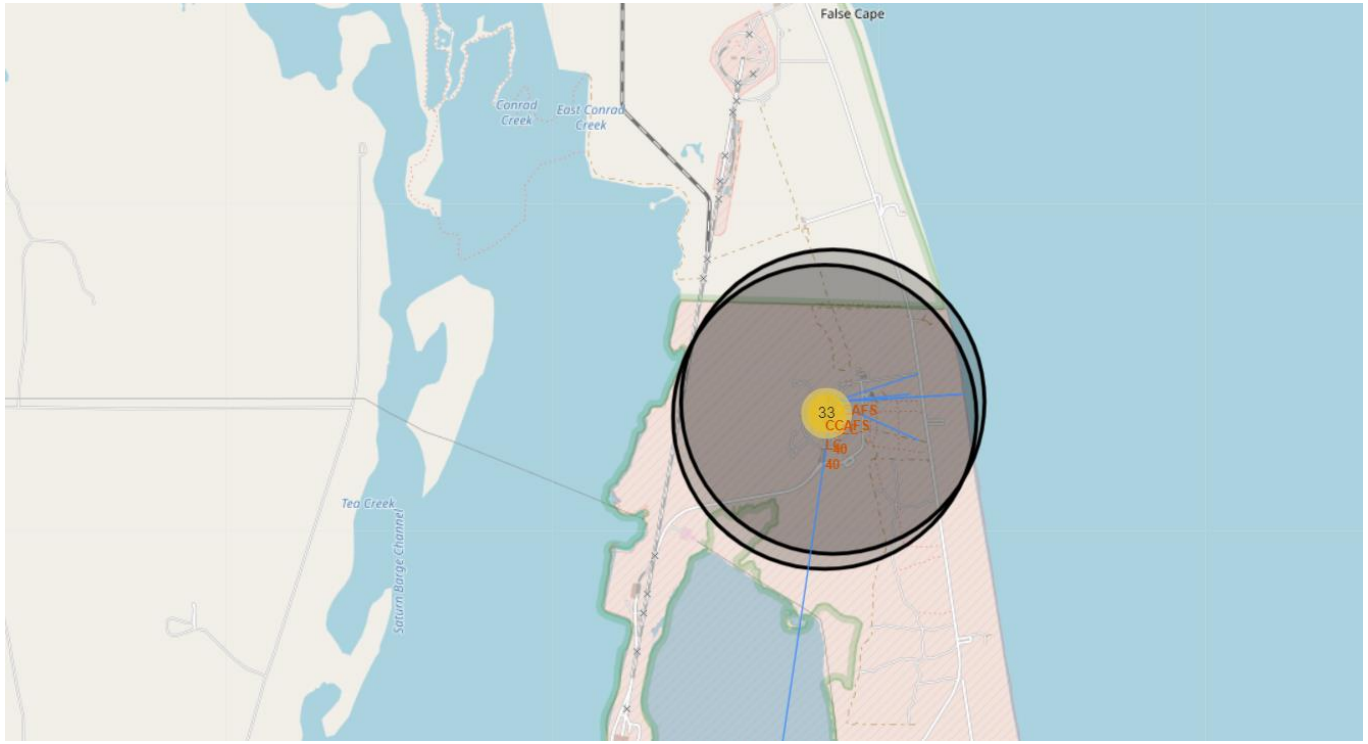
All the launch site are close to the coastline

Marking the success/failed launches for each site on the map

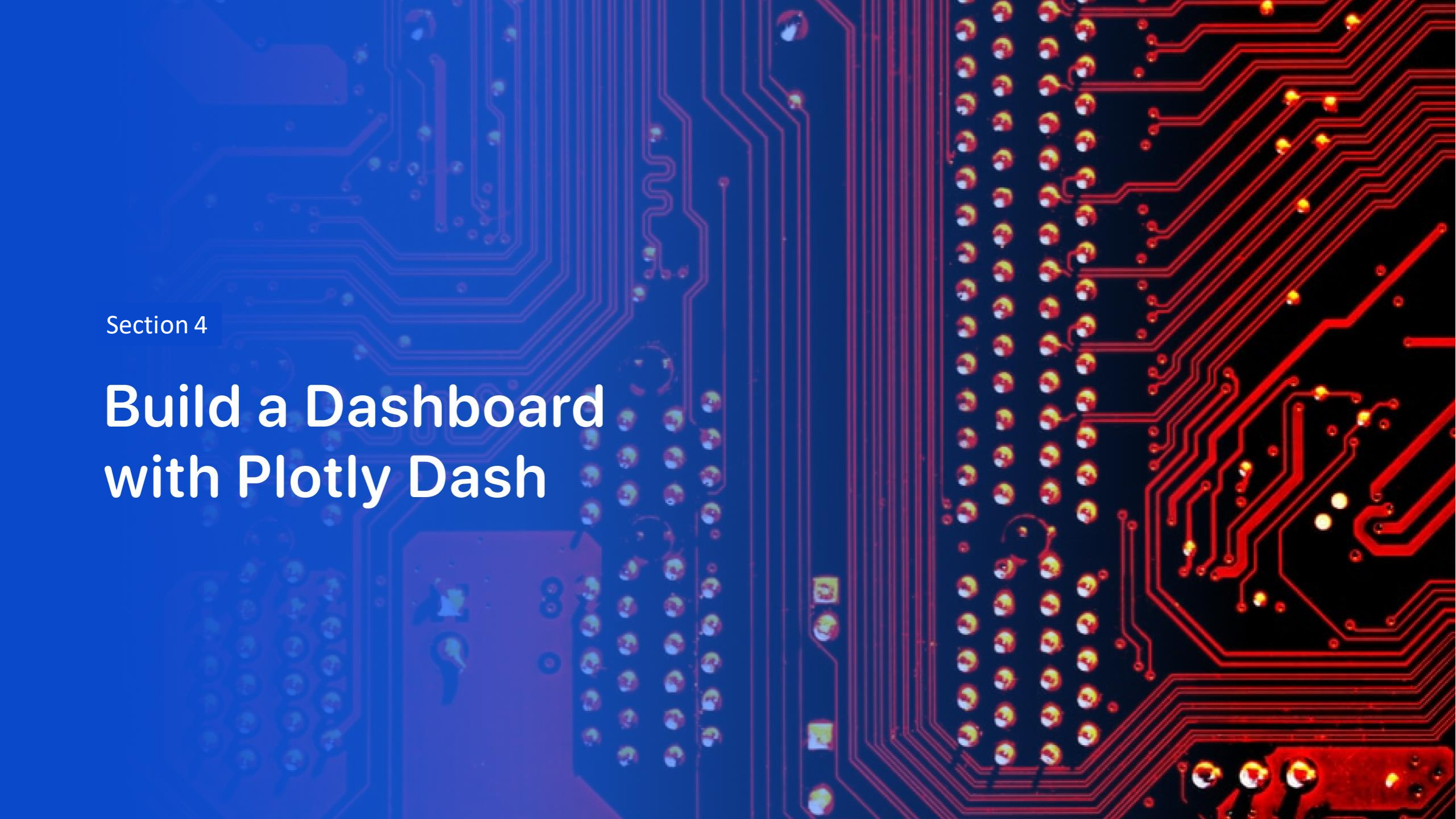
	Launch Site	Lat	Long	class
46	KSC LC-39A	28.573255	-80.646895	1
47	KSC LC-39A	28.573255	-80.646895	1
48	KSC LC-39A	28.573255	-80.646895	1
49	CCAFS SLC-40	28.563197	-80.576820	1
50	CCAFS SLC-40	28.563197	-80.576820	1
51	CCAFS SLC-40	28.563197	-80.576820	0
52	CCAFS SLC-40	28.563197	-80.576820	0
53	CCAFS SLC-40	28.563197	-80.576820	0
54	CCAFS SLC-40	28.563197	-80.576820	1
55	CCAFS SLC-40	28.563197	-80.576820	0

The KSC LC-39A site records successful landings through out

Launch Site Proximities



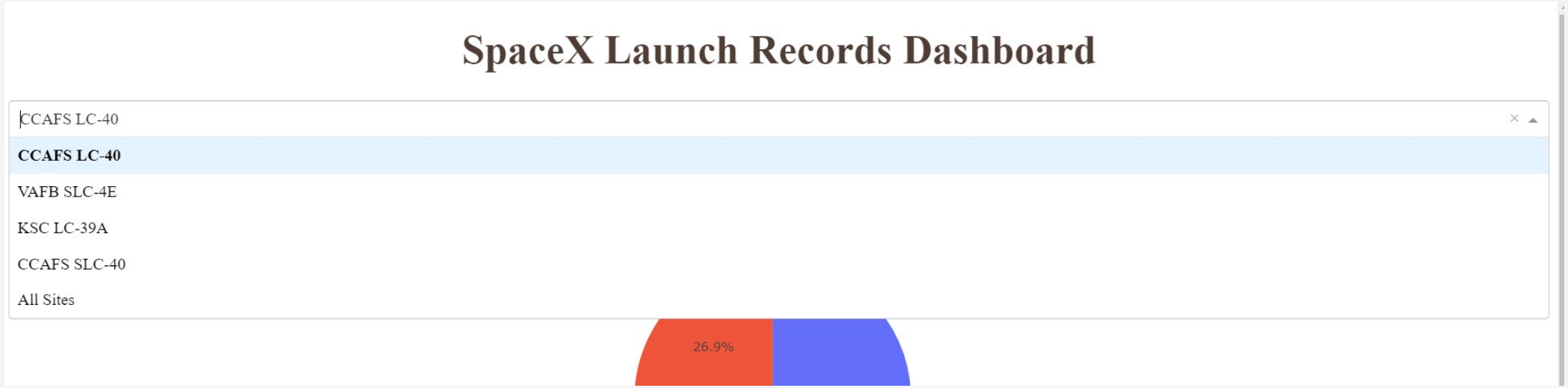
The distance between launch sites with the roads, railways and coastal line has been calculated.


Section 4
Build a Dashboard with Plotly Dash

Git hub link

https://github.com/chandana-priya-g/DataScienceTools/blob/main/finaldataviz.txt

Interactive Dashboard



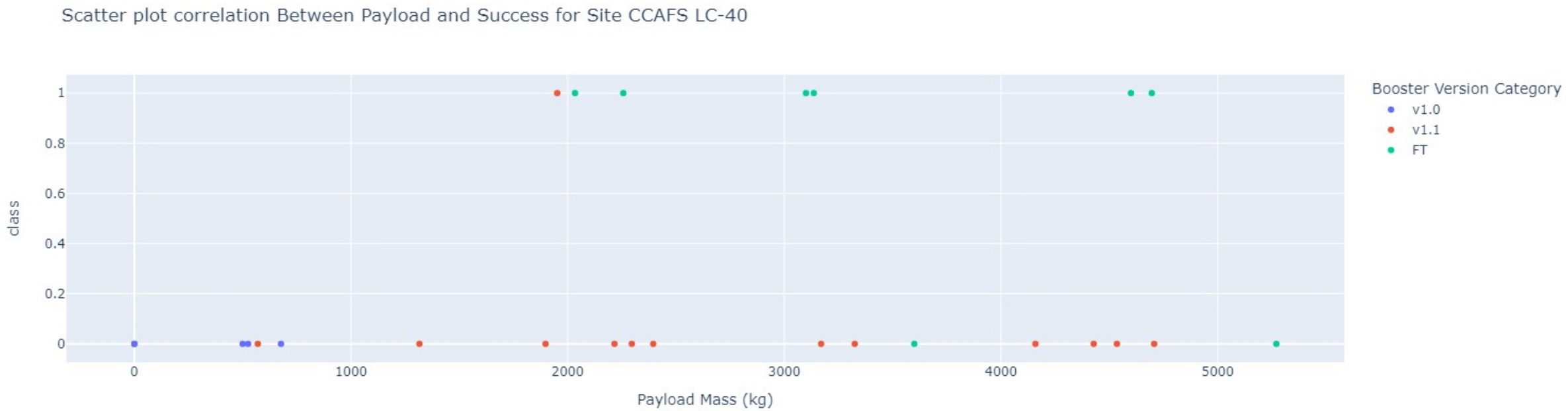
The launch sites are provided as a dropdown list according to which pie chart regarding their success rate are generated

Successful Landing with respect to launch sites

According to the pie chart KSC LA-39A recorded highest successful landings. Other sites have comparable rates.

Payload Scatter plot

A scatter plot regarding payload and success rate is generated for each selected launch site.



Section 5

Predictive Analysis (Classification)

Git hub link

https://github.com/chandana-priya-g/DataScienceTools/blob/main/FinalNotebooks/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Classification Accuracy

All models have similar accuracy while the F Score of decision tree is higher

	Algorithm	Accuracy	F Score
0	Logistic Regression	0.833333	0.888889
1	K Nearest Neighbour	0.833333	0.888889
2	Support Vector Machine	0.833333	0.888889
3	Decision Tree	0.833333	0.960000

Confusion Matrix

Decision tree is the best performing model. It shows zero false negatives. The percentage of correctly predicted values is higher, with accuracy of 83.33%

Conclusions

THE LANDING SUCCESS RATE HAS SIGNIFICANTLY INCREASED IN THE PAST DECADE

THE ABILITY TO PREDICT COST OF LAUNCHES IS SET AT 83.33%

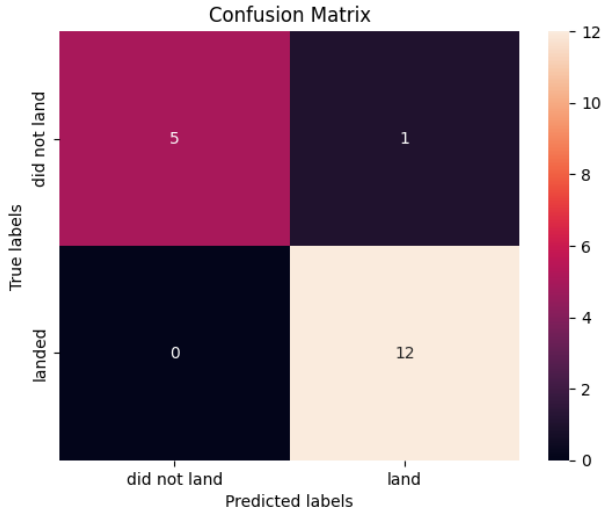
ORBIT LEVEL SEEMS TO PLAY A SIGNIFICANT ROLE ON THE SUCCESS OF LANDING.

SOME LAUNCHING SITES RECORDED 100% SUCCESS. THIS COULD MEAN FAVORABLE GEOGRAPHICAL CONDITIONS.

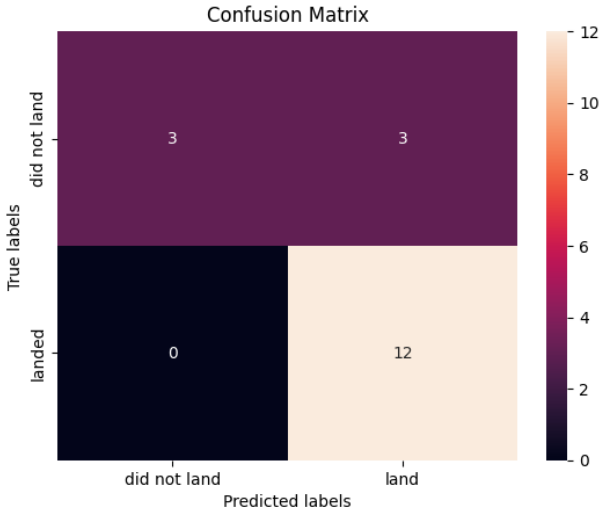
Appendix

Confusion matrices of other models

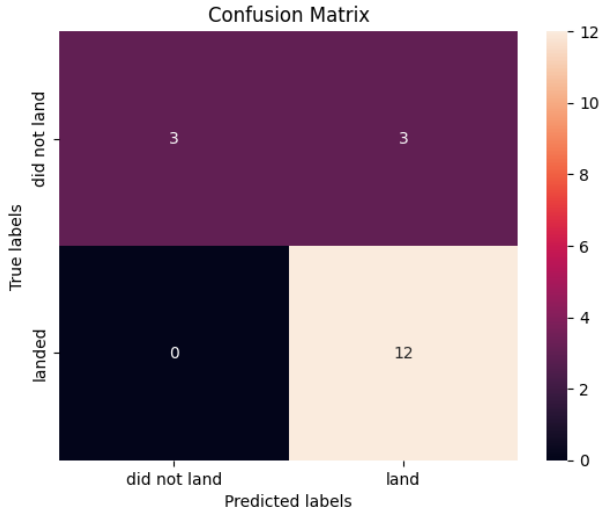
Logistic Regression



SVM



KNN



Thank you!