

Final project on data science company- space y

SAHIL BANERJEE

19-06-2023

Outline:-

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

Executive Summary

- Methodologies that were used in this project are:
 1. Web scrapping from space x website- It helped in collecting huge amount of unstructured data
 2. Exploratory data analysis- It helped in which to analysis which rocket launched successfully
 3. Machine learning- It helped us to predict the optimum classifiers which can be used in our project

Introduction:

- Our main objective is to compute whether or not will the first stage of SPACE X rocket will land safely as it is our competitor of company SPACE Y

Methodology



Data collection

- Data in this project has been primarily been extracted from the SPACE X api(<http://api.spacex.com/v4//rockets/>) and from wikipedia from using web scraping techniques using BeautifulSoup

Data Wrangling

- Initially we perform some EDA(exploratory analysis) on the given dataset to make predictions on launch_sites and all columns
- After that we will give landing outcome through exploratory analysis

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
4	2010-06-04	Falcon 9	NaN	LEO	CCSPFS SLC 40	None	1	False	False	False
5	2012-05-22	Falcon 9	525.0	LEO	CCSPFS SLC 40	None	1	False	False	False
6	2013-03-01	Falcon 9	6070.0	ISS	CCSPFS SLC 40	None	1	False	False	False
7	2013-09-29	Falcon 9	500.0	PO	WAFB SLC 4E	False Ocean	1	False	False	False
8	2013-12-03	Falcon 9	3170.0	GTO	CCSPFS SLC 40	None	1	False	False	False
--	--	--	--	--	--	--	--	--	--	--
89	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	Set
90	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	Set
91	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	Set
92	2020-10-24	Falcon 9	15600.0	VLEO	CCSPFS SLC 40	True ASDS	3	True	True	Set

EDA with Data Visualization

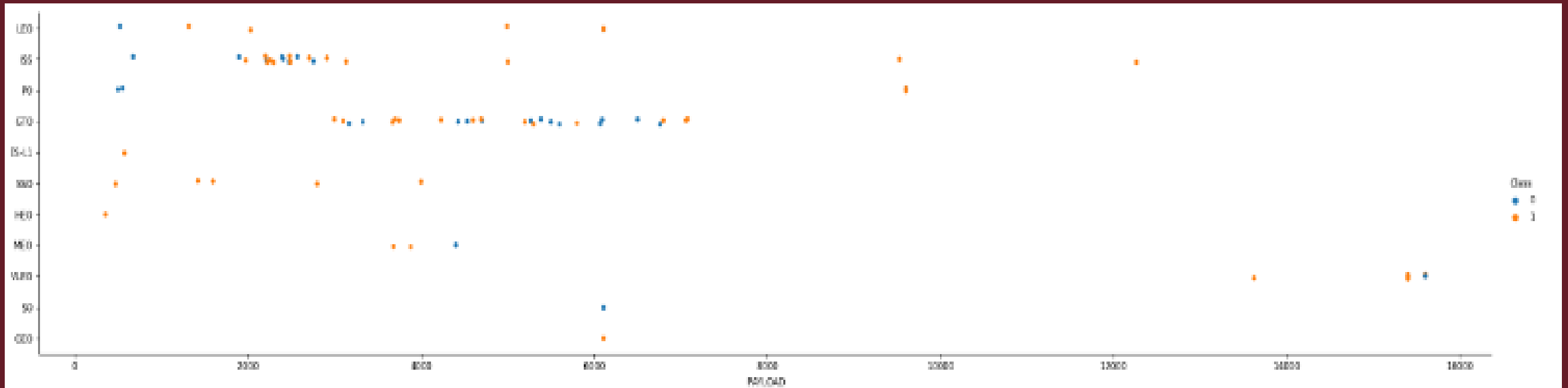
- To find the relationship between various terms bar graphs, scatterplots were used to show the relationship

EDA with SQL

- We performed various steps with the help of SQL to clean out and manipulate the following data. Which may include-
- Unique names of the launch sites in the space mission
- Top 5 launch sites in which the name begins with 'CCA'
- Total number of successful and failure outcomes
- Rank of the landing outcomes and many more..

Relation between payload and orbit-

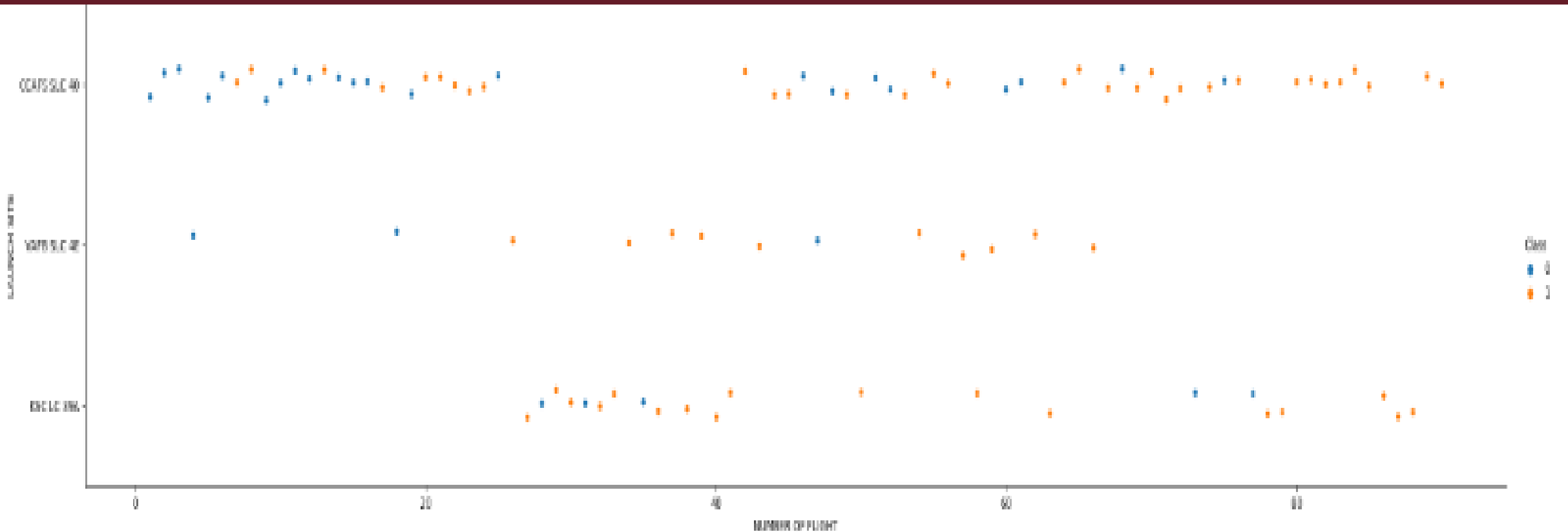
- In this slide we discuss about how we created scatterplot between payloadmass of the booster with that of the orbit



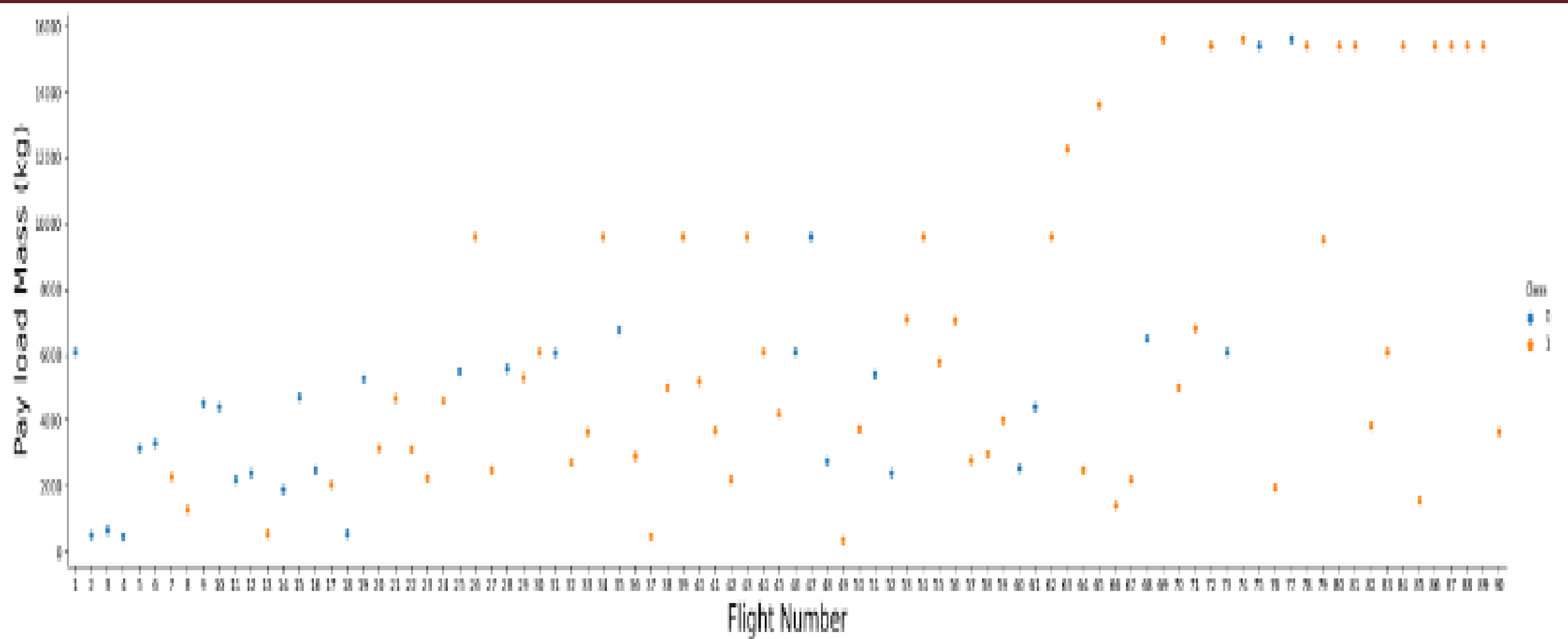
Success Rate vs Orbit type-

- The biggest success rates happens to orbits-
- ES-I1
- GEO
- HEO followed by
- SSO and LFO

Relation between launch site and flight number



Relation between payload and flight number



Launch Site names-

- All launch site names are-
- CCFAS LC-40
- CCAFS SLC-40
- KSC LC-39A
- VAFB SLC-4E

Launch sites name with CCA(5)

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	M
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	

Total payload mass

- Total payload mass carried out by NASA booster is
- 111.268 kg
- Average payload mass by F9 v1.1
- 2.928 kg

Landing outcome

- First successful landing outcome was accomplished on
- 2015-12-22

Payload mass carried out by NASA(CRS)

SUM(PAYLOAD_MASS_KG_)

45596.0

Total number successful and failure outcomes

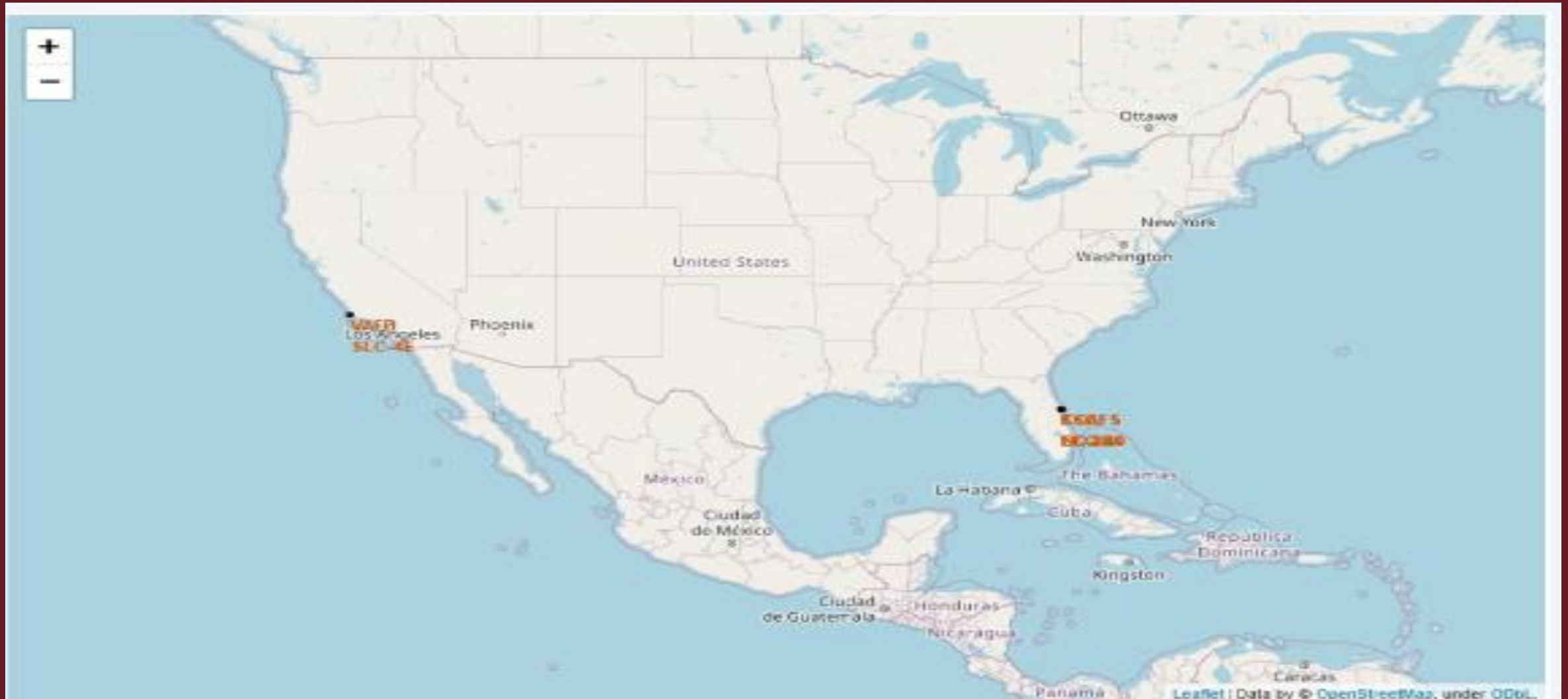
```
COUNT(MISSION_OUTCOME)
```

99

Launch sites with EDA



Launch sites are near water and far away from roads in order to avoid accidents



KSC LC-39A

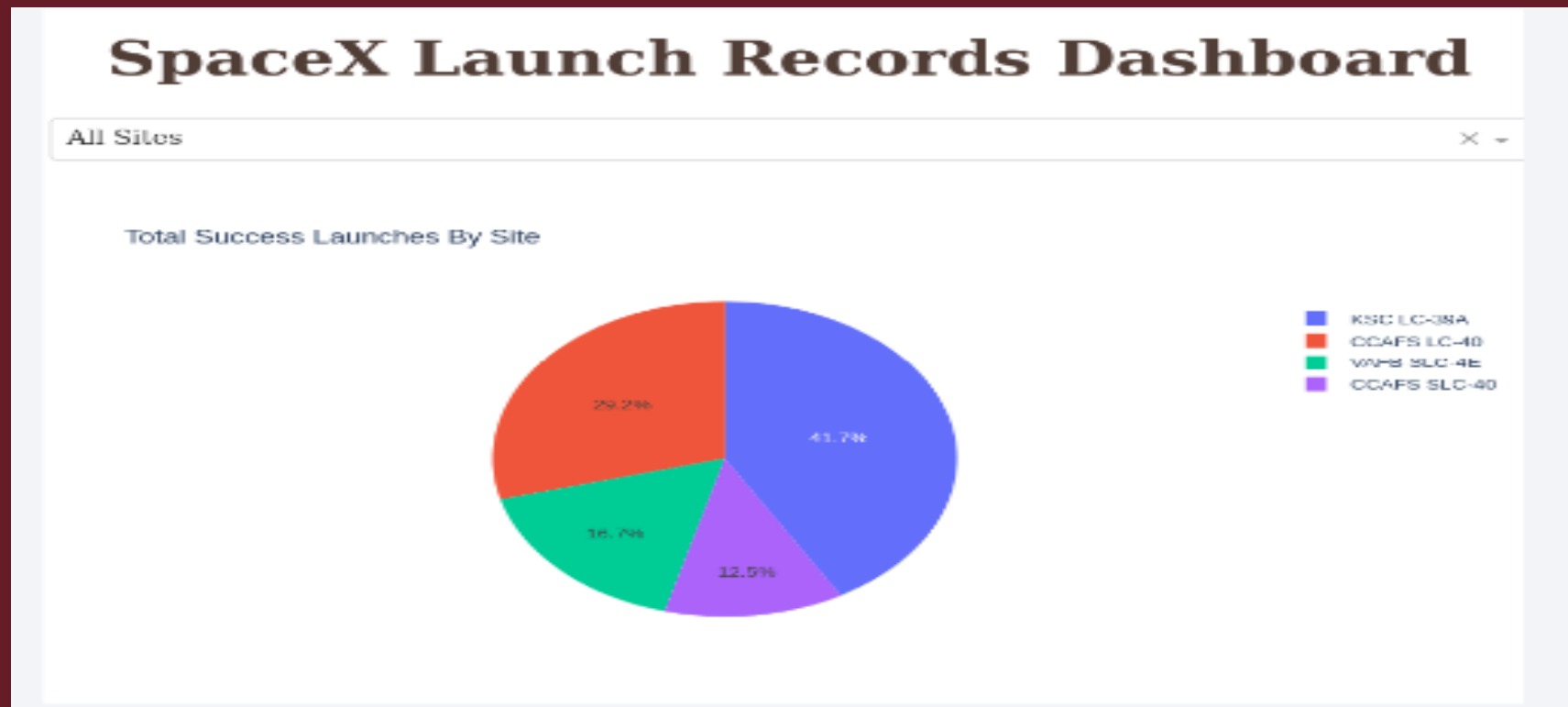
- This launch site is more suitable as it near to railroad and road and far from inhabited areas



DASHBOARD WITH PLOTLY DASH



- We have created a dash dropdown so that we can analysis our problem in much more efficient way

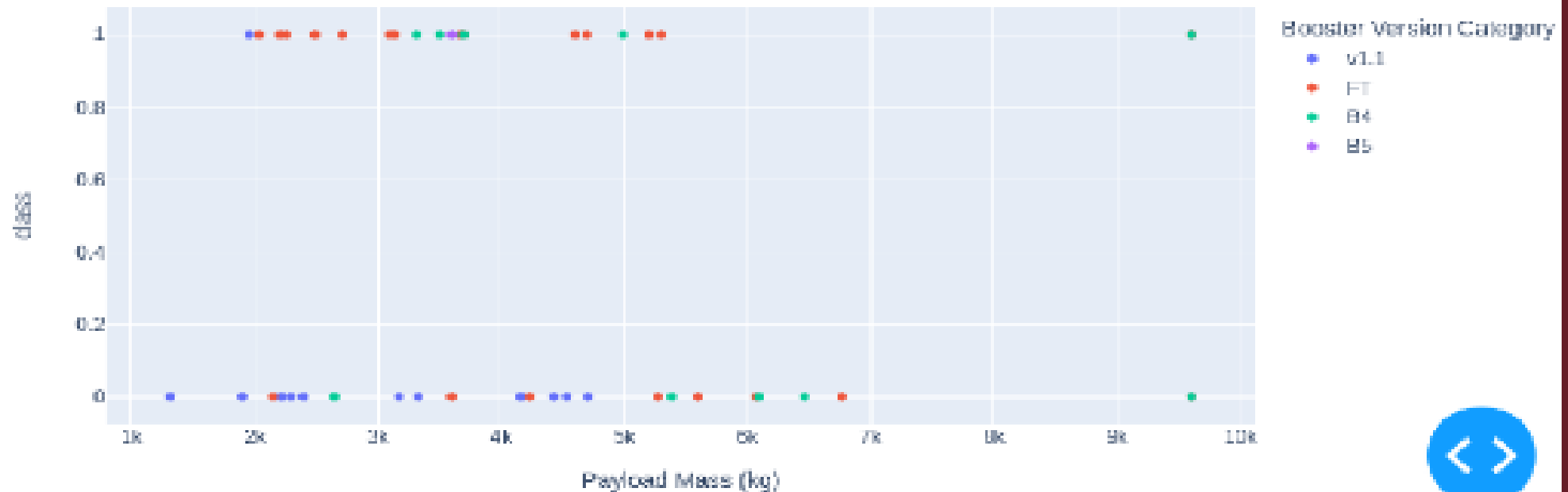


Payload vs Launch Outcome

Payload range (Kg):

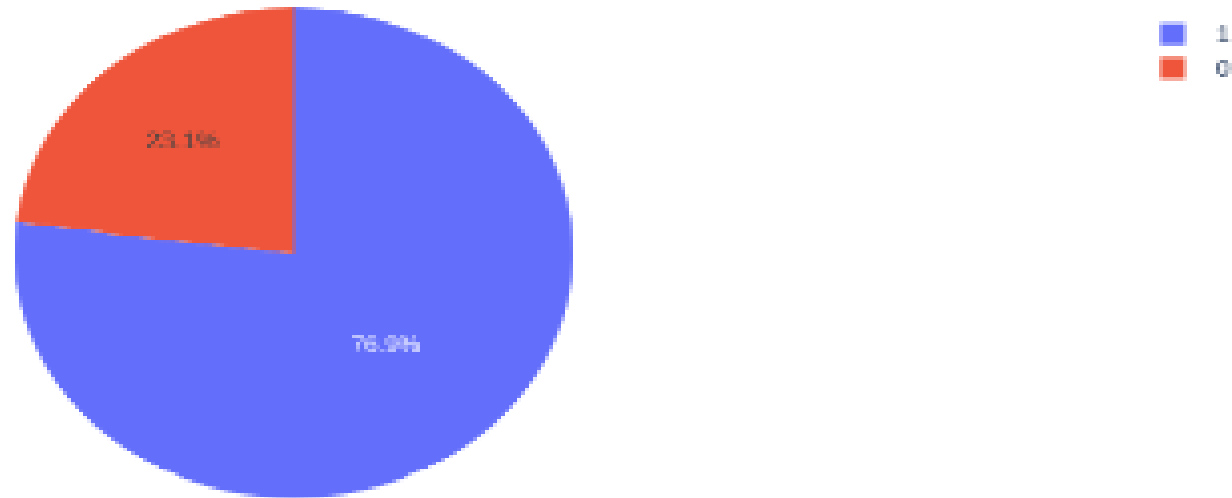


All sites - payload mass between 1,000kg and 10,000kg



Launch success ratio KSC LC- 39A

Total Launches for site KSC LC-39A



PREDICTIVE ANALYSIS

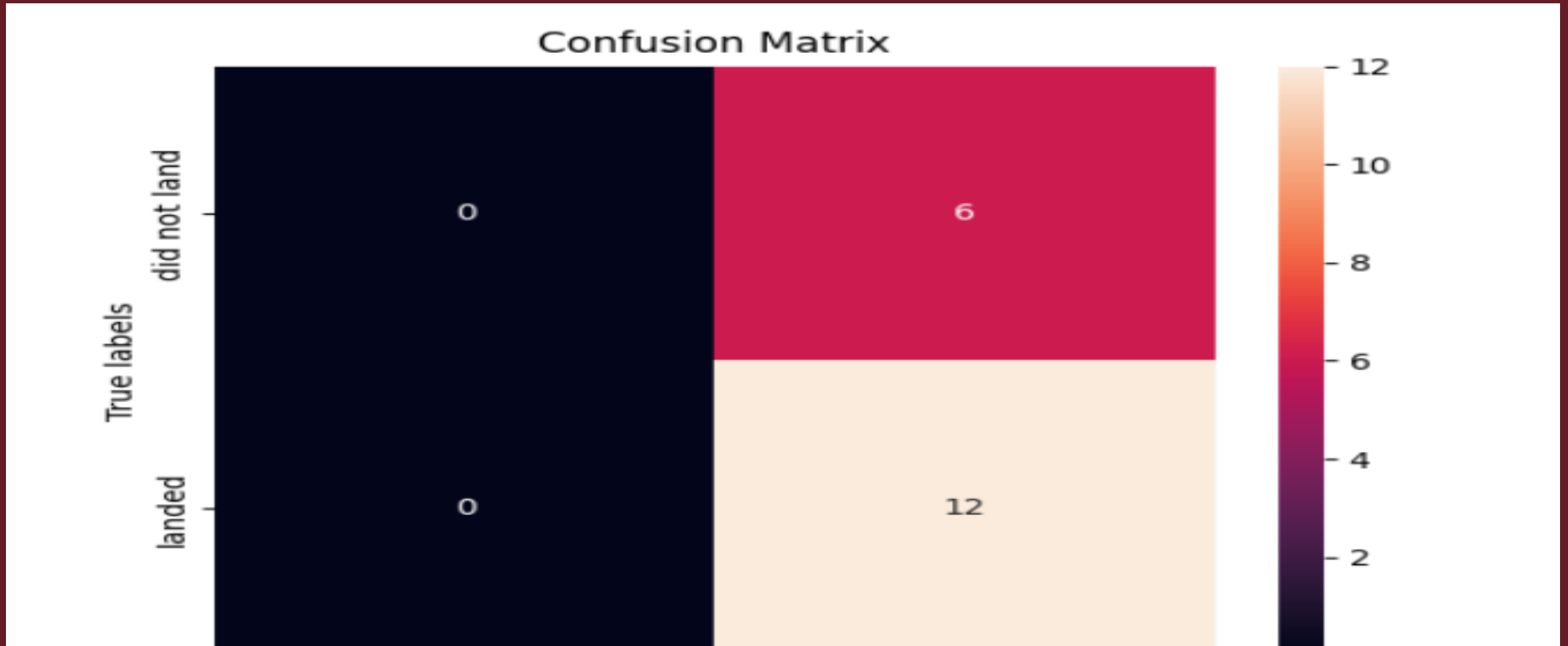


Classification accuracy

- Here decision tree classifier is more viable because it has higher accuracy


Model	Accuracy	TestAccuracy
LogReg	0.8464285714285713	0.8333333333333334
SVM	0.8482142857142856	0.8333333333333334
Tree	0.8892857142857145	0.6666666666666666
KNN	0.8482142857142858	0.8333333333333334

Confusion matrix of Decision tree classifier



CONCLUSION



- 
- The best launch site is KSC LC-39A
 - Launches that have a payload mass of more than 7000 kg are less risky to launch
 - We can use decision tree classifier for calculation of landing