IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Himanshu Goswami
10-02-2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Built interactive map with Folium
  - Built Dashboard with Plotly Dash
  - Predictive Analysis (Classification)
- Summary of all results
  - EDA Result
  - Interactive Analytics
  - Predictive Analysis

Introduction

- Project background and context
  - The commercial space age is here, and companies are making space travel affordable for everyone. One of them is SpaceX which is appearing as revolutionary.
  - SpaceX advertises Flacon 9 rocket launches on its website, with a cost of 62 million dollars whereas other charge around 165 million dollars which is way too much if we compare.
  - It's possible because SpaceX can reuse the first stage of rocket launch.
- Answer we are trying to find
  - Predict whether the first stage of SpaceX Falcon 9 rocket will land successfully?
  - What are the factors that impact the success and failure of the landing?

Section 1

Methodology

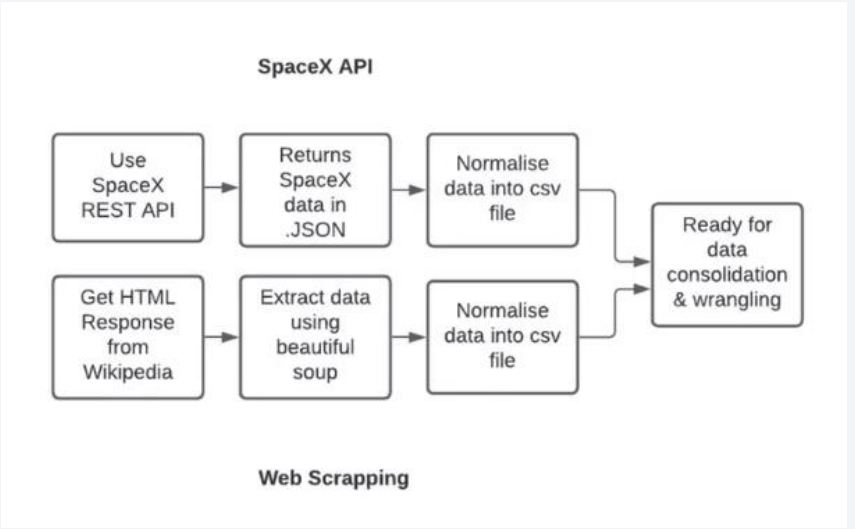
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - One Hot Encoding data fields for Machine Learning Algorithms.
  - Data cleaning - Removed 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
  - Logistic Regression, KNN, SVM and Decision Tree, these 4 models were built and evaluated for best model out of all

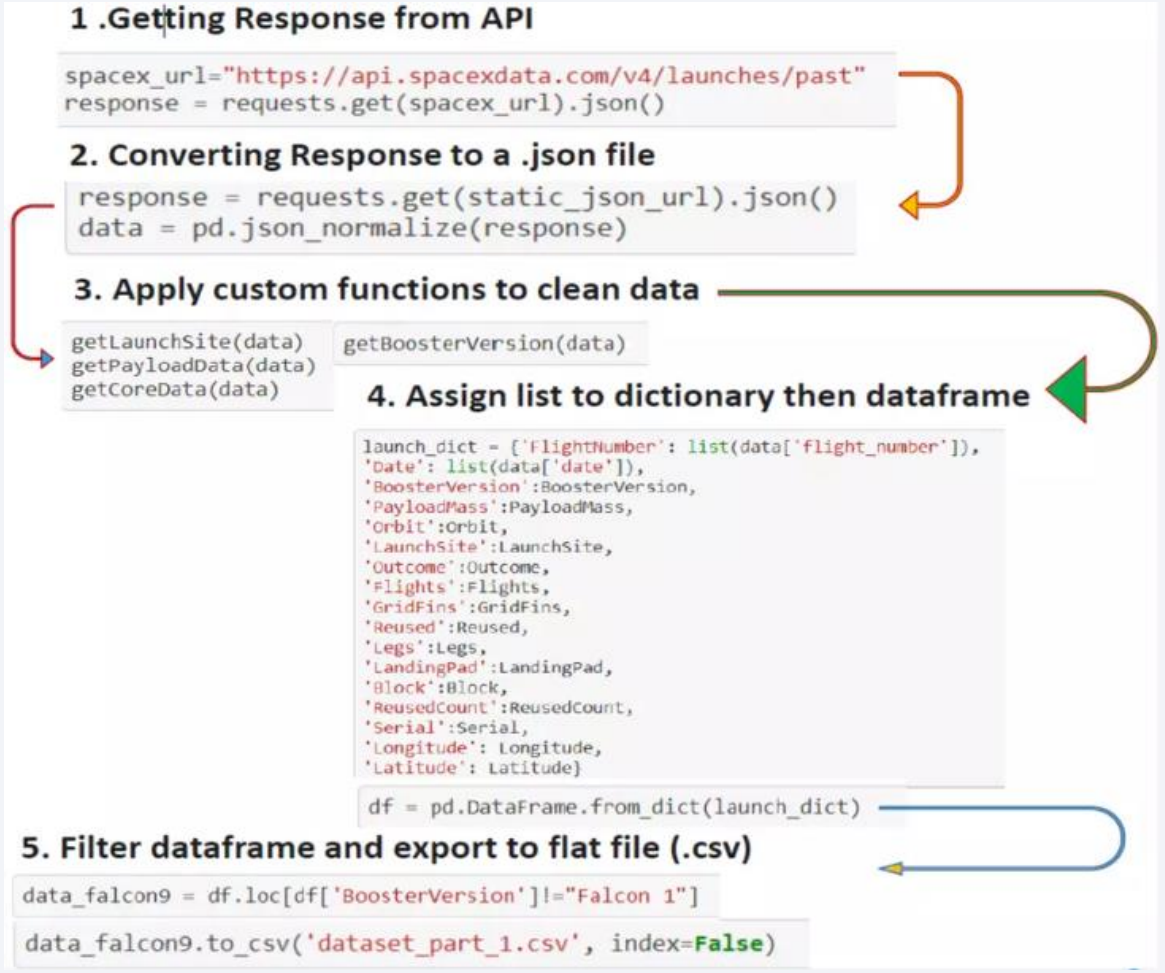
Data Collection

- The following datasets were collected:-
  - SpaceX launch data that was gathered from SpaceX Rest API.
  - This API gave us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications and landing outcome.
  - The SpaceX REST API endpoint, or URL, starts with api.spacexdata.com/v4/
  - Another popular data source for obtaining Falcon 9 Launch data is web scraping using Wikipedia using Python Package – BeautifulSoup.



Data Collection – SpaceX API

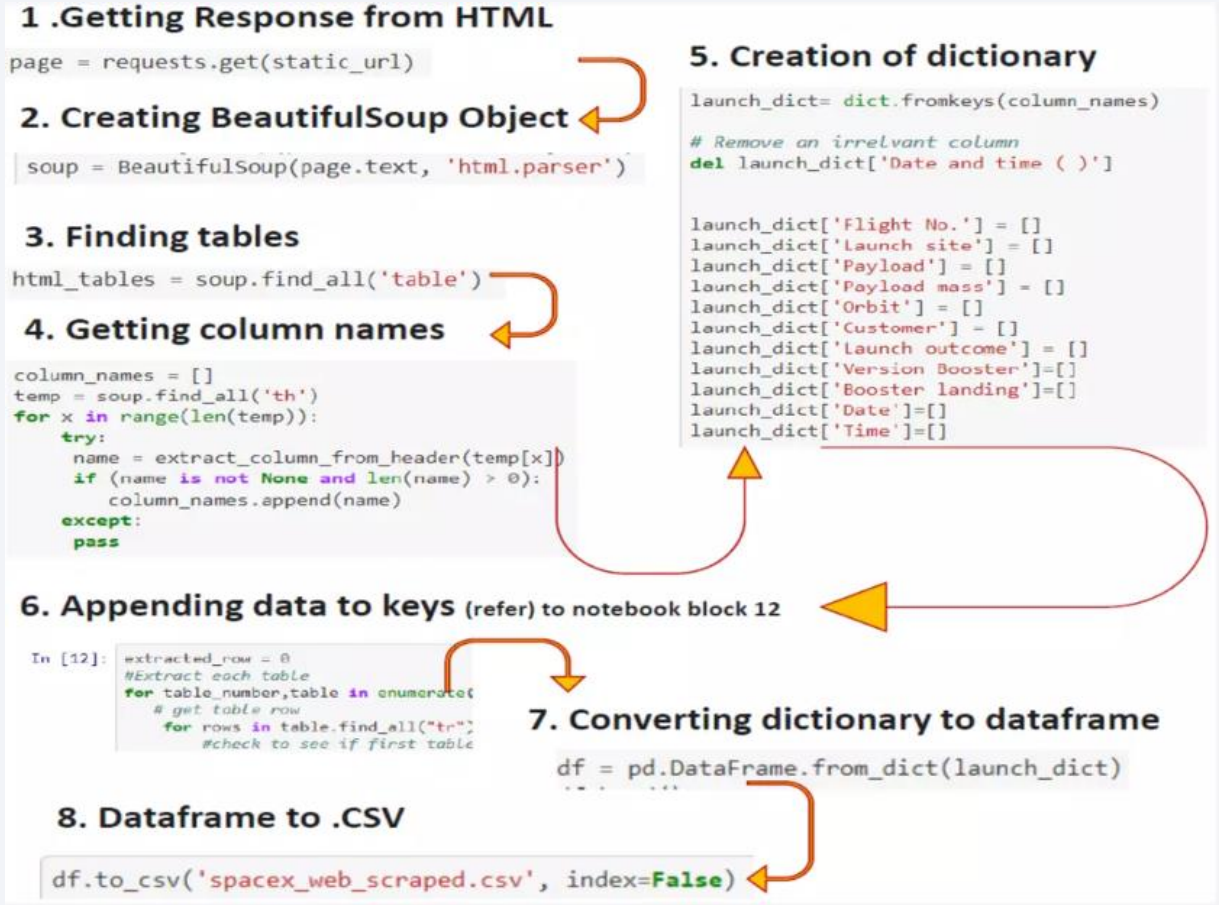
- Data collection with SpaceX REST calls.
- GitHub URL – Click Here



Data Collection - Scraping

- Web Scrapping from Wikipedia

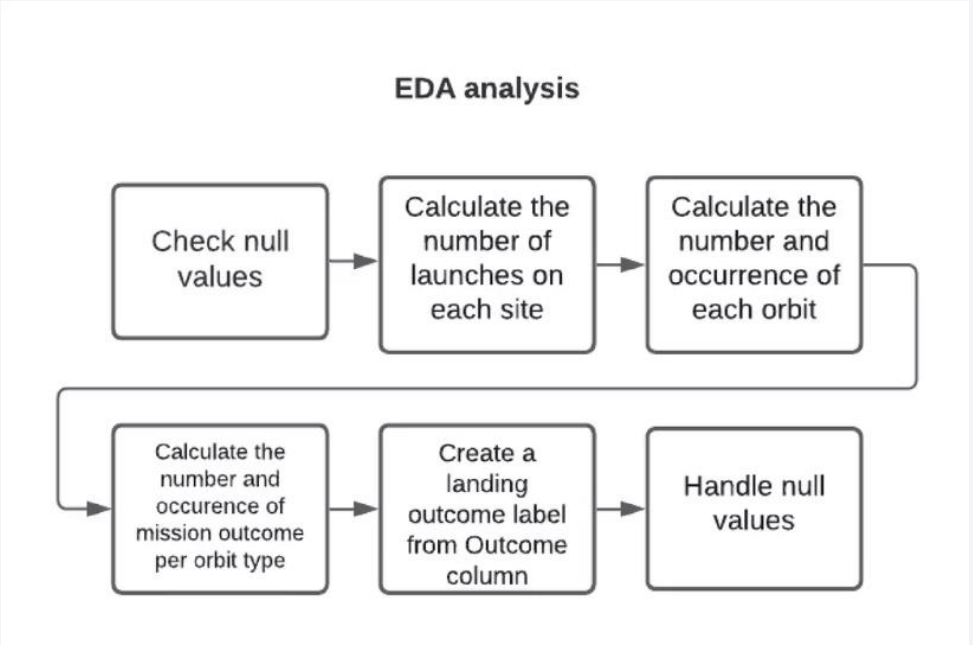
- GitHub URL – [Click Here]



Data Wrangling

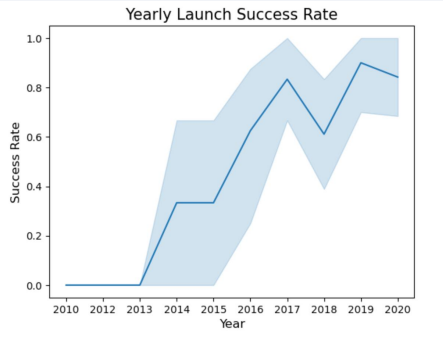
- Checked Null Values
- Performed EDA on various features
- Handled Null Values
- Removed irrelevant features

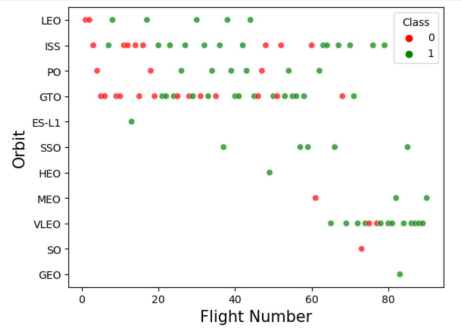
- GitHub URL – [Click Here]

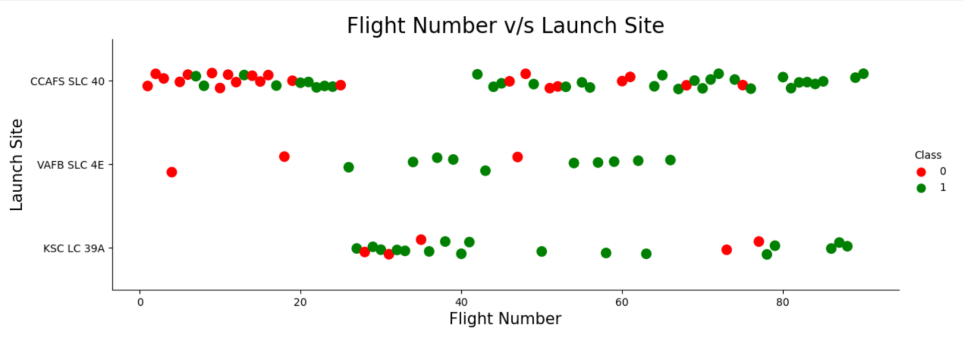


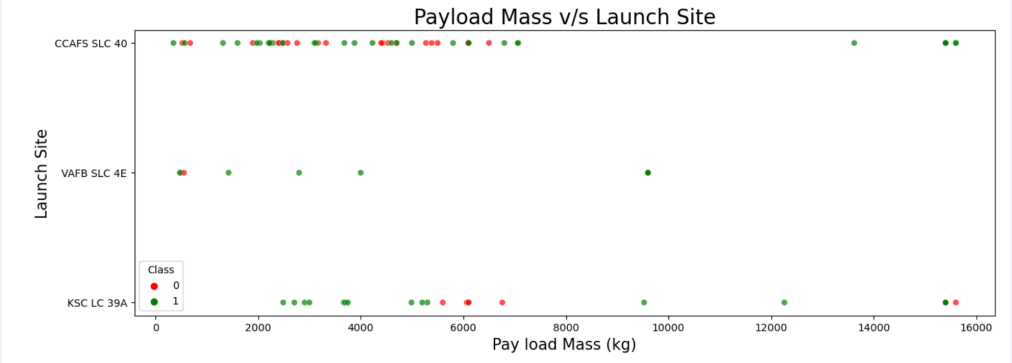
EDA with Data Visualization

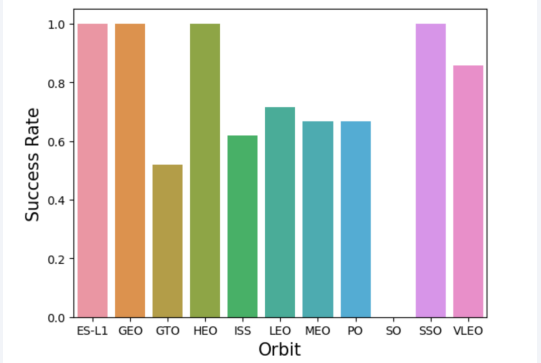
- GitHub URL – [Click Here]

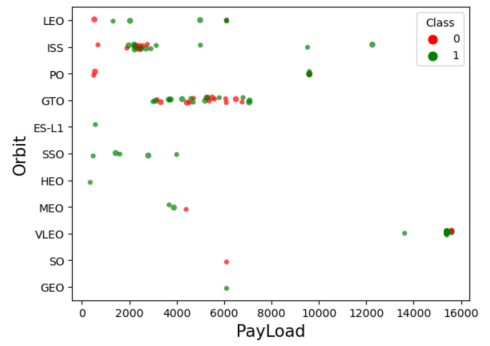








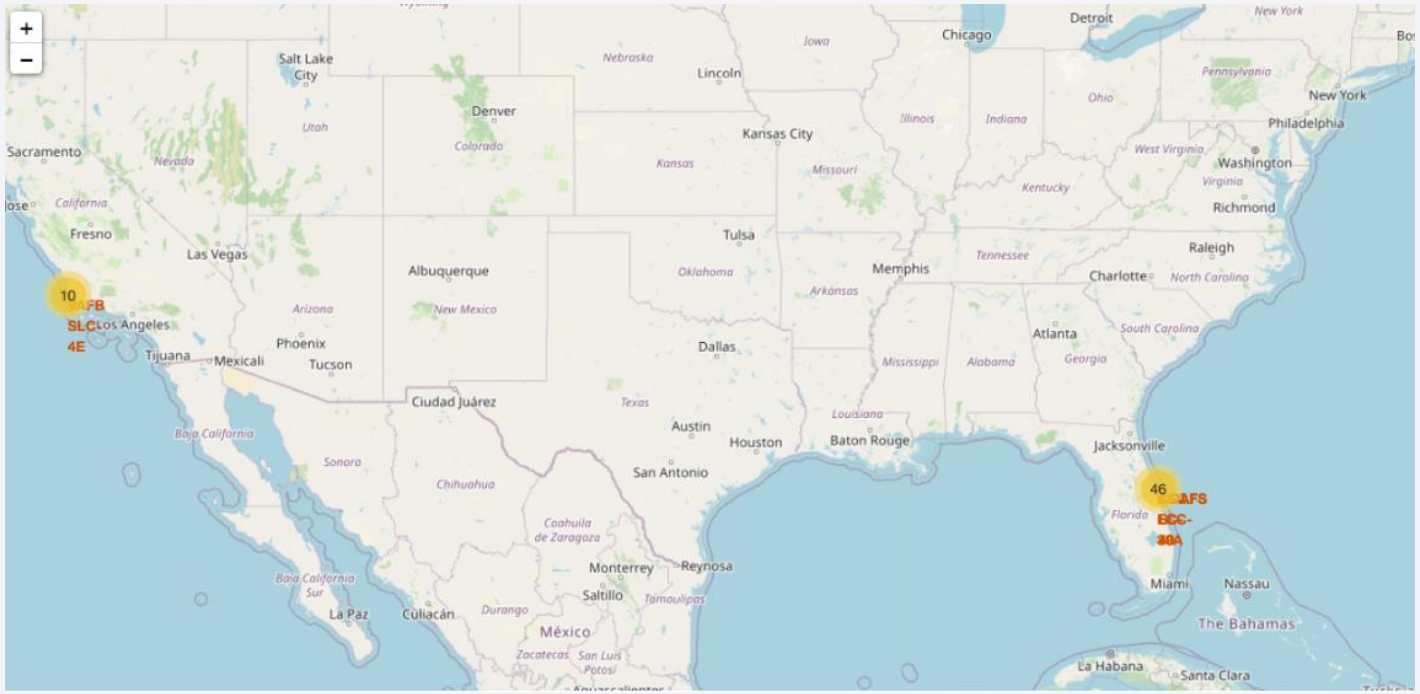




EDA with SQL

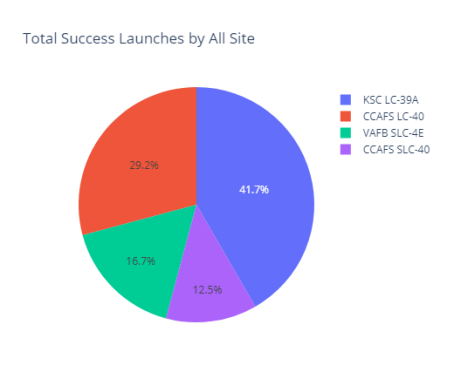
- SQL queries performed:
  - Displaying the names of the unique launch sites in the space mission
  - Displaying 5 records where launch sites begin with the string 'CCA'
  - Displaying the total payload mass carried by boosters launched by NASA (CRS)
  - Displaying average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - 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.
  - Ranking the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.
- GitHub URL - Click Here

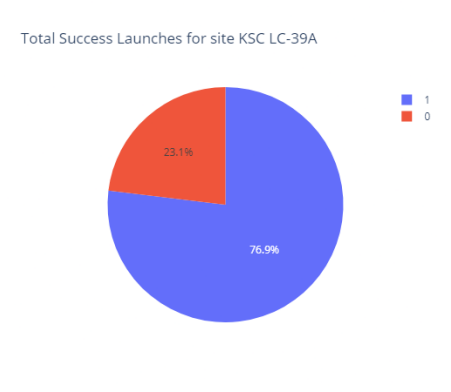
Build an Interactive Map with Folium

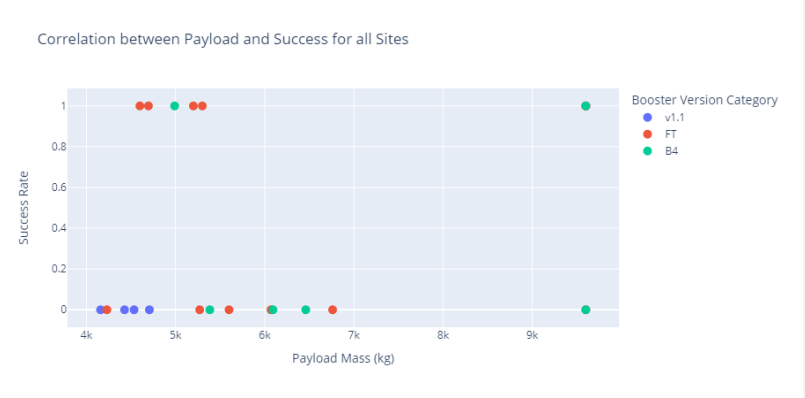


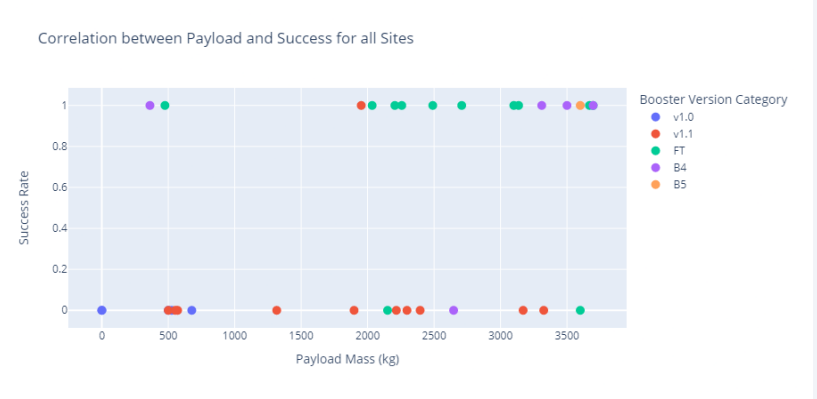
- GitHub URL – Click Here
- Folium HTML Map – Click Here

Build a Dashboard with Plotly Dash





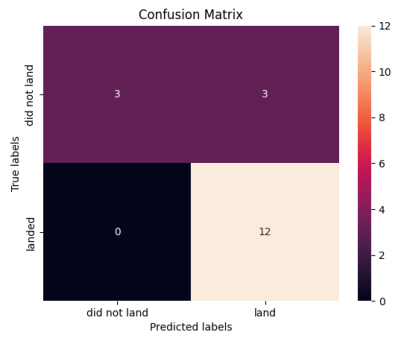


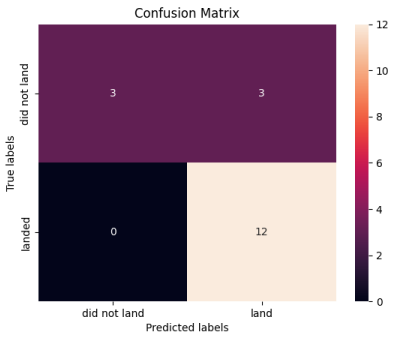


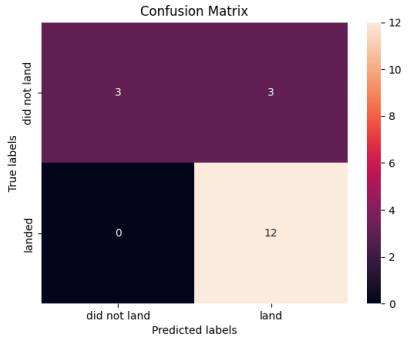
- GitHub URL – [Click Here]
- Charts showing success rates of Launch Sites and Payloads.

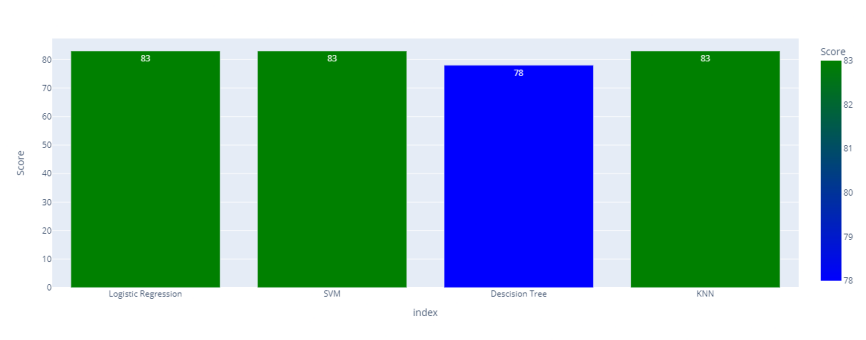
Predictive Analysis (Classification)

- SVM, KNN and Logistic Regression models gave highest and same accuracy of 83% while Decision Tree gave 78% Accuracy.
- Top 3 models have similar confusion matrix too.









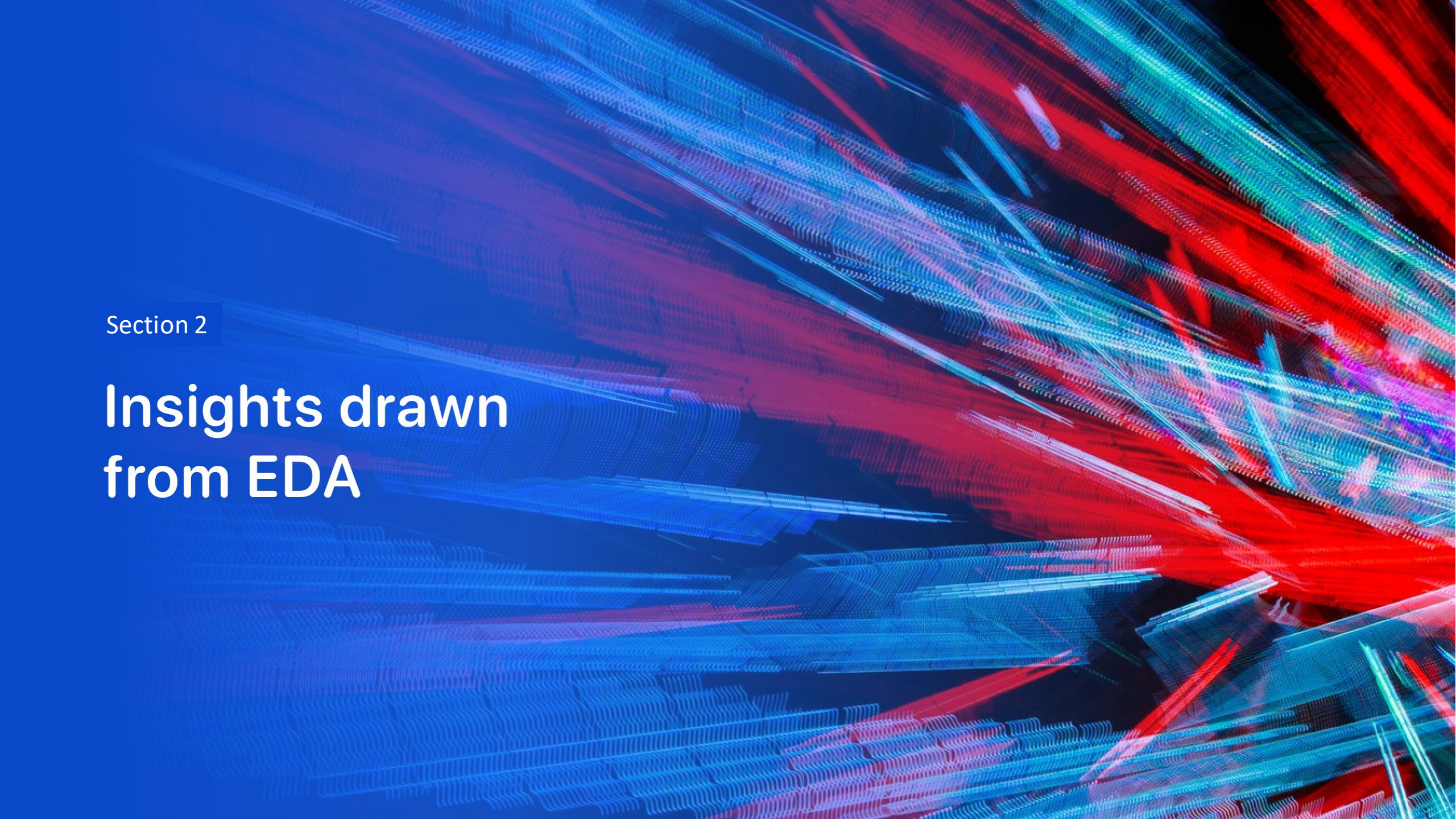
- GitHub URL – [Click Here]

Results

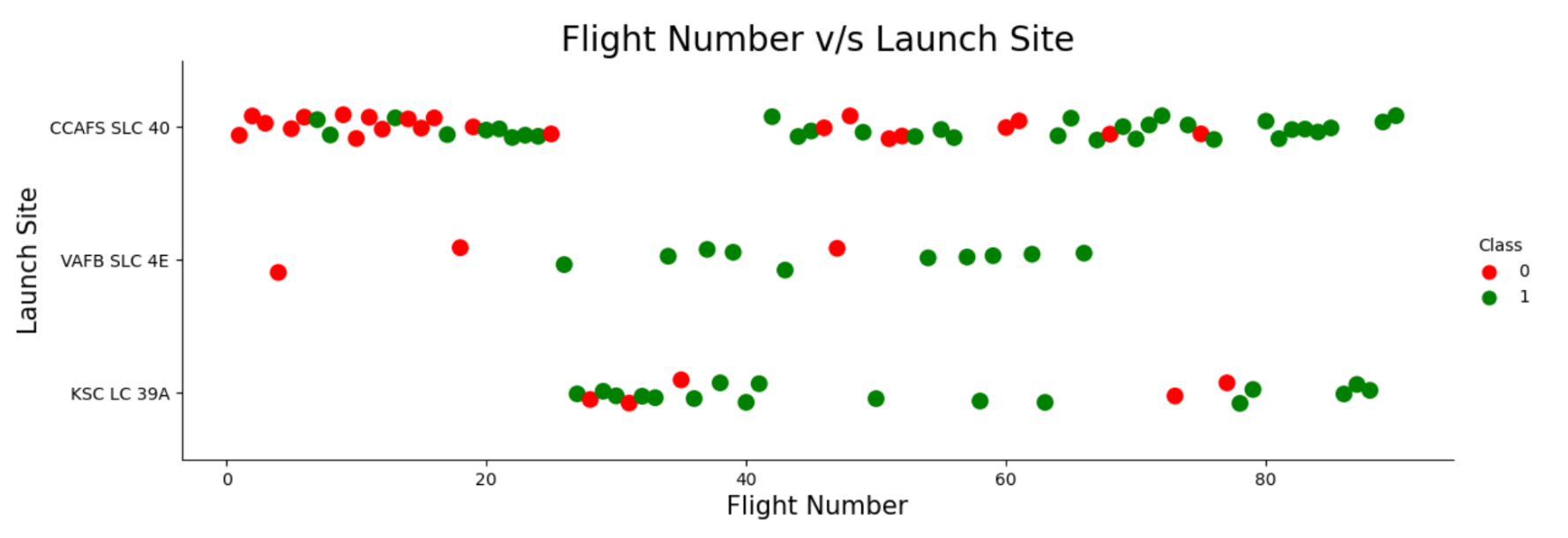
- KSC LC-39A launch site has highest success rate of 77% as compared to other 3 launch sites.
- The success of SpaceX launches is directly proportional to the time in years. They eventually got better and improved their success rate.
- Launch sites are at a safer distance from cities, well connected to roads, railways and right beside coastal.
- Low weighted payloads have higher success rate than heavier ones.
- SVM, KNN and Logistic Regression models performed best in terms of accuracy so proceeding with any of them will work well.

Section 2

Insights drawn from EDA

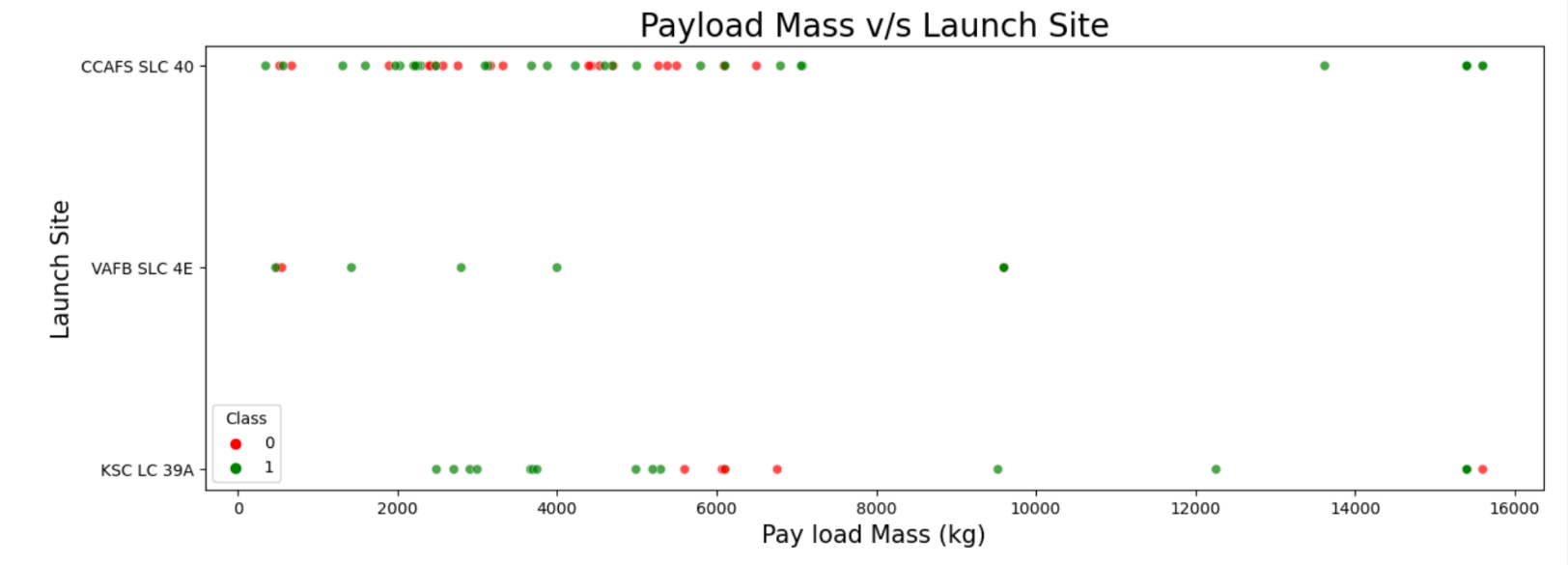


Flight Number vs. Launch Site



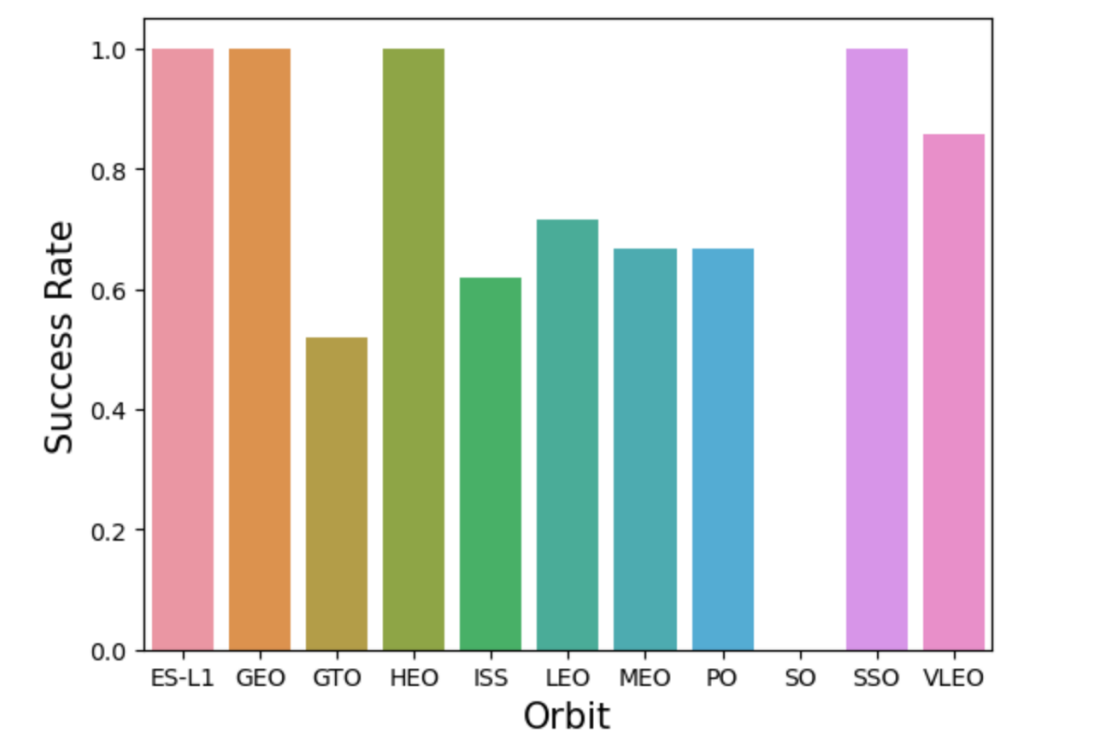
- Initial launches of rockets has far less success rate than later ones.
- Sites VAFB SLC-4E and KSC LC 39A , have success rate of more than 77%. Considering them as our launch site would be a good preference.

Payload vs. Launch Site



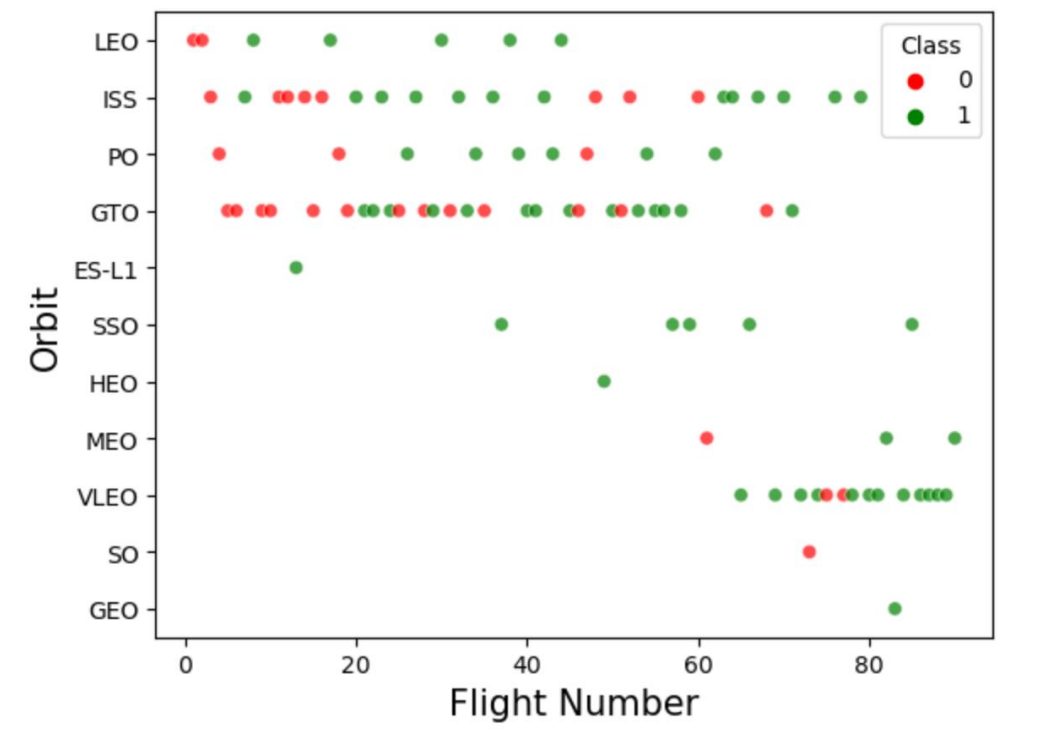
- At VAFB SLC-4E site, no rockets were launched for payload more than 10,000 Kg.
- Even though CCAFS SLC-40 site has less success rate, but it faced no failure after 6500 Kg payload.
- Site KSC LC-39A is highly successful for weights under 5500 Kg.

Success Rate vs. Orbit Type



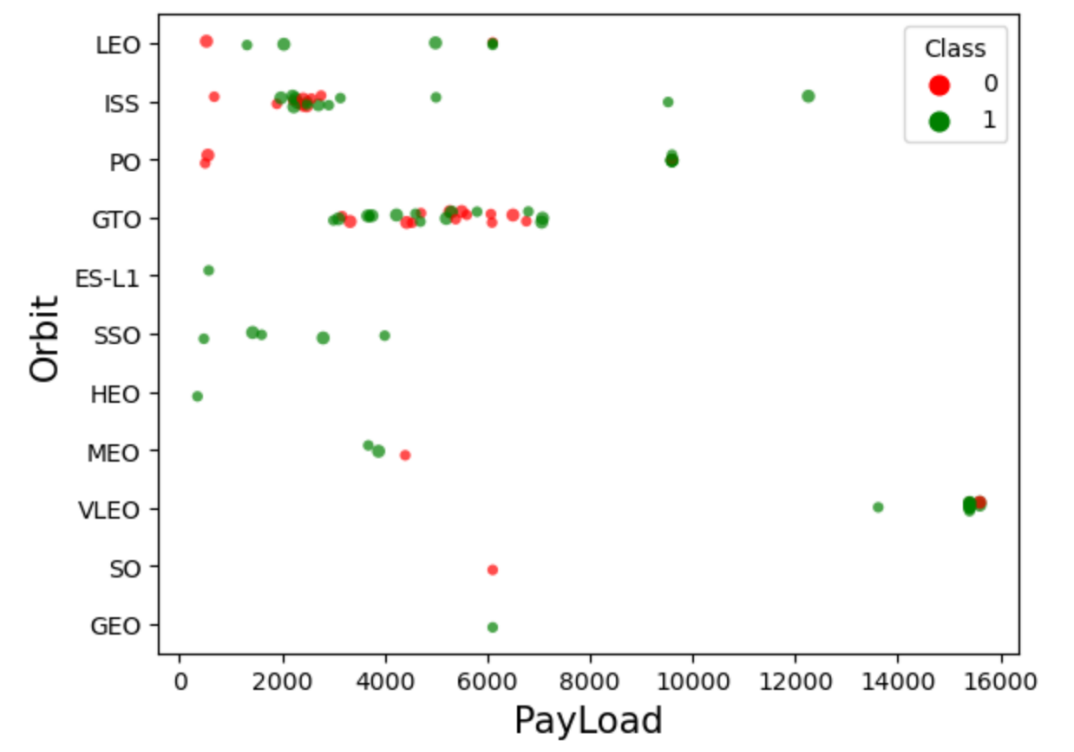
- ES-L1, GEO, HEO are orbits with high altitude (35,000+ Km) and all of them have success rate of 100%.
- Strangely, GTO is also a high-altitude orbit but with least success rate out of all!
- SSO orbit is a low level (600 Km) orbit with 100% success rate.

Flight Number vs. Orbit Type



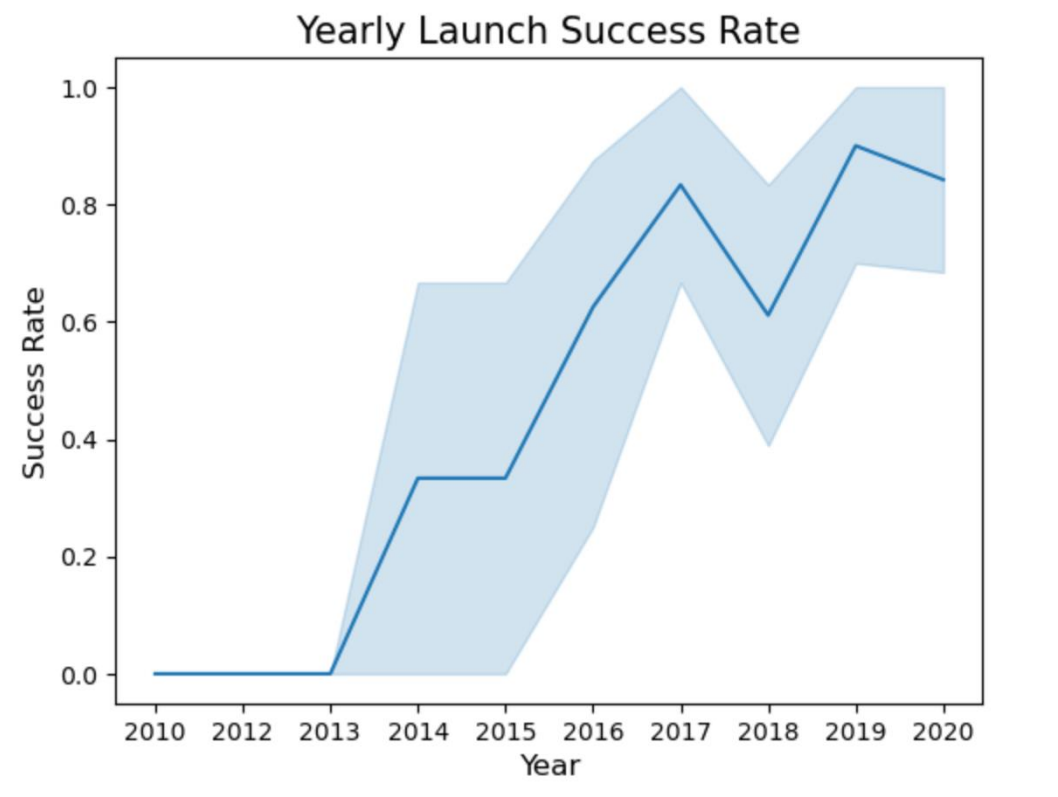
- LEO orbit faced 2 initial failures then success rate stay high after that.
- SSO orbit never faced any failure. 5/5 successful launches.
- GTO and ISS orbits are important ones and that's why we see many launches for them.
- VLEO has become SpaceX favorite lately because of Starlink Project.

Payload vs. Orbit Type



- SpaceX has sent heaviest payloads on VLEO orbit which has low altitude.
- Light payload rockets has been sent on farthest orbits.
- SpaceX has been more successful with payloads over 7000kg than lighter payloads.

Launch Success Yearly Trend



- For 3 years 2010-12, there was not even a single successful launch.
- 2013 was the breakthrough year for SpaceX.
- SpaceX had seen growing success rate from 2013 to 2020.

All Launch Site Names

- Query :- *SELECT DISTINCT Launch_Site FROM SPACEXTBL*
- Result :-

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

Launch Site Names Begin with 'CCA'

- Query :- *SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;*
- Result :-

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass Launched by NASA (CRS)

- Query :- *SELECT SUM(PAYLOAD_MASS__KG_)*
  FROM SPACEXTBL WHERE Customer = 'NASA (CRS)'
- Result :-

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- Query :- *SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL*
  WHERE Booster_Version LIKE 'F9 v1.1';
- Result :-

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Query :- *SELECT Date FROM SPACEXTBL*
  WHERE `Landing_Outcome` = "Success (ground pad)" LIMIT 1;
- Result :-

Date
22-12-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query :- *SELECT Booster_Version FROM SPACEXTBL*
  WHERE `Landing _Outcome` = "Success (drone ship)"
  AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
- Result :-

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful Mission Outcomes

- Query :- *SELECT COUNT(Mission_Outcome) AS Success_Outcome*
  FROM SPACEXTBL WHERE Mission_Outcome LIKE 'Succ%';
- Result :-

Success_Outcome
100

Boosters Carried Maximum Payload

- Query :- *SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTBL*
  WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_)
  FROM SPACEXTBL)
- Result :-

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- Query :- *SELECT * FROM SPACEXTBL*
  WHERE Landing_Outcome LIKE 'Success%'
  AND Date BETWEEN ('2015-01-01' and '2015-12-31') ORDER BY Date DESC;
- Result :-

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)
		CCAFS LC-				SKY Perfect JSAT		

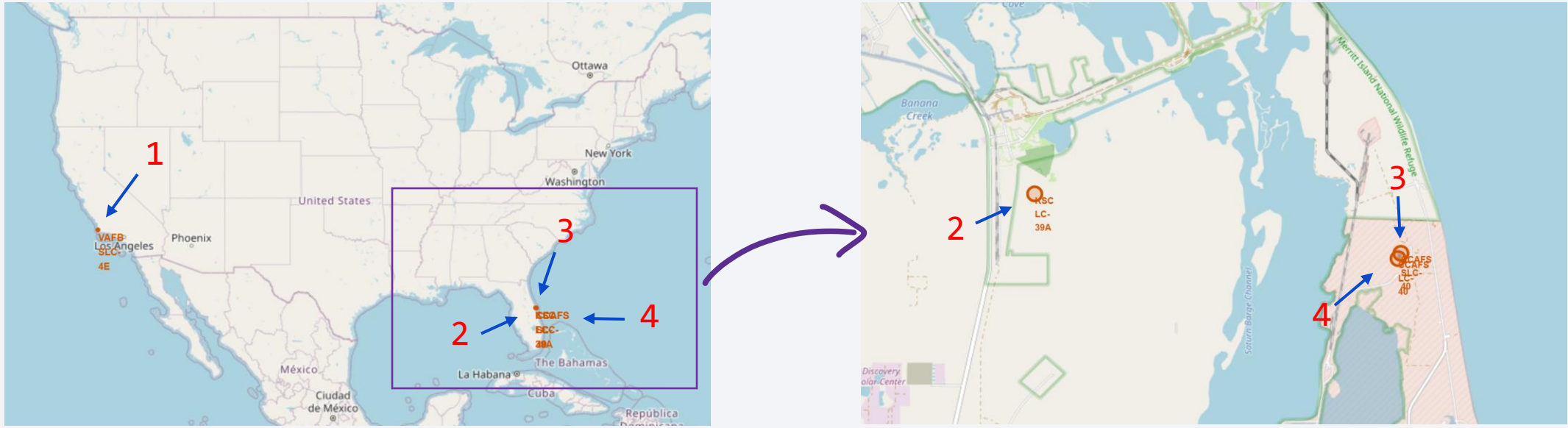
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Query :- *SELECT * FROM SPACEXTBL*
  WHERE `Landing_Outcome` LIKE 'Success%'
  AND Date BETWEEN '2010-06-04' AND '2017-03-20' ORDER BY Date DESC;
- Result :-

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)

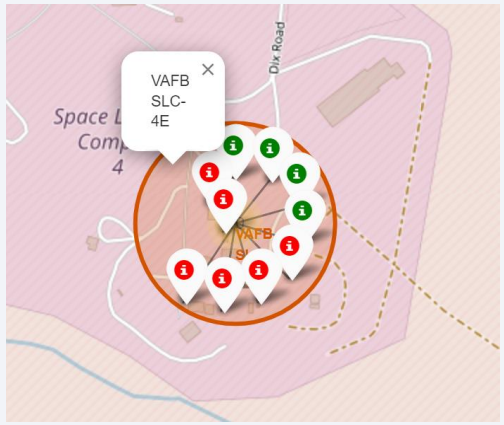

Section 3
Launch Sites
Proximities Analysis

Location of Launch Sites

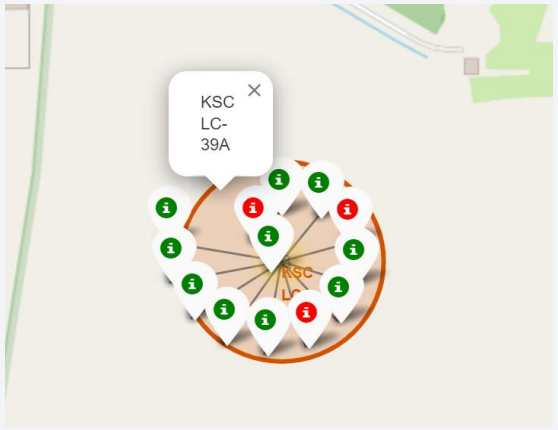


- Safest place to launch a rocket is none other but a coastline.
- In above pics we can see 1 Site (VAFB SLC-4E) is located at western coast of USA while other 3 are located on the eastern coast of USA.

Success/Failed Launches of Sites

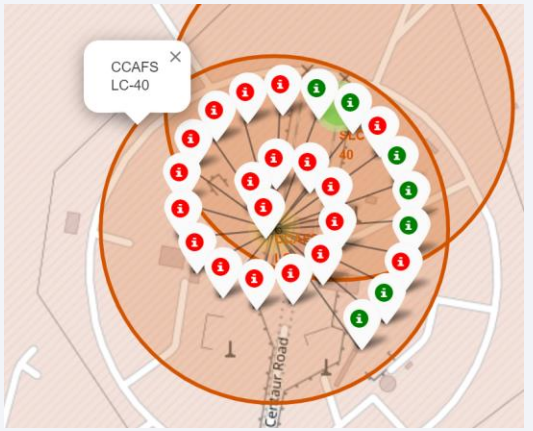


VAFB SLC-4E Site



KSC LC-39A Site

CCAFS SLC-40 Site

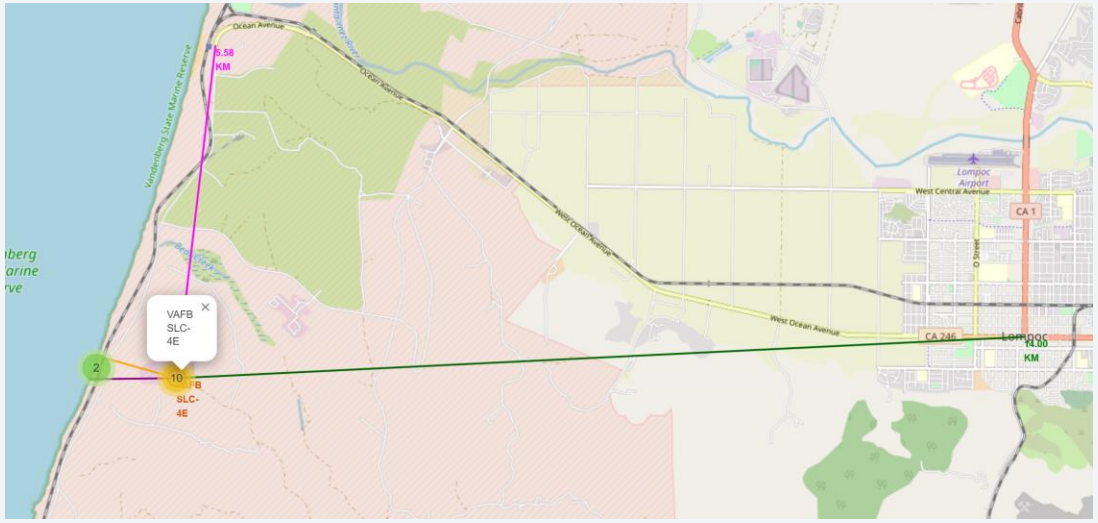


CCAFS LC-40 Site

- VAFB SLC-4E : 10 Rockets launched with 40% Success Rate.
- KSC LC-39A : 13 Rockets launched with 77% Success Rate.
- CCAFS SLC-40 : 7 Rockets launched with 43% Success Rate.
- CCAFS LC-40 : 26 Rockets launched with 27% Success Rate.

Proximities of Launch Site : VAFB SLC-4E

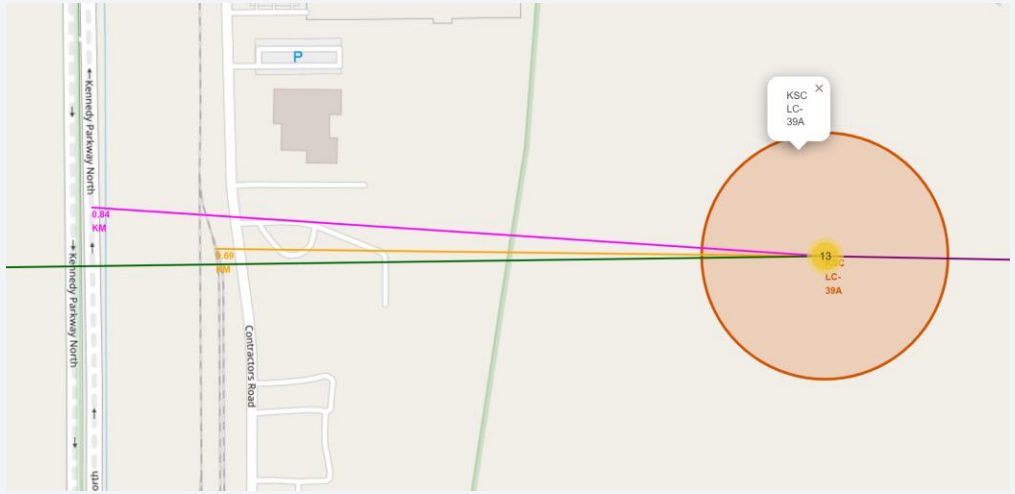
Distance from **Coastline** & **Railway**



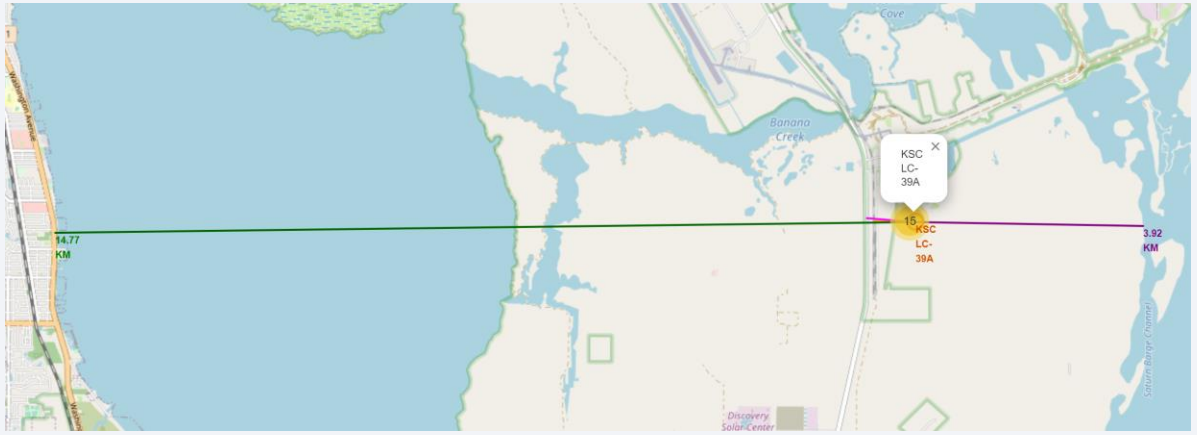
Distance from **Highway** & **City**

- 1.44 Km away from Coastline
- 1.26 Km away from Railway
- 5.58 Km away from Major Highway (Ocean Avenue)
- 14 Km away from Nearest City (Lompoc)

Proximities of Launch Site : KSC LC-39A



Distance from **Highway** & **Railway**

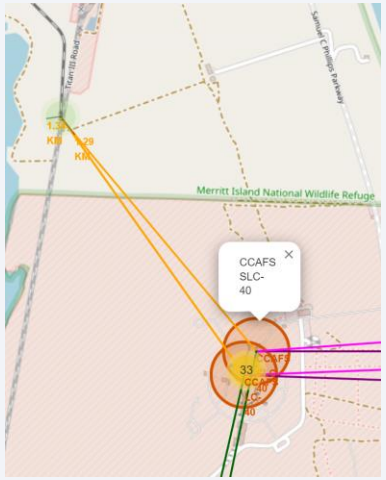


Distance from **Coastline** & **City**

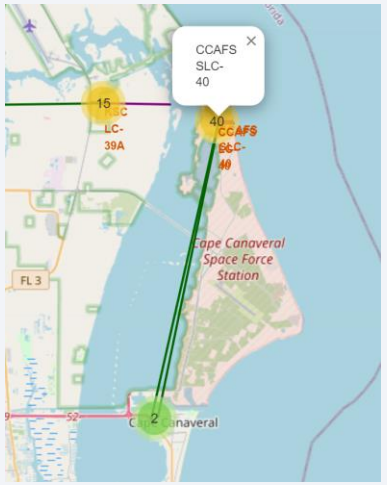
- 3.92 Km away from Coastline
- 0.69 Km away from Railway
- 0.84 Km away from Major Highway (Kennedy Parkway North)
- 14.77 Km away from Nearest City (Titusville)

Proximities of Launch Site : CCAFS SLC-40

Distance from **Highway** & **Coastline**



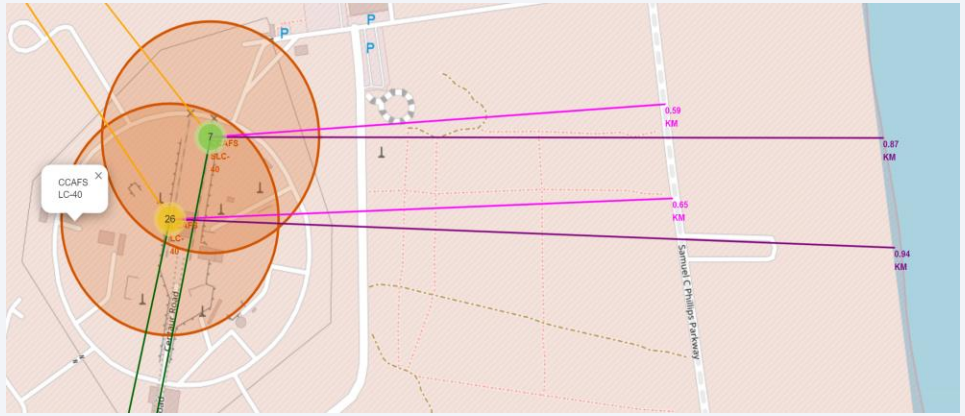
Distance from **Railway**



Distance from **City**

- 0.87 Km away from Coastline
- 1.29 Km away from Railway
- 0.59 Km away from Major Highway (Samuel C Phillips Parkway)
- 18 Km away from Nearest City (Canaveral)

Proximities of Launch Site : CCAFS LC-40

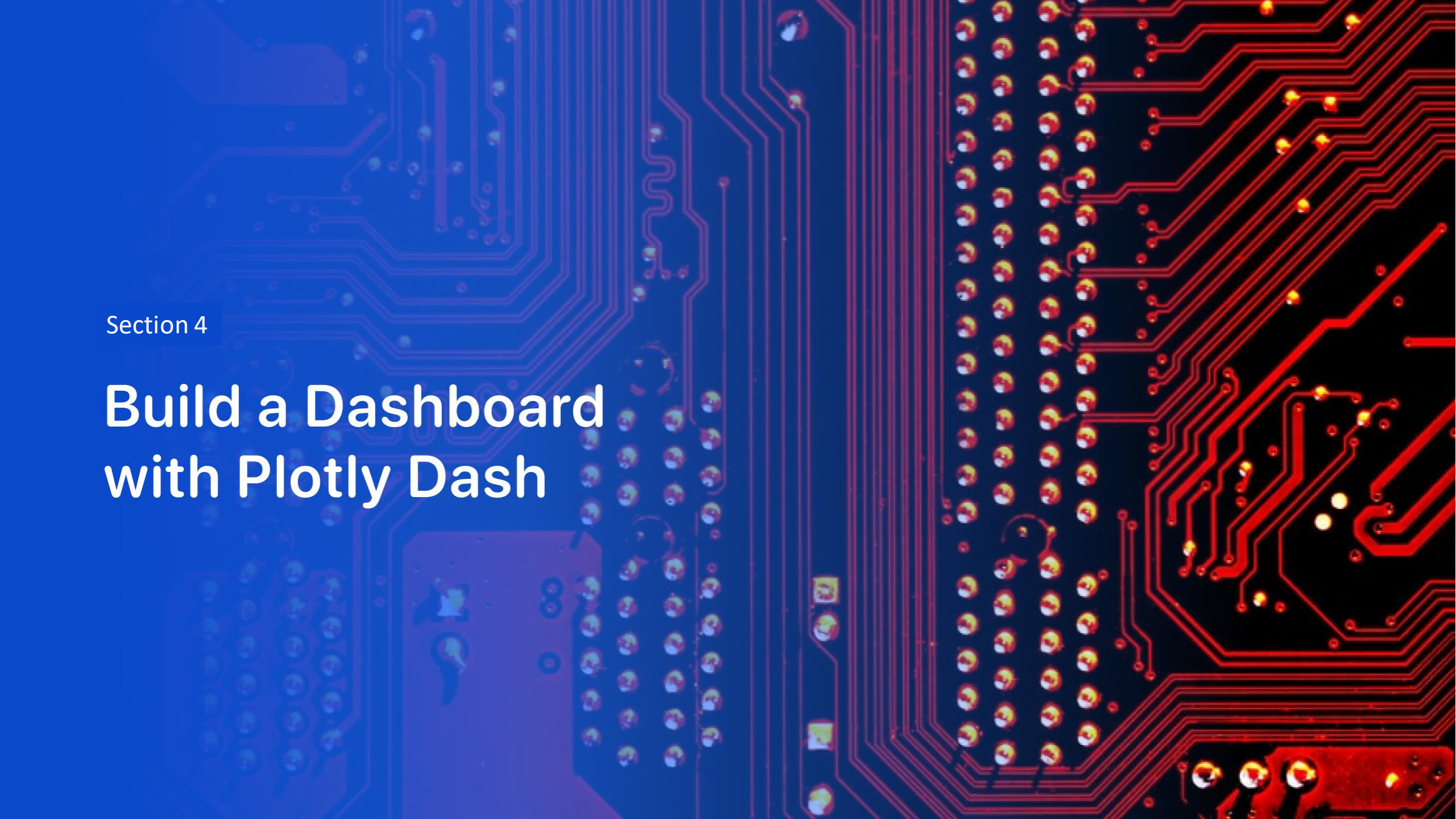


Distance from **Highway** & **Coastline**

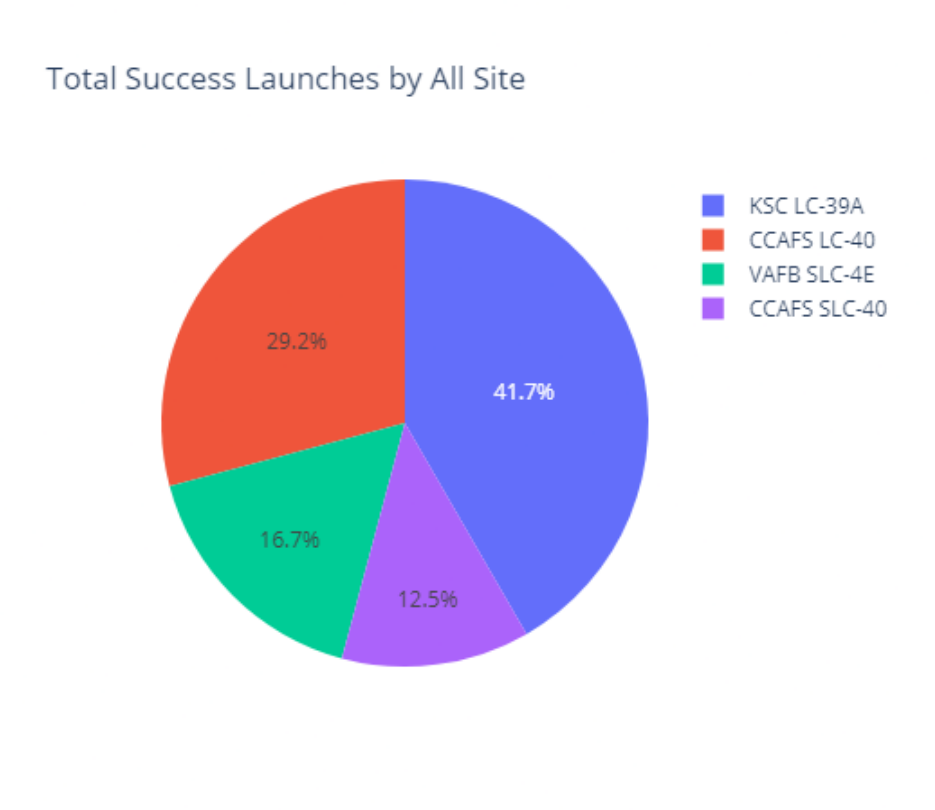
Distance from **Railway**

Distance from **City**

- 0.94 Km away from Coastline
- 1.34 Km away from Railway
- 0.65 Km away from Major Highway (Samuel C Phillips Parkway)
- 18 Km away from Nearest City (Canaveral)

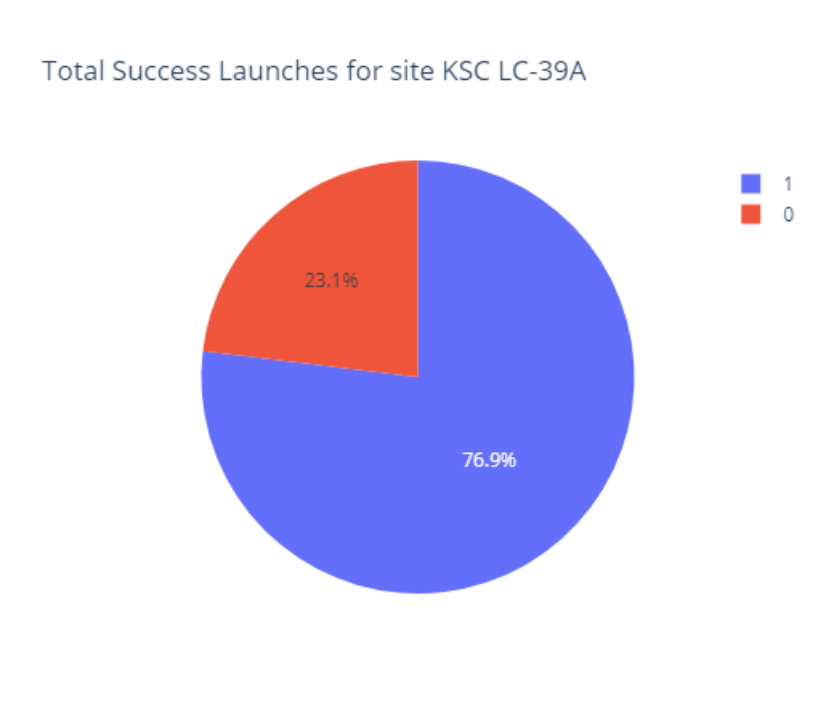

Section 4
Build a Dashboard with Plotly Dash

Success Launch Counts For All Sites



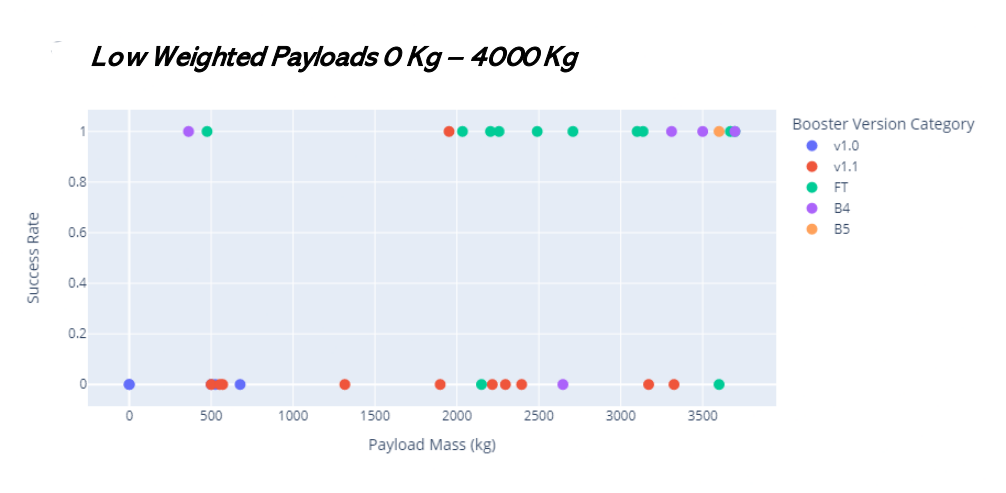
- Site KSC LC-39A has highest success launch rate as compared to other 3 sites.
- CCAFS SLC-40 has lowest success rate. It could be because initially this site was used when SpaceX started.

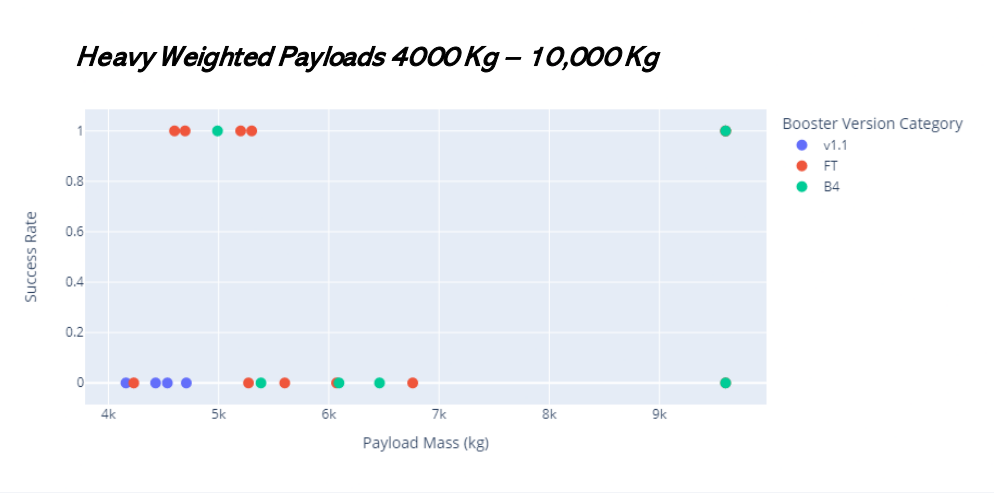
Launch Site with Highest Launch Success Ratio



- Out of all sites, KSC LC-39A came out with highest Launch Success Ratio.
- 77% Launches had been successful for SpaceX at this site.
- 23% Launches failed here.

<Dashboard Screenshot 3>



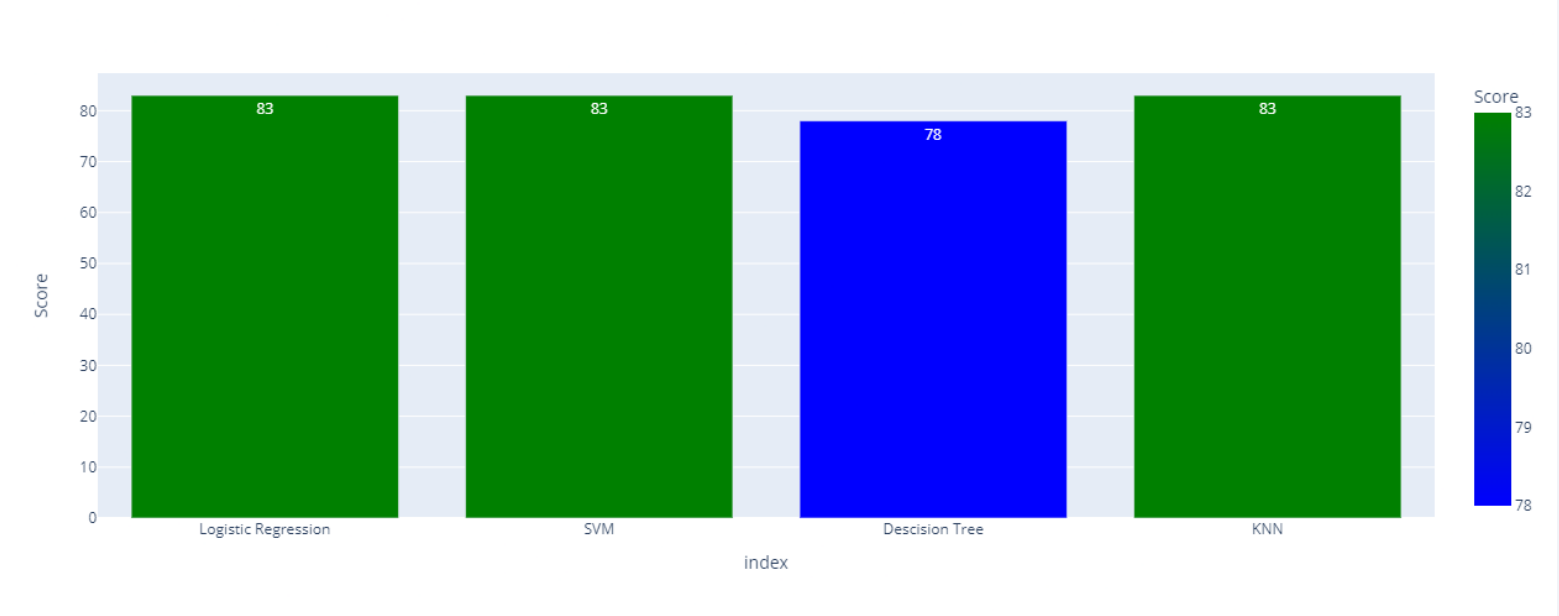


- Low Weighted Payloads have higher success chances than heavy weighted payloads.

Section 5

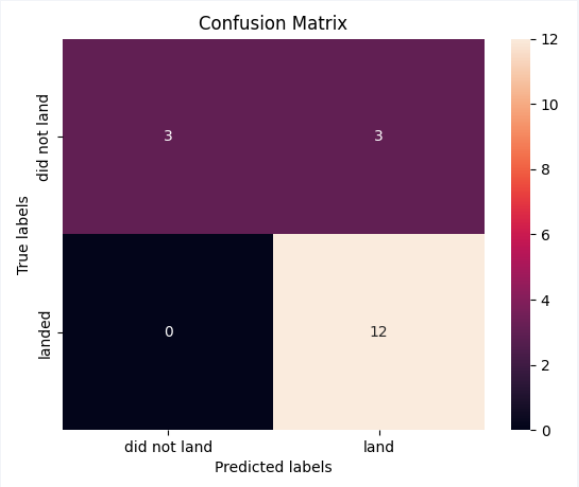
Predictive Analysis (Classification)

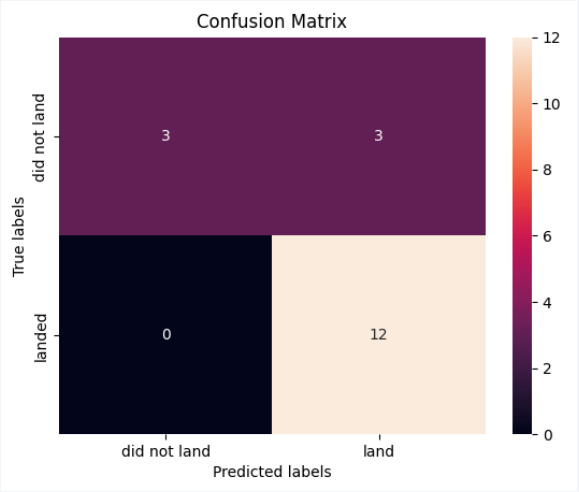
Classification Accuracy

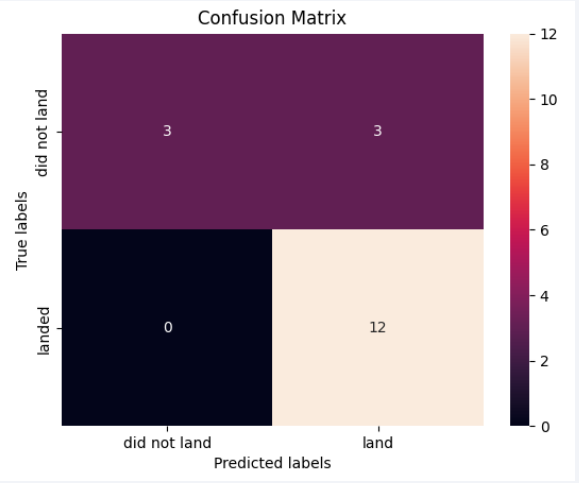


- Logistic Regression, SVM and KNN algorithms topped the chart with 83% Accuracy each.
- Decision Tree couldn't cope up with other 3 and gave 78% accuracy.

Confusion Matrix







- Confusion Matrix of winner models – SVM, LR & KNN – are same.
- We can see an error rate in top right quadrant where it is predicting "Rocket will Land" but it didn't land.

Conclusions

- KSC LC-39A site has highest rate of launching rockets successfully. Picking this site would be preferable.
- For high altitude we can choose these orbits - ES-L1, GEO, HEO. These three orbits have very high success rate.
- For low altitude we can choose this orbit – SSO.
- Boosters - F9 FT B1021.2, 1022, 1026, 1031.2 - are successful with for Drone Ship Landing with weights between 4000Kg and 6000Kg.
- Low weighted payloads perform better than heavier ones.
- SVM, KNN and Logistic Regression models were best in terms of prediction accuracy with 83% accuracy each.

Appendix

- GitHub repository with all files -

https://github.com/HimGos/IBM-Data-Science-Project

- Interactive Map – Map Link

Thank you!