IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

TVLAN

12 April 2024

Outline

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

Executive Summary

This analysis aimed to emulate SpaceX's success by predicting the Falcon 9 rocket's Stage 1 landing outcomes. Leveraging data from the SpaceX API and supplementary sources, the study explored key variables affecting landing success through visualizations and a dashboard. A predictive model was developed, with Support Vector Machines (SVM) emerging as the most accurate at 83%, thanks to its robustness against outliers. The findings provide actionable insights into optimizing launch conditions and improving future landing predictions.

Introduction

- SpaceX offers rocket launches for $62 million at approximately **38%** of its competitors' price , thanks to its innovative first stage reuse capability for its Falcon 9 rockets.

- Predicting the successful landing of the first stage of a rocket can significantly impact the overall cost of a launch, which is valuable information for our company to bid against SpaceX.

Analysis Objective :

- Create a predictive model to decide if SpaceX launches would be successful.

Section 1

Methodology

Methodology

Executive Summary

1. Data collection methodology:

    1.1 Extract data through Web scraping and API .

2. Perform data wrangling

    2.2 Filter and replace inappropriate data.

3 Perform exploratory data analysis (EDA) using visualization and SQL

4 Perform interactive visual analytics using Folium and Plotly Dash

5 Perform predictive analysis using classification models

    5.1 Perform optimization on the hyperparameters.

    5.2 Check Accuracy of models.

Data Collection – Data Sources

SpaceX API

(https://api.spacexdata.com/v4)

Falcon 9 Historical Launch Data Wikipedia Page

(https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

[hide] Flight No.	Date and time (UTC)	Version, booster[c]	Launch site	Payload[d]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
195	3 January 2023 14:56[20]	F9 B5 △ B1060.15[21]	CCSFS, SLC-40	*Transporter-6*: (115 payloads Smallsat Rideshare)	Unknown[e]	SSO	Various	Success	Success (ground pad)
	Dedicated SmallSat Rideshare mission to Sun-synchronous orbit. It included six space tugs, also known as orbital transfer vehicles (OTV), which are two of D-Orbit's ION Satellite Carriers, Epic Aerospace's Chimera LEO 1, Momentus's Vigoride-5, Skykraft's OTV and Launcher's Orbiter SN1.[22][23] Orbiter SN1 failed shortly after deployment from Falcon and before deploying payloads. One of the payloads, EWS RROCI failed to deploy from Falcon 9 and the satellite re-entered with the upper stage.[24] This was not a SpaceX failure as brokered dispensers and deployers are used on Transporter missions.[25]								
196	10 January 2023 04:50[26]	F9 B5 △ B1076.2	CCSFS, SLC-40	OneWeb Flight #16 / SpaceX Flight 2[27][28] (40 satellites)[29]	6,000 kg (13,000 lb)	Polar LEO	OneWeb	Success	Success (ground pad)
	Following the Russian invasion of Ukraine, OneWeb suspended launches on Soyuz rockets.[30] In March 2022, OneWeb announced that they had signed an agreement with SpaceX to resume satellite launches.[31]								
FH 5	15 January 2023 22:56[32]	Falcon Heavy B5 B1070 (core)	KSC, LC-39A	USSF-67 (CBAS-2 & LDPE-3A)[33]	~3,750 kg (8,270 lb)	GEO	USSF	Success	No attempt
		B1064.2 (side) △							Success (ground pad)
		B1065.2 (side) △							Success (ground pad)

Figure 1 : Launch Data From Wikipedia

Data Collection - Scraping

- Obtain the general data on the parameters of Falcon 9 launches.

- **Parameters :** Date & Time , Booster Version , Launch site , Payload , Payload Mass, Orbit , Customer, Launch Outcome ,Booster Landing Status.

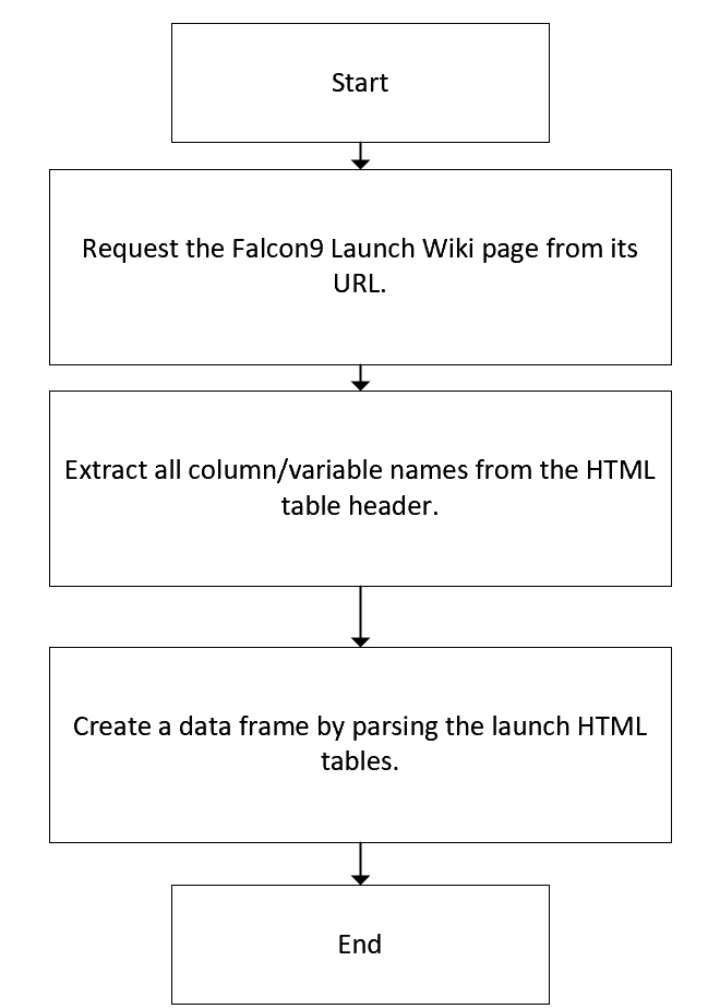


Figure 2 : Flowchart for SPACEX data web scraping.

Data Collection – SpaceX API

- Obtain detailed data on the launch parameters/variables of the rockets , payload , launchpad and cores of the previous Falcon 9 launches.

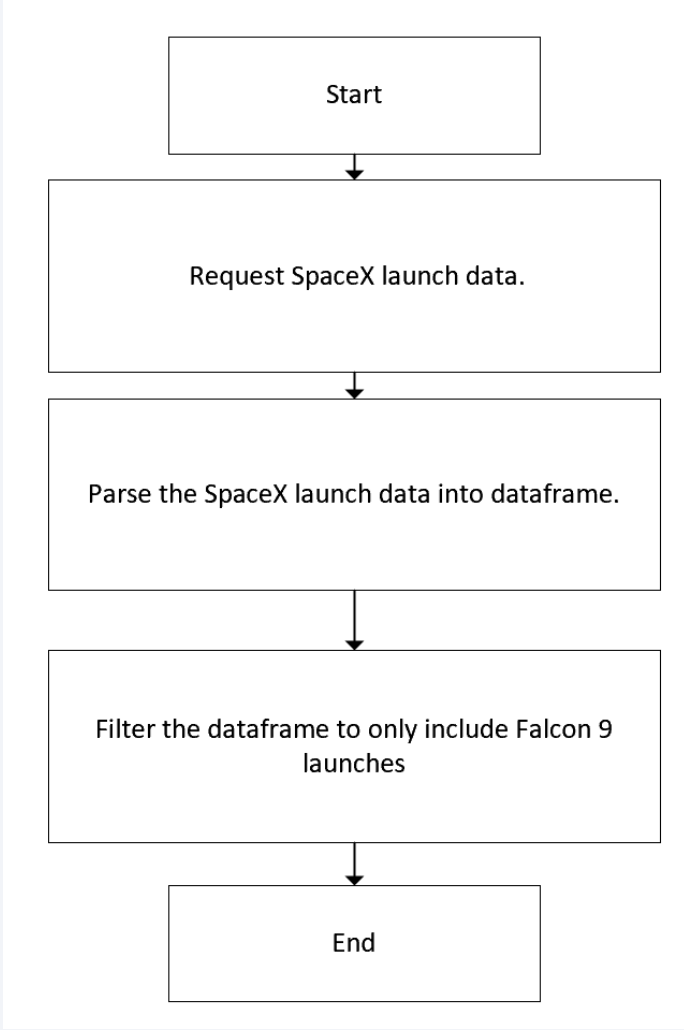


Figure 3 : Flowchart for SPACEX API data collection.

Note:

Detailed list of the variables are listed in the appendix

Data Wrangling

- Replace the null data or other inappropriate data with the mean , max or min of the column.

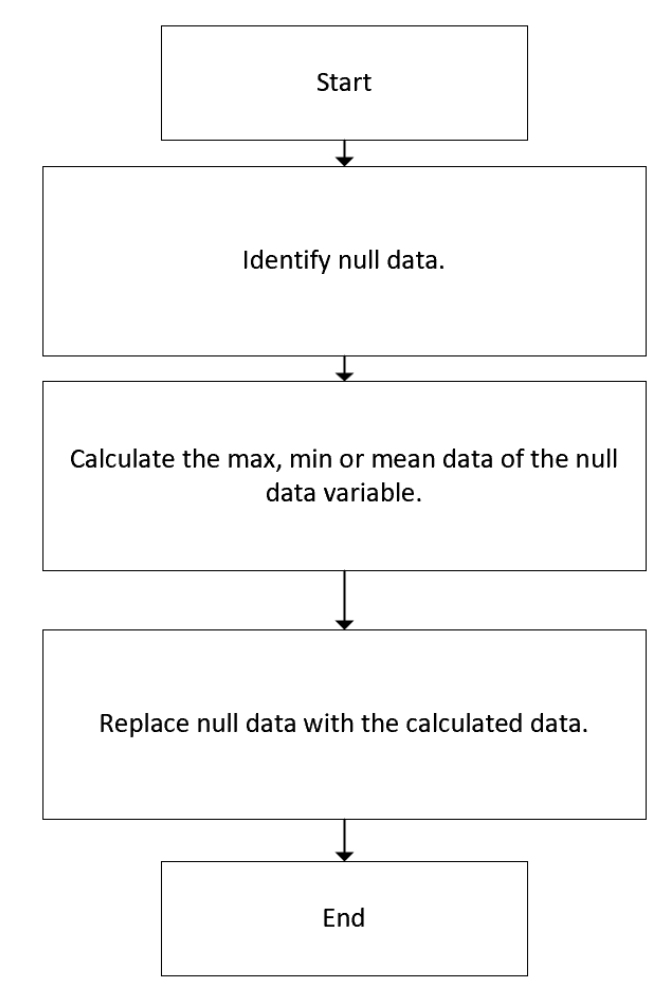


Figure 4 : Flowchart for SPACEX API data wrangling.

EDA with Data Visualization

Categorical Data

1. **Scatter Plot**

- To observe the relationship between payload mass and flight number on the mission outcome.
- To observe the relationship between launch sites and flight number along with the success rate of the rockets.
- To observe the relationship between payload mass and launch orbit on the mission outcome.

2. **Box plot**

- To observe the success rate for each orbit the payload has been launched

Continues Data

1. **Line Plot**

- To study the trend of success rate of the mission with progressing years.

EDA with SQL

SQL Search Queries :

- Names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Payload mass carried by booster version F9 v1.1
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Number of successful and failure mission outcomes
- Names of the booster versions which have carried the maximum payload mass
- Month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.

Build an Interactive Map with Folium

- Folium Map is created to display the various launch sites on a map.
- Map Features :

  Launch Site Marker – To identify each launch site and its vicinity

  Success/Fail Launches Marker- To visualize the number of launches on each site and its success/failure rate.

  Proximity Measurements – To identify the distance of seashore ,rail tracks and roads from the launch site.

Build a Dashboard with Plotly Dash

- Create a dashboard to visualize relationship between mission outcome and other independent variables.
- Dash Features :

  Pie Chart : Displays the mission success percentage for all sites and individual sites separately.

  Scatter Plot : Displays the relationship between payload mass and mission outcome.

Predictive Analysis (Classification)

- Classification Model Tested :

  Multiple Regression , Decision Tree , SVM & KNN.

- 20% of the data has been set as test data while remaining 80% is the training data.
- Optimal hyperparameters for each model is selected through GridSearchCV function.
- Cross validation is set to 10 for all models.

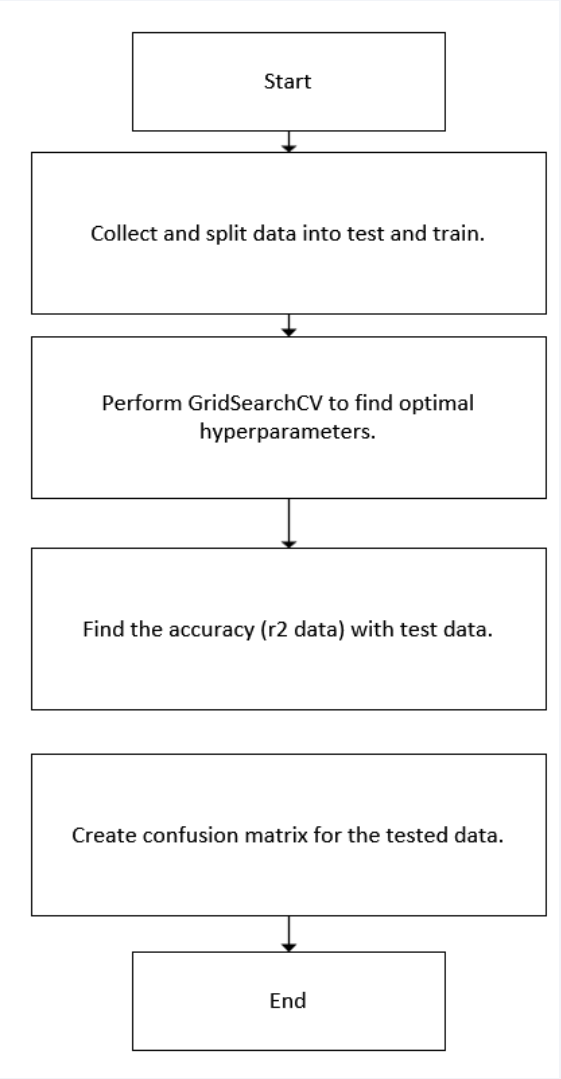


Figure 5 : Flowchart for predictive analysis of each model.

Results And Discussion

Section 2.1

Insights drawn from EDA

Flight Number vs. Launch Site

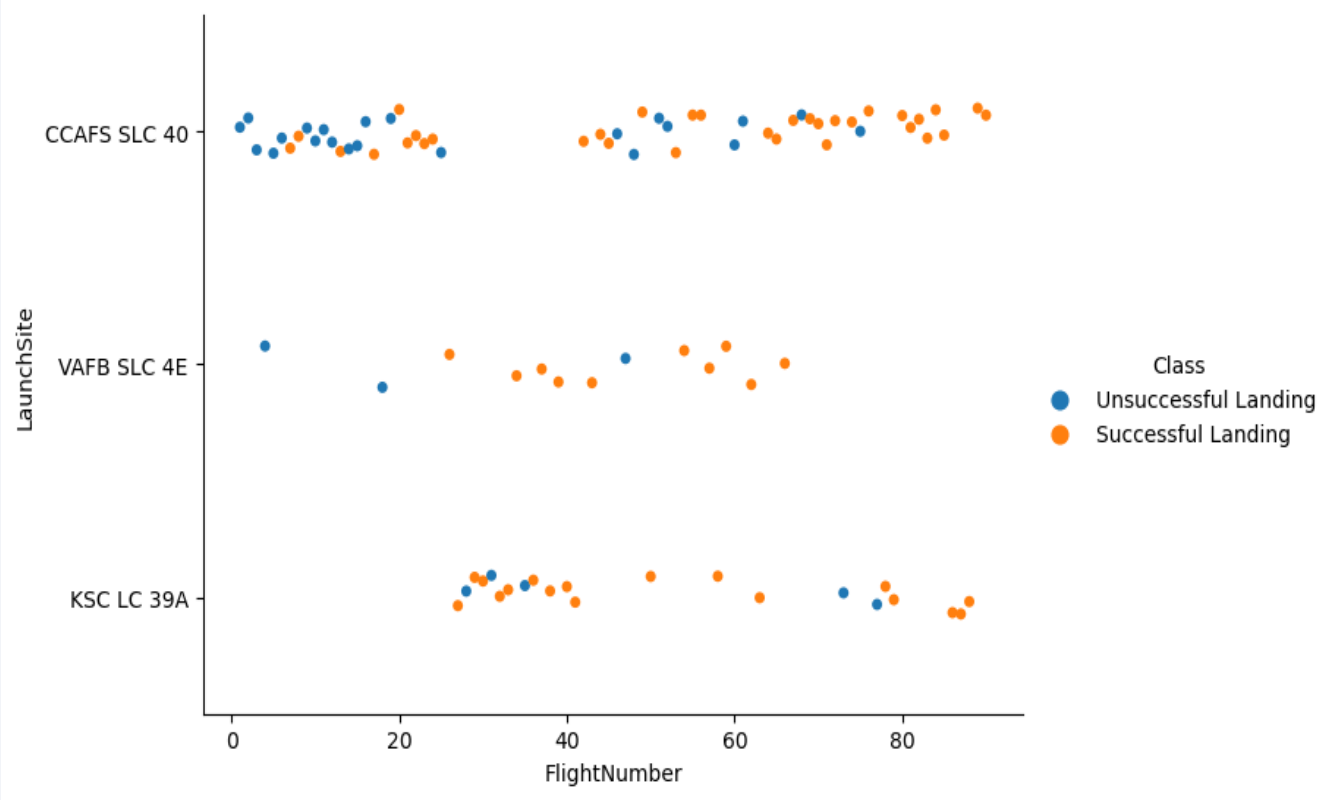


Figure 6 : Scatterplot of flight number against launch site.

- Most launches is observed on CCAFS while the least is on VAFB .
- No launches have been done on launch site VAFB past approximately flight number 70.
- All launches past flight number 78 is observed to be successful.

Payload vs. Launch Site

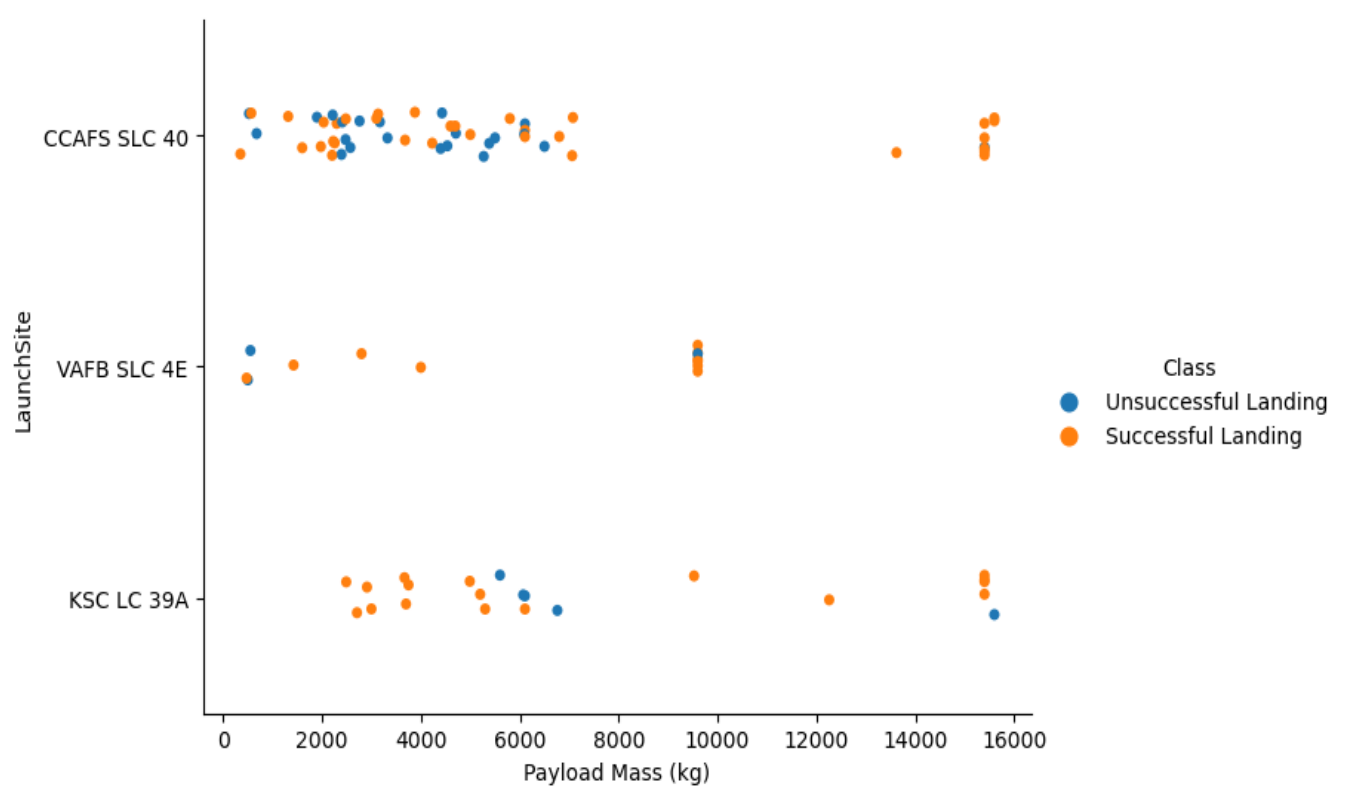


Figure 7 : Scatterplot of payload mass against launch site.

- Payload mass primarily ranges from 100 kg to 7600 kg for CCAFS, 2500 kg to 7000 kg for KSC and 900 kg for VAFB
- Payload mass for VAFB is sporadic.
- Both CCAFS and KSC has the highest payload mass of approximately 16000 kg.

Success Rate vs. Orbit Type

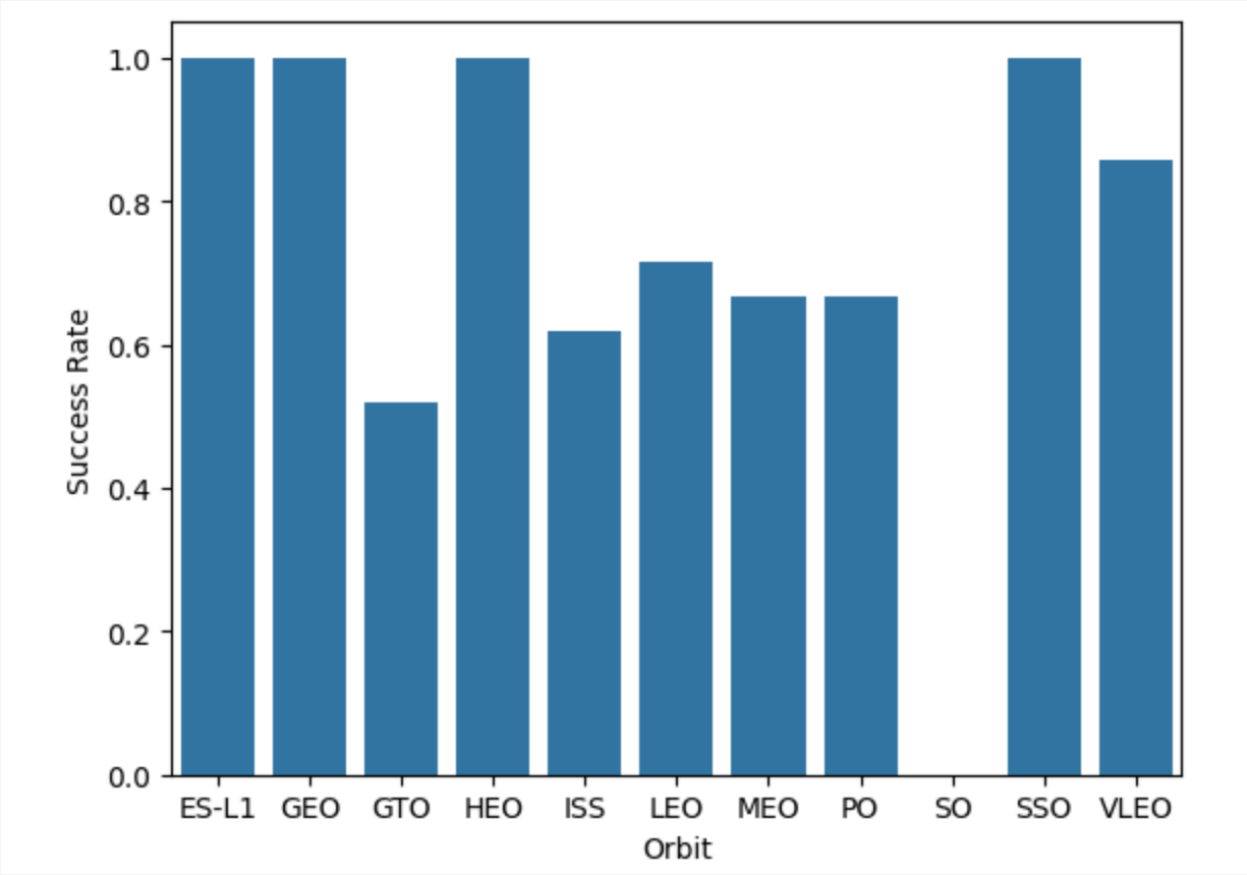


Figure 8 : Bar plot of success rate for each orbit.

- ES-L1 ,GEO, HEO & SSO has 100% success rate.
- SO has the lowest success rate at 0%.

Flight Number vs. Orbit Type

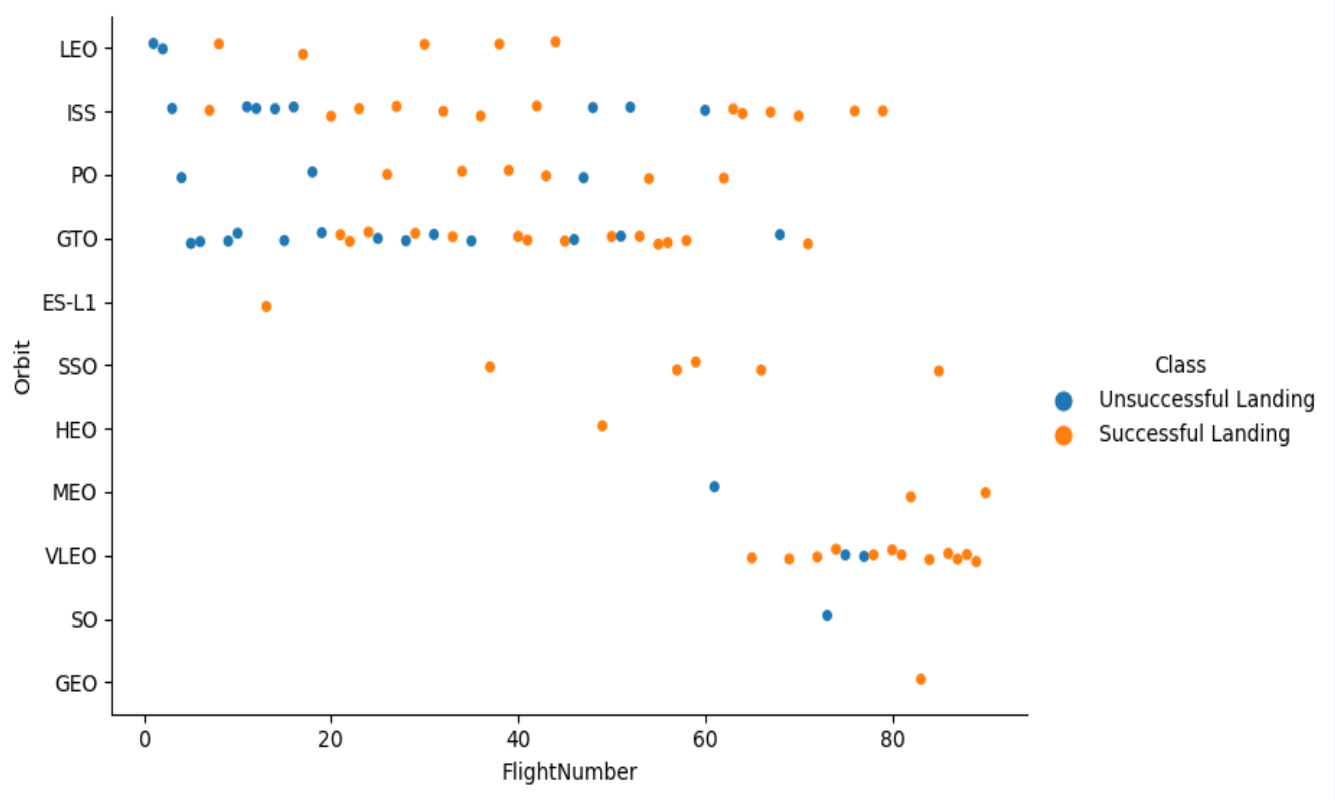


Figure 9 : Scatterplot of flight number against orbit.

- ISS , GTO and VLEO is observed to have the highest number of launches to.
- The least number of launches are ES-L1,HEO,SO and GEO with 1 launches each.
- Most launches past flight number 61 is to VLEO followed with ISS.

Payload vs. Orbit Type

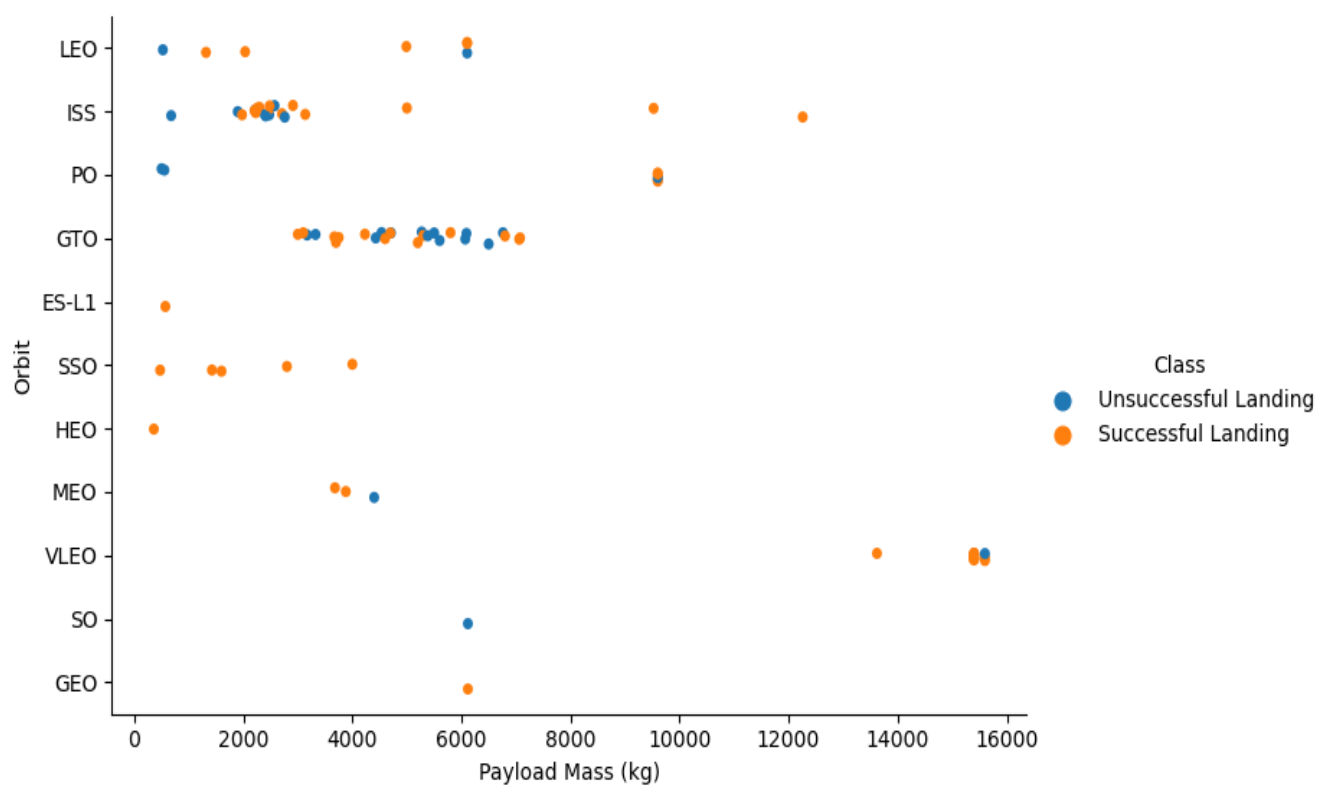


Figure 10 : Scatterplot of payload mass against orbit.

- Majority of launches to ISS has a payload mass ranging from 2000 kg to 3800 kg.
- Most launches to GTO ranges from 3000 kg to 7500 kg.
- Heaviest payloads at approximately15500 kg is launched to VLEO.

Launch Success Yearly Trend

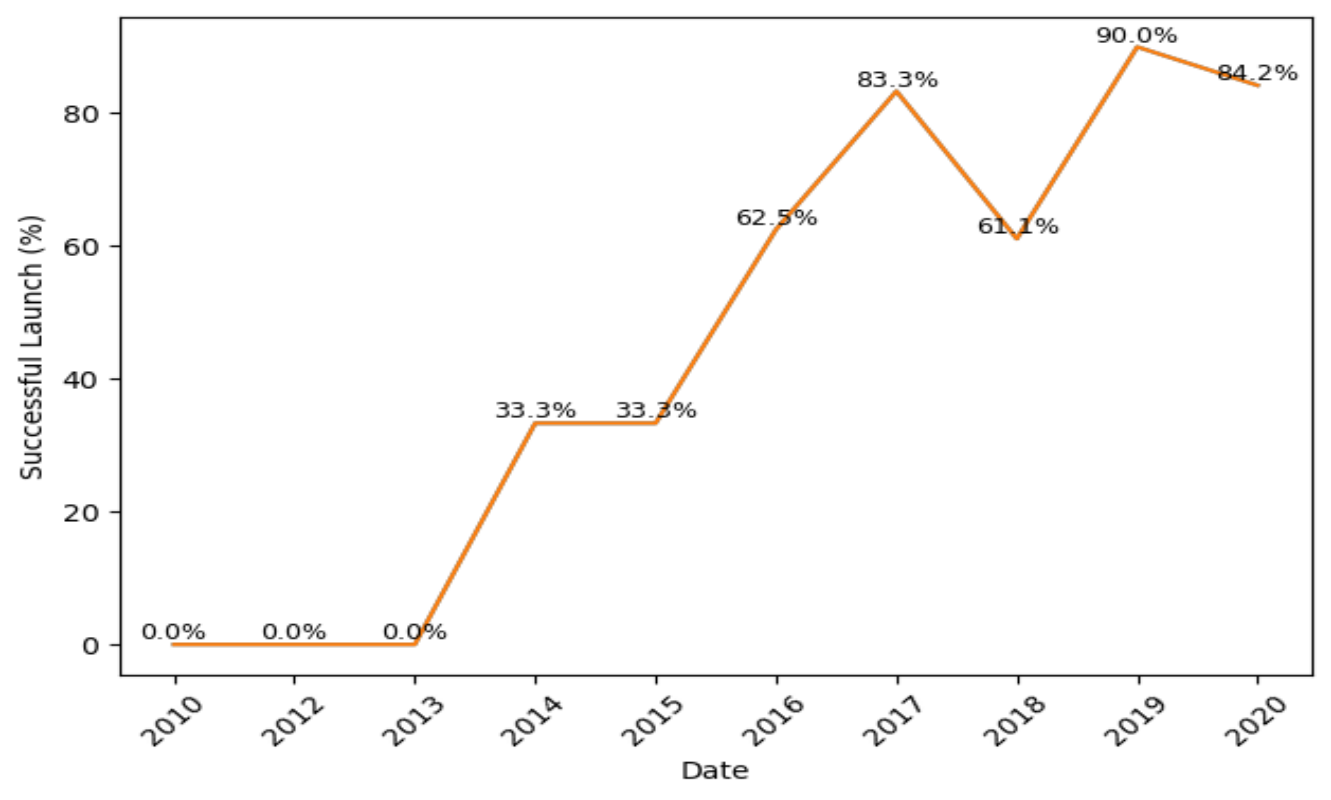


Figure 11 : Space X Launch Success % from 2010 to 2020.

- The general trend of launch success is observed to increase with from 2013 to 2017.
- The highest decrease of 22% in successful launch percentage is observed from 2017 to 2018.
- Highest success rate is recorded in 2019 followed with a minor dip by 6% in 2020.

All Launch Site Names (SQL)

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

Figure 12 : SQL Query results for all results.

Table 1 : Launch site code and its respective names.

Launch Site Code	Launch Site Name
CCAFS LC-40	Cape Canaveral Space Launch Complex 40
VAFB SLC-4E	Vandenberg Space Launch Complex 4
KSC LC-39A	Kennedy Space Center Launch Complex 39A
CCAFS SLC-40	Cape Canaveral Space Launch Complex 40 (Old Site)

Information on Launches from CCA (SQL)

Table 2 : Information on Launches from 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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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 From NASA (SQL)

SUM(PAYLOAD_MASS__KG_)
45596

Figure 13 : Query results for total payload mass by NASA.

- Total Payload Mass by NASA (Customer) is **45596.00 Kg**

Average Payload Mass by F9 v1.1 Booster

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

Figure 14 : SQL Query results for average payload mass by F9 v1.1 Booster.

- Average Payload Mass by F9 v1.1 booster is **2535.67 Kg**

First Successful Ground Landing Date

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Figure 15 : SQL Query results for first successful ground landing date.

- First Successful Ground Landing Date is on **2015-12-22**

Successful Drone Ship Landing with Payload between 4000 and 6000 Kg

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

Figure 16 : SQL Query results of booster version landed on drone ship with payload ranging from 4000 kg to 6000 kg.

- Booster versions that have successfully landed with payloads ranging between 4000 kg and 6000 kg include F9 FT versions B1022, B1026, B1021.2, and B1031.2.

Total Number of Successful and Failure Mission Outcomes

Table 3 : Measure of mission outcome.

Mission Outcome	Counts
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Mission_Outcome	COUNT(Mission_Outcome)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Figure 17 : SQL Query results for mission outcome.

- 99% of the mission outcome have been successful.

Boosters Carried Maximum Payload

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

Figure 18 : SQL Query results of booster version carrying maximum load.

2015 Launch Records

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

Figure 19 : SQL Query results of launch record for failed drone ship landing.

- Two launches were conducted in 2015. In January 2015, the F9 v1.1 booster B1012 was launched, followed by the F9 v1.1 booster B1015 in March 2015.
- Both boosters were launched from Cape Canaveral Space Launch Complex 40.

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

Landing_Outcome	COUNT(Landing_Outcome)
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Figure 20 : Query results for landing outcome from 2010-06-04 to 2017-03-20 in descending order.

- 60% of the attempted landing has been successful.
- Ground pad landing has had 100% landing success rate while drone ship has 60% success rate.

Section 2.2

Launch Sites Proximities Analysis

Launch Site Locations.

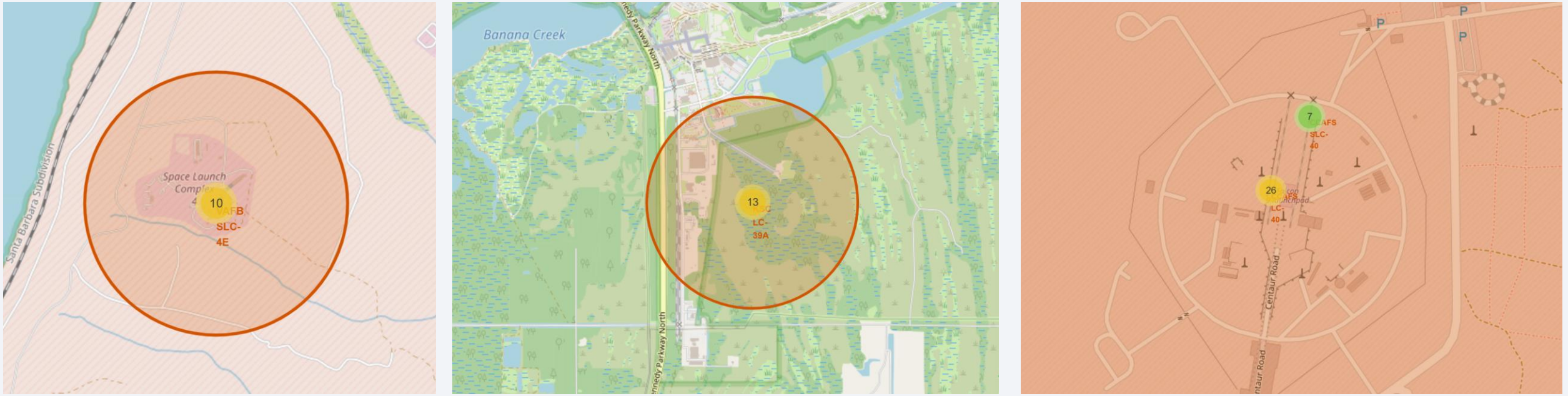


Figure 21 : VASF , KSC , CCAFS SLC & CCAFC LC (left to right) marking on a map.

- All sites are in close proximation to transportation routes particularly seashore
- CCAFS SLC has the most launches as shown In the third image.

Landing Outcome Visualization.

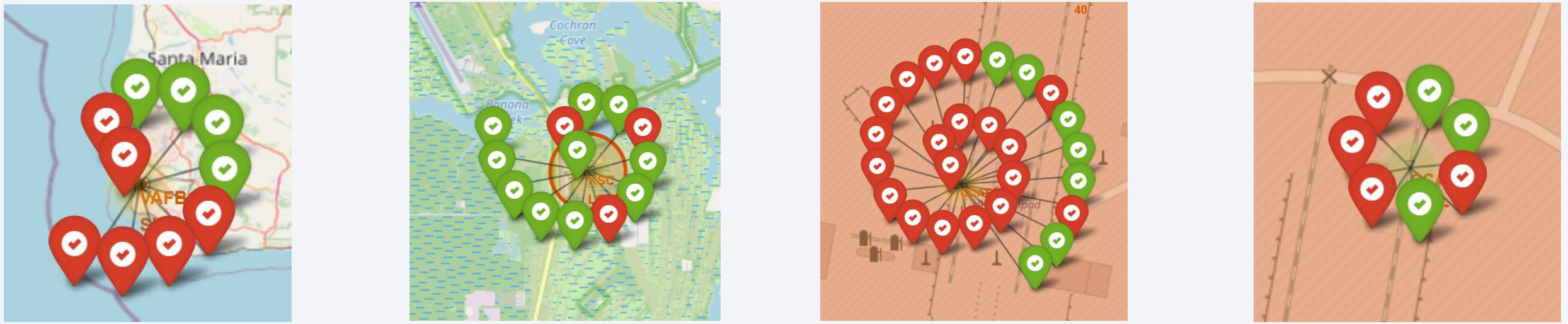


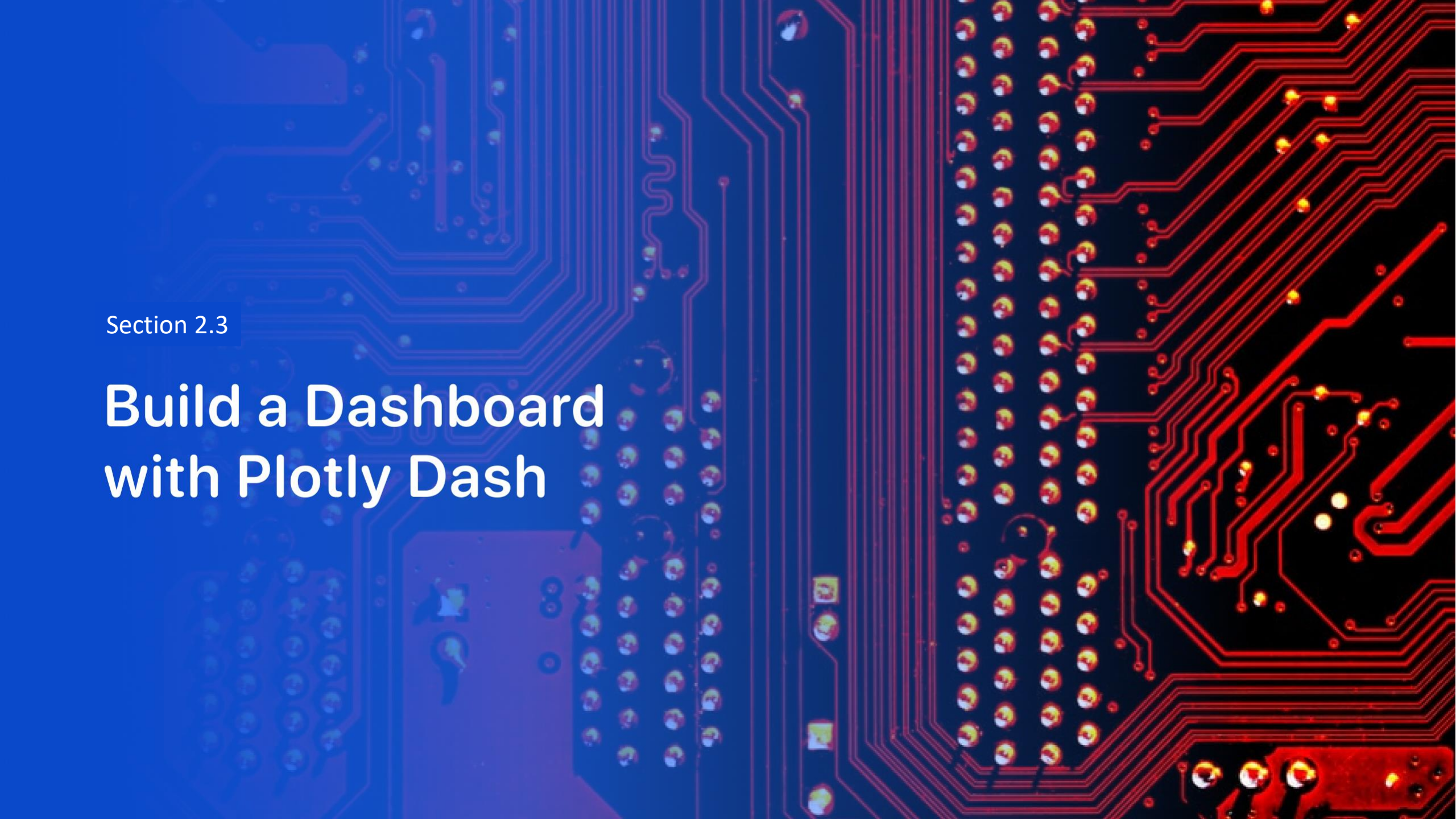
Figure 22 : Visualization of successful landing (green) for each site , VASF , KSC , CCAFS SLC & CCAFC LC (left to right).

- KSC is seen to show to have the highest successful landing.

Seashore Distance From Seashore

Figure 23 : Visualization of site proximity from seashore , VASF , KSC , CCAFS SLC & CCAFC LC (left to right).

- KSC has the longest distance from seashore at 7.8 KM while CCAFS SLC has the lowest distance at 0.9 KM.


Section 2.3
Build a Dashboard with Plotly Dash

Successful Launches at All Sites

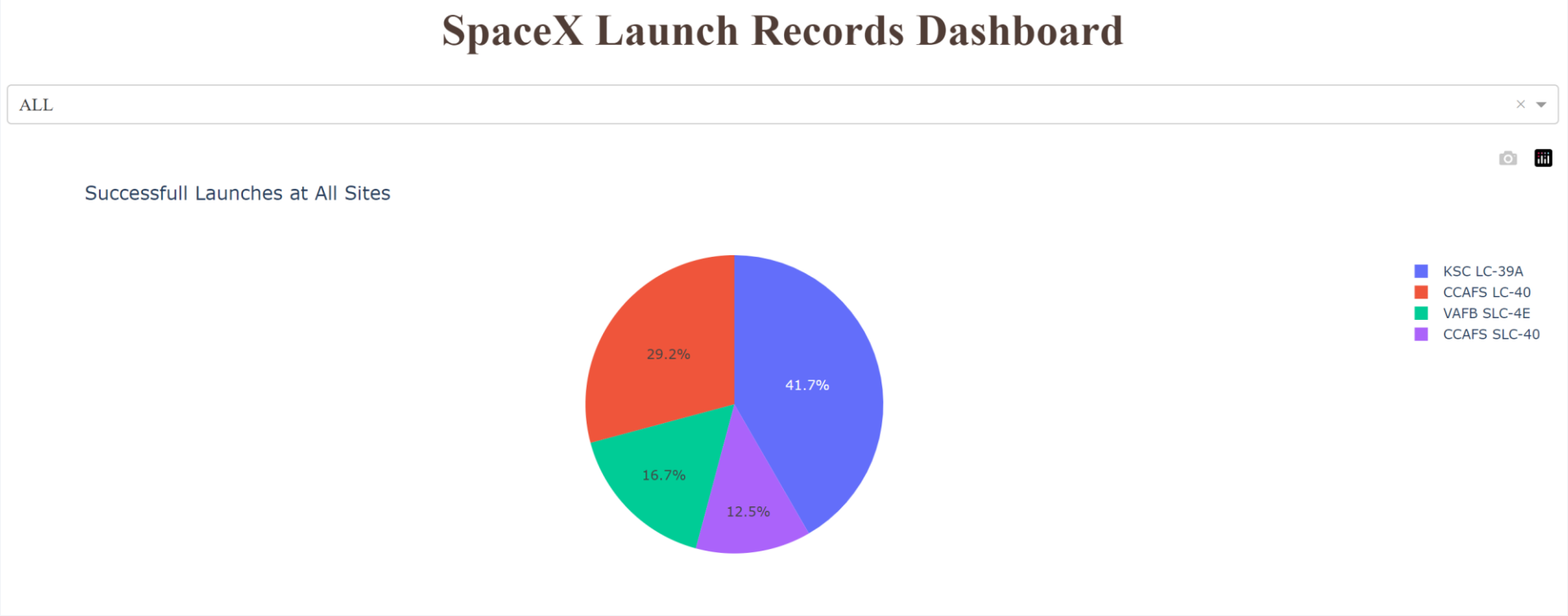


Figure 24 : Pie chart feature on the SpaceX launch record dashboard .

- Highest successful launches happened at KSC-LC -39A.

Success Rate for KSC LC-39A

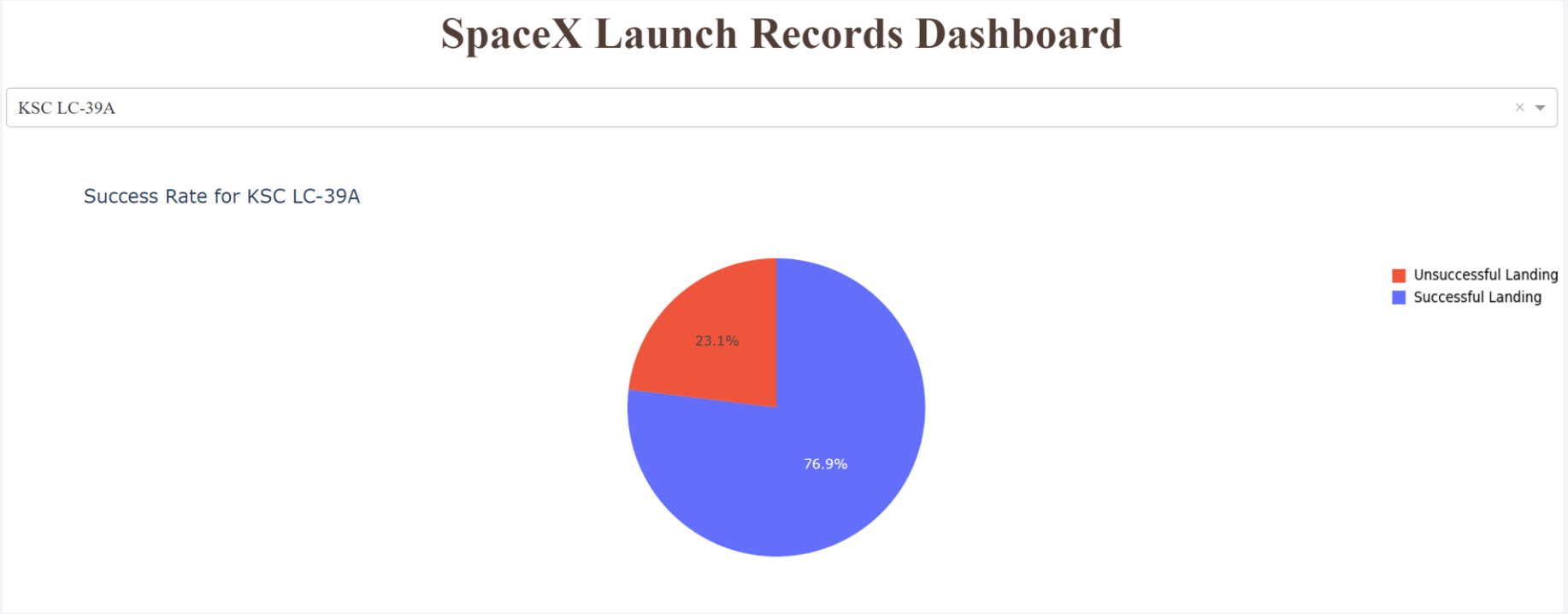


Figure 25 : Pie chart feature on the SpaceX launch record dashboard .

- KSC-LC -39A has a success percentage of 76.9%

Payload mass against launch site

Figure 26 : Scatter plot of payload mass against its launch site , payload mass ranging from 0kg to 1000 kg (left) , payload mass ranging from 1000kg to 6000 kg.

- Highest payload mass is 9600 kg at VAFB site.
- Payload mass ranges :

  CCAFS LC = 1 kg to 5200 kg , KSC = 2500 kg to 6800 kg

  CCAFS SLC = 370 kg to 6100 kg , VAFB = 480 kg to 9600 kg

Section 2.4

Predictive Analysis (Classification)

Classification Accuracy

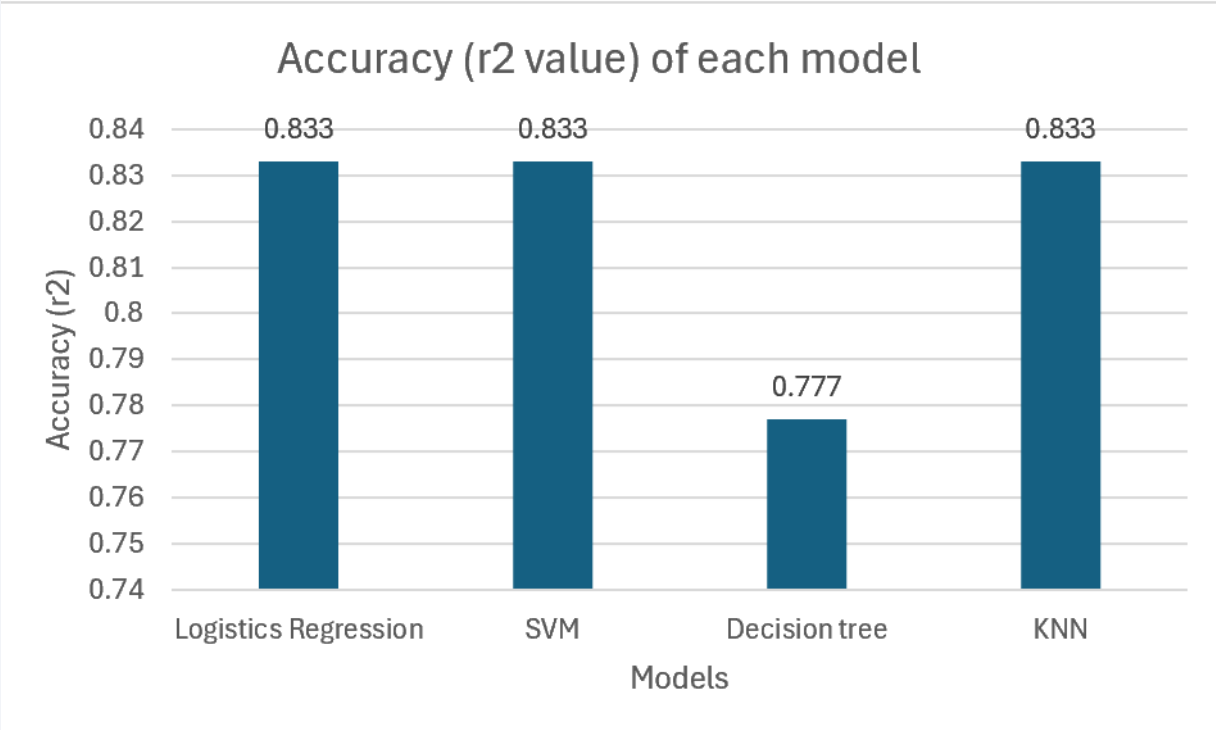


Figure 27 : Accuracy of each model.

- LR , SVM & KNN has a highest accuracy at a value of 0.833.
- Larger test data set would be optimal in acquiring a more realistic accuracy value.

Confusion Matrix

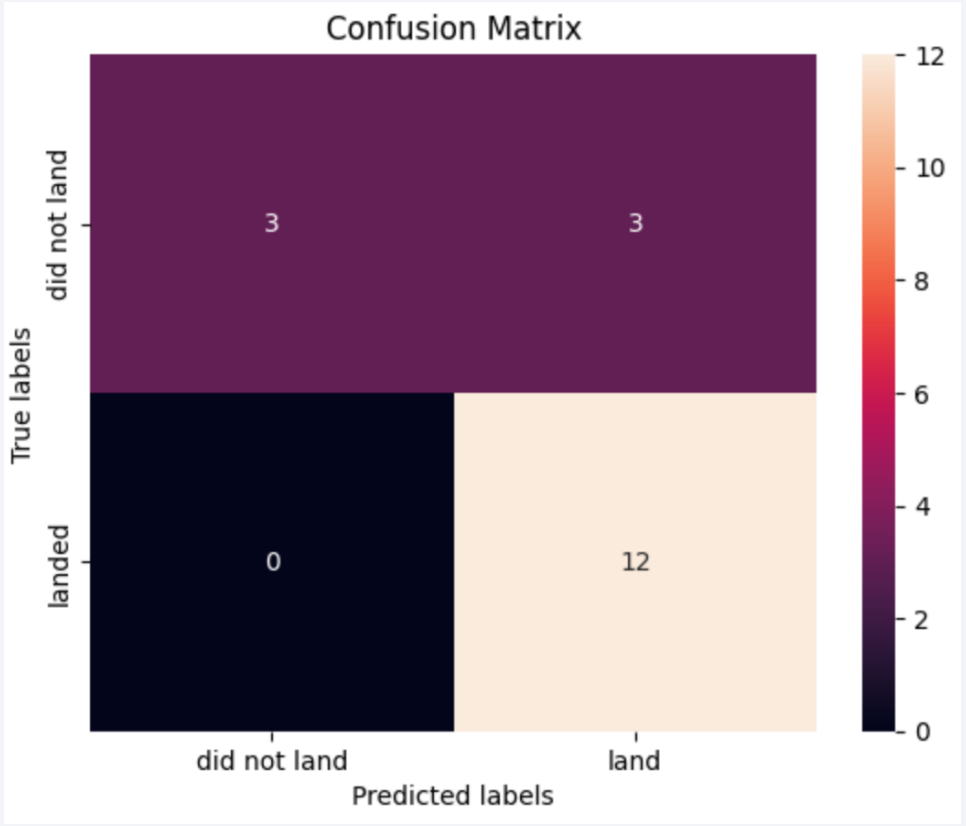


Figure 28 : Confusion matrix for svm model.

- SVM is chosen over the other models as KNN is not suitable for larger data sets and LR is less tolerant to outliers .
- The model produced 3 false negative results out of 18 data tested.

Optimal parameter for SVM is given in the appendix.

Conclusions

- KSC LC 39A launch site is seen to have the highest successful landing outcome.
- ES-L1 ,GEO, HEO & SSO orbit has a 100% landing success rate.
- The SVM model was selected as the top choice because of its impressive 83% accuracy and its robustness in handling outliers.
- The accuracy of the models can be improved in future studies by increasing the test sample size.

Appendix

- **Variables from SpaceX API -** Flight Number ,Date , Booster Version , Payload Mass, Orbit, Launch Site, Outcome, Flights, Grid Fins , Reused Legs, Landing Pad , Block, Reused Count ,Serial ,Longitude , Latitude

- **Optimal SVM parameter** - C : 1.0, gamma : 0.0316, Kernel : sigmoid

Thank you!