IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

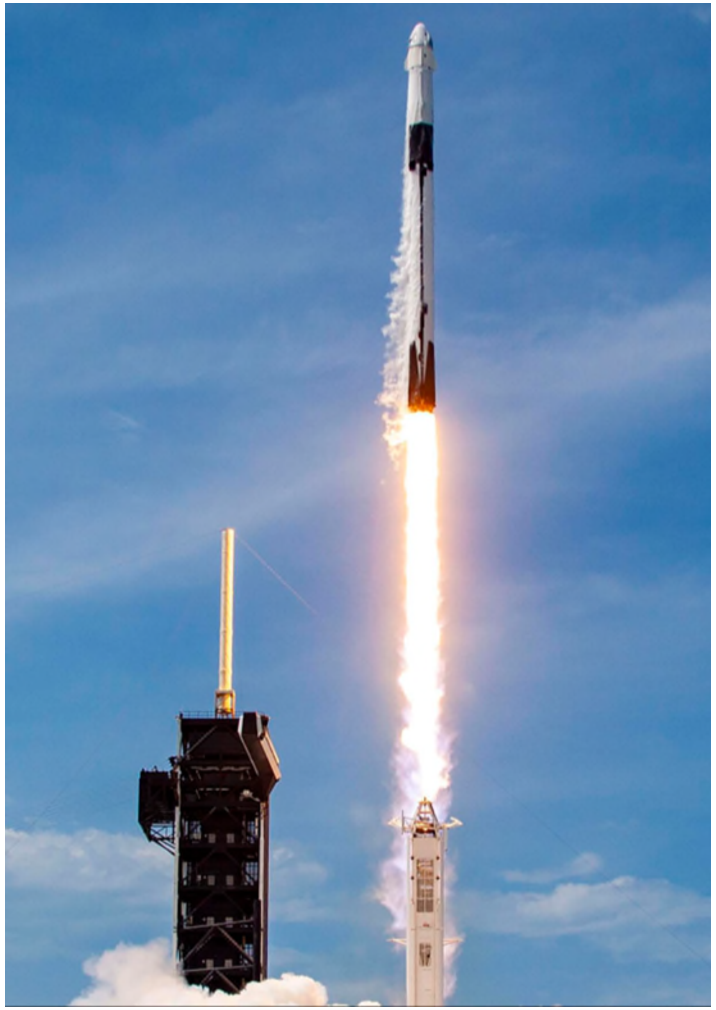
20 Jan 2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- The presentation aims to illustrate the factors for a successful rocket landing. The following methodologies were used:
- Collect data using SpaceX Rest API and web scraping techniques
- Wrangle data to create success/fail outcome variable
- Explore data with data visualization techniques
- Analyse the data via SQL queries
- Explore launch site success rates and proximity to geographical markers
- Visualise the launch sites with the most success
- Build ML Models to predict landing outcomes using logistic regression, support vector machines (SVM), decisions tree and K-nearest neighbour (KNN)

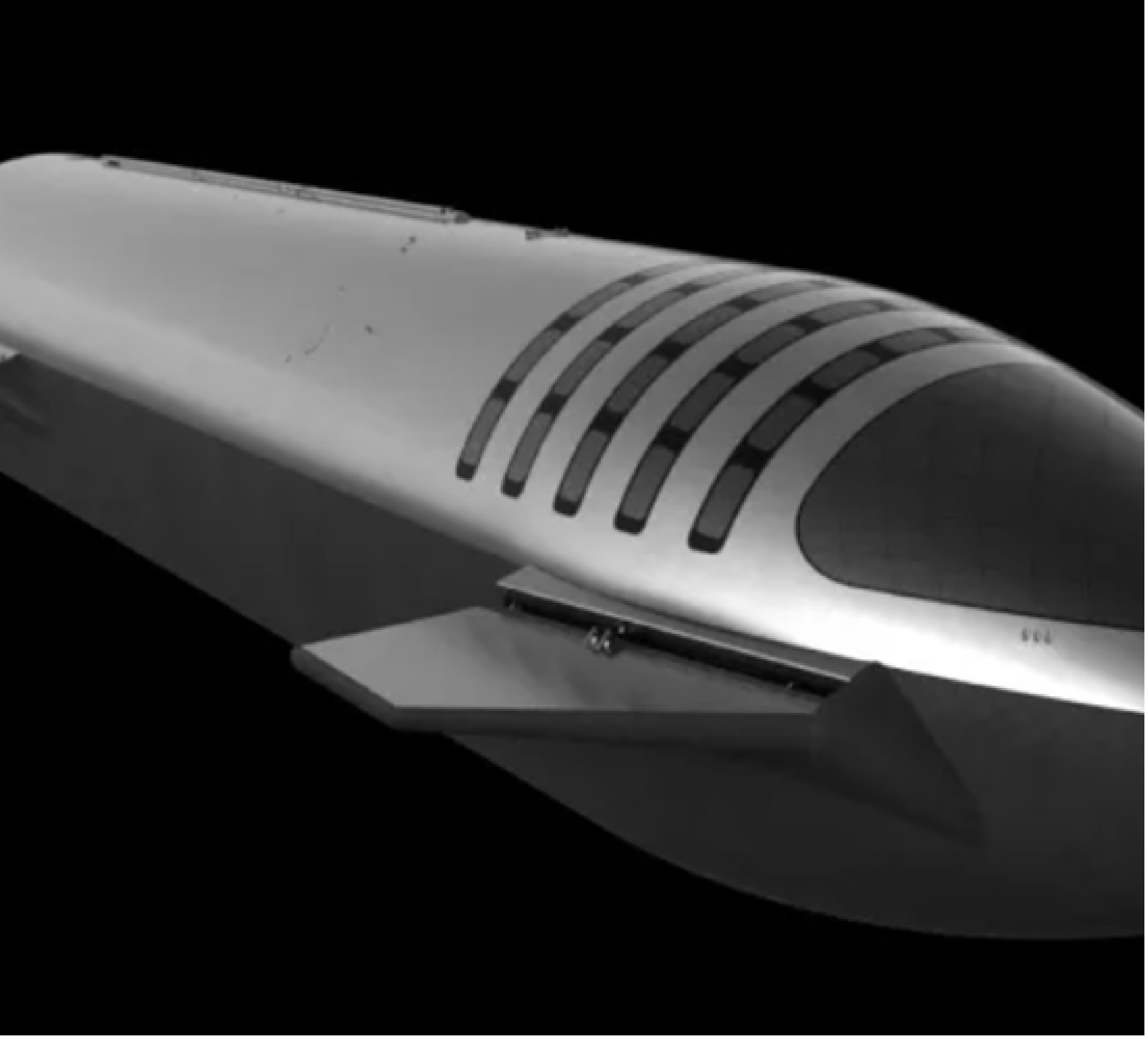


Introduction

- Falcon 9 is a reusable, two-stage rocket designed and manufactured by SpaceX for the reliable and safe transport of people and payloads into Earth orbit and beyond. We can determine the price of the launch based on whether the first stage will land. We will use public data, resources and ML models to predict whether Space X can reuse the first stage.
- **Main Goals:**
- Whether variables such as payload mass, number of flights, orbits affect first-stage landing success
- Rate of success landings over time
- Best ML model for successful landing

Section 1

Methodology



Methodology

- The data was collected via SpaceXAPI. The information included payload, rocket used, launch specifications, landing specifications and landing outcome.
- Transforming data by filling the missing values, filtering and calculations
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Build, tune, evaluate classification models

Data Collection

[SpaceX API Link]

- Request: Request data from SpaceX API
- Create: Use json_normalize method to convert the json result into a dataframe
- Create: Create dataframe from the data
- Export: Filter the dataframe to only include Falcon 9 launches
- Create: Export data to csv

Data Collection
WebScraping Link

Request	•Request publicly sourced data
Create	•Create BeautifulSoup object from HTML response
Extract	•Extract column names from the HTML table
Collect	•Collect data from parsing HTML tables
Create	•Create dictionary from the data
Export	•Export data to csv
Create	•Create Dataframe from dictionary

Data Wrangling

- Calculated the number of launches on each site¶
- Calculated the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome of the orbits
- Created a landing outcome/class label from Outcome column
- 1 means successful landing and 0 means unsuccessful landing
- Link:

EDA with Data Visualization

- Plotted charts to illustrate how each of these variables (Flight Number, Payload Mass, Orbit, Launch Sites) affect the success rate of first launch.
- Flight Number vs Payload
- Flight Number vs Launch Site
- Payload Mass (kg) vs Launch Site
- Payload Mass (kg) vs Orbit type
- We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.
- We can also see that average launch success trend increases over the years by plotting a line chart
- Link

EDA with SQL

Performed the following queries:

- 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 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 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 (desc)
- SQL Link

Build an Interactive Map with Folium

- Added yellow circles to indicate all the launch site coordinates.
- Marked the success/failed launches for each launch site on the map – Green marker for success launches and Red marker for failure launces.
- Have also calculated the distances between the launch site to its proximities to nearest coastline, railway, highway and city using coloured lines.
- Folium Link

Build a Dashboard with Plotly Dash

- Made a drop down list containing all the Launch Sites
- Create a Pie Chart showing the % of successful launches amongst sites
- Slide of Payload Mass to select the payload range
- Scatter Chart to showcase the correlation between Payload and Launch Success along with Booster versions.
- Link

Predictive Analysis (Classification)

- Created a Numpy Array from the column Class from a datafram
- Standardized the data
- Splitted into training data and test data
- Created a logistic regression, SVC, KNN then created a GridSearchCV (c = 10 for parameters) object for all of the models
- Calculated the accuracy score, F1_Score and Jaccard_score
- Created confusion matrix
- Identify the best models through accuracy score, F1_Score and Jaccard_score
- Link

Results

Launch success rate since 2013 kept increasing till 2020

KSC LC 39-A has the highest success rate of all other landing sites

Most launch sites are near the coastline and near the Equator line. They are also close to railway, highway and other cities.

ES-L1, GEO, HEO, SSO are the orbits with 100% success rate

Predictive analytics

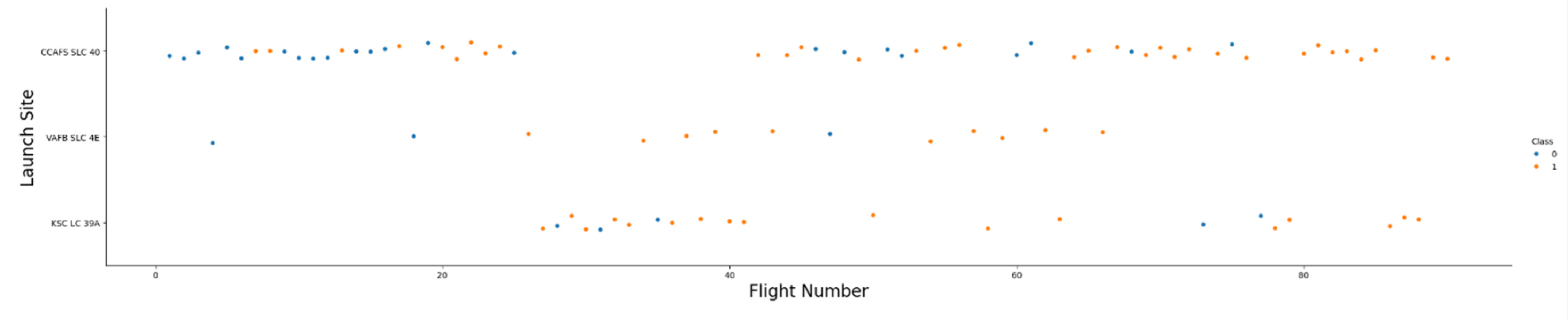
Decision Tree Outperforms other models by a slight margin

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

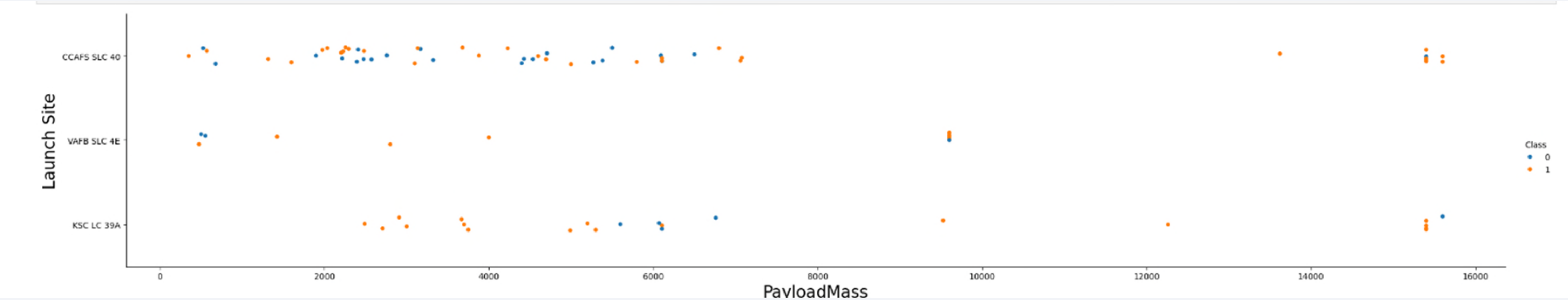
- Blue is unsuccessful and orange is successful
- CCAFS SLC 40 has the most launches
- VAFB SLC 4E and KSC LC 39 has higher success rate.
- The more the flights has launched – the more the launch site is successful.



Payload vs. Launch Site

The higher the Payload Mass, the higher chance of being successful

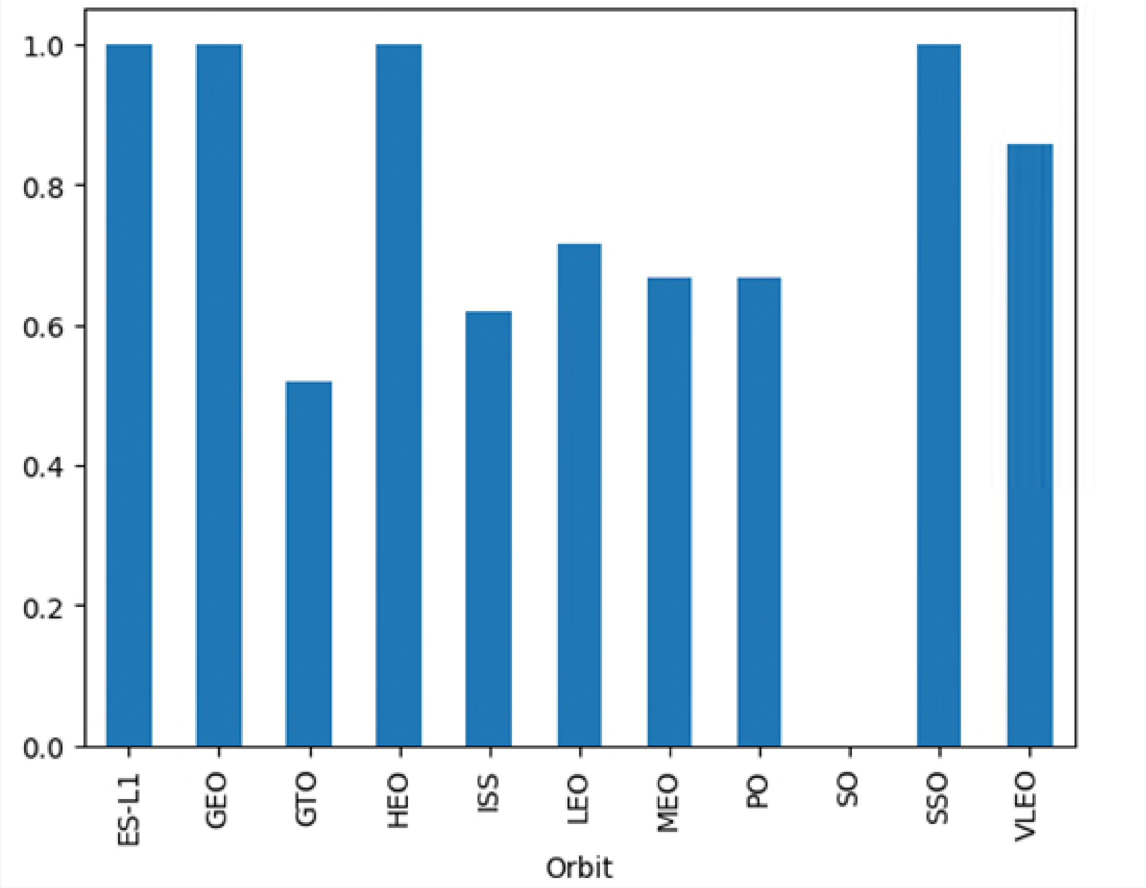
Most rockets have Mass of 2k – 6k



Success Rate vs. Orbit Type

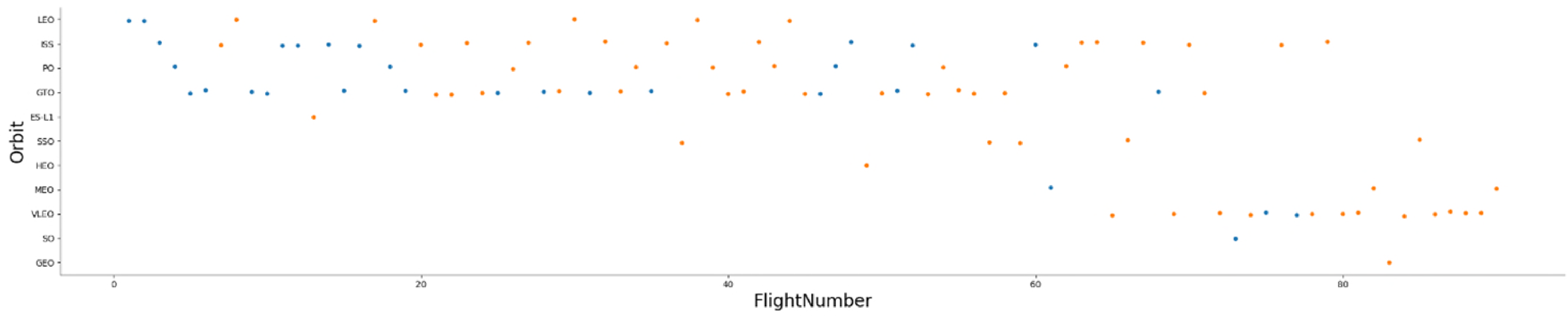
ES11, GEO, HEO and SSO has 100% success Rate

Lowest one is SO which has 0%



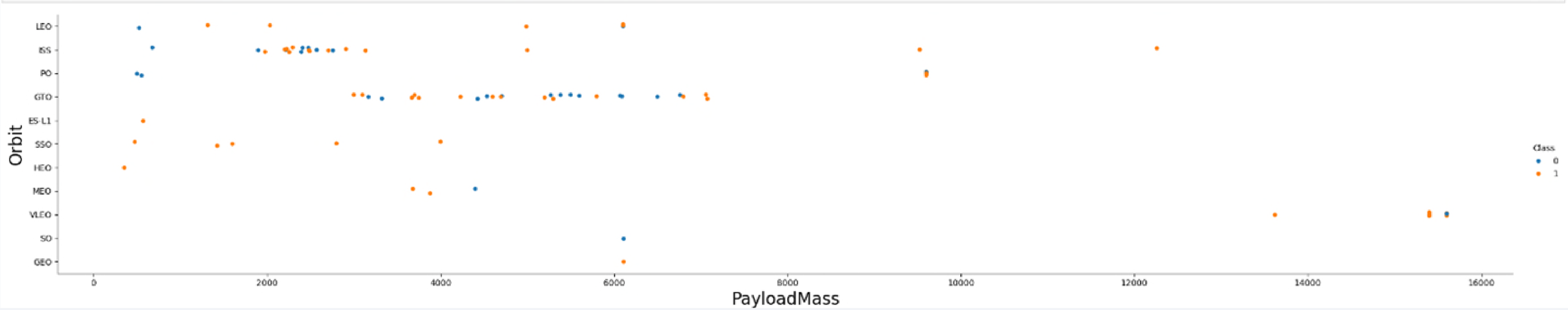
Flight Number vs. Orbit Type

- The more flight numbers , the more successful are the launches.



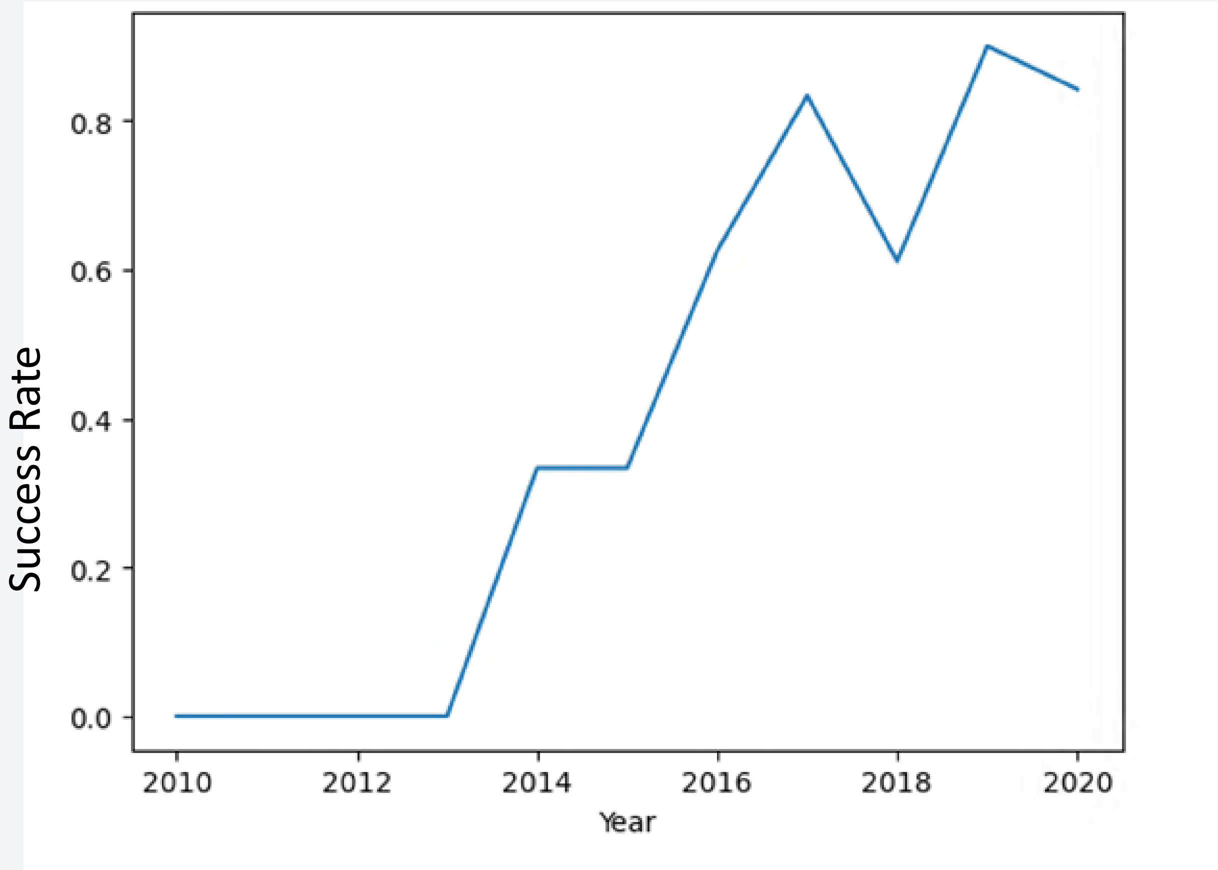
Payload vs. Orbit Type

- ISS, LEO and PO deal heavier payload mass better



Launch Success Yearly Trend

- The overall trend is increasing.
- Improving significantly from 2015 - 2014



All Launch Site Names

Launch Site List

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

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with `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

- Total payload carried by boosters from NASA in kg

sum("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 in kg

avg("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad

min("Date")
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

- List of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

- List of the names of the booster which have carried the maximum payload mass

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

2015 Launch Records

- List of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

Months	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

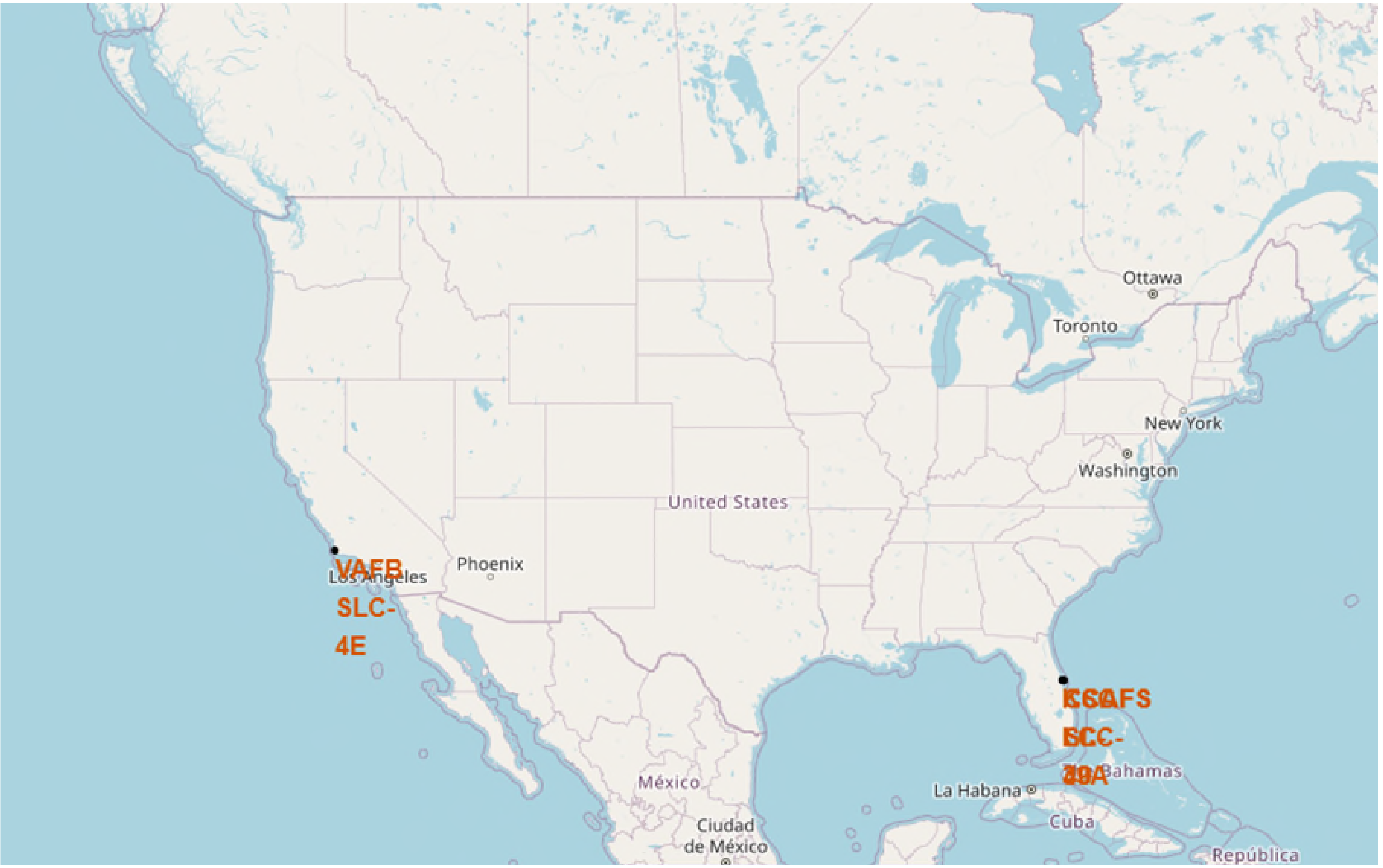
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

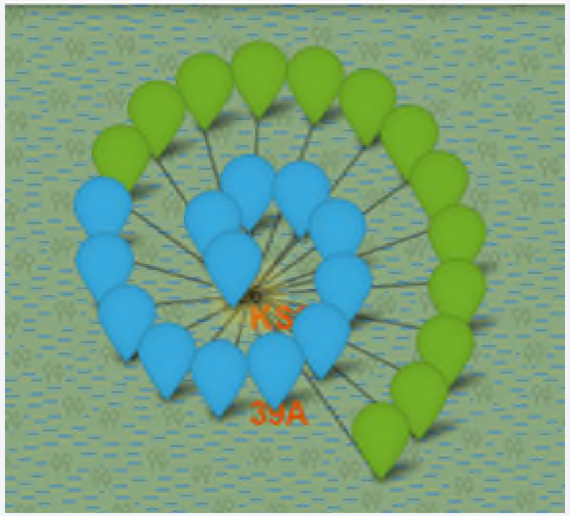
Section 3

Launch Sites Proximities Analysis

Launch Sites on the map

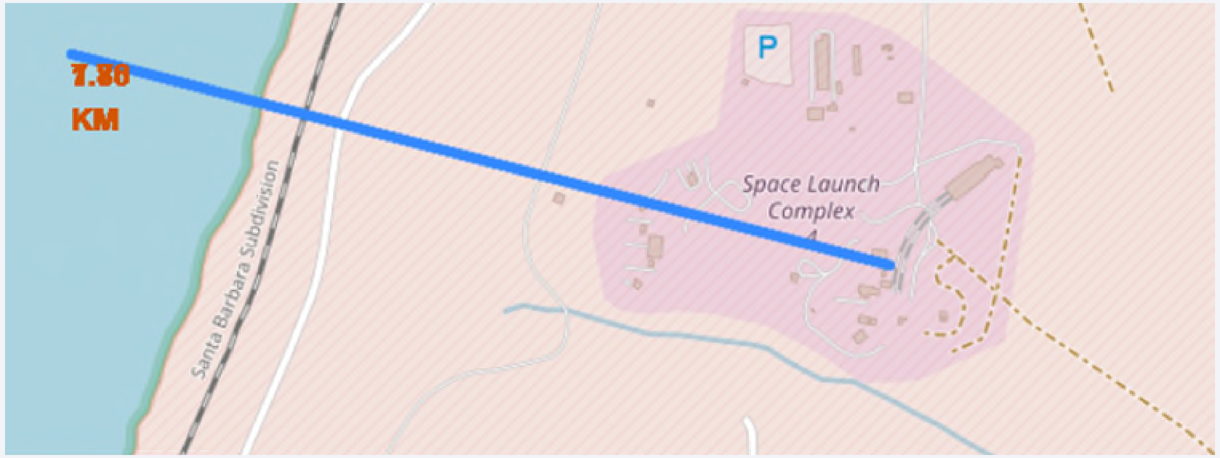


Mark the success/failed launches for each site on the map



Marked the success/failed launches for each launch site on the map – Green marker for success launches and Red marker for failure launces.

The distances between a launch site to its proximities

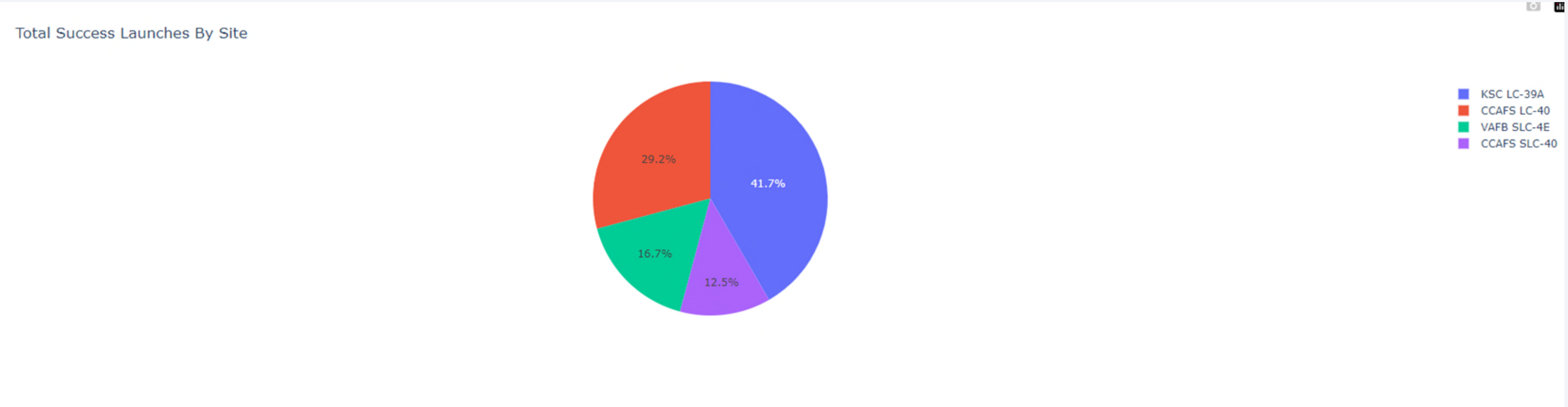


The launch site is pretty close to the railway, highway and to the city

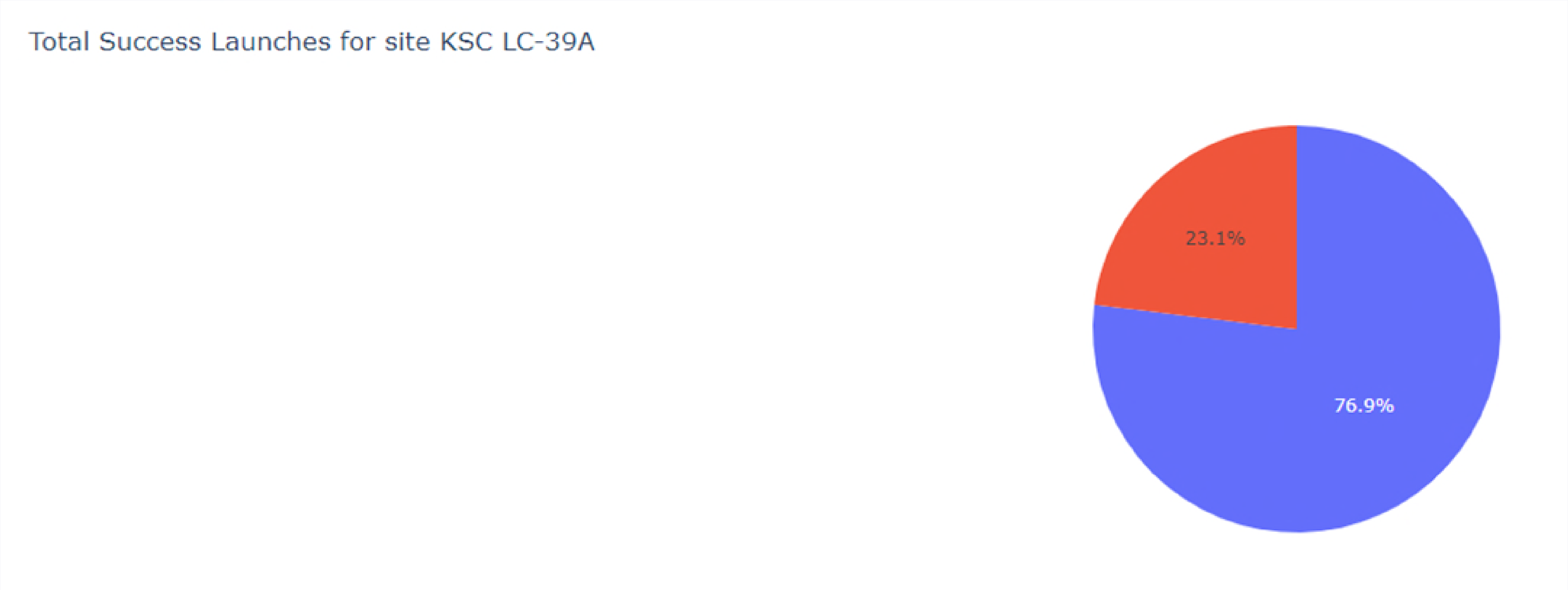
Section 4

Build a Dashboard with Plotly Dash

Total Success Launches by Site



Highest Launch Ratio



Correlation between Payload and Success for All sites

- Payload between 2k and 6k tend to launch more successful launches. FT booster versions tend to be the most successful of all Booster version

Section 5

Predictive Analysis (Classification)

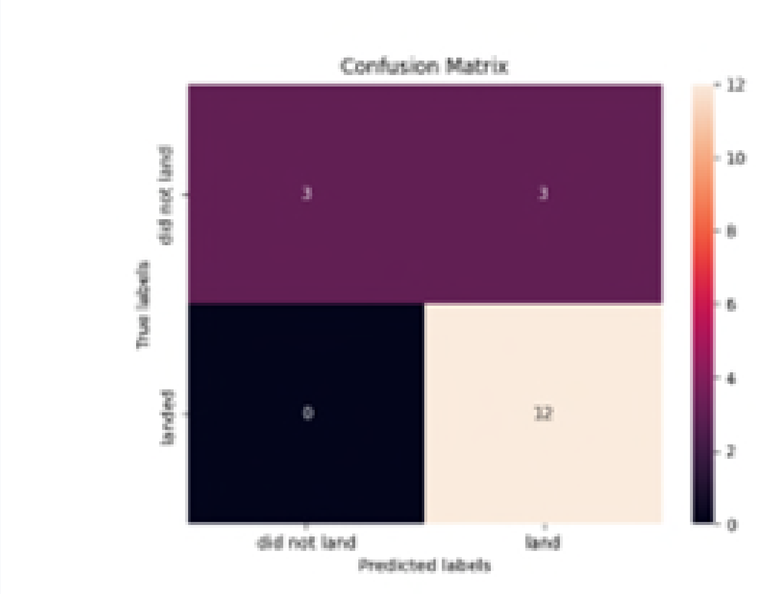
Classification Accuracy

- Decision Tree outperforms other models by slight margin

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

Confusion Matrix

- There is 3 false positive produced by the Decision Tree. Other models performed similarly in terms of Confusion Matrix



Conclusions

- Launch success rate since 2013 kept increasing till 2020
- The higher the payload mass, the higher chance of being successful
- KSC LC 39-A has the highest success rate of all other landing sites
- Most launch sites are near the coastline and near the Equator line. They are also close to railway, highway and other cities.
- ES-L1, GEO, HEO, SSO are the orbits with 100% success rate
- **Predictive analytics**
- Decision Tree Outperforms other models by a slight margin

Thank you!