IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Cristian Mauricio Romero Buitrago
30/01/2023

Outline

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

Executive Summary

The commercial space age is here, the SpaceY founded by Allon Mask was conducted a study about is rival SpaceX where the methods are used:

- Collected data from spaceX Rest API
- Web Scraping from Wikipedia databases with Python
- EDA with Data Wrangling
- EDA with SQL
- EDA with Pythons Pandas and Matplotlib
- Interactive Visual Dashboard
- Predictive analysis with Python libraries

with this methods the spaceY company that was possible collect valuable data from public source, perform EDA allow to identify which features are the best to predict success landings and with machine learning predictions showed the probability about the futures launchings success.

Introduction

The Commercial Space age is here, and many companies are making the space travel affordable for everyone. SpaceX is the most successful space travel company among them.

The reason SpaceX has been able to achieve this is because their rocket launches are inexpensive. SpaceX advertises Falcon 9 rocket launches on its website at a cost of $62 million, while the cost for other providers is about $165 million each. Much of the savings SpaceX makes is due to their ability to reuse the first stage.

the new company rival of SpaceX is SpaceY, the purpose of this study is research how spaceX perform de succes landings for reuses the landing pad and what is his launching and landing success probability.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from SpaceX API https://api.spacexdata.com/v4/launches
  - Web Scraping from List of Falcon 9 and Falcon Heavy launches - Wikipedia
- Perform data wrangling
  - The Data collected was analyzed, processed and classification by his outcome type
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The method used was GridSearchCV to tune the machine learning model and find the best parameters

Data Collection

The Data collection process about SpaceX was realized by extraction of API (API link) and Web Scraping (Wikipedia link).

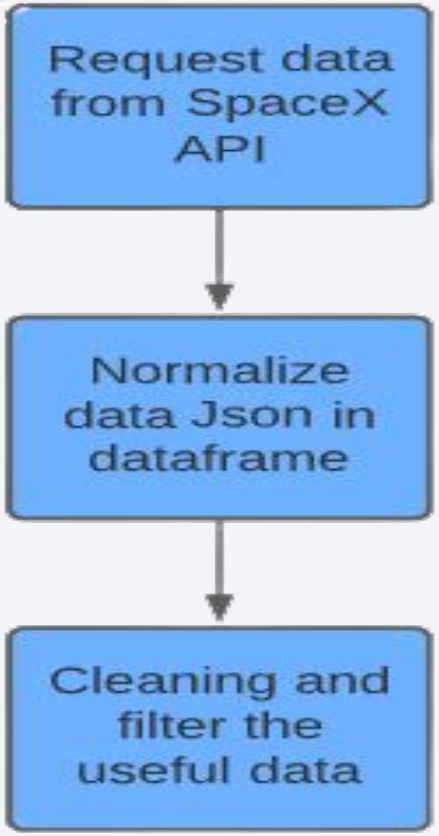
API Data Columns: FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude .

Web Scrape Data Columns: Flight No, Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time.

Data Collection – SpaceX API

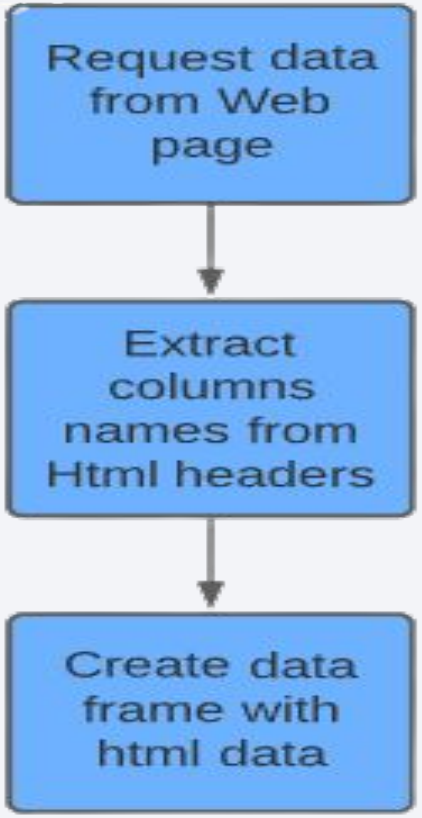
- The data were collected from API of SpaceX public link (API link)

- Source Code: API GitHub



Data Collection - Scraping

- The data were collected from wikipedia web page about SpaceX (Wikipedia link)
- Source Code: Web Scraping GitHub



Data Wrangling

- First initially with EDA was performed on the dataset.
- Second the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated
- Third and finally the landing outcome label was created from Outcome column.

Source code: [Wrangling GitHub]

EDA with Data Visualization

The EDA performed with visualization represent the relationships between columns over a period. this graphics are useful for understand the relevance of some columns compared to others the easy way.
the graphis realized was:

- Scatter Plots:
  FlightNumber vs PayloadMass, FlightNumber vs LaunchSite, LaunchSite vs PayloadMass, FlightNumber vs Orbit Type and PayloadMass vs Orbit Type
- Bar Charts:
  Success Rate vs Orbit Type
- Line Chart:
  Year vs Success Rate

Source code: EDA visualization GitHub

EDA with SQL

The following SQL queries were performed:

- 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 successful landing outcome in ground pad was achieved.
- 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.
- List the failed landing outcomes in drone ship, their booster versions, and launch site names for 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

Source code:EDA SQL GitHub

Build an Interactive Map with Folium

The differents objects in the map was created for mark the most important locations:

- Markers were used to show points on the map like all launch sites.
- Circles were used to show areas around specific coordinates on the map, like Johnson Space Center at Houston.
- Mark clusters were used to show the many points together on the map.
- Lines were used to show the route between coordinates.

Source code: Folium GitHub

Build a Dashboard with Plotly Dash

The Dashboard contains a Pie Chart and a Scatter Plot which helped to identify the best place to launch according to the payload mass.

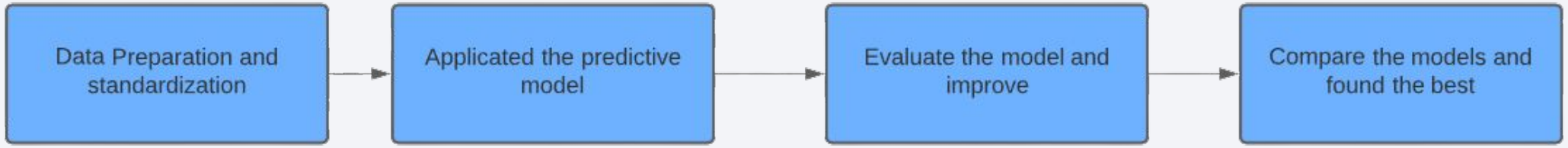
- The Pie Charts were used to show the distribution of successful landings across all launch sites. There is a drop-down option which shows each individual launch site and their success rates.
- The Scatter Plots shows these inputs: all launch sites and individual launch sites, Booster Version Category and PayloadMass (kg) on a slider between 0 and 10,000 kg. Also shows how the success rate varies across different Launch Sites, Payload Mass and booster version categories.

Source code: Dashboard Dash GitHub

Predictive Analysis (Classification)

The process of the predictive analysis was:

- The data were prepared and standardized to aplicated the predictive model.
- The model was evaluated and improve for obtain major percentage of prediction.
- When the models are refined then compared all types of model realized and select the most high predict percentage



Source code: [Predictive GitHub]

Results

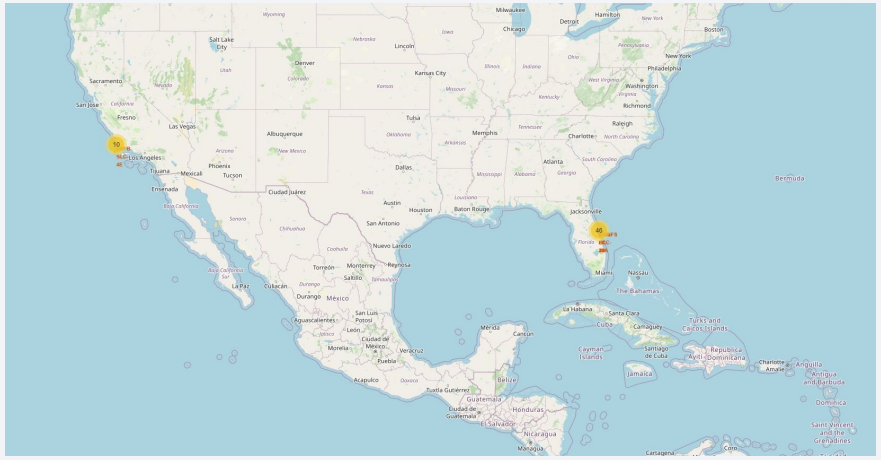
Exploratory data analysis results:

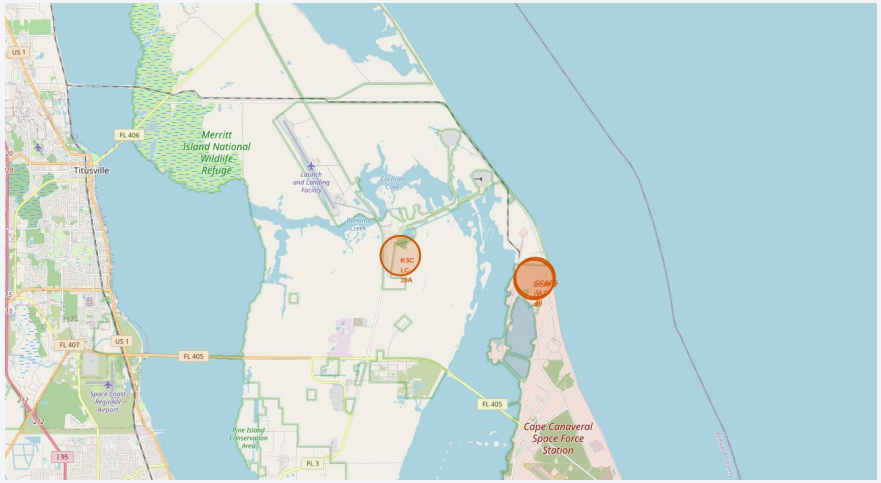
- SpaceX uses 4 different launch sites
- The first launches were done by Space X on 06/04/2010
- The average payload mass of F9 v1.1 booster is 2,928 kg
- The first success landing outcome happened in 2015, five years after the first launch in 2010
- Many Falcon 9 booster versions were successful at landing in drone ships having payload mass above the average
- Almost 100% of mission outcomes were successful
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015
- The number of landing outcomes became better as the years passed.

Results

Interactive analytics demo in screenshots:

Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around. Most launches happens at east coast launch sites.



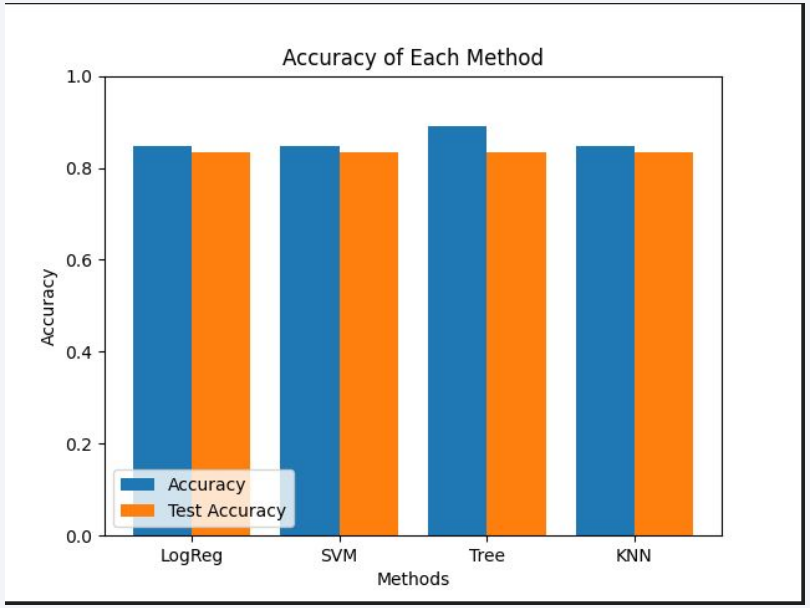


Results

Predictive Analysis results:

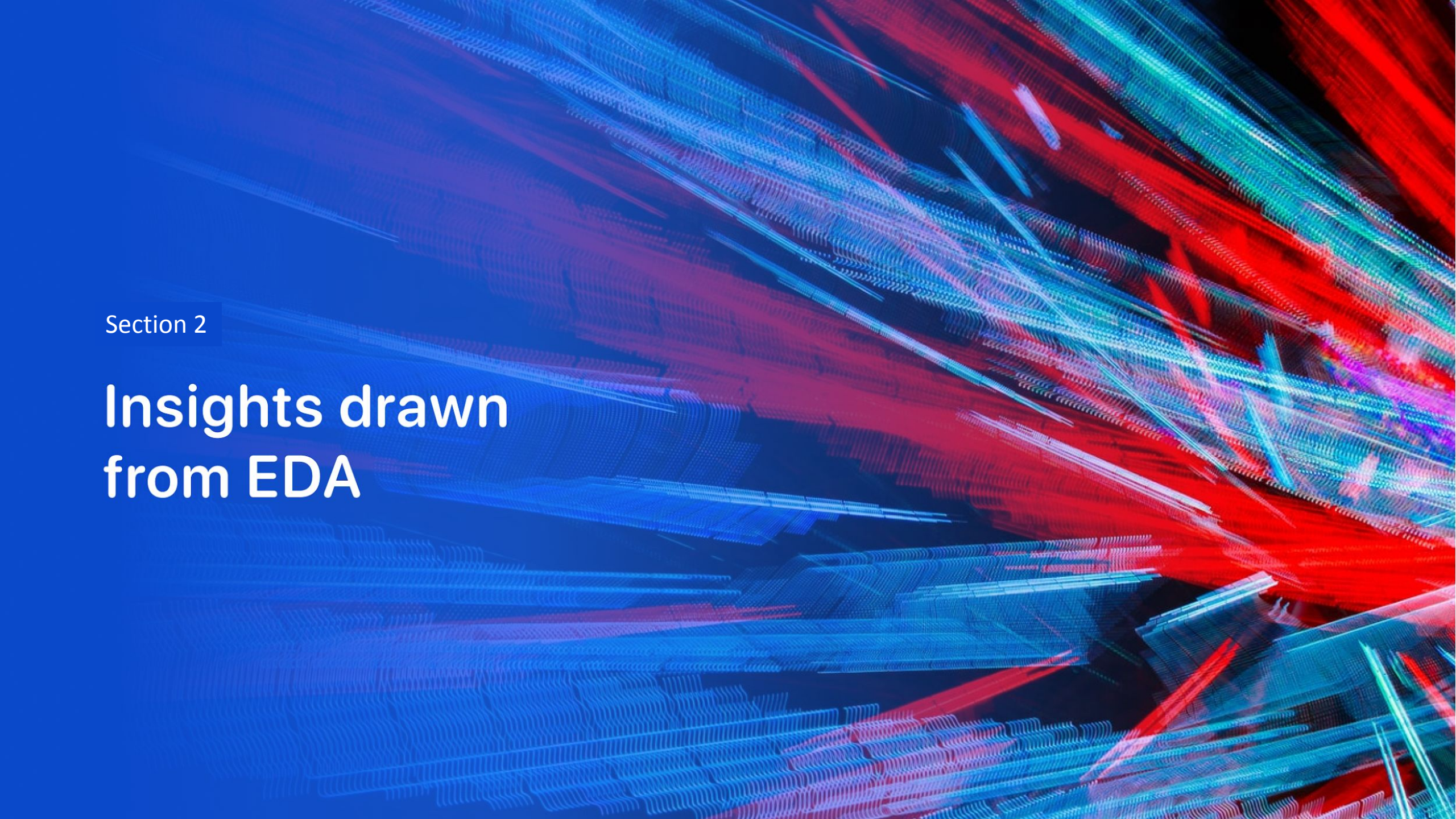
The predictive analysis showed that the Decision Tree Classifier is the best model to predict successful landings with an accuracy score of 89.1%, as compared to the other models.

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.89107	0.83333
KNN	0.84821	0.83333

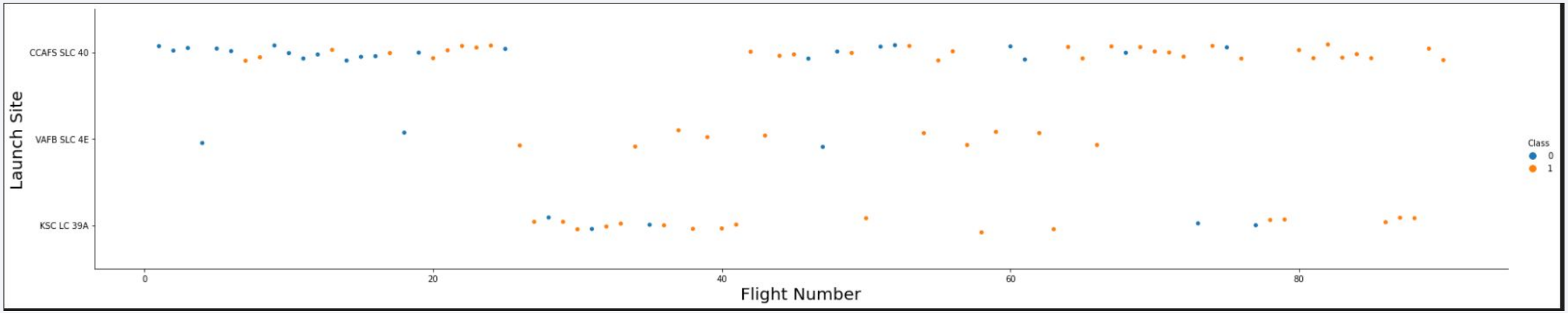


Section 2

Insights drawn from EDA

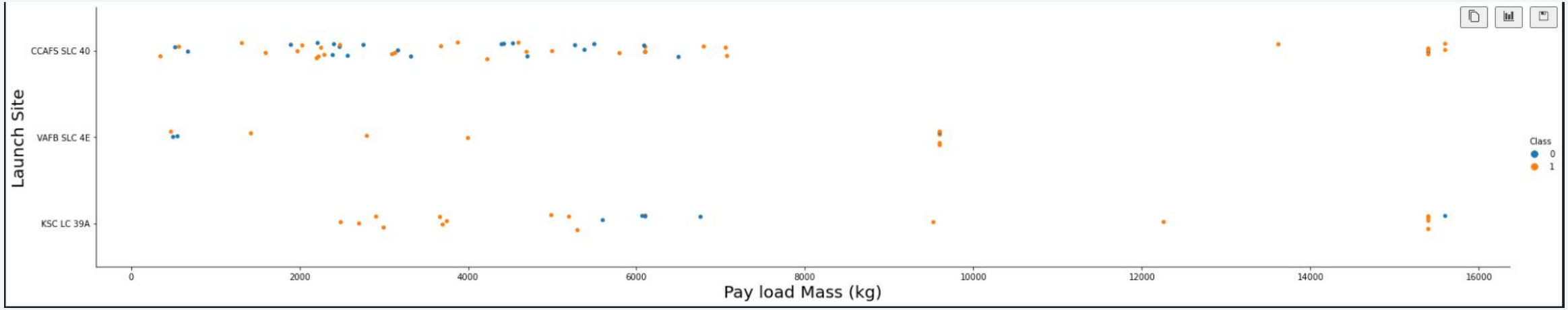


Flight Number vs. Launch Site



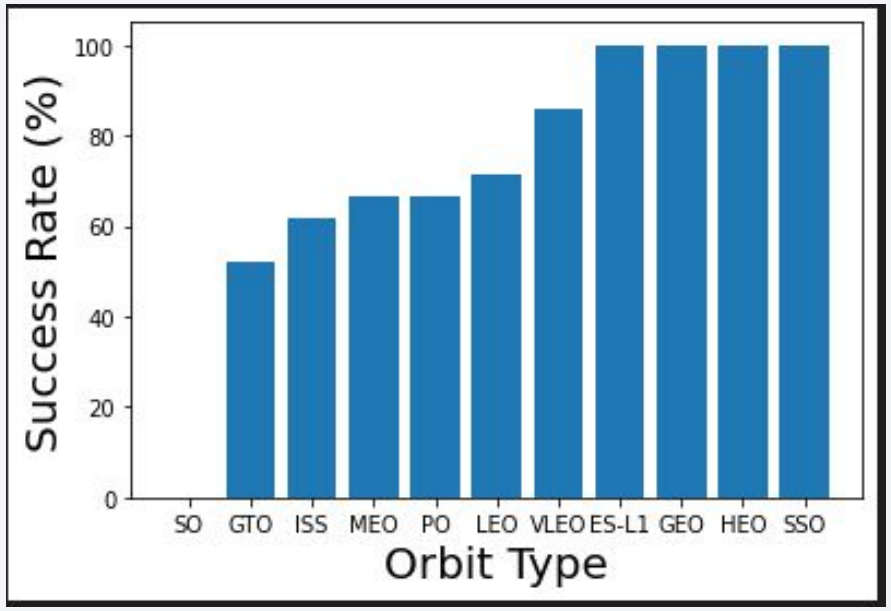
- In the graph it can be seen that the most rate of successful launches happened at CCAFS SLC 40
- also in the graph it can be seen that the general success rate improved with the launch

Payload vs. Launch Site



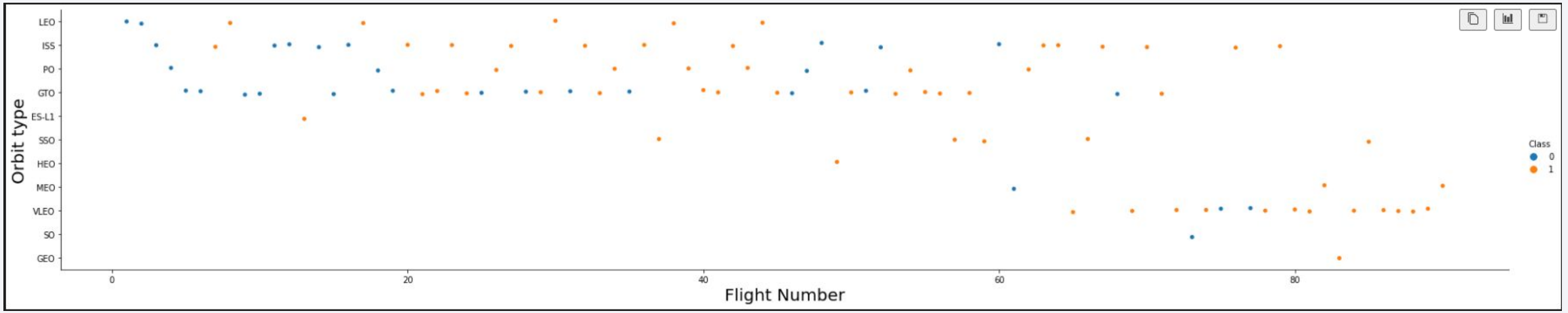
- Payloads mass over 9000 kg have high success rate
- Payloads mass over 12000kg only have tryings on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type



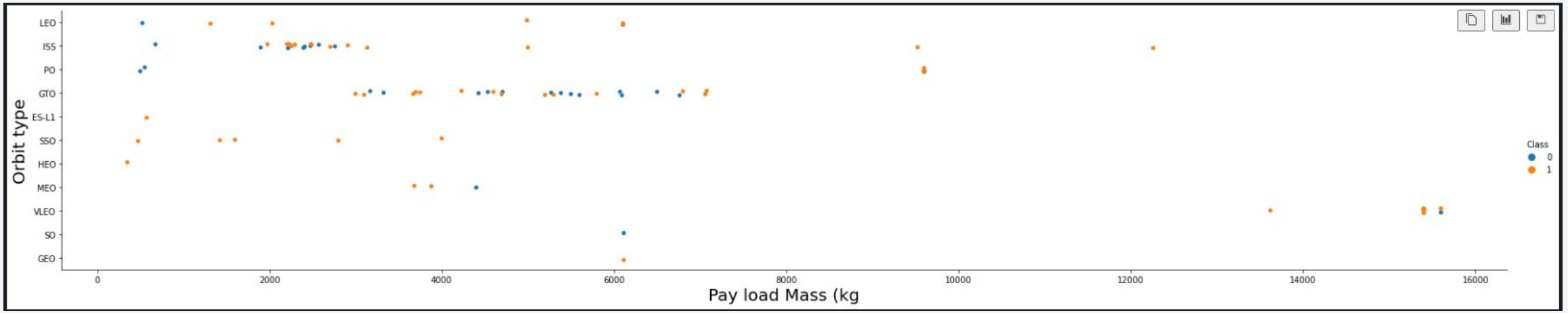
- Orbits ES-L1, GEO, HEO and SSO had success rates of 100% and 100% is the mode
- The lowest success rate orbit was SO with 0%
- Orbits MEO and PO had the qual success rate of about 70%

Flight Number vs. Orbit Type



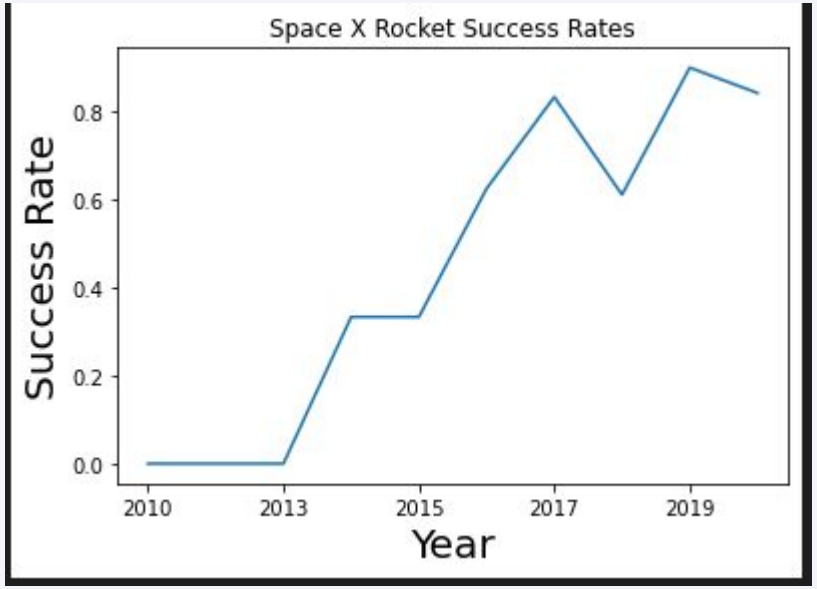
- success rate improved with time
- VLEO had many launches and most were successful
- MEO had the last launch and was successful
- The last 13 launches was successful

Payload vs. Orbit Type



- Success rate improve with the payload mass increased
- The SSO orbit had 100% successful rate
- The VLO orbit had the highest payload mass

Launch Success Yearly Trend



- The graph trend was upward
- in 2018 had the most important and unique down since 2010

All Launch Site Names

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

- This is the result from uniques launch sites from the data in table “spacextbl”

Launch Site Names Begin with 'CCA'

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

- Here we can see five samples of Cape Canaveral launches

Total Payload Mass

payload_mass_kg_total
99980

- This is the total payload mass carried by boosters from NASA according to data in table “spacextbl”

Average Payload Mass by F9 v1.1

payload_mass_kg_avg
2928

- This average payload mass is the result obtained after filtering the data by booster version ‘F9 v1.1’ and finding the average payload mass

First Successful Ground Landing Date

first_successful_landing_outcome_date
2015-12-22 00:00:00

- This is date of first successful ground landing according to data in table “spacextbl”

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- These are the boosters which have successfully landed on drone ship and had payload mass between 4000kg and 6000kg

Total Number of Successful and Failure Mission Outcomes

mission_outcome	total_number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- These are the total number of successful and failure mission outcomes according to data in table “spacextbl”

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

- These are the boosters which have carried the maximum payload mass according to data in table “spacextbl”

2015 Launch Records

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	10-01-2015
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	14-04-2015

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

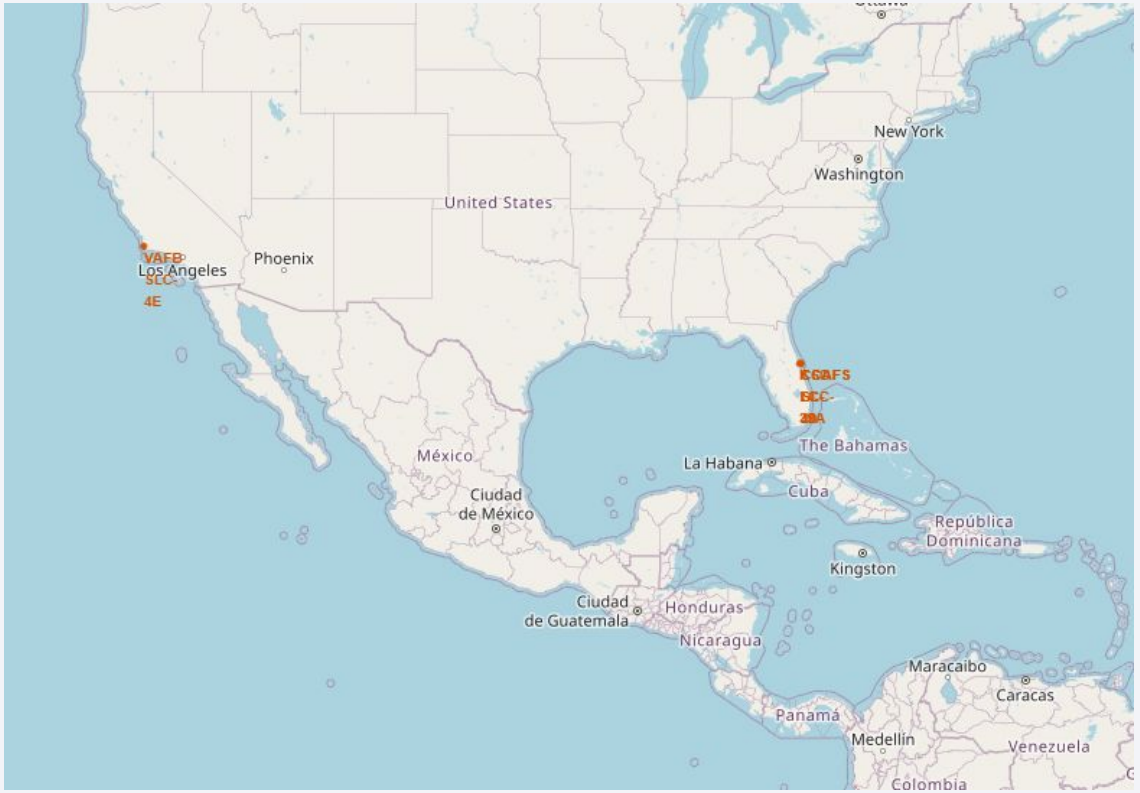
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

- The ranking of all landing outcomes between the date 2010-06-04 and 2017- 03-20

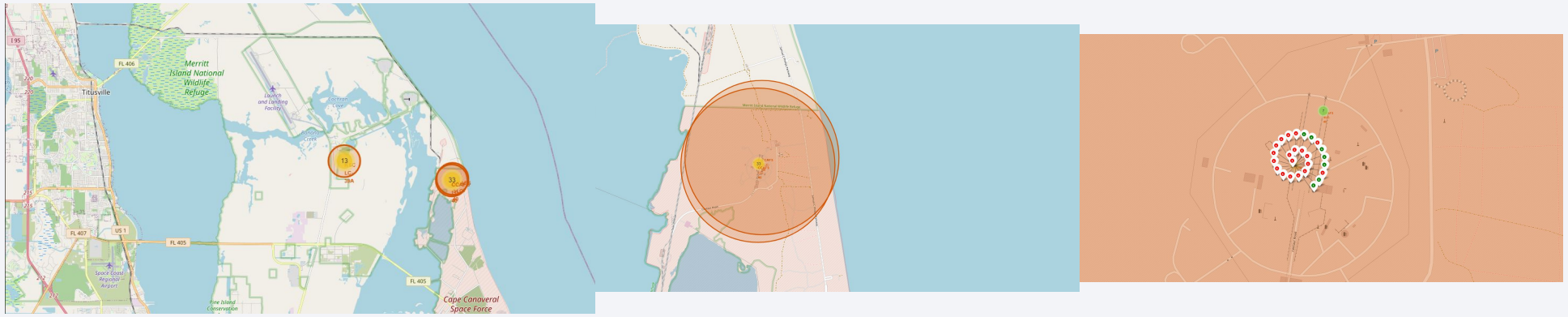

Section 3
Launch Sites
Proximities Analysis

All launch sites



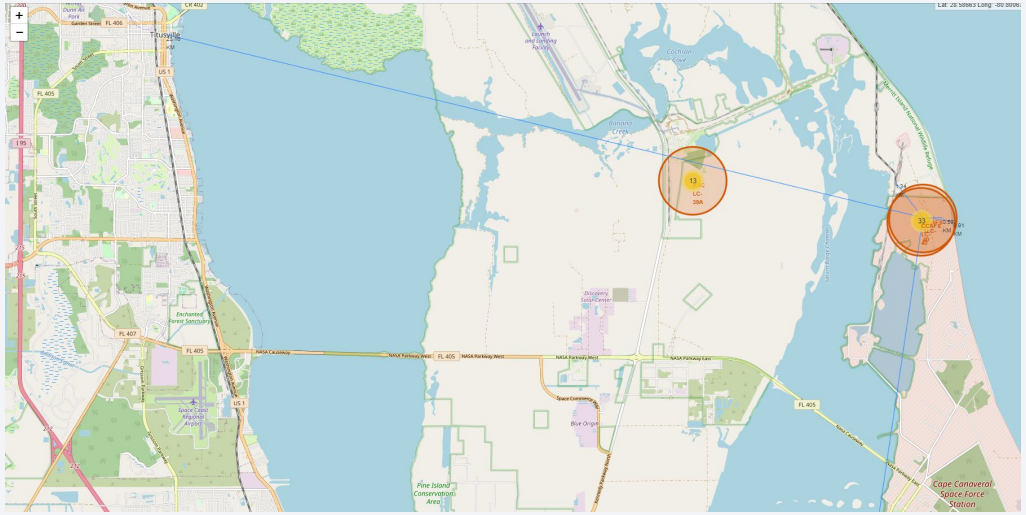
The launch sites are isolated near the sea in opposite parts of the North American continent

Color-labeled launch outcomes on the map

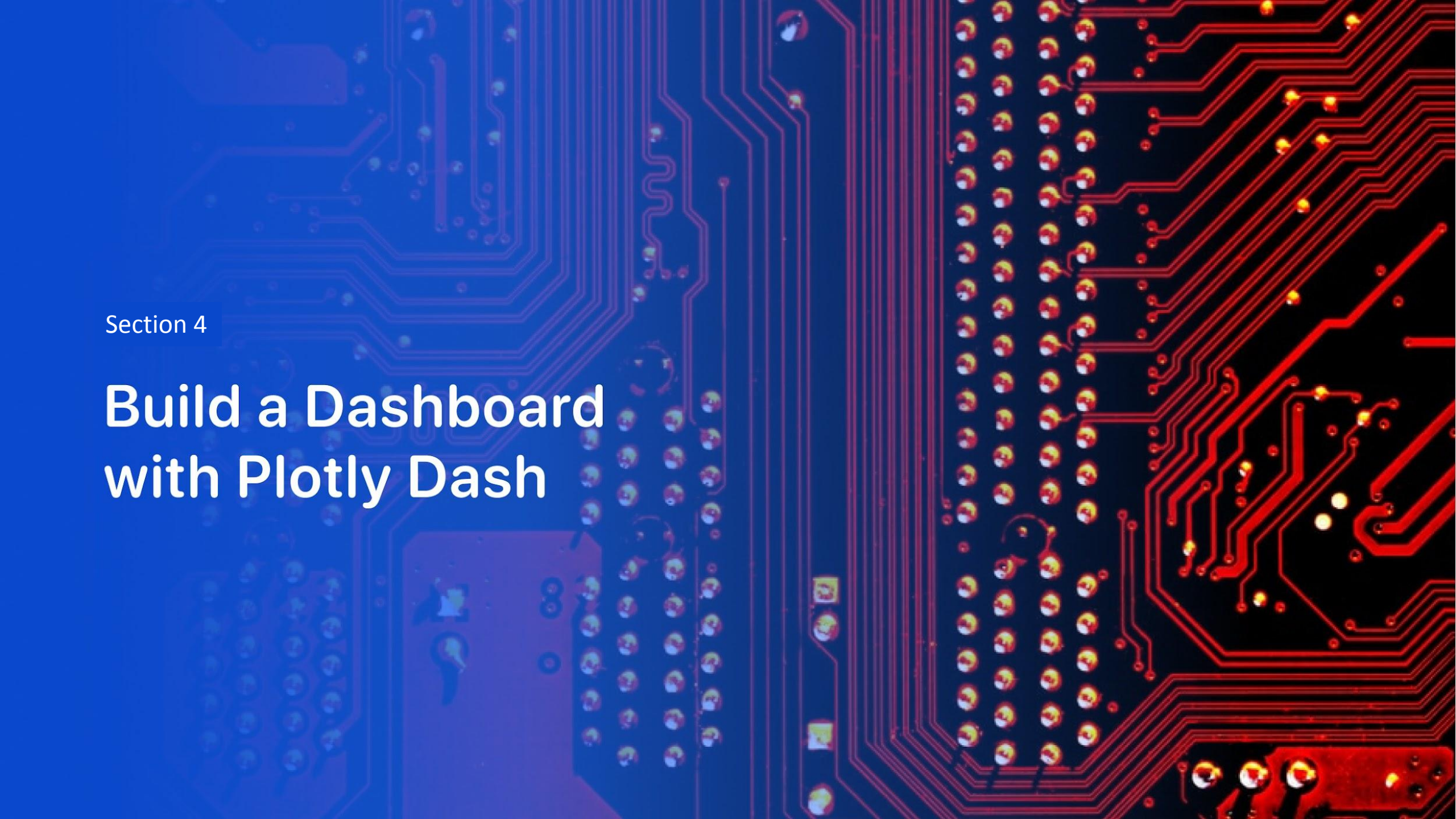


- In the KSC LC-39A launch site the green marks indicate successful launch and red marks indicate failures

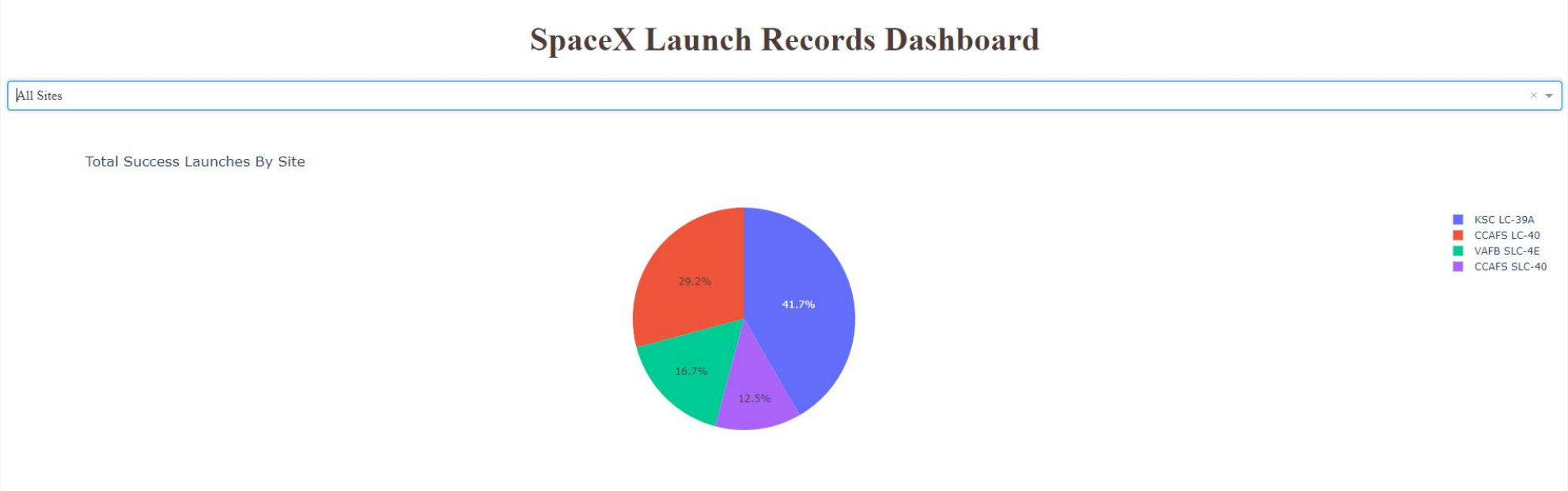
Launch Sites Proximities



- For security the launch site KSC LC-39A has a prudent distance from estructures or citys, the inhabited areas are good for launch sites

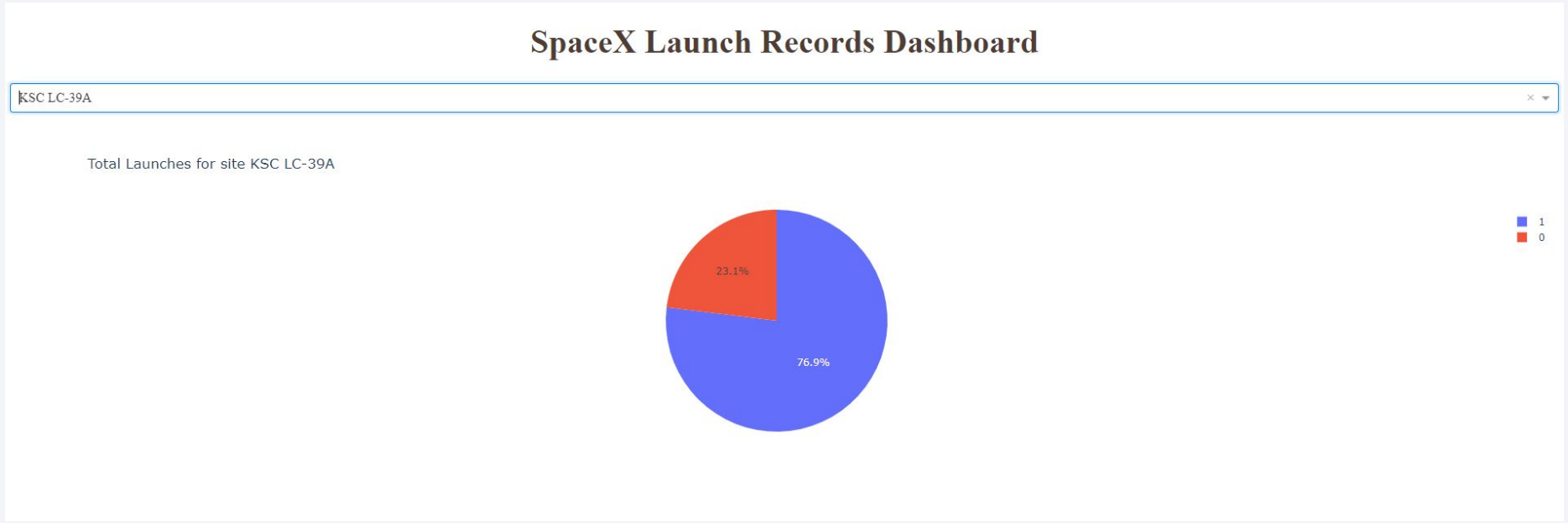

Section 4
Build a Dashboard with Plotly Dash

Total success launches



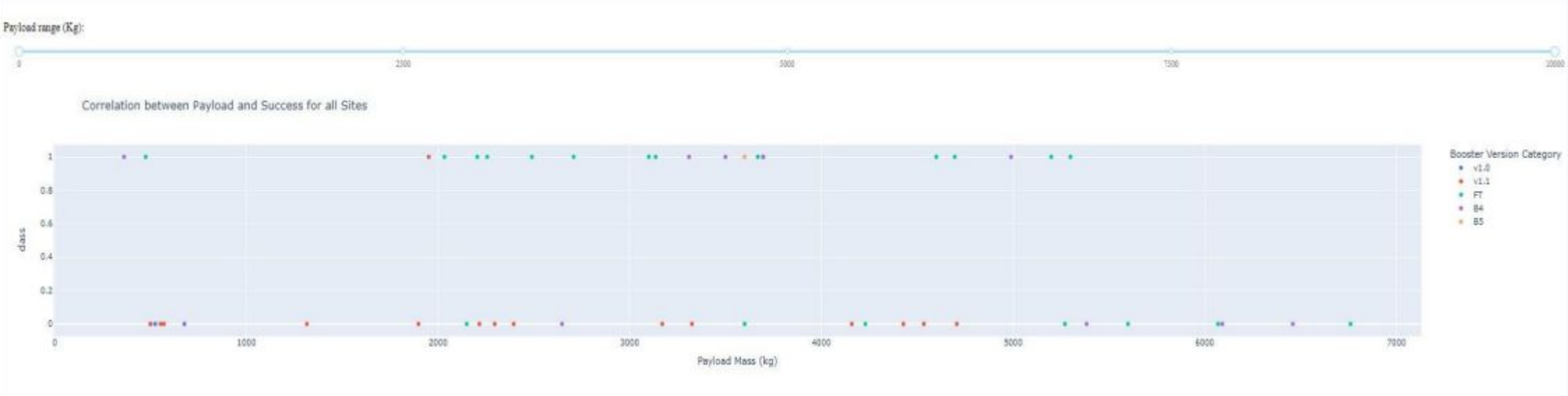
- The Pie Chart shows the distribution of successful landings across all launch sites.
- The KSLC- 39A site recorded the highest successful launches among all the launch sites.

Launch site with highest success ratio



- The KSLC- 39A launch site had the highest launch success rate with 10 successful landings (76.9%) and 3 filuries landings (23.1%)

Payload vs Launch Outcome



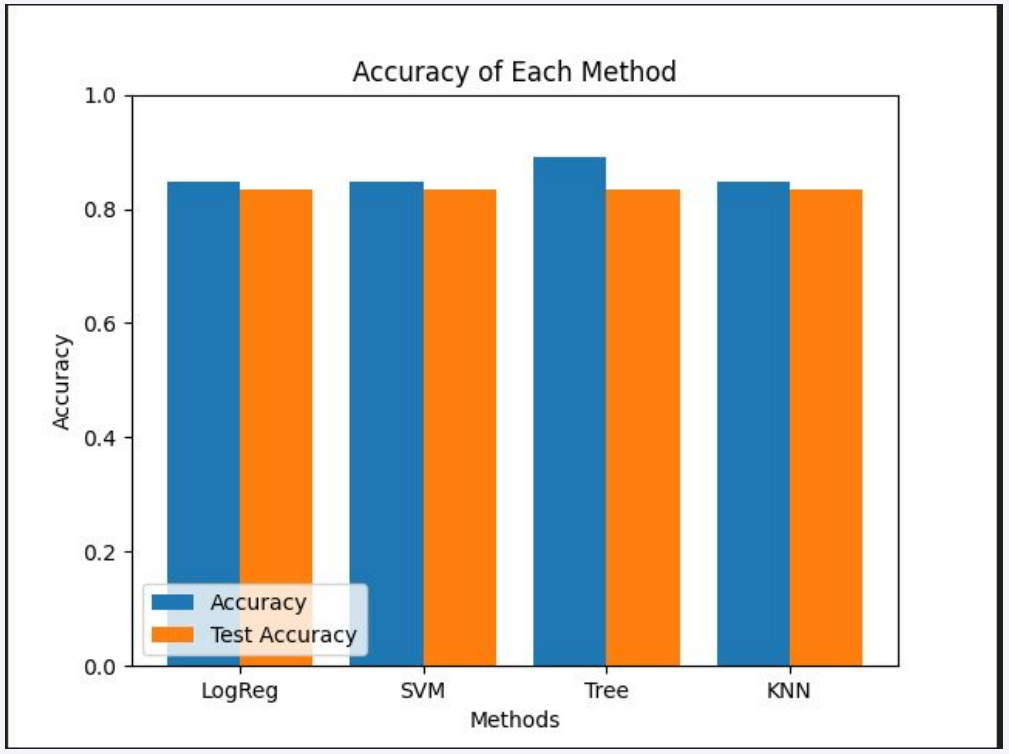
- The dashboard showed booster version FT with Payload mass under 6000kg had the major launch success rates about others

Section 5

Predictive Analysis (Classification)

Classification Accuracy

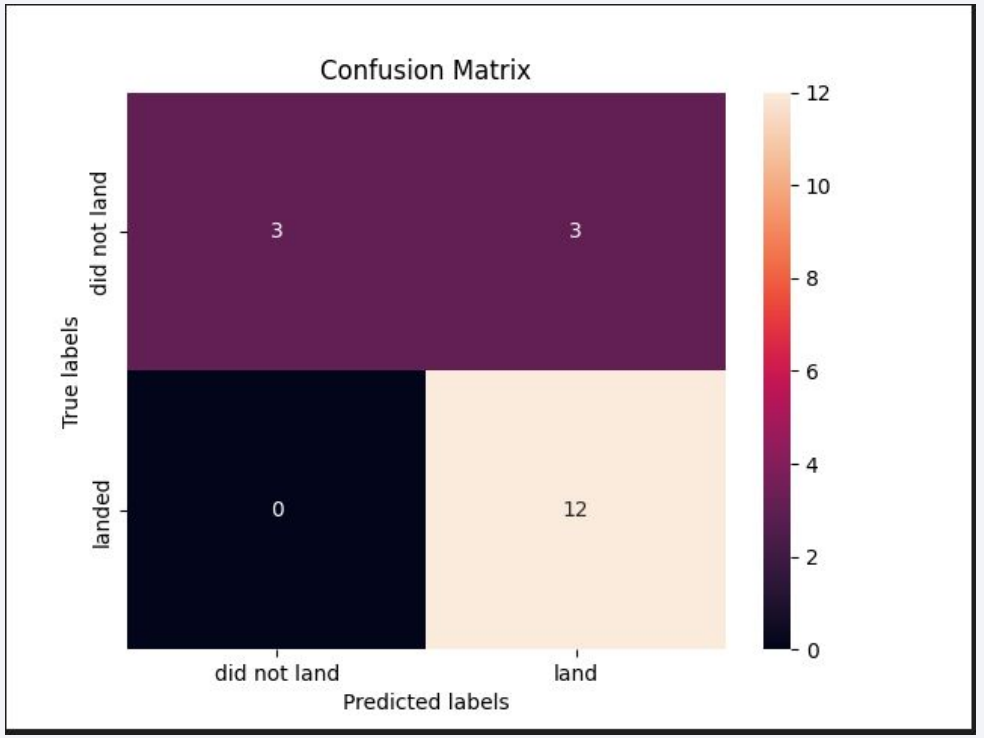
- This are different predictive models with information about landing stage successful rate, the model with best successful rate was tree model with 88.9%



Confusion Matrix

- The confusion matrix showed:
- 3 FN (False, Negatives)
- 12 TP (True, Positives)
- 3 TN (True, Negatives)

This was the result by 18 registers to predict



Conclusions

- The best model to predict the successful landing was model tree with 89%
- The best launch site was KSC-LC39A with 76.9% of launch successful
- The launches above 7000kg are less risky
- The mission are improve over time accordion with evolution of processes and rockets
- SpaceY company can use this points for increase profit and guarantee the successful mission

Appendix

- GitHub Repository : https://github.com/Crismaro19/Applied-Data-Science-Capstone-Cristian-Romero
- Coursera IBM course: https://www.coursera.org/learn/applied-data-science-capstone?specialization=ibm-data-science
- API SpaceX : https://api.spacexdata.com/v4/launches/past
- Wikipedia : https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

Thank you!