IBM Developer
SKILLS NETWORK

Data science applied to SpaceX launches

Applied Data Science Capstone

10th Courses IBM Data Science Professional Certificate and Applied Data Science Specialization

Name: Angelo Villegas

coursera

Outline

- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Introduction

- Project background and context
  - Space Exploration Technologies Corp doing business as SpaceX, is an American spacecraft manufacturer, launch service provider and satellite communications company.
  - SpaceX advertises Falcon 9 rocket launches on its website at a cost of $62 million; other suppliers cost more than $165 million each, much of the savings is due to SpaceX being able to reuse the first stage.
- Problems you want to find answers
  - Using the data we can collect on SpaceX launches, we want to predict whether the Falcon 9 first stage will land successfully.
  - Therefore, if we can determine this, we will know the cost of a launch. And thus use this information for our company and bid against SpaceX for the launch of a rocket.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Some HTML tables to Wikipedia
- The data obtained were managed and analyzed by means of
  - Data analysis (EDA) using visualization and SQL
  - Interactive visual analytics using Folium and Plotly Dash
  - Predictive analysis using classification models
    - Building, tune, evaluate classification models using *confusion matrix* and examinate de accuracy using *score function.*

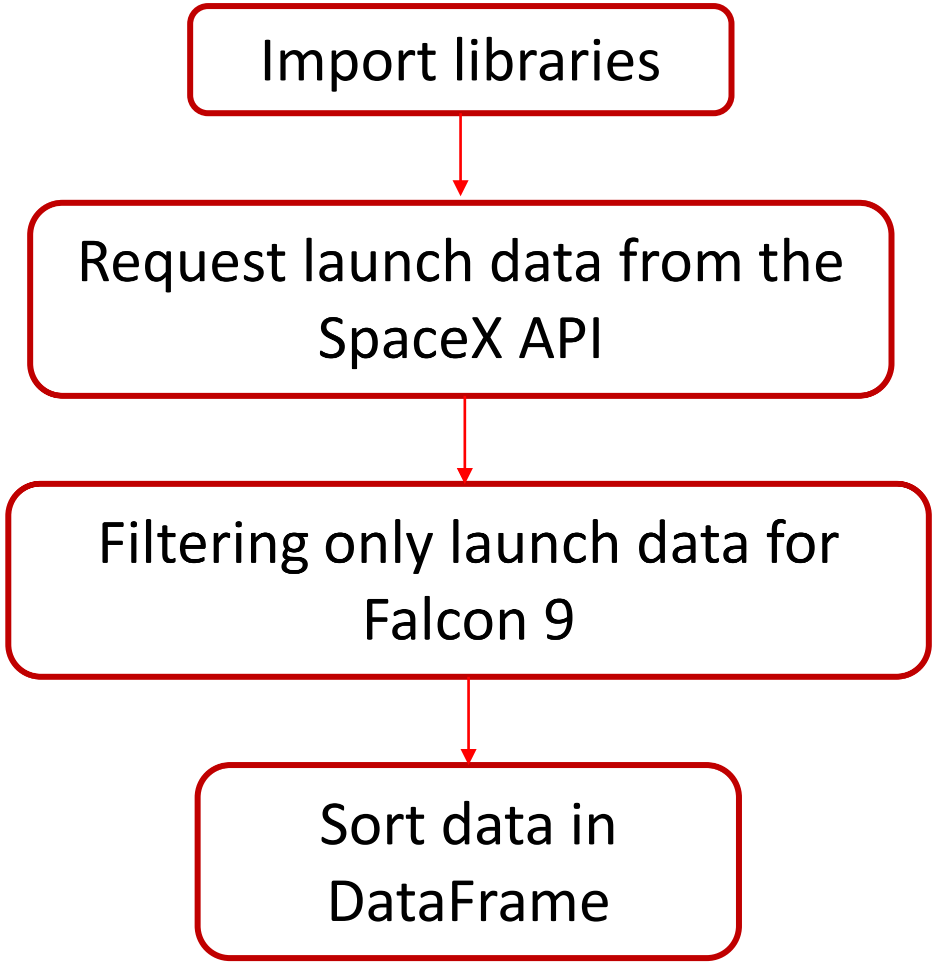
Data Collection

- Data collection with an API:
  - SpaceX REST API. We use a URL to direct us to a specific point of the API, we obtain this result through the *.json() method*, and then we visualize it in a dataframe.
- With the help of the BeautifulSoup library we extract Falcon 9 launch data from Wikipedia.

Data Collection – SpaceX API

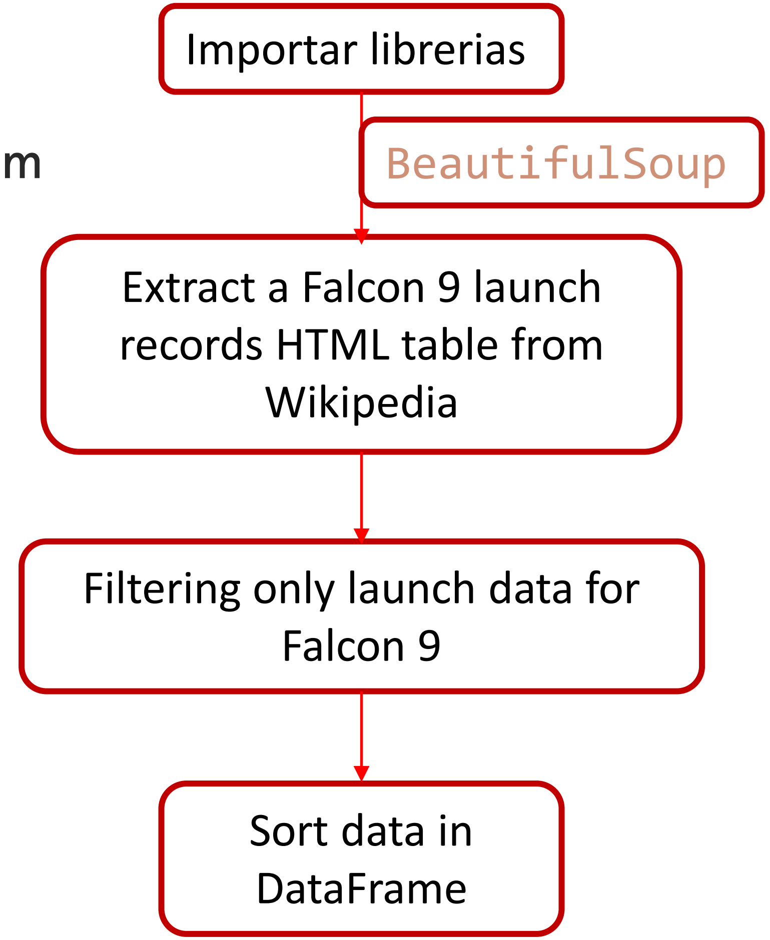
- API SpaceX – Data of Falcon 1 y Falcon 9
- Filtering only Falcon 9 's data
  - With Pandas's functions and
  - Another helpers functions

- GitHub Link of [Lab *Collecting Data to SpaceX API*]



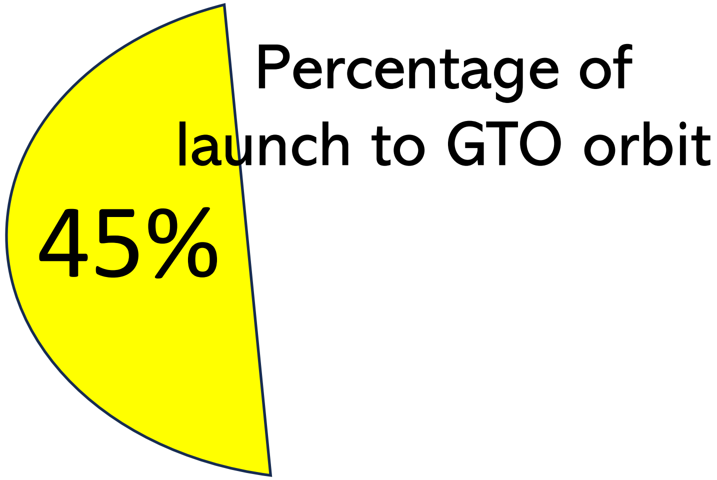
Data Collection – Scraping

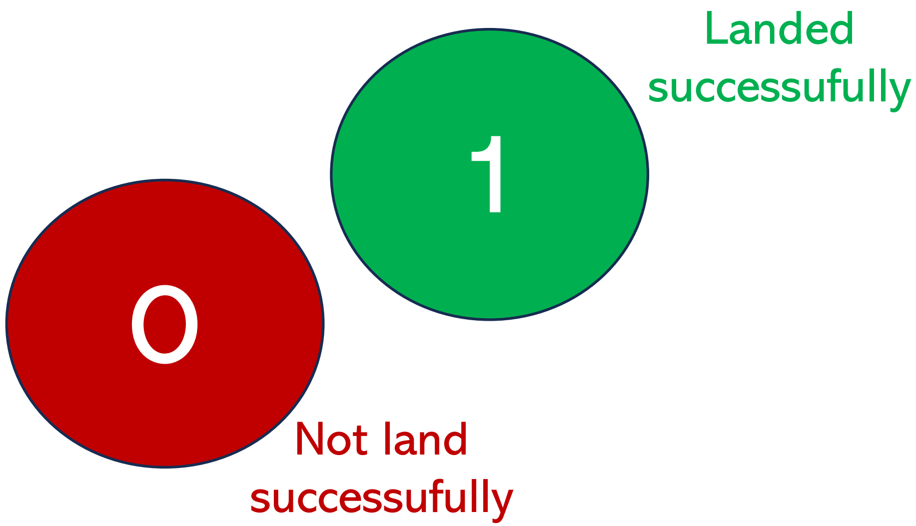
- Web scraping to collect Falcon 9 historical launch records from a ***Wikipedia.*** Link
- Filtering only Falcon 9 's data
  - With Pandas's functions and
  - Another helpers functions
- GitHub Link of Web Scraping from Wikipedia Lab



Data Wrangling

- <u>Exploratory Data Analysis (EDA)</u> to find some patterns in the data and determine what would be the label for ***training supervised models.***
- Using the SpaceX dataset
  - We examine the **Orbit** column
    - Geosynchronous orbit (GTO), was the most crowded at the launches
  - We examined the **Outcome** column, representing each release with a ratings variable:
    - **0**: not land successfully
    - **1**: landed successfully.
- GitHub Link of [Data Wrangling's Lab]()





EDA with Data Visualization

- The visualization of the data was done by observing how the different variables behaved with respect to whether the first stage had a successful landing or not.
- We perform scatter charting to see the relationship between two variables, and decipher if there is any relevant behavior in the case.
- We also visualize a bar chart to see how the target orbit of the launch influences its success.
- We then visualize at the trend in line graphs of the launches over the last few years.

- GitHub Link of Data Visualization Lab

EDA with SQL

- Summarize the SQL queries performed:

1. `SELECT Launch_site, COUNT(Launch_site) AS COUNT FROM SPACEXTABLE GROUP BY Launch_site`
2. `SELECT * FROM SPACEXTABLE WHERE Launch_site LIKE 'CCA%' LIMIT 5`
3. `SELECT SUM(PAYLOAD_MASS__KG_) AS 'Total Pay Mass by NASA' FROM SPACEXTABLE WHERE Customer LIKE '%(CRS)'`
4. `SELECT AVG(PAYLOAD_MASS__KG_) AS 'AVG Pay Mass in KG' FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1%'`
5. `SELECT * FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)' ORDER BY Date LIMIT 1`
6. `SELECT * FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000`
7. `SELECT Mission_Outcome, COUNT(Mission_Outcome) AS COUNT FROM SPACEXTABLE GROUP BY Mission_Outcome`
8. `SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)`
9. `SELECT substr(Date, 6, 2) AS Months, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE Date LIKE '2015%' AND Landing_Outcome LIKE '%(drone ship)’`
10. `SELECT Date, Landing_Outcome FROM SPACEXTABLE WHERE Landing_Outcome IN ('Failure (drone ship)','Success (ground pad)') AND Date BETWEEN '2010-06-04' AND '2017-03-20'ORDER BY Date DESC`

- GitHub Link of [EDA with SQL Lab]()

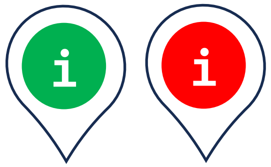
Build an Interactive Map with Folium

Folium library to visualize launch site locations on a map.

- Map objects:

Launch site locations

Cluster to view the exact coordinates of the launches at each of the Launch Site Locations



Successful and unsuccessful launches

Distance to *coastline* marker

Distance to *nearest city* marker

- GitHub Link of [Lunch Sites Analysis with Folium Lab]

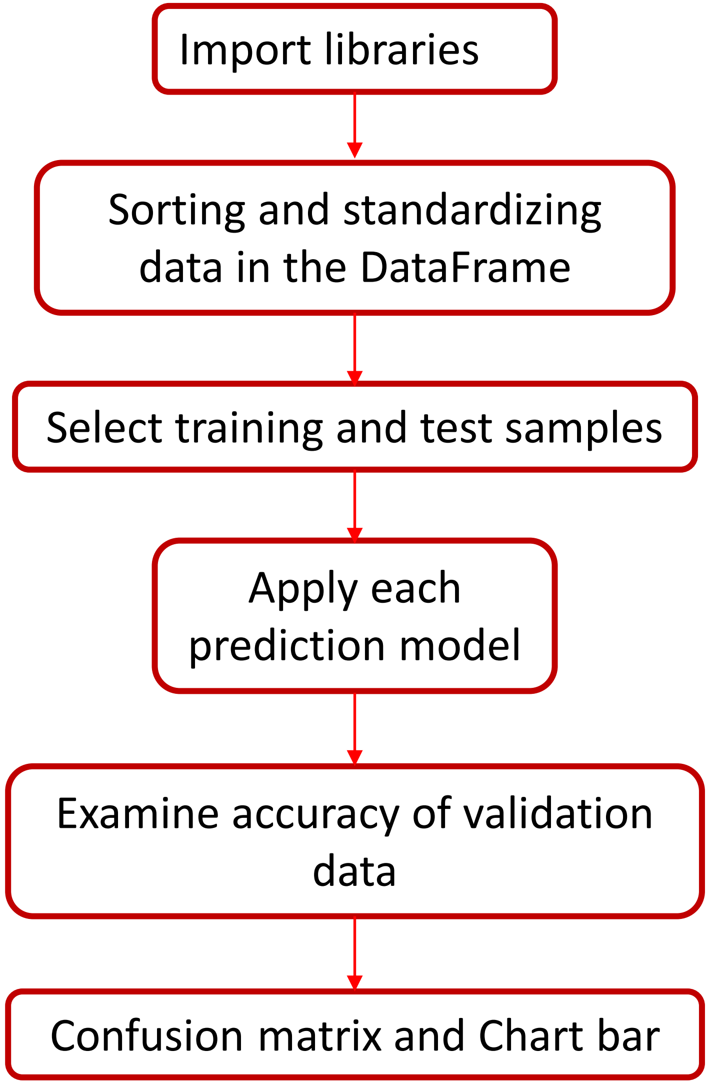
Build a Dashboard with Plotly Dash

Predictive Analysis (Classification)

- **4 modelos** de ML para predecir el exito o el fracaso del primer aterrizaje dados los datos de los laboratorios precedentes.

  - Logistic Regression
  - SVM
  - Classification Trees
  - KNN

- GitHub Link of [Machine Learning Prediction Lab]



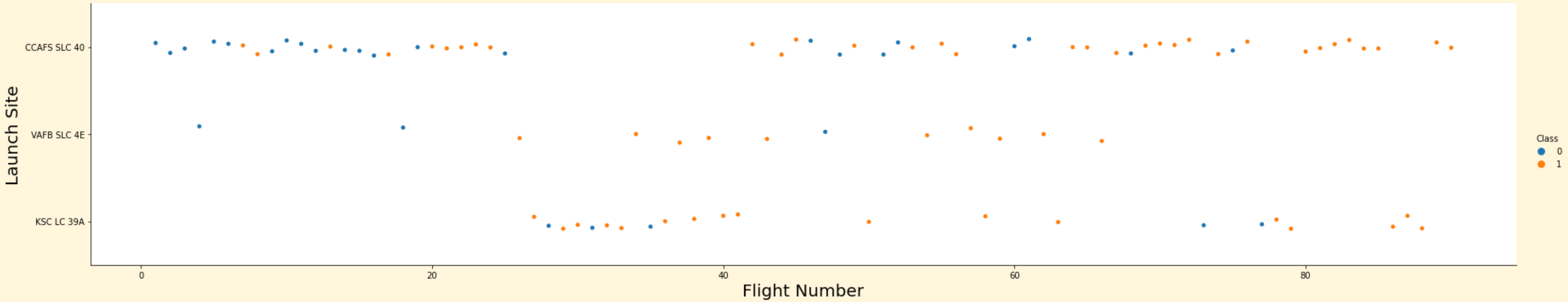
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

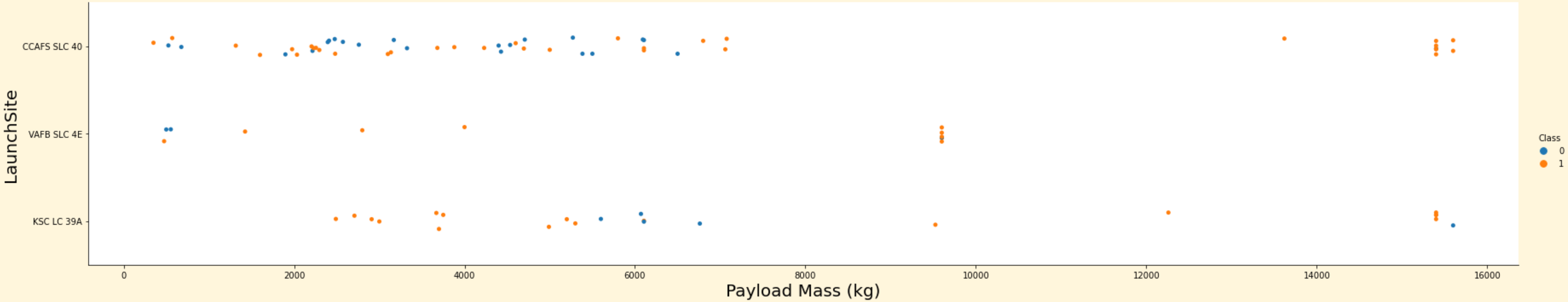
Insights drawn from EDA

Flight Number vs. Launch Site



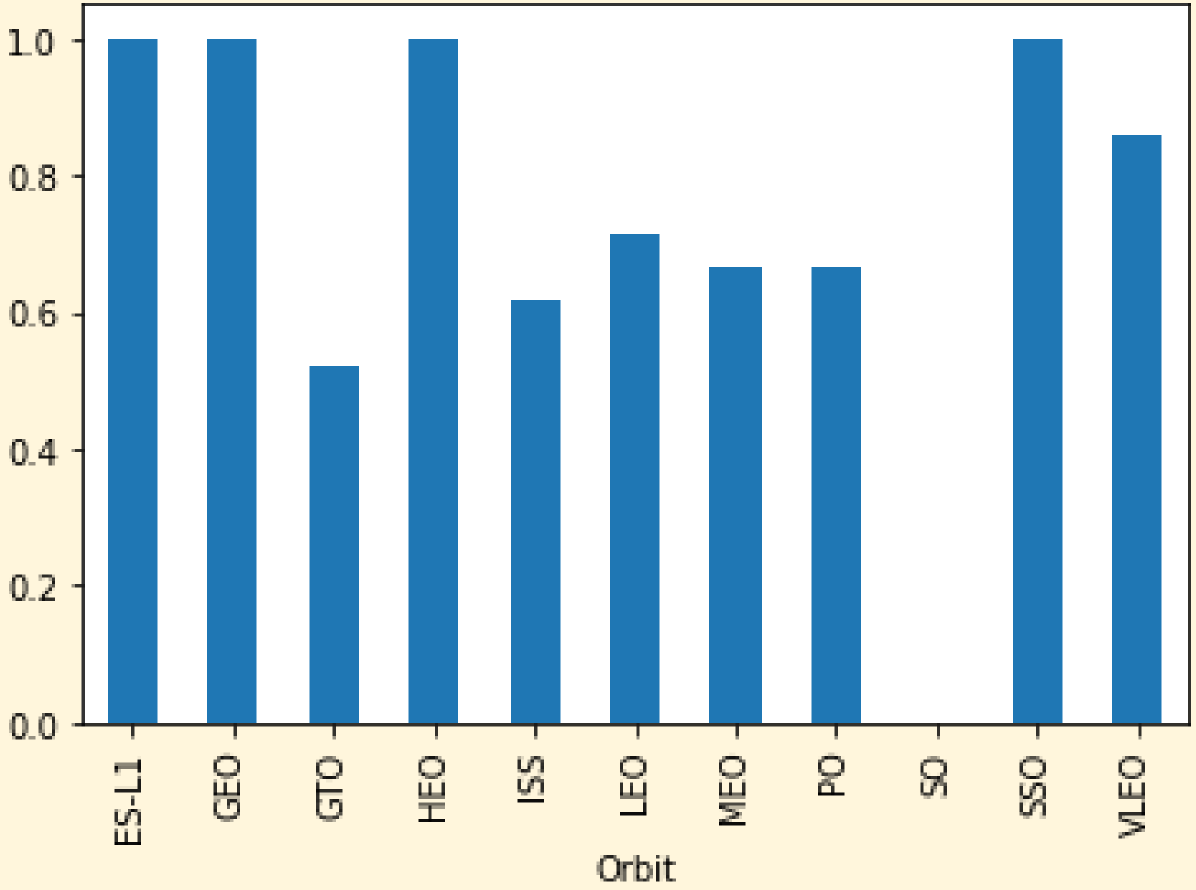
- We see that as the number of flights increases, it is more likely that the first stage will land successfully for all 3 launch sites.
- In addition, the largest number of launches were made in CCAFC SLC 40.

Payload vs. Launch Site



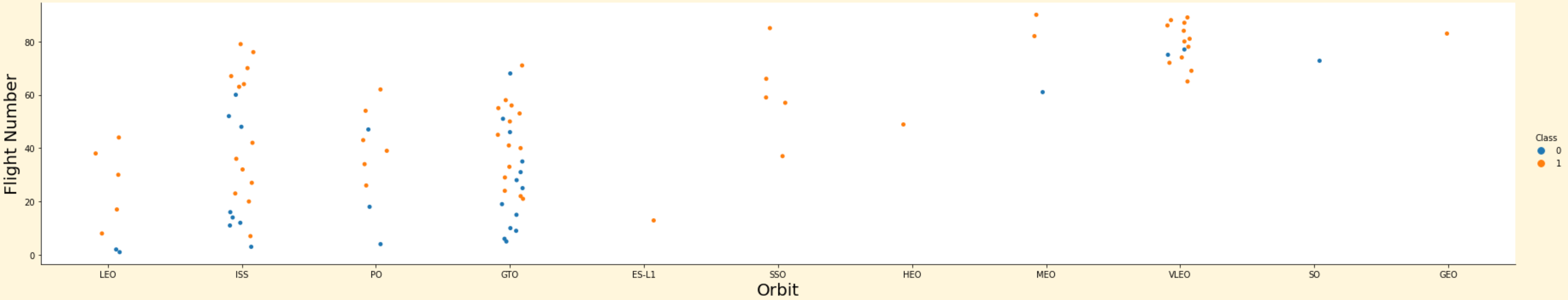
- No heavy payload rockets (greater than 10,000) are launched at the VAFB-SLC launch site. This results in a higher number of successful launches compared to unsuccessful launches at that site.
- We also note that the higher the payload, the more failed launches are almost nil.

Success Rate vs. Orbit Type



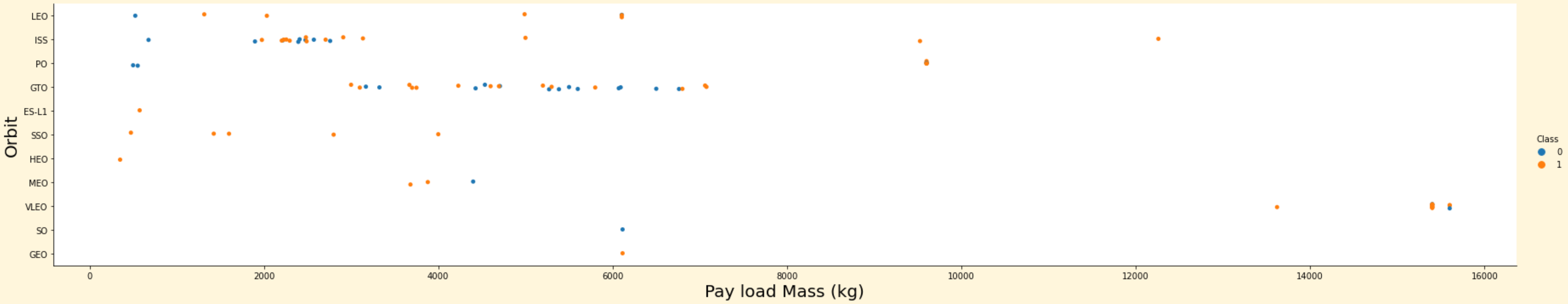
- The most successful launches are those to the ES-L1, GEO, HEO and SSO orbits.

Flight Number vs. Orbit Type



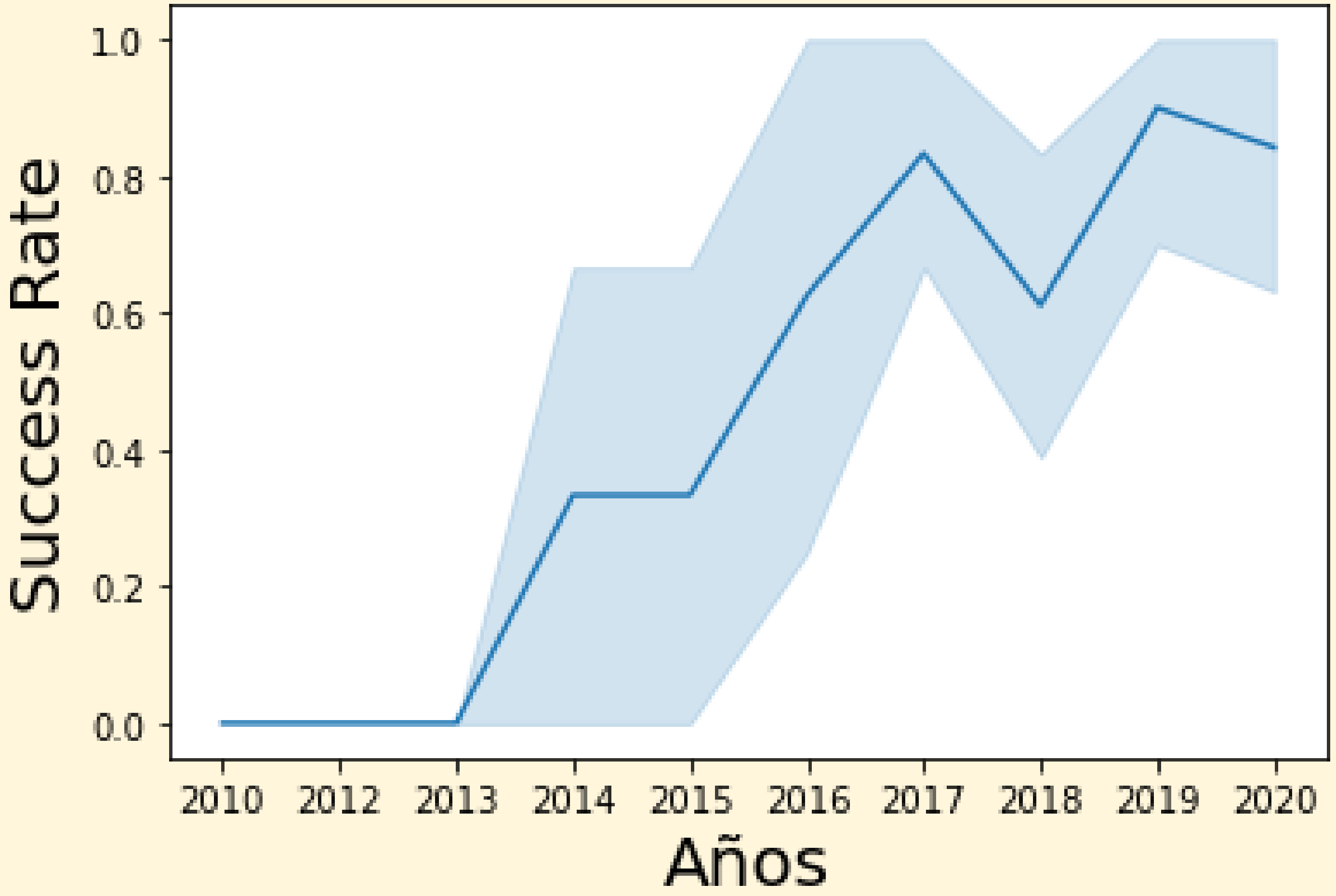
- For LEO orbit, the success of the launches seems to be related to the number of flights.
- The 100% efficiency of the launch to GEO orbit is given that there is only one launch to that orbit.
- Therefore, there seems to be no relation between the success of the launch and the number of flights to the ISS and GTO orbits.

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



- We can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

Launch_Site	COUNT
CCAFS LC-40	26
CCAFS SLC-40	34
KSC LC-39A	25
VAFB SLC-4E	16

- The table shows the name of each launch site with the number of launches made at each site.
- We see that the highest number of launches was in CCAFS SLC-40.

Launch Site Names Begin with 'CCA'

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

- The table shows 5 records where the launch site begins with 'CCA'.
- We can observe that all of them have had successful mission results, but with failed landings in 2 of them.
- In addition, all launches were directed to LEO orbit.

Total Payload Mass

TOTAL PAY MASS BY NASA : 45569 Kg

Average Payload Mass by F9 v1.1

Average Payload Mass : 2534.66 Kg

First Successful Ground Landing Date

- *Date :* 2015/12/22
- *Time:* 1:29:00
- *Booster Version:* F9 FT B1019
- *Launch Site:* CCAFS LC-40
- *Payload:* OG2 Mission 2 11 Orbcomm-OG2 satellites
- *Payload mass:* 2034 kg
- *Orbit:* Low Earth Orbit (LEO)
- *Customer:* Orbcomm

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- All launches were made into GTO orbit from the CCAFS LC-40 and KSC LC-39A sites.

Total Number of Successful and Failure Mission Outcomes

- *Failure in flight :* 1
- *Success :* 100

Boosters Carried Maximum Payload

- *F9 B5 B1048.4*
- *F9 B5 B1049.4*
- *F9 B5 B1051.3*
- *F9 B5 B1056.4*
- *F9 B5 B1048.5*
- *F9 B5 B1051.4*
- *F9 B5 B1049.5*
- *F9 B5 B1060.2*
- *F9 B5 B1058.3*
- *F9 B5 B1051.6*
- *F9 B5 B1060.3*
- *F9 B5 B1049.7*

Maximum Payload Mass : 15600 Kg

2015 Launch Records

- *Failure (drone ship):* F9 v1.1 B1012 and
  F9 v1.1 B1015
- *Precluded (drone ship):* F9 v1.1 B1018

Launch Site
CCAFS LC-40

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

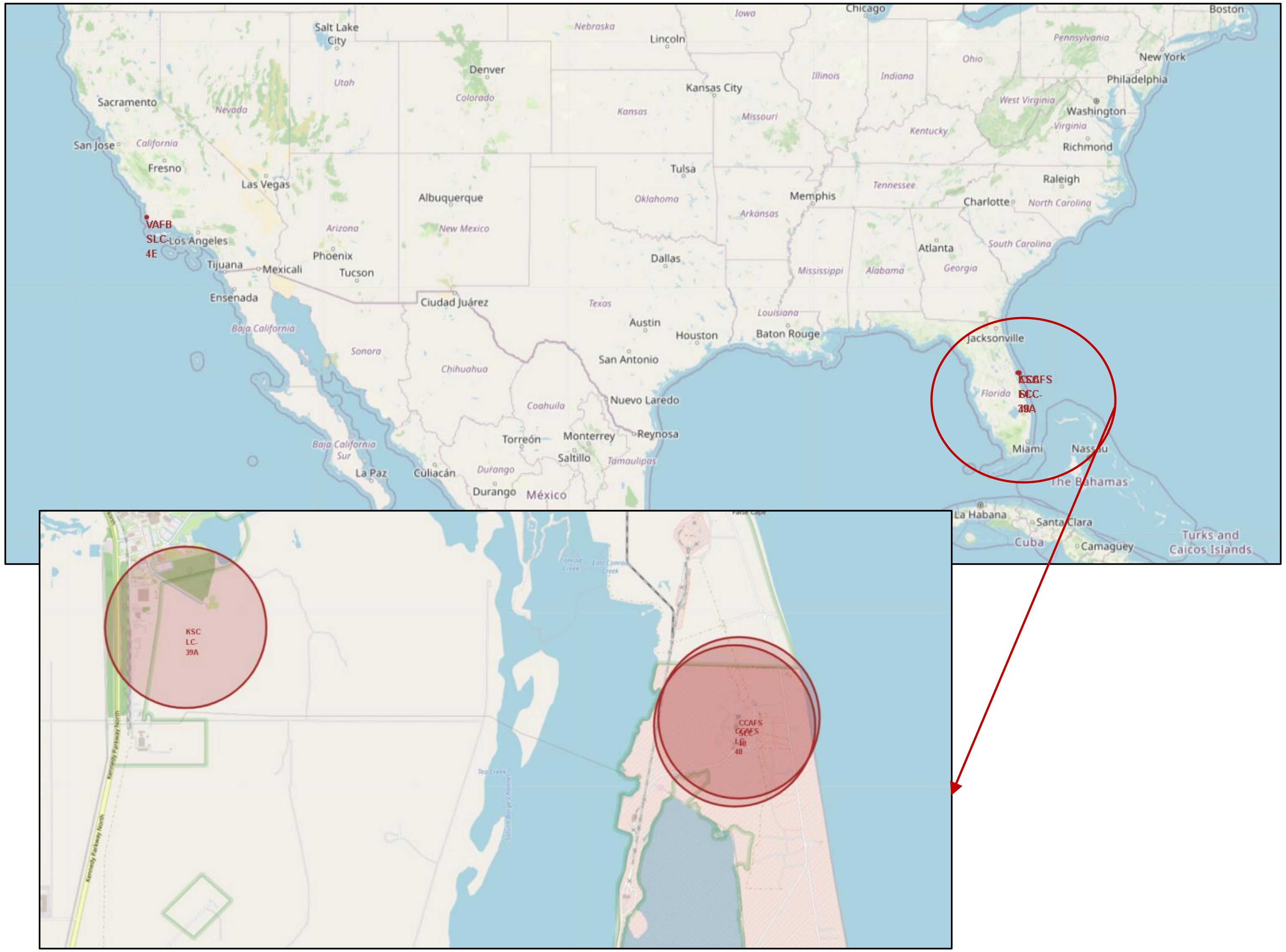
Date	Landing_Outcome
2017-03-06	Success (ground pad)
2017-02-19	Success (ground pad)
2017-01-05	Success (ground pad)
2016-07-18	Success (ground pad)
2016-06-15	Failure (drone ship)
2016-04-03	Failure (drone ship)
2016-01-17	Failure (drone ship)
2015-12-22	Success (ground pad)
2015-10-01	Failure (drone ship)
2015-04-14	Failure (drone ship)

- The table shows successful landings for ground platforms and unsuccessful landings for drone ship.

Section 3

Launch Sites Proximities Analysis

Lunch Sites visualization map

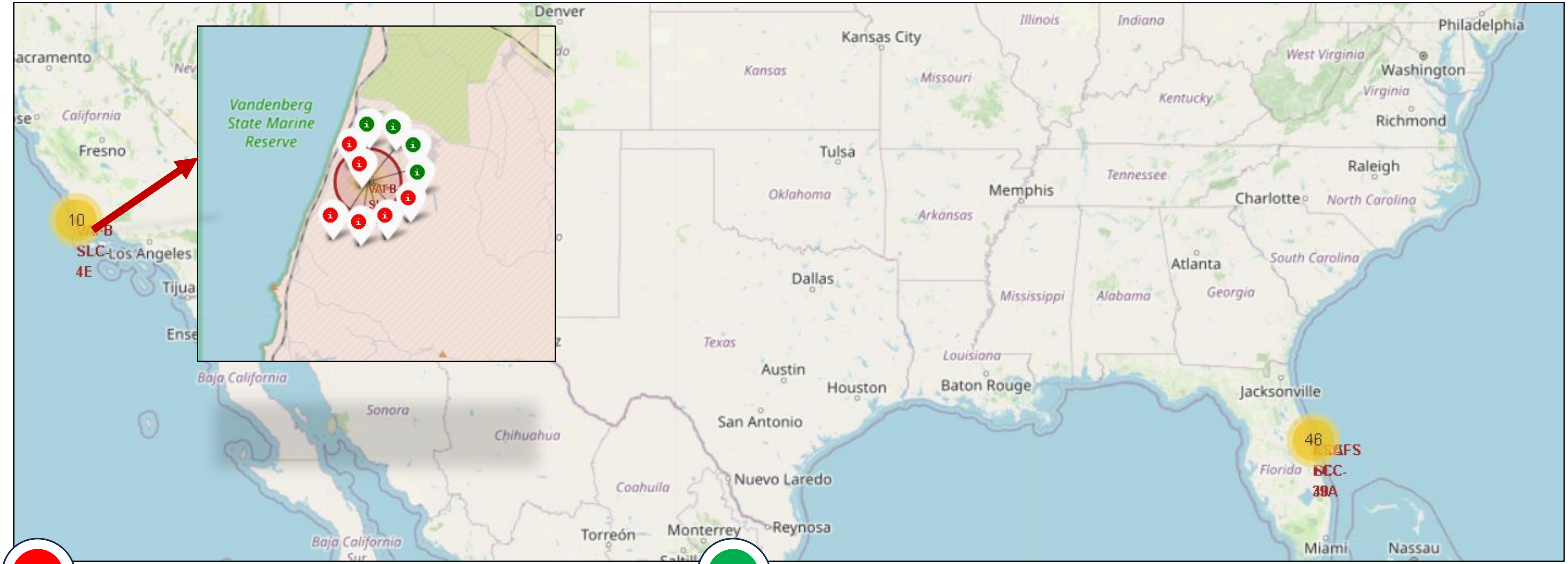


We note two important points:

- Launch sites are close to the planet's equator.
- All of them are near the coast of an ocean.

Success/failed lunches visualization map

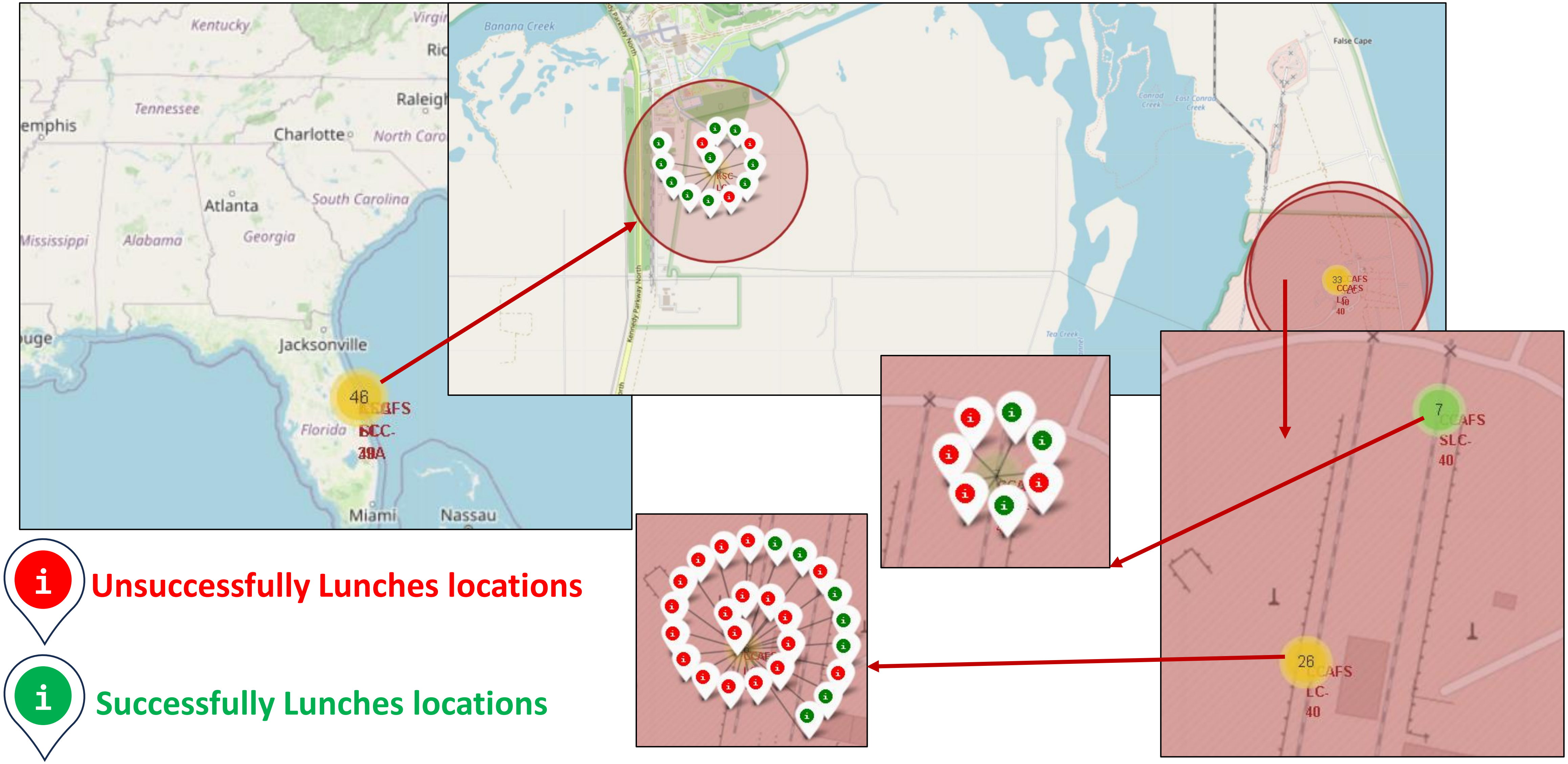
Cluster with original Zoom of the number of launches for the West Coast and East Coast of the country.



Unsuccessfully Lunches locations

Successfully Lunches locations

Success/failed lunches visualization map



Unsuccessfully Lunches locations

Successfully Lunches locations

Success/failed lunches visualization map

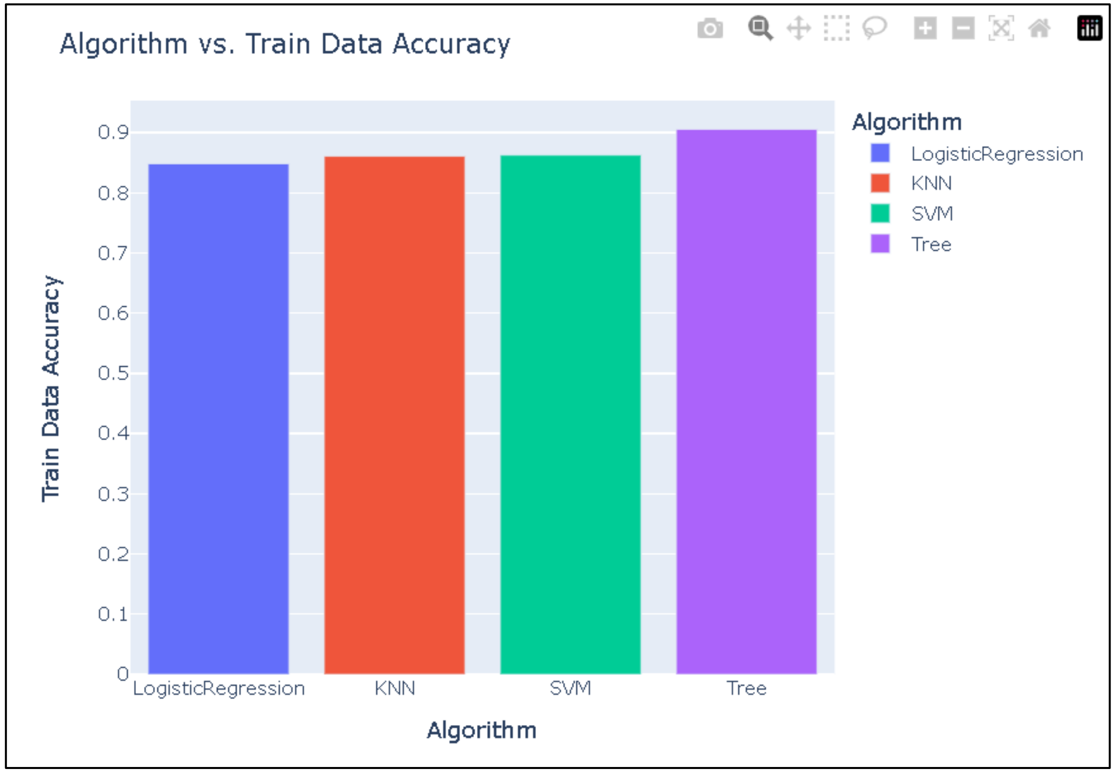
Some line traces for the calculation of distances to the costline and highway vicinity for the **CCAFS LC-40** launch site.

Distances for each launch site

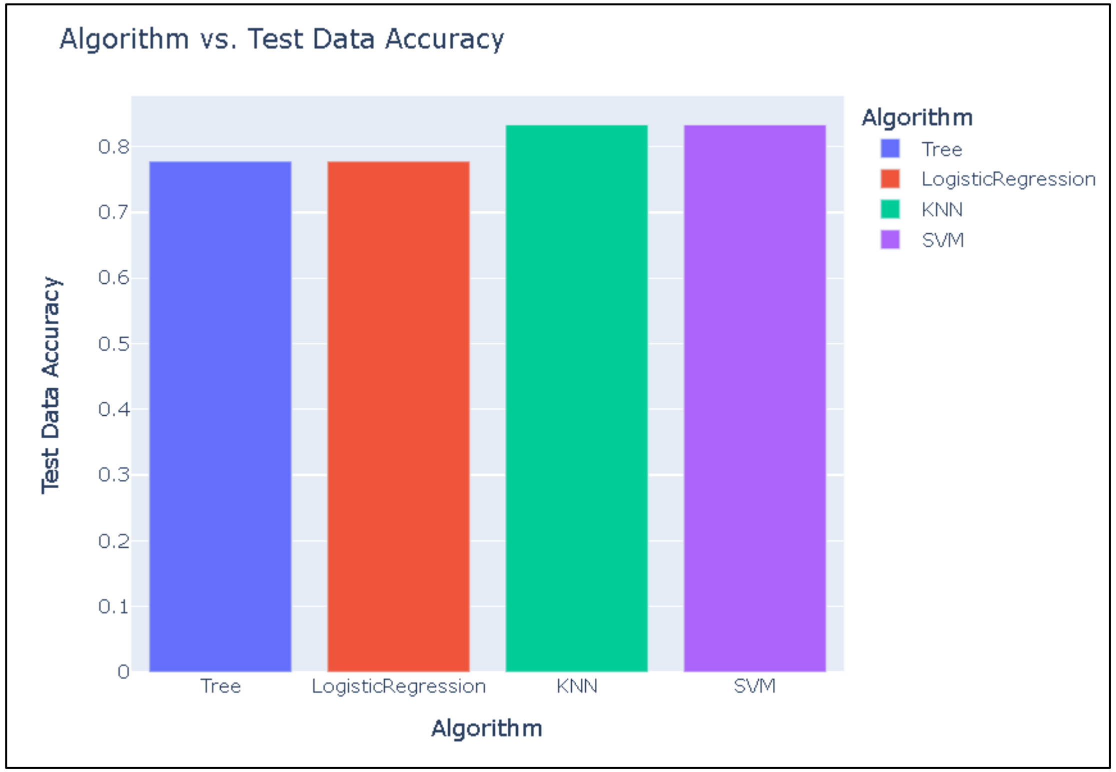
	Lunch_Site	Coastline [km]	Highways[km]	Rilways[km]	Nearest_City[km]
0	CCAFS LC-40	0.863105	0.589100	1.281903	62.035694
1	CCAFS SLC-40	0.863105	0.589100	1.281903	62.035694
2	KSC_LC 39A	7.319666	0.833236	0.719326	53.161906
3	VAFB SLC-4E	1.360011	6.155040	1.266939	14.030300

- Launch Sites are in close proximity to coast.
- Launch Sites are also close to Major Highways and Railway for logistic purposes.
- Launch sites are far from dense human habitats like cities.

Classification Accuracy – Score values



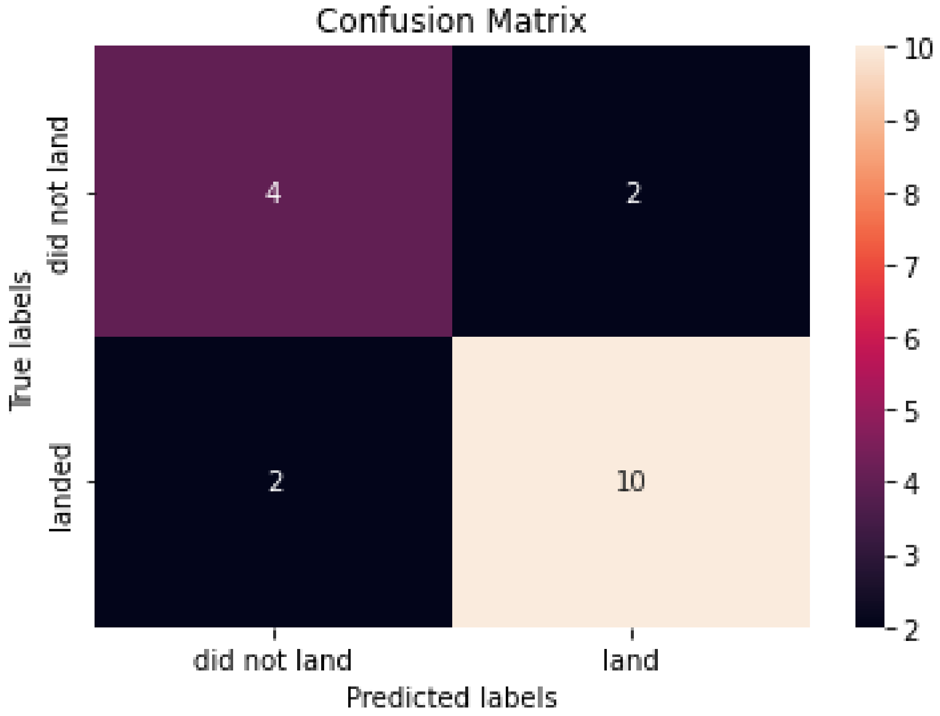
For the training data the model that best fits the data is ***TREE.***



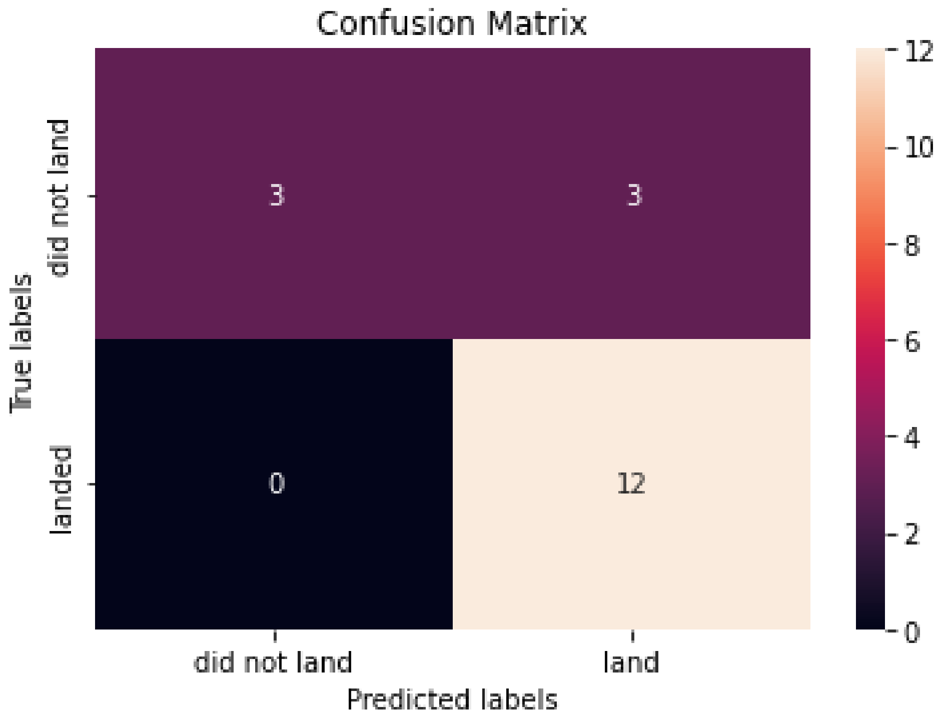
For the testing data the model that best fits the data is ***KNN.***

Classification Accuracy – Confusion Matrix

Tree Model



KNN Model



Conclusions

- The data could be obtained effectively using SpaceX API and web Scraping.
- Using the launch classification tool, we were able to distinguish between successful and unsuccessful launches.
- In addition, we note that the busiest Earth orbits are GTO, ISS, VLEO. Where the most successful missions were drone launches
- The best launch site is the CCAAFS LC -40
- We see that all launch sites are far from urban sites, and with greater proximity to the ocean and to highways and railways.