IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hasan Soufan
09/12/2022

Outline

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

Executive Summary

- SpaceX Falcon 9 First Stage Landing Prediction.
- SpaceY want to reduce its spending by studding SpaceX data and figure out how they can return the first stage of their rocket safely and try to predict the outcome of this process with different situation.
- Starting by collecting data from their apis and from scrapping Wikipedia, then doing a data wrangling find patterns in the data like (Launches Sites, Orbits and Landing Outcomes).
- Using Exploratory Data Analysis (EDA) and interactive visual analytics find relations and information between the standards and the Outcomes like (Flight Number, Payload and Orbit).
- Performing a predictive analysis then building a classification models and evaluate them, and choice the best one based on metrics like Accuracy and confusion matrix, which gives 83% of certainty when its prediction the Outcomes.
- The results are SpaceY can figure out what standards that has an effect on the landing process and predict whether their Rocket first stage will landing safety or not with 83% of certainty.

Introduction

- SpaceX (a sciences space company) spend less than others on their space flights because they try to return and land the first stage of their rockets safely, but this process doesn’t work every time, it’s depend on many features, like (Launch Site, Destination, the Rocket Version … etc.), this project is to figure out how SpaceY can reduce its spending based on studding SpaceX data.
- Depending on the previous standards we want to study SpaceX data trying to figure out:
  - What and how features effect the landing process.
  - build a prediction mechanism to find wither the first stage will land successfully or not based on different situation.

Section 1

Methodology

Methodology

Executive Summary

- **Data collection**, Data was collected using Spacex Apies and web scraping.
- **data wrangling** found patters in data like Launches Sites, Orbits and Landing Outcomes.
- **Exploratory data analysis (EDA)** using visualization and SQL show the features that have a relation with the Landing Outcomes like (Flight Number, Payload and Orbit).
- **interactive visual analytics**: using Folium map and Plotly Dash appear the distribution of the Launches Sites and the Landing Outcomes for each site, their percent of success and failure, the relation with the payload and the booster.
- **Performing predictive analysis**: by building a classification models like Logistic Regression, Support Vector Machine and Others models, evaluated by Confusion Matrix and Accuracy.

Data Collection

- data sets were collected from Spacexs apis and Wikipedia.
- At first downloading data from the sources.
- Next extracting data that relate to the issues.
- Then save the data in relational form.

Data Collection – SpaceX API

- Data Collected by Spacex APIs, with just Falcon 9* Booster Version.
- [GitHub URL for SpaceX API calls file.]

Flight Number	Date	BoosterVersion	Payload Mass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1.0	False	False	False	None	NaN	0	Merlin2C	-80.577366
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1.0	False	False	False	None	NaN	0	Merlin3C	-80.577366
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1.0	False	False	False	None	1.0	0	B0003	-80.577366
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1.0	False	False	False	None	1.0	0	B0005	-120.610829
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1.0	False	False	False	None	1.0	0	B0007	-80.577366

* Falcon 9 is a space rocket for tesla

Data Collection - Scraping

- Data Collected by scrapping Wikipedia Falcon 9* historical launch records.
- GitHub URL for web scraping notebook file

Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Launch Site	Version Booster	Booster landing	Date	Time
0	1	NaN	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	CCAFS	F9 v1.0B0003.1	Failure	4 June 2010
1	1	NaN	Dragon	0	LEO	NASA	Success	CCAFS	F9 v1.0B0003.1	Failure	4 June 2010
2	2	NaN	Dragon	525 kg	LEO	NASA	Success	CCAFS	F9 v1.0B0004.1	No attempt\n	8 December 2010
3	3	NaN	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	CCAFS	F9 v1.0B0005.1	No attempt	22 May 2012
4	4	NaN	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	CCAFS	F9 v1.0B0006.1	No attempt\n	8 October 2012

* Falcon 9 is a space rocket for tesla

Data Wrangling

- Dealing with Null values.
- Figuring out the following:
  - Launches in each Site.
  - Launches to each Orbit*.
  - Outcomes per Orbit type.
- Create label Outcome column with 1 for success and 0 for failure by encode the (Outcome) values.

- GitHub URL for Data Wrangling file

* Orbit is the curved path of a celestial object or spacecraft

EDA with Data Visualization

- **Scatter Plots**: figuring out the relationships between the following features and their effect on the Launch Outcomes
  - (Flight Number) vs (Payload).
  - (Flight Number) vs (Launch Site).
  - (Payload) vs (Launch Site).
  - (Flight Number) vs (Orbit)
  - (Payload) vs (Orbit).
- **Bar Chart**: determine the relationship between success rate of each (Orbit) type.
- **Line Chart**: visualize the launch success yearly trend
- GitHub URL for EDA visualization file.

EDA with SQL

- the names of the unique launch sites in the space mission.
- the first 5 records where the Launch Site begin with ‘CCA’.
- the total Payload Carried by Boosters Launched by NASA.
- average Payload carried by Booster Version F9 v1.1.
- the Date of the first success landing Outcome achieved.
- the Boosters that have a Payload between 4000 – 6000 and Landed successfully in drone ship.
- the total number of successful and failure mission Outcome.
- booster versions which have carried the maximum Payload.
- failed Landing outcome in drone ship, their booster version and Launch Site in 2015
- the total Landing Outcome between 2010 – 2017 in descending.
- GitHub URL for EDA with SQL file.

Build an Interactive Map with Folium

- **Cluster Markers** were created for each (Launch Site) to cluster its information.
- **Circles** on each Site to show its location.
- **Markers** for each success/failed Launches with green/red colors to display each Site Launches success rate.
- **Lines** between Launch Sites and railways, highways, cities and coastlines, to study its location.

- GitHub URL for Folium map file.

Build a Dashboard with Plotly Dash

- Plots, graphs and interactions added to a dashboard
  - **Dropdown** list for enable (Launch Site) selection.
  - **Slider** select the (Payload) range.
  - **Pie Chart** show the total successful launch for all sites and if specific site was selected view the sites success vs failed launches.
  - **Scatter chart** show the correlation between the (Payload) and the launch success with determining the (Booster Version).

- GitHub URL for Plotly Dash file.

Predictive Analysis (Classification)

- Standardize and Normalize the Data to keep it in the same range because the models used minkowski distance like Euclidean and Manhattan .
- Split the data set into train and test data to test the models and have a certainty with samples out of the range.
- Using Grid Search and Cross-Validation to test each model with multiple situations to figure out the best one.
- Using metrics like Confusion Matrix and Accuracy to evaluate the models and select the best model.

- GitHub URL predictive analysis file.

Results

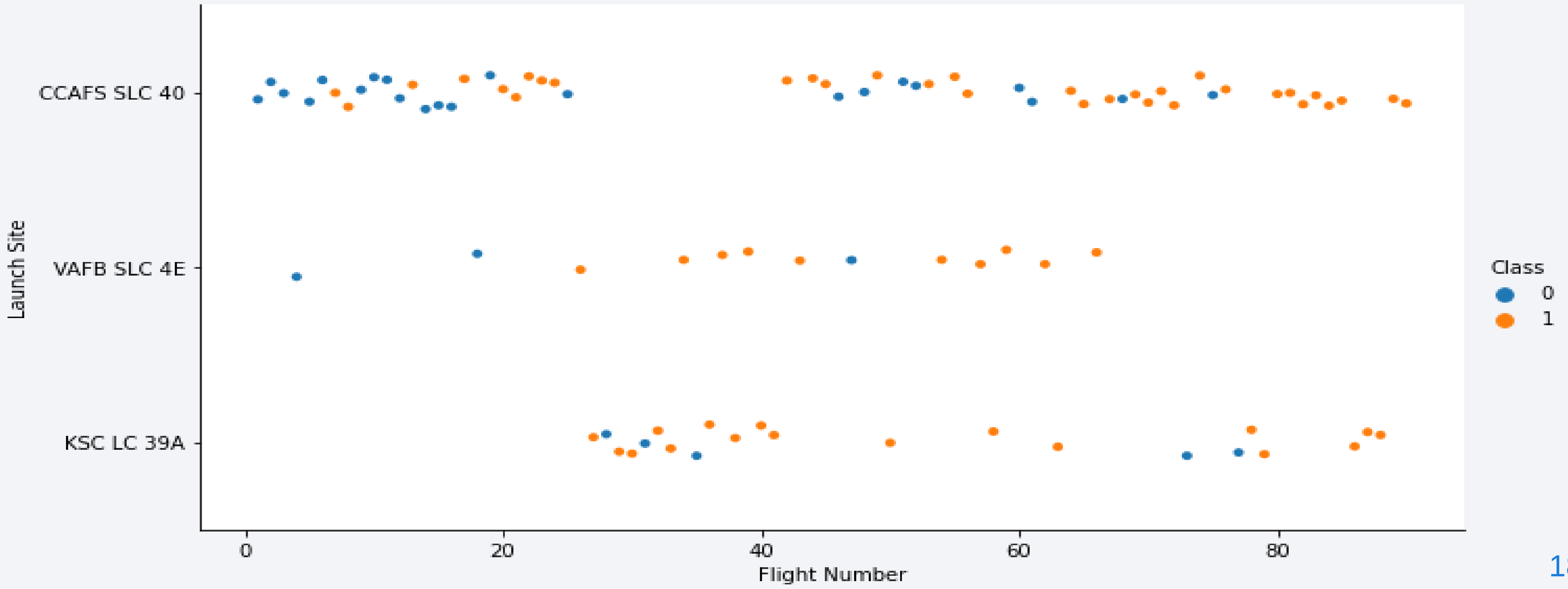
- **Exploratory data analysis** show that:
  - the first stage is more likely to landing successfully when (flight number) increase and (Payload) decrease, When (launches sites) are (VAFB SLC 4E – KSC LC 39A), when the Orbit type is (LEO – VLEO - SSO) and the average rate increase as the years increase.
- **Interactive analytics** show that:
  - the first stage is more likely to landing successfully when (launches sites) are (KSC LC 39A), the Payload decrease, and version of the Booster is (FT).
  - Fail Landing is more likely to happen when (v1.1) booster version is use.
  - All Launch sites has more than 50% of success Outcome
- **Predictive analysis** show that:
  - more than one classification model can predict the Outcomes status with 83% of certainty.

Section 2

Insights drawn from EDA

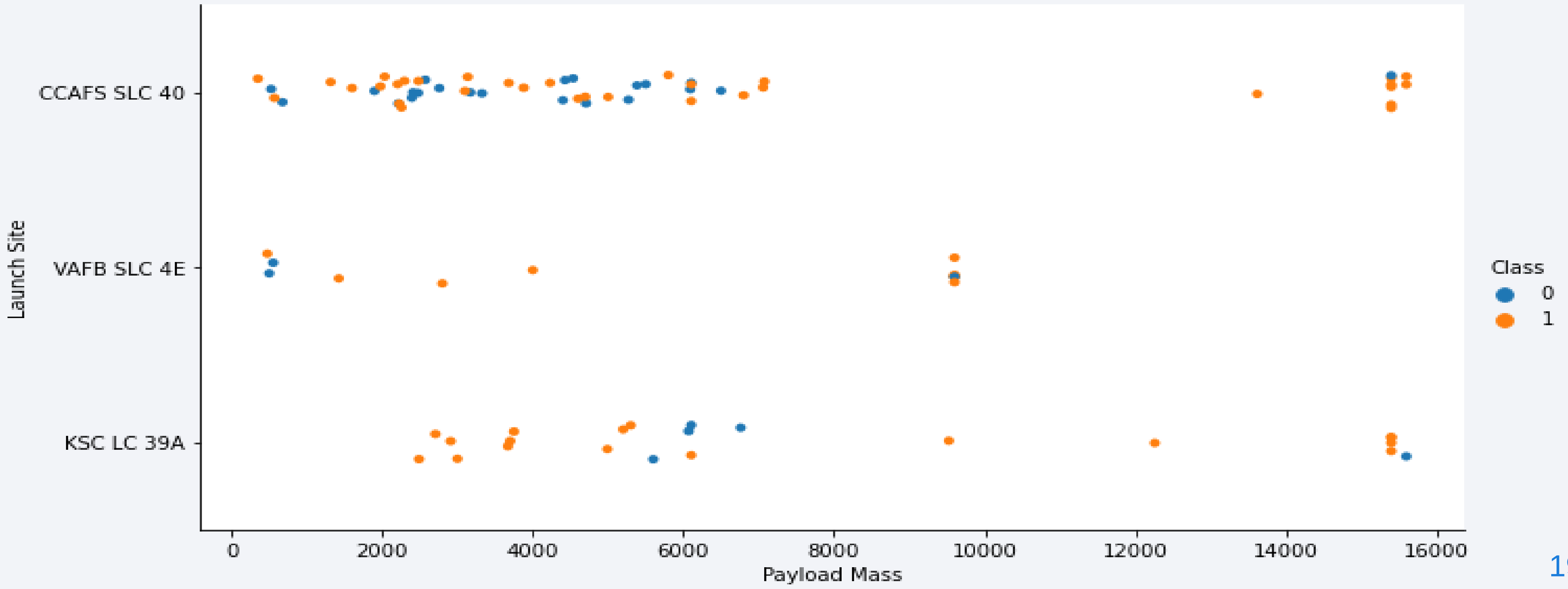
Flight Number vs. Launch Site

- When (Flight Number) increase the first stage is more likely to landing successfully specially with (VAFB SLC-4E) and (KSC LC-39A) (Launch Sites).



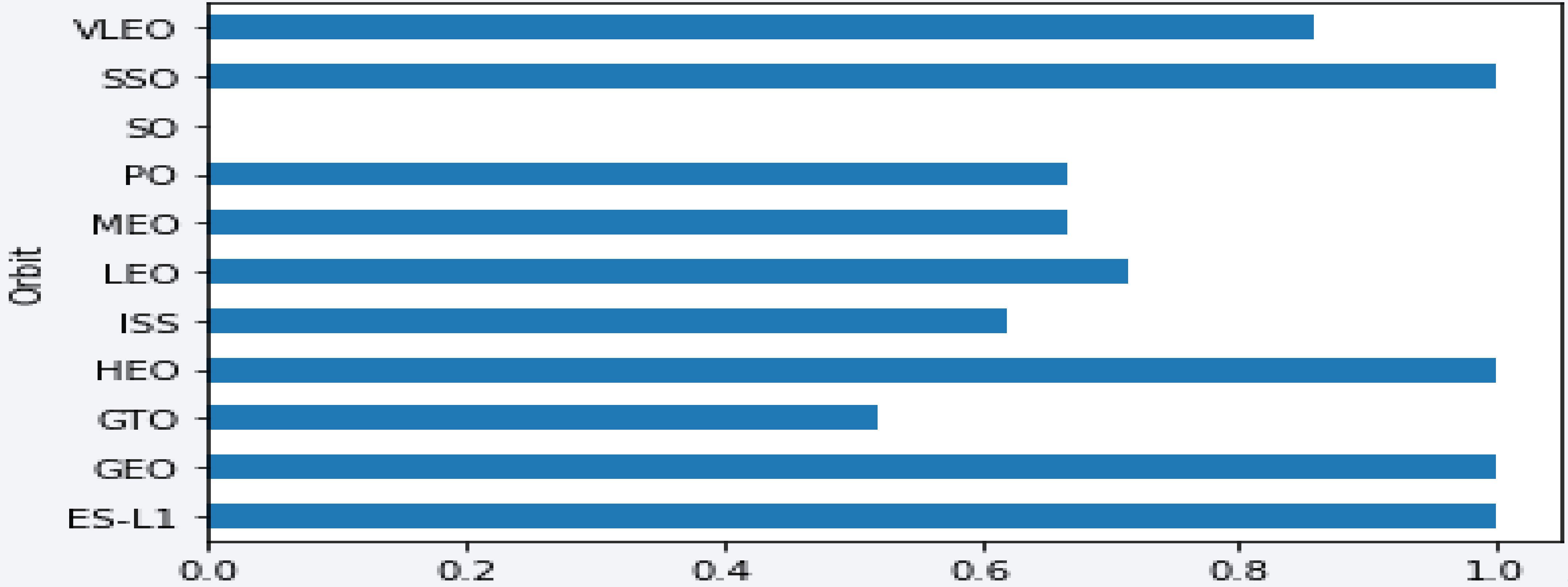
Payload vs. Launch Site

- the first stage is more likely to landing successfully when using (KSC LC 39A) (Launch Sites), and when the (Payload) decrease.



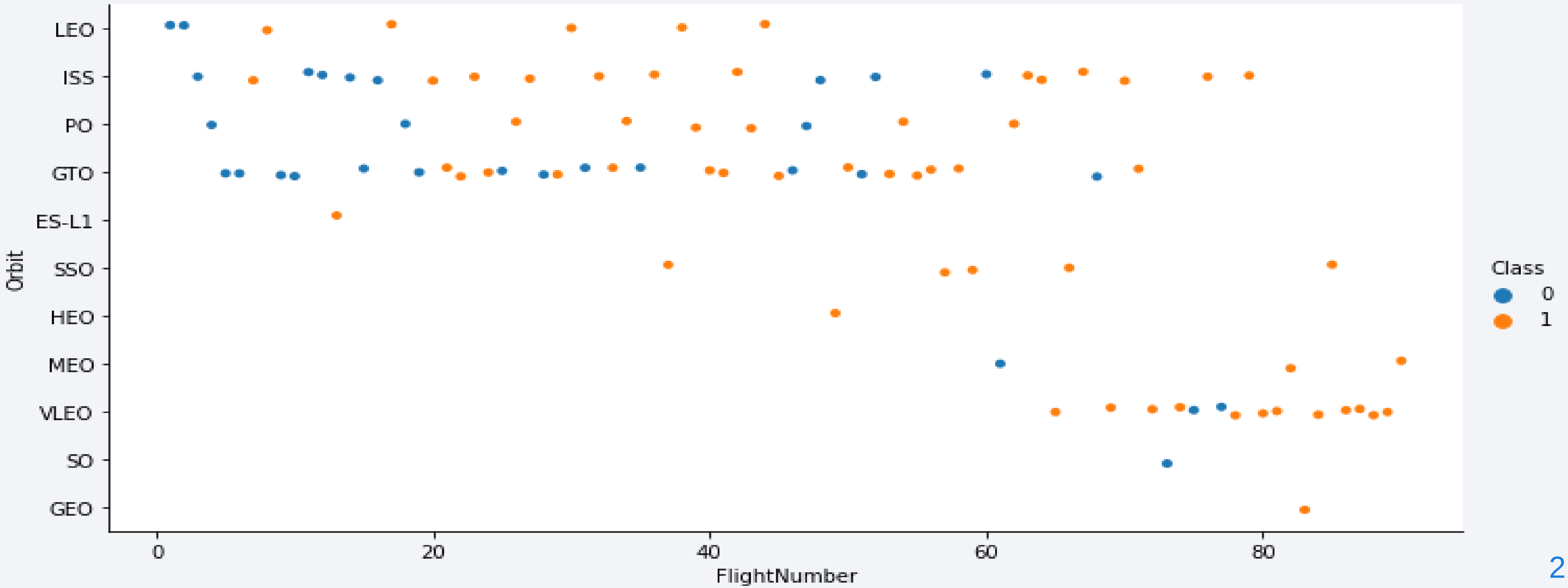
Success Rate vs. Orbit Type

- the probability of landing successfully with (Orbit),it is more likely when the Orbit types are (SSO – HEO – GEO – ES-L1 - VLEO)



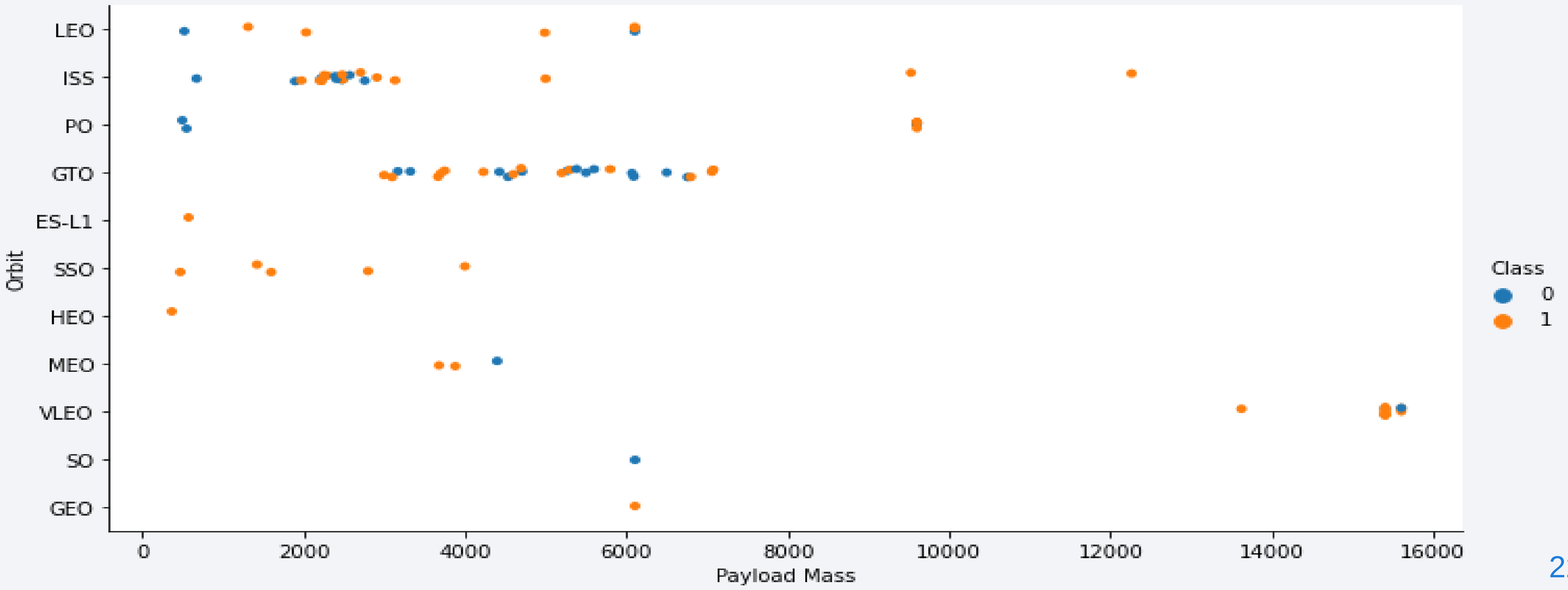
Flight Number vs. Orbit Type

- the first stage is more likely to landing successfully when (Flight Number) increase and the (Orbit) is one of (VLEO – LEO - SSO).



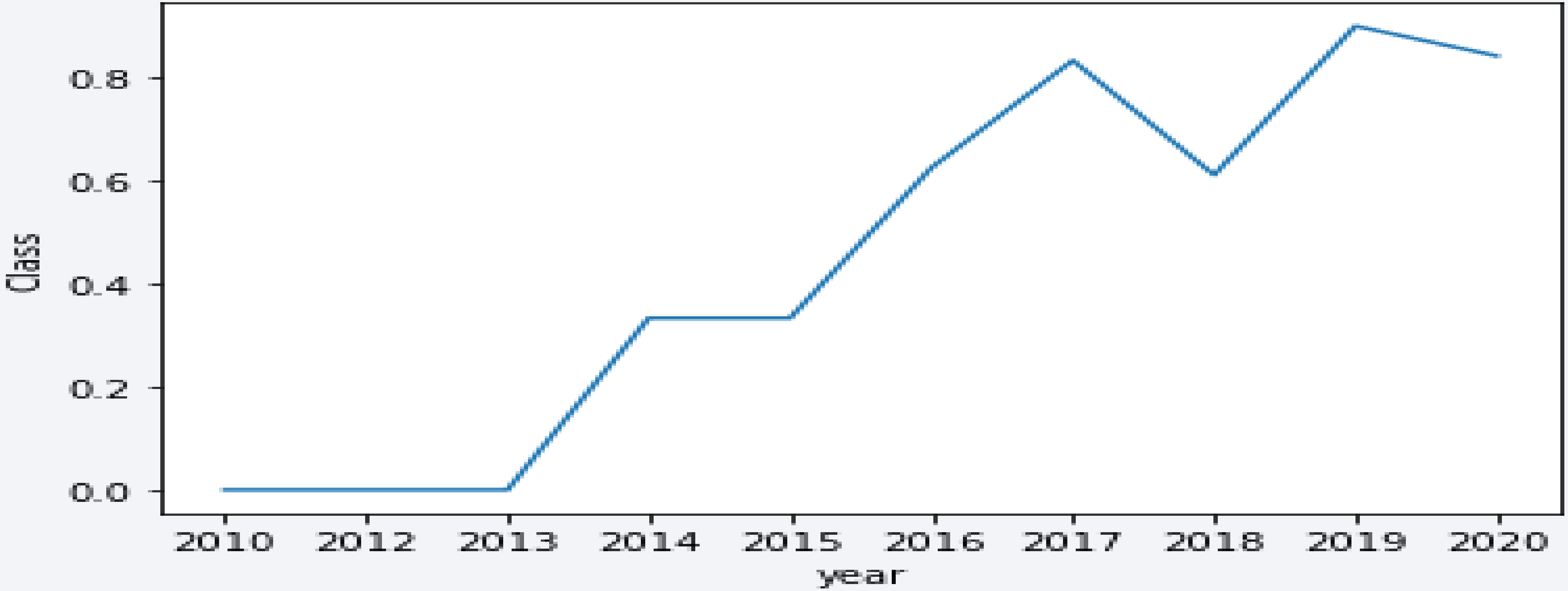
Payload vs. Orbit Type

- the first stage is more likely to landing successfully when (Payload) decrease and the (Orbit) is one of (SSO - LEO)



Launch Success Yearly Trend

- As the years increase the first stage is more likely to landing successfully.



All Launch Site Names

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

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

Total Payload Mass

NASA TOTAL PAYLOADS
45596

Average Payload Mass by F9 v1.1

AVG PAYLOADS
2928 KG

First Successful Ground Landing Date

Min DATE
01-05-2017

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

Total Number of Successful and Failure Mission Outcomes

Mission Outcomes
101

Boosters Carried Maximum Payload

Booster version
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

2015 Launch Records

Landing outcome	Booster version	Launch site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

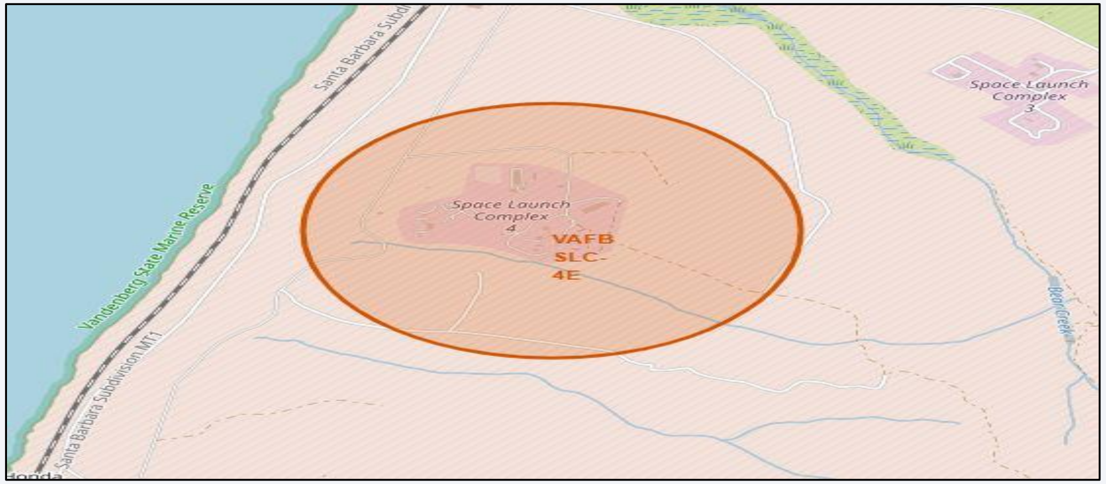
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing outcome	Count
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	22
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

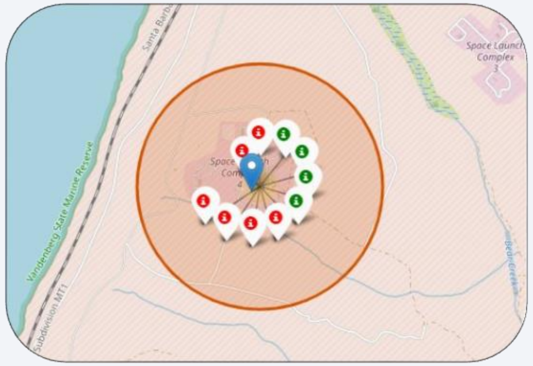
Section 3

Launch Sites Proximities Analysis

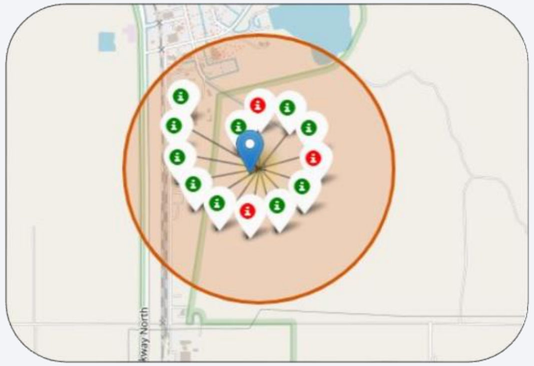
Markers of Launch Sites on the Map



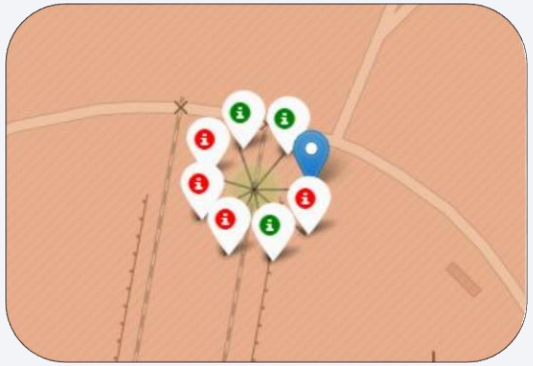
Color Labeled Launch Outcomes on the Map



VAFB SLC-4E



KSC LC-39A

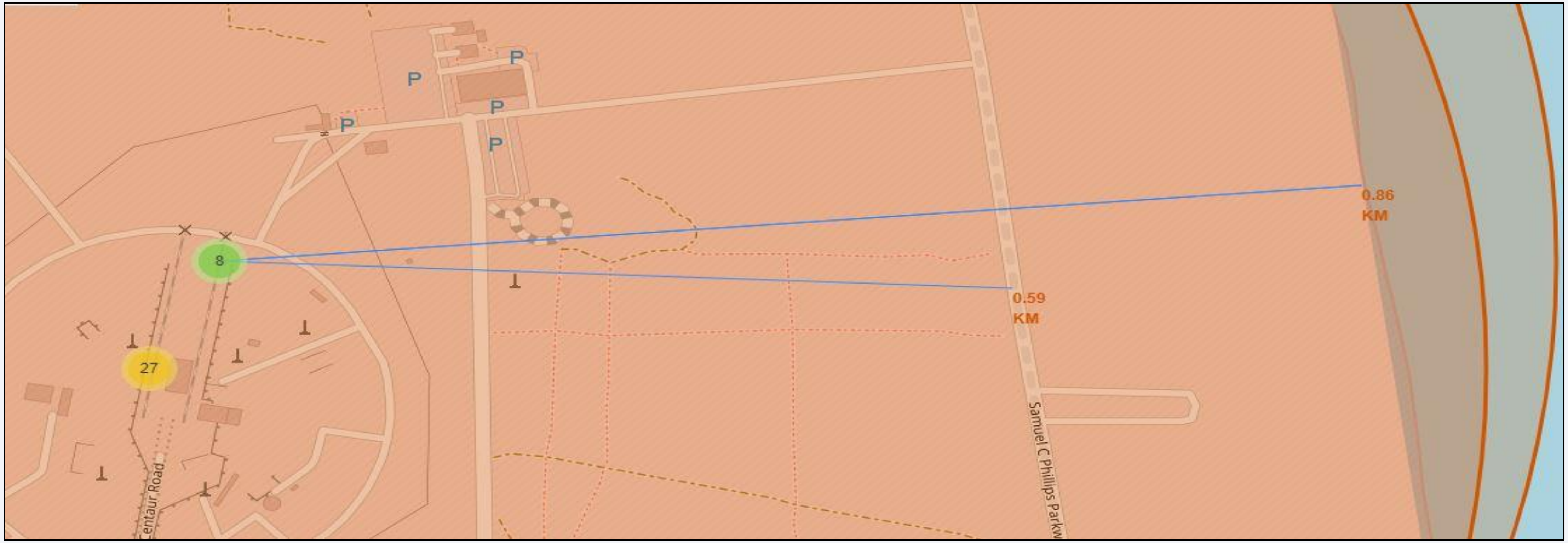


CCAFS SLC-40

CCAFS LC-40

Proximities for a Launch Site

- The launch site (CCAFS SLC-40 and CCAFS LC-40) has Proximities like highway and coastline.

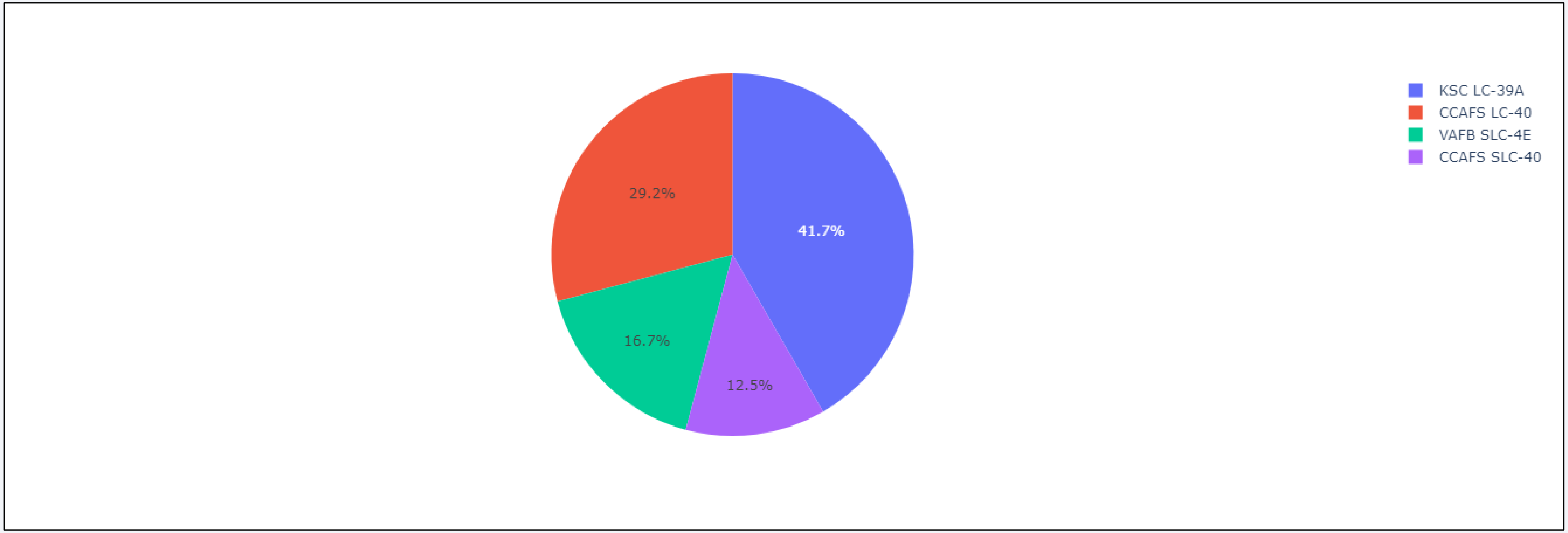


Section 4

Build a Dashboard with Plotly Dash

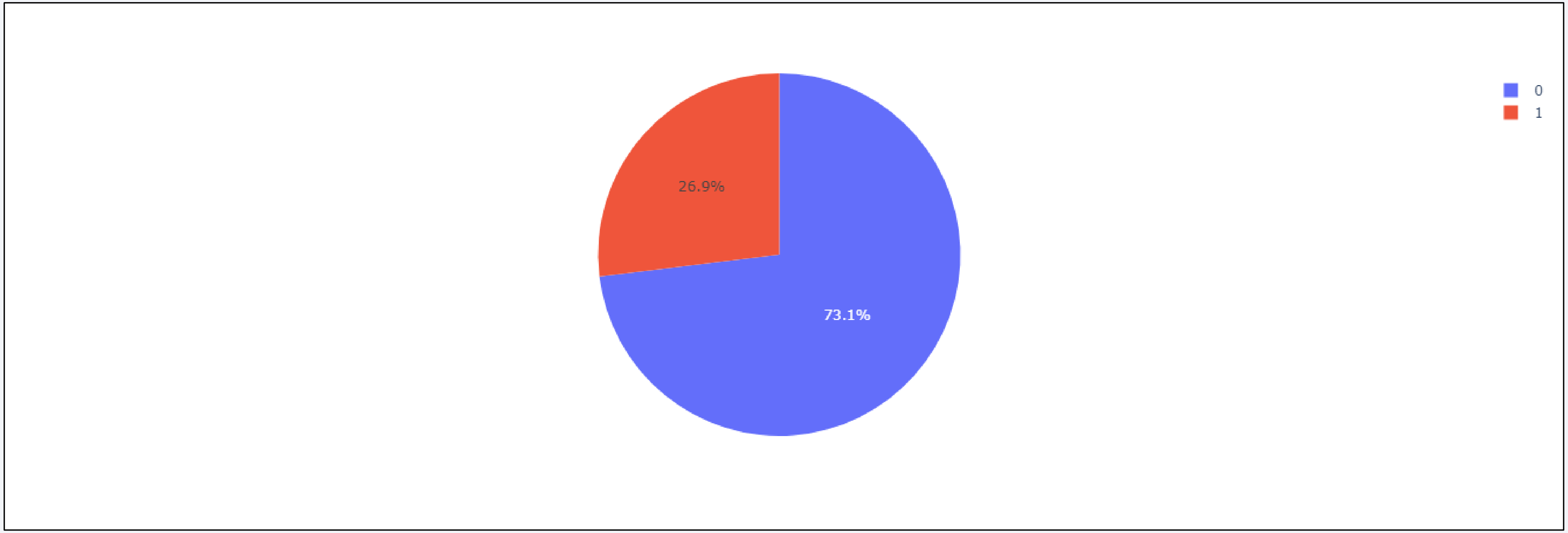
The Launch Success Count for All Sites

- Each site and its landing success outcomes per percent.



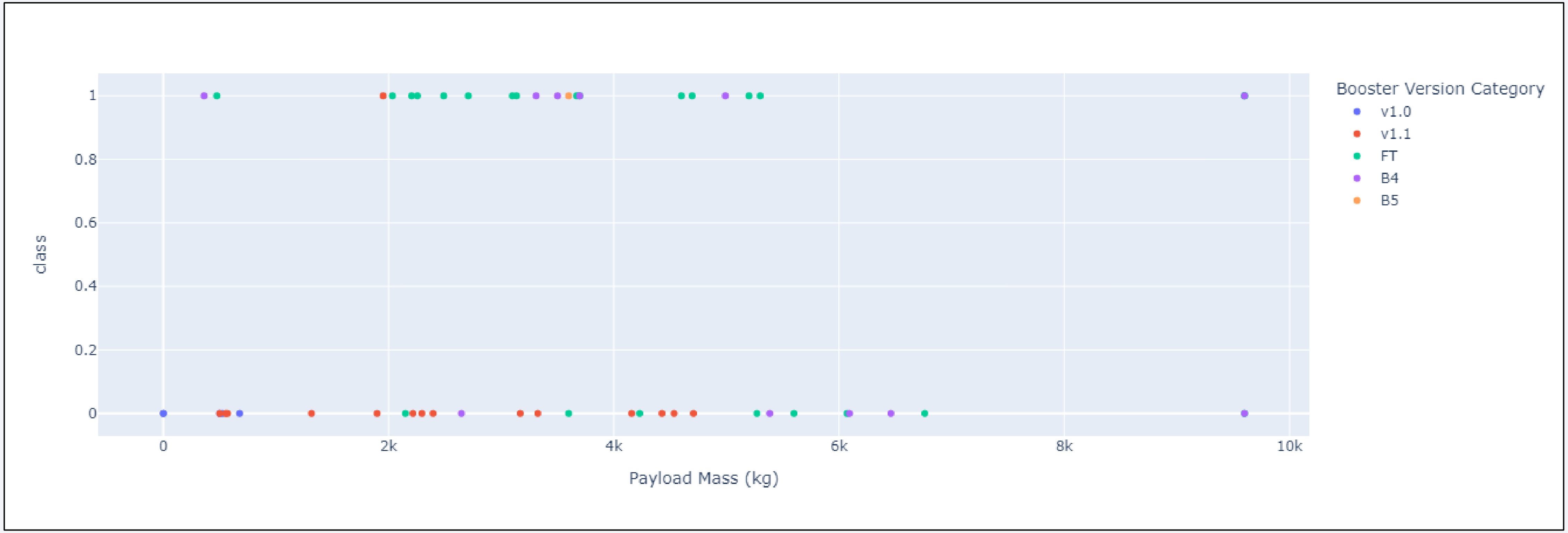
Highest Launch Success Ratio Site

- (CCAFS LC-40) has 73.1% of Landing success Outcomes.



All Sites Payload vs. Launch Outcome

- The success Outcomes is more when Booster version is (FT) with low Payload, and less when it is (v1.1)

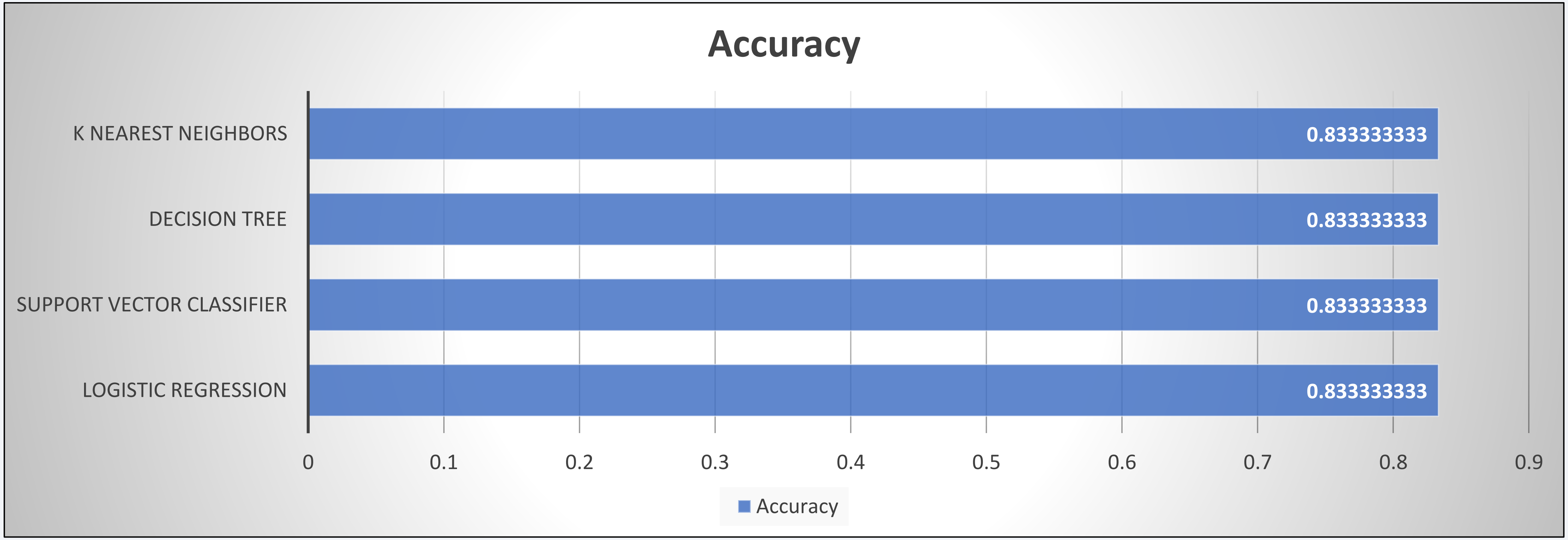


Section 5

Predictive Analysis (Classification)

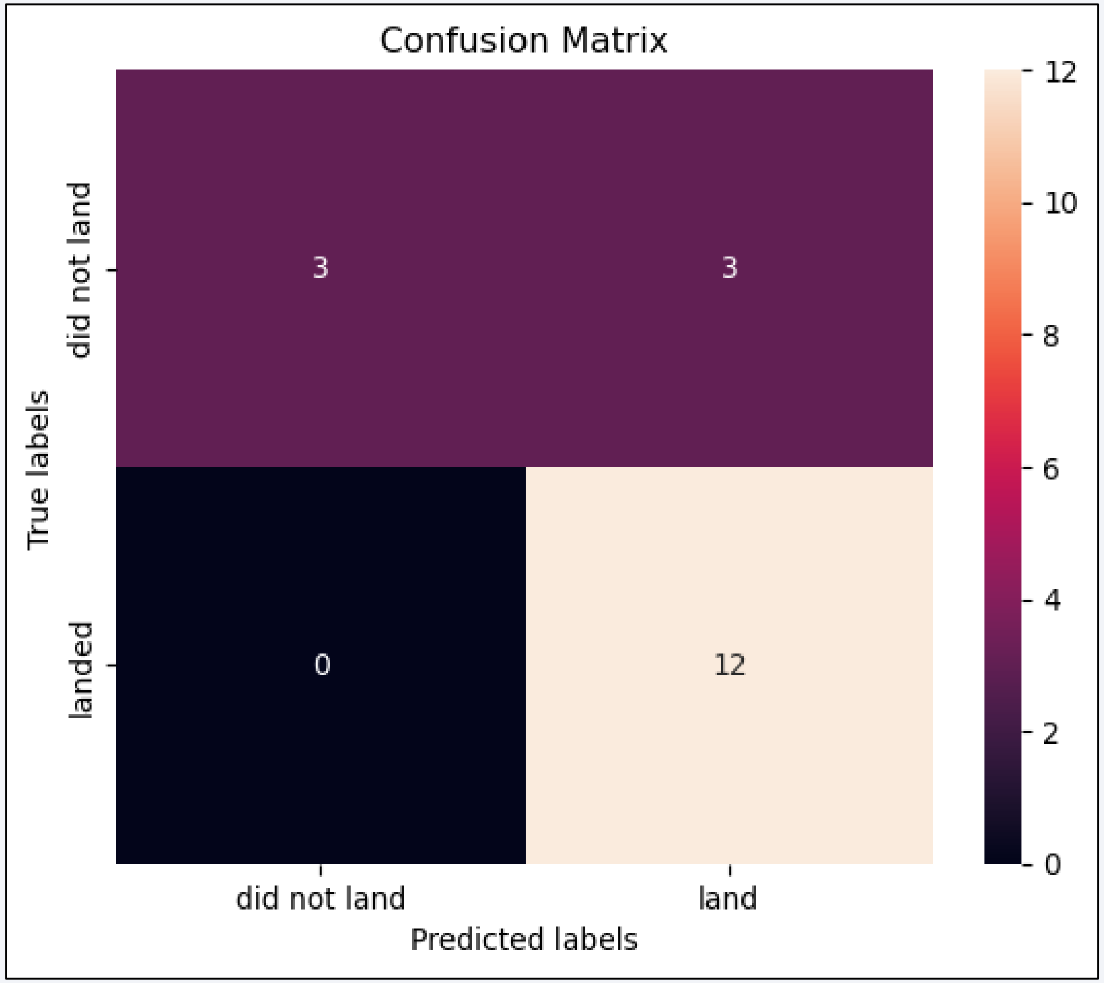
Classification Accuracy

- The Accuracy of the all models has nearly the same value.



Confusion Matrix

- The model predict all success landing samples rightly.
- The model predict a half of the failure landing samples rightly.



Conclusions

- Studied SpaceX data to figure out what and how features effect the return of the first stage of their rockets and build a prediction model for this process, Found that there is relation between (Flight Number), (Payload), (Orbit), (Booster version) and (Launch Site) against the Landing Outcomes.
- As much as the (Flight Number) increase the Landing is more to success, such like Orbits (VLEO – LEO – SSO), Launch sites (VAFB SLC-4E – KSC LC-39A) and Booster versions (FB) on the other hand (Payload) increase cause the opposite.
- Performed a predictive analysis and built a classification models like Logistic Regression, Support Vector Machine and Others models, evaluated by Confusion Matrix and Accuracy metrics, afford 83% of certainty to give the right Outcome, depending on the studied features.
- The results are SpaceY can figure out what standards that has an effect on the landing process and predict whether their Rocket first stage will landing safety or not with 83% of certainty.

Appendix

Thank you!