Winning Space Race with Data Science

Amruth GL

20/11/2023

Outline

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

Executive Summary

- To determine if first stage of SpaceX Falcon 9 we have collected data from public SpaceX API and by scrapping SpaceX Wikipedia page and use Beautiful Soup library for it.
- Created labels column ‘Class’ which denotes all the successful landings. We performed EDA(Exploratory Data Analysis) using SQL concepts like magic sql.
- We explored many queries using Group by aggregate functions.
- We have visualized the results of the queries using Seaborn and pandas libraries. Also created dashboards using the queries for details EDA.
- Later we have used different algorithms like logistic Regression, Destion Tree, KNN, SVM to predict the class labels and we found that the Logistic Regression predicts the class labels with highest accuracy.

Introduction

- Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. Space X
- From this project we would like to find if the first stage will land, and we want to predict this with highest accuracy to determine the cost of a future launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - We have collected data from public SpaceX API and by scrapping SpaceX Wikipedia page and use Beautiful Soup library for it.
- Perform data wrangling
  - Here we find some patterns in the data and determine the label for training supervised models. We convert the successful landings outcomes into training labels with 1 means booster successfully landed and 0 means it was unsuccessful.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Split the data into Train and test data and find the best Hyperparameter for SVM, Trees and Logistic Regression and find which method performs best.

Data Collection

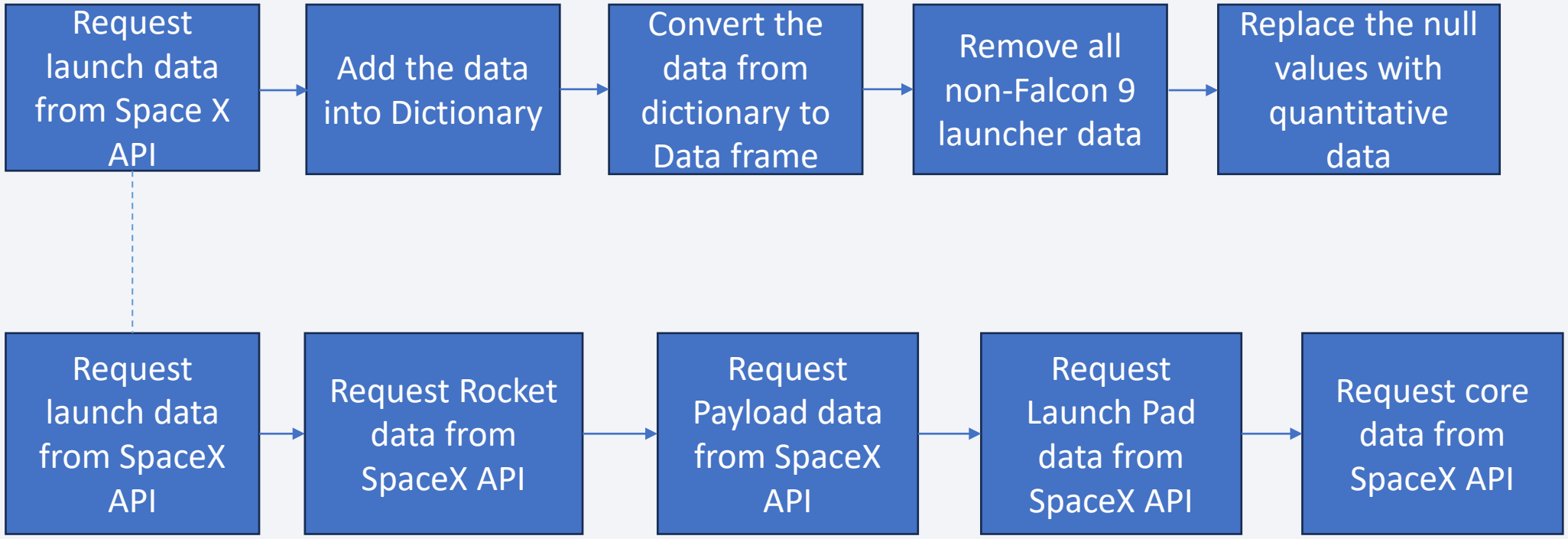
- Data was collected we have used following method.

1. Using public SpaceX API
2. Using SpaceX Wikipedia page using Beautiful soup library.

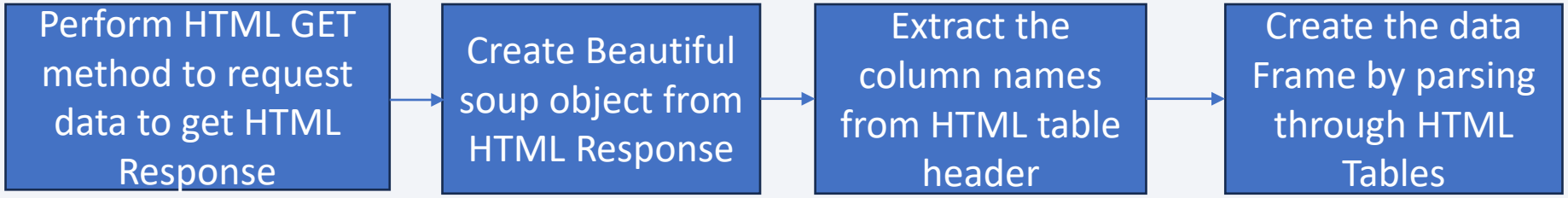
- At first, we need to extract the SpaceX Falcon 9 launch records HTML table from Wikipedia.
- Parse the table and convert the table into pandas Dataframe

GitHub Link Repo

Data Collection – SpaceX API



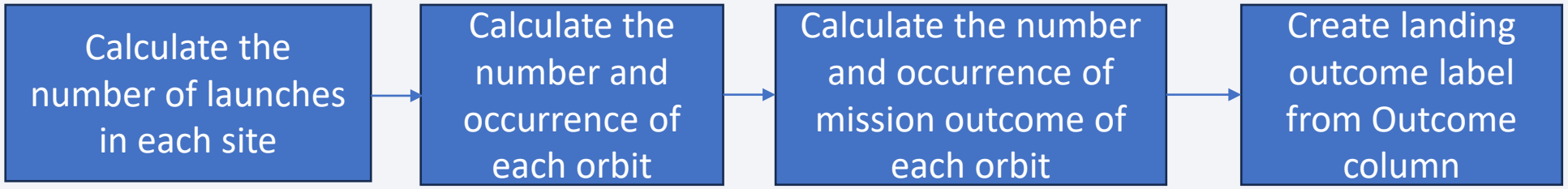
Data Collection - Scraping



GitHub Link Repo

Data Wrangling

- In Data Wrangling stage we need to categorize the data as successful launch and unsuccessful lanch
- Based on the “Outcome” column we create one more label column “Class” where successful outcomes are marked as 1 and unsuccessful outcomes are marked as 0



GitHub Link Repo

EDA with Data Visualization

- Visualized the relationship between Flight number and Launch site using scatter point chart
- Visualized the relationship between Payload and Launch site using scatter point chart
- Visualized the relationship between success rate of each orbit type using bar chart
- Visualized the relationship between Flight number and orbit type using scatter point chart
- Visualized the relationship between Payload and Orbit type using scatter point chart
- Visualized the launch success yearly trend using line chart

GitHub Repo Link

EDA with SQL

- At first, we need to load SQL extension and establish a connection with the database
- We did create basic queries to get the better understanding of the relationships between the variables in the data.
- Mainly we worked on the Launch site, Payload Mass, Mission outcome, and booster versions.

GitHub Repo Link

Build an Interactive Map with Folium

- Using Folium python library, we created a interactive map where the launch sites have been marked in circle.
- Along with the launch site we check how many launch have been done in each launch site. These are added using markers from Folium libraries.
- Created Line markers to understand the if the launch sites are nearby to the cities, Highways, Railway tracks.

- GitHub Repo Link

Build a Dashboard with Plotly Dash

- Using Plotly Dash we created Dashboard which shows the following:

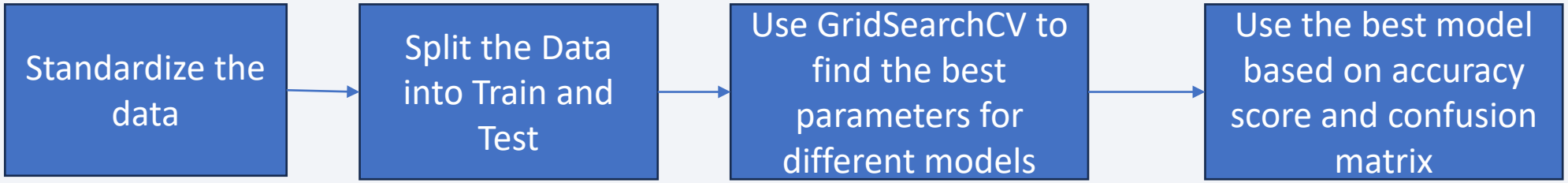
1. Pie chart which shows the different successful launches by launch site.
2. A scatter plot which show the correlation between Payload Mass and suc

This Interactive dashboard provides the relationship between launch sites, its success rates and Payload Mass

GitHub Repo Link

Predictive Analysis (Classification)

- Before classification and predicting we need standardize the data.
- Data is then split into Train and Test and Train the data with different algorithms and find the best performing algorithms and their parameters.
- We used accuracy score and Confusion Matrix to validate the best performing models.
- GitHub Repo Link



Results

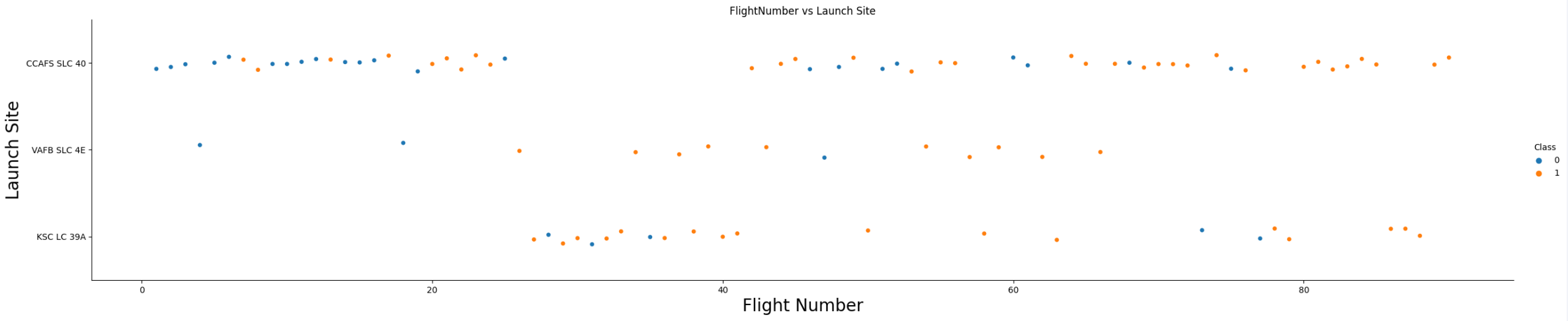
- Exploratory data analysis results:
- Launches after 2017 seems to be more successful compared to previous years.
- When the payloads are lighter recovery of the first stage seems to be more successful.
- KSC LC-39A appears to be most successful with the highest success rate of 75%

- Screen shots are available in the upcoming sections.
- Predictive analysis results: Each model performed about equally. Predicting accuracy is 83.33%

Section 2

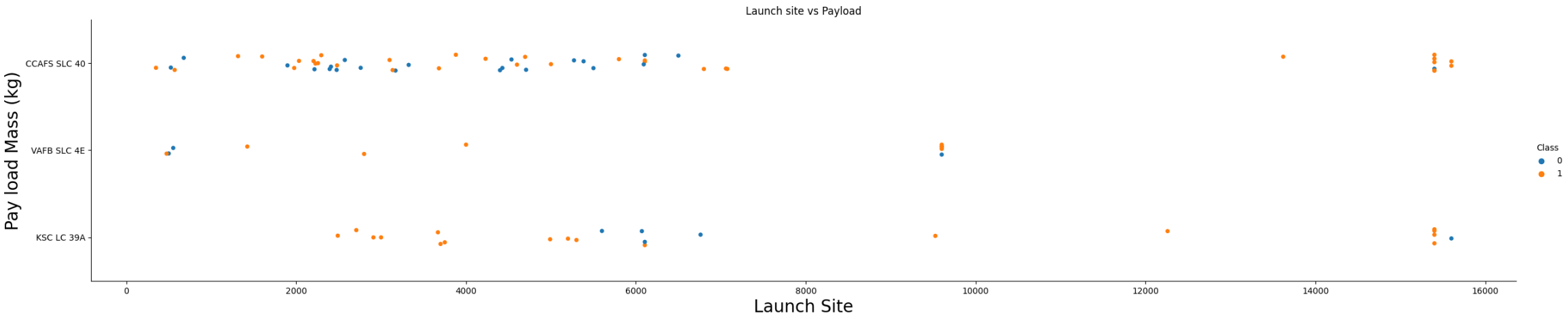
Insights drawn from EDA

Flight Number vs. Launch Site



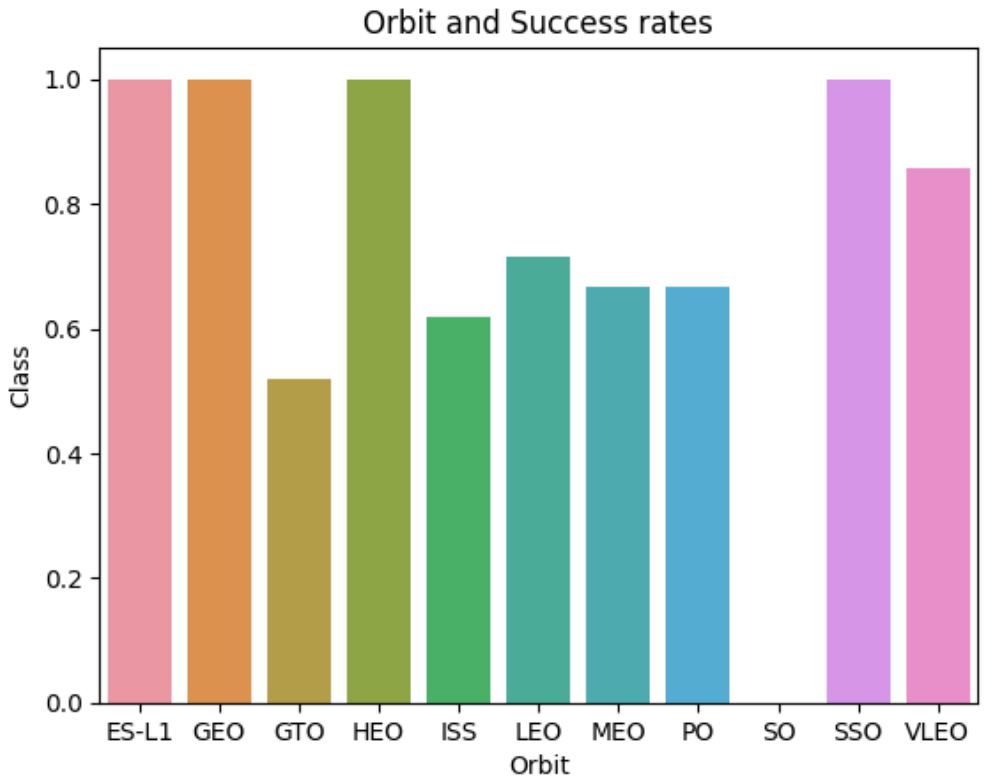
- From the graph we can understand the that as the flight number increases the success rates also increases.
- CCAFS SLC 40 and KSC LC 39A have more success when the flight number has increased above 80

Payload vs. Launch Site



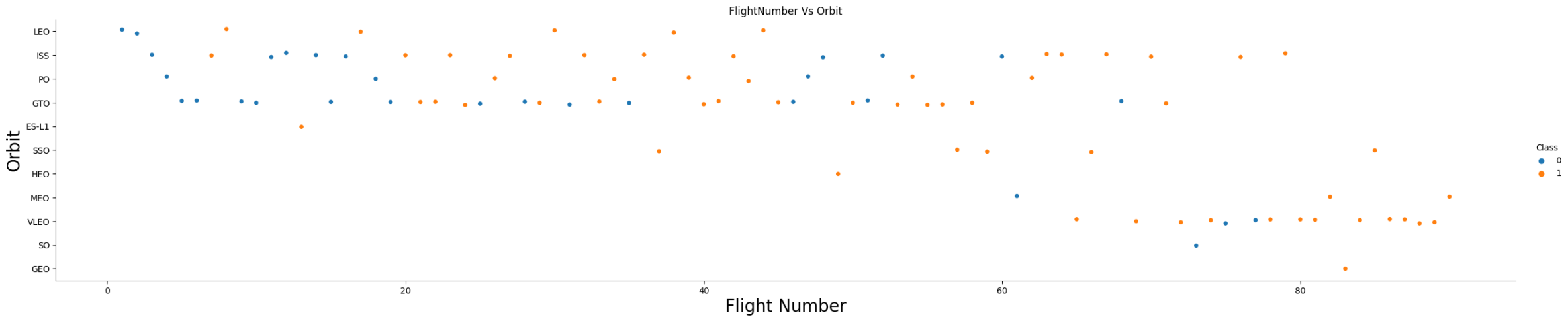
- From the chart we can find that the for VAFB-SLC launchsite there are no rockets launched for heavy payload mass greater than 10000.

Success Rate vs. Orbit Type



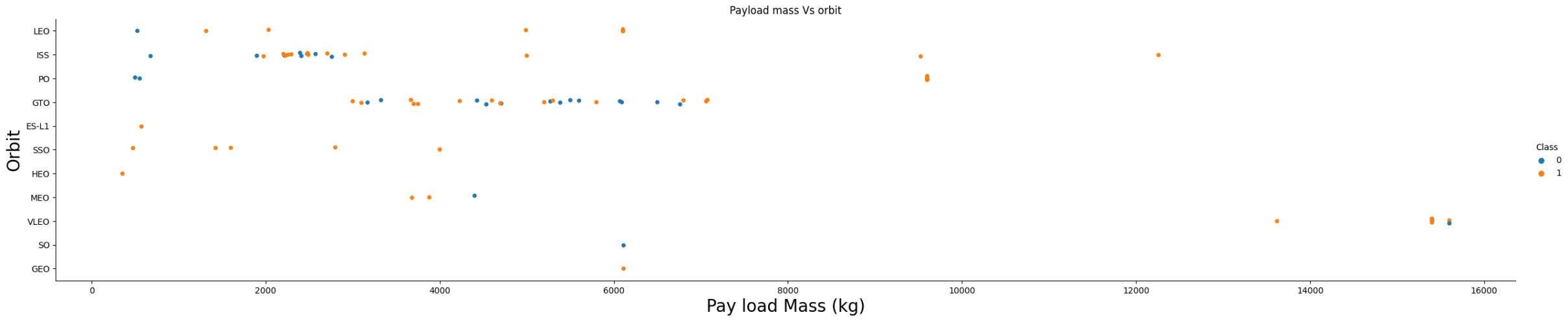
- From the bar graph we ca see that ES-L1, GEO, HEO and SSO has the highest success rate compared to other orbits.
- SO orbit has the least success rates

Flight Number vs. Orbit Type



- For LEO orbit the success rate seems to relate to the flight numbers
- On the other hand, for GTO there seems to be no relation with Flight Number

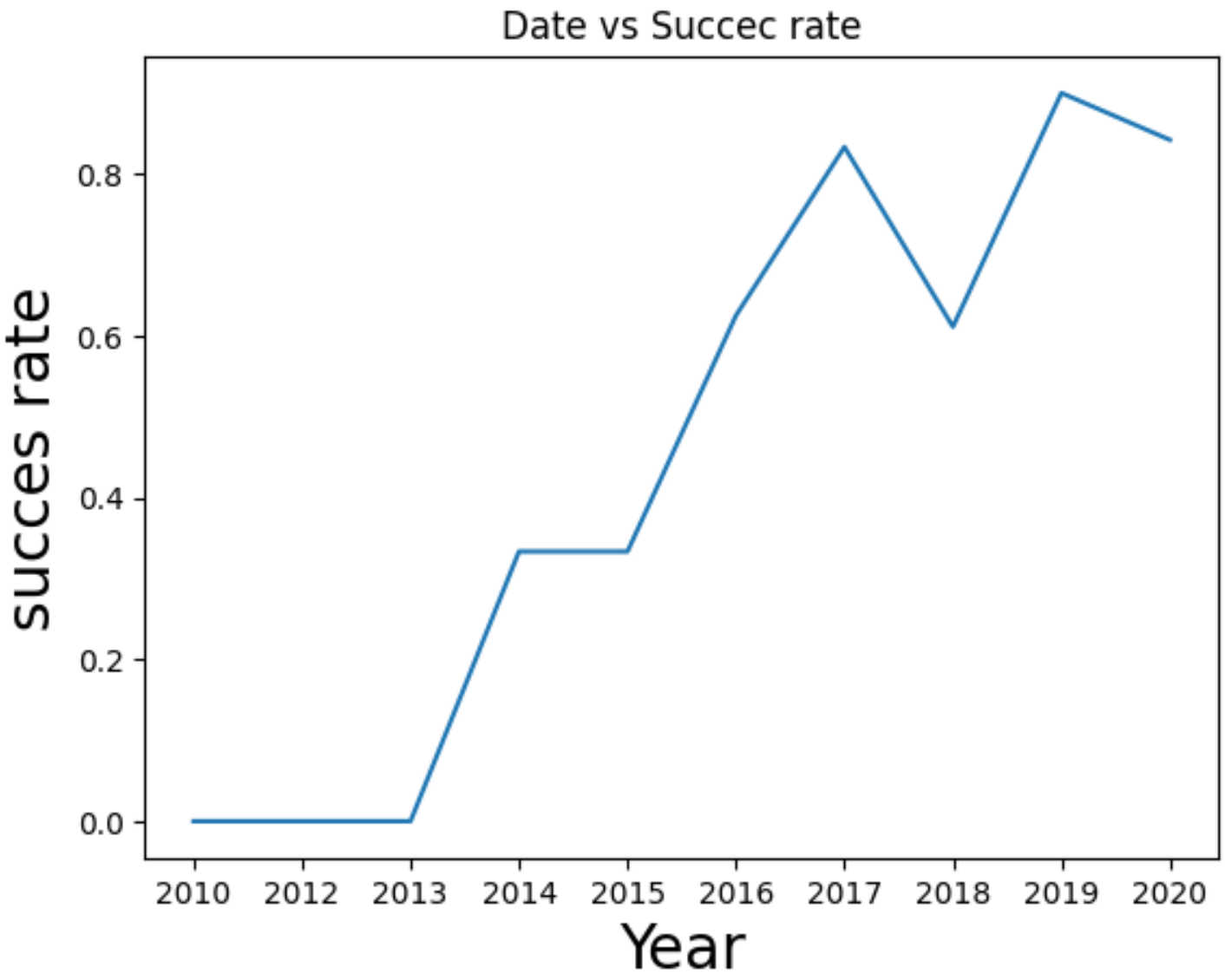
Payload vs. Orbit Type



- With Heavy Payload Mass for LEO, PO seems have higher success rates.

Launch Success Yearly Trend

- From 2013 the success rates kept on increase except some dip in 2018



All Launch Site Names

- From SQL query we found the there are 4 launch sites

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

Launch Site Names Begin with 'CCA'

- From SQL query following are the first 5 records which have launch site whose name start 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

Total Payload Mass

- Total payload Mass caried by booster launched by NASA (CRS)

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

MIN(Date)	Landing_Outcome
2015-12-22	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- Table shows the Launches which were successful when drone ship was used for landing which has the payload mass between 4000 to 6000

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

- There are large no of successful missions compare to failure.
- Out of 101 records there are 100 successful missions

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

Boosters Carried Maximum Payload

- Following list is the list of boosters which have carried the 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

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

substr(Date, 6,2)	Booster_Version	Launch_Site
10	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40
01	F9 v1.1 B1017	VAFB SLC-4E
04	F9 FT B1020	CCAFS LC-40
06	F9 FT B1024	CCAFS LC-40

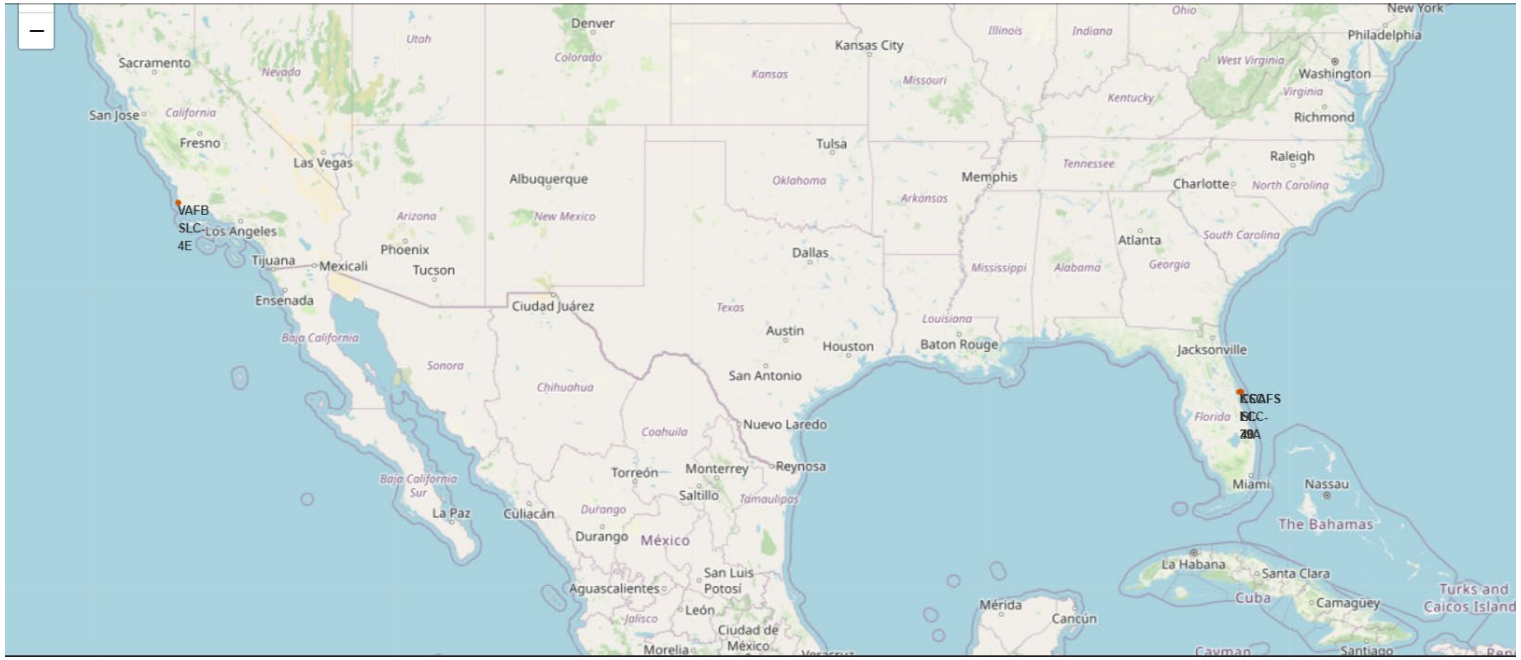
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Table show the 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	OUTCOME_COUNT
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1


Section 3
Launch Sites
Proximities Analysis

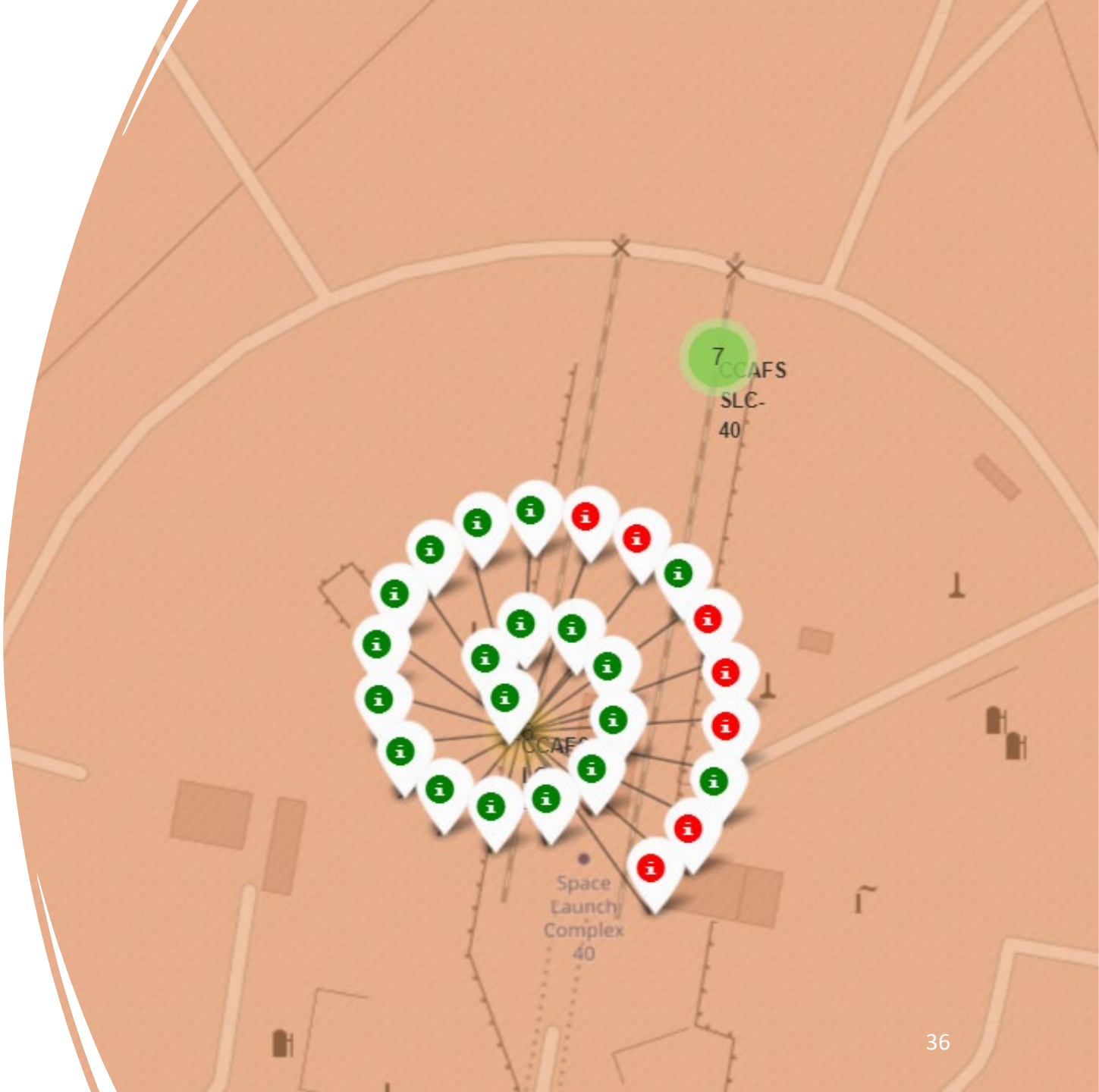
Launch Sites

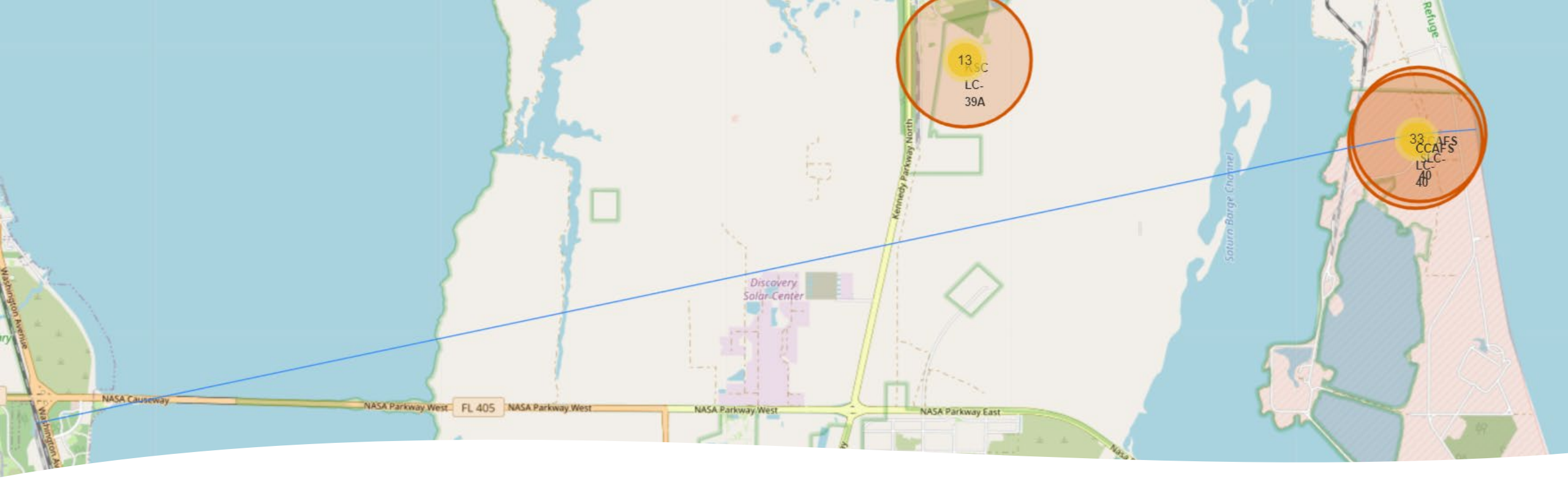


- Launch sites are mostly near the coastal area
- Launch site are mostly near to equator

Launch outcomes

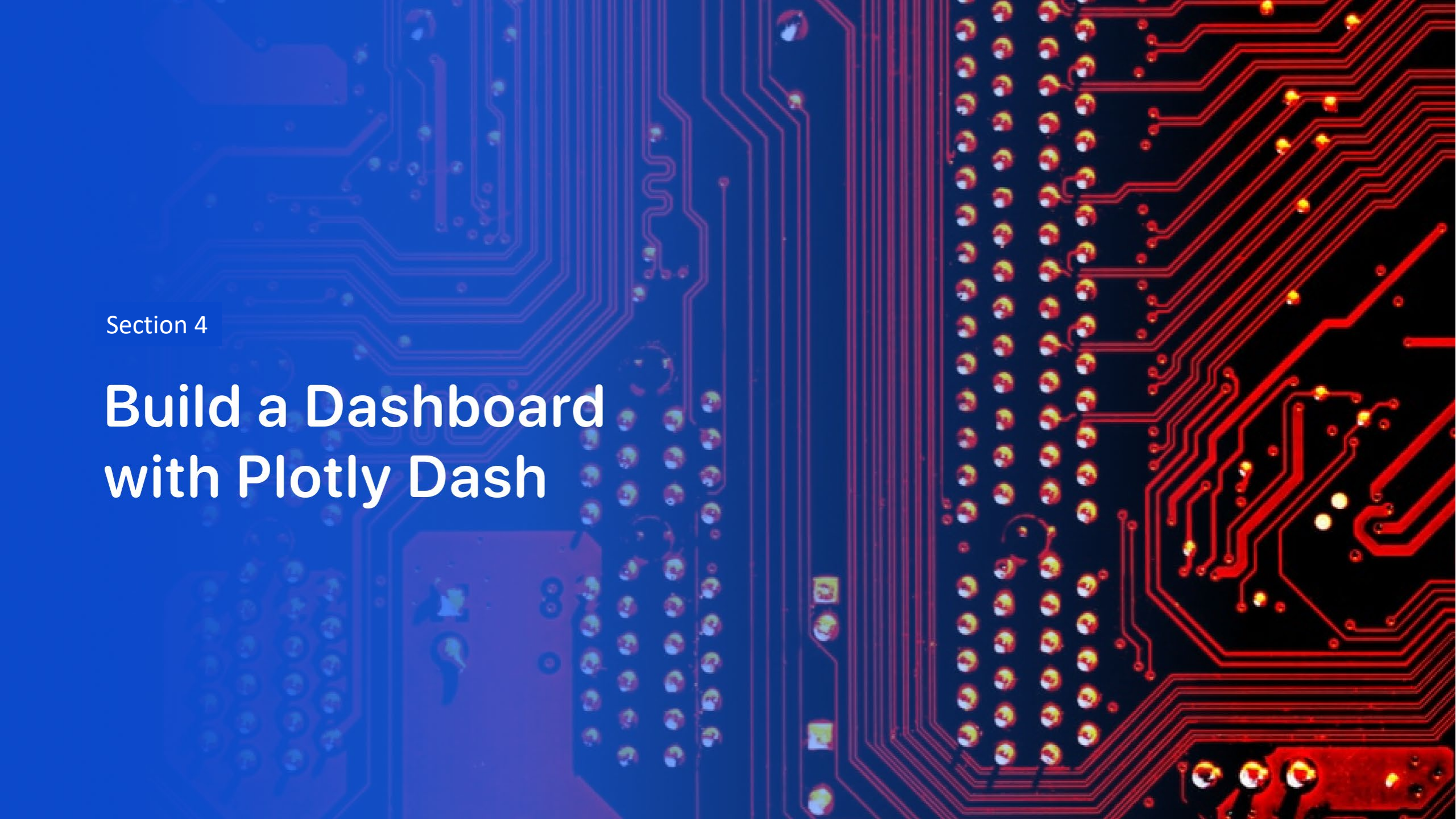
- Each marker represents the Falcon 9 recovery outcomes
- Red Marker indicate – Failed recovery
- Green Marker indicate – Successful recovery



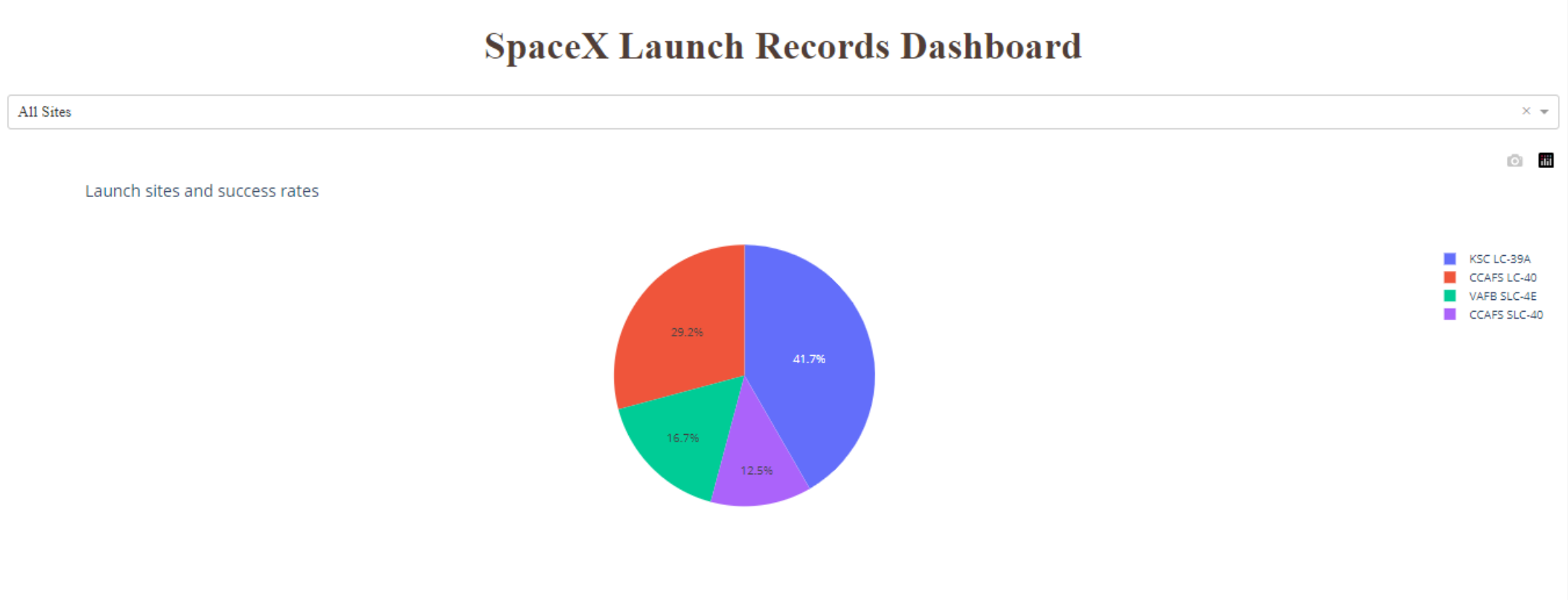


Near by locations

- Blue line shows the nearest location from launch station towards nearest coast line and to near city and highway

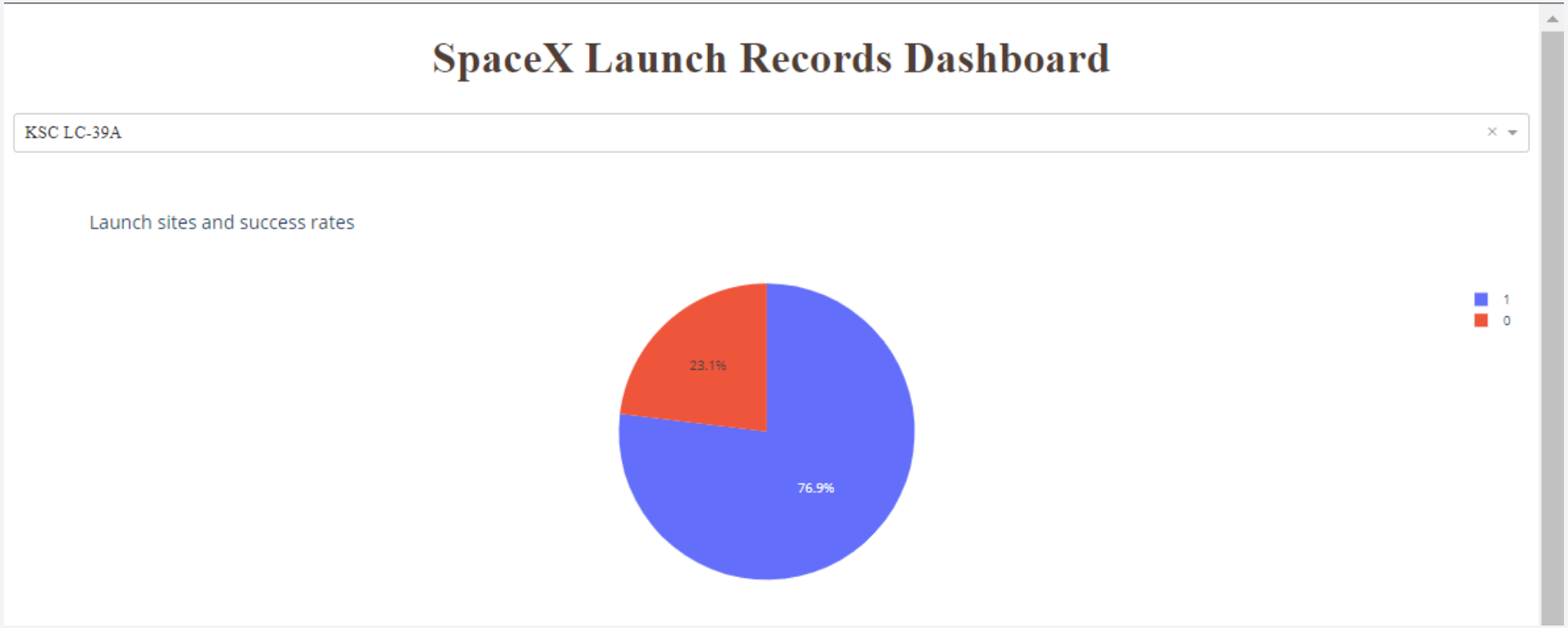

Section 4
Build a Dashboard with Plotly Dash

Launch sites and Success rates



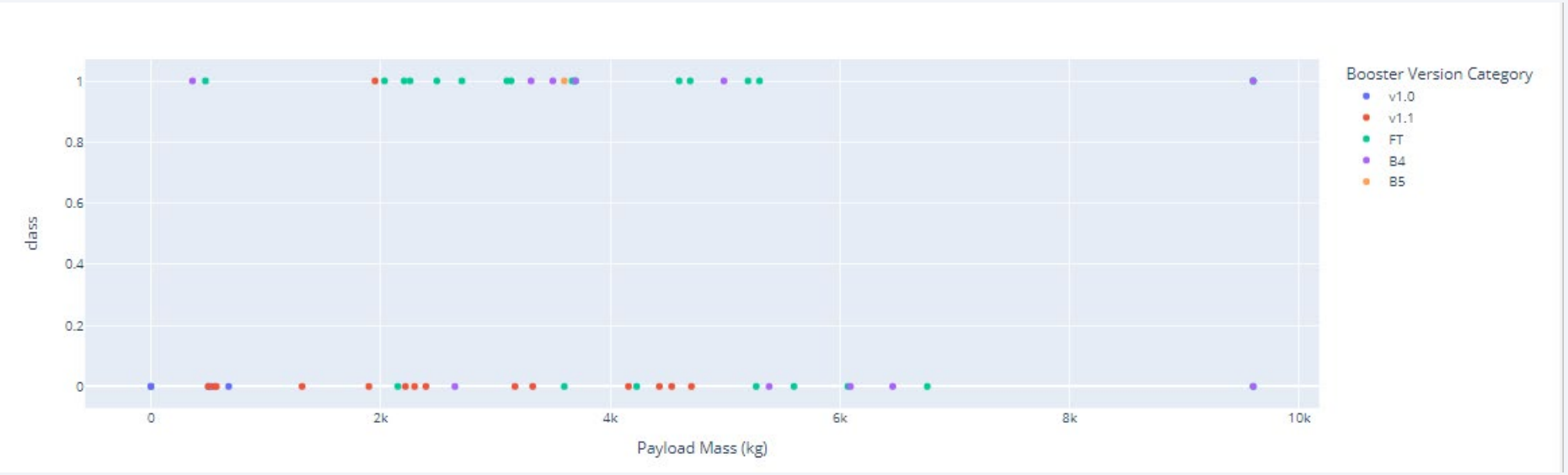
- Pie Chart shows the success rates of each Launch Site
- KSC LC-39A Launch site has the highest success rate.

Lauch site with highest success



- Pie Chart shows the Launch site success rate of most successful launch site
- It has 76.9% of success rate and 23.1% of failures in recovering first stage of Falcon 9 launch vehicles

Payload vs Launch outcome Scatter plot



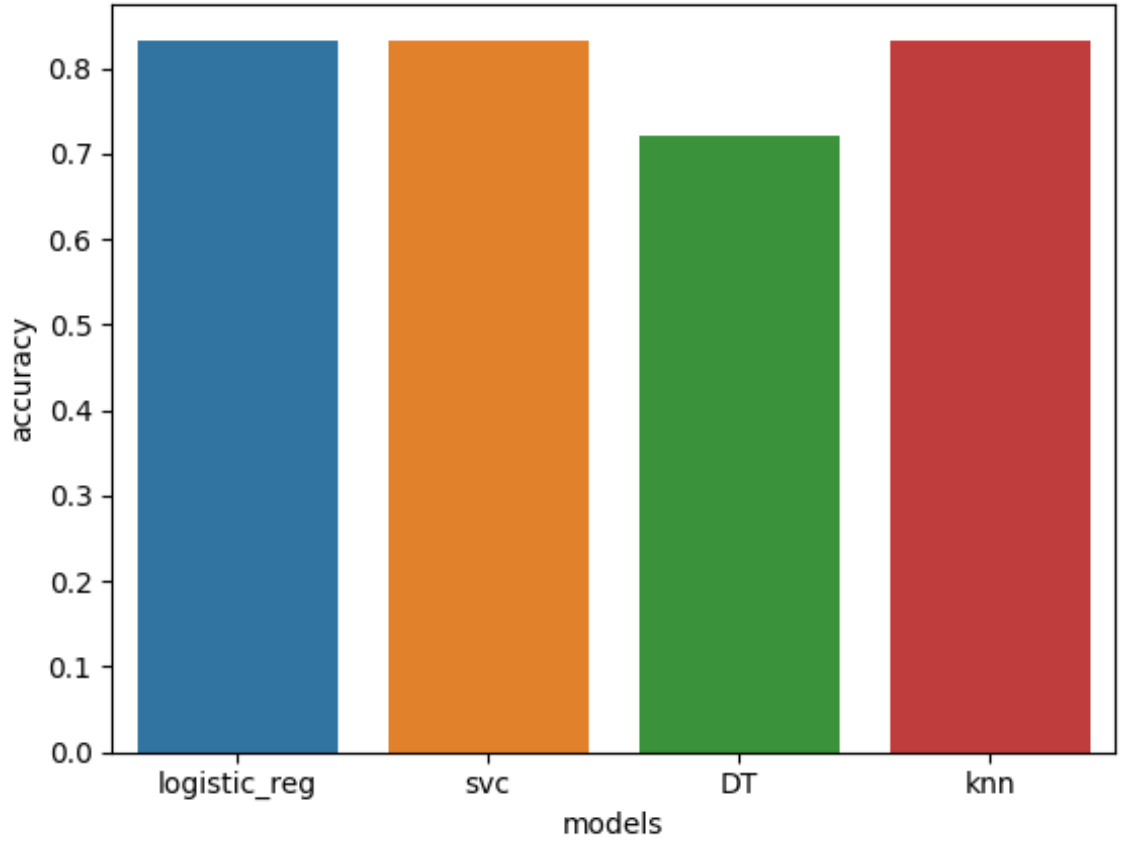
- Booster Version v1.1 seems to be having least success
- FT seems to be having more success

Section 5

Predictive Analysis (Classification)

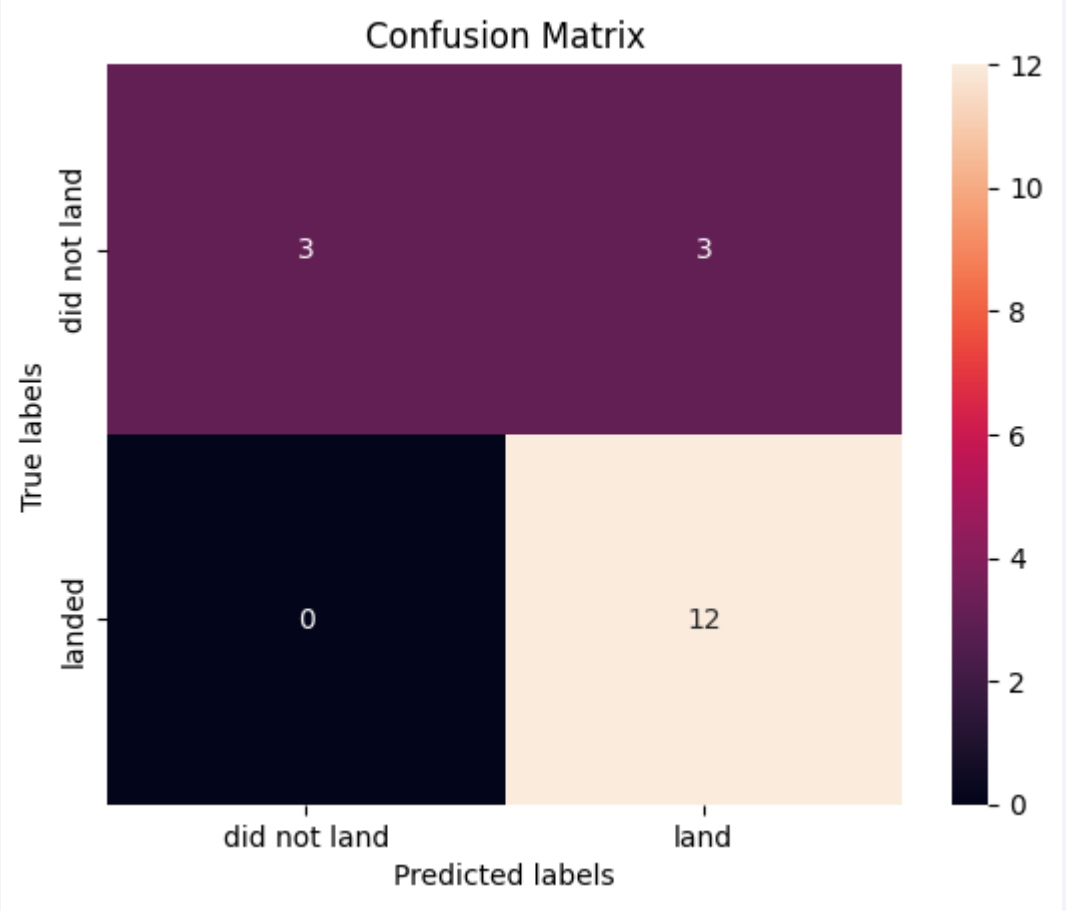
Classification Accuracy

- Bar plot show the Test Data accuracy of each model.
- All the model have almost same test data accuracy (83%)except Decision Tree.



Confusion Matrix

- Models predicted correctly for most of the case but did not predict correct outcome for 3 cases.
- Model predicted that it did not land for 3 launches which has successfully landed. Accuracy is 83.33%



Conclusions

- Space X has the successful recovery for the following

1. If the launch site is from KSC LC-39A
2. If the payload is lighter that is Payload mass is in between 2000-4000kg
3. If the recovery is done in drone ship

The Best model can predict the outcome of the recovery with a reasonable accuracy of 83.33%.

Thank you!