IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sebastian Juri Aguirre
Oct 2024

OUTLINE

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

EXECUTIVE SUMMARY

Used methodologies like:

- Data Collection SpaceX API
- Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia
- Exploratory Data Analysis (EDA) using SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Results:

- EDA results
- Visual analytics and dashboards
- Predictive analysis (classification model)
- Confusion matrix
- Conclusions

INTRODUCTION

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX 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 SpaceX can reuse the first stage.
- 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 SpaceX for a rocket launch.

Section 1

Methodology

Methodology

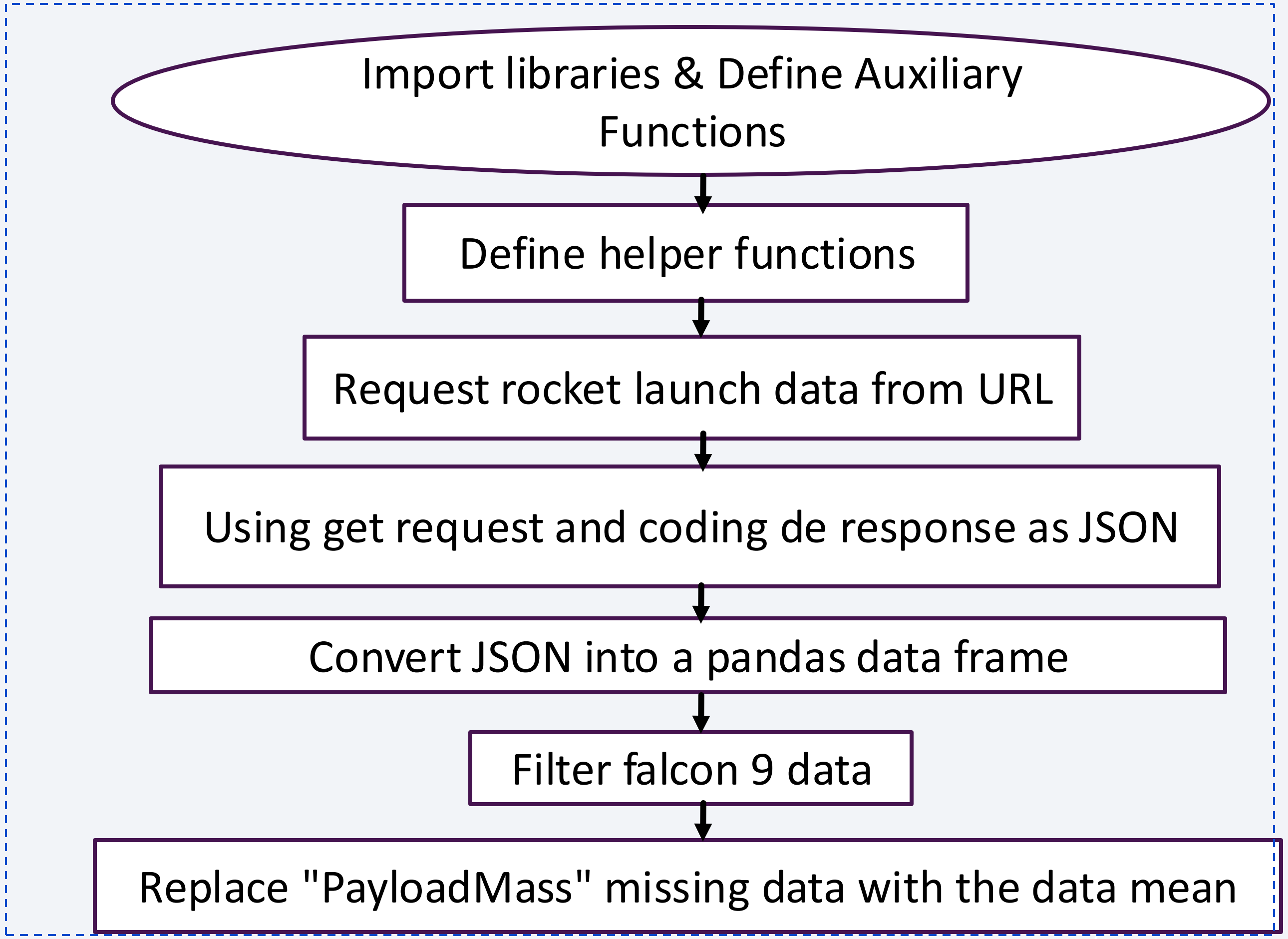
Executive Summary

- Data collection methodology:
  - SpaceX API get request
  - web scraping Falcon 9 and Heavy launches records from wiki
- Perform data wrangling
  - Filter information from Falcon 9 only, change empty data form mass with mass mean
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API

- Data collection with SpaceX REST calls using key phrases and flowcharts SpaceX API calls notebook

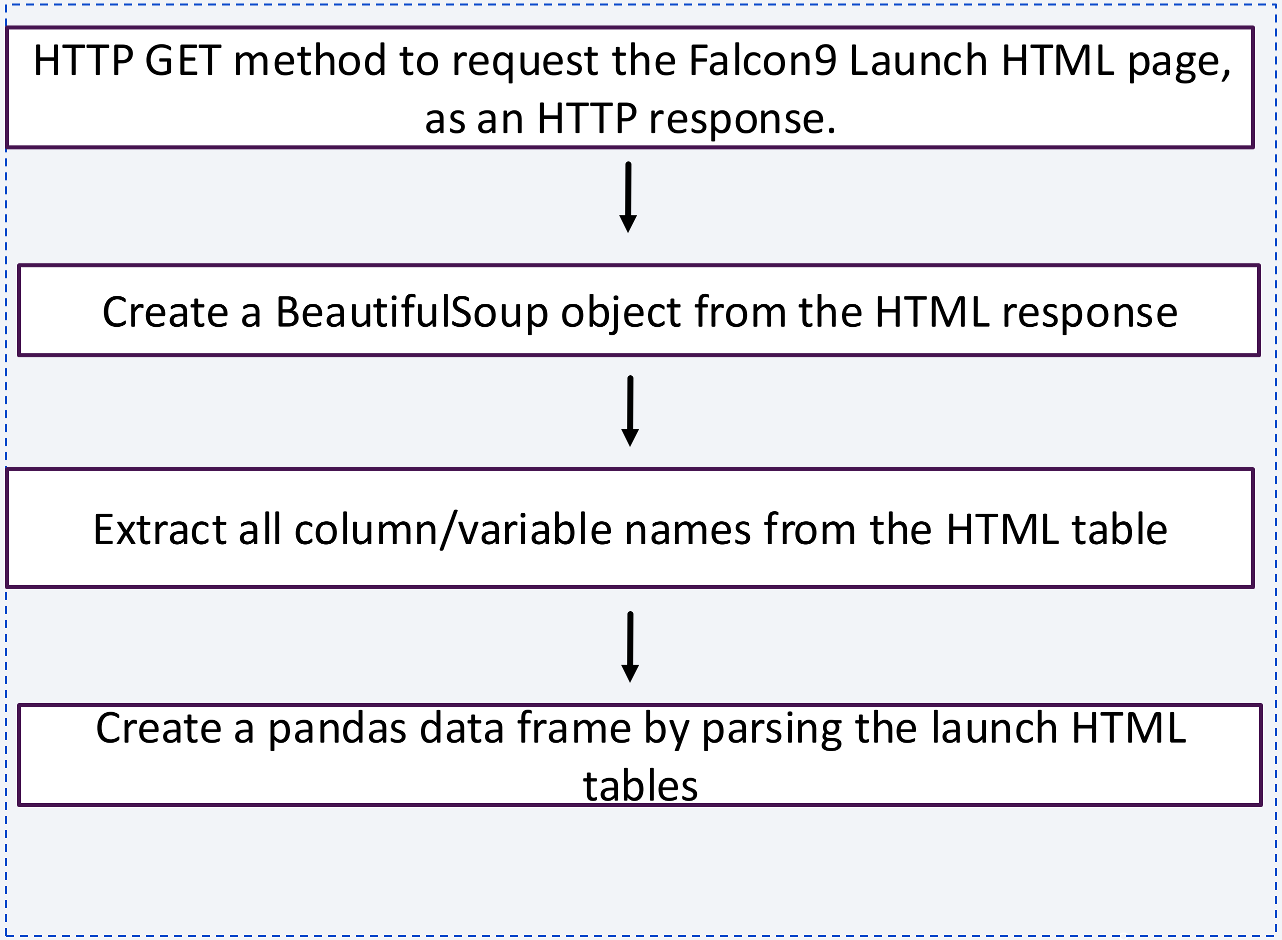
- **GitHub URL of the completed SpaceX API calls notebook:** (https://github.com/Jurexar24/Final-SpaceX-Falcon-9-Success-Landing-Prediction/blob/main/SpaceX%20API%20Data%20collection.ipynb)



Data Collection - Scraping

- Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia. Web scrap Falcon 9 launch records with BeautifulSoup. Extract a Falcon 9 launch records HTML table from Wikipedia

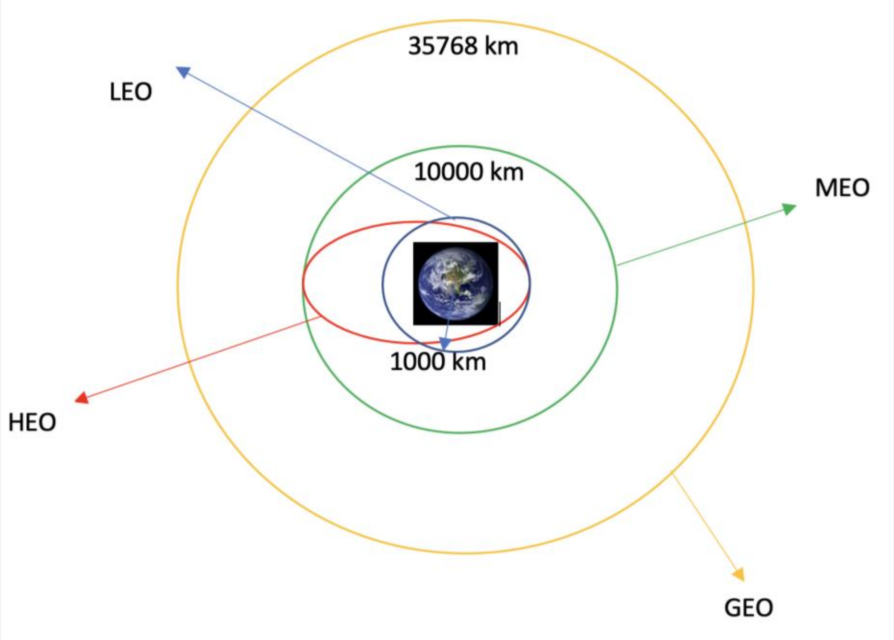
- **GitHub URL of the completed web scraping notebook:** (https://github.com/Jurexar24/Final-SpaceX-Falcon-9-Success-Landing-Prediction/blob/main/2.%20SpaceX%20Web%20scraping%20Wikipedia.ipynb)



Data Wrangling

- From SpaceX dataset data collection. Using the method value_counts() on the column LaunchSite to determine the number of launches on each site, Calculate the number and occurrence of each orbit, Calculate the number and occurence of mission outcome per orbit type. Create a landing outcome label from Outcome column. **We can use the landing outcome label mean to determine the success rate = 67%**

- Add the GitHub URL of your completed data wrangling related notebooks: (https://github.com/Jurexar24/Final-SpaceX-Falcon-9-Success-Landing-Prediction/blob/main/3.%20SpaceX%20Data%20Wrangling%20.ipynb)



	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Bl
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	

EDA with SQL

- **Names of the unique launch sites in the space mission:**
- %sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
- **Total payload mass carried by boosters launched by NASA (CRS):**
- %sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)';
- **Average payload mass carried by booster version F9 v1.1:**
- %sql SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version LIKE 'F9 v1.1%';
- **Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.**
- %sql SELECT * FROM SPACEXTBL WHERE "Landing _Outcome" LIKE 'Success%' AND (Date BETWEEN '04-06-2010' AND '20-03-2017') ORDER BY Date DESC;
- **GitHub URL of completed EDA with data visualization notebook:** (https://github.com/Jurexar24/Final-SpaceX-Falcon-9-Success-Landing-Prediction/blob/main/4.%20SpaceX%20EDA%20SQL.ipynb)

EDA with Data Visualization

- **Scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class.** Help to see the relation between Flight Number in each launch site outcome.
- **Scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hue to be the class value.** We can see the relation between mass and outcome for each launch site.
- **Bar chart relationship between success rate of each orbit type.** Helps to find which orbits have high sucess rate and compare them.
- **Scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hue to be the class value.** Visualize the relationship between FlightNumber and Orbit type on outcomes
- **Scatter point chart with x axis to be Payload and y axis to be the Orbit, and hue to be the class value.** Visualize the relationship between Payload and Orbit type.
- **Line chart with x axis to be Year and y axis to be average success rate, to get the average launch success trend.**
- GitHub URL of completed EDA with data visualization notebook (https://github.com/Jurexar24/Final-SpaceX-Falcon-9-Success-Landing-Prediction/blob/main/5.%20SpaceX%20EDA%20DataV.ipynb)

Build an Interactive Map with Folium

- On the Map it were included a circle mark for the launch sites.
- A color mark for the success rate in each site.
- Also it was calculated the distance to strategic locations such as railways, highways, coastline and cities. It can be observed as a line in the map

- GitHub URL of completed interactive map with Folium map: (https://github.com/Jurexar24/Final-SpaceX-Falcon-9-Success-Landing-Prediction/blob/main/6.SpaceX%20Launch%20Sites%20Folium.ipynb)

Build a Dashboard with Plotly Dash

- Add a Launch Site Drop-down Input Component
- Add a callback function to render success-pie-chart based on selected site dropdown
- Add a Range Slider to Select Payload
- Add a callback function to render the success-payload-scatter-chart scatter plot

- GitHub URL of completed Plotly Dash lab: (https://github.com/Jurexar24/Final-SpaceX-Falcon-9-Success-Landing-Prediction/blob/main/7.%20Interactive%20Dashboard.py)

GitHub URL of completed predictive analysis lab: (https://github.com/Jurexar24/Final-SpaceX-Falcon-9-Success-Landing-Prediction/blob/main/8.%20SpaceX%20ML%20Model.ipynb)

Predictive Analysis(logisticR,SVM,decision tree, KNN)

Import libraries & function to plot the confusion matrix

↓

Create a NumPy array from the column Class in data, by applying the method to_numpy() then assign it to the variable Y

↓

Standardize the data in X then reassign it to the variable X

↓

Use the function train_test_split to split the data X and Y into training and test data. Set the parameter test_size to 0.2 and random_state to 2

↓

Repeat (1,2,3) per each model

1.Create a Model** object then create a GridSearchCV object **_cv with cv = 10

↓

2.Calculate the accuracy on the test data using the method score

↓

3.Look at the confusion matrix

↓

Find the method performs best

Results

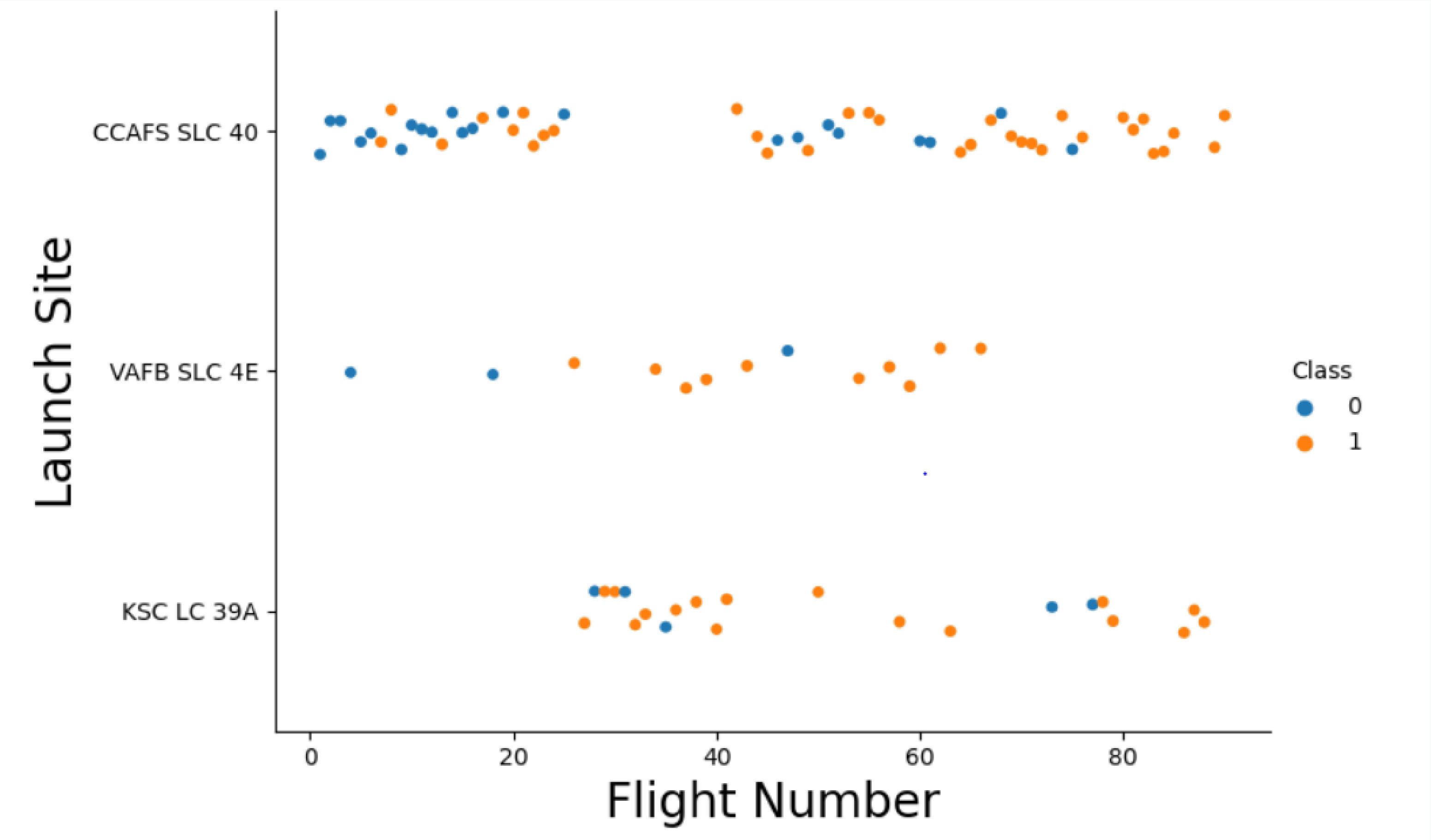
- Exploratory data analysis results:
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

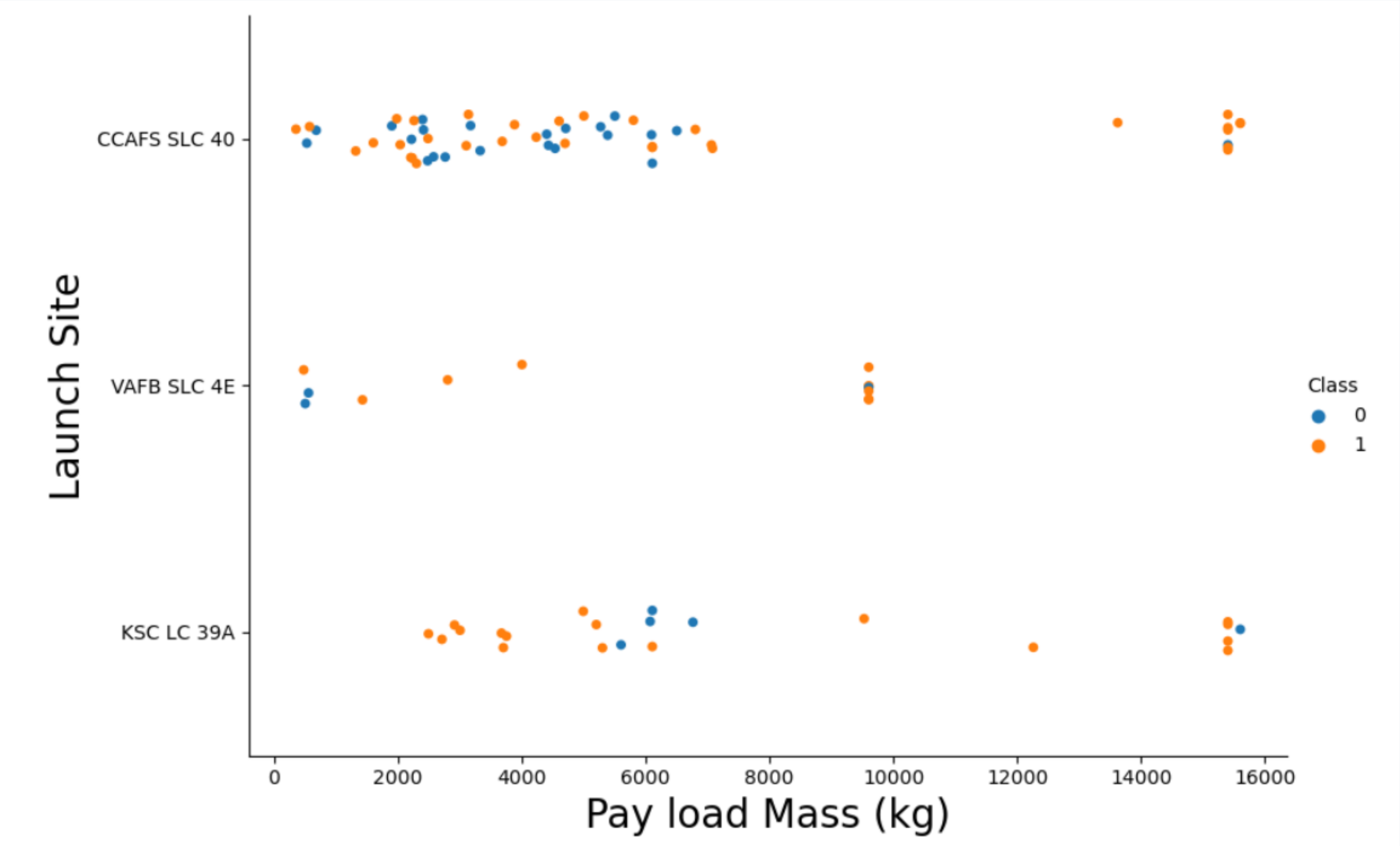
Flight Number vs. Launch Site

- As the flight number increases in each of the 3 launch sites, so does the success rate. Both KSC LC 39A and CCAFS SLC 40 have a 100% success rates after 80th flight.



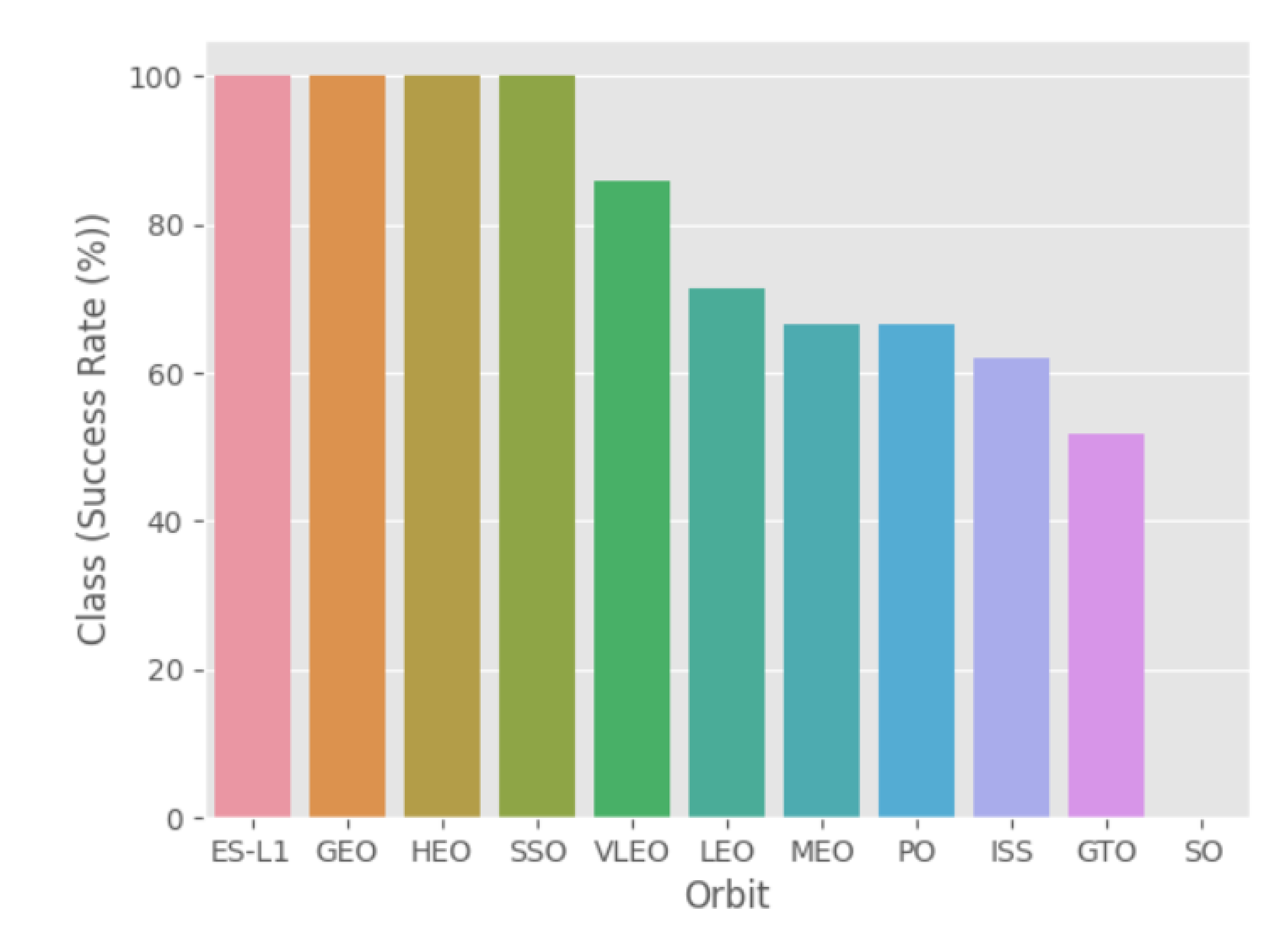
Payload vs. Launch Site

- If you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).
- KSCLC39A show a 100% success rate for payload below (5000)



Success Rate vs. Orbit Type

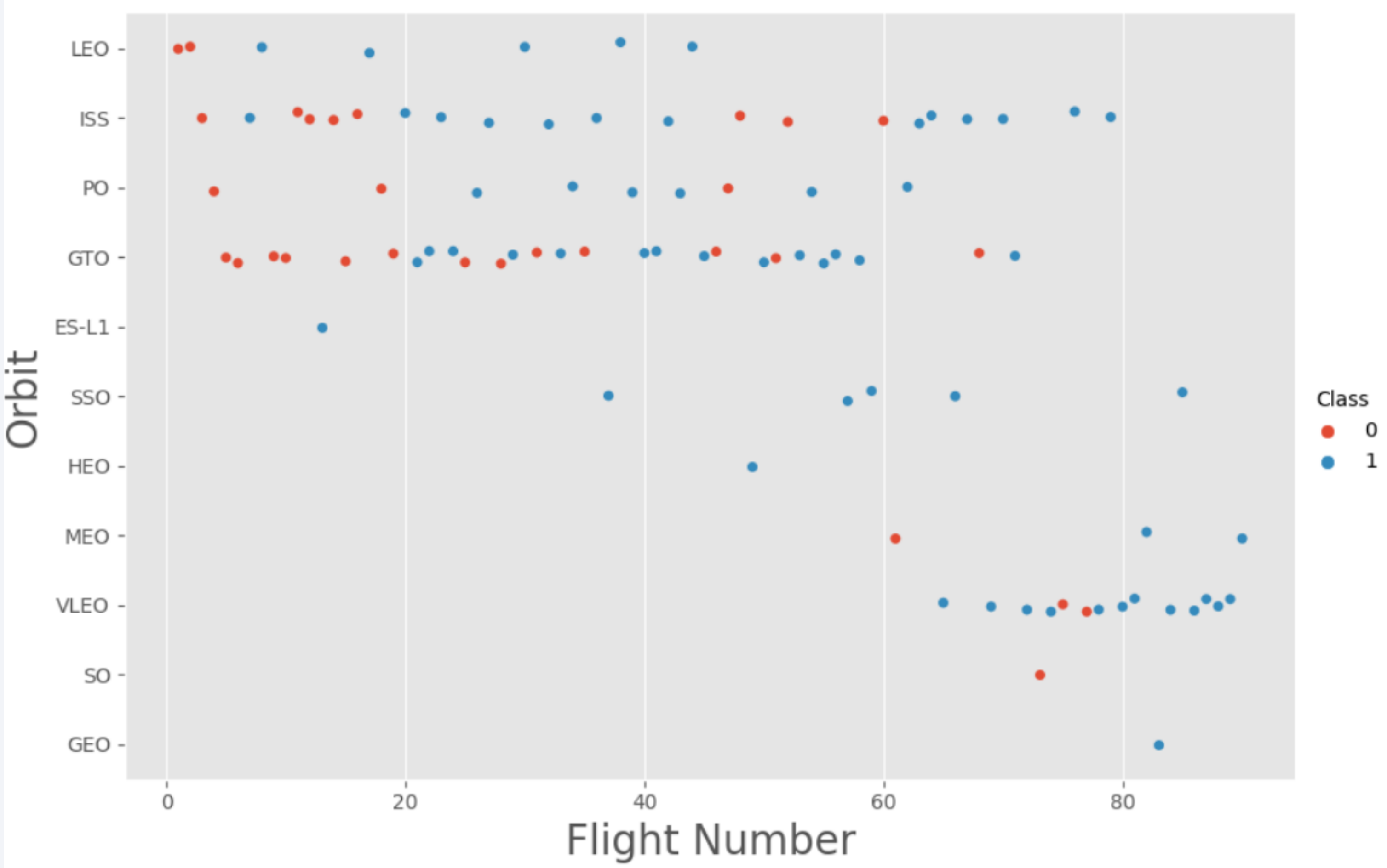
- From this bar chart we can compare the success rate for each orbit type.
- Orbits ES-L1, GEO, HEO & SSO have the highest success rates at 100%, with GTO orbit having almost the lowest success rate at ~50%. Orbit SO has 0% success rate.



Flight Number vs. Orbit Type

- We can observe the flight number have a high impact on outcome by orbit type. Specially after 60 the Success increased and above 80 there is a 100% success.

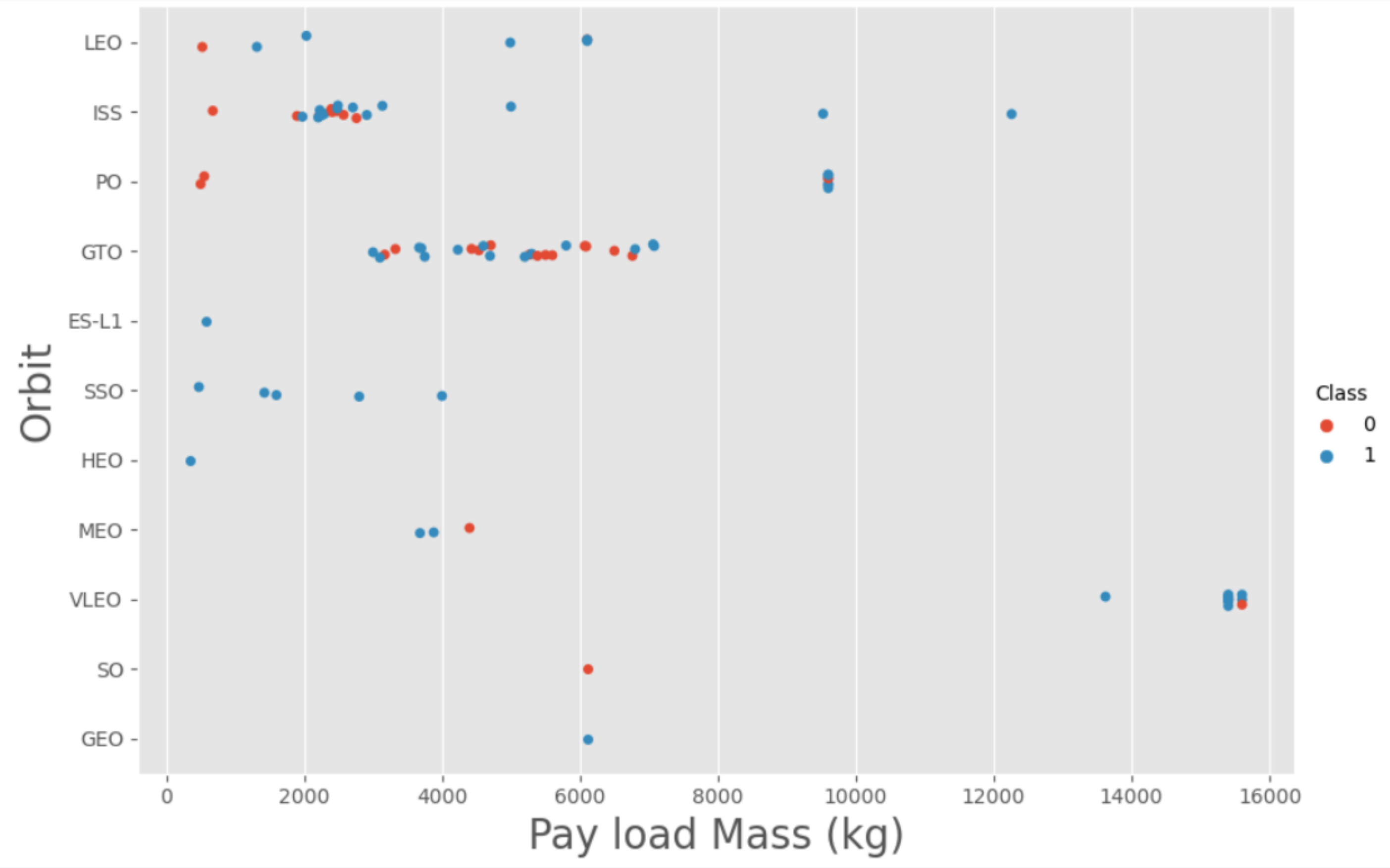
- You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



Payload vs. Orbit Type

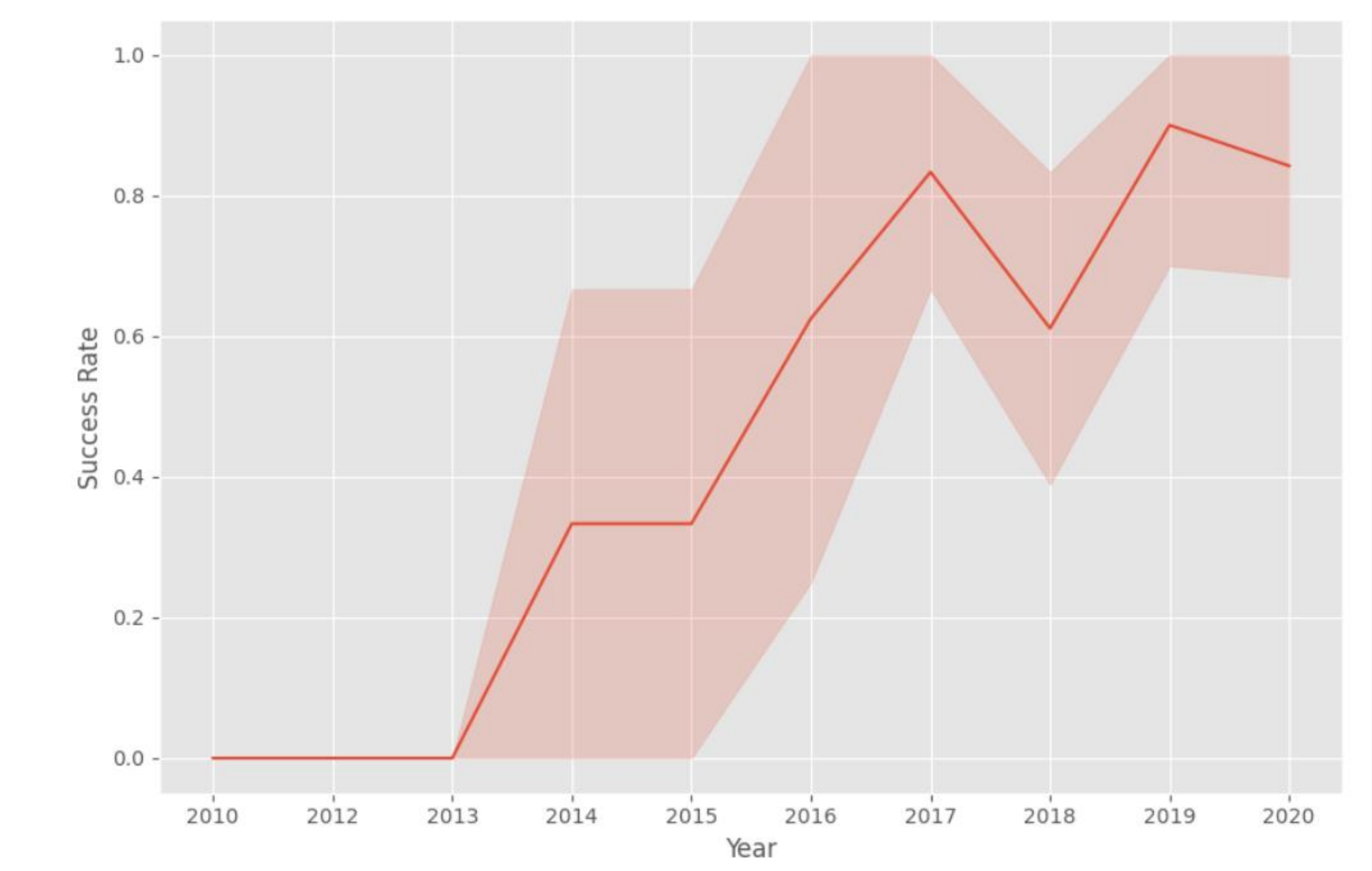
- This plot can show the relation between pay load mass and orbit on success. We can observe some Orbits with better performance at low Pay load mass in some cases is difucult to see a patron cause there aren't enough data to make a conclusion as SO,GEO,GTO.

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.



Launch Success Yearly Trend

- We can see a line chart of yearly average success rate
- You can observe that the success rate since 2013 kept increasing till 2020



All Launch Site Names

- %sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
- Display the names of the unique launch sites in the space mission

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

Launch Site Names Begin with 'CCA'

- %sql SELECT * FROM 'SPACEXTBL' WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
- Display 5 records where launch sites begin with the string '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

- %sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM 'SPACEXTBL' WHERE Customer = 'NASA (CRS)';

- Display the total payload mass carried by boosters launched by NASA (CRS)

Total Payload Mass(Kgs)	Customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

- %sql SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Version FROM 'SPACEXTBL' WHERE Booster_Version LIKE 'F9 v1.1%';
- Display average payload mass carried by booster version F9 v1.1

Payload Mass Kgs	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

First Successful Ground Landing Date

- %sql SELECT MIN(DATE) FROM 'SPACEXTBL' WHERE "Landing _Outcome" = "Success (ground pad)";

- Date when the first successful landing outcome in ground pad was achieved.

MIN(DATE)
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql SELECT DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes
- %sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- %sql SELECT "Booster_Version",Payload, "PAYLOAD_MASS__KG_" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- %sql SELECT substr(Date,7,4), substr(Date, 4, 2),"Booster_Version", "Launch_Site", Payload, "PAYLOAD_MASS__KG_", "Mission_Outcome", "Landing _Outcome" FROM SPACEXTBL WHERE substr(Date,7,4)='2015' AND "Landing _Outcome" = 'Failure (drone ship)';

substr(Date,7,4)	substr(Date, 4, 2)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Mission_Outcome	Landing _Outcome
2015	01	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	Success	Failure (drone ship)
2015	04	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	Success	Failure (drone ship)

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

- 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

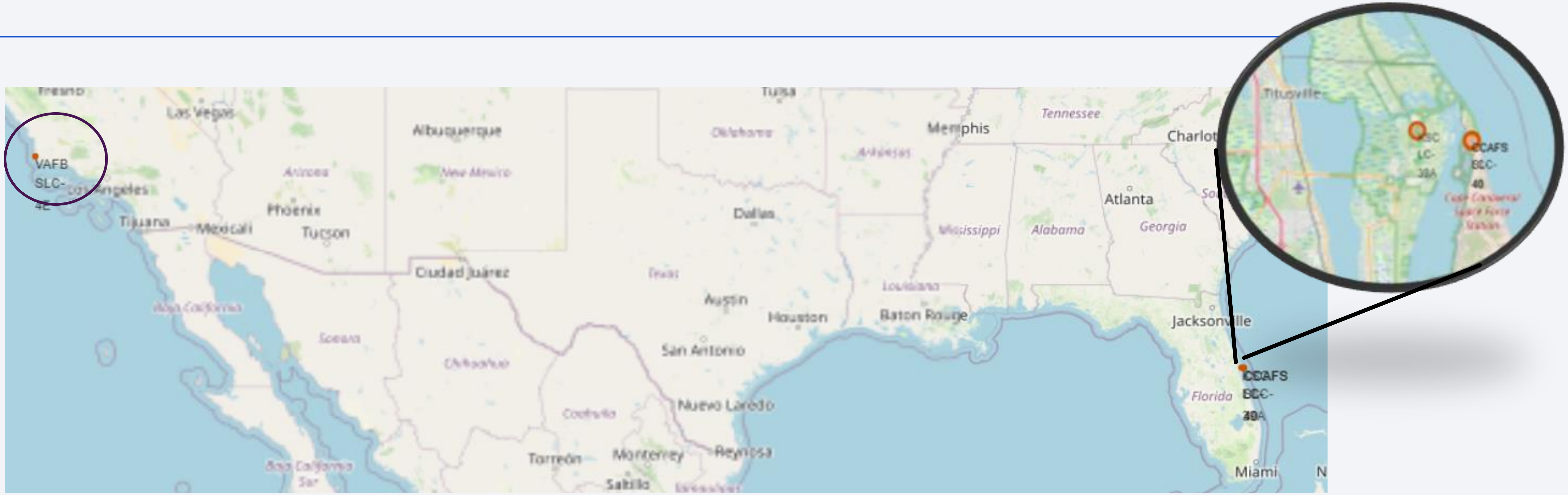
- %sql SELECT * FROM SPACEXTBL WHERE "Landing _Outcome" LIKE 'Success%' AND (Date BETWEEN '04-06-2010' AND '20-03-2017') ORDER BY Date DESC;

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-10-2020	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
18-08-2020	14:31:00	F9 B5 B1049.6	CCAFS SLC-40	Starlink 10 v1.0, SkySat-19, -20, -21, SAOCOM 1B	15440	LEO	SpaceX, Planet Labs, PlanetIQ	Success	Success
18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
18-04-2018	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362	HEO	NASA (LSP)	Success	Success (drone ship)
17-12-2019	00:10:00	F9 B5 B1056.3	CCAFS SLC-40	JCSat-18 / Kacific 1, Starlink 2 v1.0	6956	GTO	Sky Perfect JSAT, Kacific 1	Success	Success
16-11-2020	00:27:00	F9 B5B1061.1	KSC LC-39A	Crew-1, Sentinel-6 Michael Freilich	12500	LEO (ISS)	NASA (CCP)	Success	Success
15-12-2017	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)

Section 3

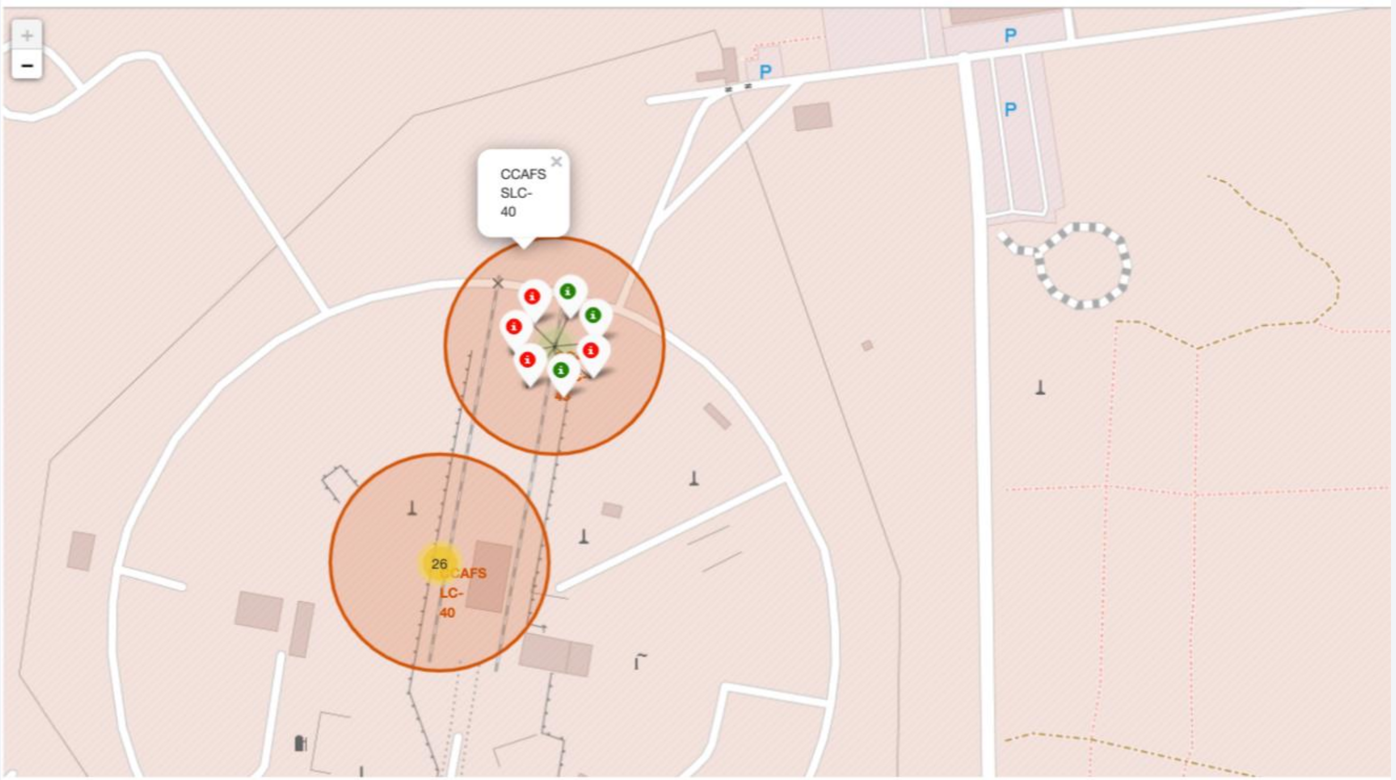
Launch Sites Proximities Analysis

Map with marked launch sites



- Launch sites are near the Ecuador line it can be cause the more stable weather. Also they are near the sea because of security they need near water spaces in the coast for landing and provably with reasonable distance from cities and towns.

Color marked success/failed launches for each site on the map



- From the color-labeled markers in marker clusters, we should be able to easily identify which launch sites have relatively high success rates.

Launch site to its proximities

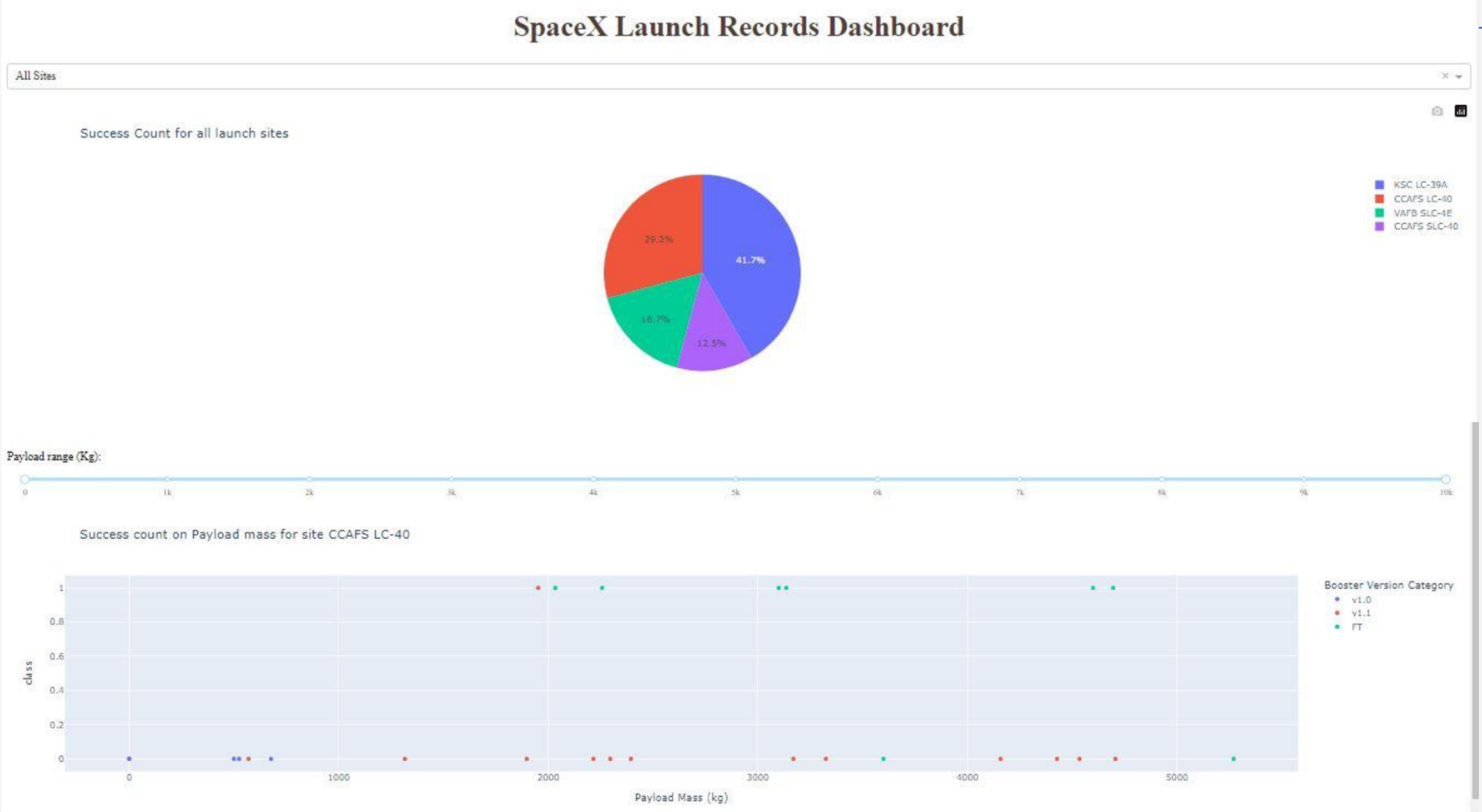
- We can confirm that the launch sites are near railway, highway, and coastline. But they are far from cities

Section 4

Build a Dashboard with Plotly Dash

Dash APP Launch records

SpaceX Dash App



Section 5

Predictive Analysis (Classification)

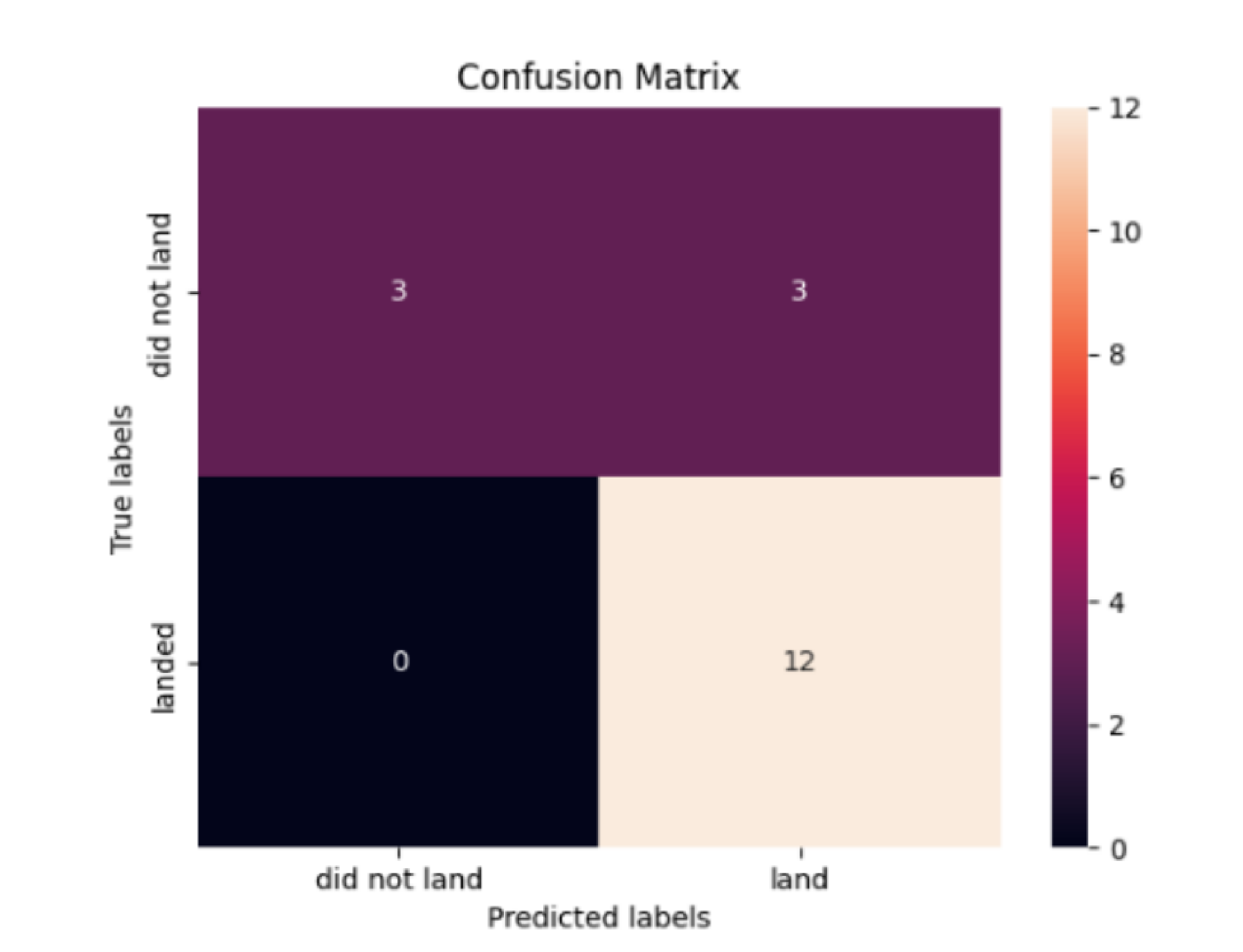
Classification Accuracy

Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

All the methods perform equally on the test data: i.e. They all have the same accuracy of 0.833333 on the test Data

Confusion Matrix

- All 4 model gave de same confusion matrix result. The matrix show that the model has 100% accuracy on not landed prediction but there were 3 mistakes on land prediction that were fail in real data from the test group.



Conclusions

- Point 1: On the initial analysis it can be observed a general success rate of 67% from the real data.
- Point 2: Different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%
- Point 3: There is a clear relation between the outcome and variables like: flight number, payload mass, launch site and orbit. It can be observed on the exploratory data analysis using SQL and the visual analytics and dashboards presented on this project.
- Point 4: On the predictive analysis It was discovered that all 4 models (Logistic regression, SVM, decision tree, and KNN had the same accuracy on the test data = 83% also the same confusion matrix.
- For now we can use any of the 4 models to predict the outcome of a next mission with variables related in this project. It is recommend to repeat the 4 models evaluation by the time there are new data to compare.

Thank you!