IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alejandro Larios
September 09, 2023

Outline

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

Executive Summary

- Predict the Falcon 9 launch outcome of a booster require different types of methodologies. In this project we use, Data Wrangling, Exploratory Data Analysis (EDA), Visual Analytics and Perform predictive analysis using classification models
- The accuracy of each model is 83.33%, this result give us a high performance.

Introduction

The space race has several competitors, the problem to solve is the high cost of different parts of a booster. SpaceX has an idea to save money that consist in reusing the Stage 1 booster.

With all the data available the SpaceX page and with the web data about the previous launches we try to predict the Falcon 9 launch outcome for the next mission

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Using SpaceX API and Web Scrap
- Perform data wrangling
  - Treat with missing values, discarding noise rows and columns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Accuracy model comparison

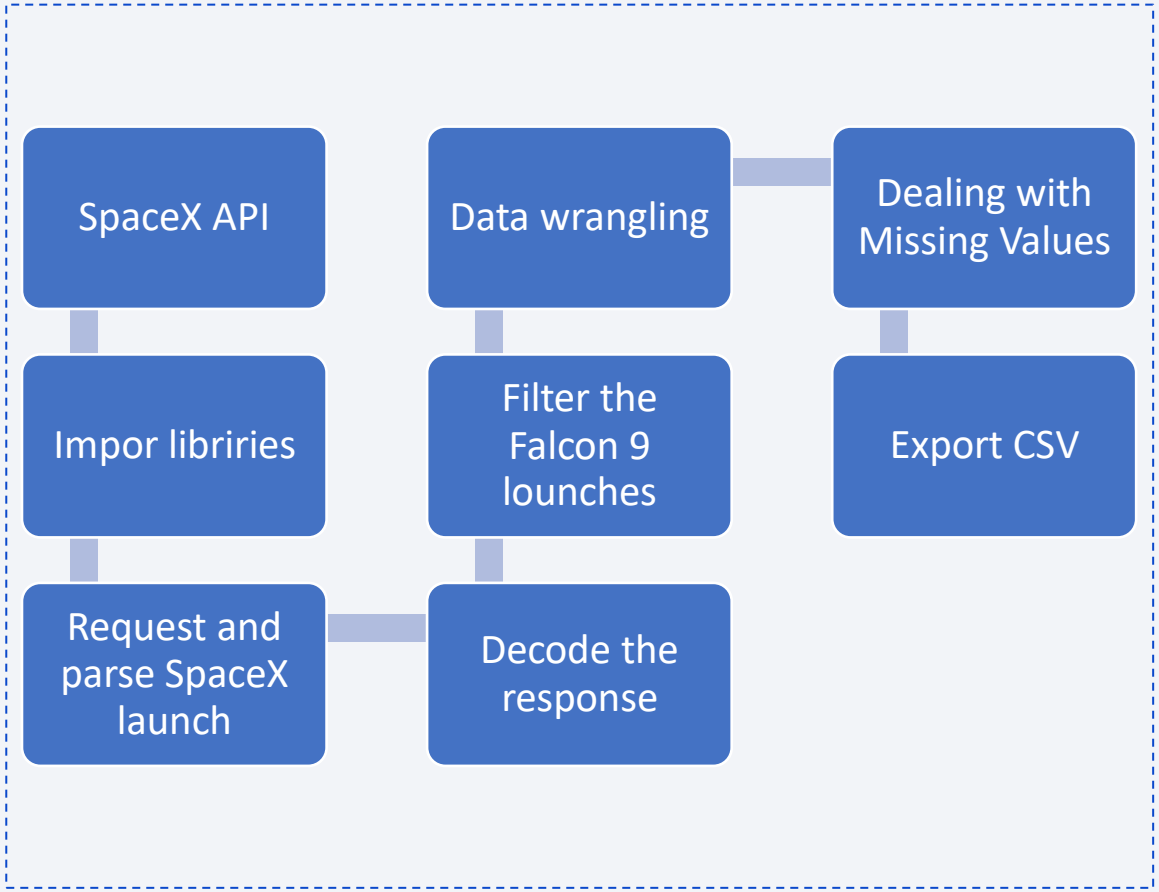
Data Collection

- Data had been collected from different sources, first one, we had made a get request to the SapceX API and second one, we made a web scrap to get information. The raw information was treated to obtain our clear information

Data Collection – SpaceX API

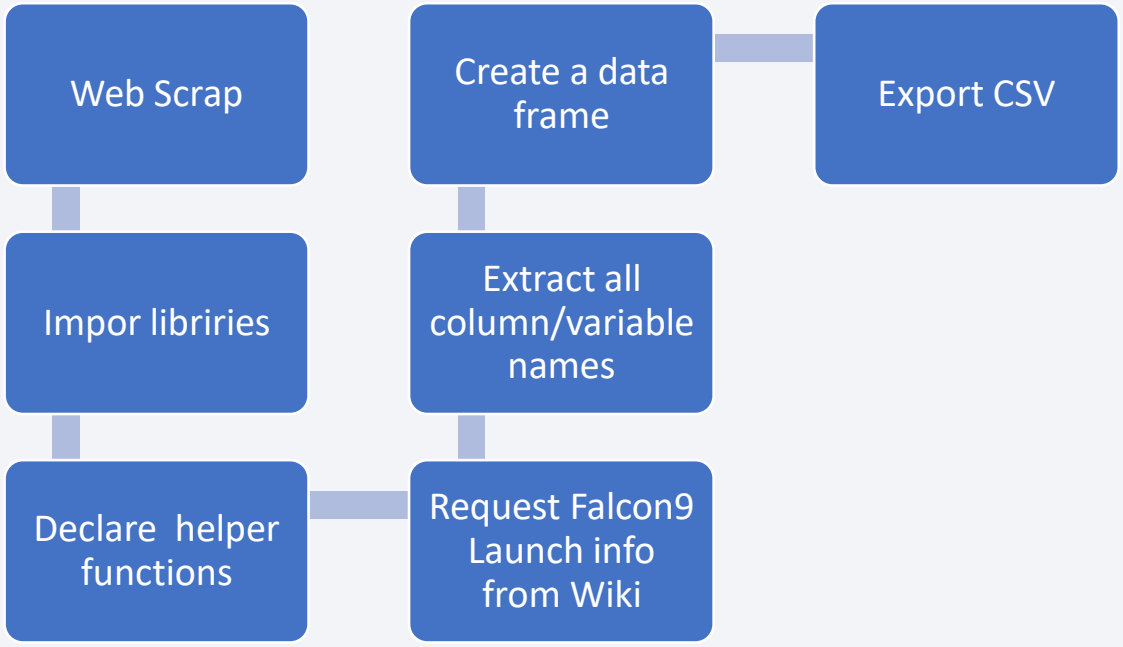
- For SpaceX API we use:
  - Respond to get the content
  - Use Json to normalize the data
  - Data wrangling

- https://github.com/AlejandroLariosC/Data-Collection.git



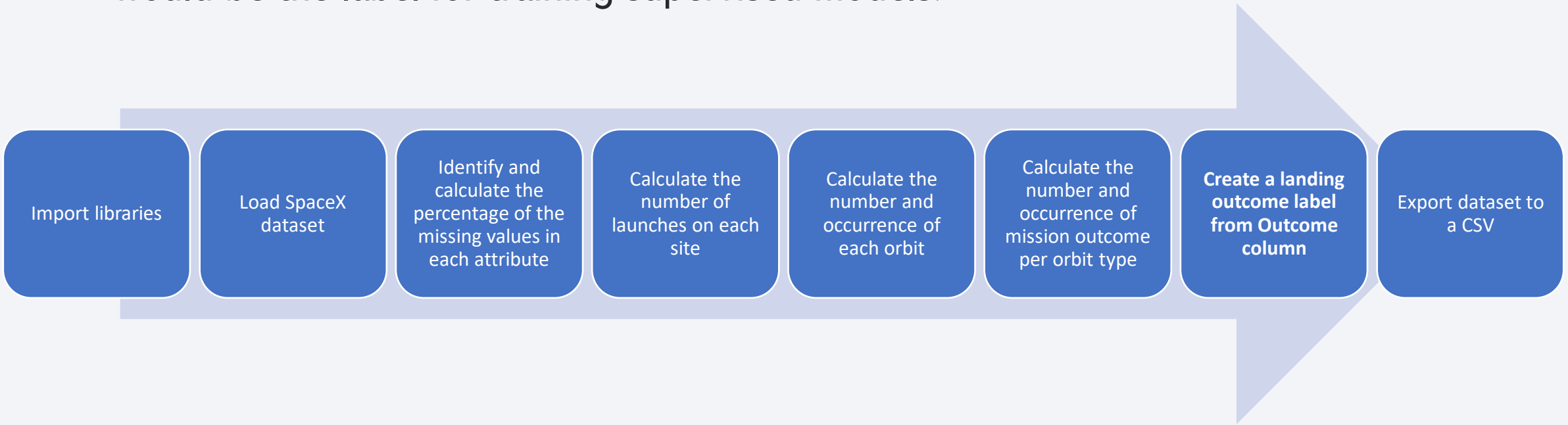
Data Collection - Scraping

- Web scrap consist of extracting information from a specific we page. In this case we use Wikipedia
- https://github.com/AlejandroLariosC/Web-scraping.git

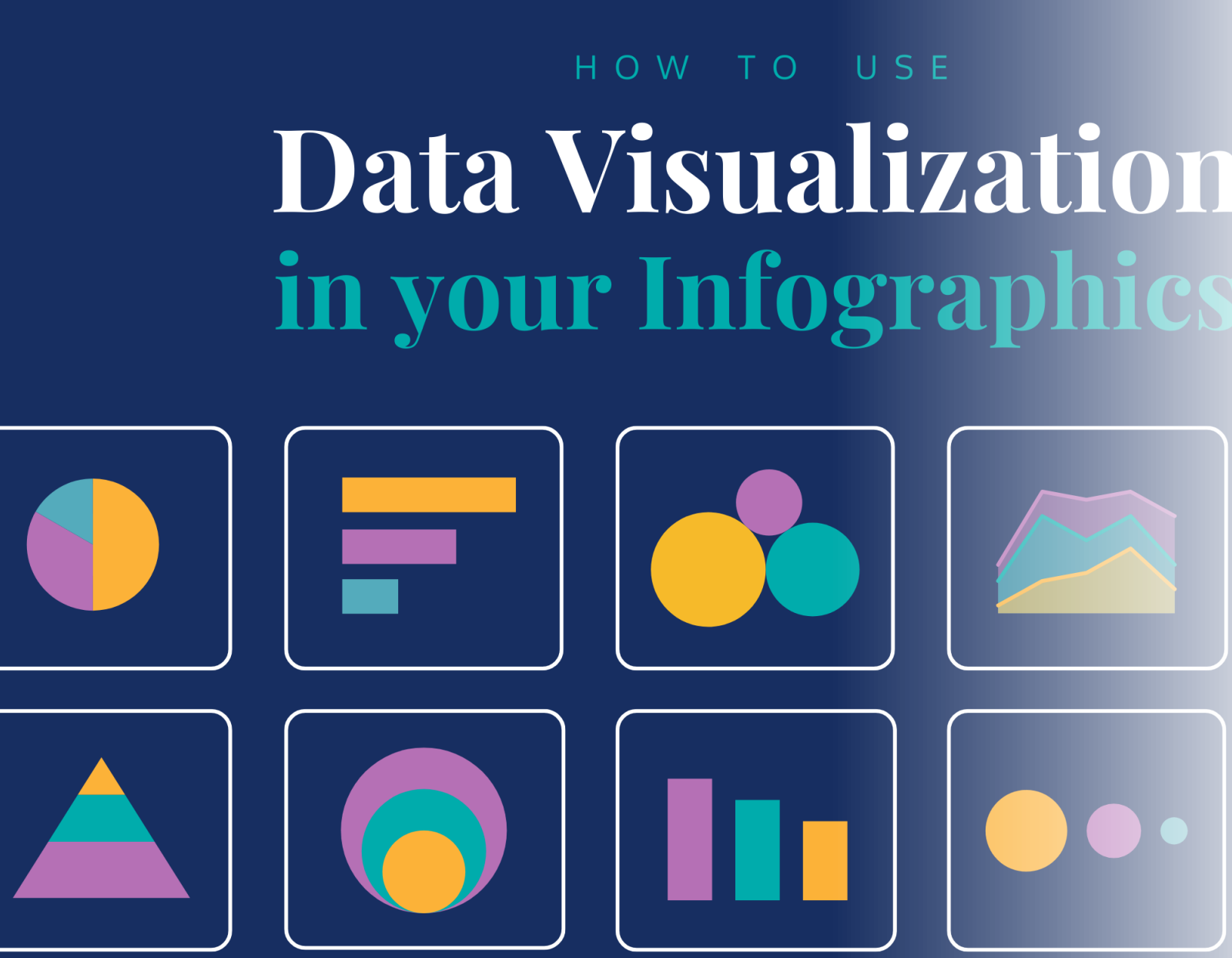


Data Wrangling

- Exploratory Data Analysis consist to find patterns in the data and determine what would be the label for training supervised models.



- https://github.com/AlejandroLariosC/Data-Wrangling.git



EDA with Data Visualization

- The data obtained, trough the SpaceX API and the web data, has a lot of information that we don't use it. In this section we compare different variables that it can show us insights about the behavior of the data like Pay load mass, Booster version, orbit, Launch site, launch outcome.

- https://github.com/AlejandroLariosC/EDA.git

EDA with SQL

- SQL queries give us general information with different conditions. Some of the queries that we apply are listed:
  - Flight Number vs Pay Load Mass. I could see some relationship between this variables
  - Launch site vs Flight number. This could show us if there is a better launch place.
  - Orbit vs Launch Outcome. This relationship could show us the performance according the orbit.
  - Orbit vs Flight number. The performance between both
  - Year vs Success /Failure. This is the more important relationship due to show us the general performance trough the time
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

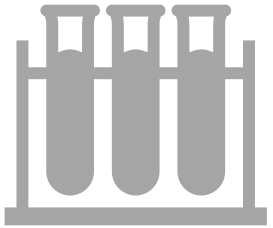
THE FOLIUM MAP SHOW THE GENERAL DISTRIBUTION OF THE LAUNCH SITES AND THE GRAPIC LAUNCH OUTCOME FOR BETTER VISUALIZATION

THE LAUNCH SITE IS A STRATEGIC SITE BECAUSE YOU NEED TO TRANSPORT DIFFERENT PARTS OF EACH BOOSTER AND ALSO YOU HAVE TO RETURN THE STAGE 1 LANDED, SO THE RAILWAY AND COASTLINE IS CRITICAL

HTTPS://GITHUB.COM/ALEJANDROLARIOSC/FOLIUM.GIT

Build a Dashboard with Plotly Dash

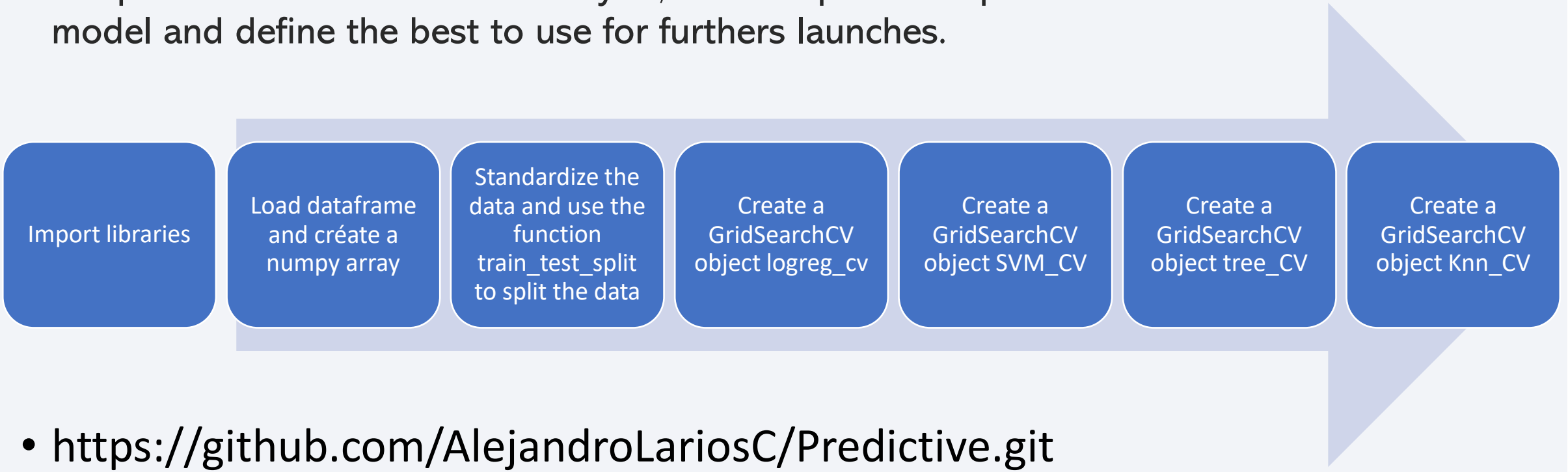
The dashboard tool help us to interact with the data, In this case, we create two graphs. The first one is about the general launch outcome and individual launch site outcome. The second one, is an interact scatter plot for manipulate the Play Load mass vs lunch sites and its outcome.



Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

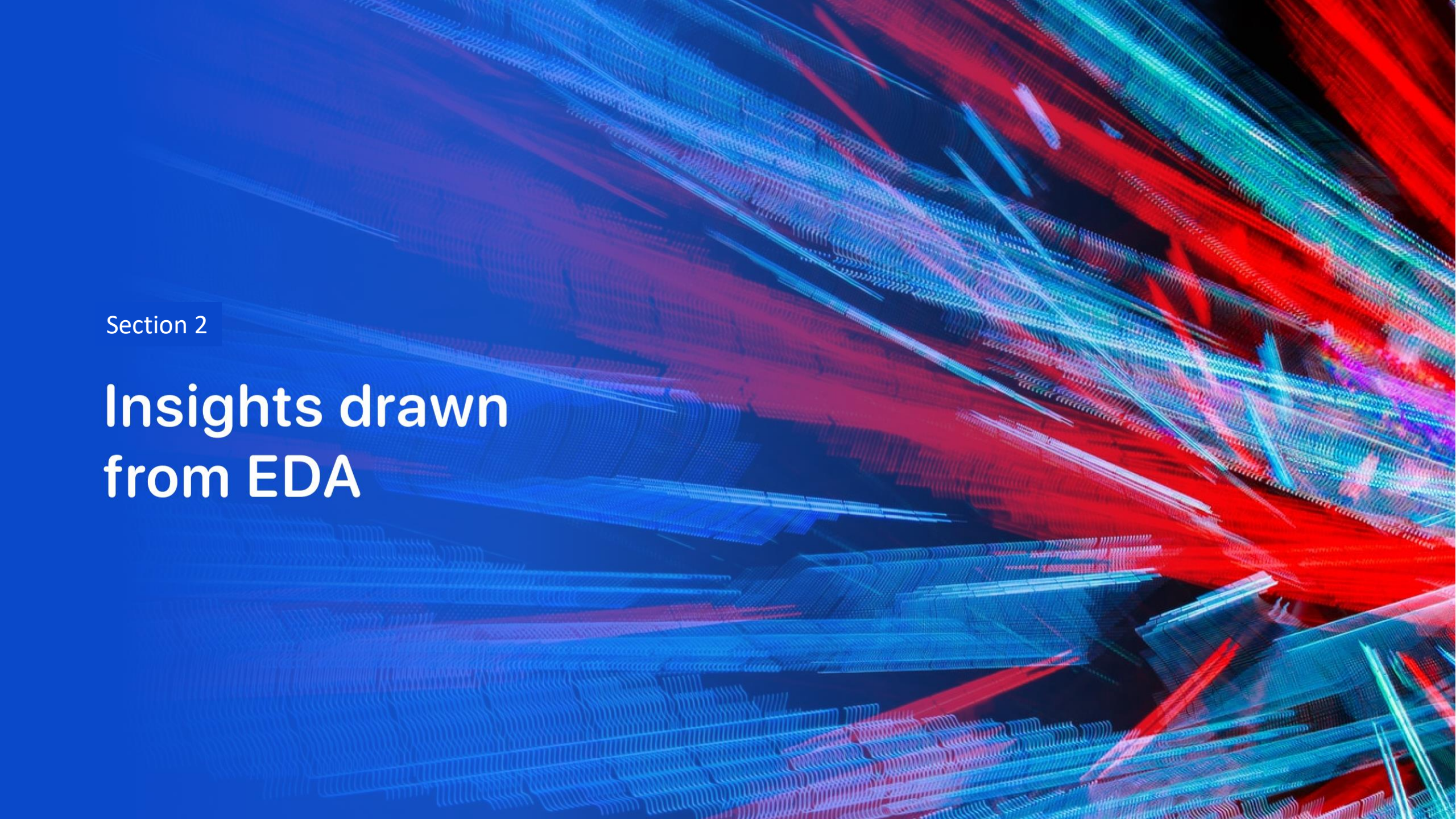
- Predictive analysis is crucial in the project evaluation. We run different models and the we evaluate its performance according with the test and train samples. In the end of the analysis, we compare the performance of each model and define the best to use for furthers launches.



- https://github.com/AlejandroLariosC/Predictive.git

Results

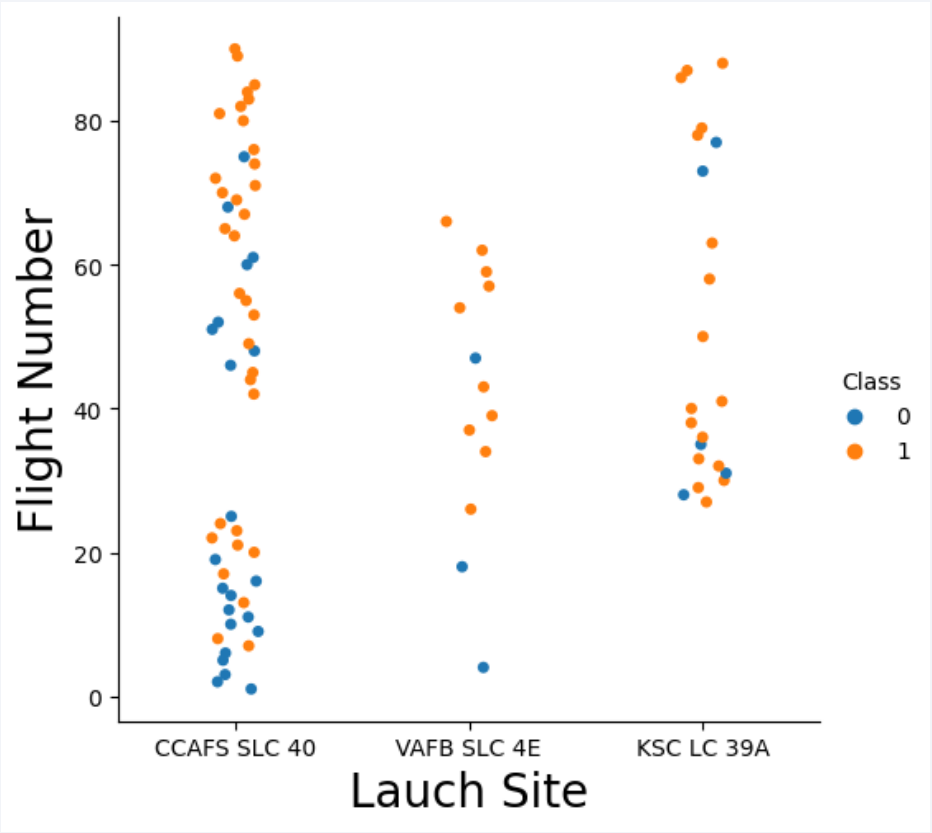
- After the data wrangling was complete, we can say that there is a relationships between variables. This point is essential because we expect a high performance during the predict models' performance


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site

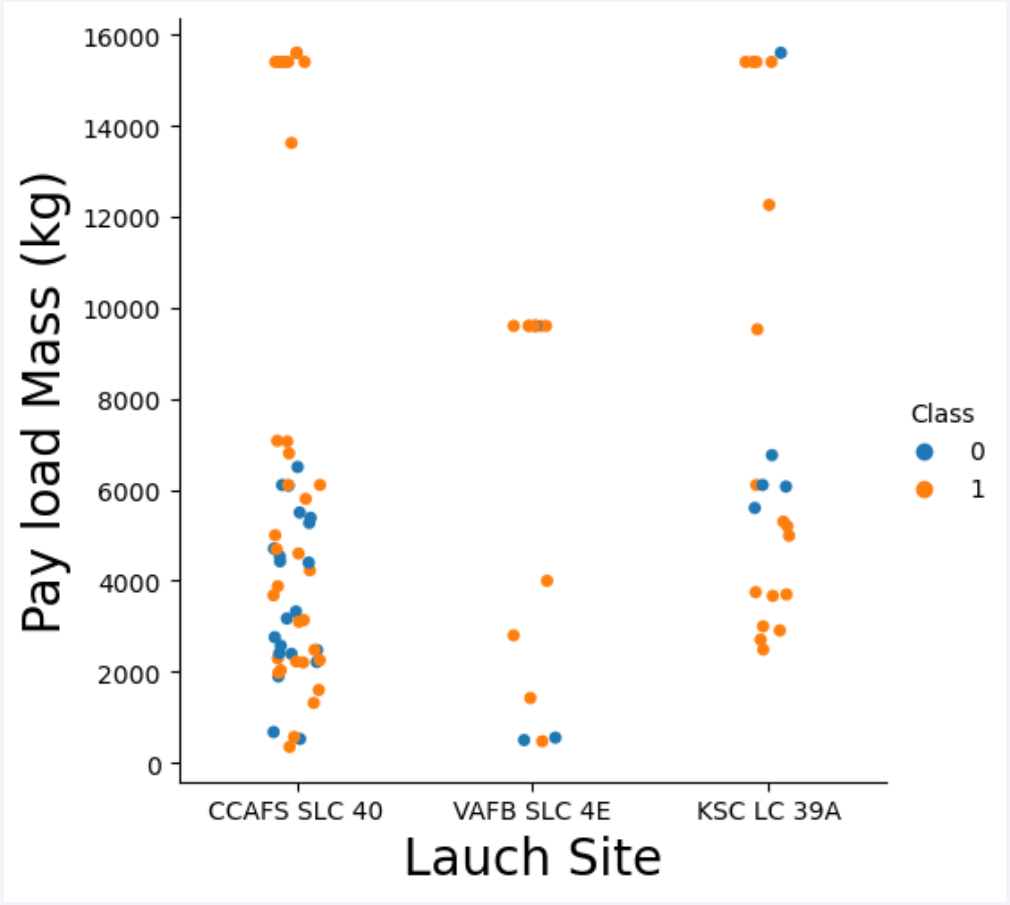
- As we can see, VAFB SLC 4E has the lowest flight number on the contrary, CCAFS SLC 40 has the highest flight number



Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

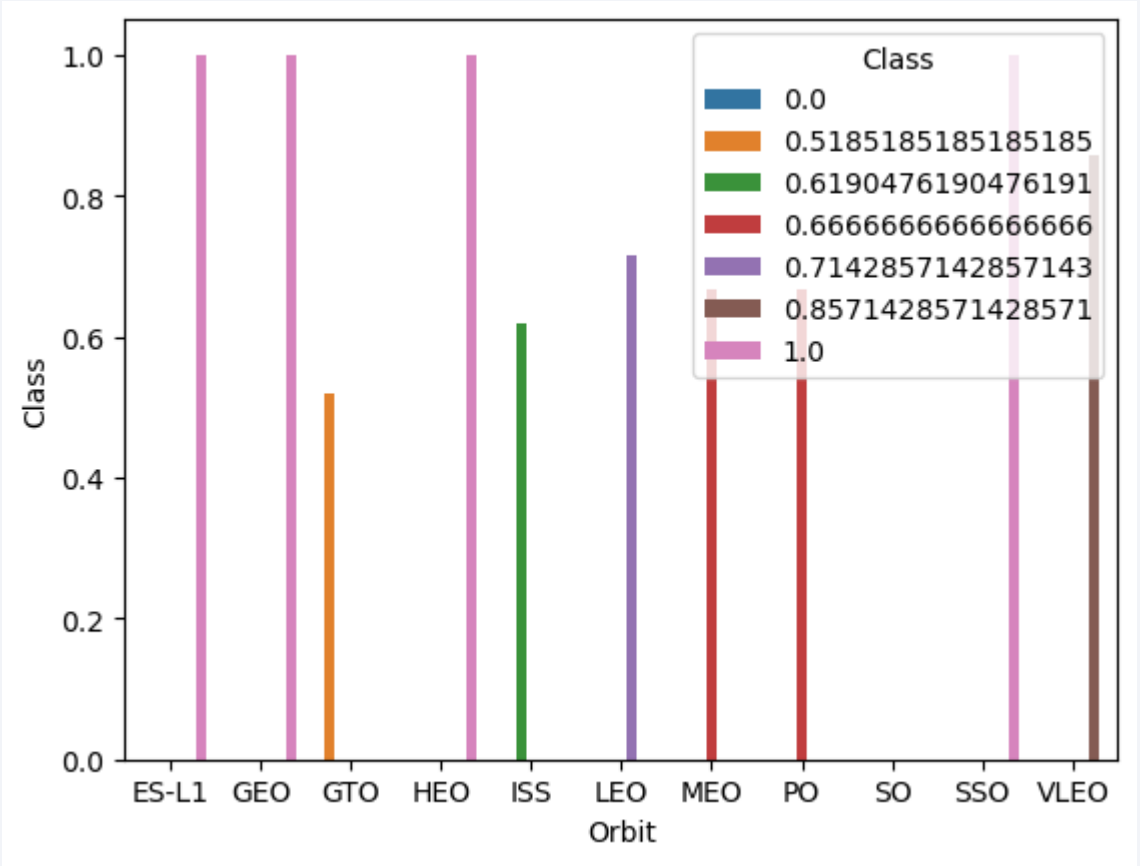
- As we can see in the graph, VAFB SLC 4E launch site can transport until 10,000 kg pay load mass, in CCAFS-SLC-40 launch site has 100% success rate transporting >120,000 kg pay load mass



Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

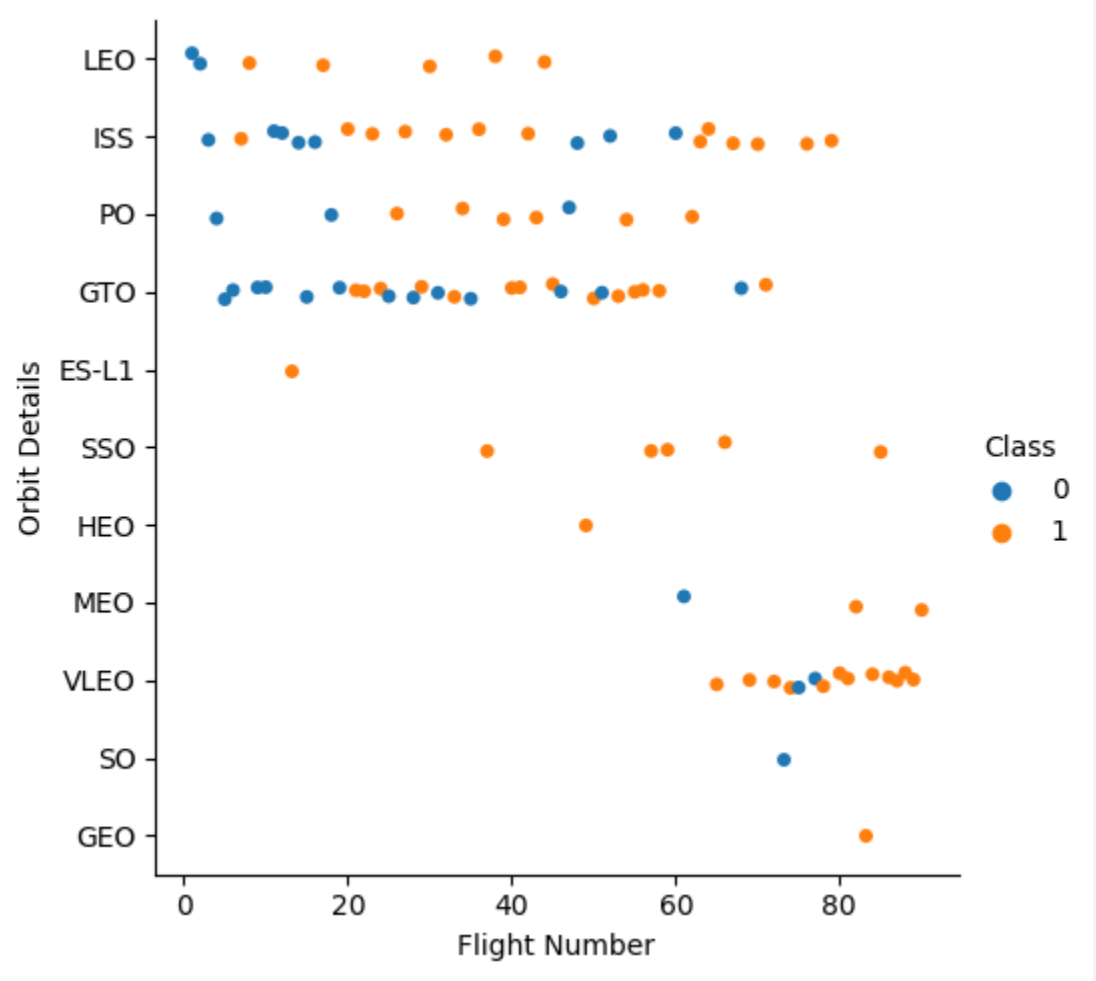
- ES-L1, GEO, HEO and SSO orbits has 100% success rate. GTO orbit has the lowest success rate with only 51.85%



Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type

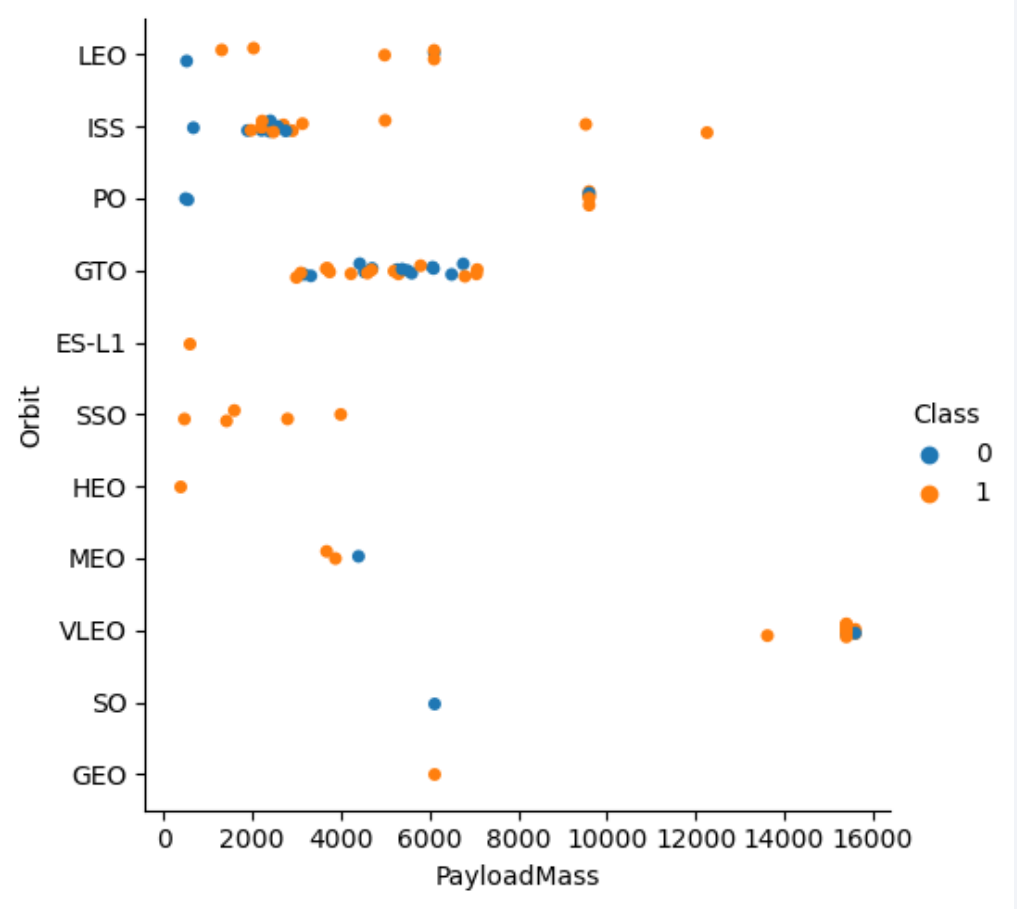
- LEO orbits success is related with the flight number, on the contrary, GTO orbit success has not related with the flight number.



Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

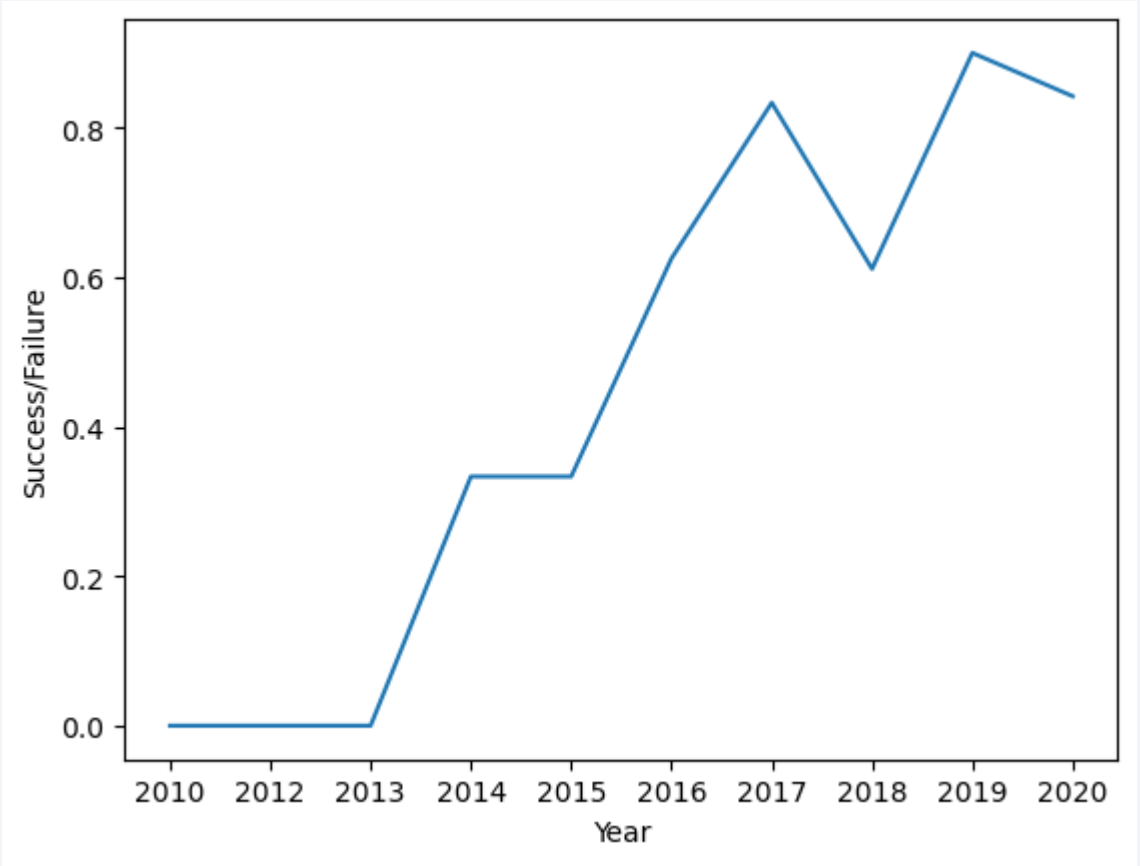
- LEO and ISS orbit success is related with the Pay Load mass while GTO has no related with the same criteria



Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- As we cans see in the graph, the Falcon 9 launch success rate has increasing tend since 2013 to 2020



All Launch Site Names

- Find the names of the unique launch sites
    1. CCAFS LC-40
    2. VAFB SLC-4E
    3. KSC LC-39[a]
    4. CCAFS SLC-40
- Present your query result with a short explanation here
    - %sql select DIS“%” is sql select” magic function to call a SQL query
        - Distinct (Launch Site) is our parameter from the column “Launch Site”
        - “SPACEXTABLEthe TINCT (Launch_Site) from SPACEXTABLE
        - name of the table where it is”Launch Site” column

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites 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

- Present your query result with a short explanation here

I. %sql select * from SPACEXTABLE where Launch_Site LIKE '%CCA%' LIMIT 5

We use a query where, we select fetch the string “CCA” in the “Launch Site” column. Finally, we limited the result at 5 rows

Total Payload Mass

- Calculate the total payload carried by boosters by NASA (CRS)

SUM (PAYLOAD_MASS__KG_)
45596

- Present your query result with a short explanation here

  %sql select SUM (PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Customer like 'NASA (CRS)'

  This is an easy query. We use the function “SUM” the column PAYLOAD_MASS_KG), where the customer is: NASA (CRS) from SPACEXTABLE table

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

AVG (PAYLOAD_MASS__KG_)
2534.6666666666665

- Present your query result with a short explanation here

  %sql select AVG (PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Booster _Version like '%F9 v1.1%'

  We use the Average function for the Pay_Load_Mass_kg column where the booster version is F9 v1.1

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

- Present your query result with a short explanation here

  %sql select * from SPACEXTABLE where Mission_Outcome = 'Success' and Landing_Outcome like'%ground pad%' order by Date limit 1

  In this query, we select all the columns from SPACEXTABLE where the Mission Outcome where “success” and the Landing outcome where “ground pad”. Finally, data is order by the “Date” column and we limit the result to the first one.

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version	PAYLOAD_MASS__KG_	Mission_Outcome	Landing_Outcome
F9 FT B1020	5271	Success	Failure (drone ship)
F9 FT B1022	4696	Success	Success (drone ship)
F9 FT B1026	4600	Success	Success (drone ship)
F9 FT B1021.2	5300	Success	Success (drone ship)
F9 FT B1031.2	5200	Success	Success (drone ship)

- Present your query result with a short explanation here
  - %sql select Booster_Version, PAYLOAD_MASS__KG_, Mission_Outcome, Landing_Outcome from SPACEXTABLE where Mission_Outcome = 'Success' and Landing_Outcome like'%drone ship%' and PAYLOAD_MASS__KG_>4000 and PAYLOAD_MASS__KG_<6000
  - For this query, we select Booster_Version, PAYLOAD_MASS__KG_, Mission_Outcome, Landing_Outcome columns from the SPACEXTABLE where Mission outcome where success, Landing Outcome where “Drone Ship” and the PayLoad is between 4,,000 and 6,000 kg

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

count (Mission_Outcome)
101

- Present your query result with a short explanation here

  %sql select count (Mission_Outcome) from SPACEXTABLE where Mission_Outcome like '%Success%' or Mission_Outcome like '%Failure %'

  In this query, we use the function “Count” the number of “Mission Outcome” where the result where “Success” or “Failure”

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2019-11-11	14:56:00	F9 B5 B1048.4	CCAFS SLC-40	Starlink 1 v1.0, SpaceX CRS-19	15600	LEO	SpaceX	Success	Success
2020-07-01	02:33:00	F9 B5 B1049.4	CCAFS SLC-40	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600	LEO	SpaceX	Success	Success
2020-01-29	14:07:00	F9 B5 B1051.3	CCAFS SLC-40	Starlink 3 v1.0, Starlink 4 v1.0	15600	LEO	SpaceX	Success	Success
2020-02-17	15:05:00	F9 B5 B1056.4	CCAFS SLC-40	Starlink 4 v1.0, SpaceX CRS-20	15600	LEO	SpaceX	Success	Failure
2020-03-18	12:16:00	F9 B5 B1048.5	KSC LC-39A	Starlink 5 v1.0, Starlink 6 v1.0	15600	LEO	SpaceX	Success	Failure
2020-04-22	19:30:00	F9 B5 B1051.4	KSC LC-39A	Starlink 6 v1.0, Crew Dragon Demo-2	15600	LEO	SpaceX	Success	Success
2020-04-06	01:25:00	F9 B5 B1049.5	CCAFS SLC-40	Starlink 7 v1.0, Starlink 8 v1.0	15600	LEO	SpaceX, Planet Labs	Success	Success
2020-03-09	12:46:14	F9 B5 B1060.2	KSC LC-39A	Starlink 11 v1.0, Starlink 12 v1.0	15600	LEO	SpaceX	Success	Success
2020-06-10	11:29:34	F9 B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-18	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-24	15:31:34	F9 B5 B1060.3	CCAFS SLC-40	Starlink 14 v1.0, GPS III-04	15600	LEO	SpaceX	Success	Success
2020-11-25	02:13:00	F9 B5 B1049.7	CCAFS SLC-40	Starlink 15 v1.0, SpaceX CRS-21	15600	LEO	SpaceX	Success	Success

- Present your query result with a short explanation here

%sql select * from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select MAX (PAYLOAD_MASS__KG_) from SPACEXTABLE)

The max Pay_LoadMass_kg where 15,600 kg so we obtain 12 rows as result

2015 Launch Records

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

month	Date	Booster_Version	Launch_Site	Landing _Outcome
01	10-01-2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	14-04-2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- Present your query result with a short explanation here
  - %sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, Landing_Outcome FROM SPACEXTABLE where Landing_Outcome like '%drone ship%' and substr(Date,7,4)='2015';
  - In this query, we use substr(Date, 4, 2) as month and columns of our interest, the we use substr(Date,7,4) for year. The result is showed above.

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

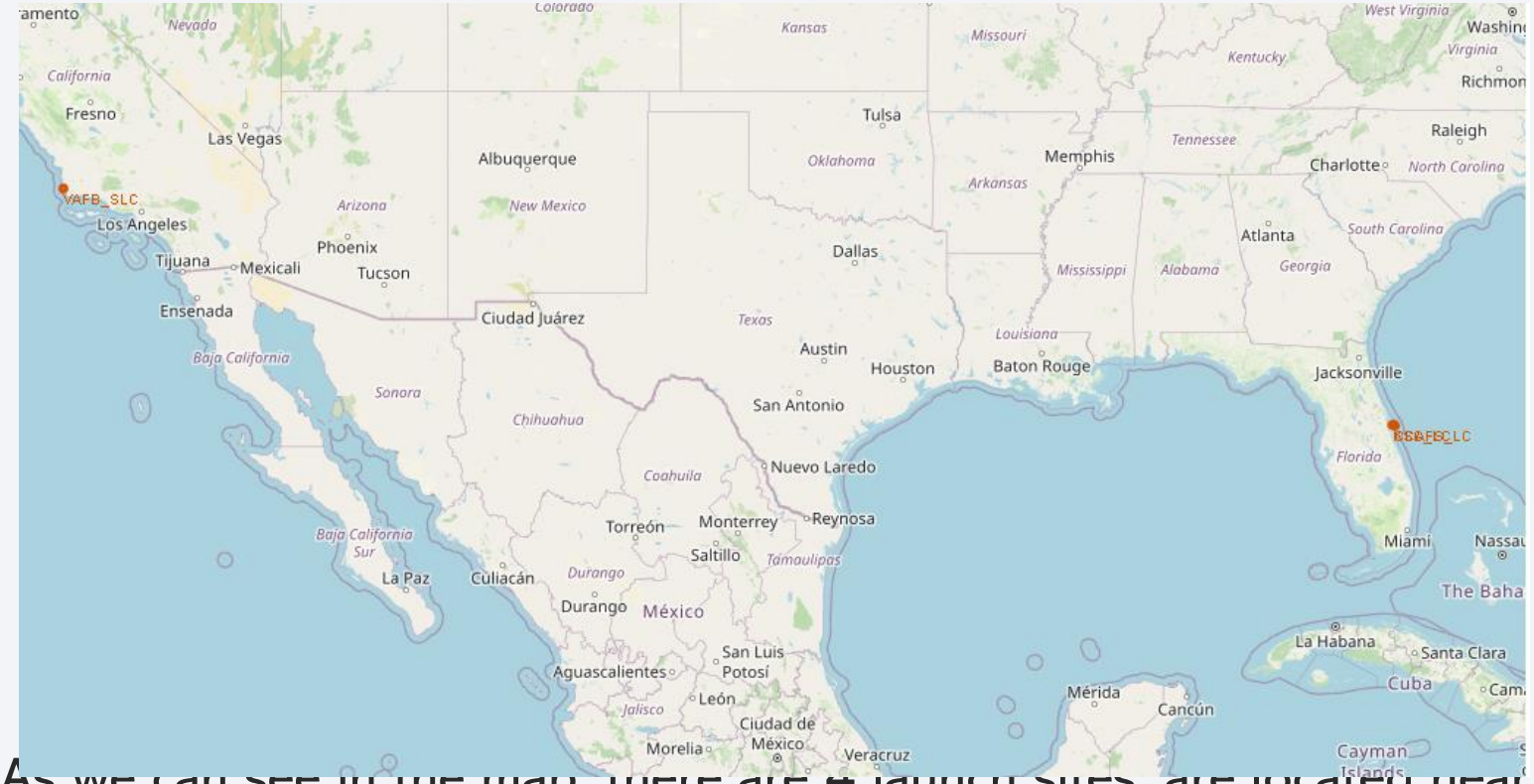
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing _Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

- Present your query result with a short explanation here
  - %sql SELECT "DATE", COUNT(Landing_Outcome) as COUNT FROM SPACEXTBL WHERE "DATE" BETWEEN '2010-06-04' and '2017-03-20' AND Landing_Outcome LIKE '%Success%' GROUP BY "DATE" ORDER BY COUNT(Landing_Outcome) Desc
  - For this query, we use the function “Count” for the Landing Outcomes column, where the records are between the dates given and the Landing Outcome where “Success” and group by date. Finally, the result where presented in a descending order.

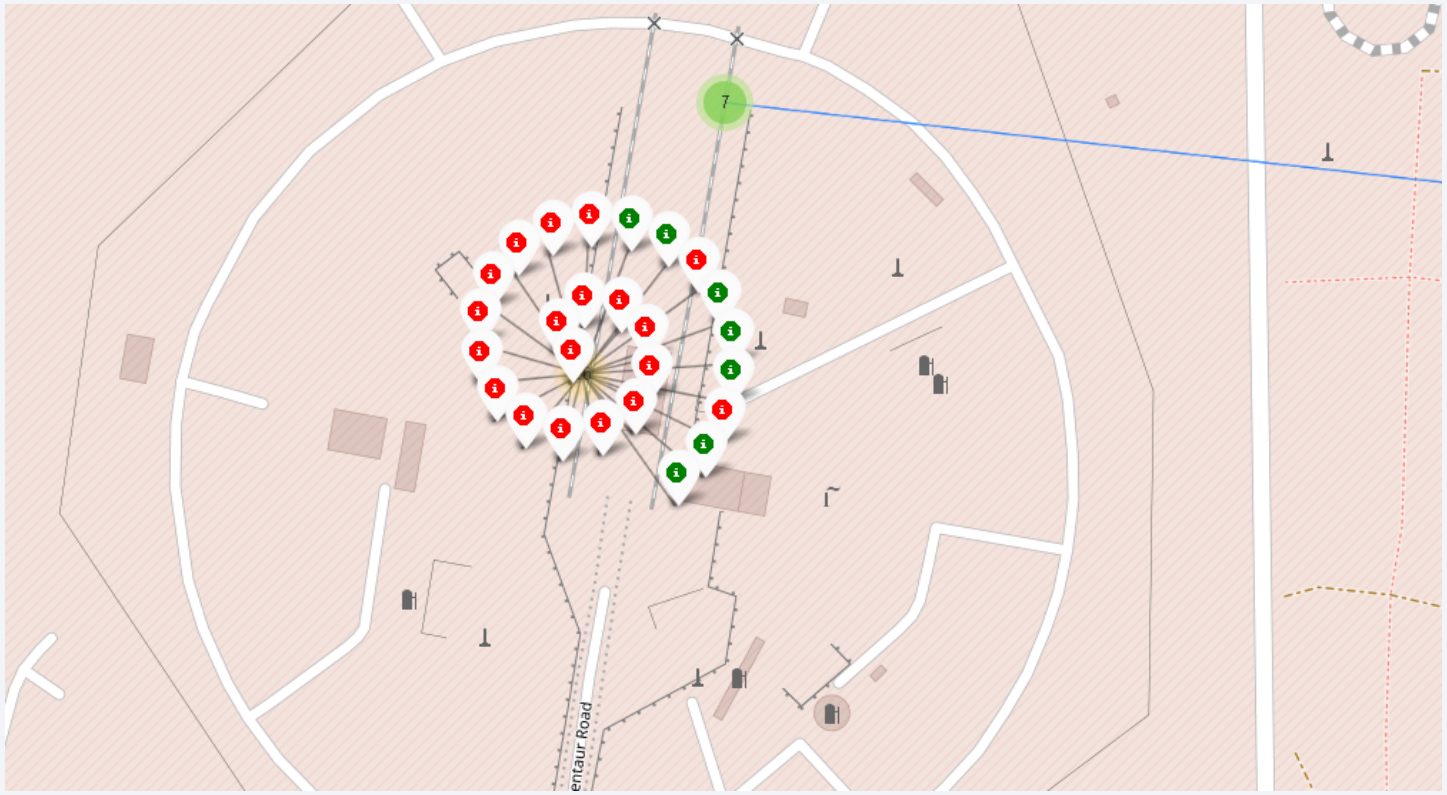

Section 3
Launch Sites
Proximities Analysis

Launch sites location



- As we can see in the map, there are 4 launch sites, are located near by the coastline due to the transportation, and also there are at least launch site in each ocean (Pacific and Atlantic)

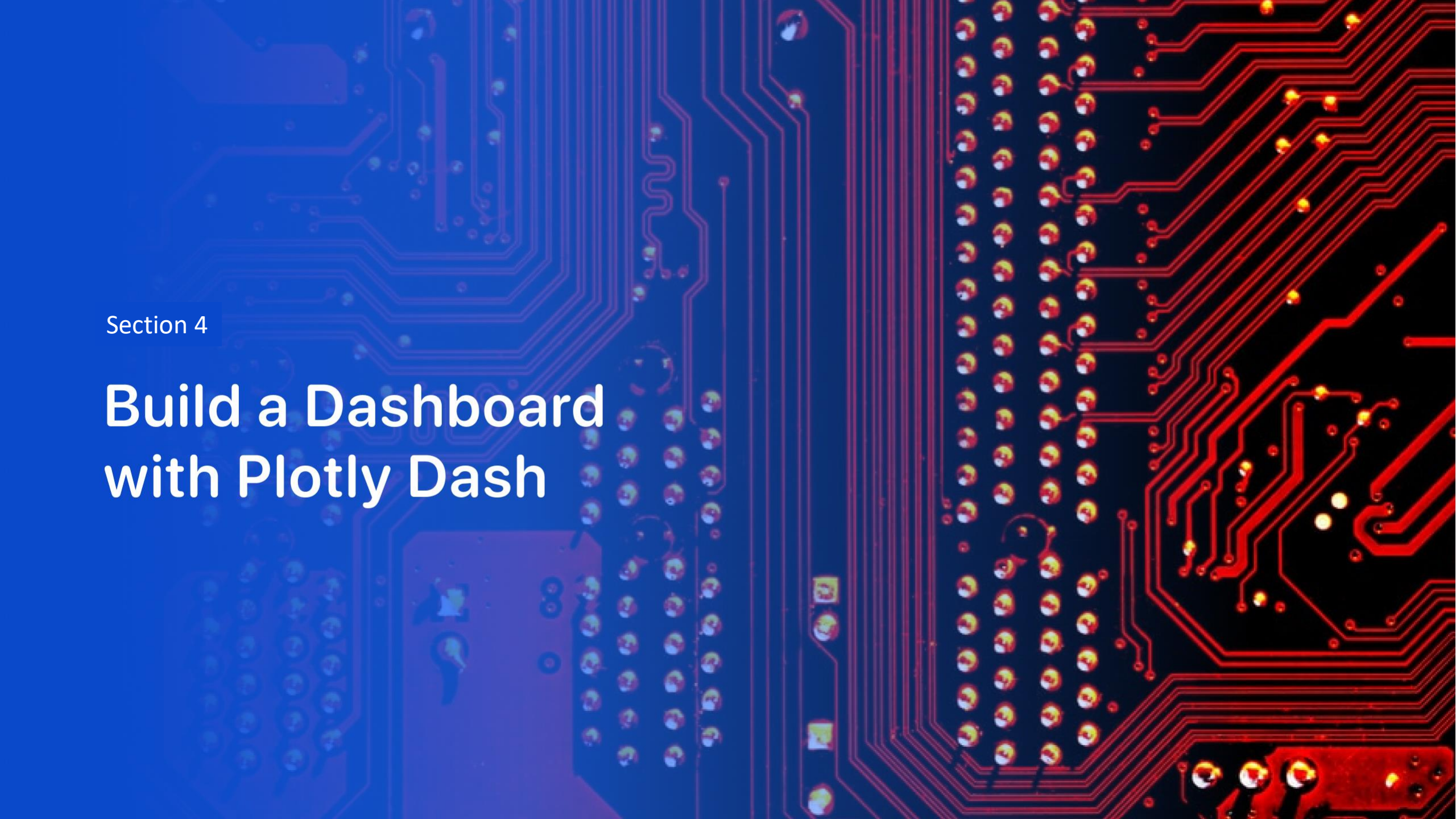
Launch outcomes map



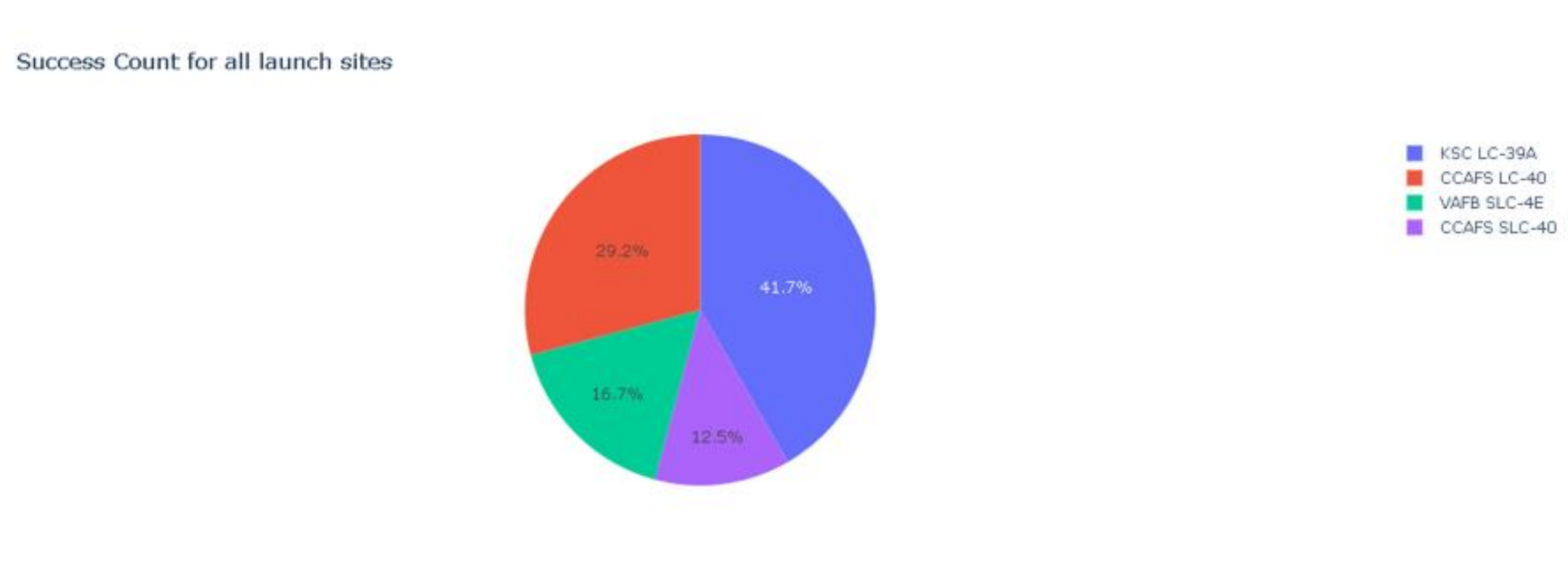
- The launch outcome map shows the success and failure outcomes from each launch site. This is a great visual help to show the success/failure rate

Distances from CCAFS launch site to closest railways, highways, coastline and Melbourne

- The CCAFS launch site is near by the highway by 0.58, also from the railways by 1.2 km and from the coastline by 0.9 km. This is so important due to the booster transportation

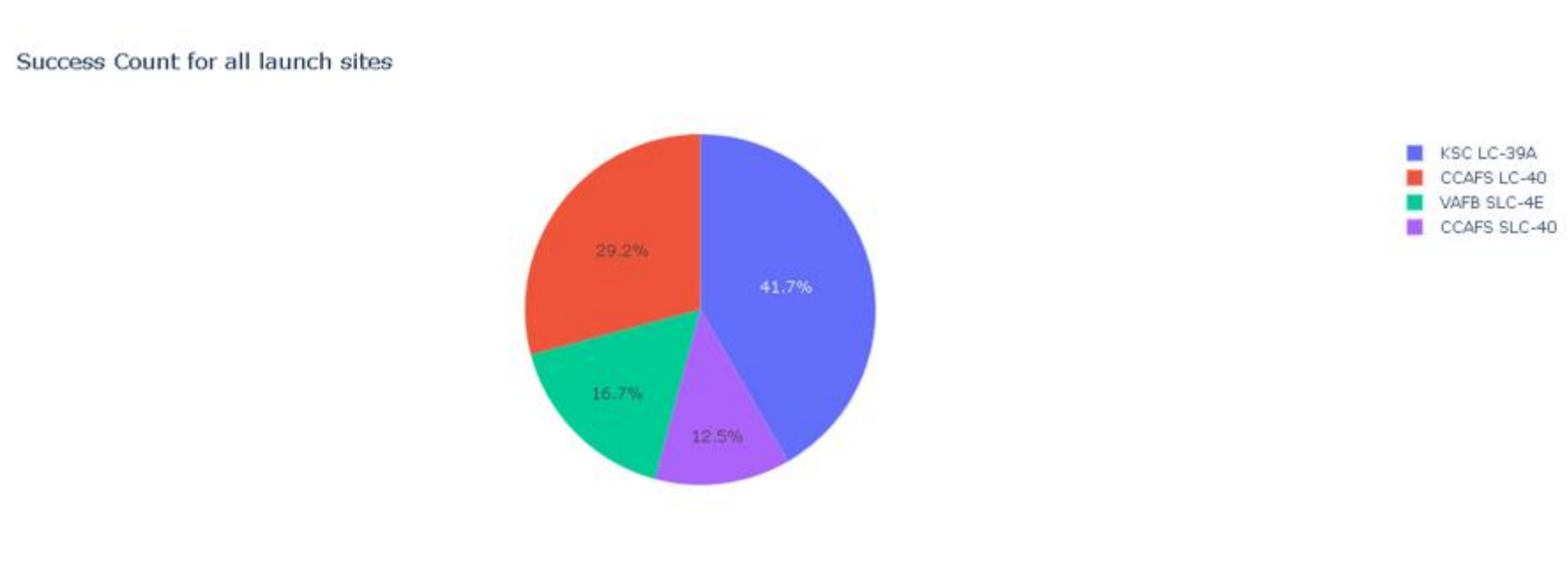

Section 4
Build a Dashboard with Plotly Dash

Success Count for all launch sites



- As we can see in the graph, the KSC LC-39A has the higher success rate with 41.7% while VAFB SLC-4E has the lowest success rate with 16.7%

KSC LC-39A launch site



- For Falcon 9 KSC LC-39A launch site has the highest success rate between all the launch sites, I didn’t take screenshot but, we know que we can see every success / failure ratio

PayLoad Mass vs Outcome

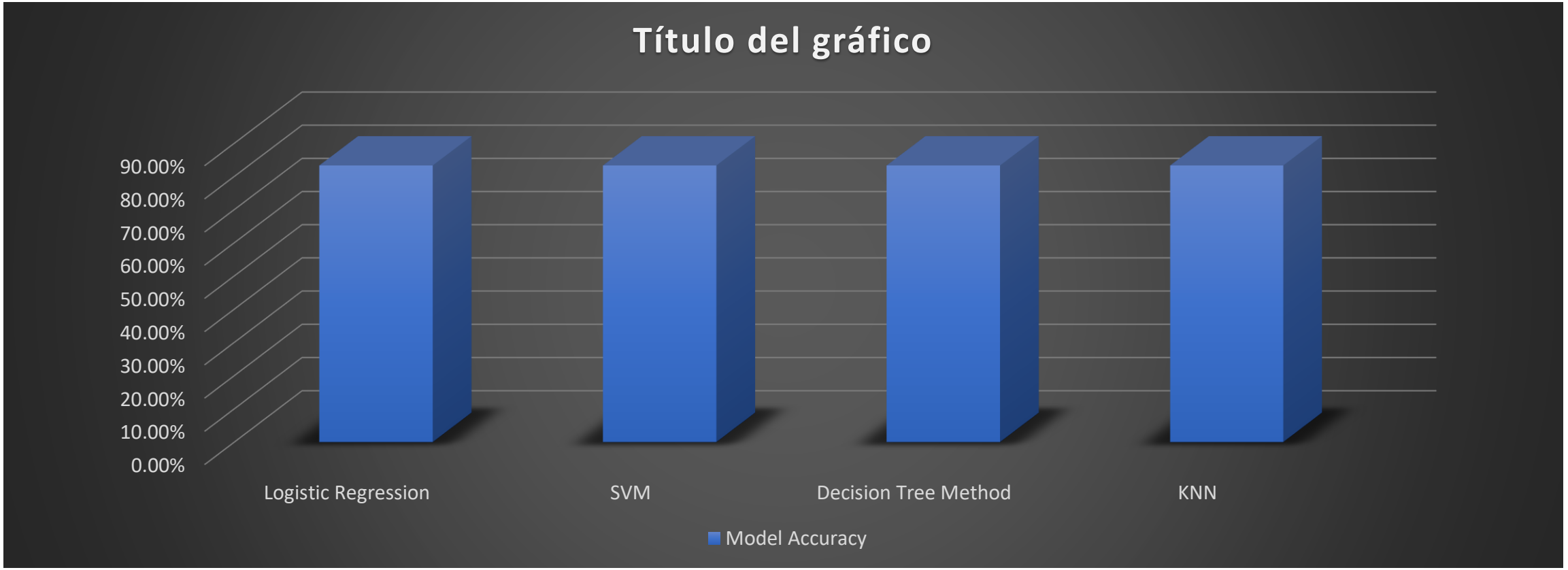
- In this graph, we can manipulate different range of mass, so the X axis auto scale itself, with this tool, it easier to see the launch outcome with different Pay Load mass

Section 5

Predictive Analysis (Classification)

Classification Accuracy

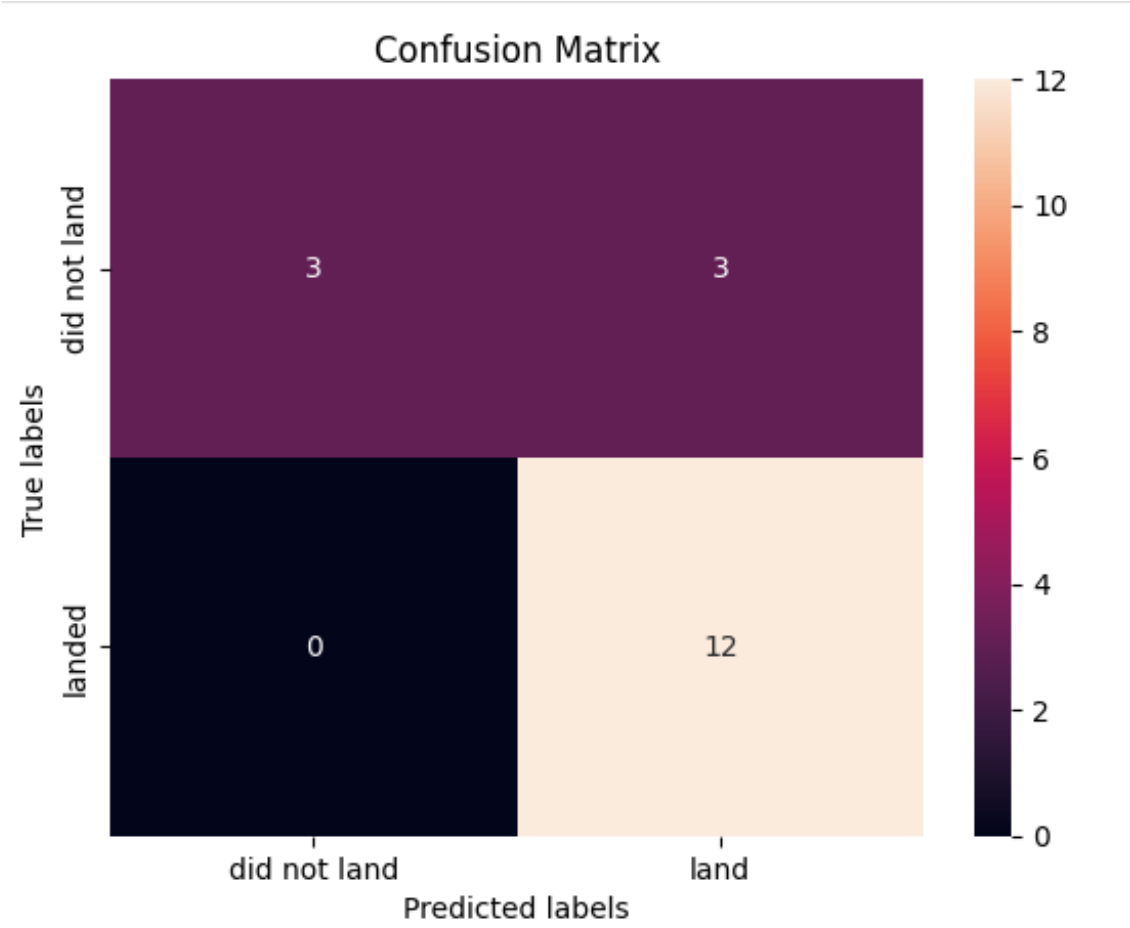
- Visualize the built model accuracy for all built classification models, in a bar chart
- In this case, Logistic regression, Support Vector Machine, Decision tree Method and KNN model give us the same accuracy



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

- We can distinguish between different classes. 15 values are correct while 3 are wrong, so, we can say that the accuracy is high

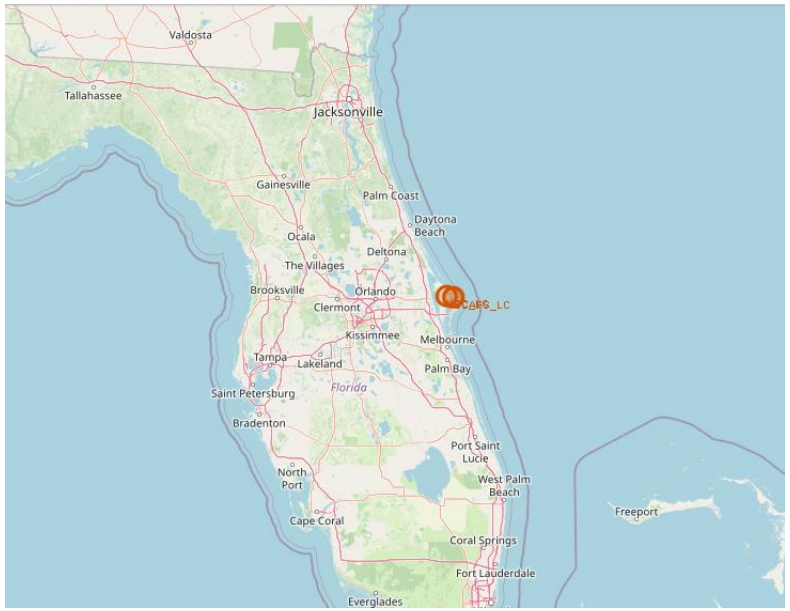


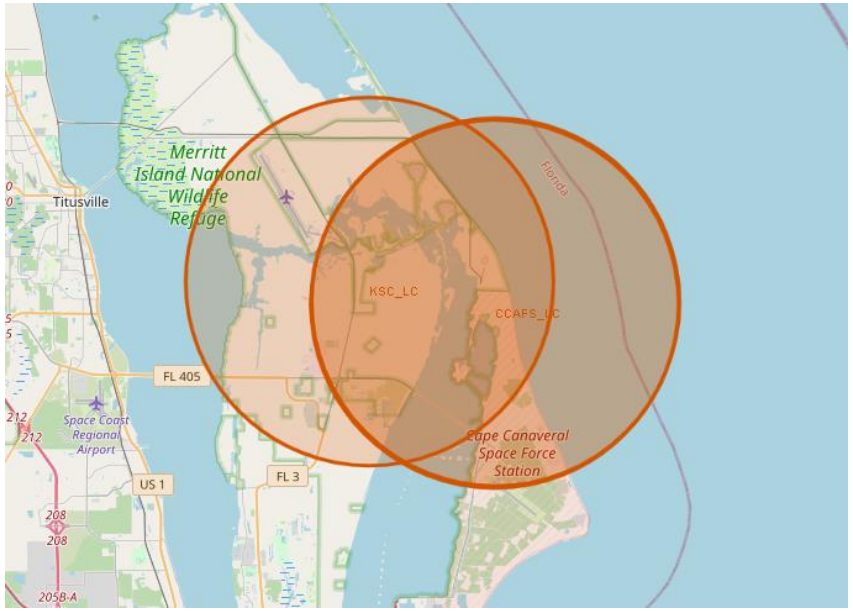
Conclusions

- For the Falcon-9 launch sites are located near by a coastline and railways
- KSC LC-39A sites has the higher success rate
- ES-L1, GEO, HEO and SSO orbits has 100% success rate
- The success ratio has increase through the time
- According with the model evaluation, we can use any predict model, with 83.33 accuracy everyone

Appendix

- Detail of KSC_LC and CCAFS_LC launch sites due to they are so close each other in the east coast





Thank you!