IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Murat Çelik
15.01.2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data were extracted, analyzed and pre-processed.
  - The rocket to be sent with the space rocket data were analyzed.
  - It has been studied with many different data such as the success status, reusability and weight of the rockets sent.
  - Detection of missing data, use of SQL, data analysis and inferences were made.
- Summary of all results
  - GridSearchCV is used. Logistic regression, Support Vector Machine, Decision Tree, K Nearest-Neighbor models were used. The results are close to each other. The overall result is 83.3%.

Introduction

- By working with space rocket data, operations were made on it. Which rockets are reusable, their weight and various information and flight successes are defined.
- Due to the diversity of data, different problems can be addressed. Different problems such as launch cost, location of first standing, reusability and success can be addressed.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Worked with spacexdata. First of all, the main data was drawn with the json data. Afterwards, additional operations were performed on the data with request commands. JSON and Request methods are learned.
- Perform data wrangling
  - Missing data was checked, data analysis was performed to generate new 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
  - We divided the data into train and test. Then we gave it to the model. We analyzed it with the confusion matrix and compared the scores.

Data Collection

- Worked with SpaceX REST API and Wikipedia,List of Falcon 9 and Falcon Heavy launches, data. Data was taken with request and worked with various columns.

2020 [edit]

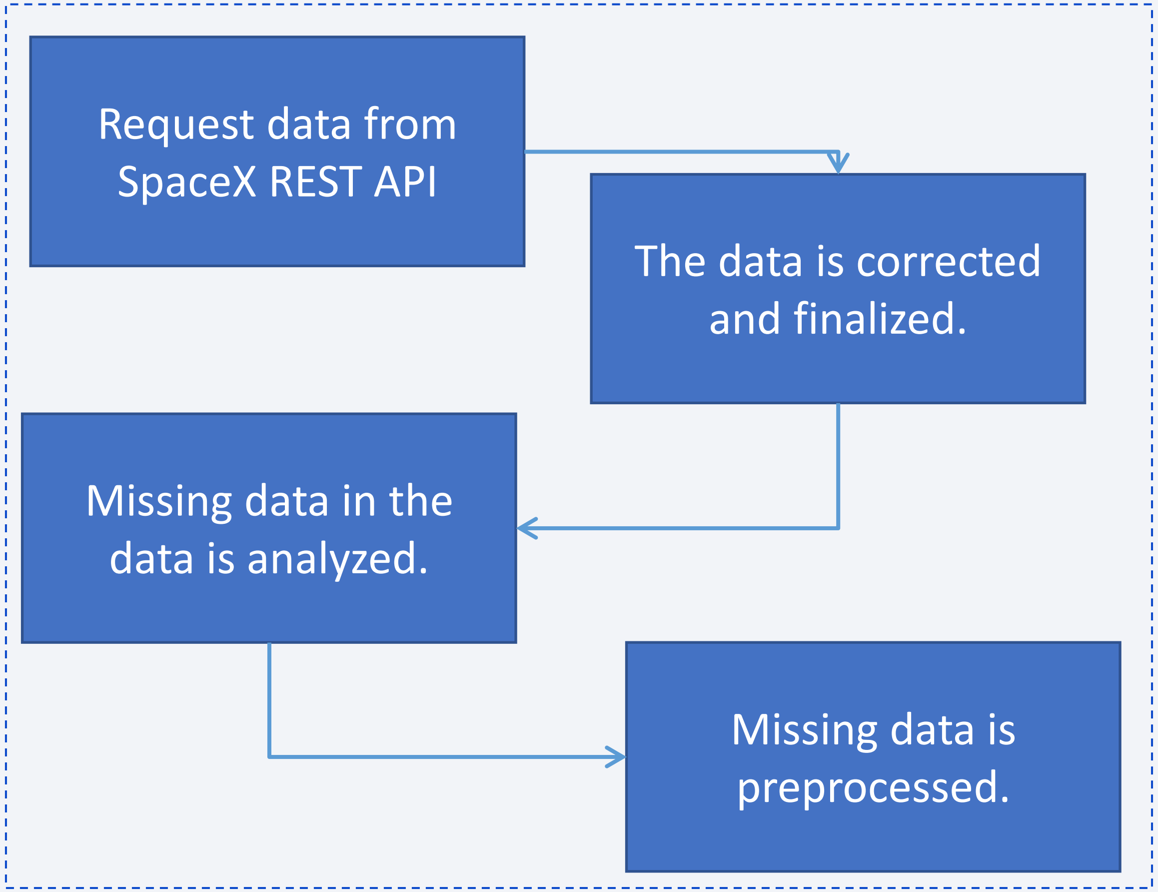
In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	Landin
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb23

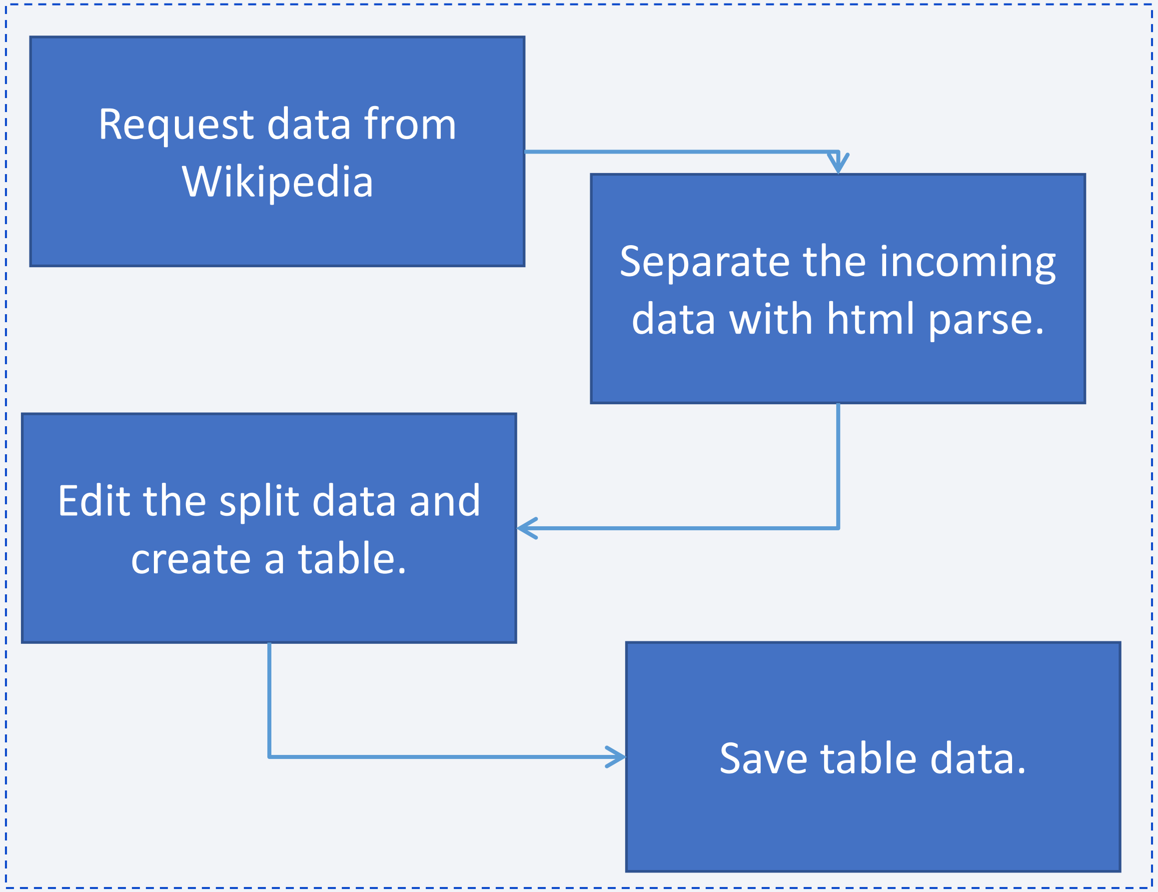
Data Collection – SpaceX API

- https://github.com/MuratCelik3506/IBM-DSCP/blob/main/Week1/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- https://github.com/MuratCelik3506/IBM-DSCP/blob/main/Week1/jupyter-labs-webscraping.ipynb

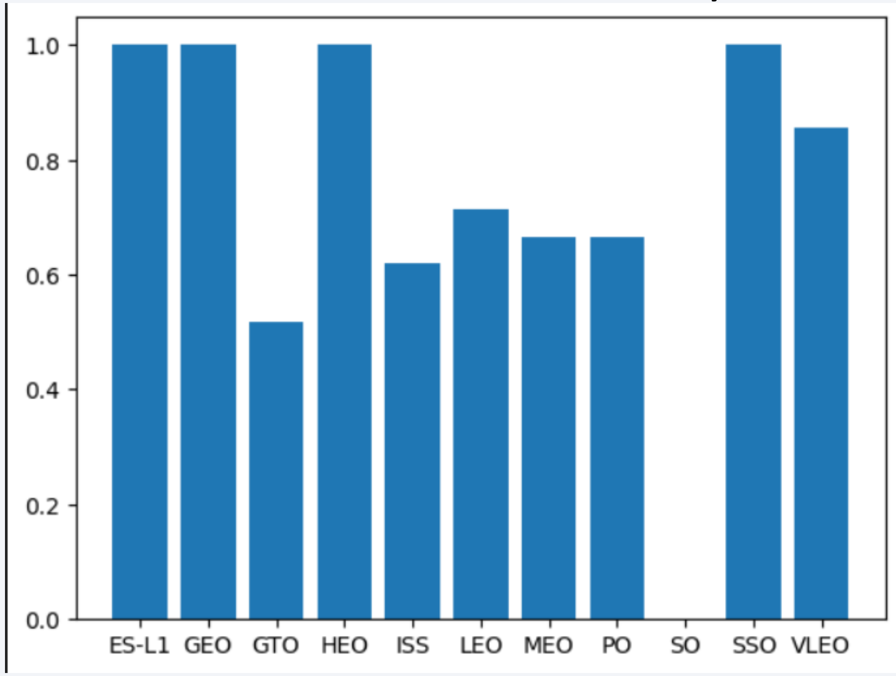


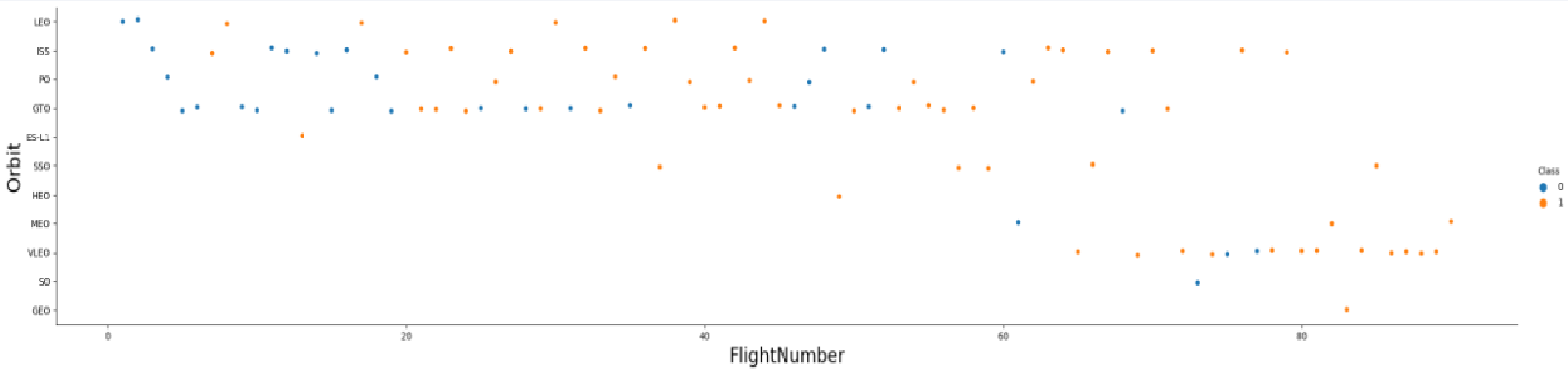
Data Wrangling

- 1) Data are analyzed.
  - 1.1) The content of the data is checked.
  - 1.2) The number of contents in the data is checked.
  - 1.3) The types of data are checked.
- 2) Landing column is added using outcome data.
- 3) Save the new data.
- https://github.com/MuratCelik3506/IBM-DSCP/blob/main/Week1/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Different graphics such as catplot, barchart, lineplot were drawn. These are tools that make it easy for us to analyze the data
- Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Trend





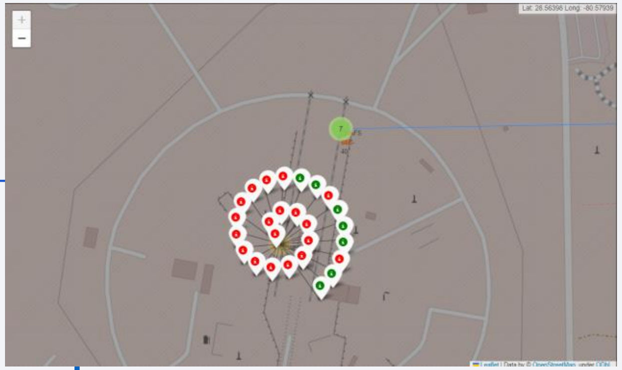
https://github.com/MuratCelik3506/IBM-DSCP/blob/main/Week2/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the   names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

https://github.com/MuratCelik3506/IBM-DSCP/blob/main/Week2/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium



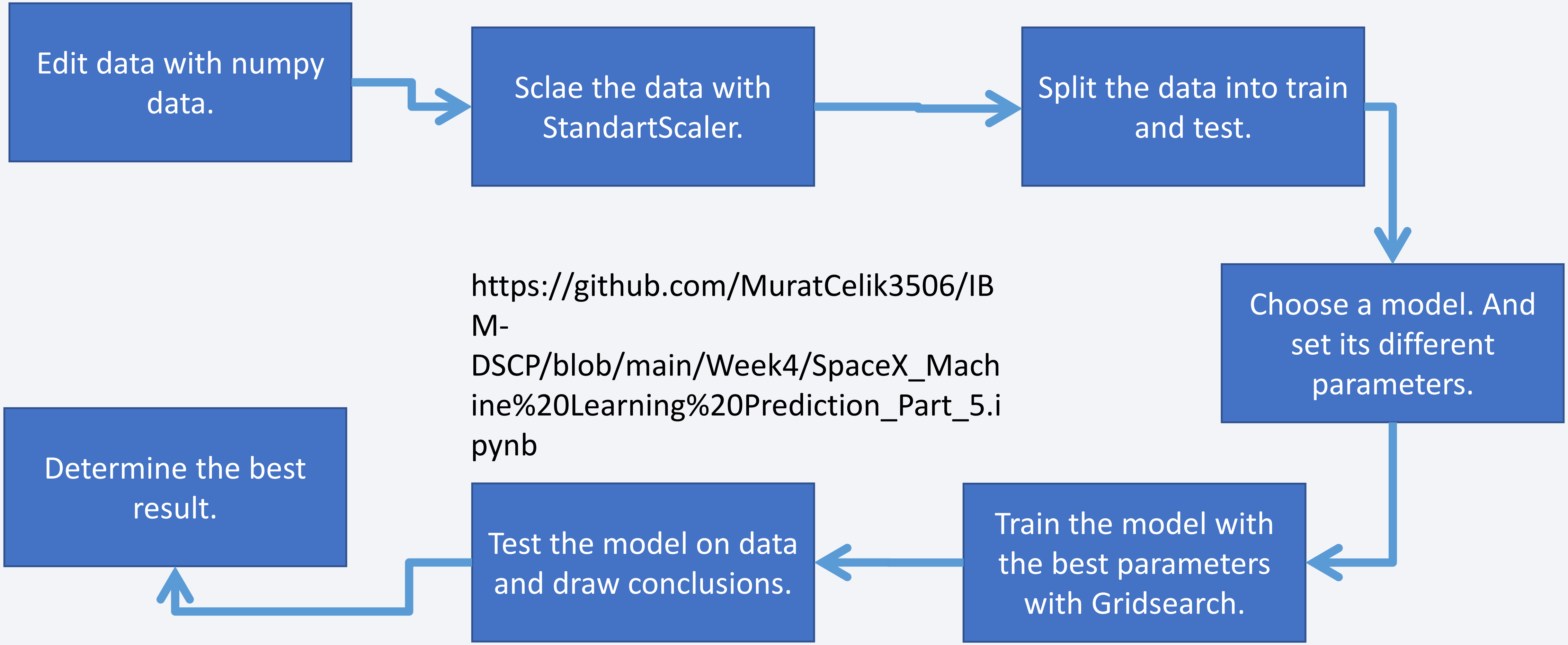
- https://github.com/MuratCelik3506/IBM-DSCP/blob/main/Week3/lab_jupyter_launch_site_location.ipynb

- Markers, circles, lines and marker clusters were used with Folium Maps
- • Markers indicate points like launch sites;
- • Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space
- Center;
- • Marker clusters indicates groups of events in each coordinate, like launches in a launch site;
- and
- • Lines are used to indicate distances between two coordinates.

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
- • Percentage of launches by site
- • Payload range
- • This combination allowed to quickly analyze the relation between payloads
- and launch sites, helping to identify where is best place to launch according
- to payloads.

Predictive Analysis (Classification)

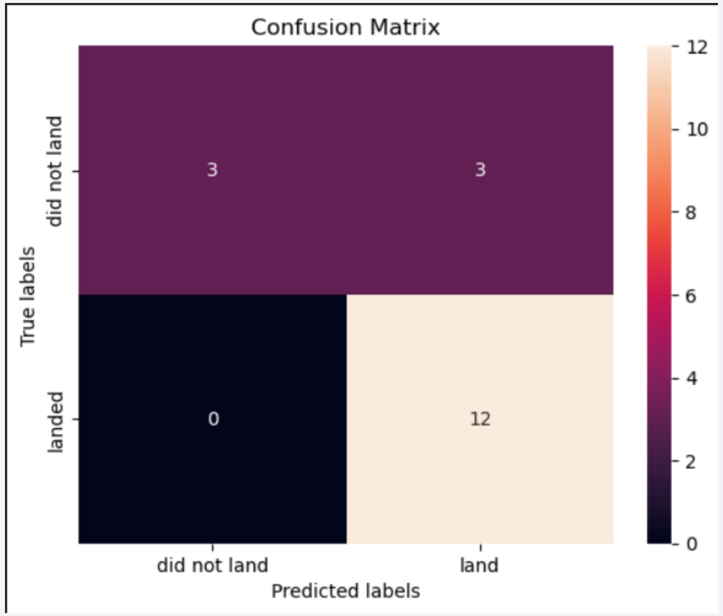


https://github.com/MuratCelik3506/IBM-DSCP/blob/main/Week4/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Logistic regression, support vector machine, decision tree classifier, k nearest neighbors

Results

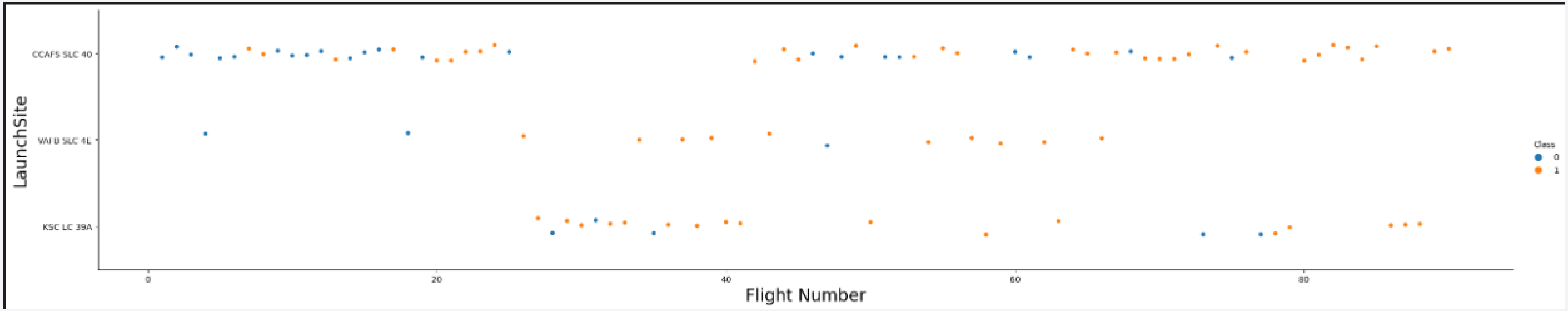
- Exploratory data analysis results
- Score : 0.8333333333333334
- Predictive analysis results



Section 2

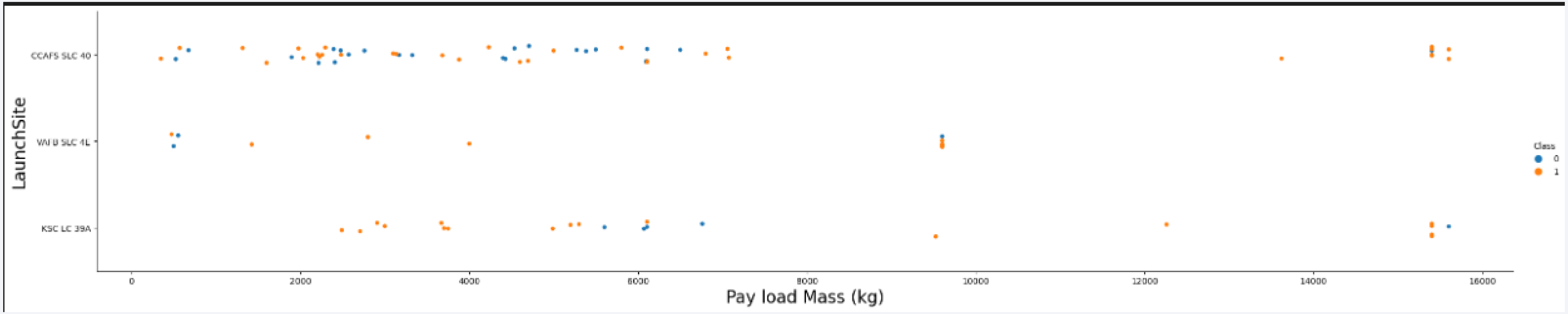
Insights drawn from EDA

Flight Number vs. Launch Site



- According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful
- VAFB SLC 4E and KSC LC 39A have higher success rates.

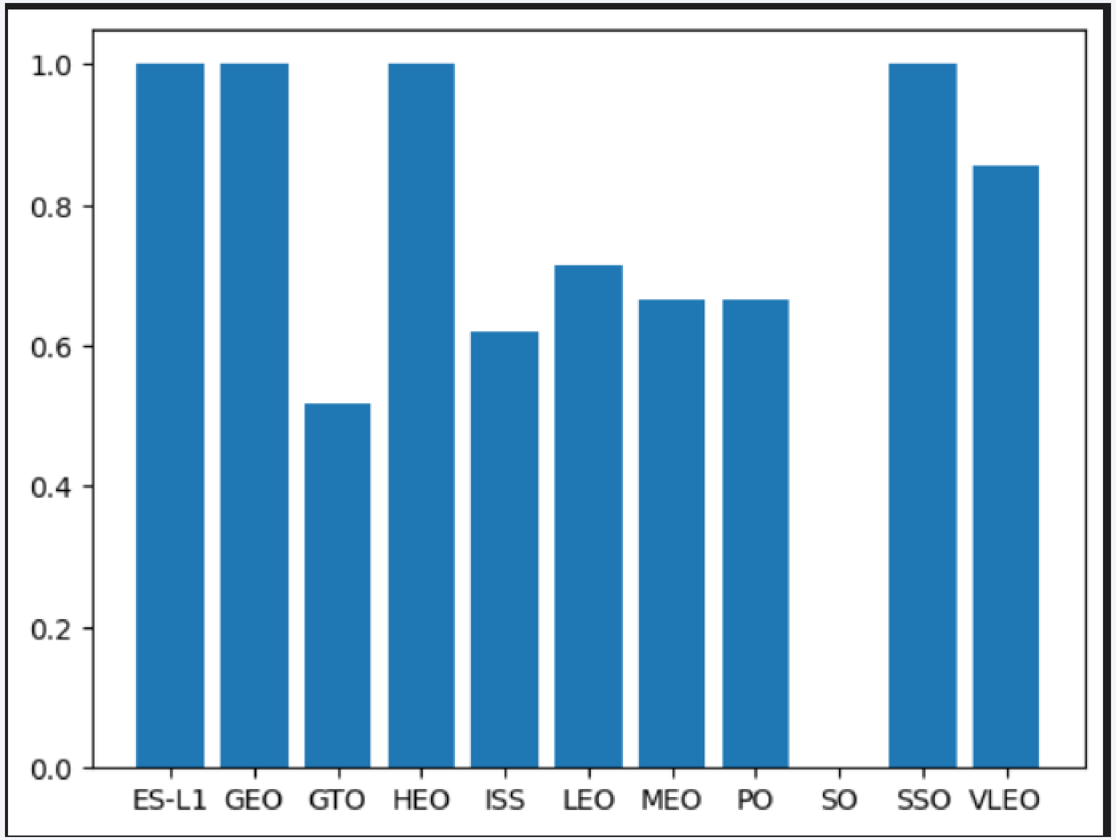
Payload vs. Launch Site



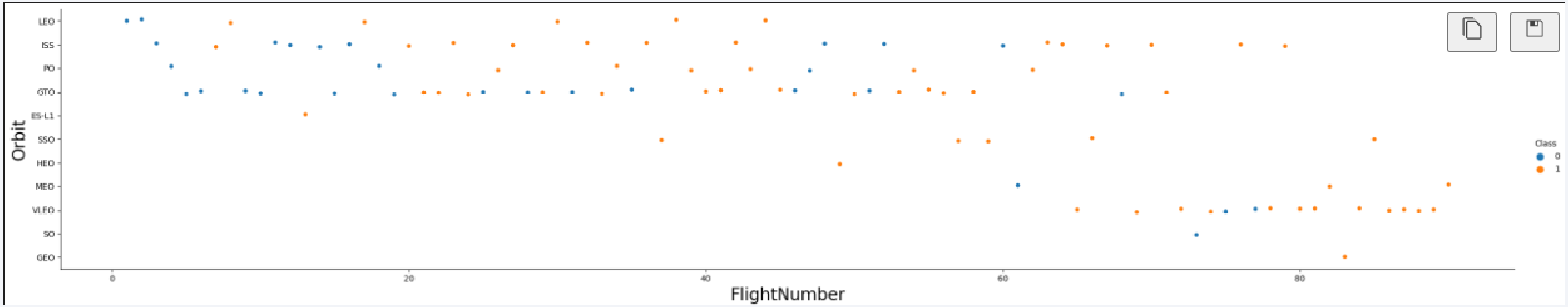
- Most of the launches with payload mass over 7000 kg were successful.
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- Explanation:
  - Orbits with 100% success rate:
    - ES-L1, GEO, HEO, SSO
  - Orbits with 0% success rate:
    - SO
  - Orbits with success rate between 50% and 85%:
    - GTO, ISS, LEO, MEO, PO

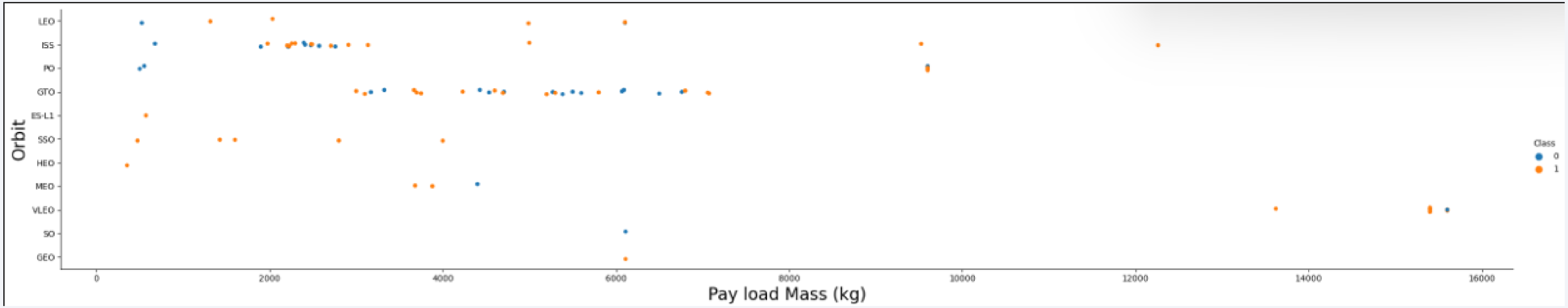


Flight Number vs. Orbit Type



- LEO orbit the Success appears related to the number of flights
- VLEO orbit seems a new business opportunity, due to recent increase of its frequency.

Payload vs. Orbit Type

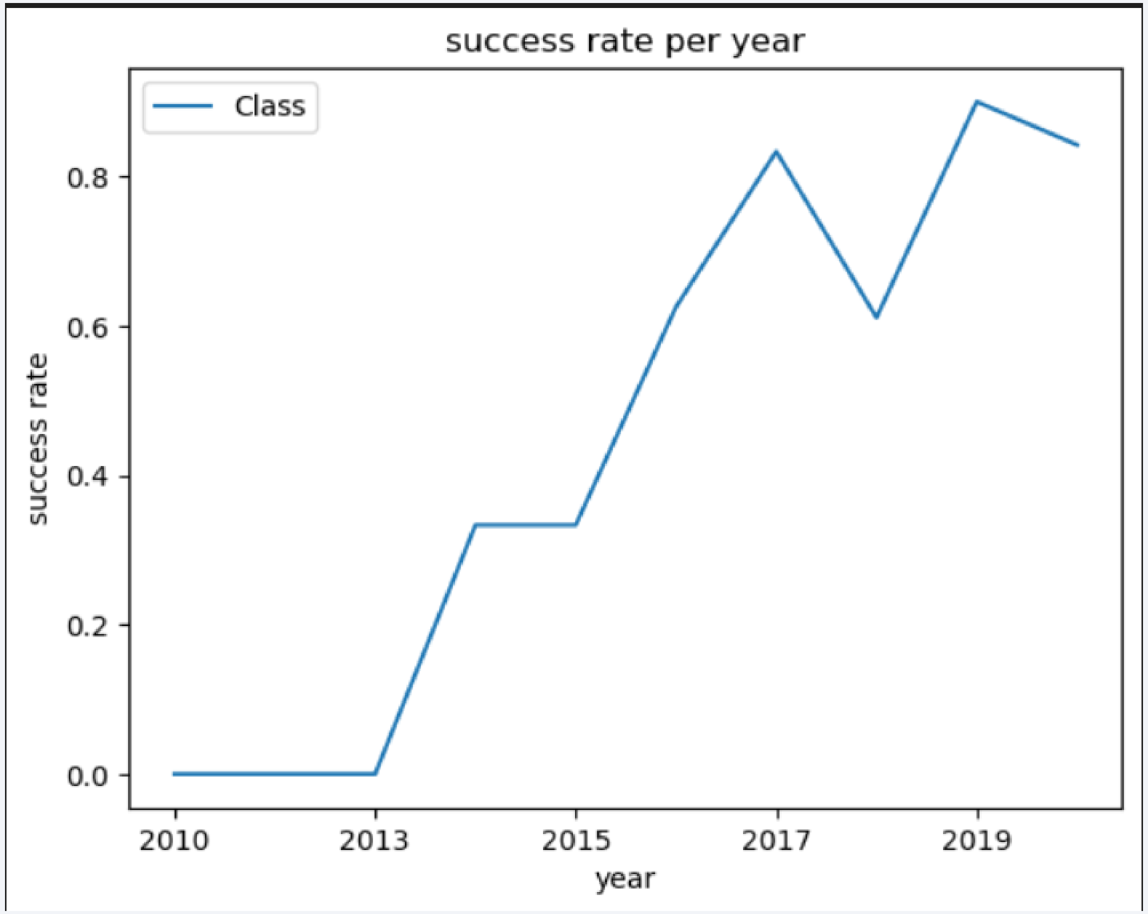


- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.
- There are few launches to the orbits SO and GEO.

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- Displaying the names of the unique launch sites in the space mission.

```
array(['CCAFS LC-40', 'VAFB SLC-4E', 'KSC LC-39A', 'CCAFS SLC-40'],
      dtype=object)
```

Launch Site Names Begin with 'CCA'

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
0	04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	08-12-2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	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

- **45596**

Average Payload Mass by F9 v1.1

- 2928.4

First Successful Ground Landing Date

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Successful Drone Ship Landing with Payload between 4000 and 6000

```
array(['F9 v1.1', 'F9 v1.1 B1011', 'F9 v1.1 B1014', 'F9 v1.1 B1016',
       'F9 FT B1020', 'F9 FT B1022', 'F9 FT B1026', 'F9 FT B1030',
       'F9 FT  B1021.2', 'F9 FT B1032.1', 'F9 B4 B1040.1',
       'F9 FT  B1031.2', 'F9 FT  B1032.2', 'F9 B4  B1040.2',
       'F9 B5 B1046.2', 'F9 B5 B1047.2', 'F9 B5 B1048.3',
       'F9 B5 B1051.2 ', 'F9 B5B1060.1', 'F9 B5 B1058.2 ', 'F9 B5B1062.1'],
      dtype=object)
```

Total Number of Successful and Failure Mission Outcomes

```
Success                               98
Failure (in flight)                    1
Success (payload status unclear)       1
Success                                1
Name: Mission_Outcome, dtype: int64
```

Boosters Carried Maximum Payload

```
array(['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 '], dtype=object)
```

2015 Launch Records

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
18	2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)

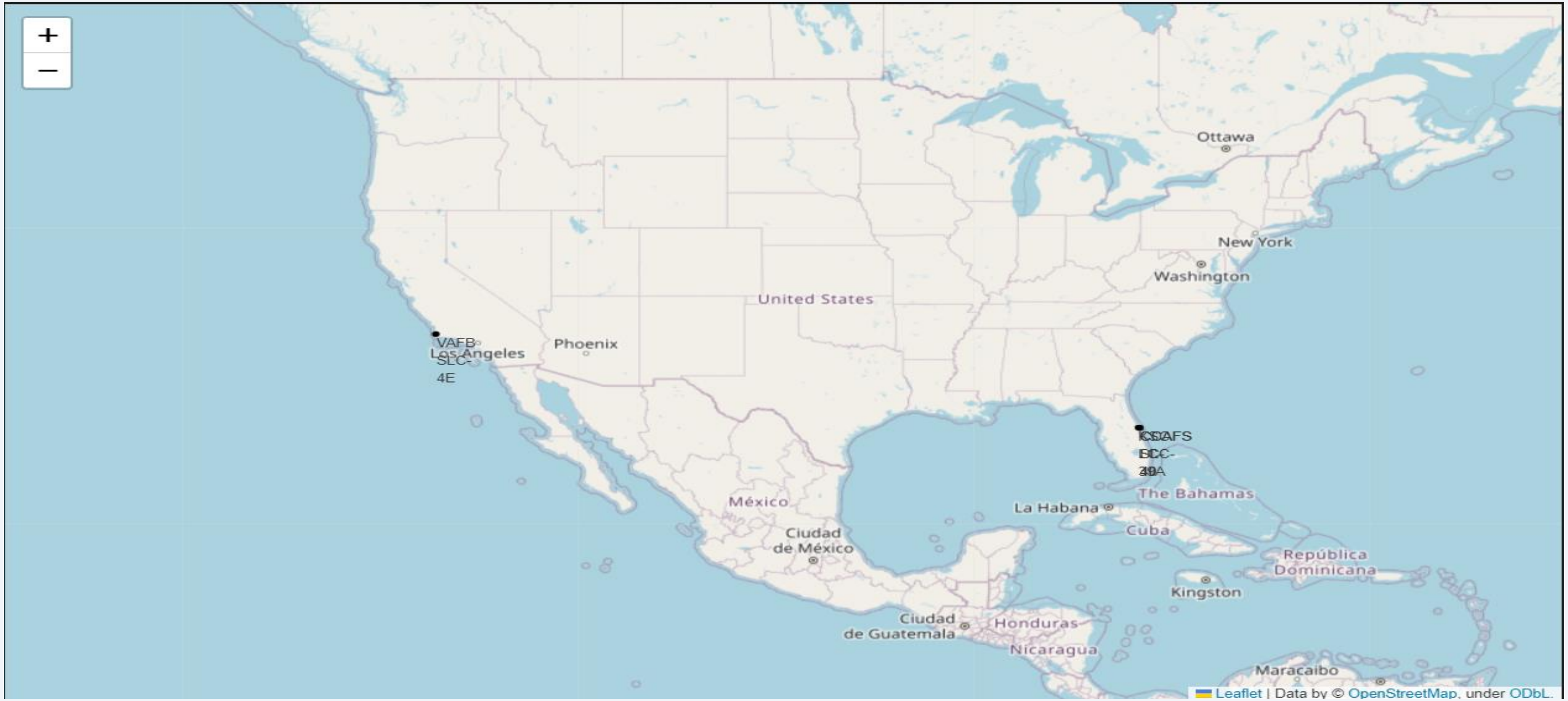
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

```
Success              31
Failure (in flight)   1
Name: Mission_Outcome, dtype: int64
```

Section 3

Launch Sites Proximities Analysis

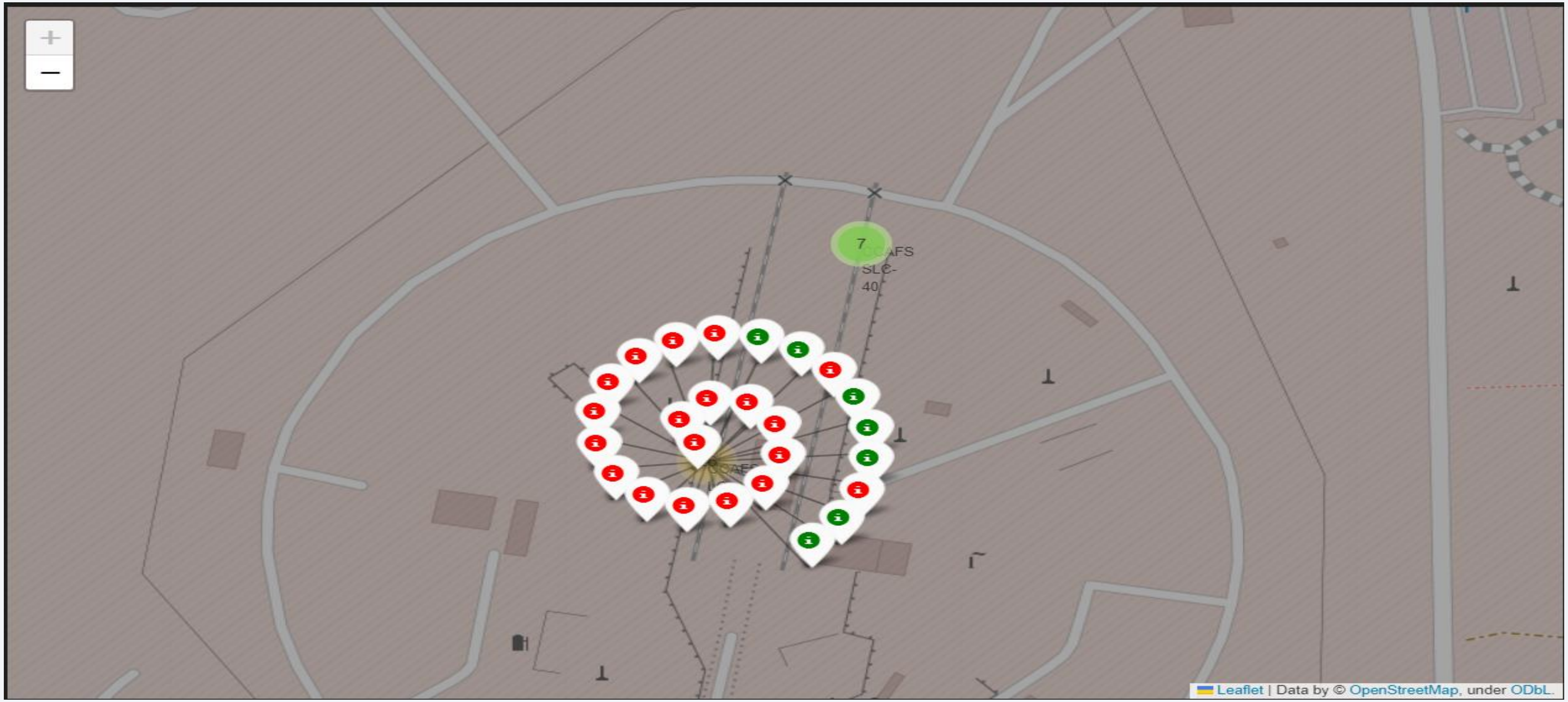
All launch sites



Colour-labeled launch records on the map

Green Marker = Successful Launch

- Red Marker = Failed Launch



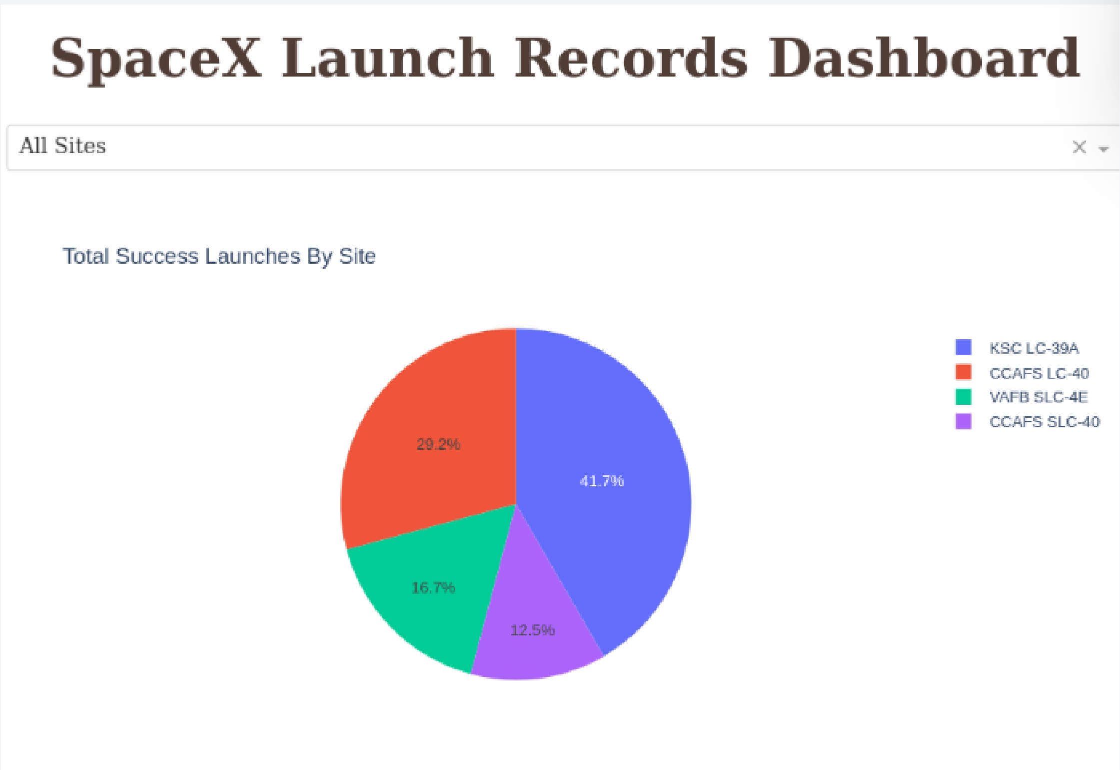
Distance from the launch site

Section 4

Build a Dashboard with Plotly Dash

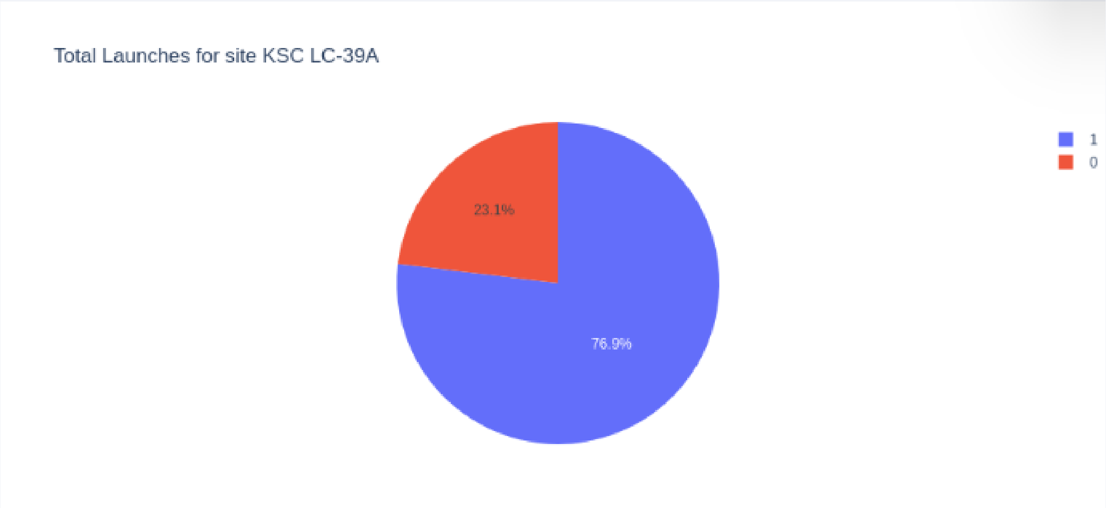
Successful Launches by Site

- Being successful has a close relationship with the landing position.



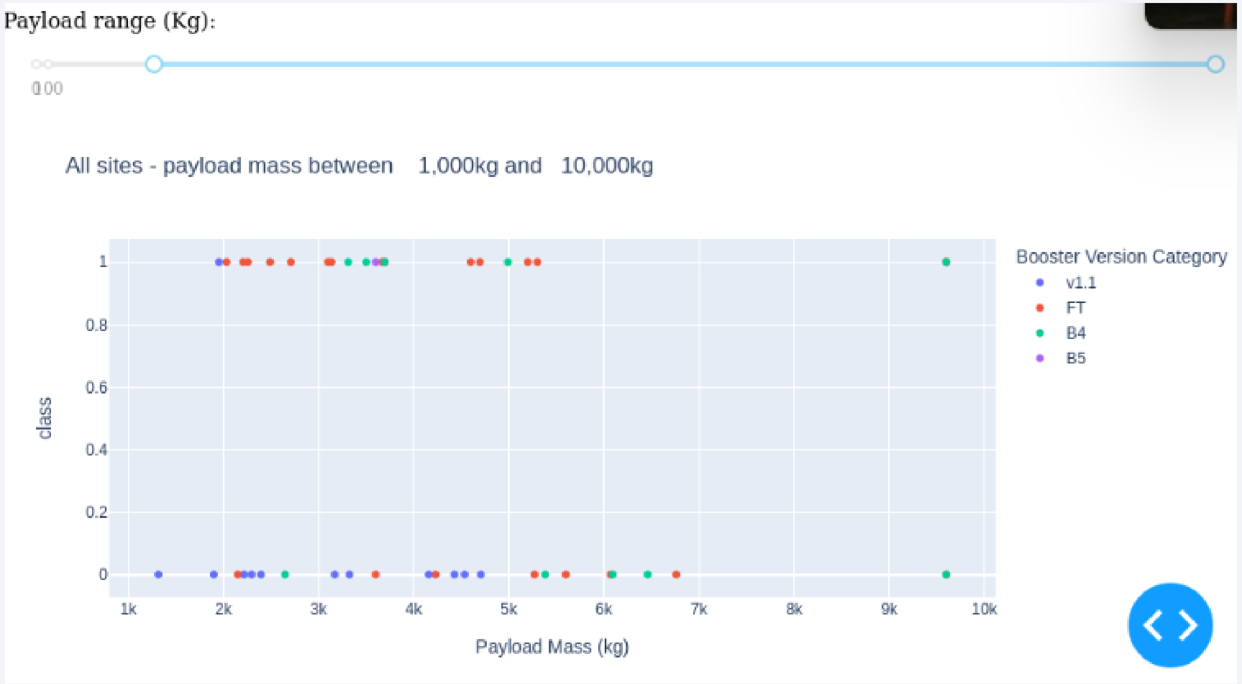
Launch Success Ratio for KSC LC

- 76.9% of launches are successful in this site.



Payload Mass vs. Launch Outcome for all sites

- The charts show that payloads between 2000 and 5500 kg have the highest success rate.

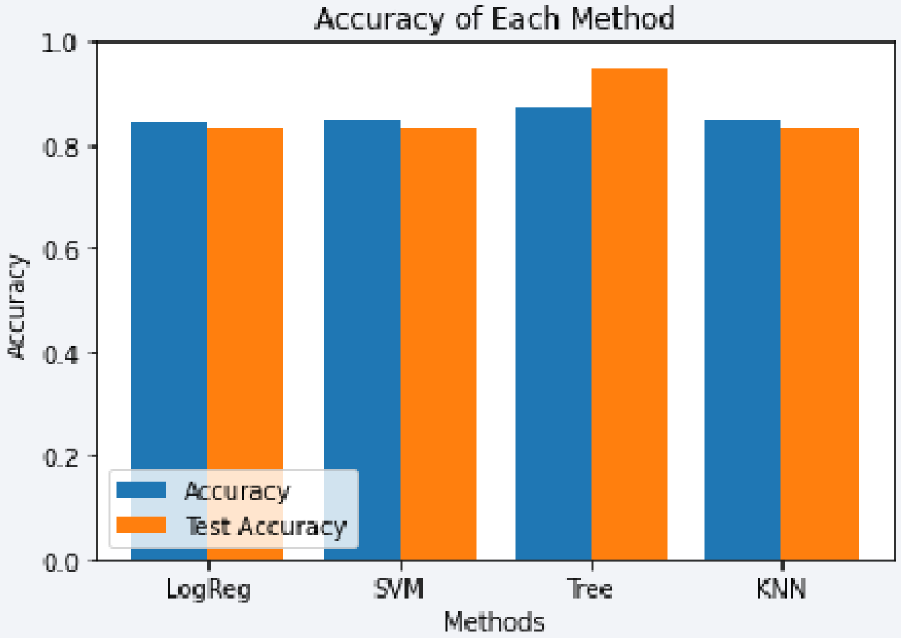


Section 5

Predictive Analysis (Classification)

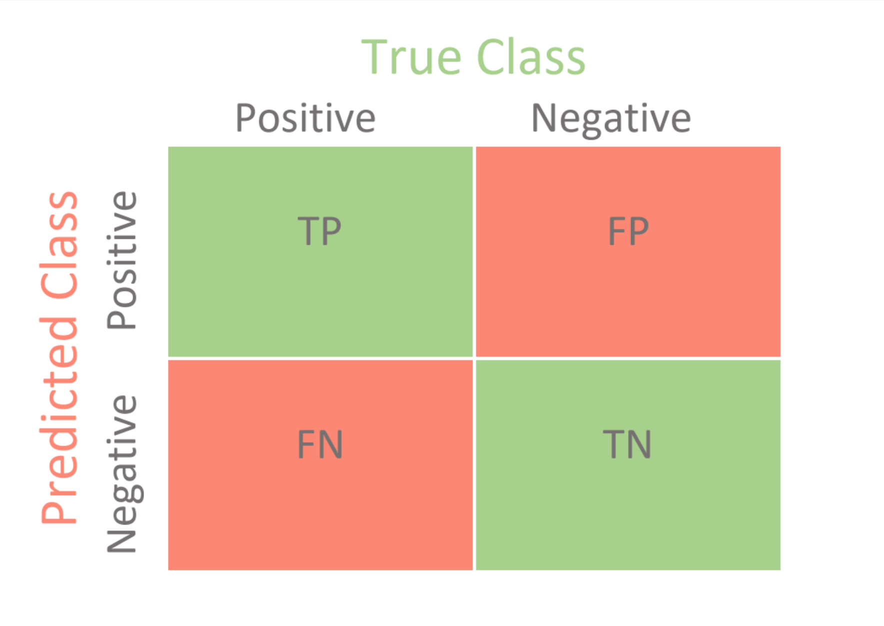
Classification Accuracy

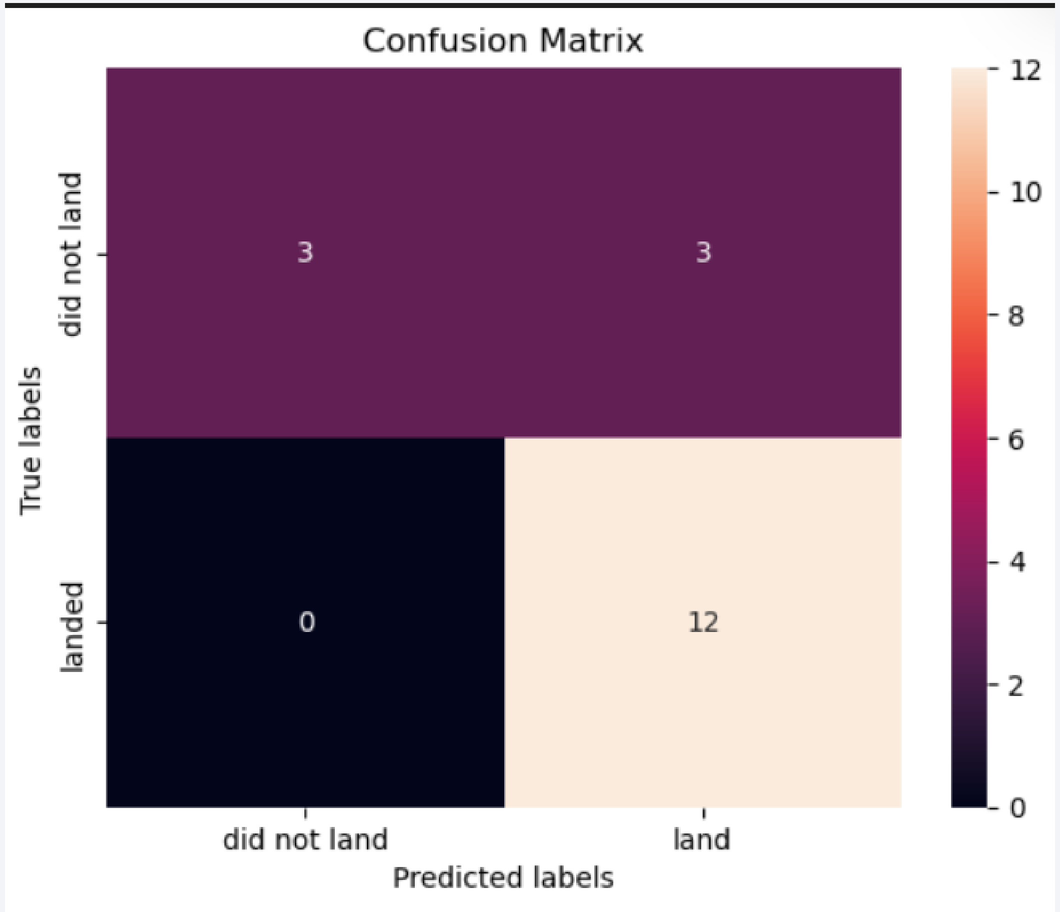
- Score : 0.8333333333333334
- K nearest neighbors
- Decision tree classifier
- Support vector machine
- Logistic regression



Confusion Matrix

- Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.





Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky.
- Most of launch sites are in proximity to the Equator line and all the sites are in very close proximity to the coast.
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate.
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- I think, Support Vector Machine can be used to predict successful landings and increase profits.

Appendix

- https://www.coursera.org/
- https://www.w3schools.com

Thank you!