IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Stephane de Lima
09/01/2022

Outline

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

Executive Summary

- By using the methodology latter detailed, I analyzed the history of launches and outcomes of the Falcon 9 from SpaceX in order to determine what are the best parameters to increase the chance of a success landing.
- With a high success rate, the launch cost will be reduced and the profit increased as the Stage 1 will be used over and over.
- Using the tree model, we can predict the outcome of a launch with an accuracy of 87%.
- I observed that the launch site, payload mass and the booster version have an impact on the probability of a success landing.

Introduction

- SpaceX created the Falcon 9, the first two-stages rockets that can be reusable. It allows to transport people and payload in Earth Orbit and more.
- In order to be used again, the first stage of the rocket has to land successfully and not be damaged.
- In this analysis, I analyze which are the best parameters to have more chance of a successful landing. By doing so, we can increased the number of launches with the same 1st stage and thus increase its return on investment.
- By reducing the costs, a launch to space is much more affordable.

Section 1

Methodology

Methodology

Executive Summary

- Data collection
- Data wrangling
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models

Data Collection

- The data was collected from 2 sources exclusively:
  - SpaceX website (API):

    After collecing all the data from the SpaceX website, I transformed it to a dataframe in order to be more manageable for analysis and filtered the Booster Version only for the Falcon 9.

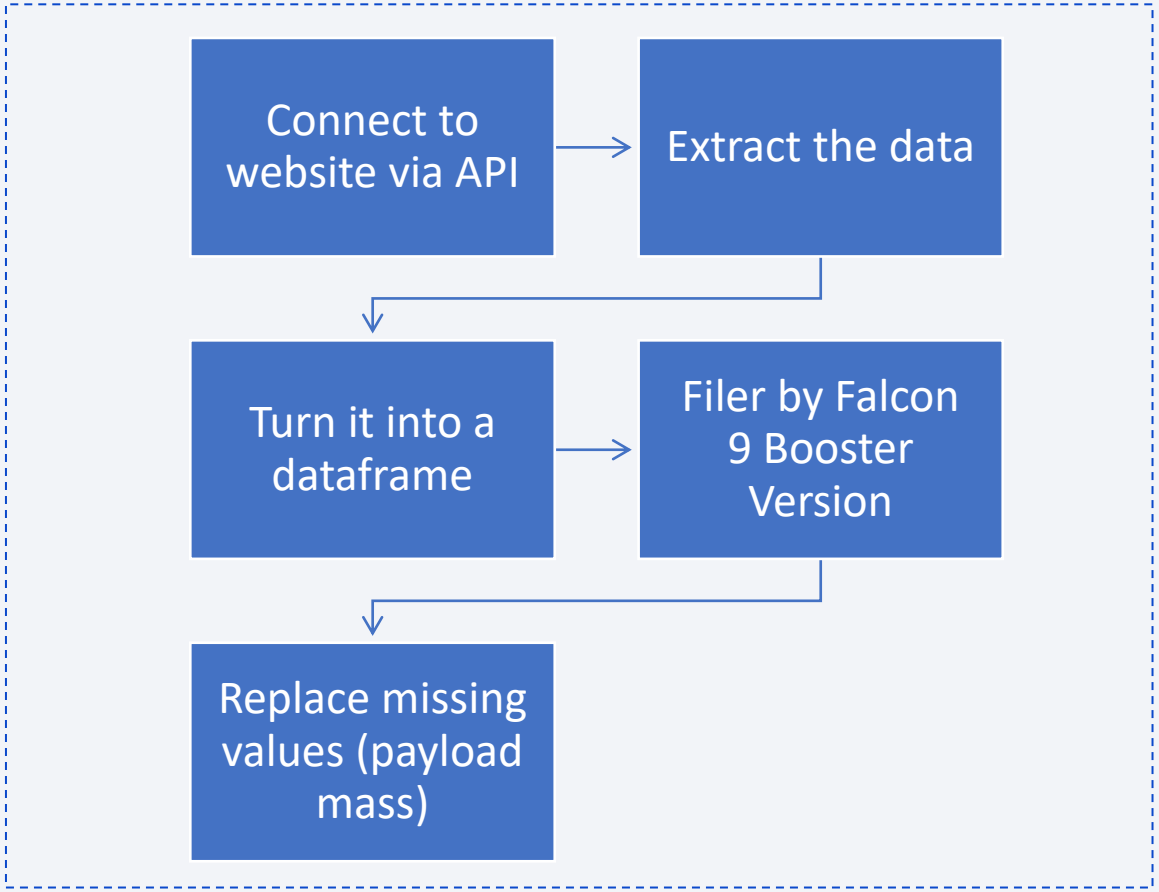
    For some records, Payloadmass was missing. We filled it with the mean Payloadmass of Falcon 9 version.

  - Wikipedia's Falcon 9 launches website, the version as of June 9th (Available here)

    Useful to get the details about the outcomes of the different launches and landing if existing

Data Collection – SpaceX API

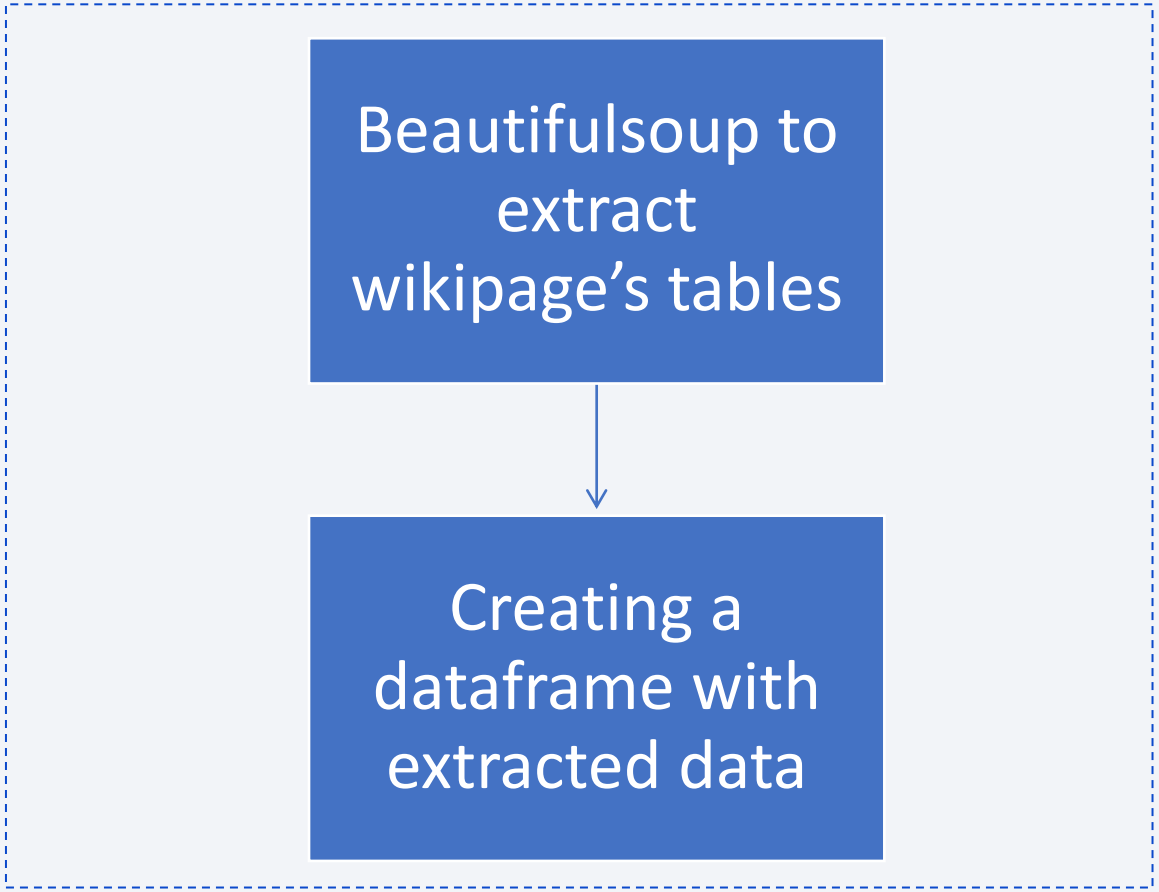
- Data was collected directly from SpaceX website using an API
- GitHub URL available here



Data Collection - Scraping

- Webscraping was applied to Wikipedia's Falcon 9 launches website (June 9th 2021 version, available here), in order to get details about the outcomes of the different launches and its landing if existing

- GitHub URL available here



Data Wrangling

- After I extract the data, the aim here was to understand it and especially the different possible outcomes for a launch to determine whether it was successful or not.

- GitHub URL available here

EDA with Data Visualization

- Data Visualization has been performed in order to detect patterns. Differents scatter plots and graphs were created, such as:
  - Flight Number and Payload Mass
  - Flight Number and Launch Site
  - Payload Mass and Launch Site
  - Success rate by Orbit
  - Flight Number and Orbit
  - Payload Mass and Orbit
  - Success Rate over the years

- GitHub URL available here

EDA with SQL

- SQL queries performed to display (GitHub URL for queries available here):

1. Names of the unique launch sites in the space mission	6. Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
2. 5 records where launch sites begin with the string 'CCA'	7. Total number of successful and failure mission outcomes
3. Total payload mass carried by boosters launched by NASA (CRS)	8. Names of the booster_versions which have carried the maximum payload mass (with a subquery)
4. Average payload mass carried by booster version F9 v1.1	9. Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
5. Date when the first successful landing outcome in ground pad was acheived.	10. Ranked 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

Build an Interactive Map with Folium

- In the folium map, the location of the launch sites were added. We can osberve that they are almost all in the Eastcoast, in Florida.
  - For each launch site, we add colored markers to identify if the launch was successful or not (green for success, red for failure).
  - Then we locate the closest highway, railway and city

- GitHub URL available here

Build a Dashboard with Plotly Dash

- An interactive dashboard has been created to see the success rate:
  - According to each launch site in a pie chart
  - According to its payload mass and its booster version category

  I aimed to see which launch site has the best success rate

- GitHub URL available here

Predictive Analysis (Classification)

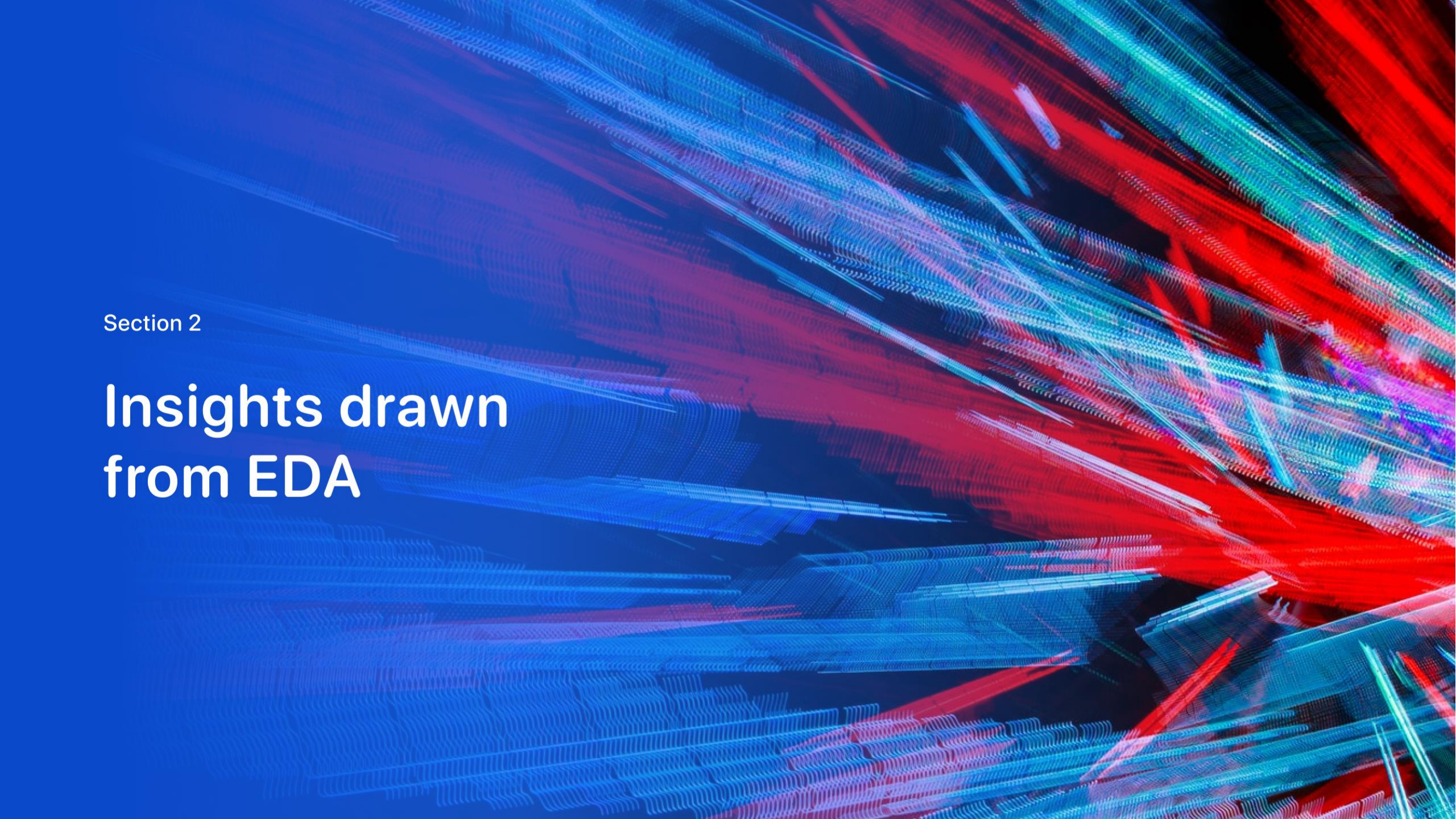
- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- GitHub URL available here

Results

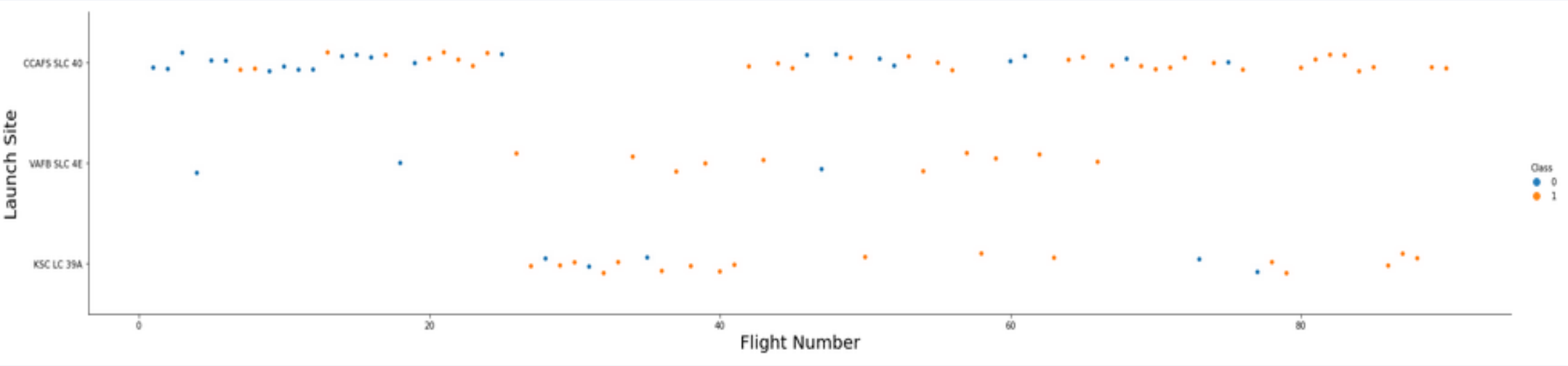
- 67% of the launches landed successfully
- With a payload mass above 9000kg, 87% of successful landing
- For ES-L1, SSO, HEO and GEO orbits, the success rate is 100%, but the payload is under 9000kg, and even under 6000kg
- The success rate increased significantly since 2013 and then 2015
- Predictive analysis results

Section 2

Insights drawn from EDA



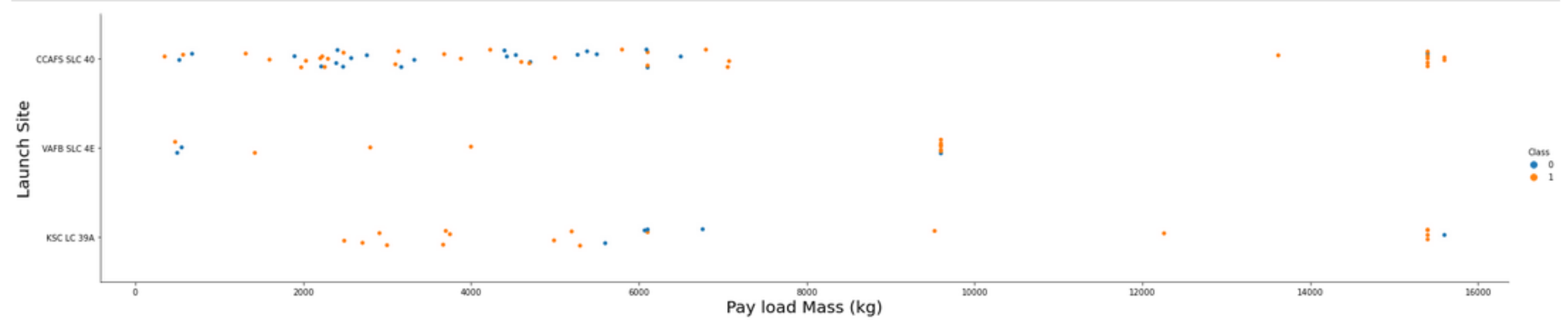
Flight Number vs. Launch Site



- With time, success rate increase. From CCAFS SLC 40 launch site, until flight number 30 a few were successful but the trend has been reverted

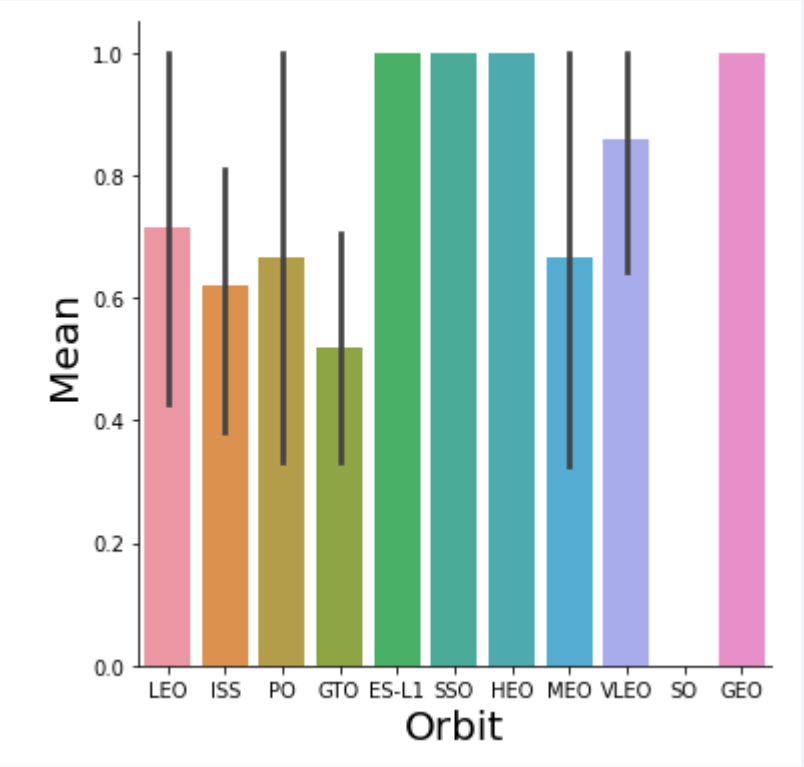
- VAB SLC 4E launch site has a good success rate (77%)

Payload vs. Launch Site



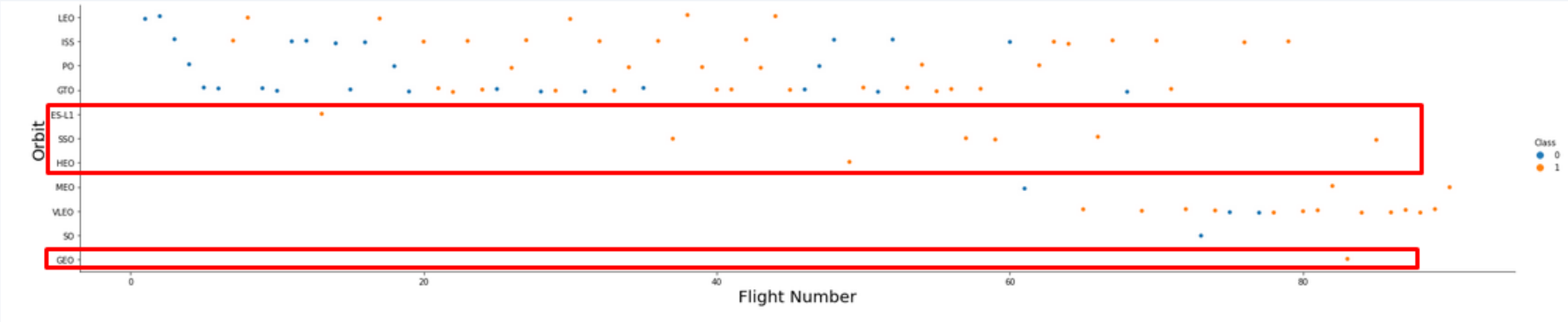
- Indenpendently of the launch site, when payload mass is above 9000kg, the success rate is much higher.

Success Rate vs. Orbit Type



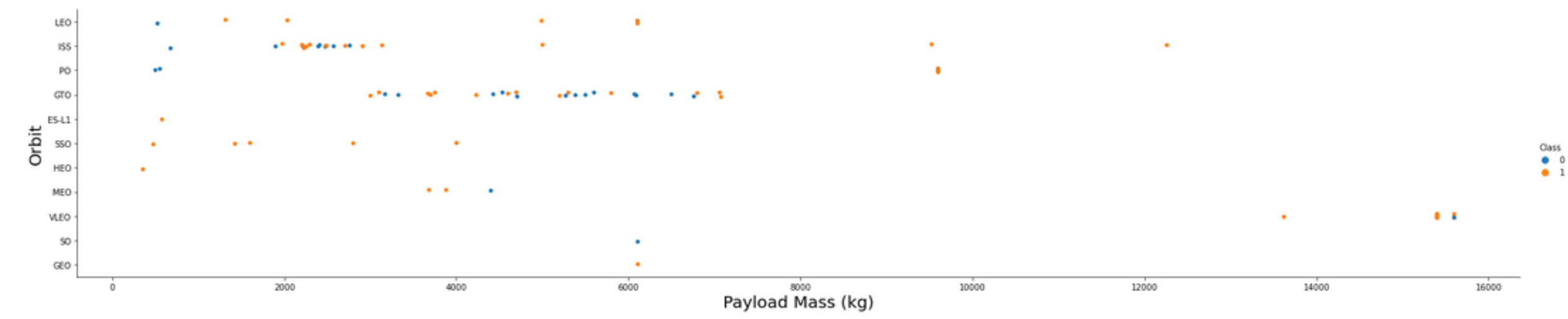
- When orbit is either ES-L1, SSO, HEO or GEO, the success rate is 100%.

Flight Number vs. Orbit Type



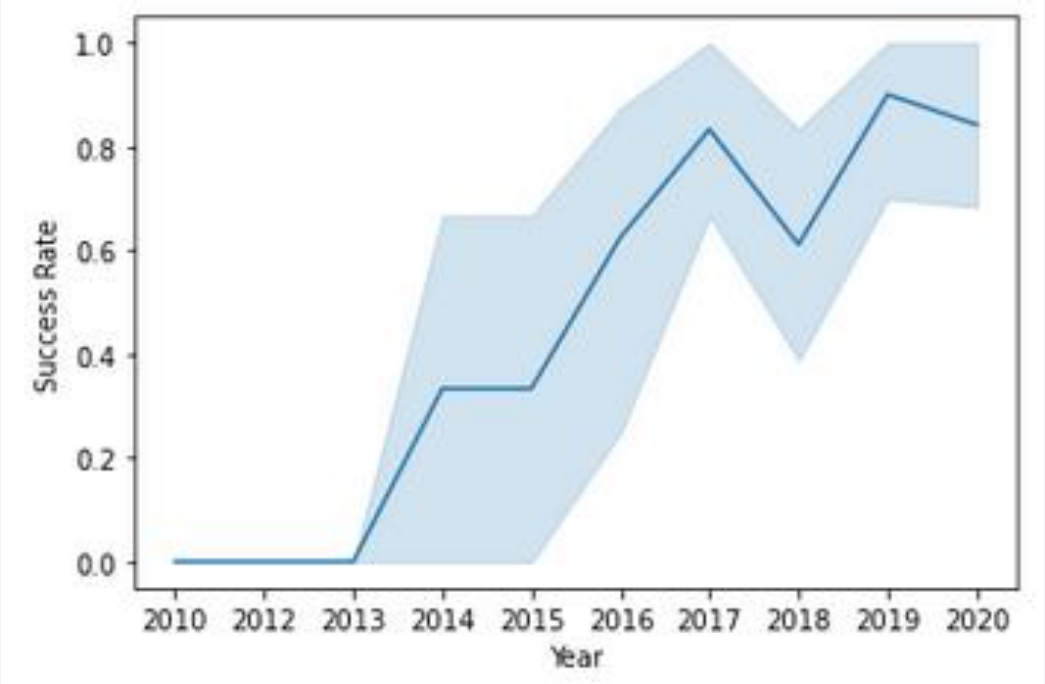
- The scatter plot shows the barchart explained in the previous slide for the mentioned orbit. In this chart, we can see that there were few launches (5 for SSO orbit and 1 for the others), which partly explains the high successful rate

Payload vs. Orbit Type



- The above-mentioned orbits with 100% successful did not respond to a high payload which we saw has a good successful rate

Launch Success Yearly Trend



- Success Rate increased with time, especially since 2013 and then 2015.

All Launch Site Names

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

- Database only has records for these launch site

Launch Site Names Begin with 'CCA'

Out[29]:

DATE	time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- 5 first records of the database where launch site was CCAFS (either LSC or LC)

Total Payload Mass

- In total, the NASA has a payload mass of 45.596kg in all launches they made.

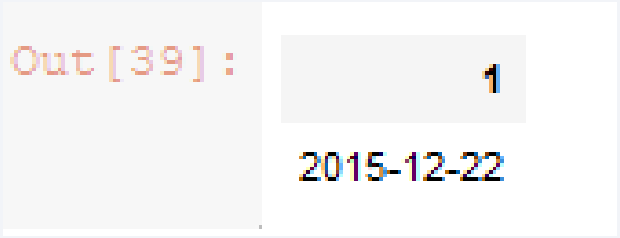
Average Payload Mass by F9 v1.1

Out[8]:

	1
	2928

- For booster version F9 v1.1, the average payload mass is 2.928 kg

First Successful Ground Landing Date



- The date of the first successful landing outcome on ground pad was on December 22th of 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

Out[43]:

booster_version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

- Those are the booster version with a successful drone ship landing, when the payload was between 4000 and 6000kg

Total Number of Successful and Failure Mission Outcomes

Number of Successful outcome	Number of failure outcome
100	1

- The mission outcome is successful for almost 100% of the time. Of 101 attempt, SpaceX had just 1 failure

Boosters Carried Maximum Payload

Out[73]:

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- The max payload is 15.600kg, and we are showing the booster version that carried on this payload. We can observe that they all from B5 category

2015 Launch Records

Out[80]:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- In 2015, the booster version with a failed landing at drone ship are those two mentioned above. Both were launched from CCAFS LC-40 site.

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

Out[13]:

landing__outcome	number_of_ocurrence
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- The list above is the ranked landing outcome for launches between 2010-06-04 and 2017-03-20. Almost 50% of the time, it was a failure (no attempt + failure drone ship)

Section 4

Launch Sites Proximities Analysis

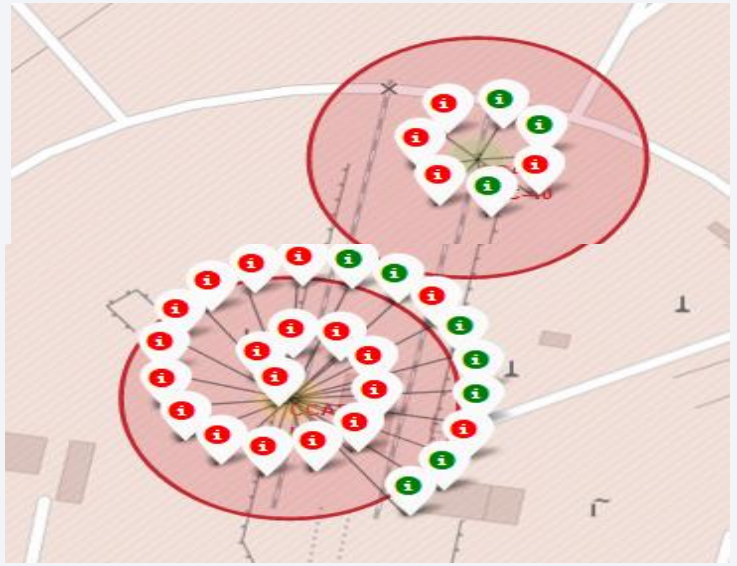
Location of Launch Sites

- We observe that all the launch sites are located in the US coasts. 1 in the West Coast (with 10 launches) and 3 in the East Coast (46 launches). Most of the launches are made in the East Coast

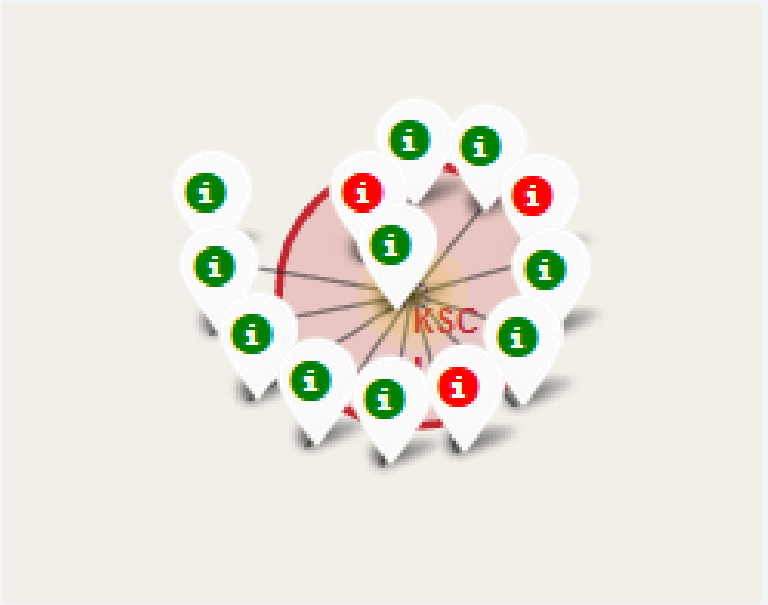
Outcome of landing in each launch site

VAFB SLC-4E outcomes

CCAFS outcomes



KSC LC outcomes



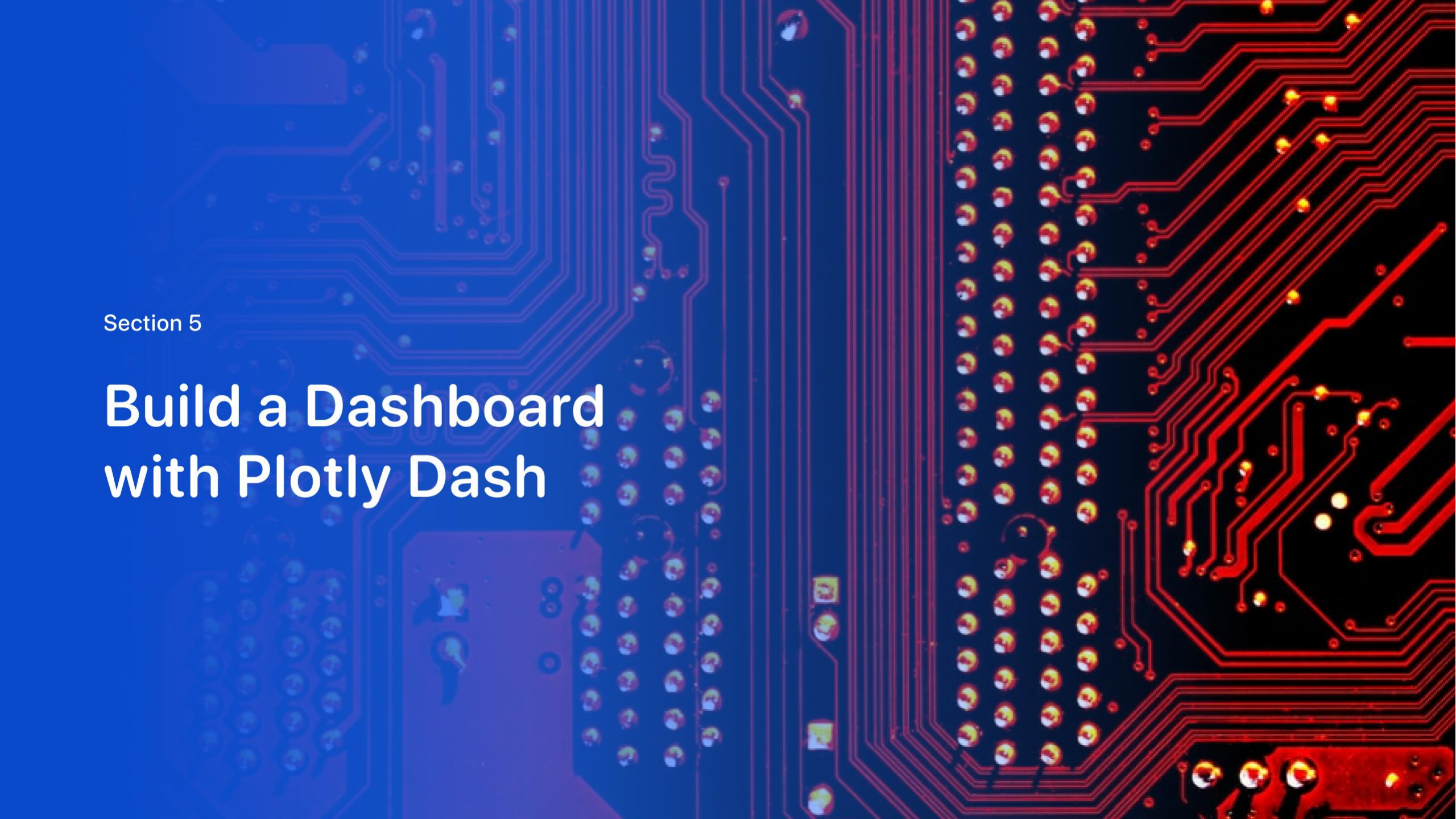
- KSL LC have much higher success rate than the other 3 launch sites
- Between both CCAFS launch sites, SLC-40 (above) performs better

Distances of VAFB launch site to possible crowded locations

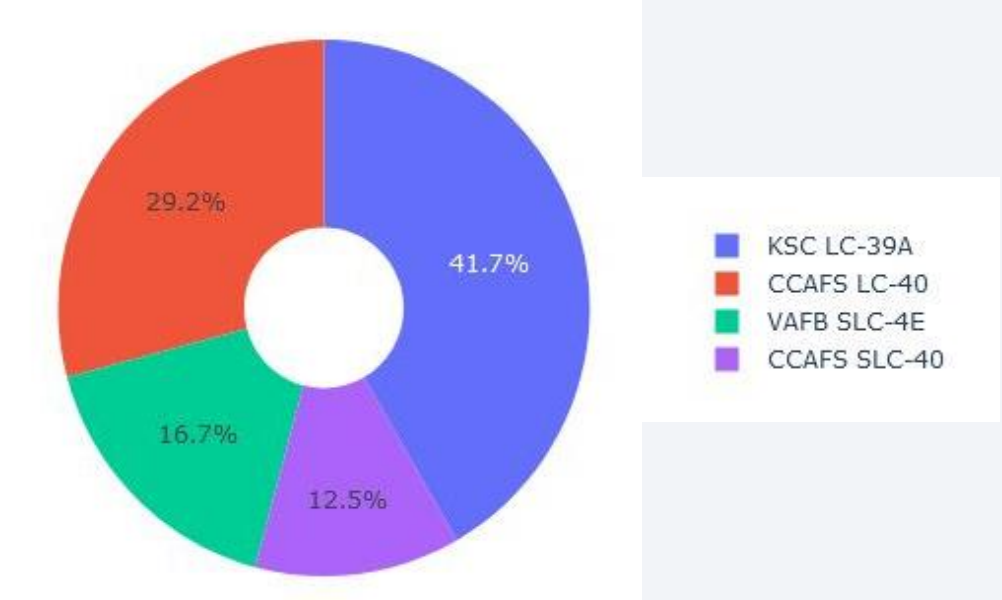
- VAFB launch site is at few kilometers from Lompoc city (14 km) which has an airport, and the highway is located 6.59km from the launch site. These distance can be taken into account for security when a rocket is launched.

Section 5

Build a Dashboard with Plotly Dash

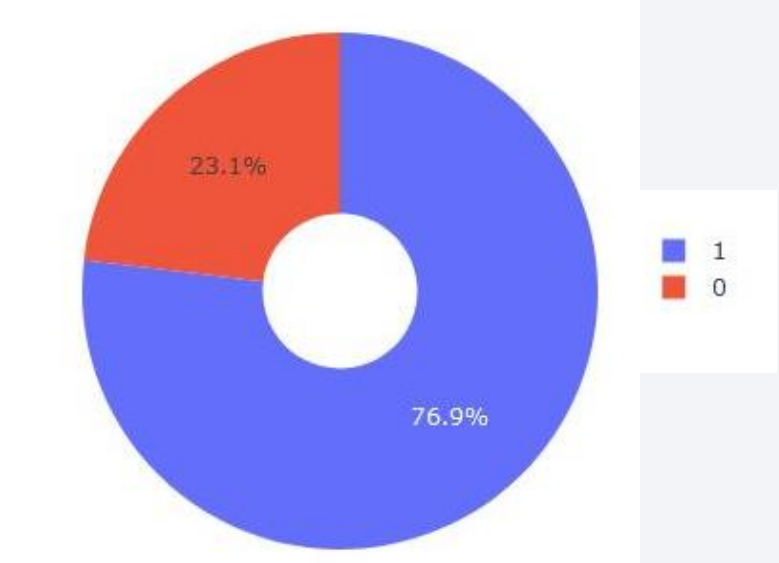


Successful launches by launch site



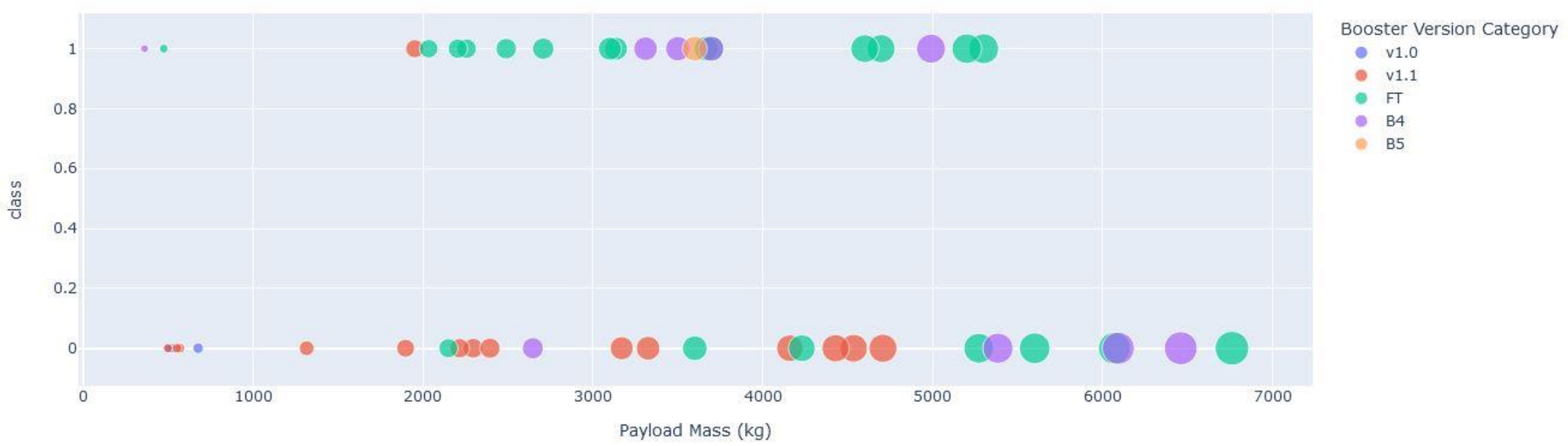
- The most successful launch site is KSL LC-39A with 41.7% of all successful launches. More than 4 of 10 successful launches were made at this launch site

Success Rate at KSC LC 39A



- For KSC LC-39A launch site, 76,9% of launches were successful

Outcome of Launches by Booster Version Category and Payload

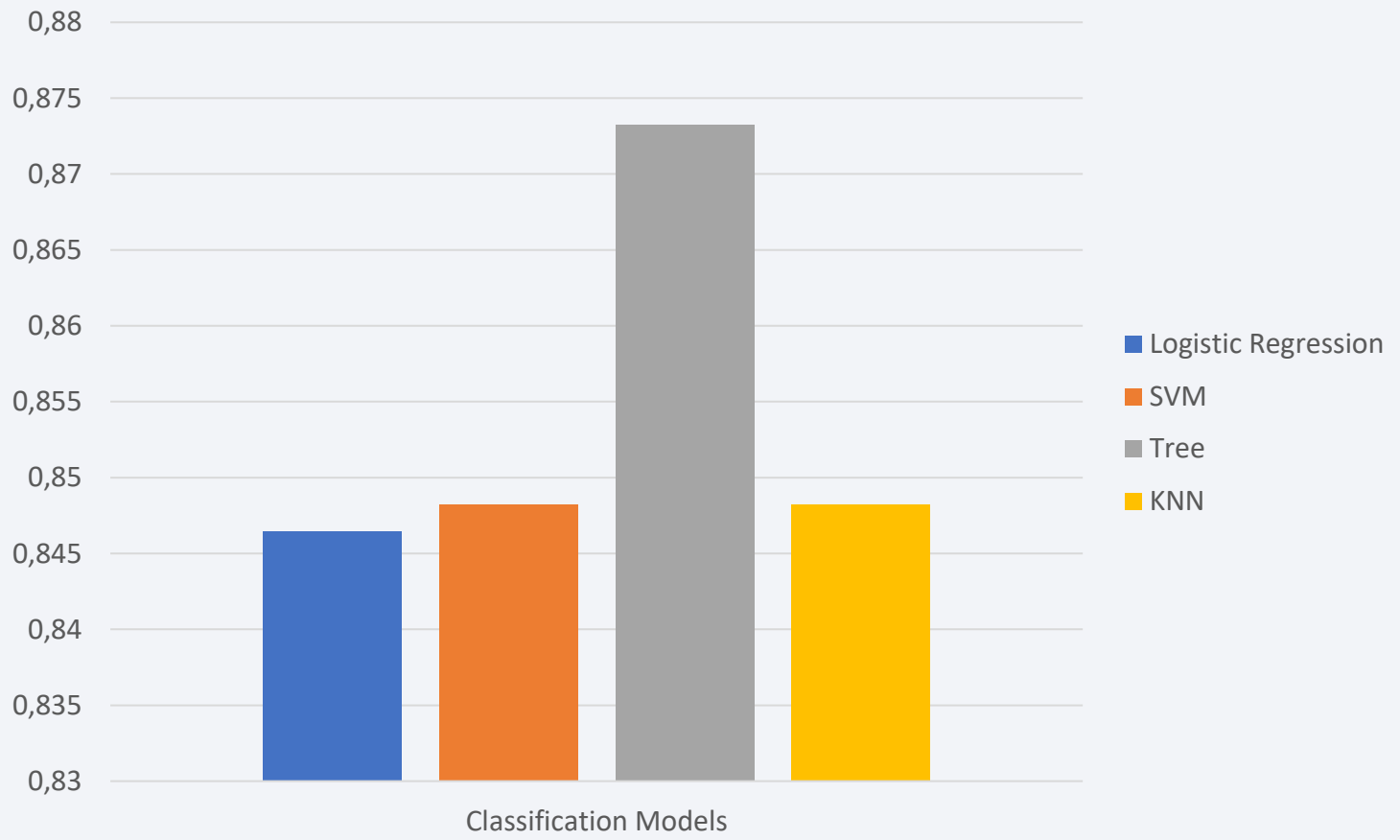


- The booster version category with highest number of successful launches is the FT category
- When payload mass is below 2000 kg or over 6000kg, the launch will most likely fail

Section 6

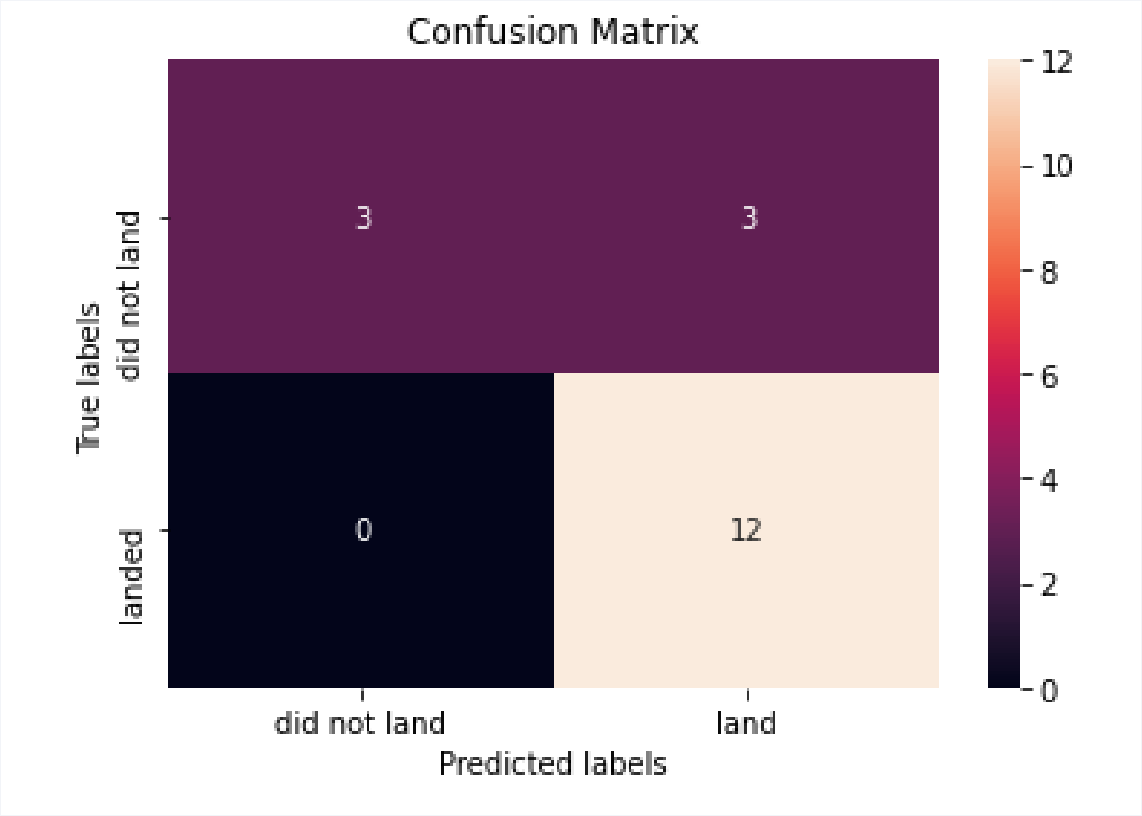
Predictive Analysis (Classification)

Classification Accuracy



- The tree model has the best accuracy among the 4 models tested.

Confusion Matrix



- Confusion matrix shows the number of false estimation of the Tree model. X-axis shows predicted label and Y-axis the true label. We only have 3 wrong predictions from the tree model, which predicted the rocket would land when it did not.

Conclusions

- Even if all the tested models have good accuracy (+84%), the tree model has a much better accuracy with +87%, making it the model to use in order to predict future outcome
- For future outcome prediction, we would use the tree model to have a better accuracy of prediction
- We can take into account the insights of the previous slides to increase the chances of success such as KSC launch site, +9000kg payload, FT booster version etc.

Thank you!