

Winning Space Race with Data Science

Rogério Yamada
22 September 2025

1 Executive Summary

2 Introduction

3 Methodology

4 Results

5 Conclusion

6 Appendix

Executive Summary

Introduction



Section 1

Methodology

- Data collection methodology
 - Data about rocket launches was obtained from a SpaceX API and web scraping Wikipedia pages
- Perform data wrangling
 - Missing data was handled, a preliminary Exploratory Data Analysis was performed, and the variable Outcome Class was defined for training the supervised models
- Perform Exploratory Data Analysis (EDA) using visualisation and SQL
- Perform interactive visual analysis using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

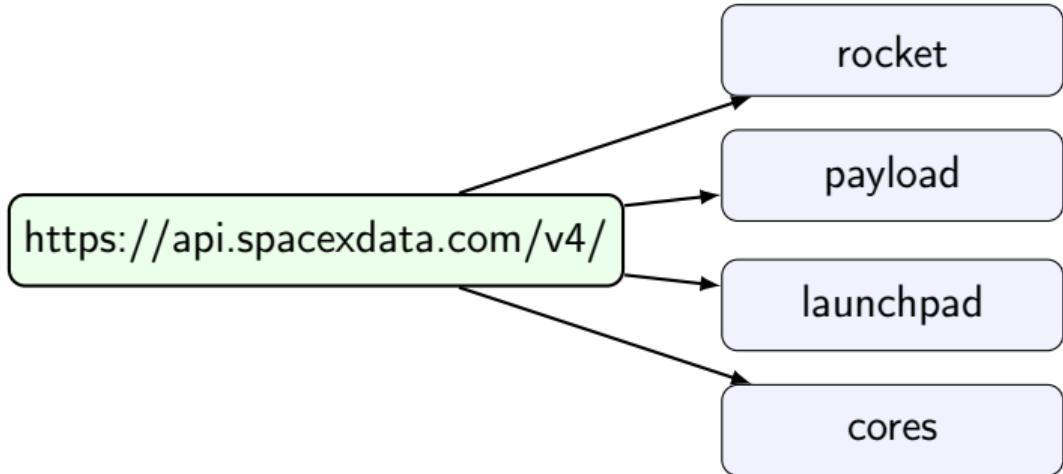
- SpaceX API data extraction and Wikipedia pages web scraping were combined to produce a dataset of SpaceX Falcon 9 landings information

11 rows ✓ 90 rows x 17 cols
[34]

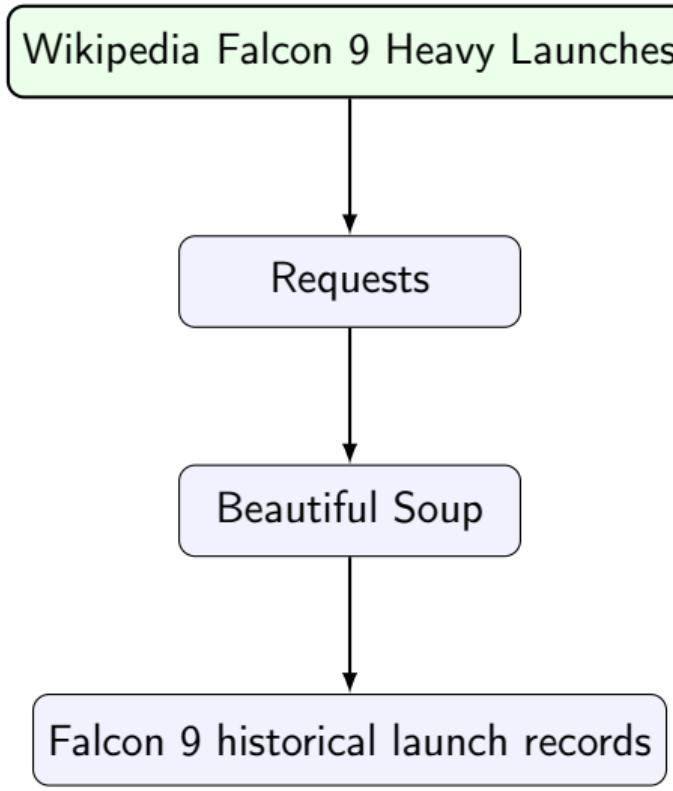
#	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2010-06-04	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
1	2	2012-08-22	Falcon 9	528.000000	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
3	4	2013-09-29	Falcon 9	508.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1093	-120.01829	34.632093
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...
85	86	2020-09-03	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	2	True	True	True	Se9e3032383ecb6bb234e7ca	5.0	12	B1068	-80.603956	28.608058
86	87	2020-10-06	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	3	True	True	True	Se9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
87	88	2020-10-18	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	6	True	True	True	Se9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
88	89	2020-10-24	Falcon 9	15600.000000	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	Se9e3033583ecbb9e534e7cc	5.0	12	B1068	-80.577366	28.561857
89	90	2020-11-05	Falcon 9	3681.000000	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	Se9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

Data Collection – SpaceX API

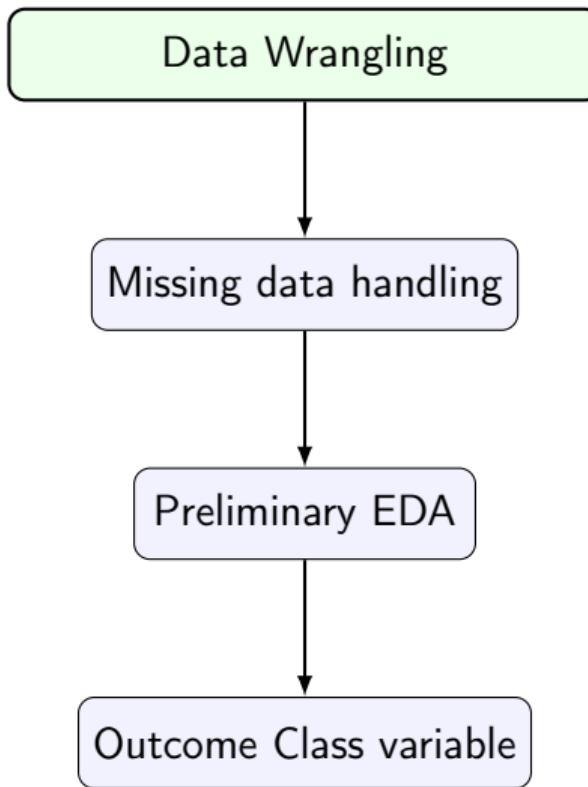
- From the SpaceX API endpoint <https://api.spacexdata.com/v4/> we probed the following data sources:
 - rocket
 - payload
 - launchpad
 - cores
- Jupyter Notebook's GitHub URL



- From the Wikipedia List of Falcon 9 and Falcon Heavy launches web page we collected Falcon 9 historical launch records
- Jupyter Notebook's GitHub URL



- Through data wrangling, the variable Outcome Class was defined for training the supervised models
- [Jupyter Notebook's GitHub URL](#)



- Several SQL queries have been processed to gain insights about the landing outcomes:
 - Launching sites
 - Total payload mass carried by specific boosters in specific sites
 - Successful and failed landing outcomes
- Jupyter Notebook's GitHub URL

- Several charts have been drawn to gain insights about the landing outcomes:
 - Flight Number versus Launch Site by Class
 - Payload Mass versus Launch Site by Class
 - Success Rate by Orbit
 - Flight Number versus Orbit by Class
 - Payload Mass versus Orbit by Class
 - Yearly Launch Success Rate
- Jupyter Notebook's GitHub URL

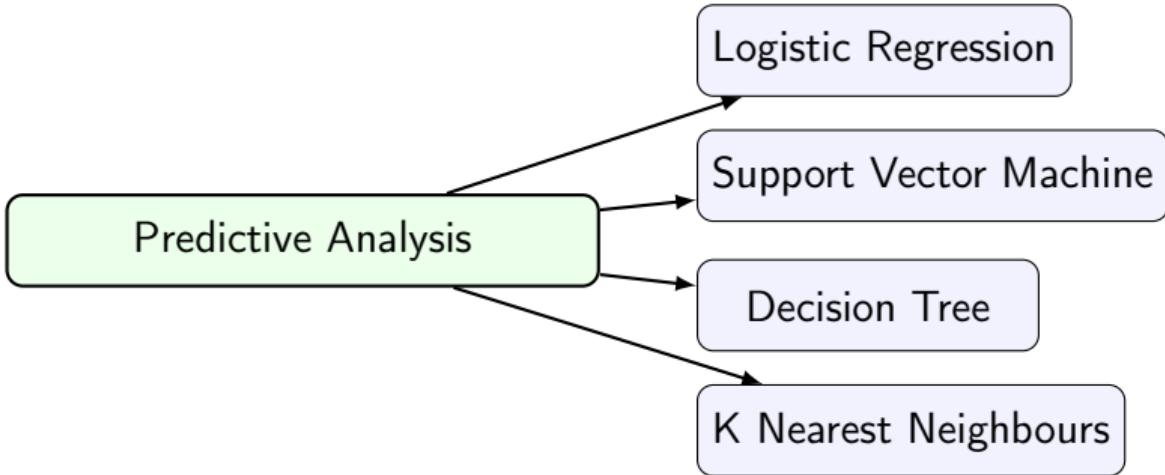
Build an Interactive Map with Folium

- A geographical analysis has been performed to gain insights about the dependencies of landing outcomes and location and surrounding of launching sites:
 - Marking all launch sites on a map
 - Marking the success/failed launches for each site on the map
 - Calculating the distances between a launch site to its proximities
- Jupyter Notebook's GitHub URL

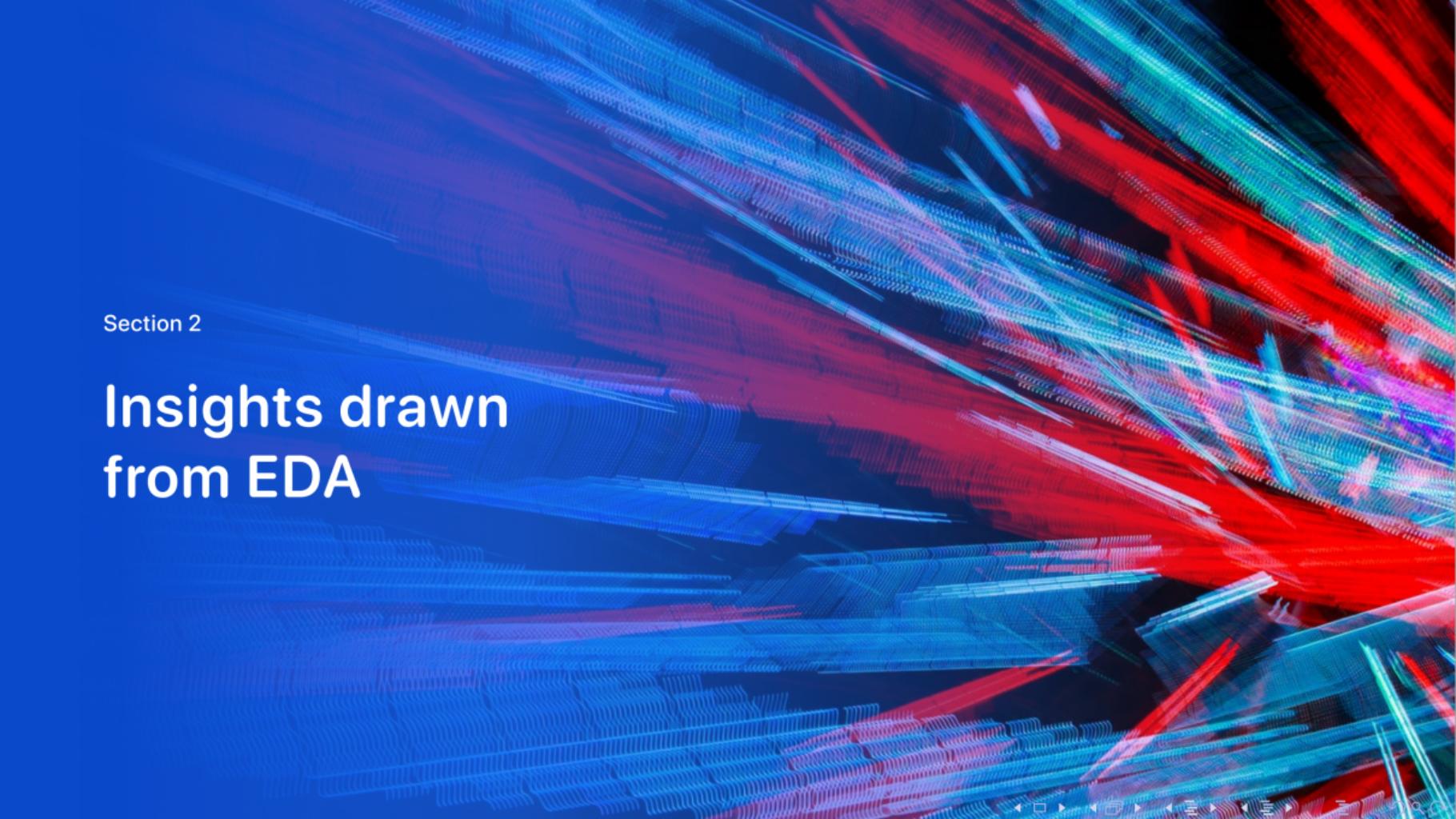
- A dashboard has been implemented to perform real-time analysis about landing outcomes considering:
 - Launch site drop-down menu
 - Interactive successful landing outcome pie chart
 - Range slider for selecting payload mass
 - Interactive successful landing outcome scatter plot
- Jupyter Notebook's GitHub URL

Predictive Analysis (Classification)

- The following predictive models have been considered:
 - Logistics Regression
 - Support Vector Machine
 - Decision Tree
 - K Nearest Neighbours
- Jupyter Notebook's GitHub URL



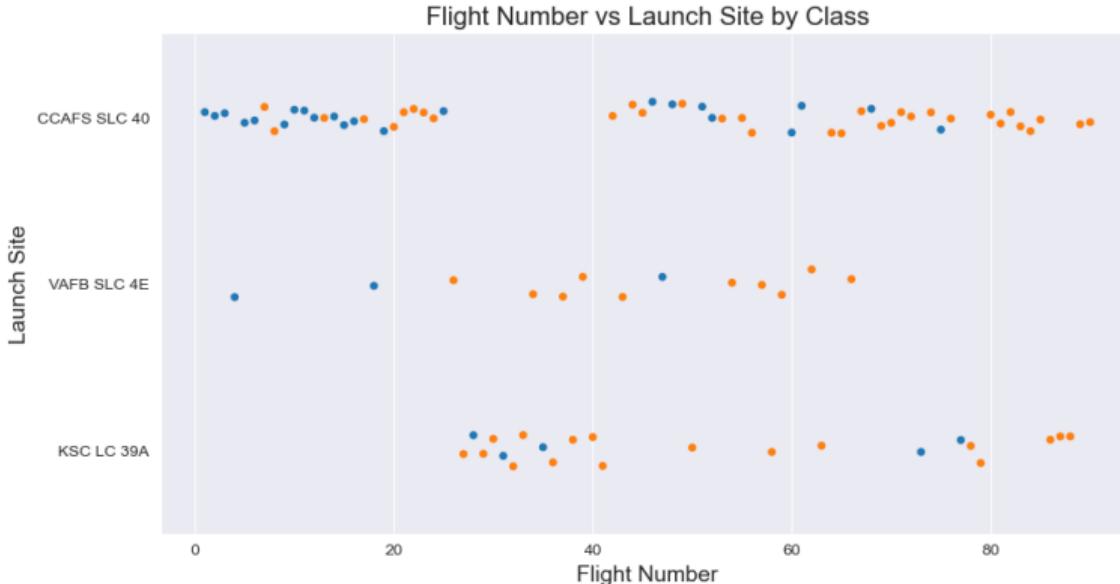
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

The background of the slide features a complex, abstract pattern of glowing lines in shades of blue, red, and green. These lines are arranged in a way that suggests depth and motion, resembling a digital or quantum landscape. The overall effect is futuristic and dynamic.

Section 2

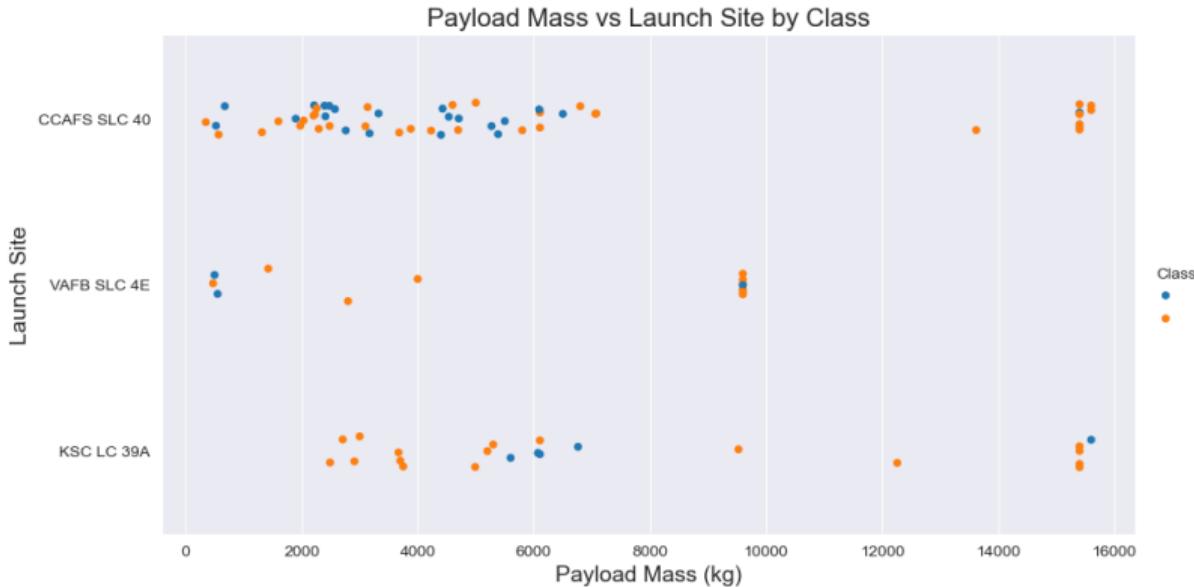
Insights drawn from EDA

Flight Number versus Launch Site



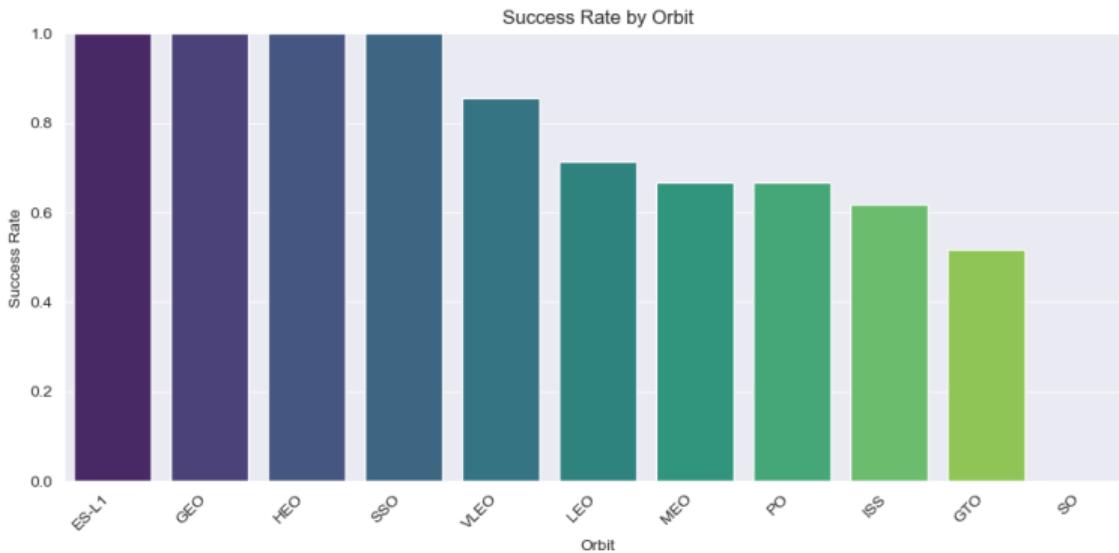
- There is a concentration of unsuccessful landing outcomes in the site CCAFS SLC 40
- As the number of flights increase, the successful landing outcomes are likely to increase

Payload versus Launch Site



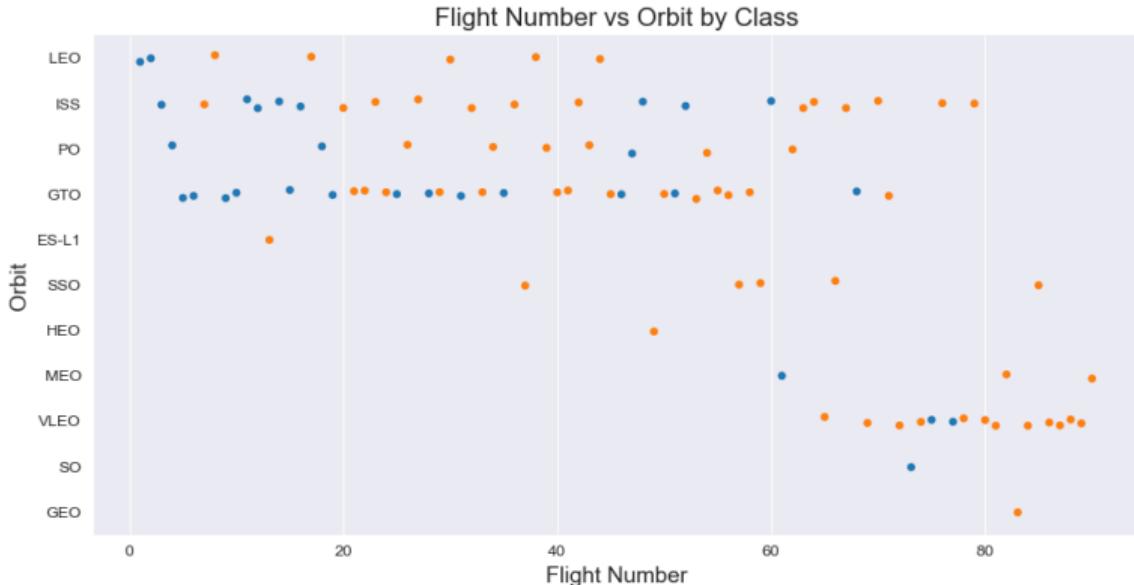
- There is a concentration of unsuccessful landing outcomes in CCAFS SLC 40 for payloads lighter than 7000 kg

Success Rate versus Orbit Type



- The success rate is higher when the orbit type is ES-L1, GEO, HEO, and SSO
- The success rate is lower when the orbit type is GTO and SO

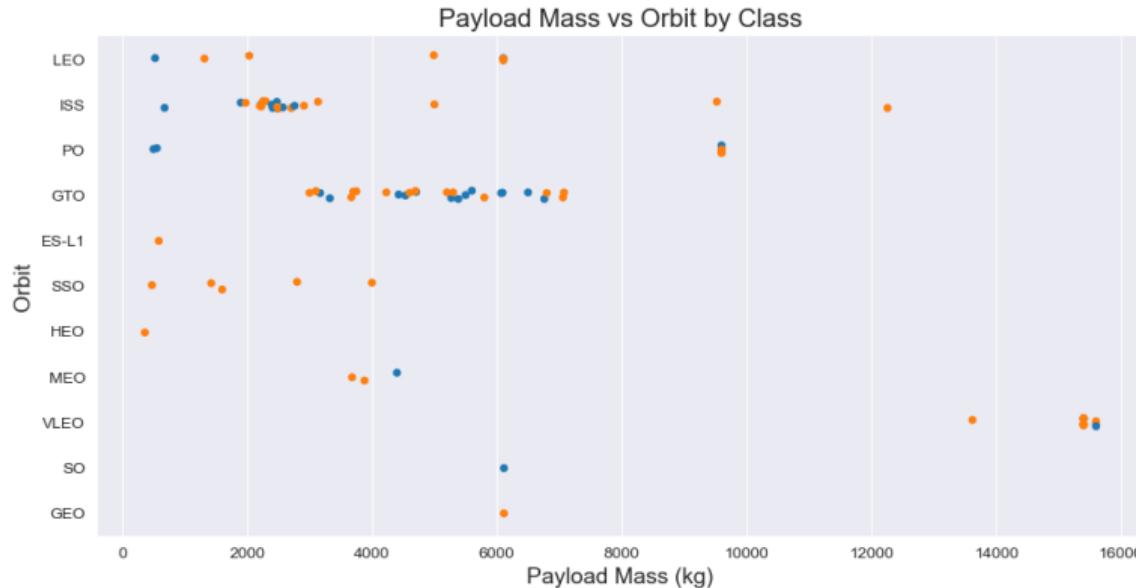
Flight Number versus Orbit Type



- The success rate is higher as the number of flights increases for orbit type LEO, ISS, and PO
- There seems to be no relationship between flight number when the orbit type is GTO



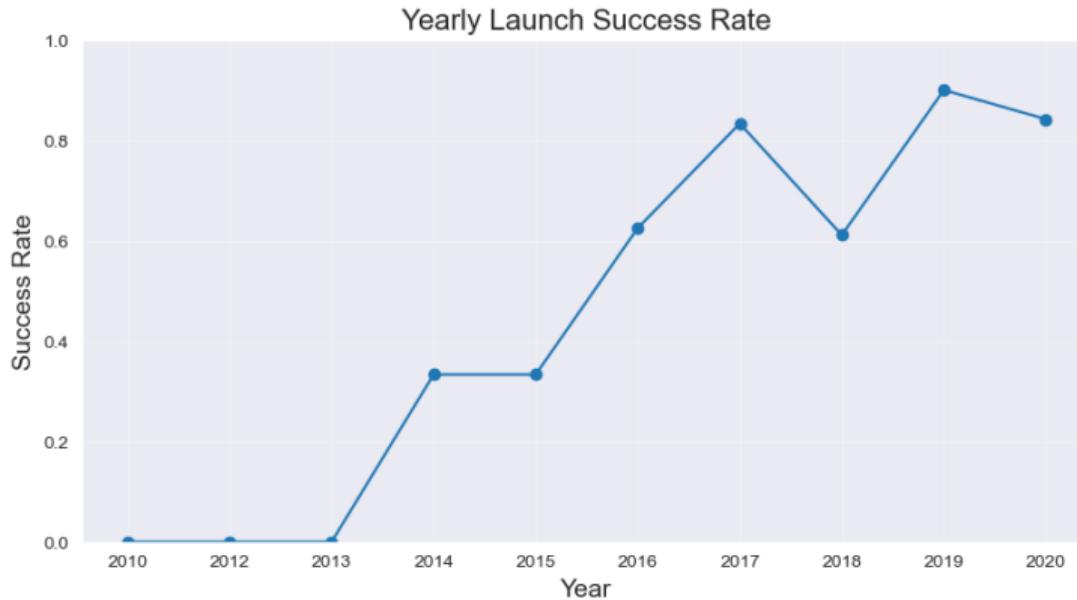
Payload versus Orbit Type



- With heavy payloads, the successful landing rate is higher for PO, LEO, and ISS orbit types
- However, regarding GTO, we cannot distinguish this well as both positive landing rate and negative landing (unsuccessful mission) occur almost evenly



Launch Success Yearly Trend



- The success rate since 2013 kept increasing until 2017 (stable in 2014) and after 2015 it started increasing

All Launch Site Names

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

- This is a sorted list of unique, non-null values from the Launch_Site column in the SPACEXTABLE table

Launch Site Names Begin with 'CCA'

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- This is a set of up to 5 rows from the SPACEXTABLE table where the Launch_Site column starts with the text "CCA" (e.g., "CCAFS ..."). The LIKE 'CCA%' filter matches any value beginning with "CCA"

Total Payload Mass = 45596

- This is the total payload mass (sum of PAYLOAD_MASS__KG_) for all rows in SPACEXTABLE where the Customer is exactly 'NASA (CRS)'. The result is a single value labeled total_payload_mass

Average Payload Mass by F9 v1.1

Average Payload Mass = 2534.666666666665

- This is the average payload mass (PAYLOAD_MASS__KG_) over rows in SPACEXTABLE whose Booster_Version starts with “F9 v1.1”, returning a single value named avg_payload_mass

First Successful Ground Landing Date

First Ground Pad Landing Success Date = 2015-12-22

- This is the earliest Date (minimum) among rows where Landing_Outcome is exactly "Success (ground pad)", labeled first_ground_pad_success_date.

Successful Drone Ship Landing with Payload between 4000 and 60000

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

- These are the unique booster versions (as booster_name) that had a “Success (drone ship)” landing and carried payloads strictly between 4000 kg and 6000 kg, sorted alphabetically

Total Number of Successful and Failure Mission Outcomes

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

- These are rows grouped by Mission_Outcome, counting how many are successes or failures (strings starting with “Success” or “Failure”), and returns each outcome with its count, sorted by count descending

Boosters Carried 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

- These are the unique booster versions that carried the maximum payload mass found in the table. The subquery gets the global `MAX(PAYLOAD__MASS__KG_)`, and the outer query lists distinct `Booster_Version` rows matching that value

2015 Launch Records

month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- This is a list of 2015 records where Landing_Outcome is “Failure (drone ship)”

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

Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precudled (drone ship)	1

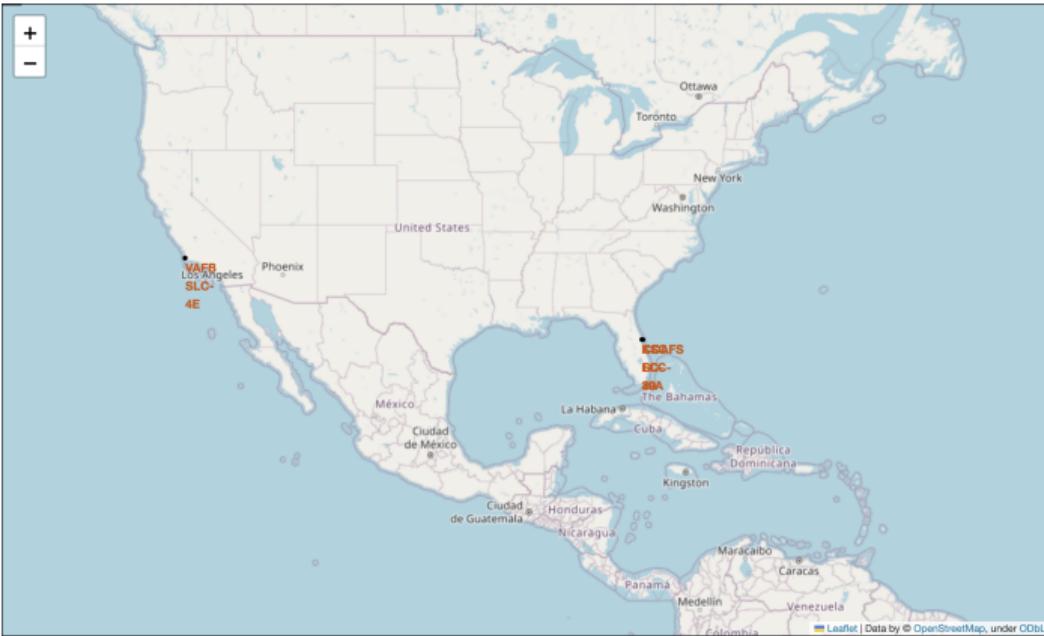
- This is a calculation of how many launches fall into each `Landing_Outcome` between 2010-06-04 and 2017-03-20, returning each outcome with its count, sorted by count descending

The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth against a dark blue-black void of space. City lights are visible as numerous small white and yellow dots, primarily concentrated in coastal and urban areas. In the upper right quadrant, there is a bright, horizontal band of light, likely the Aurora Borealis or Southern Lights. The overall atmosphere is dark and mysterious.

Section 4

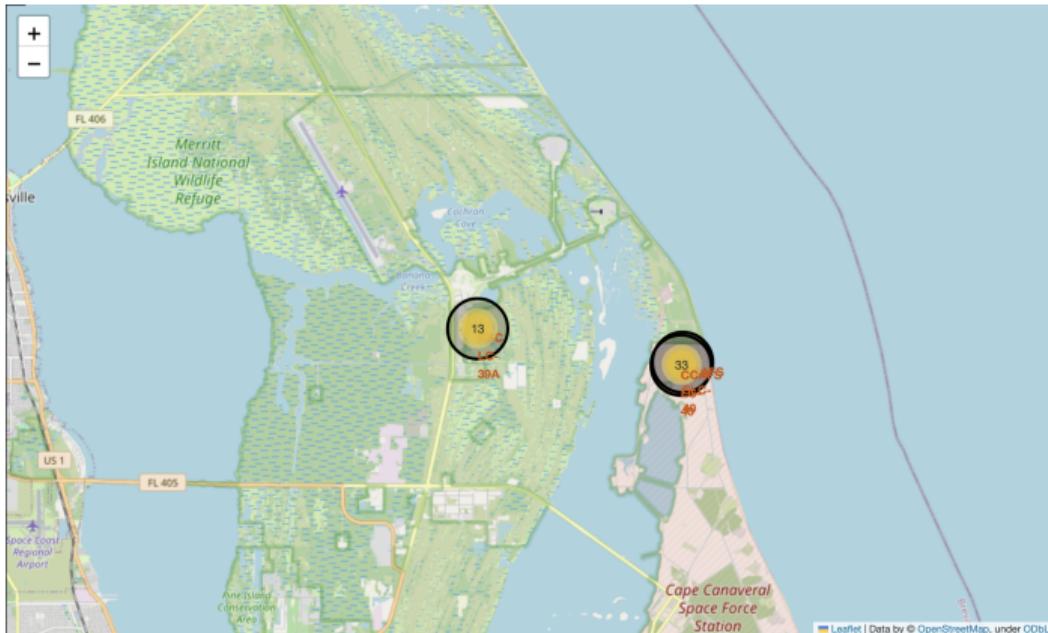
Launch Sites Proximities Analysis

All launch sites map



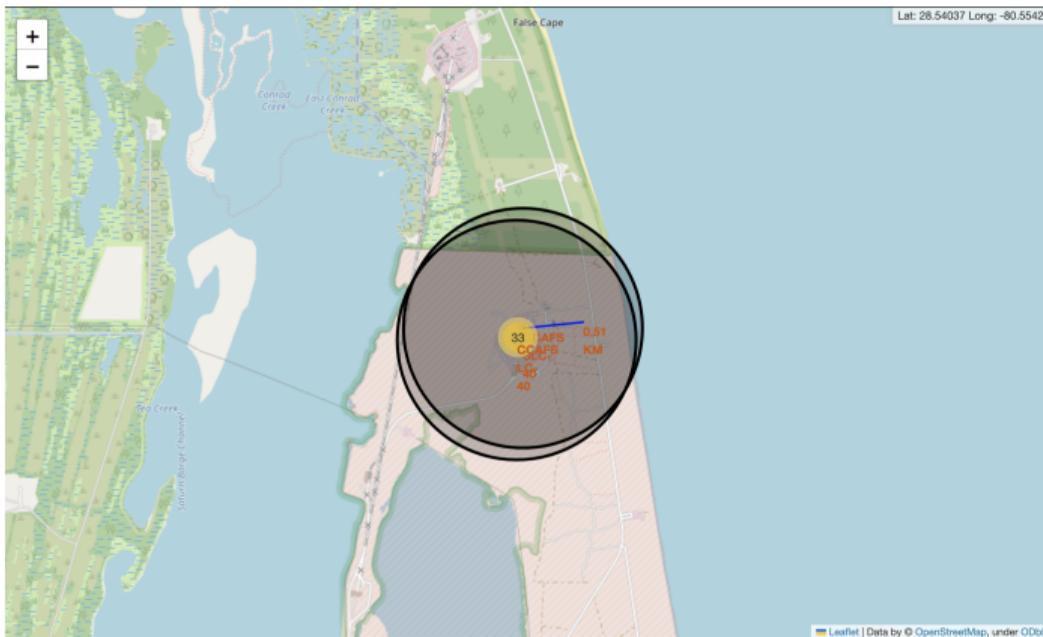
- All launch sites are plotted on the map

Success/Failed launches



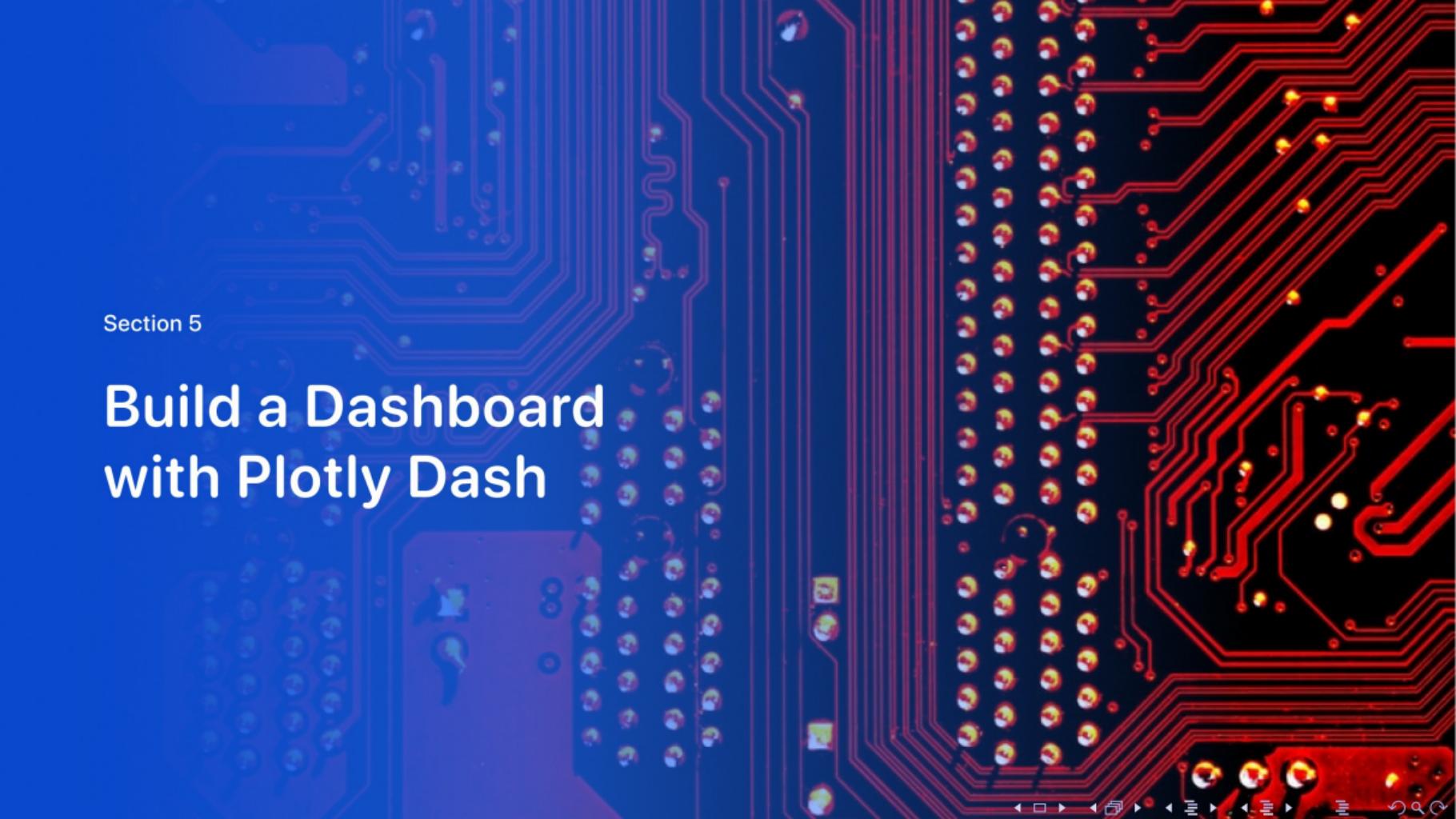
- The successful and failed launches are clustered on the map

Coastline point map



- There is a distance of 0.51 km between the CCAFS SLC-40 launch site and the coastline



The background of the slide features a close-up photograph of a printed circuit board (PCB). The board is primarily black, with intricate red and blue patterns of conductive tracks and components. A vertical column of circular pads is visible on the left side. The overall aesthetic is high-tech and futuristic.

Section 5

Build a Dashboard with Plotly Dash

Launch site drop-down menu

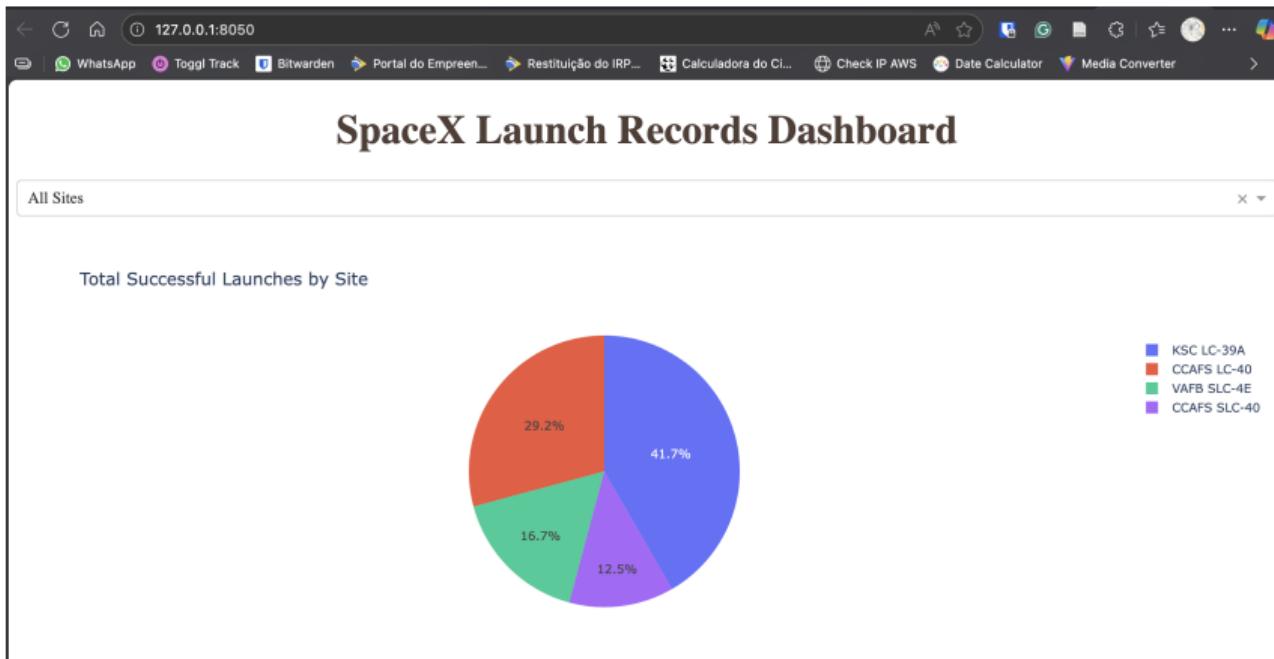
SpaceX Launch Records Dashboard

All Sites

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

- Launch site drop-down menu

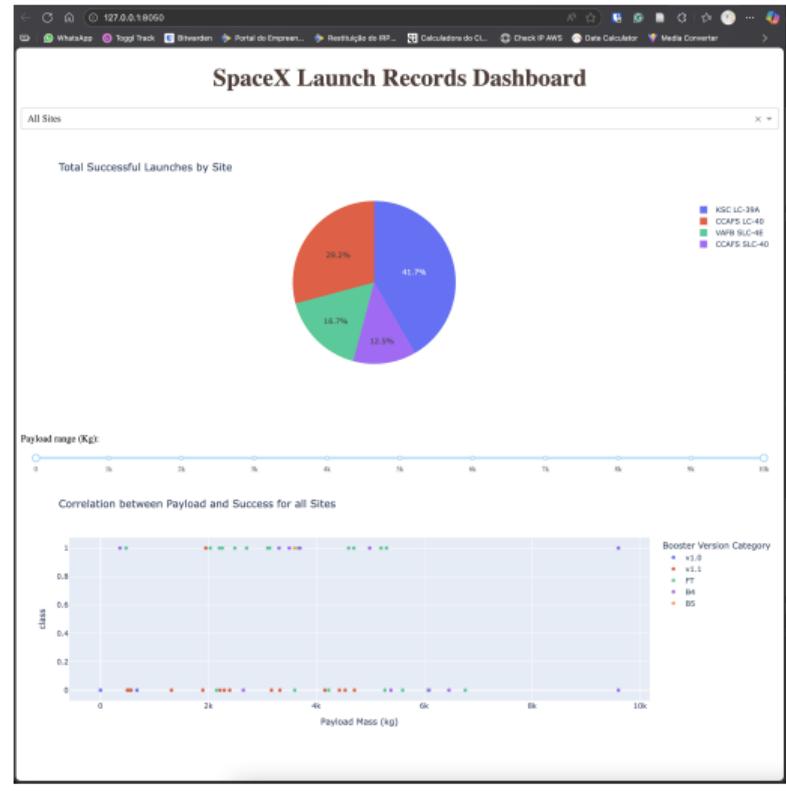
Interactive successful landing outcome pie chart



• Interactive successful landing outcome pie chart

Complete Dashboard

- This is the complete dashboard with the following elements:
 - Launch site drop-down menu
 - Interactive successful landing outcome pie chart
 - Range slider for selecting payload mass
 - Interactive successful landing outcome scatter plot



The background of the slide features a dynamic, abstract design composed of several thick, curved lines. These lines are primarily in shades of blue and yellow, creating a sense of motion and depth. The curves are smooth and organic, resembling the path of a train or a high-speed vehicle through a tunnel. The overall effect is modern and professional.

Section 6

Predictive Analysis (Classification)

Classification Accuracy

Confusion Matrix

Conclusions

-

Appendix



Thank you!

