

# 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

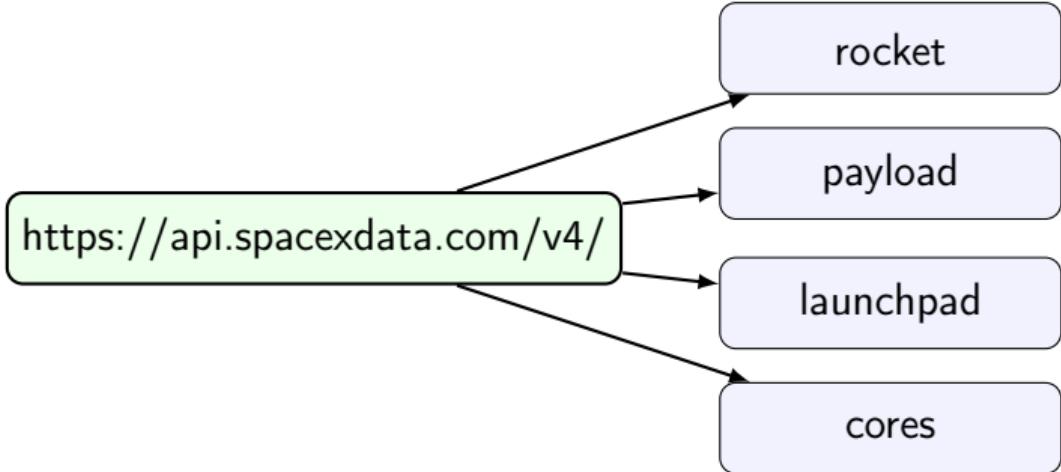
- SpaceX API data extraction and Wikipedia pages web scraping were combined to produce a dataset of SpaceX Falcon 9 landings information

11 rows ✓ 90 rows x 17 cols  
[34]

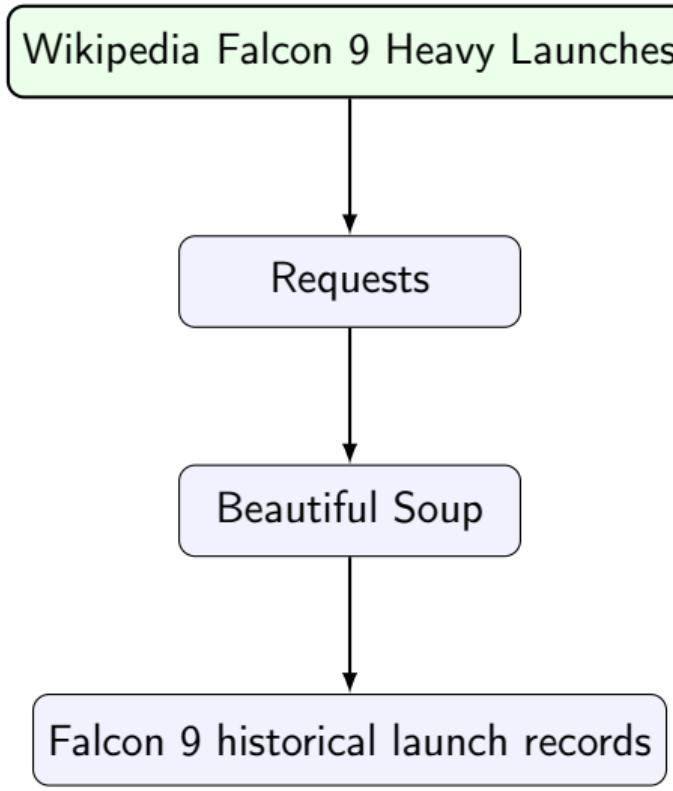
#	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2010-06-04	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
1	2	2012-08-22	Falcon 9	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

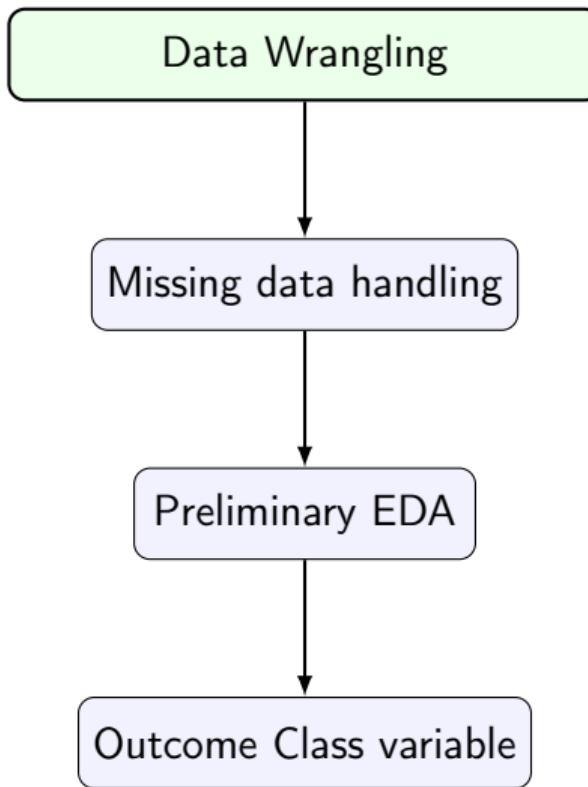
- From the SpaceX API endpoint <https://api.spacexdata.com/v4/> we probed the following data sources:
  - rocket
  - payload
  - launchpad
  - cores
- Jupyter Notebook's GitHub URL



- 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

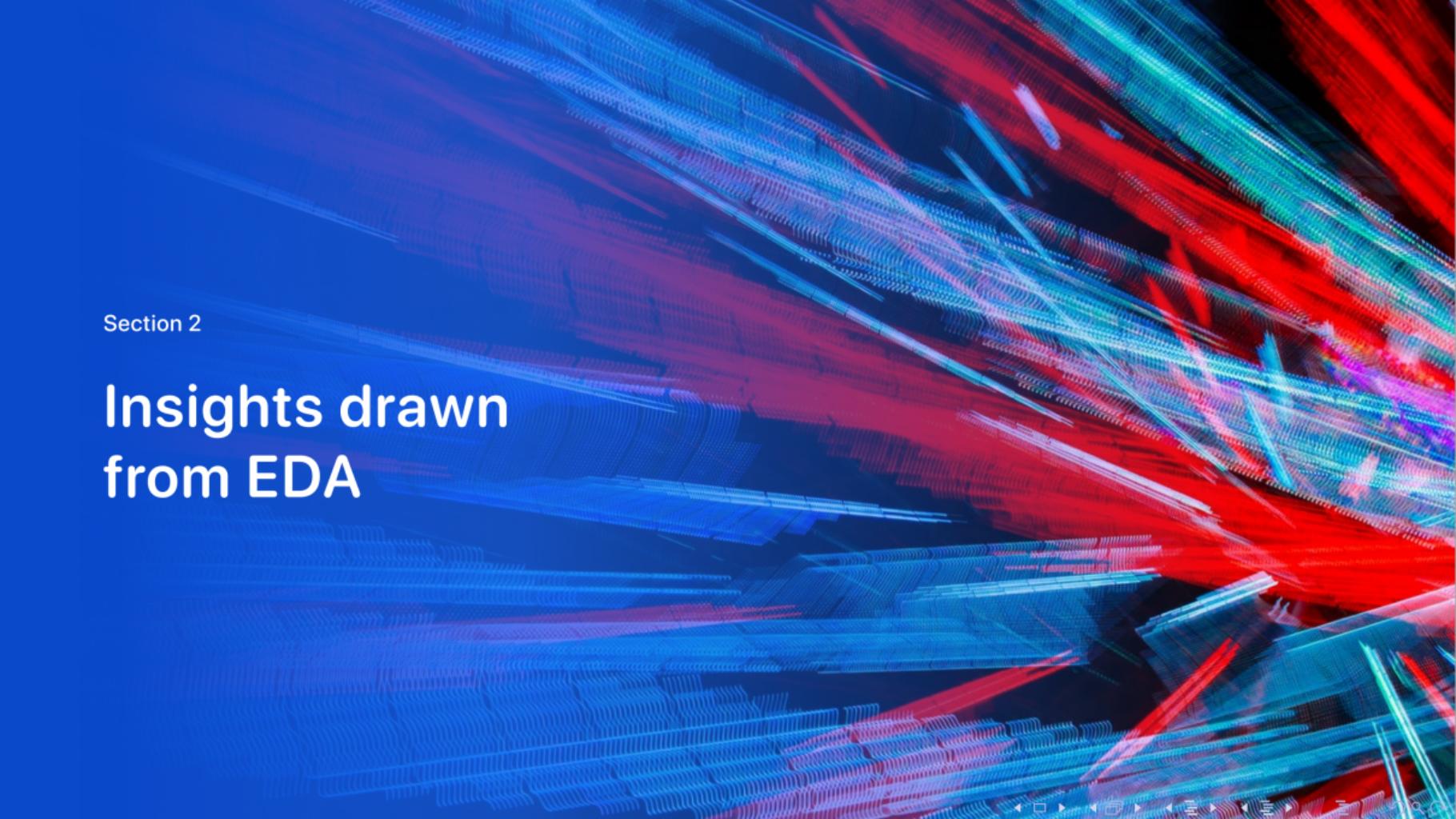
# EDA with Data Visualisation

# Build an Interactive Map with Folium

# Build a Dashboard with Plotly Dash

# Predictive Analysis (Classification)

- 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

# Payload versus Launch Site

# Success Rate versus Orbit Type

# Flight Number versus Orbit Type

# Payload versus Orbit Type

# Launch Success Yearly Trend

# All Launch Site Names

# Launch Site Names Begin with 'CCA'

# Total Payload Mass

# Average Payload Mass by F9 v1.1

# First Successful Ground Landing Date

# Successful Drone Ship Landing with Payload between 4000 and 60000

# Total Number of Successful and Failure Mission Outcomes

# Boosters Carried Maximum Payload

# 2015 Launch Records



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

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.

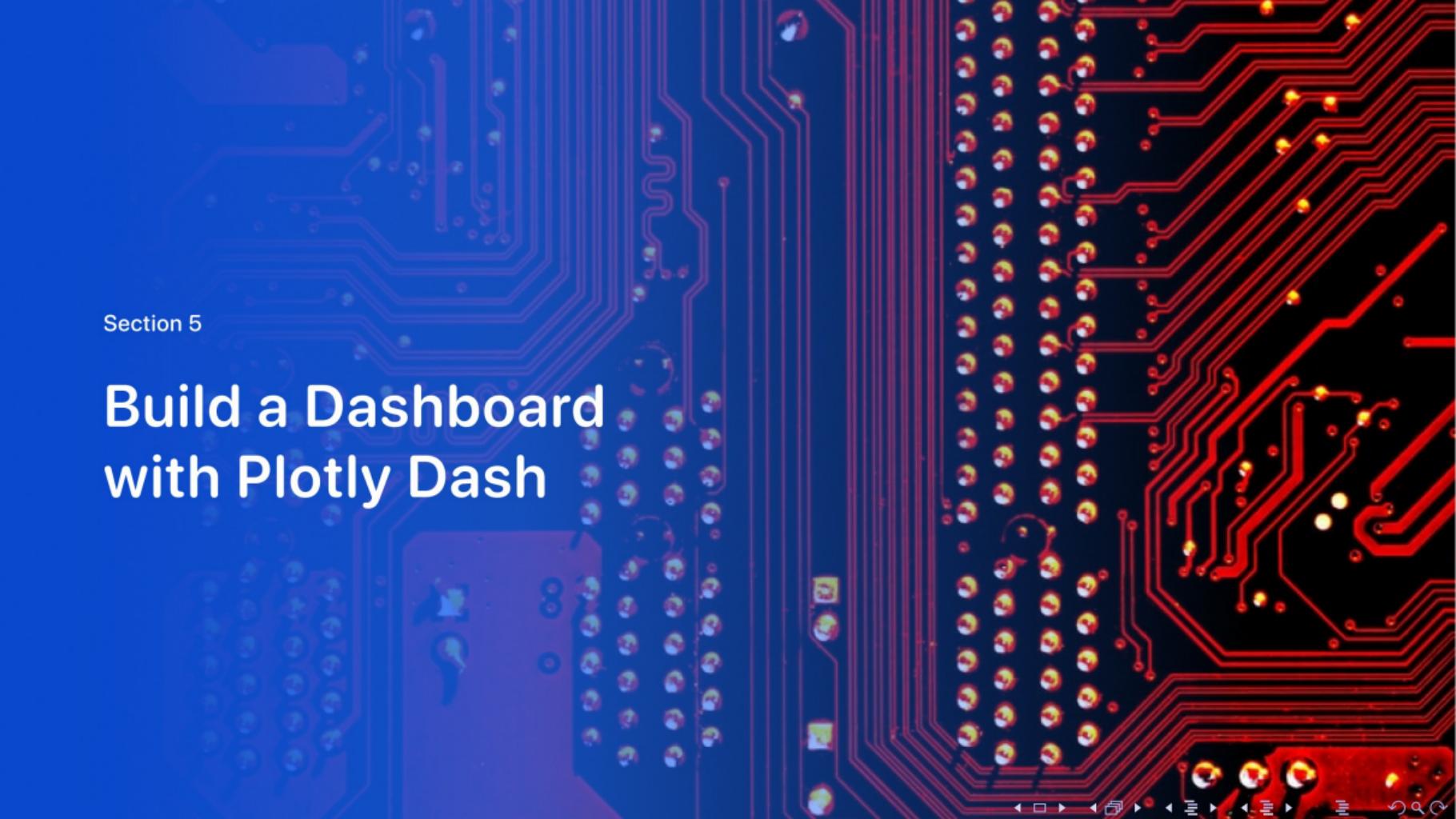
Section 4

# Launch Sites Proximities Analysis

# <Folium Map Screenshot 1>

## <Folium Map Screenshot 2>

# Folium Map Screenshot 3

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

# Dashboard Screenshot 1

# Dashboard Screenshot 2

# Dashboard Screenshot 3

The background of the slide features a dynamic, abstract design composed of several thick, curved lines in shades of blue and yellow. These lines create a sense of motion and depth, resembling a tunnel or a high-speed railway track curving through space. The overall aesthetic is modern and professional.

Section 6

# Predictive Analysis (Classification)

# Classification Accuracy

# Confusion Matrix

# Conclusions

-

# Appendix



Thank you!

