IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Oprea Jaydon George>
<30/3/2024>

Outline

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

Executive Summary

- Various methods and data science techniques were used to find the success rate of SpaceX launches and interactive visualization techniques were also used to better visualize the data
- SpaceY retains a certain degree of competitiveness against SpaceX

Introduction

- Gauge competitiveness of SpaceY against SpaceX
- Probability of success of SpaceX launches

Section 1

Methodology

Methodology

Executive Summary

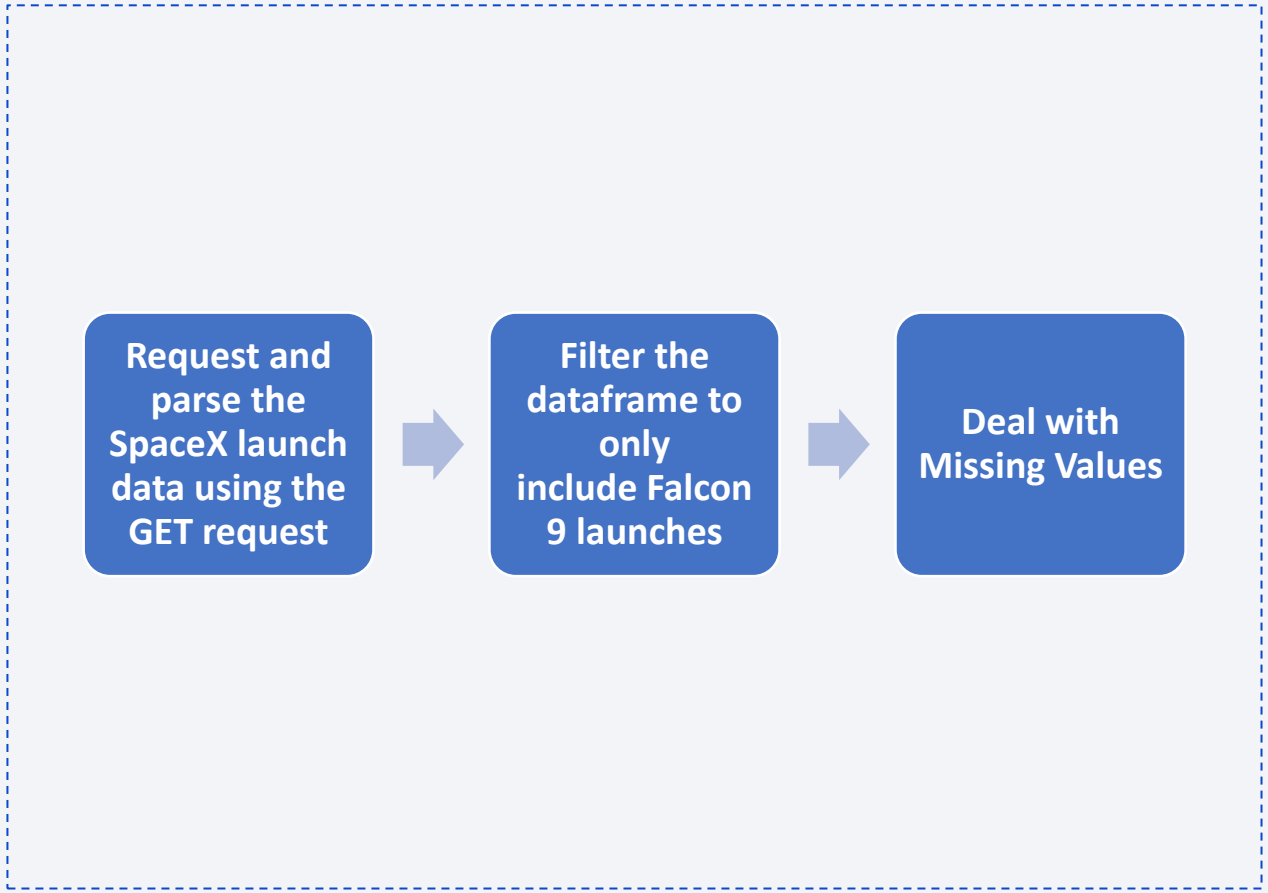
- Data collection methodology:
  - Data was collected by scraping web sources and using the SpaceX REST API
- Perform data wrangling
  - Exploratory Data Analysis is performed to determine what would be the label for training supervised models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building, tuning, evaluating classification models

Data Collection

- Data for SpaceX launches were collected by using the SpaceX REST API and scraping Wikipedia tables that contained launch data.

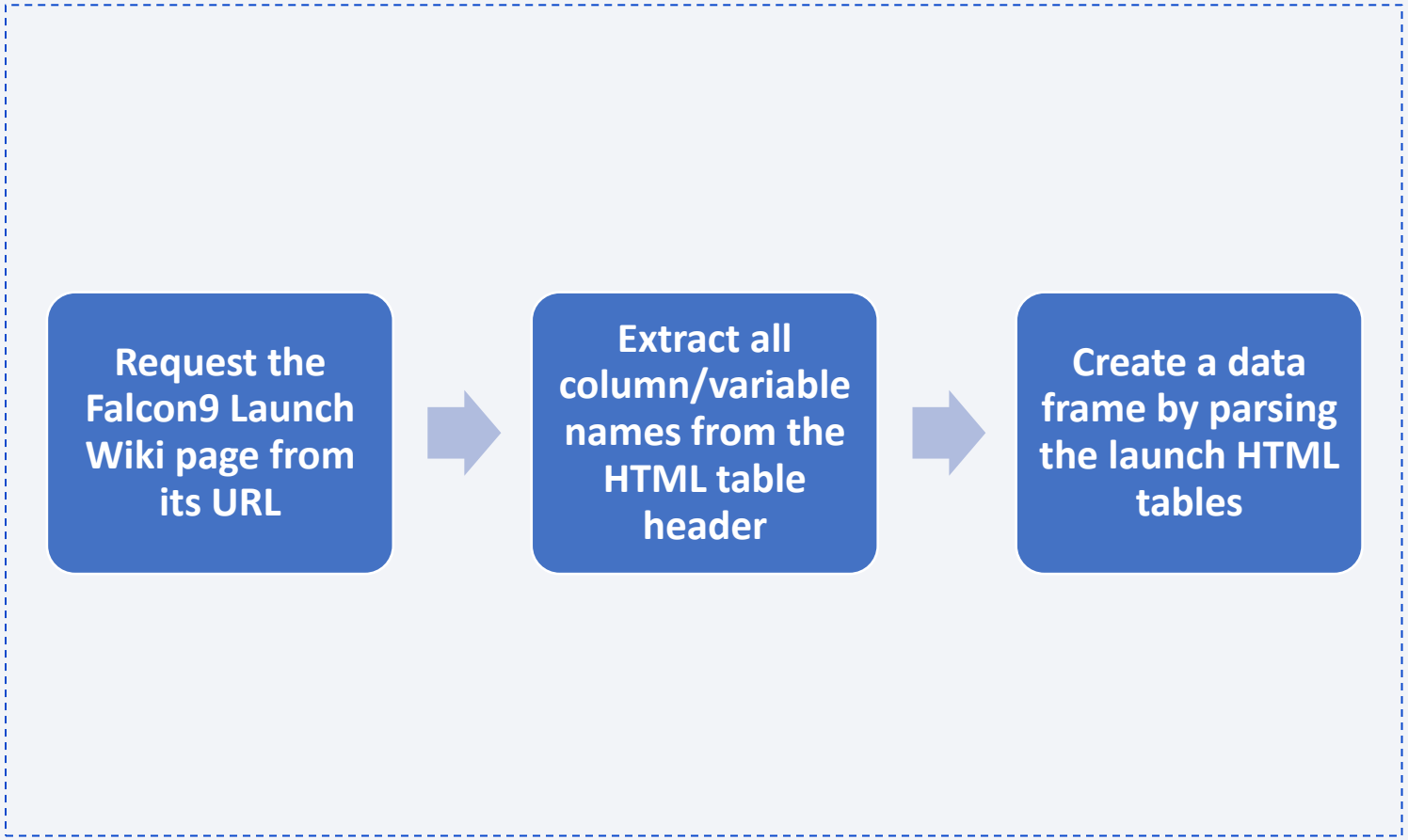
Data Collection – SpaceX API

- GitHub URL: https://github.com/EonsY/IBM-spacex-data-col



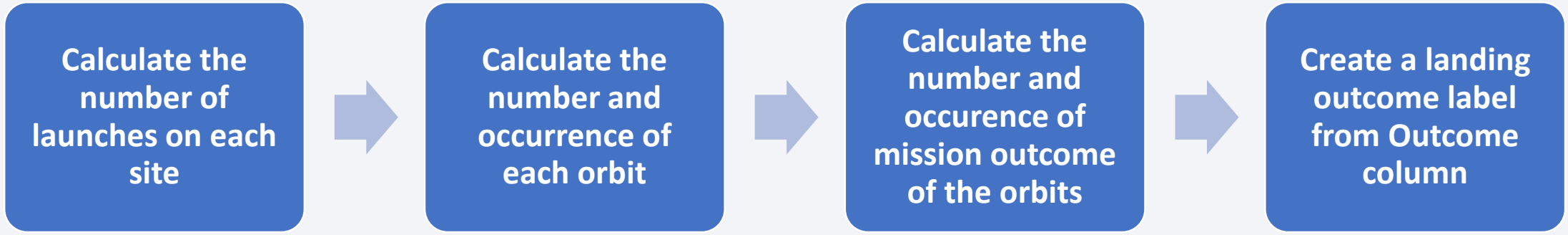
Data Collection - Scraping

- GitHub URL: https://github.com/EonsY/IBM-spacex-data-scraping



Data Wrangling

- Data that were collected is pre-processed for easier calculations for later
- GitHub URL: https://github.com/EonsY/IBM-spacex-data-wrangling



EDA with Data Visualization

- GitHub URL: https://github.com/EonsY/IBM-spacex-EDA-VL
- Plots that were used:
  - Categorical Distribution Plots, used to find insight from categorical and numerical values
  - Bar Plots, used to find success rate of different orbit launches
  - Line Plots, to find a trend of success rates throughout the years

EDA with SQL

- GitHub URL: https://github.com/EonsY/IBM-spacex-EDA-SQL

Find names of target launch sites

- **Display the names of the unique launch sites in the space mission**
- **Display 5 records where launch sites begin with the string 'CCA'**

Find out data related to launches

- **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 achieved**
- **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.**

Find out data related to launch outcomes

- **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 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

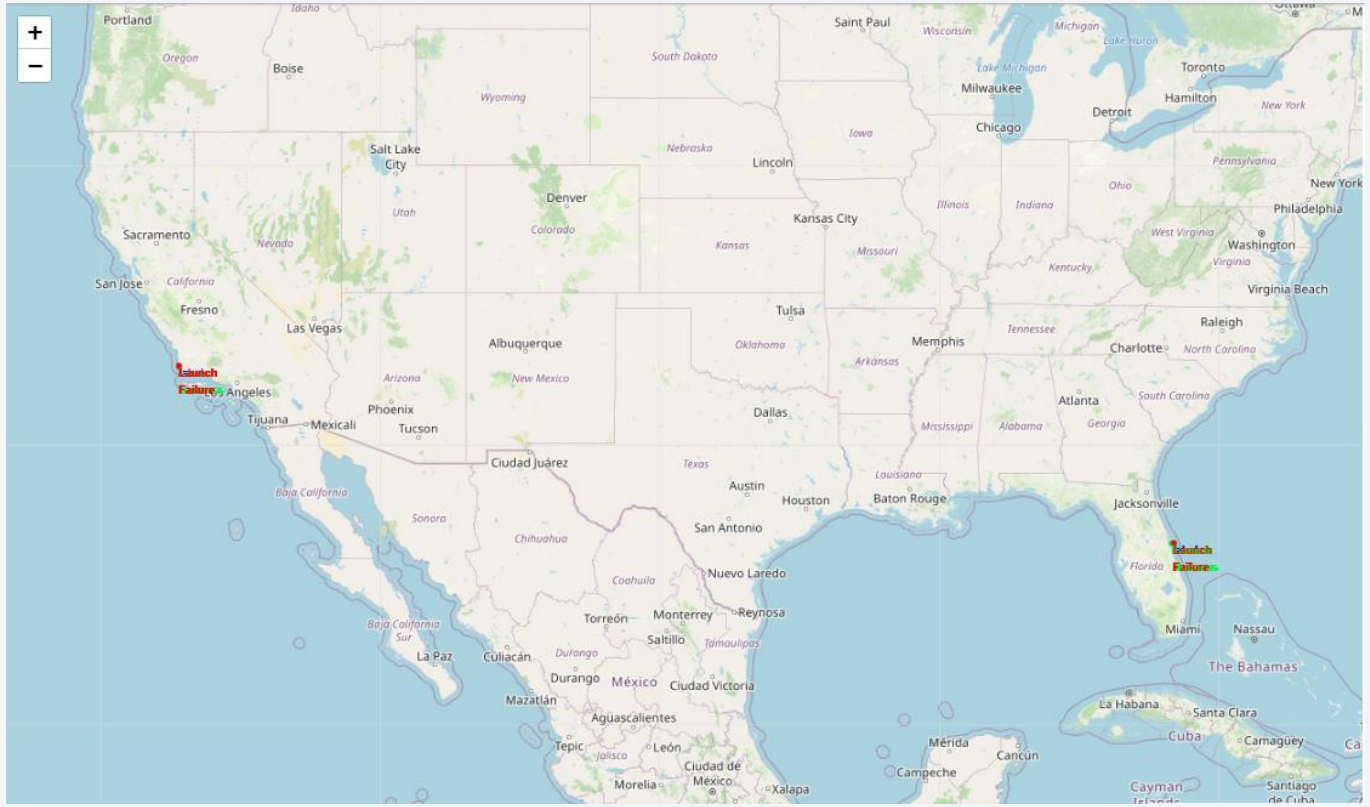
- Added markers on launch sites and lines leading to nearest coastlines
- Correlation between launch sites and their proximity to coastlines may be useful data
- GitHub URL: https://github.com/EonsY/IBM-spacex-IVA-Folium

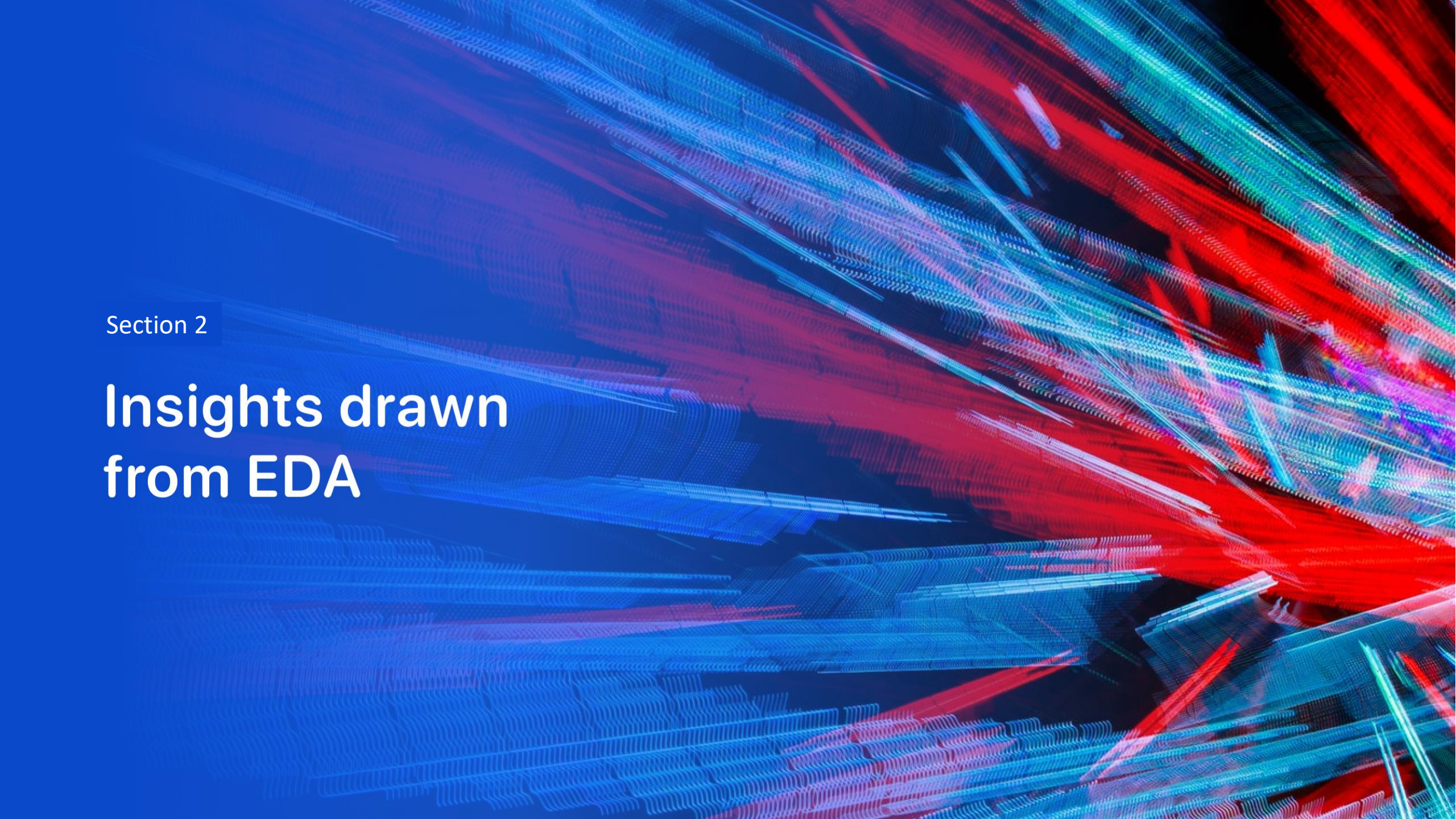
Predictive Analysis (Classification)

- Different predictive algorithms were used and compared to find the best performing ones
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

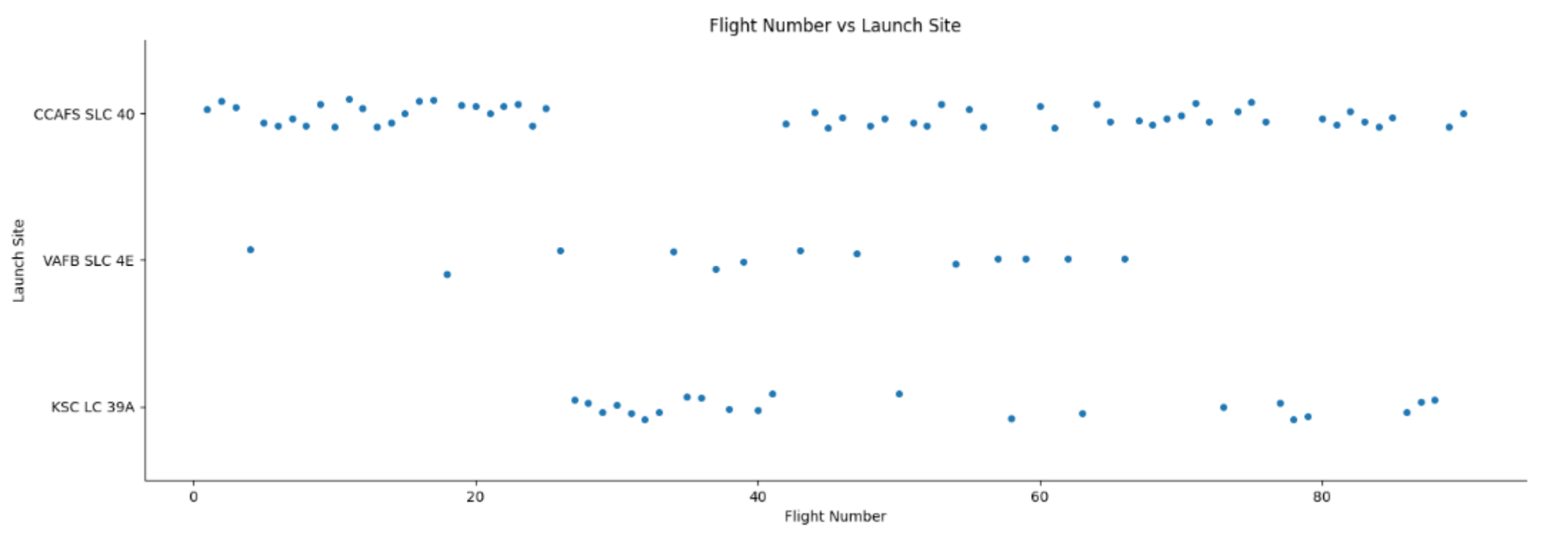
Results

- Through exploratory data analysis we gained insights about the landing outcomes of SpaceX
- Predictive analysis showed us that the success rate of landing outcomes were 83.44%



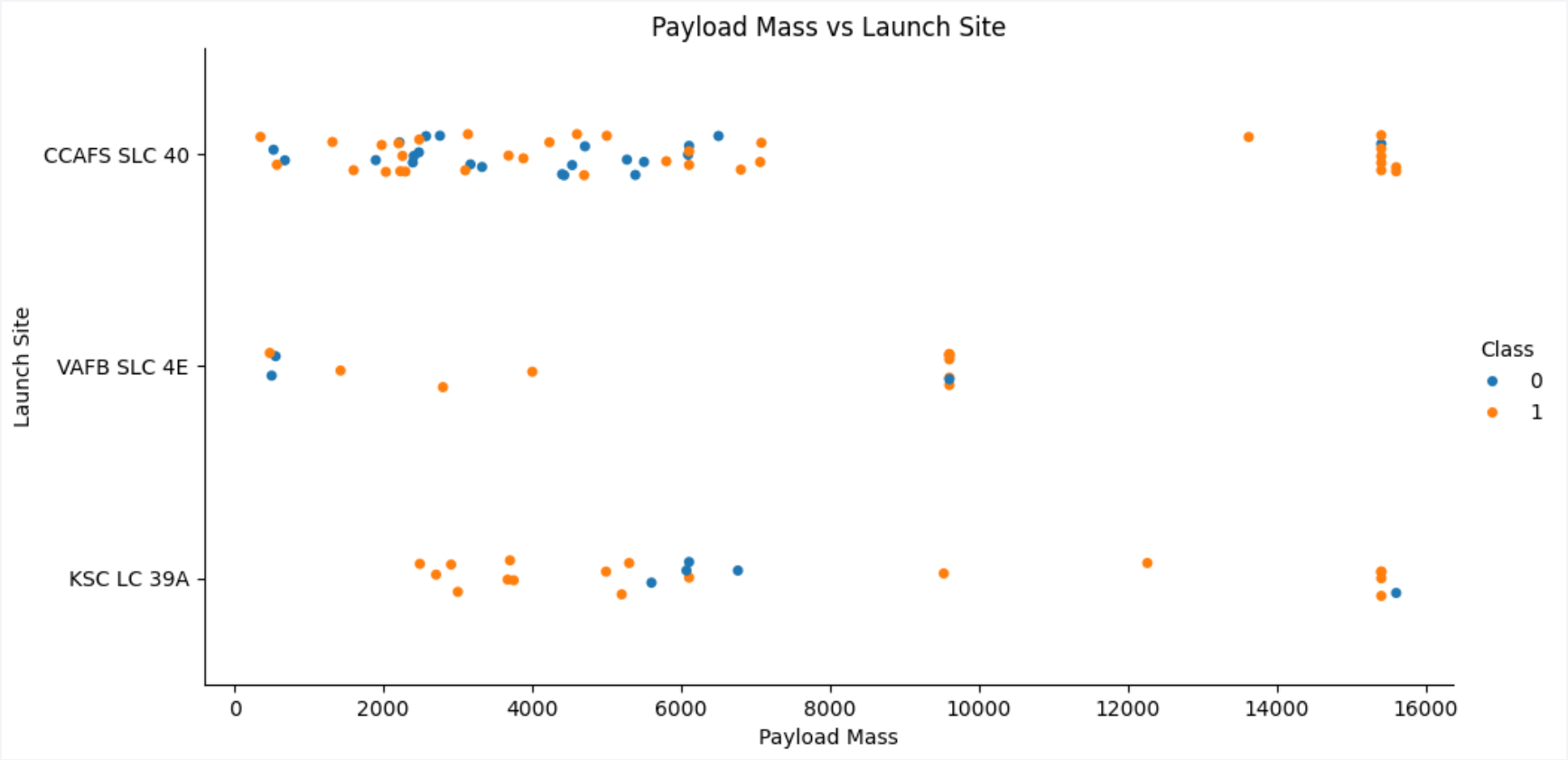

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



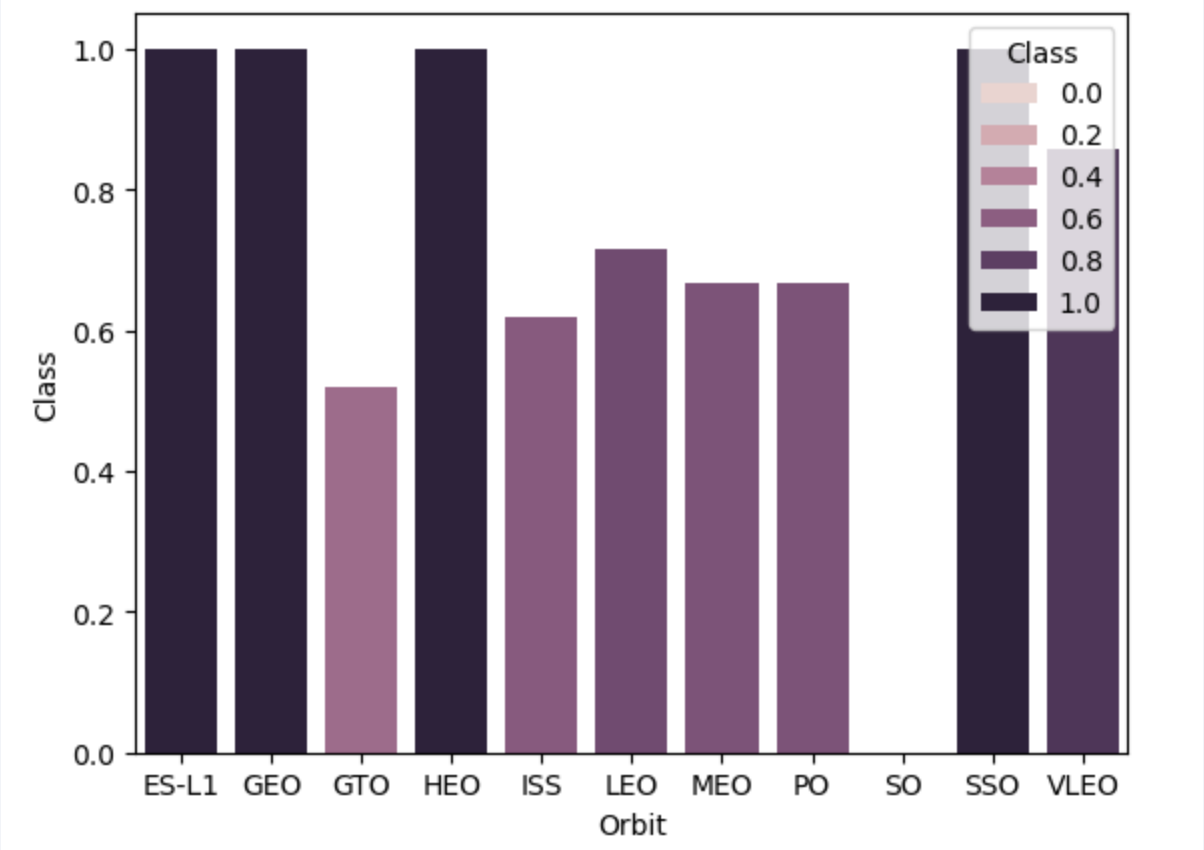
- Different groups of flight numbers might be tested in different launch sites

Payload vs. Launch Site



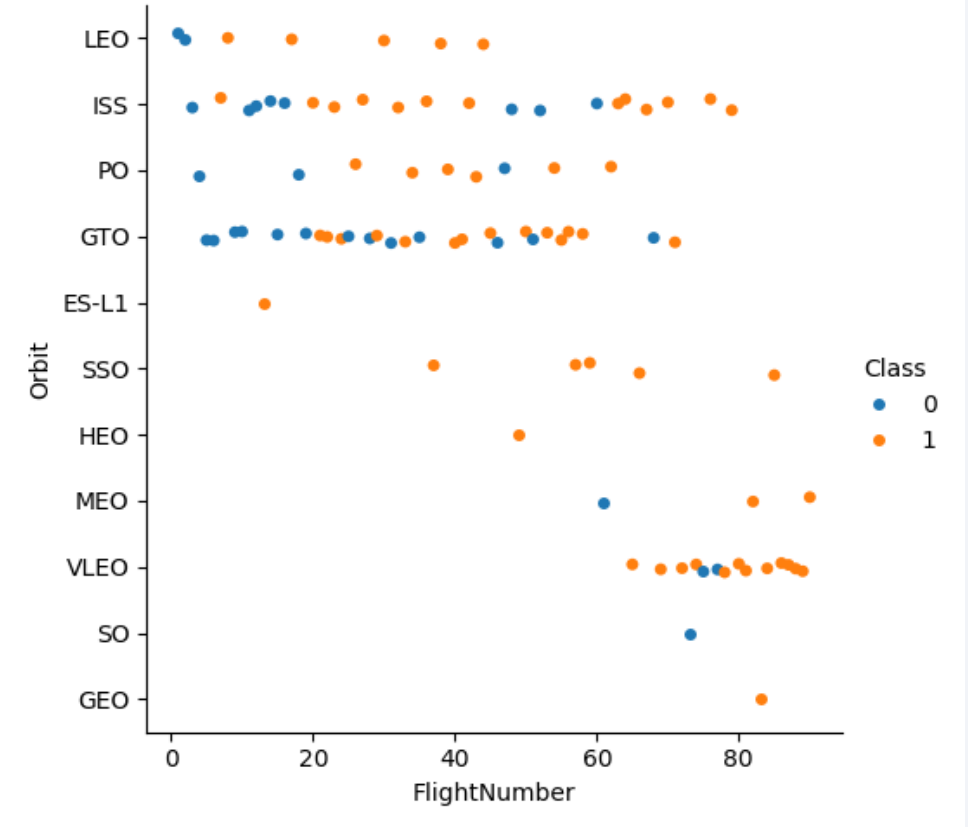
- VAFB SLC 4E was not used for launches of payloads higher than 10000kg

Success Rate vs. Orbit Type



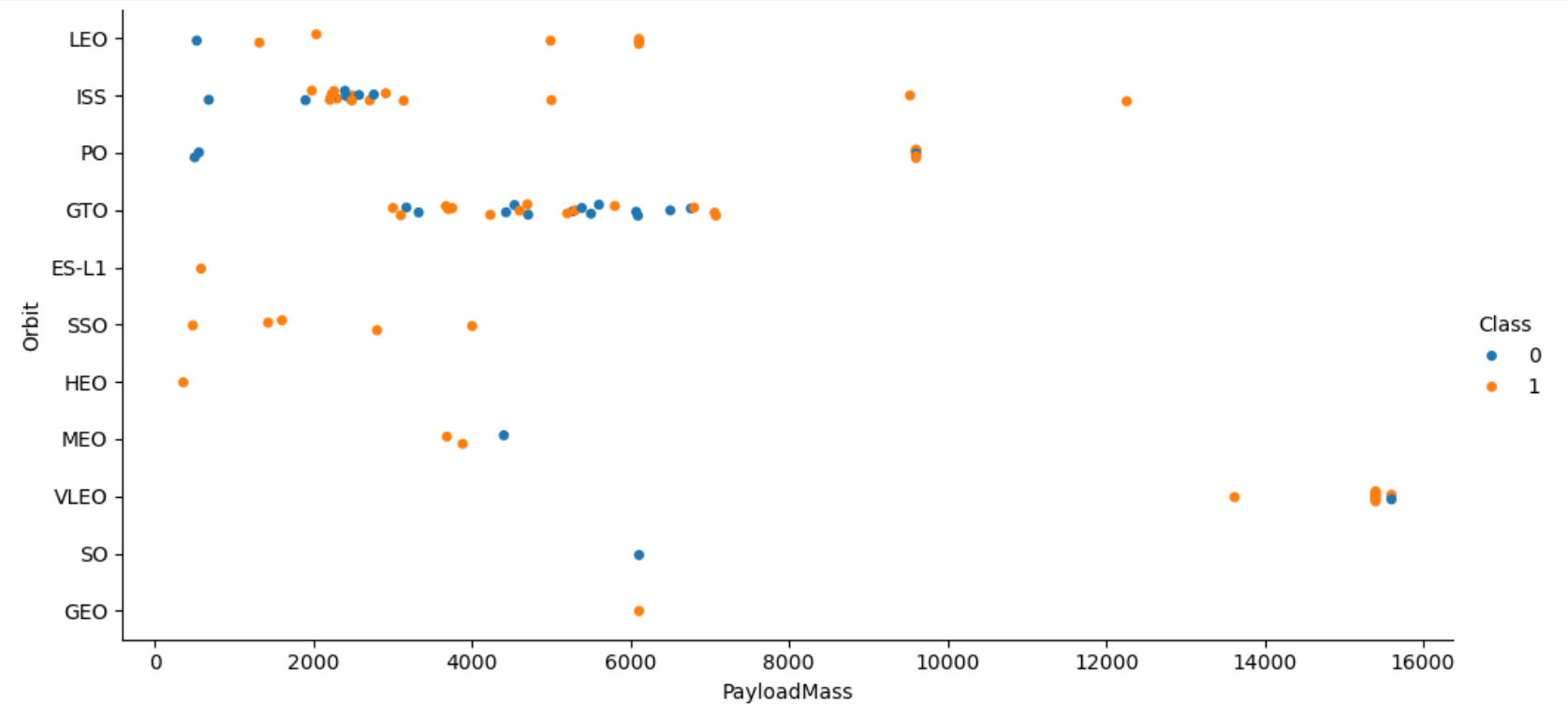
- Orbit types ES-L1, GEO, HEO, SSO has the highest success rates

Flight Number vs. Orbit Type



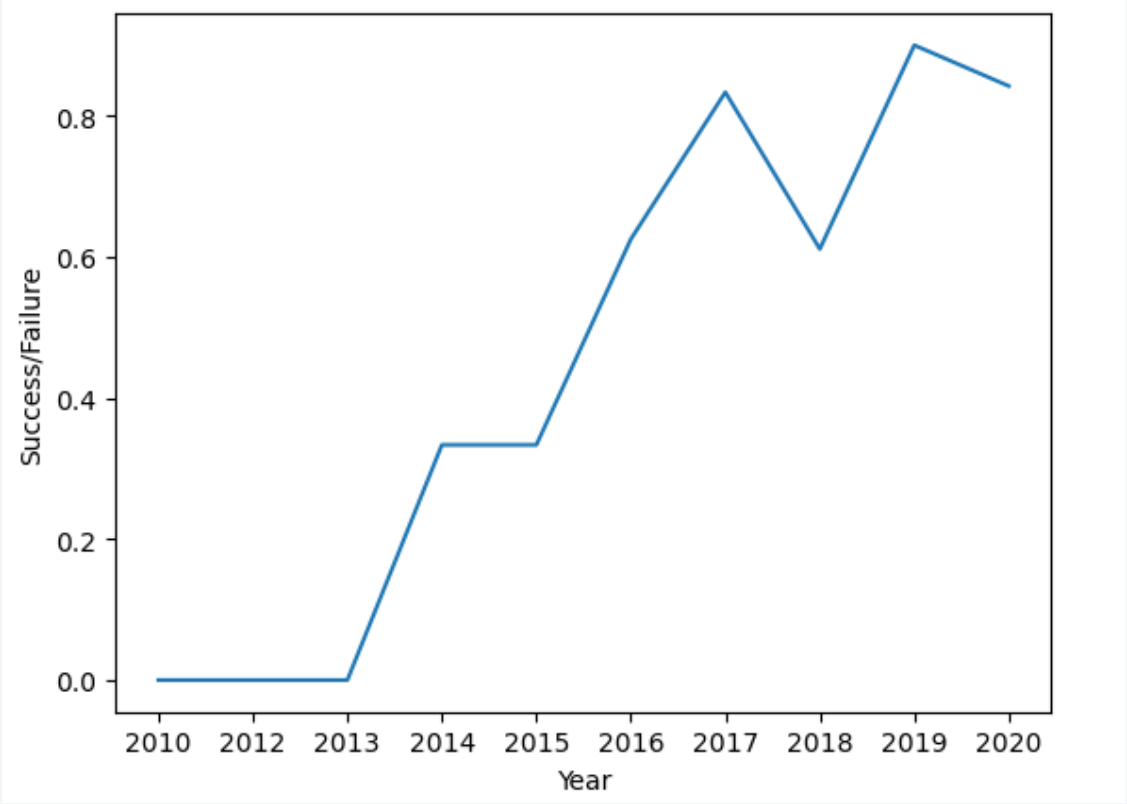
- VLEO orbits only have flight numbers above 60

Payload vs. Orbit Type



- Payload masses over 14000 kg are placed in VLEO orbit only

Launch Success Yearly Trend



- There is an increase in Success rates throughout the years

All Launch Site Names

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

- Launch sites used by SpaceX

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

- Five launch sites with names that begin with ‘CCA’ used by SpaceX
- Orbits were all to LEO

Total Payload Mass

Customer	sum(PAYLOAD_MASS__KG_)
NASA (CRS)	45596

- Total payload mass by NASA

Average Payload Mass by F9 v1.1

Booster_Version	avg(PAYLOAD_MASS__KG_)
F9 v1.1	2928.4
F9 v1.1 B1003	500.0
F9 v1.1 B1010	2216.0
F9 v1.1 B1011	4428.0
F9 v1.1 B1012	2395.0
F9 v1.1 B1013	570.0
F9 v1.1 B1014	4159.0
F9 v1.1 B1015	1898.0
F9 v1.1 B1016	4707.0
F9 v1.1 B1017	553.0
F9 v1.1 B1018	1952.0

- Average payload mass by booster version F9 v1.1 and its variants

First Successful Ground Landing Date

2015-12-22 Success (ground pad)

- First successful ground landing date by SpaceX was in 2015/12/22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- Booster versions with successful Drone Ship Landing with Payload between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

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

- Numbers of successful and failure mission outcomes

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

- Boosters that carried the maximum payloads

2015 Launch Records

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

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

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

- Different landing outcomes between 2010-06-04 and 2017-03-20

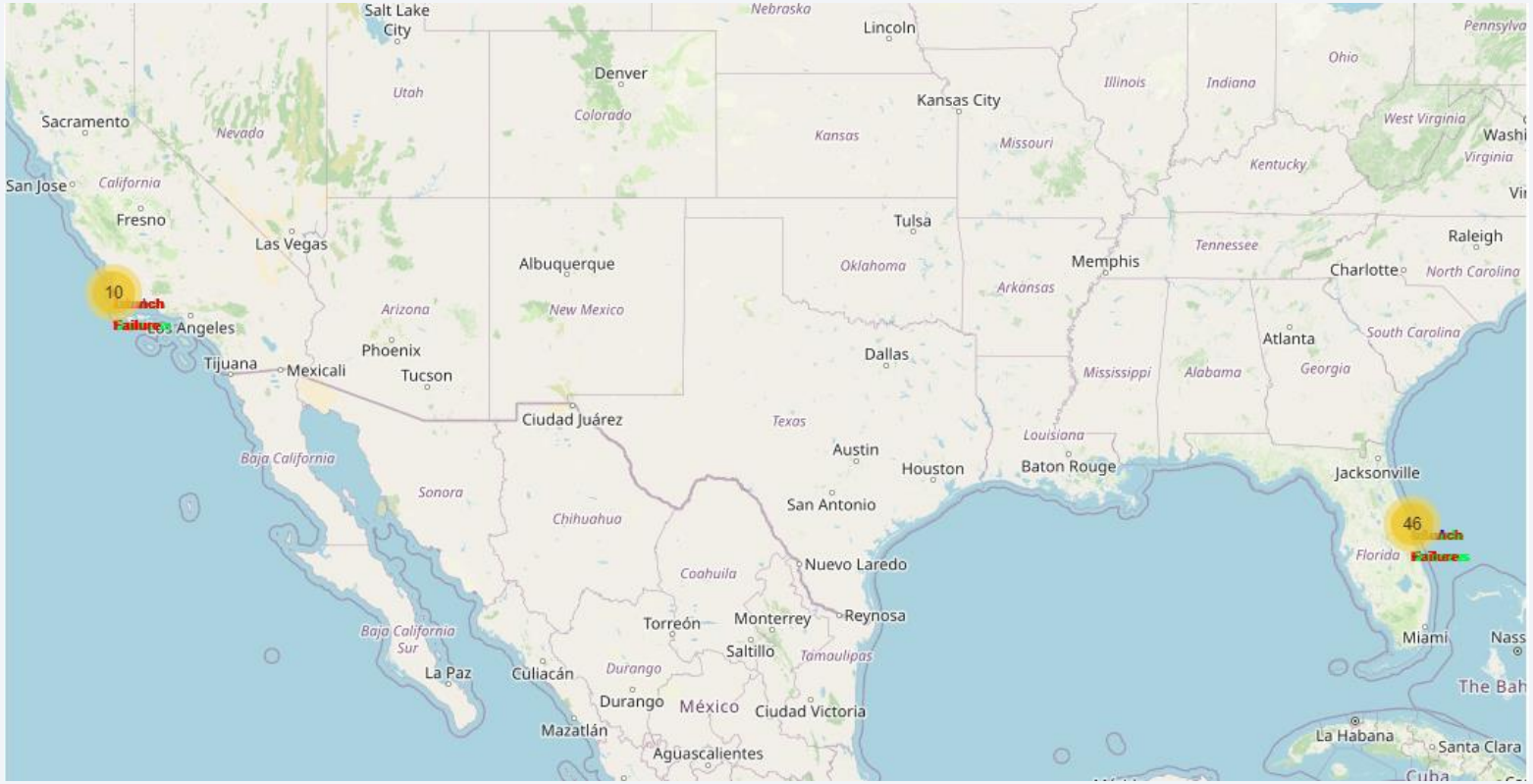
Section 3

Launch Sites Proximities Analysis

Launch Site Locations

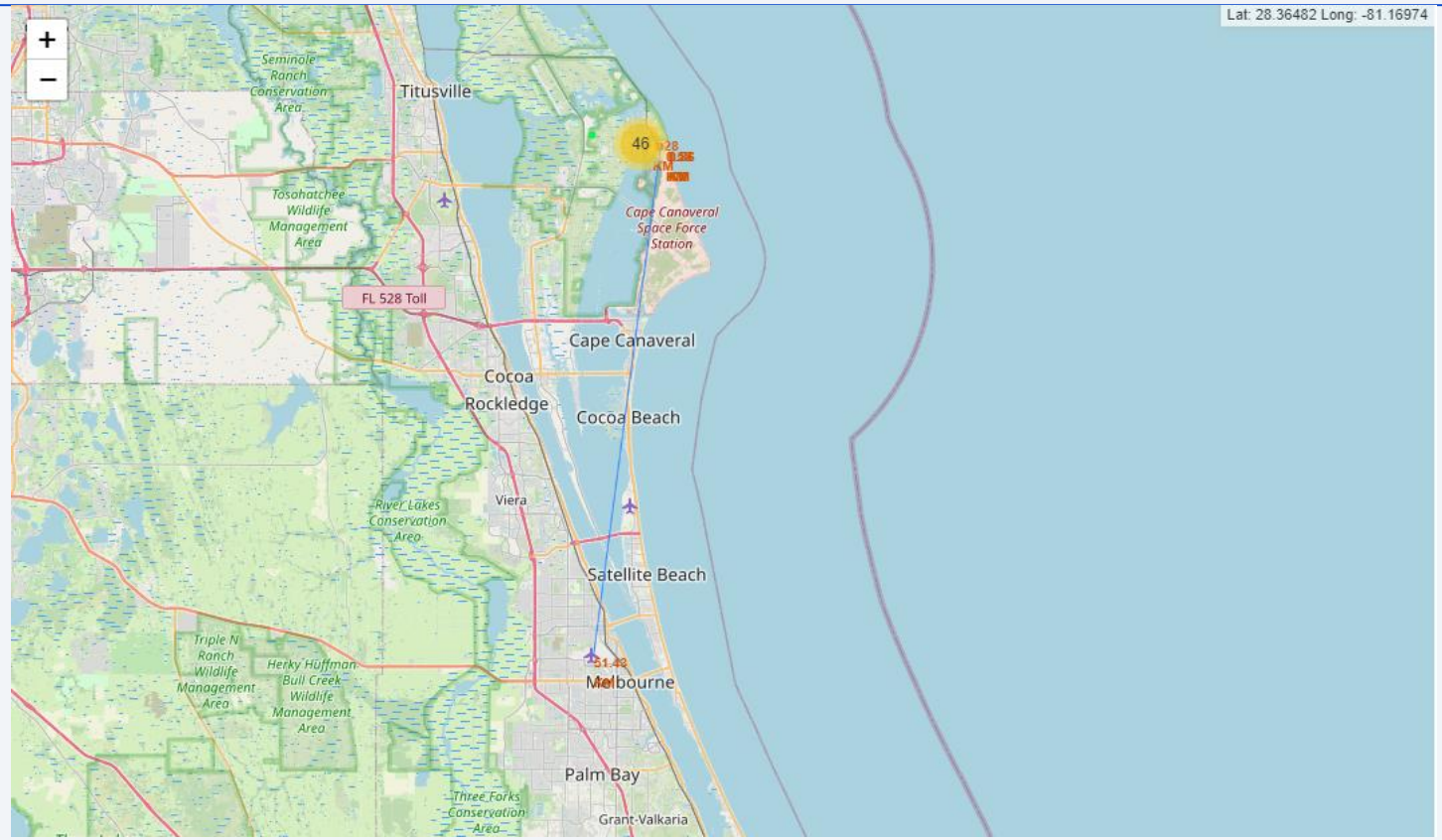
- Launch sites are located in California and Florida only

Number of launches in each launch site

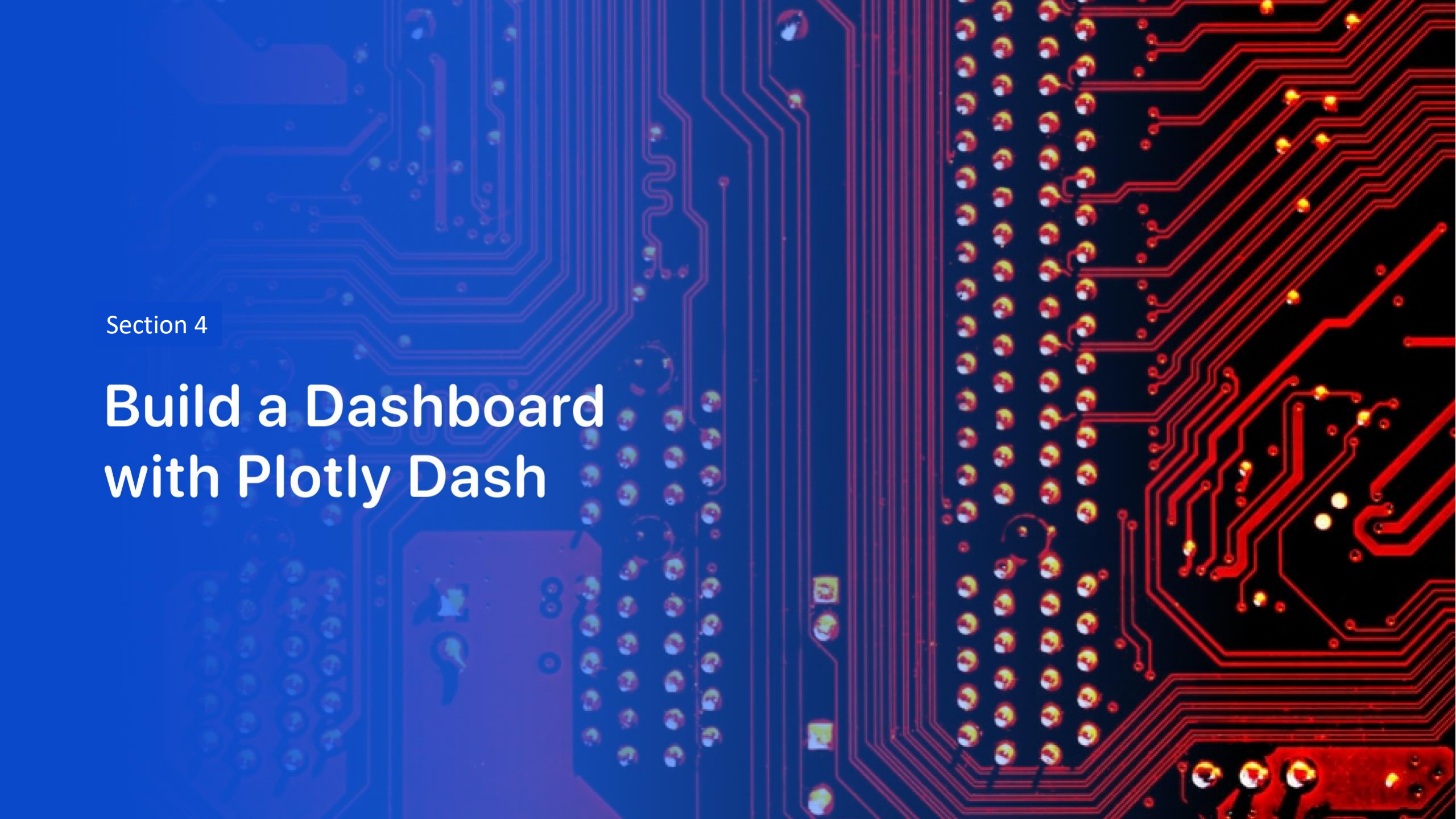


- 10 launches in California, 46 launches in Florida totaling up to 56 launches made by SpaceX

Proximity of Florida Launch Site to Airport



- The distance from the Florida launch site to its nearest airport is 51.43km


Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

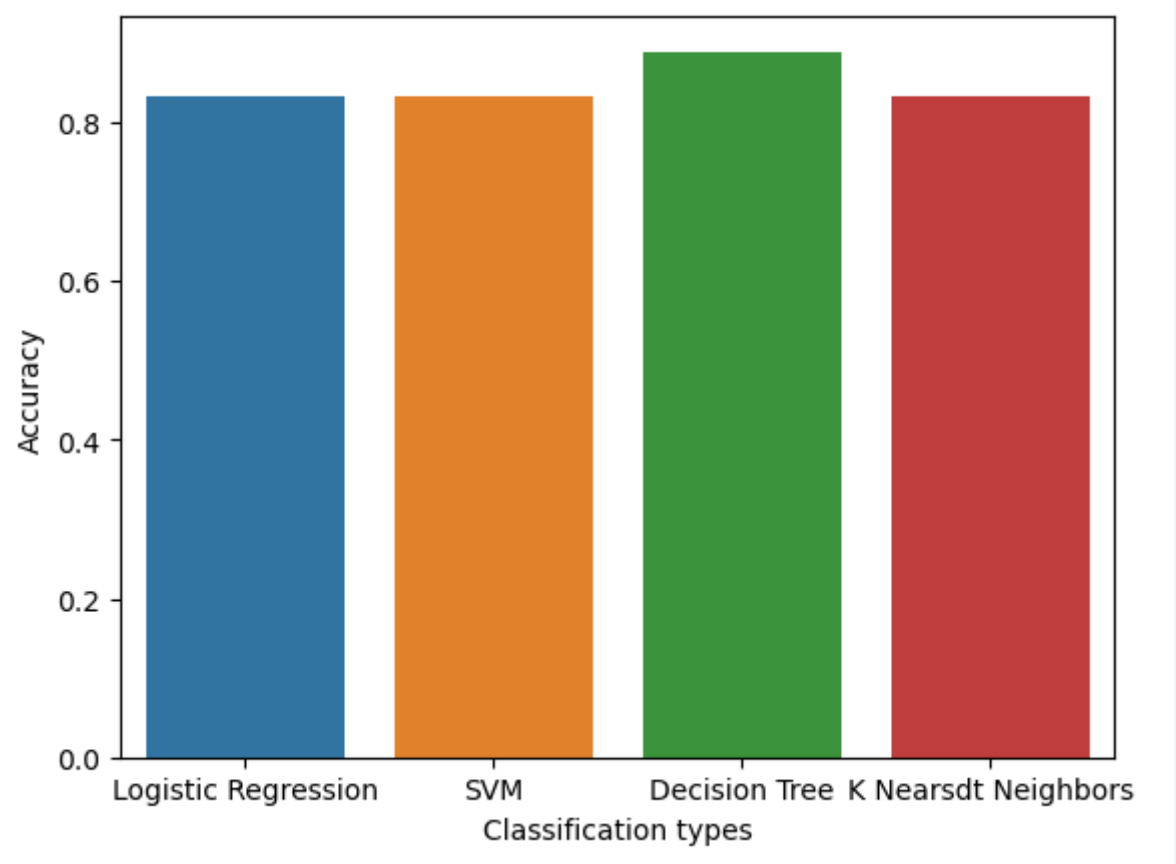
<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

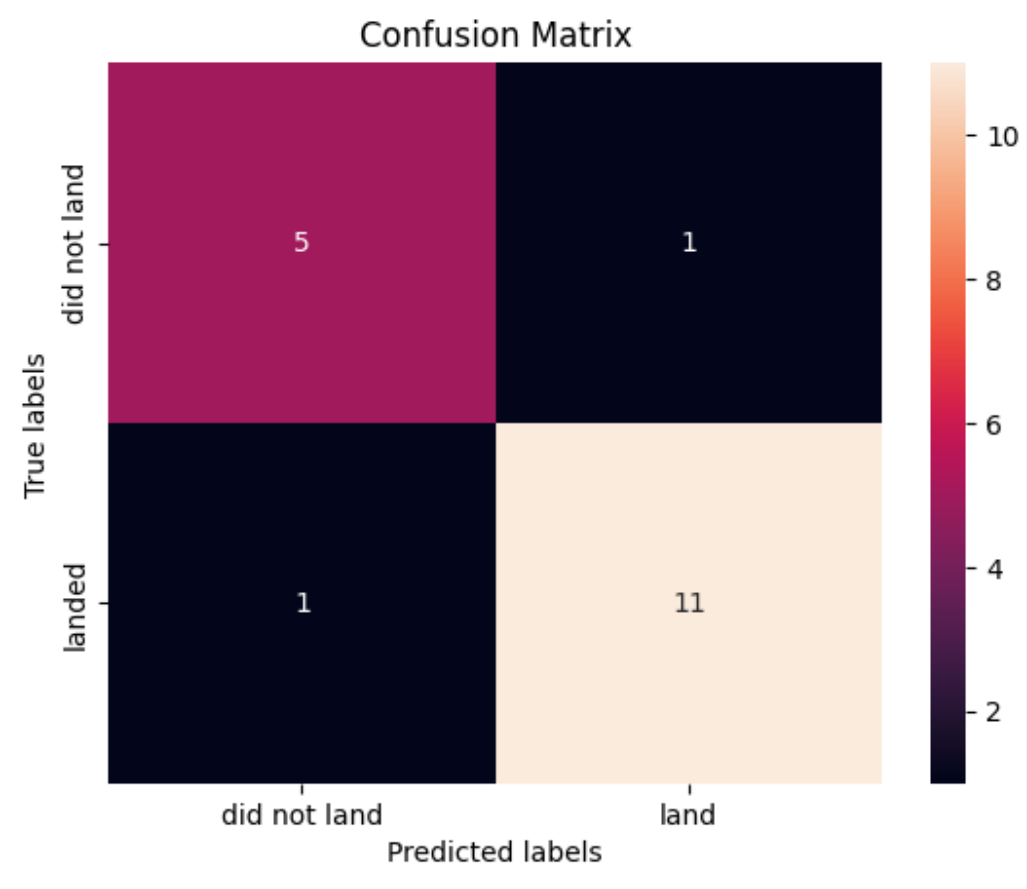
Predictive Analysis (Classification)

Classification Accuracy



- Decision trees have the highest classification accuracy

Confusion Matrix



- Confusion matrix of decision tree model shows that it is quite accurate

Conclusions

- The success rate of SpaceX launches has increased throughout the years
- SpaceX mainly focuses on ISS and GTO orbits
- If we wish to beat SpaceX we must deliver payloads of 14000kg and above to even further orbits than just VLEO
- We can also work on sending payloads to GTO orbits because SpaceX has a lower change of success in GTO missions

Thank you!