IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Felipe Morales
27/05/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Collection Data with API and Web Scraping
  - Analysis using SQL and Python Librarys
  - Visual Analytics
  - Prediction with Machine Learning
- Summary of all results

Introduction

- SpaceX designs, manufactures and launches advanced rockets and spacecraft. The company was founded in 2002 to revolutionize space technology, with the ultimate goal of enabling people to live on other planets (spacex.com, 2024).
- Predict if the Falcon 9 first stage will land successfully. Determine the cost of a launch.

Section 1

Methodology

Methodology

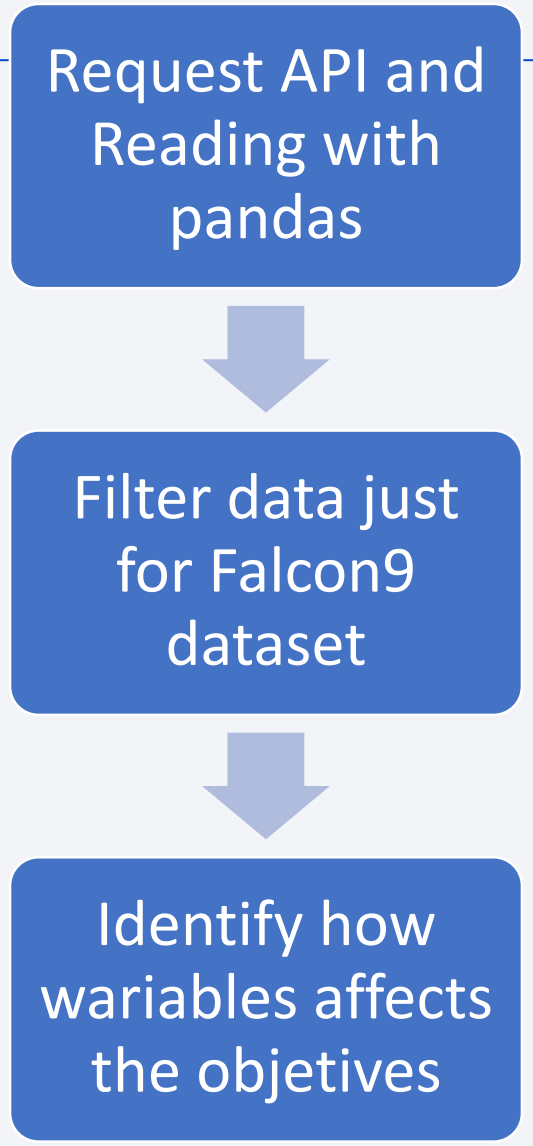
Executive Summary

- Data collection methodology:
  - SpaceX API
  - WebScraping (Wikipedia)
- Perform data wrangling
  - Transforming the raw data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection – SpaceX API

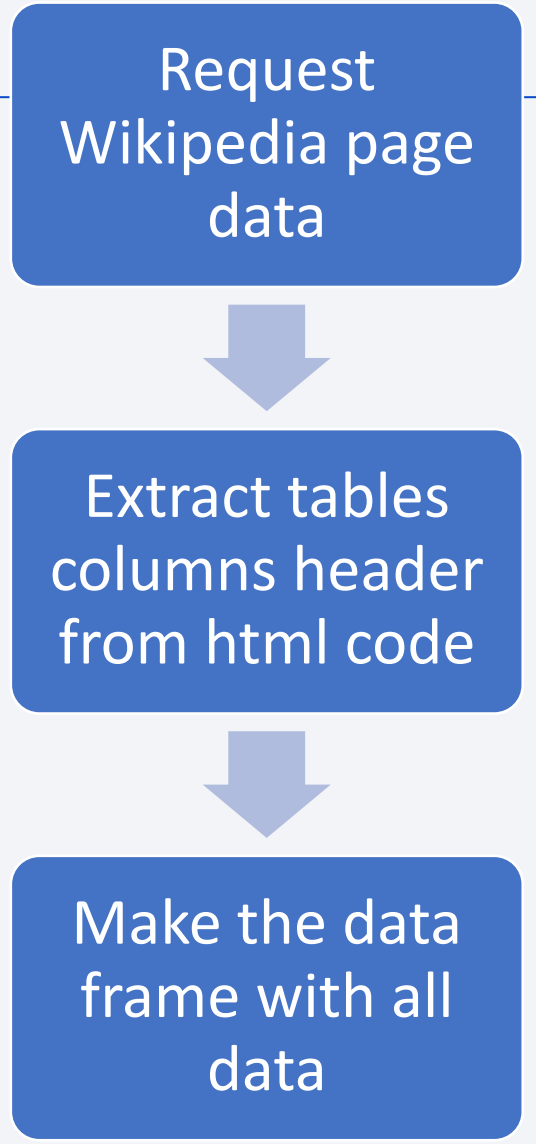
- SpaceX provides a publicly accessible interface, known as an API (Application Programming Interface), that allows users to retrieve data for their own use.
- Reading the SpaceX dataset into a Pandas dataframe.

- Code github**



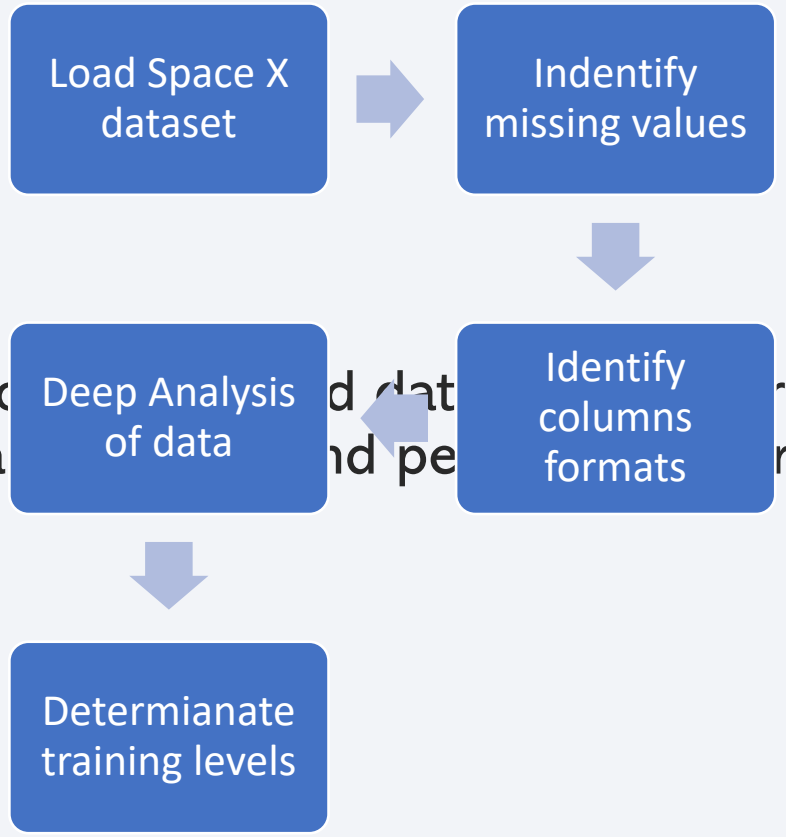
Data Collection - Scraping

- Data from Wikipedia site, extracted from html site code.
- The main variables extracted are the tables columns



Data Wrangling

- Firt, Exploratory Data Analysis (EDA) was used on the Dataset
- The principals stages of explotary data analysis are:



- Add the GitHub URL of yo d dat related notebooks, as an externa nd pe rpose

EDA with Data Visualization

- Charts plotted for visualization with EDA

1. Payload Mass and Flight Number: see how the FlightNumber (indicating the continuous launch attempts.) and Payload variables would affect the launch outcome.
2. Launch Site and Flight Number: detailed launch records
3. Launch Site and Payload Mass: relationship between Payload and Launch Site
4. Orbit and Flight Number: relationship between success rate and orbit type
5. Payload and Orbit type: relationship between FlightNumber and Orbit type

EDA with SQL

- SQL queries performed

1. Names of the unique launch sites in the space mission;
2. Top 5 launch sites whose name begin with the string 'CCA’
3. Total payload mass carried by boosters launched by NASA (CRS);
4. Average payload mass carried by booster version F9 v1.1
5. Date when the first successful landing outcome in ground pad was achieved
6. Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg
7. Total number of successful and failure mission outcomes
8. Names of the booster versions which have carried the maximum payload mass
9. Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
10. Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

Build an Interactive Map with Folium

- For building an interactive map with fusion
    1. Markers indicate points like launch sites
    2. Circles indicate highlighted areas around specific coordinates
    3. Marker clusters indicates launches and launch site
    4. Lines to indicate distances between two coordinates like lauch site and highway

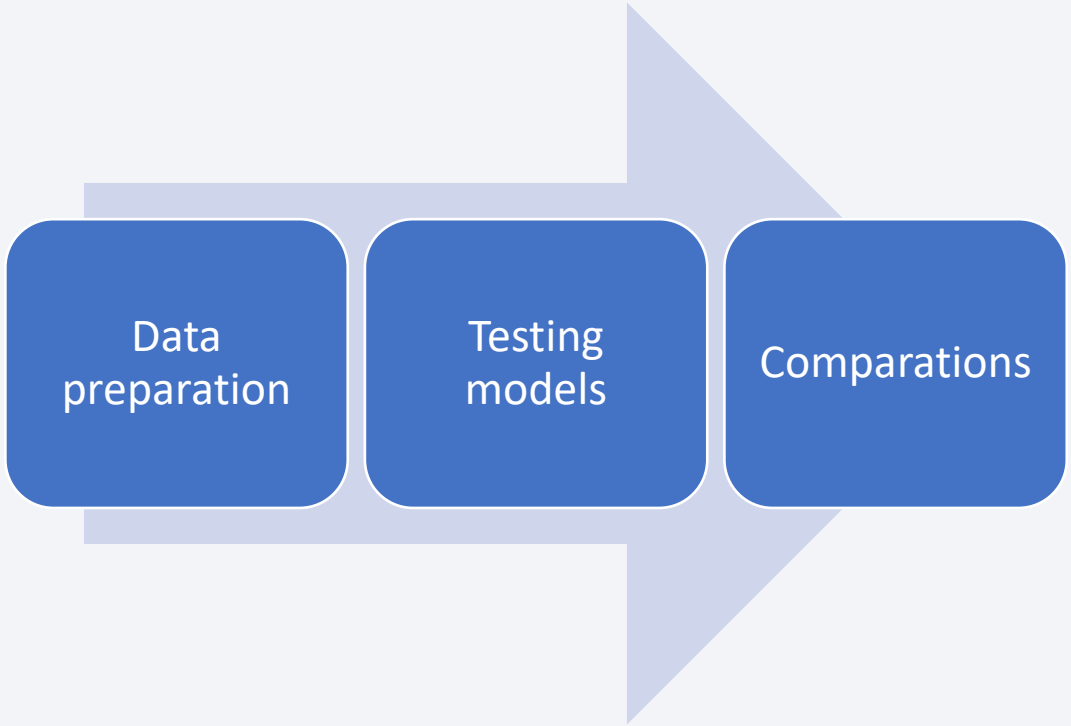
Build a Dashboard with Plotly Dash

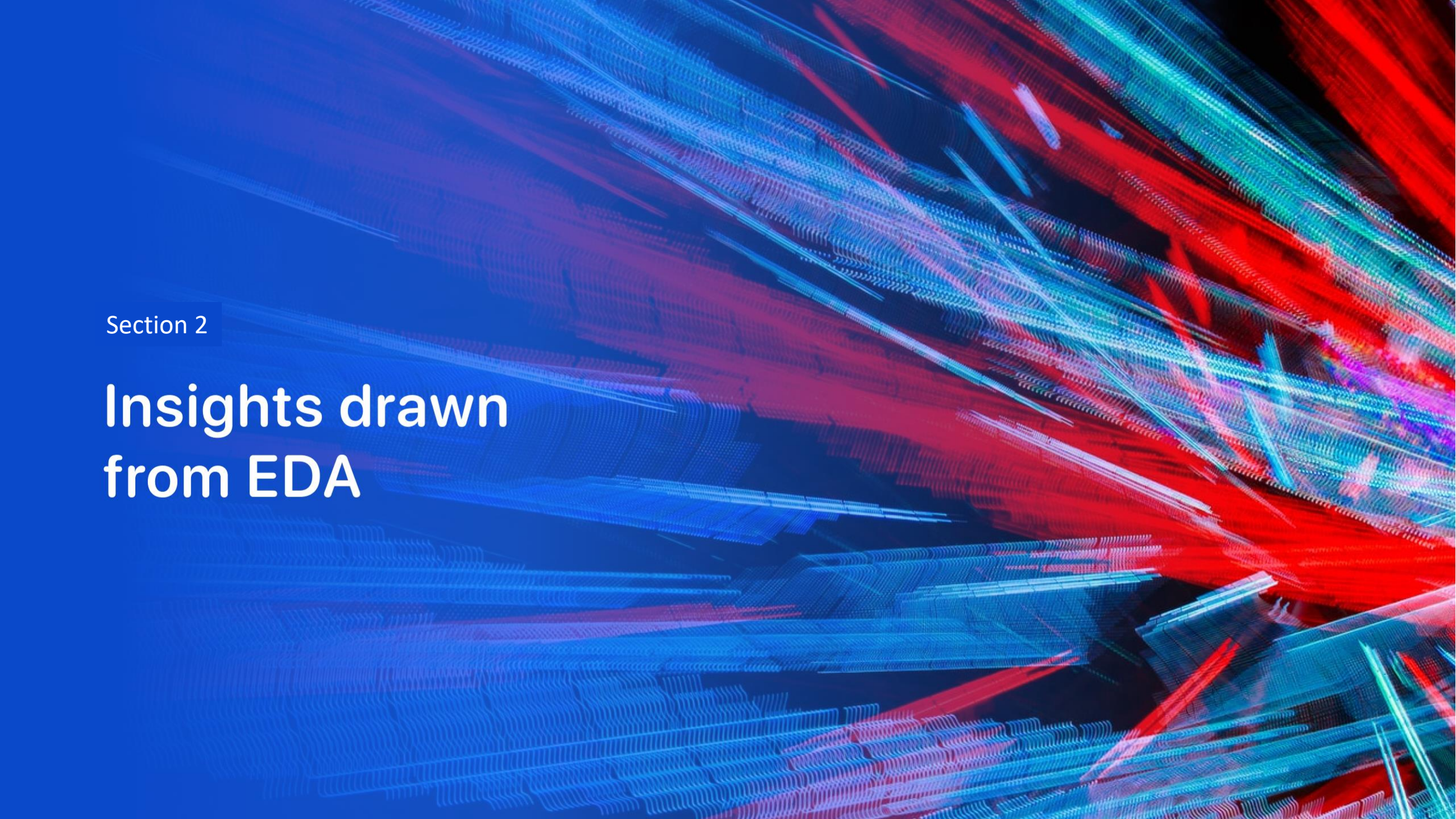
- The following plots/graphs and interactions are added to the dashboard
  - Percentage of launches by site
  - Payload range

Help to identify where is best place to launch according to payloads

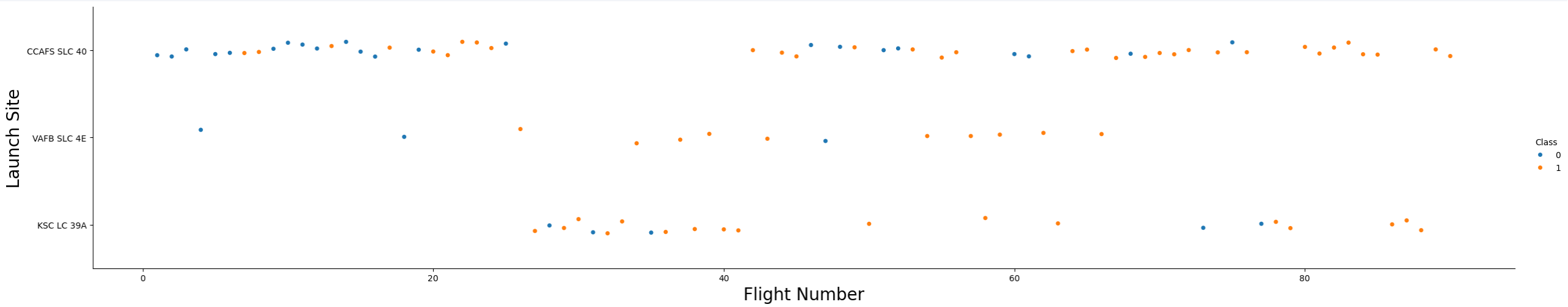
Predictive Analysis (Classification)

- Logistic regression, support vector machine, decision tree and k nearest neighbors are compared




Section 2
Insights drawn from EDA

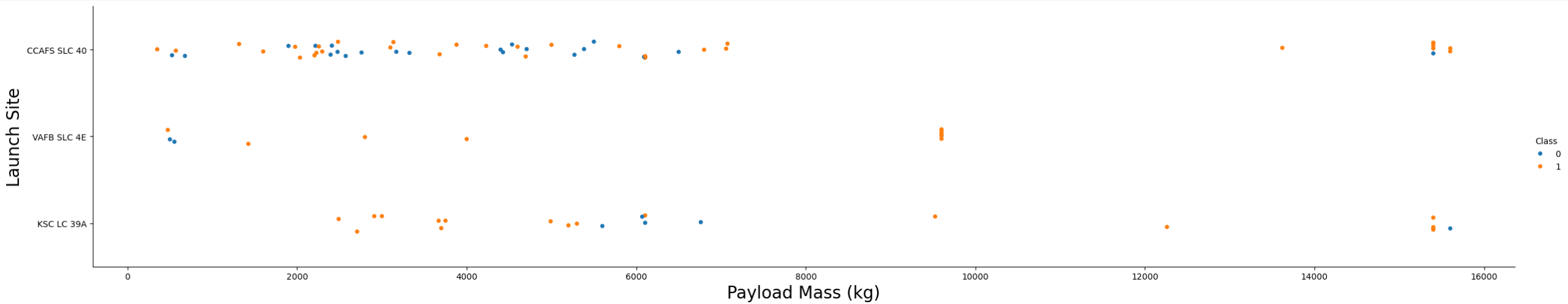
Flight Number vs. Launch Site



Conclusions:

1. The most number of launchers are in CCAFS SLC 40
2. Most successful launchers are in CCAFS SLC 40
3. Most failed lauchers are in KSC LC 39A

Payload vs. Launch Site



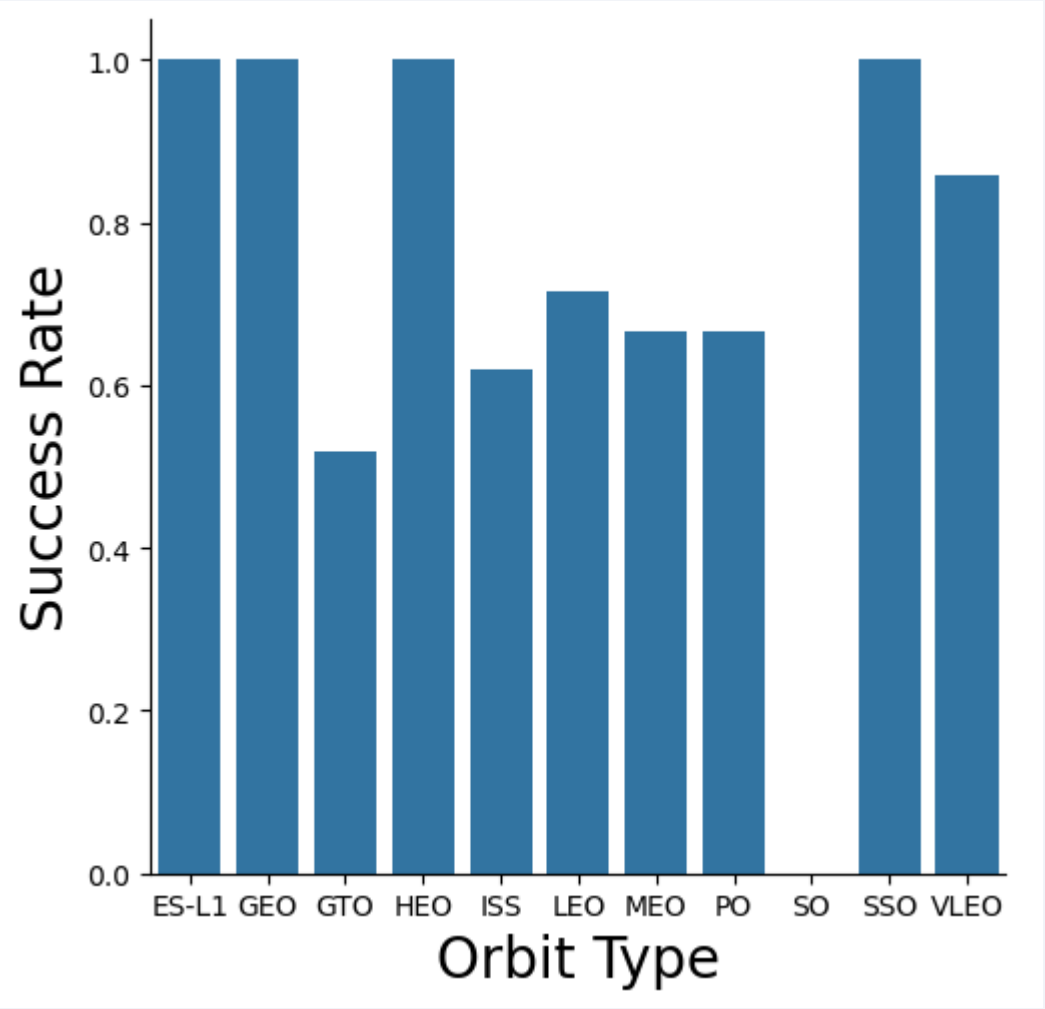
Conclusions:

1. Most >8000 payloads are successful
2. VAFB SLC 4E lauch site have good results
3. Payloads >14000 launches only in CCAFS SLC40 and KSC KC 39A

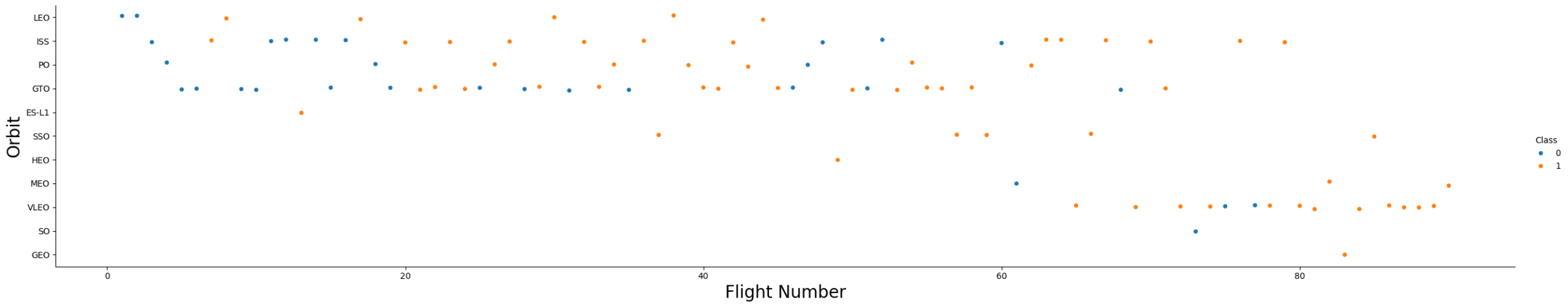
Success Rate vs. Orbit Type

Explanation of the chart

1. Orbits with 100% success rate are:
   - ES-L1
   - GEO
   - HEO
   - SSO
2. Orbits with 0% success rate are just: SO
3. Orbits with success rate between 50% and 85%:
   - GTO, ISSLEO,MEO,PO

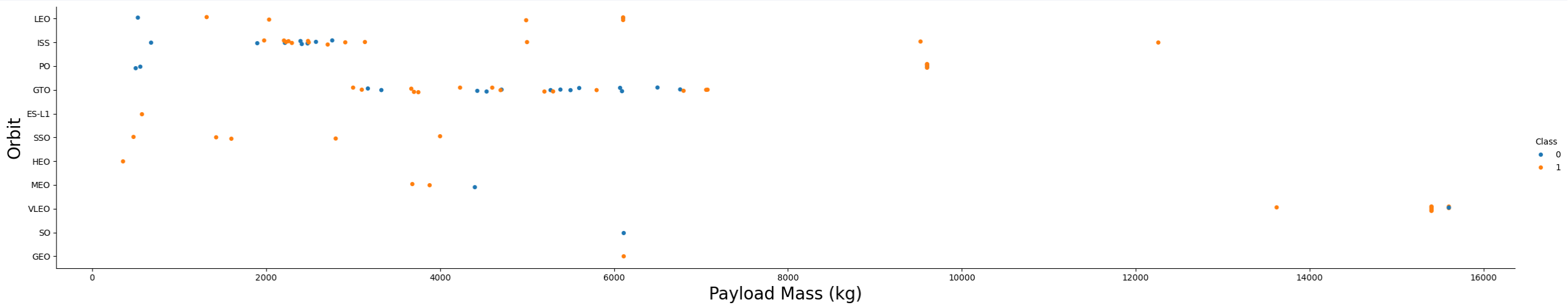


Flight Number vs. Orbit Type



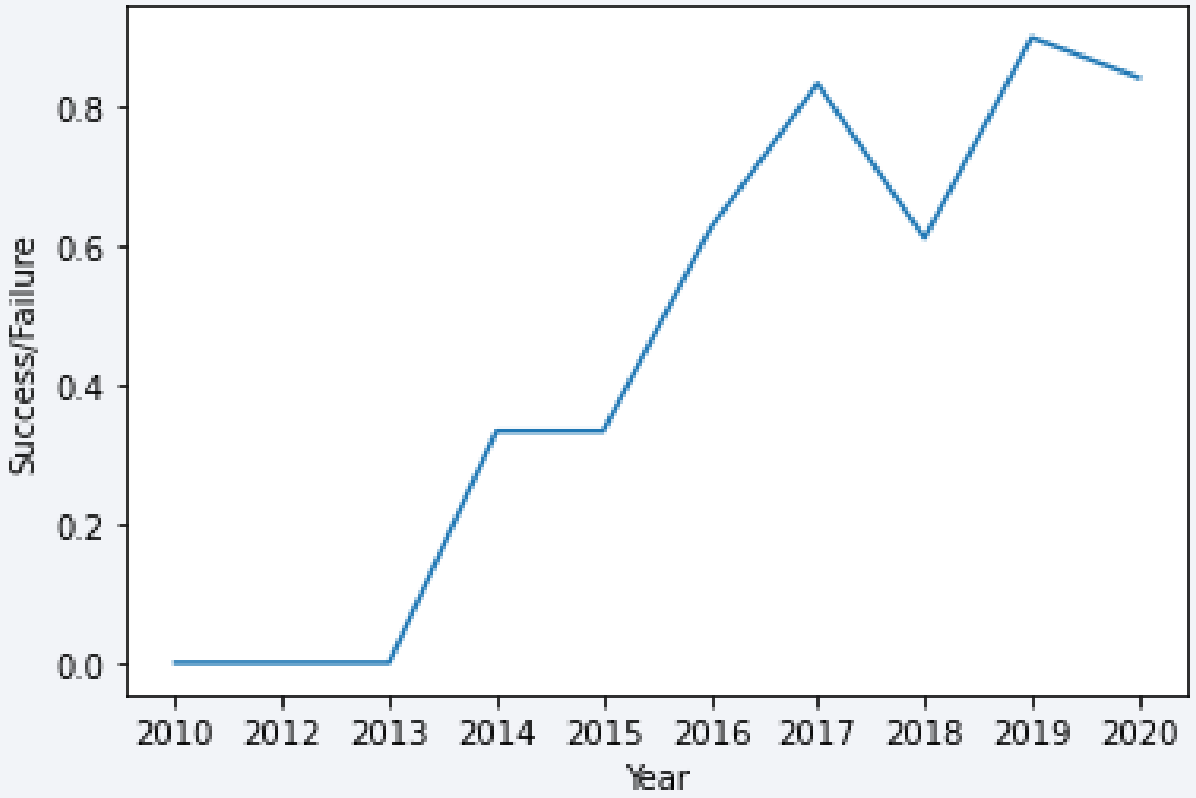
- Best successful orbits are ES-L1,GEO,HEO,SSO
- Successful have a direct relation with flight number

Payload vs. Orbit Type



1. There is no relation between payload and success rate

Launch Success Yearly Trend



1. Success rate started increasing since 2013

All Launch Site Names

- There are four launch sites
- CCAFS LC-40
- VAFB SLC-4E
- KSC LC-39A
- CCAFS SLC-40

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

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

- Randoms records where launch sites begin with `CCA`

Total Payload Mass

- The total sum payload (all customers)

619967 kg

Done.

SUM_PAYLOAD
619967

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1

  Is 2928.4 kg

```
 * sqlite:///my_data1.db
Done.
```

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad
  2015/December/22

MIN_DATE
2015-12-22

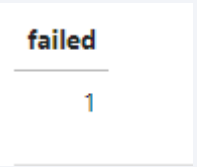
Successful Drone Ship Landing with Payload between 4000 and 6000

- List of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes are:
- Total missions == 101
- Successful missions == 100
- Failed missions == 1



Boosters Carried Maximum Payload

- Extract of table with the names of the booster which have carried the maximum payload mass
- The maximum payload mass was 15600 kg

Booster_Version	(SELECT MAX(payload_mass__kg_) FROM SPACEXTABLE)
F9 v1.0 B0003	15600
F9 v1.0 B0004	15600
F9 v1.0 B0005	15600
F9 v1.0 B0006	15600
F9 v1.0 B0007	15600
F9 v1.1 B1003	15600
F9 v1.1	15600
F9 v1.1	15600
F9 v1.1	15600
F9 v1.1	15600
F9 v1.1	15600
F9 v1.1 B1011	15600
F9 v1.1 B1010	15600
F9 v1.1 B1012	15600
F9 v1.1 B1013	15600
F9 v1.1 B1014	15600
F9 v1.1 B1015	15600
F9 v1.1 B1016	15600
F9 v1.1 B1018	15600
F9 FT B1019	15600
F9 v1.1 B1017	15600
F9 FT B1020	15600
F9 FT B1021.1	15600
F9 FT B1022	15600
F9 FT B1023.1	15600
F9 FT B1024	15600
F9 FT B1025.1	15600
F9 FT B1026	15600
F9 FT B1029.1	15600
F9 FT B1031.1	15600
F9 FT B1030	15600
F9 FT B1021.2	15600
F9 FT B1032.1	15600
F9 FT B1034	15600
F9 FT B1035.1	15600
F9 FT B1029.2	15600
F9 FT B1036.1	15600
F9 FT B1037	15600
F9 B4 B1039.1	15600
F9 FT B1038.1	15600
F9 B4 B1040.1	15600
F9 B4 B1041.1	15600
F9 FT B1031.2	15600
F9 B4 B1042.1	15600
F9 FT B1035.2	15600
F9 FT B1036.2	15600
F9 B4 B1043.1	15600
F9 FT B1032.2	15600

2015 Launch Records

- List of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

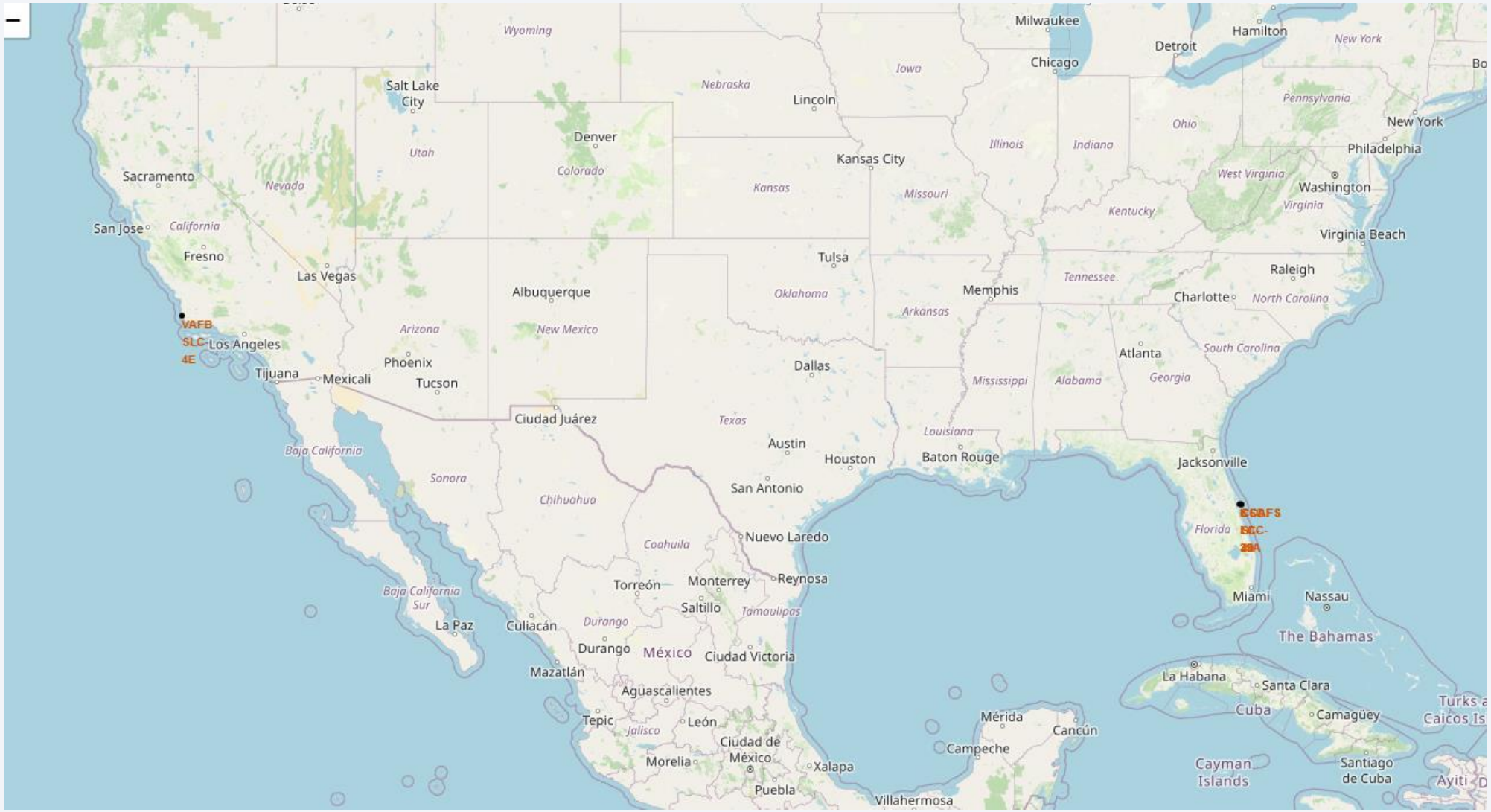
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Rank of 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

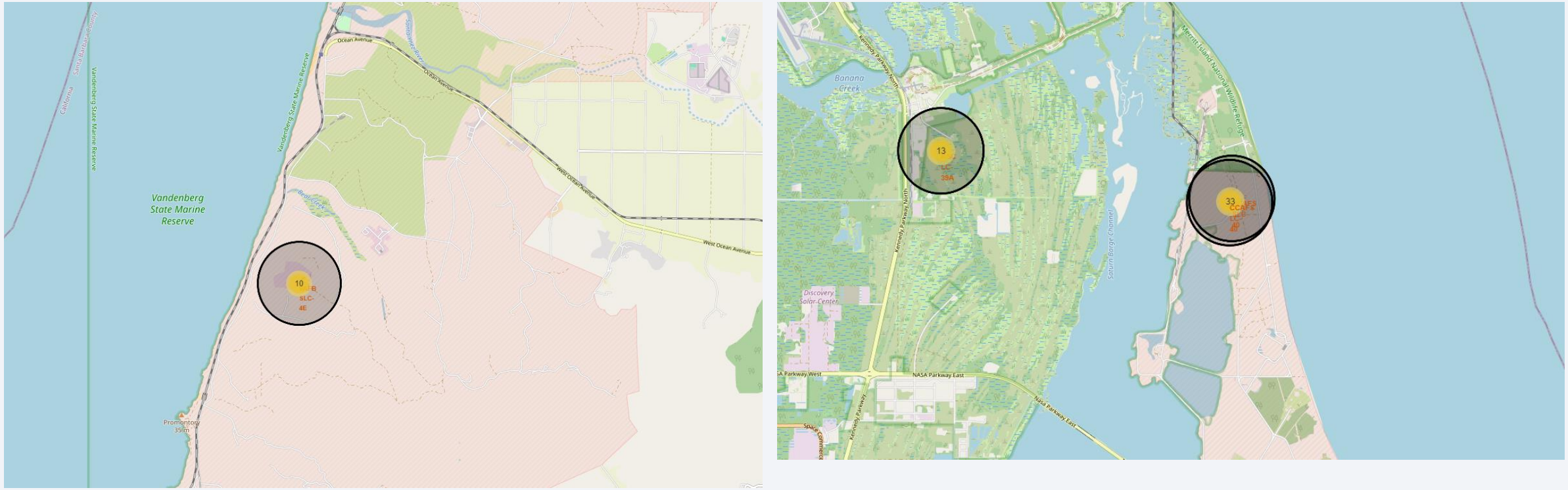
Date	Landing_Outcome
2015-01-10	Failure (drone ship)
2015-04-14	Failure (drone ship)
2015-12-22	Success (ground pad)
2016-01-17	Failure (drone ship)
2016-03-04	Failure (drone ship)
2016-06-15	Failure (drone ship)
2016-07-18	Success (ground pad)
2017-02-19	Success (ground pad)


Section 3
Launch Sites
Proximities Analysis

All launch sites

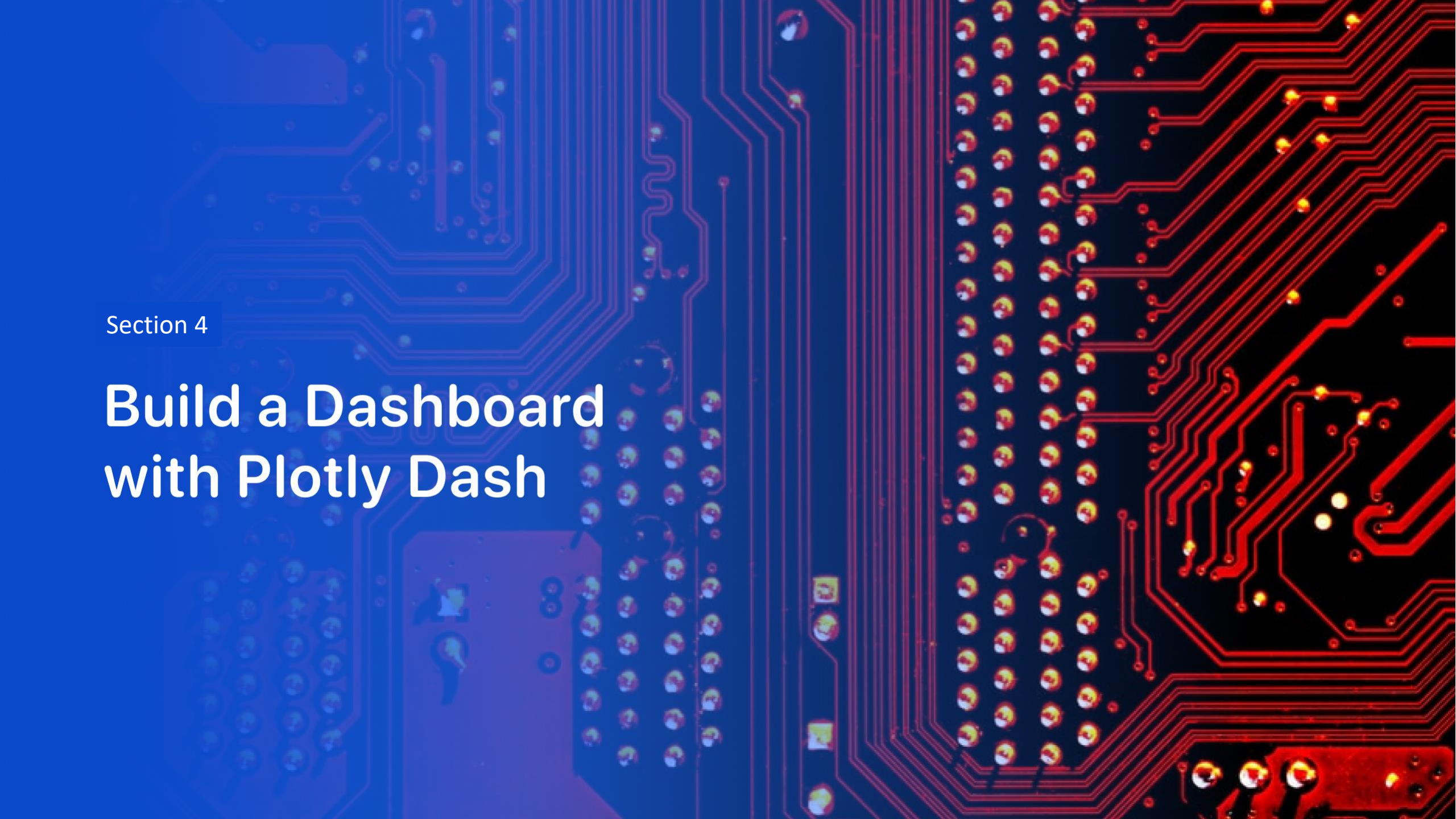


color-labeled launch outcomes on the map

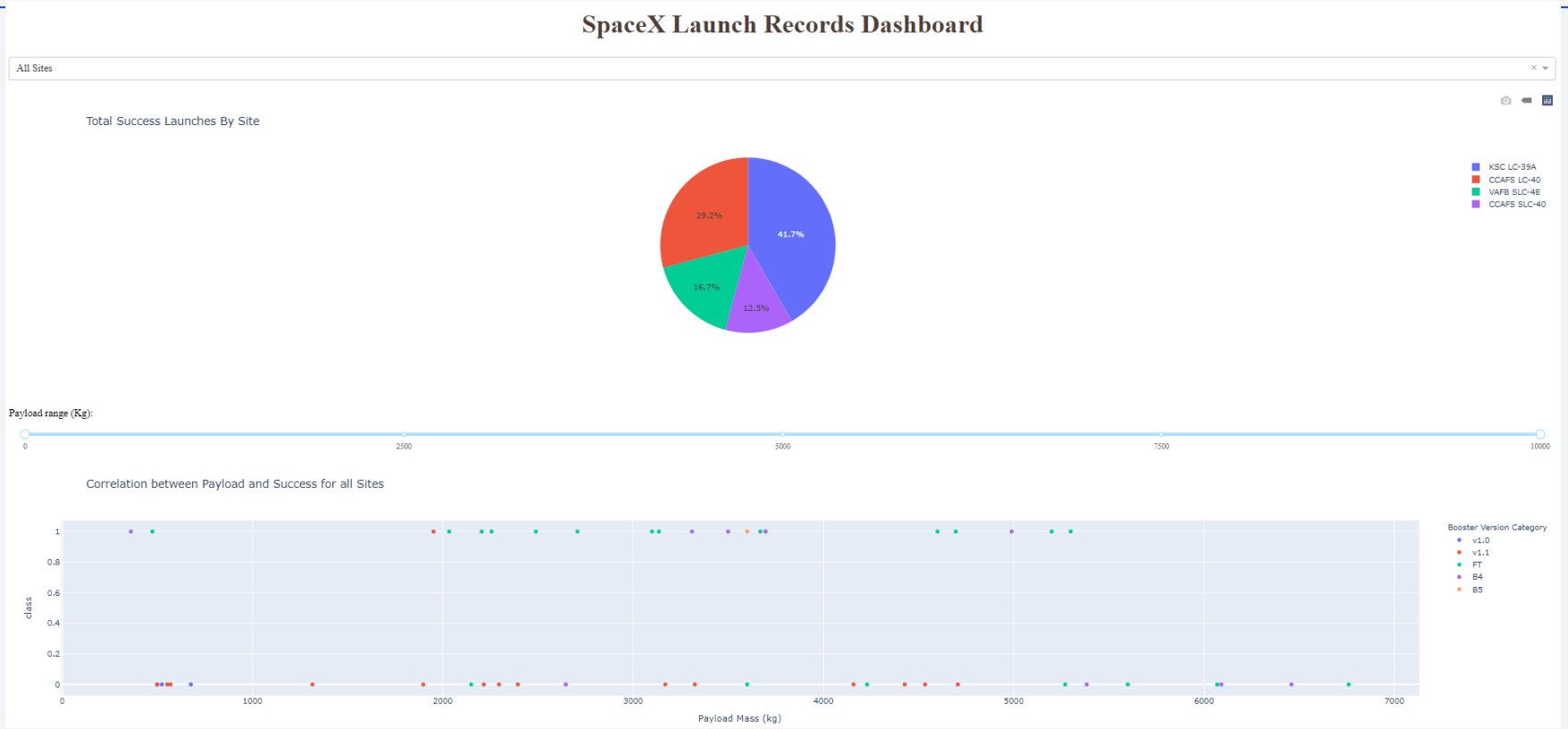


Red ones indicate failure

Distance between launch site and coast


Section 4
Build a Dashboard with Plotly Dash

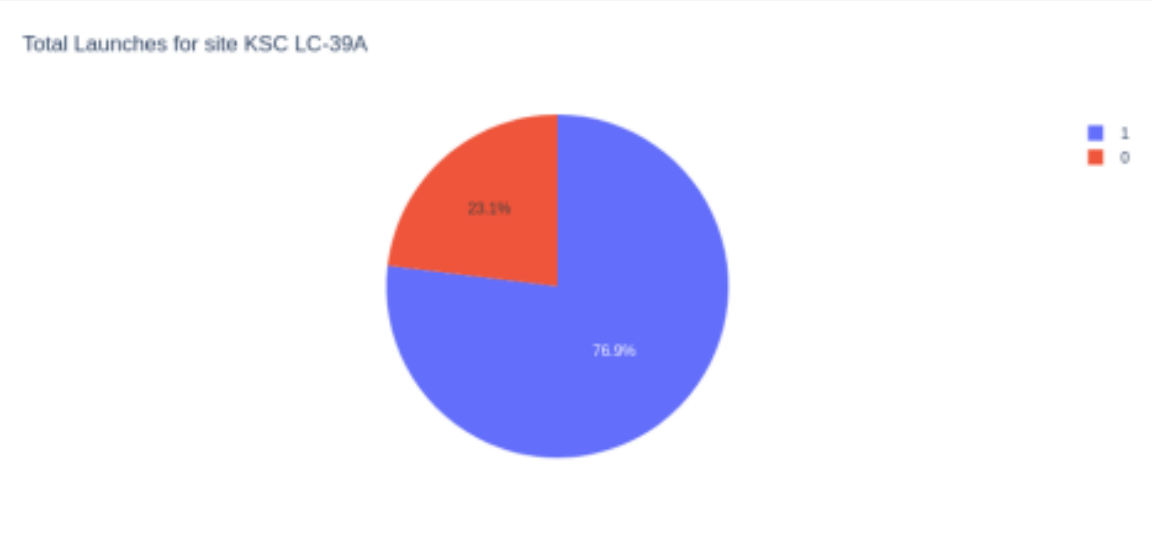
SpaceX Launch records dashboard



Tiltle, drop down input, pie chart of total launches, payload barm corration between payload and success

Launch site with highest launch success ratio

- Screenshot of the piechart for the launch site with highest launch success ratio
- With 76.9% the site is KSC LC-39A



Thank you!