Winning Space Race with Data Science

Gabriel Bessette
17.12.1998

Outline

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

Executive Summary

- Methodologies

  - Data extraction (API/Web-scraping) of Falcon F9 launches & landings
  - Data manipulation and EDA (SQL/Pandas)
  - Interactive data visualization (Folium/Plotly/Dash)
  - Predictive analysis using classification models (Scikit-learn)

- Results

  - Payload mass and success rate (2010-2020)
  - Best performing orbit type
  - Launch sites' locations
  - Best performing launch site
  - Best performing booster version
  - Best classification model

Introduction

Project background and context

- Reduced costs in space exploration are mainly due to reusability of rocket launchers
- Hence the cost of a SpaceX Flacon 9 is only 62M$ compared to 165M$ for other providers
- However, this cost is dependent on the success/failure rate of the operations

Project's aim:

- Predict if Falcon 9 will land successfully,

by using available records of past rocket launches (2010-2020)

Section 1

Methodology

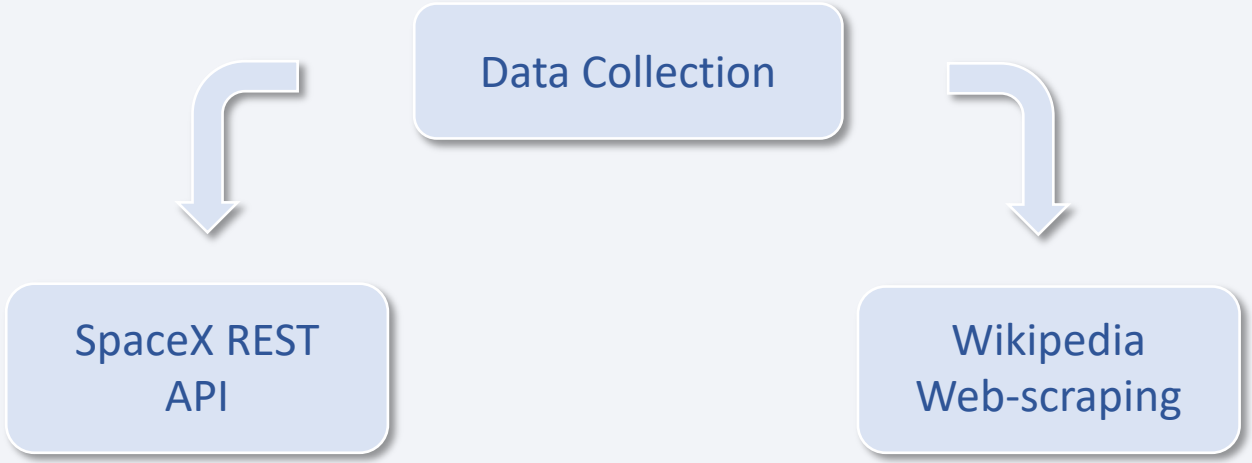
Methodology

Executive Summary

- Data collection methodology:
  - Collect data from SpaceX data base using REST API
  - Collect data from Falcon F9 launches from Wikipedia tables using Web-scraping
- Perform data wrangling
  - Cleaning data and classifying data for successful landings
- 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, and evaluating classification models (Logreg, SVM, Tree, KNN) using cross-validation, and grid search for hyper-parameter optimization

Data Collection

Data sets on Falcon F9 launches have been collected using two methods:



Data Collection – SpaceX API

- SpaceX REST calls URL API_url

- Dataset information

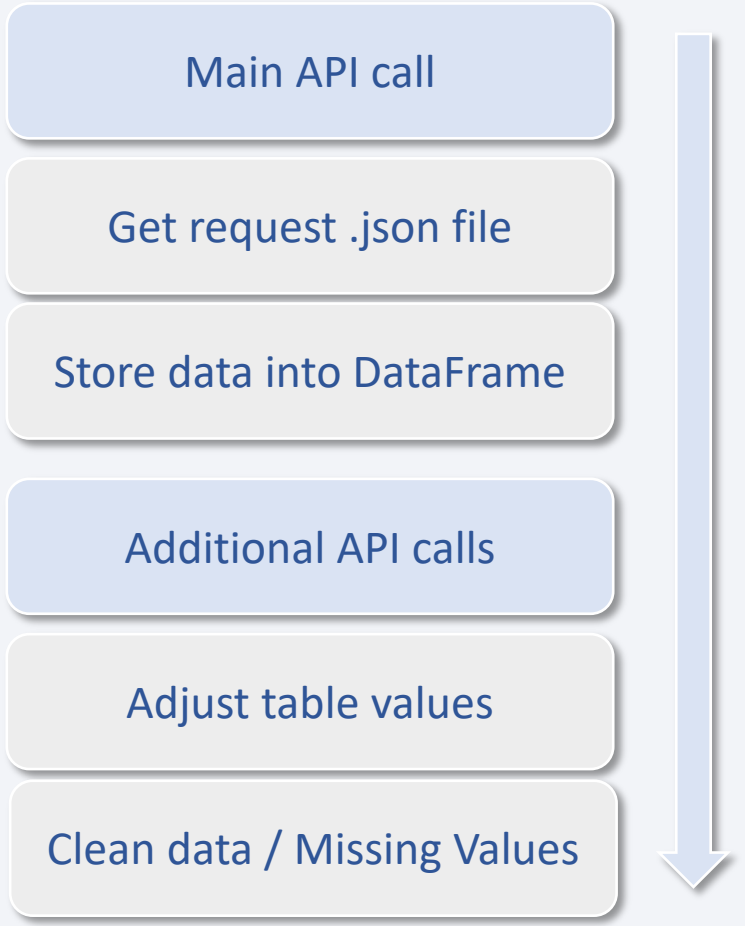
  Categorical

  Date, Booster version, Orbit type, Launch site, Launch Outcome, Grid fins, Reused booster, Legs, Landing pad, Serial number

  Numerical

  Flight number, Payload mass, Number of flights, Block, Reused count, Longitude, Latitude

Ref: SpaceX_API_jupyter_notebook



Data Collection - Scraping

- Web scraping URL webscraping_url

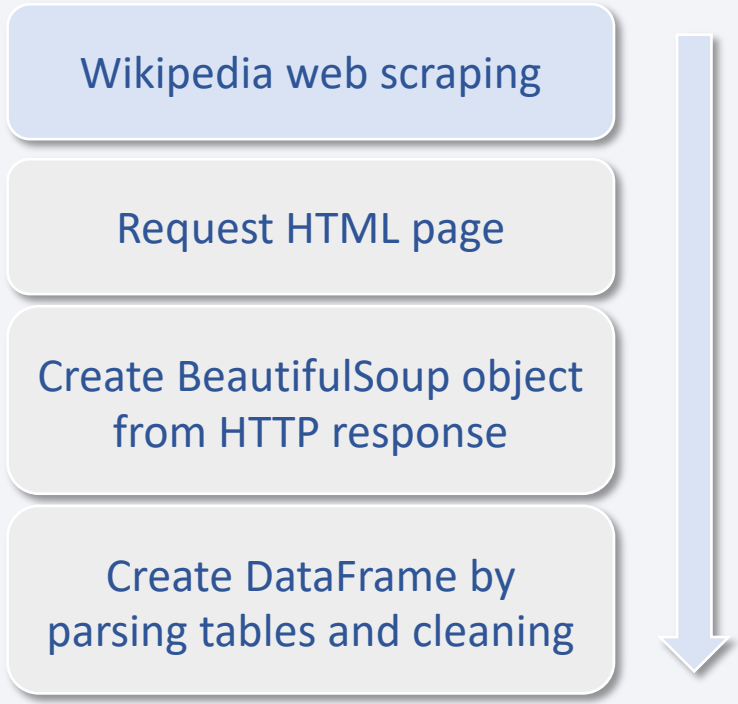
- Dataset information

  Categorical

  Date, Time, Booster version, Payload, Orbit type, Launch site, Launch Outcome, Customer, Landing Outcome

  Numerical

  Flight number, Payload mass



Ref: Web_scraping_jupyter_notebook

Data Wrangling

Constructing labels for (un)successful Falcon 9 landings

Dataset → Differentiate good from bad landing outcomes from a set of features → Create a new column of flags for successful (1) /unsuccessful (0) landing outcomes

	Outcome
0	True ASDS
1	None None
2	True RTLS
3	False ASDS
4	True Ocean
5	False Ocean
6	None ASDS
7	False RTLS

	Class
0	1
1	0
2	1
3	0
4	1
5	0
6	0
7	0

Ref: Data_wrangling_jupyter_notebook

EDA with Data Visualization

Influence of Flight Number, Payload Mass, and Launch Site on successful/failing landings

- Payload Mass (kg) vs Flight Number
- Launch Site vs Flight Number
- Launch Site vs Payload Mass (kg)

Influence of Orbit Type, Flight Number and Payload Mass on successful/failing landings

- Success Rate vs Orbit Type
- Orbit Type vs Flight Number
- Orbit Type vs Payload Mass (kg)

Evolution of landings' success rate from 2010 to 2020

- Success Rate vs Year

Ref: EDA_mpl_jupyter_notebook

EDA with SQL

Summary of the SQL queries

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- 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 acheived.
- 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. Use a subquery
- 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

Ref: EDA_SQL_jupyter_notebook

Build an Interactive Map with Folium

- Spotting SpaceX launch sites on USA map:

  Marking launch sites on map using **Circles** and **Markers**

- Marking success/failure landings for each site on map using **MarkerCluster**:

  Specify green color for successful landings and red color for unsuccessful landings

- Study the distances between a launch site to its proximities

  Specifying the distances between a launch site to its proximities using **Lines** and distance value using **Markers**

Ref: Viz_Folium_jupyter_notebook

Build a Dashboard with Plotly Dash

Summary of Graphs

- **Pie chart**: specify proportion of successful launches for each launch site
- **Scatter plot** : determine the correlation between Payload Mass and Launch Success Rate for each Falcon 9 booster version

Summary of Interactions

- **Markdown** for selection of specific launch site (including “all sites”), applied to both graphs
- **Payload mass range slider** for scatter plot to differentiate heavy from light launches

Ref: Viz_Dash_python_script

Predictive Analysis (Classification)

Model development process

Separate success/failure labels from input (all other features)

Normalise input

Split labels and inputs into training (80%) and testing sets (20%)

Train each model for different hyper-parameters and using cross-correlation

Evaluate and compare best models on test set using the accuracy metric

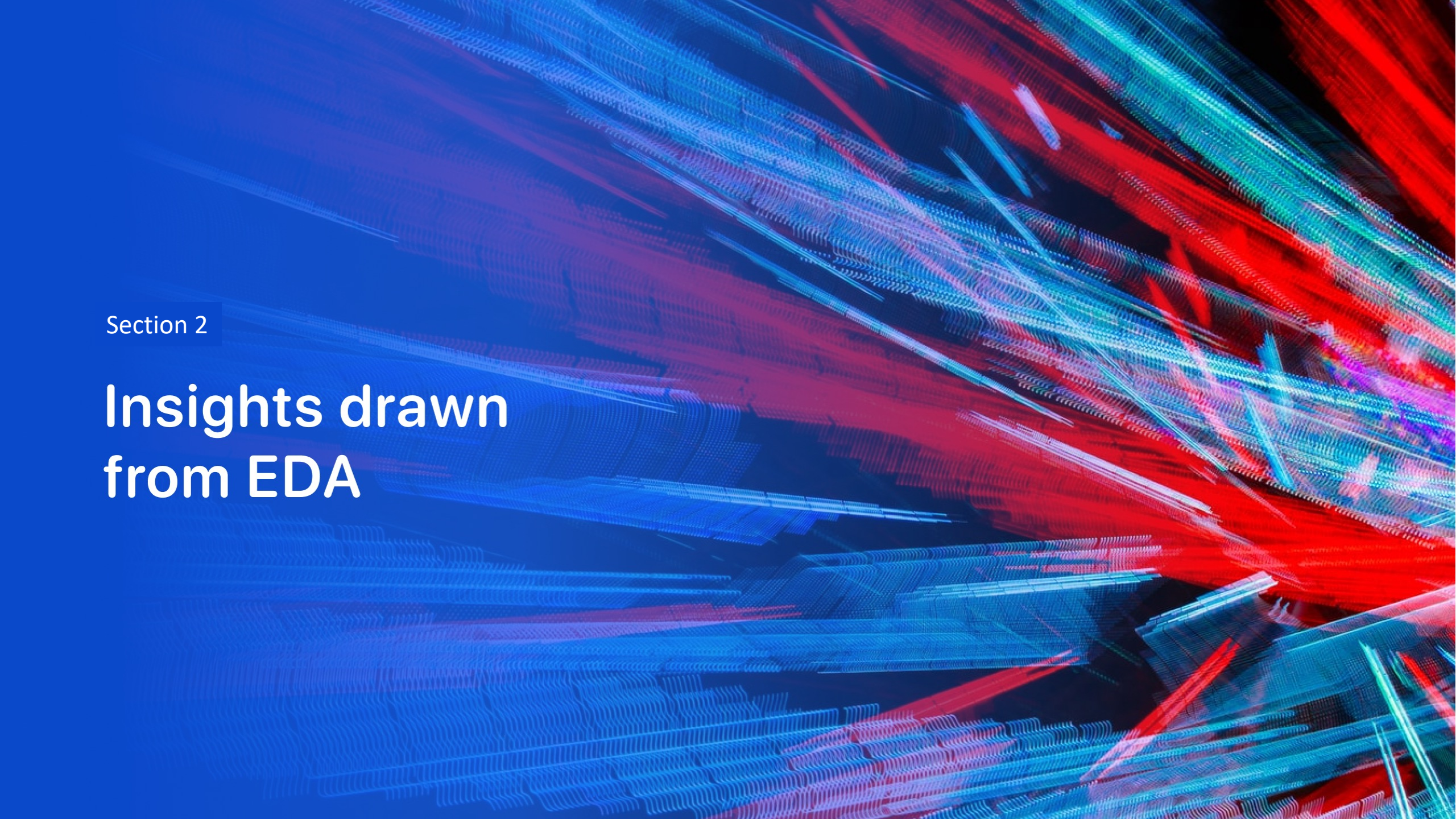
Types of Models Studied

- Logistic Regression
- Support Vector Machines
- Decision Tree
- K Nearest Neighbors

Ref: ML_jupyter_notebook

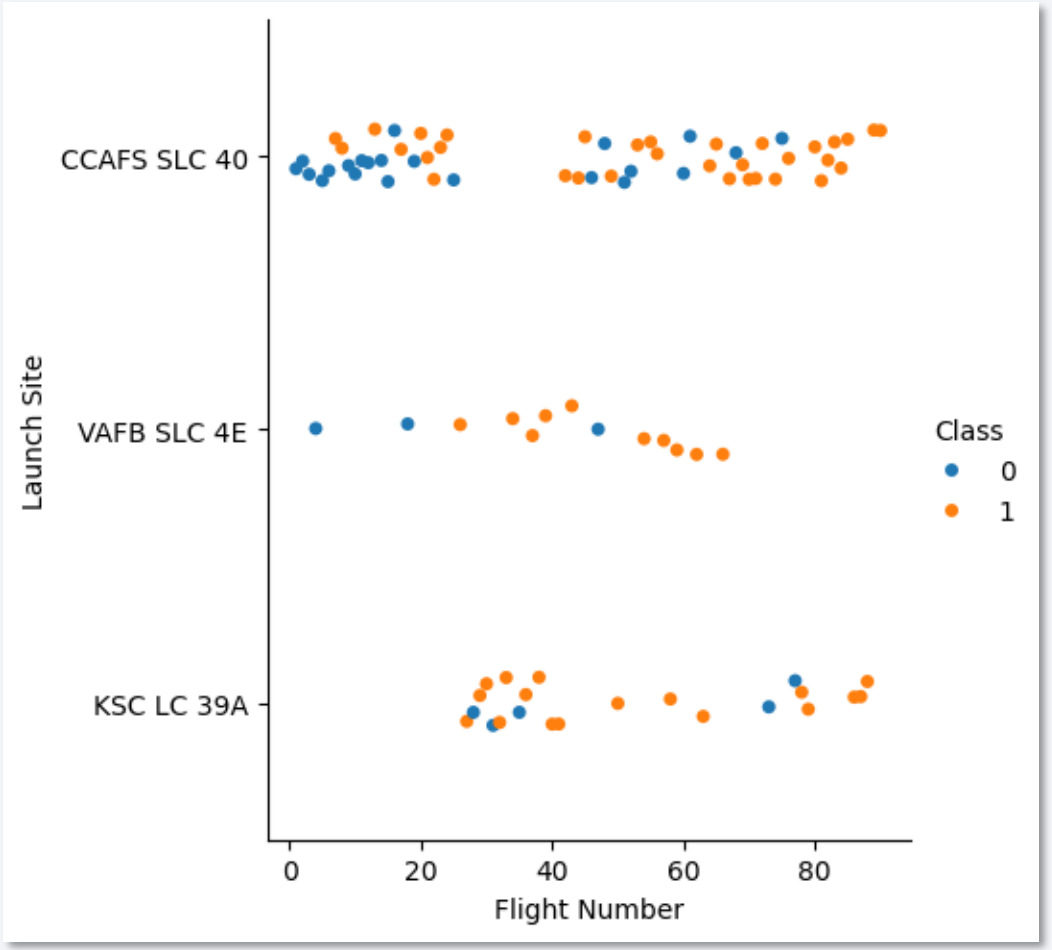
Results

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


Section 2
Insights drawn from EDA

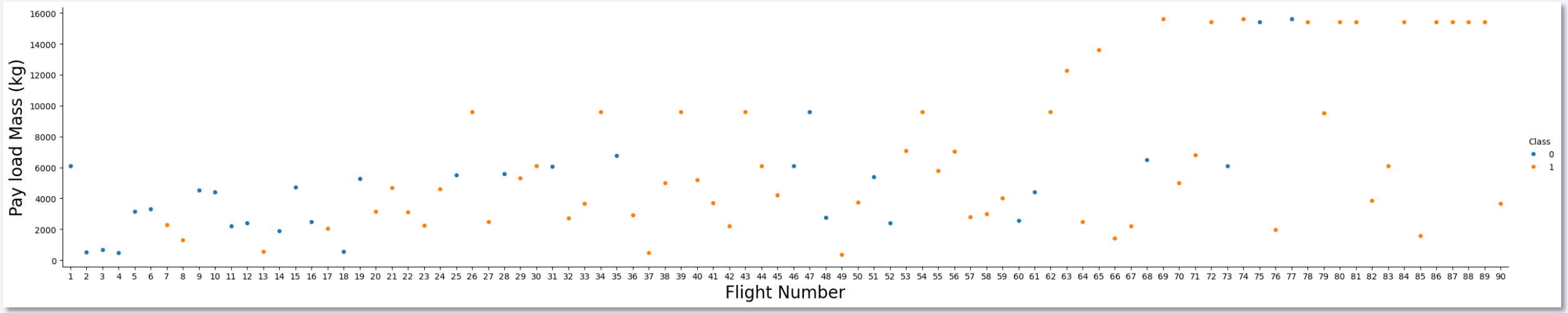
Flight Number vs. Launch Site

- Most of the 20 first launches departed from CCAFS site
- Rockets launched from CCAFS mostly failed to land during the first tests (for low FN). However, the scatter plot indicates less and less failure throughout the years, as the FN increases.
- This tendency can also be observed for other launch sites, such as VAFB and KSC.



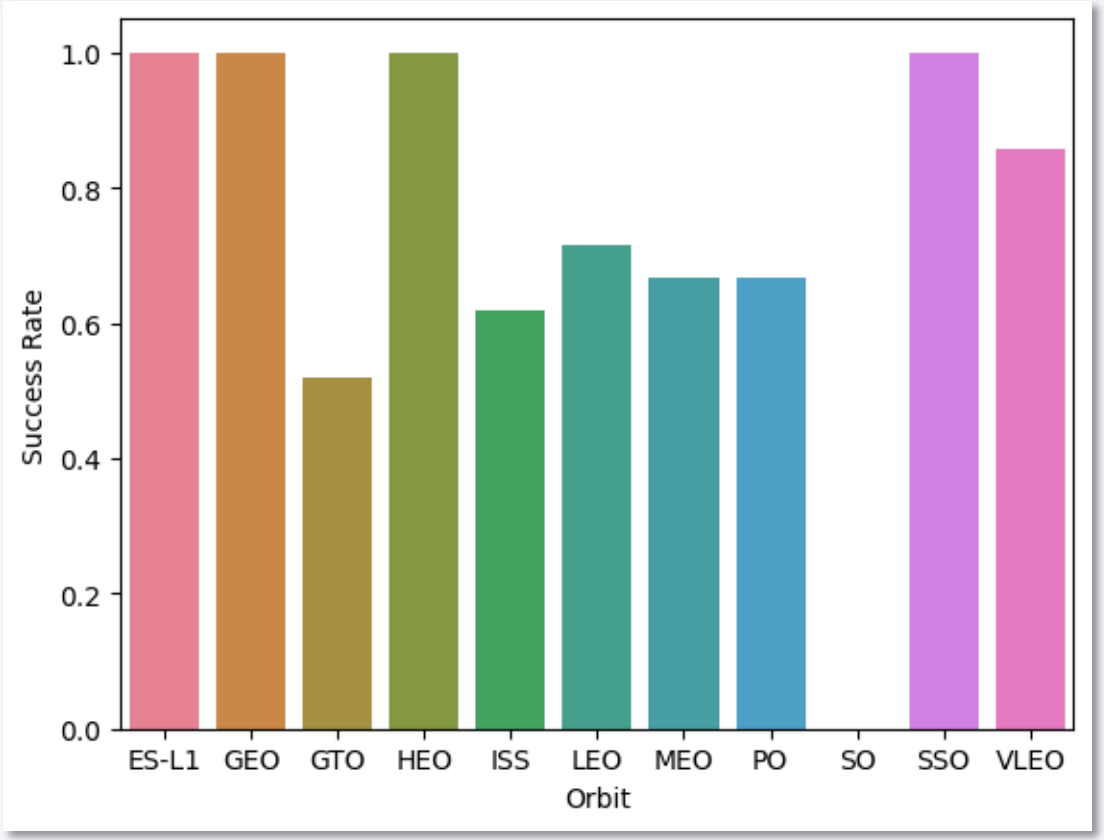
Payload vs. Launch Site

- Throughout the years, heavy launchers (high payload mass) outnumber lightweight launchers
- Success rate increases with FN



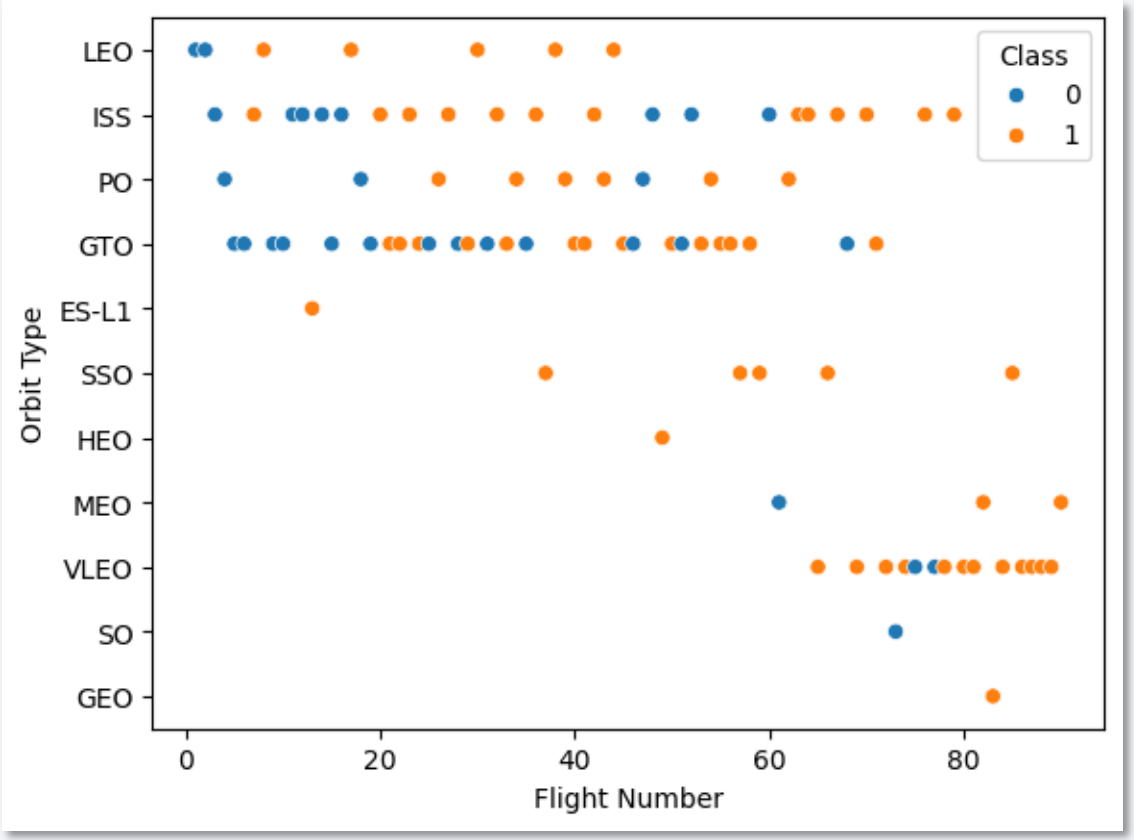
Success Rate vs. Orbit Type

- Orbit types ES-L1, GEO, HEO, SSO and VLEO show highest success rate, above 80%.
- The average success rate for all other orbits is around 60%, except for the SO orbit which shows 0% success rate (outlier).



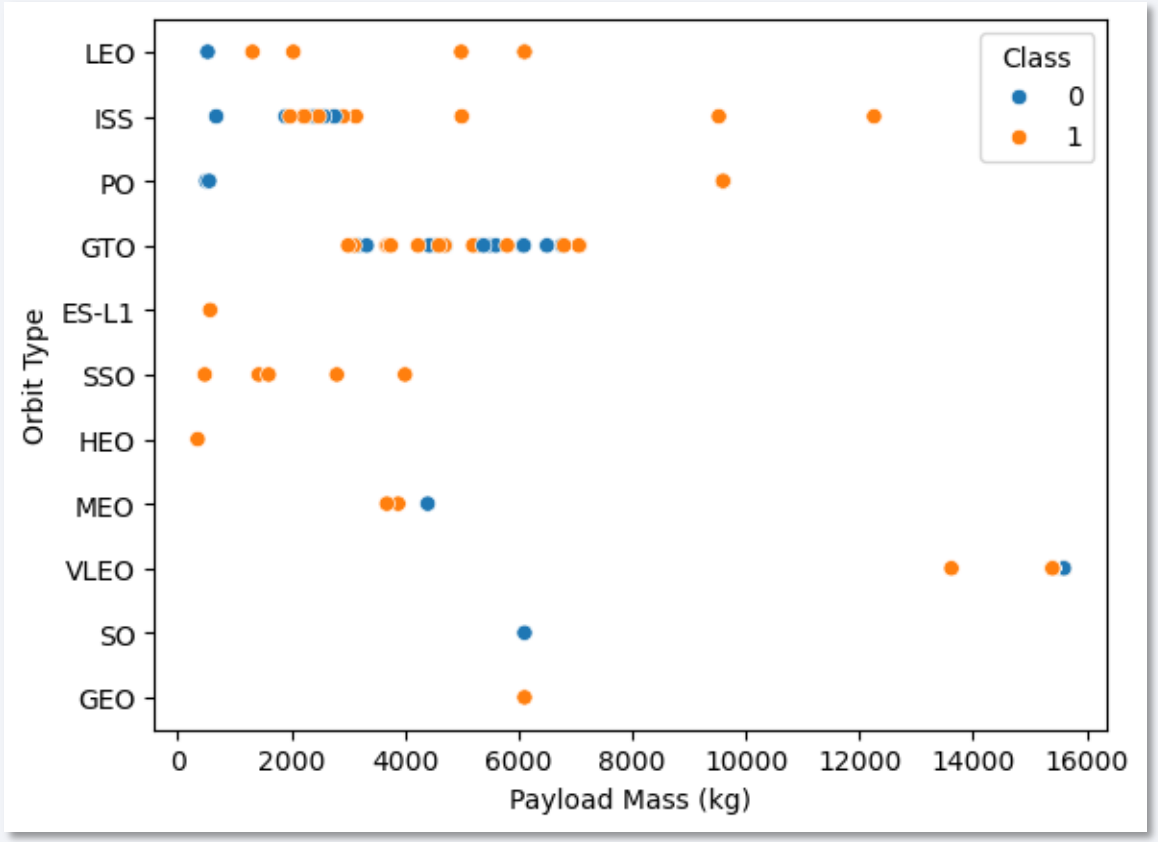
Flight Number vs. Orbit Type

- Data is not representative for the ES-L1, HEO, GEO and SO orbits, since only 1 test was attempted on these orbits.
- For all other orbit types, we observe a progression in success rate throughout the years (or FN).
- Note that SSO orbit shows 100% accuracy for more than 1 attempt.



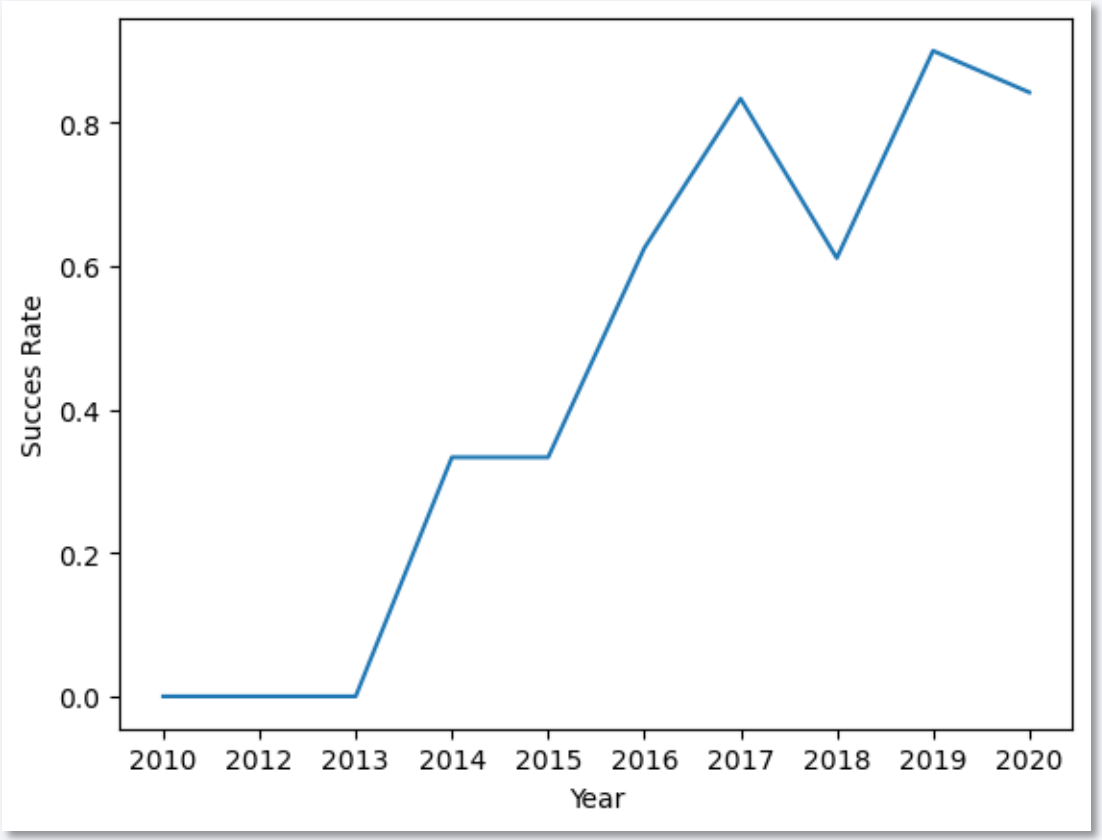
Payload vs. Orbit Type

- From the scatter plot, most of the payload carried by the F9 is below 10’000 (kg)
- Except for the ISS orbit, each orbit type seems to have a small payload range



Launch Success Yearly Trend

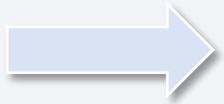
- Success rate clearly increases throughout the years
- A small drop in success rate occurs in 2018, while the general tendency for the success rate is increasing with the years



All Launch Site Names

Names of the unique launch sites

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



Falcon 9 launchers have been launched from 4 different sites from 2010 to 2020

Launch Site Names Begin with 'CCA'

The landing phase of the boosters at CCAFS resulted in either failure or no attempt

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

Total Payload Mass

Total payload mass (kg) carried by Falcon 9 boosters from NASA (2010-2020)

Total_Payload_Mass
107010

Although Falcon 9 is a SpaceX booster, NASA benefitted from this booster to send some payload into space

Average Payload Mass by F9 v1.1

Average payload mass (kg) carried by booster version F9 v1.1 (2010-2020)

AVG_Payload_Mass_F9_v1.1
2534.6666666666665

Typical weight carried by a F9 v1.1 booster

First Successful Ground Landing Date

Although tests started in 2010, it took 5 years for boosters to successfully land on a ground pad!

MIN("Date")
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

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

- Missions only failed once
- The payload status was unclear only once
- Else, every mission was a success!

Boosters Carried Maximum Payload

- The F9 B5 is one of the biggest boosters from SpaceX F9 series
- It carried the maximum payload to space

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

2015 Launch Records

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

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

F9 v1.1 **failed twice** to land on drone ship when launched from CCAFS

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

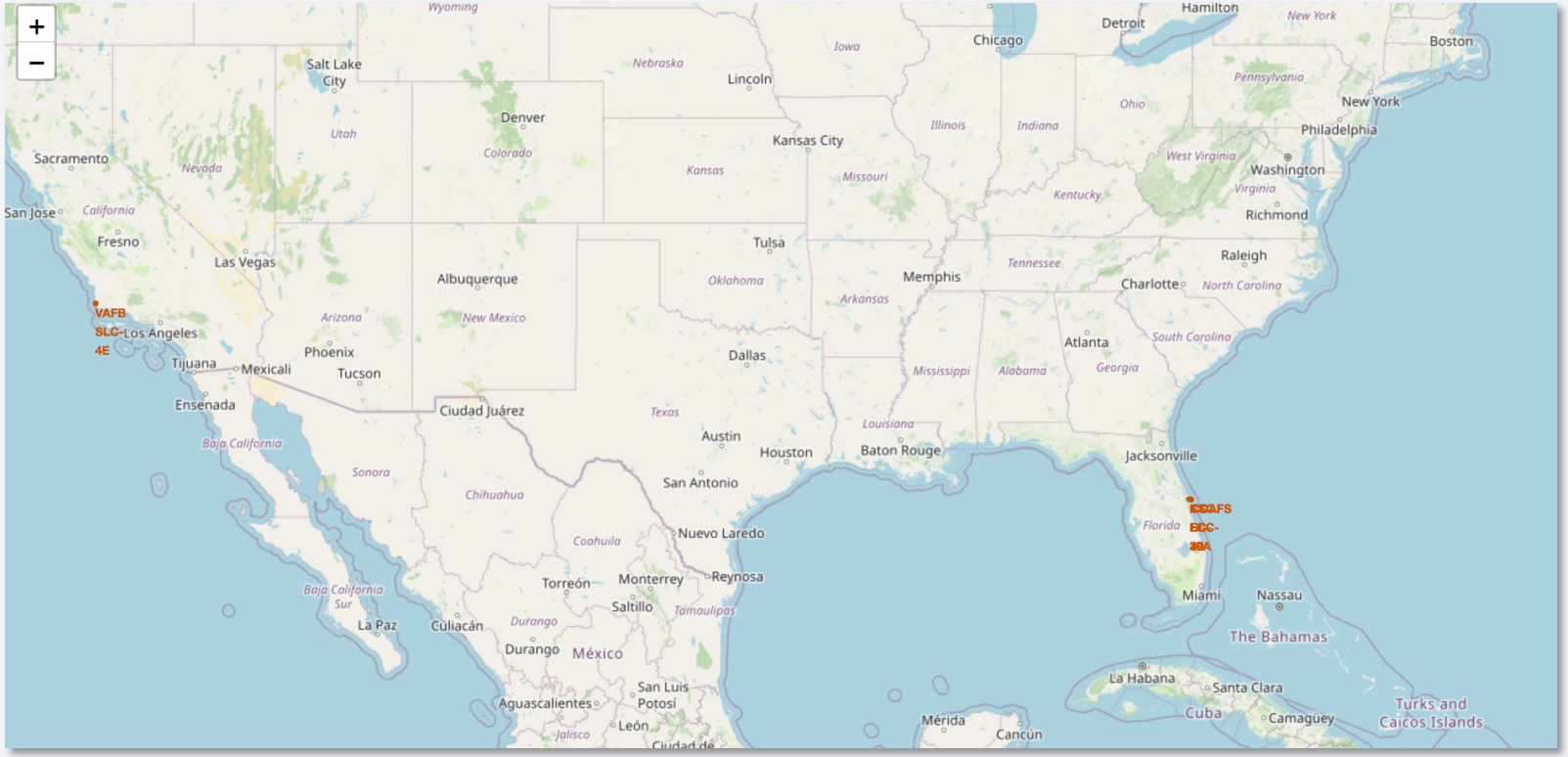
Landing_Outcome	COUNT(*)
Failure (drone ship)	5
Success (ground pad)	3

There is more failure on drone ships than success on ground pads between 2010-06-04 and 2017-03-20.


Section 3
Launch Sites
Proximities Analysis

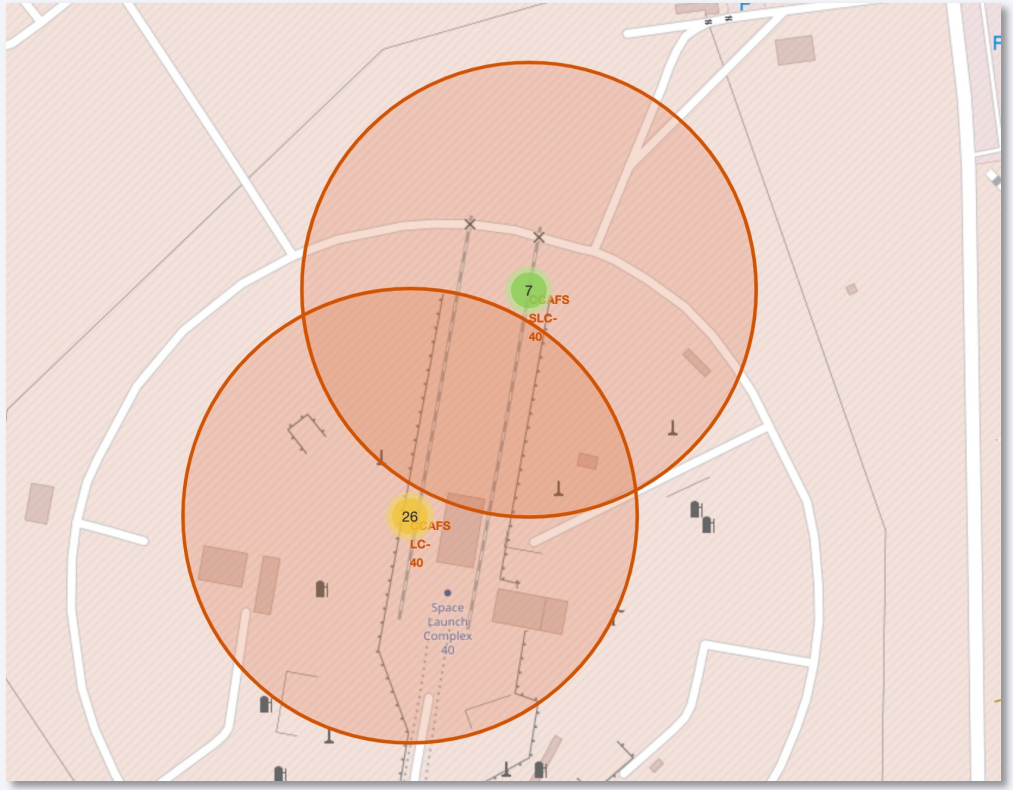
Launch Sites Location

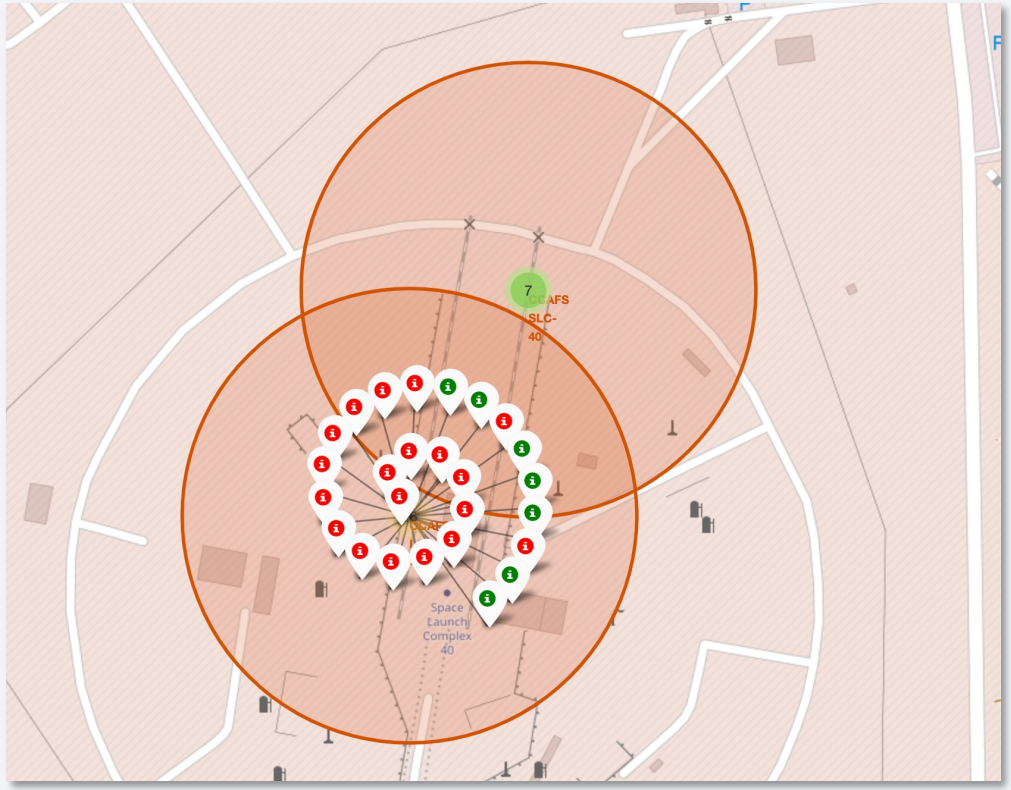
Launch sites are located close to the equator on the east or west coast of the USA



Launch Outcomes on CCAFS Sites

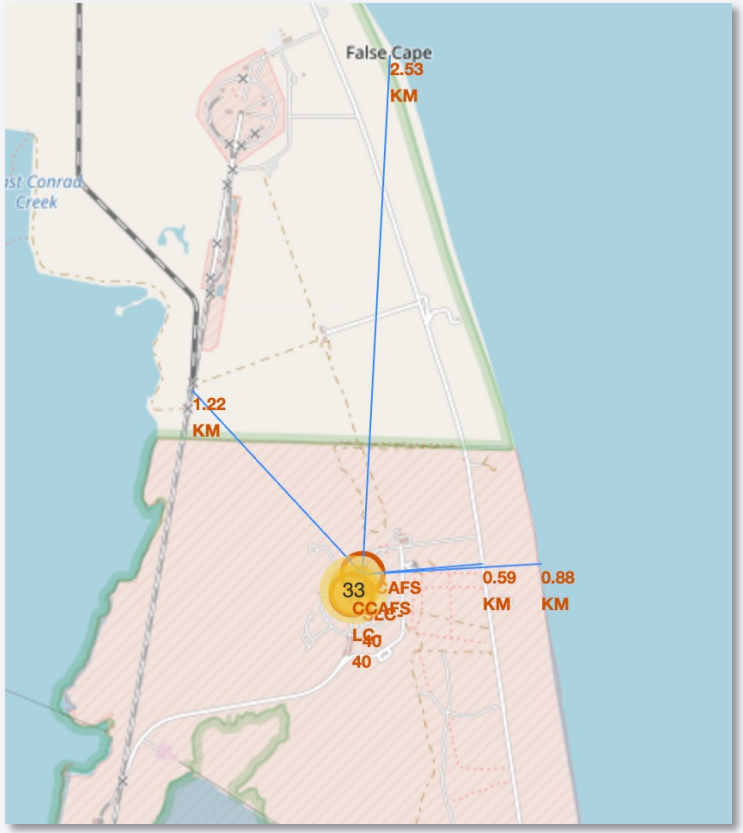
Launch outcomes indicated as grouped markers: green (success), red (failure)

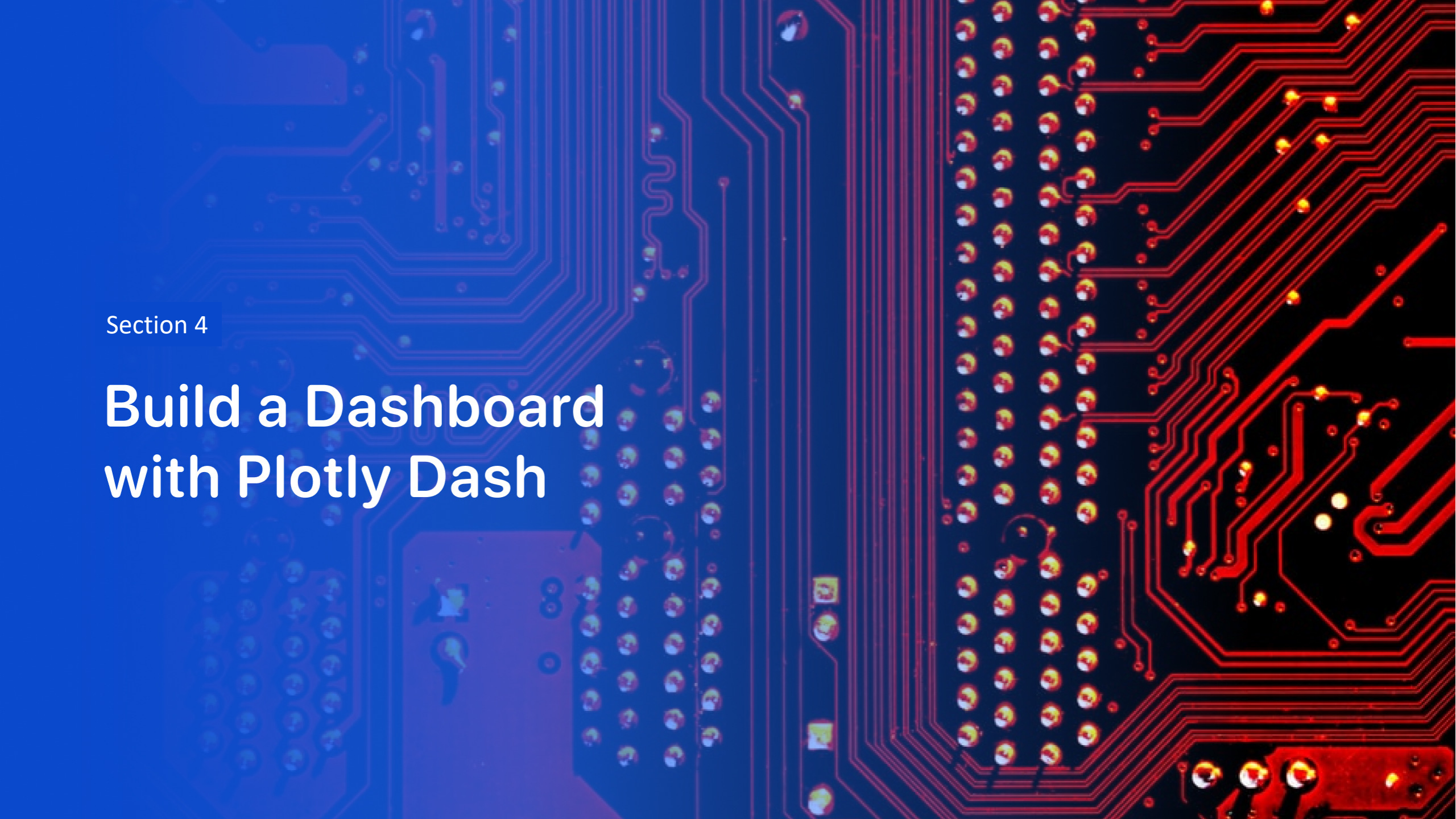




CCAFS Launch Site and Distance to its Proximities

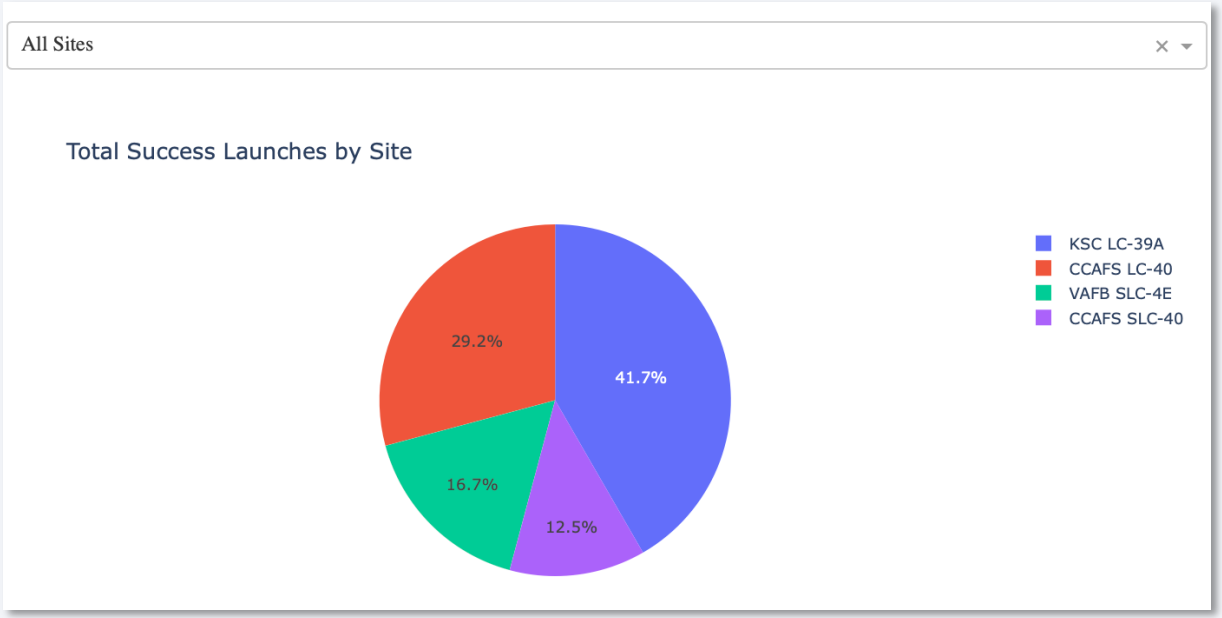
- CCAFS launch sites are located at less than 1km from the coastline for safety purposes.
- CCAFS launch sites are well connected to railways (1.22km) and highways (0.59km) for better logistics.
- Surprisingly, the closest city is only 2.5km away from the launch sites.



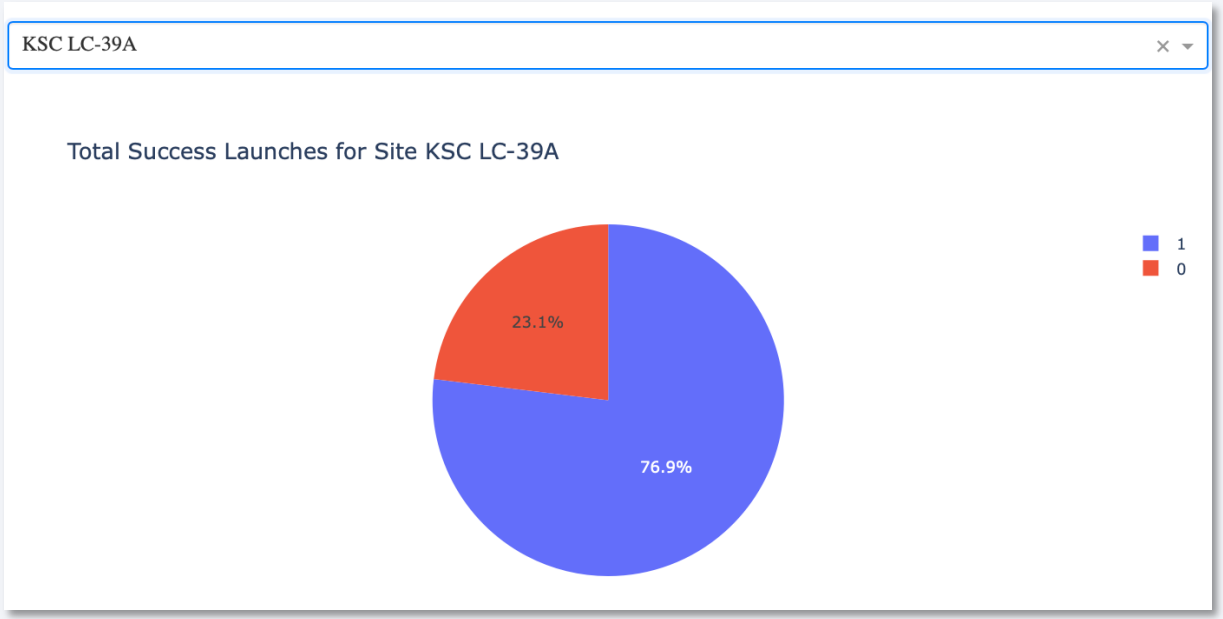

Section 4
Build a Dashboard with Plotly Dash

Launch Success Count by Site

- Majority of successful launches at KSC LC-39A site
- Minority of successful launches at CCAFS SLC-40 site



Launch Site with Highest Launch Success Ratio



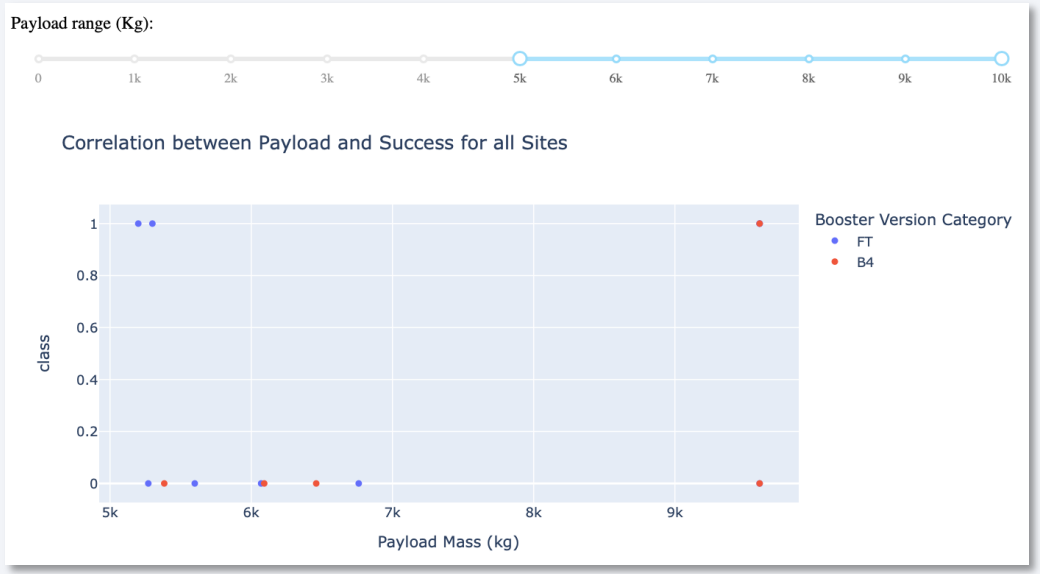
- With 76.9% of launch success rate (blue), KSC LC-39A site has the largest launch success ratio of all sites

Payload Mass vs Launch Outcome for All Sites

Payload Mass range (0kg - 5’000kg)

Boosters v1.0, v1.1, FT, B4 and B5 are used for carrying lightweight payload. The most successful candidate (class = 1) is the FT booster (green).

Payload Mass range (5’000kg - 10’000kg)



Boosters FT and B4 carry the highest payload mass with only 2 successful landings for FT (blue) and 1 for B4 (red) on the specified range.

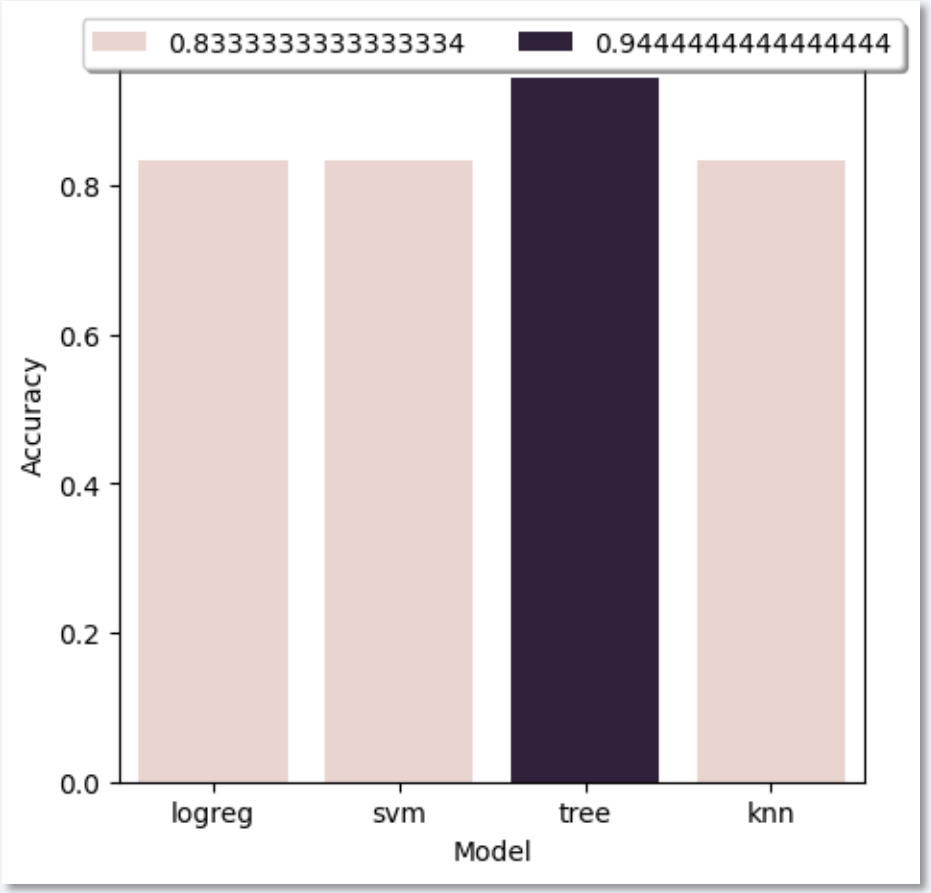
Section 5

Predictive Analysis (Classification)

Classification Accuracy

The model with highest accuracy (94%) is the **decision tree model.**

All other models have accuracy of 83%.

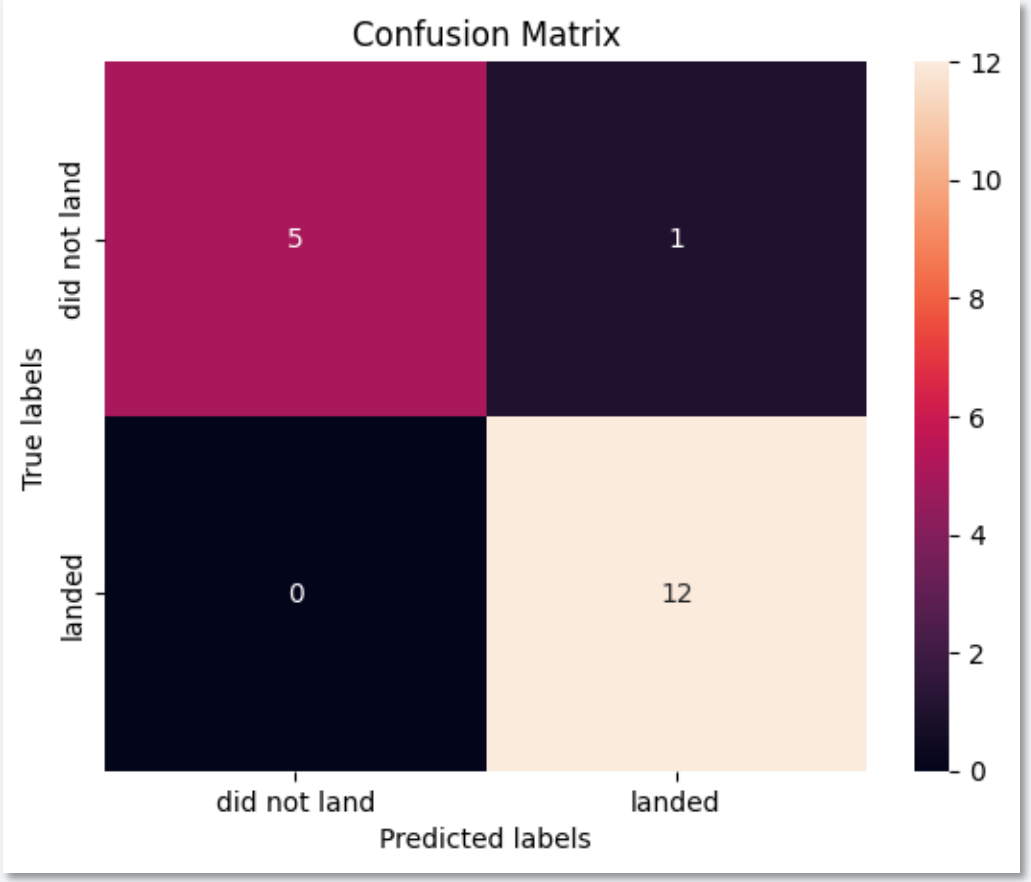


Confusion Matrix of Best Performing Model

Decision Tree Confusion Matrix
(test_size = 18 samples)

- 0 FN
- 1 FP

Model performs well except for 1 test sample (FP).



Conclusions

Key aspects for the success of SpaceX F9 launching and landing are listed below:

- Payload mass and success rate increased throughout the years from 2010 to 2020
- Best performing orbit type is the SSO orbit
- Launch sites are located close to the equator and close to coastlines
- Best performing launch site is KSC LC-39A with 76.9% success rate
- Best performing booster version is the FT for lightweight payload, whereas the success rate for heavy payload is low for both B4 and FT
- Falcon 9 launchers can be classified using a decision tree with 94% accuracy score

Appendix

Link to Github working directory: SpaceX_Project_Github

Thank you!