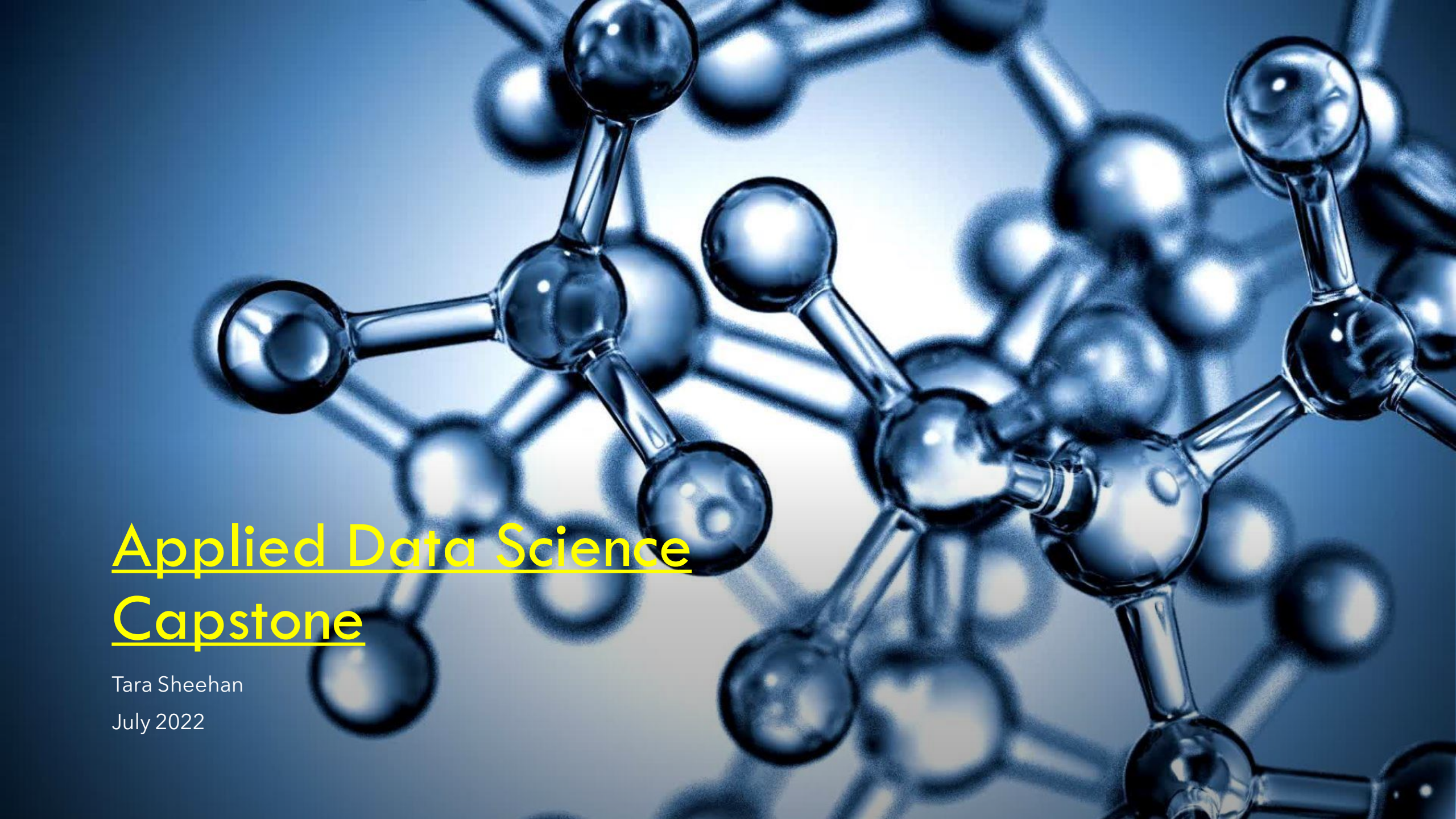


Applied Data Science Capstone

Tara Sheehan

July 2022

OUTLINE

Executive Summary
Introduction
Methodology
Results
Conclusion
Appendix

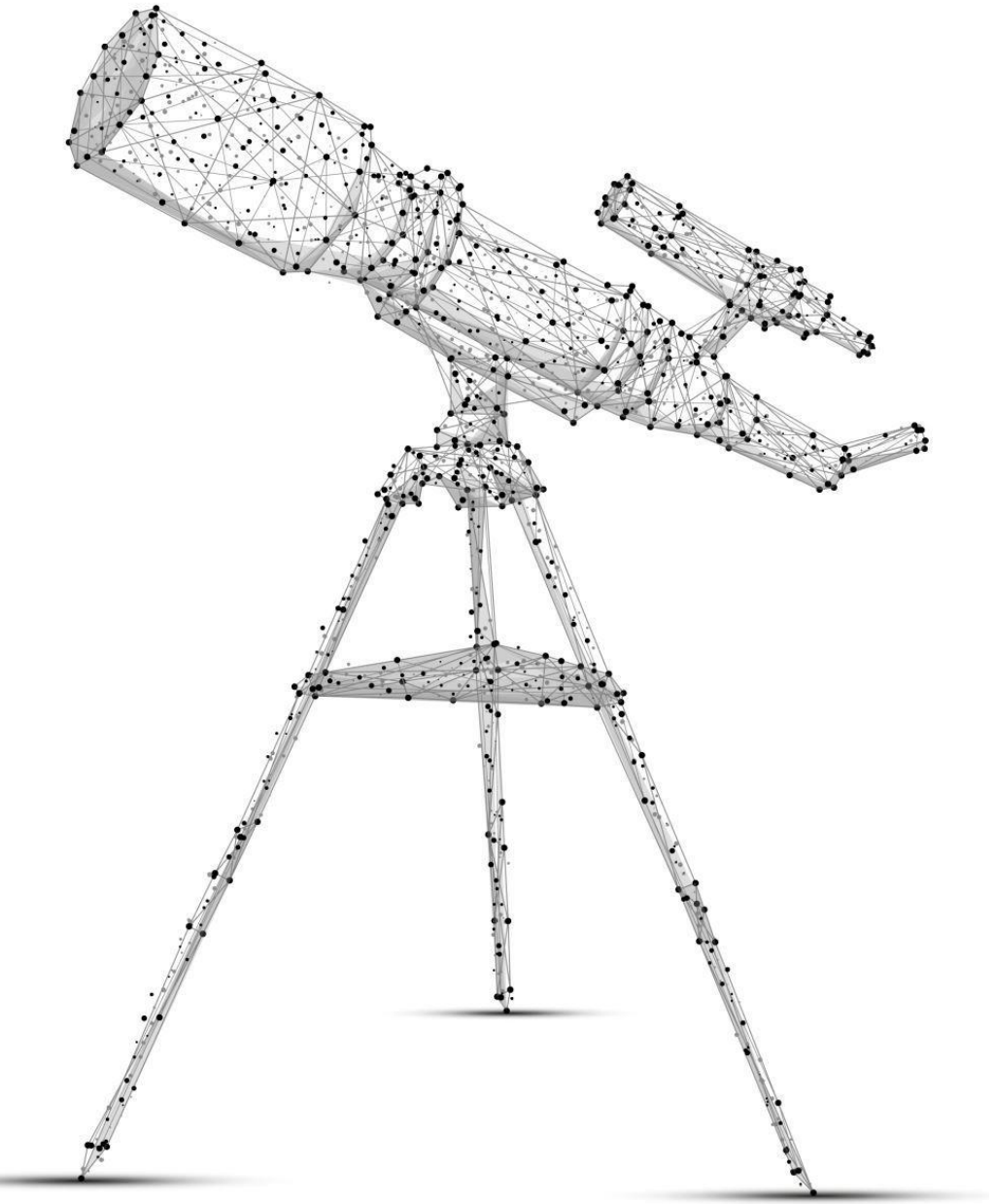
EXECUTIVE SUMMARY

The following methodologies were used to analyze data.

- Data collection using web scraping and SpaceX API
- Exploratory Data Analysis including data wrangling, data visualization and interactive visual analytics
- Machine Learning Prediction

Summary of all results

- It was possible to collect valuable data from public sources
- EDA allowed to identify which features are the best to predict success of launchings
- Machine Learning Prediction showed the best model to predict which characteristcs are important to drive this opportunity by the best way, using all collected data



INTRODUCTION

- Objective - evaluate viability of new company Space Y to compete with Space X
- Answers:
  - Best way to estimate total cost for launches via predicting successful landings of the first stage rockets
  - Where is the best place to launch

METHODOLOGY

METHODOLOGY

Executive Summary

Data collection methodology

Data from Space X came from 2 sources

Space X API (https://api.spacexdata.com/v4/rockets/)

WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)

Perform data wrangling

Collected data was created through a landing outcome label based on outcome data after summarizing and analyzing features

Perform EDA using visualization and SQL

METHODOLOGY

Executive Summary

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

Data that was collected until step was normalized, divided and evaluated by 4 classification models

Accuracy of models evaluated via various parameters

DATA COLLECTION

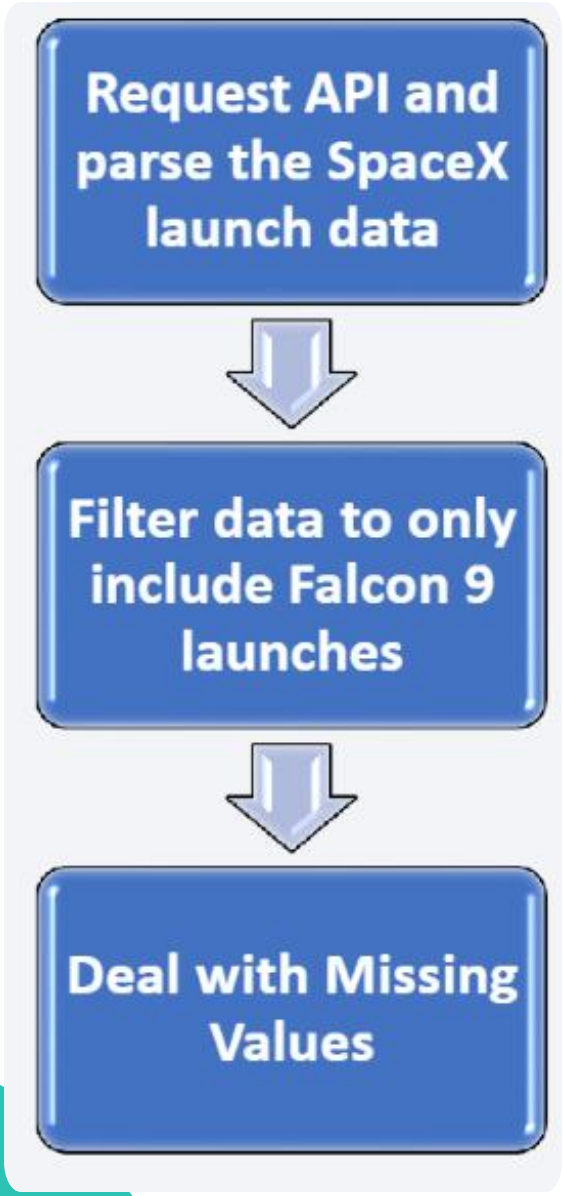
Data sets collected using web scraping from :

Space X API

Https://api.spacexdata.com/v4/rockets/

Wikipedia

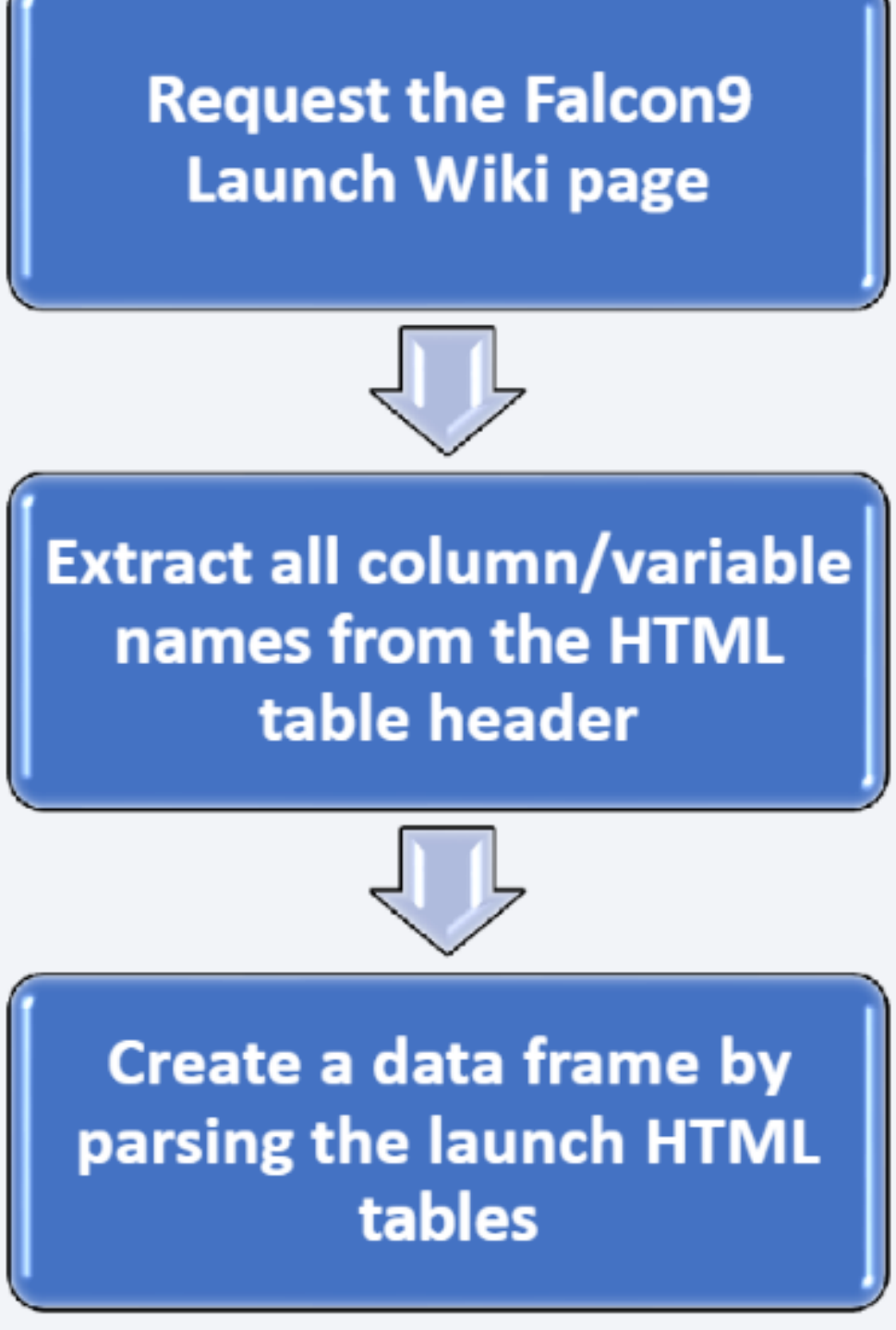
https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches



DATA COLLECTION – SPACEX API

SpaceX offers a public API where data can be obtained and used

This flowchart shows how data was used



DATA COLLECTION - SCRAPING

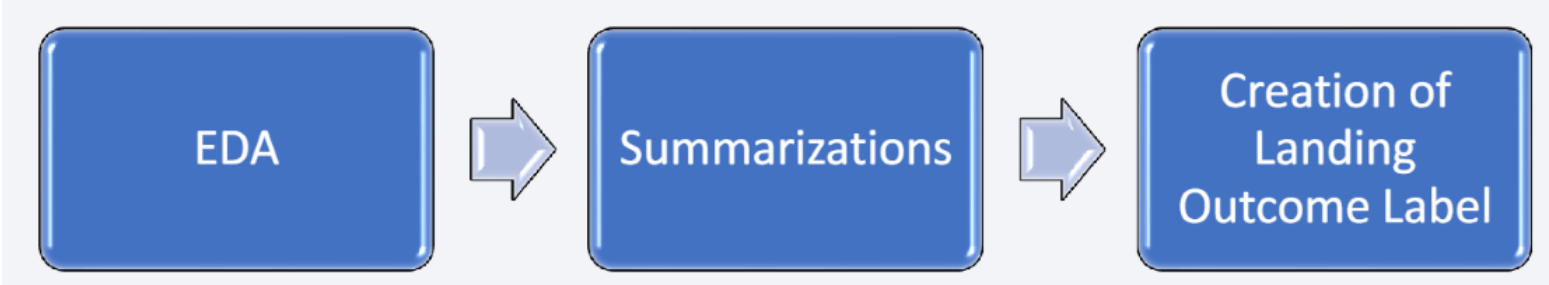
Data from SpaceX launches obtained from Wikipedia according to flowchart

DATA WRANGLING

EDA was first performed on the dataset

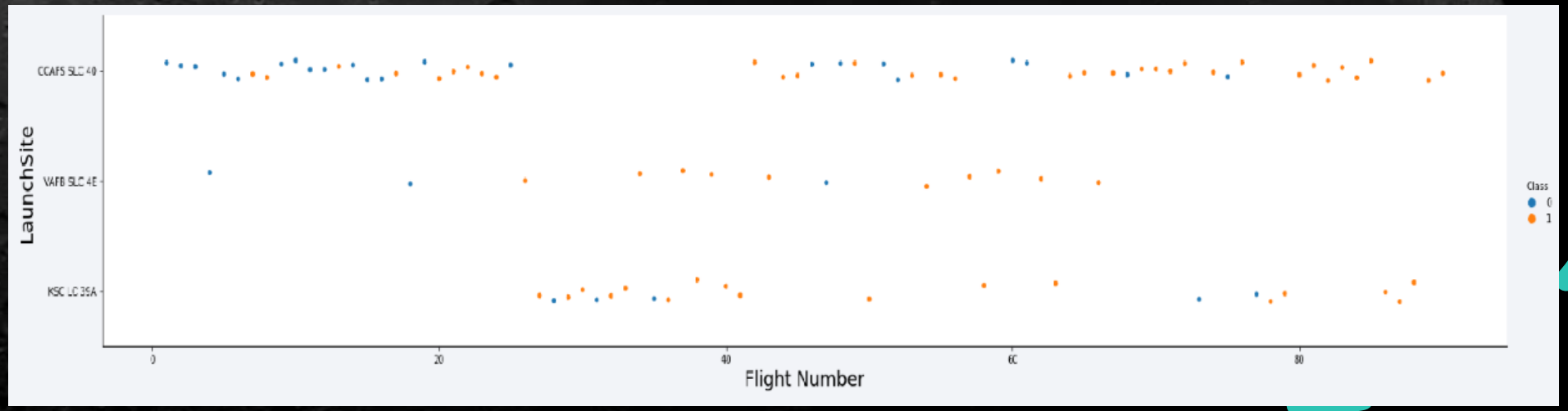
Then summary launches, occurrences of each orbit and mission outcome per orbit type were calculated

Label for landing outcome created from outcome column



EDA WITH DATA VISUALIZATION

Scatterplots and Barplots were created to explore the relationships – Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit & Flight Number, Payload & Orbit



EDA WITH SQL

SQL queries performed:

Names of launch sites in space mission

Top launch sites whose name begins with the string 'CCA'

Total payload mass carried by boosters launched by NASA

Average payload mass carried by booster

Date when the first successful landing outcome in ground pad achieved

Names of the boosters which have a payload between 4000-6000kg

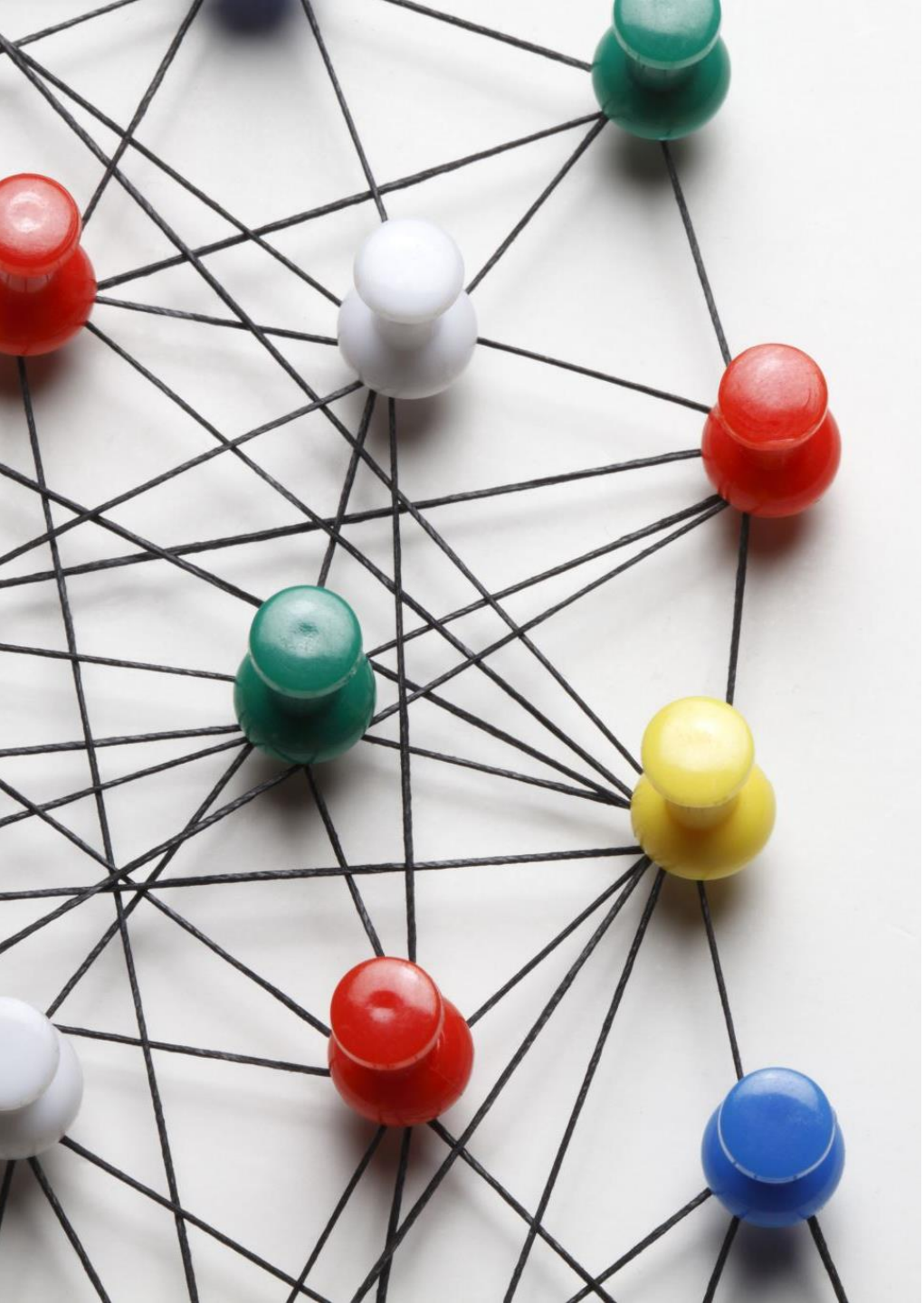
Total number of successful mission outcomes

Total number of failed mission outcomes

Names of boosters which carried maximum payload mass

Failed landing, booster versions and launch names for year 2015

Rank of landing outcomes from 2010-207



BUILD AN INTERACTIVE MAP WITH FOLIUM

Markers, circles, lines and marker clusters were used with Folium Maps

Markers indicate launch sites

Circles are highlighted areas for specific coordinates

Marker clusters indicate groups of events

Lines are used to indicate distances

BUILD A DASHBOARD WITH PLOTLY DASH

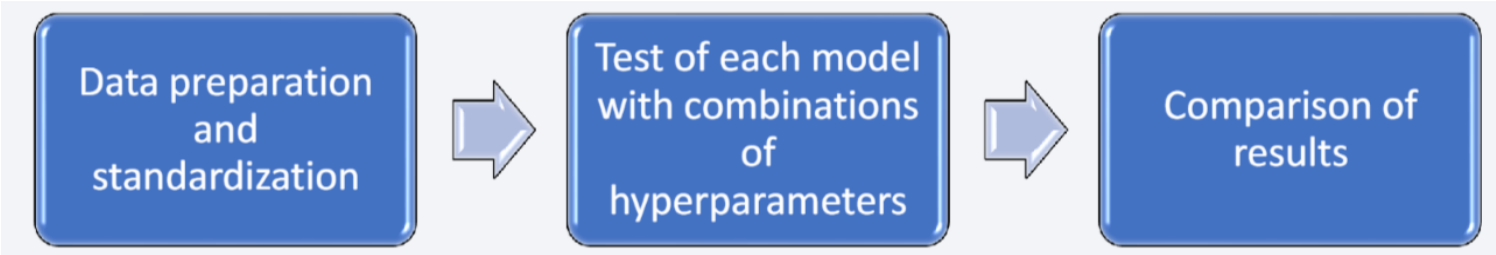
Percentage of launches by site and Payload Range were showed via graphs and plots

These visuals allow analyzation of payloads and launch sites which shows the best place to launch per payload

PREDICTIVE ANALYSIS (CLASSIFICATION)

Four classification models were compared

- logistic regression
- support vector machine
- decision tree
- k nearest neighbors



RESULTS

Exploratory data analysis results:

SpaceX uses 4 launch areas

1st launch – SpaceX and NASA

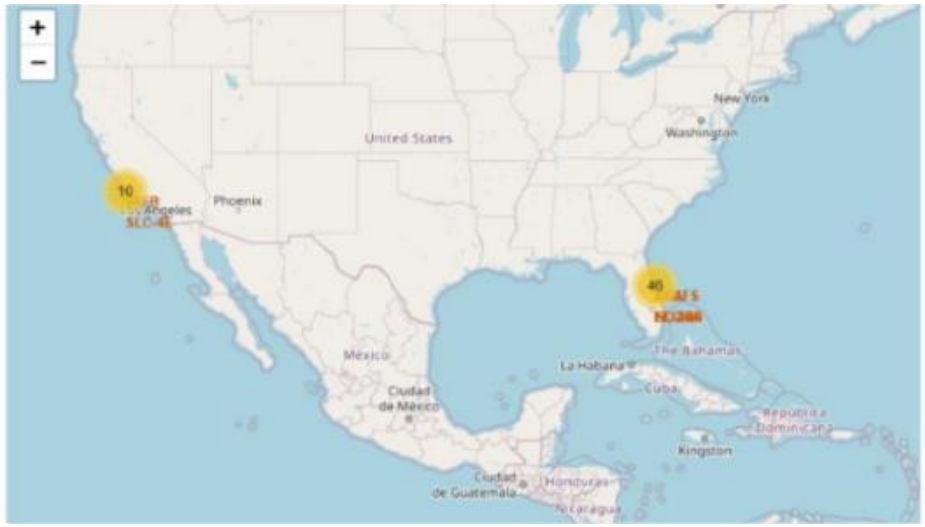
Average payload of F9 v1.1 booster is 2928 kg

1st successful landing – 2015; 5 yrs after 1st launch

Many had successful drone ship landings with payloads above average

Nearly all missions were successful

In 2015 there were 2 failed booster type landings (F9 v1.1 B1012 and F9 v1.1 B1015

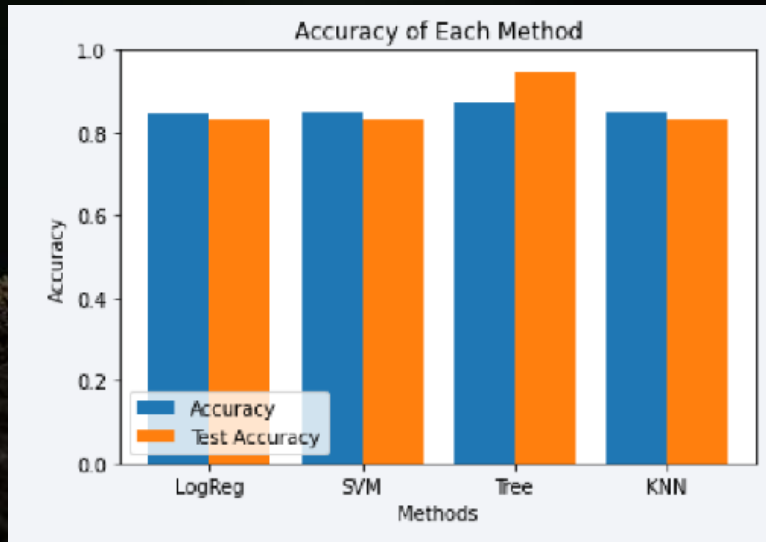


RESULTS

- Interactive analysis identified launch sites and showed most occurred near the east coast

RESULTS

- Predictive analysis showed the best model is the Decision Tree when it comes to predicting successful landings
- Provides 87%+ accuracy




INSIGHTS
DRAWN
FROM EDA

FLIGHT NUMBER VS. LAUNCH SITE

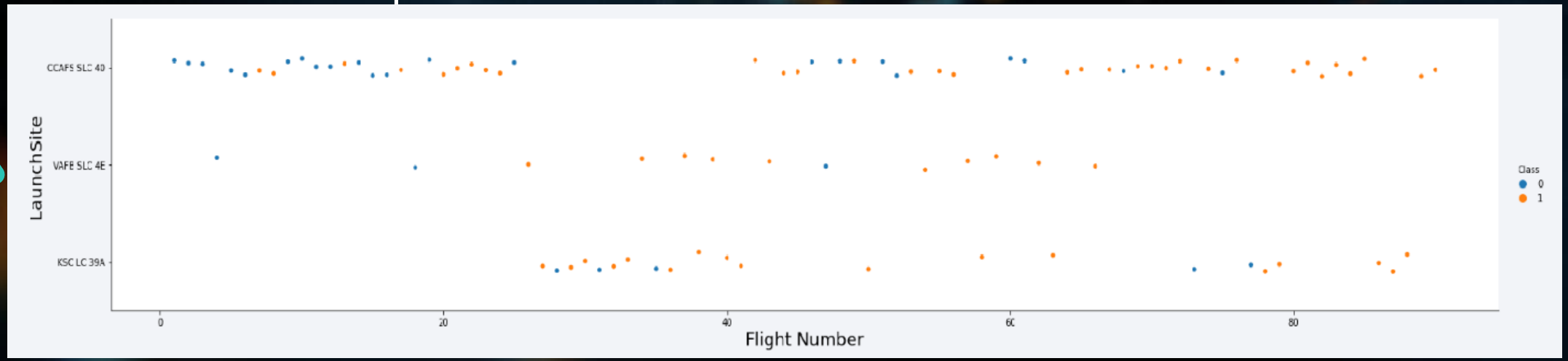
This plot shows :

-Best launch site is CCAFS SLC 40

-2nd best is VAFB SLC 4E

-3rd best is KSC LC 39A

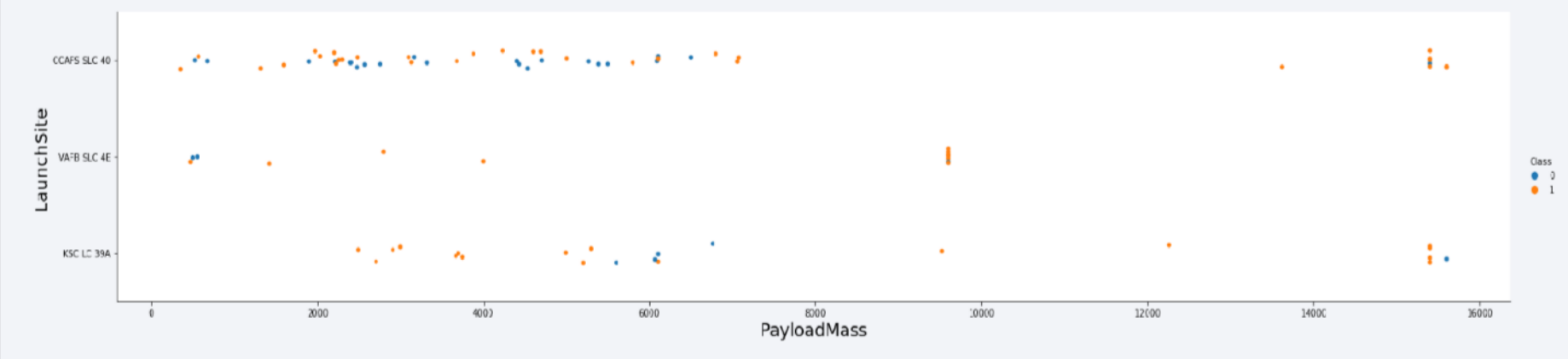
-Success rate improved over time



PAYLOAD VS LAUNCH SITE

Payloads of 9000+ kg have increased success rate

Payloads of 12000+kg only see success on CCAFS SLC 40 and KSC LC 39A launch areas



SUCCESS RATE VS ORBIT TYPE

The following orbits had the best success rates:

-ES-L1

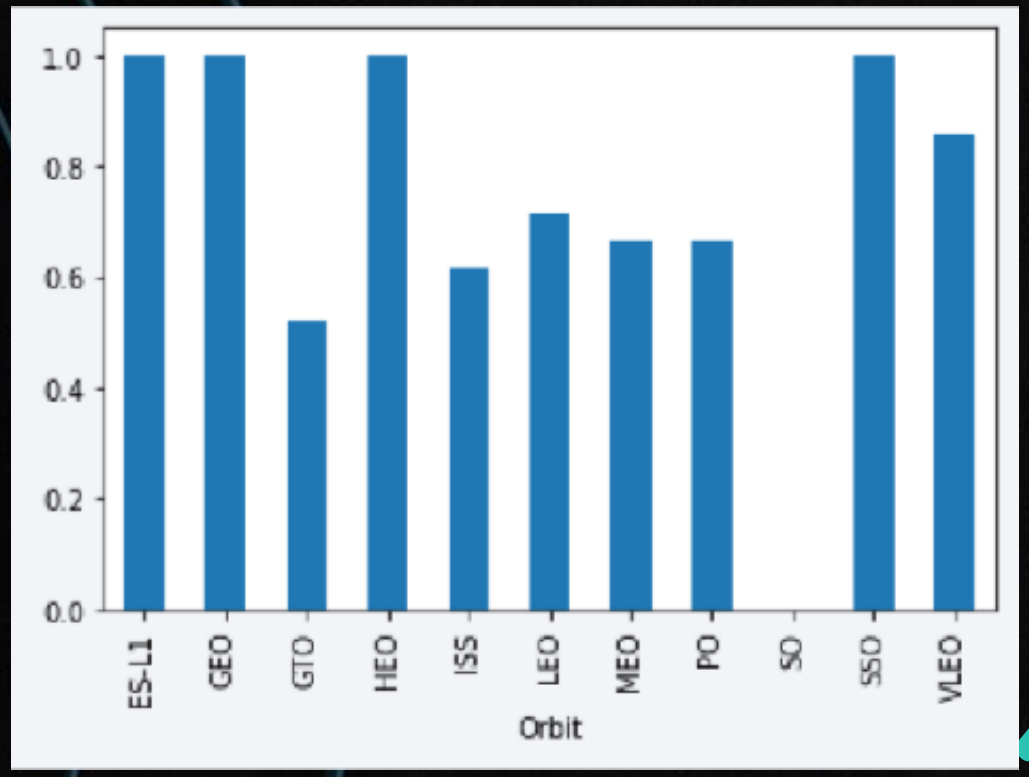
-GEO

-HEO

-SSO

-VLEO

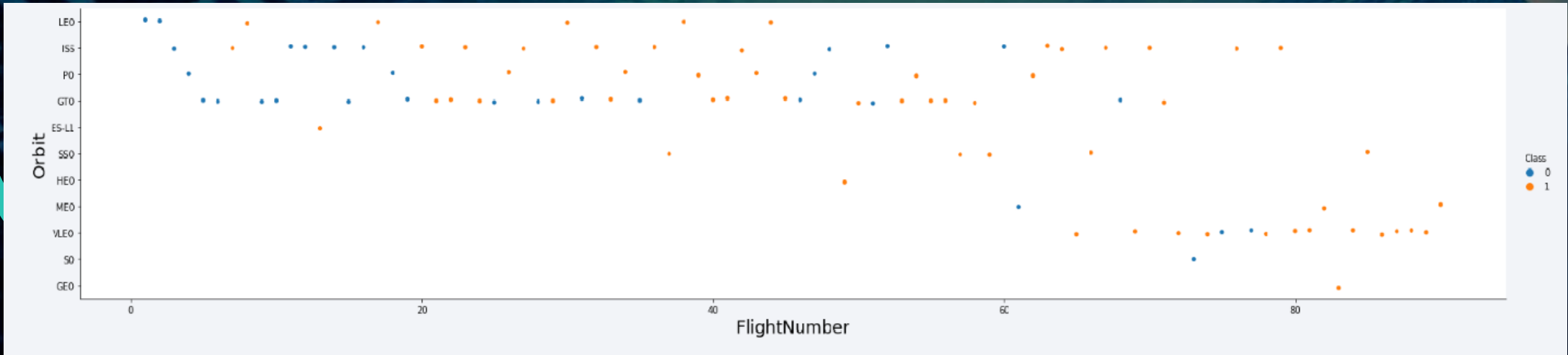
-LFO



FLIGHT NUMBER VS ORBIT TYPE

Success rates improved for all orbits over time

VLEO orbit is seeing increased success

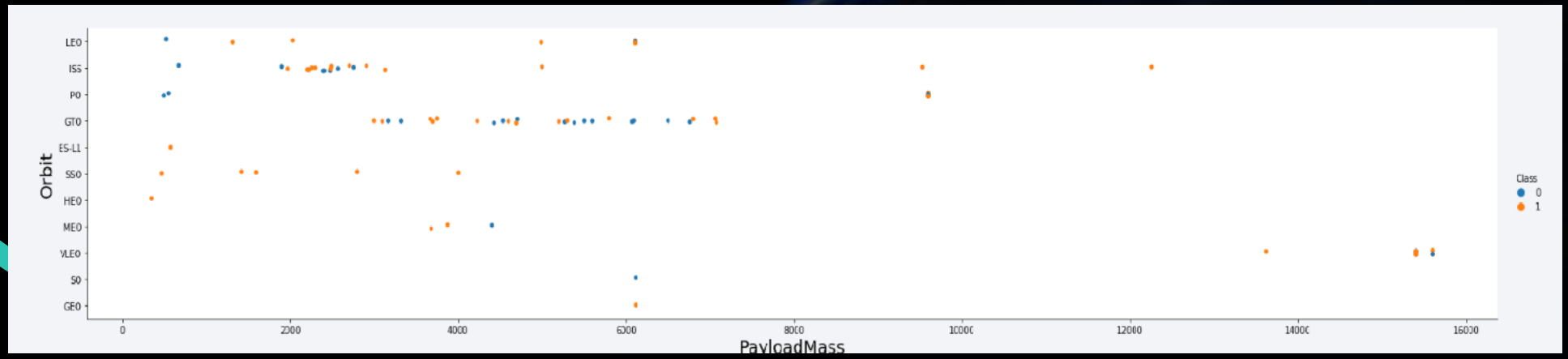


PAYLOAD VS ORBIT TYPE

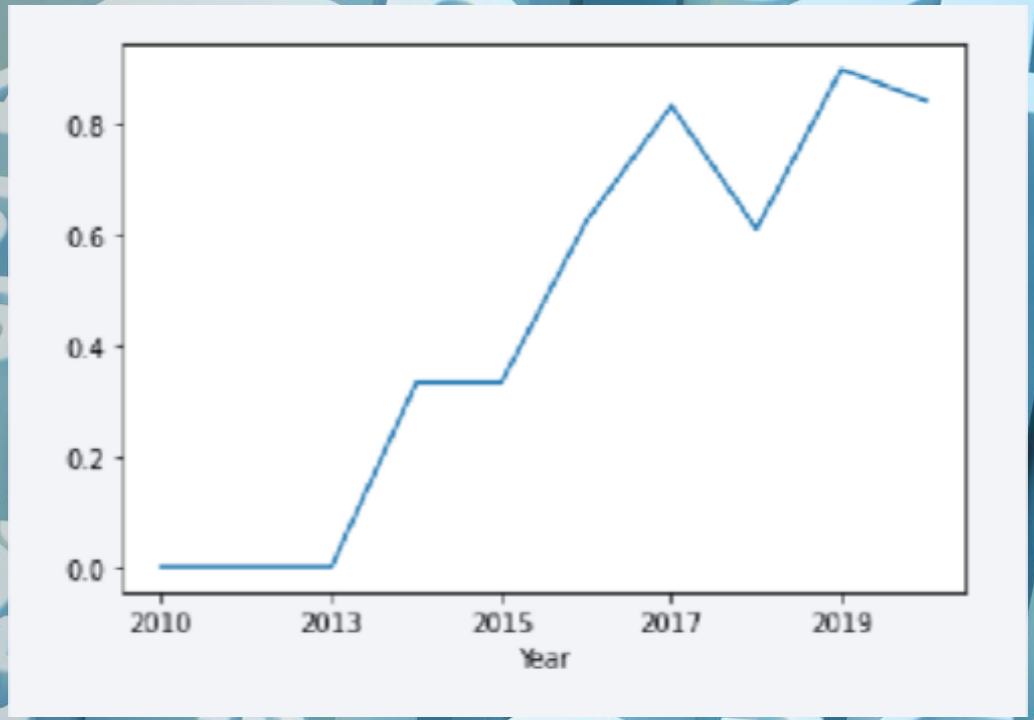
No relationship between payload and success rate in GTO orbit

ISS orbit has positive success rate and widest payload range

Few launches to SO and GEO orbits



LAUNCH SUCCESS YEARLY TREND



2010 – 2013 was and adjustment period until tech improved

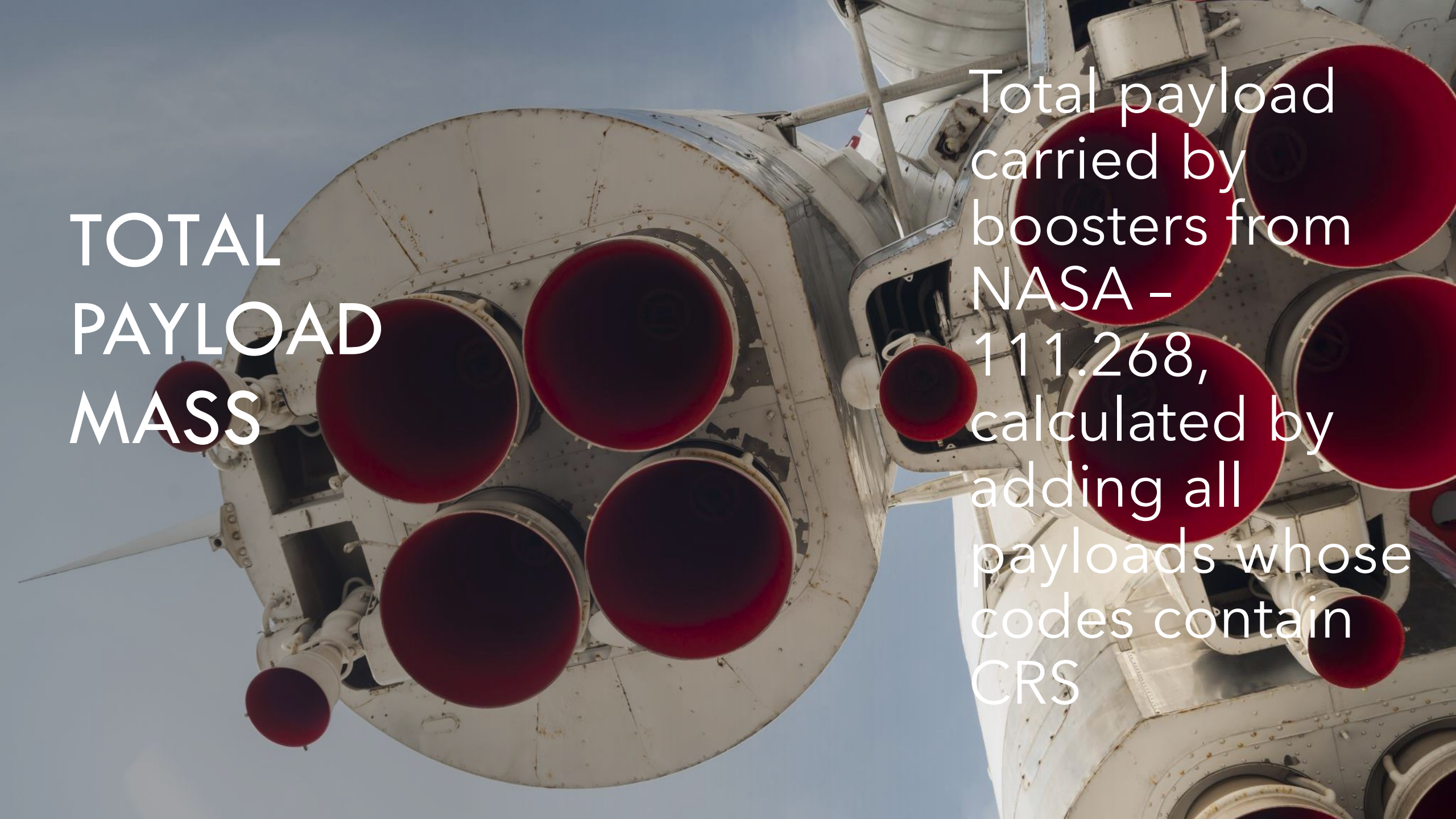
2013 – 2020 saw increased success

LAUNCH SITES

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

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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 attemp



TOTAL PAYLOAD MASS

Total payload carried by boosters from NASA – 111.268, calculated by adding all payloads whose codes contain CRS

AVERAGE PAYLOAD MASS BY F9 V1.1.

Average payload mass carried by booster version F9 v1.1 = 2.928, was found by filtering booster version then calculating the avg

FIRST SUCCESSFUL GROUND LANDING DATE

1st successful landing on ground pad was 22 Dec 2015

- By filtering data of successful landing outcomes on ground pad then finding minimum value date provided this data

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD 4000-6000

Boosters with a successful landing rate on drone ships had a payload mass between 4000-6000

Using these parameters resulted in finding these 4

F0 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

SUCCESS – 99
SUCCESS BUT W/ UNCLEAR PAYLOAD STATUS – 1
FAILURE IN FLIGHT –1

By grouping missions and counting records the above data was produced

BOOSTER CARRIED MAXIMUM PAYLOAD

Boosters which have carried the maximum payload mass registered in the dataset are listed below

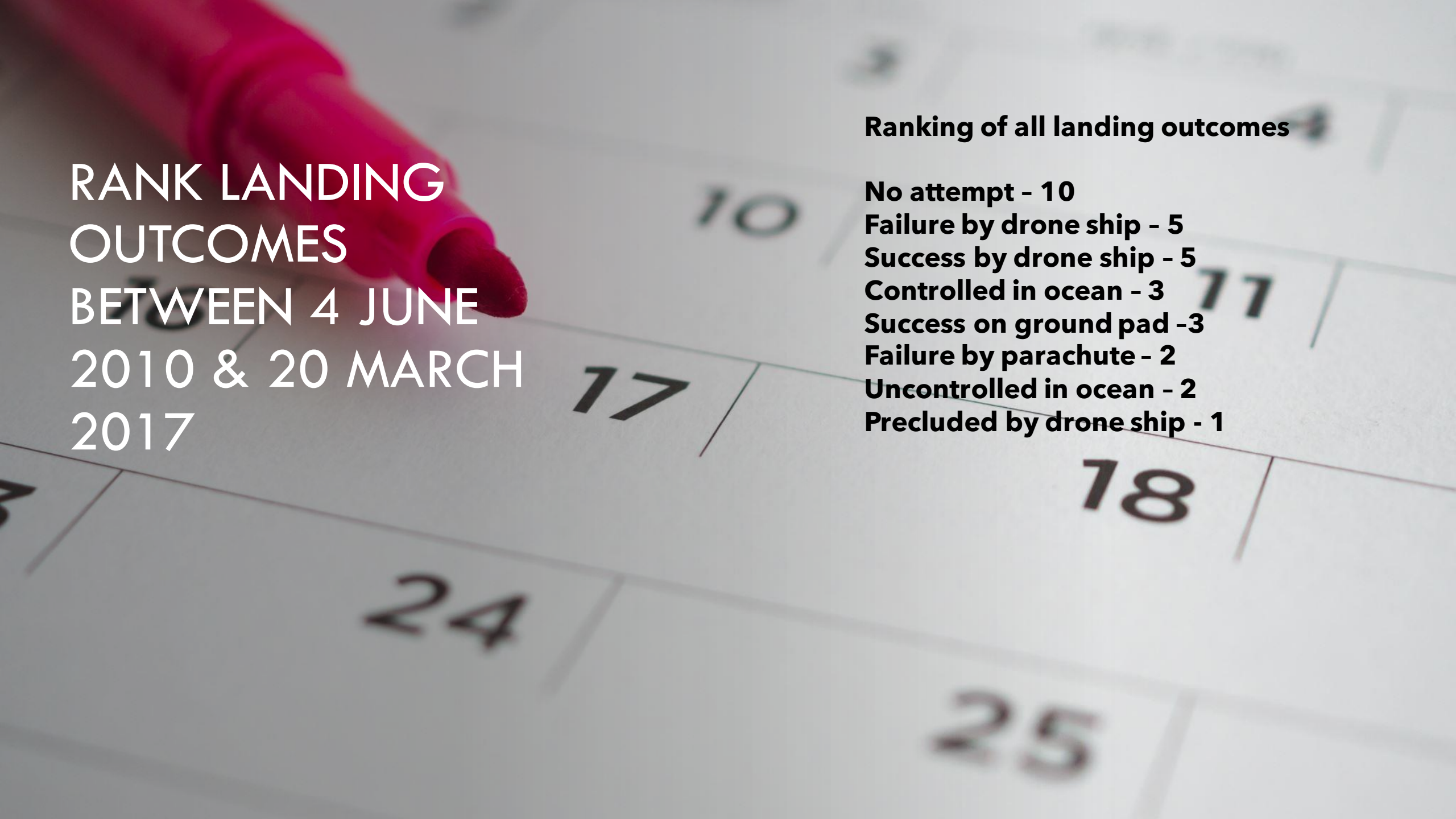
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 LAUNCH RECORDS

Failed landing outcomes from drone ships, their booster versions and launch sites:

F9 V1.1 B1012 – CCAFS LC-40

F9 V1.1 B1015 – CCAFS LC-40



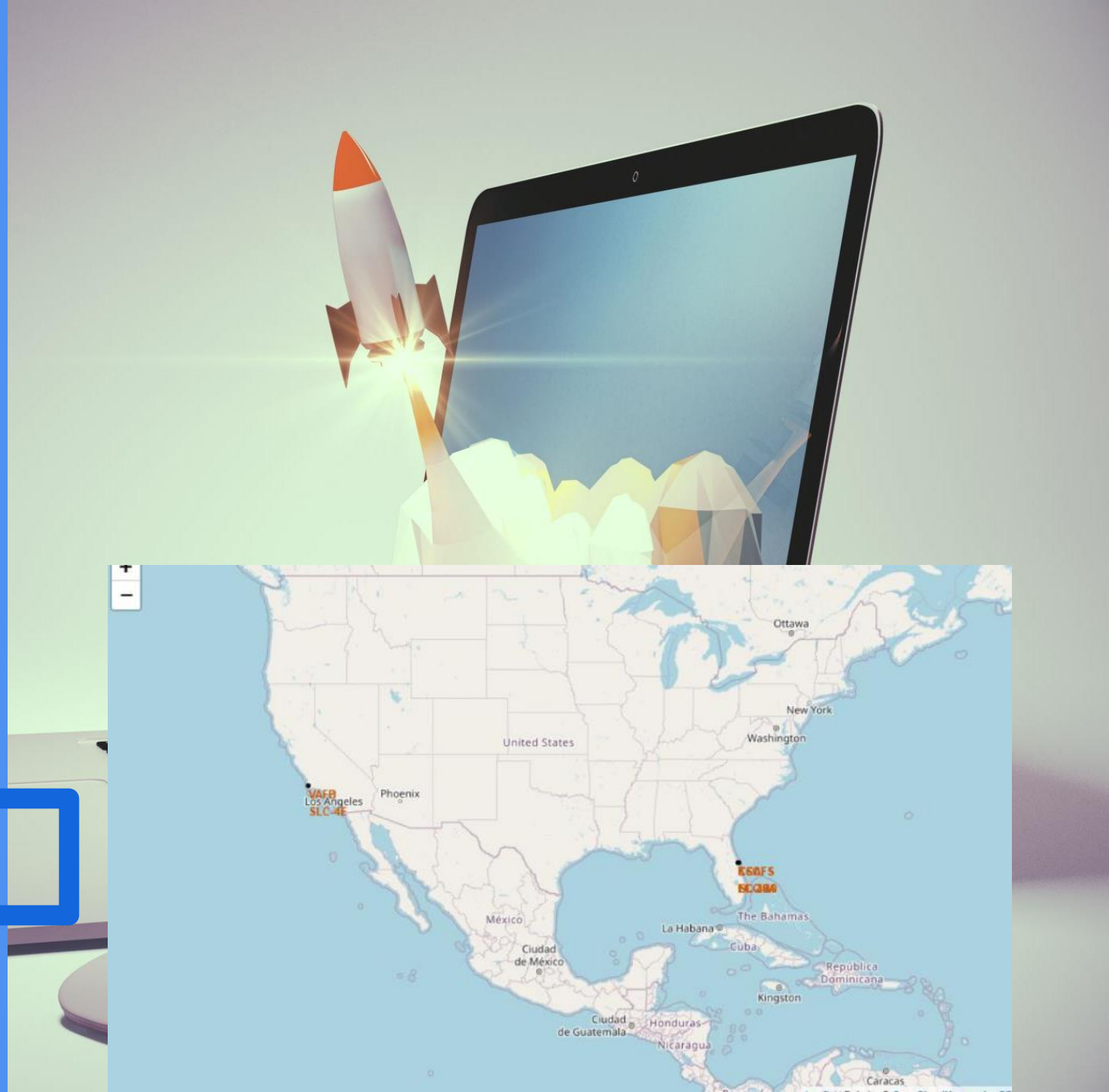
RANK LANDING OUTCOMES BETWEEN 4 JUNE 2010 & 20 MARCH 2017

Ranking of all landing outcomes

No attempt – 10
Failure by drone ship – 5
Success by drone ship – 5
Controlled in ocean – 3
Success on ground pad –3
Failure by parachute – 2
Uncontrolled in ocean – 2
Precluded by drone ship - 1

LAUNCH SITES PROXIMITIES ANALYSIS

ALL LAUNCH SITES

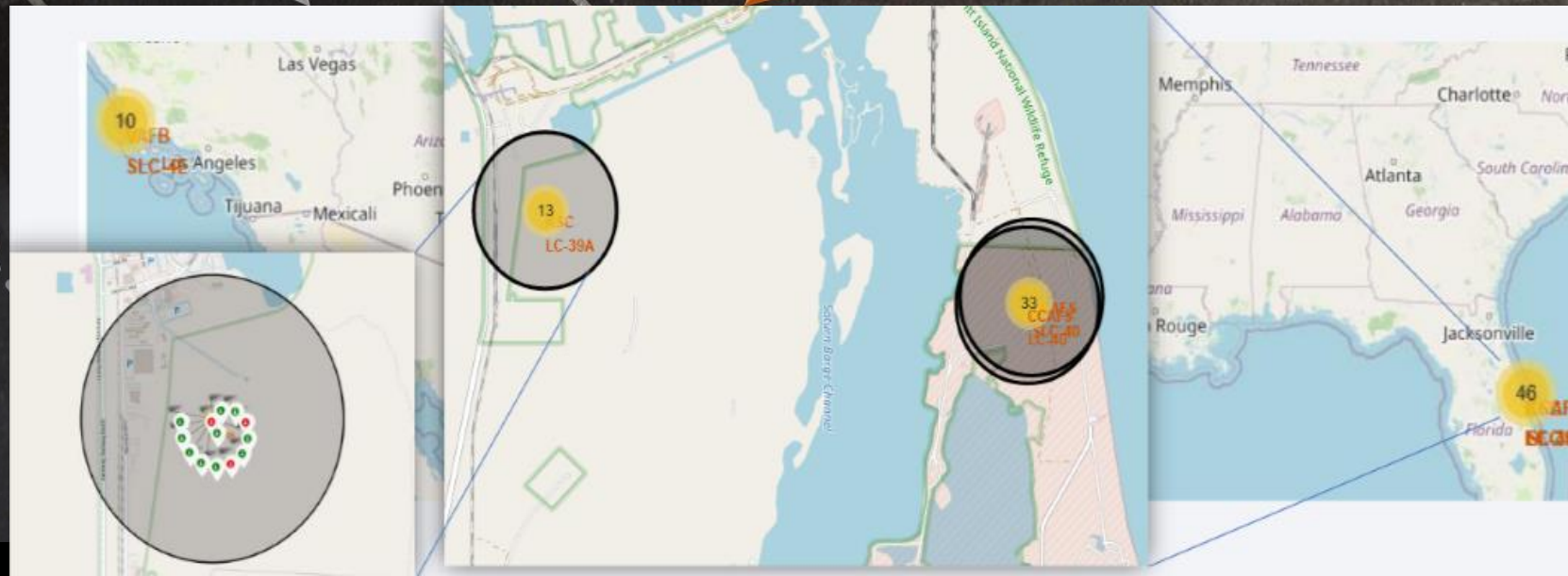


Launch sites are near sea but too far from roads and rails

LAUNCH OUTCOMES BY SITE

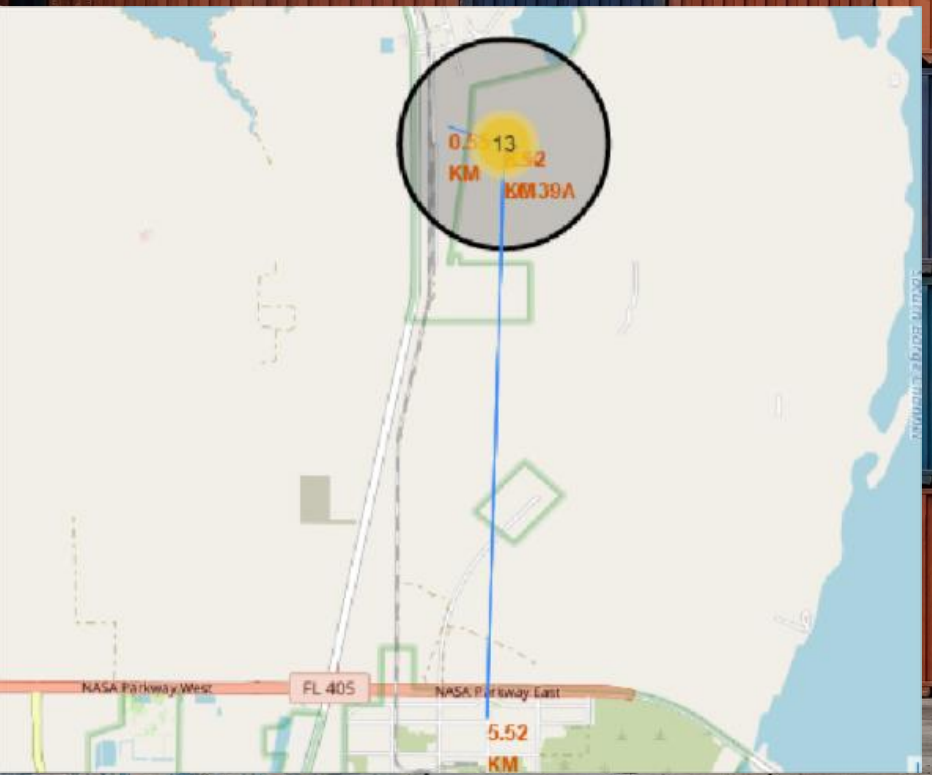
EXAMPLE OF KSC LC-39A LAUNCH SITE OUTCOMES

GREEN MARKERS INDICATE SUCCESS AND RED IS FAILURE



LOGISTICS AND SAFETY

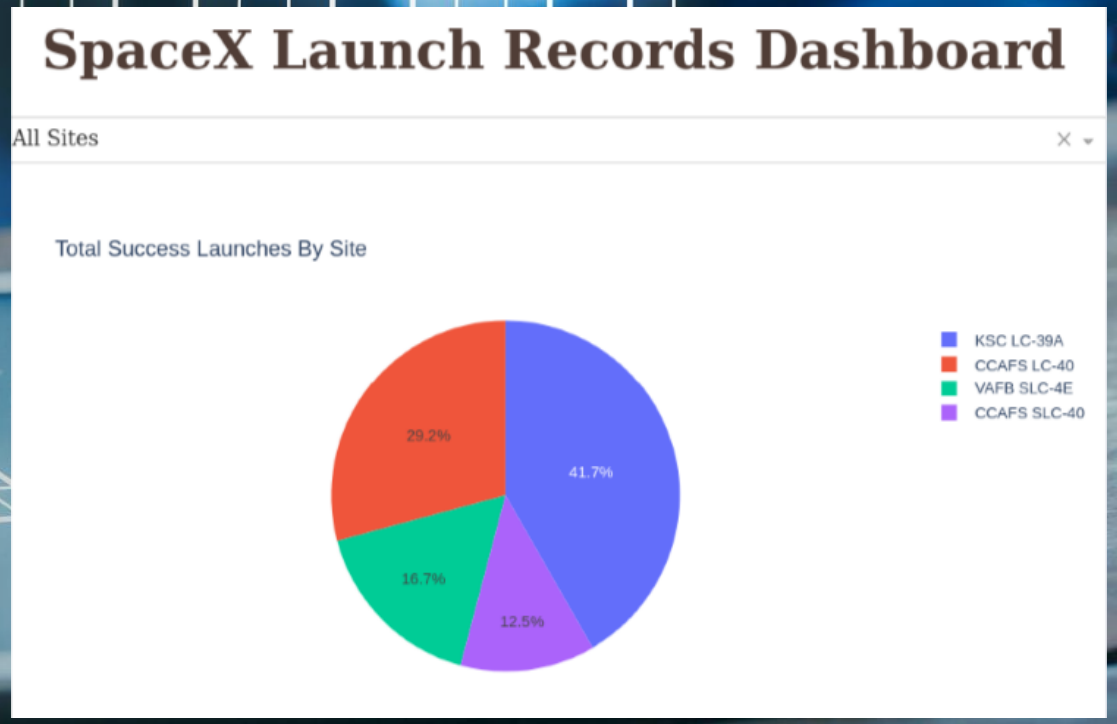
Launch site KSC LC-39A has good logistics being near rail, road and relatively far from inhabited areas



BUILD A DASHBOARD WITH PLOTLY DASH

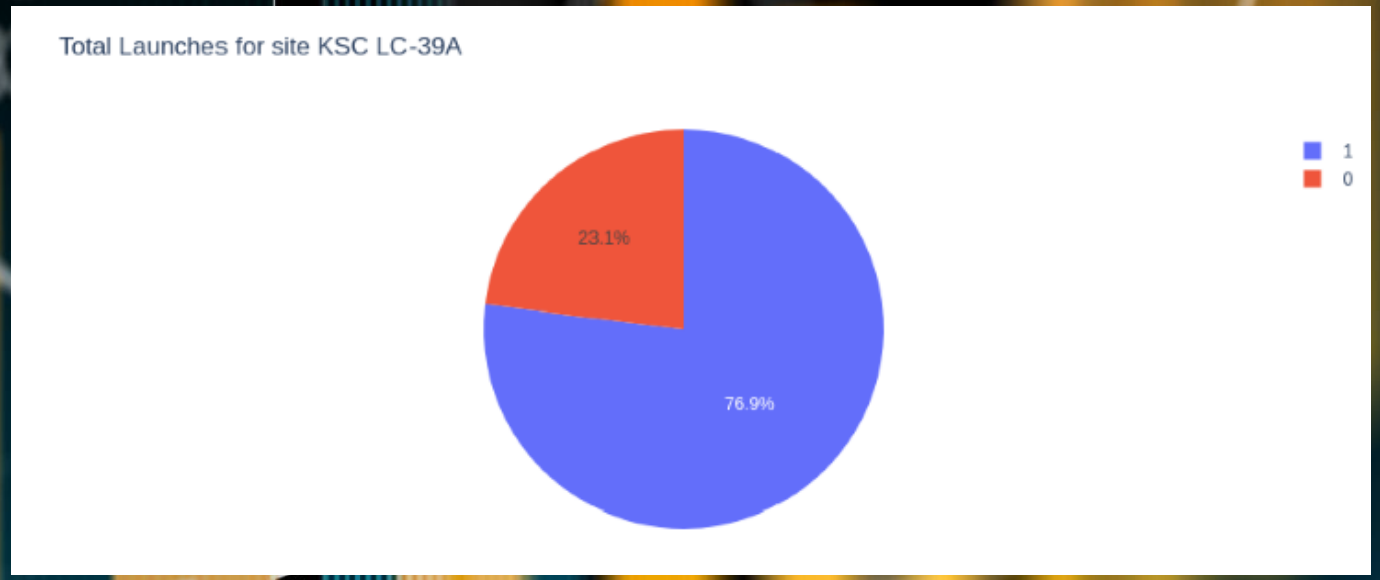
SUCCESSFUL LAUNCHES BY SITE

AN IMPORTANT FACTOR IN SUCCESS SEEMS TO BE WHERE THE LAUNCH TAKES PLACE



LAUNCH SUCCESS RATIO FOR KSC LC-39A

76.9% success rate at this site

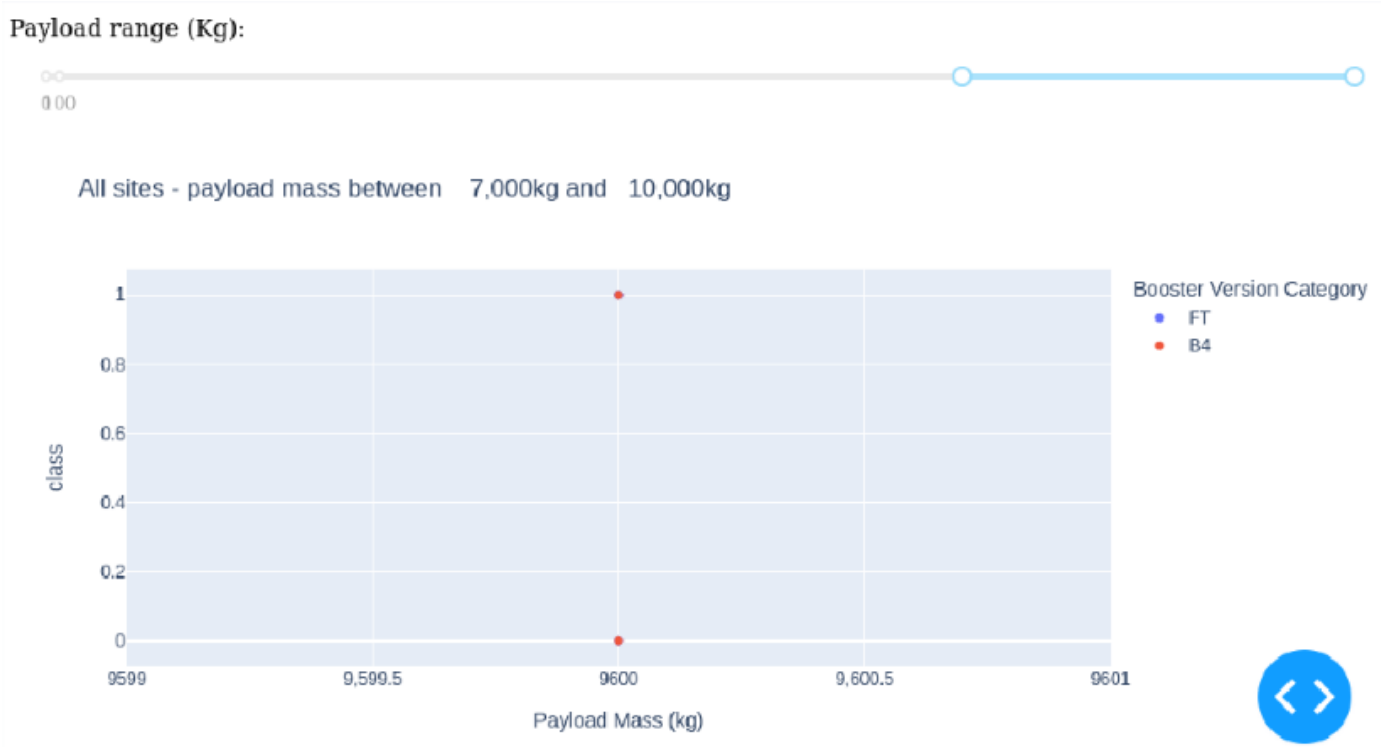


PAYLOAD VS LAUNCH OUTCOME

PAYLOADS UNDER 6000KG AND FT BOOSTERS ARE THE BEST COMBO

PAYLOAD VS LAUNCH OUTCOME

Not enough data exists to estimate launches over 7000kg

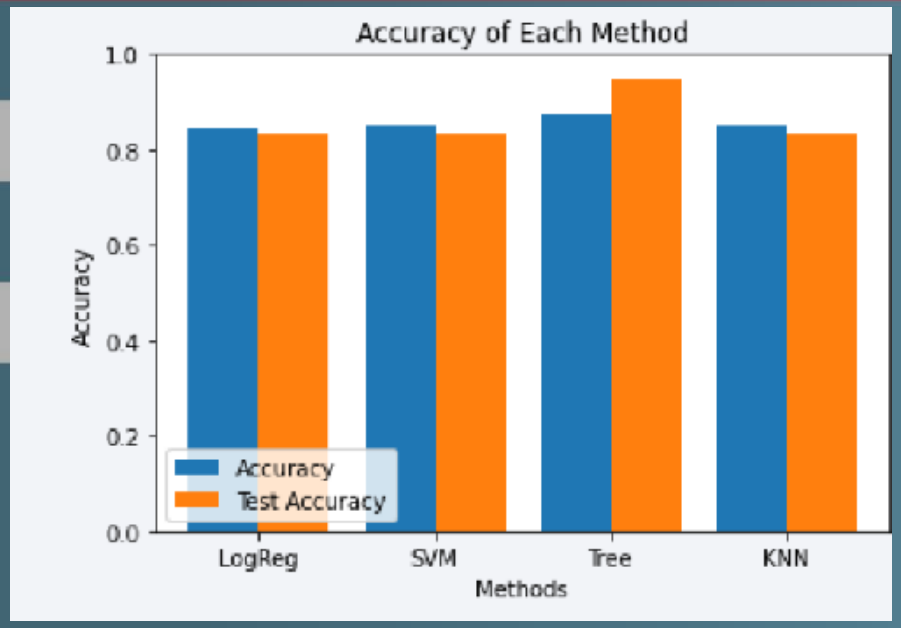


PREDICTIVE ANALYSIS (CLASSIFICATION)

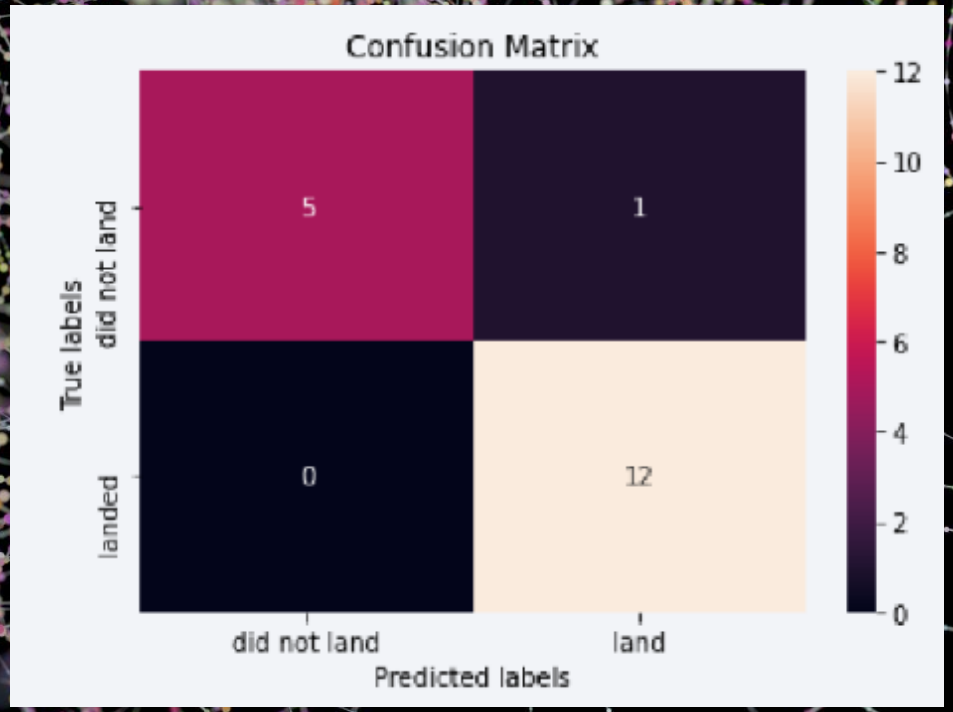
CLASSIFICATION ACCURACY

4 classification models were tested and their accuracies plotted

The model with the highest classification accuracy is the Decision Tree Classifier with an 87%+ accuracy rate



CONFUSION MATRIX OF DECISION TREE CLASSIFIER



This proves its accuracy By showing the big Numbers of true positive and true negative compared to the false ones

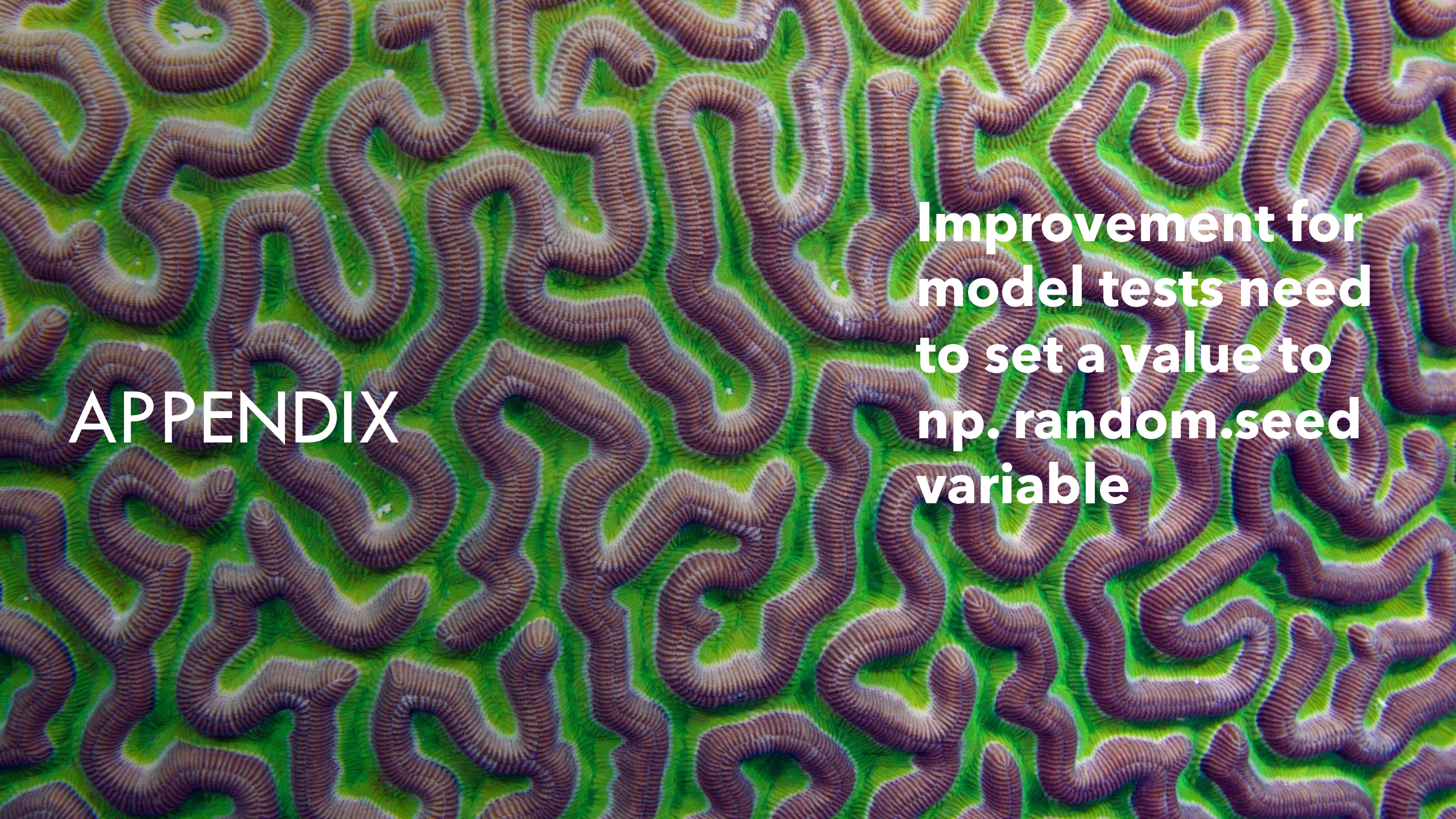
Different data sources were analyzed and refining conclusions

Best launch site is KSC LC39-A

Launches 7000+kig are less risky

Most mission outcomes are successful which is proved over time according to the evolution of processes and rockets

Decision Tree Classifier can predict successful landings and increase profits


APPENDIX
Improvement for model tests need to set a value to np. random.seed variable