DATA SCIENCE CAPSTONE PROJECT; SPACE'X

CLARIOND TIOGUEU
22.01.2023

OUTLINE

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

EXECUTIVE SUMMARY

Summary of Methodologies

- Data Collection via API, SQL and Web Scraping
- Data Wrangling and Analysis
- Interactive Maps with Folium
- Predictive Analysis for each Classification model
- Summary of all results
- Data Analysis along with Interactive Visualizations
- Best model for Predictive Analysis

INTRODUCTION

- Project background
- Predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on ist website with a cost of 62 million dollars while providers cost upward of 165 Million dollars each. SpaceX can reuse the first stage. Therefore determining the successfull landing can reduce cost.
- Problems to be solved
- Factors for sucessfull landy of the rocket
- Influence of features on outcome

METHODOLOGY

- Data Collection is the process of gathering and measuring information on targeted variables ,. This enables to evaluates outcomes and answer the relevant question.

METHODOLOGY

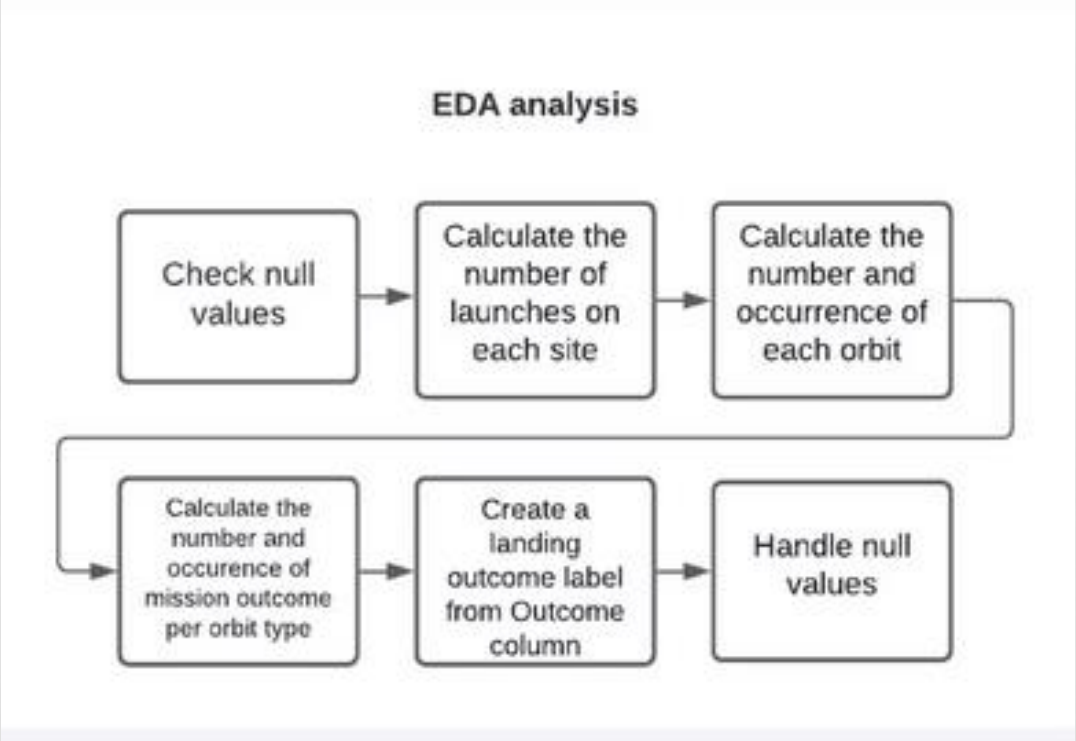
Data collection methodology:

- Collect rocket launch data from SpaceX API
- Perform data wrangling
- Modify and add columns, useful for training 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
- Break dataset into training and test data
- Develop various models (KNN, Decision Tree, etc…) using train data
- Assess accuracy scores of models and their best parameters using test data

DATA COLLECTION

- Data was collected from several maners
- Via SpaceX API
- Via Web Scraping

DATA WRANGLING



EDA

- EDA with Data Visualization
- EDA with SQL

EDA WITH DATA VISUALIZATION

- Here, a number of plots showing relationships between different variables:
  - Flight Number vs. Payload (Cat plot)
  - Flight Number vs. Launch Site (Cat plot)
  - Launch Site vs. Payload (Scatter plot)
  - Success Rate vs. Orbit type (Bar plot)
  - Orbit type vs. Flight Number (Scatter plot)
  - Orbit type vs. Payload (Scatter plot)
  - Success rate vs. Time in years (Line plot)
- An explanation of these various plots are shown in Section 2
- We also perform some one-hot encoding in this section

EDA WITH SQL

- Using SQL allows us to make complicated queries without much difficulty.
- Here, we make some basic queries on our data to get a better sense for

the relationships between variables, particularly the following:

- Launch Site
- Payload Mass (kg)
- Mission Outcome
- Booster Version
- Date

BUILD AN INTERACTIVE MAP WITH FOLIUM

- With the Python package Folium, we created an interactive map:
  - View where each Falcon 9 launch site is located, represented by a circle
  - Learn how many launches occurred at each location, represented by markers. Green markers represent a successful recovery while red markers represent unsuccessful one

BUILD A DASHBOARD WITH PLOTLY DASH

- Using Plotly Dash, we made a dashboard that shows:
  - A pie chart showing the proportion of successful recoveries to unsuccessful ones for each site
  - A Recovery Outcome vs. Payload Mass scatter plot with a range (0-10000kg)
- The dashboard provides insight into the launch sites' and payload masses' relationships with the recovery outcomes.

BUILD AN INTERACTIVE MAP WITH FOLIUM

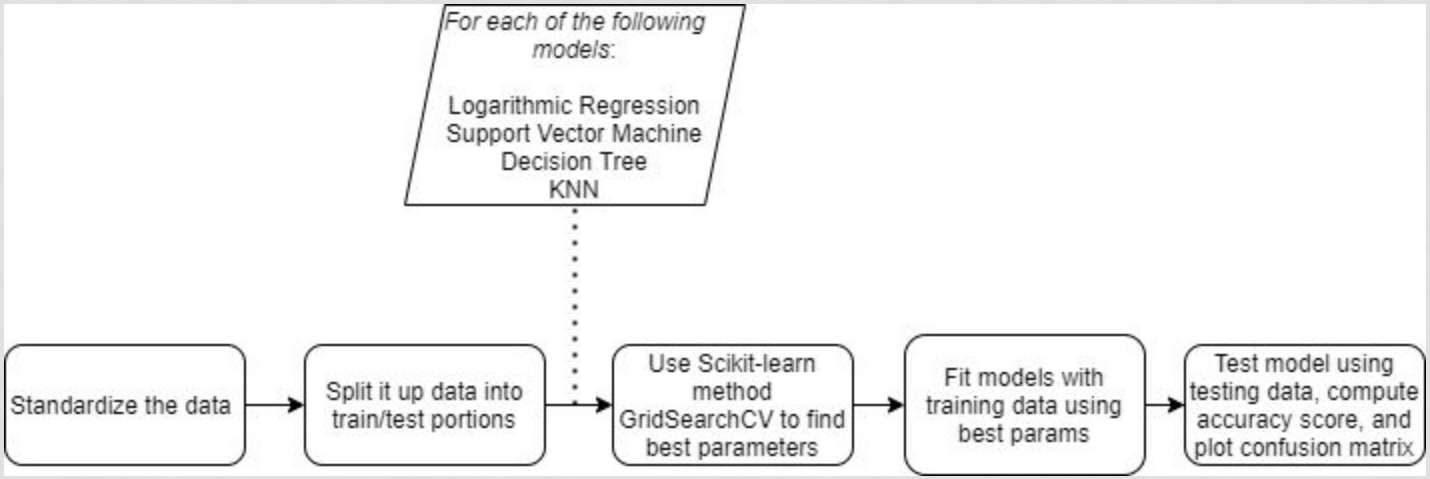
- Using the Python package Folium, we created an interactive map where one can:
  - View where each Falcon 9 launch site is located, represented by a circle
  - Learn how many launches occurred at each location, represented by
  - markers. Green markers represent a successful recovery while red markers
  - represent unsuccessful one
  - Determine distances to the closest coastline, city, railway, and highway,
  - each represented by a blue line.

DASHBOARD WITH PLOTY DASH

- Through Plotly Dash, we made a dashboard that shows:
  - A **pie chart** showing the proportion of successful recoveries to
  - unsuccessful ones for each site (which can be changed via a
  - dropdown menu)
  - A Recovery Outcome vs. Payload Mass **scatter plot** with a range
  - (0-10000kg) with bounds that can be changed by the user
- This dashboard provides insight into the launch sites' and payload masses'

relationships with the recovery outcomes.

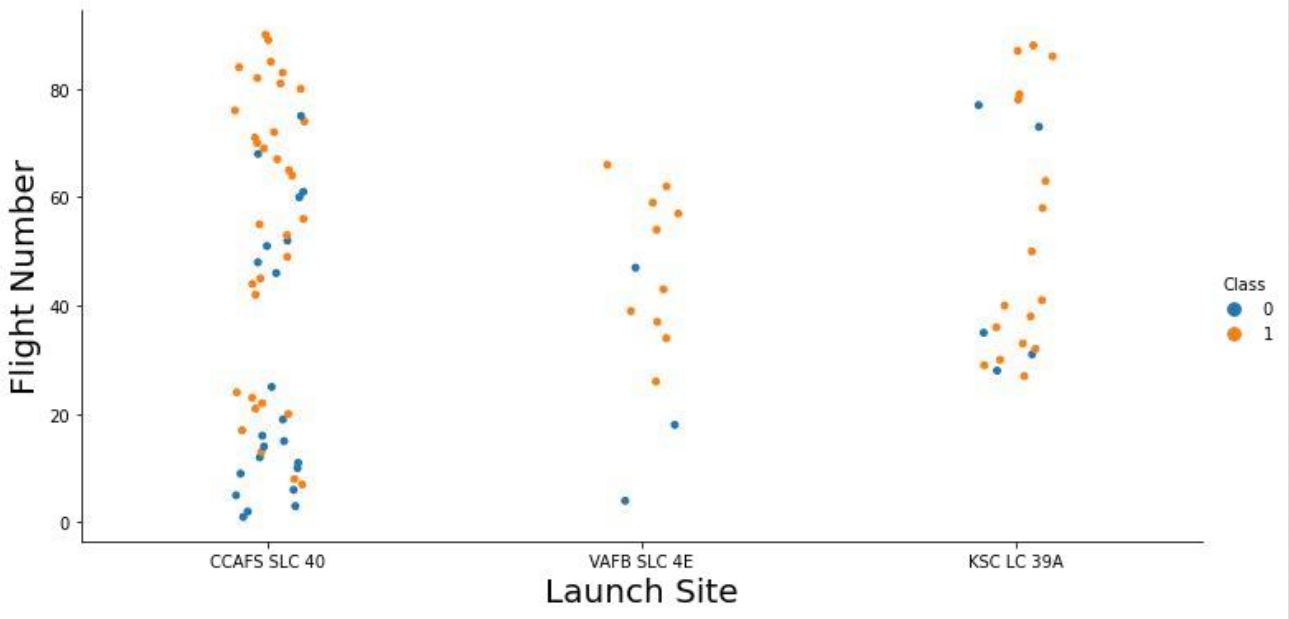
PREDICTIVE ANALYSIS (CLASSIFICATION)



INSIGHTS DRAWN FROM EDA

EDA WITH DATA VISUALIZATION

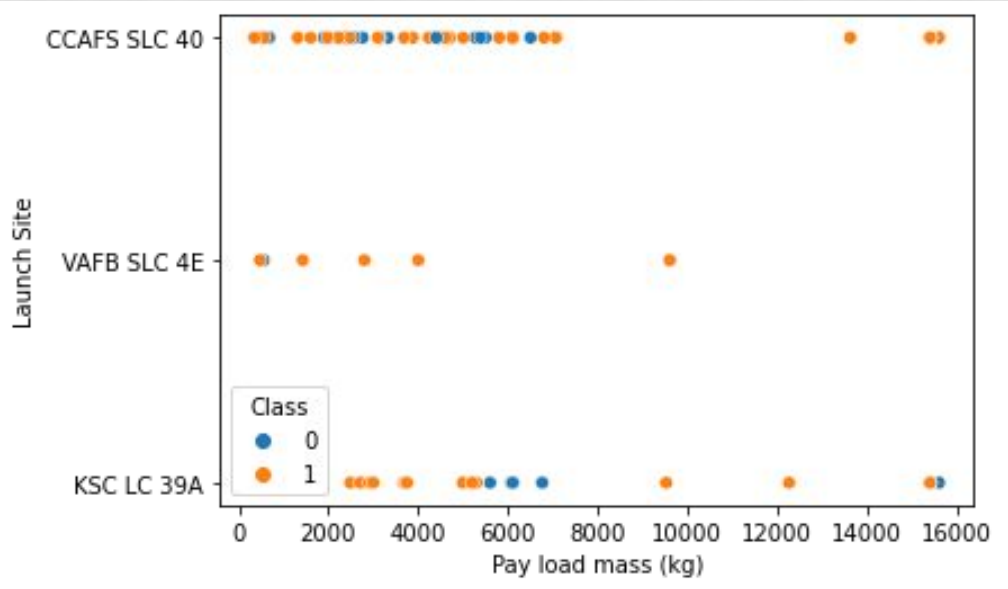
Flight Number vs Launch Site (Cat Plot)



- Rate of success has grown over time at each site, though KSC LC-39A seems to be the most consistent

EDA WITH DATA VISUALIZATION

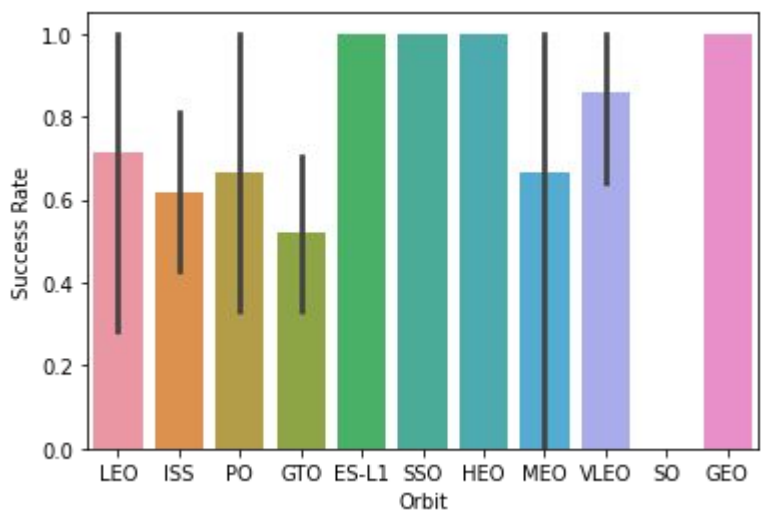
Launch Site vs Payload Mass (Scatter Plot)



- Smaller payloads (<6000 kgs) seem to correlate with higher success rate

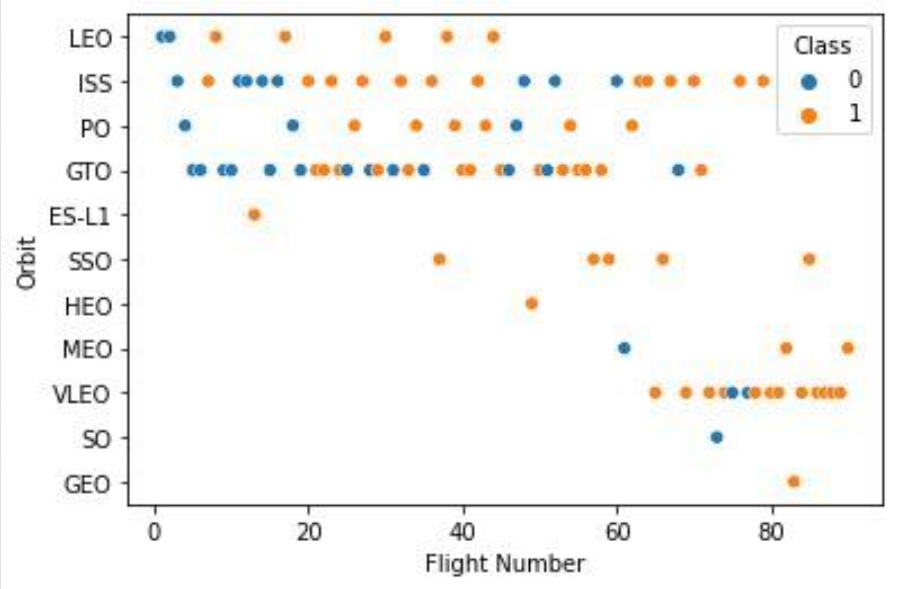
EDA WITH DATA VISUALIZATION

Success Rate vs. Orbit Type (Bar plot)



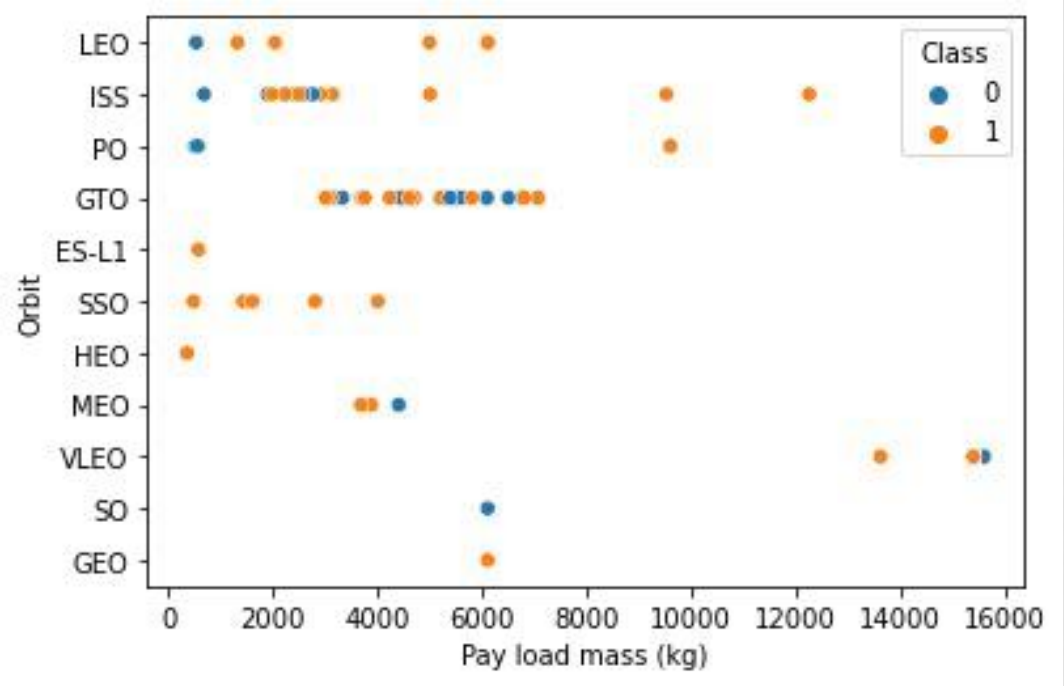
ES-L1, SSO, HEO, and GEO orbits are very reliable

FLIGHT NUMBER VS. ORBIT TYPE



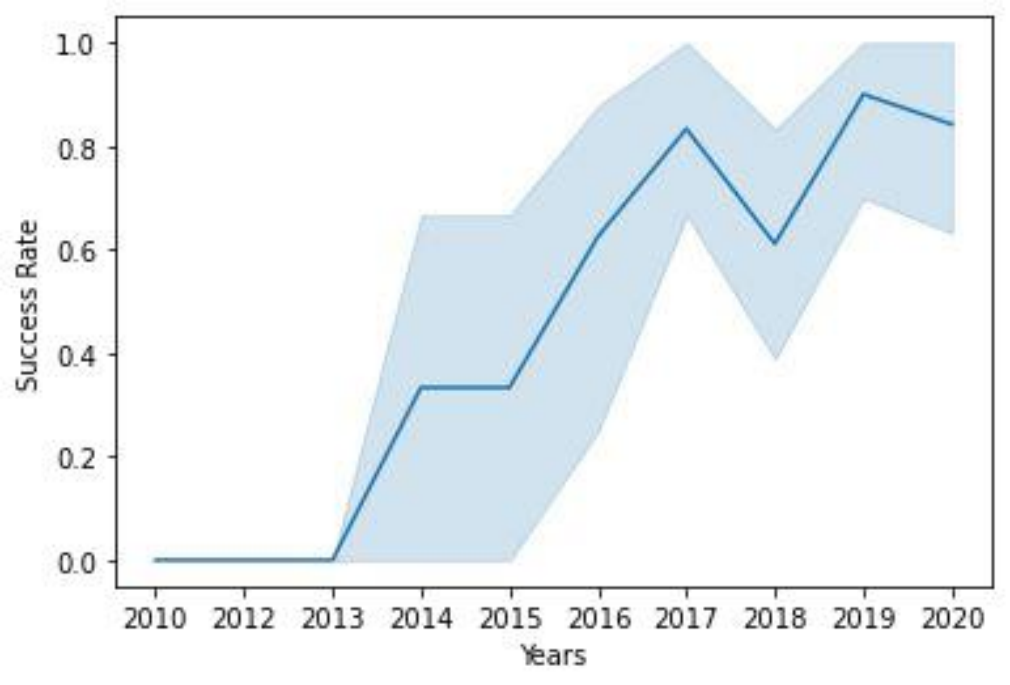
- However, ES-L1, HEO, and GEO success rate may be skewed due to each having only 1 launch
- LEO, SSO, and VLEO seem all have high success rates while having high sample sizes

PAYLOAD VS. ORBIT TYPE



- LEO and SSO orbits' success may be due to light payloads

LAUNCH SUCCESS YEARLY TREND



- Overall success rate increased from 2013 to 2017, where it has more or less stagnated

ALL LAUNCH SITE NAMES

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

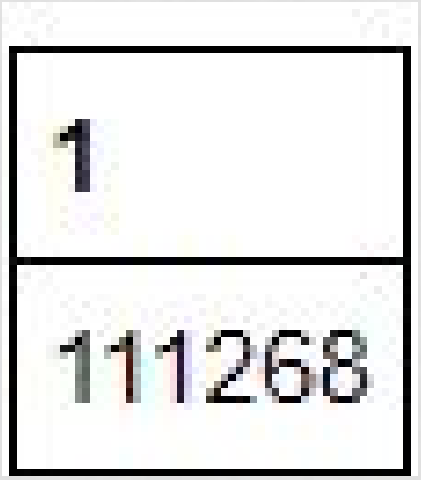
- we have 4 unique launch site locations

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 attempt

First 5 records for launch sites that begin with ‘CCA’

TOTAL PAYLOAD MASS



AVERAGE PAYLOAD MASS BY F9 V1.1

FIRST SUCCESSFUL GROUND LANDING DATE

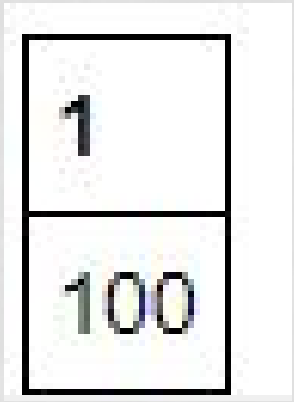
1
2010-06-04

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

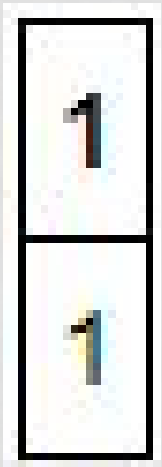
booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

- Success



- Failiure



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

2015 LAUNCH RECORDS

DATE	landing__outcome	booster_version	launch_site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

RANK LANDING OUTCOMES BETWEEN 2010-06-04 AND 2017-03-20

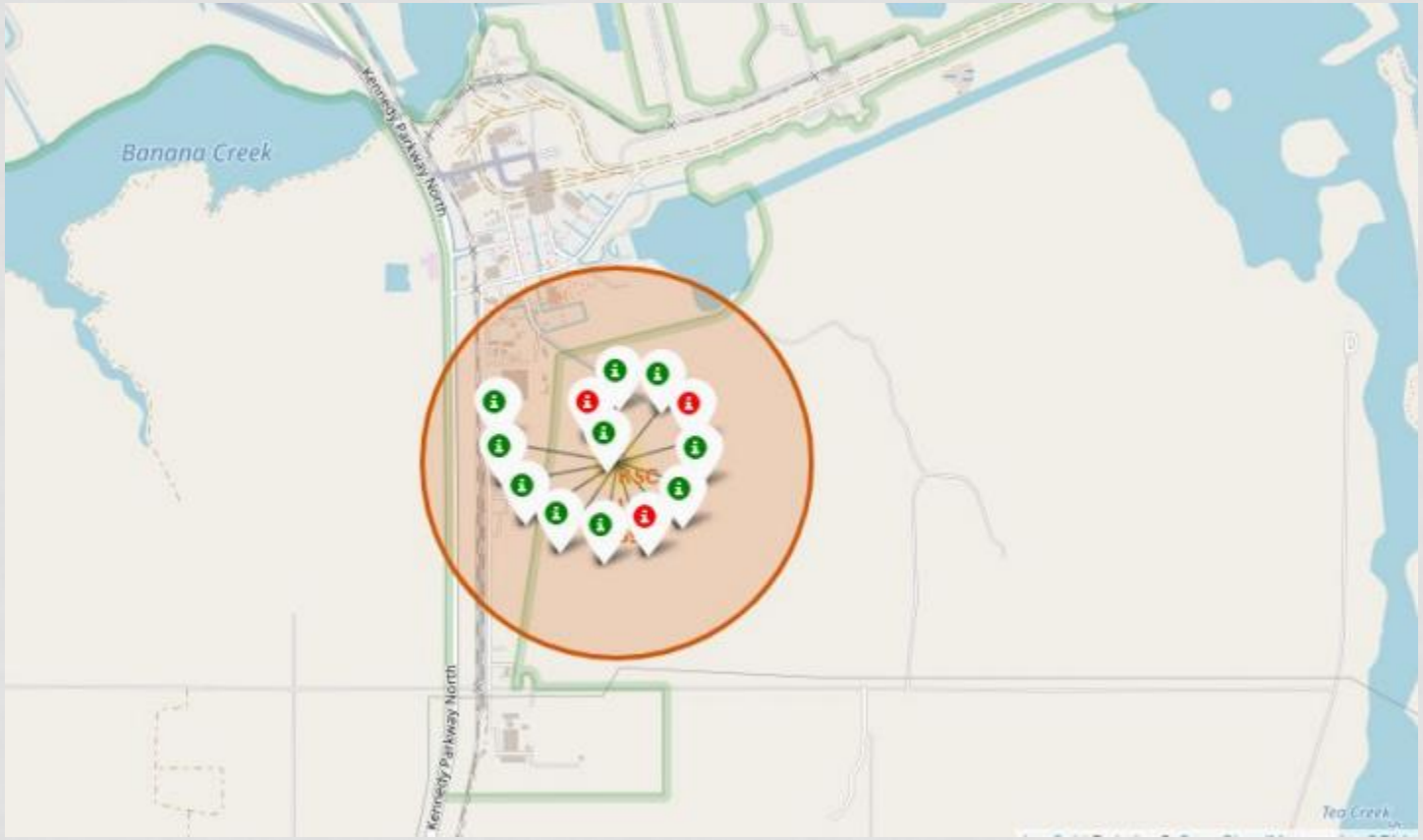
landing__outcome	counts
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

LAUNCH SITES PROXIMITIES ANALYSIS

MAP OF LAUNCH SITE LOCATIONS

VAFB SLC-4E RECOVERY OUTCOMES

KSC LC-39A RECOVERY OUTCOMES

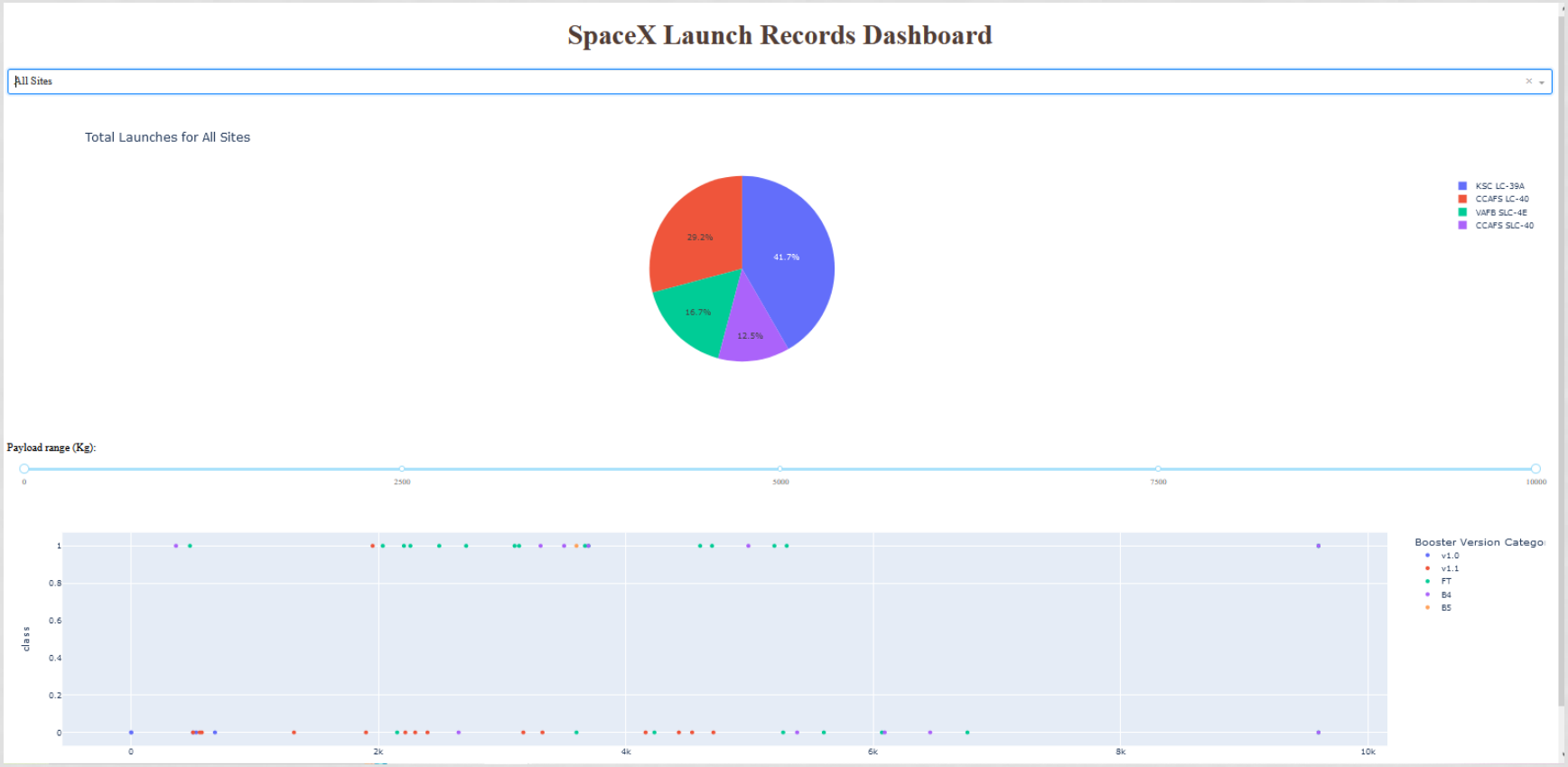


CCAFS LC-40 RECOVERY OUTCOMES

CCAFS SLC-40 RECOVERY OUTCOMES

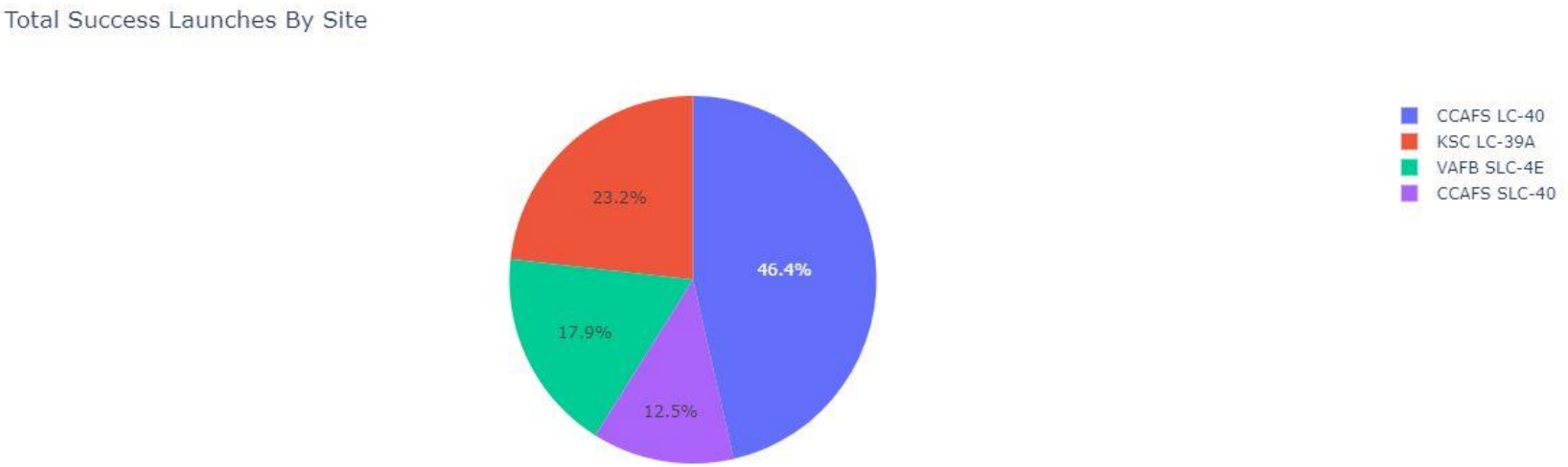
BUILD A DASHBOARD WITH PLOTLY DASH

SUCCESS LAUNCHES BY ALL SITES



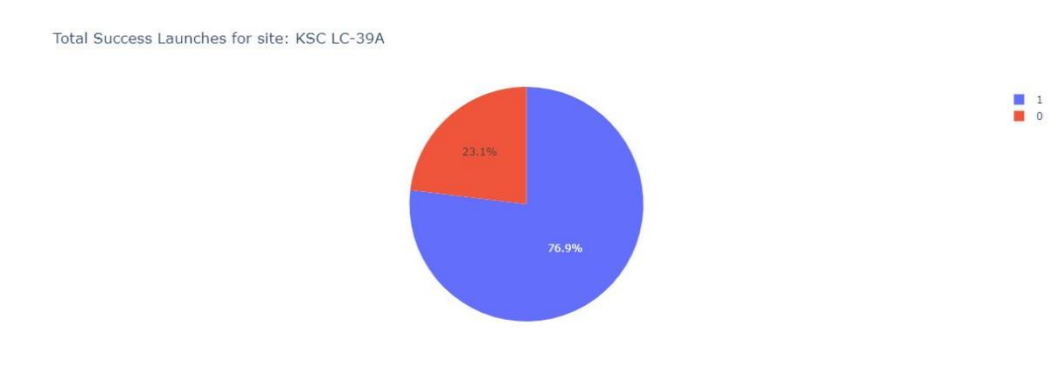
PAYLOAD VS. LAUNCH OUTCOME SCATTER FOR ALL SITES

PIE CHART OF SUCCESSFUL LAUNCHES BY SITE



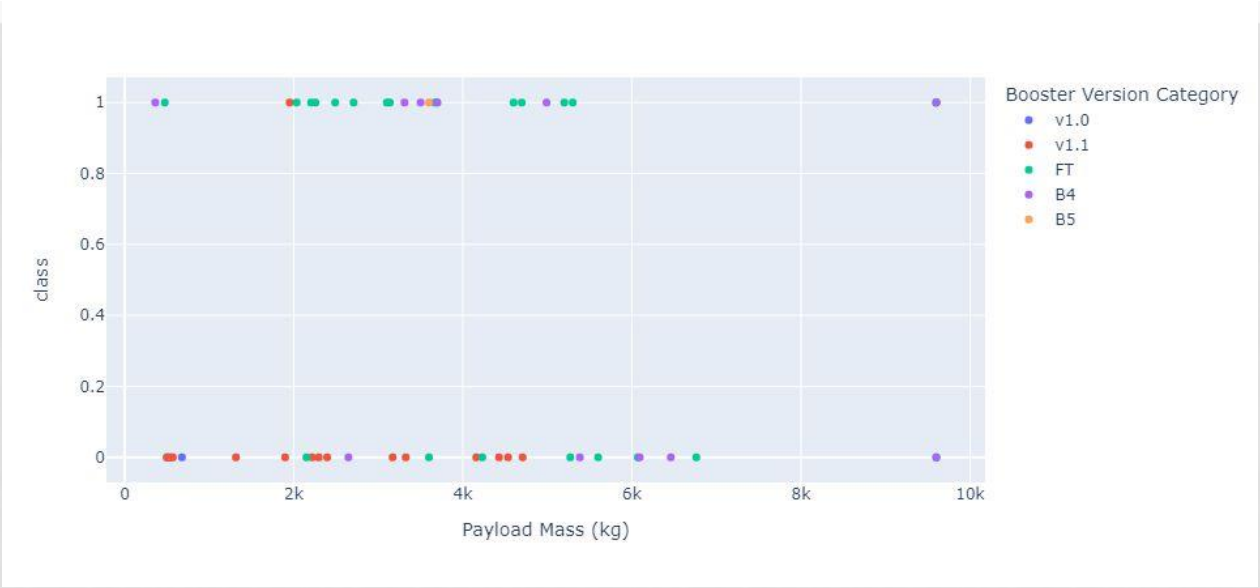
- CCAFS LC-40 is the location of over half of SpaceX's successful launches
- However, this does not tell the full story

MOST PROPORTIONALLY SUCCESSFUL LAUNCH SITE



- CCAFS LC-40 has the most launches
- KSC LC-39A, comparatively

RECOVERY OUTCOME VS. PAYLOAD MASS SCATTER PLOT

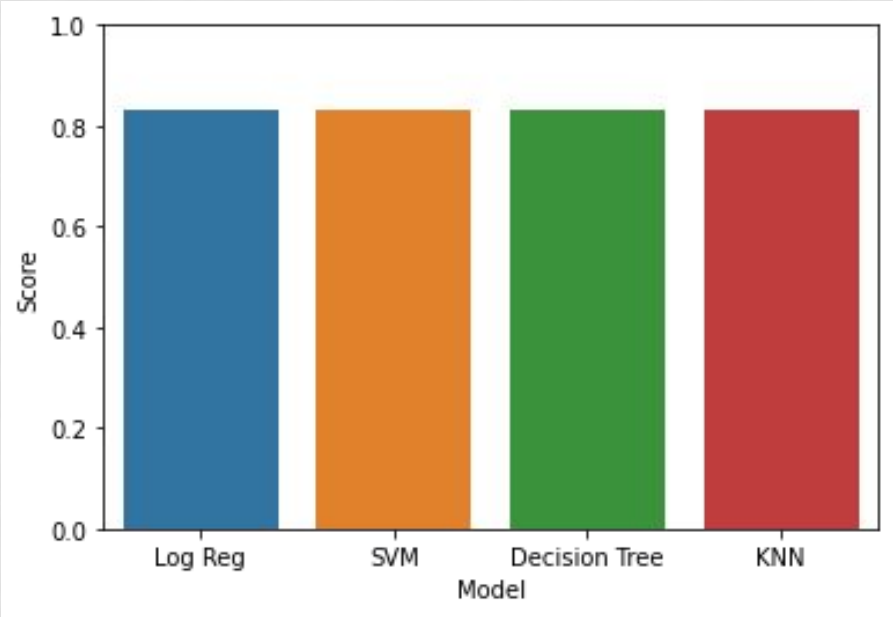


- The most successful payload range appears to be from 2000-4000kg
- The v1.1 booster seems to have failed the most while the FT

booster appears to be very successful

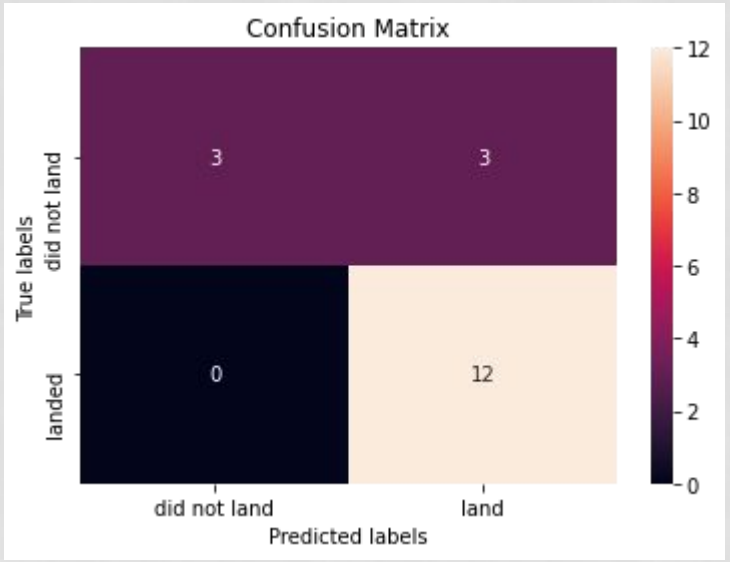
CLASSIFICATION ACCURACY

Here is a barplot for the accuracy score of each model. As can be seen, when testing them on our test data, they all exhibit identical scores (83.33%)



CONFUSION MATRIX

Our model was mostly accurate, however it wrongly predicted three landings as being successful where they were not



CONCLUSIONS

SpaceX's successful recoveries generally have the following properties:

- A launch date in the year 2017 or later
- Light payload (in the range 2000-4000kg)
- Launched from site KSC LC-39A
- Successfully recovered via drone ship

Our model predict an outcome of a given recovery with a degree of accuracy equal 83.33%

THANK YOU