Winning the space race – How SpaceY can outcompete SpaceX

Yannick Deepen

17.11.2022

IBM Developer

SKILLS NETWORK

Table of content

EXECUTIVE SUMMARY

- Space X is considered the market leader among the commercial rocket launch providers globally with many competitors. This report focuses on the factors leading to successful landing outcomes, allowing upcoming providers such as the newly founded SpaceY to conduct data-driven decision making.

- **Key findings include:**
  - Launch site locations are best situated in close proximity to railroads and highways, along the coastline and not to close to the next city
  - Through the use of mission data we were able to predict the success rate of a first stage booster landing with 83% accuracy

INTRODUCTION: BUISNESS PROBLEM

- SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. In this module, you will be provided with an overview of the problem and the tools you need to complete the course.
- **The goal of this report is to predict the likelihood of successful first stage booster rocket landings as a proxy for the cost of a launch**

METHODOLOGY

Section 1

METHODOLOGY

- Data collection
- Data wrangling
- Exploratory data analysis
- Data visualization
- Model deployment
- Presentation of results towards stakeholders

METHODOLOGY: DATA COLLECTION

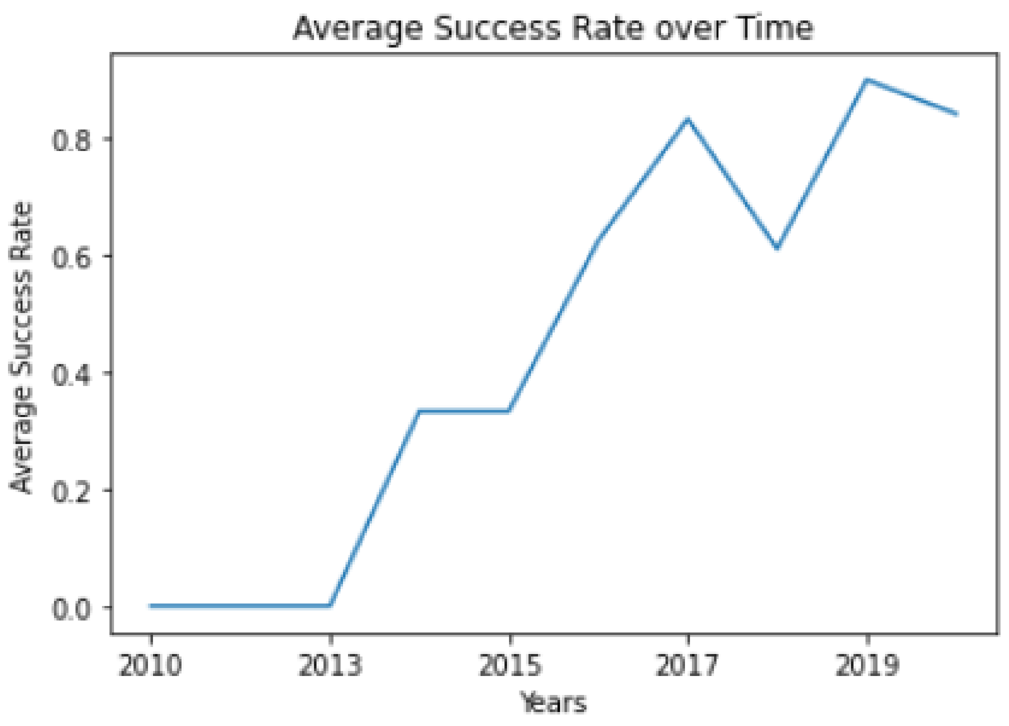
- **API**
  - Obtain historcial launchdata from Open Source REST API for SpaceX
    - Filtered for Falcon 9 launches
    - Mean imputation for payload mass values
- **Web Scraping**
  - Obtain historical launch data from WIpikpedia page
    - Extraction of column names from HTML table header and parsed table
    - Coverted to Pandas data frame

METHODOLOGY: DATA WRANGLING

- **Explore data and deterimne labels for training supervised models**
    - Calculation of number of launches on each site
    - Calculation of number and occurence of each orbit type reached
    - Calculation of number and occurence of mission outcome per orbit type

- **Creation of of landing outcome training label "Class"**
    - Dichotonomous variable
        - 0 means failure or not attempted
        - 1 means succesful landing

METHODOLOGY: EXPLORATORY DATA ANALYSIS (EDA)



- **Exploratory data Analysis with SQL**
  - Use of SQL to display and gather information of various variables of interest, including:
    - Launch Site
    - Payload Masses
    - Boosterversions
    - Mission outcomes
    - Boosterlandings
- **EDA Vidualizations**
  - Use of matplotlib and seaborn to show variety of correlations, for instance Succes Rate by orbit type or Year (see example on the left side)

METHODOLOGY: DATA VISUALIZATION

- **Location of Launch Sites**
  - Create interactive Map of Launch Site with Folium
    - Markers for all launch sites on a map
    - Show number of successful/failed launches per site on the map
    - Calculate and plot distances between lauch site and:
      - Coastlines
      - Railyways
      - Highways
      - Cities
- **Dashboard of Launch Records**
  - Use of Plotly to create Dashboard to facilitate data exploration and communication with stakeholders
  - Dashboard includes:
    - Pie-chart for records per launch site, including dropdown menu with all sites
    - Scatterplot to show success/failure of landing outcome for different payloads through a slider

METHODOLOGY: MODEL DEPLOYMENT

- Preparation of data for ML Analysis
  - Standardization
  - Split in test and training data
- Hyperparameter tuning and fitting of training data with ML Models including ten-fold CV for:
  - Logistic Regression
  - SVM
  - Decision tree
  - KNN
- Evaluation of accuracy to choose best model for our predictions

EDA insights

Section 2

EDA WITH SQL I

SpaceY queries to answer:

Which launch sites did SpaceX use
- CCAFS LC-40, VAFB SLC-4E, KSC LC-39A,CCAFS SLC-40

Examine launch cites beginning with 'CCA'
- Renaming over course of time for CCAFS-LC 40 to CCAFS SLC-40

What's the total payload mass carried by boosters launched by NASA (CRS)
- 45596 KG

What is the average payload mass carried by booster version F9 v1.1?
- 2928.4

When was the first successful landing in ground pad achieved?
- 01-05-2017

EDA WITH SQL II

Which boosters have success in drone ship and have payload mass greater than 4000 but less than 6000?

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

List the number of total failures and successes:

Failure (in flight)	1
Success	99
Success (payload status un clear)	1

List the names of the booster versions which have carried the maximum payload mass:

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

EDA WITH SQL III

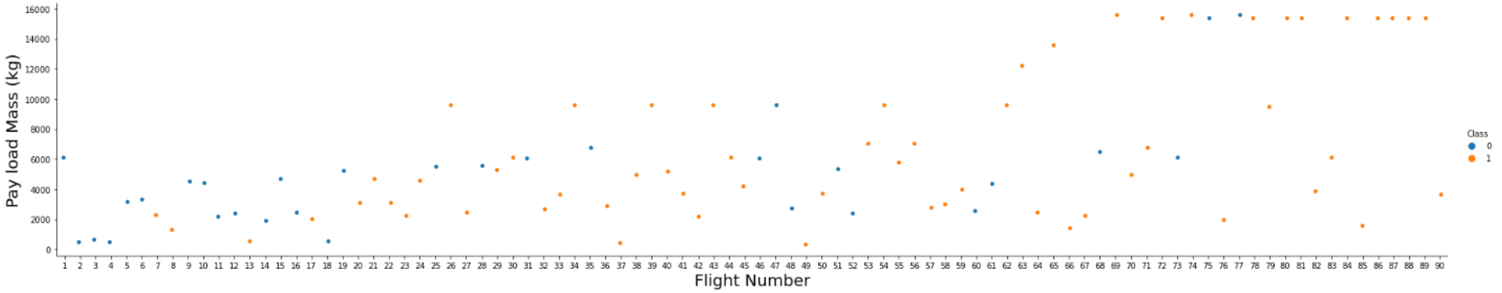
List the records which will display the month names, failure landing_outcomes in drone ship, booster versions, launch_site for the months in year 2015:

Month s	Landing _Outcome	Launch_Site	Booster_Version
01	Failure (drone ship)	CCAFS LC-40	F9 v1.1 B1012
04	Failure (drone ship)	CCAFS LC-40	F9 v1.1 B1015

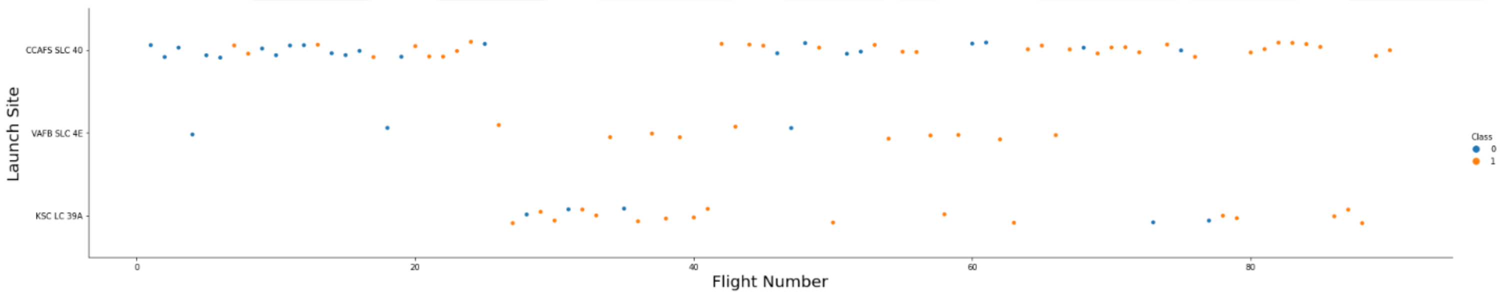
Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Landing _Outcome	Success_Count
Success	20
Success (drone ship)	8
Success (ground pad)	6

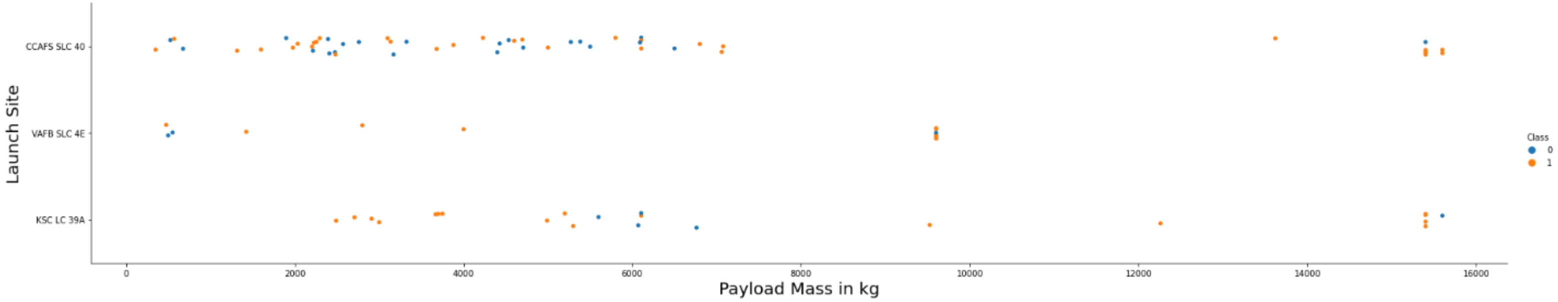
EDA VISUALIZATION I



With increasing launches the success is more common, while payload is particularly for very high loads successful

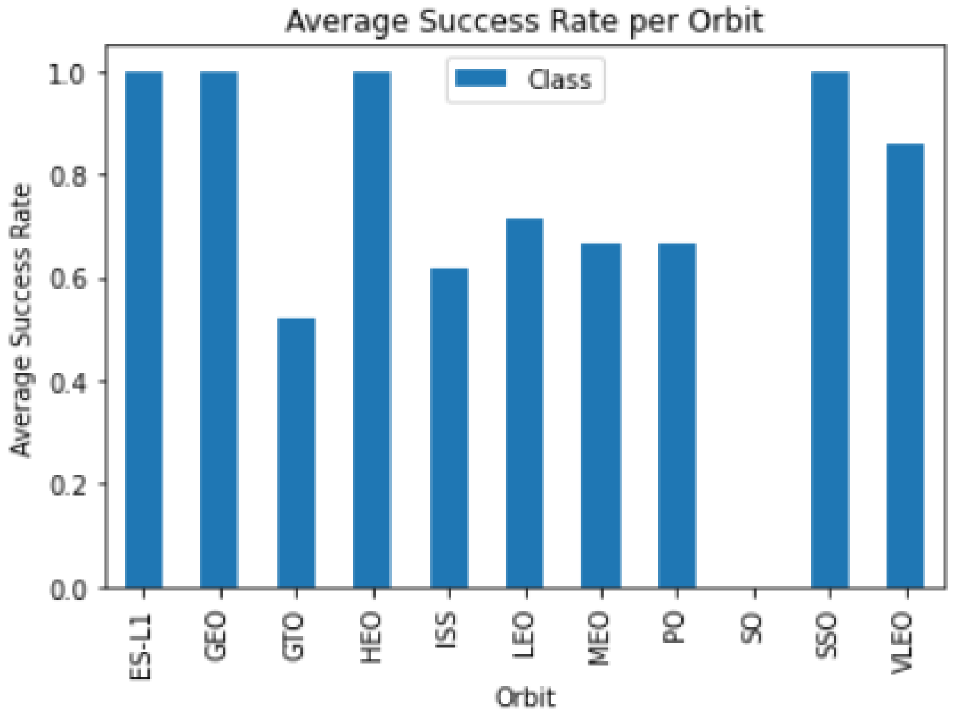


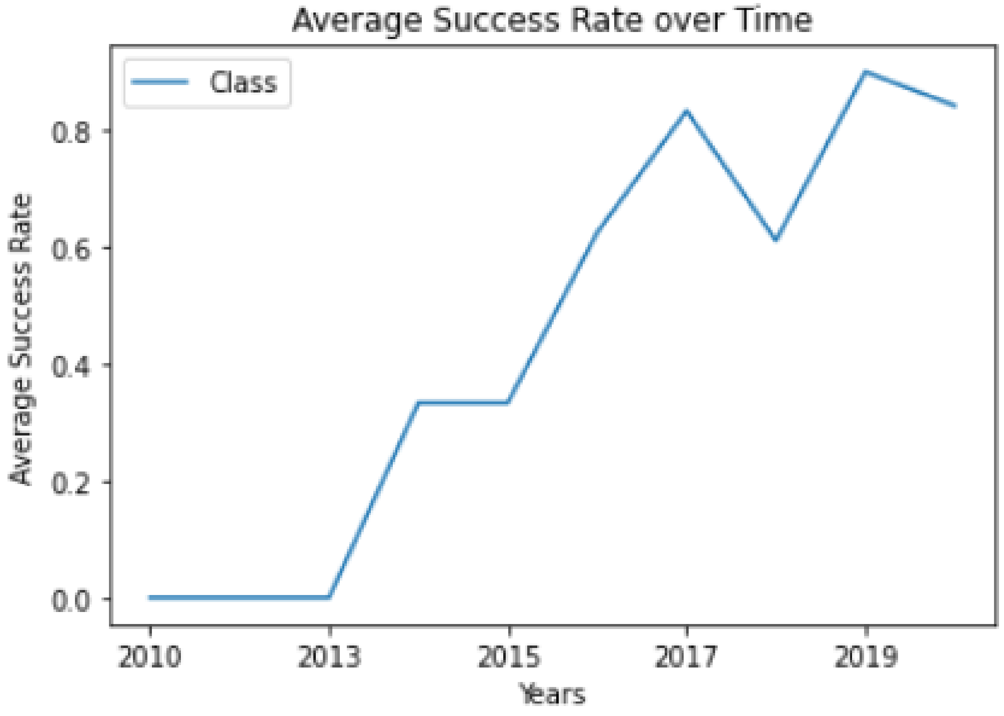
Most failures occured during the initial launces on site CCFAS SLC-40



CCFAS SLC-40 and KSC LC 39A seem to be more favourable for heavy loads

EDA VISUALIZATION II

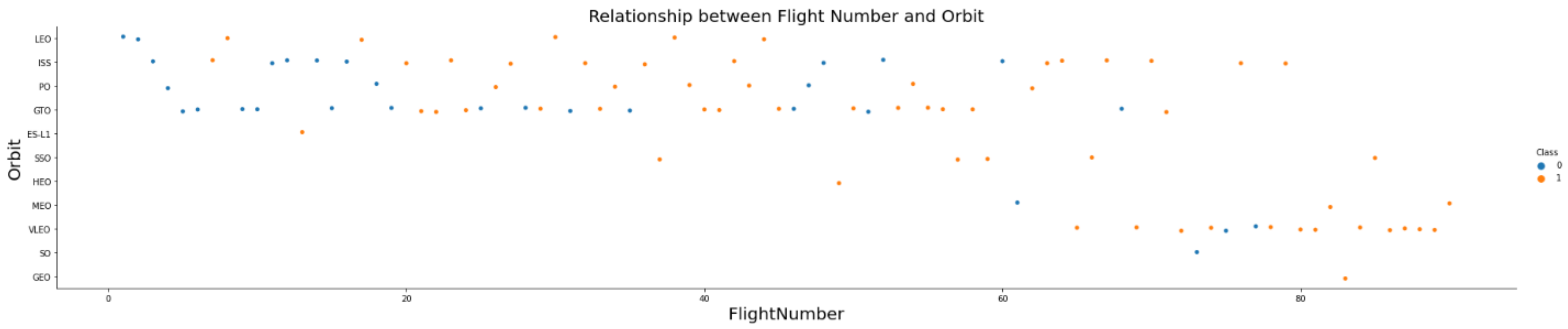


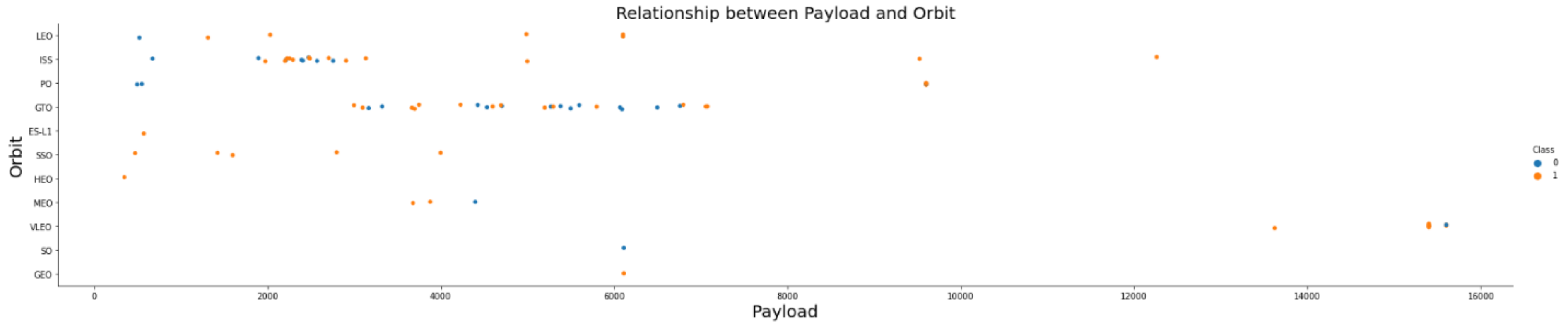


While some orbits had a success rate of 100%, the SO Orbit had 0 % success

We can see a clear growth pattern over the years since the beginning in 2013

EDA VISUALIZATION III





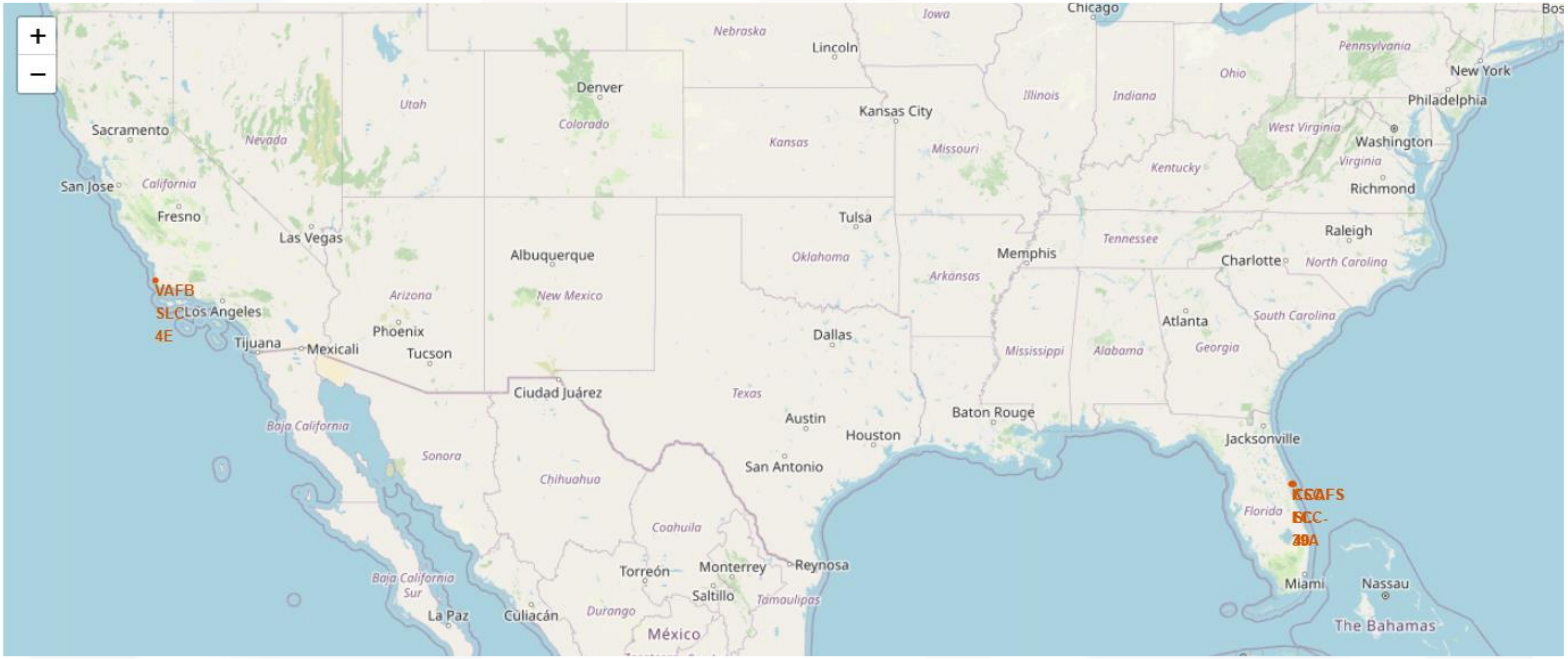
For the LEO orbit the Success appears related to the number of flights; however, there seems to be no relationship between flight number when it's the GTO orbit.

Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits

LAUNCH SITE ANALYSIS

Section 3

RESULTS LAUNCH SITE VISUALIZATION I

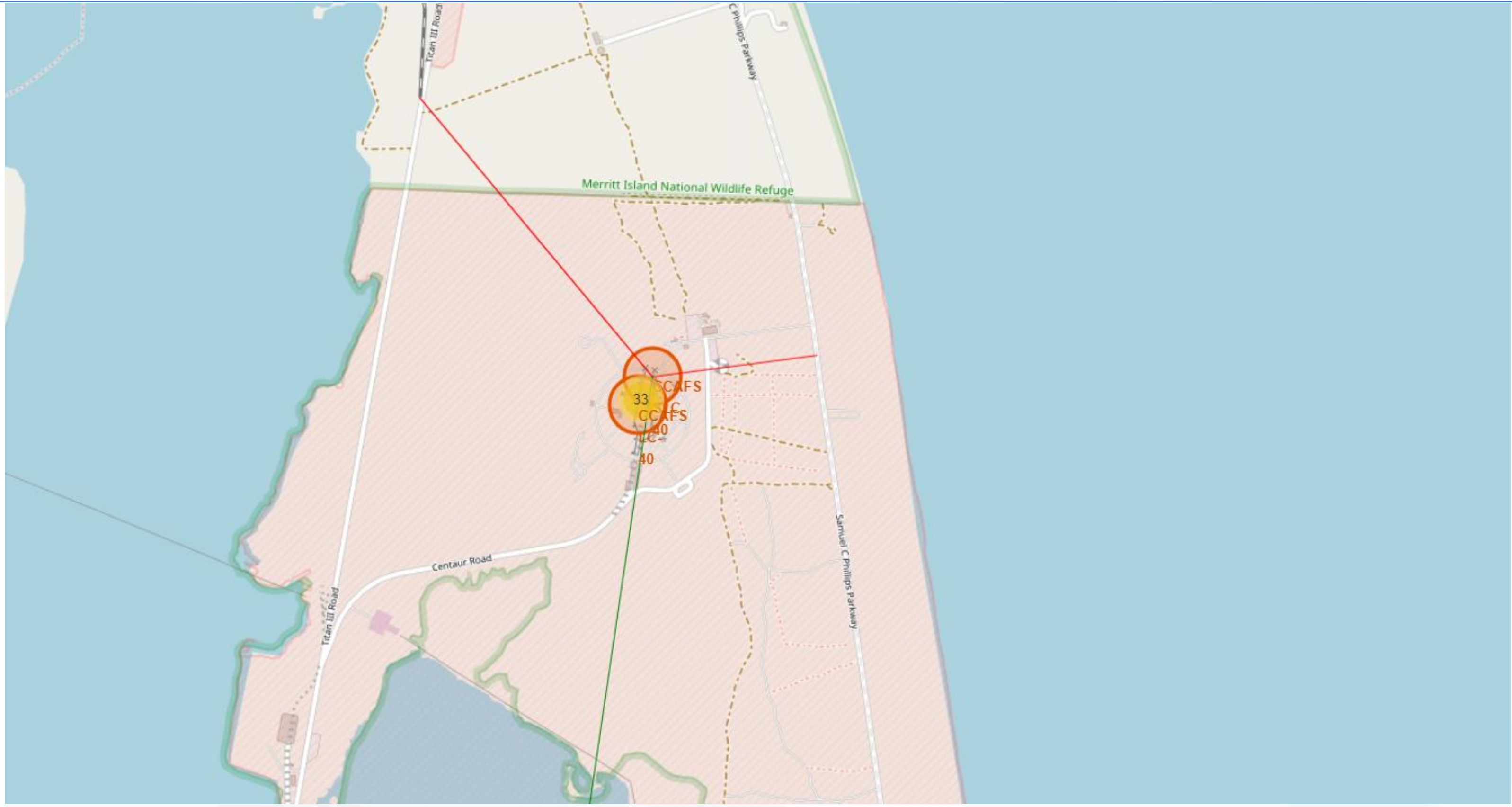


Inspection of launch site locations and proximity to the Equator line, all sites appear to be close to the coast

RESULTS LAUNCH SITE VISUALIZATION II

Visualization of successful/failed launches and launch site name

RESULTS LAUNCH SITE VISUALIZATION III



Proximity analysis of launch site CCAFS SLC-4E to closest coastline railroads, highways and cities

Proximity of lauch site to the clostest:

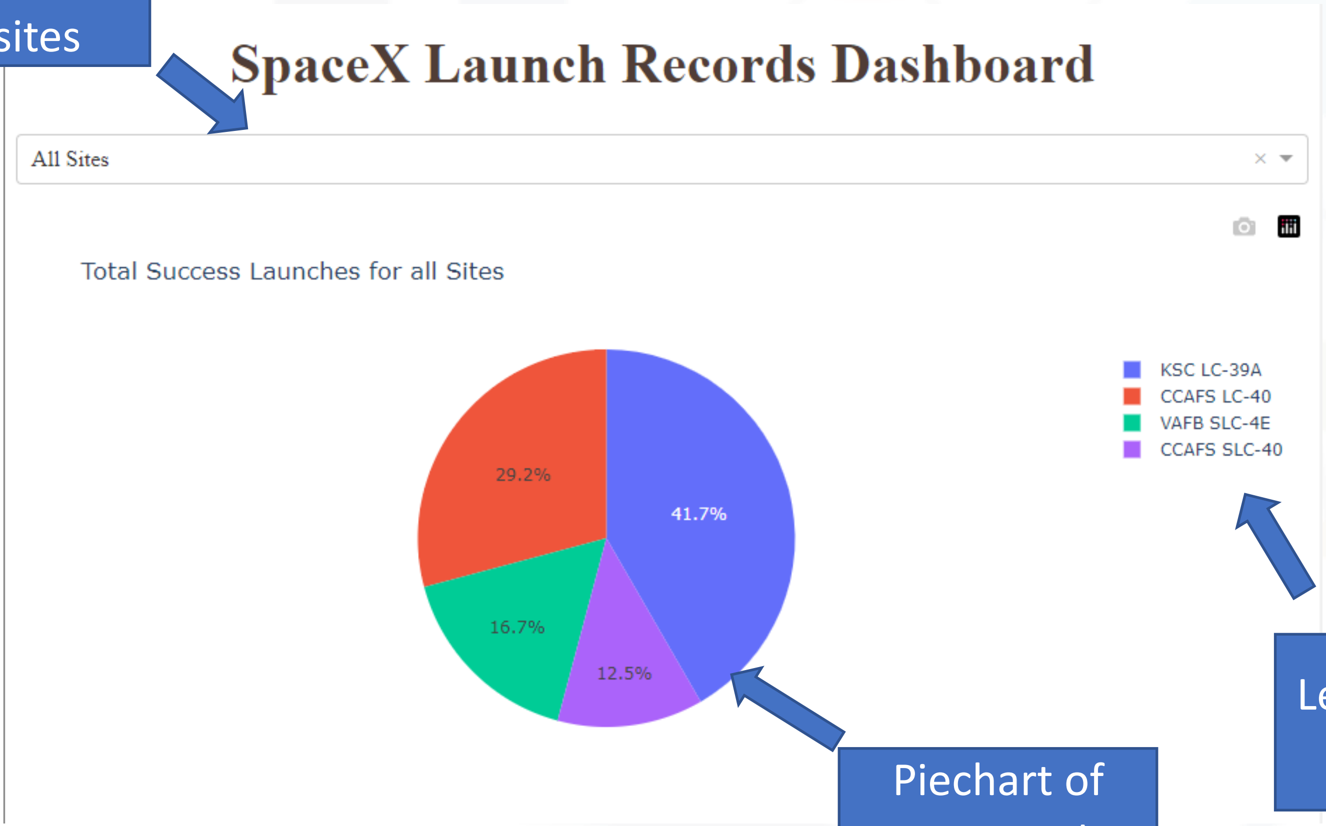
- **City: 17,43 km**
  - Less danger to population
- **Highway: 0,59 km**
  - Facilitates transport of personal and cargo
- **Railroads: 1,28 km**
  - Facilitates transport of heavy cargo
- **Coastline: 0,85 km**
  - Ideal in case of emergency abort

DASHBOARD VISUALIZATION

Section 4

RESULTS: DASHBOARD I

Dropdown menu with all sites



- KSC LC-39A has the most successful launches
- CCAFS LC-40 has the highest launch success rate with rate of all site

Piechart of success rate per site

Legend with sites

RESULTS: DASHBOARD II

Rangeslider for payload

Legend of booster versions

- The highest succes rate seems to be higher for a payload mass lower than 5000 kg, for higher payloads the success rate seems to be lowest

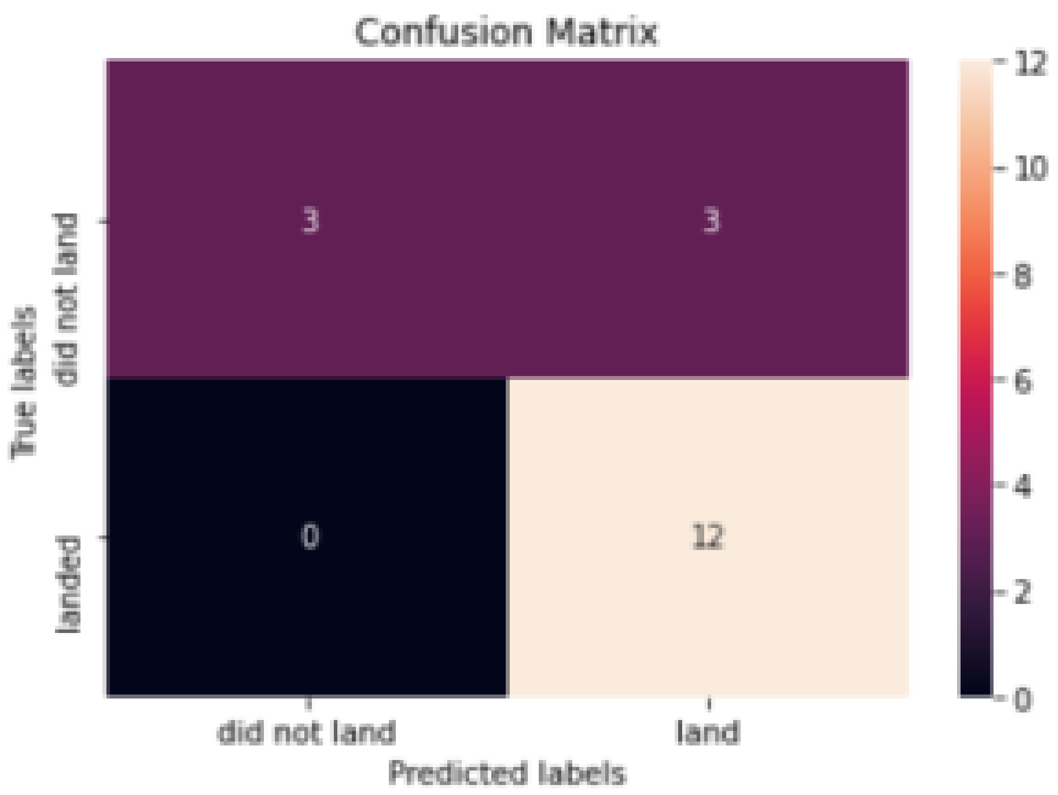
Predictive Analysis (Classification)

Section 6

IBM Developer

SKILLS NETWORK

RESULTS: PREDICTION (CLASSIFICATION)



- Comparing the accuracy of Logistic regression, KNN, SVM and decision trees we reach the same value of roughly 0,83
- Looking at the confusion matrix we can see that the models suffers from false positive predictions

CONCLUSION

- Based on the analysis we are able to predict the successful landing of the first stage boosters with 83,3 % accuracy
- SpaceY can use this information to get an idea when landings will likely succeed and when to include possible failures in their cost analysis

- **Potential improvements:**
  - Include data and analysis about reused 1st stage boosters, since they will reduce the price of the rocket launch

APPENDIX

Notebooks to recreate dataset, analysis, and models:

https://github.com/Tacatico/IBM-Capstone-Project-SpaceX-Analysis/tree/main/Jupyter%20Notebooks