IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

FEBIN MATHAI
FEBRUARY 25,2022

Outline

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

Executive Summary

- SpaceY is commercial rocket launch provider who wants to compete with SpaceX ,leader in the sector
- SpaceX launches rocket at lower cost ,starting at $62 million
- SpaceX reuses the first stage rocket booster so they are able to provide there service at a lower cost
- Using given mission parameters developed model were able to predict the first stage rocket landing successfully with an accuracy 83,3%

Introduction

- This report has been prepared as part of Applied capstone project in coursera provided by IBM
- In this capstone I take the role of data scientist working for new rocket company called spaceY
- With the help of data analysis SpaceY would be able to make more informed bids against SpaceX

Introduction

- SpaceX lunches Falcon 9 rocket a cost of $62 million when they could reuse first stage rocket
- Sometimes SpaceX sacrifice he first sage due to mission parameters like payload orbit etc.
- Therefore aim is to accurately predict the likelihood of thee landing oOf first stage as a proxy for calculating the cost of launch

Section 1

Methodology

Methodology

- Data collection
- Perform data wrangling
- 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

- API
- Acquired historical data from Open source REST API for SpaceX

- Web Scraping
- Acquired historical data from Wikipedia page 'List of Falcon 9 and Falcon Heavy Launches'

[hide] Flight No.	Date and time (UTC)	Version, Booster [b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
1	4 June 2010, 18:45	F9 v1.0[7] B0003[8]	CCAFS, SLC-40	Dragon Spacecraft Qualification Unit	No payload (excl. Dragon Mass)	LEO	SpaceX	Success	Failure[9][10] (parachute)
	First flight of Falcon 9 v1.0.[11] Used a boilerplate version of Dragon capsule which was not designed to separate from the second stage.(more details below) Attempted to recover the first stage by parachuting it into the ocean, but it burned up on reentry, before the parachutes even got to deploy.[12]								
2	8 December 2010, 15:43[13]	F9 v1.0[7] B0004[8]	CCAFS, SLC-40	Dragon demo flight C1 (Dragon C101)	Classified (excl. Dragon Mass)	LEO (ISS)	NASA (COTS) NRO	Success[9]	Failure[9][14] (parachute)
	Maiden flight of SpaceX's Dragon capsule, consisting of over 3 hours of testing thruster maneuvering and then reentry.[15] Attempted to recover the first stage by parachuting it into the ocean, but it disintegrated upon reentry, again before the parachutes were deployed.[12] (more details below) It also included two CubeSats,[16] and a wheel of Brouère cheese. Before the launch, SpaceX discovered that there was a crack in the nozzle of the 2nd stage's Merlin vacuum engine. So Elon just had them cut off the end of the nozzle with a pair of shears and launched the rocket a few days later. After SpaceX had trimmed the nozzle, NASA was notified of the change and they agreed to it.[17]								
3	22 May 2012, 07:44[18]	F9 v1.0[7] B0005[8]	CCAFS, SLC-40	Dragon demo flight C2+[19] (Dragon C102)	525 kg (1,157 lb)[20] (excl. Dragon mass)	LEO (ISS)	NASA (COTS)	Success[21]	No attempt
	The Dragon spacecraft demonstrated a series of tests before it was allowed to approach the International Space Station. Two days later, it became the first commercial spacecraft to board the ISS.[18] (more details below)								
4	8 October 2012, 00:35[22]	F9 v1.0[7] B0006[8]	CCAFS, SLC-40	SpaceX CRS-1[23] (Dragon C103)	4,700 kg (10,400 lb) (excl. Dragon mass)	LEO (ISS)	NASA (CRS)	Success	No attempt
				Orbcomm-OG2[24]	172 kg (379 lb)[25]	LEO	Orbcomm	Partial failure[26]	
	CRS-1 was successful, but the secondary payload was inserted into an abnormally low orbit and subsequently lost. This was due to one of the nine Merlin engines shutting down during the launch, and NASA declining a second reignition, as per ISS visiting vehicle safety rules, the primary payload owner is contractually allowed to decline a second reignition. NASA stated that this was because SpaceX could not guarantee a high enough likelihood of the second stage completing the second burn successfully which was required to avoid any risk of secondary payload's collision with the ISS.[27][28][29]								
5	1 March 2013, 15:10	F9 v1.0[7] B0007[8]	CCAFS, SLC-40	SpaceX CRS-2[23] (Dragon C104)	4,877 kg (10,752 lb) (excl. Dragon mass)	LEO (ISS)	NASA (CRS)	Success	No attempt
	Last launch of the original Falcon 9 v1.0 launch vehicle, first use of the unpressurized trunk section of Dragon.[30]								
6	29 September 2013, 16:00[31]	F9 v1.1[7] B1003[8]	VAFB, SLC-4E	CASSIOPE[23][32]	500 kg (1,100 lb)	Polar orbit LEO	MDA	Success[31]	Uncontrolled (ocean)[d]
	First commercial mission with a private customer, first launch from Vandenberg, and demonstration flight of Falcon 9 v1.1 with an improved 13-tonne to LEO capacity.[30] After separation from the second stage carrying Canadian commercial and scientific satellites, the first stage booster performed a controlled reentry,[33] and an ocean touchdown test for the first time. This provided good test data, even though the booster started rolling as it neared the ocean, leading to the shutdown of the central engine as the roll depleted it of fuel, resulting in a hard impact with the ocean.[31] This was the first known attempt of a rocket engine being lit to perform a supersonic retro propulsion, and allowed SpaceX to enter a public-private partnership with NASA and its Mars entry, descent, and landing technologies research projects.[34] (more details below)								
7	3 December 2013, 22:41[35]	F9 v1.1 B1004	CCAFS, SLC-40	SES-8[23][36][37]	3,170 kg (6,990 lb)	GTO	SES	Success[38]	No attempt [39]
	First Geostationary transfer orbit (GTO) launch for Falcon 9,[36] and first successful reignition of the second stage.[40] SES-8 was inserted into a Super-Synchronous Transfer Orbit of 79,341 km (49,300 mi) in apogee with an inclination of 20.55° to the equator.								

Data Wrangling

- Explored data to determine the labels for training supervised models
  - Calculated he number of launches on each site
  - Calculated outcomes per orbit sand launch site

- Created a landing outcome training label from' Outcome' column
  - Training label was named 'Class'
  - Class was given value based on outcome
    - 0 for failure
    - 1 for success

EDA with Data Visualization

- Read the dataset in pandas data frame
- Used Matplotlib and Seaborn visualization libraries to plot following graphs
  - Flight Number Vs Payload Mass
  - Flight Number Vs Launch Site
  - Payload Vs Launch Site
  - Orbit type Vs Success Rate
  - Flight Number Vs Orbit Type
  - Flight Number Vs Orbit Type
  - Year Vs Success Rate
- test/EDA with Data Visualization.ipynb at master · FebinMathai/test (github.com)

EDA with SQL

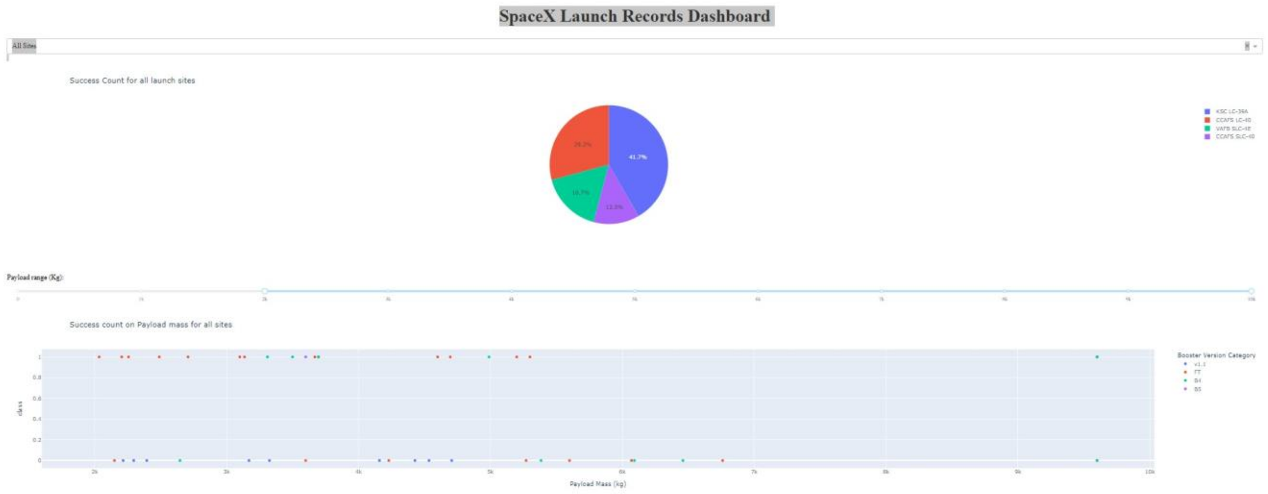
- Loaded data into an IBM DB2 instance
- Ran SQL queries to display and list information about
  1. Launch Sites
  2. Payload Masses
  3. Booster Versions
  4. Mission Outcomes
  5. Booster Landings
- test/EDA with SQL.ipynb at master · FebinMathai/test (github.com)

Build an Interactive Map with Folium

- Launch Site Location was analyzed
  - Used python interactive mapping library called folium
  - Marked all launch sites on a map
  - Marked the successful and failed launches
  - Calculated the distance to various land marks from the launch sites
    - Railways
    - Highways
    - Coastlines
    - Cities
  - test/Interactive Visual Analytics with Folium lab.ipynb at master · FebinMathai/test (github.com)

Launch Records Dashboard

- Used python Intractive dashboarding library called Plotly Dash to explore data in real time
- Pie Chart showing success rate
- Scattered chart Showing payload mass Vs landing outcome
- Drop down menu to choose between all sites and individual launch sites



Predictive Analysis (Classification)

Defined Required functions → Imported the data frame created during the collection → Standardized the data → Developed various models → Data was fitted and tested on the model → Accuracy of each model was checked using testing data

Results

EXPLORATORY DATA

INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS

PREDICTIVE ANALYSIS RESULTS

Section 2

Insights drawn from EDA

RESULTS: EDA WITH SQL

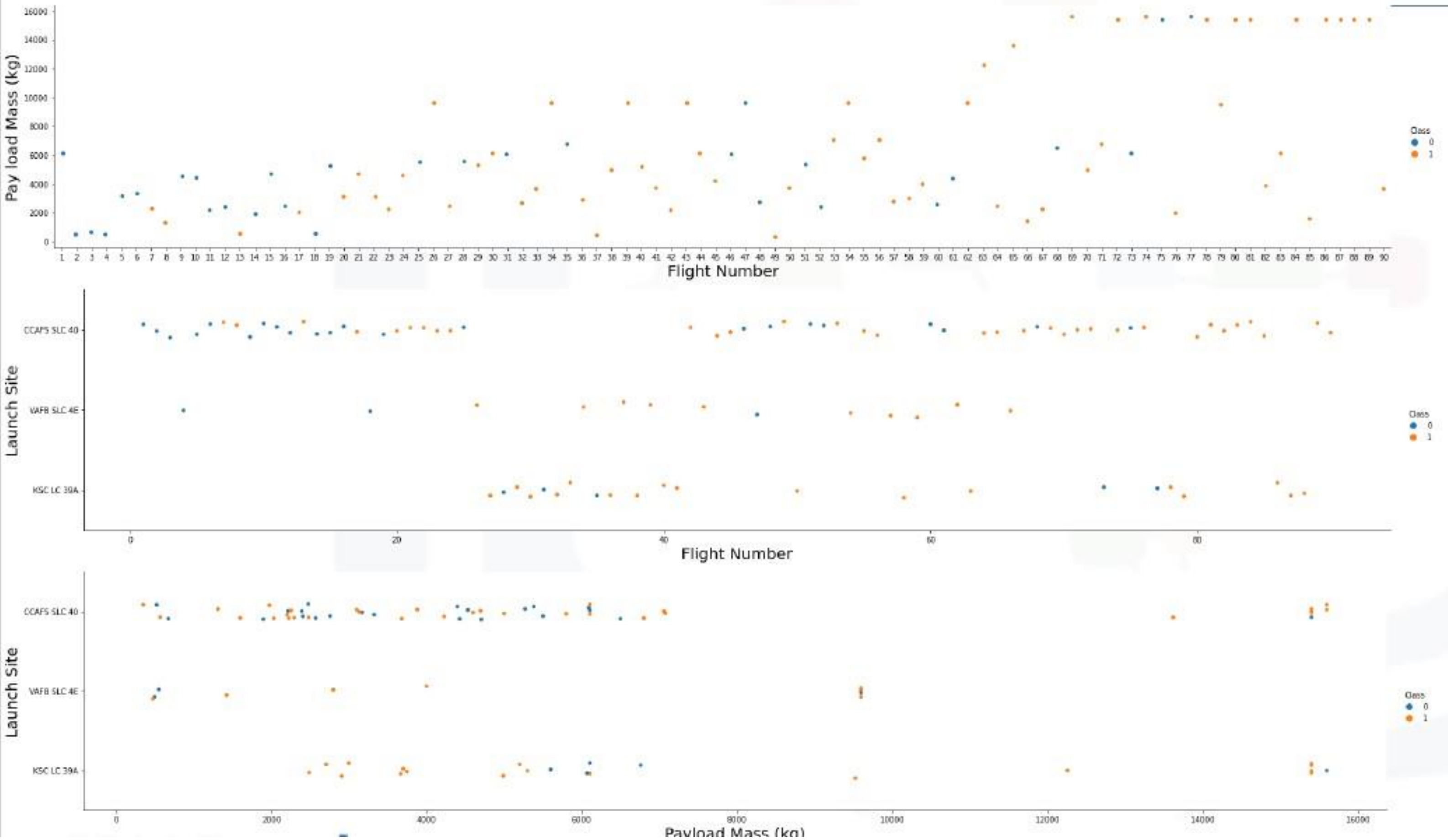
The team at SpaceY had some very specific questions to answer with SQL:

- What launch sites has SpaceX used?
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E
- Examine launch site and date records where launch sites begin with the string 'CCA', do they overlap?
  - Last launch from CCAFS LC-40 was 2016-08-14
  - First launch from CCAFS SLC-40 was 2017-12-15
  - Wikipedia confirms Cape Canaveral Space Launch Complex 40 was renamed in 2017
- Display the total payload mass carried by boosters launched by NASA (CRS)
  - 45,596 KG, total
- Display average payload mass carried by booster version F9 v1.1
  - 340 KG, average
- List the date when the first successful landing outcome in ground pad was achieved.
  - 2015-12-22, more than 5 years after the first Falcon 9 launch on 2010-06-04

RESULTS: EDA WITH SQL (CONTINUED)

- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - F9 FT B1021.1
  - F9 FT B1023.1
  - F9 FT B1029.2
  - F9 FT B1038.1
  - F9 B4 B1042.1
  - F9 B4 B1045.1
  - F9 B5 B1046.1
- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order
  - 10 - No attempt5 - Failure (drone ship)
  - 5 - Success (drone ship)
  - 3 - Controlled (ocean)
  - 3 - Success (ground pad)
  - 2 - Failure (parachute)
  - 2 - Uncontrolled (ocean)
  - 1 - Precluded (drone ship)
- List the names of the booster_versions which have carried the maximum payload mass.
  - 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
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Failure (drone ship) F9 v1.1 B1012 CCAFS LC-40
  - Failure (drone ship) F9 v1.1 B1015 CCAFS LC-40
- List the total number of successful and failure mission outcomes
  - 1 - Failure (in flight)
  - 99 - Success
  - 1 - Success (payload status unclear)

RESULTS: EDA WITH VISUALIZATION

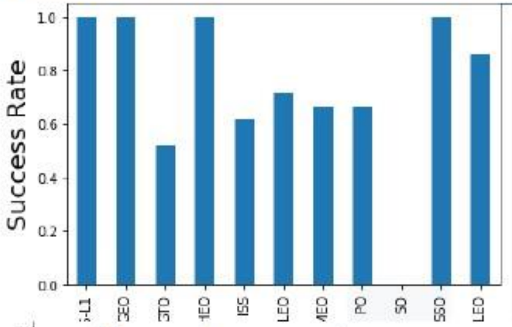


FlightNumber x PayloadMass,
1st stage landing success positively correlated with continuous launch attempts, while negatively correlated with payload mass

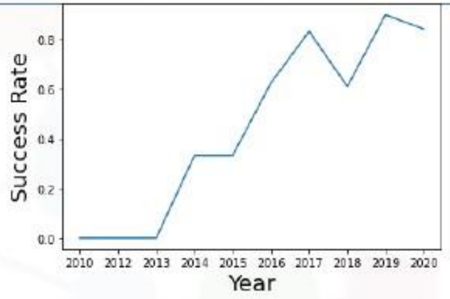
FlightNumber x LaunchSite,
CCAFS SLC 40 appears to have been where most of the early 1st stage landing failures took place

PayloadMass x LaunchSite,
CCAFS SLC 40 and KSC LC 39A appear to be favored for heavier payloads

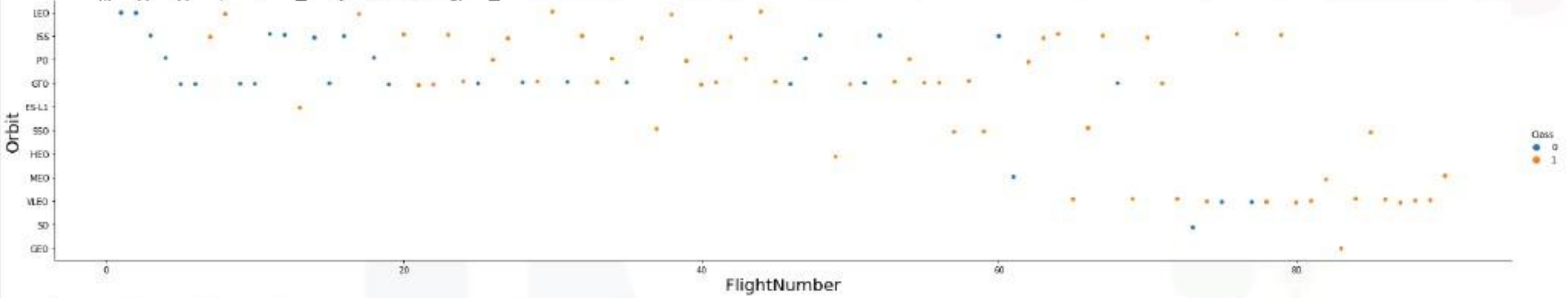
RESULTS: EDA WITH VISUALIZATION (CONTINUED)



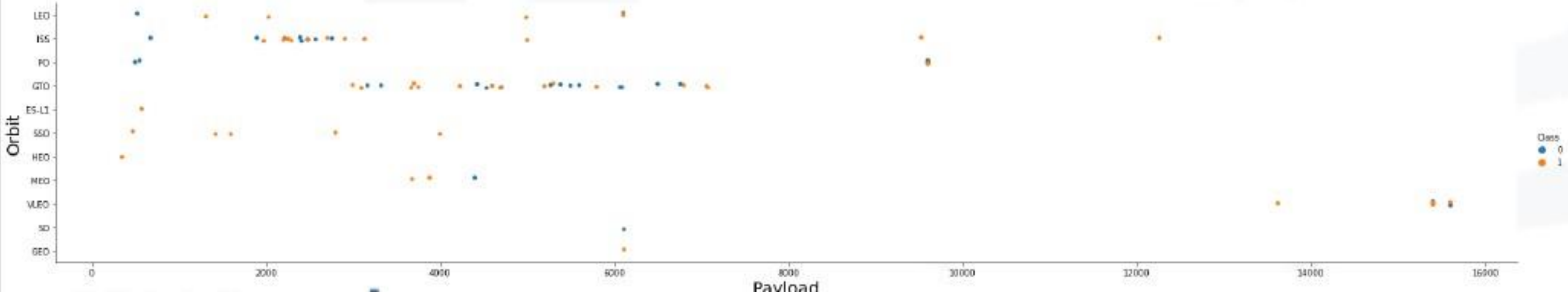
Orbit type x Success rate,
All orbit types except 'SO' have had successful 1st stage landings



Year x Success rate,
success rate trending positively on a yearly basis since 2013



FlightNumber x Orbit type,
flight number positively correlated with 1st stage recovery for all orbit types



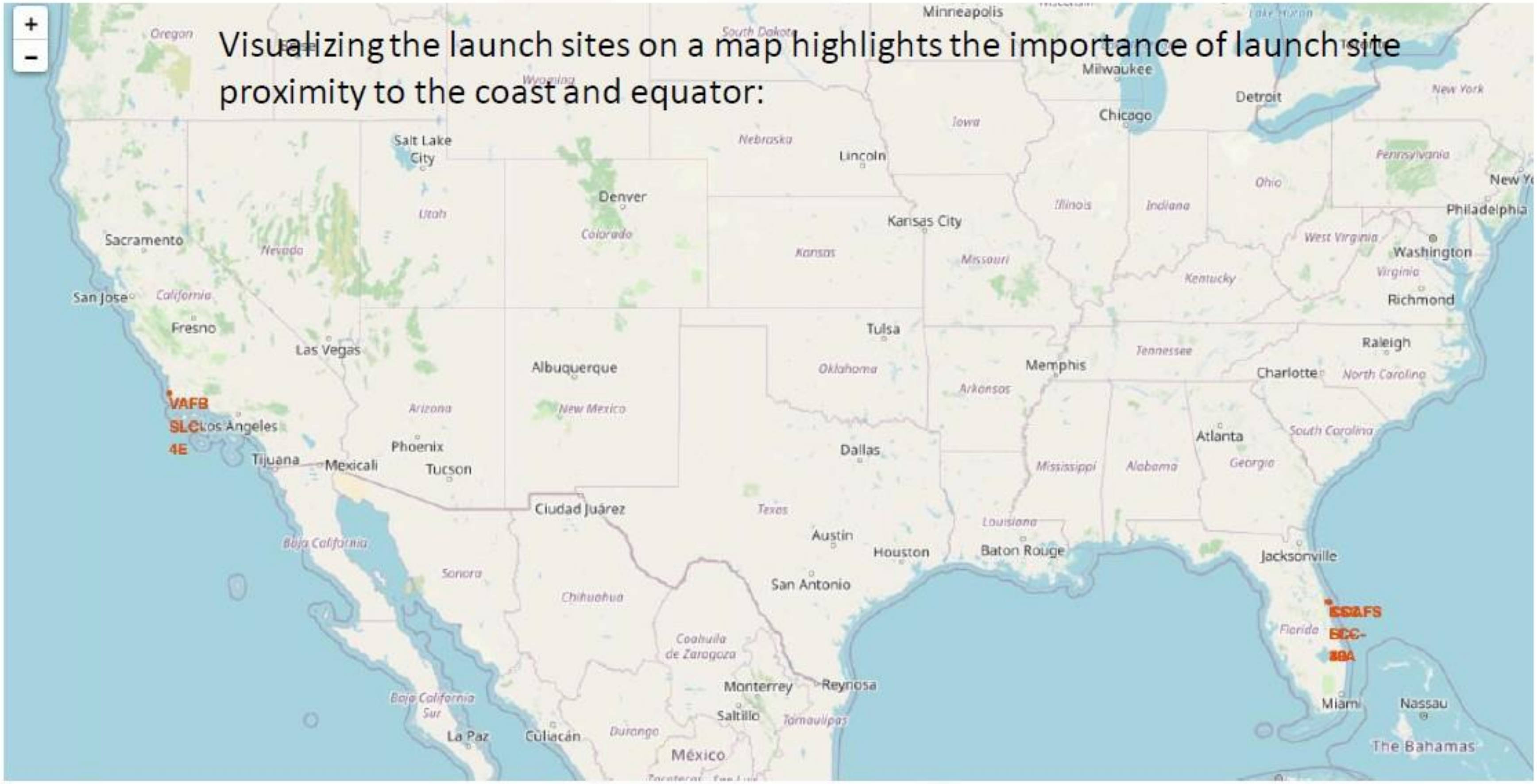
PayloadMass x Orbit type,
heavier payloads have a negative influence on GTO orbits and positive influence on ISS orbits

Section 3

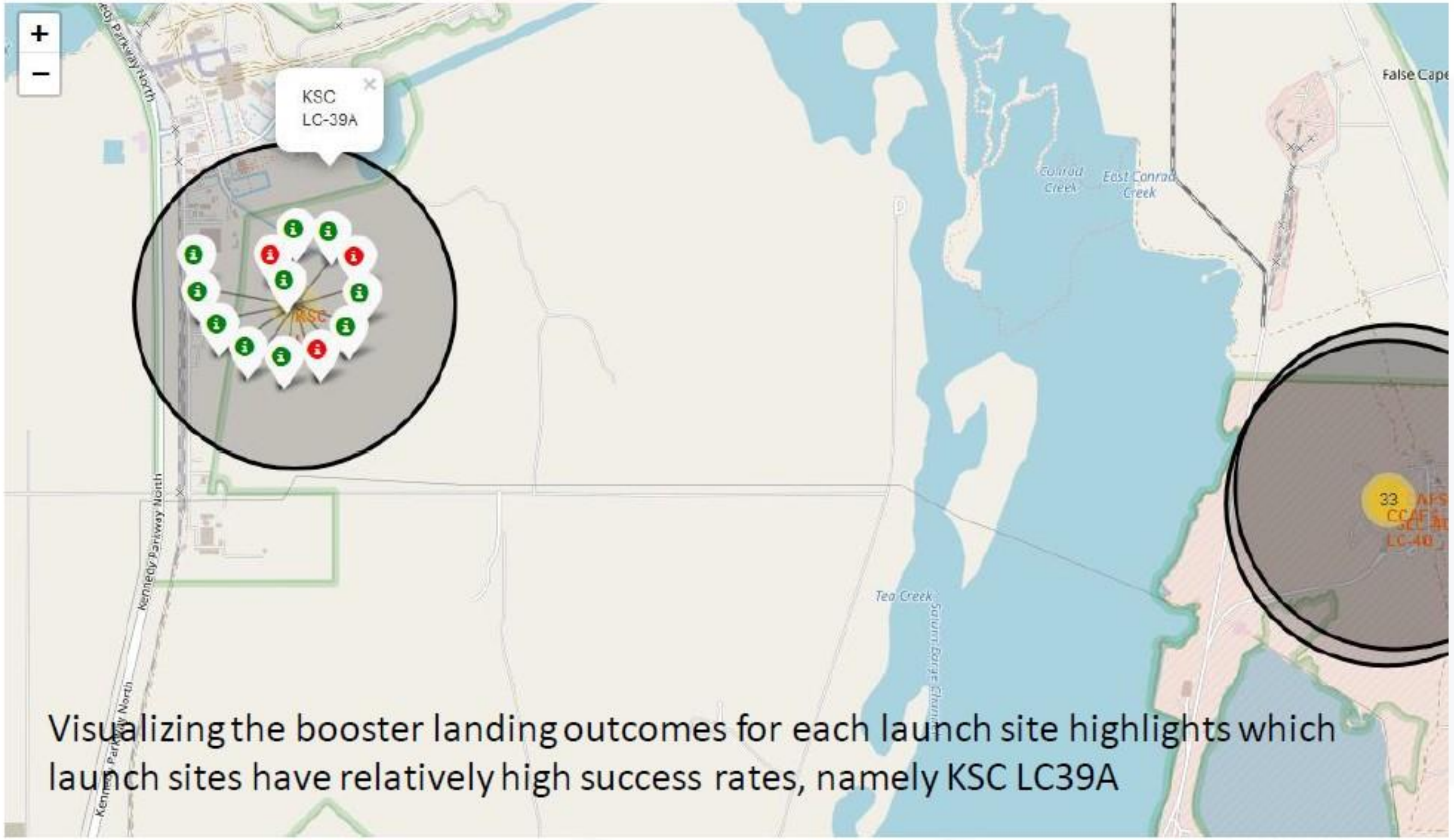
Launch Sites Proximities Analysis

RESULTS: LAUNCH SITE LOCATION ANALYSIS

Visualizing the launch sites on a map highlights the importance of launch site proximity to the coast and equator:

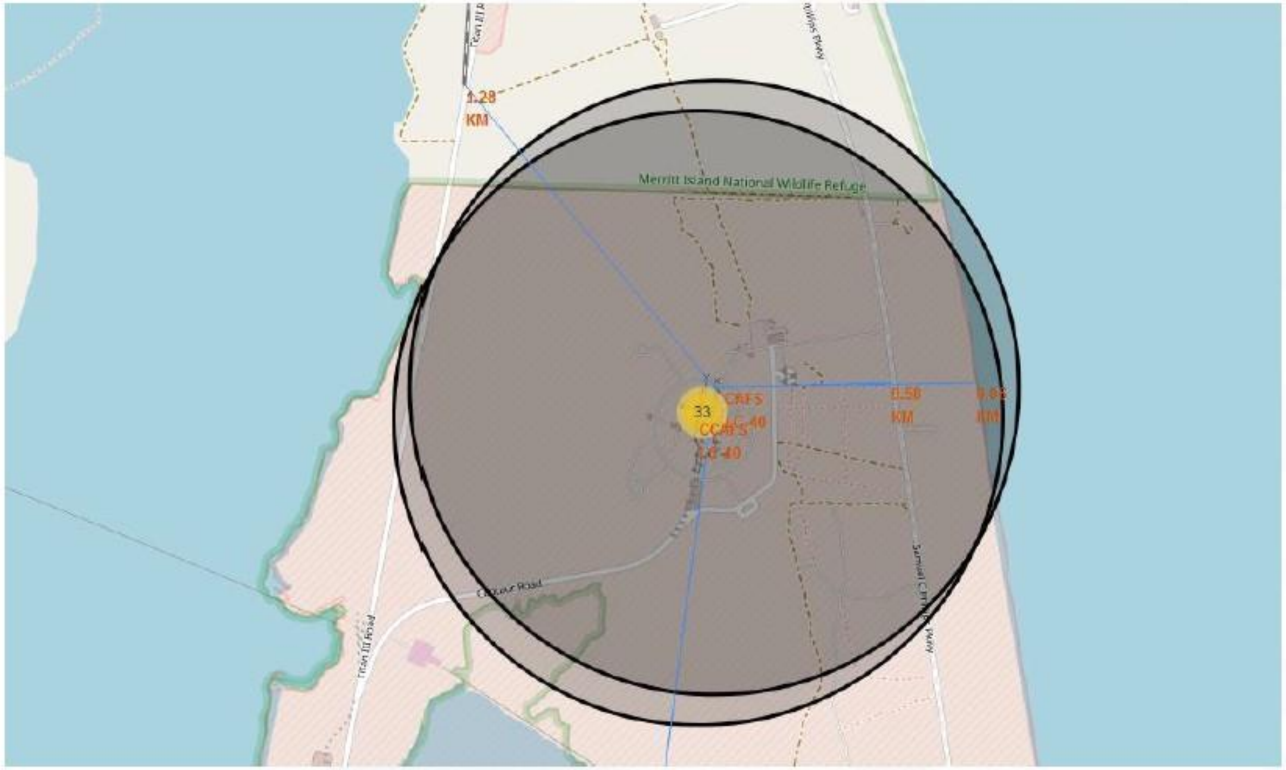


RESULTS: LAUNCH SITE LOCATION ANALYSIS (CONTINUED)



Visualizing the booster landing outcomes for each launch site highlights which launch sites have relatively high success rates, namely KSC LC39A

RESULTS: LAUNCH SITE LOCATION ANALYSIS (CONTINUED)



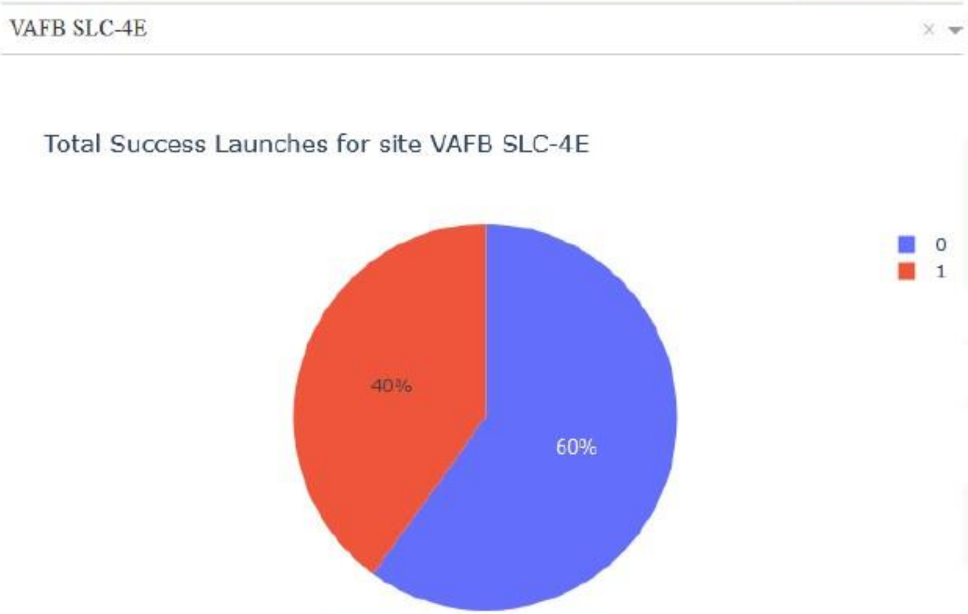
- Visualizing the railway, highway, coastline, and city proximities for each launch site allows us to see how close each is, for example:
- Proximities for CCAFS SLC-40:
  - railway: 1.28 km
    - transporting heavy cargo
  - highway: 0.58 km
    - transporting personel and equipment
  - coastline: 0.86 km
    - optionality to abort launch and attempt water landing
    - minimizing risk from falling debris
  - city: 51.43 km
    - minimizing danger to population dense areas.

Section 4

Build a Dashboard with Plotly Dash

RESULTS: LAUNCH RECORDS DASHBOARD

RESULTS: LAUNCH RECORDS DASHBOARD (CONTINUED)



Example dashboard view:
Booster landing sucess rate for VAFB SLC-4E

- Explore the dashboard yourself:
  - https://ibm-applied-data-science-capst.herokuapp.com/
  - Enabling stakeholders to explore and manipulate the data in an interactive and real-time way
- Dashboard observations:
  - FAFB SLC-4E had the heaviest successful booster landing success
  - KSC LC-39A has the highest booster landing success rate
  - Payloads < 5,300 kg had the highest booster landing success rate
  - Payloads > 5,300 kg had the lowest booster landing success rate

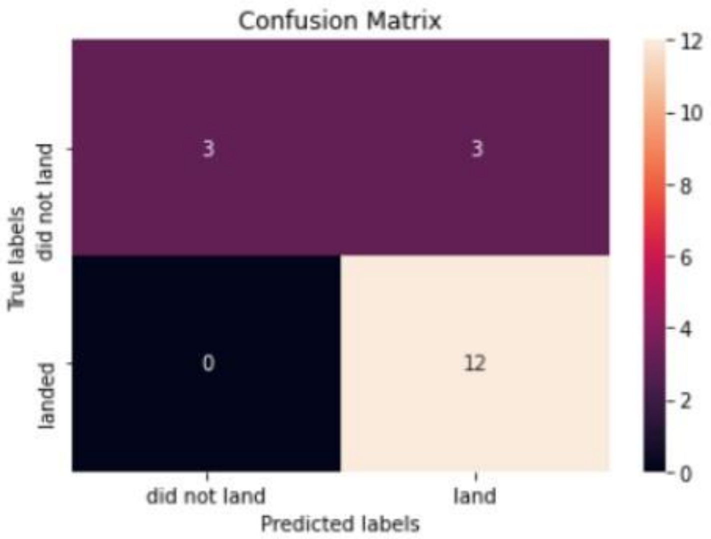
Section 5

Predictive Analysis (Classification)

RESULTS: CLASSIFICATION ACCURACY

- Each of the four models built came back with the same accuracy score, 83.33%

RESULTS: CONFUSION MATRIX



- The confusion matrices of the best performing models (4-way-tie) are the same
- The major problem is false positives as evidenced by the models incorrectly predicting the 1st stage booster to land in 3 out of 18 samples in the test set

CONCLUSION

- Using the models from this report SpaceY can predict when SpaceX will successfully land the 1st stage booster with 83.3% accuracy
- SpaceX public statements indicate the 1st stage booster costs upwards of $15 million to build
- This will enable SpaceY to make more informed bids against SpaceX, since they will have a good idea when to expect the SpaceX bid to include the cost of a sacrificed 1st stage booster
- With a list price of $62 million per launch, sacrificing the $15+ million 1st stage, would put the SpaceX bid at upwards of $77 million
- Biggest opportunities going forward to make even more informed bids:
  - Freeze the best performing combination of model and hyperparameters and re-fit using the whole dataset instead of just the training data
    - Potentially better than using only part of the data to fit the model, but you would no longer be able to measure the accuracy of the resulting model
  - Incorporate additional launch data to the dataset and model as it becomes available
  - Subdivide the current model into two models
    - Predict if SpaceX will ATTEMPT to land the 1st stage
    - Predict if SpaceX will SUCCEED in their attempt
  - Create a related model that predicts if SpaceX will launch using a previously-flown 1st stage booster
    - Would enable SpaceY to take into account when the SpaceX bid would likely include a discount

APPENDIX

- Notebooks to recreate dataset, analysis, and models:
  - https://github.com/brt-h/Applied-Data-Science-Capstone/blob/main/Hands-on%20Lab_%20Complete%20the%20Data%20Collection%20API%20Lab.ipynb
  - https://github.com/brt-h/Applied-Data-Science-Capstone/blob/main/Hands-on%20Lab_%20Data%20Collection%20with%20Web%20Scraping.ipynb
  - https://github.com/brt-h/Applied-Data-Science-Capstone/blob/main/Hands-On%20Lab_%20Data%20Wrangling.ipynb
  - https://github.com/brt-h/Applied-Data-Science-Capstone/blob/main/Hands-on%20Lab%20Complete%20the%20EDA%20with%20SQL.ipynb
  - https://github.com/brt-h/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb
  - https://github.com/brt-h/Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb
  - https://github.com/brt-h/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py
  - https://github.com/brt-h/Applied-Data-Science-Capstone/blob/main/Hands-on%20Lab_%20Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash.ipynb
  - https://github.com/brt-h/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

- Acknowledgments
  - Thank you to Joseph Santarcangelo at IBM for creating the course and materials
  - Thank you to Lakshmi Holla at IBM for assisting me with questions and troubleshooting

- References
  - https://aviationweek.com/defense-space/space/podcast-interview-spacexs-elon-musk
    - Interview with Elon Musk where he discloses the 1st stage booster to cost upwards of $15 million
  - https://datascience.stackexchange.com/a/33050
    - Explanation of why you would rebuild your model using the full dataset
  - https://www.spacex.com/vehicles/falcon-9/
    - Source of SpaceX's advertised $62 million launch price

Thank you!