IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Arav Parikh
May 1, 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Exploratory data analysis (data visualization + querying)
  - Interactive map creation
  - Dashboard creation
  - Predictive analysis
- Summary of all results
  - EDA findings
  - Dashboard insights
  - Trained machine learning classifier results

Introduction

- Project Background and Context
  - SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars while other providers cost upward of 165 million dollars each
  - Much of the savings are because SpaceX can reuse the first stage; therefore, if whether the first stage will land can be determined, then the cost of a launch can also be determined
  - This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Guiding Questions
  - Will the Falcon 9 first stage land successfully?
  - What factors (e.g. payload mass, orbit, launch site, etc.) impact first stage landing success?
  - What is the best way to model the data for this classification problem?
  - How effective are these modeling techniques in predicting landing success?

Section 1

Methodology

Methodology

Executive Summary

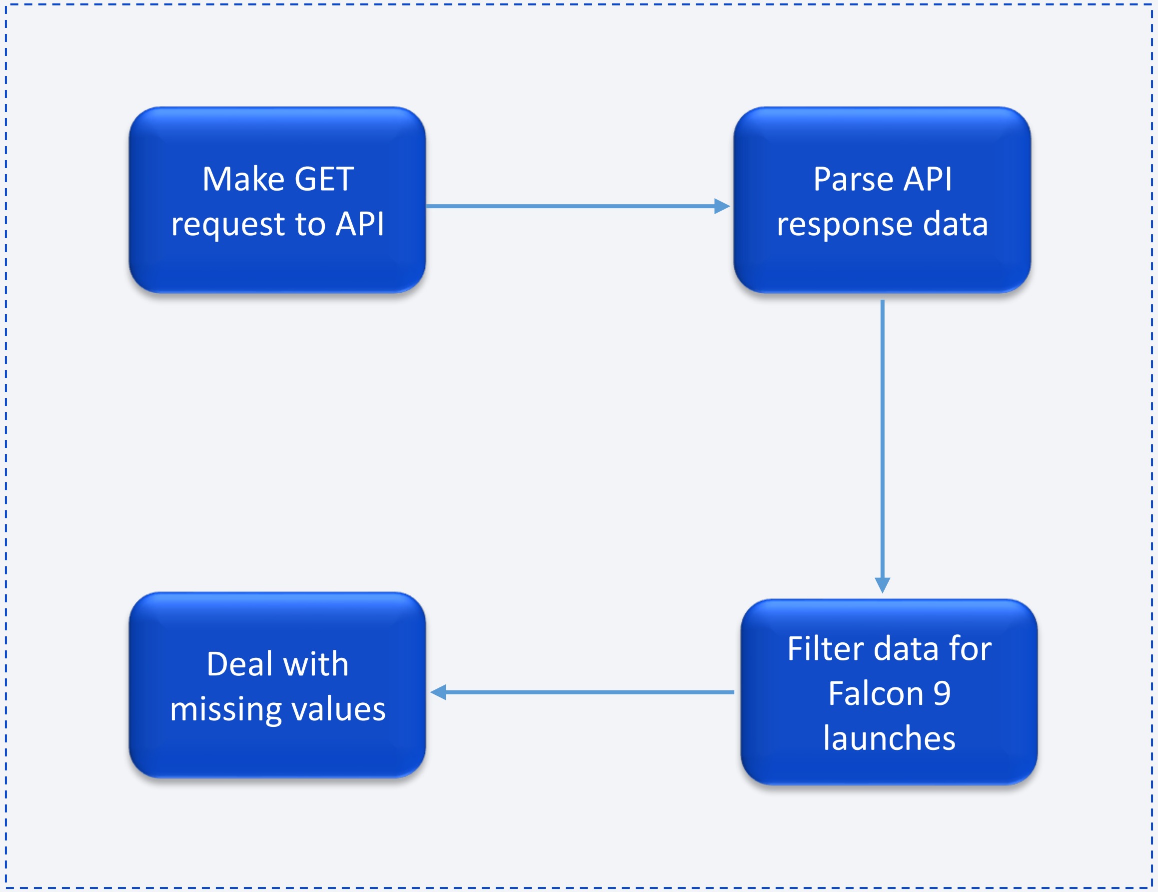
- Data Collection Methodology
  - Data Sources: SpaceX REST API + Wikipedia Web Scraping
- Data Processing Methodology
  - Data Wrangling → Binary Classification Problem Conversion
- Exploratory Data Analysis (EDA) using Visualization and SQL
- Interactive Visual Analytics using Folium and Plotly Dash
- Predictive Analysis using Classification Models
  - Data Partitioning → Model Training + Hyperparameter Tuning → Model Testing

Data Collection

- Dataset Collection
  - SpaceX REST API (https://api.spacexdata.com/v4/rockets/)
    - Columns Extracted: FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude
  - Wikipedia Web Scraping (https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922)
    - Columns Extracted: Flight No., Launch site, Payload, Payload mass, Orbit, Customer, Launch outcome, Version booster, Booster landing, Date, Time

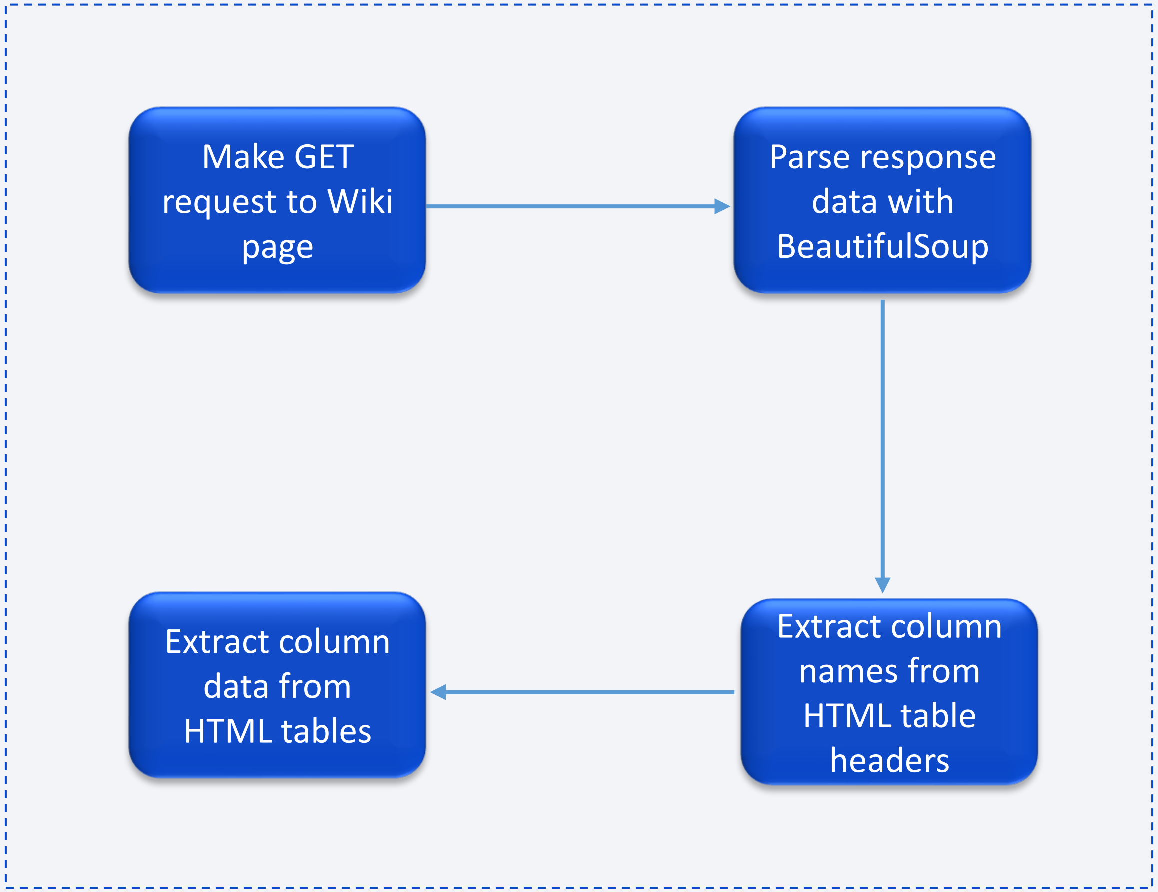
Data Collection – SpaceX API

- Source Code:



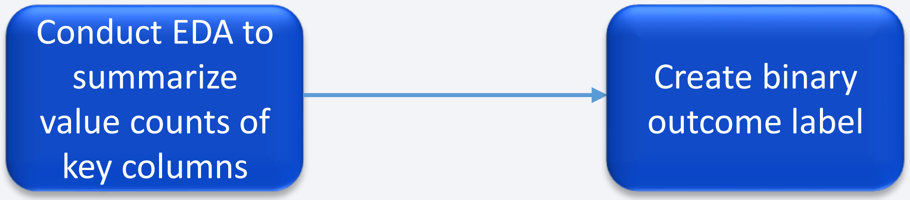
Data Collection - Scraping

- Source Code:



Data Wrangling

- Rationale for Binary Classification Problem Conversion
  - There are several different cases where the booster did not land successfully
    - Sometimes a landing was attempted but failed due to an accident; for example, True Ocean means the mission outcome was a successful landing in a specific region of the ocean while False Ocean means the mission outcome was an unsuccessful landing in a specific region of the ocean
    - True RTLS means the mission outcome was a successful landing on a ground pad while False RTLS means the mission outcome was an unsuccessful landing on a ground pad
    - True ASDS means the mission outcome was a successful landing on a drone ship while False ASDS means the mission outcome was an unsuccessful landing on a drone ship



EDA – Data Visualization

- Plots Used
  - Scatter Plots: Flight Number vs. Launch Site, Payload vs. Launch Site, Flight Number vs. Orbit Type, Payload vs. Orbit Type
  - Bar Chart: Success Rate vs. Orbit Type
  - Line Chart: Launch Success Yearly Trend
- Source Code:

EDA – SQL

- Queries Performed
  - Unique launch sites
  - Launch sites beginning with ‘CCA’
  - Total payload mass of NASA boosters
  - Average payload mass of F9 v1.1 booster
  - First successful ground pad landing
  - Successful boosters with payload mass in specific range…
- Source Code:

Interactive Map Creation – Folium

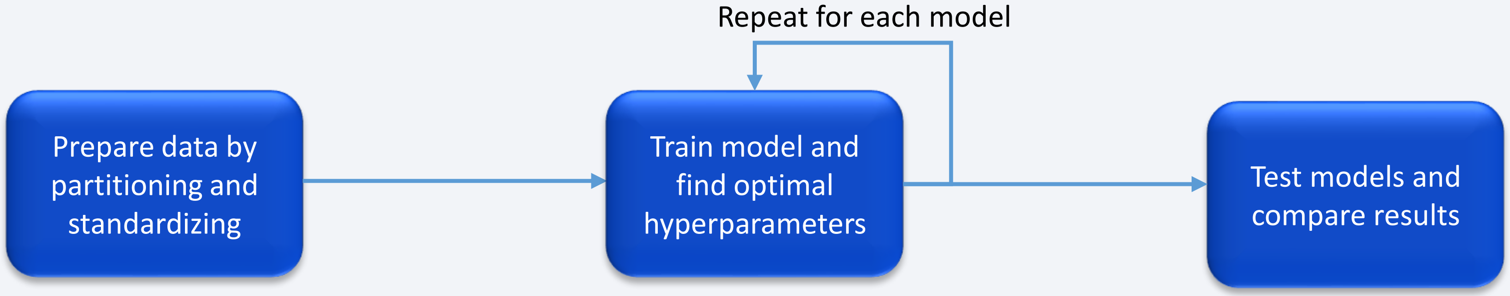
- Launch Site Markers
  - Markers with circles and labels at each launch site to visualize geographic distribution of sites
- Launch Outcome Markers
  - Marker clusters with color-coded outcomes at each launch site to gauge site success rate
- Proximity Distance Lines
  - Lines from launch site to proximities (railroads, highways, cities, etc.) to measure distances
- Source Code:

Dashboard Creation – Plotly Dash

- Launch Site Dropdown List
  - Selection of specific launch site or default of all sites
- Launch Success Pie Chart
  - Pie chart displaying success rate of selected site(s)
- Payload Mass Slider
  - Selection of range for payload mass
- Payload Mass vs. Success Rate Scatter Plot by Launch Site and Booster Version
  - Scatter plot demonstrating correlation between payload mass and success rate
- Source Code:

Predictive Analysis (Classification)

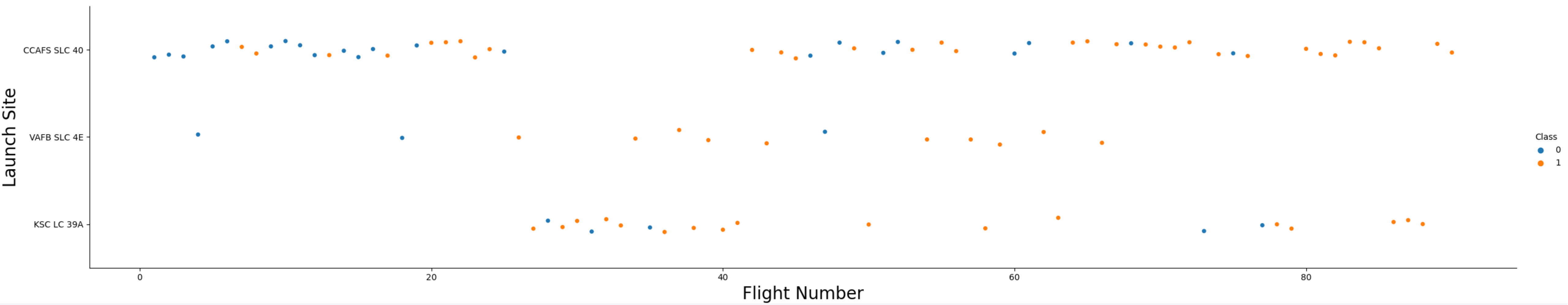
- Models Used (Alongside Grid Search Cross Validation)
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbors
- Metrics Used
  - Accuracy Score
  - Confusion Matrix
- Source Code:



Section 2

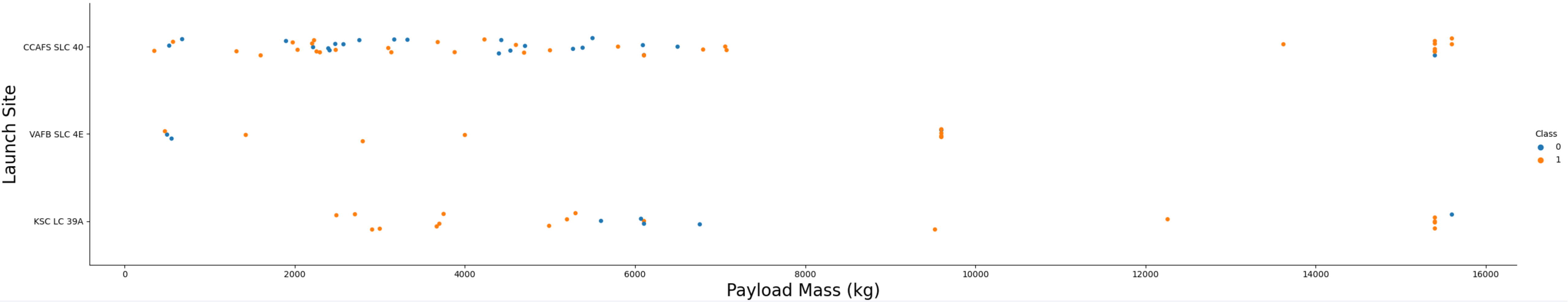
Insights drawn from EDA

Flight Number vs. Launch Site



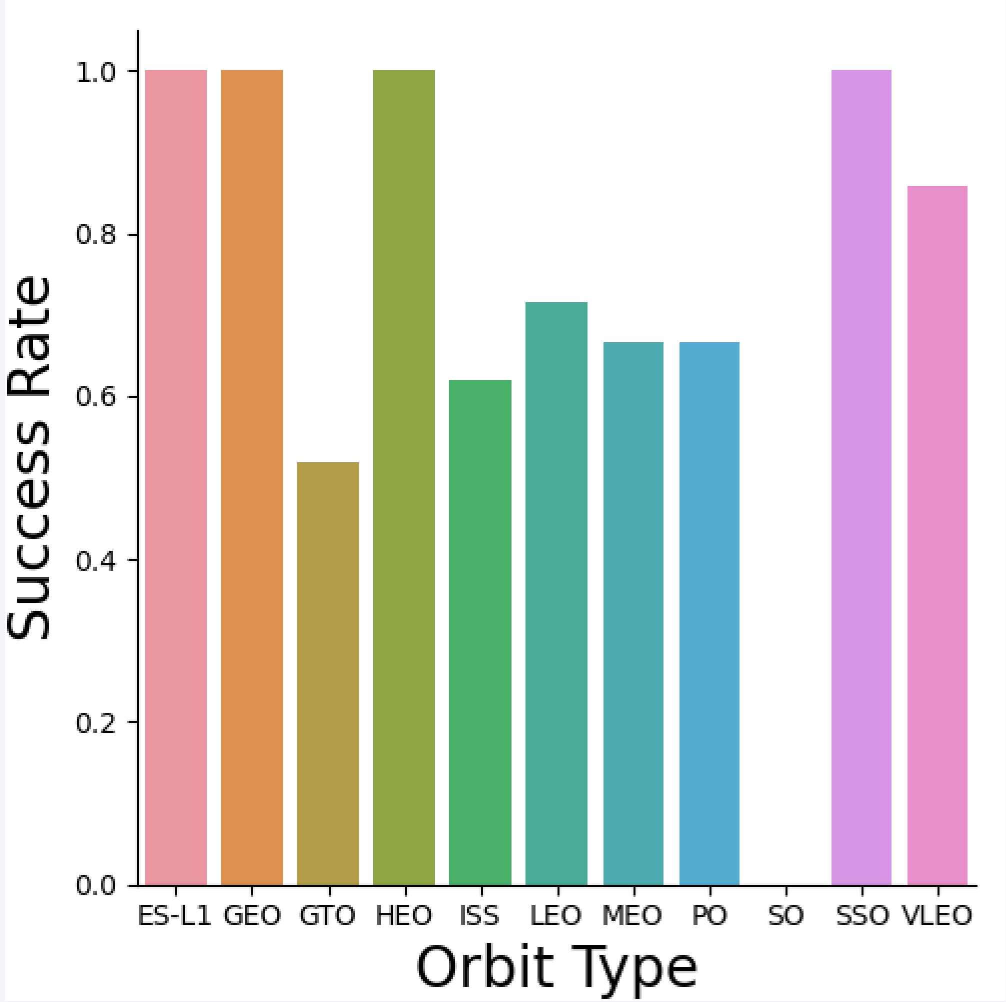
- First few missions failed consistently but most recent missions have been successful
- VAFB SLC 4E is being used less recently and has seen the least number of launches in general
- CCAFS SLC 40 has seen the most launches, accounting for its high failure rate

Payload vs. Launch Site



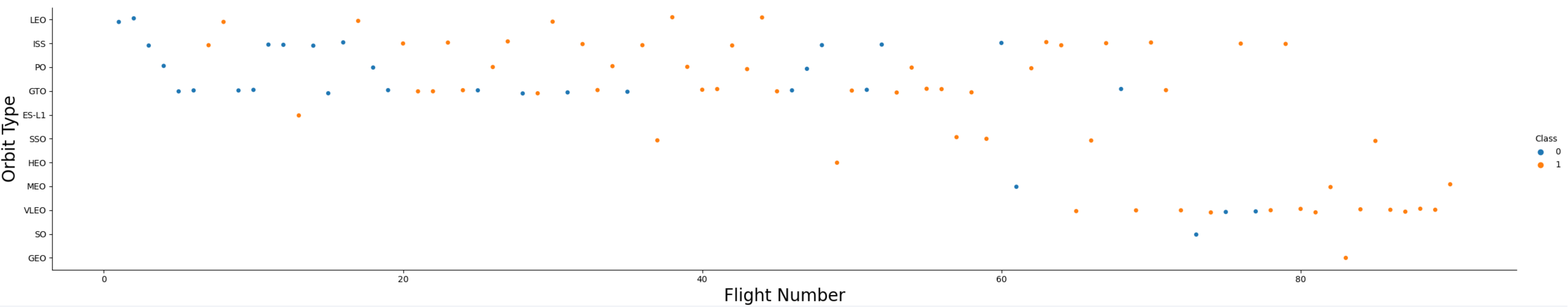
- Missions with payloads over 9000 kg are more likely to be successful
- Missions with payloads less than 6000 kg launched from KSC LC 39A are also highly successful
- VAFB SLC 4E does not seem to launch mission with payloads greater than 10000 kg

Success Rate vs. Orbit Type



- Missions with ES-L1, GEO, HEO, and SSO orbit types have 100% success rate
- Missions with SO orbit type have 0% success rate
- All other mission types fall somewhere in between but typically have success rate greater than 50%

Flight Number vs. Orbit Type

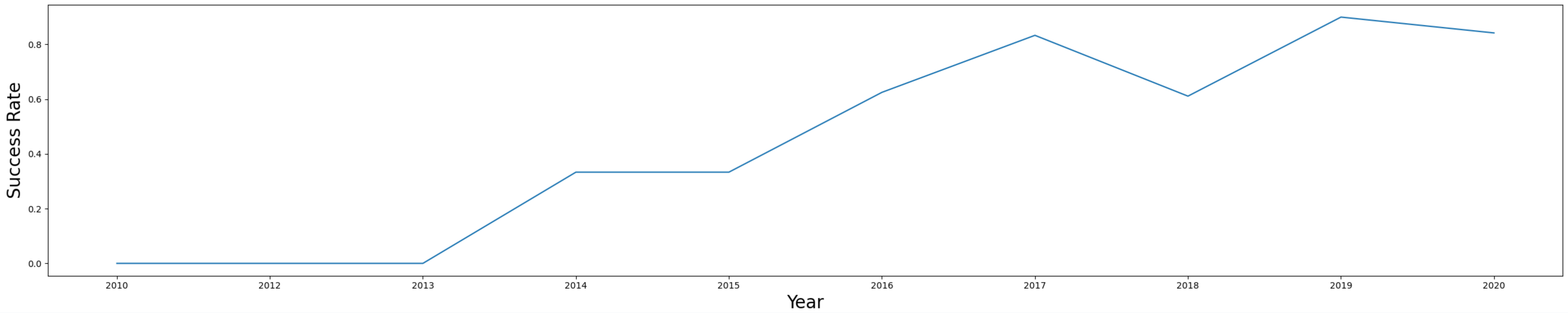


- Success rate appears to improve over time for LEO orbiting missions
- Most recent missions have been successful regardless of orbit type and most recent missions have been in VLEO orbit
- SO and GEO orbiting missions only have one failure and success respectively

Payload vs. Orbit Type

- VLEO orbiting missions have the largest payloads
- There is no clear relationship between payload and GTO orbiting missions
- LEO, ISS, and PO missions with heavier payloads are more successful

Launch Success Yearly Trend



- Success rate has been steadily increasing since 2010 with slight dips in 2017 and 2019
- Currently, success rate sits at around 80%

All Launch Site Names

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

- Select all distinct (unique) launch sites from dataset

Launch Site Names Beginning with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

- Select 5 full records of missions launched from sites that start with ‘CCA’

Total Payload Mass

Total_Payload_Mass
111268.0

- Select sum of all payload masses launched by NASA (‘CRS’ in payload description)

Average Payload Mass by F9 v1.1

Average_Payload_Mass
2534.6666666666665

- Select average payload mass of payloads carried by booster version F9 v1.1

First Successful Ground Landing Date

First_GroundPad_Landing
01/08/2018

- Select minimum (first) date where ground pad landing was success

Successful Drone Ship Landing with Payload Between 4000 and 6000 kg

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

- Select booster versions that successfully landed on drone ship and had payload mass between 4000 and 6000 kg

Total Number of Successful and Failure Mission Outcomes

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

- Select mission outcome counts grouped by outcome types

Boosters Carrying 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

- Select booster versions that have carried max payload mass

2015 Launch Records

Date	Month	Booster_Version	Launch_Site	Landing_Outcome
01/10/2015	10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
14/04/2015	04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- Select date, month, booster version, launch site, and mission outcome of all 2015 launches that failed to land on drone ship

Landing Outcomes Ranked Between 6/4/2010 and 3/20/2017

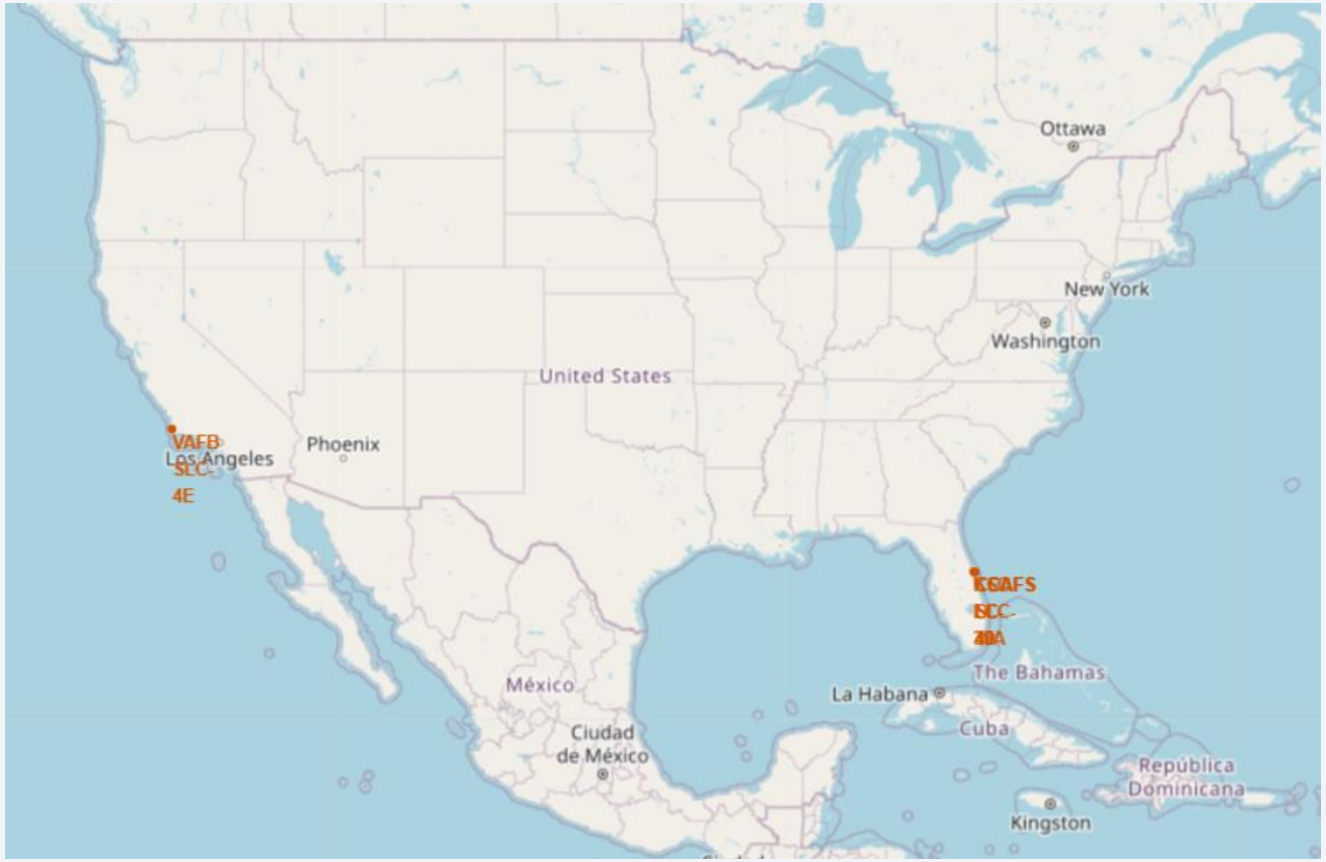
Landing_Outcome	Total
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

- Select and rank count of each mission outcome between the dates 6/4/2010 and 3/20/2017 in descending order

Section 3

Launch Sites Proximities Analysis

All Launch Sites Map



- Launch sites are located close to Equator and coastline to ensure proper rocket takeoff and prevent malfunction from impacting populace

VAFB SLC 4E Launch Outcome Map

- VAFB SLC 4E has a relatively low success rate given that it has more red (failure) markers than green (success) markers

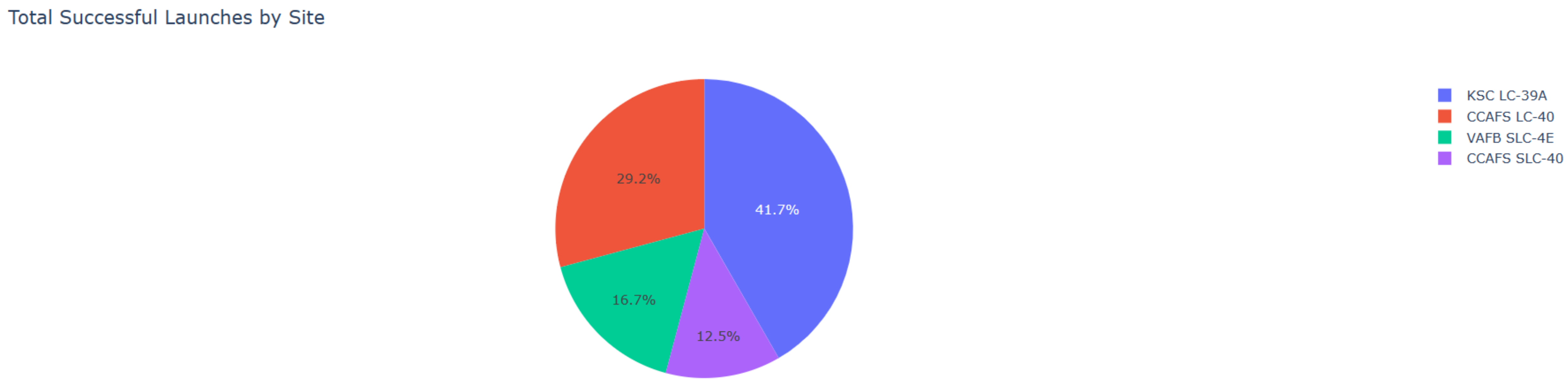
CCAFS SLC 40 Proximity Map

- CCAFS SLC 40 is relatively close to major points of interest as it is only 0.59 km away from major roadway and 0.86 km from coastline

Section 4

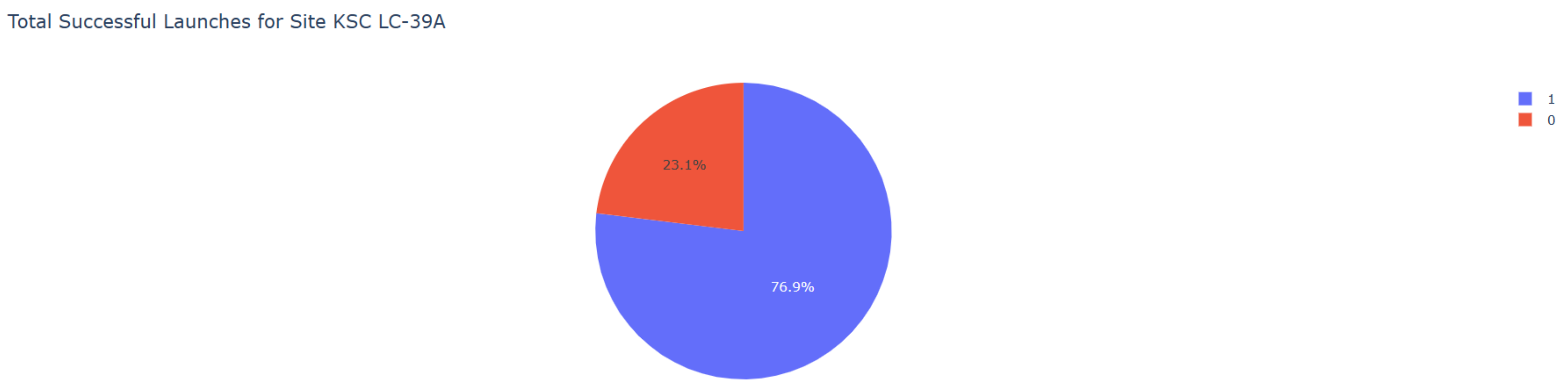
Build a Dashboard with Plotly Dash

Success Count for All Launch Sites



- KSC LC 39A is launch site with most successful launches

Launch Site with Highest Success Ratio



- KSC LC 39A has had 10 successful launches (76.9%) and only 3 failed launches (23.1%)

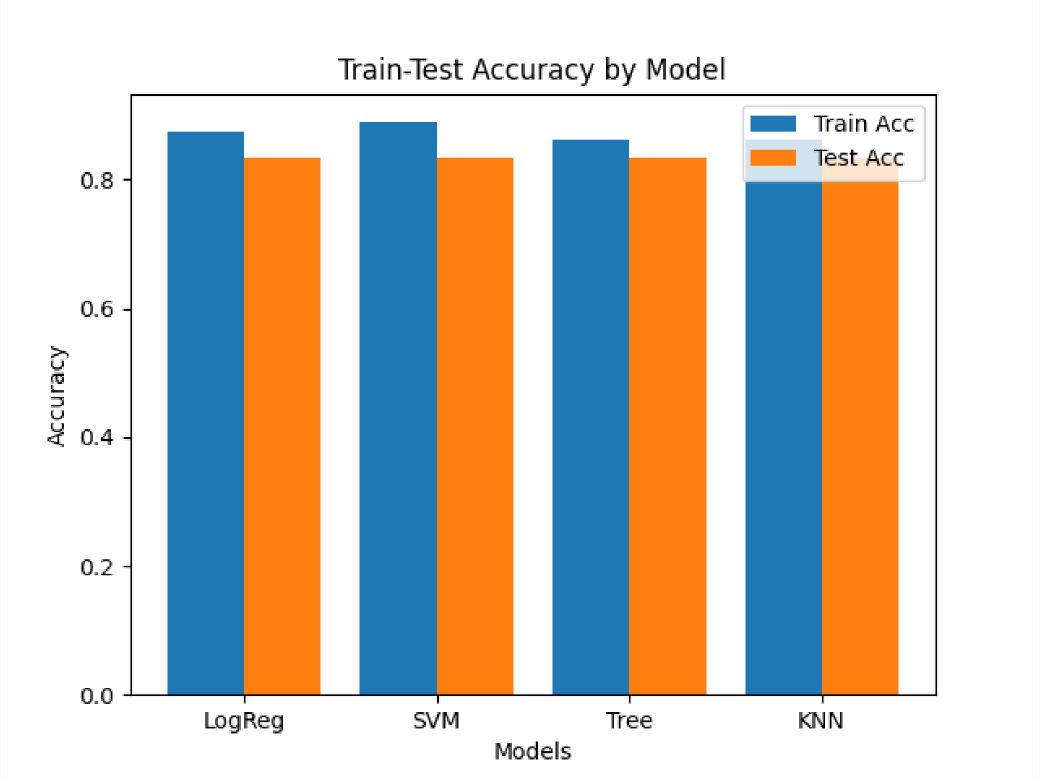
Payload Mass vs. Launch Outcome for All Launch Sites

- Payloads under 6000 kg have greater success along with payloads carried by FT boosters but not v1.1 boosters

Section 5

Predictive Analysis (Classification)

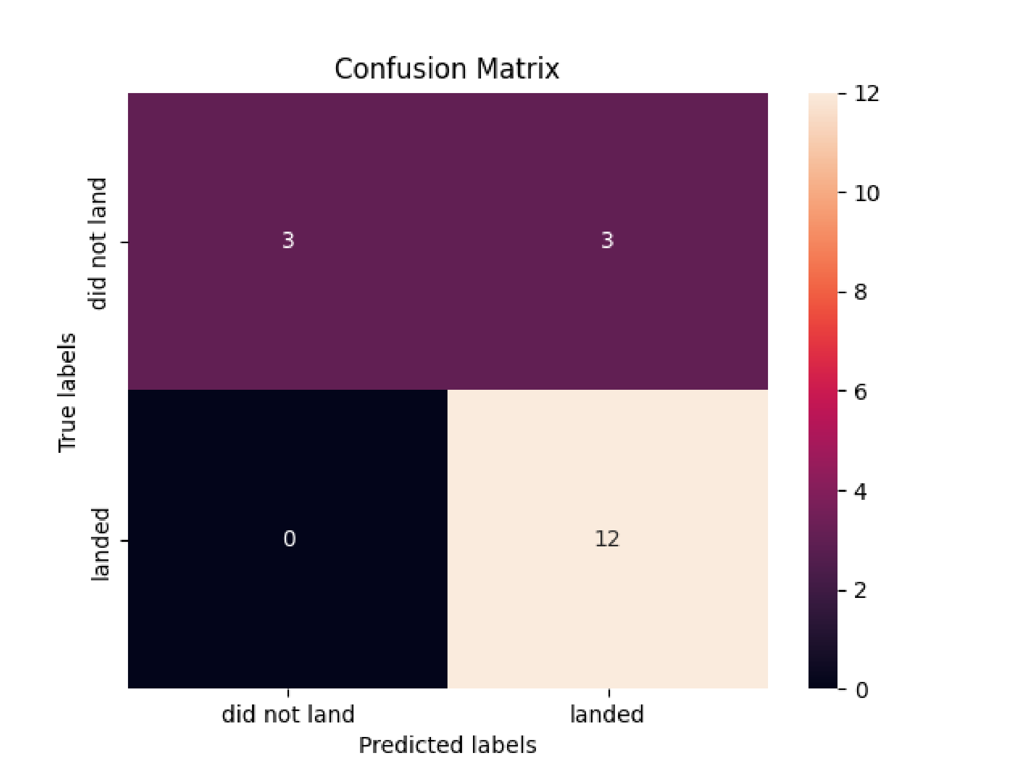
Classification Accuracy



- Same test accuracy across board oddly; SVM is best model based on tiebreaker of training accuracy
  - ~89% training accuracy and ~83% testing accuracy
  - Not ideal but impressive nonetheless given small dataset with decently large number of features

Confusion Matrix

- SVM Confusion Matrix
  - Algorithm struggles slightly with false positives but hard to correct given dataset size and class imbalance



Conclusions

- Landing success rates improved over time likely due to process and rocket refinement
- KSC LC 39A appears to be the most successful launch site
- Missions with an orbit type of ES-L1, HEO, or SSO have all been successful thus far
- Missions with heavier payloads tend to have better landing outcomes but the impact of payload mass on launch outcomes is harder to generalize
- The support vector machine algorithm can predict successful landings adequately but further testing and tuning is required to optimize model predictive power and, thus, maximize profits

Thank you!