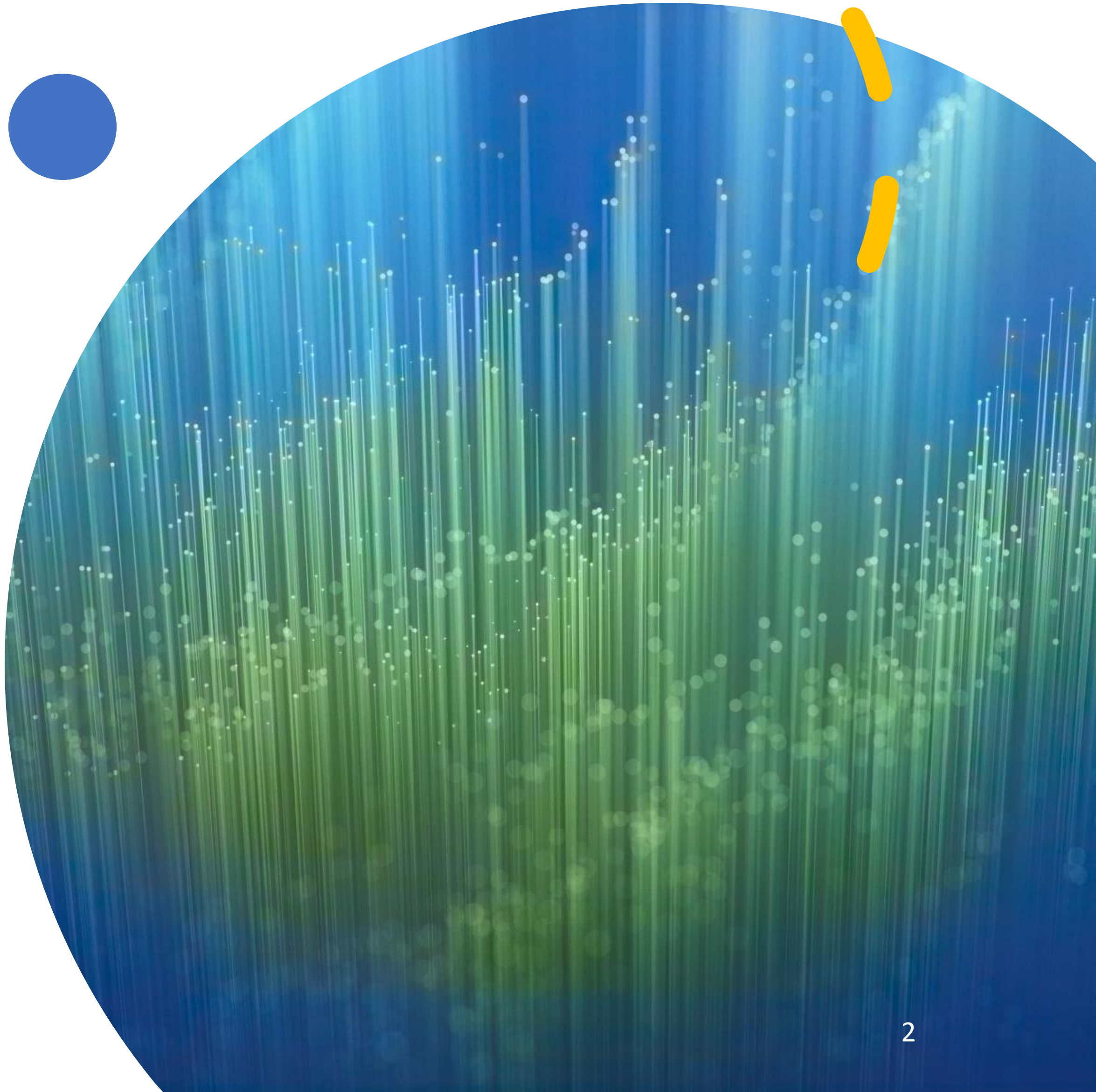
IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Michelle Z.
August 26, 2023

Outline



Executive Summary

- Methodologies: data is collected from SpaceX REST API and webscraping the Falcon 9 Launch data
  - Data is explored through SQL, visualization, and folium maps
- Logistic Regression, Support Vector Machine, and K-Nearest Neighbors machine learning models were produced. All performed the same, with a score of 0.8334

Introduction

Space Y seeks to compete with Space X

Determine price of each launch, if Space X will reuse the first stage

Done by training a machine learning model

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium
- Perform predictive analysis using classification models

Data Collection

Request data from SpaceX API

Convert URL into a Pandas dataframe with .json_normalize()

Combine columns of dataset into a dictionary

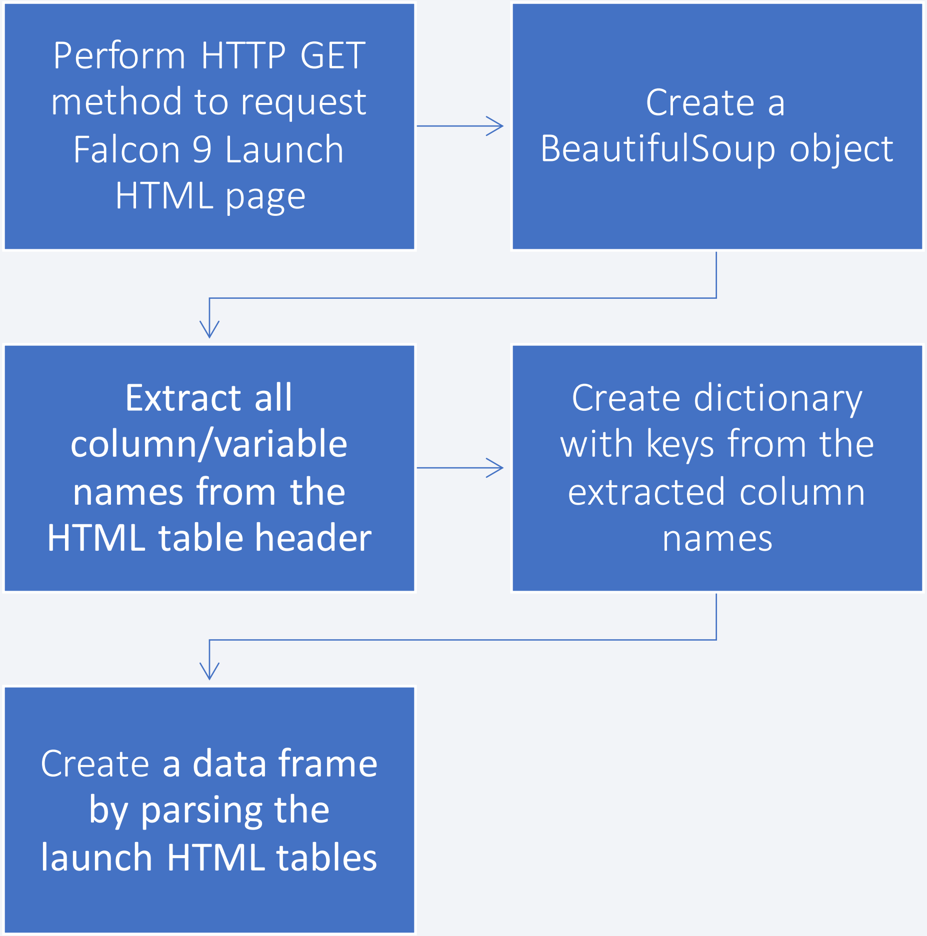
Filter dataframe to include only Falcon 9 launches

Place missing values in the dataset

Github URL

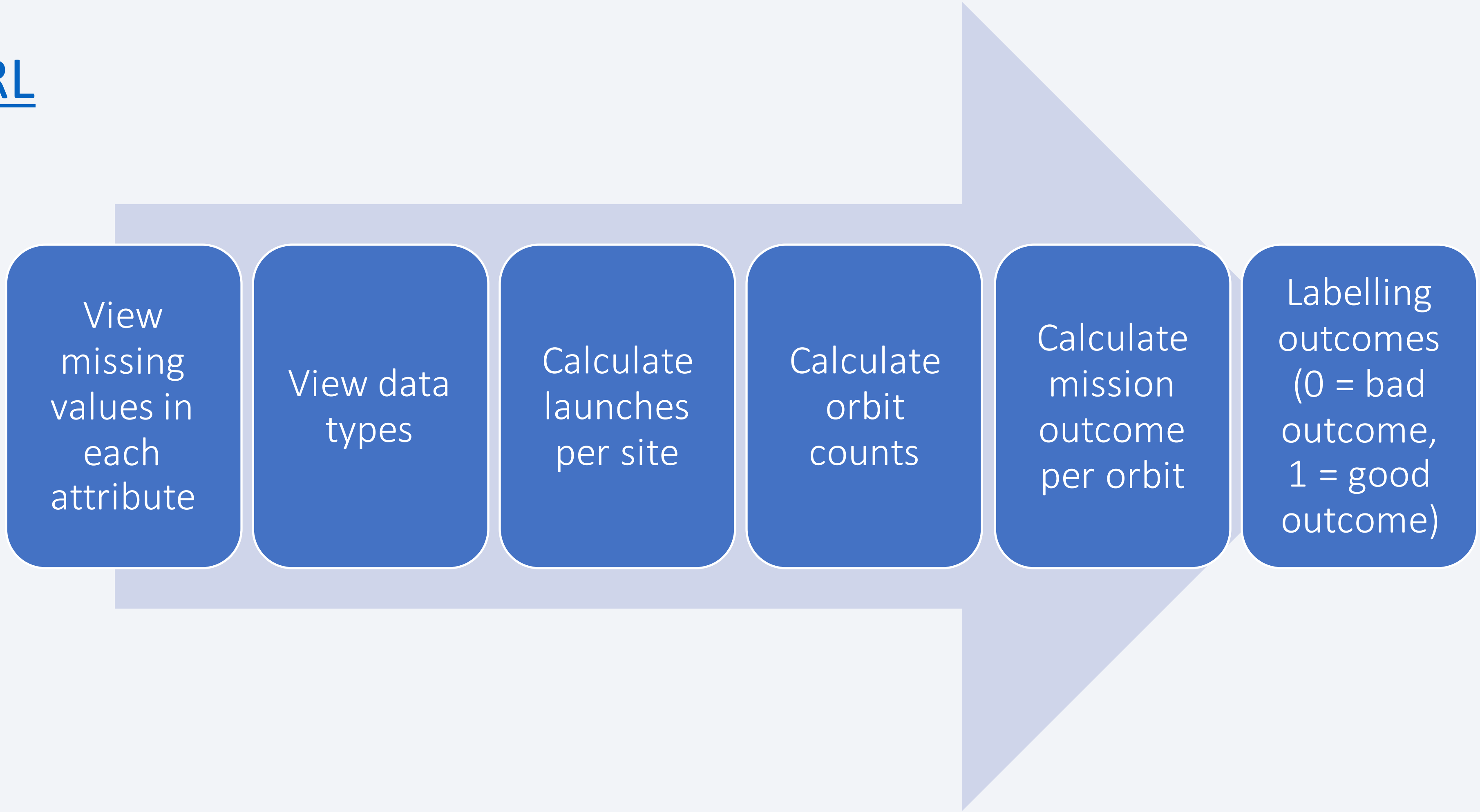
Data Collection - Scraping

Github URL



Data Wrangling

- Github URL



EDA with Data Visualization

- Charts comparing Payload Mass and Flight Number, Launch Site and Flight Number, Orbit types and success rates, Orbit types and Flight Number, and Orbit types and Payload were made
  - Information regarding Space X's decisions could be deduced through observing patterns seen in the visualizations
- A variety of charts were produced using Seaborn, including scatterplots, bar charts, and line charts
- Grouping and filtering the dataset was performed when necessary
- Github URL

EDA with SQL

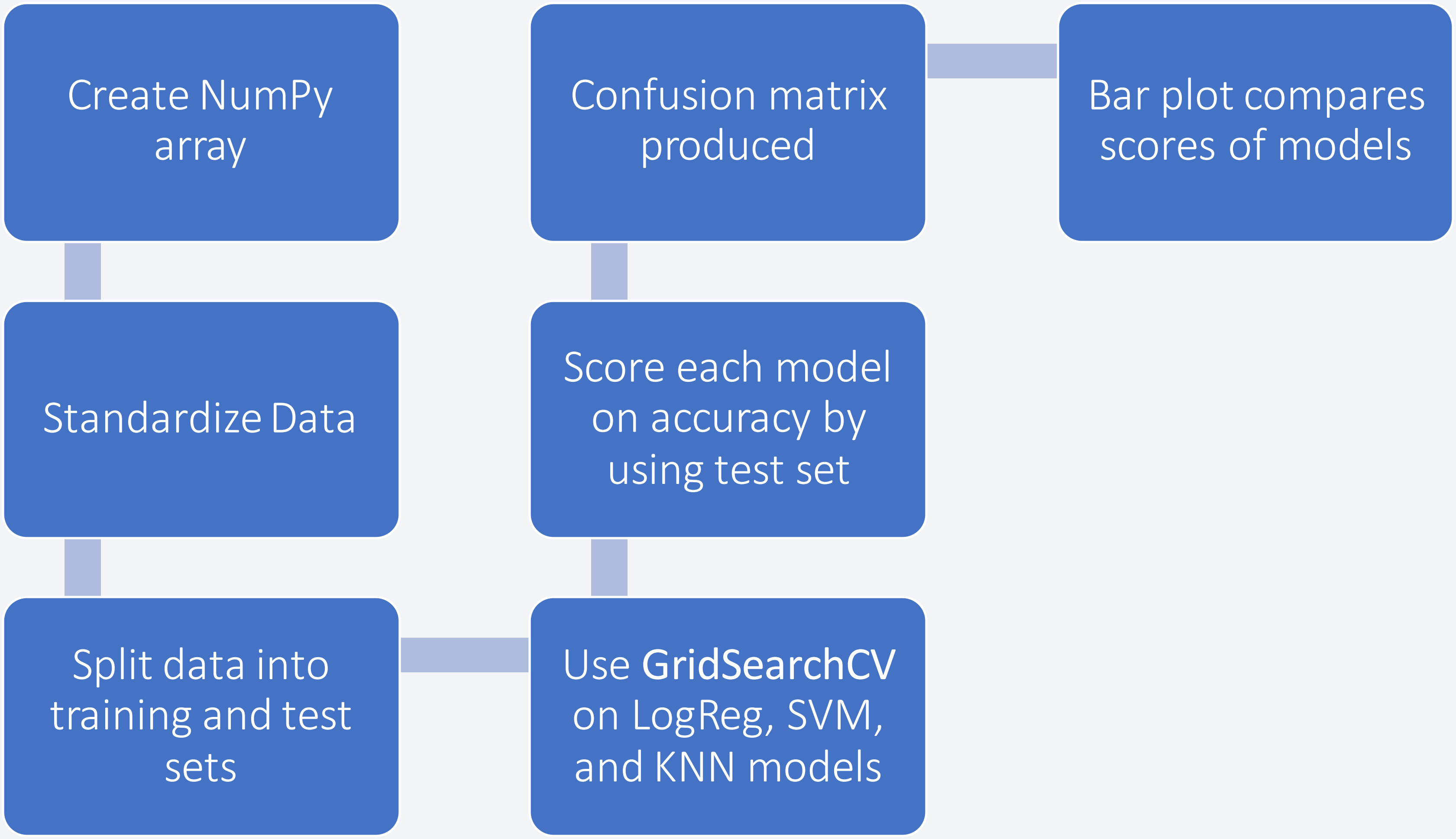
- The dataset was loaded into a Db2 database and SQL queries were performed to perform tasks such as:
  - Finding launch site names
  - Display average payload mass for certain boosters
  - Find counts for successful and failed missions
  - Display total payload mass
- Github URL

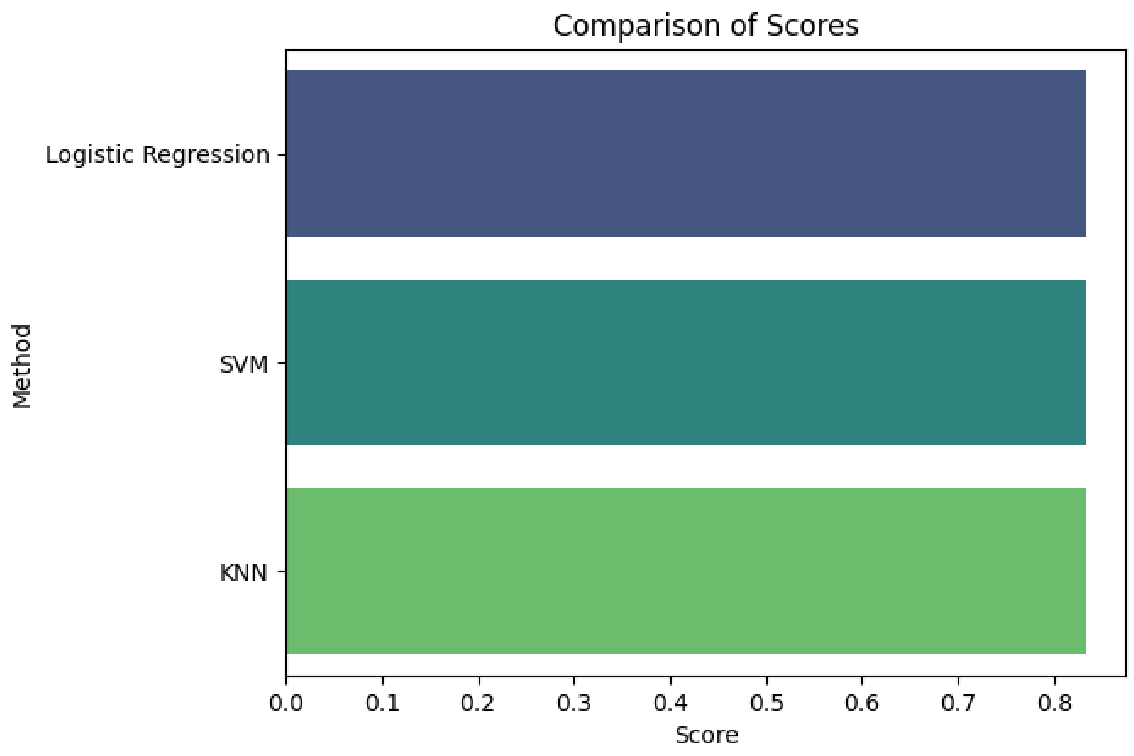
Build an Interactive Map with Folium

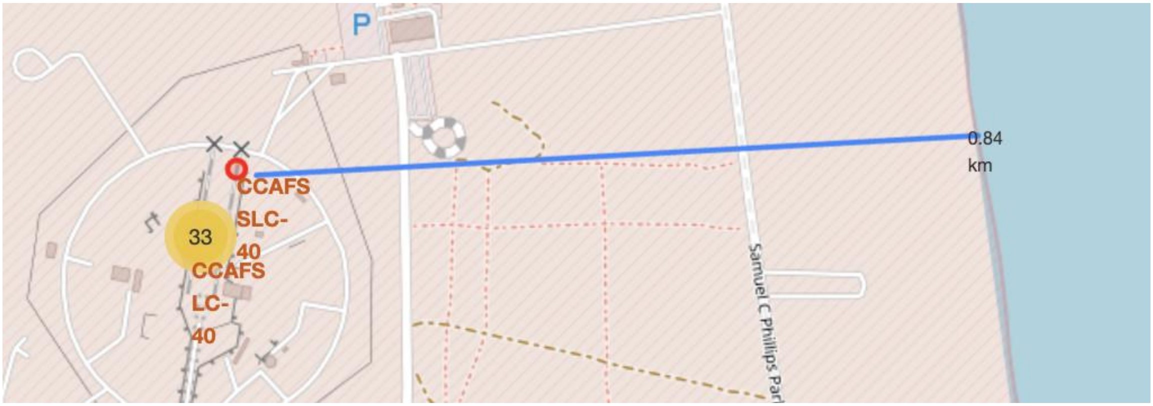
- A map with launch sites labeled by name was created
- Marker cluster was produced in order to return counts and color code success (0/1) at launch sites (could be seen if zoomed in on map)
- A distance calculation was made to demonstrate the proximity of launch sites to coastlines
- Github URL

Predictive Analysis (Classification)

Github URL







Results

- Refer to section 2.1 and 2.2 for exploratory data analysis results
- Folium map demonstrated proximity of launch sites to coastlines, possibly to avoid danger of harming people in cities if failed mission
- All models had the same accuracy score of 0.833

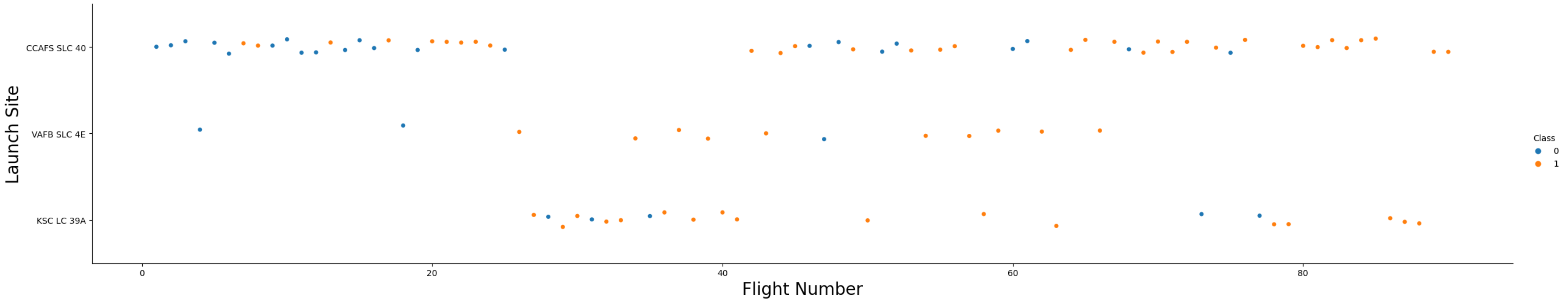
Section 2

Insights drawn from EDA

Section 2.1

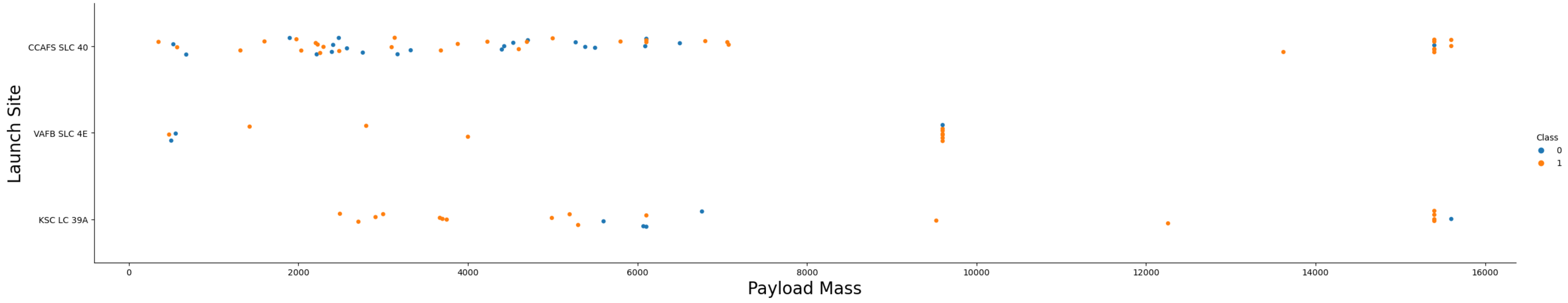
EDA with Visualization

Using Seaborn plots



Flight Number vs. Launch Site

- Blue indicates an unsuccessful launch, Orange indicates a successful launch.
- There is an increase in success over time; the ratio of blue points to orange points prior to flight 30 is greater than after

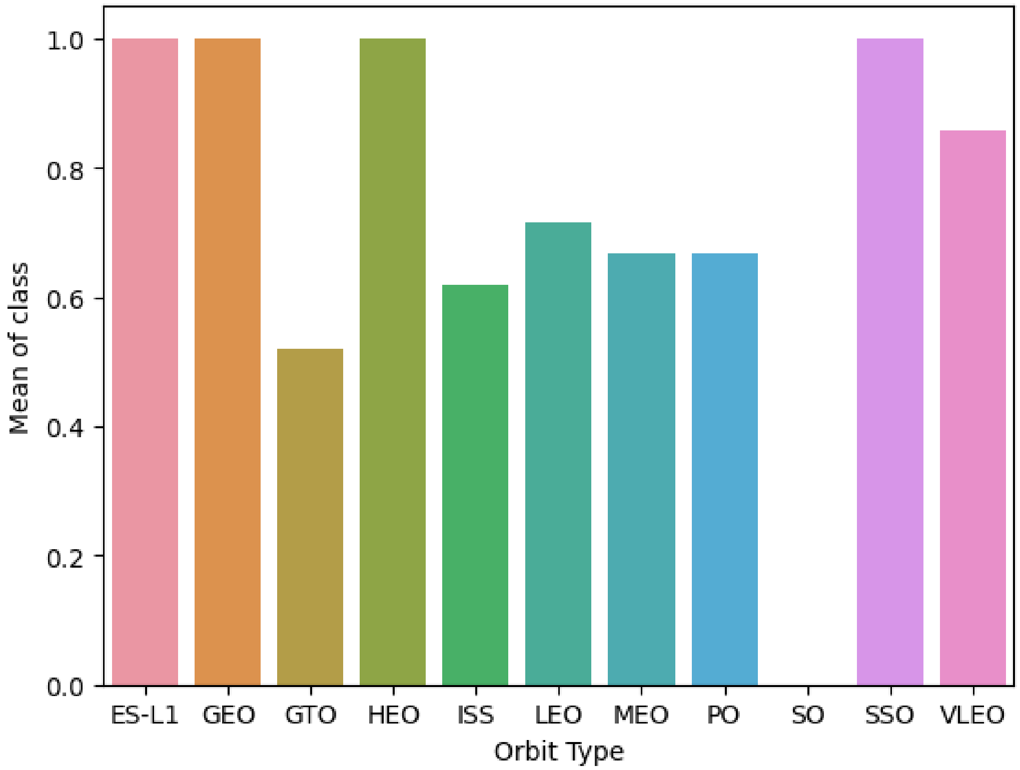


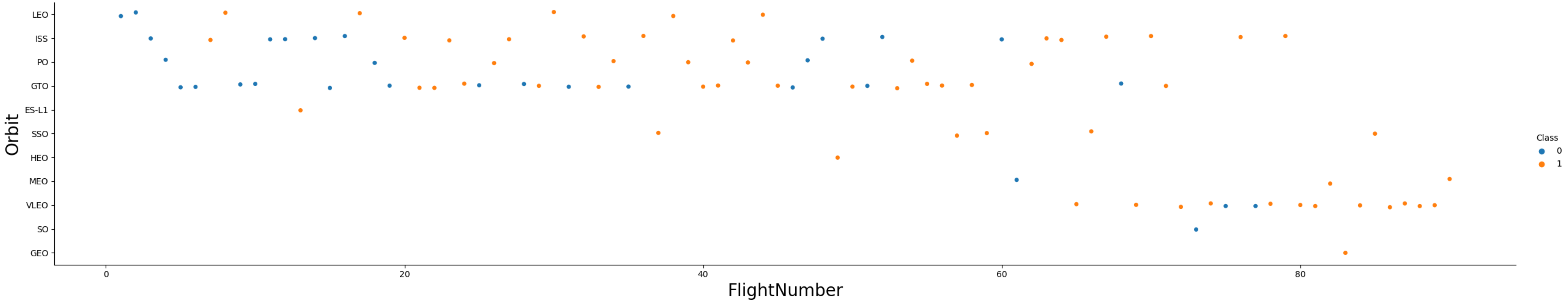
Payload vs. Launch Site

Most payload seems to be less than 6000 kg

Success Rate vs. Orbit Type

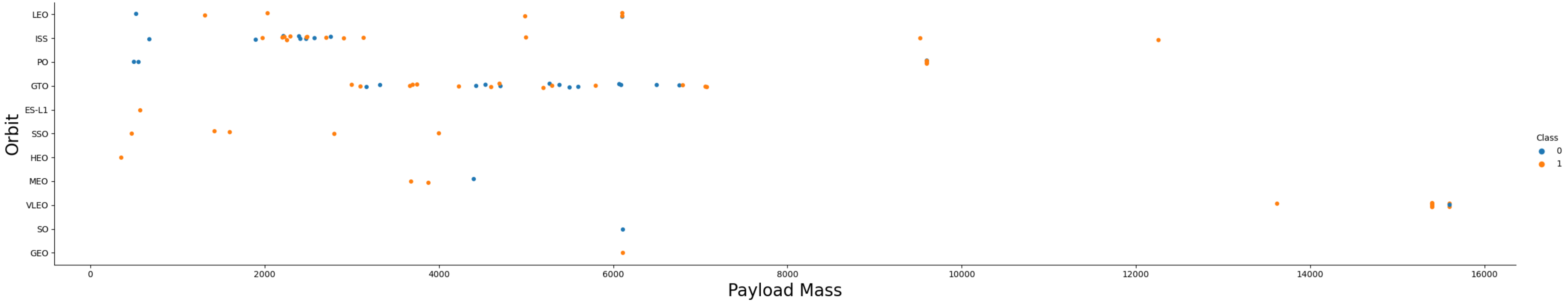
- SO had the lowest success rate (0%)
- Four orbit types had a 100% success rate (ES-L1, GEO, HEO, SSO)
- The remaining orbit types had a success rate from 50-70%





Flight Number vs. Orbit Type

- After 60 flights, there are more flights with VLEO orbit type. Prior to 60 flights, most orbit types are LEO, ISS, or GTO.

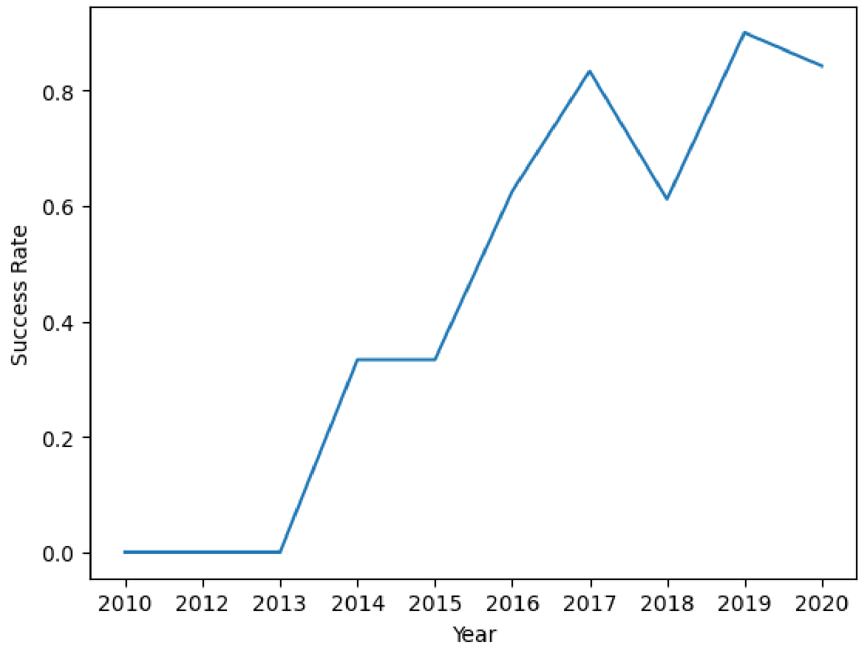


Payload vs. Orbit Type

- SSO is the most successful orbit, and its payload is always under 6000 kg. VLEO payloads are always above 12000 kg.
- GTO was attempted with a wide range of payloads

Launch Success Yearly Trend

- Launches were not successful from 2010-2013 (first three years)
- Launch success has generally improved over time, however there was a plateau from 2014 to 2016, and a drop in 2018



Section 2.2

EDA with SQL

All Launch Site Names

- Four unique launch sites: CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLAC-40

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

Launch Site Names Begin with 'CCA'

- Displayed the first 5 Launch Site Names that begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
2010-08-12	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
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

total_payload_mass
45596

Total Payload Mass

- The total payload carried by boosters from NASA is 45596 kg

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 2534.67 kg
- This is on the lower side of payload masses

Average
2534.6666666666665

First Successful Ground Landing Date

Date
2010-04-06

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

Successful Drone Ship Landing with Payload between 4000 and 6000

- There are four boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

- An overwhelming majority of mission outcomes are successful
- Thus, failed landings are intentional

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

boosterversion
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

Boosters Carried Maximum Payload

- 13 boosters, all of similar F9 B5 B104x.x type, carried the maximum payload

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- The landing outcome that occurred the most was no attempt at landing. There were two types of failures, by drone ship and by parachute.

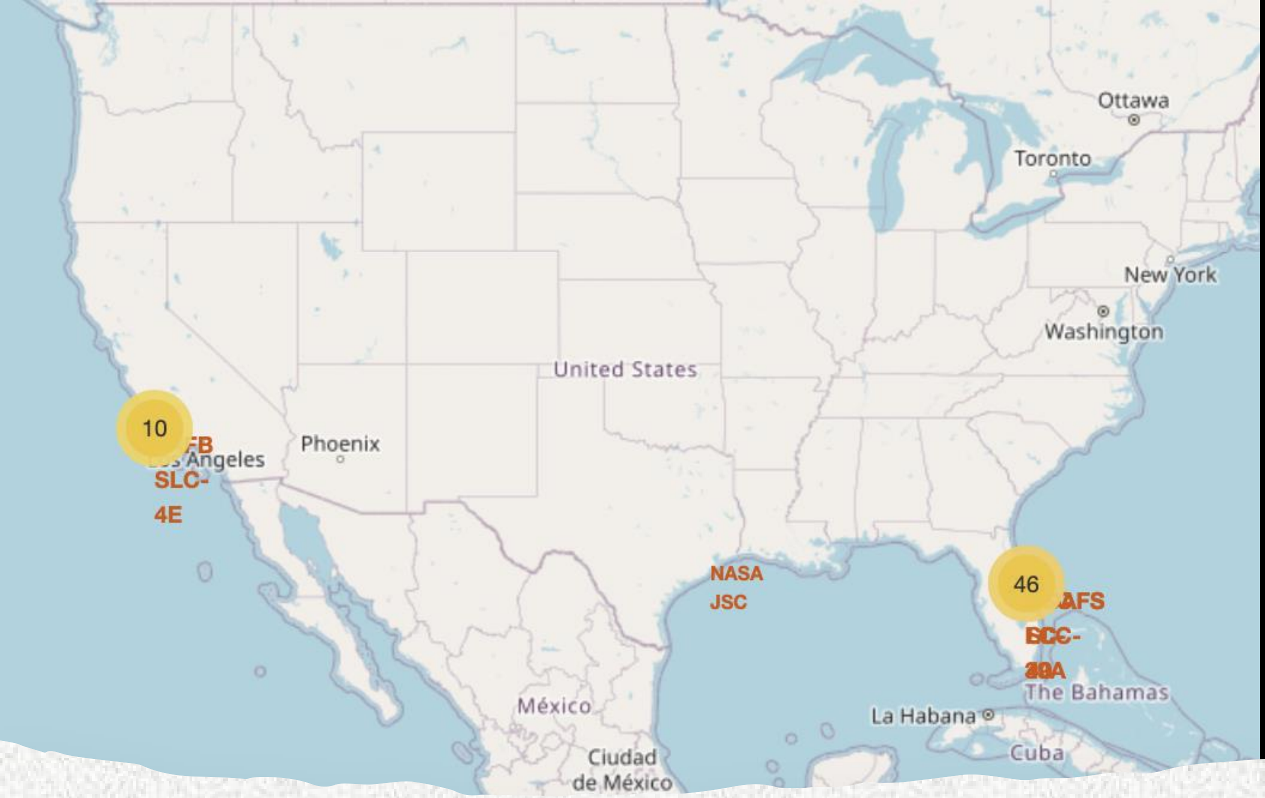
Landing_Outcome	count
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

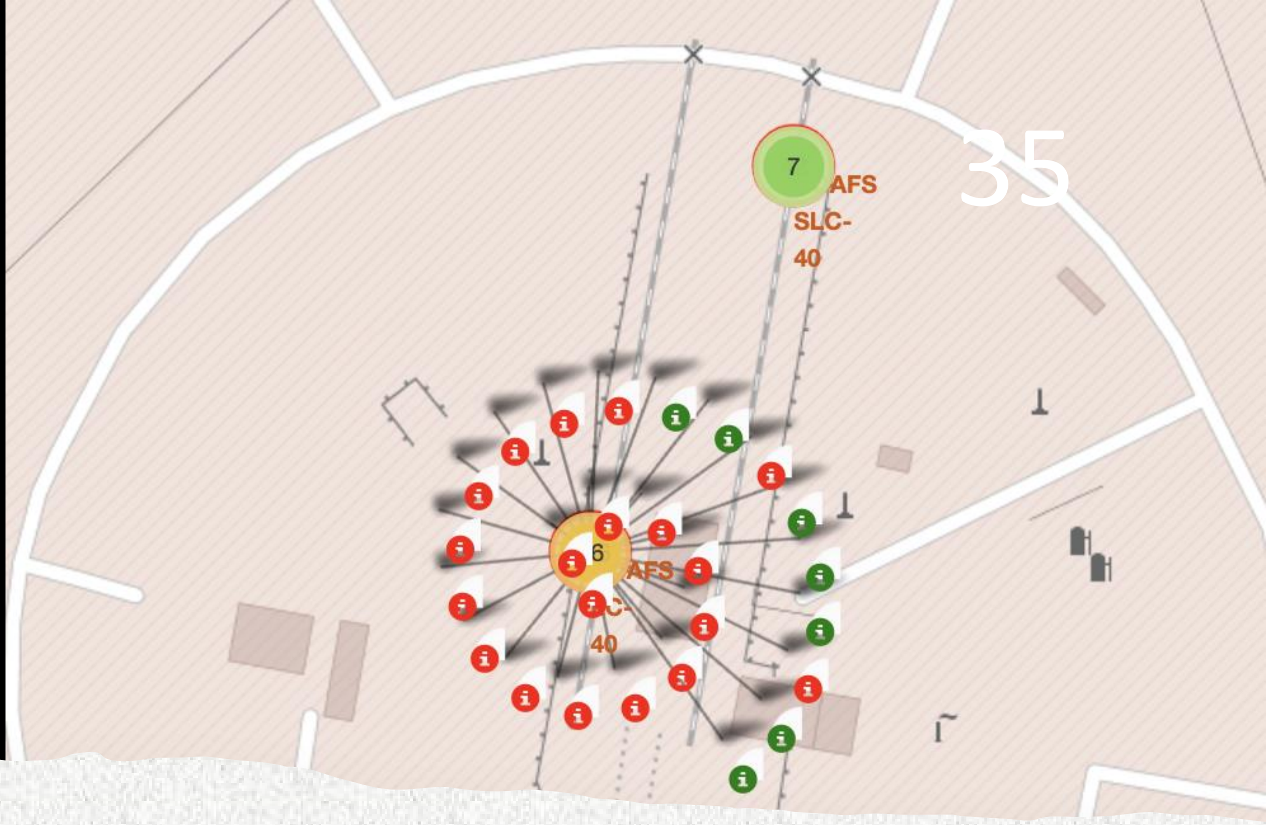
Section 3

Launch Sites Proximities Analysis

Launch site locations

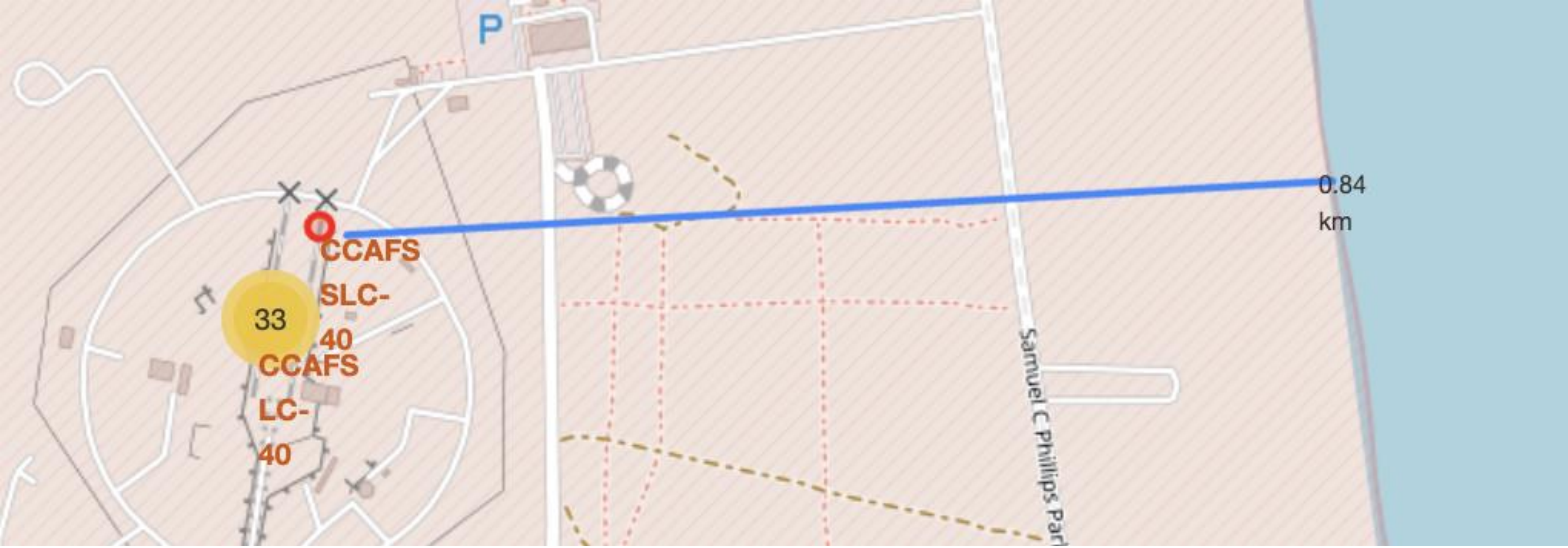
- Launch sites are concentrated on the east and west of the US, specifically in the states of Florida and California.
- These launch sites are close to large bodies of water
- There are two launch sites in Florida, very close to each other





Launch site counts and clusters

- The number of launches per launch site is displayed
- Upon zooming in further to the map, the clusters of launches can be seen
- Colors (red/green) used to differentiate between successful and unsuccessful landings

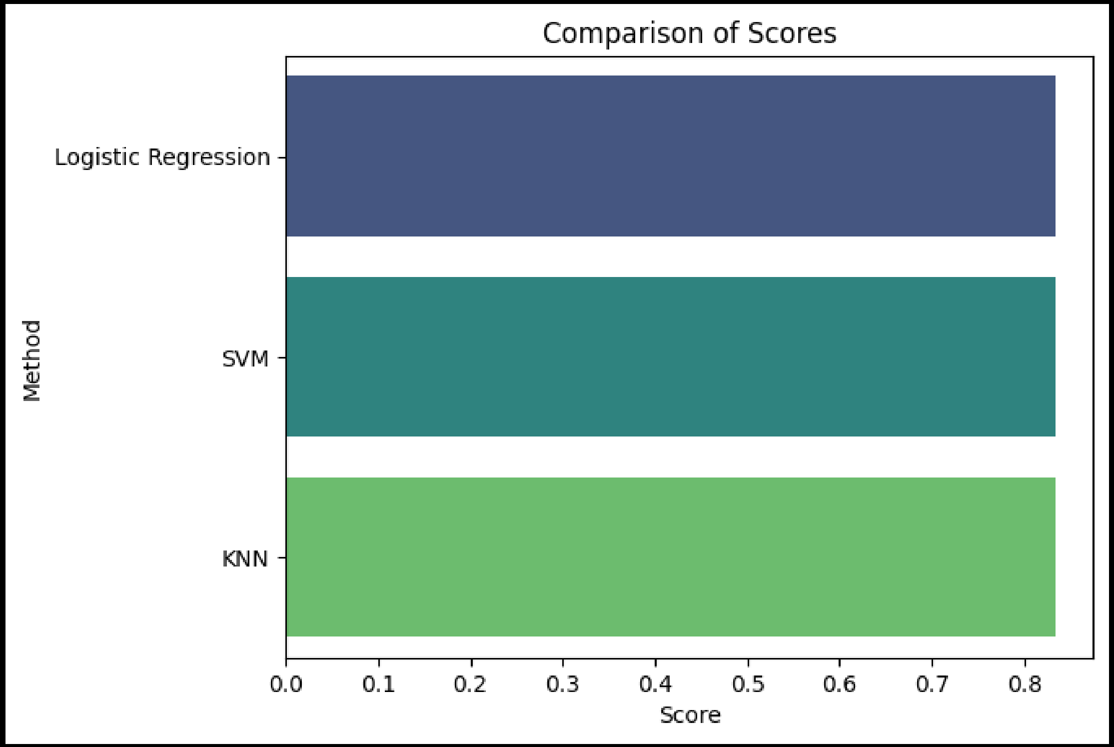


Distance between chosen launch site and coast line

- The distance from CCAFS SLC-40 to the eastern Florida coastline is 0.84 km
- Launch sites tended to be close to coasts; this makes sense as launches can fall into water instead of causing damage to land.

Section 4

Predictive Analysis (Classification)

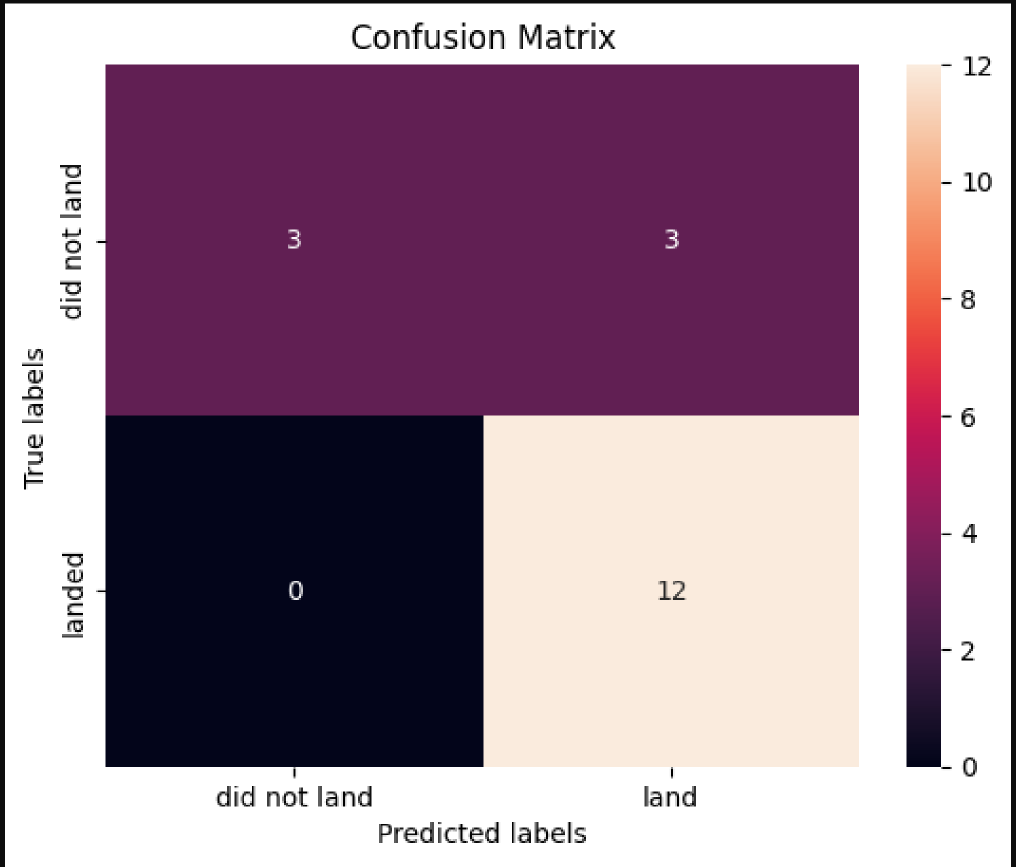


Classification Accuracy

- All models perform at the same classification accuracy of 0.833

Confusion Matrix

- Because all models performed the same, they all produced the same confusion matrix



Conclusions

- Data collected from SpaceX REST API and webscraping Falcon 9 Launch data from wikipedia
- Data explored with SQL and visualization is produced to compare and deduce information; found that missions rarely 'failed', so failed landings are intentional
- Folium maps developed and demonstrate how launch sites are clustered in Florida and California, near the coast (distance calculation demonstrated a <1 km distance from the coast)
- Logistic regression, support vector machine, and k-nearest neighbor machine learning models developed, scored for accuracy from test set of data, and confusion matrix produced. All models resulted in the same accuracy of 0.833
- Thus, the models developed would all be appropriate for use by Space Y, but more data (or a larger sample size) could be collected in order to result in a higher score

Appendix

Acknowledgements to IBM Applied Data Science Capstone staff for developing base material:

- **Primary Instructors**
- Joseph Santarcangelo
- Yan Luo
- **Other Contributors & Staff**
- **Project Lead:** Rav Ahuja **Instructional Designer:** Lakshmi Holla **Lab Authors:** Joseph Santarcangelo, Yan Luo, Azim Hirjani, Lakshmi Holla **Technical Advisor:** Yan Luo
- **Production Team**
- **Publishing:** Grace Barker, Rachael Jones **Project Coordinators:** Kathleen Bergner **Narration:** Bella West **Video Production:** Simer Preet, Lauren Hall, Hunter Bay, Tanya Singh, Om Singh
- **Teaching Assistants and Forum Moderators**
- Malika Singla
- Duvvana Mrutyunjaya Naidu
- Lakshmi Holla
- Anita Verma

Thank you!