Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Scrape the Data from the SpaceX website with Requests
- Clean the Data using Pandas
- Use SQL to explore the Data
- Visualize the Data using Matplotlib and Seaborn
- Visualize a Map of the Data with Folium
- Create a Dashboard of the Data with Dash and Plotly
- Build Machine Learning Model to predict successful landing of rocket with Scikit-Learn

Introduction

- Have data on 90 rocket launches, plus whether they successfully landed or not
- Given the features, can we build a machine learning model to predict if a rocket will successfully land or not

Section 1

Methodology

Methodology

Executive Summary

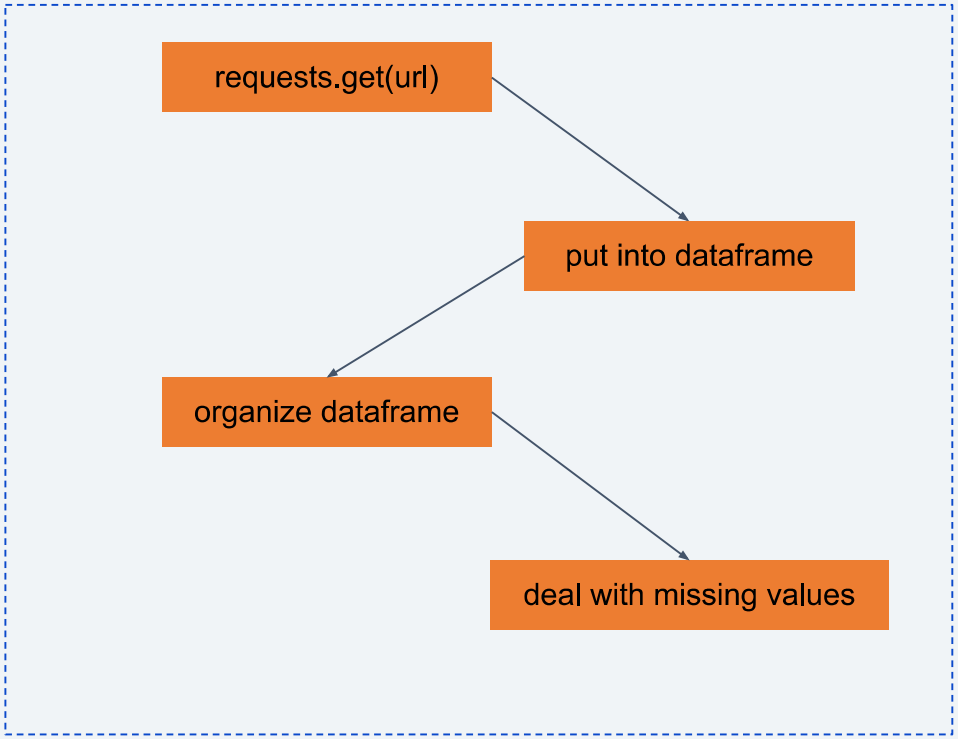
- Data collection methodology:
  - Data collected from SpaceX website using Requests API
- Perform data wrangling
  - Cleaned out missing data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Used Scikit-learn API and compared many different models to find most accurate machine learning model

Data Collection

- Used the requests API
- Get SpaceX website
- Convert to JSON and then to dataframe using pandas
- Create new and more organized dataframe from this dataframe
    - more informative column names
    - keep important columns
- Filter dataframe for only Falcon 9 launches
- Clean Data
    - Fill in missing numerical data with column average
- Added a 0/1s column for landing success (0 for failed landing, 1 for successful landing)

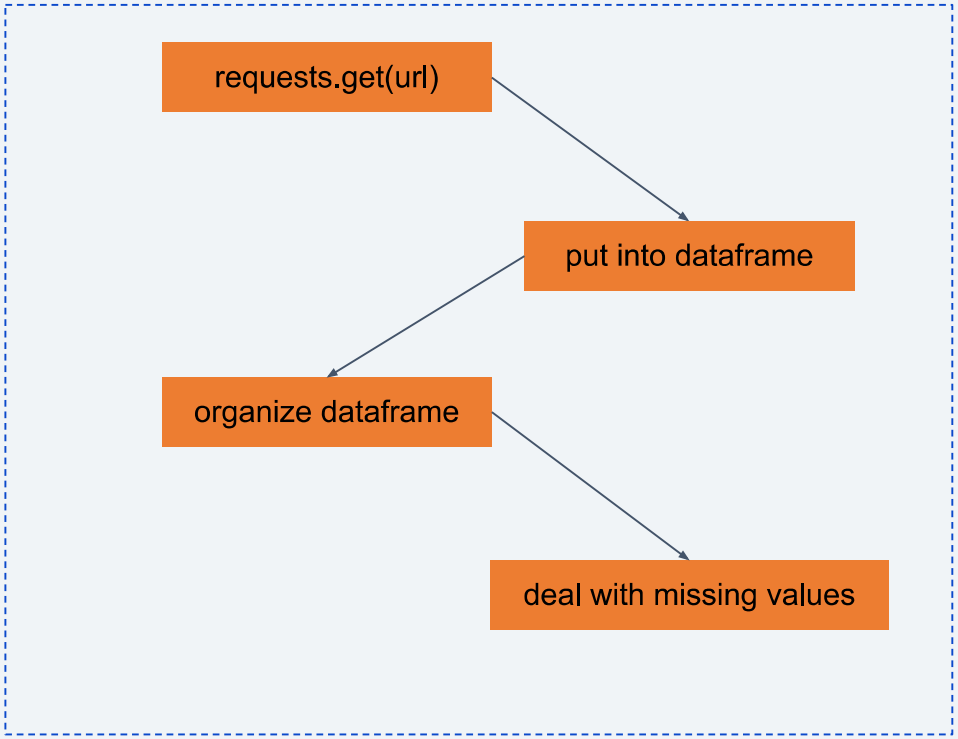
Data Collection – SpaceX API

Notebook



Data Collection – Scraping

Notebook



Data Wrangling

- Notebook
- Added extra column to dataframe
  - All negative outcomes were labelled with a 0
  - All positive outcomes were labelled with a 1

EDA with Data Visualization

- Data Visualization Notebook
- Flight number vs Payload
- Flight number vs Launch Site
- Payload vs Launch Site
- Success Rate of each Orbit Type
- Flight number vs Orbit Type
- Payload vs Orbit Type
- Launch success yearly trend

Also applied OneHotEncoding to the categorical features

EDA with SQL

- SQL Notebook
- Unique launch sites
- Launch Sites starting with 'CCA'
- Total Payload for boosters launched by NASA (CRS)
- Average Payload for booster version F9 v1.1
- First Successful Landing
- Successful landings where Payload between 4000kg and 6000kg
- Total number of successful and failed landings
- Months with failed landings in 2015
- Ranked the count of different types of successful landings between 04-06-2010 and 20-03-2017

Build an Interactive Map with Folium

- Folium Map
- Circle object creates a circle on the map at a certain location
    - can have a popup addition when mouse hovers over
- Marker object creates a marker on the map
- Marker cluster is a collection of markers
    - can be added to the map first, and then add markers to the marker cluster

Build a Dashboard with Plotly Dash

- Plotly Dash Dashboard Notebook
- A dropdown menu to select launch site, or select all launch sites
- A slider to select range of payload considered; default 0kg to 10,000kg
- A pie chart that displays the success vs failure percentages for the selected launch site
- A scatter plot that displays payload vs success/failure for the selected launch site and payload range
    - data points color coded to display booster type

Predictive Analysis (Classification)

- Classification Notebook
- Ran Logistic Regression, SVM, Decision Tree, and KNN to predict success/failure of landing given features
- Determined best model
    - Logistic Regression, SVM, and KNN tied for the best model

Results

- Exploratory data analysis results
  - There are 4 Launch Sites
  - Landing success rate increases with time
- Predictive analysis results
  - Logistic Regression, SVM, KNN give the most accurate models

Dashboard on next slide

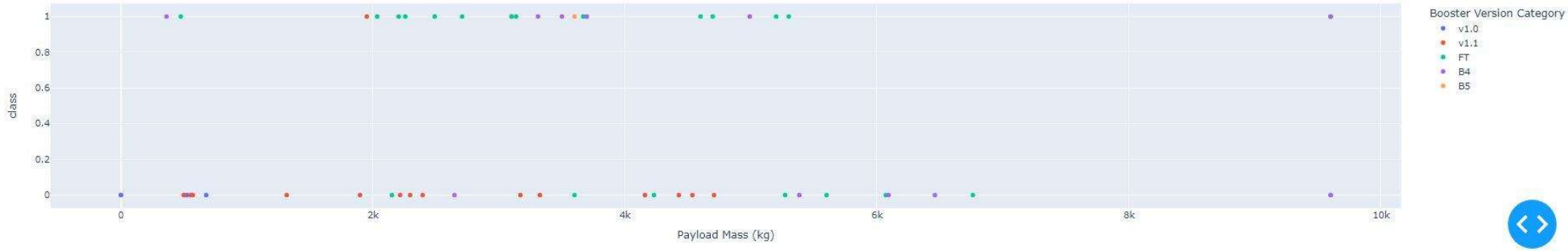
All Sites

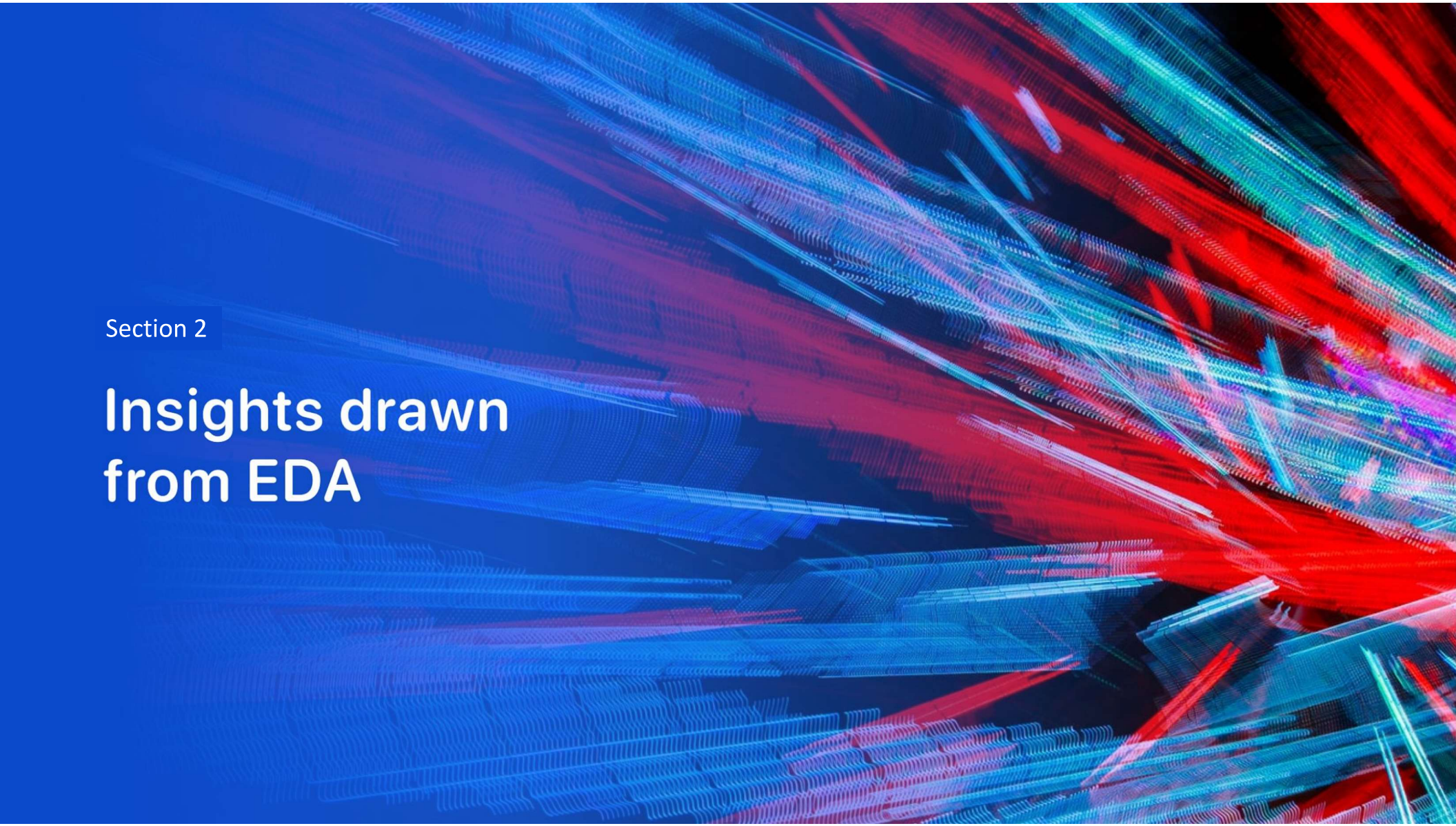
% of Successful Landings from Each Launch Site

Payload range (Kg):

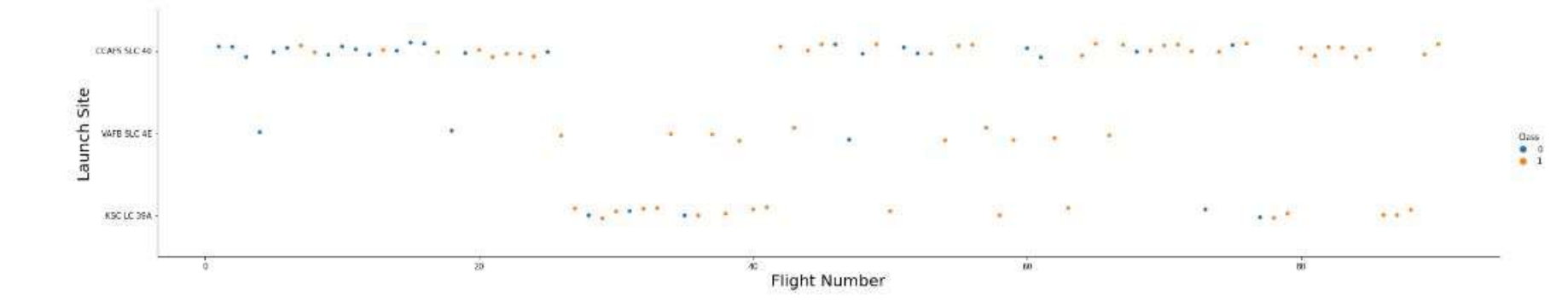
0 100

Class vs Payload Mass



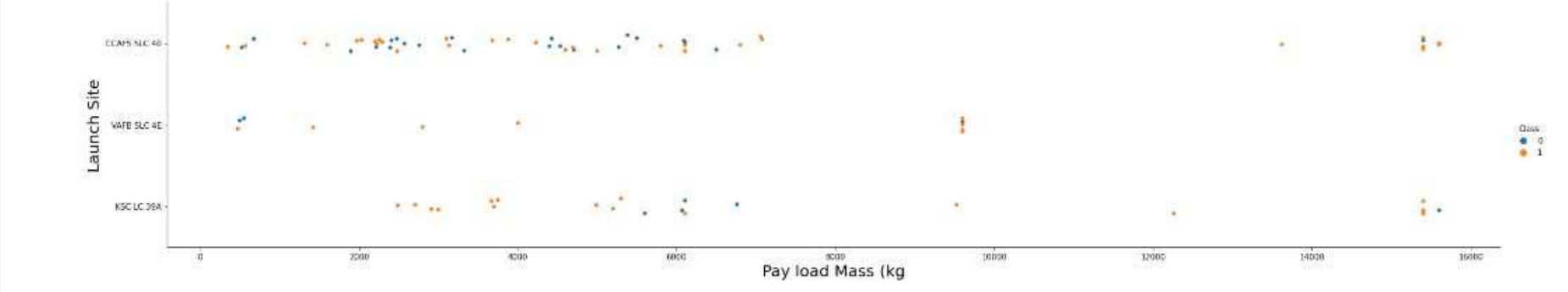

Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site



- KSC LC 39A had later flight numbers
- VAFB SLC 4E had earlier flight numbers
- CCAFS SLC 40 had early and later flight numbers
  - with a hiatus in the middle

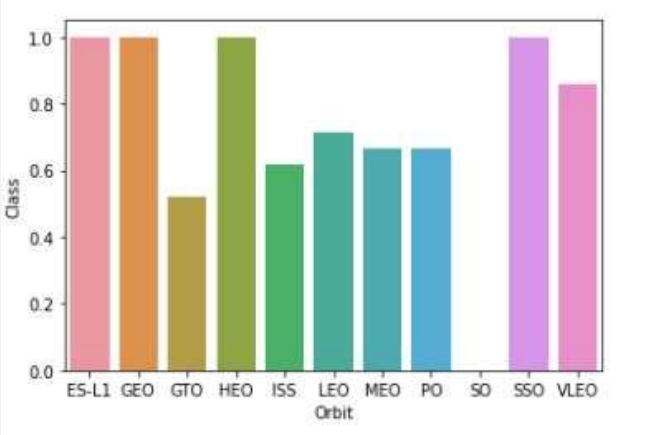
Payload vs. Launch Site



- CCAFS SLC 40 and KSC LC 39A had the outliers with heavy payloads
- VAFB SLC 4E had the lowest average payload

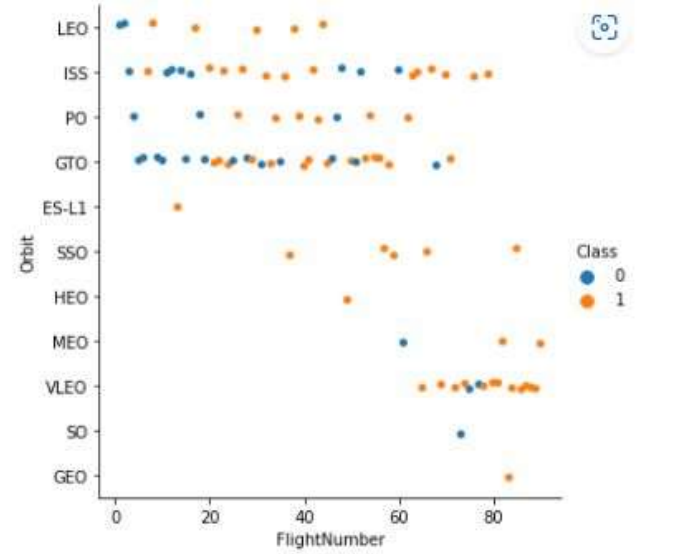
Success Rate vs. Orbit Type

- 4 orbit types had a 100% success rate for landing
- The orbit types with the lowest success rates are SO, GTO, and then ISS



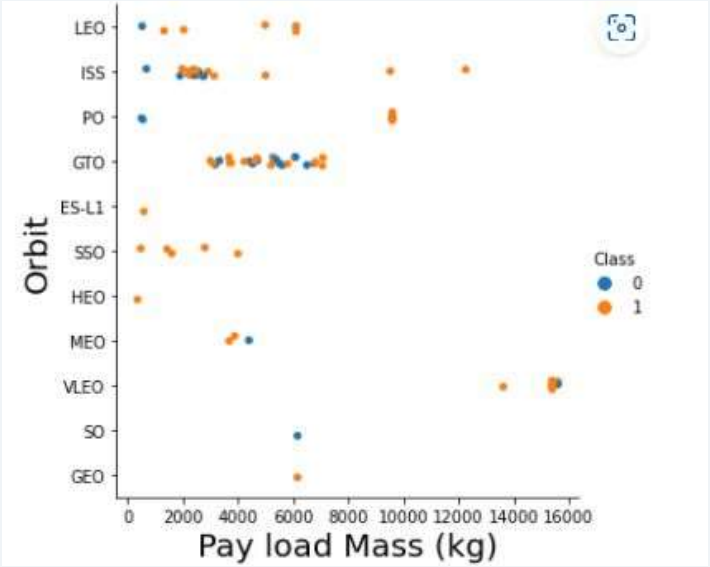
Flight Number vs. Orbit Type

- MEO, VLEO, SO, and GEO were used the later flights
- ES-LI, HEO, SO, and GEO were only used once
- Most commonly used were LEO, ISS, GTO, and VLEO



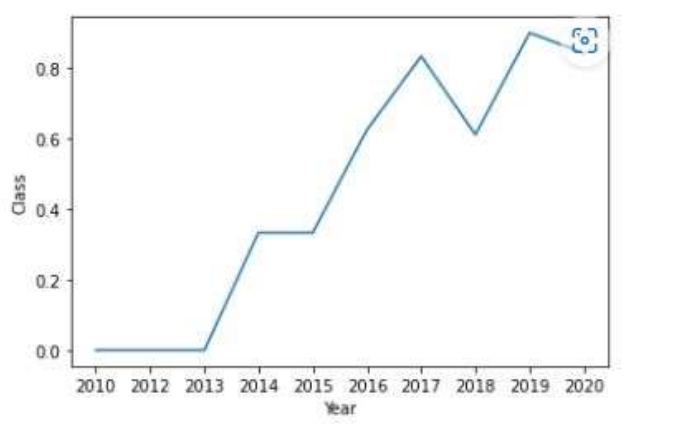
Payload vs. Orbit Type

- ISS had the greatest range of payload
- VLEO was mainly used for heavy payloads
- MEO, SSO, GTO were used for lighter payloads



Launch Success Yearly Trend

- Landing Success Rate generally increases with year
- In 2018, the landing success rate decreased from 2017



All Launch Site Names

- 4 unique Launch Site Names
  - CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40

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

Launch Site Names Begin with 'CCA'

- CCAFS LC-40 starts with 'CCA'

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- 45596kg of payload were launched by NASA (CRS) in total

TOTAL_PAYLOAD_MASS_NASA_CRS
45596

Average Payload Mass by F9 v1.1

- Average payload mass for F9 v1.1 boosters was 2928.4 kg

AVG_PAYLOAD_MASS_F9_V1_1
2928.4

First Successful Ground Landing Date

- The first date with a successful landing on a ground pad was 01-05-2017 (May 1st, 2017)

MIN_DATE_SUCCESS_GROUND_PAD
01-05-2017

Successful Drone Ship Landing with Payload between 4000 and 6000

- These 4 boosters had a successful drone ship landing while having a payload between 4000 kg and 6000 kg

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

Total Number of Successful and Failure Mission Outcomes

- 61 Successes and 10 Failures
- Note that some missions had 'None' for outcome, and we can't interpret anything from that data

Successes	Failures
61	10

Boosters Carried Maximum Payload

- These boosters carried a payload was that equivalent to the 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

- These were launched in 2015 and failed to land

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

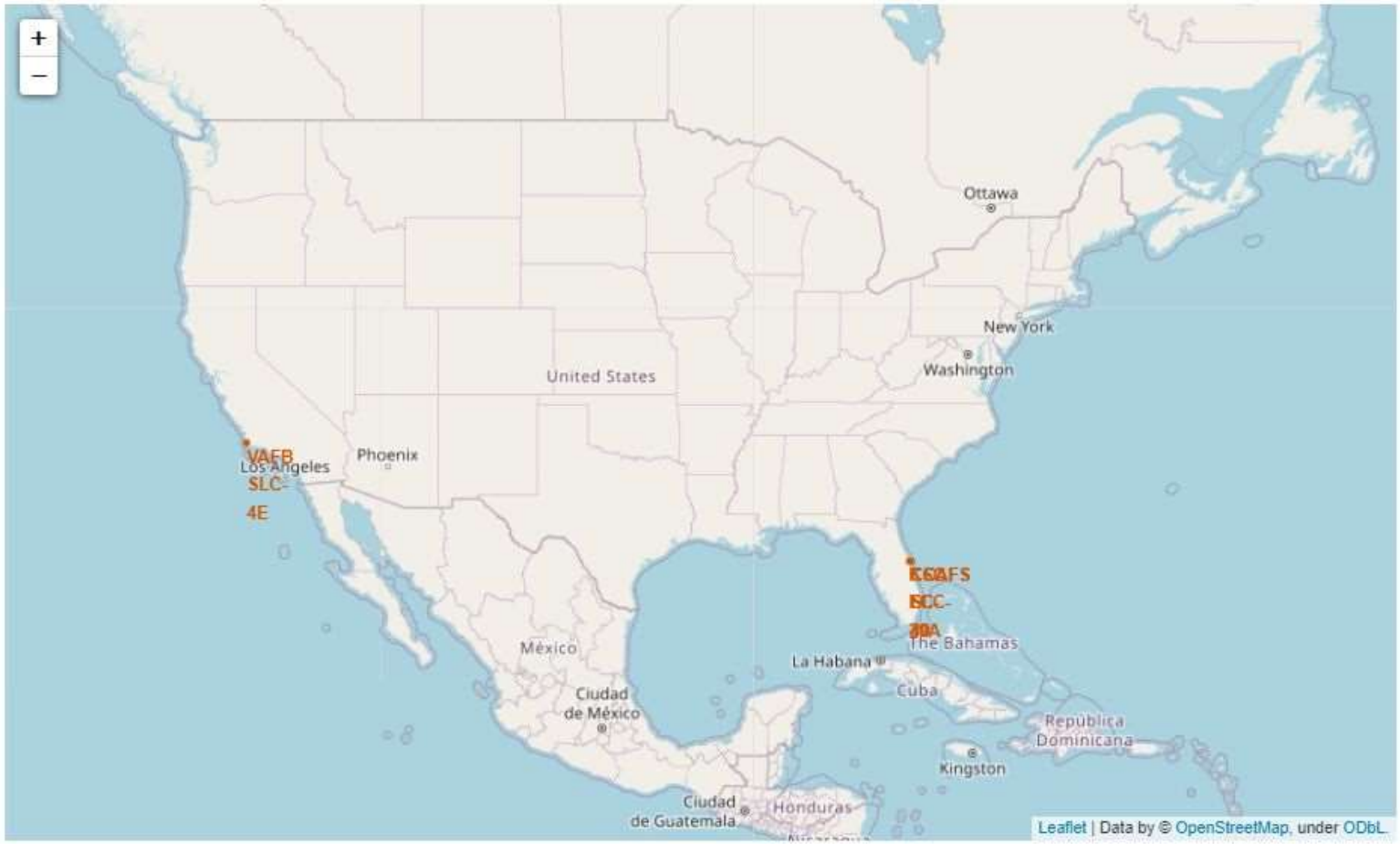
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- These are the counts for successful landing outcomes between 2010-06-04 and 2017-03-20, sorted in descending order
- Out of 34 successful landings, 20 were a plain success, 8 were on a drone ship, and 6 were on a ground pad

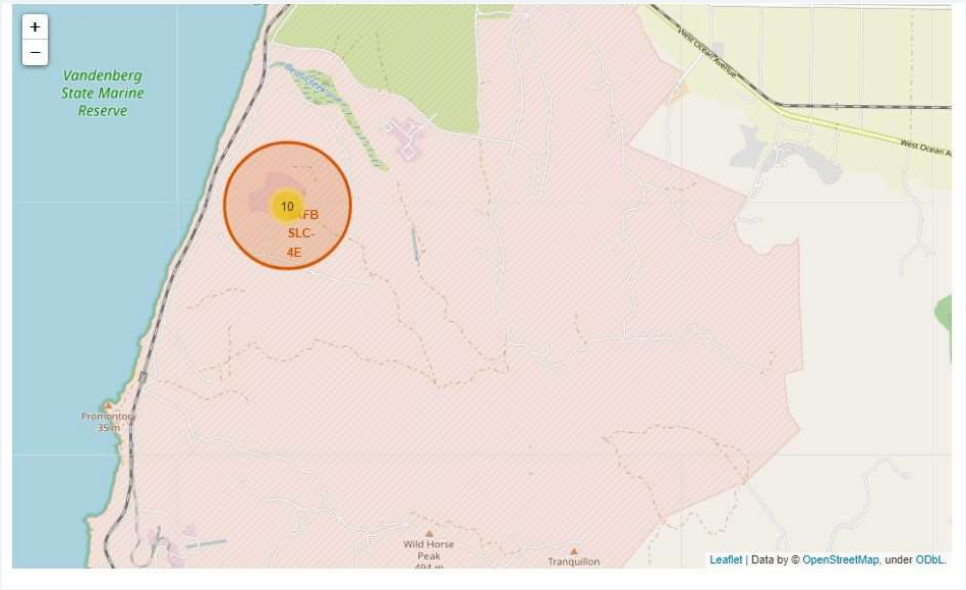
Landing _Outcome	successful
Success	20
Success (drone ship)	8
Success (ground pad)	6

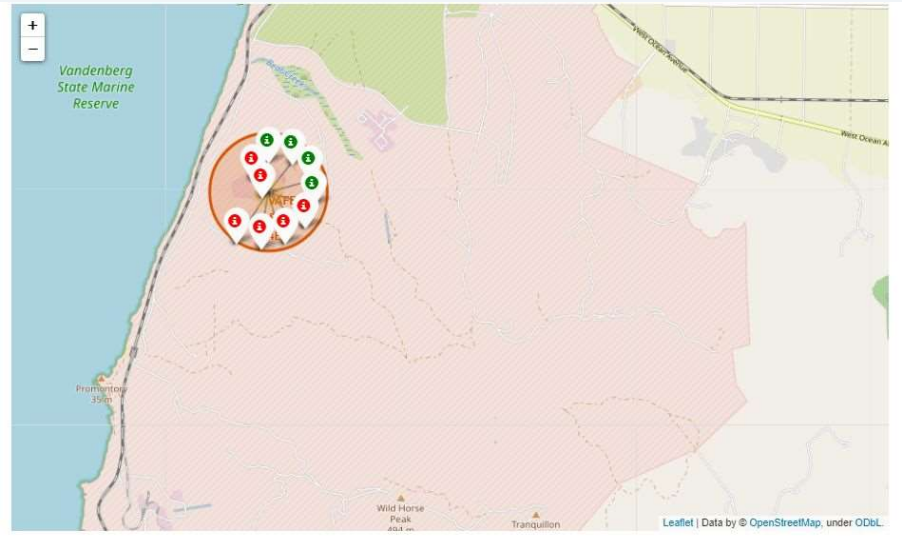

Section 3
Launch Sites
Proximities Analysis

Locations of Launch Sites



Successes and Failures at each Launch Site

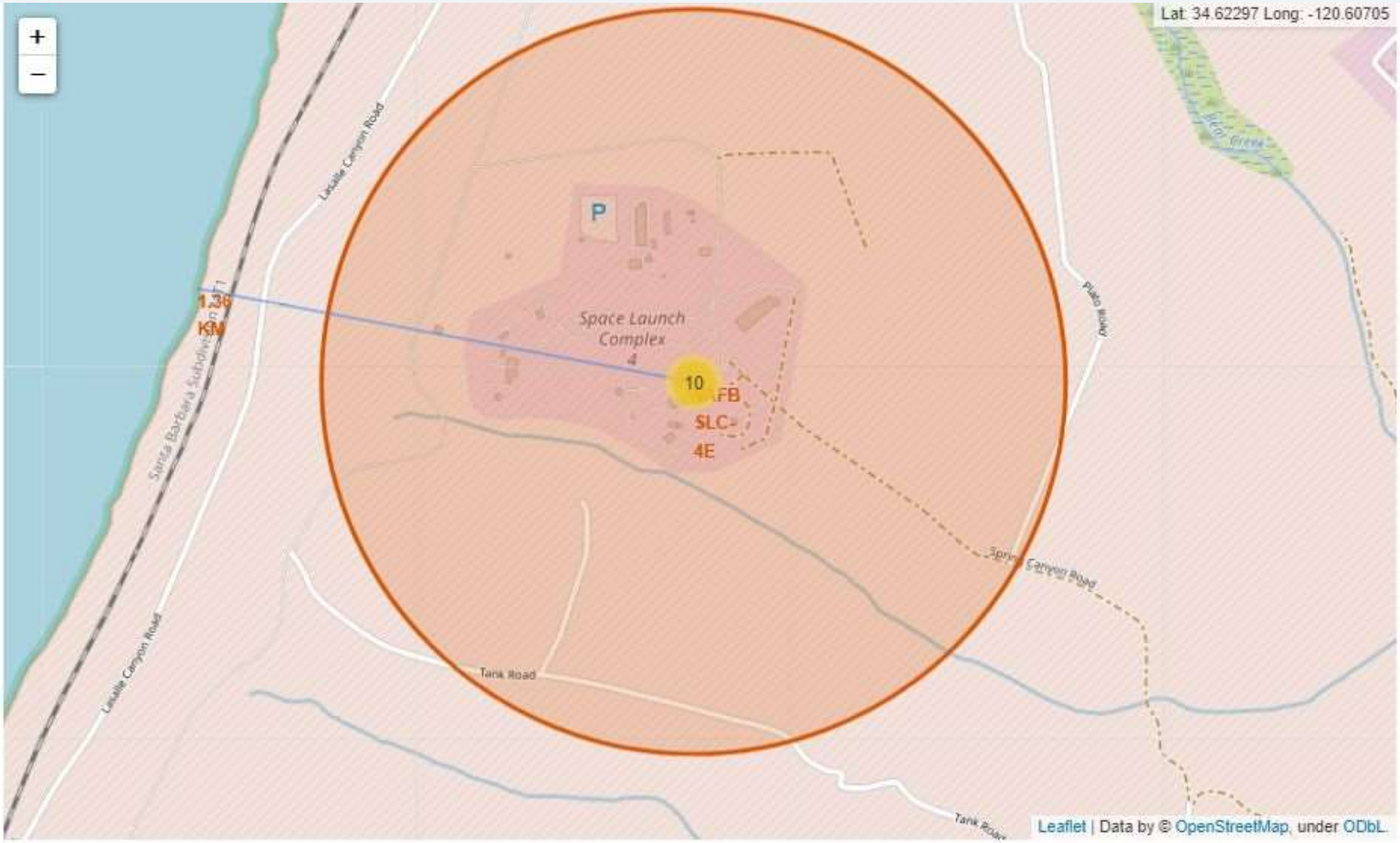




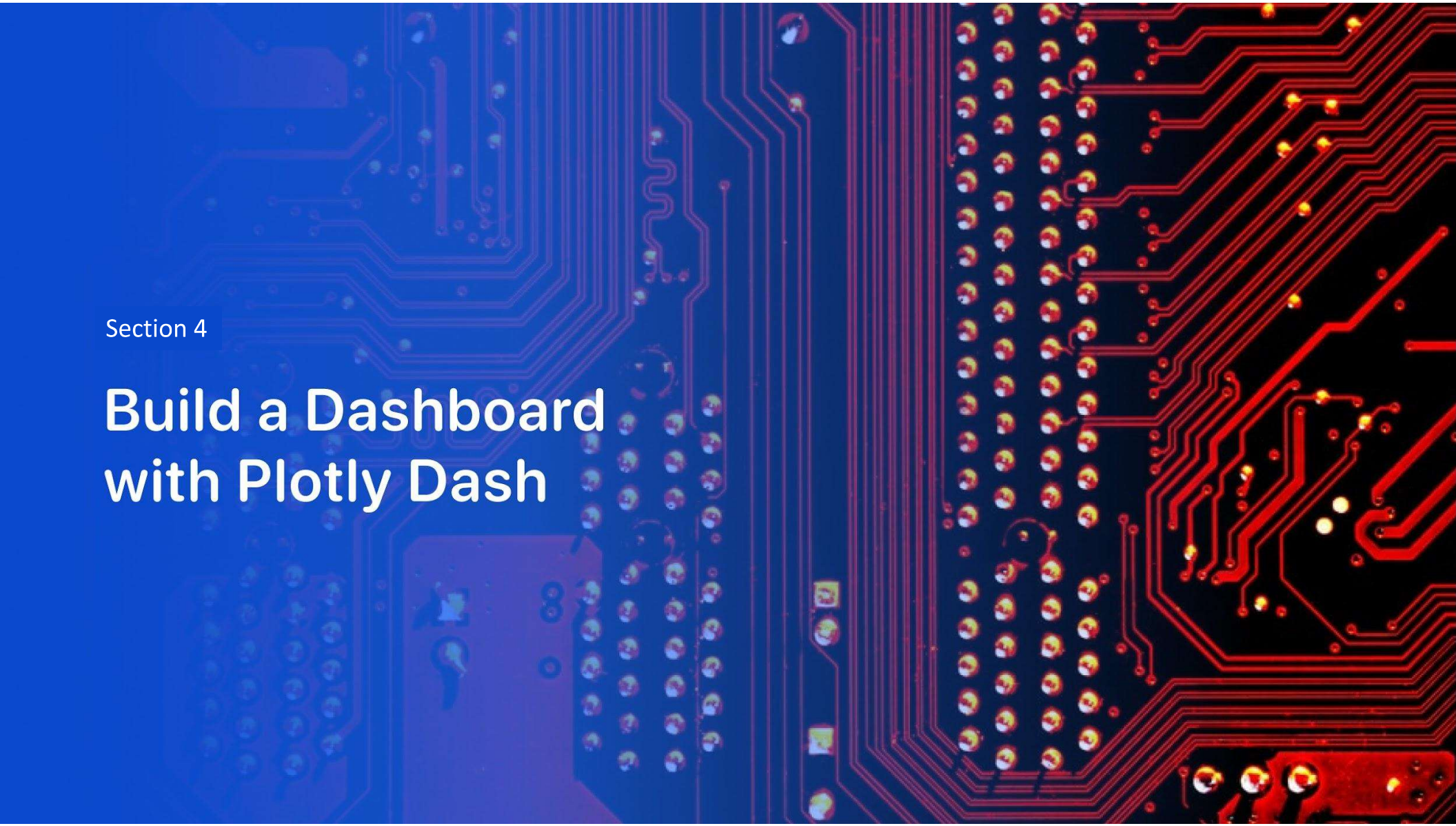
10 launches at this launch site

- green = success, red = failure
- 4 successful landings out of 10

Distances and Lines



- We see that the coast i 1.36 km from the space launch site, and we can see the line from the space launch site to the coast



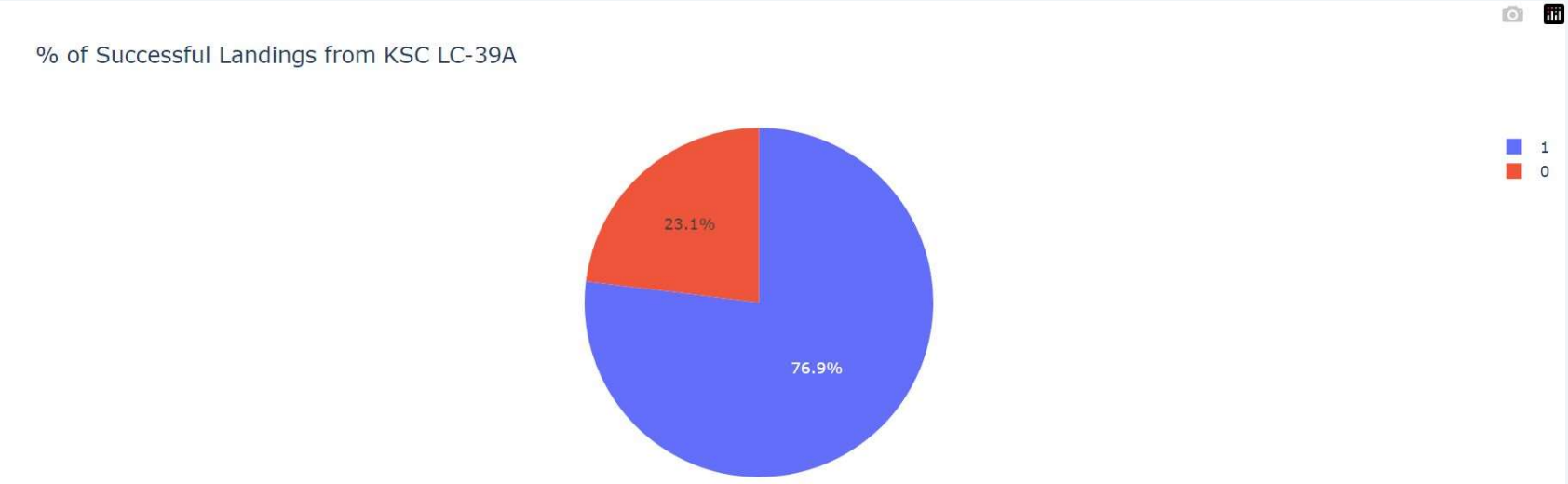
Section 4

Build a Dashboard with Plotly Dash

Pie chart for all sites

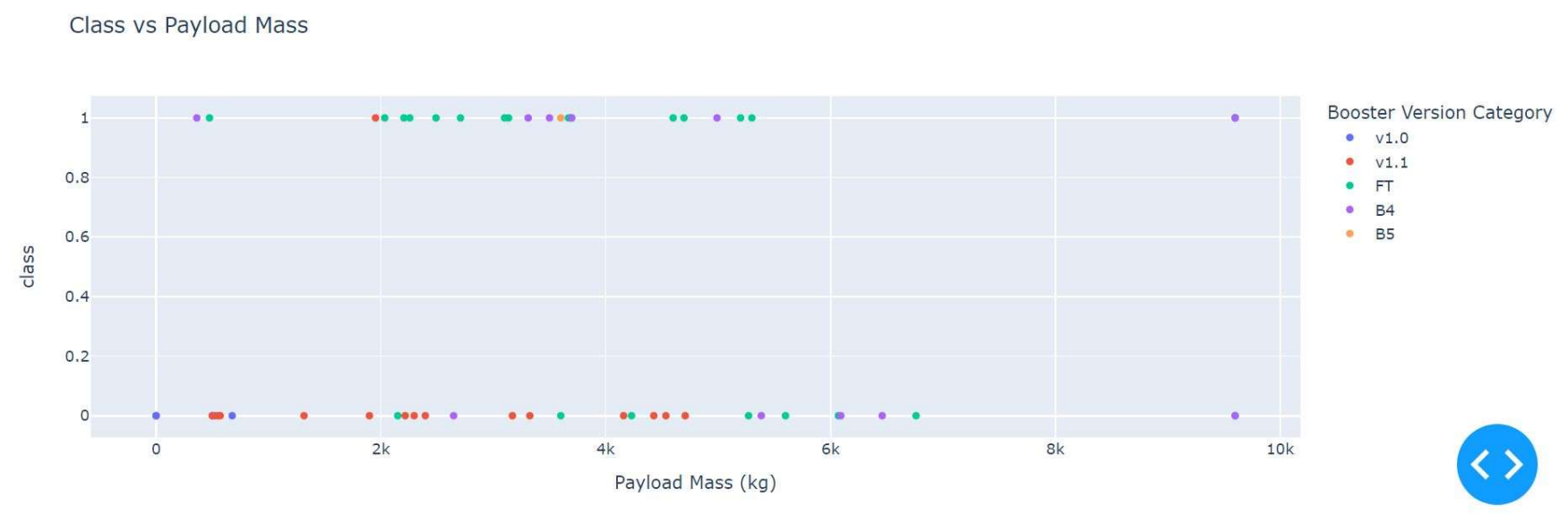
41.7% of all successful landings came from KSC LC-39A

Pie Chart for KSC LC-39A



KSC LC-39A had the greatest successful landing rate, at 76.9%. CCAFS LC-40 was close at 73.1%

Payload vs Launch Outcome



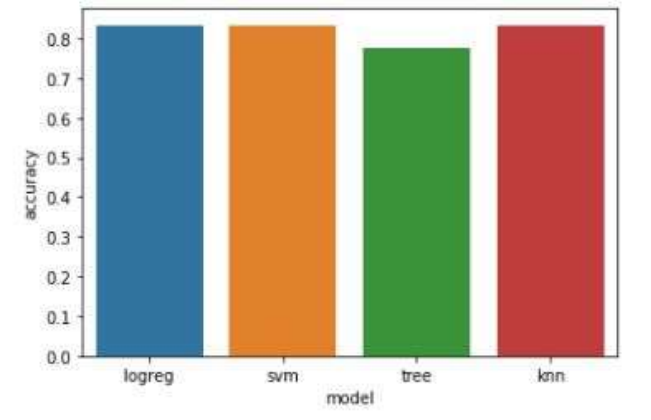
- v1.1 seems to perform poorly
- FT generally performs well
- heavier payloads tend to fail the landing more than lighter payloads

Section 5

Predictive Analysis (Classification)

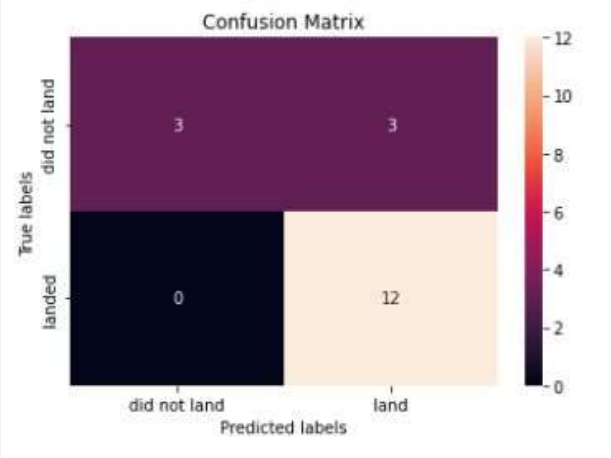
Classification Accuracy

- Logistic Regression, SVM, and KNN are tied with 83.33% accuracy
- Decision Tree is slightly lower at 77.78% accuracy



Confusion Matrix

- This is the confusion matrix for Logistic Regression
- For 3 launches in the test set that did not land, the model predicted it would land.
- For the other 15 launches in the test set, the model predicted it correctly
    - 12 landed and 3 did not land



Conclusions

- Heavier payloads were more likely to fail the landing
- Some booster versions were significantly better than others
- KSC LC-39A was the best launch site
- Our model used features such as launch site, serial number, booster type, payload mass to try to predict landing
    - worked with about 83.33% accuracy

Appendix

- Github

Thank you!