IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Oleg Zhdanenia
14 May 2024

Outline

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

Executive Summary

“Winning Space Race with Data Science” aims to analyse SpaceX launches to optimize future launches.

In this project were applied data science methodologies: data collection and wrangling, EDA, ML. Applying these methodologies gave us insight on factors that influences launches.

Introduction

SpaceX provides solution that makes rocket launches cheaper by reusing the first stage of rocket. But this requires to successfully land the first stage after launch.

Instead of rocket science we will use Data Science to predict if SpaceX can reuse the first stage

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collecting data using SpaceX REST API and Web Scraping using BeautifulSoup
- Perform data wrangling
  - Using JSON normalization
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - SVM, Logistic Regression, Decision Tree Classifier
  - Cross-Validation for hyperparameter tuning

Data Collection

- Data was collected using 2 methods:
- SpaceX REST API
- Web Scraping using BeautifulSoup

Data Collection – SpaceX API

1. Load JSON data into Pandas DataFrame
2. Extract and filter useful features
3. Handle missing values

Data Collection - Scraping

1. Call Falcon 9 Launch Wiki Page
2. Extract and filter useful features from HTML
3. Convert to Pandas DataFrame

Data Wrangling

Our target variable is “Outcome” which has values like “None None”, “True ASDS”, “True RTLS”, “False ASDS”, etc.

We will create a new variable and call it “Class”. It will have values 0 for bad outcome or 1 for good outcome.

Values “False” or “None” represents bad outcome.

EDA with Data Visualization

- Scatter plots
  to identify underlying trends in data
- Bar plot
  to visualize success rates for each orbit
- Line plot
  to visualize successful launches throughout years

EDA with SQL

Using SQL we have:

- identified launch sites
- found out average payload mass for Falcon 9 booster
- identified booster that has successful launch with max payload mass
- total number of successful and failed launches

Build an Interactive Map with Folium

- Circles and Markers for visualizing launch sites location
- Marker clusters for marking launches for each launch site. Each launch marker has color that depends on success of mission
- Lines from launch site to nearest objects like coastline, railway, city with distances

Build a Dashboard with Plotly Dash

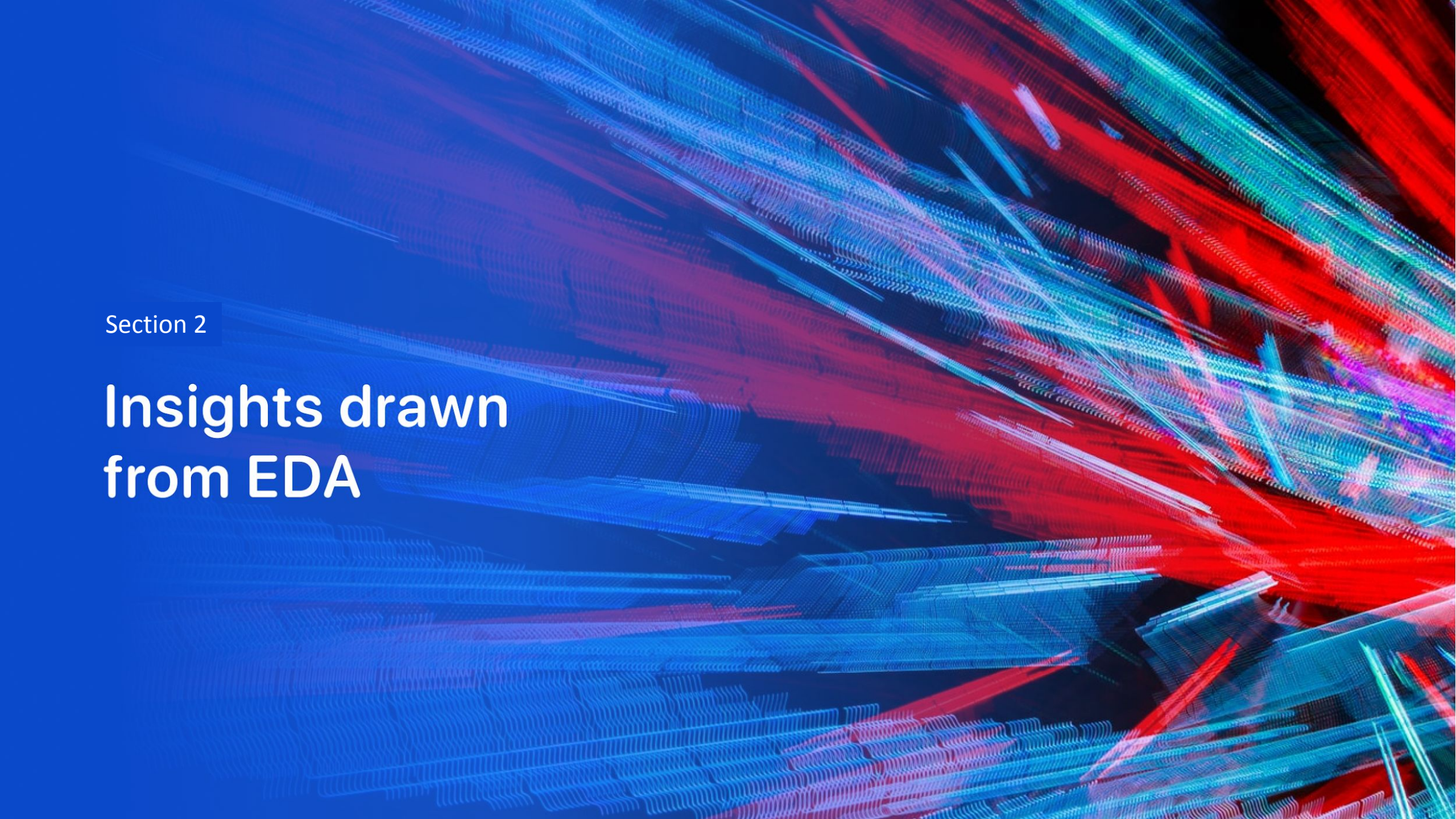
- Pie chart shows successful launches by launch site
- Scatter plot shows trends between success rate and payload mass, booster version
- Dashboard helps to answer questions such as:
    - which launch site have highest success rate?
    - which payload mass have highest success rate?
    - which Falcon 9 booster version have highest success rate?

Predictive Analysis (Classification)

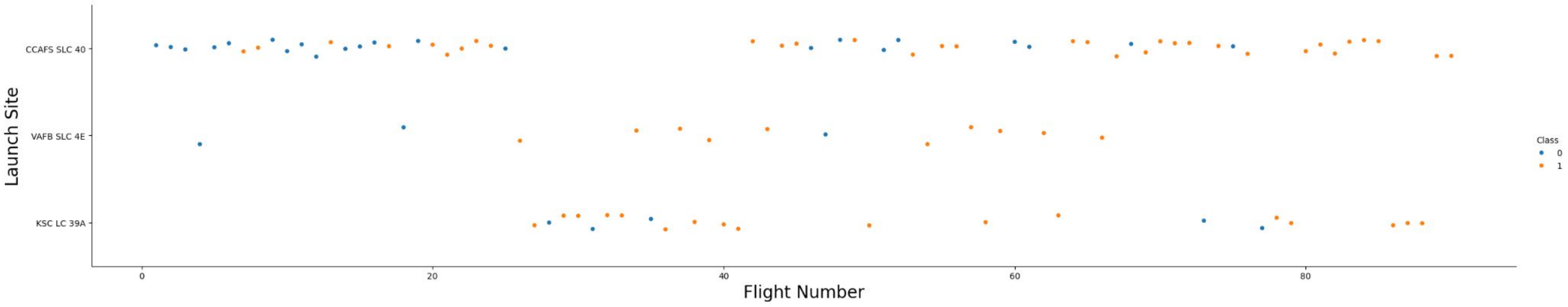
- Data preprocessing for model training: standardization
- Models for training: Logistic Regression, SVM, Decision Tree Classifier, KNN
- Each model pass through the same pipeline:
    - Cross Validation of Parameters
    - Fitting model with optimal parameters
    - Evaluation of metrics on test data

Section 2

Insights drawn from EDA

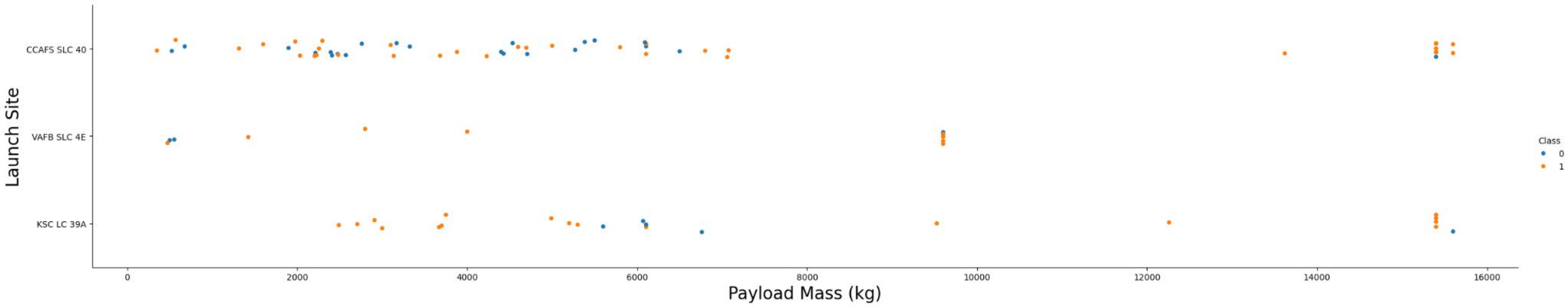


Flight Number vs. Launch Site



Later flights mostly launched from CCAFS SLC-40 and has more successful outcomes

Payload vs. Launch Site

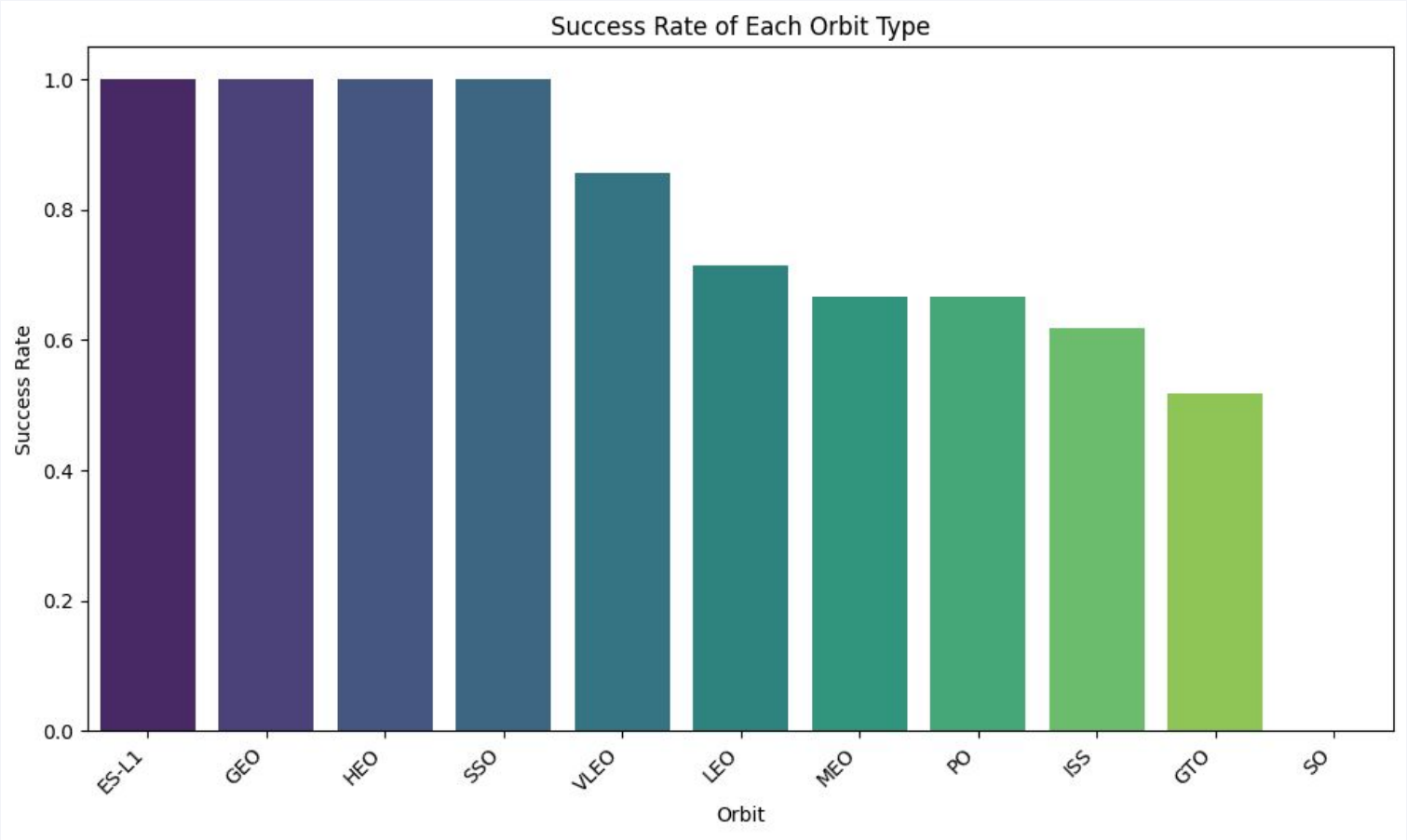


CCAFS SLC-40 and KSC LC-39A has launches (including successful) with largest payloads

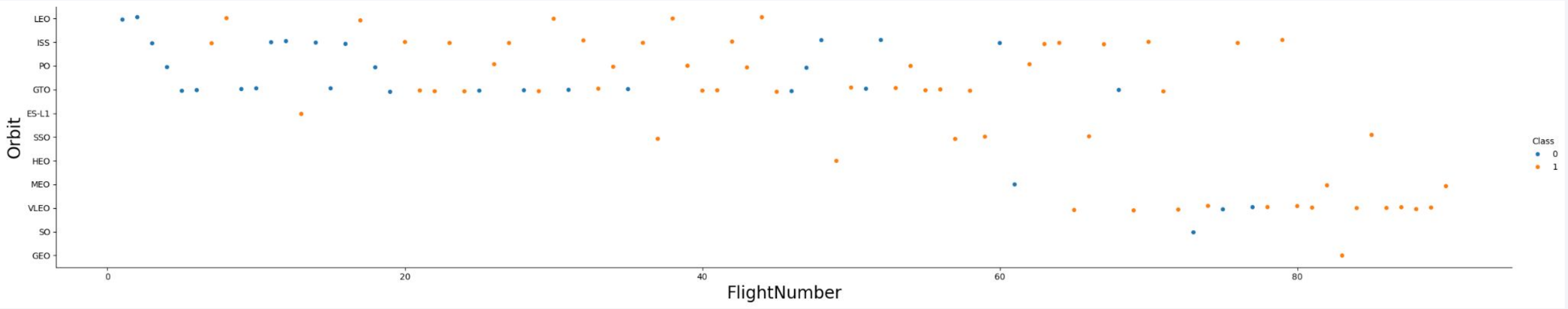
WAFB SLC 4E possible has payload limits

Success Rate vs. Orbit Type

- Highest success rate has ES-L1, GEO, HEO and SSO orbits

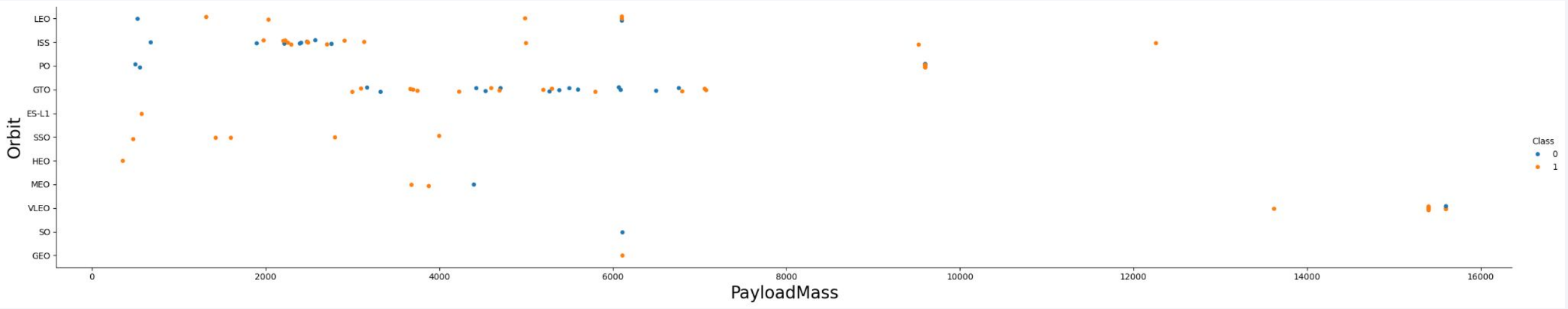


Flight Number vs. Orbit Type



Later flights mostly has been launched to VLEO orbit

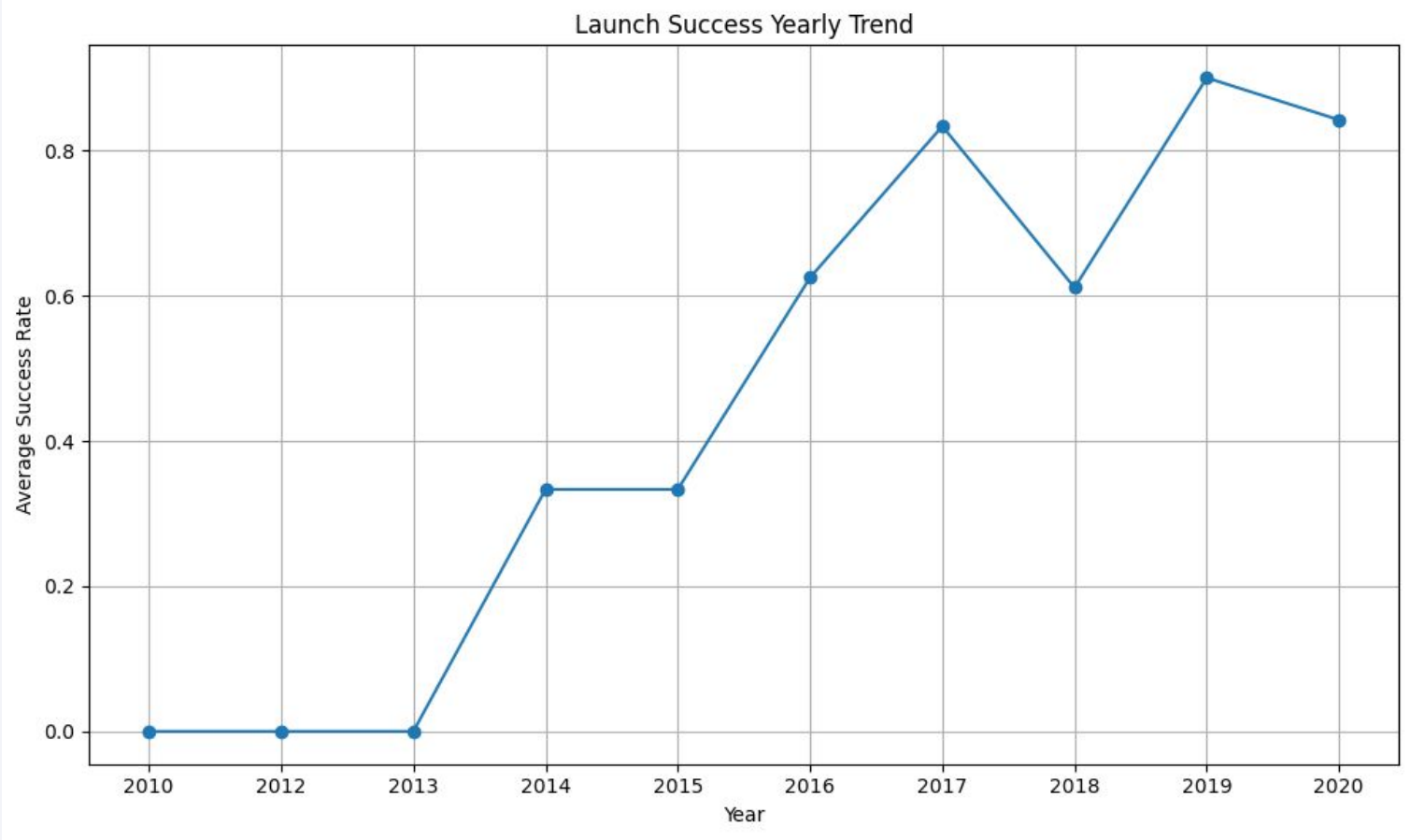
Payload vs. Orbit Type



The heaviest launches has carried payload to VLEO orbit

Launch Success Yearly Trend

- Launch success rate is improving through the time
- 2019 was the most successful year by the moment



All Launch Site Names

- Using SQL Query we got 4 launch sites:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with `CCA`

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outc
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (paracl
2010-12-08	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	Failure (paracl
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No att
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No att
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No att

Total Payload Mass

The total payload carried by boosters from NASA

NASA total payload mass
45596

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1

avg payload mass booster F9 v1.1
2928.4

First Successful Ground Landing Date

Date of the first successful landing outcome on ground pad

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

]:

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

Total Number of Successful and Failure Mission Outcomes

The total number of successful and failure mission outcomes

]:

Landing_Outcome	Outcome_Count
Failure	16
Other	24
Success	61

Boosters Carried Maximum Payload

Names of the booster which have carried the maximum payload mass

]:

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

Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

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

Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20

Landing_Outcome	Outcome_Count
Failure (drone ship)	5
Success (ground pad)	3

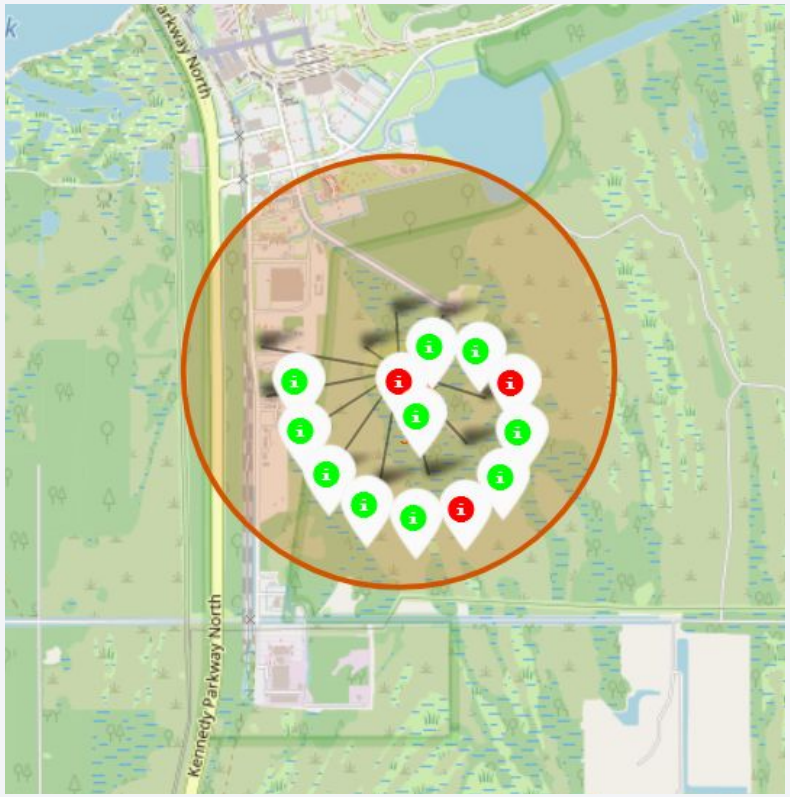

Section 3
Launch Sites
Proximities Analysis

Launch Sites

We have 4 launch sites:

- 1 in California
- 3 in Florida

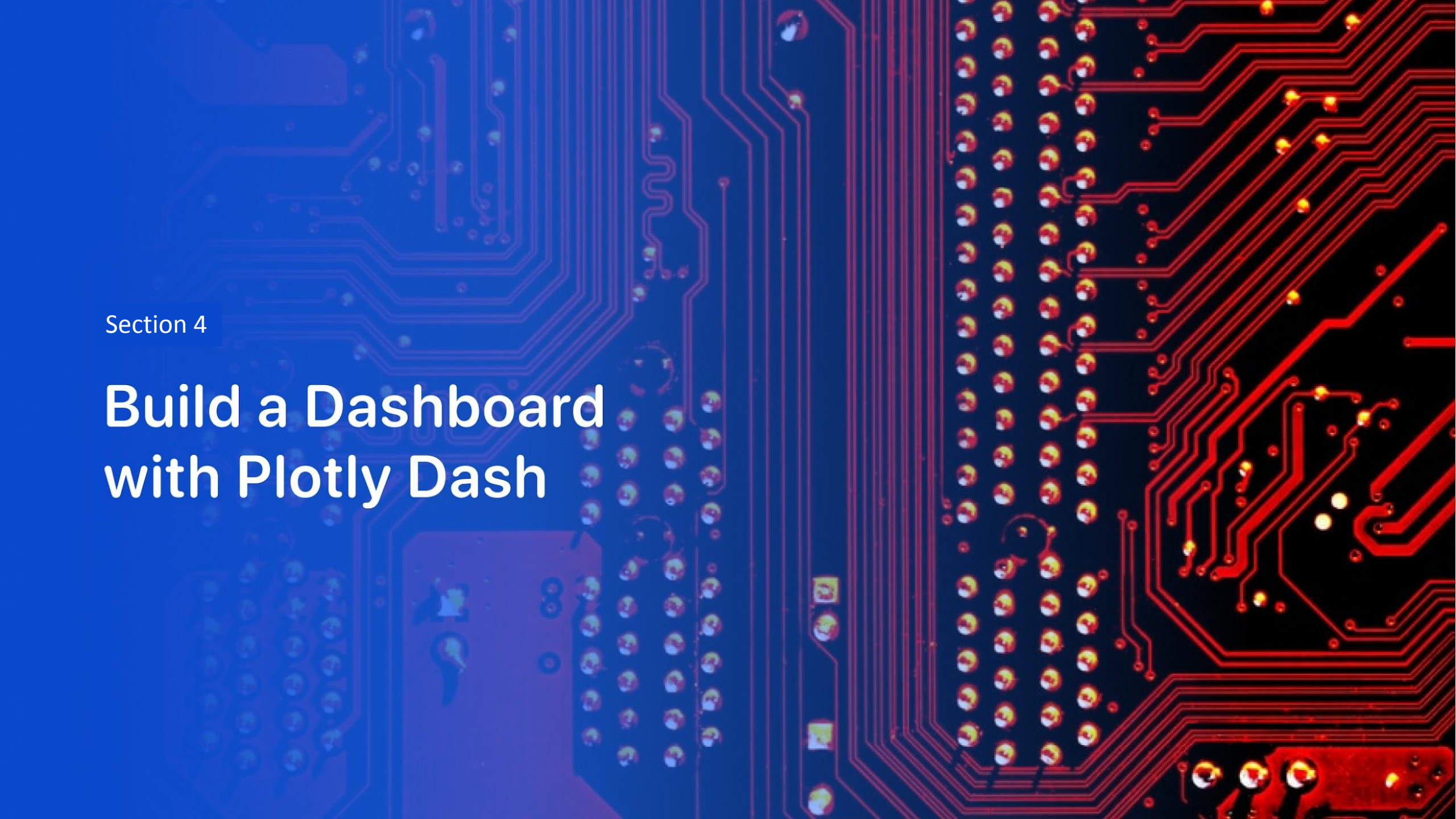
Launch Outcome Labels



With folium interactive map we can see launches with outcome for each site.

Site Proximities

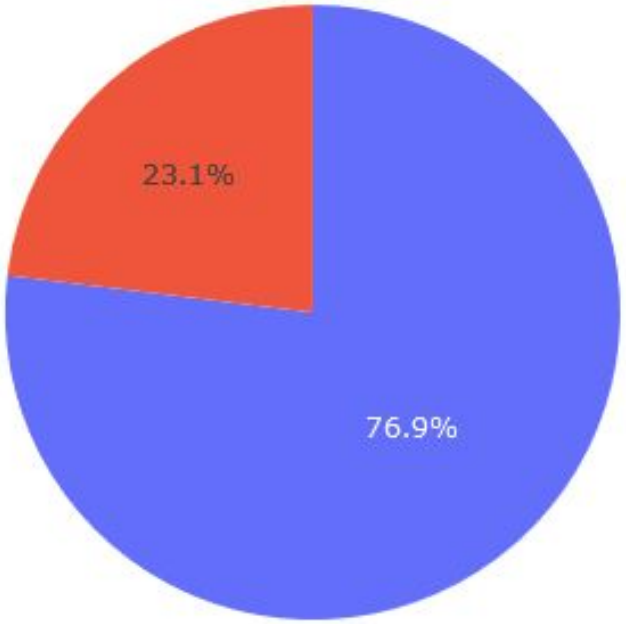
We can find out proximities of site to such objects as coast line, cities, railways, etc.


Section 4
Build a Dashboard with Plotly Dash

Launch Sites Success Ratio

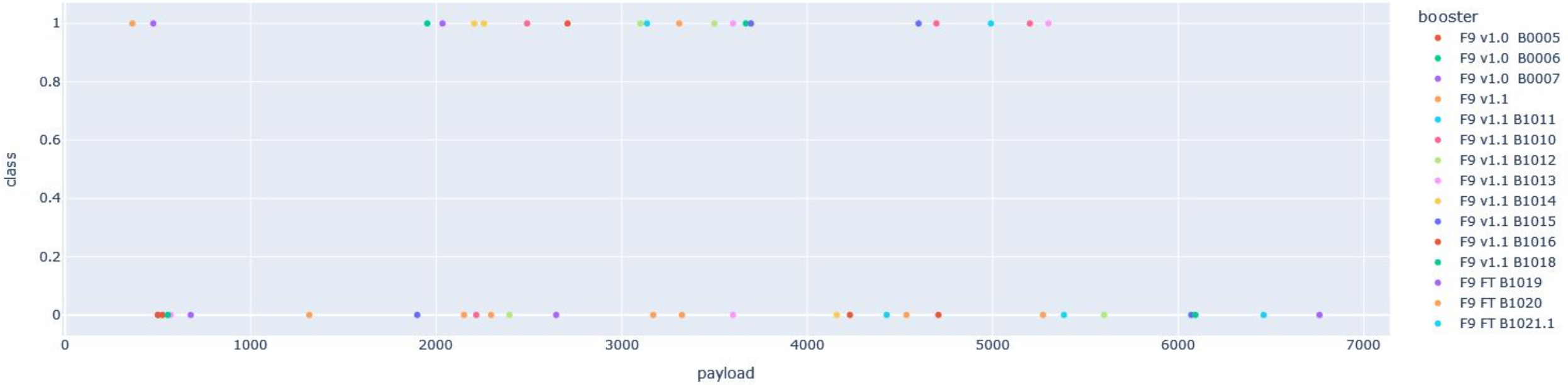
On the pie chart we can see ration on successful launch from different launch sites. The most successful launch site is KSC LC-39A

KSC LC-39A Launches Ratio



Success rate for KSC LC-39A is 76.9%

Scatter plot for success rate

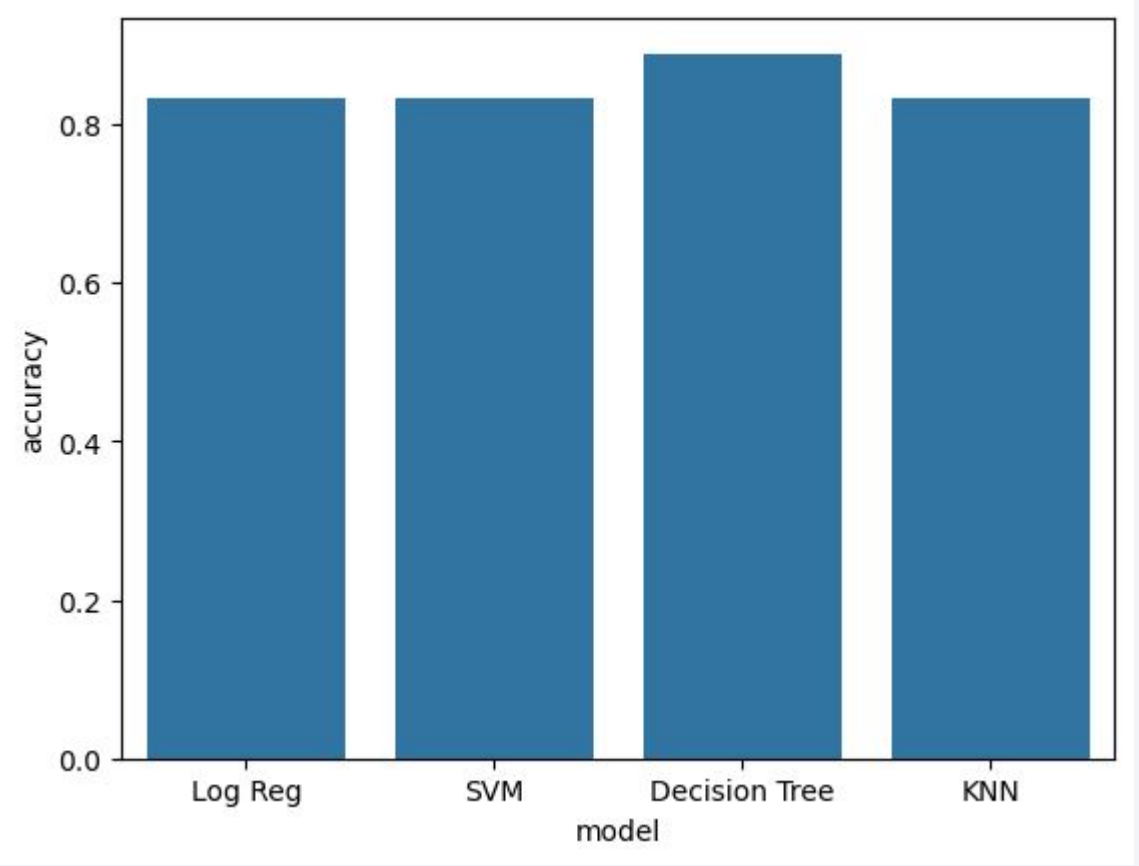


Section 5

Predictive Analysis (Classification)

Classification Accuracy

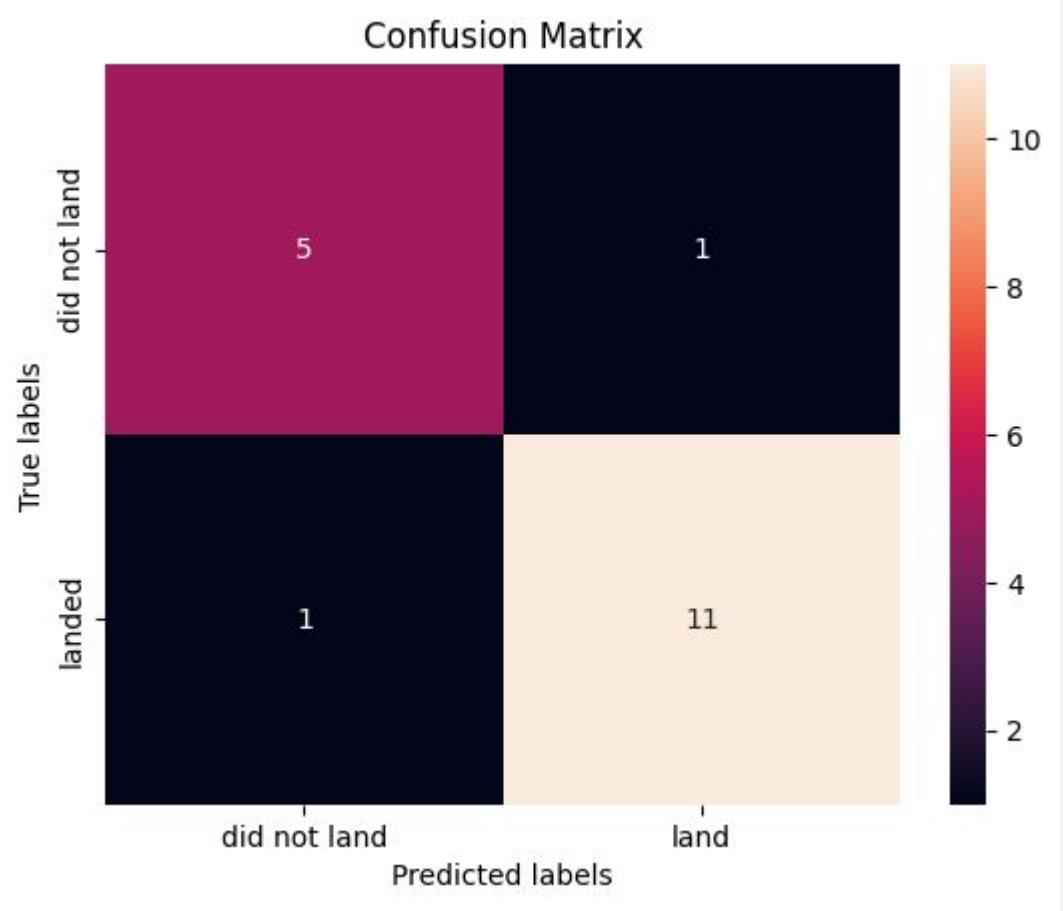
Decision Tree model has best results with accuracy 88.88%



Confusion Matrix: Decision Tree

From confusion matrix we see:

- True positives: 11
- True negatives: 5
- False positives: 1
- False negatives: 1



Conclusions

- We performed necessary data preprocessing
- Visualized data to get meaningful insights
- Performed EDA
- Applied cross validation on multiple ML models
- Identified best performing model for launch success prediction

Thank you!