IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Chow Kwok Ho
April 9, 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

Introduction

- Project background and context

  •New company Space Y that would like to compete with SpaceX founded by Billionaire industrialist Allon Musk.

- Problems you want to find answers

  •Determine the price of each launch.

  •Determine if SpaceX will reuse the first stage

  •Train a machine learning model and use public information to predict if SpaceX will reuse the first stage

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API and WebScraping
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Normalization, training and test data sets and evaluation by four different classification models by accuracy rate

Data Collection

- SpaceX REST API

  https://api.spacexdata.com/v4/rockets

- Wikipedia Website

  https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

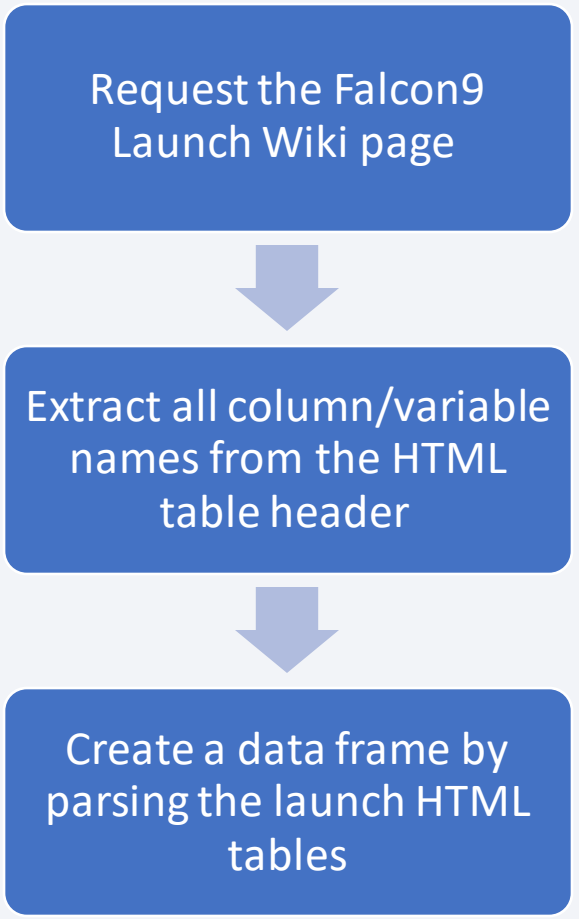
Data Collection – SpaceX API

- SpaceX REST is available to help obtaining data

- Source code: https://github.com/chowkwokho/AppliedDataScienceCapstone/blob/main/Week 1 Task 1 Data Collection API.ipynb

Data Collection - Scraping

- There is data available in Wikpiedia
- Scrapping tool BeautifulSoup is used



- Source code: https://github.com/chowkwokho/AppliedDataScienceCapstone/blob/main/Week 1 Task 2 Data Collection with Web Scraping.ipynb

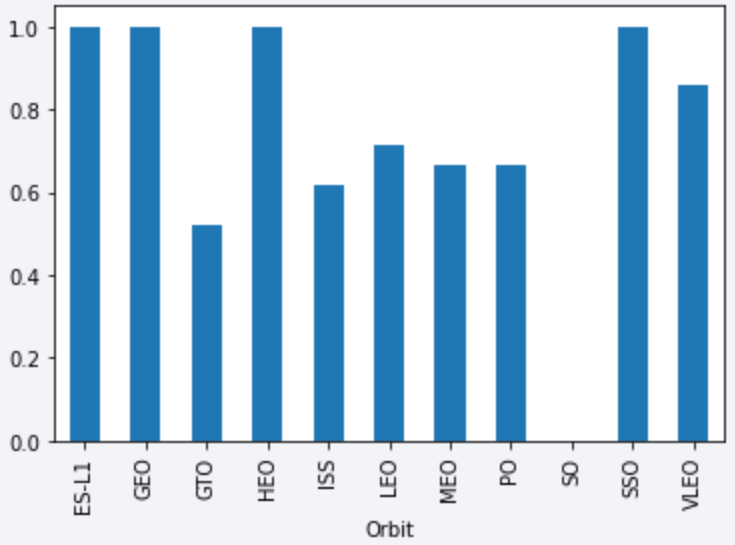
Data Wrangling

- Exploratory Data Analysis is conducted.
- Summary tables such as occurrence by launch site were outputted.
- Outcome Label is created based on landing outcome label

- Source code: https://github.com/chowkwokho/AppliedDataScienceCapstone/blob/main/Week 1 Task 3 Data Wrangling.ipynb

EDA with Data Visualization

- Visual relationship between features:
  - Scatter plot
  - Bar plot



- Source code: https://github.com/chowkwokho/AppliedDataScienceCapstone/blob/main/Week 2 Task 2 EDA with Visualization.ipynb

EDA with SQL

- Following task performed:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass.
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- Source code: https://github.com/chowkwokho/AppliedDataScienceCapstone/blob/main/Week 2 Task 1 EDA with SQL.ipynb

Build an Interactive Map with Folium

- Markers: Indicate points of location such as launch sites;
- Circles: Highlight areas around specific coordinates such as NASA Johnson Space Center;
- Marker clusters: Indicates groups of events in each coordinate such as launches in a launch site; and
- Lines: Indicate distances between two coordinates.

- Source code: https://github.com/chowkwokho/AppliedDataScienceCapstone/blob/main/Week 3 Task 1 Interactive Visual Analytics.ipynb

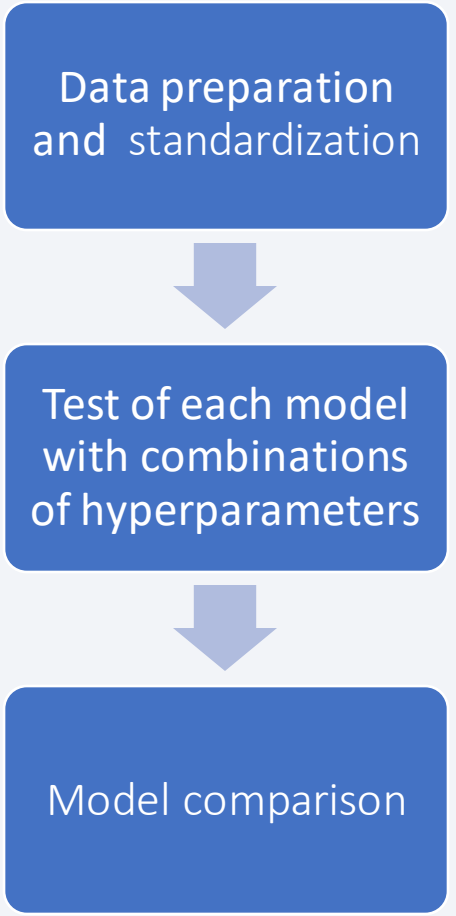
Build a Dashboard with Plotly Dash

- Visualize data with graphs and plots
  - Percentage of launches by site
  - Payload range
- Help to analyze the relation between payloads and launch sites, and identify where is best place to launch according to payloads.

- Source code: https://github.com/chowkwokho/AppliedDataScienceCapstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree Classifier
  - k Nearest Neighbors (kNN)



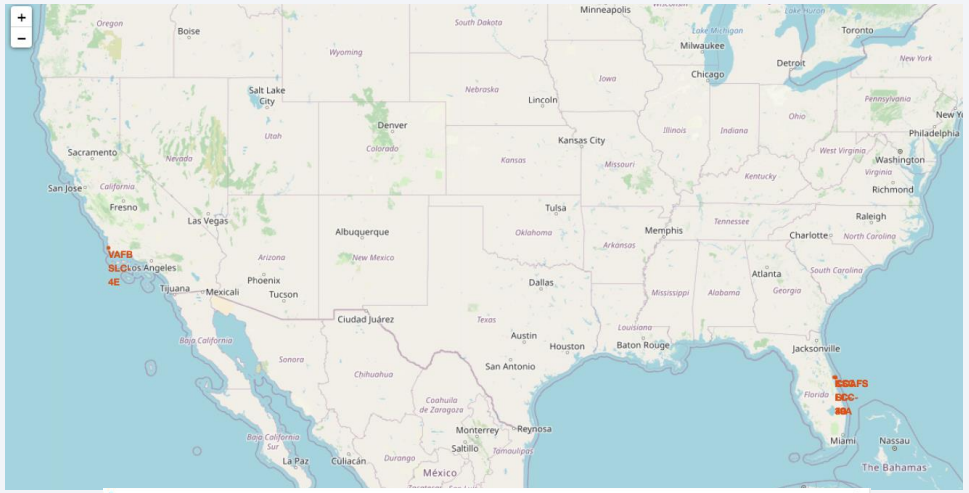
- Source code: https://github.com/chowkwokho/AppliedDataScienceCapstone/blob/main/Week 4 Task 1 Machine Learning Prediction.ipynb

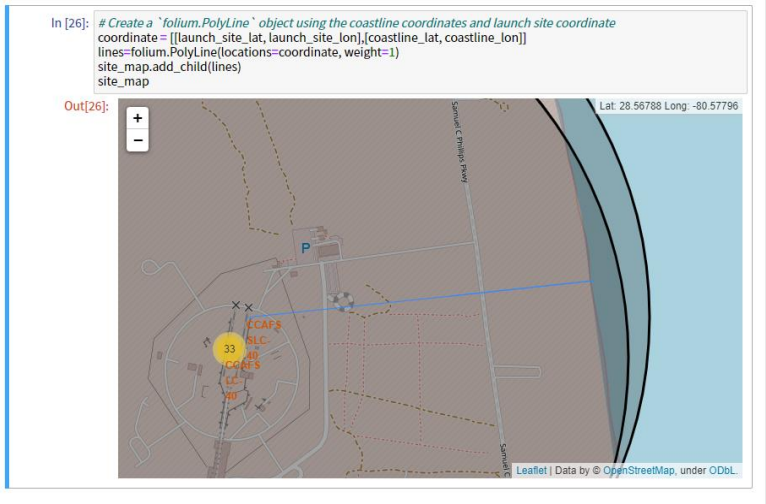
Results

- Exploratory data analysis results:
  - Space X uses 4 different launch sites;
  - The first launches were done to Space X itself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
  - Almost 100% of mission outcomes were successful;
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
  - The number of landing outcomes became as better as years passed.

Results

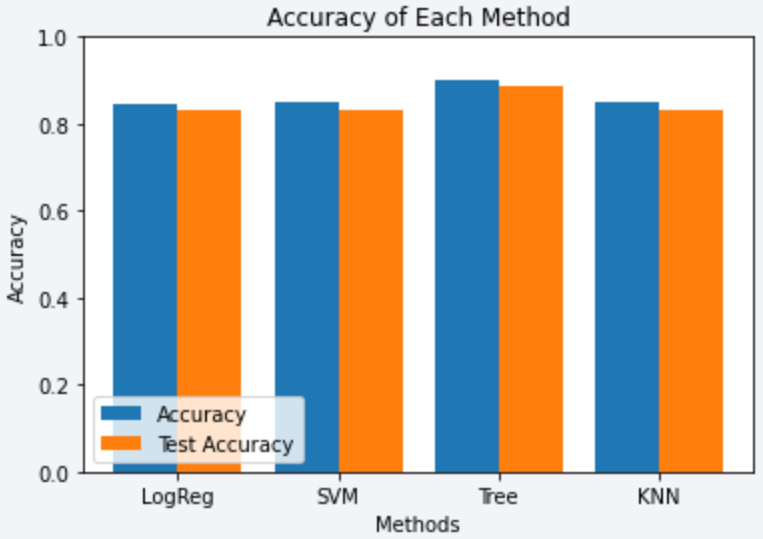
- Interactive analytics helps to identify launch sites use to be in safety places, near sea.
- Most launches happens at east coast launch sites.





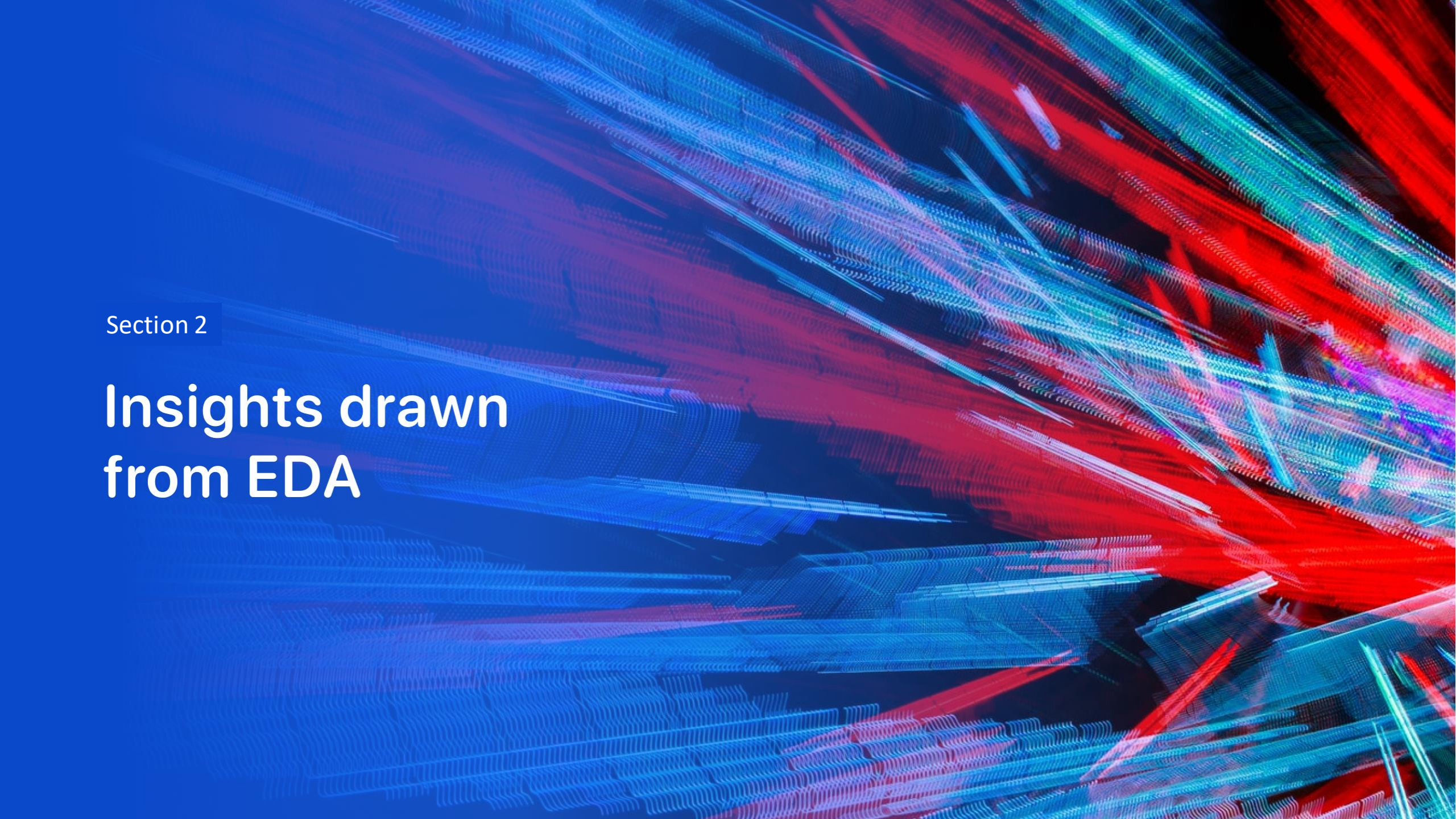
Results

- Predictive analysis results:
  - Decision Tree Classifier is the best model to predict successful landings, having accuracy over 90% and accuracy for test data over 88%

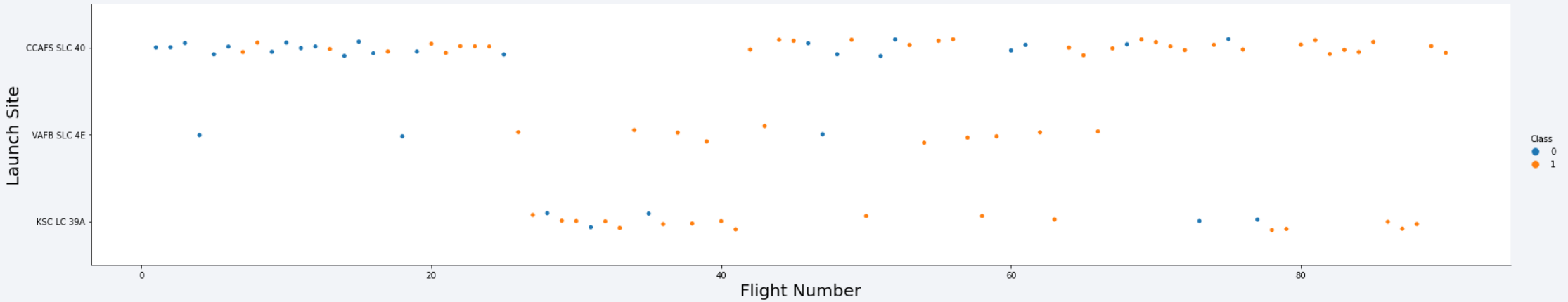


Section 2

Insights drawn from EDA

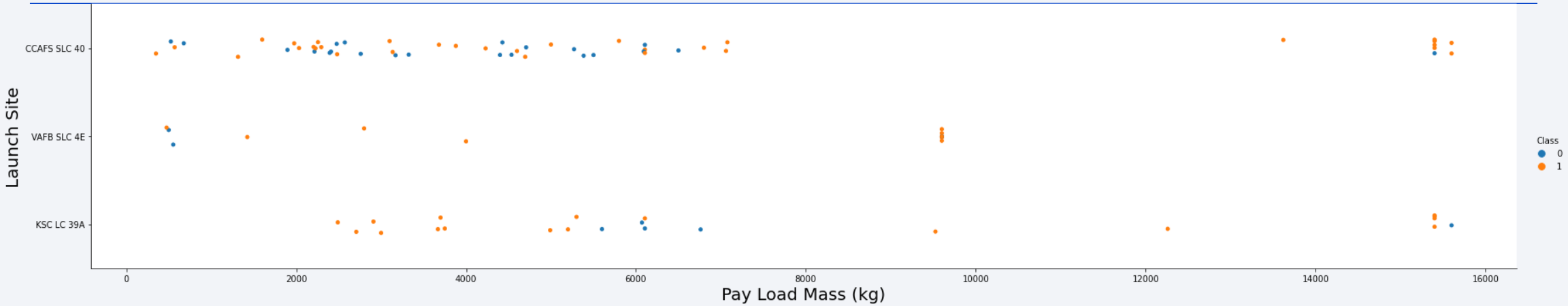


Flight Number vs. Launch Site



- The best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful
- Second place VAFB SLC 4E and third place KSC LC 39A
- The general success rate were improved over time

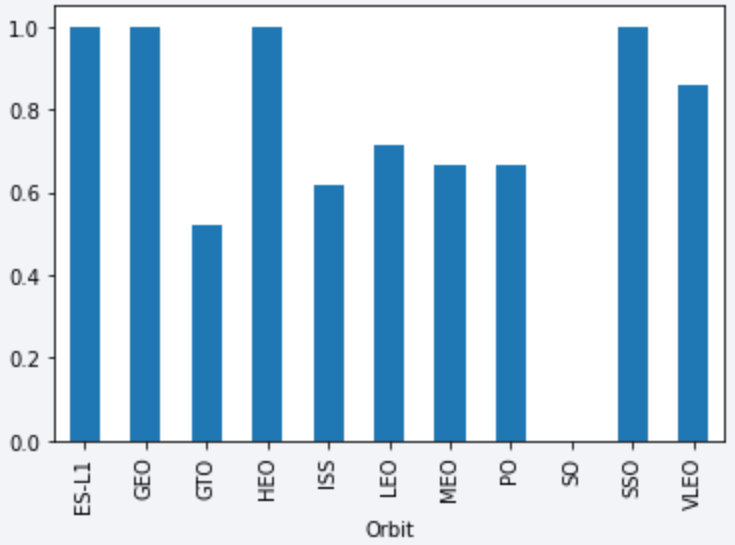
Payload vs. Launch Site



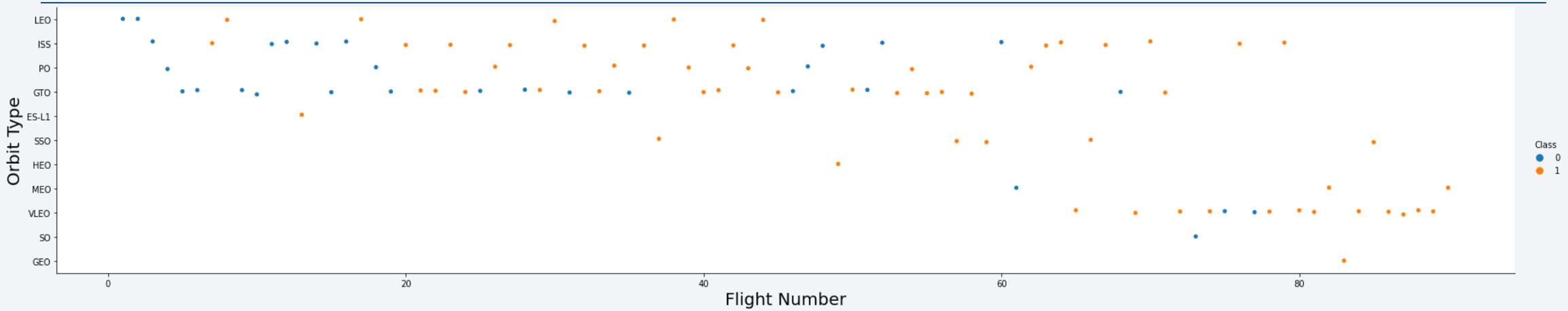
- The excellent success rate are those payloads over 9,000kg;
- For Payloads over 12,000kg, only available CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- The Orbit with highest success rates 100%:
  - ES L1;
  - GEO;
  - HEO; and
  - SSO.

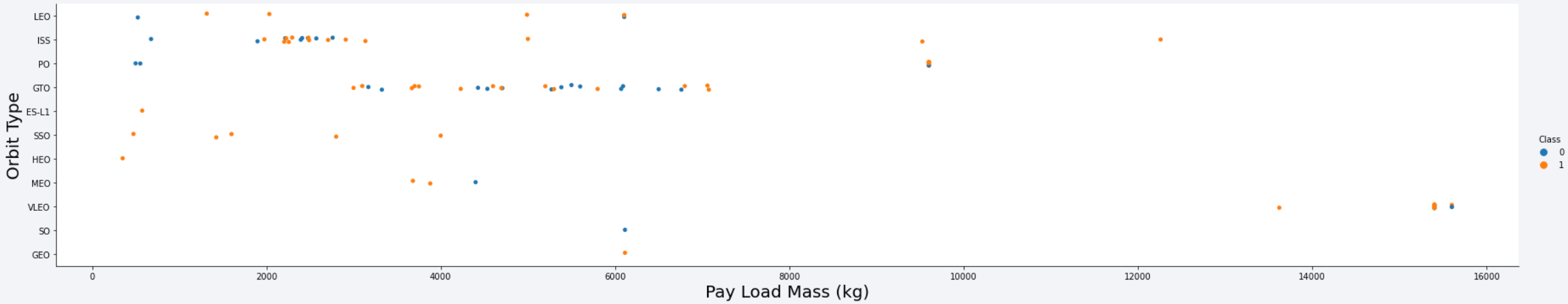


Flight Number vs. Orbit Type



- Success rate improved over time.

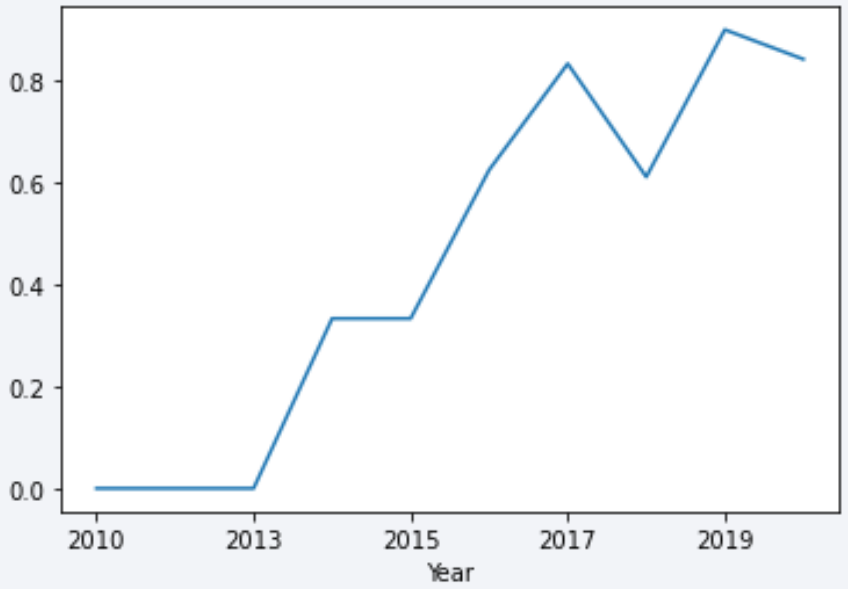
Payload vs. Orbit Type



- Relation between payload and success rate to orbit GTO is not strong;
- ISS orbit get the widest range of payload and a good rate of success;
- SO and GEO launch not so much.

Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020



All Launch Site Names

- 4 Launch sites

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- All mission outcome is success with Orbit LEO

Date	Time UTC	Booster version	Launch site	Pay Load	Pay Load Mass (kg)	Orbit	Customer	Mission Outcome	Landing Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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 (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40		525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40		500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40		677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried by boosters from NASA (code with "CRS") is 111,268

Total Payload Mass
111,268

Average Payload Mass by F9 v1.1

- Applied filter booster version "F9 v1.1", the average payload mass by F9 v1.1 is 2,928

Average Payload Mass by F9 v1.1
2,928

First Successful Ground Landing Date

- The date first successful landing outcome on ground pad is 2015-12-22

Earliest Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Booster version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes
- Most missions are success

Mission outcome	Occurence
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

Boosters Carried Maximum Payload

- 12 booster type which have carried the maximum payload mass

Booster version		
F9 B5 B1048.4	F9 B5 B1049.7	F9 B5 B1056.4
F9 B5 B1048.5	F9 B5 B1051.3	F9 B5 B1058.3
F9 B5 B1049.4	F9 B5 B1051.4	F9 B5 B1060.2
F9 B5 B1049.5	F9 B5 B1051.6	F9 B5 B1060.3

2015 Launch Records

- In 2015, there are two launches with landing outcomes in drone ship

Booster version	Launch site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

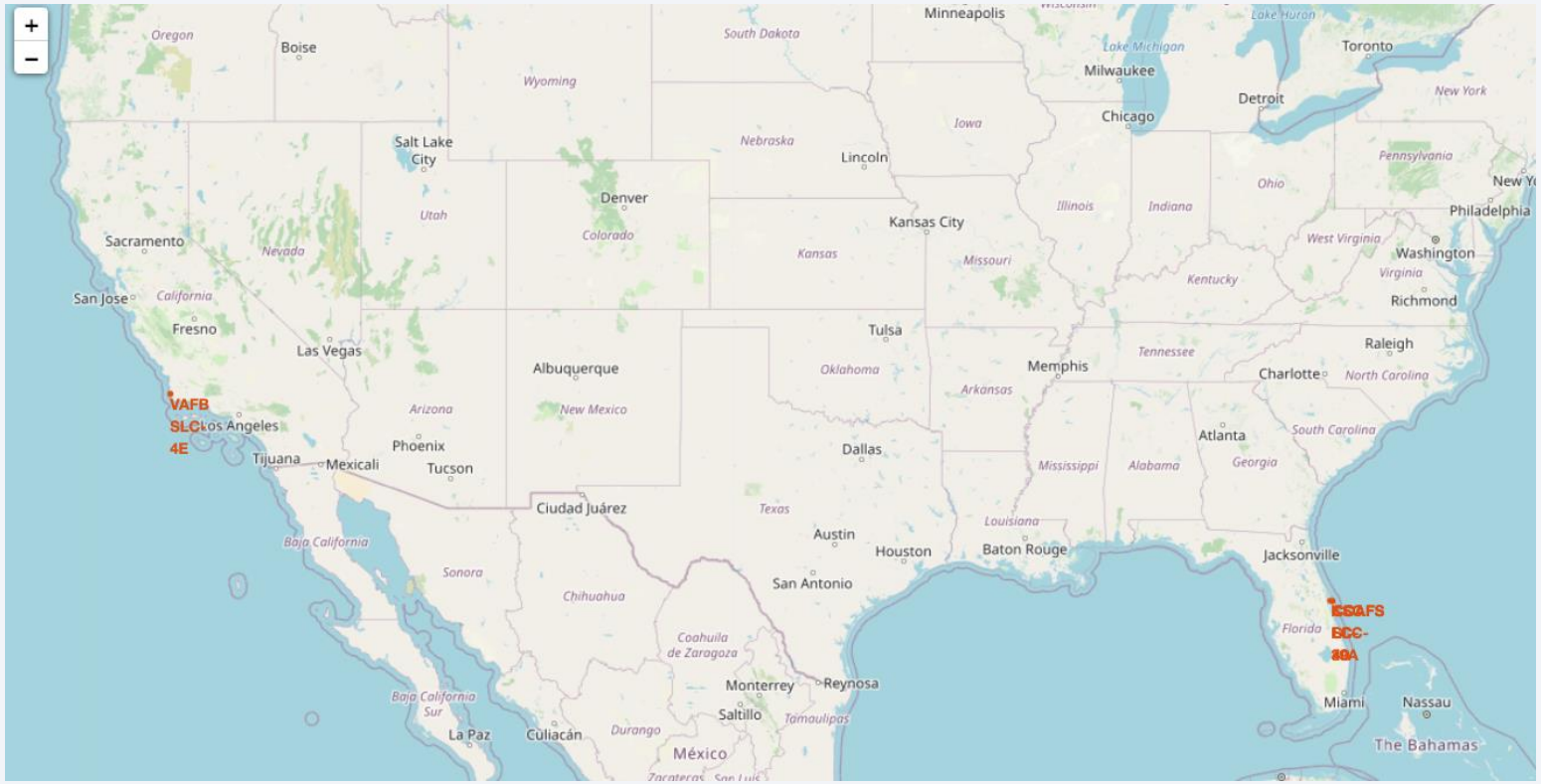
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Ranking of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order in count

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


Section 3
Launch Sites
Proximities Analysis

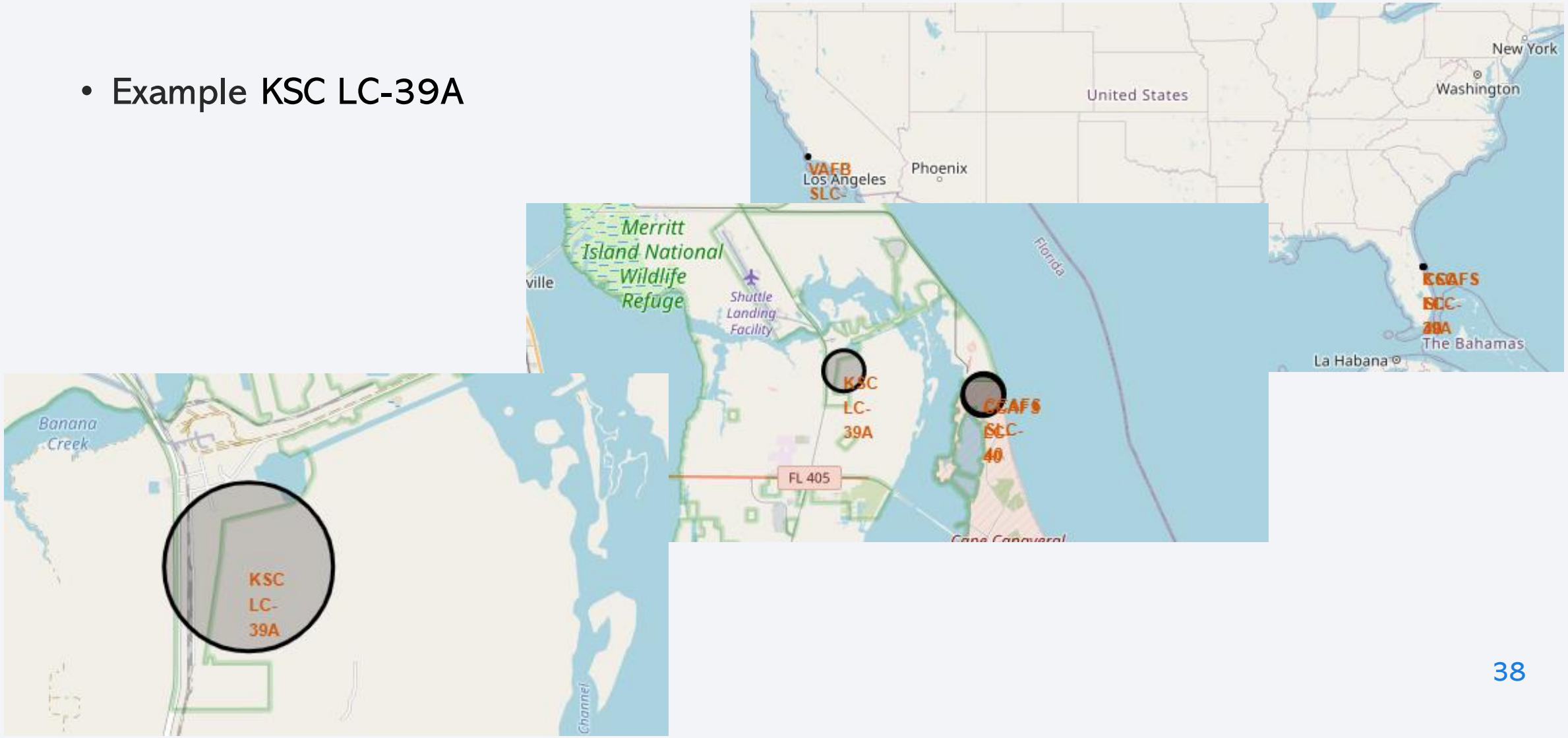
All Launch Site in USA



- Launch sites are near coastline of east and west

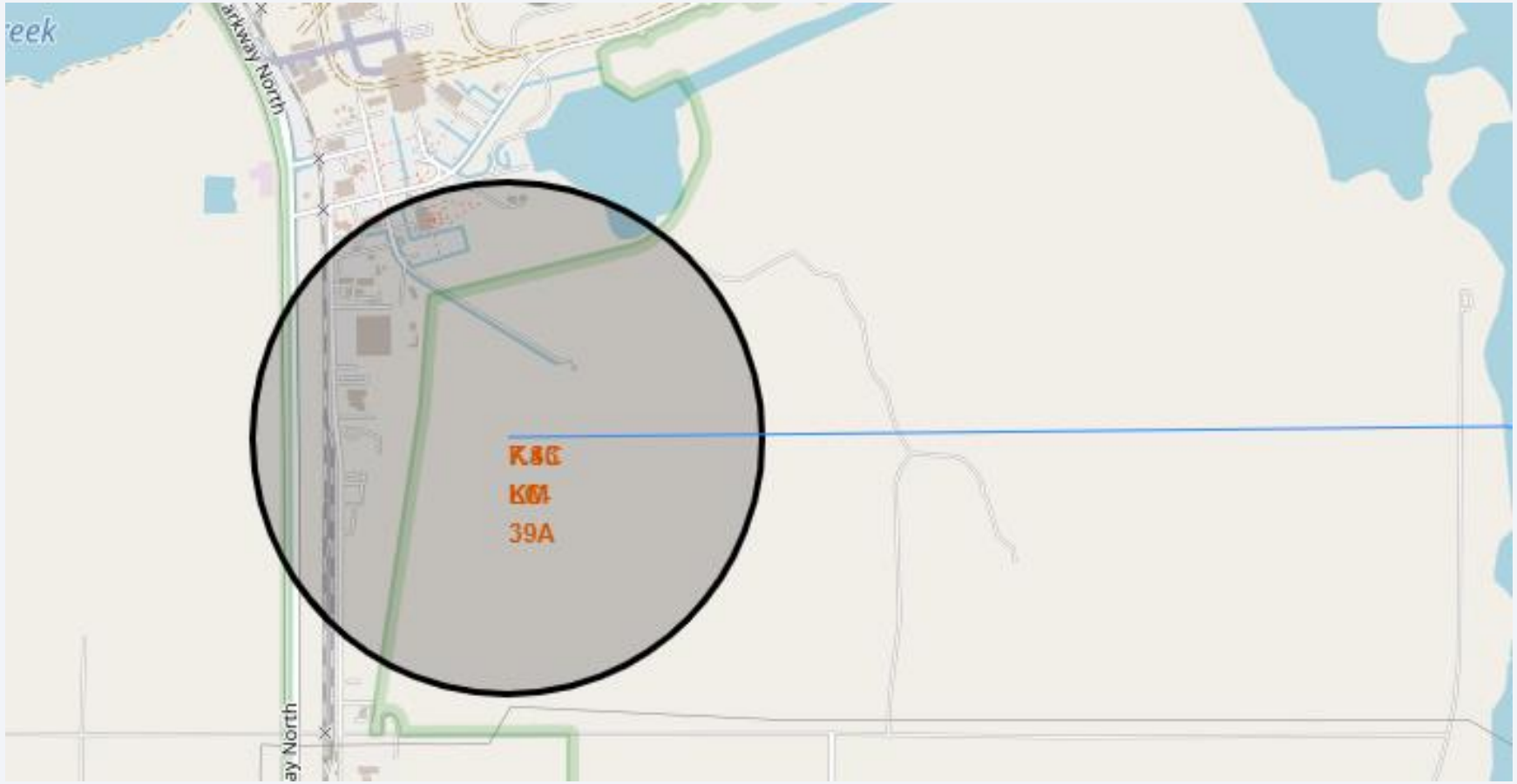
Launch Outcomes by Site

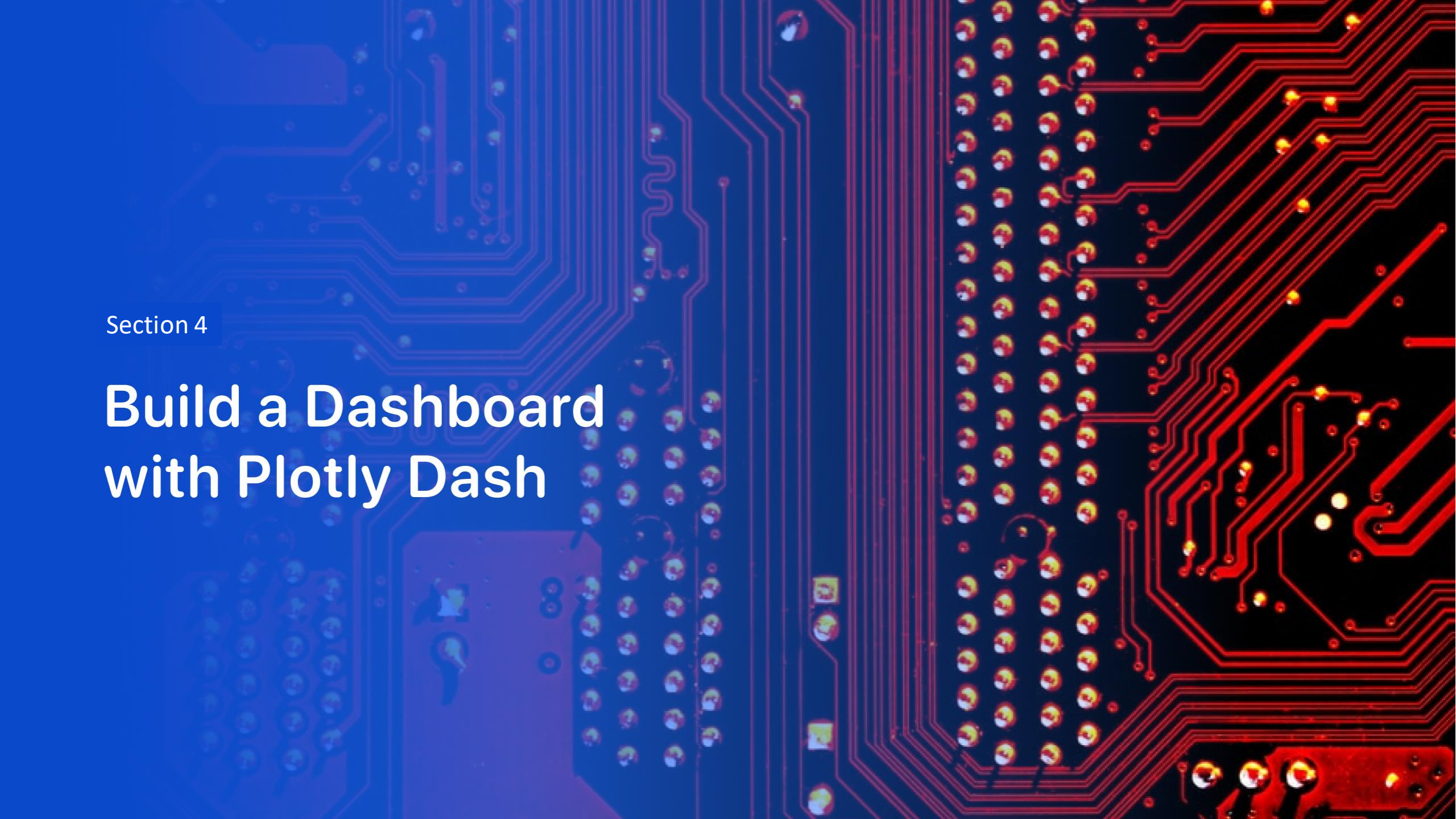
- Example KSC LC-39A



KSC LC-39A

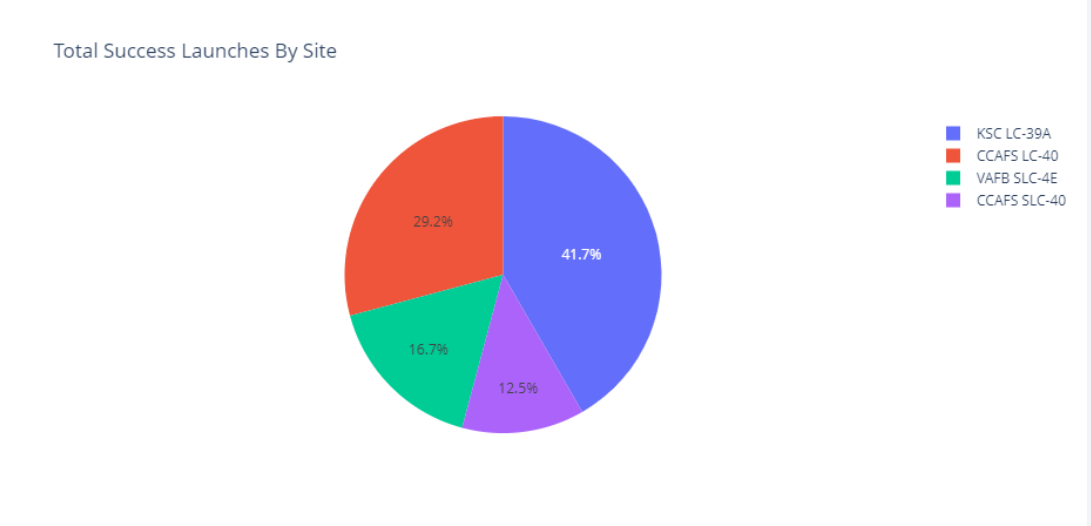
- KSC LC-39A is good location where there is no building nearly




Section 4
Build a Dashboard with Plotly Dash

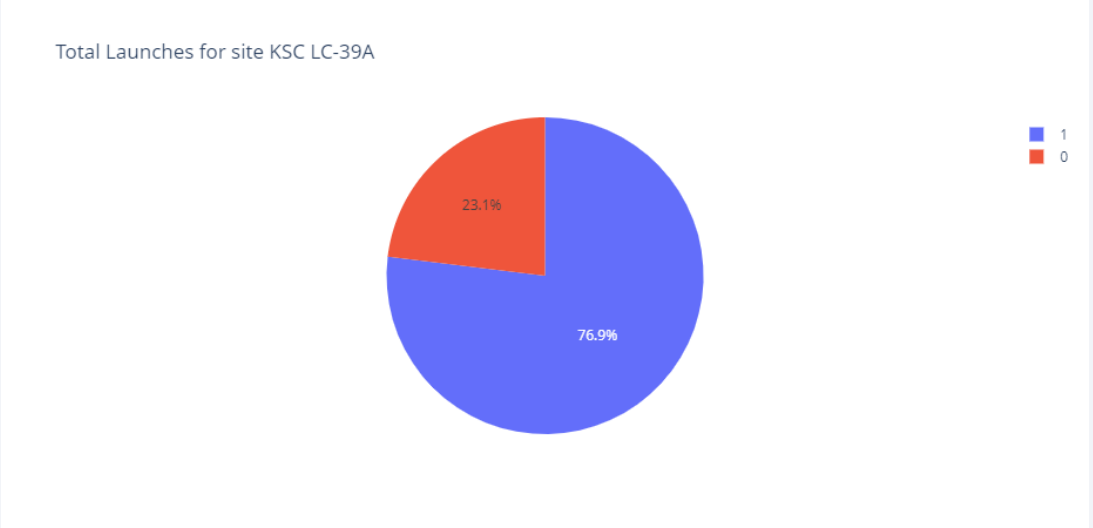
Successful launch by site

- The key factor of success of mission is the location of launch site



Launch success rate

- The KSC LC-39A has highest 76.9% of launch success rate.



Payload vs Launch Outcome

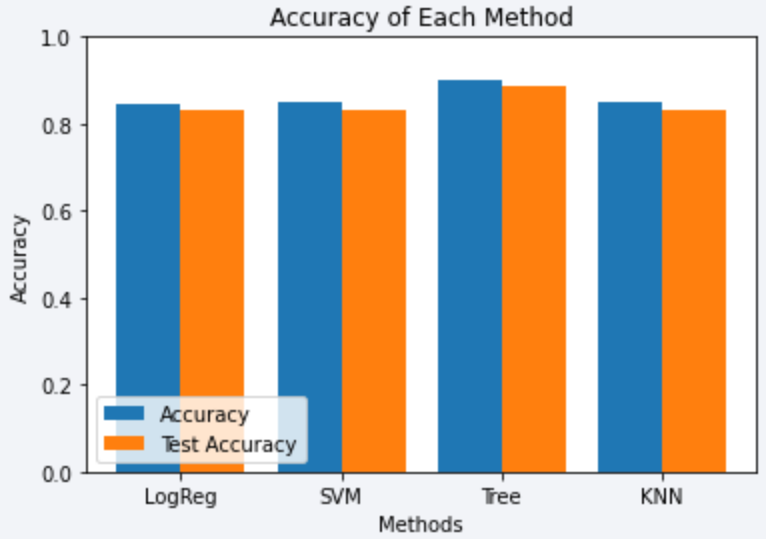
- The most successful combination is:
  - Payloads under 6,000kg
  - FT boosters

Section 5

Predictive Analysis (Classification)

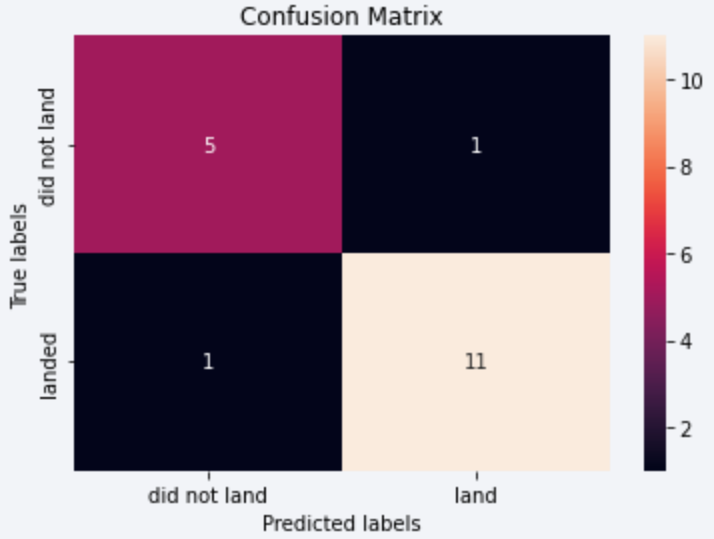
Classification Accuracy

- Accuracy rate of 4 models are plotted
- Decision Tree Classifier give the highest classification accuracy rate 88.8%



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the large numbers of true positive and true negative compared to the false ones.



Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC 39A
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time
- Decision Tree Classifier is the best model to do prediction

Appendix

- Working GitHub: https://github.com/chowkwokho/AppliedDataScienceCapstone

Thank you!