IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ailar EsmaeiliZadeh
2024-02-17

Outline

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

Executive Summary

- Summary of methodologies
  - Data collecting using web scraping and Space X API
  - Exploratory Data Analyze (EDA)
  - Machine Learning prediction
- Summary of all results
  - Machine Learning Prediction showed the best model for predict successes of lunching

Introduction

- Project background and context
  - Our goal is to use data to predict whether SpaceX will attempt to land a rocket or not, This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Vital answers
  - Estimate total cost for lunches, by predicting successful landings
  - Find best sites to make lunches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data are collected from SpaceX website https://api.spacexdata.com/v4/launches/past
  - And Web scraping via Wikipedia https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was collected normalize and divided in train and test data sets
  - Four classification models test accuracy of each model

Data Collection

- Data set were collected from Space X site and from Wikipedia with web scraping. They are listed below:
  - https://api.spacexdata.com/v4/launches/past
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

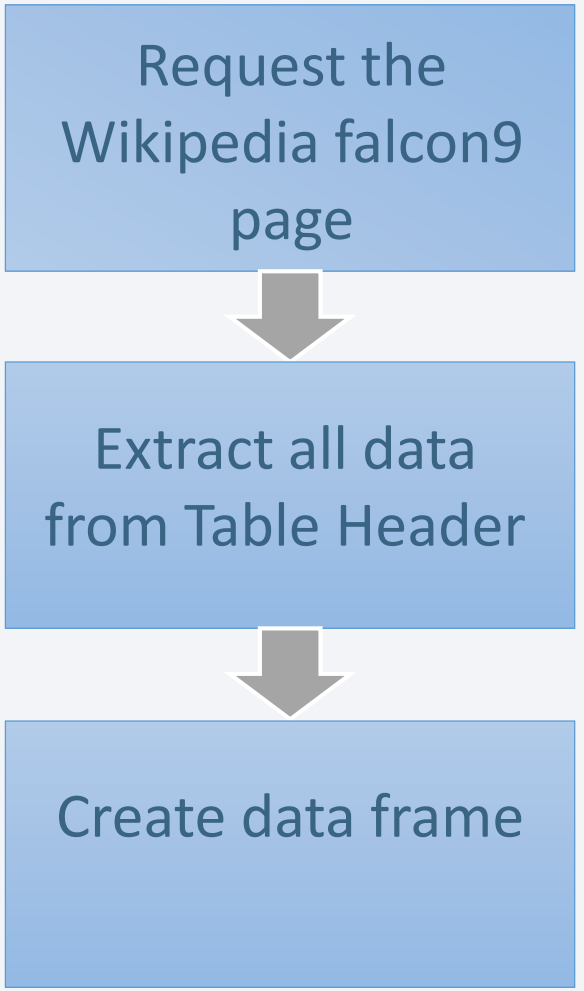
Data Collection – SpaceX API

- SpaceX offers a public API and how data can obtained and used. Based on this flow chart data collected.

- Notebook link

Data Collection - Scraping

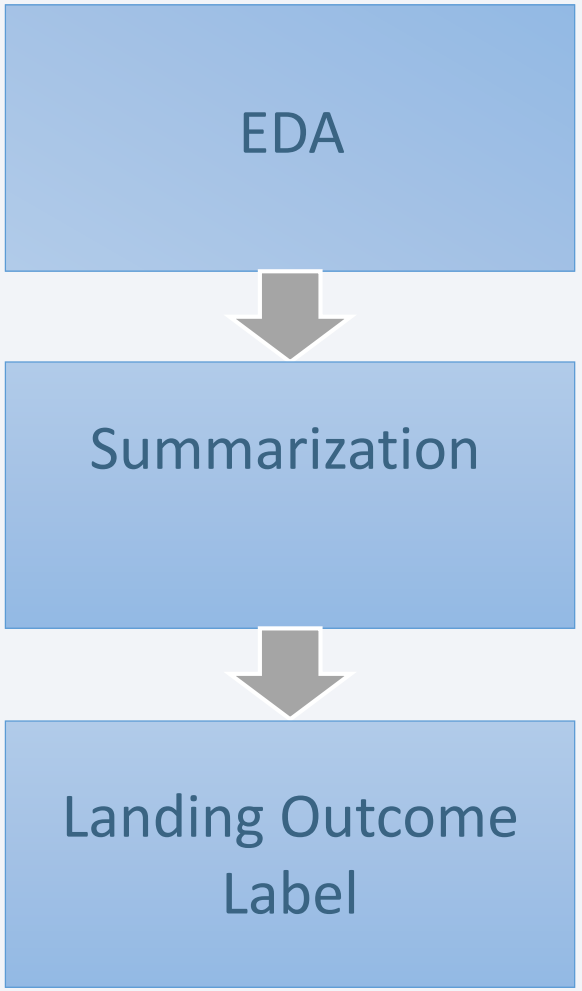
- Obtain some Data from Wikipedia about SpaceX lunches. Data mines from Wikipedia and uses as this flowchart.



- Notebook Link

Data Wrangling

- Firs some Exploratory data Analysis performed on data
- Then occurrence of each orbit and mission outcome per orbit calculated and summarized.
- Outcome label created.



- Notebook Link

EDA with Data Visualization

- Scatter and Bar chart were used to visualize relationship between different values:
  - Payload Mass & Flight Number
  - Lunch site & Flight Number
  - Lunch Site & Payload Mass
  - Orbit & Flight Number
  - Payload & Orbit

- Notebook Link

EDA with SQL

- names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA
- total payload mass carried by boosters launched by NASA (CRS)
- average payload mass carried by booster version F9 v1.1
- first successful landing outcome in ground pad was achieved
- names of the boosters which have payload mass greater than 4000 but less than 6000
- number of successful and failure mission outcomes
- names of the booster versions which have carried the maximum payload mass
- records which will display the month names, failure landing outcomes in drone ship in 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
- Notebook Link

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters used to visualize data on Folium map
- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

Notebook Link

Build a Dashboard with Plotly Dash

- Various ways to visualize data that used in this part
  - Pie chart to visualize per lunch site outcome percentage
  - Payload range
  - Scatter plot

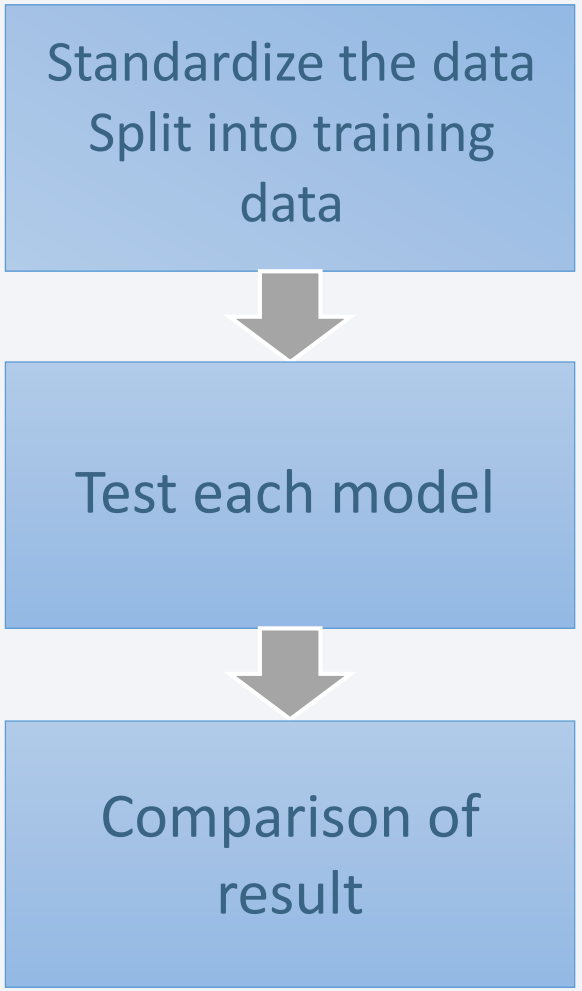
This outputs help to easily analyze data and relation between values “ Payload “ and “Lunch site’

The main result is find the best place to lunch

Notebook Link

Predictive Analysis (Classification)

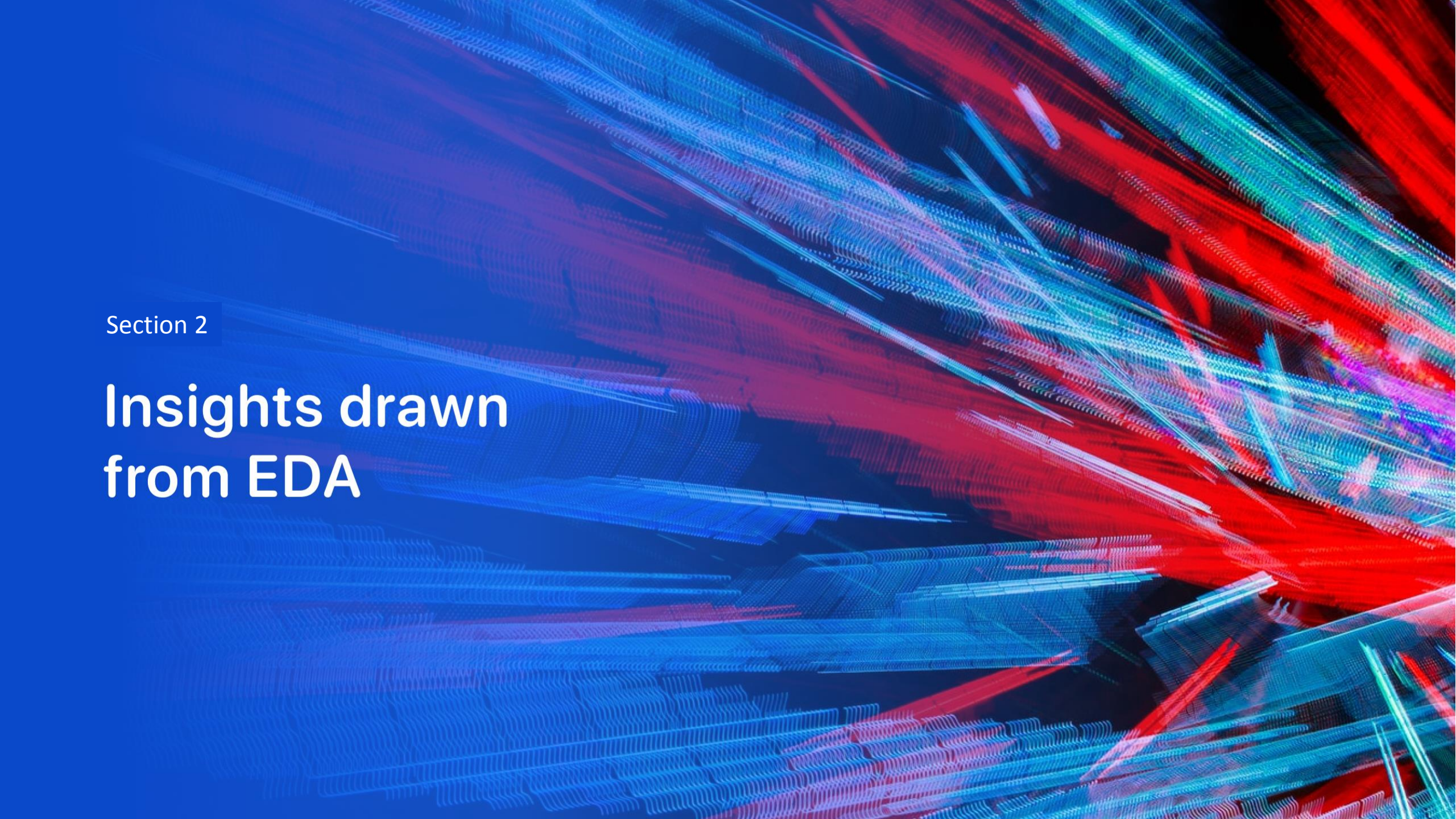
- Four classification models were used
  - Logistic regression
  - K Nearest Neighbors
  - Decision Tree
  - SVM



Notebook Link

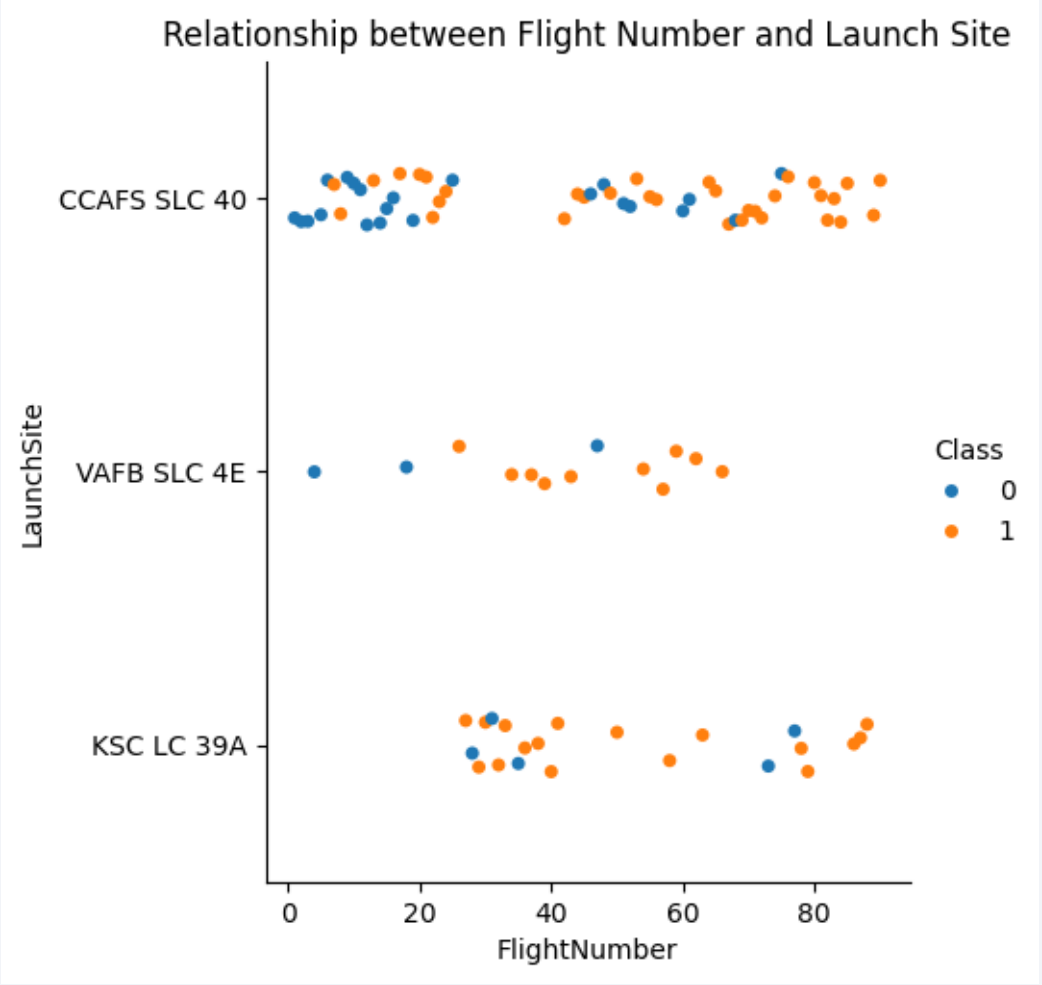
Results

- 4 sites were uses for SpaceX lunches.
- The first successful landing outcome happened in 2015
- Almost 100% mission outcome were successful
- The number of landing outcomes becomes as better as years passed
- Based on interactive Analytics it is clear that lunch sites were near sea, and most lunches happened at east coast.
- Predictive analysis showed that Decision tree in the best model to prediction with accuracy over 87%


Section 2
Insights drawn from EDA

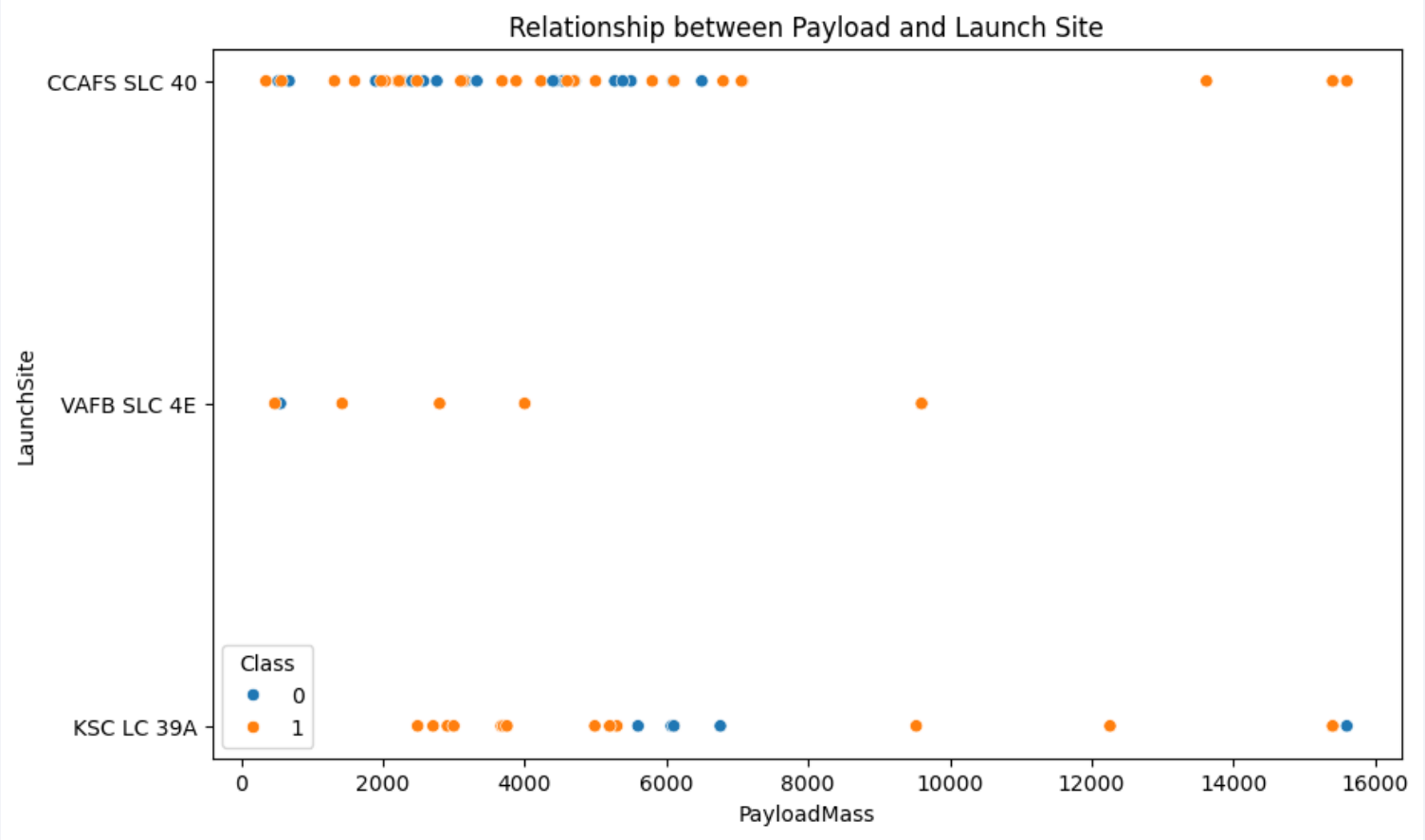
Flight Number vs. Launch Site

- Based on this chart it is possible to verify the best lunch site is upper one, CCAFS SLC 40, the next one could be KSC 39A and less success lunch site could be SLC 4E.
- It is showed that success rate improved over time.



Payload vs. Launch Site

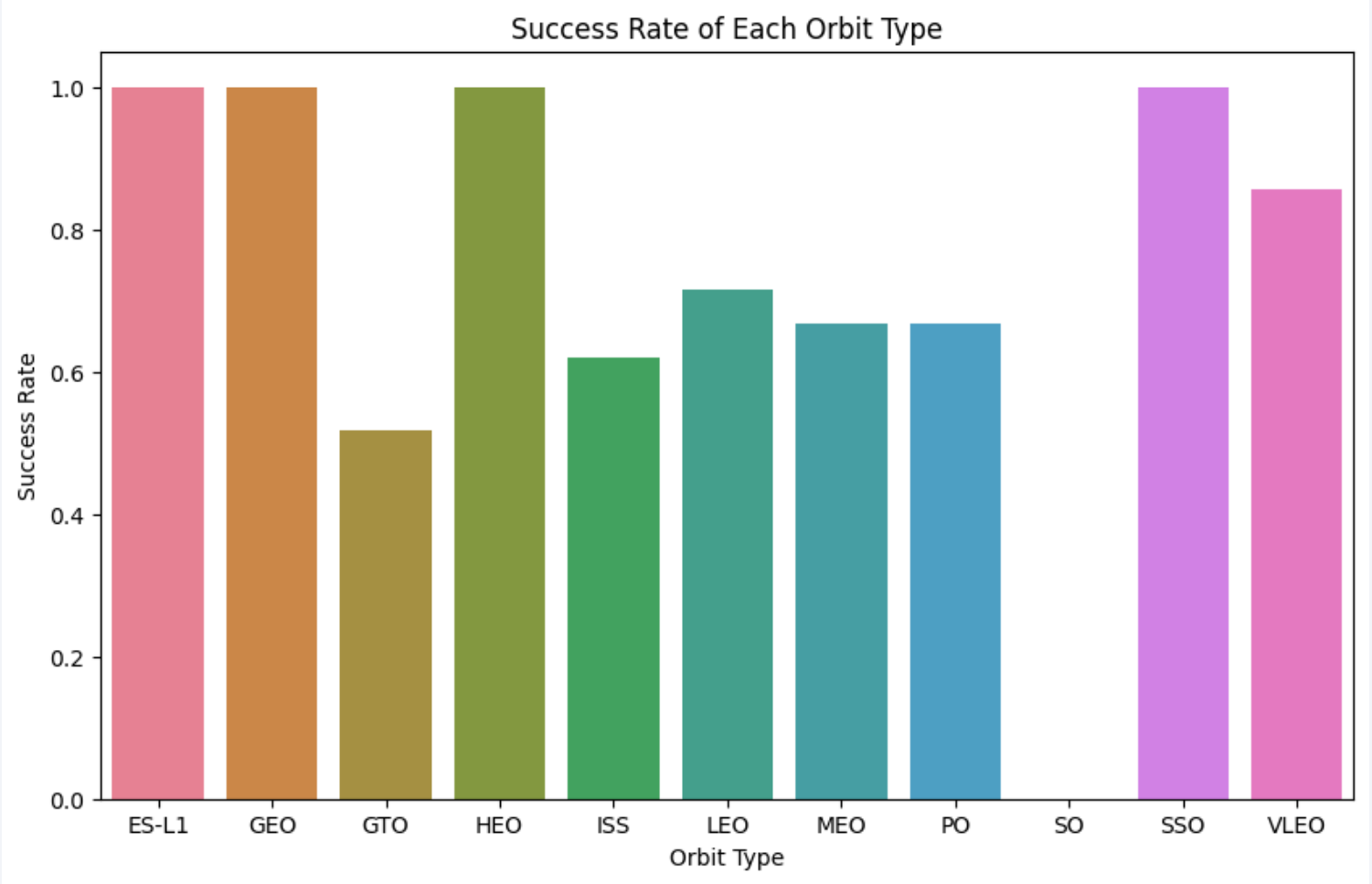
- Payloads over 9000 were in great success rate, and in other side, payloads over 12000 aren’t use in SLC 4E lunch site but it was completely successful in SLC 40.



Success Rate vs. Orbit Type

- Highest success rate is for:
  - ES-L1
  - GEO
  - HEO
  - SSO

  And after them they followed by
  - VLEO
  - LEO
  - ….

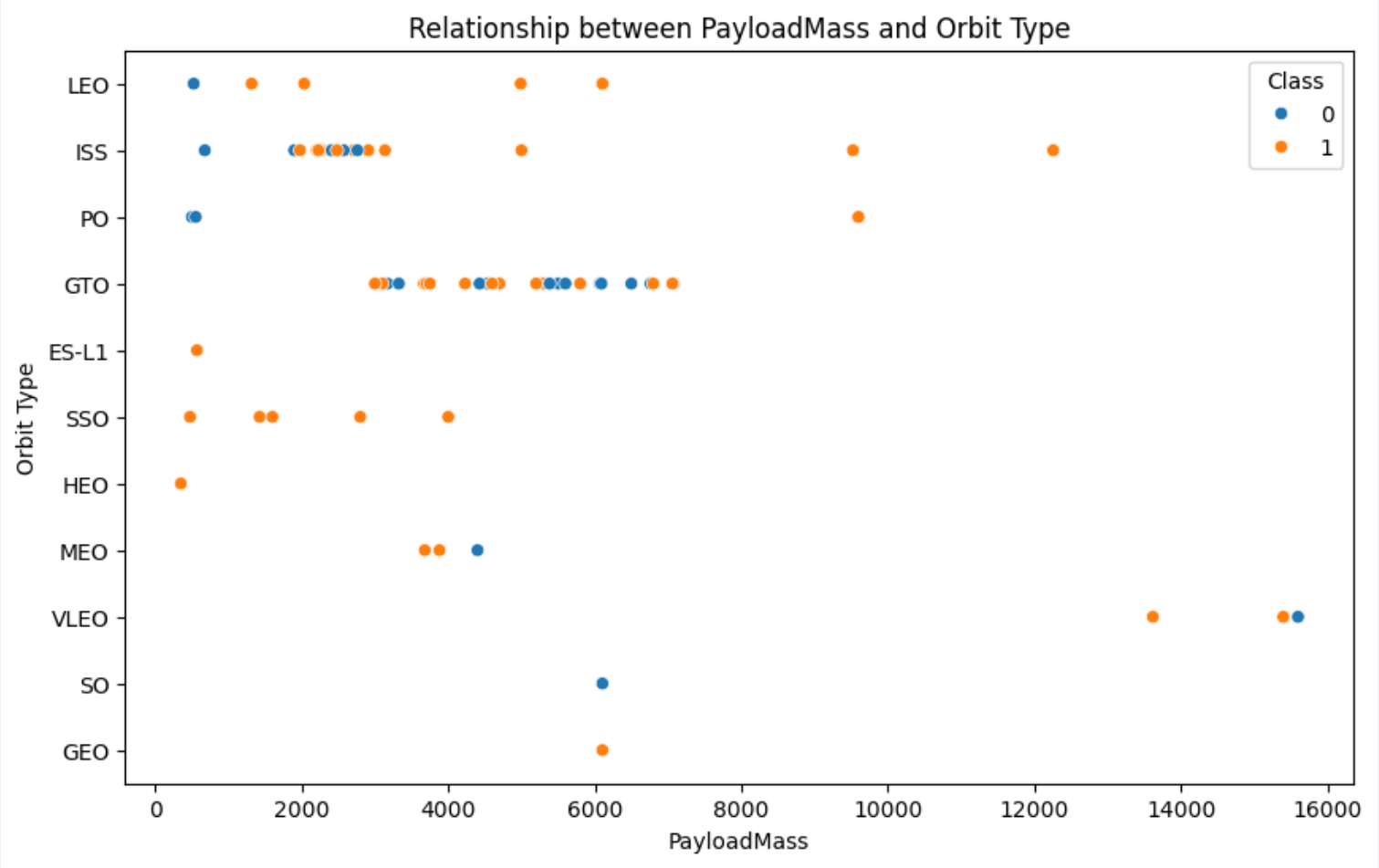


Flight Number vs. Orbit Type

- Success lunches are getting more over time in every Orbit.
- VLEO's frequency is too high and seems there is a focus on it.

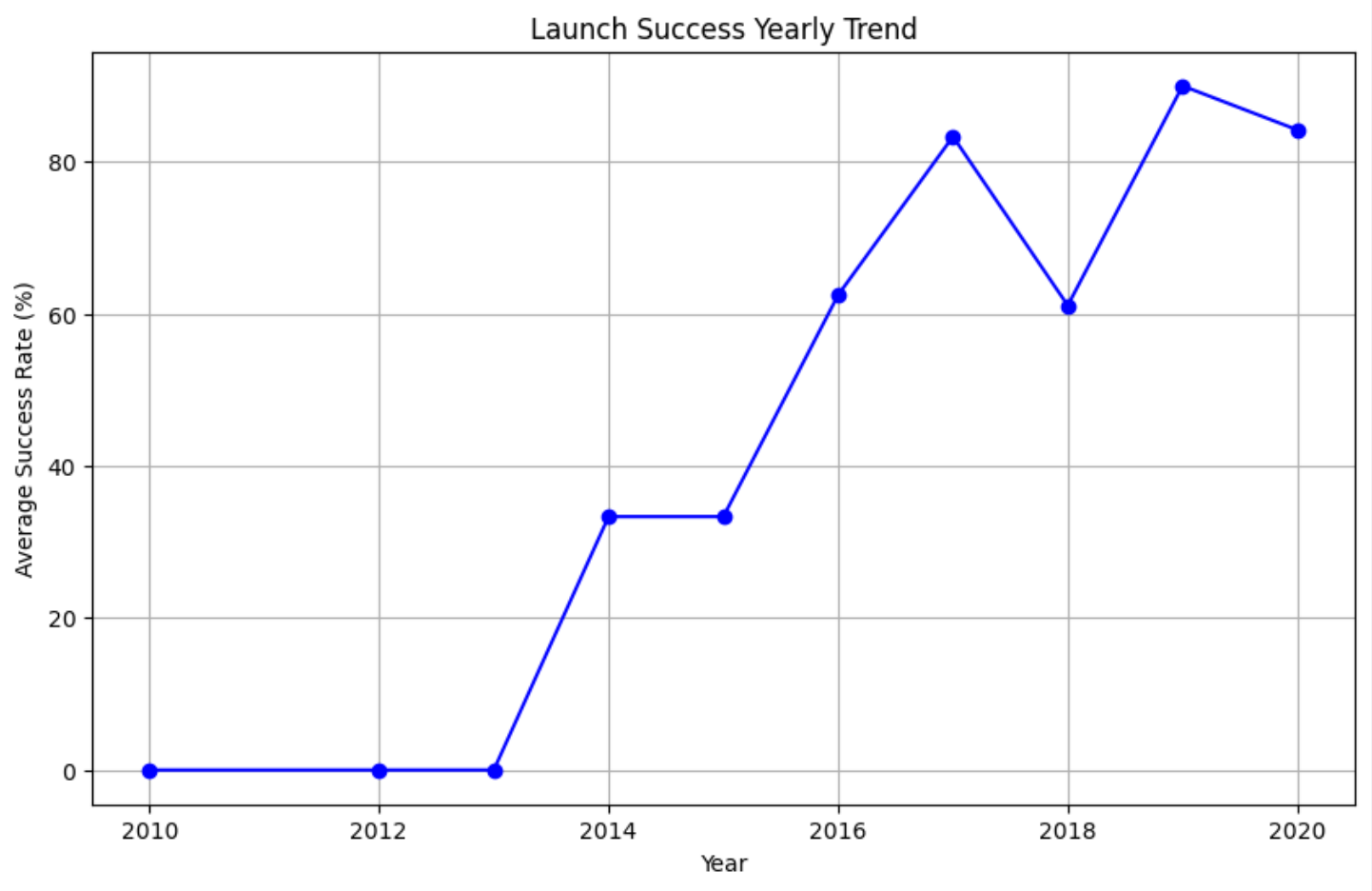
Payload vs. Orbit Type

- ISS has wide range of payload mass with improving success rate.
- GTO and SO orbits has less lunch.
- At all it seems there is no relation between payload mass and orbit type.



Launch Success Yearly Trend

- Average success rate is improved over years in range of 2013-2020.



All Launch Site Names

- By using Unique method in “Lunch_site” values, we can find that there is four unique location for this data.

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

Launch Site Names Begin with 'CCA'

- There is a list of lunch site names that start with “CCA”

]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- By summing all payloads, The Total payload mass is calculated as bellow

TotalPayloadMass
45596

Average Payload Mass by F9 v1.1

- This value is a average of payload mass by booster F9 V1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- This date is earliest time that Ground landing was successful.

Successful Drone Ship Landing with Payload between 4000 and 6000

- List of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000. this list could help to find best booster that specific range.

Booster_Version
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- In this out come there is vividly show that successful missions are extremely hight and failure one isn’t too matter.
- Grouping mission outcomes and counting records for each grouo can lead a chart

Mission_Outcome	Total
Failure (in flight)	1
Success	98

Boosters Carried Maximum Payload

- Based on this list, maximum payload caring boosters can clearly chooses:

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

- List has just only 2 failed launch

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

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

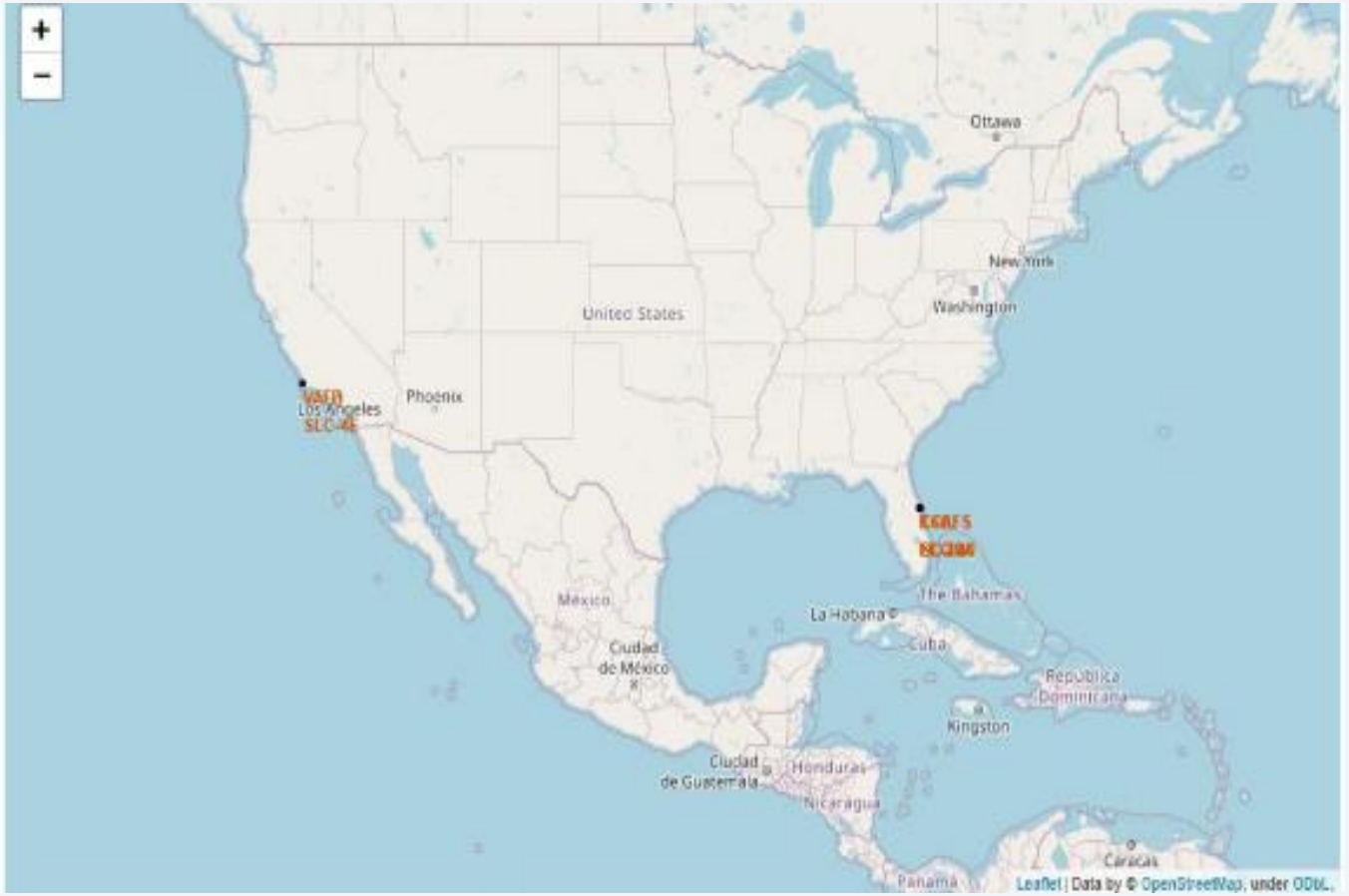
- Based on this view, we realize that “No attempt” must add to account

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


Section 3
Launch Sites
Proximities Analysis

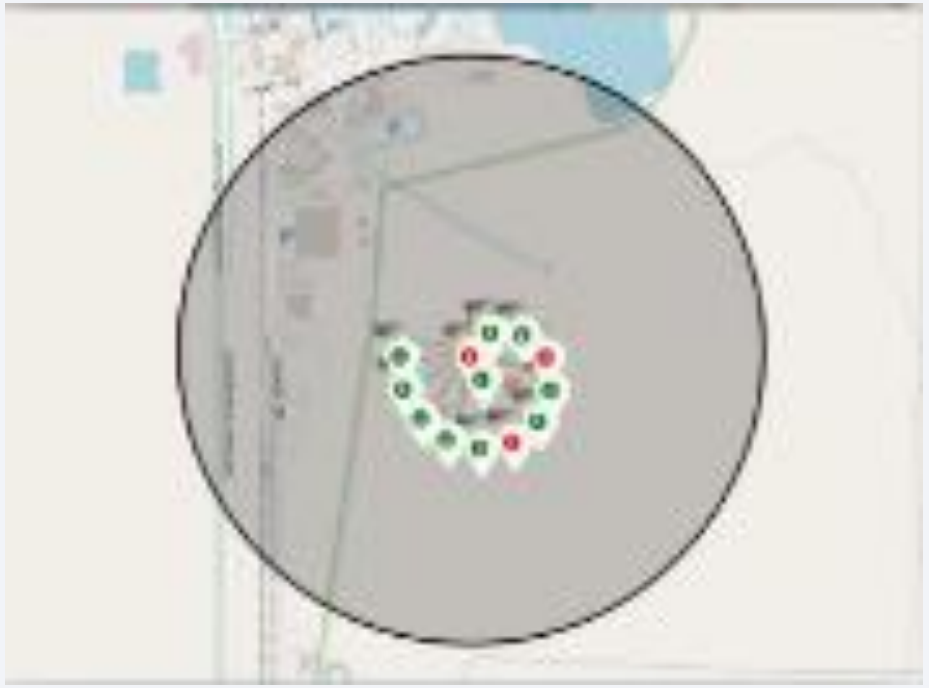
All Lunches Sites

- Lunch sites are near the sea and although near roads



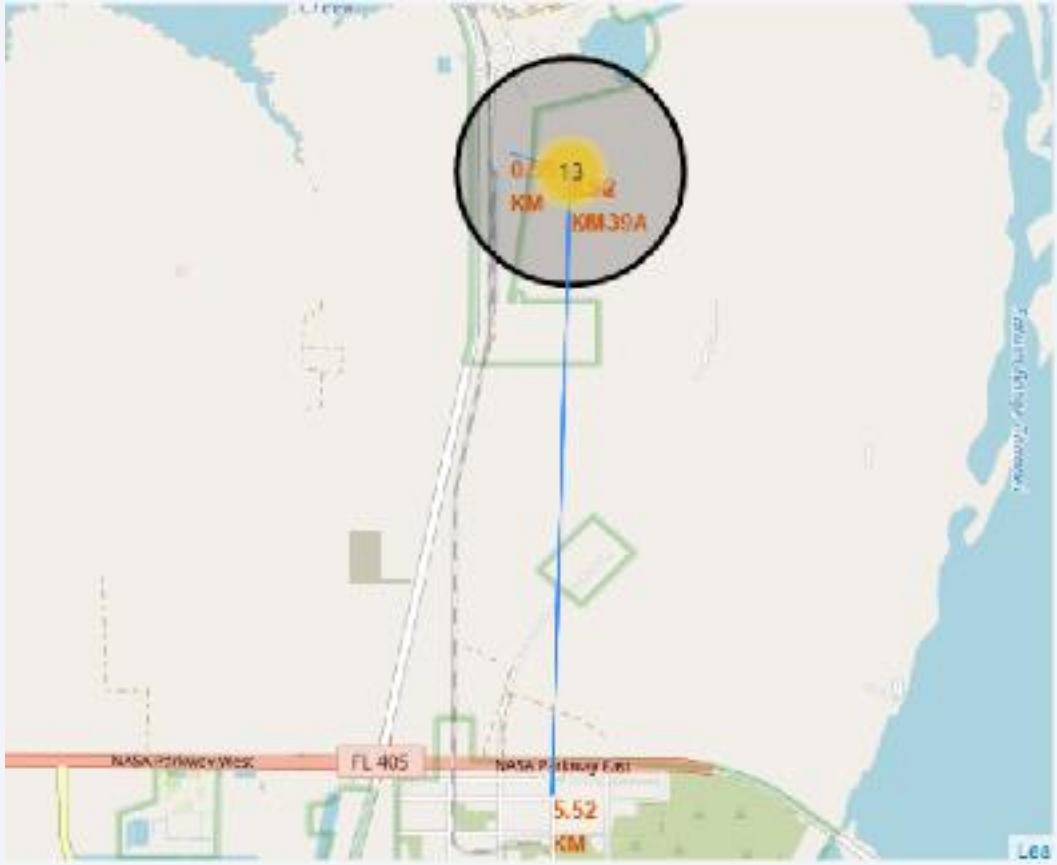
Lunch outcomes

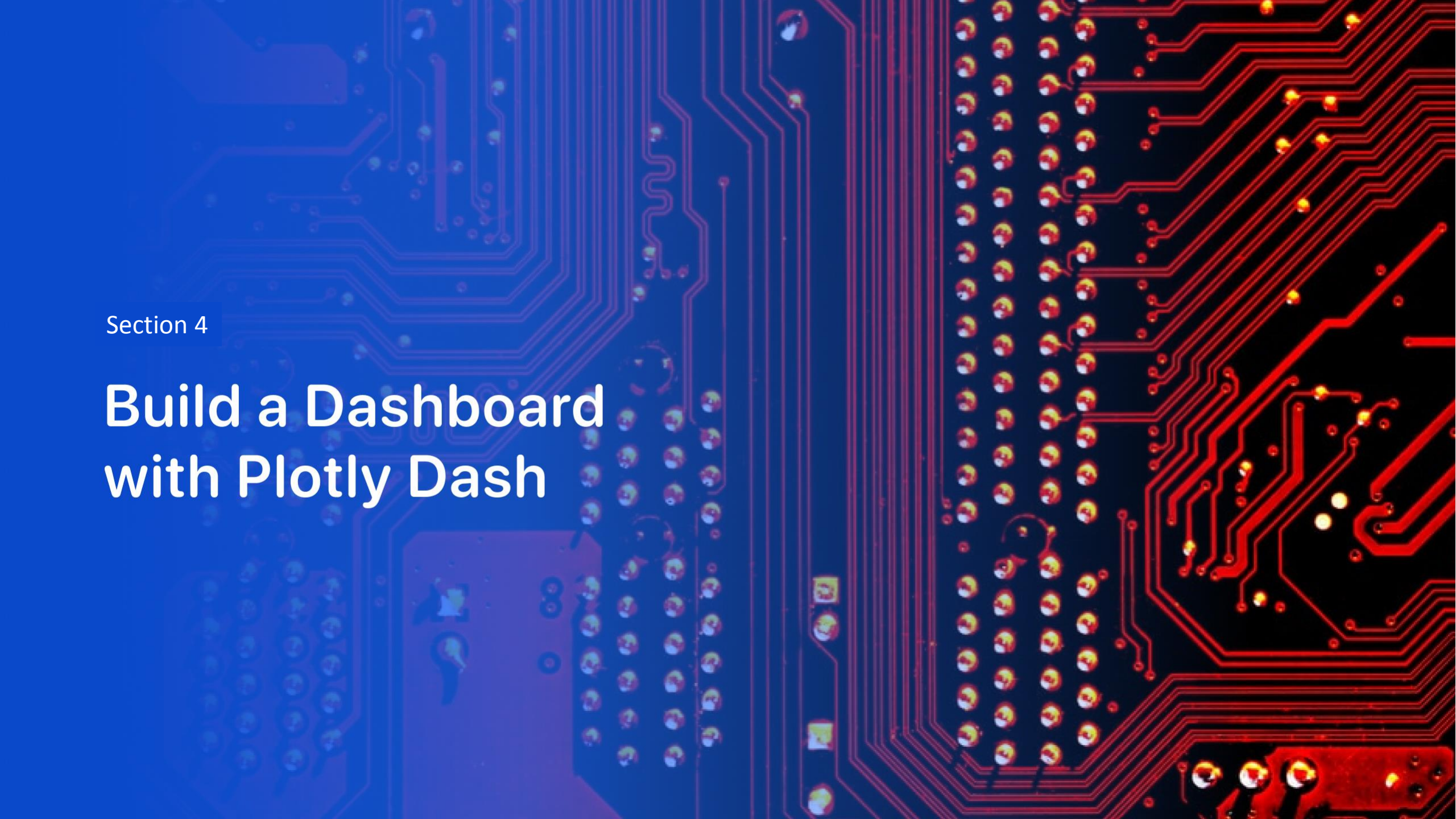
- Green dots are successful sites and red ones are unsuccessful



Distance and Logestic

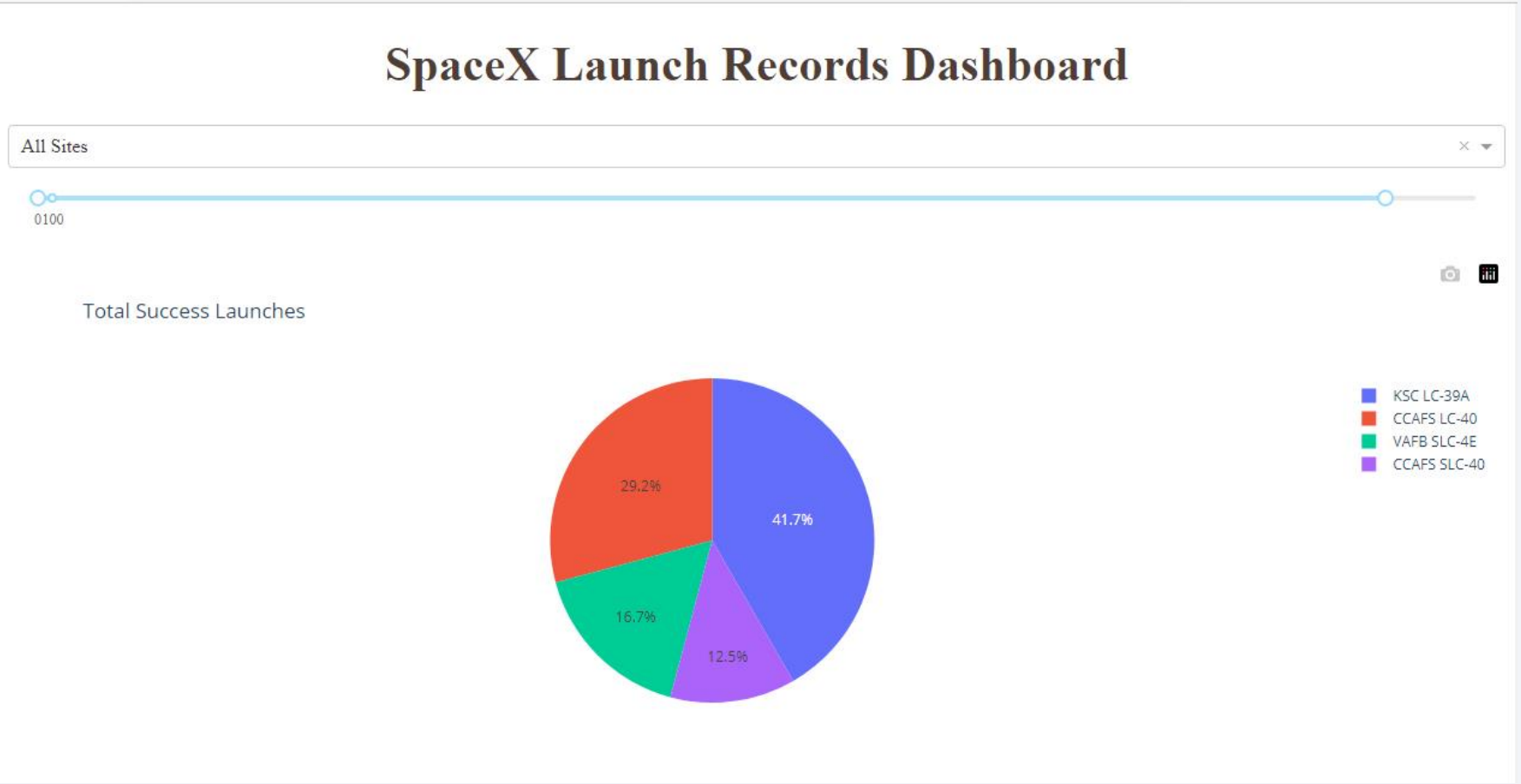
- There complete connection with road and railway transportation but sites are far from cities.




Section 4
Build a Dashboard with Plotly Dash

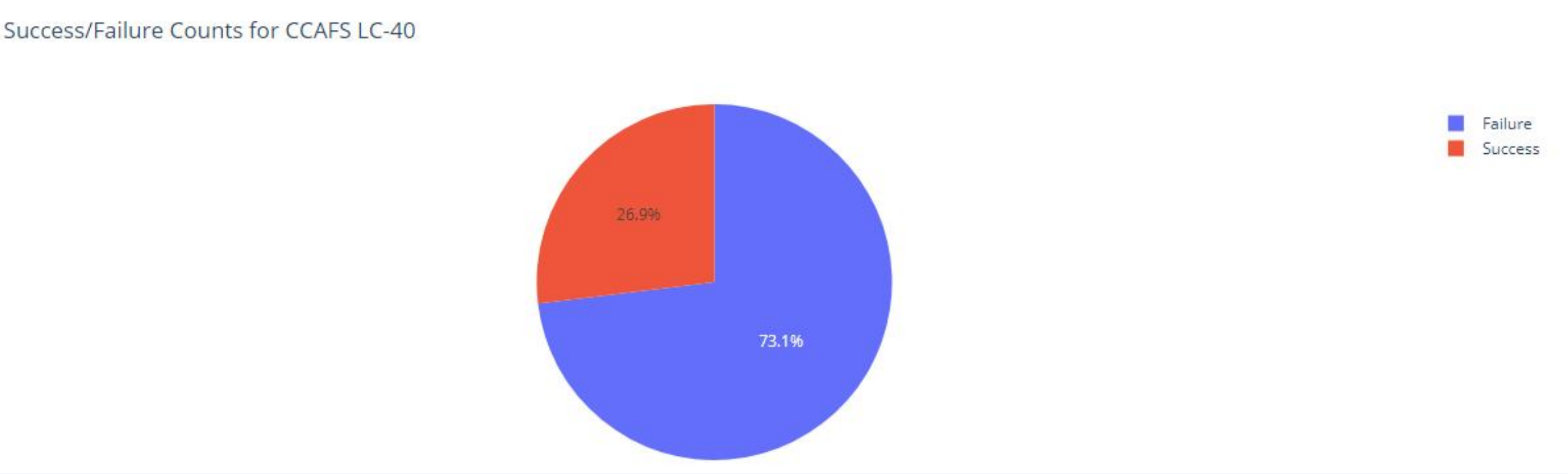
Successful Lunches

It seems that places play key roles on lunches result and success of mission.



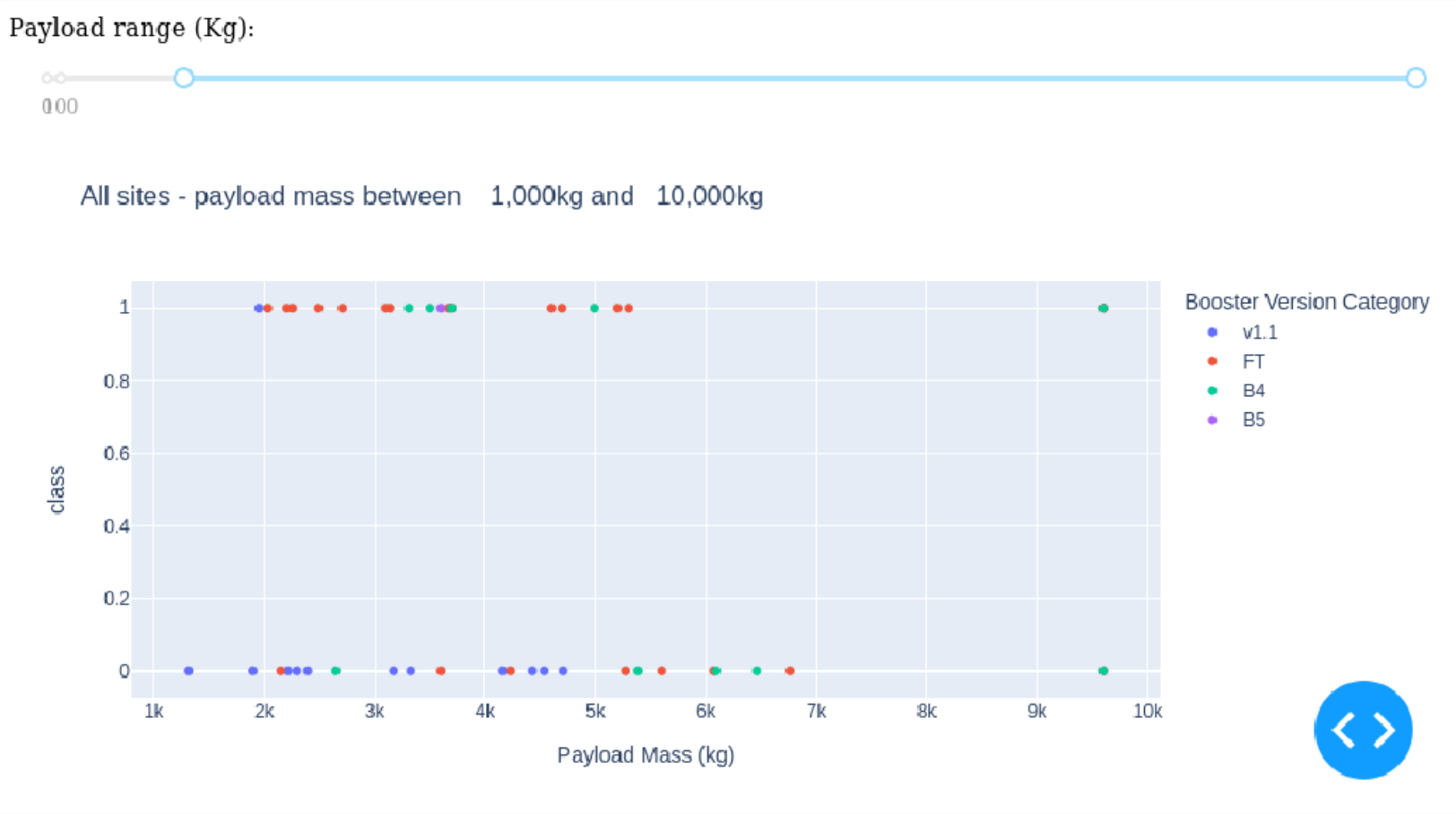
Highest success lunches

Almost 73% of lunches were successful in this site



Payload vs. Lunch outcomes

Payload under 6kb are more success ones.
FT boosters are more successful in this range

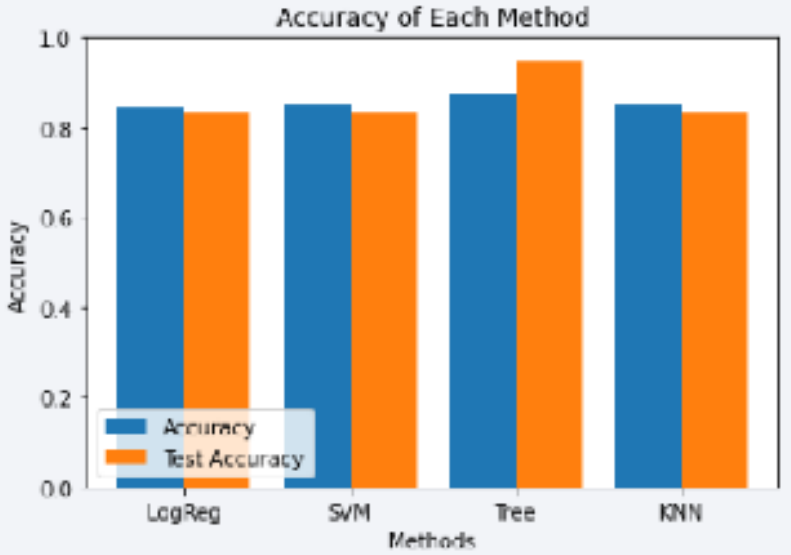


Section 5

Predictive Analysis (Classification)

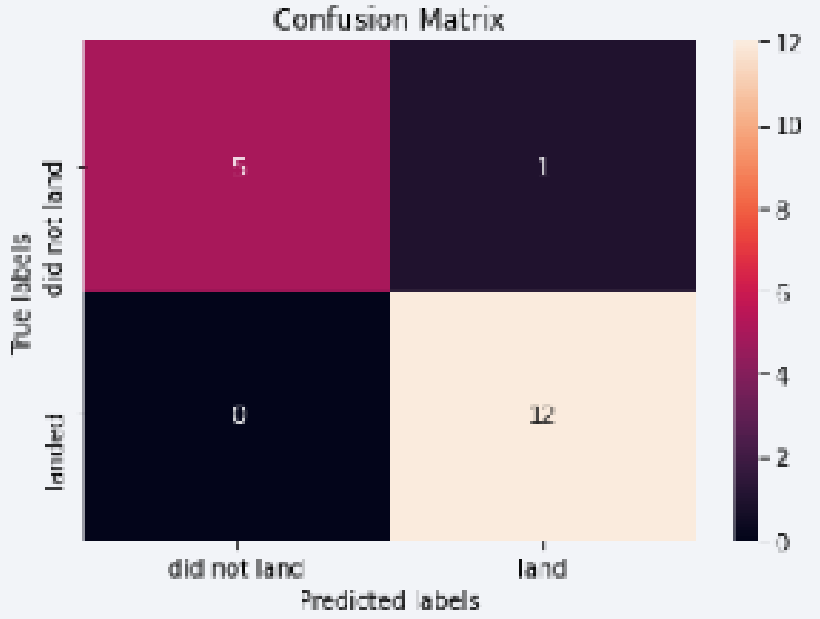
Classification Accuracy

- Four classification method were tested as attached chart . Based on this visualize, decision tree could make more accurate results.



Confusion Matrix

- Based on this matrix, true positive and true negatives are bigger and they prove accuracy of decision tree



Conclusions

- Various data analyzed, among this procces
- The best lunch site is CCAFS Lc-40
- Most of the lunches were successful at all
- Decision tree classifier can use for predict with better accuracy

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!