IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Worawit Saetan>
<12/30/2021>

Outline

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

Executive Summary

- In this project, we create the model to predict the probability of a successful landing of the first stage of SpaceX's Falcon 9 rocket.
- The related data was collected from public resources.
- Our model has 80%+ accuracy based on test data.

Introduction

- SpaceX can offer orbital launches at less than half the price of its competitors due to the fact that SpaceX can reuse the Stage 1 with high success rate.
- By predicting if the booster will land successfully, SpaceY will gain a competitive advantage for new launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Web scraping from the SpaceX data portal API and Wikipidia.org
- Perform data wrangling
  - Formatting the data and identification of target variable
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - 4 models was tuned and evaluated

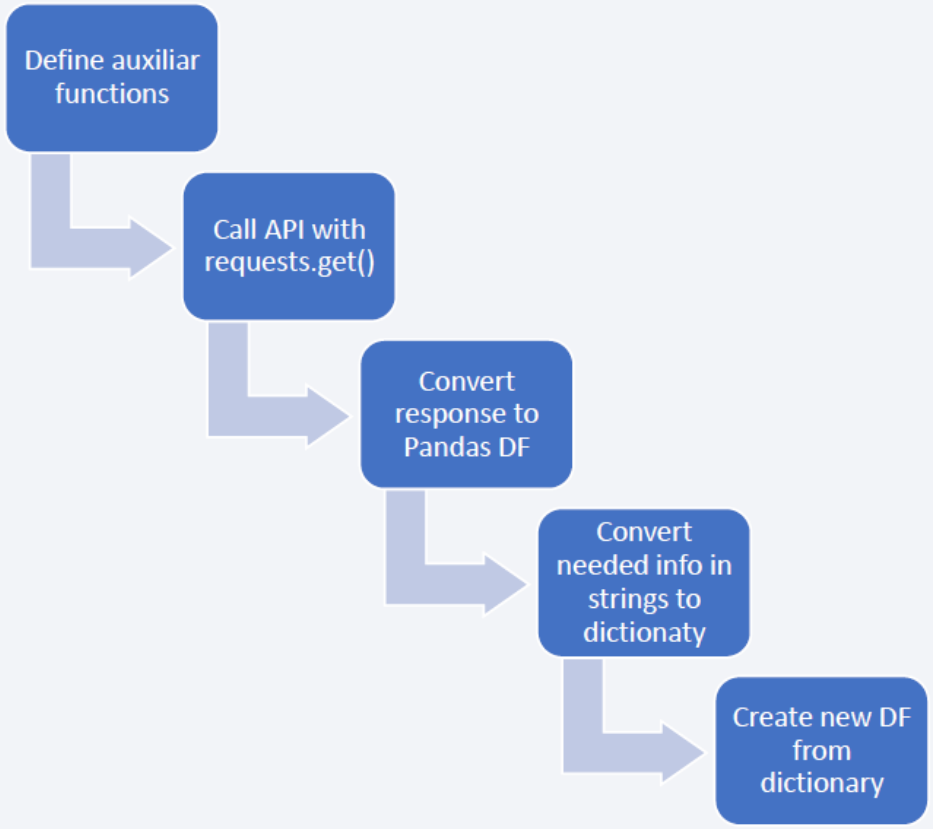
Data Collection

- Data sets were collected from 2 different public sources
  - Wikipedia
  - SpaceXdata.com
- For Wikipedia
  - List of Falcon/ 9/ and Falcon Heavy launches - Wikipedia
- For SpaceX data portal API
  - https://api.spacexdata.com/v4/launches/past

Data Collection – SpaceX API

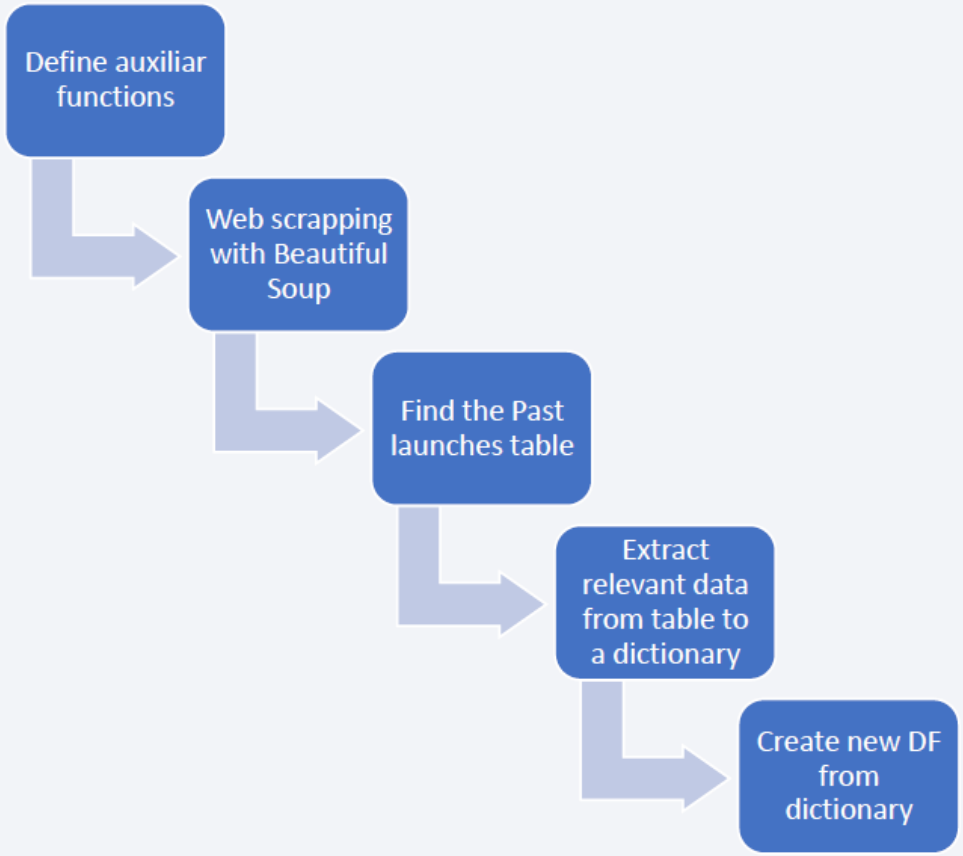
- Data from 2010 – 2020 was collected from SpaceX API

- A series of functions were used to convert entries with information in string format to a more manageable format.
- Unnecessary columns were removed



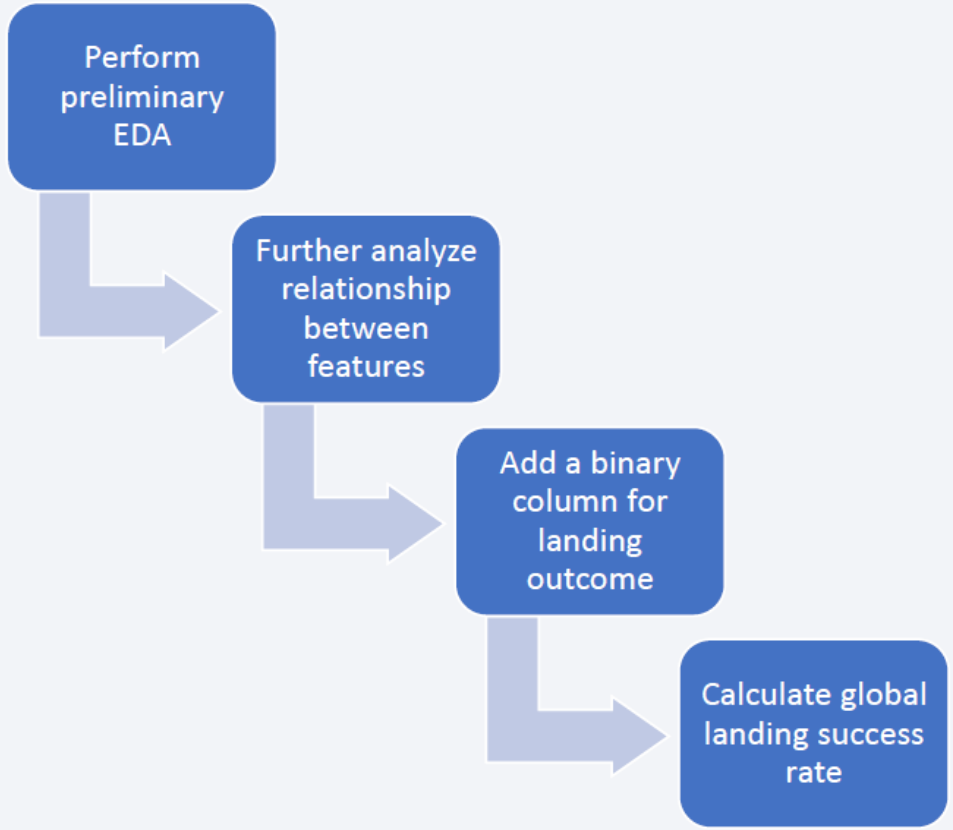
Data Collection - Scraping

- Data was collected form Wiki
- Web scraping by using Beautiful Soup library
- Data will be converted into panda frame for further use propose



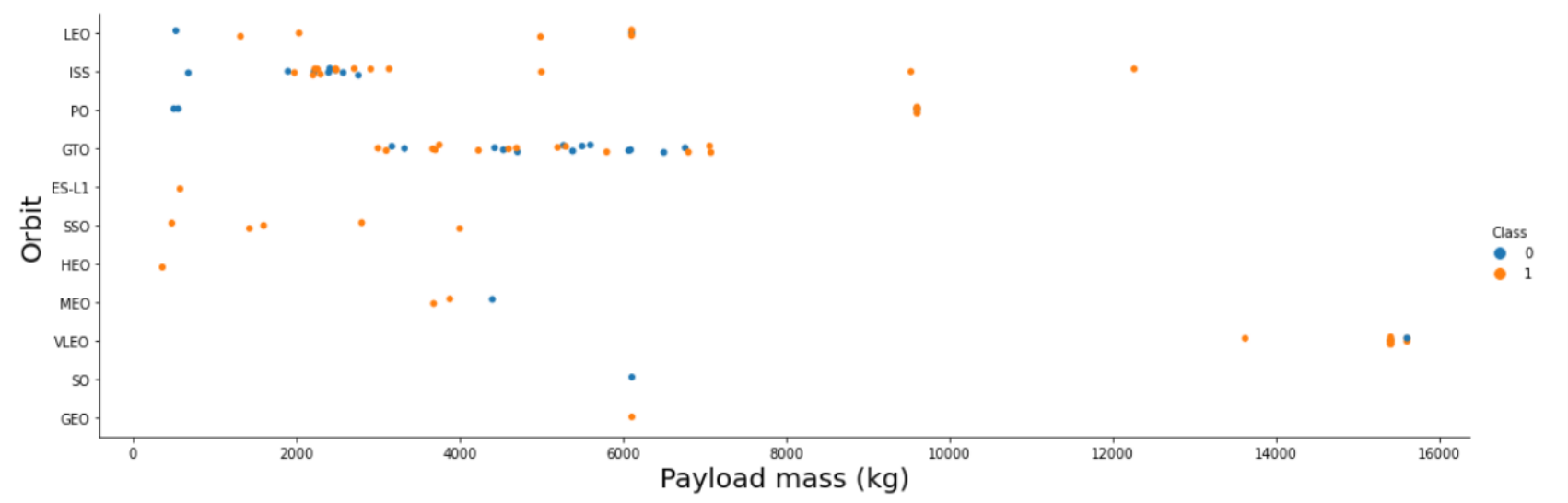
Data Wrangling

- Adding binary column to dataframe to indicate the success outcome
- The landing success rate was 67%.



EDA with Data Visualization

- EDA was performed to select the important features for our models.
- The relationship between launch site, payload mass, orbit and landing success was analyzed via scatter plot.
- The success rate was plotted for payload mass and each orbit.



EDA with SQL

- 10 queries were performed for understanding the data
  1. List all launch sites
  2. List launch sites beginning with “CCA”
  3. Total payload
  4. Average payload for Booster F9
  5. First successful landing on ground
  6. Boosters landed successfully on drone ships with payload between 4000-6000 kg
  7. Total successful and failed missions
  8. Booster that carried maximum payload 15,600 kg
  9. Failed landings in 2015
  10. Most frequent landing outcomes

Build an Interactive Map with Folium

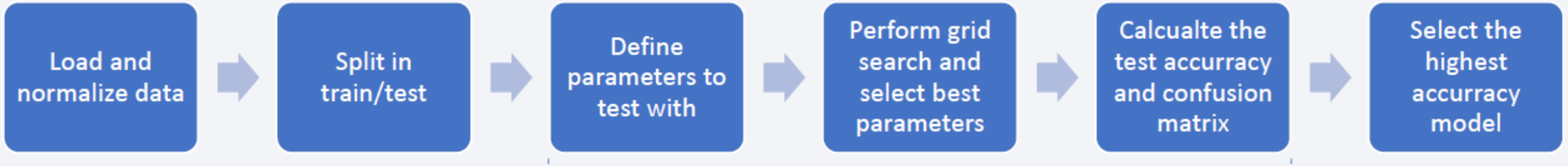
- Geospatial data analysis was performed.
- A map was created with a marker of location of each launch
- The distance to the surrounding elements (cities, highways) was calculated.

Build a Dashboard with Plotly Dash

- Create a dynamic dashboard.
- Add dropdown list for selecting launch site or other things.
- Add a scatter plot to show landing success rate and launch site.
- We can gain some insights from this dashboard.

Predictive Analysis (Classification)

- The data was split into training data 80% and test data 20%
- 4 models were built. Including logistic regression, SVM, Decision tree and KNN
- We use the GridSearchCV to find the best hyperparameter sets for each model
- Then , the accuracy was calculated on test data to compare the models

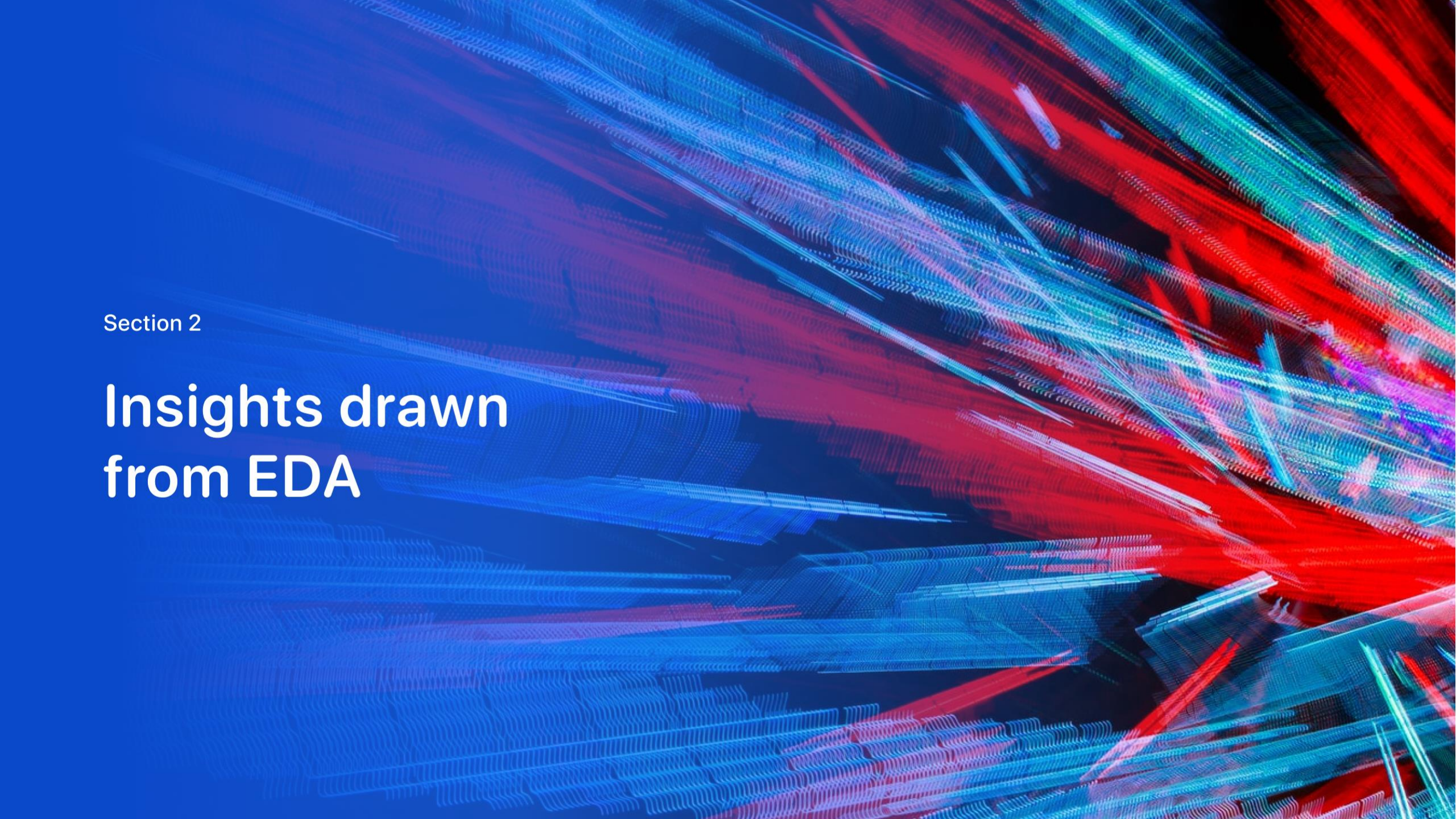


Results

- In the next slides, the following results are presented.
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

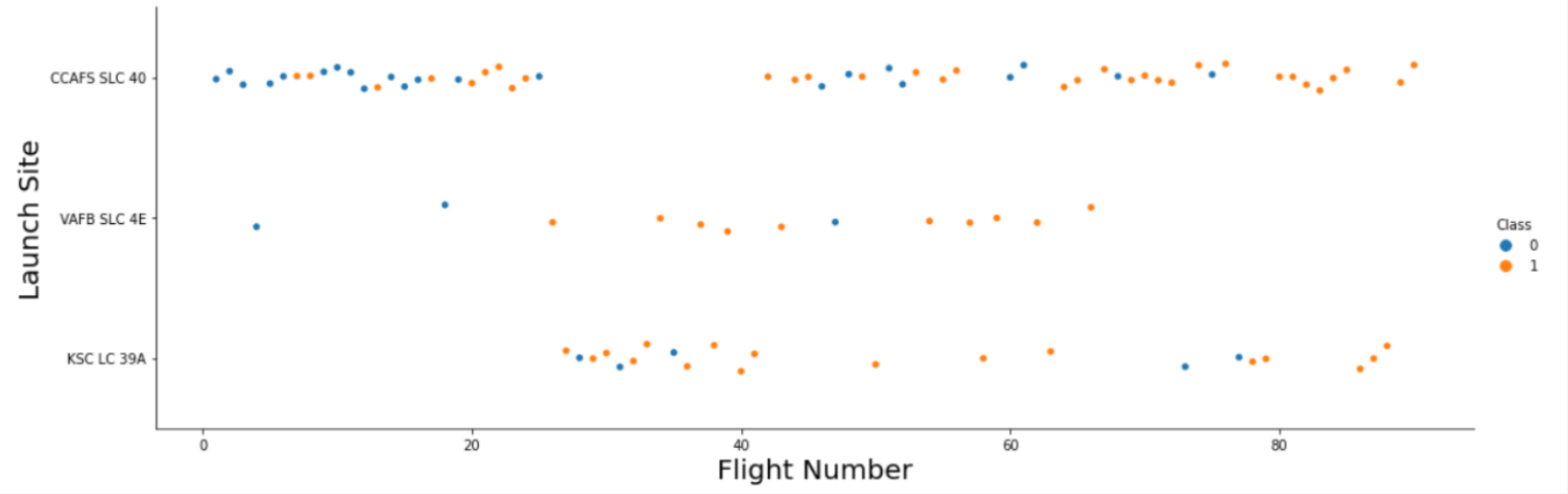
Section 2

Insights drawn from EDA



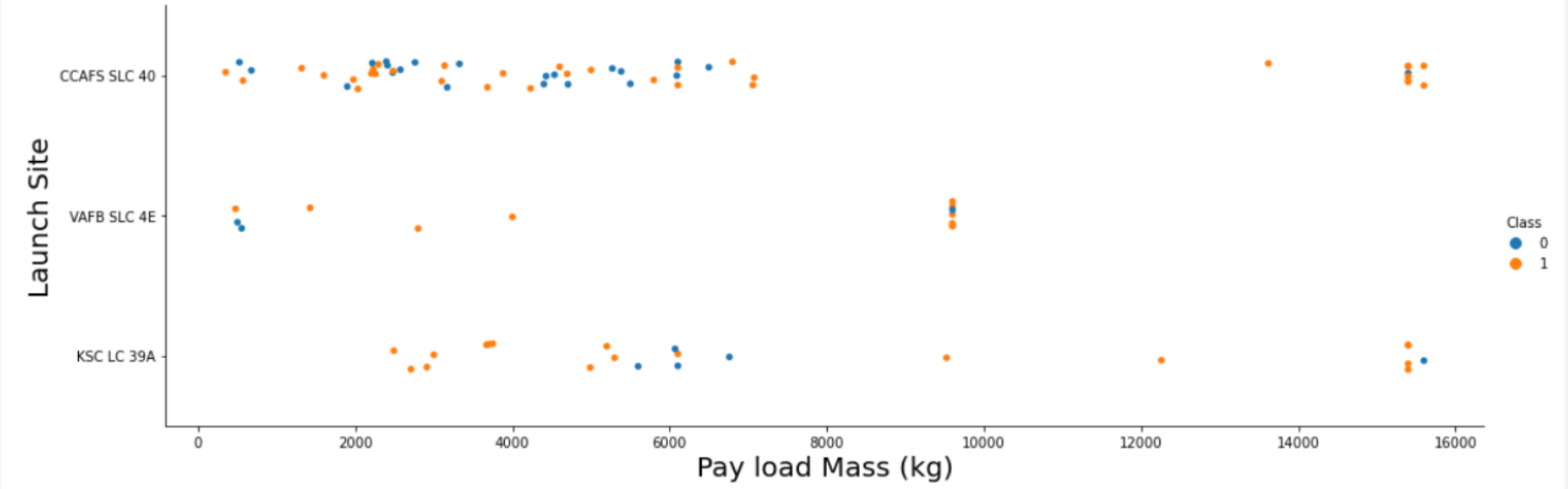
Flight Number vs. Launch Site

- For launch sites, CCAFS SLC40 and VAFB SLC 4E, successful landings increase as flight number increases. For KSC LC 39A site, successful landings don't.



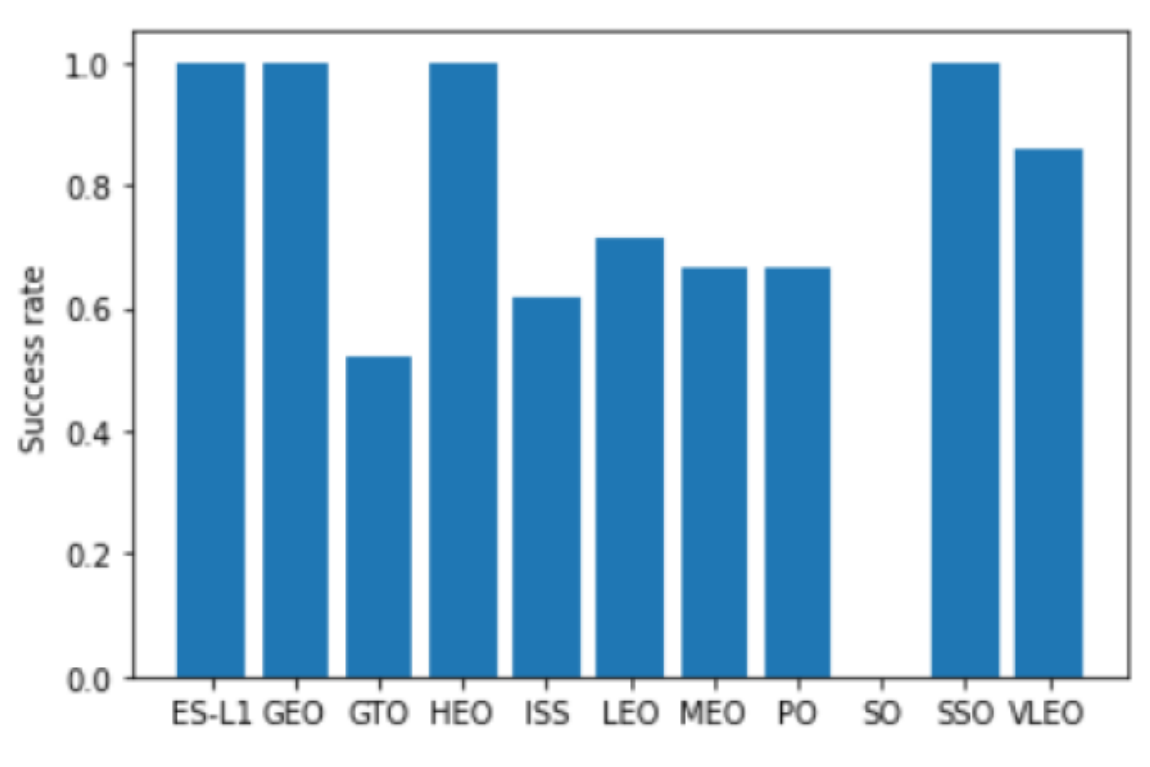
Payload vs. Launch Site

- Most frequent payloads for all sites are in the range of 2000-7000 kg.



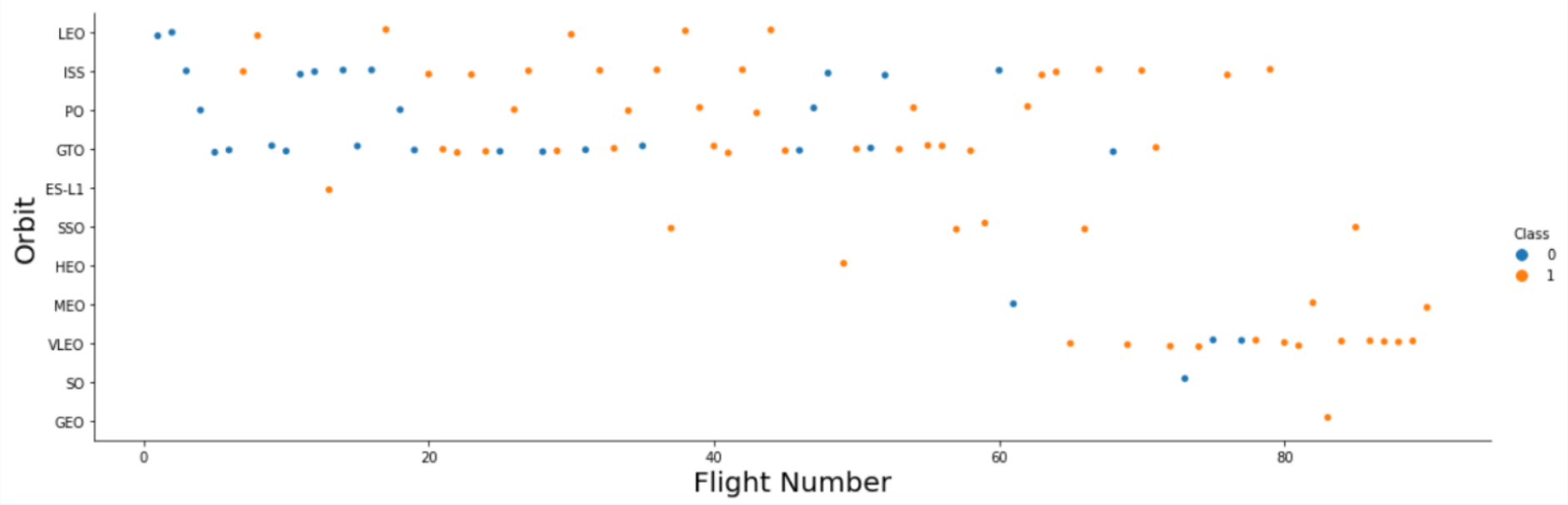
Success Rate vs. Orbit Type

- 4 orbits have 100% success
- The SO orbit has 0% success (1 mission)



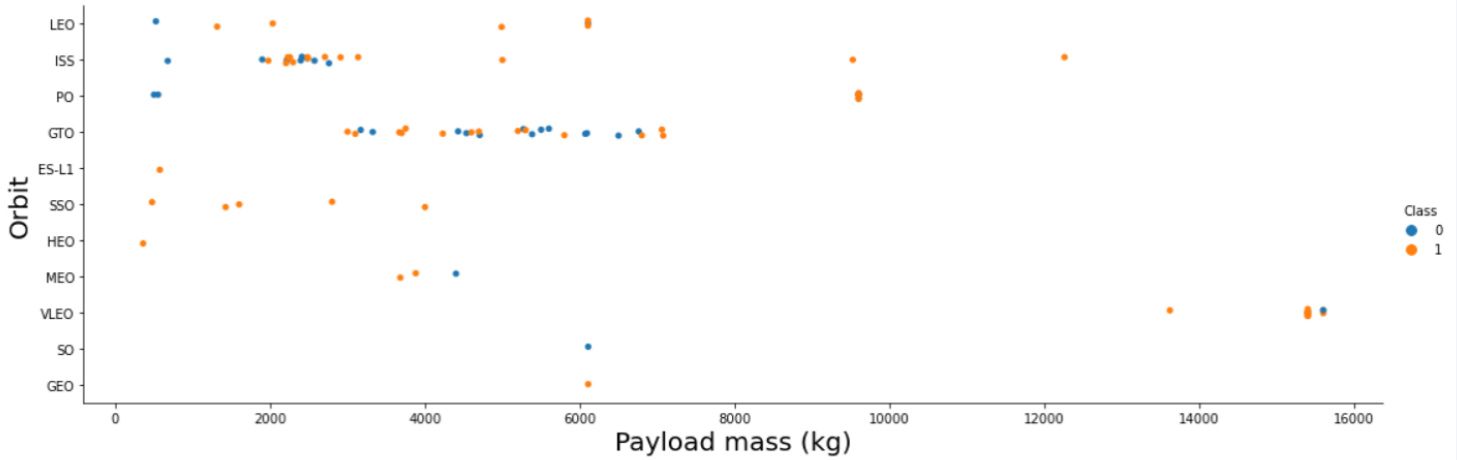
Flight Number vs. Orbit Type

- Since flight number 0-50, Orbit types (LEO, ISS, PO, GTO, LS-L1) are most frequently used.
- Since flight number 50-90, Orbit types (SSO, MEO, VLEO) are most frequently used.



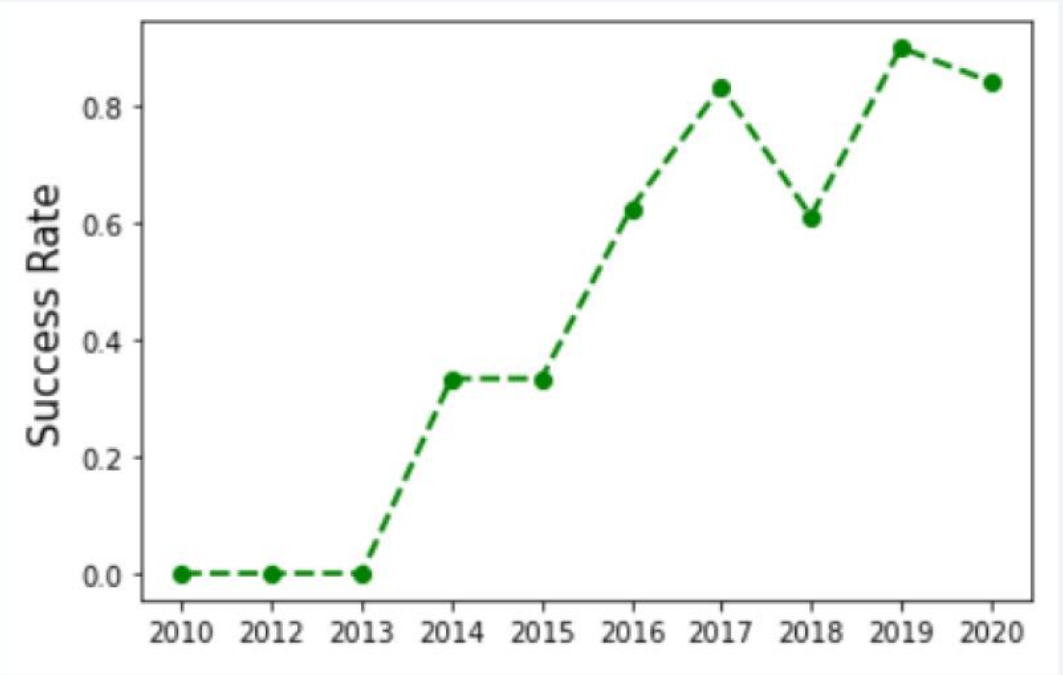
Payload vs. Orbit Type

- Payloads sent to ISS and GTO are quite close, while other orbits have high variance in payloads.



Launch Success Yearly Trend

- Success rate improves over time



All Launch Site Names

- 4 launch sites were listed with an SQL query:

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

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 _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	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- the total payload carried by boosters from NASA was 45,596 kg

```
Out[25]:    1
            45596
```

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 2,928 kg

Out[28]:

	1
	2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

- Here is the 4 boosters

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 are as follows

mission_outcome	number_of_missions
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass
- This mass 15600 kg corresponds to 60 Starlink satellites, delivered to the VLEO.

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- The failed landings in drone ship in year 2015 were listed as follows

booster_version	launch_site	Landing _Outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- There are 2 events, both with the F9 v1.1 launched from CCAFS LC-40

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

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

- The most frequent outcome was no attempt. This changes in 2017-2020.

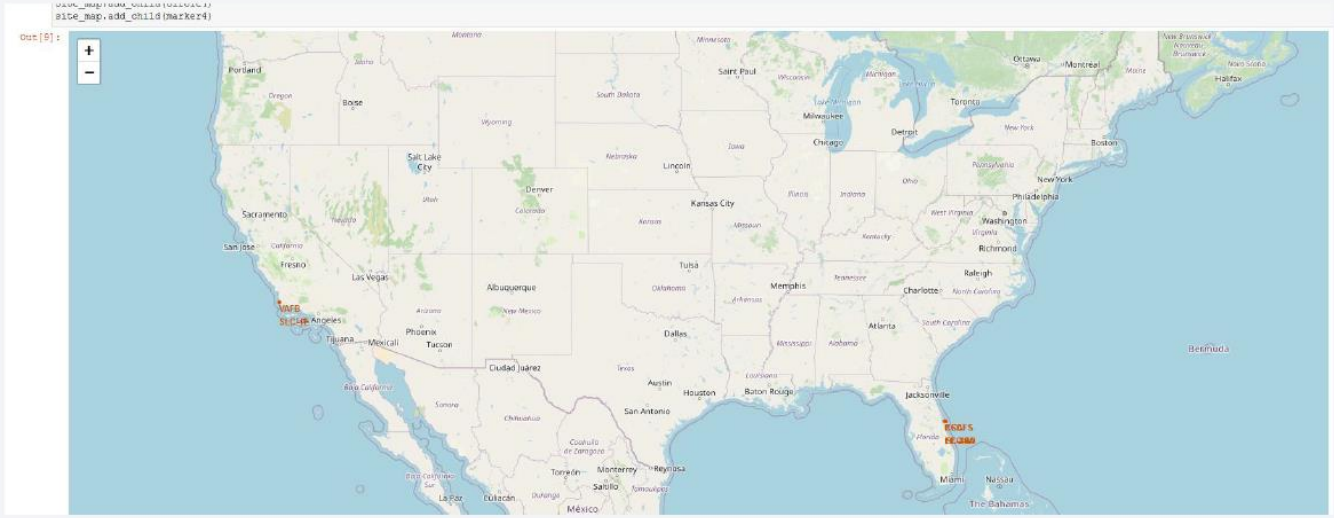
Landing outcome	Number of occurences
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 4

Launch Sites Proximities Analysis

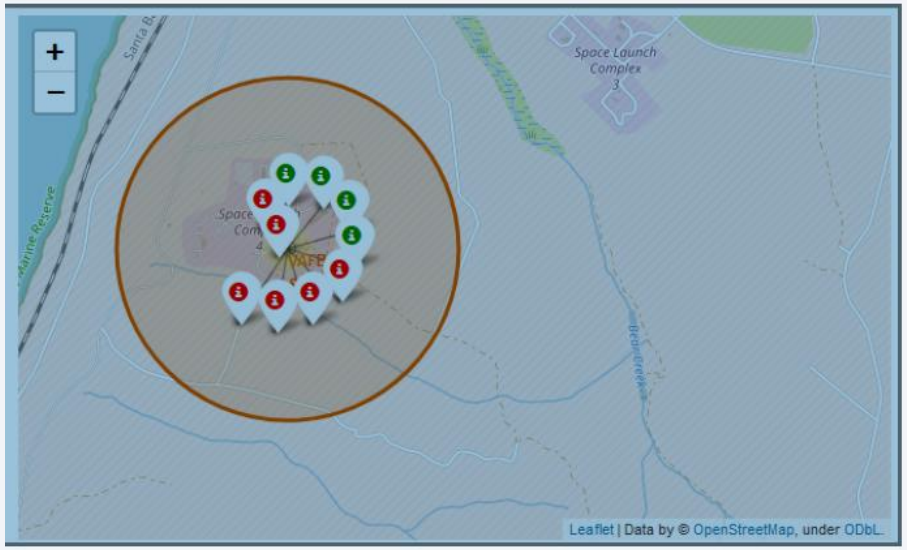
Location of launch sites

- This map shows the 4 launch sites
- All launch sites are in the south of US.



Detail of landing outcomes on the launch sites map

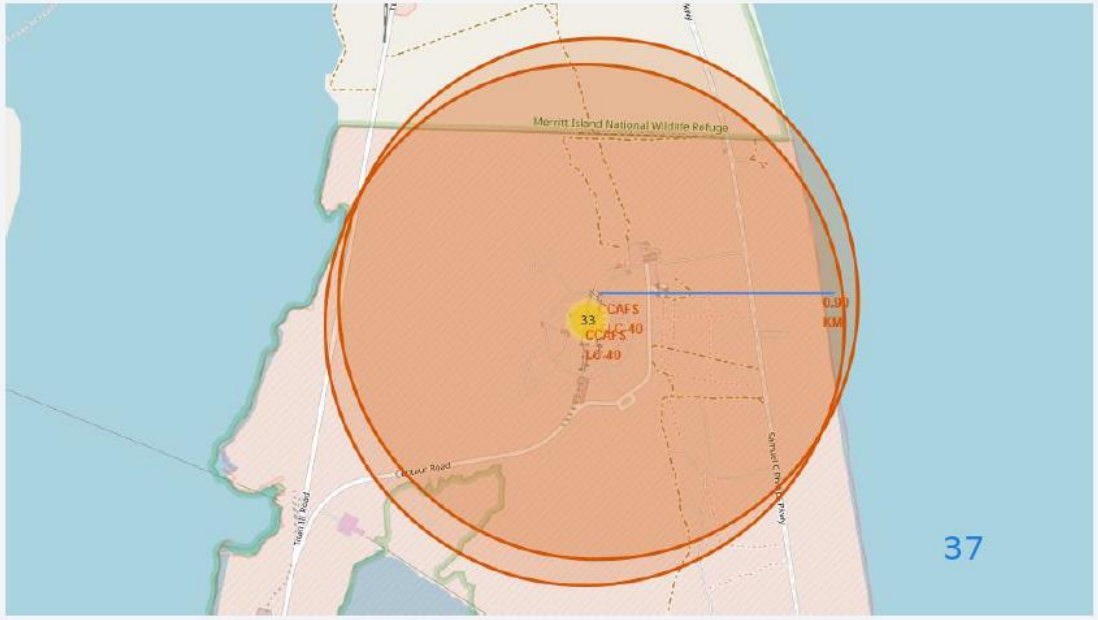
- The markers in the map represent each mission launched from 4 sites. The color of the marker represents the landing outcomes.



Distance of launch site to proximity elements

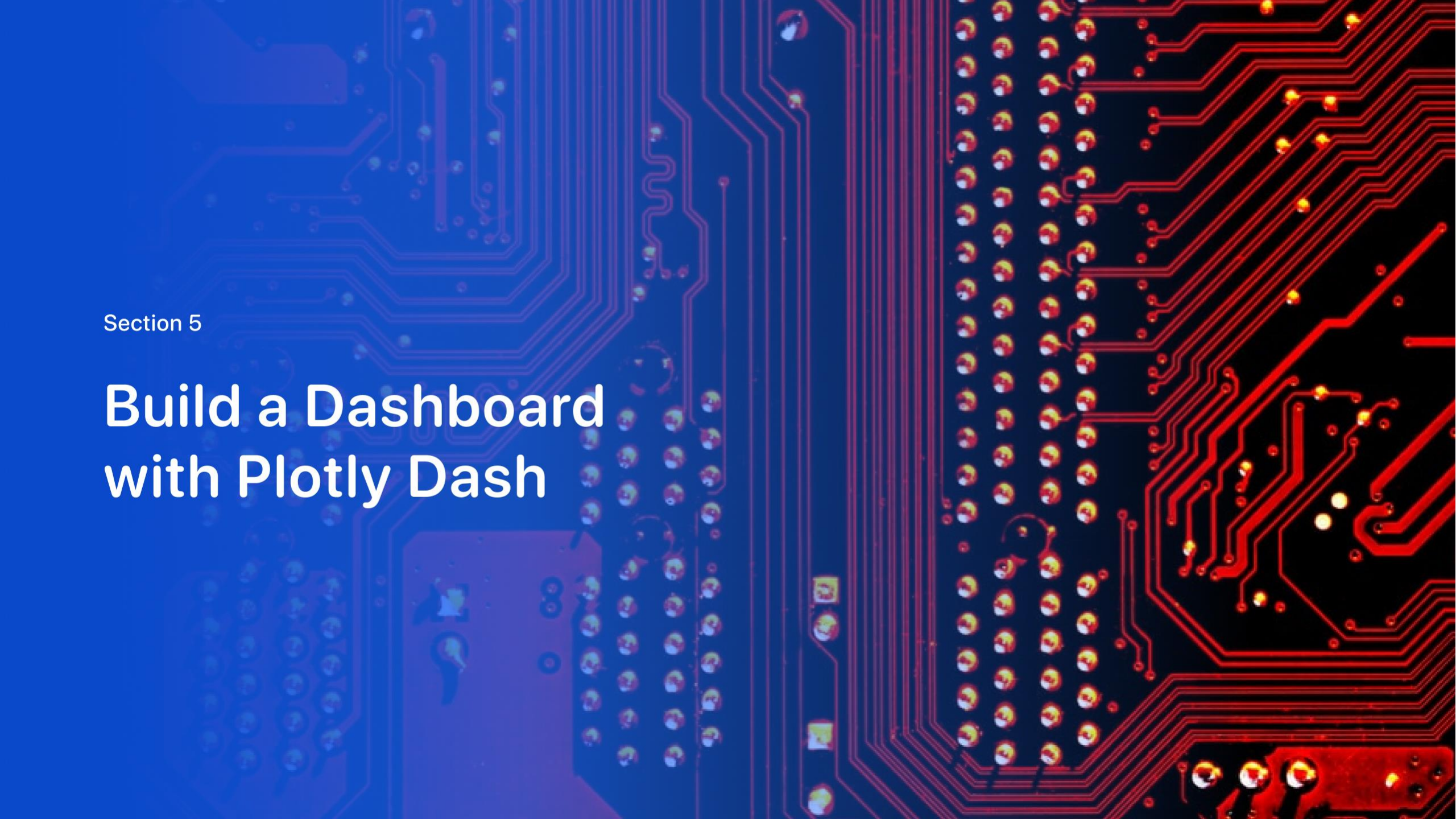
- This figure shows the distance from CCAFS SLC-40 the nearest coastline is 0.9 km.

- All launches sites are near the coast and have highways close to them.



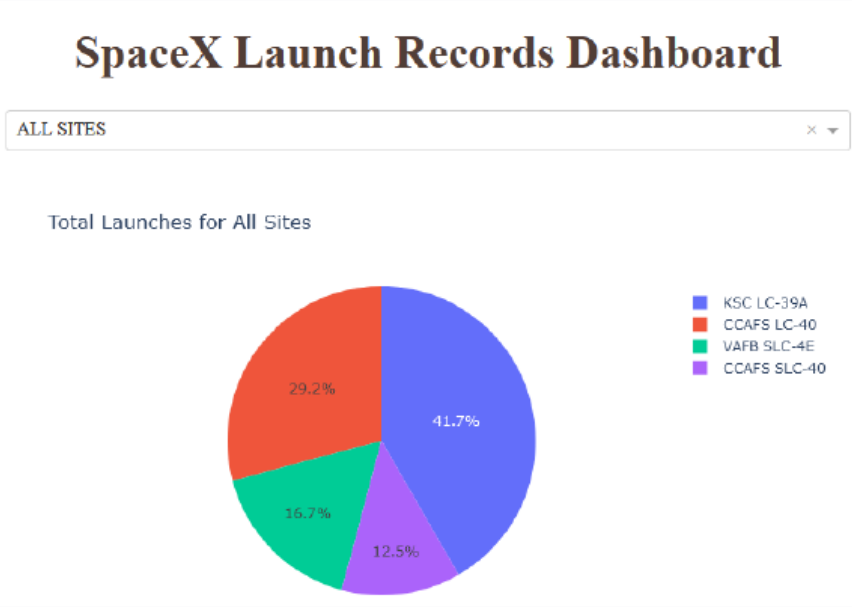
Section 5

Build a Dashboard with Plotly Dash



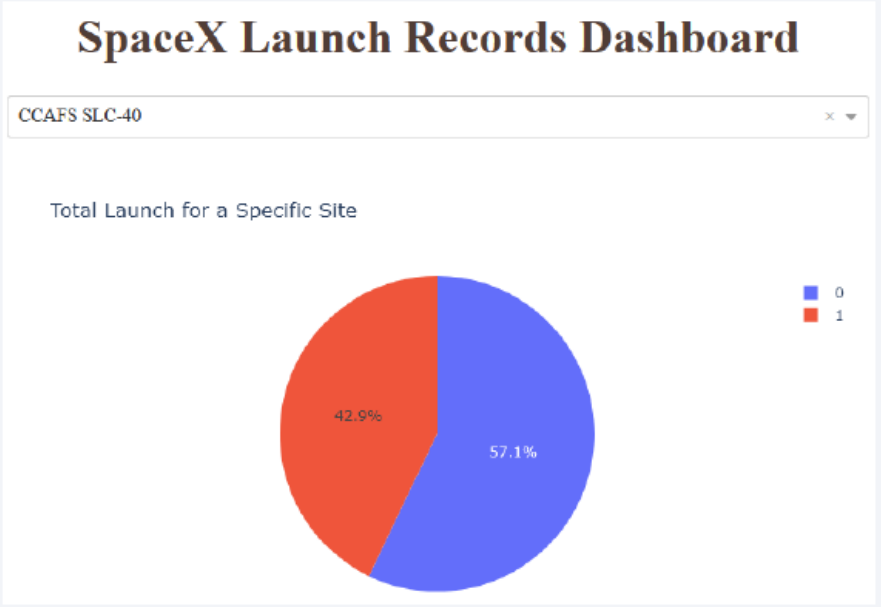
Dashboard – All site pie chart

- 4 launch sites, KSC LC-39 A has the most launches with successful landing.



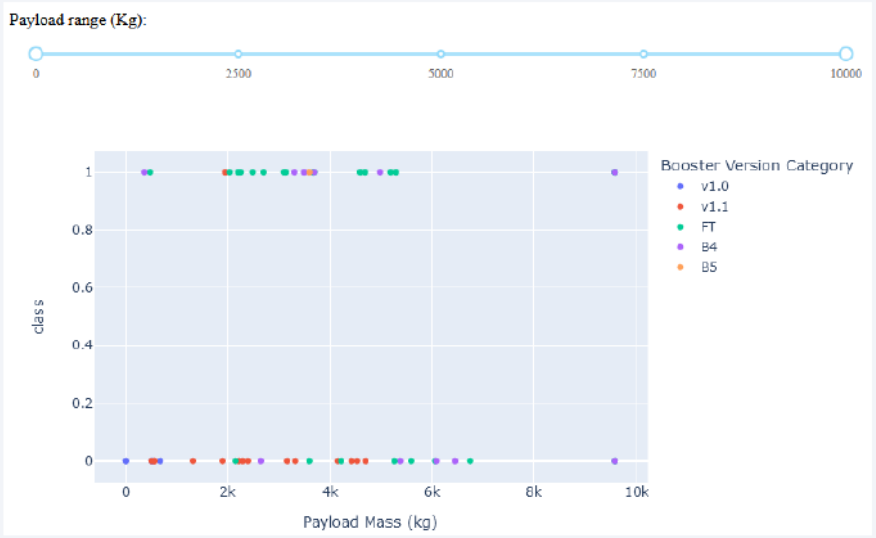
Dashboard – specific type pie chart

- The highest landing success rate corresponds to CCAFS SLC-40 at 42.9%



Dashboard – the effect of payload

- The highest success rate is achieved by booster FT and payload between 2000-3500kg

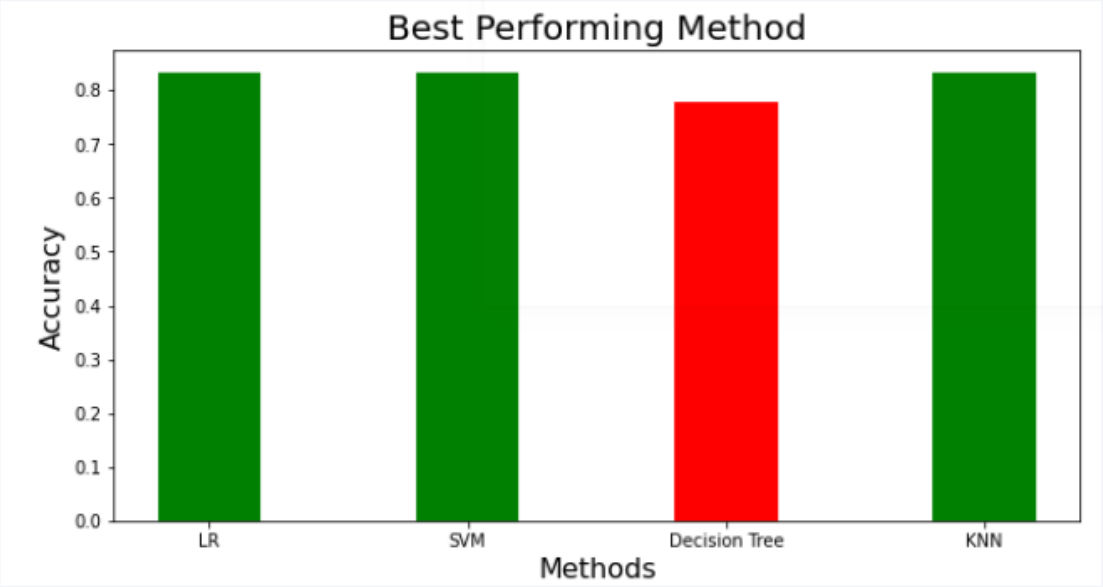


Section 6

Predictive Analysis (Classification)

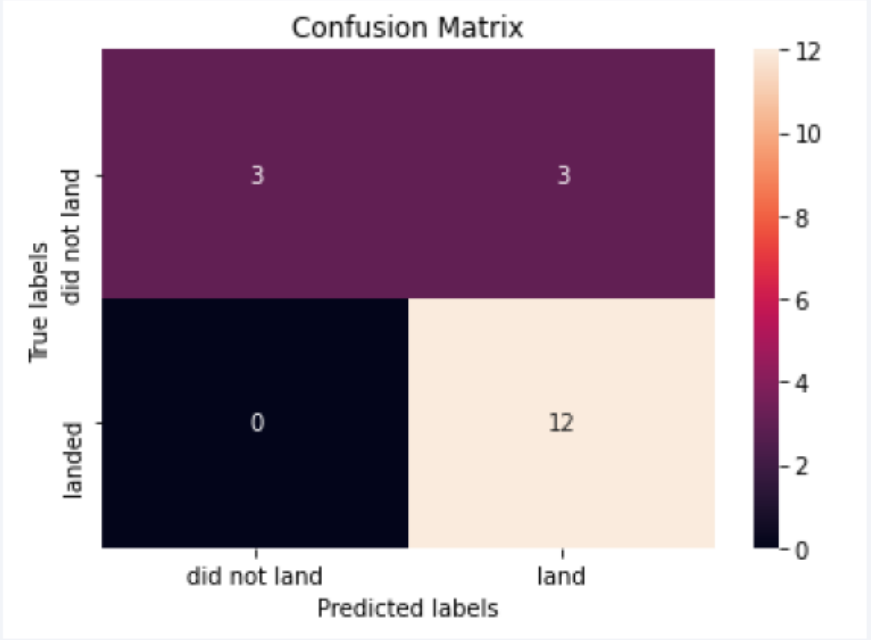
Classification Accuracy

- Logistic regression model has the best accuracy of 0.83.



Confusion Matrix

- Show the confusion matrix of the best performing model.
- The model can predict the failure to land in 3/3 cases completely.



Conclusions

- The successful landing of SpaceX' Falcon 9 stage 1 can be predicted with significant accuracy (above 80%) by using logistic regression model built on public data.
- The model can be improved if there are more public data.
- The landing success of SpaceX keeps increasing, SpaceY will need to double its effort in developing reusable booster.

Thank you!