IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dmytro Zherlitsyn
2023-01-02

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

- GitHub Link: https://github.com/dzherlitsyn/IBM_DS_Capstone.git

Executive Summary

- In the capstone project, there are applied the main Data Science methodology for predict the SpaceX Falcon 9 successful landing.
- The main results are
  - The row data was collected within the web-scraping and data wrangling methods
  - The statistical and visualization methods were applied for estimated the SpaceX Falcon 9 successfully landing
  - The SpaceX Falcon 9 successfully landing was predicted by the categorical models within the accuracy 0,83

Introduction

- The Falcon 9 first-stage landing tests were a series of controlled-descent flight tests conducted by SpaceX between 2013 and 2016. Since 2017, the first stage of Falcon 9 missions has been routinely landed if the rocket performance allowed it, and if SpaceX chose to recover the stage.
- The goal of the project is to predict whether the Falcon 9 first stage will land successfully
- The problem of this work is to analyzed and predict the SpaceX Falcon 9 successful landing based on open-data recourses.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Web-scraping and csv-files processing
- Perform data wrangling
  - Clearing the data (None detecting), means estimating and counting the data values
- Perform exploratory data analysis (EDA) using visualization (Seaborn library) and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - classification models and parameter tuning

Data Collection

- Getting data from the Falcon 9 official wiki
- Using GET requests
- Extracting table data with *beautifulsoup* library
- Filtering data (Falcon 9 only)
- Working with missing values
- Saving data as CSV format

Data Collection

- Data collection with SpaceX
  - Read open data
  - Using BeautifulSoup content processing
  - Parsing htmp by `BeautifulSoup tools`
  - Creating Data.Frame by the Features
  - Saving csv-file

- GitHub Link: https://github.com/dzherlitsyn/IBM_DS_Capstone/blob/d200516226d7d4c904d6295cd203f7a73638c5bb/001.Webscraping.ipynb

EDA with Data Visualization

- Visualize the relationship between Flight Number and Launch Site
- Visualize the relationship between Payload and Launch Site
- Visualize the relationship between success rate of each orbit type
- Visualize the relationship between FlightNumber and Orbit type
- Visualize the relationship between Payload and Orbit type
- Visualize the launch success yearly trend
- GitHub Link: https://github.com/dzherlitsyn/IBM_DS_Capstone/blob/751343437138 95b352205d1124e114d0e9b0edd2/003.Eda-dataviz.ipynb

Build an Interactive Map with Folium

- 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

- GitHub Link: https://github.com/dzherlitsyn/IBM_DS_Capstone/blob/751343437138 95b352205d1124e114d0e9b0edd2/004.Interactive%20Visual%20Analytics%20with%20Folium.ipynb

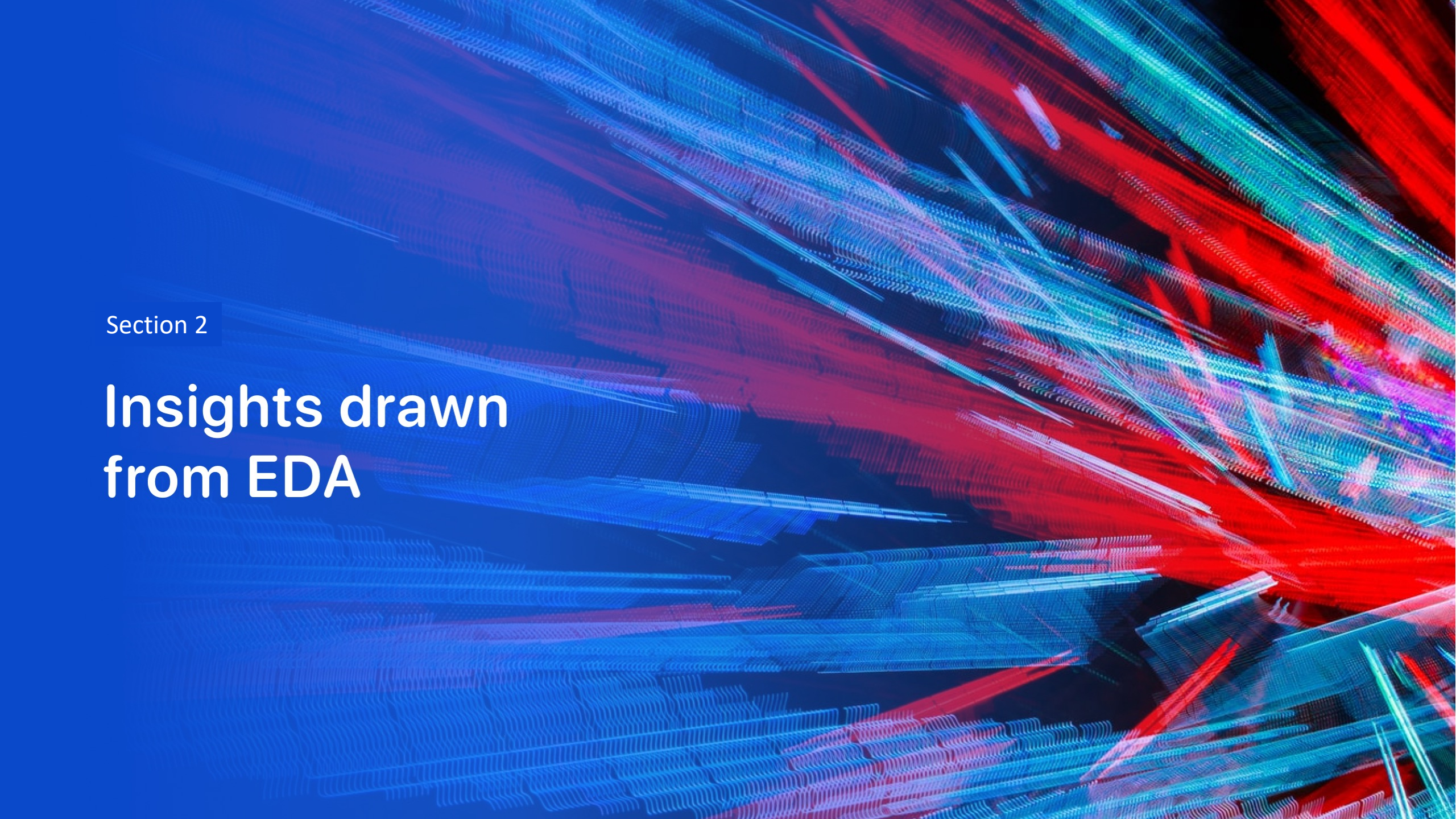
Predictive Analysis (Classification)

- Create a NumPy array from the column Class in data.
- Standardize the data in X then reassign it to the variable X using the StandardScaler transform.
- Use the function train_test_split to split the data X and Y into training and test data.
- Create models objects within the GridSearchCV outputs:
  - a logistic regression object
  - a support vector machine object
  - a decision tree classifier object
  - a k-nearest neighbors object
- Find the method performs best

- GitHub Link: https://github.com/dzherlitsyn/IBM_DS_Capstone/blob/5f99e59da3fbbee1fc4640211b52f1b8ea67b277/005.Machine_Learning_Prediction.ipynb

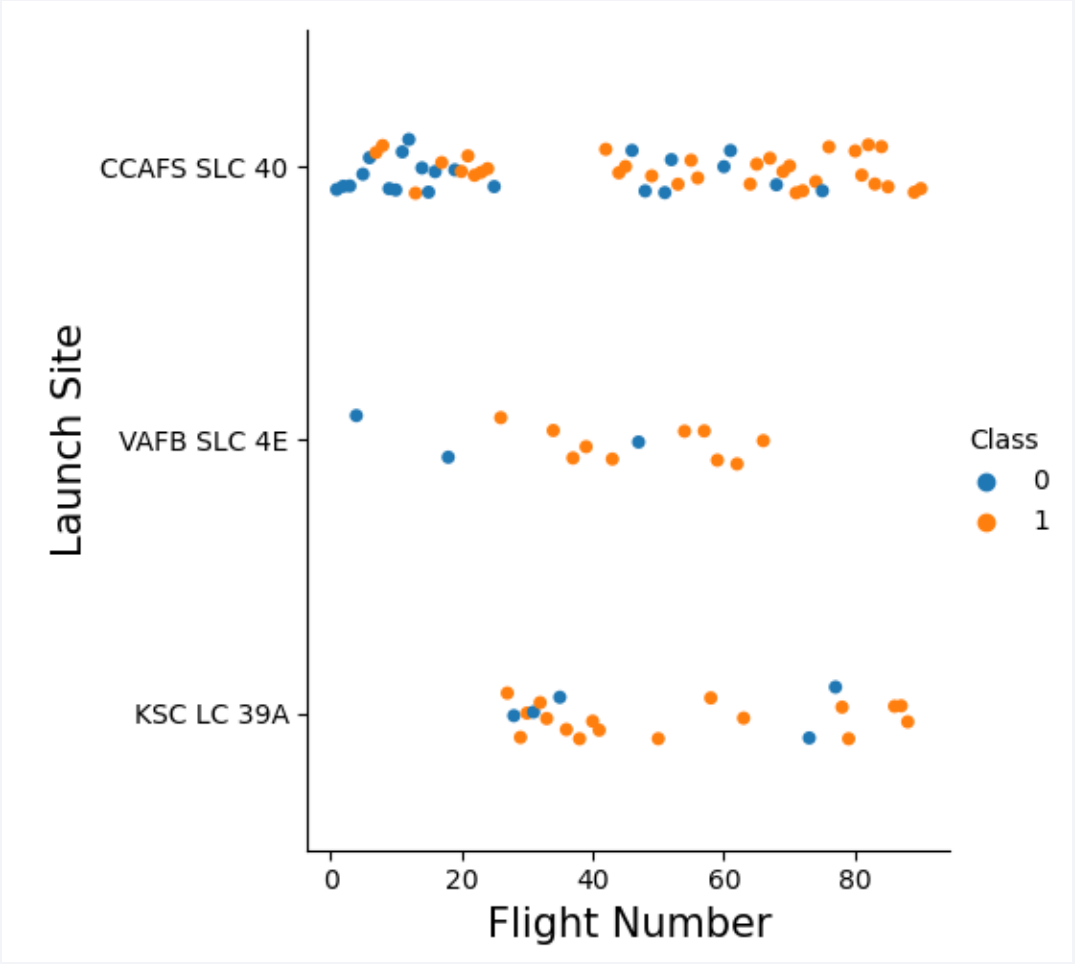
Results

- The best result has an accuracy of 0.83
- All categorical forecasting methods have the same accuracy on the test data set.


Section 2
Insights drawn from EDA

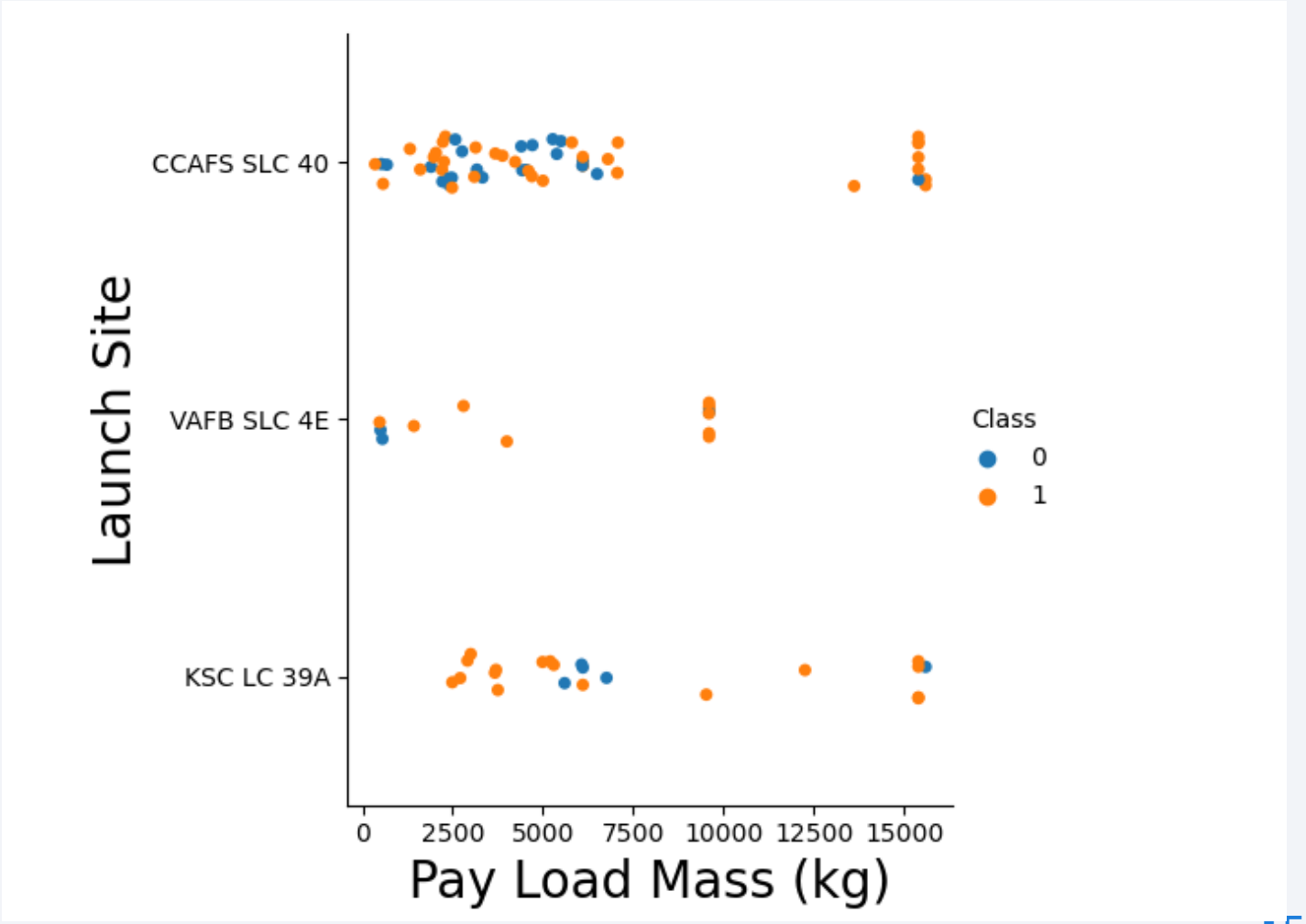
Flight Number vs. Launch Site

- The success rate (blue and red) increased as the number of flights increased.



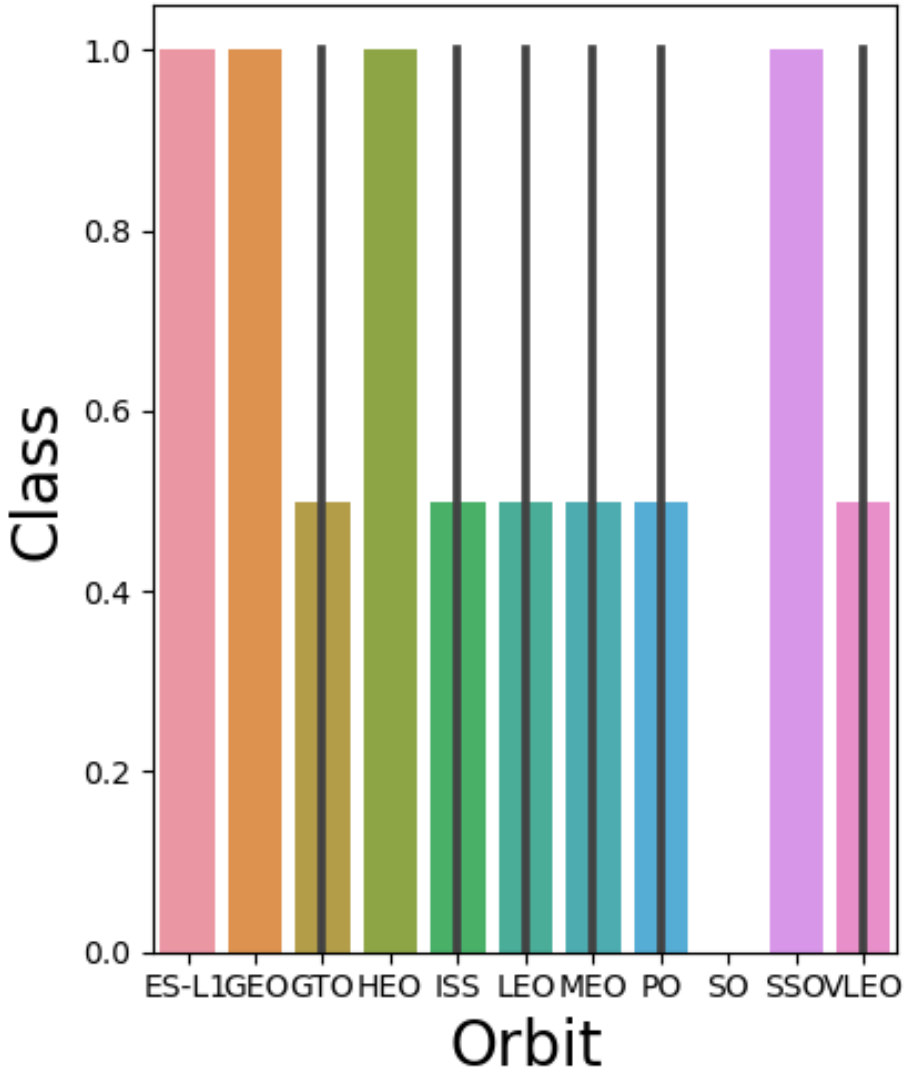
Payload vs. Launch Site

- It should be a weak correlation between the variables



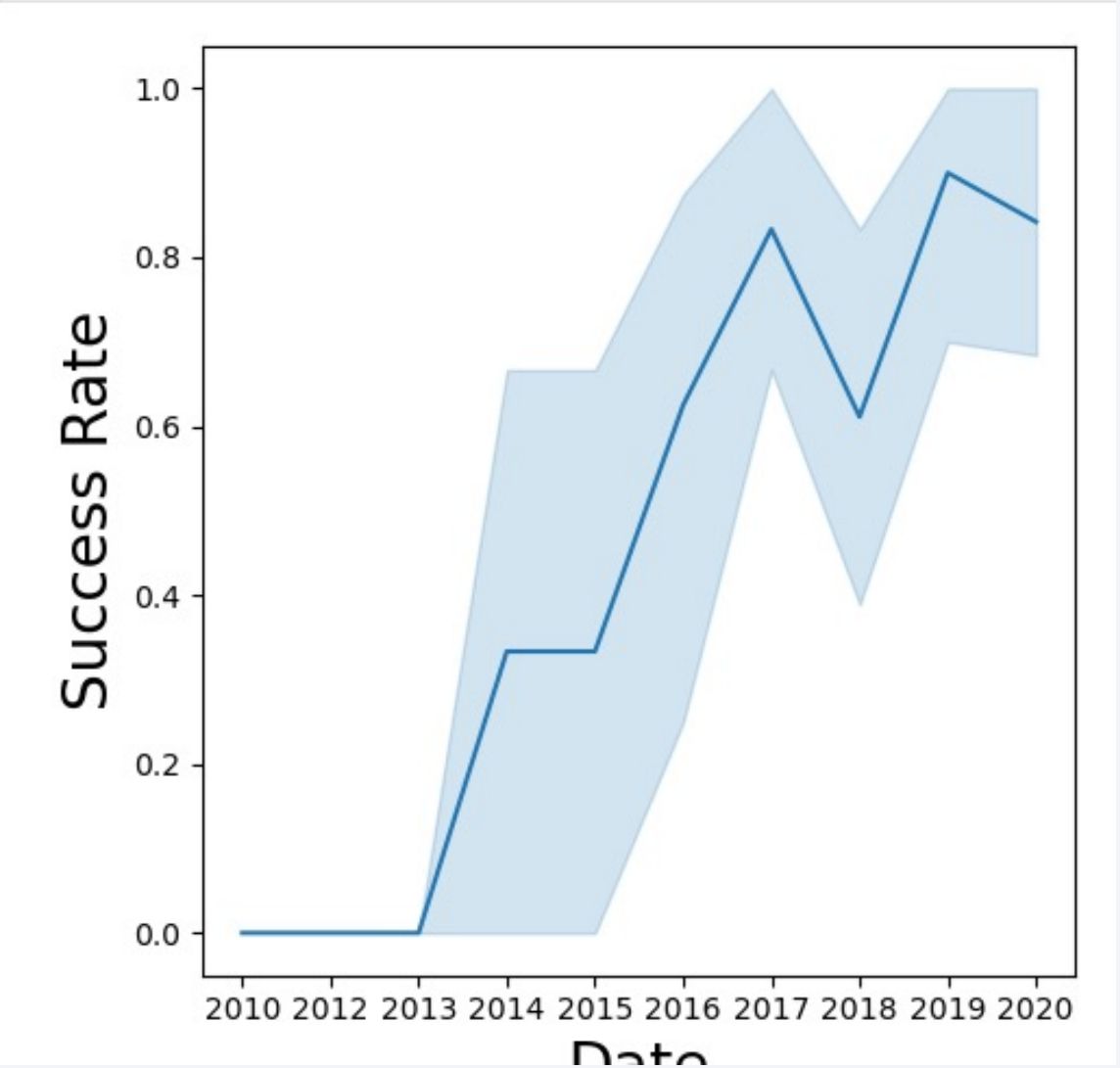
Class vs. Orbit Type

- Only SO orbit has a lowest success rate



Launch Success Yearly Trend

- The success rate has stabilized since 2017
- From 2013 to 2017 the the success rate is increasing strongly



All Launch Site Names

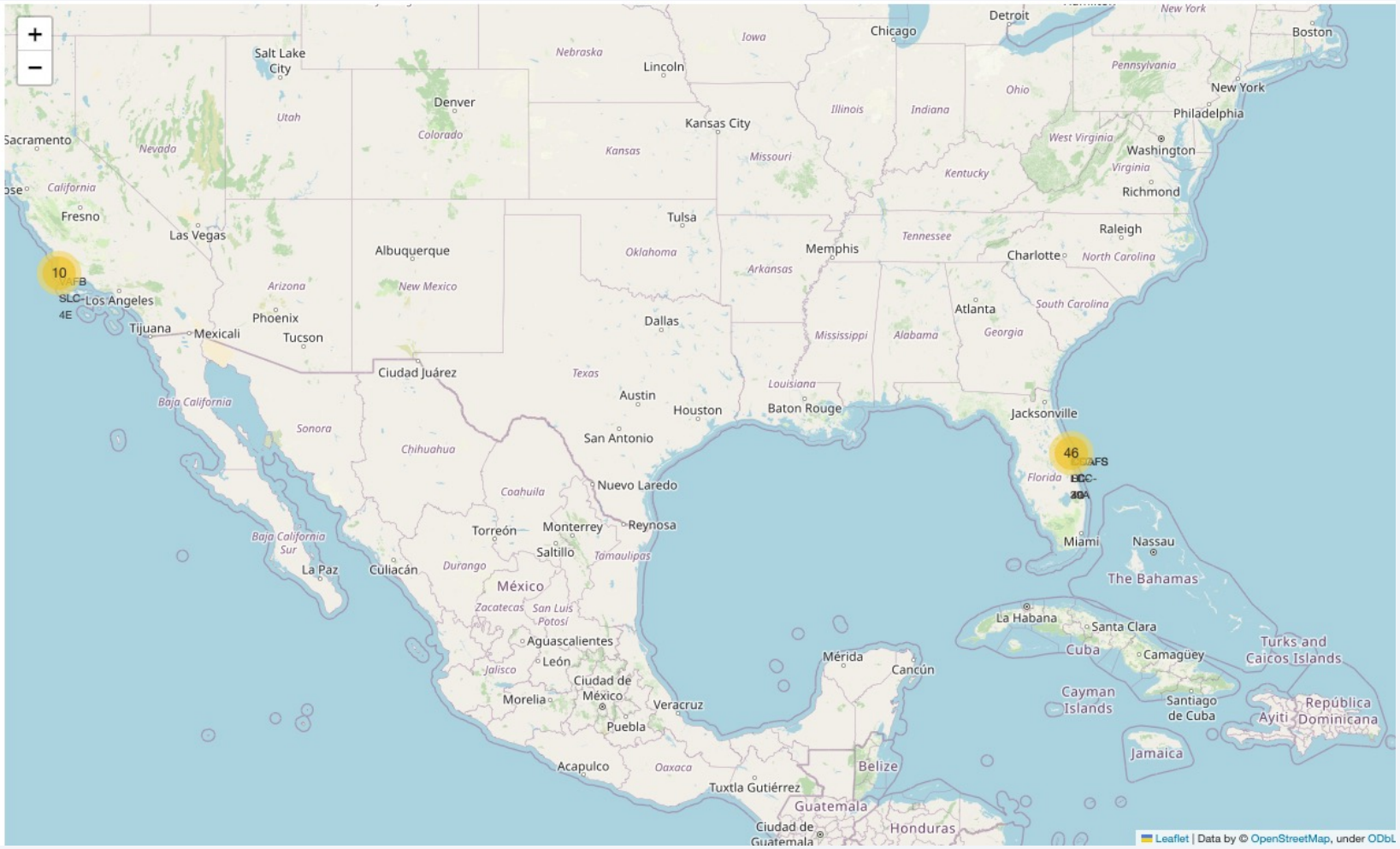
- The names of the launch sites are CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E

:

	Launch Site	Lat	Long	class
0	CCAFS LC-40	28.562302	-80.577356	0
1	CCAFS SLC-40	28.563197	-80.576820	1
2	KSC LC-39A	28.573255	-80.646895	1
3	VAFB SLC-4E	34.632834	-120.610745	0

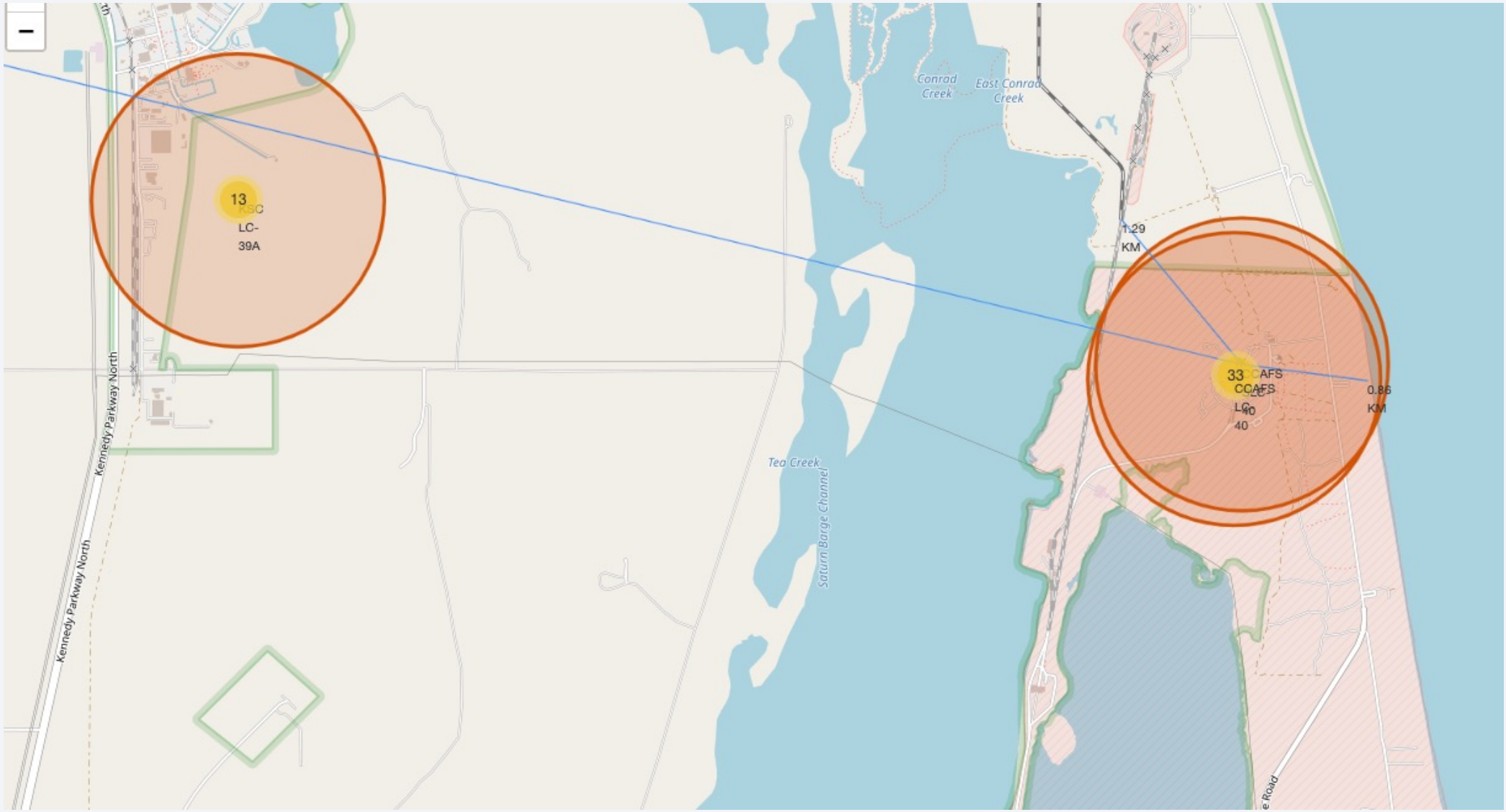

Section 3
Launch Sites
Proximities Analysis

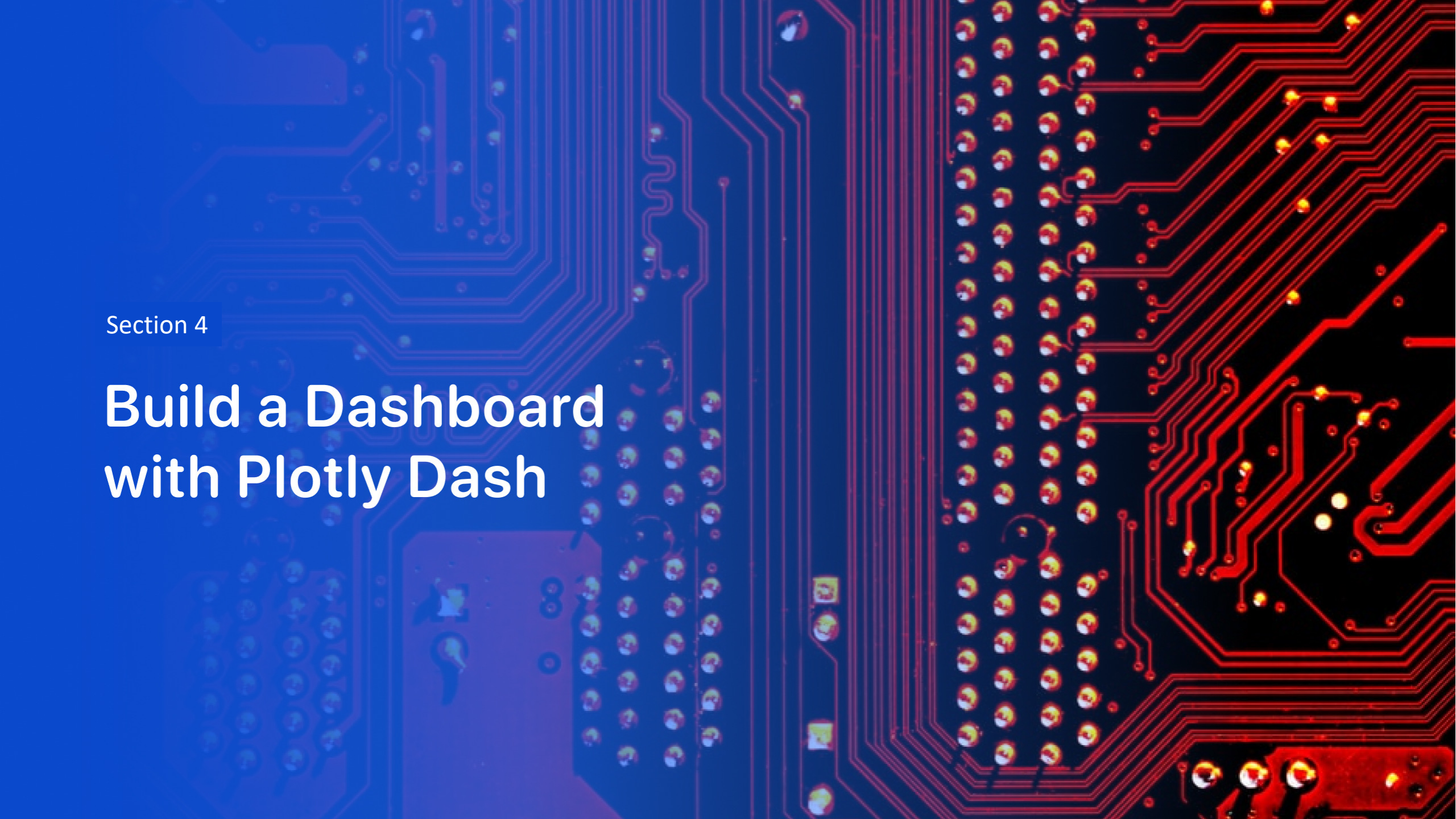
SpaceX Launch Sites Locations



The launch sites were near the coast

Landing points connections

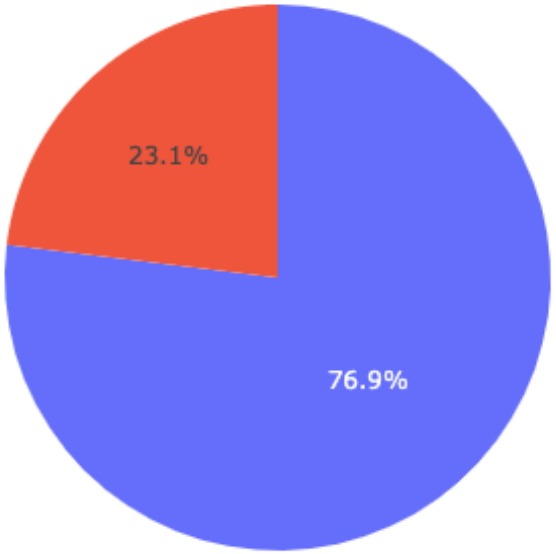



Section 4
Build a Dashboard with Plotly Dash

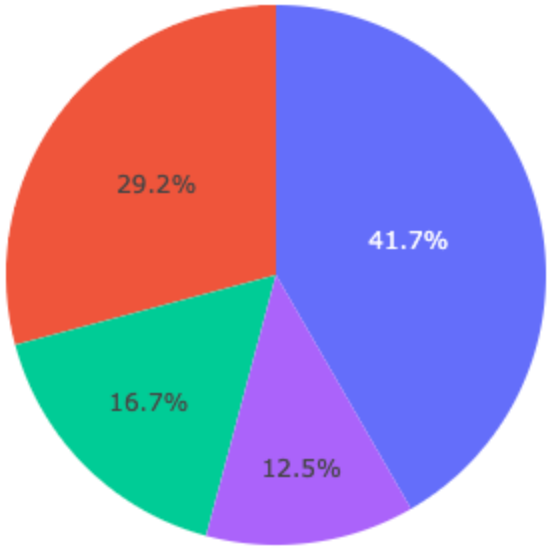
Total Launch

KSC LC-39A

Total Launch for a Specific Site

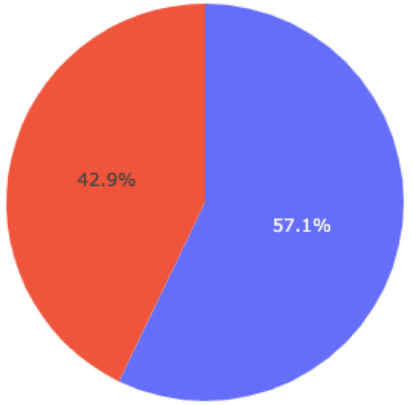


Total Launches for All Sites



CCAFS SLC-40

Total Launch for a Specific Site



Payload Mass, Booster Version category and Success

- For Payload range (Kg): 5000 to 10000

Section 5

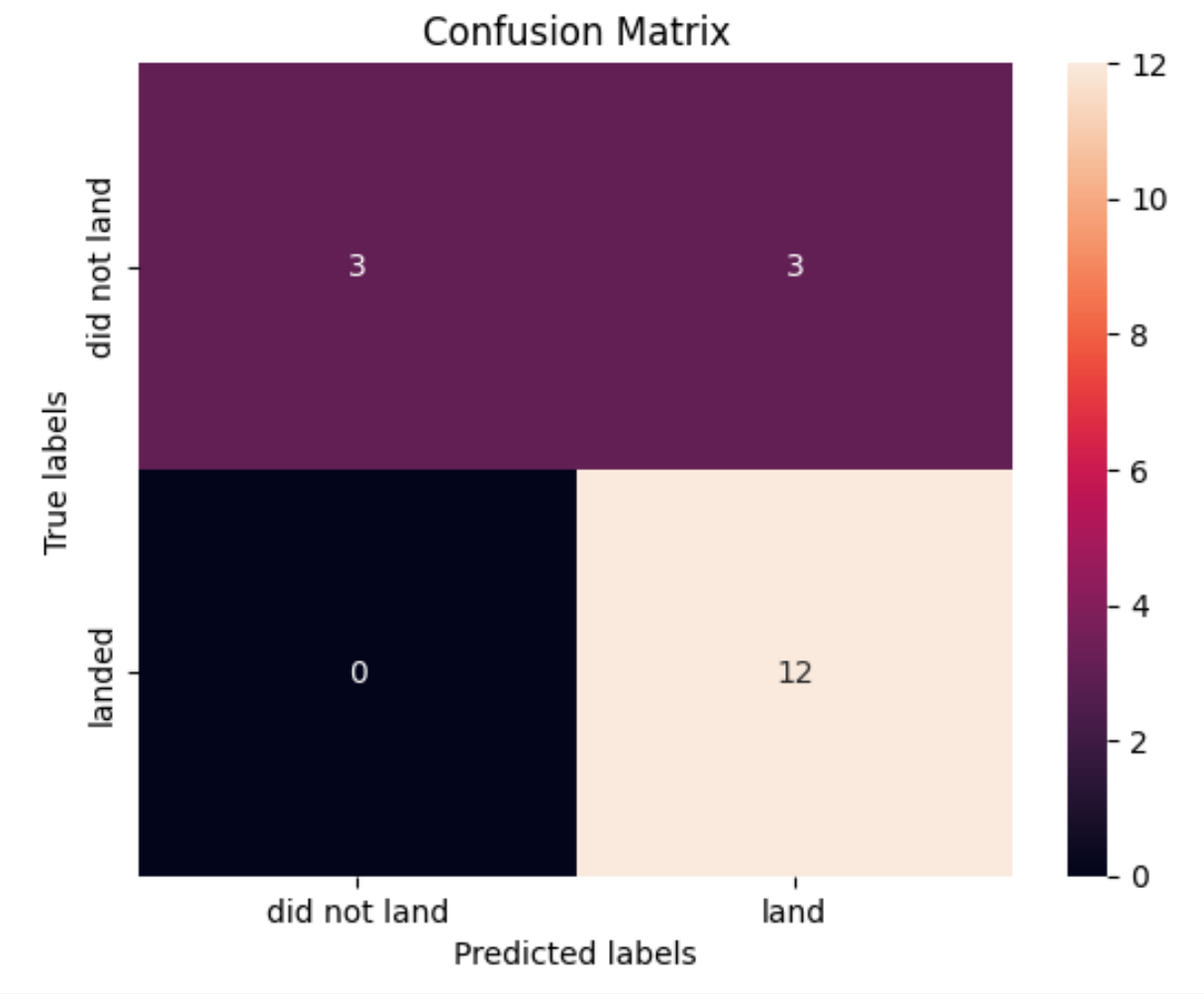
Predictive Analysis (Classification)

Classification Accuracy

- The highest training score for the decision tree classifier method is 0.88.
- The final accuracy for the test data is the same for all methods

	method	accuracy	train_score
0	Logistic regression	0.833333	0.846429
1	SVM	0.833333	0.848214
2	Decision tree classifier	0.833333	0.875000
3	k-neighbors	0.833333	0.848214

Confusion Matrix



- The best perdition for the 12 successful landings in case of True Positive Case.

Conclusions

- The project allowed to test many tools and methods of data science
- The SpaceX Falcon 9 successfully landing was predicted by the categorical models within the accuracy 0,83
- To improve the accuracy of the forecast, it is possible to use other combinations of test, train and validation samples.

Thank you!