Self-development Data Science Professional

Yuan Feng

Dec 2023

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- This is a comprehensive program focused on data collection, processing and visualization. Based on the SpaceX Falcon 9 dataset, we experienced the entire process of treatment in different ways

INTRODUCTION

- This presentation is geared towards my fellow data scientists. We're going to dive into the dataset of SpaceX rocket that's packed with the kind of technical details we thrive on. It's not just about learning new tricks; it's about enhancing our skills and staying at the top of our game in the ever-evolving world of data science, which include:
- Data Collection
- Data Wrangling
- EDA (Exploratory Data Analysis)
- Data Visualization

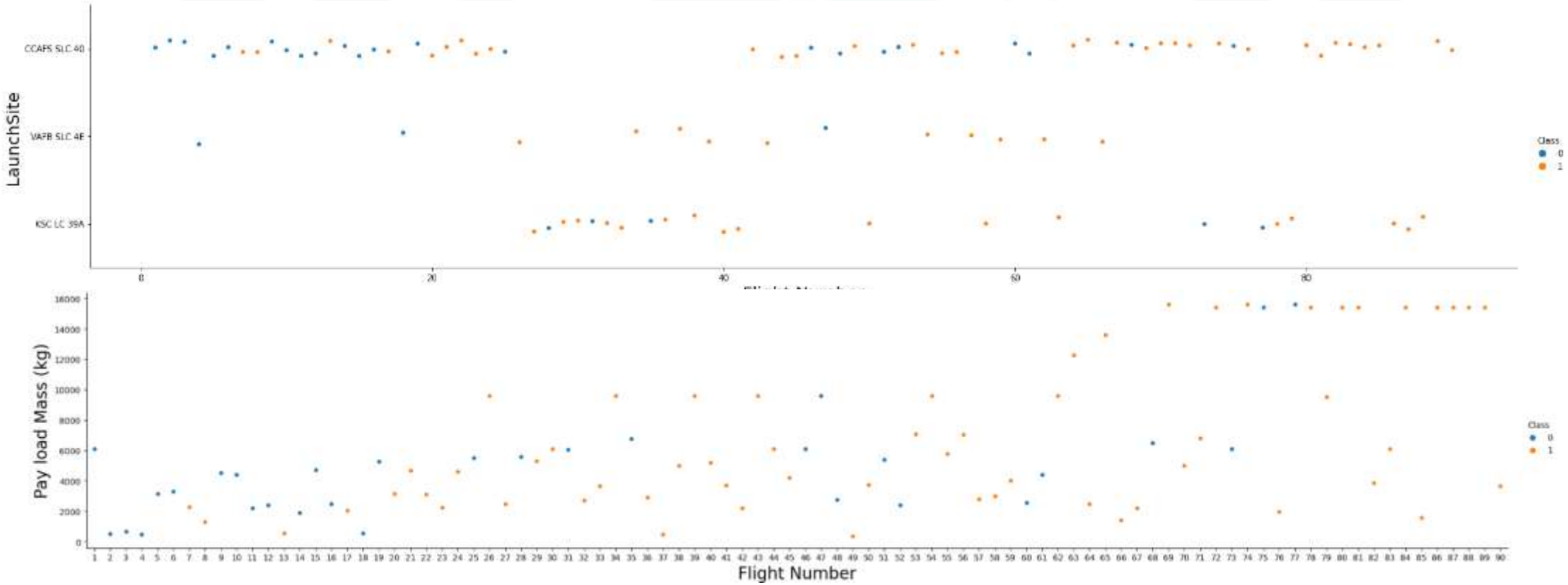
METHODOLOGY

- Data Collection:
  - Rest API Lab (URL technique)
- Data Wrangling
  - Remove duplicate, missing value
  - Normalization
- EDA
  - Distribution, Outlier, Correlation
  - Using Data Visualization or SQL
- Data Visualization
  - Chart (histogram, boxplot, scatter plot, bar, line, pie, etc)
- Prediction
  - Regreesion
  - Cross-validation

RESULTS

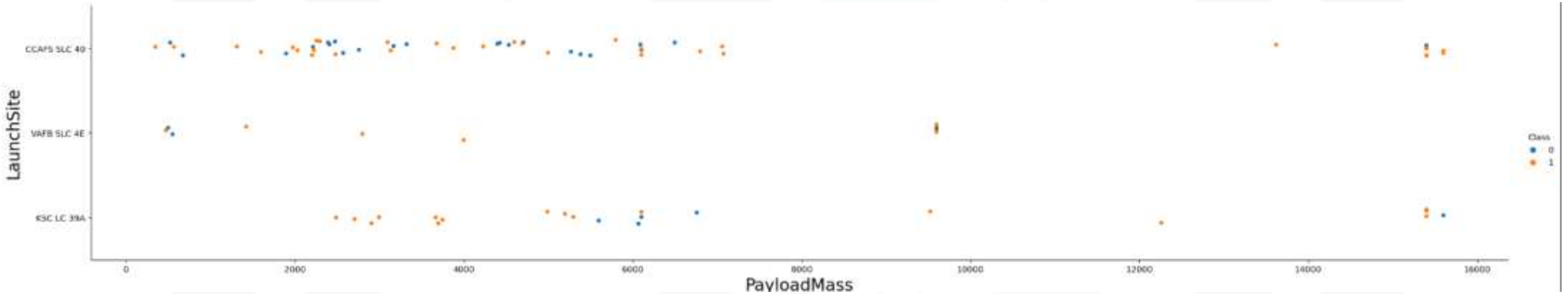
EDA with Data Visualization

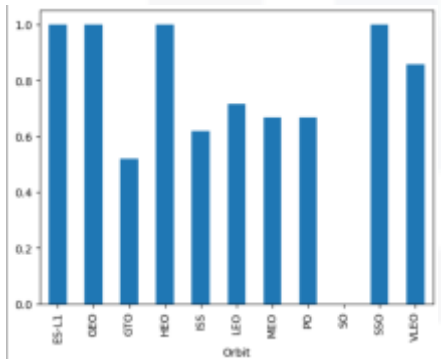
Usually prefer scatter plots or bar plots



EDA with Data Visualization

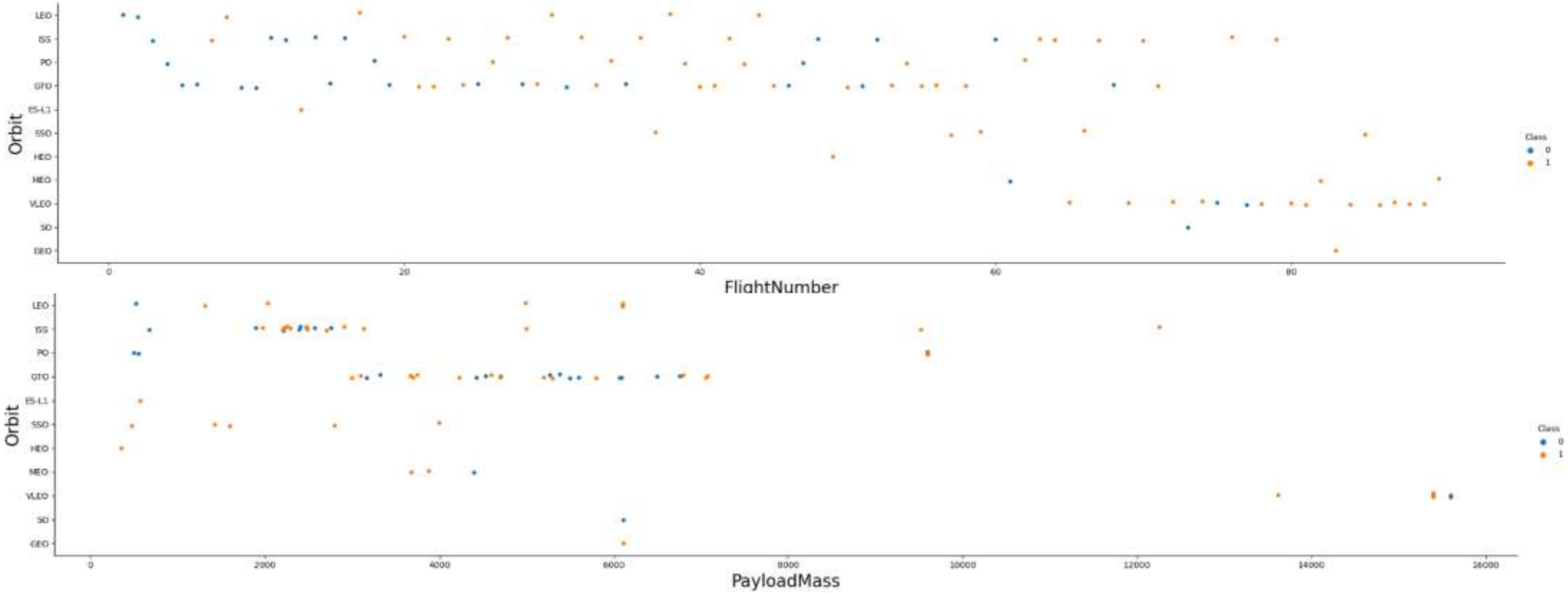
Usually prefer scatter plots or bar plots





EDA with Data Visualization

Usually prefer scatter plots or bar plots



EDA with SQL

Including the following:

- 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 succesful 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
- etc…

EDA with SQL

	launchsite
0	KSC LC-39A
1	CCAFS LC-40
2	CCAFS SLC-40
3	VAFB SLC-4E

	date	time	boosterversion	launchsite	payload	payloadmasskg	orbit	customer	missionoutcome	landingoutcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

	total_payloadmass
0	45596

	avg_payloadmass
0	2928.4

	firstsuccessfull_landing_date
0	2015-12-22

	boosterversion
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

The total number of successful mission outcome is:

	successoutcome
0	100

The total number of failed mission outcome is:

	failureoutcome
0	1

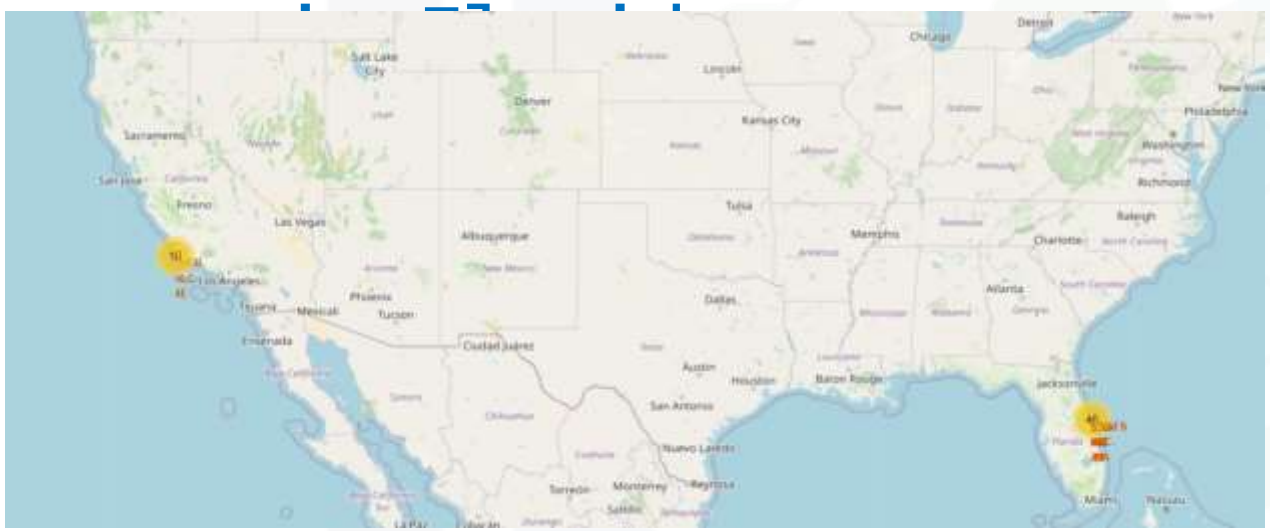
	boosterversion	payloadmasskg
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
5	F9 B5 B1051.3	15600
6	F9 B5 B1051.4	15600
7	F9 B5 B1051.6	15600
8	F9 B5 B1056.4	15600
9	F9 B5 B1058.3	15600
10	F9 B5 B1060.2	15600
11	F9 B5 B1060.3	15600

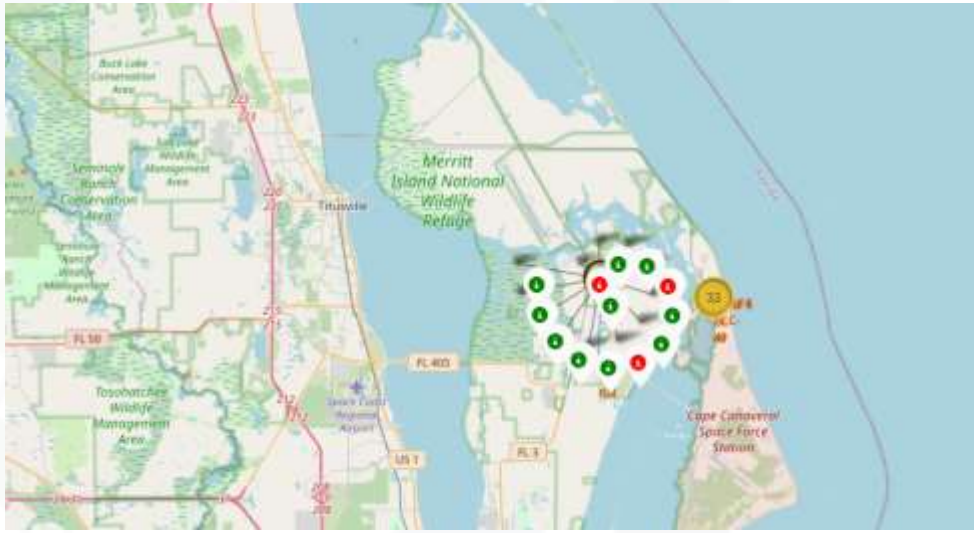
	landingoutcome	count
0	No attempt	10
1	Success (drone ship)	6
2	Failure (drone ship)	5
3	Success (ground pad)	5
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Precluded (drone ship)	1
7	Failure (parachute)	1

	boosterversion	launchsite	landingoutcome
0	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Interactive map with Folium

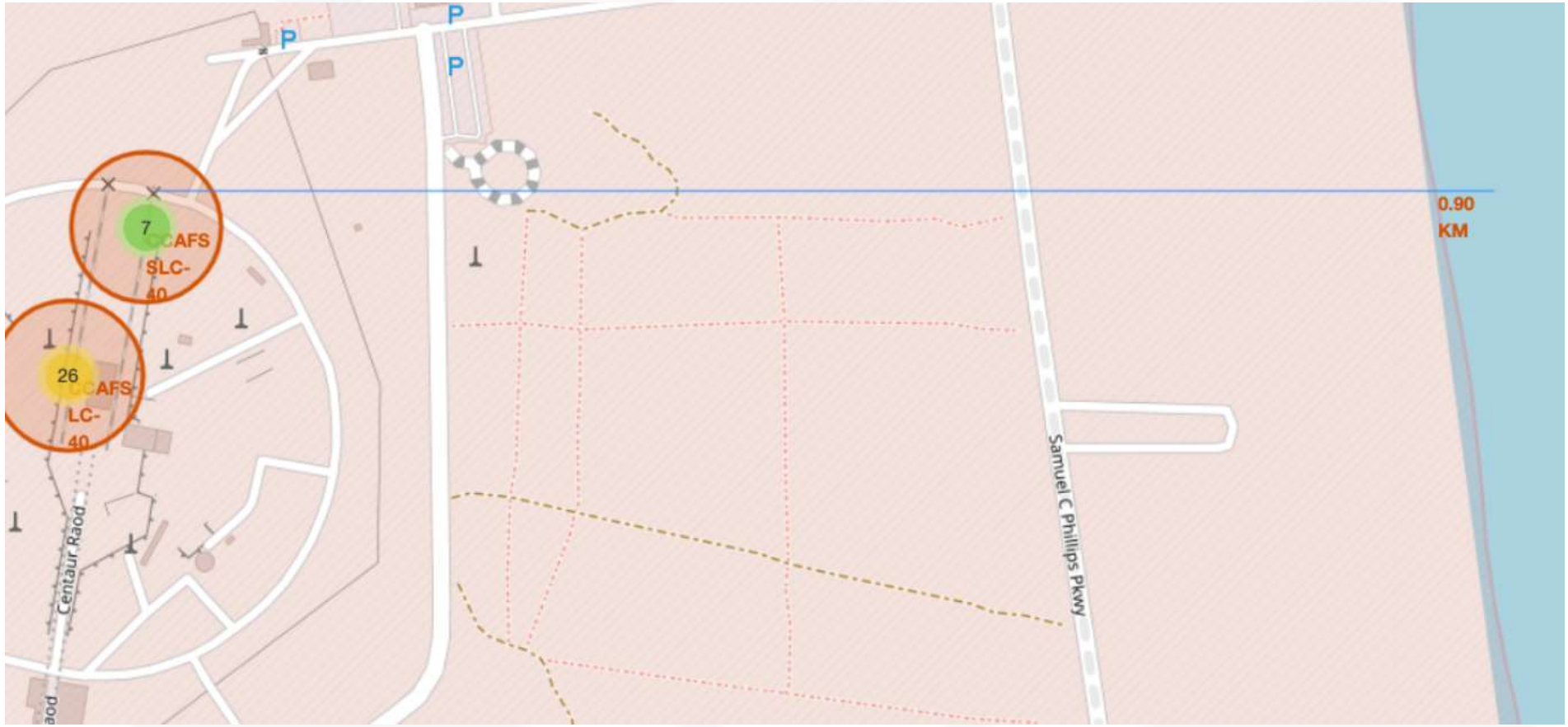
Two main place for SpaceX, one in California and





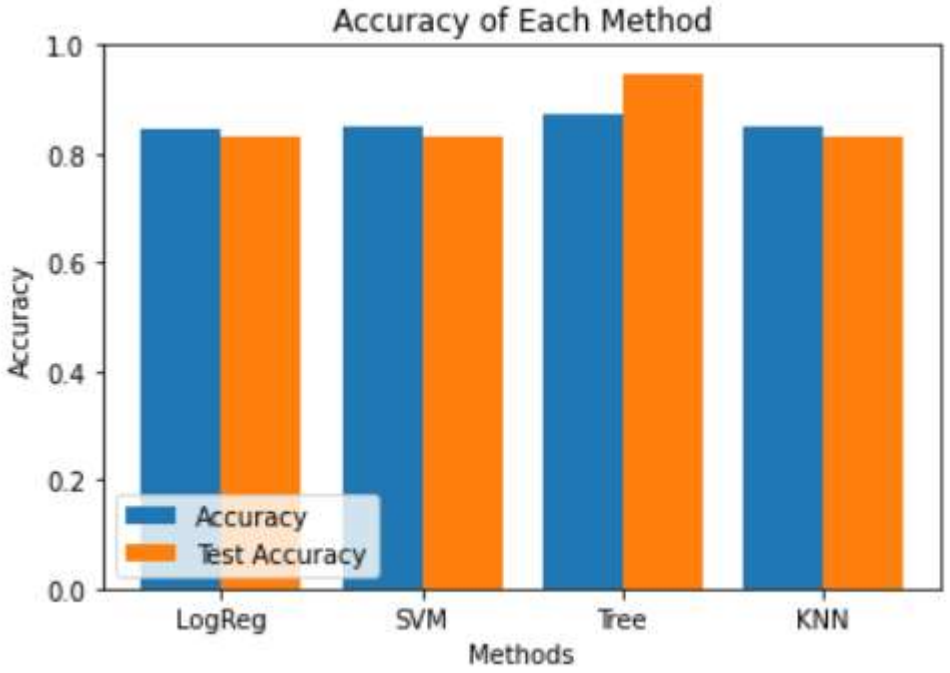
The major one is Florida, with numbers of successes ang failures

Interactive map with Folium



Predictive Analysis (Classification)

Use four different ways to classifier the result



Shows that the Decision Tree would be the best method for this task

CONCLUSION

- Entire process of data analysis for a specific case related to the business.
- Learning, reviewing the fundamental skills of data analysis.

APPENDIX