SpaceX Falcon 9 first stage Landing Prediction

Valeria Solozobova

IBM Capstone Project

12.07.2024

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- Introduction - The aim of the project
- Data collecting and cleaning: two sources of data
  - API
  - Web Scraping
- Exploratory Data Analysis
- Geospatial Visualization
- Dashboard example
- Machine Learning
- Discussion
- Conclusion

INTRODUCTION

Falcon 9 is the world's first orbital class reusable rocket. Reusability allows SpaceX to refly the most expensive parts of the rocket, which in turn drives down the cost of space access.

1. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

2. We will gather information about SpaceX , perform data analysis and visualizations.

3. We will also determine if SpaceX will reuse the first stage. Instead of using rocket science to determine if the first stage will land successfully, you will train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

METHODOLOGY

- API request and Web Scraping with BeautifulSoup (with Python)
- Data Analysis with SQL
- Data Visualization with Python
  - Seaborn charts
  - Geospatial data with Folium
  - Plotly Dashboard
- Machine Learning – Prediction of Landing Outcome (with Python)
  - Classification models

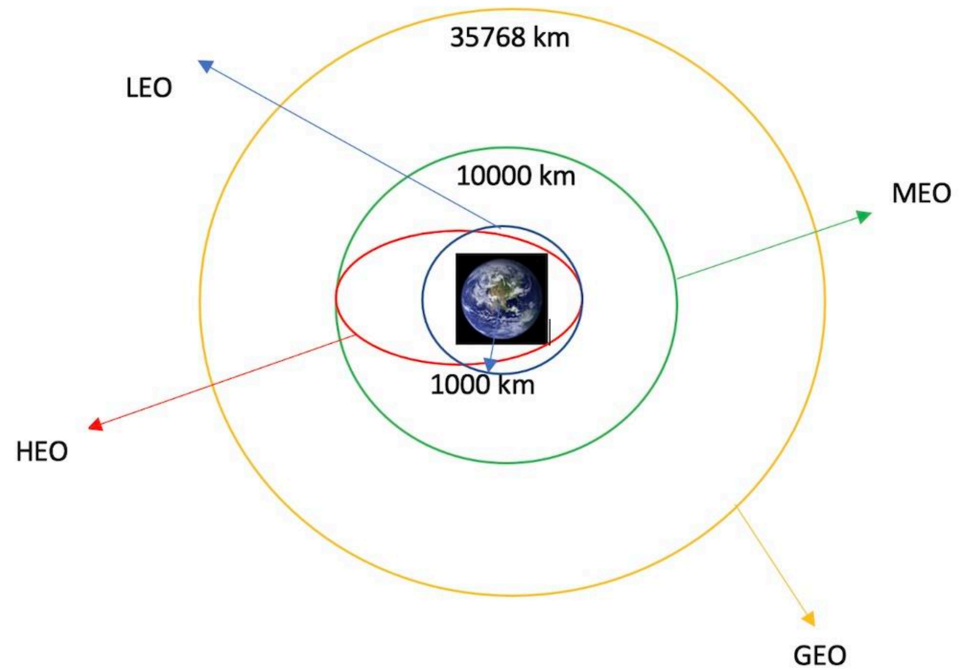
RESULTS

From database I found an information about four Launch Sites - **CCAFS LC-40** and **CCAFS SLC-40**: are together <u>Cape Canaveral Space</u> Launch Complex 40 , **VAFB SLC 4E** Vandenberg Air Force Base Space Launch Complex 4E **(SLC-4E)**, and Kennedy Space Center Launch Complex 39A **KSC LC 39A**

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

LaunchSite	Orbit	
CCAFS SLC 40	ES-L1	1
	GEO	1
	GTO	18
	HEO	1
	ISS	16
	LEO	5
	MEO	3
	SSO	1
	VLEO	9
KSC LC 39A	GTO	9
	ISS	5
	LEO	2
	SO	1
	VLEO	5
VAFB SLC 4E	PO	9
	SSO	4

The launches were dedicated for various orbits



Orbits types

RESULTS

Some curious insights obtained with SQL queries:

SUM(PAYLOAD_MASS__KG_)
619967

AVERAGE_PAYLOAD_F9_11
2534.6666666666665

SUCCESSFUL	FAILED
100	1

Mission_Outcome
Success
Failure (in flight)
Success (payload status unclear)
Success

Landing_Outcome
Failure (parachute)
No attempt
Uncontrolled (ocean)
Controlled (ocean)
Failure (drone ship)
Precluded (drone ship)
Success (ground pad)
Success (drone ship)
Success
Failure
No attempt

First_time_ground_pad
2015-12-22

Landing_Outcome	CASES
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

between 2010 and 2017

RESULTS

Some curious insights obtained with SQL queries:

COUNT(DISTINCT Booster_version)
97

The heaviest boosters:

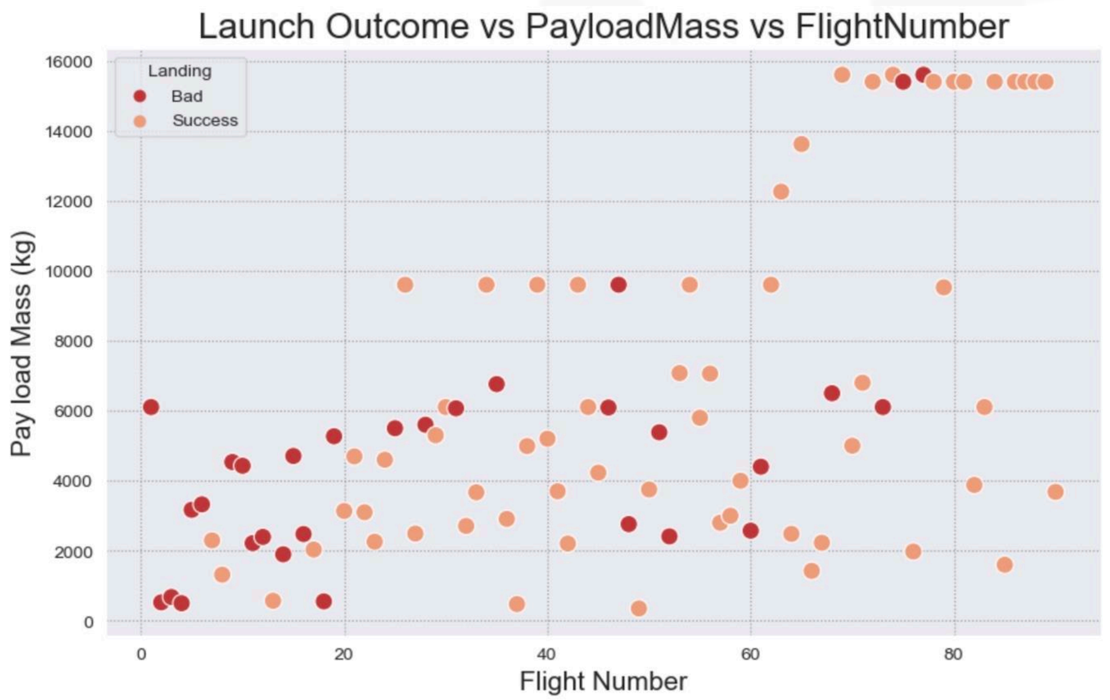
Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

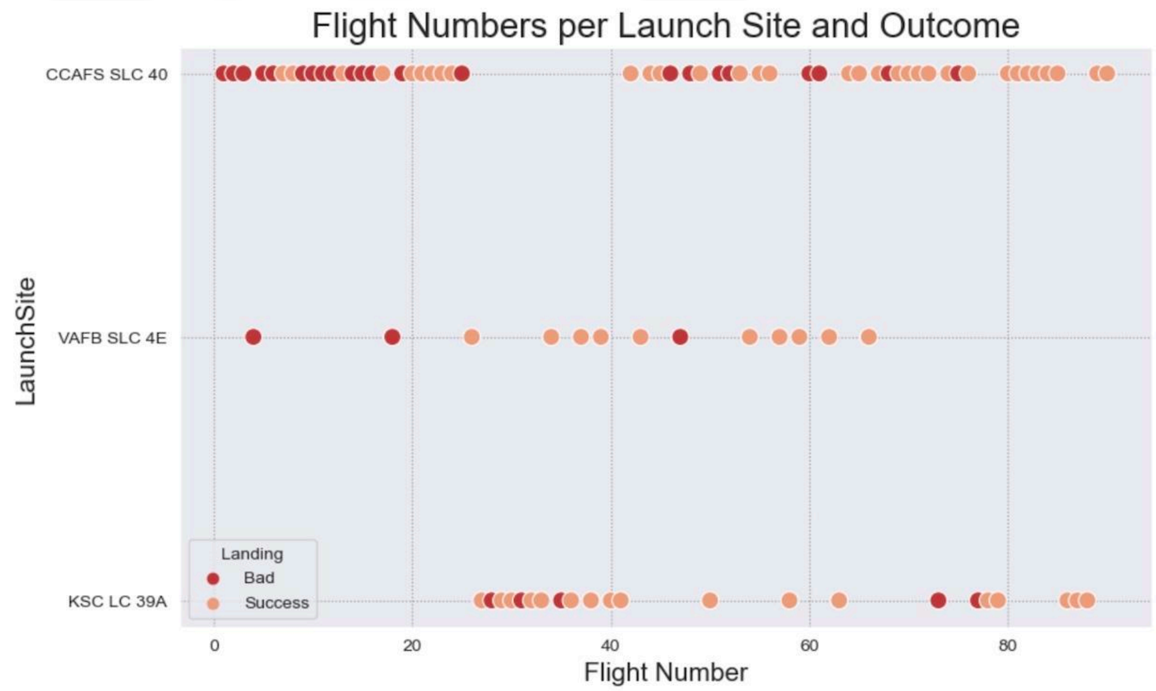
The booster (with success(drone_ship) with Payload Mass 4000 –6000 kg:

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

RESULTS

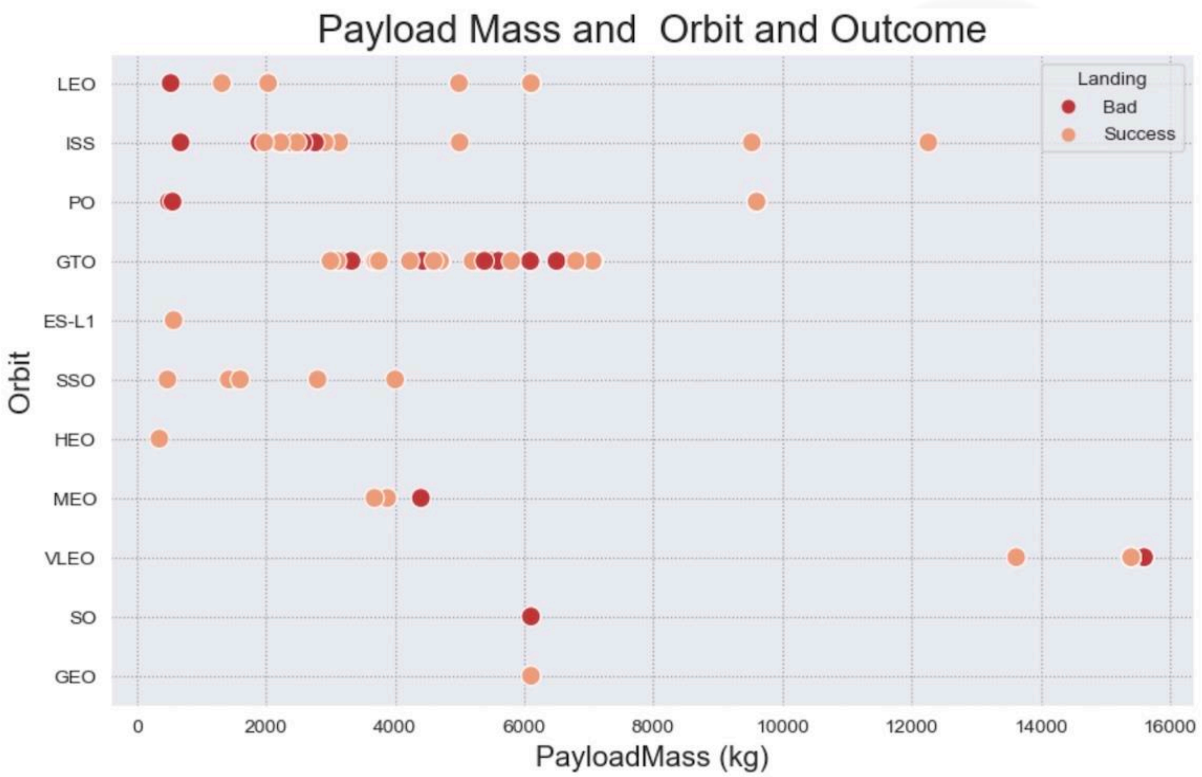
Launch Outcomes in dependance of Flight Number vs Payload Mass and Flight Number vs Launch Site

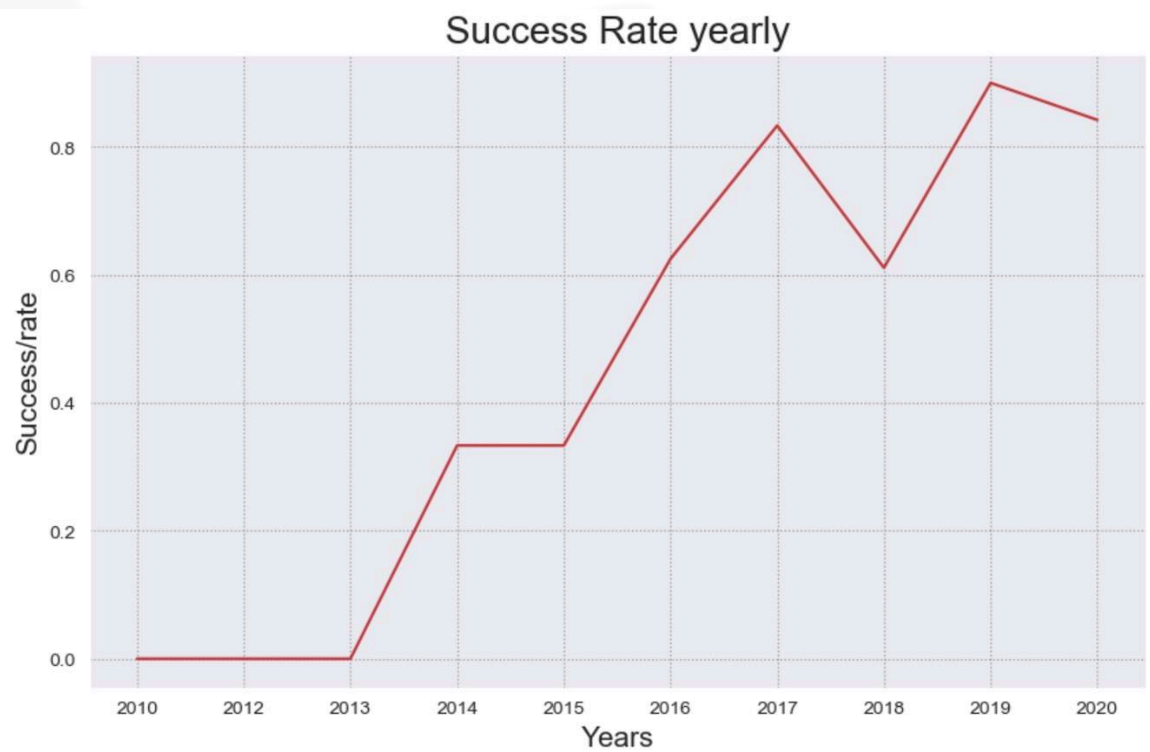




Higher Payload Mass correlates with successful outcome, as well as KSC LC 39A Launch Site, and higher Flight Number

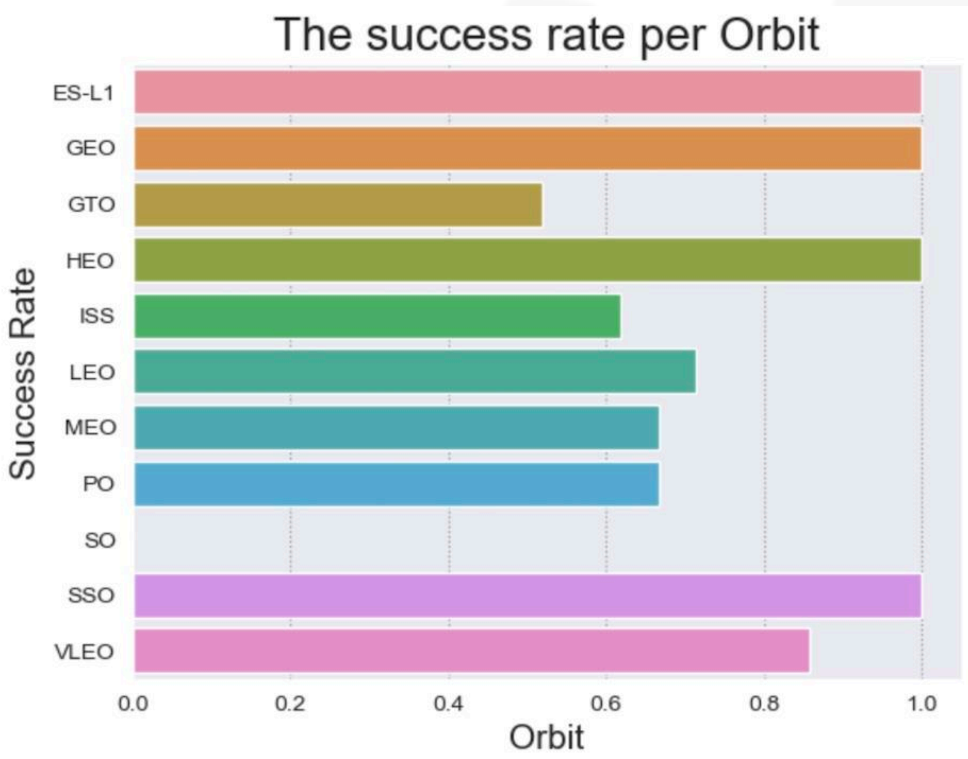
RESULTS

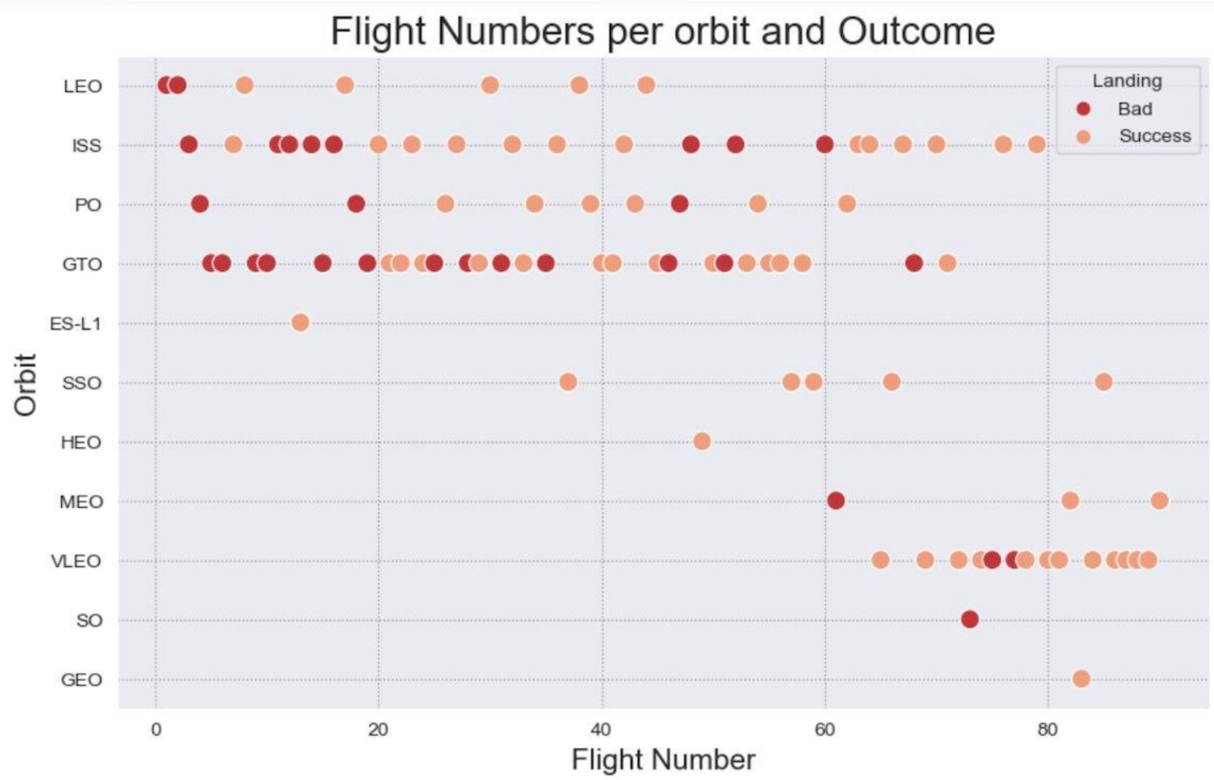




The heaviest Booster were started only form launch sites on Atlantic coast.
The success rate was growing with time from 2010 to 2020.

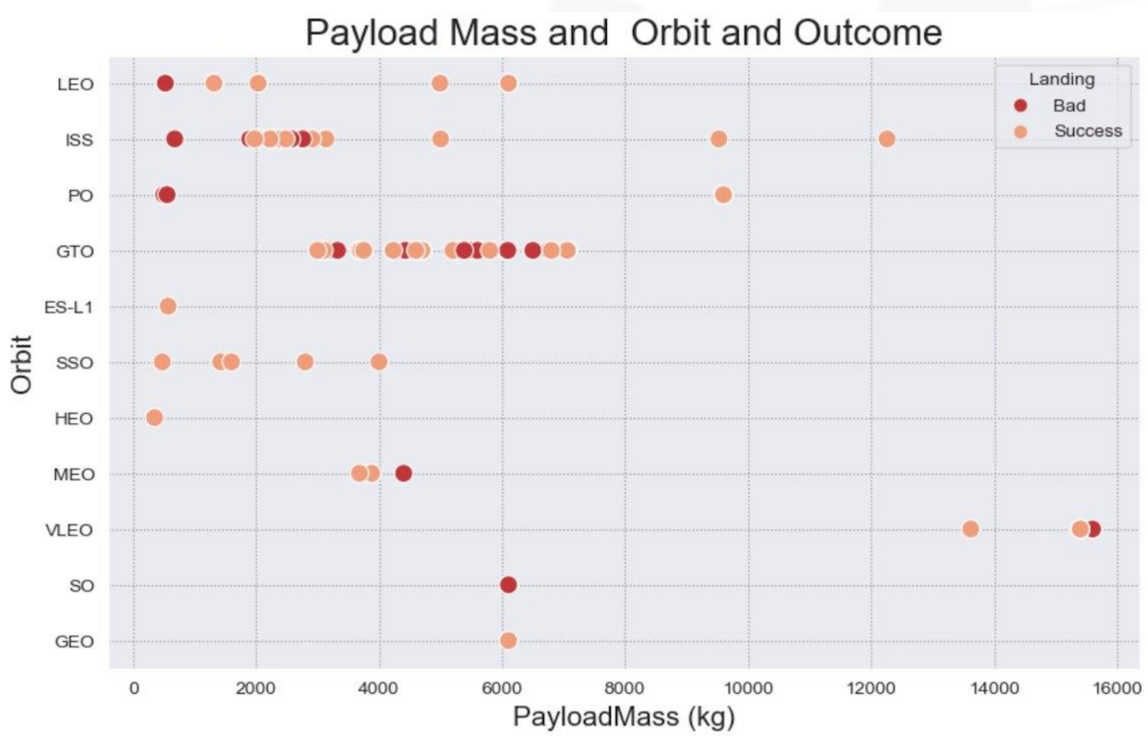
RESULTS





The launches to SSO are the most successful, however they are only 5, another orbit with good successful rate is VLEO

RESULTS

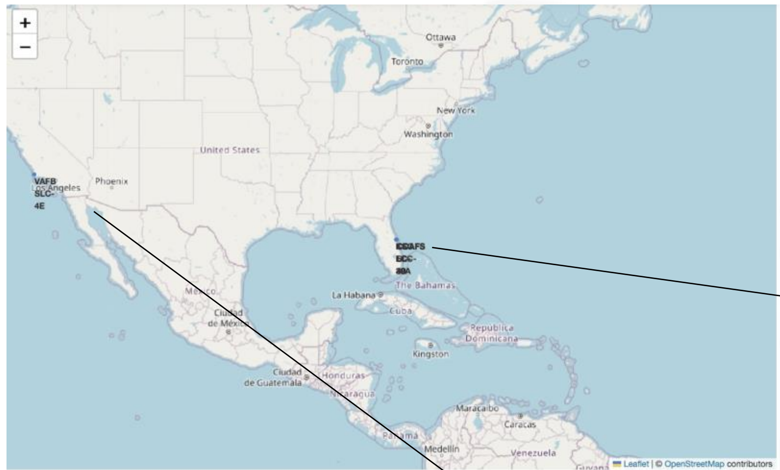


To orbit VLEO were sent the heaviest boosters, sometiems to ISS

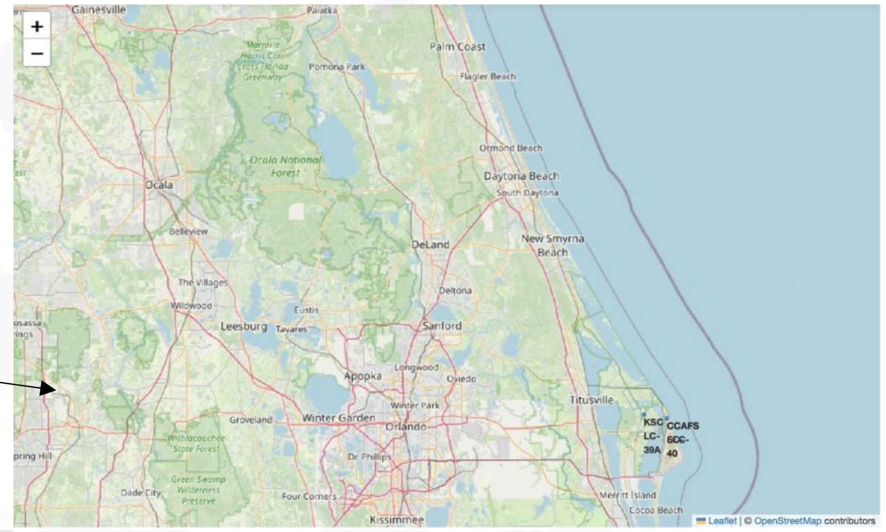
RESULTS

Geospatial Data with Folium: SpaceX Launch Sites

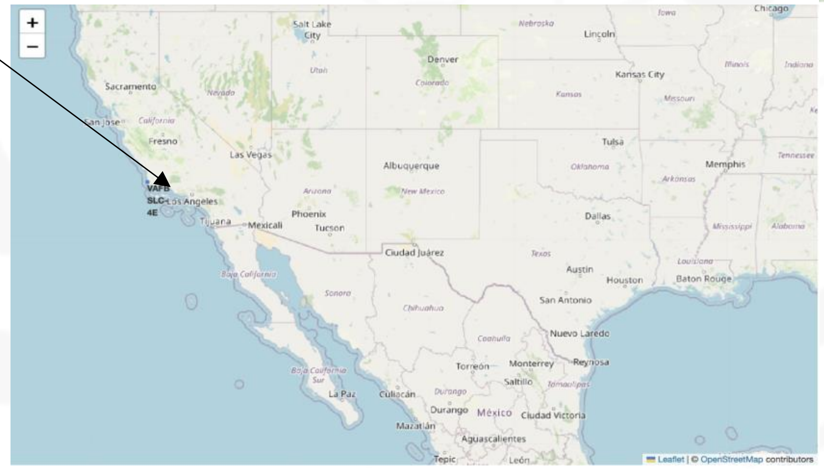
***The data provided here are for the 56 flights between 2010 and 2018**



KSC LC 39A,
CCAFS SLC-40,
CCAFS LC-40:
by Atlantic Ocean



VAFB SLC 4E:
by Pacific Ocean

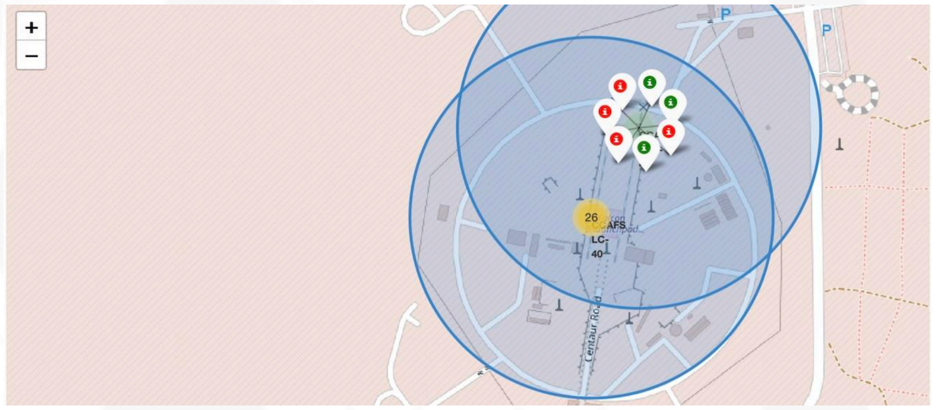


RESULTS

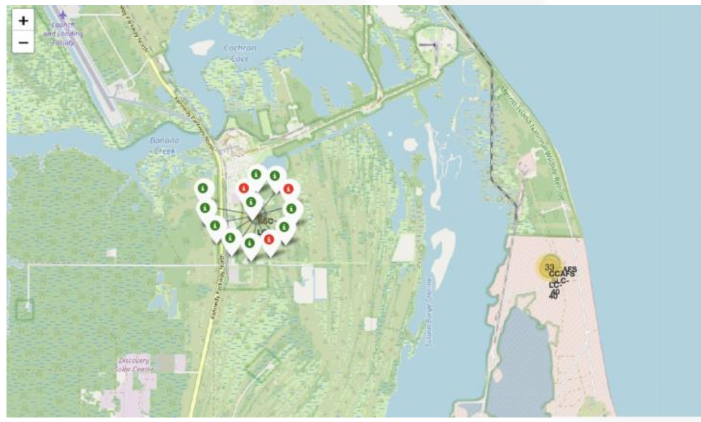
Launch Outcomes from various Launch Sites:

***The data provided here are for the 56 flights between 2010 and 2018**

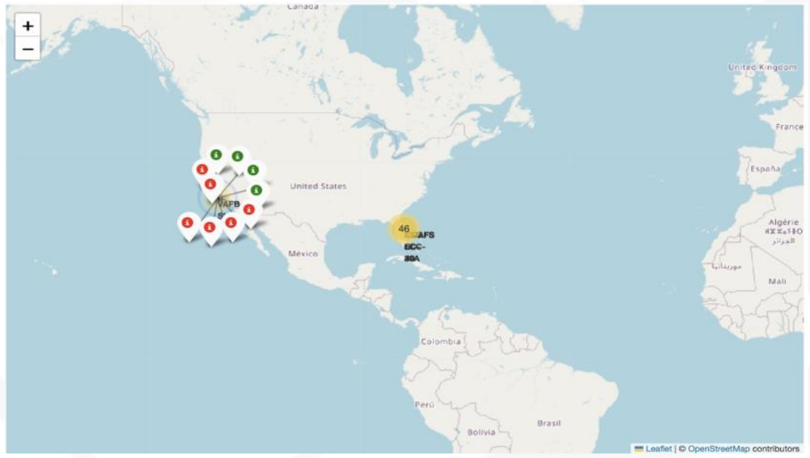
Launches from CCAFS LC-40



Launches from CCAFS SLC-40



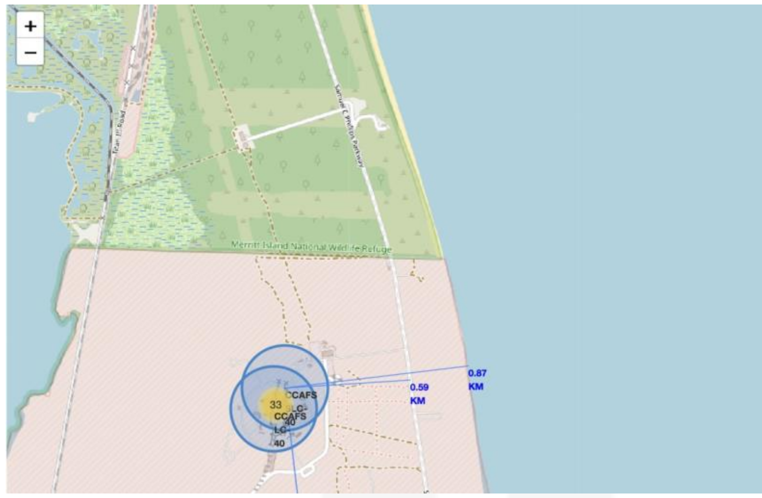
Launches from KSC LC 39A LC-40

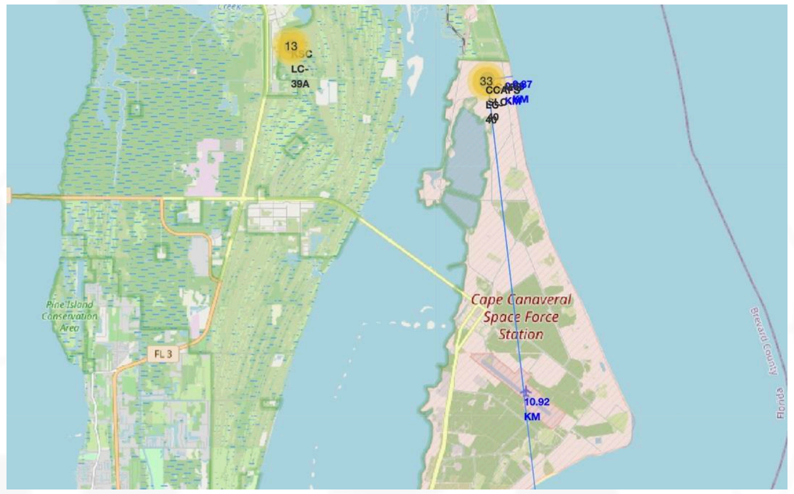


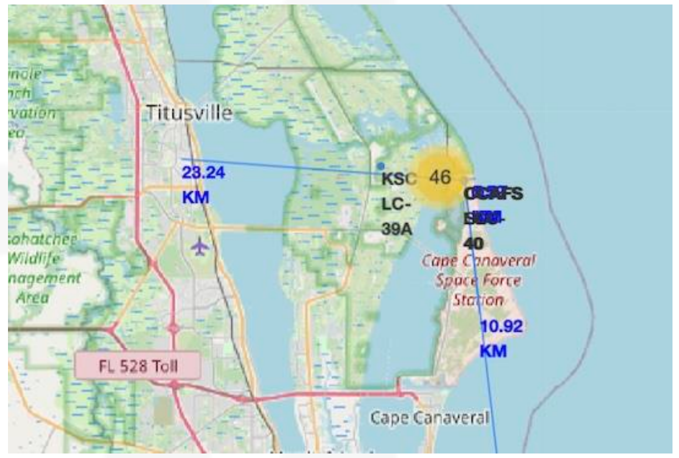
Launches from VAFB SLC 4E

RESULTS

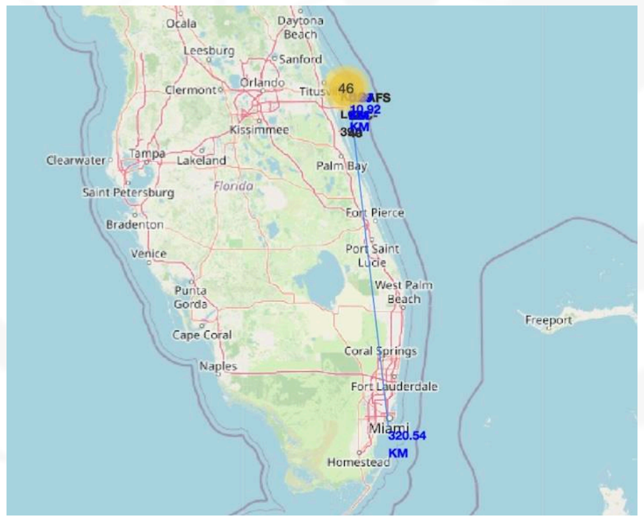
The distances from Launch Site to Strategic Objects: CCAFS LC-40





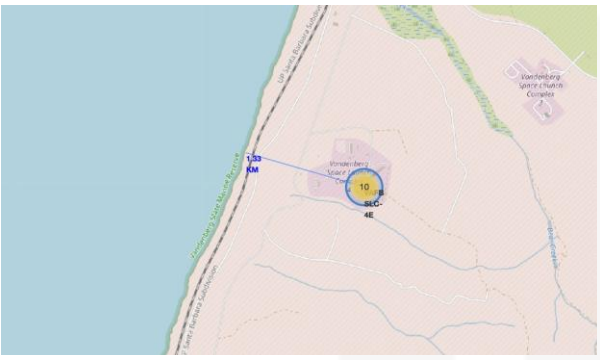


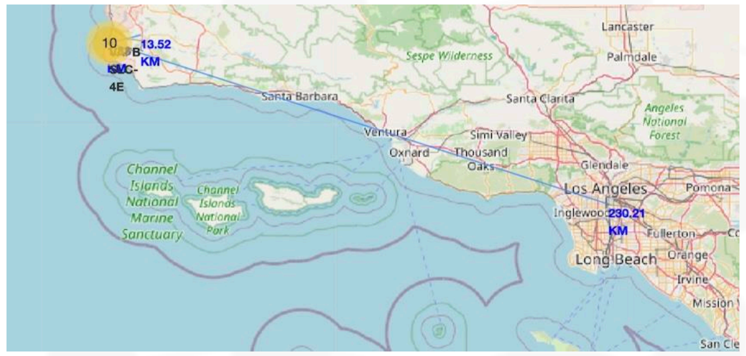
To ocean coast line: 0,87 km
To highway: 0,59 km
To airport: 10,92 km
To Miami (the megapolis): 320 km
To Titusville (the nearest small town): 23,24 km



RESULTS

The distances from Launch Site to Strategic Objects: VAFB SLC 4E





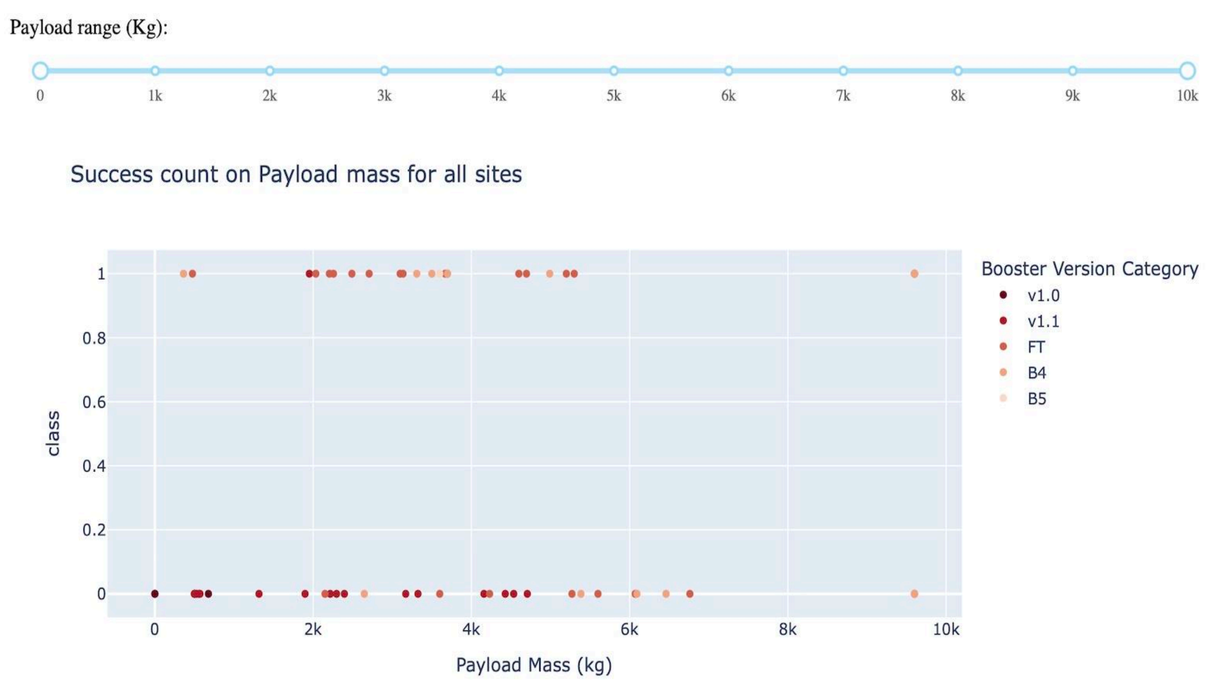
To highway: 1,33 km
To Lompoc airport: 13,52
To Los-Angeles (the megapolis): 230 km
To Lompoc (the nearest small town): 13,52 km

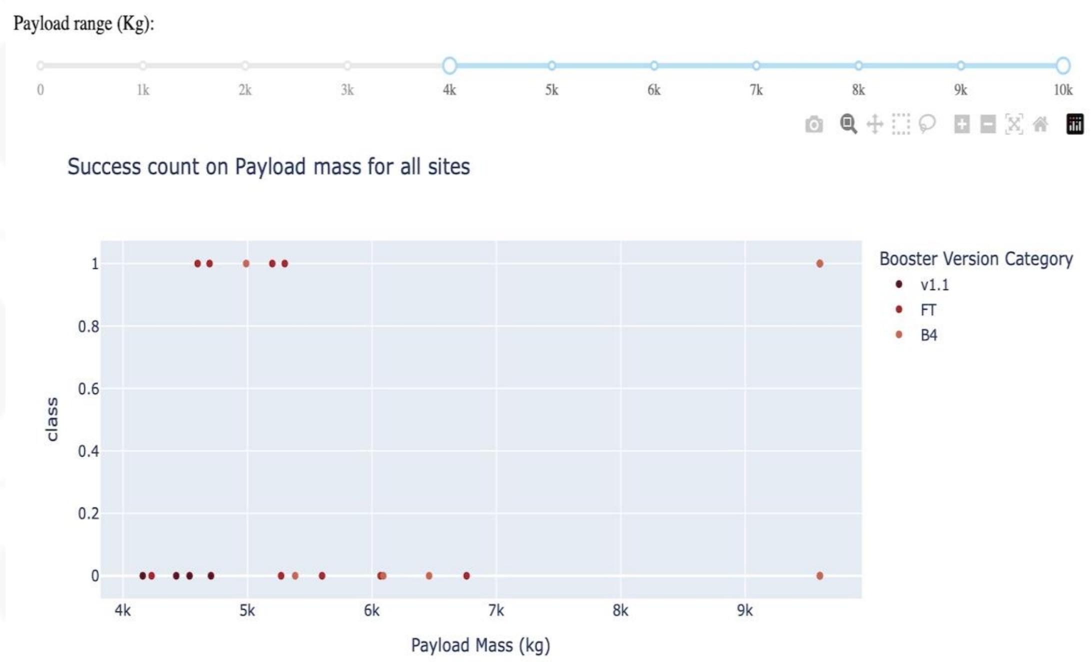
For all of the Launch Sites we can observe:
1)very close to ocean (safety reason and for landing of the first stage)
2)very close to highway and airport (for logistic reasons)
3)more far from small towns (safety and logistic) and very far from megapolises as Miami or Los-Angeles (safety)

RESULTS

The dashboard:

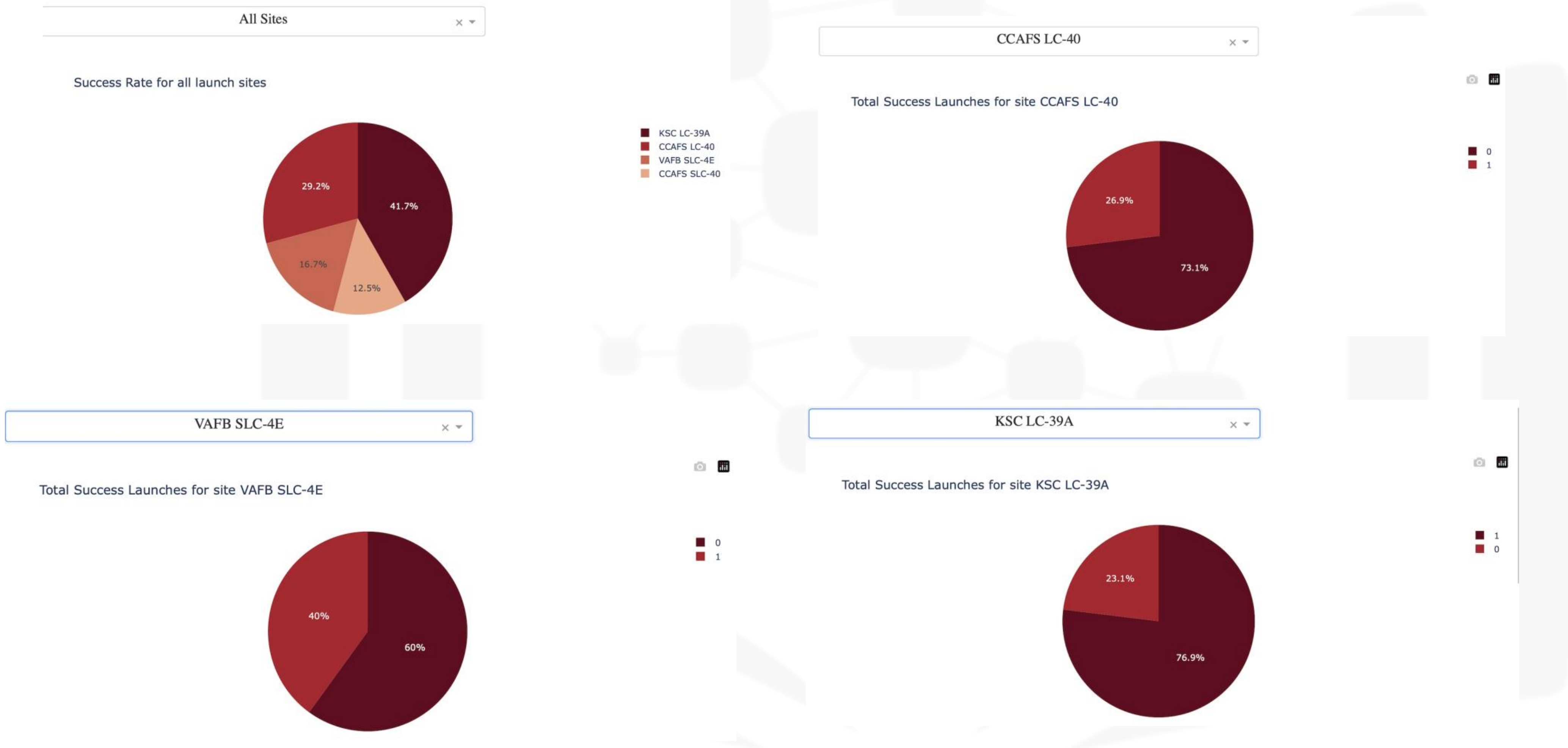
***The data provided here are for the 56 flights between 2010 and 2018**





RESULTS

Dashboard: success rate for launch sites



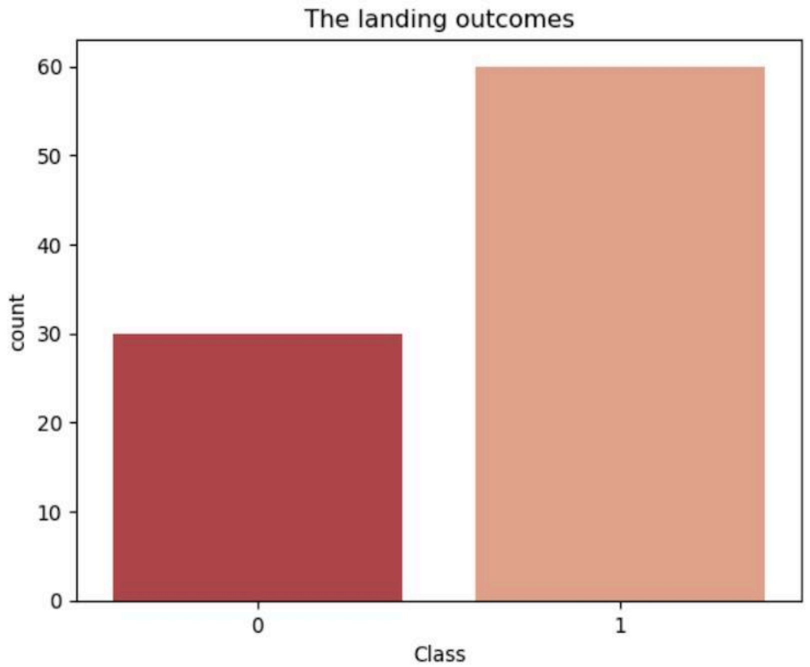
IBM Developer

SKILLS NETWORK

RESULTS

Prediction of Launch Outcome: Machine Learning

Binary classification problem
90 flights , data obtained from API, there is some imbalances

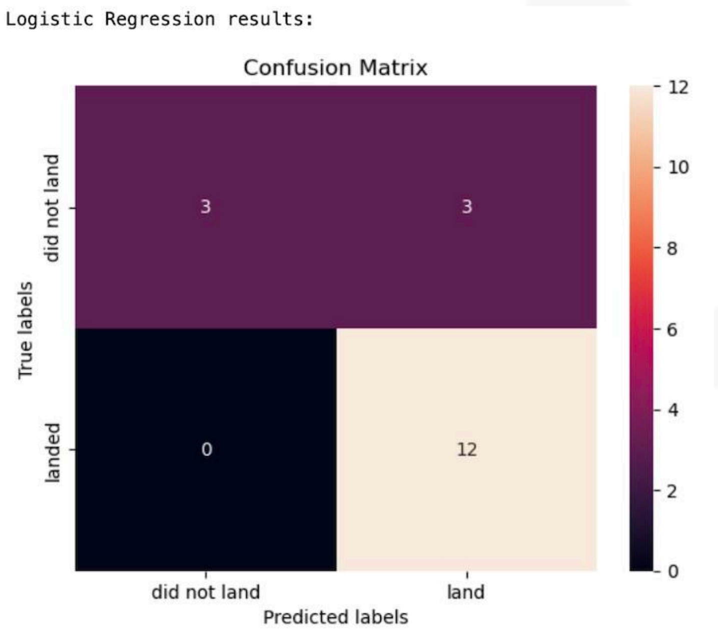


0: no success
1: success

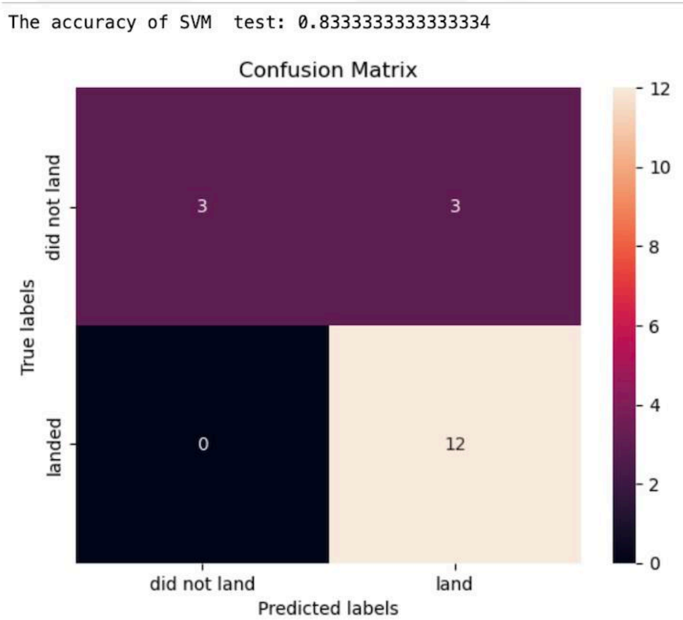
Models to use:
Logistic Regression
K- nearest neighbors
Decision Tree
Support Vector Machines
Random Forest

RESULTS

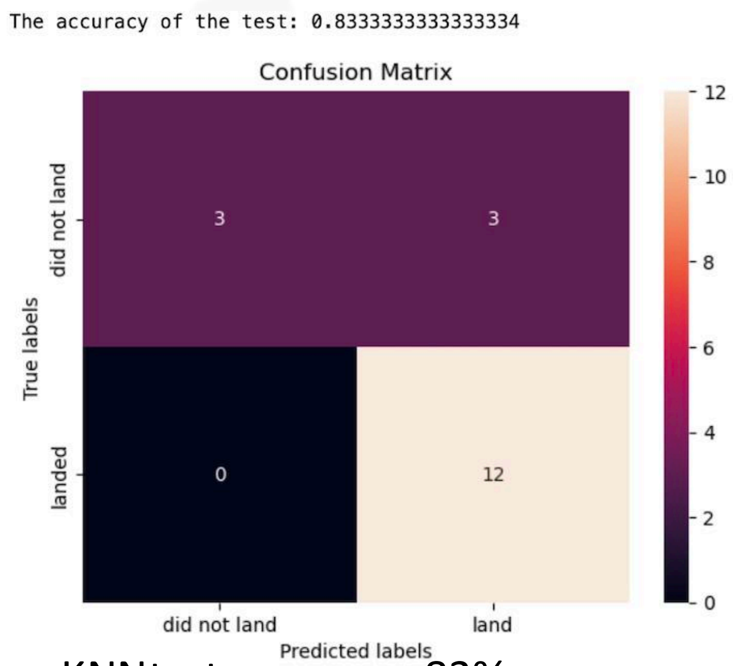
Prediction of Launch Outcome: Logistic Regression(LR) and Support Vector Machine(SVM), K-nearest neighbors(KNN)



LR: test accuracy 83%



SVM: test accuracy 83%

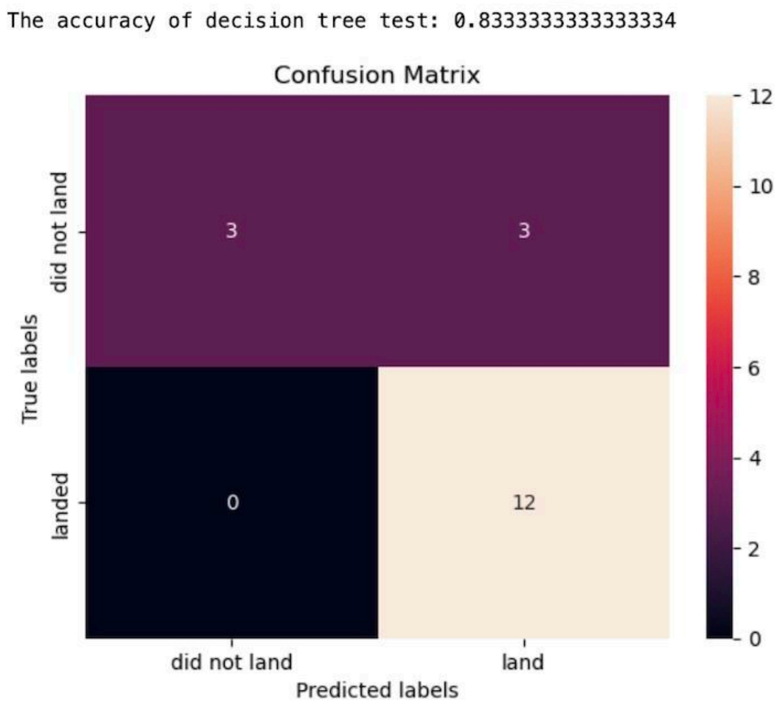


KNNtest accuracy 83%

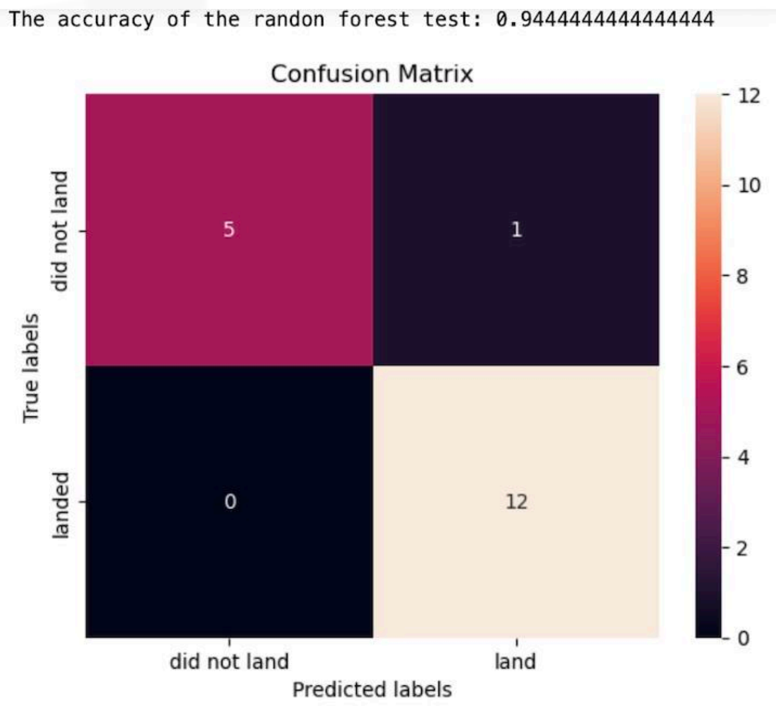
The predicted successful launches were all correct (12 of 12), but not the not-successful missions (3 of 6)

RESULTS

Prediction of Launch Outcome: Decision Trees (DT) and Random Forest (RF)



DT: test accuracy 83%



RF: test accuracy 94%

Decision Trees accuracy is simialar to LR, KNN, SVM, however, Random Forest could predict the launch outcome with a better accuracy, 94, and the prediction of not-successful launches was increased (5 of 6)

DISCUSSION

Data visualization revealed some trends in launches of SpaceX Falcon9

- It is becoming better and better with a time
- The best succes rate was achieved with a starts from **KSC LC 39A**
- Launches to the orbit SSOS and VLEO are the most successfull
- Mapping of launch sites reveal some common features for them, despite they are located on various paces in USA; close proximity to ocean, to highway, to airport, and far distance to large cities, and relative close distance to smaller cities

- Some problems were caused that the data provided were not consistent (especially for Dashboard and Folium, it was a dataset for the flights between 2010 and 2018, and more outcomes were not- successful). That is why, the results of visualization are contradictory with other visualization and with dataset used for Machine learning

*For sure one can use the other datasets for Dashboard and for Mapping with Folium (results would be different that in checkpoint)

DISCUSSION

Results of Machine Learning:

- Machine Learning algorithms were showing high accuracy score, Random Forest was the best, it predicted the outcome with 94% of accuracy
- However, the selected features for prediction are not entirely correct. It was proposed to use all the featares from the dataset, but it has to be reconsidered. We cannot use such parameters as 'Reused' (booster was reused or not, ReuseCount' (how many times it was reused), 'Flights' (how many flights made this booster), or 'LandingPad'(because if it is indicated, that means, the launch was successfull), etc. All these feautaures are indirect Outcomes.

- Maybe I will try to make new algorithms with reduced features set, or with dimensions reductions techniques as PSA), because it is interesting,how would be the results of classification

- Alternatively: I could use for the prediction the dataset obtained with scraping from Wikipedia, there are more than 200 launches, instead of 90, and some other features for prediction

CONCLUSION

- I scraped the data from Wikipedia and perform an API request to create the dataset
- Data were cleaned, wrangled, reorganized
- Data were visualized with Seaborn Charts
- The usage of Folium and Dashboard with Plotly were demonstrated
- The machine learning algorithms were developed and obtained one of the best : Random Forest

We can use this algorithm to predict the launch for SpaceX!

APPENDIX

Data sources:
https://api.spacexdata.com/v4/

https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv