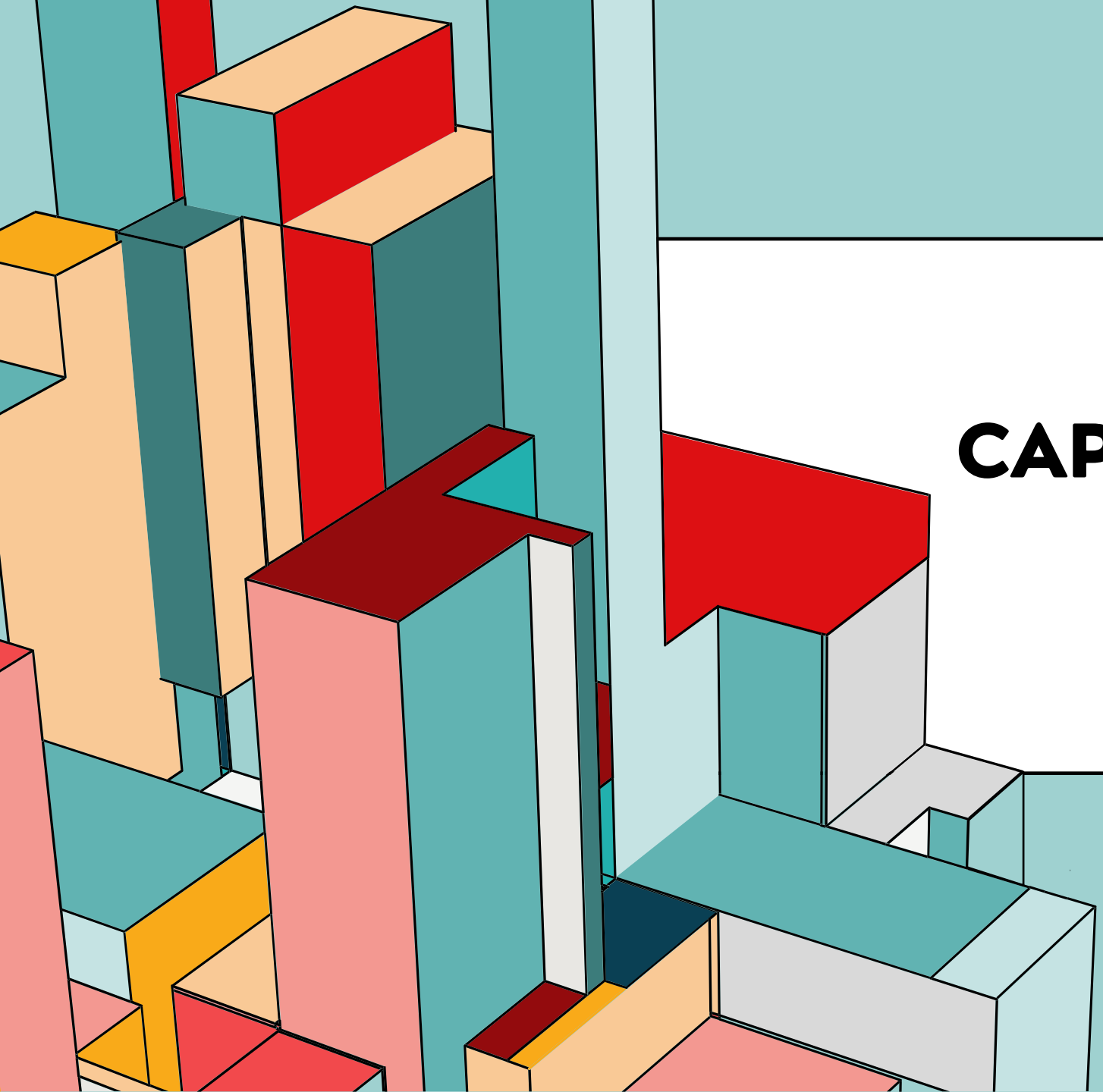


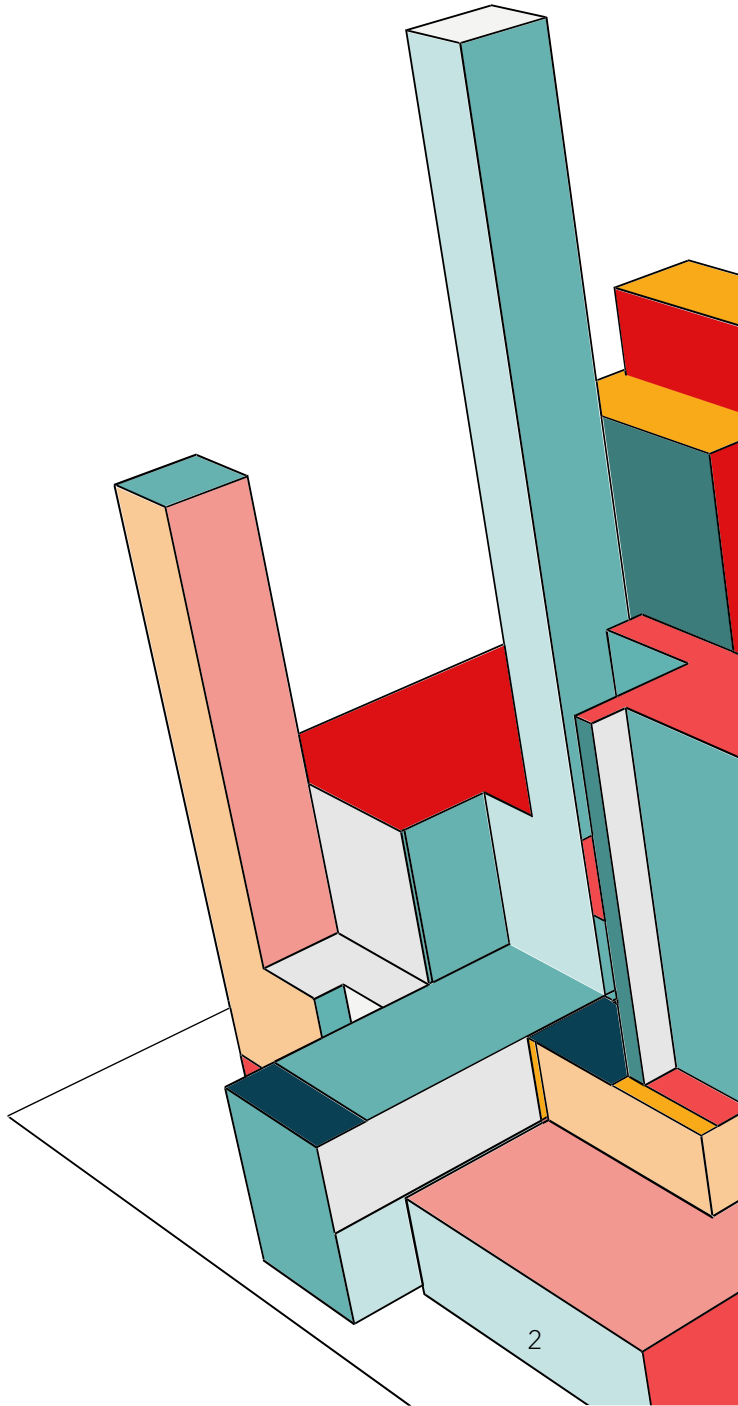
CAPSTONE PROJECT

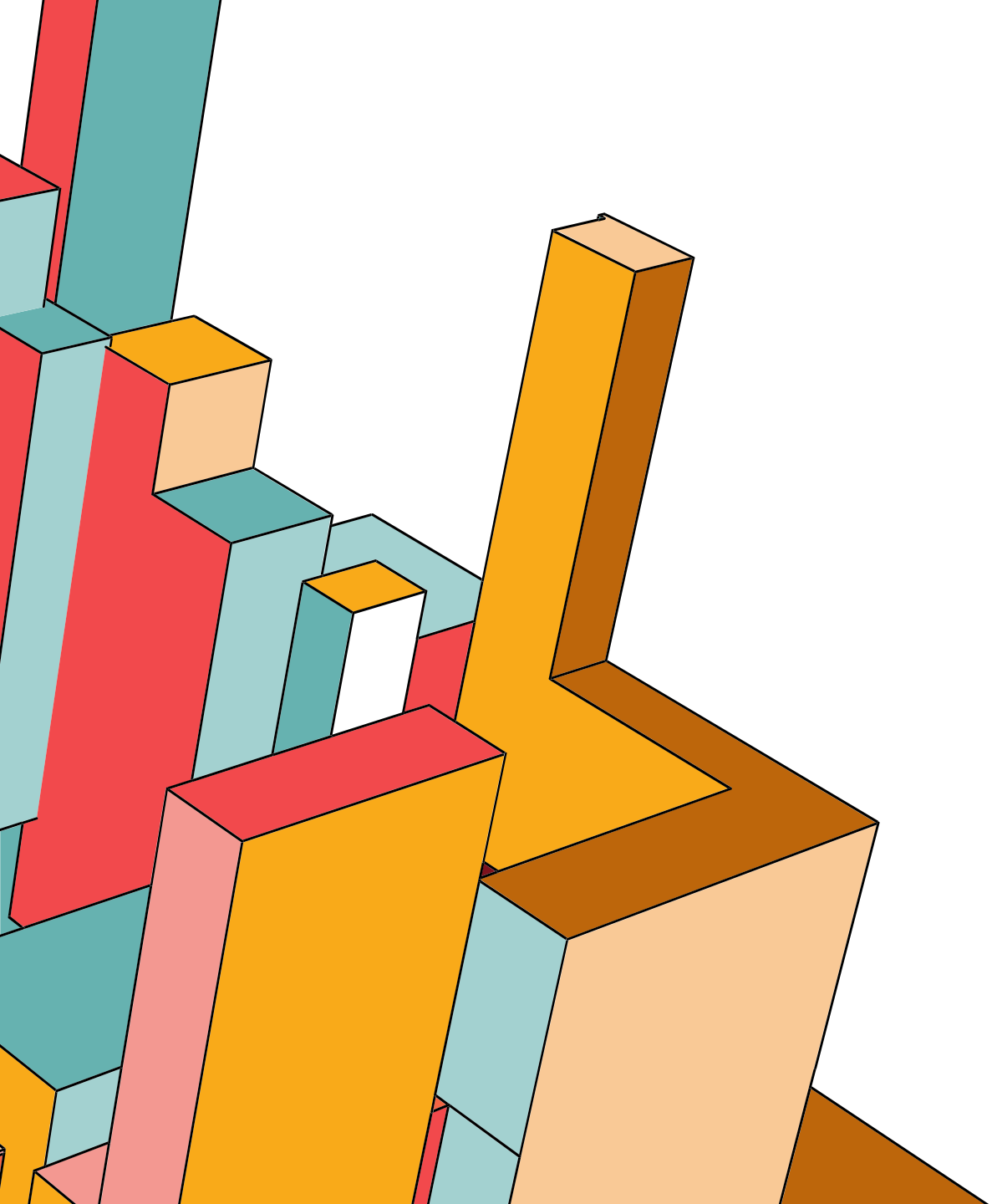
NUHA ALANEZI

09/2023

OUTLINE

1. Summary
2. Introduction
3. Methodology
4. Results
5. Conclusion





SUMMARY

- Data gathering

- handling the data

- Data Visualization and Exploratory Analysis

- Exploratory SQL Data Analysis

- Using Folium to create an interactive map

- Using Plotly Dash to Create a Dashboard

- Classification

-based predictive analysisa list of all outcomes

- Findings from Exploratory Data Analysis-

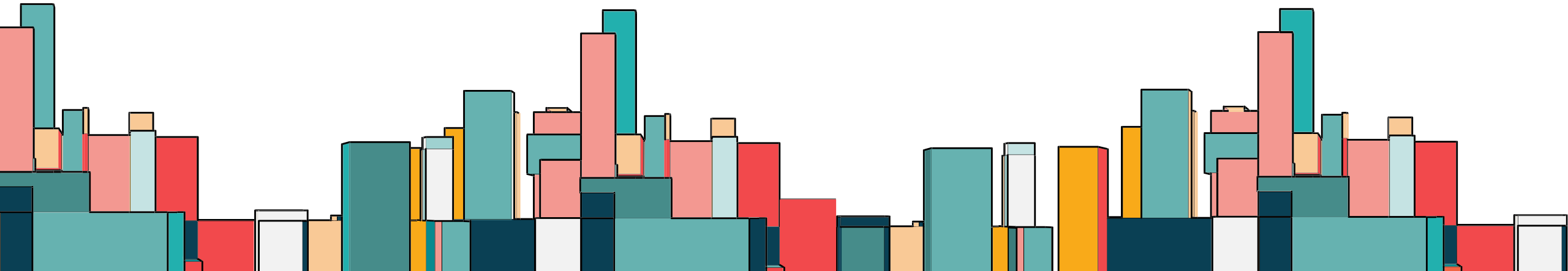
Screenshots of an interactive analytics demo- Results of the predictive analysis

INTRODUCTION

❑ Background and setting of the project

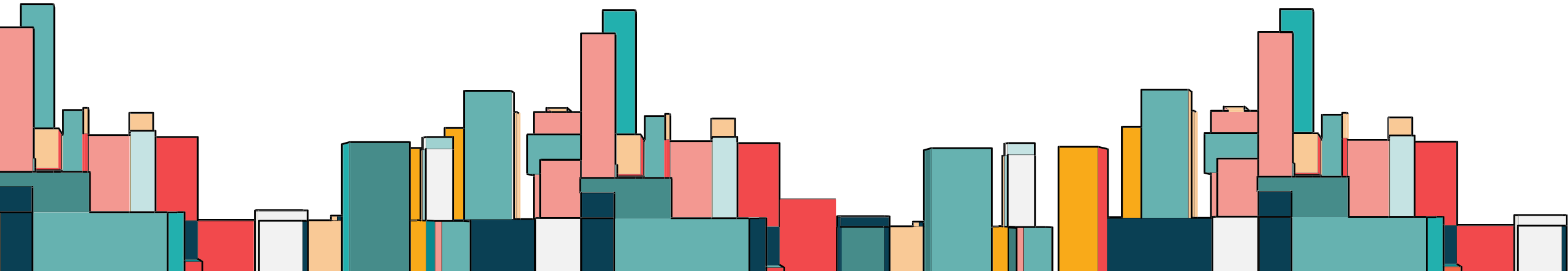
The most prosperous business of the commercial space age, SpaceX has reduced the cost of space travel. On its website, the company promotes Falcon 9 rocket launches, which are advertised for a cost of 62 million dollars. By comparison, other companies charge upwards of 165 million dollars per launch.

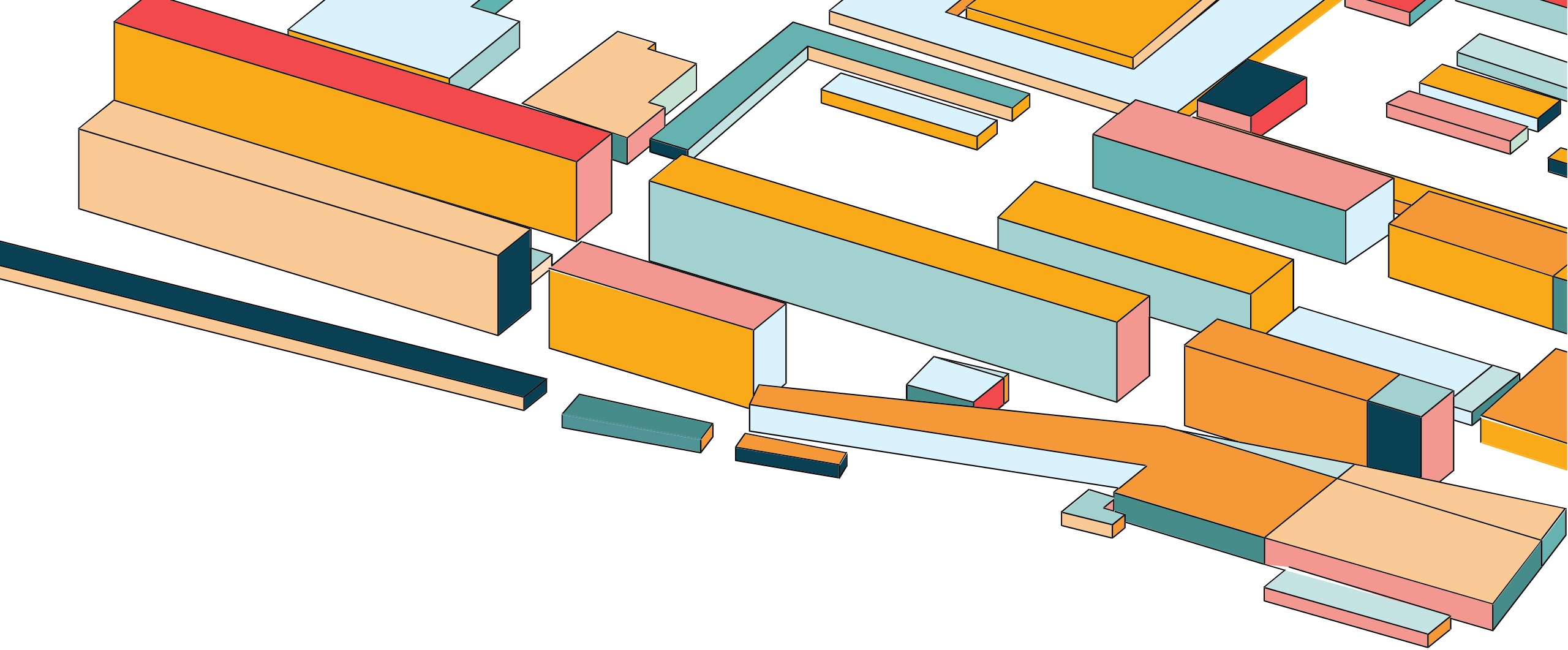
So, if we can figure out whether the first stage will land, we can figure out how much a launch will cost. We are going to make a prediction about whether SpaceX will reuse the first stage based on available data and machine learning models.



INTRODUCTION

- Query to be answered
- How does the success of the first stage landing depend on factors such cargo mass, launch site, number of flights, and orbits?
- Does the frequency of successful landings rise with time?
- In this situation, what is the best algorithm that may be utilized for binary classification?

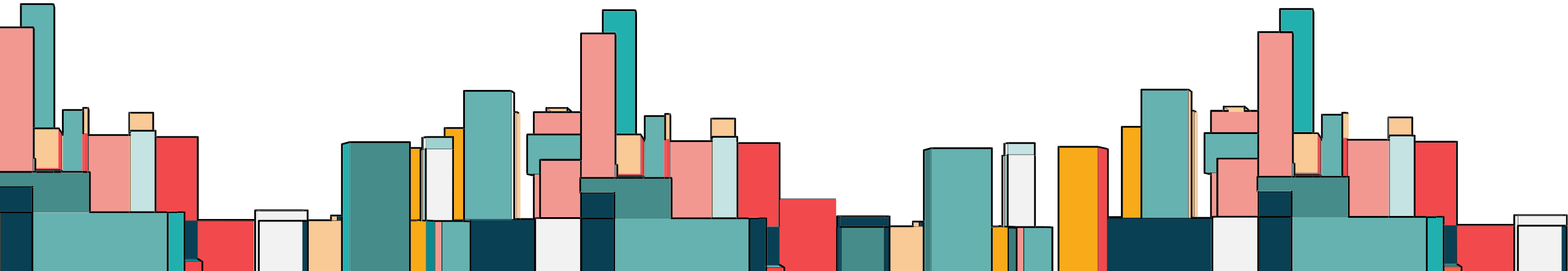




METHODOLOGY

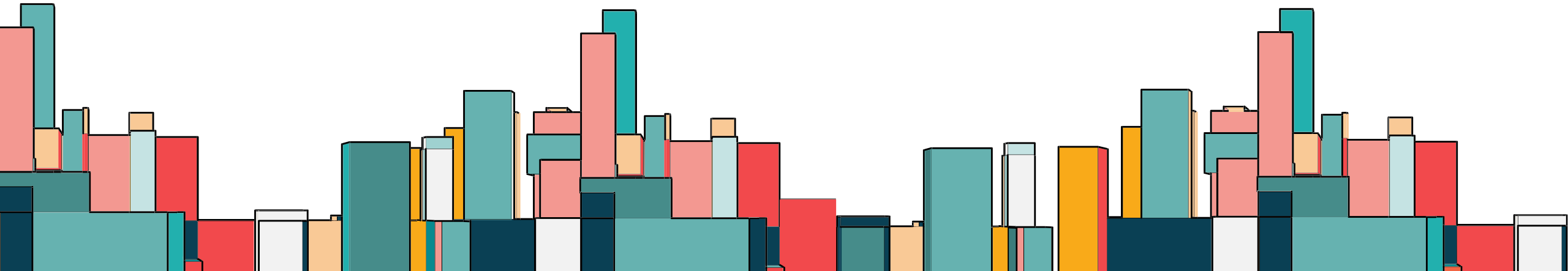
METHODOLOGY

- Data collecting techniques:
- Using the SpaceX Rest API
- Using Wikipedia web scrapingperformed data wrangling, including filtering the data, dealing with missing values, and converting the data to a binary classification using One Hot Encoding.
- carried out exploratory data analysis (EDA) utilizing SQL and graphicscarried out interactive visual analyses utilizing Plotly Dash and Foliumused classification models to perform predictive analyses.



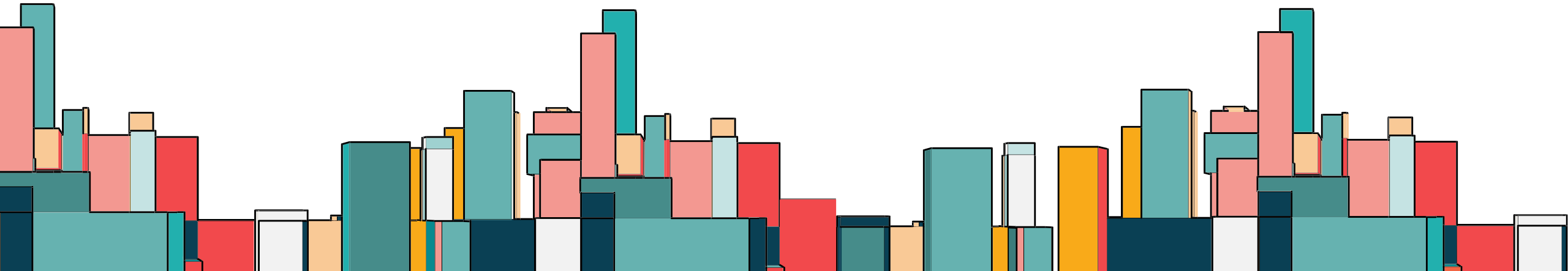
METHODOLOGY

- Data was gathered using a combination of API queries to SpaceX's REST API and web scraping of information from a table in the Wikipedia entry for the company.
- In order to obtain comprehensive data about the launches for a more in-depth analysis, we had to use both of these data collection techniques.FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, and Latitude are the data columns that can be accessed using the SpaceX REST API.
- May get data columns by utilizing Wikipedia. Web scraping includes the following information: flight number, launch site, payload, payload mass, orbit, customer, launch outcome, version booster, booster landing, date, and time.



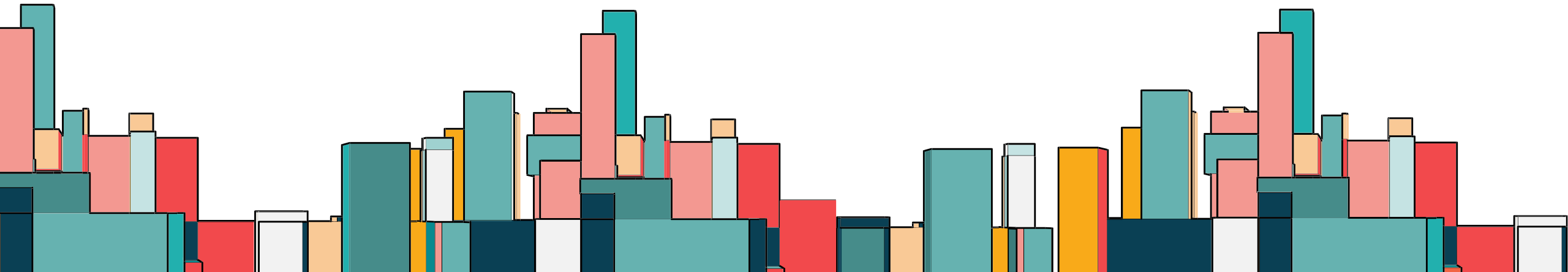
ALGORTHIMS THAT USED IN MACHINE LEARNING

- ❑ GridSearchCV
- ❑ LogReg
- ❑ SVM
- ❑ Decision Tree
- ❑ KNN models



RESULTS

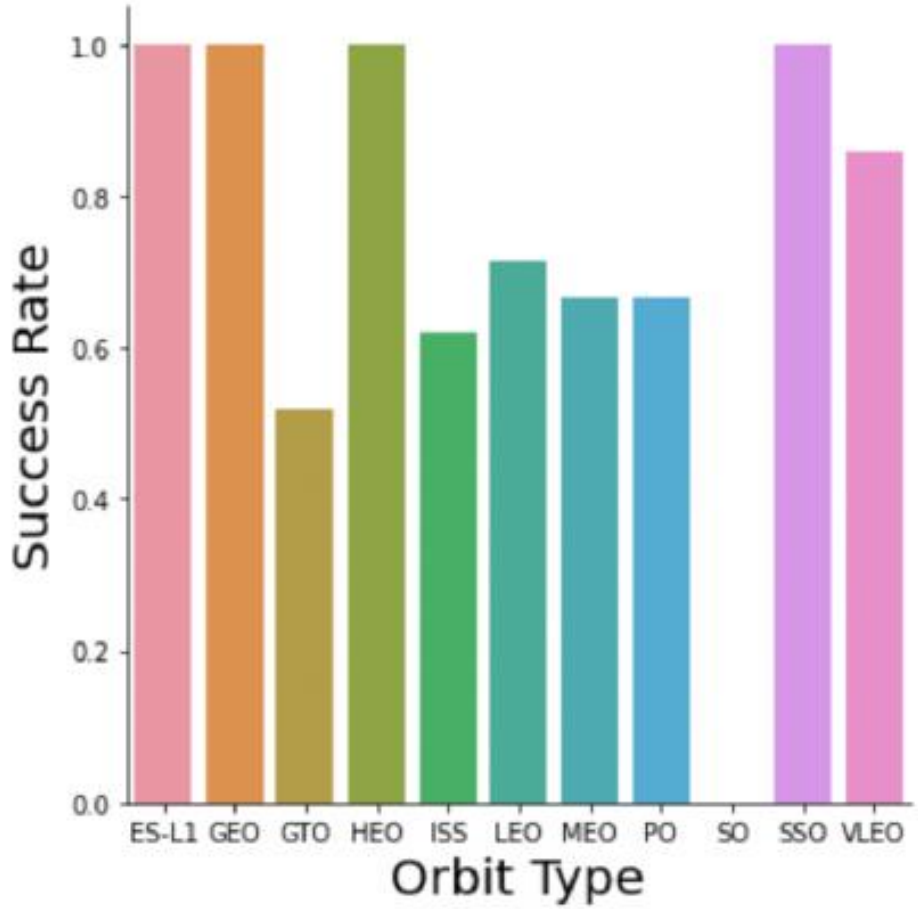
- Results of exploratory data analysis
- screenshots of an interactive analytics demonstration.
- Results of predictive analysis



RESULTS–EDA WITH VISIULAZTION

Success rate vs. Orbit

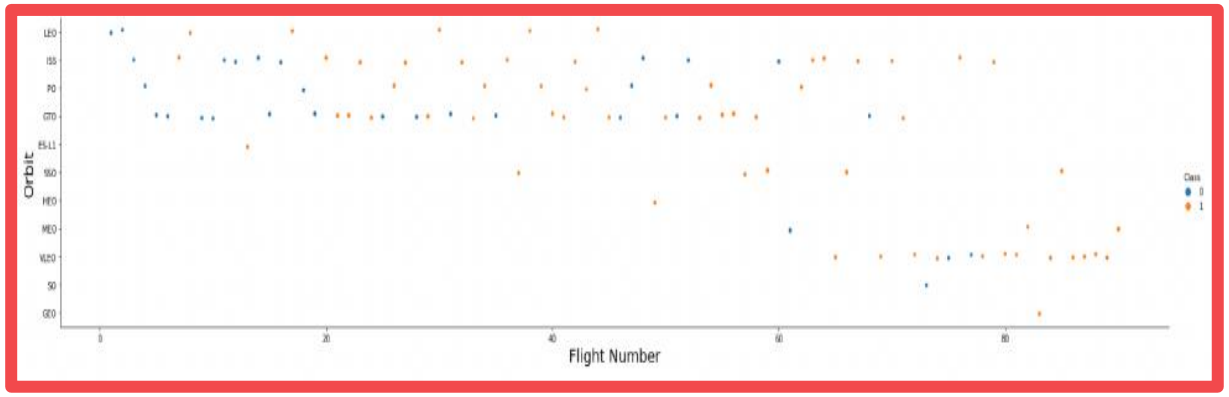
- Orbits that always succeed:GEO, HEO, SSO, and ES-L1• Orbits having a failure rate of 0%- SO• Orbits with a 50% to 85% success rate:- LEO, MEO, PO, GTO, ISS,



RESULTS-EDA WITH SQL

Flight Number vs. Orbit

- Success seems to be correlated with the number of flights when in LEO orbit, but there doesn't seem to be a correlation when in GTO orbit.



TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION RESULTS

```
In [10]: %sql select mission_outcome, count(*) as total_number from SPACEXDATASET group by mission_outcome;
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[10]:

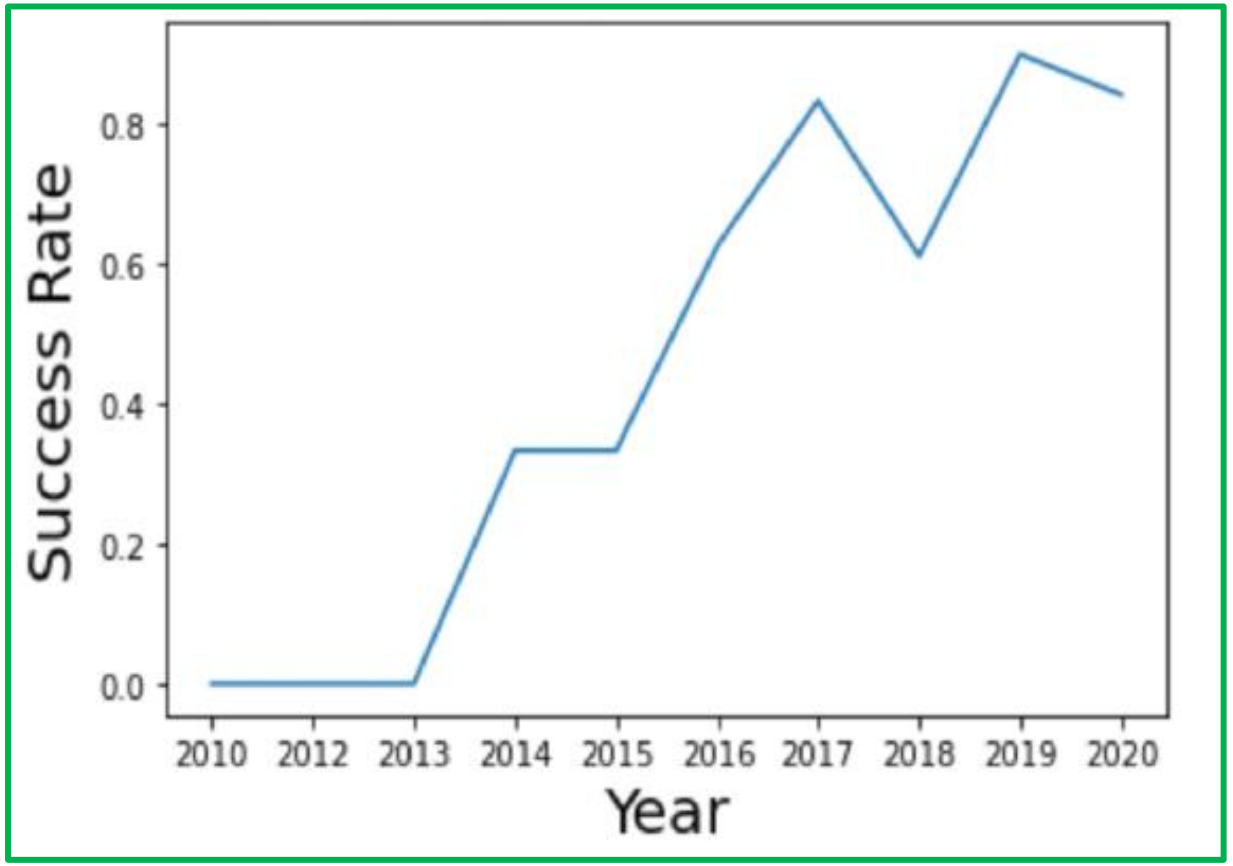
mission_outcome	total_number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

List the all number of successful and failure mission Results.

RESULTS

Launch success yearly trend

- Since 2013, the success percentage has risen continuously until 2020.



RESULTS

Launch site names begin with `CCA`

- Sbowing here 5 records that launch sites begin with the str 'CCA'.

```
In [5]: %sql select * from SPACEXDATASET where launch_site like 'CCA%' limit 5;
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[5]:

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

RESULTS

Distance from the launch site KSC LC-39A to its proximities

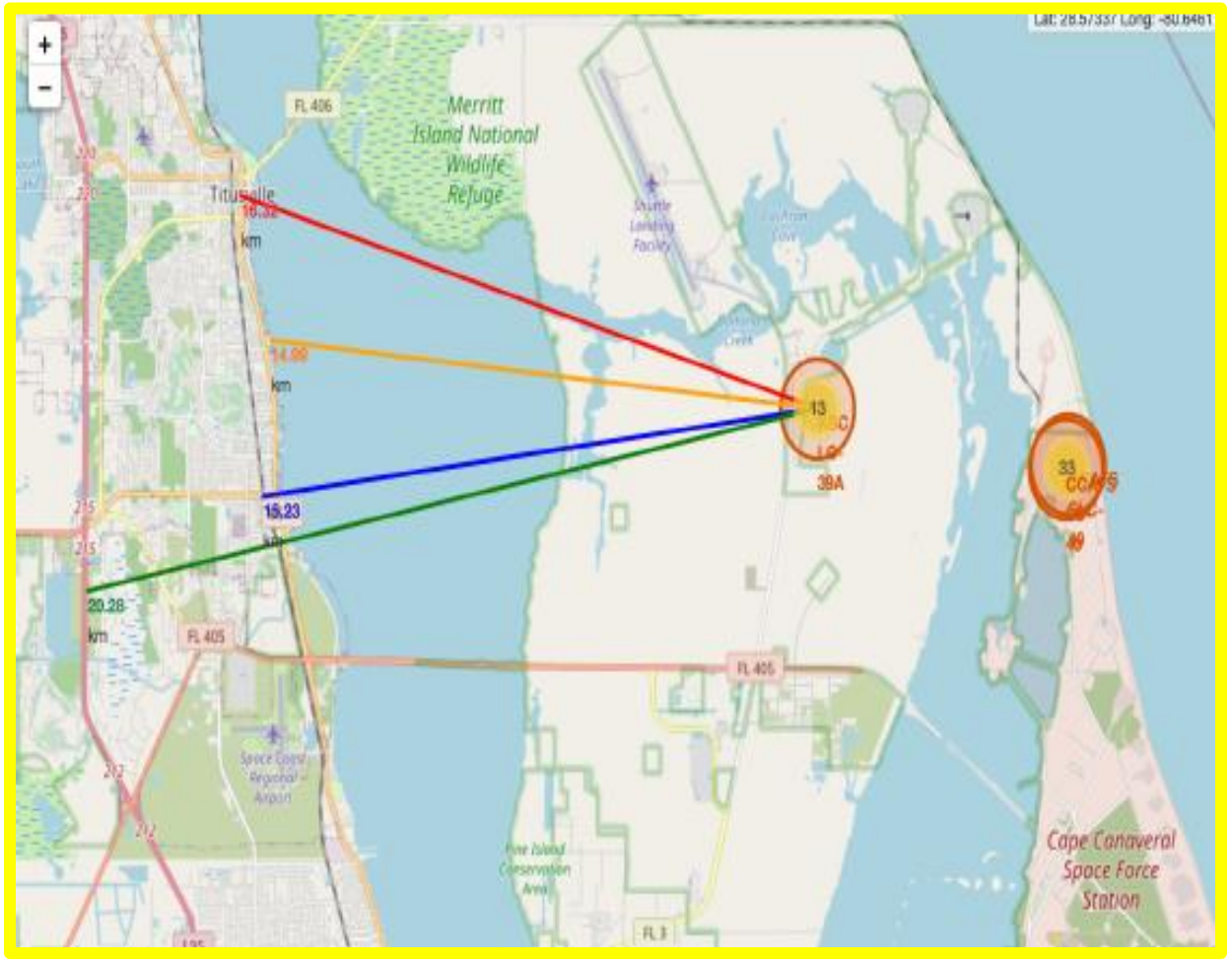
The visual study of the KSC LC-39A launch site reveals that it is:

- quite close to a railroad (15.23 kilometers)
- comparatively near a highway (20.28 km)
- quite close to the coast (14.99 km)

Additionally, Titusville (16.32 km) is relatively close to the launch location KSC LC-39A.

- A failed rocket can travel up to 15-20 km in a matter of seconds due to its high speed.

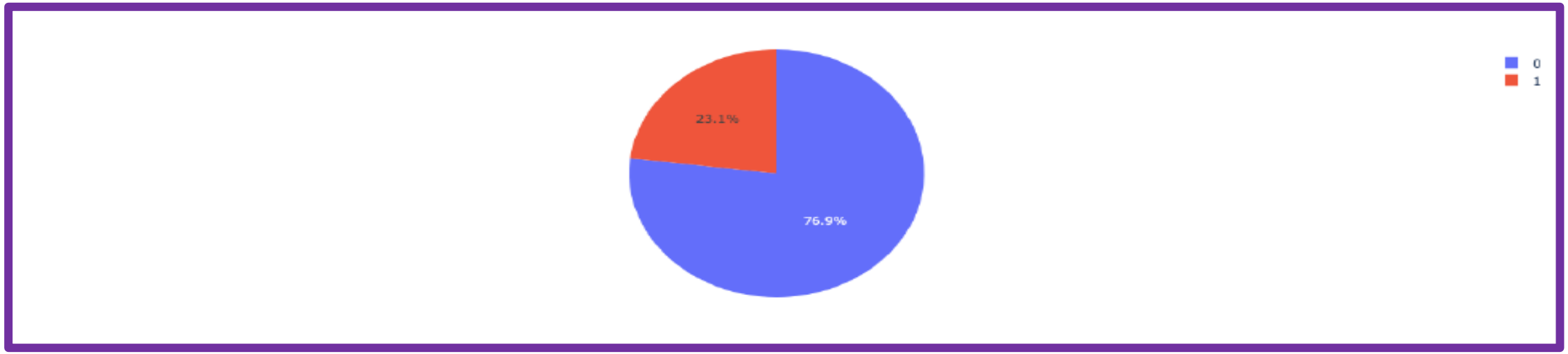
There is a chance that it could be dangerous for populated regions.



RESULTS

launch site with the highest rate of success

- The launch success percentage at KSC LC-39A is the greatest (76.9%), with 10 successful and only 3 unsuccessful landings.



RESULTS

For all sites, Payload Mass vs. Launch Outcome

- • The graphs demonstrate that the maximum success rate is achieved by payloads weighing between 2000 and 5500 kg.

RESULTS- CLASSIFICATION ACCURACY

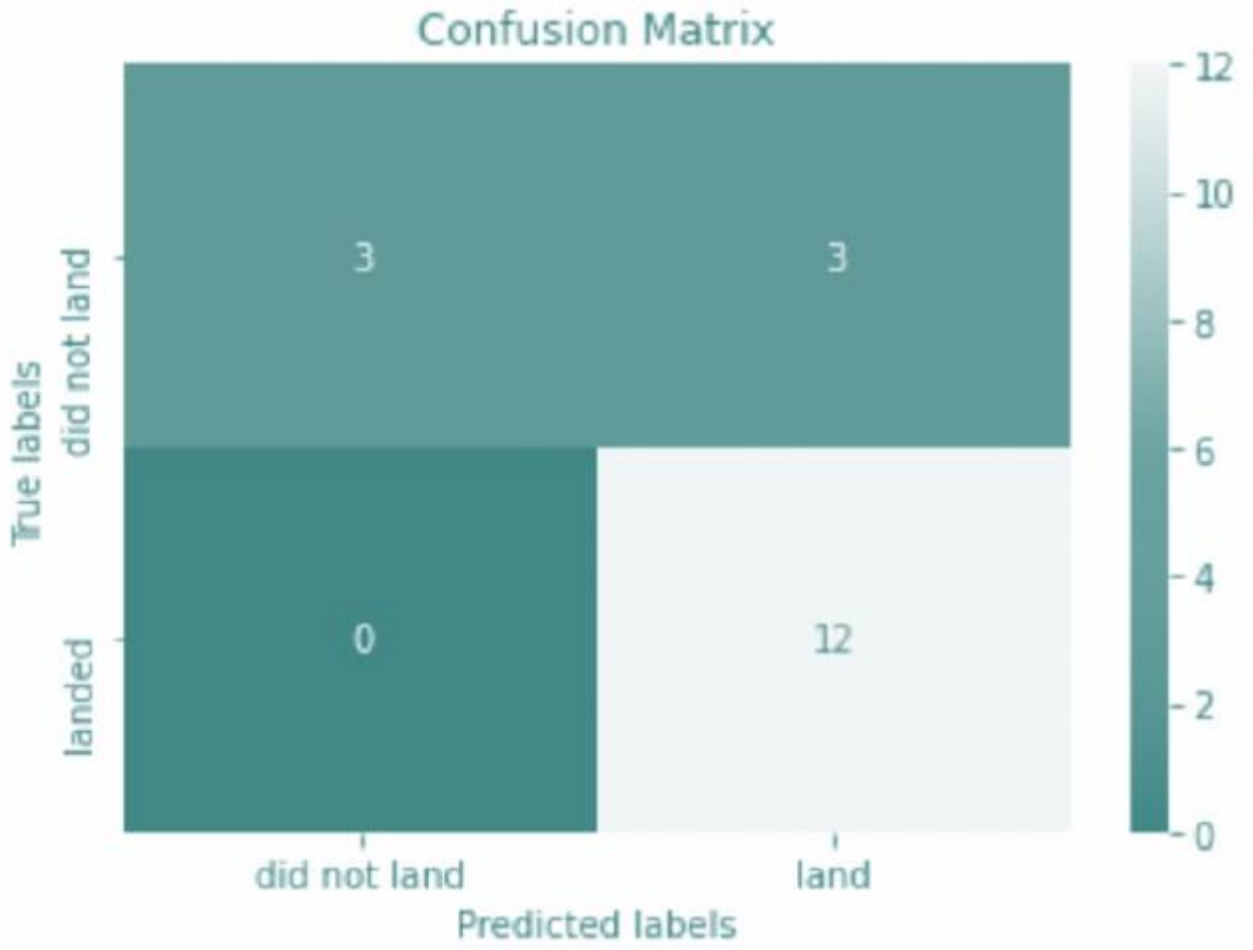
- You are unable to confirm which strategy performs the best in light of the Test Set's results.
- The low test sample size 18 samples may account for the Same Test Set results.
- As a result, we evaluated each approach using the entire dataset.
- The Decision Tree Model is the best model, according to the scores of the entire dataset. This model offers the highest accuracy in addition to higher scores.

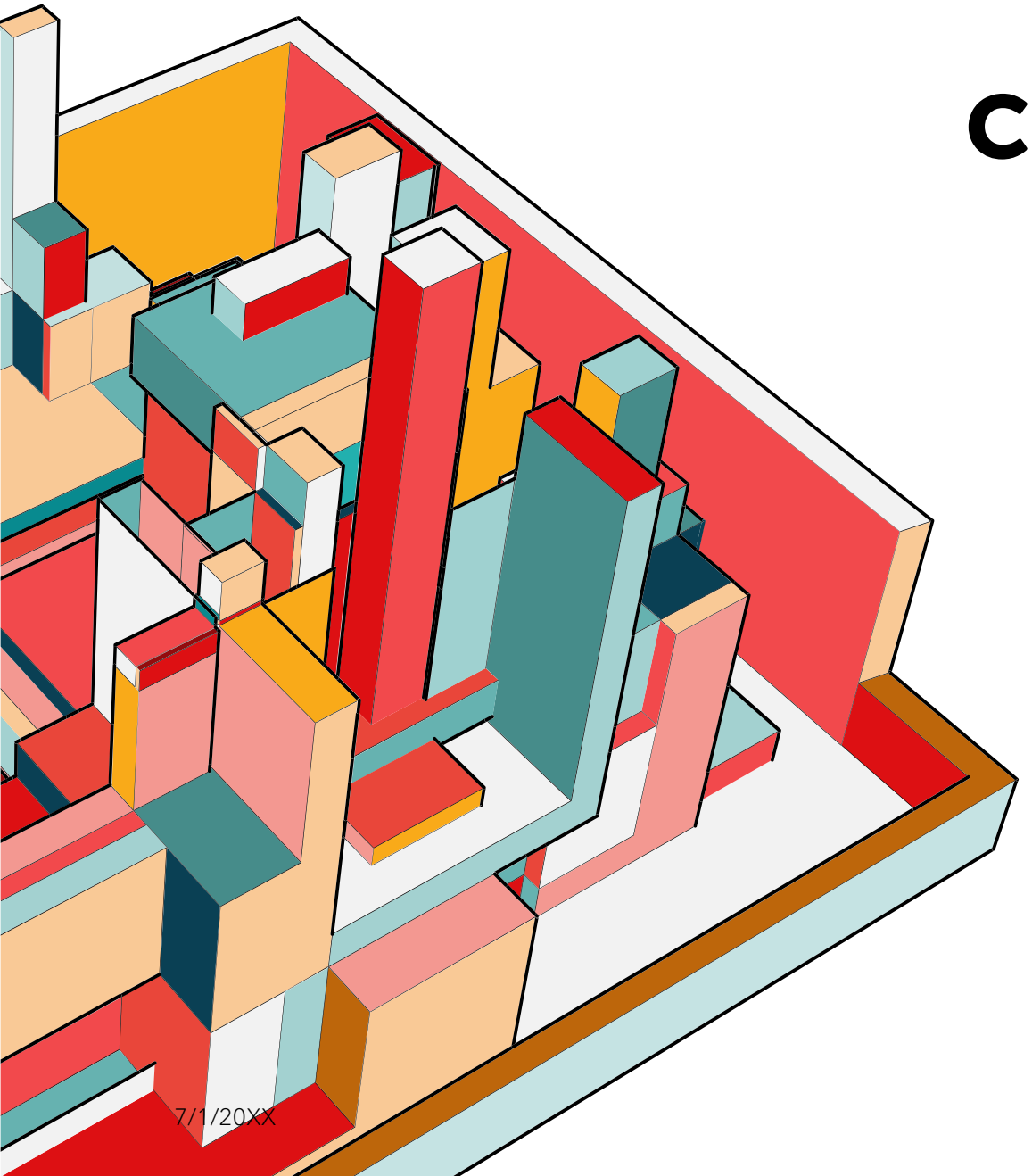
	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.882353	0.819444
F1_Score	0.909091	0.916031	0.937500	0.900763
Accuracy	0.866667	0.877778	0.911111	0.855556

RESULTS-CONFUSION MATRIX

- You are able to determine that logistic regression can discriminate between the various groups by looking at the confusion matrix. You can observe that false positives are the main issue.





CONCLUSION

- ✓ The most effective algorithm for this dataset is the decision tree model.
- ✓ Lower payload mass launches perform better than higher payload mass launches.
- ✓ The majority of launch sites are close to the equator, and every site is in close proximity to the coast.
- ✓ Over time, launches have a higher success percentage.Of all the launch locations, KSC LC-39A has the best success rate.
- ✓ The success rate for the ES-L1, GEO, HEO, and SSO orbits is 100%.

THANK YOU