Data Science, spaceY capstone project first stage reuse.

Let's Started

Author : Audrey Naila Putri

PRESENTATION CONTENT

P3. Executive Summary

P4. Introduction

P6. Methodology

P16 Results

P17. EDA with Visualization

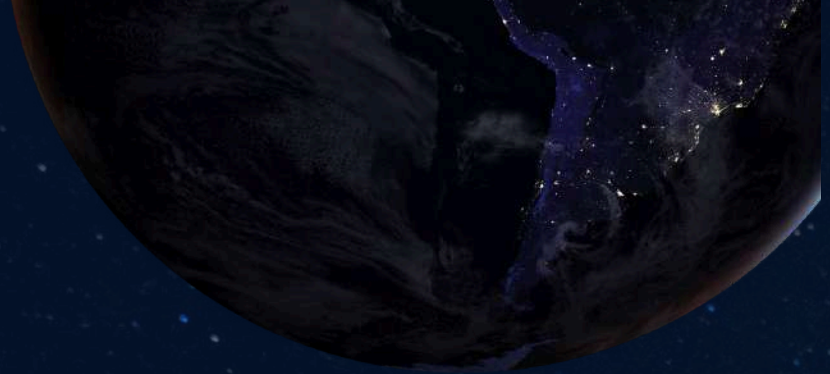
P23. EDA with SQL

P30. Interactive Maps with Folium

P35. Plotly Dash Dashboard

P39 Predictive Analytics

P41 Conclusion



EXECUTIVE SUMMARY

The research attempts to identify the factors for a successful rocket landing. To make this determination, the following methodologies where used:

- Collect data using SpaceX REST API and web scraping techniques
- Wrangle data to create success/fail outcome variable
- Explore data with data visualization techniques, considering the following factors:
- payload, launch site, flight number and yearly trend
- Analyze the data with SQL, calculating the following statistics: total payload, payload
- range for successful launches, and total # of successful and failed outcomes
- Explore launch site success rates and proximity to geographical markers
- Visualize the launch sites with the most success and successful payload ranges
- Build Models to predict landing outcomes using logistic regression, support vector
- machine (SVM), decision tree and K-nearest neighbor (KNN)

RESULTS

EDA :

- Launch success has improved over time
- KSC LC-39A has the highest success rate
- among landing sites
- Orbits ES-L1, GEO, HEO, and SSO have
- a 100% success rate

VISUALIZATION/ ANALYTICS:

- **Most launch sites are near the equator, and all are close to the coast**

INTRODUCTION

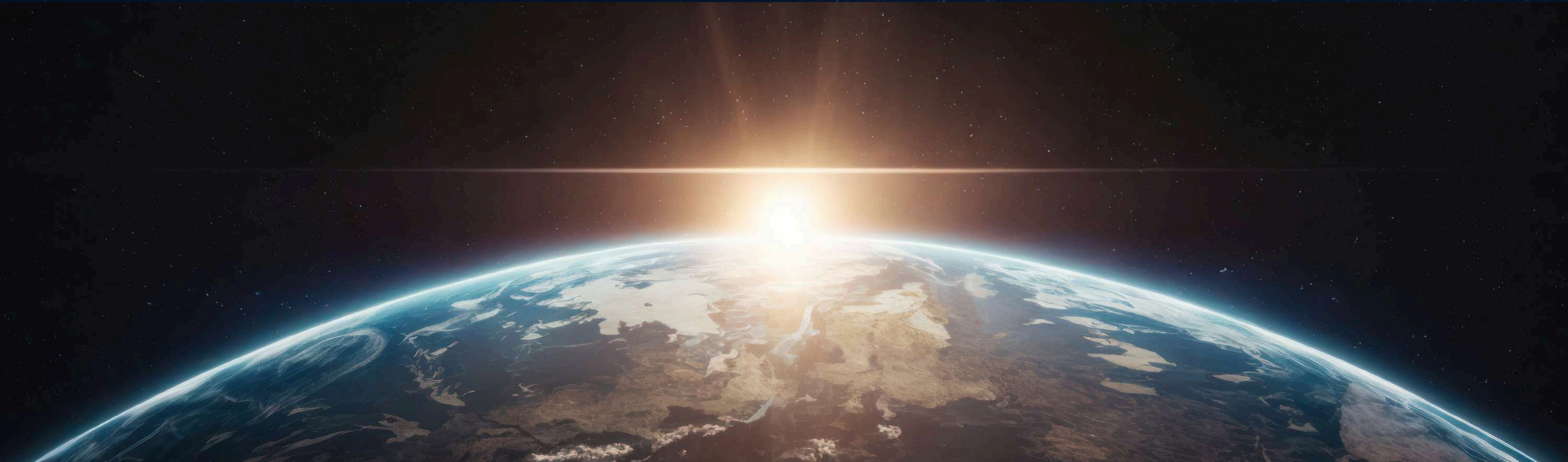
BACKGROUND :

SpaceX, a leader in the space industry, strives to make space travel affordable for everyone. Its accomplishments include sending spacecraft to the international space station, launching a satellite constellation that provides internet access and sending manned missions to space. SpaceX can do this because the rocket launches are relatively inexpensive ($62 million per launch) due to its novel reuse of the first stage of its Falcon 9 rocket. Other providers, which are not able to reuse the first stage, cost upwards of $165 million each. By determining if the first stage will land, we can determine the price of the launch. To do this, we can use public data and machine learning models to predict whether SpaceX – or a competing company – can reuse the first stage.

EXPLORE :

- Rate of successful landings overtime
- Best predictive model for successful landing (binaryclassification)
- How payload, launch site, number of flights, and orbits affect first stage landing succes

METHODOLOGY



METHODOLOGY

STEPS :

- Collect data using SpaceX REST API and web scraping techniques
- Wrangle data – by filtering the data, handling missing values and applying
- one hot encoding – to prepare the data for analysis and modeling
- Explore data via EDA with SQL and data visualization techniques
- VisualizethedatausingFoliumandPlotlyDash
- Build Models to predict landing outcomes using classification models. Tune and evaluate models to find best model and parameters

DATA COLLECTION

-API

STEPS :

- Request data from SpaceX API (rocket launch data)
- Decode response using .json() and convert to a dataframe using .json_normalize()
- Request information about the launches from SpaceX API using custom functions
- Create dictionary from the data
- Create data frame from the dictionary
- Filter dataframe to contain only Falcon 9 launches
- Replace missing values of Payload Mass with calculated .mean()
- Exportdatatocsvfile

DATA WRANGLING

-API

STEPS :

- Perform EDA and determine data labels
- Calculate:
  - # of launches for each site
  - # and occurrence of orbit
  - # and occurrence of mission
  - outcome per orbit type]
- Createbinarylandingoutcome
- column (dependent variable)
- Exportdatatocsvfile

LANDING OUTCOME :

- Landing was not always successful
- TrueOcean : mission out come had a successful landing to a specific region of the ocean

LANDING OUTCOME CONT

- False Ocean: represented an unsuccessful landing to a specific region of ocean
- TrueRTLS: meantthe mission had a successful landing on a ground pad
- False RTLS: represented an unsuccessful landing on a ground pad
- TrueASDS: meant the mission outcome had a successful landing on a drone ship
- FalseASDS: representedan unsuccessful landing on drone ship
- Outcomes converted into 1 for
- a successful landing and 0 for an
- unsuccessful landing

EDA WITH SQL

QUERIES :

DISPLAY :

- Names of unique launch sites
- 5 records where launch site begins with ‘CCA’
- Total payload mass carried by boosters launched by NASA (CRS)
- AveragepayloadmasscarriedbyboosterversionF9v1.1.

LIST :

- Date of first successful landing on ground pad
- Names of boosters which had success landing on drone ship and have payload mass greater than 4,000 but less than 6,000
- Total number of successful and failed missions
- Names of booster versions which have carried the max payload
- Failed landing out come son droneship, their booster version and launch site for the months in the year 2015
- Count of landing out comes between 2010-06-04 and 2017-03-20(desc) 2023 11

MAP WITH FOLIUM

MARKERS INDICATING LAUNCH SITES

- Added blue circle at NASA Johnson Space Center's coordinate with a popup label showing its name using its latitude and longitude coordinates
- Added red circles at all launch sites coordinates with a popup label showing its name using its name using its latitude and longitude coordinates

COLORED MARKERS OF LAUNCH OUTCOMES

- Added colored markers of successful (green) and unsuccessful (red) launches at each launch site to show which launch sites have high success rates

DISTANCES BETWEEN A LAUNCH SITE TO PROXIMITIES

- Added colored lines to show distance between launch site CCAFS SLC-40 and its proximity to the nearest coastline, railway, highway, and city

DASHBOARD

WITH PLOTLY DASH

Dropdown List with Launch Sites

• Allowusertoselectalllaunchsitesoracertainlaunchsite

Pie Chart Showing Successful Launches

• Allow user to see successful and unsuccessful launches as a percent of the total Slider of Payload Mass Range

• **Allowusertoselectpayloadmassrange**

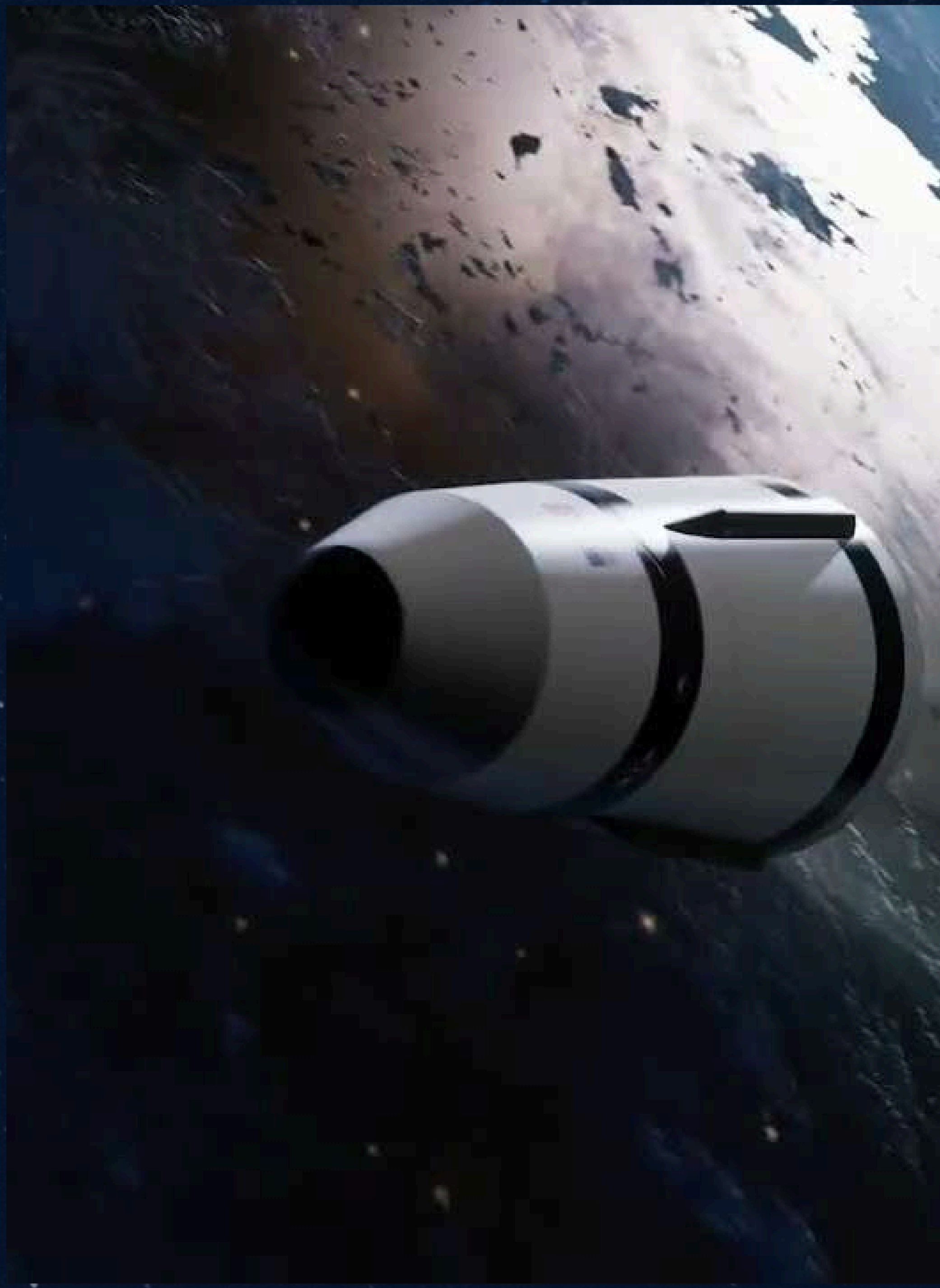
Scatter Chart Showing Payload Mass vs. Success Rate by **Booster Version**

• Allow user to see the correlation between Payload and Launch Success

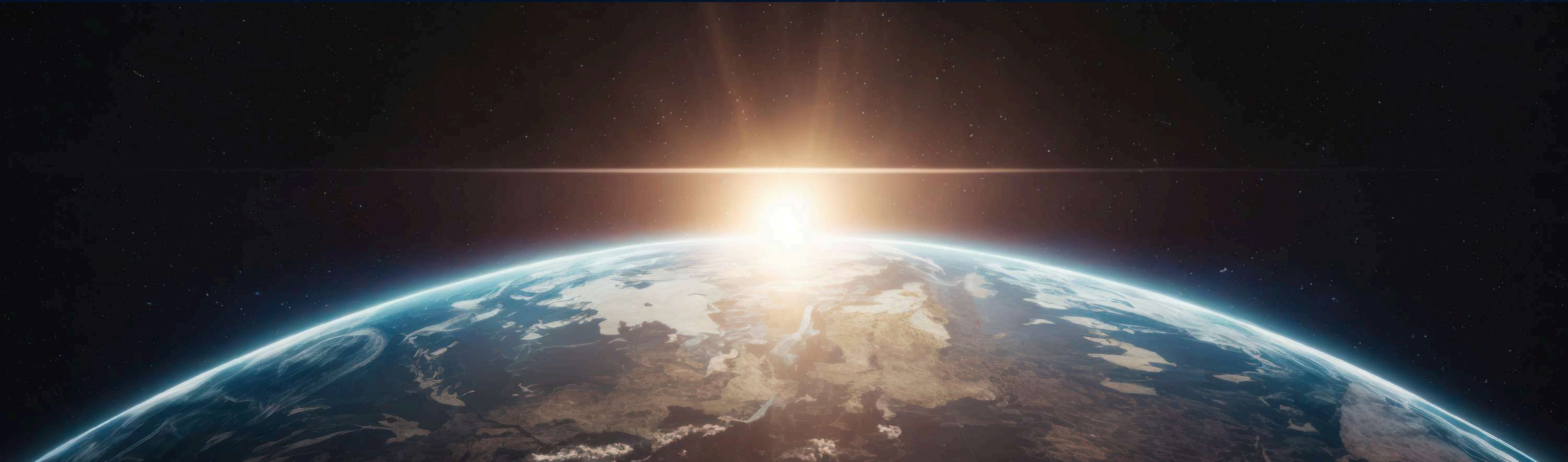
PREDICTIVE ANALYTICS

CHARTS :

- Create NumPy array from the Class column
- Standardize the data with StandardScaler.Fit and trans form the data.
- Split the data using train_test_split
- Createa Grid Search CV object with cv = 10 for parameter optimization
- Apply GridSearchCV on different algorithms: logistic regression (LogisticRegression()), support vector machine (SVC()), decision tree (DecisionTreeClassifier()), K-Nearest Neighbor (KNeighborsClassifier())
- Calculate accuracy on the test data using .score() for all models
- Assess the confusion matrix for all models
- Identify the best model using Jaccard_Score, F1_ScoreandAccuracy



RESULTS



RESULTS SUMMARY

EXPLORATORY DATA ANALYSIS

- Launch success has improved over time
- KSC LC-39A has the highest success rate among landing sites
- Orbits ES-L1, GEO, HEO and SSO have a 100% success rate

VISUAL ANALYTICS

- Most launch sites are near the equator, and all are close to the coast
- Launch sites are far enough away from anything a failed launch can damage (city, highway, railway), while still close enough to bring people and material to support launch activities

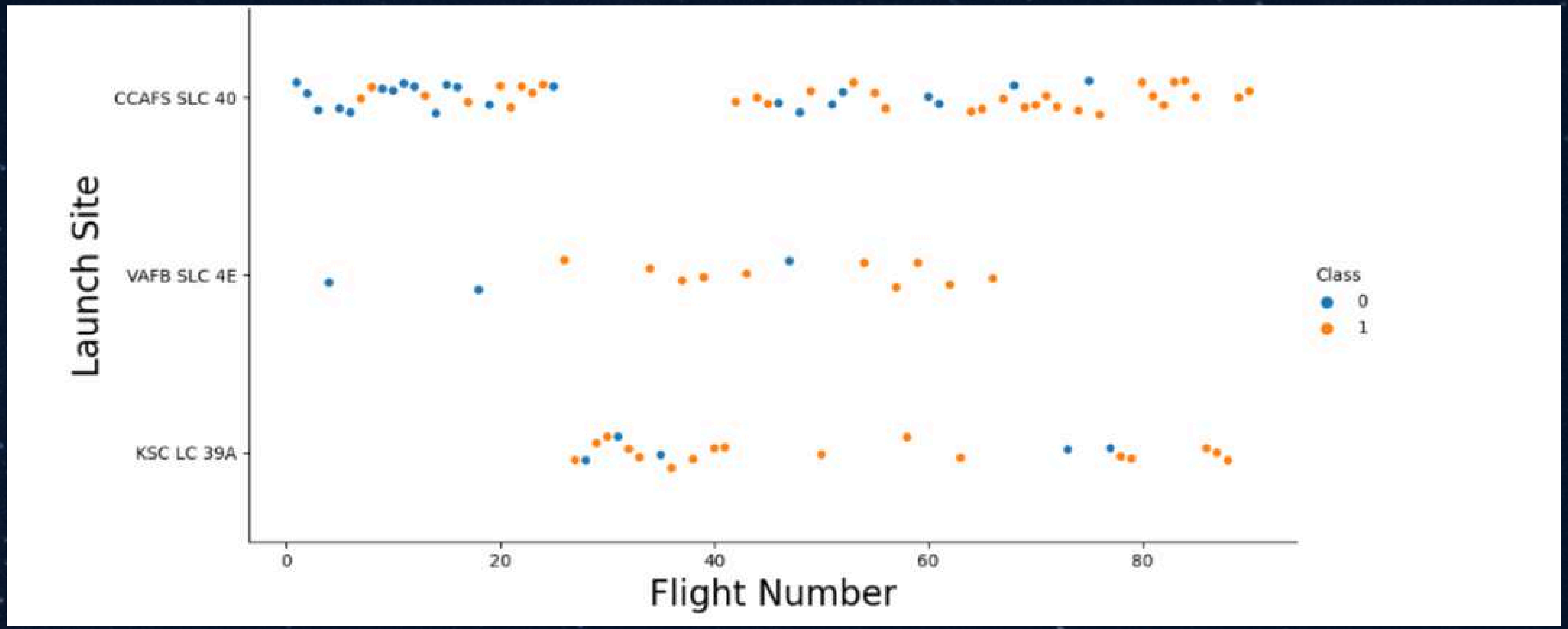
PREDICTIVE ANALYTICS

- Decision Tree mode lis the best predictive model for the dataset

FLIGHT NUMBER VS. LAUNCH SITE

EXPLORATORY DATA ANALYSIS

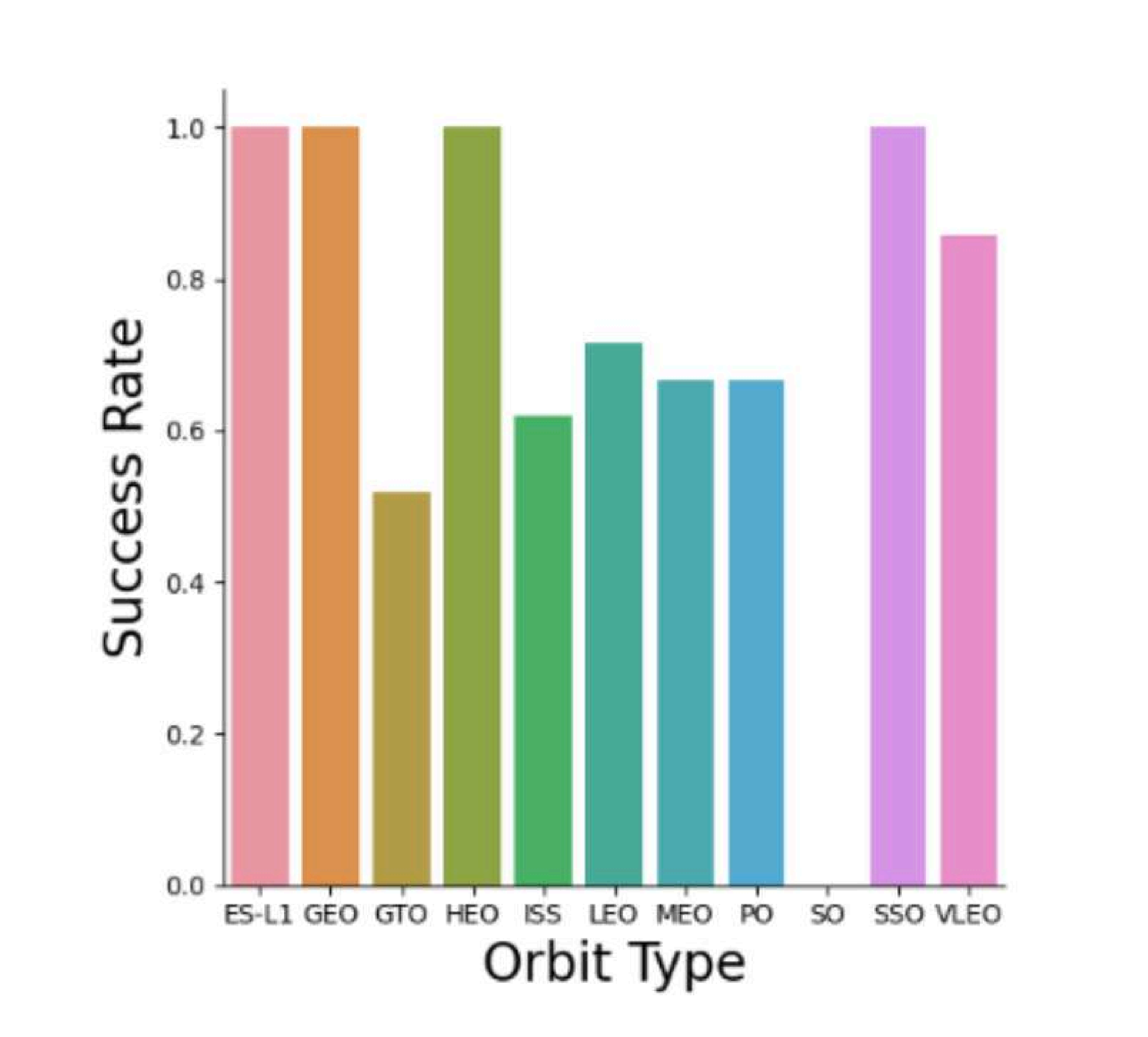
- Earlier flight shadalower success rate(blue=fail)
- Later flights had a higher success rate (orange = success)
- Around half of launches were from CCAFS SLC 40 launch site
- VAFB SLC 4E and KSC LC 39A have higher success rates
- We can infer that new launches have a higher success rate



SUCCESS RATE
BY ORBIT

EXPLORATORY DATA ANALYSIS

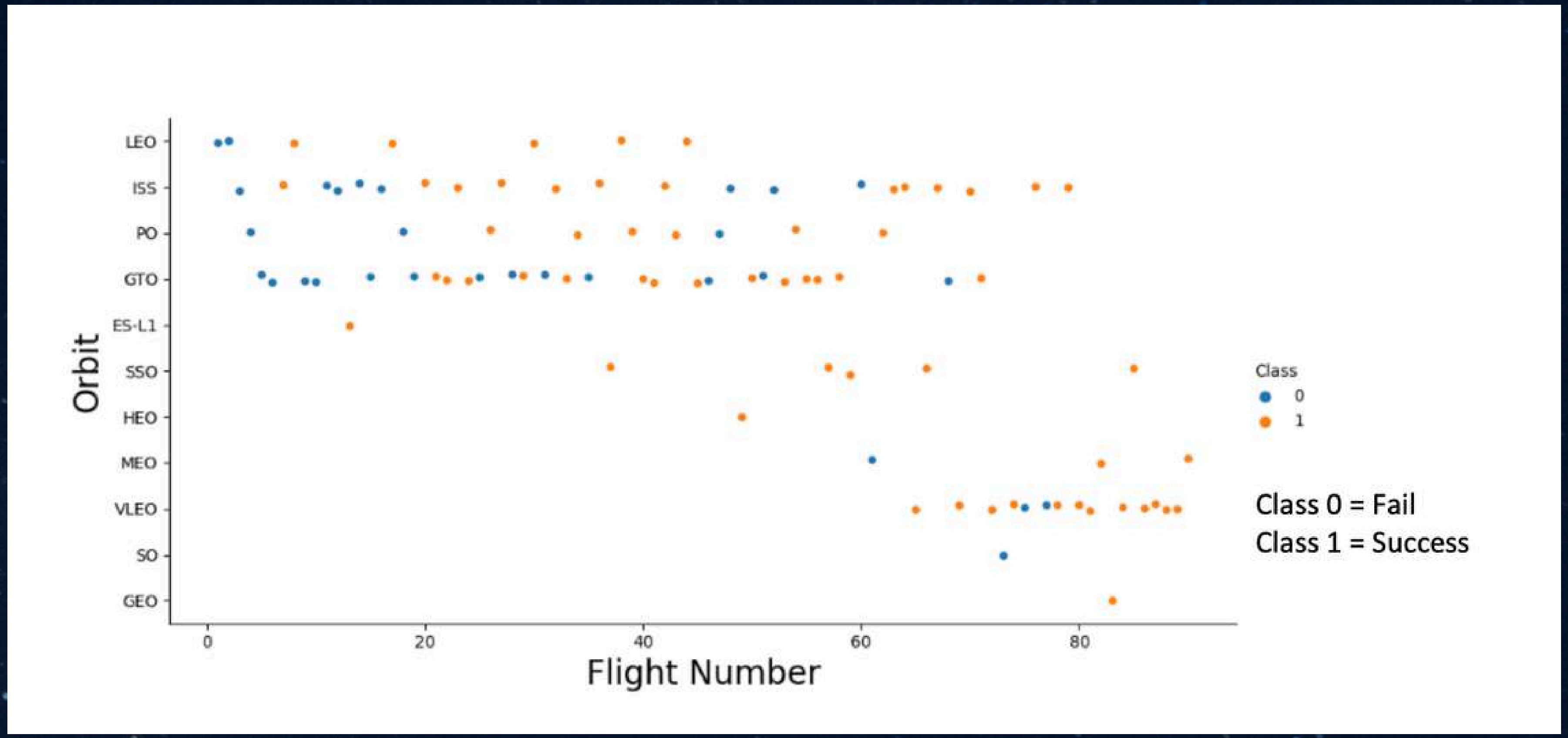
- **100% Success Rate** : ES-L1, GEO, HEO and SSO
- **50%-80% Success Rate:** GTO,I SS, LEO, MEO, PO
- **0% Success Rate**: SO



FLIGHT NUMBER VS. ORBIT

EXPLORATORY DATA ANALYSIS

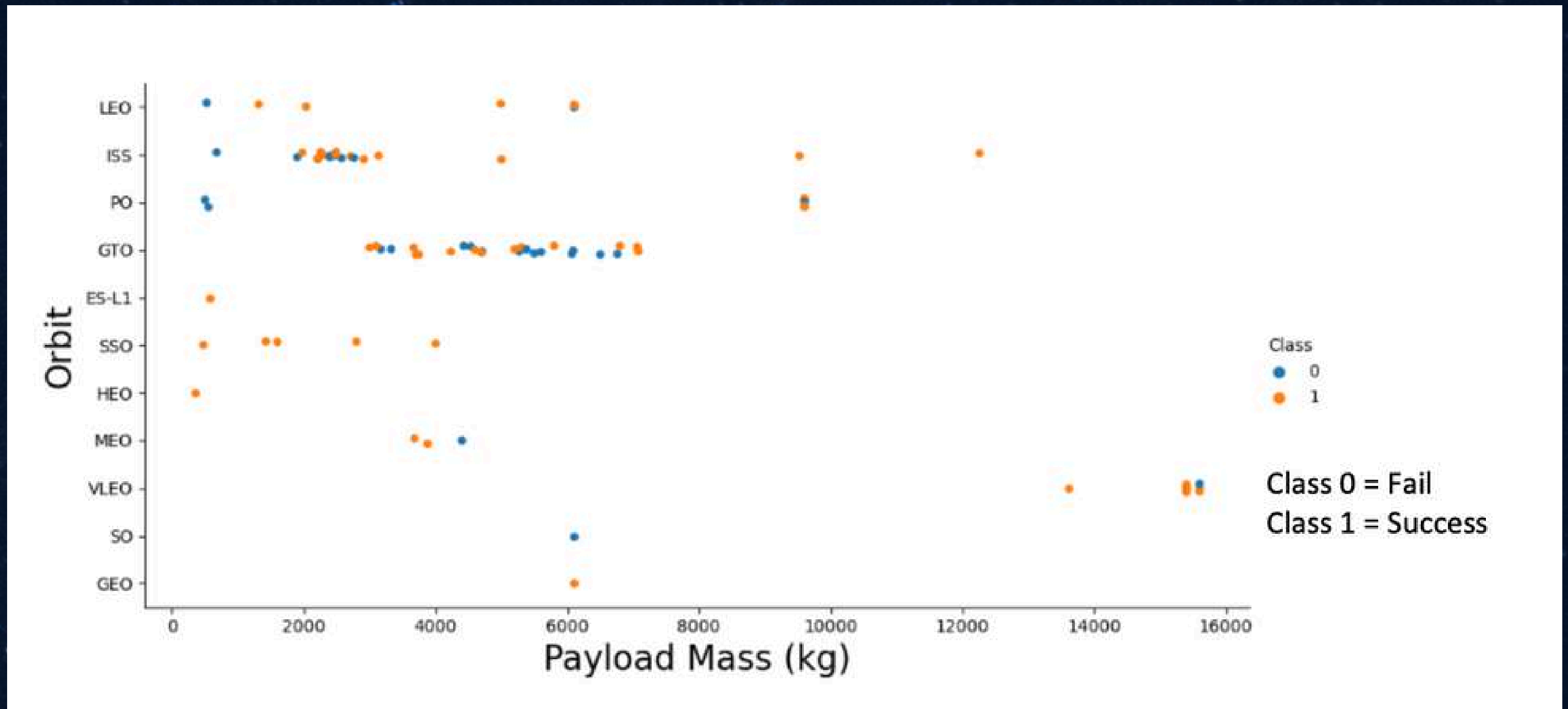
- The success rate typically increases with the number of flights for each orbit
- This relationship is highly apparent for the LEO orbit
- The GTO orbit, however, does not follow this trend



PAYLOAD VS. ORBIT

EXPLORATORY DATA ANALYSIS

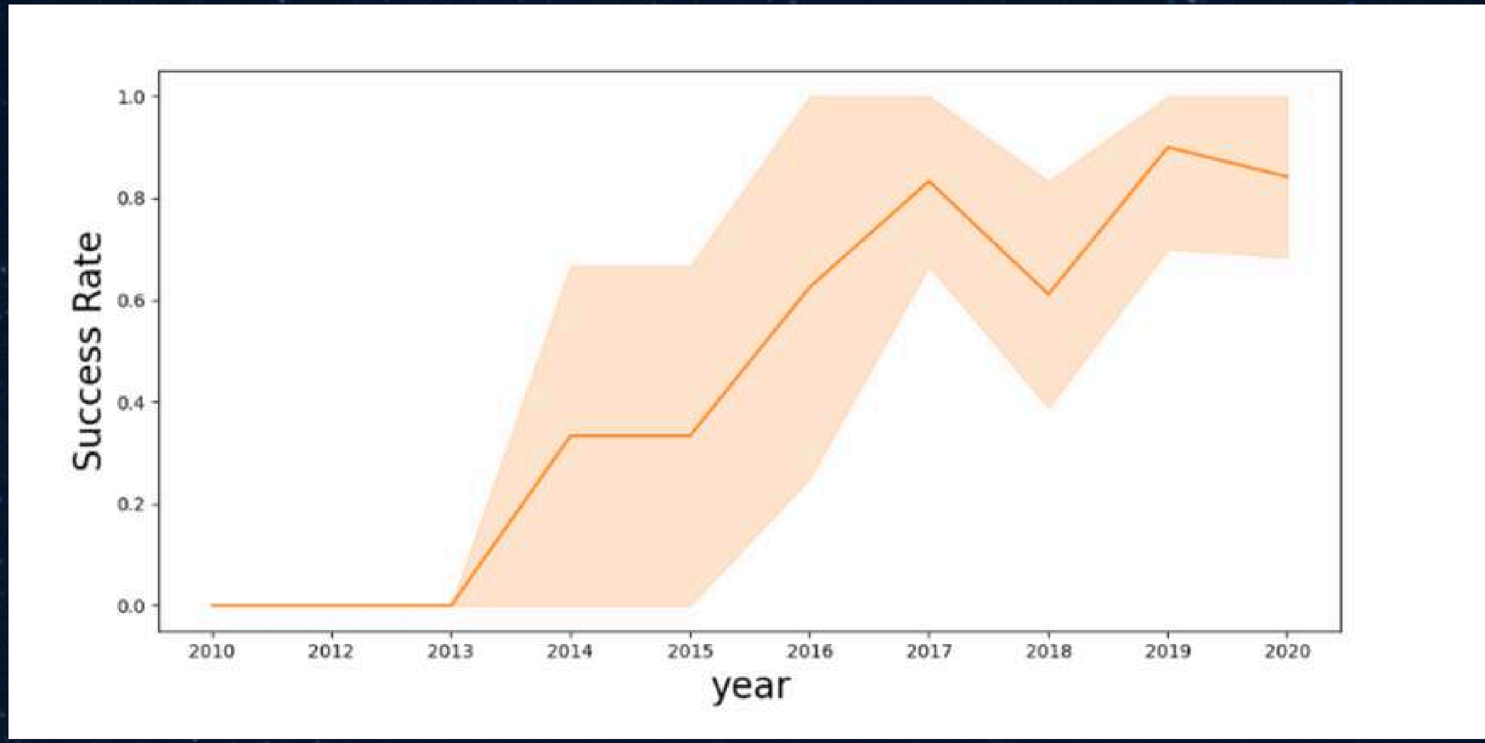
- Heavy payloads are better with LEO, ISS and PO orbits
- The GTO orbit has mixed success with heavier payloads



LAUNCH SUCCESS OVER TIME

EXPLORATORY DATA ANALYSIS

- The success rate improved from2013-2017 and 2018-2019
- The success rate decreased from 2017-2018 and from2019-2020
- Overall, the success rate has improved since
- 2013



LAUNCH SITE INFORMATION

LAUNCH SITE NAMES

• CCAFSLC-40 • CCAFSSLC-40

• KSCLC-39A. • VAFBSLC-4E

LANDING OUTCOME CONT

```
[30]: %sql ibm_db_sa://yyy33800:dwNKg8J3L0IBd6CP@1bbf73c5
      %sql SELECT Unique(LAUNCH_SITE) FROM SPACEXTBL;

 * ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4bb0-85b9
   sqlite:///my_data1.db
Done.
```

[30]:

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

RECORDS WITH LAUNCH SITE STARTING WITH CCA

Displaying 5 records below

```
%sql SELECT * \
    FROM SPACEXTBL \
    WHERE LAUNCH_SITE LIKE'CCA%' LIMIT 5;

 * ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
   sqlite:///my_data1.db
Done.
```

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

LANDING & MISSION INFO

1ST SUCCESSFUL LANDING IN GROUND PAD

- 12/22/2015

BOOSTER DRONE SHIP LANDING

- Booster mass greater than 4,000 but less than 6,000
- JSCAT-14,JSCAT-16,SES-10, SES-11 / EchoStar 105

TOTAL NUMBER OF SUCCESSFUL AND FAILED MISSION OUTCOMES

- 1 Failure in Flight
- 99 Success
- 1 Success (payload status unclear)

PAYLOAD MASS

TOTAL PAYLOAD MASS

• **45,596 kg (**total) carried by boosters launched by NASA (CRS)

TOTAL PAYLOAD MASS

• **2,928 kg** (average) carried by booster version F9 v1.1

BOOSTERS

CARRYING MAX PAYLOAD

- F9 B5 B1048.4
- F9 B5 B1049.4
- F9 B5 B1051.3
- F9 B5 B1056.4
- F9 B5 B1048.5
- F9 B5 B1051.4
- F9 B5 B1049.5
- F9 B5 B1060.2
- F9 B5 B1058.3
- F9 B5 B1051.6
- F9 B5 B1060.3
- F9 B5 B1049.7

```
%sql SELECT BOOSTER_VERSION \
FROM SPACEXTBL \
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

```
 * sqlite:///my_data1.db
Done.
```

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

FAILED LANDINGS ON DRONE SHIP

IN 2015

Showing month, date, booster version, launch site and landing outcome

```
%sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing _Outcome] \
FROM SPACEXTBL \
where [Landing _Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='2015';
```

* sqlite:///my_data1.db
Done.

month	Date	Booster_Version	Launch_Site	Landing _Outcome
01	10-01-2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	14-04-2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

COUNT OF SUCCESSFUL LANDINGS

RANKED DESCENDING

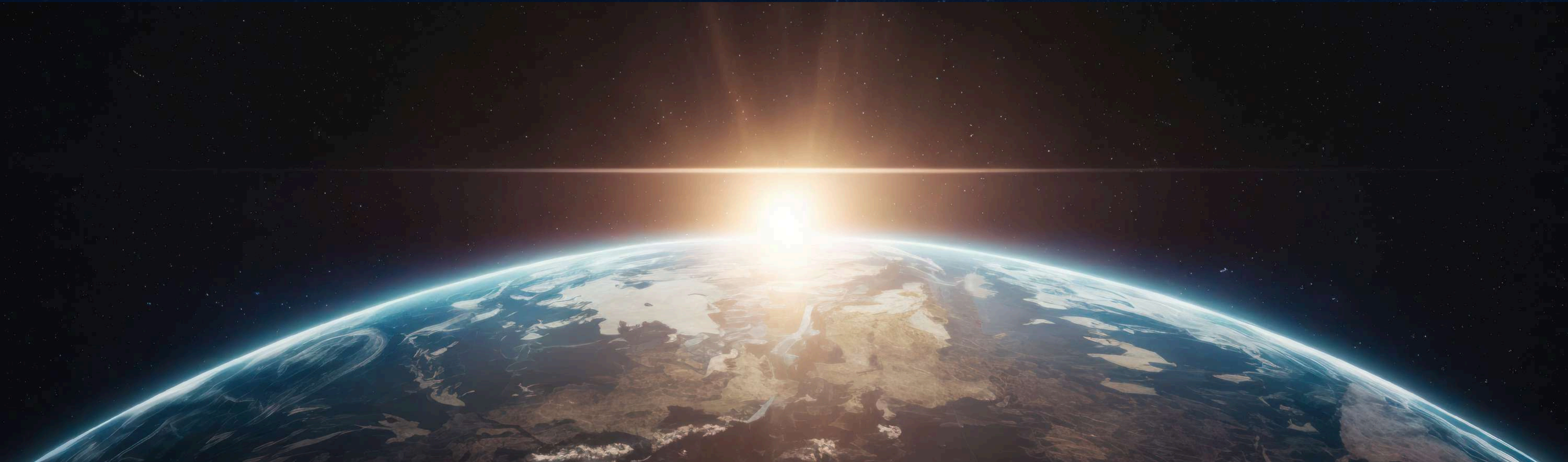
- Count of landing outcomes between 2010-06-04 and 2017-03-20 in descending order

```
%sql SELECT [Landing _Outcome], count(*) as count_outcomes \
FROM SPACEXTBL \
WHERE DATE between '04-06-2010' and '20-03-2017' group by [Landing _Outcome] order by count_outcomes DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

Landing _Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

LAUNCH SITES



LAUNCH SITES

Near Equator: the closer the launch site to the equator, the **easier** it is **to launch** to equatorial orbit, and the more help you get from Earth's rotation for a prograde orbit. Rockets launched from sites near the equator get an additional natural boost - due to the rotational speed of earth - that helps save the cost of putting in extra fuel and boosters.

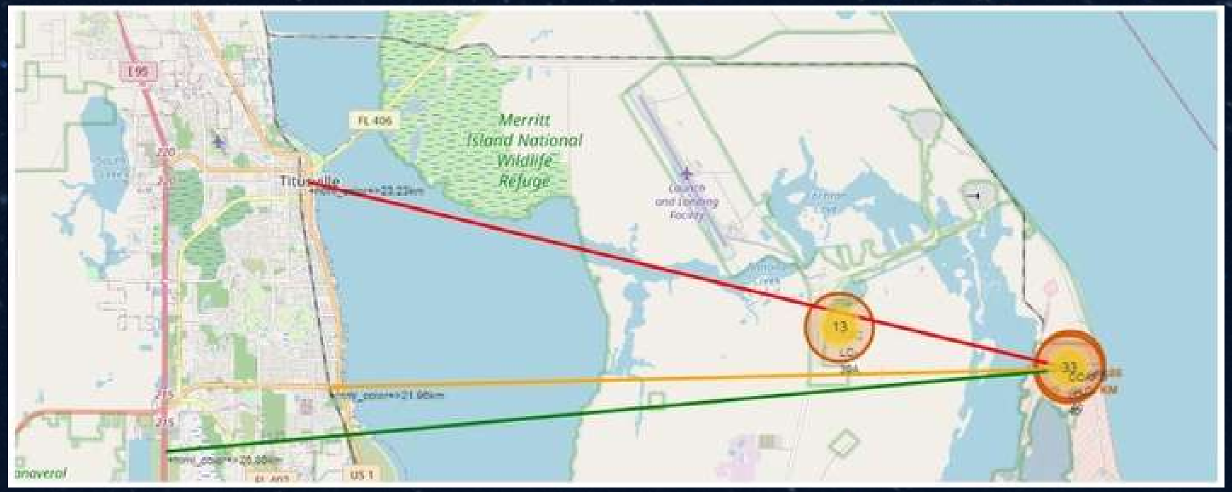
LAUNCH OUTCOMES

- Outcomes:
- Green markers for success ful launches
- Red markers for unsuccessful launches
- Launchsite **CCAFSSLC-40** hasa **3/7 success rate(42.9%)**

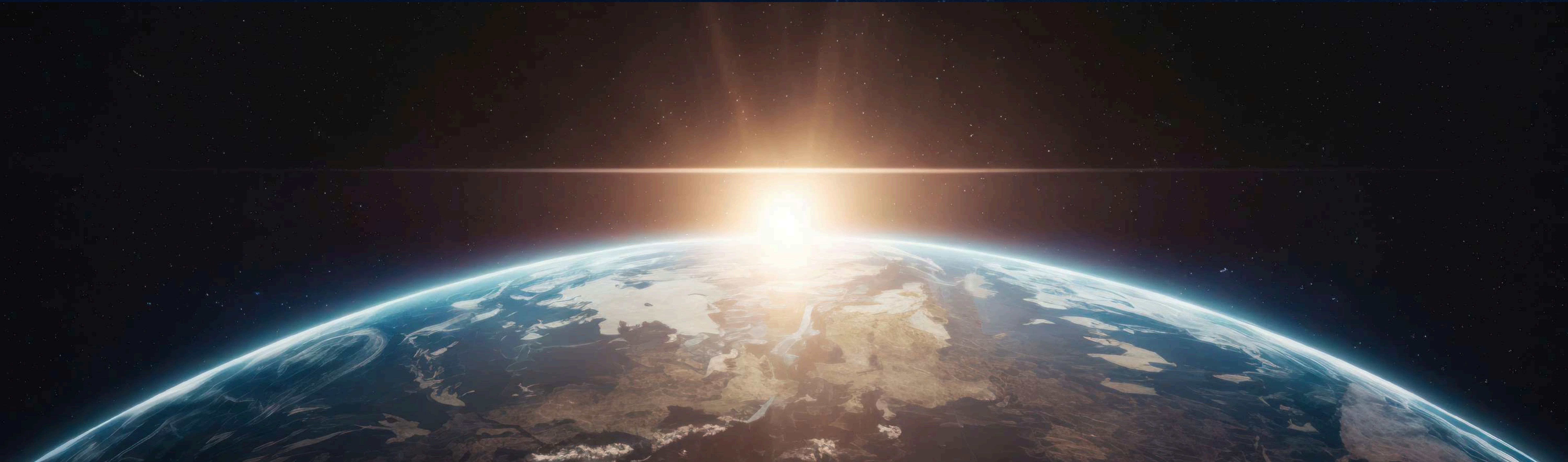
DISTANCE TO PROXIMITIES

- Coasts: help ensure that spent stages dropped along the launch path or failed launches don't fall on people or property.
- Safety / Security: needs to be an exclusion zone around the launch site to keep unauthorized people away and keep people safe.
- Transportation/InfrastructureandCities:needtobeawayfromanythinga failed launch can damage, but still close enough to roads/rails/docks to be able to bring people and material to or from it in support of launch activities.

- 86 km from nearest coastline
- 21.96 km from nearest railway
- 23.23 km from nearest city
- 26.88 km rom nearest highway



DASHBOARD WITH PLOTY



LAUNCH SUCCESS

BY SITE

- KSC LC-39A has the most successful launches amongst launch sites (41.2%)

(KSC LC-29A

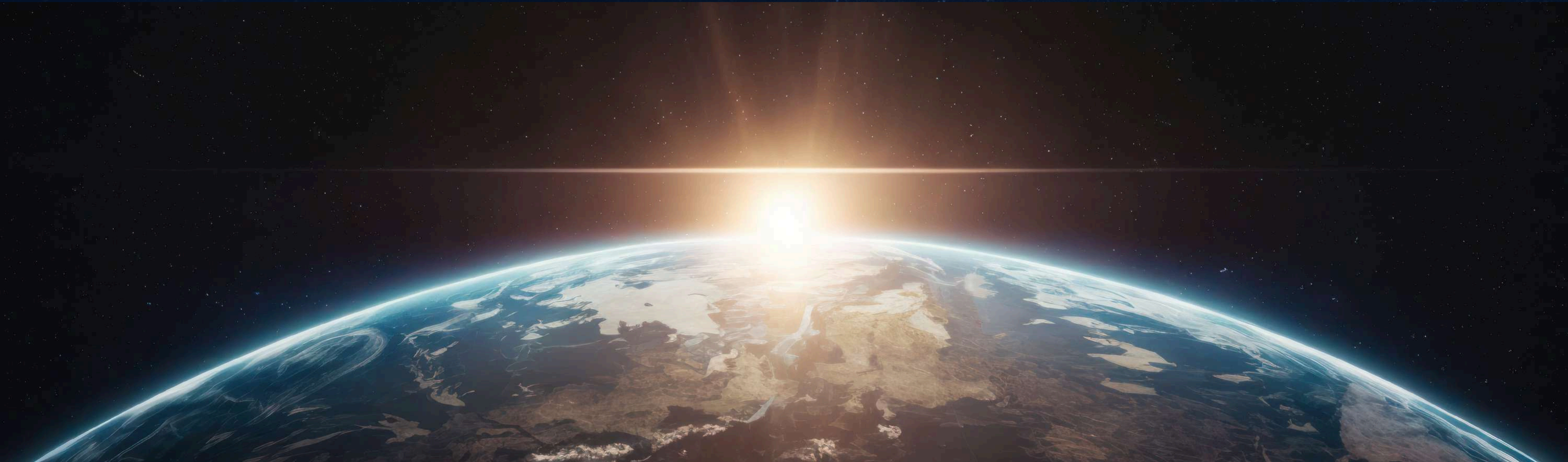
- KSC LC-39A has the highest success rate amongst launch sites (76.9%)
- 10 successful launches and 3 failed launches

BY SITE

(KSC LC-29A

MASS AND SUCCESS

PREDICTIVE ANALYSIS



CLASSIFICATION

ACCURACY

- All the models performed at about the same level and had the same scores and accuracy. This is likely due to the small dataset. The Decision Tree model slightly outperformed the rest when looking at .best_score_
- .best_score_ is the average of all cv folds for a single combination of the parameters

	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

CONFUSION MATRICES

PERFORMANCE SUMMARY

- A confusion matrix summarizes the performance of a classification algorithm
- Alltheconfusionmatricesweresidentical
- The fact that there are false positives (Type 1 error) is not good

- ConfusionMatrixOutputs:
- 12 True positive
- 3 True negative
- 3 False positive
- 0 False Negative

CONCLUSION

RESEARCH

- Model Performance: The models performed similarly on the test set with the decision tree model slightly outperforming
- Equator: Most of the launch sites are near the equator for an additional natural boost - due to the rotational speed of earth - which helps save the cost of putting in extra fuel and boosters
- Coast: All the launch sites are close to the coast
- Launch Success: Increases over time
- KSC LC-39A: Has the highest success rate among launch sites. Has a 100% success rate for launches less than 5,500 kg
- Orbits :ESL1, GEO, HEO, and SSO havea 100% success rate
- Payload Mass: Acrossall launch sites, the higher the pay load mass(kg), the
- higher the success rate

END