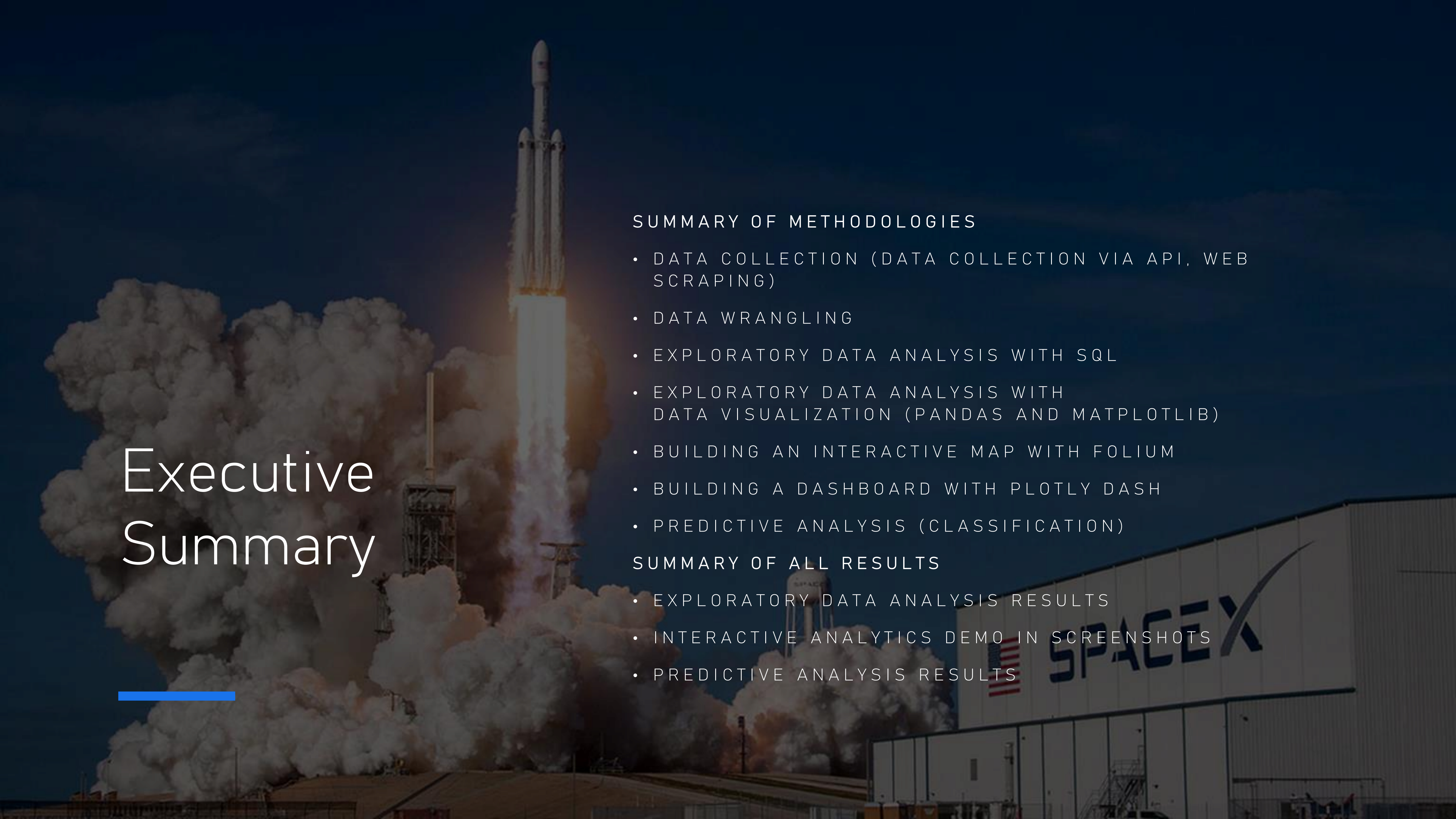
Data Science Capstone Project

HUGO SÁNCHEZ TTITO

21.01.2024

Outline

- EXECUTIVE SUMMARY
- INTRODUCTION
- METHODOLOGY
- RESULTS
- CONCLUSION
- APPENDIX



Executive Summary

SUMMARY OF METHODOLOGIES

- DATA COLLECTION (DATA COLLECTION VIA API, WEB SCRAPING)
- DATA WRANGLING
- EXPLORATORY DATA ANALYSIS WITH SQL
- EXPLORATORY DATA ANALYSIS WITH DATA VISUALIZATION (PANDAS AND MATPLOTLIB)
- BUILDING AN INTERACTIVE MAP WITH FOLIUM
- BUILDING A DASHBOARD WITH PLOTLY DASH
- PREDICTIVE ANALYSIS (CLASSIFICATION)

SUMMARY OF ALL RESULTS

- EXPLORATORY DATA ANALYSIS RESULTS
- INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS
- PREDICTIVE ANALYSIS RESULTS

Introduction

PROJECT BACKGROUND AND CONTEXT: THIS PROJECT AIMS TO PREDICT THE SUCCESSFUL LANDING OF THE FALCON 9 FIRST STAGE, A CRITICAL ASPECT CONSIDERING SPACEX'S CLAIM ON THEIR WEBSITE THAT THE COST OF THEIR FALCON 9 ROCKET LAUNCH IS 62 MILLION DOLLARS, SIGNIFICANTLY LOWER THAN THE COSTS OF OTHER PROVIDERS, MAINLY BECAUSE OF SPACEX'S ABILITY TO REUSE THE FIRST STAGE. THE PREDICTION OF THE LANDING OUTCOME IS CRUCIAL NOT JUST FOR COST ESTIMATION BUT ALSO FOR POTENTIAL COMPETITORS SEEKING INSIGHTS INTO SPACEX'S COST-EFFECTIVE APPROACH TO ROCKET LAUNCHES.

PROBLEMS SEEKING SOLUTIONS

- WHAT CHARACTERISTICS DISTINGUISH A LANDING AS SUCCESSFUL OR FAILED?
- HOW DO THE RELATIONSHIPS AMONG VARIOUS ROCKET VARIABLES INFLUENCE THE SUCCESS OR FAILURE OF A LANDING?
- WHAT CONDITIONS WILL ENABLE SPACEX TO ACHIEVE THE HIGHEST LANDING SUCCESS RATE?

Methodology

DATA COLLECTION METHODOLOGY:

- USING SPACEX REST API
- USING WEB SCRAPPING FROM WIKIPEDIA

PERFORMED DATA WRANGLING

- FILTERING THE DATA
- DEALING WITH MISSING VALUES
- USING ONE HOT ENCODING FOR CLASSIFICATION MODELS

PERFORMED EXPLORATORY DATA ANALYSIS (EDA) USING VISUALIZATION AND SQL

PERFORMED INTERACTIVE VISUAL ANALYTICS USING FOLIUM AND PLOTLY DASH

PERFORMED PREDICTIVE ANALYSIS USING CLASSIFICATION MODELS

- BUILDING, TUNING AND EVALUATION OF CLASSIFICATION MODELS

Methodology – Data collection

THE DATA COLLECTION PROCESS UTILIZED A DUAL APPROACH, INCORPORATING API REQUESTS FROM SPACEX REST API AND EMPLOYING WEB SCRAPING TECHNIQUES TO EXTRACT INFORMATION FROM A TABLE WITHIN SPACEX'S WIKIPEDIA ENTRY. THIS COMBINED METHODOLOGY WAS NECESSARY TO ENSURE COMPREHENSIVE DATA GATHERING FOR A MORE THOROUGH ANALYSIS OF LAUNCHES.

THE DATA COLUMNS OBTAINED THROUGH THE SPACEX REST API INCLUDE: FLIGHTNUMBER, DATE, BOOSTERVERSION, PAYLOADMASS, ORBIT, LAUNCHSITE, OUTCOME, FLIGHTS, GRIDFINS, REUSED, LEGS, LANDINGPAD, BLOCK, REUSEDCOUNT, SERIAL, LONGITUDE, AND LATITUDE.

IN CONTRAST, THE DATA COLUMNS ACQUIRED THROUGH WIKIPEDIA WEB SCRAPING ENCOMPASS: FLIGHT NO., LAUNCH SITE, PAYLOAD, PAYLOADMASS, ORBIT, CUSTOMER, LAUNCH OUTCOME, VERSION BOOSTER, BOOSTER LANDING, DATE, AND TIME.

THIS DUAL-SOURCED DATA COMPILATION ENSURES A COMPREHENSIVE DATASET FOR A DETAILED AND COMPREHENSIVE ANALYTICAL EXAMINATION.

Methodology – Data collection via API

1. Requesting information about rocket launches by accessing SpaceX's API.
2. Parsing the response content by utilizing the .json() function and converting it into a dataframe through the application of .json_normalize().
3. Requesting the required details regarding SpaceX launches by employing customized functions to interact with the SpaceX API.
4. Organizing the acquired data into a dictionary structure.
5. Generating a dataframe using the information stored in the dictionary.
6. Refining the dataframe to exclusively incorporate Falcon 9 launches.
7. Substituting the missing values in the Payload Mass column with the calculated mean for that specific column.
8. Saving the data to a CSV file.

Methodology – Data collection via Web scraping

1. Retrieving Falcon 9 launch information from Wikipedia
2. Initializing a BeautifulSoup object from the HTML response.
3. Extracting column names from the HTML table header.
4. Collecting data by parsing HTML tables.
5. Organizing the acquired data into a dictionary.
6. Creating a dataframe from the dictionary.
7. Exporting the data to a CSV file.

Methodology - Data wrangling

IN THE DATASET, UNSUCCESSFUL BOOSTER LANDINGS ARE CATEGORIZED BASED ON OUTCOMES LIKE TRUE OCEAN (SUCCESSFUL OCEAN LANDING), FALSE OCEAN (UNSUCCESSFUL OCEAN LANDING), TRUE RTLS (SUCCESSFUL GROUND PAD LANDING), FALSE RTLS (UNSUCCESSFUL GROUND PAD LANDING), TRUE ASDS (SUCCESSFUL DRONE SHIP LANDING), AND FALSE ASDS (UNSUCCESSFUL DRONE SHIP LANDING). FOR SIMPLICITY, THESE OUTCOMES ARE CONVERTED INTO TRAINING LABELS: "1" FOR SUCCESSFUL LANDINGS AND "0" FOR UNSUCCESSFUL ONES.

Conduct exploratory data analysis to identify and define training labels.

Determine the count of launches for each launch site.

Calculate the frequency and occurrence of each orbit in the dataset.

Calculate the number and occurrence of mission outcomes for each orbit type.

Generate a landing outcome label based on the information in the Outcome column.

Export the processed data to a CSV file.

Methodology – EDA with data visualization

VARIOUS CHARTS WERE GENERATED TO VISUALLY REPRESENT DIFFERENT ASPECTS OF THE DATASET:

- FLIGHT NUMBER VS. PAYLOAD MASS
- FLIGHT NUMBER VS. LAUNCH SITE
- PAYLOAD MASS VS. LAUNCH SITE
- ORBIT TYPE VS. SUCCESS RATE
- FLIGHT NUMBER VS. ORBIT TYPE
- PAYLOAD MASS VS. ORBIT TYPE
- YEARLY TREND OF SUCCESS RATE

SCATTER PLOTS WERE UTILIZED TO ILLUSTRATE THE RELATIONSHIPS BETWEEN VARIABLES, OFFERING INSIGHTS THAT COULD BE VALUABLE IN CONSTRUCTING MACHINE LEARNING MODELS.

BAR CHARTS WERE EMPLOYED TO COMPARE DISCRETE CATEGORIES, EMPHASIZING THE RELATIONSHIPS BETWEEN SPECIFIC CATEGORIES AND MEASURED VALUES.

LINE CHARTS WERE USED TO DEPICT TRENDS IN DATA OVER TIME, PARTICULARLY IN THE CONTEXT OF TIME SERIES ANALYSIS.

Methodology - EDA with SQL

- DISPLAYING THE UNIQUE LAUNCH SITE NAMES IN THE SPACE MISSION.
- DISPLAYING 5 RECORDS WHERE LAUNCH SITES INITIATE WITH THE STRING 'CCA.'
- DISPLAYING THE TOTAL PAYLOAD MASS CARRIED BY NASA (CRS) LAUNCHED BOOSTERS.
- DISPLAYING THE AVERAGE PAYLOAD MASS CARRIED BY BOOSTER VERSION F9 V1.1.
- LISTING THE DATE OF THE INITIAL SUCCESSFUL GROUND PAD LANDING OUTCOME.
- LISTING THE NAMES OF BOOSTERS WITH SUCCESS IN DRONE SHIP LANDINGS AND PAYLOAD MASS BETWEEN 4000 AND 6000.
- LISTING THE TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES.
- LISTING THE NAMES OF BOOSTER VERSIONS THAT CARRIED THE MAXIMUM PAYLOAD MASS.
- LISTING THE FAILED DRONE SHIP LANDING OUTCOMES, ALONG WITH BOOSTER VERSIONS AND LAUNCH SITE NAMES, FOR THE MONTHS IN THE YEAR 2015.
- RANKING THE COUNT OF LANDING OUTCOMES (E.G., FAILURE (DRONE SHIP) OR SUCCESS (GROUND PAD)) BETWEEN THE DATES 2010-06-04 AND 2017-03-20 IN DESCENDING ORDER.

Methodology – Build an Interactive Map with Folium

MARKERS FOR ALL LAUNCH SITES:

- IMPLEMENTED MARKERS WITH CIRCLES, POPUP LABELS, AND TEXT LABELS FOR NASA JOHNSON SPACE CENTER, UTILIZING ITS LATITUDE AND LONGITUDE COORDINATES AS THE STARTING LOCATION.
- INTRODUCED MARKERS WITH CIRCLES, POPUP LABELS, AND TEXT LABELS FOR ALL LAUNCH SITES, INCORPORATING THEIR LATITUDE AND LONGITUDE COORDINATES TO ILLUSTRATE THEIR GEOGRAPHICAL POSITIONS AND PROXIMITY TO THE EQUATOR AND COASTLINES.

COLORED MARKERS FOR LAUNCH OUTCOMES AT EACH LAUNCH SITE:

- INTEGRATED COLORED MARKERS REPRESENTING SUCCESS (GREEN) AND FAILURE (RED) LAUNCHES, UTILIZING A MARKER CLUSTER TO IDENTIFY LAUNCH SITES WITH RELATIVELY HIGH SUCCESS RATES.

DISTANCES FROM A LAUNCH SITE TO ITS PROXIMITIES:

- INCORPORATED COLORED LINES TO DEPICT DISTANCES BETWEEN THE LAUNCH SITE KSC LC-39A (AS AN EXAMPLE) AND ITS SURROUNDINGS, INCLUDING RAILWAYS, HIGHWAYS, COASTLINES, AND THE CLOSEST CITY.

Methodology - Dashboard with Plotly Dash

DROPDOWN MENU FOR LAUNCH SITES:

- IMPLEMENTED A DROPDOWN MENU TO FACILITATE THE SELECTION OF LAUNCH SITES.

PIE CHART DISPLAYING SUCCESS LAUNCHES (ALL SITES/SPECIFIC SITE):

- INTRODUCED A PIE CHART TO VISUALLY REPRESENT THE OVERALL COUNT OF SUCCESSFUL LAUNCHES FOR ALL SITES AND THE BREAKDOWN OF SUCCESS VS. FAILURE COUNTS IF A PARTICULAR LAUNCH SITE IS CHOSEN.

PAYLOAD MASS RANGE SLIDER:

- INTEGRATED A SLIDER FOR SELECTING THE PAYLOAD MASS RANGE.

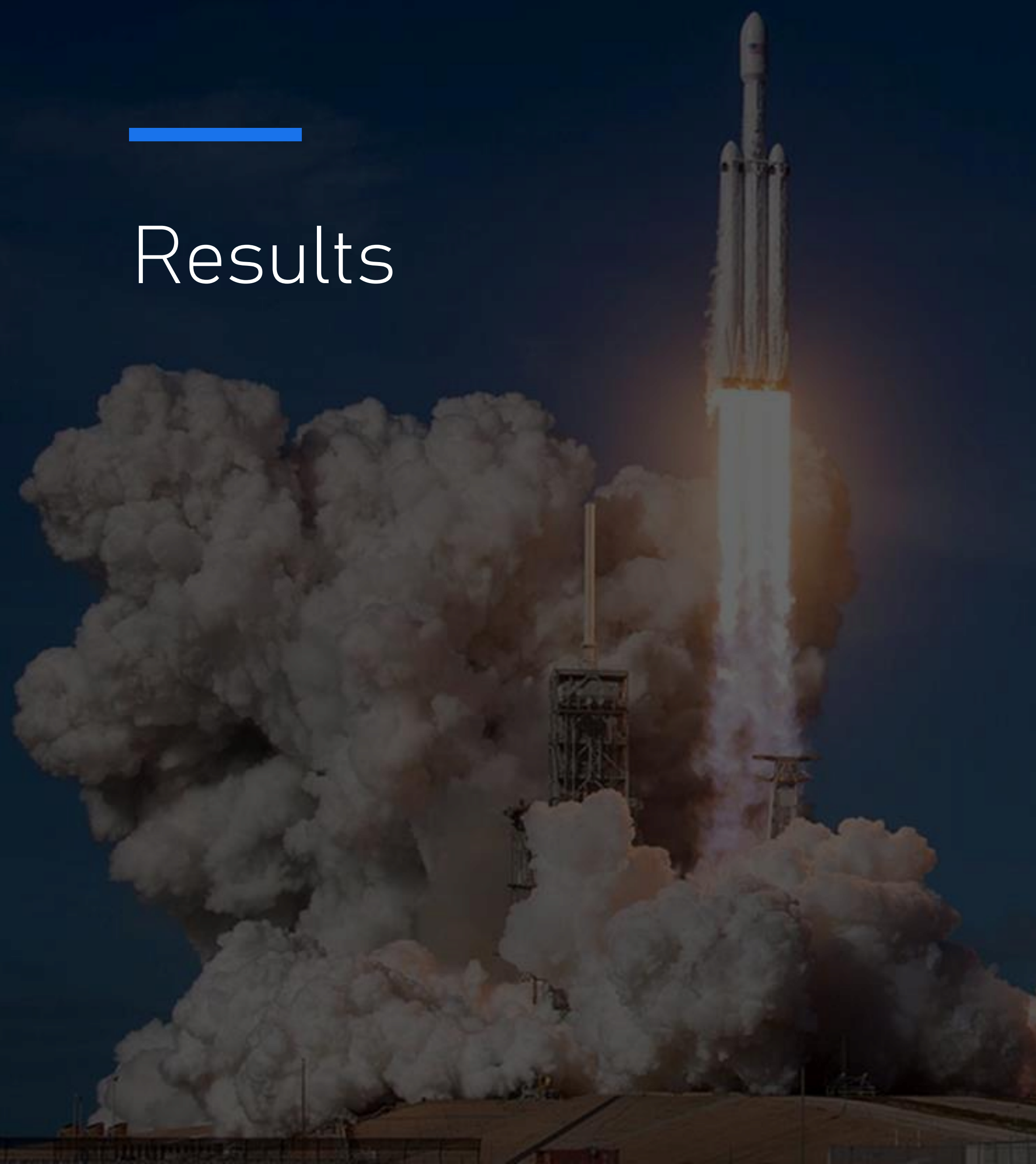
SCATTER CHART DEPICTING PAYLOAD MASS VS. SUCCESS RATE ACROSS VARIOUS BOOSTER VERSIONS:

- ADDED A SCATTER CHART TO ILLUSTRATE THE RELATIONSHIP BETWEEN PAYLOAD MASS AND LAUNCH SUCCESS FOR DIFFERENT BOOSTER VERSIONS.

Methodology - Predictive analysis (Classification)

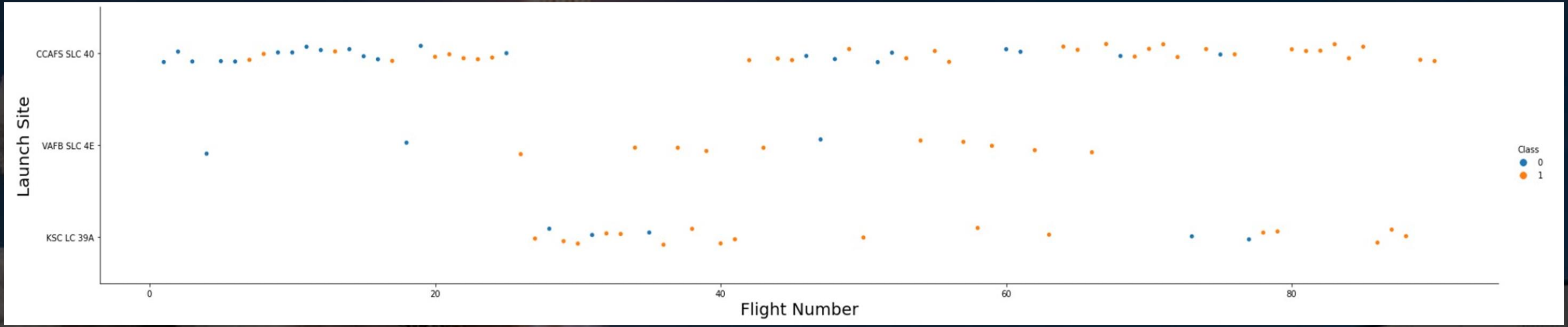
1. Creating a NumPy array by extracting the "Class" column from the dataset.
2. Standardizing the data using StandardScaler, involving both fitting and transforming the data.
3. Dividing the dataset into training and testing sets through the use of the train_test_split function.
4. Establishing a GridSearchCV object with a cross-validation value of 10 to identify the optimal parameters.
5. Applying GridSearchCV on Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN) models.
6. Calculating the accuracy on the test data for all models using the .score() method.
7. Analyzing the confusion matrix for each model.
8. Determining the best-performing method by evaluating Jaccard_score and F1_score metrics.

Results



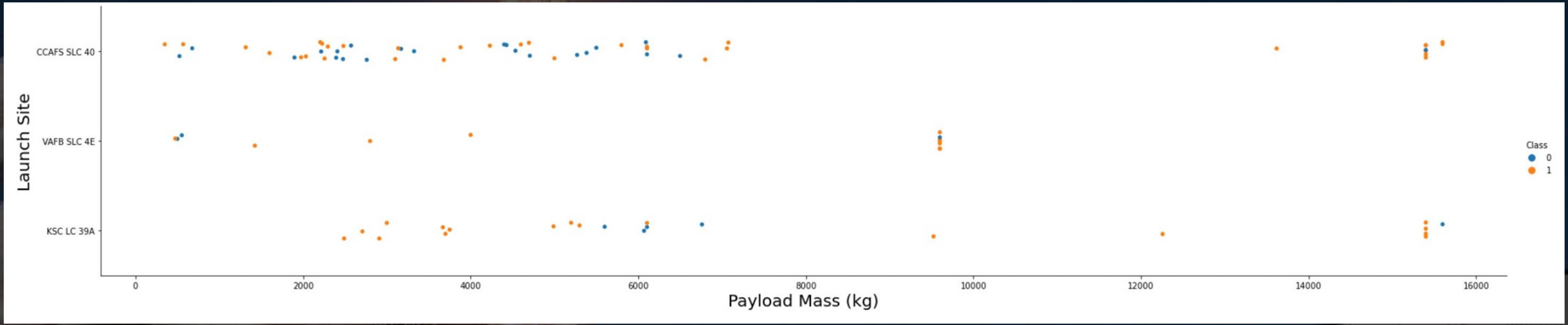
- EXPLORATORY DATA ANALYSIS RESULTS
- INTERACTIVE ANALYTICS DEMO IN SCREENSHOTS
- PREDICTIVE ANALYSIS RESULTS

EDA with Visualization (Flight Number vs. Launch Site)



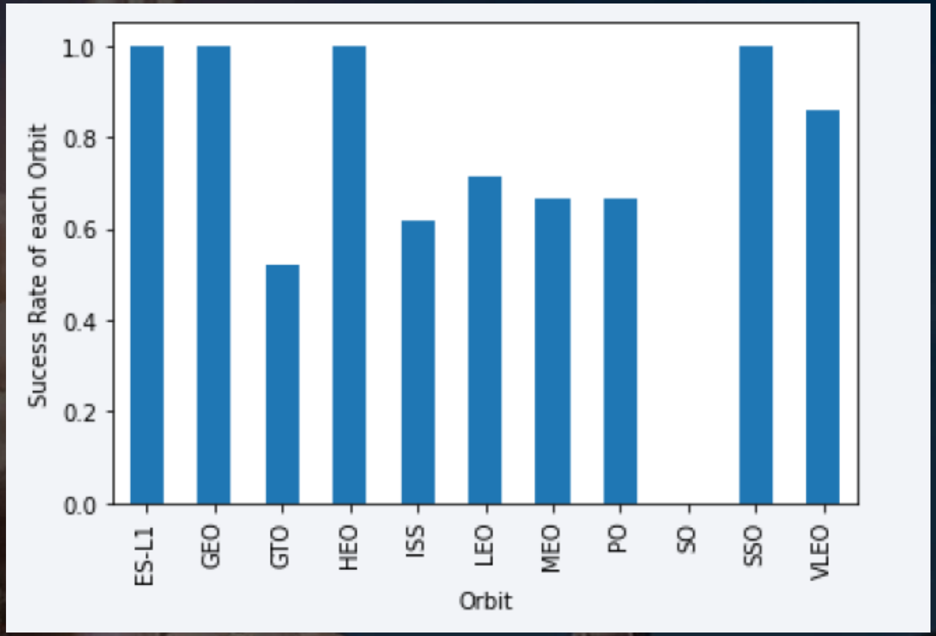
- IT IS EVIDENT THAT, FOR EACH SITE, THE SUCCESS RATE IS ON THE RISE.

EDA with Visualization (Payload vs. Launch Site)



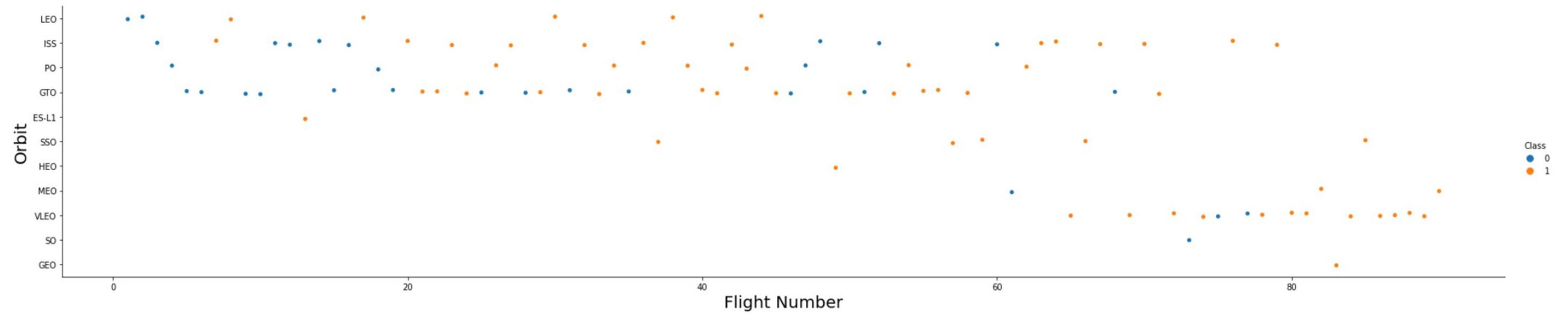
- THE SUCCESS OF A LANDING SEEMS TO BE INFLUENCED BY THE LAUNCH SITE, WITH HEAVIER PAYLOADS POTENTIALLY PLAYING A ROLE IN A SUCCESSFUL OUTCOME. HOWEVER, IT'S IMPORTANT TO NOTE THAT AN EXCESSIVELY HEAVY PAYLOAD COULD CONTRIBUTE TO A LANDING FAILURE.

EDA with Visualization (Success Rate vs. Orbit Type)



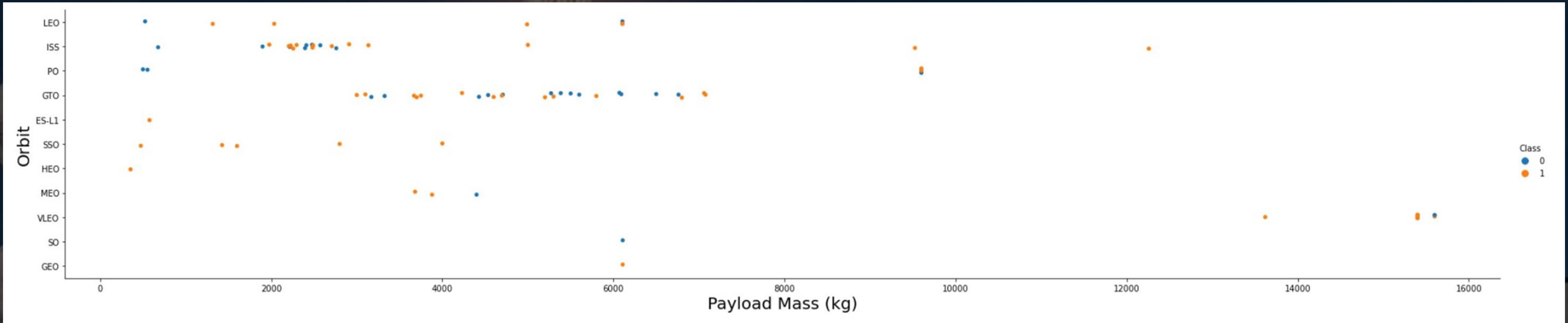
- THIS PLOT ALLOWS US TO OBSERVE THE SUCCESS RATES FOR DIFFERENT ORBIT TYPES. NOTABLY, ES L1, GEO, HEO, AND SSO EXHIBIT THE HIGHEST SUCCESS RATES.

EDA with Visualization (Flight Number vs. Orbit Type)



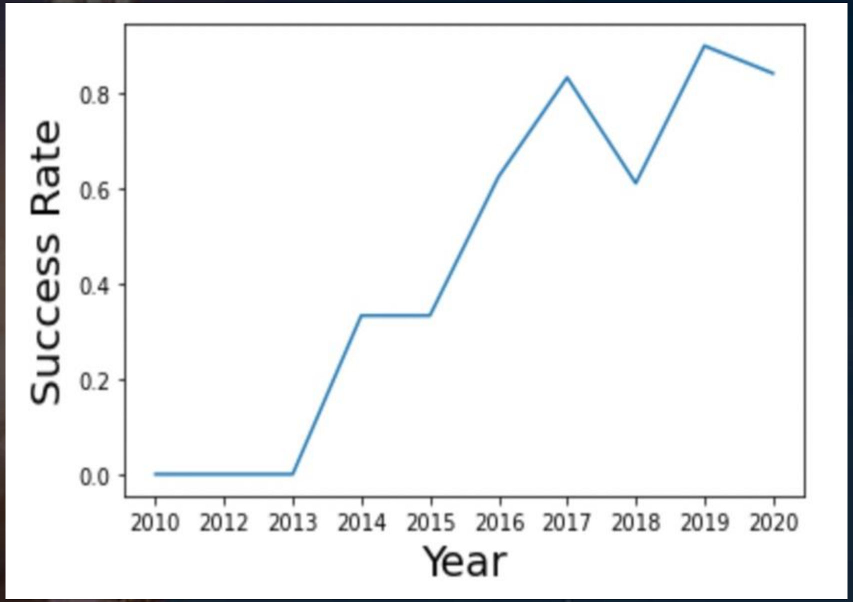
- THE SUCCESS RATE TENDS TO INCREASE WITH THE NUMBER OF FLIGHTS IN THE LOW EARTH ORBIT (LEO). WHILE NO CLEAR RELATIONSHIP IS OBSERVED FOR GEOSTATIONARY TRANSFER ORBIT (GTO), THE HIGH SUCCESS RATES IN ORBITS LIKE SUN-SYNCHRONOUS ORBIT (SSO) OR HIGHLY ELLIPTICAL ORBIT (HEO) MAY BE INFLUENCED BY KNOWLEDGE GAINED FROM PAST LAUNCHES IN OTHER ORBITS.

EDA with Visualization (Payload vs. Orbit Type)



- THE PAYLOAD WEIGHT SIGNIFICANTLY AFFECTS THE SUCCESS RATE OF LAUNCHES IN SPECIFIC ORBITS. FOR INSTANCE, IN THE LOW EARTH ORBIT (LEO), HEAVIER PAYLOADS CONTRIBUTE TO AN IMPROVED SUCCESS RATE. CONVERSELY, IN THE GEOSTATIONARY TRANSFER ORBIT (GTO), REDUCING THE PAYLOAD WEIGHT APPEARS TO ENHANCE THE LIKELIHOOD OF A SUCCESSFUL LAUNCH.

EDA with Visualization (Launch Success Yearly Trend)



- FROM 2013 ONWARD, THERE IS A NOTICEABLE UPWARD TREND IN THE SUCCESS RATE OF SPACEX ROCKETS.

EDA with SQL (All launch site names)

In [4]:
```
%sql select distinct launch_site from SPACEXDATASET;
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[4]:

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

- PRESENTING THE NAMES OF THE DISTINCT LAUNCH SITES IN THE SPACE MISSION.

EDA with SQL (Launch site names begin with '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

- SHOWCASING 5 RECORDS WHERE LAUNCH SITES COMMENCE WITH THE STRING 'CCA.'

EDA with SQL (Total payload mass)

In [6]:
```
%sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXDATASET where customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[6]:

total_payload_mass
45596

- PRESENTING THE TOTAL PAYLOAD MASS CARRIED BY BOOSTERS LAUNCHED BY NASA (CRS).

EDA with SQL (Average payload mass by F9 v1.1)

```
In [7]: %sql select avg(payload_mass__kg_) as average_payload_mass from SPACEXDATASET where booster_version like '%F9 v1.1%';

 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[7]:

average_payload_mass
2534

- ILLUSTRATING THE AVERAGE PAYLOAD MASS CARRIED BY THE BOOSTER VERSION F9 V1.1.

EDA with SQL (First successful ground landing date)

```
In [8]: %sql select min(date) as first_successful_landing from SPACEXDATASET where landing__outcome = 'Success (ground pad)';

 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
Out[8]:
```

first_successful_landing
2015-12-22

- ENUMERATING THE DATE OF THE INITIAL SUCCESSFUL LANDING OUTCOME ON THE GROUND PAD.

EDA with SQL (First successful ground landing date)

```
In [8]: %sql select min(date) as first_successful_landing from SPACEXDATASET where landing__outcome = 'Success (ground pad)';

 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[8]:

first_successful_landing
2015-12-22

- PROVIDING THE DATE OF THE FIRST SUCCESSFUL LANDING OUTCOME ON THE GROUND PAD.

EDA with SQL (Successful drone ship landing with payload between 4000 and 6000

In [9]:

```
%sql select booster_version from SPACEXDATASET where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ between 4
000 and 6000;
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[9]:

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- ENUMERATING THE NAMES OF BOOSTERS THAT ACHIEVED SUCCESS IN DRONE SHIP LANDINGS AND HAD A PAYLOAD MASS GREATER THAN 4000 BUT LESS THAN 6000.

EDA with SQL (Total number of successful and failure mission outcomes)

```
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

- TABULATING THE TOTAL NUMBER OF SUCCESSFUL AND FAILED MISSION OUTCOMES.

EDA with SQL (Boosters carried maximum payload)

```
In [11]: %sql select booster_version from SPACEXDATASET where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXDATASET);

 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
Out[11]:
```

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

- ENUMERATING THE NAMES OF BOOSTER VERSIONS THAT HAVE CARRIED THE MAXIMUM PAYLOAD MASS.

EDA with SQL (2015 launch records)

In [12]:
```
%%sql select monthname(date) as month, date, booster_version, launch_site, landing__outcome from SPACEXDATASET
      where landing__outcome = 'Failure (drone ship)' and year(date)=2015;
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[12]:

MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- PROVIDING A LIST OF FAILED LANDING OUTCOMES ON THE DRONE SHIP, ALONG WITH THEIR BOOSTER VERSIONS AND LAUNCH SITE NAMES, FOR THE MONTHS IN THE YEAR 2015.

EDA with SQL (Rank success count between 2010-06-04 and 2017-03-20)

```
In [13]: %%sql select landing__outcome, count(*) as count_outcomes from SPACEXDATASET
         where date between '2010-06-04' and '2017-03-20'
         group by landing__outcome
         order by count_outcomes desc;

 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[13]:

landing__outcome	count_outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- RANKING THE COUNT OF LANDING OUTCOMES (E.G., FAILURE (DRONE SHIP) OR SUCCESS (GROUND PAD) BETWEEN THE DATES 2010-06-04 AND 2017-03-20 IN DESCENDING ORDER.

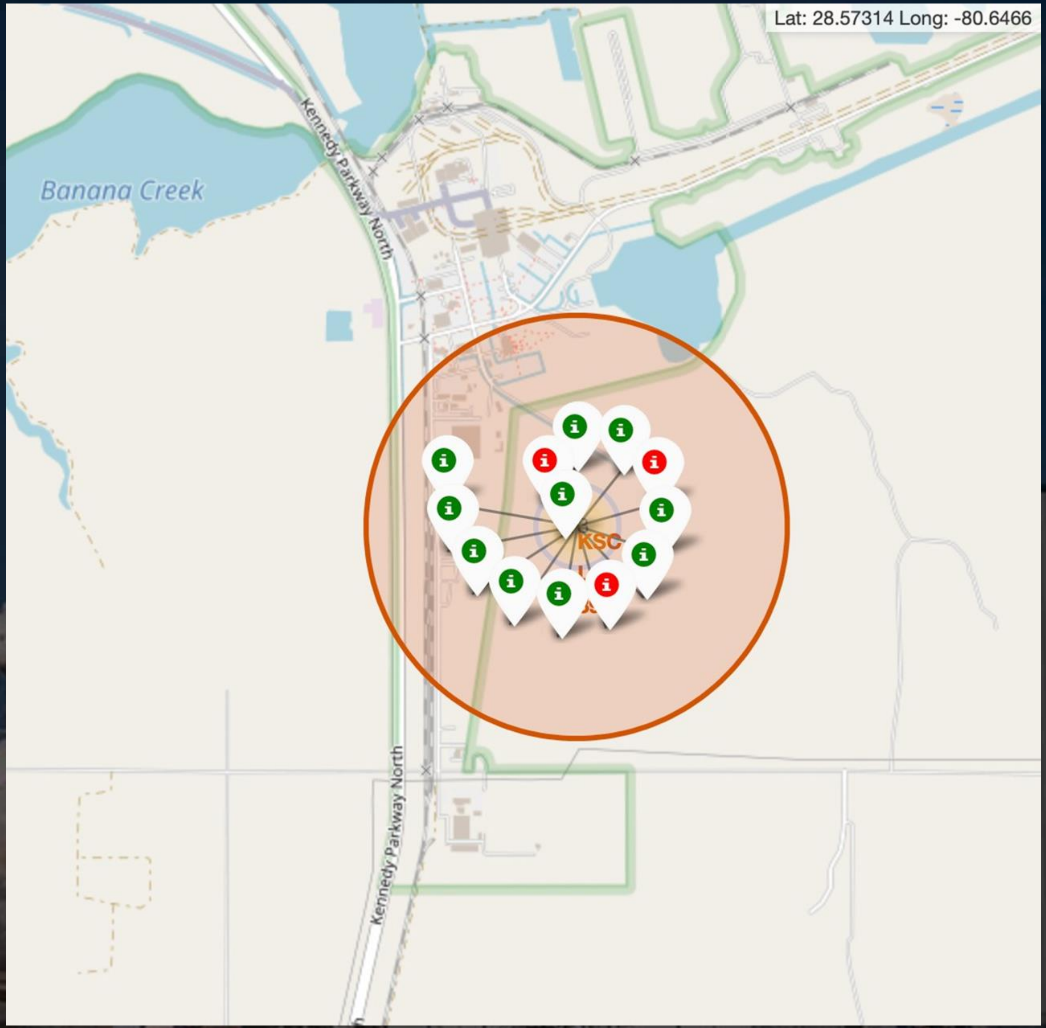
Interactive map with Folium (All launch sites' location markers on a global map)

- THE MAJORITY OF LAUNCH SITES ARE SITUATED NEAR THE EQUATOR LINE, WHERE THE EARTH'S SURFACE IS MOVING AT A HIGHER SPEED, REACHING 1670 KM/HOUR. WHEN A SPACECRAFT IS LAUNCHED FROM THE EQUATOR, IT INHERITS THIS EXISTING SPEED, AIDING IN ACHIEVING AND MAINTAINING THE NECESSARY VELOCITY FOR ORBITAL MOTION, THANKS TO THE PRINCIPLE OF INERTIA.

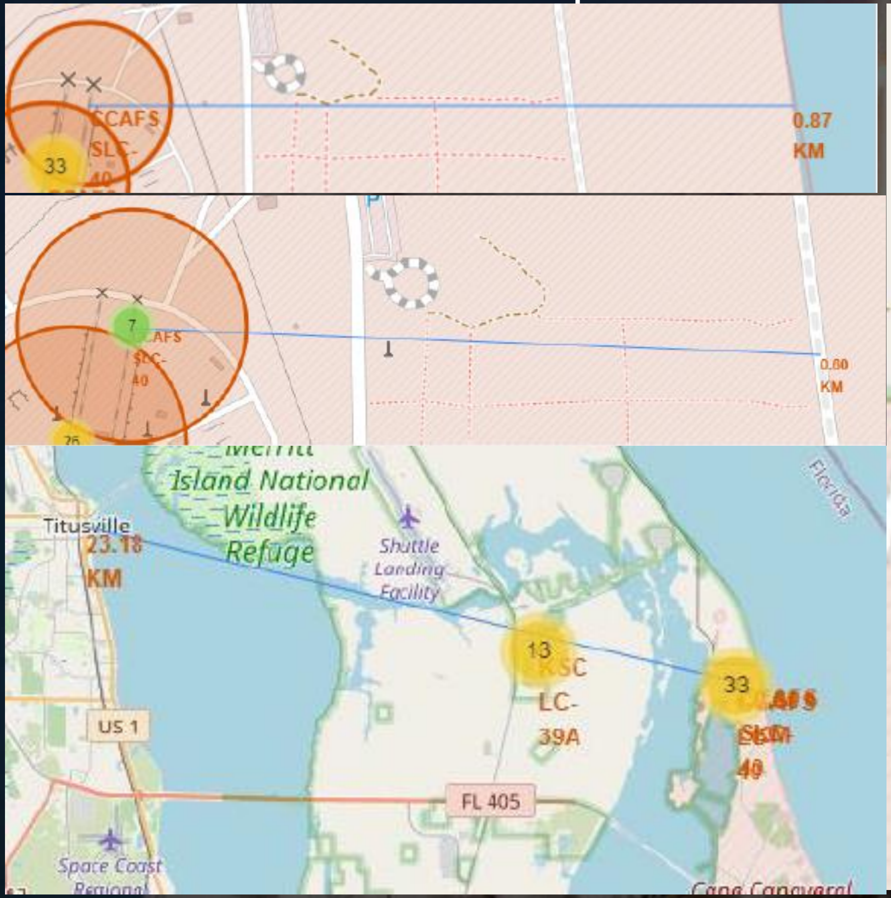
- ALL LAUNCH SITES ARE STRATEGICALLY LOCATED IN CLOSE PROXIMITY TO COASTLINES. LAUNCHING ROCKETS TOWARDS THE OCEAN SERVES TO MINIMIZE THE RISK OF DEBRIS FALLING OR EXPLODING NEAR POPULATED AREAS, ENSURING A SAFER TRAJECTORY FOR SPACE-BOUND VEHICLES.

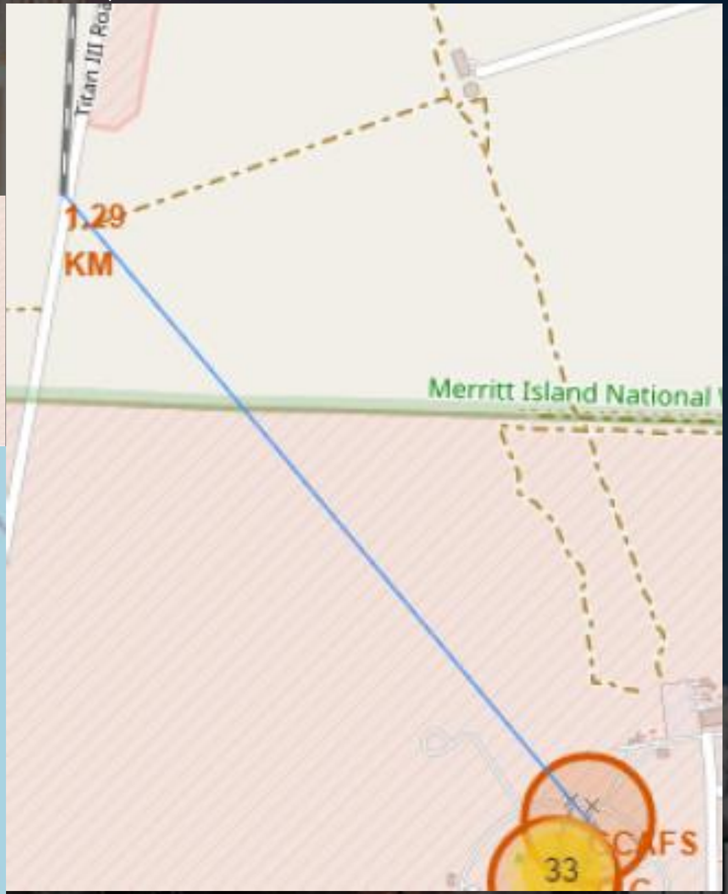
Interactive map with Folium (Color Labeled Markers)

THE GREEN MARKER SIGNIFIES SUCCESSFUL LAUNCHES, WHILE THE RED MARKER DENOTES UNSUCCESSFUL LAUNCHES. IT IS OBSERVED THAT KSC LC 39A EXHIBITS A NOTABLY HIGHER LAUNCH SUCCESS RATE.



Interactive map with Folium (Distances between CCAFS SLC 40 and its proximities)





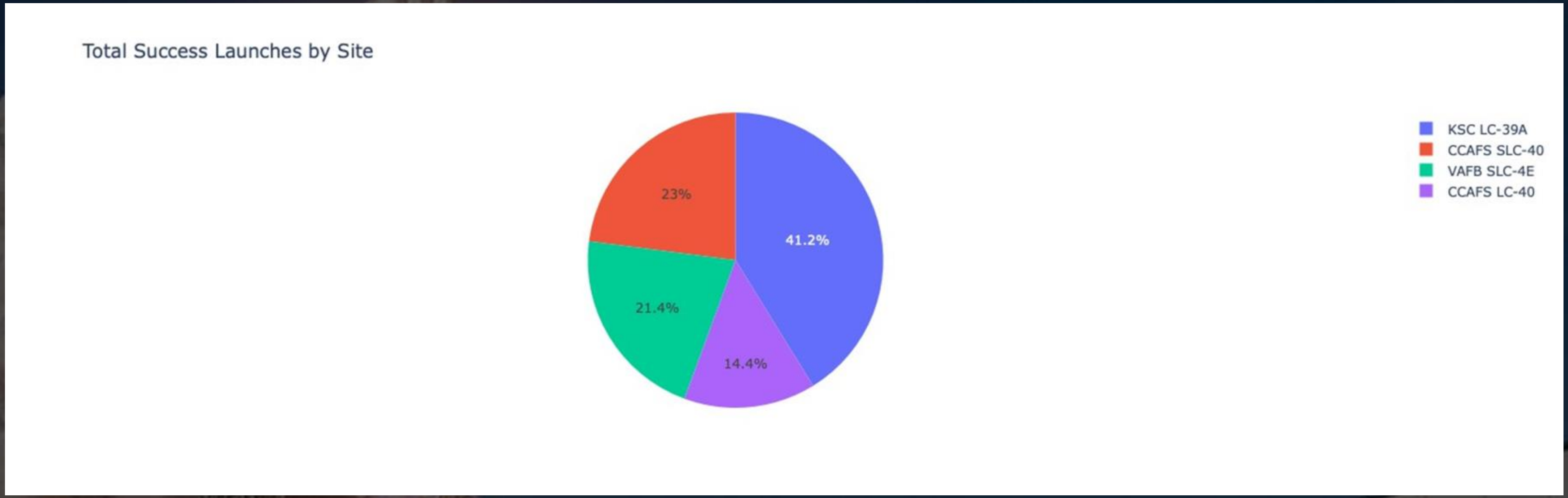
· UPON VISUALLY ANALYZING THE LAUNCH SITE KSC LC-39A, IT'S EVIDENT THAT IT IS:

- RELATIVELY CLOSE TO A RAILWAY (15.23 KM),
- RELATIVELY CLOSE TO A HIGHWAY (20.28 KM),
- RELATIVELY CLOSE TO THE COASTLINE (14.99 KM).

· ADDITIONALLY, THE LAUNCH SITE KSC LC-39A IS IN CLOSE PROXIMITY TO ITS NEAREST CITY, TITUSVILLE (16.32 KM).

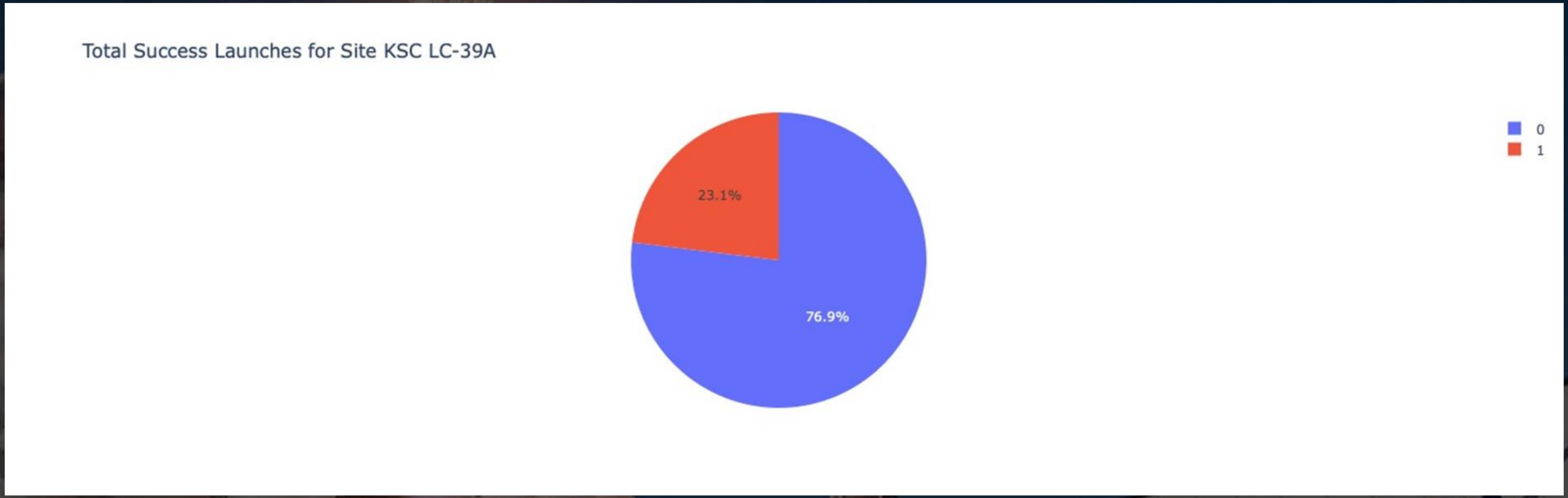
· CONSIDERING THAT A FAILED ROCKET, WITH ITS HIGH SPEED, CAN COVER DISTANCES OF 15-20 KM IN A MATTER OF SECONDS, THERE IS A POTENTIAL RISK TO POPULATED AREAS.

Build a Dashboard with Plotly Dash (Total success by Site)



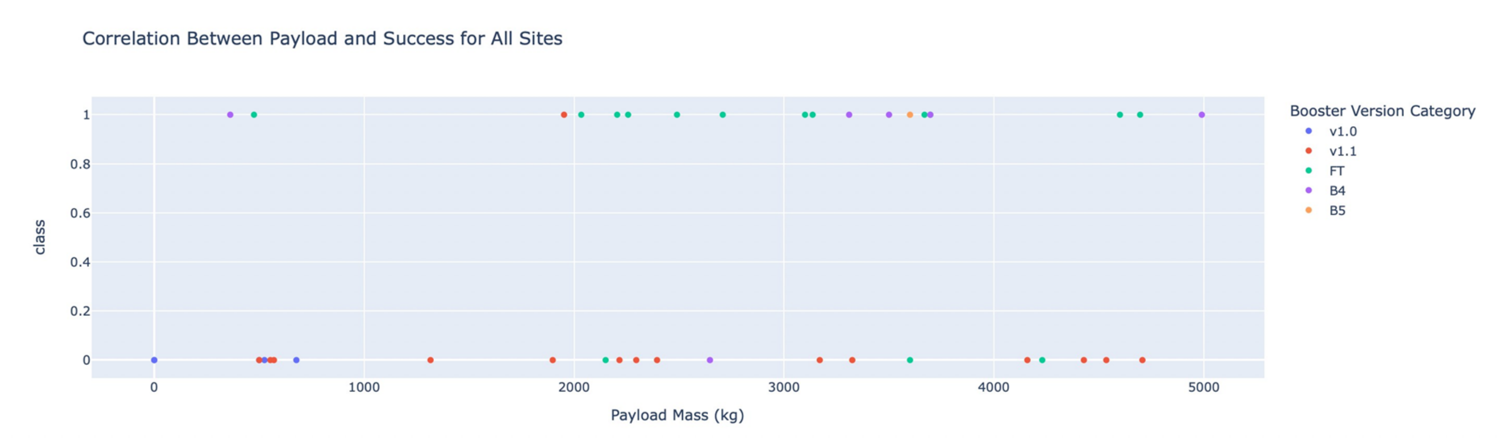
- THE CHART UNMISTAKABLY ILLUSTRATES THAT AMONG ALL THE SITES, KSC LC-39A HAS THE HIGHEST NUMBER OF SUCCESSFUL LAUNCHES.

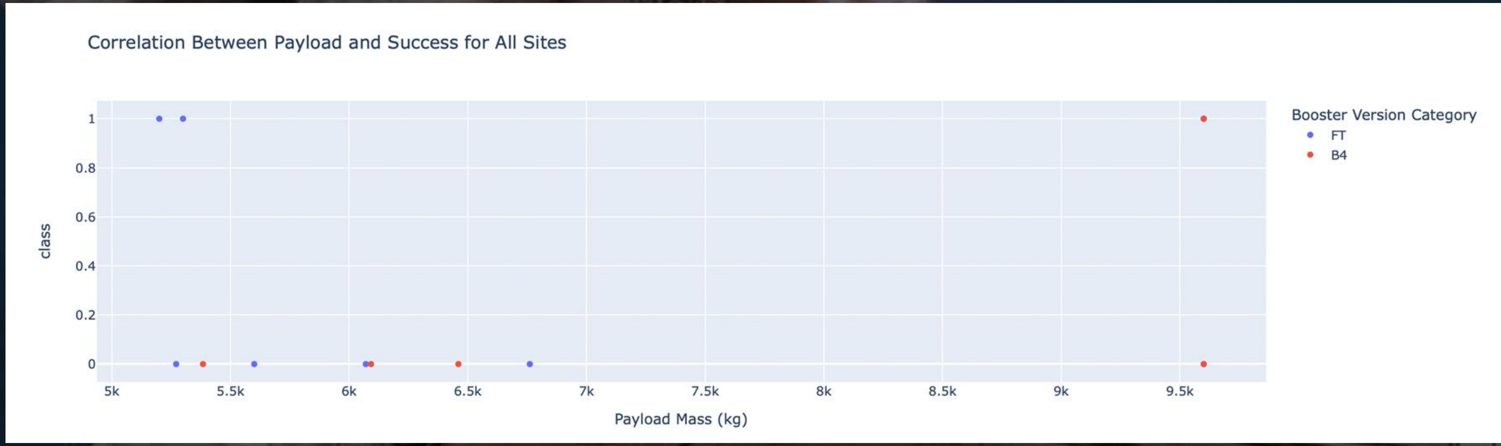
Build a Dashboard with Plotly Dash (Payload mass vs Outcome for all sites with different payload mass selected)



- KSC LC-39A BOASTS THE HIGHEST LAUNCH SUCCESS RATE AT 76.9%, ACHIEVING 10 SUCCESSFUL LANDINGS WITH ONLY 3 INSTANCES OF FAILURE.

Build a Dashboard with Plotly Dash (Total success by Site)





- THE CHARTS INDICATE THAT PAYLOADS RANGING BETWEEN 2000 AND 5500 KG EXHIBIT THE HIGHEST SUCCESS RATE.

Predictive analysis – Classification (Accuracy)

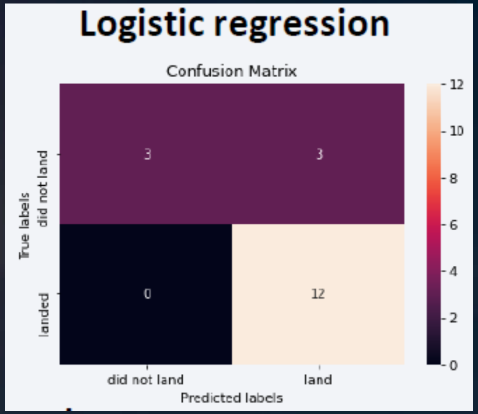
- SCORES OF THE TEST SET

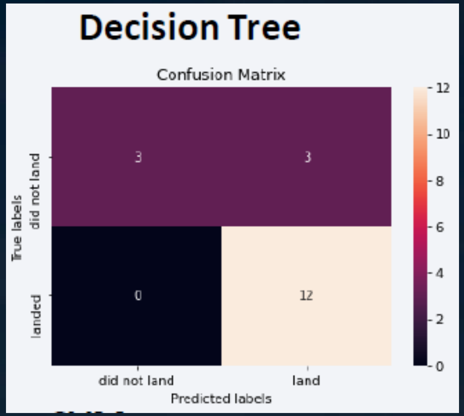
	Accuracy Train	Accuracy Test
Tree	0.876786	0.833333
Knn	0.848214	0.833333
Svm	0.848214	0.833333
Logreg	0.846429	0.833333

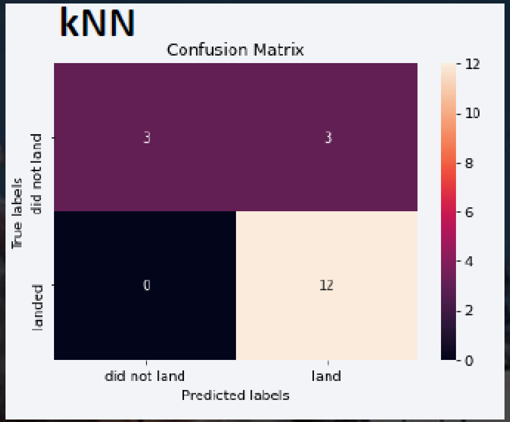
- WHILE THE TEST SET SCORES DO NOT PROVIDE CONCLUSIVE EVIDENCE REGARDING THE BEST-PERFORMING METHOD, IT'S IMPORTANT TO NOTE THAT THE SIMILARITY IN SCORES MAY BE ATTRIBUTED TO THE SMALL SAMPLE SIZE (18 SAMPLES). TO MITIGATE THIS, WE CONDUCTED TESTS ON THE ENTIRE DATASET. THE OVERALL DATASET SCORES AFFIRM THAT THE DECISION TREE MODEL STANDS OUT AS THE BEST MODEL, NOT ONLY EXHIBITING HIGHER SCORES BUT ALSO ACHIEVING THE HIGHEST ACCURACY.

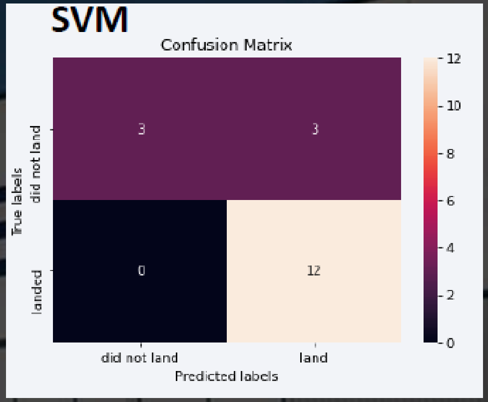
Predictive analysis – Classification (Confusion Matrix)

- GIVEN THAT THE TEST ACCURACIES ARE IDENTICAL, THE CONFUSION MATRICES ALSO EXHIBIT UNIFORMITY. THE PRIMARY ISSUE WITH THESE MODELS APPEARS TO BE FALSE POSITIVES.









Conclusions

- THE DECISION TREE MODEL PROVES TO BE THE MOST EFFECTIVE ALGORITHM FOR THIS DATASET.
- LAUNCHES WITH LOWER PAYLOAD MASSES TEND TO YIELD BETTER RESULTS COMPARED TO LARGER PAYLOAD LAUNCHES.
- MOST LAUNCH SITES ARE LOCATED IN CLOSE PROXIMITY TO THE EQUATOR LINE, AND ALL SITES ARE IN VERY CLOSE PROXIMITY TO THE COAST.
- THE SUCCESS RATE OF LAUNCHES DEMONSTRATES AN INCREASING TREND OVER THE YEARS.
- KSC LC-39A EXHIBITS THE HIGHEST SUCCESS RATE AMONG ALL LAUNCH SITES.

Thank you for your time