PREDICTIVE MODEL FOR FALCON 9 FIRST STAGE LANDING SUCCESS

BY AHMAD AJAZ

10/05/24

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Project Objective: Develop a machine learning model to predict the first stage landing success of SpaceX's Falcon 9 rocket launches.
- Approach: Utilized historical data on Falcon 9 launches, incorporating weather conditions, mission details, and technical parameters. Employed exploratory data analysis and machine learning techniques for predictive modeling.
- Key Findings: Identified critical factors influencing landing success and achieved robust predictive performance using classification algorithms.
- Impact: Successful prediction offers significant potential to enhance SpaceX's operational efficiency, mission planning, and risk management.
- Next Steps: Further refine the model and explore integration possibilities for real-time monitoring and adaptive control strategies

INTRODUCTION

Background

- SpaceX is the most successful company of the commercial space age, making space travel affordable.
- The company 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.
- Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.

Questions to be answered

- How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing?
- Does the rate of successful landings increase over the years?
- What is the best algorithm that can be used for binary classification in this case?

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 to prepare the data to a binary classification

- 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 to ensure the best results

DATA COLLECTION

URL- DATA COLLECTION

- Data collection process involved a combination of API requests from SpaceX REST API and Web Scraping data from a table in SpaceX's Wikipedia entry.

- We had to use both of these data collection methods in order to get complete information about the launches for a more detailed analysis.

- Data Columns are obtained by using SpaceX REST API:

Flight Number, Date, Booster Version, Payload Mass, Orbit, Launch Site, Outcome, Flights, Grid Fins, Reused, Legs, Landing Pad, Block, Reused Count, Serial, Longitude, Latitude

- Data Columns are obtained by using Wikipedia Web Scraping:

Flight No., Launch site, Payload, Payload Mass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

DATA COLLECTION – SPACEX API

1. Requesting Rocket launch data from SpaceX API
2. Decoding the response content using .json() and turning it into a dataframe using .json_normalize()
3. Requesting needed information about the launches from SpaceX API by applying custom functions
4. Constructing data we have obtained into a dictionary
5. Creating a dataframe from the dictionary
6. Filtering the dataframe to only include Falcon 9 launches
7. Replacing missing values of payload mass column with calculated .mean() for this column
8. Exporting the data to CSV

DATA COLLECTION – WEB SCRAPING

URL- WEB SCRAPING

1. Requesting Falcon 9 launch data from sources
2. Creating a BeautifulSoup object from the HTML response
3. Extracting all column names from the HTML table header
4. Collecting the data by parsing HTML tables
5. Constructing data we have obtained into a dictionary
6. Creating a dataframe from the dictionary
7. Exporting the data to CSV

DATA WRANGLING

URL- DATA WRANGLING

Understanding Landing Outcomes:

Various landing scenarios exist, including:

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
False ASDS: Unsuccessful drone ship landing

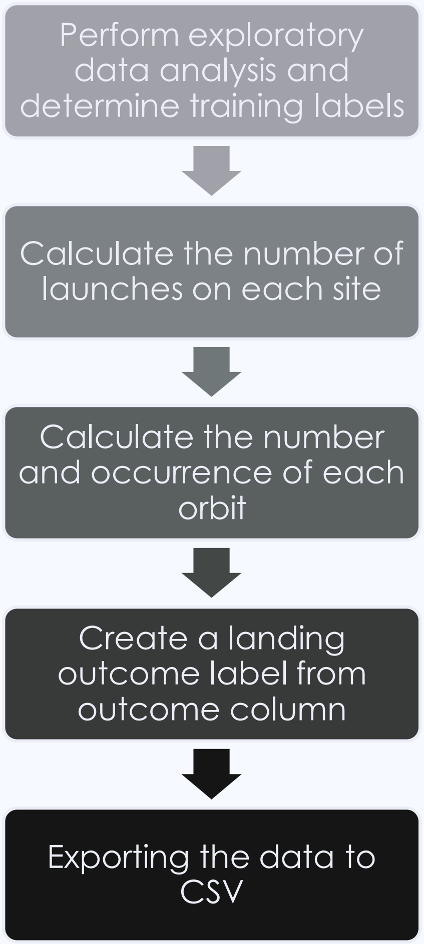
Conversion to Training Labels:

For model training, outcomes are converted to binary labels:

"1" denotes successful booster landing
"0" denotes unsuccessful landing

Purpose:

Simplifies classification task for machine learning models.
Enables accurate prediction of Falcon 9 landing success based on historical outcomes.



EDA WITH DATA VISUALIZATION

URL W/ DATA VISUALS

- **Charts Plotted:**
  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
  Success Rate Yearly Trend
- **Scatter Plots:**
  Illustrate relationships between variables.
  Potential use in machine learning models if significant relationships exist.
- **Bar Charts:**
  Compare discrete categories.
  Show relationships between specific categories and measured values.
- **Line Charts:**
  Depict trends over time (time series data).

EDA WITH SQL

URL - EDA WITH SQL

Performed SQL queries:

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'CCA'
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass
- Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015
- Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20 in descending order

BUILD AN INTERACTIVE MAP WITH FOLIUM URL

Launch Site Markers:

- Utilized latitude and longitude coordinates to mark the locations of all launch sites.
- Incorporated markers with circles, popup labels, and text labels, starting with NASA Johnson Space Center as the reference point.

Launch Outcome Visualization:

- Distinguished launch outcomes using colored markers (green for success, red for failure) across all launch sites.
- Employed marker clusters to identify launch sites with higher success rates based on launch outcomes.

Distance Visualization:

- Visualized distances between a launch site (e.g., KSC LC-39A) and its surroundings.
- Employed colored lines to represent distances to proximities such as railways, highways, coastlines, and closest cities.

BUILD A DASHBOARD WITH PLOTLY DASH URL

Launch Site Selection:
- Implemented a dropdown menu to facilitate the selection of launch sites.

Success Launches Visualization:
- Introduced a pie chart to display the total count of successful launches across all sites.
- For specific launch sites, the chart depicts the distribution of success and failure counts.

Payload Mass Range Slider:
- Incorporated a slider for selecting the payload mass range.

Payload vs. Success Rate Scatter Chart:
- Developed a scatter chart illustrating the relationship between payload mass and launch success rates.
- The chart categorizes data by different booster versions for comparative analysis.

PREDICTIVE ANALYSIS (CLASSIFICATION)

URL- MACHINE LEARNING PREDICTION

Creating a Numpy array from the column “Class” in data →

Standardizing the data with StandardScaler, then fitting and transforming it →

Splitting the data into training and testing sets with train_test_split function →

Creating a GrindSearchCV object with cv=10 to find the best parameters →

Applying GridSearchCV on LogReg, SVM, Decision Tree, and KNN models →

Calculating the accuracy on the test data using the method.score() for all models →

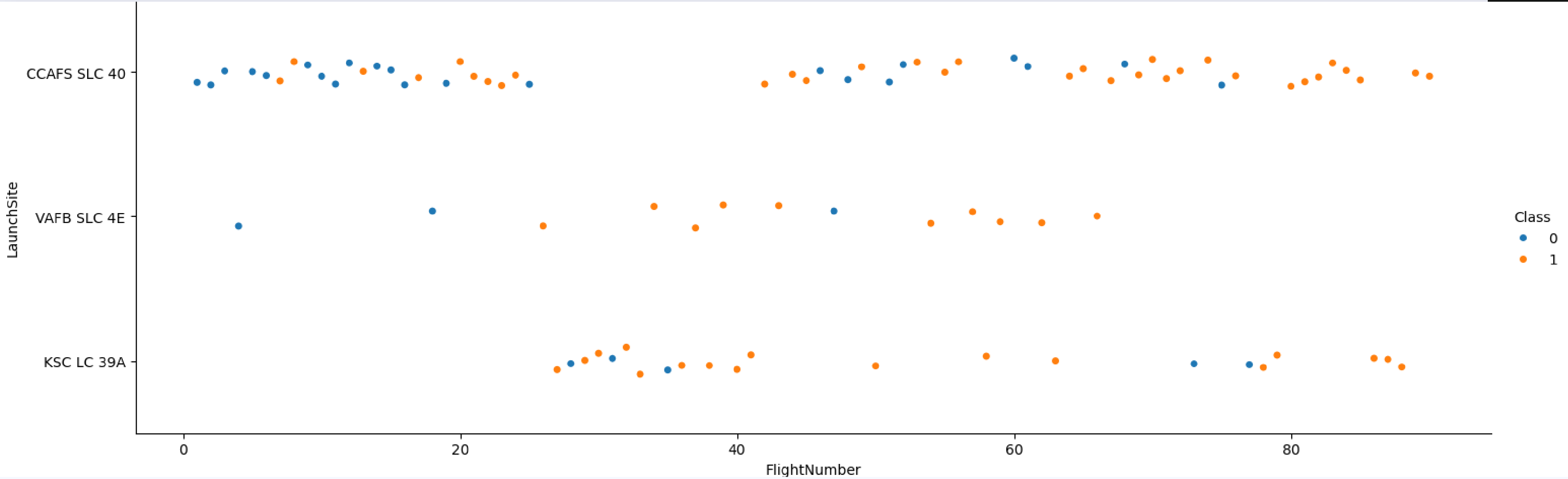
Examining the confusion matrix for all models →

Finding the method performs best by examining and Jaccard_score and F1_score metrics

RESULTS

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

EDA WITH VISUALIZATION

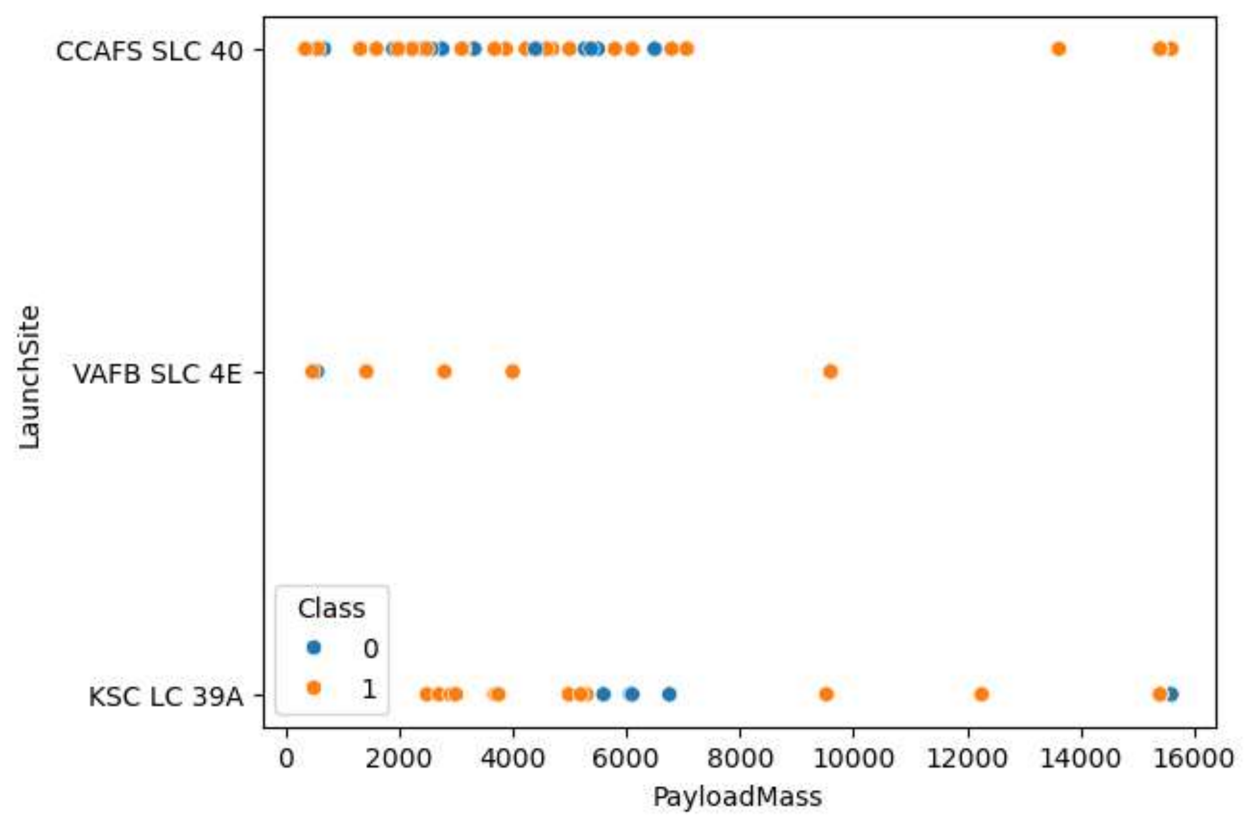


Flight Number vs. Launch Site

- The earliest flights all failed while the latest flights all succeeded.
- The CCAFS SLC 40 launch site has about a half of all launches.
- VAFB SLC 4E and KSC LC 39A have higher success rates.
- It can be assumed that each new launch has a higher rate of success.

PAYLOAD VS. LAUNCH SITE

- For every launch site the higher the payload mass, the higher the success rate.
- Most of the launches with payload mass over 7000 kg were successful.
- KSC LC 39A has a 100% success rate for payload mass under 5500 kg too.



SUCCESS RATE VS. ORBIT TYPE

EXPLANATION

- Orbits with 100% success rate:

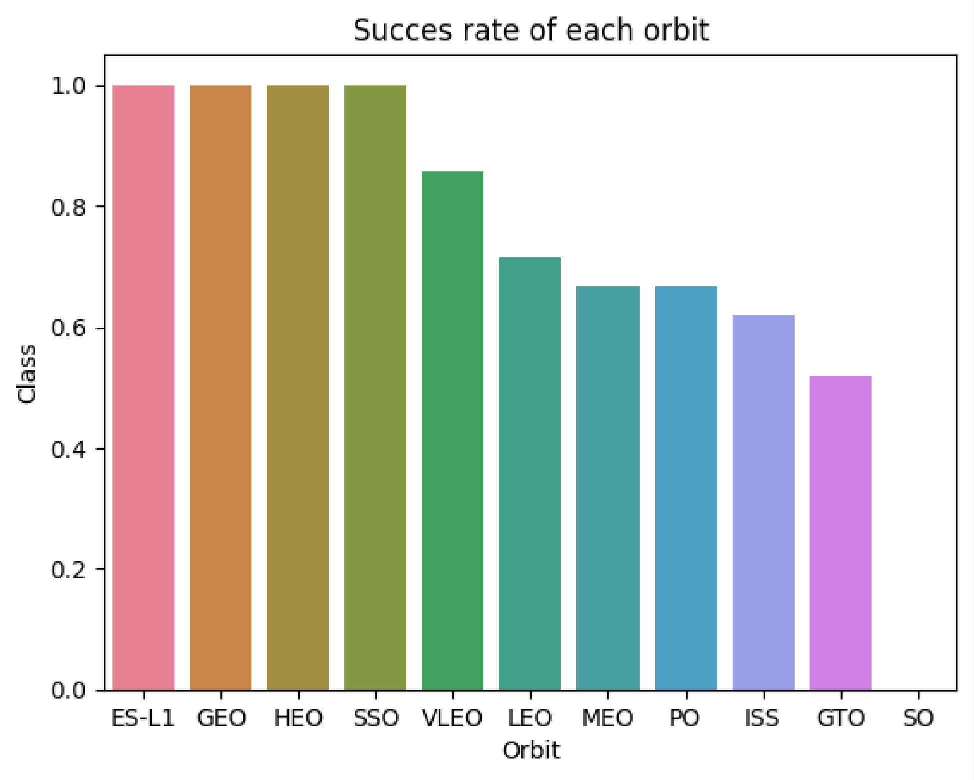
- ES-L1, GEO, HEO, SSO

- Orbits with 0% success rate:

- SO

- Orbits with success rate between 50% and 85%:

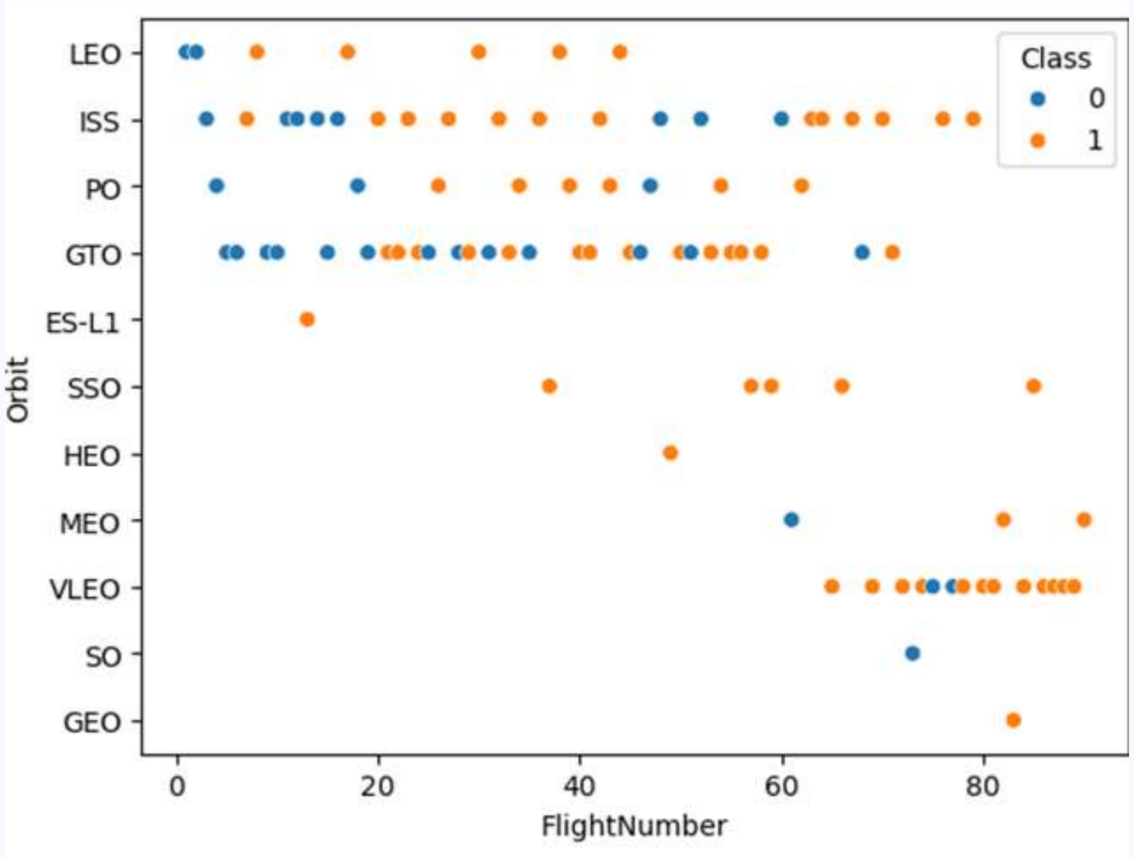
- GTO, ISS, LEO, MEO, PO



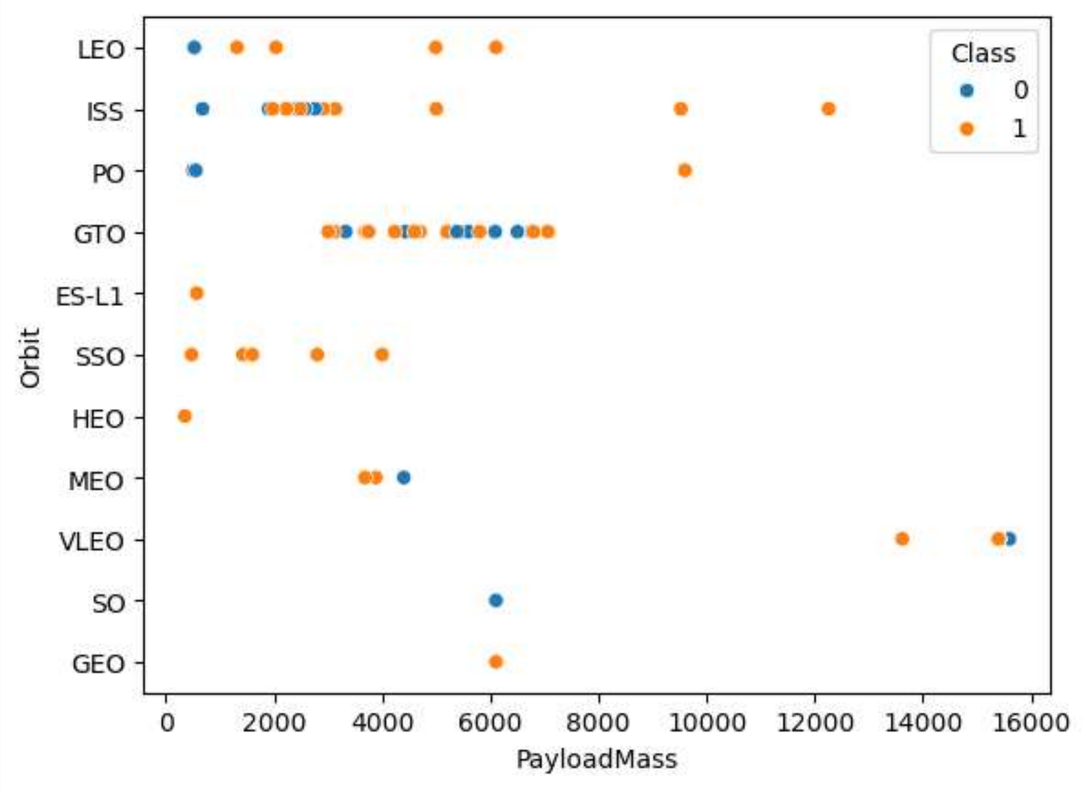
FLIGHT NUMBER VS. ORBIT TYPE

Explanation:

- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit



PAYLOAD MASS VS. ORBIT TYPE



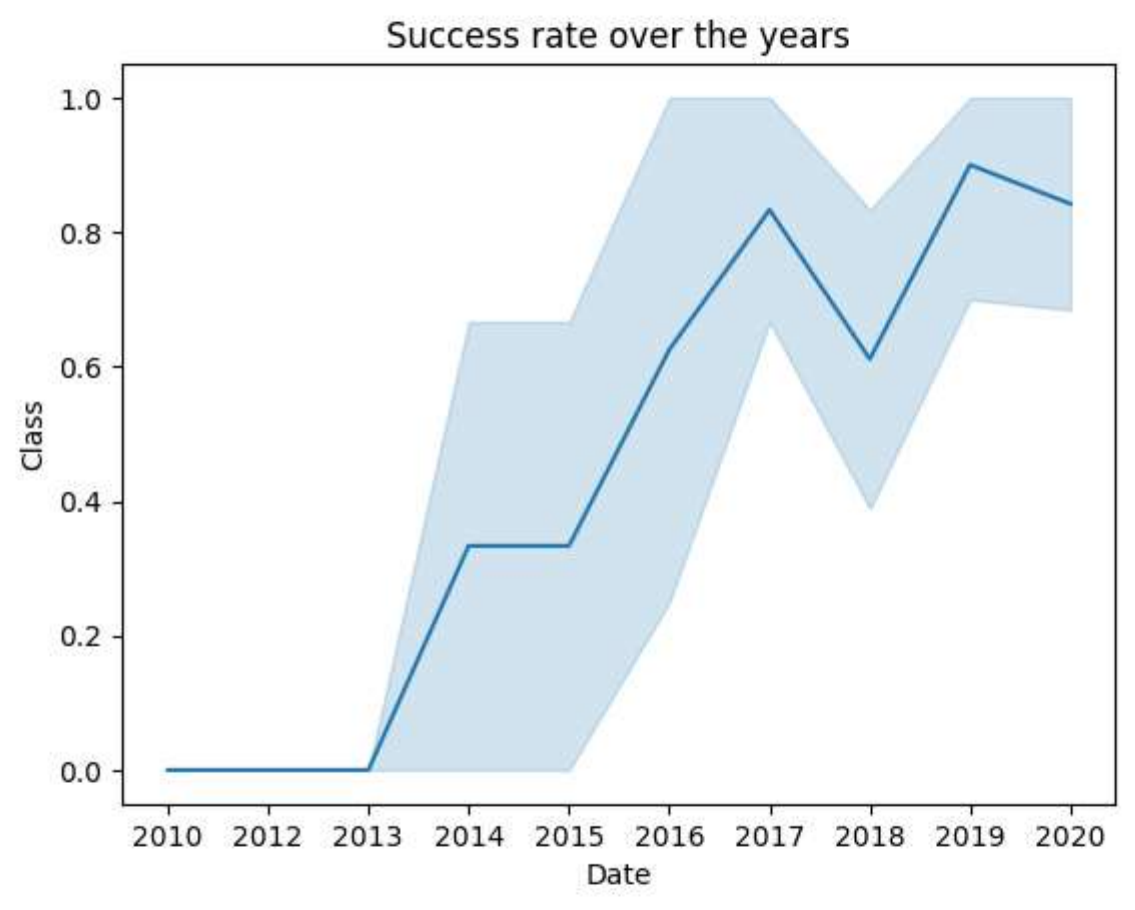
Explanation:

- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

LAUNCH SUCCESS YEARLY TREND

Explanation:

- The success rate since 2013 kept increasing till 2020.



EDA WITH SQL

LAUNCH SITE NAMES

- Displaying the names of the unique launch sites in the space mission.

```
%%sql
SELECT DISTINCT(Launch_Site) FROM SPACEXTABLE
```

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

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

LAUNCH SITE NAMES BEGIN WITH `CCA

```
%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.
```

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

Displaying 5 records where launch sites begin with the string 'CCA'.

TOTAL PAYLOAD MASS

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE
WHERE Customer == "NASA (CRS)"
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Displaying the total payload mass carried by boosters launched by NASA (CRS).

AVERAGE PAYLOAD MASS BY F9 V1.1

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE
WHERE Booster_Version == "F9 v1.1"
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

Displaying average payload mass carried by booster version F9 v1.1.

FIRST SUCCESSFUL GROUND LANDING DATE

```
%%sql
SELECT MIN(Date) FROM SPACEXTABLE
WHERE Landing_Outcome == "Success (ground pad)"
```

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

MIN(Date)
2015-12-22

Listing the date when the first successful landing outcome in ground pad was achieved.

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

```
%%sql
SELECT Payload FROM SPACEXTABLE
WHERE Landing_Outcome == "Success (drone ship)"
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

```
%%sql
SELECT COUNT(Mission_Outcome) FROM SPACEXTABLE
```

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

COUNT(Mission_Outcome)
101

Listing the total number of successful and failure mission outcomes.

MAXIMUM PAYLOAD CARRIED BY BOOSTER

Listing the names of the booster versions which have carried the maximum payload mass.

```
%%sql
SELECT Booster_Version from SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ IN (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
```

```
 * 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

LAUNCH RECORDS FROM 2015

```
%%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.
```

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)

Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015

RANK SUCCESS COUNT BETWEEN 2010-06-04 AND 2017-03-20

```
%sql SELECT Landing_Outcome, COUNT(*) AS COUNT_LAUNCHES FROM SPACEXTABLE WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GF
```

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

Landing_Outcome	COUNT_LAUNCHES
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

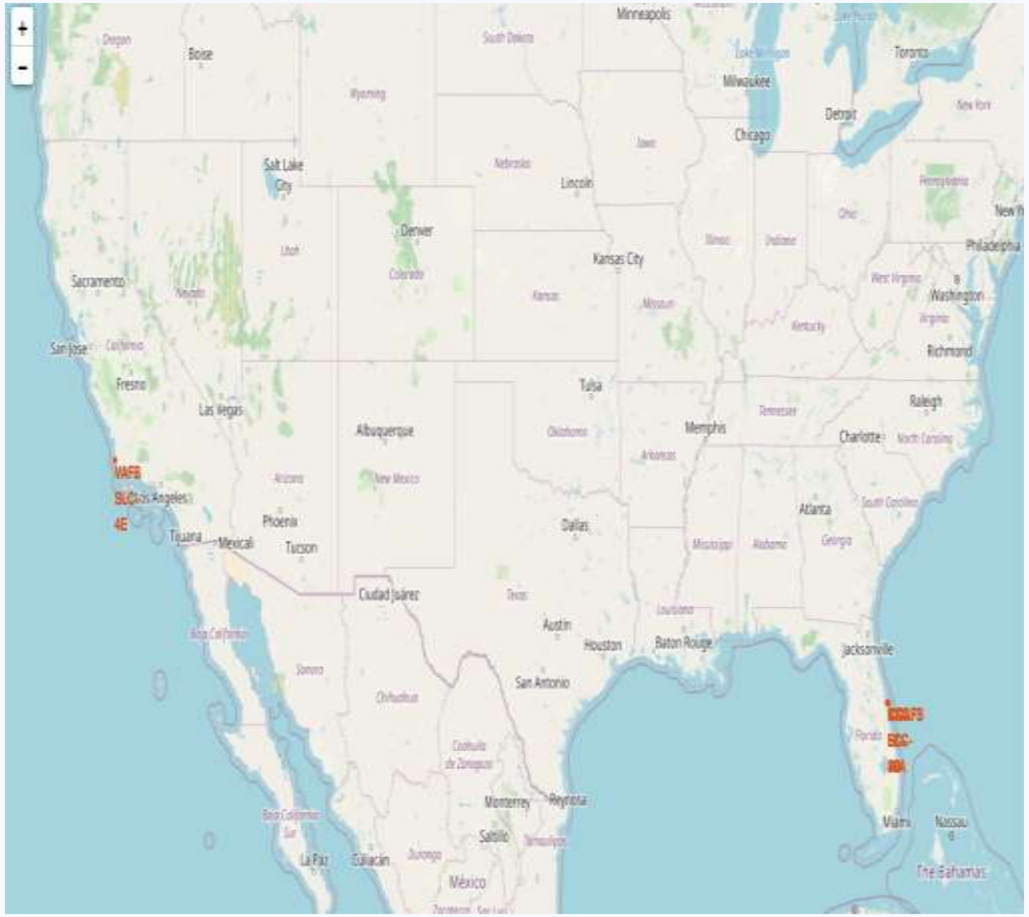
Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20 in descending order.

INTERACTIVE MAP WITH FOLIUM

MARKED LOCATIONS OF ALL THE LAUNCH SITES

- Most of Launch sites are in proximity to the Equator line. The land is moving faster at the equator than any other place on the surface of the Earth. Anything on the surface of the Earth at the equator is already moving at 1670 km/hour. If a ship is launched from the equator it goes up into space, and it is also moving around the Earth at the same speed it was moving before launching. This is because of inertia. This speed will help the spacecraft keep up a good enough speed to stay in orbit.

- All launch sites are in very close proximity to the coast, while launching rockets towards the ocean it minimizes the risk of having any debris dropping or exploding near people.



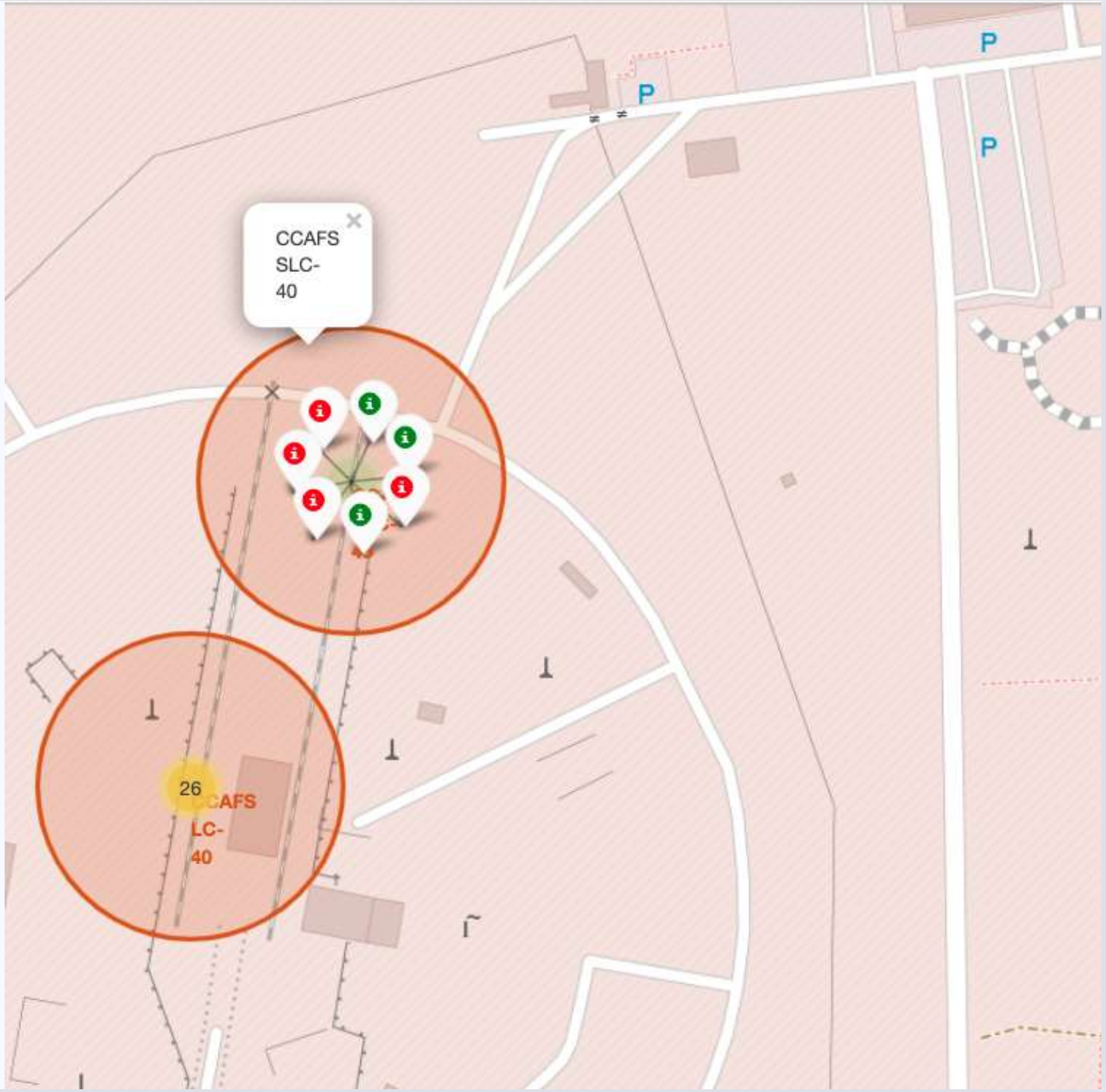
COLOUR-LABELED LAUNCH RECORDS ON THE MAP

- From the colour-labeled markers we should be able to easily identify which launch sites have relatively high success rates.

-Green Marker = Successful Launch

-Red Marker = Failed Launch

- Launch Site KSC LC-39A has a very high Success Rate.



DISTANCE FROM THE LAUNCH SITE KSC LC-39A TO ITS PROXIMITIES

- From the visual analysis of the launch site KSC LC-39A we can clearly see that it is:

 - relative close to railway (15.23 km)

 - relative close to highway (20.28 km)

 - relative close to coastline (14.99 km)

- Also the launch site KSC LC-39A is relative close to its closest city Titusville (16.32 km).

- Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially dangerous to populated areas.

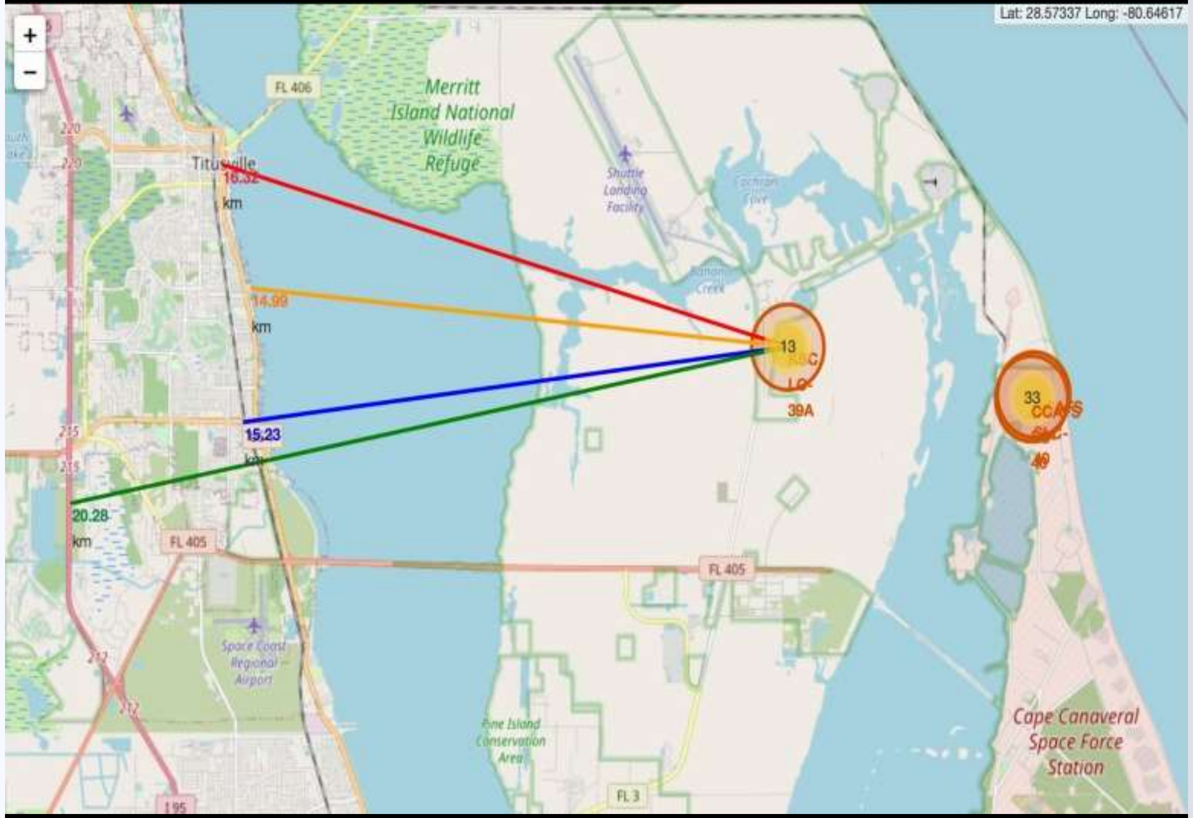
DISTANCE FROM THE LAUNCH SITE KSC LC-39A TO ITS PROXIMITIES

• From the visual analysis of the launch site KSC LC-39A we can clearly see that it is:

- relative close to railway (15.23 km)
- relative close to highway (20.28 km)
- relative close to coastline (14.99 km)

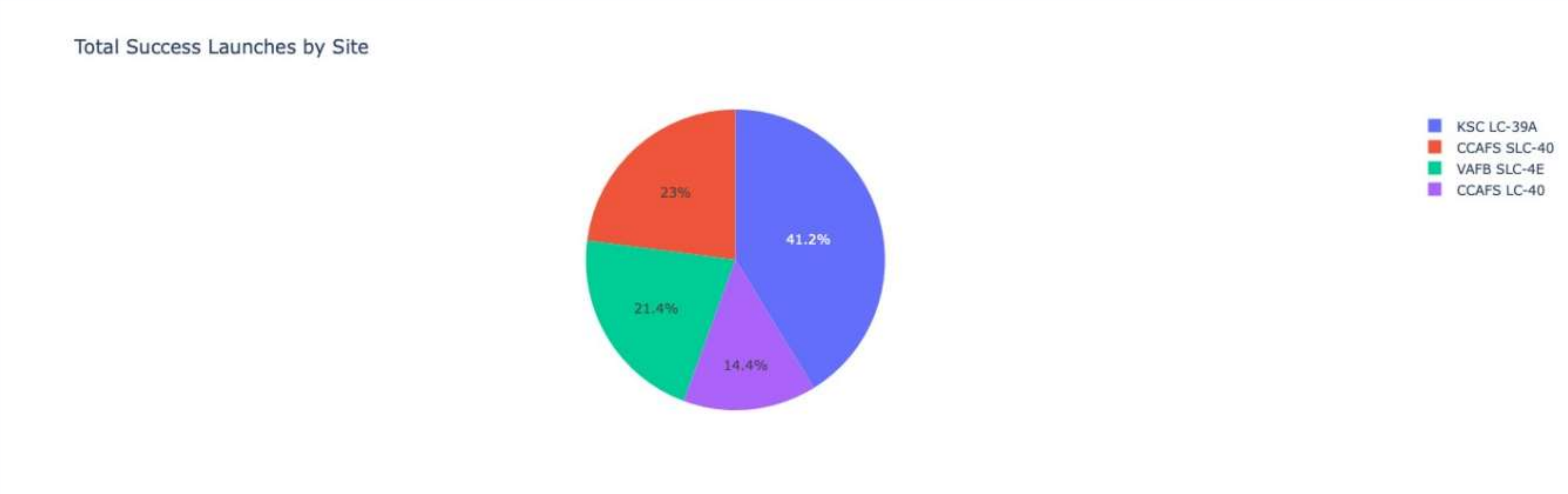
• Also the launch site KSC LC-39A is relative close to its closest city Titusville (16.32 km).

• Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially dangerous to populated areas.



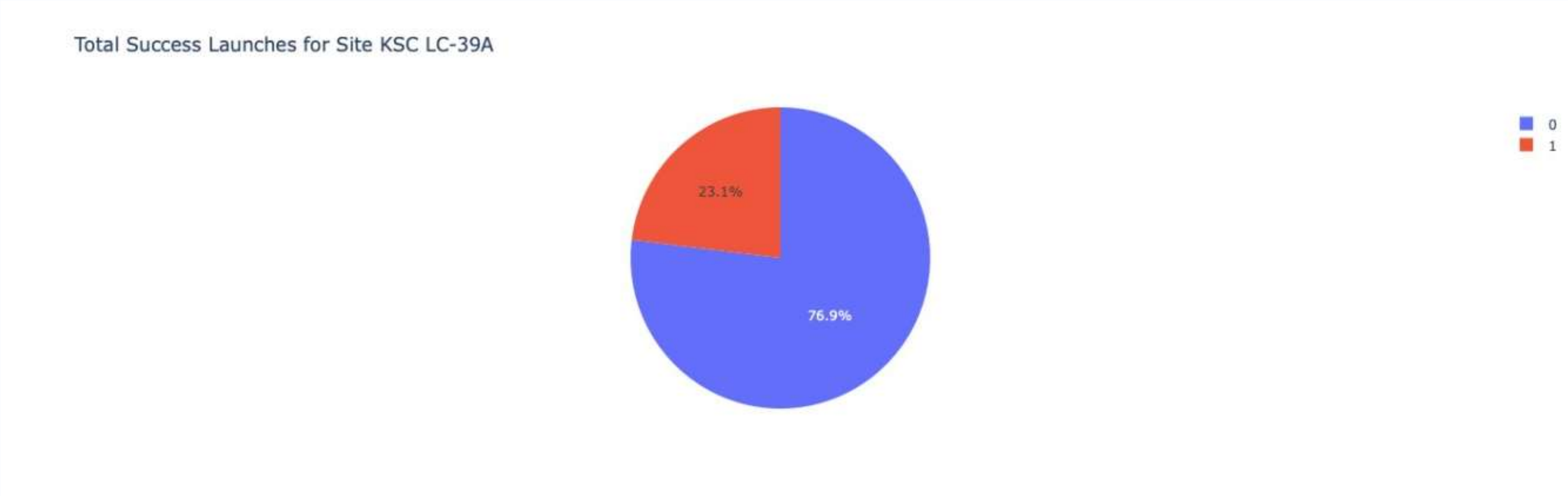
BUILD A DASHBOARD WITH PLOTLY DASH

LAUNCH SUCCESS COUNT FOR ALL SITES



The chart clearly shows that from all the sites, KSC LC-39A has the most successful launches.

LAUNCH SITE WITH HIGHEST LAUNCH SUCCESS RATIO



KSC LC-39A has the highest launch success rate (76.9%) with 10 successful and only 3 failed landings.

PAYLOAD MASS VS. LAUNCH OUTCOME FOR ALL SITES

The charts show that payloads between 2000 and 5500 kg have the highest success rate.

PREDICTIVE ANALYSIS

(CLASSIFICATION)

CLASSIFICATION ACCURACY

- Based on the scores of the Test Set, we can not confirm which method performs best.
- Same Test Set scores may be due to the small test sample size (18 samples). Therefore, we tested all methods based on the whole Dataset.
- The scores of the whole Dataset confirm that the best model is the Decision Tree Model. This model has not only higher scores, but also the highest accuracy.

Scores and Accuracy of the Test Set

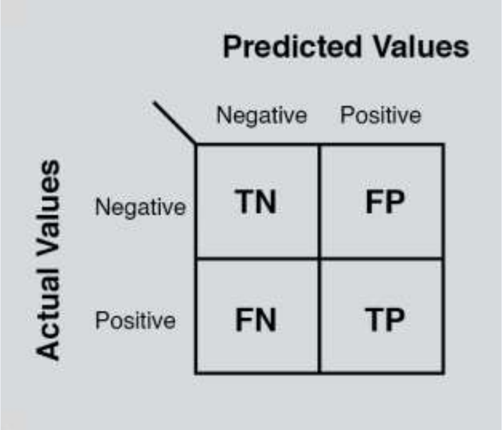
	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

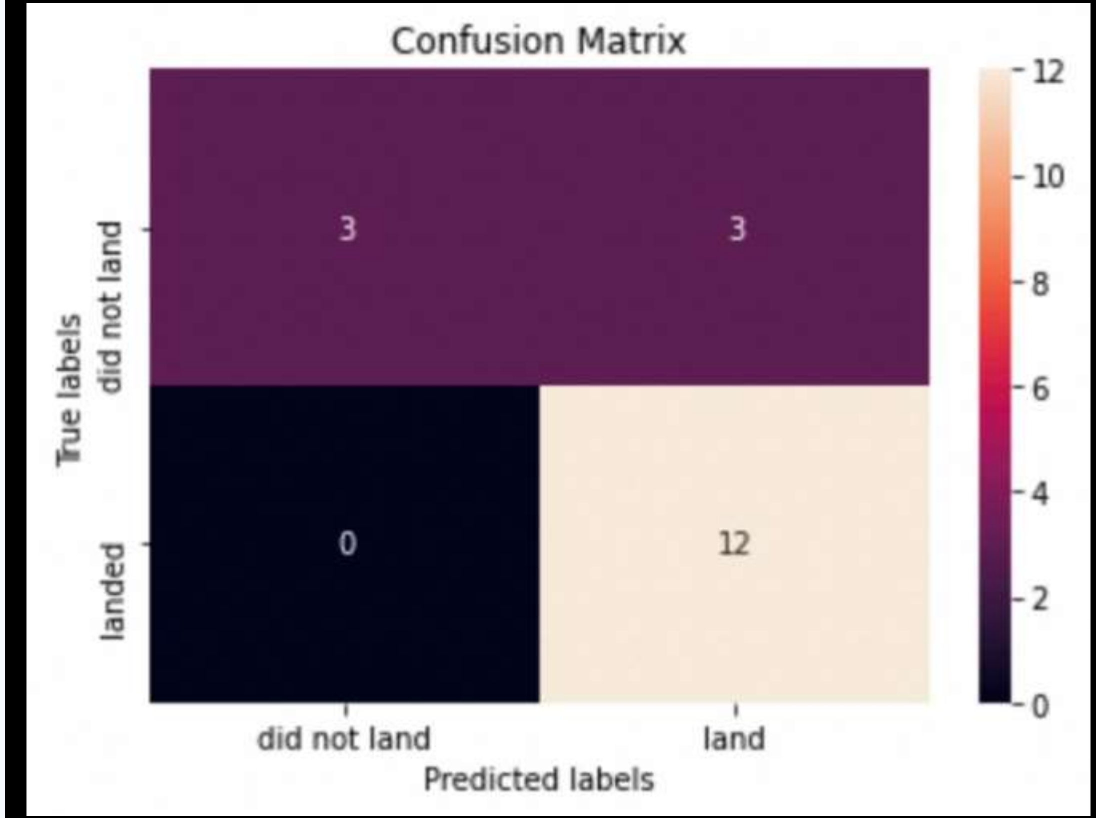
Scores and Accuracy of the Entire Data Set

	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

CONFUSION MATRIX

Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that





CONCLUSION

- The Decision Tree Model demonstrates superior performance with this dataset.
- Launches with lower payload masses exhibit higher success rates compared to those with larger payloads.
- The majority of launch sites are situated near the Equator line and in close proximity to coastlines.
- Over time, there is a noticeable upward trend in the success rate of launches.
- KSC LC-39A stands out with the highest success rate among all launch sites.
- Orbits ES-L1, GEO, HEO, and SSO achieve a 100% success rate.

APPENDIX

SPECIAL THANKS TO:
IBM & COURSERA