Success of the Falcon 9 First Stage Landings
IBM Data Science Capstone

Andrés Macedo Perea

05/04/2023

Outline

1. Executive Summary
2. Introduction
3. Methodology
4. Results
5. Discussion
6. Conclusion

Executive Summary

- Methodologies used for data analysis:
  - Data Collection using SpaceX REST API and web scrapping.
  - Exploratory Data Analysis using data wrangling, data visualization and interactive visual data.
  - Predictive Analysis using Machine Learning.
- Results summary:
  - Successfully collected data from a public source.
  - The most important factors to determine a successful launch were distinguished.
  - The best predictive model was identified.

Introduction

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, while other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. So, if we can determine if the first stage will land, we can determine the cost of a launch.

What we want to know is: What are the factors that affect the most how successful the first stage landing will be and with this, have the capability to determine the cost of launch.

Methodology

- Data Collection methods:
  - SpaceX REST API (https://api.spacexdata.com/v4/rockets/)
  - Web Scrapping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Exploratory Data Analysis methods:
  - Data wrangling for filtration, missing values and One Hot Encoding.
  - Visualization and SQL.
  - Visual analytics with Folium and Plotly Dash.
- Predictive Analysis method:
  - Classification models.

Data Collection: SpaceX REST API

Using the public API that SpaceX offers, data may be obtained by performing the following steps:

- Perform a request to the API.
- Decode with JSON with the response.
- Create a dataframe.
- Filter data for the Falcon 9 launches.
- Replace missing values.

SOURCE: https://github.com/AndresMacedo/IBMCapstone/blob/main/SpaceX%20Data%20Collection%20API.ipynb

Data Collection: Web Scrapping

Wikipedia contains information of the SpaceX launches which can be obtained by performing the following steps:

- Perform a request to the Wikipedia page.
- Create a BeautifulSoup object with the response.
- Collect data by parsing the HTML tables.
- Create a dataframe.

SOURCE: https://github.com/AndresMacedo/IBMCapstone/blob/main/SpaceX%20Web%20Scraping.ipynb

EDA: Data Wrangling

Exploratory Data Analysis is performed to obtain the following information:

- Launches on each site.
- Occurrence of each orbit.
- Occurrence of ission outcome per orbit type.

Then, a landing outcome label is created from the Outcome column.

SOURCE: https://github.com/AndresMacedo/IBMCapstone/blob/main/SpaceX%20Data%20Wrangling.ipynb

EDA: Data Visualization

Exploratory Data Analysis is performed with use of scatter plots, line charts and bar charts to visualize the relationship between:

- Payload mass and flight number.
- Launch site and flight number.
- Launch site and payload mass.
- Orbit and flight number.
- Payload and orbit.

SOURCE: https://github.com/AndresMacedo/IBMCapstone/blob/main/SpaceX%20EDA%20with%20Data%20Visualization.ipynb

EDA: SQL

Exploratory Data Analysis is performed with the following SQL queries:

- Names of the unique launch sites in the space mission.
- 5 records where launch sites begin with 'CCA'.
- Total payload mass carried by boosters launched by NASA (CRS).
- Average payload mass carried by booster version F9 v1.1.
- Date when the first successful landing outcome in ground pad was achieved.
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.

SOURCE: https://github.com/AndresMacedo/IBMCapstone/blob/main/SpaceX%20EDA%20with%20SQL.ipynb

EDA: SQL

- total number of successful and failure mission outcomes.
- names of the booster versions which have carried the maximum payload mass.
- records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.
- count of successful landing outcomes between the date 04-06-2010 and 20-03-2017.

EDA: Interactive Map with Folium

Exploratory Data Analysis is performed by adding the following features to a Folium map:

- Circle markers of the launch sites with popup labels and text labels.
- Colored markers of the launch outcomes for each launch site grouped in marker clusters.
- Lines to show distance between launch sites and points of interest.

SOURCE: https://github.com/AndresMacedo/IBMCapstone/blob/main/SpaceX%20Location%20Analysis%20with%20Folium.ipynb

EDA: Dashboard with Plotly Dash

Exploratory Data Analysis is performed by creating a Dashboard with the following features:

- Dropdown list to select launch sites.
- Pie charts showing success rate of each or all sites.
- Slider to select payload mass range.
- Scatter plot showing the relation between payload mass and success rate.

SOURCE: https://github.com/AndresMacedo/IBMCapstone/blob/main/spacex_dash_app.py

Predictive Analysis: Classification

Four classifications models were used: Logistic Regression, Support Vector Machine, Decision Tree and K Nearest Neighbors.

Predictive analysis is performed with this models by following the next steps:

- Data preparation and standarization.
- Split data in training and test sets.
- Apply the classification model.
- Obtain accuracy of the model.
- Result comparison.

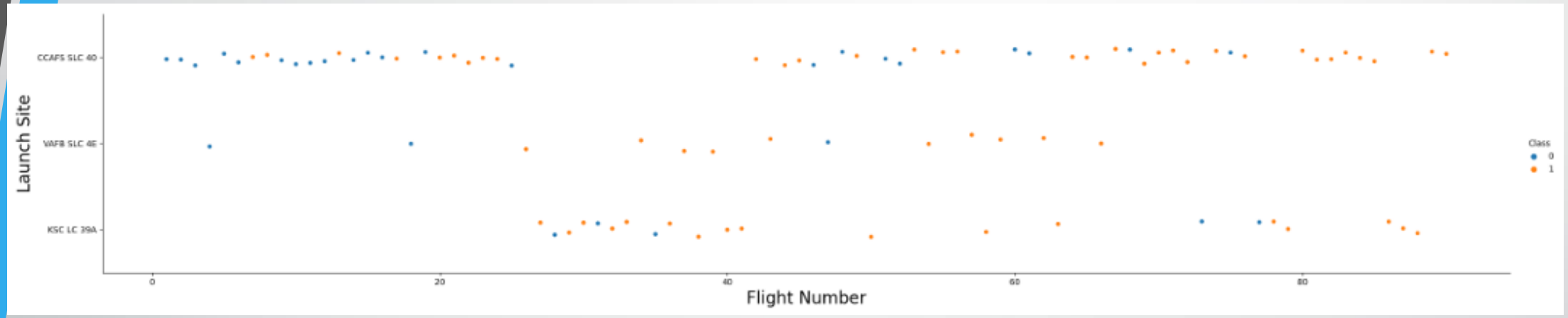
SOURCE: https://github.com/AndresMacedo/IBMCapstone/blob/main/SpaceX%20Predictions%20for%20First%20Launch.ipynb

Results

- Exploratory Data Analysis results
  - Visualization
  - SQL
  - Folium Map
  - Dashboard
- Predictive Analysis results
  - Classification

Results: Visualization

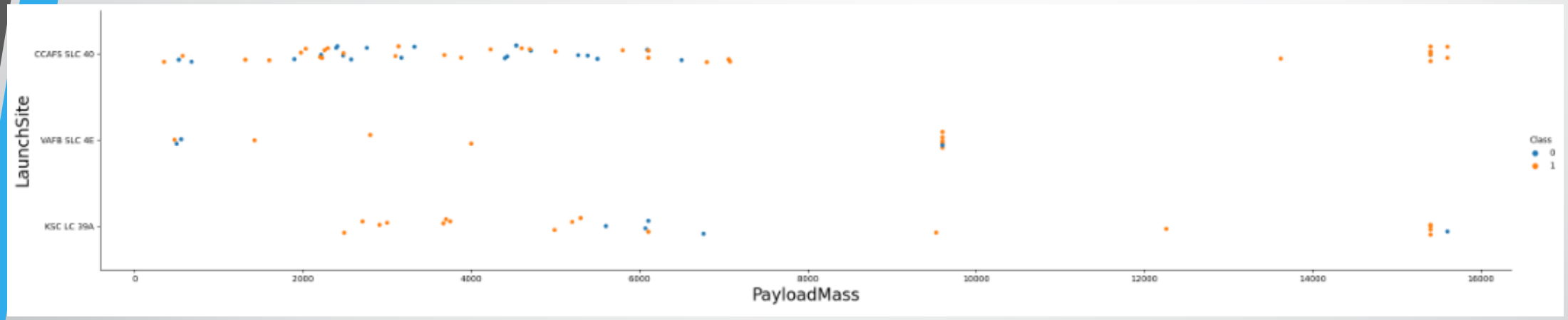
Flight Number vs. Launch Site



- The launch site with more successful launches is CCAFS LC-40, but also the site with more failed launches.
- Success rate improve over time for all sites.
- The launch site with higher success rate is KSC LC 39A.

Results: Visualization

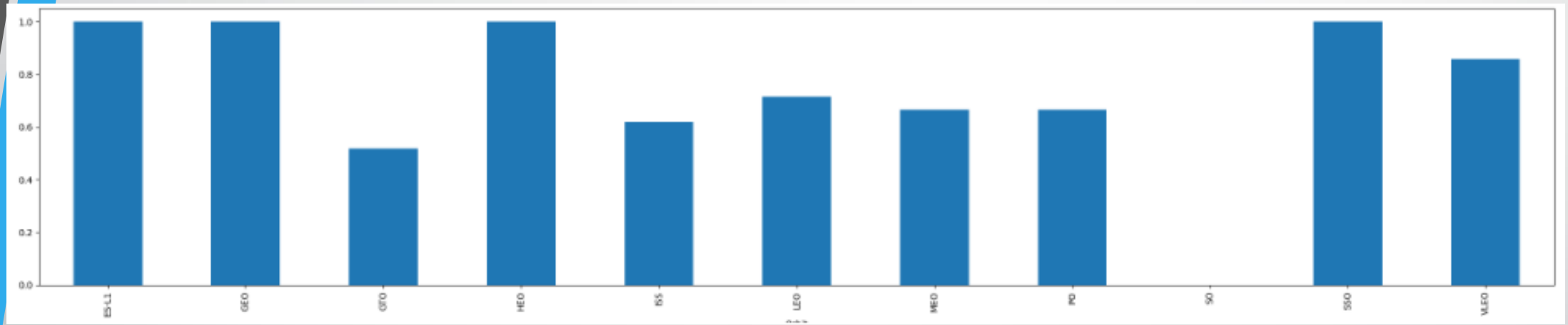
Payload Mass vs. Launch Site



- Payloads of over 9000 kg have a high success rate.
- Launches with the highest payloads are only performed in CCAFS LC-40 and KSC LC-39A.
- KSC LC-39A have a high success rate for payloads under 6000kg.

Results: Visualization

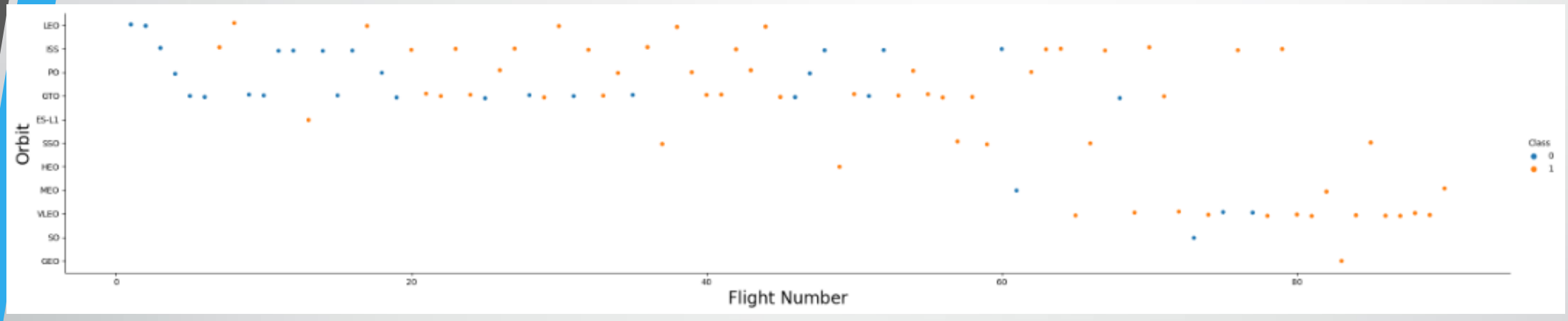
Success Rate vs. Orbit Type



- ES-L1, GEO, HEO and SSO have 100% success rate.
- SO have 0% success rate.
- ISS, LEO, MEO and PO haves success rates between 60% and 70%.

Results: Visualization

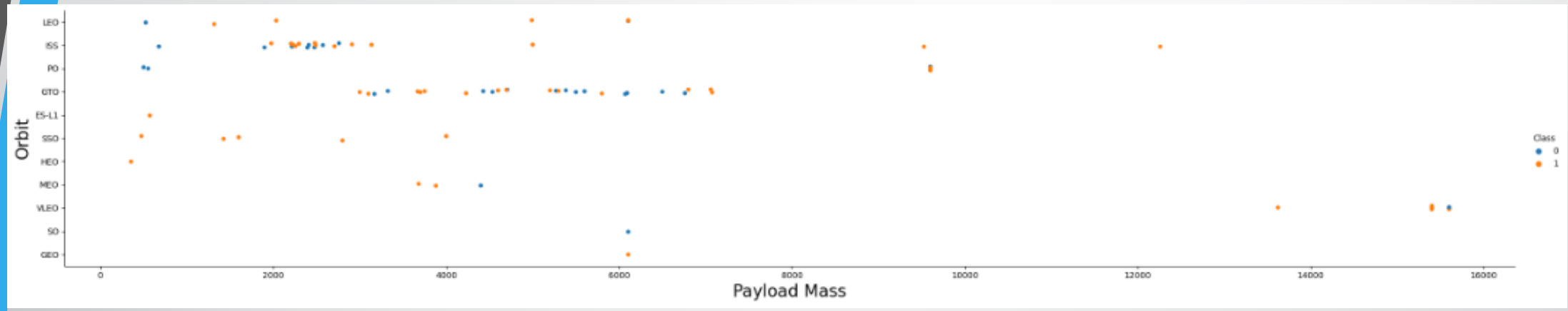
Flight Number vs. Orbit Type



- VLEO have a recent high frequency of flights.
- LEO orbit the Success appears related to the number of flights.
- GTO shows no relationship between frequency and success.

Results: Visualization

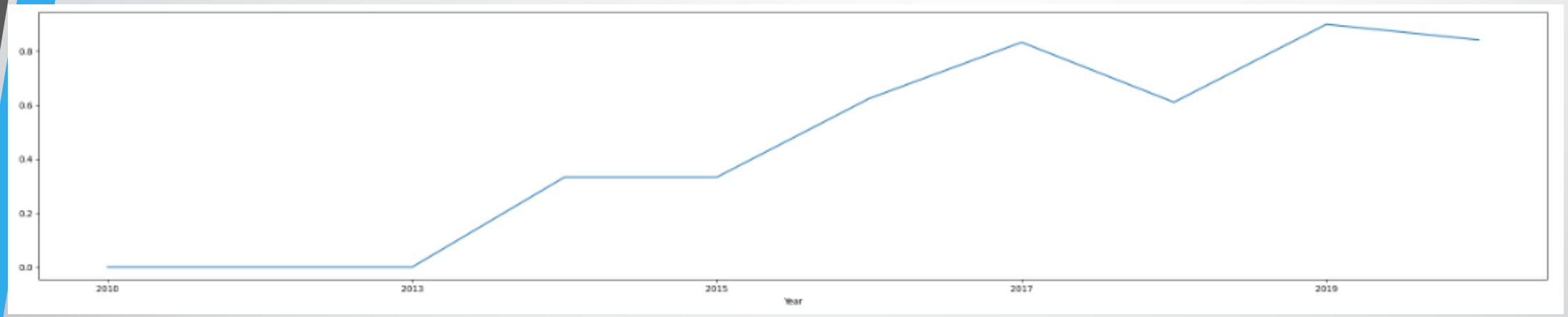
Payload Mass vs. Orbit Type



- GTO shows no relationship between payload and success.
- Heavier payloads have better success rates.
- ISS shows the widest range of payloads and success rates.

Results: Visualization

Launch Success Yearly Trend



- Success have increased since 2013.
- 2015 shows the beginning of an important improvement in success rate for the following years.
- 2018 shows a small step back in success rate.

Results: SQL

Launch Site Names

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEX;
```

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

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

Results: SQL

5 records where launch sites begin with 'CCA'

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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Results: SQL

Total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEX WHERE PAYLOAD LIKE '%CRS%';
```

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

TOTAL_PAYLOAD
111268

Results: SQL

Average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE_PAYLOAD FROM SPACEX WHERE BOOSTER_VERSION LIKE '%F9 v1.1%';
```

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

AVERAGE_PAYLOAD
2534.6666666666665

Results: SQL

Date of the first successful landing in ground pad

```
%sql SELECT MIN(DATE) AS FIRST_SUCCESS FROM SPACEX WHERE "Landing _Outcome" = 'Success (ground pad)';
```

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

FIRST_SUCCESS
01-05-2017

Results: SQL

Names of boosters which have success in drone ship landing and payload mass between 4000 kg and 6000 kg

```
%sql SELECT DISTINCT(BOOSTER_VERSION) FROM SPACEX WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND "Landing _Outcome" = 'Success (drone ship)';
```

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

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

Results: SQL

Total number of success and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) AS TOTAL FROM SPACEX GROUP BY MISSION_OUTCOME;
```

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

Mission_Outcome	TOTAL
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Results: SQL

Names of booster versions which have carried the maximum payload mass

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEX WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEX) ORDER BY BOOSTER_VERSION;
```

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

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

Results: SQL

Failure landing outcomes with drone ship and booster versions during 2015

```
%sql SELECT "Landing _Outcome", BOOSTER_VERSION FROM SPACEX WHERE "Landing _Outcome" = 'Failure (drone ship)' AND SUBSTR(DATE, 7, 4)='2015';
```

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

Landing _Outcome	Booster_Version
Failure (drone ship)	F9 v1.1 B1012
Failure (drone ship)	F9 v1.1 B1015

Results: SQL

Rank of successful landings between 04-06-2010 and 20-03-2017

```
%sql SELECT "Landing _Outcome", COUNT(*) AS TOTAL FROM SPACEX WHERE DATE BETWEEN '04-06-2010' AND '20-03-2017' AND "Landing _Outcome" LIKE '%Success%'
```

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

Landing _Outcome	TOTAL
Success	20
Success (drone ship)	8
Success (ground pad)	6

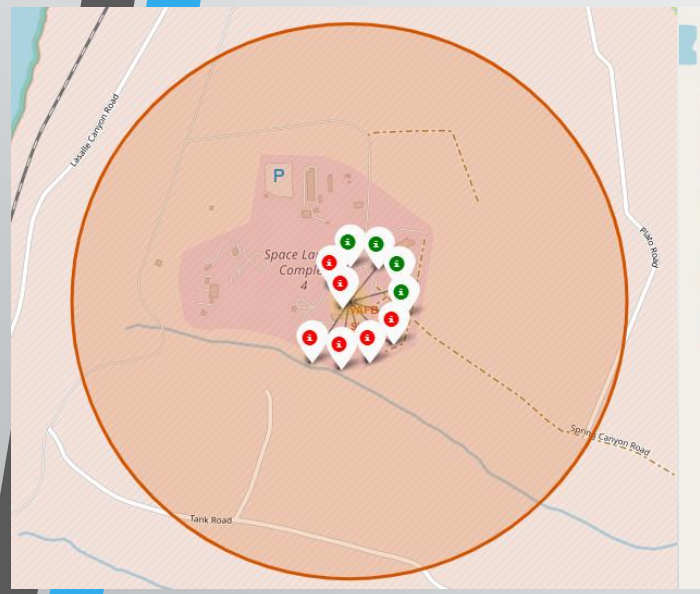
Results: Folium Map

All launch sites

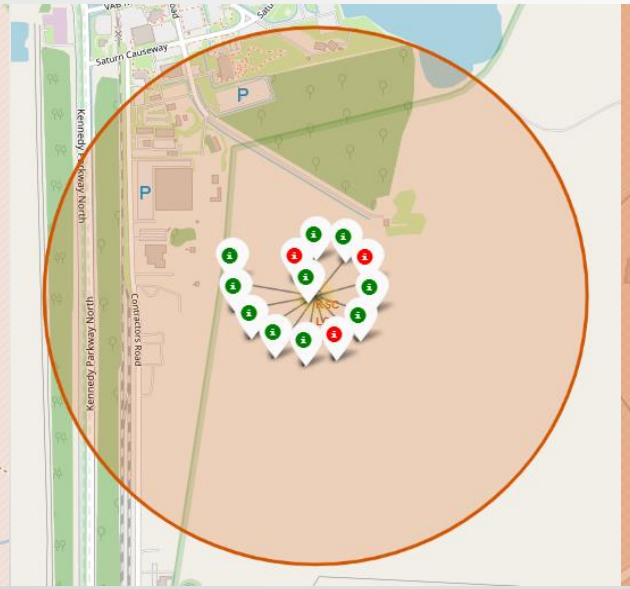
- If a spacecraft is launched from a site near Earth's equator, it can take optimum advantage of the Earth's rotational speed.
- Launch sites locate near the sea to avoid populated areas.

Results: Folium Map

Launch Outcomes

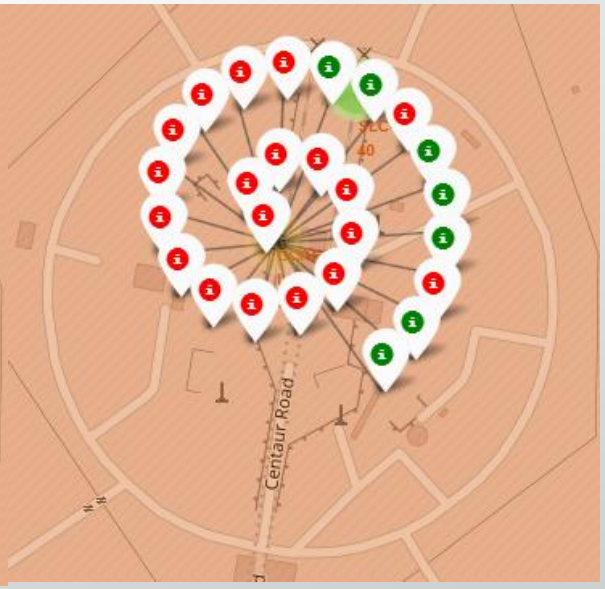


VAFB SLC-4E



KSC LC-39A

CCAFS SLC-40

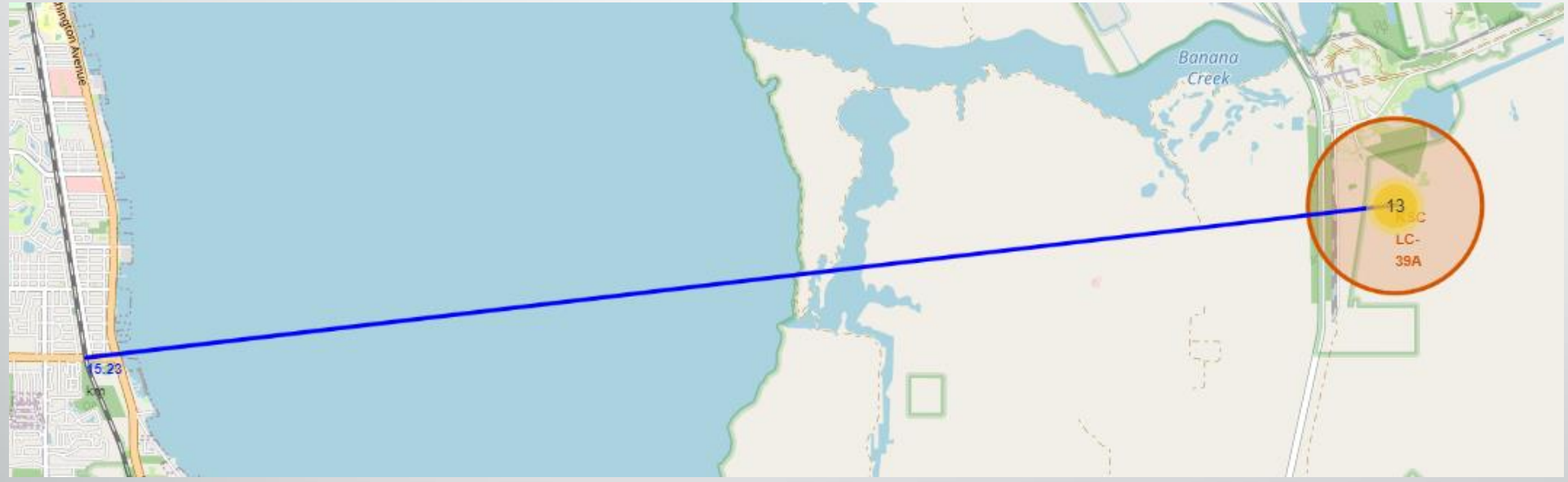


CCAFS LC-40

- Red markers show failed launches and green markers show successful launches.

Results: Folium Map

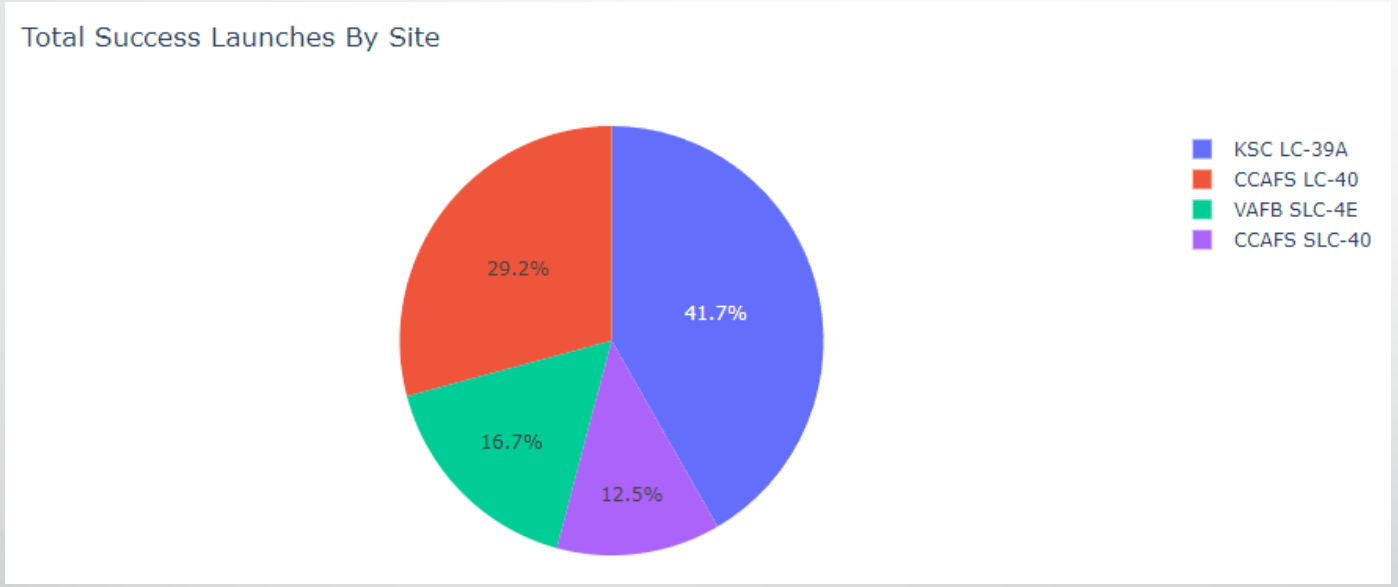
Safety



- In the particular case of KSC LC-39A, the distance between the launch site and a populated area is about 15 km.

Results: Dashboard

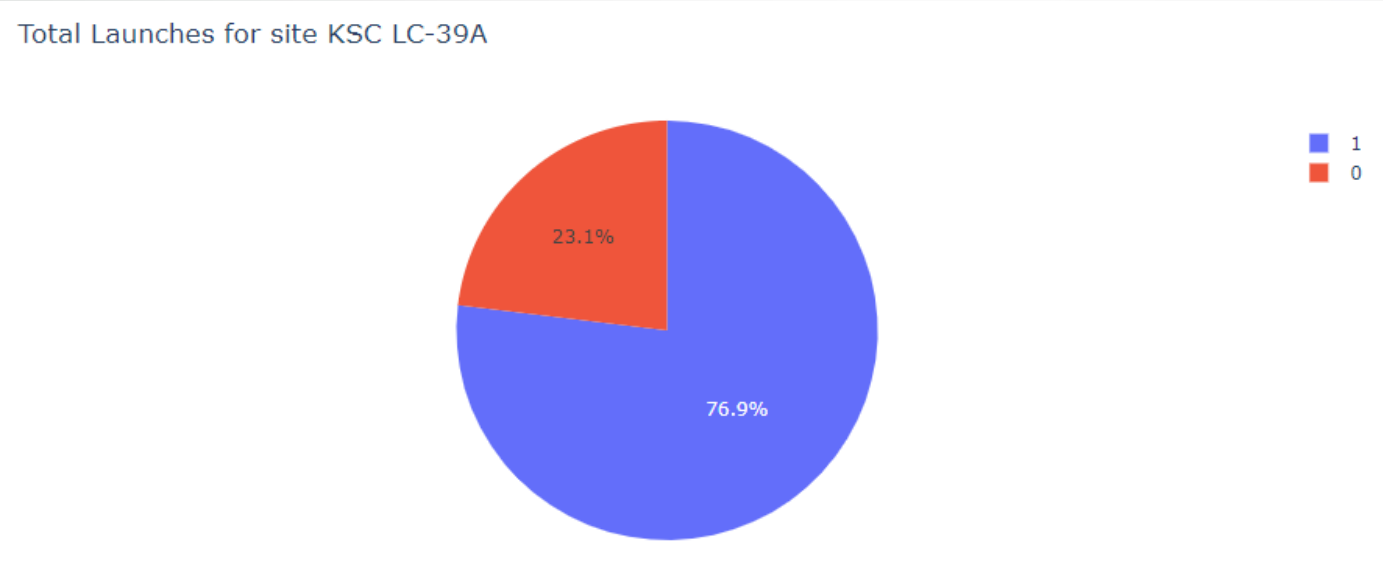
Successful launches by site



- KSC LC-39A performs the most successful launches.

Results: Dashboard

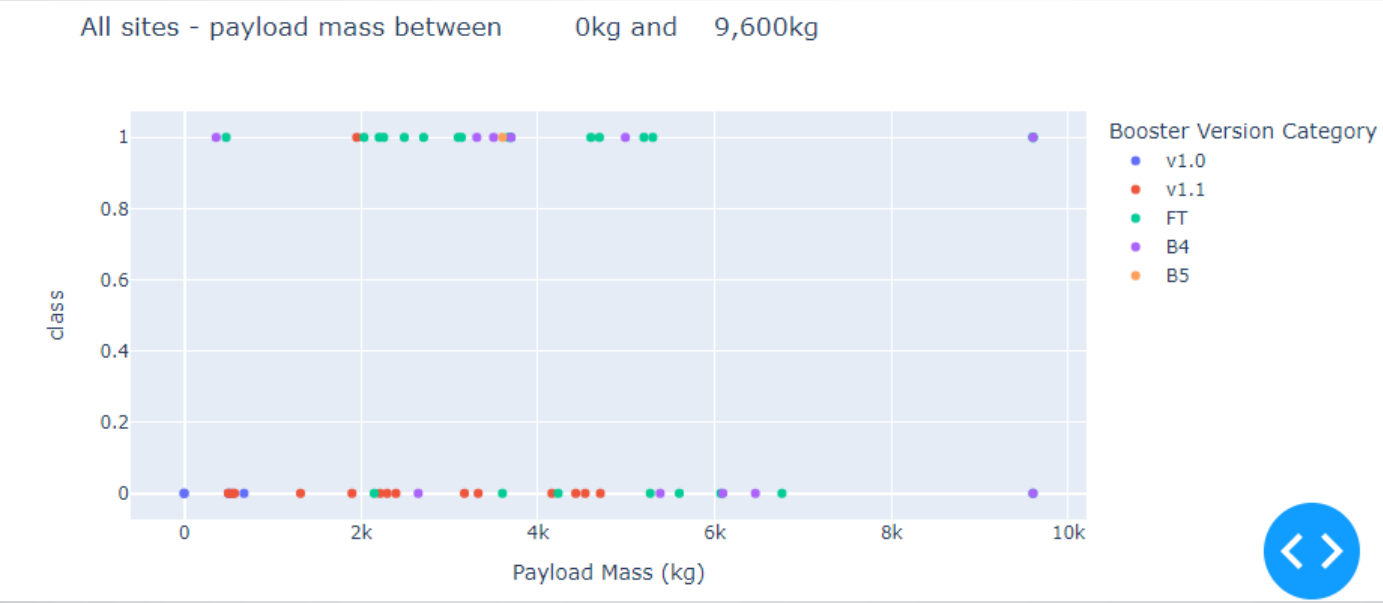
Total launches in KSC LC-39A



- KSC LC-39A counts with a success rate of ~77%.

Results: Dashboard

Payload mass vs. Class



- Payloads under 6000 kg have a high success rate.
- Payloads over 6000 kg requires more data to estimate risk.

Results: Classification

Scores for test set

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.846154	0.800000
F1_Score	0.888889	0.888889	0.916667	0.888889
Accuracy	0.833333	0.833333	0.888889	0.833333

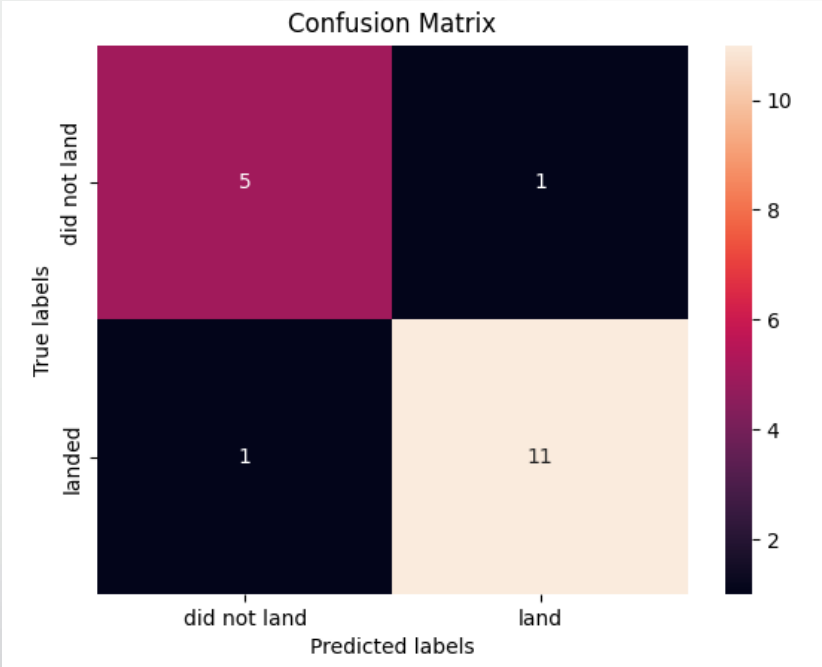
Scores for whole data set

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.841270	0.819444
F1_Score	0.909091	0.916031	0.913793	0.900763
Accuracy	0.866667	0.877778	0.888889	0.855556

- Decision tree shows to be the best classification model, both in the test set and in the whole data set.

Results: Classification

Confusion Matrix for Decision Tree



- The confusion matrix shows a high rate of true positives and true negatives.

Conclusions

- The site with the most success launches is KSC LC-39A.
- No good estimations can be made with the actual information for launches with payloads greater than 6000 kg.
- Success rate for all type of orbits have improved over time.
- Decision tree is the most optimal classification model in this case.
- ES-L1, GEO, HEO and SSO orbits have 100% success rate.

Appendix

- Folium interactive maps won't show on GitHub.
- Data set links used for the dashboard are commented in the spacex_dash_app.py code.

Thanks

- To IBM and the instructors.
- To Coursera.