IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Reinig, V.
25.08.2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies
  - Completing the Data Collection API
  - Completing the Data Collection with Web Scraping
  - Data Wrangling
  - Completing the EDA with SQL
  - EDA with Visualization
  - Interactive Visual Analytics with Folium
  - Interactive Visual Analytics with Dashboard
  - Completing the Machine Learning Prediction

Introduction

- The commercial space age is here. Companies are making space travel affordable for everyone. SpaceX is perhaps the most successful one.
- SpaceX rocket launches are relatively inexpensive. They advertise the cost of a Falcon 9 rocket launch with 62 million dollars; other providers cost upwards to 165 million dollars. This is possible because SpaceX can reuse the first stage.
- If we can determine if the first stage will land, we can determine the cost of a launch. Therefore, we will use public SpaceX data to train a machine learning model.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  - GET-Requests to SpaceX API

  - Web Scraping with Wikipedia

- Perform data wrangling

  - Clean the data

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

  - Searching for the best Machine Learning model

Data Collection

The Data sets are collected by:

- SpaceX API requests
- Web Scraping

Copy the URL of the page you want to analyze

Use GET-Request to collect and parse the SpaceX launch data

Decode the response as JSON and turn it into a data frame

Use the API to get information about the launches based on their ID

Filter the data to only include Falcon 9 launches and replace NULL-Values

Data Collection – SpaceX API

GitHub URL of the completed SpaceX API Calls notebook:

https://github.com/vire7/SpaceX-Falcon9-DataScience-Capstone/blob/07f66dd9d4659ddf0c5551c4492bdaaff1316ff7/spacex-data-collection-api.ipynb

1. Make request to SpaceX API
2. Decode response to JSON
3. Turn JSON into data frame
4. Filter the information about the launches based on their ID
5. Construct the data set

Data Collection - Scraping

GitHub URL of the completed SpaceX Web Scraping notebook:

https://github.com/vire7/SpaceX-Falcon9-DataScience-Capstone/blob/07f66dd9d4659ddf0c5551c4492bdaaff1316ff7/spacex-wikipedia-webscraping.ipynb

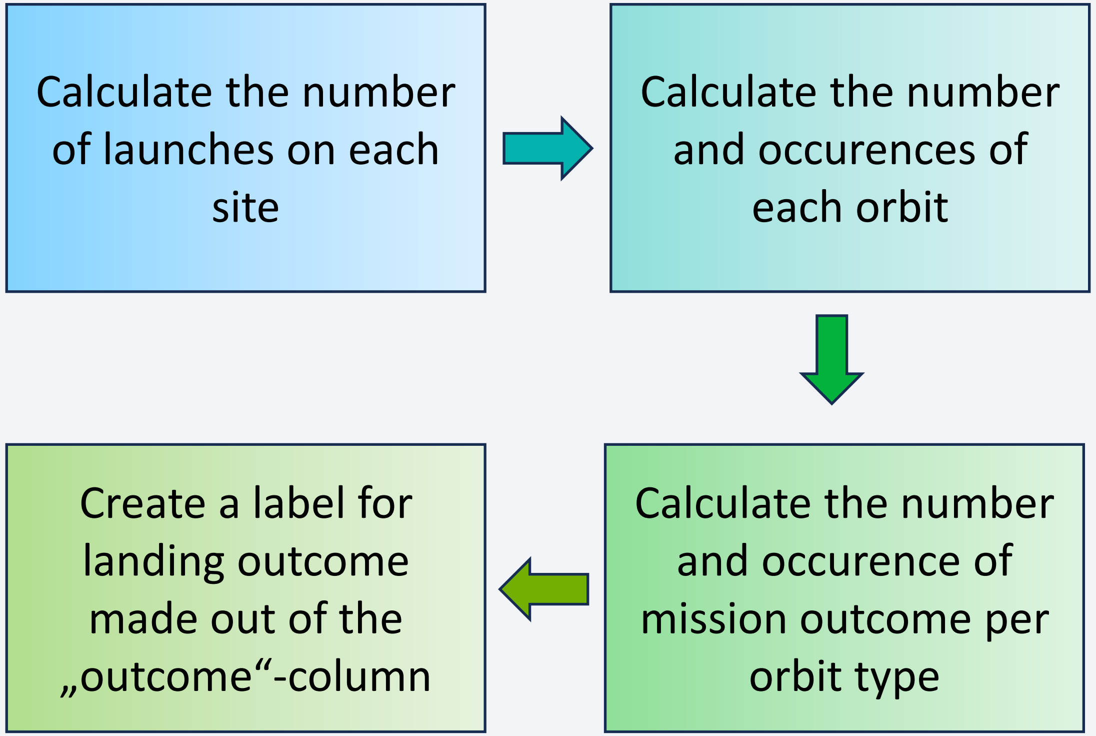
- URL of the Falcon 9 Wikipedia page
- Create a BeautifulSoup from the HTML
- Extract all column names from the HTML table and use them as keys for a dictionary
- Fill up the dictionary with launch data extracted from the rows
- Convert the dictionary into CSV

Data Wrangling

The Data Wrangling process is given in a flow chart.

GitHub URL of the completed Data Wrangling notebook:

https://github.com/vire7/SpaceX-Falcon9-DataScience-Capstone/blob/07f66dd9d4659ddf0c5551c4492bdaaff1316ff7/spacex-data-wrangling.ipynb



EDA with Data Visualization

Type of Charts used:

- **Scatter Plot:** Flight Number vs. Payload Mass, Flight Number vs. Launch Sites, Launch Sites vs. Payload Mass, Orbit vs. Flight Number, Orbit vs. Payload Mass
- **Bar Chart:** Success rate of each orbit
- **Line Plot:** Success rate and Date

GitHub URL of the EDA with Data Visualization notebook:

https://github.com/vire7/SpaceX-Falcon9-DataScience-Capstone/blob/07f66dd9d4659ddf0c5551c4492bdaaff1316ff7/eda-data-visualization.ipynb

EDA with SQL

Used SQL Queries:

- Display the names of unique launch sites and 5 records where launch sites begin with "CCA"
- Display the sum of the payload mass by boosters launched by NASA and the average payload mass carried by booster version F9 v1.1
- List the date of the first successful landing outcome in ground pad
- List the names of the boosters with success in drone ship and a payload mass between 4001 and 5999
- List the total number of successful and failure mission outcomes
- List all booster_versions that carried max payload mass, using a subquery with aggregate function
- List records which will display the month names, failure landing_outcomes in drone ship, booster_versions and launch site for 2015
- Rank the count of landing outcomes between 2010-06-04 and 2017-03-20, in descending order

https://github.com/vire7/SpaceX-Falcon9-DataScience-Capstone/blob/07f66dd9d4659ddf0c5551c4492bdaaff1316ff7/eda-sql.ipynb

Build an Interactive Map with Folium

Folium markers were used to show the SpaceX launch sites and their nearest important landmarks like railways, highways, cities and coastlines.

Polylines were used to connect the launch sites to their nearest land marks.

Green Markers represent successful launches

Red Markers represent rocket launch failures

GitHub URL for the Interactive Visual Analytics with Folium notebook:

https://github.com/vire7/SpaceX-Falcon9-DataScience-Capstone/blob/07f66dd9d4659ddf0c5551c4492bdaaff1316ff7/visual-analytics-with-folium.ipynb

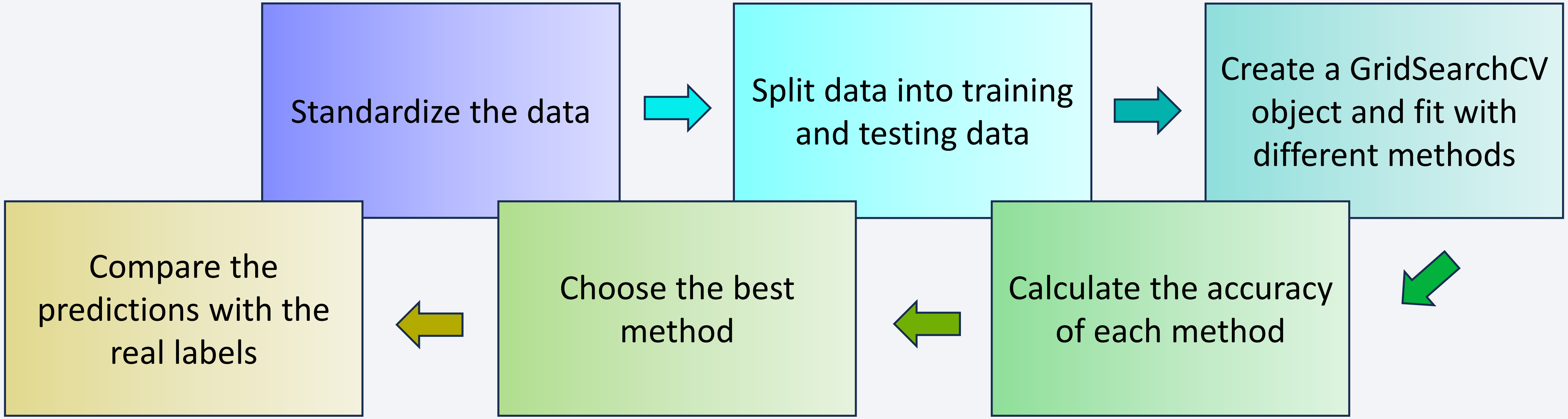
Build a Dashboard with Plotly Dash

Pie Charts and Scatter Plots were used to visualize the launch records of SpaceX.

- These charts were used to display success rate per launch site.
- Successful launches were represented with “1”, failures with “0”
- We were able to understand the factors that may influence the success rate of each site.

Predictive Analysis (Classification)

- First, we standardized the data and split them into training and testing data.
- Then build a model for each method (Logistic Regression, Support Vector Machine, Decision Tree Classifier and k-nearest neighbor)
- Followed by the calculation of the accuracy and evaluating the confusion matrix for each method (https://github.com/vire7/SpaceX-Falcon9-DataScience-Capstone/blob/07f66dd9d4659ddf0c5551c4492bdaaff1316ff7/spacex-MachineLearning-prediction.ipynb)



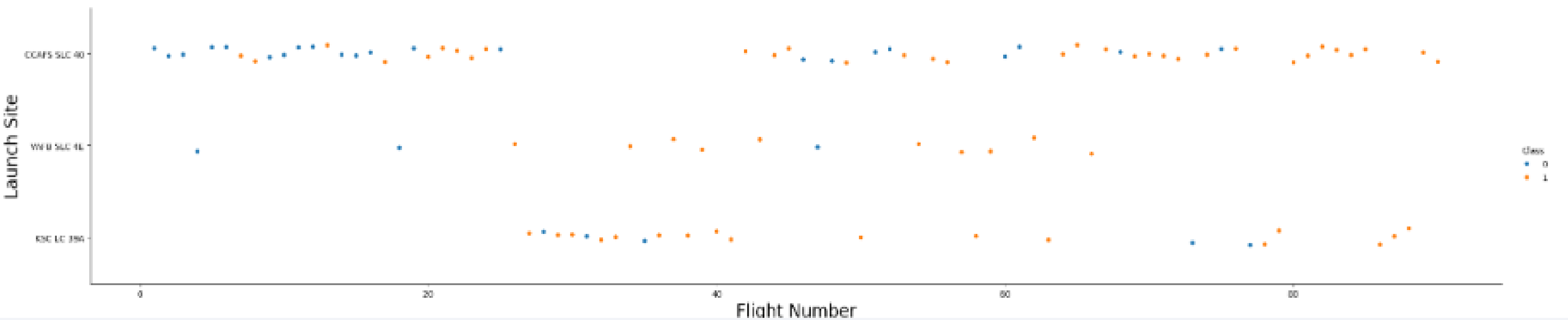
Results

- The exploratory data analysis shows that successful landing outcomes are somewhat correlated with the flight number. The successful landing outcomes have had a significant increase since 2015.
- All launch sites are located near the coast line. Perhaps, this makes it easier to test rocket landing in bodies of water.
- The launch sites are all located near highways and railways. This may facilitate better transportation of equipment and other materials.
- The machine learning models were able to predict a successful landing with an accuracy score of 83.33%

Section 2

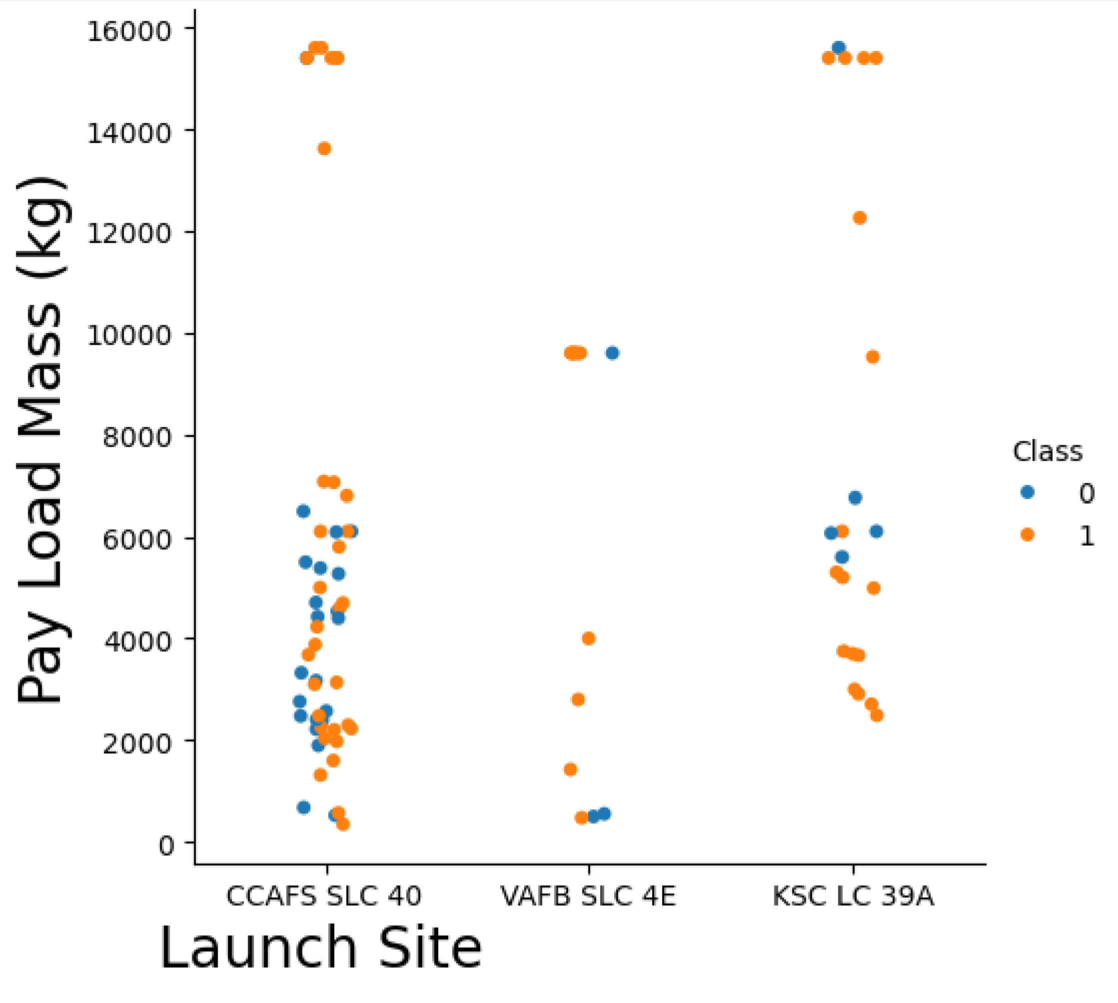
Insights drawn from EDA

Flight Number vs. Launch Site



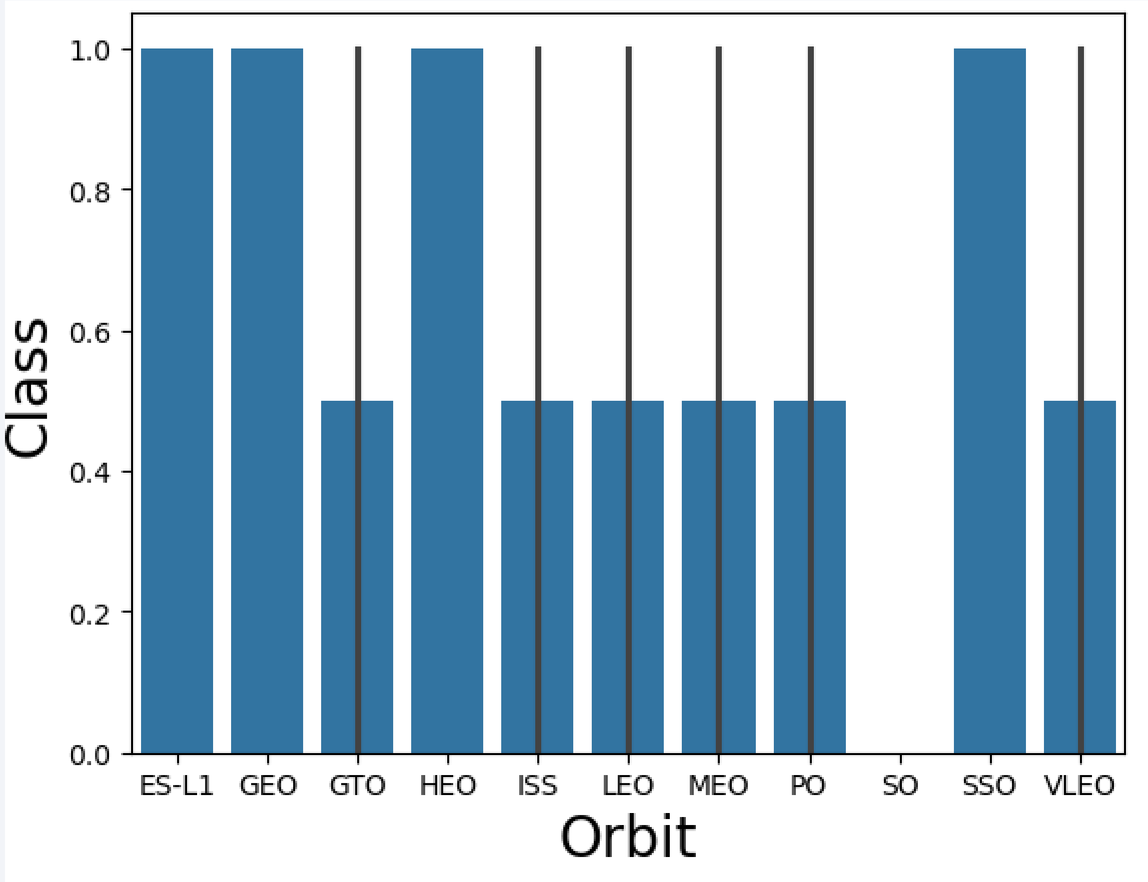
- The success rate increases the higher the flight numbers are.
- Launch site „CCAFS SLC 40“ has the highest amount of landings.

Payload vs. Launch Site



- No rockets with heavy payload mass (>10000) were launched from the „VAFB-SLC“ launch site

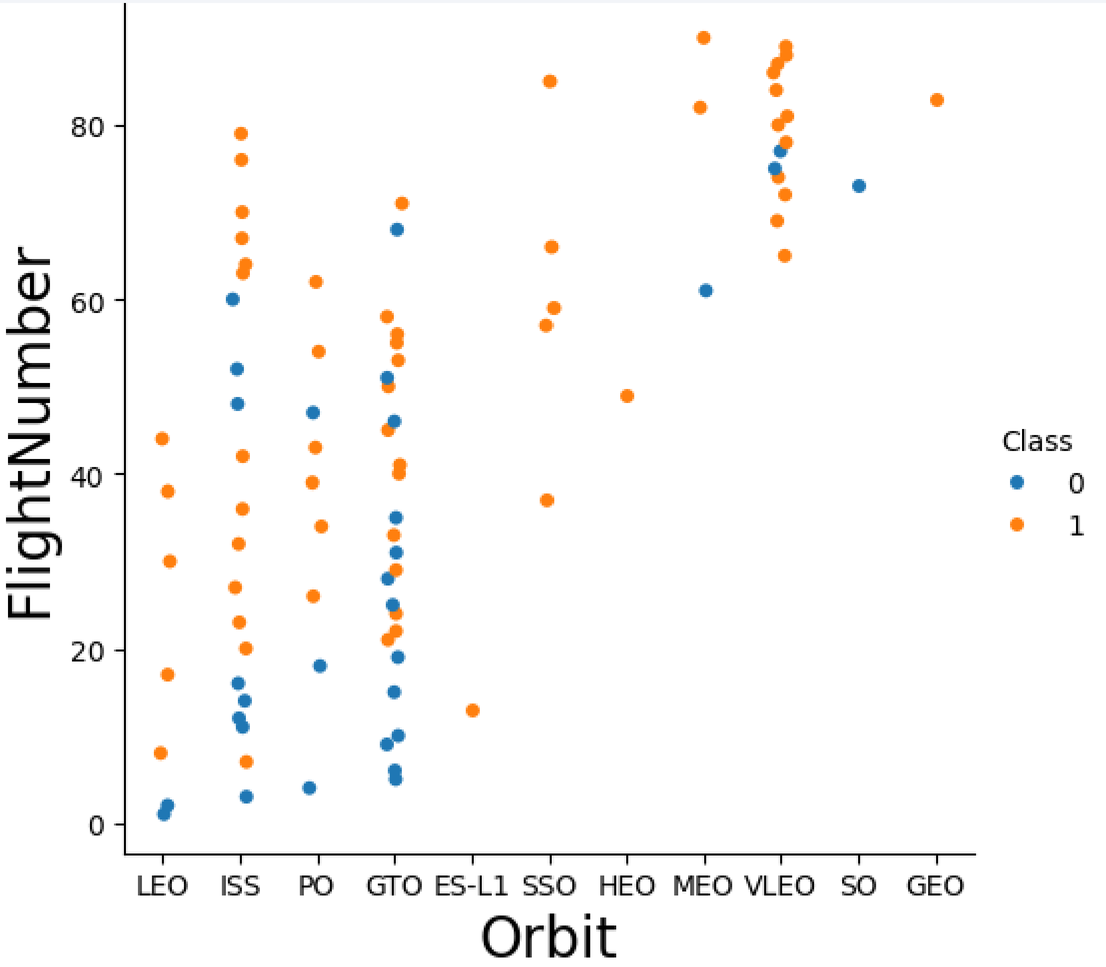
Success Rate vs. Orbit Type



The orbits with the highest success rate are:

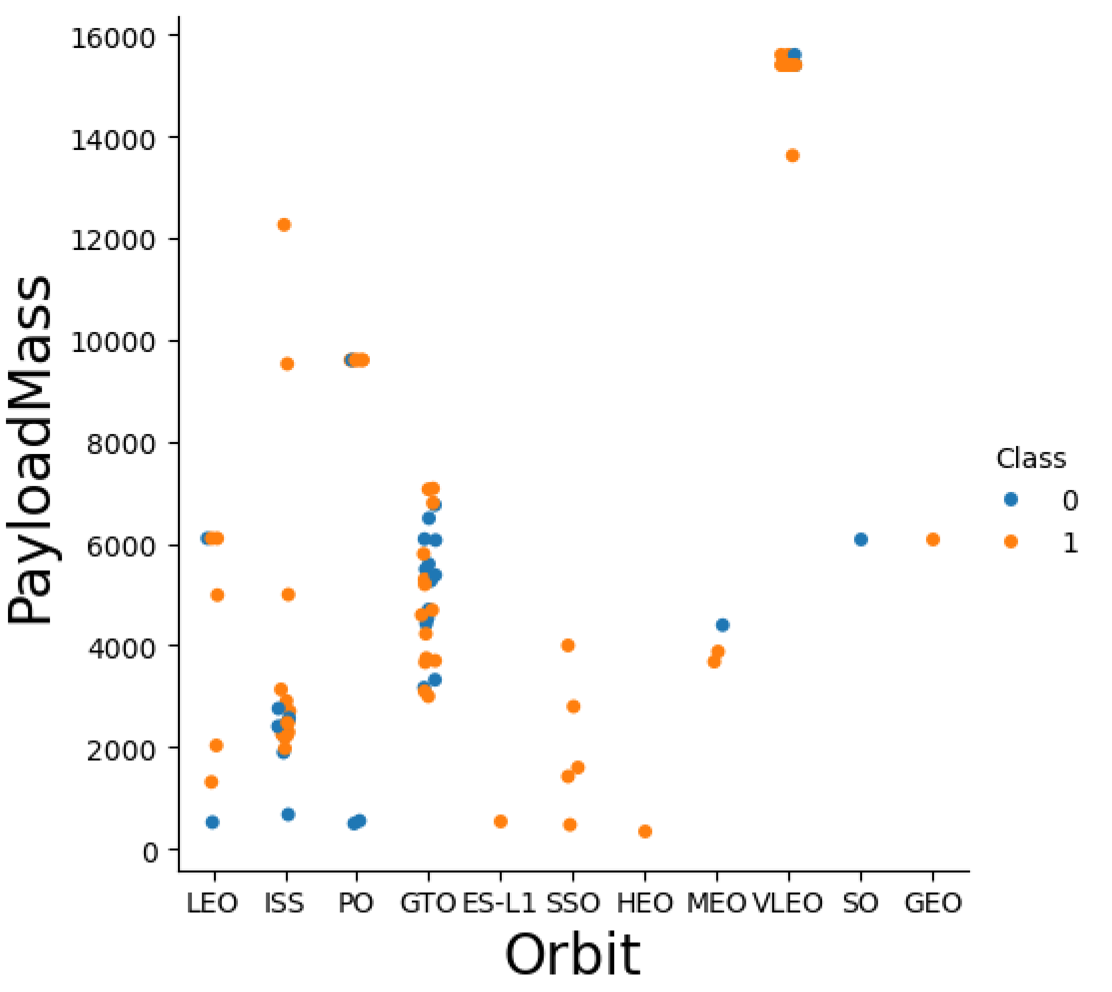
- ES-L1
- GEO
- HEO
- SSO

Flight Number vs. Orbit Type



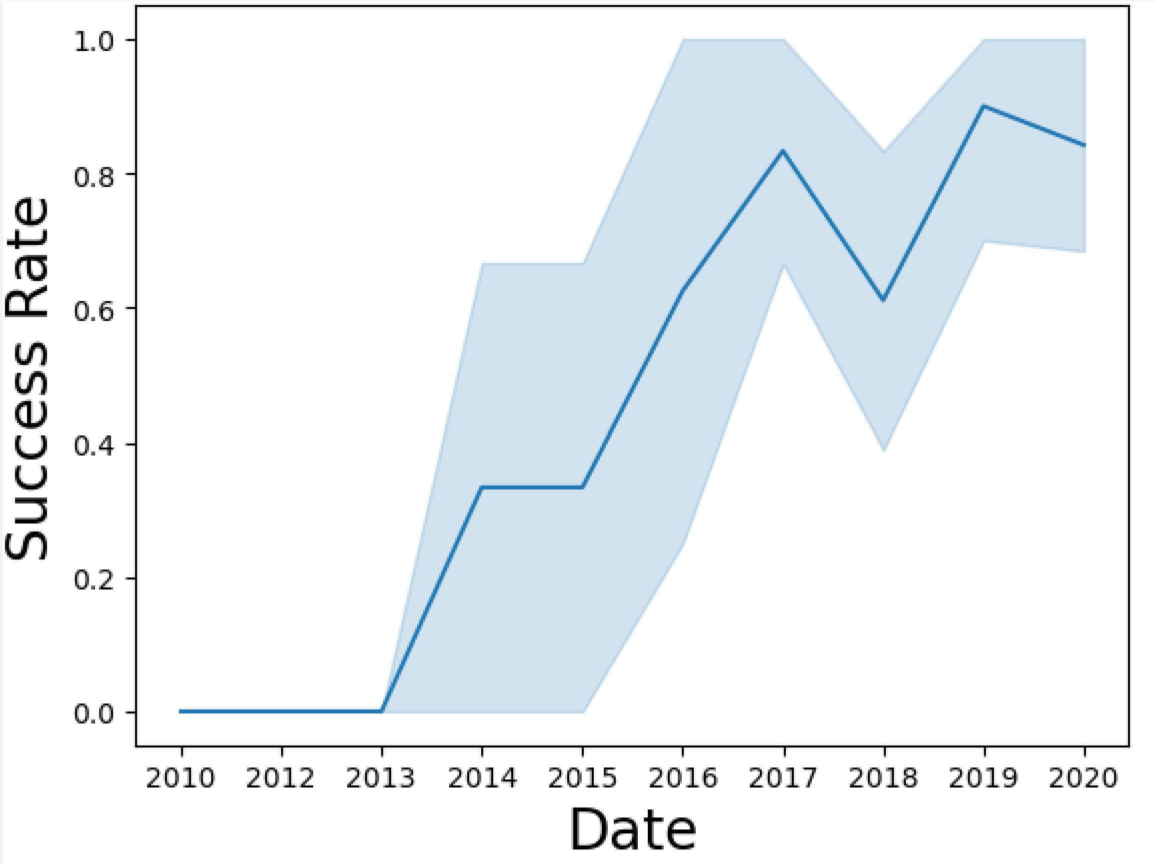
- The success rate of SEO seems to be related to the number of flights
- The success rate of the GTO orbit seems to have no relationship to the number of flights

Payload vs. Orbit Type



- Polar, LEO and ISS have a higher successrate for heavy payload
- GTO has nearly the same success rate as failure rate

Launch Success Yearly Trend



The success rate increased significantly between 2013 and 2020.

All Launch Site Names

Given the data, these are the names of the launch sites where different rocket landings where attempted:

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

Launch Site Names Beginning with 'CCA'

```
%sql SELECT * FROM SPACEXTBL 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
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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

There are 5 records where the launch site name begins with „CCA“.
As seen in the picture, there are other organizations besides SpaceX testing their rockets.

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER like 'NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

The picture displays the total payload mass by boosters launched by NASA.

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

The average payload mass carried by F9 v1.1 was 2928.4kg.

First Successful Ground Landing Date

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)';
```

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

min(DATE)
2015-12-22

The first successful ground pad landing was on the 22nd December 2015.

Successful Drone Ship Landing with Payload Mass between 4000kg and 6000kg

```
%sql SELECT BOOSTER_VERSION from SPACEXTBL WHERE LANDING_OUTCOME = 'Success (drone ship)' and  PAYLOAD_MASS__KG_  >4000 and PAYLOAD_MASS__KG_ <6000;
```

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

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

There are only 4 Boosters with a payload mass between 4000kg and 6000kg:

- F9 FT 81022
- F9 FT 81026
- F9 FT 81021.2
- F9 FT 81031.2

Total Number of Successful and Failure Mission Outcomes

```
sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'
```

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

count(MISSION_OUTCOME)
99

The total number of successful mission outcomes and failure mission outcomes is 99.

Boosters that Carried Maximum Payload

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

12 boosters have carried the maximum payload mass of 15600kg.

2015 Launch Records

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

2 boosters failed to land in 2015:

- F9 v1.1 B1012 launched from CCAFS LC-40
- F9 v1.1 B1015 launched from CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

COUNT(landing_outcome)	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
10	2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
5	2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
5	2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
3	2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)
3	2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)	Success	Controlled (ocean)
2	2013-09-29	16:00:00	F9 v1.1 B1003	VAFB SLC-4E	CASSIOPE	500	Polar LEO	MDA	Success	Uncontrolled (ocean)
2	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)

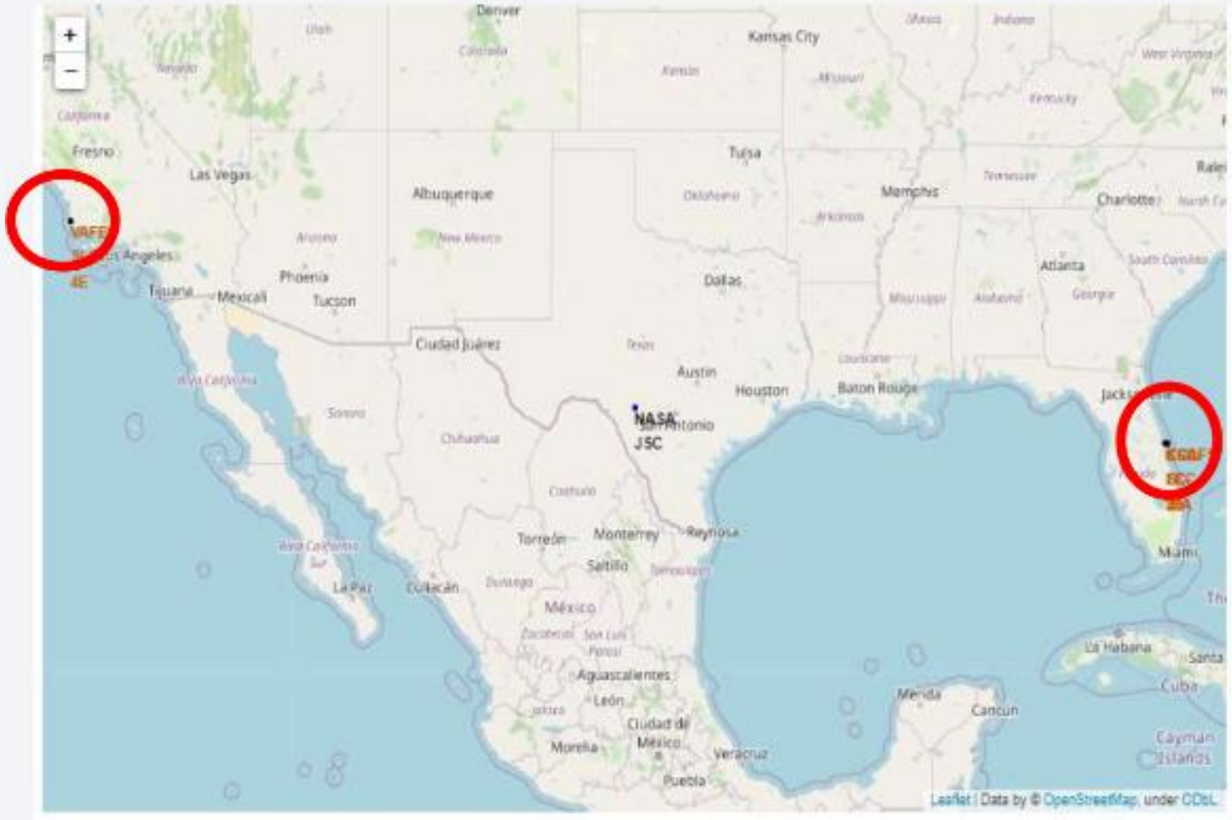
By far the most missions had no landing attempt but were deemed successful (10). Precluded drone ship landings and failed parachute landings occured the least.

Section 3

Launch Sites Proximities Analysis

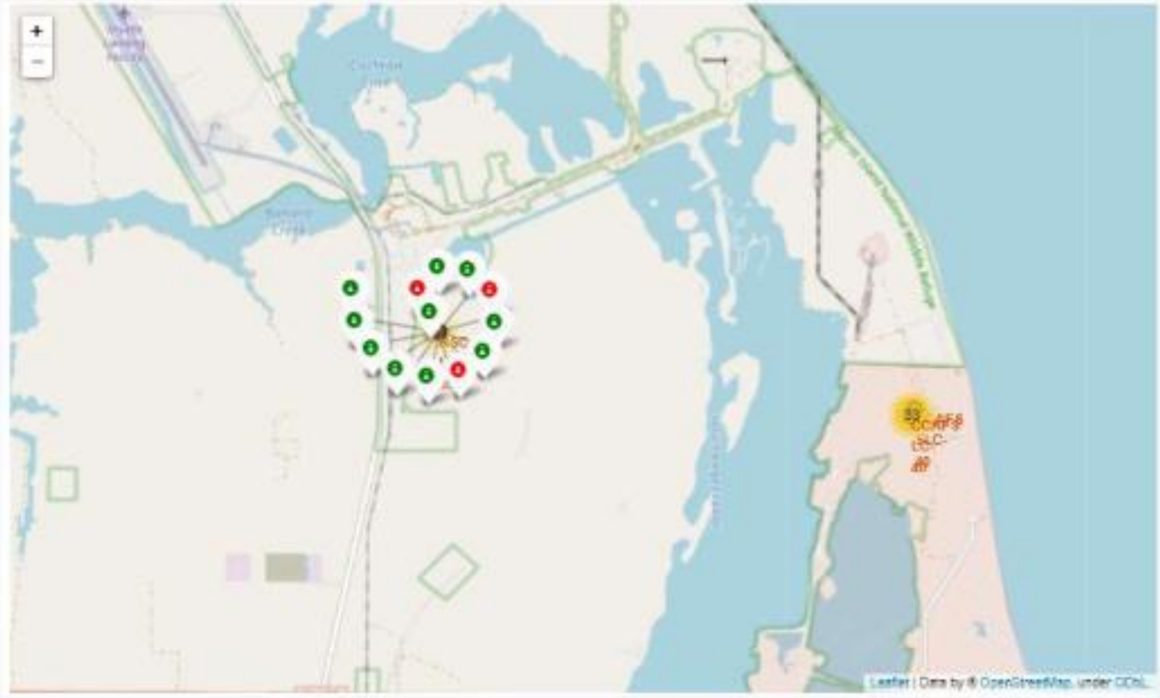
Launch Site Locations

All launch sites are near the coast line and far away from the equator line.



Success Rate of Rocket Launches

Green markers represent successful launches, red markers failed rocket launches.

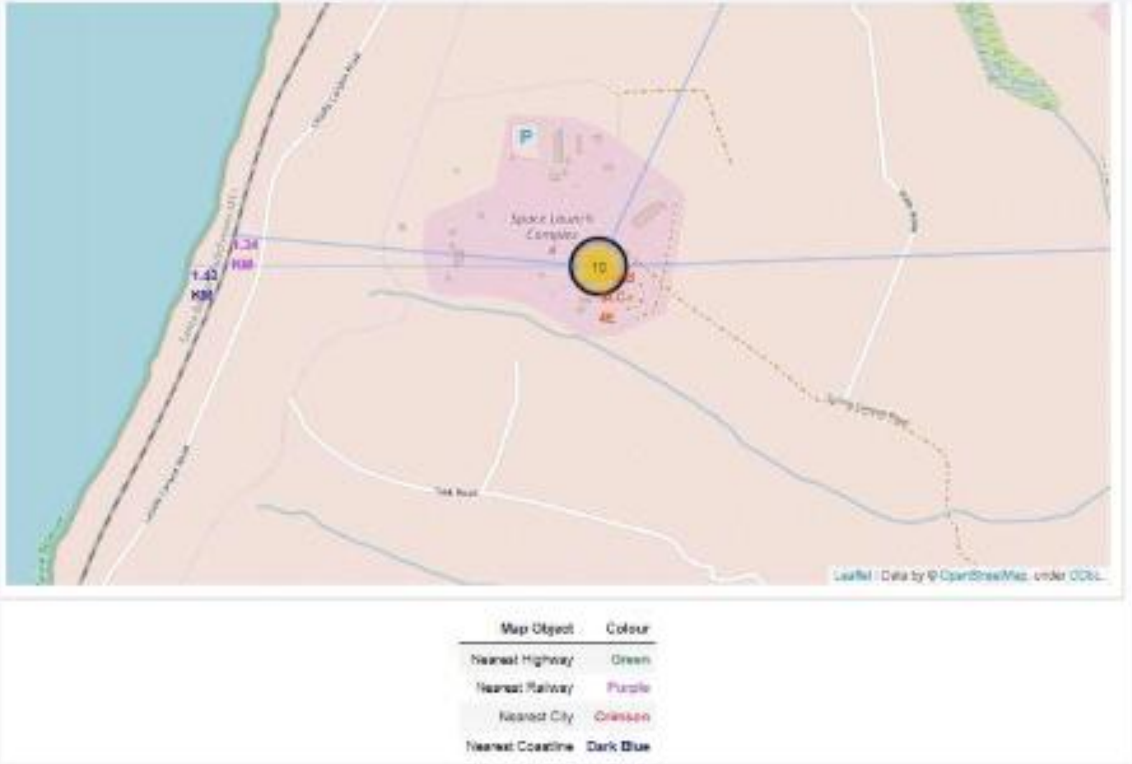


Surrounding Landmarks

Launch sites are usually more than 18 km away from any city, propably to prevent crashes in populated areas.

They are also set up near railways and highways due to the need of transportation of various materials.

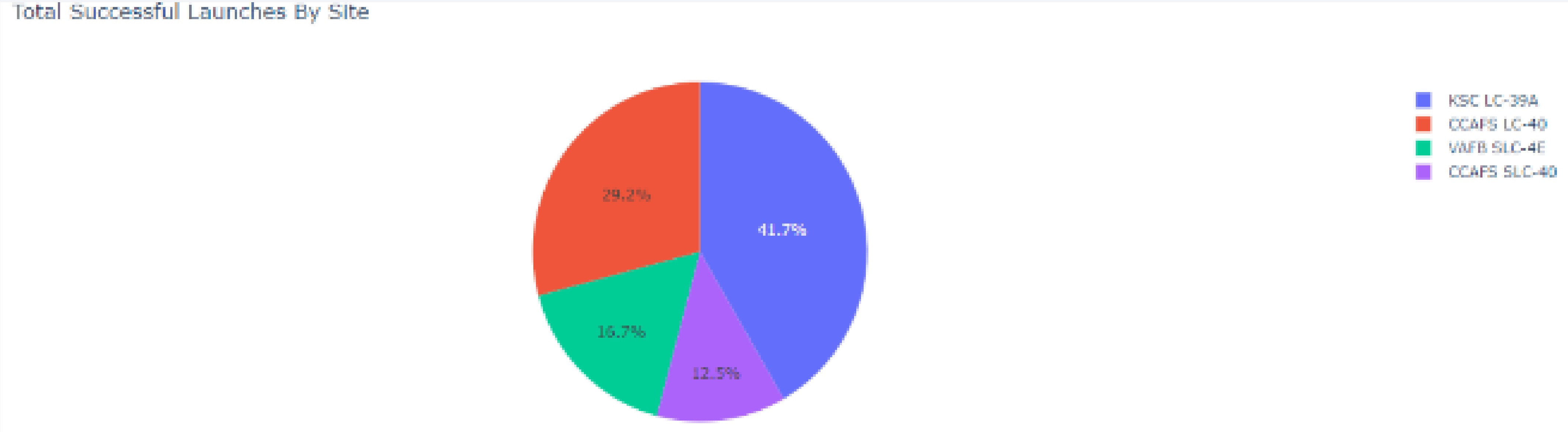
The proximity to the coast line is also notable. This is needed for landing tests near or in a body of water.



Section 4

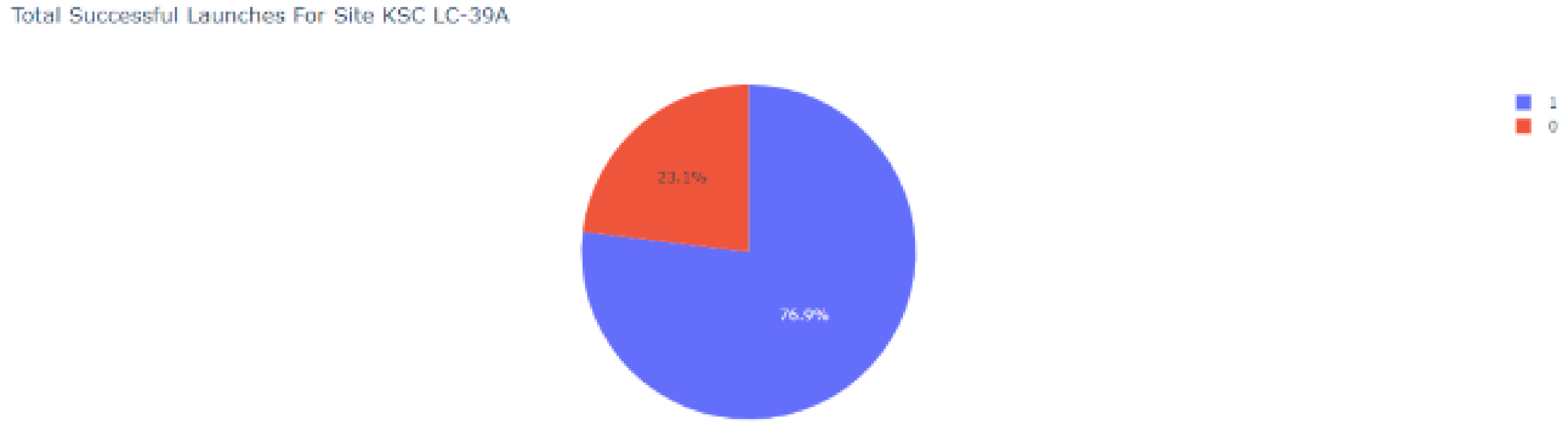
Build a Dashboard with Plotly Dash

Successful Launches by Site



KSC LC-39A has the highest number of successful launches as well as the highest success rate of launches.

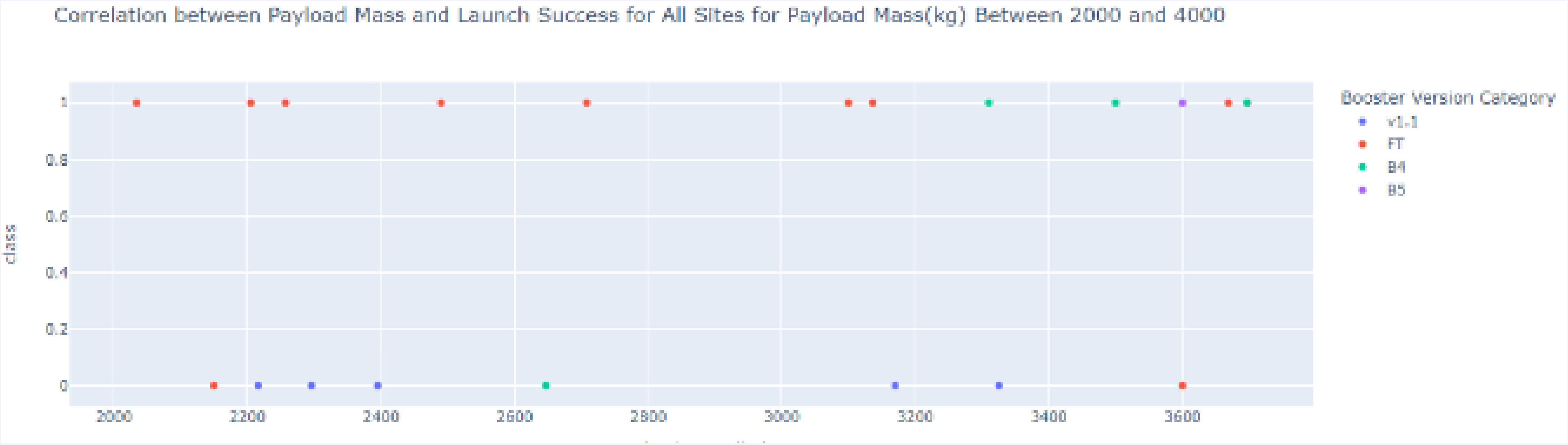
Successful Launches for KSC LC-39A



76,9% of all launches from KSC LC-39A were succesful (1).

➡ highest success rate of all launch sites

Payload Mass vs. Launch Success Rate



The payload range between 2000kg and 4000kg has the highest success rate.

Section 5

Predictive Analysis (Classification)

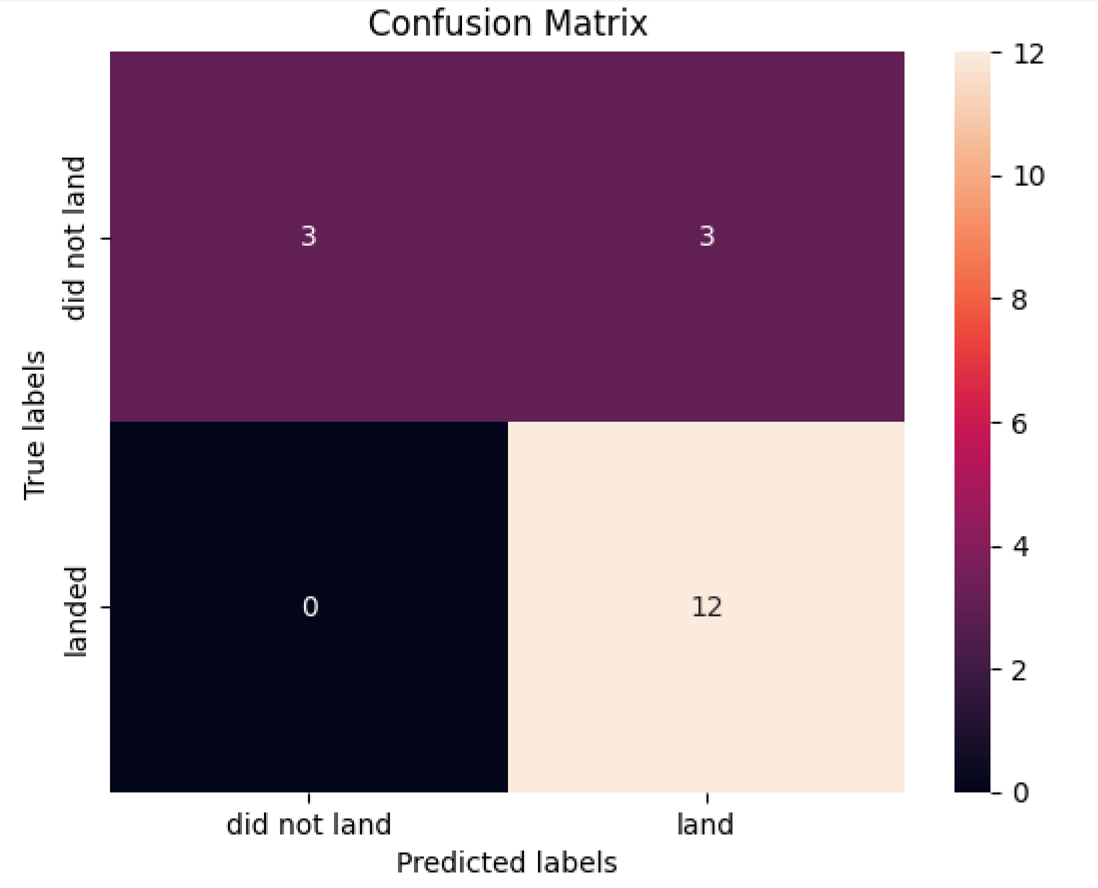
Classification Accuracy

	ML Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

All models have the same Accuracy Score (83.33%).

For the next slides we will use the Logistic Regression model for further visualization.

Confusion Matrix



The chart shows the confusion matrix for The Logistic Regression model.

The model only failed in the prediction of 3 labels.

Conclusions

- All launch sites of SpaceX are located near the coast and in proximity to railways and highways.
- The launch site KSC LC-39A had the highest success rate of all sites.
- The landing success increased with the number of flights.

All the data was used to train a machine learning model that can predict the landing outcome of rocket launches with an accuracy of 83,33%.

Thank you!