IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Andrii Voitik
30/01/2023

Outline

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

Executive Summary

- Summary of methodologies

In the market of space cargo launches, SpaceX is a leader due to the possibility of multiple use of the first stage of the rocket. In order to successfully compete with them, data was collected on the results of the landing of their rockets. This data is used for machine learning to predict the successful landing of the rocket's first stage. This became possible thanks to data collection (API requests and web scraping), their analysis using data visualization, and the application of machine learning algorithms.

- Summary of all results

The created models made it possible to determine the probability of successful landing of the first stage of the rocket depending on the launch parameters. Predictive models will save millions of dollars and attract more customers. Analytics, data visualization, and machine learning tools based on the Python programming language and using appropriate libraries allow you to achieve a high number of successful rocket landings at the level of up to 90% faster. More detailed information can be obtained from the project repository:

https://github.com/Afex81/-SpaceX_Landing_Prediction.git

Introduction

- The goal of this project is to allow the company to successfully compete with SpaceX in the space launch market by reducing their cost through reuse of the rocket's first stage. It was necessary to predict the probability of a successful landing of the first stage of the rocket using data from previous launches, such as the launch site, the weight of the cargo placed into orbit, the type of orbit, etc.
- Accurate prediction of the successful landing of the first stage of the rocket allows its reuse and, accordingly, the reduction of the launch cost from 165 to 62 million dollars (according to the SpaceX company).
- With the help of machine learning algorithms, it is possible to offer customers more attractive offers for the delivery of cargo into the orbit of our planet.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Make requests to the SpaceX AP;
  - Perform web scraping to collect Falcon 9 historical launch records on the Wikipedia page titled: List of Falcon 9 and Falcon Heavy launches.
- Perform data wrangling
  - Clean the data and explore it to find patterns in the data to determine the labels for training supervised models.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Create a machine learning pipeline to predict if the first stage will land given the data.
  - Train the best performing model to make accurate predictions.

Data Collection

An important stage in the creation of machine learning models is data collection. The accuracy of the model's prediction largely depends on the quality of the data.

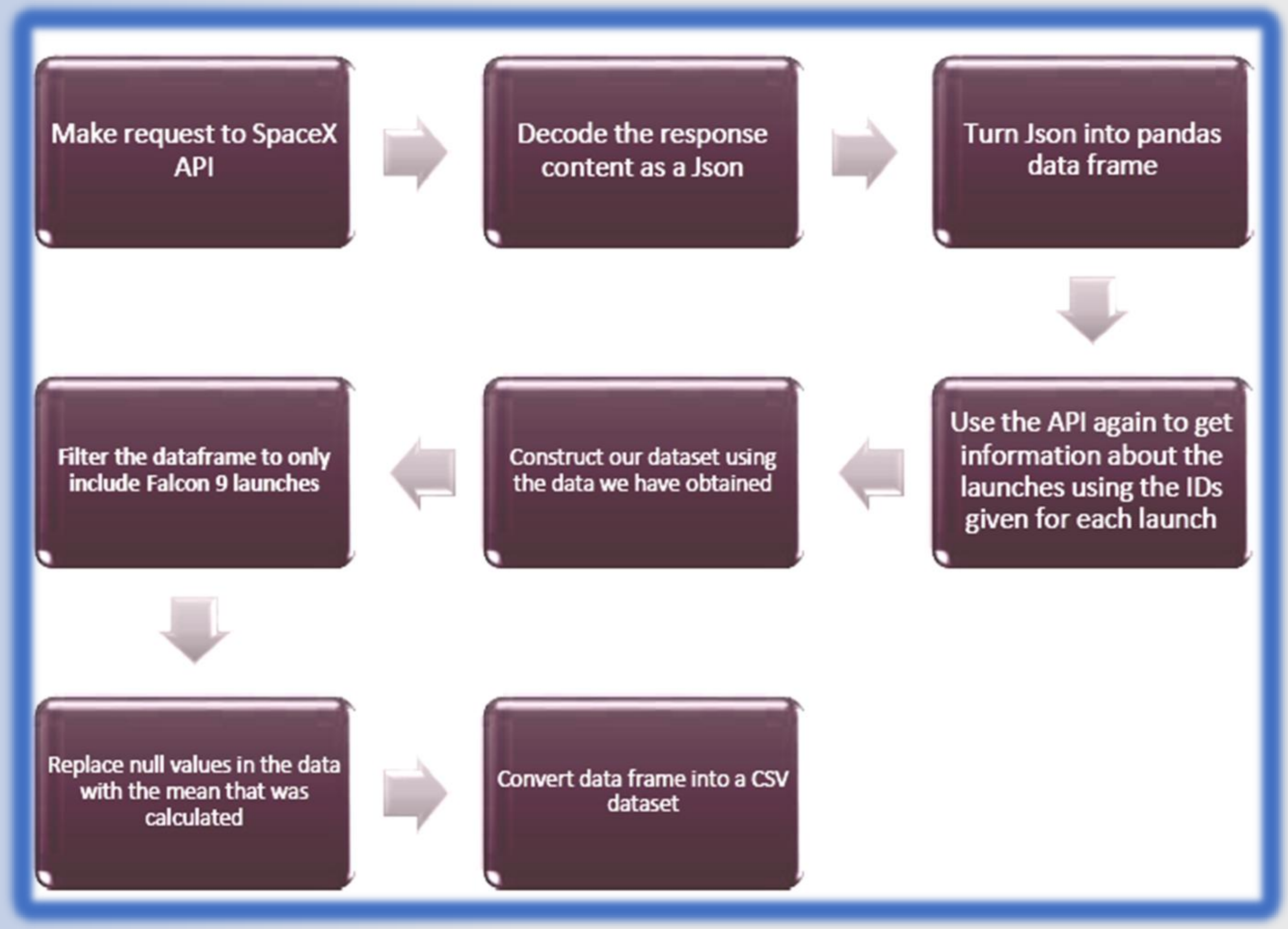
Of the existing methods, we used two methods:

- Collection of data as requested by the SpaceX API.
- Data collection using Web Scraping

These methods do not require significant resources and allow obtaining the required accuracy.

Data Collection – SpaceX API

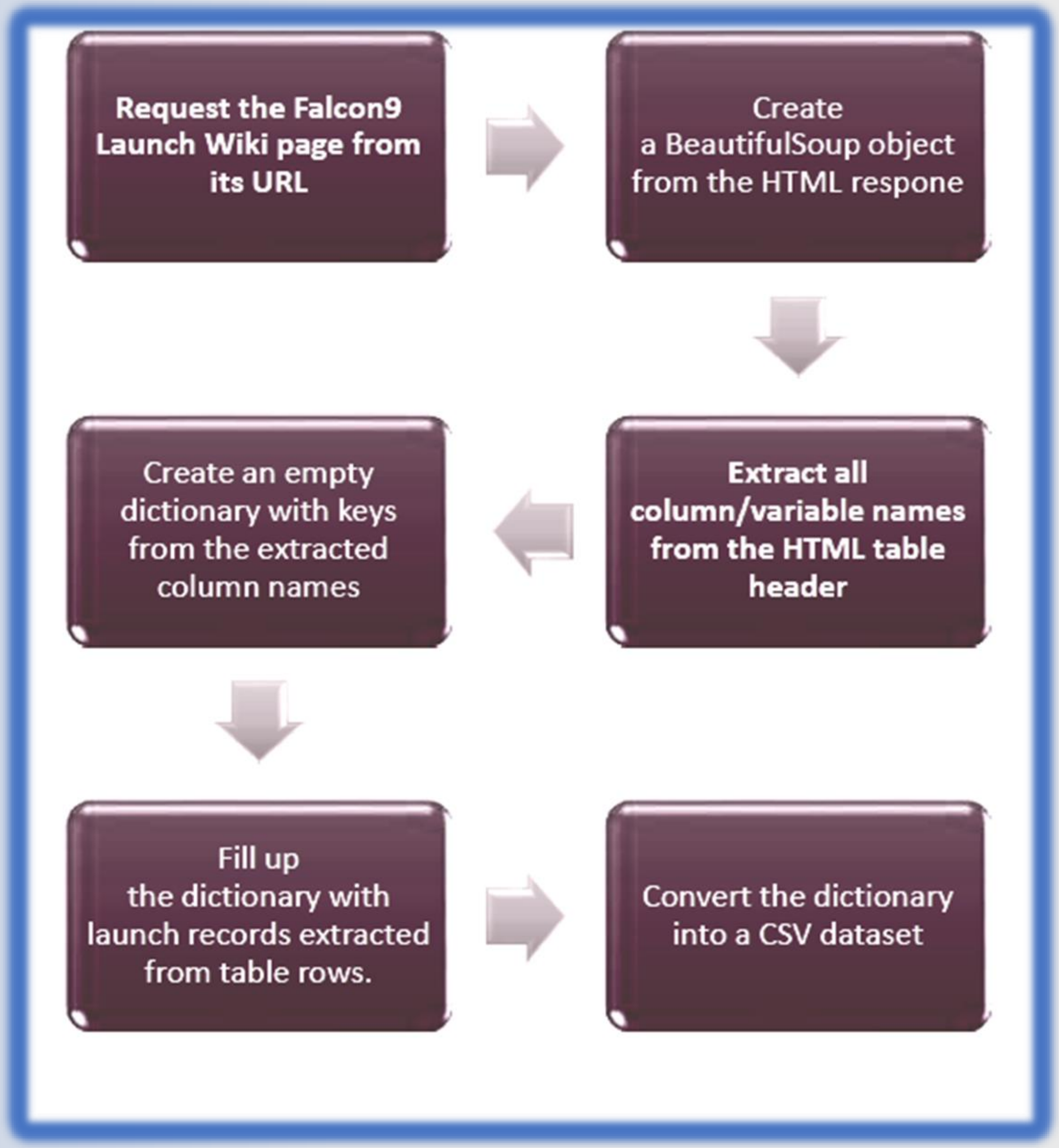
- Make a request to SpaceX API and make sure the data is in the correct format.
- Perform some basic data wrangling and formatting in order to clean the requested data.
- Convert data frame into a CSV dataset.
- URL link: https://github.com/Afex81/-SpaceX_Landing_Prediction/blob/330f5ffef3cba566c7efdf49bbd4afffaf82739e/1_API_request.py



Data Collection - Scraping

- Using BeautifulSoup, perform web scraping on the wikipedia page "List of Falcon 9 and Falcon Heavy launches"
- Store the launch records in an HTML table.
- Convert data it into a CSV dataset.

URL link: https://github.com/Afex81/-SpaceX_Landing_Prediction/blob/5d9491ca9838fe665599a97cb0a0122d7955090b/2_WEB-Scraping.py



Data Wrangling

At this stage, need to determine the dependencies between the data and set the goal for the machine learning algorithm. The data set contains a column that represents a successful or failed landing. These data were converted into labels: 1 - successful landing, 0 - failed landing. It was these labels that were later used for graphical analysis of patterns between the data and model training.

URL link: https://github.com/Afex81/-SpaceX_Landing_Prediction/blob/9fd3b454c85bd2e17d1b19270c3eef831fbb0279/3_Wrangling.py

Determine the number of launches on each facility

Calculate the number and occurrences of each orbit

Calculate the number and occurences of mission outcome per orbit type

Create a landing outcome training label and loop through all the landing outcomes

Create a "Class" column containing the information from the outcome label

Transform the data frame into a CSV dataset.

EDA with Data Visualization

Data visualization allows better understand them, see the dependencies between them, identify trends and data outliers. For visualization, we used several types of charts and graphs.

- Catplots and scatter plots were used to view the relationships of categorical variables like *Launch Site* and *Orbit*.
- A bar chart was used to visualize the success rate of each orbit type.
- A line chart was used to visualize the launch success yearly trend.

URL link: https://github.com/Afex81/-SpaceX_Landing_Prediction/blob/f71648245958308de2601ef1daf86a27df48c70b/4_EDA.py

EDA with SQL

Summary of SQL queries that were used:

Display the names of the unique launch sites in the space mission

Compare the payload mass with boosters launched by NASA (CRS)

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

List the total number of successful and failure mission outcomes

Determine the dates of different landing outcomes

URL link: https://github.com/Afex81/-SpaceX_Landing_Prediction/blob/50390ef0467fa14a12250233b1302e0ed3b7d5d1/5_SQL

Build an Interactive Map with Folium

Folium Markers allowed us to map SpaceX launch sites and nearby important landmarks such as railroads, highways, cities, and coastlines. With the help of lines, these landmarks were connected to the missile launch sites, and the circular zones of the launch sites were also highlighted. Clusters were created to mark successful/failed launches for each site. At the same time, red means failures when launching a rocket, and green symbolizes successes.

URL link: https://github.com/Afex81/-SpaceX_Landing_Prediction/blob/17e148f16cf230cb1457cd87530d10b2d7a20988/6_Folium.py

Build a Dashboard with Plotly Dash

Pie charts and dot charts were used to visualize launch data, showing the percentage of successful rocket first stage landings at each launch site, respectively. The relationship between a successful landing and the payload delivered by the rocket is also shown in the form of a linear graph.

URL link: https://github.com/Afex81/-SpaceX_Landing_Prediction/blob/2a72aed7882b1a23eecf7a89ef0932f3069ba86f/7_DASH.py

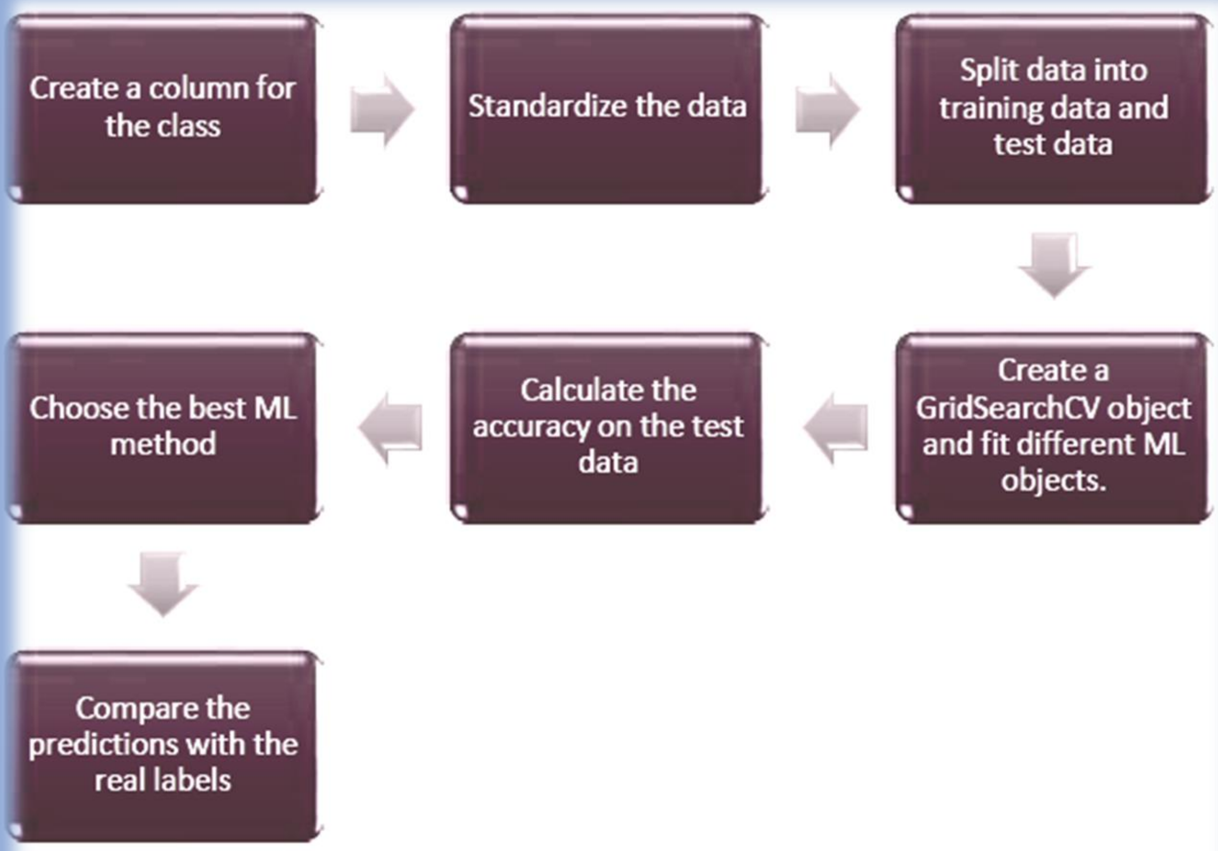
Predictive Analysis (Classification)

The Scikit-learn library was used to predict the success of the rocket's first stage landing. The data set was divided into training and test. The following machine learning methods are used:

- Logistics Regression method
- Support Vector Machine method
- Decision tree method
- K nearest neighbors method

The prediction accuracy for each method was determined and the optimal one was determined.

URL link: https://github.com/Afex81/-SpaceX_Landing_Prediction/blob/787623ad3d7fa6f80d53b8b280c85d45dd3c652e/8_MachineLearning.py



Results

Conducted research using data science methods showed that there are relationships between the successful landing of the first stage of the rocket and launch parameters. The number of successful launches has increased since 2015, which can be explained by the development of technology and the use of previous experience. Also, the launch site and the weight of the payload that the rocket launched into orbit had a certain influence on the success of the launches.

Using pre-prepared data from previous years about the successful landing of the first stage of the SpaceX Falcon 9 rocket, the machine learning methods we applied were able to predict the future landing success rate at the level of 83%.

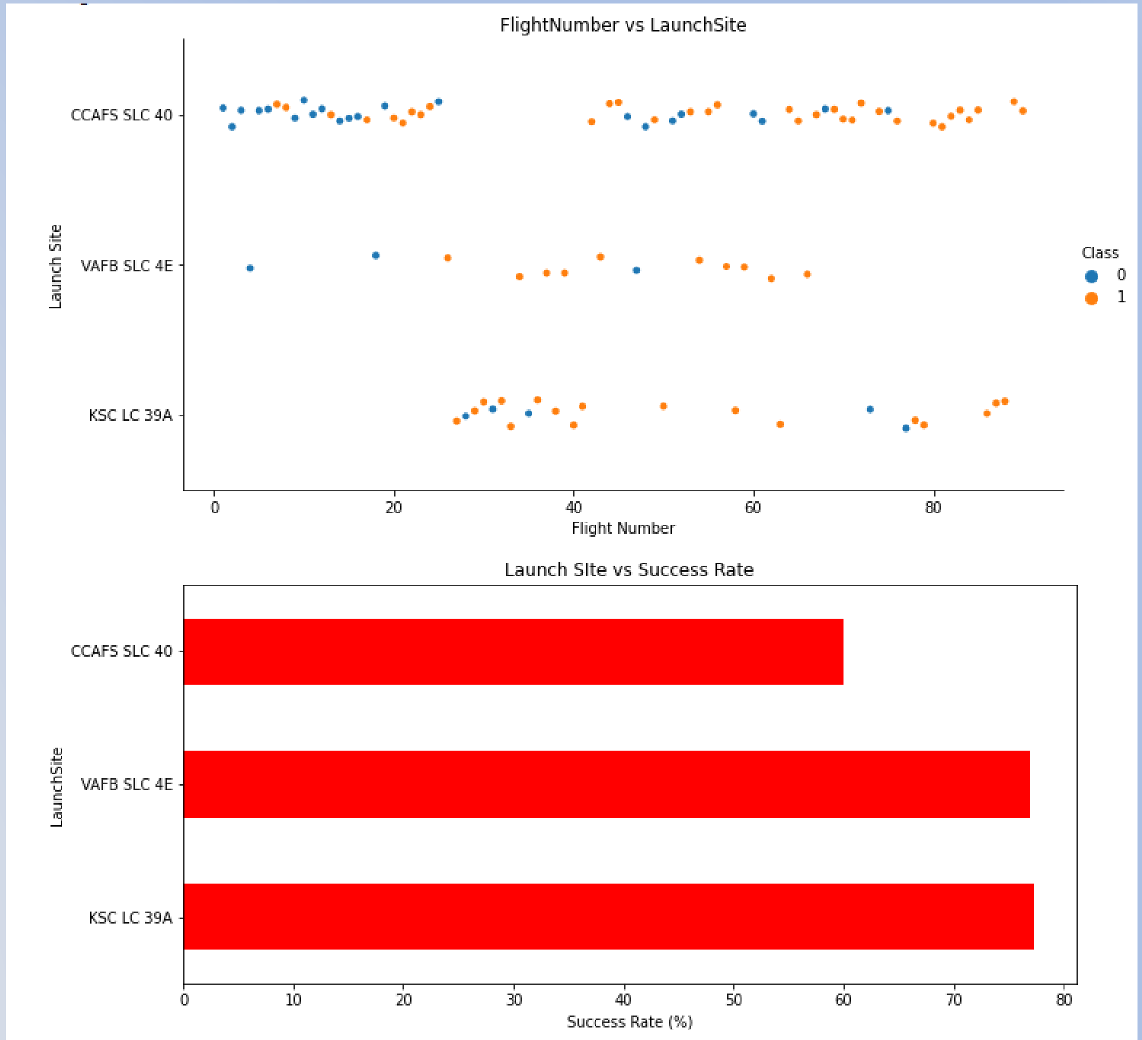
Section 2

Insights drawn from EDA

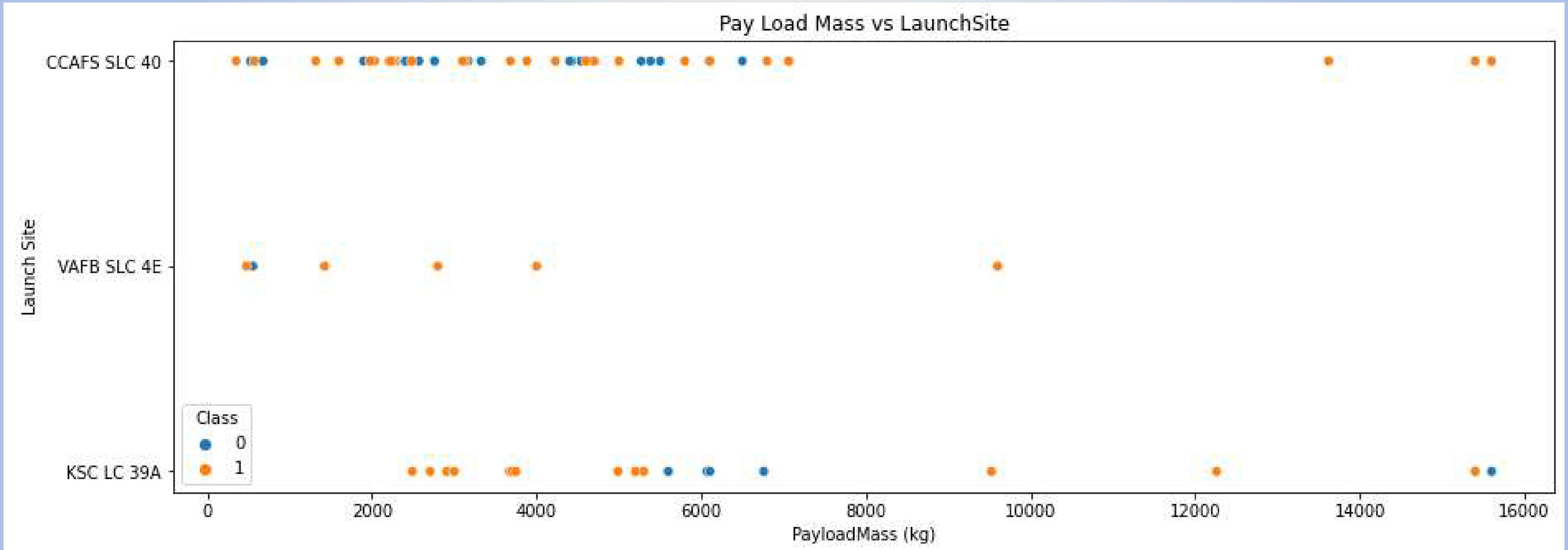
Flight Number vs. Launch Site

Over time, the number of successful landings of the first stage of the rocket increased. The most launches were made from the launch site CCAFS SLC 40. The highest percentage of successful landings was at the site VAFB SLC 4E, but there was also the lowest number of attempts.

The level of successful landings of the first stage is not lower than 60% at all launch sites.



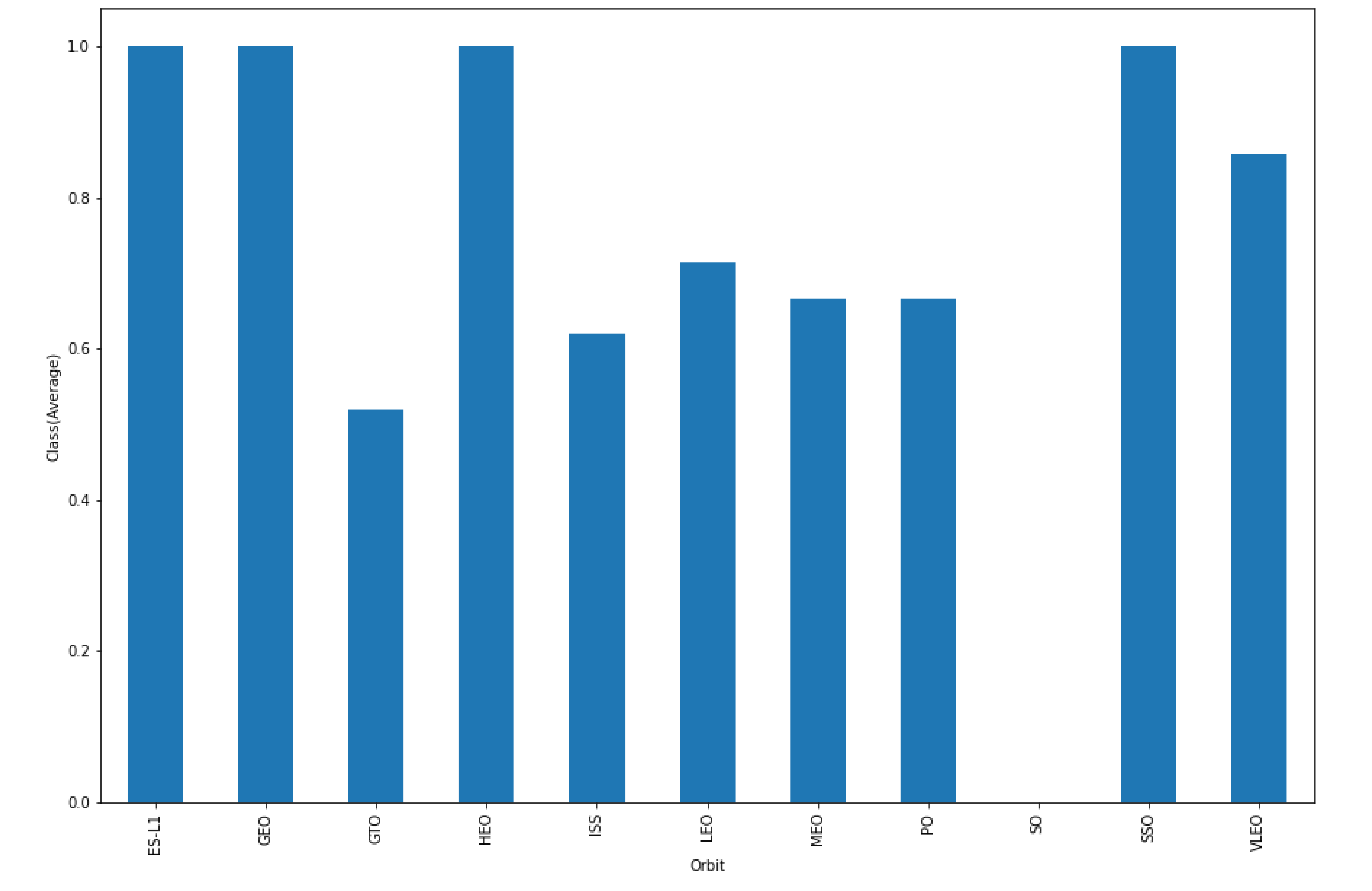
Payload vs. Launch Site



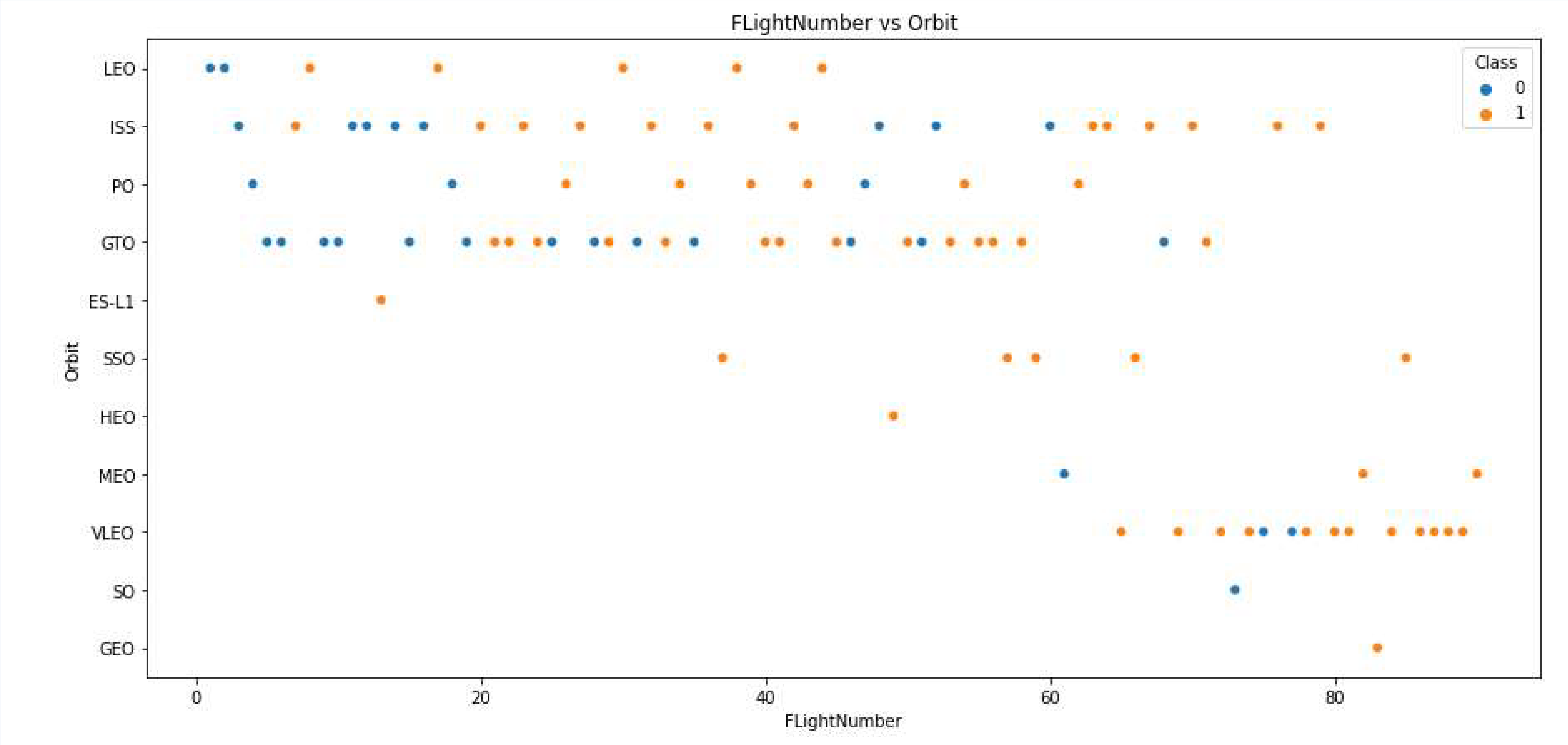
The scatterplot diagram indicates that no rockets with a payload of more than 10,000 kg were launched from the launch pad VAFB-SLC.

Success Rate vs. Orbit Type

The orbit types SSO, HEO, GEO and ES-L1 had the highest success rate.

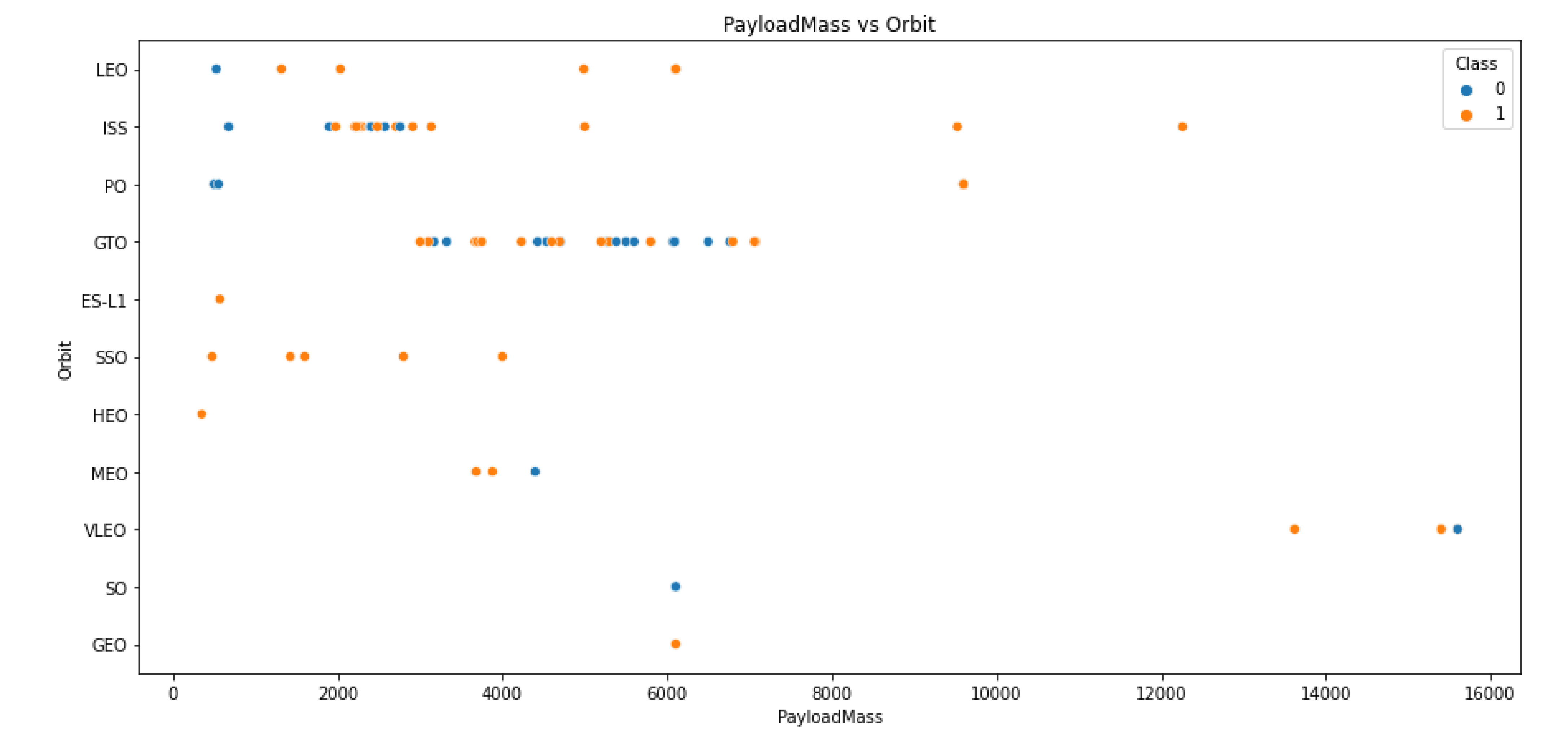


Flight Number vs. Orbit Type



In LEO orbit, there is a relationship between the number of successful rocket landings and the number of launches. However, no such relationship is observed in the GTO orbit.

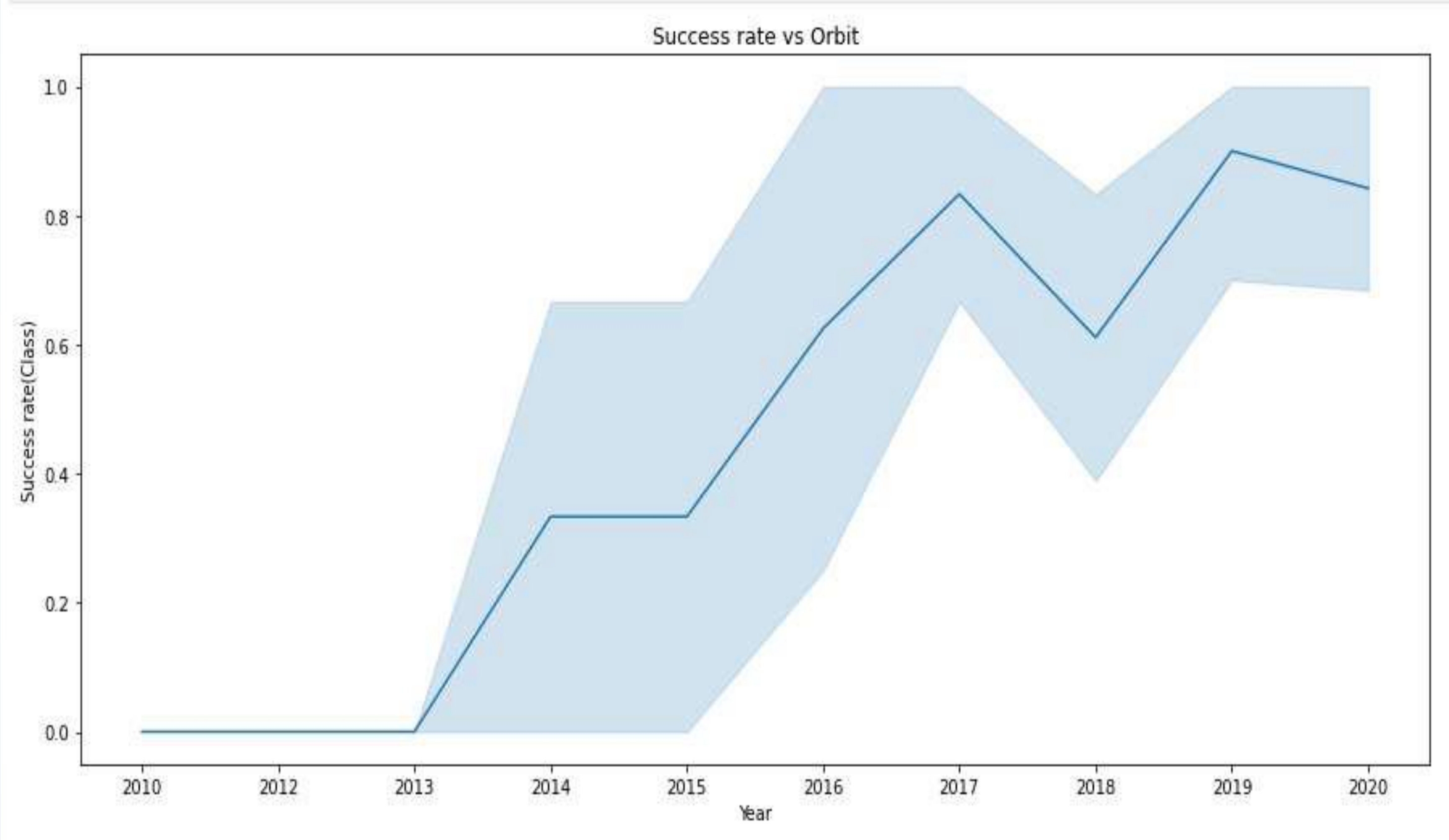
Payload vs. Orbit Type



For the Polar, LEO and ISS orbits, the relationship between increasing rocket payload and the rate of successful landings is seen. For the GTO orbit, no such clear relationship is observed, the payload does not affect the success rate.

Launch Success Yearly Trend

Success rate has significantly increased from 2013 to 2020.



All Launch Site Names

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

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

Launch Site Names Begin with 'CCA'

Date	Launch_Site	Orbit	Customer	Mission_Outcome	Landing_Outcome
04-06-2010	CCAFS LC-40	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	CCAFS LC-40	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22-05-2012	CCAFS LC-40	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	CCAFS LC-40	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	CCAFS LC-40	LEO (ISS)	NASA (CRS)	Success	No attempt

All launches were carried out from the same site, in addition to SpaceX, other companies also tested their rockets.

Total Payload Mass

- The data in the table displays the total payload mass carried by boosters launched by NASA .
- NASA (CRS) had a significantly higher total payload mass compared to the rest.

Customer	Total_Payload_Mass
NASA (CRS)	45596
NASA (CCDev)	12530
NASA (CCP)	12500
NASA (CCD)	12055
NASA (CTS)	12050
NASA (CRS), Kacific 1	2617
NASA / NOAA / ESA / EUMETSAT	1192
NASA (LSP) NOAA CNES	553
NASA (COTS)	525
NASA (LSP)	362
NASA (COTS) NRO	0

Average Payload Mass by F9 v1.1

Average_Payload_Mass (kg)	Booster_Version
2928.4	F9 v1.1

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

First Successful Ground Landing Date

Date	Landing_Outcome
22-12-2015	Success (ground pad)

The first successful landing of the rocket's first stage on the ground platform took place in December 2015. This event went down in history, as no one had been able to do such a thing before.

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

The payload weighing from 4,000 to 6,000 kg was launched into orbit only 4 times, and all attempts to land the first stage of the rocket were successful.

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Outcomes
Failure (in flight) 1	1
Success	99
Success (payload status unclear)	1

Missions generally tend to be successful with the exception of one failure.

Boosters Carried Maximum Payload

The first stage of the rocket launched the maximum payload weighing 15,600 kg into orbit 12 times.

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

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

At the beginning of 2015, an unsuccessful landing was recorded in only two cases.

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

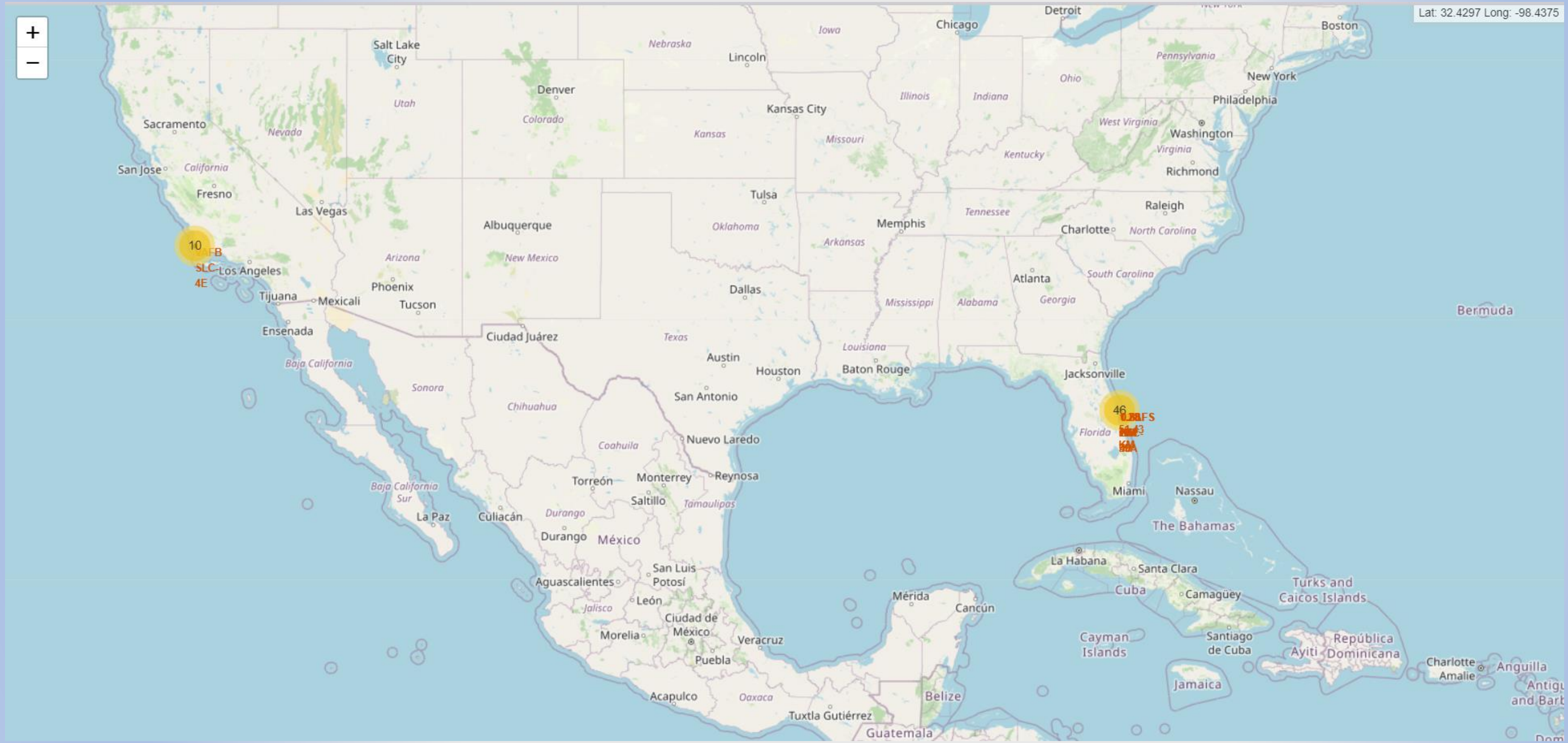
These tables indicate that the number of successful rocket first stage landings has increased since 2015. Until 2013, there were no attempts to land the first stage.

date	Landing_Outcome	Outcomes
2016-04-08	Success (drone ship)	14
2015-12-22	Success (ground pad)	9
2015-06-28	Precluded (drone ship)	1
2015-01-10	Failure (drone ship)	5
2014-04-18	Controlled (ocean)	5
2013-09-29	Uncontrolled (ocean)	2
2012-05-22	No attempt	22

Section 3

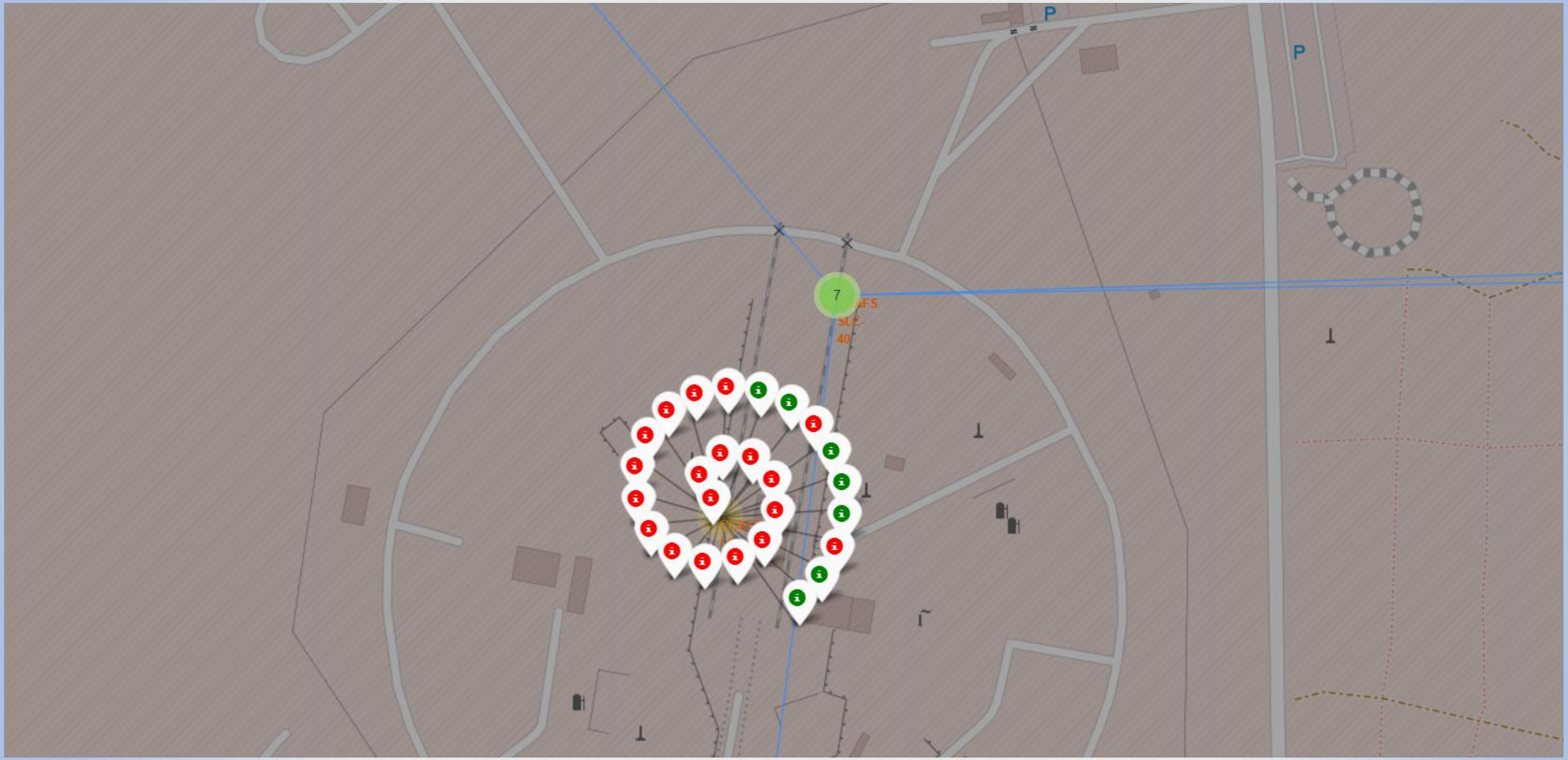
Launch Sites Proximities Analysis

Launch Site Locations



The map shows that all launch sites are near the ocean. Most of them are concentrated on the southeastern coast of the United States.

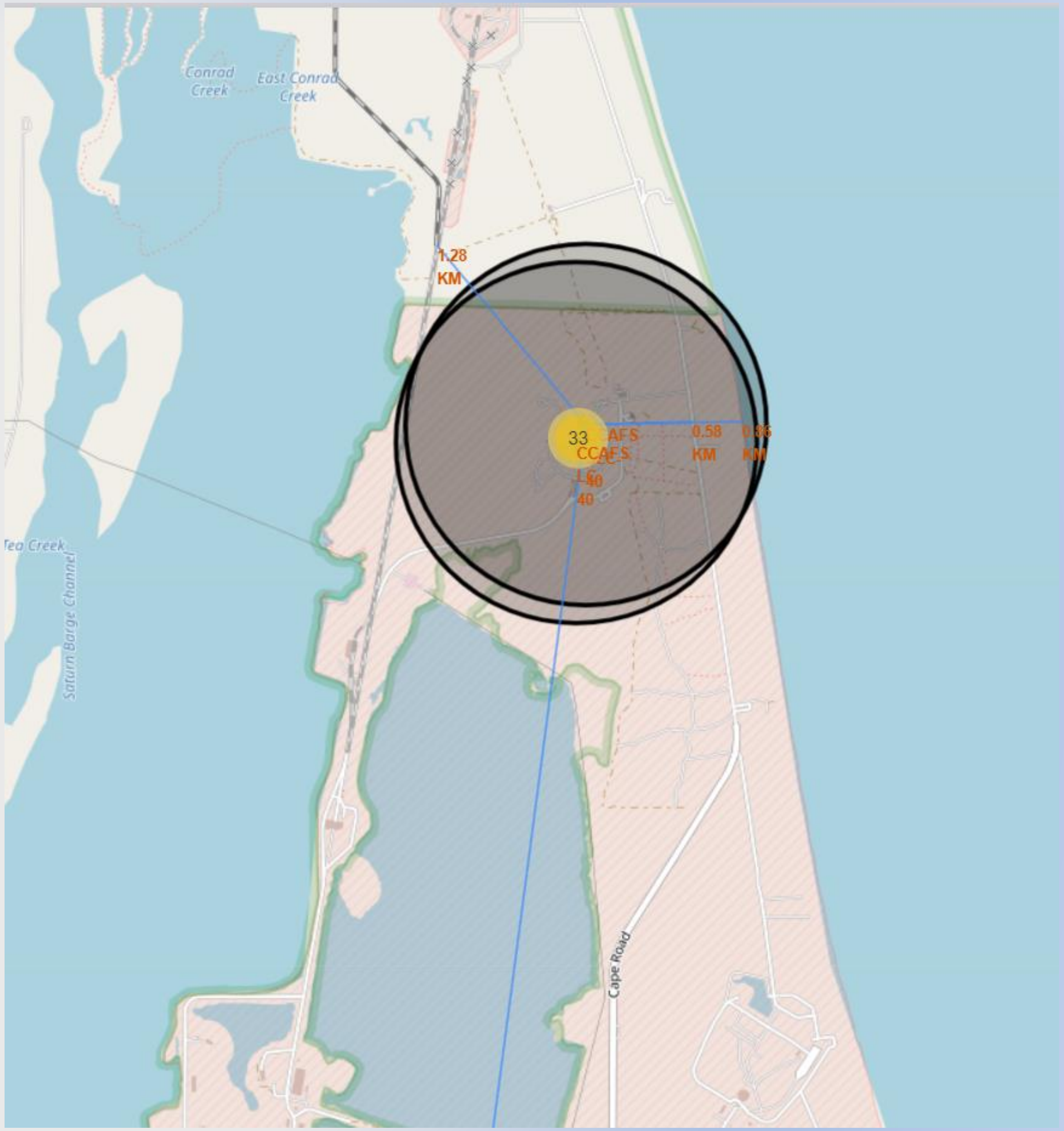
Success Rate of Rocket Launches



- Successful first stage landing attempts are indicated by a green marker, and red by failed rocket launches.
- The CAFS LC-40 launch pad did not have a high degree of missile landing success.

Surrounding Landmarks

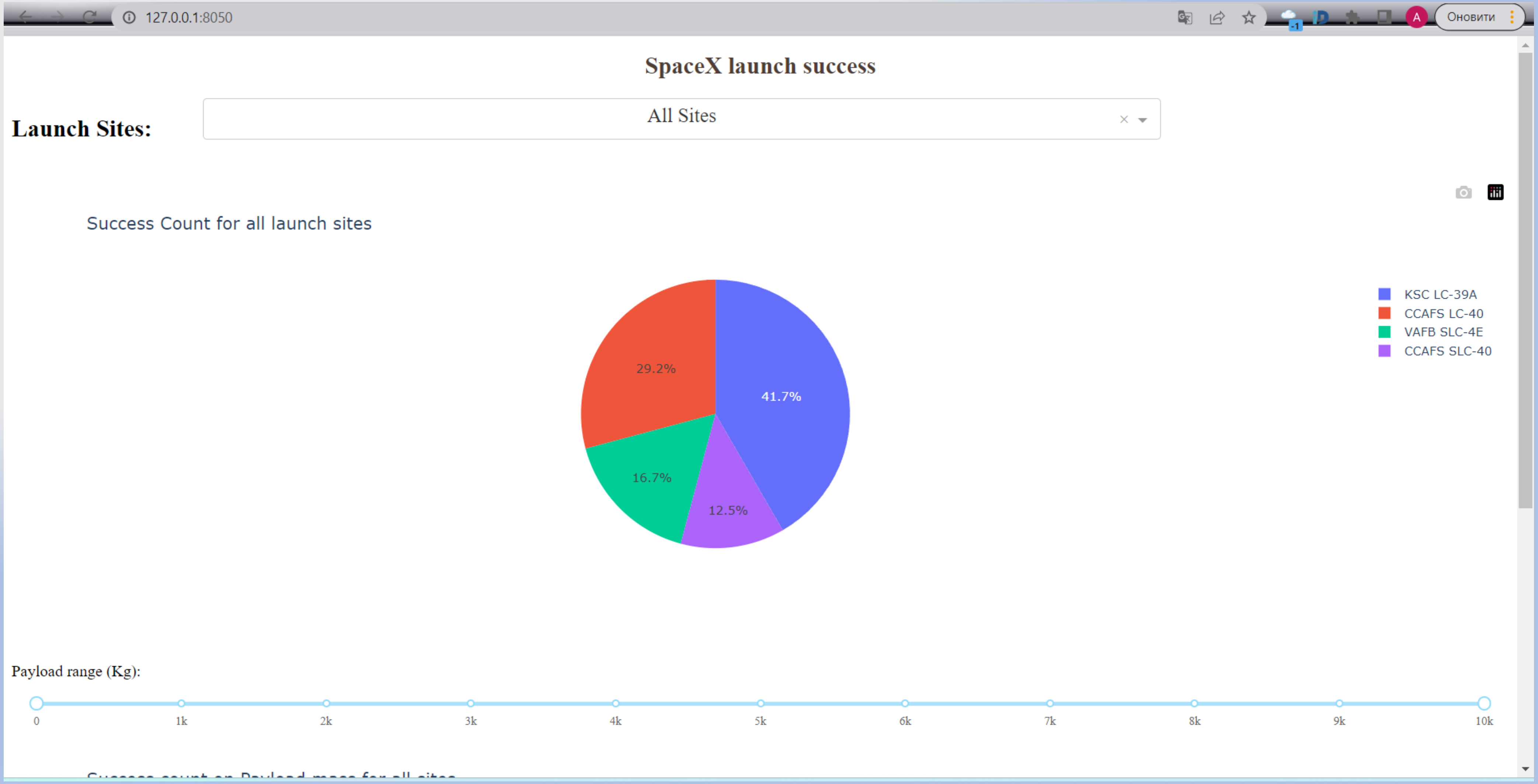
It can be concluded from the map that the launch sites are located away from large cities and at the same time close to transport infrastructure such as railways and highways.



Section 4

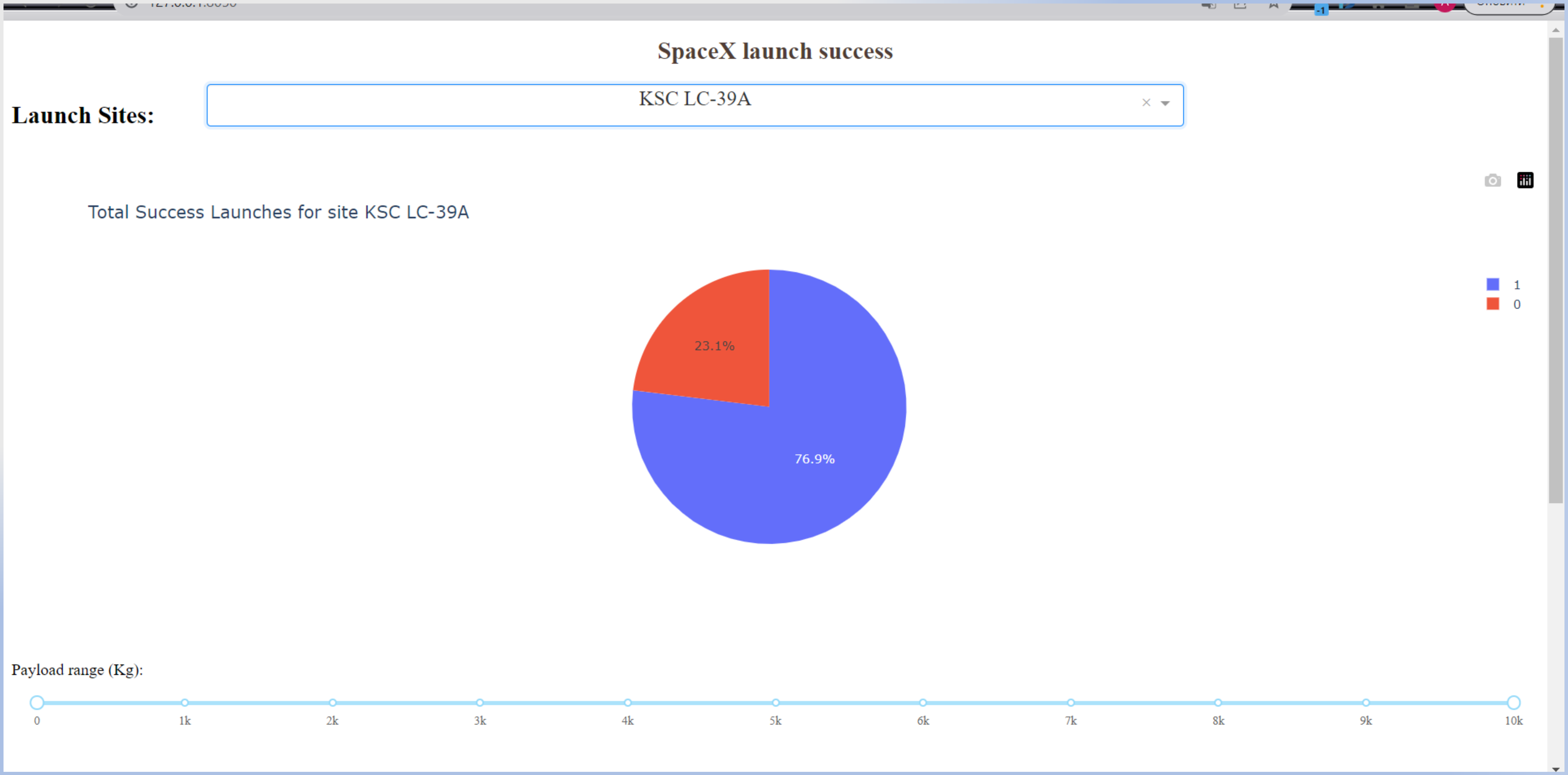
Build a Dashboard with Plotly Dash

Successful Launches by Site



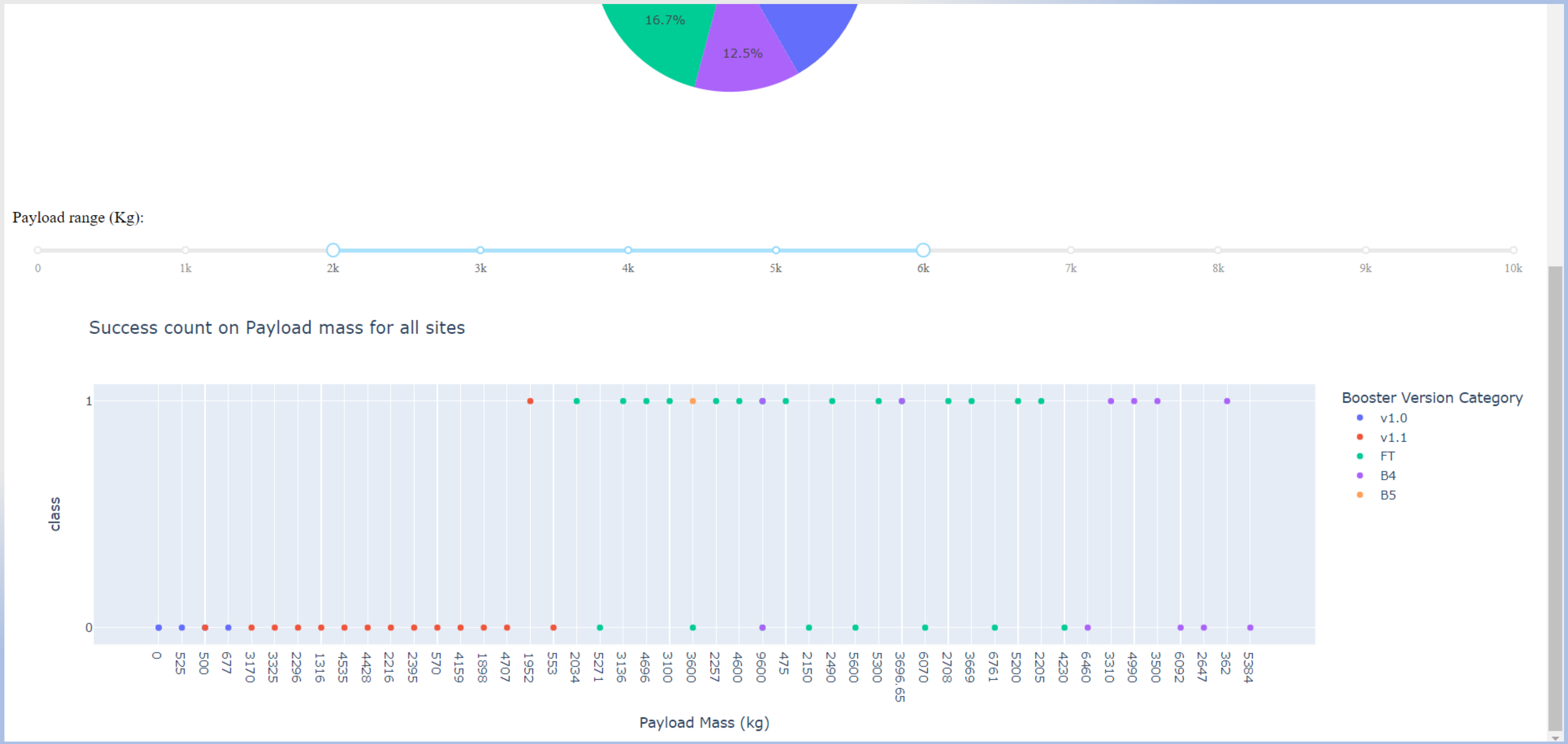
Site KSC LC-39A has the largest successful launches as well the highest launch success rate.

Total Successful Launches for Site KSC LC-39A



The KSC LC-39A launch pad has the highest percentage of successful landings at 76.9%.

Payload Mass vs. Launch Success for All Sites



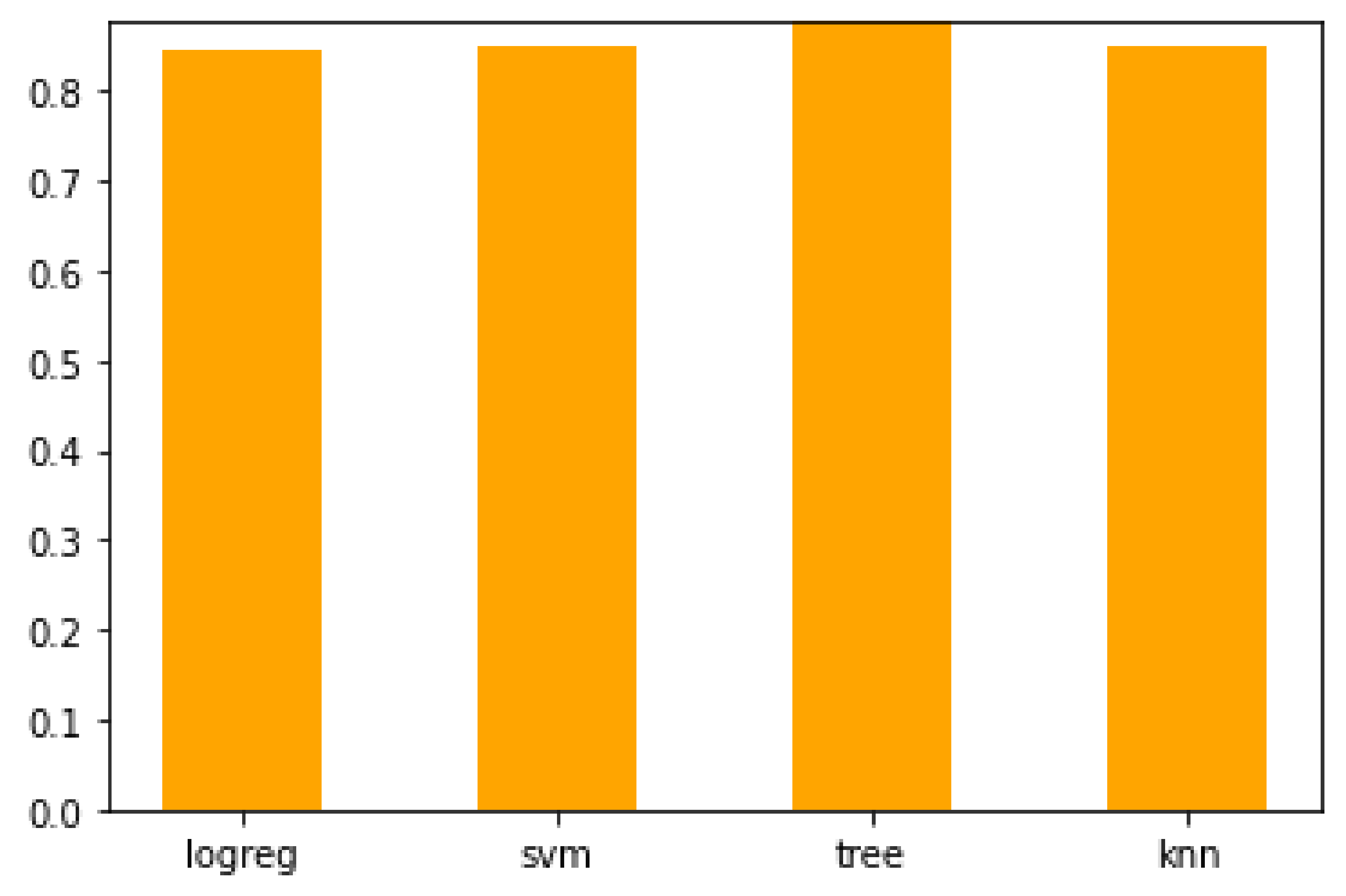
Most of the successful landings of the first stage of the rocket after the launch into orbit of the payload weighing from 2000 to 6000 kg are attributed to the FT booster.

Section 5

Predictive Analysis (Classification)

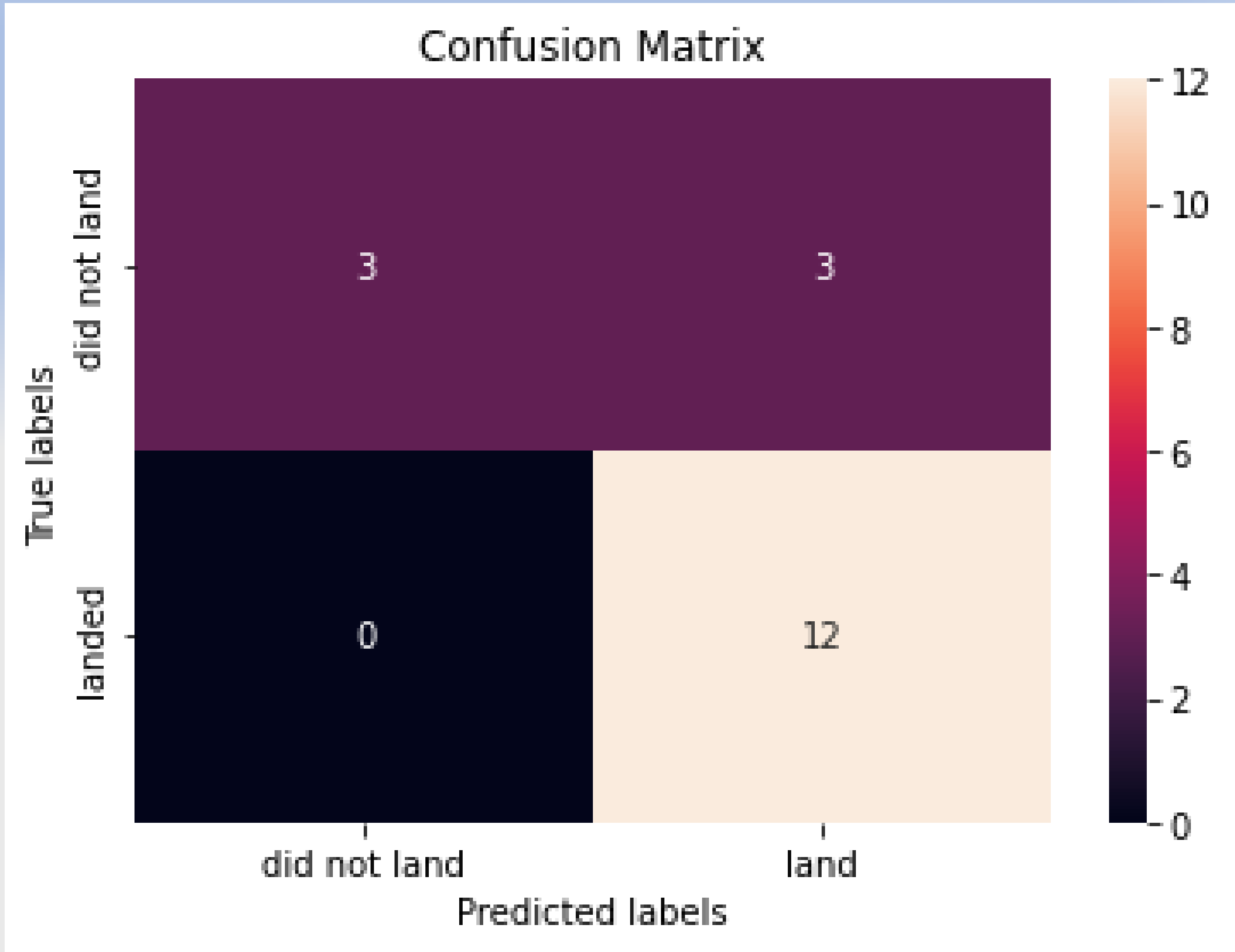
Classification Accuracy

The best indicator of prediction accuracy at the level of 87.7% was shown by the machine learning algorithm "DECISION TREE"



Confusion Matrix

The figure shows the confusion matrix for the DECISION TREE machine learning algorithm. As you can see, the model failed to predict only three cases.



Conclusions

In order to successfully compete in the space cargo launch market with such industry leaders as SpaceX, data on their rocket launches and successful landings of the first stage was analyzed.

The results of the analysis made it possible to establish that the most successful landing results began in 2015.

The rockets that were most successful in landing had a payload of 2,000 to 6,000 kg. At the same time, the FT booster was most often successful.

All launch sites are located near the ocean and transport infrastructure, which allows to reduce logistics costs and use floating platforms for the landing of the first stage of the rocket.

All these data were used to train various machine prediction algorithms, the most accurate of which was the DECISION TREE method with an accuracy of 87.7%.

This model will allow our company to be competitive on the market faster, attract more investors and customers.

Appendix

Table from the file obtained by web scraping.

URL link: https://github.com/Afex81/-SpaceX_Landing_Prediction/blob/d45cb5f9319105ec9c9553fc0b352ae6093f0a09/spacex_web_scraped.csv

main -SpaceX_Landing_Prediction / spacex_web_scraped.csv Go to file

Afex81 Add files via upload Latest commit d45cb5f 3 minutes ago History

1 contributor

228 lines (228 sloc) | 11.9 KB Raw Blame

Search this file...

Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date
1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	F9 v1.0B0003.1	Failure	4 June 2010
2	CCAFS	Dragon	0	LEO	NASA (COTS) NRO	Success	F9 v1.0B0004.1	Failure	8 December 2010
3	CCAFS	Dragon	525 kg	LEO	NASA (COTS)	Success	F9 v1.0B0005.1	No attempt	22 May 2012
4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA (CRS)	Success	F9 v1.0B0006.1	No attempt	8 October 2012
5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA (CRS)	Success	F9 v1.0B0007.1	No attempt	1 March 2013
6	VAFB	CASSIOPE	500 kg	Polar orbit	MDA	Success	F9 v1.1B1003	Uncontrolled	29 September 2013
7	CCAFS	SES-8	3,170 kg	GTO	SES	Success	F9 v1.1	No attempt	3 December 2013
8	CCAFS	Thaicom 6	3,325 kg	GTO	Thaicom	Success	F9 v1.1	No attempt	6 January 2014
9	Cape Canaveral	SpaceX CRS-3	2,296 kg	LEO	NASA (CRS)	Success	F9 v1.1	Controlled	18 April 2014
10	Cape Canaveral	Orbcomm-OG2	1,316 kg	LEO	Orbcomm	Success	F9 v1.1	Controlled	14 July 2014
11	Cape Canaveral	AsiaSat 8	4,535 kg	GTO	AsiaSat	Success	F9 v1.1	No attempt	5 August 2014
12	Cape Canaveral	AsiaSat 6	4,428 kg	GTO	AsiaSat	Success	F9 v1.1	No attempt	7 September 2014
13	Cape Canaveral	SpaceX CRS-4	2,216 kg	LEO	NASA (CRS)	Success	F9 v1.1	Uncontrolled	21 September 2014
14	Cape Canaveral	SpaceX CRS-5	2,395 kg	LEO	NASA (CRS)	Success	F9 v1.1	Failure	10 January 2015
15	Cape Canaveral	DSCOVR	570 kg	HEO	USAF NASA NOAA	Success	F9 v1.1	Controlled	11 February 2015

Thank you!