Data Science Capstone Project

LinkedIn

Julián Uribe Castañeda

Coursera

14 de enero de 2023

Outline

(✓) Executive Summary

(✓) Introduction

(✓) Methodology (Executive Summary)

(✓) Methodology

(✓) Insights drawn from EDA

(✓) Insights drawn from SQL

(✓) Launch Sites Proximities Analysis

(✓) Build a Dashboard with Plotly Dash

(✓) Predictive Analysis (Classification)

(✓) Appendix

Executive Summary

Executive Summary

Summary of methodologies

(✓) Data collection

(✓) Data wrangling

(✓) Perform exploratory data analysis using visualization and SQL

(✓) Perform interactive visual analytics using Folium and Plotly Dash

(✓) Perform predictive analysis using classification models

Summary of all results

(✓) Data analysis results

(✓) Data visualization results

Introduction

Introduction

Project background and context

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 solved

Using the relational data set, which determine some features of rocket launches such as:

- Payload mass
- Orbit
- Date
- Lat and Long (latitude and longitude)
- Launch site
- Class

We want to solve the following questions:

(1) How are these variables related to first stage landing success?

(2) As time increases, how does the landing success rate change?

(3) Among the following binary classification methods:

- LRM (logistics regression method)
- SVM (support vector machine)
- Decision tree
- K-nearest neighbors

What is the most accurate method that can be used for our data set?

Methodology (Executive Summary)

Methodology (Executive Summary).

Data collection

Our data set for this project is comprised of SpaceX launch data and Wikipedia data. More specifically, we have that:

(⋆) SpaceX launch data is a data set gathered from an API called the SpaceX REST API. This API give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.

(⋆) Wikipedia data is a collection Falcon 9 historical launch records. This data is given in the following link wiki/Falcon 9.

Data wrangling

We took the data set and we process it of the following way:

(⋆) We filter the data choosing the most notable features

(⋆) We deal with various binary classification methods and look for which of these methods is the most accurate and best fits our filtered data set.

Perform exploratory data analysis using visualization and SQL

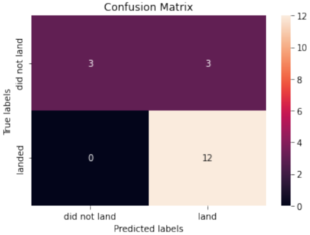
We analyzed the data and some visualizations of it using SQL and some Python libraries such as Pandas, Numpy, matplotlib and seaborn.

Perform interactive visual analytics using Folium and Plotly Dash

We describe some interesting visual things related with launch sites using Folium and Plotly Dash.

Perform predictive analysis using classification models

We trained some binary classification methods to be ensure that we get the best predictive results and we found out which one is the best classification method to work in.



Methodology

Methodology

Data Collection

Our data set for this project is composed by *SpaceX* launch data and Wikipedia data, where:

(✓) *SpaceX* launch data

SpaceX launch data is a data set gathered from an API called the *SpaceX* REST API. This API give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.

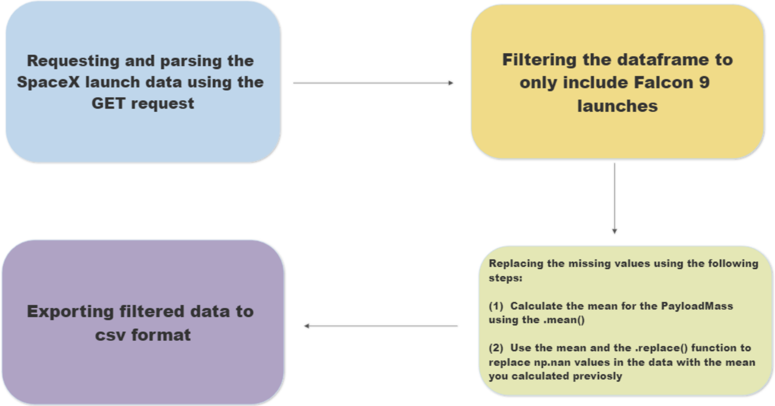
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

(✓) Wikipedia data

Wikipedia data is a collection Falcon 9 historical launch records. This data is given in the following link wiki/Falcon 9.

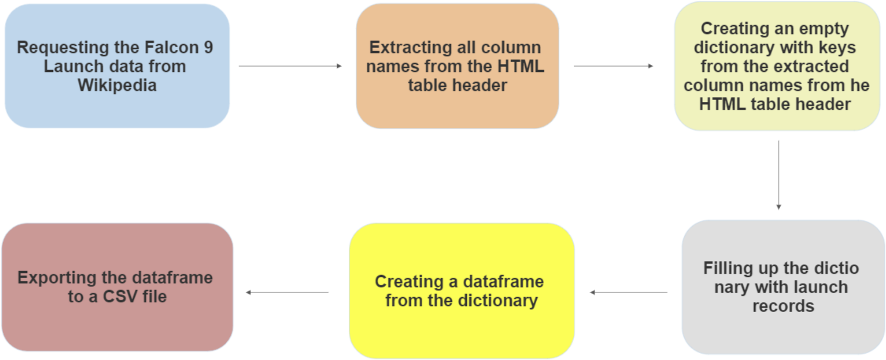
[hide] Flight No.	Date and time (UTC)	Version, booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[13]	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[15]								
79	19 January 2020, 15:30[16]	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[17] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[18]	NASA (CTS)[19]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[20] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[21] The abort test used the capsule originally intended for the first crewed flight.[22] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[23] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[24]	F9 B5 △ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[25]								
81	17 February 2020, 15:05[26]	F9 B5 △ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[27] due to incorrect wind data.[28] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[29]	F9 B5 △ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 △)	1,977 kg (4,359 lb)[30] (excl. Dragon mass)	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[31] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[32] It was SpaceX's third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
	18 March 2020, 12:16[33]	F9 B5 △ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Failure (drone ship)

Data Collection: SpaceX launch data



GitHub URL: SpaceX Launch API

Data Collection: Wikipedia data



GitHub URL: Wikipedia data

Data wrangling

In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident. In fact, we have the following things:

($\checkmark$) ***True Ocean*** means the mission outcome was successfully landed to a specific region of the ocean while

($\checkmark$) ***False Ocean*** means the mission outcome was unsuccessfully landed to a specific region of the ocean

($\checkmark$) ***True RTLS*** means the mission outcome was successfully landed to a ground pad

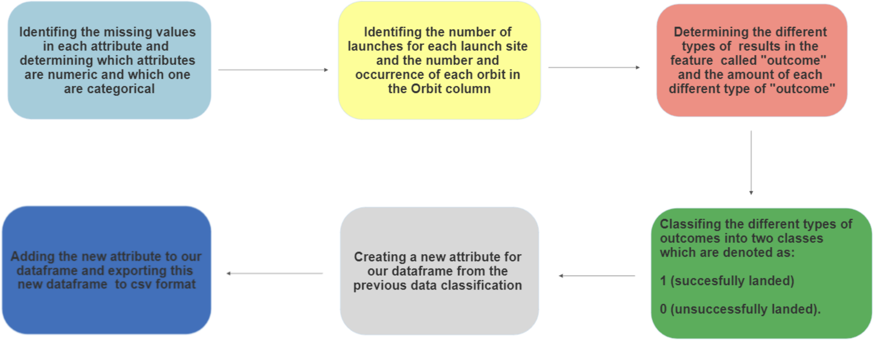
($\checkmark$) ***False RTLS*** means the mission outcome was unsuccessfully landed to a ground pad

($\checkmark$) ***True ASDS*** means the mission outcome was successfully landed on a drone ship

($\checkmark$) ***False ASDS*** means the mission outcome was unsuccessfully landed on a drone ship

So, we've taken our data, converted it to a dataframe, and then created a new column with values $\{0,1\}$ such that:

$$\text{value} = \begin{cases} 0 & \text{if the Outcome is successfully landed,} \\ \\ 1 & \text{if the Outcome is not successfully landed.} \end{cases}$$



GitHub URL: Data wrangling

Exploratory data analysis with data visualization

Summary of plotted charts

We have plotted the following charts to get the relationship between these pair of variables:

(✓) FlightNumber vs. PayloadMass (categorical scatterplots)

(✓) FlightNumber vs. LaunchSite (categorical scatterplots)

(✓) LaunchSite vs. PayloadMass (categorical scatterplots)

(✓) Orbit vs. Class (bar plot)

(✓) FlightNumber vs. Orbit (categorical scatterplots)

(✓) Payload vs. Orbit (categorical scatterplots)

(✓) Year vs. Class (line chart)

Therefore, we can see that we only have 3 types of plots which are categorical scatterplots, bar plots and line charts. These plots show the relationship between each pair of variables. Next, we remember when to use each type of plot.

(1) *Categorical scatterplots*

A categorical scatterplot is a plot that shows the relationship between a numerical and one or more categorical variables using one of several visual representations.

(2) *Bar plots*

A bar plot represents an estimate of central tendency for a numeric variable with the height of each rectangle and provides some indication of the uncertainty around that estimate using error bars. Bar plots include 0 in the quantitative axis range, and they are a good choice when 0 is a meaningful value for the quantitative variable, and you want to make comparisons against it.

(3) *Line charts*

A line chart compares two quantitative variables and one of these variables measures units of time.

GitHub URL: Exploratory data analysis with data visualization

Exploratory data analysis with SQL

Summary of the SQL queries

(✓) Display the names of the unique launch sites in the space mission.

(✓) Display 5 records where launch sites begin with the string 'CCA'.

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

(✓) 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 for the months in year 2015.

(✓) Ranking the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

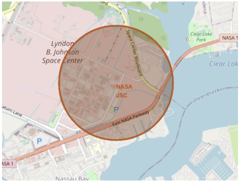
GitHub URL: Exploratory data analysis with SQL

Build an interactive map with Folium

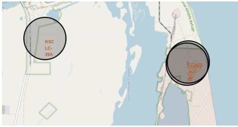
Summary maps objects

(1) *Mark all launch sites*

- I have added a marker at Nasa Johnson Space Center. This marker has a popup label and an orange circle.

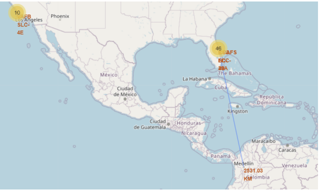


- I have added a marker for each launch site. These markers have popup label and a circle.



- I have added a marker for each launch site. These markers have popup label and a yellow circle.

(2) *Mark a line segment between two points*



I have added a line segment that describes geometrically the distance between the launch sites and Medellin (my city).

GitHub URL: Interactive map with Folium

Build a dashboard with plotly dash

Summary (plots/graphs) added to a dashboard

(1) *Launch sites dropdown list:* I have added a dropdown list to enable launch site selection.

(2) *Pie chart showing success launches:* I have added a pie chart to show the total successful launches count for all sites and the success vs failes counts for the site, if a specific lauch site was selected.

(3) *Slider of payload mass range:* I have added a slider to select payload range.

(4) *Scatter chart of payload mass vs success rate for the different booster versions:* I have added a scatter chart to show the correlation between payload and launch success.

GitHub URL: Dash

Predictive analysis (classification)

Summary of how the classification model was built, evaluated, improved and found the best performance

Next, we show the steps taken to obtain the best performance of the classification model:

(1) We took the feature «Class» in data obtained in the *lab wrangling* that represents the outcome of each launch and the the integer data set obtained in the *lab eda*.

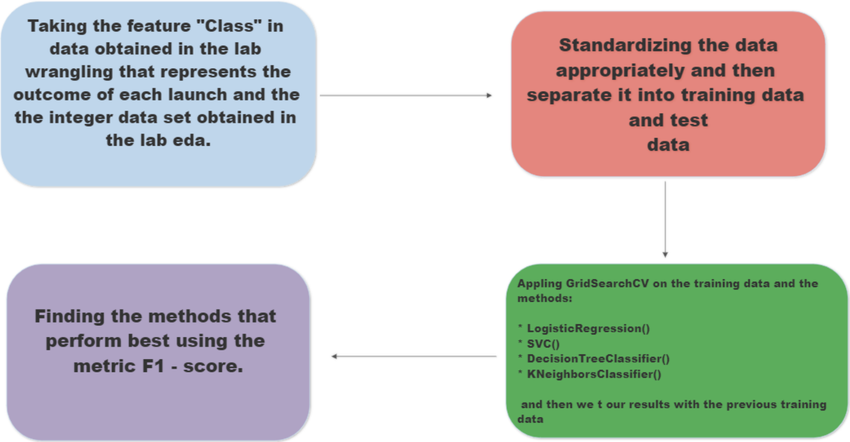
(2) We standardized the data appropriately and then separate it into training data and test data.

(3) We applied GridSearchCV on the training data and the methods:

$$\begin{cases} \text{LogisticRegression}(), \\ \text{SVC}(), \\ \text{DecisionTreeClassifier}(), \\ \text{KNeighborsClassifier}() \end{cases}$$

and then we fit our results with the previous training data.

(4) We found the methods that perform best using the metric F_1 - score.

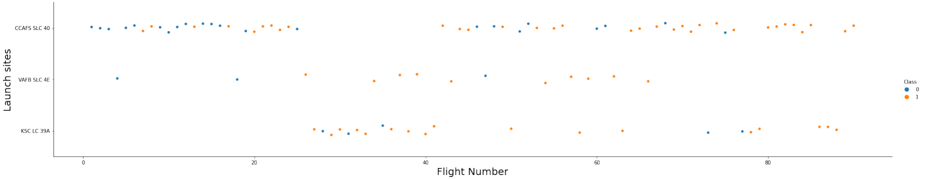


GitHub URL: Machine Learning Classification

Insights drawn from EDA

Insights drawn from EDA

Flight Number vs. Launch Site



Explanation.

For simplicity, we will denote the launch sites as follows:

$$\begin{cases} A = \text{KSC LC 39A,} \\ B = \text{VAFB SLC 4E,} \\ C = \text{The CCAFS SLC 40.} \end{cases}$$

Thus, from the scatterplot given above we can infer the following things:

- Of the first 20 flights, there were only two defective occurrences in A and B, there were no occurrences in C.

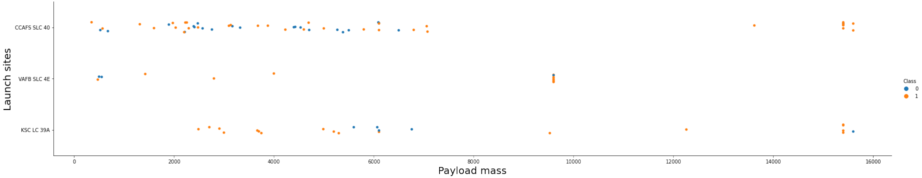
- Of the first 20 flights, there were 18 occurrences in C and only 5 were successful.

- Between flights 20 and 40 there were 11 occurrences in A and only 2 failed.

- Between flights 40 and 80 there was more frequency in C with 25 launches, of which only 17 were successful.

- As the number of flights increases, each new launch has a higher probability of success, which is normal since as time goes by, experience is gained.

Payload vs. Launch Site



Explanation.

For simplicity, we will denote the launch sites as follows:

$$\begin{cases} A = \text{KSC LC 39A}, \\ B = \text{VAFB SLC 4E}, \\ C = \text{CCAFS SLC 40}. \end{cases}$$

Thus, from the scatterplot given above we can infer the following things:

- The launch site C had more than 50% of the total launches. This means that for some reason, the site C was the preferred one for launching rockets with payload.
- The site A had a 100% success for payload mass under 5000 kg.
- The site B had a 100% success for payload mass between 1000 kg and 9000 kg.
- The site C got approximately 50% success for payload between 4000 kg and 6000 kg.

Success Rate vs. Orbit Type

Explanation.

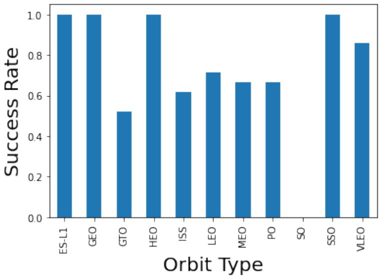
(✓) The orbits with 100% success rate are:

$$\begin{cases} \text{ES-L1,} \\ \text{GEO,} \\ \text{HEO,} \\ \text{SSO.} \end{cases}$$

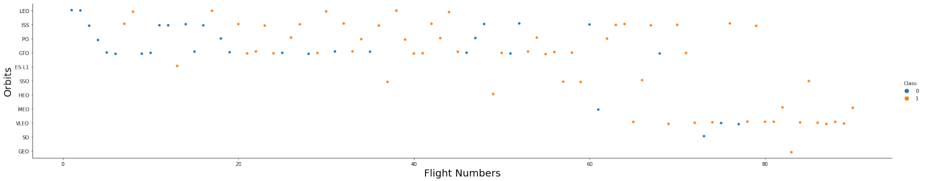
(✓) The orbit with 0% success rate is SO.

(✓) The orbits with success rate between 40% and 80% are:

$$\begin{cases} \text{GTO,} \\ \text{ISS,} \\ \text{LEO,} \\ \text{MEO,} \\ \text{PO.} \end{cases}$$



Flight Number vs. Orbit Type



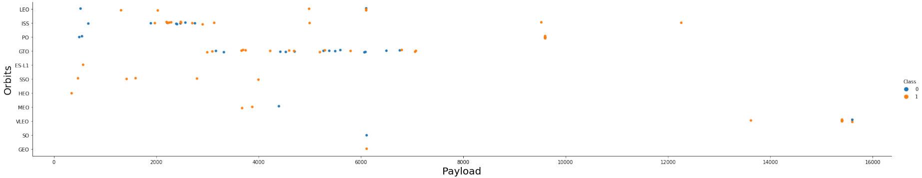
Explanation.

(✓) As the number of flights increases, it is found that flights over orbits tend to be more successful.

(✓) The first flights were made on the orbits closest to the earth.

(✓) As the number of flights increases, we have that these flights went to the farthest orbits from the earth.

Payload vs. Orbit Type



Explanation.

(✓) most of our data is in the orbits closest to the earth, since it is easier to reach and therefore we better conserve our payload mass.

(✓) If the payload tends to be very heavy (greater than 8000 km), there are fewer flights to more distant orbits.

Launch Success Yearly Trend

Explanation.

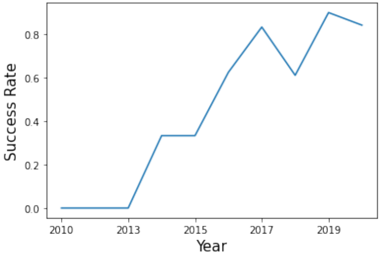
(√) No flight between 2010 and 2013 is successful, since the success rate in these periods is constant with a value of 0.

(√) As time passes between 2010 and 2017, the success rate grows.

(√) Between 2017 and 2018, the success rate decreases.

(√) Between 2018 and 2019, the success rate grows.

(√) Between 2019 and 2020, the success rate decreases.



Insights drawn from SQL

All Launch Site Names

```
[ ]  %%sql

     select distinct("Launch_Site") from "SPACEXTBL";
```

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

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

Explanation.

Displaying the names of the unique spacial launch sites.

Launch Site Names Begin 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
2010-04-06 00:00:00	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12 00:00:00	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 00:00:00	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10 00:00:00	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03 00:00:00	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Explanation.

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

Total Payload Mass

```
%%sql

select sum("PAYLOAD_MASS__KG_") as "total payload mass" from "SPACEXTBL" where "Customer" = "NASA (CRS)";
```

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

total payload mass
45596

Explanation.

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_") as "Average" from "SPACEXTBL" where "Booster_Version"  like  "%F9 v1.1%";
```

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

Average
2534.6666666666665

Explanation.

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

First Successful Ground Landing Date

```
%%sql

select min("Date") as "First successful ground landing" from "SPACEXTBL" where "Landing _Outcome" = "Success (ground pad)"
```

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

First successful ground landing
2015-12-22 00:00:00

Explanation.

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 "Booster_Version"  from "SPACEXTBL" where ("Landing _Outcome" = "Success (drone ship)") and ("PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000)
```

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

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

Explanation.

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 "mission_outcome", count(*) as "total_number" from "SPACEXTBL" group by "Mission_Outcome"
```

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

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

Explanation.

Listing the total number of successful and failure mission outcomes.

Boosters Carried Maximum Pyload

```
%%sql

select "Booster_Version" as "Boosterversion" from "SPACEXTBL" where  "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from "SPACEXTBL")
```

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

Boosterversion
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

***Explanation*.**

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

2015 Launch Records

```
%%sql

select substr("Date",6,2) as "Month", "Date", "Landing _Outcome", "Booster_Version", "Launch_Site" from "SPACEXTBL" where (substr("Date",1,4) = "2015") and ("Landing _Outcome" = "Failure (drone ship)")
```

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

Month	Date	Landing _Outcome	Booster_Version	Launch_Site
10	2015-10-01 00:00:00	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015-04-14 00:00:00	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Explanation.

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 and 2017-03-20

```
%%sql

select "Landing _Outcome", count(*) as "quantity of outcomes" from "SPACEXTBL"
where ("Date" BETWEEN "2010-06-04" and "2017-03-20")
group by "Landing _Outcome"
order by 2 desc
```

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

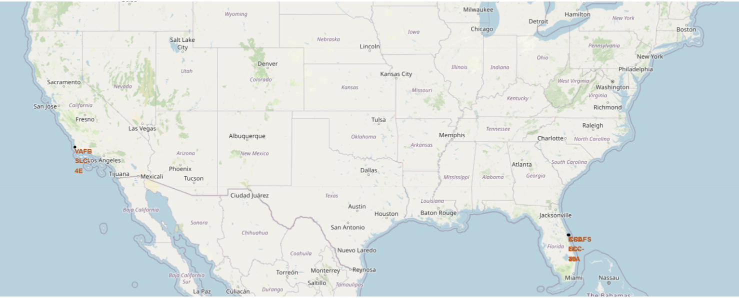
Landing _Outcome	quantity of outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Success (ground pad)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	1
Precluded (drone ship)	1

***Explanation*.**

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.

Launch Sites Proximities Analysis

All Sites Location Markers on a Global Map



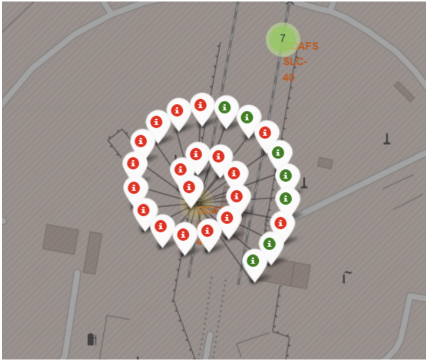
Explanation.

The launch sites shown in the figure are highly optimal and recommended for the following reasons:

(✓) These launch sites are close to the equator and the earth moves faster here than anywhere else.

(✓) These launch sites are very close to the coast, and this minimizes the risk of objects falling or exploding close to people.

All Sites Location Markers on a Global Map



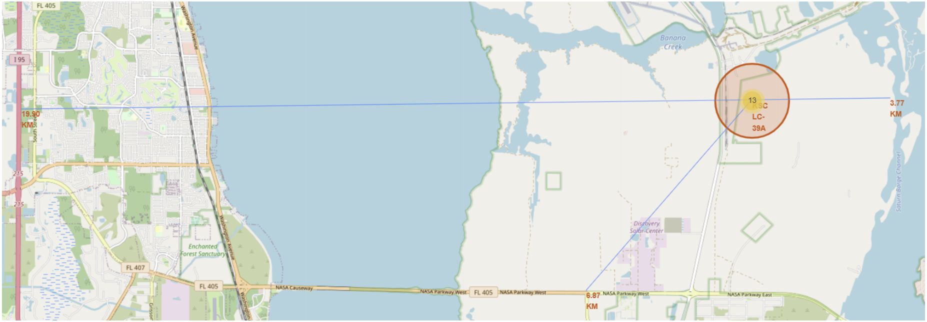
Explanation.

From the color-labeled markers we should be able to easily identify which launch sites have relatively high success rates, because:

$$\begin{cases} \textit{Green Marker} = \text{Successful Launch}, \\ \textit{Red Marker} = \text{Failed Launch}. \end{cases}$$

That's why the launch site CCAFS LC - 40 has a very low success rate.

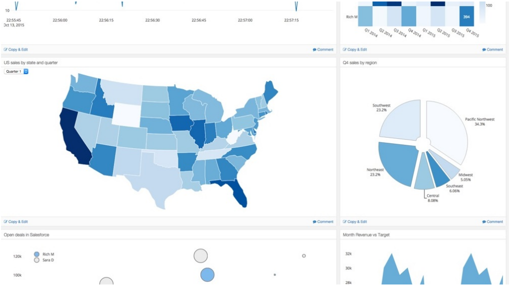
Distance Between the Launch Site KSC LC-39A to the Railway, Highway and Coastline



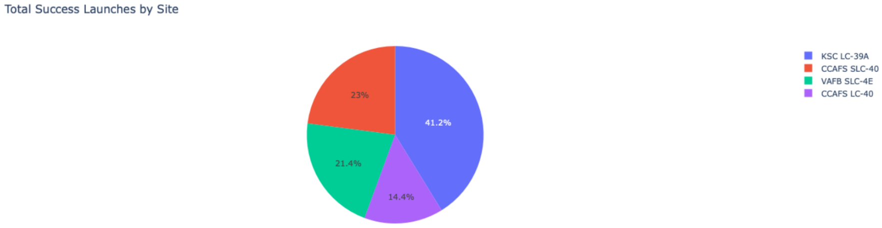
Explanation.

After having chosen the launch site KSC LC - 39A, we have decided to describe the distance between the highway, the coast and the nearest railway to this launch site. For this, we have appropriately chosen points of highway, coast and railway which apparently have minimum distance to KSC LC - 39A.

Build a Dashboard with Plotly Dash



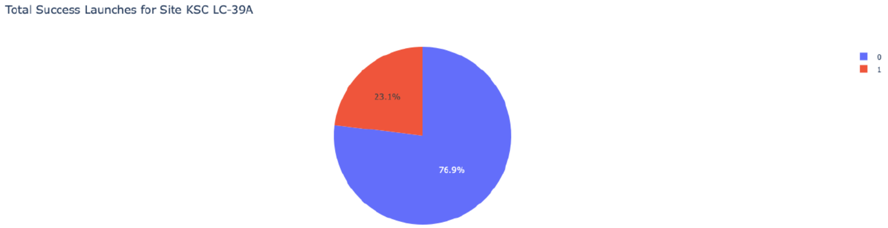
Launch Success Count for All Sites



Explanation.

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

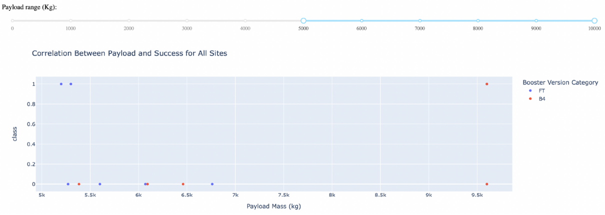
Launch Site with Highest Launch Success Ratio



Explanation.

KSC LC-39A has a launch success rate of 76,9%. Which means statistically that out of every 100 launches, approximately 76 are successful.

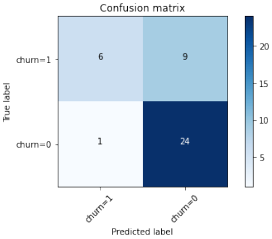
Payload vs. Launch Outcome for All Sites



Explanation.

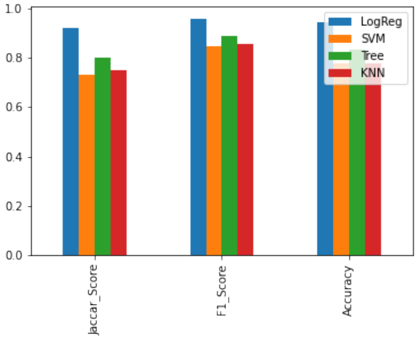
The charts show that payloads between 2000 and 5500 kg have the highest success rate.

Predictive Analysis (Classification)



Classification Accuracy

	LogReg	SVM	Tree	KNN
Jaccar_Score	0.923077	0.733333	0.800000	0.750000
F1_Score	0.960000	0.846154	0.888889	0.857143
Accuracy	0.944444	0.777778	0.833333	0.777778



Explanation.

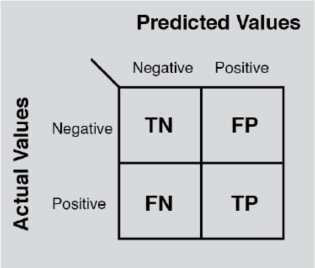
From the above results, we can conclude that the model that has the highest classification accuracy is LogReg (logistic regression).

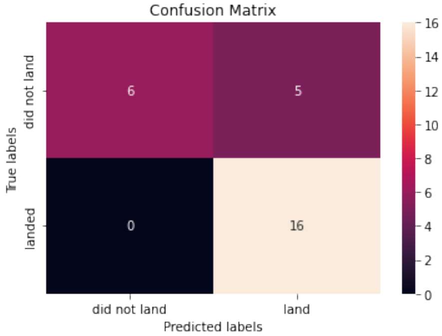
GitHub URL: Classification Accuracy

Confusion Matrix

Explanation (LogReg).

Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is true negatives





GitHub URL: Classification Accuracy

Conclusions

Conclusions

($\checkmark$) Logistic regression model is the best algorithm for our dataset.

($\checkmark$) Launches with a low payload mass show better results than launches with a larger payload mass.

($\checkmark$) The launch site *CCAFS SLC 40* had more than 50% of the total launches. This means that for some reason, the site *CCAFS SLC 40* was the preferred one for launching rockets with payload.

($\checkmark$) The site *KSC LC 39A* had a 100% success for payload mass under 5000 kg.

($\checkmark$) The site *VAFBSLC4E* had a 100% success for payload mass between 1000 kg and 9000 kg.

($\checkmark$) As the number of flights increases, each new launch has a higher probability of success

($\checkmark$) As the number of flights increases, it is found that flights over orbits tend to be more successful.

(✓) The orbits with 100 % success rate are:

$$\begin{cases} \text{ES-L1}, \\ \text{GEO}, \\ \text{HEO}, \\ \text{SSO}. \end{cases}$$

(✓) The orbit with 0 % success rate is SO.

(✓) These launch sites are close to the equator and the earth moves faster here than anywhere else.

(✓) These launch sites are very close to the coast, and this minimizes the risk of objects falling or exploding close to people.

Appendix

Appendix

(✓) SpaceX Falcon 9 first stage Landing Prediction.

(✓) Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia

(✓) Exploratory Data Analysis (EDA).

(✓) Insights drawn from EDA.

(✓) Insights drawn from SQL.

(✓) Launch Sites Locations Analysis with Folium.

(✓) Dashboard with Plotly Dash.

(✓) Machine Learning Classification.

(✓) Classification Accuracy.