IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Enrique Alcoceba
24/09/2023

Outline

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

Executive Summary

- Summary of methodologies

In this project the following methodologies have been used:

-Collecting the Data:

-**Data Collection** with **Web Scraping.**

-**Data Wrangling**

-Exploratory Data Analysis (EDA) using:

-**SQL.**

-**Data Visualization** with Pandas, Matplotlib & Seaborn.

-Predictive Analysis with **Machine Learning** using 4 types of algorithms:

-Logistic Regression.

-Support Vector Machine (SVM).

-Decission Tree.

-K Nearest Neighbors (KNN).

- Summary of all results

Executive Summary

- Summary of methodologies
- Summary of all results

In the Exploratory Data Analyssis it is first seen the correlations between variables and the landing success of the first stage.

An Interactive analytics demo was successfully performed, showing the succes rate in piecharts for each launch location and the proportion of launches between launch sites. A scatter chart of succes rate depending on payload mass, that indicates the booster version of the rocket, and let the user change the range of payload mass was performed as well.

Finally in the predictive analysis results we can determine with, at most, 0,8334 accuracy on test set whether if the stage one of the rocket will be able or not to land succesfully.

Introduction

- Project background and context
- Problems you want to find answers

The **commercial space age** is **here**, companies are making space travel **affordable** for everyone.

Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets. Perhaps the most successful is **SpaceX**.

SpaceX's accomplishments include: Sending spacecraft to the International Space Station. Starlink, a satellite internet constellation providing satellite Internet access. Sending manned missions to Space.

One reason SpaceX can do this is the **rocket launches are relatively inexpensive.** SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, **much of the savings is because SpaceX can reuse the first stage.**

Introduction

- Project background and context
- Problems you want to find answers

Determine the price of each launch through:

- **Creating dashboards for the team.**
- **Determine if the first stage will land**, so we can reuse it and determine the overall cost of the launch. This can be done through **predictive analysis with machine learning's** techniques.
- It is also important **determine launch site location and its requirements** for the new spaceship company.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - How data was collected
- Perform data wrangling
  - How data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How it was built, tuned and evaluated the classification models

Data Collection

- **The steps followed for data collection have been:**

  1. Request and parse the SpaceX launch data.
     - GET request was used in a static response object to make the response results more consistent. After the response was decoded into a Json file to turn it into a Pandas dataframe.
  2. The data was filtered to only contain Falcon 9 launches.
  3. Small data wrangling was performed to deal with Pay Load Mass missing values.

Data Collection – SpaceX API

https://github.com/gusuno/DS_Capstone/blob/c0debffec49c181438349e5b10d416030c3a9d62/jupyter-labs-spacex-data-collection-api.ipynb

jupyter-labs-spacex-data-cc × | jupyter-labs-webscraping.ip × | spacex_web_scraped.csv × | dataset_part_1.csv × | +

Delimiter: ,

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPa
1	1	2010-06-04	Falcon 9	6123.547647058824	LEO	CCSFS SLC 40	None None	1	False	False	False	
2	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	
3	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	
4	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	
5	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
6	6	2014-01-06	Falcon 9	3325.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
7	7	2014-04-18	Falcon 9	2296.0	ISS	CCSFS SLC 40	True Ocean	1	False	False	True	
8	8	2014-07-14	Falcon 9	1316.0	LEO	CCSFS SLC 40	True Ocean	1	False	False	True	
9	9	2014-08-05	Falcon 9	4535.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
10	10	2014-09-07	Falcon 9	4428.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
11	11	2014-09-21	Falcon 9	2216.0	ISS	CCSFS SLC 40	False Ocean	1	False	False	False	
12	12	2015-01-10	Falcon 9	2395.0	ISS	CCSFS SLC 40	False ASDS	1	True	False	True	5e9e3032383ecb761...
13	13	2015-02-11	Falcon 9	570.0	ES-L1	CCSFS SLC 40	True Ocean	1	True	False	True	
14	14	2015-04-14	Falcon 9	1898.0	ISS	CCSFS SLC 40	False ASDS	1	True	False	True	5e9e3032383ecb761...
15	15	2015-04-27	Falcon 9	4707.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
16	16	2015-06-28	Falcon 9	2477.0	ISS	CCSFS SLC 40	None ASDS	1	True	False	True	5e9e3032383ecb6bb...
17	17	2015-12-22	Falcon 9	2034.0	LEO	CCSFS SLC 40	True RTLS	1	True	False	True	5e9e3032383ecb267...
18	18	2016-01-17	Falcon 9	553.0	PO	VAFB SLC 4E	False ASDS	1	True	False	True	5e9e3033383ecbb9e...
19	19	2016-03-04	Falcon 9	5271.0	GTO	CCSFS SLC 40	False ASDS	1	True	False	True	5e9e3032383ecb6bb...
20	20	2016-04-08	Falcon 9	3136.0	ISS	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb...
21	21	2016-05-06	Falcon 9	4696.0	GTO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb...
22	22	2016-05-27	Falcon 9	3100.0	GTO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb...
23	23	2016-07-18	Falcon 9	2257.0	ISS	CCSFS SLC 40	True RTLS	1	True	False	True	5e9e3032383ecb267...
24	24	2016-08-14	Falcon 9	4600.0	GTO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb...
25	25	2016-09-01	Falcon 9	5500.0	GTO	CCSFS SLC 40	None ASDS	1	True	False	True	5e9e3032383ecb6bb...
26	26	2017-01-14	Falcon 9	9600.0	PO	VAFB SLC 4E	True ASDS	1	True	False	True	5e9e3033383ecbb9e...
27	27	2017-02-19	Falcon 9	2490.0	ISS	KSC LC 39A	True RTLS	1	True	False	True	5e9e3032383ecb267...
28	28	2017-03-16	Falcon 9	5600.0	GTO	KSC LC 39A	None None	1	False	False	False	
29	29	2017-03-30	Falcon 9	5300.0	GTO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb...
30	30	2017-05-01	Falcon 9	6123.547647058824	LEO	KSC LC 39A	True RTLS	1	True	False	True	5e9e3032383ecb267...
31	31	2017-05-15	Falcon 9	6070.0	GTO	KSC LC 39A	None None	1	False	False	False	
32	32	2017-06-03	Falcon 9	2708.0	ISS	KSC LC 39A	True RTLS	1	True	False	True	5e9e3032383ecb267...

Data Collection - Scraping

https://github.com/gusuno/DS_Capstone/blob/c0debffec49c181438349e5b10d416030c3a9d62/jupyter-labs-webscraping.ipynb

jupyter-labs-spacex-data-cc × | jupyter-labs-webscraping.ip × | spacex_web_scraped.csv × | dataset_part_1.csv × | +

Delimiter: ,

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
1	1	CCAFS	Dragon Spacecraft Q...	0	LEO	SpaceX	Success	F9 v1.0B0003.1	Failure	4 June 2010	18:45
2	2	CCAFS	Dragon	0	LEO	NASA (COTS) NRO	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
3	3	CCAFS	Dragon	525 kg	LEO	NASA (COTS)	Success	F9 v1.0B0005.1	No attempt	22 May 2012	07:44
4	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA (CRS)	Success	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
5	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA (CRS)	Success	F9 v1.0B0007.1	No attempt	1 March 2013	15:10
6	6	VAFB	CASSIOPE	500 kg	Polar orbit	MDA	Success	F9 v1.1B1003	Uncontrolled	29 September 2013	16:00
7	7	CCAFS	SES-8	3,170 kg	GTO	SES	Success	F9 v1.1	No attempt	3 December 2013	22:41
8	8	CCAFS	Thaicom 6	3,325 kg	GTO	Thaicom	Success	F9 v1.1	No attempt	6 January 2014	22:06
9	9	Cape Canaveral	SpaceX CRS-3	2,296 kg	LEO	NASA (CRS)	Success	F9 v1.1	Controlled	18 April 2014	19:25
10	10	Cape Canaveral	Orbcomm-OG2	1,316 kg	LEO	Orbcomm	Success	F9 v1.1	Controlled	14 July 2014	15:15
11	11	Cape Canaveral	AsiaSat 8	4,535 kg	GTO	AsiaSat	Success	F9 v1.1	No attempt	5 August 2014	08:00
12	12	Cape Canaveral	AsiaSat 6	4,428 kg	GTO	AsiaSat	Success	F9 v1.1	No attempt	7 September 2014	05:00
13	13	Cape Canaveral	SpaceX CRS-4	2,216 kg	LEO	NASA (CRS)	Success	F9 v1.1	Uncontrolled	21 September 2014	05:52
14	14	Cape Canaveral	SpaceX CRS-5	2,395 kg	LEO	NASA (CRS)	Success	F9 v1.1	Failure	10 January 2015	09:47
15	15	Cape Canaveral	DSCOVR	570 kg	HEO	USAF NASA NOAA	Success	F9 v1.1	Controlled	11 February 2015	23:03
16	16	Cape Canaveral	ABS-3A	4,159 kg	GTO	ABS Eutelsat	Success	F9 v1.1	No attempt	2 March 2015	03:50
17	17	Cape Canaveral	SpaceX CRS-6	1,898 kg	LEO	NASA (CRS)	Success	F9 v1.1	Failure	14 April 2015	20:10
18	18	Cape Canaveral	TürkmenÄlem 52°E / ...	4,707 kg	GTO	Turkmenistan Nation...	Success	F9 v1.1	No attempt	27 April 2015	23:03
19	19	Cape Canaveral	SpaceX CRS-7	1,952 kg	LEO	NASA (CRS)	Failure	F9 v1.1	Precluded	28 June 2015	14:21
20	20	Cape Canaveral	Orbcomm-OG2	2,034 kg	LEO	Orbcomm	Success	F9 FT	Success	22 December 2015	01:29
21	21	VAFB	Jason-3	553 kg	LEO	NASA (LSP) NOAA C...	Success	F9 v1.1	Failure	17 January 2016	18:42
22	22	Cape Canaveral	SES-9	5,271 kg	GTO	SES	Success	F9 FT	Failure	4 March 2016	23:35
23	23	Cape Canaveral	SpaceX CRS-8	3,136 kg	LEO	NASA (CRS)	Success	F9 FT	Success	8 April 2016	20:43
24	24	Cape Canaveral	JCSAT-14	4,696 kg	GTO	SKY Perfect JSAT Gro...	Success	F9 FT	Success	6 May 2016	05:21
25	25	Cape Canaveral	Thaicom 8	3,100 kg	GTO	Thaicom	Success	F9 FT	Success	27 May 2016	21:39
26	26	Cape Canaveral	ABS-2A	3,600 kg	GTO	ABS Eutelsat	Success	F9 FT	Failure	15 June 2016	14:29
27	27	Cape Canaveral	SpaceX CRS-9	2,257 kg	LEO	NASA (CRS)	Success	F9 FT	Success	18 July 2016	04:45
28	28	Cape Canaveral	JCSAT-16	4,600 kg	GTO	SKY Perfect JSAT Gro...	Success	F9 FT	Success	14 August 2016	05:26
29	29	VAFB	Iridium NEXT	9,600 kg	Polar	Iridium Communicati...	Success	F9 FT	Success	14 January 2017	17:54
30	30	KSC	SpaceX CRS-10	2,490 kg	LEO	NASA (CRS)	Success	F9 FT	Success	19 February 2017	14:39
31	31	KSC	EchoStar 23	5,600 kg	GTO	EchoStar	Success	F9 FT	No attempt	16 March 2017	06:00
32	32	KSC	SES-10	5,300 kg	GTO	SES	Success	F9 FT△	Success	30 March 2017	22:27
33	33	KSC	NROL-76	C	LEO	NRO	Success	F9 FT	Success	1 May 2017	11:15

Data Wrangling

- First data wrangling performed was to change all NaN values in Pay Load Mass column for the mean of said column.
- Also, convert those outcomes, where the booster did or did not land successfully, into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.

https://github.com/gusuno/DS_Capstone/blob/c0debffec49c181438349e5b10d416030c3a9d62/jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Scatter plot type was used 5 times to visualize the correlation between 2 values:
  - FlightNumber vs. PayloadMass
  - FlightNumber vs LaunchSite
  - PayloadMass vs LaunchSite
  - FlightNumber vs Orbit
  - PayloadMass vs Orbit
- Bar chart plot was used to see the relationship between success rate of each orbit type.
- Line chart plot was used to visualize the trend throughout the years of success rate.

https://github.com/gusuno/DS_Capstone/blob/dc10f8ee8d63eff0c7f28148a5de427aeb2be035/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

Summary of the SQL queries performed:

- 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).
- Display average payload mass carried by booster version F9 v1.1.
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- List the total number of successful and failure mission outcomes.
- List the names of the booster_versions which have carried the maximum payload mass.
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank 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.

https://github.com/gusuno/DS_Capstone/blob/dc10f8ee8d63eff0c7f28148a5de427aeb2be035/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers and circles were used to represent launching sites on the map.
- Also the success and failed launches were marked and clustered to be able to see how many they were in each location and, when zooming in if they succeeded or not.
- Finally, lines were added to represent the distance to the nearest services (closest city, railway, highway) from the launch site.

https://github.com/gusuno/DS_Capstone/blob/8dea84ea379a9ae89aa64f263441dfa5f55bf1be/jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- A drop down menu was added to select between different options of the launch sites so you could see different success rate's pie charts.
- Pie chart was added with a callback function.
- A slider was added to select payload range that will be displayed in the scatter chart.
- A scatter chart was added to represent the correlation between payload and launch success with a callback function.

https://github.com/gusuno/DS_Capstone/blob/8dea84ea379a9ae89aa64f263441dfa5f55bf1be/spacex_dash_app.py

Predictive Analysis (Classification)

1. The class column was extracted to creat a numpy array so we have the Y matrix.
2. The rest information was standardized to conform the X Matrix.
3. The train_test_split was performed with test_size=0.2 and random_state=2.
4. Algorithms explored are Logistic Regression, SVM, Decission Tree and KNN.
5. With the help of the GridSearchCV function the best parameters for the algorithms performed were searched. With cv=10.
6. With the best parameters found, the accuracy was calculated, first for train accuracy, and after for the test split of the data.
7. To add value and ease confusion matrixes were performed for each algorithm results.

https://github.com/gusuno/DS_Capstone/blob/069b04bc79cf88c2f3818d54946f2290c537fdd0/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

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

Results

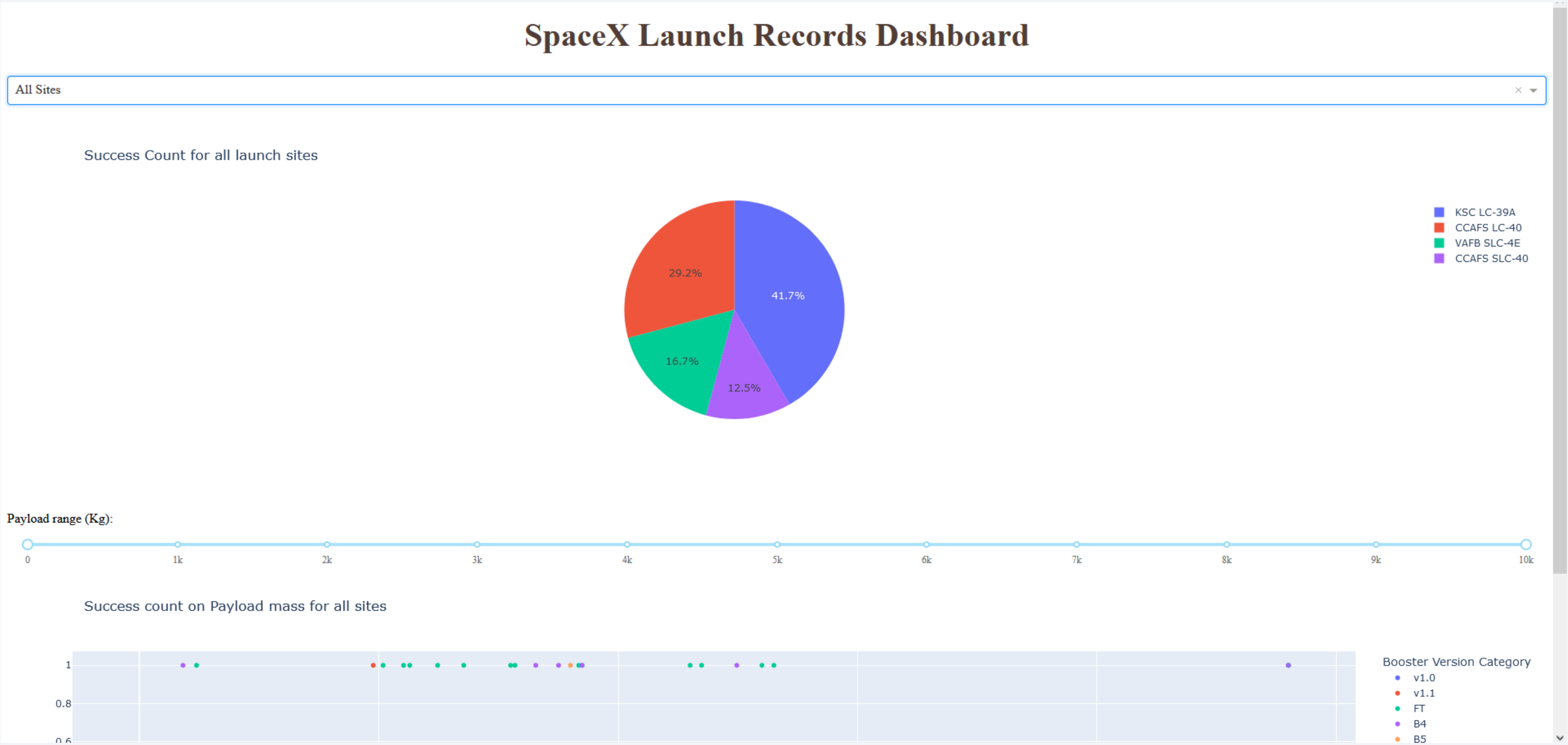
- Exploratory data analysis results
    - EDA with Data Visualization
    - EDA with SQL
- The different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- The higher the flight number is the more likely it is to succeed.
- For VAFB SLC 4E and KSC LC-39A the higher the payloadmass the more likely it is to succeed, but payloadmass for CCAFS LC-40 does not seem to be related to success.
- For the orbits ES-L1, GEO, HEO and SSO the success rate is perfect. VLEO orbit also has a great success rate.
- In the LEO orbit the success seems to be related to the number of flights. On the other hand, there appears to be no relationship between flight number when in GTO orbit.
- With heavy payloads the successful landing or positive landing rate are higher for Polar, LEO and ISS. However for GTO we cannot distinguish well, since both positive and negative landings occur.
- The sucess rate kept increasing since 2013 till 2020.

Results

- Exploratory data analysis results
    - EDA with Data Visualization
    - EDA with SQL
- The total payload mass carried by boosters launched by NASA (CRS) is 45596 kg.
- The average payload mass carried by booster version F9 v1.1 is 2534.66 kg.
- The first succesful landing outcome in ground pad was acheived in 2015-12-22.
- The mission outcome is majoritarily to be success with only 1 failure in flight.
- The landing outcomes were ranked from "No attempt" which occurred the most times (10 times).
- A list with the names of the booster_versions which have carried the maximum payload mass was created.
- A list the records which will display the month, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015 was also created, only displaying 2 occasions.

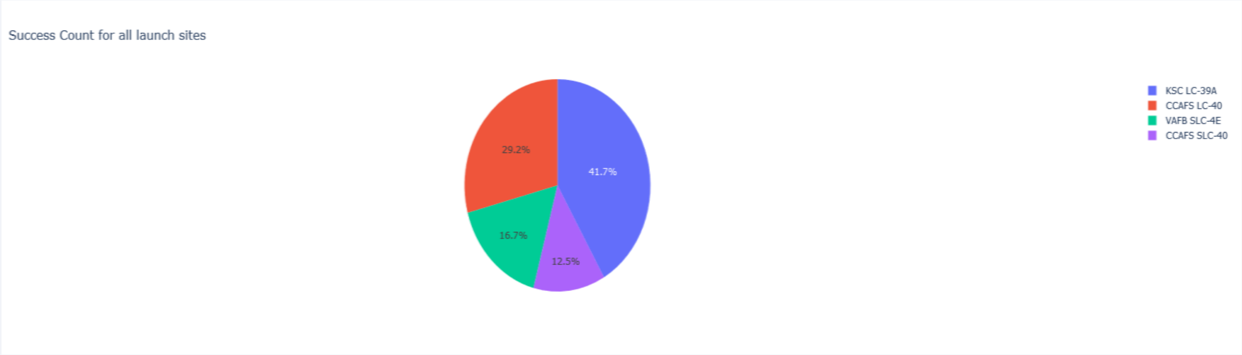
Results

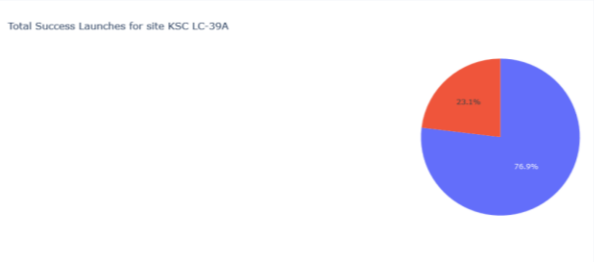
- Interactive analytics demo in screenshots

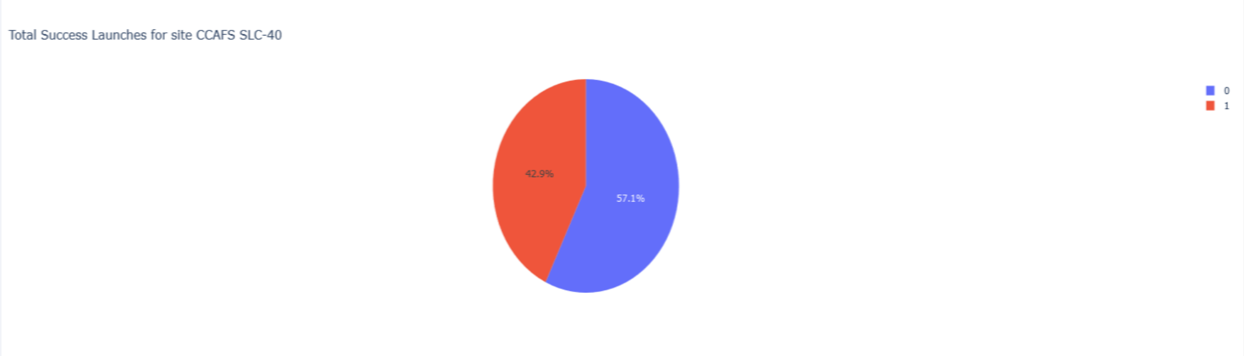


Results

- Interactive analytics demo in screenshots







Results

- Interactive analytics demo in screenshots

Results

- Predictive analysis results

In the predictive results we get basically, the results of best parameters for each algorithm tested and the accuracy and confusion matrixes of each algorithm.

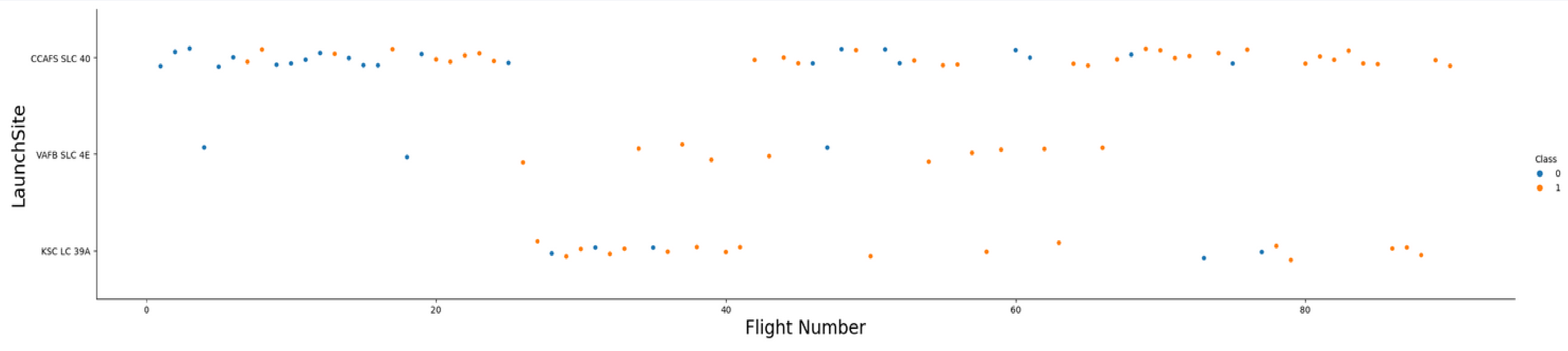
- Best parameters obtained for each:
  - **Logistic regression:** {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
  - **SVM:** {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
  - **Decission tree:** {'criterion': 'gini', 'max_depth': 4, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'random'}
  - **KNN:** {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
- Accuracy results:
  - **For all but the decission tree** the test set accuracy obtained was **0.8334**, while the **decission tree** got a **0.7778**, performing worse than the rest.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

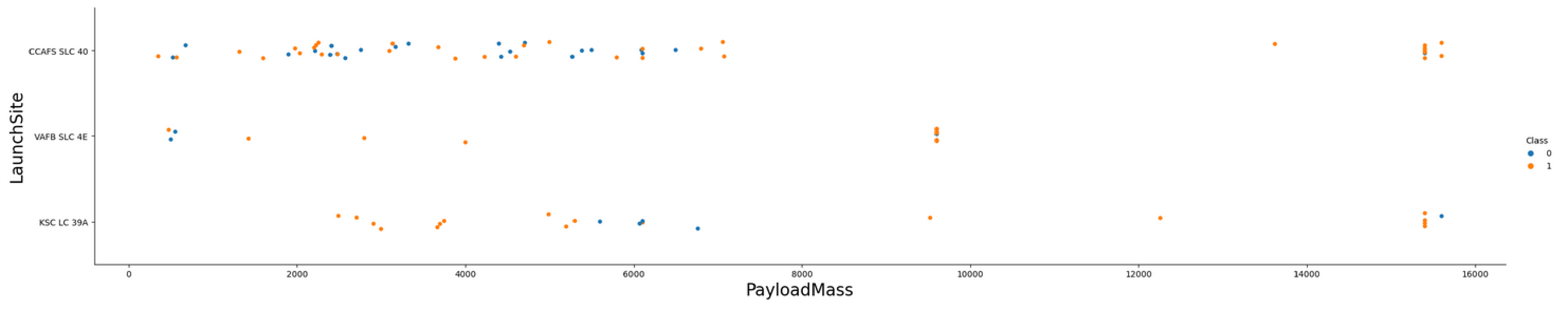
Scatter plot of Flight Number vs. Launch Site



We can see that for all three launch sites it seems that the higher the flight number is the more likely it is to succeed, specially for VAFB SLC 4E and KSC LC 39A.

Payload vs. Launch Site

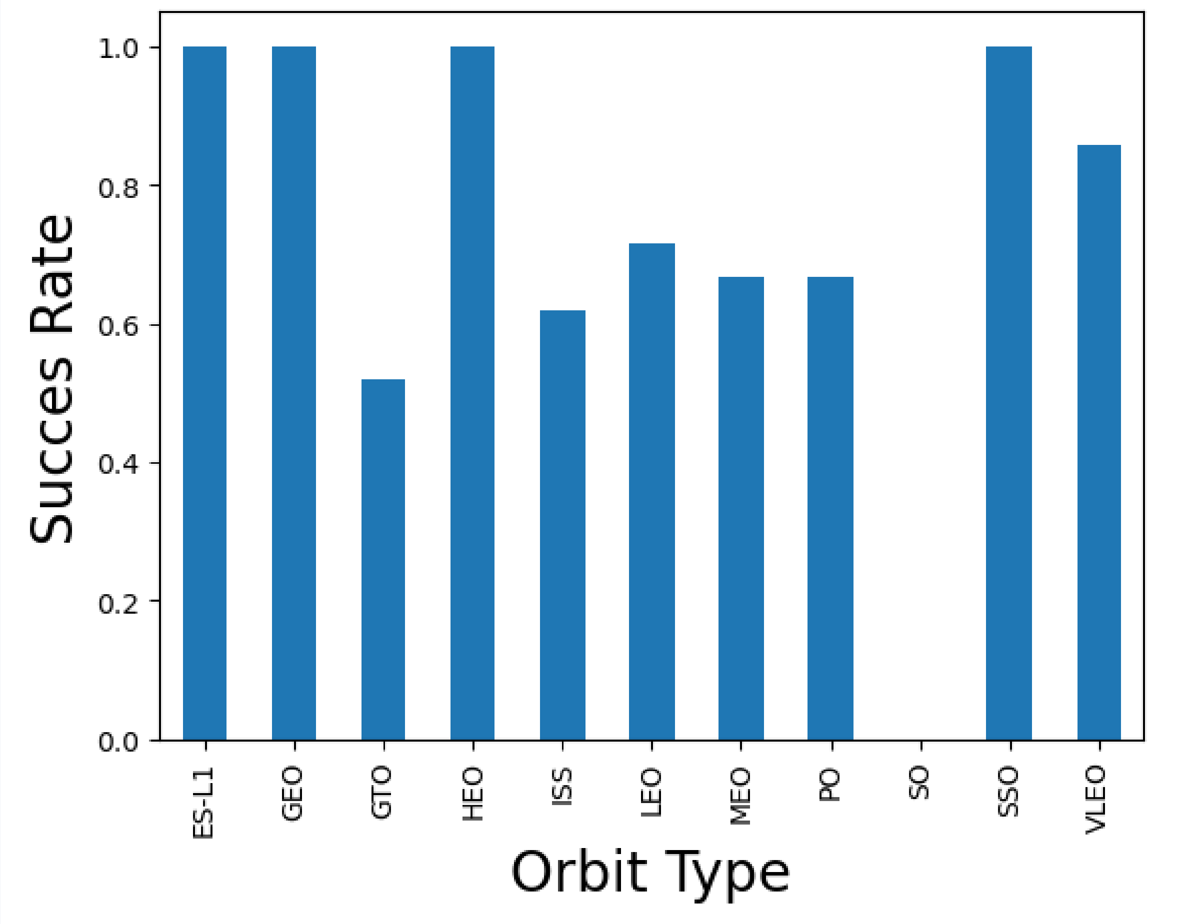
Scatter plot of Payload vs. Launch Site



For the launch sites CCAFS SLC 40 and VAFB SLC 4E the higher the payload the higher the chance to succeed. On the other hand for the KSC LC 39A launch site it seems that payloads around 6000kg were prone to failure.

Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit type



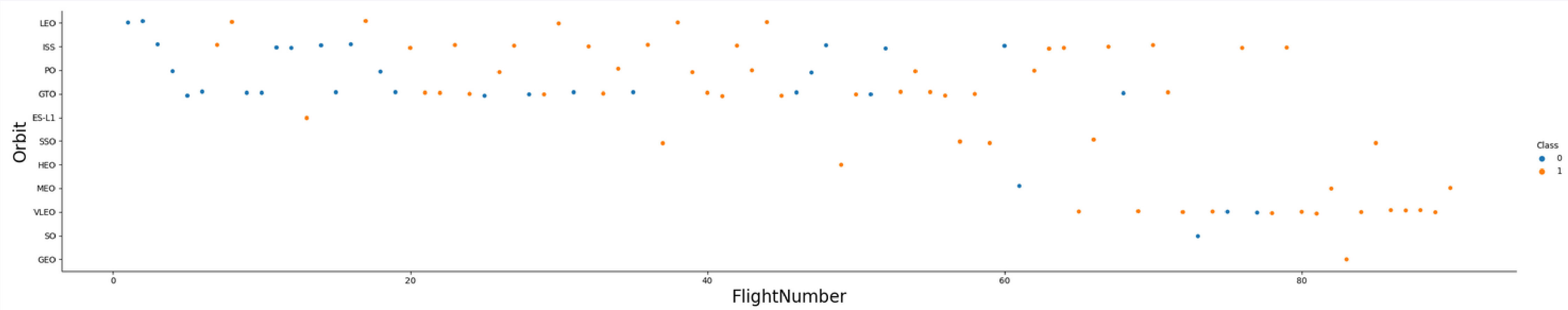
We can see that the orbits ES-L1, GEO, HEO and SSO the success rate is perfect (~100%). VLEO orbit also has a great success rate (~80%).

On the other hand we have the orbit SO with a nule success rate (~0%).

And the rest vary between a ~75% and ~55% succes rate.

Flight Number vs. Orbit Type

Scatter point of Flight number vs. Orbit type

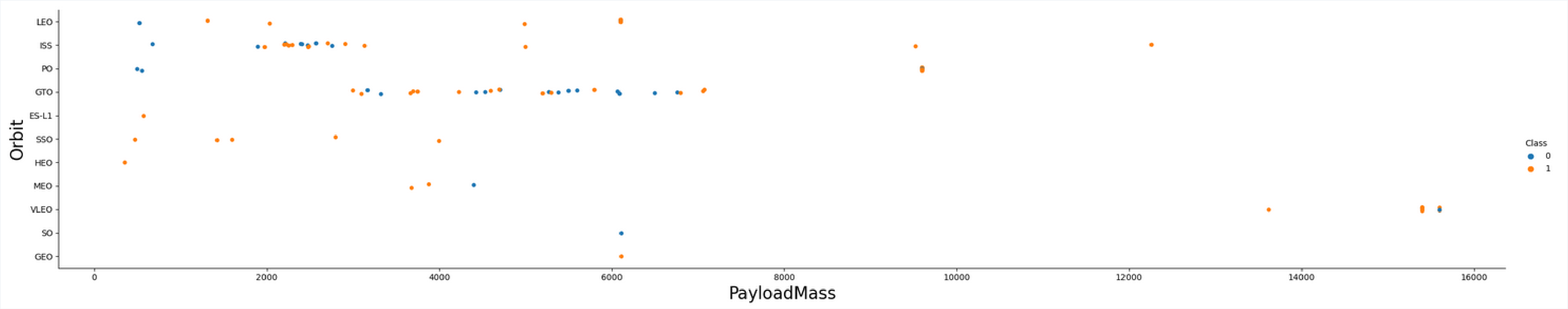


In the LEO orbit the success seems to be related to the number of flights. On the other hand, there appears to be no relationship between flight number when in GTO orbit.

And the SSO, HEO and VLEO orbit flights tend to succeed.

Payload vs. Orbit Type

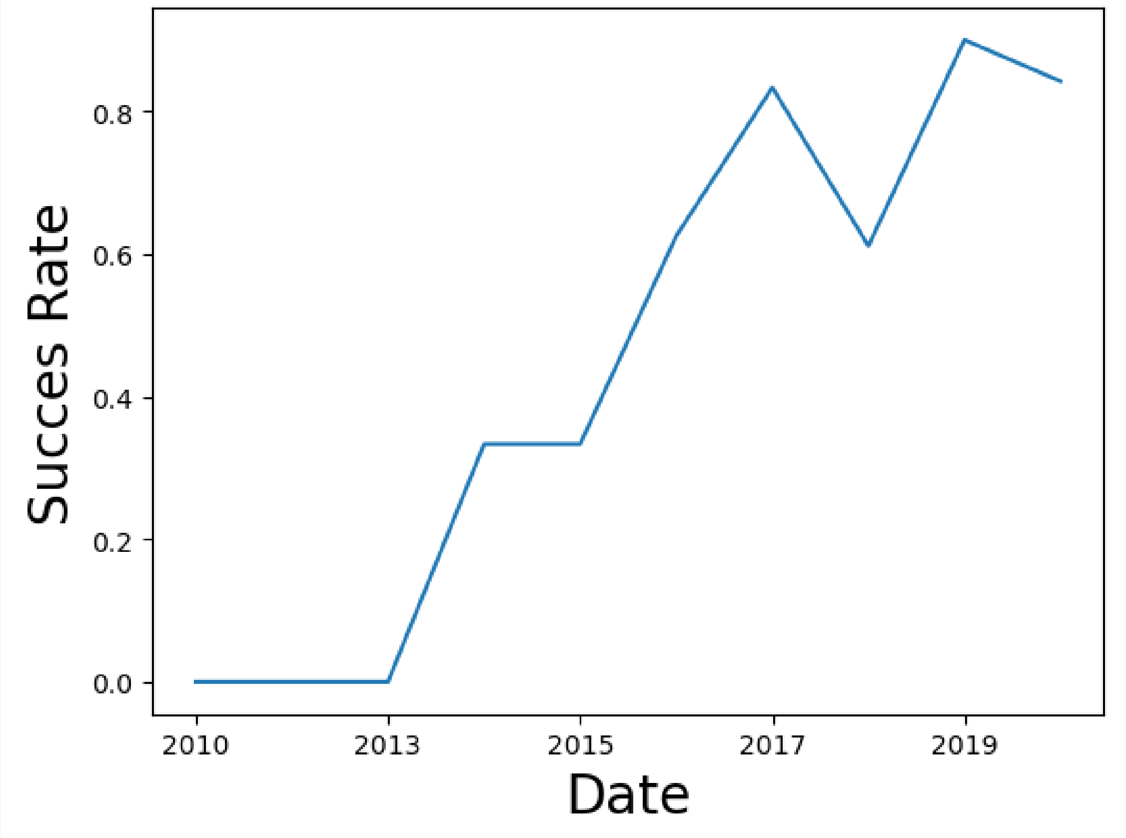
Scatter point of payload vs. orbit type



With heavy payloads the successful landing or positive landing rate are higher for Polar, LEO and ISS. However for GTO we cannot distinguish well, since both positive and negative landings occur.

Launch Success Yearly Trend

Line chart of yearly average success rate



We can observe that the success rate kept increasing since 2013 till 2020.

All Launch Site Names

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

```
[8]: %sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;
```

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

[8]:

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

Here we get a table with the launch sites' names.

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string '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	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Here we get a table with 5 records of launch sites which name begins with 'CAA'.

Total Payload Mass

Task 3

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

```
[10]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer='NASA (CRS)';
```

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

[10]:

SUM(PAYLOAD_MASS__KG_)
45596

Here the sum of all payload mass carried by boosters launched by NASA (CRS) are displayed.

Average Payload Mass by F9 v1.1

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

```
[11]: %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1%';
```

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

[11]:

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

The average payload mass carried by booster version F9 v1.1 was calculated.

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
[12]: %sql SELECT MIN(Date) FROM SPACEXTBL WHERE Landing_Outcome LIKE '%Success%';
```

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

[12]:

MIN(Date)
2015-12-22

The first date with landing outcome in ground pad succesful was listed and displayed.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

[13]:
```
%sql SELECT DISTINCT Booster_Version FROM SPACEXTBL\
WHERE Landing_Outcome = 'Success (drone ship)' AND 4000<PAYLOAD_MASS__KG_<6000;
```

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

[13]:

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

List of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000.

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

```
[14]: %sql SELECT Mission_Outcome, COUNT(Mission_Outcome) AS TOTAL_NUMBER\
FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

[14]:

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

Total number of successful and failure mission outcomes was calculated and displayed.

With an overwhelming success rate.

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

[15]:
```
%%sql
SELECT DISTINCT Booster_Version
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_=(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

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

[15]:

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

List of the names of the booster which have carried the maximum payload mass

2015 Launch Records

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

[17]:
```
%%sql
SELECT SUBSTR(Date,6,2) AS Month, Landing_Outcome, Booster_Version, Launch_Site
FROM SPACEXTBL
WHERE Landing_Outcome='Failure (drone ship)' AND Date LIKE '2015%'
ORDER BY Month;
```

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

[17]:

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

List of the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.

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

Rank 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.

[18]:
```
%%sql
SELECT Landing_Outcome, COUNT(Landing_Outcome) AS SUM_Landings
FROM SPACEXTBL
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY SUM_Landings DESC;
```

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

[18]:

Landing_Outcome	SUM_Landings
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

The rank of the count of landing outcomes between 2010-06-04 and 2017-03-20, in descending order.

We can see that the landing outcome more common is the "No attempt" one.

We can also see that between those dates the ground pad landing was the one more prone to succeed, since there is no failure in ground pad.

Section 3

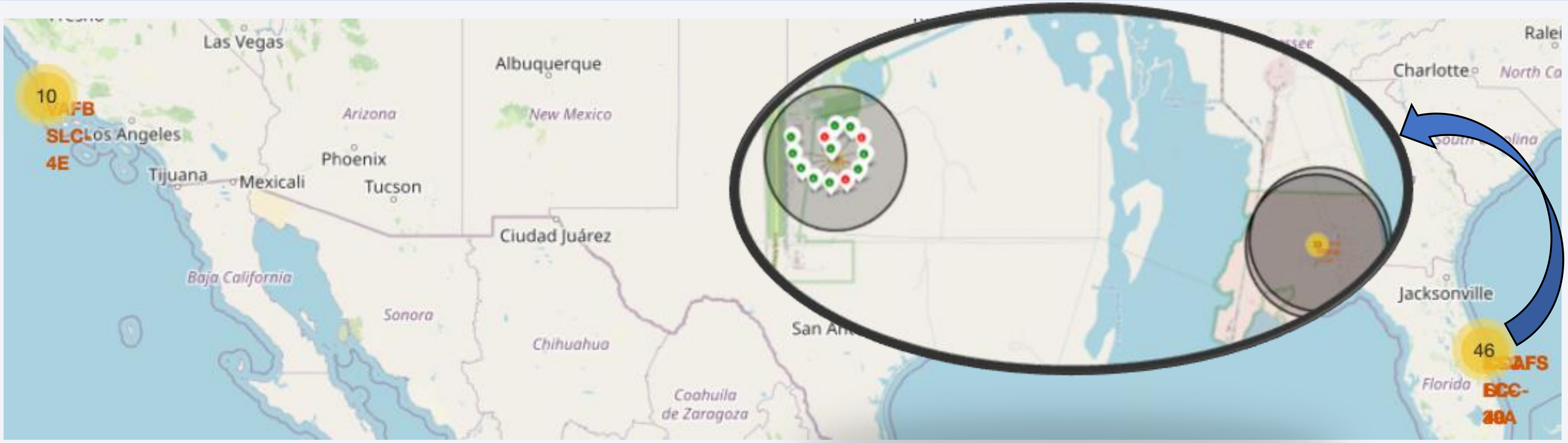
Launch Sites Proximities Analysis

Sites Location Folium Map

We can see the three launch sites, one to the left and the other 2 to the right almost together.

Here we can visualize that the launching sites are located the close enough to the equator and next to the coast.

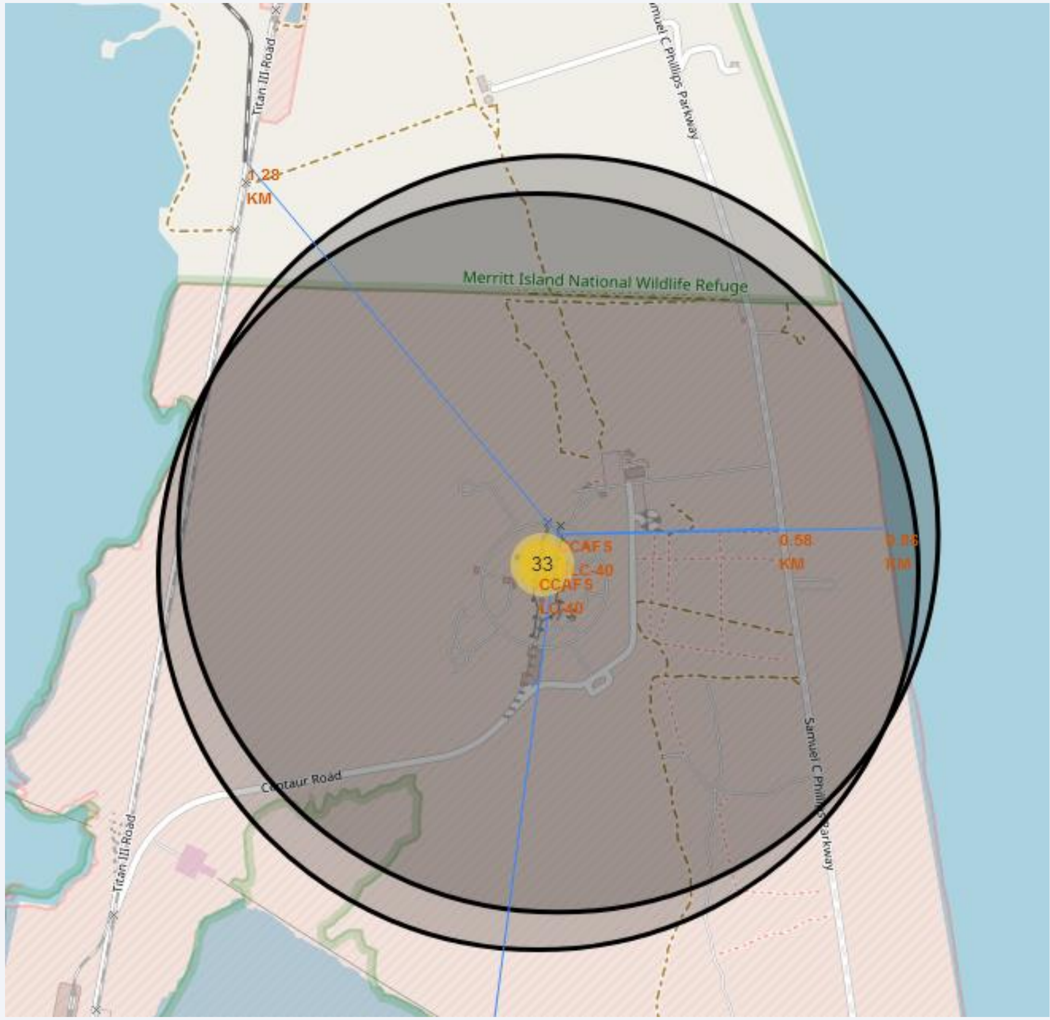
Circle sites location and success state map



In this image we can see 2 things, first circles for launching sites were implemented, thus the will tell us the number of launches per circle. And if you zoom in, it can also tell you the successful and failed launches.

Said that, we can also appreciate that the number of launches in Florida almost quintuple the ones in California, that could be due to being closer to the equator, or other variables.

Services proximity Folium Map



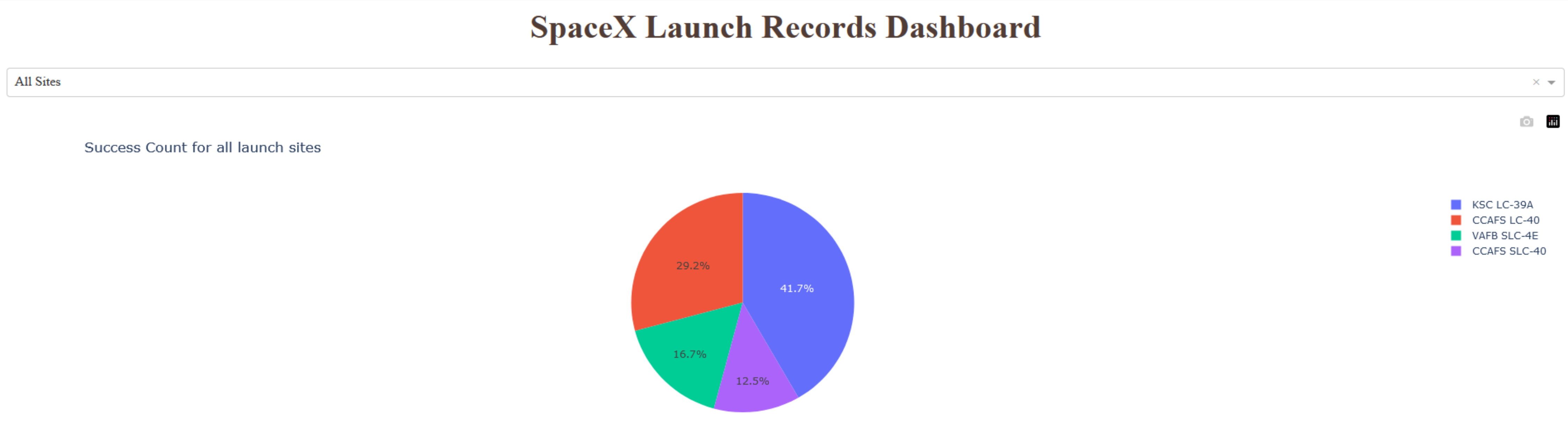
In the last and final map drawn, we can see blue lines directed to the closest coastline, city and railway or highway.

Here we can easily visualize that the launch sites are extremely close to the coast and to a railway or highway (very likely to ease the supply chain). While the closest city tend to be quite far away in comparison.

Section 4

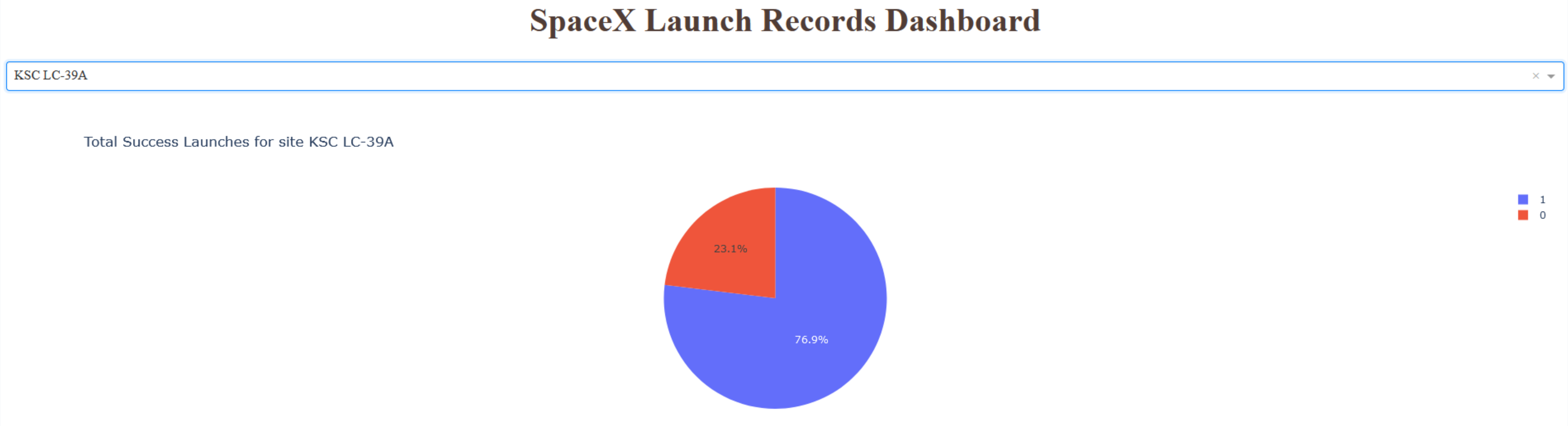
Build a Dashboard with Plotly Dash

Launch Records Dashboard Pie Chart



In this screenshot we can see that the site KSC LC-39A is the one most used, followed by CCAFS LC-40. In between these 2 sum almost the 70% of all launches.

Dashboard Pie chart with higher Success Rate



First we can see that the launch site with the higher success rate is the KSC LC-39A. And we also seen in the slide before that it is the launch site with the most launches.

Different Payload range comparison Dashboard

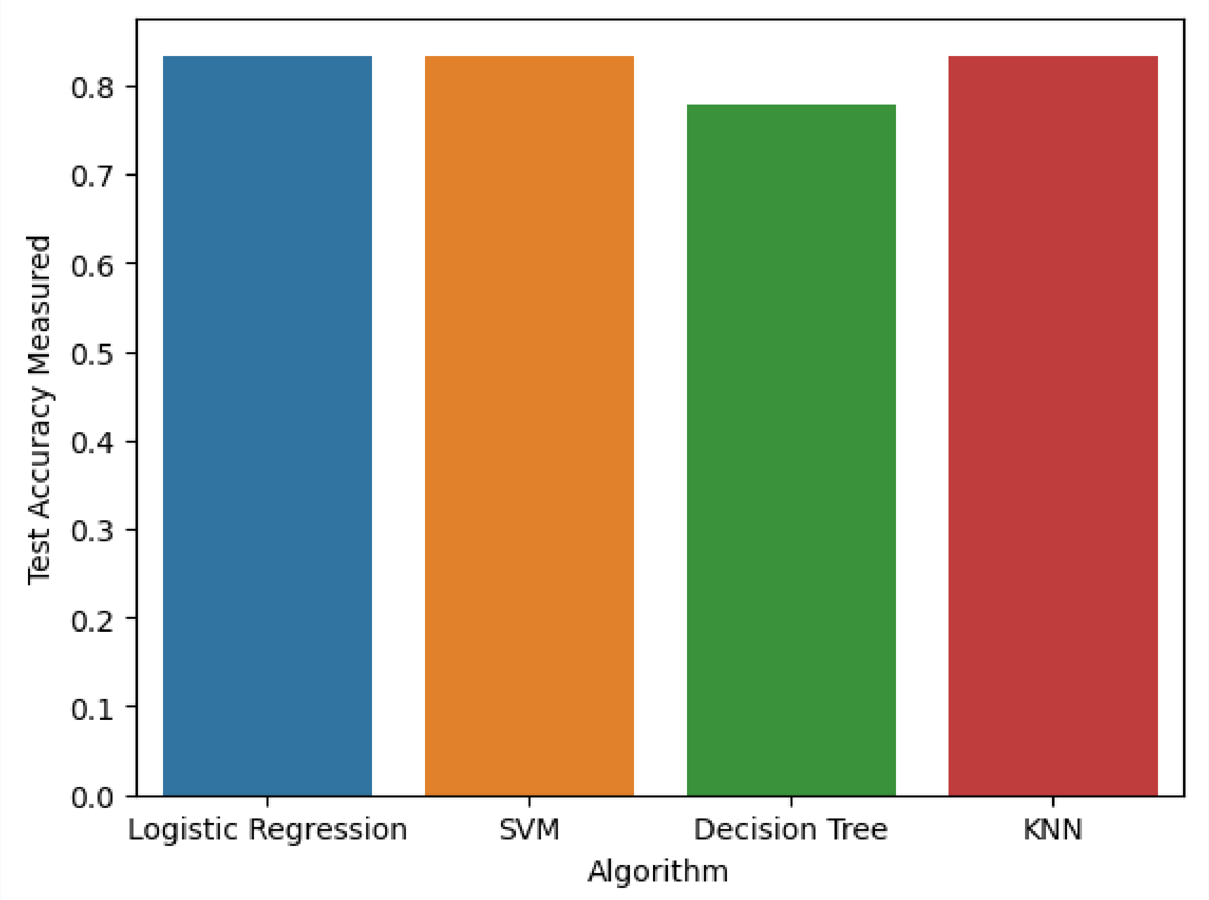
First we have a payload range to choose from. I chose the entire range, the first 5000 kg, and the last 5000kg respectively. Then we have the scatter chart representing the class vs payload and the legend of booster version on the left.

Said that, we can clearly see that all launches are a success with those three booster versions till the payload surpasses ~5500 kg. After this payload mass the class changes to 0, which means failure.

Section 5

Predictive Analysis (Classification)

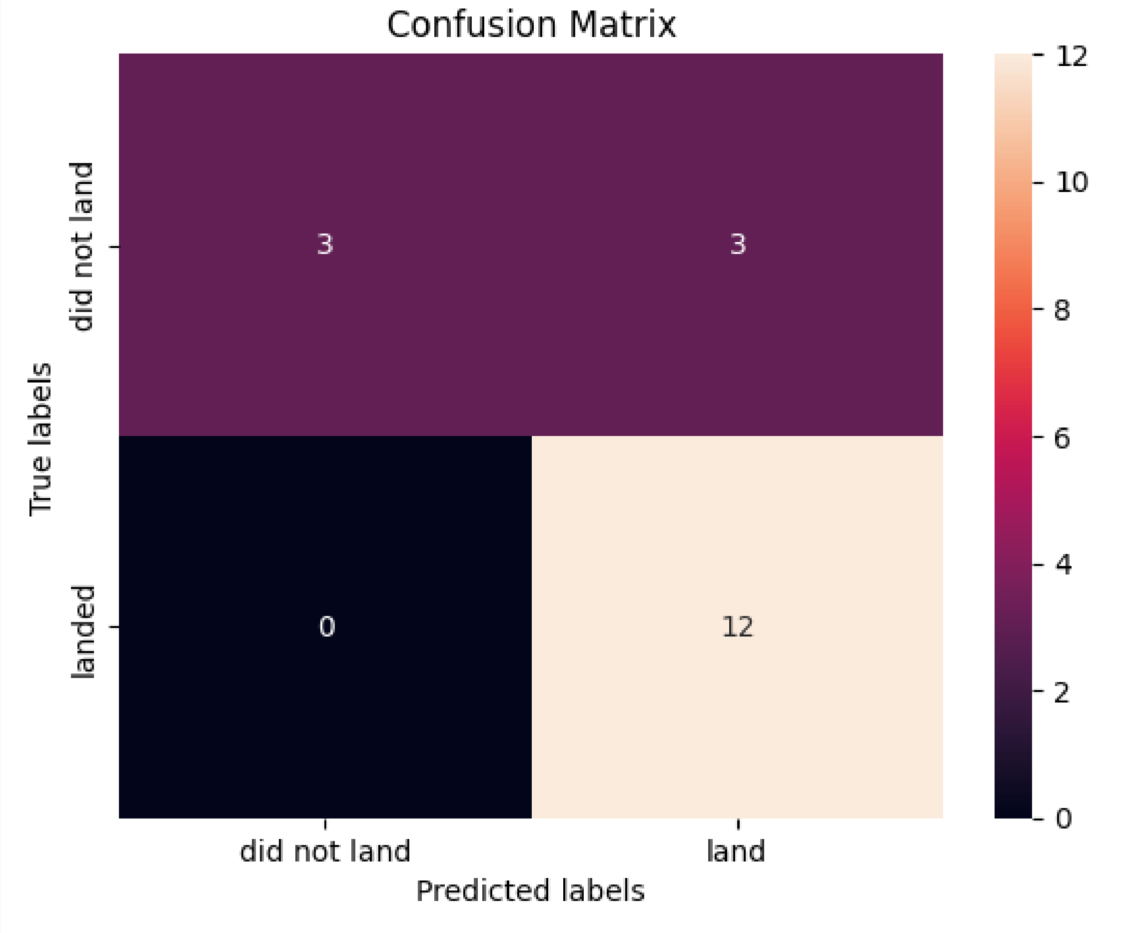
Classification Accuracy



We can see in the bar chart shown that all three Logistic Regression, SVM and KNN, performs equally in regards to test accuracy. Thus, all of them as good to perform in out of sample data.

The only one that underperforms is the decision tree algorithm.

Confusion Matrix



The resulting confusion matrix of the Logistic Regression, SVM and KNN algorithms are all the same, which is the one shown.

We can see that the landed results are categorized perfectly, whereas the ones that actually did not land are equally categorized as did not land and land.

In other words there are a lot of false negatives and no false positive.

Conclusions

- **Data collection** and **data wrangling** were successfully performed to pursue with the analysis properly.
- **EDA** performed clarify the relations between variables.
- **Launch sites proximities** analysis made clear that a route to bring resources close to the launch sites and long distance to closest city is heavily recommended and close proximity to the coast compulsory.
- **Dashboards** to inform peer team members were made. And they through important insights about launch sites as well as payload range and their success rates.
- Finally, **machine learning** was performed using 4 different algorithms, 3 of them resulted to be the best to perform the **predictive analysis** with no apparent false positives. Therefore, when success is predicted it will very likely be a success.

Thank you!