Winning Space Race with Data Science

Yann Roblin
15/01/2024

Outline

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

Executive Summary

- Data will be collected from with requested to the SpaceX API. Performing data wrangling, exploratory data analysis (EDA) using visualization and SQL, interactive visual analytics using Folium and Plotly Dash, predictive analysis using classification models, success rate for every launch configuration will be investigated.
- We can then conclude that a launch from KSC LC-39A will more likely success than from any other site, the best orbit are ES-L1, GEO, HEO and SSO, the highest success rate is reached between 1952 and 5300kg and the booster versions B5 and FT show the highest launch success rate.

Introduction

- Nowadays few companies are providing rocket launches for sending spacecraft and satellites. The main advantage of SpaceX compared to other providers for space launches is the re-use of the first stage of its Falcon 9 rocket. By knowing if the first stage will land, the cost of a launch can be determined. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- The goal of this study is to predict if the Falcon 9 first stage will land successfully, depending on launch site, pay load mass, orbit and booster. To achieve it, a machine learning model and public information will be used.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Data are requested to the SpaceX API
- Perform data wrangling
    - From the collected data, outcome (failure/success) are identified
- 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 to build, tune, evaluate classification models

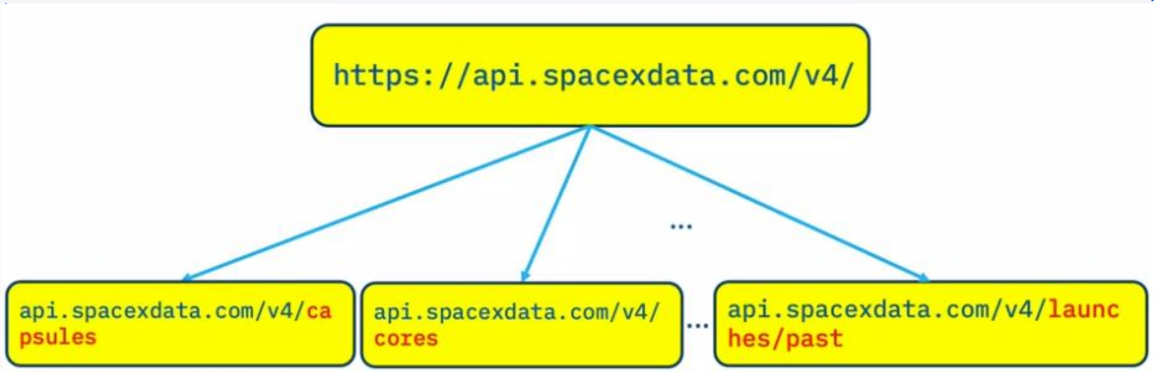
Data Collection

- Data are collected from the SpaceX REST API

- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Data are requested to the SpaceX API to do data wrangling and formatting from with api.spacexdata.com/v4/ to get capsules, cores and launches/past endpoints.
- Data will be cleaned afterwards.



- https://github.com/YanRob/IBM-Applied-Data-Science-Capstone-Project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Goal of the scraping is to extract Falcon 9 launch records HTML table from Wikipedia https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches and parse the table and convert it into a Pandas data frame

[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[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	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.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	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 after reentry, 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;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] 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[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	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.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	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[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	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.[508] 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.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second … were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the … ched the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	… 0 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

	Flight No.	Launch site		Payload	Payload mass		ustomer	Launch outcome	NaN	Version Booster	Booster landing	Date	Time
0	1	NaN	CCAFS	Dragon Spacecraft Qualification Unit	0		SpaceX	Success\n	NaN	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	NaN	CCAFS	Dragon	0	LEO	NASA	Success	NaN	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	NaN	CCAFS	Dragon	525 kg	LEO	NASA	Success	NaN	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	NaN	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	NaN	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	NaN	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	NaN	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10
...	...	...	...	...	...	...	...	...	...	...	...	...	...
116	117	NaN	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	NaN	F9 B5B1051.10	Success	9 May 2021	06:42
117	118	NaN	KSC	Starlink	~14,000 kg	LEO	SpaceX	Success\n	NaN	F9 B5B1058.8	Success	15 May 2021	22:56
118	119	NaN	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	NaN	F9 B5B1063.2	Success	26 May 2021	18:59
119	120	NaN	KSC	SpaceX CRS-22	3,328 kg	LEO	NASA	Success\n	NaN	F9 B5B1067.1	Success	3 June 2021	17:29
120	121	NaN	CCSFS	SXM-8	7,000 kg	GTO	Sirius XM	Success\n	NaN	F9 B5	Success	6 June 2021	04:26

- https://github.com/YanRob/IBM-Applied-Data-Science-Capstone-Project/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577366	28.561857	0
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577366	28.561857	1
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577366	28.561857	1
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577366	28.561857	0
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577366	28.561857	0

- The task Data Wrangling consists in writing the Landing Outcome into the new column Class with 0 in case of failure (False Ocean, False ASDS, None ASDS, None None, False RTLS) or with 1 in case of success (True ASDS, True RTLS, True Ocean)

- The mean of column class is 0.666666, i.e., the success rate is about 67%.

- https://github.com/YanRob/IBM-Applied-Data-Science-Capstone-Project/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- The following EDA charts will show the relation between flight number, launch site, orbit type, payload mass and success rate over the year.
- Goal of these charts are to identify the success rate depending on orbit, launch site and payload.
- The evolution of success rate over the year is also shown

- IBM-Applied-Data-Science-Capstone-Project/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb at main · YanRob/IBM-Applied-Data-Science-Capstone-Project (github.com)

EDA with SQL

SQL queries:

- List of launch sites
- 5 Launch sites beginning with "CCA"
- Total carried payload mass
- Average carried payload mass
- Date of the first successful landing
- Names of boosters in drone ship between 4000 and 6000 kg pay load

SQL queries:

- Number of successful and failure mission outcomes
- Name of the boosters carrying the maximum payload
- Records in drone ship ,booster versions, launch_site for the months in year 2015
- Ranking of the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order.

- [IBM-Applied-Data-Science-Capstone-Project/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · YanRob/IBM-Applied-Data-Science-Capstone-Project (github.com)]

Build an Interactive Map with Folium

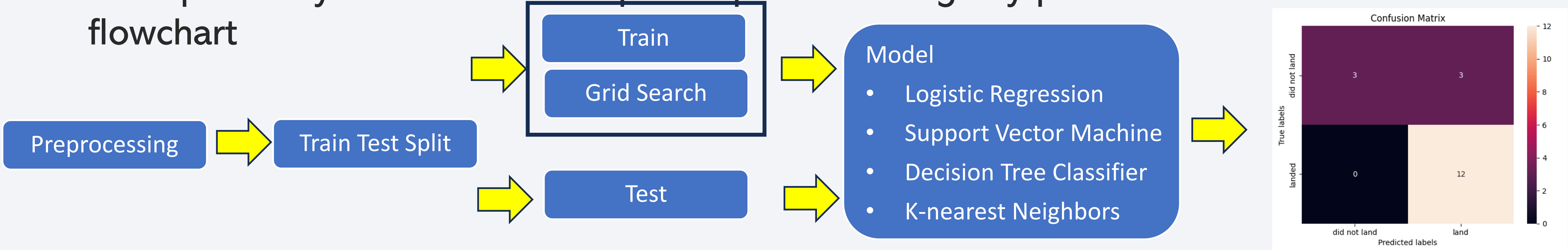
- After generating a map with Folium, one can place launch sites on the map using circles and markers.
- Using Icon allows us to visualize the number of launches performed on each site as well as the success rate of these launches.
- Adding lines makes the distance to any of geographic locations such as city, railway, coast or bridge visible.
- IBM-Applied-Data-Science-Capstone-Project/lab_jupyter_launch_site_location.jupyterlite (1).ipynb at main · YanRob/IBM-Applied-Data-Science-Capstone-Project (github.com)

Build a Dashboard with Plotly Dash

- On the Dashboard, the interactions between selected launch site, payload mass and success are highlighted.
- A pie chart allows to identify the most successful launch site, showing either the site with the most successful launches or the ratio of success when a site is selected via a dropdown menu.
- A scatter plot allows to identify success vs. failure as a function of the payload mass and site. The range of the payload mass may be adapted in order to refine the outcome.
- IBM-Applied-Data-Science-Capstone-Project/spacex_dash_app.py at main · YanRob/IBM-Applied-Data-Science-Capstone-Project (github.com)

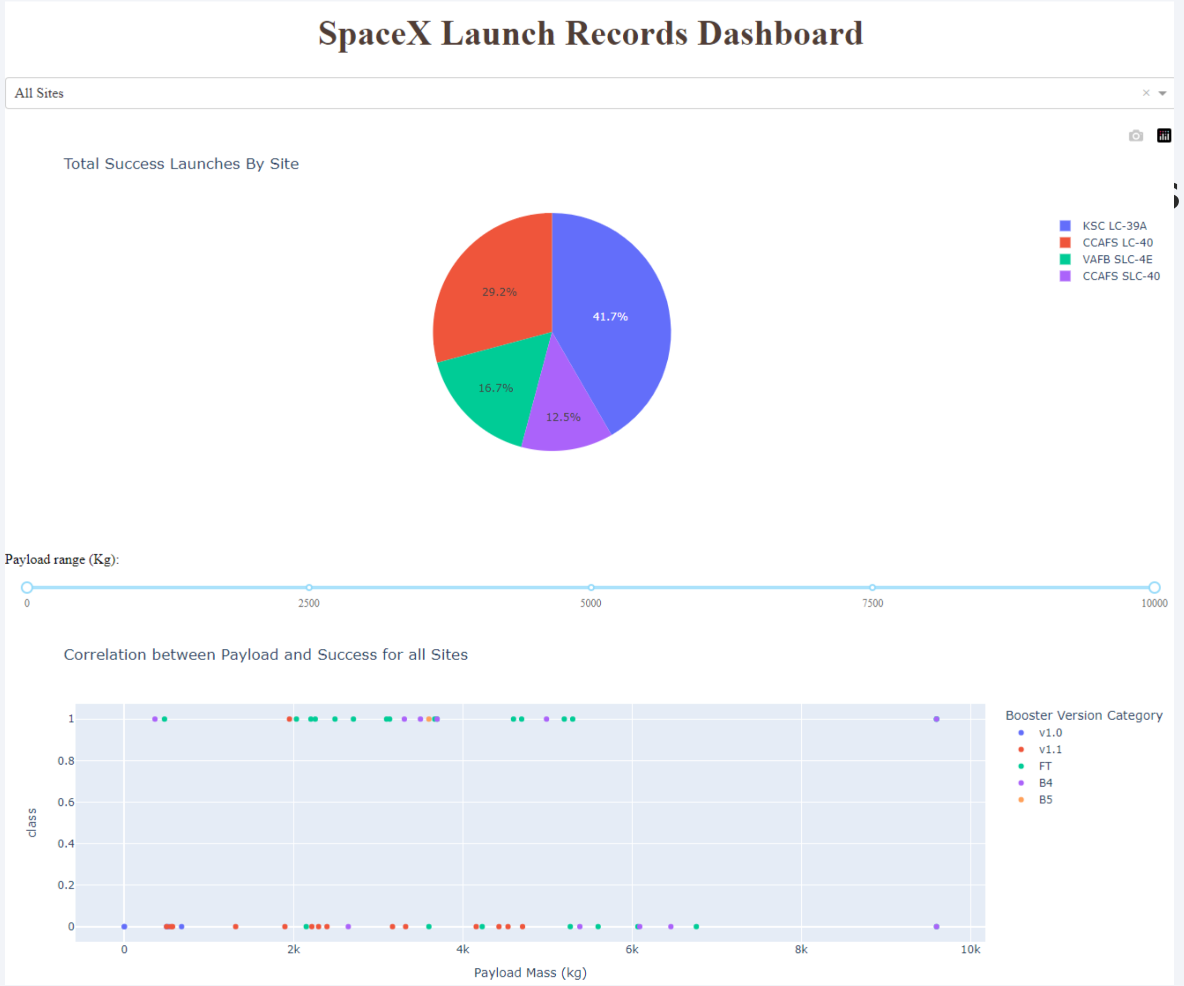
Predictive Analysis (Classification)

- After preprocessing, data are split into test and train analysis. Models (Logistic Regression, SVM, Tree decision, KNN) are trained on train data using GridSearch in order to find the best possible parameters of the model. Finally, the model is tested with the test data to obtained the confusion matrixYou need present your model development process using key phrases and flowchart

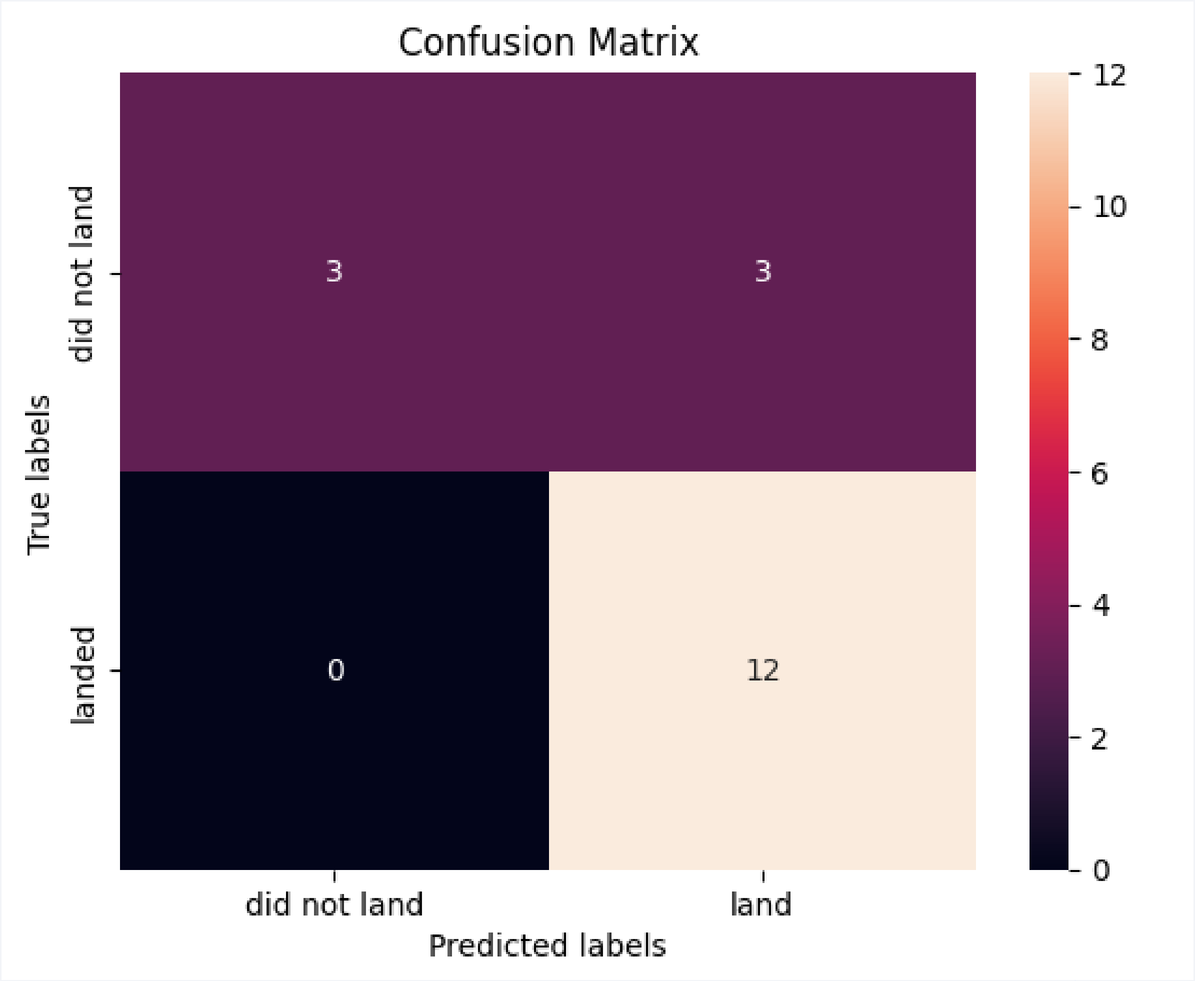


- [IBM-Applied-Data-Science-Capstone-Project/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite (1).ipynb at main · YanRob/IBM-Applied-Data-Science-Capstone-Project (github.com)]

Results



SpaceX Launch Records Dashboard
All Sites
Total Success Launches By Site
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40
41.7%
29.2%
16.7%
12.5%
Payload range (Kg):
0
2500
5000
7500
10000
Correlation between Payload and Success for all Sites
Booster Version Category
v1.0
v1.1
FT
B4
B5
class
Payload Mass (kg)



Confusion Matrix
True labels
did not land
landed
Predicted labels
did not land
land
3
3
0
12

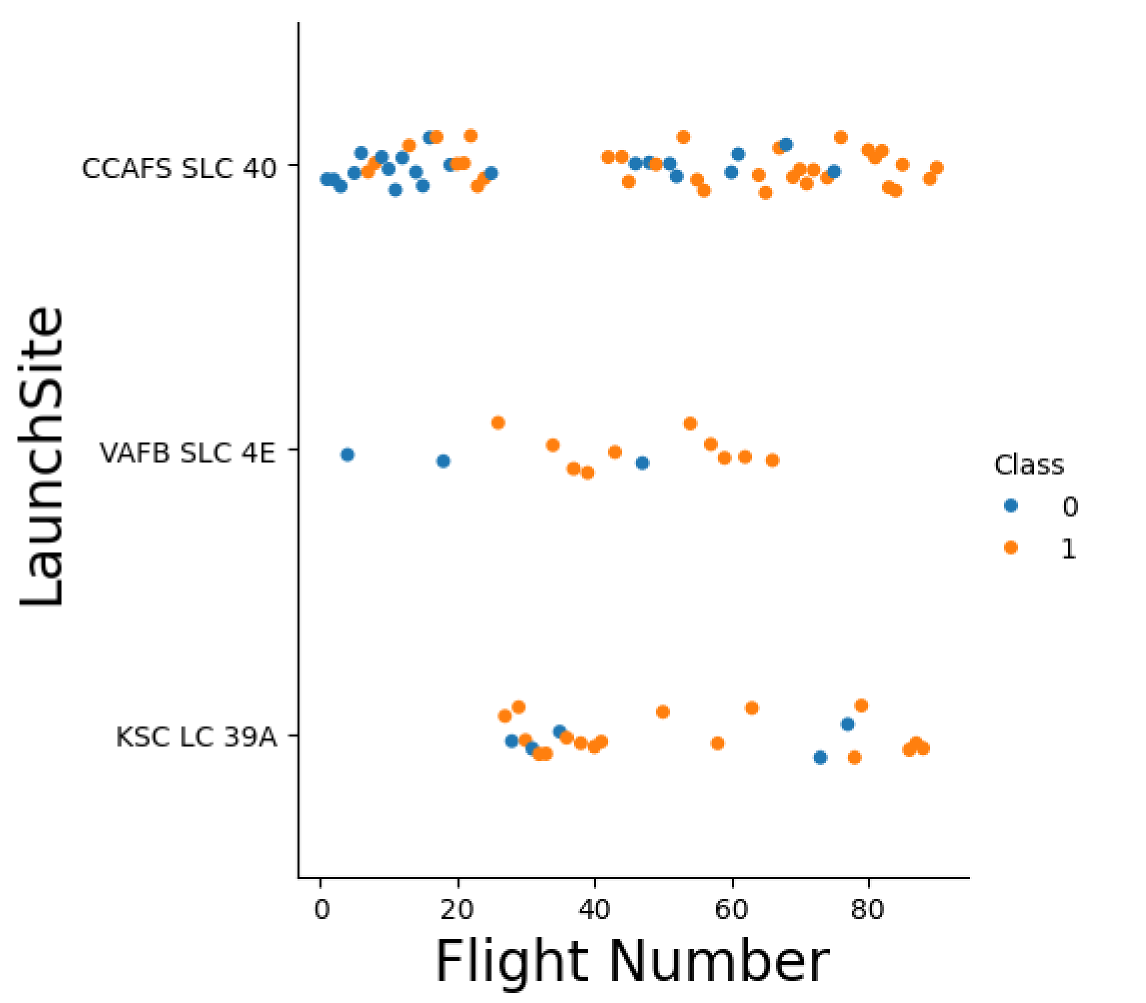
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- In this plot, failed (0) and successful (1) flights can be identified depending on their launching site.

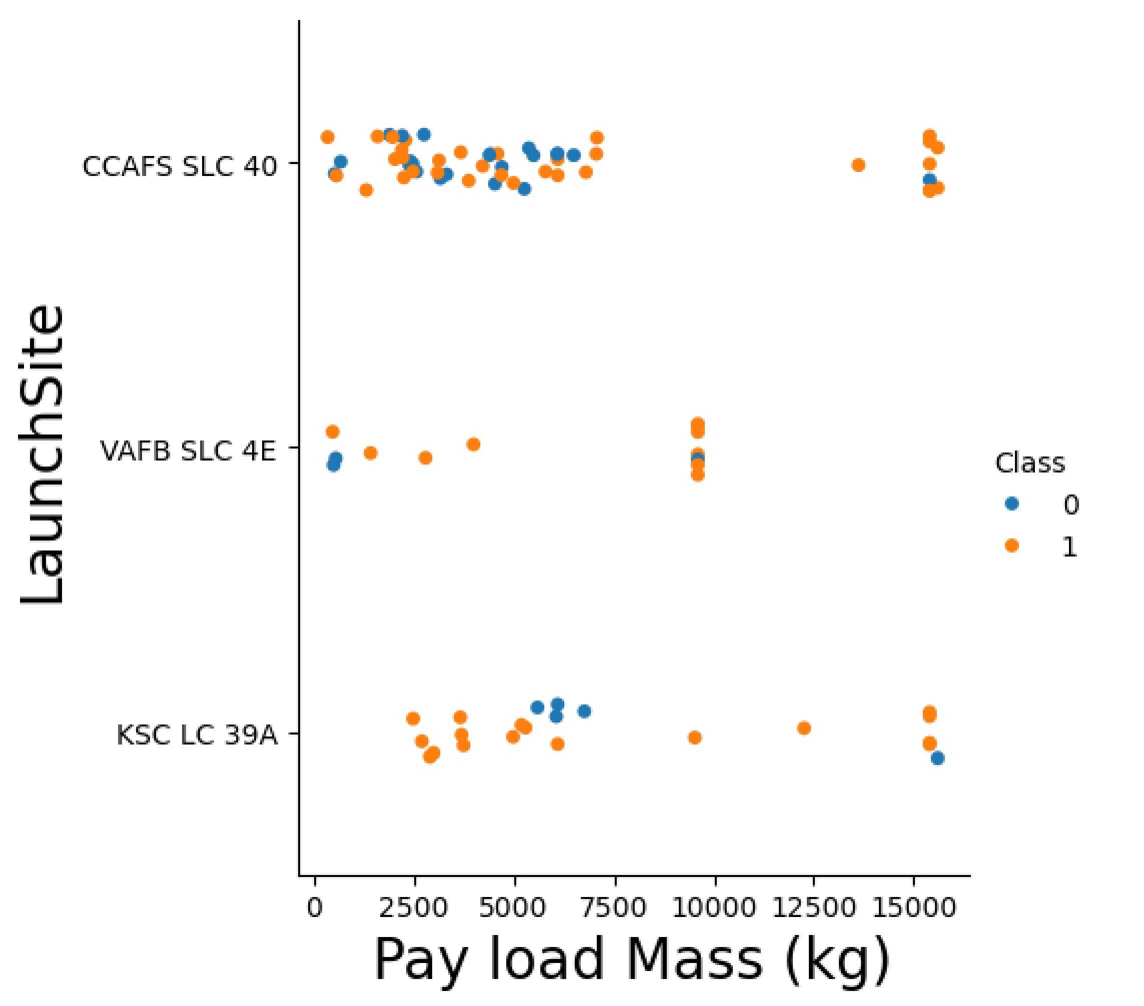
- At every launch site, the success rate improves with increasing flight number.



Payload vs. Launch Site

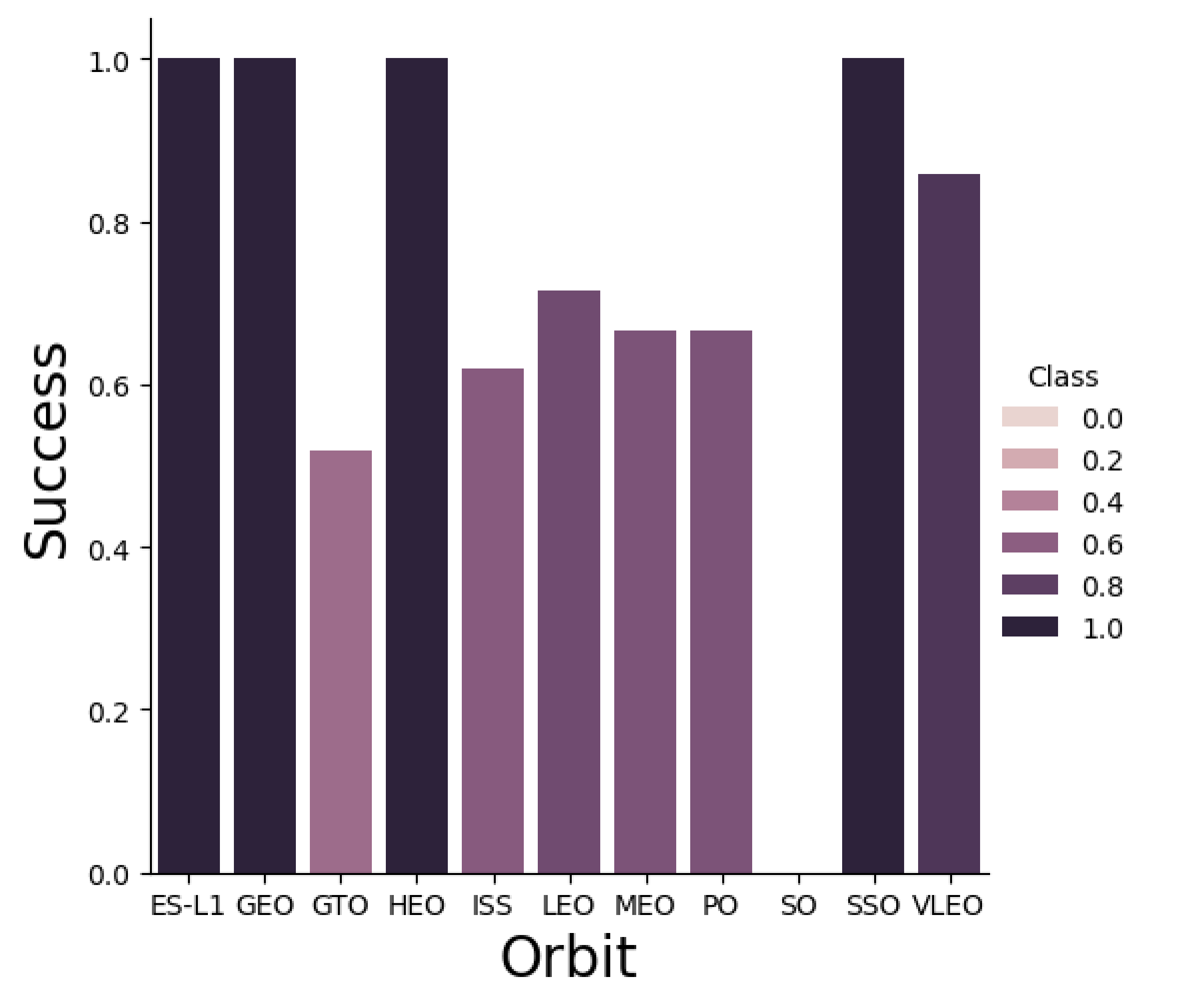
- In this plot, failed (0) and successful (1) flights can be identified depending on their launching site and their payload mass.

- No heavy rocket was launched from VAFB site.



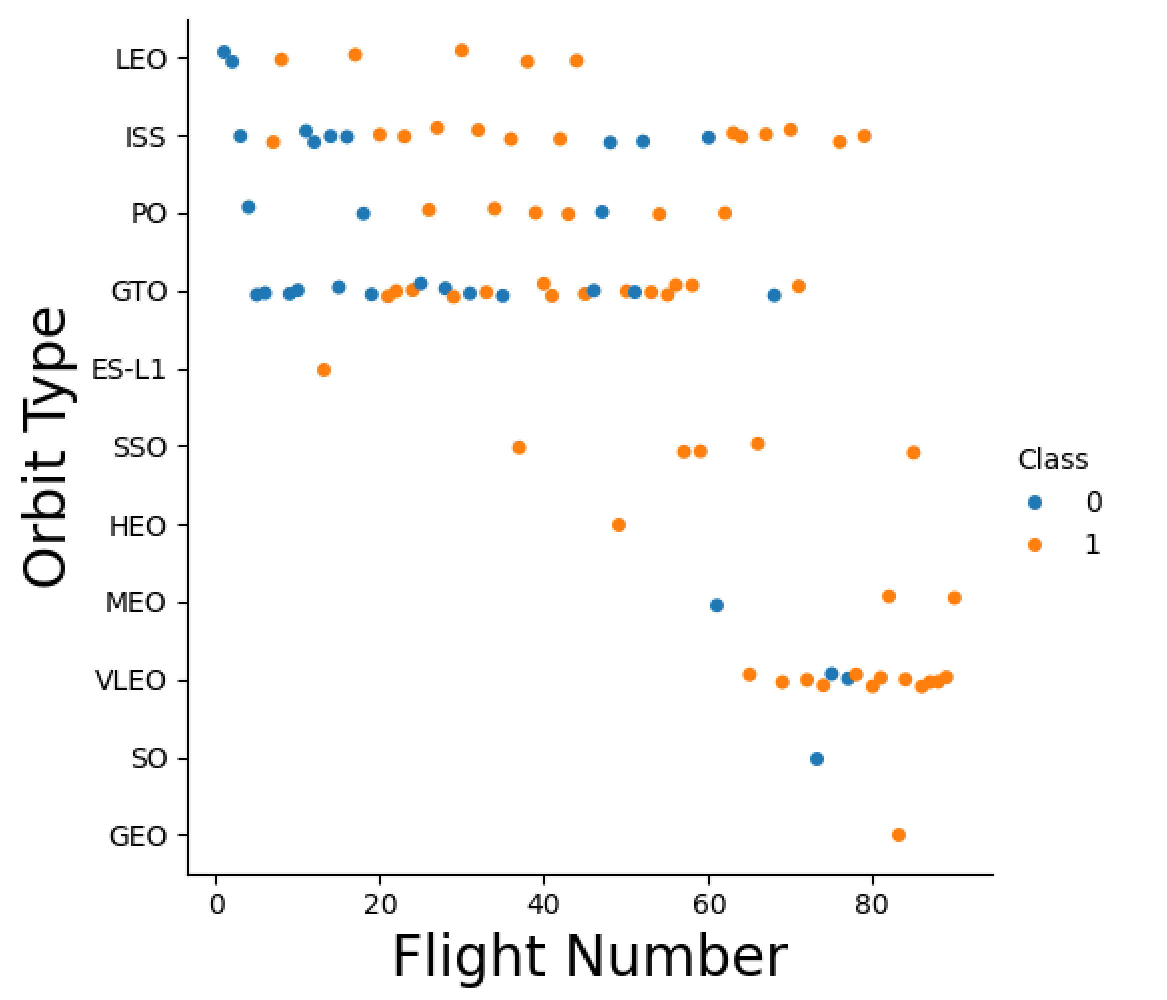
Success Rate vs. Orbit Type

- On this plot the success rate as a function of the orbit is shown.
- ES-L1, GEO, HEO and SSO are the most successful ones, where SO has no success.



Flight Number vs. Orbit Type

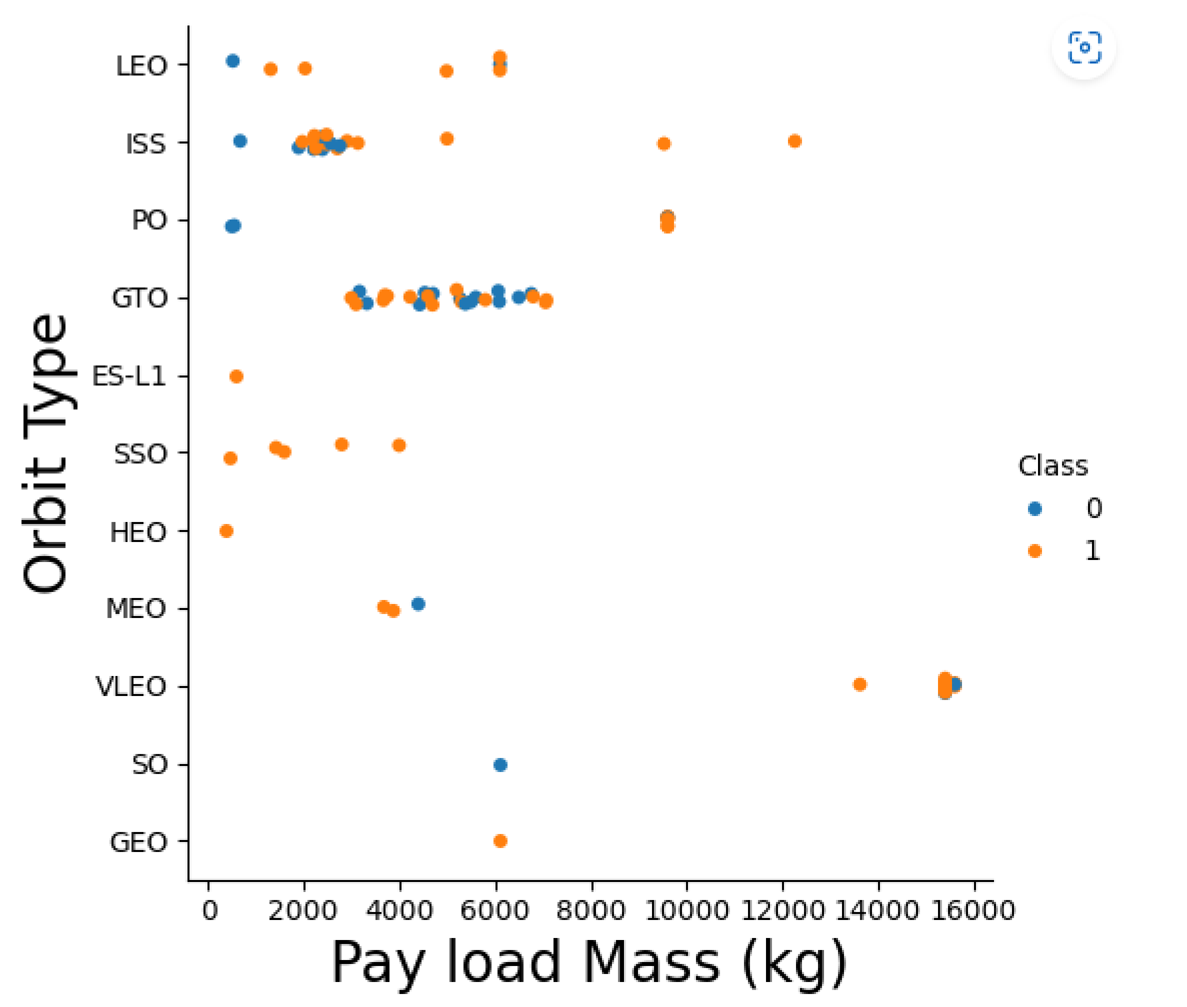
- In this plot, failed (0) and successful (1) flights can be identified depending on their orbit.
- Orbits with 100% success rate on previous slides have only few launches, and therefore not representative
- Density of failed launches decreases with increasing flight number



Payload vs. Orbit Type

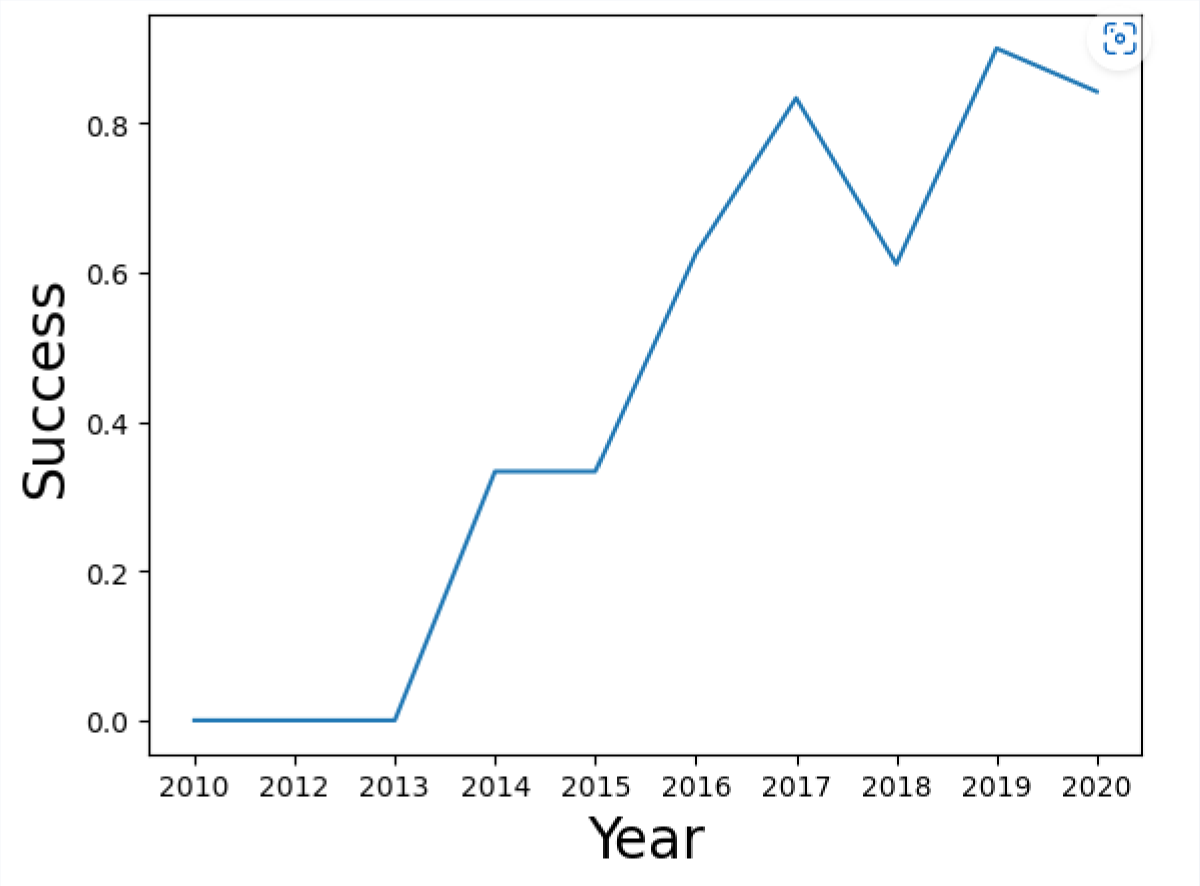
- In this plot, failed (0) and successful (1) flights can be identified depending on their orbit and their payload mass.

- An heavier payload seems to lead to a better success rate, except for GTO, where not tendency can be seen



Launch Success Yearly Trend

- The success rate is increasing from beginning of the test to react a success above 80% after 2019.



All Launch Site Names

- The unique launch sites are extracted from the SpaceX table applying the SQL query on the column Launch_Site to extract distinct sites:

```
%sql select distinct Launch_Site from SPACEXTABLE
```

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

Launch Site Names Begin with 'CCA'

- Here are 5 records where launch sites begin with `CCA`, extracted from the SpaceX table using the SQL query applied on Launch_Side column with a limitation of 5 values:

```
%sql select * from SPACEXTABLE where Launch_Site like "CCA%" limit 5
```

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

Total Payload Mass

- The total payload carried by boosters from NASA is 45,596 kg

sum(PAYLOAD_MASS__KG_)
45596

- The total payload is obtained by adding all values of the PAYLOAD_MASS_KG column of the SpaceX table with the following query:

```
%sql select * from SPACEXTABLE where Launch_Site like "CCA%" limit 5
```

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4 kg.

avg(PAYLOAD_MASS__KG_)
2928.4

- The result is obtained by averaging payload only for booster F9 v1.1 using SQL query:

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like "F9 v1.1"
```

First Successful Ground Landing Date

- The first successful landing outcome on ground pad occurred on 22th December 2015

Date
2015-12-22

- The result was obtained by selecting the first outcome of success landing.

```
%sql select Date from SPACEXTABLE where Landing_Outcome like "Success (ground pad)" limit 1
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are:

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

- The result is obtained by selected distinct boosters with a success in drone ship and a payload between 4000 and 6000kg

```
%sql select distinct Booster_Version from SPACEXTABLE where Landing_Outcome like "Success (drone ship)" and (PAYLOAD_MASS__KG_ between 4000 and 6000)
```

Total Number of Successful and Failure Mission Outcomes

- 100 missions out of 101 were successful. One of them had nevertheless an unclear status regarding payload.

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

- The result is obtained by grouping and counting Mission_Outcome

```
%sql select Mission_Outcome, count(*) from SPACEXTABLE group by Mission_Outcome
```

Boosters Carried Maximum Payload

- The boosters which have carried the maximum payload mass are

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

- The result was obtained by calculating the maximum payload mass and selecting booster with this mass:

```
%sql select Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

2015 Launch Records

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 are

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

- The result was obtained by selecting launches occurring in 2015 with a failure in drone ship.

```
%sql select substr(Date, 6, 2) as Month, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE where substr(Date,0,5)='2015' and Landing_Outcome like "Failure (drone ship)"
```

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

- The landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, are in descending order:

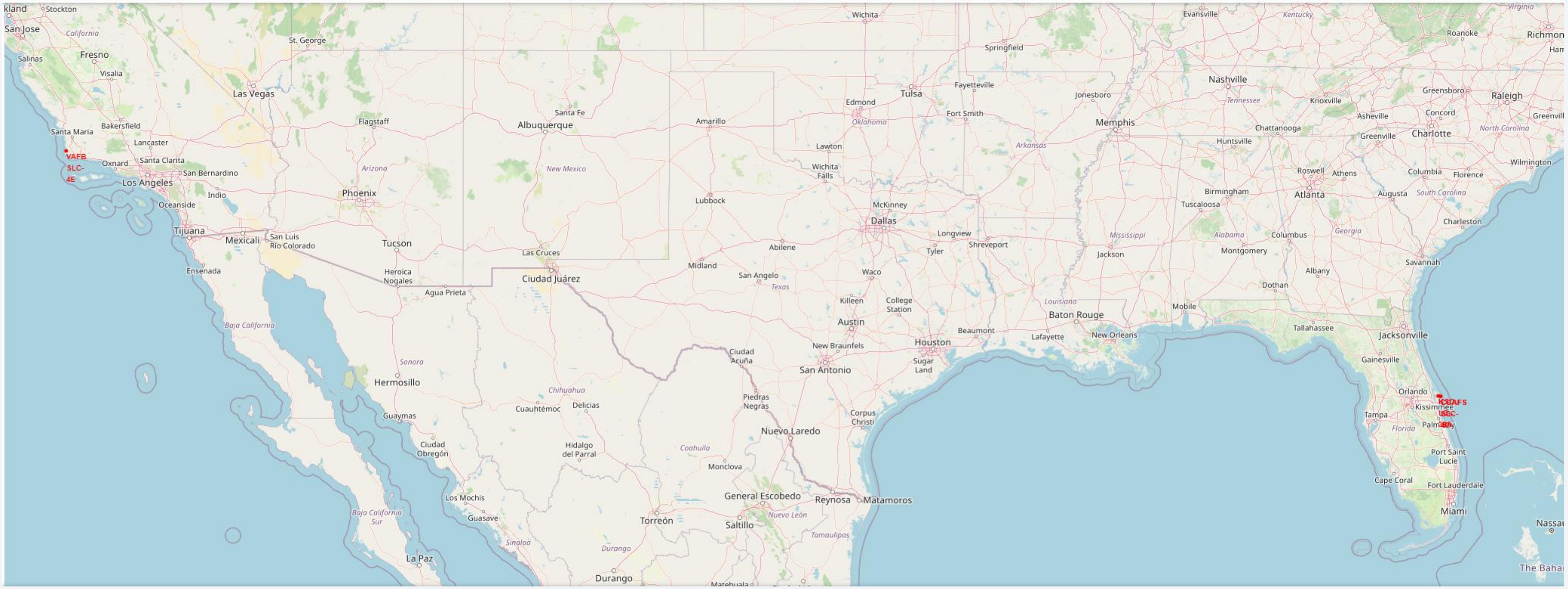
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- Only landing outcomes between 2010-06-04 and 2017-03-20 were selected and distinct outcomes were counted.

Section 3

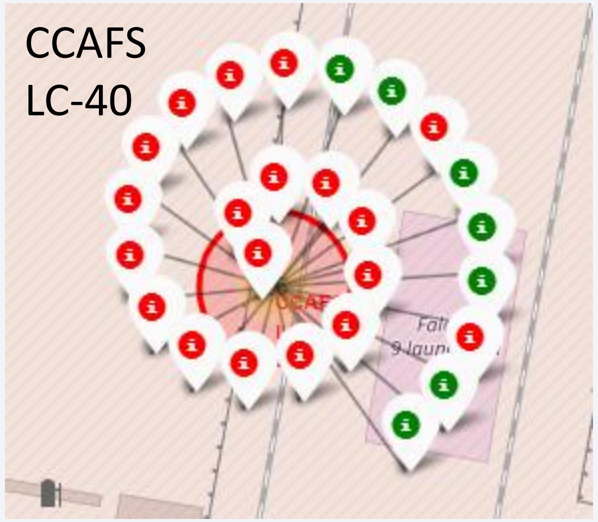
Launch Sites Proximities Analysis

Location of launch sites

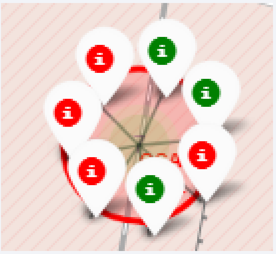


- All sites are in the South of USA, in order to be as close as possible to the equator, which is a favorable position for a launch site.
- All sites are close to the coast, either Pacific or Atlantic Oceans

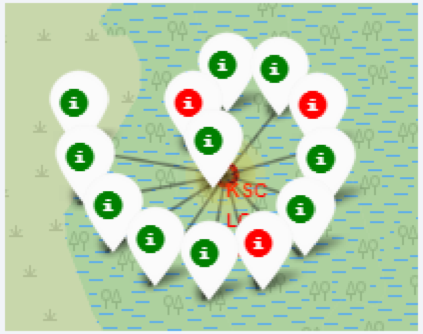
Success/failed launches per site



CCAFS
SLC-40



KSC
LC-39A



VAFB
SLC-4E

- Green marks represents success whereas red failed launches.
- The launch site KSC LC-39A show a bigger success rate than any other site

Distance between site CCAFS SLC 40 to its proximities

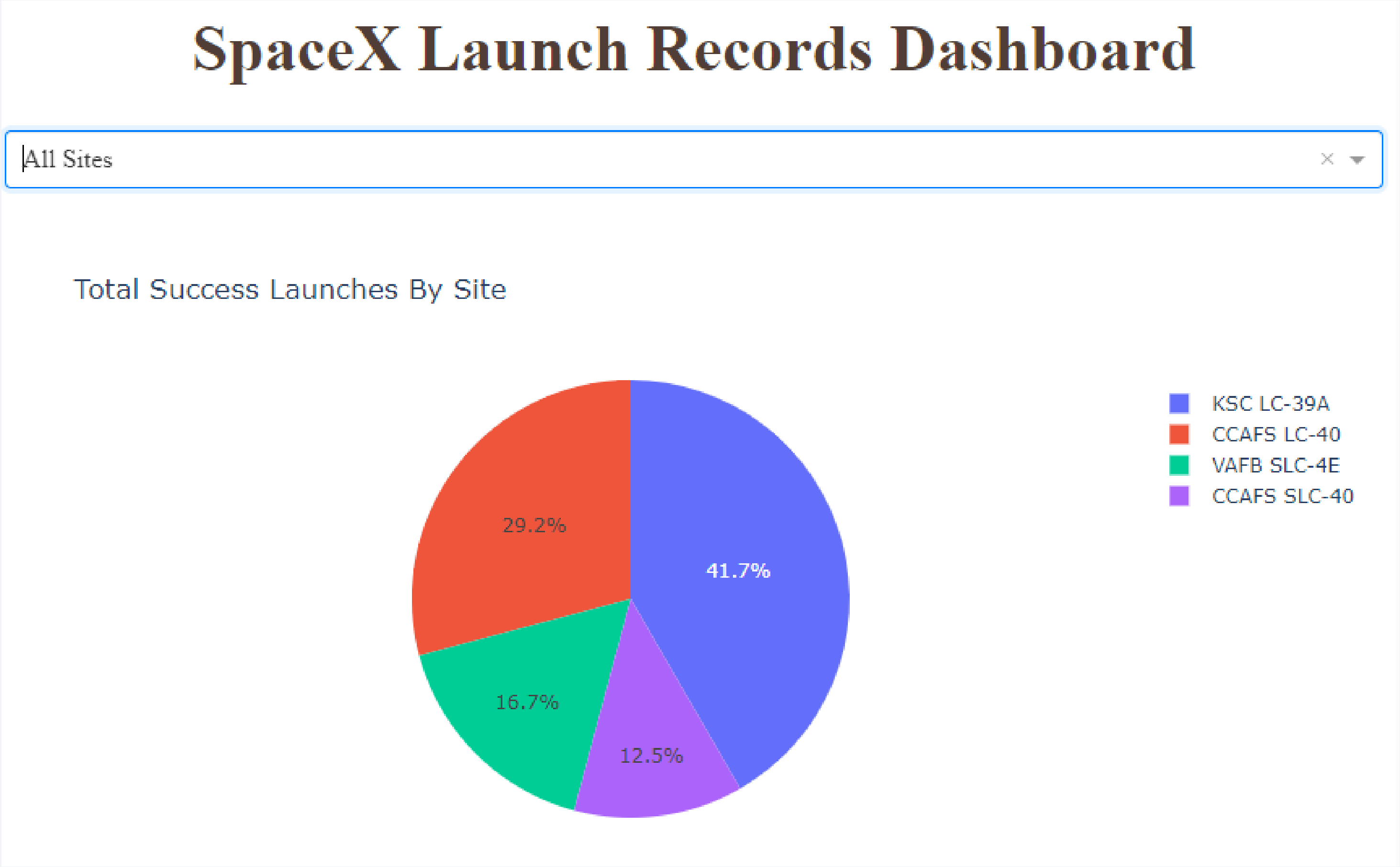
- The launch site CCAFS SLC-40 is located very close to the sea (less than 1km), which is positive
- The site is far enough from the next city (23km) and bridge (6.5km)
- But the site is very close to a railway (1.26km), which is probably nevertheless necessary for the launch site itself

Section 4

Build a Dashboard with Plotly Dash

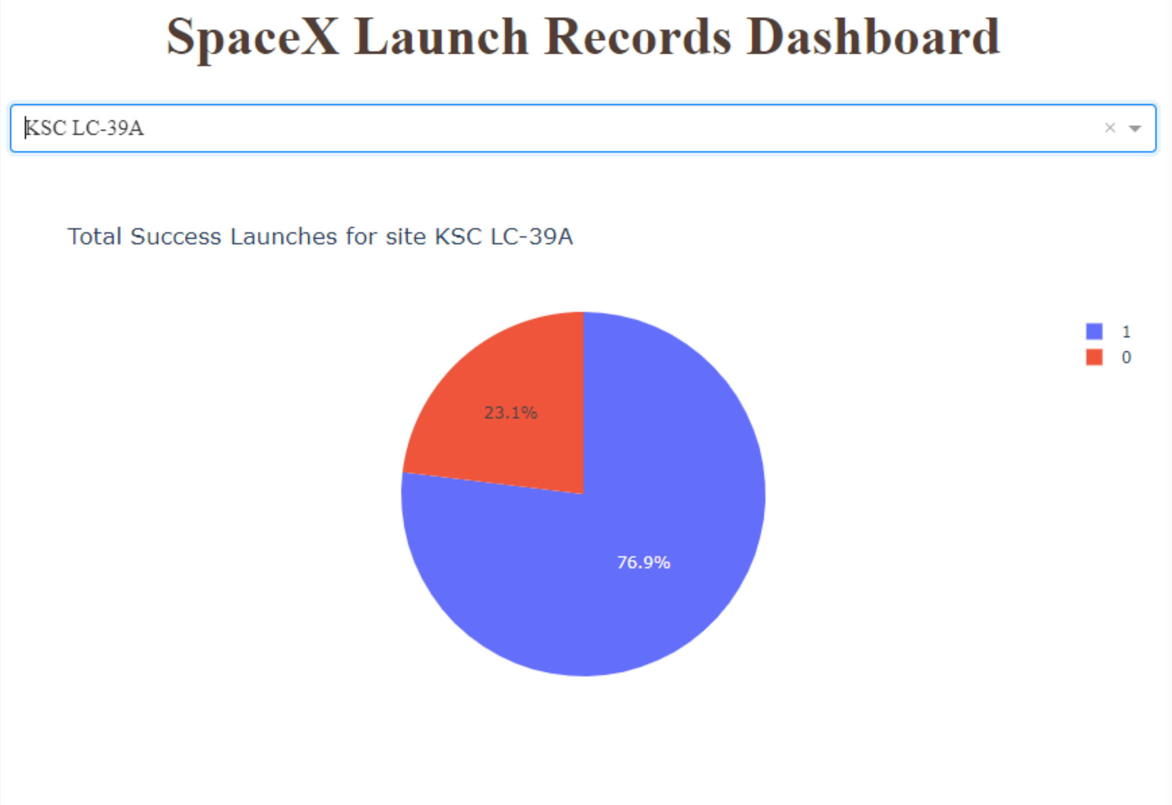
Total Success Launches by Site

- From this piechart, it is obvious that the site KSC LC-39A has the largest successful launches



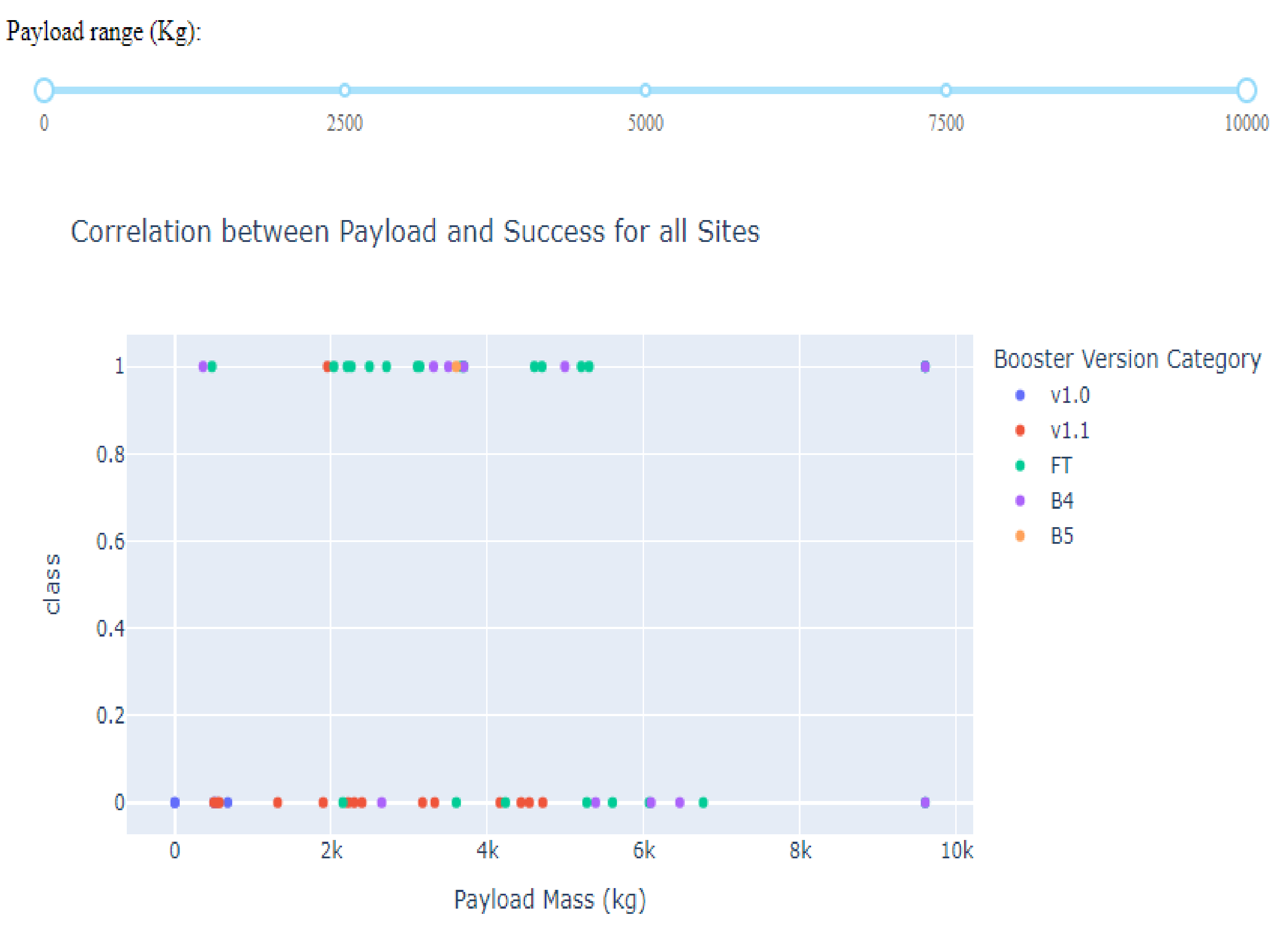
Launch success rate

- The site KSC LC-39A is also the site showing the highest launch success rate.



Payload and success rate

- The highest success rate is reached between 1952 and 5300kg
- The lowest success rate is reached below 1952 kg and above 5300 kg
- The booster versions B5 and FT show the highest launch success rate.

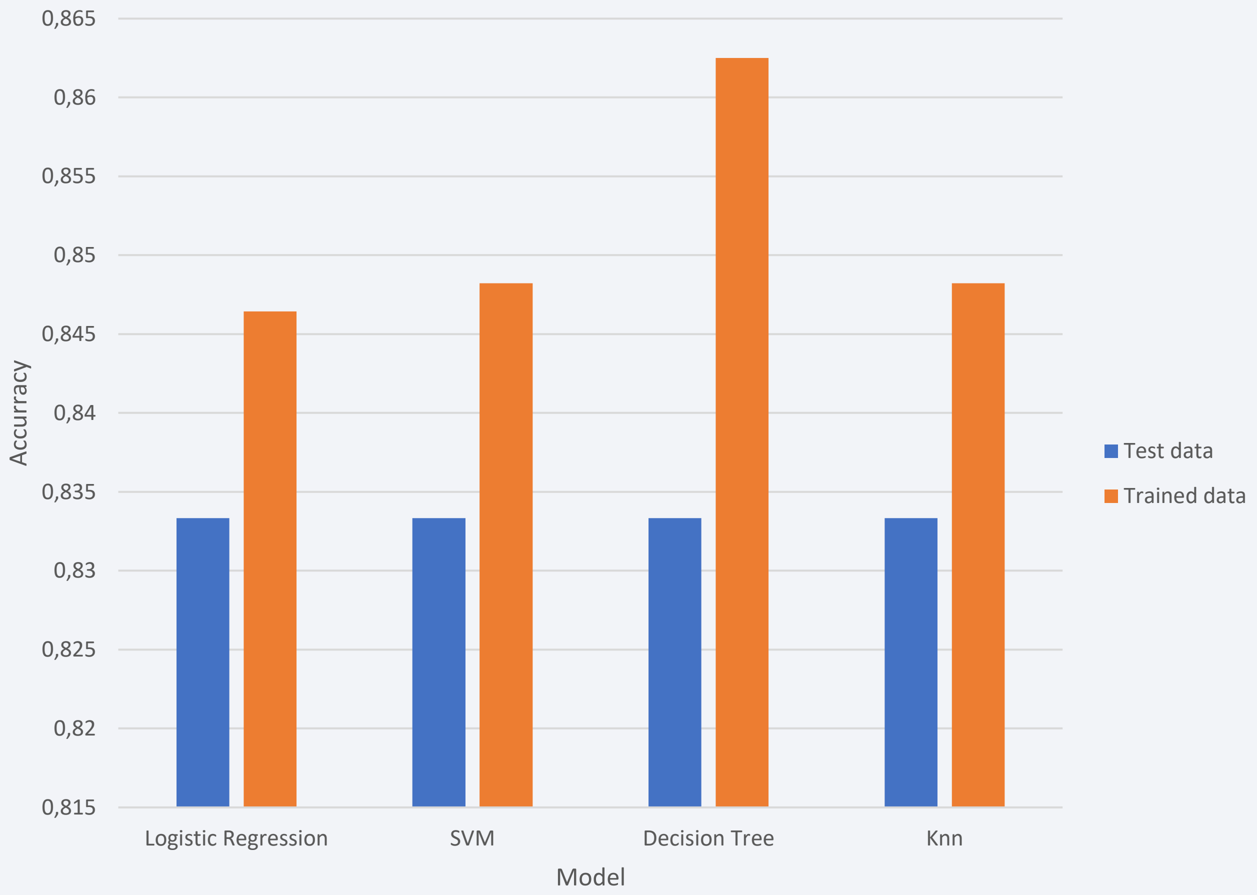


Section 5

Predictive Analysis (Classification)

Classification Accuracy

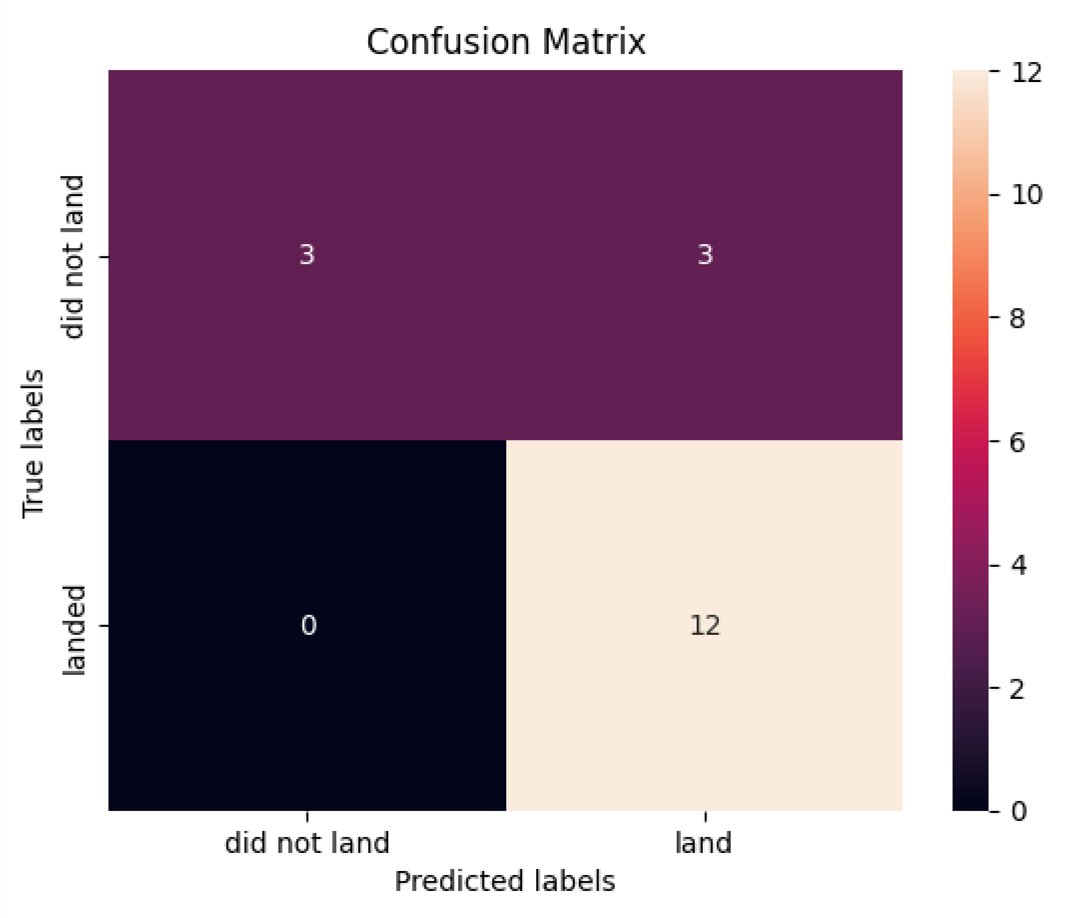
- All models have the same accuracy with the test data: 0.8333.
- Decision Tree shows nevertheless the best accuracy with train data: 0.8625.
- We can assume that Decision Tree would be the best model.
- But to confirm or not this assessment, it is required to split train and test data in other way.



Confusion Matrix

- All models lead to the same confusion matrix.

- Despite a relatively good prediction accuracy (0.83), there are still 3 false positive (predicted landing when it failed), which may be problematic.



Conclusions

- From the launch site analysis, we know that a launch from which site a launch will likely success (KSC LC-39A) or not (CCAFS LC-40).
- ES-L1, GEO, HEO and SSO are the most successful Orbit types with a success rate of 100%, where SO has no success. GTO orbits show week success.
- The highest success rate is reached for pay load masses between 1952 and 5300kg
- The booster versions B5 and FT show the highest launch success rate.

- For example, a launch from site KSC LC-39A to SSO orbit with a pay load mass between 1952 and 5300kg for a B5 launcher may be proposed at a smaller price than from another launch site.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!