IBM Data Science Capstone Project

SpaceX Landing Analysis and Predictions for cost effectiveness.

Fadhil Umar Al Farouq

20 February 2023

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

This Powerpoint is addressed to complete the **IBM Data Science Capstone Project**.

In this project, we will conduct data **collection**, **analysis**, **modelling**, and presentation of data concerning the **SpaceX landing history** to predict further landing.

The data will be collected using the **SpaceX API** for past launches and **Webscraping** from the SpaceX Wikipedia page.

The data will be processed and presented using a **dashboard** to gain insight. The dashboard will include the **site map** using folium and **web dashboard** using dash plotly.

Lastly the data will be used to create **machine learning model** for further use.

INTRODUCTION

The analysis will be divided into three steps

- Exploratory Data Analysis
- Answer questions using SQL
- Visualization using seaborn library

The main question we will try to answer is

1. What the data collected tell us about?
2. How is the data relate to each other?
3. What model is the best and is it reliable?

METHODOLOGY

The data will be collected from two sources,

1. SpaceX API endpoint using the requests method.
2. SpaceX Wikipedia web scraping using beautiful soup

The analysis will be conducted using

1. Exploratory data analysis
2. SQL queries
3. Data visualization using seaborn

Then we will create a machine learning model using sklearn and present the data using folium and plotly dash.

RESULTS

API collection

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	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
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

RESULTS

Webscraping

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

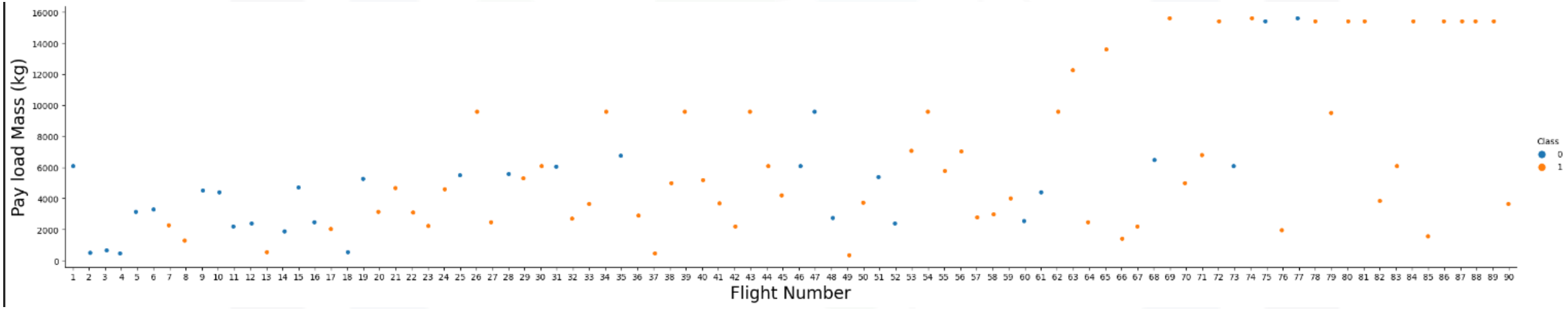
RESULTS

EDA Results

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
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
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
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
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
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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
86	2020-09-03	Falcon 9	15400.000000	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	2	B1060	-80.603956	28.608058	1
87	2020-10-06	Falcon 9	15400.000000	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	2	B1058	-80.603956	28.608058	1
88	2020-10-18	Falcon 9	15400.000000	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	5	B1051	-80.603956	28.608058	1
89	2020-10-24	Falcon 9	15400.000000	VLEO	CCAFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	2	B1060	-80.577366	28.561857	1
90	2020-11-05	Falcon 9	3681.000000	MEO	CCAFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	0	B1062	-80.577366	28.561857	1

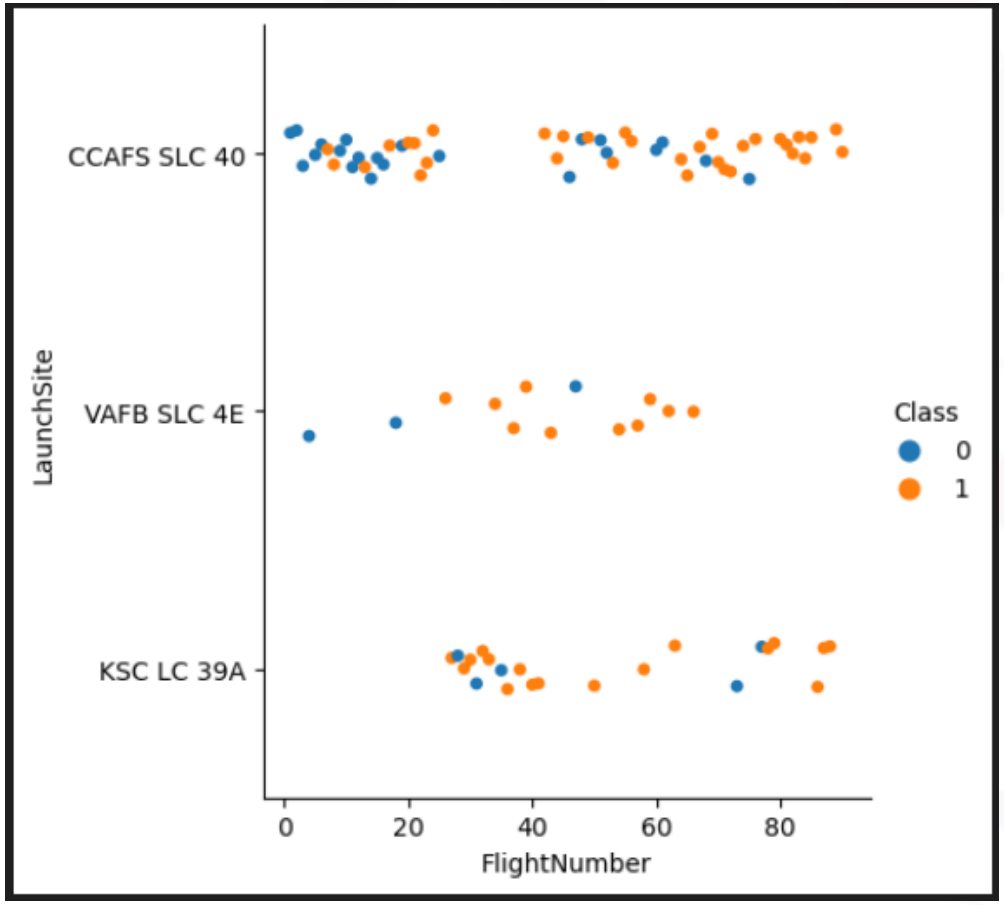
RESULTS

Payload mass per flight

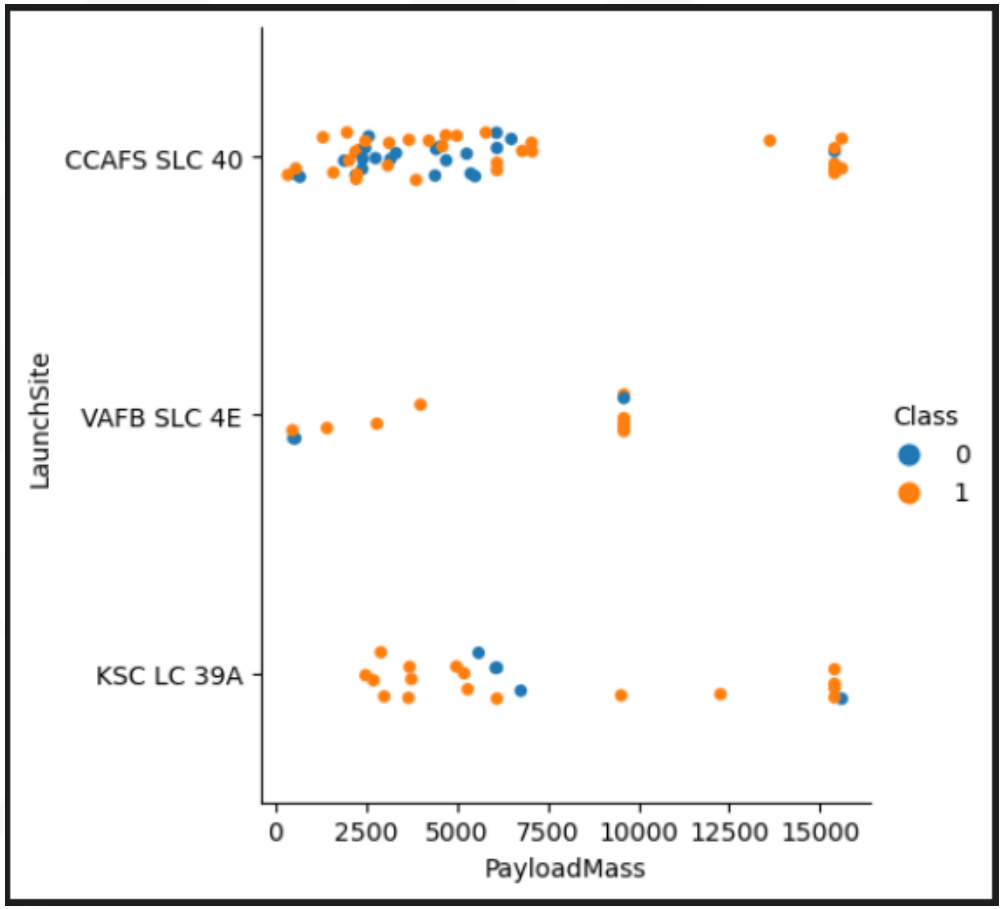


RESULTS

Success per launchsite and fightnum

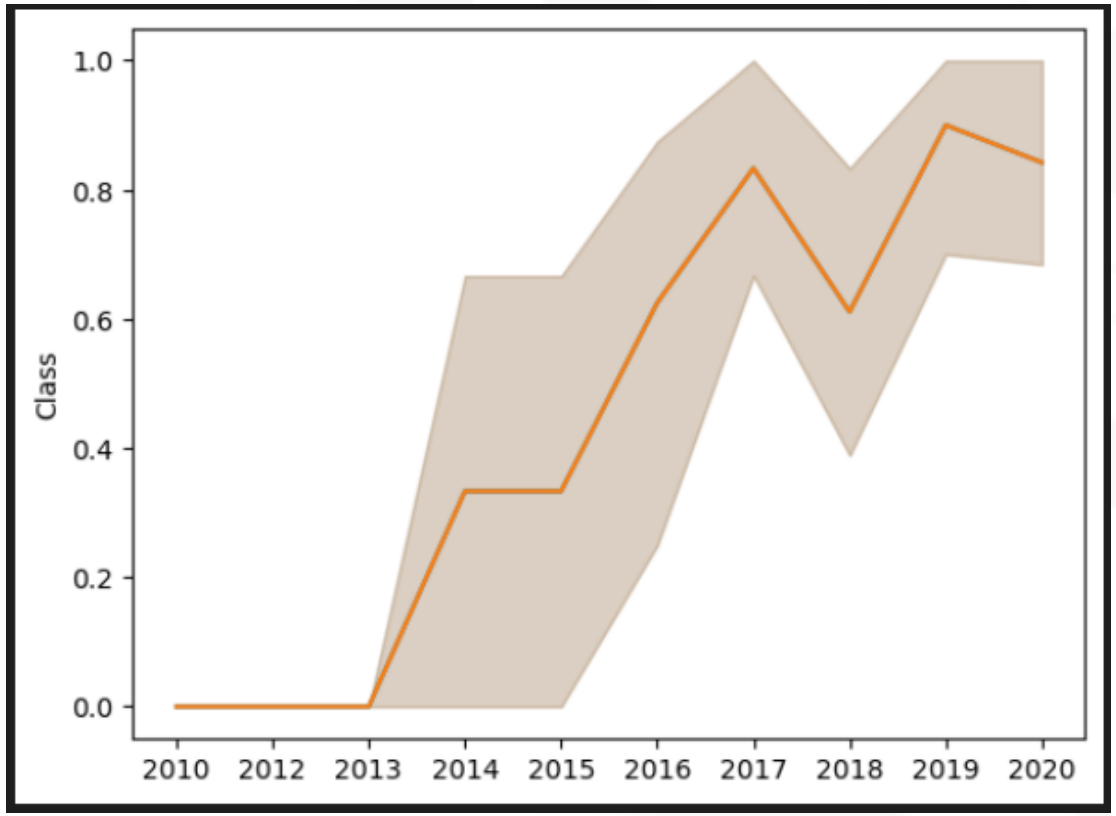


Success per laucnhsite and payload

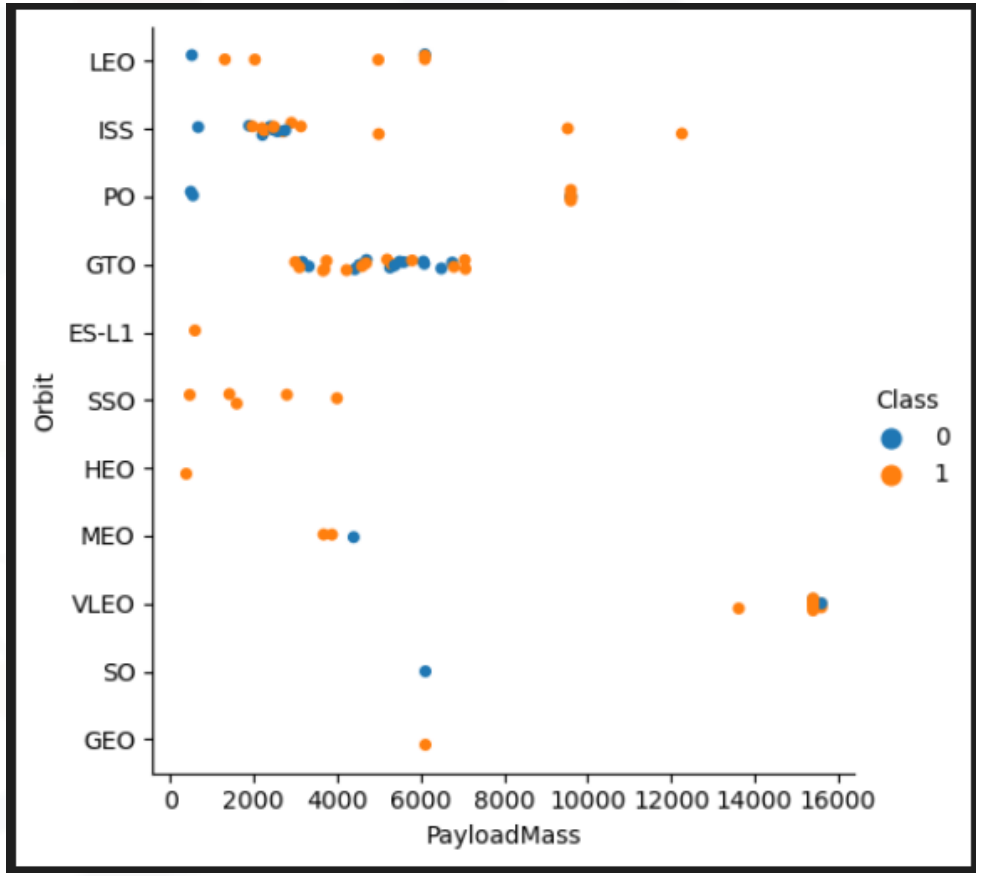


RESULTS

Success over the years

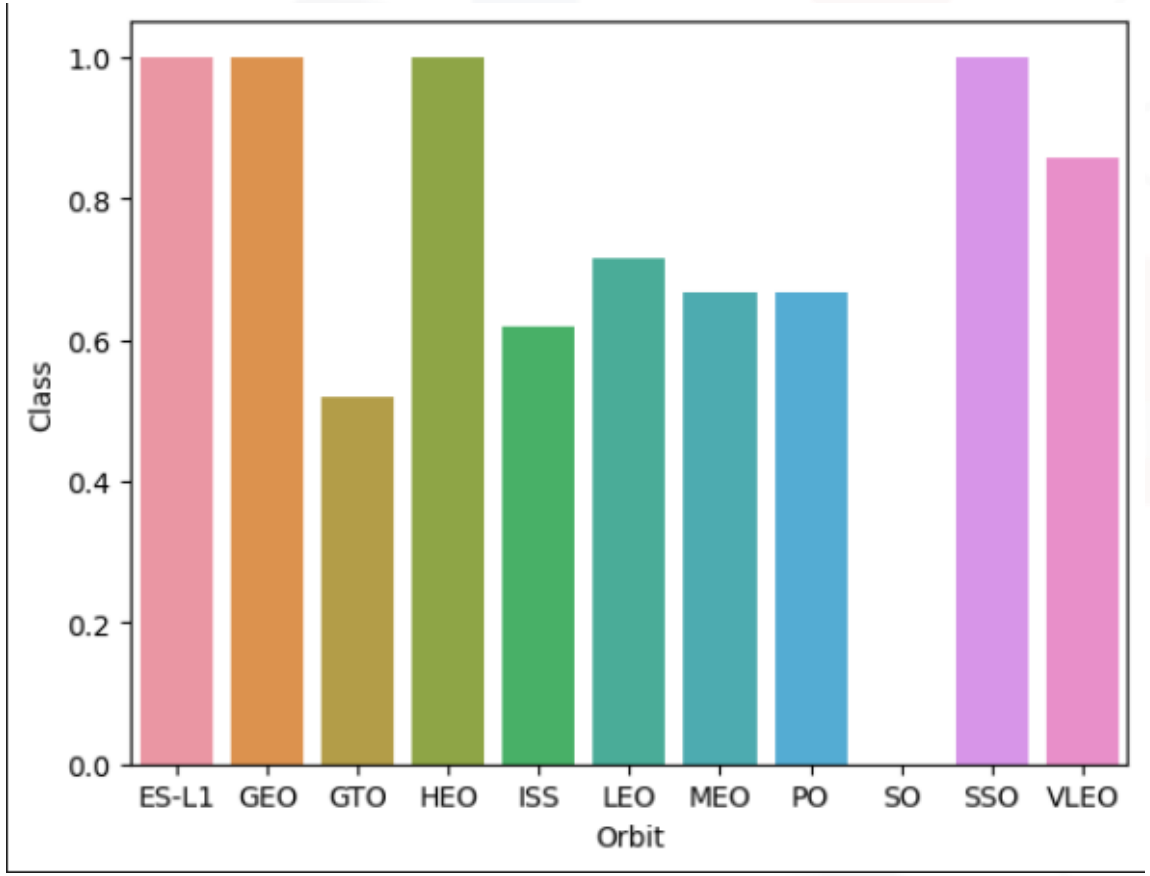


Orbit, Payload, and success relation

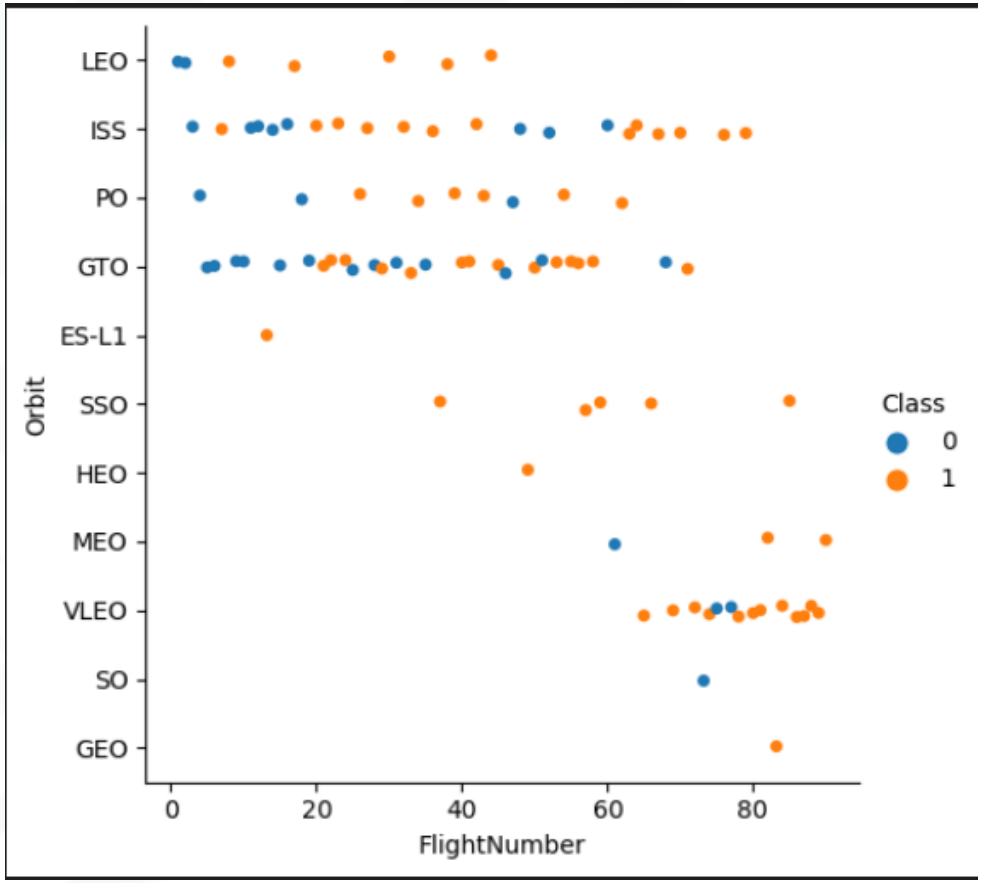


RESULTS

Success per Orbit launched



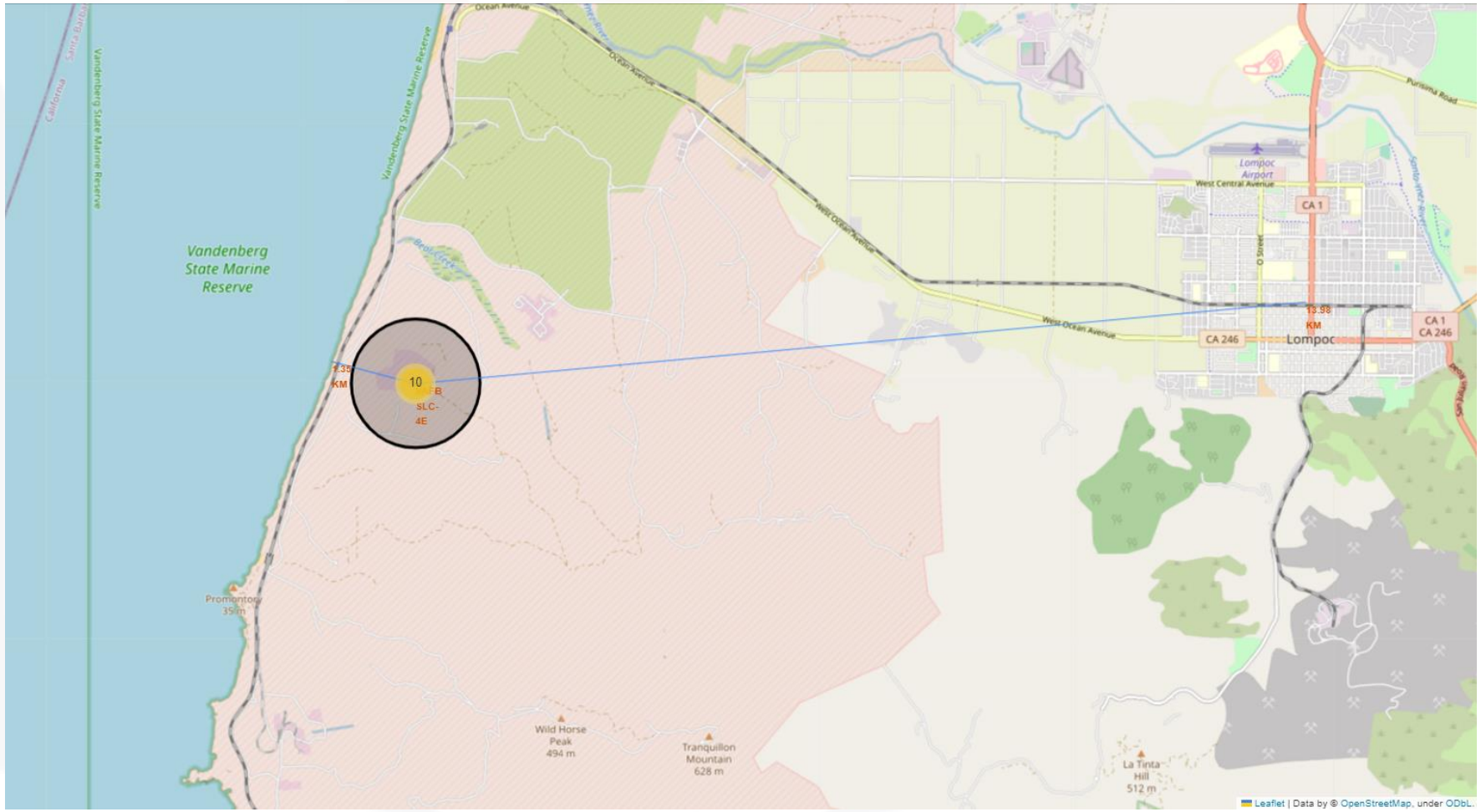
Orbit, Flight num, and success relation



RESULTS

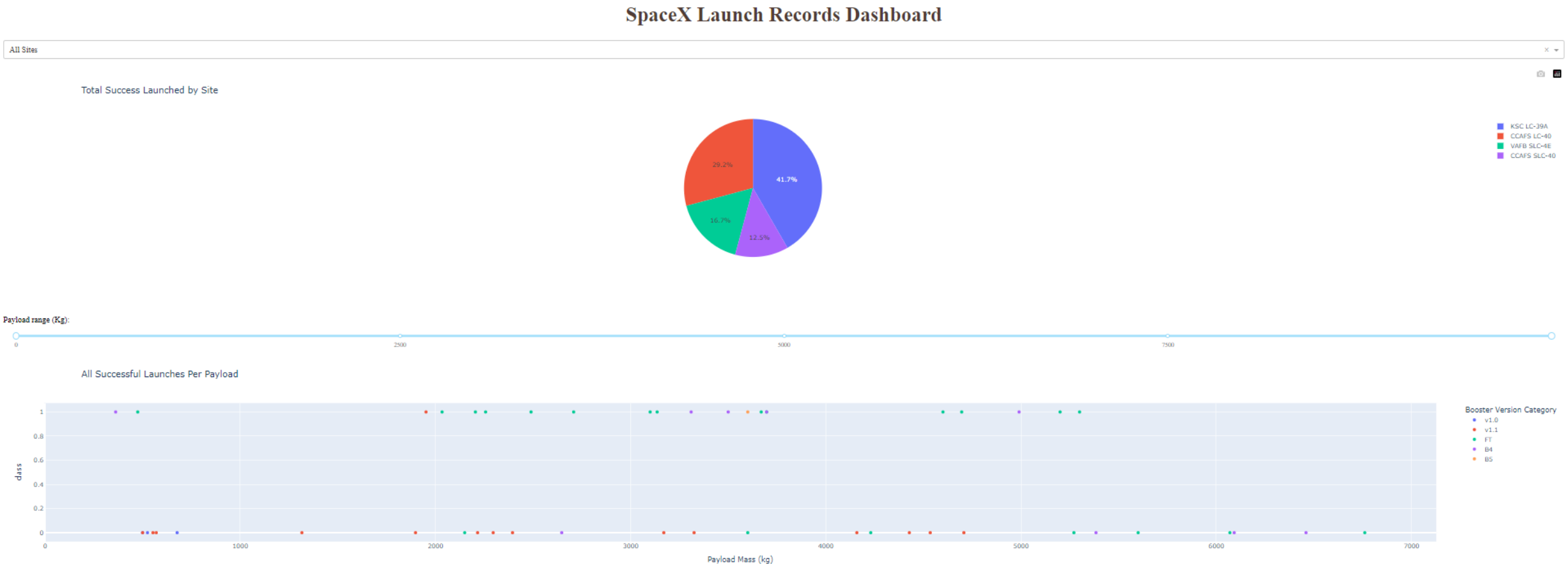
Site

Dashboard



RESULTS

Webpage Dashboard



RESULTS

Question answered using SQL

1. unique launch sites in the space mission
2. 5 records where launch sites begin with the string 'CCA’
3. total payload mass carried by boosters launched by NASA (CRS)
4. average payload mass carried by booster version F9 v1.1
5. date when the first successful landing outcome in ground pad was achieved
6. names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

RESULTS

Question answered using SQL

1. unique launch sites in the space mission
2. 5 records where launch sites begin with the string 'CCA’
3. total payload mass carried by boosters launched by NASA (CRS)
4. average payload mass carried by booster version F9 v1.1
5. date when the first successful landing outcome in ground pad was achieved
6. names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. total number of successful and failure mission outcomes
8. names of the booster_versions which have carried the maximum payload mass
9. failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
10. 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

RESULTS

1

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

3

total_payload
256163

4

avg_payload
3676

5

DATE
2017-01-05

6

booster_version	payload_mass__kg_
F9 B5 B1046.2	5800
F9 B4 B1040.2	5384
F9 FT B1032.1	5300
F9 FT B1020	5271
F9 FT B1031.2	5200
F9 B4 B1043.1	5000
F9 B4 B1040.1	4990
F9 FT B1022	4696
F9 v1.1	4535
F9 v1.1 B1011	4428
F9 B5B1062.1	4311
F9 B5 B1051.2	4200
F9 v1.1 B1014	4159
F9 B5 B1046.3	4000

2

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)
2012-08-10	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
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

RESULTS

7

mission_outcome	frequency
Success	44
Success (payload status unclear)	1

8

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3

9

booster_version	launch_site	landing__outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)

10

landing__outcome	total_outcome
No attempt	7
Failure (drone ship)	2
Success (drone ship)	2
Success (ground pad)	2
Controlled (ocean)	1
Failure (parachute)	1

RESULTS

Results of Binary Classification model

Models used :

- Logistic Regression
- SVM
- Decision Tree
- KNN

```
Logistic Regression score : 0.8333333333333334
SVM score : 0.8333333333333334
Tree score : 0.8333333333333334
KNN score : 0.8333333333333334
```

DISCUSSION

OVERALL FINDINGS & IMPLICATIONS

Findings

- The success of rocket landing is improved over the years.
- The success of rocket landing can be predicted by its characteristics.
- Small datasets cause the model to be undertrained.

Implications

- By experimenting with different parameters (mass, booster, launch site) we can determine the best for success
- Using past records, we can create a machine learning model for future deployment
- More well documented dataset is needed for better results.

CONCLUSION

- The record of past launches gives us insight on how different parameters affect success, some examples are CCAFS LC-40 launch site has better success rate.
- The success of landing can be determined using ML model but we need more well documented data for validation.
- The same accuracy presented is caused by the small dataset which imply an undertrained model

APPENDIX

- The dataset obtained is from the Wikipedia page and only include the type of object used (booster, orbit, launch site)
- For better information we might want to transform the object into a numerical value for comparison for example (energy generated and fuel efficiency by booster, height of orbit, additional equipment used in launchsite).