IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Simon BELFATMI

Outline

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

Executive Summary

- In this project we will predict if SpaceX Falcon 9 first stage will land successfully using machine learning classification algorithms as some features of the rocket launches have a correlation with the outcome of the launches that would be hard for a human to calculate. Such as Payload mass, Orbit type, Launch site, etc.
- The data tree classification algorithm may be the best machine learning algorithm to predict if the Falcon 9 first stage will land successfully.

Introduction

- SpaceX 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. If we can determine if the first stage will land, we can determine the cost of a launch and how much we save against competitors.

Section 1

Methodology

Methodology

The overall methodology includes:

- Data collecting, wrangling and formatting from the SpaceX API with web scraping.
- Exploratory data analysis (EDA) using Pandas, NumPy, and SQL
- Data visualization using, Matplotlib, Seaborn, Folium, and Dash
- Machine learning prediction, using Logistic regression, SVM*, Decision trees and KNN*

*Support vector machine
*K-nearest neighbors

Data Collection – SpaceX API

- Web scraping
  - The data is scraped from https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
  - The website contains only the data about Falcon 9 launches.
  - We end up with 121 rows or instances and 11 features per rocket. The picture below shows the first five rows:

	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

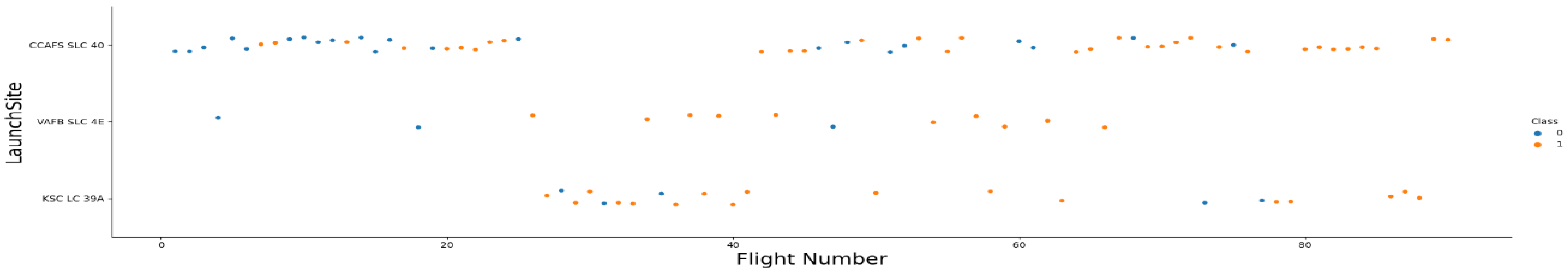
Data Collection – Scraping, Wrangling & Formatting

- The data is processed so that there are no missing or blank entries, categorical features are encoded using one-hot encoding.
- An extra column called ‘Class’ is also added to the data frame. The column ‘Class’ contains 0 if a given launch is failed and 1 if it is successful.
- In the end, we end up with 90 rows or instances and 83 columns or features, picture of the first five rows are listed below:

	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

EDA with Data Visualization

- I used both scatter and bar plots to analyze relations between success & failure and the rocket's features, such as:
  - Payload Mass & Flight Number
  - Launch Site & Flight Number
  - Launch Site & Payload Mass
  - Orbit Type & Flight Number
  - Payload & Orbit Type



EDA with SQL

- I used the following SQL Queries during the project:

  - 5 Names of unique launch sites used by SpaceX
  - First 5 Launch sites under the CCA
  - The total payload mass carried by NASA launchers
  - Avg. Payload mass carried by F9 v1.1
  - Date of first successful landing of F9
  - Names of boosters that successfully carried drone shipments of payload masses between 4000 and 6000 kg
  - Total success & failure ratio
  - How many boosters carried the maximum amt of payload mass
  - Failed landings in drone shipment, and their specifics such as: booster version and launch pads in 2015
  - Successful/Failure ratio of rockets in the dates between 06/04/2010 – 03/20/2020

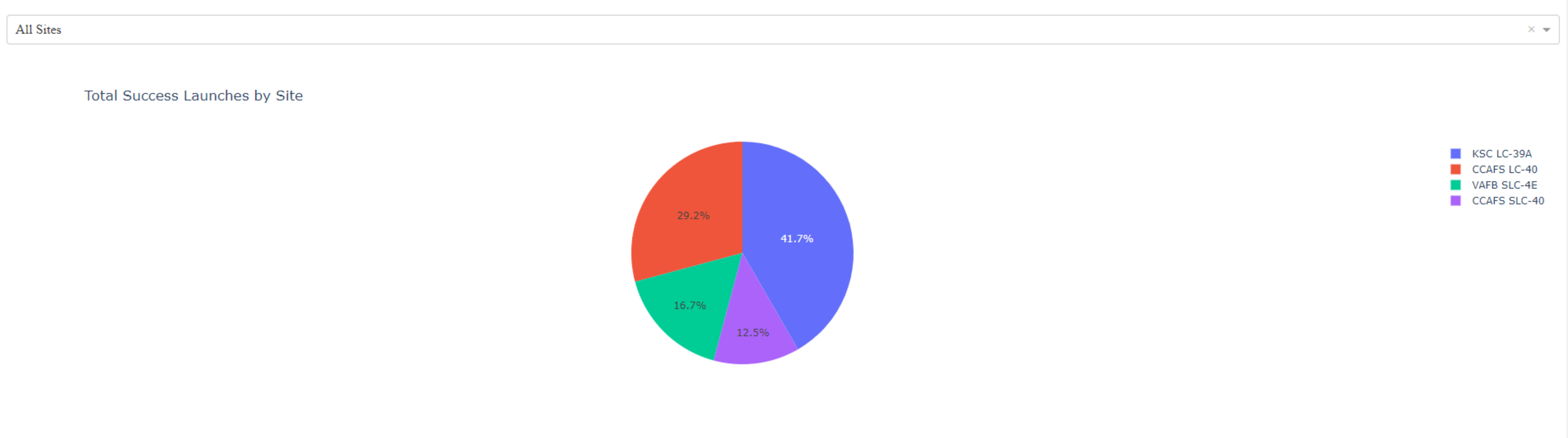
Build an Interactive Map with Folium

- I made interactive Folium maps of the launch sites used, using markers, circles, lines & marker clusters
- Markers indicate launch points
- Circles highlight areas of importance such as the NASA Johnson Space Center
- Market Clusters highlight area of an instance occurring multiple times at once, such as launches at a launch site.
- Lines are used to measure distance between two points, such as the NASA Johnson Space Center & nearby railroads/coasts.

Build a Dashboard with Plotly Dash

- I used Percentage graphs & Range graphs to visualize the relation of Payloads & the respective launch sites, helping to ID the best place to launch certain payloads.

Image is a Plotly Dashboard Chart

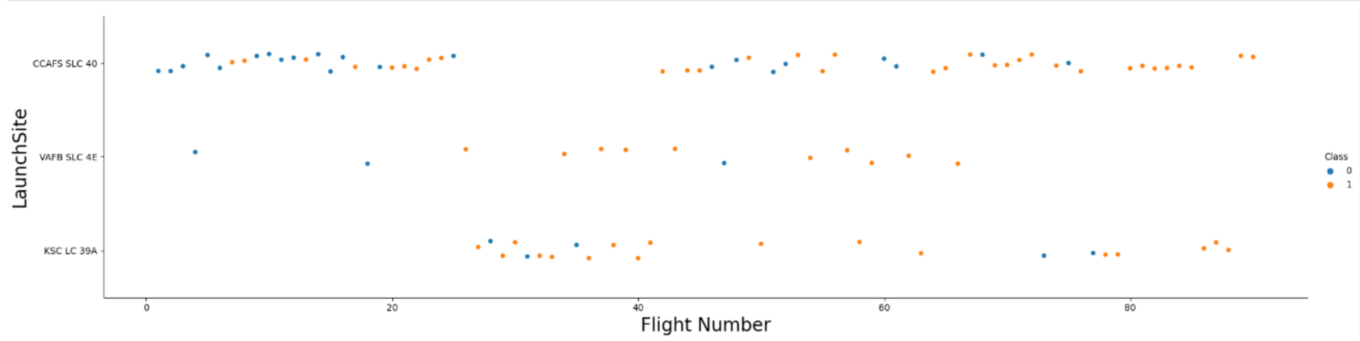


Predictive Analysis (Classification)

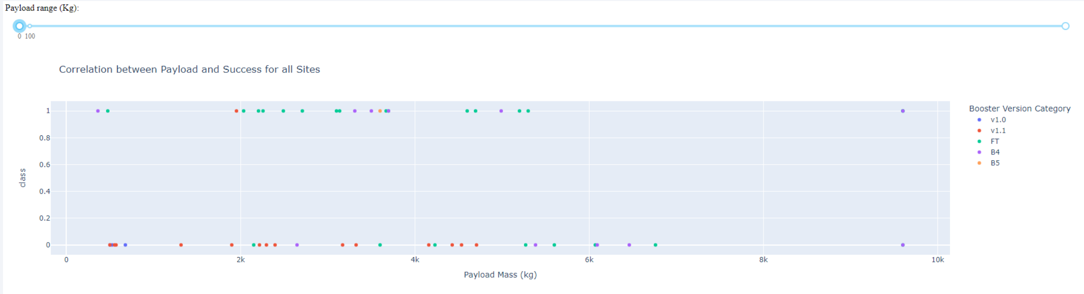
- Four classification models were compared: logistic regression, support vector
  In this method:
- First I Prepped the data and standardized it all into a single cohesive thing.
- Second I Tested the data rigorously with multiple combinations of their parameters.
- Finally I looked at the results of the output of each Classification model.

Results

- These were the results of my labor:
  - Space X uses 4 different launch sites
  - The average payload of F9 v1.1 booster is 2,928 kg
  - The first success landing outcome happened in 2015 five years after the first launch
  - Almost 100% of missions were successful
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015



	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

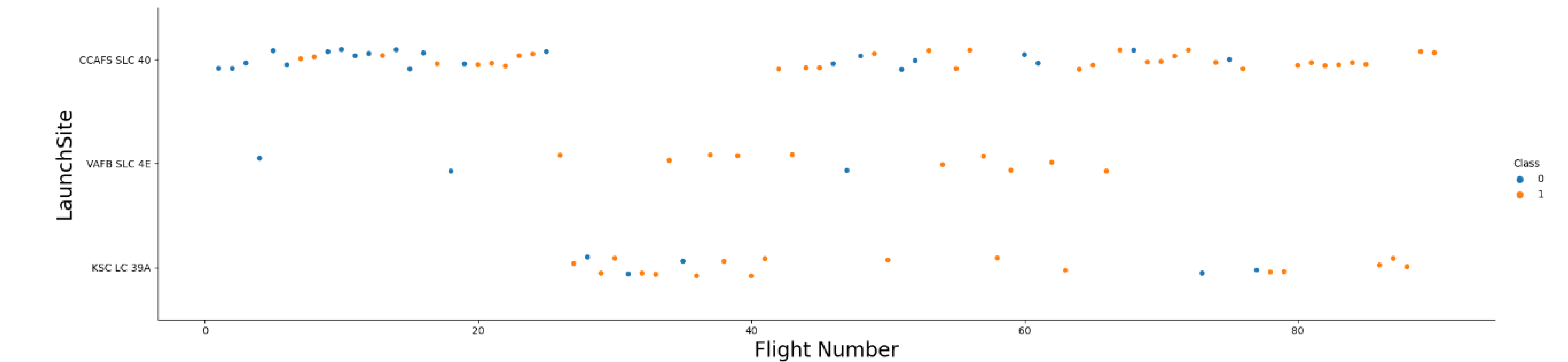


Section 2

Insights drawn from EDA

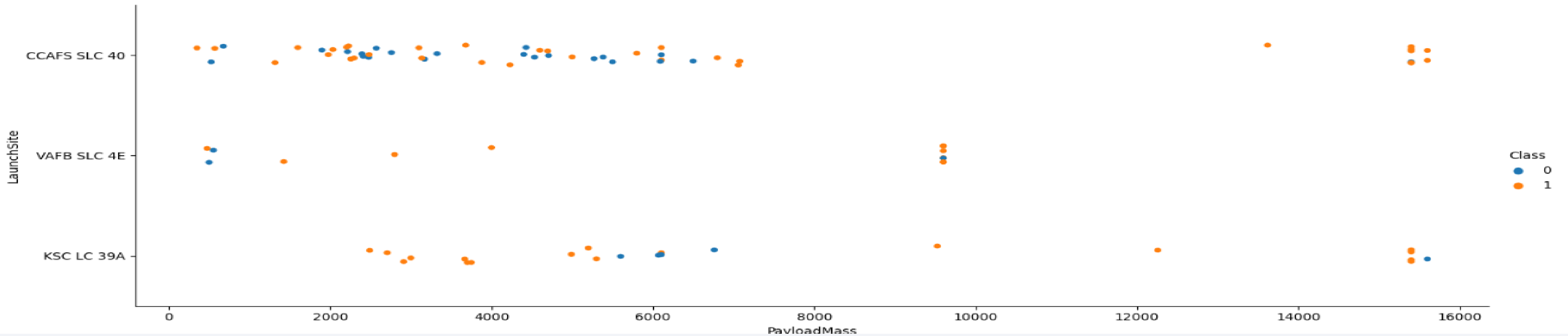
Flight Number vs. Launch Site Scatter Plot

- According to the scatter plot, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful.
- KSC LC 39A is overall the most successful one of them all from the time frame shown.
- VAFB SLC 4E has the least amount of tests and has not been tested for a while.



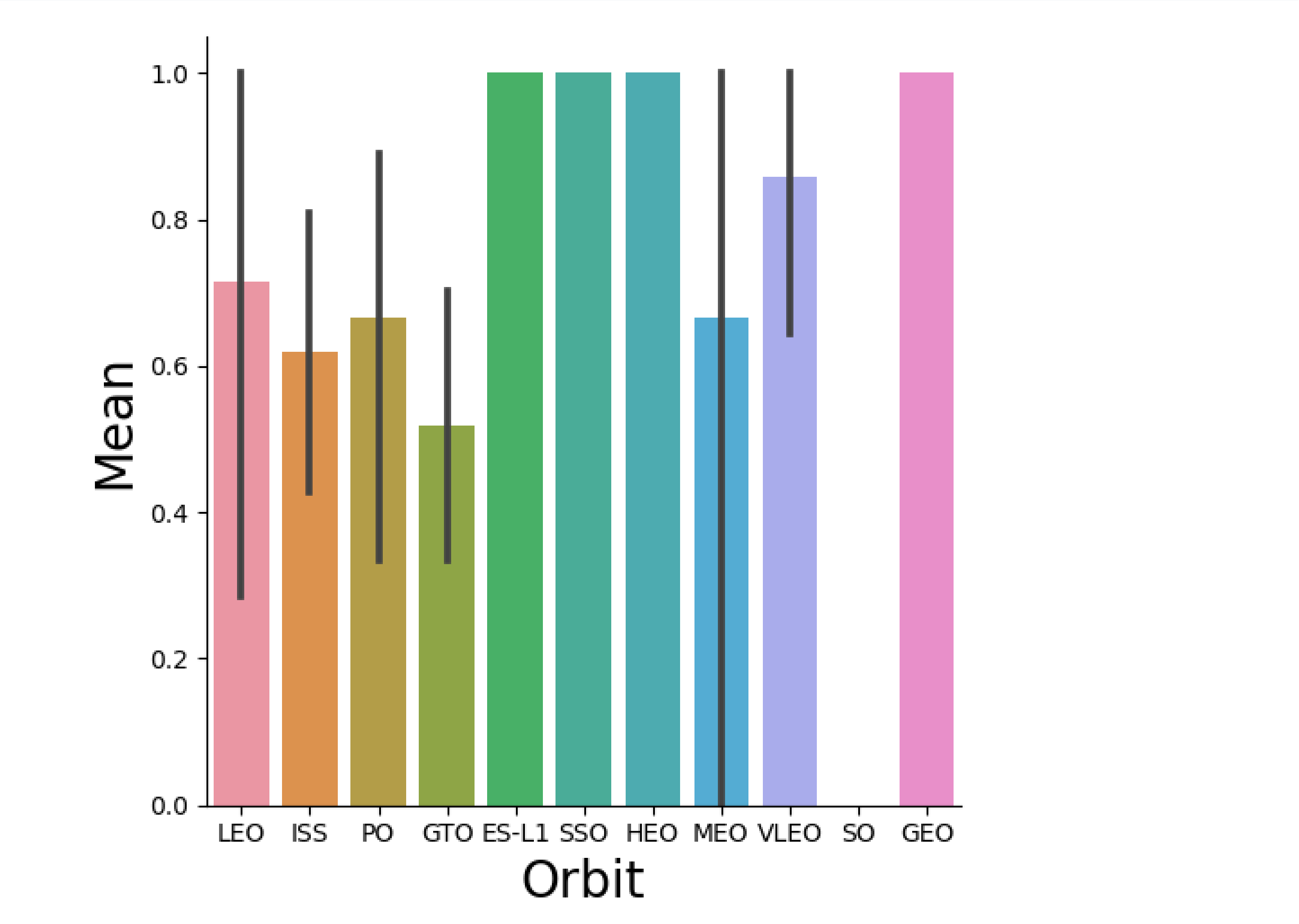
Payload vs. Launch Site Scatter Plot

- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.
- Payloads near 4,000 and 8,000 kg seem to mostly be unsuccessful while payloads over 8,000kg seem to be mostly successful while >4,000 kg is a mixed bag.



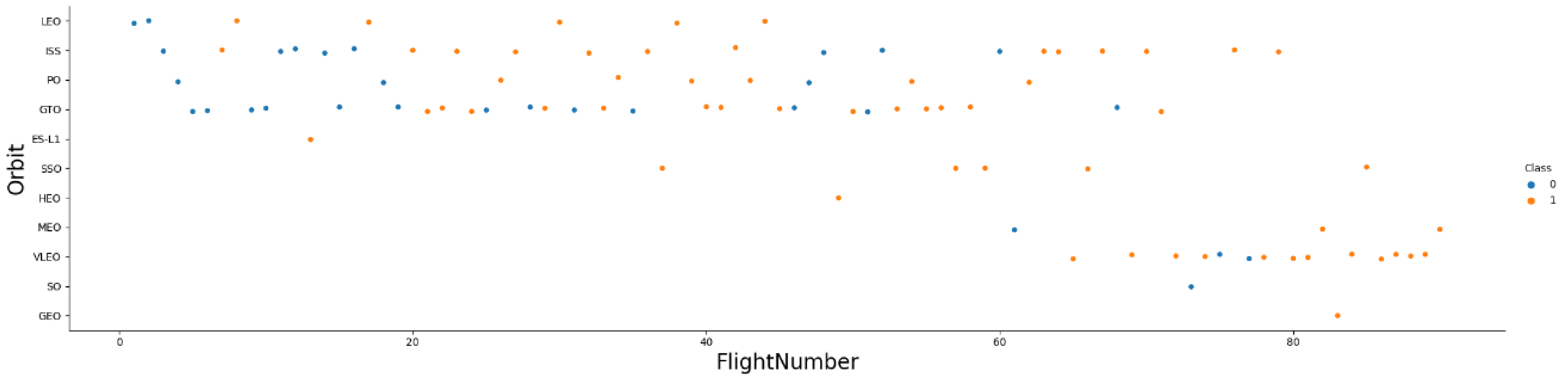
Success Rate vs. Orbit Type Bar Chart

- SSO, HEO, and ES-L1 have highest average success rates.
- The black lines represent the outliers.
- SO Appears to not be tested.



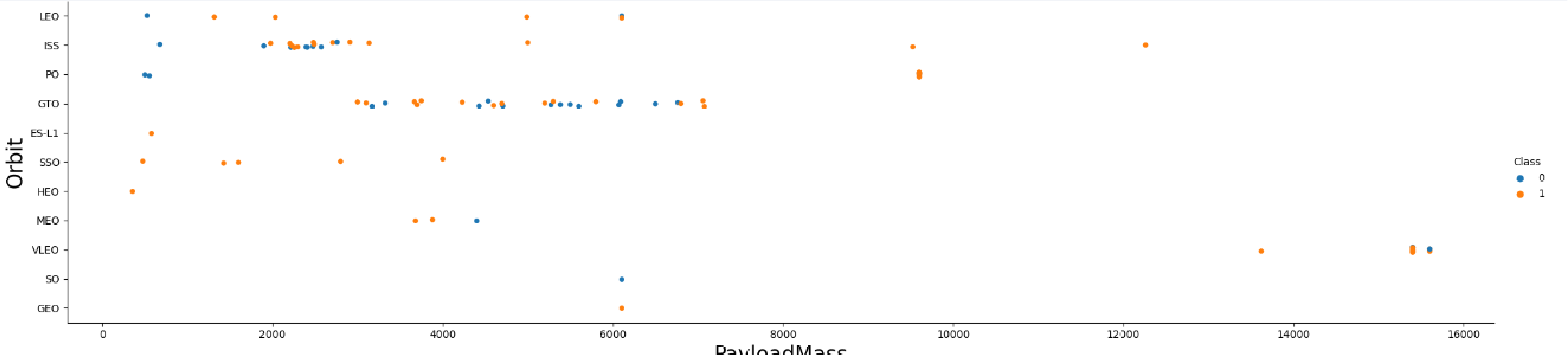
Flight Number vs. Orbit Type Scatter Plot

- Initial orbits were usually unsuccessful, SSO has the best success rate of all orbits.
- GEO and HEO Orbits have been tested only once
- ISS orbit has been the most rigorously tested, with a 62% success rate



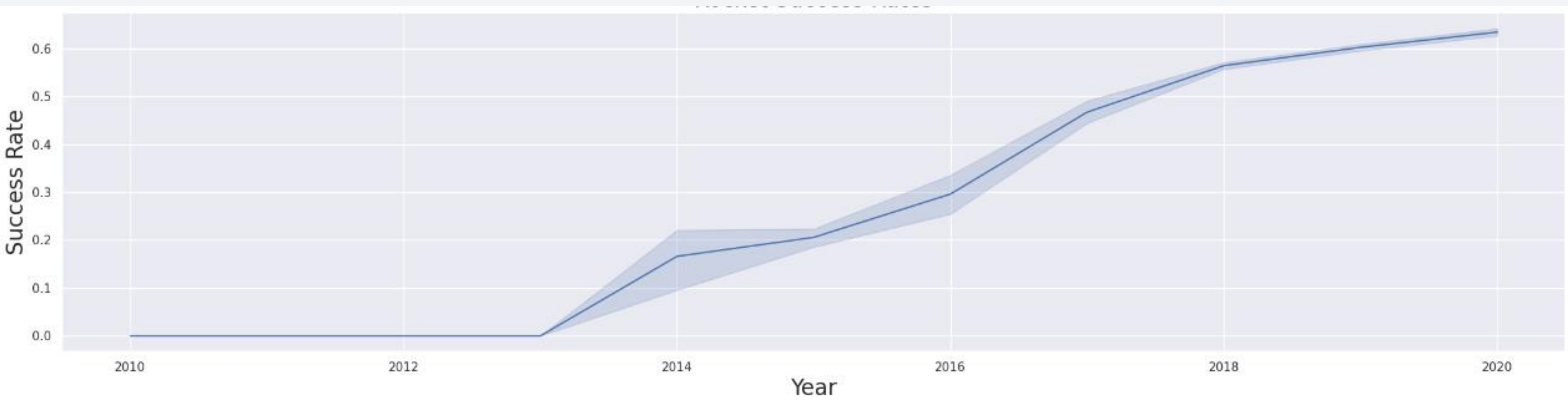
Payload vs. Orbit Type scatter plot

- LEO, ISS and PO have low successful launches with lighter payloads.
- MEO, VLEO, SO & GEO need further tests to give accurate results
- ISS & GTO have the most rigorous testing and has the best success rate of the two.



Launch Success Yearly Trend

- Launches have steadily gotten more successful with no downturn.
- Rocket launches were started in 2013 and are plateauing near 2020



Other Neat Tidbits

- Launch Site names are: CCFAS LC-40, CCFAS SLC-40, KSC LC-39A, VAFB LC-39A
- 5 Records that start with 'CCA':

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- Total Payload of NASA boosters:

TOTAL_PAYLOAD
111268

- Avg payload mass carried by booster version F9 v1.1:

AVG_PAYLOAD
2928.4

- Date of first successful Landing:

FIRST_SUCCESS_GP
01-05-2017

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

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

Other Neat Tidbits pt. 2

- Total number of successful and failure mission outcomes:
- Boosters which have carried the maximum payload mass:
- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

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

```
%sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE [LANDING _OUTCOME] = 'Failure (drone ship)' AND DATE BETWEEN '2015-01-01' AND '2015-12-31'
 * sqlite:///my_data1.db
Done.
```

Booster_Version	Launch_Site

Other Neat Tidbits pt. 3

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

```
%sql SELECT [LANDING _OUTCOME], COUNT(*) AS QTY FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY [LANDING _OUTCOME] ORDER BY QTY DESC;
```

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

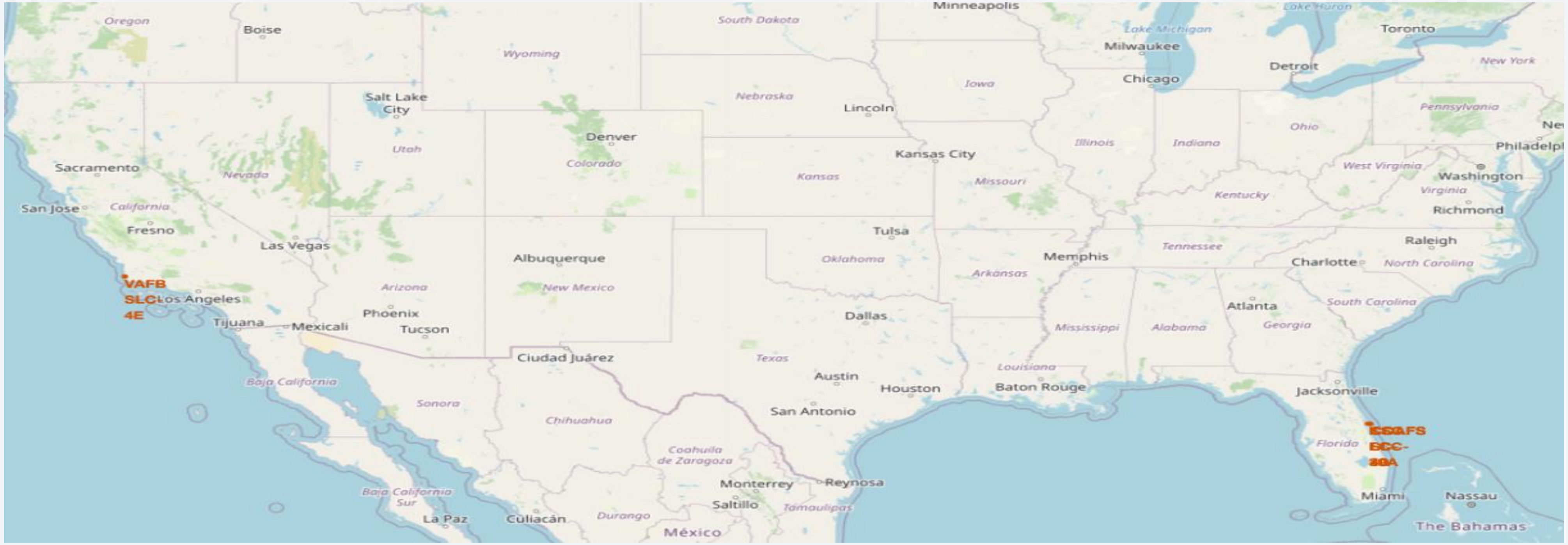
Landing _Outcome	QTY

Section 3

Launch Sites Proximities Analysis

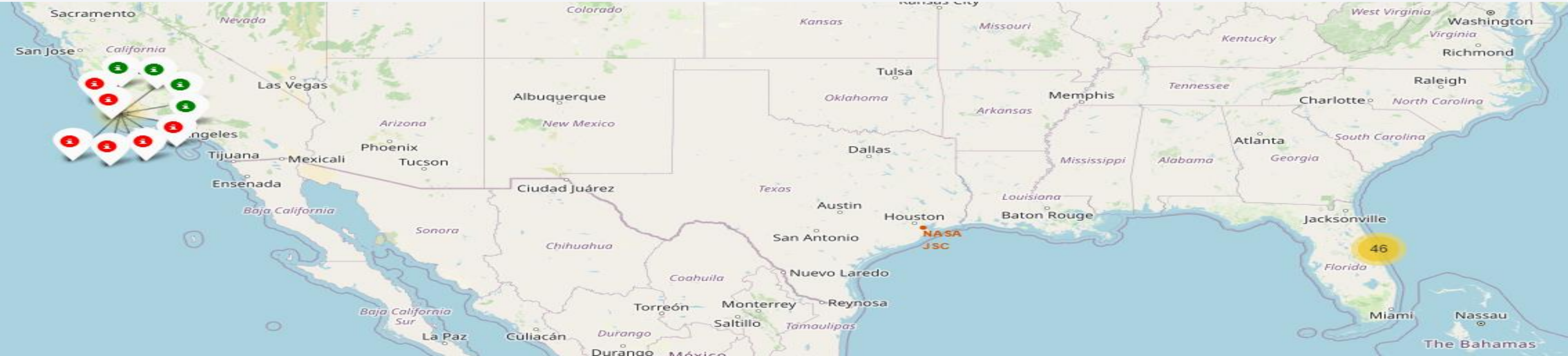
All launch sites

- Launch sites are all on the opposite sides of the country to avoid any serious rocket-on-rocket catastrophe possibly, and nearby seas to avoid big impact craters.



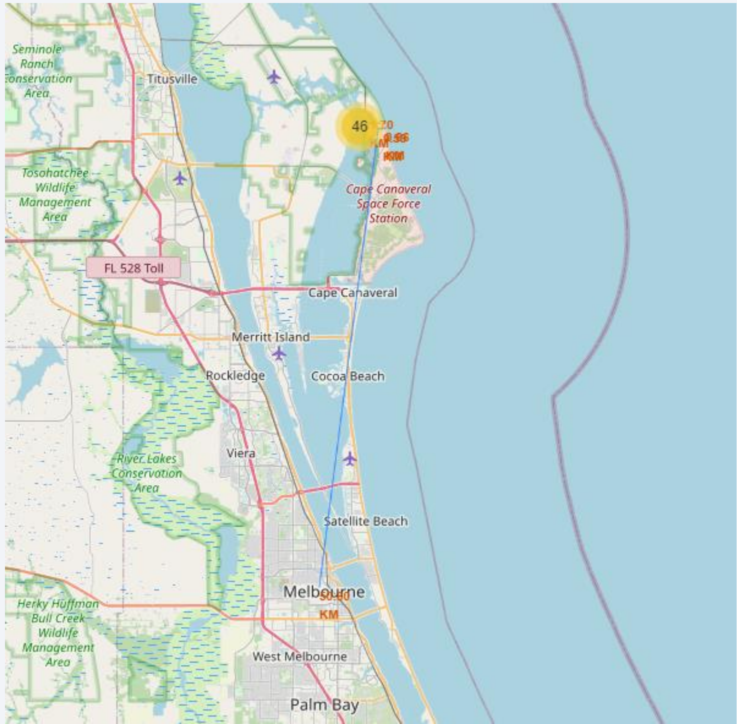
Launch Outcomes by Site

- This is an example of the VAFB SLC-4E launch site outcomes, green is successful, red is unsuccessful.



Logistics and Safety

- Cape Carnival Space station is a bit too close to a densely populated City, Melbourne.

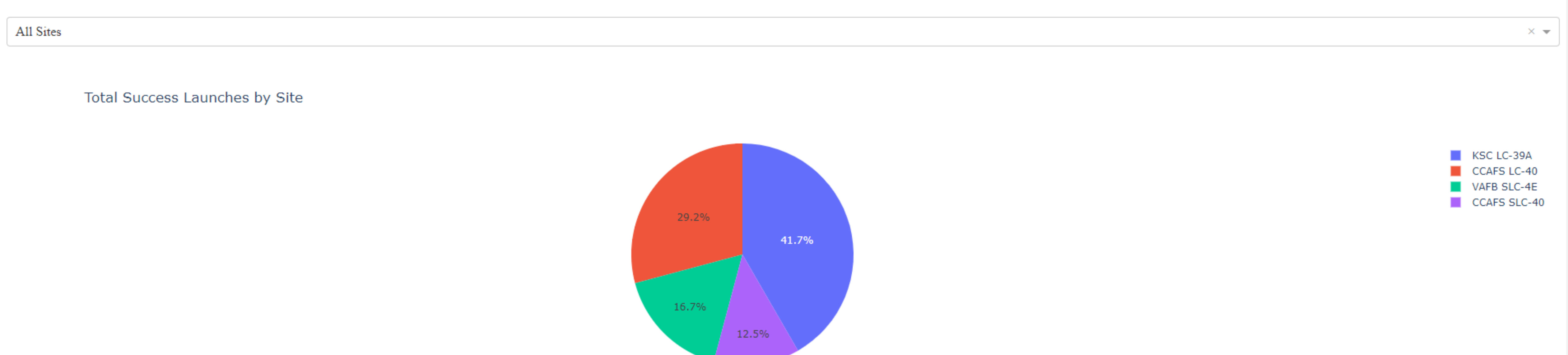


Section 4

Build a Dashboard with Plotly Dash

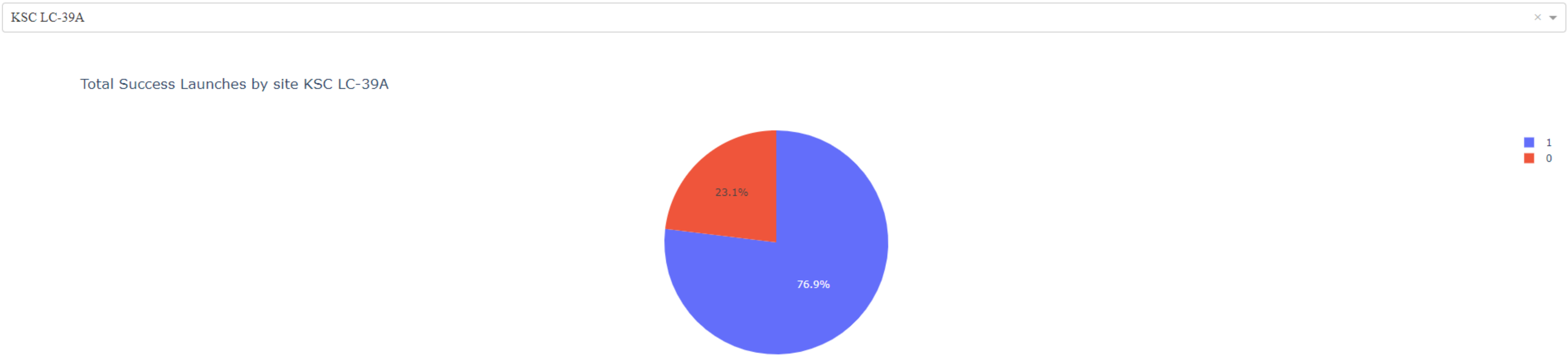
Success Count for All Sites

- KSC-LC39A has the most successful launches, so it should be replicated to the CCAFS SLC-40 and VAFB SLC-4E sites.



Highest launch success

- LC-39A has a very high success rate compared to the others, where roughly up to 40% can be failures.



Scatter Plot

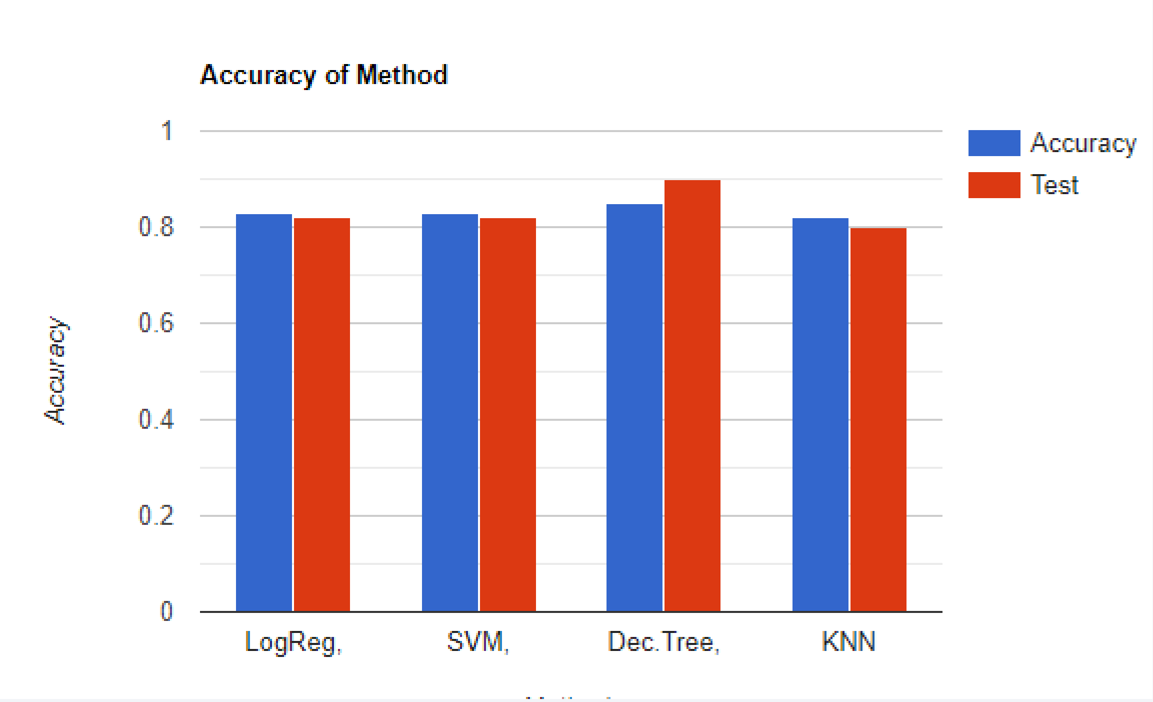
- Payloads under 6,000kg and FT boosters are the most successful combination, and there is not enough data over 7k to give valid input for these tests.

Section 5

Predictive Analysis (Classification)

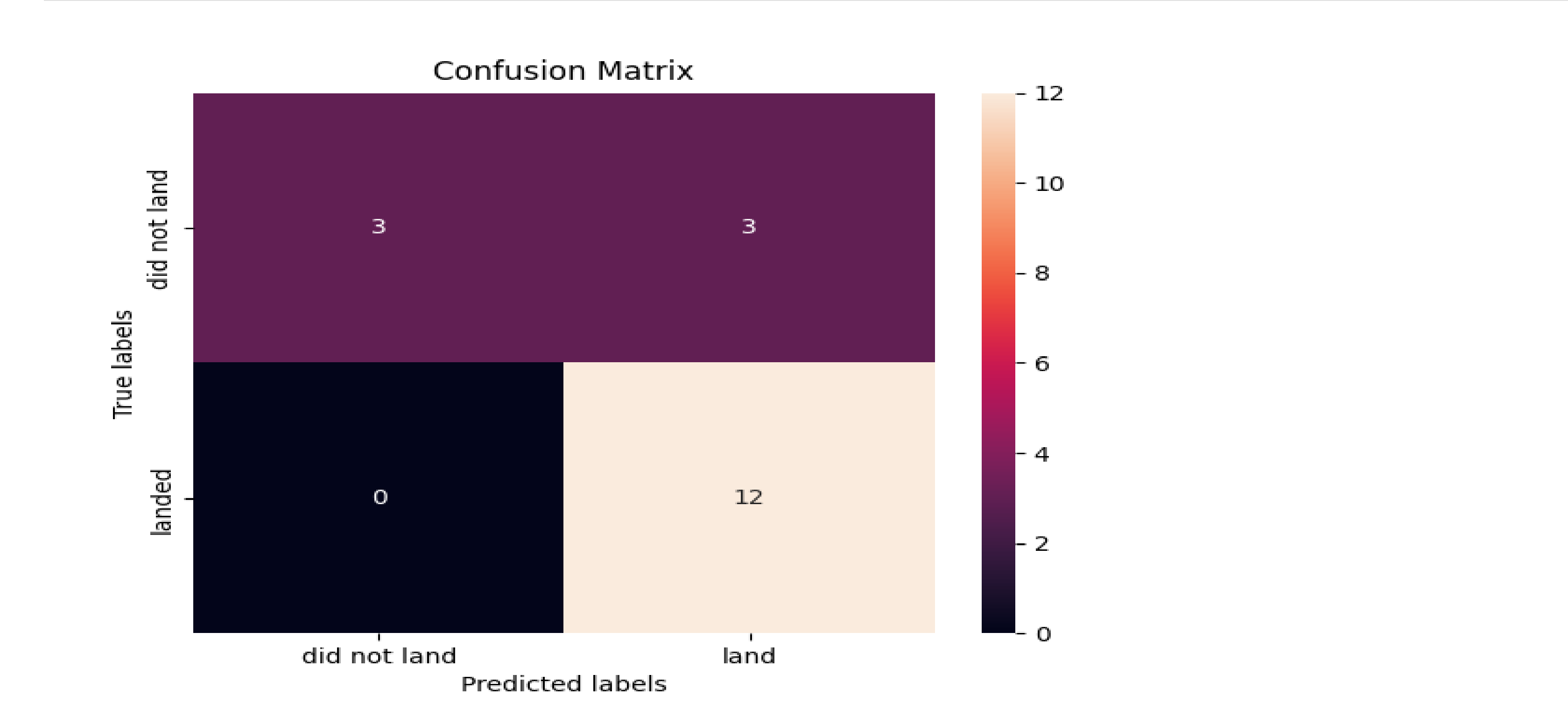
Classification Accuracy

- The model with the highest classification accuracy is Decision Tree Classifier, which is 87%.
- The model with the lowest classification accuracy is SVM, which is 83%



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.



Conclusions

- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less likely to fail.
- Successful landing outcomes seem to improve over time with technological advancements, meaning they’re learning from successful launches and are using those results.
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Thank you!