IBM Developer
SKILLS NETWORK

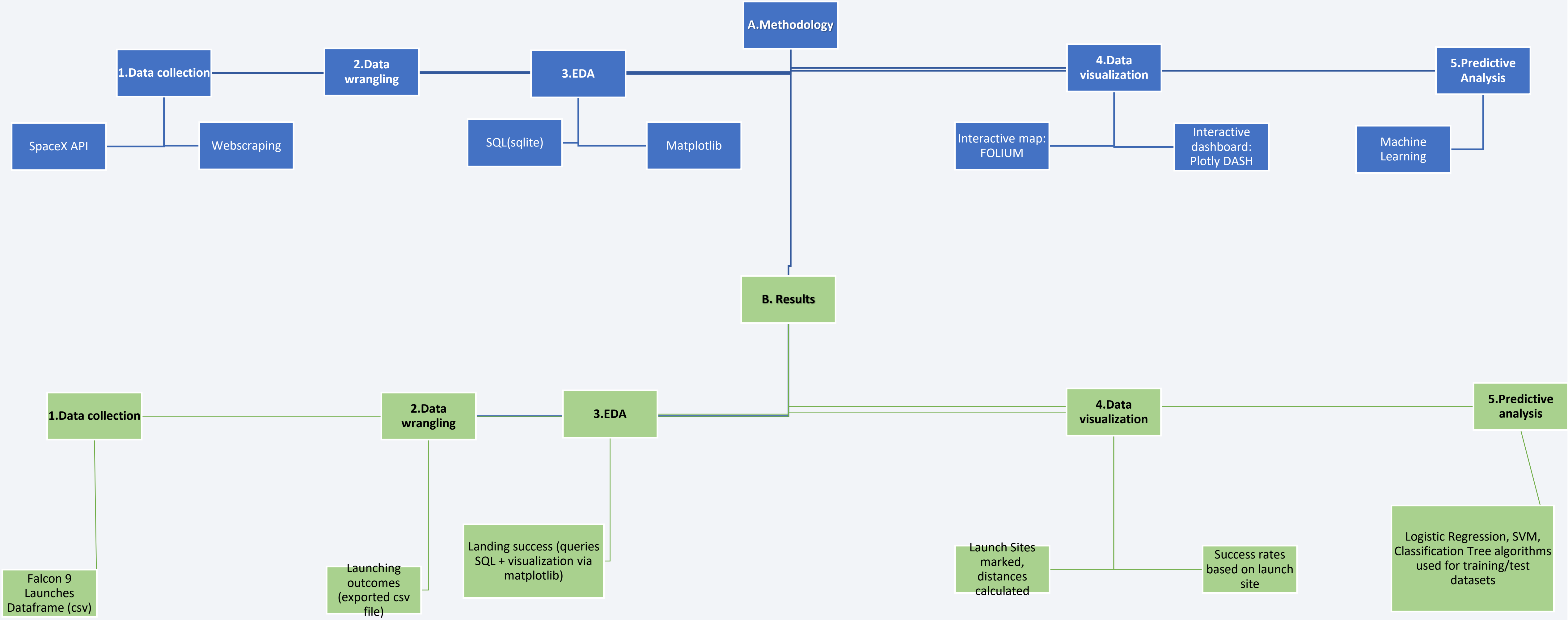
Winning Space Race with Data Science

<Stefan Liute>
<03.12.2023>

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary



Introduction

- SpaceX has deployed several spaceships over the years.
- 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. Therefore if we can determine if the first stage will land, we can determine the cost of a launch.
- In this project we predict if the Falcon 9 stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- Perform data wrangling
  - The collected data was processed using pandas, numpy transforming raw data into usable data for further analysis
- 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

SpaceX API

"https://api.spacexdata.com/v4/launches/past"

- Collected the useful data for Falcon 9. Cleaned the data and exported it in a csv file.

BoosterVersion
PayloadMass
Orbit
LaunchSite
Outcome
Flight
GridFins
Reused
Legs
LandingPad
Block
ReusedCount
Serial
Longitude
Latitude

Webscraping

"https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches"

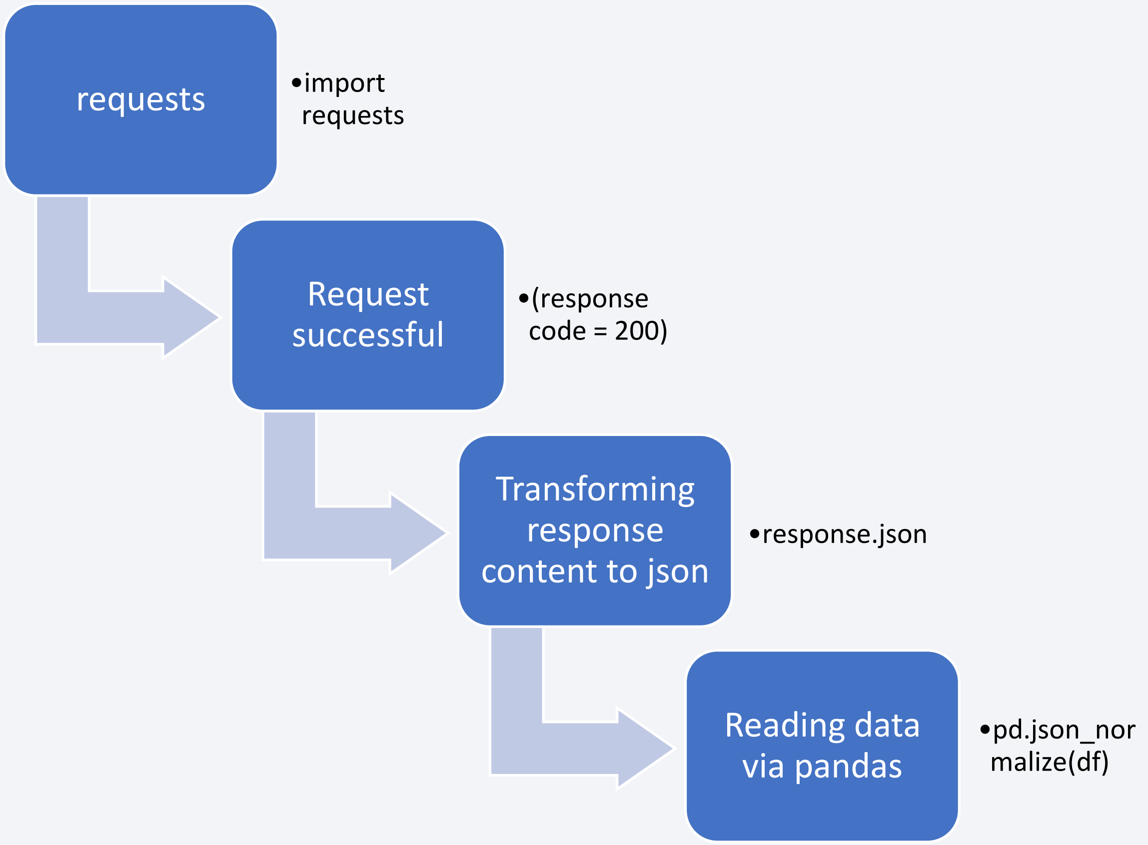
- Extracted all the data for the Falcon 9 Launch

Flight No.	Payload	Payload mass	Orbit	Customer
78	Starlink	15,600 kg	LEO	SpaceX
79	Crew Dragon in-flight abort test	12,050 kg	Sub-orbital	NASA
80	Starlink	15,600 kg	LEO	SpaceX
81	Starlink	15,600 kg	LEO	SpaceX
82	SpaceX CRS-20	1,977 kg	LEO	NASA
...	...	...	...	...
268	Starlink Group 6-25	~18,400 kg	LEO	SpaceX
269	Starlink Group 6-26	~18,400 kg	LEO	SpaceX
270	Starlink Group 6-27	~18,400 kg	LEO	SpaceX
271	SpaceX CRS-29	~9,525 kg	LEO	NASA
272	None	U	SSO	None

Launch outcome	FH 4	Date	Time	Version Booster	Launch Site	Booster landing
S	NaN	7 January 2020	02:19:21	F9 B5	CCSFS	Success
Success\n	NaN	19 January 2020	15:30	F9 B5	KSC	Not attempted\n
Success\n	NaN	29 January 2020	14:07	F9 B5	CCSFS	Success
Success\n	NaN	17 February 2020	15:05	F9 B5	CCSFS	Failure
Success\n	NaN	7 March 2020	04:50	F9 B5	CCSFS	Success
...	...	...	...	...	...	...
Success\n	NaN	30 October 2023	23:20	F9 B5	CCSFS	Success
Success\n	NaN	4 November 2023	00:37	F9 B5	CCSFS	Success
Success\n	NaN	8 November 2023	05:05	F9 B5	CCSFS	Success
Success\n	NaN	10 November 2023	01:28	F9 B5	KSC	Success
Success\n	NaN	11 November 2023	18:49	F9 B5	VSFB	Success

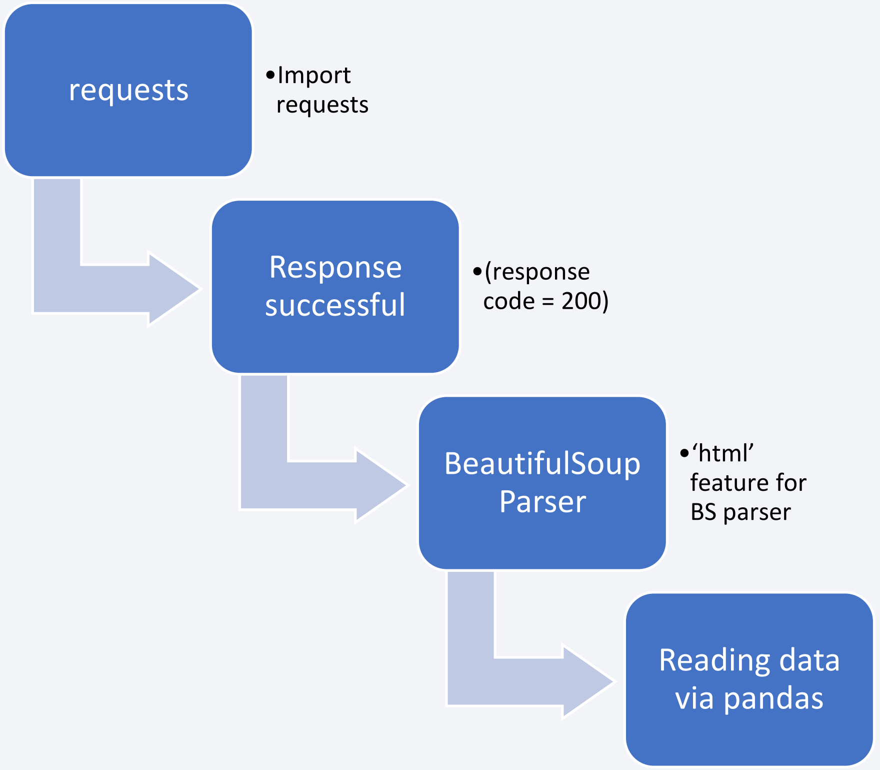
Data Collection – SpaceX API

- Link to GitHub jupyter notebook
- Collected Falcon 9 Launch information



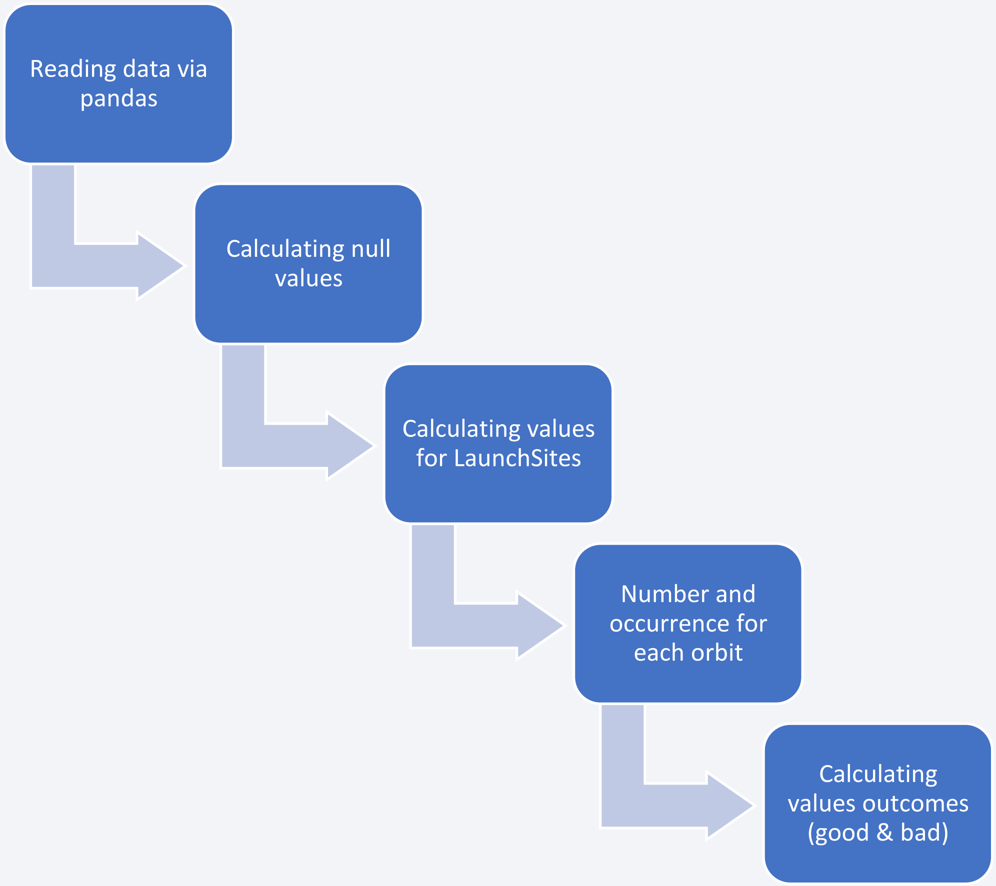
Data Collection - Scraping

- [Link to GitHub jupyter notebook]()
- Scraped Falcon 9 Launch information from Wikipedia (tables)



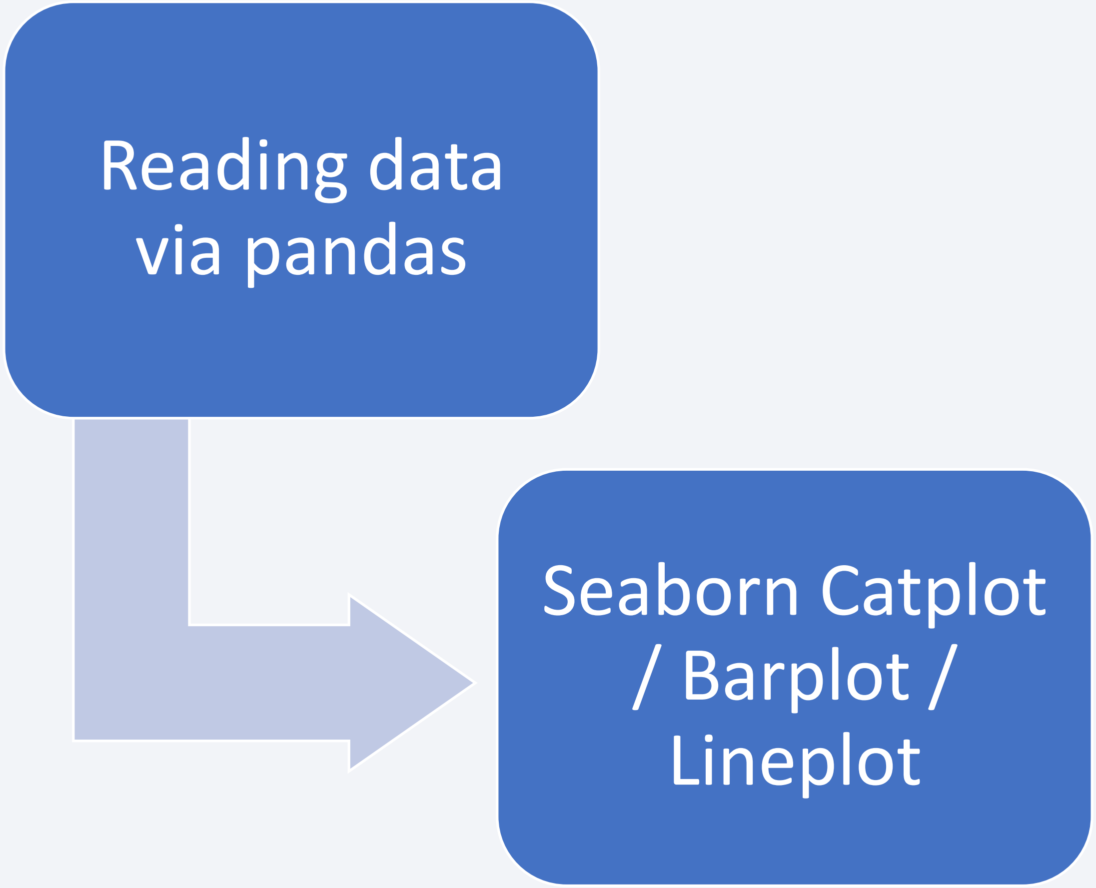
Data Wrangling

- [Link to GitHub jupyter notebook]
- Transformed raw data into useful data regarding occurrence for each orbit, outcomes of launches, etc.
- Exported it to csv file



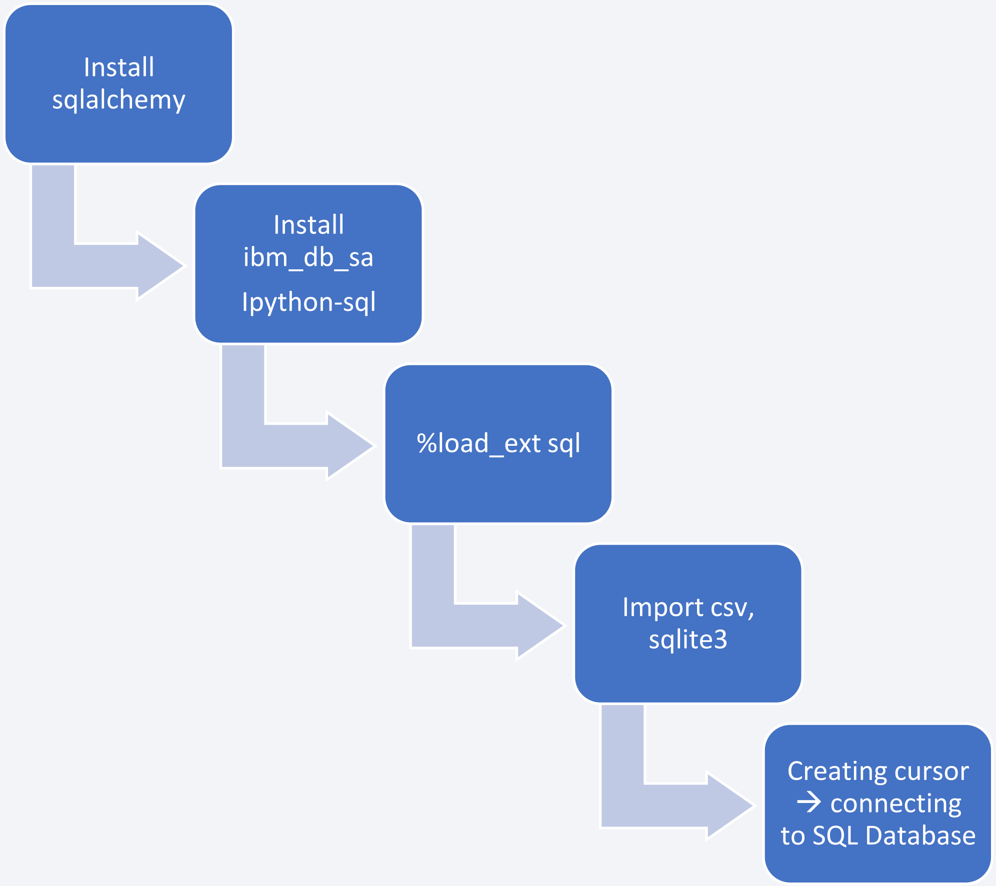
EDA with Data Visualization

- [Link to GitHub jupyter notebook]
- Visualized different relationships between Launches and Orbits, outcome information, etc



EDA with SQL

- [Link to GitHub jupyter notebook]
- Queried useful info for data visualization



Build an Interactive Map with Folium

- Created a blue circle at NASA Johnson Space Center's coordinate with a popup label showing its name
- For each launch site, I added a Circle object based on its coordinate (Lat, Long) values. In addition, I added the Launch site name as a popup label
- Calculated the distance from each launch site to the equator line

[Link to GitHub jupyter notebook]

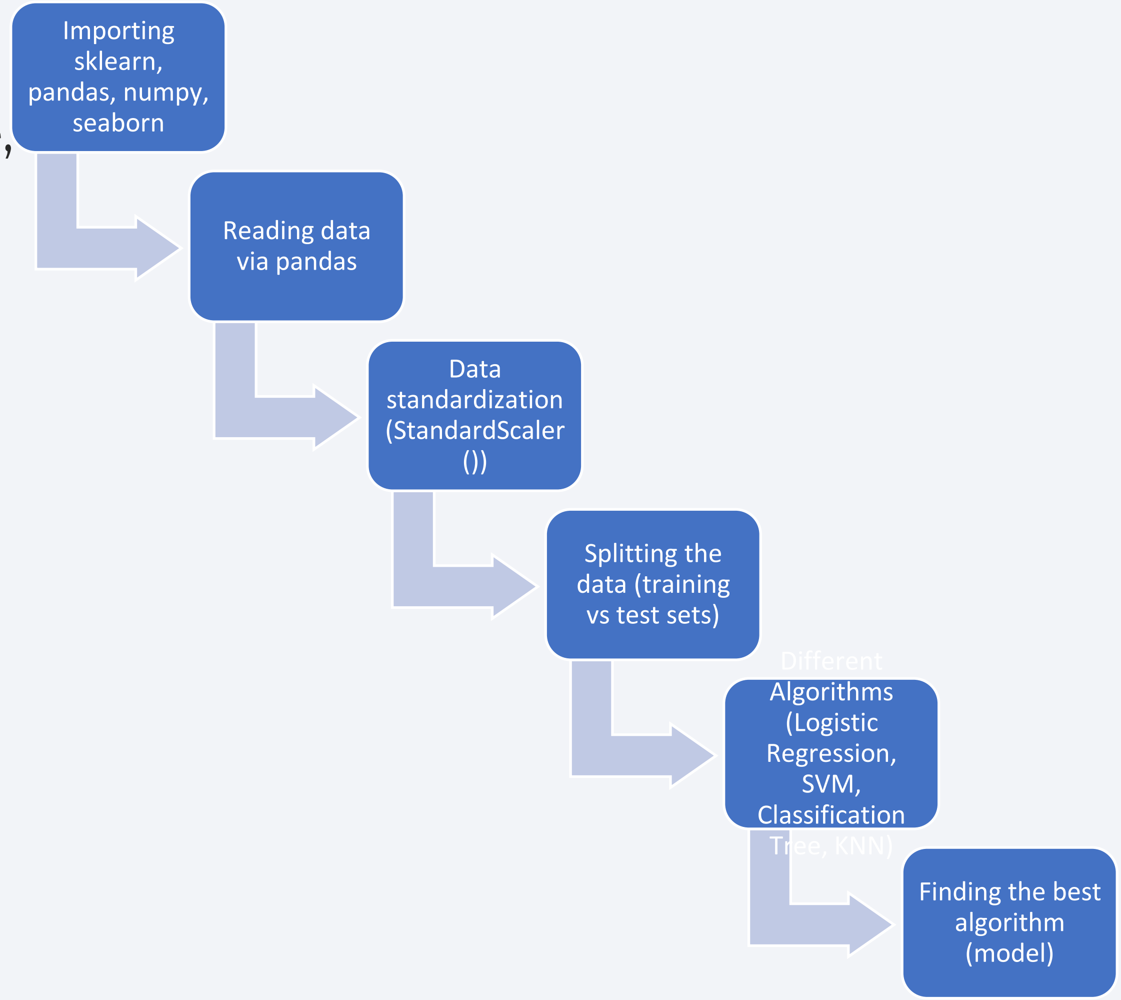
Build a Dashboard with Plotly Dash

- I added Success vs Failed Launches for all sites (combined or individual)
- Total success launches by site
- A payload range (kg) criteria with live visualization depending on the value selected
- Payload mass vs success for all sites (scatter plot)

[Link to GitHub jupyter notebook]

Predictive Analysis (Classification)

- Used Logistic Regression, SVM, Classification Tree, KNN algorithms for the split datasets (defining parameters ➔ GridSearch object ➔ fit the training sets through GridSearch ➔ print the best parameters and the accuracy ➔ plot Confusion Matrix)



[Link to GitHub]

Results

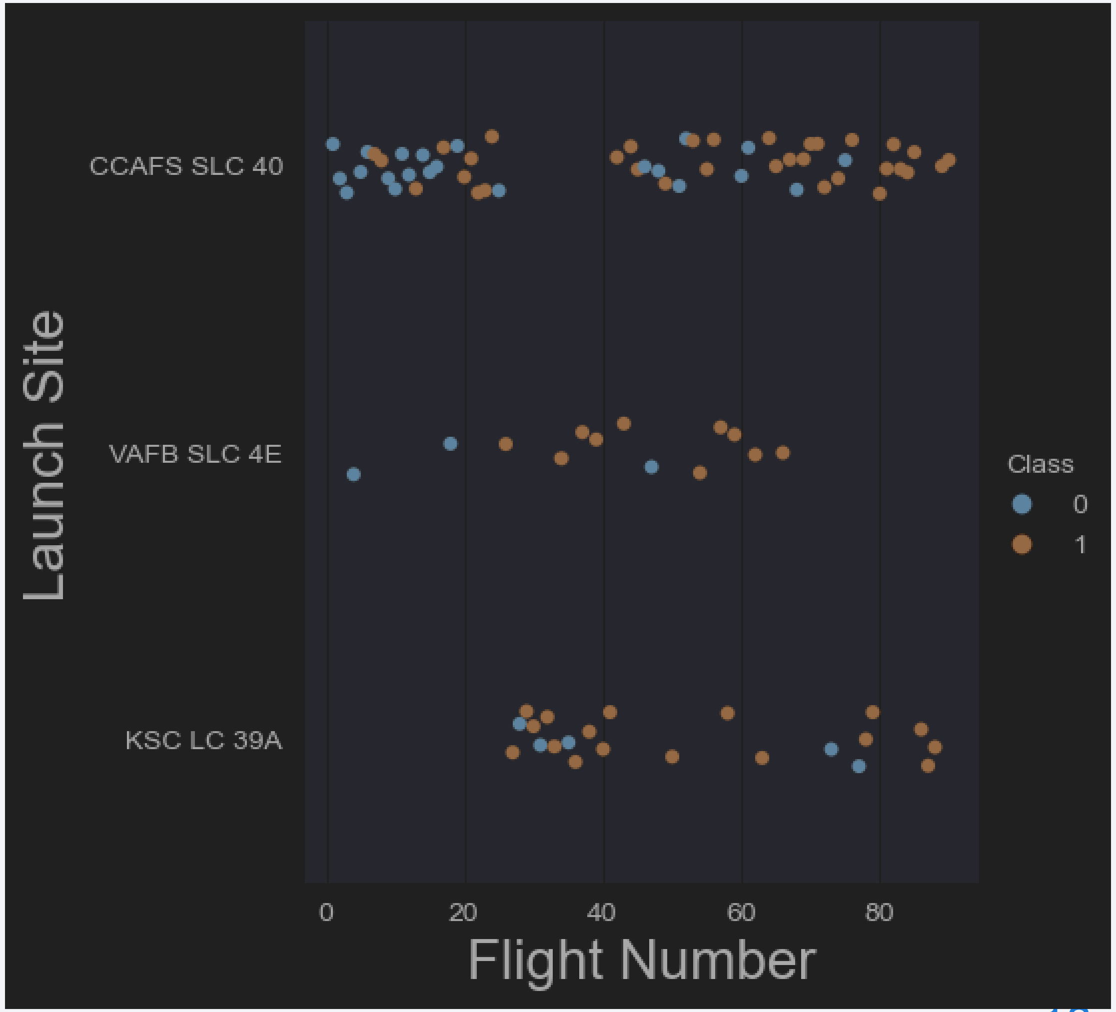
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

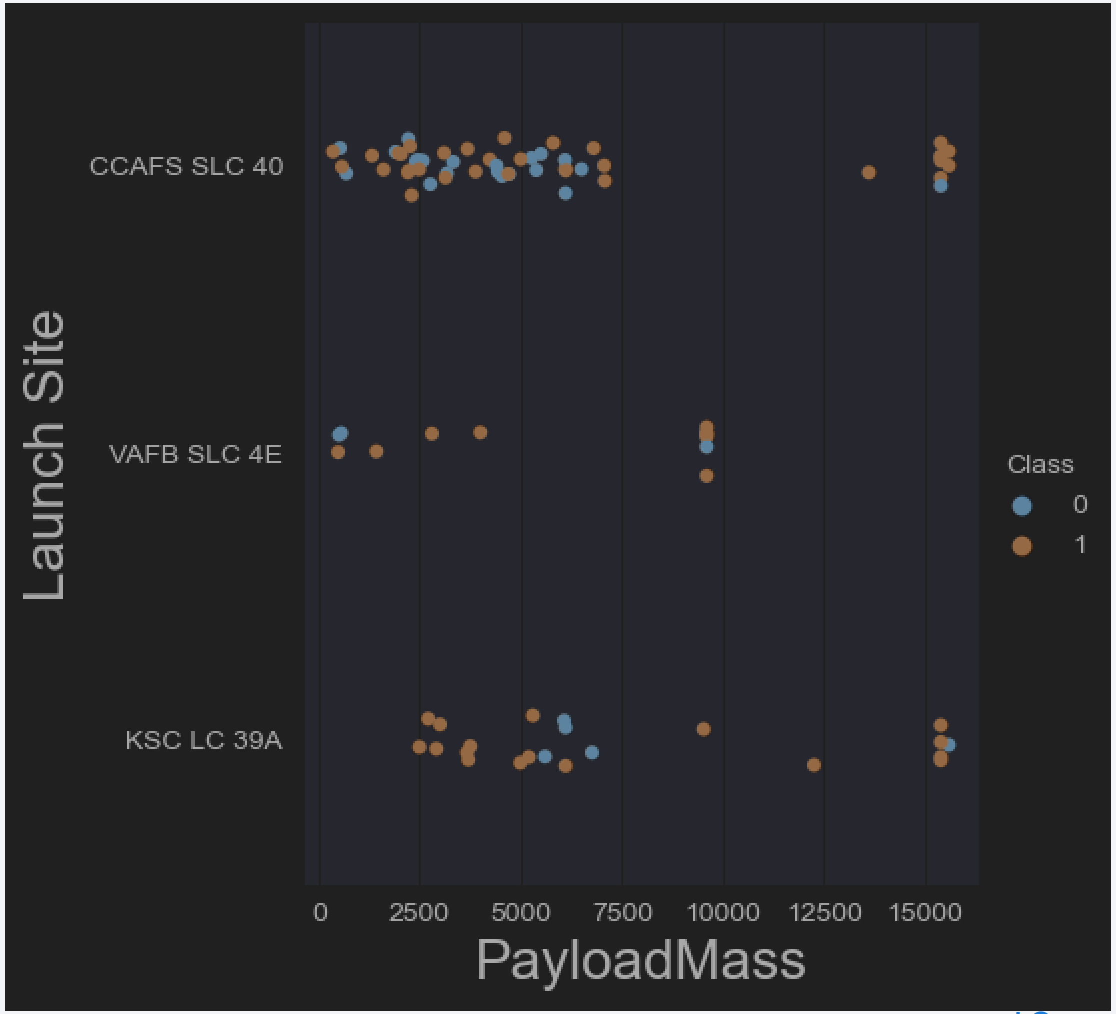
Flight Number vs. Launch Site

- We can see from this scatter plot that CCAFS SLC 40 launch site had the most flight numbers and the most failed launches. KSC LC 39A has the best success rate



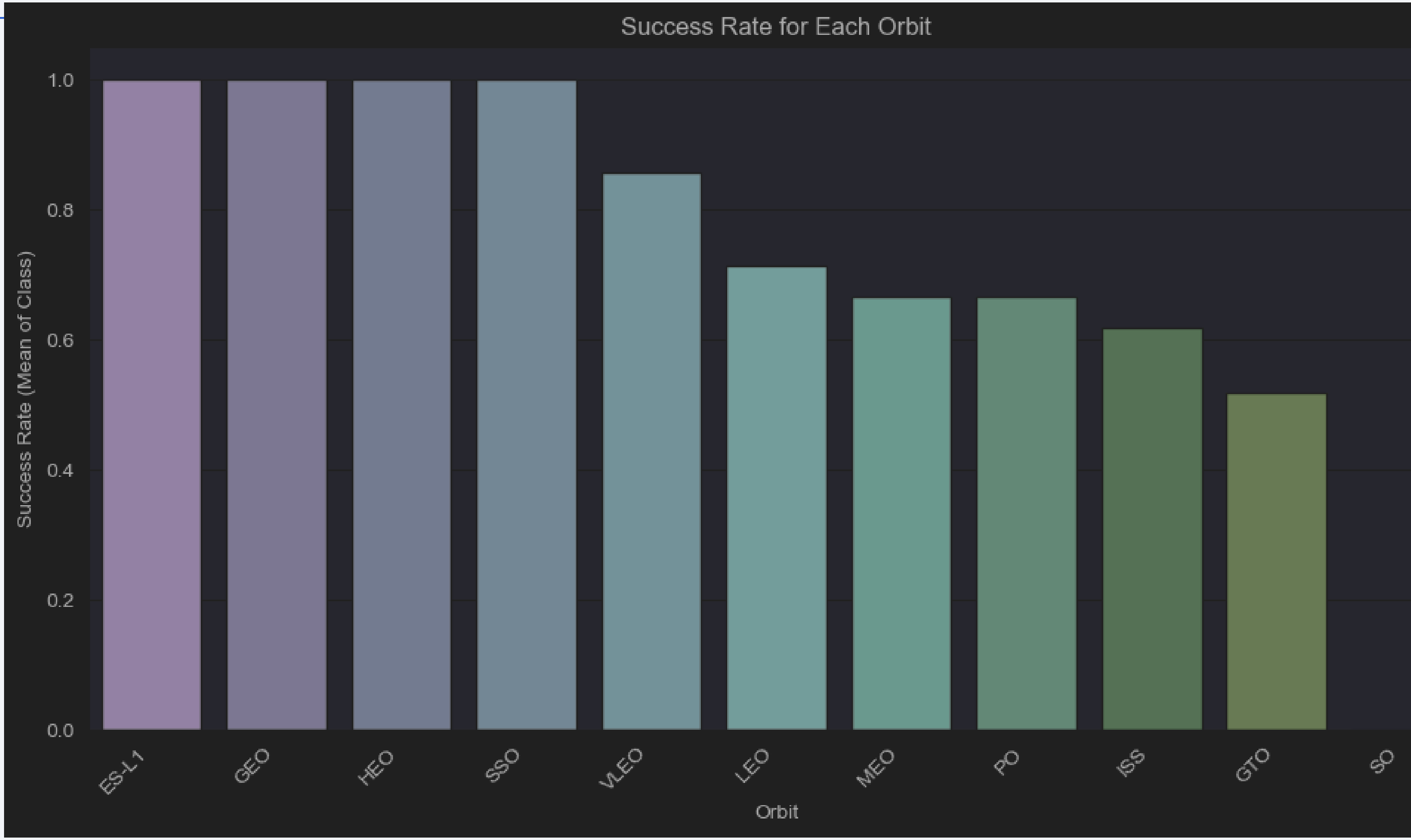
Payload vs. Launch Site

- We see that VAFB SLC 4E launch site doesn't have any launch over 10000 kg (payload mass)
- CCAFS SLC 40 doesn't have any launches between 7500 and approx. 12700 kg (payload mass), and has good launch success rate with payload mass over 15000 kgs along with KSC LC 39A launchsite



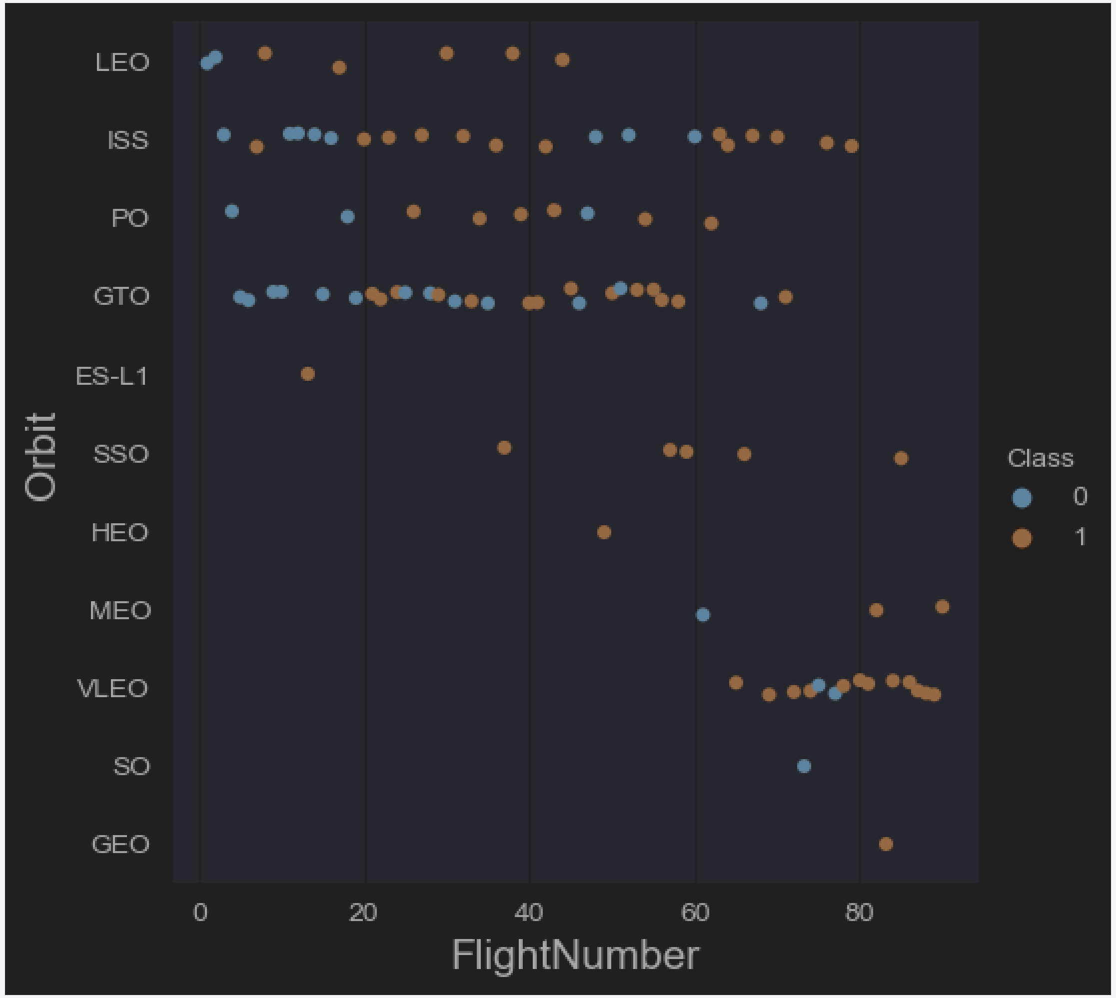
Success Rate vs. Orbit Type

- We can see that only the first 4 Orbits have the perfect success rate (1), the 5th one (VLEO) has a success rate of 0.8 and the rest of them have a success rate under 0,8 (SO orbit having the worst success rate – 0)



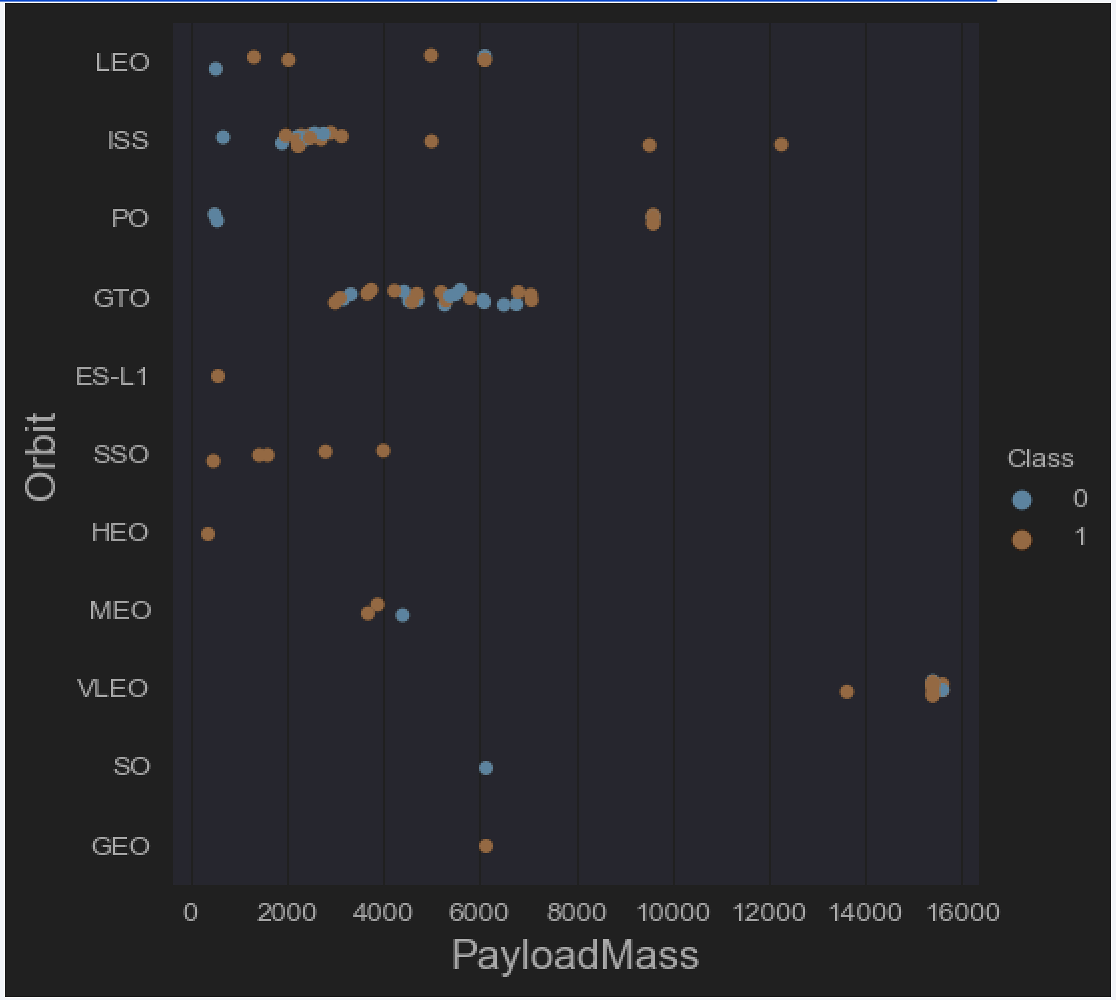
Flight Number vs. Orbit Type

- ES-L1, HEO, SO, GEO orbits had only one flight number with only SO orbit having 0 success rate.
- We can see that VLEO (the most successful orbit) started at over 60 flights (this could be the result of the improvement process over the years and flights) + it had several flights comparing to other orbits like SO, MEO, SSO, GTO which had under 3 flights launched over the 60th flight number



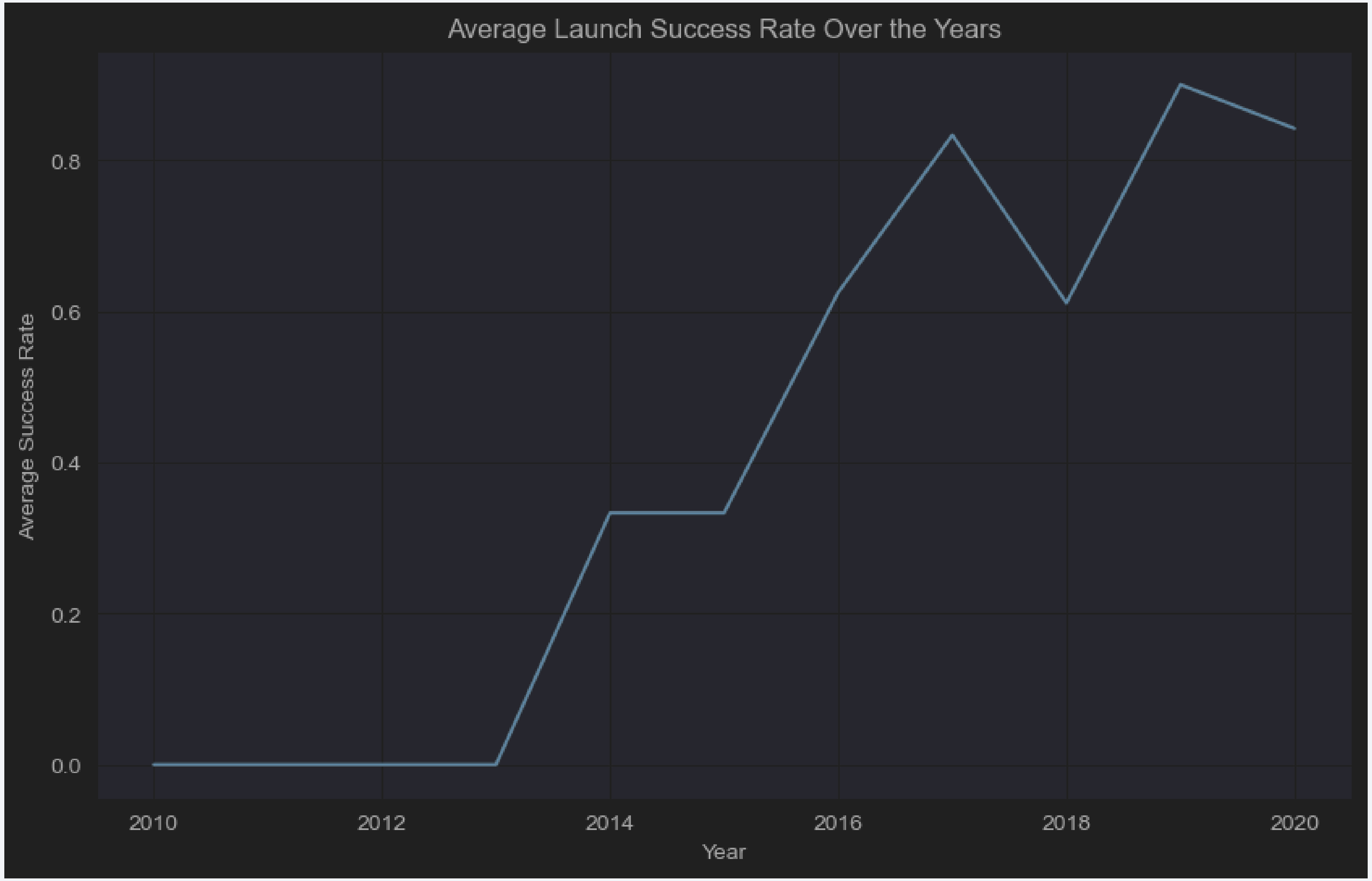
Payload vs. Orbit Type

- VLEO (the most successful orbit) had the highest Payload Mass (>14000)



Launch Success Yearly Trend

- From 2010 to 2013 the progress was stagnating. Since 2013 the success went higher until 2014-2015 (stall) then it started to progress until 2017 when it had a regress until 2018. 2018-2019 was an increasing trend until 2019 when it slightly decreased (until 2020 – the last year we have the data on)
- Overall, the average launch success rate is increasingly higher. This shows a clear improvement in technology, methodology of SpaceX launches.



All Launch Site Names

- %sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE
- 'CCAFS LC-40’,
- 'VAFB SLC-4E’
- 'KSC LC-39A’
- 'CCAFS SLC-40'

Launch Site Names Begin with 'CCA'

- 5 examples of site names starting with ‘CCA’. The query involves useful data like date&hour of the launch, BoosterVersion, site name, Outcome (success or failure)

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

[('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

- total payload carried by boosters from NASA(CRS)

```
 %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM
SPACEXTABLE

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

619967 (kg)

Average Payload Mass by F9 v1.1

- Average Payload Mass (kg) for Booster F9 v1.1

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

2928.4 (kg)

First Successful Ground Landing Date

- The first successful ground landing date

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

[('2015-12-22',)]
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- The Booster Versions where landing outcome was successful (drone ship) → output below
- * sqlite:///my_data1.db
- Done.

- [('F9 FT B1022',), ('F9 FT B1026',), ('F9 FT B1021.2',), ('F9 FT B1031.2',)]

Total Number of Successful and Failure Mission Outcomes

- * sqlite:///my_data1.db
- Done.
- * sqlite:///my_data1.db
- Done.
- +------------------------+
- | COUNT(Mission_Outcome) | SUCESSFUL
- +------------------------+
- | 100 |
- +------------------------+ +------------------------+
- | COUNT(Mission_Outcome) | FAILED
- +------------------------+
- | 1 |
- +------------------------+

Boosters Carried Maximum Payload

Selected all the distinct names of boosters that carried the max payload

('F9 v1.0 B0003',),
('F9 v1.0 B0004',),
('F9 v1.0 B0005',),
('F9 v1.0 B0006',),
('F9 v1.0 B0007',),
('F9 v1.1 B1003',),
('F9 v1.1',),
('F9 v1.1 B1011',),
('F9 v1.1 B1010',),
('F9 v1.1 B1012',),
('F9 v1.1 B1013',),
('F9 v1.1 B1014',),
('F9 v1.1 B1015',),
('F9 v1.1 B1016',),
('F9 v1.1 B1018',),
('F9 FT B1019',),
('F9 v1.1 B1017',),
('F9 FT B1020',),
('F9 FT B1021.1',),
('F9 FT B1022',),
('F9 FT B1023.1',),
('F9 FT B1024',),
('F9 FT B1025.1',),

('F9 FT B1038.1',),
('F9 B4 B1040.1',),
('F9 B4 B1041.1',),
('F9 FT B1031.2',),
('F9 B4 B1042.1',),
('F9 FT B1035.2',),
('F9 FT B1036.2',),
('F9 B4 B1043.1',),
('F9 FT B1032.2',),
('F9 FT B1038.2',),
('F9 B4 B1044',),
('F9 B4 B1041.2',),
('F9 B4 B1039.2',),
('F9 B4 B1045.1',),
('F9 B5 B1046.1',),
('F9 B4 B1043.2',),
('F9 B4 B1040.2',),
('F9 B4 B1045.2',),
('F9 B5B1047.1',),
('F9 B5B1048.1',),
('F9 B5 B1046.2',),
('F9 B5B1049.1',),
('F9 B5 B1048.2',),

('F9 FT B1026',),
('F9 FT B1029.1',),
('F9 FT B1031.1',),
('F9 FT B1030',),
('F9 FT B1021.2',),
('F9 FT B1032.1',),
('F9 FT B1034',),
('F9 FT B1035.1',),
('F9 FT B1029.2',),
('F9 FT B1036.1',),
('F9 FT B1037',),
('F9 B4 B1039.1',),
('F9 B5 B1047.2',),
('F9 B5 B1046.3',),
('F9 B5B1050',),
('F9 B5B1054',),
('F9 B5 B1049.2',),
('F9 B5 B1048.3',),
('F9 B5B1051.1',),
('F9 B5B1056.1 ',),
('F9 B5 B1049.3',),
('F9 B5 B1051.2 ',),
('F9 B5 B1056.2 ',),
('F9 B5 B1047.3 ',),
('F9 B5 B1048.4',),
('F9 B5B1059.1',),

('F9 B5 B1056.3 ',),
('F9 B5 B1049.4',),
('F9 B5 B1046.4',),
('F9 B5 B1051.3',),
('F9 B5 B1056.4',),
('F9 B5 B1059.2',),
('F9 B5 B1048.5',),
('F9 B5 B1051.4',),
('F9 B5B1058.1 ',),
('F9 B5 B1049.5',),
('F9 B5 B1059.3',),
('F9 B5B1060.1',),
('F9 B5 B1058.2 ',),
('F9 B5 B1051.5',),
('F9 B5 B1049.6',),
('F9 B5 B1059.4',),
('F9 B5 B1060.2 ',),
('F9 B5 B1058.3 ',),
('F9 B5 B1051.6',),
('F9 B5 B1060.3',),
('F9 B5B1062.1',),
('F9 B5B1061.1 ',),
('F9 B5B1063.1',),
('F9 B5 B1049.7 ',),
('F9 B5 B1058.4 ',)]

2015 Launch Records

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

- [('01', 'Failure (drone ship)', 'F9 v1.1 B1012', 'CCAFS LC-40'),
- ('04', 'Failure (drone ship)', 'F9 v1.1 B1015', 'CCAFS LC-40')]

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

- **Ranking landing outcomes (included the ones that didn’t attempt – no=10)**
- * sqlite:///my_data1.db
- Done.

- [('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)]

Section 3

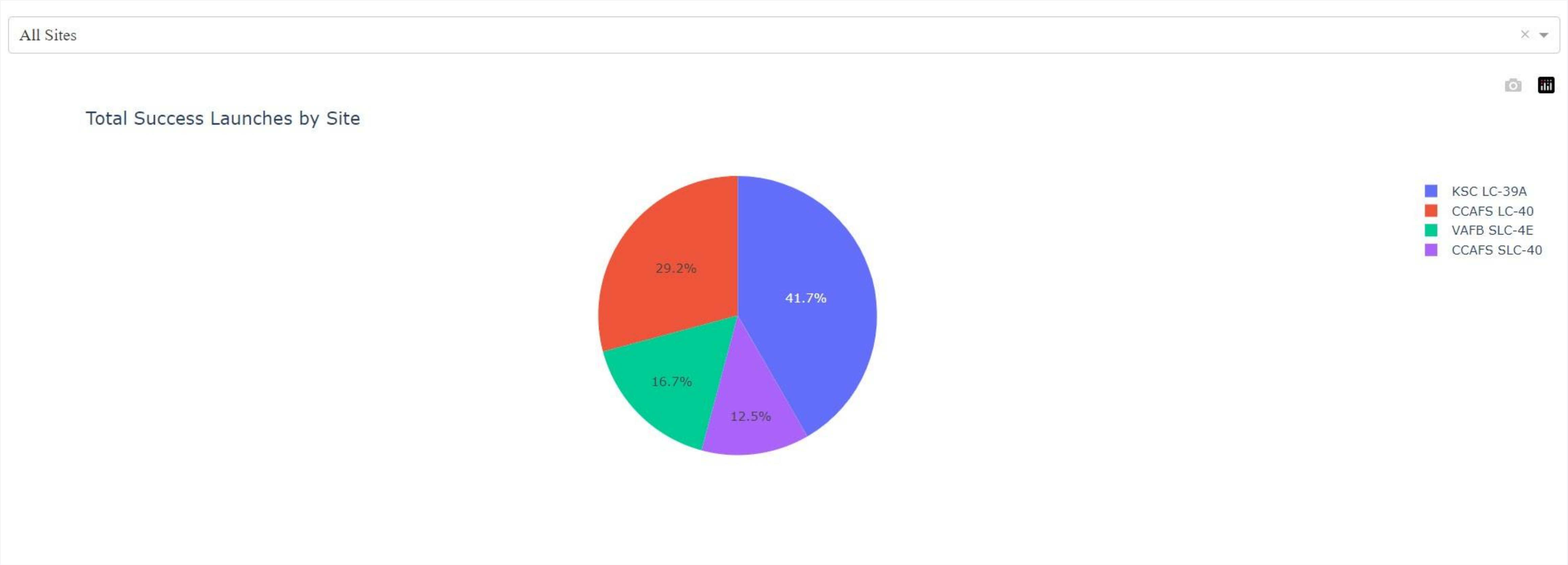
Launch Sites Proximities Analysis

Launch Sites marks on the map and the distance to equator line

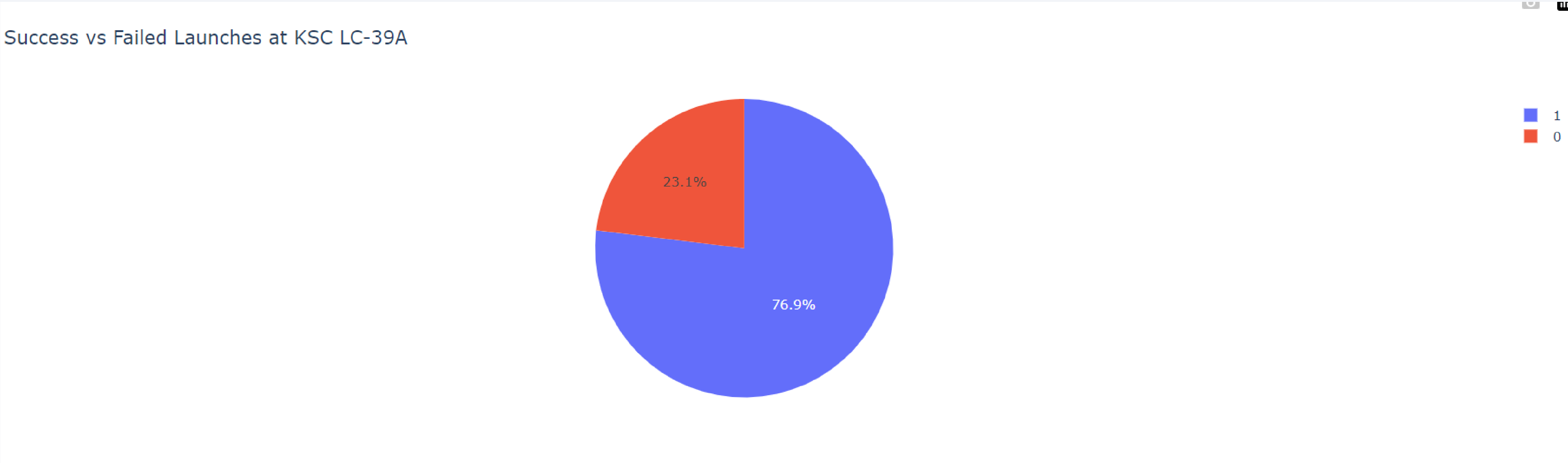
Section 4

Build a Dashboard with Plotly Dash

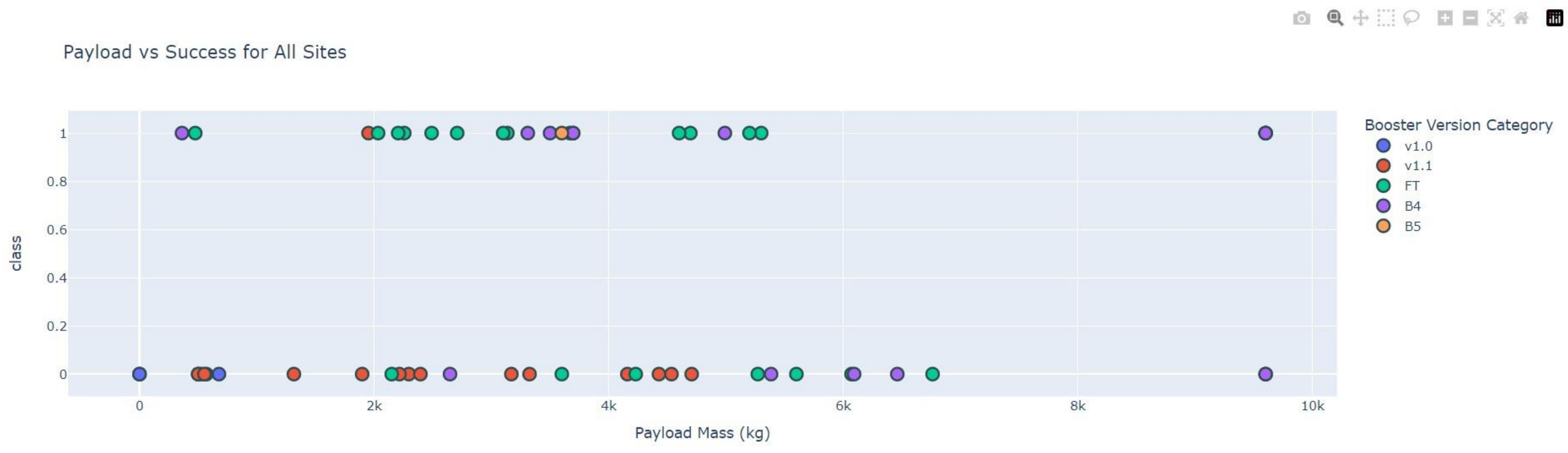
Success rate of launches by site



KSC LC-39A detailed info



Payload vs Launch outcome

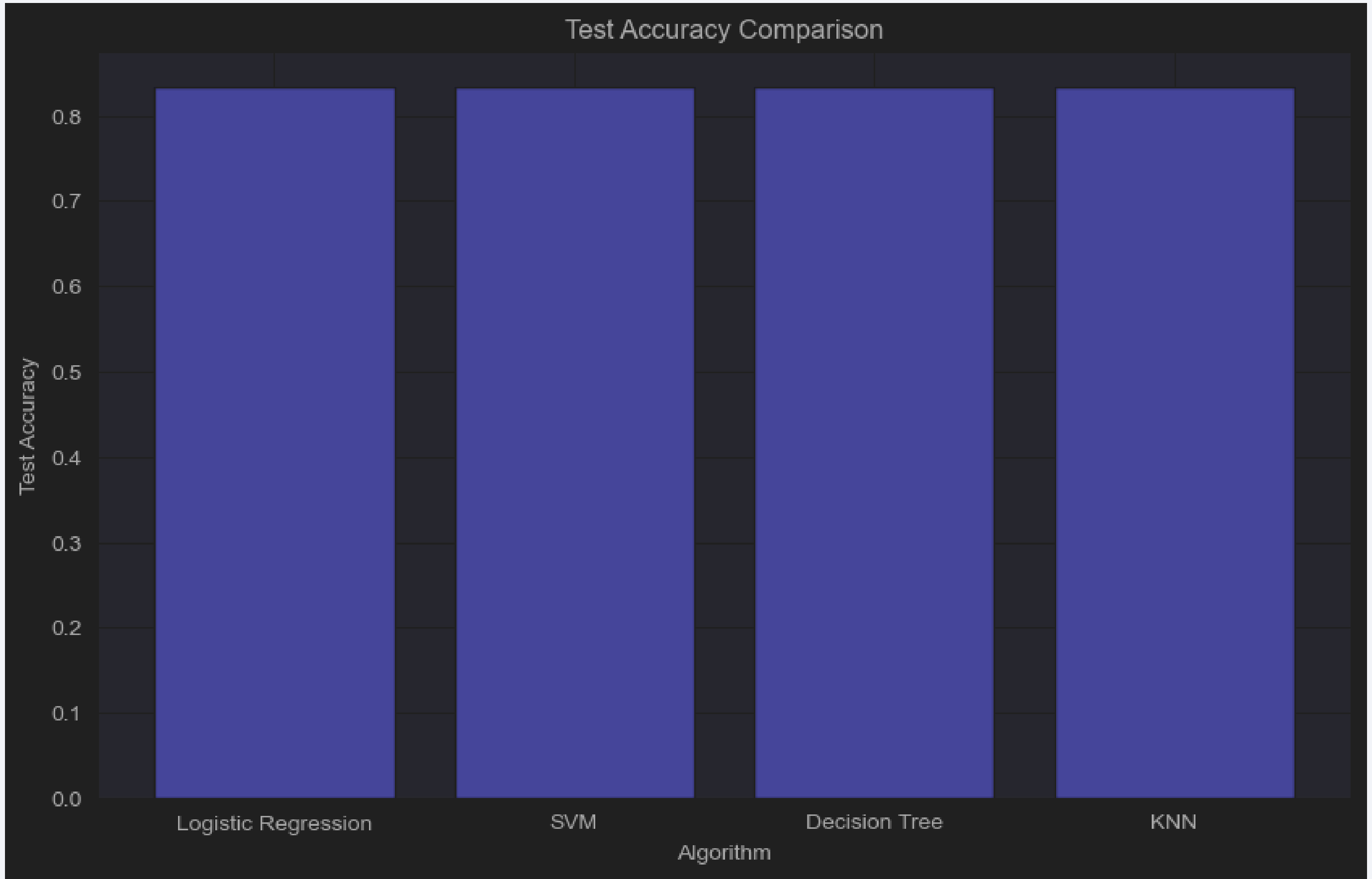


Section 5

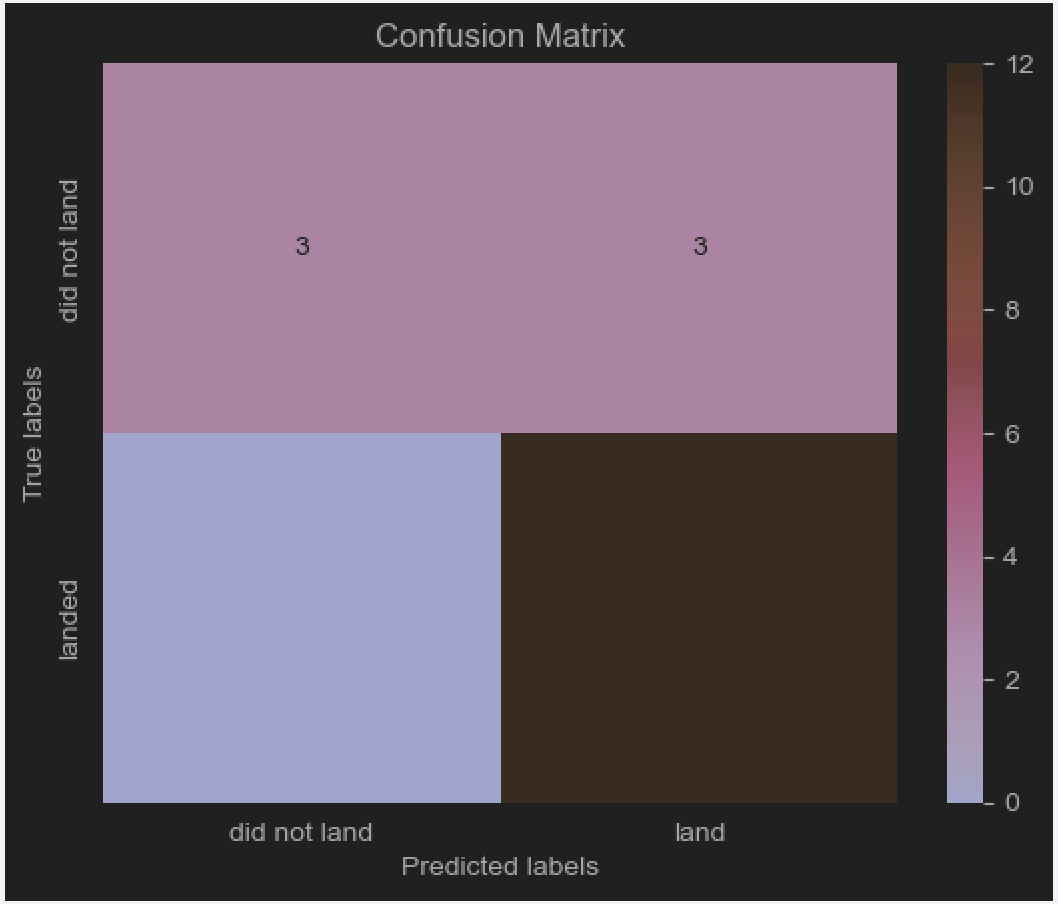
Predictive Analysis (Classification)

Classification Accuracy

	Algorithm	Precision	Recall	F1-Score	ROC AUC
0	Logistic Regression	0.800000	1.000000	0.888889	0.750000
1	SVM	0.800000	1.000000	0.888889	0.750000
2	Decision Tree	0.909091	0.833333	0.869565	0.833333
3	KNN	0.800000	1.000000	0.888889	0.750000



Confusion Matrix



Conclusions

- In this case SVM, LR, KNN models are the best

Thank you!