IBM Data Science Certification Capstone:

Cost Landing Prediction
B. Ducca

CONTENTS

Executive Summary

Introduction

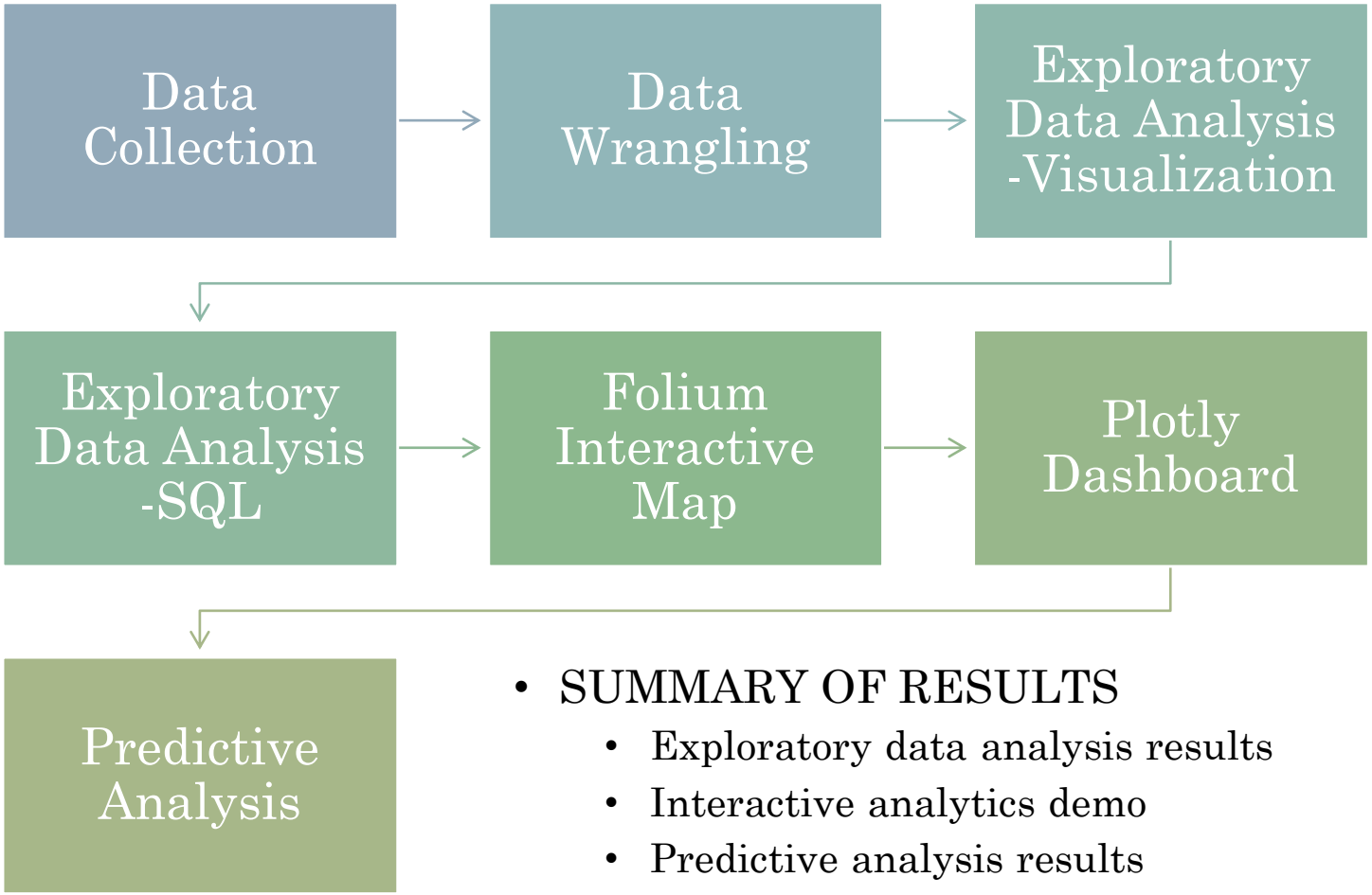
Methodology

Results

Conclusion

EXECUTIVE SUMMARY

METHODOLOGIES



- SUMMARY OF RESULTS
  - Exploratory data analysis results
  - Interactive analytics demo
  - Predictive analysis results

INTRODUCTION

PROJECT CONTEXT:

In this capstone, I will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches for a price of 62 million dollars; other providers cost upward of 165 million dollars per launch. Much of these savings are because SpaceX can reuse the first stage.

Therefore, if it can be determined if the first stage will land, I can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

PROBLEMS THAT NEED SOLVING:

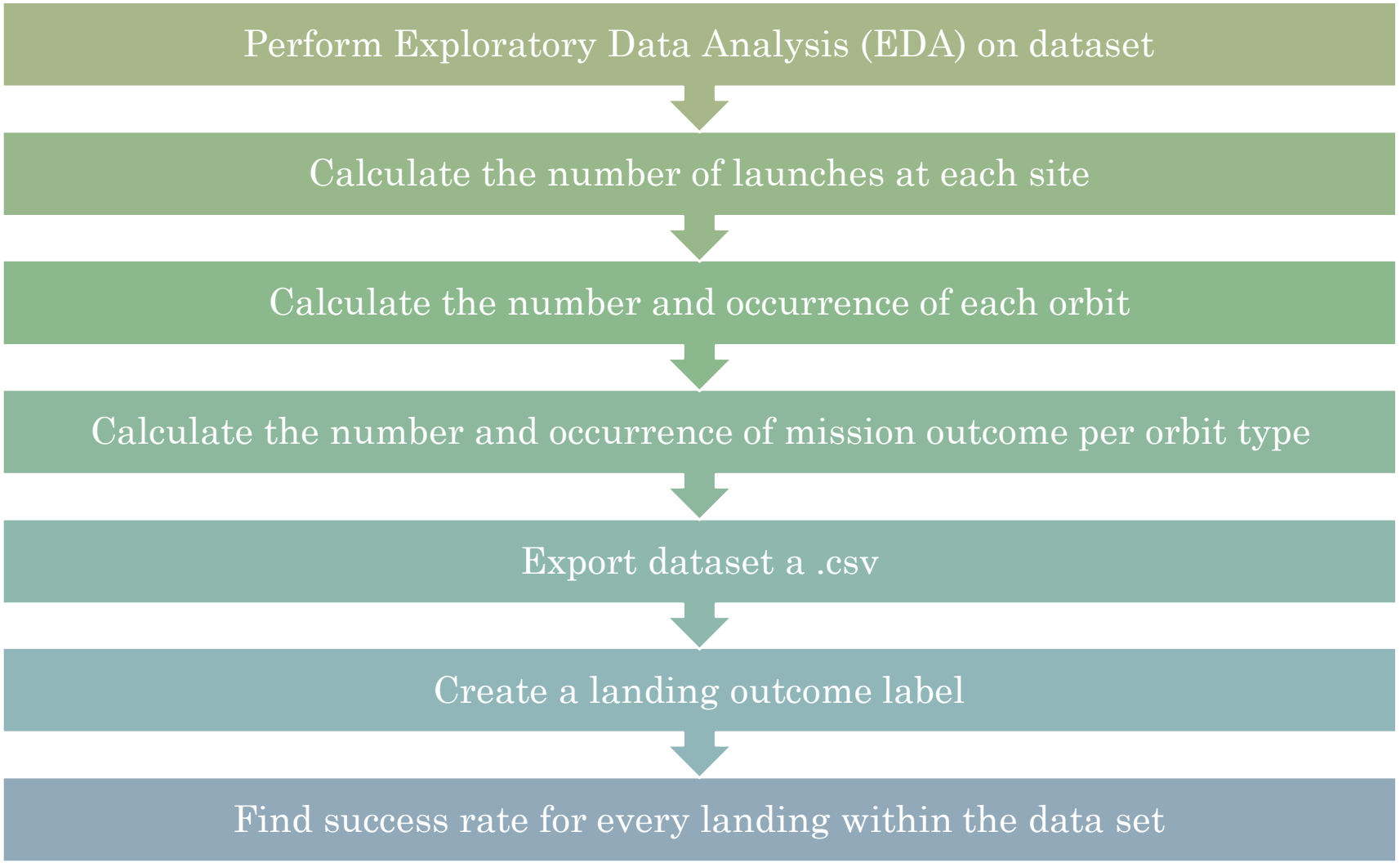
- What influences the success of the rocket landing?
- What effect does each rocket variable relationship impact the success rate?
- Which conditions get the best results and ensure the best rocket success landing rate?

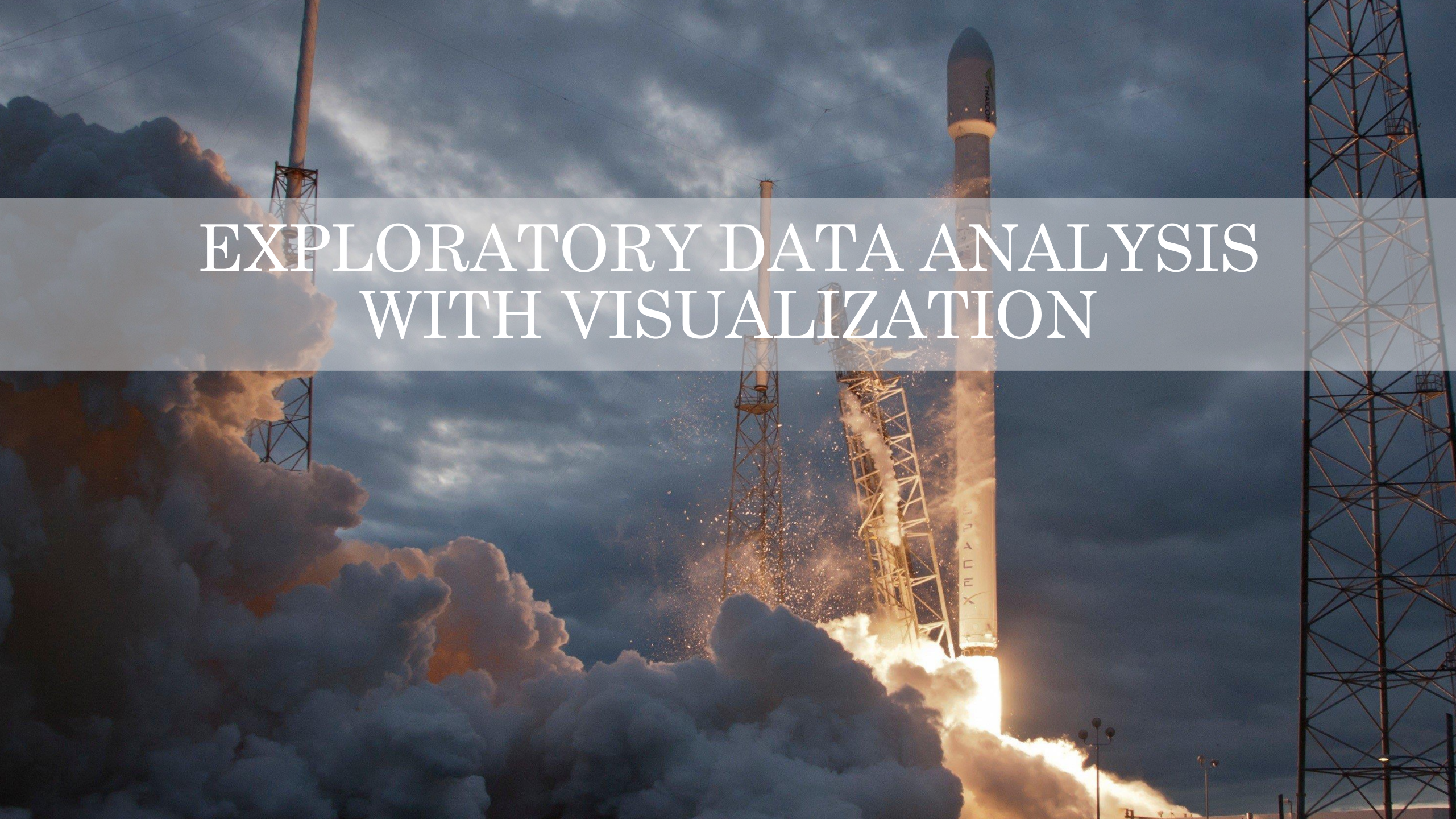
METHODOLOGY

METHODOLOGY

- Data Collection Methodology
  - SpaceX Rest API
  - Web Scraping
- Data Wrangling
  - One Hot Encoding data fields for Machine Learning and dropping irrelevant columns
- Exploratory data analysis (EDA) with visualization and SQL
  - Plotting: Scatter Graphs, Bar Graphs, to show relationships between variables and patterns of data
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - How to build, tune, evaluate classification models

DATA WRANGLING

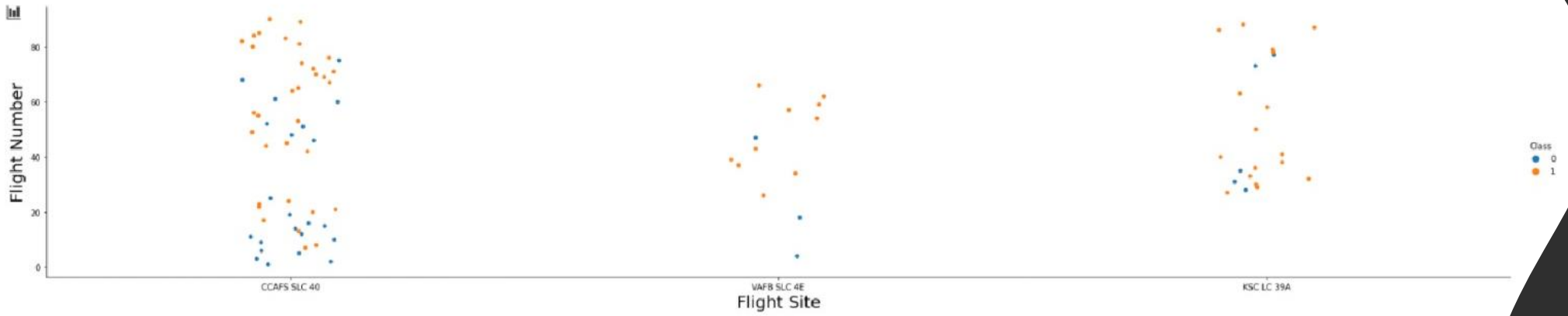




EXPLORATORY DATA ANALYSIS WITH VISUALIZATION

EDA with VISUALIZATION

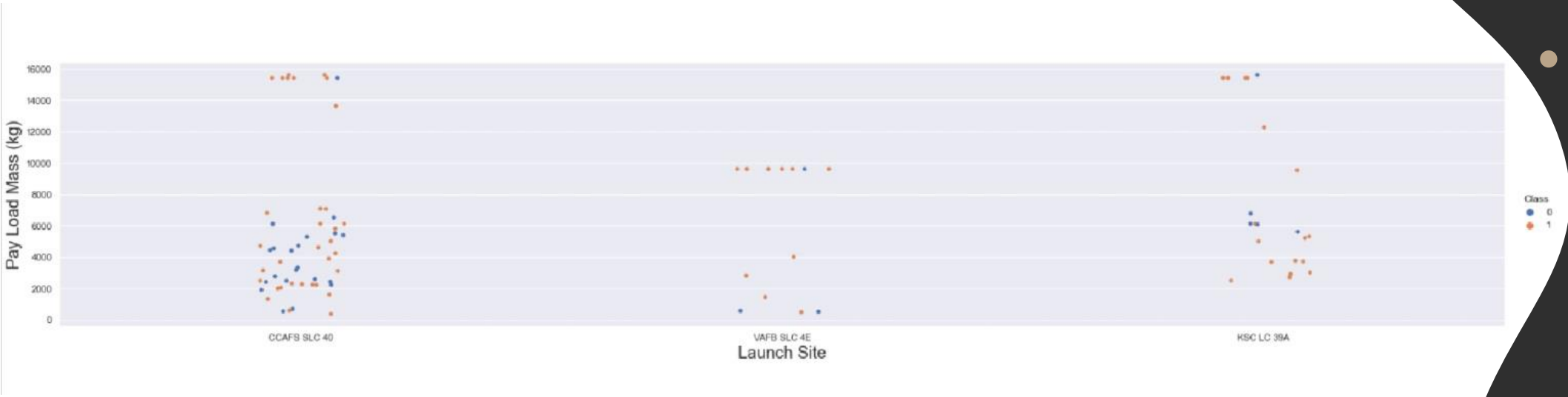
Flight Number vs Flight Site



The more flights at a launch site, the greater the success of the launch site.

EDA with VISUALIZATION

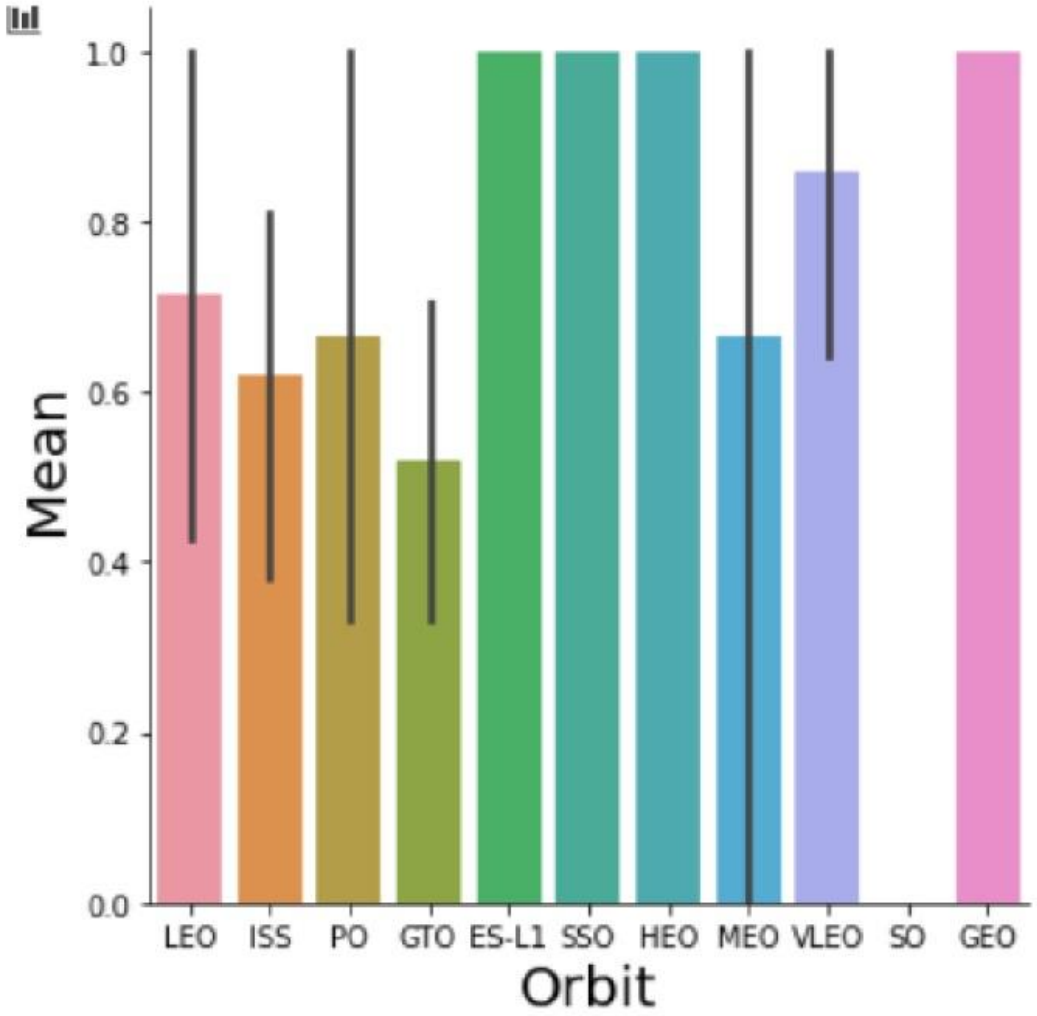
Payload Mass vs. Launch Site



Greater payload mass at launch site CCAFS SLC 40 yields a higher success rate for the rocket. Yet, No clear pattern dependent on Pay Load Mass for a successful launch based only on visualization.

EDA with VISUALIZATION

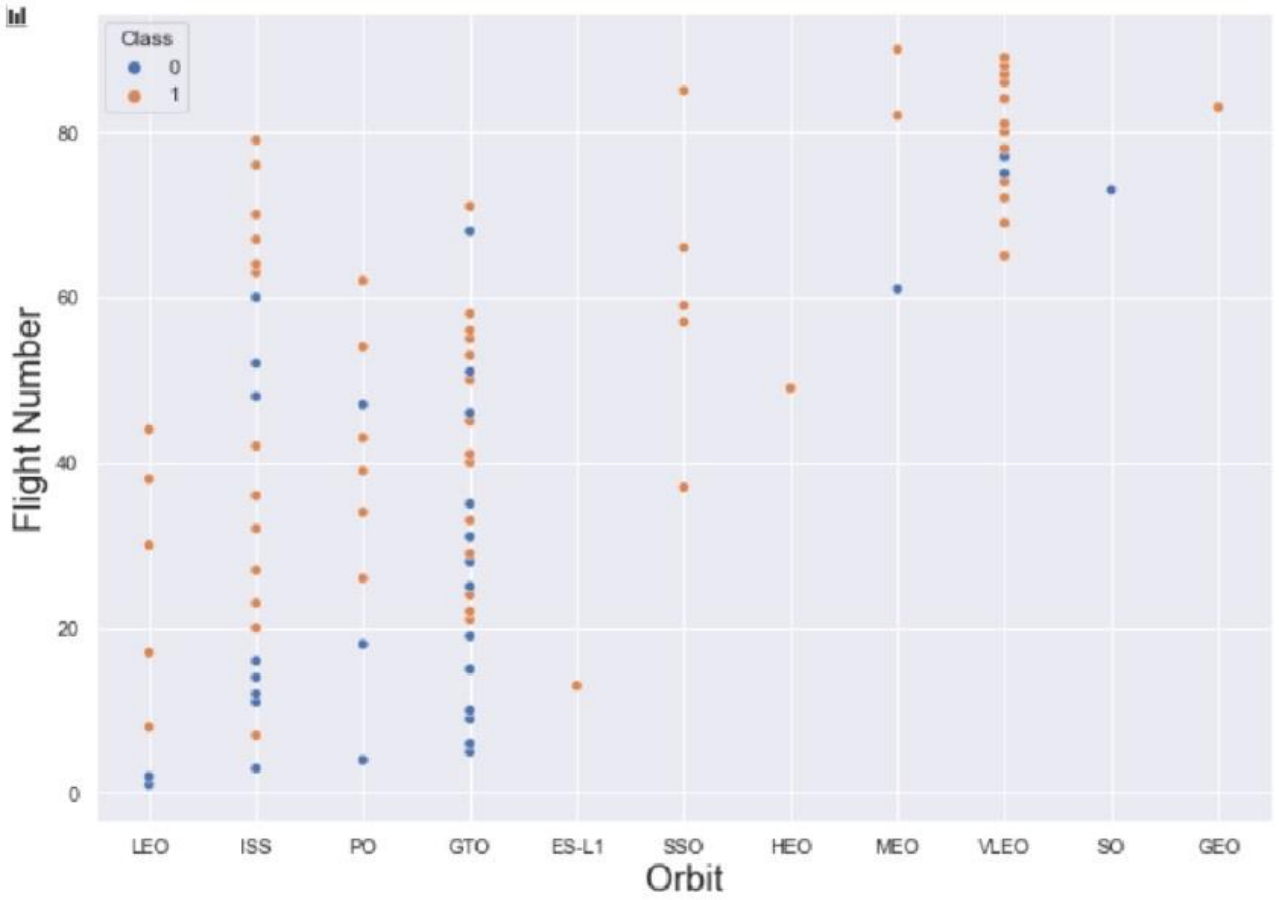
Success rate vs. Orbit Type



Orbit GEO, HEO, SSO, ES-L1 have the highest success rates

EDA with VISUALIZATION

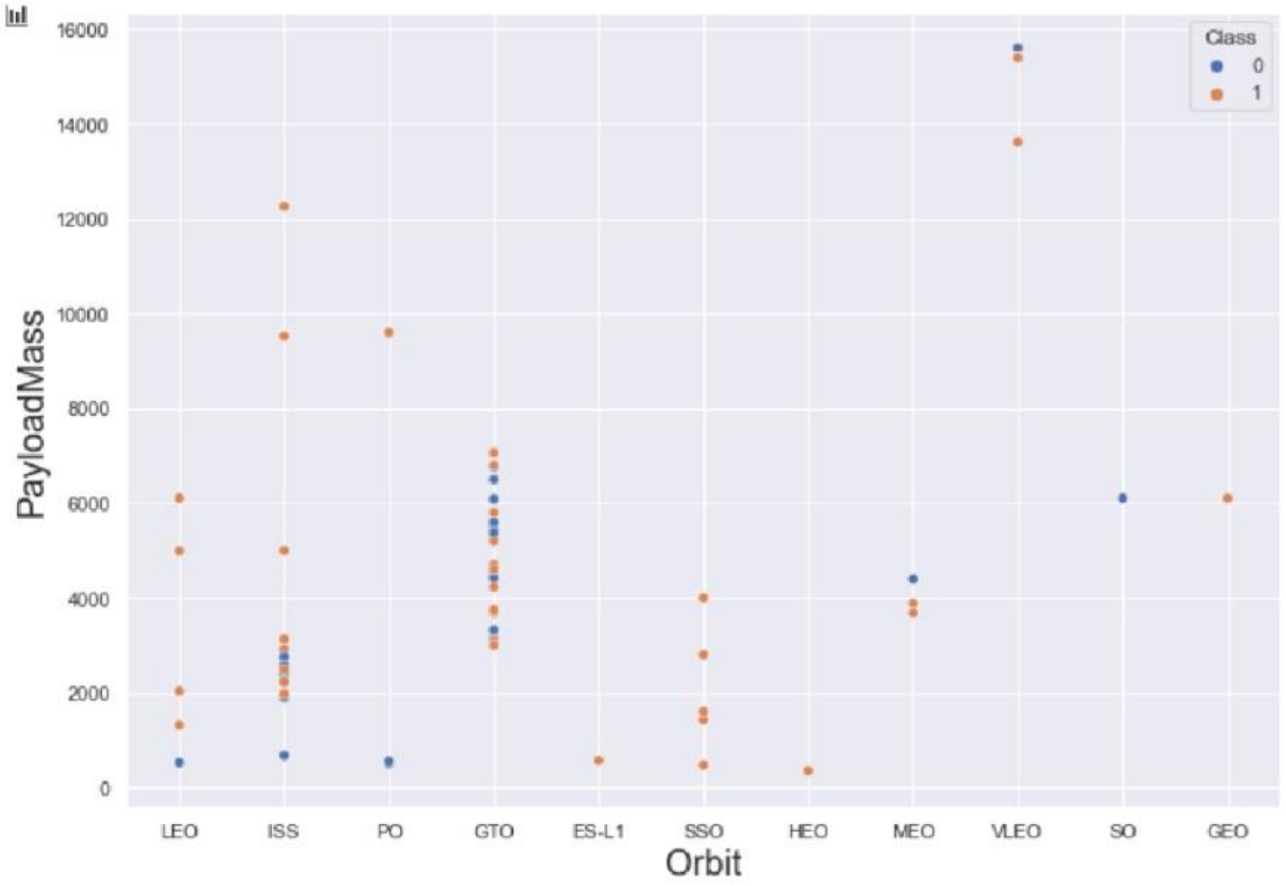
Flight Number vs Orbit Type



Success dependent on number of flights vary by orbit.

EDA with VISUALIZATION

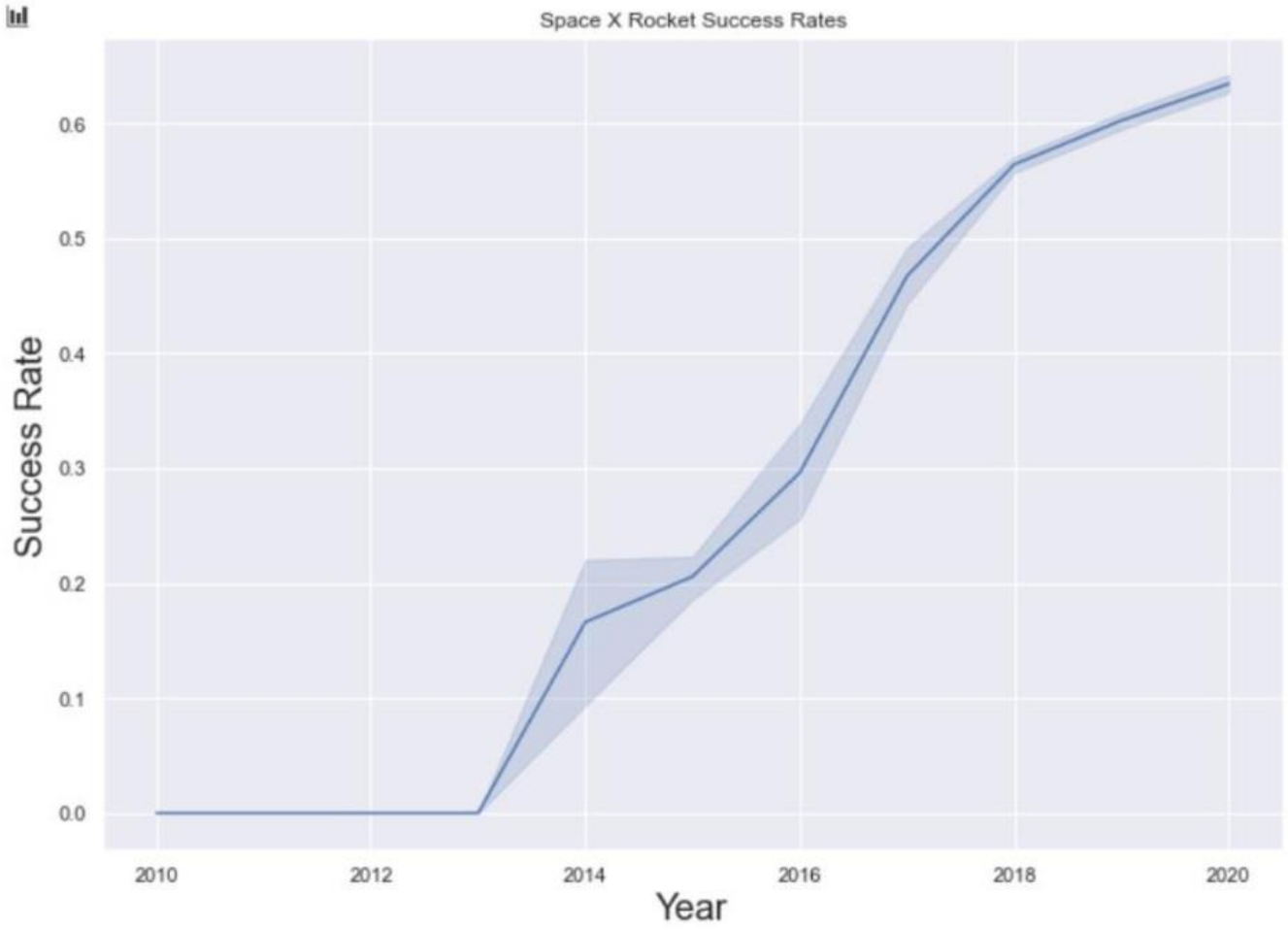
Payload vs. Orbit Type



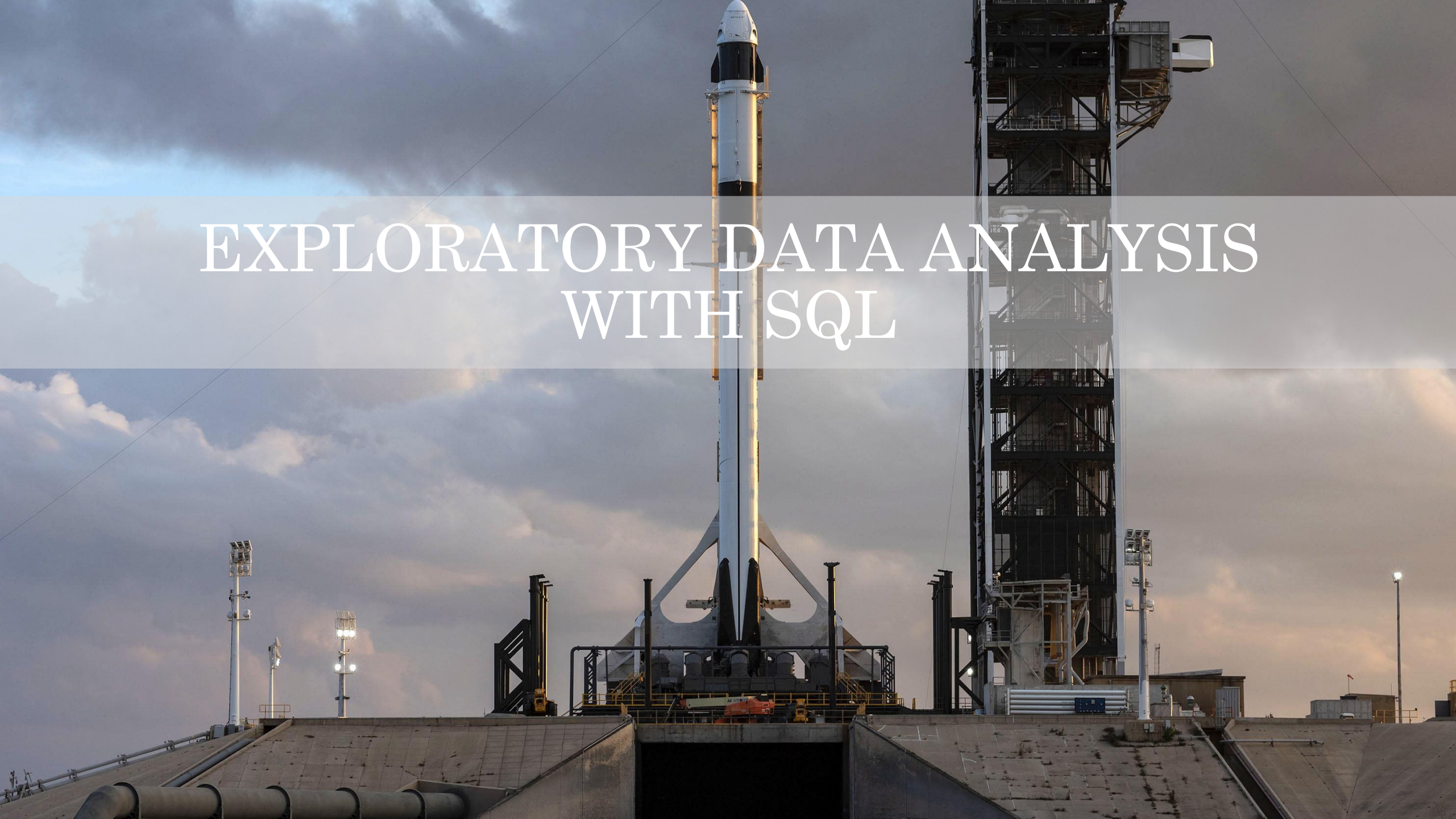
Heavy payloads have a negative influence on GTO orbits and a positive influence on GTO and Polar LEO (ISS) orbits.

EDA with VISUALIZATION

Launch Success Yearly Trend



Steady increase in successful launches since 2013



EXPLORATORY DATA ANALYSIS WITH SQL

EDA with SQL

Unique Launch Sites

Unique Launch Sites
CCAFS LC-40
CCAFS SLC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

select DISTINCT Launch_Site from tblSpaceX

Launch Site names begin with CCA

	Date	Time_UTC	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	19-02-2017	2021-07-02 14:39:00.0000000	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
1	16-03-2017	2021-07-02 06:00:00.0000000	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2	30-03-2017	2021-07-02 22:27:00.0000000	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
3	01-05-2017	2021-07-02 11:15:00.0000000	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
4	15-05-2017	2021-07-02 23:21:00.0000000	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

select TOP 5* from tblSpaceX
WHERE Launch_Site LIKE ‘KSC%’

EDA with SQL

Total Payload Mass by Customer NASA (CRS)

	Total Payload Mass
0	45596

```
select SUM(PAYLOAD_MASS_KG_) TotalPayloadMass from tblSpaceX
WHERE Customer = 'NASA (CRS)'", 'TotalPayloadMass
```

Average Payload Mass carried by booster version F9

	Average Payload Mass
0	2928

```
select AVG(PAYLOAD_MASS_KG_) AveragePayloadMass from tblSpaceX
WHERE Booster_Version= 'F9 v1.1'
```

EDA with SQL

Successful Landing Outcome in drone ship was achieved

	Date which first Successful landing outcome in drone ship was acheived.
0	06-05-2016

select MIN(Date) SLO from tblSpaceX
WHERE Landing_outcome = “Success (drone ship)”

Successful drone ship landing with payload between 4000 and 6000

	Date which first Successful landing outcome in drone ship was acheived.
0	F9 FT B1032.1
1	F9 B4 B1040.1
2	F9 B4 B1043.1

select Booster_Version from tblSpaceX WHERE Landing_Outcome = ‘Success (ground pad)’ AND Payload_MASS_KG_ > 4000 AND Payload_MASS_KG_ <6000

EDA with SQL

Booster carried maximum payload

	Booster_Version	Maximum Payload Mass
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
...	...	...
92	F9 v1.1 B1003	500
93	F9 FT B1038.1	475
94	F9 B4 B1045.1	362
95	F9 v1.0 B0003	0
96	F9 v1.0 B0004	0

97 rows × 2 columns

```
select DISTINCT Booster_Version, MAX(PAYLOAD_MASS_KG_)
AS [Maximum Payload Mass] FROM tblSpaceX GROUP BY
Booster_Version ORDER BY [Maximum Payload Mass] DESC
```

EDA with SQL

2017 Launch Records

Month	Booster_Version	Launch_Site	Landing_Outcome
January	F9 FT B1029.1	VAFB SLC-4E	Success (drone ship)
February	F9 FT B1031.1	KSC LC-39A	Success (ground pad)
March	F9 FT B1021.2	KSC LC-39A	Success (drone ship)
May	F9 FT B1032.1	KSC LC-39A	Success (ground pad)
June	F9 FT B1035.1	KSC LC-39A	Success (ground pad)
June	F9 FT B1029.2	KSC LC-39A	Success (drone ship)
June	F9 FT B1036.1	VAFB SLC-4E	Success (drone ship)
August	F9 B4 B1039.1	KSC LC-39A	Success (ground pad)
August	F9 FT B1038.1	VAFB SLC-4E	Success (drone ship)
September	F9 B4 B1040.1	KSC LC-39A	Success (ground pad)
October	F9 B4 B1041.1	VAFB SLC-4E	Success (drone ship)
October	F9 FT B1031.2	KSC LC-39A	Success (drone ship)
October	F9 B4 B1042.1	KSC LC-39A	Success (drone ship)
December	F9 FT B1035.2	CCAFS SLC-40	Success (ground pad)

```
select DATENAME(month,
DATEADD(month,MONTH(CONVERT(date, Date, 105)),
0) -1) AS Month, Booster_Version, Launch_Site, Landing
Outcome FROM tblSpaceX WHERE (Landing_Outcome
LIKE N'%Success%') AND (YEAR(CONVERT(date, Date,
105)) = '2017')
```

EDA with SQL

Total Number of Successful and Failure Mission Outcomes

	Successful_Mission_Outcomes	Failure_Mission_Outcomes
0	100	1

select(SELECT Count(Mission_Outcome) from tblSpaceX where Mission_Outcome Like '%Success%') as Successful_Mission_Outcomes,
(SELECT Count(Mission_Outcome) from tblSpaceX where Mission_Outcome LIKE '%Failure%' as Failure_Mission_Outcomes

Rank Success count between 06/04/2010-03/20/2017

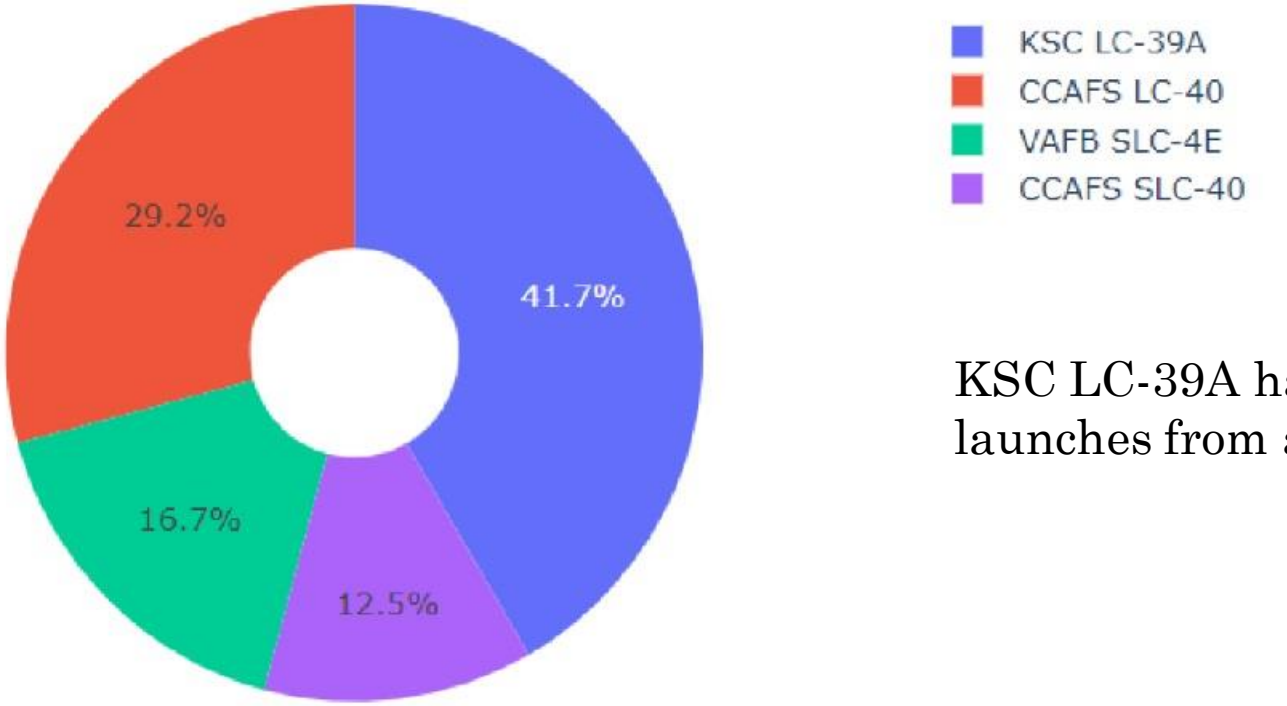
	Successful Landing Outcomes Between 2010-06-04 and 2017-03-20
0	34

select COUNT (Landing_Outcome) FROM tblSpaceX
WHERE (Landing_Outcome LIKE '%Success%') AND (Date > '06-0-2010') AND (Date < '03-20-2017')

PLOTLY DASHBOARD

Dashboard with Plotly Dash

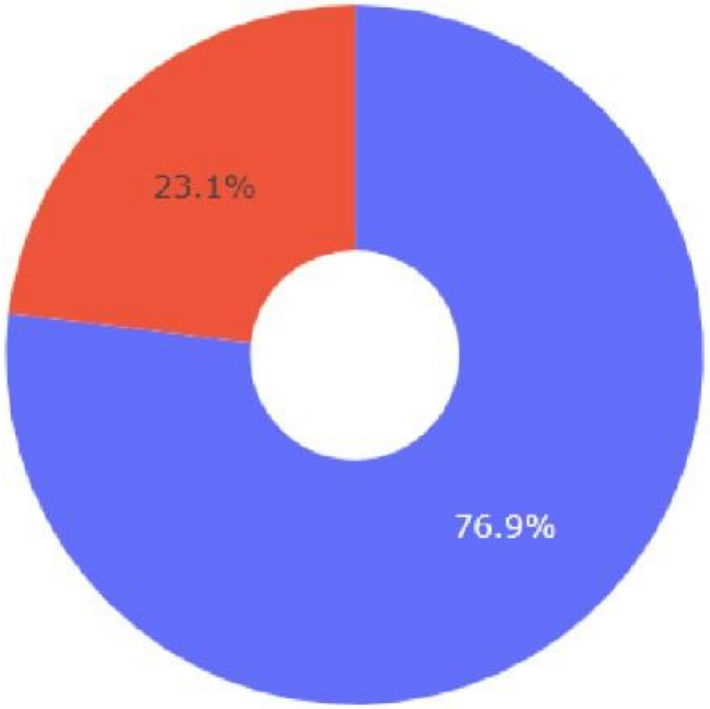
Success percentage achieved by each launch site



KSC LC-39A had the most successful launches from all sites

Dashboard with Plotly Dash

Launch site with highest launch success ratio

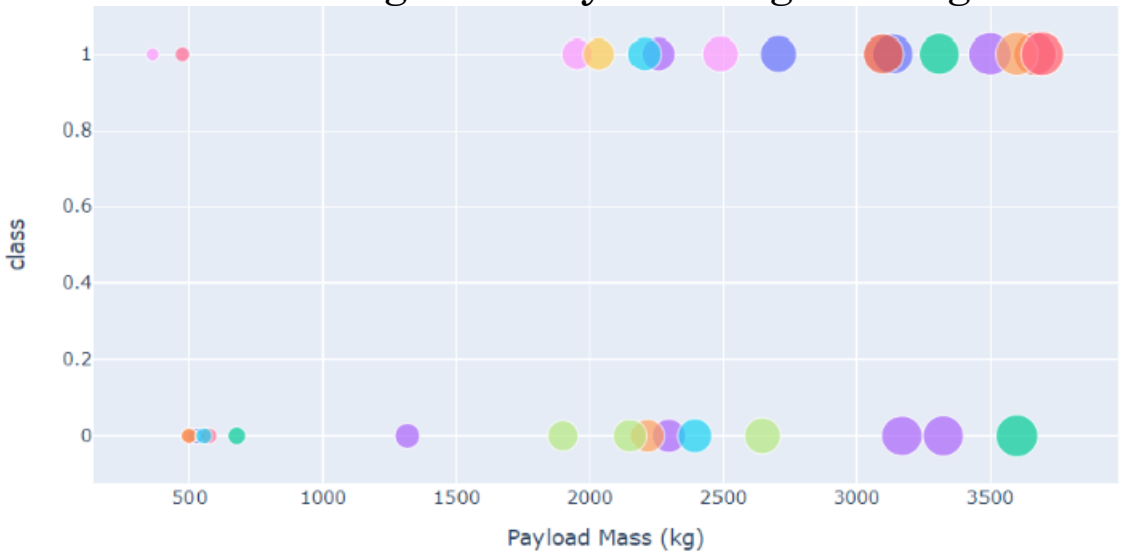


KSC LC-39A achieved a 76.9% success rate and a 23.1% failure rate

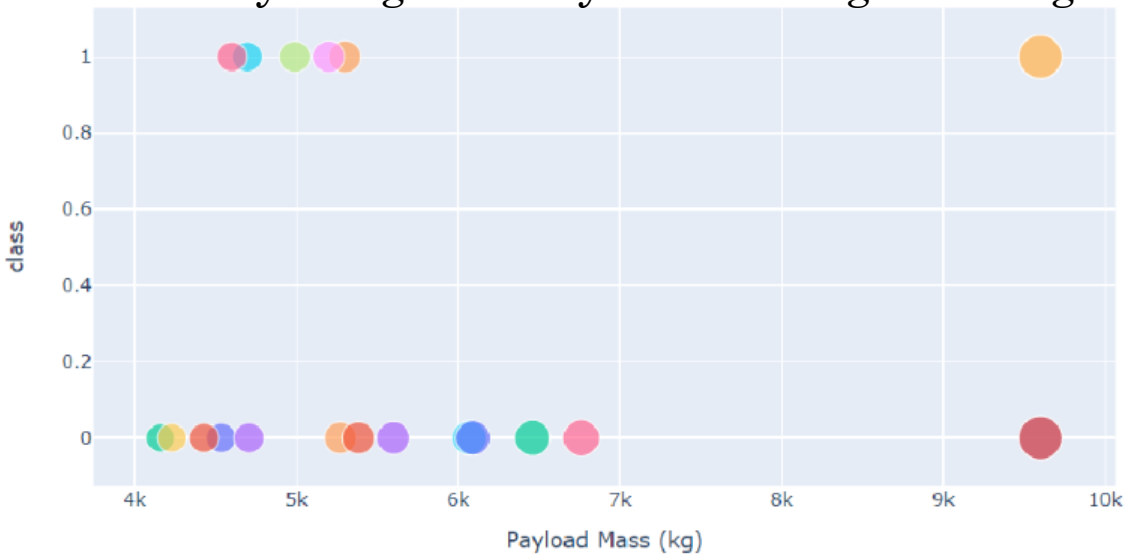
Dashboard with Plotly Dash

Payload vs Launch Outcome scatter plot for all sites

Low Weighted Payload 0kg-4000kg



Heavy Weighted Payload 4000kg-10000kg



Low weighted payloads received higher success rates when compared to weighted payloads

PREDICTIVE ANALYSIS

Classification Accuracy Using Training Data

After testing for algorithmic accuracy, it is determined that the tree algorithm has the highest accuracy by a margin.

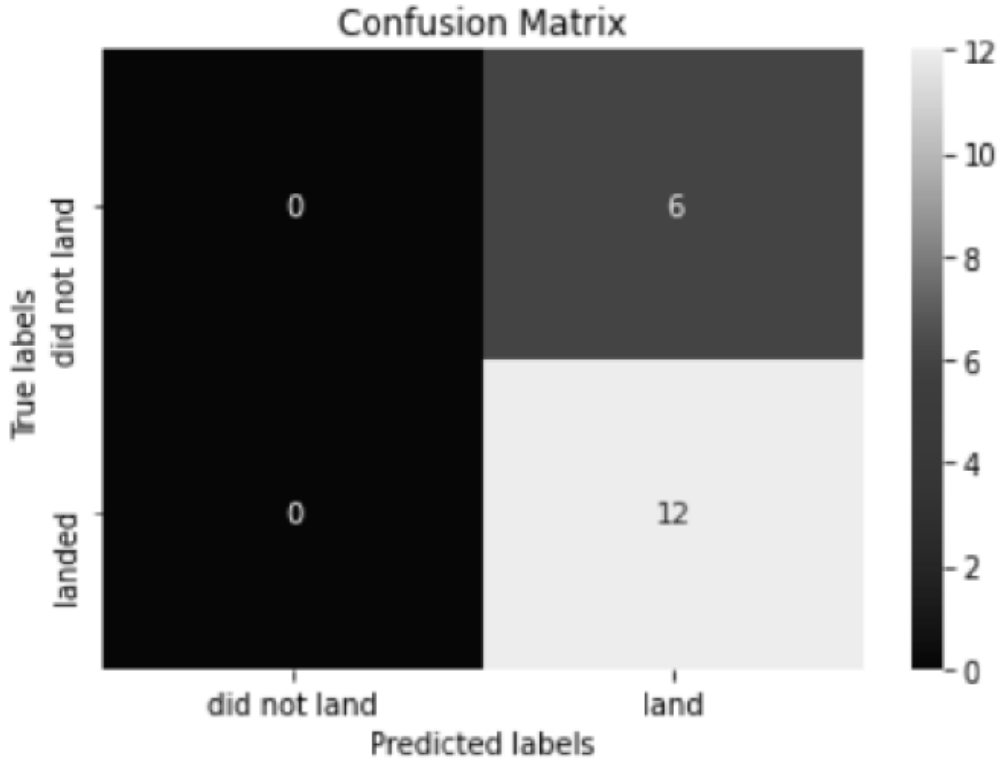
```
bestalgorithm = max(algorithms, key=algorithms.get)
```

	Accuracy	Algorithm
0	0.653571	KNN
1	0.667857	Tree
2	0.667857	LogisticRegression

Then by selecting the best hyperparameters for the decision tree classifier using the validation data, 83.33% accuracy is achieved on test data

```
Best Algorithm is Tree with a score of 0.6678571428571429
Best Params is : {'criterion': 'gini', 'max_depth': 2, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'best'}
```

Confusion Matrix for Data Tree



The Tree can distinguish between the different classes, but examination of the confusion matrix shows that false positives are an issue.

CONCLUSION

- The Tree Classifier Algorithm is best for Machine Learning for this dataset
- Low weighted payloads perform better than the heavier payloads
- The success rates for SpaceX launches is directly proportional to time in years
- KSC LC-39A had the most successful launches from all the sites
- Orbit GEO, HEO, SSO, ES-L1 has the best success rate