Winning Space Race with Data Science

Thasigan Chetty
26th June 2023

Outline

Executive Summary

Introduction

Methodology

Results

Conclusion

Appendix

Executive Summary

Summary of all results

- EDA (Exploratory Data Analysis)
- Interactive Visual analytics dashboard
- Folium Maps
- Predictive Analytics result

Summary of Methodologies

- Data Collection through API
- Data Collection with Web Scraping
- Data Wrangling
- EDA using SQL
- EDA using Pandas and Matplotlib
- Interactive Visual Analytics Dashbaord with Folium and Plotly
- Prediction Analysis (Classification)

Introduction

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of $ 62 million; other providers cost upward of $165 million each, much of the savings is because SpaceX can reuse the first stage. In order to reuse the Falcon 9 first stage rockets, the rockets need to land safely.

Problem

Determine if Falcon 9 first stage will land successfully, by using the data that has been collected,

- Identify potential factors that affect an outcome,
- Determine if these factors significantly contribute to an outcome.
- Make a prediction.

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

A SpaceX REST API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.

Perform data wrangling

One Hot Encoding

Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

Train Test Split

Data Collection

A SpaceX REST API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome. A GET request was performed,

Also fetched additional information to the data set, example booster version, site name and location, Payloads, etc.

Converted the response to Json and then converted it into a data frame using the json_normalize method.

Created Global variables and wrote data into the variables

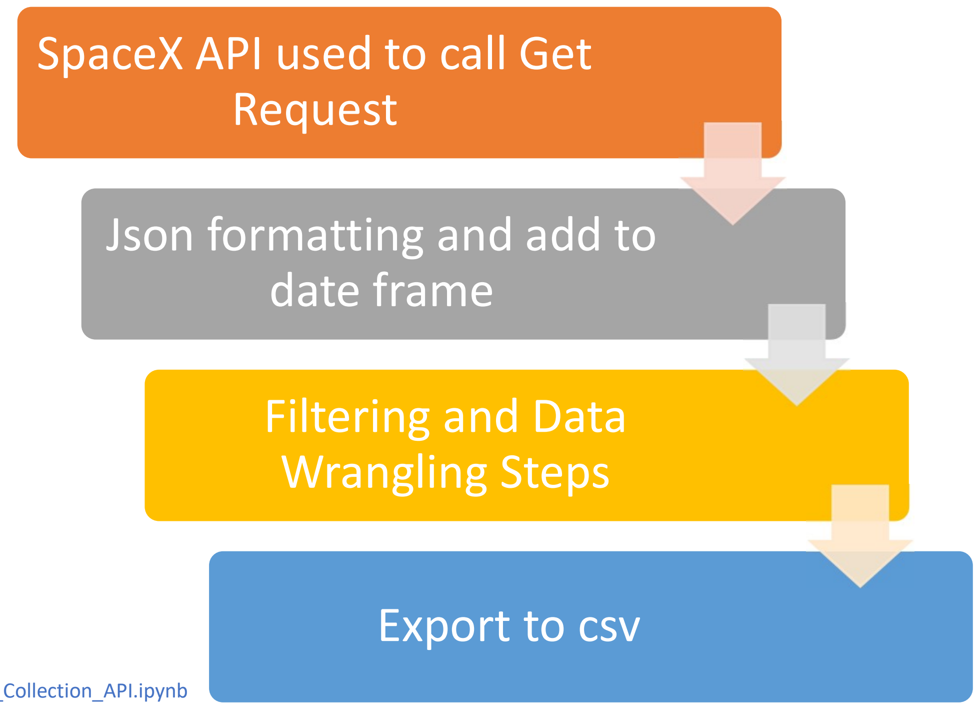
Created a dictionary called launch_dict.

Created a data frame called data_falcon9 to only include Falcon 9 launches, and then reset the flight number column.

Missing values in columns were defaulted.

Data Collection – SpaceX API

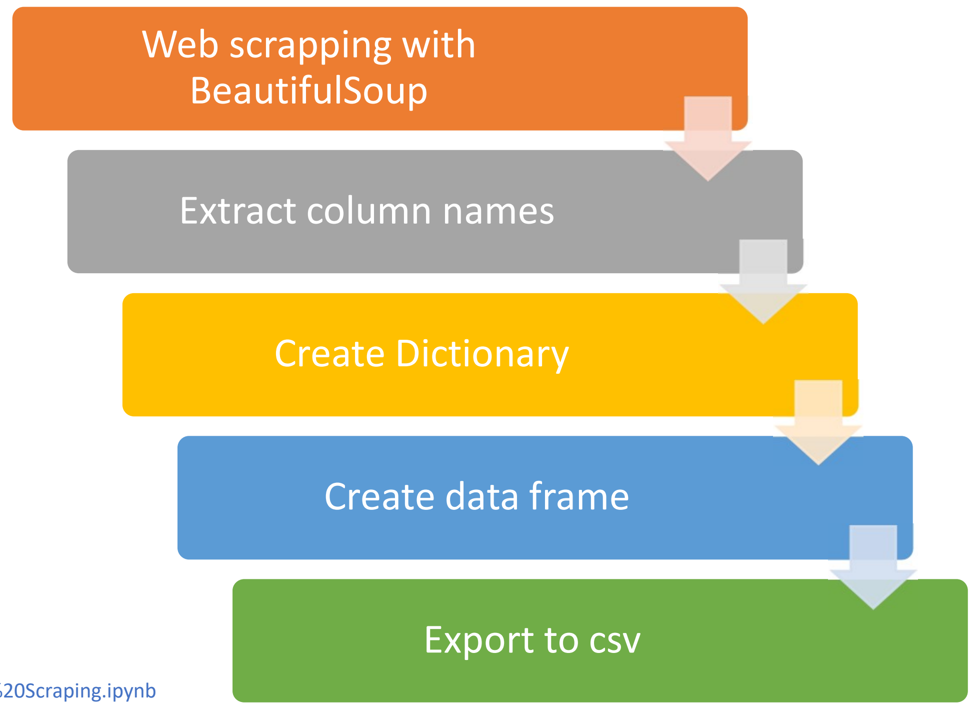
- A SpaceX REST API was call using a GET request
- Converted the response to Json and then converted it into a data frame
- Created a data frame called data_falcon9 to only include Falcon 9 launches, and performed data wrangling task (example Missing values)
- Exported to csv.



https://github.com/tessyc/SpaceX_Capstone/blob/main/Capstone_Module1_Data_Collection_API.ipynb

Data Collection – Scraping

- Web scrap Falcon 9 launch records from the wikipage with BeautifulSoup.
- iterate through the <th> elements and apply the provided extract_column_from_header() to extract column names
- Create Dictionary
- Create and populate Data frame.
- And export to csv.



https://github.com/tessyc/SpaceX_Capstone/blob/main/Capstone_Module1_Web%20Scraping.ipynb

Data Wrangling

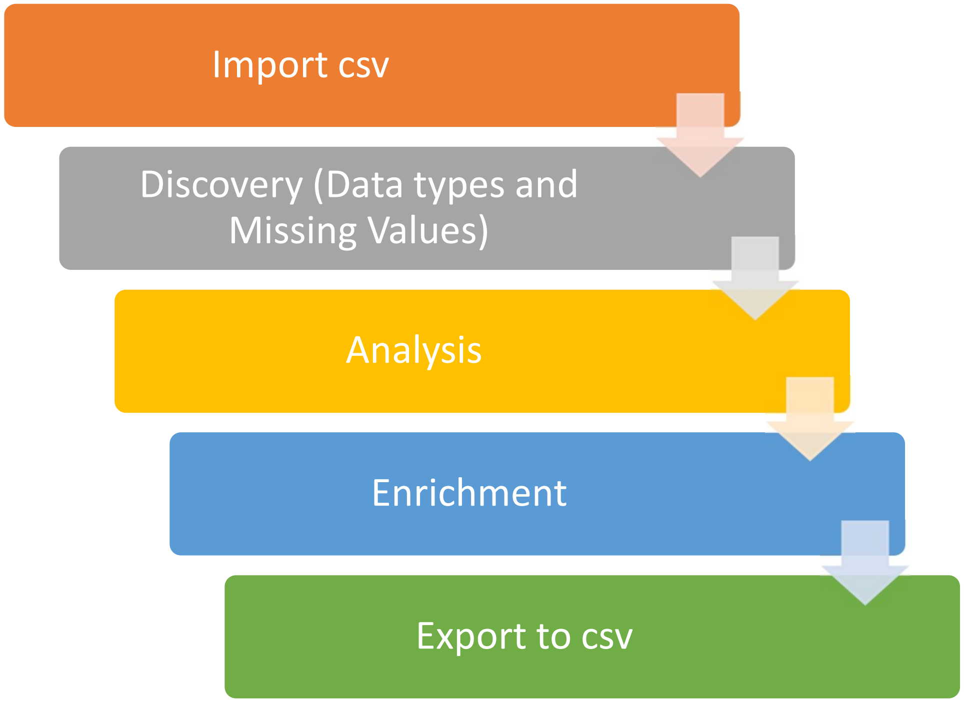
Import csv

Discovery - Data types And Missing Values were identified.

Data Analysis was performed to determine Number of Rows for attributes (Launch Sites, Orbit, Landing Outcomes)

Enrichment – added column Class

Export to csv



https://github.com/tessyc/SpaceX_Capstone/blob/main/Capstone_Module1_Data_Wrangling.ipynb

EDA with Data Visualization

- Flight Number vs Pay load Mass – Scatter plot
- Flight Number vs Launch Site – Scatter plot
- Pay load Mass vs Launch Site – Scatter plot
- Success rate of each Orbit Type – Bar Chart
- Flight Number and Orbit Type – Scatter plot
- Payload and Orbit type – Scatter plot
- Launch Success yearly trend – Line Chart

https://github.com/tessyc/SpaceX_Capstone/blob/main/Capstone_Module2_EDA_Matplotlib.ipynb

EDA with SQL

- All Launch Site Names
- Launch Site Names Begin with 'CCA'
- Total Payload Mass
- Average Payload Mass by F9 v1.1
- First Successful Ground Landing Date
- Successful Drone Ship Landing with Payload between 4000 and 6000
- Total Number of Successful and Failure Mission Outcomes
- Boosters Carried Maximum Payload
- 2015 Launch Records
- Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

https://github.com/tessyc/SpaceX_Capstone/blob/main/Capstone_Module2_EDA_SQL.ipynb

Build an Interactive Map with Folium

- Display launch sites on a map
- Display success/failed launches for each site on the map
- Calculate the distances between a launch site to the Coast
- Distance from Launch Site to the railway
- Distance from Launch Site to the highway
- Distance from Launch Site to the city

https://github.com/tessyc/SpaceX_Capstone/blob/main/Capstone_Module3_Folium.ipynb

Build a Dashboard with Plotly Dash

Launch Site Drop-down Input Component with a Default value of All Sites

Callback function to render success-pie-chart based on selected site dropdown

Range Slider to Select Payload

Callback function to render the success-payload-scatter-chart scatter plot

https://github.com/tessyc/SpaceX_Capstone/blob/main/Capstone_Module3_Plotly.py

Predictive Analysis (Classification)

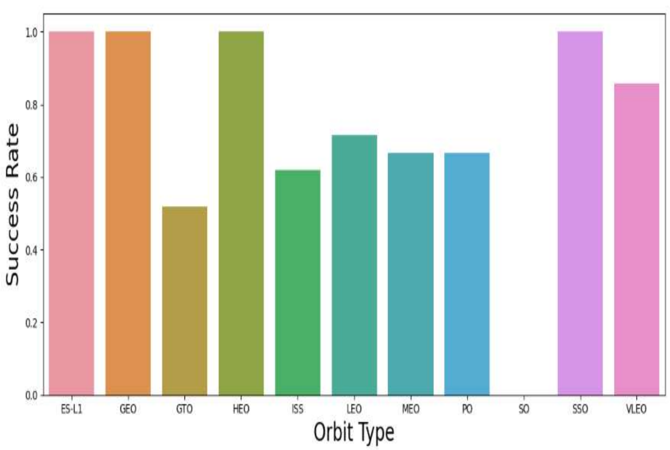
- **Data Imports**
- **Data Wrangling**
- **Perform predictive analysis using classification models (Train Test Split)**
- **Model Evaluation**
- **Model Selection**

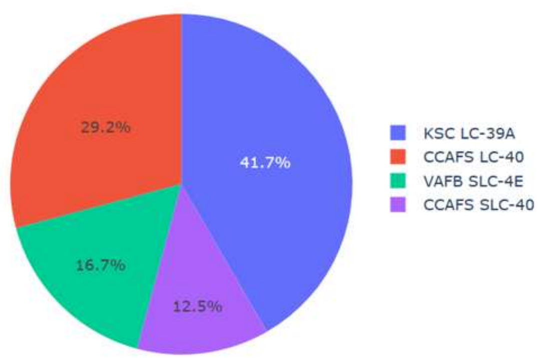
- Logistic regression
- SVM
- Decision tree
- KNN

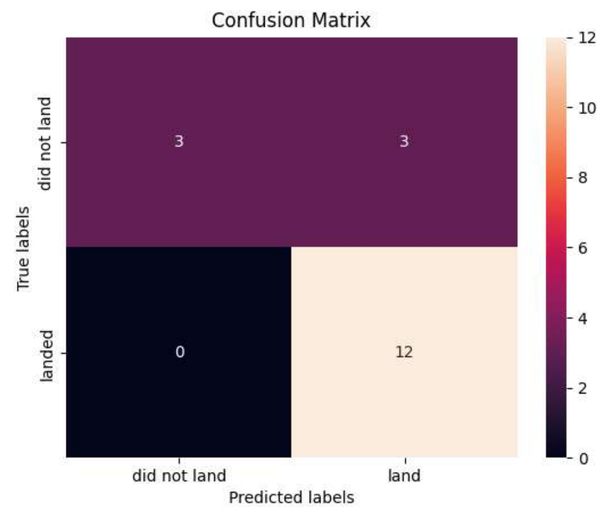
https://github.com/tessyc/SpaceX_Capstone/blob/main/Capstone_Module4_Predictive_Analysis.ipynb

Results

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





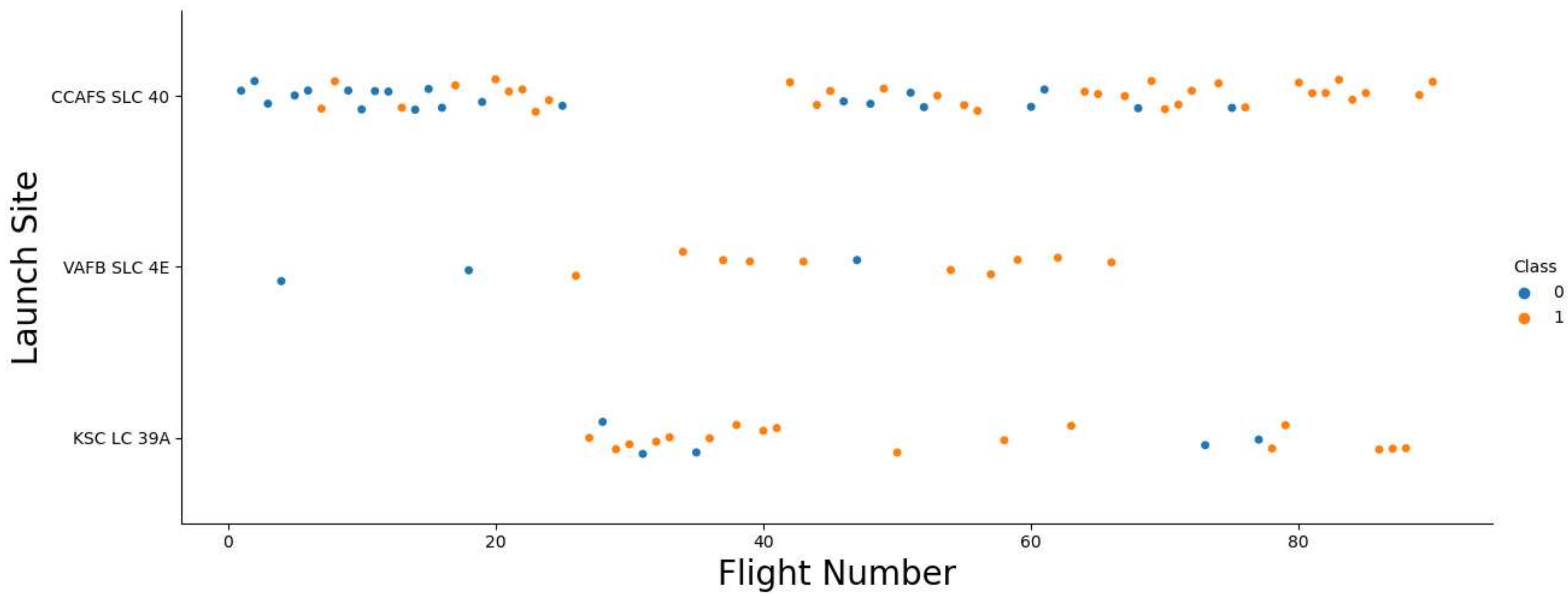


Section 2

Insights drawn from EDA

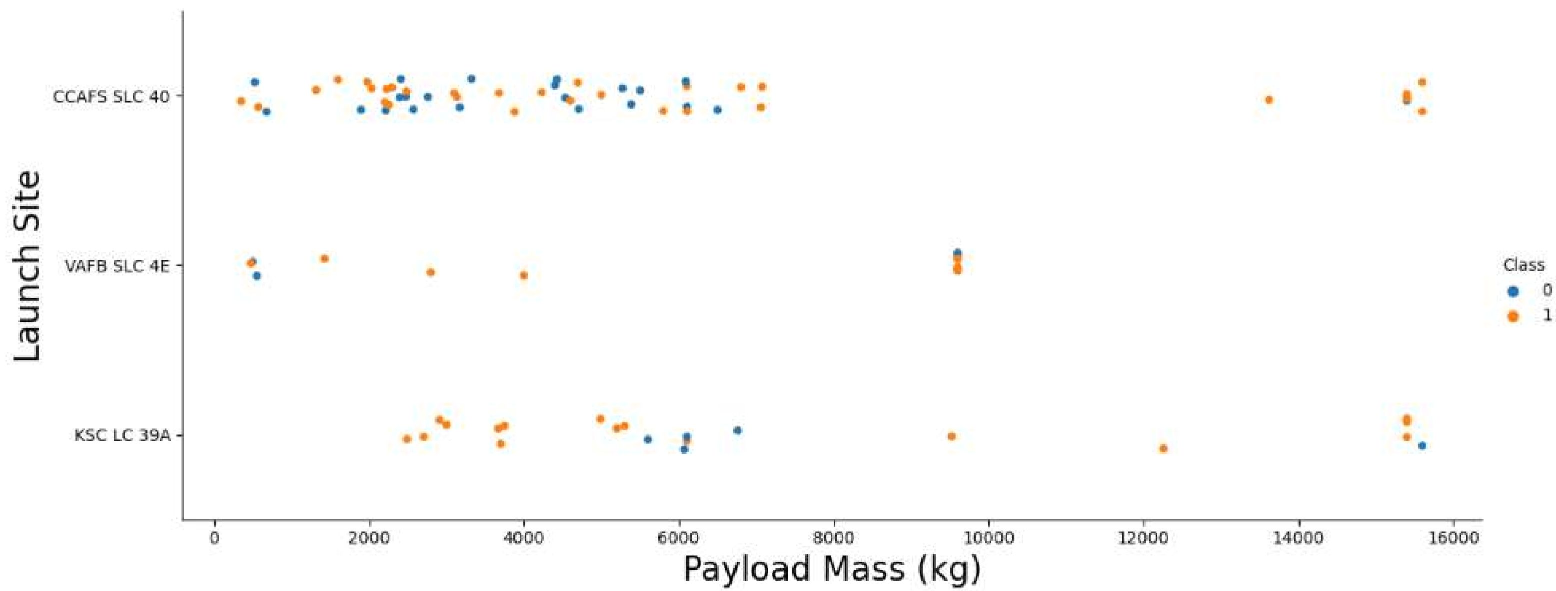
Flight Number vs. Launch Site

- The number of successful outcomes increased as more launches occurred at the site. This was especially relevant for site CCAFS SLC 40,
- If you look at CCAFS SLC 40 there are several **failed (blue)** outcomes earlier in the graph (left hand side) while more consistent **successful (orange)** outcomes later (right hand side of the Graph).



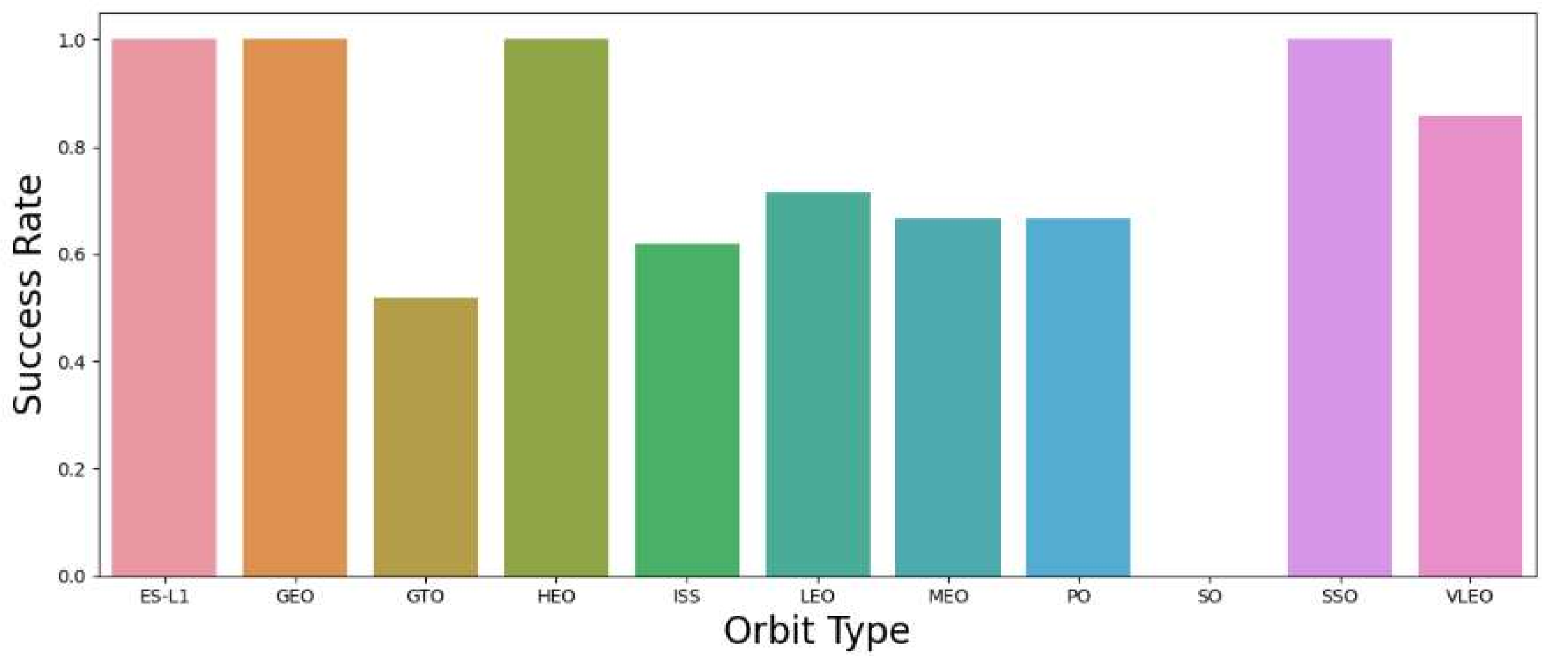
Payload vs. Launch Site

- Payload Size does not play a part in the outcome of a launch. Successful launches are evenly spread out and there is no clustering of unsuccessful launches.
- There are no launches from site VAFB SLC 4E of significant payloads, i.e is greater than 10,000 kg.



Success Rate vs. Orbit Type

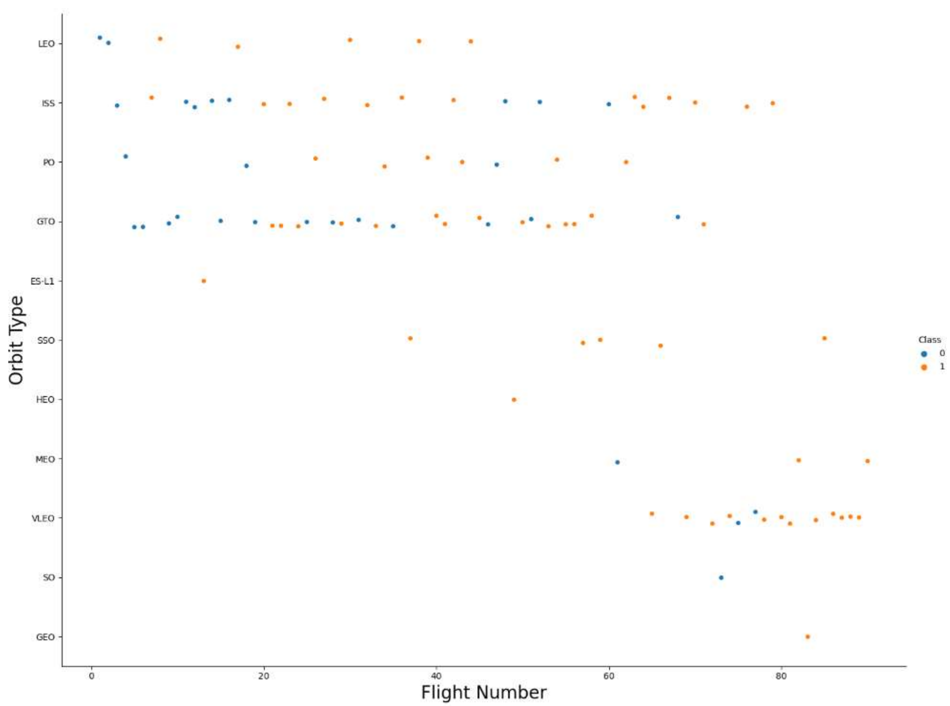
- The success rate is the mean of the class for that Orbit type.
- A success rate of 1 indicates that all launches for that Orbit type were successful while a success rate of zero means that there were no successful launches.
- Orbit type SO had zero successful launches.

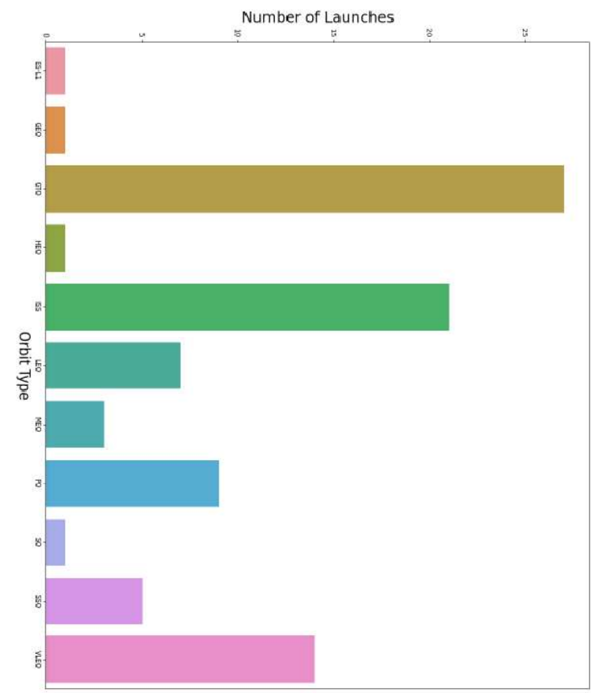


	Orbit	Class
0	ES-L1	1.000000
1	GEO	1.000000
2	GTO	0.518519
3	HEO	1.000000
4	ISS	0.619048
5	LEO	0.714286
6	MEO	0.666667
7	PO	0.666667
8	SO	0.000000
9	SSO	1.000000
10	VLEO	0.857143

Flight Number vs. Orbit Type

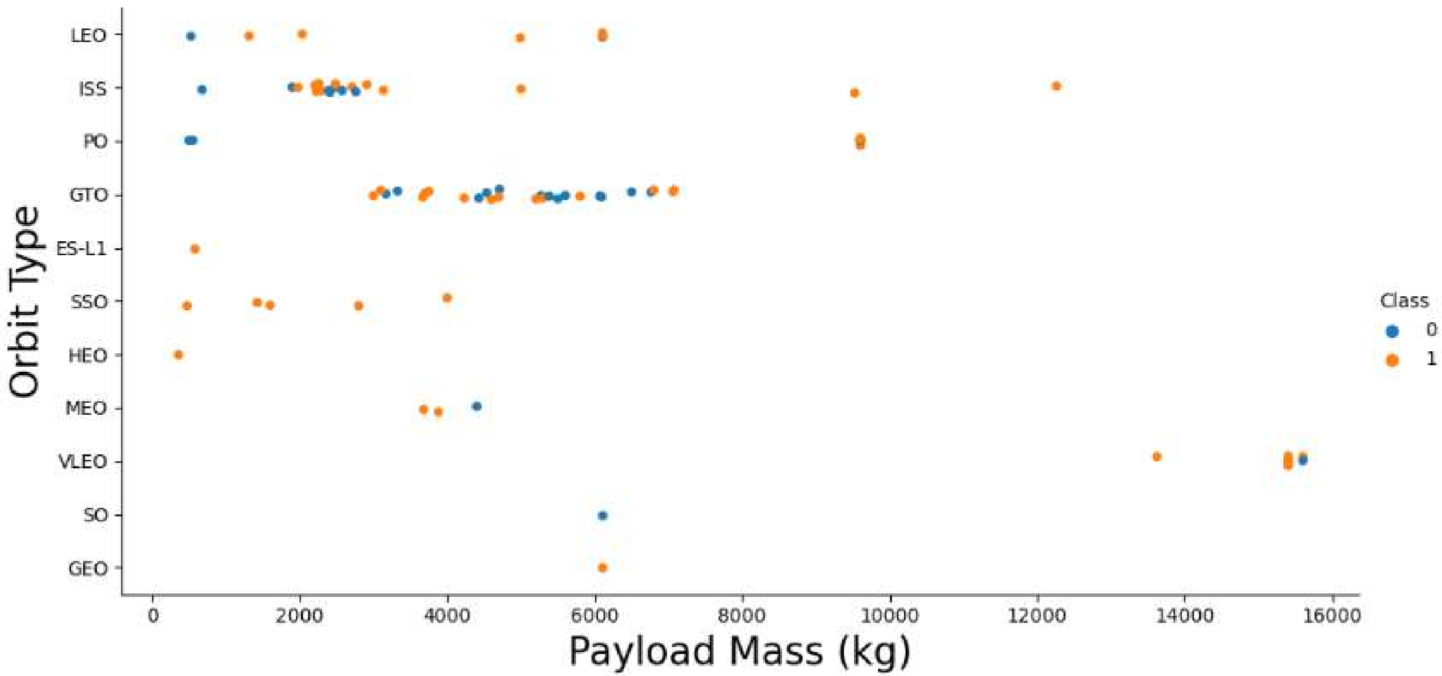
When SpaceX first started launching, only 4 orbit types were used. Later on additional orbit types were used and fazed out. VLEO is now the most commonly used orbit type, with all launches now using VLEO.





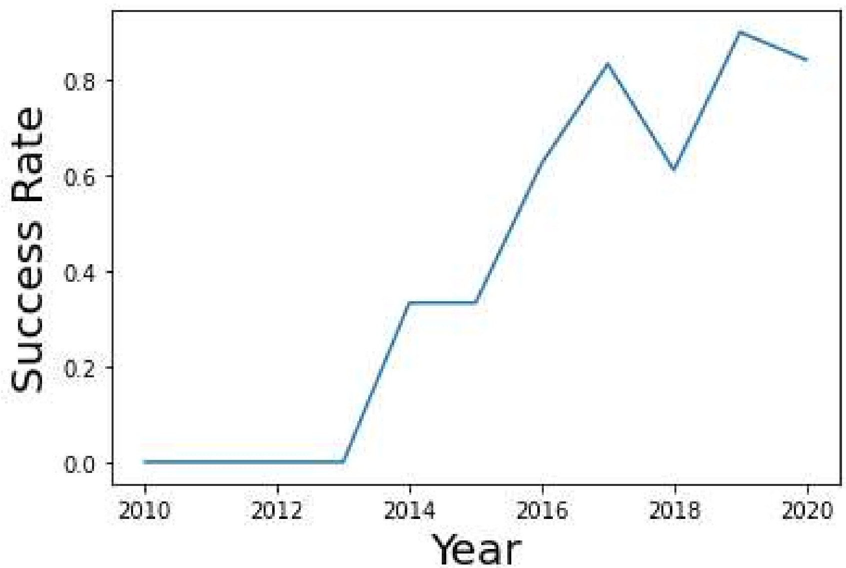
Payload vs. Orbit Type

When SpaceX first started launching, only 4 orbit types were used. Later on, additional orbit types were used and fazed out. VLEO is now the most commonly used orbit type, with all launches now using VLEO. VLEO offer SpaceX the capability to launch massive payload into space successfully, i.e., payload with masses of 12000 kg or more.



Launch Success Yearly Trend

SpaceX has seen a dramatic and sustained improvement in the success rate of launches as new orbit types, sites and boosters are used since 2014



All Launch Site Names

Unique List of launch sites

```
%sql SELECT Distinct LAUNCH_SITE FROM SPACEXTBL
```

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

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

Launch Site Names Begin with 'CCA'

Details of 5 launches from a launch site that starts with CCA.

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

The total payload mass in kilograms carried by boosters from NASA

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as 'Total Payload Mass (KG)' FROM SPACEXTBL WHERE CUSTOMER='NASA (CRS)'
```

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

Total Payload Mass (KG)
45596.0

Average Payload Mass by F9 v1.1

The average payload mass in Kilograms carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as 'Average Payload Mass (KG)' FROM SPACEXTBL WHERE BOOSTER_VERSION='F9 v1.1'
```

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

Average Payload Mass (KG)
2928.4

First Successful Ground Landing Date

Date of First successful landing outcome on ground pad

```
%sql SELECT min(DATE) as 'Earliest Landing Date' FROM SPACEXTBL WHERE LANDING_OUTCOME='Success (ground pad)'
```

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

Earliest Landing Date
01/08/2018

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ between 4000 and 6000 AND LANDING_OUTCOME='Success (drone ship)'
```

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

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

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes.

```
[31]: %sql SELECT CASE WHEN MISSION_OUTCOME LIKE  '%Success%' THEN 'Success' \
ELSE 'Failure' END AS 'Mission Outcome' \
,COUNT(*)AS 'Number Of Outcomes' FROM SPACEXTBL \
WHERE MISSION_OUTCOME LIKE '%Success%' OR MISSION_OUTCOME LIKE '%Failure%'\
GROUP BY CASE WHEN MISSION_OUTCOME LIKE  '%Success%' THEN 'Success' \
ELSE 'Failure' END
```

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

[31]:

Mission Outcome	Number Of Outcomes
Failure	1
Success	100

Boosters Carried Maximum Payload

Names of the booster which have carried the maximum payload mass.

```
%sql SELECT BOOSTER_VERSION AS 'Booster Versions' FROM SPACEXTBL WHERE \
PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

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

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

2015 Launch Records

Failed landing_outcomes in drone ship, their booster versions, and launch site names by Month in year 2015.

```
%sql SELECT SUBSTR(Date, 4, 2) as Month, \
            LANDING_OUTCOME AS 'Landing Outcome', \
            BOOSTER_VERSION AS 'Booster Version', \
            LAUNCH_SITE AS 'Launch Site' \
    FROM SPACEXTBL \
    WHERE \
        LANDING_OUTCOME = 'Failure (drone ship)' \
    AND SUBSTR(Date,7,4)='2015'
```

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

Month	Landing Outcome	Booster Version	Launch Site
10	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 of landing outcomes between the date 2010-06-04 and 2017-03-20, based on a count in descending order.

```
%sql select Rank() OVER(ORDER BY Outcomes DESC) AS Ranking, \
LANDING_OUTCOME AS 'Landing Outcomes', Outcomes AS 'Number Of Outcomes' \
from (SELECT LANDING_OUTCOME, COUNT(*) as Outcomes FROM SPACEXTBL \
    WHERE DATE BETWEEN '04/06/2010' and '20/03/2017' \
    GROUP BY LANDING_OUTCOME )\
    ORDER BY Outcomes DESC
```

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

[51]:

Ranking	Landing Outcomes	Number Of Outcomes
1	Success	20
2	No attempt	9
3	Success (drone ship)	8
4	Success (ground pad)	7
5	Failure	3
5	Failure (drone ship)	3
7	Controlled (ocean)	2
7	Failure (parachute)	2
9	No attempt	1

Section 3

Launch Sites Proximities Analysis

Map Of Global Launch Sites

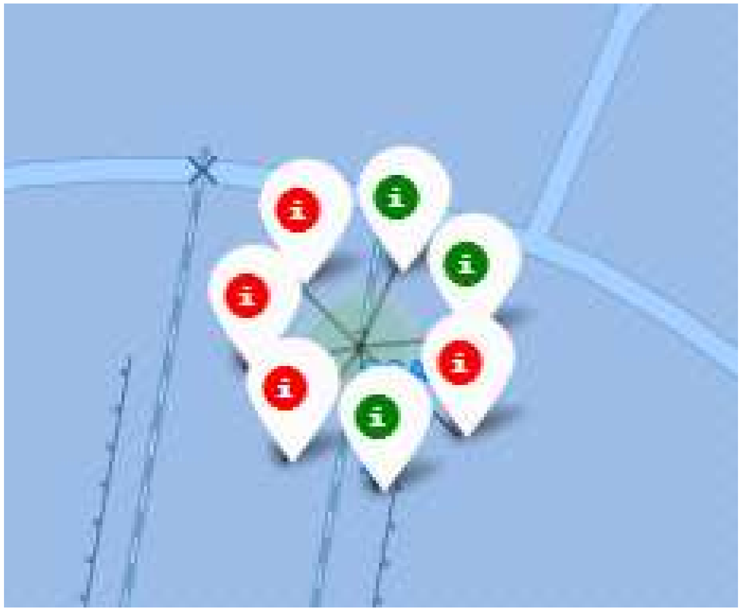
All launch sites are between Latitudes 28.56 and 34.56, north of the equator and between longitudes -80.5 and -120.61, west of the meridian.

Launching outcomes of Site CCAFS SLC-40

- Site CCAFS SLC-40 had 7 launches, 4 were unsuccessful and three were successful. This site was not suitable as more than 67 % of all launches failed.
-

The first image (on the left) show site CCAFS SLC-40 location in relation to site CCAFS LC-40, which had a launch success % of 26 %.

- The second images (on the right) indicates the Launch outcomes for site CCAFS SLC-40, The three successful launches are indicated by a Green tag and the 4 red tags represent the failed launches.

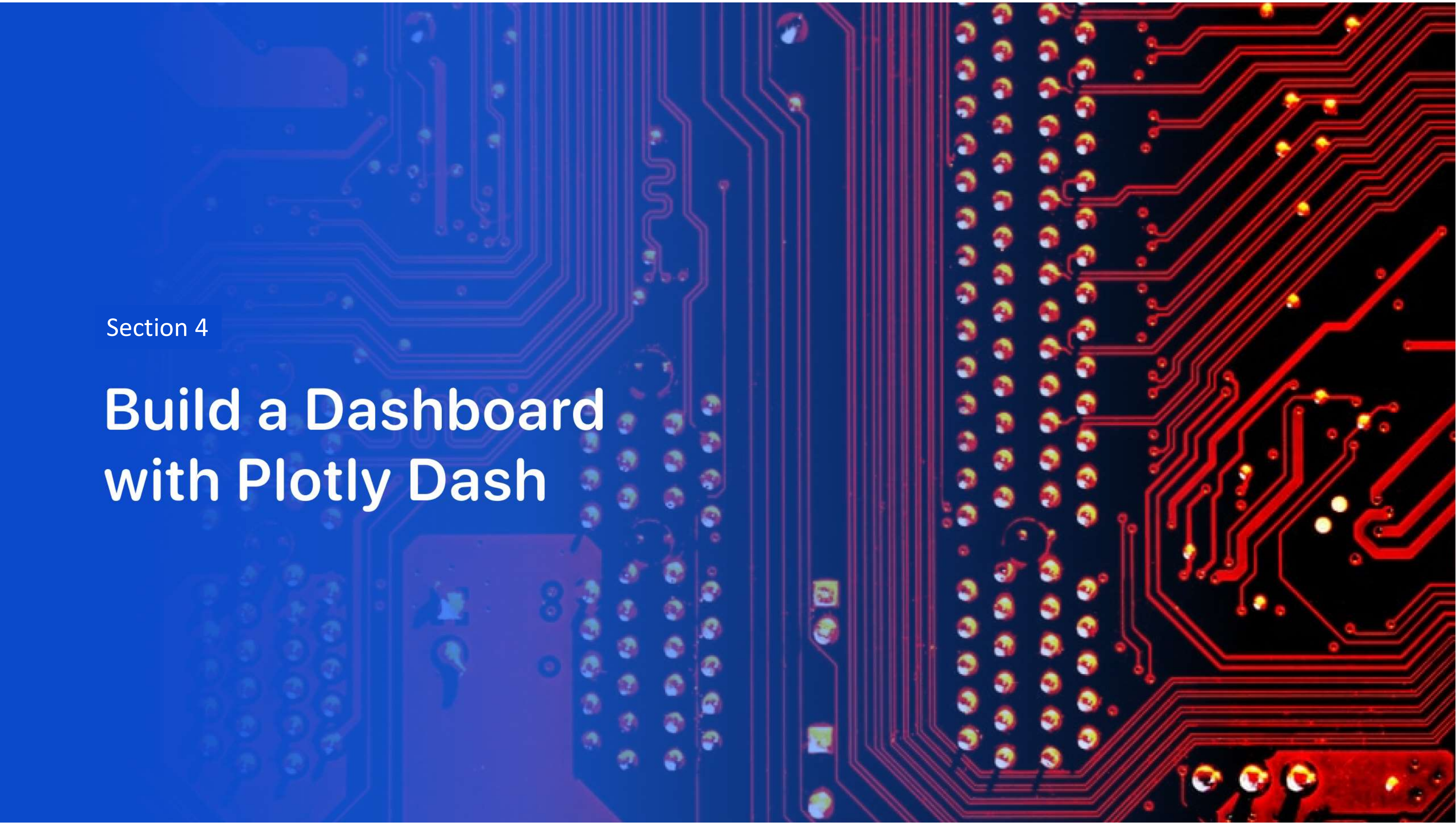


Proximity Analysis of Launch Site – CCAFS SLC-40

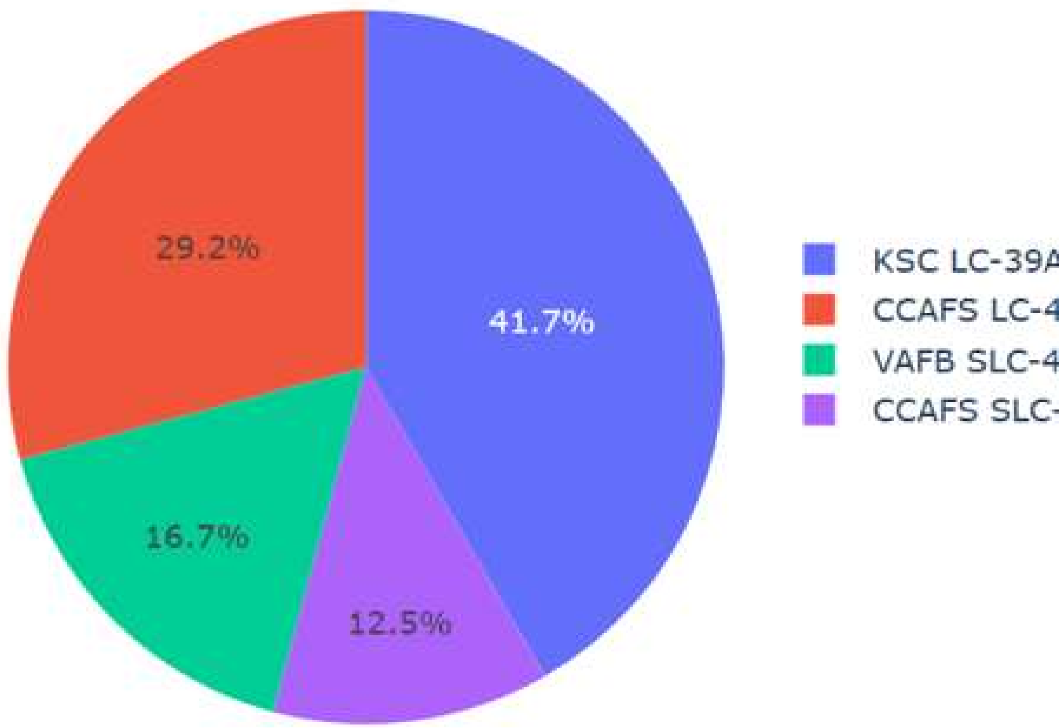
- The Launch site is far away from urban areas but close to major infrastructure (highways and railways).

- The proximity to the ocean is advantageous in the event of a rocket malfunction or if the rocket needs to self destruct over the ocean.

- Distance from Launch Site to Coastline = 0.86 km.
- Distance from Launch Site to Railways = 1.28 km.
- Distance from Launch Site to Highways = 0.59 km.
- Distance from Launch Site to cities = 52.91 km.

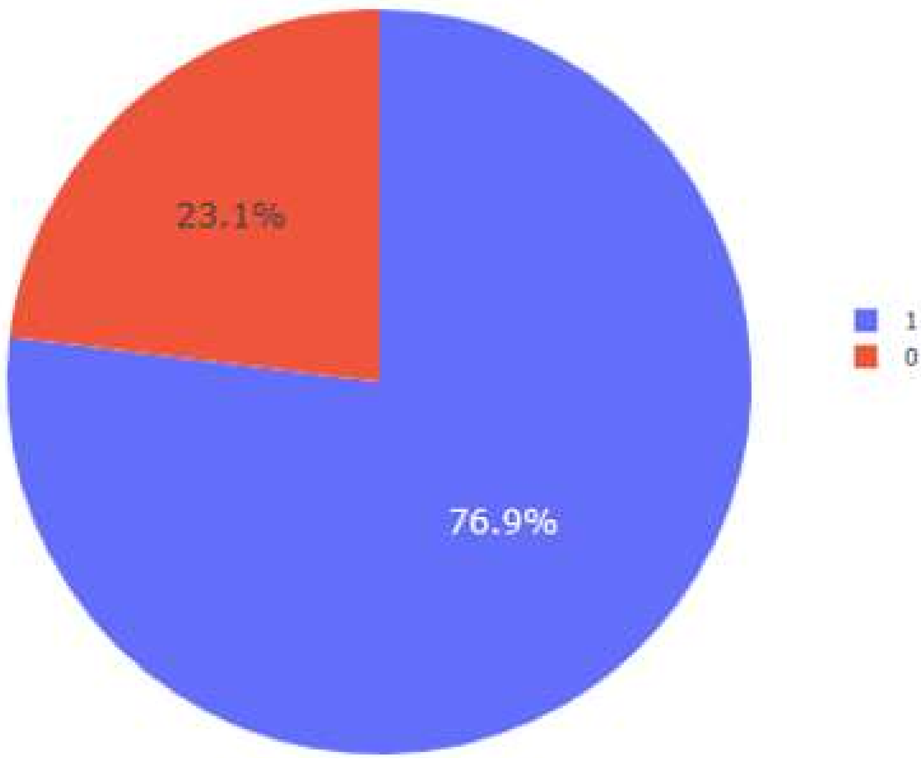

Section 4
Build a Dashboard
with Plotly Dash

% Successful Launches By Site



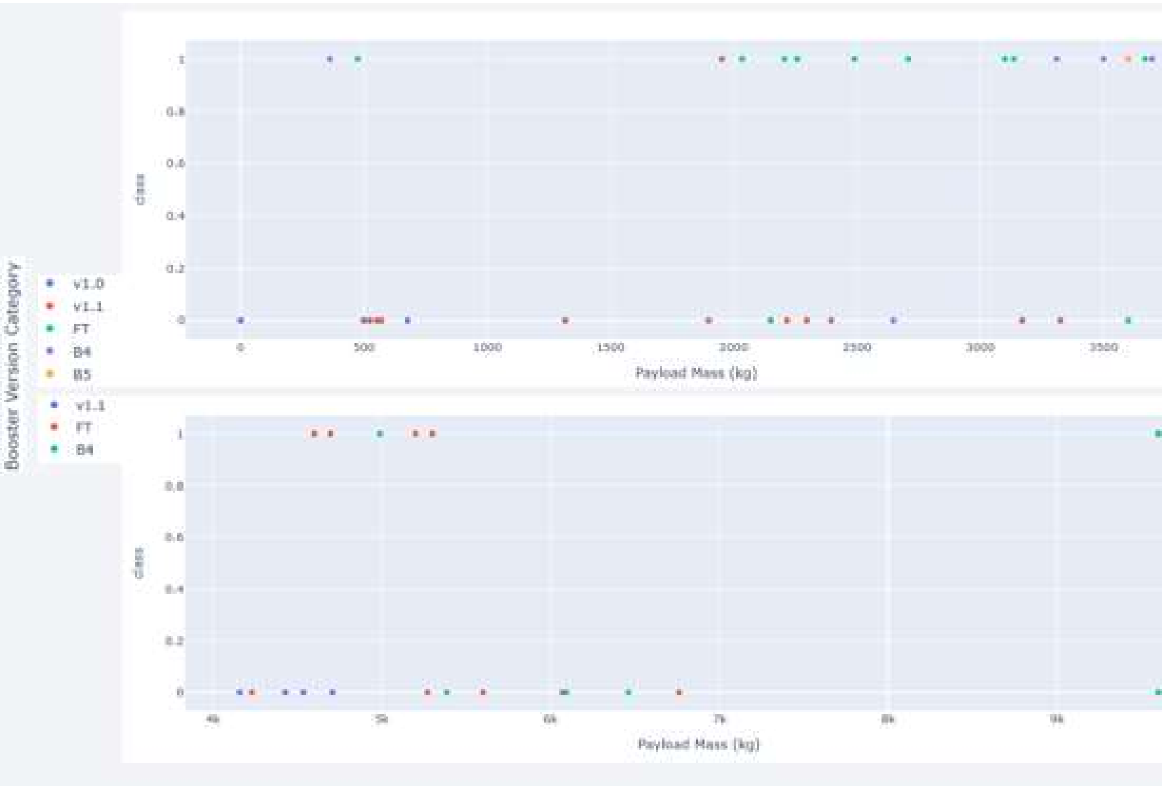
- Site KSC LC-39A had the most number of successful launches, and Accounted for 41,7 % of all successful launches.
- Site CCAFS SLC-40 had the least number of successful launches.

Most Successful Launch Site – Outcome Analysis



- Site KSC LC-39A had the most number of successful launches.
- Site KSC LC-39A has a success percentage of 76,9 % and an unsuccessful percentage of 23,10 %.
- The percentage is taken by dividing the number of successful outcomes for the site by the total number of launches for the site.

Impact of Payload on Outcome for a Booster Version

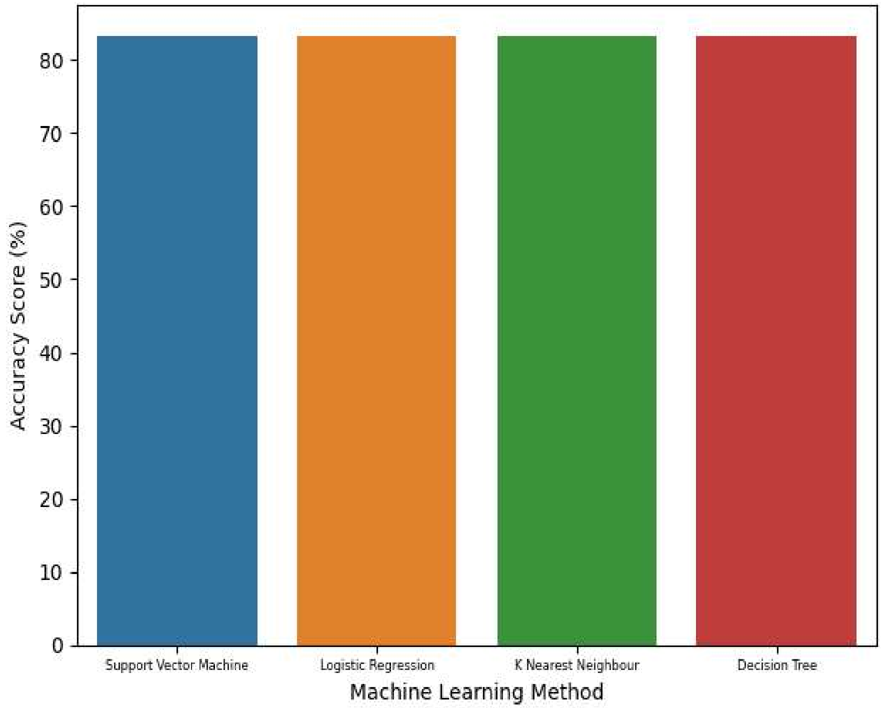


- Smaller payload have better landing outcomes regardless of booster versions.
- In the second slide the is a big disparity between success and failure. Payloads greater than 4000 kg are less likely to succeed.
- The B4 and FT booster version are the most successful booster versions.

Section 5

Predictive Analysis (Classification)

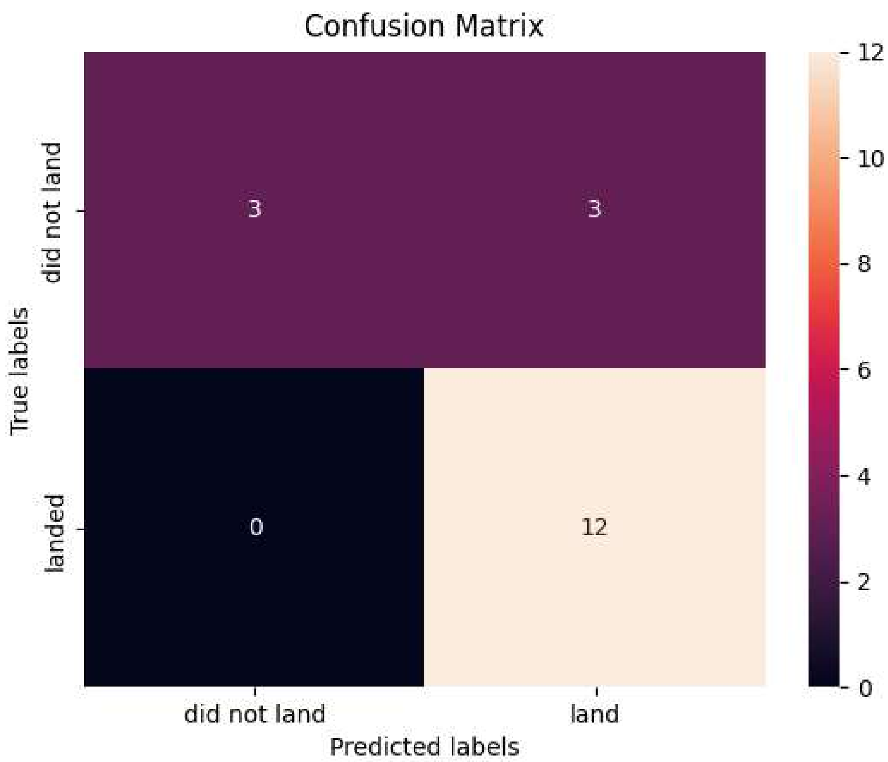
Classification Accuracy



All models have the same accuracy score %

	Machine Learning Method	Accuracy Score (%)
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

Confusion Matrix



- The best model was the k nearest neighbor model.
- The model correctly predicted 15 out of 21 outcomes,

Conclusions

- CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%
- VLEO offer SpaceX the capability to launch massive payload into space successfully, i.e. payload with masses of 12000 kg or more.
- SpaceX has seen a dramatic and sustained improvement in the success rate of launches as new orbit types, sites and boosters are used since 2014
- The best model was the KNN model with an 83,33 % Accuracy.


Thank you!