Winning Space Race with Data Science

Rebecca Aldaco
11/24/2025

Outline

- Executive Summary

- Introduction

- Methodology

- Results

- Conclusion

- Appendix

Executive Summary

The purpose of this capstone project is to determine whether SpaceX will reuse the first stage using a machine learning model and public data.

Based on data collected I have determined there is a correlation between location of launch and location of landing success and failure outcomes.

Introduction

- The purpose of this capstone project is to determine whether SpaceX will reuse the first stage using a machine learning model and public data.

- Can SpaceY be a competitor to SpaceX? What is the cost of each launch? Can we determine a successful first phase every time?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  - Created a table that displayed variations in launch such as booster version, payload mass, launch site etc.

- Perform data wrangling

  - Interpreted data sets that displayed type of orbit, success/ fail landing and landing location.

- Perform exploratory data analysis (EDA) using visualization and SQL

- Perform interactive visual analytics using Folium and Plotly Dash

- Perform predictive analysis using classification models

  - Using decision trees and best score

Data Collection

- Data was collected using an API. The API gave us data on launches, rocket specifications, launch and landing specifications.

- We used an API to acquire past launch data, we used the .json()method to acquire the data and then normalized the json data into a table. Unfortunately we ran into some null values but we replaced those null values with mean for the sake of the data.

Data Collection – SpaceX API

- The data collected from SpaceX API consisted of launch details such as payload weights, booster versions, cost per launch, success rate, location and details about launch pads and landing pads.

- https://github.com/RNALDACO/SpaceY/blob/5d2bdd025905e8276ccb0cd92afec2ef7cc46aff/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Using BeautifulSoup Falcon9 launch records were retrieved from Wikipedia and parsed and converted into a Pandas data frame.

- https://github.com/RNALDACO/SpaceY/blob/2a0e05e1f3386eea65cb90e4e9604a61f1bb56df/jupyter-labs-webscraping.ipynb

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	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
1	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
2	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
3	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
4	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
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577366	28.561857
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577366	28.561857
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577366	28.561857
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577366	28.561857
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577366	28.561857

Data Wrangling

- Data was processed by calculating:
- Number of launches on each site
- Number and occurrence of each orbit
- Number and occurrence of mission outcome of the orbit

Created a landing outcome label from the Outcome column

- https://github.com/RNALDACO/SpaceY/blob/2a0e05e1f3386eea65cb90e4e9604a61f1bb56df/labs-jupyter-spacex-Data%20wrangling.ipynb

	Outcome	Class
0	None None	1
1	None None	1
2	None None	1
3	False Ocean	1
4	None None	1

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	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	1
1	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	1
2	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	1
3	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	1
4	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	1

EDA with Data Visualization

- Exploratory Data Analysis and feature Engineering using Pandas and Matplotlib

- Visualizing the charts we can see that there are 4 sites that have high success rates, launch success rates increase by year and it seems that every flight that is successful SpaceX adds to the payload mass.

- https://github.com/R 67f3fd365/edadatavi
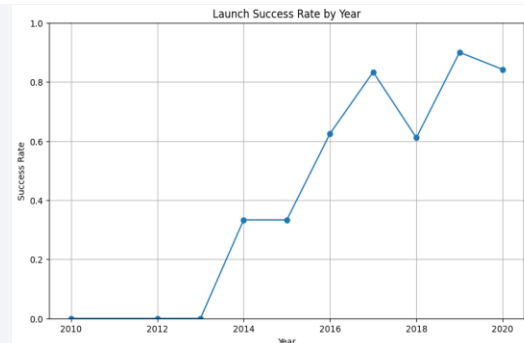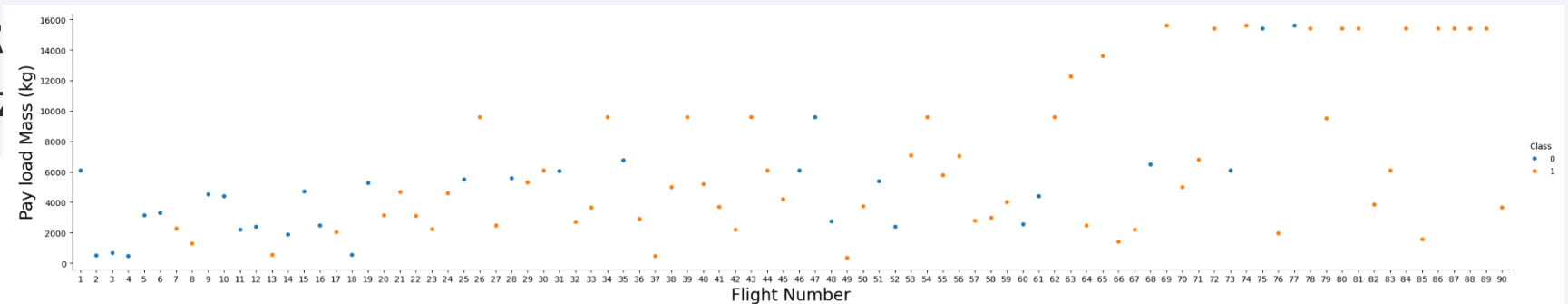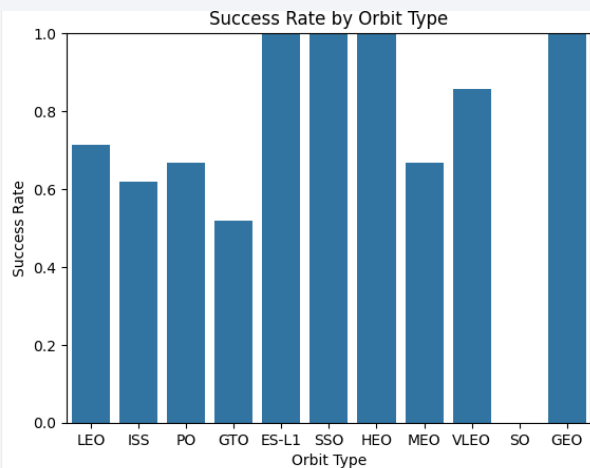
EDA with SQL

- Display the names of the unique launch sites in the space mission

- Display 5 records where launch sites begin with the string 'CCA'

- Display the total payload mass carried by boosters launched by NASA (CRS)

- Display average payload mass carried by booster version F9 v1.1¶

- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List all the booster versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- 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.

- https://github.com/RNALDACO/SpaceY/blob/2a0e05e1f3386eea65cb90e4e9604a61f1bb56df/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- A Folium Map was used to visually identify any commonalities and any landmarks that may be in close proximity using Marker Clusters, and MousePosition.

- Using MousePosition and Marker Clusters I can easily see the launch pads when I am exploring the surrounding area.

- Based on the findings I can say that launch pads are close to the coastlines and near the equator.

- https://github.com/RNALDACO/SpaceY/blob/2a0e05e1f3386eea65cb90e4e9604a61f1bb 56df/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Pie charts and scatter plots with range sliders are on the dashboard

- The pie charts and scatter plots with range sliders are interactive because they allow the exploration of the data in an easy understandable manner.

- https://github.com/RNALDACO/SpaceY/blob/953ebb14ccaee102ea3b2ec82e2d78 75041c9128/spacex-dash-app.py

Predictive Analysis (Classification)

- Loaded the data, then created a NumPy array. Standardized the data and split the X and Y into training and test data. Then created a logistic regression object, an SVM, a decision tree, a tree_cv and K nearest neighbors and then calculated the accuracy of the test data using the score method.

- https://github.com/RNALDACO/SpaceY/blob/9c2f3ec5f28d2572b6699d6f76dbbef67f3fd365/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- Falcon9 is the most common boosters used. CCFA SLC 40 is the most common launch.

- Dashboard with dropdown pie chart and scatter plot with range slider.

- Decision Tree preforms the best with the data given for this scenario.

	BEstScores
Logistic Regression	0.846429
SVM	0.848214
Decision Tree	0.889286
KNN	0.848214

Total Success Launces By All Sites

KSC LC-3
CCAFS LC
VAFB SLC
CCAFS SL

29.2%
41.7%
16.7%
12.5%

d range (Kg):

Correlation Between Payload and Mission Outcomes For All Sites

Booster Version Cate
- v1.0
- v1.1
- FT
- B4
- B5

16

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- The least amount of launches are coming from VAFB SLC 40

- KSC LC 39A has a cluster of flight numbers between 20 and 40

- Most launches are coming out of CCAFS SLC 40. Which is why they also have the most amount of failures.
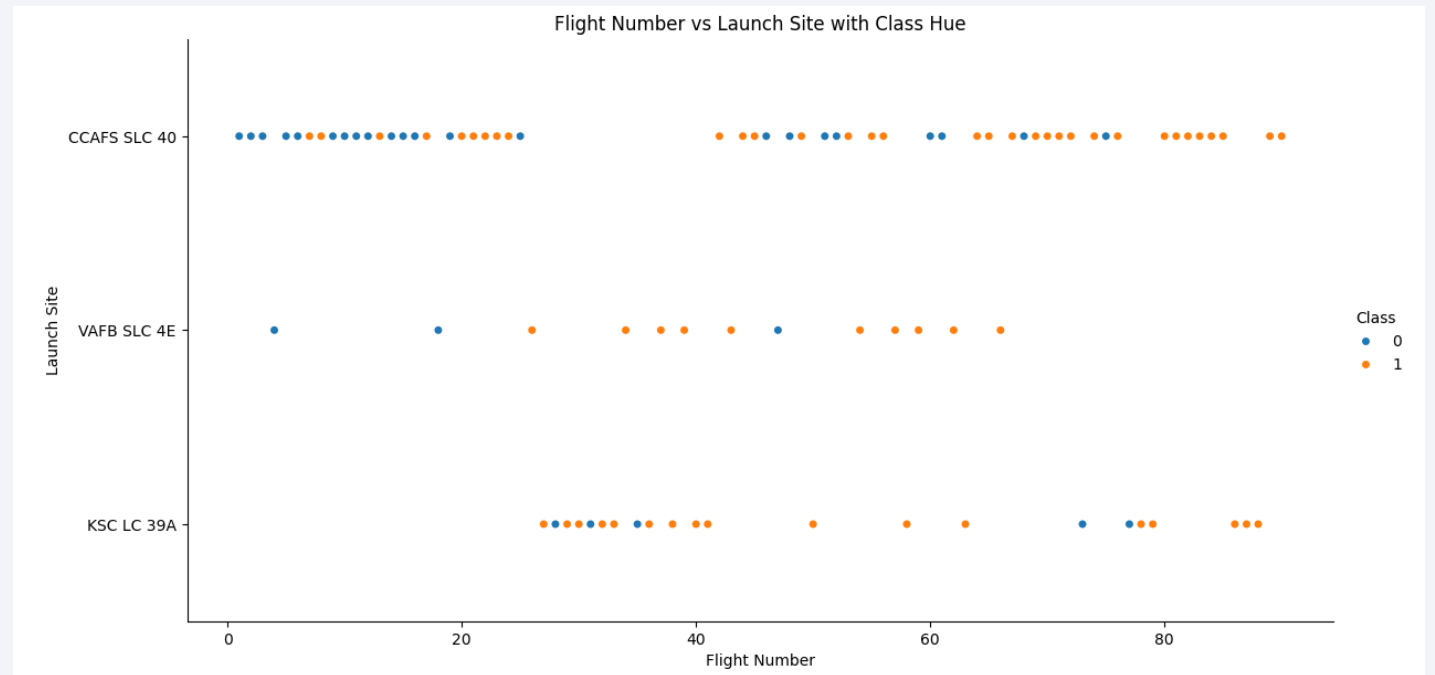
Payload vs. Launch Site

- CCAFS SLC 40 has the most amount of launches, variation in the amount of payload mass and also a success with the max payload mass

- KSC LC 39A has some variation in payload mass but 1 failure in max payload mass.
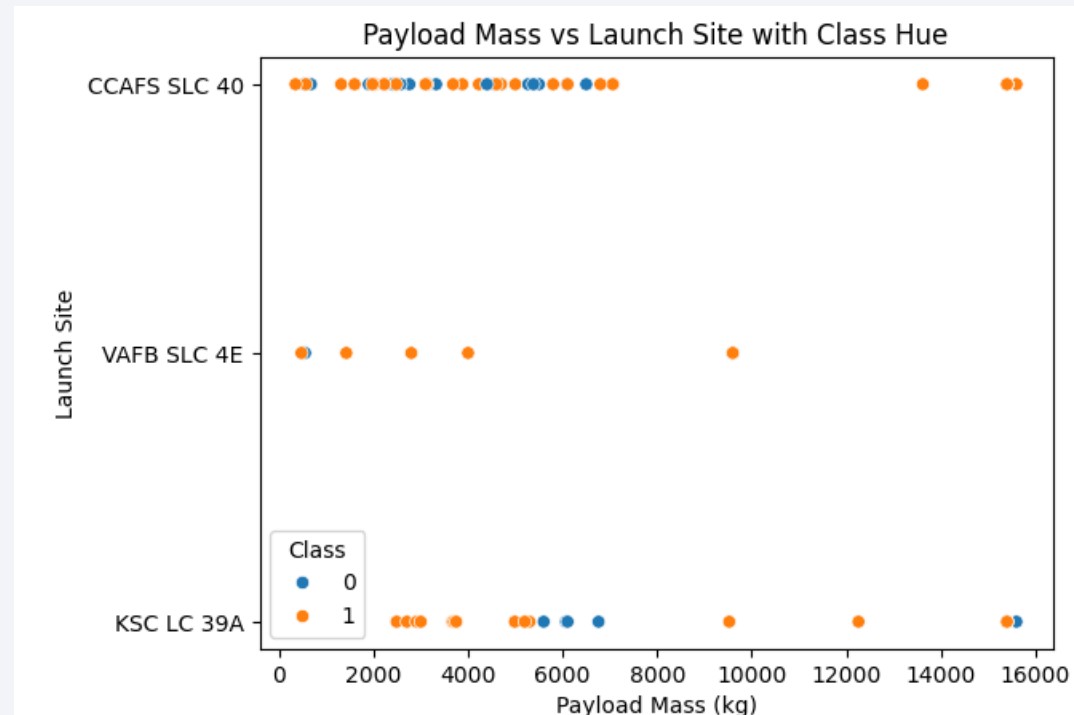
Success Rate vs. Orbit Type

- 4 types of orbit have high success rates.

- If one of these orbits can be confirmed before take-off then success rate can be determined.



Success Rate by Orbit Type
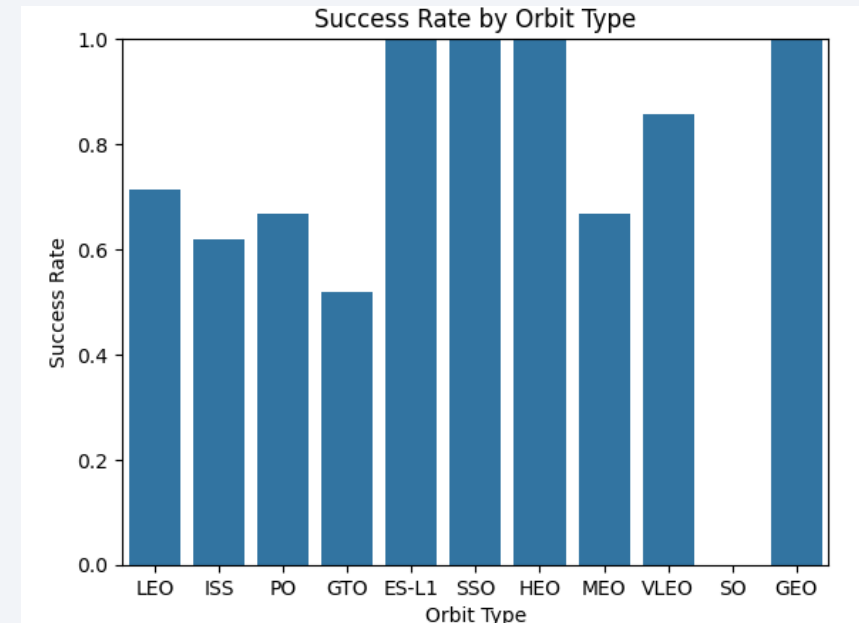
Flight Number vs. Orbit Type

- The higher the flight number than the smaller variation in orbit.

- The lower the flight number then the greater variation in orbit type. But higher rates of success.

Flight Number vs Orbit Type with Class Hue
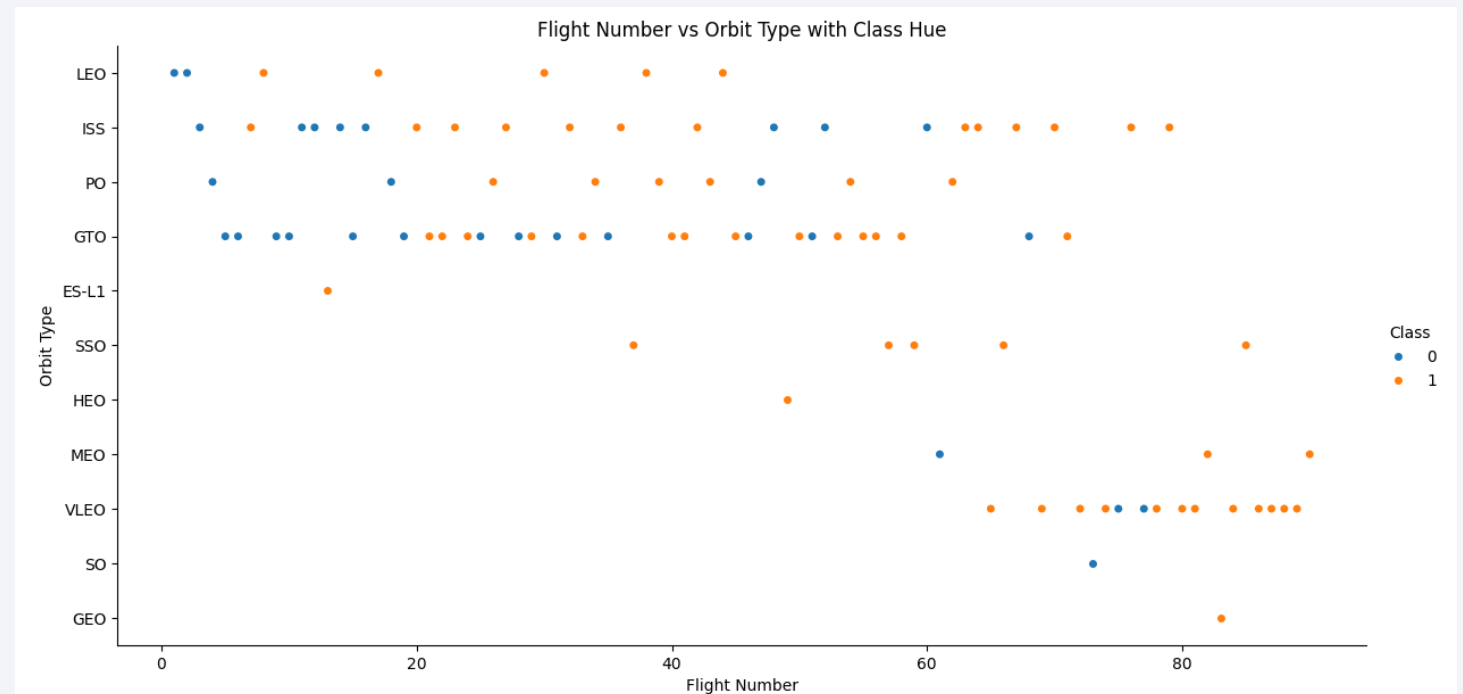
Payload vs. Orbit Type

- Geostationary transfer orbit has a seemingly high success rate with payloads between 3000kg and 7000kg.

- The ISS orbit has a tight payload mass variation between1500kg and 3000kg.
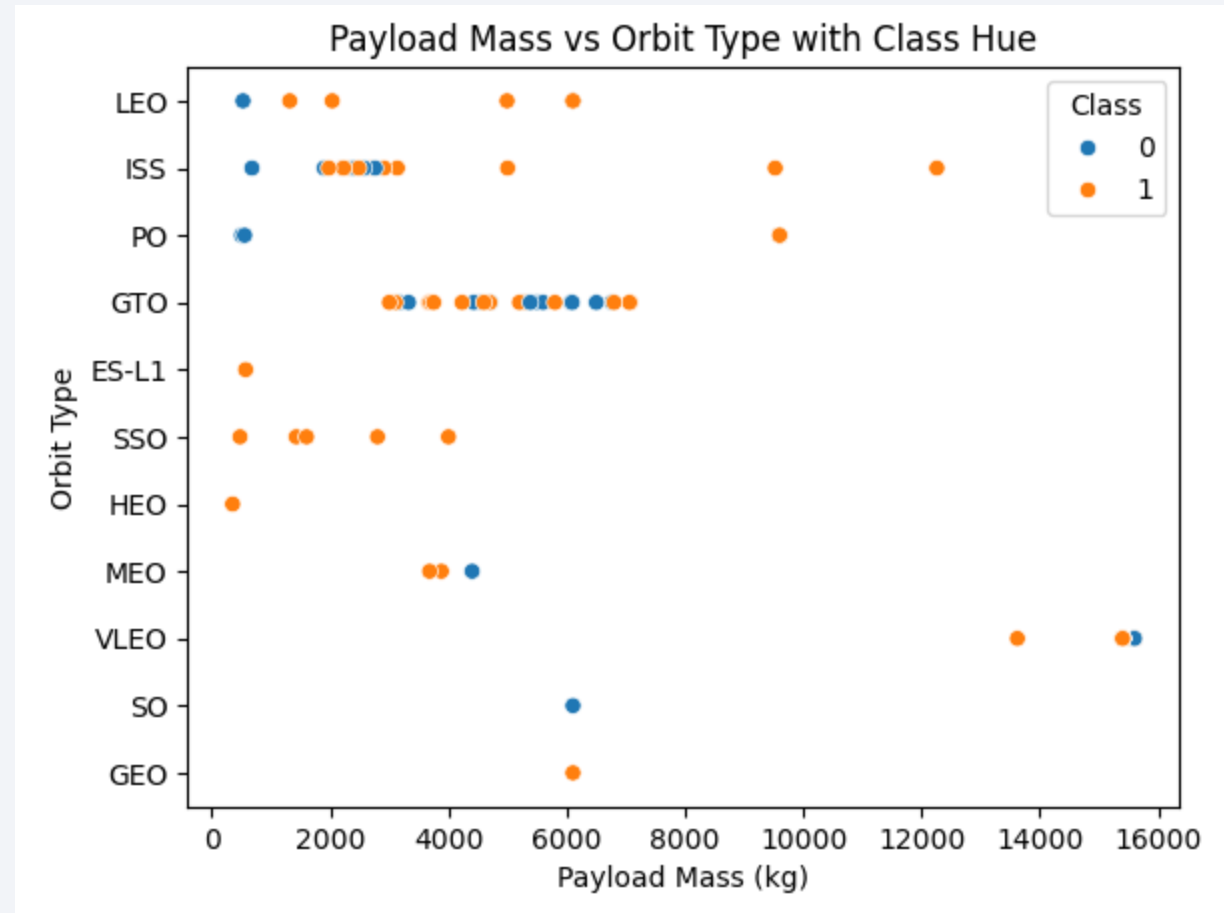
Launch Success Yearly Trend

- As time goes on and more launches take place success rates have started to increase since 2013. 2015 plateaued and 2018 had a sharp decrease in success rates.

Launch Success Rate by Year
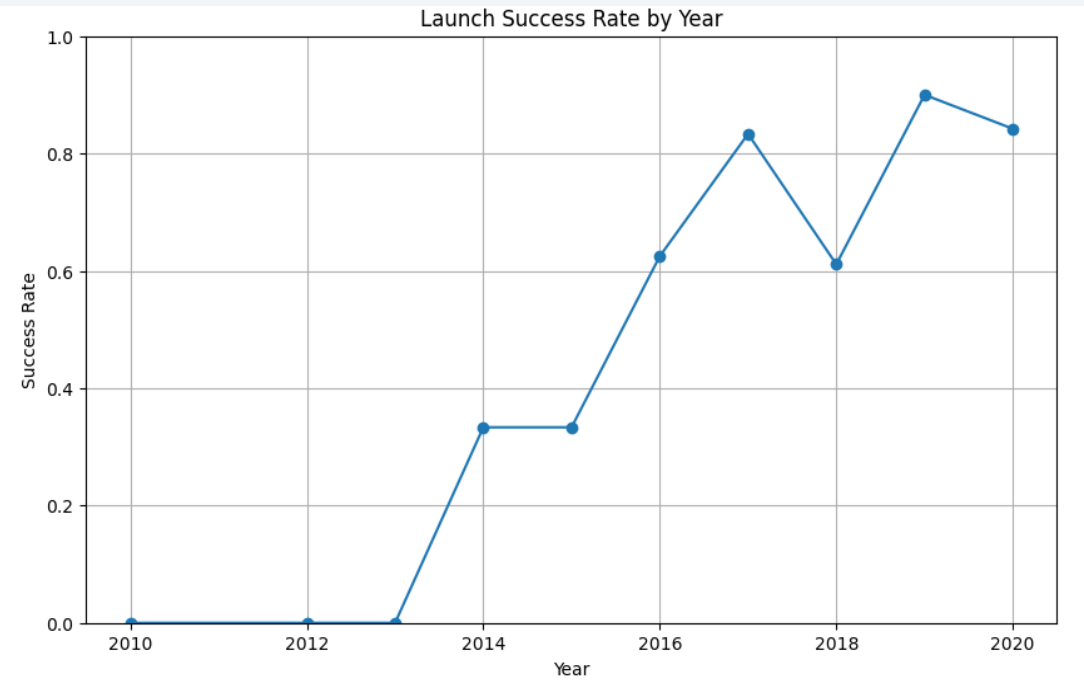
All Launch Site Names

- Below are all the names of the unique launch sites. The data was acquired by querying all the results of the launch site column.

```
('CCAFS LC-40',)
('VAFB SLC-4E',)
('KSC LC-39A',)
('CCAFS SLC-40',)
```

Launch Site Names Begin with 'CCA'

Data was found by using SQL and specifically asked to pull the first 5 launch sites starting with the letters 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
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

- Data was found by using SQL to specifically find the sum of payload mass from NASA (CRS).

According to the data the total payload mass is 45,596kg

total_payload_mass

45596

Average Payload Mass by F9 v1.1

- The Data was found by using SQL to calculate the average payload mass for the booster version F9 v1.1

- According to the data the average payload mass is 2928.4kg

average_payload_mass

2928.4

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad was found by using SQL and specifically asking for the date in which the landing outcome was successful on a ground pad.

According to the data the first successful landing date was 12/22/2015

first_successful_landing_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Data was gathered by using SQL and specifically asking to pull the booster version from a set table where the landing outcome was a successful drone ship landing with payload mass parameters higher than 4000kg but lower than 6000kg

- All of which were Falcon9

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

Total Number of Successful and Failure Mission Outcomes

- Data was gathered by using SQL and asking to specifically pull data from a table that was labeled mission outcome, counting how many of each and displaying it on a table.

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

Boosters Carried Maximum Payload

- All Boosters that have carried the maximum payload mass are Falcon9 Boosters according to the data.

- The data was gathered by using SQL by asking to display the names of the boosters that had the max payload mass.

Booster_Version
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

- There were only 2 failed launches in the year 2015 and both were from the same launch pad.

Data was gathered by using SQL and specifically asking to display the failed landing outcomes on a drone ship, with their booster versions and corresponding launch sites and display those on a table

month	Landing_Outcome	Booster_Version	Launch_Site
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

- Out of 31 launches 10 did not attempt to land 5 were successful on a drone ship. Another 5 failed to land on a drone ship. 3 had success on a ground pad.

- Data was found using SQL by specifically asked to find the landing outcome from 2010-06-04 to 2017-03-20 and sort them from descending order.

Landing_Outcome	outcome_count
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

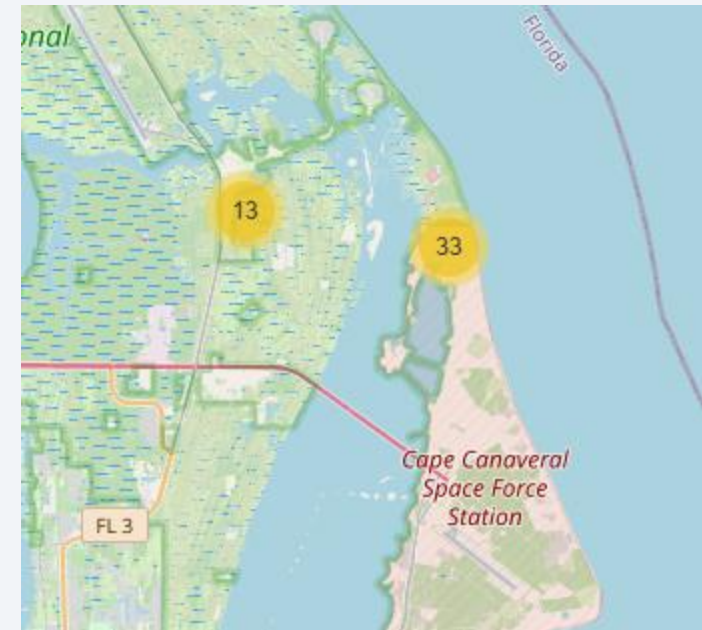# Launch Sites Proximities Analysis

Folium Global Map

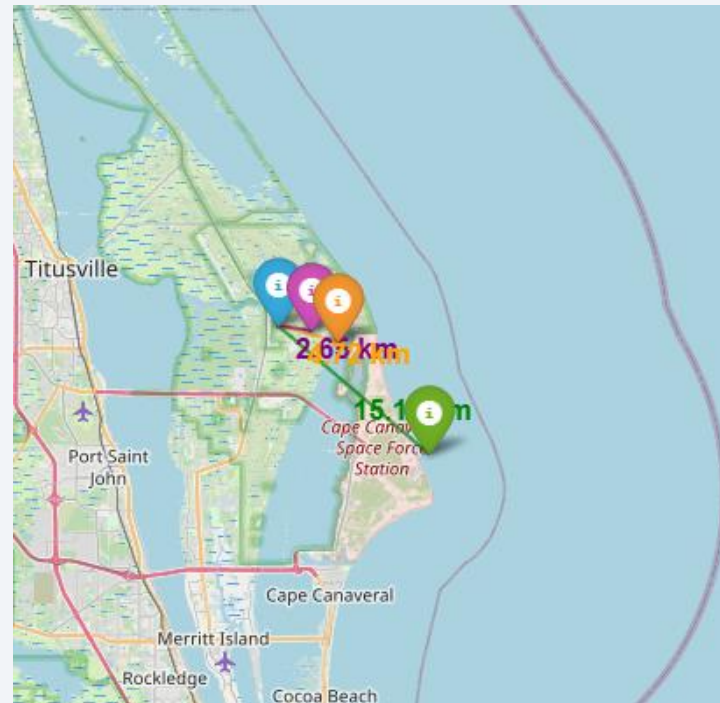All launch pads are on a coast line and located close to the equator.

Launch Outcomes

- According to the Folium Map those launch pads with higher numbers have greater success rates than those with smaller numbers

Launch Site Proximities

- Launch sites are somewhat close to railways and highways.

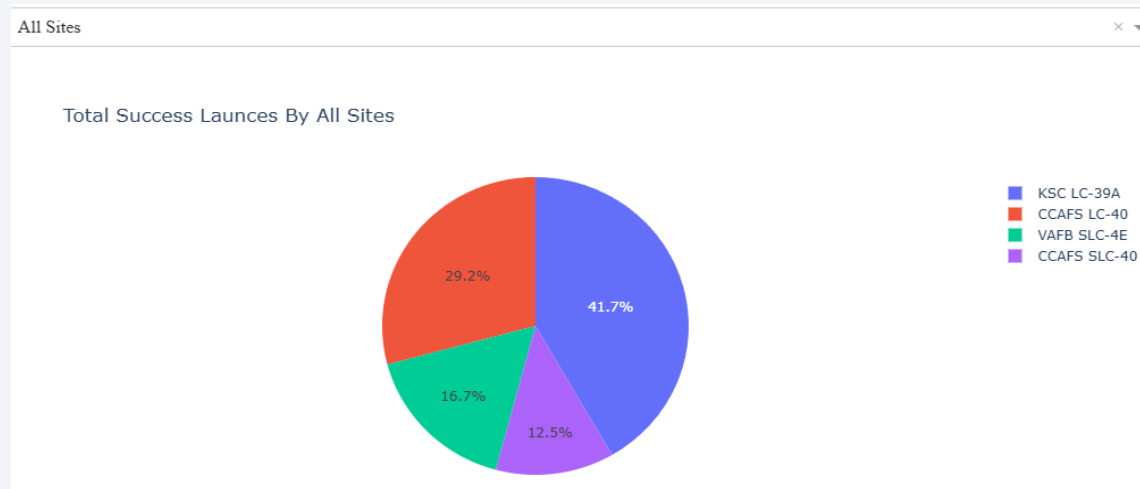- Launch sites are close to coastlines but away from cities.

Build a Dashboard with Plotly Dash

# Total Success Launches by all sites

- This interactive pie chart allows the viewer to go through each launch site and view the success and fail rate. KSC LC 39A had a higher success rate compared to other launch sites.
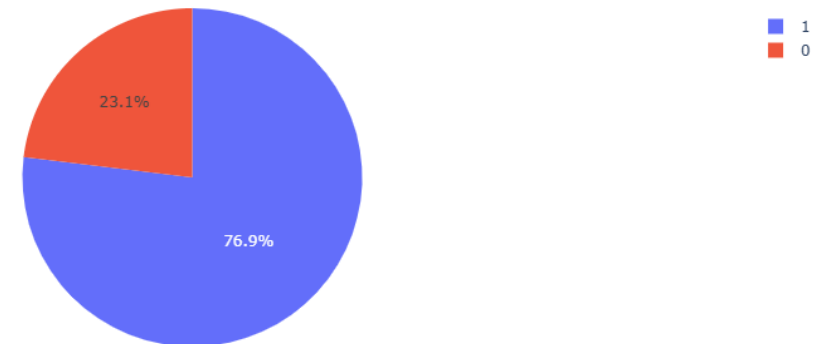
Launch Status with Highest Launch Success

- The Pie chart contradicts prior observation in believing that CCAFS SLC 40 had the highest launch success. Data was skewed due to the high volume of launches.

- KSC LC 39A has a higher success rate of 76.9% compared to CCAFS SLC 40 with a success rate of 57.1%

Launch status by: CCAFS SLC-40
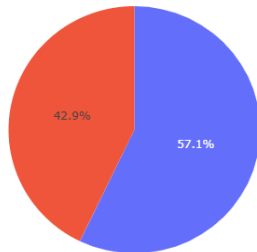
Launch status by: KSC LC-39A
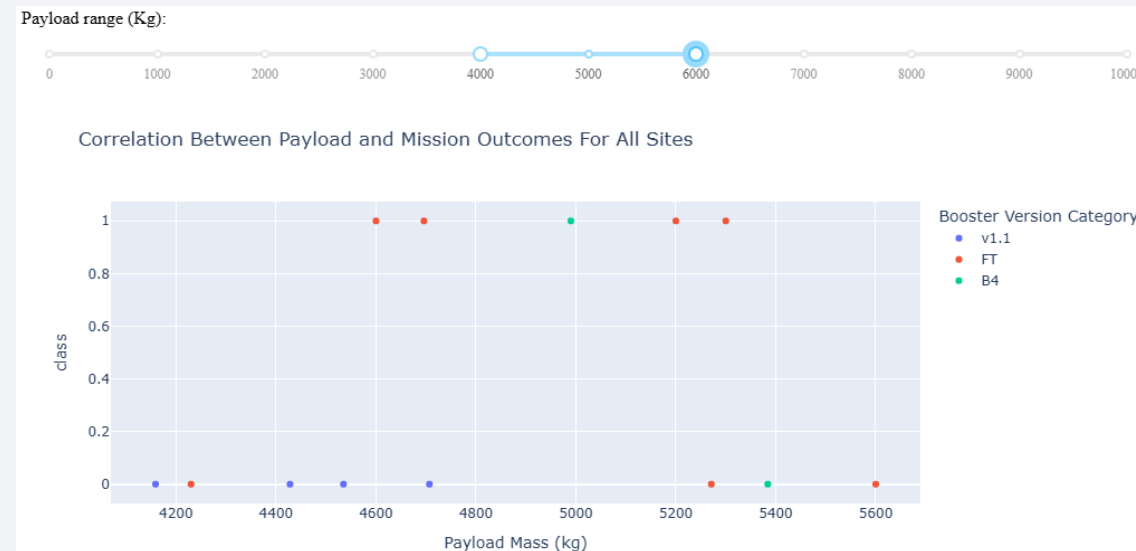
Payload vs. Launch Outcome

- The scatter plot shows the correlation bewteen the payload and mission outcomes for all sites with payloads between 4000kg and 6000kg categorized by the booster version.

- Booster version FT has higher success with higher payload mass compared to V1.1 has higher success rate with smaller payload mass.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The model with the highest classification accuracy is the decision tree with 88% accuracy.

	BEstScores
Logistic Regression	0.846429
SVM	0.848214
Decision Tree	0.889286
KNN	0.848214

Confusion Matrix

- The best preforming model of confusion matrix was a decision tree classifier with a accuracy of 88%.

True Postive - 12 (True label is landed, Predicted label is also landed)

False Postive - 3 (True label is not landed, Predicted label is landed)

Confusion Matrix
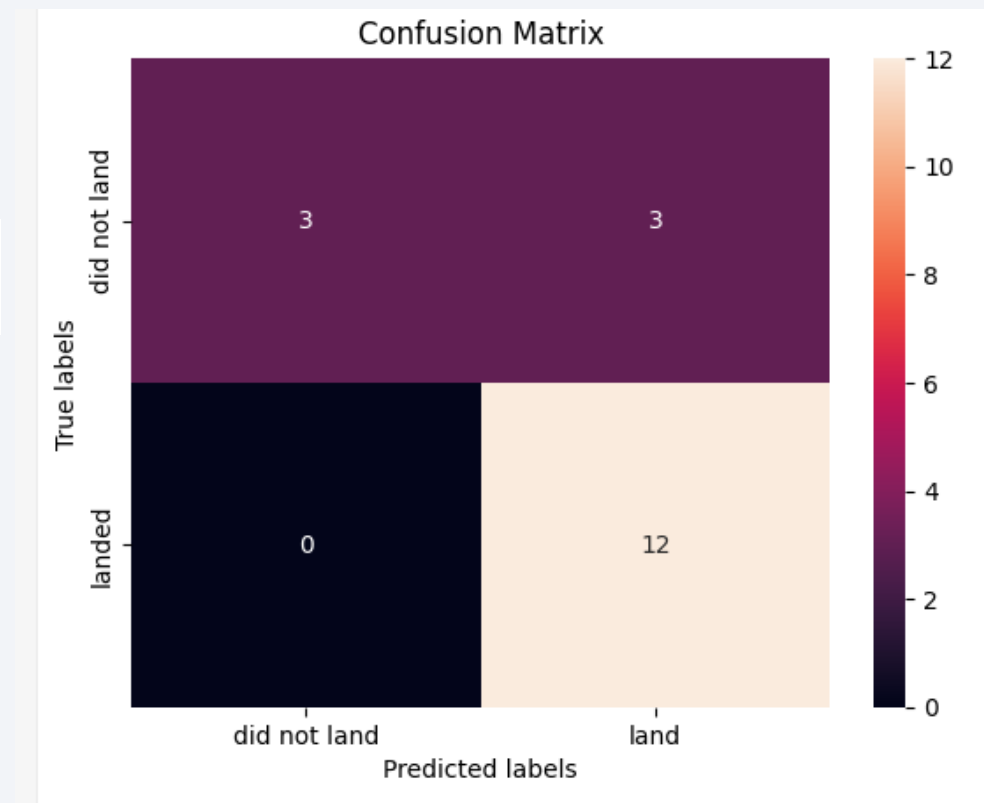
Conclusions

- If orbit type can be determined before takeoff then success rate will increase.

- CCAFS SLC 40 has the highest success rate in launches with a great variation in payload.

Appendix

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	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
1	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
2	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
3	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
4	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
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577366	28.561857
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577366	28.561857
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577366	28.561857
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577366	28.561857
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577366	28.561857

```
LaunchSite
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
```

```
Outcome
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
```

Thank you!