IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Wasakorn Pakdeesan

January 4th, 2024

Outline

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

Executive Summary

We will predict if the Falcon 9 first stage will land successfully. 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. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

We found that success rate has been increased since 2015 and Decision Tree is the most appropriate method according to the highest score for prediction.

Introduction

- SpaceX is probably the most successful company for space company. One reason is SpaceX's rocket launchings are relatively inexpensive compare to another company because they can reuse the first stage. The first stage does much of the work and is much larger to the second stage.
- We will determine the cost of each launch by gathering information and create dashboard and also determine if SpaceX will reuse the first stage.

Section 1

Methodology

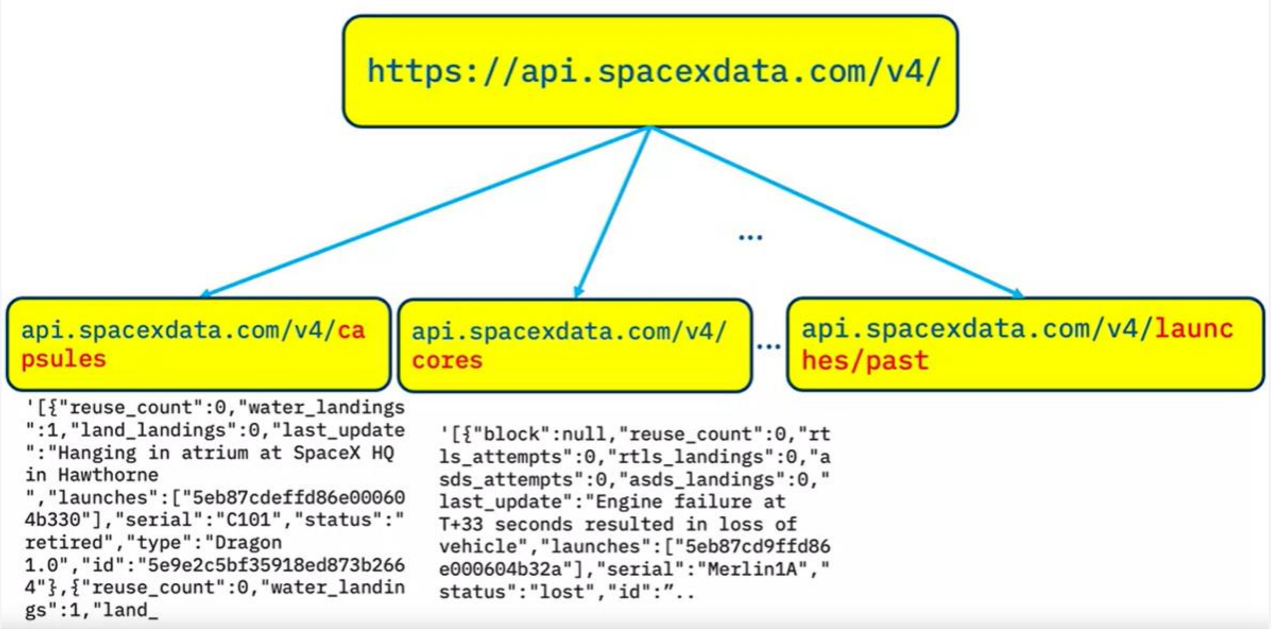
Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

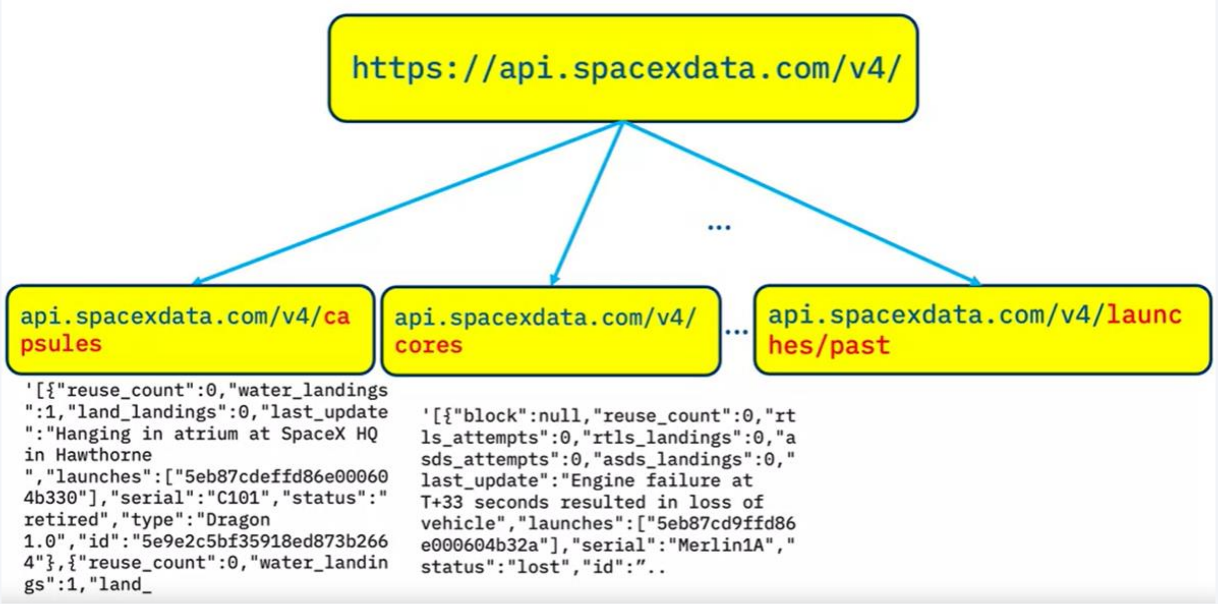
Data Collection

- Data will be gathered from an API, specifically the SpaceX REST API (https://api.spacexdata.com/v4/). This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.



Data Collection – SpaceX API

- https://github.com/wasakorn/capstoneproject/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- https://github.com/wasakorn/capstoneproject/blob/main/jupyter-labs-webscraping.ipynb

Wiki's SpaceX Launch record

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	Not attempted
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								

Web scraping with BeautifulSoup

	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	NaN
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	NaN
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	Not attempted\n	22 May 2012	NaN
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	NaN
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	Not attempted\n	1 March 2013	NaN
...	...	...	...	...	...	...	...	...	...	...	...

Data Wrangling

- Filter the dataframe to include only “Falcon 9” launches
  - Create new dataframe with BoosterVersion not equal to “Falcon 1”
- Dealing with missing values
  - Replace missing values with mean of PayloadMass
- https://github.com/wasakorn/capstoneproject/blob/main/jupyter-labs-spacex-data-collection-api.ipynb
- https://github.com/wasakorn/capstoneproject/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Categorical plot to see how the FlightNumber (indicating the continuous launch attempts.) and Payload variables would affect the launch outcome.
- Scatter plot to explain the patterns found in the Flight Number vs. Launch Site.
- Bar chart to find which orbits have high sucess rate.
- Scatter plot to see if there is any relationship between FlightNumber and Orbit type.
- Line plot to get the average launch success trend.
- https://github.com/wasakorn/capstoneproject/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Display the names of the unique launch sites in the space mission
  - select distinct "Launch_Site" from SPACEXTABLE
- Display 5 records where launch sites begin with the string 'CCA'
  - select "Launch_Site" from SPACEXTABLE where "Launch_Site" like "CCA%" limit 5
- Display the total payload mass carried by boosters launched by NASA (CRS)
  - select sum("PAYLOAD_MASS__KG_") as "Total_Payload" from SPACEXTABLE where Customer="NASA (CRS)"
- Display average payload mass carried by booster version F9 v1.1
  - List the date when the first succesful landing outcome in ground pad was acheived.

https://github.com/wasakorn/capstoneproject/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than
  - select Booster_Version from SPACEXTABLE where Landing_Outcome like "Success (drone%" and (PAYLOAD_MASS__KG_ between 4000 and 6000)
- List the total number of successful and failure mission outcomes
  - select Mission_Outcome,count(Mission_Outcome) from SPACEXTABLE group by Mission_Outcome
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
- 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.
  - select substr(Date, 6,2) as Month,Landing_Outcome,Booster_Version,Launch_Site from SPACEXTABLE where substr(Date,0,5)='2015' and Landing_Outcome like "Failure (drone%"

EDA with SQL

- 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.
    - select Landing_Outcome, count(Landing_Outcome) as Total from SPACEXTABLE where substr(Date,1,4)||substr(Date,6,7)||substr(Date,9,10) between '20100604' and '20170320' group by Landing_Outcome order by Total DESC

https://github.com/wasakorn/capstoneproject/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Each site's locations are marked on the map using site's latitude and longitude.
- Circle is used for highlighting an area with a text label on a specific coordinate
- Cluster is used when there are markers in the same location
- Explain why you added those objects
- https://github.com/wasakorn/capstoneproject/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Pie charts is used for visualizing launch success counts
- Scatter chart is used to see how payload may be correlated with mission outcomes for selected site(s).

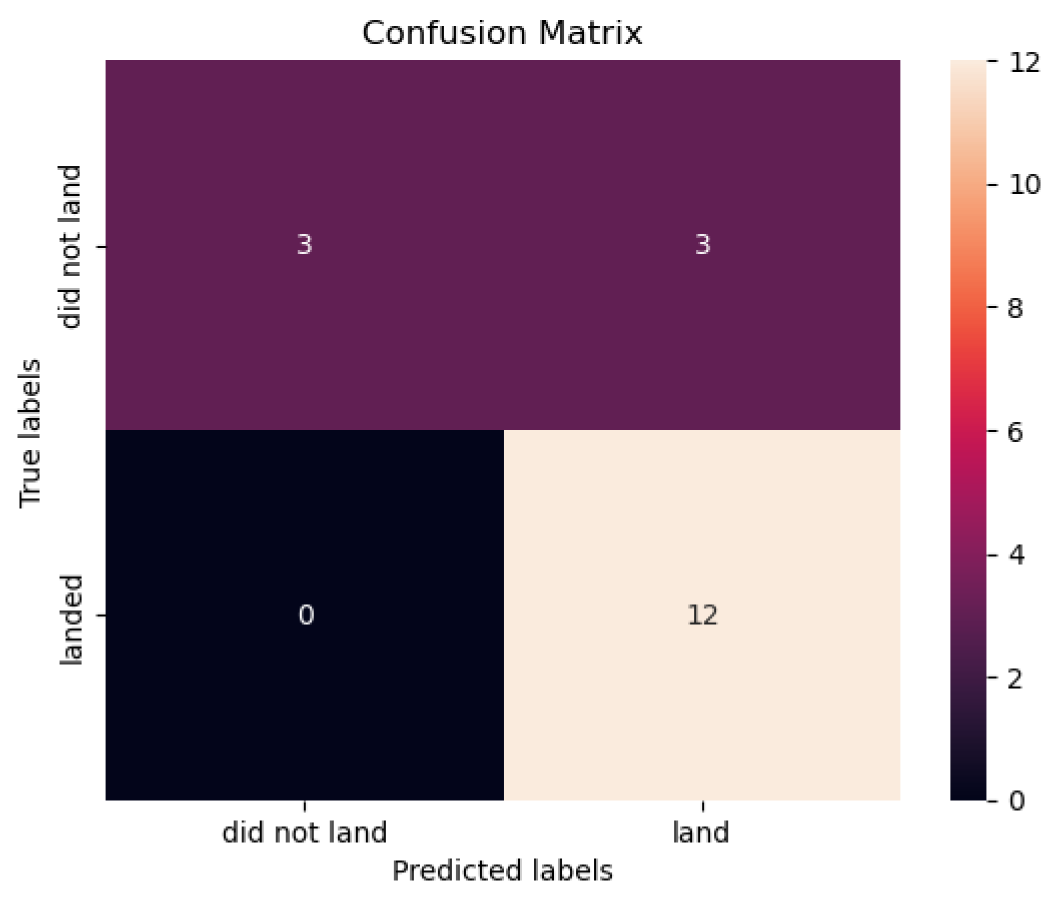
Predictive Analysis (Classification)

- Perform exploratory Data Analysis and determine Training Labels
  - Create a column for the class
  - Standardize the data
  - Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- https://github.com/wasakorn/capstoneproject/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- In this study, Decision Tree return the best score compared to another method

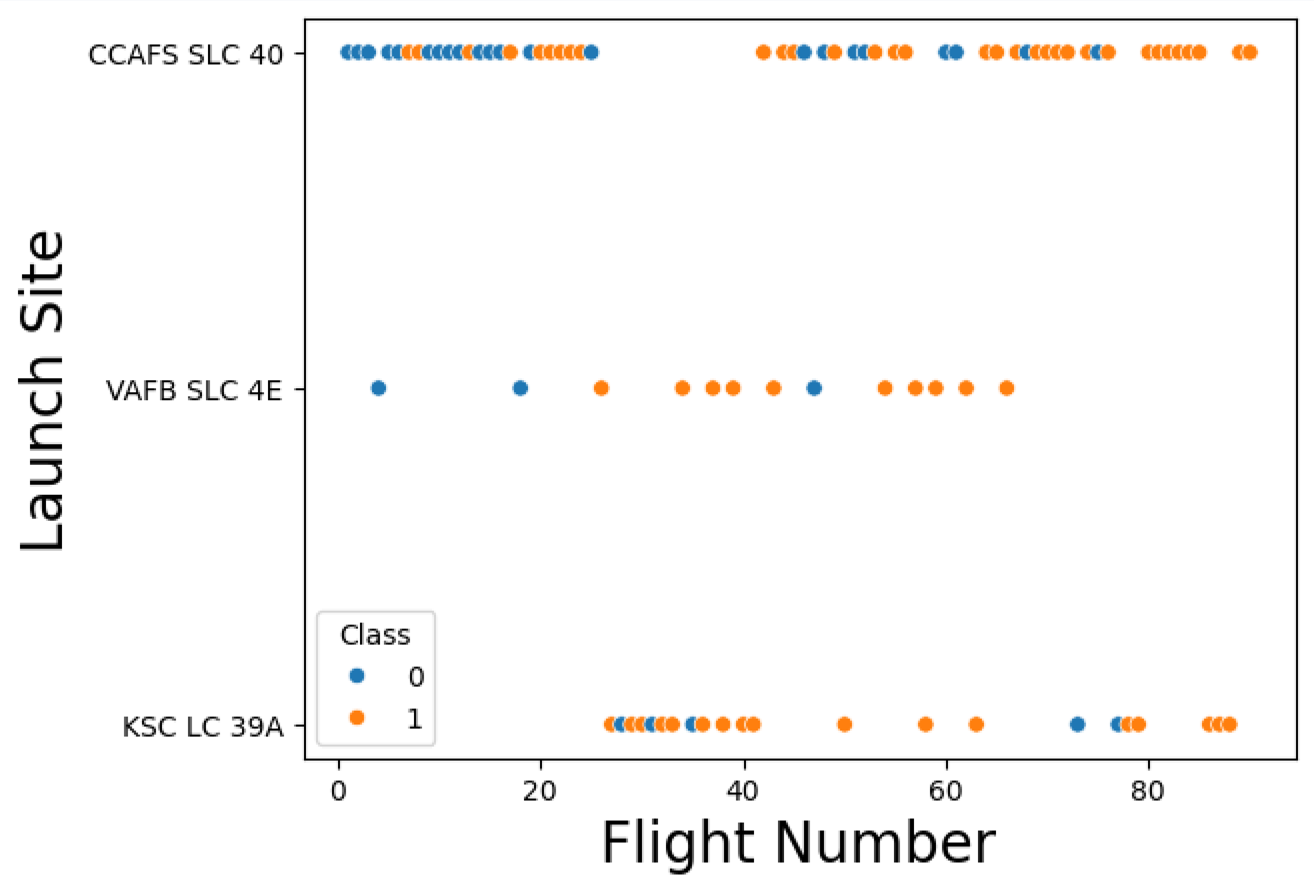
	Method	Score
0	LogisticRegression	0.846429
1	SVM	0.848214
2	DecistionTree	0.887500
3	KNN	0.848214



Section 2

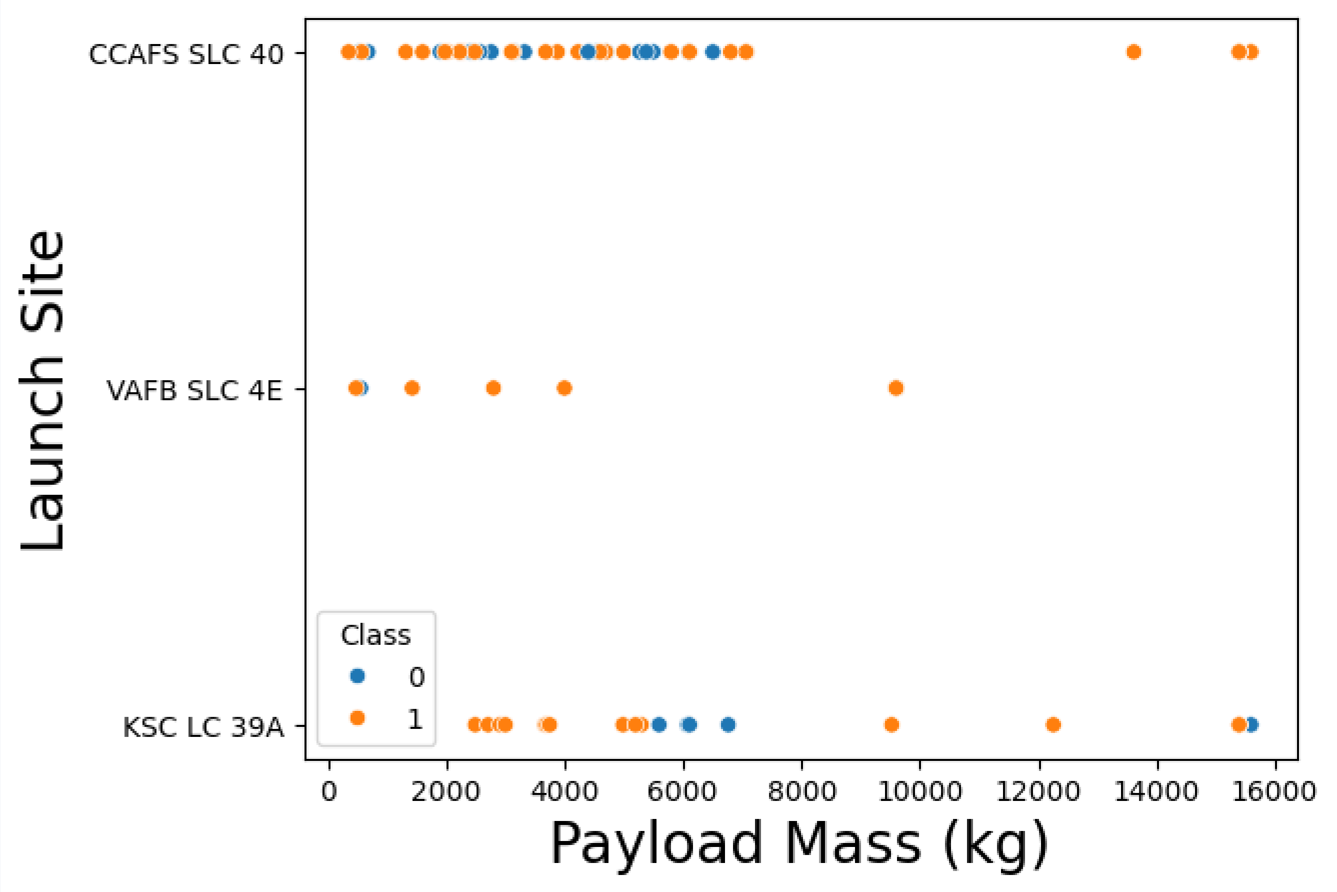
Insights drawn from EDA

Flight Number vs. Launch Site



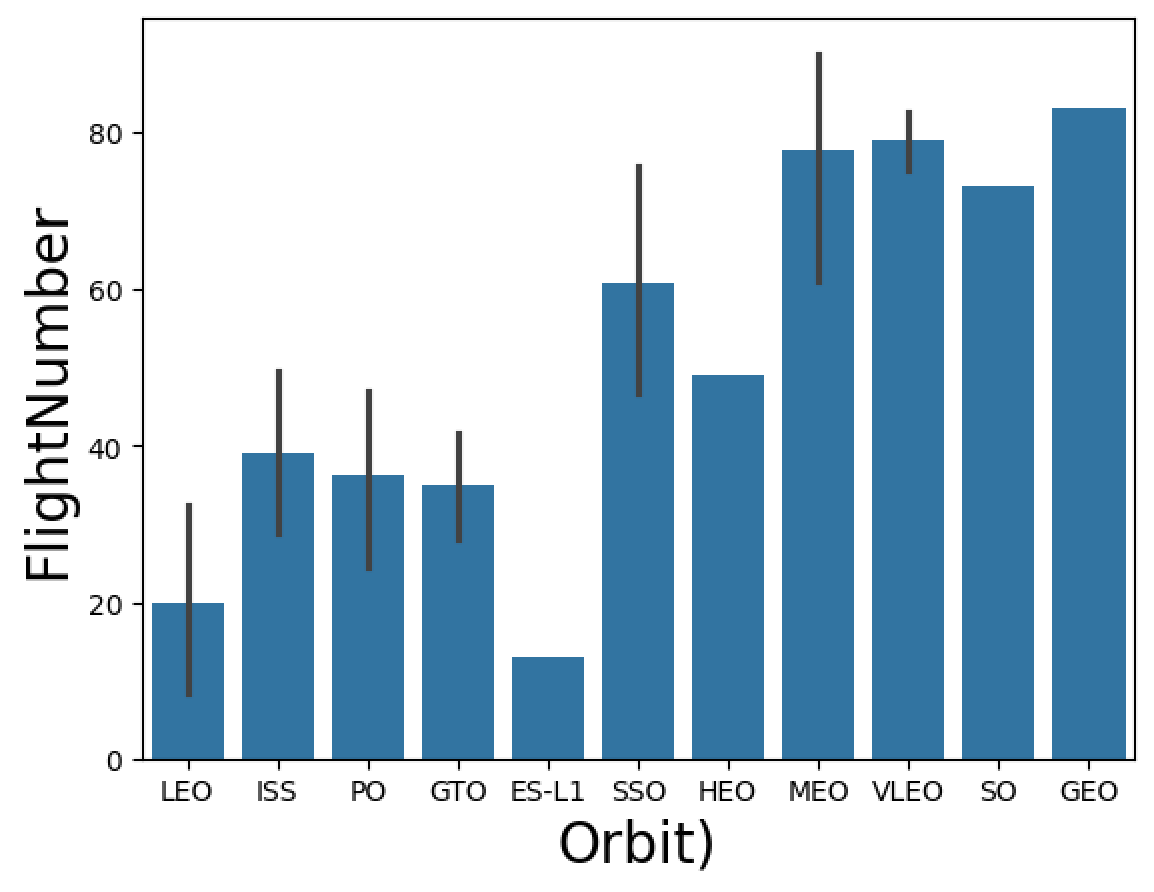
Payload vs. Launch Site

You will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

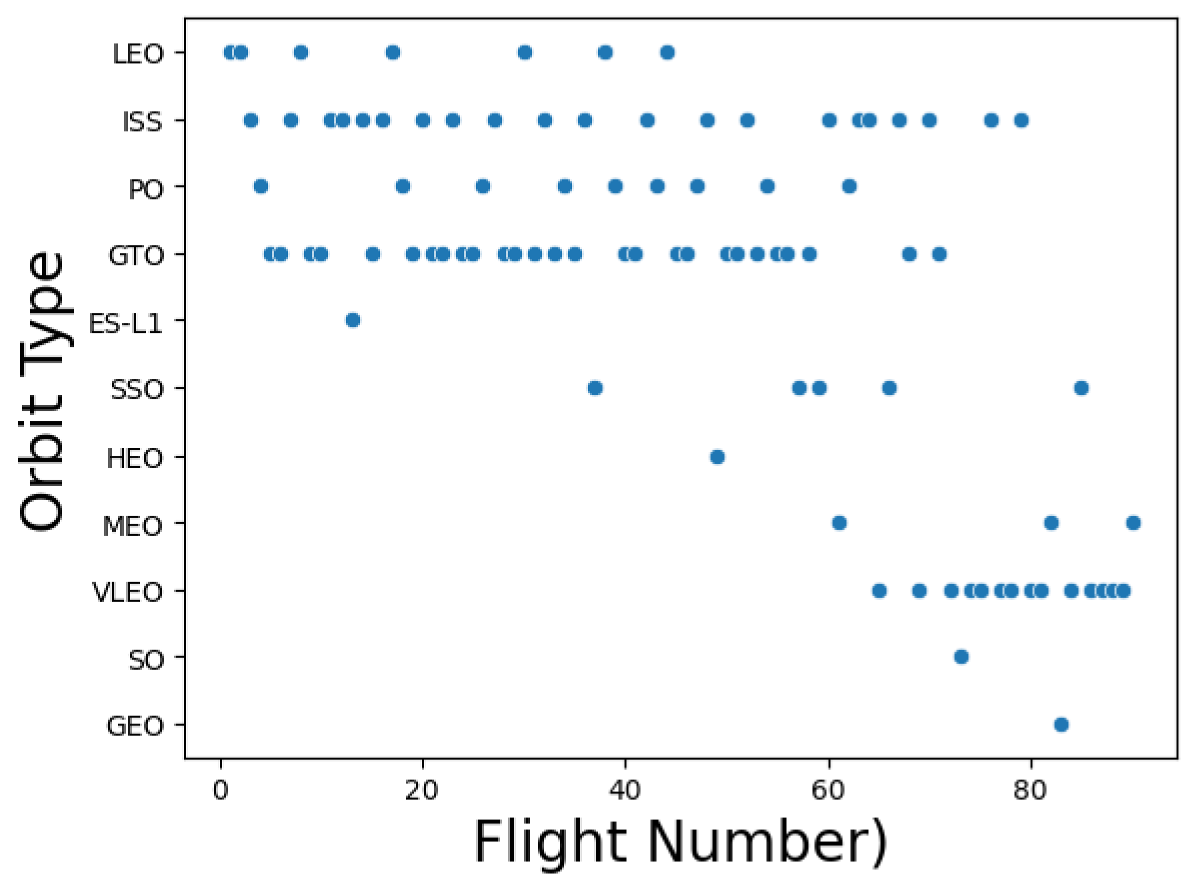


Success Rate vs. Orbit Type

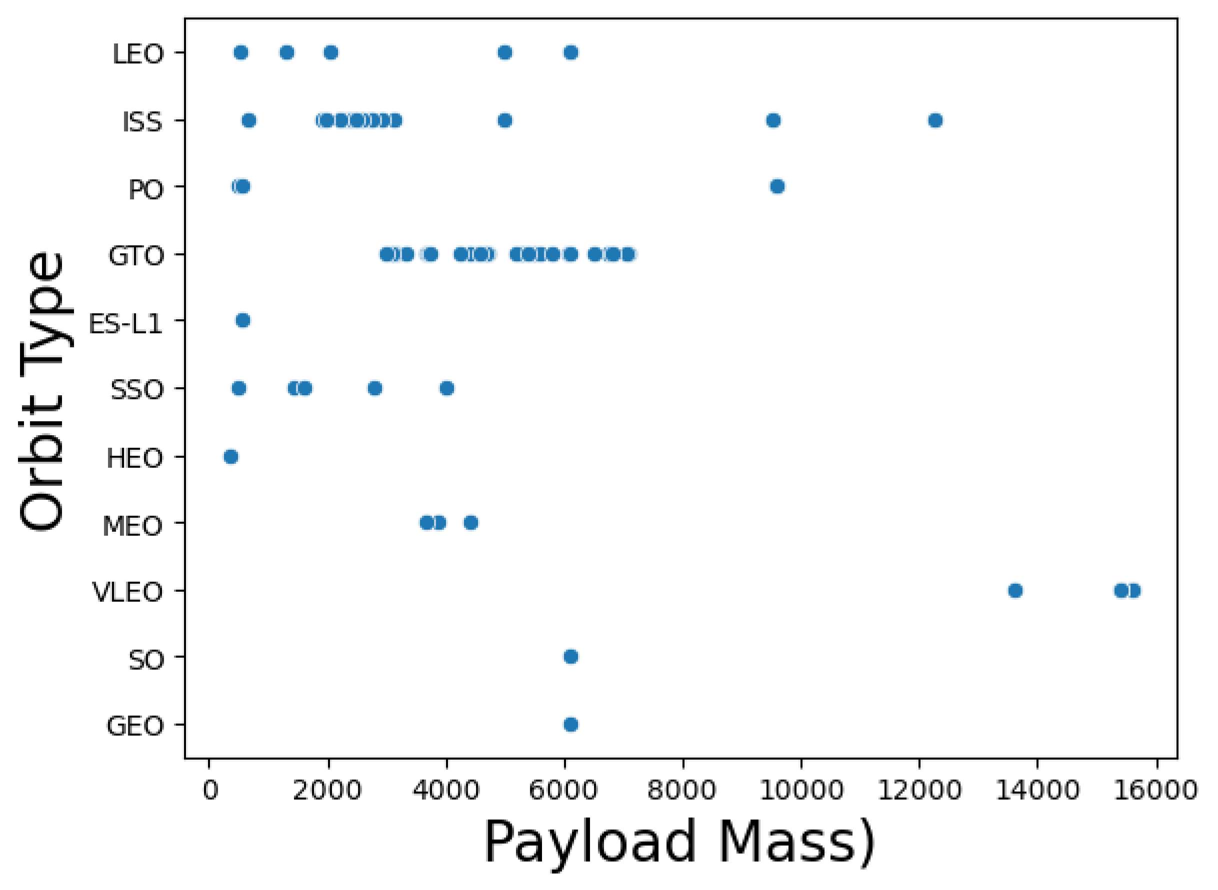
In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



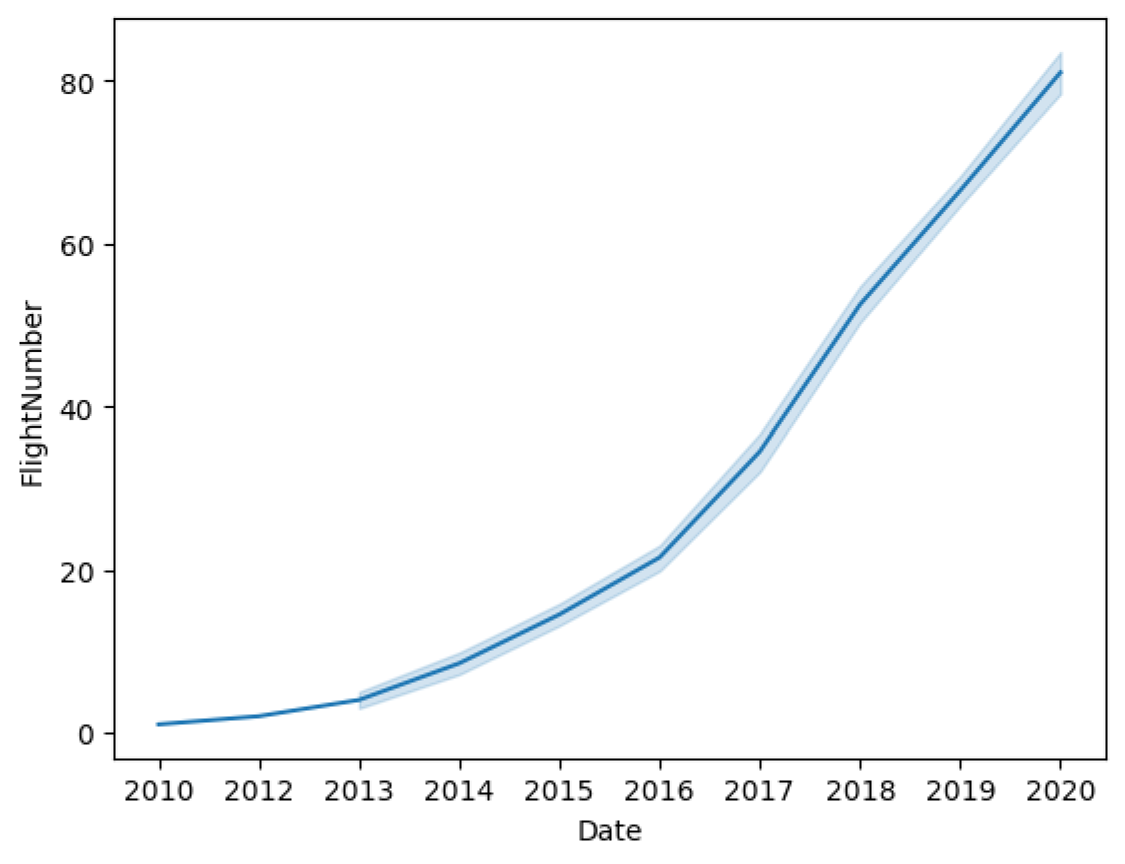
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

- Here are unique Launch Site names.

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `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

- Total payload carried by boosters from NASA is 45,596 KG
- Here is query and result.

```
In [10]:  #%sql select * from SPACEXTABLE where Customer="NASA (CRS)" limit 5
          %sql select sum("PAYLOAD_MASS__KG_") as "Total_Payload" from SPACEXTABLE where Customer="NASA (CRS)"

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

Out[10]:

Total_Payload
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is 2,534.66 KG
- Here is query and result.

```
In [11]: #%sql select * from SPACEXTABLE limit 5
         %sql select avg("PAYLOAD_MASS__KG_") as "Average_Payload_Mass" from SPACEXTABLE where Booster_Version like "F9 V1.1%"

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

Out[11]:

Average_Payload_Mass
2534.6666666666665

First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad is 2015-12-22
- Here is query and result.

```
In [12]:  #%sql select * from SPACEXTABLE where Landing_Outcome like "Success%" limit 5
          %sql select Date from SPACEXTABLE where Date=(select min(Date) from SPACEXTABLE where Landing_Outcome="Success (ground pad)'

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

Out[12]:

Date
2015-12-22

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 4000 but less than 6000
-

```
In [13]: #%sql select * from SPACEXTABLE limit 5
         %sql select Booster_Version from SPACEXTABLE where Landing_Outcome like "Success (drone%" and (PAYLOAD_MASS__KG_ between 40(

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

Out[13]:

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

Total Number of Successful and Failure Mission Outcomes

- Here is query and result

```
In [14]:
#%sql select * from SPACEXTABLE limit 5
%sql select Mission_Outcome,count(Mission_Outcome) from SPACEXTABLE group by Mission_Outcome

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

Out[14]:

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

Boosters Carried Maximum Payload

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

```
In [15]: #%sql select * from SPACEXTABLE limit 5
         %sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)

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

Out[15]:

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

- The below picture show the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Here is query and result

```
In [16]: #%sql select * from SPACEXTABLE limit 5
         %sql select substr(Date, 6,2) as Month,Landing_Outcome,Booster_Version,Launch_Site from SPACEXTABLE where substr(Date,0,5)=

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

Out[16]:

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

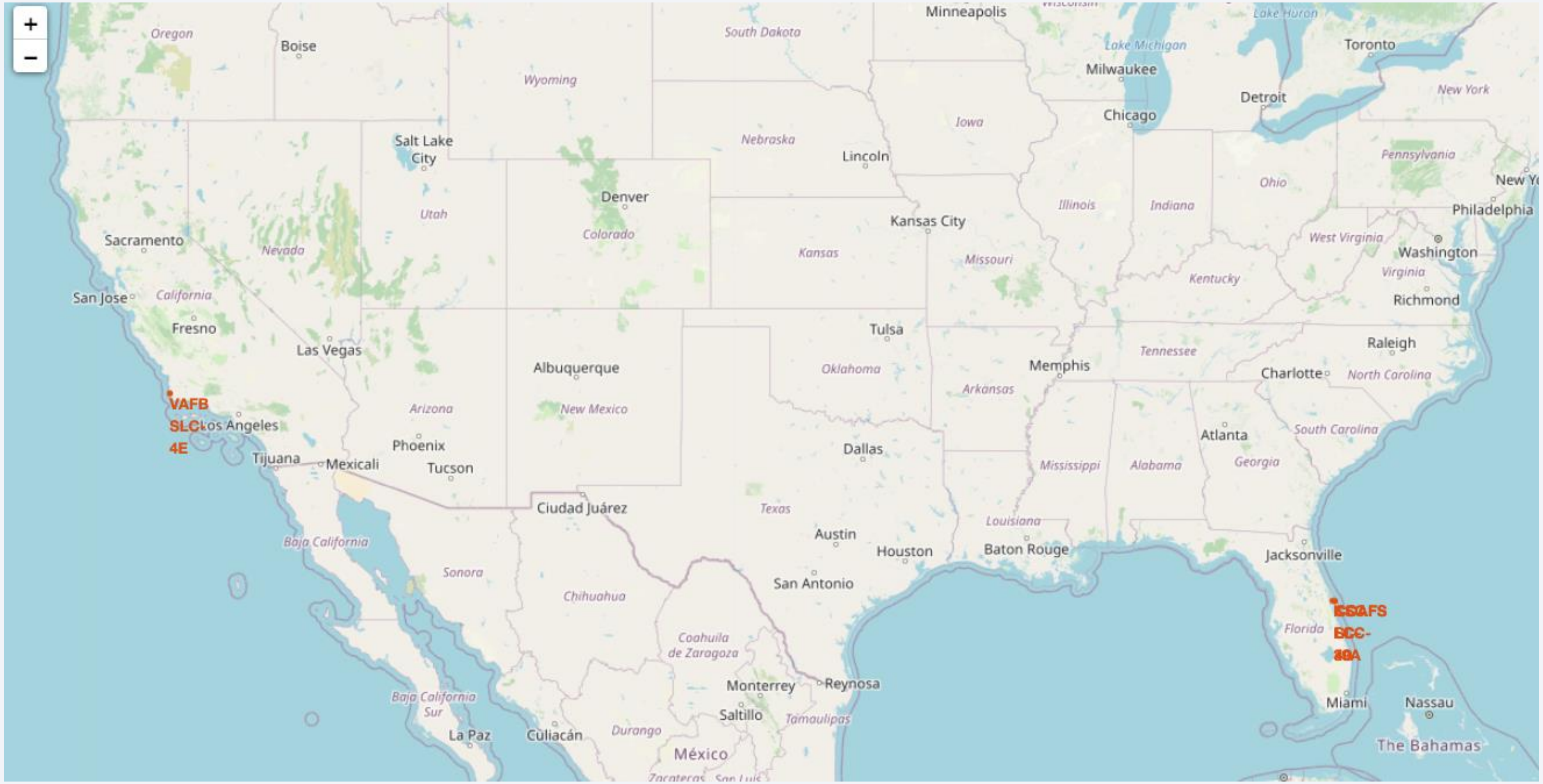
- Here is the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	Total
No attempt	10
Success (drone ship)	6
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

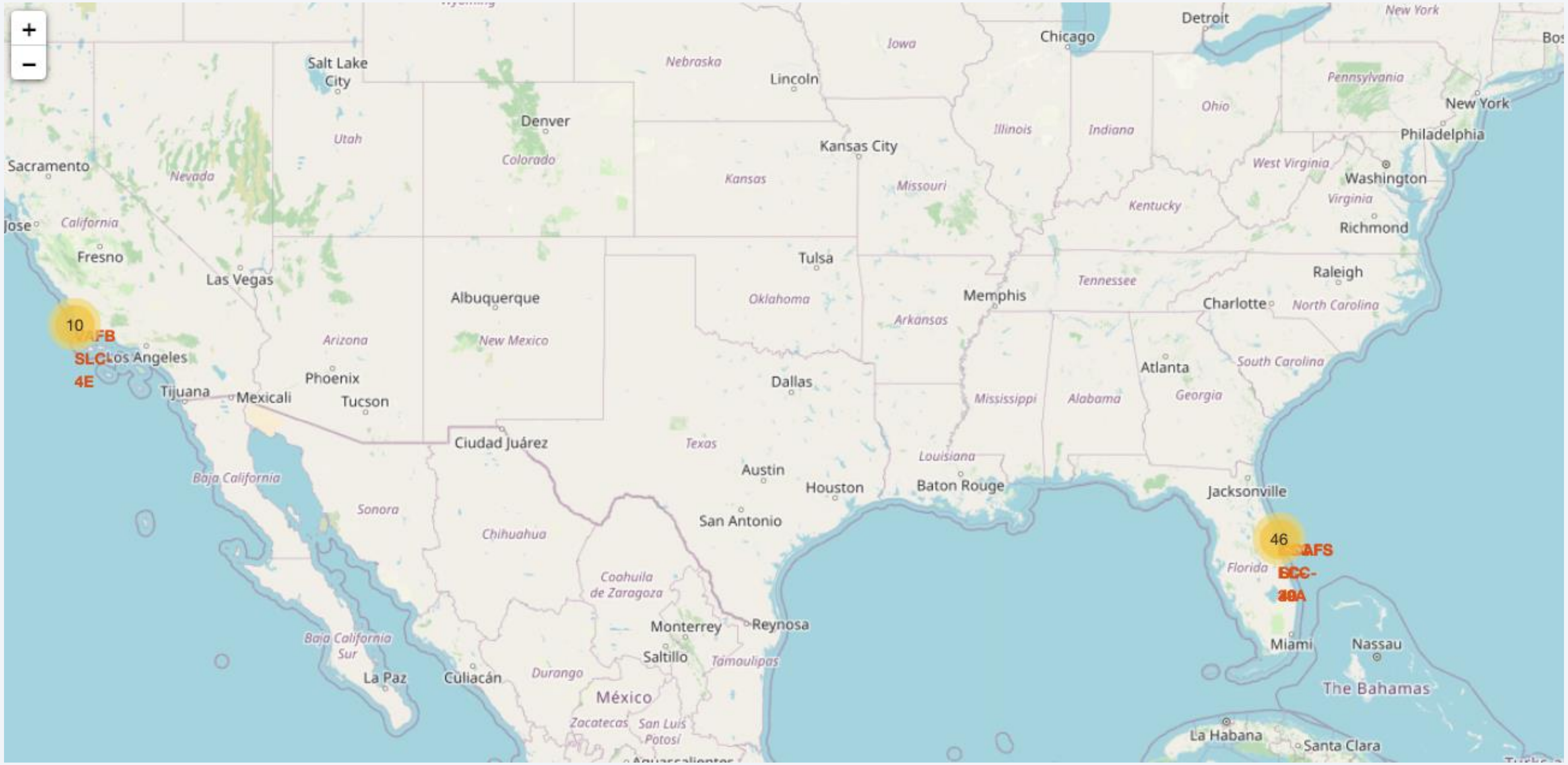
Section 3

Launch Sites Proximities Analysis

All Global Launch Sites



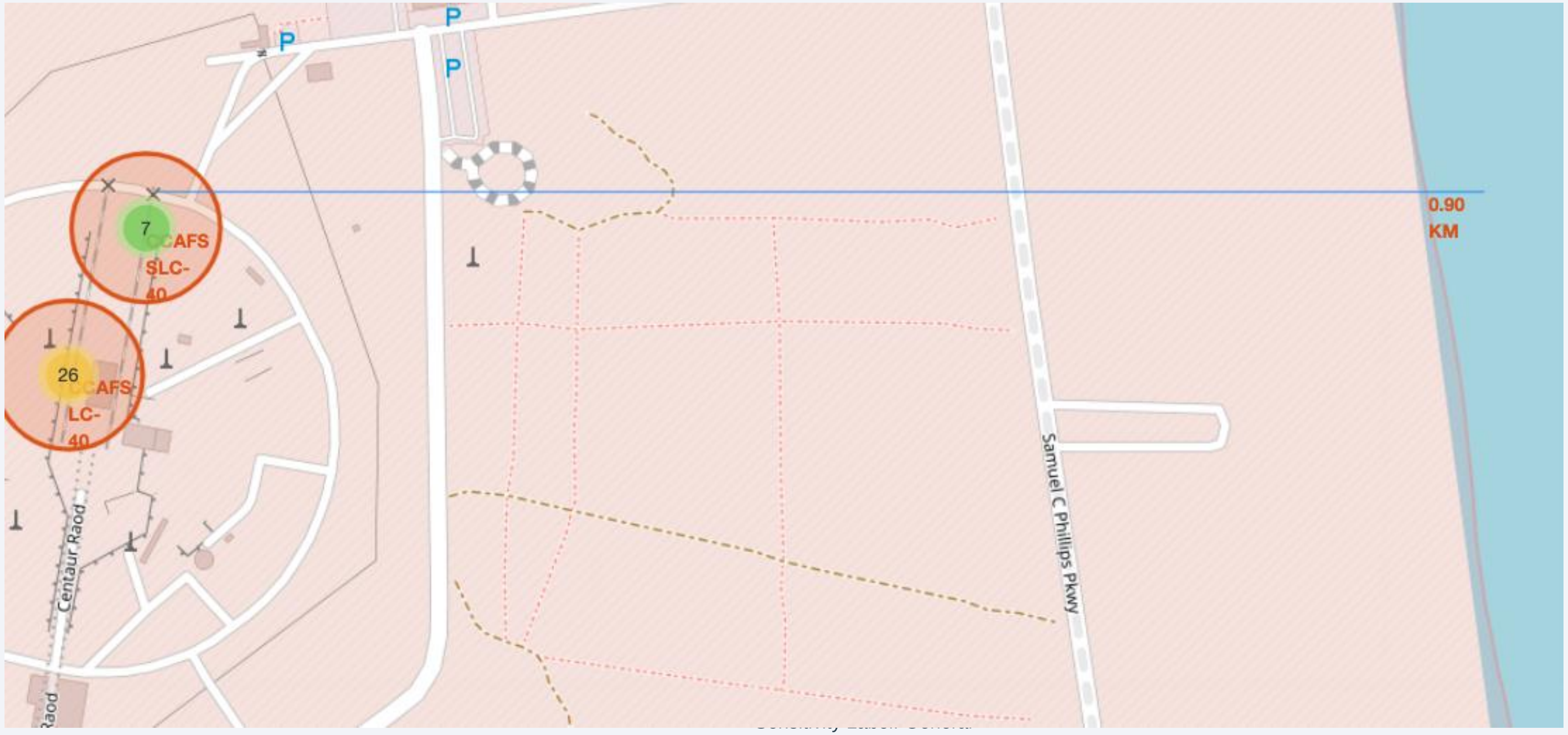
Number of Launch Sites in a location



This map show the number of Launch Site in an area

Distance from launch site to coast

This map show the distance from launch sites to coast and another location around the launch sites such as railway, highway and city



Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a pie chart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

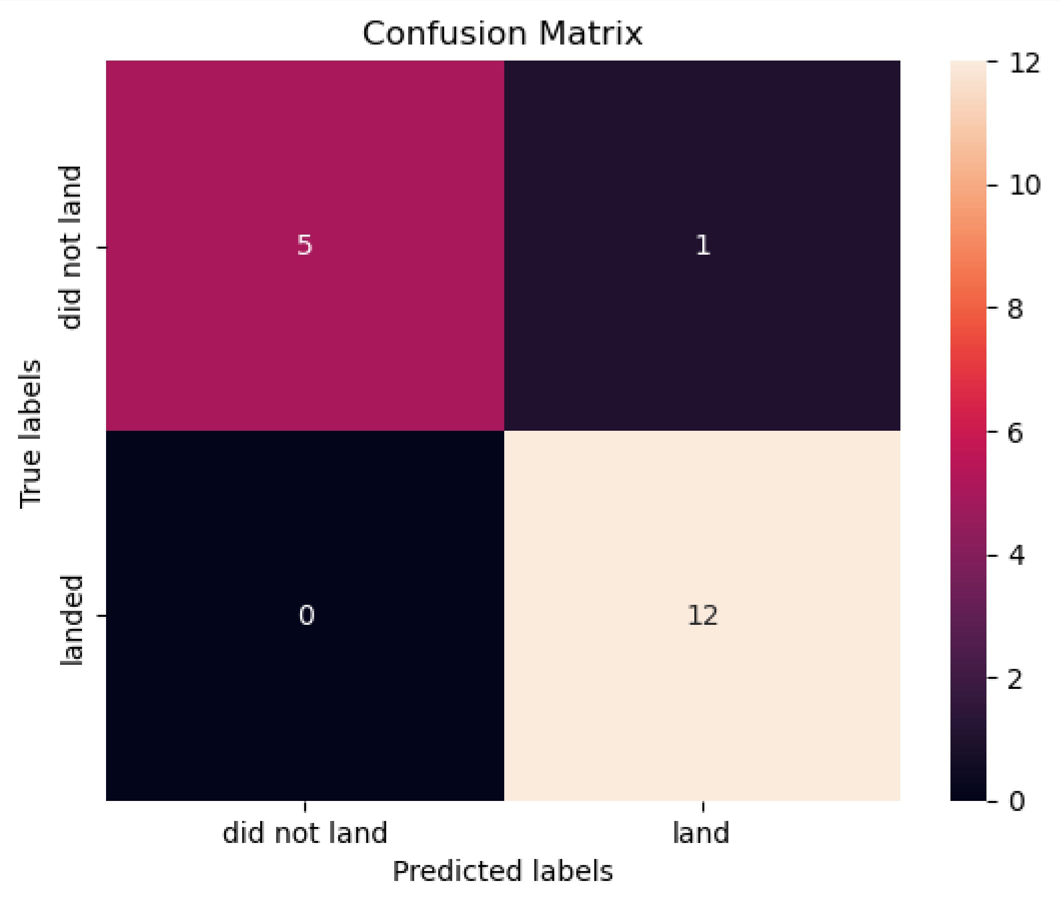
Classification Accuracy

- Decision Tree has the highest score

	Method	Score
0	LogisticRegression	0.846429
1	SVM	0.848214
2	DecistionTree	0.887500
3	KNN	0.848214

Confusion Matrix

- The model has 1 of false positive and 0 of false negative.



Conclusions

- Success rate has been increased through out the years since 2015
- GEO orbit has the highest success rate
- The highest failure is when land on drone ship
- Decision Tree is the most appropriate method in this project due to the score
- …

Appendix

- All relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets in this project can be found on the following Github repository
- https://github.com/wasakorn/capstoneproject/

Thank you!