IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rohith Muralidharan
10/09/2025

Outline

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

Executive Summary

- SpaceY is a new commercial rocket launch service provider who wants to bid against SpaceX.
- SpaceX advertises launch services starting at $62 million for missions that allow some fuel to be reserved for landing the 1st stage rocket booster, so that it can be reused.
- SpaceX public statements indicate a 1st stage Falcon 9 booster to cost upwards of $15 million to build without including R&D cost recoupment or profit margin.
- Given mission parameters such as payload mass and desired orbit, the models produced in this report were able to predict the first stage rocket booster landing successfully with an accuracy level of 83.3%.
- As a result, SpaceY will be able to make more informed bids against SpaceX by using 1st stage landing predictions as a proxy for the cost of a launch.

Introduction

- Project background and context

  - SpaceX - most successful company of the commercial space age
  - Motto is to make space travel affordable
  - Advertisement claims Falcon 9 rocket launches with a cost of 62 million dollars, on their website
  - Other providers costs a whopping 165 million dollars each
  - SpaceX claims that much of the savings is because they can reuse the first stage
  - Therefore, if we can determine if the first stage will land, we can determine the cost of a launch
  - Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

- Data collection methodology:
    - SpaceX Rest API
    - Web scraping from Wikipedia
- Perform data wrangling
    - Data filtering
    - Dealing with missing values
    - Using 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
    - Building, tuning & evaluation of models to ensure optimum results

Data Collection

- Involves a combination of API requests from SpaceX Rest API & web scraping data from Wikipedia.

  API

  - Acquired historical launch data from Open-Source Rest API for SpaceX
    - Requested and parsed the SpaceX launch data using the GET request
    - Filtered the data frame to only include Falcon 9 launches
    - Replaced missing payload mass values from classified missions with mean

  Web Scraping

  - Acquired historical launch data from Wikipedia
    - Requested the Falcon9 Launch Wiki page from its Wikipedia URL
    - Extracted all column/variable names from the HTML table header
    - Parsed the table and converted it into a Pandas data frame

Data Collection – SpaceX API

Request rocket launch data from SpaceX API

Decoding the response content using .json() and turning it into a data frame using json_normalize()

Requesting needed information about the launches from SpaceX API by applying custom functions

Constructing data we have obtained into a dictionary

Creating a data frame from the dictionary

Filtering the data frame to only include Falcon 9 launches

Replacing missing values of Payload Mass column with calculated .mean() for this column

Exporting the data to CSV

Data Collection – Scraping

Requesting Falcon 9 launch data from Wikipedia

→ Creating a BeautifulSoup object from the HTML response

→ Extracting all column names from the HTML table header

↓ Collecting the data by parsing HTML tables

↓ Constructing data we have obtained into a dictionary

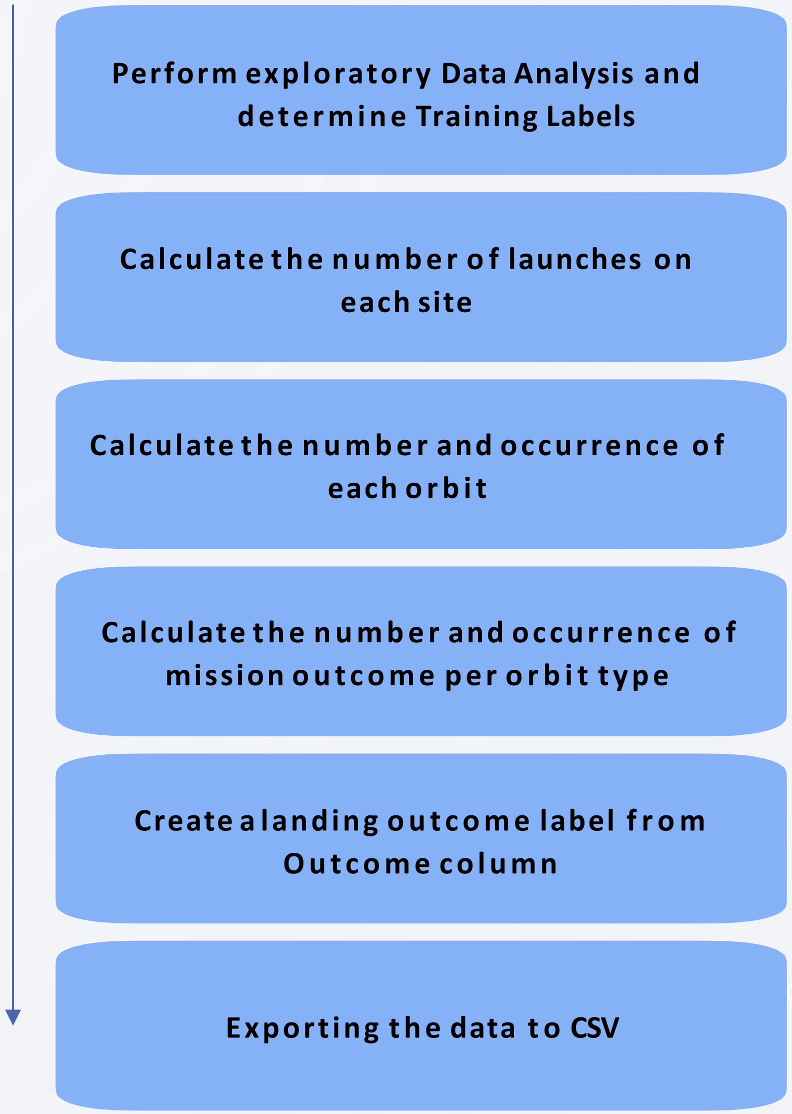
← Creating a dataframe from the dictionary

← Exporting the data to CSV

Data Wrangling

- Several cases where the booster did not land successfully is identified.
  - True Ocean:- mission outcome was successfully landed to a specific region of the ocean
  - False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean
  - True RTLS means the mission outcome was successfully landed to a ground pad
  - False RTLS means the mission outcome was unsuccessfully landed to a ground pad
  - True ASDS means the mission outcome was successfully landed on a drone ship
  - False ASDS means the mission outcome was unsuccessfully landed on a drone ship

- Convert those outcomes into Training Labels with “1” means the booster successfully landed, “0” means it was unsuccessful



EDA with Data Visualization

- Charts were plotted:
  - Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Trend

- Scatter plots show the relationship between variables
- Bar charts show comparisons among discrete categories
- Line charts show trends in data over time (time series)

EDA with SQL

Performed SQL queries:

- Displaying the names of the unique launch sites
- Displaying 5 records where launch sites begin with the string 'CCA'
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass
- Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015
- Ranking 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

Build an Interactive Map with Folium

- Markers of all Launch Sites:
  - Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location
  - Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts
- Colored Markers of the launch outcomes for each Launch Site:
  - Added colored Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates
- Distances between a Launch Site to its proximities:
  - Added coloured Lines to show distances between the Launch Site KSC LC-39A (as an example) and its proximities like Railway, Highway, Coastline and Closest City

Build a Dashboard with Plotly Dash

Dropdown List for Launch Sites:

- Added a dropdown list to enable Launch Site selection

Pie Chart showing Success Launches (All Sites/Certain Site):

- Added a pie chart to show the total successful launches count for all sites and the Success vs. Failed counts for the site, if a specific Launch Site was selected.

Slider of Payload Mass Range:

- Added a slider to select Payload range.

Scatter Chart of Payload Mass vs. Success Rate for the different Booster Versions:

- Added a scatter chart to show the correlation between Payload and Launch Success.

Predictive Analysis (Classification)

Creating a NumPy array from the column "Class" in data

Standardizing the data with Standard Scaler, then fitting and transforming it

Splitting the data into training and testing sets with train_test_split function

Creating a GridSearchCV object with cv =10 to find the best parameters

Applying GridSearchCV on LogReg, SVM, Decision Tree, and KNN models

Calculating the accuracy on the test data using the method .score() for all models

Examining the confusion matrix for all models

Finding the method performs best by examining the Jaccard_score and F1_score metrics

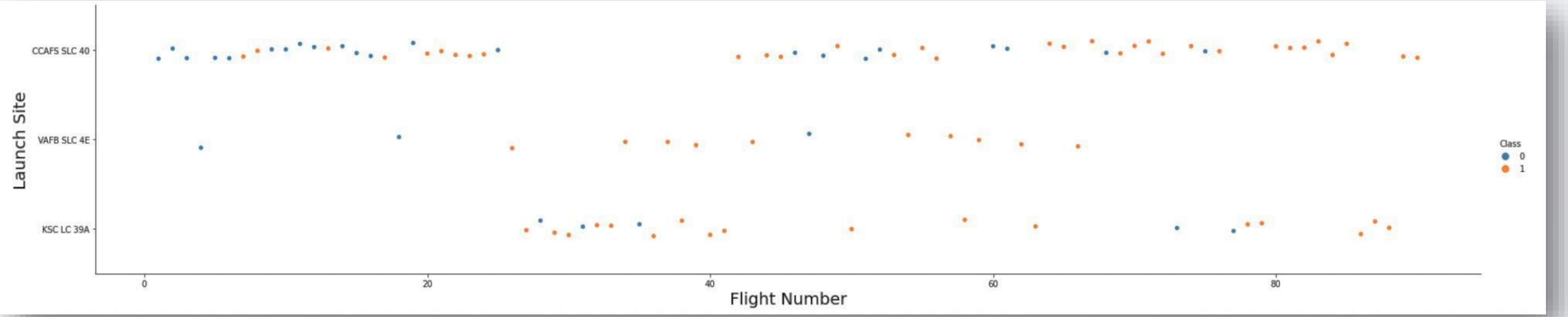
Results

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

Section 2

Insights drawn from EDA

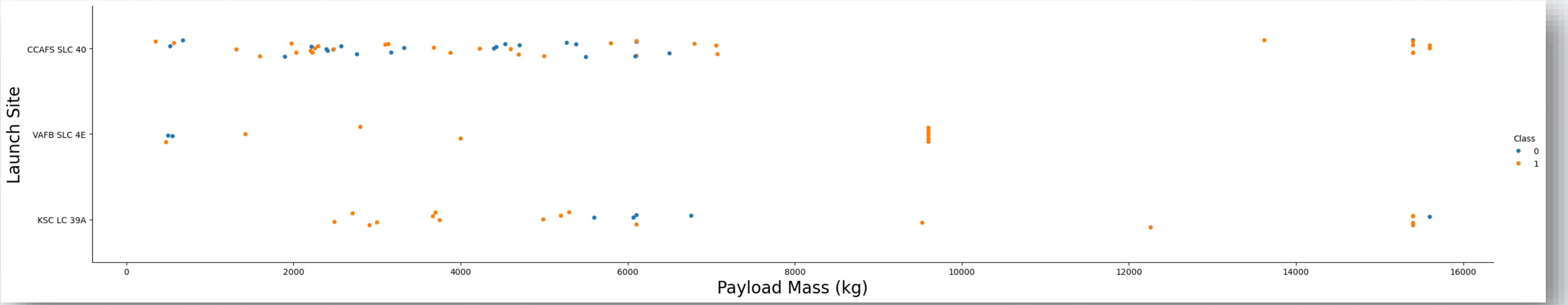
Flight Number vs. Launch Site



Explanation:

- All the earliest flights failed & all the latest flights succeeded
- Half of all launches took place at the CCAFS SLC 40 launch site
- KSC LC 39A & VAFB SLC 4E have higher success rates
- Assumption: Success rate gets higher with each new launch

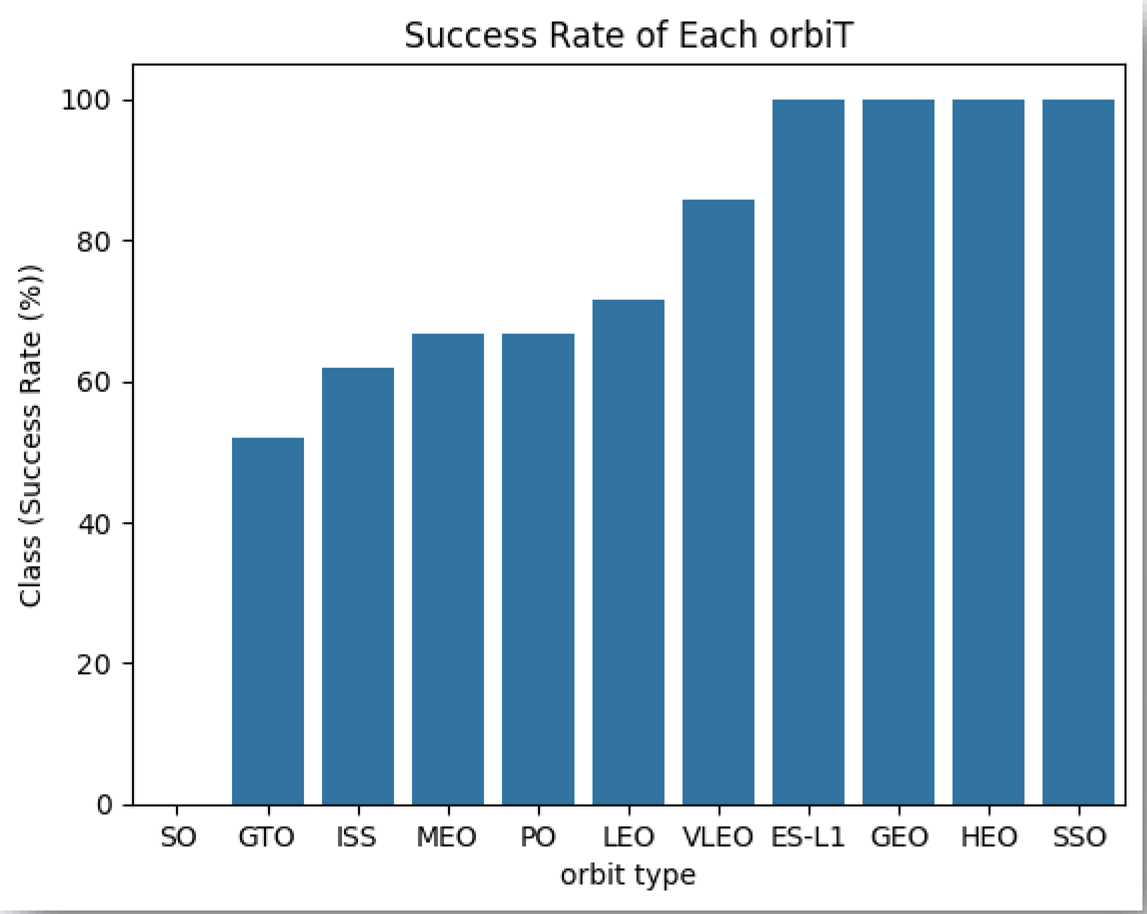
Payload vs. Launch Site



Explanation:

- For every launch site the higher the payload mass, the higher the success rate.
- Most of the launches with payload mass over 7000 kg were successful.
- KSC LC 39A has a 100% success rate for payload mass under 5500 kg too

Success Rate vs. Orbit Type



- Orbits with 0% success rate:

  SO

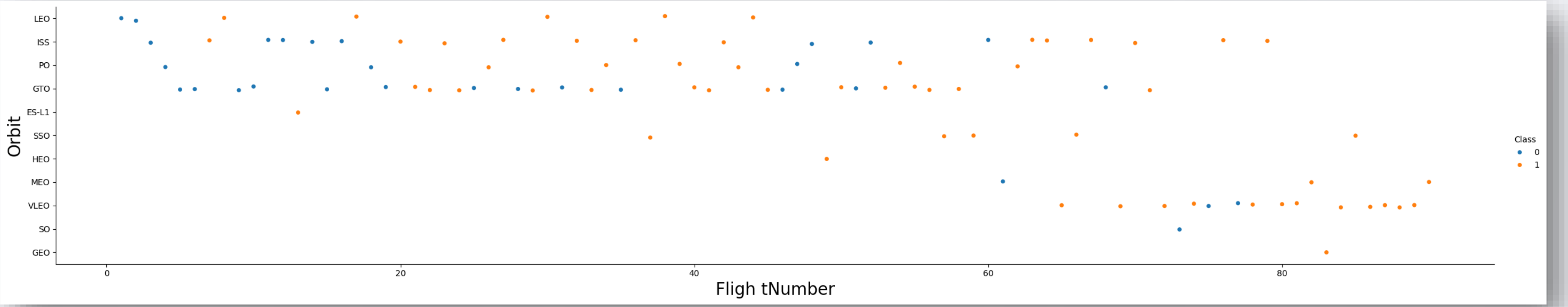
- Orbits with success rate between 50% and 85%:

  GTO, ISS, LEO, MEO, PO

- Orbits with 100% success rate:

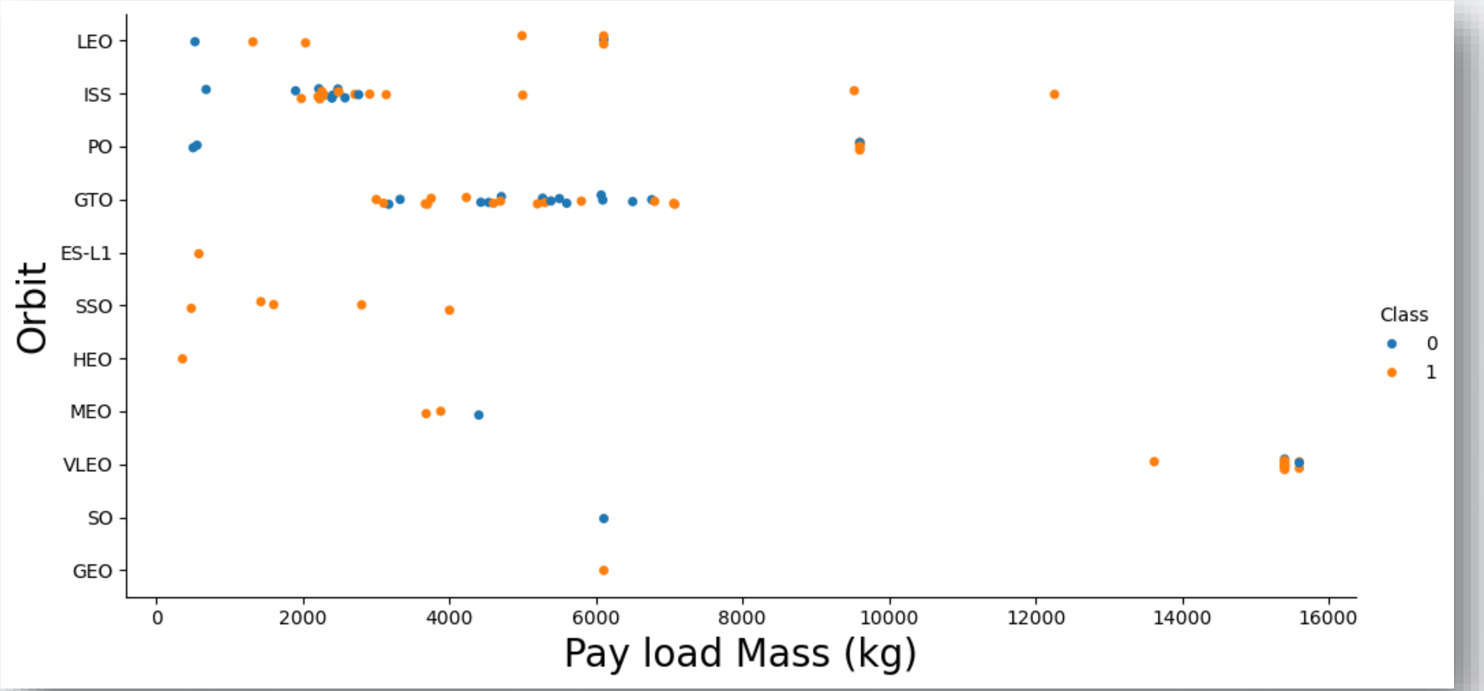
  ES-L1, GEO, HEO, SSO

Flight Number vs. Orbit Type



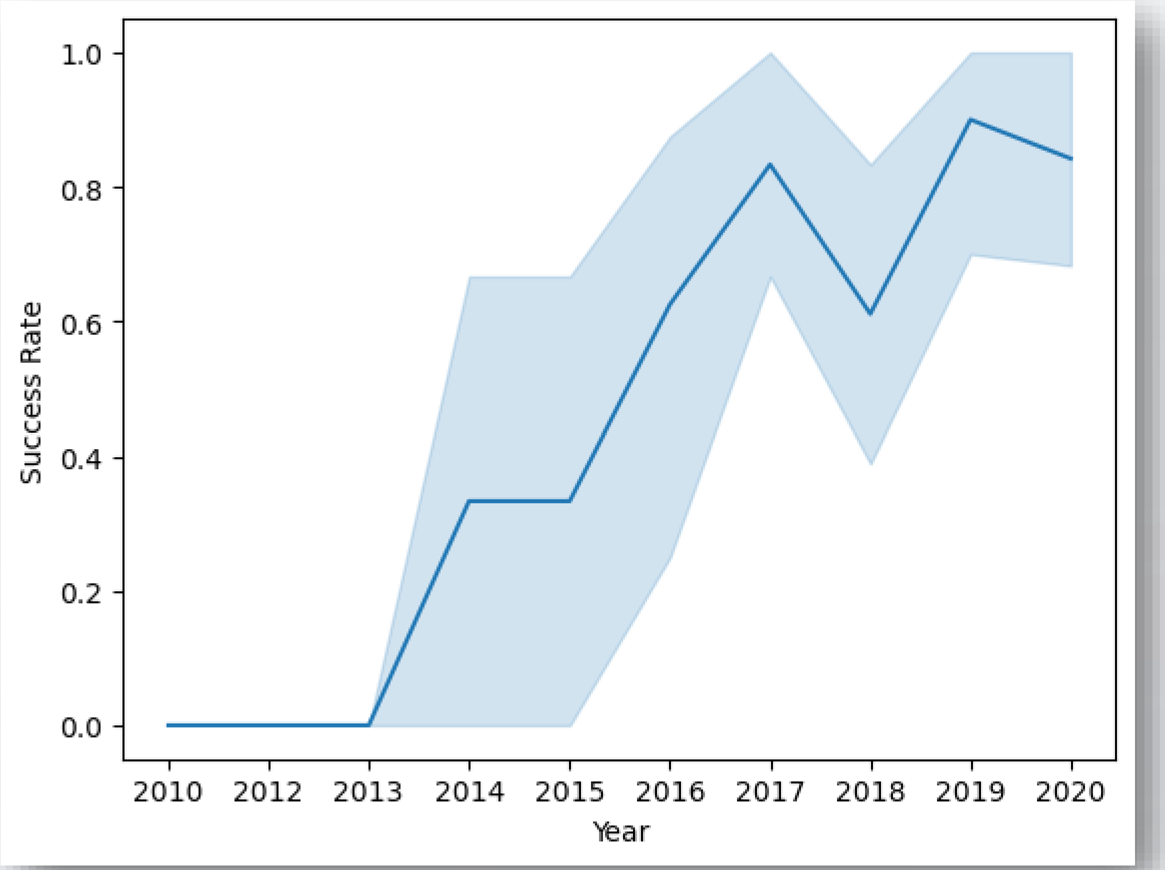
- Success is related to the number of flights in LEO orbit
- No conclusive relationship between flight number in GTO orbit

Payload vs. Orbit Type



- Heavy payloads have positive on GTO and Polar LEO (ISS) orbits
- Having negative influence on GTO orbits

Launch Success Yearly Trend



- 2010-2013:
  - No success in this period
- 2013-2014:
  - Exponential increase in success rate
- 2014-2015:
  - No change in success rate
- 2015-2017:
  - Exponential increase similar to 2013-2014
- 2017-2018:
  - Sharp dip in success rate
- 2018-2019:
  - Get back to the success rate slightly above to 2017
- 2019-2020:
  - No major change in the success rate

All Launch Site Names

```
In [39]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;

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

Out[39]:

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

Displaying the names of the unique launch sites

Launch Site Names Begin with 'CCA'

In [40]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

Out[40]:

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

Displaying 5 records where launch sites begin with 'CCA'

Total Payload Mass

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

In [41]:
```
%sql SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';
```

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

Out[41]:

Total_Payload_Mass
45596

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

Average Payload Mass by F9 v1.1

In [42]:

```
%sql SELECT AVG("Payload_Mass__kg_") AS Avg_Payload_Mass FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

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

Out[42]:

Avg_Payload_Mass
2928.4

Displaying average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

In [43]:

```
%sql SELECT MIN("Date") AS First_Successful_Ground_Pad_Landing FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground
```

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

Out[43]:

First_Successful_Ground_Pad_Landing
2015-12-22

Listing the date when the first successful landing outcome in ground pad was achieved

Successful Drone Ship Landing with Payload between 4000 and 6000

```
In [44]: %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Mission_Outcome" = 'Success' AND "PAYLOAD_MASS__KG_" > 4000 A

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

Out[44]:

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

```
In [45]: %sql SELECT "Mission_Outcome", COUNT(*) AS total_number FROM SPACEXTABLE GROUP BY "Mission_Outcom";

 * sqlite:///my_data1.db
Done.
Out[45]:
```

Mission_Outcome	total_number
Success	101

Listing the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

In [46]:

```
%sql SELECT "Booster_Version", "Payload_Mass" FROM SPACEXTABLE WHERE "Payload_Mass" = (SELECT MAX("Payload_Mass") FROM SPACE
```

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

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
F9 v1.1
F9 v1.1
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

Listing the names of the booster versions which have carried the maximum payload mass

2015 Launch Records

In [47]:

```
%sql SELECT strftime('%m', Date) AS month_number, strftime('%Y-%m-%d', Date) AS Date, Booster_Version, Launch_Site, Landing_
```

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

Out[47]:

month_number	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015

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

In [48]:

```
%sql SELECT Landing_Outcome, COUNT(*) AS Landing_Outcomes FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
```

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

Out[48]:

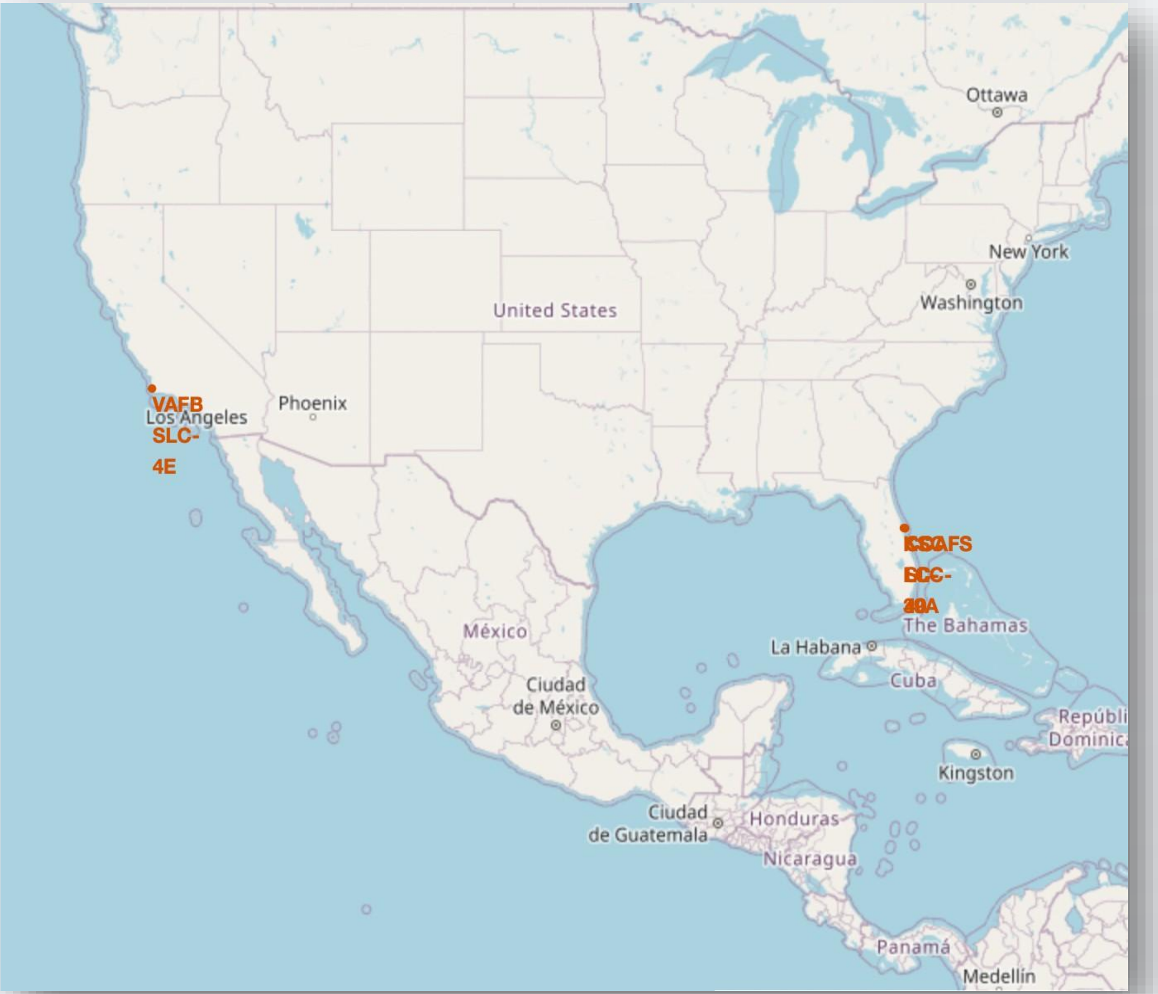
Landing_Outcome	Landing_Outcomes
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Ranking 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

Section 3

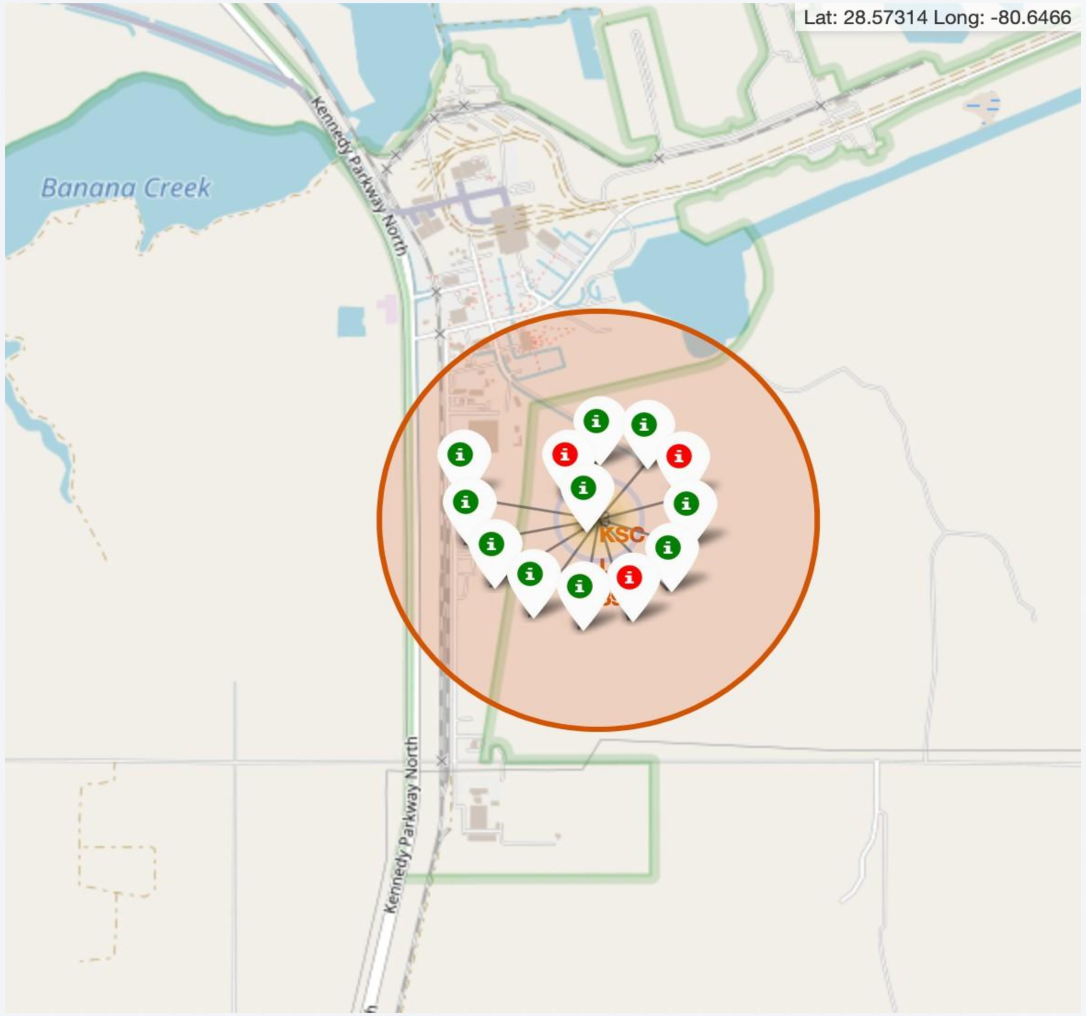
Launch Sites Proximities Analysis

All Launch site's location Markers on Global map



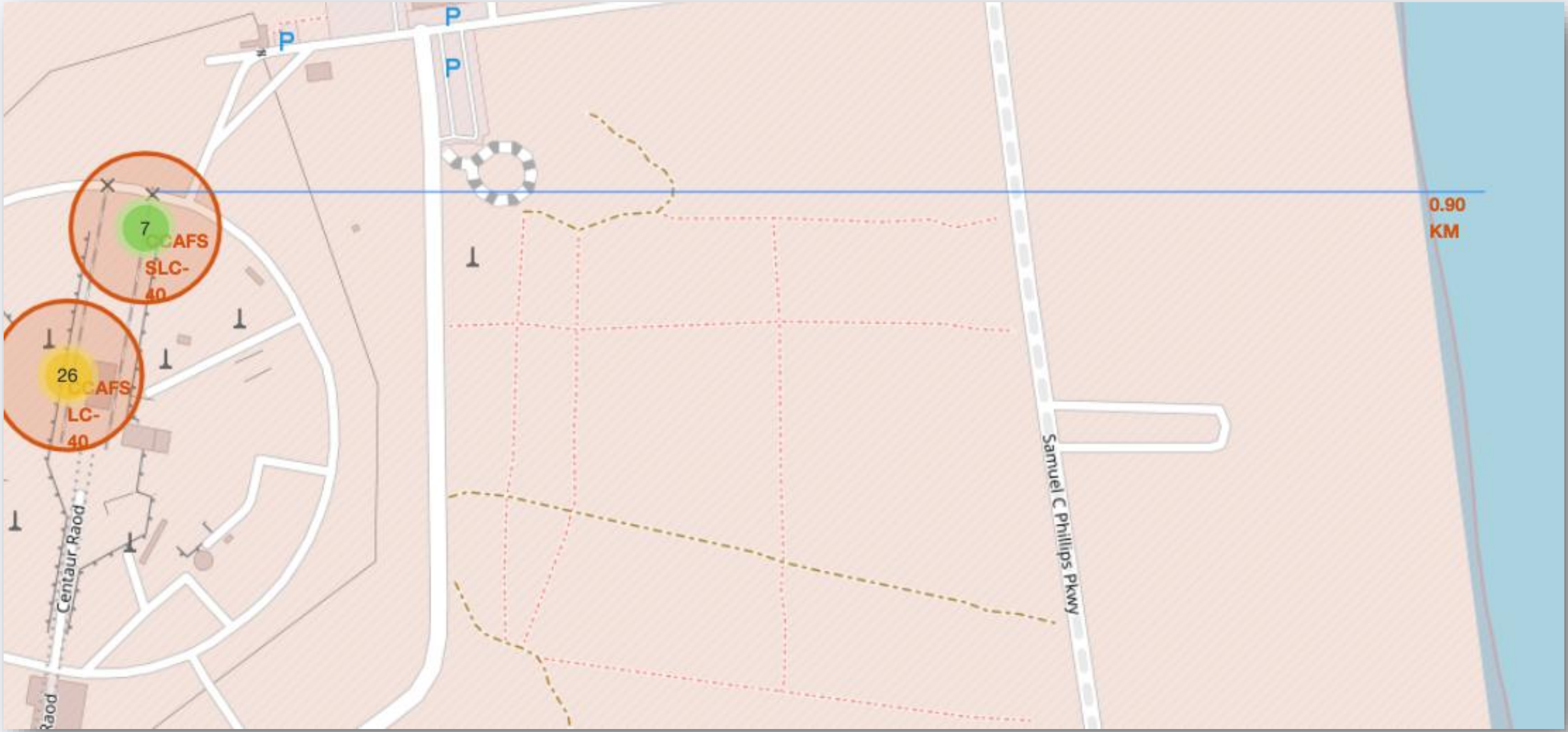
- Most of Launch sites are in proximity to the Equator line
- The land is moving faster at the equator(1670 km/hour)
- Accounting theory of inertia, this speed will help the launched spacecraft keep up a good enough speed to stay in orbit
- Also launch sites are very close to the coast
- Launching rockets towards the ocean minimizes the risk of having any debris dropping or exploding near people

Color Labeled Markers



- Using color labeled markers we can easily identify which launch sites have relatively high success rates
  - ✓ Green Marker = Successful Launch
  - ✓ Red Marker = Failed Launch
- Launch Site KSC LC-39A is having a very high Success Rate.

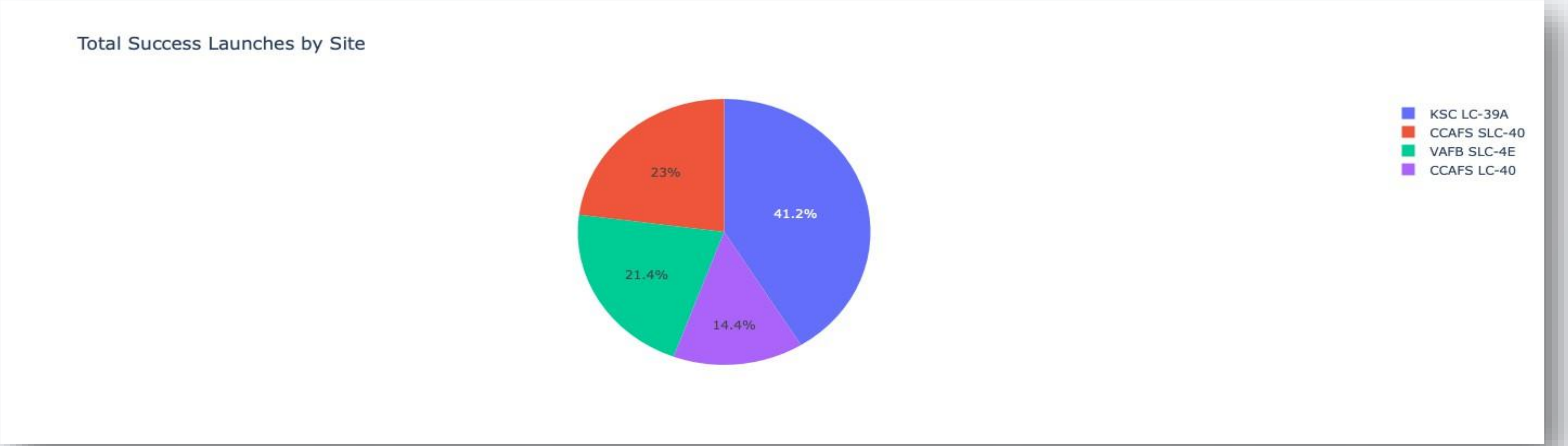
Distance between Launch site and proximities



Section 4

Build a Dashboard with Plotly Dash

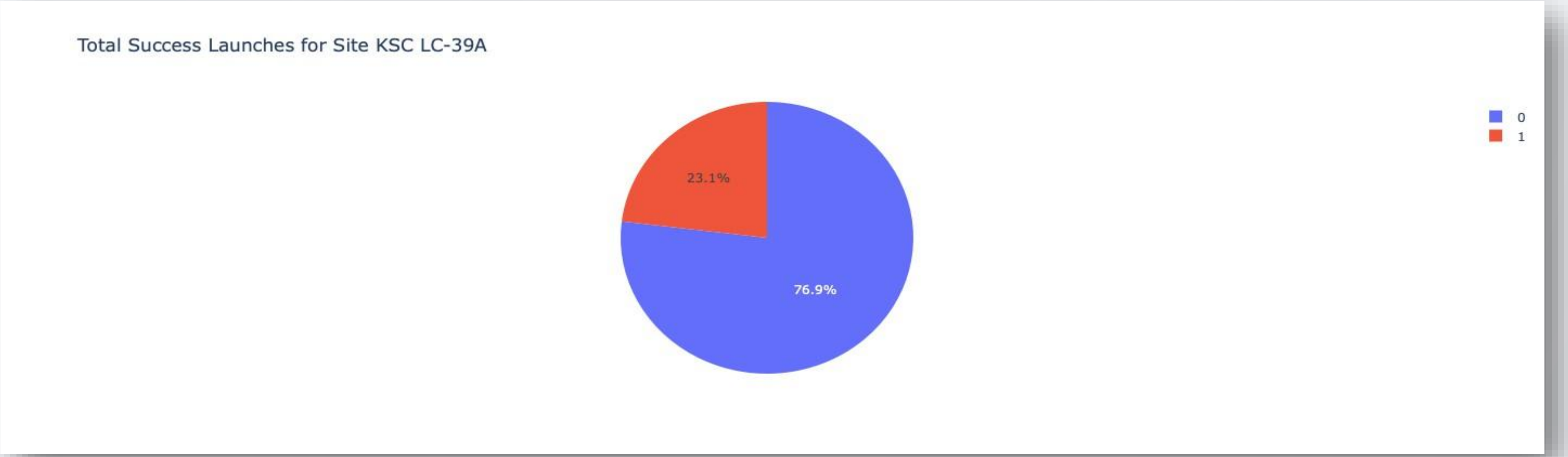
Total success launches by site



Findings:

- KSC LC-39A has the most successful launches
- CCAFS LC-40 is having the least successful launches

Launch site with highest launch success ratio



- 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

Payload Mass vs Launch Outcome of all sites

- The highest success rate is for the payloads between 2000 and 5500 kg

Section 5

Predictive Analysis (Classification)

Classification Accuracy

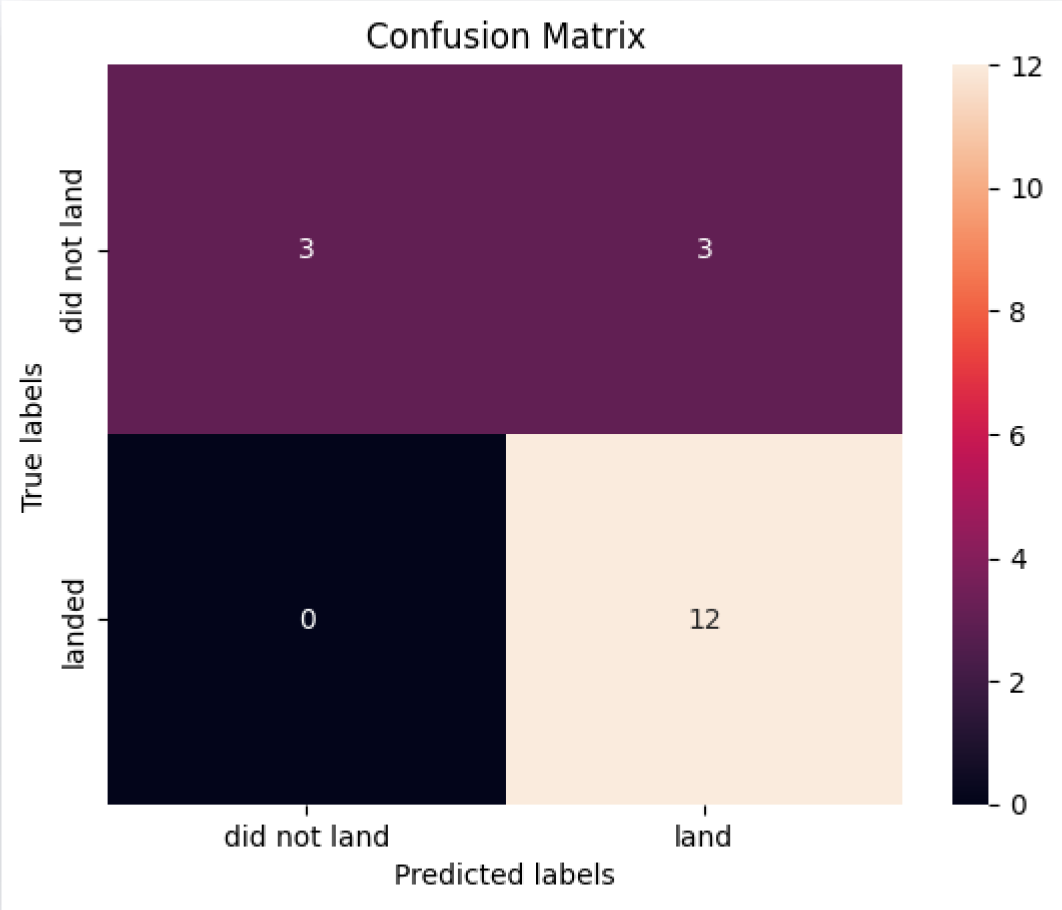
Scores & Accuracy of the entire dataset

Out[38]:

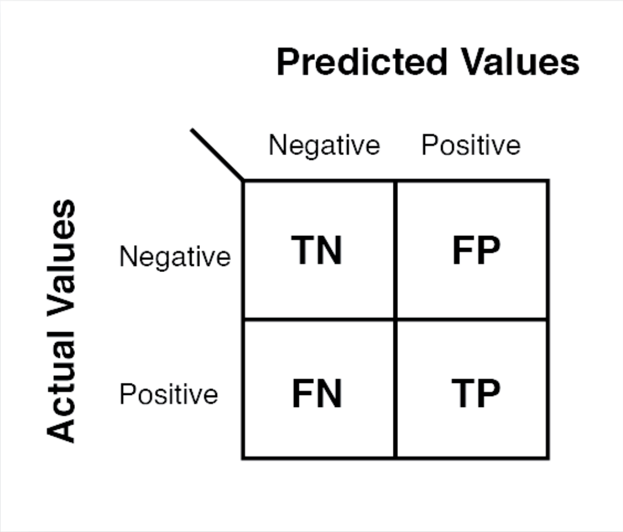
	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.882353	0.819444
F1_Score	0.909091	0.916031	0.937500	0.900763
Accuracy	0.866667	0.877778	0.911111	0.855556

- We can not confirm which method performs best, based on the scores of the Test Set
- The entire dataset score confirms that the best model is the ***Decision Tree Model***
- Model not only have higher scores, but also higher accuracy

Confusion Matrix



- LogReg can be distinguished based on different classes
- The major problem resides is false positives



Conclusions

- The best algorithm for this dataset is Decision Tree Model
- Most of launch sites are in proximity to the Equator line and very close to the coast
- KSC LC-39A has the highest success rate of the launches
- Orbits namely ES-L1, GEO, HEO and SSO have 100% of success rate
- The success rate of launches increased over the years
- Low payload mass have better results when compared to launches with larger payload mass

Appendix

- **Notebooks to recreate dataset, analysis, and models:**

  - https://github.com/Rogueit805/IBM-Applied-Data-Science-Capstone/blob/ab60d54dfb03eb04471ca29c6ac4f4068444f023/Data%20Collection%20API.ipynb
  - https://github.com/Rogueit805/IBM-Applied-Data-Science-Capstone/blob/ab60d54dfb03eb04471ca29c6ac4f4068444f023/Data%20Collection%20with%20Web%20Scraping.ipynb
  - https://github.com/Rogueit805/IBM-Applied-Data-Science-Capstone/blob/ab60d54dfb03eb04471ca29c6ac4f4068444f023/Data%20Wrangling.ipynb
  - https://github.com/Rogueit805/IBM-Applied-Data-Science-Capstone/blob/ab60d54dfb03eb04471ca29c6ac4f4068444f023/EDA%20with%20Data%20Visualization.ipynb
  - https://github.com/Rogueit805/IBM-Applied-Data-Science-Capstone/blob/ab60d54dfb03eb04471ca29c6ac4f4068444f023/EDA%20with%20SQL.ipynb
  - https://github.com/Rogueit805/IBM-Applied-Data-Science-Capstone/blob/ab60d54dfb03eb04471ca29c6ac4f4068444f023/Interactive%20Visual%20Analytics%20with%20Folium.ipynb
  - https://github.com/Rogueit805/IBM-Applied-Data-Science-Capstone/blob/ab60d54dfb03eb04471ca29c6ac4f4068444f023/spacex-dash-app.py
  - https://github.com/Rogueit805/IBM-Applied-Data-Science-Capstone/blob/ab60d54dfb03eb04471ca29c6ac4f4068444f023/Machine%20Learning%20Prediction.ipynb

Appendix

- **Acknowledgments**
  - **Primary Instructors:** Joseph Santarcangelo, Yan Luo
  - **Other Contributors & Staff**
  - **Project Lead:** Rav Ahuja, **Instructional Designer:** Lakshmi Holla, **Lab Authors:** Joseph Santarcangelo, Yan Luo, Azim Hirjani, Lakshmi Holla **Technical Advisor:** Yan Luo
  - **Production Team**
  - **Publishing:** Grace Barker, Rachael Jones **Project Coordinators:** Kathleen Bergner **Narration:** Bella West **Video Production:** Simer Preet, Lauren Hall, Hunter Bay, Tanya Singh, Om Singh
  - **Teaching Assistants and Forum Moderators:** Malika Singla, Duvvana Mrutyunjaya Naidu, Lakshmi Holla, Anita Verma

- **References**

  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
  - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv

Thank you!