# Winning Space Race with Data Science

Juan Manuel Fariñas
April 2023

# Outline

# Executive Summary

- Summary of methodologies
  - Data Collection through API and Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

# Introduction

- Project background and context

- 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.

- Problems you want to find answers

- Predict if the Falcon 9 first stage will land successfully.

- The best place to make launches

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX API and web scraping from Wikipedia.
- Perform data wrangling
  - One-hot encoding was applied to categorical features
- 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 was collected from Space X API and web scraping Wikipedia.

- Api flow, we get information about launches from Space X:

FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

- Web scraping flow, we get more information about launches from Space X

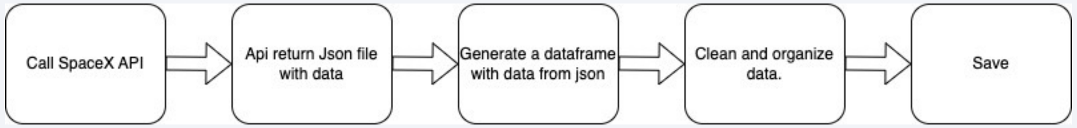
Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

# Data Collection – SpaceX API

- Space X have an API where we can ask for data about launches.

- The Space X REST API URL is api.spacexdata.com/v4/

- Complete source code is in notebook [w1_01 - data-collection-api.ipynb]

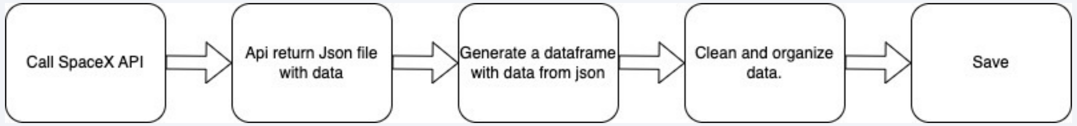


# Data Collection - Scraping

- We collected information through wikipedia.

- Wikipedia link is s https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

- Complete source code is in notebook w1_02 - data collection-webscraping.ipynb



# Data Wrangling

- Identify launches per site.
- Calculate occurrences per orbit.
- Calculate the number and occurence of mission outcome per orbit type.
- Create a landing outcome label

Using the Outcome column, we create a list where the element is zero if is bad_outcome or otherwise.

- Complete source code is in notebook [w1_03 - data-wrangling.ipynb]

# EDA with Data Visualization

- Scatter Graph.

We use this type of graph to visualize: Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload vs. Launch Site, Orbit vs. Flight Number, Payload vs. Orbit Type and Orbit vs. Payload Mass.

- Bar Graph.

We use this type of graph to visualize: Success rate vs. Orbit

- Line Graph.

We use this type of graph to visualize: Success rate vs. Year

- Complete source code is in notebook [w2_02 - eda-dataviz.ipynb]

# EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

Displaying 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 the names of the booster_versions which have carried the maximum payload mass.

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for 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

- Complete source code is in notebook [w2_01 - eda-sql-coursera.ipynb]

# Build an Interactive Map with Folium

- Folium maps mark launch sites, successful and unsuccessful landings, and a proximity to key locations like railway, highway, coast and airport.
- We add this items to understand and view the information about launch sites, where they are and the surrounding items. We can see successful and unsuccessful landing in each site.
- Complete source code is in notebook w3_01 - launch-site-location.ipynb

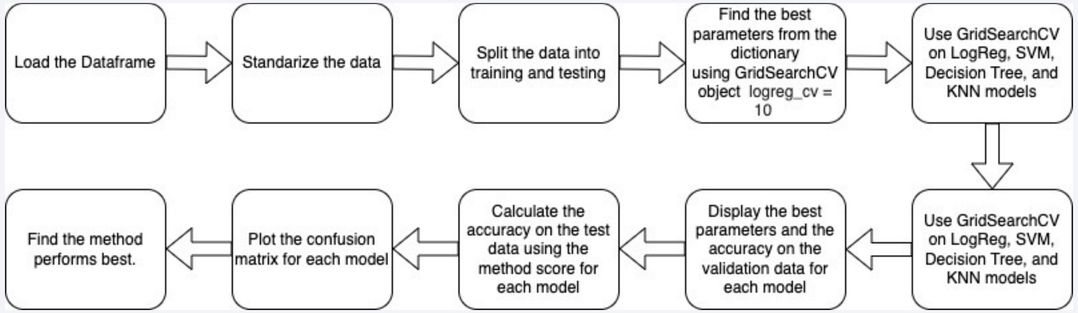
# Build a Dashboard with Plotly Dash

- The Dashboard has dropdown with all-sites, pie chart with success launches, range slider to select payload and scatter plot components with a correlation between payload and success.
- Dropdown allows a user to choose the launch site or all launch sites.
- Pie chart shows the total success and the total failure for the launch site chosen with the dropdown component.
- Range slider allows a user to select a payload mass in a fixed range.
- Scatter chart shows the relationship between two variables, in particular Payload Mass and Success.

- Complete source code is in file [spacex_dash_app.py]()

# Predictive Analysis (Classification)

- We load, normalize and split data into train and test sets.
- We train and evaluate each used models: LogReg, SVM, Decision Tree and KNN.
- We choose the best model.




- Complete source code is in notebook [w4_01 - machine-learning-prediction-part-5.ipynb]

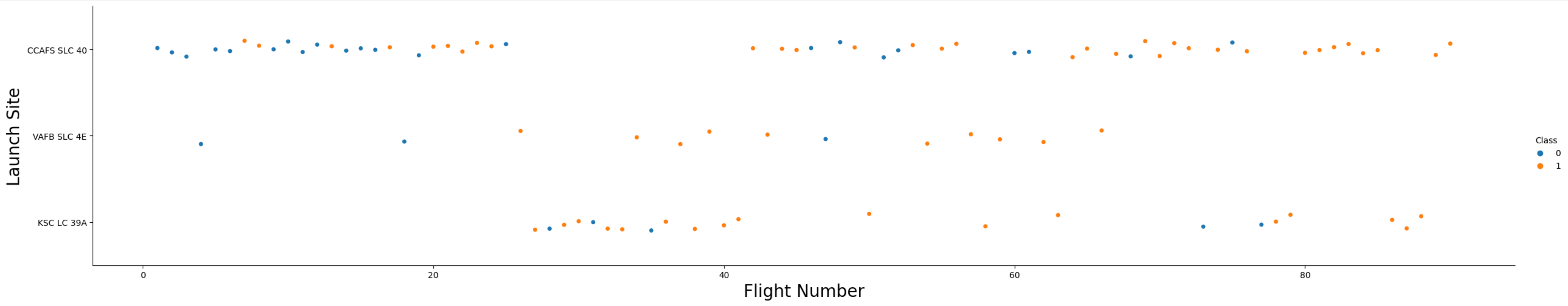
# Results

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

Section 2

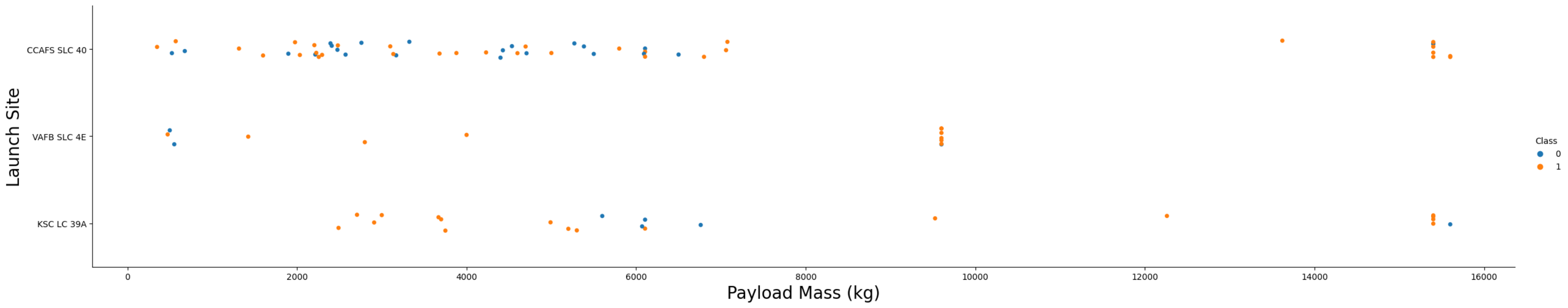
# Insights drawn from EDA

# Flight Number vs. Launch Site



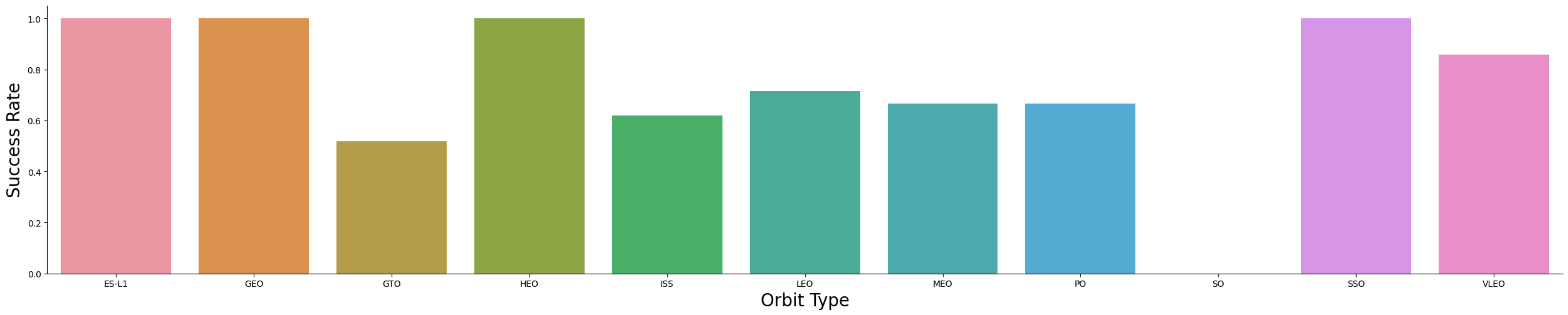

- Shows the flights number and where is launched.

# Payload vs. Launch Site



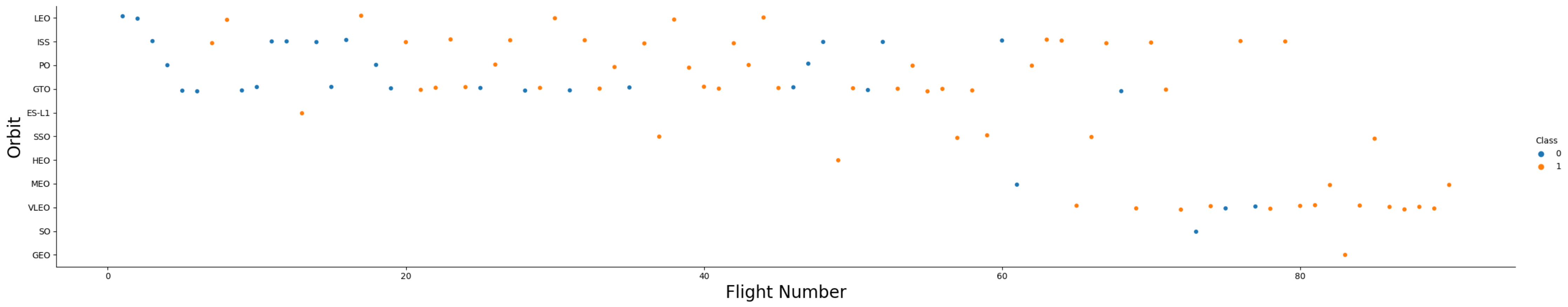

- Show the the payload mass and where are launched

# Success Rate vs. Orbit Type



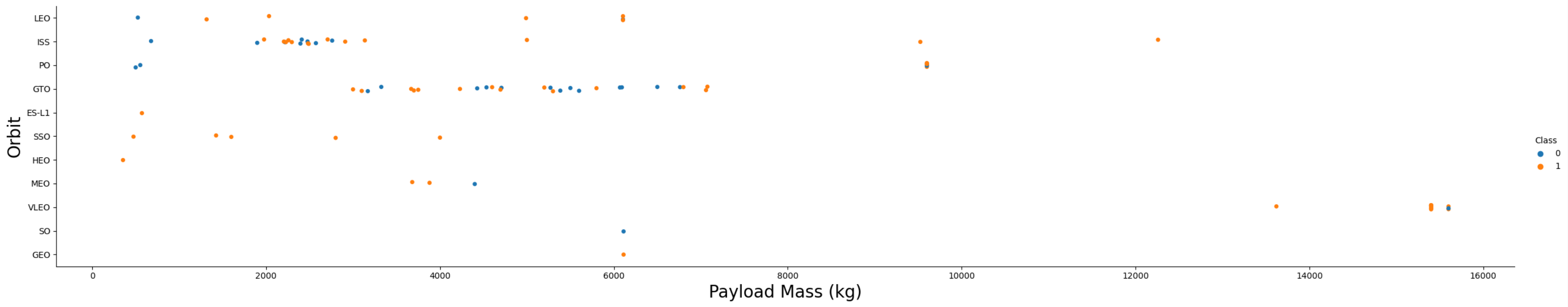

- Show the orbit type and success rate % of launches

# Flight Number vs. Orbit Type



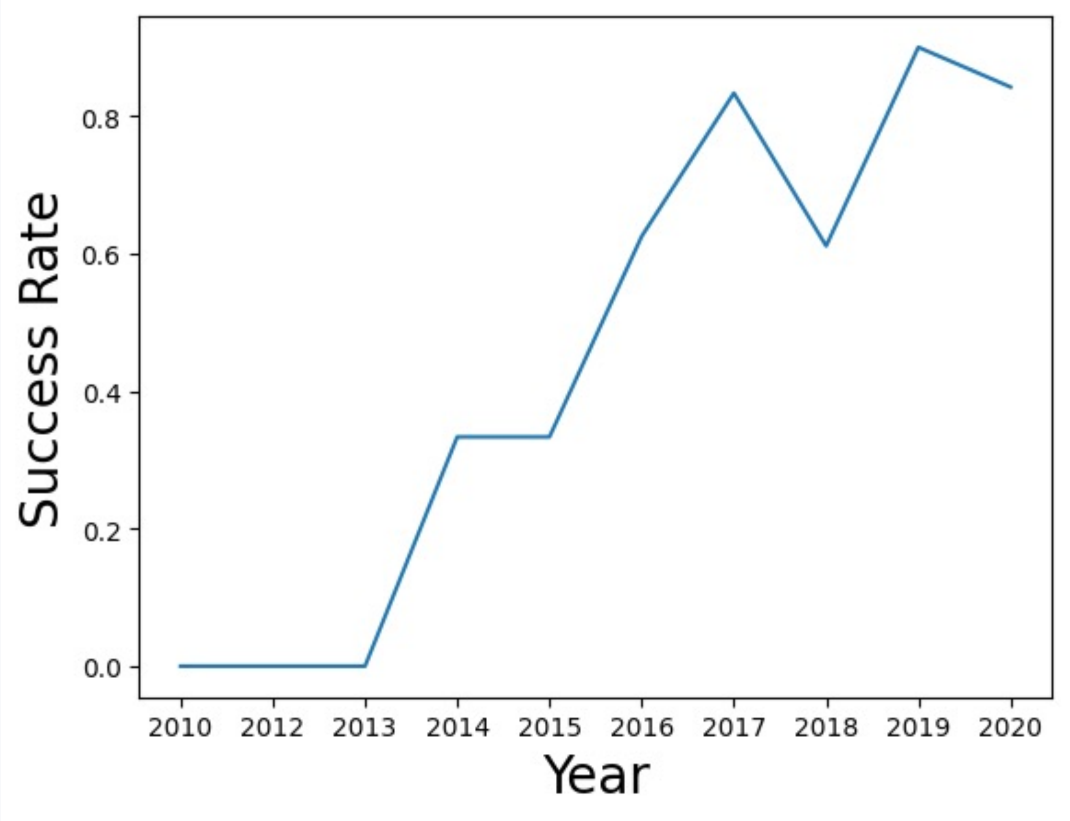

- Show the flight number and orbit type.

# Payload vs. Orbit Type




- Show the payload mass and orbit type.

# Launch Success Yearly Trend




- Show the launch success rate through the years.

# All Launch Site Names

- SQL Query:

```
SELECT DISTINCT Launch_Site FROM `e-quanta-317101.spacex.spacex` LIMIT 1000
```

- Result:

| | Launch_Site |
|---|---|
| 0 | KSC LC-39A |
| 1 | CCAFS LC-40 |
| 2 | VAFB SLC-4E |
| 3 | CCAFS SLC-40 |

# Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
SELECT * FROM `e-quanta-317101.spacex.spacex` WHERE Launch_Site like 'CCA%'
LIMIT 5
```

- Result

| | Date | Time__UTC_ | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing__Outcome |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2013-12-03 | 22:41:00 | F9 v1.1 | CCAFS LC-40 | SES-8 | 3170 | GTO | SES | Success | No attempt |
| 1 | 2014-01-06 | 22:06:00 | F9 v1.1 | CCAFS LC-40 | Thaicom 6 | 3325 | GTO | Thaicom | Success | No attempt |
| 2 | 2014-08-05 | 08:00:00 | F9 v1.1 | CCAFS LC-40 | AsiaSat 8 | 4535 | GTO | AsiaSat | Success | No attempt |
| 3 | 2014-09-07 | 05:00:00 | F9 v1.1 B1011 | CCAFS LC-40 | AsiaSat 6 | 4428 | GTO | AsiaSat | Success | No attempt |
| 4 | 2015-03-02 | 03:50:00 | F9 v1.1 B1014 | CCAFS LC-40 | ABS-3A Eutelsat 115 West B | 4159 | GTO | ABS Eutelsat | Success | No attempt |

# Total Payload Mass

- Calculate the total payload carried by boosters from NASA:

```
SELECT SUM (PAYLOAD_MASS__KG_) FROM `e-quanta-317101.spacex.spacex` WHERE
Customer='NASA (CRS)'
```

- Result:

| | f0_ |
|---|---|
| 0 | 45596 |

# Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

```
SELECT AVG (PAYLOAD_MASS__KG_) FROM `e-quanta-317101.spacex.spacex` WHERE
Booster_Version like 'F9 v1.1%'
```

- Result:

| | f0_ |
|---|---|
| 0 | 2534.666667 |

# First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
SELECT * FROM `e-quanta-317101.spacex.spacex` WHERE Date = (SELECT MIN (Date)
FROM `e-quanta-317101.spacex.spacex` WHERE Landing__Outcome = 'Success (ground
pad)')
```

- Result:

| | Date | Time__UTC_ | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing__Outcome |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2015-12-22 | 01:29:00 | F9 FT B1019 | CCAFS LC-40 | OG2 Mission 2 11 Orbcomm-OG2 satellites | 2034 | LEO | Orbcomm | Success | Success (ground pad) |

# Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

```
SELECT * FROM `e-quanta-317101.spacex.spacex` WHERE Landing__Outcome like
'Success%' AND PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000
```

- Result:

| | Date | Time__UTC_ | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing__Outcome |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2017-03-30 | 22:27:00 | F9 FT B1021.2 | KSC LC-39A | SES-10 | 5300 | GTO | SES | Success | Success (drone ship) |
| 1 | 2017-10-11 | 22:53:00 | F9 FT B1031.2 | KSC LC-39A | SES-11 / EchoStar 105 | 5200 | GTO | SES EchoStar | Success | Success (drone ship) |
| 2 | 2018-11-15 | 20:46:00 | F9 B5 B1047.2 | KSC LC-39A | Es hail 2 | 5300 | GTO | Es hailSat | Success | Success |
| 3 | 2017-05-01 | 11:15:00 | F9 FT B1032.1 | KSC LC-39A | NROL-76 | 5300 | LEO | NRO | Success | Success (ground pad) |
| 4 | 2017-09-07 | 14:00:00 | F9 B4 B1040.1 | KSC LC-39A | Boeing X-37B OTV-5 | 4990 | LEO | U.S. Air Force | Success | Success (ground pad) |
| 5 | 2016-05-06 | 05:21:00 | F9 FT B1022 | CCAFS LC-40 | JCSAT-14 | 4696 | GTO | SKY Perfect JSAT Group | Success | Success (drone ship) |
| 6 | 2016-08-14 | 05:26:00 | F9 FT B1026 | CCAFS LC-40 | JCSAT-16 | 4600 | GTO | SKY Perfect JSAT Group | Success | Success (drone ship) |
| 7 | 2019-06-12 | 14:17:00 | F9 B5 B1051.2 | VAFB SLC-4E | RADARSAT Constellation, SpaceX CRS-18 | 4200 | SSO | Canadian Space Agency (CSA) | Success | Success |
| 8 | 2018-08-07 | 05:18:00 | F9 B5 B1046.2 | CCAFS SLC-40 | Merah Putih | 5800 | GTO | Telkom Indonesia | Success | Success |
| 9 | 2019-02-22 | 01:45:00 | F9 B5 B1048.3 | CCAFS SLC-40 | Nusantara Satu, Beresheet Moon lander, S5 | 4850 | GTO | PSN, SpaceIL / IAI | Success | Success |
| 10 | 2020-07-20 | 21:30:00 | F9 B5 B1058.2 | CCAFS SLC-40 | ANASIS-II, Starlink 9 v1.0 | 5500 | GTO | Republic of Korea Army, Spaceflight Industries... | Success | Success |
| 11 | 2018-01-08 | 01:00:00 | F9 B4 B1043.1 | CCAFS SLC-40 | Zuma | 5000 | LEO | Northrop Grumman | Success (payload status unclear) | Success (ground pad) |
| 12 | 2020-06-30 | 20:10:46 | F9 B5B1060.1 | CCAFS SLC-40 | GPS III-03, ANASIS-II | 4311 | MEO | U.S. Space Force | Success | Success |
| 13 | 2020-11-05 | 23:24:23 | F9 B5B1062.1 | CCAFS SLC-40 | GPS III-04 , Crew-1 | 4311 | MEO | USSF | Success | Success |

# Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
Select Mission_Outcome, count (Mission_Outcome) FROM `e-quanta-
317101.spacex.spacex` group by Mission_Outcome
```

- Result:

| | Mission_Outcome | f0_ |
|---|---|---|
| 0 | Success | 98 |
| 1 | Failure (in flight) | 1 |
| 2 | Success (payload status unclear) | 1 |
| 3 | Success | 1 |

# Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
SELECT DISTINCT Booster_Version FROM `e-quanta-317101.spacex.spacex` WHERE
PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM `e-quanta-
317101.spacex.spacex`)
```

- Result:

| | Booster_Version |
|---|---|
| 0 | F9 B5 B1048.5 |
| 1 | F9 B5 B1051.4 |
| 2 | F9 B5 B1060.2 |
| 3 | F9 B5 B1058.3 |
| 4 | F9 B5 B1051.6 |
| 5 | F9 B5 B1048.4 |
| 6 | F9 B5 B1049.4 |
| 7 | F9 B5 B1051.3 |
| 8 | F9 B5 B1056.4 |
| 9 | F9 B5 B1049.5 |
| 10 | F9 B5 B1060.3 |
| 11 | F9 B5 B1049.7 |

# 2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
SELECT Landing__Outcome, Booster_Version, Launch_Site FROM `e-quanta-
317101.spacex.spacex` where Landing__Outcome = 'Failure (drone ship)' and
EXTRACT(YEAR FROM Date)=2015
```

- Result:

| | Landing__Outcome | Booster_Version | Launch_Site |
|---|---|---|---|
| 0 | Failure (drone ship) | F9 v1.1 B1012 | CCAFS LC-40 |
| 1 | Failure (drone ship) | F9 v1.1 B1015 | CCAFS LC-40 |

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

- 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) FROM `e-quanta-
317101.spacex.spacex` WHERE Date > cast('2010-06-04 00:00:00' as datetime) and
Date < cast('2017-03-20 00:00:00' as datetime) GROUP BY Landing__Outcome ORDER
BY count(Landing__Outcome) DESC
```

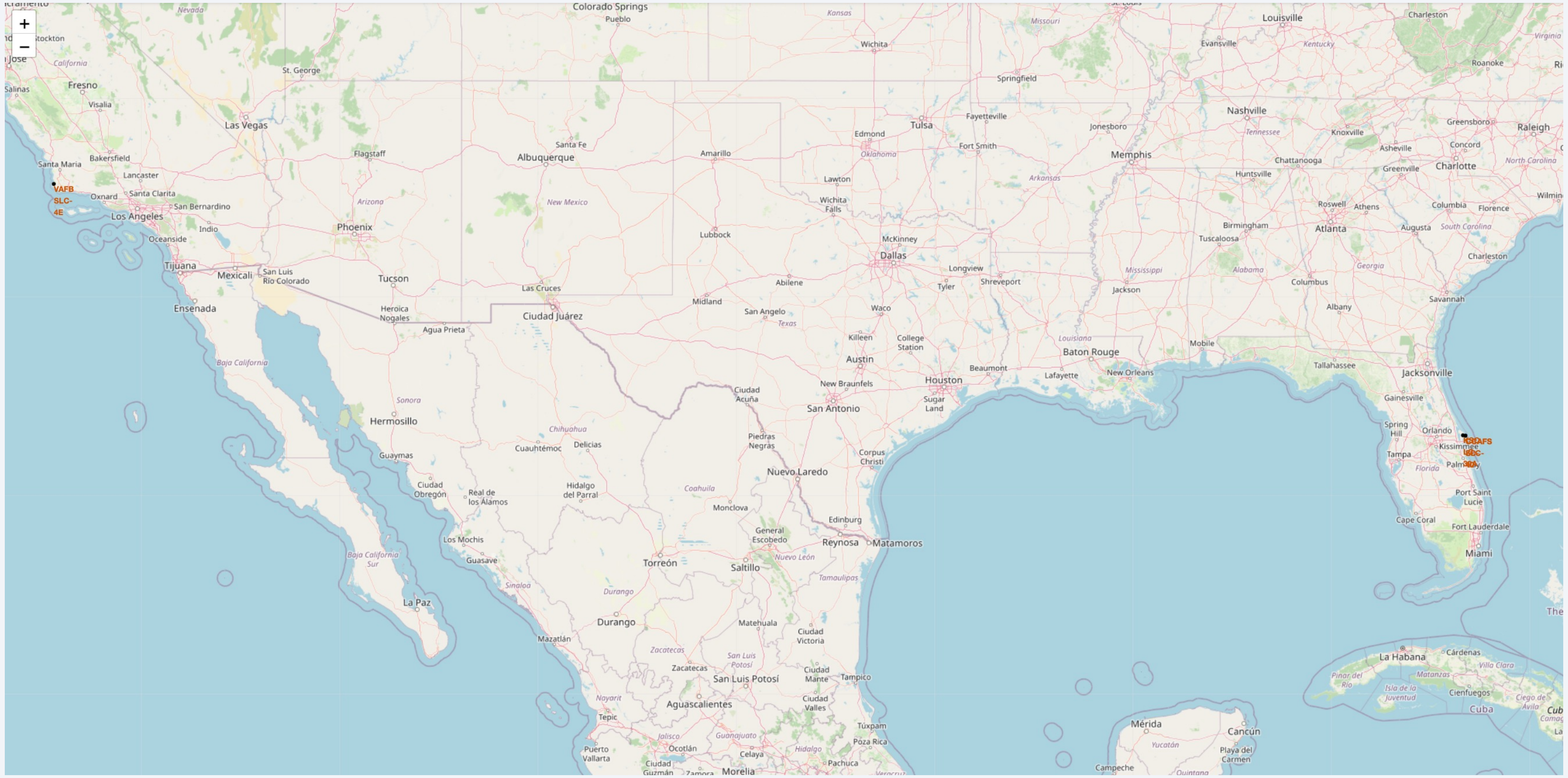
- Result:

| | Landing__Outcome | f0_ |
|---|---|---|
| 0 | No attempt | 10 |
| 1 | Failure (drone ship) | 5 |
| 2 | Success (drone ship) | 5 |
| 3 | Success (ground pad) | 3 |
| 4 | Controlled (ocean) | 3 |
| 5 | Uncontrolled (ocean) | 2 |
| 6 | Failure (parachute) | 1 |
| 7 | Precluded (drone ship) | 1 |

Section 3

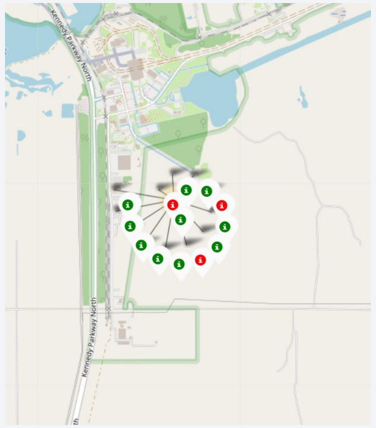
# Launch Sites Proximities Analysis

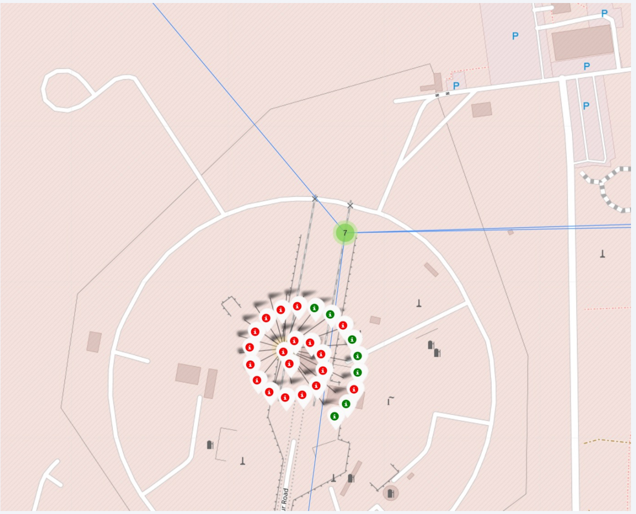
# Launch Sites

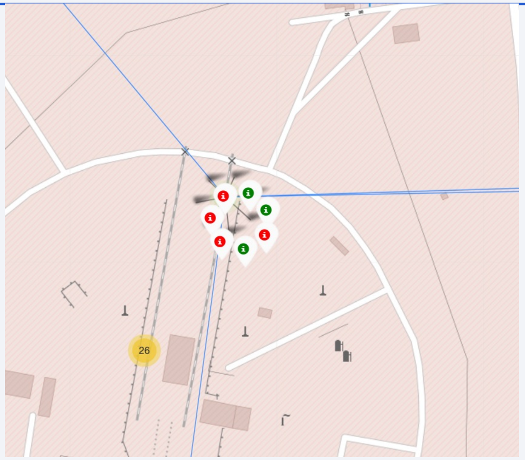


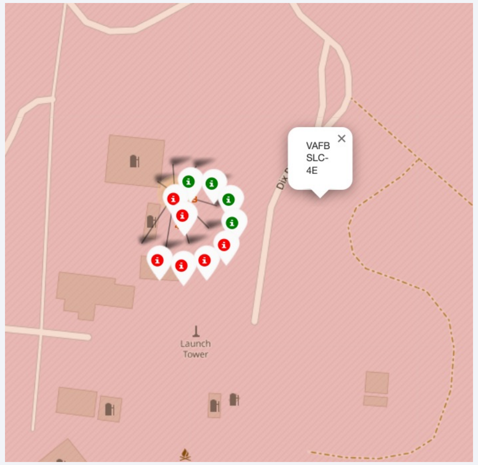
- The launch sites are near the weast and east coast of USA.

# Colored marker from every launch site.



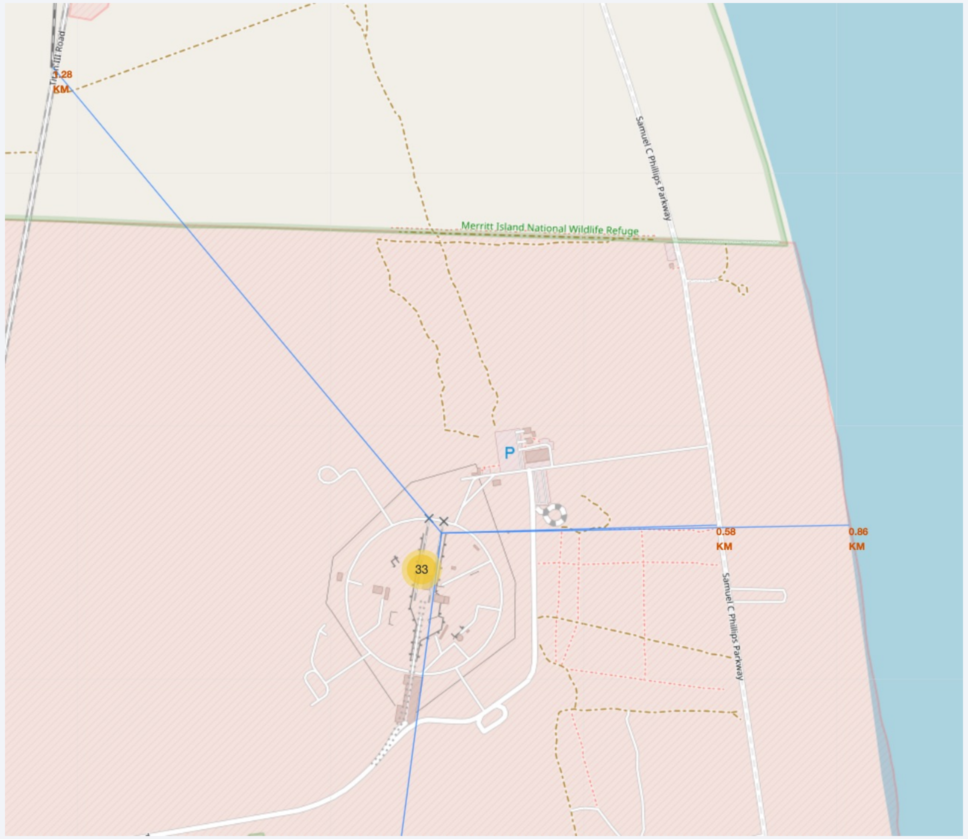







- Success count (green) and failure count (red) from every launch site.

# Distances from CCAFS SLC-40

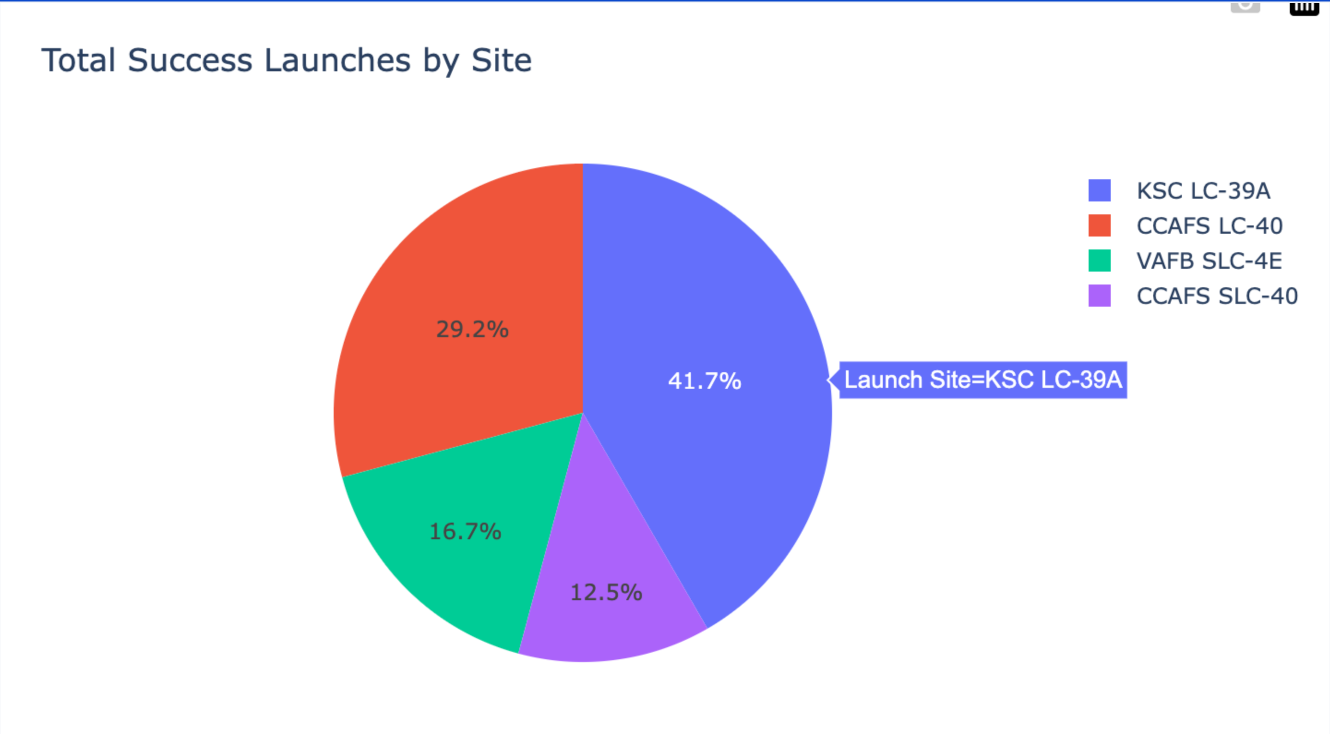



- Distance to Highway (0.58 km), Coast (0.86 km) and railway (1.28 km). Airport is the most far away (51.43 km)

Section 4

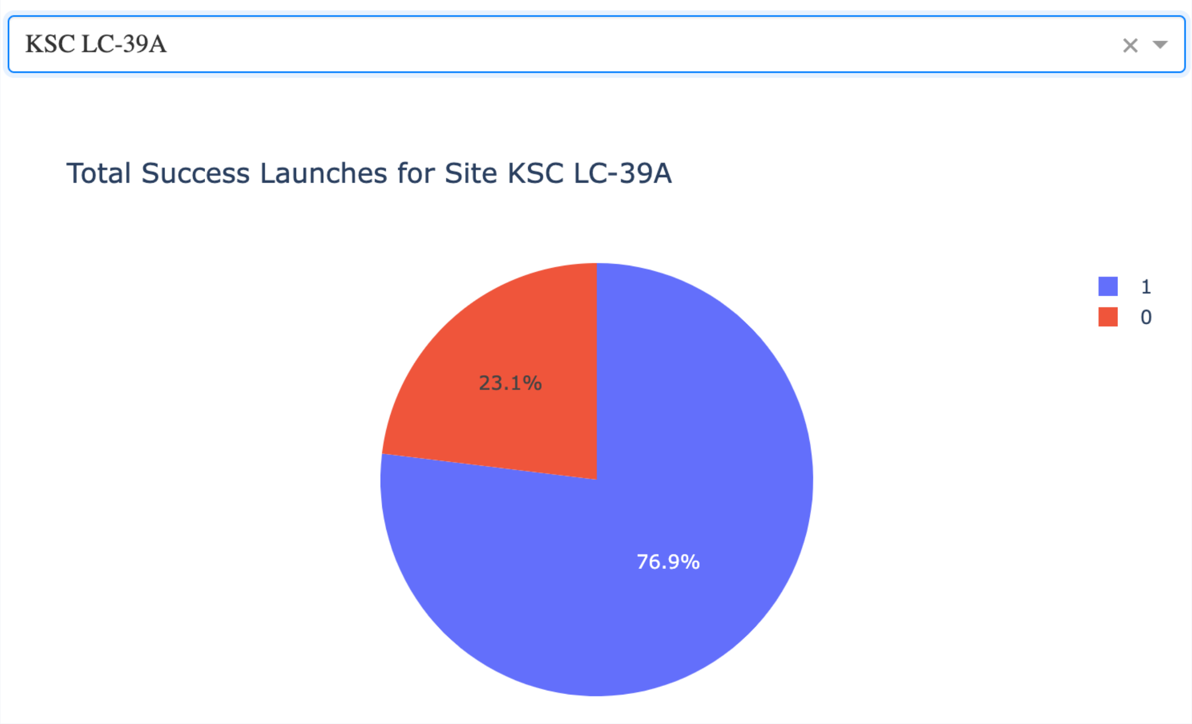
# Build a Dashboard with Plotly Dash

# Total Success Launch by Site (% of total)




- Launch Site KSC LC 39A have the most success launches (41,7%)

# Success launches by KSC LC-39A




- KSC LC-39A have the most success launches and have 76.9% of success of its launches.

# Payload and Success Rate in different ranges.

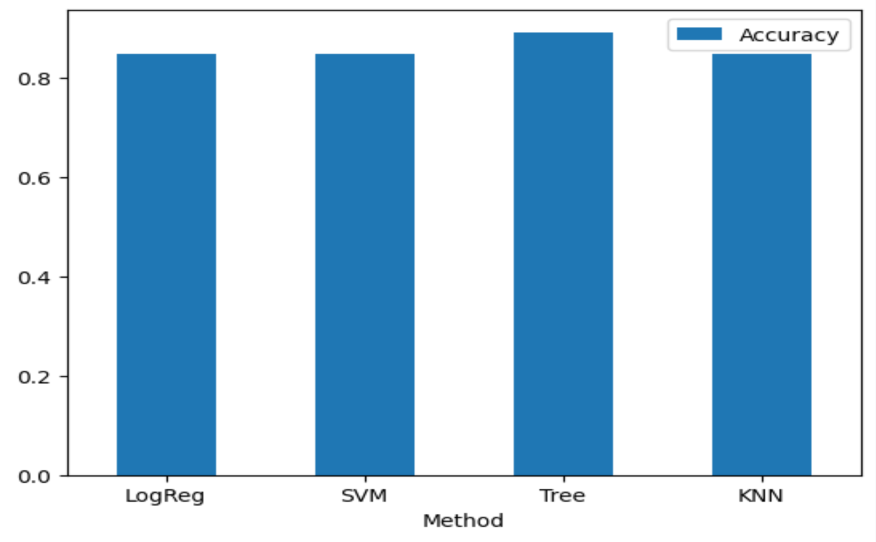


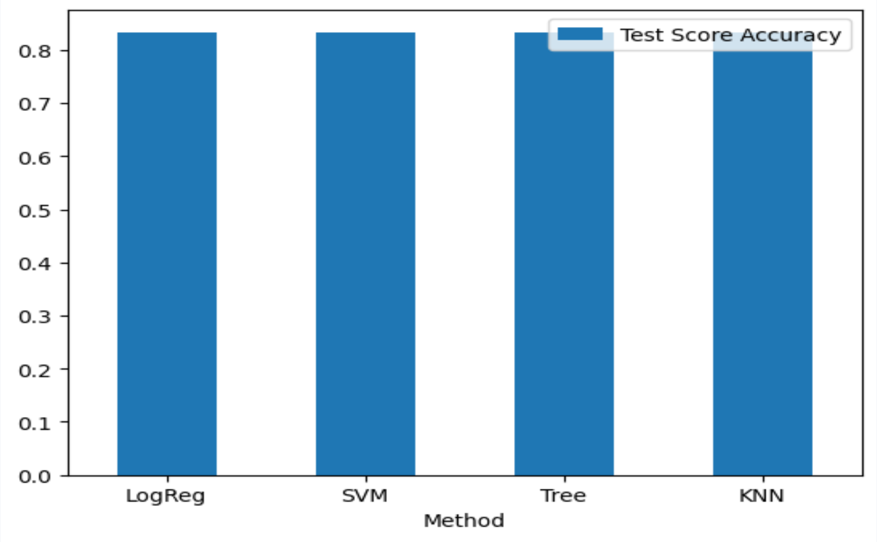
- Low weight payloads (0~4000kg) have better success rate than heavy weightpayloads (4000~10000kg).

Section 5

# Predictive Analysis (Classification)

# Classification Accuracy



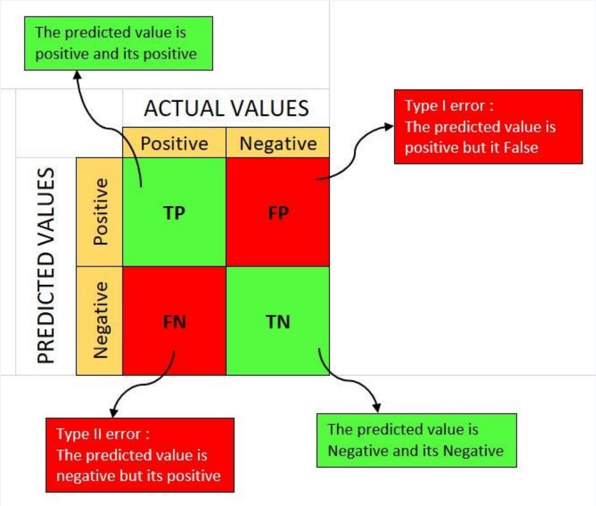



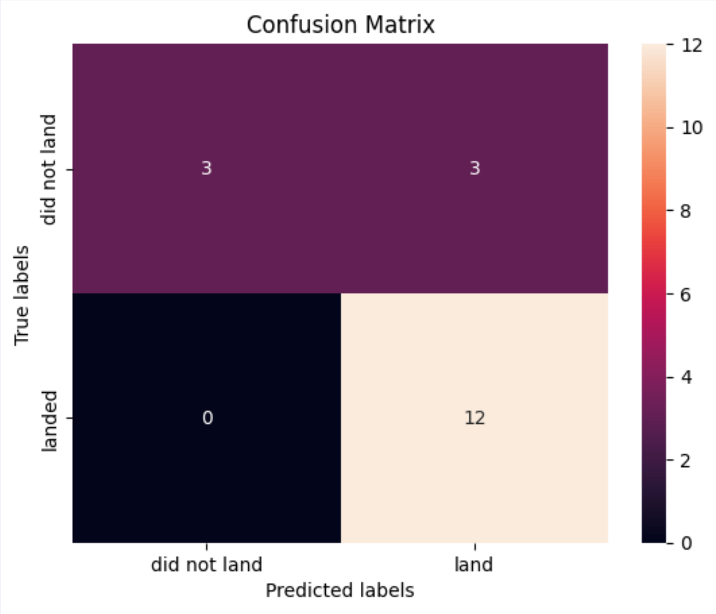

| | LogReg | SVM | Tree | KNN |
|---|---|---|---|---|
| Accuracy | 0.84643 | 0.84821 | 0.89107 | 0.84821 |
| Test Score Accuracy | 0.83333 | 0.83333 | 0.83333 | 0.83333 |

- Decision tree have better accuracy.

# Confusion Matrix

All model have the same Confusion Matrix.




The problem is, in every model, the false positives (type error I). It's predicted a positive value, but its false.

# Conclusions

- Decision Tree Classifier Algorithm is the best model.
- Low weighted payloads perform better that heavy weighted payload.
- The orbits with the best success rates are GEO, HEO, SSO and ES-L1.
- KSC-L39A is the site with the best success rate.
- The success rates increase though the years, we assume that with the experience gained the performance increase.

# Thank you!