THE RACE TO MARS

Ly Phat Viet Linh

19th Oct 2021

Outline

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

Executive Summary

- SpaceX is an American manufacturer and designer of means of transportation for space travel. Since 2008 the company has been steadily launching, orbiting and recovering spacecrafts into space.
- 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.
- Using data science, this report aims to help SpaceX determine the landing outcome for future spacecraft launches, reducing the cost of each launch.
- The report also shows insight into what aspect / property of a launch affects its landing outcome.

Introduction

- The report aims to determine and predict the first stage landing outcome of future rocket launch.
- By determining if the first stage will land successfully, 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.
- The questions:
  - What property of a launch affects landing outcome? Is it launch site, payload range, booster version or orbit type?
  - Can we predict future landing outcome base on historical launch data?

Methodology

Methodology

Executive Summary

Data collection methodology:

Data was collected through Space X API: Booster Version, Launch Site, Payload Data, Core Data

Data wrangling

Dealing with missing data

Exploratory data analysis (EDA) using visualization and SQL

Interactive visual analytics using Folium and Plotly Dash

Predictive analysis using classification models

4 models were built using Support Vector Machine, K Nearest Neighbor, Logistic Regression, and Decision Tree

Hyperparameters tuning with GridSearch CV

Data Collection

The data set was collected through both SpaceX API and web scraping its Wikipedia page.

API

- Using Request library to get data from SpaceX API

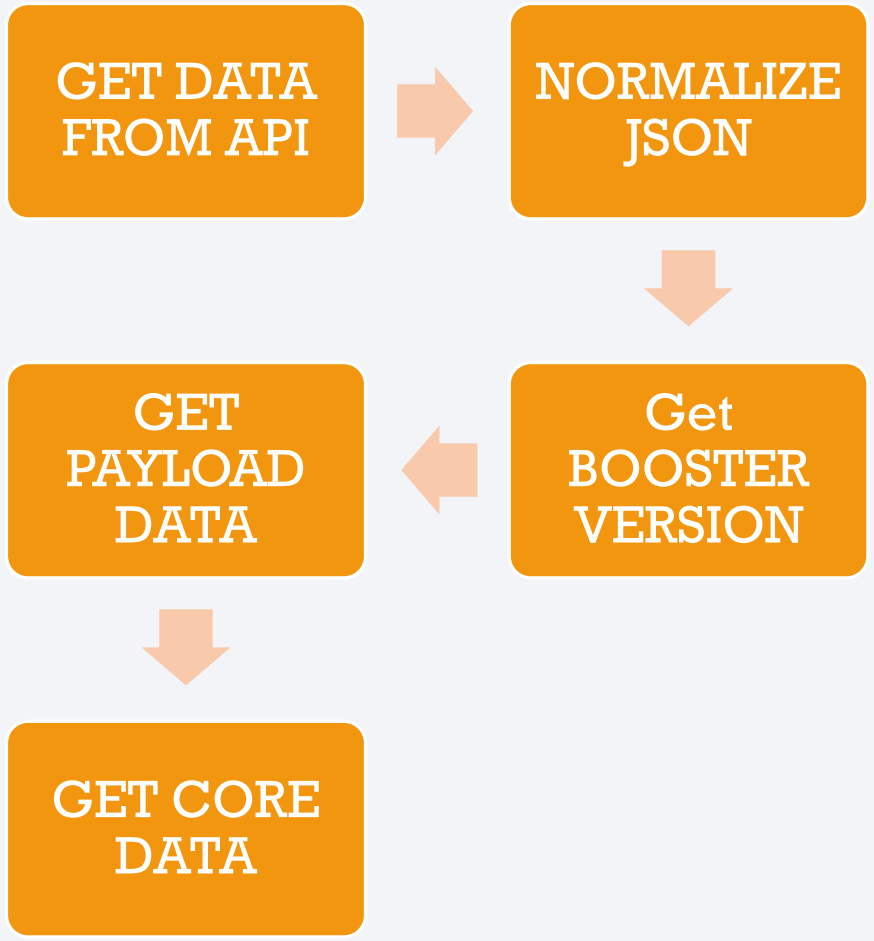
WEB SCRAPING

- Collect Falcon 9 historical launch from Wikipedia

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- GITHUB link to Notebook



Data Collection - Scraping

Request the Wiki page

- Create a request object
- Create Beautiful Soup object for scraping

Extracting Column Names

- Find all "table"
- Iterate through each header to get the column names

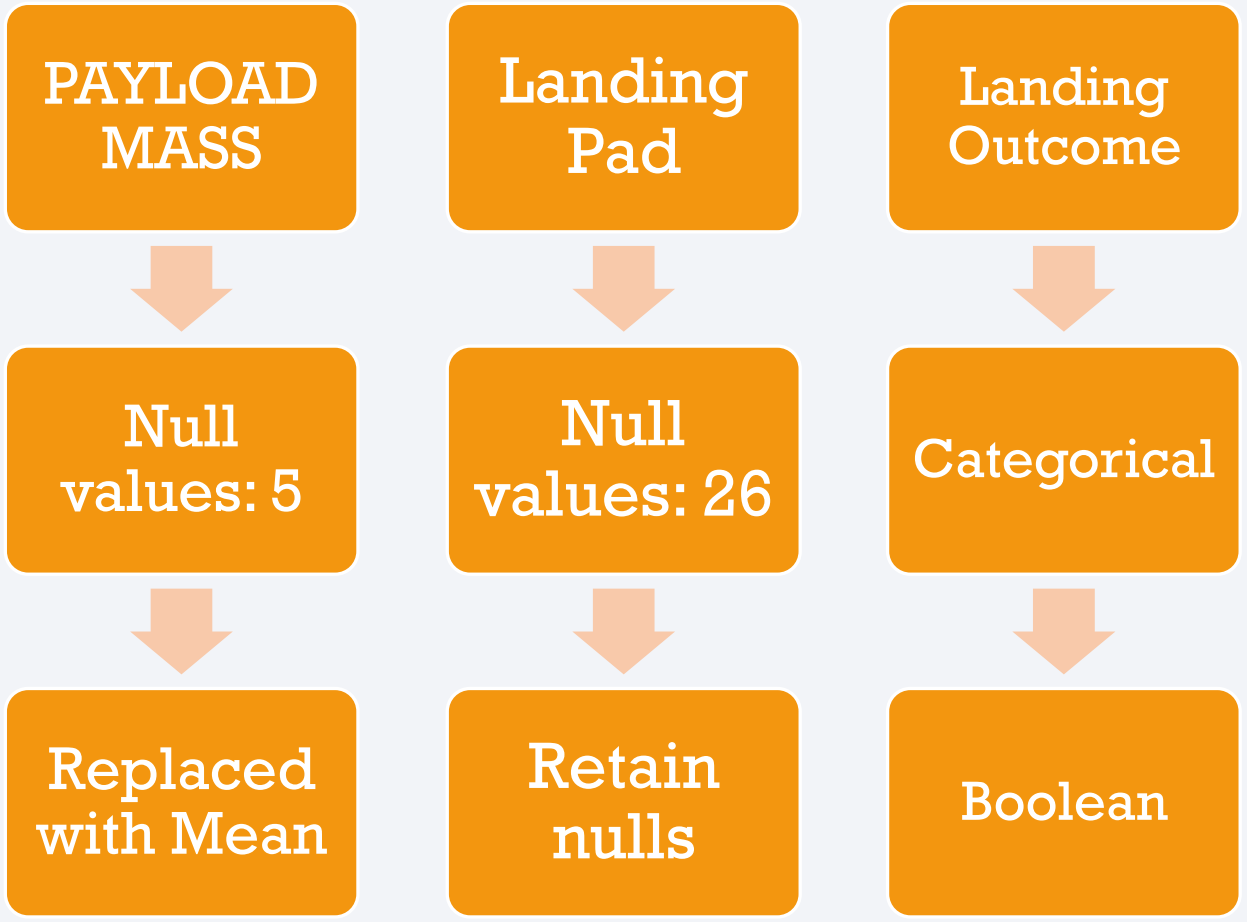
Create DataFrame

- Iterate through each row to parse data
- Append data to respective column in dataframe

- GITHUB URL to Notebook

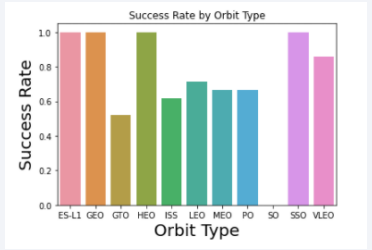
Data Wrangling

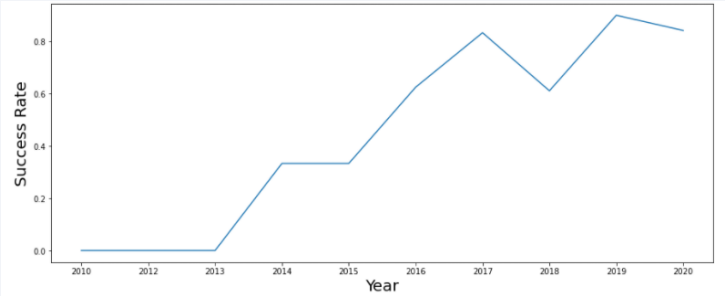
- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- GITHUB URL to Notebook

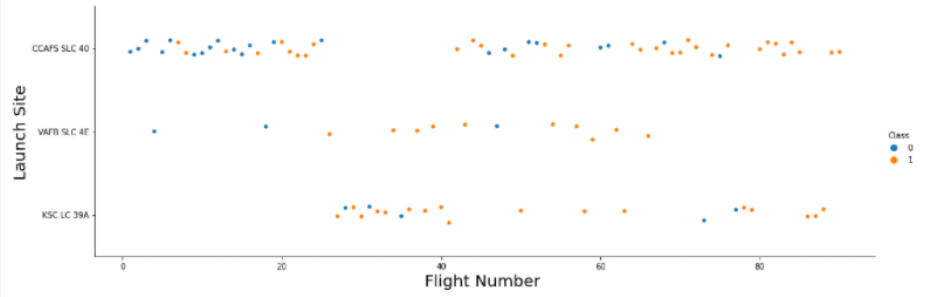


EDA with Data Visualization

- Scatter plots were plotted between Flight Number, Launch Site, Payload Mass, to try find the patterns between these 3 features with launch outcome.
- Bar chart classify success rate of launches between orbit type.
- Line chart show success rate trend over the years
- GITHUB URL to Notebook







EDA with SQL

- Exploratory Data Analysis with SQL:
  - Display unique launch sites
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was acheived.
  - 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. Use a subquery
  - 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
- GITHUB URL to Notebook

Build an Interactive Map with Folium

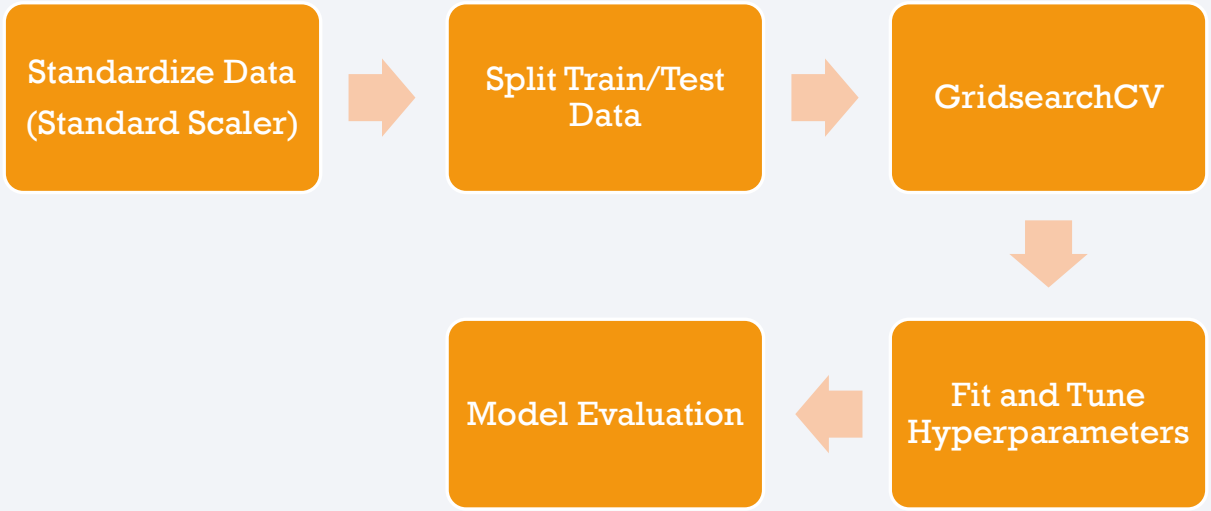
- Launch sites marker added on map
- Success / failed launches of each site added on map with color code.
- Marker and distance line was added from launch site CCAFS SLC-40 to nearest railway, highway and city (Melbourne) to see if launch sites were in close proximity of traffic or dense population area
- GITHUB URL to Notebook (note that Folium does not get rendered by Github)

Build a Dashboard with Plotly Dash

- Interactive pie-chart showing the launch success rate between launch sites, or within each launch site. Interactive scatter chart showing launch outcome by payload range.
- Conclusions draw by visual analysing with the chart:
  - Which site has the largest successful launches? VAFB SLC4E 9600kg
  - Which site has the highest launch success rate? KSC LC39A
  - Which payload range(s) has the highest launch success rate? 2000 - 4000kg
  - Which payload range(s) has the lowest launch success rate? 6000 - 7000 kg
  - Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate? FT
- GITHUB URL to .py file and screenshots of dashboard

Predictive Analysis (Classification)

- To predict the landing outcome for future launches, 4 predictive models were created using algorithm:
  - Support Vector Machine
  - K Nearest Neighbor
  - Decision Tree
  - Logistic Regression
- GITHUB URL to Notebook

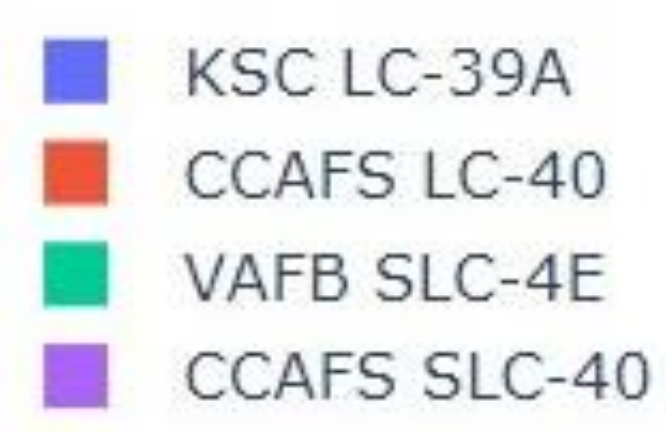


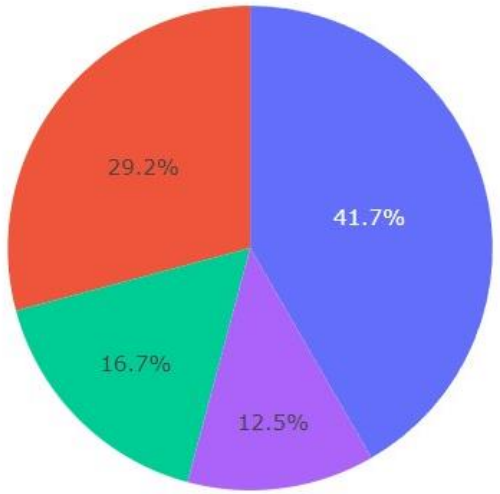
Results

Exploratory data analysis shows the success rate of a launch are affected by launch site, payload carried, orbit type and booster version.

Since 2013 the overall success rate has been increasing steadily.

4 predictive models were built and fine tuned to predict the landing outcome of future launches

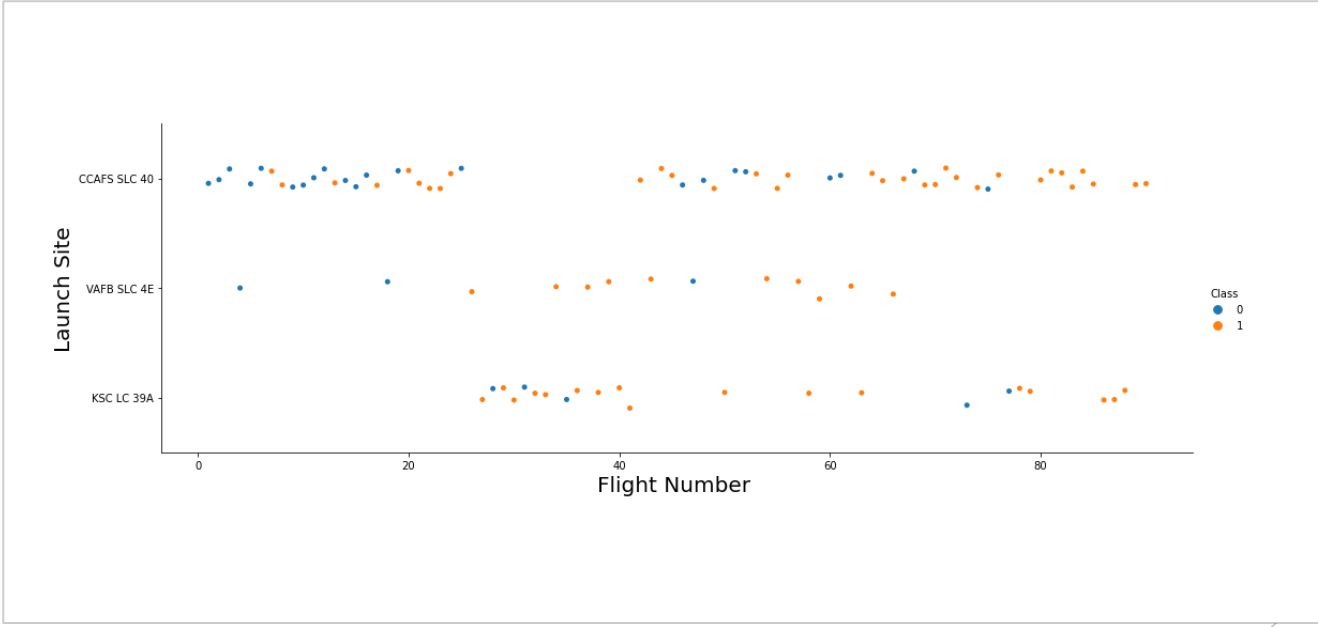





INSIGHT
DRAWN FROM
EDA

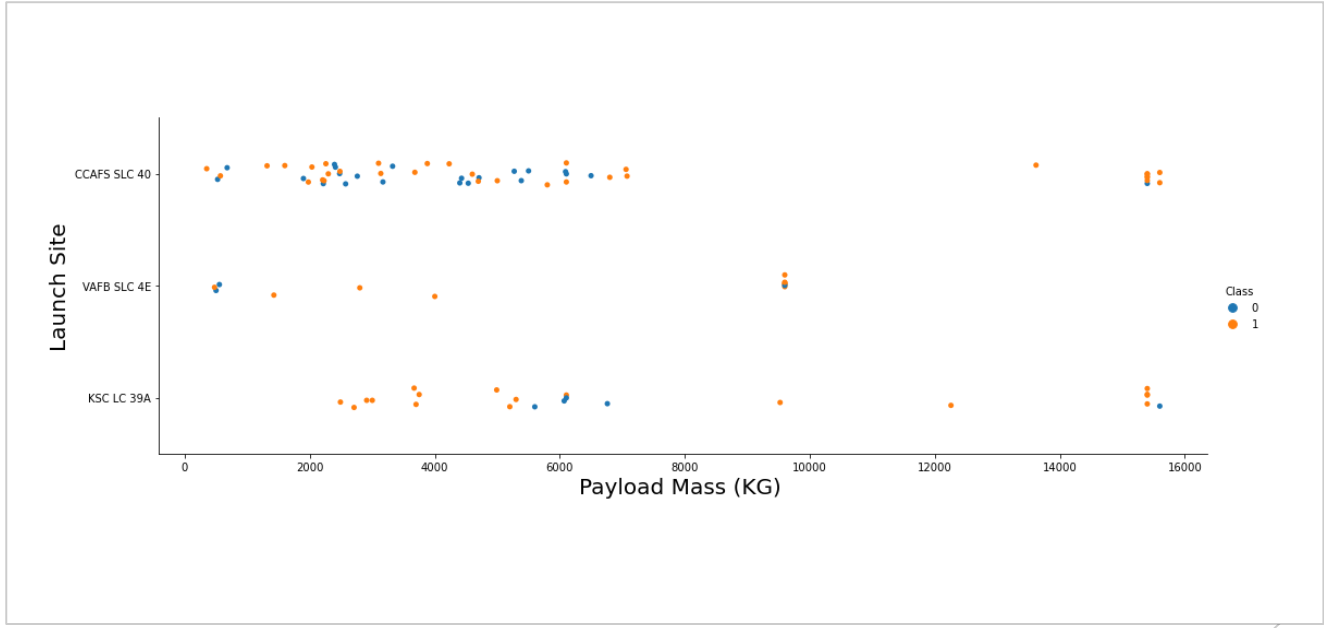
Flight Number vs. Launch Site

- As we can see, the success rate seems to be increasing with Flight Number, for each Launch Site (class 1 = succeeded, class 0 = failed)



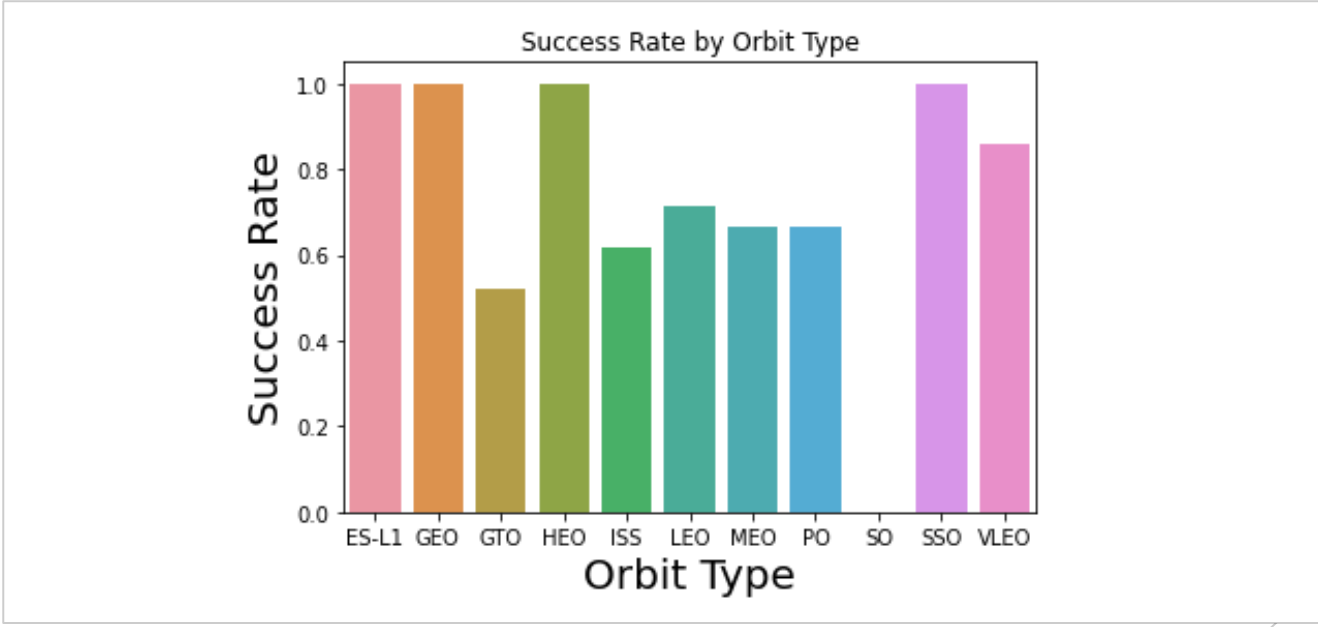
Payload vs. Launch Site

- Higher payload mass launches seem to have more success rate than low payload mass launches. (class 1 = succeeded, class 0 = failed)



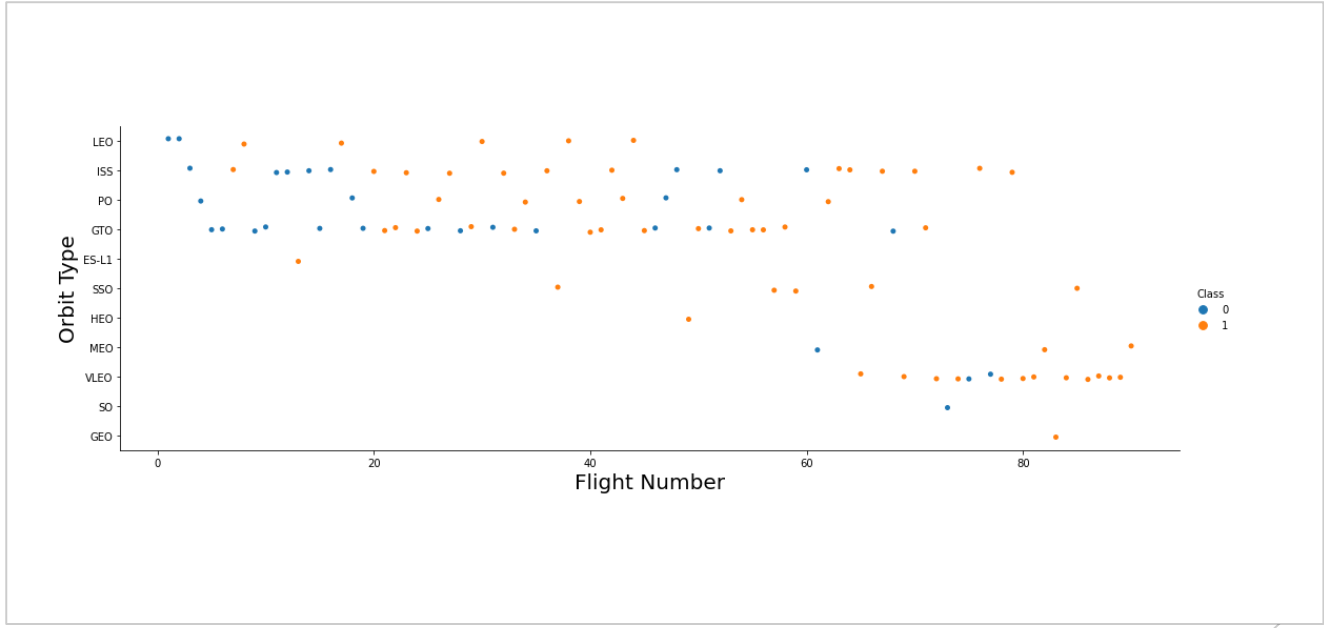
Success Rate vs. Orbit Type

- Bar chart showing success rate per each orbit type.



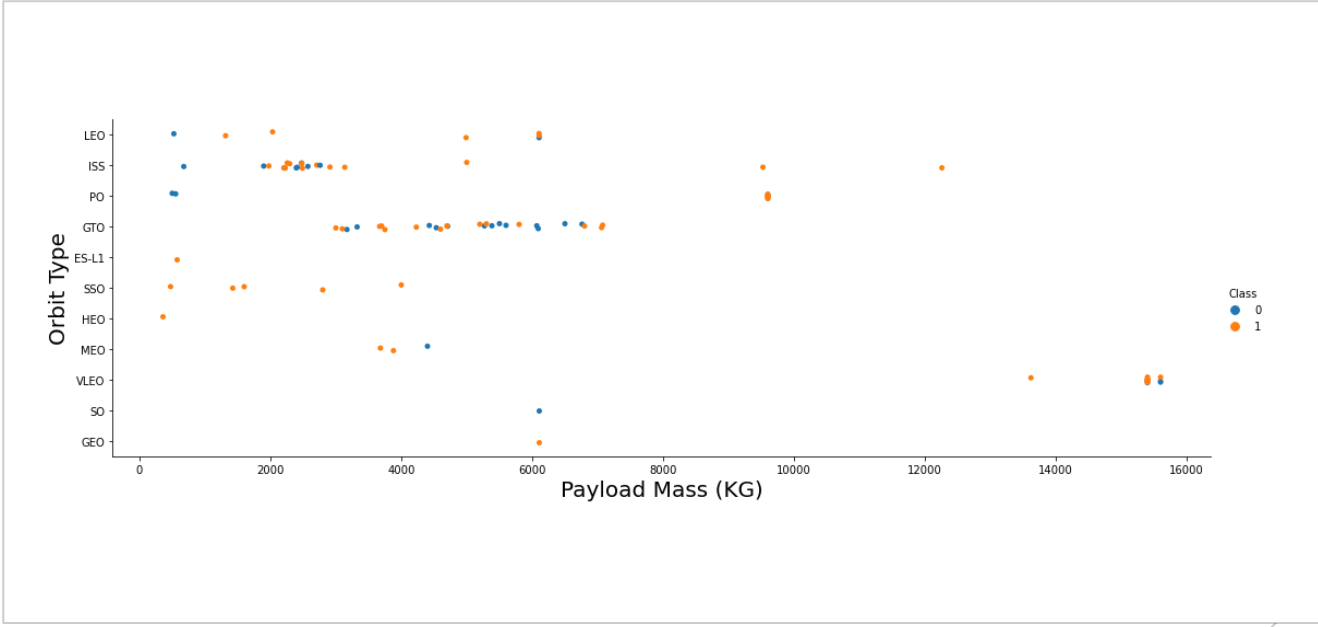
Flight Number vs. Orbit Type

- With the LEO orbit the Success appears related to the number of flight
- There seems to be no relationship between flight number when in GTO orbit.



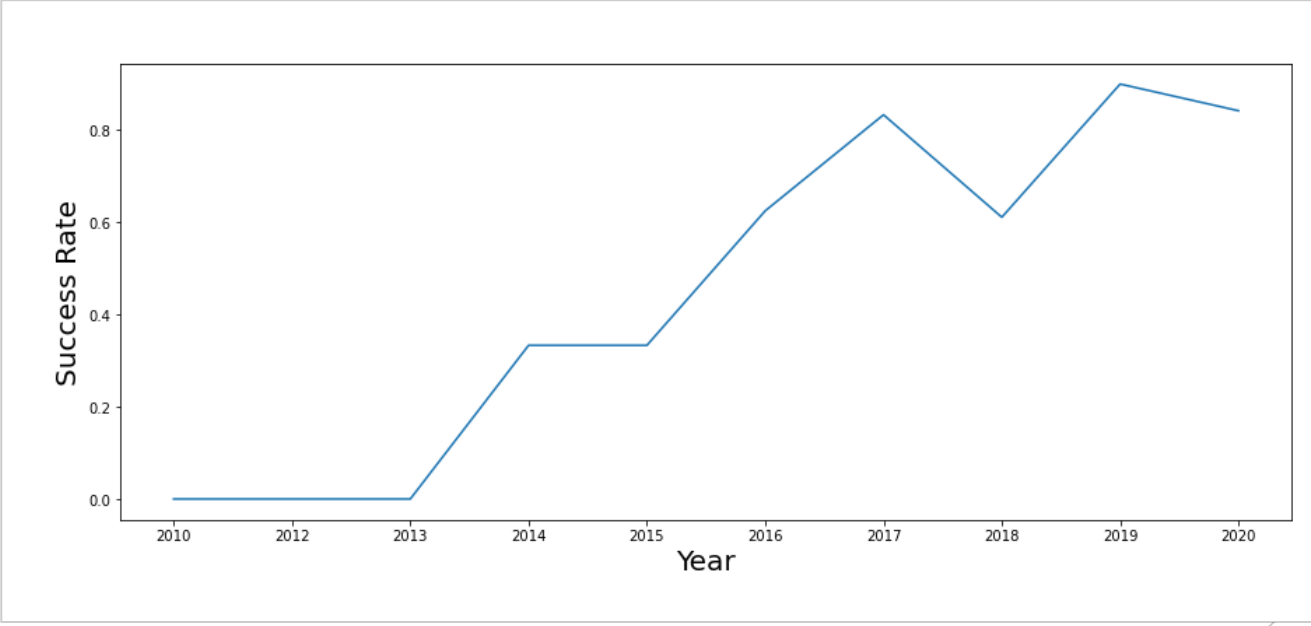
Payload vs. Orbit Type

- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits. (class 1 = succeeded, class 0 = failed)



Launch Success Yearly Trend

- The launch success rate improves significantly each year!



All Launch Site Names

- Unique launch sites can be select with a simply query

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

```
%%sql
select
unique launch_site from SPACEXTBL
```

```
 * ibm_db_sa://shd48119:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90
Done.
```

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

Launch Site Names Begin with 'CCA'

5 launches with launch sites begin with CCA:

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

```
%%sql
select
*
from
SPACEXTBL
where launch_site like 'CCA%'
limit 5
```

```
 * ibm_db_sa://shd48119:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119/bludb
Done.
```

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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: 45,596 kg

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

```
%%sql
select

sum(payload_mass__kg_)

from

SPACEXTBL

where

customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://shd48119:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119/bludb
Done.
```

1
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1: 2,928 kg

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

```
%%sql
select

avg(payload_mass__kg_)

from

SPACEXTBL

where

booster_version = 'F9 v1.1'
```

```
 * ibm_db_sa://shd48119:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:30119/bludb
Done.
```

1
2928

First Successful Ground Landing Date

First successful ground landing was achieved on:

22nd Dec 2015

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

```
%%sql
select

min(DATE)

from

SPACEXTBL

where landing__outcome = 'Success (ground pad)'
```

```
 * ibm_db_sa://shd48119:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databases.app
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters that have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

```
[39]: %%sql
select

unique booster_version

from

SPACEXTBL

where landing__outcome = 'Success (drone ship)'
and 4000 < payload_mass__kg_ < 6000
```

```
 * ibm_db_sa://shd48119:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8
30119/bludb
Done.
```

booster_version
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1
F9 FT B1029.2
F9 FT B1021.1
F9 FT B1023.1
F9 FT B1038.1

Total Number of Successful and Failure Mission Outcomes

Total number of successful versus failure mission outcomes:

100 success : 1 failure

List the total number of successful and failure mission outcomes

```
[42]: %%sql
select

mission_outcome,
count(mission_outcome) as total

from

SPACEXTBL

group by mission_outcome
```

```
 * ibm_db_sa://shd48119:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2
30119/bludb
Done.
```

[42]:

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Maximum payload carried: 15,600 kg
- 15 different booster versions have carried the maximum payload.

```
%%sql

select booster_version, payload_mass__kg_

from SPACEXTBL

where payload_mass__kg_ IN (select max(payload_mass__kg_) from SPACEXTBL)

order by booster_version
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

Failed landing_outcomes in drone ship, their booster versions, and launch site names in year 2015:

[66]:
```
%%sql
select
landing__outcome,
booster_version,
launch_site

from SPACEXTBL

where landing__outcome = 'Failure (drone ship)'
and DATE LIKE '%2015%'
```

```
 * ibm_db_sa://shd48119:***@824dfd4d-99de-440d-9991-629c01b3832d.bs2io90l08kqb1od8lcg.databas
30119/bludb
Done.
```

[66]:

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Number of landing outcomes between 2010-06-04 and 2017-03-20, in descending order.

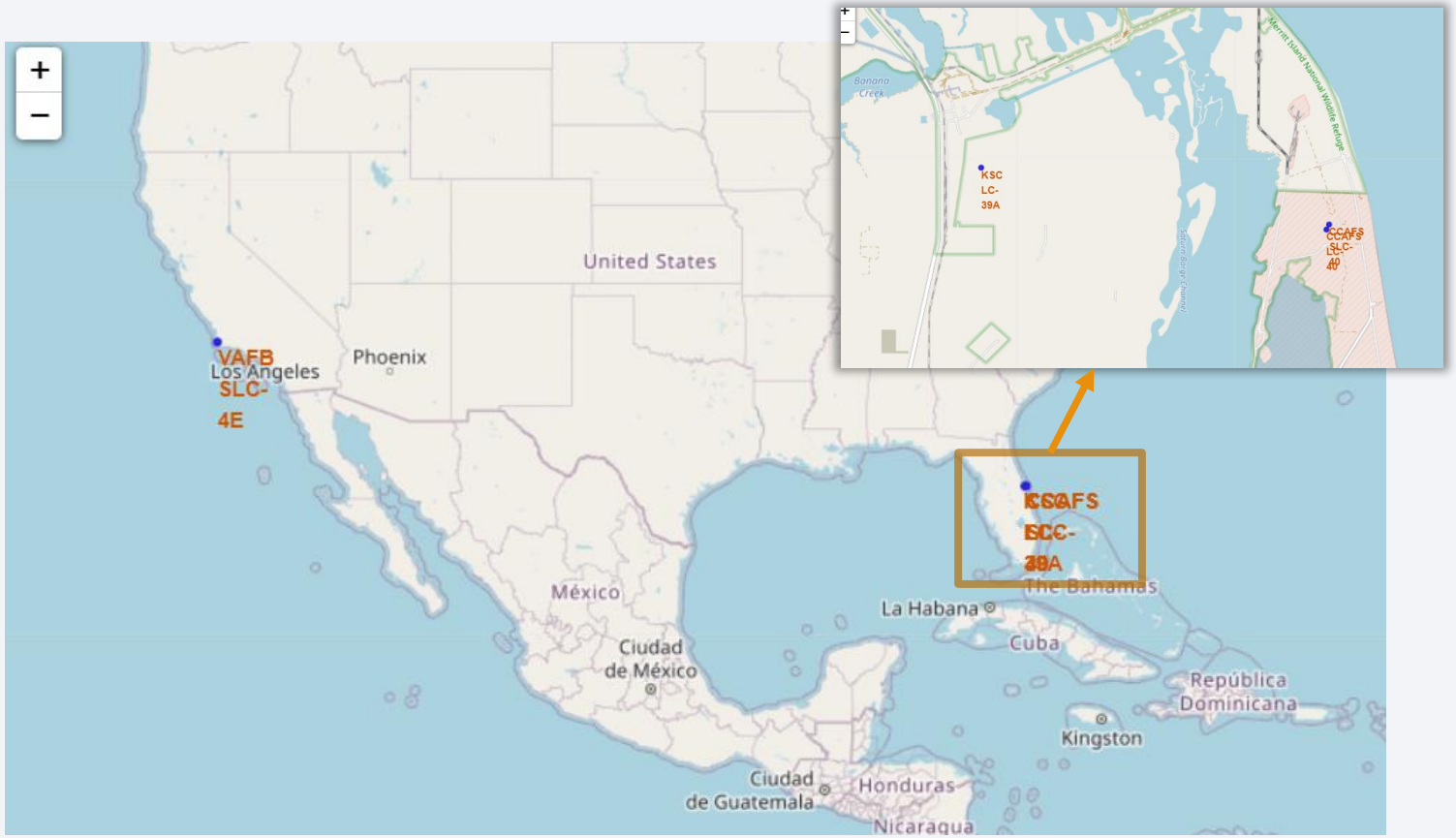
```
%%sql
select
landing__outcome,
count(landing__outcome) as count
from SPACEXTBL
where DATE between '2010-06-04' and '2017-03-20'
group by landing__outcome
order by count desc
```

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


Launch Site
Proximity Analysis

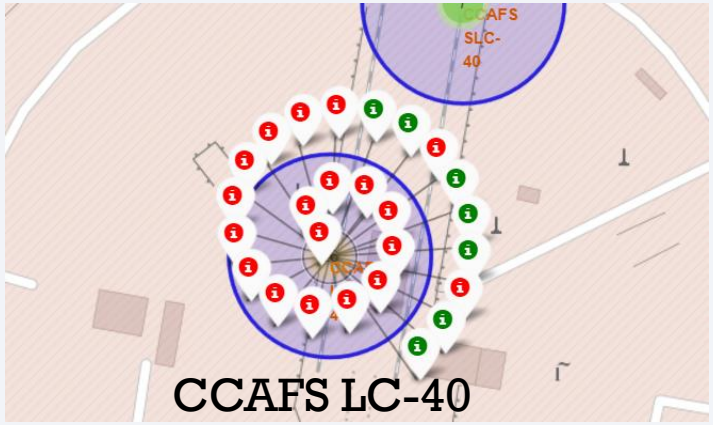
Launch Site Locations

We can see 4 launch sites are all situated on the coastline.

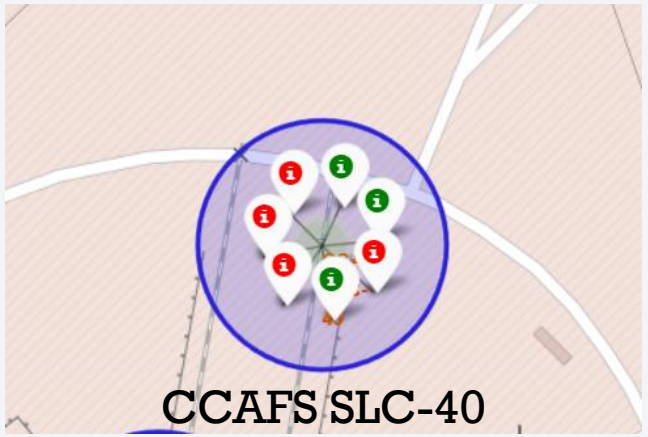


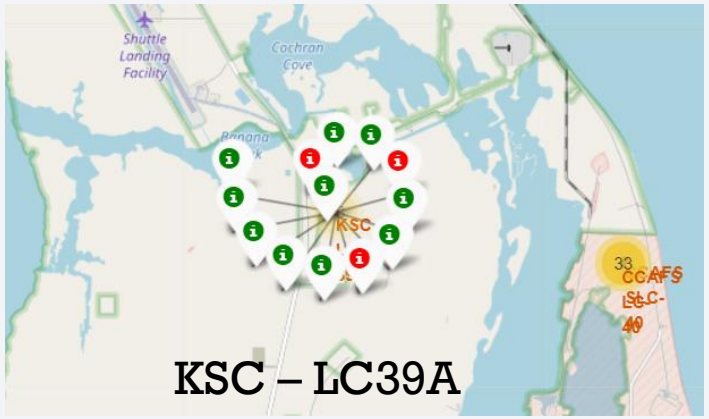
Launch Outcome By Launch Site

- The KSC – LC39A Launch site has higher success rate
- The CCAFS LC-40 has lower success rate
- Fewer launches was done on CCAFS SLC-40 and VAFB SLC-4E



CCAFS LC-40


CCAFS SLC-40

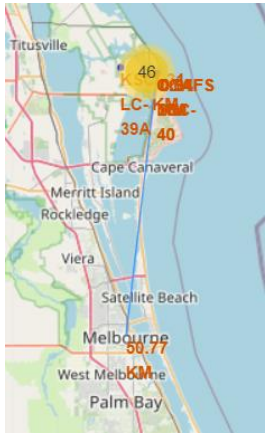


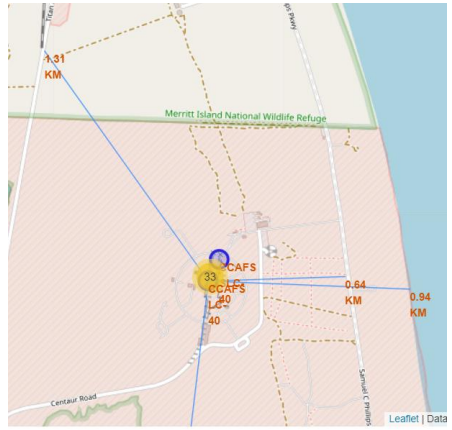
KSC – LC39A

VAFB SLC-4E

Launch Site Proximity

- Launch sites are in relatively close proximity of:
  - Railways
  - Highways
  - Coastlines
- Far away from:
  - Cities

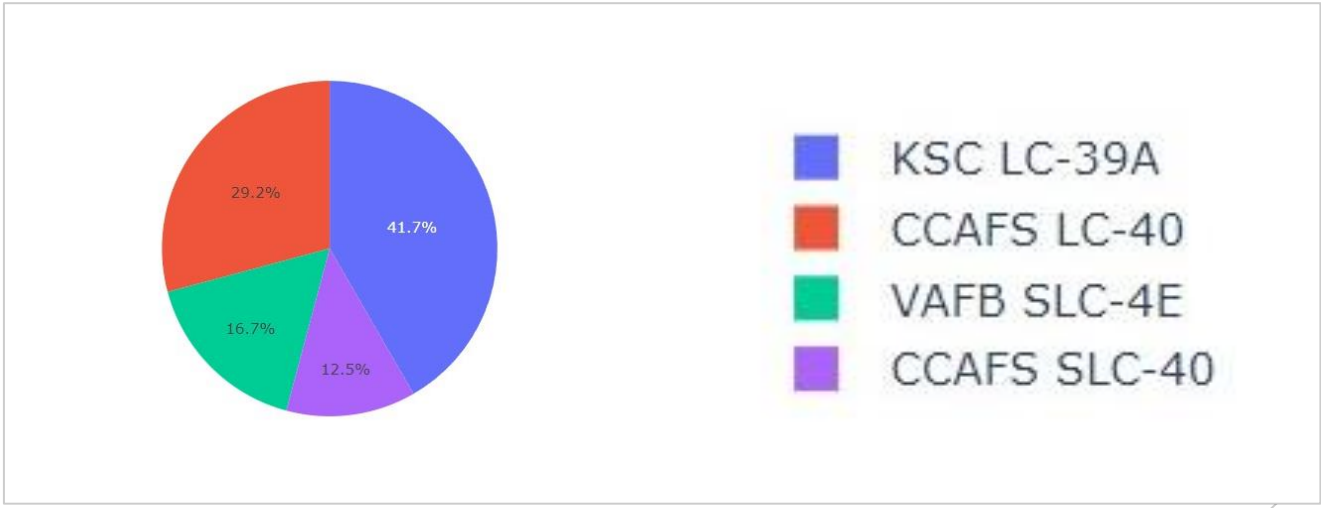




Dashboard with Plotly Dash

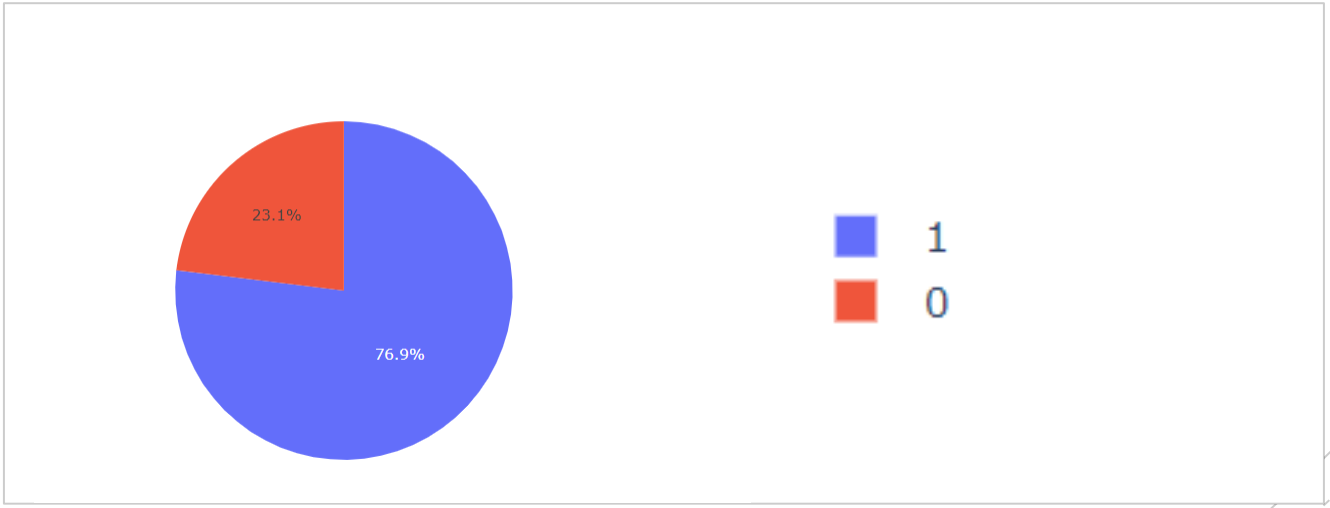
Success Rate of all Launch Sites

- KSC LC-39A has the highest success rate of all launch sites.
- CCAFS SLC-40 has the lowest success rate of all launch sites.



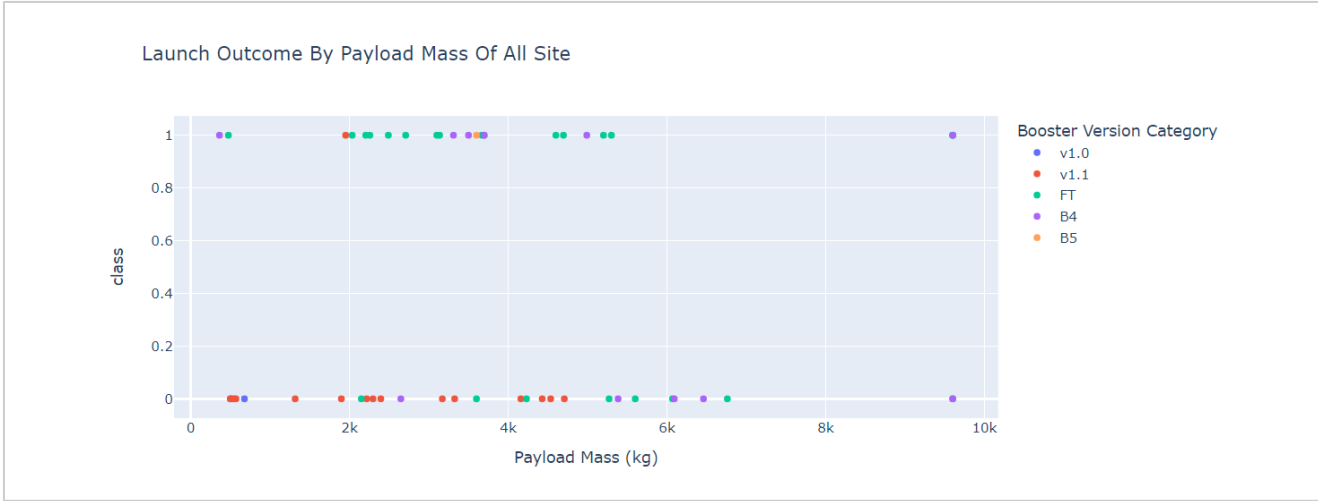
Success Launch Rate of KSC LC-39A

- Class = 1 : success rate
- Class = 0: failure rate
- KSC LC-39A has 76.9% success launch rate.



Launch Outcome by Payload Mass

- Booster FT has more success launches with payload range from 2000 – 4000kg
- Payload range from 6000-8000kg has no success launches.



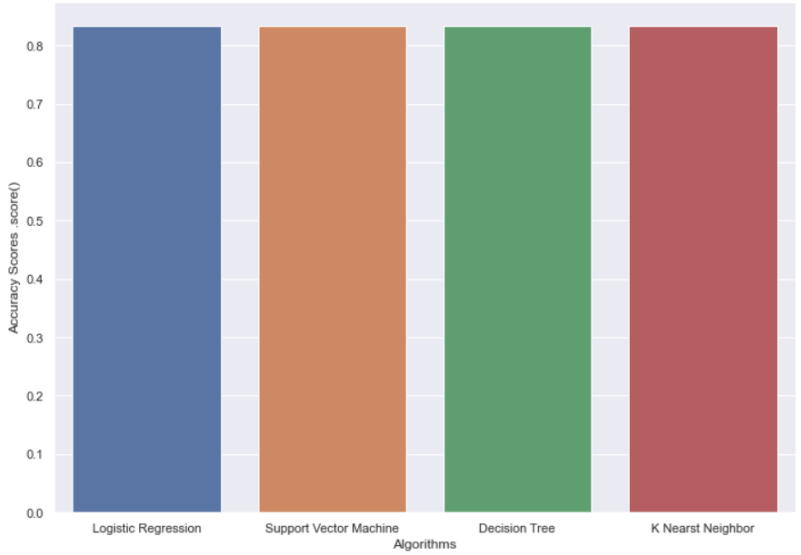
Predictive Analysis (Classification)

Classification Accuracy

4 models were built:

- Logistic Regression
- Support Vector Machine
- Decision Tree
- K Nearest Neighbor

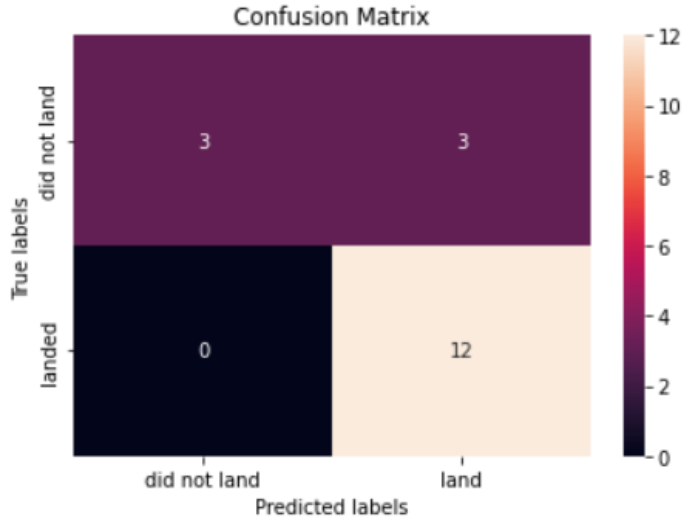
The scores are practically the same: 0.83



Confusion Matrix

We can see:

- 3 records were correctly classified as “did not land”
- 12 records were correctly classified as “landed”
- 3 records were incorrectly classified as “landed”
- 3 records were incorrectly classified as “did not land”



Conclusions

- Since 2013 the overall success rate has been increasing steadily.
- In terms of launch sites:
  - KSC LC-39A has the highest success rate of all launch sites.
  - CCAFS SLC-40 has the lowest success rate of all launch sites.
- In terms of payload range:
  - Range 2000 – 4000kg has the highest success rate
  - Range 6000-7000kg has the lowest success rate
- Booster version and orbit type's affect on landing outcome are unconclusive, as the dataset is too small.


Thank you