IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Name: Nabil Mostafa
Date: 29/09/2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Discussion
- Conclusion

Executive Summary

- Summary of Methodologies:
  - Data Collection through API
  - Data Collection with 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 Results:
  - Exploratory Data Analysis Result
  - Interactive Analytics and Dashboard
  - Predictive Analytics Result

Introduction

- Project background and context:

  SpaceX claims that a Falcon 9 rocket launch costs 62 million dollars on its website, while it costs other providers upward of 165 million dollars, much of the savings is because SpaceX can reuse the first stage of the rockets. Therefore, if we can determine if the first stage will land, we can predict the cost of a launch.

  The objective of this study is to see how the information can be used to help an alternate company -SpaceY- which wants to compete against SpaceX for. This goal of the project is to use machine learning algorithms to predict if the first stage will land successfully.

- Problems we want to answer:
  - What factors determine if the rocket will land successfully?
  - What operating conditions needs to be in place to ensure a successful landing program.

Section 1

Methodology

Methodology

Summary

- Data collection methodology:
  - Data was collected using 2 methods:
    - SpaceX API.
    - Web scraping Wikipedia page (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches).
- Perform exploratory data analysis (EDA) using visualization and SQL.
- Build interactive visual analytics using Plotly Dash.
- Perform predictive analysis using classification algorithms.

Data Collection

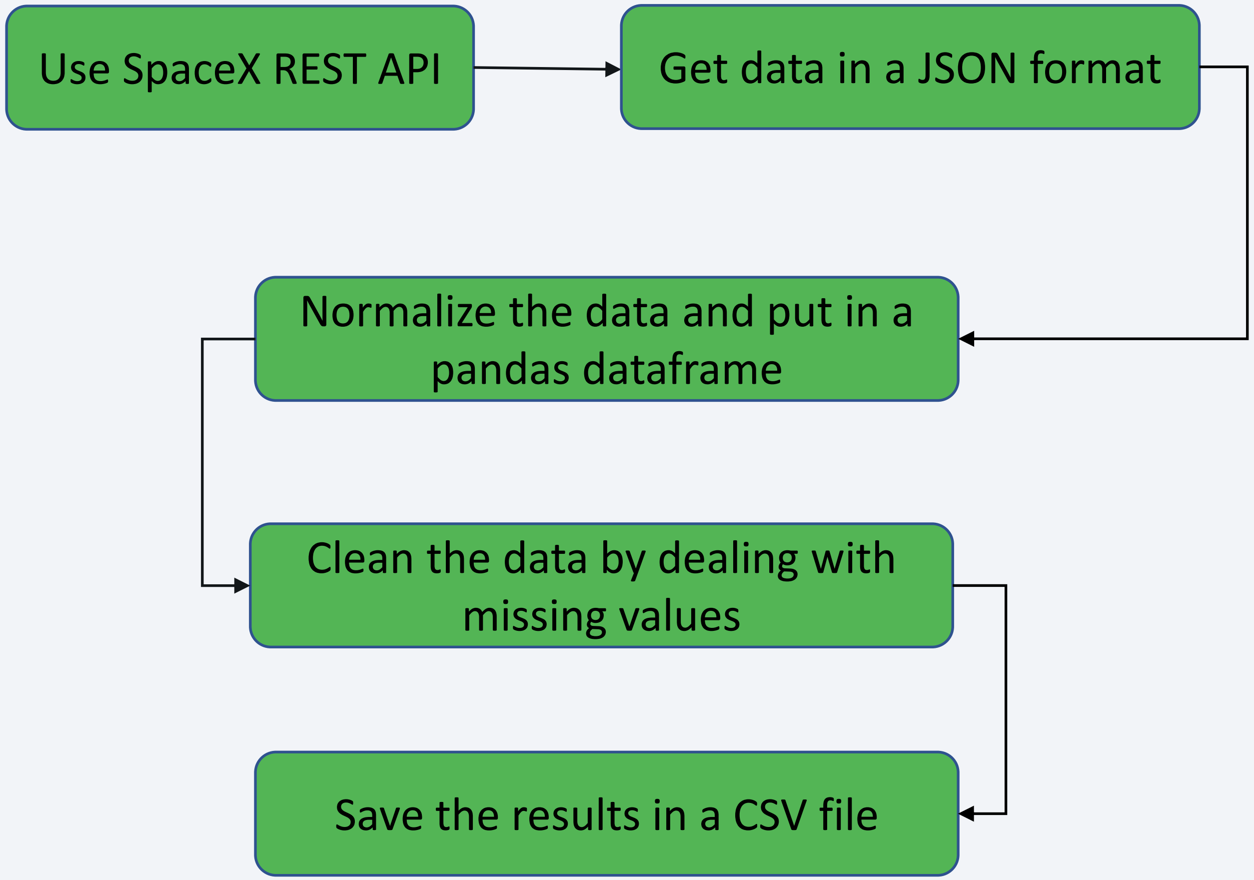
- Data was collected from 2 sources:

  1- SpaceX API.

  2- Wikipedia page using web scraping with requests and beautifulsoup liberaries.

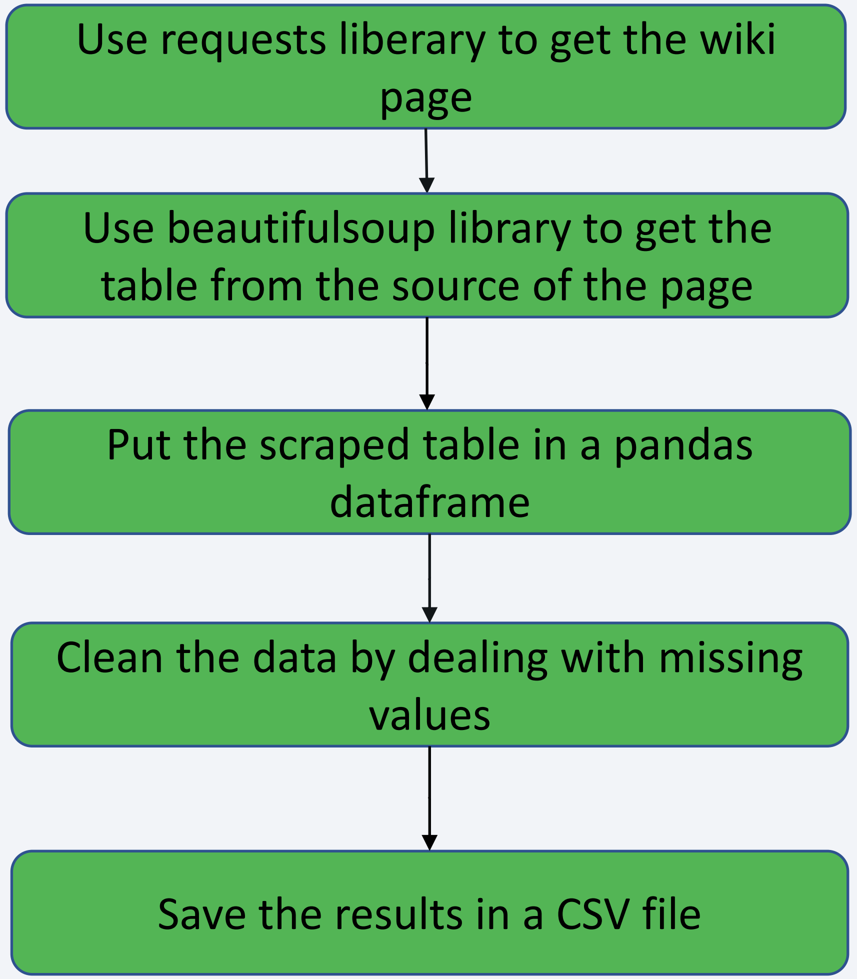
Data Collection – SpaceX API

- By using the API we get the returned data in a Json format.
- We then use that format response to create a dataframe of the results.
- Then we clean this data by removing some of the missing values.
- The collection notebook GitHub link: https://github.com/NabilMostafa1/IBM-Data-Science-Projects/blob/master/Data_Science_capstone_project/IBM_spacex-data-collection_api.ipynb



Data Collection - Scraping

- The data was collected from a Wikipedia page.
- We wanted to get the successful lunches of the Falcon 9 rocket.
- We clean the data and remove the missing values.
- The notebook GitHub link: https://github.com/NabilMostafa1/IBM-Data-Science-Projects/blob/master/Data_Science_capstone_project/IBM_spacex-data-collection_web-scraping.ipynb



Data Wrangling

- First we did some exploration of the data that we have to gather some insights on how the data is structured
- We discover what launch sites where used, what are the launch orbits, what type of booster rockets were used, and what is the range of the payload.
- Then we summariesed each of this parameters to get a broad idea of what is important in the data, like what is the most common launch site.
- Then we hot-one encoded the categorical values so we can preform the prediction later on.
- The GitHub notebook: https://github.com/NabilMostafa1/IBM-Data-Science-Projects/blob/master/Data_Science_capstone_project/IBM_spacex_EDA.ipynb

EDA with Data Visualization

- We used matplotlib to plot some attributes of the data and how they corelate with each other.
- We can visualize how some of the features are related, like the orbit and payload.
- The GitHub notebook: https://github.com/NabilMostafa1/IBM-Data-Science-Projects/blob/master/Data_Science_capstone_project/IBM_spacex_EDA-visualization.ipynb

EDA with SQL

- We used SQL to Gain some information about the data.
- We preformed some queries like:
  - **the names of the unique launch sites in the space mission**
  - **5 records where launch sites begin with the string 'CCA'**
  - **the total payload mass carried by boosters launched by NASA (CRS)**
  - **average payload mass carried by booster version F9 v1.1**
  - **the date when the first successful landing outcome in ground pad was achieved.**
  - **the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000**
  - **the total number of successful and failure mission outcomes**
  - **the names of the booster versions which have carried the maximum payload mass.**
- The GitHub notebook: https://github.com/NabilMostafa1/IBM-Data-Science-Projects/blob/master/Data_Science_capstone_project/IBM_spacex_EDA-SQL.ipynb

Build an Interactive Map with Folium

- We used Folium maps to see the locations of the launch sites.
- Then we used markers to see the number of failed and successful lunches.
- Then we calculated some distances from the launch sites to some near by loctions like roads, rails, coasts, and cities.
- The GitHub notebook: https://nbviewer.org/github/NabilMostafa1/IBM-Data-Science-Projects/blob/master/Data_Science_capstone_project/IBM_spacex_launch-sites.ipynb

Build a Dashboard with Plotly Dash

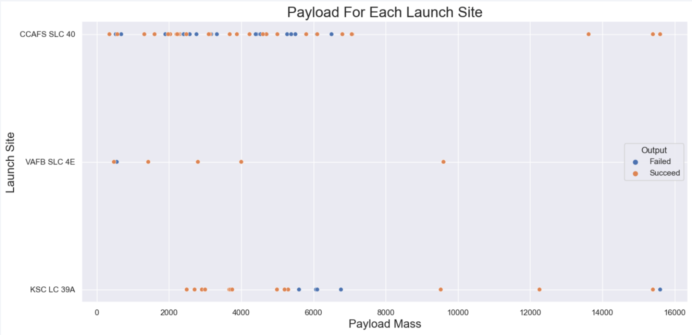
- We used Plotly Dash to build an interactive web app.
- The dashboard contains:
  - 2 dropdown selectors for the lunch site and booster type.
  - 1 range slider for the payload.
  - A scatter plot for launch site with flight number and a pie chart for the mission outcome.
- The GitHub Link: https://github.com/NabilMostafa1/IBM-Data-Science-Projects/tree/master/Data_Science_capstone_project/Plotly-Dashboard

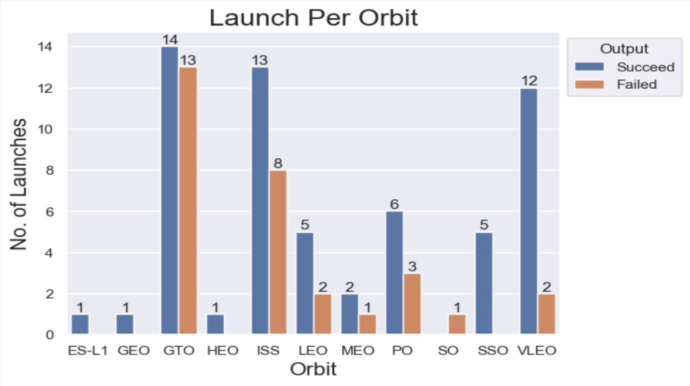
Predictive Analysis (Classification)

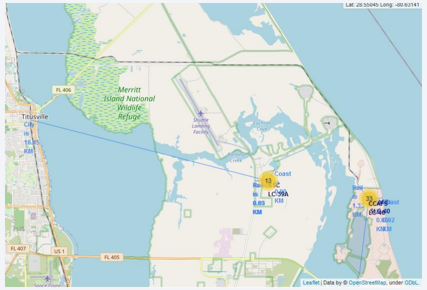
- We used different types of classification algorithms like logistic regression, decision trees, support vector machine, and k-nearest neighbor.
- We calculated the score of each classifier to determine the what would be the best algorithm for our case.
- The GitHub notebook: https://github.com/NabilMostafa1/IBM-Data-Science-Projects/blob/master/Data_Science_capstone_project/IBM_spacex_ML-prediction.ipynb

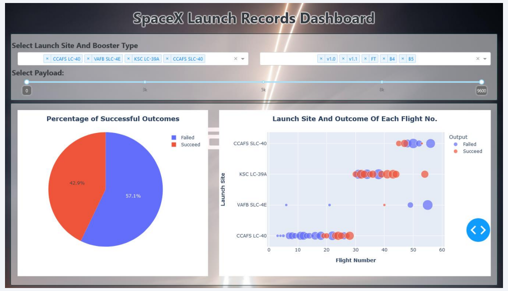
Results

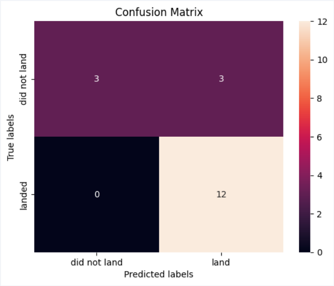
- SpaceX has used 4 different launch sites.
- The first successful landing of the first stage happened in 2015, after 5 years of trials.
- The F9 v1.1 booster are the most successful booster yet with an average of 3 tons of payload.
- Most of the missions were launched to low earth orbit.
- Al the launch locations are over 15 km from the nearest city but have a coast within 1 km.
- The classifier that returned the most accurate results was KNN.







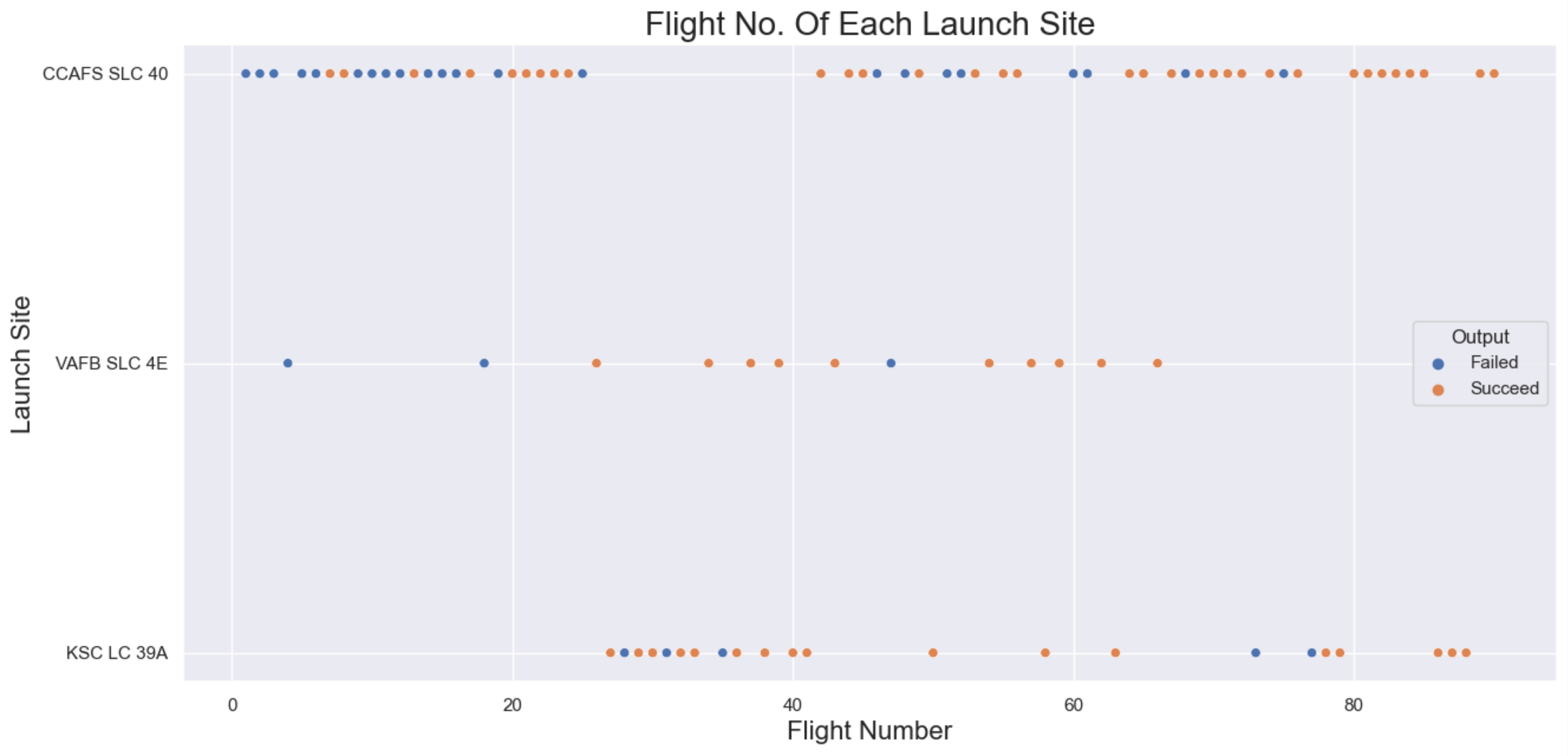




Section 2

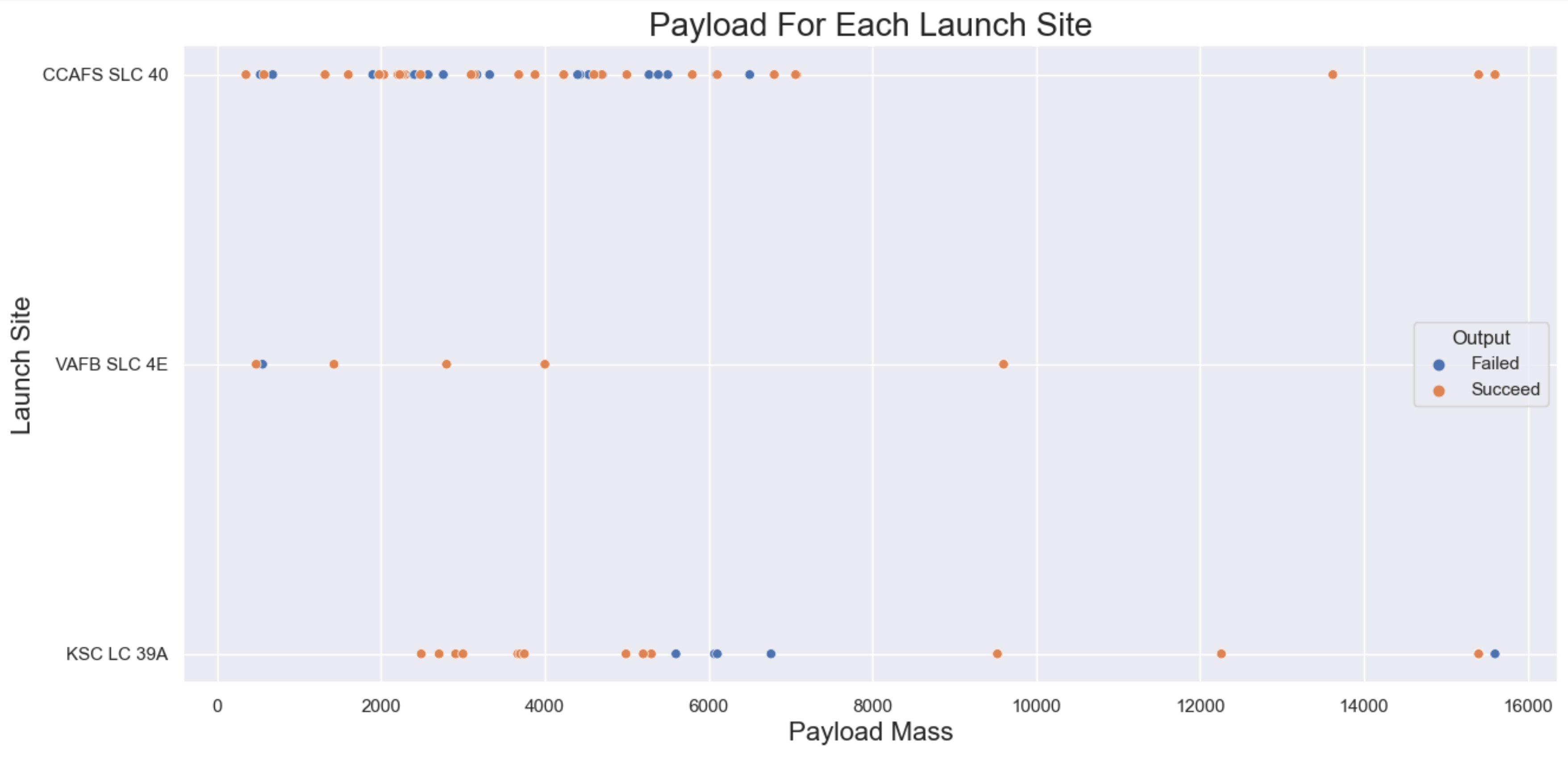
Insights drawn from EDA

Flight Number vs. Launch Site



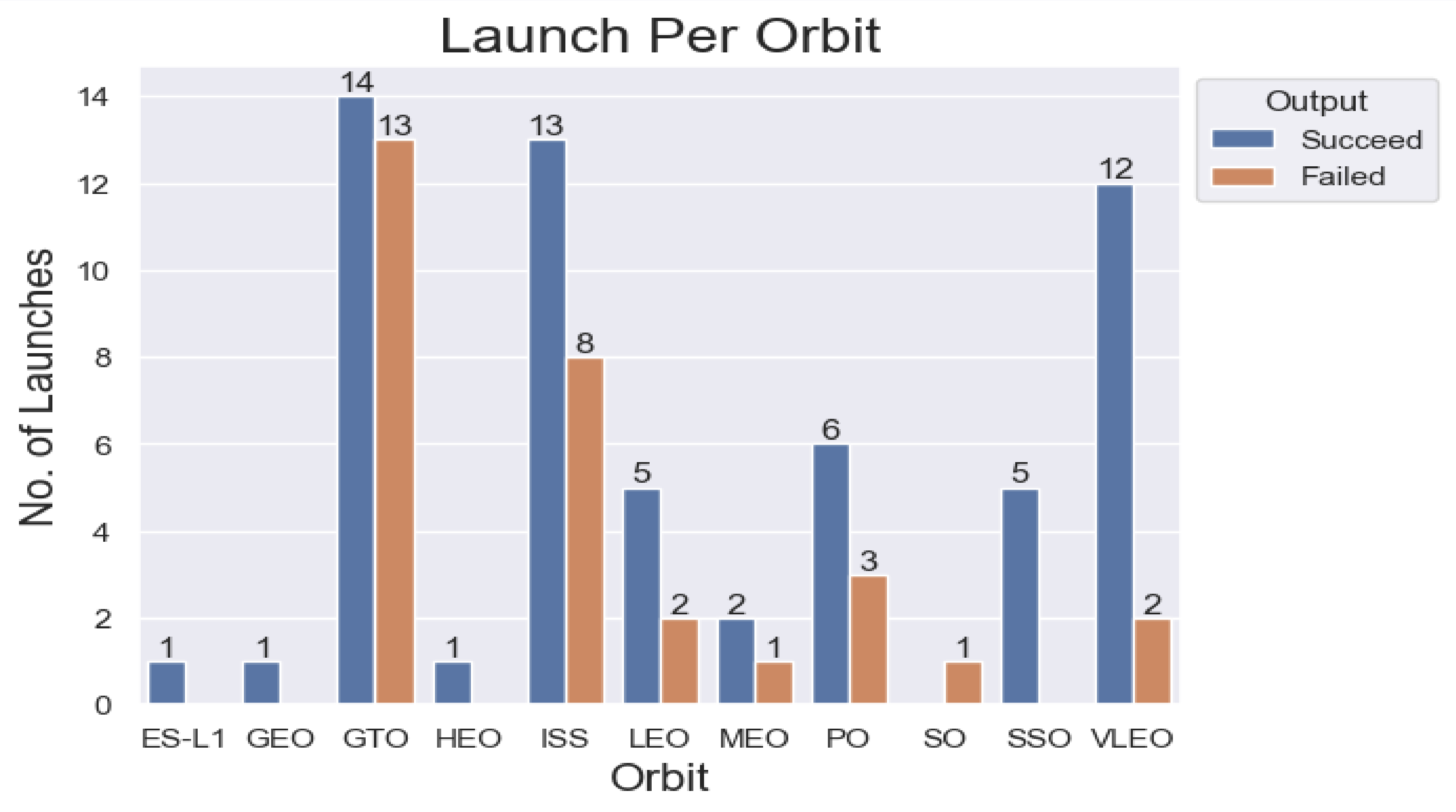
- We can notice that CCAFS SLC 40 is the most used site, except for a brief period were KSC-LC-39A was used more.

Payload vs. Launch Site



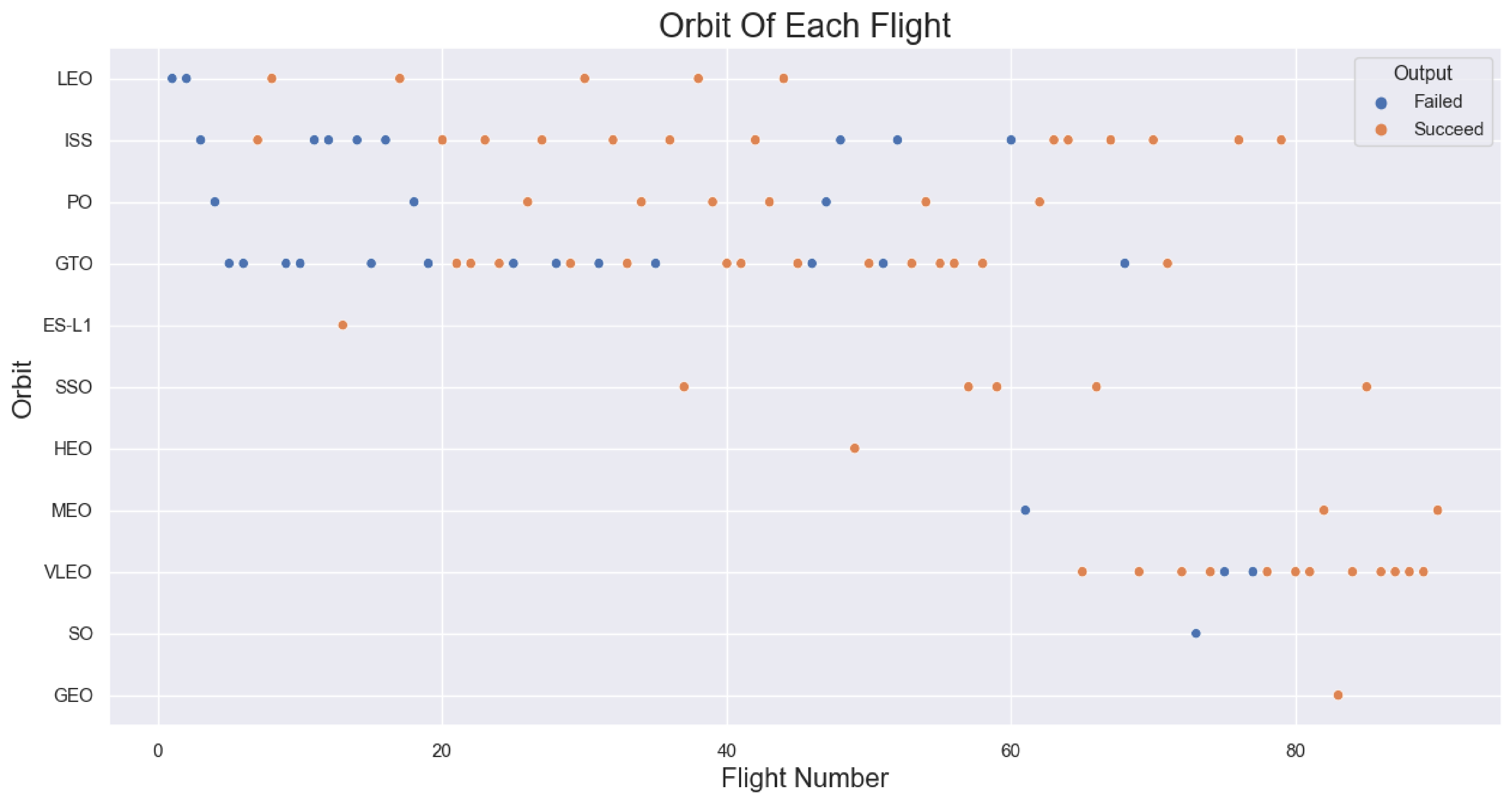
- We can notice that most missions had a payload less than 8 tons.
- Most missions with higher payload were successful.

Success Rate vs. Orbit Type



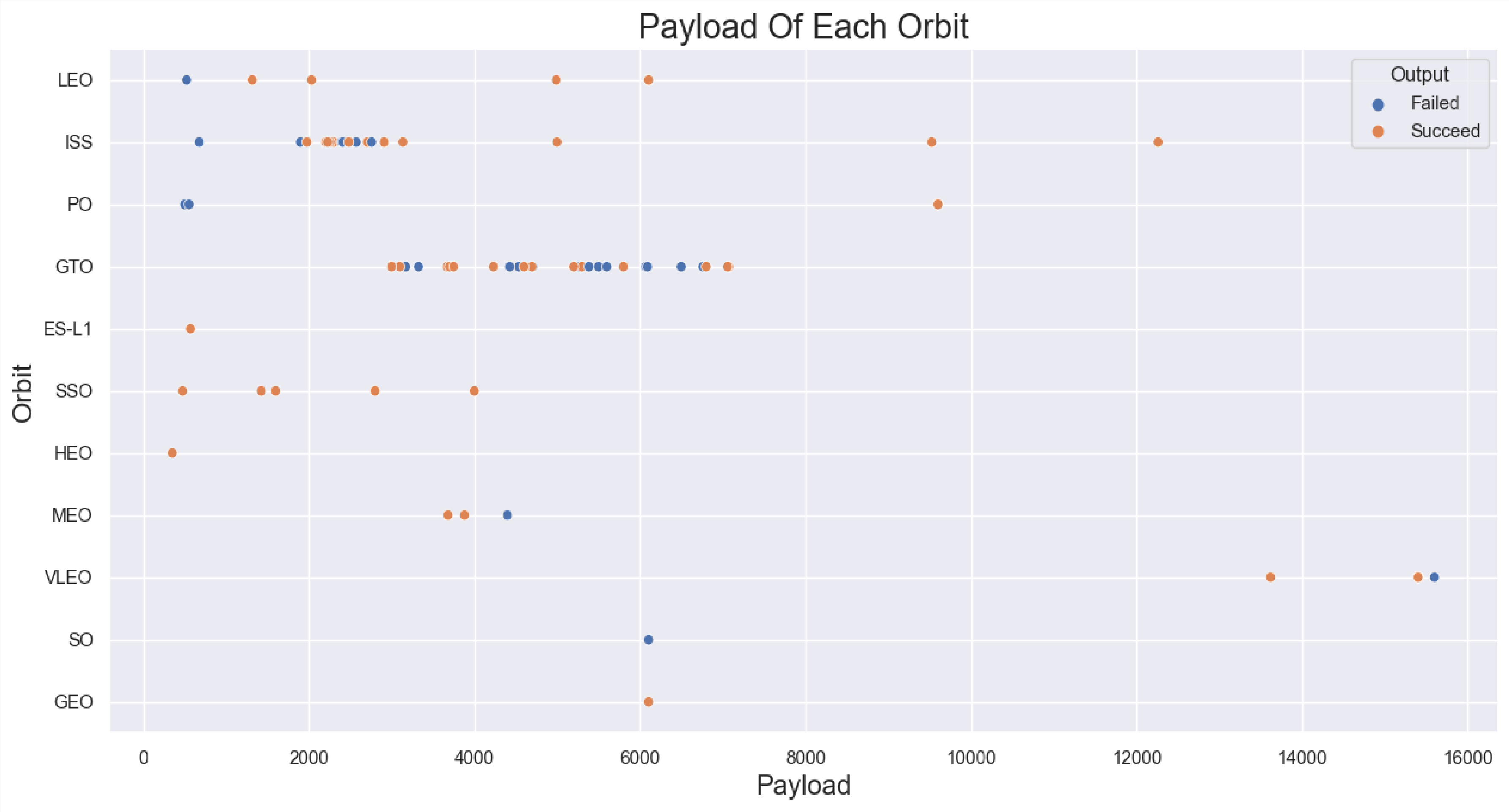
- If we ignore orbits with 1 missions we can see that SSO orbit has highest success rate followed by VLEO orbit.
- GTO orbit have approximately 50% success rate

Flight Number vs. Orbit Type



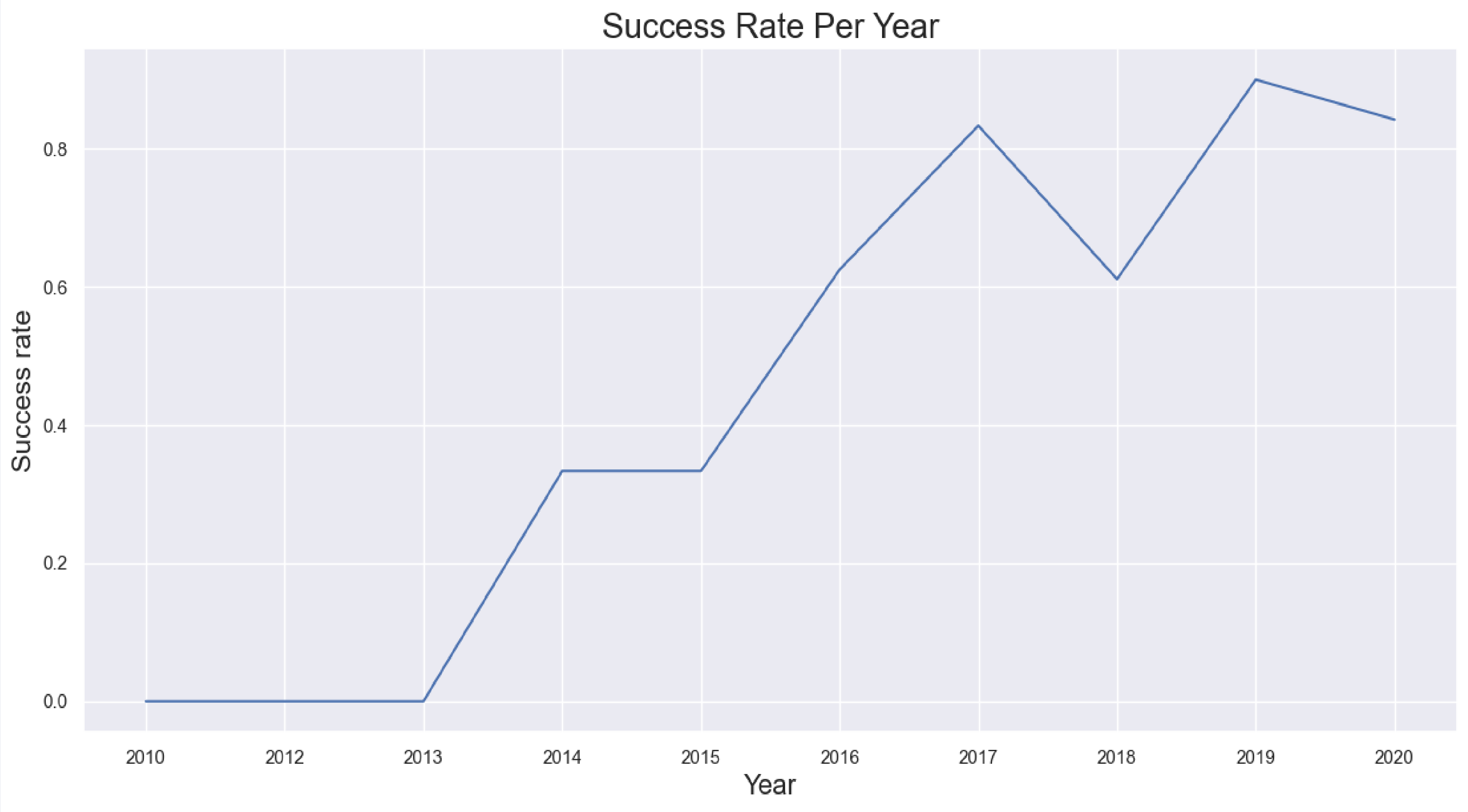
- We can notice that most missions was launched to low earth orbits like LEO, GTO and VLEO.
- Most of the latest missions were to VLEO may be for testing of the starlink project.

Payload vs. Orbit Type



- We can notice that most missions to ISS had a payload of less than 4 tons.
- All missions to GTO had a payload of less than 8 tons.
- All missions to VLEO had a payload of more than 11 tons.

Launch Success Yearly Trend



- We can see that the success rate have increased over the year with 2019 being the most successful year.

All Launch Site Names

In [6]:
```
%%sql
select distinct(launch_site) from spacex
```

```
 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[6]:

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

Launch Site Names Begin with 'CCA'

In [8]:

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

```
 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[8]:

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

In [22]:

```
%%sql
select sum(payload_mass__kg_) as payload_sum from spacex
where customer like '%CRS%'
```

```
 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[22]:

payload_sum
48213

Average Payload Mass by F9 v1.1

In [25]:

```
%%sql
select avg(payload_mass__kg_) as payload_avr from spacex
where booster_version like 'F9 v1.1%'
```

```
 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[25]:

payload_avr
2534

First Successful Ground Landing Date

```
In [27]: %%sql
         select min(date) as first_date from spacex
         where landing__outcome = 'Success (ground pad)'

 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[27]:

first_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
In [33]: %%sql
select booster_version, payload_mass__kg_ from spacex
where landing__outcome = 'Success (drone ship)' and (payload_mass__kg_ between 4000 and 6000)
```

```
 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[33]:

booster_version	payload_mass__kg_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

In [34]:

```
%%sql
select count(mission_outcome) as count_, mission_outcome from spacex
group by mission_outcome
```

```
 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[34]:

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

Boosters Carried Maximum Payload

In [38]:

```
%%sql
select distinct(booster_version), payload_mass__kg_ from spacex
where payload_mass__kg_ = (select max(payload_mass__kg_) from spacex)
```

```
 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[38]:

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

In [47]:

```
%%sql
select booster_version, landing__outcome, launch_site, date from spacex
where year(date) = '2015' and (landing__outcome like 'Failure%')
```

```
 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

Out[47]:

booster_version	landing__outcome	launch_site	DATE
F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40	2015-01-10
F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40	2015-04-14

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

In [41]:

```
%%sql
select count(landing__outcome) as count_outcome, landing__outcome from spacex
where date between '2010-06-04' and '2017-03-20'
group by landing__outcome
```

```
 * ibm_db_sa://cbw63011:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

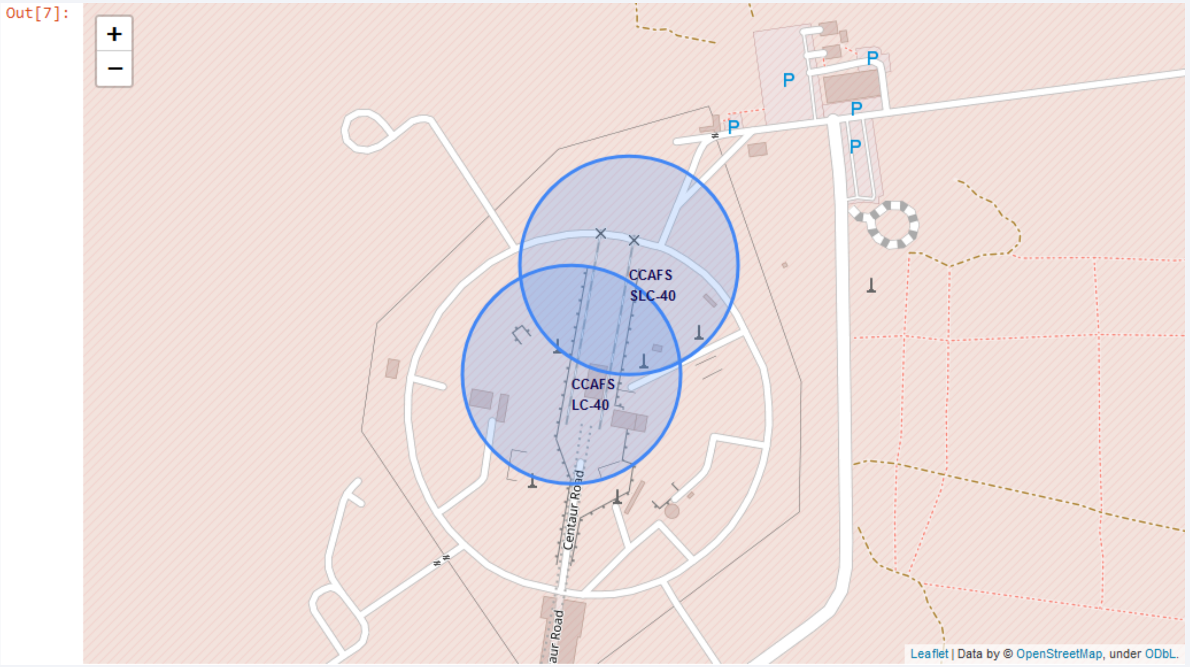
Out[41]:

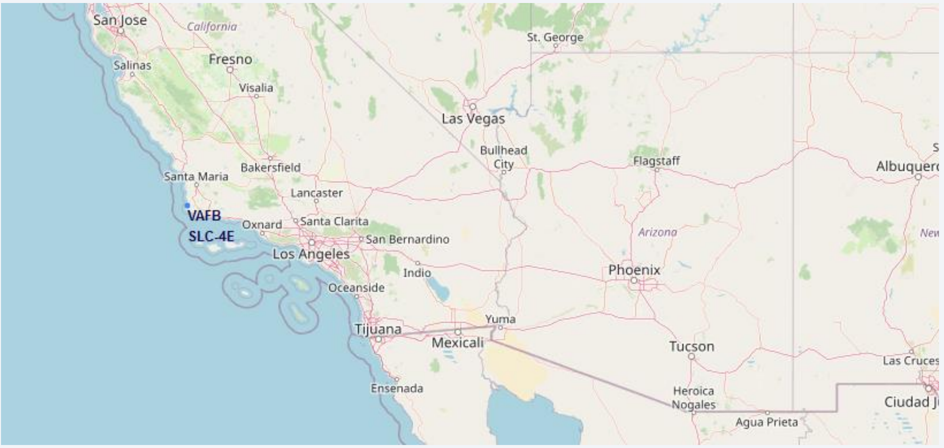
count_outcome	landing__outcome
3	Controlled (ocean)
5	Failure (drone ship)
2	Failure (parachute)
10	No attempt
1	Precluded (drone ship)
5	Success (drone ship)
3	Success (ground pad)
2	Uncontrolled (ocean)

Section 3

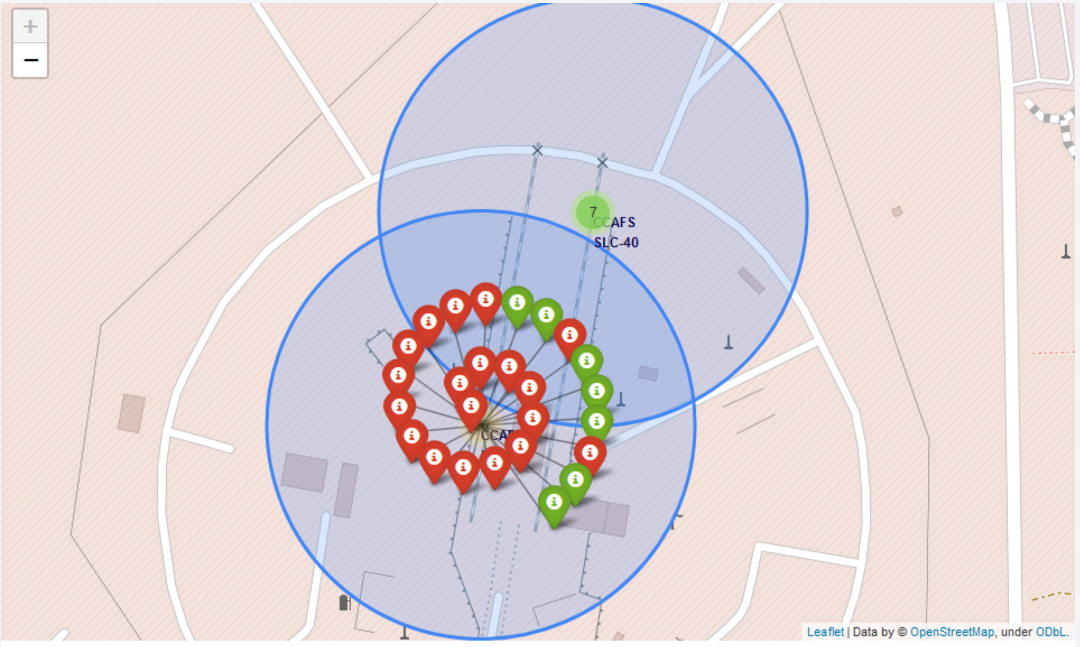
Launch Sites Proximities Analysis

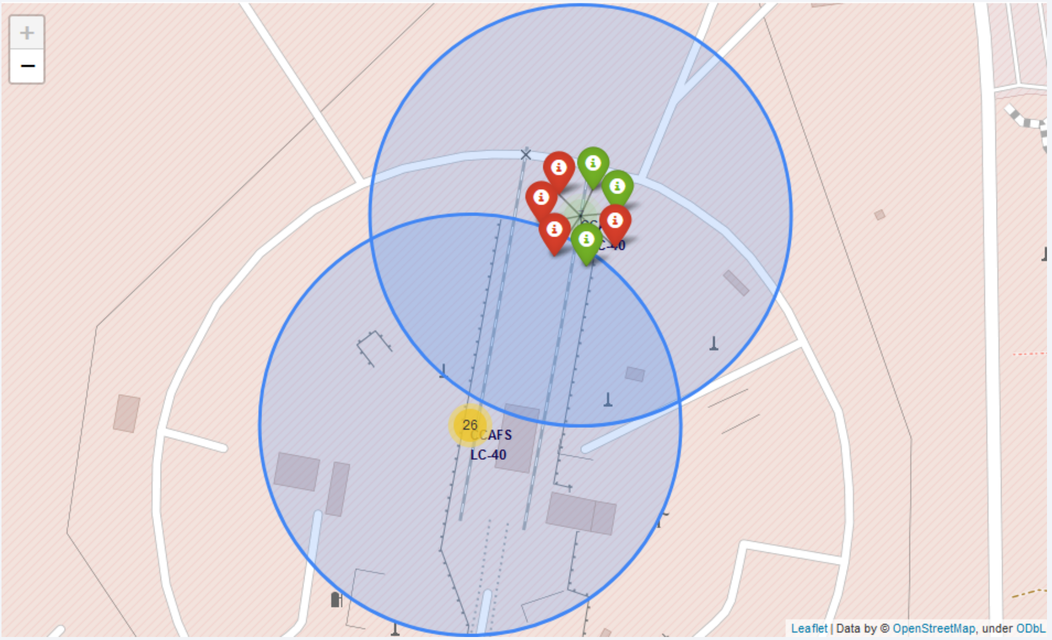
Locations of All Launch Sites.

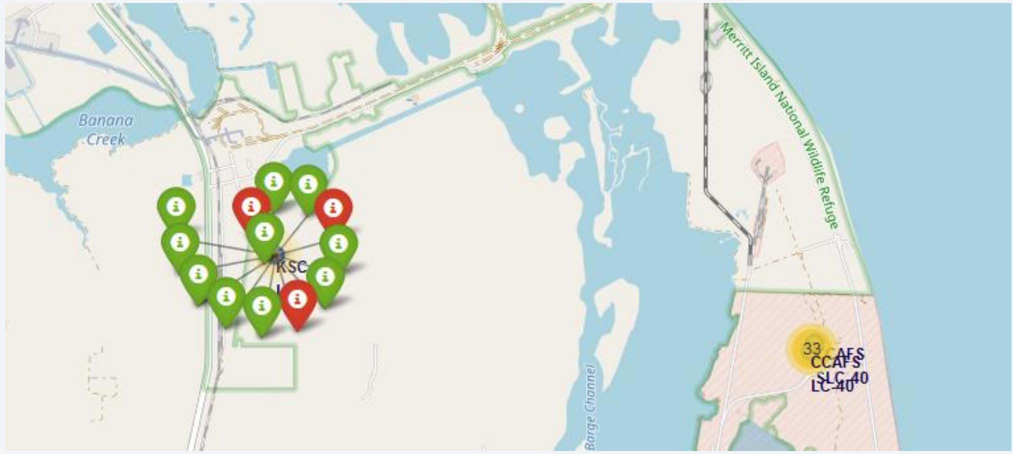




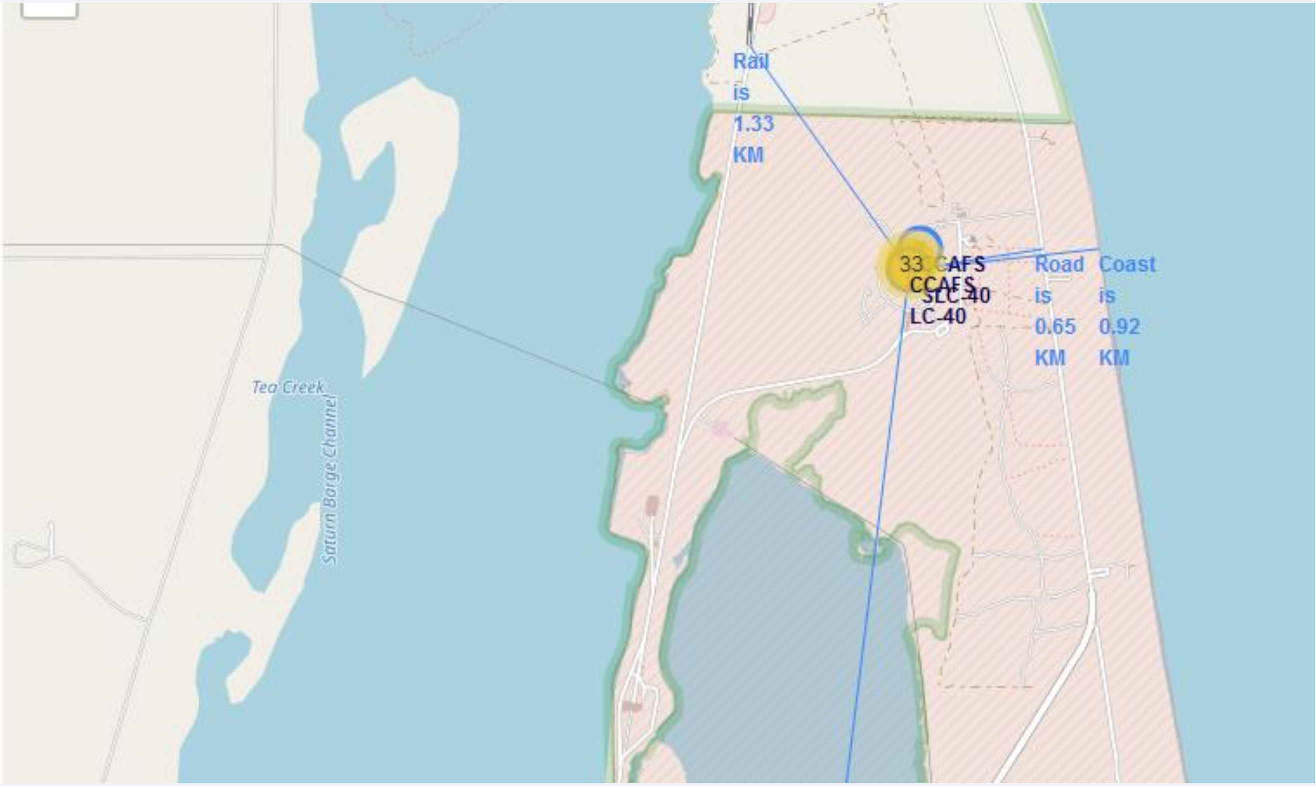
Mission Outcome Markers Added To Launch Sites.

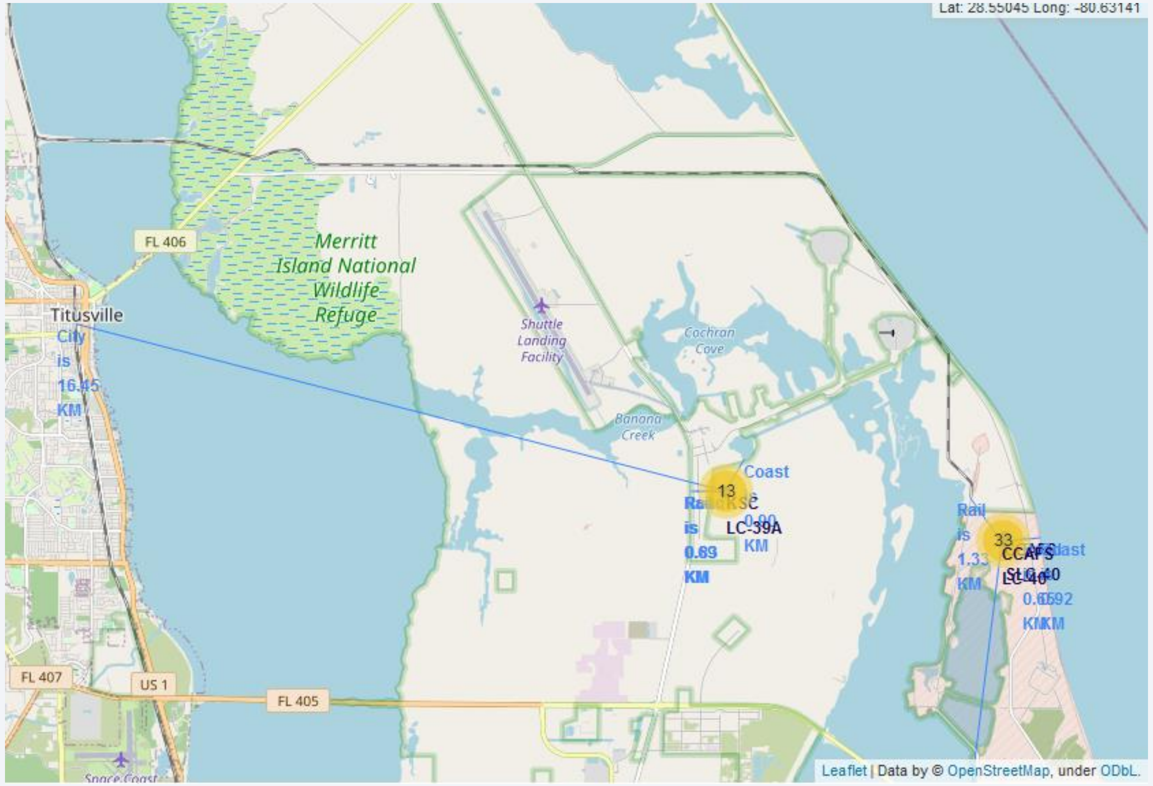






Launch Sites Distances to nearby facilities.

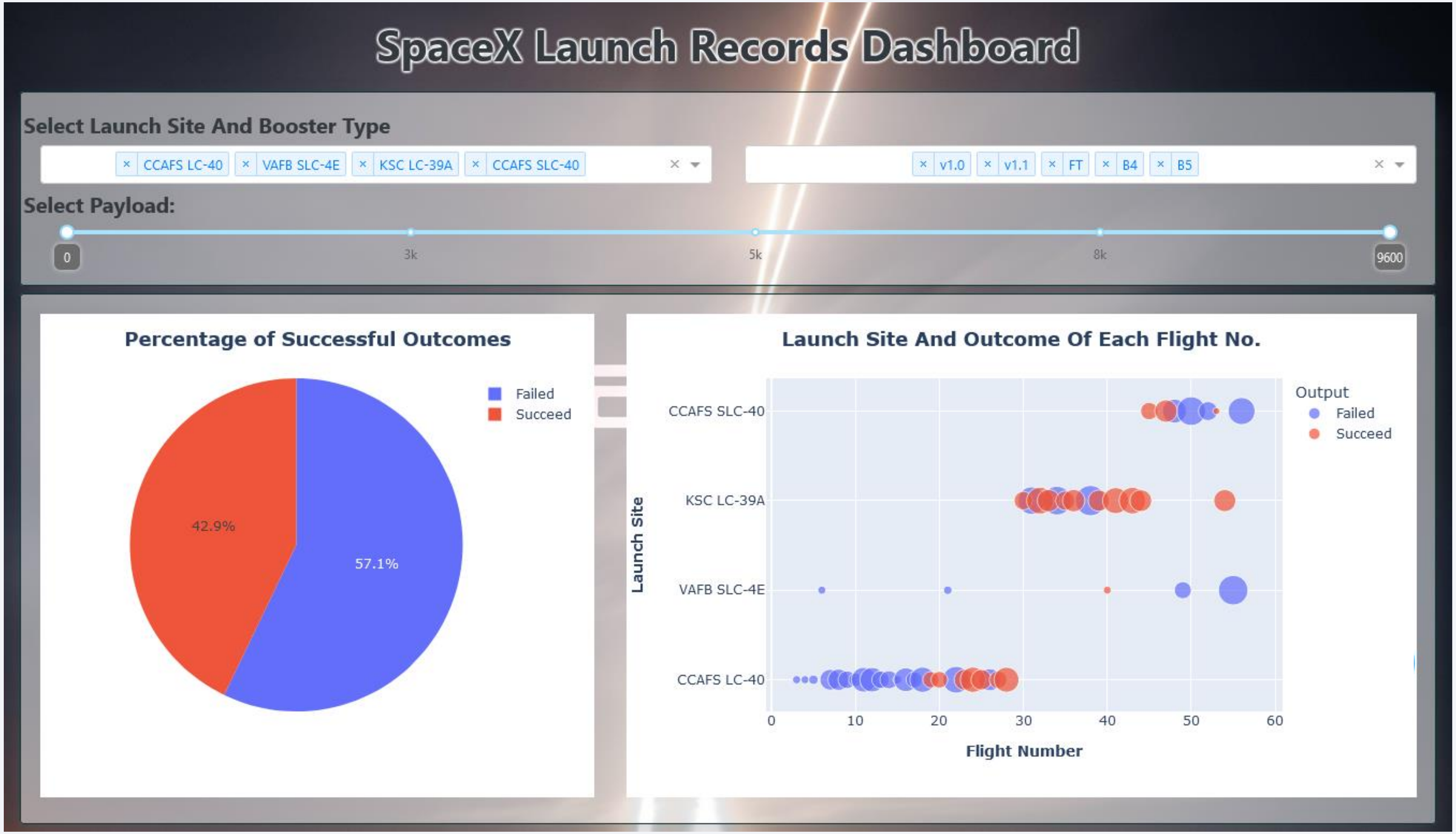




Section 4

Build a Dashboard with Plotly Dash

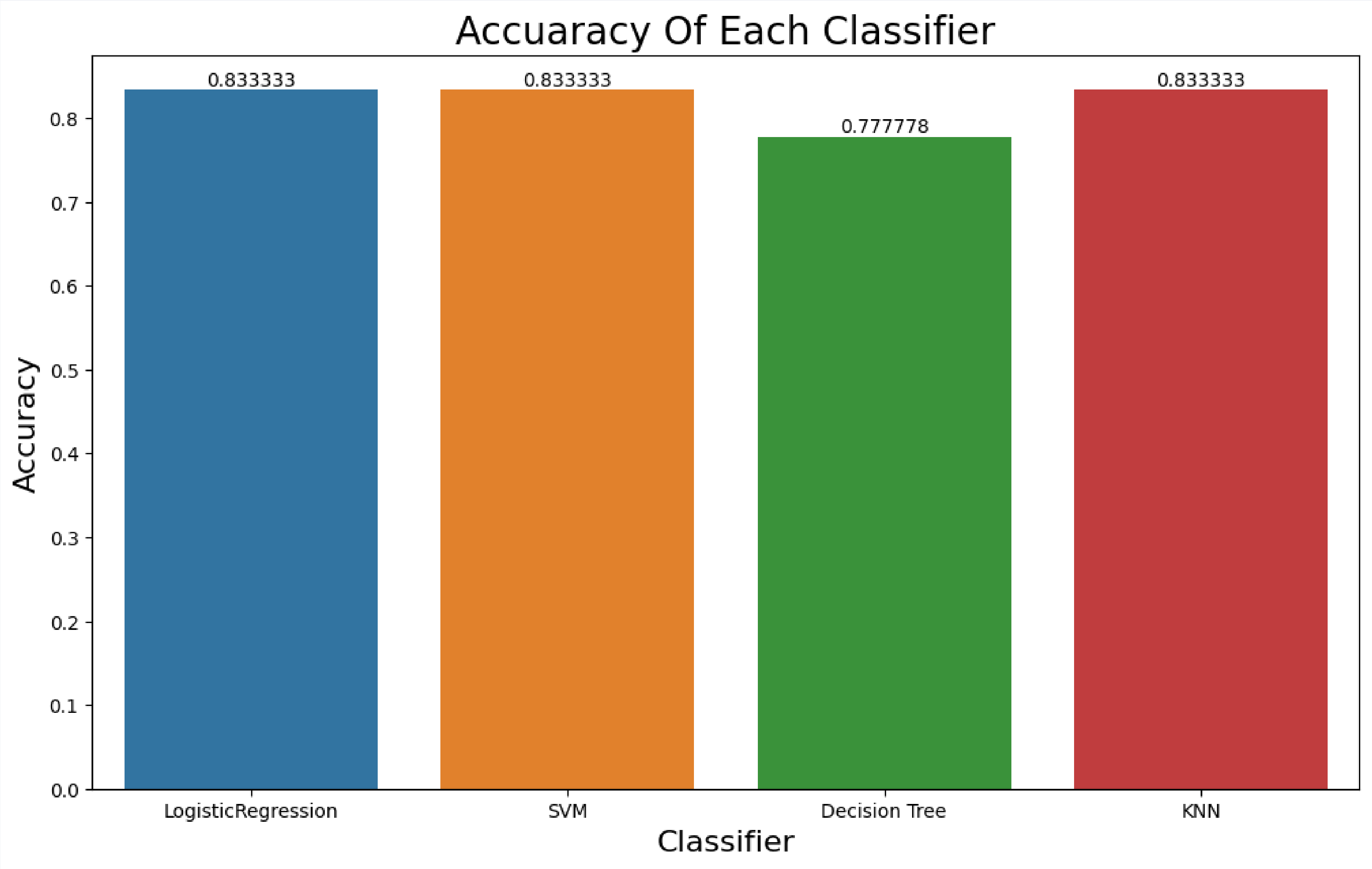
Plotly Dash Dashboard



Section 5

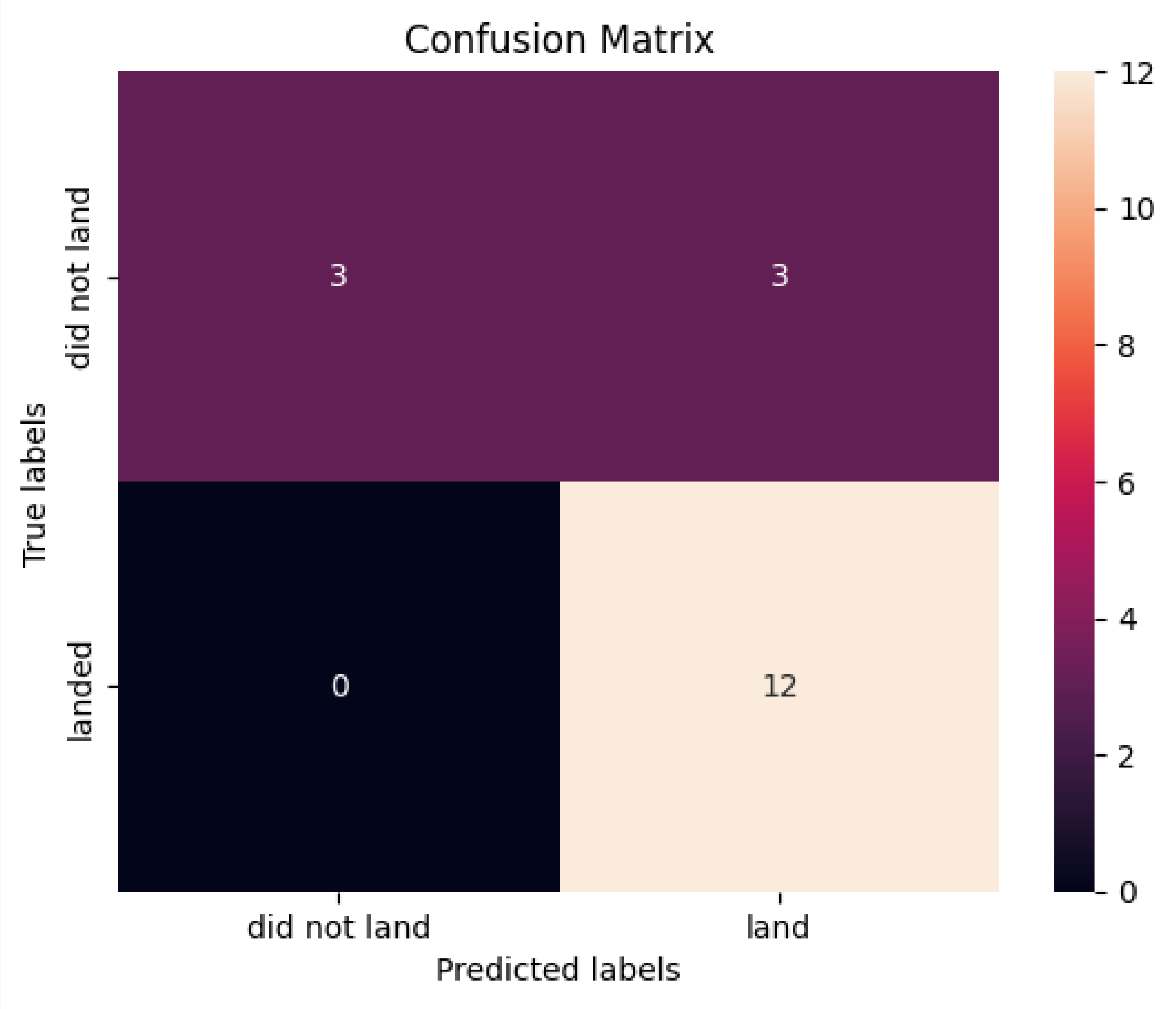
Predictive Analysis (Classification)

Classification Accuracy



- The output of Logistic Regression, SVM and KNN are all the same.

Confusion Matrix



Conclusions

- Successful landing have improved significantly over the years with the best year being 2019.
- CCAFS LC-40 is the most used launch site.
- All Launch sites are located far from the cities.
- Low earth orbit missions have more success on average
- Missions with payload more than 8 ton seem to be more successful.
- F9 v1.1 is the most powerful and successful booster yet.
- KNN was the best classifier to predict the data.

Thank you!