Winning Space Race with Data Science

Pablo Posada
<Date>

Outline

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

Executive Summary

Methodologies

- Data collection through API
- Data collection with web scraping
- Data wrangling and exploratory data analysis (EDA)
- Visualizations
- Predictive analysis with machine learning

Results

- Success/Fail plots
- SQL queries
- Launch site locations/history
- Machine learning model selection

Introduction

Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each

If we can determine if the first stage will land, we can determine the cost of a launch.

Much of the savings is because Space X can reuse the first stage.

https://github.com/p4pp19/SpaceX-Launcha

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collection using SpaceX API.
  - Collection using web scraping.
- Perform data wrangling
  - Extract Falcon 9 launch records HTML table from Wikipedia.
  - Parsing the table and conversion into a Pandas data frame.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

API

Request to the SpaceX API url → Decode the response content to create dataframe

Web scraping

Get HTTP response → Create a dataframe by parsing the HTML tables

Data Collection – SpaceX API

Start requesting rocket launch data from SpaceX API with the following URL.

To make the requested JSON results more consistent, we used a static response object.

We decoded the response content as a Json using .json and turn it into a Pandas dataframe using .json_normalize().

```
[17]: spacex_url="https://api.spacexdata.com/v4/launches/past"

[18]: response = requests.get(spacex_url)
```

```
[20]: static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork
```

```
[25]: # Use json_normalize meethod to convert the json result into a dataframe

      response=requests.get(static_json_url)
      response = response.json()
      response = pd.json_normalize(response)
```

Using the dataframe `data` print the first 5 rows

```
[26]: # Get the head of the dataframe
      data = pd.DataFrame(response)
      data.head()
```

https://github.com/p4pp19/SpaceX-Launcha/blob/main/W1-spacex-data-collection-api.ipynb

Data Collection - Scraping

TASK 1: Request the Falcon9 Launch Wiki page from its URL

Use HTTP GET method to request the Falcon9 Launch HTML page, then create a *BeautifulSoup* object from the HTML *response.*

TASK 2: Extract all column/ variable names from the HTML table header

Find all tables on the wiki page first.

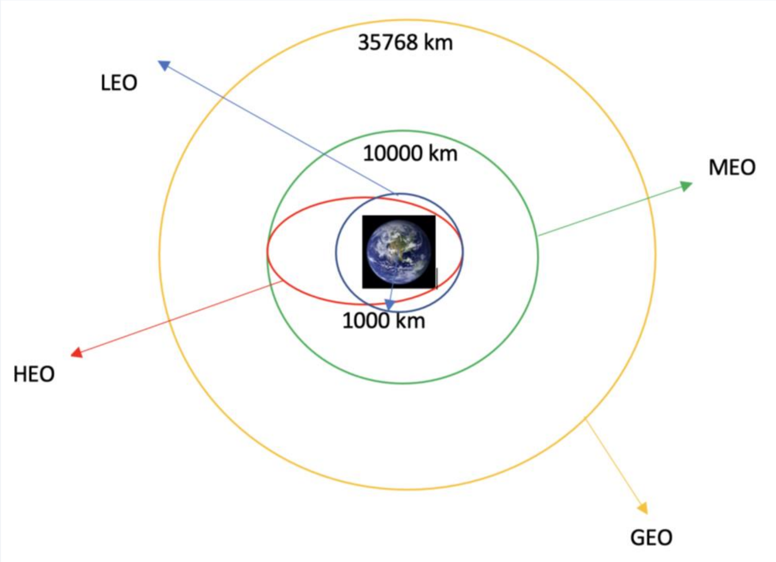
Iterate through the *<th>* elements and apply the *extract_column_from_header()*` function to extract column name one by one.

TASK 3: Create a data frame by parsing the launch HTML tables

Create an empty dictionary with keys from the extracted column names. This dictionary is converted into a Pandas dataframe.

https://github.com/p4pp19/SpaceX-Launcha/blob/main/W1-spacex-webscraping.ipynb

Data Wrangling



TASK 2: Calculate the number and occurrence of each orbit

TASK 3: Calculate the number and occurrence of mission outcome per orbit type

TASK 1: Calculate the number of launches on each site

TASK 4: Create a landing outcome label

https://github.com/p4pp19/SpaceX-Launcha/blob/main/W1-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Bar charts

To represent numerical and categorical variables grouped in intervals

Line plots

To emphasize changes in values for one variable (plotted on the vertical axis) for continuous values of a second variable (plotted on the horizontal)

Scatter plots

To express a trend

https://github.com/p4pp19/SpaceX-Launcha/blob/main/W2-ExpDatAnal-datavisualizations.ipynb

EDA with SQL

- 1. Display the names of the unique launch sites in the space mission
- 2. Display 5 records where launch sites begin with the string 'CCA'
- 3. Display the total payload mass carried by boosters launched by NASA (CRS)
- 4. Display average payload mass carried by booster version F9 v1.1
- 5. List the date when the first succesful landing outcome in ground pad was acheived.

https://github.com/p4pp19/SpaceX-Launcha/blob/main/W2-ExplDatAn-sql.ipynb

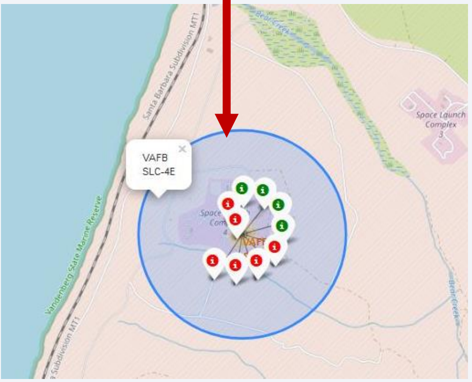
EDA with SQL

- 6. Display the names of the unique launch sites in the space mission
- 7. List the total number of successful and failure mission outcomes
- 8. List the names of the booster versions which have carried the maximum payload mass
- 9. List the failed landing outcomes in drone ship, their booster versions, and launch site names for the in year 2015
- 10. 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

https://github.com/p4pp19/SpaceX-Launcha/blob/main/W2-ExplDatAn-sql.ipynb

Build an Interactive Map with Folium

Circles for places



Markers for launchings

Color for success status

Lines to illustrate distances

https://github.com/p4pp19/SpaceX-Launcha/blob/main/W3-Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Predictive Analysis (Classification)

1. Perform exploratory Data Analysis

X Y

2. Determine Training Labels

3. Find best Hyperparameters for each:

Logistic Regression

SVM

Classification Trees

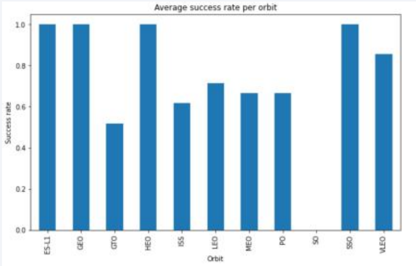
K-nearest neighbors

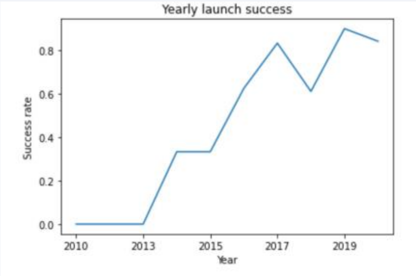
4. Choose the best model

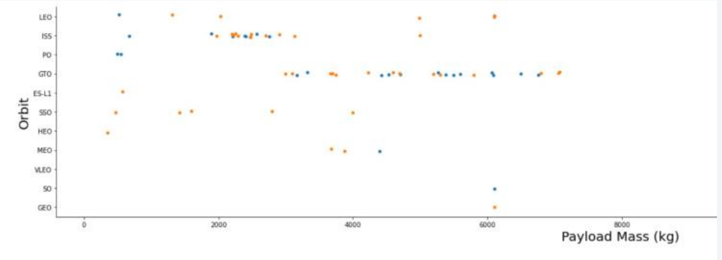
https://github.com/p4pp19/SpaceX-Launcha/blob/main/W4-SpaceX_Machine%20Learning%20Predictions.ipynb

Results

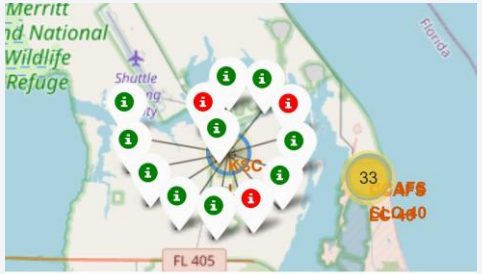
EDA



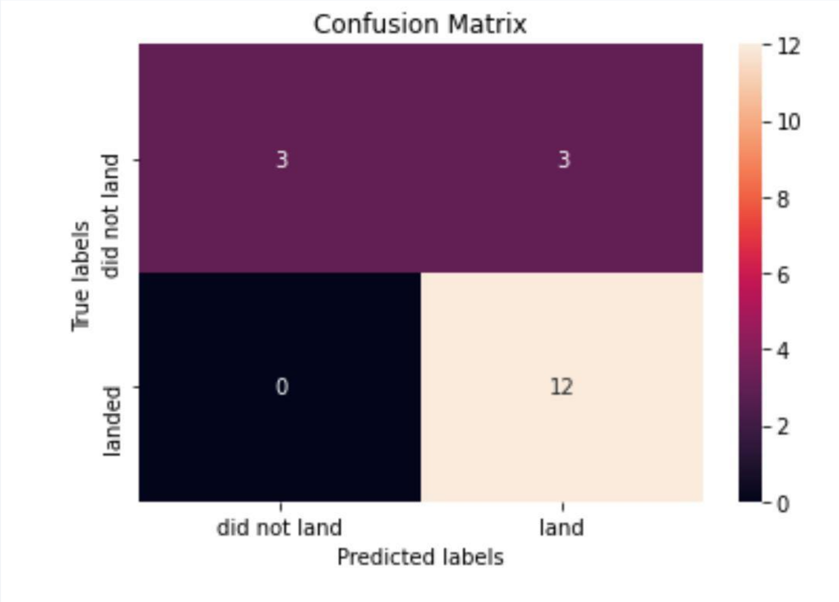




Interactive Analytics



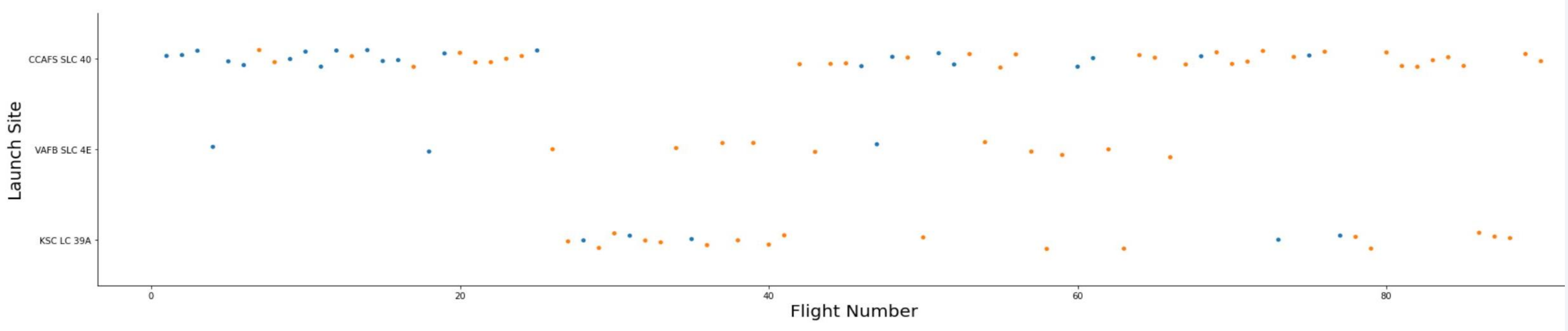
Predictive Analysis



Section 2

Insights drawn from EDA

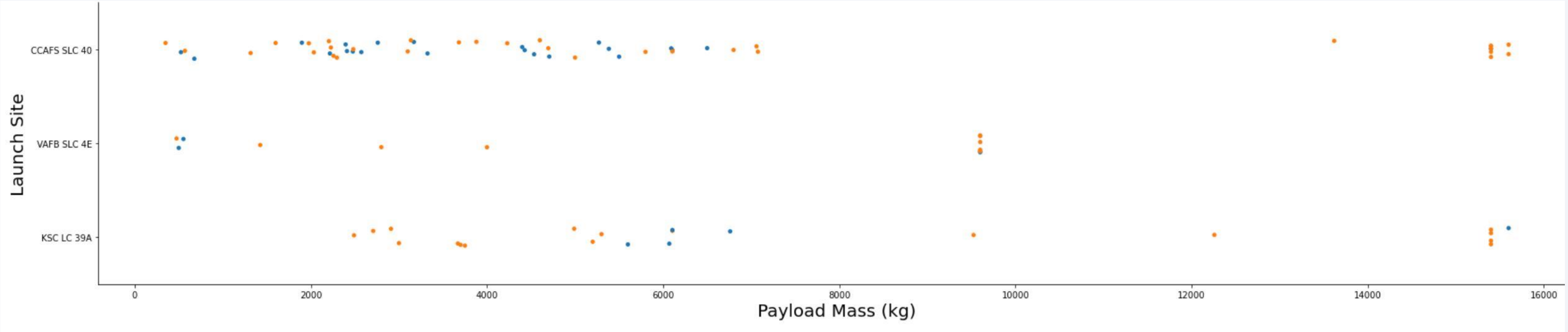
Flight Number vs. Launch Site



Fail   Success

- VAFB SLC 4E is the least used Launch site
- Most of missions were successful after Flight 35

Payload vs. Launch Site

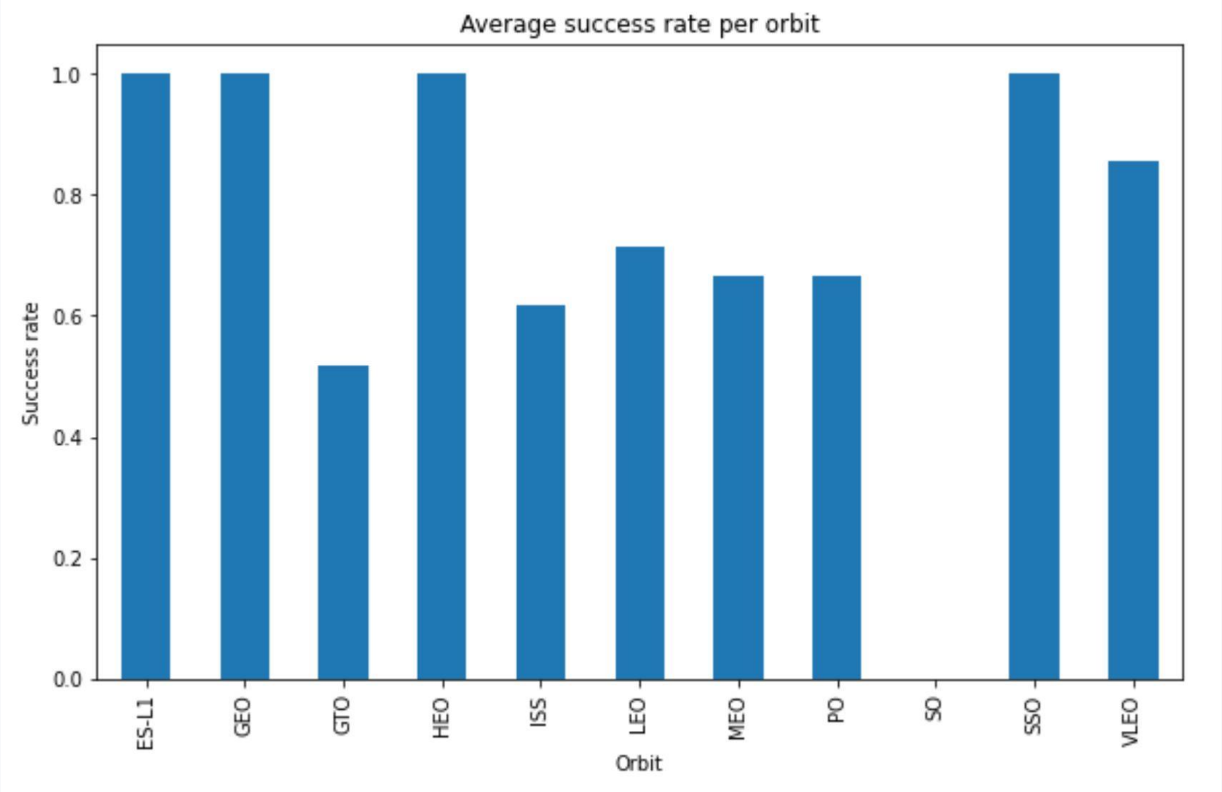


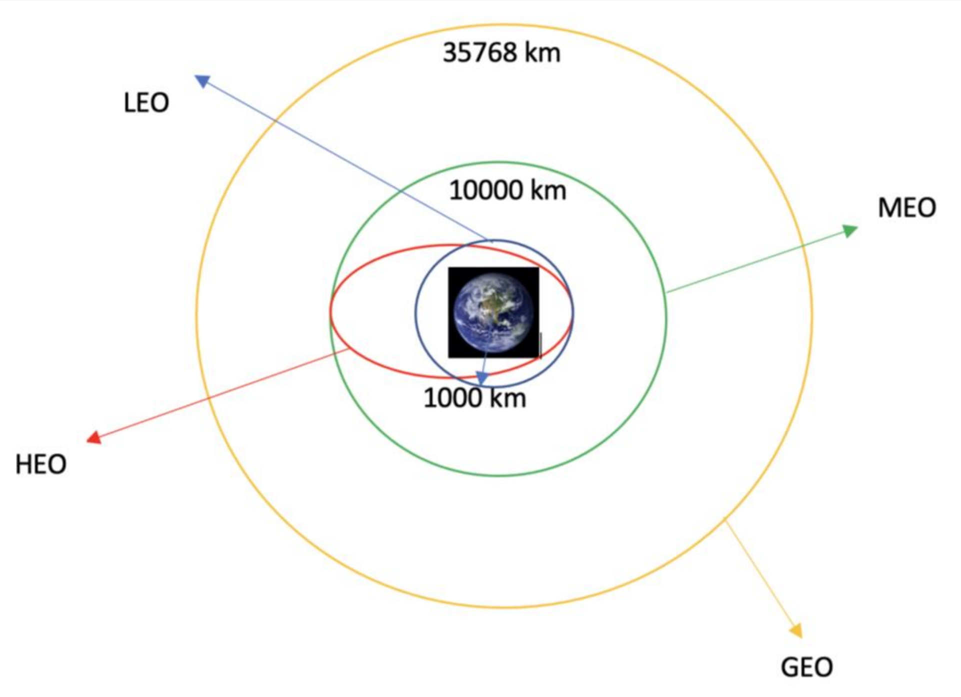
Fail

Success

- Missions carrying more than 9000 kg are more likely to succeed
- The heaviest payloads are launched from CCAFS SLC 40 and KSC LC 39A

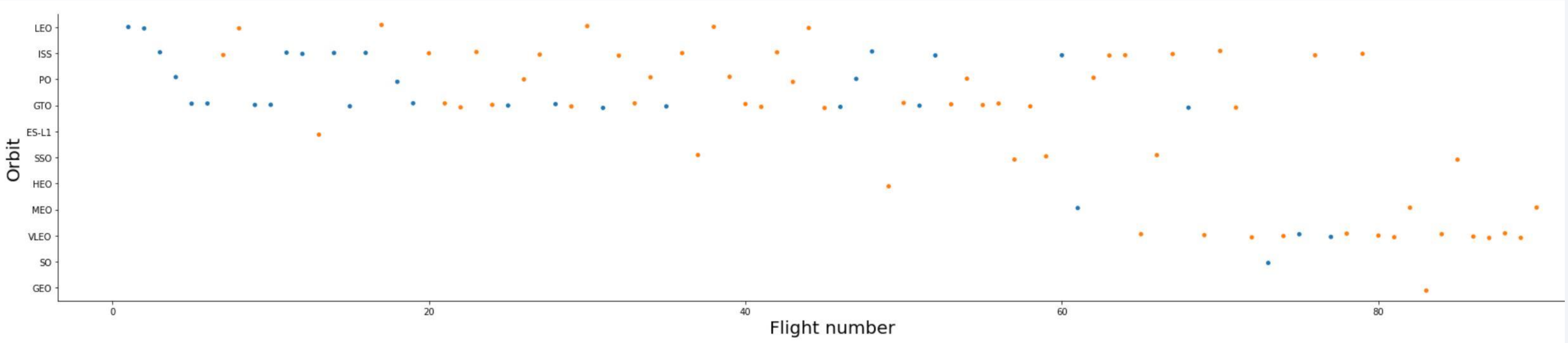
Success Rate vs. Orbit Type





- ES-L1, GEO, HEO and SSO have almost perfect success rates
- Other orbits oscillate around 60% success

Flight Number vs. Orbit Type

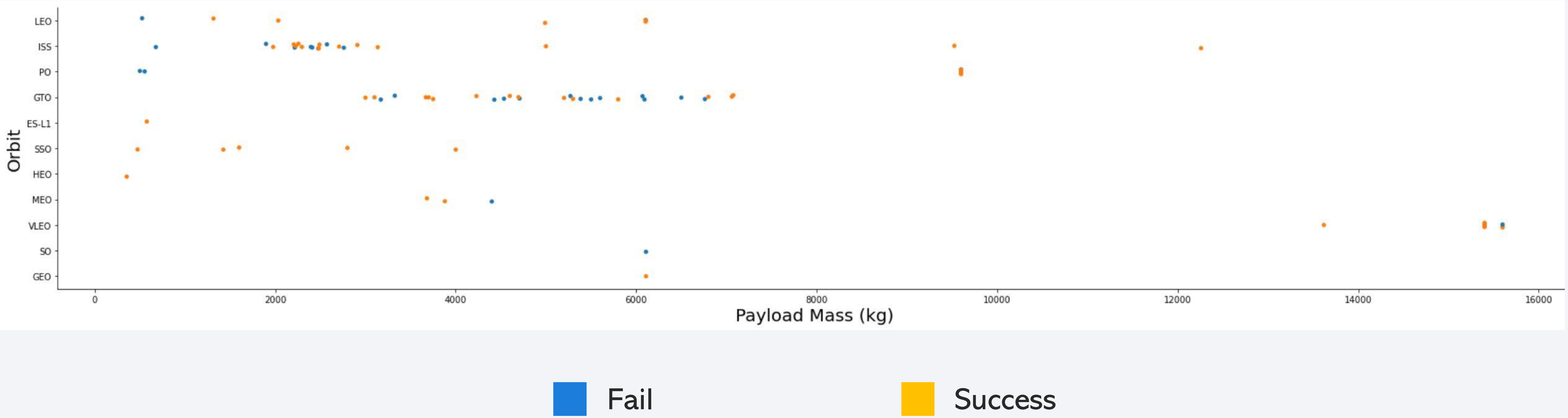


Fail

Success

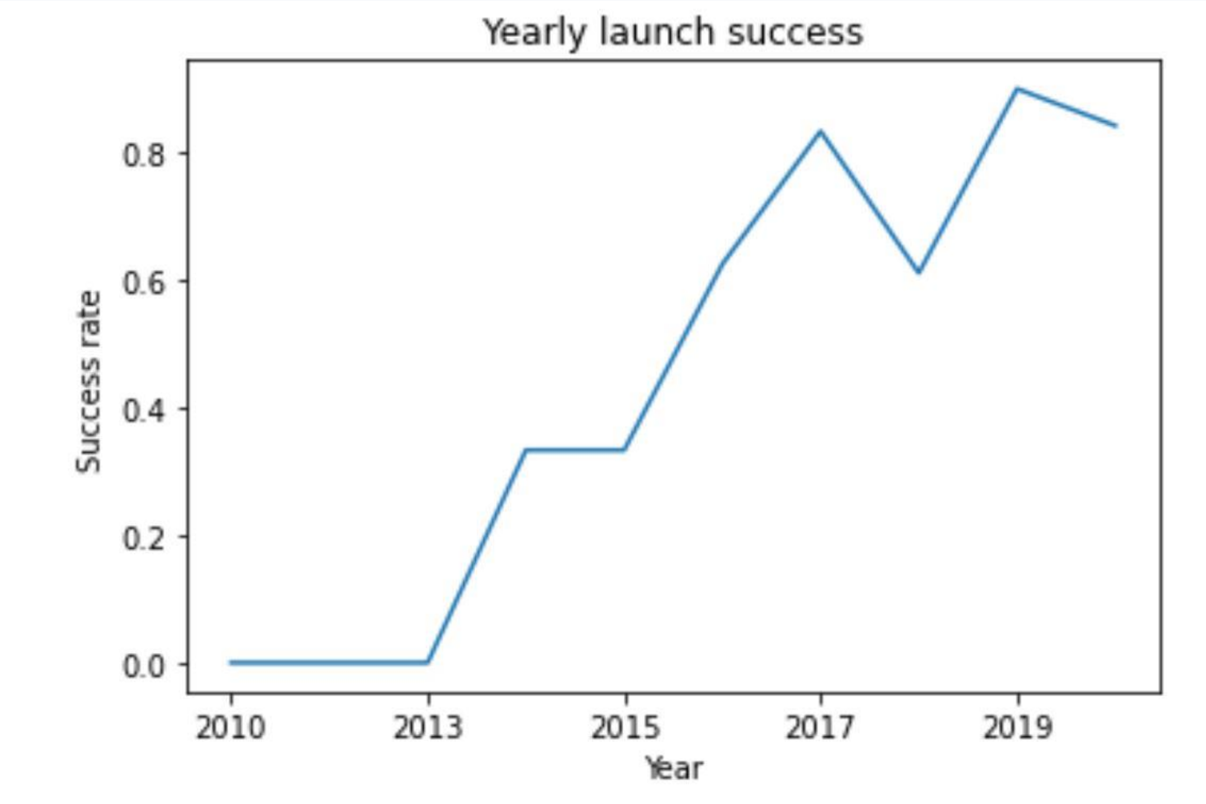
- LEO missions improved considerably over time
- There has been persistent problems related to missions going to GTO orbit

Payload vs. Orbit Type



- Heavier payloads have a negative effect on GTO missions and a positive effect on PO, ISS and LEO

Launch Success Yearly Trend



- Success rates kept increasing from 2013 to 2019

All Launch Site Names

```
%%sql
select distinct(LAUNCH_SITE) from SPAVCEXTBL
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.
Done.
```

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

- We find 4 unique launch sites

Launch Site Names Begin with 'CCA'

```
%%sql
select * from SPAVCEXTBL
where LAUNCH_SITE like 'CCA%' limit 5
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/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

- First 5 mission records where launch sites begin with `CCA`

Total Payload Mass

```
%%sql
select sum(payload_mass__kg_) as total_payload_mass  from SPAVCEXTBL
where customer ='NASA (CRS)'
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3s
Done.
```

total_payload_mass
45596

- The total payload (in kg) carried by boosters from NASA

Average Payload Mass by F9 v1.1

```
%%sql
select avg(payload_mass__kg_) as average_payload_mass  from SPAVCEXTBL
where booster_version ='F9 v1.1'
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3s
Done.
```

average_payload_mass
2928

- booster version F9 v1.1 has carried an average payload mass of 2928 kg

First Successful Ground Landing Date

```
%%sql
select DATE as Date, landing__outcome from SPAVCEXTBL
where landing__outcome =
(select landing__outcome from SPAVCEXTBL where landing__outcome like'%pad%' limit 1 )
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.dat
Done.
```

DATE	landing__outcome
2015-12-22	Success (ground pad)
2016-07-18	Success (ground pad)
2017-02-19	Success (ground pad)
2017-05-01	Success (ground pad)
2017-06-03	Success (ground pad)
2017-08-14	Success (ground pad)
2017-09-07	Success (ground pad)
2017-12-15	Success (ground pad)
2018-01-08	Success (ground pad)

- First successful landing outcome on ground pad achieved on 2015/12/22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%%sql
select BOOSTER_VERSION,LANDING__OUTCOME from SPAVCEXTBL
where LANDING__OUTCOME = 'Success (drone ship)'
and (PAYLOAD_MASS__KG_ between 4000 and 6000 )
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.clou
Done.
```

booster_version	landing__outcome
F9 FT B1022	Success (drone ship)
F9 FT B1026	Success (drone ship)
F9 FT B1021.2	Success (drone ship)
F9 FT B1031.2	Success (drone ship)

- There is a selection of booster versions that operates better when carrying payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

```
%%sql
SELECT
    COUNT(CASE WHEN MISSION_OUTCOME like'%Success%' then 1 ELSE NULL END) as "Success",
    COUNT(CASE WHEN MISSION_OUTCOME like'%Failure%' then 1 ELSE NULL END) as "Failure"
from SPAVCEXTBL
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.dat
Done.
```

Success	Failure
100	1

- Outstanding success rate!

Boosters Carried Maximum Payload

```
%%sql
select BOOSTER_VERSION,PAYLOAD_MASS__KG_ from SPAVCEXTBL
where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPAVCEXTBL)
order by BOOSTER_VERSION
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tg
Done.
```

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

- 12 booster versions have carried the maximum payload mass

2015 Launch Records

```
%%sql
select DATE, LANDING__OUTCOME,BOOSTER_VERSION, LAUNCH_SITE from SPAVCEXTBL
where LANDING__OUTCOME = 'Failure (drone ship)'
and YEAR(DATE) = 2015
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu
Done.
```

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

- Failed landing outcomes in drone ship during year 2015

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

```
%%sql
select LANDING__OUTCOME, count(*) as count from SPAVCEXTBL
where DATE between '2010-06-04' AND '2017-03-20'
group by LANDING__OUTCOME
order by 2 desc
```

```
 * ibm_db_sa://fyn91381:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a0
Done.
```

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

- Count of landing outcomes between 2010-06-04 and 2017-03-20

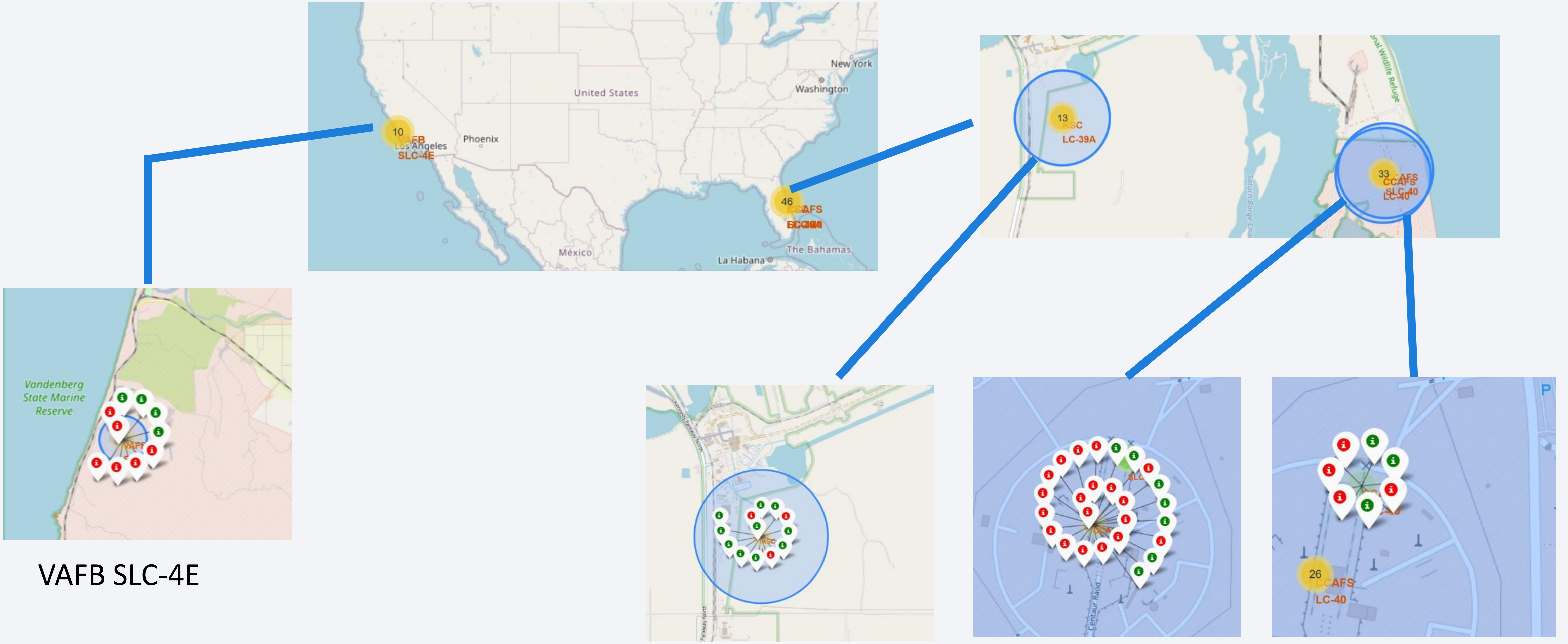
Section 4

Launch Sites Proximities Analysis

Launch sites – general

- 1 Launch site located on the west coast and 3 on the east coast
- All Launch sites as close as possible to the equator line

Mission Success



VAFB SLC-4E

KSC LC-39A

CCAFS LC-40

CCAFS SLC-40

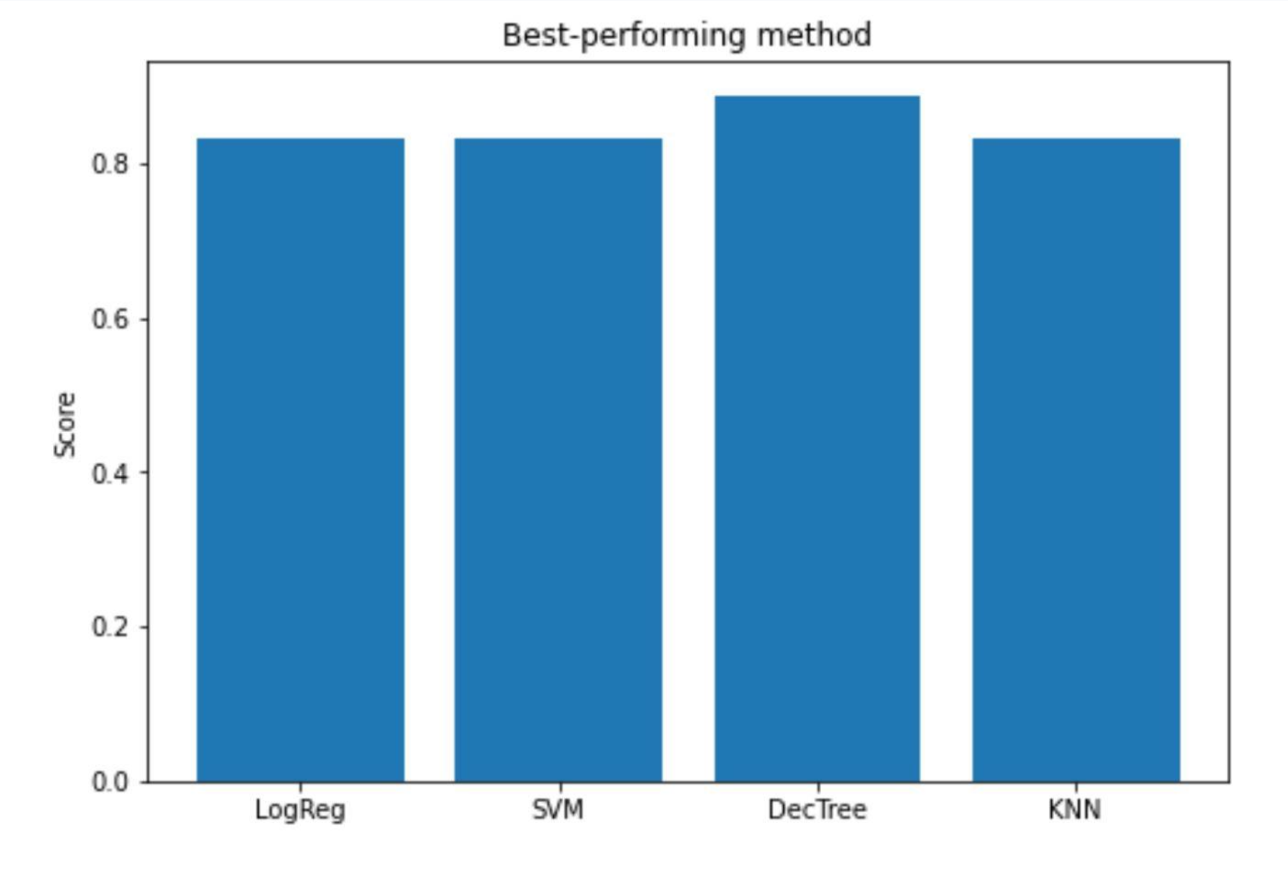
VAFB SLC-4E proximities

- Railway, highway and coastline are all less than 1.5 km apart from the launch site.

Section 6

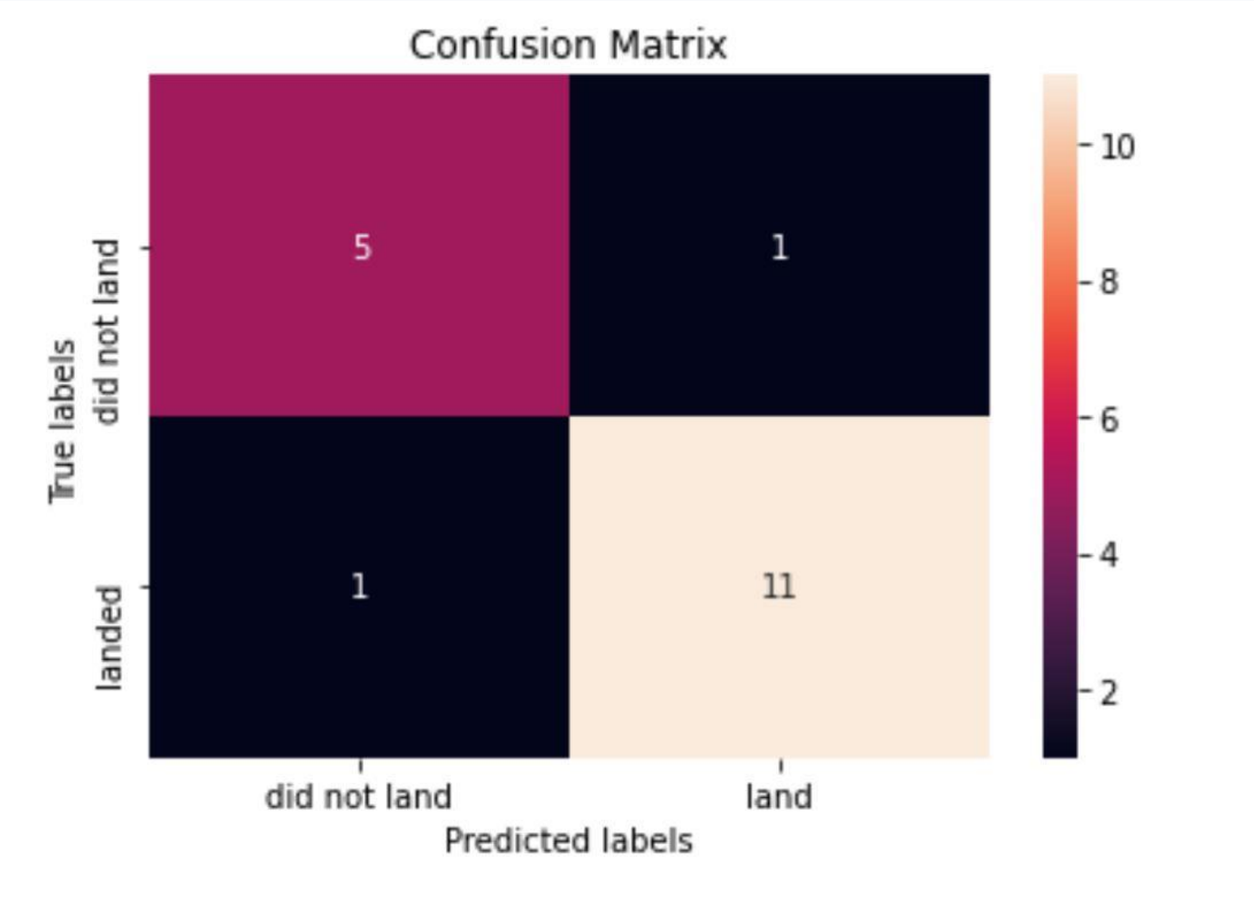
Predictive Analysis (Classification)

Classification Accuracy



- The Decision Tree model is the best-suited for our dataset with an accuracy score of 0.8888

Confusion Matrix



- Successful landings:

11 correct

1 incorrect

- Unsuccessful landings:

5 correct

1 incorrect

Conclusions

- Mission success rate has increased dramatically over the past 7 years
- It is more likely to have a successful mission if you are sending heavier payloads
- Booster version must be considered when carrying payloads less than 6000 kg
- Data suggests that heaviest payloads should be launched from CCAFS SLC 40 and KSC LC 39A
- The best-suited model to predict whether a future launching would be successful or not was determined to be the decision tree classifier

Thank you!