IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ali Kaan Bilge
22.04.2025

Outline

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

Executive Summary

Summary of methodologies

- Data collection
- Data wrangling
- Exploratory Data Analysis with Data Visualization
- Exploratory Data Analysis with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Summary of all results

- Exploratory Data Analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Introduction

Project background

- SpaceX is the most successful company of the commercial space age, making space travel affordable.
- The company 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.
- Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.

Questions to be answered

- How do variables such as payload mass, launch site, number of lights, and orbits affect the success of the first stage landing?
- Does the rate of successful landings increase over the years?
- What is the best algorithm that can be used for binary classification in this case?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Using SpaceX Rest API and Web Scrapping from Wikipedia
- Perform data wrangling
  - Filtering the data
  - Dealing with missing values
  - Using One Hot Encoding to prepare the data to a binary classification
- 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 and evaluation of classification models to ensure the best results

Data Collection – SpaceX API

Requesting Data from SpaceX API

Decoding the response content and turning into dataframe

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

Constructing data into a dictionary

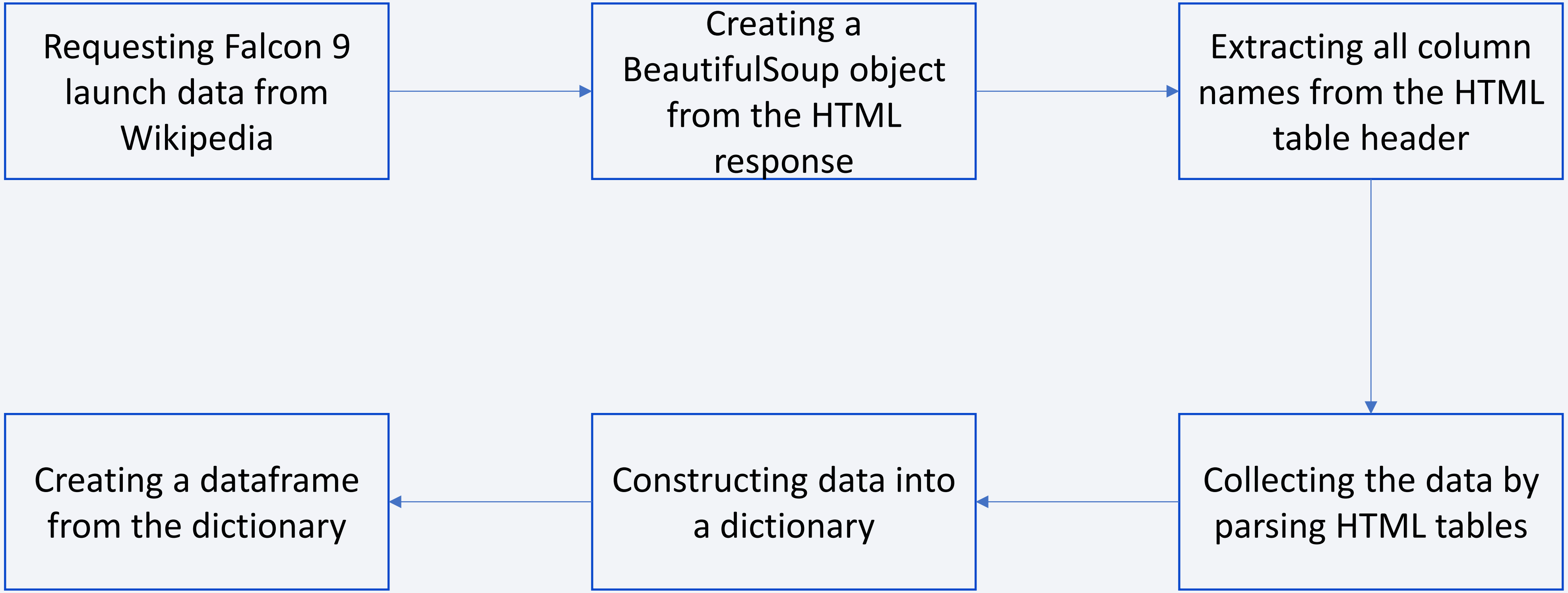
Creating a dataframe from the dictionary

Filtering the dataframe to only include Falcon 9 launches

Replacing missing values of Payload Mass column

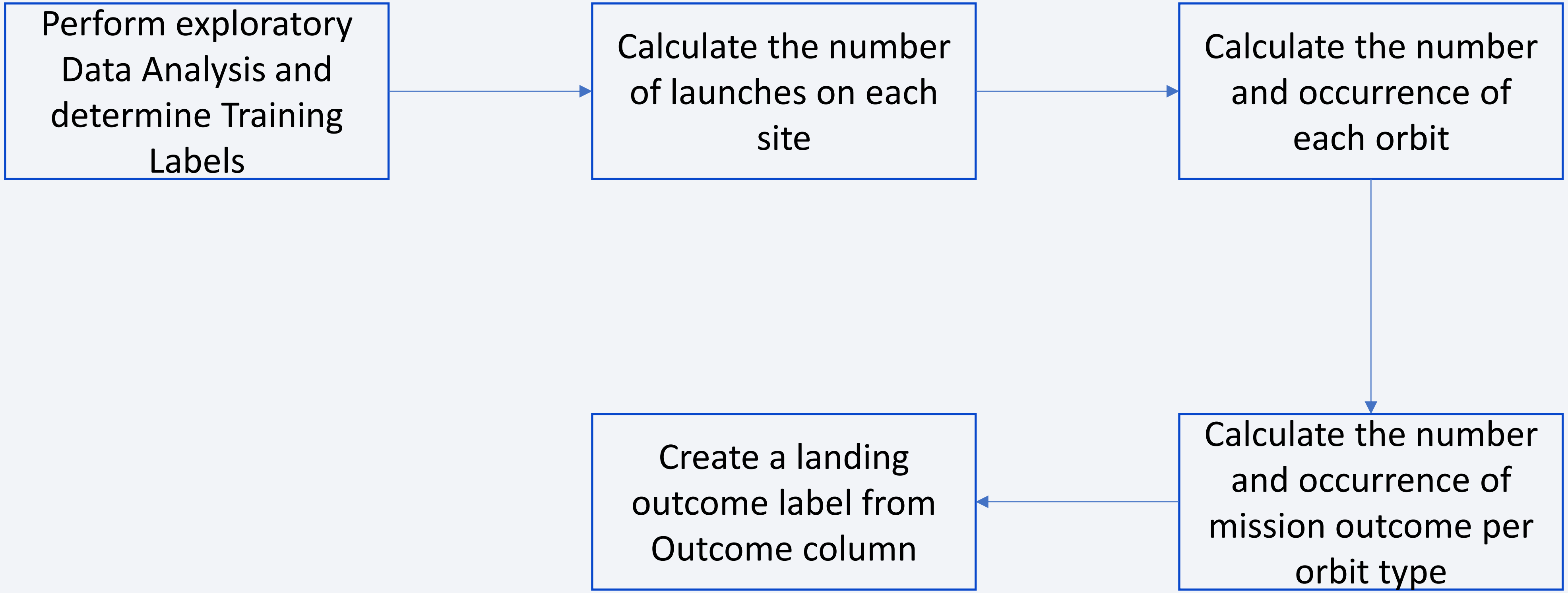
https://github.com/AliKaanBilge/IBM-Capstone-Project/blob/676981138e3664695c4296f2a0e2dd492da9b6d4/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping



https://github.com/AliKaanBilge/IBM-Capstone-Project/blob/d8fa1a377b314a18985360f37c0b39ec6fd4e02f/jupyter-labs-webscraping.ipynb

Data Wrangling



https://github.com/AliKaanBilge/IBM-Capstone-Project/blob/332b2663cd90e8a84c14b5ea95a40e08a61d0281/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Charts

- 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
- Success Rate Yearly Trend

https://github.com/AliKaanBilge/IBM-Capstone-Project/blob/d1c9035d9b7a1e908545dd2e6fe8c451640d3a9c/edadataviz.ipynb

EDA with SQL

- Displaying the names of the unique launch sites in the space mission
- Displaying 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 between the date 2010-06-04 and 2017-03-20 in descending order

https://github.com/AliKaanBilge/IBM-Capstone-Project/blob/77ea37f670b54bf28e97c30af4a6f3ede058e639/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers of all Launch Sites
- Colored Markers of the launch outcomes for each Launch Site
- Distances between a Launch Site to its proximities

https://github.com/AliKaanBilge/IBM-Capstone-Project/blob/e3bba1c0daffaef08b3e328ba09be178e9f7ebdc/lab_jupyter_launch_site_location.ipynb

Predictive Analysis (Classification)

1. Creating a NumPy array from the column “Class” in data
2. Standardizing the data with StandardScaler, then fitting and transforming it
3. Splitting the data into training and testing sets
4. Creating a GridSearchCV object with cv = 10 to find the best parameters
5. Applying GridSearchCV on LogReg, SVM, Decision Tree, and KNN models
6. Calculating the accuracy on the test data for all models
7. Examining the confusion matrix for all models
8. Finding the method performs best

https://github.com/AliKaanBilge/IBM-Capstone-Project/blob/5aed04649e307170b9e21be766990792ffa76da6/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

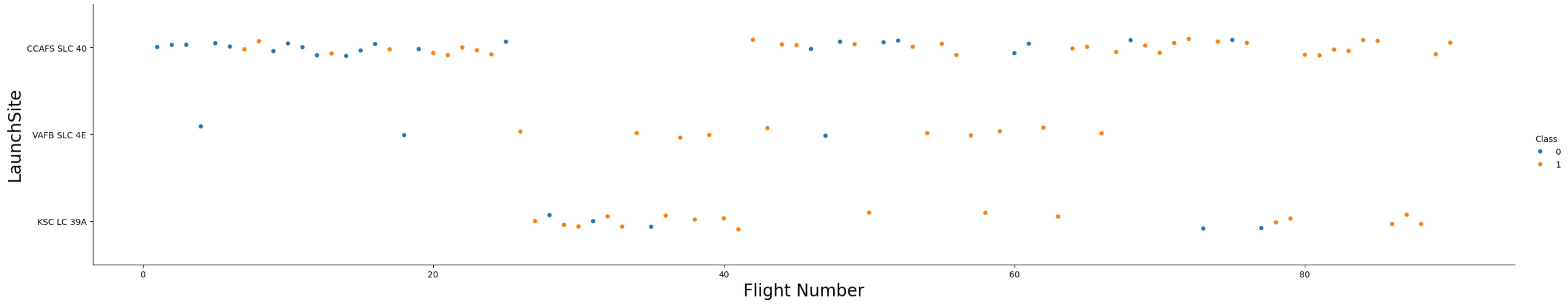
Results

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

Section 2

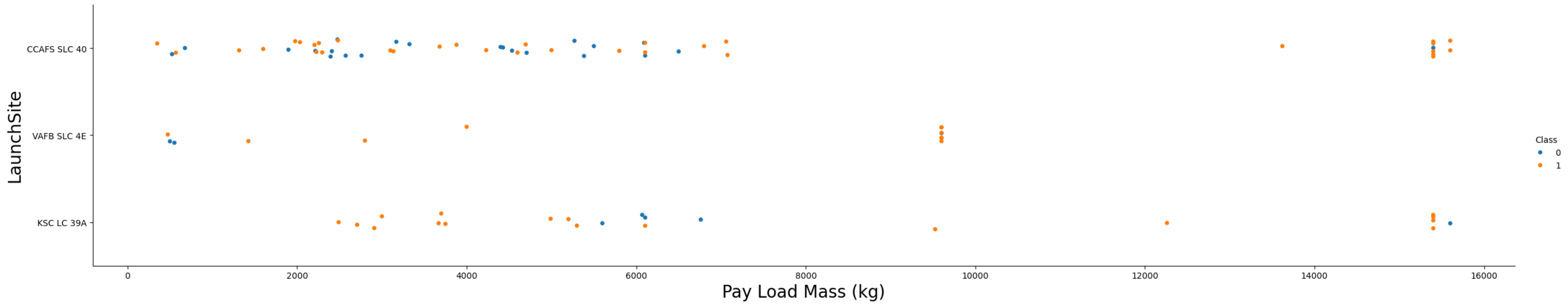
Insights drawn from EDA

Flight Number vs. Launch Site



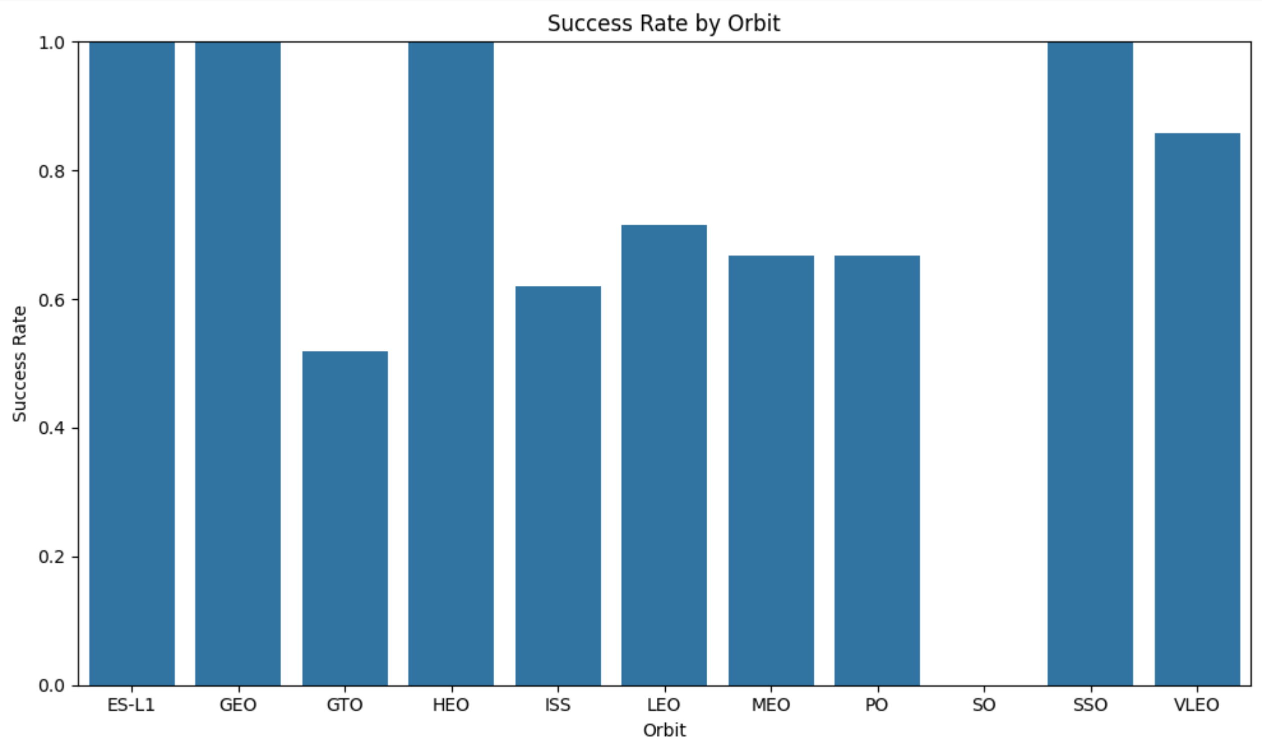
- The earliest flights all failed while the latest flights all succeeded.
- The CCAFS SLC 40 launch site has about a half of all launches.
- VAFB SLC 4E and KSC LC 39A have higher success rates.

Payload vs. Launch Site



- For every launch site the higher the payload mass means 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.

Success Rate vs. Orbit Type

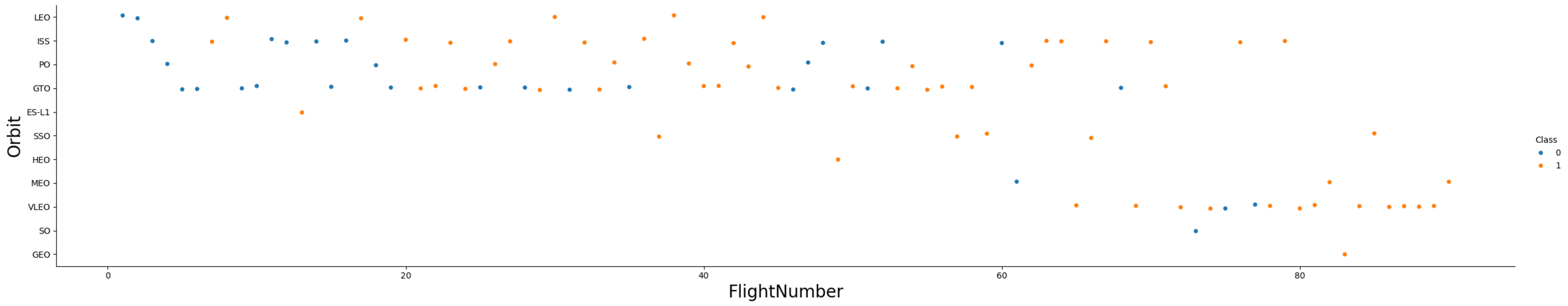


Orbits with 100% success rate
ES-L1, GEO, HEO, SSO

Orbits with success rate between 50% and 85%
GTO, ISS, LEO, MEO, PO

Orbits with 0% success rate
SO

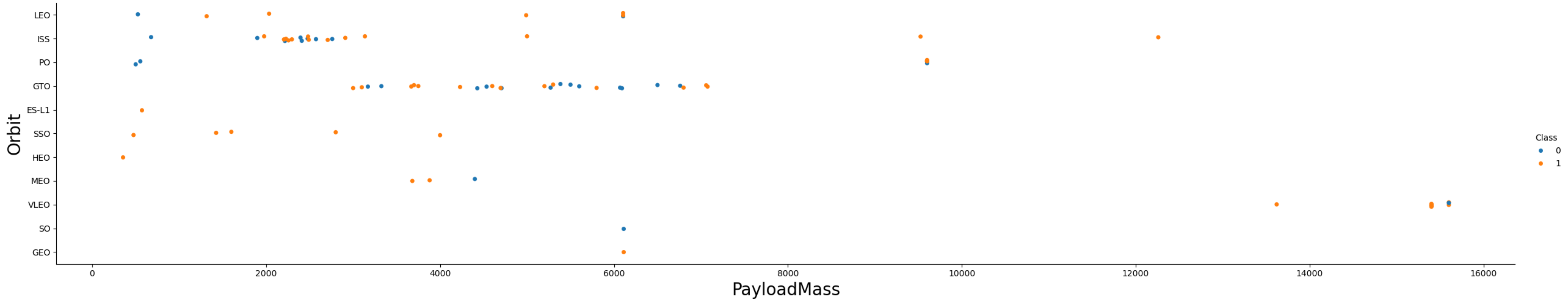
Flight Number vs. Orbit Type



In the LEO orbit the Success rate appears related to the number of flights

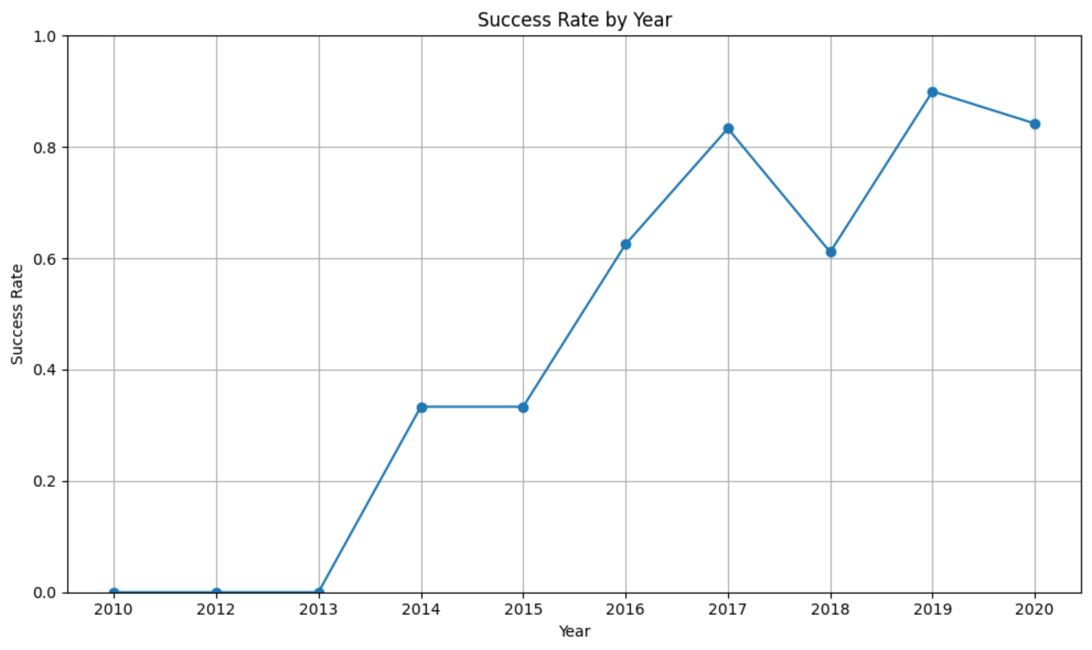
There seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend



The success rate has been increasing since 2013.

All Launch Site Names

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

In [13]:
```
%sql select distinct Launch_Site         from SPACEXTABLE
```

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

Out[13]:

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

Launch Site Names Begin with 'CCA'

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

In [14]:
```
%sql select *    from SPACEXTABLE a where  a.Launch_Site like "CCA%"
```

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

Out[14]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcor
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachu
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 (parachu
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attem
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attem
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attem

Total Payload Mass

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

In [16]:
```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE  where Customer like "%NASA%"
```

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

Out[16]:

sum(PAYLOAD_MASS__KG_)
107010

Average Payload Mass by F9 v1.1

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

In [25]:
```
%sql select Booster_Version, avg(PAYLOAD_MASS__KG_) from SPACEXTABLE  where Booster_Version = "F9 v1.1" group by Booster_Ver
```

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

Out[25]:

Booster_Version	avg(PAYLOAD_MASS__KG_)
F9 v1.1	2928.4

First Successful Ground Landing Date

In [28]:
```
%sql select min(Date) from SPACEXTABLE  where Landing_Outcome= "Success (ground pad)"
```

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

Out[28]:

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

In [31]:
```
ect Booster_Version from SPACEXTABLE  where Landing_Outcome ="Success (drone ship)" and (PAYLOAD_MASS__KG_ between 4000 and 6
```

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

Out[31]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

In [34]:
```
%sql select Mission_Outcome, count(Mission_Outcome) from SPACEXTABLE group by Mission_Outcome
```

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

Out[34]:

Mission_Outcome	count(Mission_Outcome)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

```
In [37]: %sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)

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

Out[37]:

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

In [52]:

```
%sql select * FROM SPACEXTABLE WHERE substr(Date, 0, 5) = '2015' and Landing_Outcome="Failure (drone ship)"
```

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

Out[52]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

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

In [57]:
```
%sql select Landing_Outcome, count(Landing_Outcome) FROM SPACEXTABLE where Date BETWEEN '2010-06-04' AND '2017-03-20'  grou
```

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

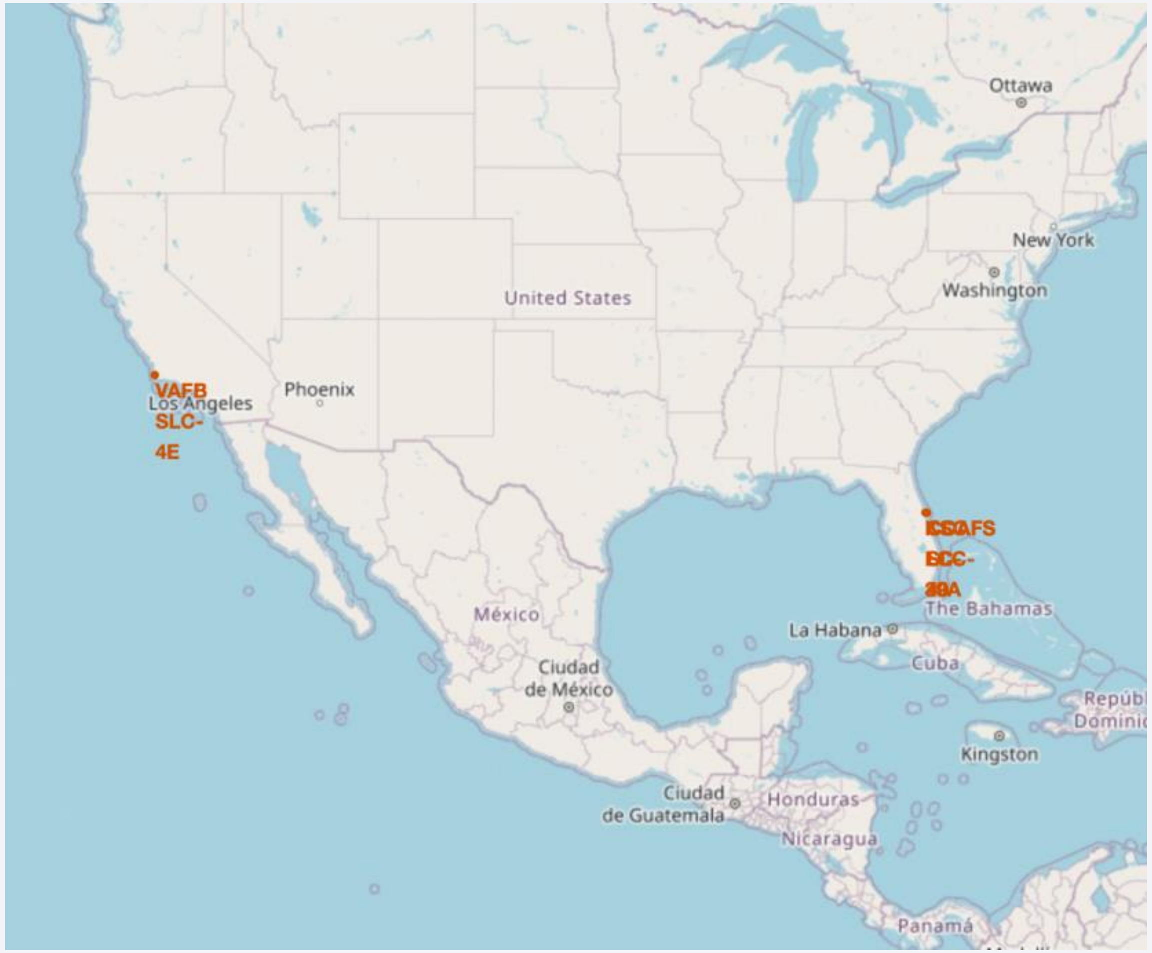
Out[57]:

Landing_Outcome	count(Landing_Outcome)
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

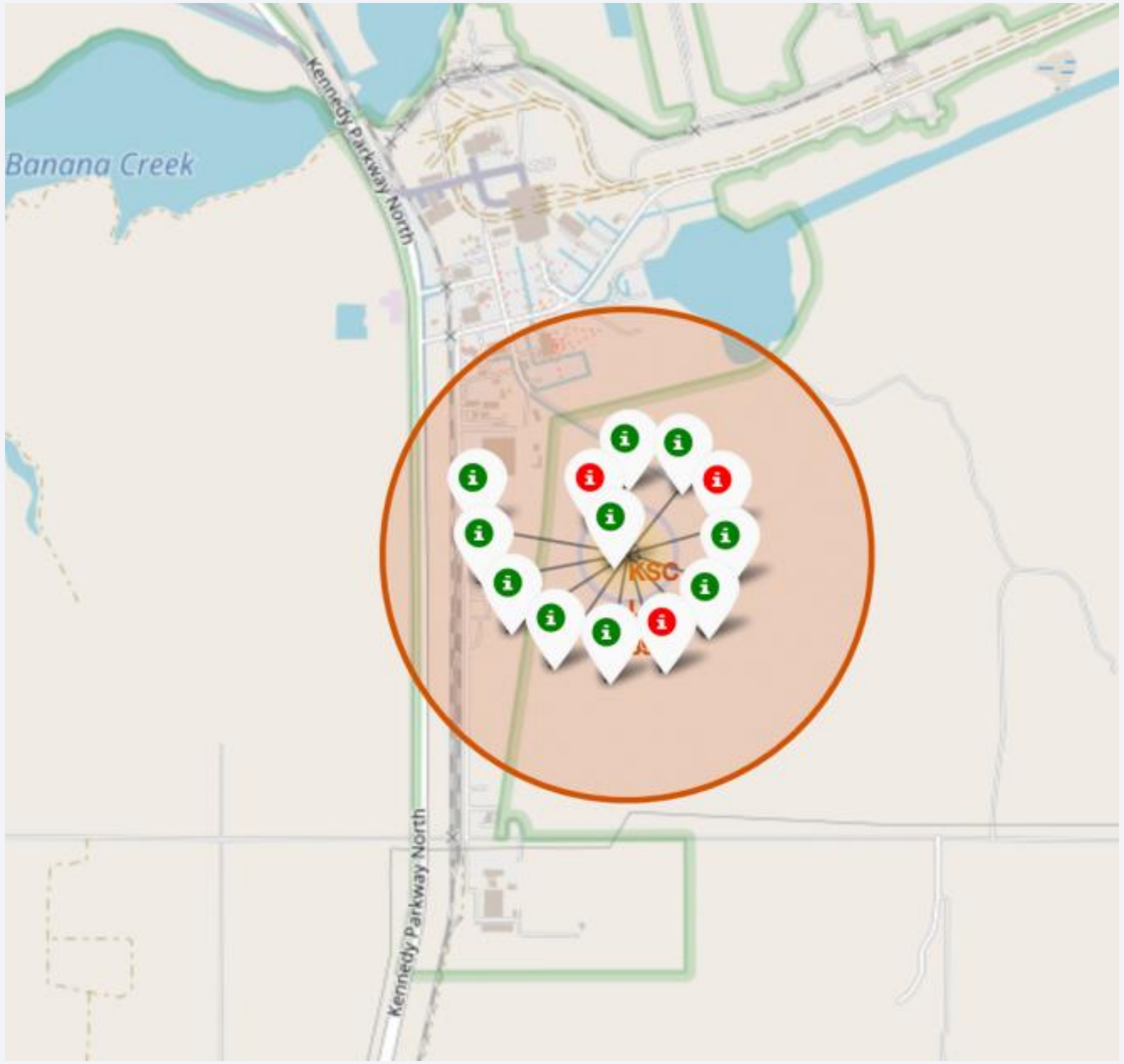
Section 3

Launch Sites Proximities Analysis

All launch sites' location markers



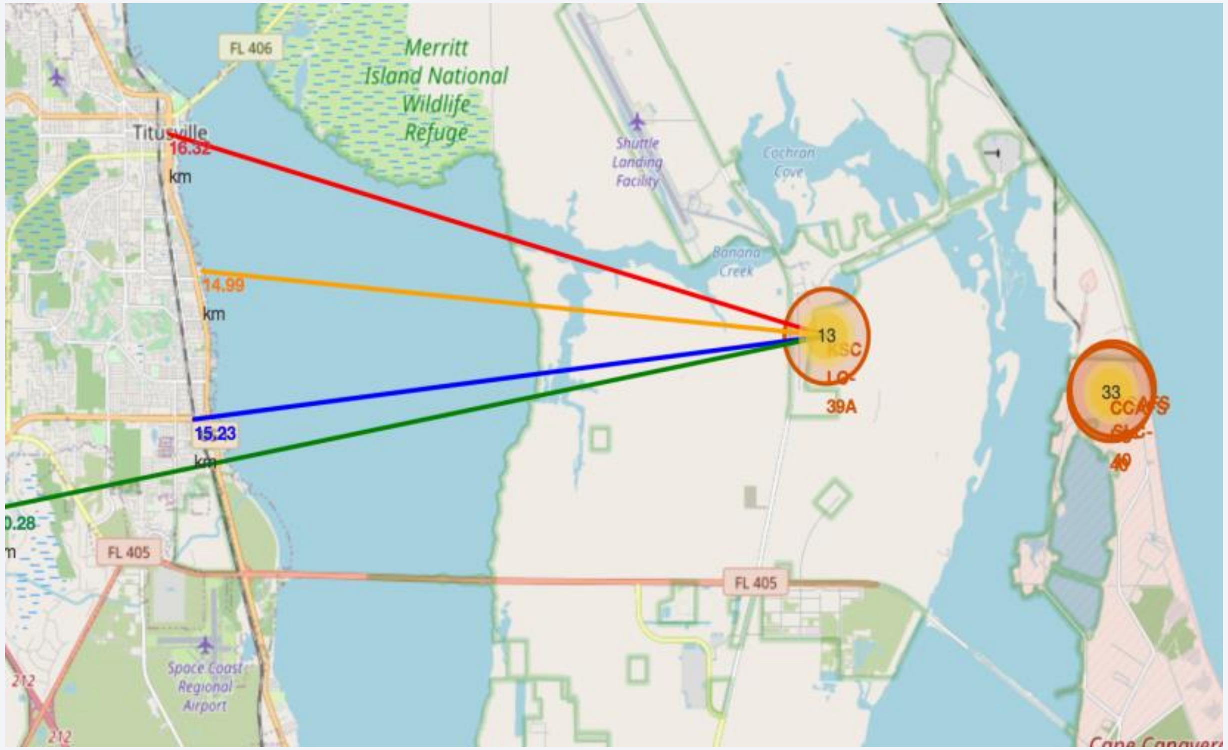
Color-labeled launch record



Green Marker = Successful Launch

Red Marker = Failed Launch

Distance from KSC LC-39A to its proximities



Relative close to railway (15.23 km)

Relative close to highway (20.28 km)

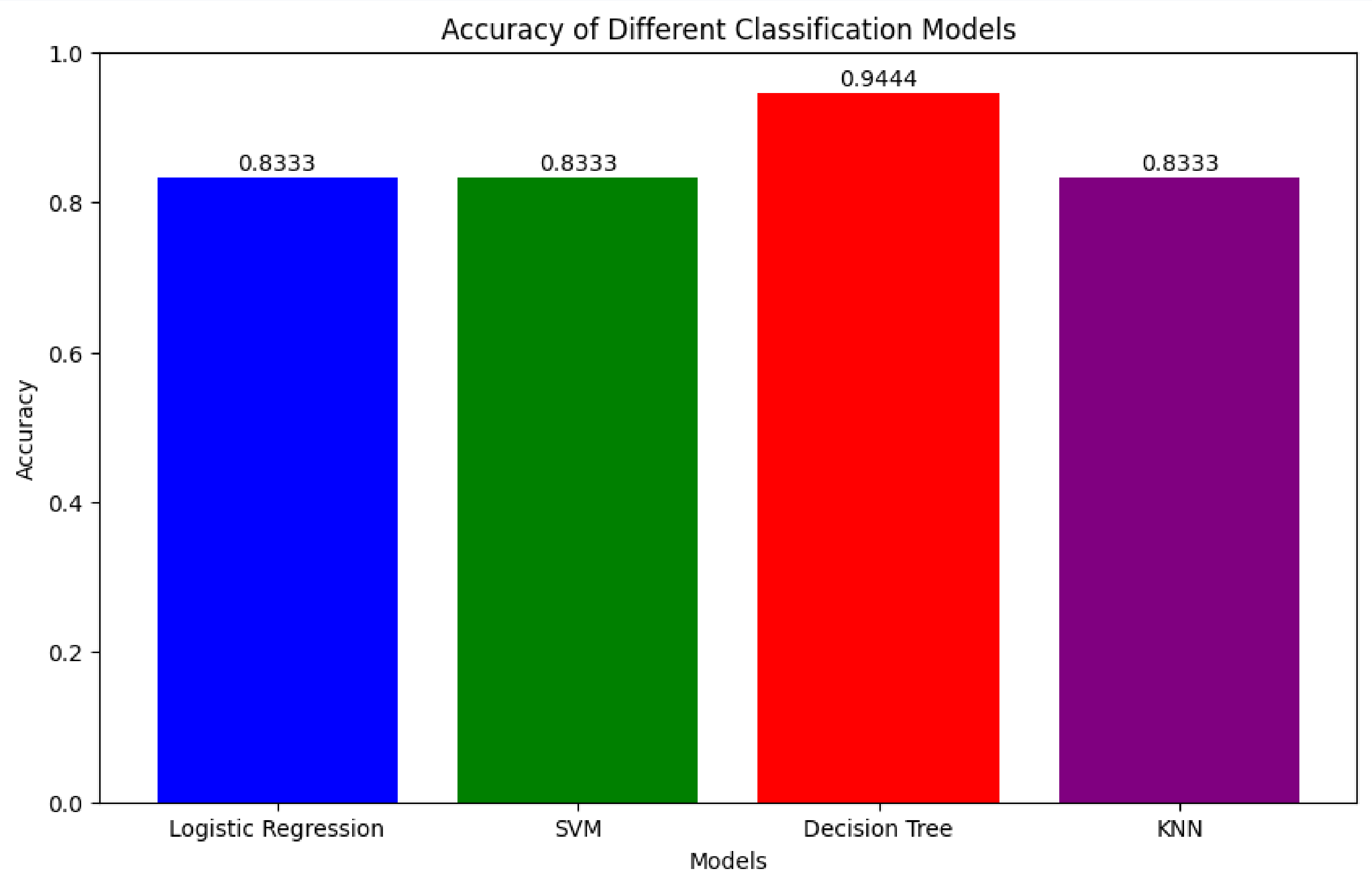
Relative close to coastline (14.99 km)

Relative close to the closest city Titusville (16.32 km)

Section 5

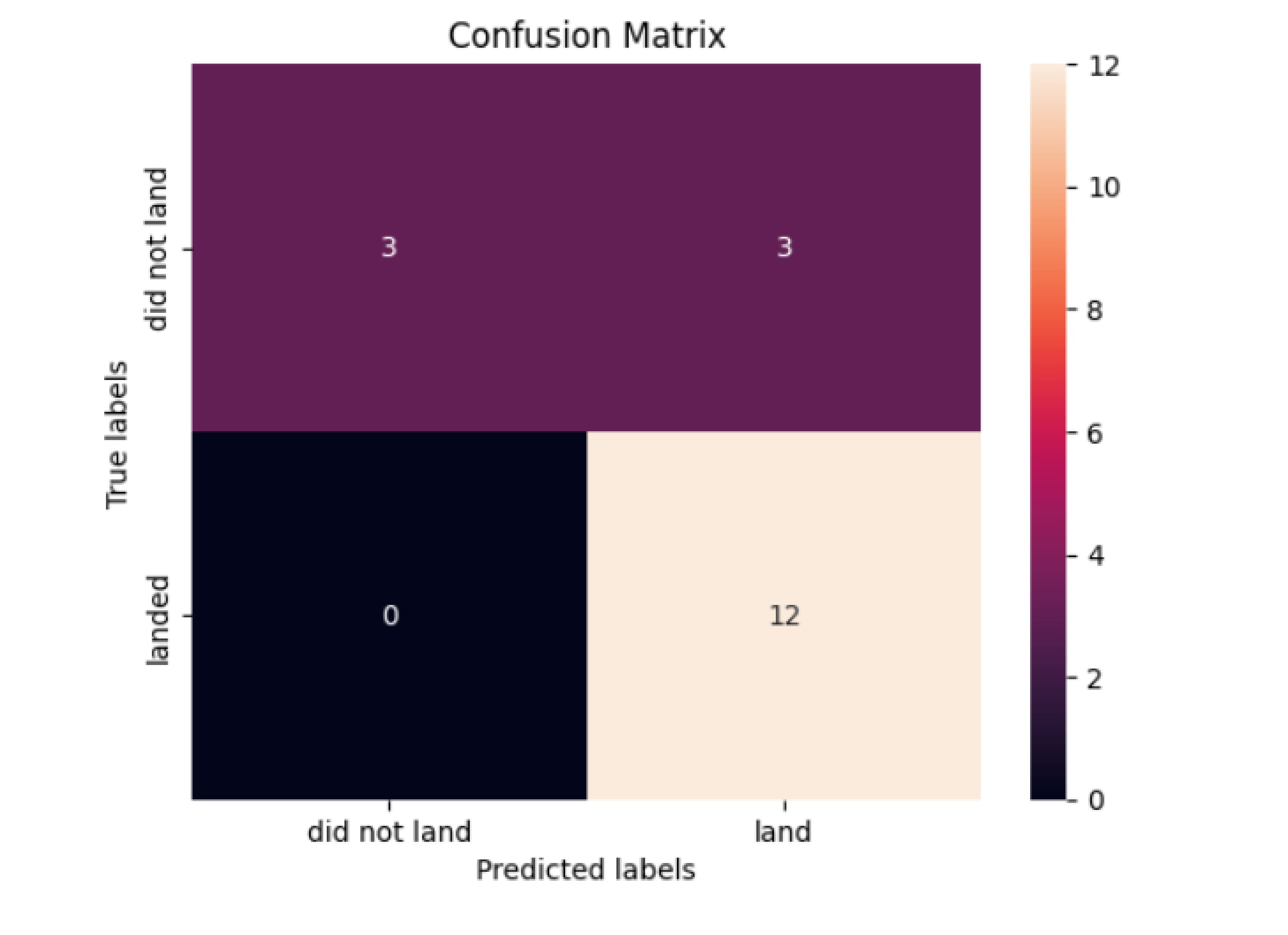
Predictive Analysis (Classification)

Classification Accuracy



The scores of the whole Dataset confirm that the best model is the Decision Tree Model.

Confusion Matrix



Conclusions

- Decision Tree Model is the best algorithm for this dataset.
- Launches with a low payload mass show better results than launches with a larger payload mass.
- Most of launch sites are in proximity to the Equator line and all the sites are in very close proximity to the coast.
- The success rate of launches increases over the years.
- KSC LC-39A has the highest success rate of the launches from all the sites.

Thank you!