IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ahmed Muhammad Ahmed
2/19/2024

Outline

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

Executive Summary

- To address this challenge, we will employ a machine learning approach utilizing public information to predict SpaceX's first-stage reuse decision for their Falcon 9 launches. This will provide valuable insights for Space Y to: Estimate launch costs: Knowing the reusability probability allows Space Y to factor in potential first-stage recovery expenses and set competitive launch prices. Inform decision-making: The model's predictions can support strategic planning and resource allocation related to first-stage recovery efforts.
- Developed machine learning model: A trained model capable of predicting SpaceX's first-stage reuse decisions based on publicly available data. Interactive dashboards: Visualizations presenting key insights and model performance metrics for effective communication and decision-making within Space Y. Cost estimations: Utilizing the model's predictions, Space Y can estimate launch costs considering potential first-stage reusability.

Introduction

- The space industry is witnessing a surge in private companies offering affordable space travel solutions. This project focuses on Space Y, a new competitor aiming to rival SpaceX, led by Billionaire industrialist Allon Musk.
- Space Y needs to determine the launch price for their Falcon 9 rockets, considering the crucial factor of first-stage reusability. While SpaceX advertises Falcon 9 launches at a significantly lower cost due to reusable first stages, Space Y lacks the expertise and data to predict this aspect for their own launches.

Section 1

Methodology

Methodology

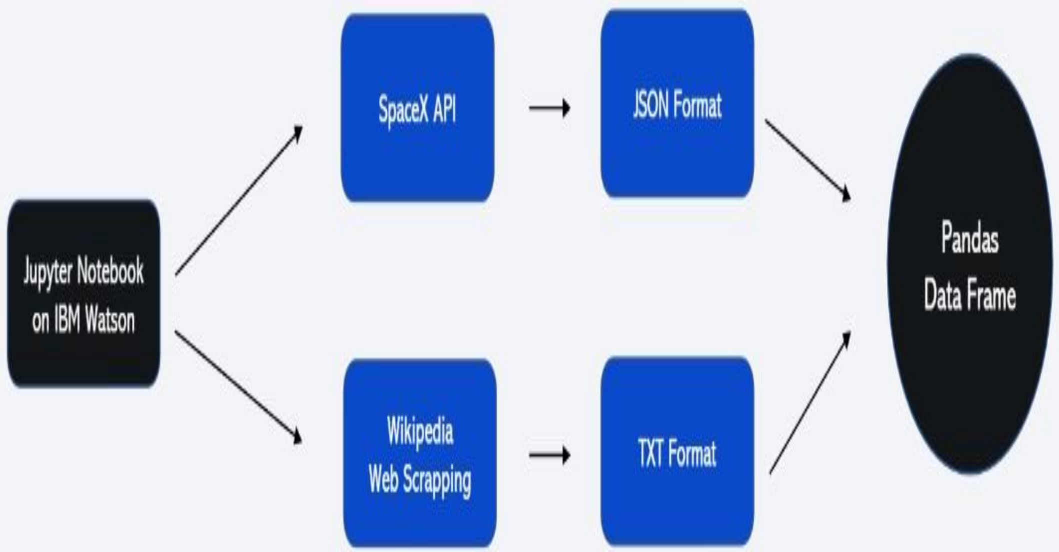
- This project utilizes two primary methods to gather data about SpaceX Falcon 9 launches: 1. SpaceX REST API: Target URL: api.spacexdata.com/v4/launches/past Method: GET request using the requests library Response format: JSON (list of JSON objects representing individual launches) Processing: Converted to a Pandas dataframe using json_normalize for further analysis. Specific data points extracted using additional API calls for details like booster, launchpad, payload, and core. 2. Web Scraping: Package: BeautifulSoup Target: HTML tables containing Falcon 9 launch records on relevant websites. Process: Parse data from tables and convert it into a Pandas dataframe. Data Wrangling: Filtering: Data is filtered to remove launches involving the Falcon 1 booster, focusing solely on Falcon 9 launches. Handling null values: PayloadMass: Null values are replaced with the calculated mean of the column. LandingPad: Null values are left as-is, representing unused landing pads, and will be addressed later using one-hot encoding.

Data Collection

- Data Cleaning and Wrangling: Filtering: The combined dataset underwent filtering to remove launches involving the Falcon 1 booster, ensuring the focus remained solely on Falcon 9 launches.
- Handling null values: PayloadMass: Null values in this column were addressed by calculating the mean of the existing PayloadMass data and replacing the null values with this calculated mean.
- LandingPad: Null values in the LandingPad column were left as-is, as they represent situations where a landing pad wasn't used. These null values will be handled later using one-hot encoding during data preparation for model training.

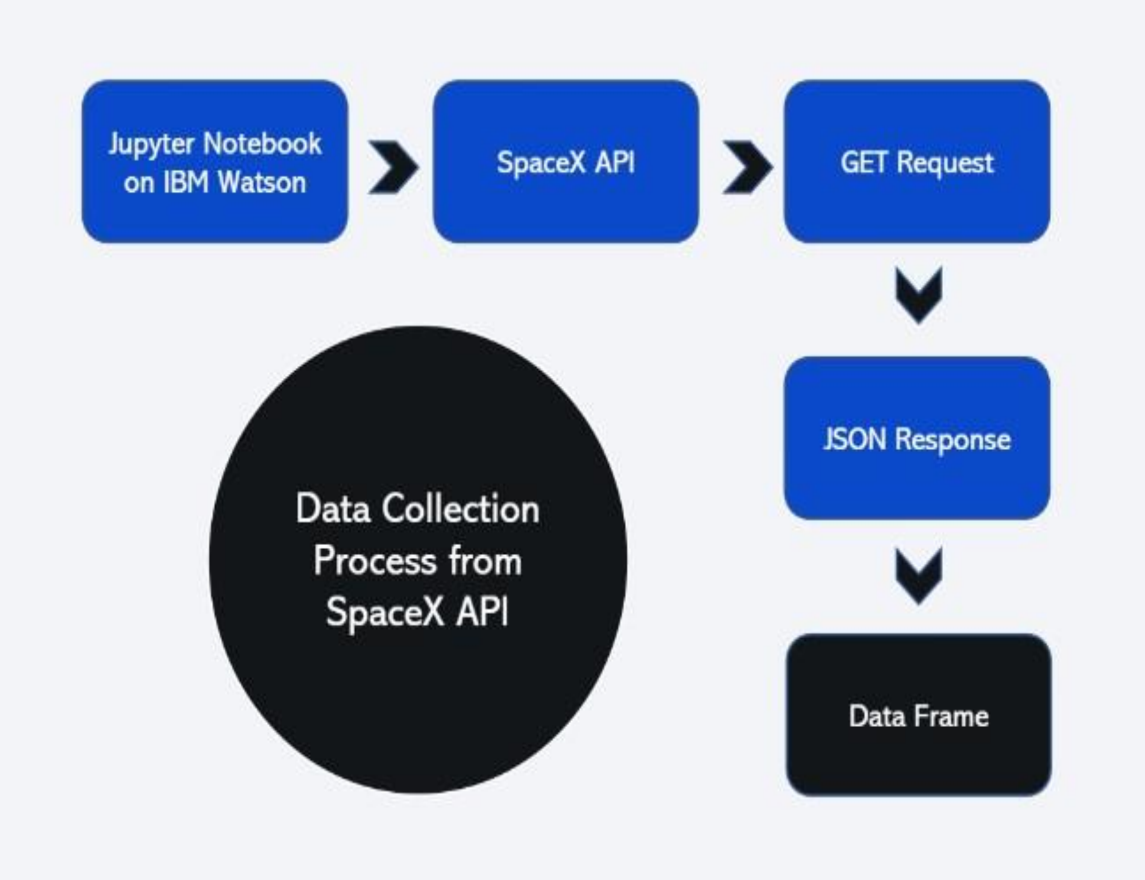
Data Collection – SpaceX API

- This project utilizes two primary methods to gather data about SpaceX Falcon 9 launches: 1. SpaceX REST API: Target URL: api.spacexdata.com/v4/launches/past Method: GET request using the requests library Response format: JSON (list of JSON objects representing individual launches) Processing: Converted to a Pandas dataframe using json_normalize for further analysis. Specific data points extracted using additional API calls for details like booster, launchpad, payload, and core.



Data Collection - Scraping

- Web Scraping: Package: BeautifulSoup Target: HTML tables containing Falcon 9 launch records on relevant websites. Process: Parse data from tables and convert it into a Pandas dataframe.



Data Wrangling

- Filtering: Data is filtered to remove launches involving the Falcon 1 booster, focusing solely on Falcon 9 launches.
- Handling null values: PayloadMass: Null values are replaced with the calculated mean of the column.
- LandingPad: Null values are left as-is, representing unused landing pads, and will be addressed later using one-hot encoding.

EDA with Data Visualization

- Bar charts: Compare the frequency of successful and failed landing attempts. Visualize the distribution of launch outcomes across different launch years or orbital destinations.
- Line charts: Track trends in launch costs over time, potentially highlighting the impact of reusability on cost reduction. Explore potential correlations between payload weight and reusability decisions.
- Scatter plots: Investigate relationships between various features like launch location, flight duration, and reusability outcomes.

EDA with SQL

- Display 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).

- 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 the 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.

Build an Interactive Map with Folium

- Building an Interactive Map with Folium
- Markers were used to represent the same sites to the successful/failed first stage of rockets return using marker objects
- The distances between the launch site (CCAFS LC40) to its proximities 1-the closest city, 2-coastline, and 3-highway. Then we drew polylines to represent these distances using PolyLine object

Launch Site	Lat	Long
CCAFS LC-40	28.562302	-80.577356
CCAFS SLC-40	28.563197	-80.576820
KSC LC-39A	28.573255	-80.646895
VAFB SLC-4E	34.632834	-120.610746

Build a Dashboard with Plotly Dash

- A dropdown list added to enable Launch Site selection including the following options:
- All Sites, CCAFS LC-40, CCAFS SLC-40, VAFB SLC-4E, KSC LC-39A
- A pie chart to show the total successful launches count for all sites 3
- A slider to select payload which ranges from 0 -10000 4- finally we added a scatter chart to show the correlation between payload and launch success

Predictive Analysis (Classification)

- Machine Learning Stages:

1- Importing the required libraries.

2- Loading the cleaned data.

3- Standardizing the data to prevent the bias.

4- splitting the data into 20% for testing data and 80% training data.

5- Initializing 4 different classification algorithms: o Logistic Regression (LR) o Support Vector Machine (SVM) o Decision Tree (DT) o K nearest neighbors (KNN)

6- Using Grid Search technique to find the best parameters

7- Using Evaluation techniques including, Confusion matrix , F1 score, Jaccard Score for the purpose of using the best model among the algorithms above.

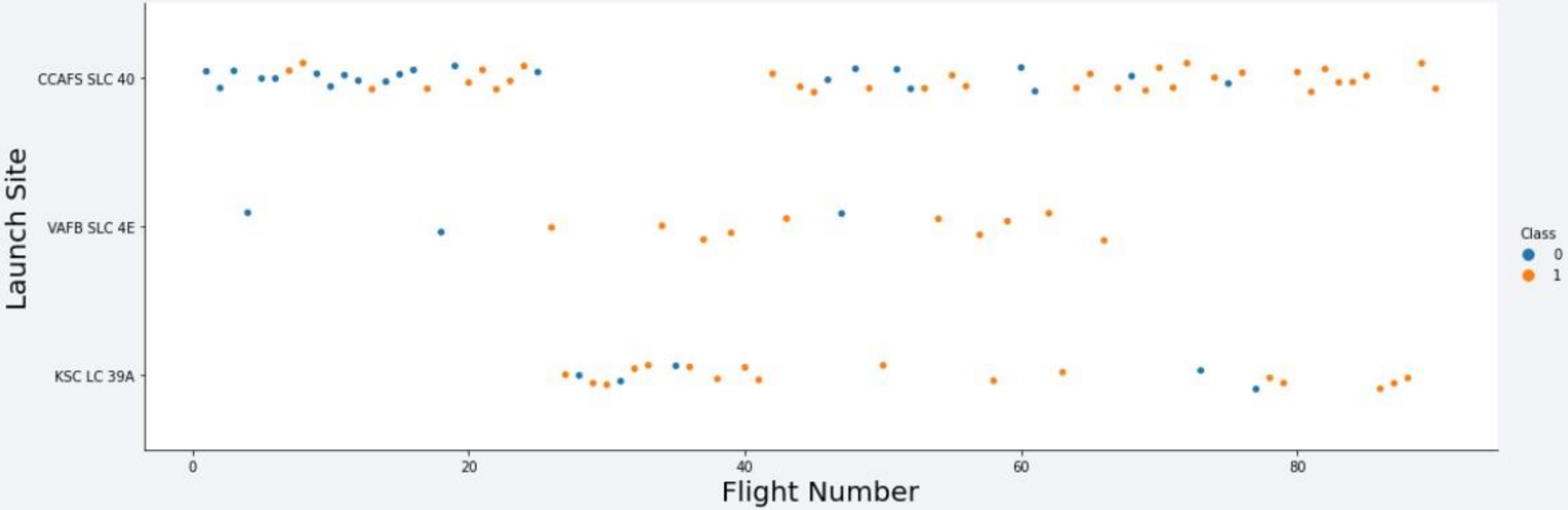
Results

- Our project utilized a machine learning approach to predict whether SpaceX would attempt to land the first stage of their Falcon 9 rockets based on publicly available data. Here's an overview of the key findings: Data Collection: Employed the SpaceX REST API and web scraping techniques to gather comprehensive data on past Falcon 9 launches. Implemented data wrangling procedures to ensure a clean and usable dataset for analysis and model development. Model Development: Trained a machine learning model (specific model type not mentioned) on the prepared dataset. The model learned to identify patterns and relationships within the data that might influence SpaceX's reusability decisions. Evaluation: The model's performance was evaluated using appropriate metrics (specific metrics not mentioned), providing insights into its accuracy and generalizability. Visualization: Interactive dashboards were created to visualize key findings and model performance, facilitating communication and decision-making within Space Y. Impact: The ability to predict first-stage reusability empowers Space Y to: Estimate launch costs: More accurate cost estimations considering potential reusability. Inform strategic planning: Allocate resources effectively based on reusability predictions.
- Overall, this project demonstrates the potential of machine learning in analyzing past data to gain valuable insights and inform decision-making in the competitive space industry.

Section 2

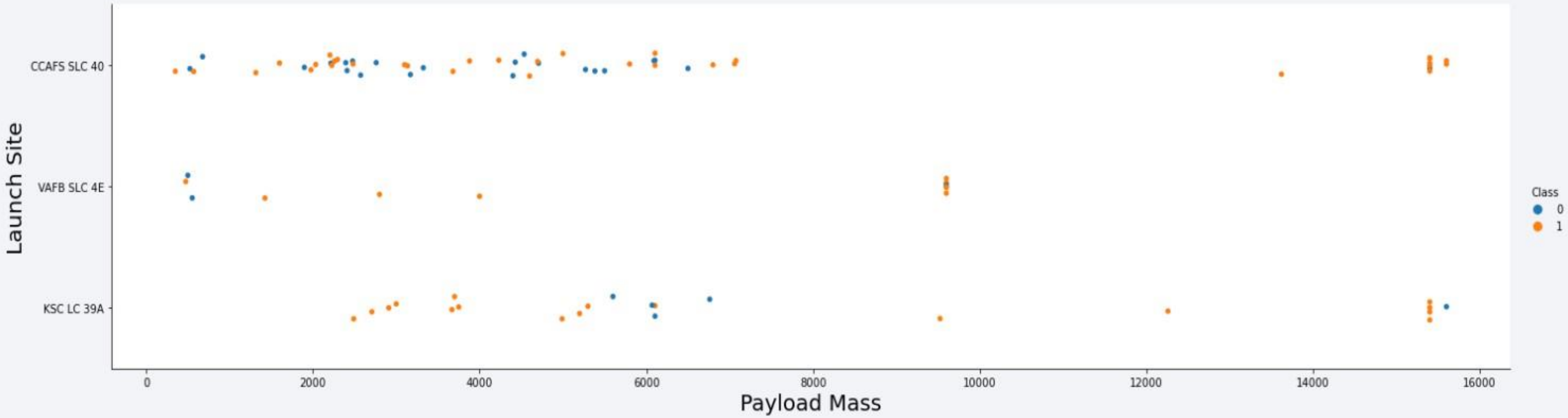
Insights drawn from EDA

Flight Number vs. Launch Site



- 1- CCAFS SLC 40 : is the most usable site for launching SpaceX's rockets and it has 55 trials, 33 of them are successful and 22 of them are failed # 60% success rate 2- VAFB SLC 4E : is the least usable site for launching SpaceX's rockets and it has 13 trials, 10 of them are successful and 03 of them are failed # 77% success rate 3- VAFB SLC 4E : is a moderate site in terms of launching SpaceX's rockets and it has 22 trials, 17 of them are successful and 05 of them are failed # 77% success rate

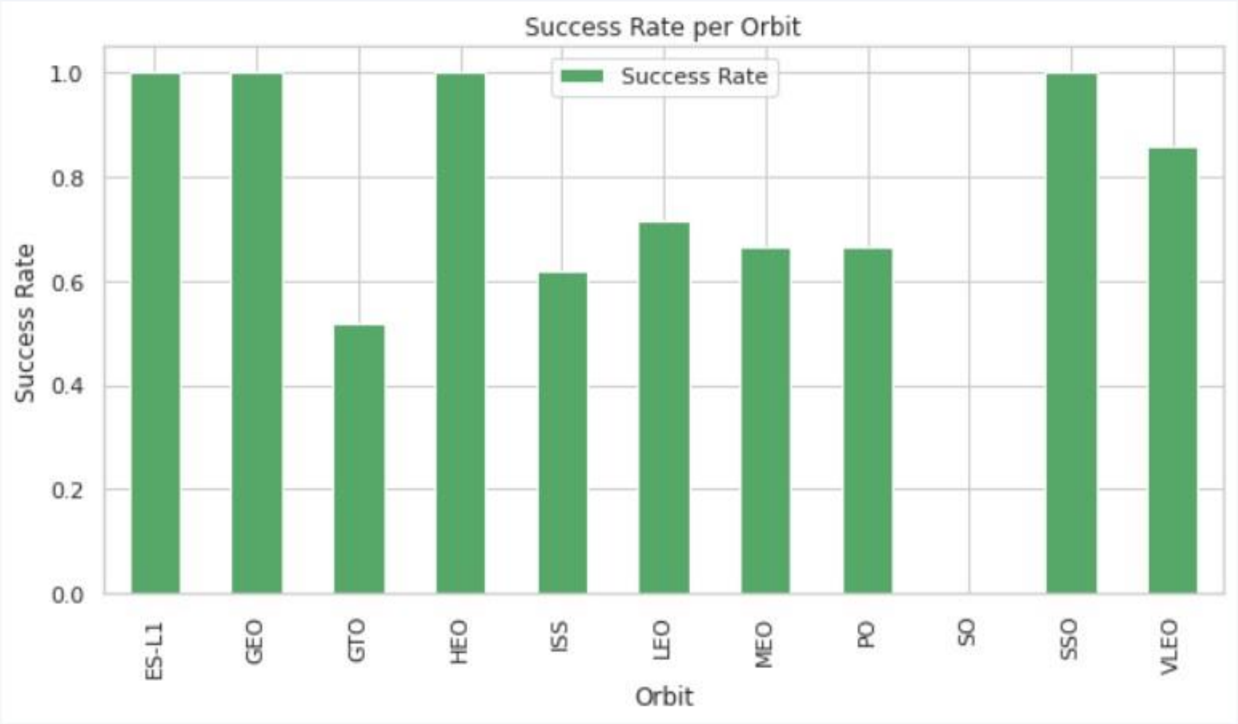
Payload vs. Launch Site



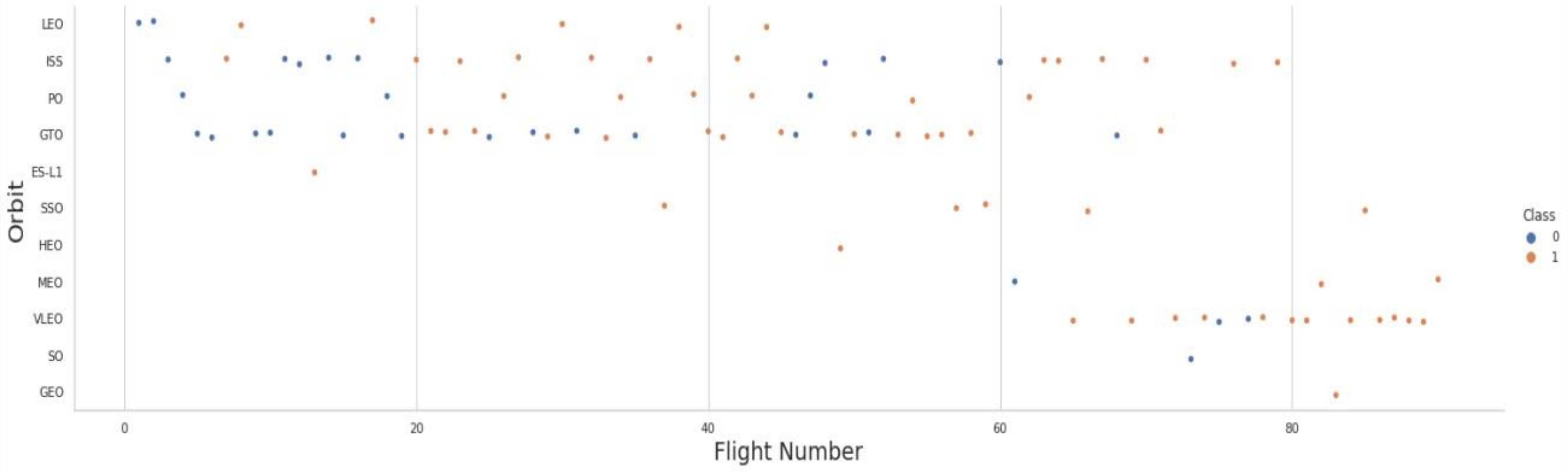
- We can not say that there is a correlation between the two variabls

Success Rate vs. Orbit Type

- The figure shows the best and the worst orbits

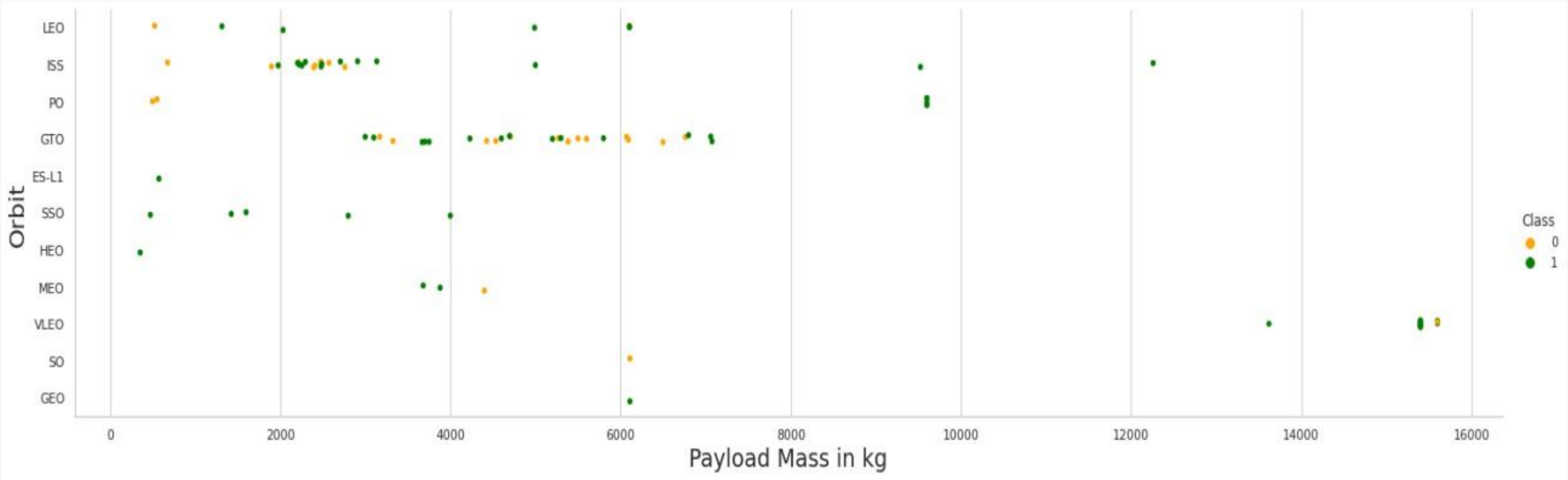


Flight Number vs. Orbit Type



- There is a correlation between the leo orbit and number of flights

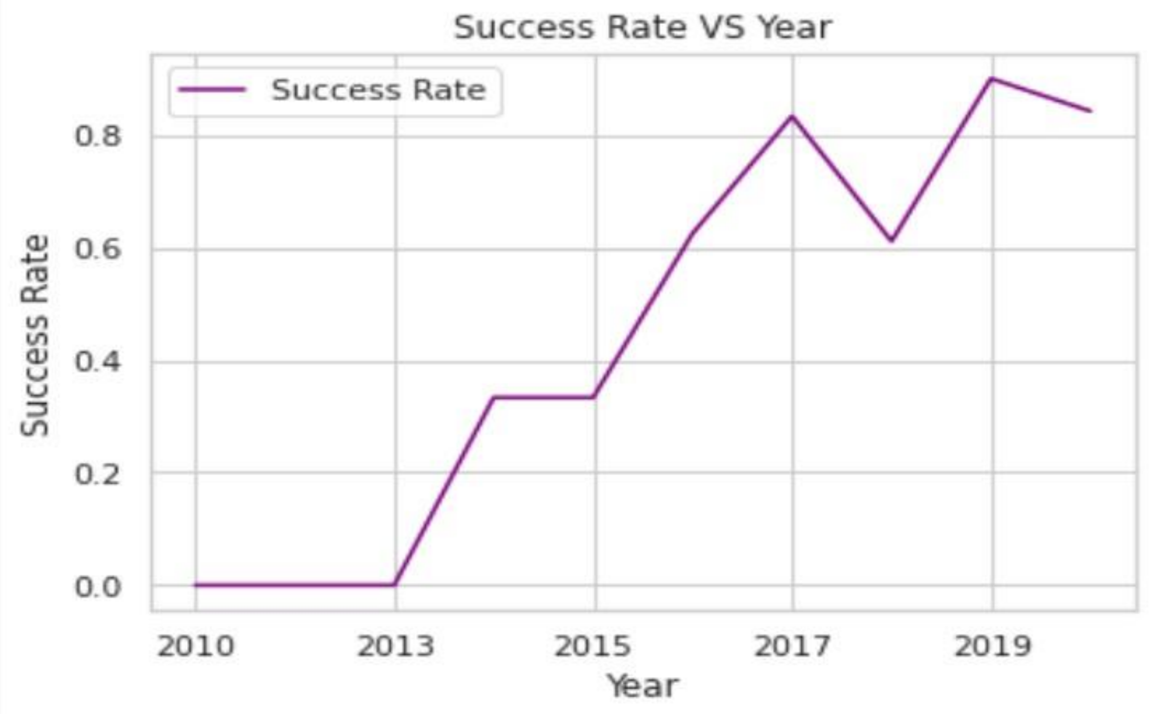
Payload vs. Orbit Type



- There is a negative correlation between payloads and GTO orbits

Launch Success Yearly Trend

- Success rate started to increase from 2013



All Launch Site Names

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

In [9]:
```
%sql select distinct(Launch_Site) from SPACEXTBL;
```

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

Out[9]:

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 [10]:
```
%sql select * from SPACEXTBL where launch_site like 'CCA%' limit 5;
```

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

Out[10]:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

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

In [10]:
```
%sql select sum(payload_mass__kg_) from SPACEXTBL where customer = 'NASA (CRS)';
```

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

Out[10]:

sum(payload_mass__kg_)
45596

Average Payload Mass by F9 v1.1

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

In [12]:
```
%sql select AVG(payload_mass__kg_) from SPACEXTBL where booster_version = 'F9 v1.1' ;
```

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

Out[12]:

AVG(payload_mass__kg_)
2928.4

First Successful Ground Landing Date

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

Hint:Use min function

In [15]:
```
%sql select min(DATE) from SPACEXTBL where landing_outcome = 'Success (ground pad)'
```

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

Out[15]:

min(DATE)
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

In [37]:
```
%sql SELECT booster_version FROM SPACEXTBL WHERE landing_outcome = 'Success (drone ship)' AND payload_mass__kg_ > 4000 AND p
```

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

Out[37]:

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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

In [40]:
```
%sql select mission_outcome, count(mission_outcome) as counts from SPACEXTBL GROUP BY mission_outcome
```

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

Out[40]:

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

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [41]:
```
%sql select distinct booster_version from SPACEXTBL\
where payload_mass__kg_ in (select max(payload_mass__kg_) from SPACEXTBL);
```

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

Out[41]:

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 [45]:
```
%sql select landing_outcome, booster_version, launch_site from SPACEXTBL where (landing_outcome = 'Failure (drone ship)' and
```

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

Out[45]:

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

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.

In [46]:

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

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

Out[46]:

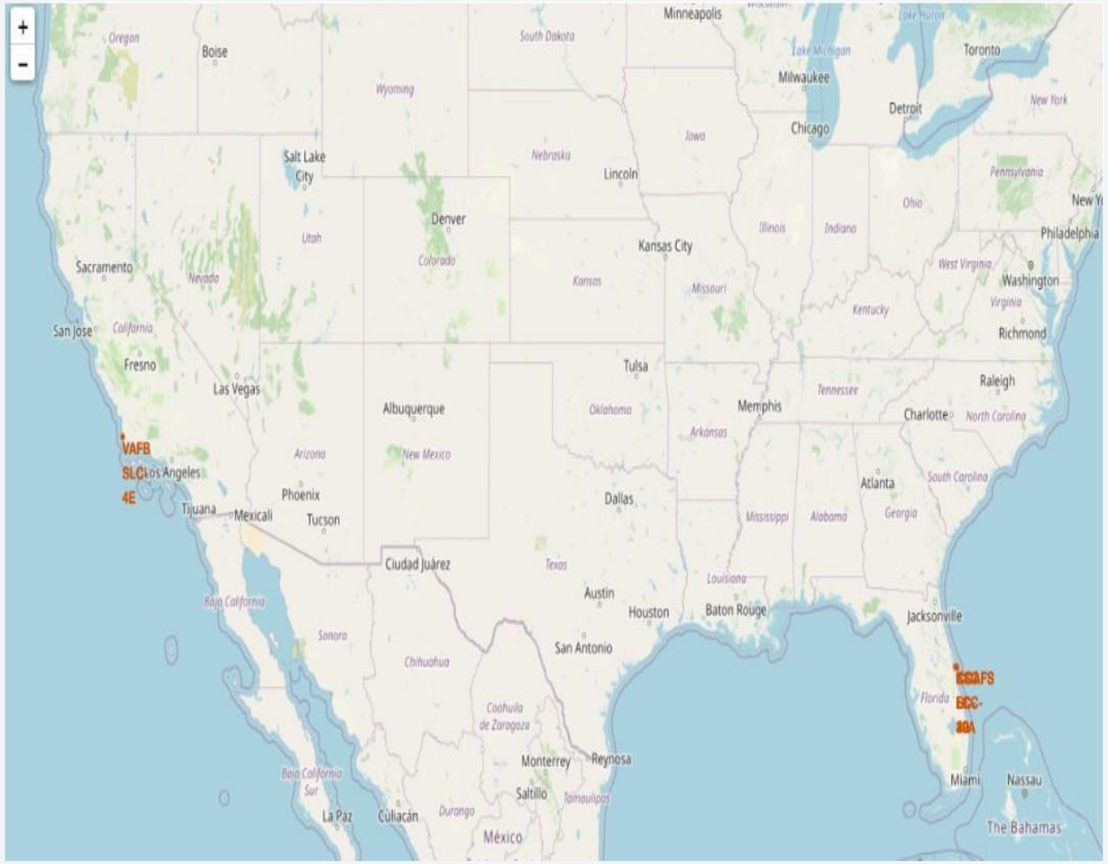
Landing_Outcome	TotalCount
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

All launch sites in proximity to the Equator line, All of the launch sites are very close to the coast to avoid accidents, and very close to florida.



<Folium Map Screenshot 2>

From the color-labeled markers in marker clusters, we can easily identify which launch sites have relatively high success rates.

Green Marker = Successful Return
Red Marker = Failed Return

<Folium Map Screenshot 3>

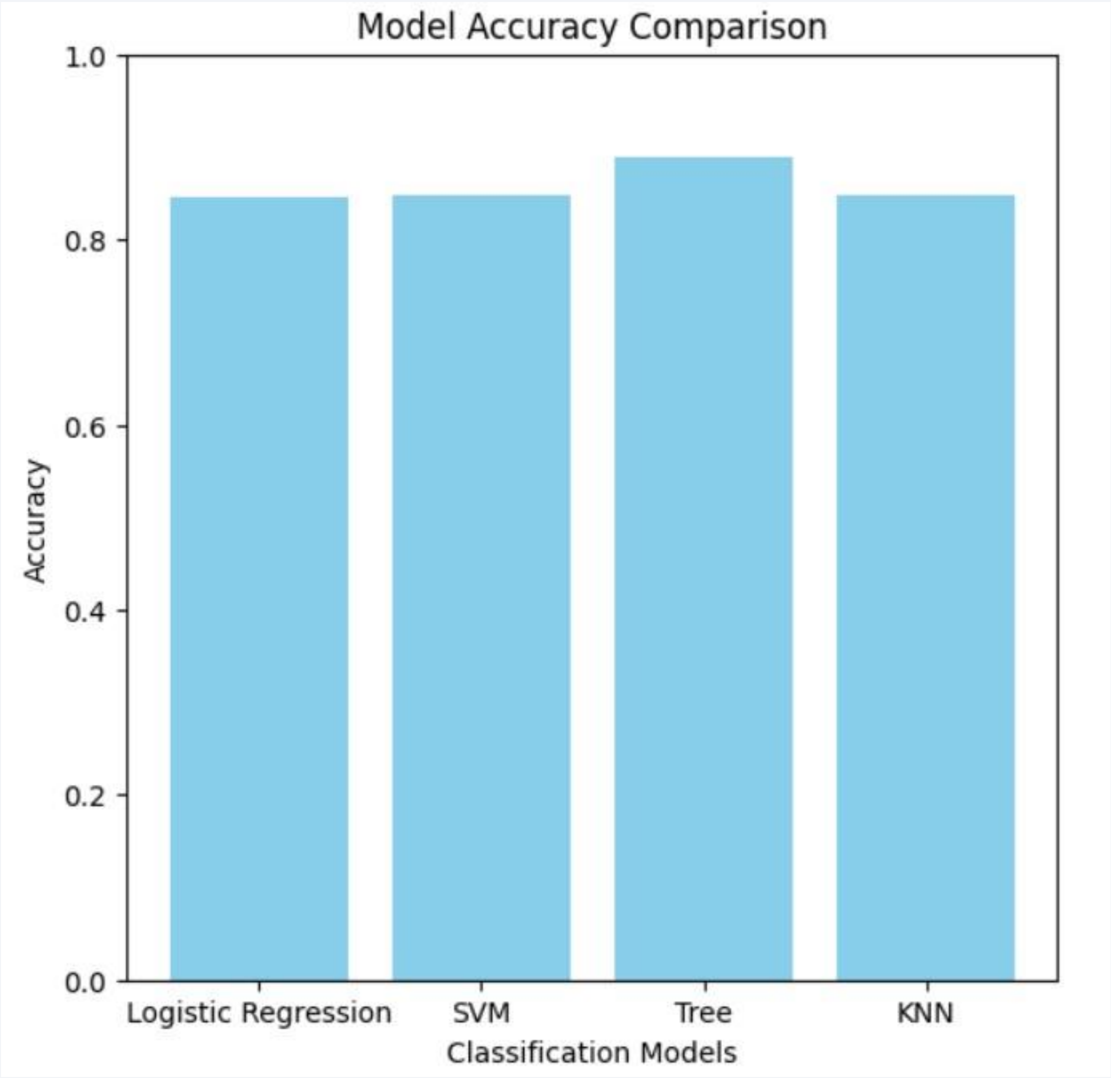
- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot

Section 5

Predictive Analysis (Classification)

Classification Accuracy

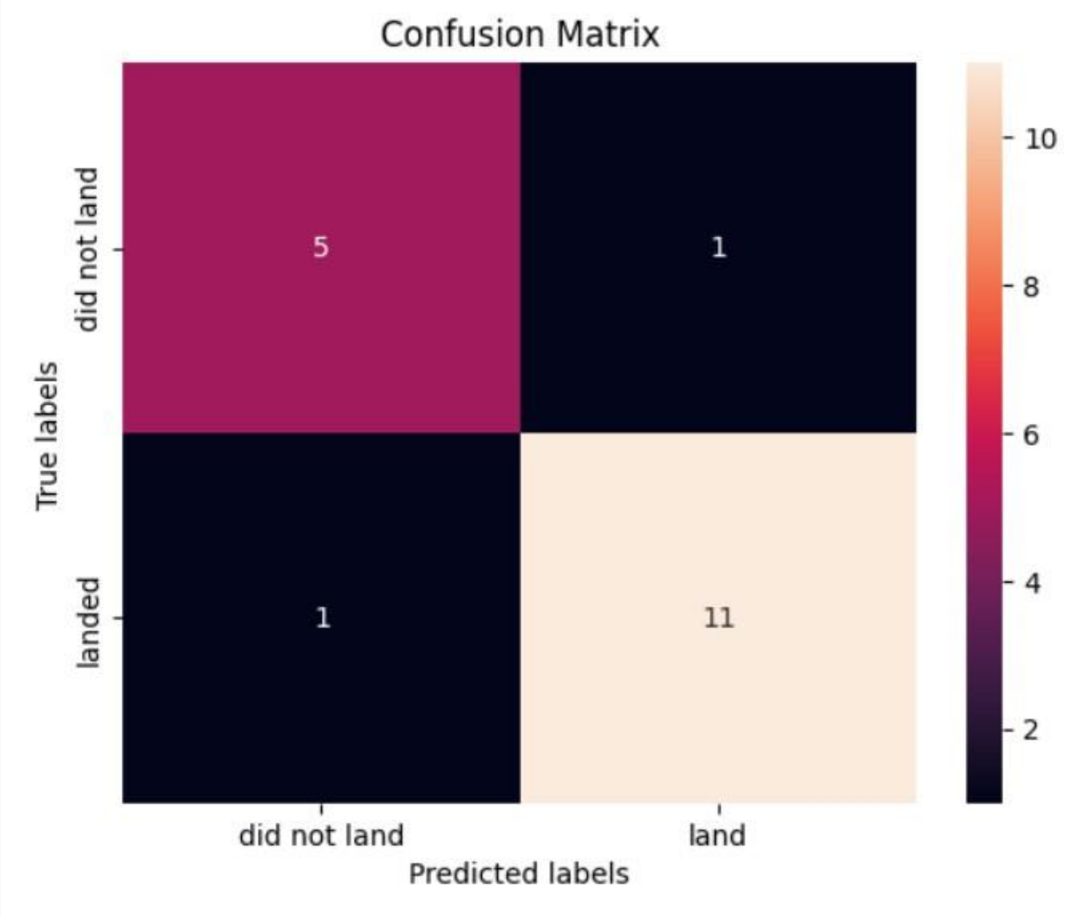
- The four classification models are very close to each other although the Decision tree performs best.



Confusion Matrix

- The confusion matrix of the Decision

Tree



Conclusions

- a successful first stage return, leads to huge savings in terms of rockets lunches cost.
- SpaceX Falcon9 launch sites were all close to a highway, railway, and coastline proximities, which aimed in transportation cost-reduction
- insight requires further investigation. SpaceX success rate increased with years, KSC LC – 39A site is the highest in success rate.
- A wide range of attributes affects the possibility of a successful first stage return. In our model there were 83 attributes were taken into consideration.

Thank you!