Analysis of SpaceX Falcon9 first stage successful landing

Alpesh Gorde

27-08-2021

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Using the historical data from all Falcon-9 launches, the future success of the Falcon9 1st stage landing predicted.

- Factors considered include launching sites, payload mass, orbits, landing pads, and features of the rockets such as whether it has grid-fins, landing legs, reuse history, and its version numbers.

- By using Python, Machine Learning algorithms and data related to Space X rocket launches, analyzed the successful landings of the stage 1 boosters in order to predict future successful landings.

INTRODUCTION

- Reusable 1st stage represent a potential cost reduction with an average cost of 65 million per launch as compared to 162 million dollar per launch for others.
- SpaceX Falcon-9, is the first rocket that has ever achieved the relanding success of its first stage rocket.
- The relanding and reusability of the first stage enables huge cost saving, marking space exploration affordable.

METHODOLOGY

1. Data collection:

Using BeautifulSoup, data regarding SpaceX Falcon 9 launch records from 2010 to 2020, were obtained from Wikipedia by parsing the HTML flight table.

2. Perform data wrangling:

Data from the HTML table was loaded into a pandas data frame and each Column frequencies were analyzed. Missing numbers were imputed as means of their respective columns.

3. Perform exploratory data analysis (EDA) using visualization and SQL:

A combination of SQL queries and plot visualizations via Matplotlib, Plotly, Folium, and DASH integrations were used to visualize different aspects of the data.

METHODOLOGY

4. Perform interactive visual analytics using Folium and Plotly Dash:

Utilizing scikit.learn, the data was split into trining and test data after variables were dummy coded. Various models were to predict success.

5. Perform predictive analysis using classification models

- Create a column for the class
- Standardize the data
- Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data

1. Data collection

- Data from the SpaceX API was requested and obtained via a JSON file.
- The data was appended to a table and filtered for Falcon 9 launches.
- Utilizing BeautifulSoup, obtained the publicly available data regarding SpaceXFalcon 9 launches via Wikipedia.

Flowchart of SpaceX API

Web Scrapping

- Request the Falcon9 Launch Wiki page from its URL
  - static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=102 7686922”
  - response = requests.get(static_url).text
  - soup = BeautifulSoup(response, 'html.parser')
- Extract all column/variable names from the HTMLtable header
  - html_tables = soup.find_all('table’)
  - first_launch_table = html_tables[2]

  - Create a data frame by parsing the launch HTML tables

Data wrangling

- In this part, perform some explanatory data analysis and create the outcome label
- The steps are as follows:
  - Calculate the number of launches on each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurrence of mission outcome per orbit type
  - Create a landing outcome label from Outcome column

EDA with data visualization

- To understand the relationship between the relanding success rate and various factors, created a number of charts, overlay the outcome of the relanding.

- PayloadMass vs. FlightNumber: as the flight number increases, the firststage is more likely to reland successfully.

- PayloadMass vs. Booster: VLEO has relatively high success rate to reland successfully.

EDA with SQL

- SQL allows us to explore the data directly
- Observed that there are four launch sites
- Obtain the situation with different booster versions

Build an interactive map with Folium

- Identifies all the launch sites on the map

- Marked the launch sites with different colors based on success and failure of the landing outcomes

- Demonstrated distance to important surrounding logistic features like railroads, coasts and highways.

Build a Dashboard with Plotly Dash

- Pie graph was used to demonstrate launch site success/failures i.e. to see if one.

- Site is more associated with successful landing.

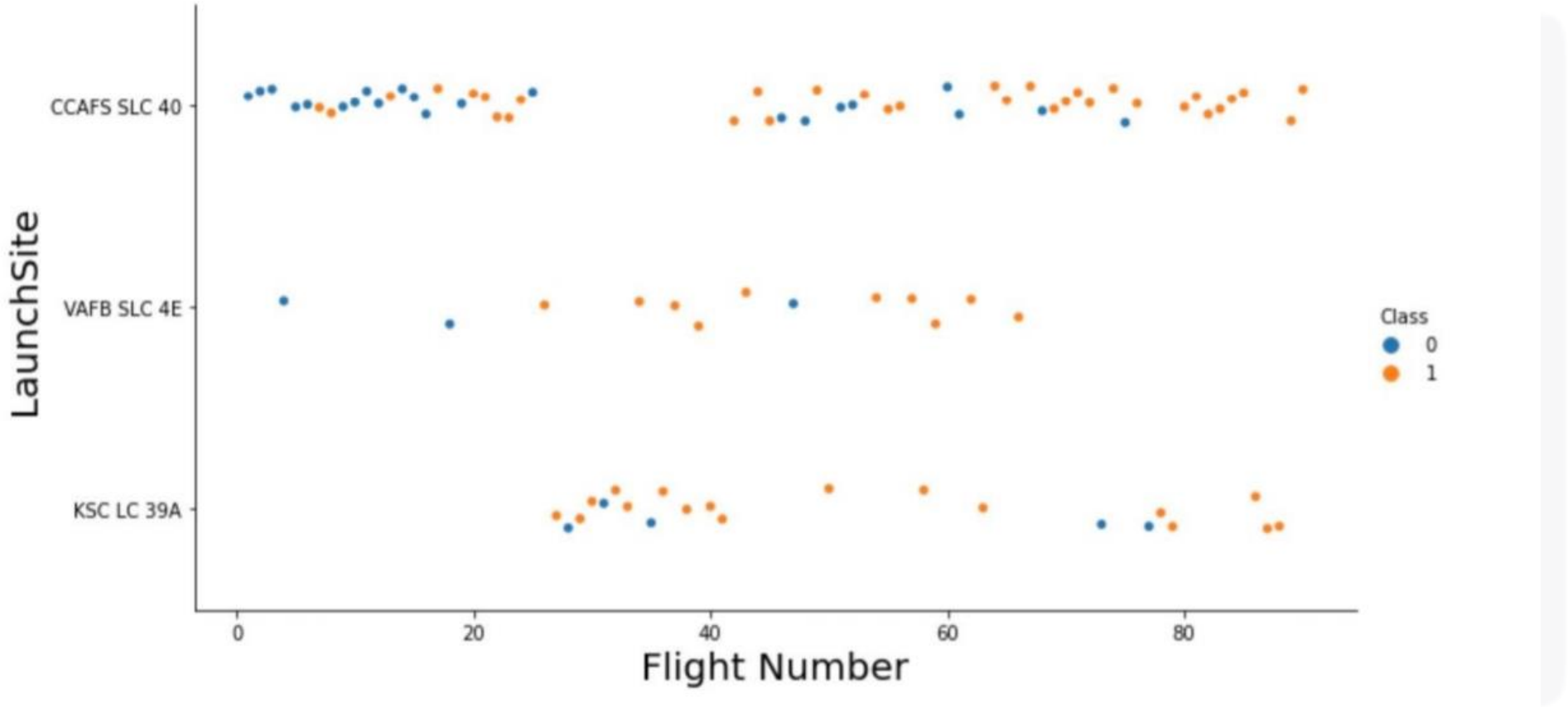
- Scatter plots were used to analyze the relationship between payload and outcome of each launch.

Predictive analysis (Classification)

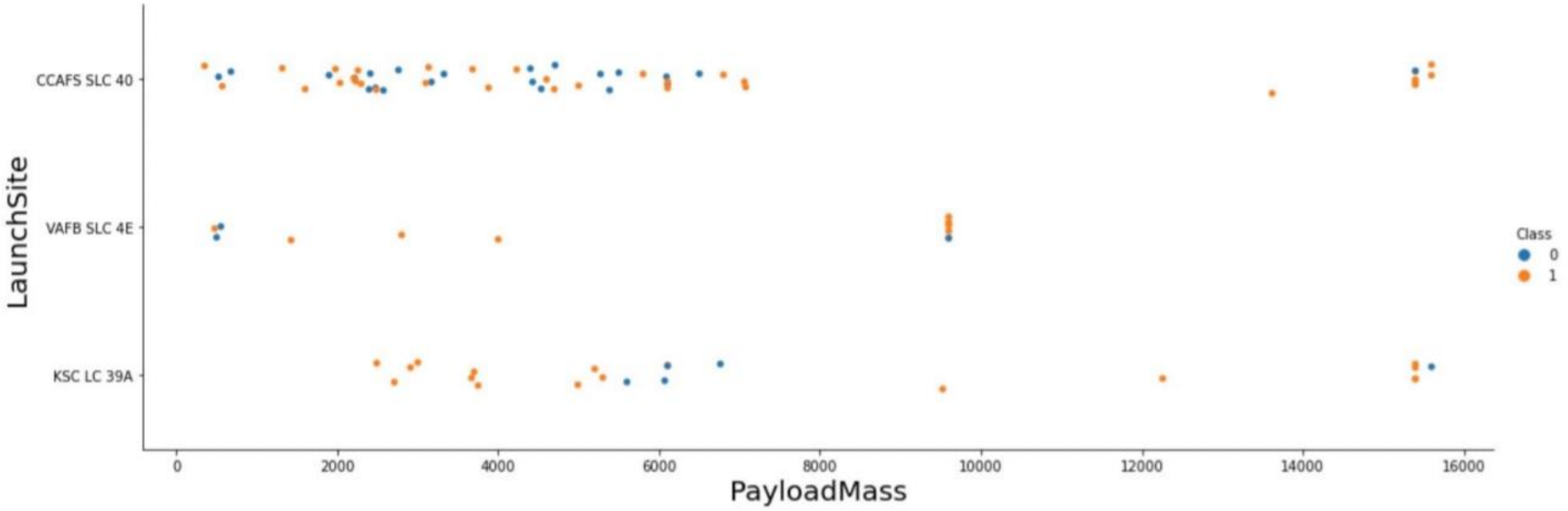
- The data was split into two parts. One part was used for training the data and the other was used for testing the data on the algorithm.
- Used Gridsearch to identify key hyperparameters and best predictive score.
- Using logistic regression, SVM, Decision Tree and KNN. Each regression was fit using the training data and then the model was tested against a separate test data set.
- Confusion matrix was used to demonstrate prediction accuracy

EDA with Visualization

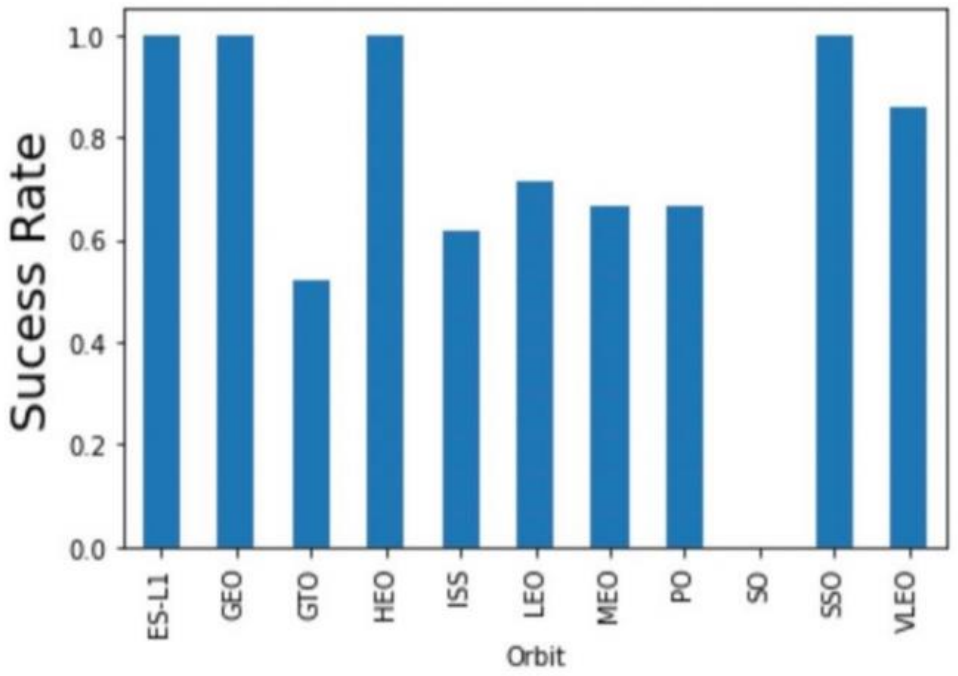
Flight Number vs. Launch Site



Payload vs. Launch Site

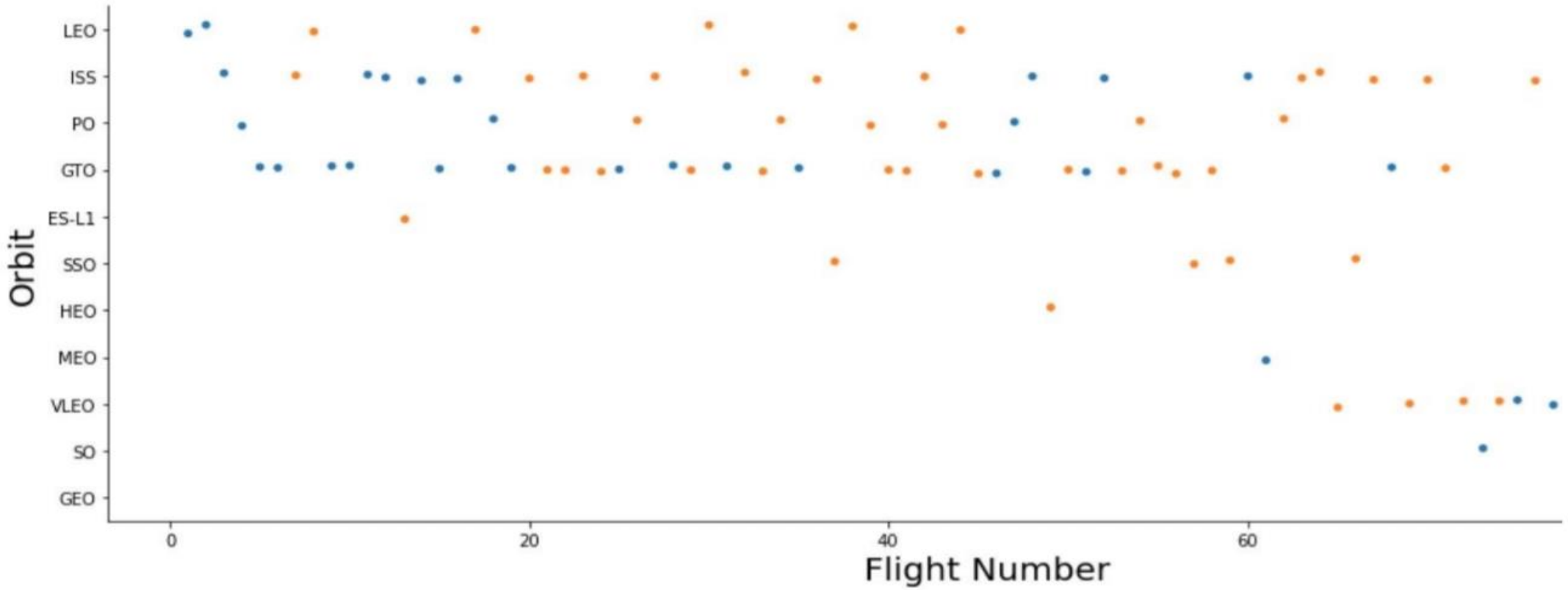


Success rate vs. Orbit type

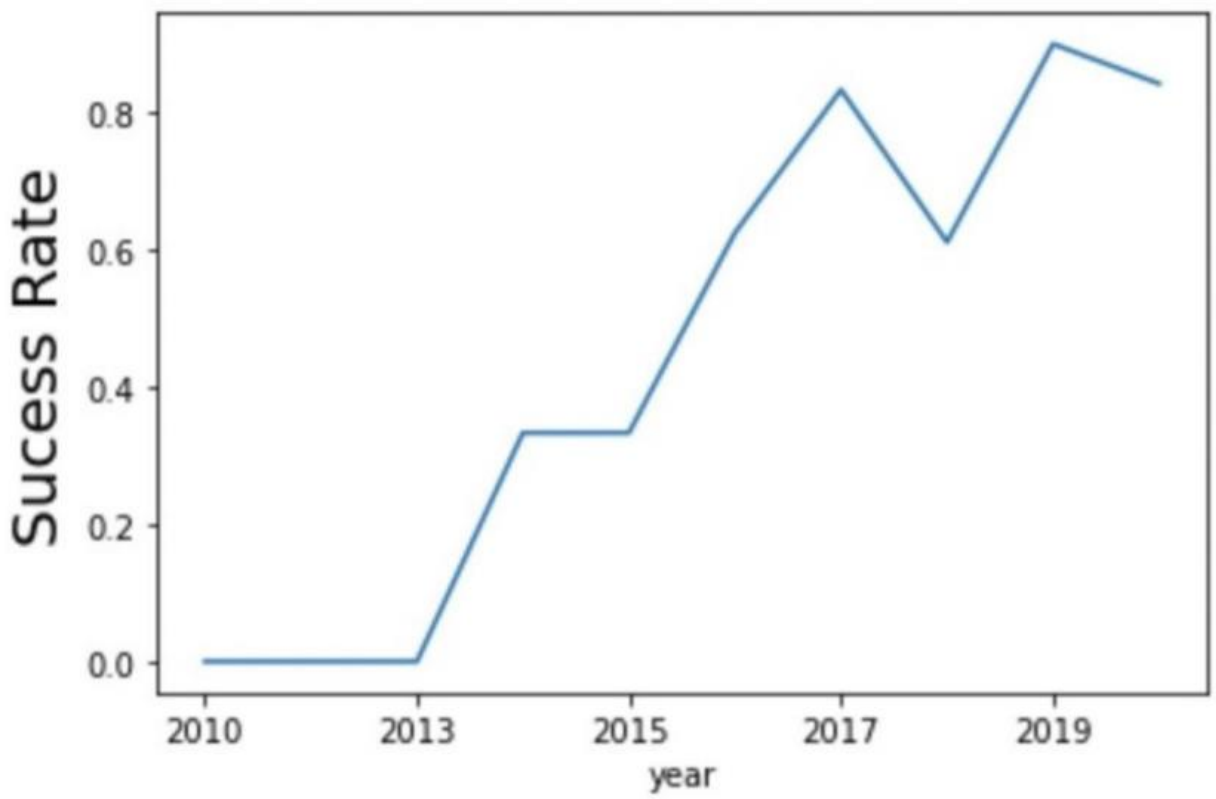


- Success rate is high (close to 100%) for orbit types ES-L1, GEO, HEO, SSO, VLEO.
- Success rate is high (close to 100%) for orbit types FTO, ISS, LEO, MEO and PO
- Success rate is high (close to 100%) for orbit type SO (only one flight)

Flight Number vs. Orbit type



Success Rate vs Year



EDA with SQL

All launch site names

- The unique launch sites are:
- Cape Canaveral Space Launch Complex 40 in California
- CCAFS LC-40, CCAFS SLC -40
- Kennedy Space Center Launch Complex 39 in Florida
- Vandenberg Space Launch Complex 4 in Florida

```
%sql SELECT DISTINCT(launch_site) FROM spacex
```

```
 * ibm_db_sa://bjm34039:***@fbd88901-ebdb-4a4f-a32
1/bludb
Done.
```

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

Total payload mass

- Total payload carried by boosters from NASA is 45,596 lb

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

```
]: %sql SELECT SUM(payload_mass__kg_) FROM spacex WHERE customer = 'NASA (CRS)'

 * ibm_db_sa://bjm34039:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqd
1/bludb
Done.
```

]:

1
45596

Average payload mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is 2,534lb

```
%sql SELECT AVG(payload_mass__kg_) FROM spacex WHERE booster_version LIKE 'F9 v1.1%'
```

```
 * ibm_db_sa://bjm34039:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.app
1/bludb
Done.
```

1
2534

First successful ground landing date

First successful landing outcome in ground pad is December 22, 2015

```
%sql SELECT MIN(DATE) FROM spacex where landing__outcome LIKE '%Success%'
```

```
 * ibm_db_sa://bjm34039:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.
1/bludb
Done.
```

1
2015-12-22

Successful drone ship landing with payload between 4000 and 6000

```
: %sql SELECT booster_version FROM spacex \
  where landing__outcome = 'Success (drone ship)' \
  AND 4000 < payload_mass__kg_ < 6000;

   * ibm_db_sa://bjm34039:***@fbd88901-ebdb-4a4f-a32e-9822b9fb
  Done.
```

3]:

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

Total number of successful and failure mission outcomes

- Calculate the total number of successful and failure mission outcomes
- 99 missions were successful
- 1 mission was successful but payload status unknown.
- 1 mission failed

```
%sql SELECT mission_outcome, COUNT(*) FROM spacex GROUP BY mission_outcome
```

```
 * ibm_db_sa://bjm34039:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tg
Done.
```

]:

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters carried maximum payload

- List the names of the booster which have carried the maximum payload mass

```
%sql SELECT DISTINCT(booster_version) FROM spacex \
WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) from spacex)
```

```
 * ibm_db_sa://bjm34039:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0
1/bludb
Done.
```

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

Rank success count between 2010-06-04 and 2017-03-20

Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order

```
%sql SELECT landing__outcome, COUNT(landing__outcome) AS COUNT_OUTCOME FROM spacex \
WHERE date BETWEEN '2016-06-04' AND '2017-03-20' \
GROUP BY landing__outcome \
ORDER BY COUNT(landing__outcome) DESC
#year(date) =2016
#BETWEEN '04-06-2016' AND '20-03-2017'
```

```
 * ibm_db_sa://bjm34039:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.dat
1/bludb
Done.
```

landing__outcome	count_outcome
Success (drone ship)	2
Success (ground pad)	2
Failure (drone ship)	1
No attempt	1

Rank success count between 2010-06-04 and 2017-03-20

Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order

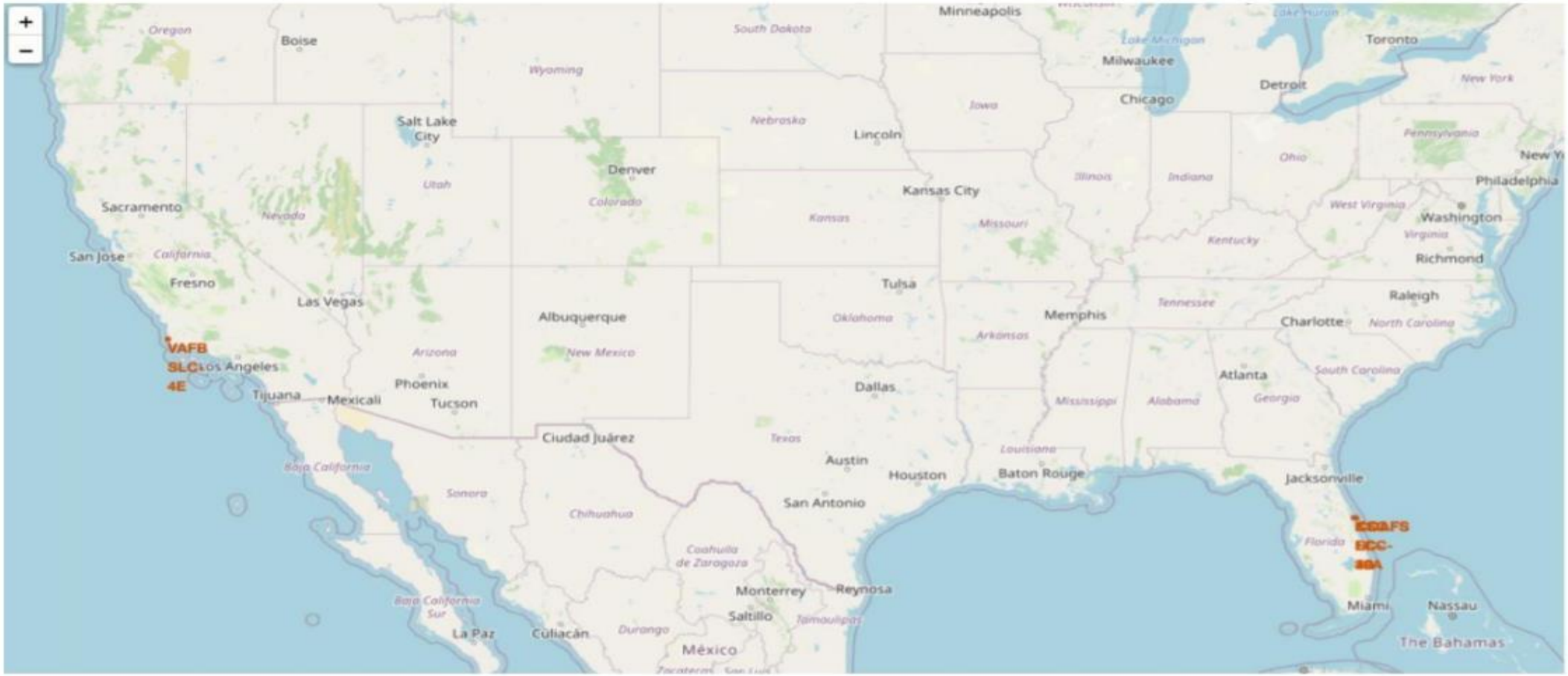
```
%sql SELECT landing__outcome, COUNT(landing__outcome) AS COUNT_OUTCOME FROM spacex \
WHERE date BETWEEN '2016-06-04' AND '2017-03-20' \
GROUP BY landing__outcome \
ORDER BY COUNT(landing__outcome) DESC
#year(date) =2016
#BETWEEN '04-06-2016' AND '20-03-2017'
```

```
 * ibm_db_sa://bjm34039:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.dat
1/bludb
Done.
```

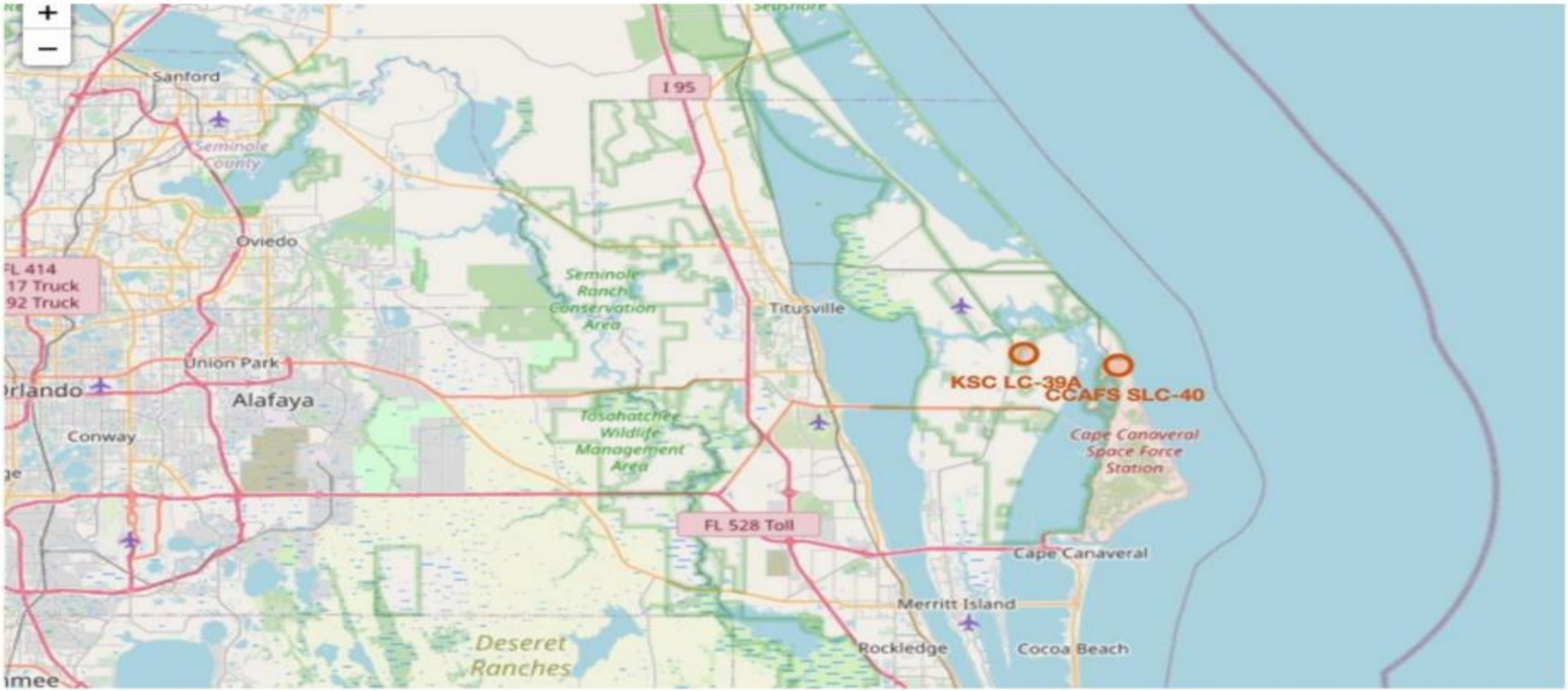
landing__outcome	count_outcome
Success (drone ship)	2
Success (ground pad)	2
Failure (drone ship)	1
No attempt	1

Interactive map with Folium

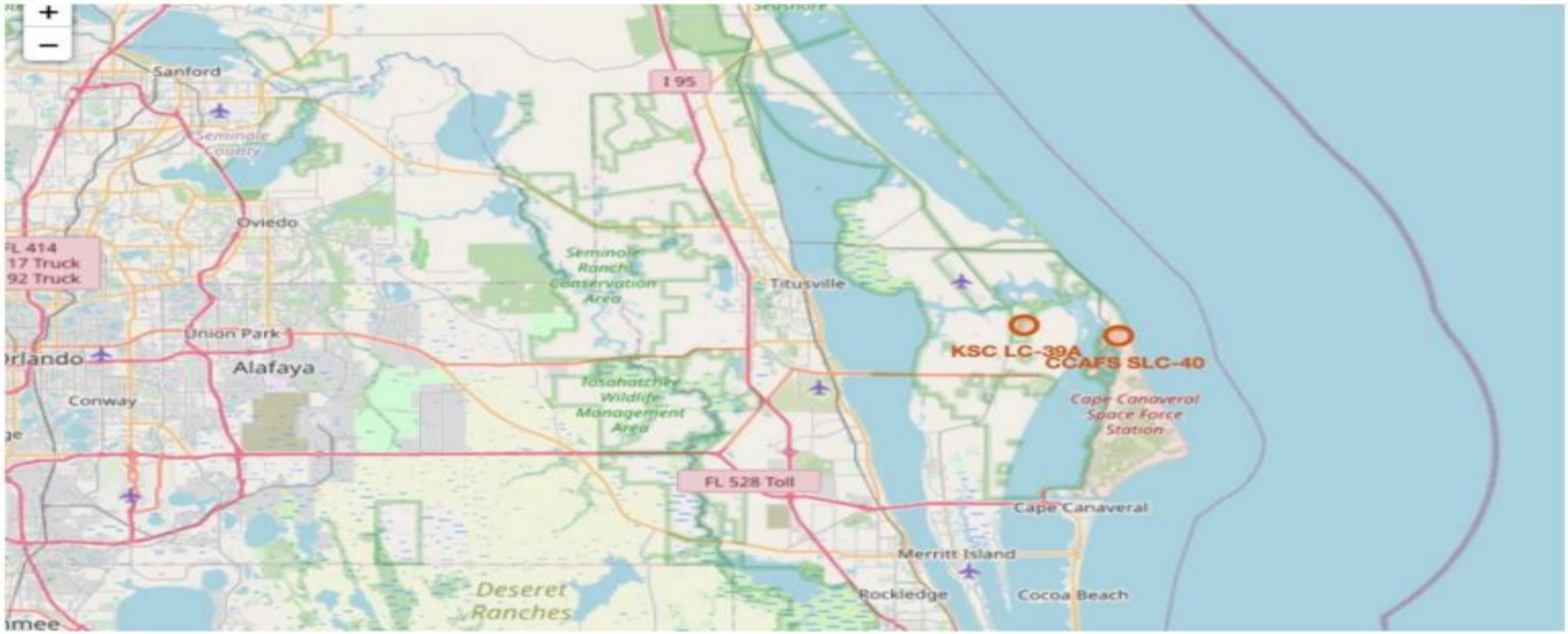
Launch Site Location:



Successful Launches by site



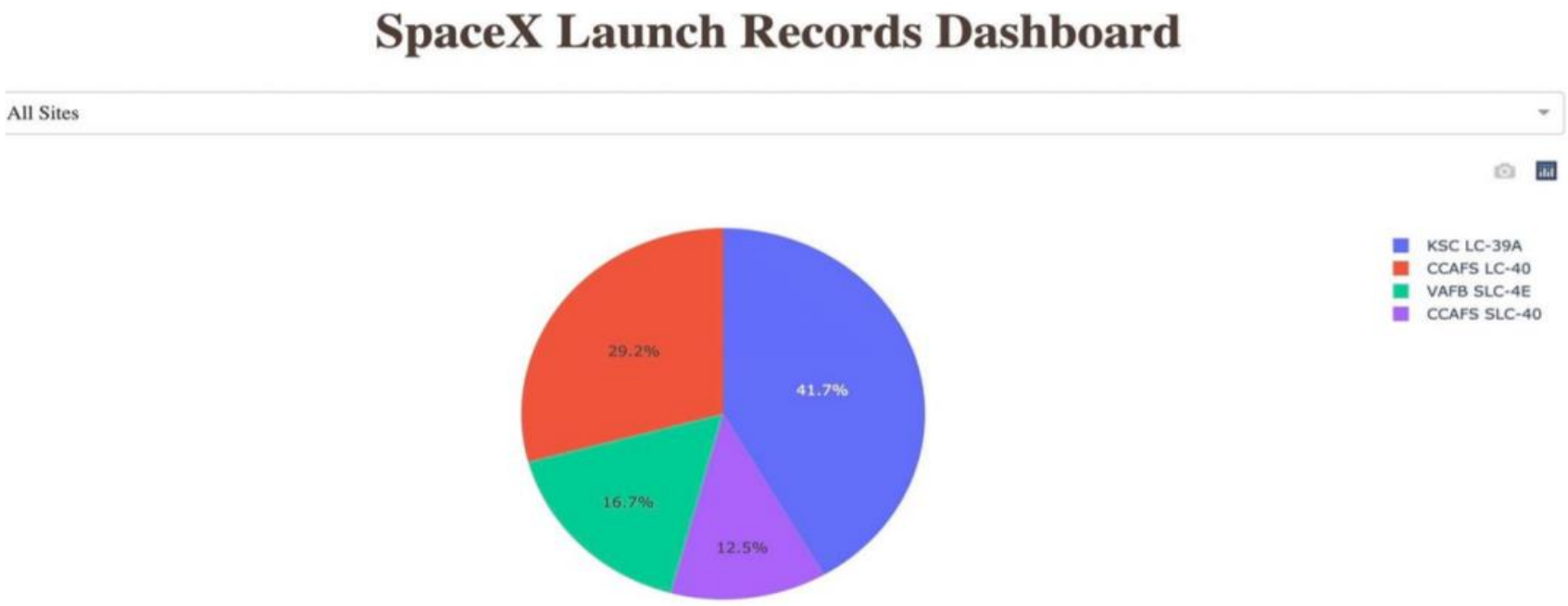
Distance Between Launch Site and Logistic Features



Build a Dashboard with Plotly Dash

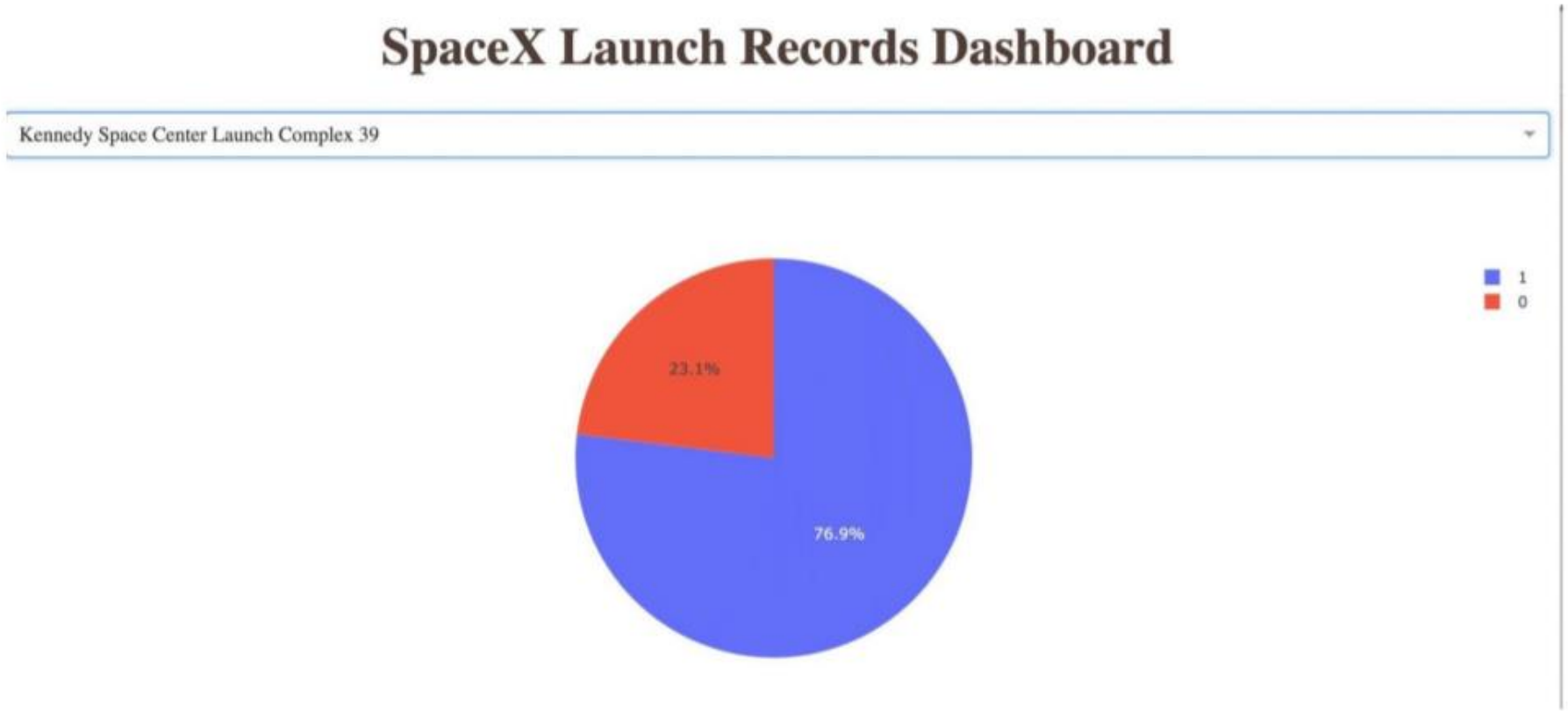
Relanding Success for All Sities

- KSC his 41.7% of the success counts, CCAFS LC-40 has 29.2% of the success counts



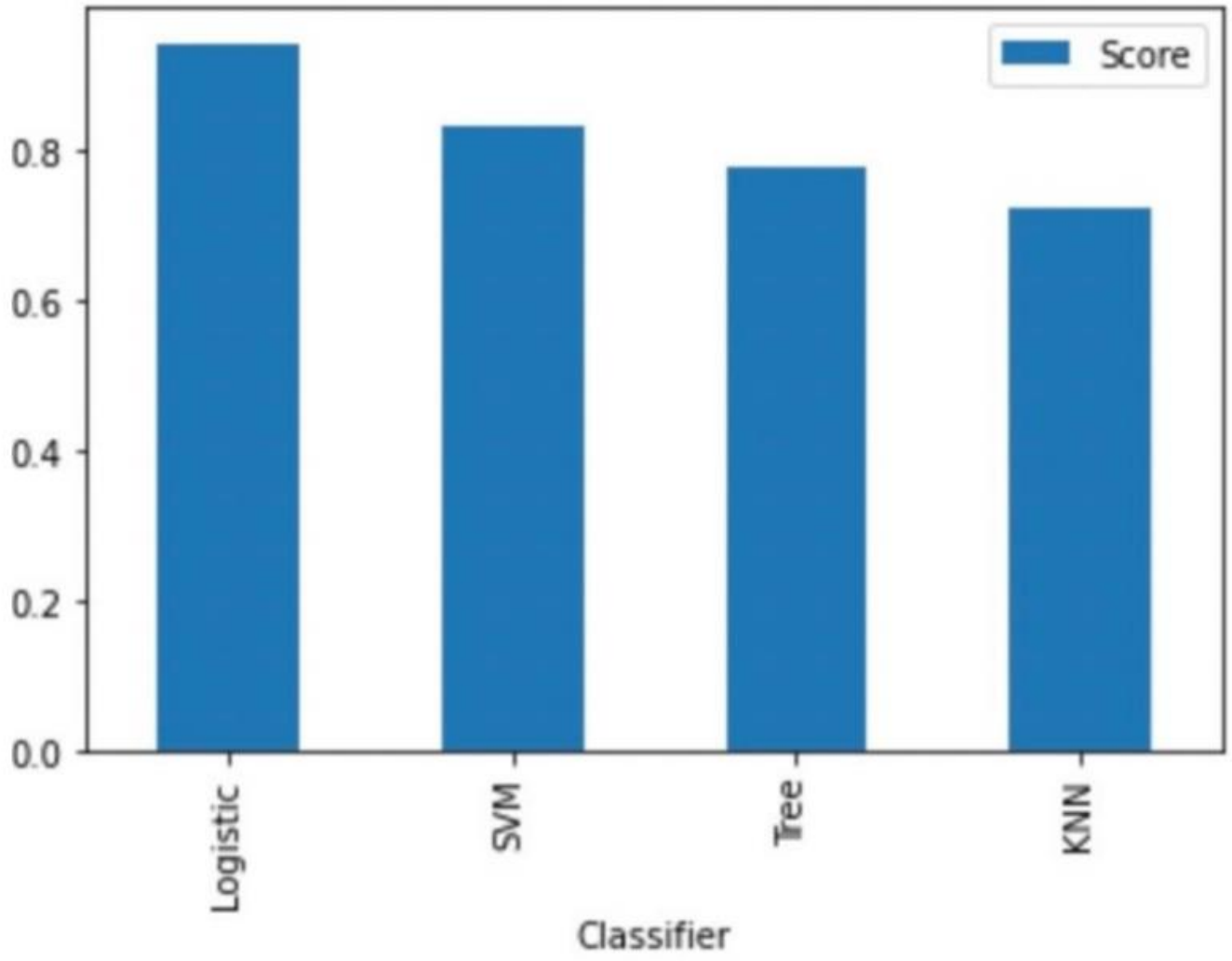
Launch Site with Highest Relanding Success Ratio

- KSC has the highest relanding success ratio, at 76.9%.

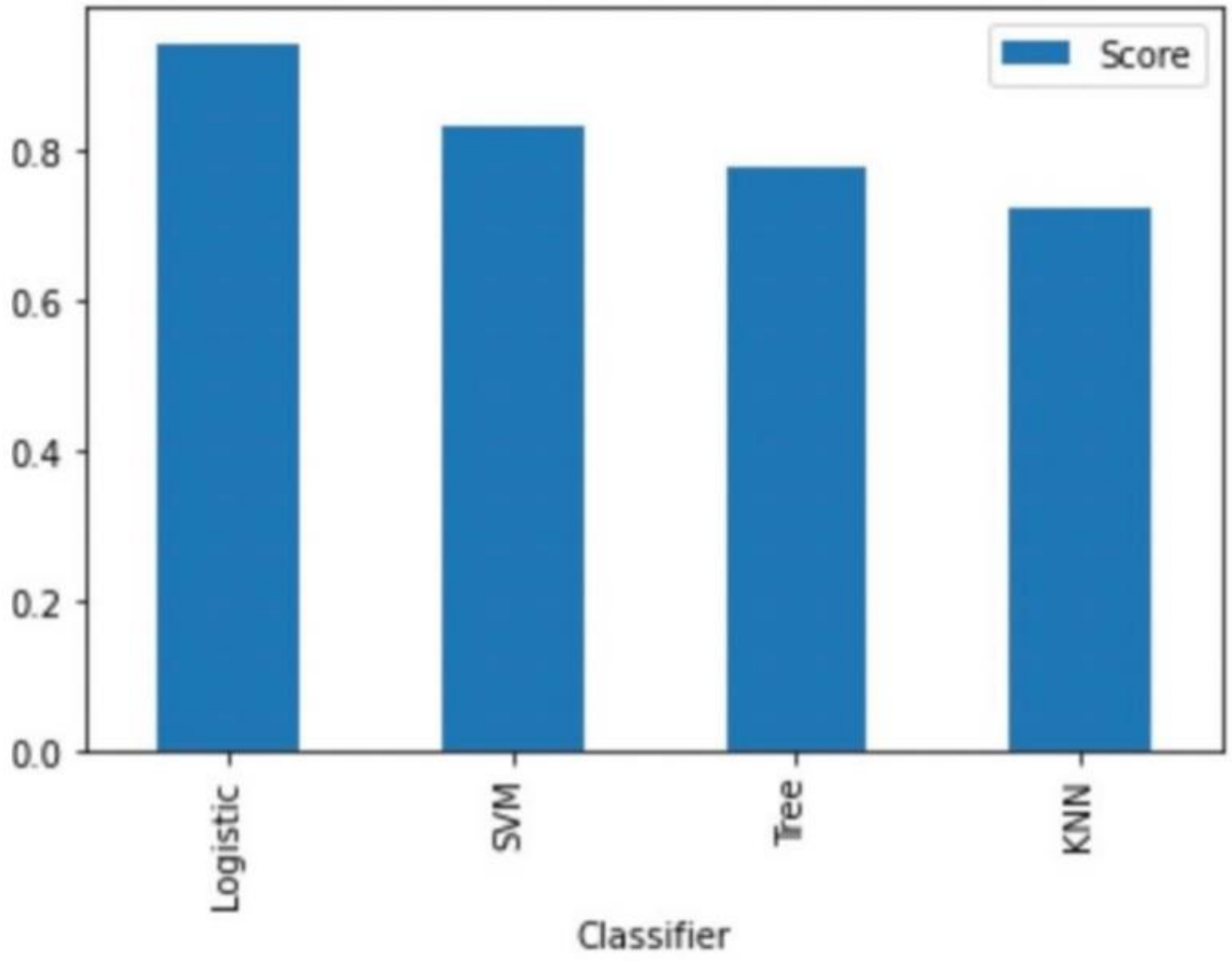


Predictive analysis (Classification)

Classification Accuracy

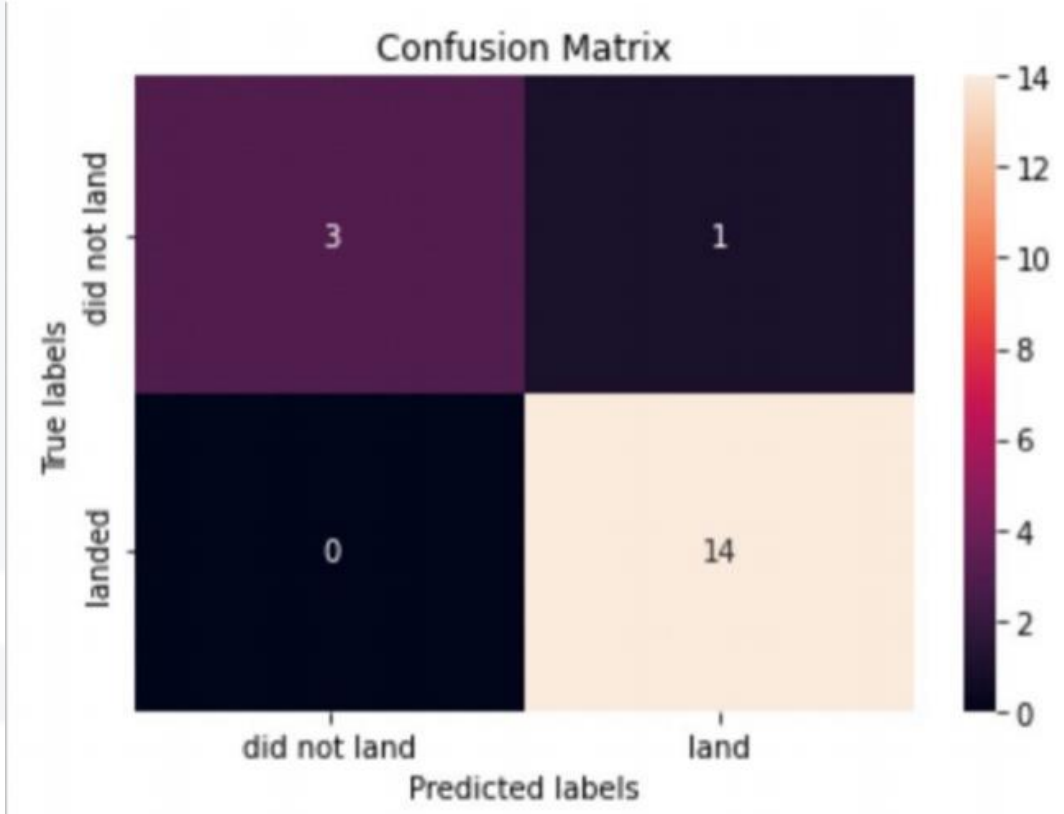


Classification Accuracy



Confusion Matrix

The best accuracy score is from logistic regression classifier. Its accuracy is 94.4%, that is, it makes only one mistake among 18 test cases.



CONCLUSION

- Important factor to predict success include: Launch Number, Desired Orbit, Booster Version and Payload Mass.
- Overall east coast launches are more successful compared to west coast.
- Given the high success rate, SpaceX Falcon 9booster represents a reliable and cheaper alternative to single use rockets.