IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

J Z

Jan 6th, 2024

Outline

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

Executive Summary

Summary of methodologies

- Data Collection
- Data Wrangling
- EDA with Visualization
- EDA with SQL
- Building an Interactive maps with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis

Summary of all results

- Preliminary analysis with based on EDA
- Interactive maps and dashboards
- Predictive results

Introduction

Project background and context

- The project is to predict if the Falcon 9 first stage will land successfully. SpaceX 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. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Problems you want to find answers

- What are the conditions for a successful landing?
- What are the outcome dependent on different variables with success rate?
- What conditions does SpaceX have to achieve to get the best rocket success landing rate?

Section 1

Methodology

Methodology

- Data collection methodology:
  - Using SpaceX Rest API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - Data was cleaned from irrelevant columns and transformed using one hot encoding for Machine Learning
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Plotting: Scatter, Bar and Line graphs to show patterns of data
- Perform interactive visual analytics using Folium and Plotly Dash
  - Using Folium and Plotly Dash Visualization to build interactive maps and dashboards
- Perform predictive analysis using classification models
  - Compare logistic regression model, support vector machine tree decision classifier, KNN by using GridSearchCVto select the best fit model

Data Collection

- The datasets were collected from two sources: API and web page.

1. API

The process includes requesting for data, converting info into dataframe, data wrangling, and exporting to CSV file.

2. Web page

The process includes requesting for data, extracting data, parsing data, and exporting to CSV file.

Data Collection – SpaceX API

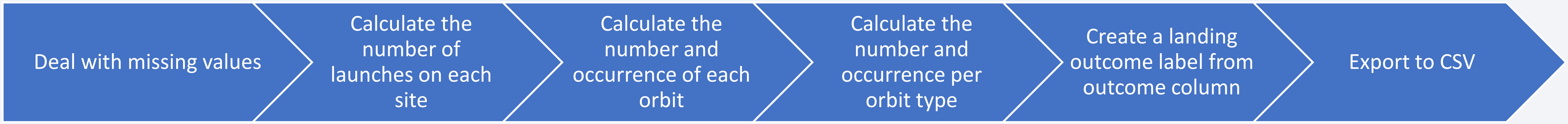
GitHub URL: https://github.com/zj11217/IBMDataScience/blob/main/Week%201%20-%20Data%20Collection%20API.ipynb

Data Collection - Scraping

GitHub URL:

https://github.com/zj11217/IBMDataScience/blob/main/Week%201%20-%20Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling



GitHub URL:

https://github.com/zj11217/IBMDataScience/blob/main/Week%201%20-%20Data%20Wrangling.ipynb

EDA with Data Visualization

- The relationship between Flight Number and Launch Site -> scatter plot
- The relationship between Payload and Launch Site -> scatter plot
- The relationship between success rate of each orbit type -> bar plot
- The relationship between Flight Number and Orbit type -> scatter plot
- The relationship between Payload and Orbit type -> scatter plot
- The launch success yearly trend -> line chart

The scatter plot can best describe the relationship between two categorical fields

The bar plot can best compare different groups

The line plot can best show the time series data and trends

GitHub:

https://github.com/zj11217/IBMDataScience/blob/main/Week%202%20-%20EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

The SQL queries performed:

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 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 in year 2015
- Ranking 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

GitHub: https://github.com/zj11217/IBMDataScience/blob/main/Week%202%20-%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Circle marker around each launch site was added with a label of the name of the launch site
- A green marker was added if a launch was successful and a red marker was added when a launch was failed
- Lines were used to show the distance between launch site and its proximities.

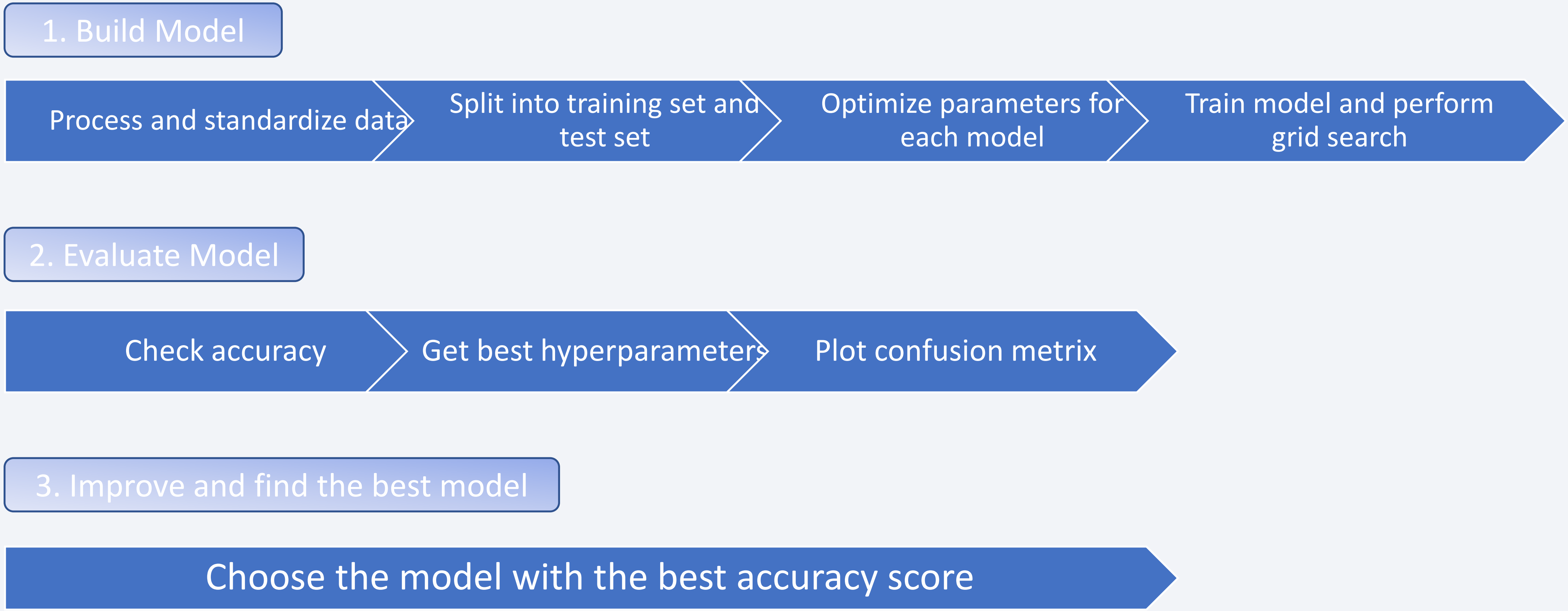
GitHub URL: https://github.com/zj11217/IBMDataScience/blob/main/Week%203%20-%20Data%20Visualization%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Pie chart was added, as it shows the success rate of the launch sites, and displays and proportion between success and fails of the launch site.
- Scatter Plot was also added, as it shows the correlation between Mission Outcome and Payload Mass for different Booster versions for all sites or a given launch site. The payload Mass can be filtered by a range slicer.

GitHub URL: https://github.com/zj11217/IBMDataScience/blob/main/Week%204%20-%20Machine%20Learning%20Prediction.ipynb

Predictive Analysis (Classification)



GitHub URL: https://github.com/zj11217/IBMDataScience/blob/main/Week%204%20-%20Machine%20Learning%20Prediction.ipynb

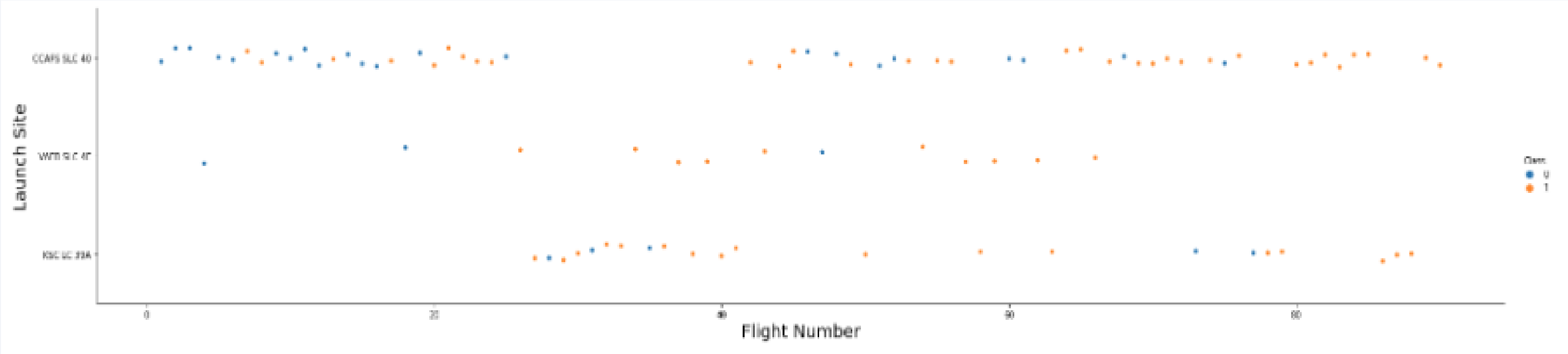
Results

- Exploratory data analysis results
  - KSC LC-39AandVAFB SLC 4Ehas a success rate of 77%
  - VAFB SLC 4E has no payload above 10000 kg
  - In the LEO orbit the Success appears related to the number of flights
  - With heavy payloads the successful landing or positive landing rate are more for Polar,LEOand ISS
  - The sucessrate since 2013 kept increasing till 2020
- Interactive analytics demo in screenshots
- Predictive analysis results
  - 4 ture positive, 7 true negative, 5 false positive and 2 false negative. The accuracy of the mode is around 89% with the best parameters.

Section 2

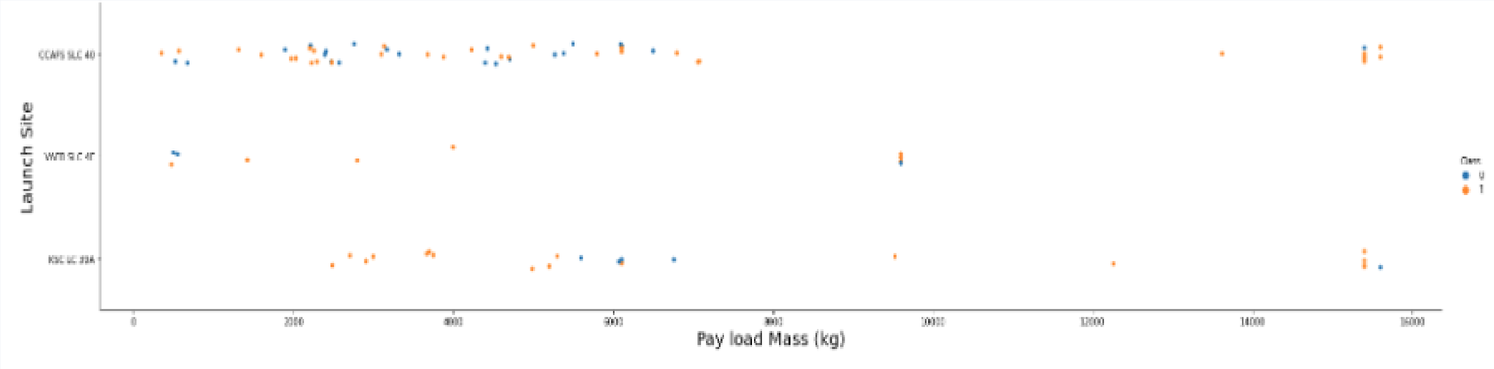
Insights drawn from EDA

Flight Number vs. Launch Site



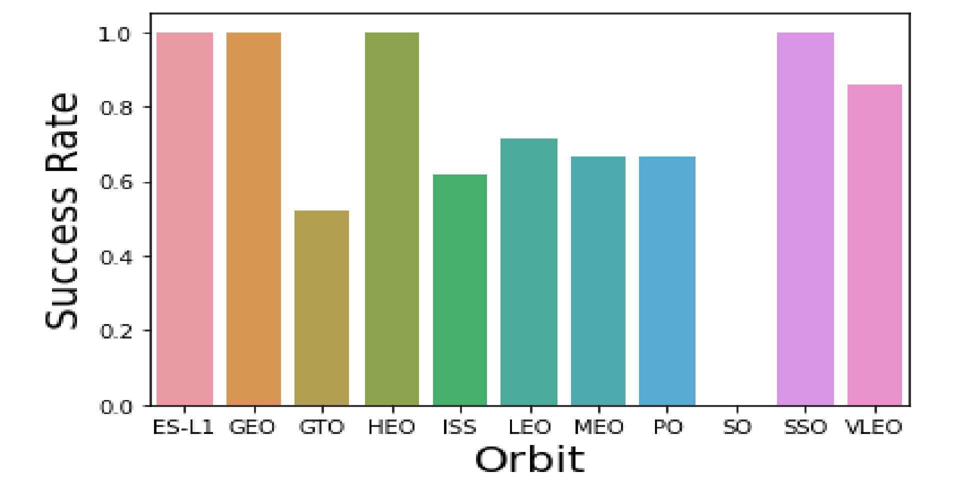
- KSC LC 39A has the highest successful rate, CCAFS SLC 40 has the lowest

Payload vs. Launch Site



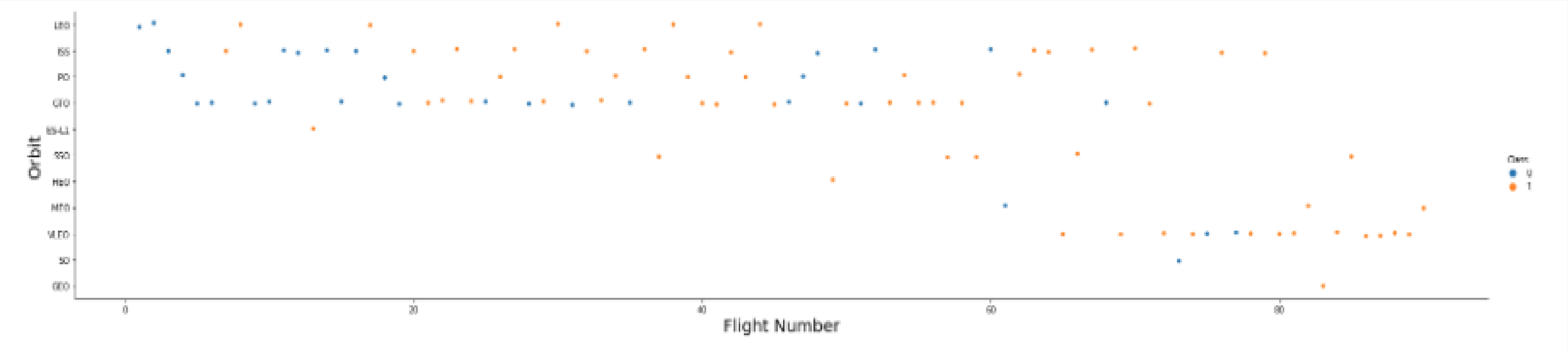
- VAFB SLC 4E has no payload above 10000 kg

Success Rate vs. Orbit Type



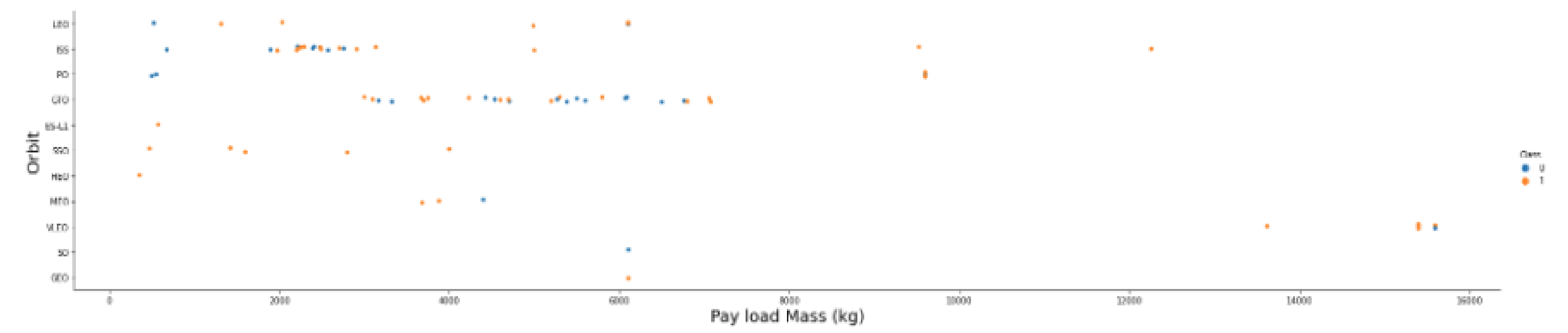
- ES-L1, GEO, HEO, SSO have the highest success rate

Flight Number vs. Orbit Type



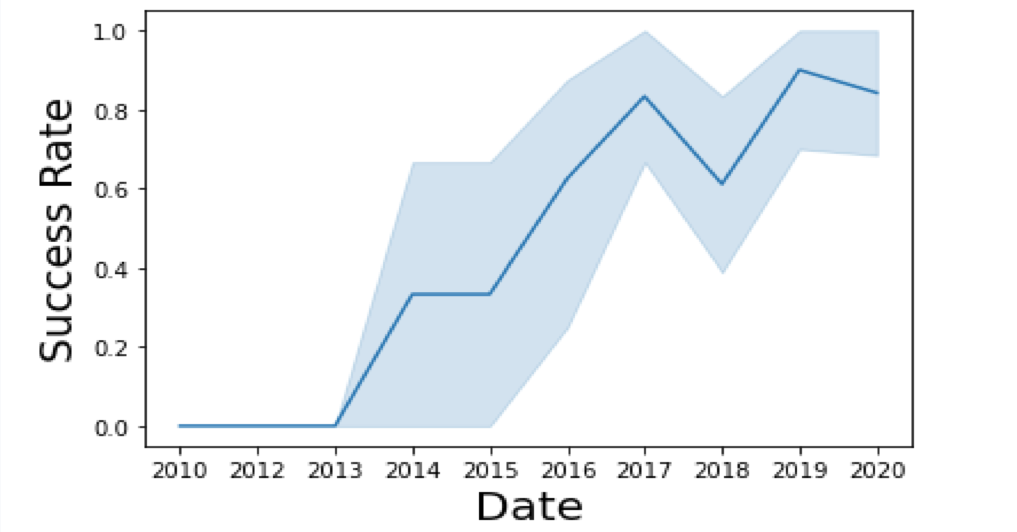
- LOE orbit success appears to relate to the flight numbers
- GEO orbit appears to have no correlation with flight numbers
- VLEO orbit has high success rate

Payload vs. Orbit Type



- GTO orbit appears to have no relationship with payload
- ISS and LEO orbits appear to have higher success rate when payload increase
- SSO has higher success rate when payload decrease

Launch Success Yearly Trend



- There was no success in and before 2013
- The success rate increased from 2013 to 2017, and reached the highest in 2019

All Launch Site Names

```
%sql select distinct launch_site from spacextbl
```

```
 * ibm_db_sa://zvg47118:***@ba99a9e6-d59e-4883-8fc0-c
Done.
```

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

- There are 4 unique launch sites

Launch Site Names Begin with 'CCA'

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

```
* ibm_db_sa://zvg47118:***@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

- The 5 records are from launch site CCAFS LC-40, the landing outcomes are failure/no attempt.

Total Payload Mass

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

```
* ibm_db_sa://zvg47118:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7
one.
```

total_payload_mass
45596

- The total payload carried by boosters from NASA is 45596 kg

Average Payload Mass by F9 v1.1

```
%%sql select avg(payload_mass__kg_) avg_payload_mass
from spacextbl where booster_version = 'F9 v1.1'
```

```
* ibm_db_sa://zvg47118:***@ba99a9e6-d59e-4883-8fc0-d6a8
Done.
```

avg_payload_mass
2928

- The average payload mass carried by booster version F9 v1.1 is 2928kg

First Successful Ground Landing Date

```
%%sql select min(date) min_date
from spacextbl where landing__outcome = 'Success (ground pad)'
```

```
* ibm_db_sa://zvg47118:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1
Done.
```

min_date
2015-12-22

- The date of the first successful landing outcome on ground pad is Dec 22, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql select booster_version from spacextbl
where landing__outcome = 'Success (drone ship)'
and payload_mass__kg_ > 4000 and payload_mass__kg_ < 6000
```

```
 * ibm_db_sa://zvg47118:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a
Done.
```

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

- There are 4 boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

```
%%sql select
(case when mission_outcome like '%Success%' then 'Success' else 'Failure' end) mission_outcomes,
count(*) qty
from spacextbl group by (case when mission_outcome like '%Success%' then 'Success' else 'Failure' end
```

```
 * ibm_db_sa://zvg47118:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomai
Done.
```

mission_outcomes	qty
Failure	1
Success	100

- There are 1 failure and 99 successes.

Boosters Carried Maximum Payload

```
%%sql select booster_version from spacextbl
where payload_mass__kg_ = (select max(payload_mass__kg_) from spacextbl)
```

```
* ibm_db_sa://zvg47118:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tg
Done.
```

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

- Different booster version has different max payload mass.

2015 Launch Records

```
%%sql select booster_version,launch_site from spacextbl
where landing__outcome = 'Failure (drone ship)' and year(date) = 2015
```

```
 * ibm_db_sa://zvg47118:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgt
Done.
```

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- In January and April, 2015 there are launch failure by booster B1012 and B1015.

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

```
%%sql select landing__outcome,count(landing__outcome) qty from spacextbl
where (date between '2010-06-04' and '2017-03-20')
group by landing__outcome order by 2 desc
```

```
* ibm_db_sa://zvg47118:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu
Done.
```

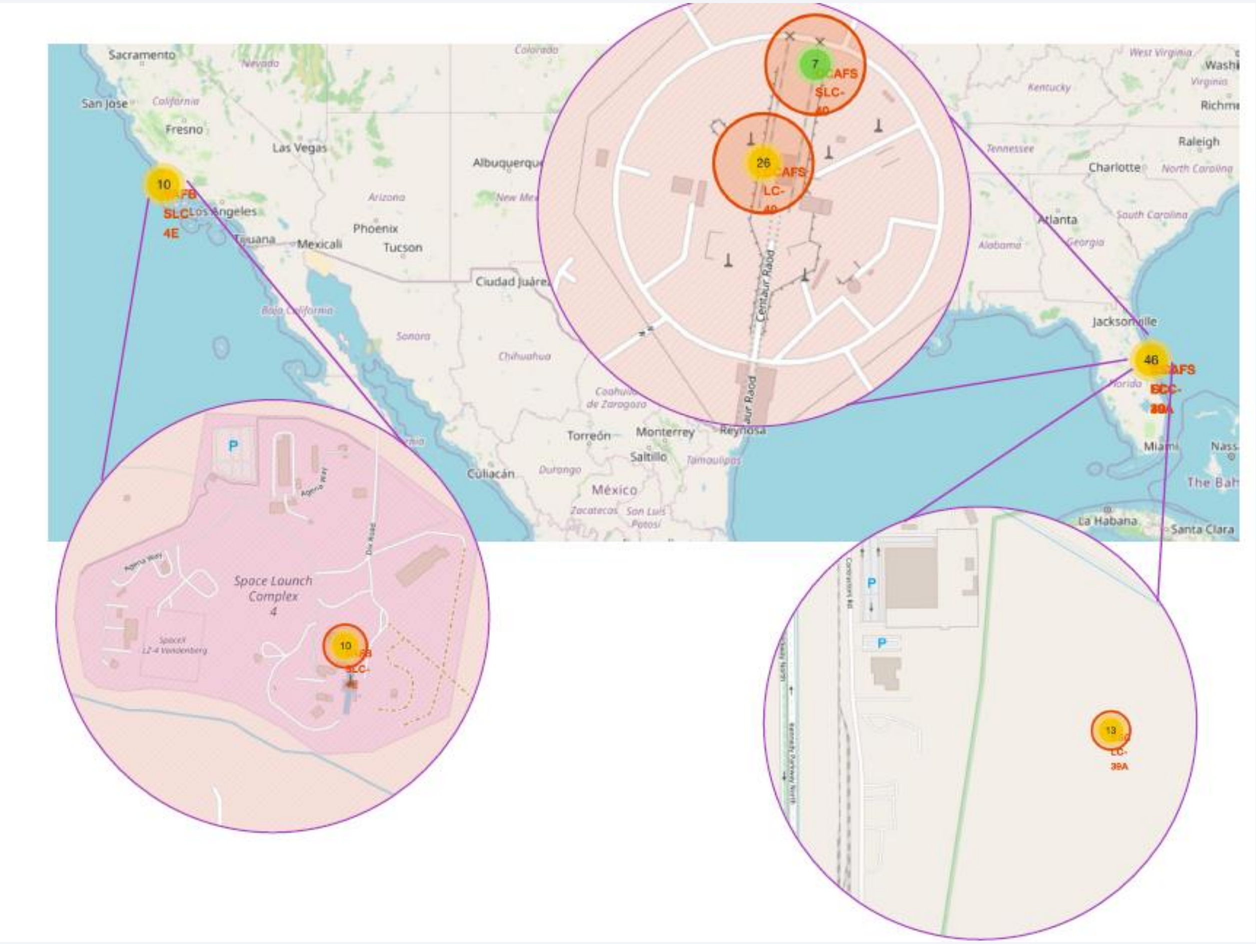
landing__outcome	qty
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- There are total 31 outcomes listed as above

Section 3

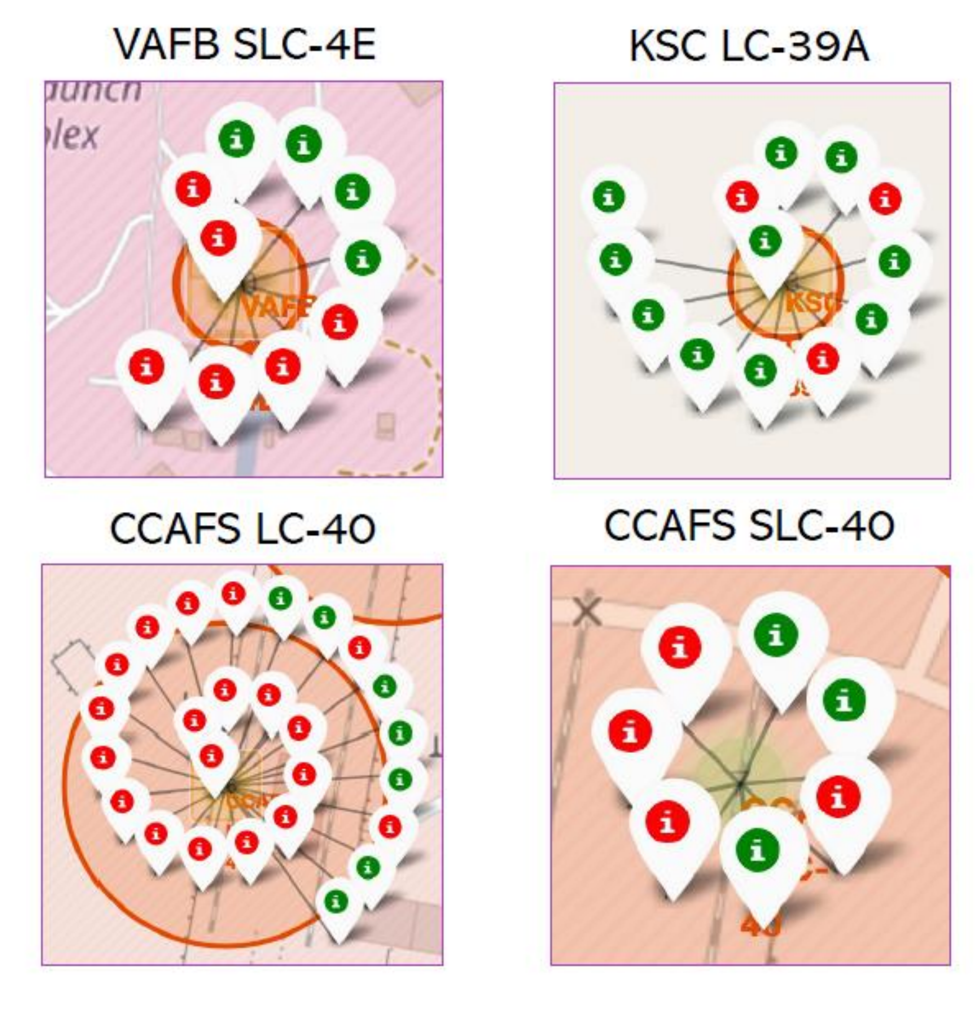
Launch Sites Proximities Analysis

SpaceX launch sites



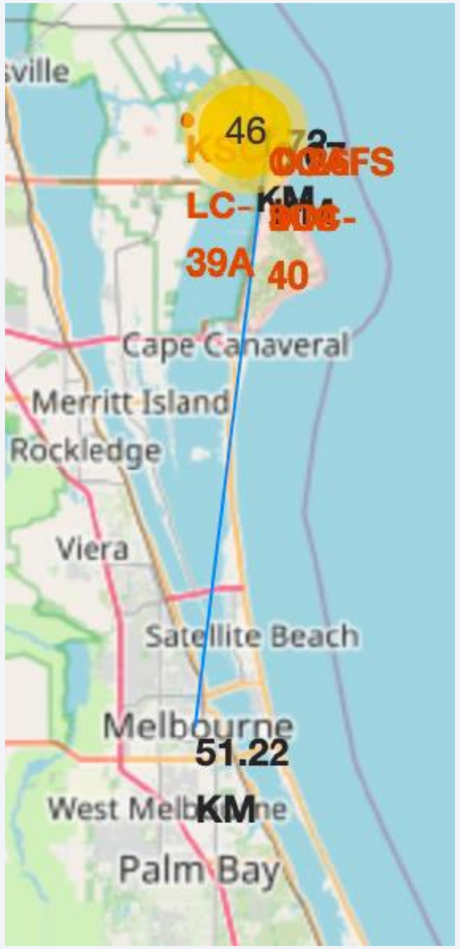
- All the launch sites are close to the coast, to minimize the risk of any incident.

SpaceX launch site success rate



- Green mark shows successful landing
- Red mark shows failed landing
- KSC LC-39A has the highest success rate

SpaceX launch site success rate

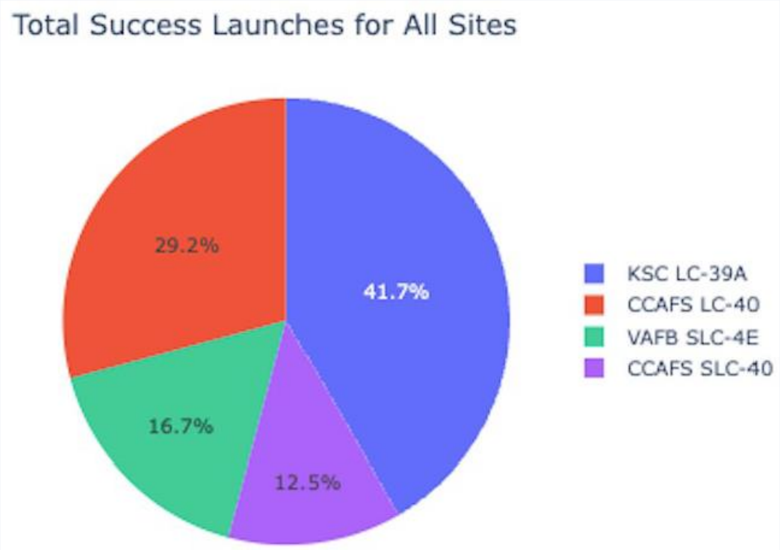


- CCAFS SLC-40 is located close to the coast, to minimize the risk in case of failure
- CCAFS SLC-40 is located close to highway and railway, to enable access to transportation

Section 4

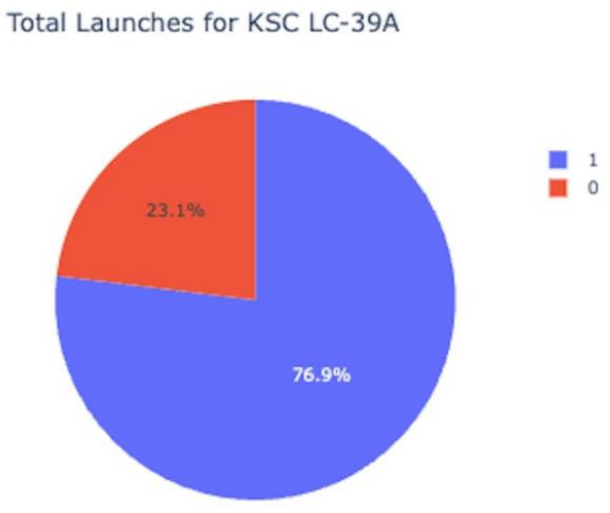
Build a Dashboard with Plotly Dash

Success launch for all sites



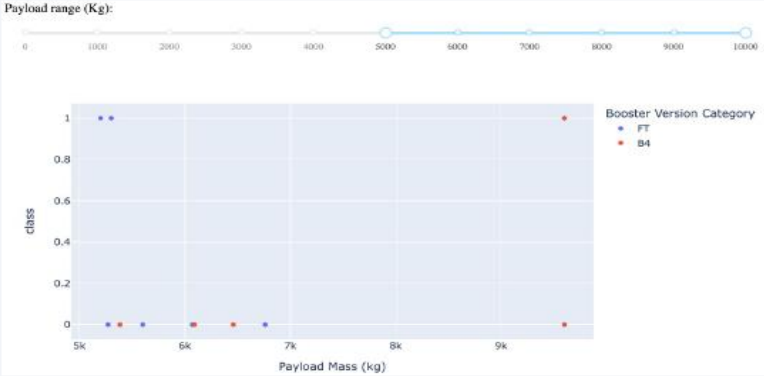
- KSC-LC-39A has the most successful launches
- CCAFS LC-40 has the second successful launches

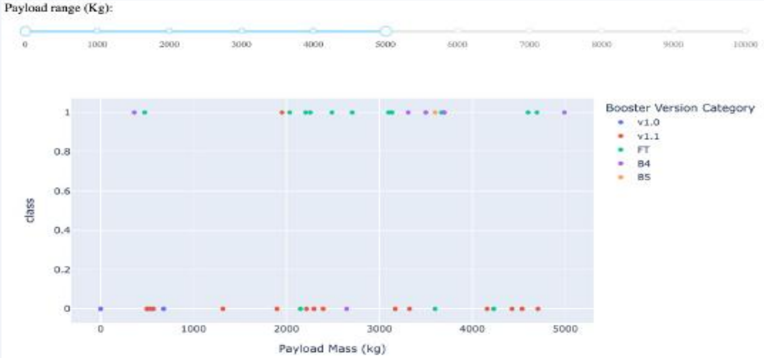
Launch success ratio

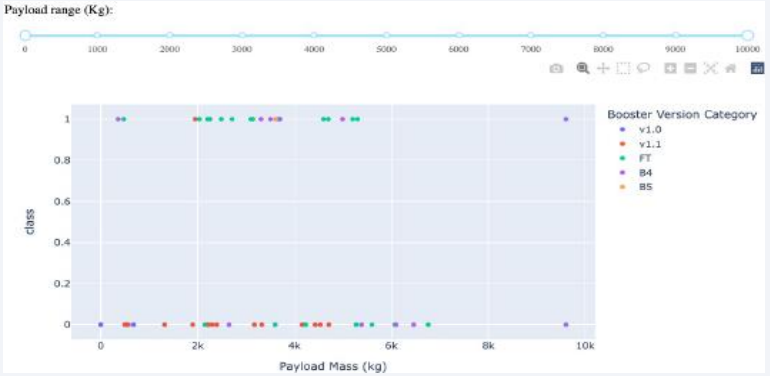


- KSC LC-39A has the success rate 76.9%

Payload and launch outcomes





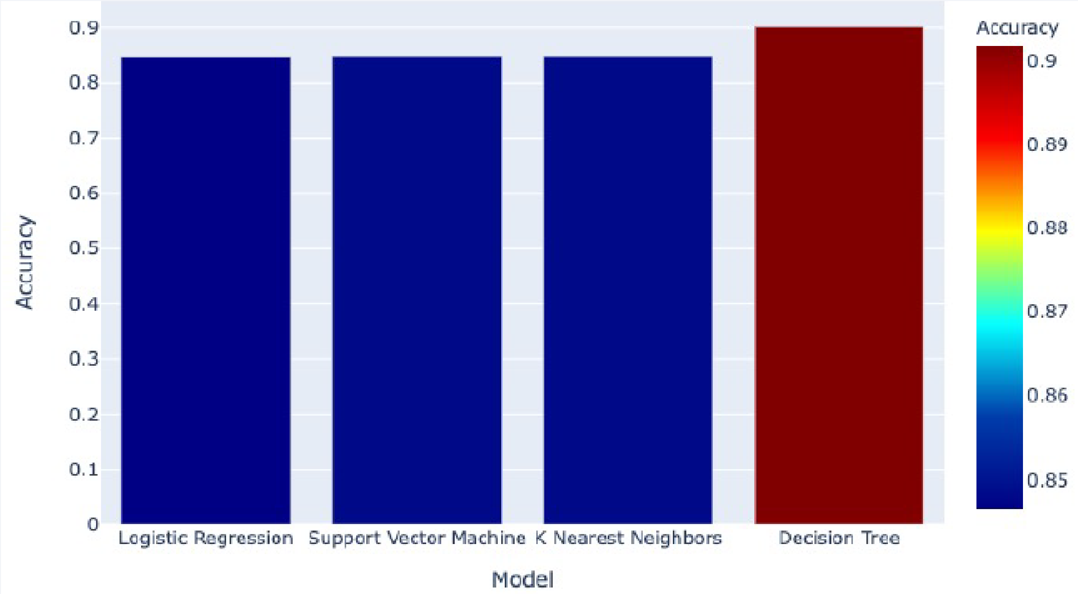


- Payload up to 5000kg has a higher success rate than payload over 5000kg
- FT booster has the highest launch success rate

Section 5

Predictive Analysis (Classification)

Classification Accuracy



- Decision tree has the best accuracy

Confusion Matrix

- Decision tree has the best balance

Conclusions

- There is a relationship between launch site and success rate. Payload mass is also related with the success rate
- The success rate has increased since 2013
- The decision tree model has the highest accuracy.

Appendix

- https://www.coursera.org/learn/applied-data-science-capstone/home/week/5

Thank you!