IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Oscar Quiroga
2024-06-17

Outline

Executive Summary

- A series of exploratory analysis and predictive classification were done to landings of the first phase of the rocket launcher.

- Accompanied with visualization in charts, graphs, dashboards and maps.

- Some candidate landing sites, booster types and orbit types for future missions were found.

Introduction

- SpaceX highly competitive because they can recover the first phase of the rocket launcher

- We want to find the best parameters of a successful landing to predict the price of a landing of the first phase.

Section
1
Methodology

Methodology

Executive Summary

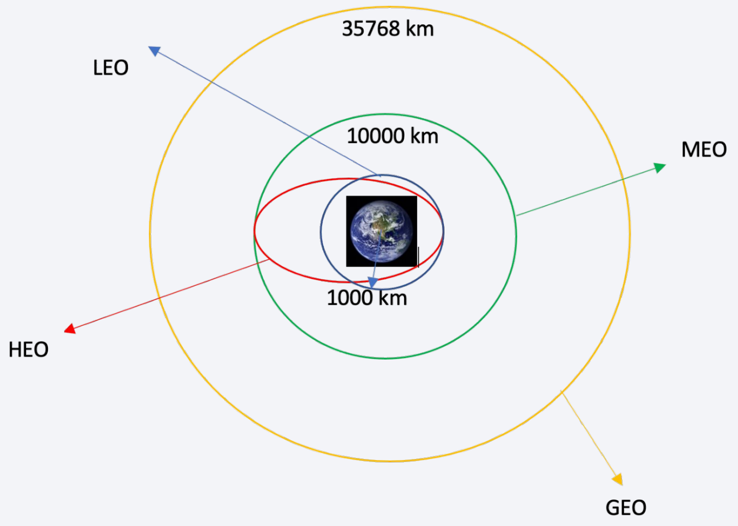
- Data collection methodology:
  - Data was collected from SpaceX Launch data own library and webscraping.
- Perform data wrangling
  - Data was cleaned for missing values, preprocessed and new features were added for analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Grid search was used with different methods evaluated using accuracy and confusion matrix splitting the data in training and test datasets.

Data Collection

- From the API and web scraping the fields were:

  - Launch site
  - Payload mass in Kg
  - Orbit: the target orbit
  - Date
  - Booster version
  - Launch Outcome
  - Booster landing

- API URL: https://api.spacexdata.com/v4
- Scraping source: https://en.wikipedia.org



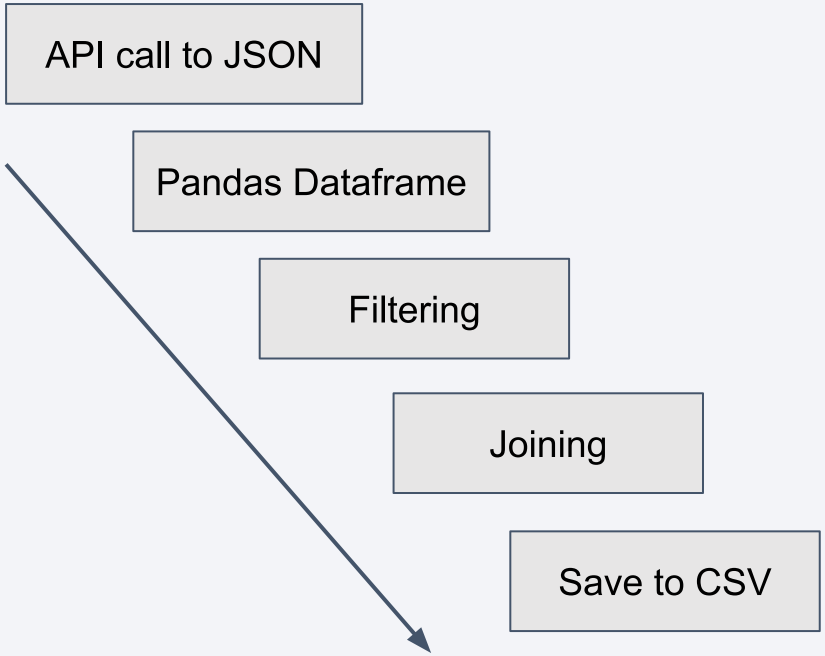
Earth Orbits

Data Collection – SpaceX API

- Filtering:
  - Launches with one payload
  - Falcon 9 launches only
  - Remove unnecessary fields

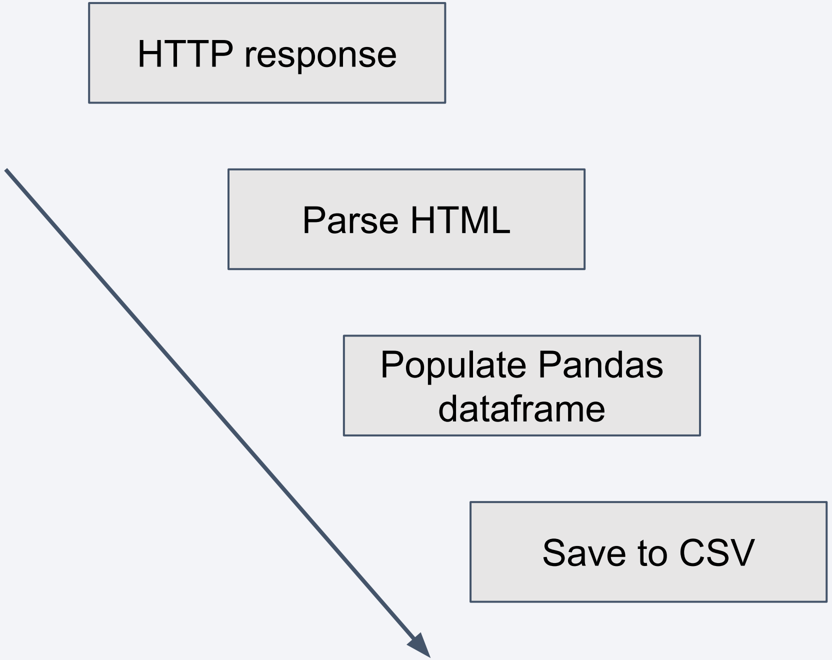
- Joining fields: IDs used to extract fields from other tables in API

- GitHub URL of data collection https://github.com/osjuquiga/IBM-Data-Sciene-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Beautifulsoup to parse HTML
- Data frame populated with iteration over all relevant HTML tables with launch entries
- GitHub URL of webscrapping https://github.com/osjuquiga/IBM-Data-Sciene-Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- From Collected SpaceX API
- Null values replaced with mean of field.
- Exploratory data analysis (EDA)
- Data wrangling GitHub URL https://github.com/osjuquiga/IBM-Data-Sciene-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA target labels for ML classification

Landing Outcome	Outcome Label
True ASDS	True
None None	False
True RTLS	True
False ASDS	False
True Ocean	True
False Ocean	False
None ASDS	False
False RTLS	False

EDA with Data Visualization

- Fields Orbit, Year, Flight number, Launch Site and Payload were visualized against each other to find patterns in data.

- Mostly Seaborn charts

- Prepare data for feature engineering

- GitHub URL https://github.com/osjuquiga/IBM-Data-Sciene-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Webscrapped data loaded into sqlite, Jupyter magic statements for querying

- General Payload statistics about some clients like NASA, booster type and launch site. Also statistics in time.
- Determine the cost of a launch if first stage will land.

- GitHub URL
  https://github.com/osjuquiga/IBM-Data-Sciene-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Markers and pop ups for launch locations
- Find patterns in proximity to other nearby locations like beaches, railways or streets
- Find success rate of location

- GitHub URL of interactive map https://github.com/osjuquiga/IBM-Data-Sciene-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Build a Dashboard with Plotly Dash

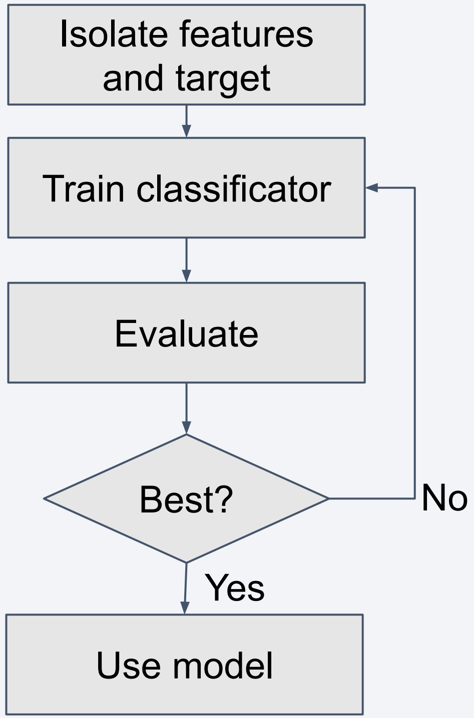
- interactive dashboard with dropdown launching station and date ranges for successful launches

- The dashboard focus on the most viable launching station in time.

- GitHub URL of Plotly Dash https://github.com/osjuquiga/IBM-Data-Sciene-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- A series of machine learning methods for classification, SVM, KNN, Logistic regression and decision trees were tested
- One hot vector for categorical variables
- GitHub URL predictive analysis lab https://github.com/osjuquiga/IBM-Data-Sciene-Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

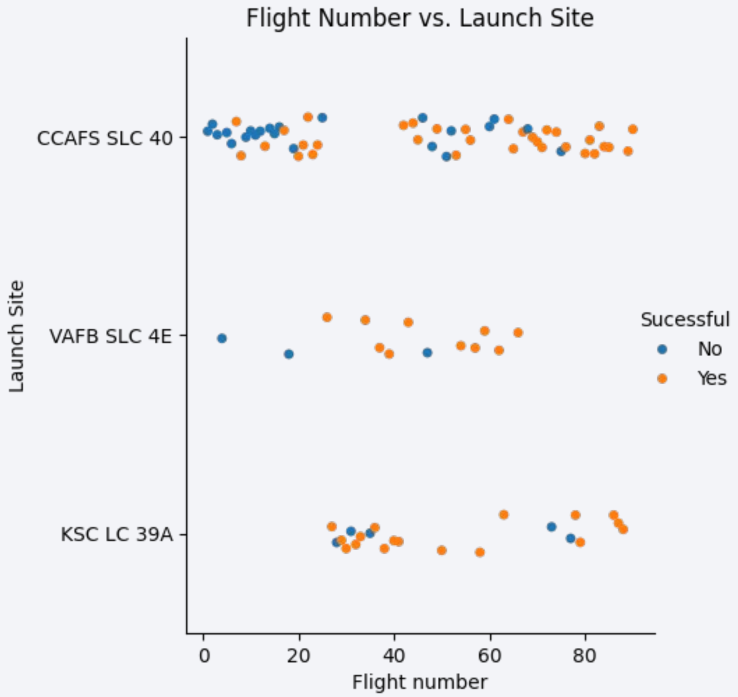


Section
2

Insights drawn from EDA

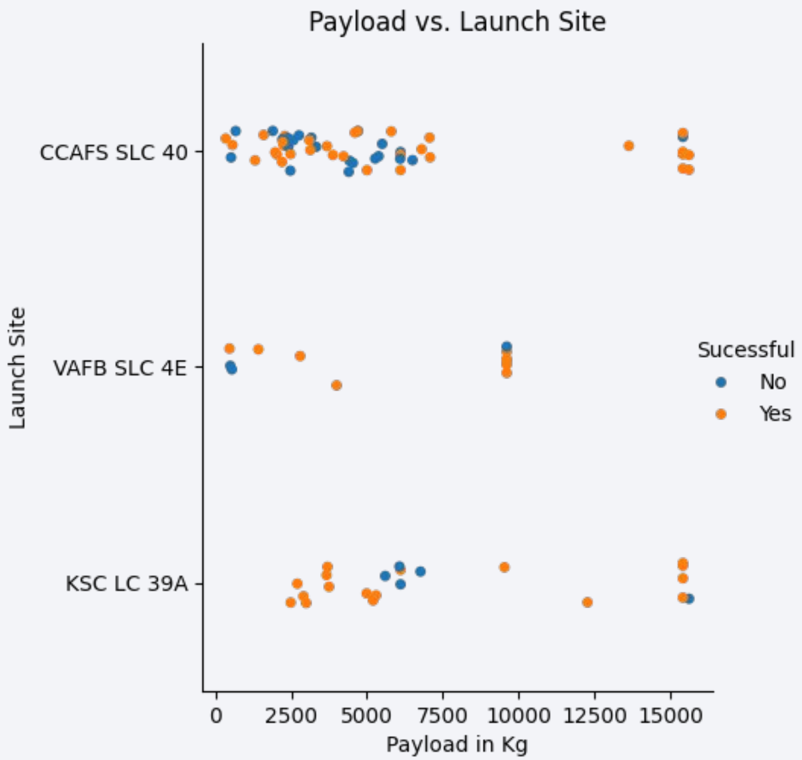
Flight Number vs. Launch Site

- There is a gap of launches around number 20 and 40
- Launch site KSC LC 39A has a high ratio of successful launches



Payload vs. Launch Site

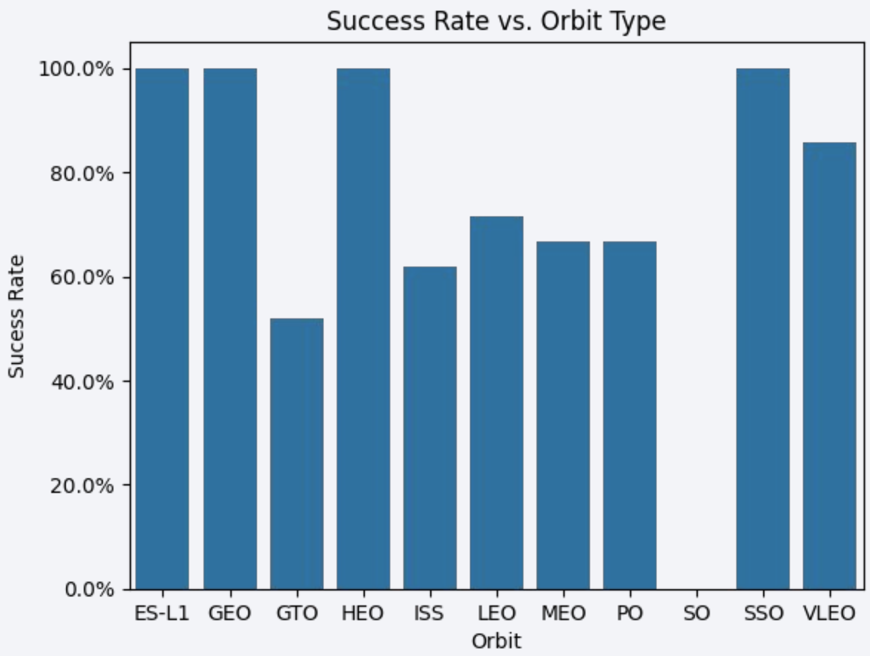
- VAFB SLC 4E was not used with high payloads
- CCAFS SLC 40 was highly used for low payloads



Success Rate vs. Orbit Type

- VLEO and SSO have the highest rate of success, the other 100% orbits lack samples

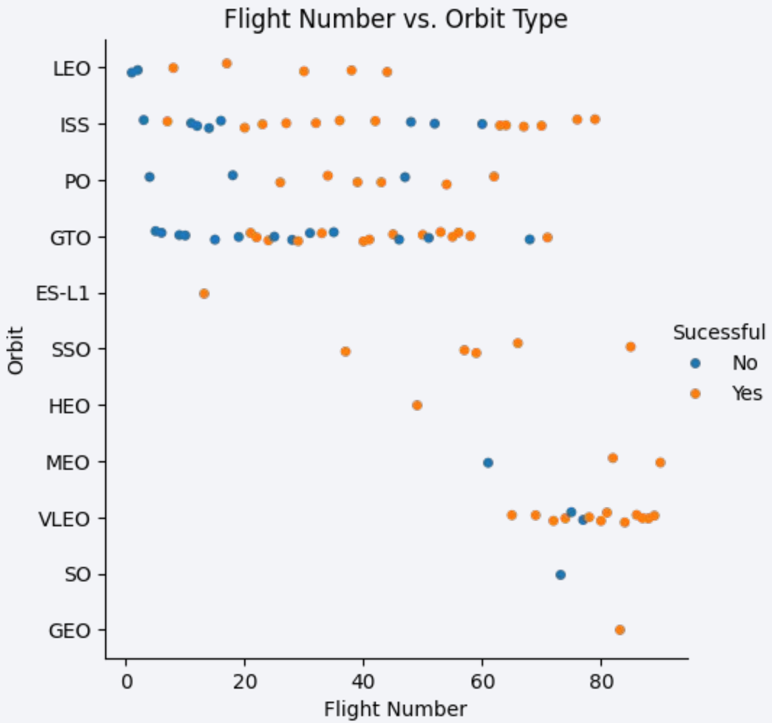
- One launch was to SO orbit and it failed.



Flight Number vs. Orbit Type

- Latter flights were done in VLEO orbit with a high success rate

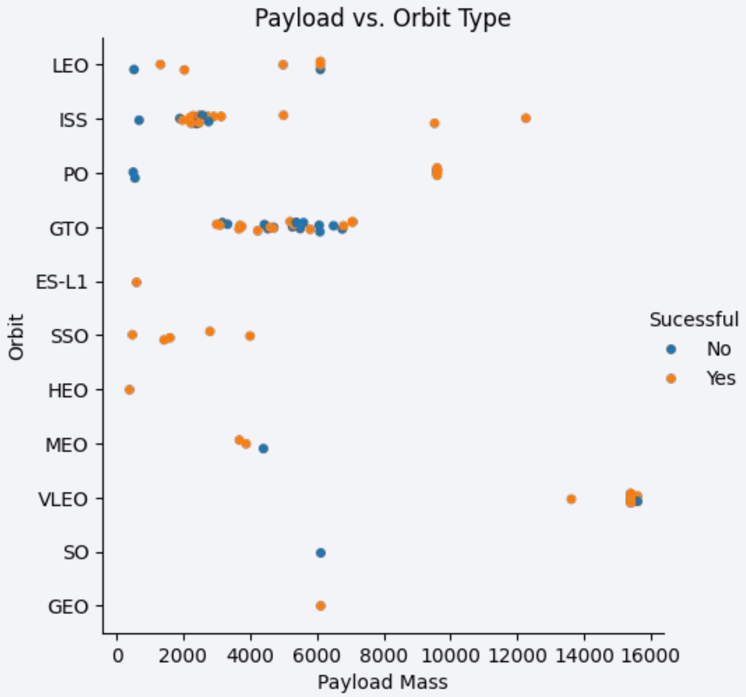
- Earlier flights were done in LEO orbit



Payload vs. Orbit Type

- With heavy payloads the successful landing rate is more for Polar,LEO and ISS.

- GTO we cannot distinguish this well as both positive landing rate and negative



Launch Success Yearly Trend

- The success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

All Launch Site Names

```
%sql select DISTINCT(Launch_Site)
from spacextable
```

- There are 4 unique launch sites

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

Launch Site Names Begin with 'CCA'

```
%sql select * from spacextable
where Launch_Site like 'CCA%'
limit 5
```

- These are early launches with NASA as clients with failed landings

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)

Total Payload Mass

- Over 45000 kg has been launched to space with one payload per launch for NASA (CRS) customer
- This NASA customer requires high payloads

```
%sql select sum(payload_mass__kg_)
as Total_payload_mass__kg from
spacextable where customer = "NASA
(CRS)"
```

Total_payload_mass__kg
45596

Average Payload Mass by F9 v1.1

- On average a payload with F9 booster v1.1 is 2928.4 Kg
- This booster deals with low payloads on average

```
%%sql
    select avg(payload_mass__kg_) as mean_payload_mass_kg
    from spacextable where booster_version like 'F9 v1.1'
```

Mean_payload_mass_kg
2928.4

First Successful Ground Landing Date

- SpaceX started launching in 2010
- It took 5 years to get a successful landing

```
%sql select min(date) as Date from
spacextable where Landing_Outcome =
'Success (ground pad)'
```

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql select Booster_Version from spacextable
        where
        payload_mass__kg_ > 4000 and
        payload_mass__kg_ < 6000 and
        Landing_Outcome = 'Success (drone ship)'
```

- There are 4 versions of boosters with average payloads

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

Total Number of Successful and Failure Mission Outcomes

```
%sql select Mission_Outcome, count(*)
as frequency from spacextable group
by Mission_Outcome
```

- Almost all missions have been successful flying payloads to their destination

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

Boosters Carried Maximum Payload

- F9 B5 B series are the ones used to carry the maximum heaviest payloads.

```
%%sql
    select Booster_Version from spacextable
        where payload_mass__kg_ = (
            select max(payload_mass__kg_)
from spacextable
        )
```

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 B1060.2	
F9 B5 B1058.3	
F9 B5 B1051.6	
F9 B5 B1060.3	
F9 B5 B1049.7	

2015 Launch Records

- The year 2015 the successful landings where F9 v1.1 B series with launching site CCAFS LC-40

month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
02	F9 v1.1 B1013	CCAFS LC-40
03	F9 v1.1 B1014	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40
04	F9 v1.1 B1016	CCAFS LC-40
06	F9 v1.1 B1018	CCAFS LC-40
12	F9 FT B1019	CCAFS LC-40

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

- The 2 landings with parachutes failed
- And using a drone ship for landing has the same odds of failing and succeeding
- Ocean and ground pads all are successful

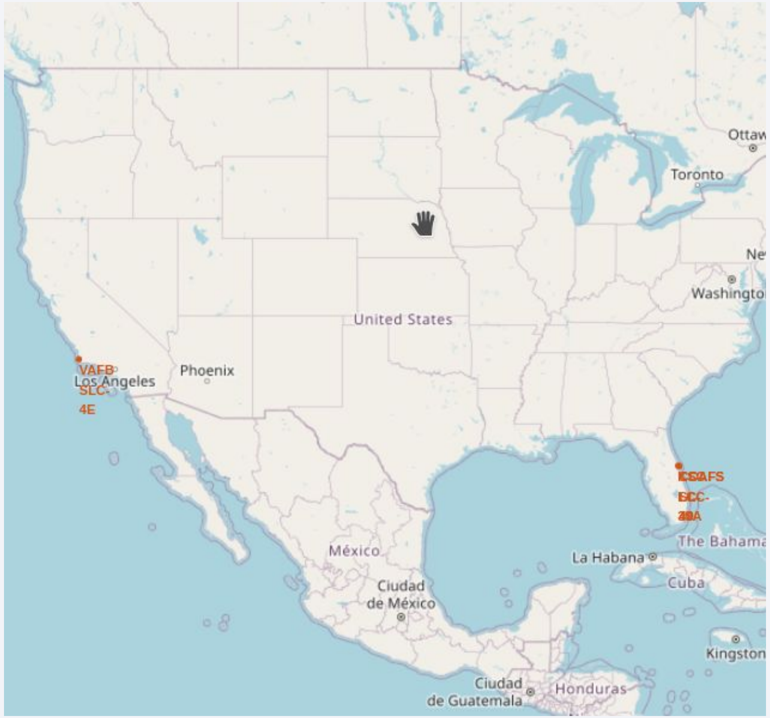
Landing_Outcome	frequency
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section
3

Launch Sites Proximities Analysis

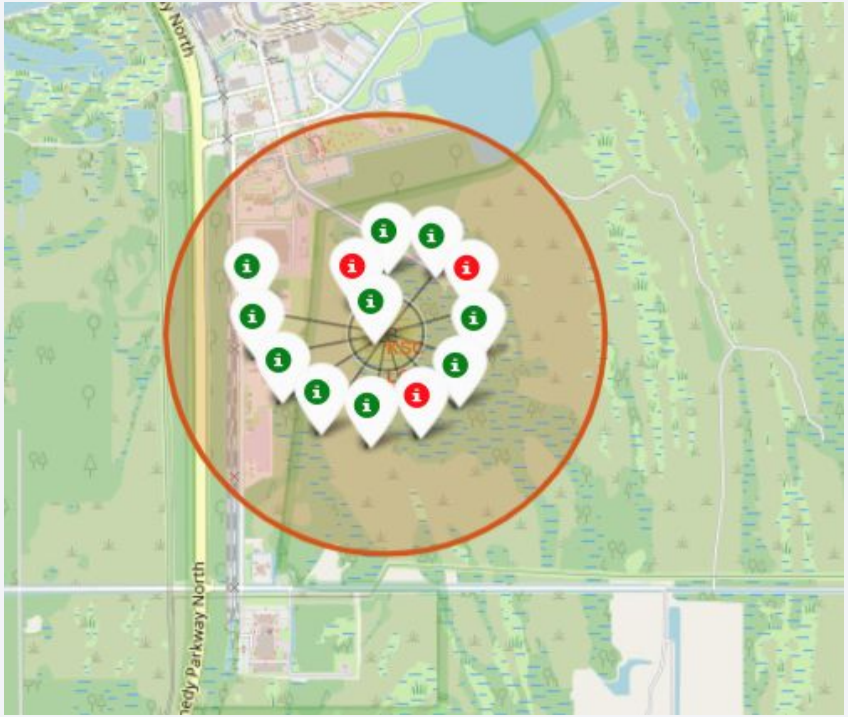
Launching sites SpaceX

- 1 Launching site is located in California.
- 3 are crumpled up in Florida at the other side of the continent.
- Very close to the equator and near an ocean.



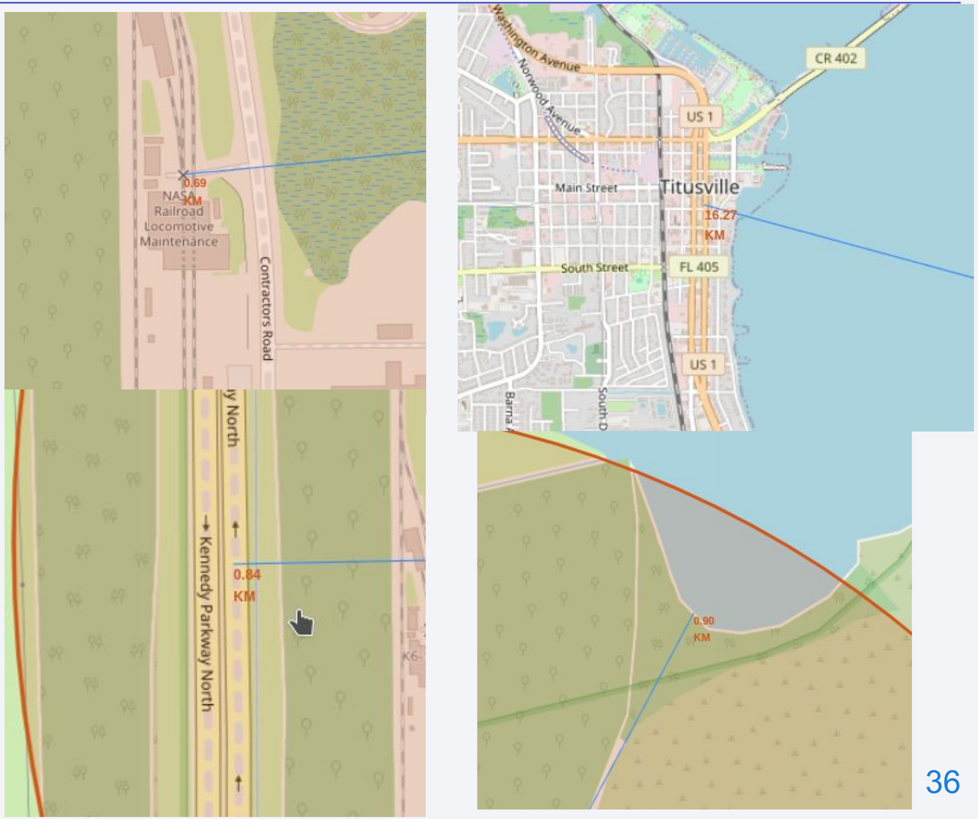
Success/Failure of KSC LC-39A

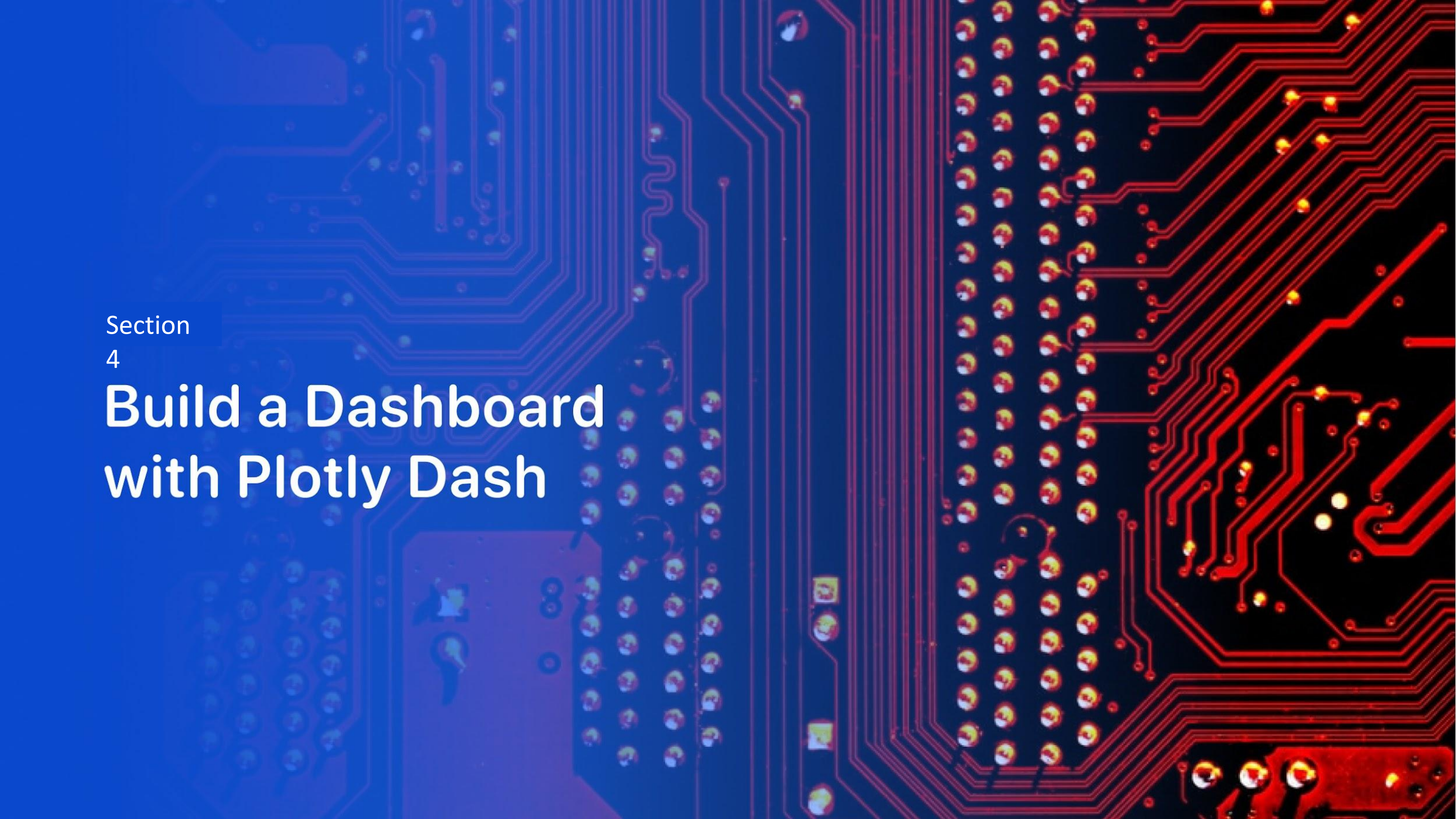
- The site is inside a natural park, also has a high success rate compared to the other sites.



Landmarks of KSC LC-39A

- This launch site has pretty close access to railways and highways, less than 1Km
- The closest body of water is less than 1Km
- The closest city is far away, around 14Km



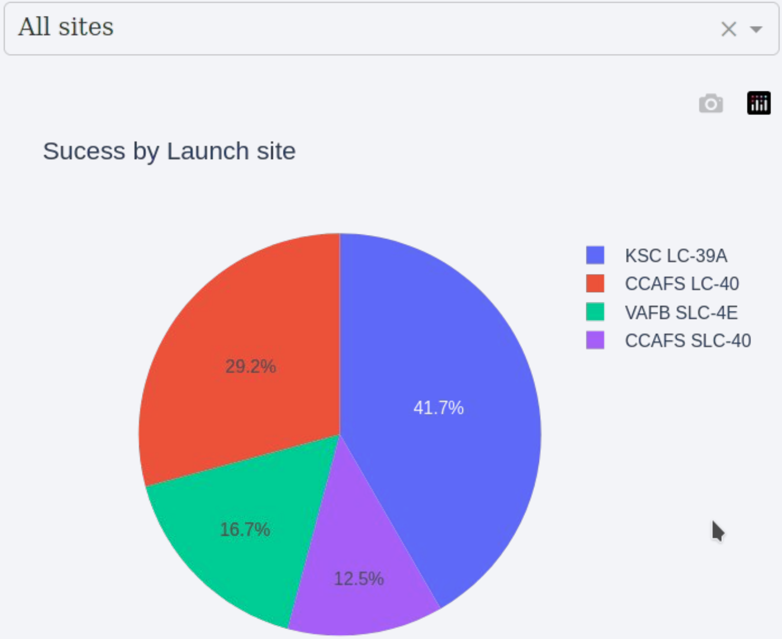


Section
4

Build a Dashboard with Plotly Dash

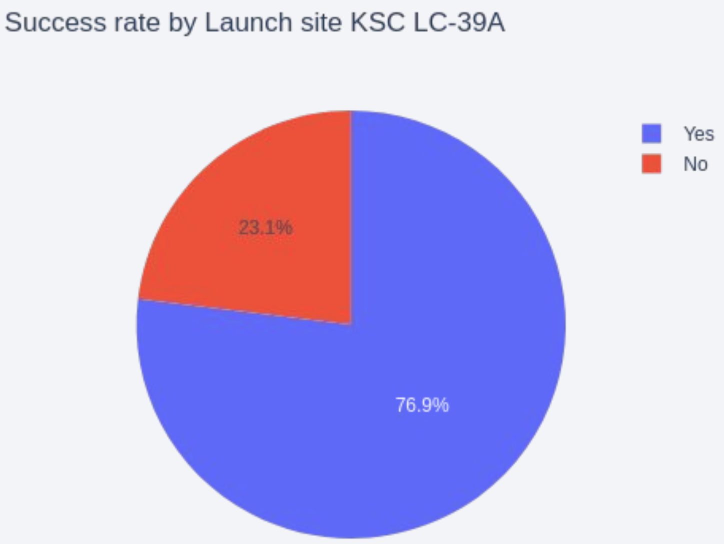
Success Landing by Launch Site

- KSC LC-39A Launch site has the most success rate, 41.7%
- CCAFS SLC-40 has the lowest success rate, 12.5%



Success Landing of KSC LC-39A

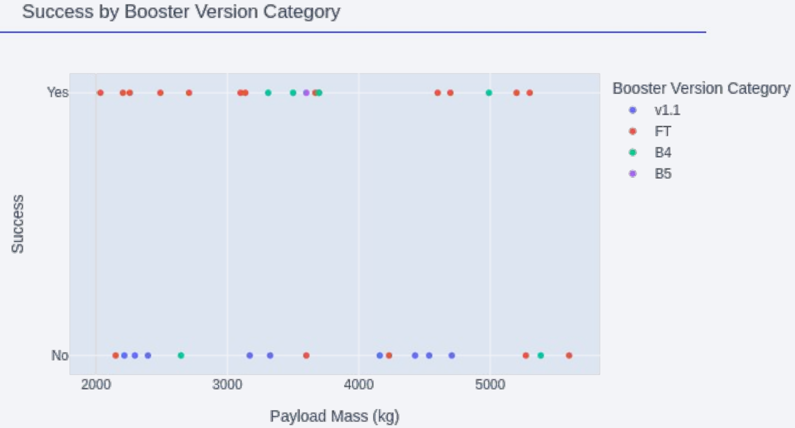
- Almost 3 of 4 landings have been successful at KSC LC-39A

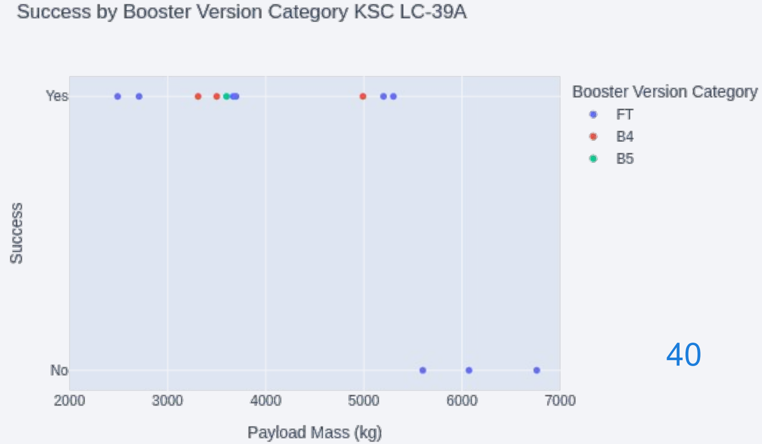


Success by Booster version

- The booster with the highest rate of success is B4

- For launch site KSC LC-39A the FT booster is less successful with payloads above 5500kg



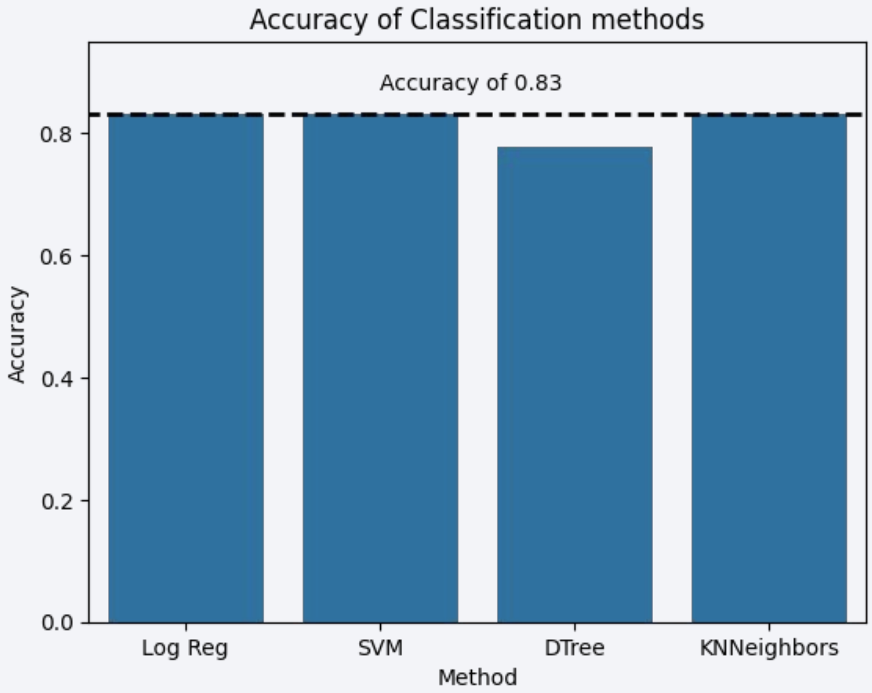


Section
5

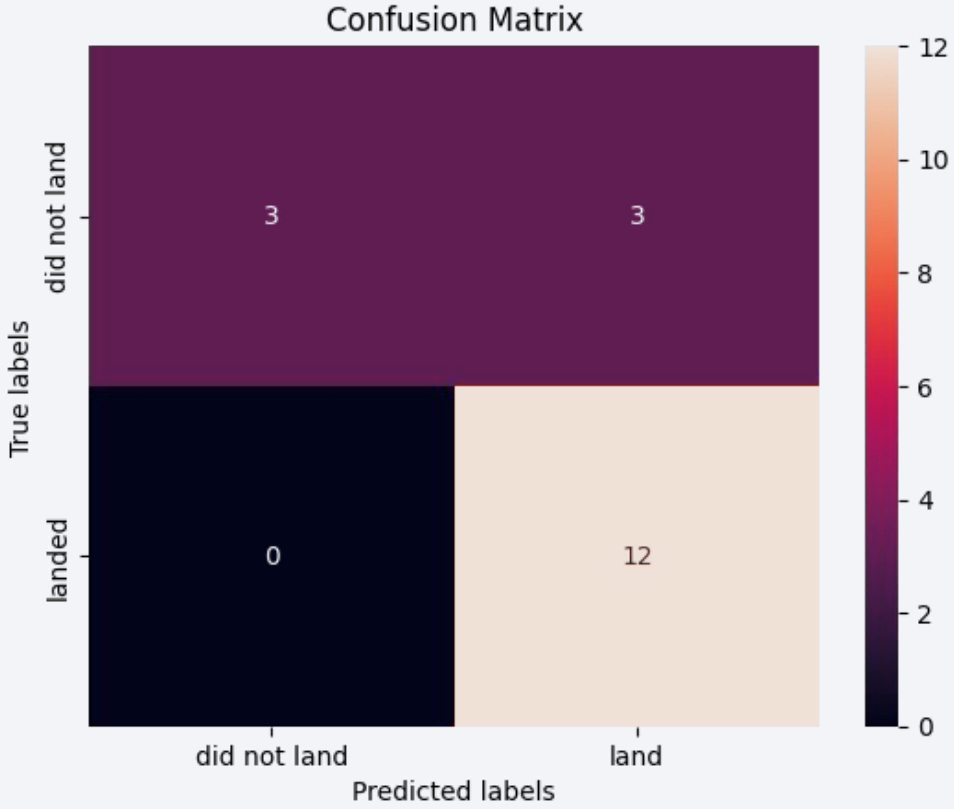
Predictive Analysis (Classification)

Classification Accuracy for Landing success

- Only the decision tree didn't reach the max accuracy of 0.83
- Train/Test split for accuracy
- Logistic regression was chosen because it also provides probability predictions



Confusion Matrix of Logistic Regression



- The model has a precision of 100% for landings

- 50% for unsuccessful landings, model is not good for predicting a failed landing.

Conclusions

- The best landing site is KSC LC-39A located in Florida, USA with highways and railways less than 1km away
- The best booster type is B4 in the above location
- With available data, the prediction model is best suited to predict successful landings but not failed landings
- The best orbit to target is VLEO, though others have higher success rate, there's not enough data

Thank you!