Winning Space Race with Data Science

Aswini Dileep
08/12/2025

Outline

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

Executive Summary

- Summary of methodologies
  - This project follows these steps:
    - Data Collection
    - Data Wrangling
    - Exploratory Data Analysis
    - Interactive Visual Analysis
    - Predictive Analysis (Classification)
- Summary of all results
  - This project produced the following outputs and visualizations:
    - EDA results
    - Interactive dashboard
    - Predictive analysis of classification models

Introduction

- Project background and context:
  - SpaceX launches Falcon 9 rockets at a cost of around $62m. This is considerably cheaper than other providers (which usually cost upwards of $165m), and much of the savings are because SpaceX can land, and then re-use the first stage of the rocket.

- Problems you want to find answers:
  - What factors influence the successful landing of the Falcon 9 first stage?
  - How can we accurately predict the landing outcome using machine learning models?
  - Which machine learning model performs best in predicting the landing success?

Section 1

Methodology

Methodology

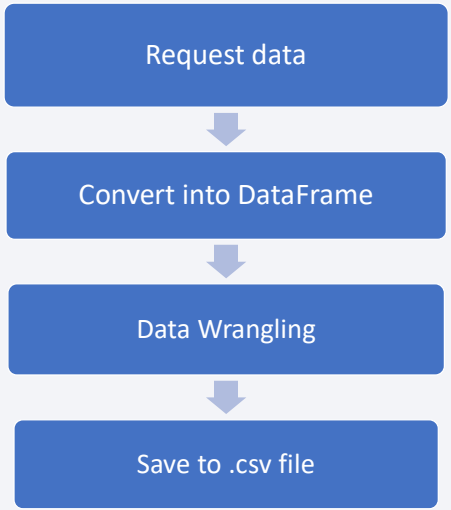
- Data collection methodology:
  - Require the data from SpaceX API
  - Collect data from a Wikipedia page
- Perform data wrangling
  - Perform EDA to find some patterns
  - Determine what would be the label for training supervised model
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Compare logistic regression model, SVM, decision tree classifier, kNN by using GridSearchCV to select the best fit model

Data Collection

- Step 1: SpaceX API Request
  - Initiate API Request
  - Fetch Launch Data
  - Store Data Locally
- Step 2: Web Scraping Wikipedia
  - Extract HTML Table
  - Parse with BeautifulSoup
  - Convert to DataFrame
- Step 3 : Data Integration
  - SpaceX API data and Wikipedia data
  - Merge Datasets
  - Final Integrated Data

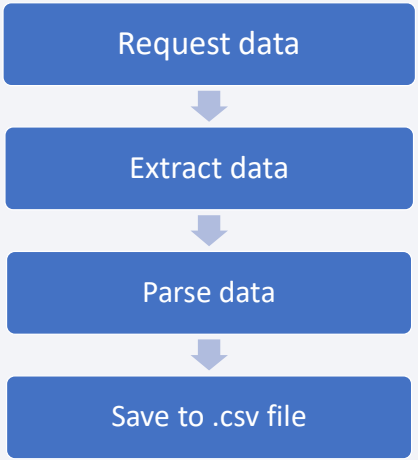
Data Collection – SpaceX API

- GitHub URL
  - https://github.com/Aswini-Dileep/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



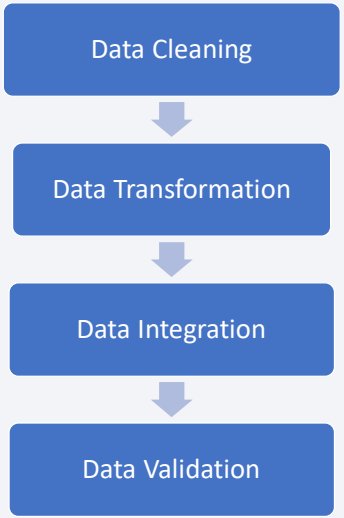
Data Collection - Scraping

- GitHub URL
  - https://github.com/Aswini-Dileep/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- GitHub URL
  - https://github.com/Aswini-Dileep/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Scatter plot - Best to describe the relation between two categorical data.
  - The relationship between Flight number and Launch Site.
  - The relationship between Payload and Launch Site.
  - The relationship between Payload and Orbit type.
  - The relationship between Payload and Orbit type.
- Bar Plot – Best to compare several categorical data
  - The relationship between Success rate of each Orbit type
- Line Chart – Best to show the time series data
  - The launch success yearly trend
- GitHub URL
  - https://github.com/Aswini-Dileep/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/edadataviz.ipynb

EDA with SQL

1. Display the names of the unique launch sites in the space mission
2. Display 5 records where launch sites begin with the string 'CCA'
3. Display the total payload mass carried by boosters launched by NASA (CRS)
4. Display average payload mass carried by booster version F9 v1.1
5. List the date when the first successful landing outcome in ground pad was achieved.
6. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. List the total number of successful and failure mission outcomes
8. List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.
9. List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
10. 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.

https://github.com/Aswini-Dileep/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Mark all launch sites on the map
- Mark the success/failed launches for each site on map
- Calculate the distance between a launch site to its proximities
- To find some geographical patterns about launch sites

- GitHub URL
  - https://github.com/Aswini-Dileep/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/lab_jupyter_launch_site_location.ipynb

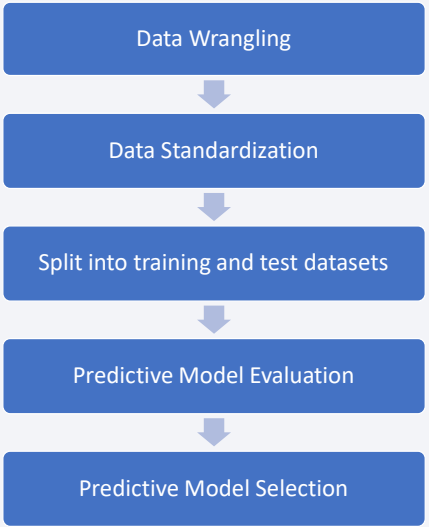
Build a Dashboard with Plotly Dash

- A launch site dropdown input component
- A success pie-chart based on the selected site dropdown
- A range slicer to select payload
- A success payload scatter chart scatter plot based on the selected site dropdown
- To inspect the relationship of success rate between launch site and payload

Predictive Analysis (Classification)

- Logistic Regression Model
- Support Vector Machine Model
- Decision Tree Classifier
- k-Nearest Neighbor Model

- GitHub URL
  - https://github.com/Aswini-Dileep/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/7.%20SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

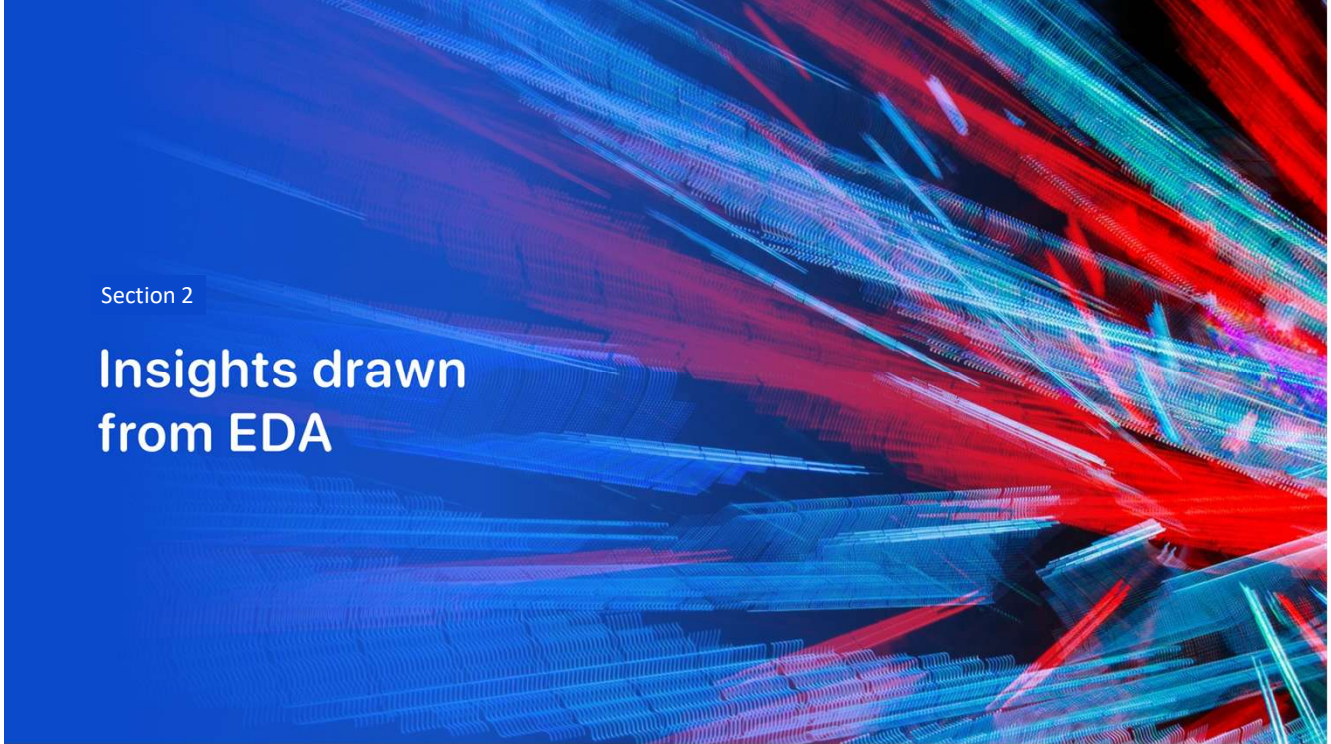


Results

Exploratory Data Analysis

Interactive Analytics

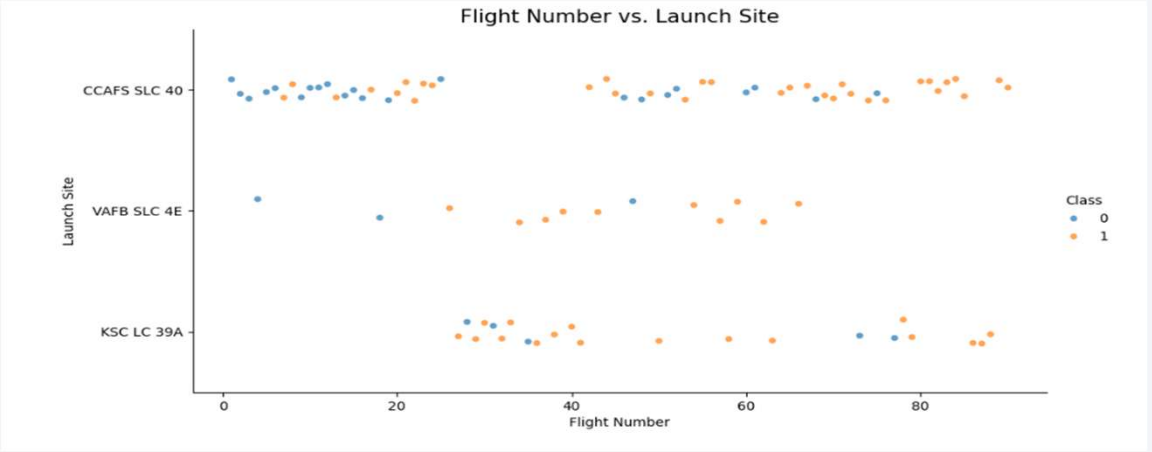
Predictive Analysis



Section 2

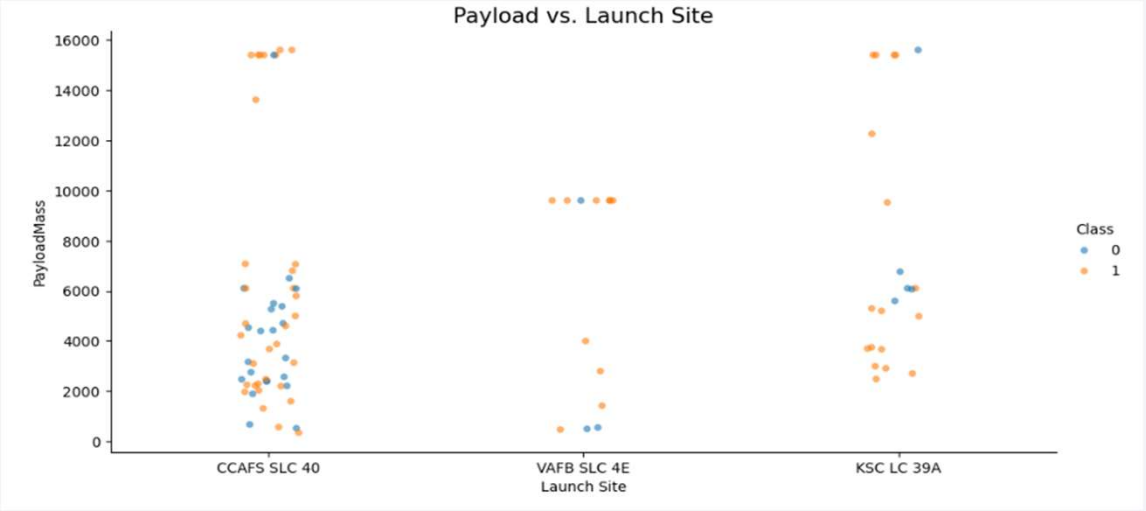
Insights drawn from EDA

Flight Number vs. Launch Site



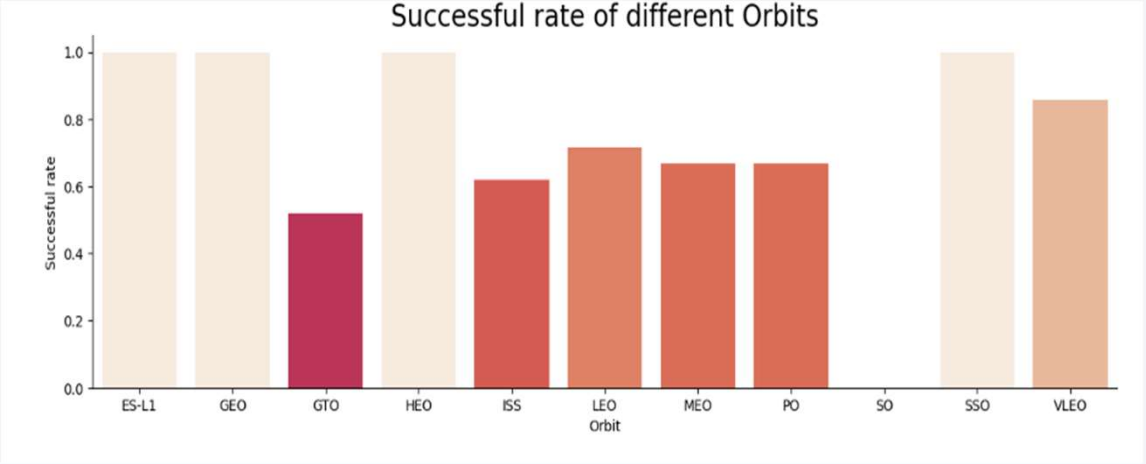
KSC LC 39A has the highest success rate and CCAFS SLC 40 has the lowest.

Payload vs. Launch Site



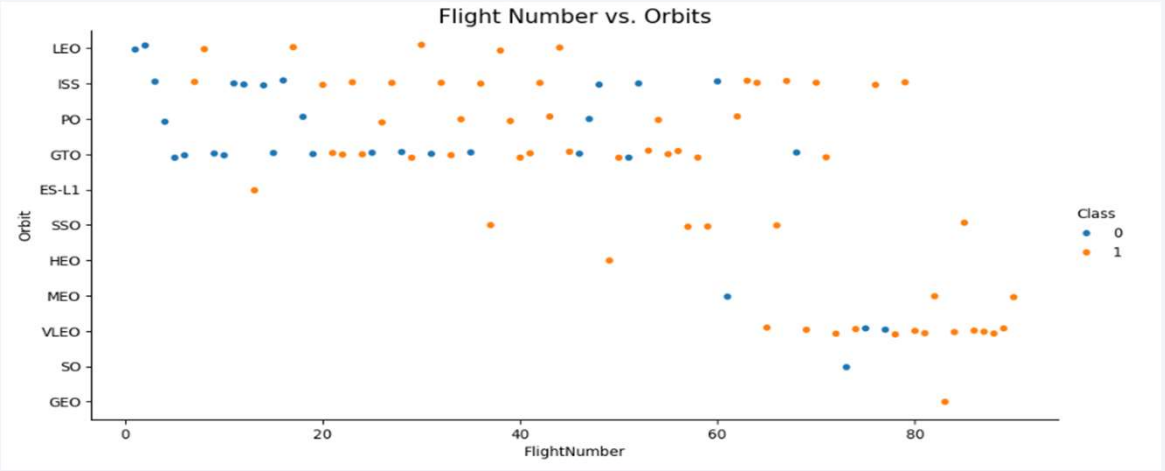
- VAFB SLC 4E has no payload above 10000 kg

Success Rate vs. Orbit Type



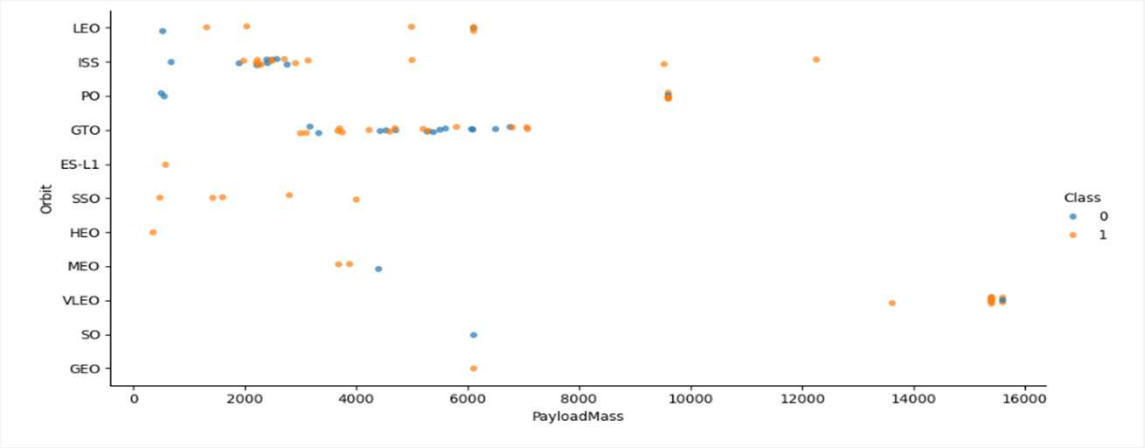
- ES-L1, GEO, HEO and SSO have the highest success rates

Flight Number vs. Orbit Type



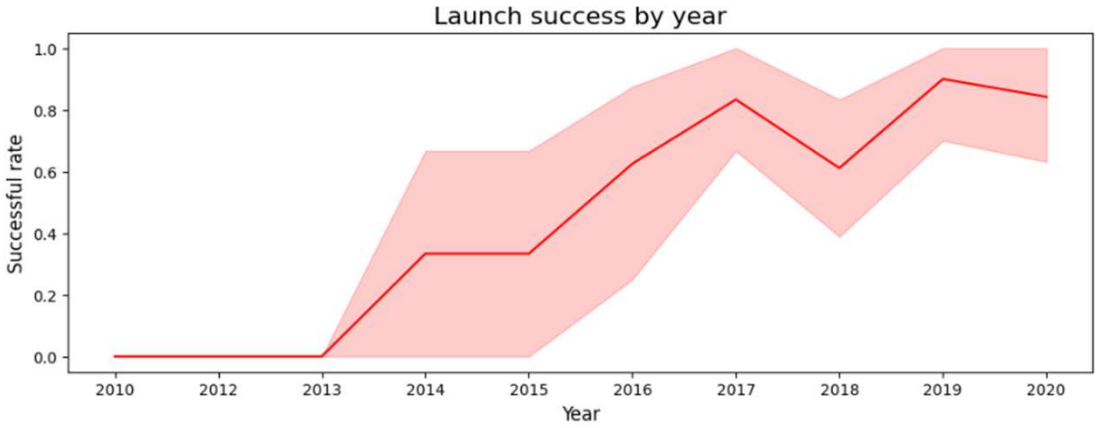
- In the LEO orbit the Success appears related to the number of flights. On the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- With heavy payloads the successful landing or Positive landing rate are more for Polar, LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing (Unsuccessful mission) are both there here.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

```
%sql Select distinct Launch_Site from SPACEXTBL
```

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

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

Launch Site Names Begin with 'CCA'

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

```
 * sqlite:///my_data1.db
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	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
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
%sql select count(Payload) from SPACEXTBL where Customer = 'NASA (CRS)'
```

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

count(Payload)
20

Average Payload Mass by F9 v1.1

```
%sql SELECT avg(PAYLOAD_MASS__KG_) AS Avg_Payload FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1';
```

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

Avg_Payload
2928.4

First Successful Ground Landing Date

```
%sql select min(Date) from SPACEXTBL where Landing_Outcome like 'Success (ground pad)' ;
```

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT DISTINCT Customer, Landing_Outcome,PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE Landing_Outcome ='Success (drone ship)
```

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

Customer	Landing_Outcome	PAYLOAD_MASS__KG_
SKY Perfect JSAT Group	Success (drone ship)	4696
SKY Perfect JSAT Group	Success (drone ship)	4600
SES	Success (drone ship)	5300
SES EchoStar	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

```
%sql select  Mission_Outcome, count(*) from SPACEXTBL group by mission_outcome
```

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

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

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version, Max_Payload FROM (SELECT Booster_Version, MAX(PAYLOAD_MASS__KG_) AS Max_Payload FROM SPACEXTBL
```

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

Booster_Version	Max_Payload
F9 B4 B1039.2	2647
F9 B4 B1040.2	5384
F9 B4 B1041.2	9600
F9 B4 B1043.2	6460
F9 B4 B1039.1	3310
F9 B4 B1040.1	4990
F9 B4 B1041.1	9600
F9 B4 B1042.1	3500
F9 B4 B1043.1	5000
F9 B4 B1044	6092
F9 B4 B1045.1	362
F9 B4 B1045.2	2697

2015 Launch Records

```
%sql SELECT SUBSTR(Date,6,2) AS Month, Booster_Version, Launch_site FROM SPACEXTBL WHERE Landing_Outcome LIKE 'Failure%dron
```

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

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

```
%sql SELECT Landing_Outcome, COUNT(*) AS Numbers FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)' or Landing_Ou
```

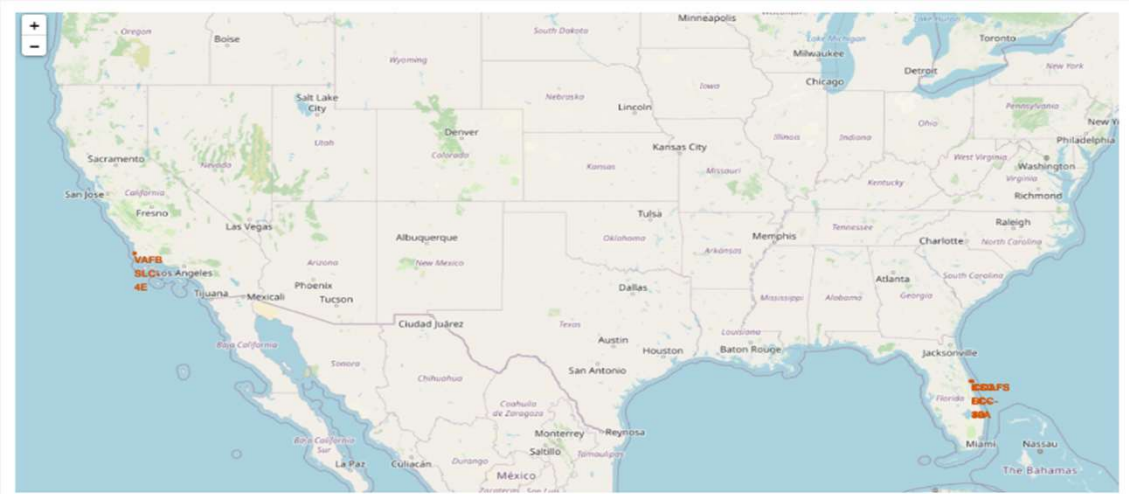
```
* sqlite:///my_data1.db
Done.
```

Landing_Outcome	Numbers
Success (ground pad)	9

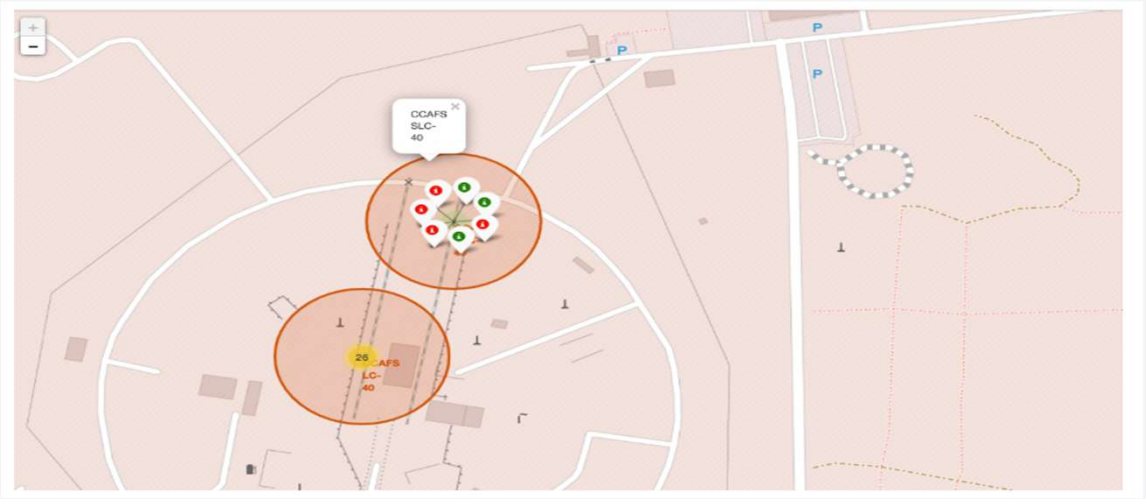
Section 3

Launch Sites Proximities Analysis

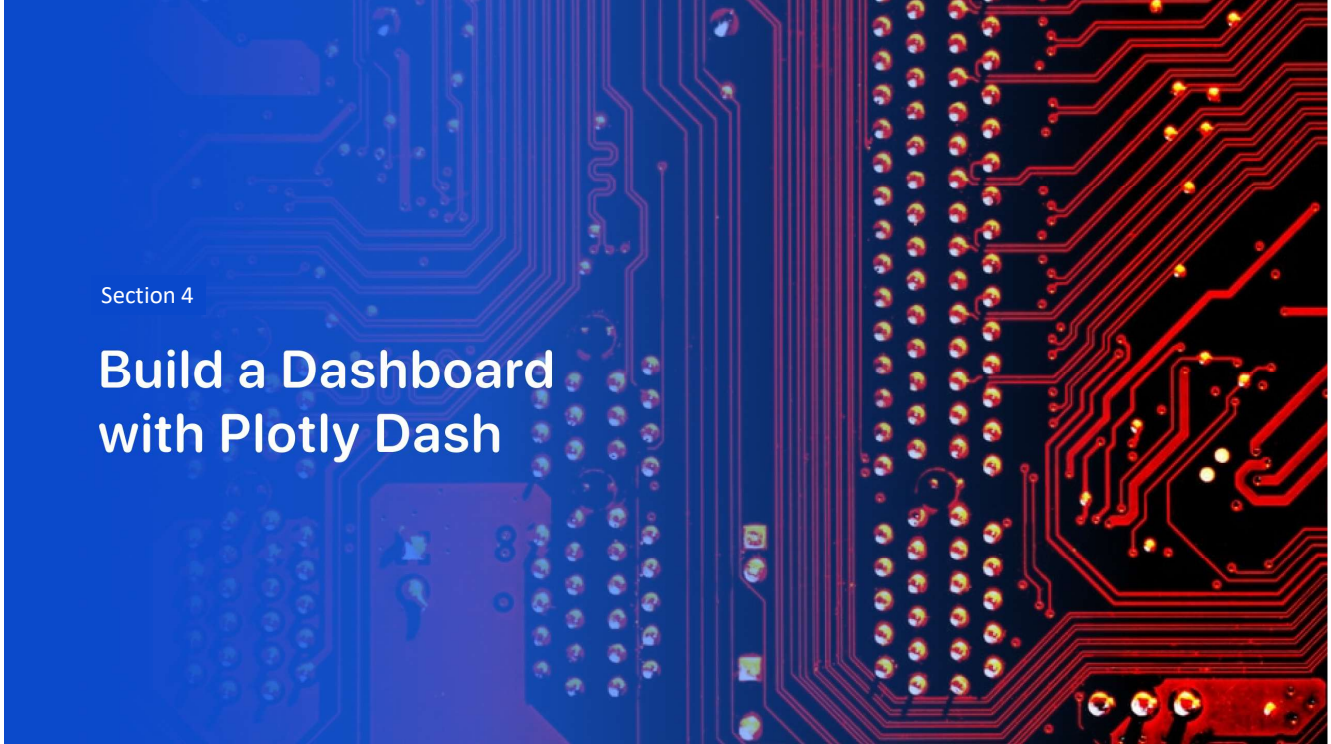
All Launch Sites



Success/Failed Launches For Each Site On The Map



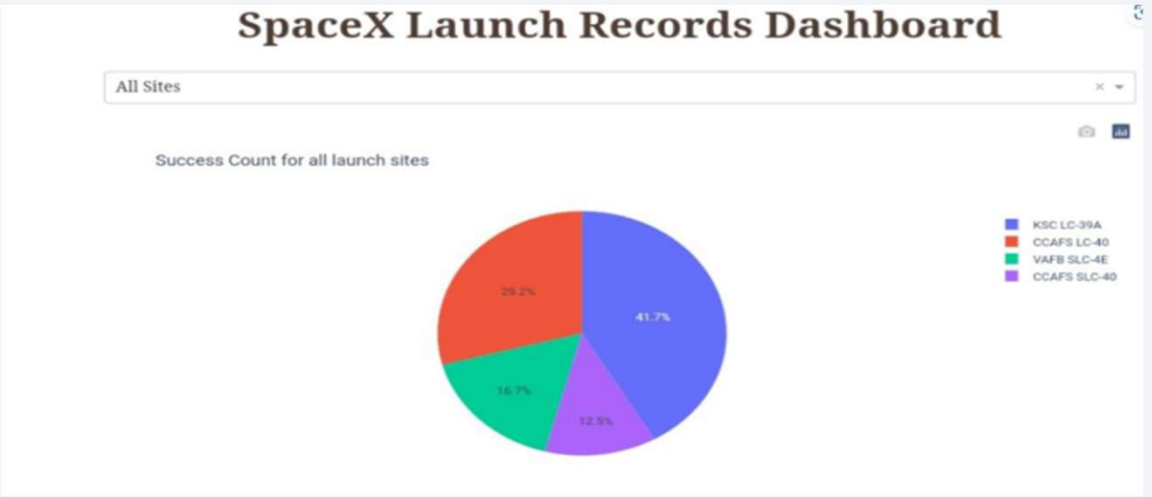
Distances Between a Launch Site to its Proximities



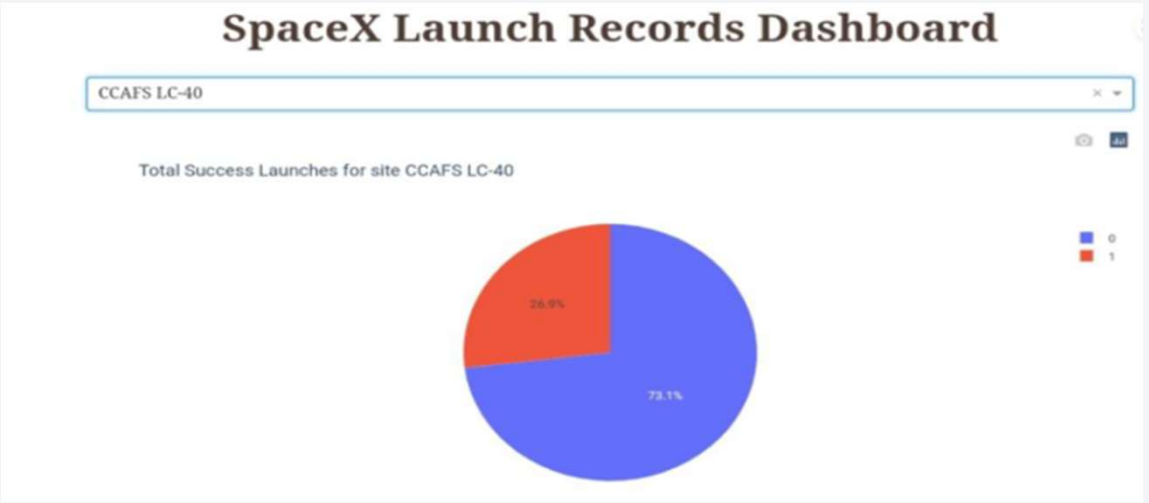
Section 4

Build a Dashboard with Plotly Dash

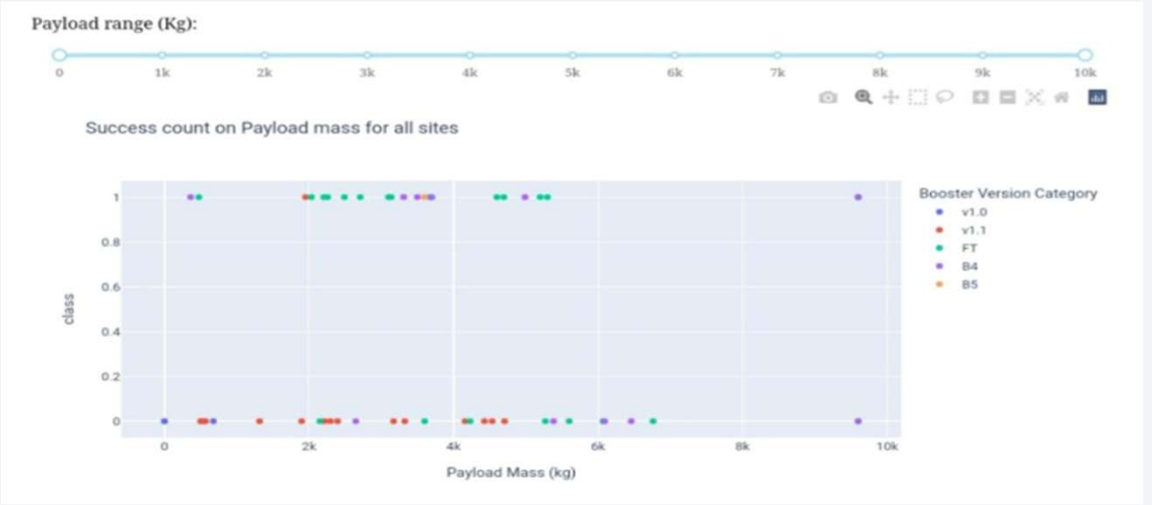
All Launch Sites



Highest Success Launch Ratio



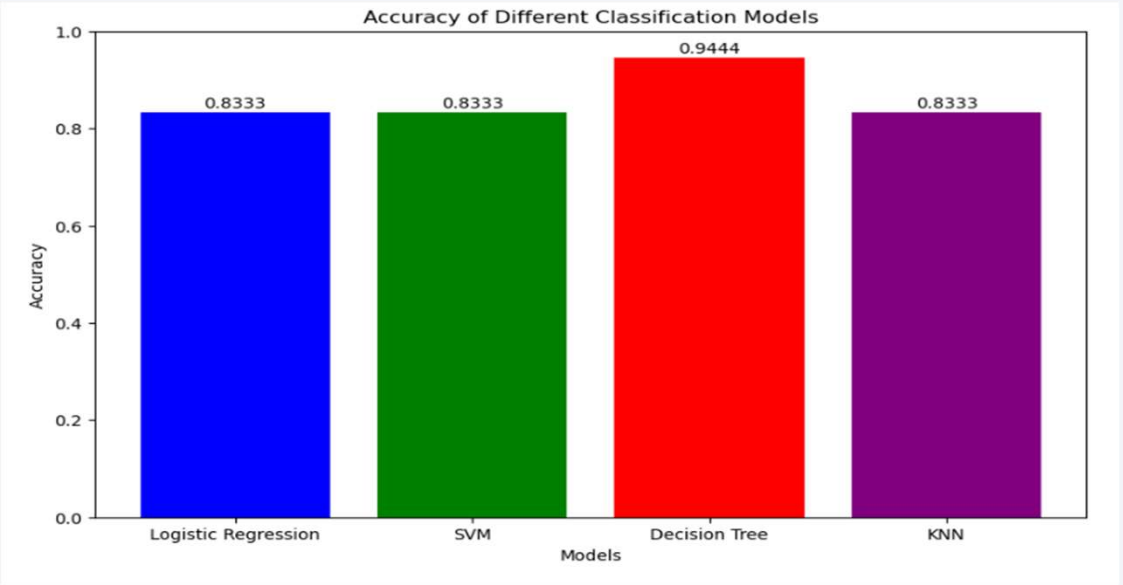
Payload Vs Launch Outcome



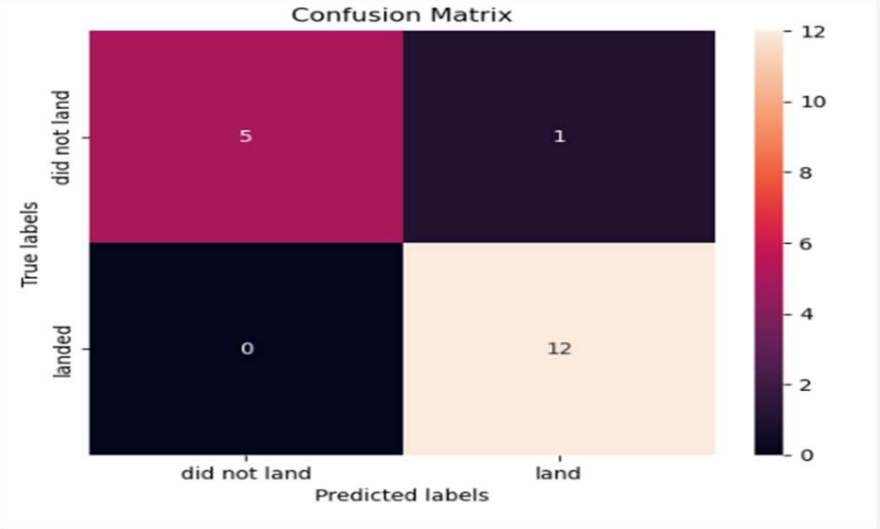
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- There is a correlation between launch site and success rate Payload mass, which is also associated with the success rate. The more massive the payload the less likely the first stage will return.
- For orbit, SO has the least success rate while ES-L1, GEO, HEO, and SSO have the highest success rate according to the yearly trend.
- There has been an increase in the success rate since 2013 kept increasing till 2020
- With best parameter provided, decision tree classifier used in prediction yielded the highest accuracy of 89%

Appendix

- https://github.com/Aswini-Dileep/SpaceX-Falcon-9-first-stage-Landing-Prediction

Thank you!