Winning Space Race with Data Science

Emad Mohammed Habibi
15/11/2024

Outline

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

Executive Summary

- **Summary of methodologies**

- Data was gathered from the publicly available SpaceX API and the SpaceX Wikipedia page. A new column called "Class" was created to categorize successful landings. Various exploratory techniques were employed, and only the relevant columns were selected as features for further analysis. The analysis showed important relationships between variables and identified the location of the launch sites. Categorical variables were converted into numerical format to prepare the data for modeling. Subsequently, the data was standardized, and GridSearchCV was utilized to identify the optimal parameters for the machine learning models. The accuracy scores of all models were visualized to gauge their performance.
- In this capstone project, we will predict if the SpaceX Falcon 9 first stage will land successfully using several machine learning classification algorithms.
- The main steps in this project include:

    - Data collection, wrangling, and formatting
    - Exploratory data analysis
    - Interactive data visualization
    - Machine learning prediction

Executive Summary

- **Summary of all results**

- The analysis yielded four machine learning models: Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K Nearest Neighbors. Interestingly, all models exhibited a similar accuracy rate of approximately 83.3%. However, it was noted that all models tended to overpredict successful landings. This suggests a potential need for additional data to refine the models and improve their predictive accuracy. Further data collection efforts could enhance the robustness and reliability of the machine learning models in predicting SpaceX's first stage reuse outcomes.

- Our graphs show that some features of the rocket launches have a correlation with the outcome of the launches, i.e., success or failure.

- It is also concluded that decision tree may be the best machine learning algorithm to predict if the Falcon 9 first stage will land successfully.

Introduction

- **Project background and context**

- In this capstone, we will 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.
- Most unsuccessful landings are planned. Sometimes, SpaceX will perform a controlled landing in the ocean.
- This is because of the Space X can reuse the first stage.
- A new company Space Y wants to join in the business, and the only way they have to do this is to predict which rockets are going to land safely and be reuse in the first stage.

- **Problems you want to find answers**

- Determine if the first stage of SpaceX Falcon 9 will land successfully
- Impact of different parameters/variables on the landing outcomes (e.g., launch site, payload mass, booster version, etc.)
- Correlations between launch sites and success rates

Section 1

Methodology

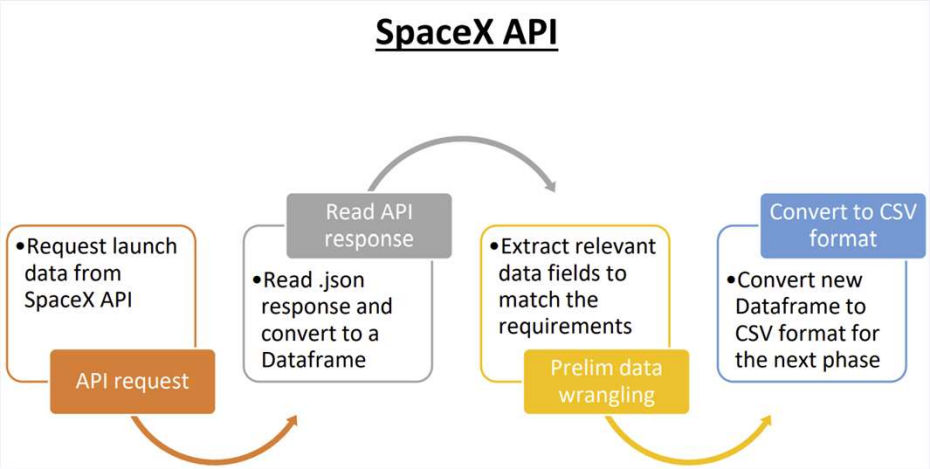
Methodology

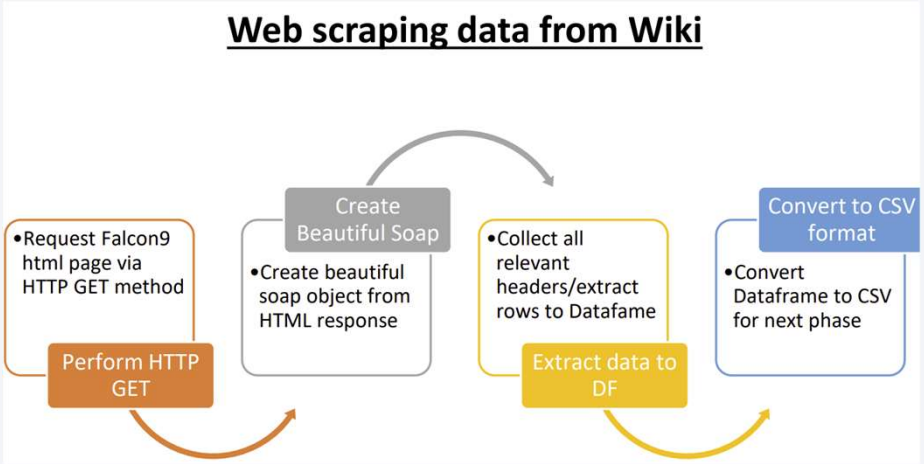
Executive Summary

- Data collection methodology:
  - Use of SpaceX API and SpaceX Wikipedia Page, Web scraping
- Perform data wrangling
  - Adjusting variable types and labeling each lunch as a successful landing or unsuccessful landing
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Pandas and NumPy
  - SQL
- Perform interactive visual analytics using Folium and Plotly Dash
  - Matplotlib and Seaborn
  - Folium
  - Dash
- Perform predictive analysis using classification models
  - Create different Machine Learning Models and Tune them using GridSearchCV
    - Logistic regression, Support vector machine (SVM), Decision tree, K-nearest neighbors (KNN)

Data Collection

- Describe how data sets were collected.
- Data collection is the process of gathering data from available sources. This data can be structured, unstructured, or semi-structured. For this project, data was collected via SpaceX API and Web scrapping Wiki pages for relevant launch data.
- You need to present your data collection process use key phrases and flowcharts





Data Collection – SpaceX API

- **1. API Request and read responseinto DF**
  - Create API GET request, normalize data and read in toa Dataframe:

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

```
# Use json_normalize method to convert the json result into a dataframe
data = pd.json_normalize(resp)
```

- **2. Declare global variables**
  - lists that will store data returned by helper functions with additional API calls to get relevant data

```
#Global variables
BoosterVersion = []
PayloadMass = []
Orbit = []
LaunchSite = []
Outcome = []
Flights = []
GridFins = []
Reused = []
Legs = []
LandingPad = []
Block = []
ReusedCount = []
Serial = []
Longitude = []
Latitude = []
```

3. Call helper functions with APIcalls to populate global vars

Call helper functions to get relevant data where columns have IDs(e.g., rocket column is an identification number)
getBoosterVersion(data)
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)...

https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/Data%20Collection%20API.ipynb

Data Collection – SpaceX API

- **4. Construct data using dictionary**
- Construct dataset from received data & combine columns into a dictionary:

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

5. Convert Dict to Dataframe, filterfor Falcon9 launches, covert to CSV

Create Dataframe from dictionary and filter to keep only the Falcon9 launches

```
# Create a data from launch_dict
launch_data = pd.DataFrame(launch_dict)
```

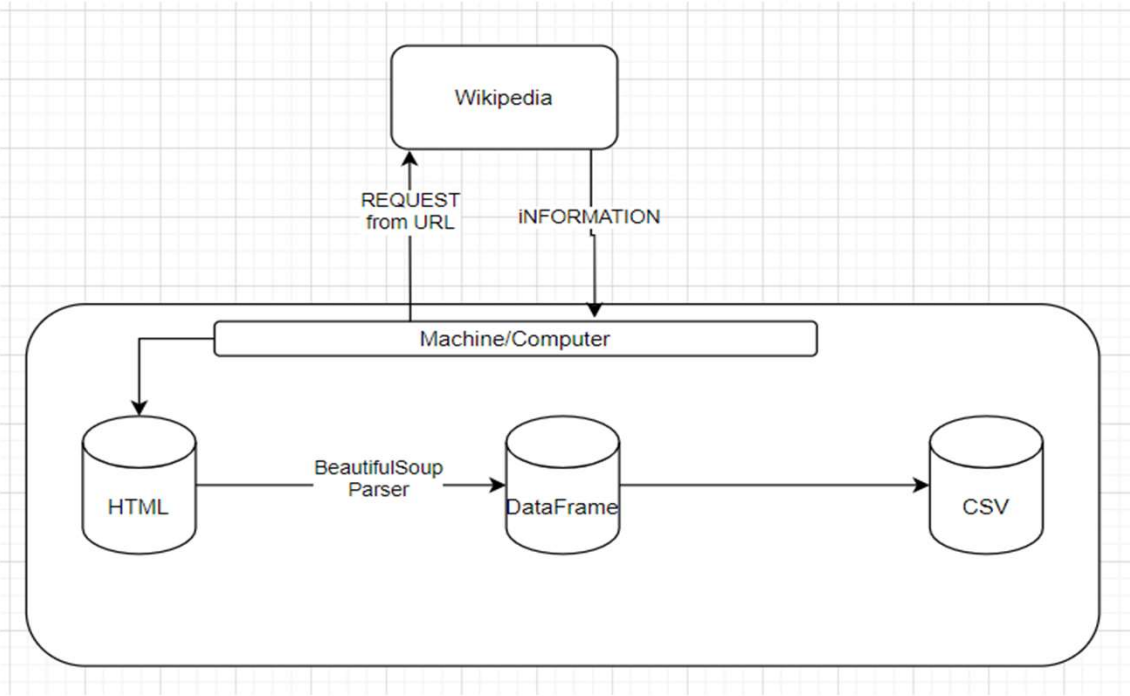
```
# Hint data['BoosterVersion']!='Falcon 1'
data_falcon9 = launch_data[launch_data['BoosterVersion'] != 'Falcon 1']
```

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/Data%20Collection%20API.ipynb

Data Collection - Scraping

- **1. Getting Response from HTML**
- response = requests.get(static_url)
- **2. Create BeautifulSoup Object**
- soup = BeautifulSoup(response.text, "htm15lib")
- **3. Find all tables**
- html tables = soup.findAll('table')



- https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Collection - Scraping

- **4. Get column names**

```
for th in first_launch_table.find_all('th'):
    name = extract_column_from_header(th)
    if name is not None and len(name) > 0 :
        column_names.append(name)
```

- **5. Create dictionary**

```
launch_dict = dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster'] = []
launch_dict['Booster landing'] = []
launch_dict['Date'] = []
launch_dict['Time'] = []
```

- **6. Add data to keys**

```
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all
   # get table row
    for rows in table.find_all("tr"):
        #check to see if first table heading is a
        if rows.th:
            if rows.th.string:
                flight_number=rows.th.string.stri
                flag=flight_number.isdigit()
```

- **7. Create dataframe from dictionary**
- df=pd.DataFrame(launch dict)

- **8. Export to filed**
- f.to_csv('spacex_web_scraped.csv', index=False)

- https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

- In the dataset, there are several cases where the booster did not land successfully..True Ocean, True RTLS, True ASDS means the mission has been successful.. False Ocean, False RTLS, False ASDS means the mission was a failure..
- We need to transform string variables into categorical variables where 1 means themission has been successful and O means the mission was a failure.

- **1. Calculate launches number for each**
- df['LaunchSite'].value_counts()

```
LaunchSite
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: count, dtype: int64
```

- **2. Calculate the number and occurenceof each orbit**
- df['Orbit'].value_counts()

```
Orbit
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
HEO       1
ES-L1     1
SO        1
GEO       1
Name: count, dtype: int64
```

- **3. Calculate number and occurrence of mission outcome per orbit type**
- landing_outcomes = df['Outcome'].value_counts()landing_outcomes

```
Outcome
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: count, dtype: int64
```

- **4 Create landing outcome label from Outcome**

```
landing_class = []
for key,value in df["Outcome"].items():
     if value in bad_outcomes:
        landing_class.append(0)
     else:
        landing_class.append(1)
df['Class']=landing_class
```

- **5 Export to file**

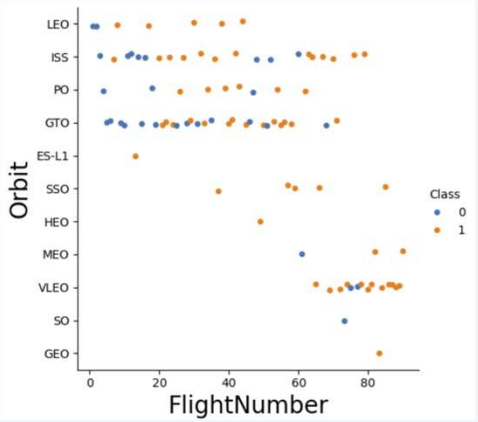
```
df.to_csv("dataset_part_2.csv", index=False)
```

- https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- **Scatter Graphs :**

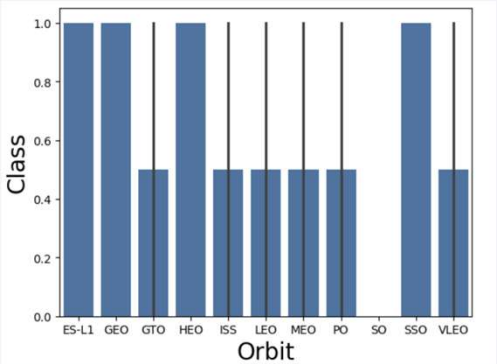
- Flight Number vs. Payload Mass
- Flight Number vs. Launch Site
- Payload vs. Launch Site
- Orbit vs. Flight Number
- Payload vs. Orbit Type
- Orbit vs. Payload Mass



- Scatter plots show the relationship between variables. If a relationship exists, they could be used in machine learning model.

- **Bar Graph**

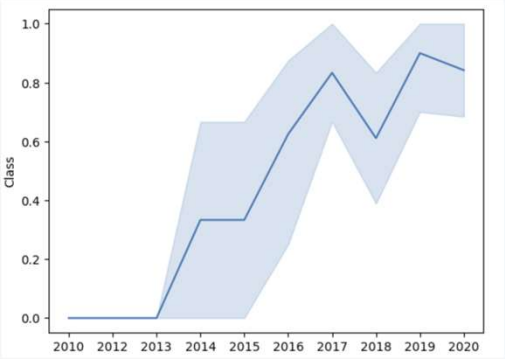
- Success rate vs. Orbit



- Show the relationship between numeric and categoric variables.·

- **Line Graph·**

- Success rate vs. Year



- Line graphs show data variables and their trends.
- Line graphs can help to show global behavior and make prediction for unseen data.

- https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- **Performed SQL queries:**

- 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 the months 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.

- https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- **Folium map object is a map centered on NASA Johnson Space Center at Houson, Texas.**
- Red circle at NASA Johnson Space Center's coordinate with label showing its name (folium.Circle, folium.map.Marker).
- Red circles at each launch site coordinates with label showing launch site name (folium.Circle, folium.map.Marker, folium.features.DivIcon).
- The grouping of points in a cluster to display multiple and different information for the same coordinates (folium.plugins.MarkerCluster).
- Markers to show successful and unsuccessful landings. Green for successful landing and Red for unsuccessful landing. (folium.map.Marker, folium.Icon).
- Markers to show distance between launch site to key locations (railway, highway, coastway, city) and plot a line between them. (folium.map.Marker, folium.PolyLine, folium.features.DivIcon) .
- These objects are created in order to understand better the problem and the data. We can show easily all launch sites, their surroundings and the number of successful and unsuccessful landings
- https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- **Built a Plotly Dash web application to perform interactive visual analytics on SpaceX launch data in real-time. Added Launch Site Drop-down, Pie Chart, Payload range slide, and a Scatter chart to the Dashboard.**

- Added a Launch Site Drop-down Input component to the dashboard to provide an ability to filter Dashboard visual by all launch sites or a particular launch site.
- Added a Pie Chart to the Dashboard to show total success launches when 'All Sites' is selected and show success and failed counts when a particular site is selected.
- Added a Payload range slider to the Dashboard to easily select different payload ranges to identify visual patterns.
- Added a Scatter chart to observe how payload may be correlated with mission outcomes for selected site(s). The color-label Booster version on each scatter point provided missions outcomes with different boosters.

- **Dashboard helped answer following questions:**

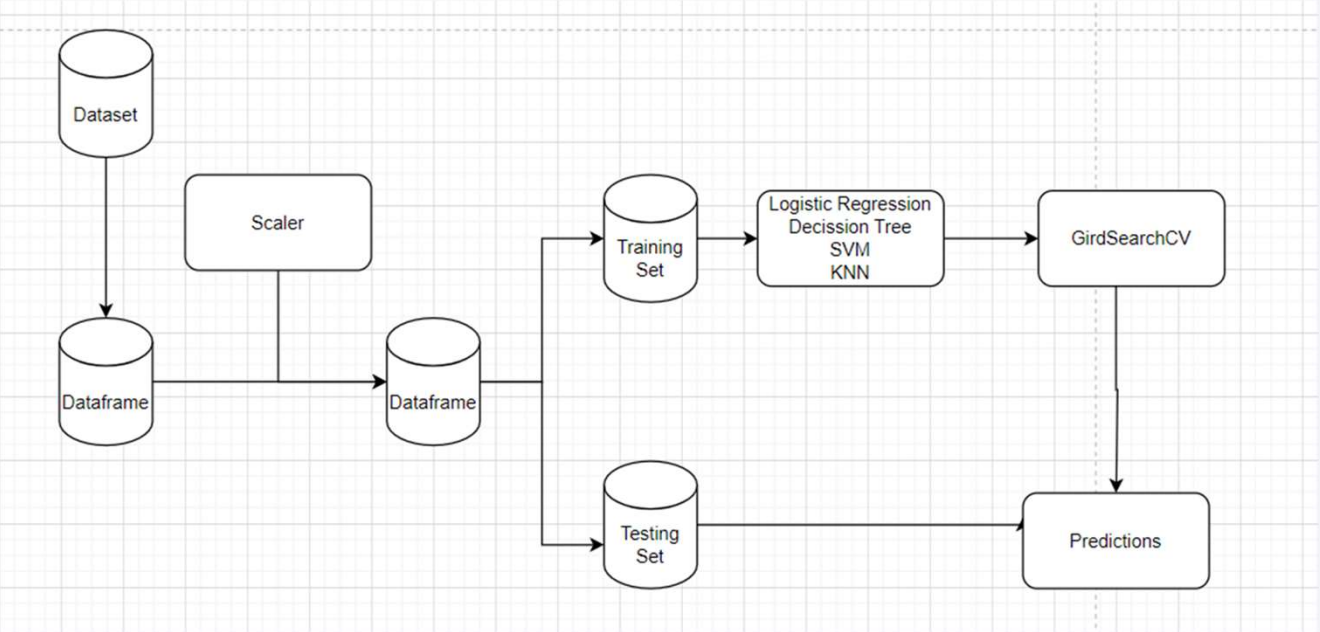
- Which site has the largest successful launches? KSC LC-39A with 10 .
- Which site has the highest launch success rate? KSC LC-39A with 76.9% success.
- Which payload range(s) has the highest launch success rate? 2000 – 5000 kg.
- Which payload range(s) has the lowest launch success rate? 0-2000 and 5500 – 7000.
- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate? FT

- https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- **We prepare the data, then training multiple models, and tunning their parameters using GridSearchCV.**

- Read dataset into Dataframe and create a 'Class' array.
- Standardize the data.
- Train/Test/Split data in to training and test data sets.
- Create and Refine Models.
- Find the best performing
- Finally, we use the best performing model with the test set.

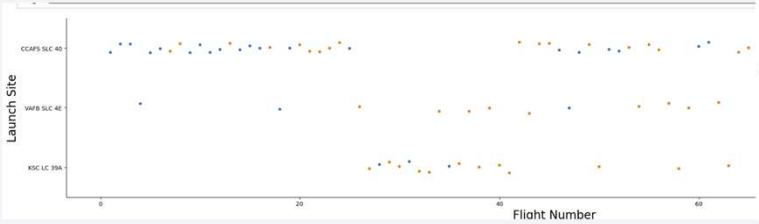


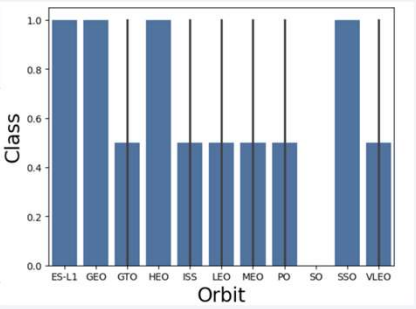
https://github.com/cmenapaz/Applied_Data_Science_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

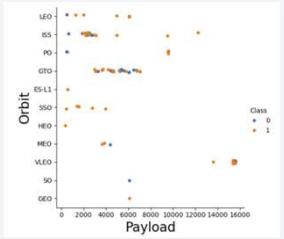
Following sections and slides explain results for:

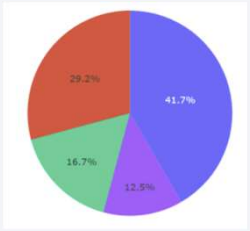
- Exploratory data analysis results
  - Samples



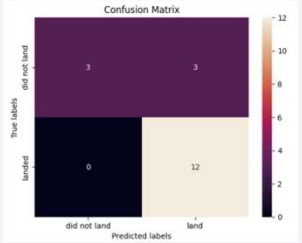


- Interactive analytics demo in screenshots
  - Samples

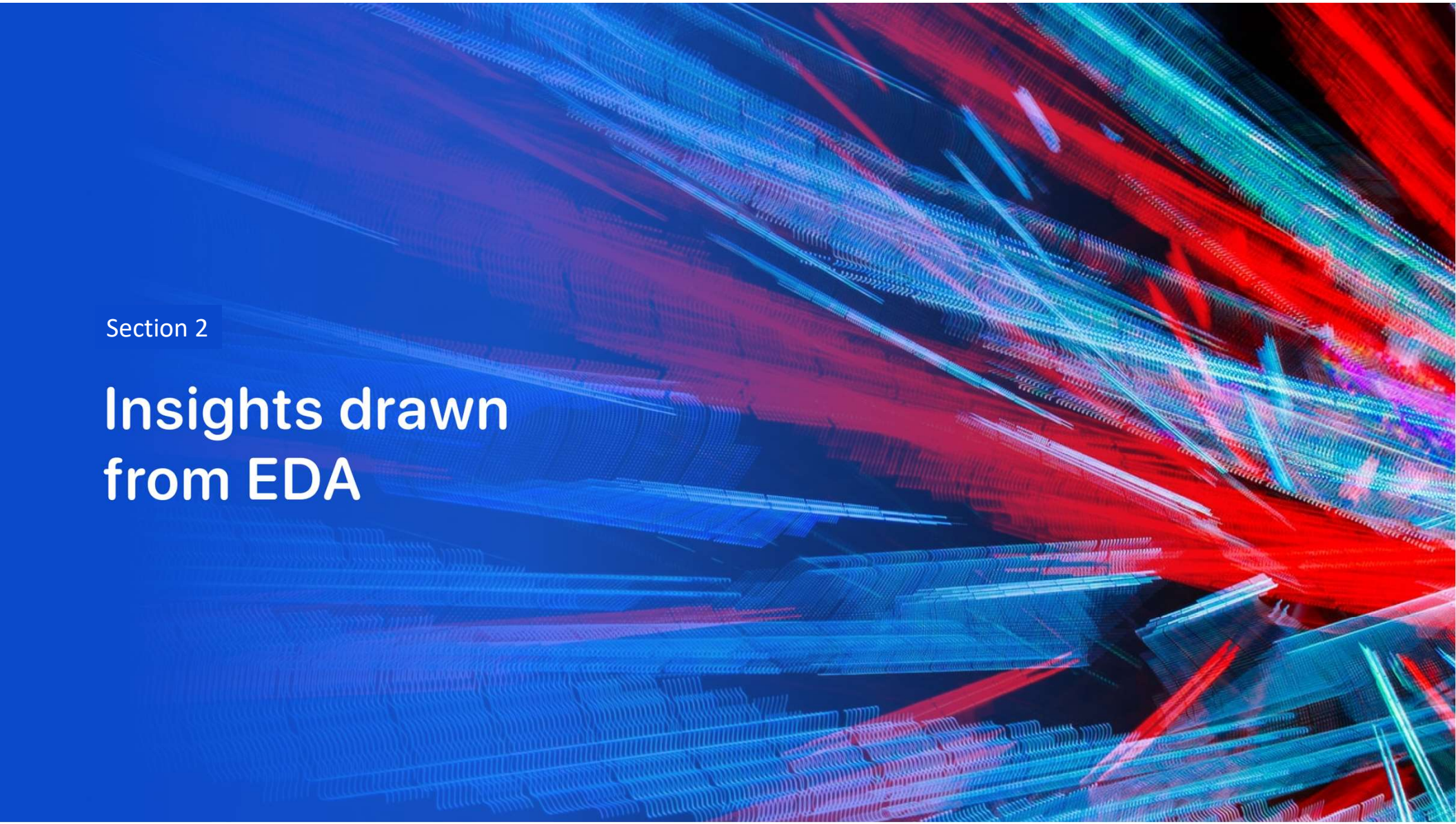




- Predictive analysis result
  - Samples



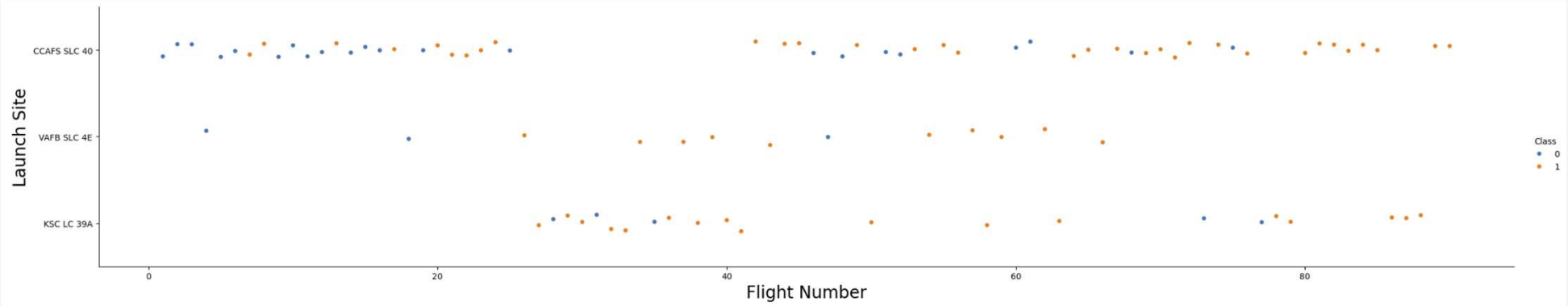
	Algo Type	Accuracy Score
2	Decision Tree	0.903571
3	KNN	0.848214
1	SVM	0.848214
0	Logistic Regression	0.846429



Section 2

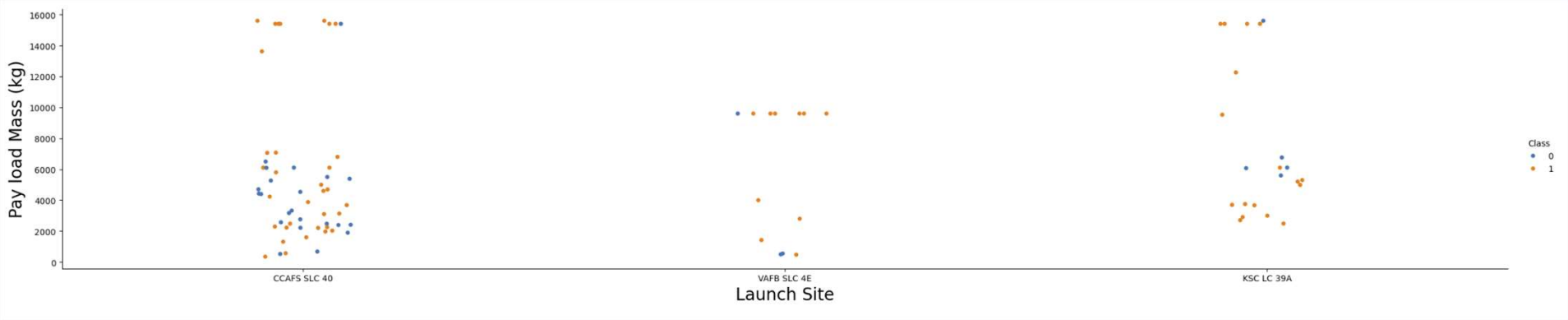
Insights drawn from EDA

Flight Number vs. Launch Site



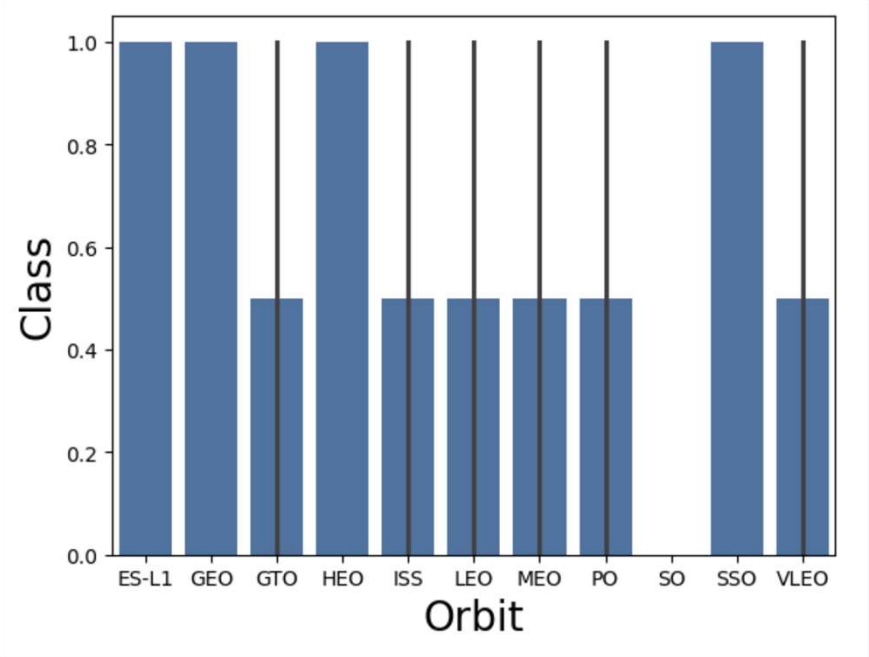
- According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful;
- In second place VAFB SLC 4E and third place KSC LC 39A;
- It's also possible to see that the general success rate improved over time.

Payload vs. Launch Site



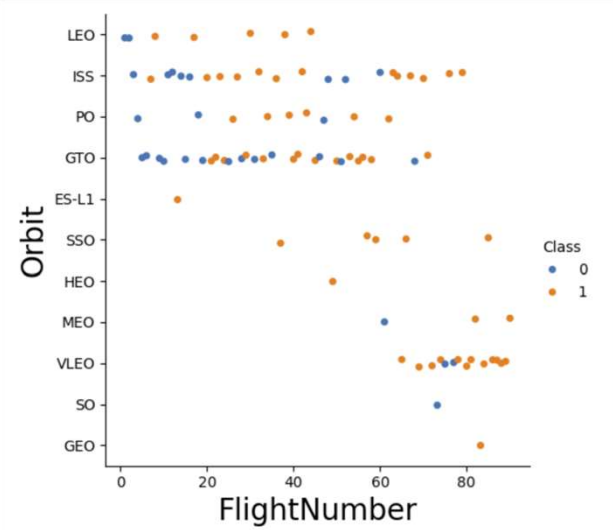
- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type



- Orbits ES-LI, GEO, HEO, and SSO have the highest success rates.
- GTO orbit has the lowest success rate.

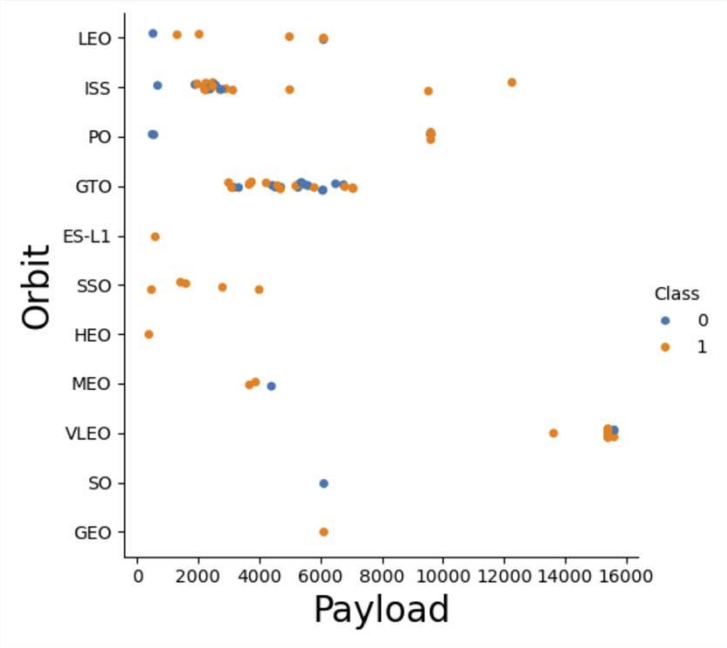
Flight Number vs. Orbit Type



- Launch Orbit preferences changed over FlightNumber.
- Launch Outcome seems to correlate with this preference.
- SpaceX started with LEO orbits which saw moderate success LEO and returned to VLEO in recent launches.

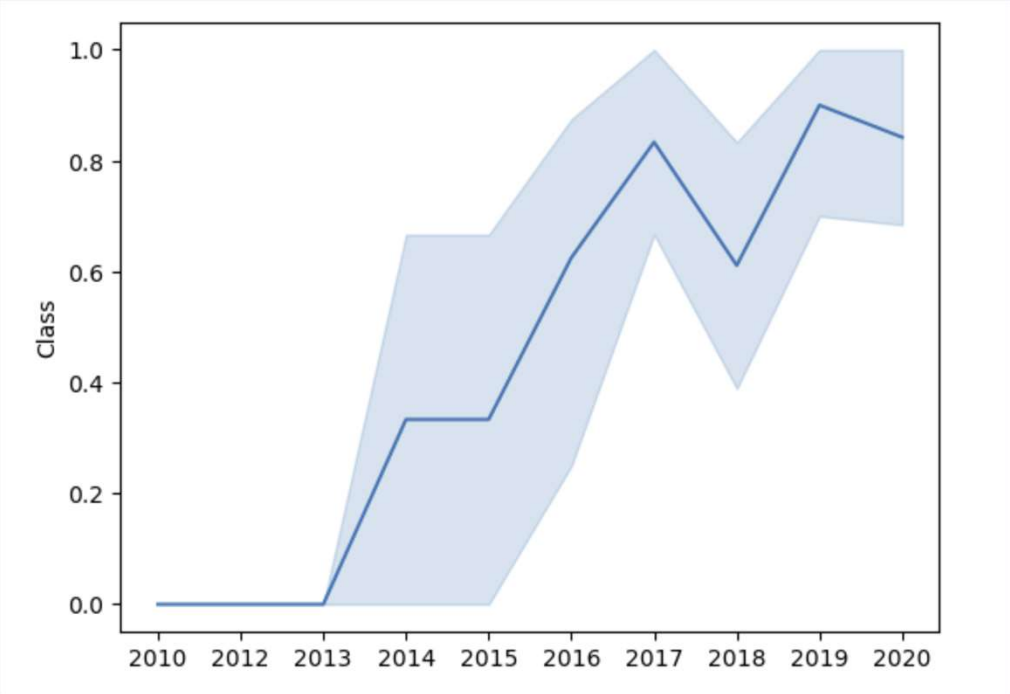
- SpaceX appearsto perform betterin lower orbits or Sun-synchronous orbits

Payload vs. Orbit Type



- Payload massseems to correlatewith orbit LEOand SSO seem to haverelatively lowpayload mass The other most successful orbit VLEO only has payloadmass valuesin the higher end of the range.

Launch Success Yearly Trend



- Success rate (Class=1) increased by about 80% between 2013 and 2020
- Success rates remained the same between 2010 and 2013 and between 2014 and 2015
- Success rates decreased between 2017 and 2018 and between 2019 and 2020

All Launch Site Names

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

- Query unique launch site names from database.
- CCAFS SLC-40 and CCAFSSLC-40 likely all represent the same.
- Launch site with data entry errors.
- CCAFS LC-40 was the previous name. Likely only 3 unique launch_site values: CCAFS SLC-40.

Launch Site Names Begin with 'CCA'

- **Query:** %sql select * from SPACEXTBL where launch_site like 'CCA%' limit 5.

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

- **Description:** Using keyword ‘Like’ and format ‘CCA%’, returns records where ‘Launch_Site’ column starts with “CCA”.
- Limit 5, limits the number of returned records to 5.

Total Payload Mass

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

```
%sql select sum(payload_mass__kg_) as sum from SPACEXTBL where customer like 'NASA (CRS)'
```

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

sum
45596

- This query sums the total payload mass in kg where NASA was the customer.
- CRS stands for Commercial Resupply Services which indicates that these payloads were sent to the International Space Station (ISS)

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2.928 kg

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

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

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

average
2534.6666666666665

- This query calculates the average payload mass or launches which used booster versionF9 v1.1
- Average payload mass of F9 1.1 is on the low end of our payload mass rang

First Successful Ground Landing Date

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

Hint:Use min function

```
%sql select min(date) as date from SPACEXTBL where mission_outcome like 'Success'
```

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

date
2010-06-04

- We use the min() function to find the result.
- We observed that the dates of the first successful landing outcome on ground pad was 22nd December 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- The list of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 is the next one:

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

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

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

- Selecting distinct booster versions according to the filters above, these 4 are the result.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

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

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

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

- This query returns a count of each mission outcome.
- SpaceX appears to achieve its mission outcome nearly 99% of the time.
- This means that most of the landing failures are intended.
- Interestingly, one launch has an unclear payload status and unfortunately one failed in flight

%

Boosters Carried Maximum Payload

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

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

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

- Query:
- **%sql** select booster_version from SPACEXTBL where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL)
- Description:
- The sub query returns the maximum payload mass by using keywork ‘max’ on the pay load mass column
- The main query returns booster versions and respective payload mass where payload mass is maximum with value of 15600

2015 Launch Records

- Query: •

```
select Landing__Outcome, Booster_Version, Launch_Site from spacextbl where Landing__Outcome = 'Failure (drone ship)' and year(Date) = '2015'
```

- Description:
- The query lists landing outcome, booster version, and the launch site where landing outcome is failed in drone ship and the year is 2015 • The ‘and’ operator in the where clause returns booster versions where both conditions in the where clause are true
- The ‘year’ keywork extracts the year from column ‘Date
- The results identify launch site as ‘CCAFS LC-40’ and booster version as F9 v1.1 B1012 and B1015 that had failed landing outcomes in drop ship in the year 2015
- Result

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

- The ranking 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 is:

```
%sql select LANDING_OUTCOME, count(*) as count from SPACEXTABLE where Date >= '2010-06-04' AND Date <= '2017-03-20' GRO
```

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

Landing_Outcome	count
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

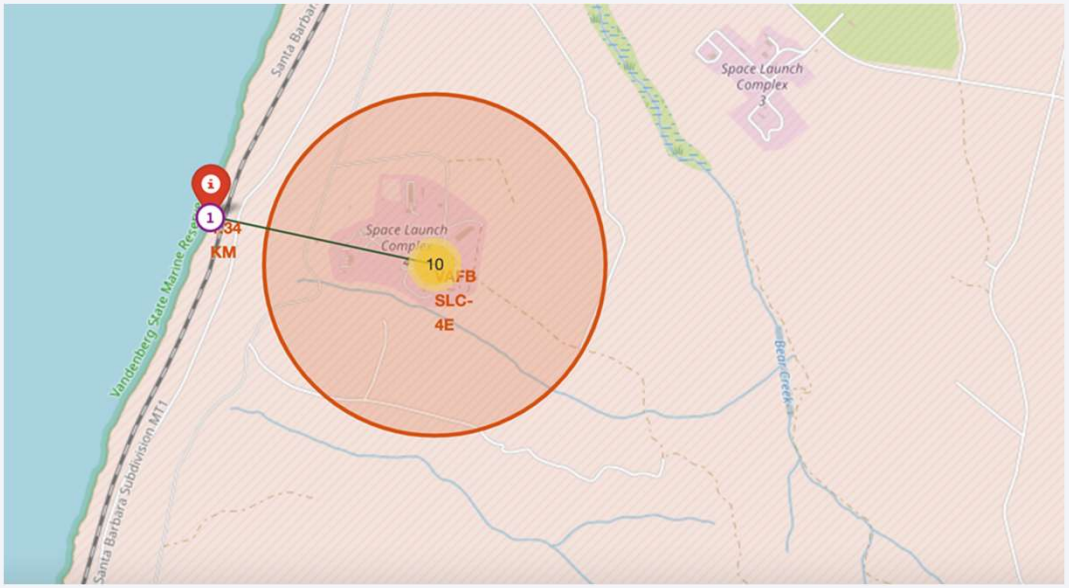
- We can see that all the SpaceX launch sites are located inside the United State.
- The left map shows all launch sites relative US map. The right map shows the two Florida launch sitessince they are very close to each other.All launch sites arenear the ocean.

Folium Map - Success Color Markers

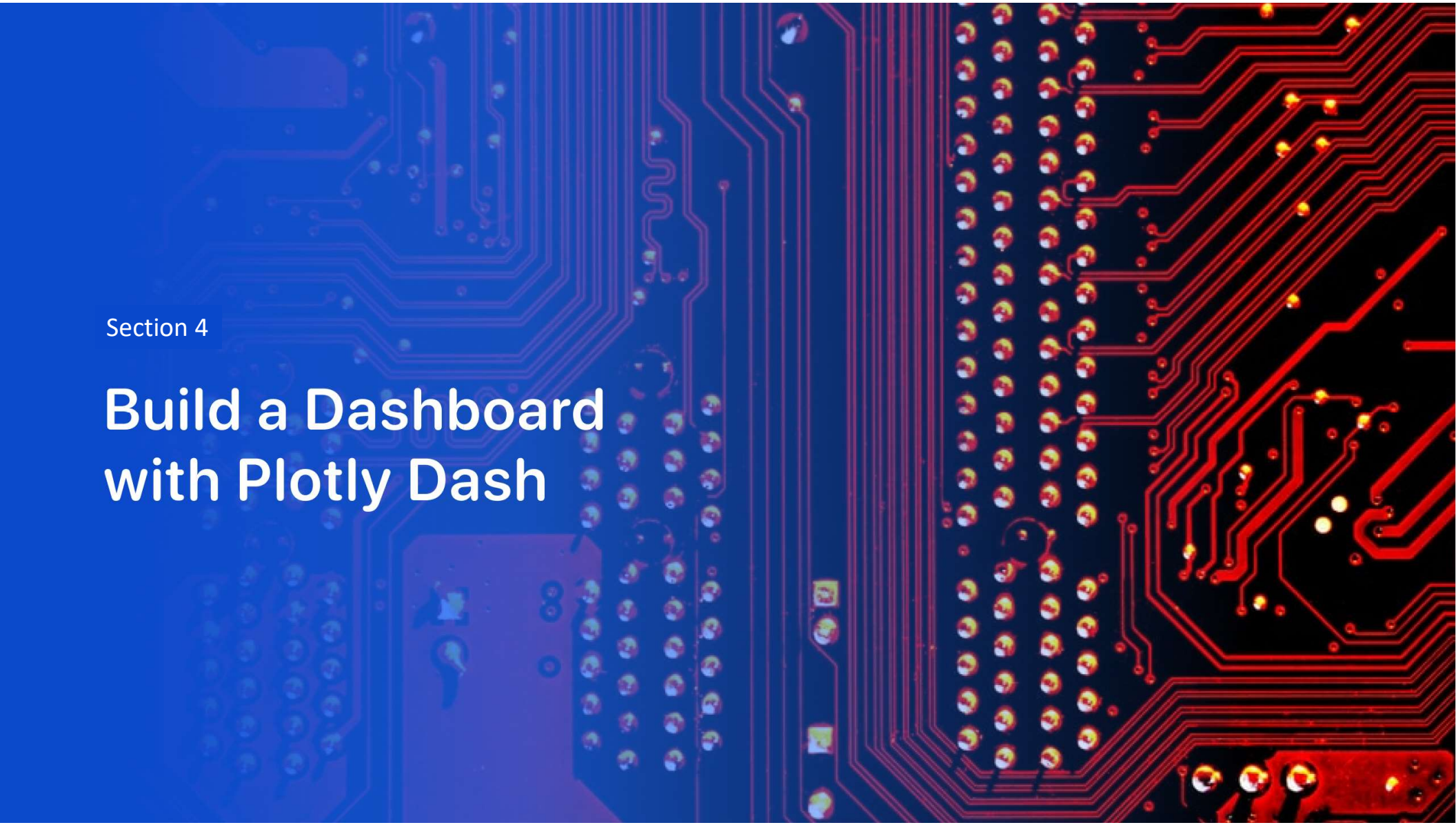
- Most of the launch sites are in Florida, while only one is in California
- If we click each Launch Site, we can see the landing sites too, where those in red failed and those in green were successful.
- Clusters on Folium map can be clicked on to display each successful landing (green icon) and failed landing (red icon). In this example VAFB SLC-4E shows 4 successful landings and 6 failed landings.

Folium Map

SpaceX Falcon9 – Launch Site to proximity Distance Map



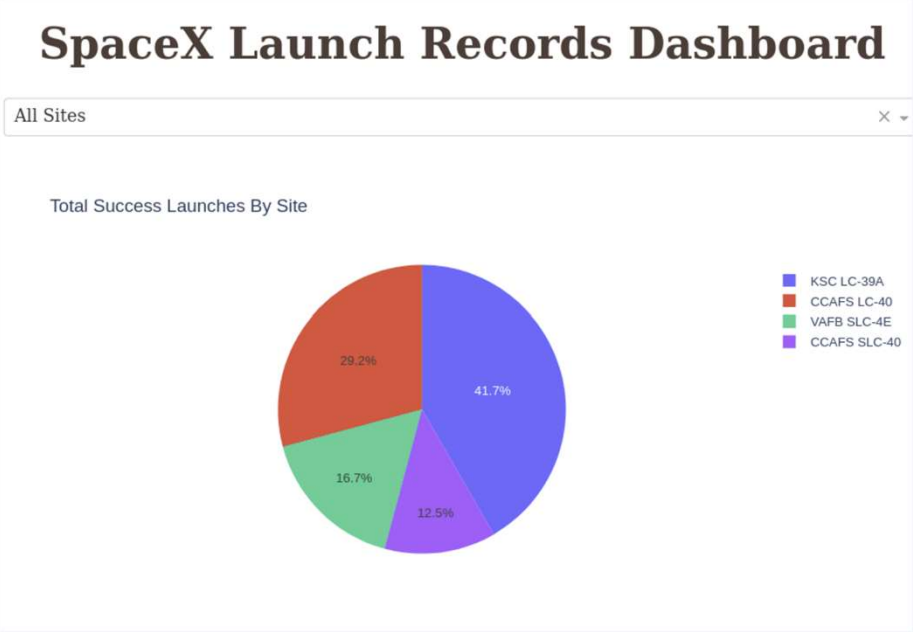
- The distances between a launch site to its proximities such as the nearest city, railway, or highway
  - The picture below shows the distance between the VAFB SLC-4E launch site and the nearest coastline



Section 4

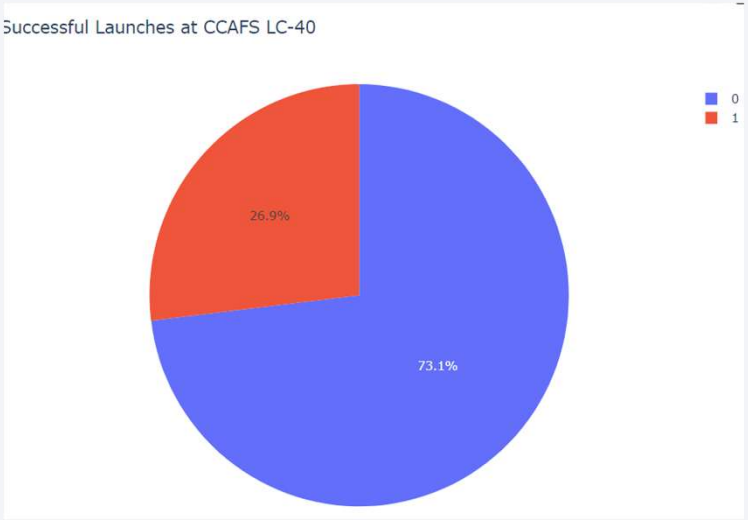
Build a Dashboard with Plotly Dash

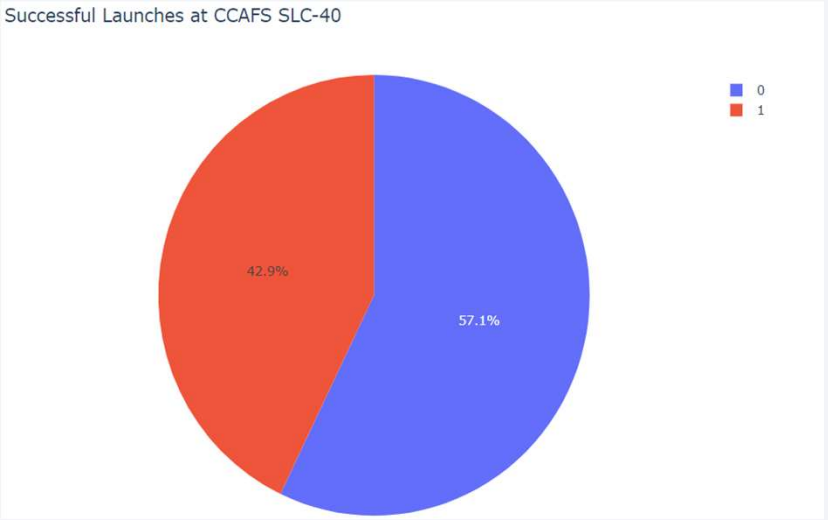
Distribution of successful launch sites



- The place from where launches are done seems to be a very important factor of success of missions
- We can see that the Launch Site CCAFS LC is the one with more successful Launches.
- The smallest one is CCAFS SLC, which can be explain by the proximity that we saw previously.

Comparation of Successful Rate





- As we saw before, LC-40 has the most successful Launches and SLC-40 the least. However, is we take a closer look, we can see that LC-40 only have 27% of successful rate, vs the 43% successful rate of SLC-40.

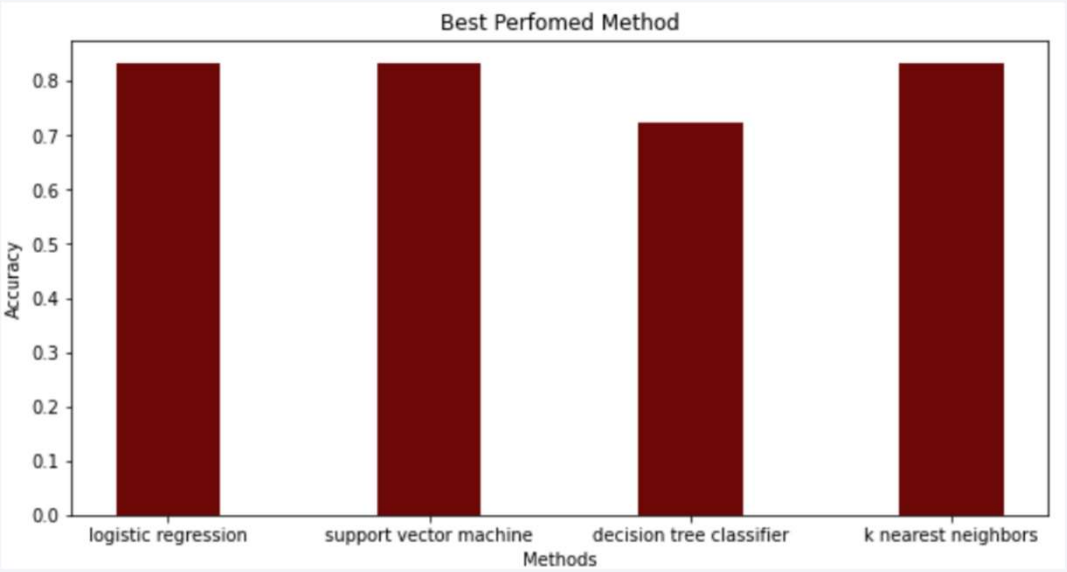
Payload vs. Launch Outcome

- Most successful launches are in the payload range from 2000 to about 5500
- Booster version category ‘FT’ has the most successful launches
- Only booster with a success launch when payload is greater than 6k is ‘B4’
- Payloads under 6,000kg and FT boosters are the most successful combination.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

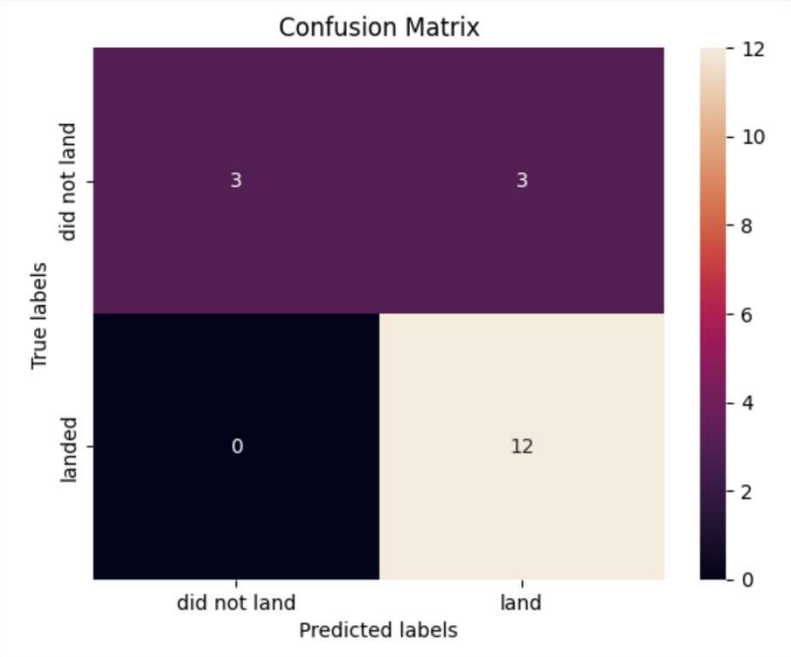


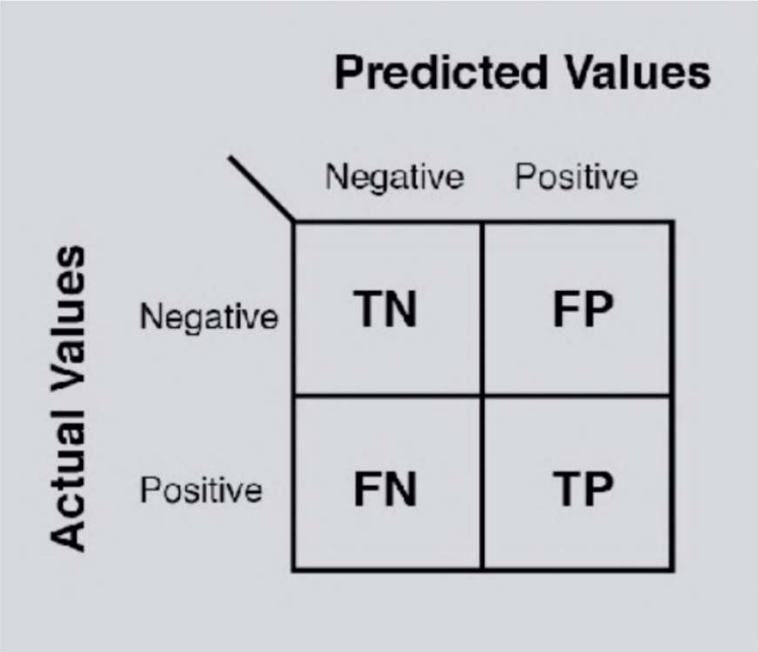
	Algo Type	Accuracy Score	Test Data Accuracy Score
2	Decision Tree	0.875000	0.833333
3	KNN	0.848214	0.833333
1	SVM	0.848214	0.833333
0	Logistic Regression	0.846429	0.833333

- Based on the Accuracy scores and as also evident from the bar chart, Decision Tree algorithm has the highest classification score with a value of .8750
- Given that the Accuracy scores for Classfication algorithms are very close and the test scores are the same, we may need a broader data set to further tune the models

- The models had virtually the same accuracy on the test set at 83.33%accuracy, except the decision tree classifier with 72,23 %.
- It should be noted that test size is small at only sample size of18.
- This can cause large variance in accuracy results, such as those in Decision Tree Classifier model in repeated runs.
- We likely need more data to determine the best model

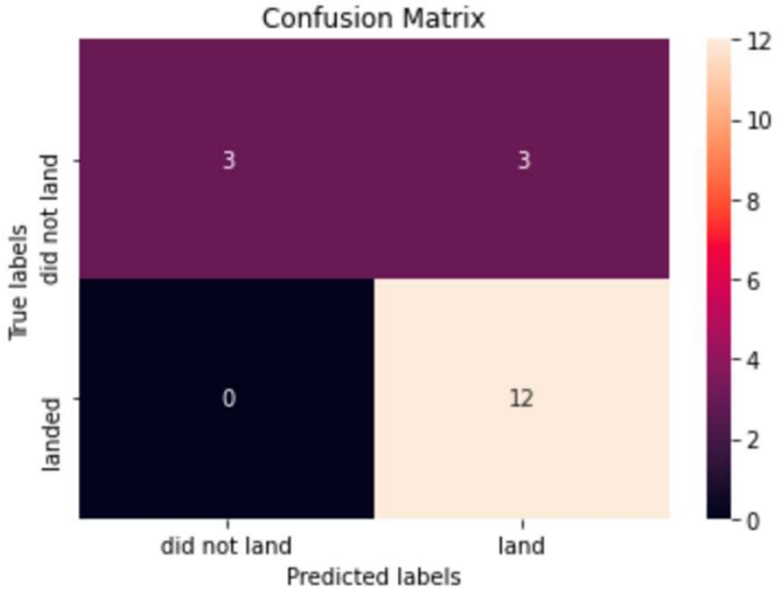
Confusion Matrix





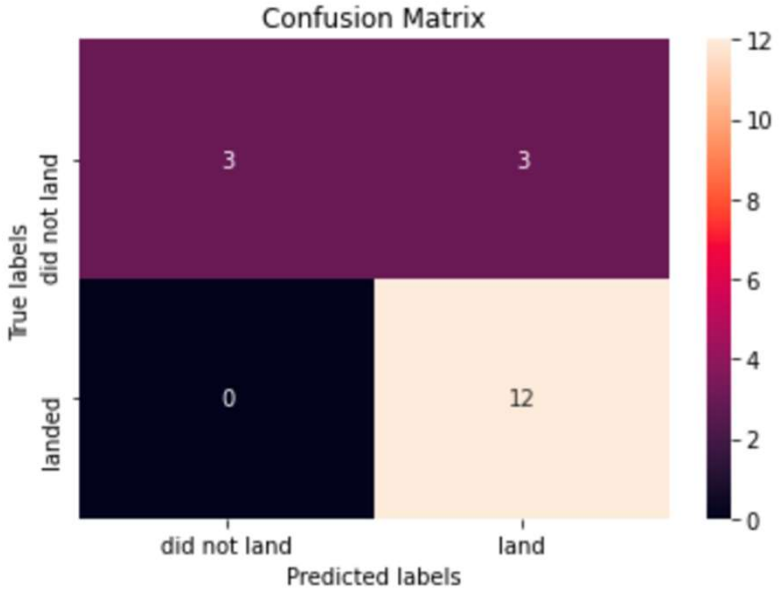
- Since all the models have the same accuracy, this confusion Matrix works for all of them. Here we can see that it has no false negative, however is have the same true negative values as false positive, which might indicate that the model is **unbalance and over predicts successful landings.**

Confusion Matrix



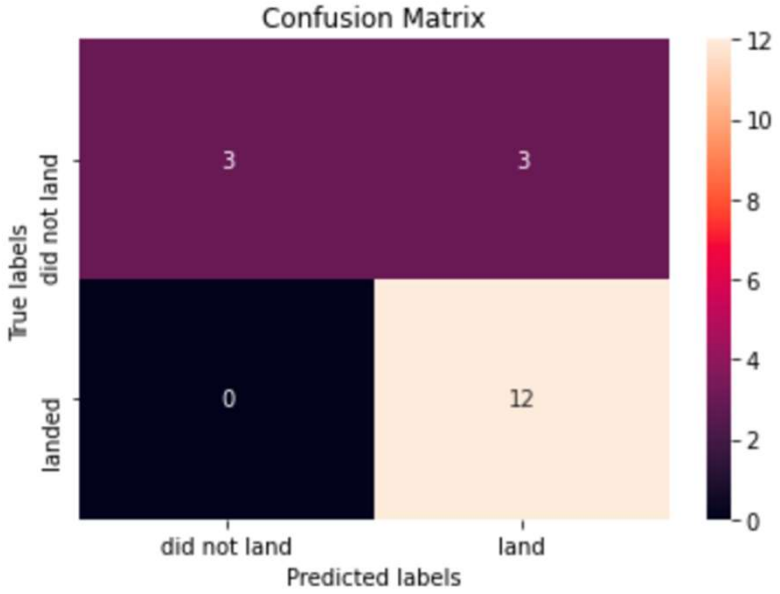
- **Logistic regression**
  - GridSearchCV best score: 0.8464285714285713
  - Accuracy score on test set: 0.8333333333333334
  - Confusion matrix:

Confusion Matrix



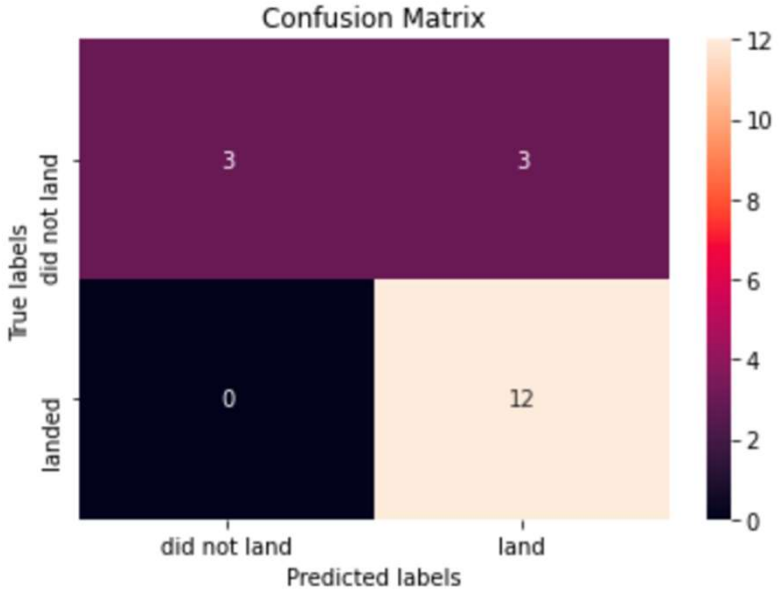
- **Support vector machine (SVM)**
  - GridSearchCV best score: 0.8482142857142856
  - Accuracy score on test set: 0.8333333333333334
  - Confusion matrix:

Confusion Matrix



- **Decision tree**
  - GridSearchCV best score: 0.8892857142857142
  - Accuracy score on test set: 0.8333333333333334
  - Confusion matrix:

Confusion Matrix



- **K nearest neighbors (KNN)**
  - GridSearchCV best score: 0.8482142857142858
  - Accuracy score on test set: 0.8333333333333334
  - Confusion matrix:

Conclusions

- The exploratory analysis showed that the best orbits to go to are ES-L1, GEO, HEO and SSO, since they have 100% success rate.
- 2019 was the year with the highest average success rate of all years.
- The launch site with highest success rate is CCAFS-SCL40
- Most of the flight have less than 8.000 kg payload mass. However, those that have more than that, present a high success rate.
- All the models had 83.33% accuracy, and all of them over predicted the successful landings, which might be a problem when estimating the cost of the project.
- The Tree Classifier Algorithm is the best Machine Learning approach for this dataset.
- The low weighted payloads (which define as 4000kg and below) performed better than the heavy weighted payloads.
- Starting from the year 2013, the success rate for SpaceX launches is increased, directly proportional time in years to 2020, which it will eventually perfect the launches in the future.
- KSC LC-39A have the most successful launches of any sites; 76.9%
- SSO orbit have the most success rate; 100% and more than 1 occurrence.

Appendix

- If you would like to see the entire notebook, please go to this address:
- https://github.com/cmenapaz/Applied_Data_Science_Capstone

Thank you!