SPACE X FALCON 9

LaShonda Henderson

EXECUTIVE SUMMARY

Executive Summary:

Using web scraping and other Python solutions, an analysis was performed using data to tell the story of Falcon 9. As defined by the source data collected:

“Falcon 9 is a partially reusable two-stage-to-orbit medium-lift launch vehicle designed and manufactured by SpaceX in the United States. The latest version of the first stage can return to Earth and be flown again multiple times.”

Data was gathered, analyzed, and predictions were made. Through the analysis, and the information produced, I was able to explore where landings took place in the United States and determine the success and failure rate of Falcon 9 launches.

INTRODUCTION SLIDE

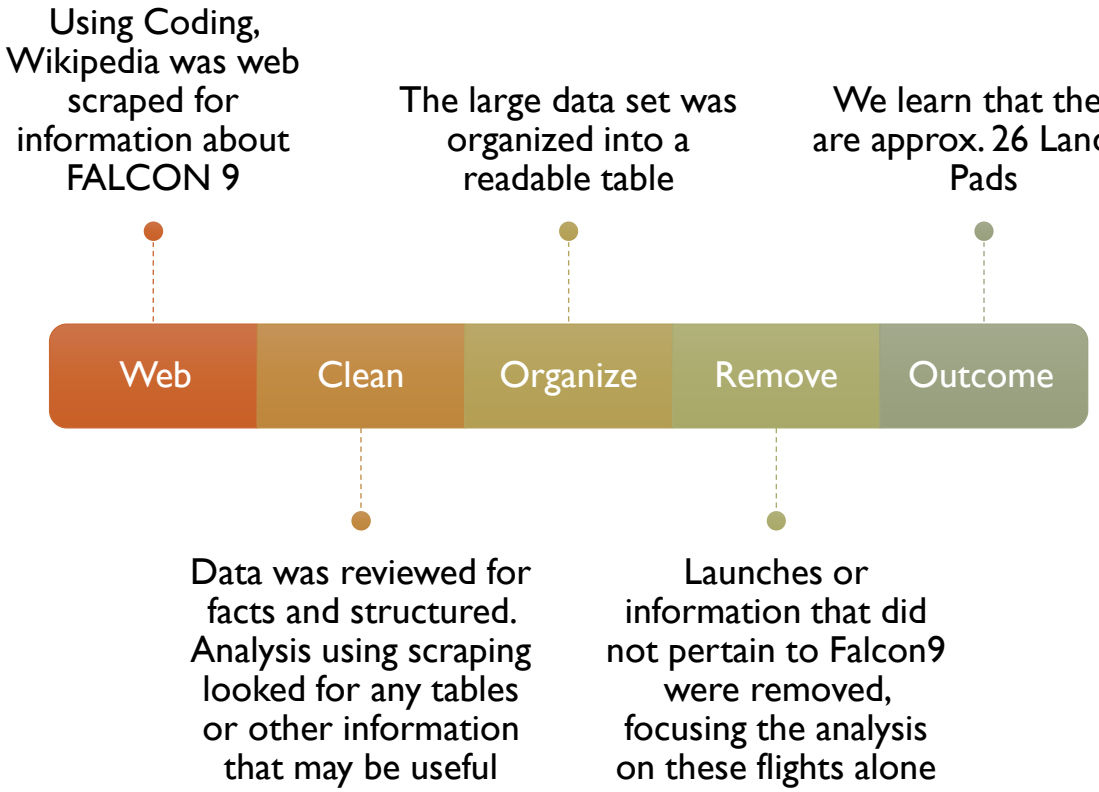
Space X is a highly anticipated and watched project. This case seeks to follow Falcon 9 rocket's success and opportunities to reconsider costs spent on future ventures. This report gives insight to opportunities for creating a similar project or enhance the current one.

We are seeking to understand and explain through visualizations and Machine Learning:

1. Are there factors that make the launch successful?
2. What data can be captured to help determine a successful landing
3. What factors must be considered to produce an environment conducive to a successful landing operation?

All results will be displayed first, with the methodology displayed next slide for the presentation

DATA COLLECTION



Git Hub:
https://github.com/lashondahenderson/CapstoneDS/blob/a231221f4169a6d937c5e2446bf4908d17e923be/Lab1.ipynb

METHODOLOGY

Data collection methodology:
Data was collected using SpaceX API and web scraping from Wikipedia.

Multiple steps had to take place to get the data:

- Appending to existing tables
- Pulling information from the URL
- "Calling" the website and "receiving" a response
- Parsing the data
- Creating a dataframe filtering the results to focus on Falcon 9

DATA WRANGLING

Using the Data compiled, landings were more than 50% successful for hitting their predicted targets.

GTO	27
ISS	21
VLEO	14
PO	9
LEO	7
SSO	5
MEO	3
GEO	1
ES-L1	1
SO	1
HEO	1

Wrangled Data years were:
2010-2016

There were 5 successful landings to the region of the Ocean predicted and 2 Occurrences outside of the predicted calculated Ocean Region

1.0

METHODOLOGY

Data wrangling methodology:
Data was collected using HTML from Wiki page.

Multiple steps had to take place to get the data:

- Creating a Beautiful Soup object
- Extracting columns and HTML table headers
- Created a Data Frame to house the information collected

```
# use requests.get() method with the provided static_url
# assign the response to a object

data = requests.get(static_url).text
```

Create a BeautifulSoup object from the HTML response

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup=BeautifulSoup(data,'html5lib')
print ('BeautifulSoup')
```

```
BeautifulSoup
```

Print the page title to verify if the BeautifulSoup object was created properly

```
# Use soup.title attribute
print(soup.title)
```

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

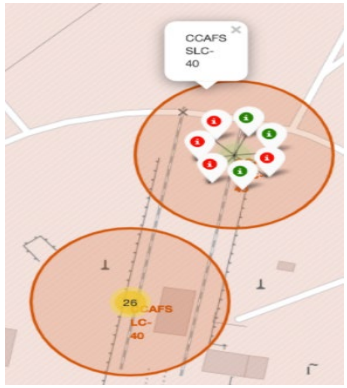
FOLIUM ANALYSIS

Launch Site Analysis

Although the data gathered only showed two sites, one for California, and one for Florida, we learn about the details around those sites, such as street roads, railroads and other highways that were close to the site. We can determine if any of those factors are important for landing outcomes.

The last 5 launches took place at site CCAFS SLC40

Launches | Pinning the Locations



https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/d8f287f4-3674-4d57-953d-fcd8202d817f/view?access_token=443a21e093a052fc88ca30ba4a7050c5639a35abe1417a59f0a1bc6e9678d2cc

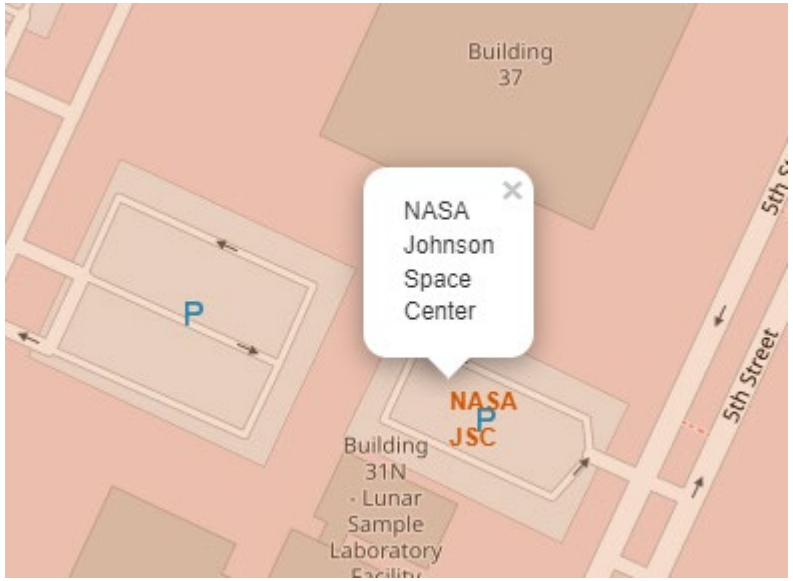
METHODOLOGY

Folium methodology:
Using GEO mapping, the launch site locations were mapped and reviewed.

Multiple steps had to take place to get the data:

- Coordinates were identified Latitude and Longitude
- Map Zoom function allowed the actual building to be viewed
- Transformed data to mark successful versus failed launches

	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610746

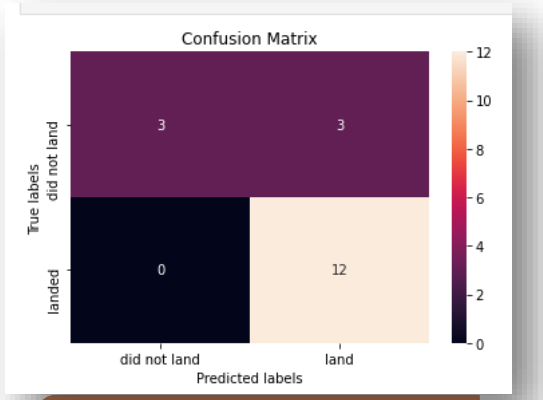


PREDICTIVE ANALYSIS

Machine Learning Predictions

There were approx. 18 test samples to begin this predication cycle.

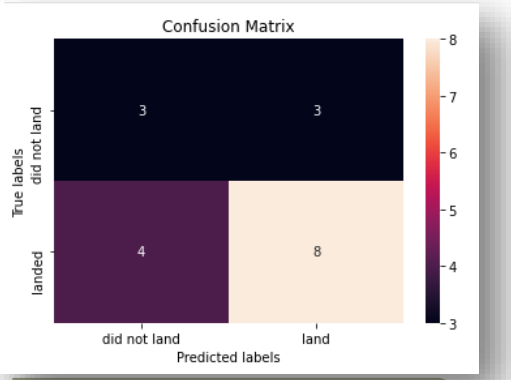
The Accuracy of the test Data using score was 0.8333



Accuracy of Test Data

12 Landings

The Accuracy of the test Data using Tree CV was 0.61111



Accuracy of Tree CV

8 Landings

Both methods predicted 3 non landings

Git HUB: CapstoneDS/Machine Learning PA.ipynb at a231221f4169a6d937c5e2446bf4908d17e923be · lashondahenderson/CapstoneDS (github.com)

METHODOLOGY

Predictive Analysis methodology:
Data was organized into methods that could leverage Machine learning. The code was manipulated, and data cleaned so the systems can make guesses that were as close to real time accuracy as possible.

Multiple steps had to take place to get the data:

- NumPy array was created
- Logistic regression objects were created
- Accuracies were determined
- Matrixes were plotted for visual considerations

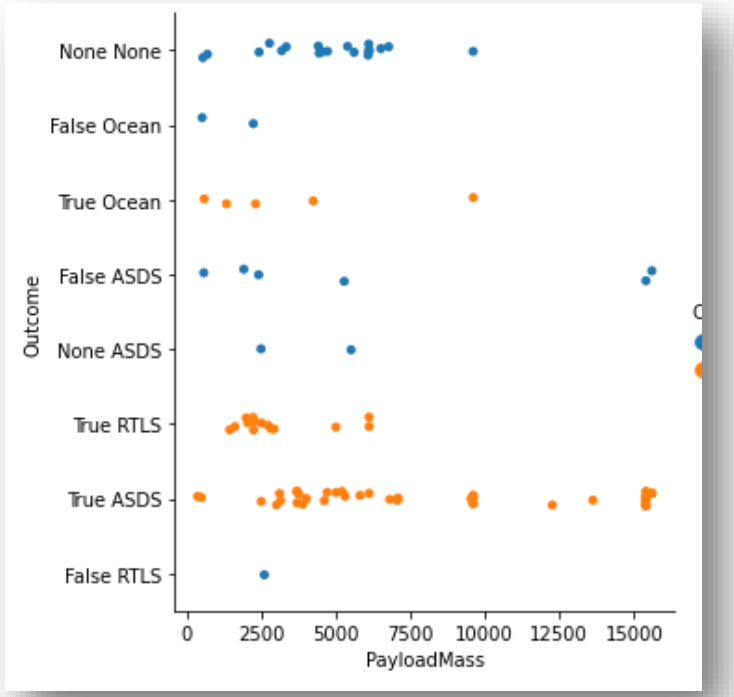
```
Y=data['Class'].to_numpy()
Y
```

```
array([0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1,
       1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1,
       1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1])
```

```
methods = []
accuracy = []
methods.append('Logistic regression')
accuracy.append(logreg_cv.score(X_test,Y_test))
print ("Accuracy :",logreg_cv.score(X_test,Y_test))
```

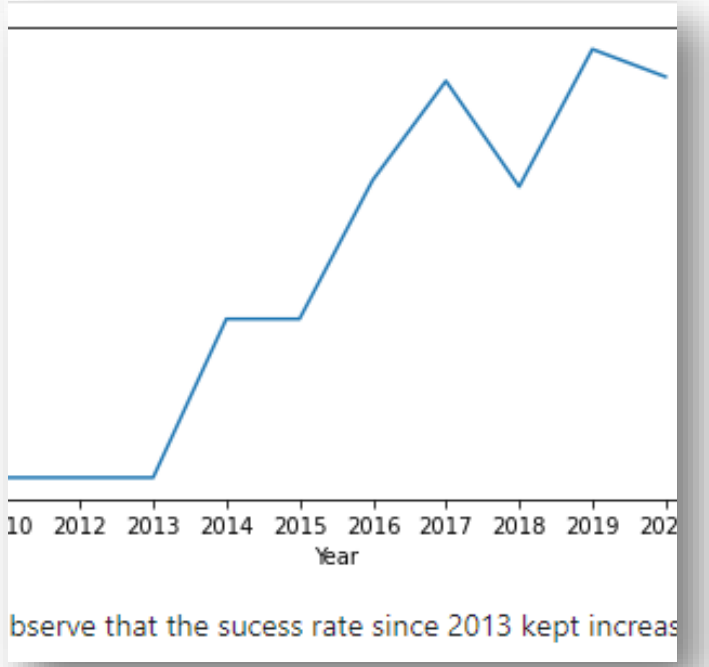
```
Accuracy : 0.8333333333333334
```

EDA WITH VISUALIZATION



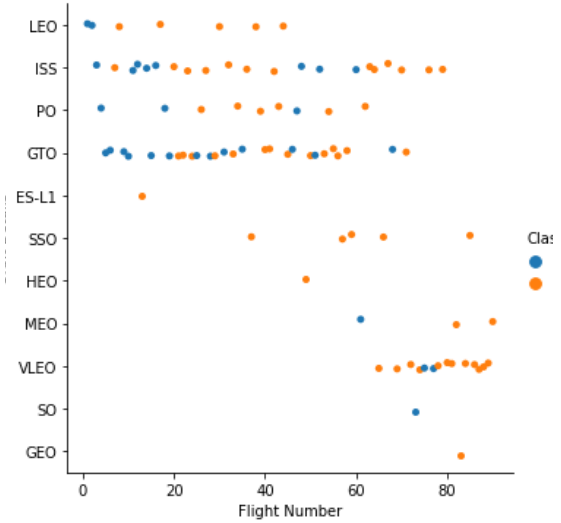
The payload mass, the weight of the Rocket and the landmarks near, produced outcomes that were not surprising. The heavier the rocket, the more landings met their targets.

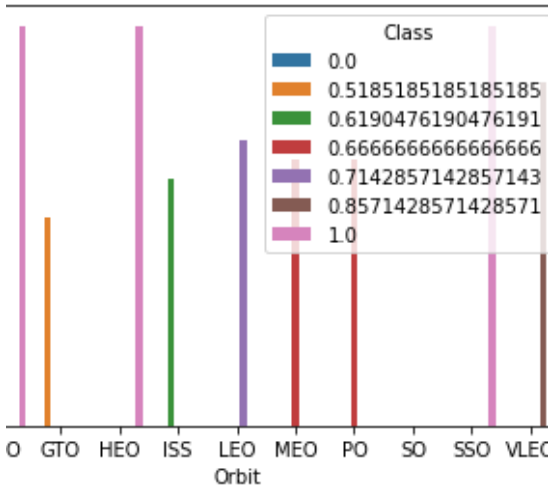
We find a steady increase of launches from 2013, until 2020. 2010, and 2012 saw no successful launches.



EDA WITH VISUALIZATION

As the number of flights increased, the number expanded metrics increases as well. Plain language the more launches, flights became more frequent.





Orbit types place a role of successful launches. SSO, HEO, GEO, and ES-L1 Being the most successful

METHODOLOGY

EDA Visuals methodology:
Data was pulled from a CSV file, then placed into visualizations that could tell the flight stories.

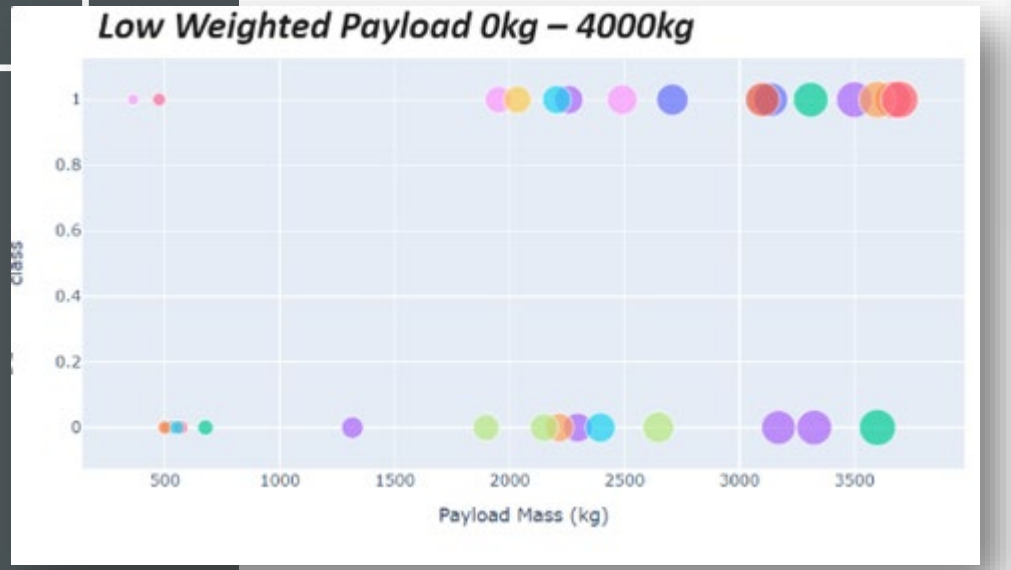
Multiple steps had to take place to get the data:

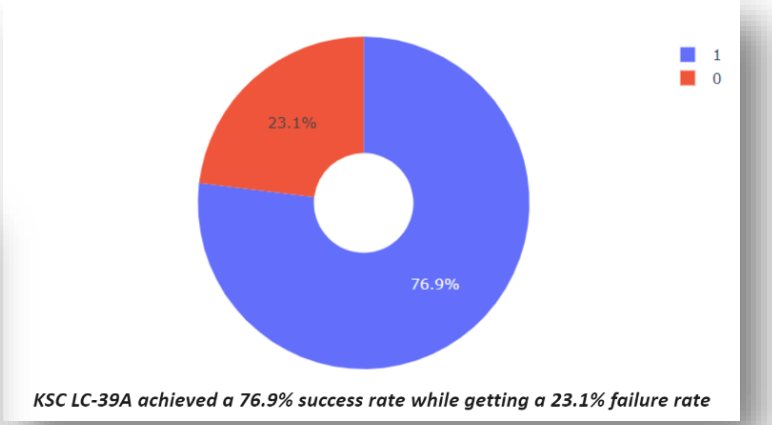
- Creating a Table from the CSV file
- Plots were created showing flights vs payloads mass
- Success rates by Orbit were plotted
- Finally, a trend analysis was created

```
# A function to Extract years from the date
year=[]
def Extract_year(date):
    for i in df["Date"]:
        year.append(i.split("-")[0])
    return year
Extract_year(1)
df["Year"]=year
average_by_year = df.groupby(by="Year").mean()
average_by_year.reset_index(inplace=True)
```

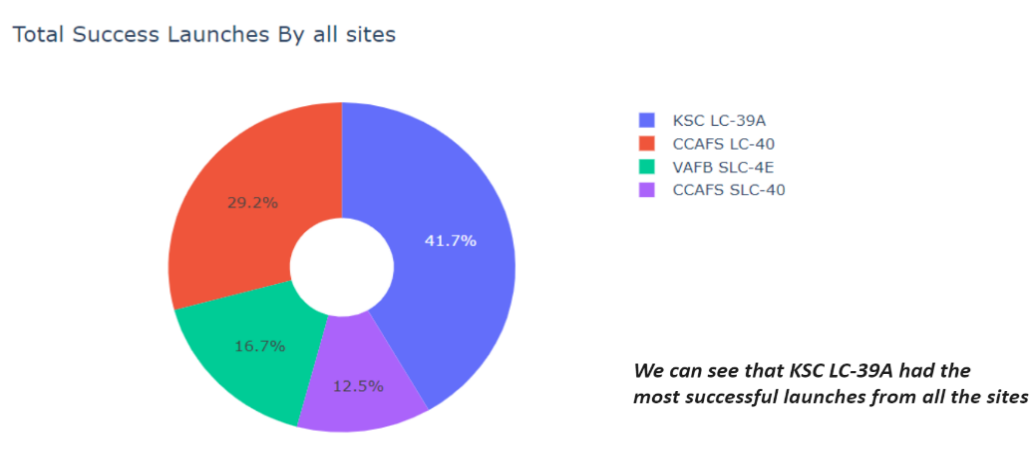
```
# Plot a line chart with x axis to be the extracted year and y axis to be the success rate
plt.plot(average_by_year["Year"],average_by_year["Class"])
plt.xlabel("Year")
plt.ylabel("Success/Failure")
plt.show()
```

PLOTLY DASH





- Low Weighted Payloads had the most success
- KSC LC-39a had a high success rate but there are areas for improvement with a 23% failure rate



CONCLUSION

While the initial launches did not see success in 2010, and 2012, by 2020 there were many successful launches that had markers, milestones, and measures that could be used to assess what "SUCCESS" looks like for a decades old program.

Falcon 9 can be used as a model to indicate areas of change, methods of predication, and geographical outcomes to be used for success.