IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Patrick Demario
22 February 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection via API and Web Scraping
  - Data Wrangling
  - Data analysis with Python & SQL
  - Machine Learning Prediction

- Summary of all results
  - Result of Data Analysis
  - Predictive Analysis
  - Interactive

Introduction

- Space X's Falcon 9 cost $62 Million much cheaper than other providersWe want to know if this information can be used if an alternate company wants to bid against space X for a rocket launch. In this lab, you will create a machine learning pipeline to predict if the first stage will land given the data from the preceding labs.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX API and Wikipedia
- Perform data wrangling
  - Hot Encoding
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

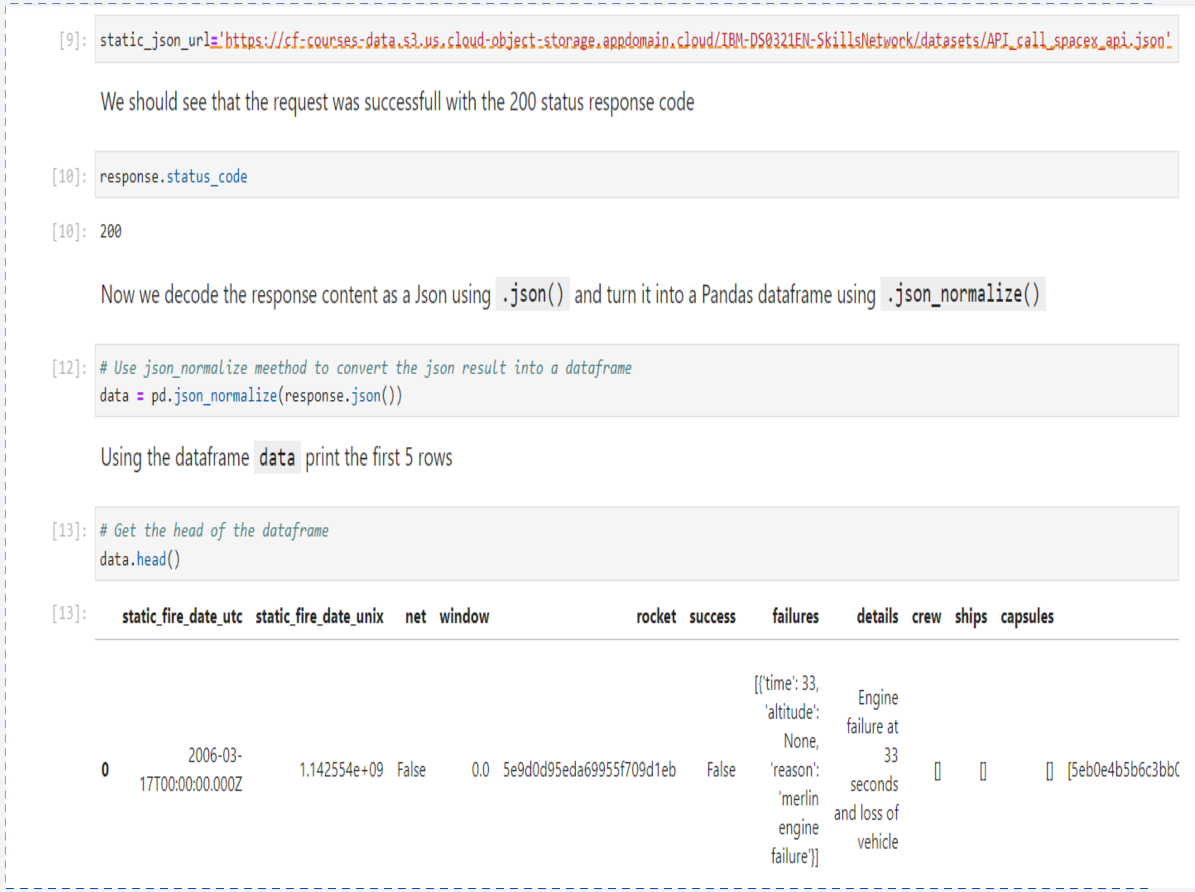
Data Collection

- Get Request of the SpaceX API
- Decode the JSON content using .json() function
- Clean the data

Data Collection – SpaceX API

- This is how data collection with SpaceX REST calls using key phrases and flowcharts

- https://github.com/vellos0x1/IBM-Coursera-Final-Course/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



```
[9]: static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

We should see that the request was successfull with the 200 status response code

```
[10]: response.status_code
[10]: 200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
[12]: # Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

Using the dataframe `data` print the first 5 rows

```
[13]: # Get the head of the dataframe
data.head()
```

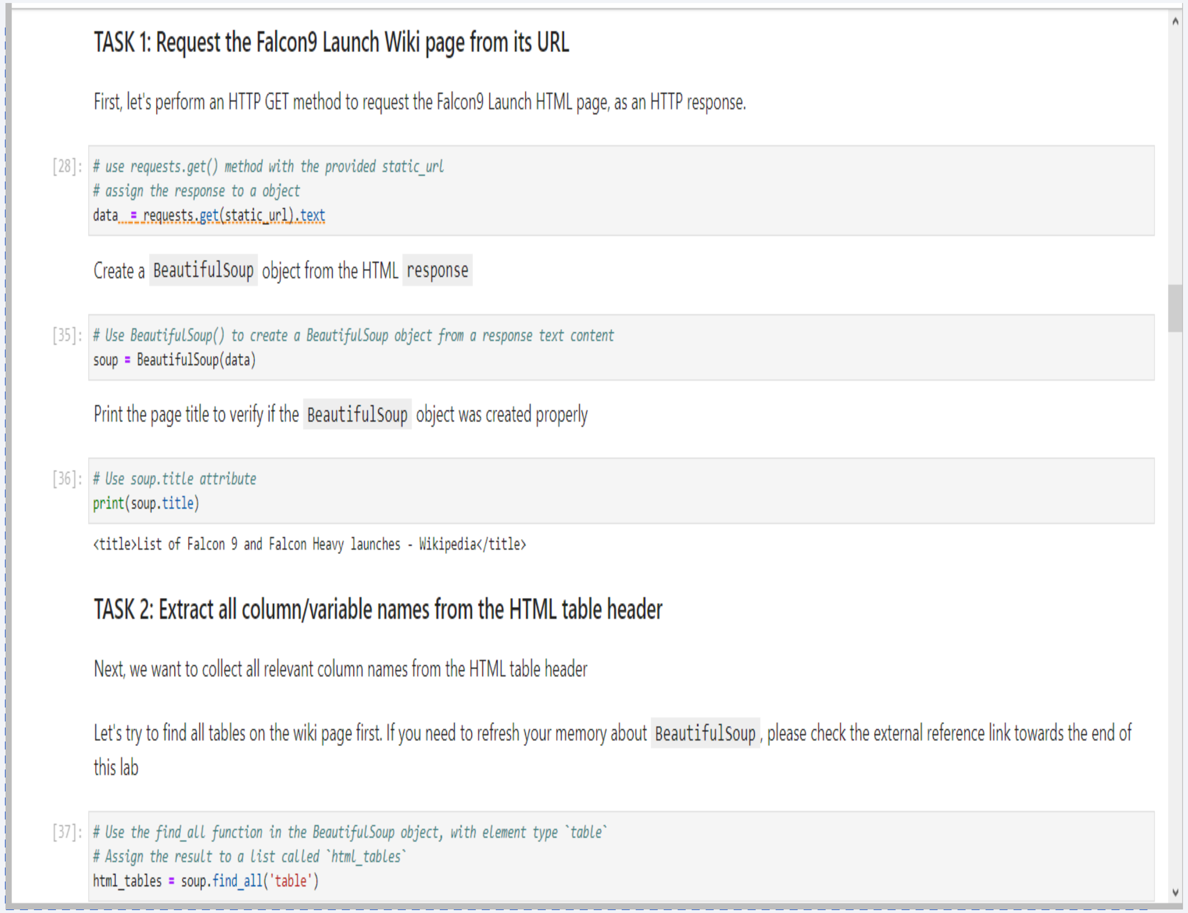
[13]:

	static_fire_date_utc	static_fire_date_unix	net	window	rocket	success	failures	details	crew	ships	capsules
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}]	Engine failure at 33 seconds and loss of vehicle	[]	[]	[] [5eb0e4b5b6c3bb0

Data Collection - Scraping

- This is how web scraping process using key phrases and flowcharts

- https://github.com/vellos0x1/IBM-Coursera-Final-Course/blob/main/jupyter-labs-webscraping%20(3).ipynb



Data Wrangling

- Perform Data Analysis and determined the train label
- Calculate the launches, and quantity of orbits
- Create the landing outcome label and then export as .csv
- https://github.com/vellos0x1/IBM-Coursera-Final-Course/blob/main/jupyter-labs-webscraping%20(3).ipynb

EDA with Data Visualization

- We use scatter plots because it’s the best chart to visualize the data of flight number and site, payload, success rate, and the orbit type
- https://github.com/vellos0x1/IBM-Coursera-Final-Course/blob/main/jupyter-labs-webscraping%20(4).ipynb

EDA with SQL

- ~~Using bullet point format, summarize the SQL queries you performed~~
- ~~Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose~~
- **IBM WATSON STUDIO stopped working so theres nothing in here**

Build an Interactive Map with Folium

- The map objects such as markers, circles, and lines to mark the success or failure of each launch
- https://github.com/vellos0x1/IBM-Coursera-Final-Course/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- We use Pie Charts to show total launches of sites and Scatter Graphs for Payload mass for different boosters
- https://github.com/vellos0x1/IBM-Coursera-Final-Course/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Load numpy and pandas, transform data, split into train and test.
- You need present your model development process using key phrases and flowchart
- https://github.com/vellos0x1/IBM-Coursera-Final-Course/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations

IBM WATSON STUDIO stopped working so theres nothing in here

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

- Show the screenshot of the scatter plot with explanations

IBM WATSON STUDIO stopped working so theres nothing in here

Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type
- Show the screenshot of the scatter plot with explanations

IBM WATSON STUDIO stopped working so theres nothing in here

Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations

IBM WATSON STUDIO stopped working so theres nothing in here

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations

IBM WATSON STUDIO stopped working so theres nothing in here

Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- Show the screenshot of the scatter plot with explanations

IBM WATSON STUDIO stopped working so theres nothing in here

All Launch Site Names

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

Launch Site Names Begin with 'CCA'

- I have issues with my IBM Watson Studio so here is the top 5 Launch sites with names begin with 'CCA'

	A	B	C	D	E	F	G	H	
1	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission
2	04/06/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
3	08/12/2010	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
4	22/05/2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
5	08/10/2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
6	01/03/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

- I have issues with my IBM Watson Studio so here is the total sum of payload mass in the spreadsheet. The total payload Mass is 619967 KGs or 619.967 Tonnes

=SUM([PAYLOAD_MASS__KG_])

oster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_
B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15600
B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600
B5B1061.1	KSC LC-39A	Crew-1, Sentinel-6 Michael Freilich	12500
B5 B1058.4	KSC LC-39A	SpaceX CRS-21	2972
v1.1 B1003	VAFB SLC-4E	CASSIOPE	500
v1.1 B1017	VAFB SLC-4E	Jason-3	553
FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600
FT B1036.1	VAFB SLC-4E	Iridium NEXT 2	9600
FT B1038.1	VAFB SLC-4E	Formosat-5	475
B4 B1041.1	VAFB SLC-4E	Iridium NEXT 3	9600
FT B1036.2	VAFB SLC-4E	Iridium NEXT 4	9600
FT B1038.2	VAFB SLC-4E	Paz Tintin A & B	2150
B4 B1041.2	VAFB SLC-4E	Iridium NEXT 5	9600
B4 B1043.2	VAFB SLC-4E	Iridium NEXT 6 GRACE-FO 1, 2	6460
B5B1048.1	VAFB SLC-4E	Iridium NEXT-7	9600
B5 B1048.2	VAFB SLC-4E	SAOCOM 1A	3000
B5 B1046.3	VAFB SLC-4E	SSO-A	4000
B5 B1049.2	VAFB SLC-4E	Iridium NEXT-8	9600
B5 B1051.2	VAFB SLC-4E	RADARSAT Constellation, SpaceX CRS-18	4200
B5B1063.1	VAFB SLC-4E	Sentinel-6 Michael Freilich, Starlink 15 v1.0	1192
			619967

Average Payload Mass by F9 v1.1

- I have issues with my IBM Watson Studio so here is the average payload by F9 v1.1 mass in the spreadsheet. The average payload Mass is 2843.6 KGs or 2.84 Tonnes

7	03/12/2013	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170
8	06/01/2014	22:06:00	F9 v1.1	CCAFS LC-40	Thaicom 6	3325
9	18/04/2014	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296
10	14/07/2014	15:15:00	F9 v1.1	CCAFS LC-40	OG2 Mission 1 6 Orbcomm-OG2 satellites	1316
11	05/08/2014	08:00:00	F9 v1.1	CCAFS LC-40	AsiaSat 8	4535
12	07/09/2014	05:00:00	F9 v1.1 B1011	CCAFS LC-40	AsiaSat 6	4428
13	21/09/2014	05:52:00	F9 v1.1 B1010	CCAFS LC-40	SpaceX CRS-4	2216
14	10/01/2015	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395
15	11/02/2015	23:03:00	F9 v1.1 B1013	CCAFS LC-40	DSCOVR	570
16	02/03/2015	03:50:00	F9 v1.1 B1014	CCAFS LC-40	ABS-3A Eutelsat 115 West B	4159
17	14/04/2015	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898
18	27/04/2015	23:03:00	F9 v1.1 B1016	CCAFS LC-40	Turkmen 52 / MonacoSAT	4707
19	28/06/2015	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952

First Successful Ground Landing Date

- I have issues with my IBM Watson Studio so here is the first successful ground landing date. Which is on 22/12/2015

92	22/12/2015	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2	11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground

Successful Drone Ship Landing with Payload between 4000 and 6000

- I have issues with my IBM Watson Studio so here is the successful drone ship landing with payload between 4000 & 6000
- F9 FT B1022
- F9 FT B1026
- F9 FT B1021.2
- F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- I have issues with my IBM Watson Studio so here is the total number of successful and failure missions
- Success 100 (1 success with payload unknown status)
- Failure 1 (failure inlfight)

Date	Time (UTC)	Booster_Versior	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outco
28/06/2015	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (dron
04/06/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachu
08/12/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere chees	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachu
18/04/2018	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362	HEO	NASA (LSP)	Success	Success (drone s
24/08/2017	18:51:00	F9 FT B1038.1	VAFB SLC-4E	Formosat-5	475	SSO	NSPO	Success	Success (drone s
08/10/2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
29/09/2013	16:00:00	F9 v1.1 B1003	VAFB SLC-4E	CASSIOPE	500	Polar LEO	MDA	Success	Uncontrolled (oc
22/05/2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
17/01/2016	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553	LEO	NASA (LSP) NOAA CNES	Success	Failure (drone sh
11/02/2015	23:03:00	F9 v1.1 B1013	CCAFS LC-40	DSCOVR	570	HEO	U.S. Air Force NASA NOAA	Success	Controlled (ocea
01/03/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
21/11/2020	17:17:08	F9 B5B1063.1	VAFB SLC-4E	Sentinel-6 Michael Freilich, Starlink 15 v1.0	1192	LEO	NASA / NOAA / ESA / EUMETSAT	Success	Success
14/07/2014	15:15:00	F9 v1.1	CCAFS LC-40	OG2 Mission 1 6 Orbcomm-OG2 satellites	1316	LEO	Orbcomm	Success	Controlled (ocea
14/04/2015	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone sh
07/03/2020	04:50:00	F9 B5 B1059.2	CCAFS SLC-40	SpaceX CRS-20, Starlink 5 v1.0	1977	LEO (ISS)	NASA (CRS)	Success	Success
22/12/2015	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground
22/02/2018	14:17:00	F9 FT B1038.2	VAFB SLC-4E	Paz Tintin A & B	2150	SSO	Hisdesat exactEarth SpaceX	Success	No attempt
15/12/2017	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205	LEO (ISS)	NASA (CRS)	Success	Success (ground
21/09/2014	05:52:00	F9 v1.1 B1010	CCAFS LC-40	SpaceX CRS-4	2216	LEO (ISS)	NASA (CRS)	Success	Uncontrolled (oc
18/07/2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground
25/07/2019	22:01:00	F9 B5 B1056.2	CCAFS SLC-40	SpaceX CRS-18, AMOS-17	2268	LEO (ISS)	NASA (CRS)	Success	Success
18/04/2014	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)	Success	Controlled (ocea
10/01/2015	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone sh
19/02/2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground
04/05/2019	06:48:00	F9 B5B1056.1	CCAFS SLC-40	SpaceX CRS-17, Starlink v0.9	2495	LEO (ISS)	NASA (CRS)	Success	Success
05/12/2018	18:16:00	F9 B5B1050	CCAFS SLC-40	SpaceX CRS-16	2500	LEO (ISS)	NASA (CRS)	Success	Failure
05/12/2019	17:29:00	F9 B5B1059.1	CCAFS SLC-40	SpaceX CRS-19, JCSat-18 / Kacific 1	2617	LEO (ISS)	NASA (CRS), Kacific 1	Success	Success
02/04/2018	20:30:00	F9 B4 B1039.2	CCAFS SLC-40	SpaceX CRS-14	2647	LEO (ISS)	NASA (CRS)	Success	No attempt
29/06/2018	09:42:00	F9 B4 B1045.2	CCAFS SLC-40	SpaceX CRS-15	2697	LEO (ISS)	NASA (CRS)	Success	No attempt
03/06/2017	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708	LEO (ISS)	NASA (CRS)	Success	Success (ground

Boosters Carried Maximum Payload

- I have issues with my IBM Watson Studio so here are the boosters that carried the maximum payload

	Date	Time (UTC)	Booster_Versior	Launch_Site	Payload	PAYLOAD_MASS__KG_
2	17/02/2020	15:05:00	F9 B5 B1056.4	CCAFS SLC-40	Starlink 4 v1.0, SpaceX CRS-20	15600
3	18/03/2020	12:16:00	F9 B5 B1048.5	KSC LC-39A	Starlink 5 v1.0, Starlink 6 v1.0	15600
4	11/11/2019	14:56:00	F9 B5 B1048.4	CCAFS SLC-40	Starlink 1 v1.0, SpaceX CRS-19	15600
5	07/01/2020	02:33:00	F9 B5 B1049.4	CCAFS SLC-40	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
6	29/01/2020	14:07:00	F9 B5 B1051.3	CCAFS SLC-40	Starlink 3 v1.0, Starlink 4 v1.0	15600
7	04/06/2020	01:25:00	F9 B5 B1049.5	CCAFS SLC-40	Starlink 7 v1.0, Starlink 8 v1.0	15600
8	24/10/2020	15:31:34	F9 B5 B1060.3	CCAFS SLC-40	Starlink 14 v1.0, GPS III-04	15600
9	25/11/2020	02:13:00	F9 B5 B1049.7	CCAFS SLC-40	Starlink 15 v1.0, SpaceX CRS-21	15600
10	22/04/2020	19:30:00	F9 B5 B1051.4	KSC LC-39A	Starlink 6 v1.0, Crew Dragon Demo-2	15600
11	03/09/2020	12:46:14	F9 B5 B1060.2	KSC LC-39A	Starlink 11 v1.0, Starlink 12 v1.0	15600
12	06/10/2020	11:29:34	F9 B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15600

- F9 B5 B1056.2
- F9 B5 B1056.3
- F9 B5 B1056.4
- F9 B5 B1056.5
- F9 B5 B1056.7

2015 Launch Records

- I have issues with my IBM Watson Studio so here is the name of the boosters that has failure while landing in drone ship in the year 2015
- F9 v1.1 B1015
- F9 v.1.1 B1012

14/04/2015	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6		1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone s
10/01/2015	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5		2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone s

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

- 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

- Present your query result with a short explanation here

Section 3

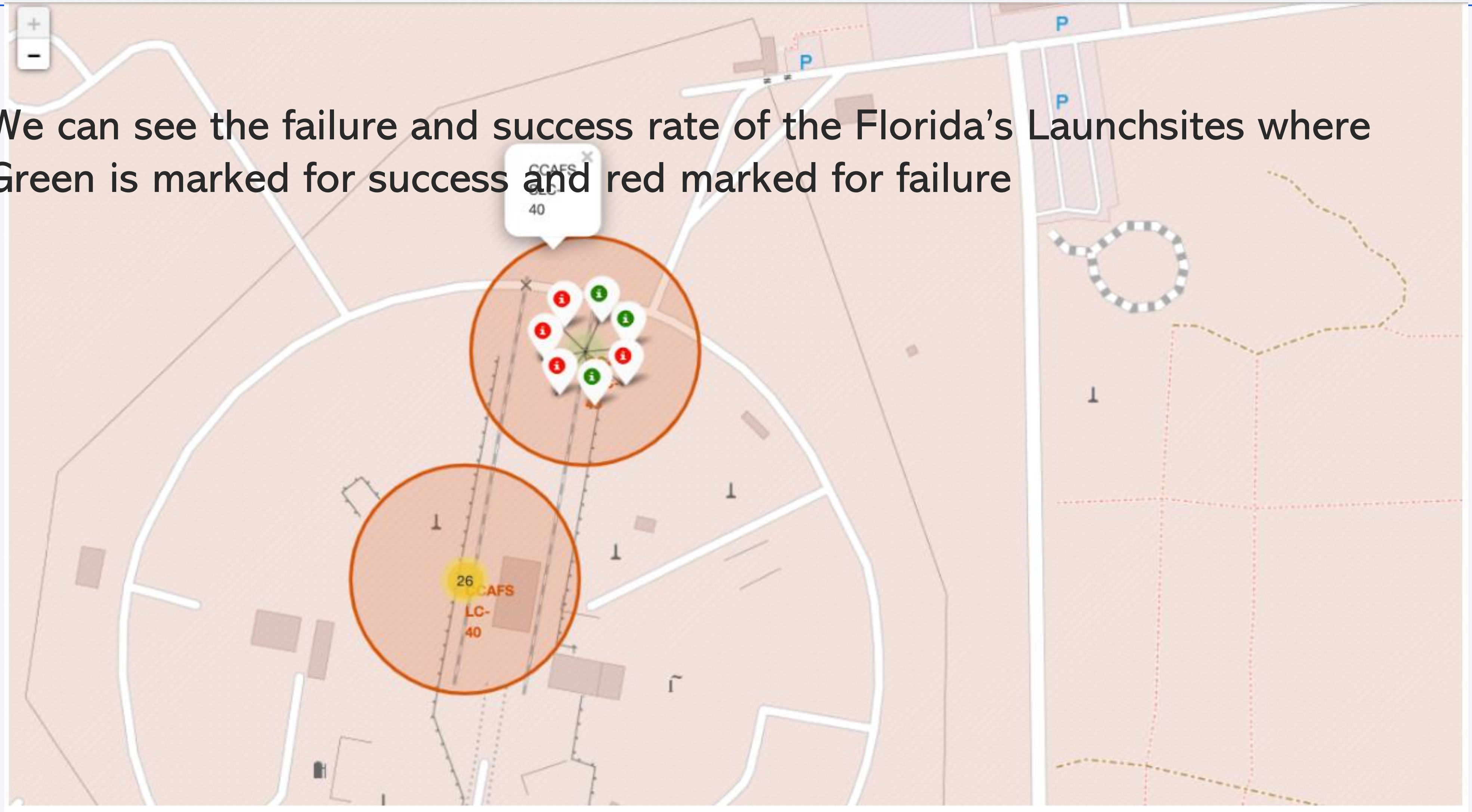
Launch Sites Proximities Analysis

Map Markers

- We can see all of the Launch sites of SpaceX

Color Labels

We can see the failure and success rate of the Florida's Launchsites where Green is marked for success and red marked for failure



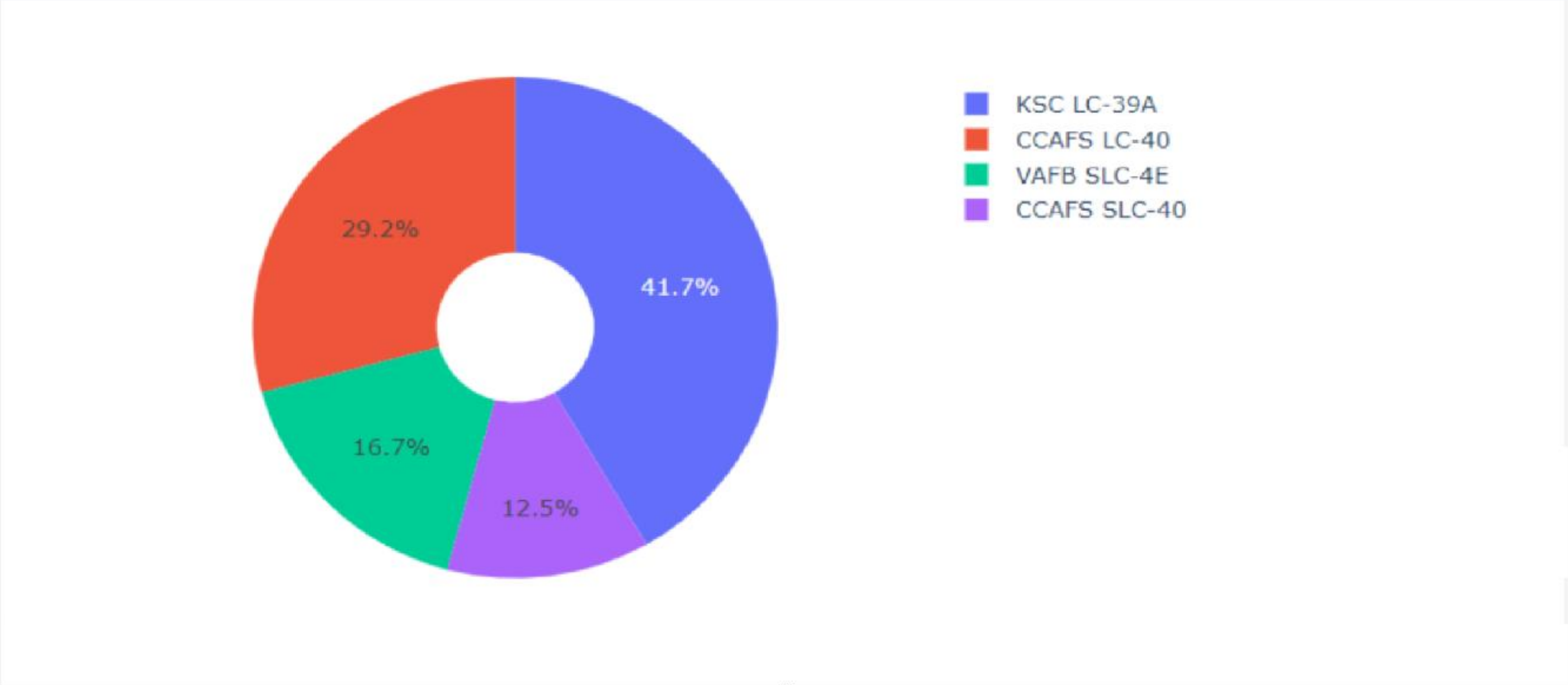
Landmarks near the Launch site

- Here we can see the landmarks near the launch site. Therfe is Railways, Highways, City, and Coastline

Section 4

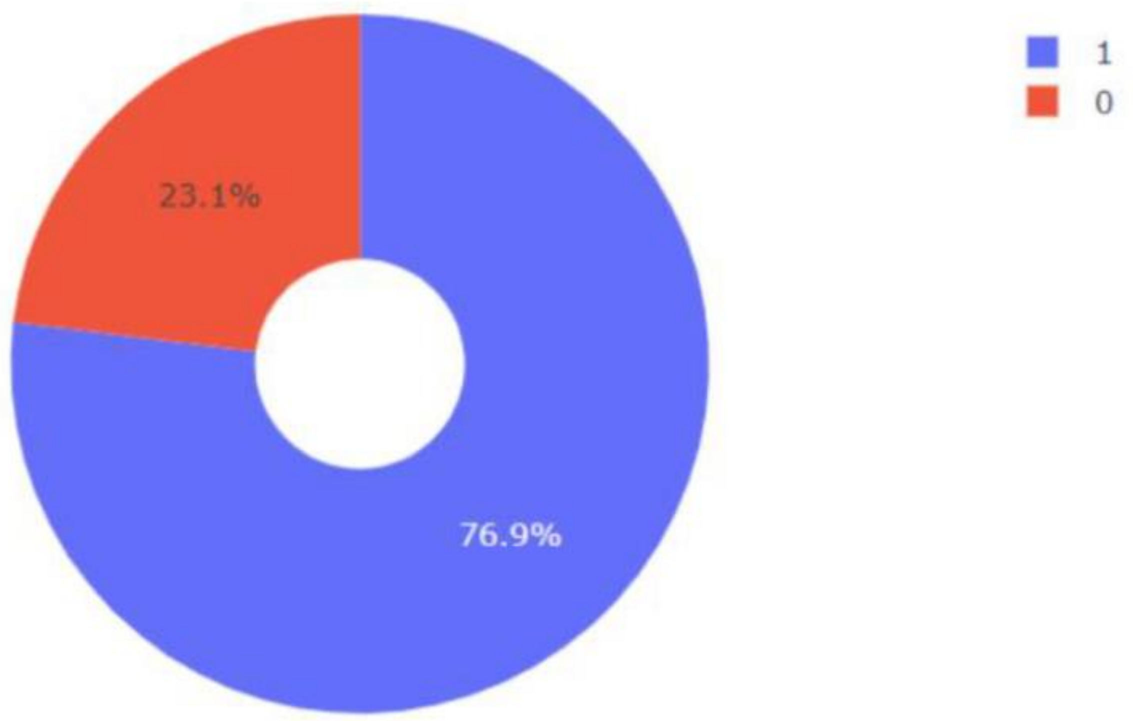
Build a Dashboard with Plotly Dash

Pie Chart of the Successful Launch Rate

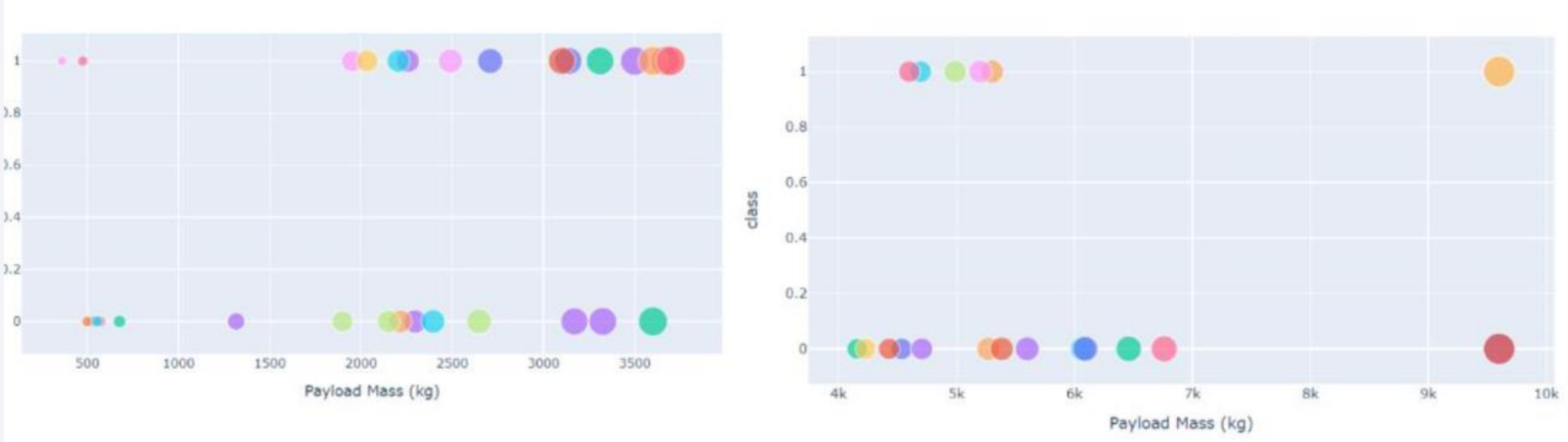


Pie Chart of the launch site with the highest success rate

- KSC LC-39A has the highest success rate



Payload vs Launch Outcome Scatter Plot



- We can see the success rate for lower payload is higher than higher payload

Section 5

Predictive Analysis (Classification)

Classification Accuracy

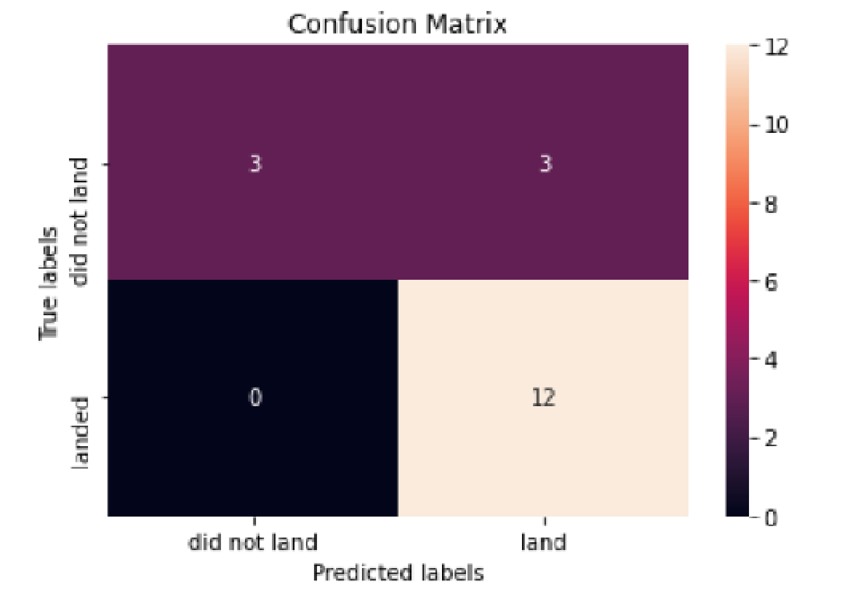
```
models = {'KNeighbors':knn_cv.best_score_,
          'DecisionTree':tree_cv.best_score_,
          'LogisticRegression':logreg_cv.best_score_,
          'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.8732142857142856
Best params is : {'criterion': 'gini', 'max_depth': 6, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split': 5, 'splitter': 'random'}
```

- Decision tree is the highest accuracy

Confusion Matrix

- This Matrix shows the classifier can distinguish between different classes of success rates of the launch site.



Conclusions

- SpaceX is improving their launch success rates as seen from 2013-2020 with increasing in launch rates.
- Site KSC LC-39A has the most launch success rate
- Decision tree classifier is the most accurate ML method

Thank you!