CAPSTONE PROJECT DE DATA SCIENCE

Fawad Awan

4/18/23

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Web Scraping
- Data analysis
- Data visualization
- Machine learning

INTRODUCTION

- Web Scraping of SpaceX's web API and created a data frame out of it.
- Data analysis based on the wrangled data
- Data visualization to find the relationships b/w different params
- Machine learning best algorithm’s determination.

METHODOLOGY:

The following python libraries were used:

- Request
- BeautifulSoup4
- Pandas
- NumPy
- Matplotlib
- Seaborn
- Dash
- SkLearn

RESULTS

- The results are with their respective topic name.
- Starting from the web scraping to the machine learning.

Web scraping

```
headings = []
for key,values in dict(launch_dict).items():
    if key not in headings:
        headings.append(key)
    if values is None:
        del launch_dict[key]

def pad_dict_list(dict_list, padel):    "padel": Unknown word.
    lmax = 0    "lmax": Unknown word.
    for lname in dict_list.keys():
        lmax = max(lmax, len(dict_list[lname]))    "lmax": Unknown word.
    for lname in dict_list.keys():
        ll = len(dict_list[lname])
        if  ll < lmax:    "lmax": Unknown word.
            dict_list[lname] += [padel] * (lmax - ll)    "padel": Unknown word.
    return dict_list

pad_dict_list(launch_dict,0)

df = pd.DataFrame(launch_dict)
df
```

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	<generator object Tag._all_strings at 0x7f1a1d...	Success\n	F9 v1.0	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	<generator object Tag._all_strings at 0x7f1a1e...	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	<generator object Tag._all_strings at 0x7f1a1e...	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	<generator object Tag._all_strings at 0x7f1a1e...	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	<generator object Tag._all_strings at 0x7f1a1d...	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10
...	...	...	...	...	...	...	...	...	...	...	...
101	102	CCSFS	SXM-7	7,000 kg	GTO	<generator object Tag._all_strings at 0x7f1a1d...	Success\n	F9 B5 △	Success	13 December 2020	17:30:00
102	103	KSC	NROL-108	C	LEO	<generator object Tag._all_strings at 0x7f1a1d...	Success\n	F9 B5 △	Success	19 December 2020	14:00:00
103	104	CCSFS	Türksat 5A	3,500 kg	GTO	<generator object Tag._all_strings at 0x7f1a1d...	Success\n	F9 B5	Success	8 January 2021	02:15
104	105	KSC	Starlink	15,600 kg	LEO	<generator object Tag._all_strings at 0x7f1a1d...	Success\n	F9 B5B1051.8	Success	20 January 2021	13:02
105	106	CCSFS	Transporter-1	~5,000 kg	SSO	0	0	F9 B5B1058.5	0	24 January 2021	15:00

Data analysis

```
df.head(8)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	0	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	0	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	0	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	0	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	0	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	0	1	False	False	False	NaN	1.0	0	B1005	-80.577366	28.561857	0
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	1	1	False	False	True	NaN	1.0	0	B1006	-80.577366	28.561857	1
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	1	1	False	False	True	NaN	1.0	0	B1007	-80.577366	28.561857	1

EDA with SQL:

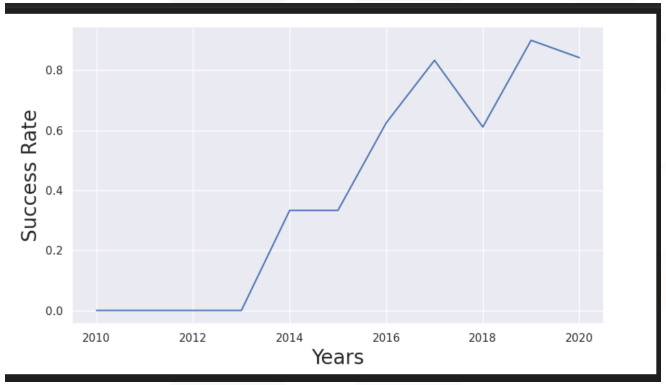
```
%sql select * from spacextbl where launch_site like 'cca%' limit 5;    "spacextbl": Unknown word.
```

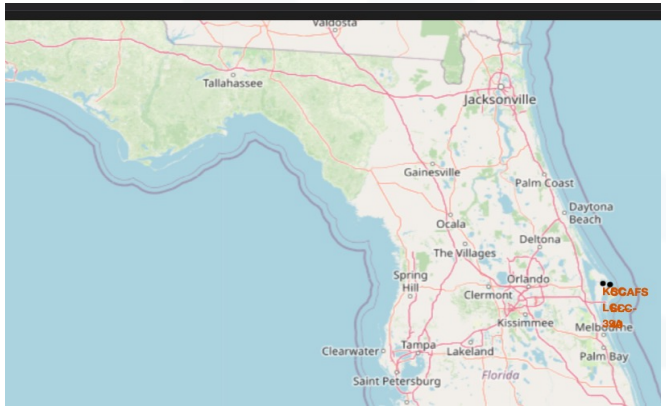
```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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	Failure (parachute)
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
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

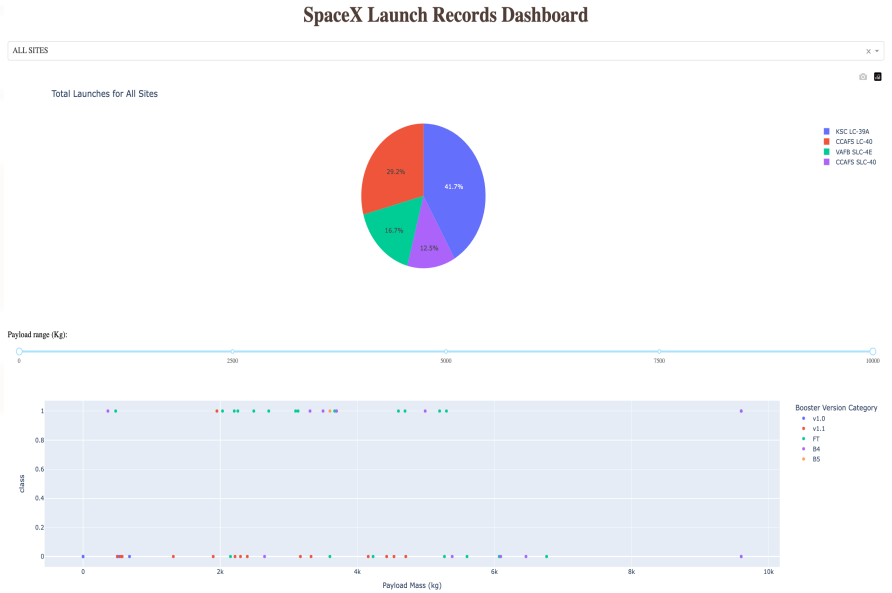
Data visualization

- Matplotlib, Seaborn & Folium:





- Dash & plotly:



Machine learning best algorithm's determination:

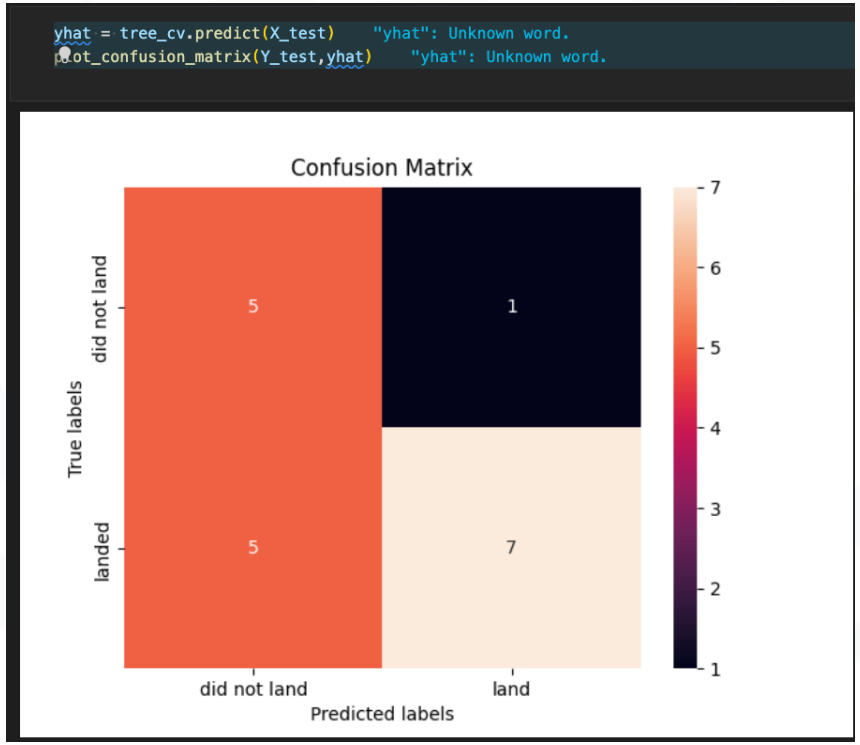
```
print('Accuracy for Logistics Regression method:', logreg_cv.score(X_test, Y_test))
print('Accuracy for Support Vector Machine method:', svm_cv.score(X_test, Y_test))
print('Accuracy for Decision tree method:', tree_cv.score(X_test, Y_test))
print('Accuracy for K nearsdt neighbors method:', knn_cv.score(X_test, Y_test))
```

```
Accuracy for Logistics Regression method: 0.8333333333333334
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.6666666666666666
Accuracy for K nearsdt neighbors method: 0.8333333333333334
```

ML Classification algos:

The tree classification technique had the least of the score (0.66), so it's confusion matrix based on the pre-determined training and testing data acquired by train-test-split is as under:

```
yhat = tree_cv.predict(X_test)    "yhat": Unknown word.
plot_confusion_matrix(Y_test,yhat)    "yhat": Unknown word.
```



DISCUSSION

None

OVERALL FINDINGS & IMPLICATIONS

Findings

- Decision tree's score is way less than the other classification algorithms' scores

Implications

- The prediction whether the launch is going to be or not, can't be a decision with multiple option but rather an event.

CONCLUSION

Given in the finding and implications