SDAIA ACADEMY

5 DEC 2021 |

Saudi Stock Market

INTRODUCTION

Using data science in the stock market is not new, but that doesn't apply for Saudi Stock Exchange (Tadawul), It needs to be explored and studied deeply, so we can cluster companies based on its behavior during the good and bad days.

Also, we can identify the days with a very large number of trades and try to understand the reason behind it.

Finally, we can predict the stocks prices.

DATA

The dataset is in csv format

It contains 600,000 rows, each rows and 14 features.

This dataset can be found at Kaggle

Data features

- **Symbol (Integer):** The symbol or the reference number of the company.
- **Name (String):** Name of the company.
- **Trading_Name (String):** The trading name of the company.
- **Sectoer (String):** The sector in which the company operates.
- **Date (Date):** The date of the stock price.
- **Open (Decimal):** The opening price.
- **High (Decimal):** The highest price of the stock at that day.
- **low (Decimal):** The lowest price of the stock at that day.
- **Close (Decimal):** The closing price.
- **Change (Decimal):** The change in price from the last day.
- **Perc_Change (Decimal):** The percentage of the change.
- **Volume_Traded (Decimal):** The volume of the trades for the day.
- **Value_Traded (Decimal):** The value of the.
- **No_Trades (Decimal):** The number of trades for the day.

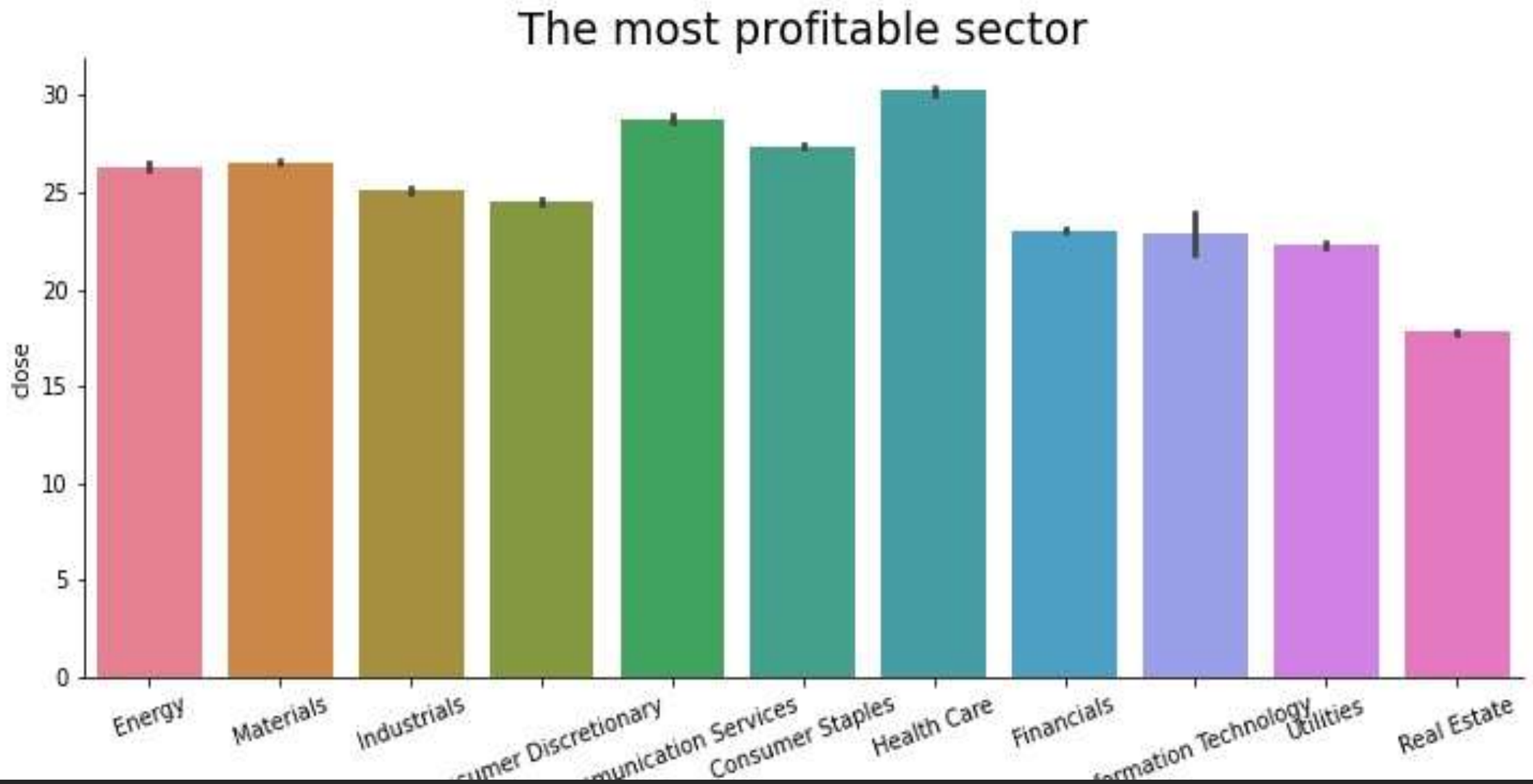
GOALS

I. predict the stocks prices

II. What is the highest Profit section?

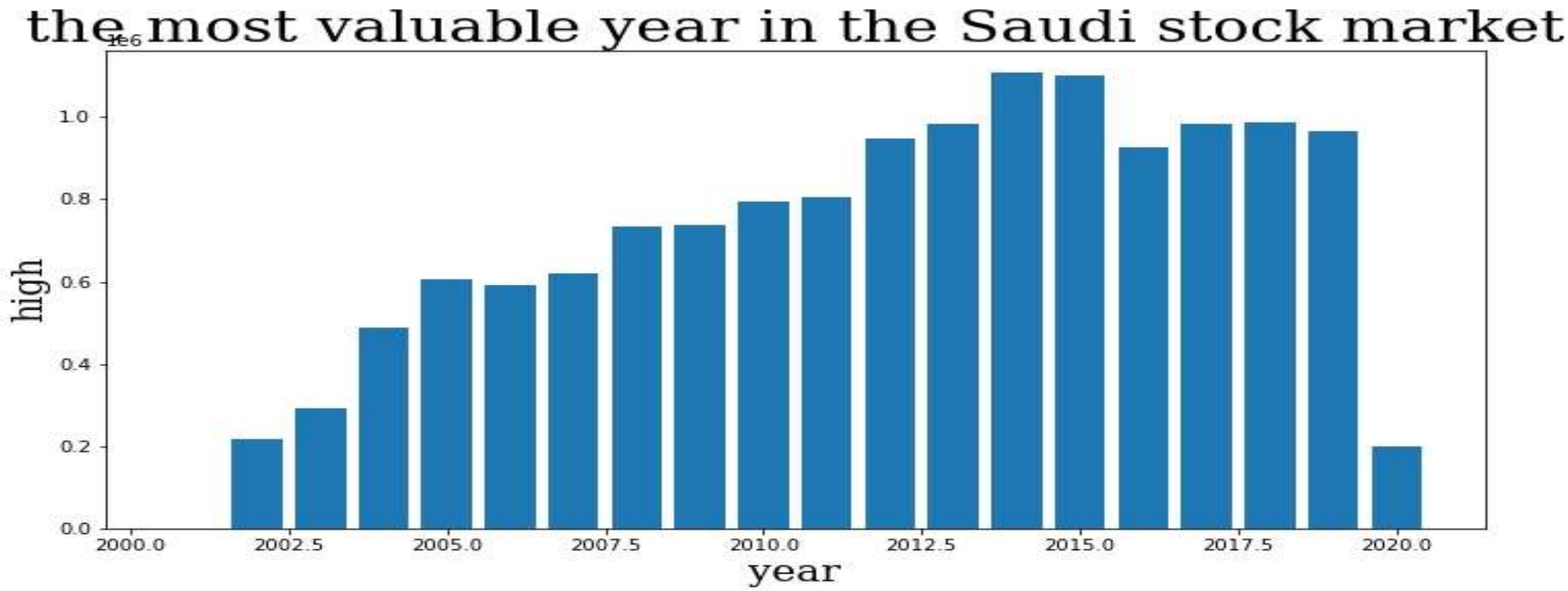
III. What is the less and most valuable year in the Saudi Stock market ?

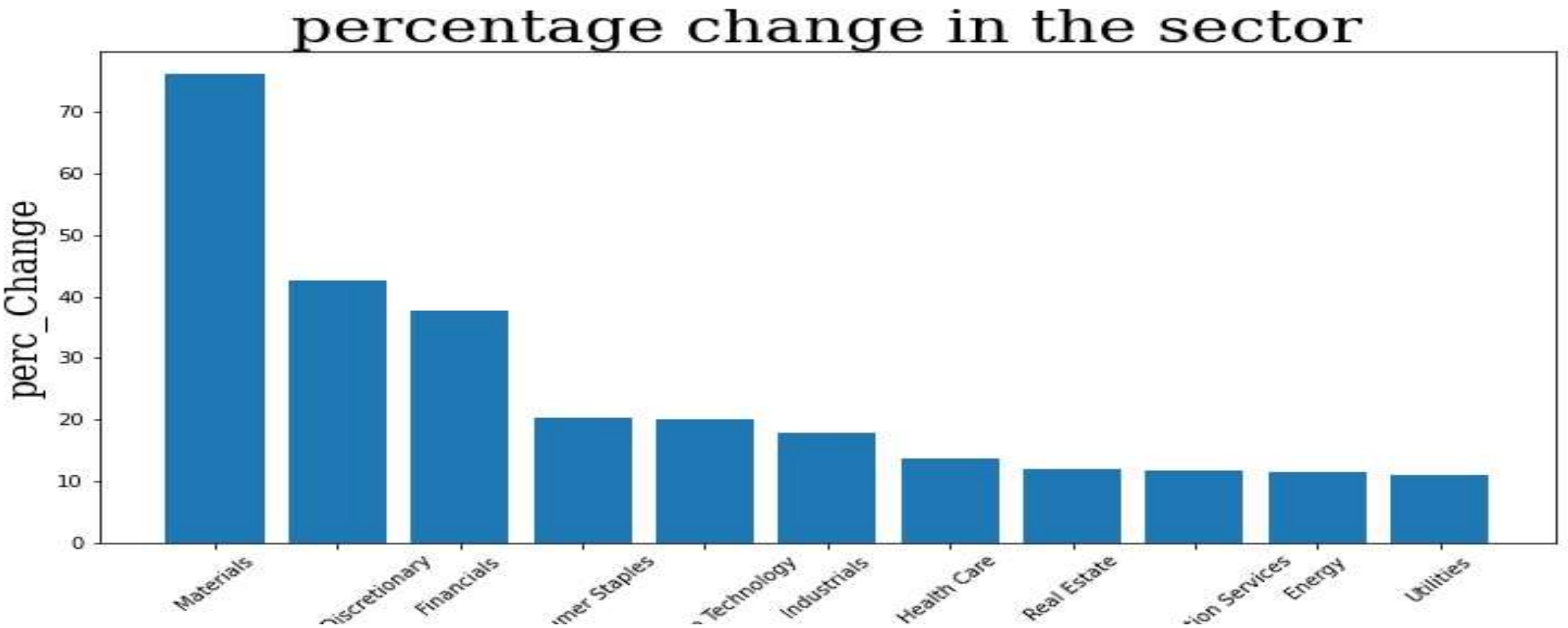
IV. What is the percentage change in the sector, and what is the highest sector ?



What is the highest Profit section

What is the less and most valuable year in the Saudi Stock market



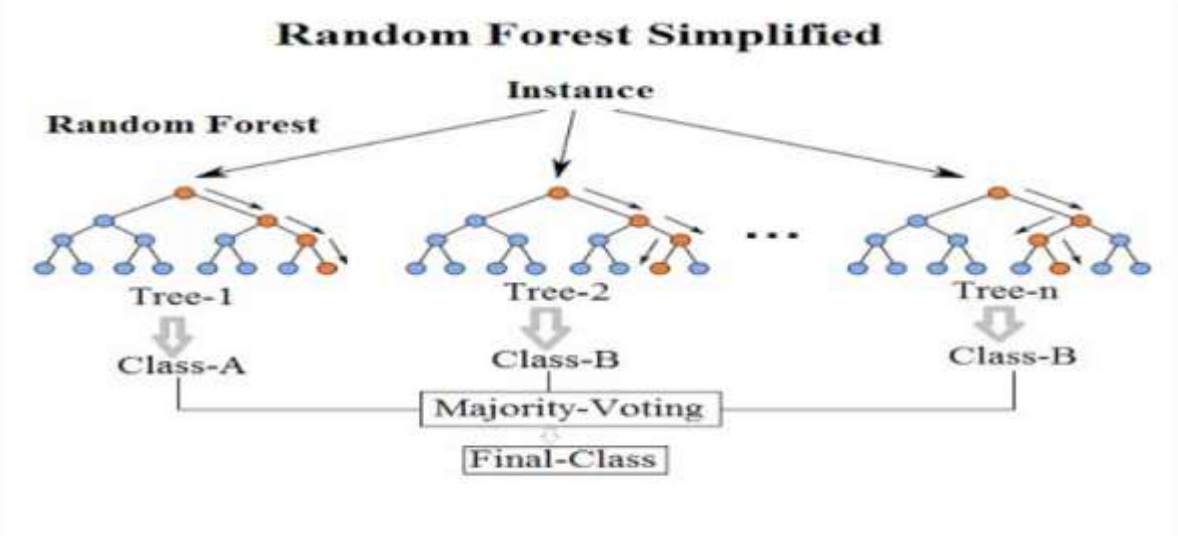

percentage change in the sector
perc_Change
0
10
20
30
40
50
60
70
Materials
Discretionary
Financials
mer Staples
Technology
Industrials
Health Care
Real Estate
ion Services
Energy
Utilities

Predict The Stocks Prices

1. **Linear Regression**

A solution to avoid overfitting is using decision trees.

2. **Random Forest Model Evaluation**



	prediction
0	32.325453
1	31.855085
2	32.747660
3	33.355873
4	33.343089
5	33.463612
6	34.221564
7	34.852940
8	35.432032
9	35.018065

```
1  y_train.head(10)
```

```
19895    31.15
5956     34.03
24702    10.26
6587     22.00
2553     49.00
31974    16.49
2863     24.30
9644     11.80
2067     24.30
26800    30.70
Name: close, dtype: float64
```

	prediction
0	32.004659
1	35.020005
2	10.972651
3	22.672885
4	47.149408
5	17.180586
6	25.139758
7	12.476548
8	25.138601
9	31.360030

```
1  y_test.head(10)
```

```
0    31.60
1    31.35
2    31.70
3    32.45
4    32.70
5    32.20
6    33.10
7    33.60
8    34.25
9    34.35
Name: close, dtype: float64
```

TOOLS

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Sklearn preprocessing
- Sklearn model selection
- LinearRegression
- DecisionTreeRegressor
- Mean squared error
- Jupyter notebook

Thank for your time

ANY QUESTIONS