Stocks’ pricing

Introduction about data set

- Daily trade data for around 200 companies listed in the Saudi Arabia exchange market (Tadawul) for the last 4 years.
- Financial data for these companies including the balance sheet and cash flow statements.

How many rows and columns and where you find your data :

The data set is expected to have around 200,000 (1,000 for each company) rows with 50 columns.

What are the questions that you are interested to answer :

- What are the variables impacting the change in stock price?
- How could we utilize the historical **trade** data to estimate the stock price in the future?
- How could we utilize the historical **financial** data to estimate the stock price in the future?

Which tools you will use in this project :

- Colab, pandas, numpy, seaborn, matplotlib, sklearn

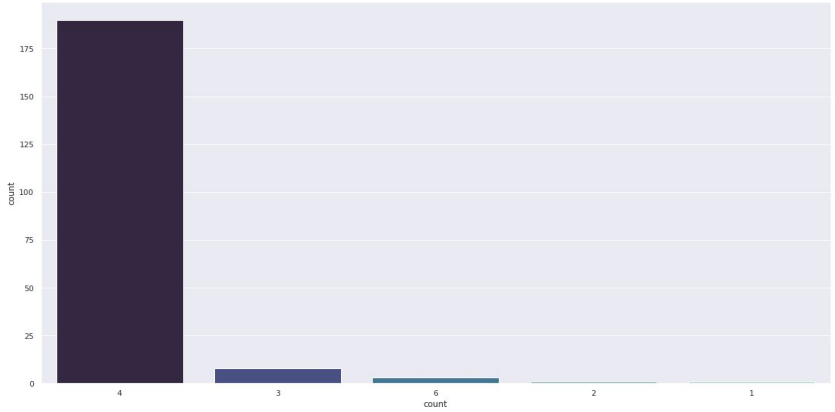
What you need to do to achieve the goal :

- Understanding the trading mechanism in Tadawul.
- Understanding basic finance for companies
- Merging, exploring , cleaning and transforming the data

Findings:

~120 companies have more than 1000 day trade data

~180 companies have 4 years financial data



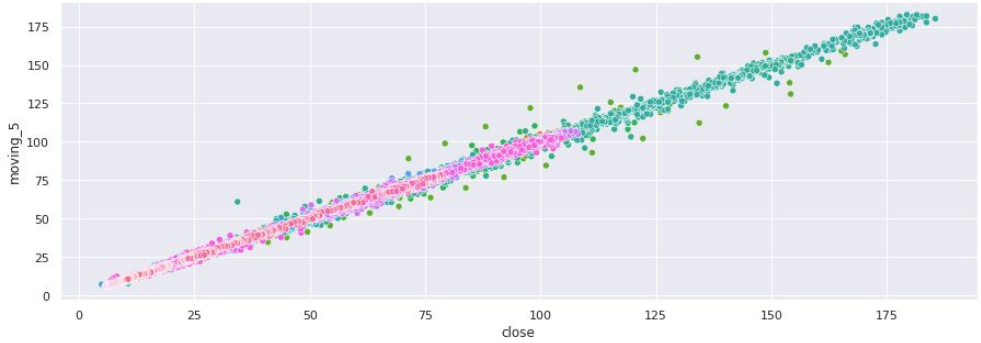
Findings:

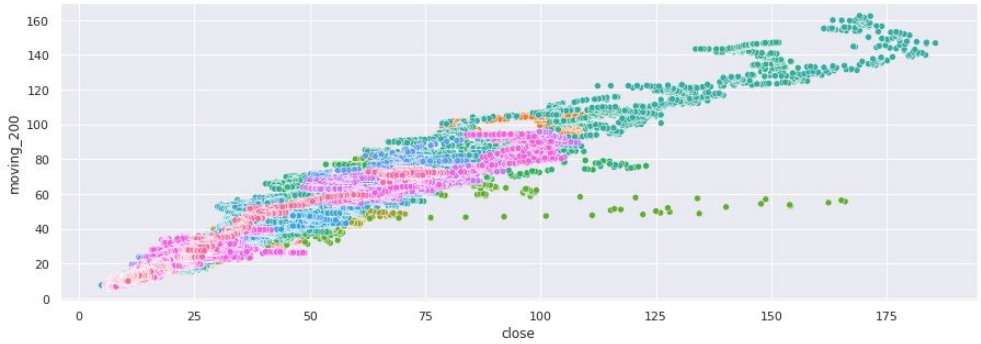
Close price mostly affected by day trade data and balance sheet data

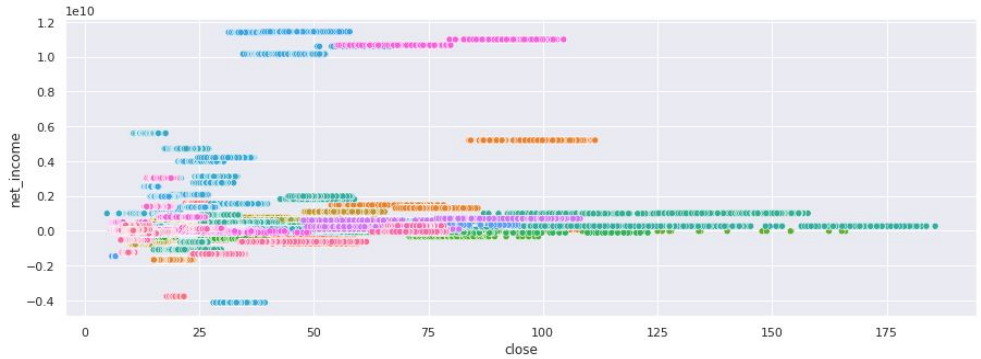
Feature	Correlation
high	0.9998
low	0.9998
open	0.9997
moving_5	0.9995
moving_10	0.9986
moving_100	0.9882
moving_200	0.9814
retained_earnings	0.2975
inventory	0.2673
net_income	0.2370
minority_interest	0.2314
deferred_long_term_asset_charges	0.2149
cash	0.2091
dividends_paid	-0.3416

Findings:

Close price correlation varies across the different companies, specifically for the financial data







Models:

Different Models have been created to predict the stock price

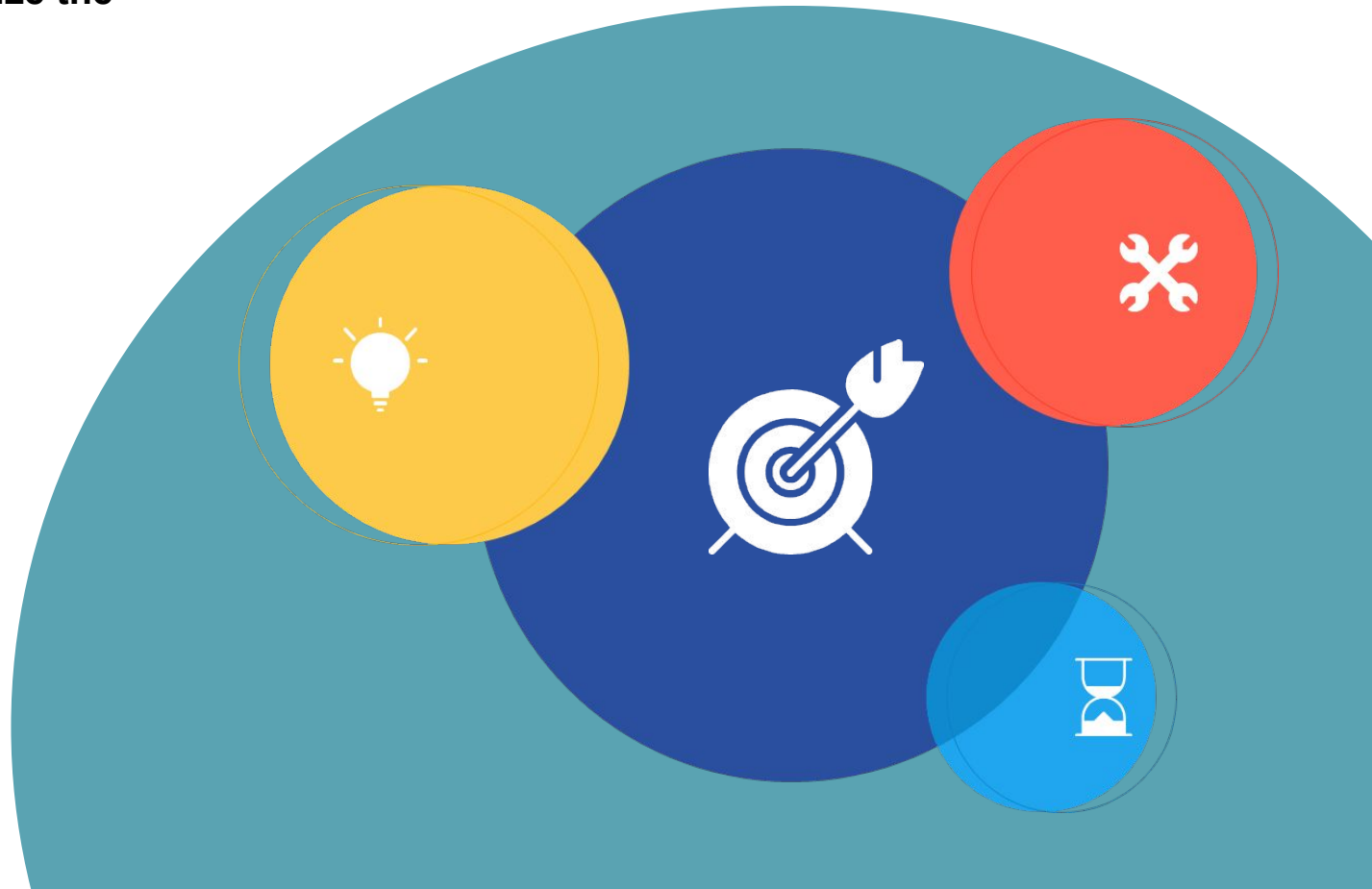
Features	All		
Model Name	LinearRegression	Lasso	Ridge
Absolute percentage error	0.96%	3.71%	0.96%

Features	High Correlated (All)		
Model Name	LinearRegression	Lasso	Ridge
Absolute percentage error	0.95%	3.71%	0.95%

Features	High Correlated (Financial)		
Model Name	LinearRegression	Lasso	Ridge
Absolute percentage error	72.09%	73.31%	72.09%

Recommendation:

- **Linear regression model is better to value the stock price using the both day trade and financial data**
- **Further toning to the model could be done to minimize the impact of the velocity of day trade data.**



Thank You!