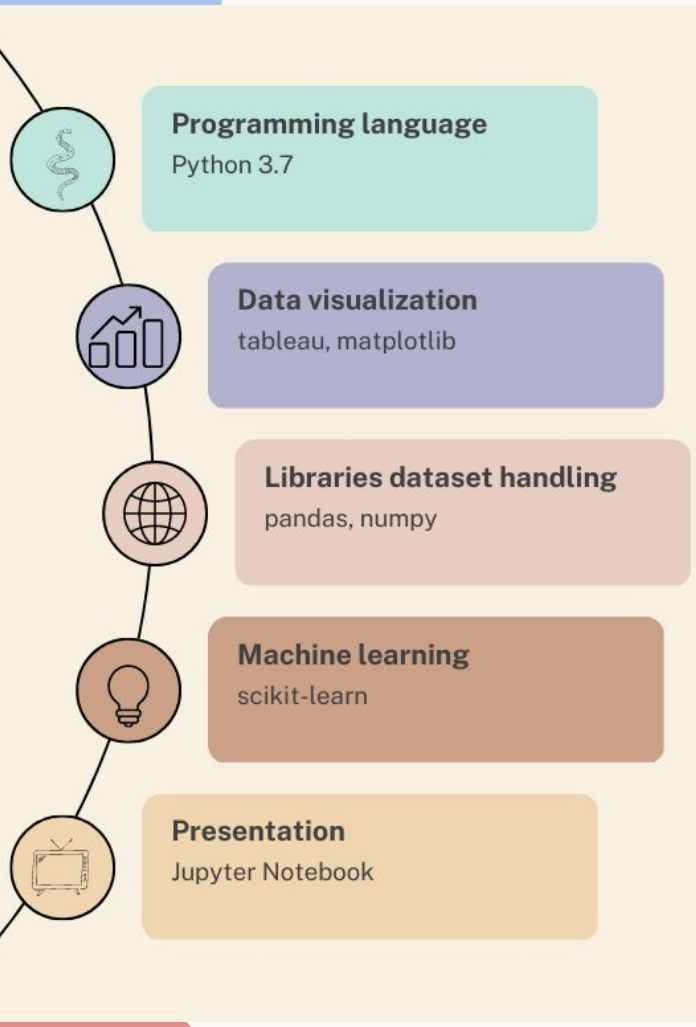
STOCK ANALYSIS

Seemab Haider
Shreya Akotiya
Mia Shi
James Qiu
Anjali Skilton

GOALS

- To understand the effect of stock prices on trading volume for different business sectors.
- What is the correlation between stock sectors and cryptocurrency?

WHAT WE USED



Programming language
Python 3.7

Data visualization
tableau, matplotlib

Libraries dataset handling
pandas, numpy

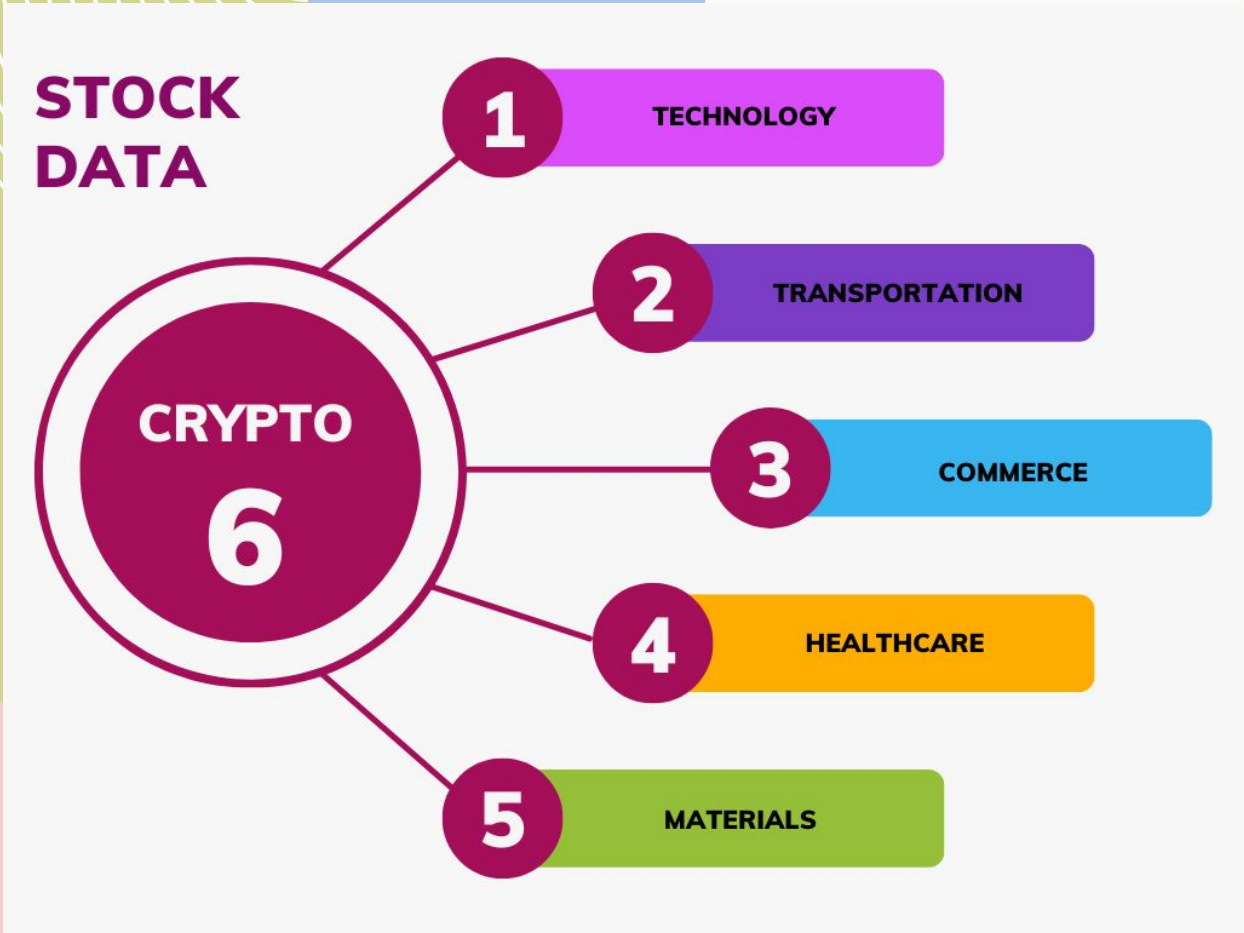
Machine learning
scikit-learn

Presentation
Jupyter Notebook

Background & Purpose Overview

- **Why we chose these 5 industries for our research?**
  - Technology industry developed greatly over last decades and dramatically impact on life styles, e.g. apple watch starts to check health!
  - Transportation industries refers to industries in the for-hire transportation and warehousing sector, such as air, rail, water, and truck transportation.
  - Consumer Cyclical & Commerce Consumer cyclicals are a category of stocks that rely heavily on the business cycle and economic conditions. Consumer cyclicals include industries such as automotive, housing, entertainment, and retail.
  - Healthcare The healthcare industry is an aggregation and integration of sectors within the economic system that provides goods and services to treat patients with curative, preventive, rehabilitative, and palliative care.
  - The Materials Sector encompasses a wide range of commodity-related manufacturing industries. Included in this sector are companies that manufacture chemicals, construction materials, glass, paper, forest products and related packaging products, and metals, minerals and mining companies, including producers of steel.

- **Relationship with the macroeconomic**



DESCRIPTION OF DATA

PROCESS

1 DATA EXPLORATION

ANALYSIS 2

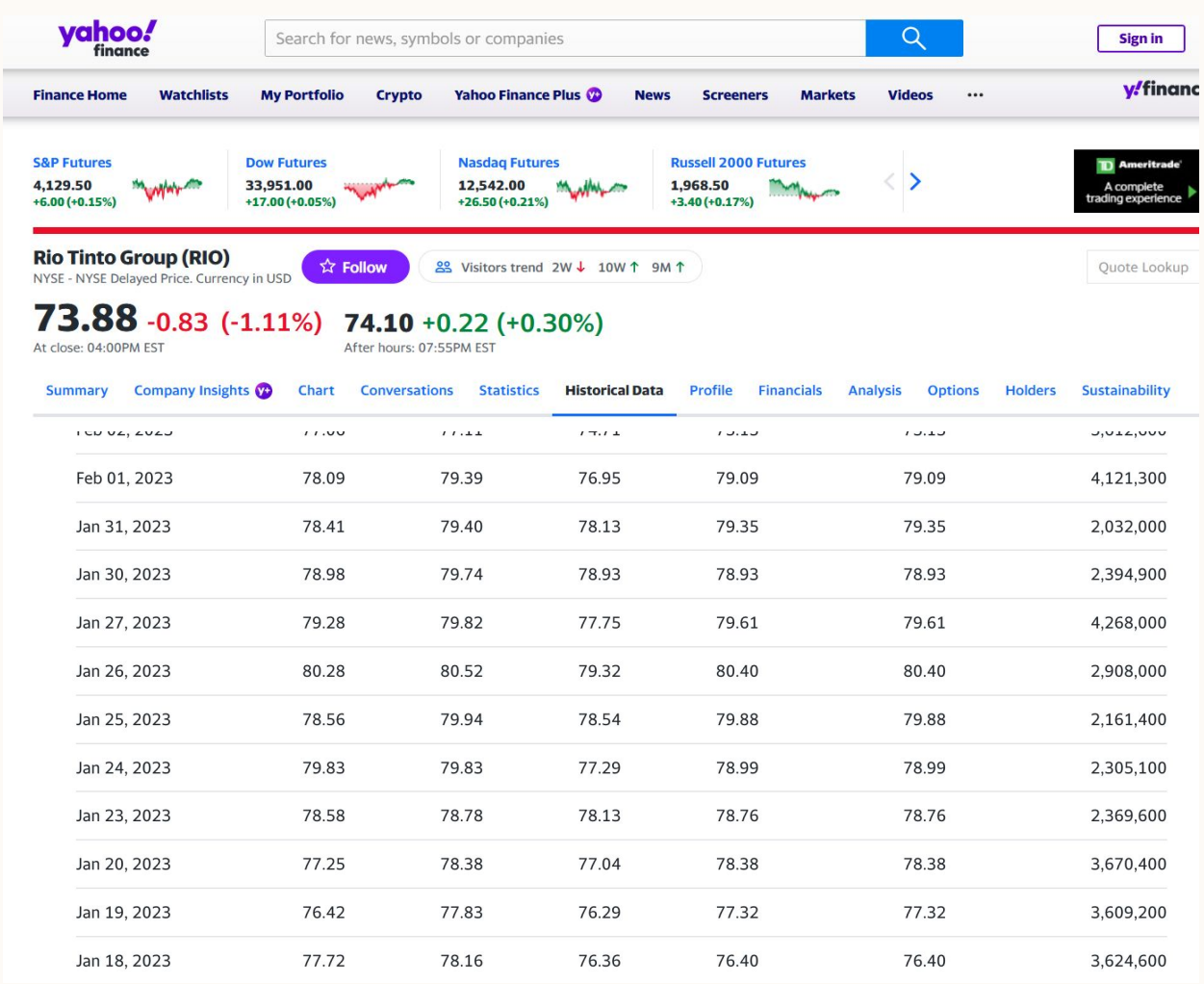
3 DATA VISUALIZATION

CONCLUSION 4

DATA EXPLORATION

- Data was collected from yahoo finance https://finance.yahoo.com/
- Historical data of 5 different sectors and Bitcoin were collected.
- 10 different companies of similar market cap were selected from each sector, resulting in 51 different stocks.
- Time span of 10 years from 2013-2023
- Overall more than 1 million rows.

Data Source



Feb 01, 2023	78.09	79.39	76.95	79.09	79.09	4,121,300
Jan 31, 2023	78.41	79.40	78.13	79.35	79.35	2,032,000
Jan 30, 2023	78.98	79.74	78.93	78.93	78.93	2,394,900
Jan 27, 2023	79.28	79.82	77.75	79.61	79.61	4,268,000
Jan 26, 2023	80.28	80.52	79.32	80.40	80.40	2,908,000
Jan 25, 2023	78.56	79.94	78.54	79.88	79.88	2,161,400
Jan 24, 2023	79.83	79.83	77.29	78.99	78.99	2,305,100
Jan 23, 2023	78.58	78.78	78.13	78.76	78.76	2,369,600
Jan 20, 2023	77.25	78.38	77.04	78.38	78.38	3,670,400
Jan 19, 2023	76.42	77.83	76.29	77.32	77.32	3,609,200
Jan 18, 2023	77.72	78.16	76.36	76.40	76.40	3,624,600

FLOWCHART

10 TICKERS TO ONE CSV

NORMALIZE HIGH, LOW, AND VOLUME

MACHINE LEARNING PREDICTION

MODELING AND VISUALIZATION

COMPARISION

ANALYSIS

Multi Linear Regression

- Multi Linear Regression Model was used for 6 different sectors for building statistical models that characterize relationship among two features “High price” and “Low price” of Stocks with dependent variable in this case “Volume Trading”.
- Model is validated by looking at it’s coefficient of determination(R2).
- We rebuild our model using OLS() function and provided model summary

OLS REGRESSION AND R2 SCORE

OLS Regression Results

Dep. Variable:	Volume	R-squared:	0.453
Model:	OLS	Adj. R-squared:	0.452
Method:	Least Squares	F-statistic:	930.8
Date:	Fri, 03 Feb 2023	Prob (F-statistic):	2.41e-295
Time:	18:44:28	Log-Likelihood:	1830.6
No. Observations:	2255	AIC:	-3655.
Df Residuals:	2252	BIC:	-3638.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.5481	0.005	103.068	0.000	0.538	0.559
High	8.5450	0.266	32.086	0.000	8.023	9.067
Low	-8.8182	0.266	-33.145	0.000	-9.340	-8.296

Omnibus:	523.810	Durbin-Watson:	0.967
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1534.207
Skew:	1.188	Prob(JB):	0.00
Kurtosis:	6.268	Cond. No.	186.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

]:

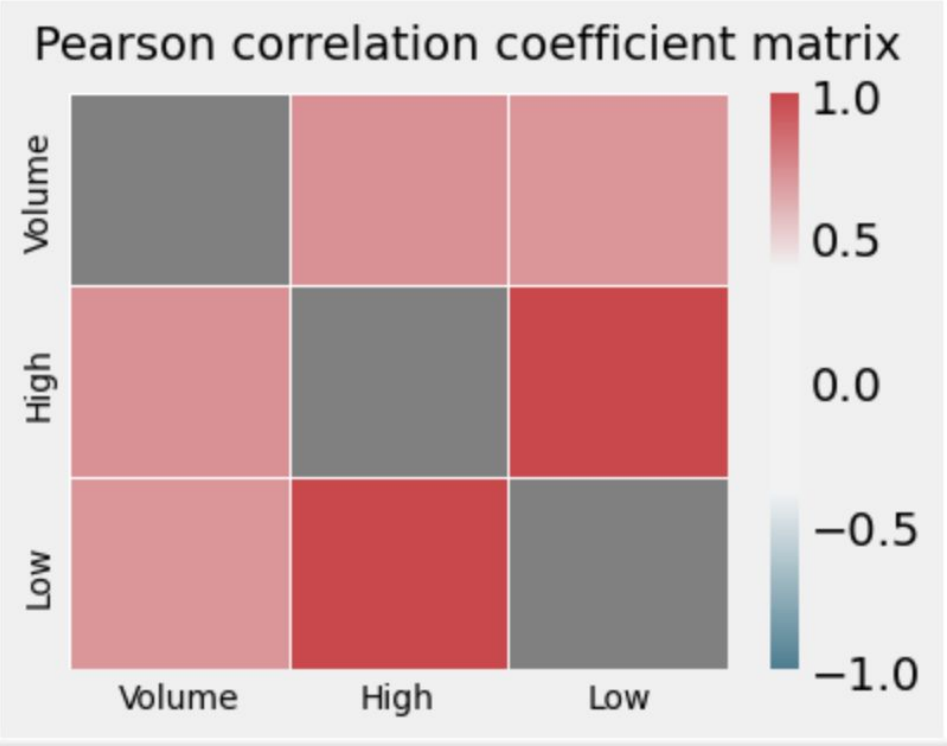
```
# R squared sroce. If equals 1 then perfect correlation between independent and dependent.
# If 0 no correlation.

print('R2 score:', olsmod.rsquared)
```

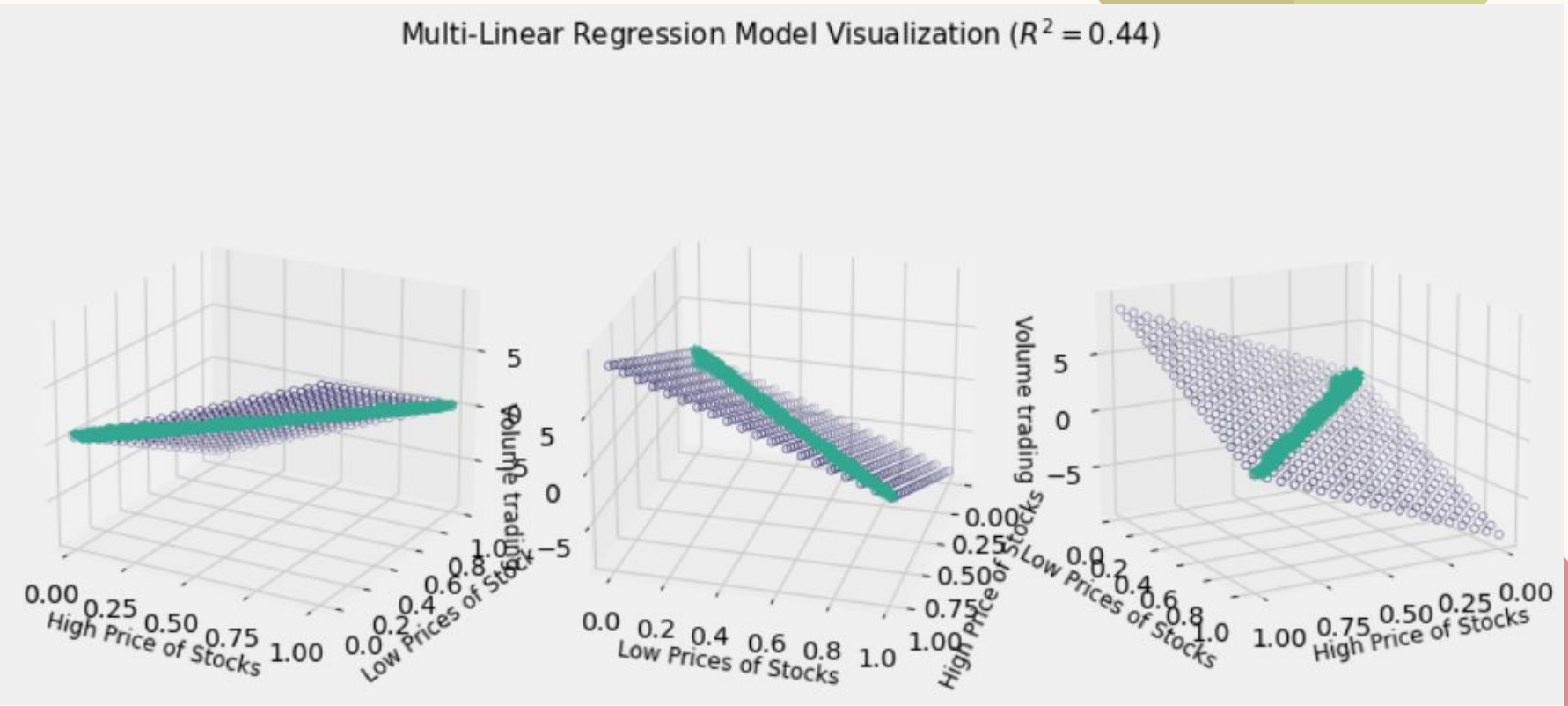
```
R2 score: 0.4525440917539122
```

Pearson Correlation Coefficient Matrix

	Volume	High	Low
Volume	1.000000	0.720484	0.708519
High	0.720484	1.000000	0.999007
Low	0.708519	0.999007	1.000000



Scatter Plot



Data Visualization

OVERALL FINDINGS

1. Based on the model, we can see that price change in stocks affects the volume trading mostly in two sectors that include Consumer and Bitcoin(Cryptocurrency)
2. Technology and Healthcare sectors are the least affected by stock price changes.
3. Since we have on average under 50% correlation from our predictive model, we find that outside factors also affect the volume trading.

Recommendations

- For the future analysis, it's always good to compare different models, e.g. we may use the normal or lognormal distribution model for prediction.
- We may also improve the machine learning model, e.g. increasing the layers to increase the accuracy so that better predict the price trendency.
- We may try to use a supervised machine learning model such as the Random Forest Classifier combining a multitude of decision trees to predict the stock price and compare the accuracy.

THANK YOU!