Speaker Notes

**Tori**

Stock Analysis VS US Economic Factors

Team Members: Alex, Zahra, Misa/Celine, Olympia and myself

Purpose:

The purpose of this analysis is to look at how US macro economic factors impact stock prices.  The topic was selected because team members are interested in which industries are the most affected by various economic variables.

Objective:

Using a basic neural network model for Machine Learning along with historical stock and selected economic data, the economic factors are analyzed for their affect within each chosen industry sector.  The objective is to predict which industries may be the most stable for investment purposes.

Question: Is there any industry that is more or less affected by the economy?

Data Sources: OECD website, and YFinance

**Olympia**

On this slide our summary of how our process flow:

ETL  => Database =>  Machine Learning

Extract Transform Load comes in 2 parts. The first part is pulling out our raw data from yfinance and Wikipedia

Yfinance It's an open-source tool that uses Yahoo's publicly available APIs, and is intended for research and educational purposes

Wikipedia is a free online encyclopedia, created and edited by volunteers around the world and hosted by the Wikimedia Foundation.

Since our project is about Stock Trends VS US Economic Factors.

We decided the stock market we will use is S&P500.

S&P 500, is a stock market index tracking the performance of 500 large companies listed on stock exchanges in the United States. It is widely considered one of the best gauges of large U.S. stocks, and even the entire equities market.

=>  next slide please

Developed Python script to extract SP&500 for each ticker performance and Wikipedia for ticker information and then a created dataframe. We need Wikipedia information since it has the GCIS (Global Industry Classification Standard) Sector which has a total of 11 Sectors.

Afterwards,  transformed the data, by renaming column names, keeping columns we only need, removing null data, joining dataframes and creating a csv file.

**Zahra**

Once Olympia finished the initial clean up on the ticker data, I was then responsible for completing the final ETL process to prepare the tables to be loaded into our database.

To prepare the stock data, I used pandas and the random library to randomly select 12 tickers for each sector.  In thinking back to our question – we knew we wanted to categorize stocks in different sectors by their stability. To account for this I calculated the daily percentage change of the adjusted close price for each ticker in each sector. Since we know that the OECD data is available annually from 2013 – 2020, I then averaged each ticker’s percentage change for each year.

Regarding the OECD data  the original dataset contained over 100 economic statistics across 22 different categories. Understanding economic health is a complicated  However,  we narrowed down which features to use based on eliminating stats with null values, and by considering which broad categories and statistics we believe provide the best over of the US’ economic health throughout the years. In total we ended up with 14 features.

Once both tables were cleaned and ready to be transformed, I exported them as CSVs and created an AWS database. I connected the AWS database to PosgresSQL and created the appropriate tables for the stock data and the econ data, then imported the cleaned CSVs into the tables.

In Postgres and based on our ERD the stock table and econ table were joined using a left outer join on year to create a new table which was shared through the AWS endpoint for the machine learning model.

**Alex**

To predict whether a stock will be stable, we used a Basic Neural Network Machine Learning Model. US economic data such as the Federal Reserve Annual Interest Rates, Tax on Corporate Profits, Unemployment Rate, etc. was used in the Machine Learning Model to determine what features affect stock stability. We used these features in our model because, as previously stated, these factors greatly impact the US economy which then in turn should affect the stock market.

Before implementing our model, the data needed some preliminary preprocessing. First, we dropped the columns; “Year” and “Ticker”, because they are not necessary for predicting. Next, we encoded the categorical variable, which was the different sectors of stocks. Finally, we created a new column, the Is\_Stable column. This new column would indicate whether the daily percent change in stock price was less than 0.1%. Our research showed that a change of less than 0.1% is considered a stable stock.

Following our data preprocessing the model was ran using our new Is\_Stable column for predictions. Originally, it was overfitting, so through trial and error we decided as a team that the best way to stop over fitting was to lower the epoch count from 50 to 15. This is because the model stopped overfitting but still had a suitable accuracy score (80-90%) once the change was made.

Here we can see, our model loss and model accuracy change as the epochs ran.

**Celine**

In conclusion, upon analyzing the data and running machine learning, the top 3 stable sectors are real estate, consumer staples & energy. As looking at our visual charts, it tells these top 3 are the highest frequency of stable stocks which helps your decision-making for your investment. The outcome is based on the unemployment rates for 8 years in each sector which means these top 3 sectors are relatively most stable between minimum and maximum rates of unemployment that affect the company’s values. As looking at the charts of economic growth and Prices & Inflation Rates,  
When the Covid-19 hit the GDP growth fell, these top 3 sectors were fairly stable. Although the energy went down in 2016, the value of energy got back quickly and gradually went up comparing to other sectors. Our analysis based on 8 years historical sectors and our machine learning predicts the nearest future for your investment. Also, other thing to consider for investing, the average adjusted closing price will help to predict stock’s value since it helps to evaluate stock performance which helps investors understand how much they would’ve made by investing in a given asset as well as comparing the performance of two or more assets. Although there are a lot of unexpected economic issues like world war or factory problems, this analysis helps to see the close prospect in your investment. Thank you all!