BIZ-INSIGHT HUB

Your Destination for Exploring Corporate Performance,

Viewing Stock Prices, and Accessing Financial Metrics

RDB Final Project Group 7

2018310018 ChanHwi Jeong

2018312407 JungYong Son

2023319502 Zhu Xukun

Outline

1 Introduction of Our Database

2 About Our Data

3 GUI Introduction

4 Lesson & Learned

5 Q & A

Part 1

Introduction of Our Database

Introduction of Our Database

Our database provides :

- Relevant data on the financial and investment markets
- Investors with investment suggestions based on basic information on different stocks, companies, and industries

Target Users

1. Financial consultants

   Brief but sufficient company fundamental information can meet the needs of desk research

2. Novice investors

   Help investors choose suitable investment targets in different industries

3. Business analysts

   Identify the list of competitors in the same industry and quickly obtain basic data of the corresponding companies

Introduction of Our Database

Language & Software

Python: Pandas for reconstructing the database and handling, importing data
Tkinter for building GUI
Matplotlib for drawing the picture of stock price

MySQL: Workbench of our database

Stata: Used for regression by group(prediction part)

Excel: Row data cleaning

Main Function

- Obtain company's financial data and stock price trends
- Select the industry to obtain the ranking of stocks with the most investment value
- Query the basic information of the industry and the overview of the model used for estimation
- Express their opinions on stocks, and can also see other users' opinions of stocks

Part 2

About Our Data

About Our Data

- Raw data from the CSMAR database: in CSV format, and can be imported into the database through python after simple processing

- CSMAR database is the first domestic economic and financial database developed by Shenzhen CSMAR Data Technology Co., Ltd.(Commercial)

- Covers 19 major series such as character characteristics, bank research, stock market, and company research

- When users have data needs, they can manually search for data in the system (when the amount of data is small), or send a complete list of needs to the company. The data specialist will process the needs as soon as possible and send a rough form to the user via email

Data Cleaning

- **Missing data handling**

Examined each company's financial statements to determine how missing values should be filled in:

If the company does not have the column value (for example, a retail company has no R&D expenses), the value will be given 0.(A vacant value inside the annual report means that no transactions or business activities related to this accounting item were generated, so 0 is desirable)

If the company fails to report in time, the average value in this industry will be used instead (This is because the financial statements of each listed company are required to be published at a unified time. It is possible that some companies have not had time to calculate all the financial data)

These missing values are data from the company's financial statements, which are completely different from the data used for prediction, which are data from the company's stock information, such as the opening price, the closing price, and so on

→ Errors in handling missing values in the financial statements do not affect the prediction results

Data Cleaning

- **Format standardization**

Ensure that all financial values follow a unified currency format and unit.

At the same time, the stock code and time format are processed uniformly to ensure the continuity and comparability of time series data.

Finally, the subsidiary data is removed from the table to ensure the identity of comparable companies

(The number of subsidiaries owned by different firms varies, making the data at the subsidiary level non-comparable, so we only keep the data for the parent firm)

- **Prediction (← click here to check the details of the prediction model)**

Prediction: About "**Fama-French 3 Factor Model**"

$$E(R_i) - R_f = \alpha + \beta_{i,M}(E(R_M) - R_f) + \beta_{i,SMB}E(SMB) + \beta_{i,HML}E(HML)$$

Firstly, Stocks are divided into two groups, Small and Big, according to market capitalization; and three groups, Value, Middle and Growth, according to Book-to-market ratio.

So we get six group(portfolio), After they are constructed, the returns of the six group are calculated each month(This month's value – last month's value), which are recorded as SH, SM, SL, BH, BM and BL. Then calculate the SMB factor and HML factor according to the following expressions:

		Value		
		Value (top 30%)	Middle (middle 40%)	Growth (bottom 30%)
Size	Small (bottom 50%)	S/V	S/M	S/G
	Big (top 50%)	B/V	B/M	B/G

$$SMB = \frac{1}{3}(S/V+S/M+S/G) - \frac{1}{3}(B/V+B/M+B/G)$$

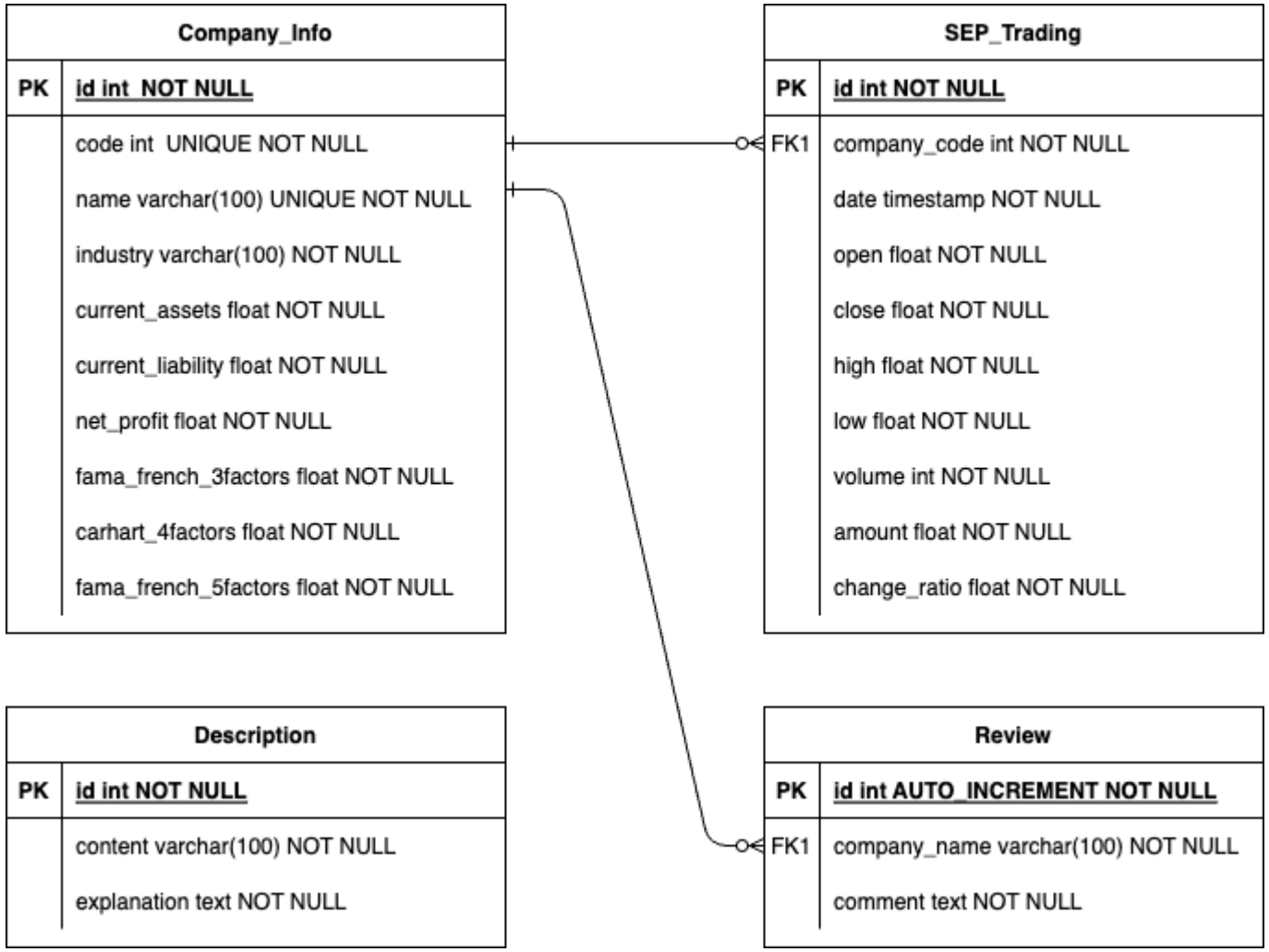
$$VMG = \frac{1}{2}(S/V+B/V) - \frac{1}{2}(S/G+B/G)$$

Secondly, we calculate $E(R_M) - R_f$: CSI 300 return rate - risk-free interest rate(10-year Treasury bond interest rate) and $E(R_i) - R_f$: each stocks return rate- risk-free interest .

(For each set of regressions, SMB, HML, $E(R_M) - R_f$ are same because they are calculate by all the stocks return rate. CSI300(like Chinese version of S&P500,) and R_f is nearly constant.)

Finally, For each stock, we collect its data for the last 30 days and calculate its return rate. And do regression to estimate parameter α,which indicates potential value. And sort by parameterα size in descending order.

Database ER Diagram



Columns for the presentation of the company‘s key information

→ Reflects the performance of the company's recurring business

→ Whether or not the company is operating on a regular basis and whether or not its main business is performing well

Part 3

GUI Introduction

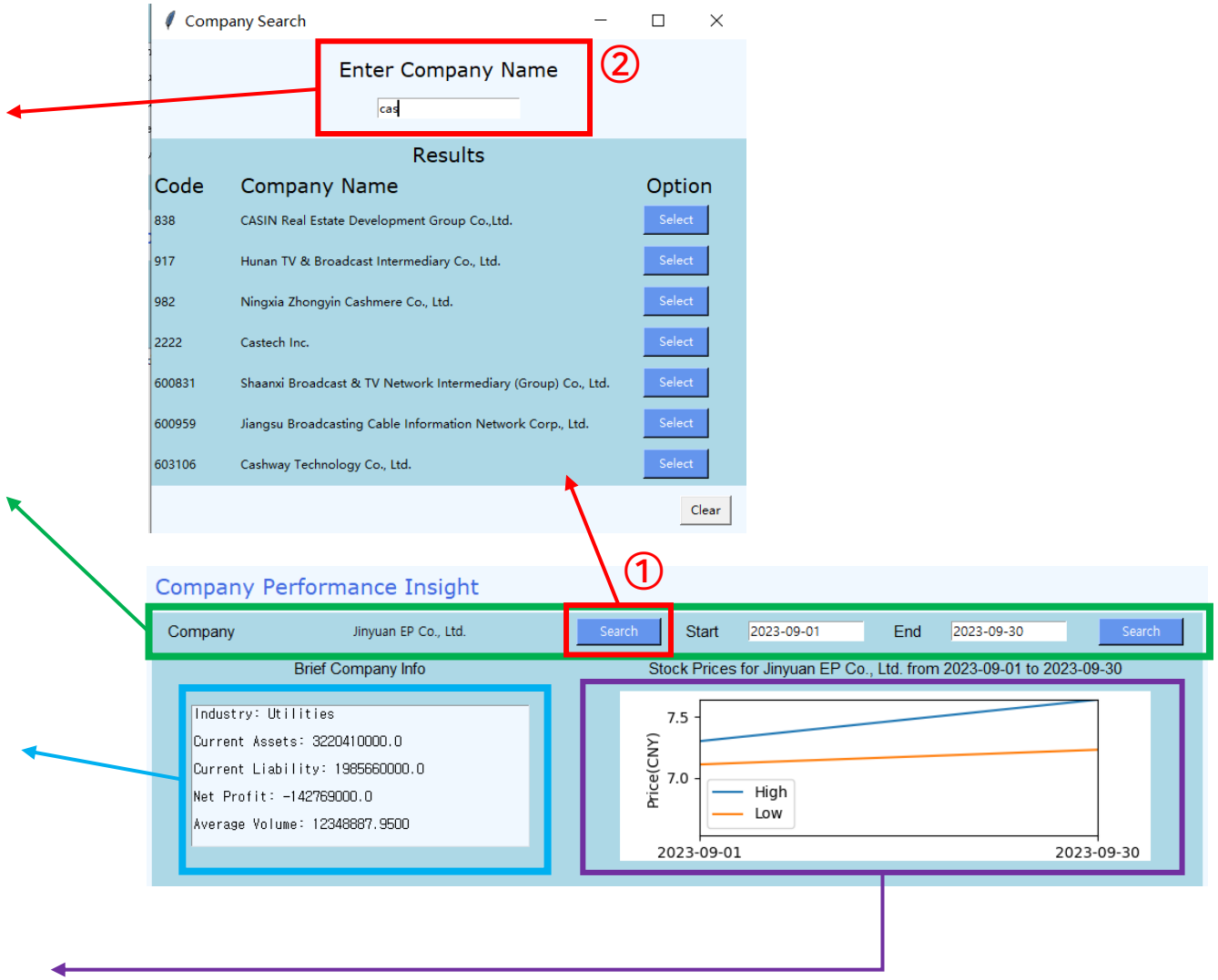
‘Company Performance Insight’ Part

- From the main screen, click on the ‘Search’ button next to the ‘Company’ label to open the ‘Company Search’ window. In the new window, enter the name of the company you want to search for and select the desired company.

- After selecting the company, set the desired start date and end date you want to query. Click the ‘Search’ button on the right side.

- The information of the required company in the time period will be returned, including: company name, industry, and current assets, current liabilities, net profit, etc.

- At the same time, the right side will draw the change rate of the company's stock price during the entered time period.

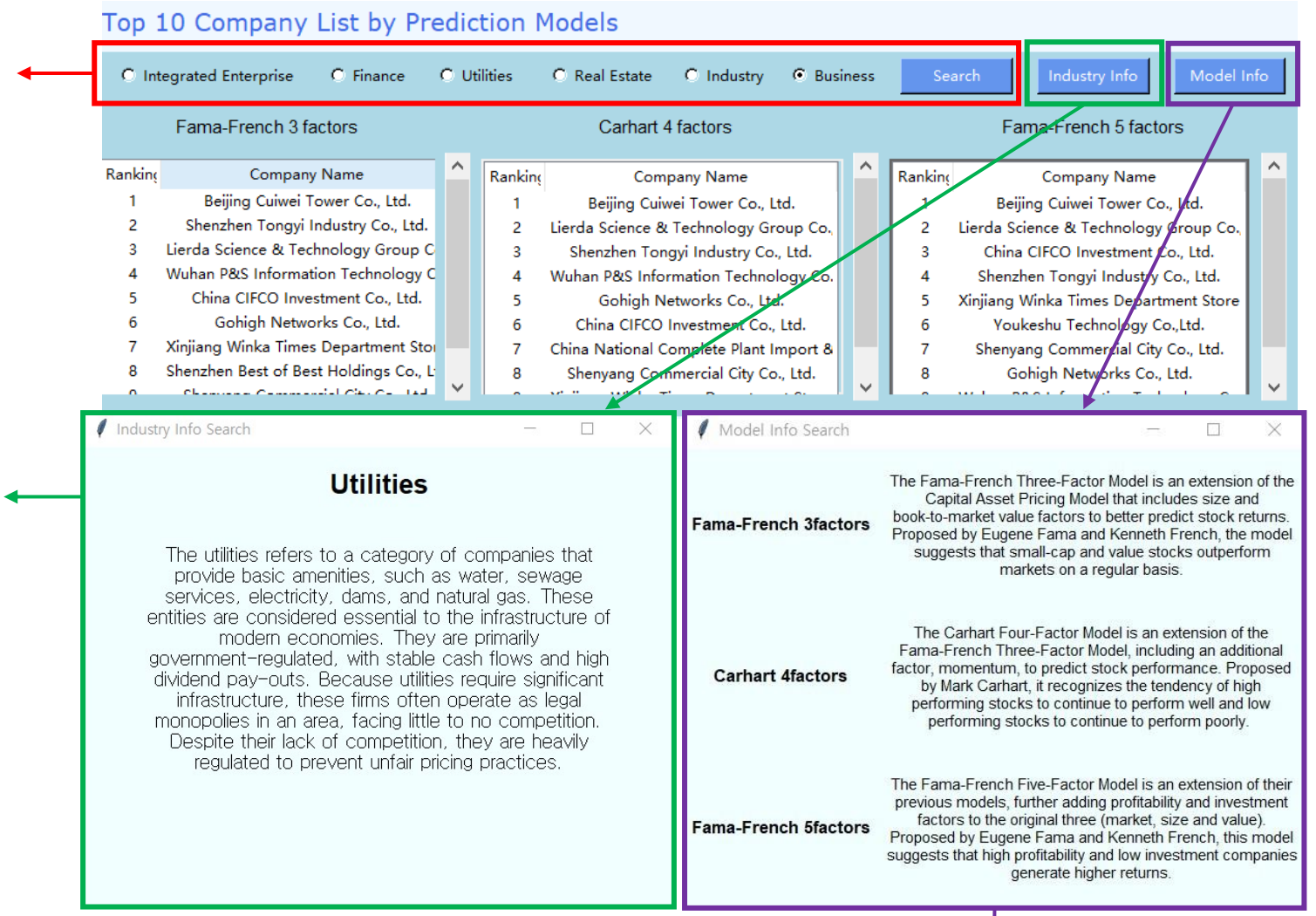


‘Top 10 Company List by Prediction Models’ Part

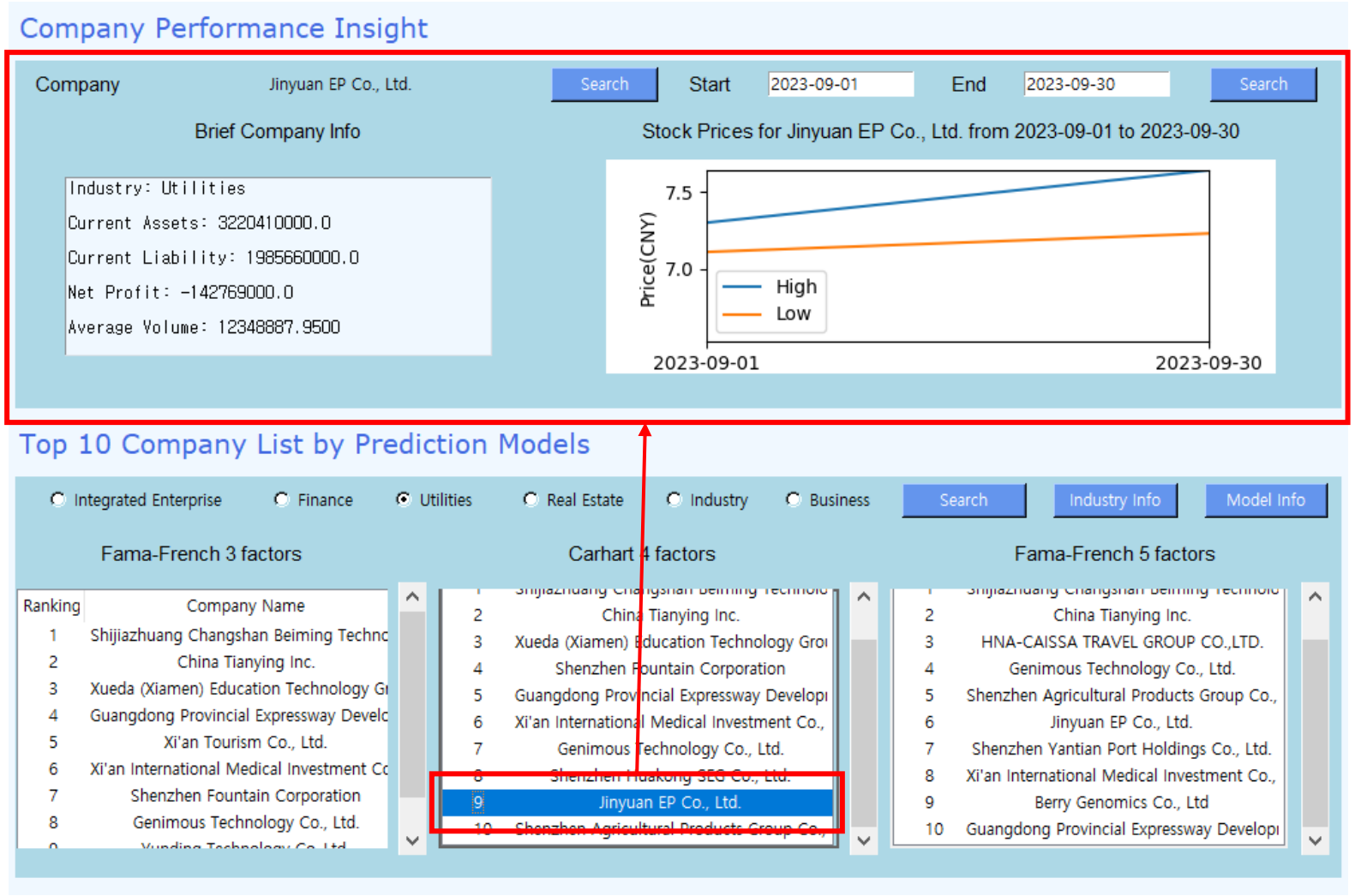
- Select the industry of interest in the Radio Buttons, click the ‘Search’ button, and then the 10 most suitable stocks for investment in the industry selected based on the three models will be displayed in the three boxes below.

- Click ‘Industry Info’ button and a new window will open, showing a brief description of the industry to which these companies belong.

- Clicking on ‘Model Info’ will open a new window showing a brief description of the three models used.



‘Top 10 Company List by Prediction Models’ Part



- In the list of companies generated by each prediction model, double-click on the name of the company if you wish to obtain information.

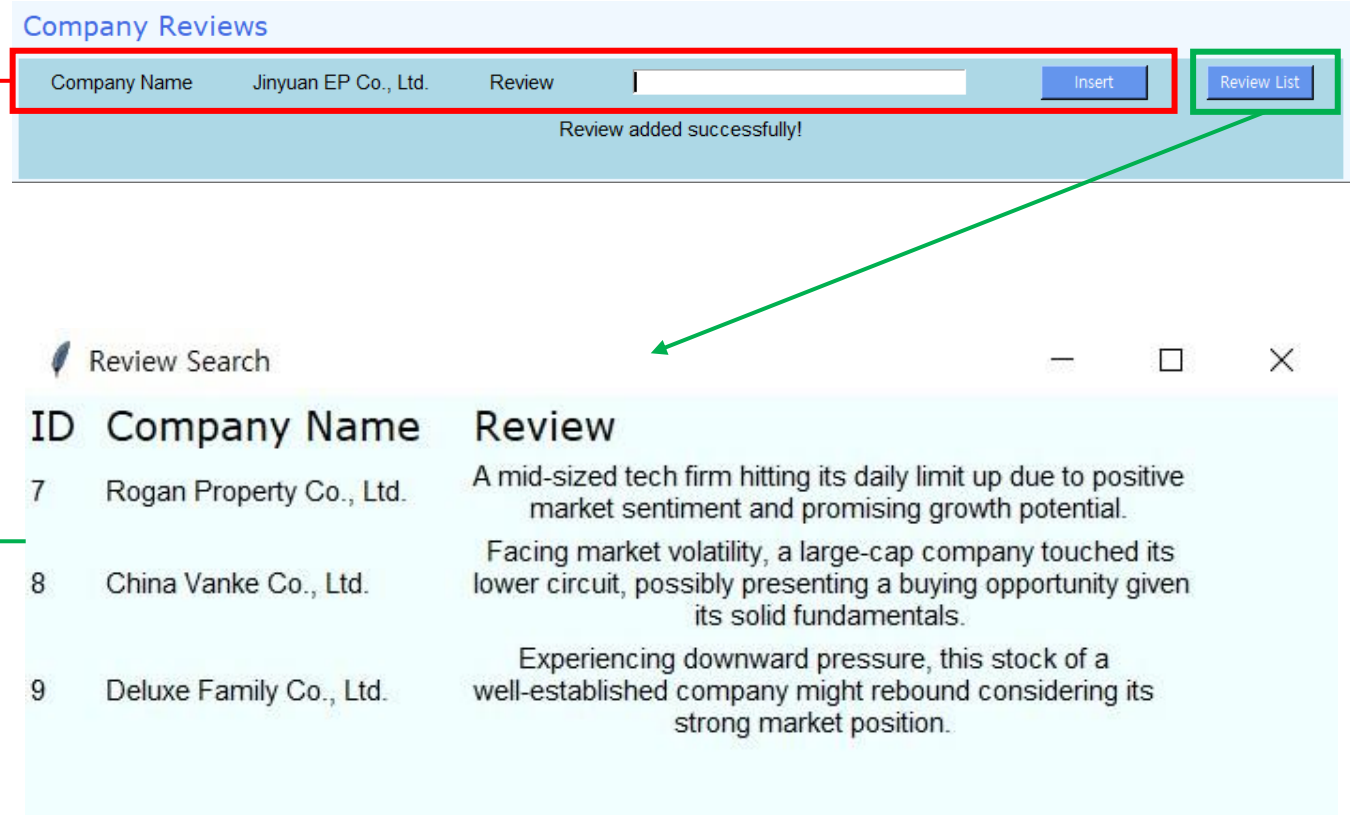
- Then, you'll be able to see information about that company displayed in the 'company performance insight' frame according to the date set. (company name, brief info, stock price chart)

‘Company Reviews’ Part

- In the ‘Company Reviews' section, the name of the company you selected above will be pre-selected. Enter the review for that company and click the insert button. You will then see the label 'Review Added Successfully!’

- Alternatively, if you want to view already written reviews, click the ‘Review List’ button to open a new ‘Review Search' window.

- you can see the ID, company code, and corresponding comments.

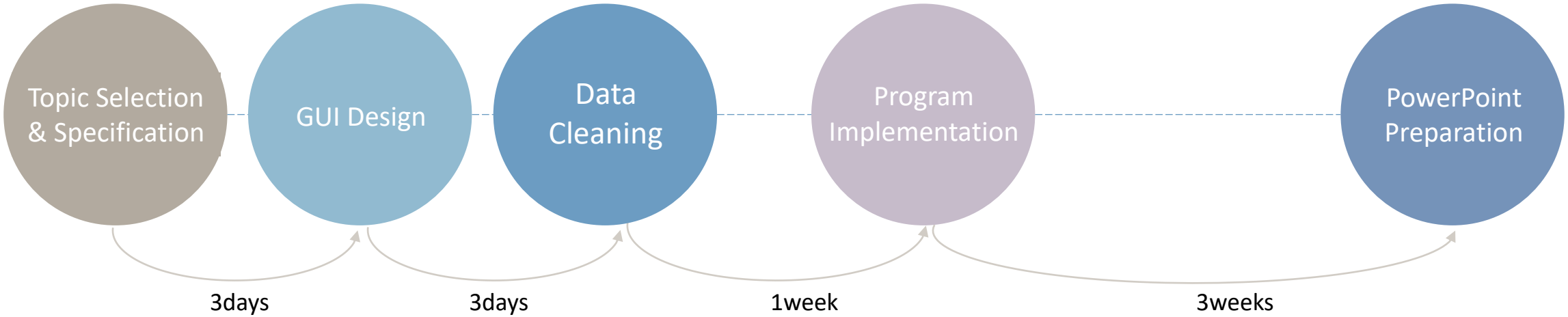


Part 4

Lesson & Learned

Lesson & Learned

Timeline



Challenges & What We Learned

1. **Diverse features**, stock data, daily charts, prediction

2. **Database modification**: errors, effort in building program

3. **Valuable learning**: feature prioritization, program development


THANK YOU
FOR LISTENING !


Q & A

Appendix-Prediction

- Based on the raw data obtained, we use three classic asset pricing models to judge whether a certain tar get is worth investing in:

$$E(R_i) - R_f = \alpha + \beta_{i,M}(E(R_M) - R_f) + \beta_{i,SMB}E(SMB) + \beta_{i,HML}E(HML) \text{(FF3)}$$

$$E(R_i) - R_f = \alpha + \beta_{i,M}(E(R_M) - R_f) + \beta_{i,SMB}E(SMB) + \beta_{i,HML}E(HML) + \beta_{i,MOM}E(MOM) \text{(Carhart4)}$$

$$E(R_i) - R_f = \alpha + \beta_{i,M}(E(R_M) - R_f) + \beta_{i,SMB}E(SMB) + \beta_{i,HML}E(HML) + \beta_{i,RMW}E(RMW) + \beta_{i,CMA}E(CMA) \text{(FF5)}$$

$E(R_i) - R_f$: Expected return of the investment- Risk-free rat

$E(R_M) - R_f$: Market risk premium, the expected excess return on the market portfolio over the risk-free rate

$E(SMB)$: The expected excess return of small-cap stocks over large-cap stocks

$E(HML)$: The expected excess return of high book-to-market (value) stocks over low book-to-market (growth) stocks

$E(MOM)$: The expected excess return of stocks that have high returns over past 3 to 12 months over those have had low returns

$E(RMW)$:The expected excess return of firms with high operating profitability over firms with low operating profitability

$E(CMA)$: The expected excess return of firms with low investment growth over firms with high investment growth

α: Excess returns that cannot be explained by the market

- Calculate the corresponding factor data through the stock trading data in the past 12 months, bring it into the above equation for group regression, and rank stocks according to the size of the estimated parameters α