PDA Final assessment report

Predictive Data Analytics Course

IBAT College, Dublin. Lecturer: John Rowley

by Hossein Yazdi

Even though one option is enough, just to learn more, I have tried to work on all four chapters of our course, Python, MySQL, RapidMiner, and Power BI, in two different Scenarios to cover two distinct data, one from Siemens company and the other from ASX Australian market.

# First Scenario: Siemens Financial Statements Analysis in Python and MySQL

Siemens is a global powerhouse focusing on the areas of electrification, automation, and digitalization. One of the world's largest producers of energy-efficient, resource-saving technologies, Siemens is a leading supplier of systems for power generation and transmission as well as medical diagnosis. (Siemens B, 2022)

Three reports from its financial statements have been downloaded in Python at the first step and after analyses to provide various results and interpretations in a file named “Hossein Yazdi. PDA Final Assessment in Python” has been exported to CSV files and then imported to MySQL to extract some specific results. The second part in MySQL is minor duties, while the main focus had been performed in Python.

**To see the full report in both parts please refer to the repository’s folder “First Scenario” as in the Python part and the SQL part files, the complementary reports, are provided.**

However, the following section is a summary extracted from the mentioned folder:

Some libraries like Panda and matplotlib are imported

Three primary files are as follows:

Siemens Consolidated Income Statement.csv from 2020 and 2021

Siemens Consolidated Statements of Financial Position.csv from 2017 to 2020

Siemens Consolidated Statements of Comprehensive income.csv from 2020 and 2021

The files are cleaned and prepared by removing some unwanted rows, filling out some empty data by other elements, and changing the names of columns.

After that, to establish the vertical and horizontal analyses we need some new columns to calculate the ratios based on the primary or major items like revenue in the income statement.

Then to be able to interpret the results there are different graphs, pie charts, and bar charts.

For every diagram, the relevant interpretation is written under them.

Finally, it is noticeable that the company had managed the costs, especially the variable costs, in a great manner during these years, because of notable results in some ratios like income per revenue.

Then the final data are transferred to MySQL to extract some minor particular tables, such as the five larger amounts, or joining data from two or more tables.

# Second Scenario: ASX 200 Analysis in MySQL, RapidMiner, and Power BI

## MySQL:

ASX 200, is an Australian share market index comprising the 200 largest companies (by market capitalization) listed on the Australian stock market. (Staff, 2022)

Extracting the age of every transaction by using the “TIMESTAMPDIFF(DAY, Date, CURDATE())”

Then by applying the “select, from, where” command, the companies with high performances in some specific sectors are extracted.

**The full report of this part is in another distinct file as a SQL TEXT format. Also, the files related to this section are in the “Second Scenario/MySQL Part” folder.**

## RapidMiner

In RapidMiner two dataset, one complete file like a training one and the other unlabeled file which does not include the “Sector” information. The prediction for this latter one is achieved by using the apply model and other tools in RapidMiner.

**The full report of this part is in another distinct file. Also, the files related to this section are in the “Second Scenario/RapidMiner Part” folder.**

## Power BI

The same file from asx200, training one, is loaded as csv to Power BI.

**The full report of this part is in another distinct file. Also, the files related to this section are in the “Second Scenario/Power BI Part” folder.**

# Reference

CFI Team, 2022. Financial Controls. [online] Corporate Finance Institute. Available at: <https://corporatefinanceinstitute.com/resources/knowledge/finance/financial-controls/> [Accessed 26 August 2022].

Hotz, N., 2018. What is CRISP DM? Data Science Process Alliance. Available at: <https://www.datascience-pm.com/crisp-dm-2/> [Accessed 21 August 2022].

ibm, 2022. Predictive Analytics | IBM. [online] Available at: <https://www.ibm.com/analytics/predictive-analytics> [Accessed 20 August 2022].

Siemens, 2022. Siemens Aktiengesellschaft (SIE.DE) Aktienpreis, Nachrichten, Kurs und Verlauf – Yahoo Finanzen. [online] Available at: <https://de.finance.yahoo.com/quote/SIE.DE/> [Accessed 2 July 2022].

Siemens A, 2022. Investor Relations | Company | Siemens Global. [online] Available at: <https://new.siemens.com/global/en/company/investor-relations.html> [Accessed 20 August 2022].

The link for asx200.csv: <http://www.asx200list.com/wp-content/uploads/csv/20170401-asx200.csv>"