End semester practical examination

Course: Programming for Analytics (MPBA 507)

Date: 2:30-4:30pm, 6 December 2021

Note: The solution requires cleaning and combining two files: (1) "Top Management Salary.xlsx", and (2) "Companies Data.xlsx"

1. Import the "Top Management Salary.xlsx" file into a data-frame "**tpm_salaries**".
2. Clean the columns names, and remove unnecessary rows from "**tpm_salaries**".
3. Remove the duplicate names in the 'Executive Name" column in "**tpm_salaries**".
4. Import the "Companies Data.xlsx" file into a data-frame "**companies_data**".
5. Clean the columns names, and remove unnecessary rows from "**companies_data**".
6. Remove the duplicate names in the 'Company Name" column in "**companies_data**".
7. Sort the Top Management names by alphabetical order in "**tpm_salaries**"
8. Sort the Company names by alphabetical order in "**companies_data**"
9. Identify the top five paid top management. Ensure that the salaries are not in scientific format.
10. Identify the oldest top management leader and the youngest top management leader.
11. Create a group column "profitability", and assign values "1" for companies with PAT > 0, and values "0" for PAT <0 in "companies_data"
12. Create a new data frame that has columns "company_name, "tpm_name", "marketcap", "PAT", profitability, salary and designation.
13. Extract the salaries, PAT, market cap, profitability and designation of top management of companies that have "Bajaj" name grouped by year.
14. Extract the salaries, PAT, market cap, profitability and designation of top management of companies that with more than 1 lakh crores of market cap grouped by year.
15. Identify the longest serving top management leader, and how the salary changed for this leader.
16. Submit the final response at this link: https://forms.gle/hC2KCr31tZ29h7ATA