Head of Analytics Task

Task Description: This task evaluates your SQL querying and Python data analysis skills.
You will work with two datasets:
Your goal is to extract, analyze, and present insights accurately, efficiently, and with well-documented code.

- Due Date/Time: Mentioned in the email
- Estimated Time to Complete: 3- 4 hours

Task 1

Data: https://github.com/datacharmer/test_db
ER Data Model (if needed):
https://github.com/datacharmer/test_db/blob/master/images/employees.png

You can use any dialect of SQL
Output format should be in firstname_lastname_sql_test.sql

Part 1

Show the maximum annual salary in the company after 01/06/1995.

1. Show the TOP 10 annual salaries in the company initiated (from_date) after 01/06/1995 (June)
2. Show employees who satisfy the following description: He (gender male) was 45 when hired, born in October and was hired on Sunday.
3. In the dept_emp table, show the quantity of employees by department (dept_no). To_date must be greater than current_date. Show departments
with more than 13,000 employees. Sort by quantity of employees.
4. Show the minimum and maximum salaries by employee. Also show the quantity of salary changes. Use aliases.
5. Show employees who were hired in 1985. Display the quantity of employees by day (hint: use function - dayname).
Sort result to show data from Monday to Sunday (reminder use something other than dayname)

Part 2

Display lines only for employees whose salary changed 3 or more times. Use window functions as a priority.

Columns we are interested in:

- Emp_no,
- from,
- to,
- salary,
- Salary_change by how much the salary increased during promotion in absolute value, - Salary_change_all_time by the amount increased from the very first letter in the row in percentage,
- Days_to_salary_increase (the number of days that have passed since the previous salary

increase),
- Salary_change_num_during_year employee salary change number within the year, - Rapid_promotion Field or rapid growth of salary (If the current promotion is more than 15+% from the previous one to 'rapid' otherwise 'slow').

Task 2

Data: https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci
File: online_retail_II.xlsx (use 2010–2011 sheet)
Output format should be in firstname_lastname_python_test.ipynb

Part 1. Data Cleaning & Summary

Load the dataset.

Filter out:

▸ Missing CustomerID
▸ Negative quantities and unit prices
▸ Cancelled orders (InvoiceNo starts with "C")

Add:

▸ Revenue = Quantity × UnitPrice
▸ InvoiceDate as proper datetime

Produce a short data quality summary:

▸ Total rows removed by each cleaning step
▸ Total unique customers and products
▸ Total revenue

2. Quick Descriptive Analytics

Find:

▸ Top 5 products by total revenue
▸ Top 5 countries by total revenue
▸ Top 5 customers by total revenue

Plot:

▸ A bar chart of top 5 products by revenue
▸ A line chart of daily total revenue (aggregated over time)

3. Business Insight

Based on the results, write:
▸ 2 observations on data quality (e.g., many cancellations, pricing anomalies) ▸ 2 observations on revenue concentration (e.g., top products or customers dominate) ▸ 1 potential business action you'd take as Head of Analytics.

Submission Instructions

▸ Submission format (e.g., document, code file (.sql, .ipynb))
▸ Where to submit - send it to ani@propfirmmatch.com
▸ Subject line - Task / Head of Analytics