**INTERVIEW MINI-PROJECT: EMPLOYEE DATA MART**

**Purpose:**

Test candidate’s ability to:

· Write code to perform simple data manipulation

· Access data via an API

· Plan a data loading execution

**Included Datasets:**

(see ‘explanation’ tab within each .xlsx file for explanation of variables)

· Employee\_Roster\_Data.xlsx

· Email\_Data.txt

· hours.xlsx

· skills.xlsx

**Instruction:**

* Read below Problem Statement and familiarize yourself with the attached datasets
* Create a root folder with the name “InterviewProject\_<YourName>” (replace <YourName> with your actual full name).
* Inside the root folder include a single folder per task (task1, task2, task3)
* Please use Python as your programming language.
* Upload all the files to your github personal account
* Send the Git URL to [pablo.giuffrida@rga.com](mailto:pablo.giuffrida@rga.com) including all the files before the deadline.

**Deadline: Thursday, April 23, 2021** (early submission is ok too!)

**Submission Checklist:**

Remember to include in the right folder the following items:

* Copy of formatted Employee\_Roster, Hours and Skills tables (exports after second task)
* Copy of all the code files used to perform this project.

**Problem Statement:**

Company ABC has over 1,300 employees located in four offices around the globe. ABC would like to understand what factors, if any, correlate with employee performance. ABC would like you to use email communication data as a proxy for employee collaboration and worked hours and skills to understand performance.

ABC has data on its employees contained within the Employee\_Roster\_Data.xlsx file, employee email data contained within Email\_Data.txt, employee worked hours on each client within hours.csv and employee skills data within skills.xlsx (see each file for explanation of variables).

* Each row within Employee\_Roster\_Data describes attributes about an individual employee
* Each row within Email\_Data describes an individual email sent by a person (either an employee or person not employed at ABC).
* Each row within hours describes the hours spent by an employee in admin tasks and on each different client. Also includes the target total hours.
* Each row within skills describes a skill level for a given employee.

You will see within Employee\_Roster\_Data that each employee has a unique Email\_ID which corresponds to the from\_ID and to\_ID within Email\_Data. If an ID within Email\_Data does not appear within the Email\_ID column of Employee\_Roster\_Data, it is because that individual person is not an employee at ABC (e.g. a vendor sending an email to an employee, or an employee sending an email to a customer).

The below tasks describe in detail what Company ABC would like you to do.

**Task #1 - Format Data**

Salary information is contained within **Employee\_Roster\_Data.xlsx** but is in different currencies. Please convert all salaries to USD using the **API** from<http://fixer.io/>. You should use the exchange rates as of **April 3, 2017** (do not use today’s exchange rate). Important note: You must access the exchange rate data using Python; do not simply access the API from your web-browser and manually copy the exchange rate results. Please import the resulting dataset into a table in MS SQL Server Express Edition.

Task # 1 Answer:

Please create a folder inside “InterviewProject\_<YourName>” with the name “task1” including the .py /.sql files you used to get data from the API and the upload script from Python into SQL Server.

**Task #2 - Upload and transform data**

Upload all the data into SQL Server Express Edition and create the data model. Please follow these steps:

* **Create a table** in SQL Server for each data source.
* **Upload** all the remaining files into MS SQL Server Express Edition.
* **Clean** each data source. Remove, if any, useless columns and special characters.
* Create a unified **dimension** for each attribute: Create a single and unified table for each attribute to represent all the possible values for an attribute in all the different data sources (For example, Users, Departments, Titles, etc.).
* **Alter** the existing tables by adding the following **fields using SQL Server**. Please create them inside the table you think they will fit:
  + Total Working Hours by Employee: 40 hours per week.
  + Utilization: % of client hours against: working hours minus admin hours
  + Client time: % of client hours against total working hours
  + Admin Time: % of admin hours against total working hours
  + Skills Level Description (numeric value is the Attribute column in skills spreadsheet:
    - 0: wants to learn.
    - 1-2: Heavy supervision
    - 3-4: Light supervision
    - 5: Expert
* **Export** the final versions of hours, skills and Employee\_Roaster\_Data tables.

Task # 2 Answer:

Please create a folder inside ““InterviewProject\_<YourName>”” with the name “task2” including the .py /.sql files with the scripts used to create the tables, upload the files and transform the data. Also, include the export of hours, skills and employee\_roaster\_data tables.

**Task #3 - Plan an execution**

Describe how you would execute and schedule the whole data loading process from the previous two tasks.  
Detail what steps you would follow from the start, when you pull the data, until the end, having the processed and transformed data in the final set of tables. Describe how each script would be executed and in what order.

If you know any tools to do some of the work with, you can name them, but it is not necessary to describe the way the tool works, just a general idea of the concepts you are using.

You can also use pseudo code if you want to, to get some structure to the data loading scripts.

Task # 3 Answer:

Please create a folder inside “InterviewProject\_<YourName>” with the name “task3” including a .docx document with the plan’s detailed steps