Data Engineering Assessment

Question 1:

a. By referring to **Question1-Material.xlsx**, provide us with a normalized data model design to minimize data redundancy.

Based on your normalized model, answer the below questions by providing us with the respective SQL queries:

b. Which are the 10 largest completed deposit transactions? Extract the amount, customer email, customer brand name and calendar date.

c. What is the total number and amount of failed deposit transactions per brand? Extract total number, amount, customer brand name, paymentstatus.

d. How much daily turnover and accounting revenue did each brand generate per product in the first 6 days of the year? Extract turnover, accountingrevenue, brand.

e. What is the average gamewin per product? Extract the average gamewin and product.

f. Which customers had a lifetime total turnover of 500 EUR or more and what was this total turnover amount? Extract turnover, customeremail, customer

Question 2:

By referring to the **Question2-Material.zip** folder, write a console application/program in C# to read all the csv files in the mentioned folder. Read the data from each file and store in a local database with the correct data types for each field. You can unzip the given file manually or through the program.

Additional points:

- Take care of error handling
- Cannot read file using hard-coded file names
- First row of each file is the header
- CSV Delimiter is comma (,)

Extra:

- Once file is processed, move file into a PROCESSED folder, and if file has an ERROR and cannot be parsed, move to ERROR folder. For this exercise a file with no data rows should be sent to the ERROR folder.

Question 3:

As Data Engineer you are requested to write a Python application in order to connect to Adventure Works database on SQL Server and do data manipulation.
https://docs.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver15&tabs=ssms

Steps include:

- Restore **AdventureWorks2019.bak** database on SQL Server. It is important to use the 2019 OLTP version.
- Connect to the database using pyodbc or any other library you prefer
- Using tables Sales.SalesOrderHeader and Sales.SalesOrderDetail create and populate a Fact table called Fact.Sales to the best of your abilities
- Using Sales.Customer and tables in Person schema create and populate a Dimension called Dim.Customer. You do not need to use all the tables in Person schema to create this dimension.
- You can create Fact.Sales and Dim.Customer using either DDL or Python

Notes:

- Refer to **AdventureWorks ERD.jpg** which can help you in the above assignment
- It is highly recommended to write structured and well documented code

https://docs.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver15&tabs=ssms

https://github.com/Microsoft/sql-server-samples/releases/download/adventureworks/AdventureWorks2019.bak