**Data Analyst Task – Mrinmoy Ghosh**

Github location - <https://github.com/mrinmoyghosh09/Assessment1/tree/main>

1. There are 3 files given as a source, which I downloaded in my system.
2. Converted Shipping JSON to CSV file using python and loaded it into MS SQL Server using SQL Server Management Studio (SSMS). Loaded Order.csv & Customer.xls to Local SQL Server under Test DB the same way. It can also be loaded via multiple other way. Like using SSIS, python etc. I am attaching the method using SSMS & SSIS here.

Json to CSV conversion - <https://github.com/mrinmoyghosh09/Assessment1/blob/main/json_to_csv.py>

A screenshot of a computer

AI-generated content may be incorrect.

SSIS Method –

<https://github.com/mrinmoyghosh09/Assessment1/blob/main/Load_Customer_File.dtsx>

**Technical Specifications for the Data Engineering Team**

1. Below link contains the DDL queries for creating the tables.

<https://github.com/mrinmoyghosh09/Assessment1/blob/main/Metadata.txt>

**Data Accuracy Check**

1. Data Quality –
   1. There is no duplicate in the data.
   2. There are impurities in the First name column of Customer data, as it contains special characters & numeric value in some cases.
   3. Added a flag to the rows having data quality issue. A call can be taken to discard or modify those records as per the Business requirement. For this assessment, I kept the rows as it is. The details is here -  
      <https://github.com/mrinmoyghosh09/Assessment1/blob/main/Customer_updated.xlsx>

**Data Model**

1. Entity Relationship data model –
   1. Customer to Order: One-to-Many (A customer can place multiple orders)
   2. Order to Shipping: There is no direct relationship in this case. However, one Order can have multiple shipment.
   3. Customer to Shipping: One-to-Many

**Insights for the Customer/Stakeholder**

I have attached the query to figure out the data for the 5 questions mentioned and also added the results adjacent to it.  
<https://github.com/mrinmoyghosh09/Assessment1/blob/main/Query%20%26%20Results.xlsx>

I am listing down the Insight Stories based on the data shared –

1. The orders are shipped in 3 countries, where UK saw the highest sales and UAE is the lowest.
2. Average Age of the Customer is higher in UAE than UK & USA.
3. Among countries, UK has a higher pending Shipment to be delivered. Used the query in below link to get the result.

<https://github.com/mrinmoyghosh09/Assessment1/blob/main/Insight.xlsx>

1. Maximum amount of order came from Monitor & Harddisk.

**Sample Report/Dashboard**

I have tried to create a sample PowerBI summary page out of the data I got. Attaching a screenshot and the pbix file from github.

<https://github.com/mrinmoyghosh09/Assessment1/blob/main/Assessment.pbix>

A close-up of a graph

AI-generated content may be incorrect.