1. DATA ENRICHMENT RECOMMENDATION:.
   Null Values and Missing Entries: While the dataset is rich on insights it has a significant number of nulls especially on key indicators like Year. This makes it challenging to make time based modeling. Accurate capture of the year the vehicle was acquired and or sold is key in driving critical sales decisions.
   Sales Persons/staff information: It will be interesting to be able to link the sales to the staff for the purpose of coaching, performance improvement training or knowing which staff are more critical in driving more sales.
   CVM Campaign data: Are there any promotions that have been done over the past periods. It will be great to link the promotions data with the sales data so as to measure the specific value of the promotional campaign.

2. DATA WAREHOUSE STRUCTURE:
   From the simple design I have come up with in doing this assignment, I will proceed with a **DATALAKE** over the traditional data warehouse.
   My Datalake will have a four tier architecture namely:
   - Landing - Receiving layer.
   - Bronze - Datalake rules layer
   - Silver - Featurestore layer layer
   - Gold - Reporting layer

   The entire Datalake will sit in an object storage database such as a local server/machine directory, Blob storage for Microsoft Azure or S3 Bucket for AWS. The reason for choosing an object storage and not an RDMS system is I want to be able to handle all types of files from csv files, log files, text files, excel files and even data from other databases.
   Object storage file systems give me this degree of freedom.

   The Landing Layer: Data coming into the lake is received here first. At this stage the data is handled and stored as is (No transformation applied.)

   The next three layers of Bronze, Silver and Gold is what conforms to the traditional data warehouse concept of ***star schema architecture***.

   Bronze Layer: Data from Landing is then transformed according to the rules of the lake and stored in delta format.

   Silver Layer: This is similar to fact tables in traditional data warehouses only that it can help provide 360 degrees view of the data sets for example 360 degree view of each customer.

   Gold Layer: Reporting happens from this layer. Its the equivalent of aggregate or reporting tables in the traditional data warehouse.
   Ideally business intelligence tools will connect to the lake at this layer.

3. Design Principle:
   I usually avoid tightly designing solutions that are tightly coupled to a particular system or service provider. For example designing a system that can only run in AWS and will require a complete overhaul should the migration to on-premise or Azure become necessary.
   My codebase will mostly sit in for example Github,Bitbucket, Azure Devops or any versioning tool available.
   I will use Jupyter notebooks or Databricks for development.
   For archastration I can use Azure data factory, Nifi, Airflow, depending on what tool is available.
   With this kind of design migrating from one platform to another will mostly need configuration changes only.

GAYLORD ODHIAMBO