**Narratives:**

**Information details provided:**

File A – File A is a Text file. This file holds information regarding the vehicles that are/were Part of the Auto dealers Inventory.

* VIN Number
* Year the vehicle for made
* Vehicle Model
* Color of the Vehicle
* All Wheel v/s Front Wheel Drive
* Number of Doors in the Vehicle
* Engine Type – Electric v/s IC engine
* Manufacturer’s suggested retail price

File B – File B is an excel format. It has the Customer, Sales and vehicle information that was sold. Following are the data fields provided in the file.

* Transaction Id
* Last Name, First name and Middle Initials of the Customer
* Address details of the Customer
* Date the Vehicle was sold
* Vehicle information like Model, Year of make, Color, Engine Type and VIN
* Sales information like MSRP, Discount if applicable, Trade in (if applicable), Trade in value, Purchase price, If the customer is a Repeat customer or not.

File C – File C has data in Word Document. It has information regarding the Customer who brought a vehicle from the Auto Dealer. Following information is present in the file.

* Last Name, First name and Middle Initials of the Customer
* Address details of the Customer
* Additional information pertaining to the Customer like profession and special financing needs etc.

**Below are the Assumption made in Designing and Developing the Schema, Relational Model and Database.**

1. The Auto dealer only sells brand new vehicles.
2. Auto Dealer has only 3 Departments i.e. Inventory, sales and Customer relations.
3. The auto dealer is only involved in Sales. Post sales servicing is not been done by the Auto Dealer.
4. The E-R Model attached does not re-present all the attributes for a given entity. Please refer to the Logical data Model or Physical table to know all the attributes in a given entity.

As part of the Design I have normalized the information provided in 3 different logical Models mentioned below:

1. Logical Model for information related to Inventory Management
2. Logical Model to represent information regarding the Vehicle Sales
3. Logical Model to store Customer relation information

During the process of designing the Schema and Database all of the information provided as part of the input files were used. I did introduce new fields like the Customer Id, Vehicle Availability Status.

Following factors and processes were considered to represent the data:

* Information provided on the input files was analyzed.
* Entities and attributes were identified.
* Relationships between the entities were identified and defined.
* Cardinality was established between the identified entities.

I have chosen **VIN number** as a Primary Key for the inventory management model. VIN number will help uniquely identify a vehicle in the Auto Dealers inventory.

I have chosen **Customer id** as a Primary key for the Customer management model. Customer id will uniquely identify auto dealer’s customers and their corresponding details.

The Sales management model has **Transaction Id** as Primary key. VIN Number and Customer Id are the foreign key in the Sales Management Model; these foreign keys will help establish relationship with Inventory management and Customer management Models. The Transaction id attribute will help uniquely identify the sales transaction done by the auto dealer. The VIN number as a foreign key to the Inventory management model will help identify the corresponding vehicle that was sold as part of the sales transaction id. The Customer id attribute as a foreign key on the Sales management model will help uniquely identify the customer that has purchased the vehicle.

**The ER model:**

* The Inventory management Model and the Sales Model have **one to one** relationship.
* The Customer management model and the sales management model **have one to many** relationship considering that a single customer might have multiple vehicles purchased from the Auto dealer.
* The Inventory management Model and the Customer management Model have **one to many** relationship. Customer can purchase multiple vehicles from the Auto dealer’s inventory.
* Due to limited amount of the information/data provided; Understanding the information provided and logically separating the attributes was the hardest part during the course of designing the Schemas and Database.

**Data Curation considerations while designing the Schema. The below activities can sustain for future data discovery as well.**

* Revision Summary tab provided to document any changes to the Schema, logical model, entities etc.
* Scalability - New source of information can be easily brought in and integrated to this existing data Model

E.g.) Post sales servicing related transaction can be easily implemented and Stored in the sales management model.

* Storage methods can be easily changed without impacting application programs
* New attributes can be added without significant impact to the application program
* Provenance – Vehicle availability Status is derived based on the sales information
* Modifications can be tracked and audited using the Revision History
* Document any kind of re-formatting in the Revision summary

Additional Data Curation activities that I would recommend:

* Develop Access control Mechanism
* Develop Scripts for data transformation