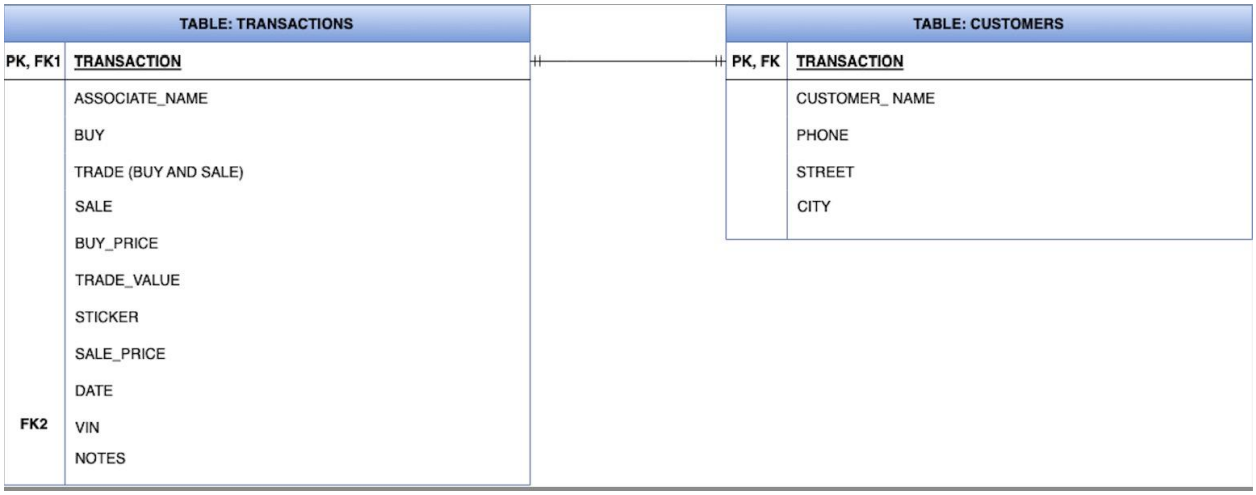
Assignment 2: Ontologies / ER Diagram Design Exercise

Felicia Liu (liu318@illinois.edu)
CS598: Foundations of Data Curation

1. Evidence of in-depth examination of data sources (steps 1 & 2)

   [10 points] Create an ER diagram for the Pre-owned dealer database, as described in the attached file.



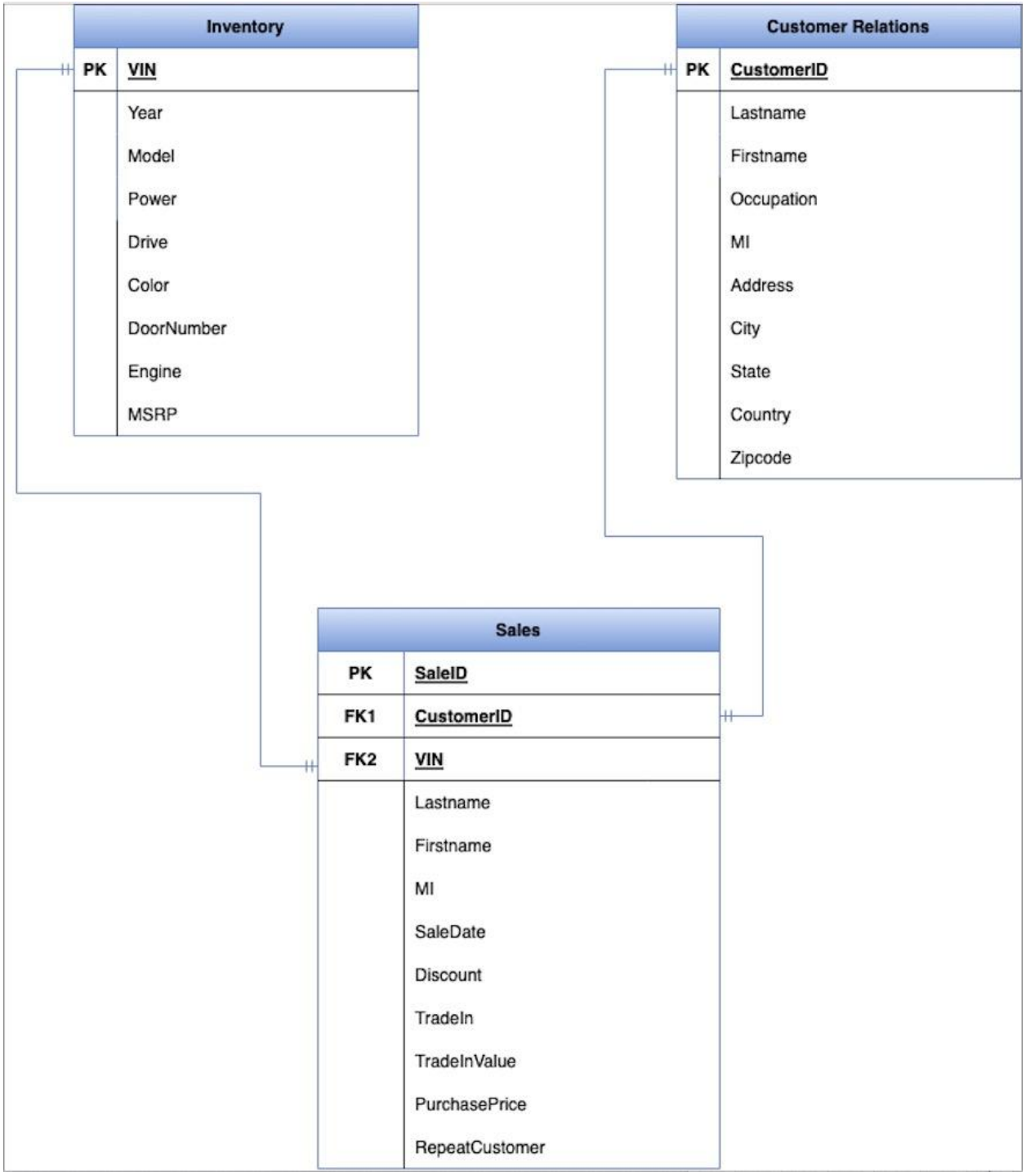
ER Diagram for pre-owned cars dealer company

“MDS_Exercise2_File A” is an Excel file and it contains table “Transactions” and table “Customers” respectively.

- In the “Transactions” table, column “Transaction” works as both primary key in this table and foreign key to link the table “Customers”. Besides the column “Transaction”, there are columns “Associate_Name”, “Buy”, ”Trade(Buy and Sell)”, “Sale”, “Buy_Price” constituting this table.
- In the “Customers” table, column “Transaction” serves as the primary key because it is the unique number to differentiate the customers. What is more, it is also the foreign key to link table “Transaction”. This table also includes customer’s personal information like “Customer_Name”, “Phone”, “Street”, “City”, etc.

2. Evidence of in-depth examination of data sources (steps 1 & 2)

   [10 points] Create a separate ER diagram that reflects the schema you designed for Assignment 1. You may update the schema based on feedback you received from instructors after submitting Assignment 1.



ER Diagram for auto dealer company

The relationship for the three schemas is presented as below.

- In the “Inventory” table, column “VIN” acts as the primary key since it is the unique number for each vehicle. In the “Sales” table, column “VIN” serves as the foreign key to link the “Inventory” table and “Sales” table.
- In the “Customer Relations” table, column “CustomerID” is the primary key because this table aims at managing the customer information, and each customer should have a unique ID. In the “Sales” table, column “CustomerID” works as the foreign key to link the “Customer Relations” table and “Sales” table.
- In the “Sales” table, the column “SaleID” is the primary key since it stands for each unique order.

3. Evidence of understanding ER diagrams and data integration (steps 3 & 4)

   Create at least 3 intermediate ER diagrams that showcase your integration process [15 points]. These diagrams should be accompanied by narrative prose (either in a separate document or as annotations directly to the diagram) that describe each of the integration steps taken on the diagram [20 points]. See the integration process described in the data integration slides for examples of what this might look like, and follow the example shown in the “Schema integration: an example” lecture. There is no one right way to do this, but your decisions should be justifiable, and should minimize the potential for information loss. Be sure to justify your design decisions in your narrative prose! Discussion of both various curatorial objectives and the pros and cons of various integration steps is necessary.

For the auto dealer company from assignment 1, we mainly created 3 different tables: “Sales”, “Customers” and “Inventory”.

For the pre-owned cars dealer company, we mainly created 2 different tables: “Transactions” and “Customers”.

As a consequence, I decided to integrate table “Sales” from auto dealer company and table “Transactions” from pre-owned cars dealer company as integrated table 1, and integrate table “Customers” from auto dealer company and table “Customers” from pre-owned cars dealer company as integrated table 2, and leave the table “Inventory” separately as integrated table 3.

In order to generate the 3 intermediate ER diagrams, the following procedures are performed:

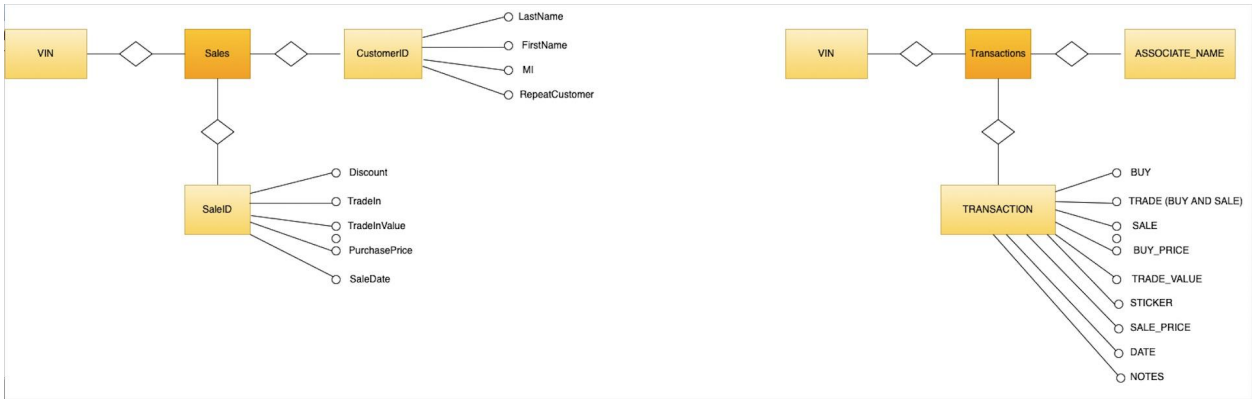
- Step 1:

To aggregate table “Sales” from auto dealer company and table “Transactions” from pre-owned cars dealer company, we needed to compare these two databases.

- Step 2:

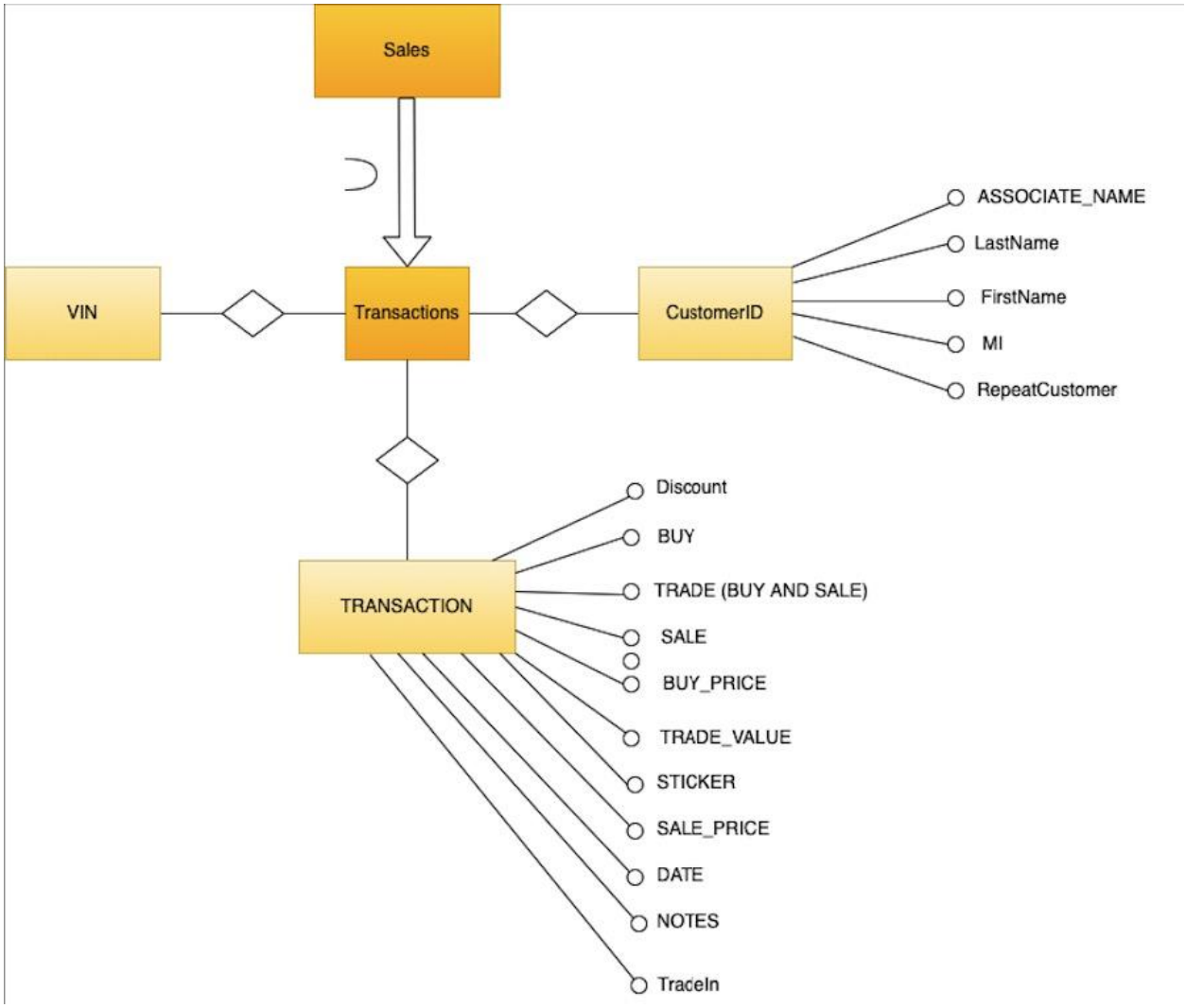
As can be observed in the below picture, the “Sales” table is included in the “Transaction” table because the “Transaction” table includes both “buy” and “sales” transactions. In addition, the column “SaleID” in the “Sales” table and the column “Transaction” in the “Transactions” table appear to refer to the same attributes. Hence, while integrating these two tables, we are able to replace column “SaleID” with the title “Transaction”. Furthermore, the column “CustomerID” and its corresponding derivative columns from the “Sale” table and the column “Associate_Name” from the “Transactions” table are also pointing to the customer’s personal information. Therefore, we can add the columns “CustomerID”, “Last name”, “First name” etc to table “Transactions”. Additionally,

the column “VIN” from table “Sales” and column “VIN” from table “Transactions” are both referring to the inventory reference number, which are deemed as the same attribute.



- Step 3:

Based on the step 2 analysis, the “Sales” table is included in the “Transactions” table. What is more, we combined the column “SaleID” and the respective columns into column “Transaction”, and expanded “Associate Name” to “CustomerID” to include more customer’s information as is indicated in the below picture. No information loss was identified after integrating these two tables.



- Step 4:

After integrating the two databases, the integrated ER Diagram for the integrated table 1 will be shown as the following picture:

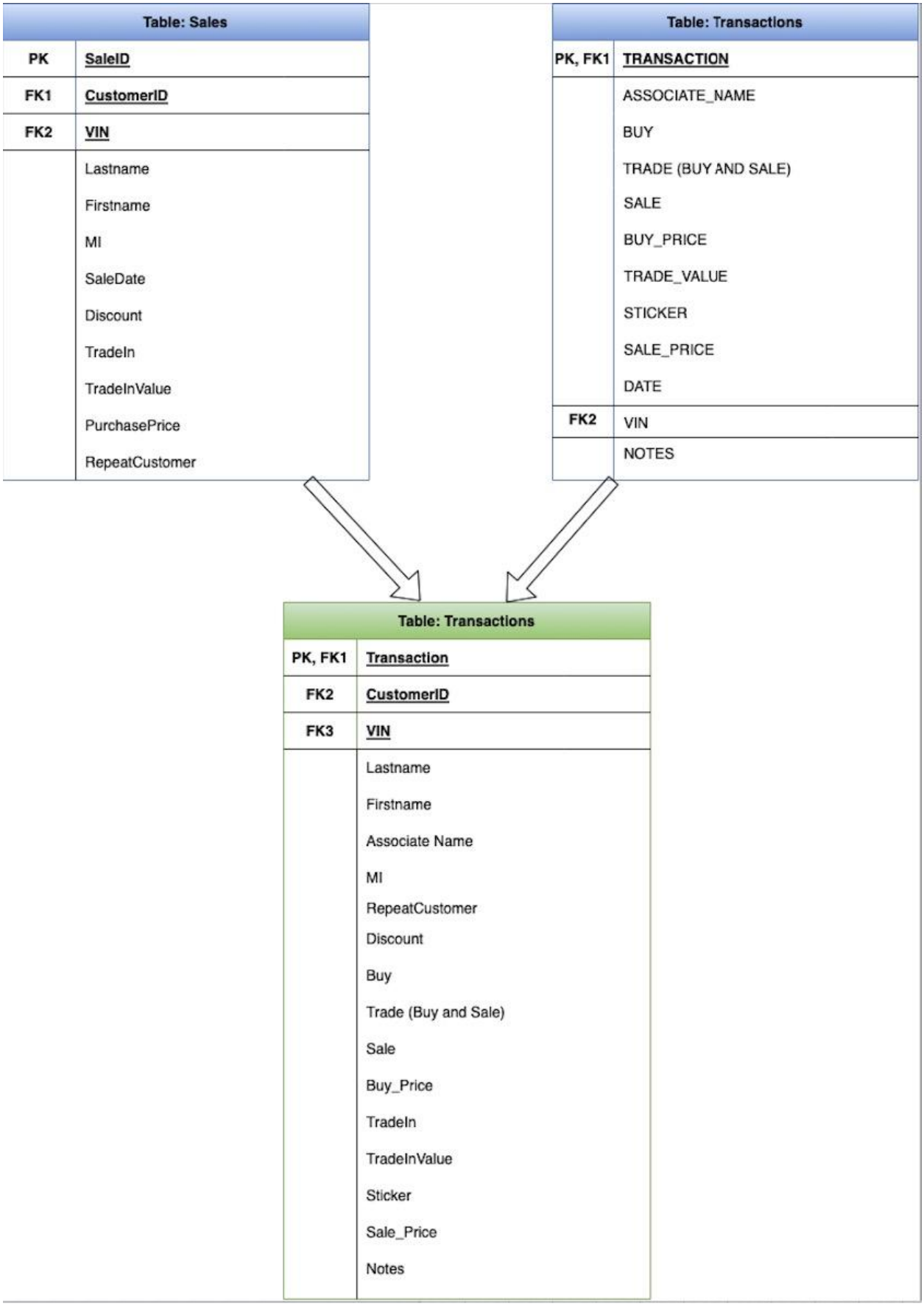
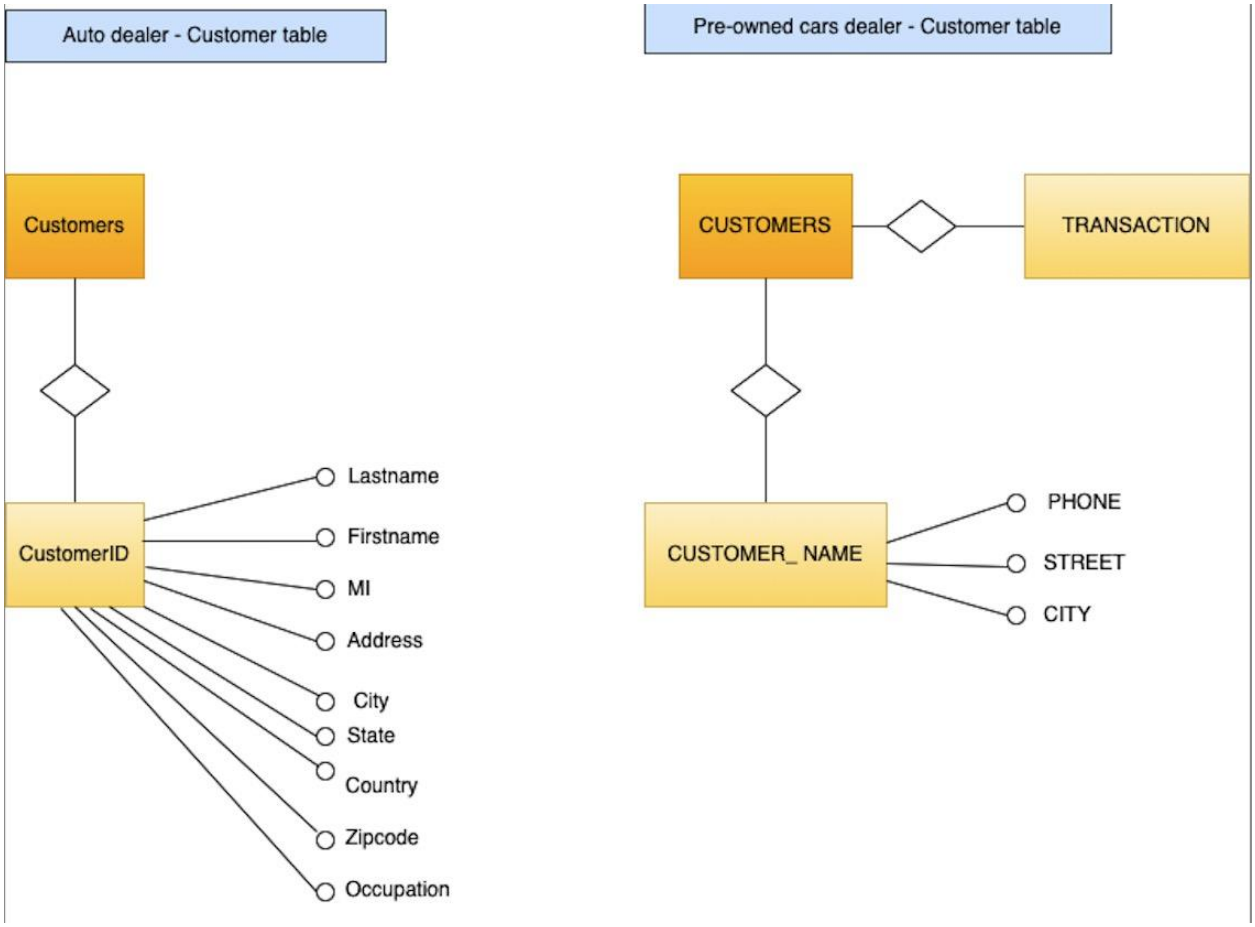


Table: Sales
PK
SaleID
FK1
CustomerID
FK2
VIN
Lastname
Firstname
MI
SaleDate
Discount
TradeIn
TradeInValue
PurchasePrice
RepeatCustomer
Table: Transactions
PK, FK1
TRANSACTION
ASSOCIATE_NAME
BUY
TRADE (BUY AND SALE)
SALE
BUY_PRICE
TRADE_VALUE
STICKER
SALE_PRICE
DATE
FK2
VIN
NOTES
Table: Transactions
PK, FK1
Transaction
FK2
CustomerID
FK3
VIN
Lastname
Firstname
Associate Name
MI
RepeatCustomer
Discount
Buy
Trade (Buy and Sale)
Sale
Buy_Price
TradeIn
TradeInValue
Sticker
Sale_Price
Notes

- Step 5:

To aggregate table “Customer” from auto dealer company and table “Customers” from pre-owned cars dealer company, we made a comparison between these two databases.

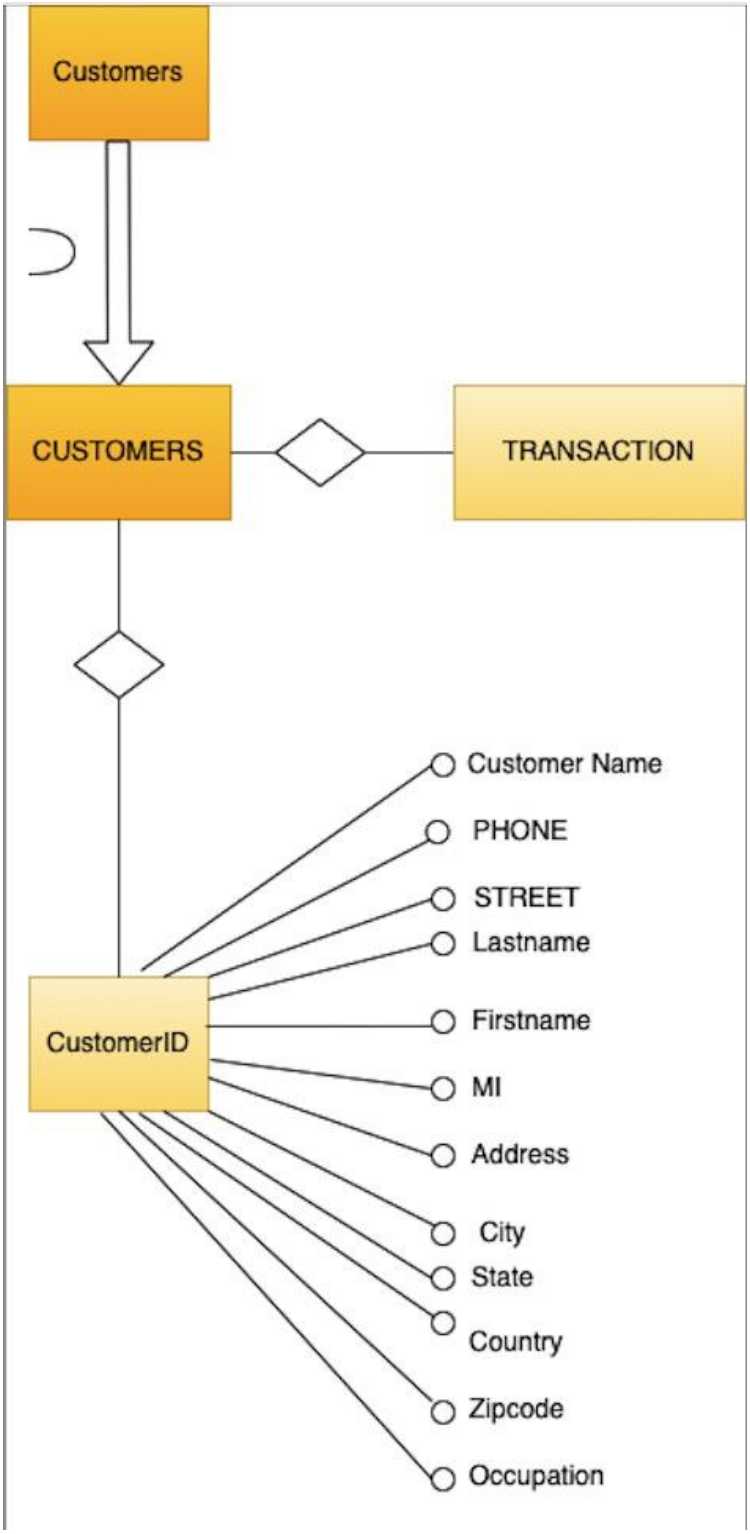
As is shown in the below picture, the “Customers” table from the Auto dealer company contains customer’s personal information including the customer's last name, first name, address, and etc. Similarily, the “Customers” table from the Auto dealer company also contains customer’s information including customer’s name, phone, street, and etc. Specifically, the “Customers” table from the Auto dealer company also includes “Transaction” information.



- Step 6:

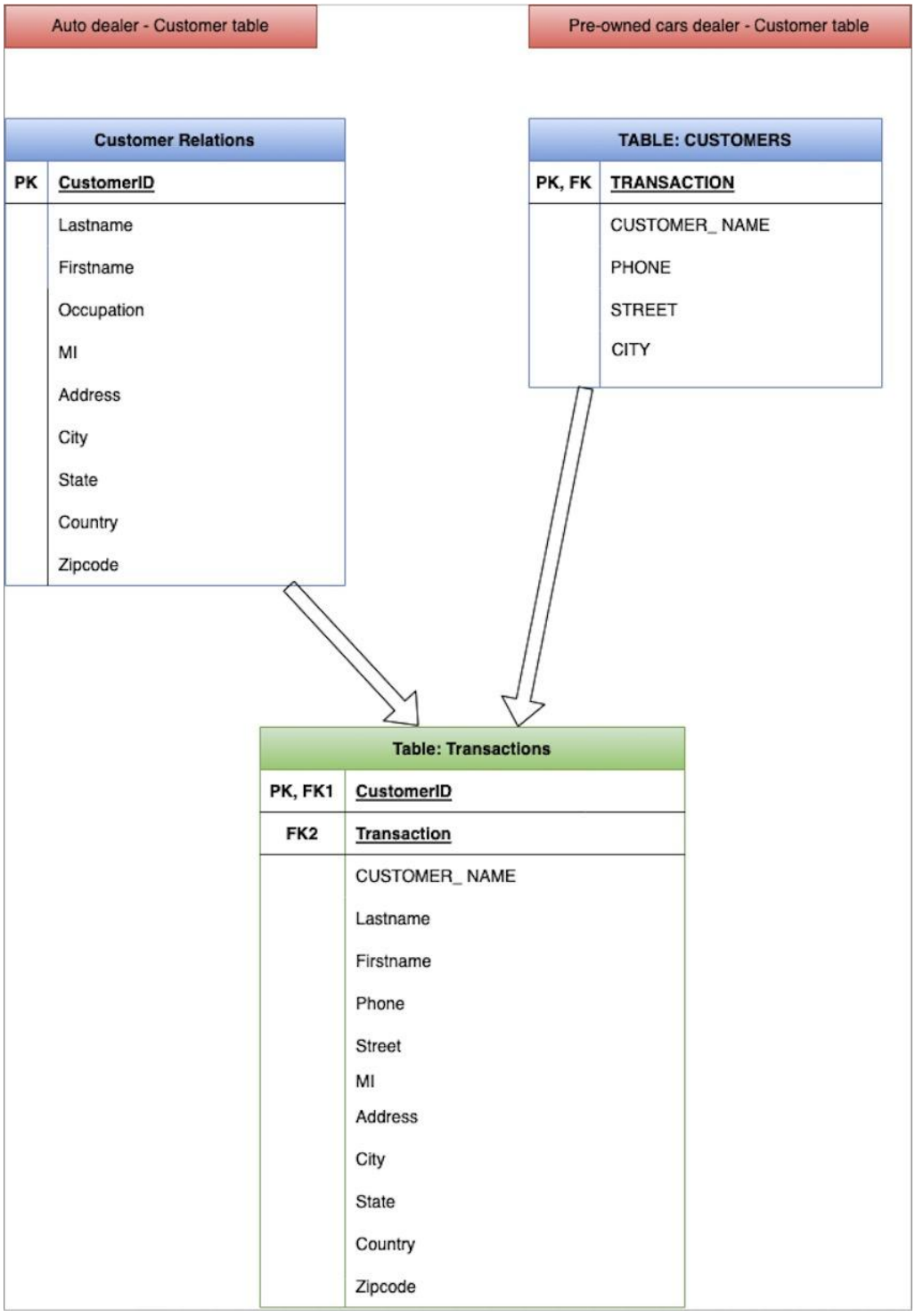
Based on the step 5 analysis, the integrated table 2 which combines the “Customers” table from the auto dealer company and the “Customers” table from the pre-owned cars dealer company should include columns “Transaction”, “CustomerID”, “Phone”, “Last name”, “First name” and etc. Since the original “Customers” table from the auto dealer company does not contain “Transaction”, we should obtain the “SaleID” from the “Sales” table as the “Transaction”. In addition, we should assign a unique number as customer ID to the customers from the pre-owned cars dealer company.

There is no information loss for the below integrated “Customers” table.


Customers
CUSTOMERS
TRANSACTION
CustomerID
Customer Name
PHONE
STREET
Lastname
Firstname
MI
Address
City
State
Country
Zipcode
Occupation

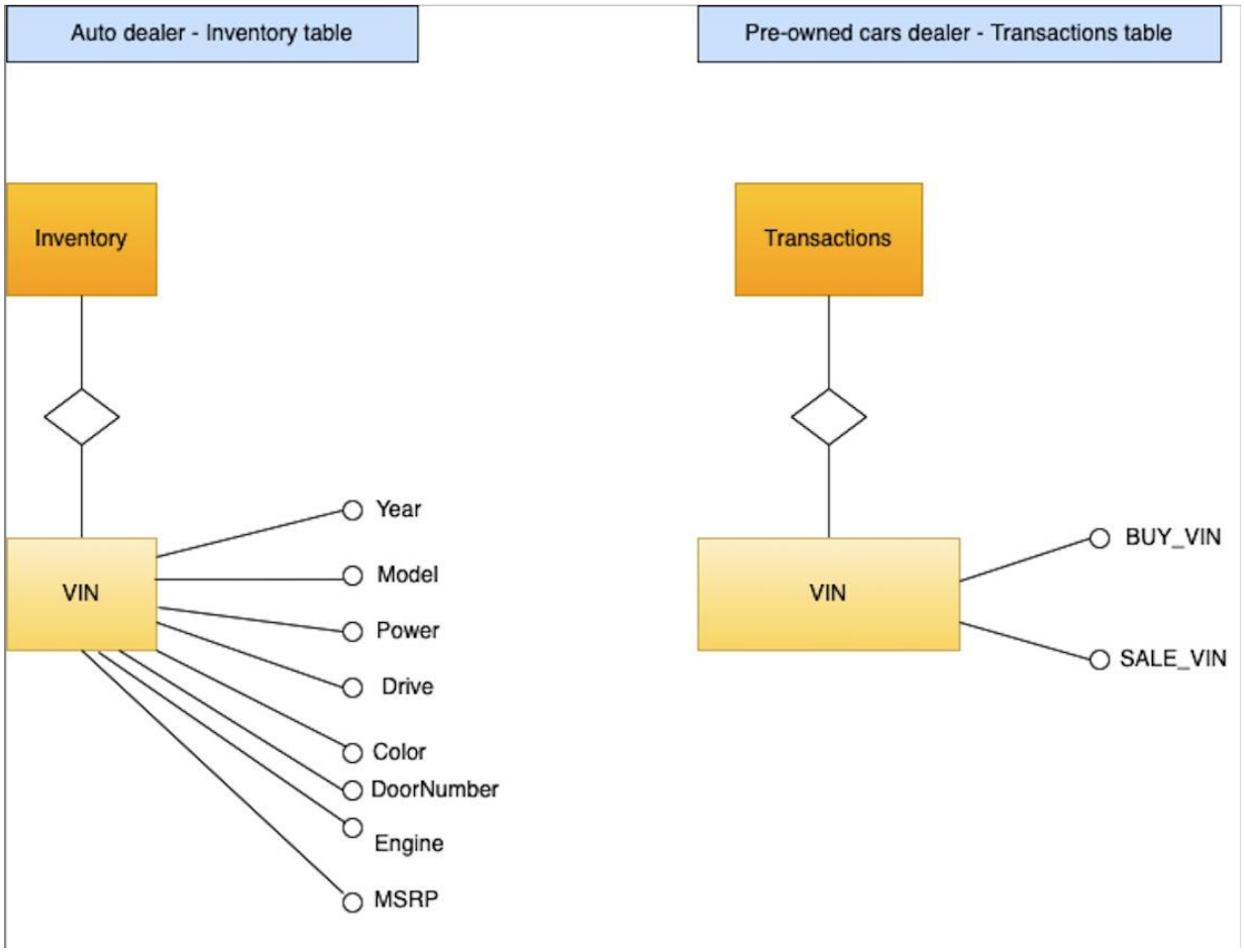
- Step 7:

After integrating the two databases, the integrated ER Diagram for the integrated table 2 will be shown as the following picture:



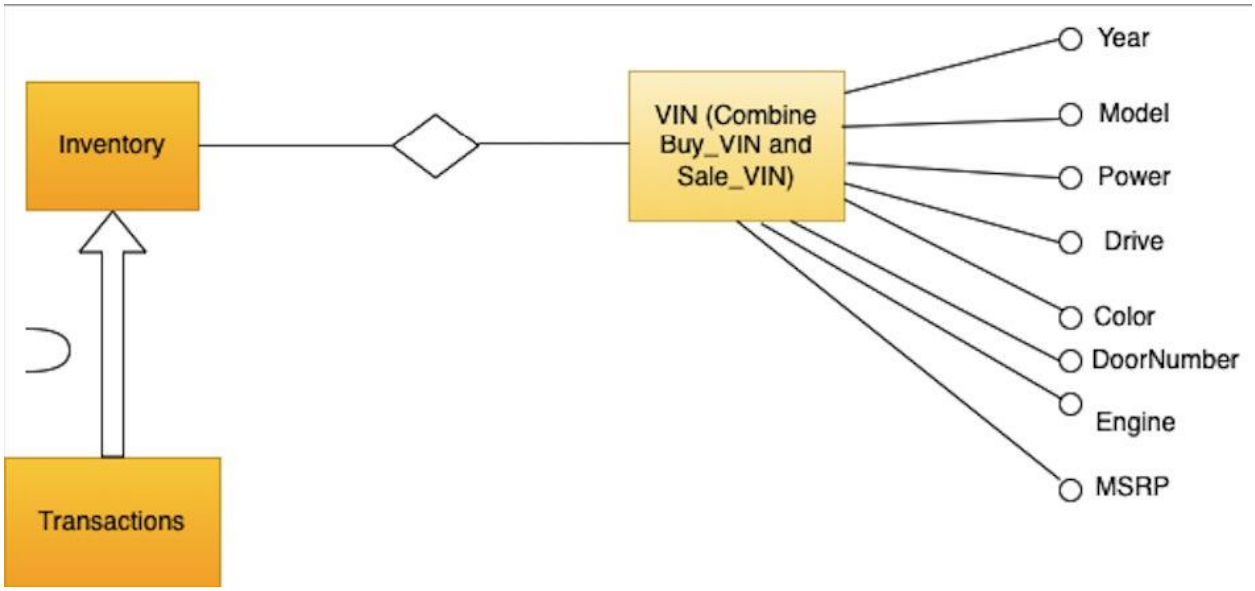
- Step 8:

There is an “Inventory” table from the auto dealer company, however, there is no similar “Inventory” table from the pre-owned cars dealer company. Nevertheless, from the “Transactions” table of the pre-owned cars dealer company, we can identify “Buy_VIN” and “Sale_VIN” columns from the “Transactions” table, which is considered the same as “VIN” column from the “Inventory” table.



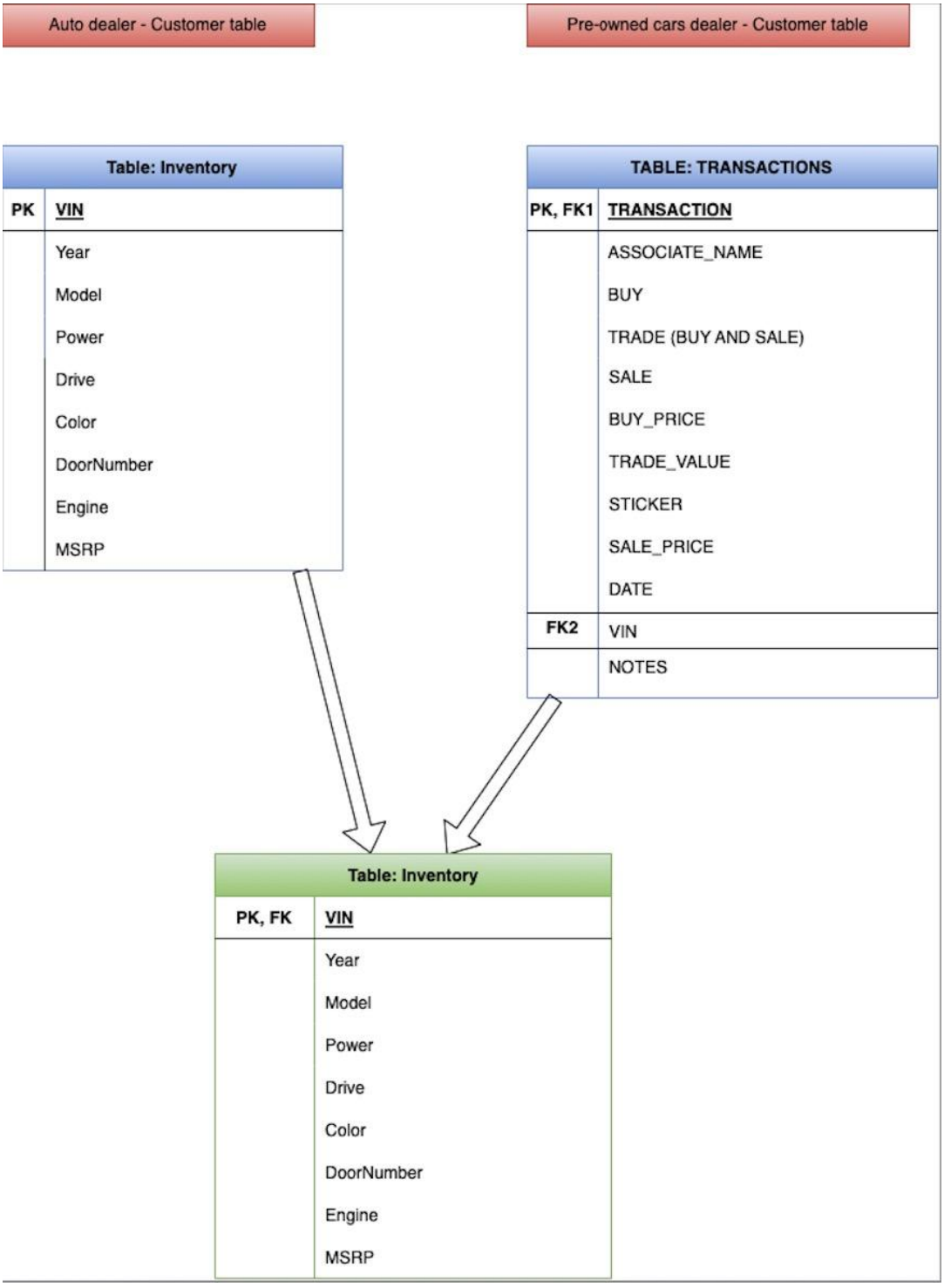
- Step 9:

Based on the step 8 analysis, the integrated table 3 which combines the “Inventory” table from the auto dealer company and the certain columns from “Transactions” table of the pre-owned cars dealer company should include columns “VIN”, “Year”, “Model”, “Power”, and etc. After further analyzing the two databases, the “Transactions” table should be included in the “Inventory” table since both the “Buy_VIN” and “Sale_VIN” columns should be set up as “VIN” in the “Inventory” table. What is more, the “Inventory” table also contains the car information like “Model”, “Power”, “Drive”. We determined that these columns should also be included in the integrated table 3 in order not to lose critical information.



- Step 10:

After integrating the two databases, the integrated ER Diagram for the integrated table 3 will be shown as the following picture:

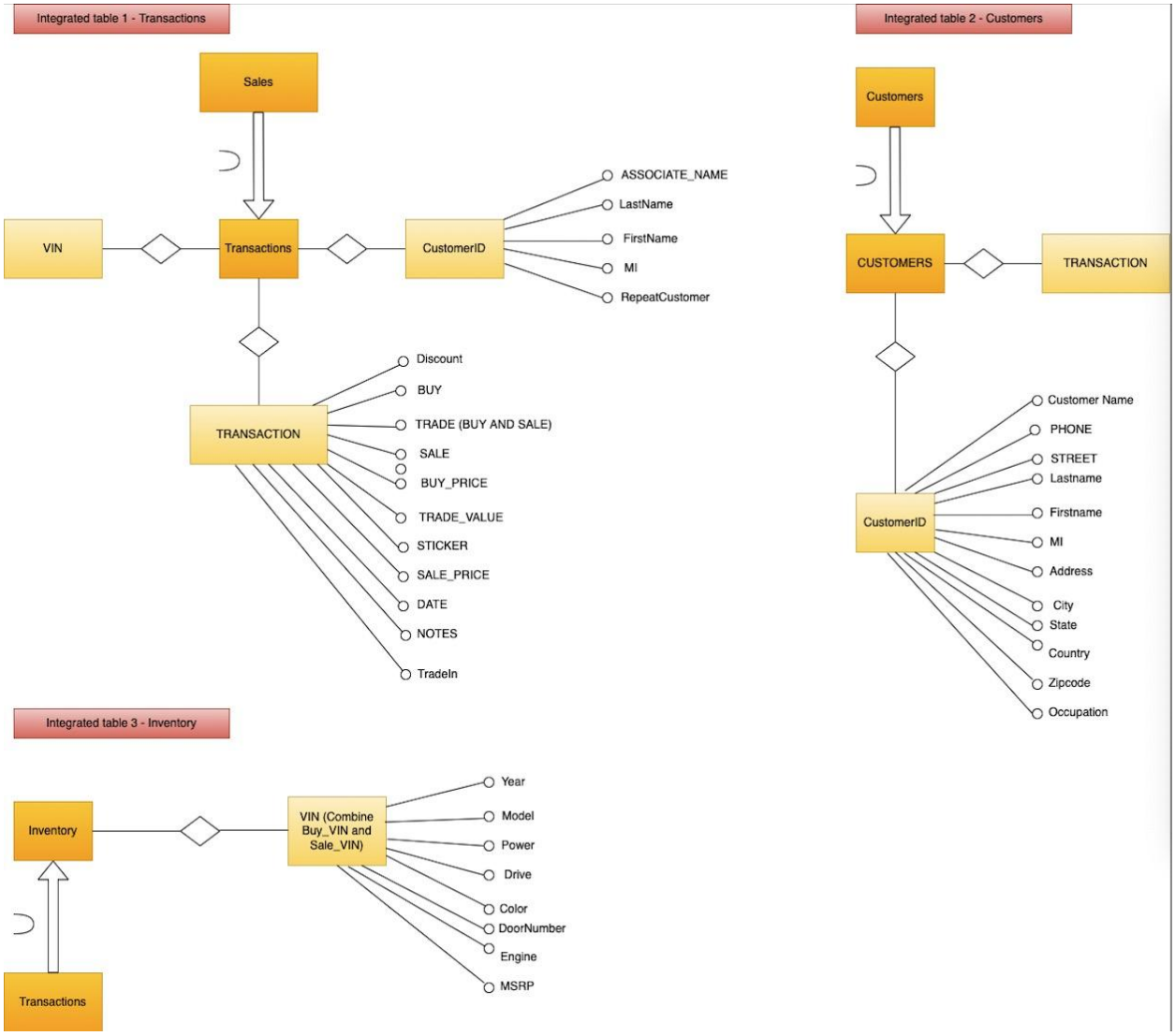


4. Thoughtful discussion of data integration objectives and decisions (steps 3 & 4)

   Finally, create one integrated ER diagram representing the merged schemas of the two dealerships (i.e., the final product of the integration process) [10 points]. Be sure to describe any final integration steps taken at this point (as described in step 3 above) [10 points]. Be sure to justify your design decisions in your narrative prose! Discussion of both various curatorial objectives and the pros and cons of the final design is necessary. Consider how you could have done things differently and in which areas the design can still be improved.

   - Step 1:

   As is indicated in the below picture, 3 intermediate ER diagrams are generated in the previous step.



- Step 2:

To aggregate the 3 intermediate ER diagrams into 1 integration ER Diagram, we need to analyze the current table structures.

As can be seen in the diagram below, the 1 integration diagram is constituted by 3 intermediate diagrams, which are “Inventory” table, “Transactions” table, and “Customers” table respectively.

- Step 3:

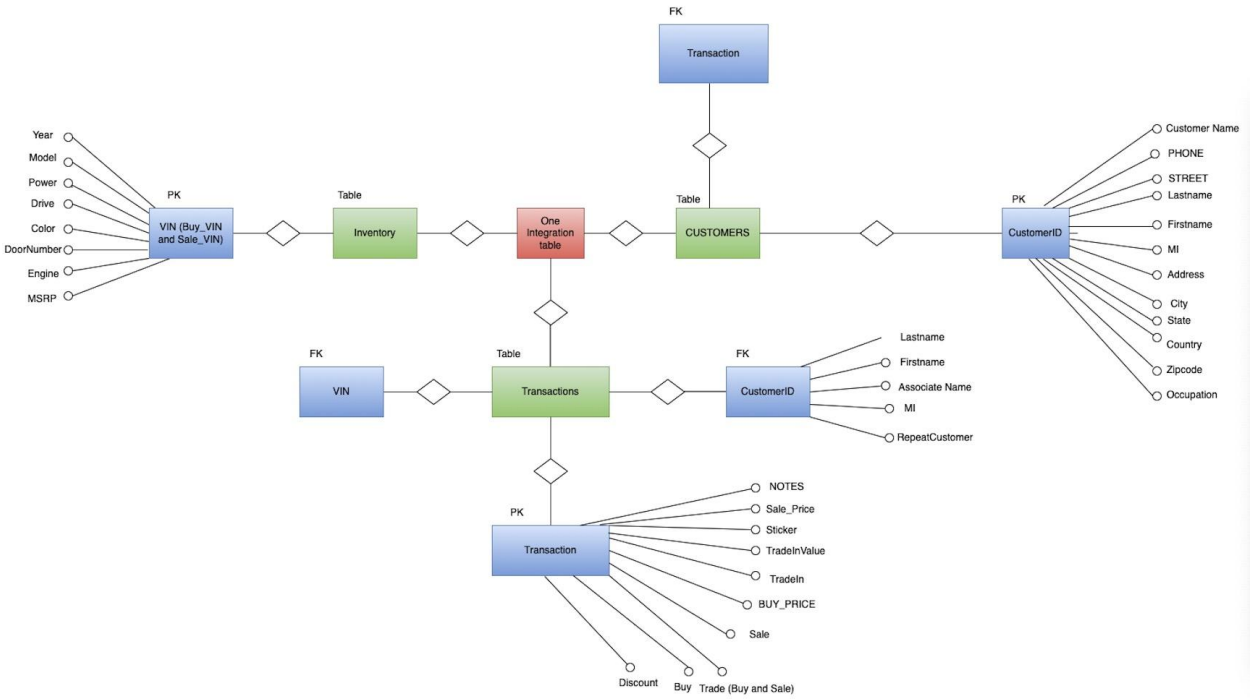
In the “Inventory” table, the column “VIN” serves as the primary key, and it includes vehicle information like “Year”, “Model”, “Drive”, “DoorNumber” and etc.

- Step 4:

In the “Customers” table, the column “CustomerID” acts as the primary key, and the database will store customer’s personal information including “First Name”, “Last Name”, “City”, “Address” and etc.

- Step 5:

In the “Transactions” table, the column “Transaction” works as the primary key, what is more, columns “VIN” and “CustomerID” are also included in this database to link the 3 intermediate tables. Additionally, transaction information like “Discount”, “TradeIn”, “Trade In Price”, “Buy Price”, “Sale Price” are presented in this table.



- Step 6:

After integrating the the 3 intermediate ER diagrams into 1 integration ER Diagram, the final integrated ER Diagram is presented as the following picture:

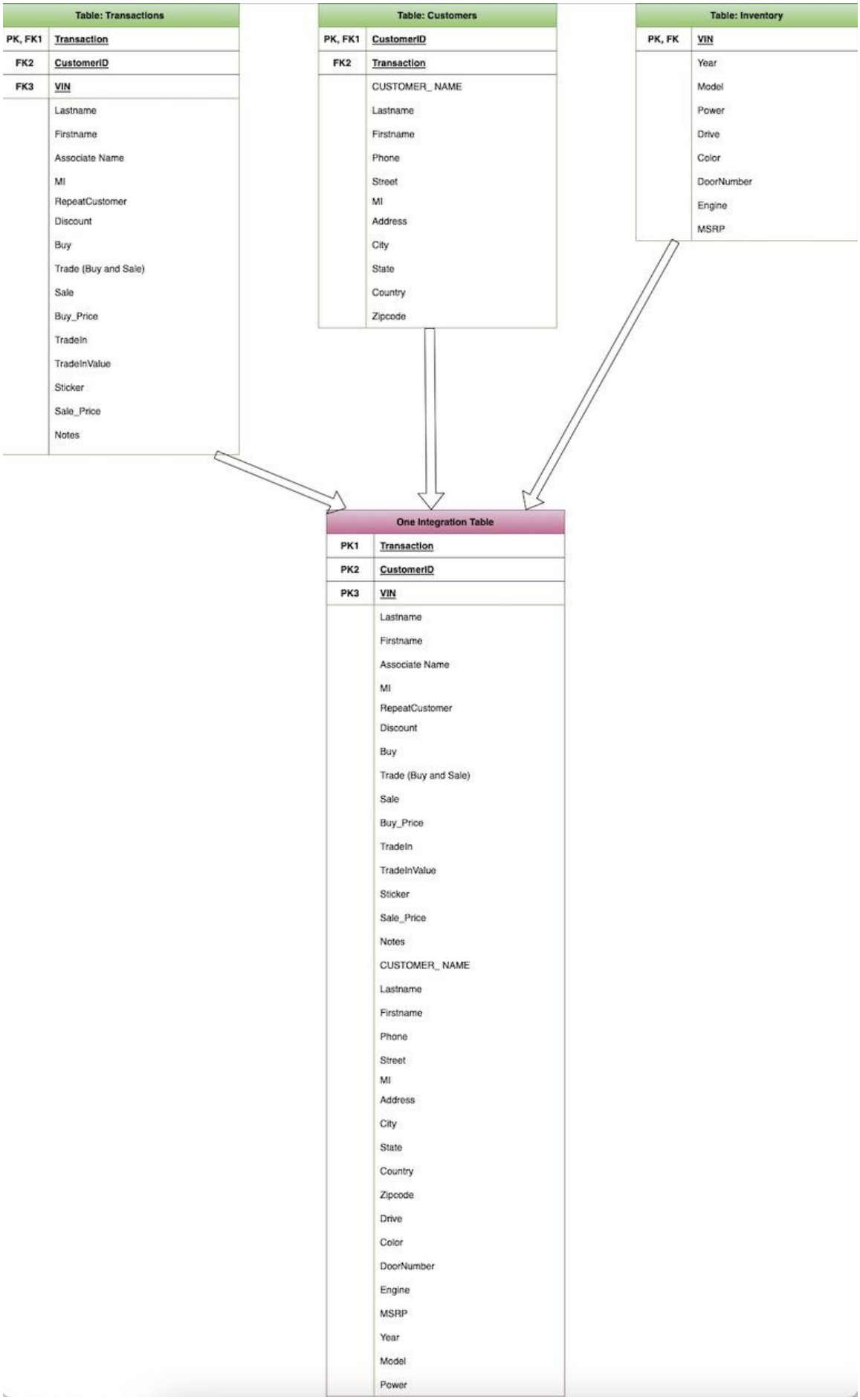


Table: Transactions
PK, FK1 Transaction
FK2 CustomerID
FK3 VIN
Lastname
Firstname
Associate Name
MI
RepeatCustomer
Discount
Buy
Trade (Buy and Sale)
Sale
Buy_Price
TradeIn
TradeInValue
Sticker
Sale_Price
Notes
Table: Customers
PK, FK1 CustomerID
FK2 Transaction
CUSTOMER_ NAME
Lastname
Firstname
Phone
Street
MI
Address
City
State
Country
Zipcode
Table: Inventory
PK, FK VIN
Year
Model
Power
Drive
Color
DoorNumber
Engine
MSRP
One Integration Table
PK1 Transaction
PK2 CustomerID
PK3 VIN
Lastname
Firstname
Associate Name
MI
RepeatCustomer
Discount
Buy
Trade (Buy and Sale)
Sale
Buy_Price
TradeIn
TradeInValue
Sticker
Sale_Price
Notes
CUSTOMER_ NAME
Lastname
Firstname
Phone
Street
MI
Address
City
State
Country
Zipcode
Drive
Color
DoorNumber
Engine
MSRP
Year
Model
Power