

DATABASE PROJECT

Normalization and Functional Dependencies

Submitted To: Dr. Irum Matloob

Submitted By: Laiba Imran(2021-BSE-072)

Laiba Shahid(2021-BSE-051)

Hira Shakeel (2021-BSE-049)

Date: 12th -June-2023

Structured Query Language

Queries typically fall under the category of Data Definition Language (DDL) and Data Manipulation Language (DML) in the context of database management systems. Let's break down the two categories:

- 1. Data Definition Language (DDL):
 - DDL statements are used to define or modify the structure of database objects, such as tables, indexes, or views.
 - Examples of DDL statements include CREATE, ALTER, and DROP.
 - DDL statements do not directly manipulate the data within the tables; they focus on defining the database schema.
- 2. Data Manipulation Language (DML):
 - DML statements are used to manipulate the data stored in the database.
 - Examples of DML statements include SELECT, INSERT, UPDATE, and DELETE.
 - DML statements allow you to retrieve, insert, modify, and delete data from tables.

Created Tables using queries

In SQL, a CREATE query is used to create new database objects such as tables, views, indexes, or procedures.

Following queries are used to create the following tables:

CAR_comm,COMPANY_comm,Sale_comm,CONSUMER_comm,FEATURES and RATINGS_comm.

- Attributes in CAR comm Table are VIN, condition, price, mileage, stock #, Drivetrain.
- Attributes in COMPANY_comm Table are Make, State, Seller_Type.
- Attributes in Sale_comm are Stock_#,Year.
- Attributes in Consumer_comm Table are Consumer_Rating, Consumer_Review_#.
- Attributes in FEATURES Table are Model,
- Exterior_Color,Interior_Color,Engine,MPG,Fuel_Type,Transmission.
 Attributes in RATINGS comm are Consumer Rating,
- Confort_Rating,Interior_Design_Rating,Performance_Rating,Value_For_Money_Rating, Exterior_Styling_Rating,Reliability_Rating.

Table - 1

CAR TABLE

Query:

```
CREATE TABLE CAR_comm
```

```
VIN INT PRIMARY KEY,
condition VARCHAR(50) NOT NULL,
price INT NOT NULL,
mileage INT,
stock_# INT,
Drivetrain VARCHAR(50),
);
Table – 2
    • COMPANY TABLE
Query:
CREATE TABLE COMPANY_comm
Make VARCHAR(50),
State VARCHAR(50),
Seller_Type VARCHAR(50) NOT NULL,
);
Table – 3
    • SALE TABLE
Query:
CREATE TABLE Sale_comm
Stock_# VARCHAR,
Year INT,
);
Table - 4

    CONSUMER TABLE

Query:
CREATE TABLE Consumer_comm
Consumer_Rating VARCHAR(50),
Consumer_Review_# INT,
Table - 5
    • FEATURES TABLE
Query:
CREATE TABLE FEATURES
Model VARCHAR PRIMARY KEY,
Exterior_Color VARCHAR(50) NOT NULL,
Interior_Color VARCHAR(50) NOT NULL,
Engine VARCHAR(100),
MPG INT,
```

```
Fuel_Type VARCHAR(50),
Transmission VARCHAR(50),
);
Table - 6
```

• RATINGS TABLE

Query:

```
CREATE TABLE RATINGS_comm
(
Consumer_Rating VARCHAR(50) NOT NULL,
Confort_Rating VARCHAR(50) NOT NULL,
Interior_Design_Rating VARCHAR(50) NOT NULL,
Performance_Rating VARCHAR(100),
Value_For_Money_Rating VARCHAR(50),
Exterior_Styling_Rating VARCHAR(50),
Reliability_Rating VARCHAR(50),
);
```

Inserted Data Using queries

In SQL, an INSERT query is used to insert data into a table.

By the below given queries the data has been inserted into above created tables.

Table - 1

CONSUMER TABLE INSERTION

Query:

```
USE [Honda_sell_database]

GO

INSERT INTO [dbo].[Consumer_comm]

([Consumer_Rating]

,[Consumer_Review_#])

SELECT

[Consumer_Rating]

,[Consumer_Review_#]

FROM honda_sell_data$
```

Table – 2

• SALE TABLE INSERTION

```
Query:
GO
USE [Honda_sell_database]
GO
INSERT INTO [dbo].[Sale]
     ([Stock_#]
     ,[Year])
  SELECT
           [Stock_#]
     ,[Year],
                FROM honda_sell_data$
GO
USE [Honda_sell_database]
GO
Table - 3
       RATINGS TABLE INSERTION
Query:
INSERT INTO [dbo].[RATINGS_comm]
     ([Consumer_Rating]
     ,[Comfort_Rating]
     ,[Interior_Design_Rating]
     ,[Performance_Rating]
     ,[Value_For_Money_Rating]
     , [Exterior\_Styling\_Rating] \\
     ,[Reliability_Rating])
  SELECT
           [Consumer_Rating]
     ,[Comfort_Rating]
```

```
,[Interior_Design_Rating]
     ,[Performance_Rating]
     ,[Value_For_Money_Rating]
     ,[Exterior_Styling_Rating]
     ,[Reliability_Rating]
                FROM honda_sell_data$
GO
CREATE TABLE COMPANY_comm
Make VARCHAR(50),
State VARCHAR(50),
Seller_Type VARCHAR(50) NOT NULL,
);
Table - 4

    COMPANY TABLE INSERTION

Query:
USE [Honda_sell_database]
GO
INSERT INTO [dbo].[COMPANY_comm]
     ([Make]
     ,[State]
     ,[Seller_Type])
  SELECT
            [Make]
     ,[State]
     ,[Seller_Type]
                FROM honda_sell_data$
```

DESIGN VIEW OF TABLES AND PRIMARY KEYS

Primary Key:

In a relational database, a primary key is a column or a set of columns that uniquely identifies each row in a table. Its purpose is to ensure data integrity and provide a way to uniquely identify and retrieve individual records in a table. Here are some key characteristics of a primary key:

- ➤ Uniqueness: Every value in the primary key column(s) must be unique. No two rows can have the same value(s) in the primary key column(s).
- Non-nullability: The primary key column(s) cannot contain null values. Each row must have a valid, non-null value in the primary key column(s).
- ➤ **Immutability:** The value(s) of the primary key column(s) should generally remain unchanged over the lifetime of a row. Modifying the primary key value(s) is not recommended, as it can lead to data inconsistencies and potential issues with referential integrity.
- ➤ **Indexing:** By default, most database systems automatically create an index on the primary key column(s). This helps improve the performance of searching, joining, and retrieving data from the table.
- ➤ **Relationship establishment:** A primary key can be used to establish relationships (such as foreign keys) with other tables. This enables the creation of logical connections between related data across multiple tables.

Primary Keys in Tables:

- In **CAR** comm table VIN (Vehicle Identification Number) is the Primary key.
- In **COMPANY_comm** table ID is the primary key.
- In **CONSUMER_comm** table Consumer_ID is the primary key.
- In **RATINGS_comm** table RATING_ID is the primary key.
- In **Sale_comm** table sale_ID is the primary key.

Table – 1

• CAR TABLE

Design View:

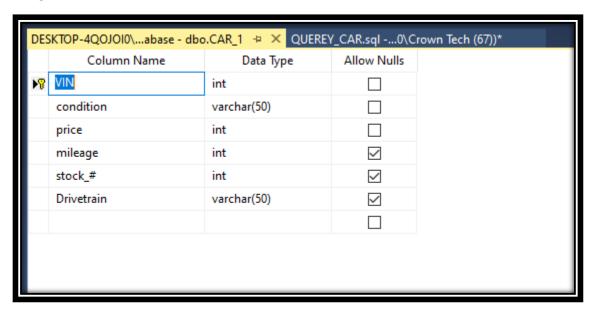


Table – 2

• **COMPANY TABLE**

Design View:

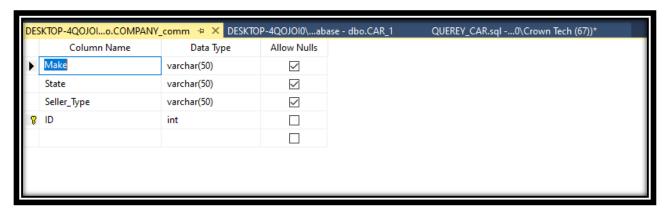


Table - 3

• **CONSUMER TABLE**

Design View:

DES	SKTOP-4QOJOI0o.Consumer	_comm × DESKTO	P-4QOJOIo.CO	OMPANY_comm	DESKTO
	Column Name	Data Type	Allow Nulls		
Þ	Consumer_Rating	varchar(50)	\checkmark		
	Consumer_Review_#	int	\checkmark		
P	Consumer_ID	int			

Table – 4

RATINGS TABLE

Design View:

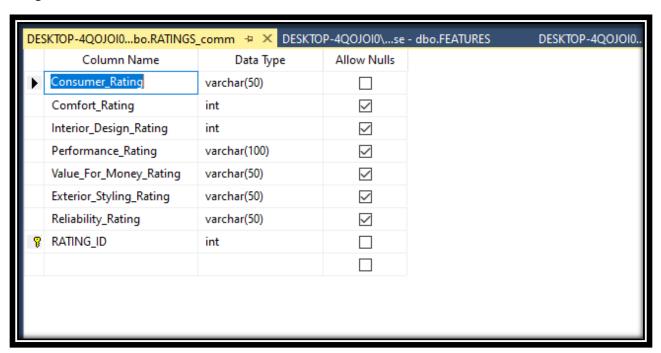
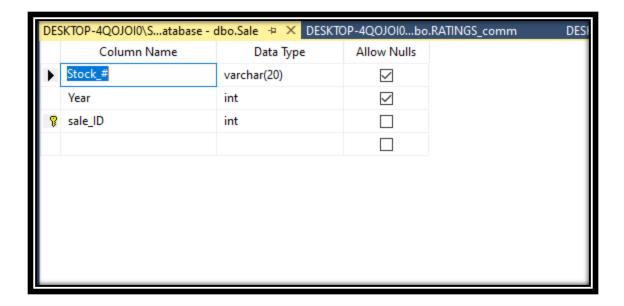


Table – 5

• SALE TABLE

Design View:



REMOVING REDUNDANCY FROM TABLES

By running below query redundancy has been removed from CAR_comm table.

Similarly, the same procedure is applied to all other tables accordingly to remove redundancy.

Table - 1

CAR TABLE

Duplication of data:

No duplication.

Table - 2

COMPANY TABLE

Duplication of data:

Removing redundancy.

Query:

Select Distinct Make, State, Seller_Type

From COMPANY_comm;

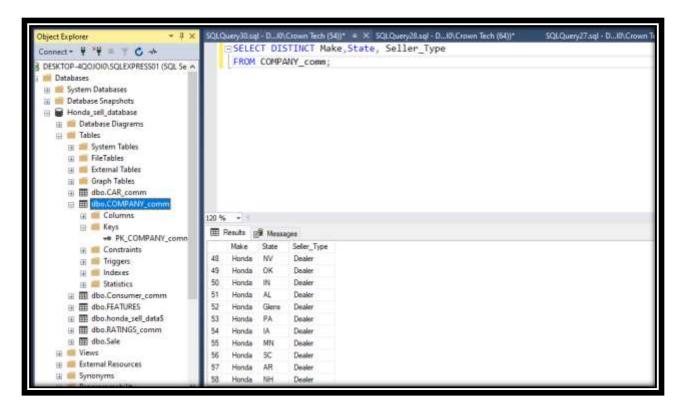


Table - 3

• **CONSUMER TABLE**

Duplication of data:

Removing redundancy.

Query:

DELETE FROM CONSUMER_COMM

WHERE CONSUMER_ID NOT IN(

SELECT MIN(CONSUMER_ID)

FROM CONSUMER_COMM

GROUP BY CONSUMER_RATING, CONSUMER_REVIEW_#

);

SELECT * FROM CONSUMER_COMM;

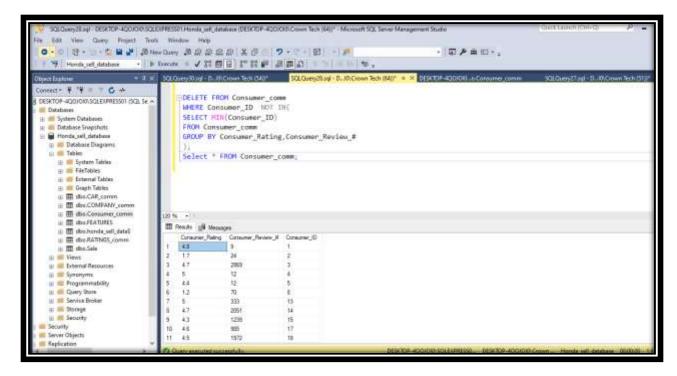


Table - 4

RATINGS TABLE

Duplication of data:

Removing redundancy.

Query:

```
DELETE FROM RATINGS_comm
WHERE RATING_ID NOT IN(
SELECT MIN(RATING_ID)
FROM RATINGS_comm
GROUP BY
Consumer_Rating,Comfort_Rating,Interior_Design_Rating,Performance_Rating,Value_For_Money_
Rating,Exterior_Styling_Rating,Reliability_Rating
);
Select * FROM RATINGS_comm;
```

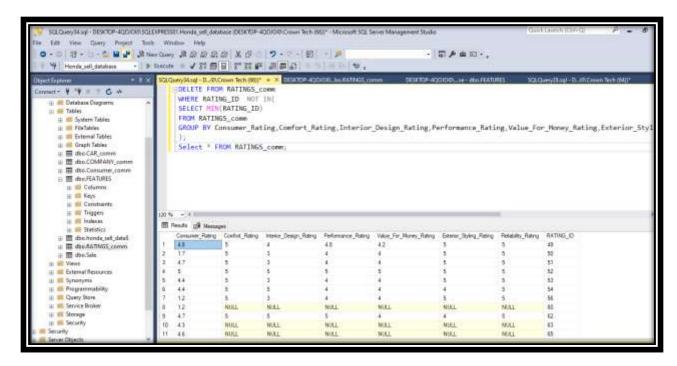


Table - 5

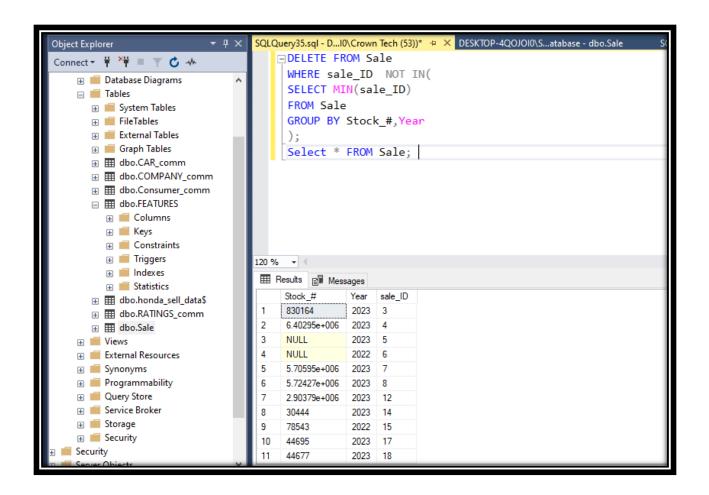
SALE TABLE

Duplication of data:

Removing redundancy.

Query:

```
DELETE FROM Sale
WHERE sale_ID NOT IN(
SELECT MIN(sale_ID)
FROM Sale
GROUP BY Stock_#,Year
);
Select * FROM Sale;
```



Constraints in Tables

Foreign Keys:

Foreign keys serve to establish relationships between tables in a relational database. They ensure referential integrity by enforcing constraints that maintain consistency and prevent inconsistencies in the data.

1. Foreign Key of COMPANY AND CAR (COMPANY HAS A CAR)

```
SQLQuery3.sql - DE...IO\Crown Tech (51))* DESKTOP-4QOJOI...o.COMPANY_comm

SQLQuery2.sql - DE...IO\Crown Tech (78))

alter table [dbo].[CAR_comm]

add constraint FK_Comp_car FOREIGN key(ID) REFERENCES [dbo].[COMPANY_comm](ID);
```



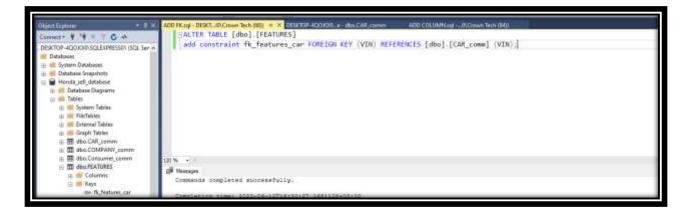
2. Foreign key of FEATURES (CAR HAS FEATURES)

```
ADD FK.sql - DESKT...IO\Crown Tech (68))

DESKTOP-4QOJOIO...e - dbo.CAR_comm

ALTER TABLE [dbo].[FEATURES]

ADD [VIN] nvarchar (255);
```



3. Foreign key of CONSUMER AND CAR (CONSUMER BUYS A CAR)

```
DESKTOP-4QOJOI0...o.Consumer_comm

SQLQuery4.sql - DE...I0\Crown Tech (60))*

SQLQuery3.sql - DE...I0\Crown Tech (54))* * × 4

ADD [Consumer_ID] int not null;
```



4. Foreign key of car and rating (car has rating)



5. Foreign key of CAR AND SALE (CAR IS SALED OUT)





FUNCTIONAL DEPENDENCIES

Functional dependencies describe the relationships between attributes (columns) within a database table. They establish a dependency relationship where the value of one or more attributes determines the value of another attribute.

CAR_comm

```
DESKTOP-4QOJONO...e-dbo.CAR_comm

select [VIN]

from [dbo].[CAR_comm]

group by [VIN]

having count ([VIN]) > 1;
```

```
DESKTOP-4QOJOIO...e - dbo.CAR_comm

| select [condition] |
| from [dbo].[CAR_comm] |
| group by [condition] |
| having count ([condition]]) > 1;
```

```
| Results | Messages | (0 rows affected) | | Completion time: 2023-06-12T19:36:20.2681096+05:00 |
```

```
DESKTOP-4QOJOIO...e - dbo.CAR_comm

| select [mileage] |
| from [dbo].[CAR_comm] |
| group by [mileage] |
| having count ([mileage]) > 1
```

```
| Results | Messages | (0 rows affected) | | Completion time: 2023-06-12T19:48:33.3214668+05:00 |
```

```
DESKTOP-4QOJOIO...e-dbo.CAR_comm

select [stock_#]

from [dbo].[CAR_comm]

group by [stock_#]

having count ([stock_#]) > 1
```

```
| select [Make] | from [dbo].[COMPANY_comm] | group by [Make] | having count ([Make]) > 1
```



COMPANY_comm

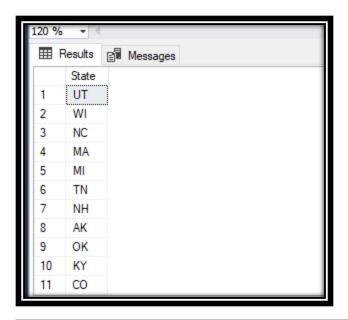
```
DESKTOP-4QOJOI...o.COMPANY_comm

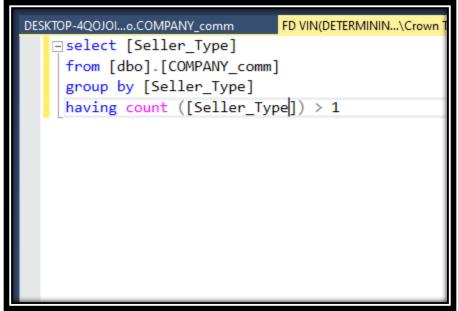
select [State]

from [dbo].[COMPANY_comm]

group by [State]

having count ([State]]) > 1
```



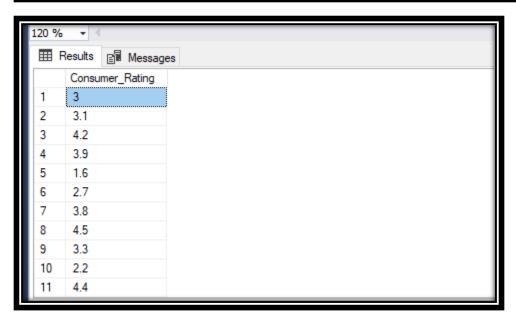




Consumer_comm

```
DESKTOP-4QOJOIO...o.Consumer_comm

| Select [Consumer_Rating] |
| from [dbo].[Consumer_comm] |
| group by [Consumer_Rating] |
| having count ([Consumer_Rating]) > 1
```



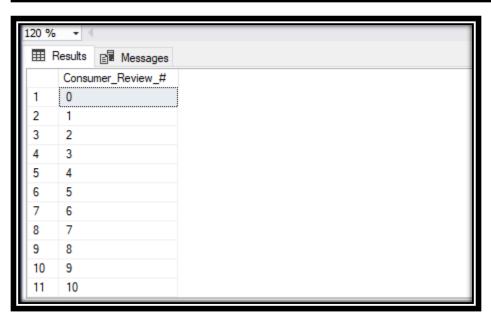
```
DESKTOP-4QOJOI0...o.Consumer_comm

Select [Consumer_Review_#]

from [dbo].[Consumer_comm]

group by [Consumer_Review_#]

having count ([Consumer_Review_#]) > 1
```



```
DESKTOP-4QOJOIO...o.Consumer_comm

select [Consumer_ID]

from [dbo].[Consumer_comm]

group by [Consumer_ID]

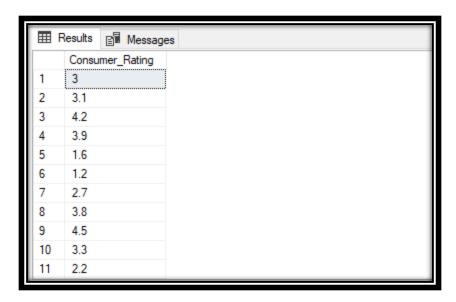
having count ([Consumer_ID]) > 1
```



RATINGS_comm

```
DESKTOP-4QOJOIO...bo.RATINGS_comm

| select [Consumer_Rating] |
| from [dbo].[RATINGS_comm] |
| group by [Consumer_Rating] |
| having count ([Consumer_Rating]) > 1
```





```
DESKTOP-4QOJOIO...bo.RATINGS_comm

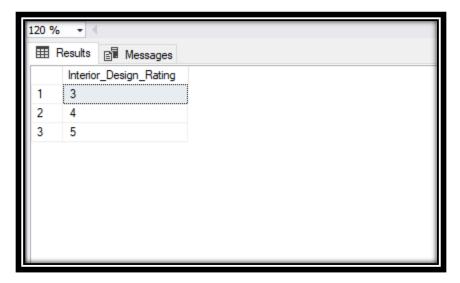
FD VIN(DETERMININ...\Crown Tech (52))* ** X

Select [Interior_Design_Rating]

from [dbo].[RATINGS_comm]

group by [Interior_Design_Rating]

having count ([Interior_Design_Rating]) > 1
```



```
DESKTOP-4QOJOIO...bo.RATINGS_comm

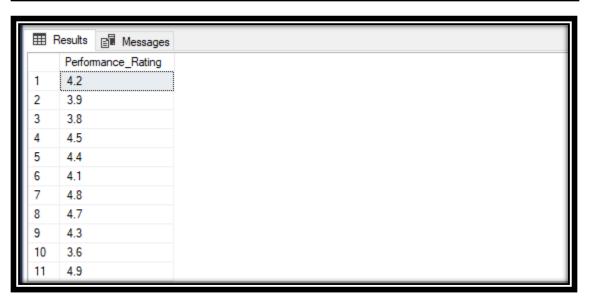
FD VIN(DETERMININ...\Crown Tech (52))* 

Select [Performance_Rating]

from [dbo].[RATINGS_comm]

group by [Performance_Rating]

having count ([Performance_Rating]) > 1
```



```
DESKTOP-4QOJOIO...bo.RATINGS_comm

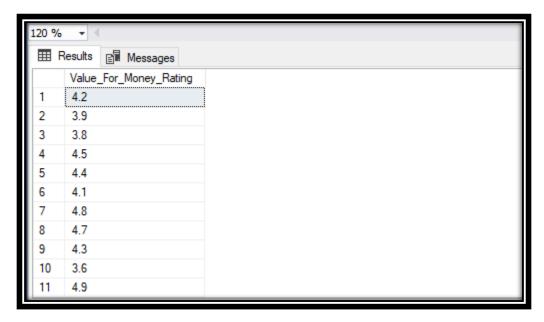
FD VIN(DETERMININ...\Crown Tech (52))* * X

Select [Value_For_Money_Rating]

from [dbo].[RATINGS_comm]

group by [Value_For_Money_Rating]

having count ([Value_For_Money_Rating]) > 1
```



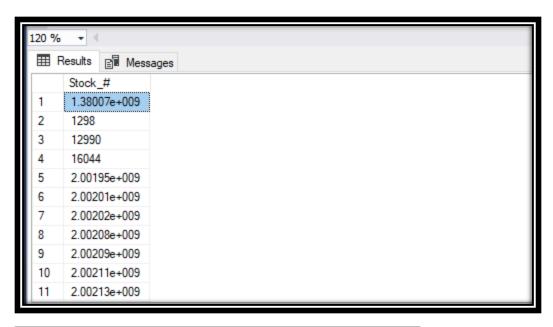
```
■ Results ■ Messages

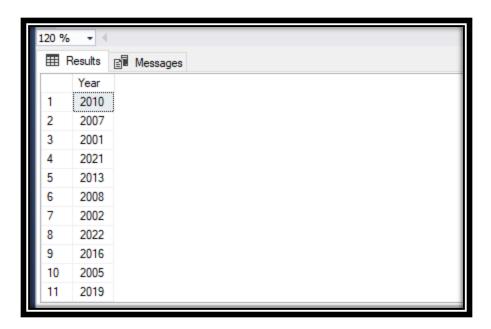
RATING_ID
```

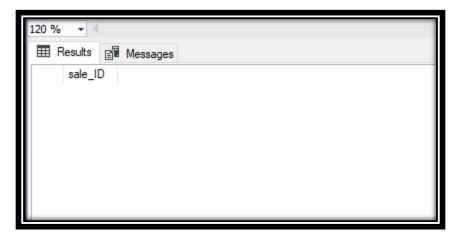
Sale_comm

```
DESKTOP-4QOJOIO\S...atabase - dbo.Sale

| select [Stock_#]
| from [dbo].[Sale]
| group by [Stock_#]
| having count ([Stock_#]) > 1
```







NORMALIZATION

1. First Normal Form: (1NF)

• Atomic Values:

Each column in a table should contain only atomic (indivisible) values. This means that each value in a column should represent a single piece of data. If a column contains multiple values or a list of values, it should be split into separate columns.

Unique Column Names:

Each column in a table should have a unique name, and no two columns should have the same name. This ensures that each column is uniquely identifiable.

• Rows and Columns:

Each row in a table should represent a unique record or entity, and each column should contain a single attribute of that entity. There should be no duplicate rows, and each column should have a distinct purpose.

Primary Key:

Each table should have a primary key that uniquely identifies each row in the table. The primary key should consist of one or more columns that have unique values for each row. This helps in distinguishing one row from another.

No Repeating Groups:

A table should not contain repeating groups or arrays of values. Each column should contain only single values, and if there are multiple values related to an entity, they should be placed in separate rows.

Key Points:

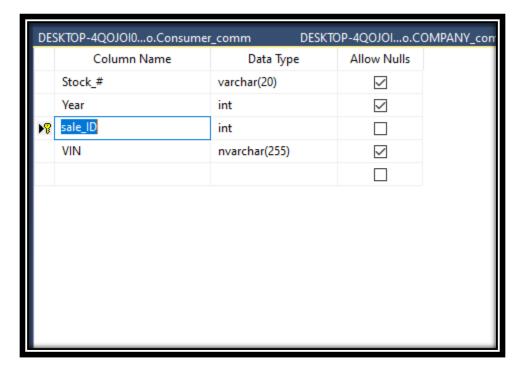
In short there should not be multi valued attribute for any attribute and Unique record should exist.

	Column Name	Data Type	Allow Nulls
₽₽	VIN	nvarchar(255)	
	condition	varchar(50)	
	price	int	
	mileage	int	$\overline{\checkmark}$
	stock_#	int	\checkmark
	Drivetrain	varchar(50)	$\overline{\checkmark}$
	ID	int	
	Consumer_ID	int	

DE:	SKTOP-4QOJOIo.COMPANY	_comm	P-4QOJOI0e - (dbo.CAR_
	Column Name	Data Type	Allow Nulls	
×	Make	varchar(50)	\checkmark	
	State	varchar(50)	\checkmark	
	Seller_Type	varchar(50)	\checkmark	
P	ID	int		

DES	SKTOP-4QOJOI0o.Consumer	_comm	P-4QOJOIo.C0	DMPANY_com
	Column Name	Data Type	Allow Nulls	
Þ	Consumer_Rating	varchar(50)	\checkmark	
	Consumer_Review_#	int	\checkmark	
P	Consumer_ID	int		

DES	SKTOP-4QOJOI0o.Consumer	r_comm DESKT0	DP-4QOJOIo.COMPA
	Column Name	Data Type	Allow Nulls
	Consumer_Rating	varchar(50)	
	Comfort_Rating	int	\checkmark
	Interior_Design_Rating	int	\checkmark
	Performance_Rating	varchar(100)	$\overline{\checkmark}$
	Value_For_Money_Rating	varchar(50)	\checkmark
	Exterior_Styling_Rating	varchar(50)	\checkmark
	Reliability_Rating	varchar(50)	\checkmark
₽®	RATING_ID	int	
	VIN	nvarchar(255)	\checkmark



Results:

So, in our database there is no multivalued attribute so, it is in *first normal form*.

- 2. Second Normal form:(2NF)
- Meet the criteria for 1NF:

The table must already satisfy the criteria for 1NF, including having atomic values, unique column names, rows and columns, and a primary key.

• Single-Attribute Primary Key:

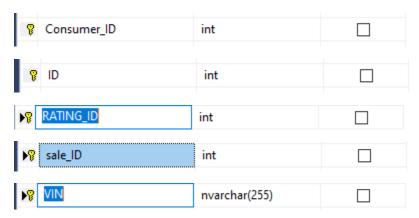
The table should have a single-attribute primary key, meaning it consists of only one column. If the primary key is composite (consisting of multiple columns), it should be broken down into separate tables, each with a single-column primary key.

• Partial Dependencies:

Ensure that each non-key column in the table depends on the entire primary key, not just a part of it. If any non-key column depends on only a subset of the primary key, it should be moved to a separate table along with the subset of the primary key it depends on.

Results:

So, as the tables given above there is *no partial dependencies* in this database because no composite primary key exists.



- 3. Third normal form: (3NF)
- Meet the criteria for 2NF:

Table should be in 2nf.

• Transitive Dependency:

No transitive dependency exists.

• Non-Prime Attributes:

Non-prime attributes should not depend on other non-prime

Results:

All the tables in this Honda data sell *only prime attributes can determine all other attributes*.

4. (Boyce-Codd Normal Form): BCNF

BCNF stands for Boyce-Codd Normal Form, which is a higher level of database normalization.

• Check on Prime attributes:

it is strictly checked that only attributes.	prime attributes or the candidate can	determine all the other non-prime
As this database is in 3nf and	also in BCNF.	
	L.H.S =ck, sk	