CAR SALES DATA

-- this step of creating car sales table to help in data importation

CREATE TABLE car\_sales\_data (

Car\_Name VARCHAR(255),

Year INT,

Present\_Price DECIMAL(10, 2),

Kms\_Driven INT,

Fuel\_Type VARCHAR(50),

Seller\_Type VARCHAR(50),

Transmission VARCHAR(50),

Owner INT,

Selling\_Price DECIMAL(10, 2)

);

-- data table created and have data in my databases

-- next step is querying the data to do EDA

-- This is a single-line comment

-- This selects all records from the car\_sales\_data table

SELECT \* FROM car\_sales\_data;

-- check missing values in the dataset

SELECT

COUNT(\*) AS total\_rows,

SUM(CASE WHEN Car\_Name IS NULL THEN 1 ELSE 0 END) AS missing\_car\_name,

SUM(CASE WHEN Year IS NULL THEN 1 ELSE 0 END) AS missing\_year,

SUM(CASE WHEN Present\_Price IS NULL THEN 1 ELSE 0 END) AS missing\_present\_price,

SUM(CASE WHEN Kms\_Driven IS NULL THEN 1 ELSE 0 END) AS missing\_kms\_driven,

SUM(CASE WHEN Fuel\_Type IS NULL THEN 1 ELSE 0 END) AS missing\_fuel\_type,

SUM(CASE WHEN Seller\_Type IS NULL THEN 1 ELSE 0 END) AS missing\_seller\_type,

SUM(CASE WHEN Transmission IS NULL THEN 1 ELSE 0 END) AS missing\_transmission,

SUM(CASE WHEN Owner IS NULL THEN 1 ELSE 0 END) AS missing\_owner,

SUM(CASE WHEN Selling\_Price IS NULL THEN 1 ELSE 0 END) AS missing\_selling\_price

FROM car\_sales\_data;

-- there are 10k reords and no missing values present in the data

-- check for duplicate values

SELECT

Car\_Name,

Year,

COUNT(\*) AS duplicate\_count

FROM car\_sales\_data

GROUP BY Car\_Name, Year

HAVING COUNT(\*) > 1

Order by duplicate\_count DESC;

-- clearly the majority of cars are Cty 2015 and city 2016.

-- Check for Outliers

SELECT \*

FROM car\_sales\_data

WHERE Selling\_Price > 12000;

--- Extreme values by KMs driven

SELECT \*

FROM car\_sales\_data

WHERE Kms\_Driven > 210000

Order by Kms\_driven;

-- Car with the highest kms driven is Honda karizma with 213104 kms and costing 6705.05

-- consistency of data

SELECT DISTINCT Fuel\_Type FROM car\_sales\_data;

--there is diesel, petrol and CNG cars

SELECT DISTINCT Seller\_Type FROM car\_sales\_data;

-- the seller type in this data are individual and dealers

SELECT DISTINCT Transmission FROM car\_sales\_data;

-- the transmission type for majority of these cars is manual or automatic

-- data looks fine we can now import it to python jupyter notebook for ml

-- Thank You