Usecase 4.2

By: eng. Esraa Madhi

So we defined data science as: It's the process of asking interesting questions, and then answering those questions using data.

For any **Data project** we will go through these steps:

1. Defining the Problem Statement
2. Collecting Data
3. Data Quality Checking and Remediation
4. Exploratory Data Analysis
5. Building Machine Learning Models
6. Model Evaluation
7. Communicating Results
8. Model Deployment
9. Model Performance Maintenance in Production

Usecase 4.2

Step 1: Defining the Problem Statement

- The first phase of our project, we gained insights into the used car market in Saudi Arabia, uncovering key aspects such as:
  - [x] ~~The region within Saudi Arabia that holds the largest share of the used car market.~~
  - [x] ~~The most frequently listed car brand for sale within the observed timeframe.~~
  - [x] ~~The car features that have the most substantial impact on a used car's price.~~
  - [x] ~~The typical price range for used vehicles around four years of age.~~

- In the 2nd phase:
  - [x] ~~we developed a predictive model for used car prices~~
- For the 3rd phase, our objective is to develop a predictive model for used car prices **category**

Step 2: Collecting Data

Used cars data is collected from scrapping online website called Sayarah for selling cars in 2023 and posted in kaggle

Step 3: Data Quality Checking and Remediation

Done in the notebook

Step 4: Exploratory Data Analysis

Done in the notebook

Step 5: Building Machine Learning Models

To do in the notebook. But before constructing your predictive model using the used car dataset sourced from the Sayarah website in 2023 and shared on Kaggle, you'll need to perform some feature engineering on the `price` column to

create a new target variable: `car_price_category`. This new feature will categorize car prices into three distinct groups:

- **Cheap_Price**: Cars that fall below the 35th percentile of the dataset's price distribution.
- **Good_Price**: Cars with prices ranging between the 35th and 75th percentiles.
- **High_Price**: Cars that exceed the 75th percentile in terms of price.

This engineered feature will serve as the new target variable for your machine learning model.

Step 6: Model Evaluation

To do in the notebook

Step 7: Communicating Results

To do in the notebook

Step 8: Model Deployment

Not applicable

Step 9 : Model Performance Maintenance in Production

Not applicable