📱 Cellphone Price Prediction — Domain Analysis

🎯 Problem Statement

The objective of this project is to **predict the price range of mobile phones** based on their hardware specifications.
This is a **multi-class classification** task with four price categories.

🎯 Target / Dependent Variable

`price_range`

- **0** → Low cost
- **1** → Medium cost
- **2** → High cost
- **3** → Very high cost

This is the class label the model must predict.

📩 Input / Independent Variables

Below are all independent features with their exact meanings:

1. Performance & Hardware

- **battery_power** → Total energy a battery can store at one time (mAh)
- **clock_speed** → Speed at which the microprocessor executes instructions
- **int_memory** → Internal memory in gigabytes
- **n_cores** → Number of cores in the processor
- **ram** → Random Access Memory in megabytes
- **m_dep** → Mobile depth in cm
- **mobile_wt** → Weight of the mobile phone

2. Connectivity & Technology

- **blue** → Has Bluetooth or not (0/1)
- **dual_sim** → Has dual SIM support or not (0/1)
- **four_g** → Has 4G or not (0/1)
- **three_g** → Has 3G or not (0/1)
- **wifi** → Has WiFi or not (0/1)
- **touch_screen** → Has touchscreen or not (0/1)

3. Camera

- **fc** → Front camera megapixels
- **pc** → Primary camera megapixels

4. Display

- **px_height** → Pixel resolution height
- **px_width** → Pixel resolution width
- **sc_h** → Screen height in cm
- **sc_w** → Screen width in cm

5. Battery Usage

- **talk_time** → Longest time a single battery charge will last during usage

📊 Feature Impact (Domain Understanding)

Some features affect price more than others.

🔥 High-Impact Features

- **RAM** — Biggest contributor to price category
- **Battery Power**
- **Screen Resolution** (`px_height × px_width`)
- **Internal Memory**
- **Processor Cores / Clock Speed**

⚠ Low-Impact Features

- **Bluetooth**, **WiFi**, **Dual SIM**, **Touchscreen**, **Talk Time**, **3G/4G**
  These are near-universal features and don't strongly differentiate price categories.

📌 Overall Summary

- The dataset is ideal for **classification models** (Random Forest, XGBoost, SVM).
- **price_range** is a 4-class target label.
- Hardware features like RAM and screen resolution dominate prediction power.
- Connectivity features contribute very little.
- This is a good academic ML dataset but not a literal real-world pricing system.