**FINDINGS**

1. **Data Characteristics**:
   * The dataset contains pizza-related features such as company, diameter, topping, variant, size, extra sauce, extra cheese, and extra mushrooms.
   * The dataset required significant preprocessing to clean and format the data correctly.
   * **The dataset has 129 rows and 9 columns.**

**UNIVARIATE ANALYSIS**

1. **Price Distribution:**
   * Prices were adjusted from Indonesian Rupiah to Kes, showing a varied distribution of pizza prices.
   * The price distribution is skewed to the right, with a majority of prices falling between **20,000 and 80,000.**
   * The **mode is around 40,000**, and there are a few **outliers above 120,000**. Overall, most items are affordable, with fewer expensive items.
2. **The diameter distribution:**
   * The diameter distribution is skewed to the right, with most pizzas being small **(8.5-14 inches).**
   * There are a few outliers **(20-22 inches).**
3. **Topping distribution**
   * Overall, chicken, mushrooms, and mozzarella are the dominant toppings with 29, 22, and 21 occurrences, respectively in the dataset.
   * Less popular toppings: Tuna, vegetables, and meat are less popular, with 9, 9, and 8 occurrences, respectively.
4. **Variant distribution**
   * The classic pizza variant is the most popular, followed by "meat lovers," "double\_mix," "crunchy," and "new\_york." Other variants have fewer occurrences.
5. **Size distribution**
   * Most popular size is medium size with 41 occurrences.
   * Other popular sizes: "Small," "large," and "reguler" are also relatively popular, with 22, 21, and 20 occurrences, respectively.Less popular sizes: "Jumbo" and "XL" are less popular, with 15 and 10 occurrences, respectively.
6. **Extra Sauce & Extra Cheese distribution**
   * Over 60% of customers preferred extra sauce and **Extra Cheese** on the pizza.

# **BIVARIATE ANALYSIS**

### **Comparing the target variable (Price) with company,topping and size**

1. **Price By Company:**
   * Company A has the highest average pizza price, while Company E has the lowest. Companies B, C, and D have similar prices.
   * This information can help understand pricing strategies and make informed purchasing decisions.
2. **Price By Toppings:**
   * The price of a pizza is influenced by the topping, with chicken and mozzarella being more expensive options.
   * Sausage and onion pizzas have the lowest prices.
   * There is some variation in prices within each topping category, as indicated by the size of the boxes and whiskers.
   * Black pepper and beef pizzas have some outliers with very high prices.
3. **Price By Size:**
   * Jumbo pizzas tend to have the highest prices, followed by regular and large pizzas.
   * Small and medium pizzas have lower prices, with small pizzas having the lowest average price.
   * There is some variation in prices within each size category, as indicated by the size of the boxes and whiskers.
   * Large and XL pizzas have some outliers with very high prices**.**

**These findings can be useful for understanding the pricing structure,customer preferences and making informed decisions about pricing strategies or target markets.**

1. **Find the most expensive pizza**
   * The most expensive pizza in the dataset is from Company A, costs 162,092.
   * Has a diameter of 18 inches, is topped with mozzarella, has the double\_signature variant, is jumbo sized, has extra sauce, but no extra cheese or extra mushroom**s.**
2. **Feature Importance Relationships:**
   * Diameter and topping showed noticeable relationships with the price of the pizzas.
3. **Model Performance:**
   * The Gradient Booster Regressor model provided the best R2 score of 93%, indicating it as the most accurate model for predicting pizza prices.
   * Feature importance analysis highlighted diameter and topping as significant contributors to the model's predictions.

14. Price Prediction with the saved model

I tested the model to predict the price of the pizza with the following parameter

* + 'company':1,
  + 'diameter':22.0,
  + 'topping':2,
  + 'variant':8,
  + 'size':1,
  + 'extra\_sauce':1,
  + 'extra\_cheese':1,
  + 'extra\_mushrooms':1

**The model predicted the price to be 151,782.52**

**LINK TO THE DATA**

<https://www.kaggle.com/datasets/knightbearr/pizza-price-prediction?select=pizza_v2.csv>