Voici quelques points généraux que l'on peut déduire des histogrammes :

1. **Répartition des Données :**
   * Pour toutes les catégories (Robusta, Arabica, Espresso, Lungo, Latte, Cappuccino), la majorité des valeurs se concentrent dans les premières plages de valeurs (0 à environ 10 000).
   * Les données montrent une forte asymétrie à droite, indiquant que la majorité des observations ont des valeurs basses, tandis que quelques observations ont des valeurs très élevées.

**Distribution Asymétrique :**

* Les histogrammes montrent une distribution fortement asymétrique à droite pour chaque type de café. Cela signifie que la majorité des données se trouvent dans les petites valeurs, avec quelques valeurs très élevées.
* Ce type de distribution n'est pas anormal en soi, surtout dans des contextes où il y a une grande variation entre les observations.

Voici une analyse basée sur ces graphiques de densité :

1. **Distribution Asymétrique :**
   * Les graphiques de densité montrent également une forte asymétrie à droite, similaire à ce que nous avons observé dans les histogrammes.
   * La majorité des données se concentre autour des valeurs faibles, avec des queues longues s'étendant vers des valeurs élevées.
2. **Comparaison des Catégories :**
   * **Robusta** et **Espresso** ont des pics très similaires, avec des densités maximales autour de valeurs basses (proches de 0) et des queues longues qui s'étendent jusqu'à environ 100 000.
   * **Arabica** montre une densité maximale un peu plus large mais suit une distribution similaire aux autres types de café.
   * **Lungo**, **Latte**, et **Cappuccino** montrent des pics de densité autour de valeurs basses, avec des queues plus courtes par rapport aux autres types de café. Leurs densités diminuent rapidement après environ 10 000 à 20 000.
3. **Densité Maximale :**
   * Les densités maximales sont très élevées pour des valeurs proches de zéro, indiquant une forte concentration de données dans cette plage.
   * La densité diminue rapidement à mesure que les valeurs augmentent.

Pour analyser les données à partir des histogrammes fournis, nous allons créer des box plots (boîtes à moustaches) qui visualiseront la distribution des données pour chaque catégorie de café (Robusta, Arabica, Espresso, Lungo, Latte, Cappuccino). Ces box plots nous permettront de voir les médianes, les quartiles, les valeurs minimales et maximales, ainsi que les éventuels outliers.

### **Étapes de l'Analyse**

1. **Création des Box Plots :**
   * Utilisation des données fournies pour créer des box plots.
   * Observation des éléments clés des box plots : médiane, quartiles, moustaches, et outliers.
2. **Interprétation des Box Plots :**
   * Analyse de la dispersion et de la centralité des données.
   * Identification des outliers et leur impact sur la distribution.

### **Exemple de Code Python pour Créer des Box Plots**

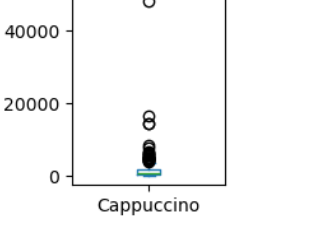
Voici un exemple de code Python pour créer et visualiser les box plots pour les catégories de café :

### **Interprétation des Box Plots**

Voici ce que nous allons rechercher dans les box plots pour chaque catégorie de café :

1. **Médiane (Ligne au milieu de la boîte) :**
   * Indique la valeur centrale des données 50000.
2. **Quartiles (Q1 et Q3) :**
   * La boîte montre l'intervalle interquartile (IQR), qui contient les 50% des valeurs centrales.
   * Q1 25000
   * Q3 75000
3. **Moustaches (Whiskers) :**
   * Indiquent la plage dans laquelle se trouvent la majorité des données, souvent définie par 1,5 fois l'IQR en dessous de Q1 et au-dessus de Q3.

Par exemple



Par exemple la la majorite des moustaches sont entre 0 et 20000

### **Analyse de Corrélation**

L'analyse de corrélation permet de comprendre la relation entre différentes variables quantitatives. Les coefficients de corrélation varient entre -1 et 1 :

* **1** indique une corrélation positive parfaite.
* **-1** indique une corrélation négative parfaite.
* **0** indique aucune corrélation.

**Robusta** :

* Corrélation modérée avec **Lungo** (0.35).
* Corrélation faible avec **Cappuccino** (0.24) et **Arabica** (0.10).
* Corrélation négative très faible avec **Latte** (-0.10).
* Aucune corrélation significative avec **Espresso** (-0.01).

**Arabica** :

* Corrélation forte avec **Espresso** (0.73) et **Latte** (0.66).
* Corrélation modérée avec **Cappuccino** (0.41).
* Corrélation faible avec **Lungo** (0.12).
* Corrélation très faible avec **Robusta** (0.10).

**Espresso** :

* Corrélation très forte avec **Latte** (0.92).
* Corrélation forte avec **Arabica** (0.73).
* Corrélation faible avec **Cappuccino** (0.21).
* Corrélation très faible avec **Robusta** (-0.01) et **Lungo** (-0.04).

**Lungo** :

* Corrélation modérée avec **Cappuccino** (0.39) et **Robusta** (0.35).
* Corrélation faible avec **Arabica** (0.12).
* Corrélation très faible avec **Espresso** (-0.04) et **Latte** (-0.13).

**Latte** :

* Corrélation très forte avec **Espresso** (0.92).
* Corrélation forte avec **Arabica** (0.66).
* Corrélation très faible avec **Robusta** (-0.10), **Lungo** (-0.13), et **Cappuccino** (0.07).

**Cappuccino** :

* Corrélation modérée avec **Arabica** (0.41), **Lungo** (0.39), et **Robusta** (0.24).
* Corrélation faible avec **Espresso** (0.21).
* Corrélation très faible avec **Latte** (0.07).