**لمحة:**

تم تحميل الداتاسيت من موقع kaggle وهنا [رابط](https://www.kaggle.com/datasets/ahmedabbas757/coffee-sales)، وهي بيانات لمبيعات متجر قهوة يحتوي على عدة فروع

لم أجد قيم null في الجدول، ولكن في حال وجدنا قيم null سيتم حل المشكلة بعدة طرق:

1. حذف الأسطر إذا كان عدد القيم null قليل.

2. استبدال قيم null ب mean إذا كان توزيع القيم غاوسي، وإذا غير غاوسي نستخدم median، اذا كانت القيم numerical.

3. أو في حالة كانت القيم categorical يتم استبدالها بالقيم الاكثر تكرارا.

4. يمكن استخدام التنبؤ (Imputation) للقيم المفقودة باستخدام خوارزميات مثل K-Nearest Neighbors (KNN) أو الانحدار، إذا كانت البيانات تحتوي على الكثير من القيم المفقودة بدلاً من الاعتماد فقط على الأكثر تكرارًا.

5. استبدال القيم المفقودة باستخدام قيم ثابتة أو فئات محددة: يمكن استبدال القيم المفقودة بقيمة ثابتة مثل 0 أو "غير متاح" أو فئة "مجهول" في حالة المتغيرات categorical.

من خلال هذه التعليمية**:**

df.groupby(['product\_category','product\_type']).size().sort\_values(ascending=False)

ستعطيني كل category ما هي أكثر type  مبيعا، مثلا:

 فئة Tea من نوع Brewed Chai tea تم بيعه 17183 مرة

فئة Coffee من نوع Gourmet brewed coffee تم بيعه 16912 مرة

من خلال هذه التعليمية**:**

df.groupby(['transaction\_date','product\_category']).size().sort\_values(ascending=False)

أكثر الأيام مبيعًا كانت لأصناف القهوة (Coffee) في شهر يونيو.

بعض الأيام كان فيها مبيعات قليلة جدًا (مثلاً: صنف Branded أو Packaged Chocolate تم بيعه مرة واحدة فقط في ذلك اليوم).

هذا التحليل مفيد لمعرفة أفضل الفترات مبيعًا حسب كل فئة.

من خلال هذه التعليمية**:**

**df\_transaction\_date=df['transaction\_date'].dt.to\_period('M').astype('str').sort\_values(ascending=False)**

**df.groupby([df\_transaction\_date,'product\_category']).size().sort\_values(ascending=False)**

شهر يونيو 2023 كان أكثر الشهور نشاطًا، خصوصًا لفئة القهوة (Coffee) ثم الشاي (Tea).

فئات المنتجات الأكثر مبيعًا عمومًا:

* Coffee
* Tea
* Bakery
* Drinking Chocolate
* Flavours

فئات المنتجات ذات المبيعات الأقل تشمل:

* Packaged Chocolate
* Branded
* Loose Tea
* Coffee beans

من خلال هذه التعليمية**:**

**df\_transaction\_date=df['transaction\_date'].dt.to\_period('M').astype('str').sort\_values(ascending=False)**

**df.groupby([df\_transaction\_date,'product\_category'** **'product\_type']).size().sort\_values(ascending=False)**

المنتج الأعلى مبيعًا في يونيو 2023 هو:

* Gourmet brewed coffee (عدد 4107 مبيعات)
* يليه Brewed Chai tea (4068 مبيعات)

منتجات مثل Green beans و Clothing كانت الأقل مبيعًا خلال بعض الأشهر.

هذا التحليل يساعد في**:**

* تحديد المنتجات الرائجة شهريًا.

من خلال هذه التعليمية**:**

**df.groupby(['transaction\_qty','store\_location']).size().sort\_values(ascending=False)**

تحليل العلاقة بين الكميات المباعة والموقع الجغرافي للمتجر.  
تحديد المواقع التي تسجل مبيعات مرتفعة بكميات معينة.

من خلال هذه التعليمية**:**

**df\_transaction\_time = pd.to\_datetime(df['transaction\_time'], format='%H:%M:%S').dt.hour.sort\_values(ascending=False)**

**df.groupby([df\_transaction\_time, 'product\_category', 'product\_type']).size().sort\_values(ascending=False)**

* الساعة **10** صباحًا تشهد أعلى مبيعات، وخاصةً**:**
  + **Barista Espresso (2081** عملية بيع**)**
  + **Brewed Chai tea (2063** عملية بيع**)**
* الأنواع ذات المبيعات المنخفضة ظهرت في ساعات متأخرة أو مبكرة، مثل الساعة **6** أو **18** أو **20.**

هذا التحليل يساعد على**:**

* معرفة الذروة الزمنية للمبيعات**.**
* اتخاذ قرارات تسويقية ذكية **(**مثل العروض في ساعات الذروة**).**

من خلال هذه التعليمية**:**

**df.groupby(['transaction\_qty','store\_location']).size().sort\_values(ascending=False)**

تحليل العلاقة بين الكميات المباعة والموقع الجغرافي للمتجر.  
تحديد المواقع التي تسجل مبيعات مرتفعة بكميات معينة.

من خلال هذه التعليمية**:**

**df.groupby(['product\_category', 'product\_type'** **,'unit\_price']).size().sort\_values(ascending=False)**

Brewed Chai tea من فئة Tea بسعر 2.50 هو الأكثر مبيعًا بـ 5981 عملية.

بعض المنتجات (مثل Barista Espresso) تباع بأكثر من سعر، مما يدل على وجود اختلافات في الحجم أو العروض.

بعض المنتجات مثل Housewares أو Clothing تُباع بأعداد قليلة وبأسعار أعلى.

**من خلال هذه التعليمية:**

**df.groupby(['product\_category', 'product\_type' ,'product\_detail']).size().sort\_values(ascending=False)**

Chocolate Croissant من فئة Bakery ونوع Pastry بعدد 3076 مرة.

Earl Grey Rg من نوع Brewed Black tea بعدد 3053 مرة.

التحليل يوضح أداء كل منتج بتفصيله الدقيق، مما يساعد في اتخاذ قرارات دقيقة حول تفضيلات العملاء