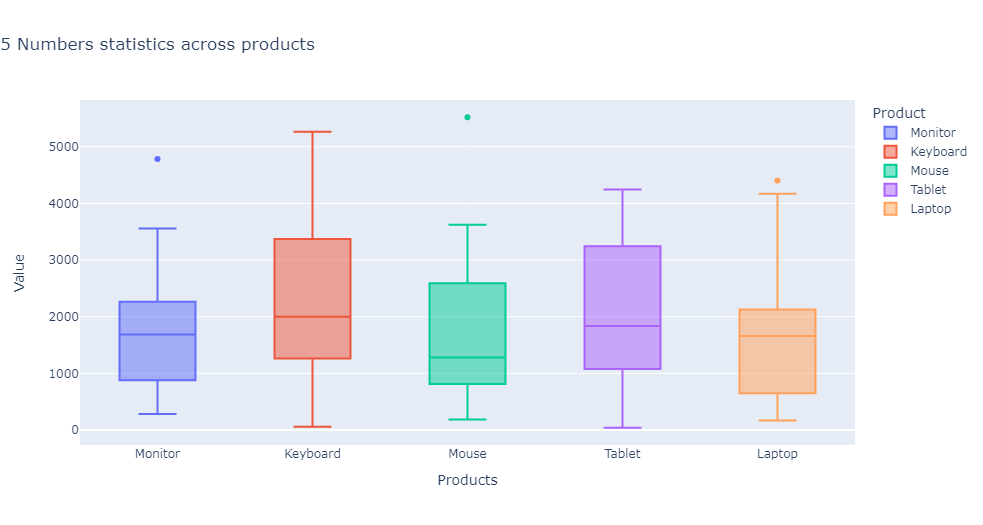
**هذا الملخص مقدم من: سهيل سلّام**

| # | اسم العمود | ملخص تنظيف البيانات | عدد التكرارات | عدد الفراغات | عدد التعديلات |
| --- | --- | --- | --- | --- | --- |
| 1 | Order | تم إعادة الترقيم بناء على قيم عمود Order Date، بدءاً من الرقم 101 ولغاية 200  تم جعل هذا الحقل بصيغة رقم بدون فواصل عشرية | لا يوجد | لا يوجد | 100 سجل |
| 2 | Product | تم ملئ الفراغات حسب قرب سعر المنتج لمعدل سعر المنتجات التي كانت كالتالي:  **Product Average Unit Price**  Keyboard 1956.68  Laptop 1648.60  Mouse 1519.03  Monitor 1633.35  Tablet 1834.03  أما الفراغات فكانت حسب اسعارها المذكورة في حقل “CR Price per Unit” وبعد إعادة تعبئتها كالتالي:  **Product CR Price per Unit**  Mouse 972  Laptop 1666.666667  Mouse 1133.333333  Mouse 1444.814815  Tablet -3824.444444  Mouse 260  Mouse 343.5087719  Mouse 991  يرجى العلم أن السعر المستخدم “CR Price per Unit” هو سعر الوحدة بعد إعادة احتسابه كما سيأتي لاحقاً.  تم توحيد مسمى labtop ليصبح Laptop  تم جعل هذا الحقل حقل نصي | التكرار لا معنى له هنا | 8 | توحيد مسمى Laptop عدد (1) |
| 3 | Category | تم توحيد الفئات حسب اسم المنتج باعتماد التالي:  **Product Category**  Keyboard Accessories  Mouse Accessories  Laptop Peripherals  Tablet Peripherals  Monitor Electronics  تم جعل هذا الحقل حقل نصي |  |  | |  |  | | --- | --- | | **Product** | **Count** | | Keyboard | 21 | | Mouse | 32 | | Laptop | 16 | | Tablet | 10 | | Monitor | 21 | |
| 4 | Customer ID | يوجد تكرار لرقم الزبون في كل السجلات عدا السجلات الفارغة، ولذلك تم أعادة ترقيم رقم الزبون اعتماداً على الاسم بدءاً من رقم 1827 ولغاية 1832حسب أسماء الزبائن وهم (ستة).  Ahmad 17 تكرار اخذ الترقيم 1827  Ali 28 تكرار منهم 2 كانوا فراغ أخذ الترقيم 1828  Fatima 21 تكرار منهم 2 كانوا فراغ اخذت الترقيم 1829  Mona 7 تكرارات منهم 1 كان فارغاً أخذت الترقيم 1830  Sarah 5 تكرارات منهم 1 كان فارغاً أخذ الترقيم 1831  Yousef 10 تكرارات أخذ الترقيم 1832  الفراغات الباقية وعددها فراغ واحد تم تعبئته اعتماداً على رقم زبون أكثر الزبائن تكراراً وهو Ali بترقيم 1828.  تم جعل هذا الحقل حقل رقمي بدون فواصل عشرية |  |  |  |
| 5 | Customer Name | تم ملئ الفراغات اعتماداً على أكثر الزبائن تكراراً وهو Ali  تم تعديل حقل Customer ID لهذه الفراغات لتصبح رقم الزبون Ali  تم جعل هذا الحقل حقل نصي | التكرار لا معنى له هنا | 12 | 12 |
| 6 | Quantity | تم ملئ الفراغات بقيمة 1  تم تعديل الكتابة التي تعبر عن أرقام (three) إلى أرقام فعلية  تم ابقاء القيم السالبة كما هي على اعتبار أنها ستعتبر مردودات المبيعات وبغض النظر عن المكتوب في حالة التسليم لغاية الان وعدد تكرارات هذه القيم في هذا العمود هو 12 لغاية اللحظة.  تم جعل هذا الحقل حقل رقمي بخانتين بعد الفاصلة العشرية. | التكرار لا معنى له هنا | 6 | 10 |
| 7 | Price per Unit | بمقارنة عمود اجمالي السعر مع هذا العمود يتضح أن ضرب الكمية بهذا السعر في كل سجل لا يعطي اجمالي السعر وبغض النظر عن وجود أو عدم وجود خصم، لذلك، وضعت فرضية أن هذه المعلومات في هذا العمود خاطئة لسبب ما، والفرضية الثانية أن عمود اجمالي السعر هو الصحيح لأنها نقود مقبوضة، وبالتالي يجب إعادة احتساب سعر الوحدة أخذين بعين الاعتبار حقل الخصم بعد التنظيفات التي جرت عليه وباستخدام هذه المعادلة  تم وضع القيم بعد الاحتساب في عمود جديد باسم “CR Price per Unit”  تم استبدال كلمتي "one thousand" بالرقم 1000 في حالتي تكرار.  تم استبدال الفراغات برقم 1 في 3 حالات تكرار.  تم جعل هذا الحقل حقل رقمي بخانتين بعد الفاصلة العشرية. | التكرار لا معنى له هنا |  |  |
| 8 | Discount (%) | تم ملئ الفراغات بقيمة (0)  تم تعديل الكتابة التي تعبر عن أرقام (none) إلى أرقام فعلية (0).  تم جعل هذا الحقل حقل رقمي بدون فواصل عشرية. | التكرار لا معنى له هنا | 5 | 4 |
| 9 | Total Price | يوجد قيم فارغة في هذا العمود، تم تعبئتها بنتيجة (معدل سعر الوحدة مضروباً في الكمية ومخصوم منه قيمة الخصم) حسب نوع المنتج وكما ذكر أعلاه وحسب المعادلة التالية: | التكرار لا معنى له هنا | 11 | 11 |
|  |  | بالنسبة لـ CR Price Per Quantity هو معدل سعر المنتج لهذا السجل، فمثلاً معدل سعر الـ Mouse هو 1956.68 ، أنظر جدول معدلات أسعار المنتجات أعلاه.  تم جعل هذا الحقل حقل رقمي بخانتين بعد الفاصلة العشرية. | | | |
| 10 | CR Price per Unit | هو عمود جديد تم اعادة احتسابه وكما ورد سابقاً.  تم جعل هذا الحقل حقل رقمي بخانتين بعد الفاصلة العشرية. | التكرار لا معنى له هنا | لايوجد | 100 |
| 11 | CR Discount Value | هو عمود جديد تم إضافته لكي يحتسب قيمة الخصم حسب السعر الجديد المذكور في رقم 10 أعلاه، وذلك ليكون جاهزاً من أجل تحليل قيم الخصومات التي أعطيت للمنتجات.  تم جعل هذا الحقل حقل رقمي بخانتين بعد الفاصلة العشرية. | التكرار لا معنى له هنا | لايوجد | 100 |
| 12 | Order Date | بمقارنة تواريخ الطلبات مع رقم الطلب يلاحظ أن ترقيم الطلبات غير صحيح، لذلك تم:   1. ترتيب جميع سجلات هذا الملف اعتماداً على تاريخ الطلب. 2. اعادة ترقيم الطلبات بدءاً من الرقم 101 ولغاية 200.   تم جعل هذا الحقل حقل تاريخ طويل. | التكرار لا معنى له هنا | لا يوجد | 100 |
| 13 | Region | عدد القيم Unknown هي 9، تم استبدالهم بأكثر المناطق تكراراً وهي North-West  حيث كانت تكرارات المناطق كالتالي:   |  |  | | --- | --- | | **Region** | **Count** | | Central | 12 | | North-East | 21 | | North-West | 28 | | South-East | 15 | | South-West | 15 |   تم جعل هذا الحقل حقل نصي | التكرار لا معنى له هنا | 9 | 9 |
| 14 | Payment Method | عدد الفراغات هو 15 تم تعبئتها بأكثر طرق الدفع استخداماً وهي Credit Card  حيث كانت تكرارات طرق الدفع كالتالي:   |  |  | | --- | --- | | **Payment Method** | **Count** | | Cash | 20 | | Cheque | 4 | | Credit Card | 41 | | Online Payment | 20 |   تم جعل هذا الحقل حقل نصي | التكرار لا معنى له هنا | 15 | 15 |
| 15 | Delivery Status | عدد الفراغات هو 6 تم تعبئتها بأكثر حالة للتسليم توفراً وهي Pending  حيث كانت تكرارات حالات التسليم كالتالي:   |  |  | | --- | --- | | **Delivery Status** | **Count** | | Delivered | 36 | | Lost | 8 | | Pending | 37 | | Returned | 13 | |  |  |   بمقارنة حالة التسليم مع الكميات التي تحتوي قيم سالبة، فلا تدل الحالة المذكورة على إرجاع للمبيعات، لذلك، تم تعديل حالة التسليم لها لتصبح Returned، وعدد هذه الحالات هو 12 حالة.  هذه الحالات الـ 12 تم تعديل قيم الكميات لها بإزالة قيم السالب منها، وقد تم تظليل أسطرها باللون الأزرق الفاتح، وسيتم الاعتماد على حالة التسليم لمعرفة مردودات المبيعات وليس القيم السالبة.  تم جعل هذا الحقل حقل نصي | التكرار لا معنى له هنا | 6 | 6  12 |
| 16 | Feedback | تم ملئ الفراغات للتغذية الراجعة بـ Average وعددها 11 حالة  تم جعل هذا الحقل حقل نصي | التكرار لا معنى له هنا | 11 | 11 |
| 17 | Serial | تم إضافة هذا العمود قبل أعادة ترتيب السجلات ليحافظ على الترتيب الأصلي.  تم جعل هذا الحقل حقل رقمي بدون فواصل عشرية. | لا ينطبق | لا يوجد | لا يوجد |
| 18 | Sales\_Project | تم تخزين الملف باسم جديد وهو:  Sales\_Project\_Cleaned.xlsx | لا ينطبق | لا ينطبق | لا ينطبق |

**ملاحظات حول أسعار المنتجات:**

من الرسم البياني التالي:

يلاحظ أن أسعار المنتجات غير حقيقية، أو ربما تم وضعها بشكل عشوائي، أنظر للوحة المفاتيح، أنظر أدنى سعر وأعلى سعر والوسيط، وأنظر أن توزيع البيانات منحني لليمين، بمعنى أن أكثر الأسعار أعلى من الوسيط.

**ملاحظات حول مجموعة البيانات:**

عدد السجلات قليل جداً والكثير من المعلومات مفقودة وبالتالي لن نستطيع التعمق في عمليات تحليل البيانات ولا الوصول لمرحلة تعلم الألة واستخدام النمذجة ولا ربما استخدام مؤشرات الأداء المتعلقة بعمليات البيع وربما يعود سبب ذلك إلى أن واضع التحدي لن يصل لذلك العمق بمجموعة البيانات هذه، أو أنه سيقدم غيرها إذا كان يخطط لتعليم معمق في تحليل البيانات، لكن ربما يرى أن لا يعيدنا لمرحلة التنظيف مرة أخرى، وإن كنت أفضل أن نبدأ بمجموعة بيانات كبيرة وفيها من الأعمدة ما يمكننا من الابحار فيها واستخدامها في كل مراحل البيانات ومن ضمنها تعلم الآلة والتعلم العميق وغيرها.

ملاحظات حول الهدف من تحليل مجموعة البيانات هذه.

بالعادة يتم توضيح هدف عملية التحليل الذي يطلبه صاحب هذه البيانات، فهل يريد مراقبة أداء الأعمال في المناطق؟ أو يريد تصنيف الربحية سواء للفروع أو المنتجات، أم هل يريد معرفة أنسب سعر للمنتجات أم شيء آخر؟  
يفضل تحديد هدف محدد مع التحدي ليكون واضحاً للجميع لماذا أقوم بتحليل مجموعة البيانات هذا وما الهدف الذي أسعى لتحقيقه لكي يتم توجيه تفكير المتدربين نحو التعمق في الحقول وتشكيلاتها وتفصيلاتها التي تساعد في الوصول لتحقيق الهدف.

شكراً.