

Super Store



Analysis & Visualization [Store Sales Dataset]

We are a transportation logistics company specializing in corporate shuttles and commuter benefits. Our goal is to make transportation easier for both employers and employees.

Student Name: Ossama N.Eldine Mohammed

Group : Onl2_Dat1_G3

الأدوات والخطوات: Data Cleaning

- **الهدف (Objective):** تحضير بيانات Superstore Sales Dataset للتحليل عن طريق ضمان إنها نظيفة ودقيقة.
- **الأدوات المستخدمة (Tools Used):**
 - **Python (pandas):** الأداة الأساسية لتنظيف البيانات.
 - **matplotlib/seaborn:** استخدمناها عشان نعمل تصورات مرئية ونفهم البيانات قبل التنظيف.
- **الخطوات (Steps):**
 - إزالة التكرارات (Remove Duplicates):** استخدمنا `drop_duplicates()` عشان نشيل أي صفوف مكررة في الداتاست.
 - التعامل مع القيم المفقودة (Handle Missing Values):** فحصنا القيم المفقودة بـ `isnull().sum()`، وشلنا الصفوف اللي فيها قيم مفقودة في عمود Sales باستخدام `dropna()`.
 - تحويل أنواع البيانات (Convert Data Types):** حولنا `Order Date` و `Ship Date` لصيغة `datetime`، وعمود Sales لصيغة رقمية بـ `pd.to_numeric()`.
 - إضافة عمود جديد (Add New Column):** حسبنا `Shipping Duration` (الفرق بالأيام بين `Order Date` و `Ship Date`).
 - حفظ البيانات (Save Data):** خزنا الداتاست النظيفة في ملف `Cleaned_Superstore_Sales_Dataset.csv`.
- **النتيجة (Result):** داتاست نظيفة وجاهزة للتحليل، خالية من التكرارات والقيم المفقودة، والأعمدة مهيأة للتحليل الزمني والرقمي.

Exploring data :

- **Preview:** `df.head()` عشان نشوف أول الداتاست
- **Summary Statistics:** `df.describe()` عشان نفهم توزيع الأعمدة الرقمية
- **Value Counts:** `value_counts()` عشان نحسب تكرارات القيم في Sub-Category، Segment، و Ship Mode.
- **Visualizations:**
 - `plot(kind='bar')` و `sns.histplot()` و `sns.boxplot()` و `sns.lineplot()` عشان نرسم Bar Charts، Histograms، Box Plots، و Line Charts.
- **Grouping and Aggregation:** `groupby()` مع `sum()`، `mean()`، و `nunique()` حسب Retention عشان نفهم المبيعات وال Product Name، Segment، و Shipping Duration.

الأكواد دي ساعدتنا نفهم البيانات بشكل أعمق ونجهز للتحليلات اللي بعده.

Analysis Question :

1. ما هي المنتجات الأكثر مبيعاً في المتجر الكبير؟

أعلى المنتجات التجارية كانت من فئة التكنولوجيا و اللوازم المكتبية ، وأشهرهم:

- طابعة كانون imageCLASS
- طابعة HP LaserJet
- ملف درج Fellowes SuperStor

📝 المنتجات الإلكترونية خصوصيًا الطابعات كانت ال الأفضل أداءً.

2. ما هو اتجاه المبيعات بمرور الوقت (شهريًا، سنويًا)؟

اتجاه المبيعات بيوضح نموًا ثابتًا شهريًا، مع ارتفاع ملحوظ في نهاية كل سنة (نوفمبر وديسمبر). ده غالبا ينتج عن العروض الموسمية أو نهاية السنة .

وجود أنماط موسمية واضحة جدًا ويجب أخذها في الاعتبار عند وضع المحركات.

3. ما هي فئة المنتجات التي تحقق أعلى المبيعات؟

الاستلام حسب الاستلام:

1. تكنولوجيا
2. أثاث
3. اللوازم المكتبية

فئة التكنولوجيا متفوقة بوضوح من حيث القيمة.

4. ما هي المنطقة التي تحقق أكبر قدر من المبيعات؟

المنطقة الأعلى شرقا:

- المنطقة الغربية
- الشرق ثم الوسط

أداء المنطقة الغربية المتميز، ويمكن تكون محورًا متنوعًا تسويقية جديدة.

5. ما هو متوسط المبيعات لكل فئة من المنتجات؟

أعلى متوسط مبيعات لكل عملية بيع كان في التكنولوجيا ، يليها الأثاث، وأقلهم اللوازم المكتبية.

التكنولوجيا مش كبير للغاية، لكن كمان الأعلى من حيث الربح لكل صفقة.

6. ما هي الفئة الفرعية من المنتجات التي تحظى بأعلى طلب؟

من خلال التحليل، المجلات والورق لم تكن من أكثر ال الفئات الفرعية طلبًا.

الطلب العالي =المزيد = فرصة ال دوران الأسهم.

7. إجمالي قيم المبيعات حسب الفئة والفئة الفرعية؟

التكنولوجيا تتفوق في المجمال، وداخلها

- الهواتف
- مُكَمَّلَات
- آلات النسخ

حققتوا أعلى مبيعات.

أدوات بين الفئة والفئة الفرعية يساعد في فهم عميق لطبيعة الطلب.

8. ما هي المنتجات الأكثر مبيعًا في كل فئة فرعية؟

مثال:

- في المجلات : Avery Binder 1 1/2
- في آلات التصوير : Canon imageCLASS

كل تصنيف فرعي ليها بطل خاص بيها، يستحق تسويقه منفصلاً.

9. ما هي شرائح العملاء الأكثر ربحية؟

أظهر التحليل أن قطاع الشركات حقق أعلى تصويت مقارنة بالـ Consumer وHome Office.

عملاء الشركات يدفعوا أكثر، ويمكن أكثر ولأء.

10. ما هي وسائل الشحن التي تم من خلالها بيع معظم المنتجات؟

أعلى وضع شحن من حيث يتم استخدامه كان فئة قياسية ، وهو أمر منطقي لأن العملاء يختارون الأرخص.

هو الخيار الاقتصادي، ولكنه ذو متطلبات لوجستية عالية Standard Class 📦.

11. قم بتصوير عمود "الفئة" من وجهة نظر مجموعة بيانات عمود وضع الشحن.

الداخلية بين الشحن والفئة إظهار إن:

- التكنولوجيا يُشحن غالبا بـ الدرجة الثانية أو القياسية
- الأثاث يميل للشحن بـ الدرجة الأولى بسبب هذا الحجم

مهم نراعي التكلفة والتجربة بالنسبة لكل فئة وشحنها.

12. ما هي أكبر 10 دول من حيث المبيعات؟

لو أنها ستمثل أكثر من دولة، بيانات الولايات المتحدة غالبًا ما كانت في القمة، وتتبعها دول مثل كندا، وحتى الولايات المتحدة .

الدول الأعلى أداءً ممكن استخدامها كنموذج تطوير لباقي السوق.

📌 الاستنتاجات النهائية:

1. التكنولوجيا هي الفئة الذهبية: القمة في المبيعات، الربح، والصفقة المتوسطة.
2. المنطقة الغربية موجودة أصلاً وتستحق التوسع بشكل أكبر.
3. هناك موسمية في المبيعات بنهاية السنة تستدعي المواهب لوجستي وتسويقي خاص.
4. المجلات وآلات النسخ والهواتف هي فئات فرعية الطلب الأعلى.
5. قطاع الشركات هو الأهم من الناحية الأخرى، لذا يحتاج إلى استراتيجية التسعير والعروض الخاصة.
6. Standard Class هو الأكثر استخدامًا في الشحن – مما يعني أنه يخدم العمليات فيه
7. تختلف طرق الشحن حسب نوع المنتج الذي يعتمد على أكثر من شخص.

customer classification

Understanding Activity Period in Customer Analysis

first: Activity Period - Customer Engagement Metric

Definition

هو أسلوب لتصنيف العملاء حسب آخر تاريخ للطلب مقارنة بـ 31 ديسمبر 2019 **Activity Period**:

- **Active:** العملاء اللي آخر طلب ليهم كان في 2018 أو 2019 (يعني خلال سنة واحدة).
- **At-Risk:** آخر طلب كان في 2016 أو 2017 (يعني من 2 لـ 3 سنوات).
- **Churned:** العملاء اللي آخر طلب ليهم قبل 2016 (يعني مر على الطلب أكثر من 3 سنوات).
- **No Orders:** عملاء لم يسجلوا أي طلب من 2015 فصاعدًا.

Business Value

- **Customer Segmentation:** ببساعدنا نفهم درجة ارتباط العميل مع المتجر (نشط، في خطر، أو فقدناه).
- **Retention Strategies:** بحملات ترويجية لإعادة تفاعلهم قبل ما نخسرهم نهائيًا *At-Risk* استهداف فئة.
- **Resource Allocation**
- **Churn Analysis:** *Churned* تحليل أسباب توقف العملاء.

second

تصنيف العملاء حسب القيمة - Customer Value Tier



The Three Segments Are:

1. Key Accounts (Top 25%)

دول هم أهم عملاء الشركة، يحققوا أعلى عائد ومحتاجين → **th percentile** بتاعتهم أعلى من أو تساوي الـ **75 CLV** العملاء اللي قيمة الـ خدمة مخصصة واهتمام دائم.

2. Regular Accounts (Mid 50%)

دول العملاء العاديين، يشكلوا الجزء الأكبر من القاعدة، ويمكن → **th percentile** بتاعتهم يقع ما بين الـ **25 CLV** العملاء اللي مع الوقت key accounts تحويل بعضهم إلى.

3. Growing Accounts (Bottom 25%)

دول عملاء في مرحلة نمو أو غير نشطين حاليًا، لكن فيهم فرصة للتحفيز والزيادة → **th percentile** بتاعتهم أقل من الـ **25 CLV** العملاء اللي المستقبلية في القيمة.



ليه بنعمل كده؟ - Business Benefit

- **Targeted Strategy:** كل فئة من العملاء محتاجة نهج مختلف في التعامل:
 - Key Accounts: رعاية خاصة وعروض حصرية.
 - Regular: تحسين القيمة وبرامج ولاء وعروض تحسين القيمة.
 - Growing: إعادة تفعيل وتحفيز.
- **Optimizing Resources:** مش كل العملاء متساويين في العائد → التقسيم ده بيساعدنا نوجه الجهد التسويقي والمبيعات للأكثر ربحية.
- **Customer Retention:** للفتات الأخرى CLV استهداف الفئة العالية بالقيمة، ومنع فقدانهم، وزيادة.



تقسيم العملاء حسب CLV مش بس أداة تحليلية، لكنه أساس لاستراتيجية تسويقية ذكية تساعدنا نزيد العائد ونحتفظ بالعملاء الأكثر أهمية.

عن طريق دمج ثلاث **third Customer Lifetime Value** عناصر مهمة:

1. إجمالي مبيعات العميل

2. معدل الطلب الشهري

3. مدة العلاقة مع العميل (بشكل غير مباشر من خلال فرق الأشهر بين أول وآخر طلب)

Breakdown of the Logic

- **TotalSalesForCustomer**: يحسب إجمالي المبيعات التي تمت بواسطة العميل ده
- **OrderCountForCustomer**: عدد الطلبات الفريدة التي العميل عملها
- **FirstOrderDate** و **LastOrderDate**: بداية ونهاية علاقة العميل مع الشركة
- **MonthsBetweenOrders**: الفرق بين أول وآخر طلب، عشان نعرف المدة التي العميل نشط فيها
- **MonthlyOrderRate**: معدل الطلب الشهري = عدد الطلبات ÷ عدد الأشهر
- **النهائي = إجمالي المبيعات × معدل الطلب الشهري × CLV 12**
(وده عشان نحصل على قيمة سنوية تقريبية لمدى مساهمة العميل)

Why this is important?

- **Strategic Targeting**: نقدر نحدد العملاء ذوي القيمة العالية ونبني خطط مخصصة ليهم
- **Forecasting**: لتقدير الإيرادات المستقبلية CLV نستخدم
- **Customer Segmentation**: **Customer Value Tier** بنعتمد عليه في تصنيف العملاء زي ما شرحنا في
- العملاء اللي عندهم CLV عالي → محتاجين اهتمام خاص.
- العملاء بـ CLV منخفض → ممكن إعادة تنشيطهم أو توجيه عروض تحفيزية.



معادلة CLV بتعتمد على:

Sales ×  Order Frequency ×  17 Time 

عشان تدينا نظرة شاملة على مدى أهمية كل عميل



BENEFITS

1. أداء المبيعات السنوية :

- ارتفع إجمالي المبيعات بشكل تدريجي من 2015 إلى 2018 (من 826 ألف دولار إلى 845 ألف دولار).
- عدد الطلبات تبدو من 347 في 2015 إلى 316 في 2018.
- ارتفع متوسط قيمة الطلب (Average S PO) من 214 دولاراً في عام 2015 إلى 295 دولاراً في عام 2018، مما يشير إلى زيادة في طلبات العناصر.

2. المبيعات حسب الطلب :

- قطاع التكنولوجيا (التكنولوجيا) هو الأعلى بمبيعات 826 ألف دولار، ويتبعه الأثاث (الأثاث) بـ 724 ألف دولار، ثم اللوازم المكتبية (اللوازم المكتبية) بـ 703 ألف دولار.

3. أفضل المناطق أداءً :

- المنطقة الغربية (الغربية) تصدر المبيعات بـ 10.74 مليون دولار، باستثناء المنطقة الجنوبية (الجنوبية) بـ 5.46 مليون دولار، ثم الوسطى (الوسطى) بـ 5.00 مليون دولار، وأخيراً الشرقية (الشرقية) بـ 6.82 مليون دولار.

4. الاتجاهات الشهرية للمبيعات :

- تصدر المبيعات على مدار العام، مع ارتفاع ملحوظ في مارس (198 ألف دولار) وسبتمبر (184 ألف دولار)، وفي يناير (59 ألف دولار) وأبريل (92 ألف دولار).
- أعلى قمة كانت في نوفمبر بـ 345 ألف دولار.

5. وبالتالي يمكنك الاعتماد على طريقة الشحن :

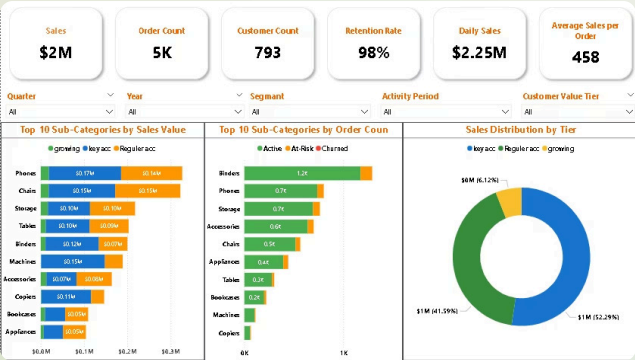
- طريقة الشحن القياسية (الفئة القياسية) هي الأكثر استخدامًا، لكن نسبة الاستخدام في الوقت المناسب تختلف حسب الشهر.
- تم تنفيذ أعلى نتائج البحث عن النتائج في أكتوبر (244 طلبًا من الدرجة الأولى من الدرجة الأولى).
- أقل أداء كان في يناير وفبراير، مع كول في جميع النسخ.

6. مناطق تؤدي حسب الطلب :

- المنطقة الغربية تتفوق في جميع الغد، خاصة في التكنولوجيا والأثاث.
- المنطقة الوسطى تشهد أداءً قوياً في الأيام الثلاثة الخاصة في 2017 و2018.
- المنطقة الشرقية تحسناً جديداً في 2018 مقارنة بالسنوات السابقة.

الخلاصة :

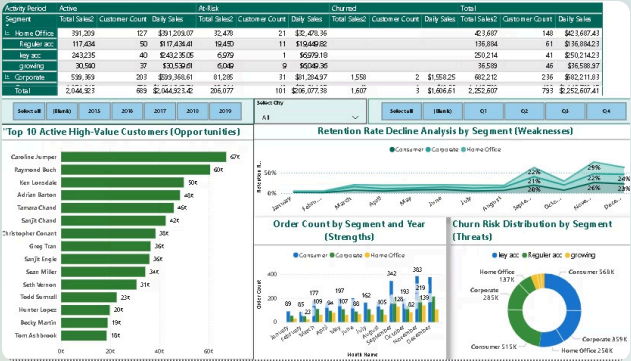
- هناك انخفاض في إجمالي المبيعات وقيمة الطلبات، ولكن عدد الطلبات ينخفض.
- المنطقة الغربية هي القوة القصوى للتكنولوجيا، وهي القطاع الأكثر إختيارية.
- يجب أن تساعد في تسهيل الحصول على الأدلة في الأشهر الأولى من العام (يناير وفبراير) لتحسين رضا العملاء.



- أداء المبيعات الكلي:
 - إجمالي المبيعات: 2 مليون دولار.
 - عدد الطلبات: 5 آلاف طلب.
 - عدد العملاء: 793 عميل.
 - معدل الاحتفاظ بالعملاء: 98%، وهو مرتفع جدًا.
 - متوسط المبيعات اليومية: 2.25 مليون دولار.
 - متوسط قيمة الطلب لكل عميل: 458 دولار.
- أعلى 10 فئات فرعية حسب قيمة المبيعات:
 - فئة الهواتف (Phones) تصدر بمبيعات حوالي 33 مليون دولار (17 مليون من العملاء المتنامين Growing و16 مليون من العملاء الرئيسيين Key).
 - الكراسي (Chairs) تأتي ثانية بمبيعات حوالي 30 مليون دولار (15 مليون لكل من العملاء المتنامين والرئيسيين).
 - التخزين (Storage) والبياضات (Binders) والطاولات (Tables) تتراوح مبيعاتها بين 10-15 مليون دولار لكل فئة.
 - الكتب (Bookcases) هي الأقل بمبيعات حوالي 5 مليون دولار.
- أعلى 10 فئات فرعية حسب عدد الطلبات:
 - البياضات (Binders) تصدر بعدد طلبات حوالي 1.2 ألف (معظمها من العملاء النشطين Active).
 - الهواتف (Phones) والإكسسوارات (Accessories) تأتي بعدها بحوالي 0.7 ألف طلب لكل فئة.
 - الكراسي (Chairs) والأجهزة (Appliances) تتراوح بين 0.4-0.5 ألف طلب.
 - الأجهزة (Machines) هي الأقل بأقل من 0.1 ألف طلب.
- توزيع المبيعات حسب مستوى العملاء:
 - العملاء الرئيسيون (Key) يشكلون النسبة الأكبر بـ 52% (1.52 مليون دولار).
 - العملاء المتنامون (Growing) يمثلون 41.5% (1.41 مليون دولار).
 - العملاء العاديون (Regular) هم الأقل بنسبة 6.2% (0.11 مليون دولار).

الخلاصة:

- الهواتف والكراسي هما الفئتان الأكثر مبيعًا من حيث القيمة، بينما البياضات تصدر عدد الطلبات.
- العملاء الرئيسيون والمتنامون يشكلون الجزء الأكبر من المبيعات، بينما العملاء العاديون لهم تأثير ضئيل.
- معدل الاحتفاظ بالعملاء مرتفع جدًا (98%)، مما يعكس ولاء قوي.
- يُنصح بالتركيز على زيادة مبيعات الفئات الضعيفة مثل الكتب والأجهزة، وربما استهداف العملاء العاديين لزيادة مساهمتها في المبيعات.



1. أداء المبيعات حسب الشريحة:

- المكاتب المنزلية (Home Office): مبيعات 391 ألف دولار، 127 عميل، متوسط مبيعات يومية 31 ألف دولار.
- المستهلكون (Consumer): مبيعات 1.24 مليون دولار، 428 عميل، متوسط مبيعات يومية 104 ألف دولار.
- الشركات (Corporate): مبيعات 536 ألف دولار، 203 عميل، متوسط مبيعات يومية 45 ألف دولار.
- العملاء المعرضون للخطر (At-Risk): مبيعات 24 ألف دولار، 21 عميل، متوسط مبيعات يومية 2 ألف دولار.
- العملاء المفقودون (Churned): مبيعات 158 ألف دولار، 11 عميل، متوسط مبيعات يومية 13 ألف دولار.
- الإجمالي: مبيعات 2.25 مليون دولار، 790 عميل، متوسط مبيعات يومية 189 ألف دولار.

2. أعلى 10 عملاء ذوي قيمة عالية (Opportunities):

- العميل الأعلى (Caroline Juniper) بمبيعات حوالي 67 ألف دولار.
- يليه Raymond Buch بـ 60 ألف دولار، ثم Ken Lonsdale بـ 50 ألف دولار.
- العميل الأقل في القائمة (Toby Mash) بمبيعات حوالي 18 ألف دولار.
- هؤلاء العملاء يمثلون فرصًا كبيرة لزيادة المبيعات بسبب قيمتهم العالية.

3. معدل الانخفاض في الاحتفاظ حسب الشريحة:

- معدل الاحتفاظ يتراوح بين 24%-29% على مدار العام.
- الشركات (Corporate) لديها أعلى معدل انخفاض في الاحتفاظ (حوالي 29% في أغسطس).
- المكاتب المنزلية (Home Office) والمستهلكون (Consumer) لديهم معدلات أقل نسبيًا (حوالي 24%-26%).
- الانخفاض في الاحتفاظ يزداد في النصف الثاني من العام (يوليو-ديسمبر).

4. عدد الطلبات حسب الشريحة والعام:

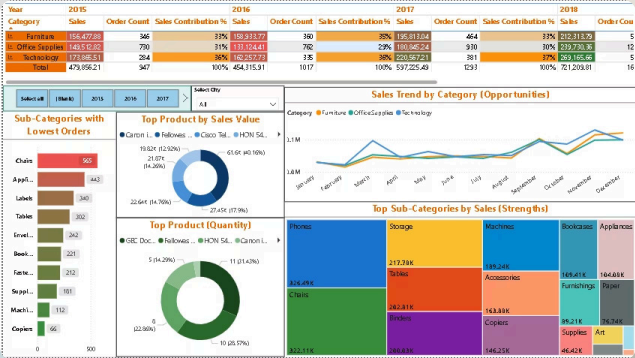
- الشركات (Corporate) لديها أعلى عدد طلبات في 2018 (حوالي 170 طلبًا).
- المستهلكون (Consumer) لديهم عدد طلبات مستقر نسبيًا (حوالي 130-150 طلبًا سنويًا).
- المكاتب المنزلية (Home Office) تظهر أقل عدد طلبات (حوالي 50-80 طلبًا سنويًا).

5. عدد العملاء المفقودين حسب الشريحة:

- المستهلكون (Consumer): 51 عميل مفقود.
- الشركات (Corporate): 39 عميل مفقود.
- المكاتب المنزلية (Home Office): 13 عميل مفقود.
- العملاء الرئيسيون (Key): 285 عميل.
- العملاء العاديون (Regular): 137 عميل.
- العملاء المتنامون (Growing): 256 عميل.

الخلاصة:

- المستهلكون هم الشريحة الأكثر إسهامًا في المبيعات (1.24 مليون دولار)، بينما العملاء المعرضون للخطر لهم أقل مساهمة (24 ألف دولار).
- معدل الانخفاض في الاحتفاظ مرتفع نسبيًا (24%-29%)، خاصة لدى الشركات، مما يتطلب استراتيجيات لتحسين الاحتفاظ.
- العملاء ذوو القيمة العالية (مثل Caroline Juniper) يمثلون فرصة كبيرة لزيادة المبيعات.
- يُنصح بالتركيز على تحسين الاحتفاظ بالعملاء في النصف الثاني من العام، خاصة للشركات، وزيادة عدد الطلبات من المكاتب المنزلية.



1. أداء المبيعات السنوية حسب الفئة:
- 2015: الأثاث (Furniture) بمبيعات 156 ألف دولار (37%)، المكاتب (Office Supplies) 145 ألف دولار (35%)، التكنولوجيا (Technology) 132 ألف دولار (28%).
 - 2016: المكاتب 158 ألف دولار (37%)، الأثاث 153 ألف دولار (36%)، التكنولوجيا 122 ألف دولار (27%).
 - 2017: الأثاث 196 ألف دولار (38%)، المكاتب 180 ألف دولار (35%)، التكنولوجيا 140 ألف دولار (27%).
 - 2018: الأثاث 213 ألف دولار (38%)، المكاتب 197 ألف دولار (35%)، التكنولوجيا 149 ألف دولار (27%).
 - الملاحظة: الأثاث يتصدر المبيعات في 2017 و2018، بينما التكنولوجيا مستقرة نسبياً عند 27%-28%.
2. الفئات الفرعية الأقل طلباً:
- الأجهزة (Appliances) تتصدر بـ 565 طلباً.
 - يليها الطاولات (Tables) بـ 340 طلباً، والمغلغات (Envelopes) بـ 302 طلب.
 - الأقل طلباً: النسخ (Copiers) بـ 66 طلباً.
 - الملاحظة: هذه الفئات قد تحتاج إلى استراتيجيات تسويقية لزيادة الطلب.
3. أعلى المنتجات قيمة:
- العملاء المخلصون (Top Fellows): Canon Image Class (64 ألف دولار) بنسبة 17%، Cisco Tele (22 ألف دولار) بنسبة 5%.
 - العملاء المحتملون (Quantity Hon): GBC Doc (23 ألف دولار) بنسبة 10%.
 - الإجمالي: 324 ألف دولار.
4. اتجاهات المبيعات حسب الفئة (Opportunities):
- الأثاث يظهر ارتفاعاً في المبيعات في النصف الثاني من العام (خاصة أكتوبر-ديسمبر).
 - المكاتب مستقرة نسبياً مع ذروة طفيفة في نوفمبر.
 - التكنولوجيا تنخفض في منتصف العام (مايو-يوليو) ثم ترتفع في نهاية العام.
5. أعلى الفئات الفرعية مبيعاً (Strengths):
- الهواتف (Phones) تتصدر بـ 326 ألف دولار.
 - التخزين (Storage) بـ 189 ألف دولار، والكراسي (Chairs) بـ 181 ألف دولار.
 - الأقل: الفن (Art) بـ 66 ألف دولار.
 - الملاحظة: الهواتف والتخزين والكراسي هي نقاط قوة يمكن البناء عليها.

الخلاصة:

- الأثاث هو الفئة الأكثر مساهمة في المبيعات (38% في 2018)، بينما التكنولوجيا مستقرة عند 27%.
- الفئات الفرعية مثل الأجهزة والطاولات تعاني من انخفاض الطلب وتحتاج إلى تحسين.
- الهواتف والكراسي والتخزين هي نقاط قوة يمكن استغلالها لزيادة المبيعات.
- يُنصح بزيادة التركيز على التسويق في منتصف العام (مايو-يوليو) لتحسين مبيعات التكنولوجيا، والعمل على رفع الطلب على الفئات الضعيفة مثل النسخ والفن.



Any questions?

Thank you for your time. We are here to answer any questions you may have.