

تقرير تحليلي - Superstore

نتائج تحليل ملف البيانات SampleSuperstore.csv

مقرر: نكاء اصطناعي

المهندس: زكريا عبدالعزيز محمد الحنيني

جدول المحتويات

2	مقدمة التحليل: -	
3	قراءة وتنظيف البيانات: -	(1)
7	التحليل الأساسي: -	(2)
9	التحليل الإضافي: -	(3)
11	الرسوم البيانية: -	(4)

مقدمة التحليل: -

يهدف هذا التكميف إلى تحليل بيانات متجر Superstore باستخدام مكتبات Pandas و NumPy و Matplotlib ، وذلك لاستخراج مؤشرات تساعد في فهم المبيعات والأرباح واتخاذ قرارات أفضل.

1) قراءة وتنظيف البيانات: -

يشمل هذا الجزء حساب:

- ❖ اقرأ ملف البيانات باستخدام `pandas.csv_read`.
- ❖ تحقق من وجود قيم ناقصة، وقم بمعالجتها (تعويض أو حذف).
- ❖ تحقق من صحة البيانات مثل (التواريخ والقيم الرقمية).

1. اقرأ ملف البيانات باستخدام `pandas.csv_read`؟

الحل

الشفرة (Code) 1

```
# قراءة البيانات
SuperstoreData = pd.read_csv("SampleSuperstore.csv", encoding='cp1252')
SuperstoreData.head()
```

نتيجة تنفيذ Code 1

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	
0	1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	42420	South	FUR-SO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600	2	0.00	41.9136
1	2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs	731.9400	3	0.00	219.5820
2	3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	90036	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b...	14.6200	2	0.00	6.8714
3	4	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	33311	South	FUR-TA-10000577	Furniture	Tables	Bretford C44500 Series Slim Rectangular Table	957.5775	5	0.45	-383.0310
4	5	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	33311	South	OFF-ST-10000780	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.3600	2	0.20	2.5164

ملاحظات	استنتاجات	التوصيات
نلاحظ ان هناك خطأ CSV. في حال وجود مشكلة في الترميز، نستخدم 'cp1252'	يمكن قراءة ملف البيانات واستخدامه في التحليل	استخدام 'encoding='cp1252' لحل مشكلة الترميز

الشفرة (Code) 2

```
# معلومات عن البيانات
SuperstoreData.info()
```

نتيجة تنفيذ Code 2

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Row ID              9994 non-null   int64
1   Order ID            9994 non-null   object
2   Order Date          9994 non-null   object
3   Ship Date           9994 non-null   object
4   Ship Mode           9994 non-null   object
5   Customer ID         9994 non-null   object
6   Customer Name       9994 non-null   object
7   Segment            9994 non-null   object
8   Country             9994 non-null   object
9   City               9994 non-null   object
10  State              9994 non-null   object
11  Postal Code        9994 non-null   int64
12  Region            9994 non-null   object
13  Product ID         9994 non-null   object
14  Category           9994 non-null   object
15  Sub-Category       9994 non-null   object
16  Product Name       9994 non-null   object
17  Sales              9994 non-null   float64
18  Quantity           9994 non-null   int64
19  Discount            9994 non-null   float64
20  Profit             9994 non-null   float64
dtypes: float64(3), int64(3), object(15)
memory usage: 1.6+ MB

```

ملاحظات	استنتاجات	التوصيات
الملف يحتوي على 9994 صف من البيانات	نستنتج ان البيانات كبيرة	الاهتمام بتحليل البيانات قل تدريب نموذج الذكاء الاصطناعي عليها

2. تحقق من وجود قيم ناقصة، وقم بمعالجتها (تعويض أو حذف)؟

الحل

الشفرة (Code) 1

```

# القيم الفارغة أو المفقودة
SuperstoreData.isnull().sum()

```

نتيجة تنفيذ Code 1

```

Row ID      0
Order ID    0
Order Date  0
Ship Date   0
Ship Mode   0
Customer ID  0
Customer Name 0
Segment     0
Country     0
City        0
State       0
Postal Code  0
Region      0
Product ID  0
Category    0
Sub-Category 0
Product Name 0
Sales       0
Quantity    0
Discount    0
Profit      0
dtype: int64

```

ملاحظات	استنتاجات	التوصيات
البيانات لا تحتوي على قيم فارغة	البيانات نظيفة	التأكد من ان البيانات نظيفة وجاهزة للمرحلة التالية

الشفرة (Code) 2

```

# حذف الصفوف التي تحتوي على قيم فارغة
SuperstoreData.dropna(inplace=True)
# معلومات عن البيانات
SuperstoreData.info()

```

نتيجة تنفيذ Code 2

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Row ID              9994 non-null   int64
1   Order ID            9994 non-null   object
2   Order Date          9994 non-null   object
3   Ship Date           9994 non-null   object
4   Ship Mode           9994 non-null   object
5   Customer ID         9994 non-null   object
6   Customer Name       9994 non-null   object
7   Segment             9994 non-null   object
8   Country             9994 non-null   object
9   City                9994 non-null   object
10  State               9994 non-null   object
11  Postal Code         9994 non-null   int64
12  Region              9994 non-null   object
13  Product ID          9994 non-null   object
14  Category            9994 non-null   object
15  Sub-Category        9994 non-null   object
16  Product Name        9994 non-null   object
17  Sales               9994 non-null   float64
18  Quantity            9994 non-null   int64
19  Discount            9994 non-null   float64
20  Profit              9994 non-null   float64
dtypes: float64(3), int64(3), object(15)
memory usage: 1.6+ MB

```

ملاحظات	استنتاجات	التوصيات
فيظنننا يوتحتى لى 9994 اقصد تانايب دعب	البيانات لا تحتوى على قيم فارغة	فحص أنواع البيانات

3. تحقق من صحة البيانات مثل (التواريخ والقيم الرقمية)؟

الحل

شفرة (Code)

```

# تحويل التواريخ
SuperstoreData['Order Date'] = pd.to_datetime(SuperstoreData['Order Date'])
SuperstoreData['Ship Date'] = pd.to_datetime(SuperstoreData['Ship Date'])

# تحويل الارقام
SuperstoreData['Sales'] = pd.to_numeric(SuperstoreData['Sales'])
SuperstoreData['Profit'] = pd.to_numeric(SuperstoreData['Profit'])

# وصف البيانات
SuperstoreData.describe()

```

نتيجة تنفيذ Code

	Row ID	Order Date	Ship Date	Postal Code	Sales	Quantity	Discount	Profit
count	9994.000000	9994	9994	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	4997.500000	2016-04-30 00:07:12.259355648	2016-05-03 23:06:58.571142912	55190.379428	229.858001	3.789574	0.156203	28.656896
min	1.000000	2014-01-03 00:00:00	2014-01-07 00:00:00	1040.000000	0.444000	1.000000	0.000000	-6599.978000
25%	2499.250000	2015-05-23 00:00:00	2015-05-27 00:00:00	23223.000000	17.280000	2.000000	0.000000	1.728750
50%	4997.500000	2016-06-26 00:00:00	2016-06-29 00:00:00	56430.500000	54.490000	3.000000	0.200000	8.666500
75%	7495.750000	2017-05-14 00:00:00	2017-05-18 00:00:00	90008.000000	209.940000	5.000000	0.200000	29.364000
max	9994.000000	2017-12-30 00:00:00	2018-01-05 00:00:00	99301.000000	22638.480000	14.000000	0.800000	8399.976000
std	2885.163629	NaN	NaN	32063.693350	623.245101	2.225110	0.206452	234.260108

التوصيات	استنتاجات	ملاحظات
إدخال التاريخ بشكل كامل لكي لا يعتبر التاريخ رقم	جميع القيم سليمة	لا يوجد أي قيم خاطئة, الخطأ الذي يظهر في عمودي (Order Date, Ship Date) لان القيم عاب عن تاريخ وليست ارقام

(2) التحليل الأساسي: -

يشمل هذا الجزء حساب:

- ❖ حساب إجمالي المبيعات (Sales) لكل فئة م نتج (Category).
- ❖ حساب متوسط الربح (Profit) لكل منطقة (Region).
- ❖ تصنيف العملاء حسب حجم مشترياتهم بناءً على مجموع المبيعات يمكنك تقسيمهم إلى مجموعات حسب النسبة المئوية أو حدود تحددها.

1. حساب إجمالي المبيعات (Sales) لكل فئة منتج (Category)?

الحل

شفرة (Code)

```
SuperstoreData.groupby('Category')['Sales'].sum()
```

نتيجة تنفيذ Code

```
Category
Furniture      741999.7953
Office Supplies 719047.0320
Technology      836154.0330
Name: Sales, dtype: float64
```

ملاحظات	استنتاجات	التوصيات
نلاحظ ان اجمالي المبيعات لثلاثة الفئات متقارب	لكن نستنتج ان فئة Tech هي التي تملك الإجمالي الأكبر	الاهتمام لفئات التي تمتلك إجمالي مبيعات قليل

2. حساب متوسط الربح (Profit) لكل منطقة (Region)؟

الحل

شفرة (Code)

```
SuperstoreData.groupby('Region')['Profit'].mean()
```

نتيجة تنفيذ Code

```
Region
Central    17.092709
East       32.135808
South      28.857673
West       33.849032
Name: Profit, dtype: float64
```

ملاحظات	استنتاجات	التوصيات
نلاحظ ان على متوسط ربح هو في الفرع West	نستنتج ان هناك ثلاثة مناطق نسبة المتوسط الربح جيدة	دراسة جدوى لمنطقة Central

3. تصنيف العملاء حسب حجم مشترياتهم بناءً على مجموع المبيعات (يمكنك تقسيمهم إلى مجموعات حسب النسبة المئوية أو حدود تحددها)؟

الحل

شفرة (Code)

```
# حساب إجمالي المبيعات لكل عميل
customer_sales = SuperstoreData.groupby('Customer Name')['Sales'].sum()

def classify_customer(sales):
    if sales <= 1000:
        return 'صغير'
```



```

elif sales <= 5000:
    return 'متوسط'
else:
    return 'كبير'

customer_category_alt2 = customer_sales.apply(classify_customer)
customer_category_alt2.value_counts()

```

نتيجة تنفيذ Code

```

Sales
509    متوسط
167    صغير
117    كبير
Name: count, dtype: int64

```

ملاحظات	استنتاجات	التوصيات
نلاحظ ان نسبة المشتريات من الفئة المتوسطة هي الاكثر	نستنتج ان اغلب العملاء من الفئة المتوسطة	أستهدف الفئتين الأقل شران ودراسة سلوكهم

(3) التحليل الإضافي:-

يشمل هذا الجزء حساب:

- ❖ تحليل العلاقة بين كمية المبيعات (Quantity) والربح (Profit) مثل المتوسط، الانحراف المعياري باستخدام إحصائيات وصفية.
- ❖ تحليل أداء الفروع (Segment) حسب المبيعات وال أرباح.

1) تحليل العلاقة بين كمية المبيعات (Quantity) والربح (Profit) مثل المتوسط، الانحراف المعياري باستخدام إحصائيات وصفية؟

الحل

شفرة (Code)

```

SuperstoreData[['Quantity','Profit']].describe()

```

نتيجة تنفيذ Code

	Quantity	Profit
count	9994.000000	9994.000000
mean	3.789574	28.656896
std	2.225110	234.260108
min	1.000000	-6599.978000
25%	2.000000	1.728750
50%	3.000000	8.666500
75%	5.000000	29.364000
max	14.000000	8399.976000

ملاحظات	استنتاجات	التوصيات
نلاحظ انهو في اقل قيمة مبيعات هناك خسائر او نهو عند اعلا قيمة مبيعات يوجد ارباح هائلة	نستنتج انهو كلما قلت المبيعات تحولت الأرباح إلى خسائر والعكس صحيح	اتخاذ اجراء سريع بخصوص أقل قيمة مبيعات لتقليل الخسائر الناتجة عنها

(2) تحليل أداء الفروع (Segment) حسب المبيعات وال أرباح؟

الحل

شفرة (Code)

```
SuperstoreData.groupby('Segment')[['Sales', 'Profit']].mean()
```

نتيجة تنفيذ Code

	Sales	Profit
Segment		
Consumer	223.733644	25.836873
Corporate	233.823300	30.456667
Home Office	240.972041	33.818664

ملاحظات	استنتاجات	التوصيات
نلاحظ انهو كولما زادت المبيعات زادت الارباح	مبيعات الفروع متقاربه لذلك الأرباح متقاربه ايضا	دراسة وضع الفروع الاقل مبيعا

3) الرسوم البيانية:- المخططات المطلوبة:

- ❖ مخطط شريطي (Chart Bar) لإظهار إجمالي المبيعات لكل فئة منتج.
- ❖ مخطط دائري (Chart Pie) لنسبة المبيعات لكل منطقة.
- ❖ مخطط نقاط (Plot Scatter) للعلاقة بين كمية المبيعات والربح.
- ❖ مخطط خطي (Chart Line) لعرض تطور المبيعات عبر الوقت إن وُجد عمود تاريخ.
- ❖ أي رسم بياني آخر يمكن ك إنشاؤه يدعم تحليل.

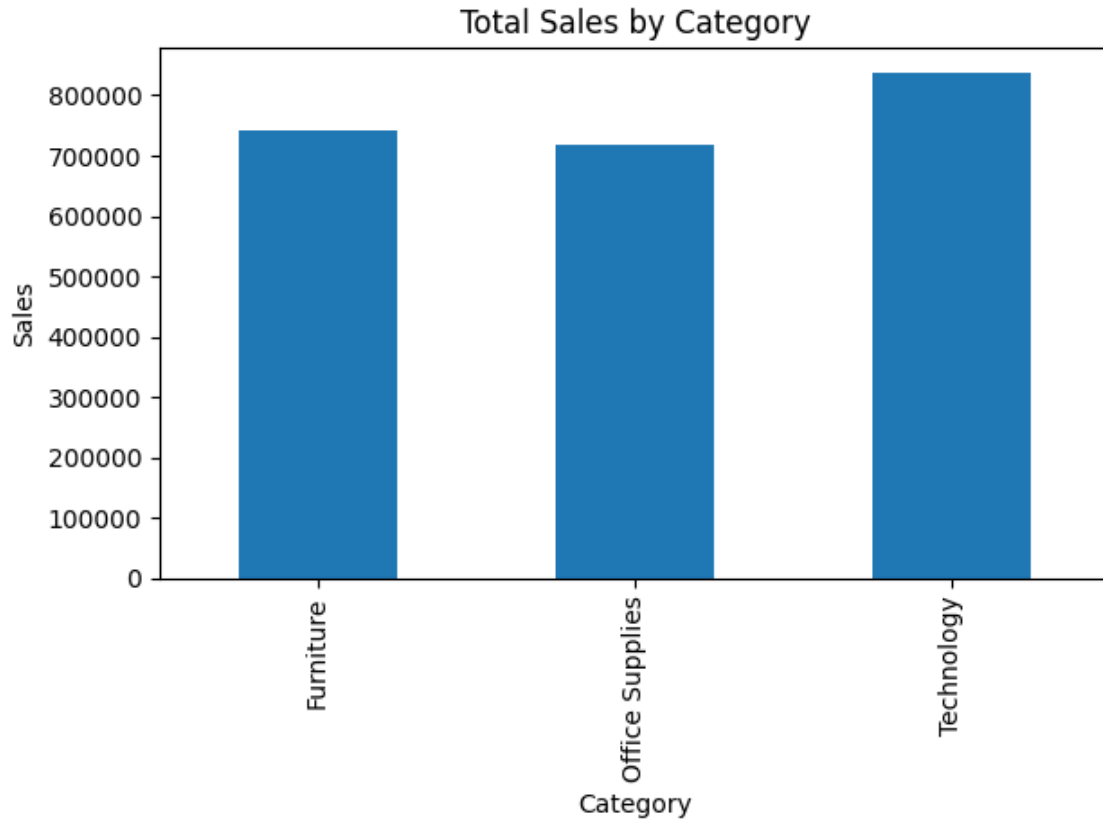
1. مخطط شريطي (Chart Bar) لإظهار إجمالي المبيعات لكل فئة منتج؟

الحل

الشفرة (Code)

```
sales_by_category = SuperstoreData.groupby('Category')['Sales'].sum()
plt.figure()
sales_by_category.plot(kind='bar')
plt.title('Total Sales by Category') # عنوان الرسم البياني
plt.xlabel('Category') # تسمية المحور الأفقي (الفئة التصنيفية)
plt.ylabel('Number of Sales') # تسمية المحور الرأسي (عدد المبيعات)
plt.show() # عرض الرسم البياني
```

نتيجة تنفيذ Code



التوصيات	استنتاجات	ملاحظات
الاستمرار في اتباع نفس الاستراتيجية	الفئات مبيعاتها ممتازة	نلاحظ ان فئة Tech أكثر الفئات مبيعا ومتقاربة مع الفئات الخراء

2. مخطط دائري (Chart Pie) لنسبة المبيعات لكل منطقة؟

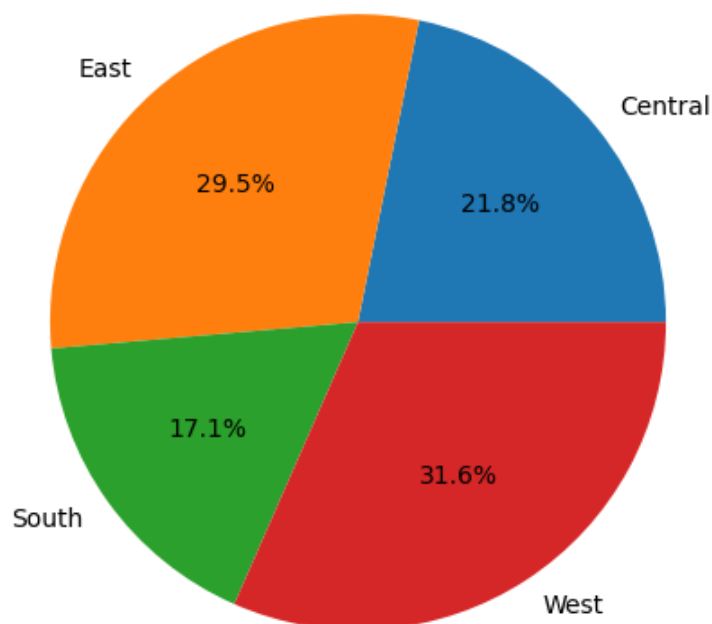
الحل

شفرة (Code)

```
region_sales = SuperstoreData.groupby('Region')['Sales'].sum()
plt.figure()
region_sales.plot(kind='pie', autopct='%1.1f%%')
plt.title('Sales Percentage by Region')
plt.show()
```

نتيجة تنفيذ Code

Sales Percentage by Region



ملاحظات	استنتاجات	التوصيات
نلاحظ ان فرعين مبيعاتهم أكثر من 60%	توجد مبيعات منخفضة بشكل كبير	اتخاذ الاجراء اللازمه لزيادة المبيعات في الفروع متدنية العمل

3. مخطط نقاط (Plot Scatter) للعلاقة بين كمية المبيعات والربح؟

الحل

شفرة (Code)

```
plt.figure()
plt.scatter(SuperstoreData['Quantity'], SuperstoreData['Profit'])
plt.title('Relationship Between Quantity and Profit')
plt.xlabel('Quantity')
plt.ylabel('Profit')
plt.show()
```

نتيجة تنفيذ Code



التوصيات	استنتاجات	ملاحظات
إعادة النظر في سياسات المتجر لتلافي الخسائر المحتملة	هناك احتمال إفلاس المتجر وخروجه من السوق بدين كبير	كلما زادت المبيعات قلت الخسائر

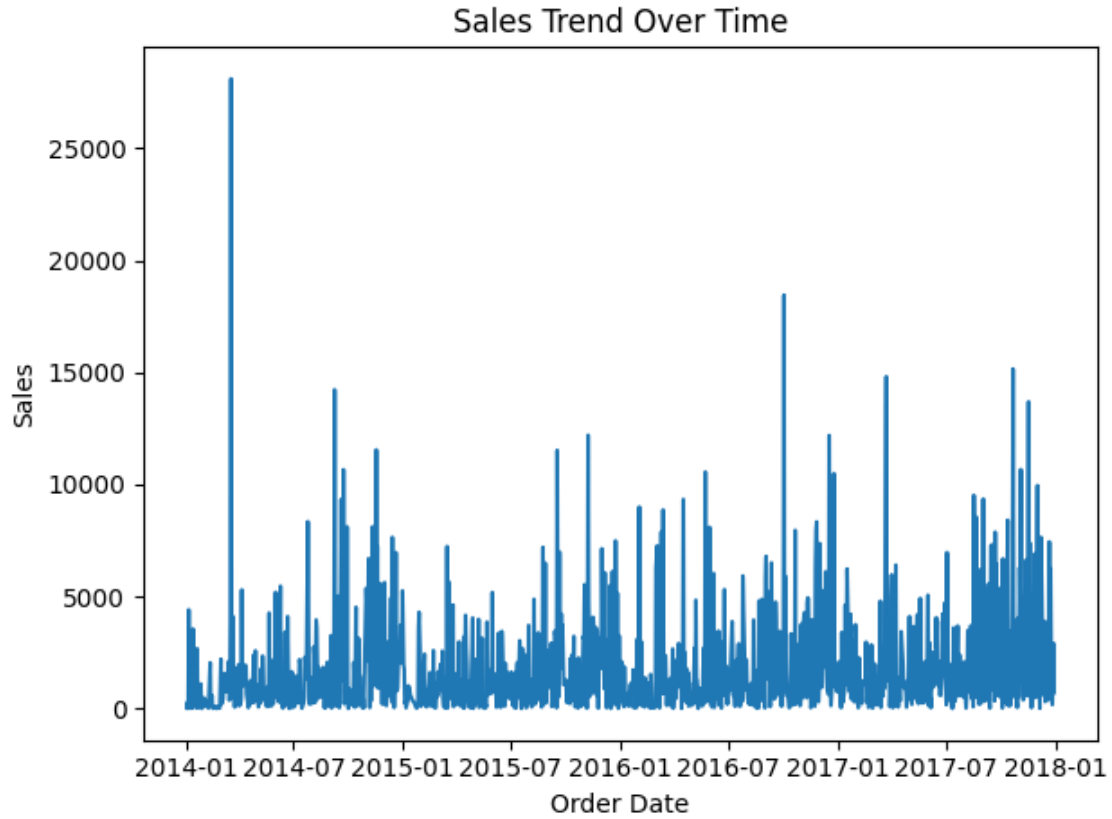
4. مخطط خطي (Chart Line) لعرض تطور المبيعات عبر الوقت إن وُجد عمود تاريخ؟

الحل

شفرة (Code)

```
sales_over_time = SuperstoreData.groupby('Order Date')['Sales'].sum()
plt.figure()
plt.plot(sales_over_time)
plt.title('Sales Trend Over Time')
plt.show()
```

نتيجة تنفيذ Code



التوصيات	استنتاجات	ملاحظات
التوسع الالكتروني لزيادة المبيعات	نستنتج ان المبيعات في حالت نمو بطيء ومضطرب	نلاحظ ان المبيعات تزداد كل عام من 1% إلى 5% سنوياً

5. أي رسم بياني آخر يمكن ك إنشاءه يدعم تحليل؟

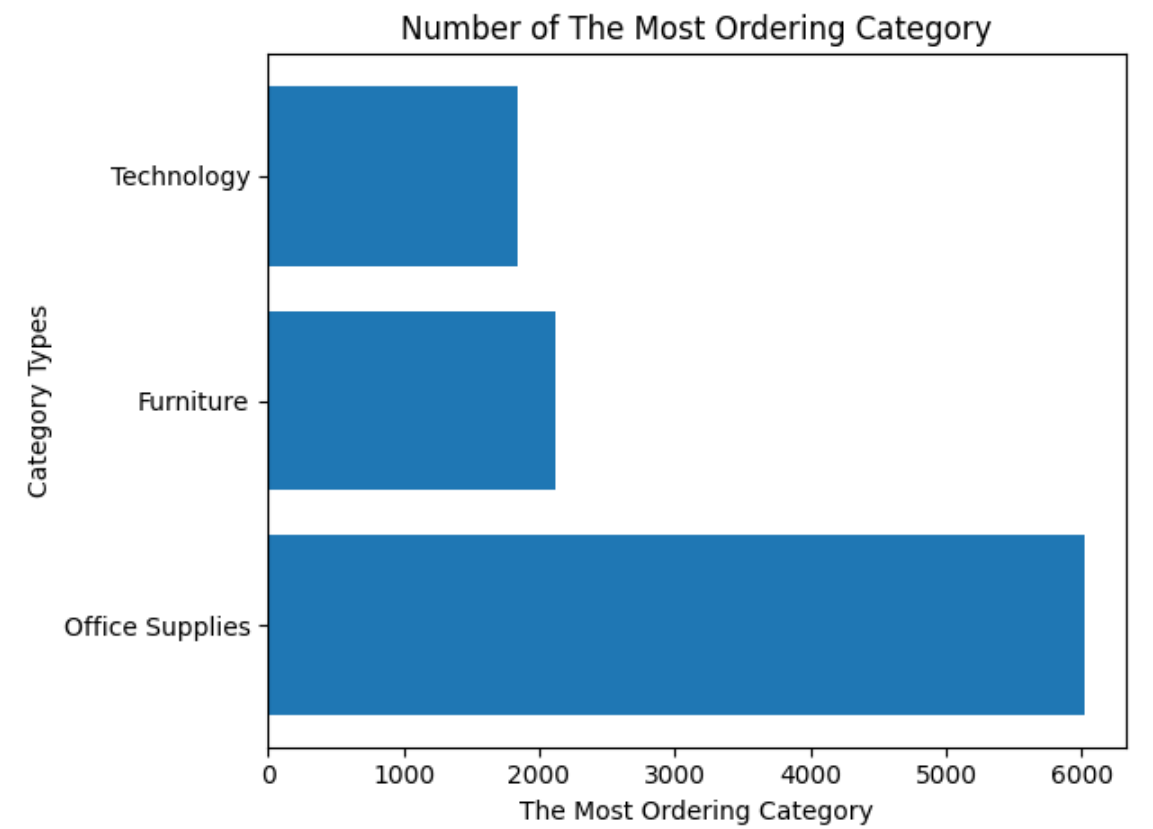
الحل

شفرة (Code)

```
# عد أنواع الفئات
Category_counts = SuperstoreData['Category'].value_counts()
# print(Category_counts)
# print(Category_counts.index)
# print(Category_counts.values)
# رسم المخطط الشريطي الأفقي
plt.barh(Category_counts.index, Category_counts.values)
```

```
plt.title('Number of The Most Ordering Category')
plt.xlabel('The Most Ordering Category')
plt.ylabel('Category Types')
plt.show()
```

نتيجة تنفيذ Code



التوصيات	استنتاجات	ملاحظات
عمل حملات إعلانية لجذب عملاء جدد لبيع الفئات الأخرى الأقل مبيعا	نستنتج ان معظم الزبائن هم من أصحاب الشركات ونو الاعمال الخاصة	نلاحظ ان Office Supplies الأكثر طلبا من بين الفئات الاخرى