

تقرير تحليلي - Superstore

نتائج تحليل ملف البيانات SampleSuperstore.csv

مقرر: ذكاء اصطناعي

المهندس: زكريا عبدالعزيز محمد الحنيني

جدول المحتويات

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

مقدمة التحليل:-

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

١) قراءة وتنظيف البيانات: -

يشمل هذا الجزء حساب:

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

١. اقرأ ملف البيانات باستخدام `pandas.csv_read`.

الحل

١ (Code) الشفرة

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

نتيجة تنفيذ ١ Code

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-2016-132155	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	-	42420	South	FUR-BQ-1000045	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600	2	0.00	41.9136
1	2	CA-2016-132156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	-	42420	South	FUR-ES-1000045	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	Darin Van Huff	Corporate	United States	Los Angeles	-	90036	West	OFF-LA-10000454	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b..	14.6200	2	0.00	8.8714
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	Brettford CH4500 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-10000760	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.3680	2	0.20	2.5164

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

٢ (Code) الشفرة

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

نتيجة تنفيذ ٢ Code

```

<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

```

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

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

الحل

1 (Code) الشفرة

```

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

```

1 Code نتائج تنفيذ

```

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

```

الوصيات	استنتاجات	ملاحظات
التأكد من ان البيانات نظيفة وجاهرة للمرحلة التالية	البيانات نظيفة	البيانات لا تحتوي على قيم فارغة

الشفرة 2 (Code)

```

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

```

نتيجة تنفيذ 2 Code

```

<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

```

الوصيات	استنتاجات	ملاحظات
فحص أنواع البيانات	البيانات لا تحتوي على قيم فارغة	فيطننا يومئذ 9994 ملخص بيانات دعوة

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

الحل

(Code) شفرة

```

# تحويل التاريخ
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	9994.000000	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	55190379428	229.858001	3.789574	0.156203
min	1.000000	2014-01-03 00:00:00	2014-01-07 00:00:00	1040.000000	0.444000	1.000000	0.000000
25%	2499.250000	2015-05-23 00:00:00	2015-05-27 00:00:00	23223.000000	17.280000	2.000000	0.000000
50%	4997.500000	2016-06-26 00:00:00	2016-06-29 00:00:00	56430.500000	54.490000	3.000000	0.200000
75%	7495.750000	2017-05-14 00:00:00	2017-05-18 00:00:00	90008.800000	209.940000	5.000000	0.200000
max	9994.000000	2017-12-30 00:00:00	2018-01-05 00:00:00	99301.000000	22638.480000	14.000000	0.800000
std	2885.163629	NaN	NaN	32063.693350	623.245101	2.225110	0.206452

الوصيات	استنتاجات	ملاحظات
إدخال التاريخ بشكل كامل لكي لا يعتبر التاريخ رقم	جميع القيم سليمة	لا يوجد أي قيم خطأ، الخطاء الذي يظهر في عمودي (Order) (Date, Ship Date) لأن القيم عبار عن تاريخ وليس ارقام

(2) التحليل الأساسي: -

يشمل هذا الجزء حساب:

- ❖ حساب إجمالي المبيعات (Sales) لكل فئة منتج (Category).
- ❖ حساب متوسط الربح (Profit) لكل منطقة (Region).
- ❖ ترتيب العملاء حسب حجم مشترياتهم بناء على مجموعة المبيعات يمكنك تقسيمهن إلى مجموعات حسب النسبة المئوية أو حدود تحدها.

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	

الوصيات	استنتاجات	ملاحظات
الاهتمام لفنانات التي تمتلك إجمالي مبيعات قليل	لكن نستنتج ان فنانة Tech هي التي تملك الإجمالي الأكبر	نلاحظ ان اجمالي المبيعات لثلاثة الفنانات متقارب

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

الحل

(Code) شفرة

```
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
```

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

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

الحل

(Code) شفرة

```
# 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

```

الوصيات	استنتاجات	ملاحظات
أستهان الفئتين الأقل شراث ودراسة سلوكهم	نستنتج ان اغلب العملاء من الفئة المتوسطة	نلاحظ ان نسبة المشتريات من الفئة المتوسطة هي الاكثر

٣) 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 والربح.

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

الحل

(Code) شفرة

```
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

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

(2) تحليل أداء الفروع (Segment) حسب المبيعات والربح؟

الحل

(Code) شفرة

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

نتيجة تنفيذ Code

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

الوصيات	استنتاجات	ملاحظات
دراسة وضع الفروع الاقل مبيعا	مبيعات الفروع متقاربه لذلك الربح متقاربه ايضا	نلاحظ انهو كلما زادت المبيعات زادت الارباح

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 خطى (Chart Line) لعرض تطور المبيعات عبر الوقت إن وجد عمود تاريخ.
- ❖ أي رسم بياني آخر يمكن إنشاؤه يدعم تحليلاً.

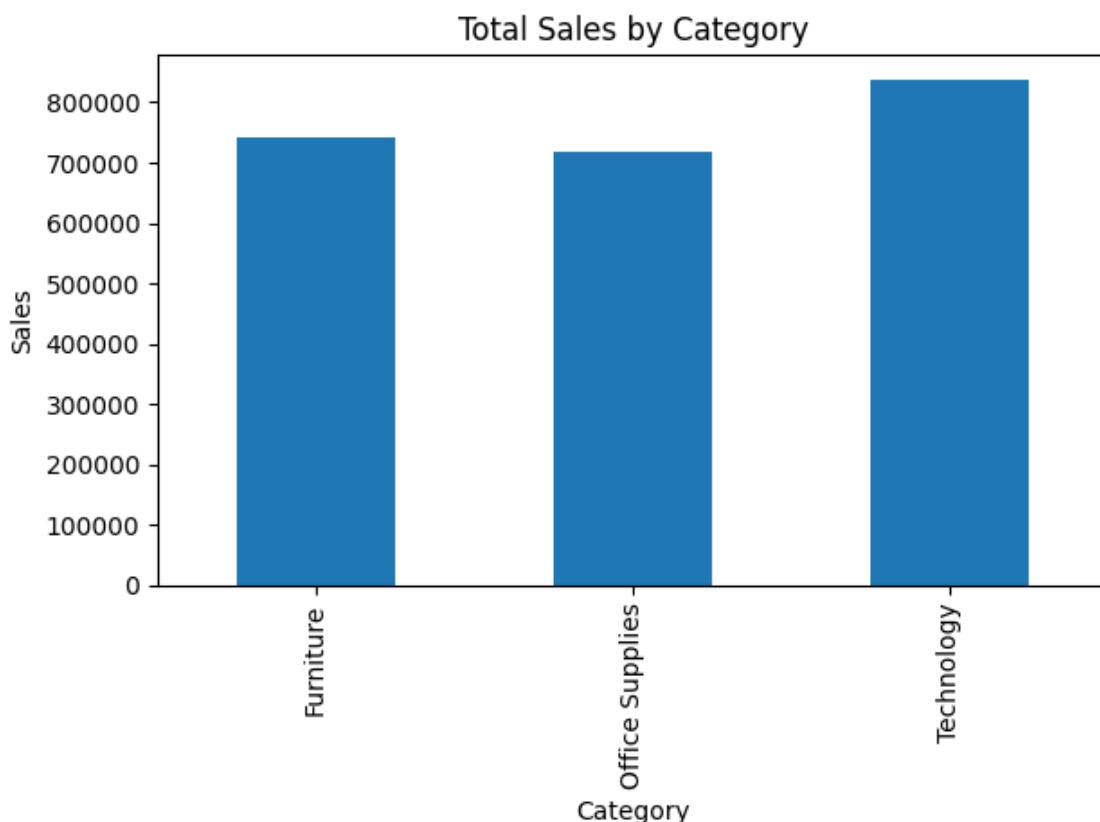
1. مخطط شريطي(Chart Bar) لإظهار إجمالي المبيعات لكل فئة منتج؟

الحل

الشفرة (Code)

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

نتيجة تنفيذ Code



الوصيات	استنتاجات	ملاحظات
الاستمرار في اتباع نفس الاستراتيجية	الفئات مبيعاتها متباينة	نلاحظ أن فئة Tech أكثر الفئات مبيعاً ومتقاربة مع الفئات الخرقاء

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

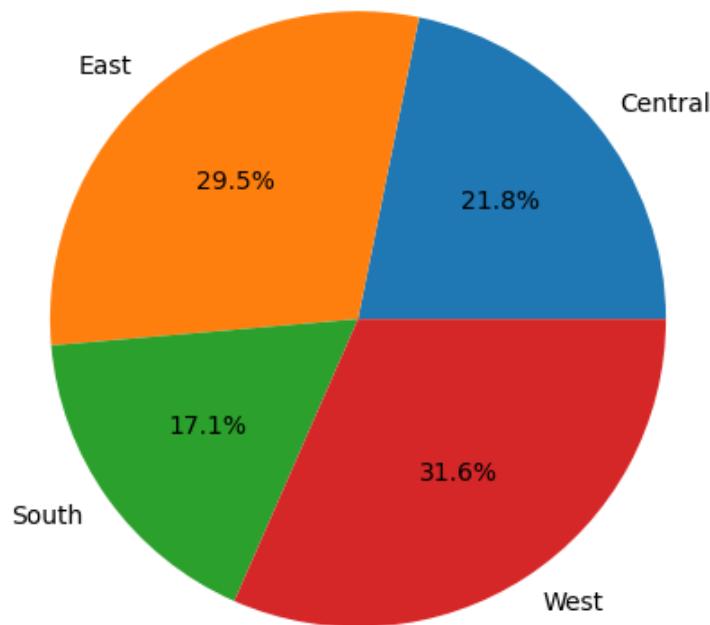
الحل

شفرة (Code)

```
region_sales = SuperstoreData.groupby('Region')['Sales'].sum()
plt.figure()
region_sales.plot(kind='pie', autopct='%.1f%%')
plt.title('Sales Percentage by Region')
plt.show()
```

نتيجة تنفيذ Code

Sales Percentage by Region



الوصيات	استنتاجات	ملاحظات
اتخاذ الاجراء الازمه لزياد المبيعات في الفروع متعددة العمل	توجد مبيعات منخفضه بشكل كبير	نلاحظ ان فرعين مبيعاتهم أكثر من %60

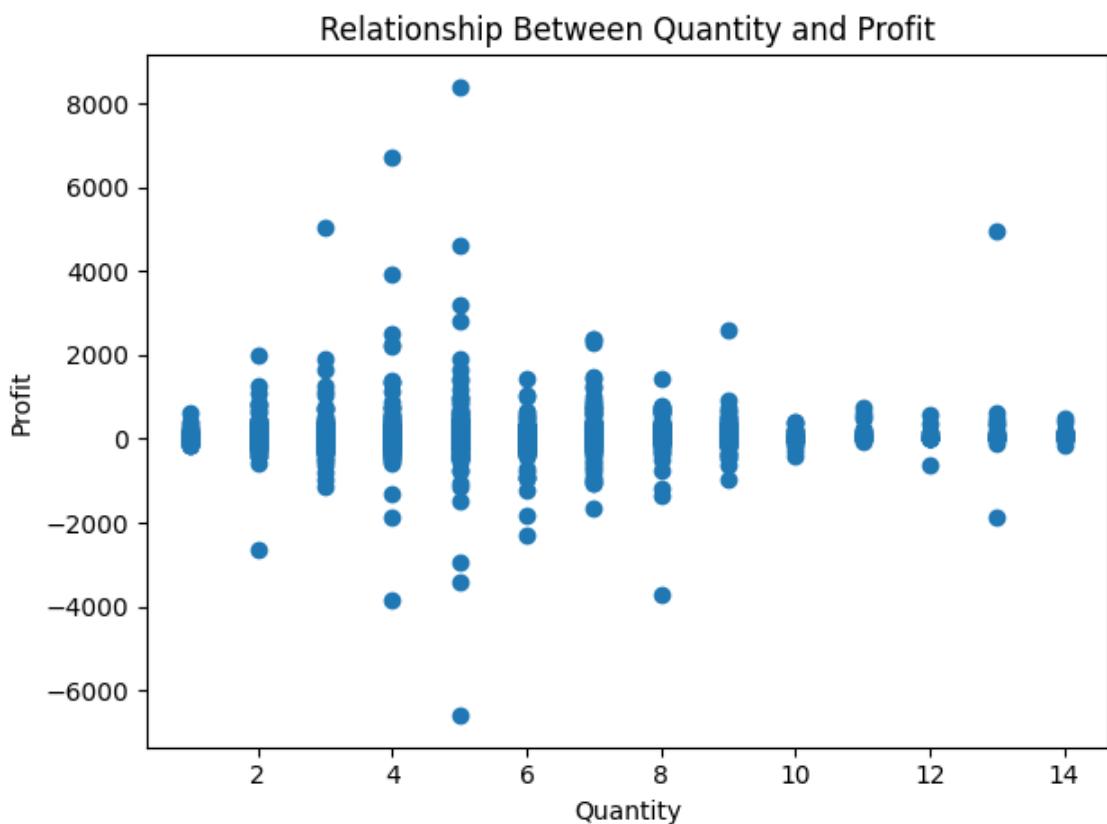
3. مخطط نقاط (Plot Scatter) للعلاقة بين كمية المبيعات والربح؟

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



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

4. مخطط خطى (Chart Line) لعرض تطور المبيعات عبر الوقت إنْ وجد عمود تاريخ؟

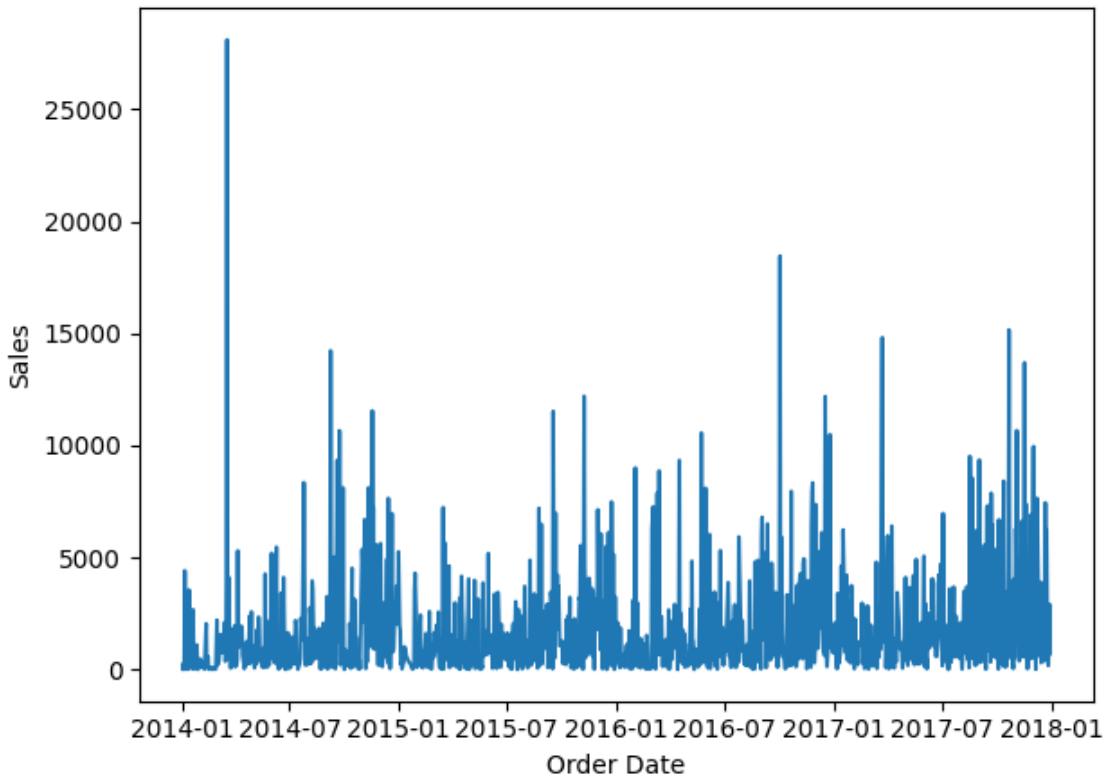
الحل

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

Sales Trend Over Time



الوصيات	استنتاجات	ملاحظات
التوسيع الإلكتروني لزيادة المبيعات	نستنتج ان المبيعات في حالت نمو بطيء ومطرد	نلاحظ ان المبيعات تزداد كل عام من 1% إلى 5%

5. أي رسم بياني آخر يمكن إنشاؤه يدعم تحليل؟

الحل

(Code) شفرة

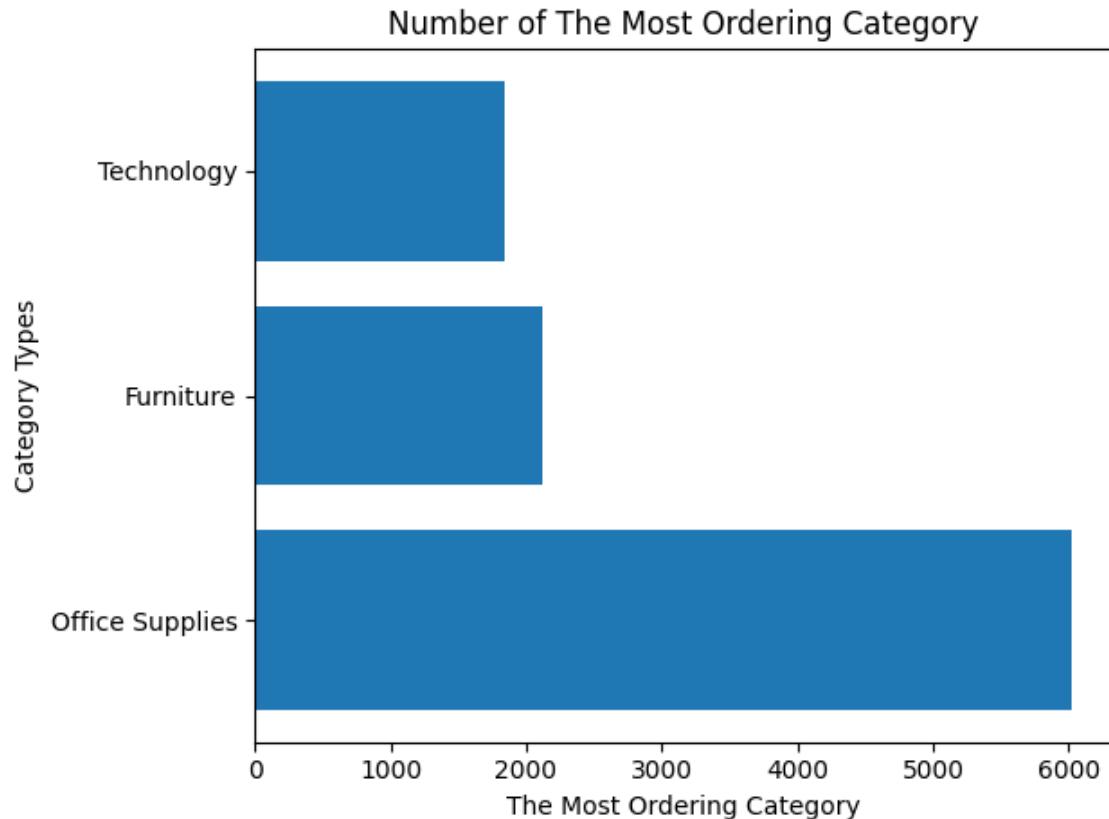
```
# عد أنواع الفئات
Category_counts = SuperstoreData['Category'].value_counts()
# print(Category_counts)
# print(Category_counts.index)
# print(Category_counts.values)
# رسم المخطط الشريطي الأفقي
plt.barh(Category_counts.index, Category_counts.values)
```

```

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

```

نتيجة تنفيذ Code



الوصيات	استنتاجات	ملاحظات
عمل حمالات إعلانية لجذب عملاء جدد لبع الفئات الأخرى الأقل مبيعا	نستنتج أن معظم الزبائن هم من أصحاب الشركات وذو الاعمال الخاصة	نلاحظ ان Office Supplies الأكثر طلبا من بين الفئات الاخرياء