تقرير :TP03 تحليل البيانات باستخدام Python وPower BI

1. مقدمة:

يهدف هذا العمل التطبيقي إلى تطبيق خطوات تحليل البيانات باستخدام لغةPython ، بدايةً من تحميل البيانات ومعالجتها، وصولاً إلى تحليلها وعرض النتائج باستخدام.Power BI

تم الاعتماد على ملف بيانات بصيغة CSV ضمن مشروع التنبؤ بالمبيعات(Store Sales) ، حيث قمنا بالعمل على ملف.test.csv

2. وصف البيانات:

البيانات المستخدمة مأخوذة من Dataset خاص بالمتاجر، وتحتوي على المعلومات التالية:

تاريخ العملية :date

رقم المتجر :store_nbr

نوع المنتج :family

هل المنتج ضمن عرض ترويجي :onpromotion

.ملاحظة:

ملف test.csv لا يحتوي على عمود المبيعات

لذلك لا يمكن تحليل الأرباح أو الأداء المالي يستخدم هذا الملف أساساً في التنبؤ بالمبيعات.

3 . الأدوات المستعملة

Python :لغة البرمجة

المكتبات :

لمعالجة البيانات Pandas

أداة العرض :

Power BI Desktop

4. خطوات العمل:

1.4 . جمع البيانات(Data Collection)

تم تحميل البيانات من ملف CSV باستخدام مكتبةPandas ، وذلك عبر الكود التالي:

```
import pandas as pd data = pd.read_csv('test.csv') print(data.head()) print(data.info())
```

الهدف:

قراءة البيانات

عرض أول الصفوف

معرفة أنواع الأعمدة وعدد القيم

2.4. تنظيف البيانات(Data Cleaning)

تم تنظيف البيانات من خلال حذف القيم الفارغة، وذلك باستخدام الكود:

```
data_cleaned = data.dropna()
```

الهدف:

إزالة الصفوف التي تحتوي على قيم ناقصة

تحسين جودة البيانات

3.4.حفظ البيانات

بعد التنظيف، تم حفظ البيانات في ملف جديد:

```
data_cleaned.to_csv('test_cleaned.csv', index=False)
```

الهدف:

الاحتفاظ بنسخة نظيفة من البيانات

استعمالها لاحقاً في التحليل أوPower BI

4.4تحليل البيانات(EDA)

تم إجراء تحليل بسيط لفهم بنية البيانات، من خلال:

عرض الإحصائيات العامة

التعرف على أنواع البيانات

فهم توزيع القيم

📌بما أن الملف لا يحتوي على المبيعات، فقد ركز التحليل على:

أنواع المنتجات

العروض الترويجية

توزيع البيانات

5.عرض وتحليل النتائج(Power BI)

تم استخدام Power BI لإنشاء مجموعة من الرسوم البيانية التي تساعد في فهم البيانات بشكل بصري، وكانت النتائج كما يلي:

1.5.مخطط خطي(Line Chart)

يمثل هذا المخطط تطور عدد العروض الترويجية (onpromotion) حسب نوع المنتج.(family)

الملاحظة:

يوجد تفاوت بين المنتجات من حيث عدد العروض

بعض الفئات مثل المواد الغذائية (Grocery) لديها عدد عروض أكبر مقارنة بفئات أخرى

2.5.مخطط أعمدة(Bar Chart)

يعرض هذا المخطط عدد القيم حسب نوع المنتج.(family)

✓الملاحظة:

البيانات موزعة على عدة أنواع من المنتجات

بعض الفئات تحتوي على عدد أكبر من السجلات، مما يدل على كثرة توفرها أو أهميتها

3.5.مخطط دائري(Pie Chart)

يوضح هذا المخطط توزيع البيانات حسب نوع المنتج.(family)

الملاحظة:

التوزيع متنوع بين الفئات

لا توجد فئة واحدة تهيمن بشكل كامل على البيانات

4.5 .مخطط دائري مجوف(Donut Chart)

يمثل هذا المخطط توزيع القيم حسب المعرف.(id)

الملاحظة:

البيانات تحتوي على عدد كبير من القيم المختلفة

هذا يدل على أن البيانات مفصلة وتشمل عدد كبير من العمليات

تحليل عام :

من خلال الرسوم البيانية، يمكن استنتاج ما يلي:

البيانات تحتوي على تنوع كبير في المنتجات

هناك اختلاف في عدد العروض الترويجية بين الفئات

توزيع البيانات متوازن نسبياً

البيانات مناسبة للتحليل المتقدم أو التنبؤ

خلاصة التحليل :

ساعدت الرسوم البيانية في Power BI على:

فهم توزيع البيانات

اكتشاف الفئات الأكثر تمثيلاً

تحليل العروض الترويجية

مما يسهل اتخاذ قرارات مستقبلية أو بناء نماذج تنبؤية .

.6.ملاحظات :

ملف test.csv لا يحتوي على عمود المبيعات

لذلك لا يمكن تحليل الأرباح أو الأداء المالي

يستخدم هذا الملف أساساً في التنبؤ بالمبيعات

.7.خاتمة :

في هذا العمل، تم تطبيق المراحل الأساسية لتحليل البيانات، بدءاً من جمع البيانات وتنظيفها، وصولاً إلى عرض النتائج باستخدام Power BI.

ساعد هذا المشروع في فهم كيفية التعامل مع البيانات الحقيقية وتحضيرها للاستخدام في التحليل أو التنبؤ.