

➤ الخطوة الأولى:

البحث عن البيانات وتحميلها، كانت بيانات مبيعات (Walmart Sales) وهي أكبر شركة تجزئة في العالم، تأسست عام 1962 في الولايات المتحدة وتدير سلسلة متاجر ضخمة تباع بأسعار مخفضة. تقدم الشركة منتجات متنوعة تشمل البقالة، الملابس، الإلكترونيات وغيرها، وتخدم ملايين العملاء يومياً حول العالم.

تم الحصول على البيانات من موقع (Kaggle)، وهو موقع تابع لشركة Google يوفر منصة لمجتمع علوم البيانات للمنافسة، التعلم، ومشاركة المشاريع والبيانات.

➤ الخطوة الثانية:

تم تحميل البيانات بصيغة (CSV)، وهي اختصار لـ (Comma-Separated Values) وهي تنسيق بسيط لتخزين البيانات في جداول، حيث يتم فصل القيم داخل كل صف بواسطة فواصل مثل الفاصلة "،" أو الفاصلة المنقوطة "؛".

من أبرز خصائص إستخدامها:

1. صيغة مدعومة من أغلب الأنظمة.
2. حجمها خفيف مقارنة بصيغ مثل Excel.
3. تُستخدم بكثرة في علم البيانات وتعلم الآلة لتحميل وتدريب النماذج.

وتظهر بهذا الشكل: رقم (1)

ولجعلها تكون مثل الشكل رقم (2) في إكسيل. إتبع الخطوات التالية.

(1)

1	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
2	1	05-02-2010	1643690	0	42.31	2.572	211.0964	8.106
3	1	12-02-2010	1641957	44	38.51	2.548	211.2422	8.106
4	1	19-02-2010	1611968	17	39.93	2.514	211.2891	8.106
5	1	26-02-2010	1409727	59	46.63	2.561	211.3196	8.106
6	1	05-03-2010	1554807	0	46.5	2.625	211.3501	8.106
7	1	12-03-2010	1439541	59	57.79	2.667	211.3806	8.106
8	1	19-03-2010	1472516	79	54.58	2.72	211.2156	8.106
9	1	26-03-2010	1404430	92	51.45	2.732	211.0198	8.106
10	1	02-04-2010	1594968	28	62.27	2.719	210.8204	7.808
11	1	09-04-2010	1545418	53	65.86	2.77	210.6229	7.808
12	1	16-04-2010	1466058	28	66.32	2.808	210.4887	7.808
13	1	23-04-2010	1391256	12	64.84	2.795	210.4391	7.808
14	1	30-04-2010	1425101	0	67.41	2.78	210.3895	7.808
15	1	07-05-2010	1603955	12	72.55	2.835	210.34	7.808
16	1	14-05-2010	1494252	5	74.78	2.854	210.3374	7.808
17	1	21-05-2010	1399662	0	76.44	2.826	210.6171	7.808
18	1	28-05-2010	1432070	0	80.44	2.759	210.8968	7.808
19	1	04-06-2010	1615525	0	80.69	2.705	211.1764	7.808
20	1	11-06-2010	1542561	0	80.43	2.695	211.4561	7.808
21	1	18-06-2010	1503284	0	84.11	2.637	211.4538	7.808
22	1	25-06-2010	1422712	0	84.34	2.653	211.3387	7.808
23	1	02-07-2010	1460418	0	80.91	2.669	211.2235	7.787
24	1	09-07-2010	1546074	0	80.48	2.642	211.1084	7.787
25	1	16-07-2010	1448939	0	83.15	2.623	211.1004	7.787
26	1	23-07-2010	1385065	0	83.36	2.608	211.2351	7.787
27	1	30-07-2010	1371986	0	81.84	2.64	211.3699	7.787
28	1	06-08-2010	1605491	78	80.87	2.627	211.5047	7.787
29	1	13-08-2010	1508238	0	87	2.692	211.0394	7.787
30	1	20-08-2010	1513080	0	86.65	2.664	211.6034	7.787
31	1	27-08-2010	1448143	0	84.77	2.618	211.6273	7.787

(2)

الخطوات: (EXCEL)

1	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
2	1	05-02-2010	1643691	0	42.31	2.572	211.0964	8.106
3	1	12-02-2010	1641957	1	38.51	2.548	211.2422	8.106
4	1	19-02-2010	1611968	0	39.93	2.514	211.2891	8.106
5	1	26-02-2010	1409728	0	46.63	2.561	211.3196	8.106
6	1	05-03-2010	1554807	0	46.5	2.625	211.3501	8.106
7	1	12-03-2010	1439542	0	57.79	2.667	211.3806	8.106
8	1	19-03-2010	1472516	0	54.58	2.72	211.2156	8.106
9	1	26-03-2010	1404430	0	51.45	2.732	211.0198	8.106
10	1	02-04-2010	1594968	0	62.27	2.719	210.8204	7.808
11	1	09-04-2010	1545419	0	65.86	2.77	210.6229	7.808
12	1	16-04-2010	1466058	0	66.32	2.808	210.4887	7.808
13	1	23-04-2010	1391256	0	64.84	2.795	210.4391	7.808
14	1	30-04-2010	1425101	0	67.41	2.78	210.3895	7.808
15	1	07-05-2010	1603955	0	72.55	2.835	210.34	7.808
16	1	14-05-2010	1494252	0	74.78	2.854	210.3374	7.808
17	1	21-05-2010	1399662	0	76.44	2.826	210.6171	7.808
18	1	28-05-2010	1432070	0	80.44	2.759	210.8968	7.808
19	1	04-06-2010	1615525	0	80.69	2.705	211.1764	7.808
20	1	11-06-2010	1542561	0	80.43	2.695	211.4561	7.808
21	1	18-06-2010	1503284	0	84.11	2.637	211.4538	7.808
22	1	25-06-2010	1422712	0	84.34	2.653	211.3387	7.808
23	1	02-07-2010	1460418	0	80.91	2.669	211.2235	7.787
24	1	09-07-2010	1546074	0	80.48	2.642	211.1084	7.787
25	1	16-07-2010	1448939	0	83.15	2.623	211.1004	7.787
26	1	23-07-2010	1385065	0	83.36	2.608	211.2351	7.787
27	1	30-07-2010	1371987	0	81.84	2.64	211.3699	7.787
28	1	06-08-2010	1605492	0	87.16	2.627	211.5047	7.787
29	1	13-08-2010	1508238	0	87	2.692	211.0394	7.787
30	1	20-08-2010	1513080	0	86.65	2.664	211.6034	7.787
31	1	27-08-2010	1448143	0	84.77	2.618	211.6273	7.787

بإستخدام ميزة "Text to Columns"

- 1- حدد العمود الذي يحتوي على البيانات غير المنسقة، (عادةً يكون العمود الأول).
- 2- اذهب إلى "Data" في شريط الأدوات، ثم اختر "Text to Columns".
- 3- اختر "Delimited" ثم اضغط Next.

4- في قائمة **Delimiters**، حدد الفاصلة (,) إذا كانت البيانات مفصولة بفواصل، ثم اضغط **Next**.

5- اضغط **Finish**، وستظهر البيانات مقسمة بشكل صحيح في الأعمدة.

ولكن قمت بالعمل على إستدعاء البيانات مباشرةً بصيغة (CSV) داخل PowerBi من خلال الخطوات التالية:

الخطوات: (Power Bi)

- 1- إفتح **Power Bi**
- 2- إذهب إلى **Home** من ثم **Get Data**
- 3- إختار صيغة الملف وهي **Text/Csv**
- 4- حدد الفاصلة من خلال (**Delimiter**) في الأغلب تكون (**Comma**)
- 5- إضغط (**Load**) أو **transform data** لتحويلك لـ **Power Query**
- 6- **File Origin** يحدد ترميز النصوص في الملف مثل UTF-8 لضمان قراءة البيانات بشكل صحيح.
- 7- **Data Type Detection** يساعد Power BI في تحديد نوع البيانات مثل (نص، رقم، تاريخ) لكل عمود بشكل تلقائي أو يدوي.

➤ الخطوة الثالثة:

فهم البيانات والأعمدة بشكل جيد.

Store Number	يتم تسجيل مجموع المبيعات إسبوعياً	المبيعات الأسبوعية لكل متجر	علامة على وجود أو عدم وجود عطلة، 1: يوجد عطلة 0: لا يوجد.	درجة حرارة الجو في منطقة المتجر خلال الأسبوع، لمعرفة تأثير الطقس على حجم المبيعات.	متوسط سعر الوقود في المنطقة خلال الأسبوع، قد يؤثر على حركة المستهلكين والمبيعات، خاصة في المواسم.
--------------	-----------------------------------	-----------------------------	---	--	---

معدل البطالة في المنطقة خلال نفس الفترة، يُظهر الحالة الاقتصادية العامة التي قد تؤثر على المبيعات.

Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
2	يوليو, 02 2010	2003940.64	0	82.74	2.669	210.8803726	8.099
2	يوليو, 09 2010	1880902.62	0	82.59	2.642	210.7657317	8.099
2	يوليو, 16 2010	1845879.79	0	85.32	2.623	210.7577954	8.099

Consumer Price Index:

يُستخدم لتقييم التضخم وتأثيره على القوة الشرائية. يعني كلما زادت أسعار السلع قد تقل المبيعات، لعدم قدرة المشتري على شراء نفس السلع كما كان يشتريها قبل التضخم لأن نفس المبلغ الذي خصصه الشخص لشراء السلع لا يمكنه شراء نفس الكميات لإرتفاع أسعارها.

التأكد من جاهزية البيانات باستخدام Power Query Editor.

كانت معظم البيانات جاهزة بدون أي أخطاء أو خلايا فارغة، أو تفاوت في الأرقام.

باستثناء تعديلات بسيطة، زي عمود (Weekly_Sales) حولته من: Fixed Decimal Number لـ Whole Number

وعمود (CPI) قللت الأرقام العشرية (decimal places) .. وبقوا بالشكل التالي:

\$ Weekly_Sales	1.2 CPI
Valid 100%	Valid 100%
Error 0%	Error 0%
Empty 0%	Empty 0%
1,643,690.90	211.096
1,641,957.44	211.242
1,611,968.17	211.289
1,409,727.59	211.32
1,554,806.68	211.35
1,439,541.59	211.381
1,472,515.79	211.216
1,404,429.92	211.018

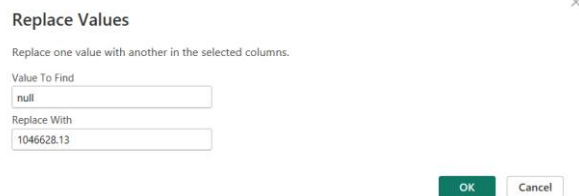
ولكن على سبيل التدريب على: Power Query، قمت بجعل بعض الخلايا فارغة في عمود: Weekly_Sales وهنا ييجي دور إزاي هعالج البيانات دي (الخلايا الفارغة). وكان الحل إني أعوض الخلايا الفارغة بمتوسط المبيعات.

قمت أولاً بإستبدال بعض الخلايا لجعلها (null) من خلال: الوقوف على الخلية من ثم:

Right Click → Replace Value → Replace With = null وقمت بإستبدال عدد من الخلايا وجعلها تكون (null).

بعد ذلك أريد تعويض القيم الفارغة وحل المشكلة وذلك من خلال متوسط المبيعات الأسبوعية. عن طريق التالي: أولاً نريد الحصول على متوسط المبيعات لدينا، من خلال **Average → Statistics → Transform → Select column** ونقوم بأخذ الرقم وهو المتوسط لدينا. ونعود إلى البيانات من ثم:

Select column → Right Click → Replace Value → Value To Find = null → Replace With = Average number



بعدين باستخدام (DAX) حولت عمود (Holiday_Flag) بمعنى إن رقم 1 يبقى يوجد عطلة و 0 لا يوجد.
وده علشان تسهل عملية القراءة مش أكثر..

من خلال التالي: Modeling → New Column

```
1 Holiday_Description =
2 IF(
3     'Walmart_Sales'[Holiday_Flag] = 1,
4     "Holiday",
5     "No Holiday"
6 )
```

إسم الجدول. ← 1
إسم العمود اللي هيبحت فيه. ← 'Walmart_Sales'[Holiday_Flag]
لو كان الشرط = 1 يبقى يوجد عطلة لو غير كده يبقى لا يوجد عطلة
إسم العمود = Holiday_Description

وكمان حولت عمود الـ (Store) وكان به عدد 45 فرع مرقمين 1 - 2 - 3 حولتهم للتالي:

Branch 1 Branch 2 Branch 3 ...

من خلال التالي: Modeling → New Column

```
1 Branch_Name = "Branch " & 'Walmart_Sales'[Store]
```

نص ثابت، هيطهر في بداية كل قيمة في العمود، + أي نص ثابت لازم يكتب بين "علامتي تنصيص"
Concatenations عامل للدمج يعني ضم النصوص مع بعض.
القيمة من عمود Store في جدول Walmart_Sales

وبكده بقوا بالشكل التالي: أسهل وضوحاً وللقراءة

Store	Branch_Name	Holiday_Flag	Holiday_Description
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
2	Branch 2	0	No Holiday
15	Branch 15	0	No Holiday
15	Branch 15	0	No Holiday

➤ الخطوة الرابعة:

أنا محتاج إيه؟ محتاج أجاب على إيه وأعرف إيه تحديداً..

- 1- هل هناك تأثير كبير من التضخم على المبيعات..
- 2- مدى تأثير معدل البطالة على المبيعات..
- 3- مدى تأثير الطقس على حالة المبيعات..
- 4- هل سعر الوقود يؤثر على حركة المستهلكين والمبيعات..
- 5- مدى قوة تأثير العطلات على المتاجر..
- 6- إجمالي المبيعات
- 7- أعلى المتاجر تحقيقاً للمبيعات
- 8- المبيعات حسب السنوات
- 9- المبيعات حسب الأرباع السنوية
- 10- مقارنة المبيعات في العطلات مقابل الأيام العادية
- 11- مقارنة المبيعات بين الفروع أو المناطق
- 12- أداء المبيعات الأسبوعي/الشهري (الاتجاه العام)
- 13- تحديد الأسابيع أو الشهور الأقل مبيعاً (فرص التحسين)

إستخدمت 3 أنواع من التحليل:

1- تحليل وصفي (Descriptive Analysis)

يهدف إلى وصف ما يحدث فعليًا، مثل عدد أيام العطلات، إجمالي المبيعات خلالها، مقارنة بالعادة.

2- تحليل الأعمال (Business Insight)

تحليل تأثير العطلات على الأداء (Holiday Impact)

- يسمى بـ "Holiday Sales Impact Analysis" أو "Event-Based Sales Analysis"
- يُستخدم لمعرفة:

- هل العطلات تؤثر إيجابياً أو سلبياً على المبيعات؟
- هل يجب زيادة المخزون أو العروض في فترات معينة؟
- ما المتاجر أو المناطق التي تستفيد أكثر من العطلات؟

3- تحليل سببي (Causal Analysis)

نحاول فيه تحديد السبب والتأثير.

نسأل: هل ارتفاع التضخم أو انخفاض درجات الحرارة يسبب تغيراً في المبيعات؟

Created by: **KAMEL ABDELWAHAB**

KAMEL ABDELWAHAB
LinkedIn & Github Account:
Profile Link:

