

صحافة البيانات

المحاضرة الخامسة

DGM 231

خطوات إنتاج القصص الصحفية
المدفوعة بالبيانات
(تنظيف البيانات)

إعداد
د/ نهلة صلاح



محاوّر المحاضرة

- أولاً: مفهوم عملية تنظيف البيانات.
- ثانياً: أهمية خطوة تنظيف البيانات.
- ثالثاً: خطوات تنظيف البيانات.
- رابعاً: التحقق ومراجعة البيانات.

أولاً: مفهوم عملية تنظيف البيانات



■ يحصل الصحفيون على البيانات من مصادر متعددة وبصيغ مختلفة، يصعب معها تحليل البيانات واستخراج القصص الخبرية منها ما لم يتم تنظيم هذه البيانات ووضعها في أشكال يمكن التعامل معها.

■ عملية تنظيف البيانات هي مرحلة إصلاح ومعالجة البيانات، والتحقق من صحتها، وقراءتها بشكل صحيح داخل الجداول بهدف تأهيلها لعملية التحليل والتصميم.

ثانيًا: أهمية خطوة تنظيف البيانات



خطوة تنظيف البيانات مهمة لعدة أسباب أهمها:

1. ضمان جودة البيانات:

تنظيف البيانات يساعد في إزالة الأخطاء والتكرارات والبيانات غير المكتملة، وهذا الذي يضمن أن البيانات المستخدمة في التحليل دقيقة.

2. تقليل المخاطر:

البيانات غير النظيفة ستؤدي إلى تحليلات وقرارات خاطئة، التي سوف تزود من هذه المخاطر، ولكن تنظيف البيانات يقلل من هذه المخاطر بشكل كبير.

ثانيًا: أهمية خطوة تنظيف البيانات

3. سرعة التحليل:

البيانات النظيفة والمنسقة بشكل صحيح بتسرع جدا عملية الحساب في الخطوات المتقدمة في التحليل.



ثانيًا: أهمية خطوة تنظيف البيانات



ويقدم خبراء صحافة البيانات عددًا من النصائح للصحفيين في مرحلة تنظيف البيانات بعد جمعها من المصادر المختلفة من أهمها:

1. الاحتفاظ بنسخة من البيانات في صورتها الأصلية للرجوع إليها في أي وقت يحتاجه الصحفي.

2. إنشاء "جدول تعريف للبيانات BIOGRAPHY DATA" يتضمن هذا الجدول سيرة ذاتية للبيانات التي جمعها الصحفي فهي بمثابة إجراء مقابلة مع البيانات حول من وماذا ومتى وأين ولماذا تم جمع البيانات؟.

ثانيًا: أهمية خطوة تنظيف البيانات

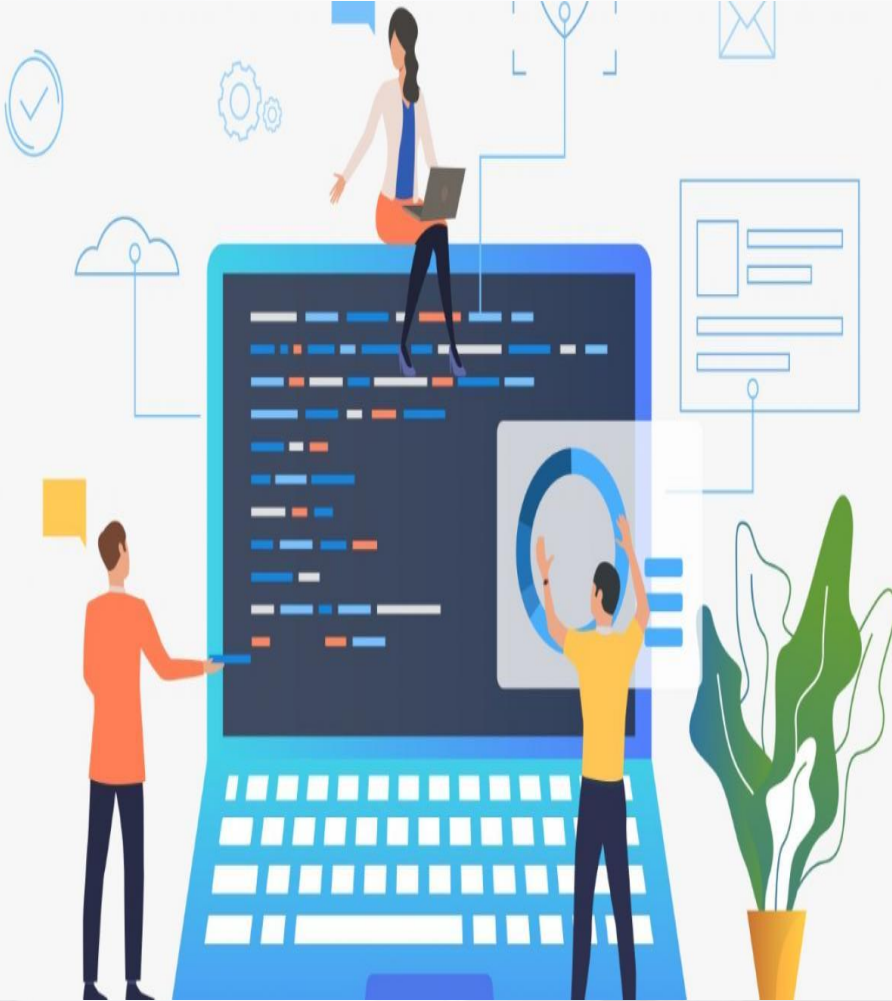


ويساعد هذا الجدول “صحفي البيانات” في فهم البيانات التي بحوزته، ويقلل الوقت والمجهود الذي يبذله خلال هذه المرحلة، وهو ما يمكن أن يجعل قطعة بيانات صحفية متوسطة أو جيدة إلى قطعة صحافة بيانات مميزة، ويتضمن هذا الجدول تاريخ وتفاصيل البيانات التي تم جمعها، مثل: اسم قاعدة البيانات التي تم تجميع البيانات منها أو الجهة التي تمتلك هذه البيانات على سبيل المثال: قاعدة بيانات الأمم المتحدة أو البنك الدولي إلخ .

3. روابط الوصول لهذه البيانات على شبكة الإنترنت على سبيل المثال (www.UN.org....)، وذلك حتى لا ينسى الصحفي من أين حصل على هذه البيانات.

ثانيًا: أهمية خطوة تنظيف البيانات

4. مكان الاحتفاظ بالبيانات على جهاز كمبيوتر الصفي لسرعة الوصول إليها ومن أفضل حفظها في ملف خاص.
5. اسماء الأشخاص الذين عملوا على جمع البيانات.
6. تاريخ جمع البيانات والفترة الزمنية التي تتحدث عنها البيانات.
7. إعادة تهيئة البيانات وإعادة تشكيلها مرة أخرى من خلال حذف الأجزاء غير المهمة من البيانات والتي لن تؤثر على القصة الصحفية، والتعرف على نوع البيانات سواء رقمية أو وصفية، حتى يسهل التعامل معها.



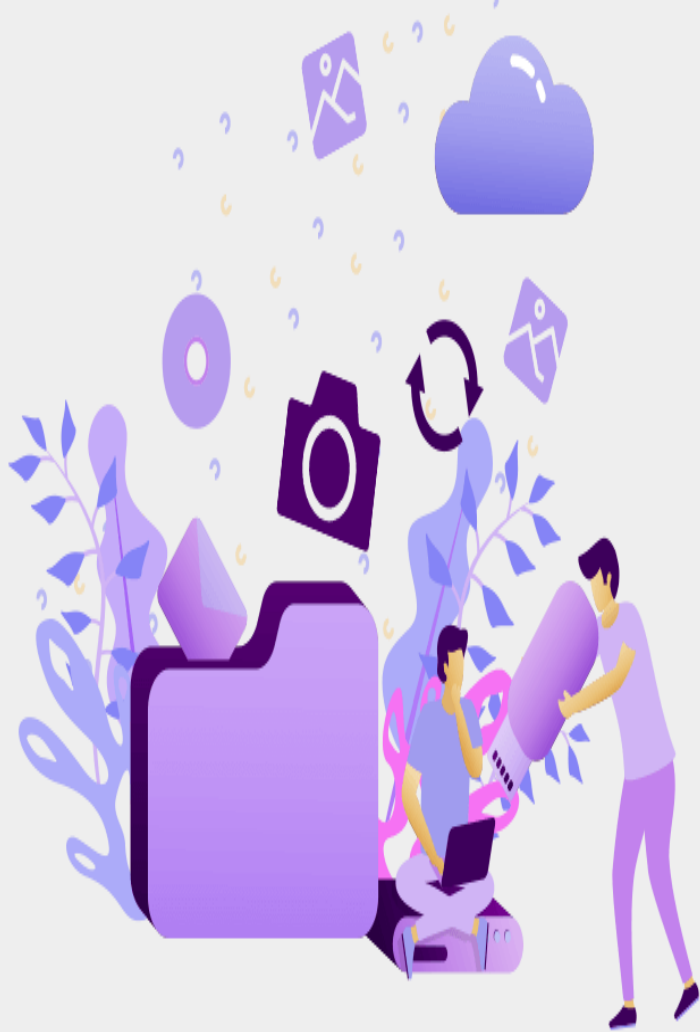
ثانيًا: أهمية خطوة تنظيف البيانات



8 . رغم أن مرحلة تنظيف البيانات تعد عملية شاقة وتتطلب الكثير من الوقت، إلا أنها تمثل فرصة جيدة للصحفيين لفهم البيانات التي بحوزتهم، وتحديد المتغيرات والأشخاص والأدوار الفاعلة والزوايا التي يجب أن تروى منها القصة الصحفية، وتسجيل جميع الملاحظات التي يراها الصحفي، وتدوين أي تغيرات جديدة قد تظهر أمامه عند تنظيف البيانات.

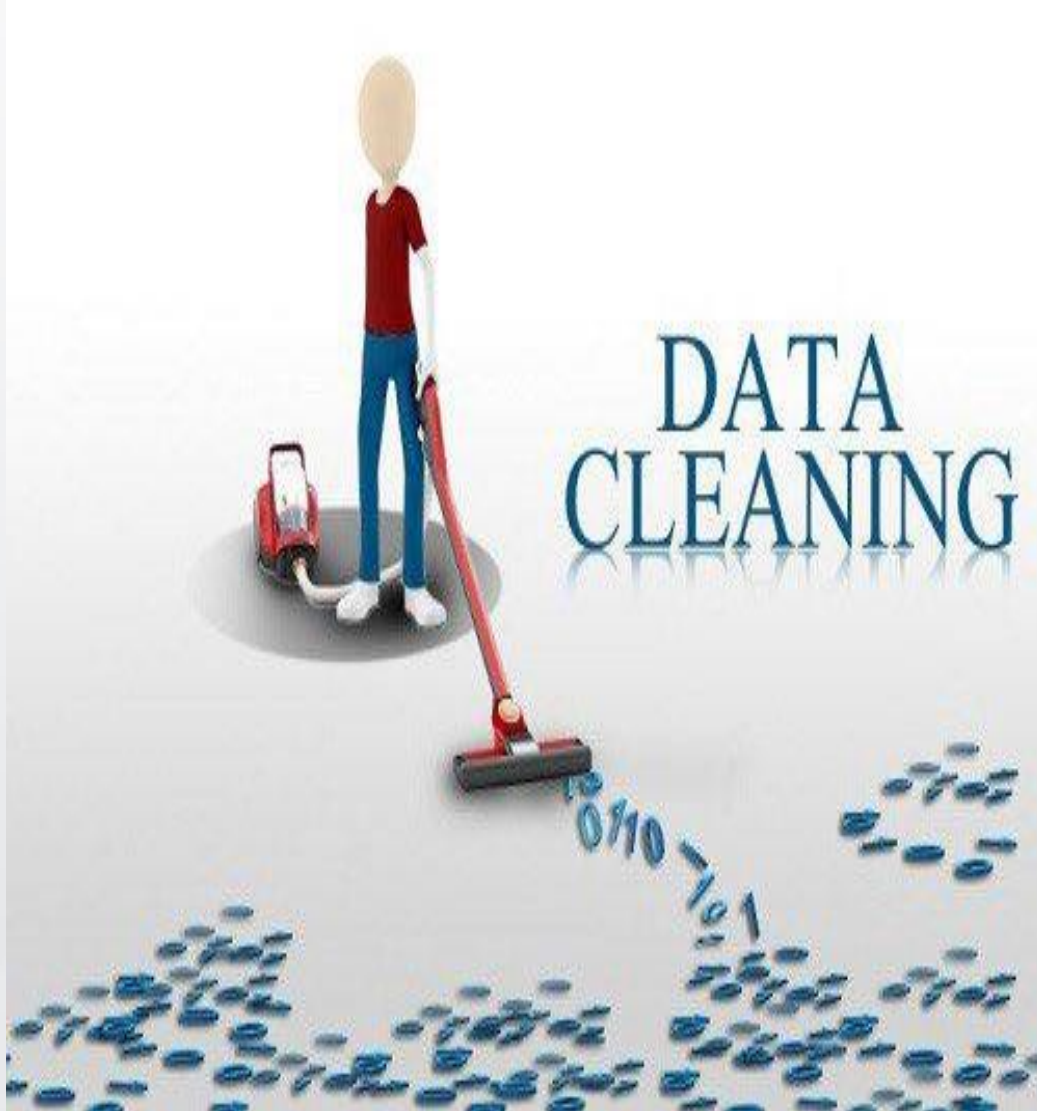
9 . التعرف على البيانات المفقودة كالأرقام والأماكن والأسماء وغيرها، ومحاولة استكمال الأجزاء والقيم المفقودة منها، وتوحيد المتغيرات الرقمية، والتحقق من القيم المتطرفة، التي توجد بسبب خطأ في القياس أو قيام الصحفي بإدخال البيانات بشكل خاطئ إلى برنامج Excel حتى لا تتسبب البيانات المفقودة في مشكلة أثناء التحليل.

ثانيًا: أهمية خطوة تنظيف البيانات



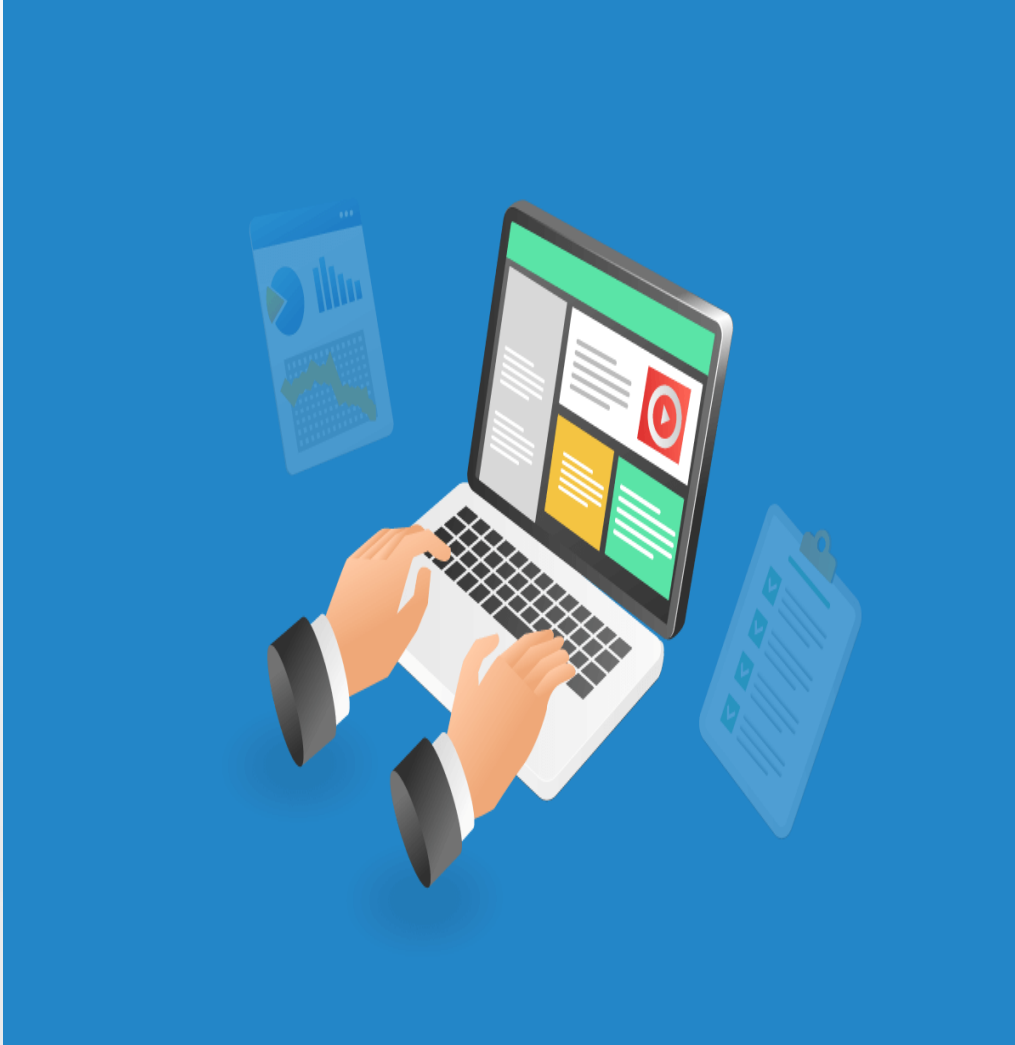
10 . ضرورة مراعاة أخلاقيات تنظيف البيانات، لاسيما عند وجود بيانات مفقودة يمكن أن تؤثر على نتائج التحليل، كما يحدث على سبيل المثال: في بعض الاستبيانات أن عدد من الناس لا تجيب على بعض الأسئلة الموجودة في الاستبيان، كالأسئلة المتعلقة بالدخل والسن وغيرها، وهنا يجب على الصحفي أن يذكر أن ٢٠% من البيانات مفقودة وغير كاملة، حتى تتضح النتائج أمام القراء بشكل صحيح وواقعي وغير متحيز، أو أن يتوجه إلى خبراء الإحصاء لمساعدته في ما يمكن فعله مع البيانات المفقودة والقيم المتطرفة، خلال مرحلة تنظيف البيانات، حتى لا تتسبب في خلق تحيزا أو معلومات غير صحيحة في القصة الصحفية.

ثالثًا: خطوات تنظيف البيانات



- إن لكل مجموعة بيانات مشاكلها، لذلك واحدة من أولى المهام هي التعامل مع هذه المشكلات فيما يعرف بعملية تنظيف البيانات، والتي تعتبر خطوة حاسمة في مجال صحافة البيانات؛ حيث تهدف إلى تصحيح الأخطاء وإزالتها والتشوهات في مجموعة البيانات، يشمل ذلك إصلاح البيانات التي قد تكون غير صحيحة وغير متناسقة وتنظيم البيانات بشكل صحيح لتتبع تنسيق محدد، والتعامل مع البيانات المكررة أو الناقصة، و دمج البيانات من مصادر متعددة في مجموعات بيانات واحدة، وأيضًا قد تظهر البيانات بتسميات غير صحيحة أو متكررة بشكل مرئي، وبالتالي فإن إجراء تنظيف البيانات تأتي لضمان أن التقارير التي تعتمد على هذه البيانات تظل موثوقة ودقيقة.

ثالثًا: خطوات تنظيف البيانات



- وليس هناك منهجية ثابتة ومحددة تنطبق على جميع الحالات في عملية تنظيف البيانات، بل يجب تكييف الإجراءات والمخططات لكل مجموعة بيانات بناءً على خصائصها الفريدة، ومع ذلك فإن إنشاء قالب أو إطار عمل لعملية تنظيف البيانات يعد أمرًا ضروريًا لضمان تنفيذ العملية بشكل صحيح ومنهجي في كل مرة، ويمكن اتباع هذه الخطوات الأساسية لوضع إطار عام لتنظيف البيانات:

ثالثًا: خطوات تنظيف البيانات

DATA CLEANSING TOOLS



- تنظيف البيانات باستخدام Microsoft Excel:

- برنامج Microsoft Excel هو برنامج جداول بيانات يتيح إنشاء وتنسيق وتحليل وتخزين البيانات بشكل فعال.

الخطوة الأولى فهم البيانات:

قبل البدء في تنظيف البيانات، من الضروري فهم طبيعة البيانات، هيكلها والمشاكل المحتملة فيها.

ثالثًا: خطوات تنظيف البيانات

الخطوة الثانية: إزالة البيانات المكررة أو غير المهمة:

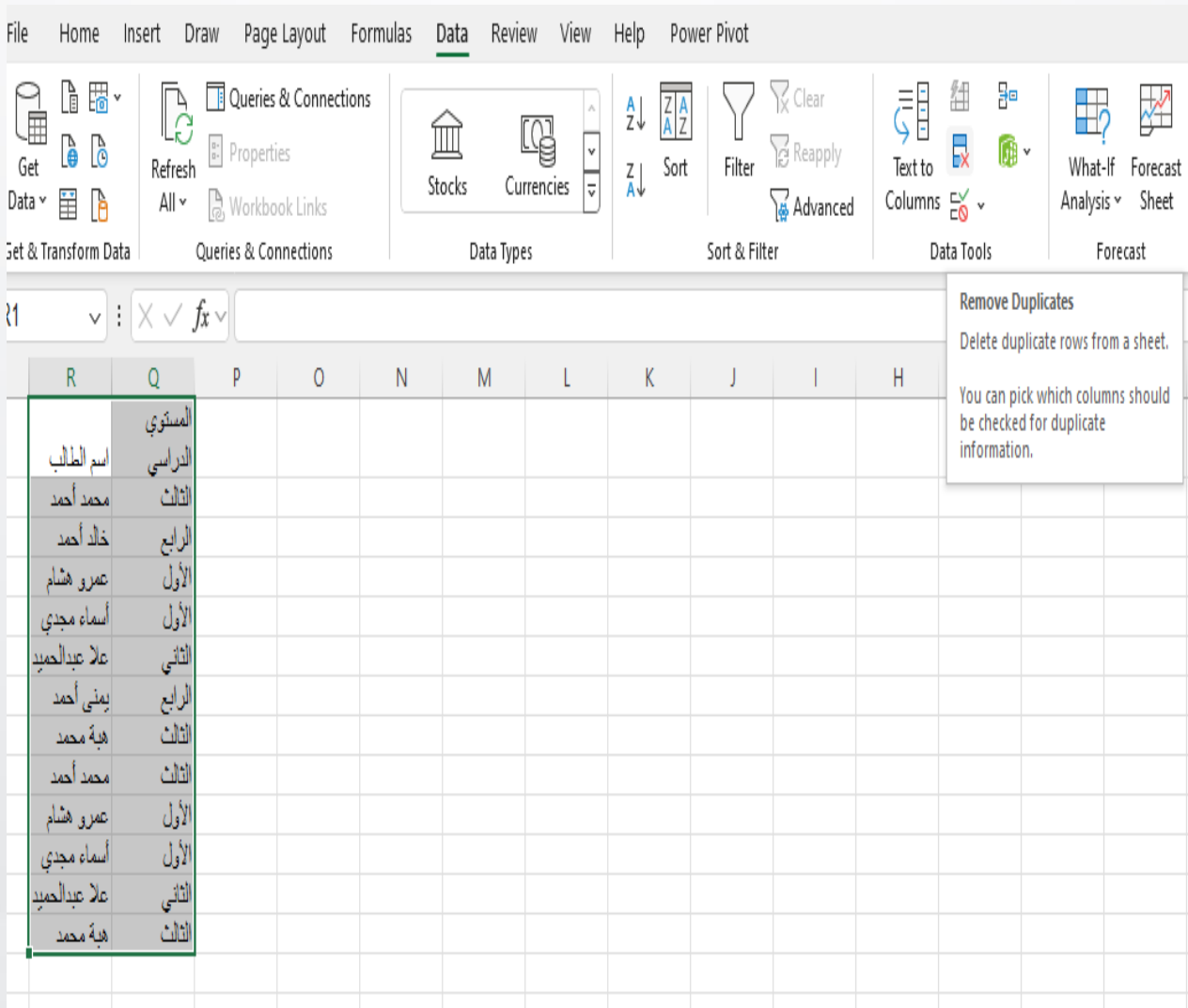
تتضمن المرحلة الأولى من عملية تنقية البيانات إزالة البيانات غير المرغوب فيها من مجموعة البيانات الخاصة بنا بما في ذلك البيانات المكررة أو تلك التي ليست لها صلة، ويحدث وجود البيانات المكررة بشكل شائع أثناء جمع البيانات عندما تجمع بيانات من مصادر متعددة أو تقوم بجمع البيانات من جمهور خلال استطلاعات الرأي، فمن الممكن أن يحدث التكرار، لابد هنا من إجراء عملية إزالة التكرار بدقة لتصحيح هذه التناقضات، إضافة إلى ذلك يمكن إزالة البيانات غير المهمة التي لا تتناسب مع المشكلة المحددة التي نحاول تحليلها.

S	T	U	V
		العمر	الاسم
		19	احمد
		30	عمرو
		25	خالد
		19	أحمد
		33	محمد
		20	زياد
		30	عمرو

ثالثًا: خطوات تنظيف البيانات

أ. إزالة القيم المكررة:

انقر على تبويب بيانات – أدوات البيانات – إزالة التكرارات.



File Home Insert Draw Page Layout Formulas Data Review View Help Power Pivot

Get Data Refresh Properties Workbook Links Queries & Connections Data Types Sort Filter Text to Columns What-If Analysis Forecast Sheet

Remove Duplicates
Delete duplicate rows from a sheet.
You can pick which columns should be checked for duplicate information.

المستوى	اسم الطالب
الثالث	محمد أحمد
الرابع	خالد أحمد
الأول	عمرو هشام
الأول	أسماء مجدي
الثاني	علا عبدالحميد
الرابع	يمنى أحمد
الثالث	هبة محمد
الثالث	محمد أحمد
الأول	عمرو هشام
الأول	أسماء مجدي
الثاني	علا عبدالحميد
الثالث	هبة محمد

ثالثًا: خطوات تنظيف البيانات

File Home Insert Draw Page Layout Formulas **Data** Review View Help Power Pivot

Get & Transform Data Queries & Connections Data Types Sort & Filter Data Tools

Q2

R	Q	P	O	N	M
اسم الطالب	المستوي الدراسي				
محمد أحمد	الثالث				
خالد أحمد	الرابع				
عمرو هشام	الأول				
أسماء مجدي	الأول				
علا عبد الحميد	الثاني				
يمنى أحمد	الرابع				
هبة محمد	الثالث				
محمد أحمد	الثالث				
عمرو هشام	الأول				
أسماء مجدي	الأول				
علا عبد الحميد	الثاني				
هبة محمد	الثالث				

Remove Duplicates

To delete duplicate values, select one or more columns that contain duplicates.

☒ Select All ☐ Unselect All ☒ My data has headers

Columns

- ☒ المستوي الدراسي
- ☒ اسم الطالب

OK Cancel

ثالثًا: خطوات تنظيف البيانات

الخطوة الثالثة: إصلاح الأخطاء في مجموعة البيانات:

- في بعض الأحيان، تحتوي القيم النصية على أحرف مسافة بادئة أو زائدة أو متعددة مضمنة، قد تتسبب هذه الأحرف أحيانًا في نتائج خاطئة عند الفرز أو التصفية أو البحث، في هذه الحالة تستخدم دالة (TRIM).
- نقوم بإدراج عمود جديد بجانب العمود الذي يحتوي على فراغات.
- في هذه الخلية الأولى من العمود، قم بكتابة الدالة TRIM وحدد الخلية التي تحتوي على النص الأصلي الذي تريد تنظيفه.

COUNT : X ✓ fx =TRIM(Q1)						
	R	Q	P	O	N	M
1		عمرو هشام	=TRIM(Q1)			
2		أسماء مجدي				
3		علا عبدالحميد				
4		يمنى أحمد				
5		هبة محمد				
6		محمد أحمد				
7		عمرو هشام				

ثالثًا: خطوات تنظيف البيانات

[illegible]

- قبل استخدام أي دالة في الإكسل نضيف علامة " = " قبل كتابة اسم الدالة ثم الأقواس ().
- نضغط على مفتاح Enter بعد كتابة الصيغة، ستقوم الصيغة بتنظيف النص في الخلية المحددة وإظهار النتيجة في الخلية الجديدة التي قمنا بتحديددها.
- يمكننا نسخ الصيغة التي أنشأناها في الخلية الجديدة ولصقها في باقي الصفوف في العمود الجديد، وسيتم تنفيذ الدالة TRIM على النصوص في الصفوف الأخرى أيضًا.

ثالثًا: خطوات تنظيف البيانات

The screenshot displays the Microsoft Excel interface. The ribbon at the top includes File, Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, Help, and Power Pivot. The Home ribbon is active, showing options for Clipboard, Font, Alignment, Number, Styles, Cells, and Editing. Below the ribbon, the formula bar shows the active cell as Q1 with the text 'countries'. The worksheet contains a table with the following data:

Names	countries	gender
nahla	Egypt	f
Amr	Syria	m
Ahmed	Syrian	m
Mohamed	Egypt	m

The 'Find and Replace' dialog box is open, showing the 'Find' tab. The 'Find what' field contains 'syrian' and the 'Replace with' field contains 'syria'. The 'Find Next' button is highlighted.

أ- تصحيح الأخطاء الإملائية باستخدام البحث والاستبدال:

- ننتقل إلى أول عمود النص وننقر على عنوان العمود لتحديده
- ننتقل إلى شريط الأدوات الرئيسي ونختار " تصحيح النص " أو " بحث واستبدال "

ثالثًا: خطوات تنظيف البيانات

ب. توحيد شكل الحروف:

□ إذا كنا نرغب في توحيد حروف النص إلى حالة كبيرة أو صغيرة يمكننا استخدام الدالة Upper أو Lower.

□ في عمود جديد أو نفس العمود المنظف ، نكتب الصيغة المناسبة.

J	I	H	G	F
		Name		
		ahmed ali	=UPPER(H4)	
		digital		
		media		
		sustainabl		
		e		
		developm		
		en		

ثالثًا: خطوات تنظيف البيانات

ج. دمج وتقسيم الأعمدة وإظهار الأعمدة المخفية:

أولاً: دمج الأعمدة :

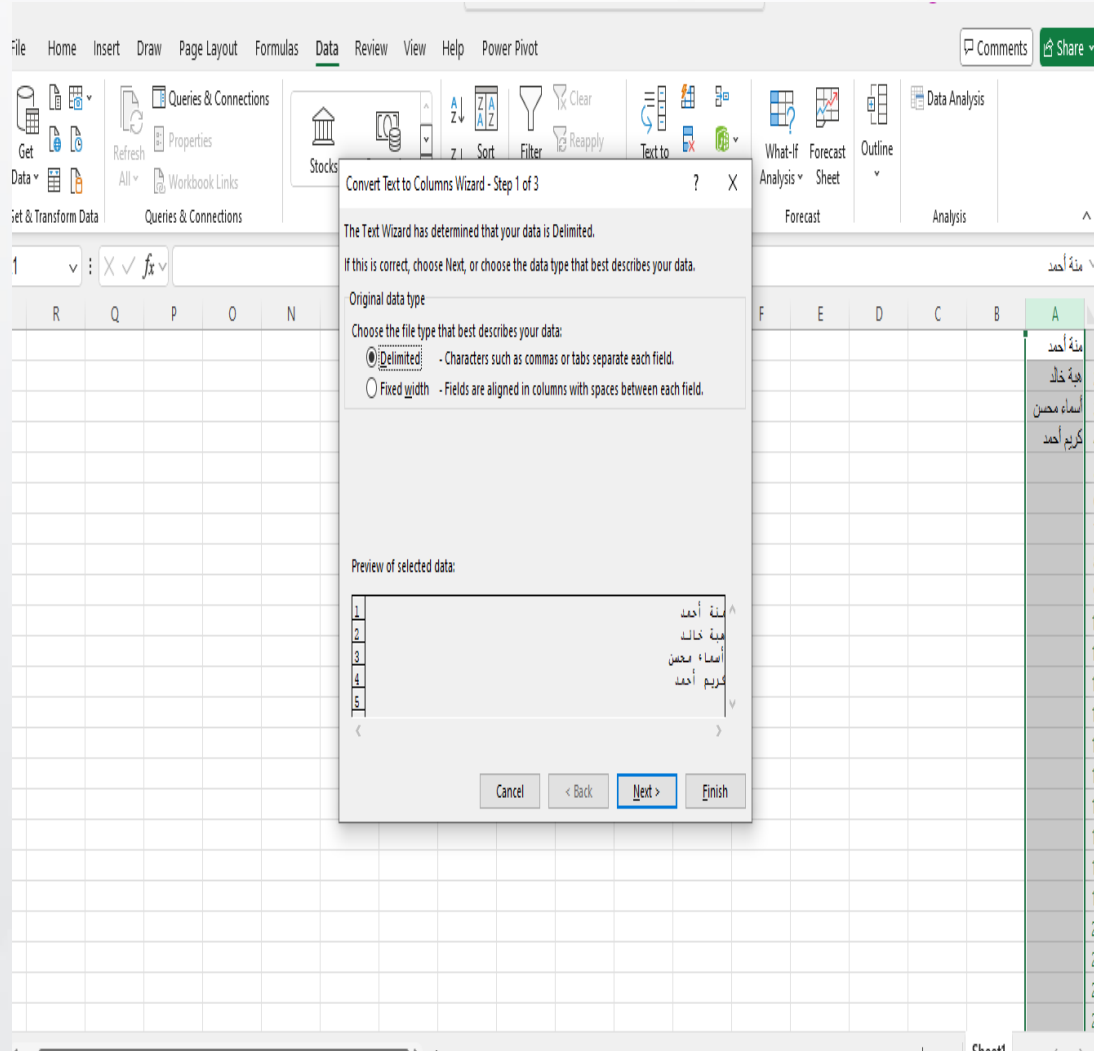
عند دمج عمودين نستخدم دالة "CONCAT" أحيانا نحتاج الجمع بين عمودين أو أكثر في عمود واحد لسهولة التحليل مثل وجود الاسم الأول والاسم الأخير في عمودين منفصلين نستخدم دالة "CONCAT" لجمع عموديين نقوم بكتابة الدالة ثم تحديد نطاق العمودين معًا.

COUNT	⌵	:	✗ ✓ f_x	⌵	=CONCAT(P2:Q2)	
	R	Q	P	O	N	M
		هشام	عمرو	عمرو هشام		
		مجدي	أسماء	=CONCAT(P2:Q2)		
		عبد الحميد	علا			
		أحمد	يمنى			
		محمد	هبة			
		أحمد	محمد			
		هشام	عمرو			

ثالثًا: خطوات تنظيف البيانات

ثانيًا: تقسيم الأعمدة:

- تقسيم البيانات في عمود واحد إلى أعمدة متعددة (مثل تقسيم الأسماء الأول والأخير):
- حدد العمود الذي يحتوي على البيانات التي تريد تقسيمها.
- من الشريط العلوي، نذهب إلى تبويب Data البيانات
- اختر Text to Columns تحويل النص إلى أعمدة.
- في مربع الحوار الذي يظهر:
- اختر Delimited محدد (إذا كانت البيانات تفصلها علامات معينة مثل مسافة أو فاصلة).



ثالثًا: خطوات تنظيف البيانات

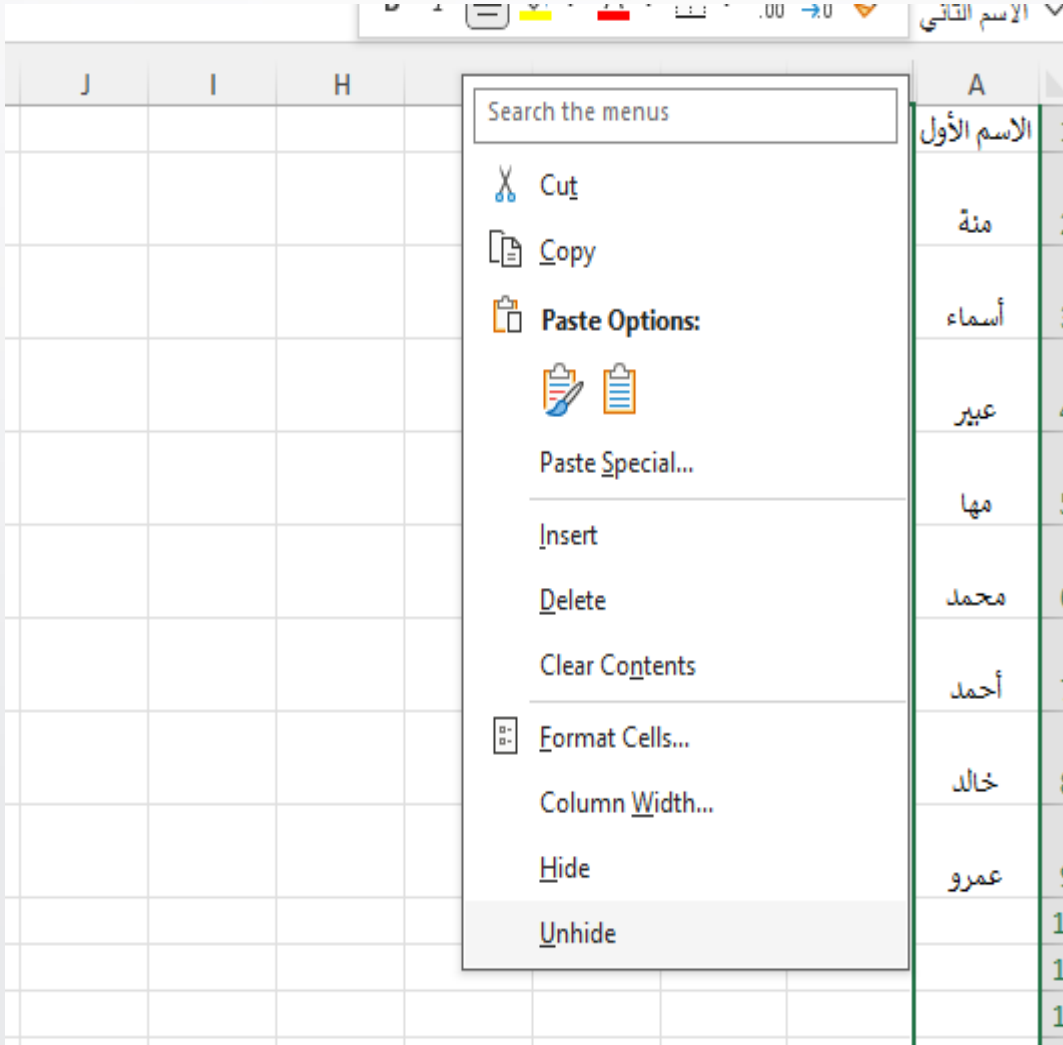
D	C	B	A
		أحمد	منة
		خالد	هبة
		محسن	أسماء
		أحمد	كريم

- اختر (Fixed Width عرض ثابت (إذا كنت تريد تقسيم النص بناءً على عدد ثابت من الأحرف.
- اضغط على (Next التالي: (إذا اخترت Delimited ، اختر نوع الفاصل (مسافة، فاصلة، إلخ).
- إذا اخترت Fixed Width ، ضع المؤشرات حيث تريد أن يتم التقسيم.
- اضغط على (Finishإنهاء (وستقسم البيانات إلى أعمدة متعددة).

ثالثًا: خطوات تنظيف البيانات

ثالثًا إظهار الأعمدة الخفية:

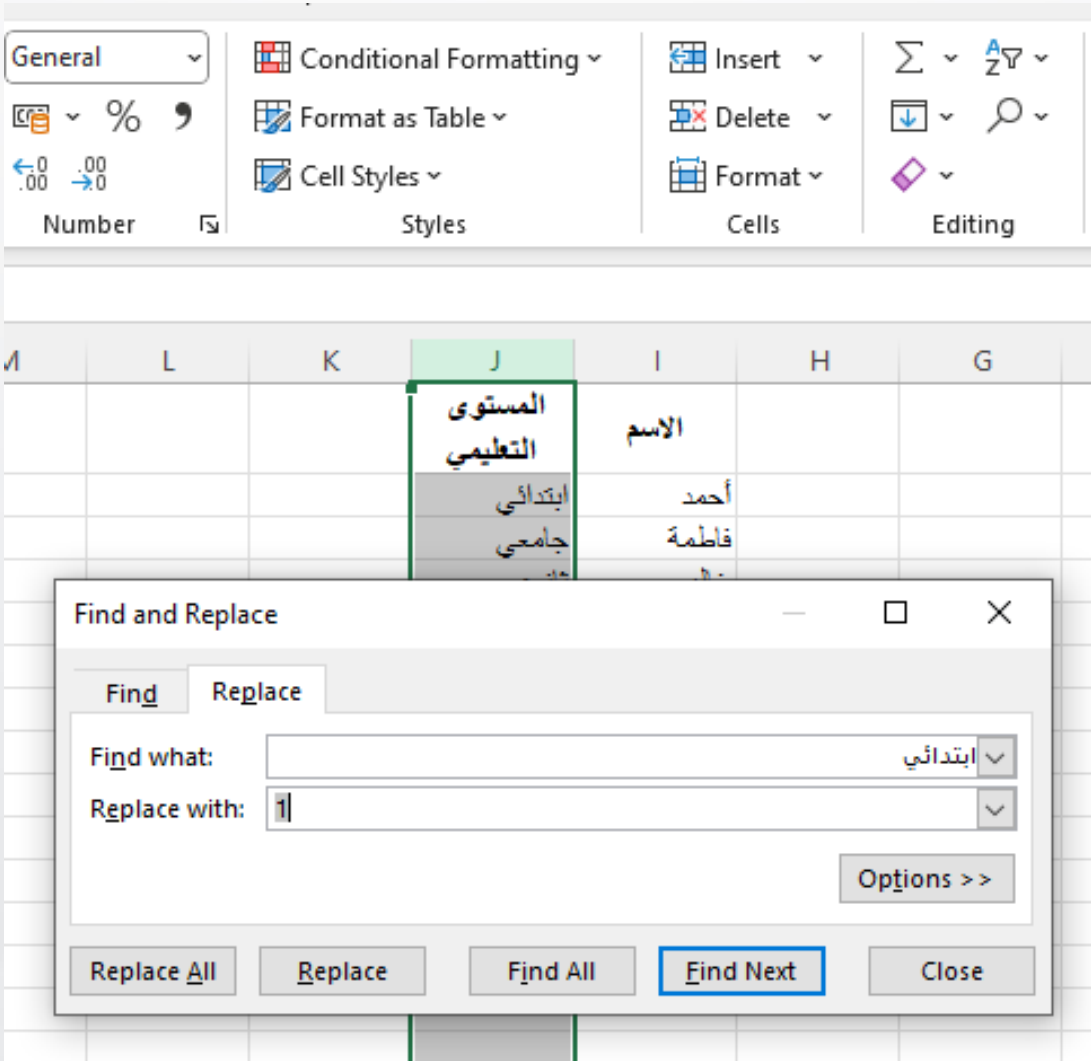
خاصية إظهار الأعمدة الخفية في مايكروسوفت إكسل تتيح للمستخدمين استعادة الأعمدة التي تم إخفاؤها مسبقًا بحيث تصبح مرئية مرة أخرى في ورقة العمل، يتم ذلك عن طريق تحديد الأعمدة المجاورة للعمود المخفي ثم اختيار "إظهار" من قائمة التنسيق أو النقر بزر الماوس الأيمن.



ثالثًا: خطوات تنظيف البيانات

د. تحويل البيانات:

أثناء عملية تحويل البيانات من الضروري تغيير البيانات النصية لبيانات رقمية، في بعض الأحيان لسهولة تحليل البيانات تحتاج إلى المدخلات الرقمية، للقيام بذلك يمكن مطابقة قيم أرقام معينة مع البيانات النصية في كل عمود، فعلى سبيل المثال: لدينا عمود يحتوي على بيانات نصية توصف المستوى التعليمي للجمهور الذي ندرسه، ولأغراض التحليل نحتاج معرفة متوسط المستوى التعليمي لكل فرد في هذه الحالة نحتاج أولاً لتحويل البيانات لتنسيق الأرقام.



ثالثًا: خطوات تنظيف البيانات

K	J	I	H	G
		المستوى التعليمي	الاسم	
		1	أحمد	
		4	فاطمة	
		3	خالد	
		5	لبنى	
		2	يوسف	
		4	منة	
		3	هبة	
		5	عمرو	
		6	محمد	

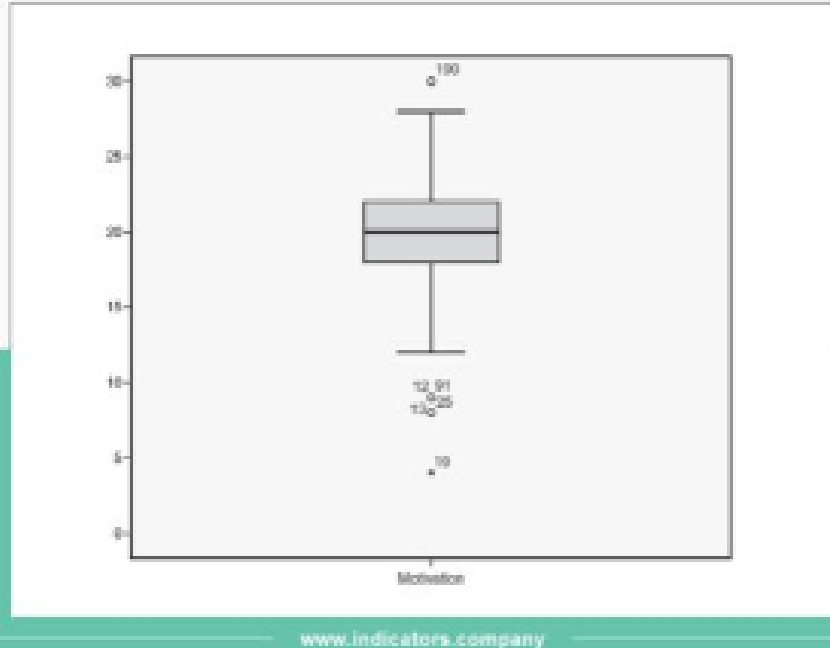
حدد البيانات في العمود الذي نريد تغيير بياناته رقميا من الشريط الرئيسي اختر البحث والاستبدال، ثم استبدل المرحلة الابتدائية برقم "1" وانقر على "استبدال الكل".

نكرر هذا الإجراء في جميع مراحل التعليم مع إعطاء كل مرحلة رقما ثم ننقر على تم.

ثالثاً: خطوات تنظيف البيانات

الخطوة الرابعة: تصفية القيم الشاذة:

معالجة القيم الشاذة



غالباً ما تكون هناك بيانات فردية تبدو غير مناسبة للبيانات التي تقوم بتحليلها، إذا كان لديك سبب مشروع لإزالة قيمة شاذة، مثل أخطاء في إدخال البيانات، فإن ذلك سيساعد في تحسين أداء البيانات التي تعمل عليها، ومع ذلك في بعض الأحيان يمكن أن يكون ظهور قيمة شاذة دليلاً على ظاهرة تقوم بالعمل عليها، عليك أن تتذكر أن وجود قيمة شاذة لا يعني بالضرورة أنها غير صحيحة، فإن هذه الخطوة ضرورية لتحديد صحة هذا الرقم، إذا ثبت أن القيمة الشاذة غير مهمة للتحليل أو هي خطأ يمكن إزالتها.

ثالثًا: خطوات تنظيف البيانات

- تصفية القيم الشاذة (Outliers) في مايكروسوفت إكسيل، يمكن استخدام بعض الأدوات البسيطة المتوفرة في البرنامج مثل الفرز والتصفية، أو استخدام المعادلات الإحصائية لتحديد القيم الشاذة، إليك كيفية القيام بذلك:

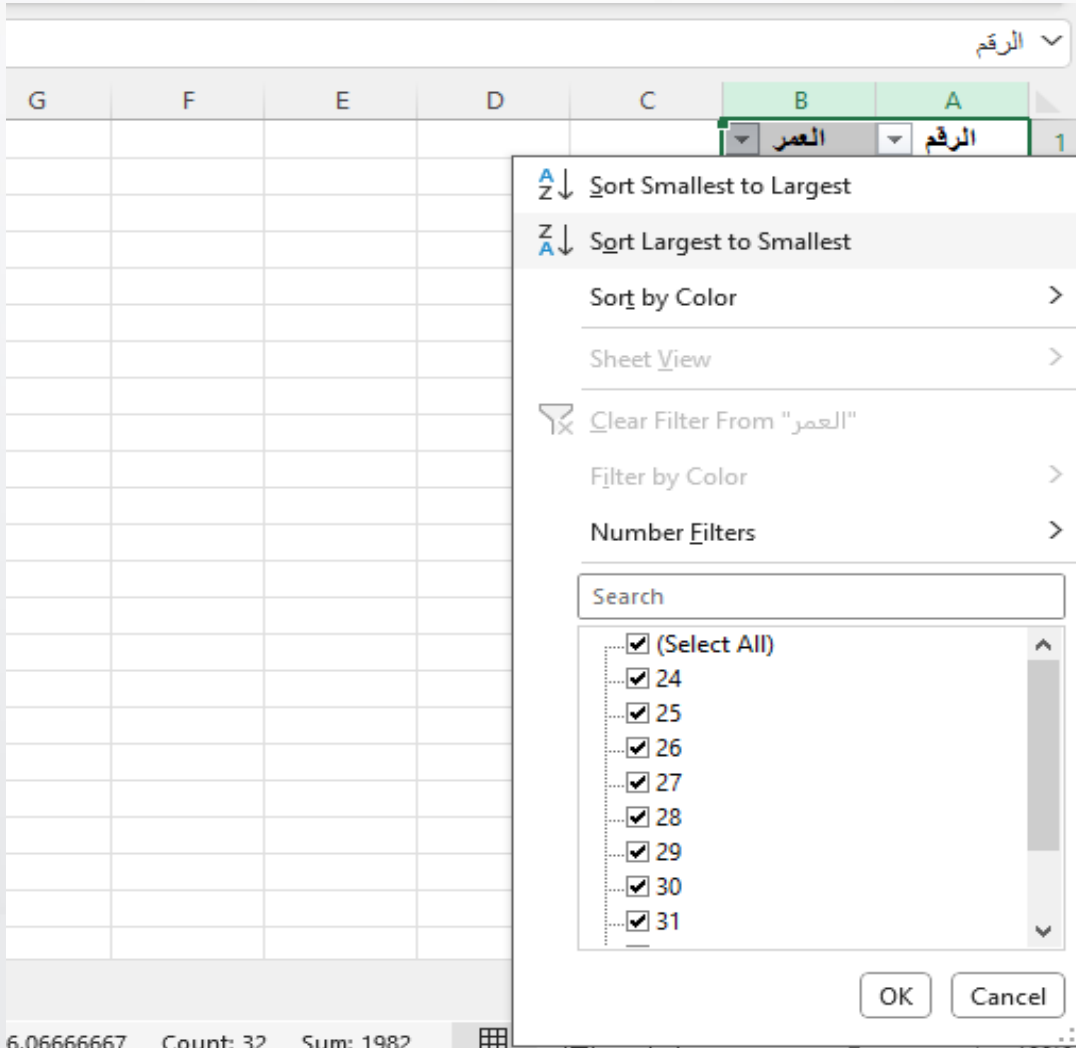
- حدد البيانات: اختر العمود الذي يحتوي على القيم التي تريد تحليلها.
- اذهب إلى تبويب Data البيانات (في الشريط العلوي).
- اضغط على Sort فرز أو Filter تصفية).

الرقم	الرقم
1	1
2	25
3	30
4	27
5	29
6	28
7	26
8	32
9	24
10	500
11	31
12	29
13	26
14	1000
15	27
16	28
17	

ثالثاً: خطوات تنظيف البيانات

■ الفرز : يمكنك ترتيب البيانات تصاعدياً أو تنازلياً لرؤية القيم الشاذة التي تقع في الأطراف.

■ التصفية : بعد الضغط على زر **Filter** تصفية، ستتم إضافة سهم في أعلى العمود، يمكنك استخدام هذا السهم لتصفية القيم المرتفعة أو المنخفضة جداً.



ثالثًا: خطوات تنظيف البيانات

الخطوة الخامسة التعامل مع البيانات المفقودة:

لا يمكن تجاهل البيانات المفقودة لأنه قد تتسبب في نتائج مضللة، وهناك عدة طرق للتعامل مع هذه البيانات المفقودة، والتي يجب تناولها بعناية نظرا لعدم وجود طريقة مثلى، ولكن يمكن النظر في الخيارات التالية:

الخيار الأول: يمكن حذف البيانات التي تحتوي على قيم مفقودة ولكن يجب عليك أن تكون حذرًا؛ لأن ذلك قد يؤدي إلى فقدان جزء من المعلومات، لذا يجب التفكير جيدا قبل القرار بالحذف.

الخيار الثاني: يمكن إدخال قيم مفقودة استنادا إلى البيانات الأخرى لو كانت البيانات نوعية.



ثالثًا: خطوات تنظيف البيانات

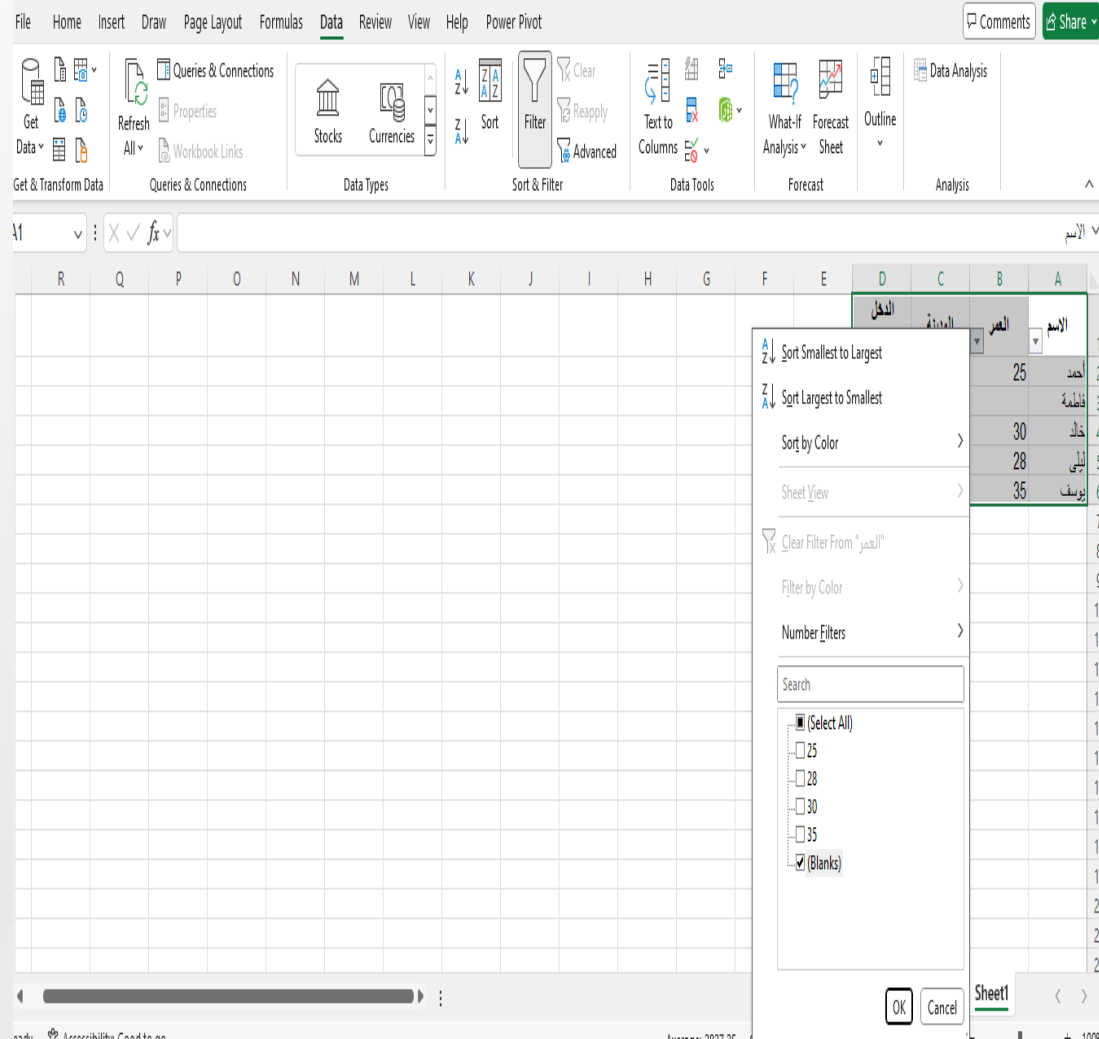
S	T	U	V
		العمر	الاسم
		19	احمد
		30	عمرو
		25	خالد
		19	أحمد
		0	محمد
		20	زياد
		30	عمرو

الخيار الثالث: يمكن استبدال القيمة الرقمية الفارغة ب (0) أو لا توجد قيمة.

- ملئ البيانات المفقودة جزء مهم من عملية تنظيف البيانات، وفي بعض الأحيان نجد خلايا فارغة في عدد من الأعمدة إذا كانت الأعمدة غير رقمية من الممكن إضافة كلمة "بدون قيمة" حتى لا يؤثر ذلك على نتائج التحليل لاحقًا.

ثالثًا: خطوات تنظيف البيانات

- من تبويب بيانات نختار تصفية، ثم يتم تمييز كل الخلايا، هناك عوامل تصفية أعلى كل عمود به بيانات، حدد العمود الذي به فراغات، قم بتصفية العمود حسب الفراغات، تظهر لك كل صفوف الورقة الفارغة، ولإصلاح ذلك استبدال القيم الفارغة بكلمة " بدون قيمة" أو "none- given" في الخلفية الفارغة، ثم ننقر مزدوجا فوق مؤشر التعبئة لملئ كافة الخلايا الفارغة بالقيمة المدخلة.



ثالثًا: خطوات تنظيف البيانات

F	E	D	C	B	A
		الدخل الشهري	المدينة	العمر	الاسم
		5000	الرياض	25	أحمد
		6000	جدة	none-giver	فاطمة
		4500	none-giver	30	خالد
		none-giver	مكة	28	لبنى
		7000	الدمام	35	يوسف

- نوقف عامل التصفية لنجد أن القيمة الجديدة قد حلت محل كافة الفراغات بالعمود المحدد، وبالمثل بإمكاننا إدخال قيم المتوسط للعمود لو كان العمود كمياً، أو الصفر ليحل محل الخلايا الفارغة بنفس الطريقة.

رابعًا: التحقق ومراجعة البيانات

في نهاية عملية تنظيف البيانات يجب أن نكون قادرين
على الإجابة على هذه الأسئلة كجزء من التحقق
الأساسي:

- هل البيانات منطقية؟
- هل تتبع البيانات القواعد المناسبة لمجالها؟
- هل تثبت أو تنفي افتراضك، أو تسلط الضوء على أي نقاط غامضة؟
- هل يمكن العثور على اتجاهات (Trends) في البيانات تساعد في تحسين الافتراضية التي يتم العمل عليها؟
- إذا لم يكن الأمر كذلك هل هذا بسبب مشكلة في جودة البيانات؟



رابعًا: التحقق ومراجعة البيانات



- يمكن أن تتسبب البيانات غير الصحيحة، وغير المتناسقة أو المفقودة إلى استنتاجات خاطئة، وبالتالي تفقد الصحيفة أو المؤسسة الإعلامية مصداقيتها، لذلك من المهم إنشاء منهج خاص للتعامل مع البيانات في المؤسسات الإعلامية، ولتحقيق ذلك يجب على الصحفيين و محرري البيانات أن يوثقوا الأدوات و الممارسات التي يمكن أن يستخدموها لبناء المنهج.

رابعًا: التحقق ومراجعة البيانات

□ التحقق من تنظيف البيانات:

بغض النظر عن الأداة المستخدمة في تنظيف البيانات، فلا يوجد نهج يتعلق بالتحقق من تنظيف البيانات، ولكل مشروع تنظيمة الخاص، وتؤدي متطلبات البيانات لقائمة فريدة بالأشياء التي لابد من التحقق منها، ولكن ضع في اعتبارك هذه القائمة عند التحقق من تنظيف البيانات :

التأكد من تحديد المشكلات الأكثر شيوعا وتصحيحها
مثل:

1. البيانات الفارغة هل قمت بالبحث عن القيم الفارغة.
2. الكلمات التي تحتوي على أخطاء إملائية.



رابعًا: التحقق ومراجعة البيانات



3. التعرف على مصادر الخطأ هل استخدمت الدوال الصحيحة للتعرف على مصدر الخطأ.
4. التحقق من الأرقام الخاطئة.
5. التحقق من الأحرف الزائدة والمسافات.
6. التحقق من أنواع البيانات غير المتطابقة مثل: التواريخ والبيانات الرقمية والنصوص.
7. التحقق أن التواريخ متوافقة مع مجموعة البيانات لديه (أي لها تنسيق موحد)
8. مراجعة تسميات الأعمدة.

ملخص المحاضرة

- أولاً: مفهوم عملية تنظيف البيانات.
- ثانياً: أهمية خطوة تنظيف البيانات.
- ثالثاً: خطوات تنظيف البيانات.
- رابعاً: التحقق ومراجعة البيانات.

شكرًا لاستماعكم