

فهم ومعاينة واستكشاف وتحليل البيانات وتلخيصها

الأهداف التعليمية:

يهدف هذا الجزء إلى تعليم المتدربين كيفية التعاطي مع البيانات وتبسيطها من أجل استخدامها إعلامياً.

نقاط المحتوى:

- اختيار البيانات اللازمة من بين مصادر عدة.
- قراءة البيانات وفهمها بعد معاينتها واستكشافها.
- تبسيط البيانات وترتيبها.
- تحليل البيانات

مدة التدريب: ٦ ساعات

أولاً: اختيار البيانات اللازمة من بين مصادر عدة (ساعة و٣٠ دقيقة)

هناك الكثير من الجهات التي توفر بيانات حول مختلف القضايا. كيف نعرف هذه الجهات؟ وكيف نتأكد من أنها الجهة ذات الاختصاص؟

لا بد أن نعد قائمة بأسماء الجهات ذات الاهتمام في حقول مختلفة. ولا بد أن تشمل القائمة روابط مواقعها الإلكترونية، وأرقام الهواتف إن احتاج الأمر، والأشخاص المسؤولين عنها.

عندما تود الكتابة عن موضوع معين، أول ما يجب أن تفعله هو أن تطرح أسئلة حول المعلومات التي تنقصك. وابحث عن أجوبة لتلك الأسئلة وعن أجوبة أخرى تحيط بالموضوع، أي ليس لها علاقة مباشرة بالموضوع.

مثال: إذا أردت أن تعالج موضوع البطالة في فلسطين في عام ٢٠١٥، لا بد أن تبحث بشكل أساسي عن مستوى البطالة في إسرائيل أيضاً، والدول العربية. كما لا بد من البحث عن مستوى النمو الاقتصادي في ذلك العام، وعن السياسات الداخلية، ومخرجات التعليم والتدريب لدى السلطة وإسرائيل.

قد تكون المعلومة بحاجة إلى بيانات تُثبتها أو تدحضها، وقد تكون البيانات بحاجة إلى معلومة تفسرها، لذلك، من المهم هنا أن تحصل على كل المتغيرات والأرقام والمستندات التي يمكن أن تحتاجها لاحقاً بدل الاكتفاء بالأسئلة الآنية. فالمصادر العديدة توفر زوايا جديدة وأفكاراً جديدة وفهماً أعمق.

ضع لائحة بالبيانات التي يوجد فيها تصريحات موثقة.

قد يُصرح أحد المسؤولين أن السوق المحلي شهد شيئاً من التحسن. لا بد أن تعرف هنا مدى صحة هذه المعلومة. إذا اطلعت على النتائج الأساسية لمسح القوى العاملة للربع الأول من عام ٢٠١٥، ستجد معلومة تدعم هذا التصريح. إذ تشير نتائج المسح إلى أن عدد العاملين في السوق المحلي ما بين الربع الثالث والرابع لعام ٢٠١٥ ارتفع من ٨٣٠٩٠٠ إلى ٨٦٤٥٠٠.

بعد مدة قليلة من البحث، ستجد نفسك أمام كمّ هائل من البيانات.

ثانياً: استكشاف البيانات ومعاينتها وفهمها وقراءتها (ساعة و٢٠ دقيقة)

بعد أن تجمع هذا الكمّ من البيانات، ستجد نفسك محتاراً. كيف ستنظر إليها؟ أهم ما تقررته هنا هو أن تتعامل مع البيانات كما تتعامل مع الحقائق الحيوية الأخرى، كالمقابلة مع مسؤول مثلاً.

فهم البيانات تعني القدرة على القراءة من أجل المعرفة، والتفكير بشكل نقدي في المواد المنشورة. وتحليل الإحصاءات والأرقام، وفهم كيفية التعامل مع المعطيات الواسعة للمعلومات

ثالثاً: تبسيط البيانات وترتيبها (ساعة)

تعامل مع الأرقام وكأنها شيء مُسلّ يكشف الأسرار ويزيل الغموض عن القصص بطريقة سهلة.

دع البيانات تتحدث دون أن تُسقط عليها اعتقاداتك.

اطرح أسئلة عن الأرقام التي تراها، فالأرقام لا يجب أن تكون صحيحة أو خاطئة. لكن، عليك أن تفهم بالضبط ما الذي نعبه. لذلك، تخيل القصة التي يمكن أن تكون وراءها. ثم اختبرها من خلال حقائق أخرى.

لا تُهمل أبداً بيانات لمجرد أنك لا تفهمها. وتذكر أنه ليس مطلوباً منك أن تكون خبير إحصاء، بل أن تفهم فقط دلالات الأرقام.

ضعها جميعاً أمامك وابدأ بتصنيفها من حيث الجزئية التي تغطيها في موضوعك، ومن حيث أهميتها لحياة الناس. وبهذا يكون لديك بيانات من مصادر عدة لجزئية معينة.

رابعاً: تحليل البيانات (ساعة و٣٠ دقيقة)

اسأل الأسئلة الخمسة: من، متى، ماذا، كيف، أين، لماذا؟

يُنصح بجمع البيانات التي تجيب على سؤال "مَنْ" الذي أصدر البيانات، و"مِنْ أَيْنَ" أتت؟ بعدها اسأل لماذا هذه البيانات مهمة؟ وهذا سؤال مهم لأنه قد يقود إلى نتيجة مفادها أن البيانات مثل أي شكل من أشكال المعلومات ممكن أن تكون خاطئة. كما أن الاحتفاظ بالمصدر الذي جاءت منه المعلومة يجعل القارئ يثق بك.

ثم اجمع البيانات التي تجيب على سؤال "ماذا" تريد أن تخبر القارئ من معلومات؟ ومتى صدرت هذه البيانات؟ البيانات القديمة لا تعني أنها غير صالحة. لكن يجب عدم تقديمها على أنها الجديد في قصتك.

لا تنسَ هنا أنك لست أكاديمياً. وقراؤك ليسوا كذلك أيضاً. فصحافة البيانات تروي القصة بوضوح وسهولة وبساطة. لكن تذكر أن التقارير الإحصائية ليست موجهة للجمهور العام. لكن ما تكتبه أنت كذلك. لهذا، لا بد من تبسيط البيانات. ابدأ بصياغة ما تعنيه البيانات التي تجيب عن الأسئلة الخمسة بلغة سهلة وبسيطة ومفهومة.

أجب على سؤال "لماذا" حدث ذلك؟ واستخدم كل المعلومات التي جمعتها إلى جانب البيانات وابدأ بإيجاد رابط بينها.

سيكون واضحاً هنا ما إذا كنت قد حققت سبقاً صحفياً، وما الجديد الذي ستأتي به قصتك.

عندها سيبدأ العمل الأهم وهو تحليل البيانات التي حصلت عليها. لإجراء هذا التحليل، ستحتاج إلى بعض المهارات السهلة في الرياضيات والإحصاء على الأرجح أنك تعلمتها في الصفوف الابتدائية في المدرسة. أهمها النسبة المئوية (percentage)، والمنسوب/المستوى (rate)، المعدل (average)، والوسيط (median)، والمدى (range)، والأرباع (quartiles).

كما يجب أن تعرف كيف تستخدم الانحراف المعياري (standard deviation) لتوضيح القيم المنحرفة (outliers)، وكذلك بعض أدوات الإحصاء مثل علاقة التبادل (correlation) والتراجع (regression).

أهمية هذه المهارات هي أنها تُمكنك من الحرص على المصادقية التي يجب ألا تتعرض للخطر. والبيانات الخاطئة قد تسبب ذلك. وعندما تُضمن أرقاماً فيما ستقدمه لقرائك، فإنك بذلك تقدّم له شواهد (evidences) دامغة تثبت القصة.

عندما نمتلك بعض المهارات في الرياضيات، فإننا سنتمكن من معرفة متى تكون البيانات هي بحد ذاتها قصة مثيرة للاهتمام. وسنستذكر هنا بعض المهارات البسيطة الضرورية حتماً عند تحليل البيانات.

النسبة المئوية للتغيير (percentage change): على الرغم من أننا نهتم عادة بالنسب الأقل وأو الأكثر، أي الرقم الأصغر أم الأكبر. إلا أن الأهم غالباً هو معرفة نسبة التغيير في الوضع الذي نحن بصدده الحديث عنه. وهذا يظهر من خلال فهم التغيير بالأرقام.

الأمر سهل جداً. ما يجب أن تفعله هو أن تقارن الرقم الجديد بالرقم القديم. القاعدة هي أن تأخذ الرقم الجديد وتطرحه من الرقم القديم. ثم تقسم الناتج على الرقم القديم، تحصل على النسبة المئوية للتغيير.

مثال: صدر عن الشرطة الفلسطينية أن عدد حالات الانتحار ارتفعت في فلسطين هذا العام من ١٩ إلى ٣٢ العام الماضي. ما هي النسبة المئوية للتغيير؟

$$٣٢ - ١٩ = ١٣$$

$$\text{نقسم } ١٣ \div ١٩ = ٠,٦٨ \%$$

فنقول إن نسبة الانتحار ارتفعت في فلسطين هذا العام بنسبة ٠,٦٨%. وهذا يدل على وتيرة مرتفعة في التغيير.

أما إذا كانت حالات الانتحار قد انخفضت من ٣٢ إلى ١٩. فكيف تكون نسبة المتغيير؟

$$٣٢ - ١٩ = ١٣$$

$$\text{نقسم } ١٣ \div ٣٢ = ٠,٤٠ \%$$

فنقول إن نسبة الانتحار انخفضت بنسبة ٤٠,٠%.

المدى (range):

هو أبسط طريقة لفهم التغير في القيم. وهو ببساطة الفرق بين القيمة الأعلى والقيمة الأدنى.

أعلى رواتب يحصل عليها موظفو القطاع الحكومي ١٠٠٠٠ شيقل. أدنى راتب هو ١٥٠٠ شيقلا. المدى ٨٥٠٠ شيقلا.

أي أن الرواتب تتراوح بين ١٥٠٠ و ١٠٠٠٠ شيقلا.

المتوسط (mean) أو المعدل (average): تكون بجمع جميع القيم ثم قسمتها على عدد القيم. نحصل على المتوسط، المعدل لجميع القيم.

مثلا: درجات الحرارة في مدن فلسطين على النحو الآتي: القدس ١٥°، بيت لحم ١٤°، الخليل ١٢°، جنين ٢٥°، طولكرم ٢٣°، رام الله ١٣°، غزة ٢٢°، قلقيلية ٢٠°، أريحا ٣٠°.

يتم حساب المتوسط (mean) أو المعدل (average):

$$١٧٤ = ٣٠ + ٢٠ + ٢٢ + ١٣ + ٢٣ + ٢٥ + ١٢ + ١٤ + ١٥$$

$$١٩,٣ = ٩ \div ١٧٤$$

متوسط/معدل درجات الحرارة في فلسطين إذا ١٩,٣

الوسيط (median): طريقة أخرى لقياس القيم المختلفة وهي القيمة المتوسطة لمجموعة قيم مرتبة تصاعديا. فمثلا إذا كان لدينا قيم ١، ٢، ٣، ٤، ٥، فالوسيط هو ٣. أما إذا كان عدد القيم زوجيا، ١، ٤، ٦، ٧، ٩، ١٠، نقسم الوسيطين ويكون الوسيط هو الوسط الحسابي لهما. أي $٦ + ٧ = ١٣$

$$٦,٥ = ١٣ \div ٢$$

أهمية الوسيط هي أنه يقسم البيانات إلى الأرباع (quartiles) إلى النصف الأعلى والنصف الأسفل من البيانات وكل نصف فيه ربعين ٤/٢. لذا، إذا وجدت

وسيط (median) البيانات، وذهبت إلى النصف الأدنى من البيانات، ستجد الربع الأول، وإذا ذهبت إلى الأعلى ستجد الربع الثالث.

القيم المتطرفة (outliers): هي القيم البعيدة عن الوسيط، وهي أكثر ما نهتم بها، لأنها مصدر المعلومات غير المعتادة وهي التي تدل على ما هو مختلف وغير مألوف.

لذلك كثيرا ما تجد القصص الجديدة في القيم المتطرفة. لكن كيف نعرف ما إذا كانت هذه القيم متطرفة فعلا أم لا؟ هنا تبرز أهمية الانحراف المعياري.

الانحراف المعياري:

هو أحد طرق معرفة الوسط الحسابي لغالبية القيم حتى في ظل وجود تباين كبير مع القيم المتطرفة منها. وبهذا يمكن الحكم على حالات تباين القيم.

مثلا: إذا كان الانحراف المعياري لمعدلات الفقر في المحافظات الفلسطينية صفرا.

هذا يعني أن نسبة الفقر في فلسطين هي نفسها. ممكن أن تكون ١٠%. أو ٥٠%. لكن لا يوجد فرق في نسب الفقر بين المحافظات.

لكن، إذا كان الانحراف المعياري لمستويات الفقر ٢٠%. يجب أن نعرف المتوسط الحسابي للقيم. إذا كان ٣٠%، فإن مستويات الفقر ستتراوح بين ١٠% و ٥٠%

$$\text{إما } ١٠ = ٢٠ - ٣٠$$

$$\text{أو } ٥٠ = ٢٠ + ٣٠$$

قد يكون مستوى الفقر في محافظة ما ٧٠، لكن هذه القيمة متطرفة، لأنها بعيدة جدا عن الوسط. الانحراف المعياري لا يستثنى عنها لكنه يقلل من أهميتها مقارنة بغالبية القيم. وبذلك نعرف أن هناك ما يُقال عن ارتفاع نسبة الفقر في تلك المحافظة. لأنه (الفقر) أكثر بكثير من بقية المحافظات.

أغلبية الجهات التي توفر بيانات كجهاز الإحصاء المركزي تذكر مقدار الانحراف المعياري في المعطيات التي تقدمها. لكن، إذا لم يكن الانحراف المعياري مذكورا، يمكن حسابه من خلال برنامج ال (excel).^١

ليس من المهم أن نذكر الإنحراف المعياري (standard deviation) عندما نحلل البيانات، لكن من المهم أن نقيس من خلاله القيم المتطرفة، أي القيم الأكثر بعدا عن المركز، كي نتمكن من التأكد من صدق المعلومة وتبسيطها.

المصادر:

- Paul Bradshaw, Alberto Cairo, Steve Doig, Simon Rogers, Nicolas Kayser-Bril. Doing Journalism with Data: First Steps, Skills and Tools.
- Jonathan Gray, Liliana Bounegrou & Lacy Chambers (editors). The Data Journalism Handbook.

ⁱ - حساب الانحراف المعياري بواسطة excel:

http://libweb.surrey.ac.uk/library/skills/Number%20Skills%20Leicester/page_20.htm