

# أساسيات علم البيانات

## ❖ علوم الرياضيات التي تحتاجها لتصبح عالم بيانات Mathematics Needed to Become a Data Scientist

تتطلب خوارزميات علم البيانات بالإضافة إلى تنفيذ التحليلات واكتشاف الأفكار من البيانات الموجودة معرفة رياضية، ففي حين أن الرياضيات لا تعتبر الأداة الوحيدة المطلوبة لعالم البيانات، ولكنها من أهم تلك الأدوات أحد العناصر الأكثر أهمية في سير عمل مشروع علم البيانات تحديد وفهم تحديات الأعمال وتحويلها إلى تحديات رياضية.

## ❖ الجبر الخطي Linear Algebra

يهتم الجبر الخطي بالمصفوفات والمتجهات، مما يُعد أمراً مهماً للغاية لأنه في نماذج علم البيانات والخوارزميات يتم تحويل جميع الأرقام والمعلومات إلى مصفوفات. ويتم استخدام تقنية أخرى ضرورية في معالجة البيانات الضخمة وترتكز على الجبر الخطي، وهي تقنية تقليص الأبعاد. وكذلك تُعد رؤية الحاسب (Computer Vision) والبرمجة اللغوية العصبية (NLP) من مجالات علم البيانات التي تعتمد بشكل كبير على الجبر الخطي.

## ❖ الرياضيات المتقطعة Discrete Mathematics

تتخصص الرياضيات المتقطعة في طرق المنطق والاستنتاج، وهي جوانب أساسية في تصميم الخوارزميات وتعد أساس علم البيانات. ومن المجالات المهمة جداً الخاصة بالرياضيات المتقطعة هي نظرية المخططات تستخدم المخططات في نمذجة شبكات معقدة للغاية مثل شبكات تنظيم الجينات، وتعتبر دراسة هذه المخططات في علم البيانات مهمة جداً للتقدم في بعض المجالات مثل الطب الدقيق وبيولوجيا الأنظمة وغيرها الكثير.

## ❖ الاحتمالات والإحصاء Probability and Statistics

عند إنشاء البيانات بعد تحليلها، يحتاج عالم البيانات إلى معرفة عملية بالإحصاء والاحتمالات لكي يتمكن من فهم وتفسير تلك البيانات. يستخدم علماء البيانات مقاييس مثل التباين والارتباط والانحراف المعياري على نطاق واسع للحصول على نظرة ثاقبة على العلاقات الأساسية لخصائص مجموعة البيانات.

## ❖ التفاضل والتكامل Calculus

يُعد تمثيل النتائج من تحليل البيانات أمراً بالغ الأهمية لتوفير معلومات مستنيرة من خلال إنشاء الرسوم والمخططات البيانية. يعتبر التفاضل والتكامل جزءاً لا يتجزأ من الخوارزميات المستخدمة في العمليات الحسابية المعقدة المطلوبة لهذه العملية، ويتم استخدام خصائص مثل الاشتقاق الجزئي، والانحدار الخطي، والنزول الاشتقاقي على نطاق واسع في التطوير والتحسين وحساب الخسارة.

## البايثون في علم البيانات Python for Data Science

يفضل محترفو علم البيانات عامة استخدام لغة البايثون في مشروعاتهم المختصة بهذا الجانب، فهي لغة برمجة عالية المستوى كائنية التوجه وسهلة التعلم، ومن السهل البدء في العمل على مشروع ما بحيث يمكنك بدء كتابة كود برمجي بسيط أو تصميم وتنفيذ حل باستخدام مبادئ البرمجة كائنية التوجه (OOP).

يوفر استخدام واجهات برمجة التطبيقات (APIS) والمكتبات القياسية الوصول إلى دوال قوية سهلة الاستخدام. توجد العديد من المكتبات الجاهزة للاستخدام في البايثون من قبل المتخصصين في المؤسسات المختلفة تغطي مجموعة متنوعة من الاحتياجات مثل: استخراج البيانات، وإعداد البيانات وتحليلها، ومعالجة البيانات والنمذجة التنبؤية، وتمثيل البيانات، وإعداد التقارير كذلك فإن مكتبات البايثون تدعم تطبيقات تعلم الآلة ومتطلبات الذكاء الاصطناعي المتقدمة بما يتجاوز تطبيقات علم البيانات التقليدية.

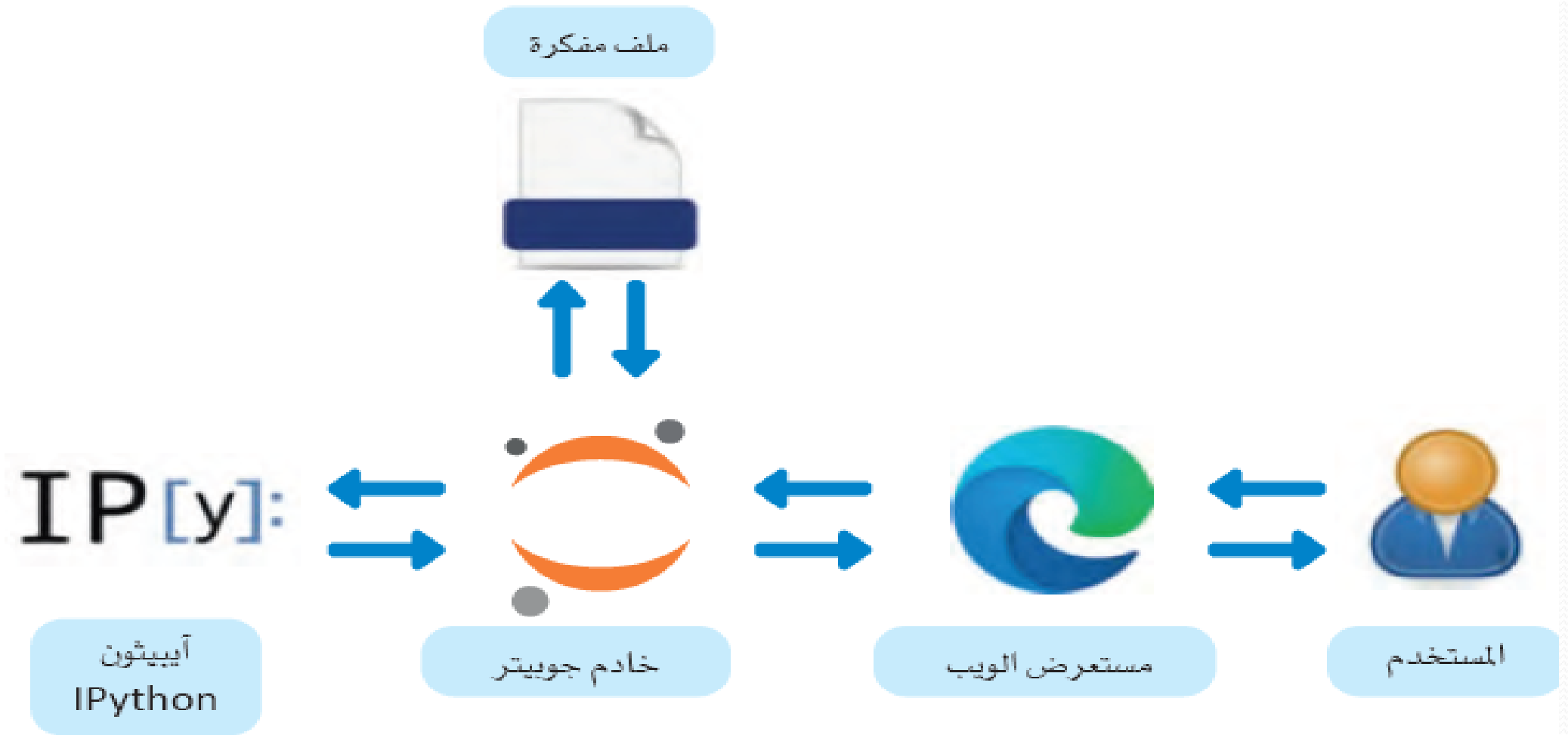
البايثون (Python) لغة برمجة عالية المستوى تستخدم لأغراض متعددة وقد اكتسبت شعبية متزايدة في علم البيانات وتعلم الآلة.

## مقدمة إلى مفكرة جوبيتر Jupyter

يمكن كتابة أوامر البايثون النصية في أحد بيئات التطوير المتكاملة (IDE) مثل فيجول ستوديو كود (Visual Studio Code) أو PyCharm أو يمكنك كتابتها في مفكرة جوبيتر (Jupyter Notebook).

تعتبر مفكرة جوبيتر أحد تطبيقات الويب مفتوحة المصدر المستخدمة لتطوير وتقديم مشروعات علم البيانات باستخدام البايثون، تتيح البيئة التفاعلية لعلماء البيانات إنشاء مفكرات محوسبة، وتدمج مفكرة جوبيتر أوامر البايثون وتخرجها في مستند واحد يجمع بين التمثيلات والمعادلات الرياضية وأشكال البيانات الأخرى بعد تثبيت البرنامج، يمكن تشغيله في متصفح الويب إما عبر الإنترنت أو على حاسب شخصي.

تدعم مفكرة جوبيتر إلى جانب البايثون أكثر من ١٠٠ لغة برمجة يطلق عليها اسم أنوية (kernels) في بيئة نظام مفكرة جوبيتر بما فيها Java و R و Julia و MATLAB و Octave و Scheme و Processing و Scala وغيرها. يقوم تطبيق مفكرة جوبيتر بتشغيل نواة IPython فقط، ولكن يمكن تثبيت أنوية إضافية.



شكل 1.9: هيكلية مفكرة جويتر

# أدوات علم البيانات Tools for Data Science

علم البيانات عملية معقدة تتطلب الكثير من الخطوات لتوفير حلول لعلم البيانات، ولكل خطوة من خطوات هذه العملية العديد من الأدوات لإنجاز المهمة المطلوبة.

جدول 1.9: الأدوات الشائعة لعلم البيانات

أدوات البرمجيات	الغرض	
MySQL و SQL Server و MongoDB و Neo4j.	قواعد البيانات حيث يتم تخزين البيانات.	تخزين البيانات
Python و SQL و Apache TinkerPop.	الأدوات التي تستعلم عن البيانات التي تريد تحليلها.	تحويل صيغة البيانات
Pandas و NumPy و Apache Spark.	تحويل البيانات المطلوبة إلى نماذج مناسبة للتحليل.	النمذجة
Tensorflow و PyTorch و IBM Watson و AWS و Sagemaker.	العملية التي تولد الرؤى المطلوبة.	التحليل
Matplotlib و D3.js و R.	تصوير النتائج في الشكل الأمثل.	التصوير



## الجدول 1.10: المهن المتعلقة بعلم البيانات

علم البيانات هو أحد أسرع التخصصات المتعلقة بعلوم الحاسب نمواً والأكثر طلباً في الوقت الراهن، وتبدو الفرص الوظيفية في علم البيانات واعدة.

عالم بيانات	مهمته هي إيجاد البيانات ومعالجتها وتحليلها للشركات والمؤسسات. يأخذ البيانات الأولية وغير المعالجة ثم يتم استخراج رؤى وأتماط من البيانات تساعد الشركات والمؤسسات على تحليل أدائها واتخاذ قرارات مهمة.
مهندس تعلم آلة	مسؤول عن تنفيذ حلول وأنظمة تعلّم الآلة لتطبيقات معينة. يجب أن يكون على دراية بهندسة البرمجيات وبالإحصاء ليتمكن من اختيار الحلول والحكم على صحة نماذج تعلّم الآلة الناتجة.
متخصص تعلم آلة	بينما يختص مهندسو تعلّم الآلة في تطبيق نماذج تعلّم الآلة، فإن متخصص تعلّم الآلة يركز على إيجاد الخوارزميات الرياضية المحددة التي تنتج النماذج التي يمكن للمهندسين استخدامها بعد ذلك.
معماري تطبيقات	يقوم بتصميم نظم المعلومات للمؤسسات والشركات.
معماري أنظمة المؤسسات الكبيرة	يجمع بين المعرفة التجارية والتقنية، وعلى تواصل مستمر بين أصحاب الأعمال والإدارات الفنية، ويكلف بترجمة احتياجات بيانات الأعمال والمؤسسات إلى مواصفات وحلول تقنية تُرسل إلى الفرق الفنية.
معماري بيانات	مسؤول عن تخزين وسير المعلومات في شركة أو مؤسسة. يعمل مع علماء ومهندسي البيانات لتطوير طرق نقل البيانات بصورة مناسبة لإدخال مجموعة البيانات وتحليلها وإخراج النتائج.
مهندس بيانات	يساعد مهندس البيانات في بناء الإطار الرقمي لجمع البيانات وتخزينها ومعالجتها، والتي سيستخدمها علماء البيانات والمحللون في عملهم.
معماري البنية التحتية	دوره يختص في إدارة البنية التحتية حيث يتم تخزين البيانات ومعالجتها، ويأخذ بالاعتبار عوامل مثل خصوصية البيانات والحماية وأداء البنية التحتية على الخوادم حيث يتم تحليل البيانات، ولقد أصبحت مشروعات علم البيانات أكثر تعقيداً بشكل عام، لذلك يحتاج مهندسو البنية التحتية إلى التأكد من اكتمال معالجة البيانات ضمن الجداول الزمنية المناسبة.
محلل بيانات	هو محترف يأخذ الرؤى من مجموعات البيانات المعالجة ويُنشئ التقارير والتصورات والتحليلات الأخرى المتنوعة والتي تتماشى مع الأهداف الرئيسية التي ينبغي للحل المبني على علم البيانات تحقيقها.





# مجتمعات علم البيانات عبر الإنترنت Data Science Online Communities

يرغب علماء البيانات في البقاء على اتصال مع أقرانهم في هذا المجال أو في المهن المماثلة لتعلم أفكار وطرق جديدة لأن منهجيات وتقنيات علم البيانات دائمة التغير. وتوفر الموارد عبر الإنترنت لعلماء البيانات الفرصة فقط في الحفاظ على وتيرة معينة، وهنا برزت الحاجة إلى وجود مجتمع من خبراء علم البيانات لدعم هذا العمل وظهور مجموعة متنوعة من المنتديات والمجموعات عبر الإنترنت والتي تمكنهم من الاتصال معا وتطوير هذا المجال بكفاءة من خلال المشاركة في مجتمعاته عبر الإنترنت.

<https://www.kaggle.com>

• كاجل Kaggle

كاجل شركة تابعة لجوجل، وهي أكبر مجتمع لعلم البيانات يضم ملايين الأعضاء النشطين ومجموعة واسعة من الموارد، ويمكن لعلماء البيانات العثور على مجموعات البيانات العامة والموارد التعليمية وبيئة العمل المستندة إلى مجموعة النظراء لدعم عمل تحليل البيانات الخاص بهم.

• مجتمع بيانات IBM Data Community IBM

مجتمع بيانات IBM عبارة عن منتدى عبر الإنترنت به مدونات مخصصة لعلم البيانات. يستضيف الأوراق البحثية والبت عبر الإنترنت والعروض التقديمية التي يتم تحديثها مع تطور هذا المجال.

<https://community.ibm.com/community/user/datascience>

## المصطلحات الرئيسية

Artificial Intelligence	ذكاء اصطناعي	Data Visualization	تصوير البيانات
Big Data	البيانات الضخمة	Data Warehouse	مستودع البيانات
Business Intelligence	ذكاء الأعمال	Exploratory Data Analysis	التحليل الاستكشافي للبيانات
Cloud Storage	التخزين السحابي	Information	معلومات
Data	البيانات	In-Memory Computing	الحوسبة في الذاكرة
Data Analysis	تحليل البيانات	Jupyter Notebook	مفكرة جوبيتر
Data Analyst	محلل البيانات	Knowledge	المعرفة
Data Cleaning	تنظيف البيانات	Python	البايثون
Data Collection	جمع البيانات	Raw Data	بيانات أولية
Data Lake	بحيرة بيانات	Value	القيمة
Data Mining	التنقيب في البيانات	Variety	التنوع
Data Preparation	تجهيز البيانات	Velocity	السرعة
Data Science	علم البيانات	Veracity	الموثوقية
Data Science Life Cycle	دورة حياة علم البيانات	Volume	الحجم
Data Scientist	عالم البيانات		

# جمع البيانات والتحقق من صحتها

- سيتعرف الطالب في هذه الوحدة على مفاهيم جمع البيانات وطرق التحقق من صحتها.
- ويشمل ذلك معنى جمع البيانات وأنماطها المختلفة ومصادرها. وسيتم التطرق إلى موضوع ترميز البيانات والتركيز على ميزات وعيوب هذه الطرق، وفي الختام سيتعرف الطالب على خطوات التحقق من صحة البيانات وأنماطها.
- أهداف التعلم
- سيكون الطالب قادر على أن:
- يعرف مصطلح جمع البيانات
- يصنف مصادر البيانات
- يصنف معايير جودة المعلومات
- يعرف مفهوم منصات البيانات المفتوحة
- يعرف الأنماط المختلفة للبيانات
- يعرف المقصود بترميز البيانات
- يطبق خطوات التحقق من صحة البيانات

# جمع البيانات Data Collection

- ✓ تعد مرحلة جمع البيانات (Data Collection) أهم مراحل البحث العلمي أو دراسة ظاهرة معينة وتمثل هذه المرحلة عملية جمع الحقائق والأرقام والمصطلحات للمتغيرات المستهدفة في ذلك البحث أو الدراسة. يمكن جمع البيانات باستخدام أجهزة مختلفة مثل أجهزة الاستشعار وتسجيل البيانات.
- ✓ تتطلب عملية جمع البيانات فهماً عميقاً للمعاملات قيد الدراسة.
- ✓ التخطيط والعمل الدؤوب للحصول على بيانات عالية الجودة، حيث تسهم البيانات عالية الجودة في إجراء التحليل المناسب لأداء المهام على الوجه المطلوب وذلك لاستخراج معلومات مفيدة حول الظاهرة قيد الدراسة.
- ✓ تختلف طرق جمع البيانات باختلاف نوع البيانات، ولكن يظل التحقق من صحة ودقة الخطوات المتعلقة بجمع البيانات أمراً هاماً.

# مصادر البيانات Sources of Data

يوجد تصنيفان أساسيان لمصادر البيانات: مصادر البيانات الرئيسة Primary data sources، ومصادر البيانات الثانوية Secondary data sources .

## ❖ مصادر البيانات الرئيسة Primary Data Source

يحتوي مصدر البيانات الرئيسة على بيانات لم تجمع من قبل ويمكن جمعها من خلال أجهزة الاستشعار وأجهزة تسجيل البيانات وكذلك من خلال الاستبانة. من الأمثلة التي تدل على مصادر البيانات الرئيسة مقياس درجة الحرارة الذي يجمع بيانات درجة حرارة الهواء، ومقياس سرعة الرياح الذي يقيس سرعة الرياح، وكذلك فإن إجراء استبانة للعملاء حول طبيعة الطقس الذي يفضلونه عند السفر للخارج يُعد مصدرًا من مصادر البيانات الرئيسة.

## ❖ مصادر البيانات الثانوية Secondary Data Source

تحصل على هذا النوع من البيانات من خلال استخدام بيانات أخرى من مصدر البيانات الرئيس. فعلى سبيل المثال، يمكنك أولاً استخدام مقاييس درجة الحرارة ومقاييس الرياح كمصادر بيانات أولية لحساب بيانات درجة الحرارة وسرعة الرياح، ثم يمكنك استخدام صيغة حساب درجة البرودة كمصدر بيانات ثانوي من أجل الحصول على بيانات درجة البرودة.

## الجدول 2.1: أوجه الاختلاف بين مصادر البيانات الرئيسية والثانوية

مصادر البيانات الثانوية	مصادر البيانات الرئيسية
لا تعتبر بيانات أصلية لأنها تعتمد على بيانات أخرى تم جمعها من قبل آخرين.	يتم جمعها مباشرة من مصادرها الأصلية.
تكون بصورة منظمة ومعالجة.	تكون بصورة خام وغير معرّفة.
أقل دقة لأنها مرتبطة بأحداث ماضية.	أكثر دقة لكونها بيانات حالية.
يتم جمعها من الكتب والمجلات والمستندات والصفحات الإلكترونية والمدونات وغير ذلك.	يتم جمعها بواسطة أجهزة الاستشعار، الاستبيانات، المقابلات والتجارب وغيرها.
أقل تكلفة وأقل استهلاكاً للوقت.	عالية التكلفة وتتطلب الكثير من الوقت لجمعها.

## • مصادر البيانات الداخلية والخارجية Internal and External Data Sources

- ✓ يمكن تصنيف مصادر البيانات إلى مصادر داخلية وخارجية.
- ✓ تعكس مصادر البيانات الداخلية تلك البيانات التي يتم جمعها داخليا من قبل جهة العمل أو الجهة المسؤولة عن جمع البيانات.
- ✓ بينما البيانات الخارجية هي التي يتم تحصيلها من خارج مجال جهة العمل.
- ✓ وعلى سبيل المثال، تعتبر البيانات التي يتم جمعها من جهاز استشعار تابع لجامعة أو مؤسسة علمية بيانات داخلية، بينما تعتبر البيانات التي يتم الحصول عليها من مؤسسات أخرى أو أفراد أو من مصادر خارج الجامعة بيانات خارجية.



## جودة المعلومات Information Quality

عندما تتم معالجة البيانات أو تنظيمها أو تقديمها في سياق معين فإنها تسمى معلومات، وتوصف قيمتها لاستخدام معين بأنها "جودة المعلومات". تعد جودة المعلومات عاملا مهما في قيمة تلك المعلومات ومدى مصداقيتها للاستخدام في اتخاذ القرارات. ومع زيادة جمع وحفظ البيانات، أصبحت جودة المعلومات الناتجة عن معالجتها ذات أهمية كبيرة ومتزايدة. يساعد ضمان جودة المعلومات على التحديد الدقيق للاحتياجات الفعلية لتنفيذ المشروعات، وكذلك توجيه الخدمات بشكل فعال، وزيادة الكفاءة في العمل. وبالمقارنة يمكن أن تتسبب المعلومات غير الدقيقة في حدوث خلل في الأعمال وتقلل من الكفاءة وتؤدي إلى التأخير في إنجاز المشروعات. يمكن التحقق من جودة المعلومات من خلال معايير محددة تسمى معايير الجودة (Quality standards)، وهي:

- الكفاية (Completeness)
- الدقة (Precision)
- الملاءمة (Appropriateness)
- مستوى التفاصيل (Level of Detail)
- التوقيت (Timeliness)

## فيما يلي بعض الأسئلة التي يمكن أن تساعدك في التحقق من دقة المعلومات:

- هل يمكن التحقق من الحقائق والإحصاءات والمعلومات من مصادر أخرى؟
- هل يمكن تكرار التجربة التي تم الحصول على البيانات منها؟ وهل لها نفس النتائج؟
- ما مصدر المعلومات؟
- لماذا تم إنشاء المعلومات؟
- هل تبدو المعلومات دقيقة بناءً على معرفتك وخبرتك؟
- هل تتضمن المعلومات أخطاء إملائية أو أحرفاً في غير مكانها؟
- هل الاقتباسات التي تم الاستشهاد بها صحيحة؟

تكمُن أهمية **معايير الجودة** الخمس هذه في أنها تساعدك على التحقق من موثوقية كل المعلومات التي تجدها في المواقع الإلكترونية.

### ● أولاً: الدقة Accuracy

الدقة هي التأكد من صحة المعلومات، ويجب أن تكون المعلومات صحيحة لكي تُعد عالية الجودة. ويتم التحقق من دقتها من خلال المصادر الموثوقة، حيث يسهل التأكد من خلوها من الأخطاء الحسابية أو اللغوية أو غيرها من الأخطاء.

### ● ثانياً: الملاءمة Appropriateness

يجب أن تكون المعلومات مرتبطة بموضوعك أو بالسؤال البحثي، فكلما كانت المعلومات متعلقة بما تبحث عنه كلما كانت ملاءمتها أفضل. يؤدي الحصول على معلومات إضافية لا تتعلق بما تبحث عنه إلى إضاعة الوقت في محاولة العثور على المعلومات المطلوبة.

### ● ثالثاً: التوقيت Timeliness

يُعد تاريخ نشر المعلومات جزءاً مهماً، حيث يوضح مدى حداثة المعلومات ومناسبتها لموضوع البحث ولذلك يجب التأكد من الحصول على آخر تحديث للمعلومات، ويجب عليك التحقق دائماً من مصدر المعلومات.

## ● رابعًا: مستوى التفاصيل Level of detail

تحدد جودة المعلومات أيضًا من خلال النظر إلى مستوى التفاصيل التي تقدمها تلك المعلومات. يعد وجود الكثير من التفاصيل على المعلومات المطلوبة أمرًا صعبًا، بينما تؤدي قلة التفاصيل إلى صعوبة فهم المعلومات. ويعتمد مستوى التفاصيل على المشكلة ودراستها، فإذا كانت بسيطة فإنك تحتاج إلى مستوى قليل من التفاصيل، وإذا كانت معقدة فإنك تحتاج إلى مستوى عالٍ من التفاصيل. إن المقدار الصحيح من المعلومات هو معيار أساسي لجودتها.

## ● خامسًا: الكفاية Completeness

تعد كفاية المعلومات مقياسًا مهمًا للشمولية المطلوبة للتأكد من أن المعلومات المقدمة تعطي صورة كاملة عن الواقع. إن عدم الحصول على جميع المعلومات المطلوبة يعني أنك لن تتمكن من استخدامها بشكل صحيح، مما يعني أن جودة تلك المعلومات ضعيفة وغير كاملة ولا يمكن اتخاذ القرارات الصحيحة بناءً على تلك المعلومات.

## منصات البيانات المفتوحة

# Open Data Platforms

- ✓ منصات البيانات المفتوحة هي منصات تدعم وصول المستخدمين لمجموعات البيانات المفتوحة.
- ✓ تقدم منصات البيانات المفتوحة النموذجية بيانات المؤسسة التي تستضيف هذه المنصة. تستضيف الجهات الحكومية أو المؤسسات غير الربحية منصات البيانات المفتوحة التي تتيح وصول الجمهور إلى البيانات.
- ✓ تقوم بجمع البيانات وتنظيمها باستمرار من مجموعة متنوعة من القطاعات العامة. ويمكن استخدام مجموعات البيانات هذه دون تكلفة الصفح المجموعات مالية أو تحديات تقنية للوصول إليها.
- ✓ يمكن إعادة استخدام البيانات المفتوحة وإعادة توزيعها مع مراعاة المتطلبات التي يفرضها ترخيص البيانات. ويمكن أيضاً استخدامها من قبل مواطني الدول الأخرى.

هذه بعض الاستخدامات الشائعة لمنصات البيانات المفتوحة:

- الشفافية في إعداد الموازنة الحكومية والإنفاق على خدمات الدولة.
- إحصائيات أداء المؤسسات الحكومية.
- بيانات أداء مختلف القطاعات العامة للدولة كالتعليم والرعاية الصحية والنقل، والتي يمكن استخدامها في إجراء الأبحاث التي تُبرز الأداء العام داخل هذه الدولة.

## خصوصية البيانات Data Privacy

مصطلح "البيانات الشخصية" يُطلق على البيانات المتعلقة بالأشخاص وهي تلك البيانات التي من خلالها يتم تحديد هوية شخص معين، مثل اسم الشخص وعائلته ورقم هاتفه ورقم الهوية. أدّت عملية توسع وسائل التواصل الإلكتروني إلى زيادة في المخاطر المتعلقة بالخصوصية والتي عليك حماية نفسك منها، وتتضمن خصوصية البيانات قدرة الشخص على أن يقرر بنفسه متى وكيف وإلى أي مدى يمكن مشاركة معلوماته الشخصية مع الآخرين أو تزويد الآخرين بتلك المعلومات.

الأذونات القانونية لجمع واستخدام البيانات

## Legal Permissions to Collect and Use Data

يتطلب جمع البيانات واستخدامها في المشروعات البحثية توفر التصاريح القانونية المناسبة، تختلف هذه التصاريح في طبيعتها، ولكن العاملين الرئيسيين اللذين يجب أخذهما بعين الاعتبار هما الموقع الذي سيتم فيه تخزين البيانات، وموقع المستفيدين الذين سيتعاملون مع تلك البيانات تحتاج الشركات والمؤسسات إلى التأكد من أن الخدمات التي تجمع البيانات وتستخدمها متوافقة قانونياً مع قوانين بلدانها.



# أنواع البيانات

## Numerical Data and Categorical Data البيانات الرقمية والبيانات الفئوية

بعد أن تعرفت على مفهوم البيانات الآن ستتعرف على أنماطها المختلفة. يمكن أن تتخذ البيانات عدة أشكال مثل عدد زوار حدث معين، أو مدة تلك الزيارة، أو غير ذلك. في الدراسات البحثية، هناك نوعان أساسيان من البيانات وهما البيانات الرقمية والبيانات الفئوية.

### ١. البيانات الرقمية Numerical Data

تتكون البيانات الرقمية من حقائق قابلة للقياس كعدد الفعاليات المختلفة في مدينة ما. ويمكن للبيانات الرقمية أن تكون منفصلة أو تسلسلية.

#### • البيانات المنفصلة Discrete Data

تمثل البيانات المنفصلة عناصر قابلة للعد يمكنها أن تأخذ قيما معينة فقط، مثل عدد الطلبة في الفصل.

#### • البيانات التسلسلية Continuous Data

البيانات التسلسلية تمثل بيانات يمكن قياسها ويمكنها أن تأخذ أي قيمة مثل طول الشخص.

## ٢-البيانات الفئوية Categorical Data

البيانات الفئوية هي نوع بيانات غير كمية، يمكن تخزينها وتحديدًا بناءً على الأسماء أو التسميات المعطاة لها. ويمكن أن توصف هذه البيانات على أنها اسمية أو ترتيبية.

### • البيانات الاسمية Nominal Data

يتم تعريف البيانات الاسمية على أنها بيانات تُستخدم لتسمية المتغيرات أو توسيمها، دون تحديد قيمة كمية لها أو وضعها في ترتيب معين. على سبيل المثال: يمكن تصنيف نتائج اختبار ما اسميًا: "ناجح" أو "راسب".

### • البيانات الترتيبية Ordinal Data

البيانات الترتيبية هي نوع من البيانات الفئوية التي تتبع ترتيبًا معينًا. يتم فرز مجموعات البيانات الترتيبية وفقًا لنظام تصنيف محدد. على سبيل المثال: يمكن فرز نتائج اختبار ما بترتيب تنازلي حسب الدرجة: A ثم B ثم C ثم D..

## • البيانات الرسومية أو بيانات مقاطع الصوت أو الفيديو

### Graphical, Video and Audio Data Types

على الرغم من أن البيانات عادة ما تكون في شكل أبجدي رقمي (نصوص وأرقام ورموز) ، إلا أنها قد تتكون من صور أو مقاطع صوتية أو مقاطع فيديو فيما يلي بعض الأنواع الأخرى من البيانات:

## • البيانات الرسومية Graphical Data

تتكون البيانات الرسومية من مخططات ورسوم بيانية وغير ذلك. على سبيل المثال: مجموعة الصور الخاصة بالمعالم السياحية لمنطقة محددة، أو الرسم البياني الخاص بأعداد الزوار لأحد الأماكن السياحية.

## • بيانات الصور ومقاطع الفيديو Image and Videos Data

قد تكون الصورة الرقمية صورة أو رسماً توضيحياً يتم تمثيله بوحدات البيكسل أو بالمتجهات. تتكون بيانات الفيديو من سلسلة من الصور المتحركة والصوت ومن الأمثلة عليها الإعلان التلفزيوني لحملة سياحية.

## • البيانات الصوتية Audio Data

تتكون البيانات الصوتية من الأصوات وكذلك المؤثرات الصوتية المختلفة كالتسجيلات الصوتية الإرشادية في المتاحف والأماكن السياحية المختلفة.

## البيانات الثابتة والمتغيرة Static and Dynamic Data

قد تبقى البيانات على حالتها الأصلية بعد تسجيلها وقد تتغير أحياناً، ولذلك يمكن تمثيل البيانات بشكل ثابت أو متغير.

### • البيانات الثابتة Static Data

البيانات الثابتة هي البيانات التي لا تتغير بعد تسجيلها، مثلاً البيانات التي تحتويها مجلة مطبوعة حول الأماكن السياحية، فبمجرد طباعة المجلة لا يمكن تغيير المعلومات فيها.

### • البيانات المتغيرة (الديناميكية) Dynamic Data

البيانات المتغيرة هي البيانات التي قد تتغير بعد تسجيلها ويجب تحديثها باستمرار. على سبيل المثال بيانات موقع إلكتروني يحتوي على معلومات عن الأماكن السياحية، حيث يمكن تحديثها عند الحاجة.

## ترميز البيانات Data Coding

إن البيانات التي تم الحصول عليها من التجارب أو من خلال الدراسات الاستقصائية هي بيانات في صورتها الأولية وفي الغالب تحتاج إلى ترميز. تتيح عملية الترميز تنظيم البيانات وترتيبها بطريقة محددة وذلك باستخدام رموز مختلفة مثل الأرقام أو الحروف أو الكلمات القصيرة، والتي تصف سياقاً معيناً أو تعبر عن عبارة أو فقرة بأكملها، وفيما يلي أمثلة من الحياة اليومية حيث تُستخدم الرموز لتمثيل البيانات:

## رموز المطارات Airport Codes

وضع اتحاد النقل الجوي الدولي (IATA) رمزاً مكوناً من ثلاثة حروف يحدد العديد من المطارات حول العالم. يمكنك البحث عن تذاكر الطيران على الإنترنت باستخدام رمز المطار، ويمكنك ملاحظة الرمز كذلك على بطاقات الأمتعة التي يتم إلصاقها بالحقائب عند تسليمها في مكاتب تسجيل المغادرة بالمطار استعداداً لركوب الطائرة مما يتيح شحن الأمتعة للوجهة المقصودة، كما يظهر الرمز على بطاقات الأمتعة المرفقة في مكاتب تسجيل الوصول بالمطار؛ لتوفير السلامة في حالة فقدان الأمتعة.

## رموز العملات Currency Codes

لكل بلد في أنحاء العالم عملة خاصة به، وتستخدم رموز العملات بدلا من اسم العملة كاختصارات متعارف عليها عند التعاملات المالية

### • مزايا ترميز البيانات:

١. إدخال أسرع للبيانات (من الأسهل كتابة الرمز CA بدلا من كتابة مطار القاهرة الدولي)
٢. تأخذ مساحة أقل (من الصعب كتابة الاسم الكامل للبلد على اللوحات أو على وسائل النقل العام مثل سيارات الأجرة والحافلات، ولكن يصبح الأمر سهلا مع رموز تسجيل المركبات الدولية).
٣. تسريع عملية البحث (لكل منطقة رمزها الخاص يُستخدم هذا الرمز للبحث عن عنوان حسب رمز المنطقة ورقم عن البيانات يستخدمه مكتب البريد لتسهيل توزيع رسائل البريد)

## عيوب ترميز البيانات:

١. معنى غامض للبيانات (قد تجد صعوبة في التمييز بين الرموز المتشابهة)
٢. صعوبة فهم الترميز (قد يكون من الصعب تفسير أو تذكر معنى الرمز)
٣. الرموز المستخدمة قد تستنفد (مثلا قد يكون عدد العناصر المراد ترميزها كبيرا جدا بحيث لا تكفي مجموعة الأحرف فتدمج الأرقام والحروف أو تستخدم الأرقام الطويلة وهذا يعقد عملية الترميز، مثل ترميز المنتجات الاستهلاكية في المتاجر)

## **الرموز الشريطية Barcodes**

ترى الرموز الشريطية في كل مكان حولك وبشكل يومي، فمثلا تجدها مطبوعة على التذاكر الإلكترونية والمنتجات في محلات البقالة وغيرها. الرمز الشريطي هو ملصق به خطوط سوداء رفيعة إلى جانب مجموعة متنوعة من الأرقام. تستخدم في تنظيم المعلومات وفهرستها أو وضع علامة على أسعار المنتجات



## رقم الكتاب المعياري الدولي ISBN (International Standard Book Number)

يوجد رقم على غلاف معظم الكتب (مثل الكتيب الإرشادي) أعلى الرمز الشريطي يُسمى رقم الكتاب المعياري الدولي (ISBN) م فريد يستخدمه الناشرون والمكتبات ومحلات بيع الكتب لتحديد عناوين الكتب وإصداراتها. يتكون رقم الكتاب المعياري الدولي من ثلاثة عشر خانة عشرية ويقسم إلى خمس مجموعات متتالية من الأرقام.

### مثال

فيما يلي مثال على رمز ISBN، حيث يشير كل عنصر إلى جزء معين من المعلومات حول الدليل.



الشكل 2.16: مثال على ISBN المكون من 13 رقمًا



## رموز الاستجابة السريعة QR Codes



رمز الاستجابة السريعة Quick Response QR هو بمثابة الجيل الثاني من الرمز الشريطي (barcode) والذي يتكون من خطوط سوداء متجاورة ومختلفة الة السمك ويحتوي على مزيد من المعلومات. قد يشير رمز الاستجابة السريعة إلى محتوى إلكتروني مثل المواقع الإلكترونية أو مقاطع الفيديو، أو الملفات الرقمية، ويمكن قراءة هذا الرمز باستخدام كاميرات الهواتف الذكية.

# التحقق من صحة إدخال البيانات

## Data Entry Validation التحقق من صحة إدخال البيانات

يشير مفهوم التحقق من صحة إدخال البيانات إلى أي نشاط يتحقق من أن البيانات المدخلة تنبثق من مجموعة من القيم المعتمدة، وتتوافق مع القواعد المقبولة للبيانات، وقد تتبع تلك البيانات بعض العمليات والإجراءات التصحيحية، وتهدف عملية التحقق من صحة البيانات إلى ضمان الدقة والجودة وتنفيذ من خلال إنشاء عدة فحوصات لضمان الاتساق المنطقي للبيانات المدخلة والمخزنة. ويتم قبول البيانات في حال توافقها مع القوانين، ويتم تجاهل ورفض البيانات

التحقق من صحة البيانات:

فحوصات للتأكد من التناسق المنطقي للبيانات المدخلة والمخزنة.

على سبيل المثال، يمكن أن يتراوح النطاق المقبول لقيم درجة حرارة الهواء المسجلة من جهاز استشعار درجة الحرارة من سالب ٨٩.٢ درجة مئوية (أدنى درجة حرارة) إلى ٥٨ درجة مئوية (أعلى درجة حرارة)، ويجب ألا يسجل جهاز الاستشعار قيم درجة حرارة الهواء مثل ١٠٠ درجة مئوية، حيث يشير ظهور مثل هذا النوع من البيانات في نتائج جهاز الاستشعار المسجلة إلى حدوث خلل ويجب رفض القيمة.

## **أنواع التحقق من صحة إدخال البيانات** Types of Data Entry Validation

يوجد العديد من أنواع التحقق التي يمكن استخدامها للتأكد من صحة البيانات المدخلة، كما يمكن استخدام برامج مثل مايكروسوفت إكسل (Microsoft Excel) للتحقق من صحة البيانات المدخلة. تتبع معظم أساليب التحقق من صحة إدخال البيانات طريقة أو أكثر من طرق التحقق أدناه للتأكد من أن البيانات صحيحة قبل تخزينها. وأكثر هذه الطرق شيوعاً هي التحقق من البحث والتحقق من التواجد، والتحقق من الطول، والتحقق من النطاق، والتحقق من الصيغة والتحقق من النوع.

## أنواع التحقق من صحة البيانات:

- التحقق من البحث (Lookup check) :

يساعد على تقليل الأخطاء باستخدام قائمة محدودة النطاق تحتوي على قيم مدخلة مسبقا، فبدلا من كتابة اسم المطار يمكن لموظف شركة الطيران اختيار المطار من قائمة تحتوي على رموز المطارات حول العالم. هذا النوع من التحقق يساعد على ادخال البيانات بسرعة وبنسبة أخطاء أقل.

- التحقق من التواجد (Presence check) :

يجعل عملية الإدخال إلزامية في الخلية مما يضمن عدم تركها فارغة. يجب إدخال البيانات الهامة وإلا فإن هناك مشكلة في تكامل البيانات، مثلا لا يمكن ترك الحقول الخاصة باسم الشخص واسم عائلته فارغة.

- التحقق من الطول (Length check) :

يهدف إلى التأكد من أن الرموز والحروف تُدخل بنطاق طول محدد. مثلا أرقام ISBN أو رموز العملات أو رموز البلدان ١٥٠ لها طول ثابت مكون من ١٣ و ٣ و ٢ رقمًا أو حرفا، على التوالي.

- التحقق من النطاق (Range check) :

يستخدم للتأكد من أن الأرقام المدخلة تقع ضمن نطاق معين ويشمل حدين هما الحد الأقصى (Maximum limit) والحد الأدنى (Minimum limit) ، فمثلا إذا كنت تريد إدخال عمر الشخص، فيجب على النظام قبول الأرقام الموجبة بحد أعلى ١٤٠ فقط، وأي شيء آخر يتجاوز هذا النطاق سيعد بيانات غير صالحة.

- التحقق من الصيغة (Format check)

يُستخدم للتأكد من أن البيانات تأتي بصيغة محددة مسبقا ولن يسمح بأي صيغة أخرى يتم إدخالها في الخلية تكون التحقق من الصيغة عمليات التحقق هذه مفيدة عند استخدام حقول البيانات للرموز البريدية أو أرقام الهواتف. في كلتا الحالتين، يتوقع النظام تنسيق بيانات محدد للغاية.

- التحقق من النوع (Type check)

يضمن إدخال المستخدمين لنوع القيمة الصحيح في حقل محدد، فمثلا إذا تم تمييز حقل البيانات على أنه حقل رقمي فلن تتمكن من تخزين قيمة نصية.

## المصطلحات الرئيسية

Accuracy	الدقة	Information Quality	جودة المعلومات
Appropriateness	الملاءمة	Isbn	رقم الكتاب المعياري الدولي
Audio Data	البيانات الصوتية	Legal Permissions	الأذونات القانونية
Barcode	الرموز الشريطية	Length Check	التحقق من الطول
Check Digit	التحقق من الرقم	Level Of Detail	مستوى التفاصيل
Completeness	الكفاية	Lookup Check	التحقق من البحث
Data	البيانات	Open Data Platforms	منصات البيانات المفتوحة
Data Coding	ترميز البيانات	Presence Check	التحقق من التواجد
Data Types	أنواع البيانات	QR Code	رمز الاستجابة السريعة
Data Validation	التحقق من صحة البيانات	Range Check	التحقق من النطاق
Fixed Data	بيانات ثابتة	Type Check	التحقق من النوع
Format Check	التحقق من الصيغة	Variable Data	بيانات متغيرة