

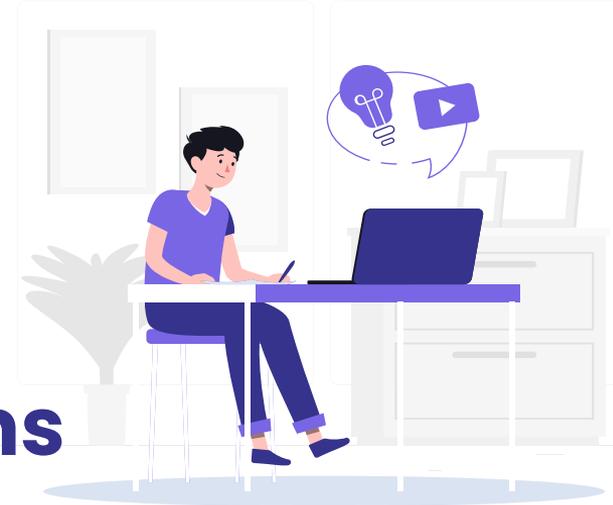
جامعة طرابلس  
كلية تقنية المعلومات  
قسم نظم المعلومات

المقرر الدراسي ITIS406

أنظمة دعم القرار

**Decision support systems**

المحاضرة السابعة



اعداد: أ.فاطمة بشير القاضي  
algadyfatma@gmail.com

# مقدمة

مع وجود كميات كبيرة من البيانات المخزنة في قواعد البيانات ازدادت الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات واستخراج المعلومات والمعارف منها حتى ظهر ما يسمى بالتنقيب في البيانات كتقنية تهدف إلى استخراج المعرفة من كميات هائلة من البيانات.

وبالرغم من أن تقنية تنقيب البيانات تعتبر مازالت في طور النمو إلا انها فرضت نفسها بقوة في عصر المعلوماتية واستخدامها يوفر للشركات والمنظمات في جميع المجالات القدرة على استكشاف أهم المعلومات في قواعد البيانات كما تركز تقنيات التنقيب على بناء التنبؤات المستقبلية واستكشاف السلوك والاتجاهات مما يسمح باتخاذ القرارات الصحيحة في الوقت المناسب.

# تتقيب البيانات Data Mining

- استكشاف وتحليل كميات كبيرة من البيانات من اجل استكشاف انماط وقواعد ذات مغزى.

- عملية استكشاف مؤتمتة لبيانات مفيدة في مصادر بيانات ضخمة. يتم نشر تقنيات التتقيب عن البيانات لصقل قواعد بيانات ضخمة بهدف ايجاد نماذج غريبة ومفيدة قد تبقى غير معروفة بدونها. كما انها تقدم امكانيات لتوقع احداث مستقبلية.

- عملية بحث مؤتمتة عن معرفة من البيانات دون وجود فرضيات مسبقة عما يمكن أن تكون هذه المعرفة. حيث يقوم على تحليل كمية كبيرة من البيانات لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة للمهتمين بهذه البيانات.

# تابع تنقيب البيانات

- خطوات غير تقليدية لإيجاد أنماط مفيدة وغير مكتشفة سابقا من بيانات خام.
- تنقيب البيانات هو استخراج الأنماط المفيدة من مصادر البيانات على سبيل المثال قواعد البيانات والويب والصور.
- الأنماط : هي العلاقات الرياضية او المنطقية التي تربط البيانات.
- يجب أن تكون الأنماط صالحة **valid** ، غير مألوفة **novel** ، يحتمل أن تكون مفيدة **potentially useful** ، مفهومة **understandable**.
- يُطلق على تنقيب البيانات أيضًا اكتشاف المعرفة و التنقيب عن البيانات **knowledge discovery and data mining (KDD)**

## Example : Analysis of purchases in a supermarket

customer1	pizza	beer	cheese	bread	chips
customer2	milk	bread	ham	cigaretts	
customer3	yoghurt	sugar	flour	cornflakes	napkins
customer4	shampoo	beer	chips	newspaper	pizza
customer5	chips	coffee	beer	pizza	cream
customer6	Jam	rolls	butter	beer	
...					

# الفوائد الرئيسية للتنقيب في البيانات

➤ اكتشاف الأنماط: يعد اكتشاف الأنماط التلقائي ميزة استراتيجية تساعد في نمذجة السلوك المستقبلي والتنبؤ به.

➤ تحليل الاتجاهات: إن فهم الاتجاهات يبيّنك على اطلاع دائم بالتطورات الحالية في الصناعة ، ويساعد على تقليل التكاليف واختيار التوقيت المناسب لطرح منتجاتك في السوق .

➤ كشف الاحتيال: تساعد تقنيات التنقيب عن البيانات في اكتشاف الاحتيال من خلال اكتشاف الحالات الشاذة في مجموعات البيانات، يستخدم لاكتشاف الاحتيال في النشاطات التي من المحتمل أن تتعرض لاحتيال..

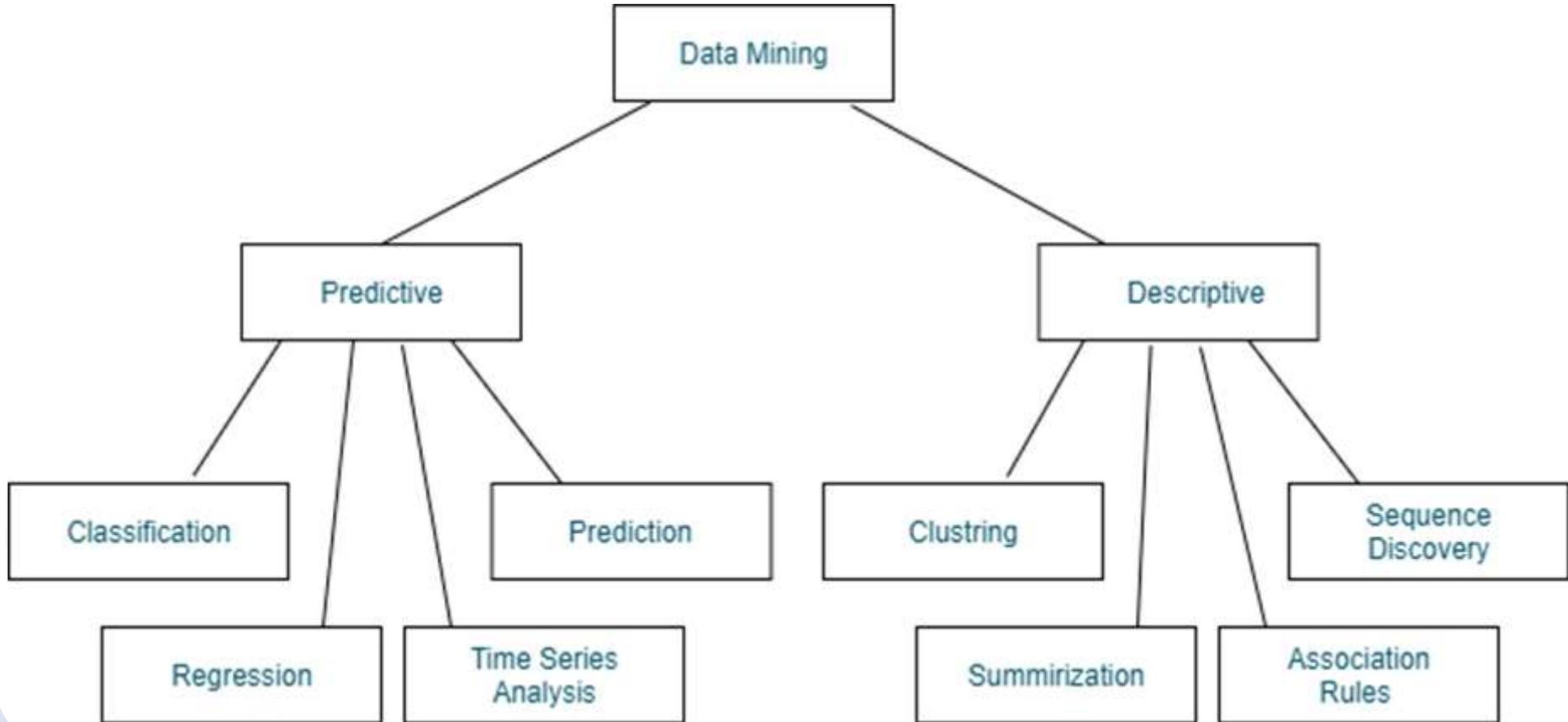
➤ التنبؤ في الأسواق المالية: تُستخدم تقنيات التنقيب عن البيانات على نطاق واسع لنمذجة الأسواق المالية والتنبؤ بالنتائج المحتملة.

# الطرق المستخدمة في التنقيب

هناك عدة أدوات تستخدم لأجل تنقيب البيانات لذا فإن اختيار الأداة المناسبة يعتمد على طبيعة البيانات التي تتم دراستها فضلا عن حجم البيانات، وهناك نموذجان لتنقيب البيانات هما:

- النموذج التنبؤي Predictive model: وهو النموذج الذي يستخدم النتائج المعروفة المستنبطة من البيانات المختلفة لأجل التنبؤ بقيم لاحقة للبيانات.
- النموذج الوصفي Descriptive model: وهو النموذج الذي يعرف الانماط والعلاقات من البيانات، وعلى عكس النموذج التنبؤي يستخدم النموذج الوصفي كطريقة لاكتشاف خصائص البيانات التي تتم دراستها وليس للتنبؤ بخصائص جديدة.

# تابع الطرق المستخدمة في التنقيب



# تابع الطرق المستخدمة في التنقيب

أشهر أدوات تنقيب البيانات التي تدرج تحت النموذج التنبؤي:

## 1. تحليل الانحدار Regression Analysis:

تحليل الانحدار هو تحليل يمكننا من إيجاد معادلة رياضية تربط بين متغير تابع ومتغير أو عدة متغيرات مستقلة. فمثلا يمكننا باستخدام تحليل الانحدار دراسة العوامل التي تؤثر في زيادة الطلب على المنتج وتحديد نموذج رياضي (معادلة) لهذه العلاقة. هذا النموذج يجعلنا قادرين ليس فقط على فهم طبيعة العلاقة وتحديد العوامل المؤثرة فعلا بل إنه يجعلنا قادرين على توقع تأثير تغير أي متغير من هذه المتغيرات المستقلة على المتغير التابع.

# تابع الطرق المستخدمة في التّقيب

## 2. التصنيف Classification

التصنيف هو أسلوب شائع يفصل مجموعة من البيانات إلى فئات (تصنيفات) مختلفة يمكن استخدامها فيما بعد لتصنيف البيانات المستقبلية، أهميته تكمن في انه يسمح بتنظيم مجموعات البيانات من جميع الأنواع ، بما في ذلك مجموعات البيانات المعقدة والكبيرة بالإضافة إلى مجموعات البيانات الصغيرة والبسيطة. مثلا باستخدام هذا التحليل يمكننا بناء نموذج لتصنيف طلبات الحصول علي القروض المصرفية اما امانة (safe) أو خطرة (risky) هذا التحليل يمكن أن يوفر لنا فهما افضل للبيانات ككل.

# تابع الطرق المستخدمة في التّقيب

## 3. تحليل السلاسل الزمنية Time Series Analysis:

تحليل السلاسل الزمنية عبارة عن مراقبة قيم ظاهرة من الظواهر في سلسلة تواريخ متلاحقة، أياما أو أشهراً أو سنوات والهدف من وراء تحليل سلاسل الزمنية هو دراسة التغيرات التي تكون قد طرأت علي الظاهرة التي تمثلها خلال فترة زمنية وتحليل أسبابها كنتائجها أو التنبؤ اعتمادا علي فكرة من حوادث الماضي بما سيحدث في المستقبل، لأن مراقبة البيانات عبر الزمن تنتج تحليلا مفيداً.

# تابع الطرق المستخدمة في التنبؤ

## 4. التنبؤ Prediction :

التنبؤ عبارة عن عملية تقدير و تخمين ذكي و مدروس مبني على الطبيعة الظاهرة و تطويرها ونموها في وضعها الحالي ودرجة النمو واتجاهاته و مداه و قوته بعد أن يتم اخضاع كل ذلك لأدوات القياس المناسبة، والتنبؤ يبني على الكيفية التي تكون عليها الظاهرة في وضعها الطبيعي دون أن يؤخذ بعين الاعتبار اي امور طارئه او استثنائية قد تأخذ مكانها ويكون لها فعل و أثر في وقت من الاوقات. وكلما كانت دراسة الظاهرة دقيقة في كافة الجوانب كلما كان التنبؤ الى حد ما صحيحا وهذا يعتمد على استقرار وثبات الشروط و الظروف الموضوعية و عدم تبديلها او تغيرها، ويرافق عملية التنبؤ الاستعداد المسبق لعملية تغيير قبل حدوثها مما يجعل عملية التكيف سهلة و مقبولة في وقت لاحق.

# تابع الطرق المستخدمة في التنقيب

أشهر أدوات تنقيب البيانات التي تدرج تحت النموذج الوصفي:

## 1. العنقدة (التجميع) Clustering:

التجميع هو عملية تحديد مجموعات البيانات المتشابهة مع بعضها البعض, حيث يساعد على تقسيم البيانات إلى عدة مجموعات فرعية تحتوي على بيانات متشابهة, وتسمى هذه المجموعات الفرعية بالعناقيد, وبعد تصنيف البيانات إلى مجموعات فرعية مختلفة يتم تعيين تسمية لكل مجموعة. الآن بعد أن تم تقسيم البيانات من قاعدة بيانات العملاء مثلا إلى مجموعات, يمكننا اتخاذ قرار مستنير بشأن ما المنتج الذي نعتقد أنه الأنسب لزبون ما.

# تابع الطرق المستخدمة في التقيب

## 2. التلخيص Summarization:

التلخيص هو العثور على وصف موجز لمجموعة البيانات او مجموعة فرعية من مجموعة البيانات، ويشير التلخيص إلى أساليب تفتتت كتل البيانات الكبيرة إلى مقاييس موجزة توفر وصفا عاما للمتغيرات و علاقاتها. ومن الأمثلة على الأساليب المستخدمة في التلخيص نذكر: المتوسطات، و المجاميع، والإحصائيات الوصفية التي تتضمن مقاييس النزعة المركزية مثل المتوسط الحسابي و الوسيط والمنوال، ومقاييس التشتت مثل الانحراف المعياري. يوفر تلخيص عدد كبير من العناصر المتعلقة بمبيعات موسم العودة على الدراسة مثلا وصفاً عاماً موجز للبيانات ، والذي يمكن أن يكون مفيدة للغاية لمديري المبيعات والتسويق.

# تابع الطرق المستخدمة في التنقيب

## 3. قواعد الربط Association Rules:

قواعد الارتباط هي احدى التقنيات الواعدة في تنقيب البيانات كأداة من ادوات ولديها القدرة على معالجة كميات هائلة من البيانات، KDD استكشاف المعرفة وتسمح باستنتاج كل القوانين الممكنة التي ترشح بعض الصفات الموجودة اعتماد على وجود الصفات الأخرى. مثلا يمكن للأطباء استخدام قواعد الارتباط للمساعدة في تشخيص المرضى. هناك العديد من المتغيرات التي يجب مراعاتها عند إجراء التشخيص ، حيث تشترك العديد من الأمراض في الأعراض . باستخدام قواعد الارتباط وتحليل البيانات المدعوم بالتعلم الآلي ، يمكن للأطباء تحديد الاحتمال الشرطي لمرض معين من خلال مقارنة علاقات الأعراض في البيانات من الحالات السابقة . مع إجراء التشخيصات الجديدة.

# تابع الطرق المستخدمة في التنقيب

## 4. اكتشاف التسلسل Sequence Discovery:

اكتشاف التسلسل يُعنى بدراسة كيفية عمل شيء ما بترتيب معين. على سبيل المثال، قد يشتري المستخدم بشكل متكرر جل الحلاقة قبل شراء ماكينة حلاقة في المتجر. يعود الأمر كله إلى الترتيب الذي يشتري به المستخدم المنتج، ويمكن لصاحب المتجر بعد ذلك ترتيب العناصر وفقًا لذلك.

# مراحل Crisp-DM: عملية التنقيب في البيانات القياسية

CRISP-DM (Cross-Industry Standard Process for Data Mining)

- 1. فهم مشكلات الأعمال Business Issue Understanding:**  
تبدأ المرحلة الأولى بفهم المشروع ومتطلبات العمل. الناتج من هذه المرحلة هو تعريف مشكلة التنقيب في البيانات والتي يمكن معالجتها باستخدام نموذج القرار.
- 2. فهم البيانات Data Understanding:**  
يتم في هذه المرحلة جمع البيانات الأولية ومتابعة أنشطتها لفهمها؛ وتحديد جودتها وتناقضاتها ومشاكل قراءتها ، لاكتشاف المستوى الأولي للأفكار.

# تابع مراحل Crisp-DM

## 3. إعداد البيانات Data Preparation:

تغطي هذه المرحلة جميع الأنشطة التي تم القيام بها لإنشاء مجموعة البيانات النهائية من البيانات الأولية. لاحظ أن تحويل البيانات الأولية إلى مجموعة بيانات تحليلية يشكل 90% من وقت المشروع.

بمجرد تحديد مصادر البيانات، نحتاج إلى اختيار وتوضيح وبناء وتنسيق البيانات في النموذج المطلوب. القيام بمهمة استكشاف البيانات يجب أن تتم بتعمق وتدقيق كبير لملاحظة الأنماط التي يستند عليها فهم الأعمال التجارية. جودة البيانات التي تم توضيحها أو البيانات النهائية ستؤثر على أداء النموذج النهائي.

# تابع مراحل Crisp-DM

## 4. النمذجة Modelling:

النماذج هي تطبيقات خوارزمية للبحث وتحديد وعرض أي أنماط أو رسائل في البيانات , تساعدنا النمذجة على فهم مشاكل الأعمال واتخاذ قرار بشأن أفضل نموذج نوصي به. كما أنه يساعدنا في إعداد النماذج والتصورات التي تجعل من السهل على العملاء فهم بياناتهم واتخاذ أفضل قرارات العمل قدر الإمكان.

## 5. التقييم Evaluation:

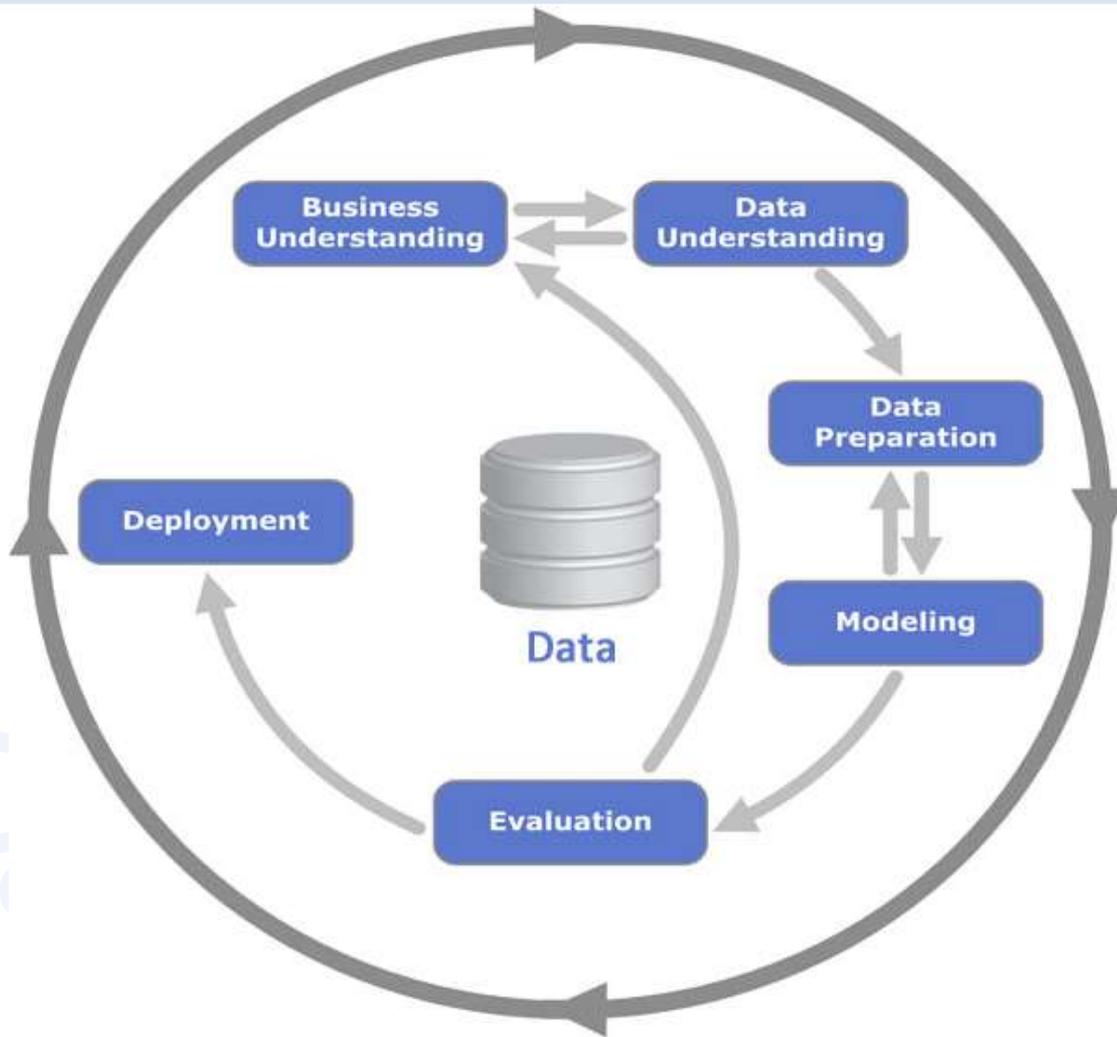
في هذه المرحلة يتم تقييم النتيجة في سياق أهداف العمل. الهدف في هذه المرحلة هو تحديد ما إذا كان هناك أي قضية تجارية مهمة لم يتم النظر فيها بشكل كافٍ. وفي نهاية هذه المرحلة، يجب اتخاذ قرار الانتقال أو عدم الانتقال إلى مرحلة النشر.

# تابع مراحل Crisp-DM

## 6. النشر Deployment

يتم في هذه المرحلة تحديد كيفية استخدام النتائج. يجب تنظيم المعرفة المكتسبة وعرضها بطريقة يمكن لأصحاب المصلحة استخدامها. ووفقًا للمتطلبات، مرحلة النشر يمكن أن تكون بسيطة مثل إنشاء تقرير، أو معقدة مثل تنفيذ عملية تنقيب في البيانات القابلة للتكرار.

يقدم CRISP-DM إطارًا موحدًا لتوثيق الخبرة والإرشادات العامة. بالإضافة إلى ذلك، يمكن لـ CRISP-DM أن يطبق في صناعات مختلفة بأنواع مختلفة من البيانات.



مراحل  
**Crisp-DM**

# تطبيقات تنقيب البيانات

يلعب التنقيب عن البيانات دوراً أساسياً في البنوك، والتأمين والنقل والتجارة، ويمكن للتنقيب عن البيانات حل كثير من المشاكل الحسابية، وزيادة الأرباح وصنع قرارات حكيمة. ولم تكن العمليات التجارية التطبيق الأولى لتقنيات التنقيب عن البيانات إنما كانت من المجالات المهمة، لأن العمليات التجارية بها الكثير من بيانات المبيعات، مثل تسجيلات الشراء الخاصة بالمستهلكين، ومعلومات المستهلكين، ومعلومات الخدمة وغيرها. يمكن للشركات استخدام تنقيب البيانات لتصنيف المستخدمين من مجموعات المستخدمين الأساسية وإيجاد الخصائص المشتركة للمستخدمين ورغباتهم المستقبلية، وتقديم منتجات كافية وخدمات تلبي رغباتهم. وعند استخدام تطبيقات برامج التنقيب عن البيانات، يجب اختيار الخوارزمية المناسبة، والمعرفة فيما وراء البيانات التي يمكن إيجادها.

# تابع تطبيقات تنقيب البيانات

وبما أن التنقيب عن البيانات يقوم بعدة مهام مثل: تجميع البيانات وتخزينها وتنظيمها لذا فهو يستخدم في عدة مجالات مثل الطب والمالية والقانون والدفاع والتعليم وغيرها، وتستخدم معظم التطبيقات التنقيب عن البيانات للدعاية والتسويق والمبيعات، كما يمكن استخدامه في التشخيص، ومن الأمثلة على هذه التطبيقات:

- البنوك: تستخدم التنقيب عن البيانات للحصول على بيانات تساعد في جذب الزبائن.

- السوبر ماركت: ينظم بضائعه طبقا لنماذج البيع والمعلومات حول التجمعات بين المنتجات، وايجاد السلع التي تباع مع بعضها.

- تطبيقات تكنولوجيا المعلومات: يساعد التنقيب عن البيانات في التأكد من جودة البيانات، فعلى سبيل المثال في يساعد التنقيب عن البيانات في التطبيقات اللوجستية في اختيار الأفراد المناسبين للعمل في مشروعات محددة.

# تصور البيانات

## Data Visualization

هل سبق لك أن استخدمت إحصائيات موجزة، أشياء مثل ما هو متوسط الأرباح؟ ما هو متوسط المبيعات؟ يمكن أن تكون هذه الأرقام ملخصًا مفيدًا لمجموعة البيانات ، ولكن يمكنها أيضًا إخفاء التفاصيل داخل هذه البيانات. هناك خطر من أن الاعتماد على الملخصات يمكن أن يؤدي إلى إجابات مضللة أو يمكن كذلك أن تكون غير صحيحة. فقط من خلال تصور البيانات يمكننا رؤية الشكل الحقيقي للبيانات، فالاعتماد على ملخصات الجداول أو الأرقام يخفي المعلومات داخل البيانات. نحتاج إلى تصورها لاستكشاف بياناتنا وفهمها وشرحها، وهذه هي قوة تصور البيانات.

# تابع تصور البيانات

## تصور البيانات : Data Visualization

- هو التمثيل الصوري أو المرئي للمعلومات والبيانات في هيئة عناصر مرئية، مثل: الأشكال البيانية والخرائط وغيرها، وذلك من خلال الأدوات المتخصصة.
- تقنية تستخدم لتفسير كميات كبيرة من البيانات واكتساب نظرة ثاقبة عليها. يتم تحقيق ذلك من خلال العروض المرئية وغالبًا ما تكون تفاعلية للبيانات الخام.
- وهناك الكثير من التعريفات التي تصف تمثيل البيانات، ولكنها جميعًا تشترك في كونها تعني تحويل البيانات إلى أشكال أو صور مرئية يمكن قراءتها بسهولة ويسر.

# تابع تصور البيانات



56298

DUIS AUTE VIRE DO  
REPERENDERIT IN  
VELT ESSE CILLUM  
FUGIAT NULLA FARI



56298

LOREM IPSUM DOLOR  
CONSECTETUR ADIPS  
SED DO EUISMOD TEMPOR  
UT LABORE ET DOLORE



56298

SINT OCCAECAT CUP  
PROIDENT, SUNT IN  
OFFICIA DESERUNT  
ANIM ID EST LABOREUM

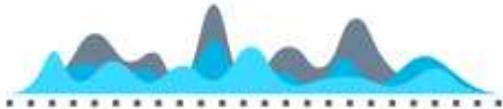


56298

IPSUM DOLOR LOREM  
CONSECTETUR ADIPS  
SED DO EUISMOD TEM  
OFFICIA DESERUNT



DUIS AUTE VIRE DOLOR REPERH  
ENDERIT IN VOL VELIT ESSE CILLUM



SIT AMET, CONSECTETUR ADIPISCI  
ELIT, SED DO EUISMOD TEMPOR  
INCIDIDUNT UT ENIM AD MINIM  
LOREM IPSUM DOLOR, CONSECTETUR  
ADIPS SED



53 43 98  
95 45



43 98 98  
33 89



23 18 27  
63 37



DUIS AUTE VIRE DO  
REPERENDERIT IN



LOREM IPSUM DOLOR  
CONSECTETUR ADIPS



SINT OCCAECAT CUP  
PROIDENT, SUNT IN



IPSUM DOLOR LOREM  
CONSECTETUR ADIPS



65496



98713



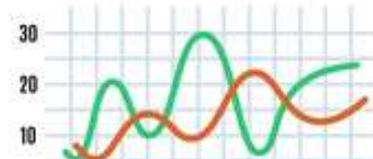
12386



49513



2576 DUIS AUTE VIRE DO  
7852 CONSECTETUR ADIPS  
1638 SINT OCCAECAT CUP  
5287 IPSUM DOLOR LOREM  
2787 PROIDENT SUNT IN



# مزايا تصور البيانات

1. معالجة كميات كبيرة من البيانات بسرعة: تعمل تقنية تصور البيانات على تمثيل مخازن البيانات الكبيرة والمعقدة في لمح البصر.
2. الحصول على البيانات الهامة في الوقت الحقيقي: تعمل تقنية تصور البيانات على تمثيل البيانات الضخمة في الوقت الفعلي ، كما تقوم باستمرار بتحديث المعلومات المقدمة.
3. فهم البيانات بشكل أفضل من خلال التفاعل: السماح للمستخدمين برؤية ما وراء الأرقام. من خلال التعمق في الأسباب والاتجاهات.
4. فحص الارتباطات بين العمليات التجارية والأداء: اكتشاف الروابط المخبئة بصرياً التي تربط الأساليب اليومية بأداء الأعمال.
5. وضع قصة ضمن البيانات: مشاهدة الأحداث والعواقب والتوصيات مصفوفة بطريقة نقل قصة.

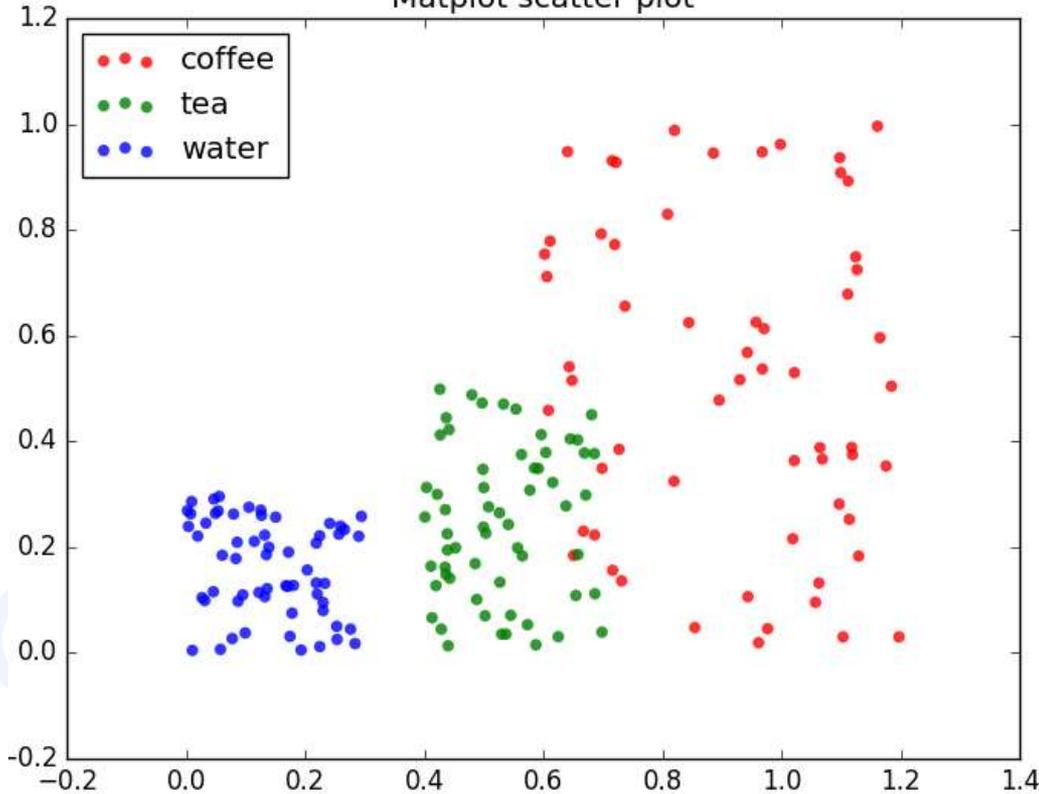
# أنواع تصور البيانات

عندما يتعلق الأمر بإنشاء تصور، فإن الحد الوحيد هو خيالك، ولكن هناك بعض الأنواع الشائعة:

- الرسوم البيانية الخطية Bar Graphs.
- الرسوم البيانية الدائرية Pie Charts.
- الجداول Tables.
- الانفوجرافيك Infographics
- الخرائط Maps
- صور GIF: يمكن ان تتضمن رسمًا بيانيًا متحركًا أو مخططًا يوضح النمو أو الحركة أو الانحدار أو التقدم أو يتغير بمرور الوقت.
- مخططات التشتت scatter diagram: هو نوع من الرسم البياني الذي يعرض الارتباط أو العلاقة بين موضوعين على محور أفقي ورأسي، وتوضح النقاط المنتشرة بينهما أمثلة على تلك العلاقة.

# تابع أنواع تصور البيانات

Matplot scatter plot



مثال على مخططات التشتت



مثال على الانفوجرافيك