



معسكر علم البيانات و تعلم الآلة

13 - 11 - 2022

نبذة عن المدرب





**Network: Student_Bootcamp
Password: Bootcamp@001**

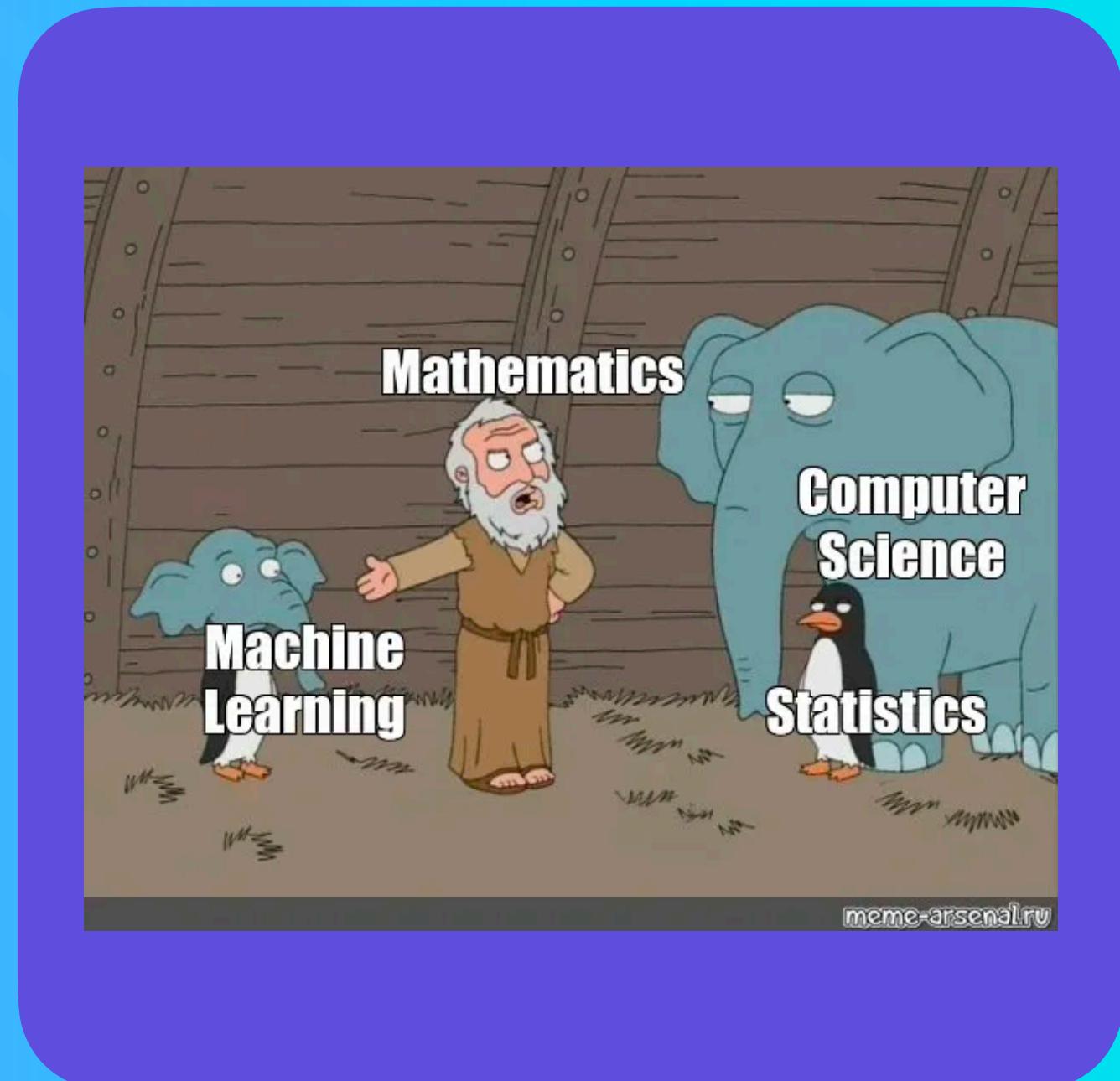
محتوى المعسكر

الأسبوع السادس Final Project	الأسبوع الخامس Modeling Interpretation in Action	الأسبوع الرابع EDA & FE in Action	الأسبوع الثالث Machine Learning	الأسبوع الثاني Data Analysis and Visualization	الأسبوع الأول Getting Started	اليوم
Final Project	Models Families: Distance & Time Series	DS Knowledge Catalog	Intro to ML	NumPy	Intro to DS	الأحد
Final Project	Models Evaluation: Regression & Classification	EDA1: Univariate & Multivariate Analysis	Supervised ML	Pandas	Git & Github	الاثنين
Final Project	Optimization Techniques	EDA2: Association Analysis & Hypothesis Construction	Supervised ML	Matplotlib	Python Review	الثلاثاء
Final Project	NLP and Text Mining Basics	Features Engineering: Scaling, Merging & Discretization	Unsupervised ML	Seaborn	Python Review	الأربعاء
Presentation	Neural Networks Basics	Models Families: Continuous & Categorical	Unsupervised ML	Plotly	Python Review	الخميس

*ملحوظة: قد تتغير الموضوعات أو أوقات طرحها بناءً على تقدم الطلاب.



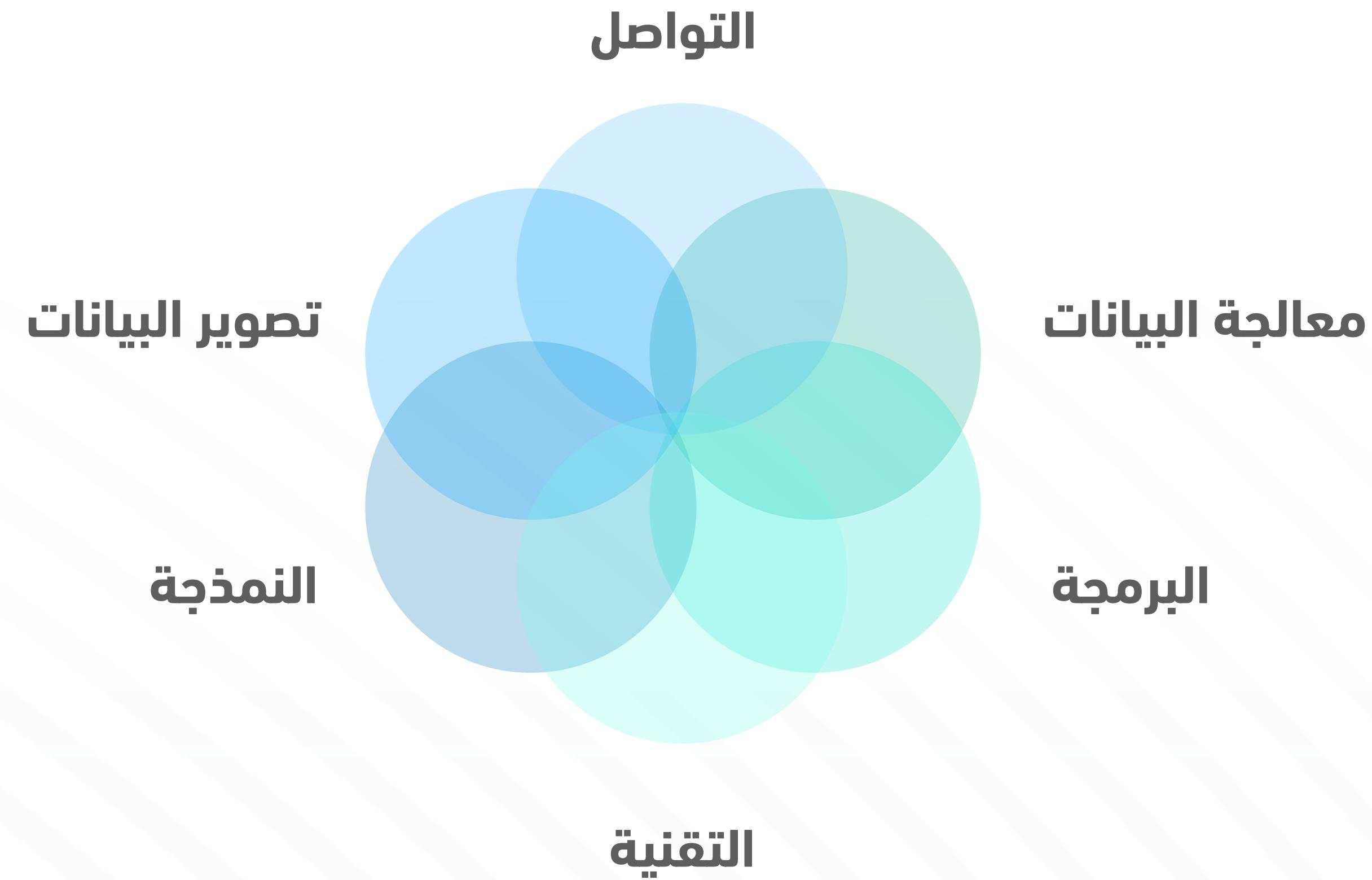
من هو عالم البيانات ؟



الإحصائي ... الاكتواري ... عالم البيانات



محلل البيانات ... عالم البيانات / الذكاء الاصطناعي





انعاش إصائي



المتغيرات العشوائية - Random Variables

خواص المتغيرات العشوائية:

- المتغير المجهول القيمة
- يحتمل المتغير العشوائي قيم متعددة من مجال مستمر
- تُستخدم لدراسة احتمالية قيم معينة لظاهرة عشوائية^٩

المتغيرات العشوائية - Random Variables

أنواع المتغيرات العشوائية:

- منفصلة

- مستمرة

التوزيعات الاحتمالية - Probability Distributions

علاقة التوزيعات الاحتمالية بالمتغيرات العشوائية:

- تأخذ المتغيرات العشوائية توزيعات احتمالية مختلفة
- التوزيع الاحتمالي يصف ظواهر تحتوي على متغيرات عشوائية

التوزيعات الاحتمالية - Probability Distributions

أمثلة على أنواع التوزيعات الاحتمالية المنفصلة:

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

- توزيع برونولي
- توزيع ثئي

التوزيعات الاحتمالية - Probability Distributions

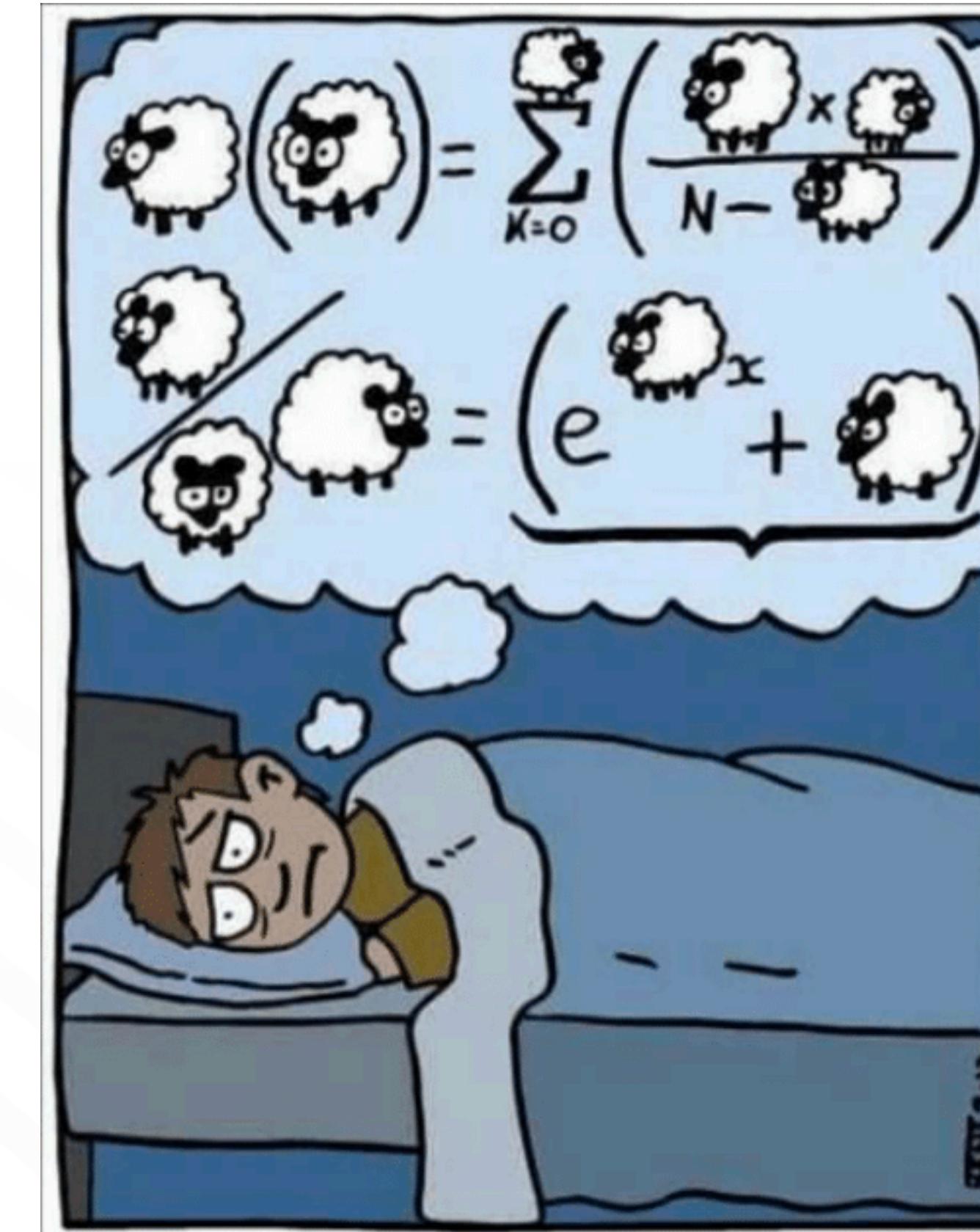
أمثلة على أنواع التوزيعات الاحتمالية المتصلة:

- التوزيع الطبيعي

- التوزيع التراكمي

ومواضيع الإدصاء في البيانات تطول

ولكن ...



ماذا تمثل النتيجة س .. ؟

{ 8, 4, 2, 2 }

س = 4

الوسط الحسابي، أو المعدل

ماذا تمثل النتيجة س .. ؟

{ 8, 4, 2, 2 }

$$س = 2^{(4 - 2)} = (2)$$

التبالين

ماذا تمثل النتيجة س .. ؟

{ 8, 4, 2, 2 }

$$س = 4 / (4+4+0+16) =$$

الانحراف المعياري

ماذا تمثل النتيجة س .. ؟

{ 8, 4, 2, 2 }

س = 2

المنوال

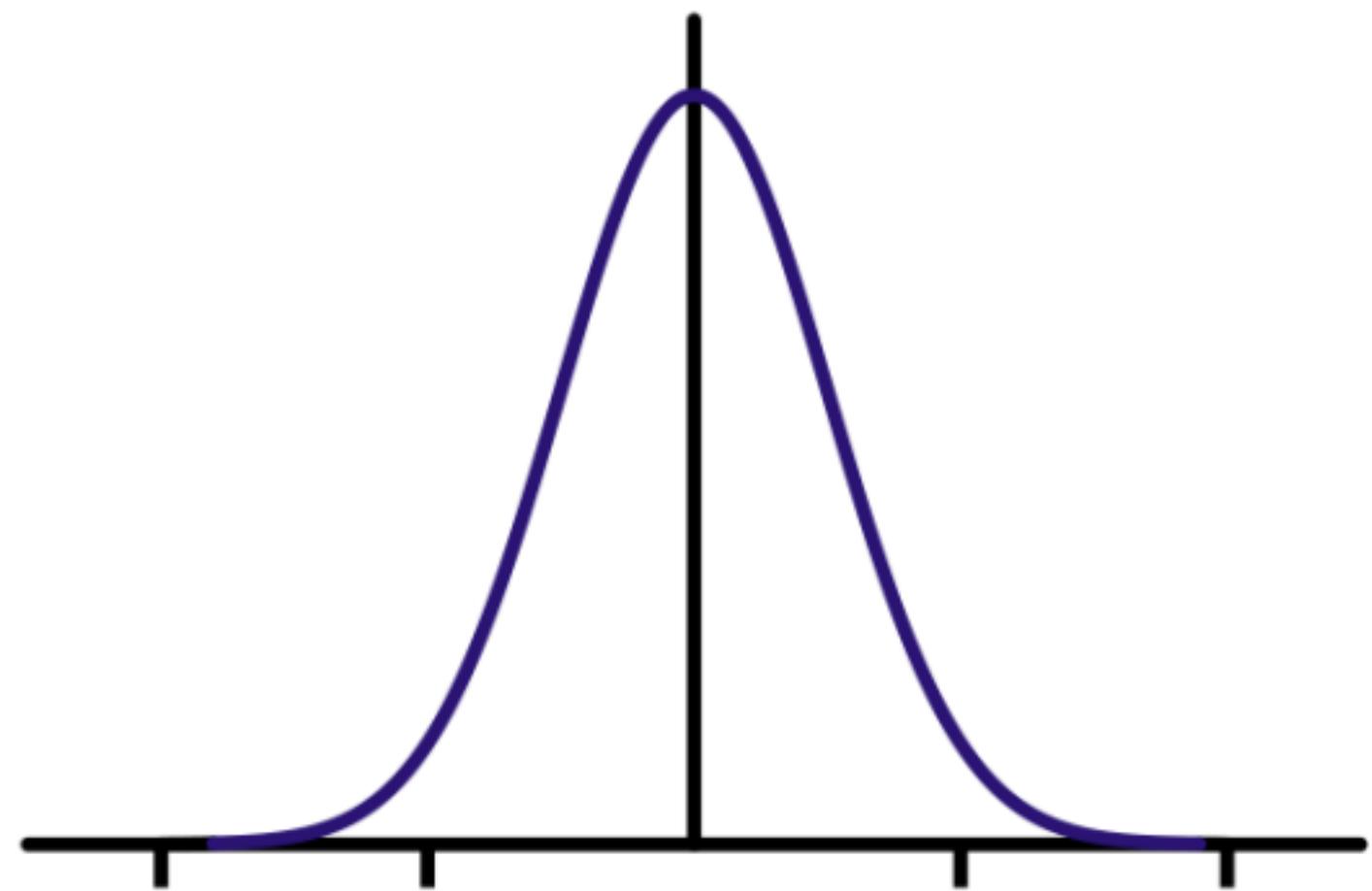
ماذا تمثل النتيجة س .. ؟

{ 8, 4, 2, 2 }

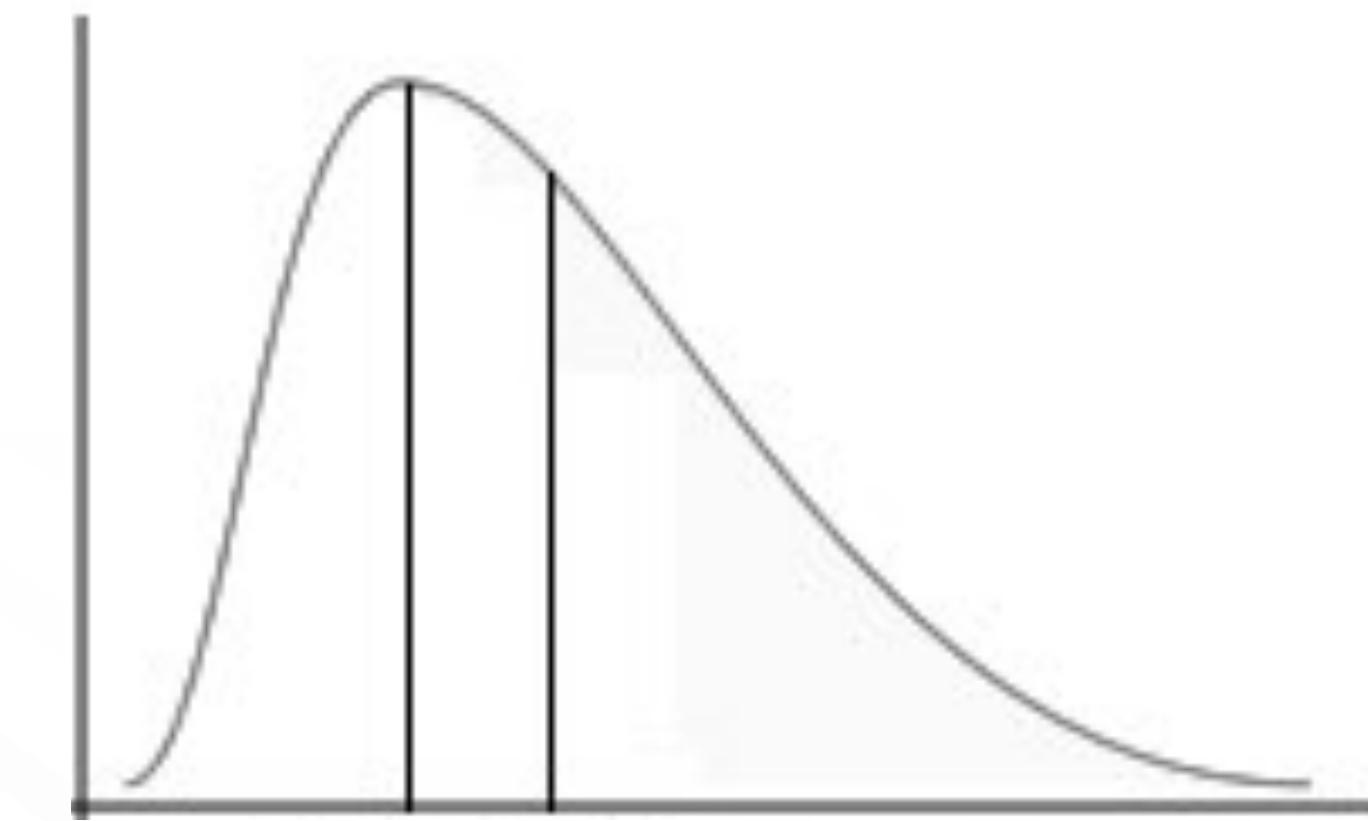
س = 3

الوسيط

التوزيع الطبيعي



التوزيع مائل لليمين





الإحصاء الوصفي - Descriptive Statistics -

الاستدلال الإحصائي - Statistical Inference -



فِوْمِ الْمَعَادِلَاتِ الرِّياضِيَّةِ

Mathematical content visible in the image:

- Circle: $A = \pi r^2$
- Circle: $C = 2\pi r$
- Cylinder: $V = \pi r^2 h$
- Cone: $V = \frac{1}{3} \pi r^2 h$
- Trigonometry (30°, 45°, 60°):

\sin	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$
\cos	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$
\tan	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$
- Right Triangle (30°, 60°, 90°):

$2x$	60°	x
45°	$x\sqrt{3}$	$\sqrt{2}$
- Integration:

$$\int \sin x dx = -\cos x + C$$

$$\int \frac{dx}{\cos^2 x} = \operatorname{tg} x + C$$

$$\int \operatorname{tg} x dx = -\ln|\cos x| + C$$

$$\int \frac{dx}{\sin x} = \ln\left|\operatorname{tg}\frac{x}{2}\right| + C$$

$$\int \frac{dx}{a^2 + x^2} = \frac{1}{a} \arctg \frac{x}{a} + C$$

$$\int \frac{dx}{x} = \frac{1}{a} \ln|x| + C$$
- Graph of $y = \operatorname{tg}(\theta)$ vs θ/rad .
- Quadratic Equations:

$$ax^2 + bx + c = 0$$

$$a(x^2 + \frac{b}{a}x + \frac{c}{a}) = 0$$

$$x^2 + \frac{b}{2a}x + (\frac{b}{2a})^2 - (\frac{b}{2a})^2 + \frac{c}{a} = 0$$

$$(x + \frac{b}{2a})^2 - \frac{b^2 - 4ac}{4a} = 0$$

المعادلات الرياضية تواجهنا في معظم - إن لم يكن جميع - تطبيقات علم البيانات

ولنفهم المعادلة ونقرأها نحتاج أن نعمل على تجربة فهم التمثيلات والأرقام

- اللوغاريتمات، الدوال الأُسّية، متعددة الحدود، الأعداد الكسرية
- مبرهنات الهندسة الرياضية مثل الزوايا المحيطية، المطابقات المثلثية
- الأعداد الحقيقية والمركبة
- المتسلسلات، المجاميع، المتباينات، المصفوفات

ماذا يمثل هذا الرقم .. ؟

2.718281828

الثابت الطبيعي e

ماذا يمثل هذا الرقم .. ؟

3.14159265359

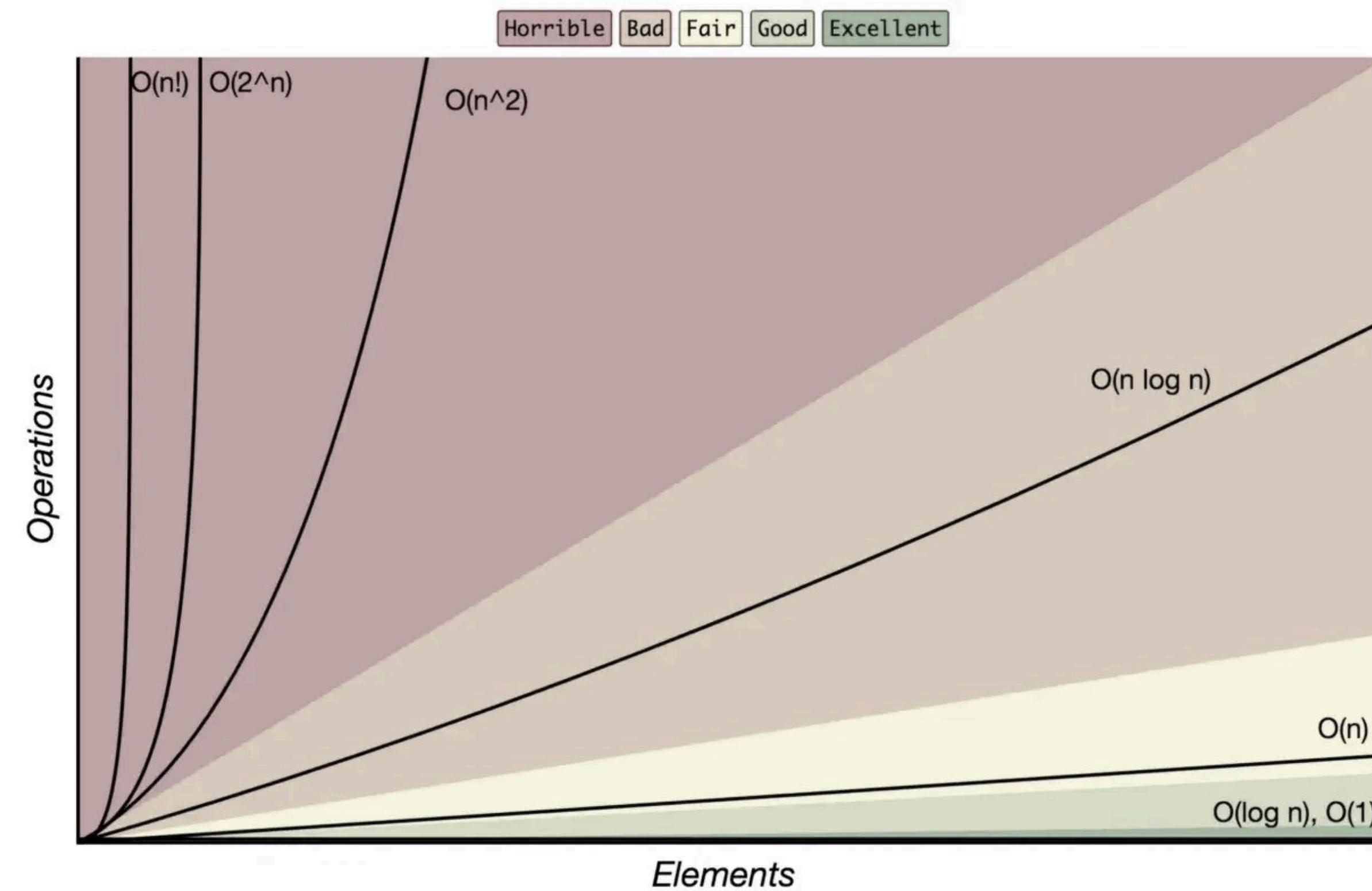
الثابت الدائرة π



خوارزميات تراكيب البيانات



الخوارزميات وسرعتها - Big O



تمثيل للتوزيع سرعة أداء أشهر الخوارزميات. المصدر: <http://bigocheatsheet.com>

الخوارزميات وسرعتها - Big O

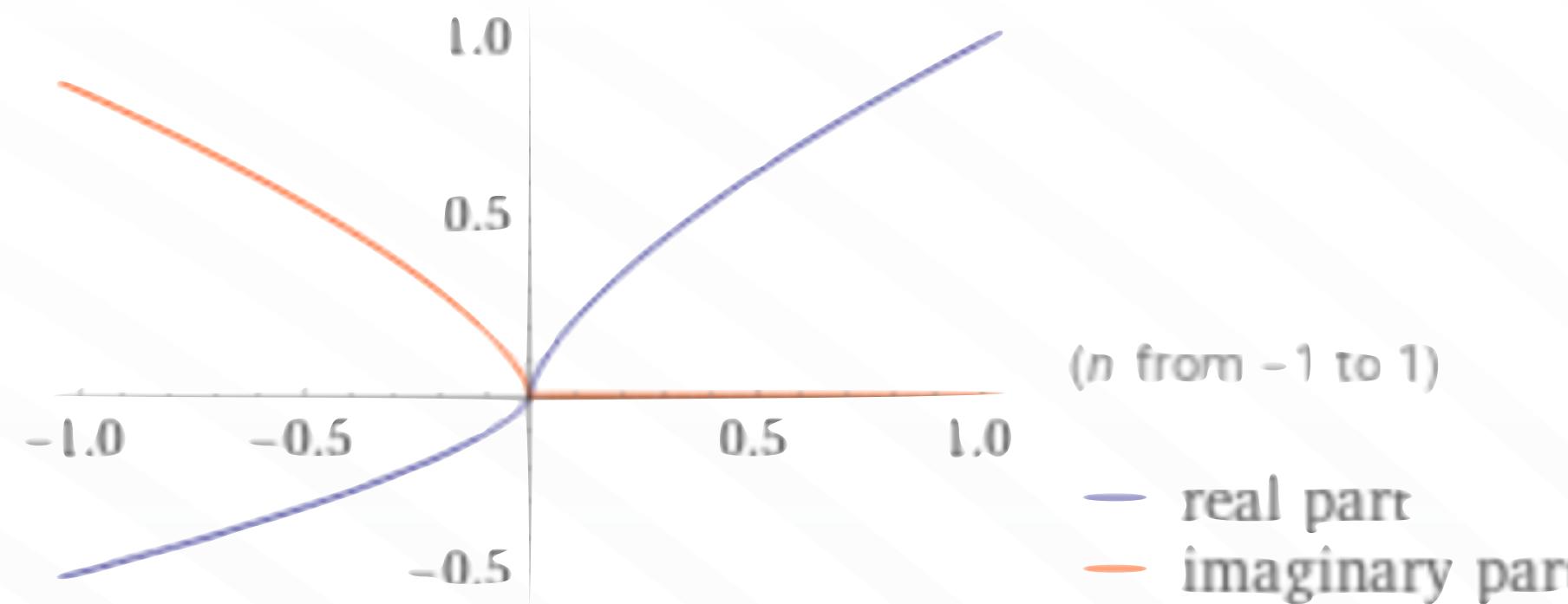
ما هو ترتيب سرعة الخوارزميات التالية:

1. $f_1(n) = O(n^{2/3})$
2. $f_2(n) = O(n \log(n))$
3. $f_3(n) = O(2^n)$
4. $f_4(n) = O(n^2 \log(n))$

الخوارزميات وسرعتها - Big O

ما هو ترتيب سرعة الخوارزميات التالية:

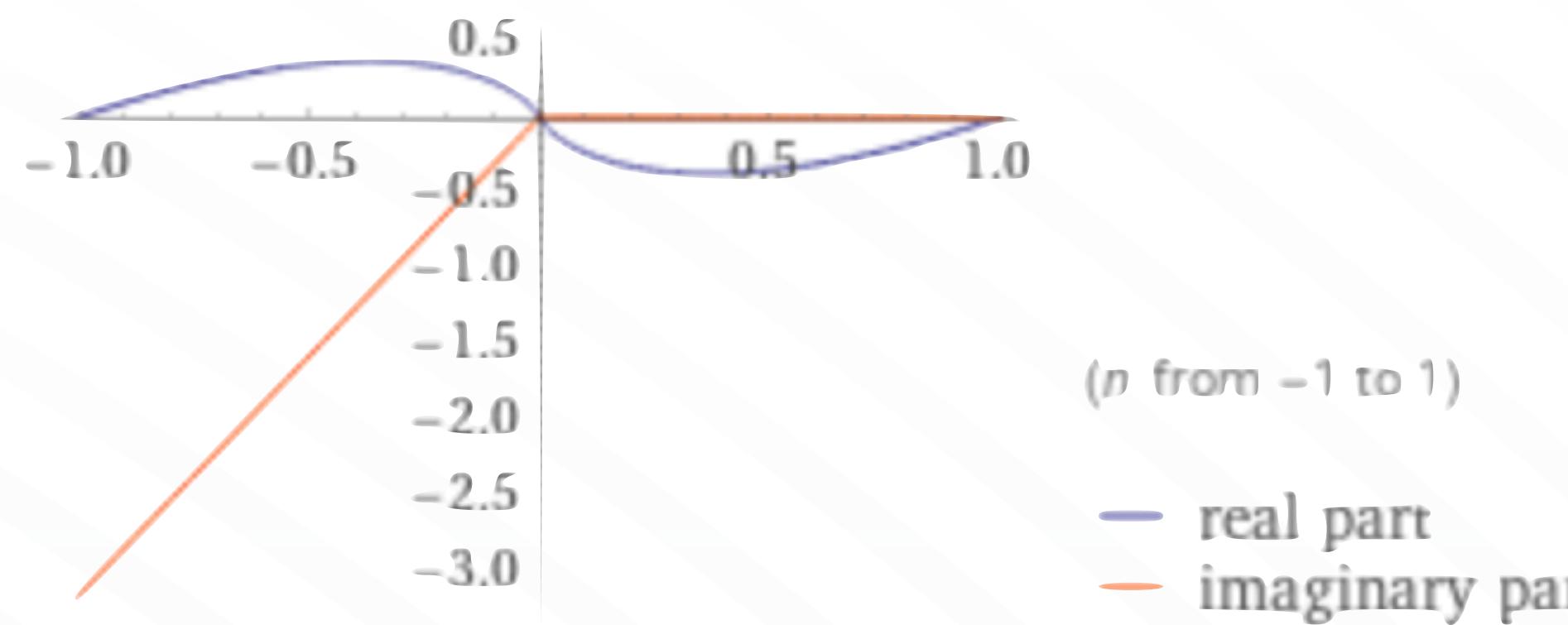
1. $f_1(n) = O(n^{2/3})$



الخوارزميات وسرعتها - Big O

ما هو ترتيب سرعة الخوارزميات التالية:

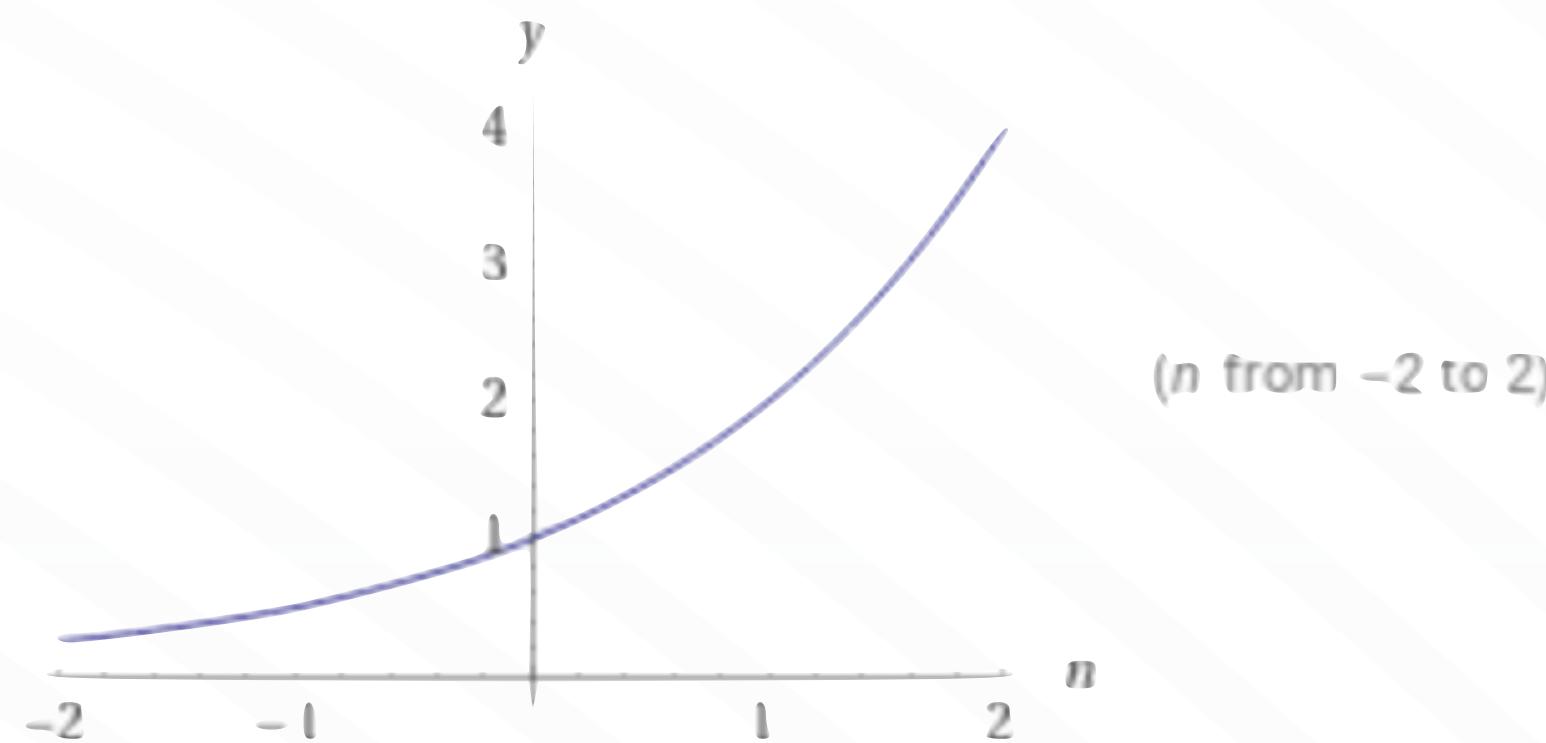
2. $f_2(n) = O(n \log(n))$



الخوارزميات وسرعتها - Big O

ما هو ترتيب سرعة الخوارزميات التالية:

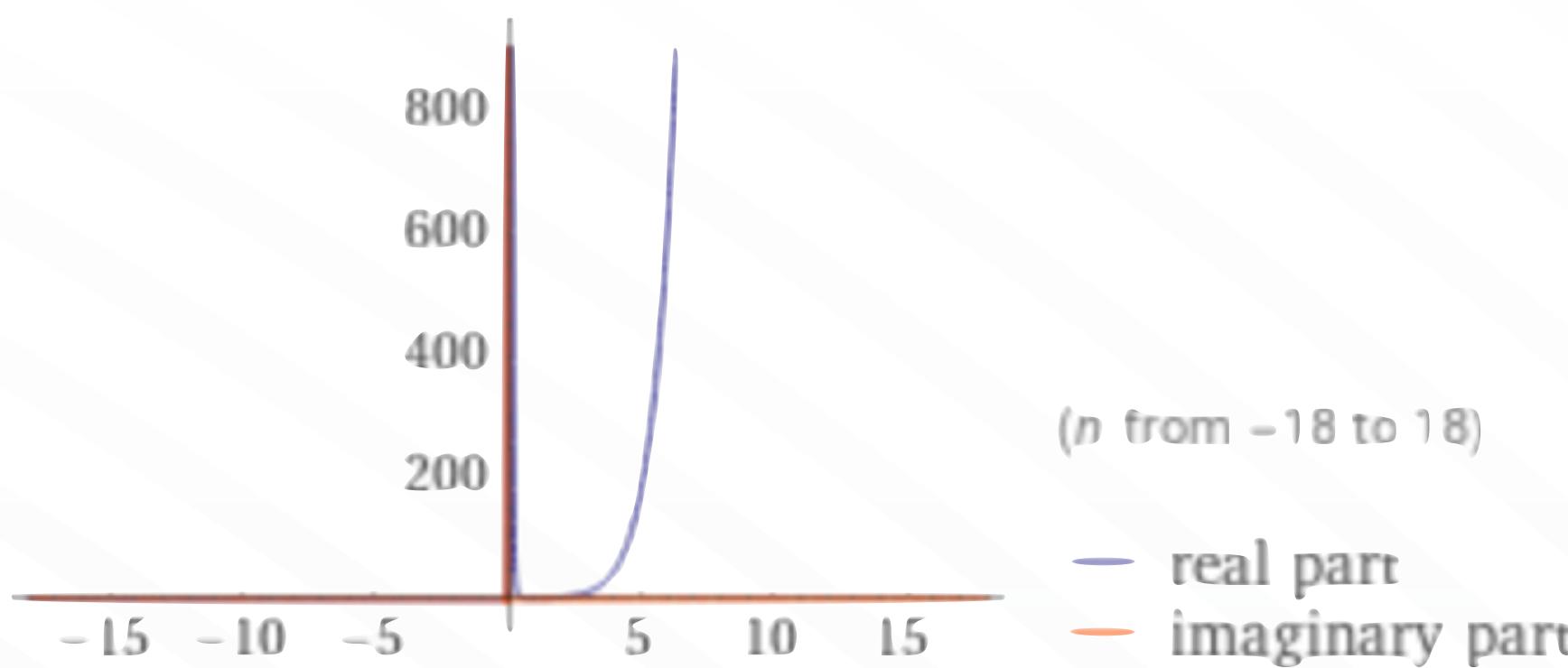
3. $f_3(n) = O(2n)$



الخوارزميات وسرعتها - Big O

ما هو ترتيب سرعة الخوارزميات التالية:

4. $f_4(n) = O(n^2 \log(n))$



أشهر الخوارزميات

أشهر خوارزميات البحث (ما هي سرعتهم؟):

1. البحث البسيط - Simple Search

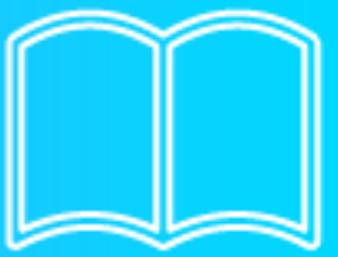
2. البحث الثنائي - Binary Search

أشهر الخوارزميات

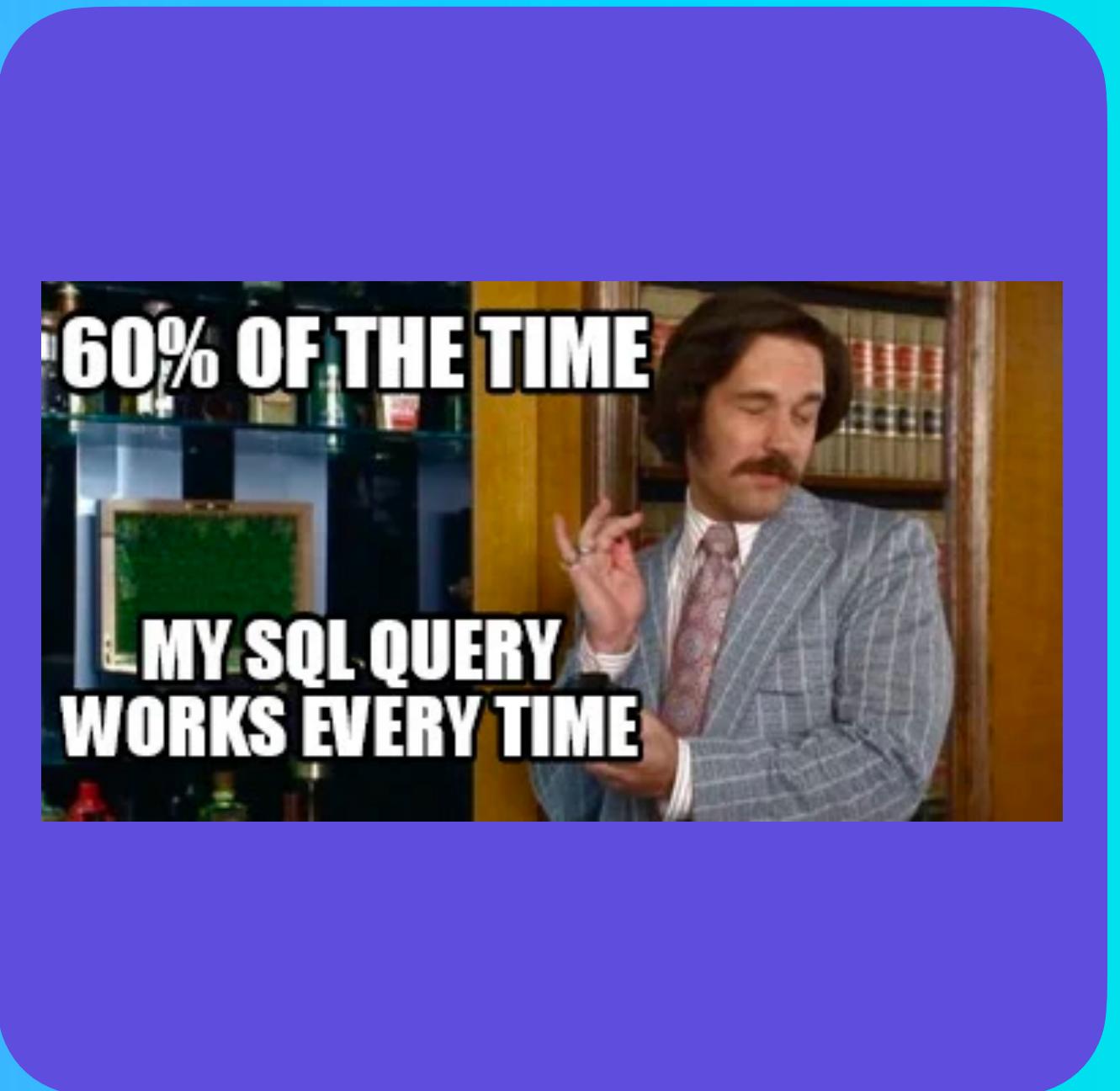
أشهر خوارزميات الترتيب (ما هي سرعتهم؟):

1. الترتيب الانتقائي - Selection Sort

2. الترتيب السريع - Quick Sort



SQL



SQL? NoSQL



أساسيات SQL

ترتيب تنفيذ الاستعلام في البيانات المهيكلة باستخدام لغة SQL:

1. اختيار ودمج الجداول From & Join
2. التصفية والشروط Where
3. التجميع Group by
4. تصفية التجمعات Having
5. اختيار الأعمدة للعرض Select
6. الترتيب Order By
7. حد عدد الأسطر للعرض Limit

البحث في النصوص SQL

التعامل مع النصوص وتركيزها:

- فمثلاً: للبحث عن السطور التي تحتوي على جزء من نص معين فتستخدم العلامة المئوية .1. بحث عن كلمة **طويق** في عمود academies للجدول bootcamps_providers يكون كالتالي:

```
SELECT *  
FROM bootcamps_providers bp  
WHERE bp.academies LIKE '%طويق%'
```

- 2. كيف نبحث عن كل الأكاديميات التي تنتهي بكلمة تقني ؟

شروط القوائم SQL

التعامل مع النصوص وتركيزها يختلف بين التقنيات اختلف طفيف:

1. البحث عن قائمة الأكاديميات في السعودية في جدول bootcamps_providers مع العلم أن بيانات المدن يتم تحديثها باستمرار بقاعدة البيانات مع الإبقاء على تاريخ التغيير يكون كالتالي:

```
SELECT *
FROM bootcamps_providers bp
WHERE bp.location IN (SELECT DISTINCT cities FROM ksa)
```

تقسيم النتائج للحصول على القيمة الأعلى

في الجداول التي تبقى على تاريخ التغيير، مع معرف للسجلات يكون غالباً التكرار فيه وارد جدًا:

1. تساعدنا هذه الطريقة في استرجاع القيمة ذات الأهمية لنا مثلاً آخر تحديث على سطر معين:

```
function( . . . ) OVER (  
    PARTITION BY . . .  
    ORDER BY . . .  
    ROWS BETWEEN . . . AND . . .  
)
```

تقسيم النتائج للحصول على القيمة الأعلى

في الجداول التي تبقى على تاريخ التغيير، مع معرف للسجلات يكون غالباً التكرار فيه وارد جداً:

1. تساعدنا هذه الطريقة في استرجاع القيمة ذات الأهمية لنا مثلاً آخر تحديث على سطر معين:

```
SELECT diagnosis_code, description_txt, effective_date, expire_date
FROM (
    SELECT diagnosis_code, description_txt, effective_date, expire_date,
           ROW_NUMBER() OVER (PARTITION BY diagnosis_code
                               ORDER BY effective_date DESC) rn
      FROM diagnosis_code
     WHERE expire_date IS NOT NULL
)
WHERE rn = 1
```



مهارات التواصل والتفسير

A black and white portrait of Albert Einstein on the right side of the slide. He has his characteristic wild hair and a gentle smile. To his left is a dark rectangular box containing a quote by him. The quote reads:
**EVERYBODY IS A GENIUS.
BUT IF YOU JUDGE A FISH BY ITS ABILITY TO CLIMB A TREE, IT WILL LIVE ITS WHOLE LIFE BELIEVING THAT IT IS STUPID.**
ALBERT EINSTEIN
WWW.WELLNESSWORKS.IN



استفساراتكم؟