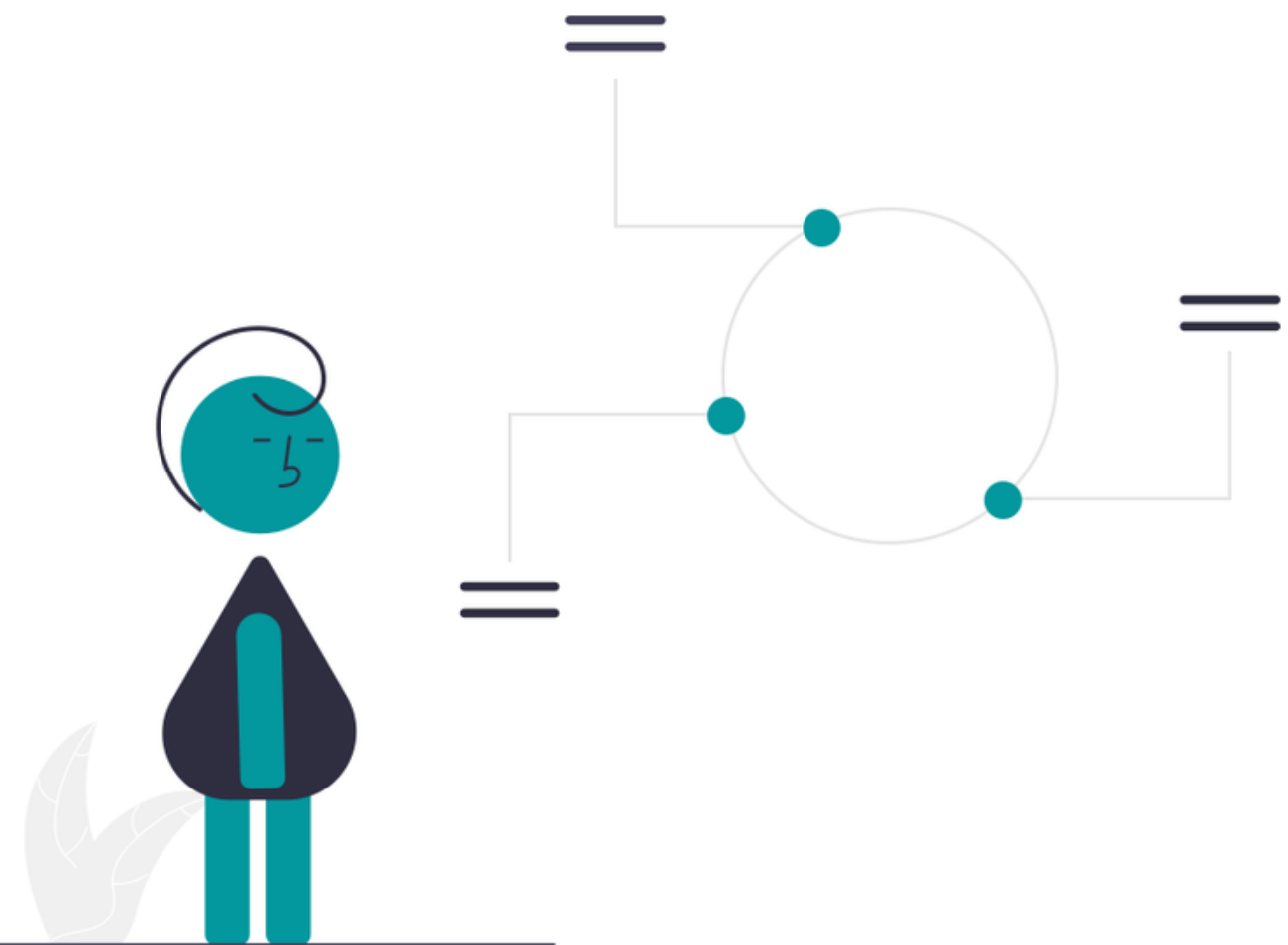


BIG DATA & ANALYTICS

TORTURE THE DATA

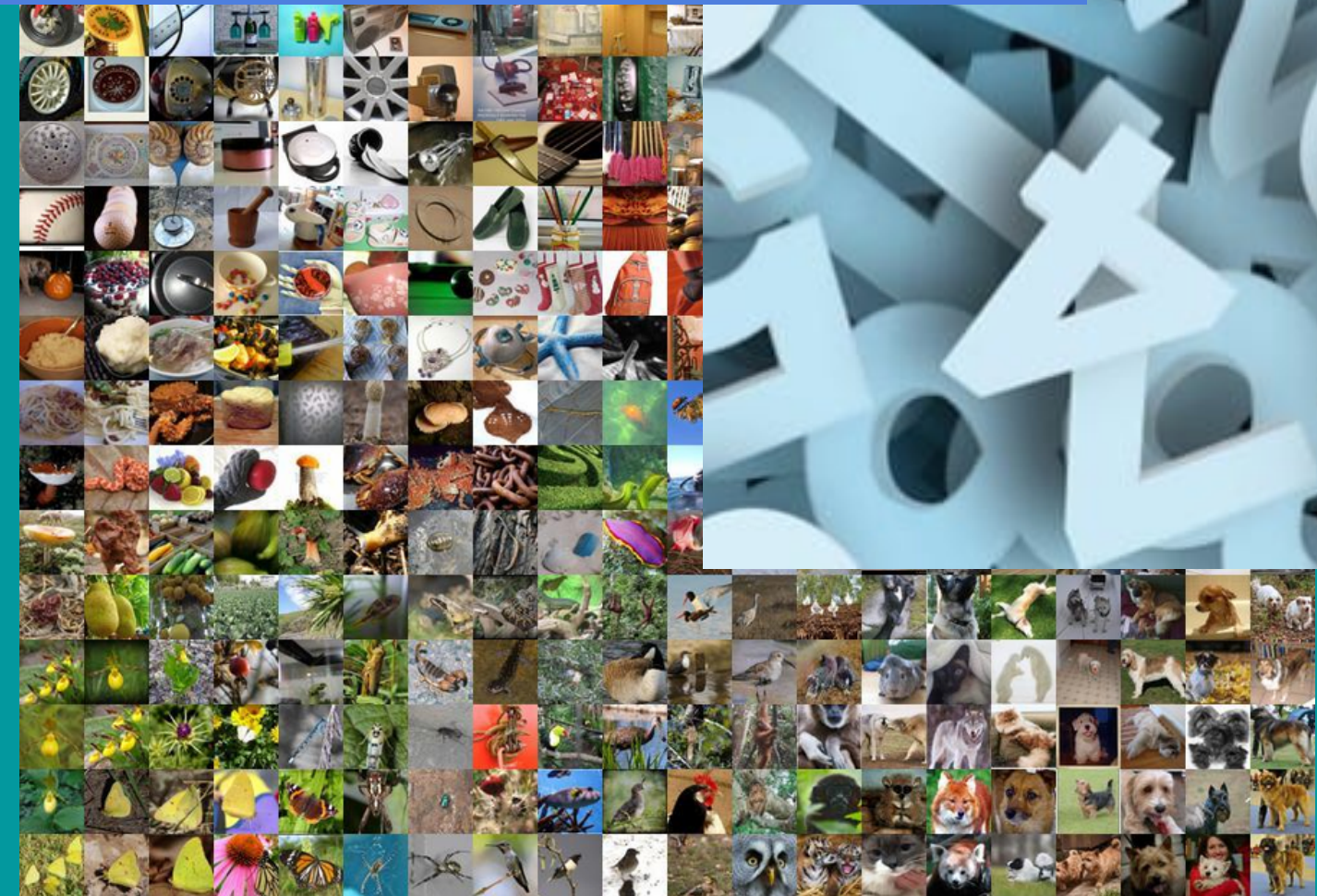
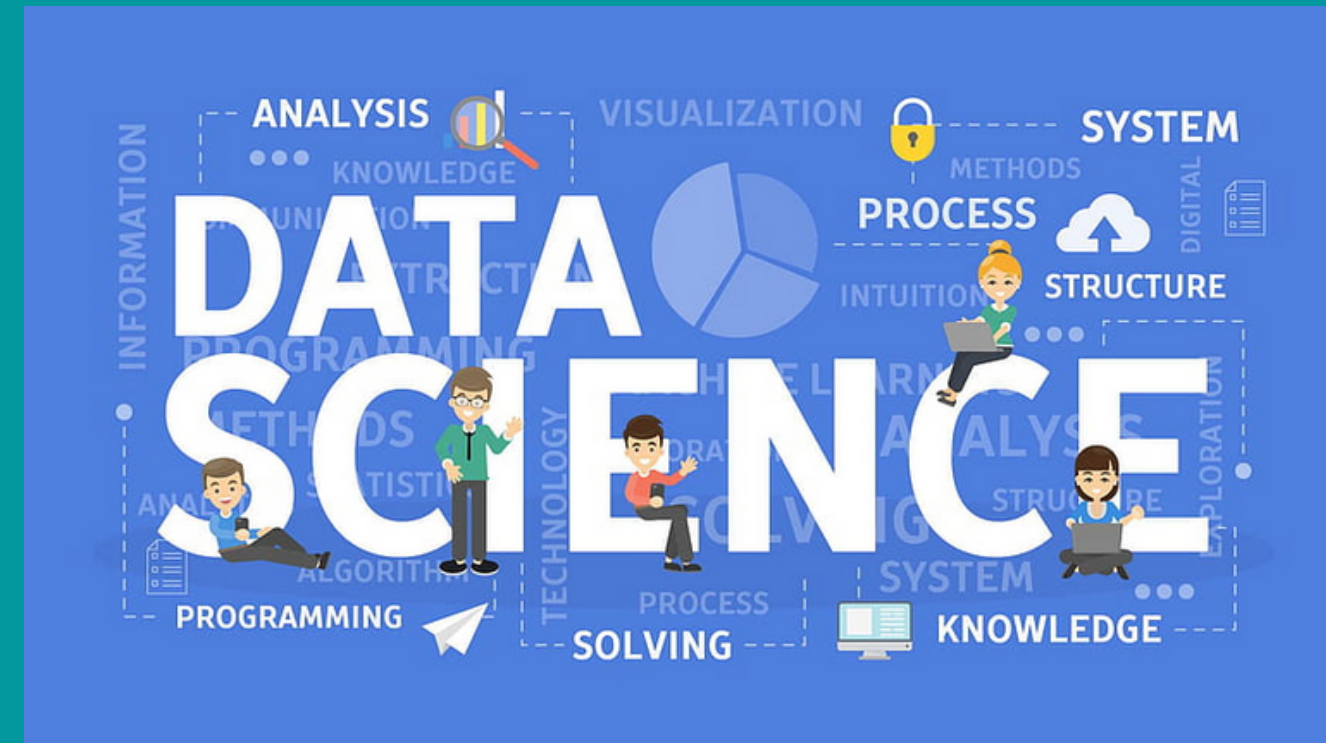
AND IT WILL CONFESS TO ANYTHING

Από τα δεδομένα στην
ανάλυση





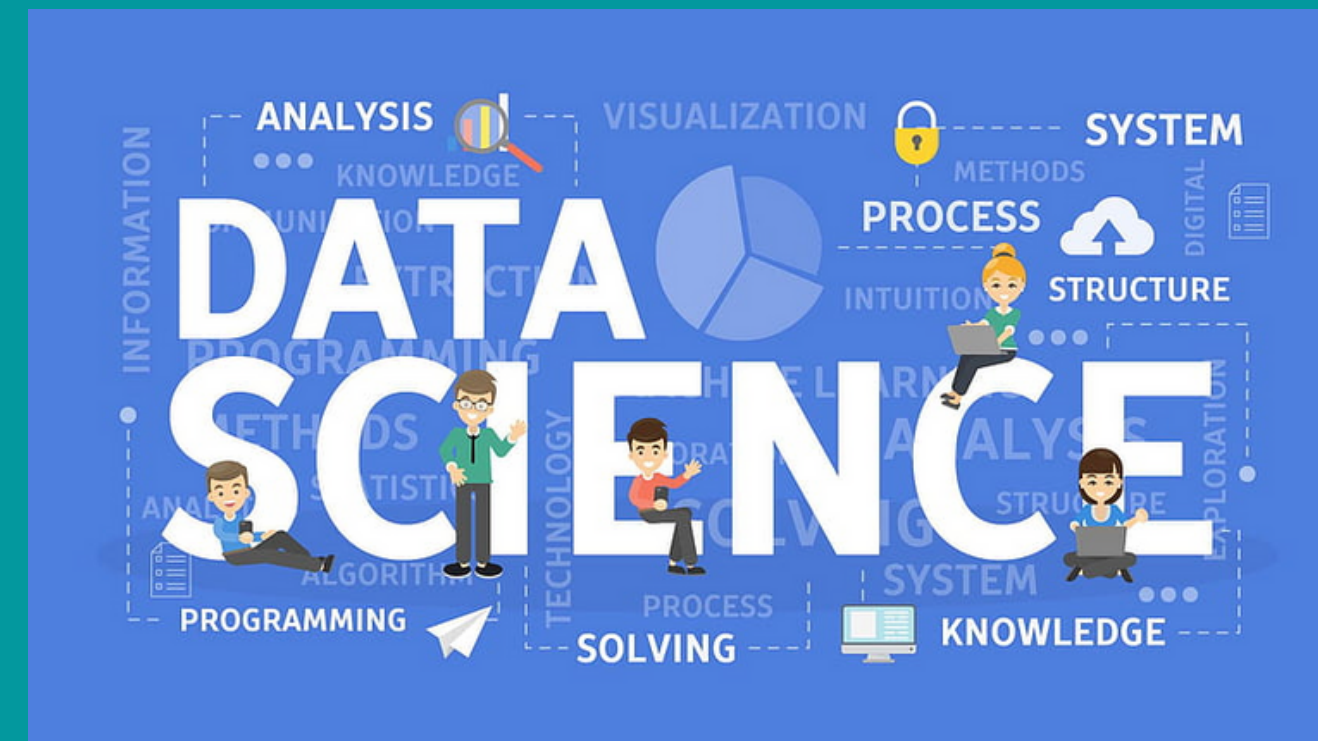
Τι είναι τα δεδομένα
(data)?





Τι είναι τα δεδομένα (data)?

Σύνολο **πληροφοριών** που συλλέγονται και μεταφράζονται για κάποιο σκοπό.





Ποιά είναι η διαφορά μεταξύ
ποιοτικών και **ποσοτικών**
δεδομένων?

Αναφέρετε ένα **ποιοτικό** και
ένα **ποσοτικό** δεδομένο για
έναν σκύλο.



Ποιά είναι η διαφορά μεταξύ
ποιοτικών και **ποσοτικών**
δεδομένων?

Κάποια δεδομένα είναι **ποιοτικά** (άρα περιγράφουν κάτι) ενώ άλλα είναι **ποσοτικά** (άρα μας δίνουν πληροφορία μέσω αριθμών)

π.χ "πέρασα ωραία στην εκδρομή" vs
"έχω 5 μήλα"



Ποιά είναι η διαφορά μεταξύ
ποιοτικών και **ποσοτικών**
δεδομένων?

Αναφέρετε ένα **ποιοτικό** και
ένα **ποσοτικό** δεδομένο για
έναν σκύλο.



Αναφέρετε ένα **ποιοτικό** και
ένα **ποσοτικό** δεδομένο για
έναν σκύλο

Είναι μαύρος
Έχει μακρύ τρίχωμα
Έχει πολλή ενέργεια

Vs

Έχει 4 πόδια
2 αδέρφια
Ζυγίζει 20 κιλά

Τύποι δεδομένων

1

Δομημένα

- **Εύκολο** να αποθηκευτούν, επεξεργαστούν και αναλυθούν
- **~5-10%** του συνόλου των δεδομένων



2

3

Τύποι δεδομένων

1

Δομημένα

- **Εύκολο** να αποθηκευτούν, επεξεργαστούν και αναλυθούν
- **~5-10%** του συνόλου των δεδομένων



2

3

Αδόμητα

- **Δύσκολο** να κατηγοριοποιηθούν
- **~80%** του συνόλου των δεδομένων



Τύποι δεδομένων

1

Δομημένα

- **Εύκολο** να αποθηκευτούν, επεξεργαστούν και αναλυθούν
- **~5-10%** του συνόλου των δεδομένων



2

Ημι-Δομημένα

- Μίξη των άλλων 2
- Έχουν **ικανότητες οργάνωσης** και είναι **ευκολότερο** να αναλυθούν



3

Αδόμητα

- **Δύσκολο** να κατηγοριοποιηθούν
- **~80%** του συνόλου των δεδομένων



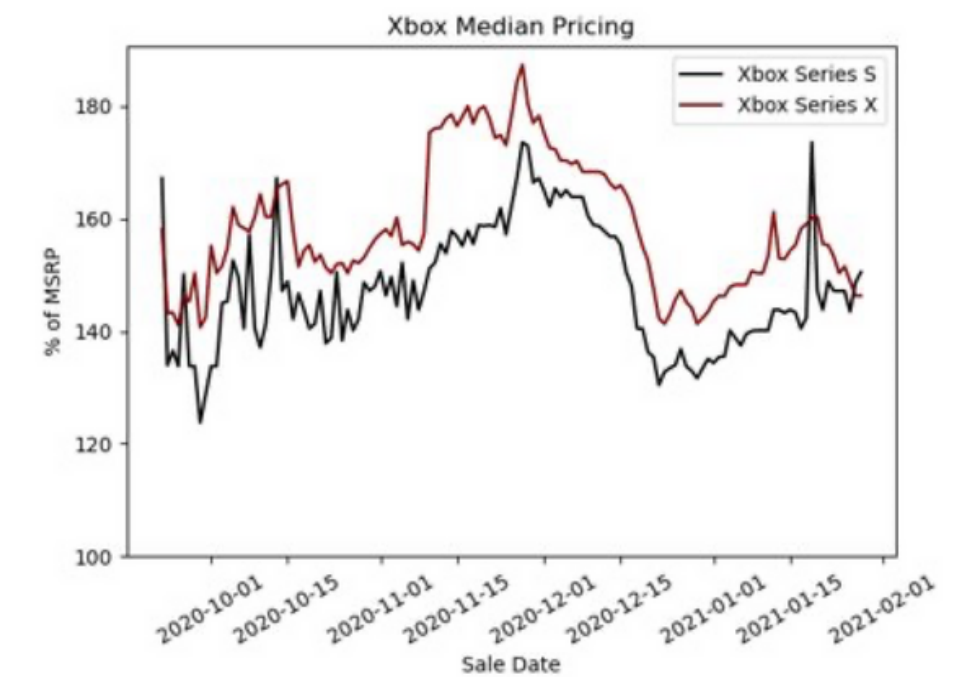
Παραδείγματα Τύπων Δεδομένων

Δομημένα

- Ώρα και Ημερομηνία
- Τηλεφωνικοί αριθμοί
- Πληροφορίες τραπεζικών συναλλαγών
- Ονόματα, διευθύνσεις και e-mail
- Τιμές προϊόντων

#	C1-D	C2-D	C3-D
	Call Received	Call Answered	Call Completed
1	Sep-07-2020 13:10:12	Sep-07-2020 13:10:47	Sep-07-2020 13:26:33
2	Sep-07-2020 15:33:36	Sep-07-2020 15:33:48	Sep-07-2020 15:43:00
3	Sep-07-2020 15:56:24	Sep-07-2020 15:56:59	Sep-07-2020 16:14:05
4	Sep-07-2020 16:05:00	Sep-07-2020 16:05:11	Sep-07-2020 16:10:34
5	Sep-07-2020 22:54:48	Sep-07-2020 22:55:03	Sep-07-2020 23:08:14
6	Sep-08-2020 0:24:24	Sep-08-2020 0:25:06	Sep-08-2020 0:42:12
7	Sep-08-2020 3:47:36	Sep-08-2020 3:48:18	Sep-08-2020 4:00:33
8	Sep-08-2020 6:02:48	Sep-08-2020 6:03:33	Sep-08-2020 6:34:45

	A	B	C
1	S No	Phone Numbers	
2	1	8046151300	
3	2	8130227245	
4	3	9899944310	
5	4	7987368321	
6	5	9457239975	
7	6	9205464773	
8	7	9818636072	
9			



- Συστήματα κρατήσεων σε ξενοδοχεία
- Λογισμικά καταγραφής πωλήσεων
- Ιατρικές συσκευές

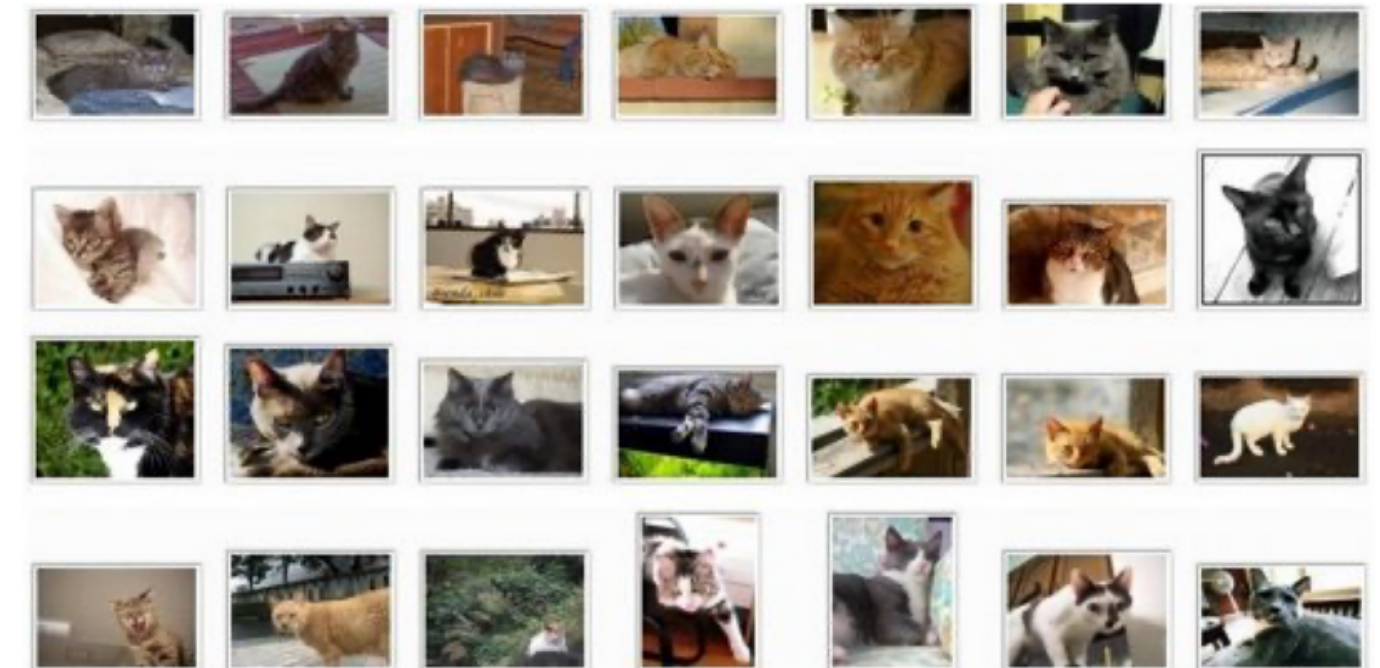
Παραδείγματα Τύπων Δεδομένων

Αδόμητα

- Αρχεία txt
- Μηνύματα σε e-mail
- Αρχεία ήχου και εικόνας
- Φωτογραφίες
- Καταγραφές από κάμερα
- Βιβλία
- Αξιολογήσεις προϊόντων



- Κοινωνικά Δίκτυα
- Ιστοσελίδες
- Λογισμικά επεξεργασίας κειμένου
- Παρουσιάσεις
- GPS, δορυφόροι
- Εφαρμογές μηνυμάτων



The PS5 is a genuine leap forward for console gaming, offering gorgeous 4K performance, stunningly fast load times and a truly game-changing controller that makes playing games more immersive and tactile than ever. It plays nearly all PS4 games, and, in many cases, allows them to run and load better than ever before. Nov 11, 2022

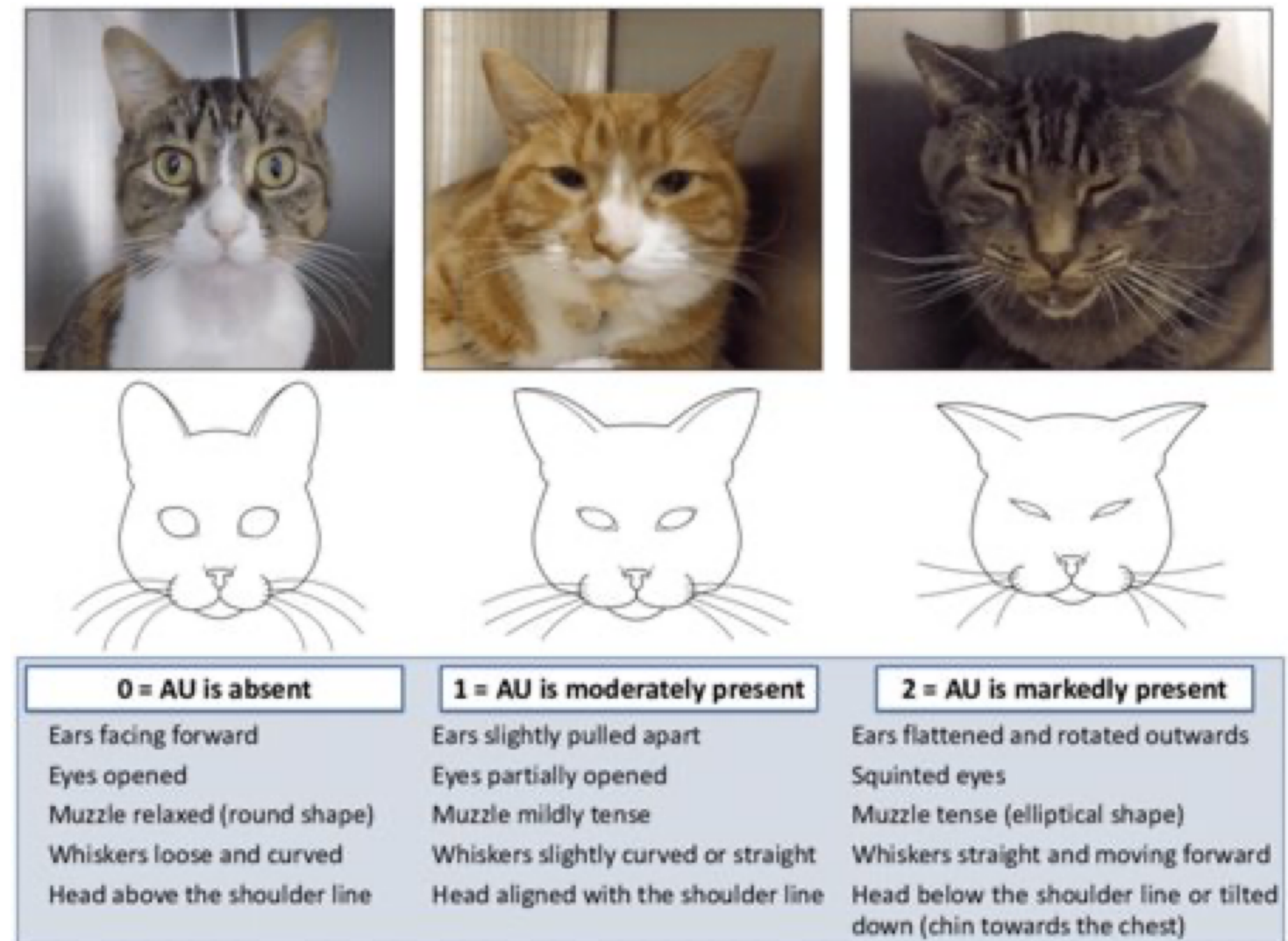
<https://www.tomsguide.com> > Reviews > Gaming

[PS5 review: The future of console gaming is here - Tom's Guide](#)

Παραδείγματα Τύπων Δεδομένων

Ημι-Δομημένα

- Ιστοσελίδα που έχει τίτλο, εικόνα, μικρή περιγραφή και βοηθά να διαφοροποιηθούν τα μέρη της
- Εικόνες στο διαδίκτυο που συνήθως συνοδεύονται από μία σύντομη λεκτική περιγραφή





Δομημένα

Θετικά

- Εύκολο να χρησιμοποιηθούν από **αλγορίθμους Μηχανικής Μάθησης**
- Εύκολο να χρησιμοποιηθούν από **χρήστες που δεν έχουν σε βάθος γνώσεις**
- Διαχειρίσιμα από **περισσότερα εργαλεία ανάλυσης**

- Περιορισμένη **χρήση**
- Περιορισμένες **επιλογές αποθήκευσης**

Αρνητικά





Αδόμητα

Θετικά

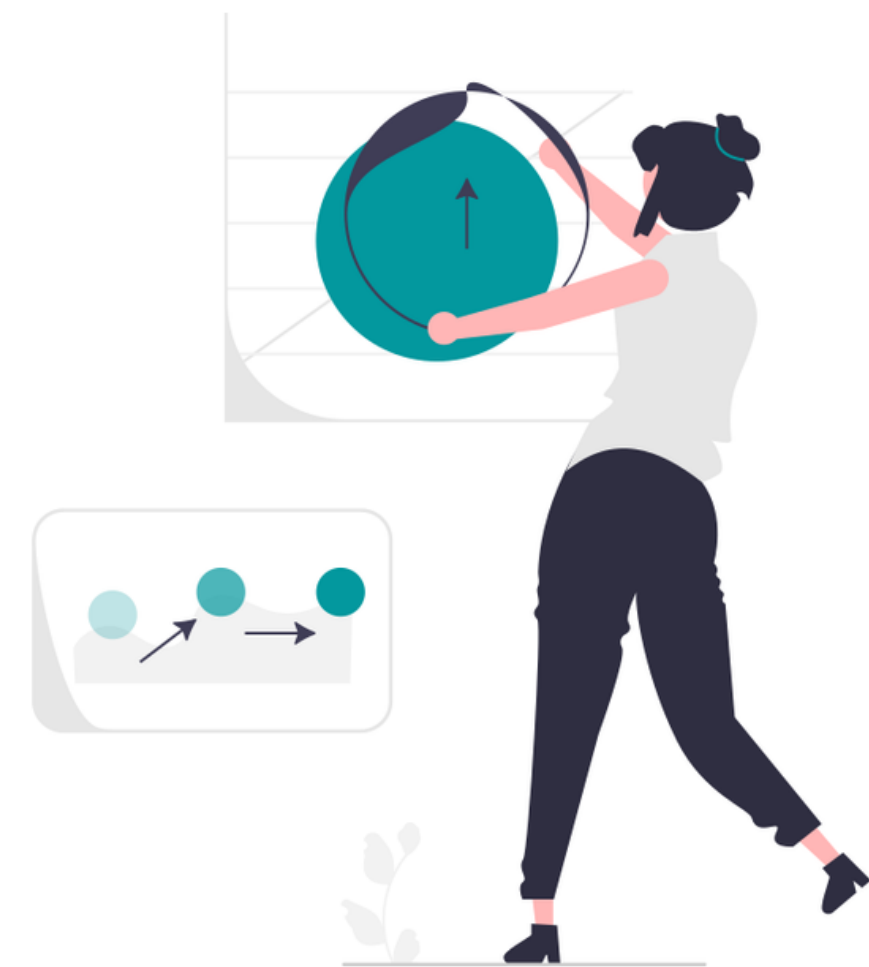
- Βρίσκονται στην **αρχική τους μορφή** και μπορούν να τροποποιηθούν από τον αναλυτή κατάλληλα
- **Γρηγορότεροι ρυθμοί** απόκτησης τους
- **Ευκολότερη και φθηνότερη αποθήκευση** μόνο όσων χρειάζονται

- Απαιτούνται **εξειδικευμένες γνώσεις**
- Απαιτούνται **ειδικά εργαλεία** ανάλυσης και χειρισμού τους

Αρνητικά



Τα Μεγάλα Δεδομένα





Τι είναι τα Μεγάλα
δεδομένα
(Big Data)?



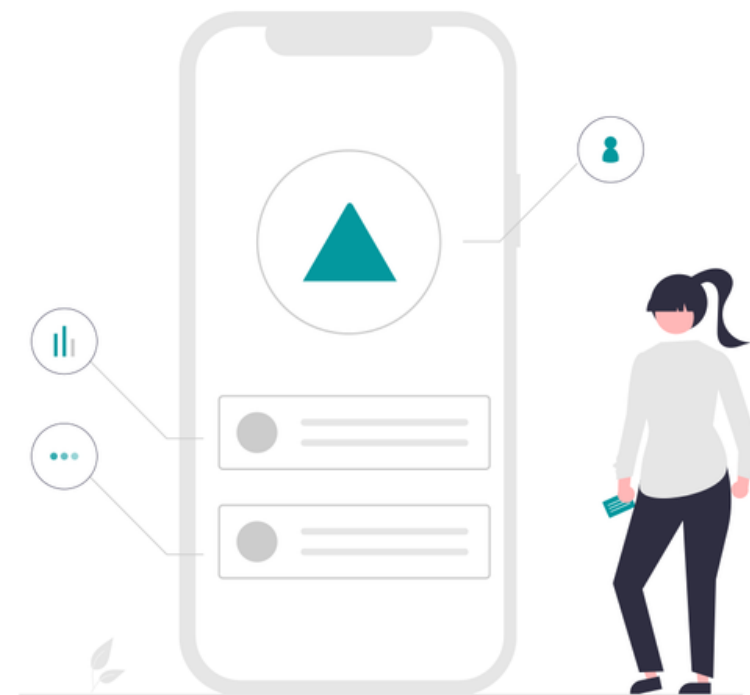
Τι είναι τα Μεγάλα δεδομένα (Big Data)?

Σύνολο δεδομένων που λόγω του πλήθους, της πολυπλοκότητας και της μεγάλης ταχύτητας παραγωγής τους είναι **δύσκολο** να επεξεργαστούν και να αναλυθούν με τις παραδοσιακές τεχνικές και εργαλεία.



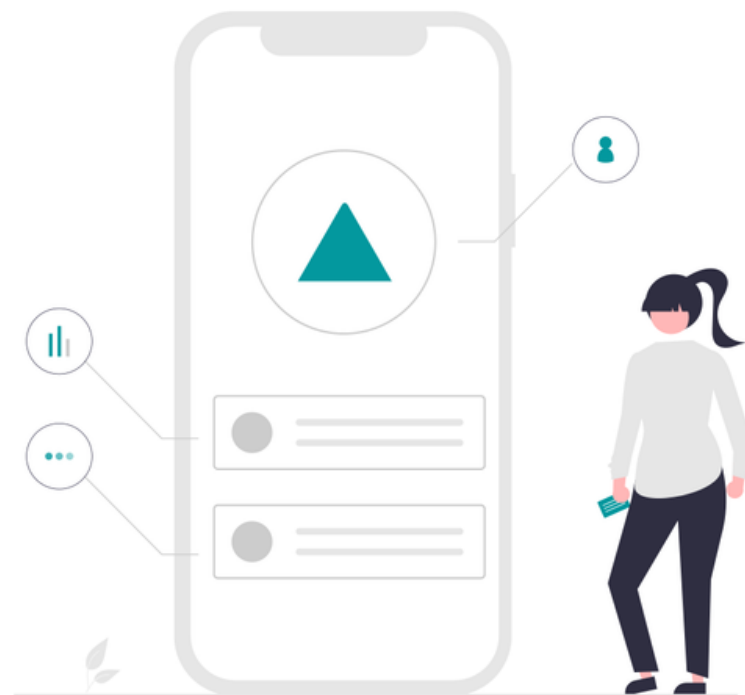


Γιατί όμως Μεγάλα
Δεδομένα?





Γιατί όμως Μεγάλα Δεδομένα?



Βοηθούν να βρίσκουμε:

- Κρυμμένα μοτίβα
- Κρυμμένες συσχετίσεις
- Αγοραστικές τάσεις
- Καταναλωτικές προτιμήσεις



Οδηγεί σε πιο καλά πληροφορημένες
αποφάσεις και στρατηγικές κινήσεις

Πως φτάσαμε στα Big Data?



Ιστορική Αναδρομή

2005

social media
αρχίζουν και
γίνονται δημοφιλή

2010

5 δισεκατομμύρια
κινητά =>
Συνειδητοποίηση
του τεράστιου
όγκου δεδομένων
που παράγονται

2015

Η Google και η
Microsoft
κατασκευάζουν
μεγάλα data centers

2021

Οι ταχύτερες
στα data
centers
ξεπερνούν τα
1000G

2022

Η Google πλέον
εξυπηρετεί 40.000
αναζητήσεις το
δευτερόλεπτο (3.5
δισ την ημέρα)

Πως φτάσαμε στα Big Data?

Fun
Fact



Facebook, Twitter,
LinkedIn, Instagram
κλπ παράγουν περί
τα **2.5 εκατομμύρια
Terrabytes** δεδομένα
ημερησίως



Ιστορική Αναδρομή

2005

social media
αρχίζουν και
γίνονται δημοφιλή

2010

5 δισεκατομμύρια
κινητά =>
Συνειδητοποίηση
του τεράστιου
όγκου δεδομένων
που παράγονται

2015

Η Google και η
Microsoft
κατασκευάζουν
μεγάλα data centers

2021

Οι ταχύτητες
στα data
centers
ξεπερνούν τα
1000G

2022

Η Google πλέον
εξυπηρετεί 40.000
αναζητήσεις το
δευτερόλεπτο (3.5
δισ την ημέρα)



Τι είναι τα Data Centers?



Τι είναι τα Data Centers?

Υποδομή μιας εταιρείας που χρησιμοποιείται για όλες τις IT επιχειρήσεις ενός οργανισμού, δηλαδή, για την αποθήκευση, επεξεργασία και ανάλυση των δεδομένων της.

Η ασφάλεια και η αξιοπιστία τους είναι πρωταρχικός στόχος του οργανισμού



Κάθε δευτερόλεπτο:

3 εκατομμύρια mails, εκ των οποίων το 67% είναι spam

Κάθε άνθρωπος παράγει 1.7MB

300 ώρες βίντεο ανεβαίνουν στο YouTube

Κάθε λεπτό:

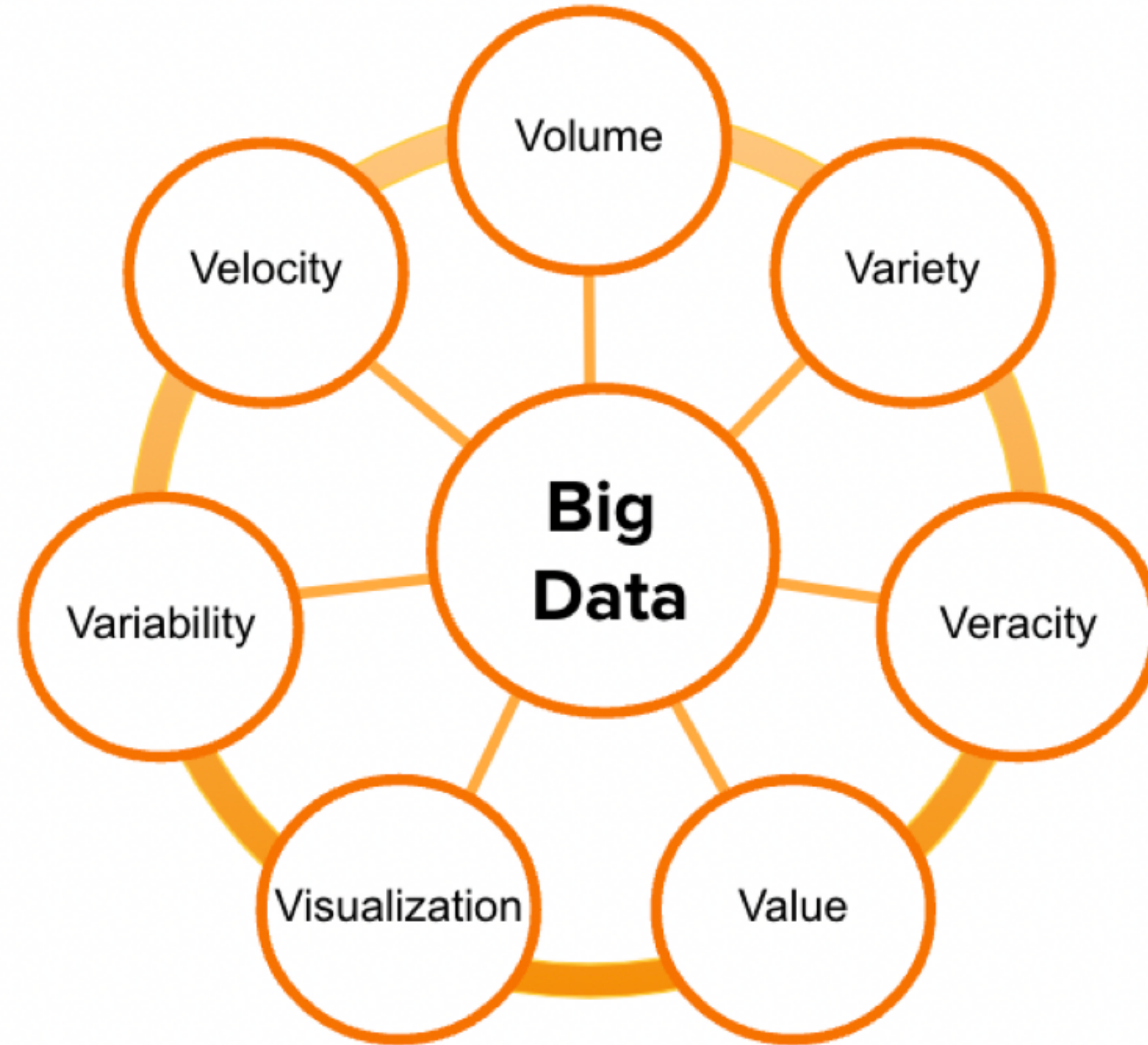
Γίνονται 1.4 εκατομμύρια κλήσεις

350.000 stories ανεβαίνουν στο Instagram

Βλέπουμε 400.000 ώρες Netflix

Η Amazon στέλνει 6.500 πακέτα

7 V'S OF BIG DATA



7 V's of Big Data

1

VOLUME

- Πόσα δεδομένα έχουμε
- Παλαιότερα σε Gigabytes (GB), πλέον σε Yottabytes (YB) ή και Zettabytes (ZB)
- Αναμένεται σημαντική αύξηση του όγκου των παραγόμενων δεδομένων

2

3

4

7 V's of Big Data

1

VOLUME

- Πόσα δεδομένα έχουμε
- Παλαιότερα σε Gigabytes (GB), πλέον σε Yottabytes (YB) ή και Zettabytes (ZB)
- Αναμένεται σημαντική αύξηση του όγκου των παραγόμενων δεδομένων

2

VELOCITY

- Πόσο γρήγορα επεξεργάζονται και γίνονται διαθέσιμα τα δεδομένα
- Σήμερα, αν δεν είναι real-time, θεωρείται αργά

3

4

7 V's of Big Data

1

VOLUME

- Πόσα δεδομένα έχουμε
- Παλαιότερα σε Gigabytes (GB), πλέον σε Yottabytes (YB) ή και Zettabytes (ZB)
- Αναμένεται σημαντική αύξηση του όγκου των παραγόμενων δεδομένων

2

VELOCITY

- Πόσο γρήγορα επεξεργάζονται και γίνονται διαθέσιμα τα δεδομένα
- Σήμερα, αν δεν είναι real-time, θεωρείται αργά

3

VARIETY

- Μία από τις μεγαλύτερες προκλήσεις
- Διάφοροι τύποι δεδομένων και διάφορες δομές
- Δυσκολία και σημασία στην οργάνωση τους

4

7 V's of Big Data

1

VOLUME

- Πόσα δεδομένα έχουμε
- Παλαιότερα σε Gigabytes (GB), πλέον σε Yottabytes (YB) ή και Zettabytes (ZB)
- Αναμένεται σημαντική αύξηση του όγκου των παραγόμενων δεδομένων

2

VELOCITY

- Πόσο γρήγορα επεξεργάζονται και γίνονται διαθέσιμα τα δεδομένα
- Σήμερα, αν δεν είναι real-time, θεωρείται αργά

3

VARIETY

- Μία από τις μεγαλύτερες προκλήσεις
- Διάφοροι τύποι δεδομένων και διάφορες δομές
- Δυσκολία και σημασία στην οργάνωση τους

4

VARIABILITY

- Διαφορετικό από την ποικιλία
- Ένα κατάστημα πώλησης καφέ έχει 6 διαφορετικές ποικιλίες. Αν κάθε μέρα παίρνεις την ίδια αλλά έχει διαφορετική γεύση, αυτό ονομάζεται μεταβλητότητα.
- Επηρεάζει την ομοιογένεια των δεδομένων.

7 V's of Big Data

5

VERACITY

- Εξασφαλίζει την ακρίβεια των δεδομένων
- Βοηθά να αντιμετωπιστεί το "garbage in, garbage out" πρόβλημα

6

7

7 V's of Big Data

5

VERACITY

- Εξασφαλίζει την ακρίβεια των δεδομένων
- Βοηθά να αντιμετωπιστεί το "garbage in, garbage out" πρόβλημα

6

VISUALIZATION

- Γραφικές παραστάσεις και διαγράμματα που βοηθούν περισσότερο από αναφορές γεμάτες αριθμούς

7

7 V's of Big Data

5

VERACITY

- Εξασφαλίζει την ακρίβεια των δεδομένων
- Βοηθά να αντιμετωπιστεί το "garbage in, garbage out" πρόβλημα

6

VISUALIZATION

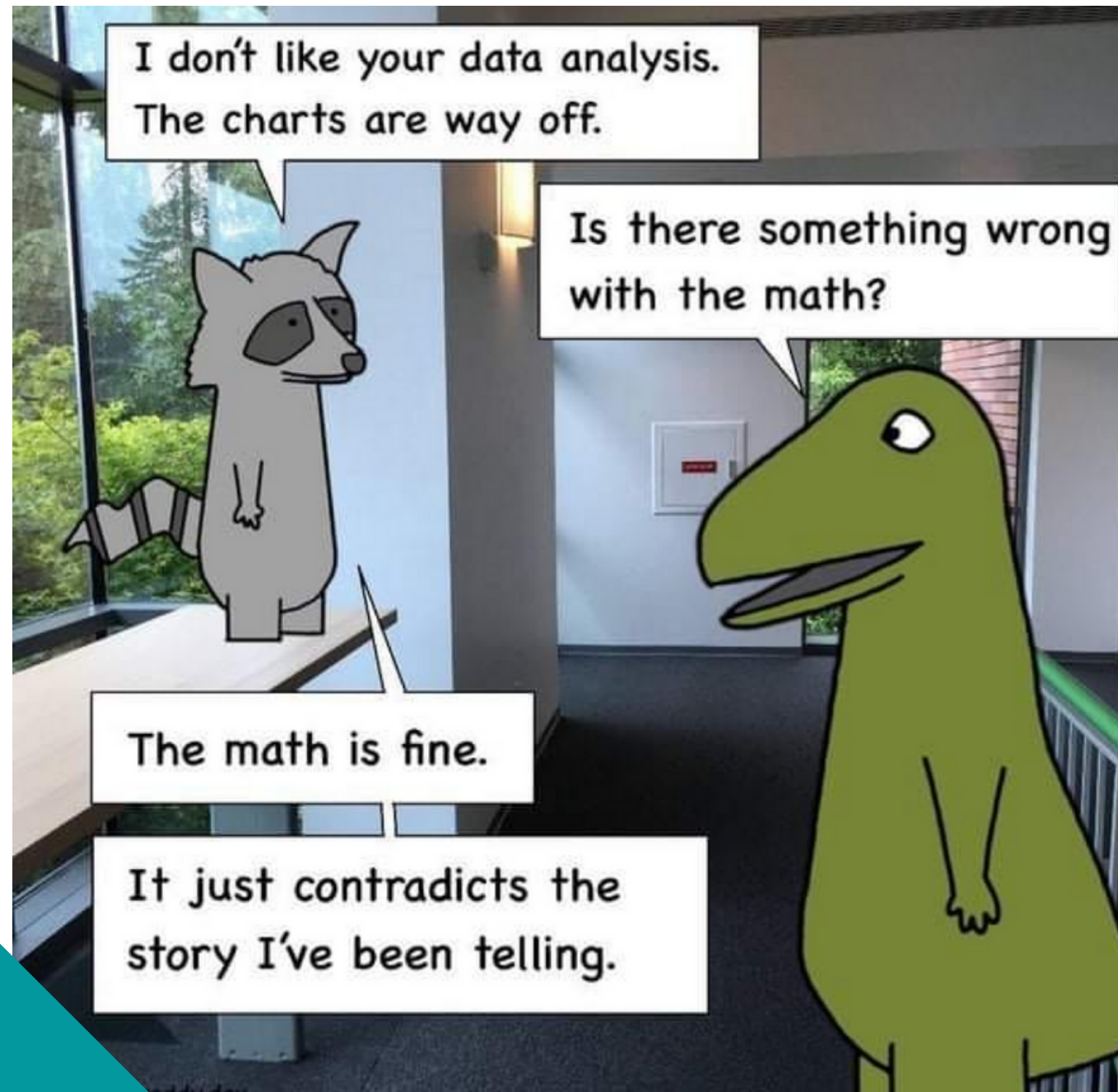
- Γραφικές παραστάσεις και διαγράμματα που βοηθούν περισσότερο από αναφορές γεμάτες αριθμούς

7

VALUE

- Αυτοσκοπός και τελικός στόχος
- Αφού γίνουν όλα τα παραπάνω, πρέπει να απορρέει κάποια αξία/κάποιο κέρδος από την επεξεργασία δεδομένων.

Επεξεργασία Μεγάλων Δεδομένων






Βήμα 1ο – Συλλογή

- **Συλλογή** δεδομένων από διάφορες πηγές
- **Αφαίρεση** λανθασμένων δεδομένων
- Σωστές ετικέτες και **κατηγοριοποίηση**
- Βασικό βήμα για την μετέπειτα σωστή επεξεργασία





Βήμα 2ο – Μετατροπή

- **Αλλαγή της μορφής** των δεδομένων π.χ. clustering
- **Κανονικοποίηση**
- **Μετατροπή** από αδόμητα σε δομημένα





Βήμα 3ο – φόρτωση

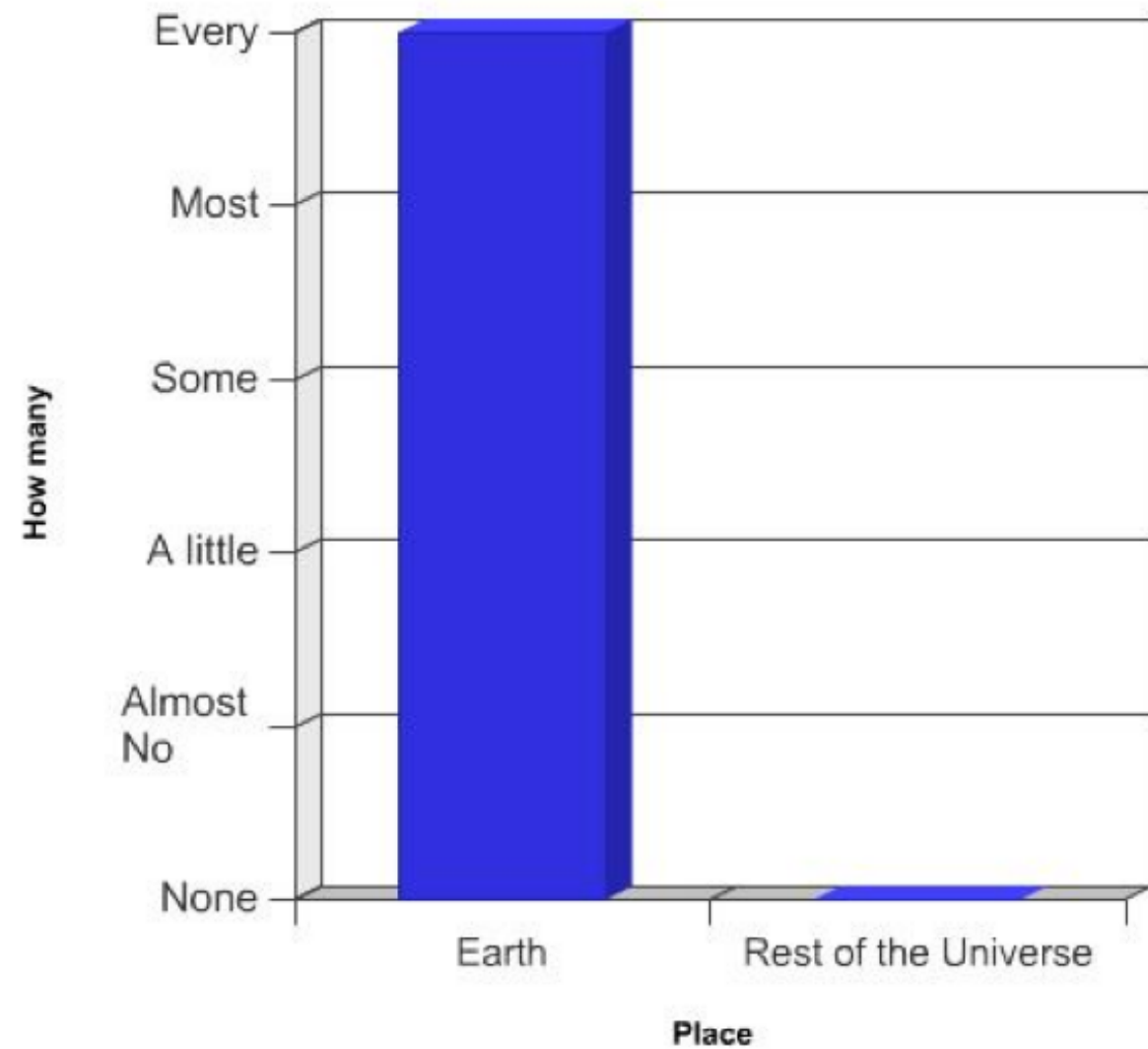
"**Ανέβασμα**" των δεδομένων στην κεντρική βάση δεδομένων



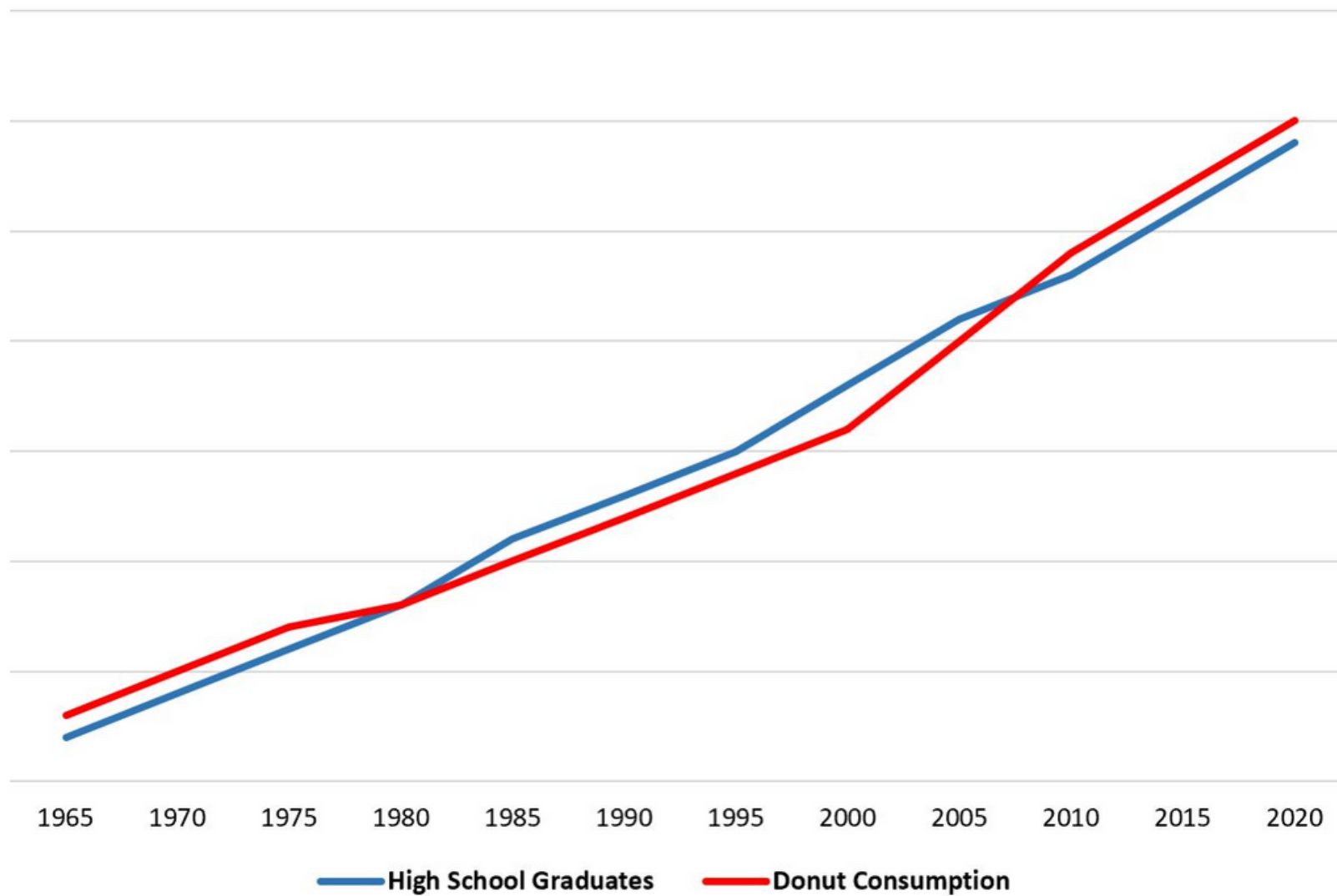


Βήμα 4ο – Οπτικοποίηση και Ανάλυση

Winners of Miss Universe



High School Graduates vs. Donut Consumption





Βήμα 5ο – Εφαρμογή Μηχανικής Μάθησης

- Δημιουργία **μοντέλων** που εξελίσσονται με νέες εισόδους
- **Μαθαίνουν** από τα δεδομένα
- Βρίσκει **μοτίβα** και κάνει προβλέψεις εξέλιξης χωρίς ανθρώπινη παρέμβαση



Η μηχανική μάθηση είναι ένα υποσύνολο της **τεχνητής νοημοσύνης (AI)**

Διδασκαλία υπολογιστών για να μαθαίνουν από τα δεδομένα και να **βελτιώνονται με την εμπειρία** – αντί να είναι ρητά προγραμματισμένοι να το κάνουν.





Ποιά είναι η "αγαπημένη"
εφαρμογή όσων ασχολούνται
με Machine Learning?

Η Μέθοδος των 5 "Γιατί"

?

- Εργαλείο ανάλυσης της **αιτίας ενός γεγονότος**
- Βοηθά να βρεθεί η αιτία ενός προβλήματος κάνοντας την ερώτηση "**Γιατί?**" αρκετές φορές
- Δημιουργείται μία **αλυσίδα αιτίας** – αποτελέσματος που μας οδηγεί στην αρχική αιτία
- Αναπτύχθηκε από τον Sakichi Toyoda, ιδρυτή της **Toyota** και αποτελεί ακόμα και σήμερα την **βάση της επιστημονικής της προσέγγισης**

Problem	There is a puddle of water on the floor.
Why?	The overhead pipe is leaking.
Why?	There is too much water pressure in the pipe.
Why?	There is a faulty control valve.
Why?	Control valves have not been tested.
Why?	Control valves are not on the maintenance schedule.

Structured Query Language

- Γλώσσα προγραμματισμού για την διαχείριση δεδομένων σε ένα σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων
- Η γλώσσα περιλαμβάνει δυνατότητες ανάκτησης και ενημέρωσης δεδομένων, δημιουργίας και τροποποίησης σχημάτων και σχεσιακών πινάκων, αλλά και ελέγχου πρόσβασης στα δεδομένα.
- Η SQL αναπτύχθηκε στην **IBM** από τους Andrew Richardson, Donald C. Messerly και Raymond F. Boyce, στις αρχές της δεκαετίας του 1970.

Εισαγωγή στην SQL



SQLITEONLINE

SQLite

0.0.4 beta

Table

demo

MariaDB

PostgreSQL

MS SQL

SQLite

```
1 SELECT * FROM demo;
```

ID	Name	Hint
1	https://SQL.BanD	The most secure, fast, efficient
2	SQL Online	Next gen SQL editor
3	Kirill N.	https://www.linkedin.com/in/sqli
4	Twitter	https://twitter.com/SqliteOnlineC
5	Donate (ERC20: ETH or USDC)	0xCcc227E5615D4FADd75822
6	Donate (BTC)	bc1q25zqmgll2fz0tyduusyfageh

BUY

ADS

SQLITEONLINE

File Owner DB Run Export **Import**

SQLite
0.0.4 beta

Table
demo

MariaDB

PostgreSQL

MS SQL

```
1 SELECT * FROM demo;
```

ID	Name	Hint
1	https://SQL.BanD	The most secure, fast, efficient
2	SQL Online	Next gen SQL editor
3	Kirill N.	https://www.linkedin.com/in/sqli
4	Twitter	https://twitter.com/SqliteOnlineC
5	Donate (ERC20: ETH or USDC)	0xCcc227E5615D4FADd75822
6	Donate (BTC)	bc1q25zqmgll2fz0tyduusyfageh

- Πατάμε το εικονίδιο **Import**

SQLITEONLINE

File Owner DB Run Export Import

SQLite
0.0.4 beta
Table
demo
MariaDB
PostgreSQL
MS SQL

BUY

ADS

Import

Book.csv

File **Open**

Type CSV

Table name Book

Delimiter ,

Escape "

Column name New-auto

Command Run

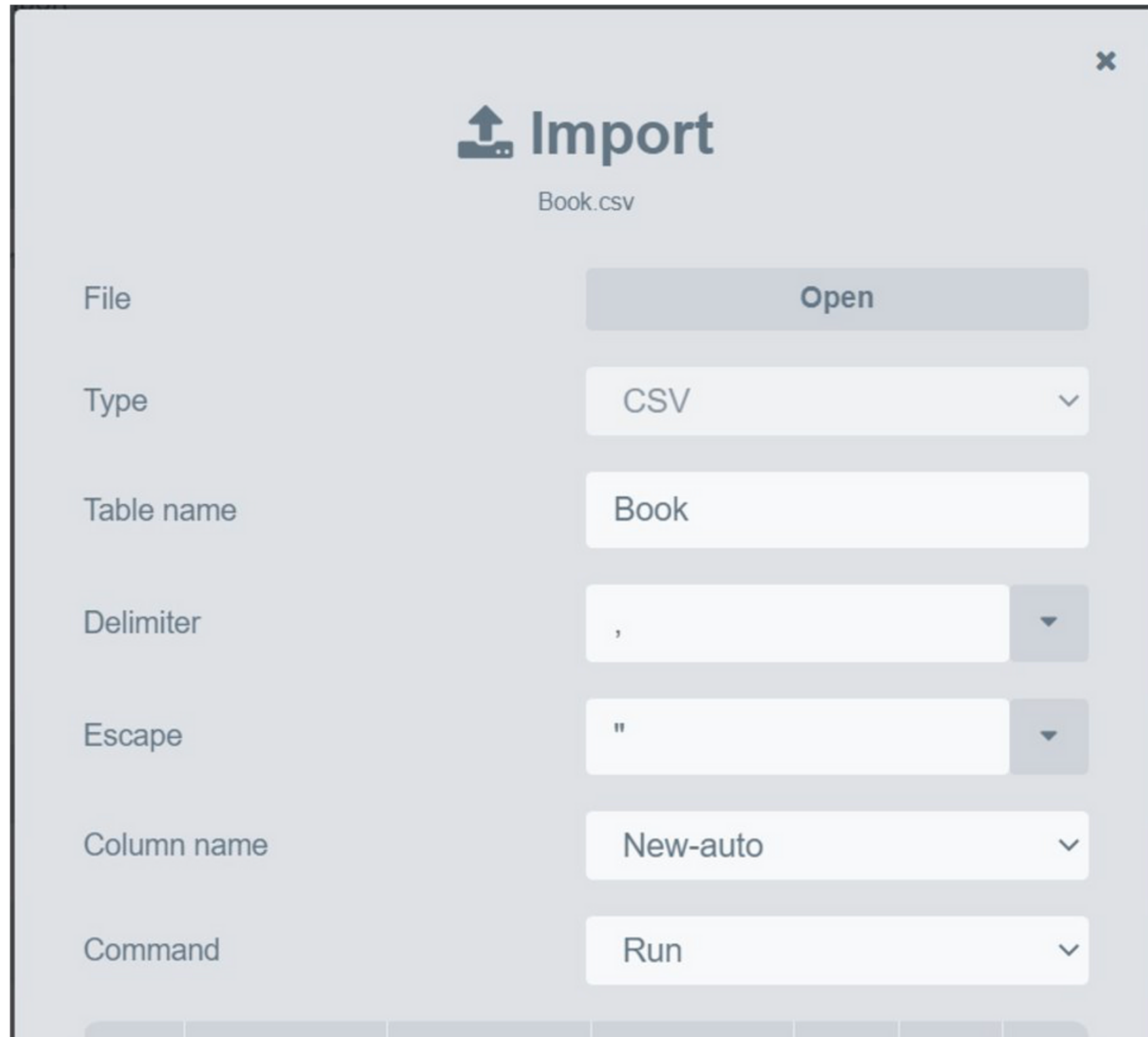
- Πατάμε το εικονίδιο **Import**
- Επιλέγουμε το **Open**
- Επιλέγουμε το αρχείο **WeatherData.csv** που κατεβάσαμε

!

Όλες οι υπόλοιπες επιλογές μας μένουν ίδιες

(SDC) 0xCcc227E5615D4FADd75822
bc1q25zqmgll2fz0tyduusyfager


SQLITEONLINE



The screenshot shows the 'Import' dialog box in SQLiteOnline. The title bar says 'Import' with a close button. Below the title, it says 'Book.csv'. The form has the following fields:

File	Open
Type	CSV
Table name	Book
Delimiter	,
Escape	"
Column name	New-auto
Command	Run

- Δίνουμε όνομα της επιλογής μας στην Βάση Δεδομένων π.χ. MyDB
- Επιλέγουμε την 2η επιλογή **First Line**



Όλες οι υπόλοιπες επιλογές μας μένουν ίδιες



SELECT...FROM

Πως διαβάζω δεδομένα από την
βάση δεδομένων?



SELECT...FROM

*Πως διαβάζω δεδομένα από την
βάση δεδομένων?*



SELECT από ποιά/ες στήλες θα πάρω δεδομένα.

FROM από ποιόν πίνακα θα επιλέξω τις στήλες. Οι στήλες θα πρέπει να βρίσκονται στον πίνακα.



SELECT..FROM

*Πως διαβάζω δεδομένα από την
βάση δεδομένων?*



```
SELECT *  
FROM WeatherDataThessaloniki;
```



LIMIT

Για να δούμε μόνο κάποιες από τις πρώτες γραμμές ενός πίνακα.

Πιο γρήγορο από το να φορτώσουμε ολόκληρο το dataset.

Είναι **ΠΑΝΤΑ** η τελευταία εντολή.



Παράδειγμα για τις 10 πρώτες γραμμές ενός πίνακα:

```
SELECT *  
FROM WeatherDataThessaloniki  
LIMIT 10;
```



ORDER BY - 1

Ταξινόμηση των αποτελεσμάτων χρησιμοποιώντας τα δεδομένα μίας στήλης.

Έχει παροδική δράση, αντίθετα πχ με το `sort` στο Excel.

Επομένως, στην επόμενη ερώτηση (query), τα δεδομένα θα είναι αταξινόμητα.

Πάντα μετά τα `SELECT` και `FROM` αλλά πριν το `LIMIT`.



ORDER BY - 1

```
SELECT *  
FROM WeatherDataThessaloniki  
ORDER BY Max_Temperature  
LIMIT 10;
```



ORDER BY - 1



Pro tip

Το DESC μετά την στήλη στην δήλωση ORDER BY ταξινομεί τα δεδομένα σε φθίνουσα σειρά.

Η αύξουσα ταξινόμηση είναι η προκαθορισμένη επιλογή.

```
SELECT *  
FROM WeatherDataThessaloniki  
ORDER BY Max_Temperature  
LIMIT 10;
```



ORDER BY - 2

Όταν δίνουμε πάνω από μία στήλες στην εντολή **ORDER BY**, η ταξινόμηση γίνεται αρχικά με βάση την πιο αριστερή στήλη, έπειτα την επόμενη στα αριστερά κλπ.

Αυτό μπορεί να αντιστραφεί με την χρήση της **DESC**.



```
SELECT *  
FROM WeatherDataThessaloniki  
ORDER BY Max_Temperature DESC  
LIMIT 10;
```



WHERE

Κοινά σύμβολα που χρησιμοποιούνται σε μία δήλωση **WHERE**:

- > (μεγαλύτερο)
- < (μικρότερο)
- >= (μεγαλύτερο ή ίσο)
- <= (μικρότερο ή ίσο)
- = (ίσο)
- != (διάφορο)



WHERE

> (μεγαλύτερο)
< (μικρότερο)
>= (μεγαλύτερο ή ίσο)
<= (μικρότερο ή ίσο)
= (ίσο)
!= (διάφορο)



```
SELECT *  
FROM WeatherDataThessaloniki  
WHERE Avg_Temperature > 20  
ORDER BY Max_Temperature  
LIMIT 100;
```

ΑΡΙΘΜΗΤΙΚΟΙ ΤΕΛΕΣΤΕΣ

Η δημιουργία μίας νέας στήλης σαν συνδυασμό άλλων είναι γνωστή ως **παραγόμενη ή υπολογισμένη στήλη**.

Συνήθως της δίνεται ένα όνομα με χρήση της δεσμευμένης λέξης **AS**.

Είναι **προσωρινή** και παύει να υπάρχει στην επόμενη αναζήτηση.

Αν η νέα στήλη έχει παραχθεί με κάποια **μαθηματική έκφραση**:

- * (Πολλαπλασιασμός)
- + (Πρόσθεση)
- - (Αφαίρεση)
- / (Διαίρεση)

ΑΡΙΘΜΗΤΙΚΟΙ ΤΕΛΕΣΤΕΣ

Η δημιουργία μίας νέας στήλης σαν συνδυασμό άλλων είναι γνωστή ως **παραγώμενη ή υπολογισμένη στήλη**.

Συνήθως της δίνεται ένα όνομα με χρήση της δεσμευμένης λέξης **AS**.

Είναι **προσωρινή** και παύει να υπάρχει στην επόμενη αναζήτηση.

Αν η νέα στήλη έχει παραχθεί με κάποια **μαθηματική έκφραση**:

- * (Πολλαπλασιασμός)
- + (Πρόσθεση)
- - (Αφαίρεση)
- / (Διαίρεση)



```
SELECT Date, ((Max_Temperature +  
Min_Temperature)/2) AS Avg_Temperature  
FROM WeatherDataThessaloniki  
LIMIT 50;
```

LIKE

Χρήσιμο όταν επεξεργαζόμαστε **αρχεία κειμένου**.

Χρησιμοποιείται μέσα σε μία εντολή **WHERE**.

Συχνά χρησιμοποιείται με το σύμβολο **%**.

Το **%** μας λέει ότι ίσως να θέλουμε οποιοδήποτε πλήθος αριθμών ή χαρακτήρων μέχρι να βρούμε ένα συγκεκριμένο κομμάτι ή μετά από αυτό.

Μονά ή διπλά εισαγωγικά για την περίπτωση χαρακτήρων αφού το **'T'** δεν είναι το ίδιο με το **'t'**.

LIKE

Χρήσιμο όταν επεξεργαζόμαστε **αρχεία κειμένου**.

Χρησιμοποιείται μέσα σε μία εντολή **WHERE**.

Συχνά χρησιμοποιείται με το σύμβολο **%**.

Το **%** μας λέει ότι ίσως να θέλουμε οποιοδήποτε πλήθος αριθμών ή χαρακτήρων μέχρι να βρούμε ένα συγκεκριμένο κομμάτι ή μετά από αυτό.

Μονά ή διπλά εισαγωγικά για την περίπτωση χαρακτήρων αφού το **'T'** δεν είναι το ίδιο με το **'t'**.



```
SELECT *  
FROM WeatherDataThessaloniki  
WHERE Weather_Description LIKE  
'%cold%';
```

IN

Χρήσιμος τόσο με στήλες που περιέχουν αριθμούς όσο και χαρακτήρες.

Επιτρέπει την χρήση του = αλλά για πάνω από ένα αντικείμενα μίας συγκεκριμένης στήλης.

Μπορούν να ελεγχούν μία, δύο και παραπάνω τιμές μίας στήλης.



Pro tip

Μονά ή διπλά εισαγωγικά για την περίπτωση χαρακτήρων αφού το 'T' δεν είναι το ίδιο με το 't'.

Διπλά αν υπάρχει απόστροφος στο κείμενο.



```
SELECT *  
FROM WeatherDataThessaloniki  
WHERE Weather_Description IN (30,35);
```

NOT

Χρησιμοποιείται μαζί με τους προηγούμενους δύο τελεστές **IN** και **LIKE**.

Με την χρήση των **NOT LIKE** ή **NOT IN**, μπορούμε να εξάγουμε όλες τις στήλες που δεν πληρούν κάποιο συγκεκριμένο κριτήριο.

NOT

Χρησιμοποιείται μαζί με τους προηγούμενους δύο τελεστές **IN** και **LIKE**.

Με την χρήση των **NOT LIKE** ή **NOT IN**, μπορούμε να εξάγουμε όλες τις στήλες που δεν πληρούν κάποιο συγκεκριμένο κριτήριο.



```
SELECT *  
FROM WeatherDataThessaloniki  
WHERE Weather_Description NOT IN  
      (30, 35)  
ORDER BY Date
```


AND & BETWEEN

Ο τελεστής **AND** χρησιμοποιείται μέσα σε μία εντολή **WHERE** για να λάβει υπόψη πάνω από έναν λογικούς όρους.

Πρέπει να αναφέρεται η **στήλη** που μας ενδιαφέρει.
Μπορούν να συνδεθούν όσες δηλώσεις μας ενδιαφέρουν.

Μπορεί να **συνδυαστεί** με όλους τους τελεστές που είδαμε μέχρι τώρα (λογικούς και αριθμητικούς).

Οι **LIKE**, **IN**, και **NOT** μπορούν επίσης να συνδεθούν χρησιμοποιώντας τον τελεστή **AND**.

AND & BETWEEN

Ο τελεστής **AND** χρησιμοποιείται μέσα σε μία εντολή **WHERE** για να λάβει υπόψη πάνω από έναν λογικούς όρους.

Πρέπει να αναφέρεται η **στήλη** που μας ενδιαφέρει.
Μπορούν να συνδεθούν όσες δηλώσεις μας ενδιαφέρουν.

Μπορεί να **συνδυαστεί** με όλους τους τελεστές που είδαμε μέχρι τώρα (λογικούς και αριθμητικούς).

Οι **LIKE**, **IN**, και **NOT** μπορούν επίσης να συνδεθούν χρησιμοποιώντας τον τελεστή **AND**.



```
SELECT *  
FROM WeatherDataThessaloniki  
WHERE Avg_Humidity >= 80 AND  
Avg_Humidity <= 100  
ORDER BY Date
```

AND & BETWEEN

Όταν η ίδια στήλη χρησιμοποιείται για διαφορετικά μέρη της δήλωσης AND, η δήλωση BETWEEN βοηθάει σε μία πιο "όμορφη" δήλωση.

Για παράδειγμα, αντί για:

```
WHERE column >= 6 AND column <= 10
```

Μπορούμε να γράψουμε:

```
WHERE column BETWEEN 6 AND 10
```

AND & BETWEEN

Όταν η ίδια στήλη χρησιμοποιείται για διαφορετικά μέρη της δήλωσης AND, η δήλωση BETWEEN βοηθάει σε μία πιο "όμορφη" δήλωση.

Για παράδειγμα, αντί για:

```
WHERE column >= 6 AND column <= 10
```

Μπορούμε να γράψουμε:

```
WHERE column BETWEEN 6 AND 10
```



```
SELECT *  
FROM WeatherDataThessaloniki  
WHERE Avg_Humidity BETWEEN 80 AND 90  
ORDER BY Date
```

OR

Ο τελεστής OR μπορεί να συνδυάσει πολλαπλές δηλώσεις.

Πρέπει να αναφέρεται η στήλη που μας ενδιαφέρει.

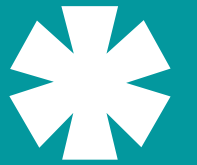
Μπορούν να συνδεθούν όσες δηλώσεις μας ενδιαφέρουν.

Μπορεί να συνδυαστεί με όλους τους τελεστές που είδαμε μέχρι τώρα (λογικούς και αριθμητικούς).

Οι LIKE, IN, NOT, AND, και BETWEEN μπορούν επίσης να συνδεθούν χρησιμοποιώντας τον τελεστή OR.

ΛΟΓΙΚΟΙ ΤΕΛΕΣΤΕΣ

OR



Ο τελεστής OR μπορεί να συνδυάσει πολλαπλές δηλώσεις.

Πρέπει να αναφέρεται η στήλη που μας ενδιαφέρει.

Μπορούν να συνδεθούν όσες δηλώσεις μας ενδιαφέρουν.

Μπορεί να συνδυαστεί με όλους τους τελεστές που είδαμε μέχρι τώρα (λογικούς και αριθμητικούς).

Οι LIKE, IN, NOT, AND, και BETWEEN μπορούν επίσης να συνδεθούν χρησιμοποιώντας τον τελεστή OR.

```
SELECT Date,  
Max_Temperature,  
Avg_Humidity,  
Avg_Wind_Speed,  
Avg_Pressure  
FROM WeatherDataThessaloniki  
WHERE Avg_Temperature > 30 OR  
Avg_Humidity > 80 OR  
Avg_Wind_Speed > 20
```



Pro tip

Όταν συνδυάζουμε πολλαπλές τέτοιες δηλώσεις, είναι καλό να χρησιμοποιούμε παρενθέσεις!

COUNT & NULL

Το NULL είναι τύπος δεδομένων που δείχνει ότι δεν υπάρχουν δεδομένα.

Στις συναρτήσεις πρόσθεσης συχνά αγνοούνται.

Μέτρηση του πλήθους των γραμμών σε ένα πίνακα

COUNT & NULL

Το NULL είναι τύπος δεδομένων που δείχνει ότι δεν υπάρχουν δεδομένα.

Στις συναρτήσεις πρόσθεσης συχνά αγνοούνται.

Μέτρηση του πλήθους των γραμμών σε ένα πίνακα



```
SELECT COUNT(*)  
FROM WeatherDataThessaloniki;
```


SUM

Αντίθετα με την COUNT, η SUM μπορεί να χρησιμοποιηθεί μόνο σε αριθμητικά δεδομένα.

Θα αγνοήσει τις τιμές NULL.

Aggregation Reminder

Οι προσθέσεις γίνονται ΜΟΝΟ κάθετα – τιμές μίας στήλης.

Αν θέλουμε πρόσθεση σε μία συγκεκριμένη γραμμή, γίνεται αριθμητικά.

SUM

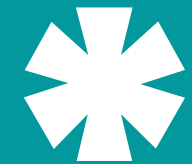
Αντίθετα με την COUNT, η SUM μπορεί να χρησιμοποιηθεί μόνο σε αριθμητικά δεδομένα.

Θα αγνοήσει τις τιμές NULL.

Aggregation Reminder

Οι προσθέσεις γίνονται ΜΟΝΟ κάθετα – τιμές μίας στήλης.

Αν θέλουμε προσθεση σε μία συγκεκριμένη γραμμή, γίνεται αριθμητικά.



```
SELECT SUM(Avg_Temperature)/COUNT(*)  
        AS avg_temp,  
        SUM(Avg_Wind_Speed)/COUNT(*) AS  
        avg_wind,  
        SUM(Avg_Pressure)/COUNT(*) AS  
        avg_press  
FROM WeatherDataThessaloniki
```

MIN & MAX

Τα MIN και MAX επίσης αγνοούν τις τιμές NULL values.

MIN & MAX

Τα MIN και MAX επίσης αγνοούν τις τιμές NULL.

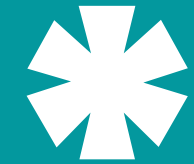
Λειτουργούν παρόμοια με την COUNT αφού μπορούν να χρησιμοποιηθούν και σε στήλες με μη αριθμητικά δεδομένα.



Pro tip

Ανάλογα με τον τύπο της στήλης, η MIN θα επιστρέψει τον μικρότερο αριθμό, την πιο μακρινή ημερομηνία ή τον χαρακτήρα που βρίσκεται πιο κοντά στο "A".

Η MAX κάνει το ανάποδο. Επιστρέφει τον υψηλότερο αριθμό, την πιο κοντινή ημερομηνία ή τον χαρακτήρα που βρίσκεται πιο κοντά στο "Z".



```
SELECT MIN(Avg_Temperature) AS
           avg_temp_min,
MIN(Min_Temperature) AS
           min_temp_min,
MIN(Min_Humidity) AS
           min_humidity_min,
MAX(Max_Temperature) AS
           max_temp_max,
MAX(Max_Humidity) AS
           max_hum_max,
MAX(Max_Wind_Speed) AS wind_max
FROM WeatherDataThessaloniki
```

AVG

Επιστρέφει τον μέσο όρο όλων των δεδομένων (άθροισμα όλων των στοιχείων σε μία στήλη διά το πλήθος τους).

Επίσης αγνοεί τις τιμές **NULL** τόσο στον αριθμητή όσο και στον παρονομαστή.

Αν θέλουμε τα **NULL** να μετρηθούν σαν μηδενικά, θα πρέπει να χρησιμοποιηθούν οι **SUM** και **COUNT**.

Αυτό δεν είναι καλή ιδέα αν οι τιμές **NULL** αναπαριστούν άγνωστες τιμές για τα δεδομένα μας.

ΛΟΓΙΚΟΙ ΤΕΛΕΣΤΕΣ

AVG

Επιστρέφει τον μέσο όρο όλων των δεδομένων (άθροισμα όλων των στοιχείων σε μία στήλη διά το πλήθος τους).

Επίσης αγνοεί τις τιμές **NULL** τόσο στον αριθμητή όσο και στον παρονομαστή.

Αν θέλουμε τα **NULL** να μετρηθούν σαν μηδενικά, θα πρέπει να χρησιμοποιηθούν οι **SUM** και **COUNT**.

Αυτό δεν είναι καλή ιδέα αν οι τιμές **NULL** αναπαριστούν άγνωστες τιμές για τα δεδομένα μας.



Pro tip

Η μετρική median ίσως να είναι πιο κατάλληλη για να βρεθεί η κεντρική τιμή των δεδομένων αλλά είναι πιο δύσκολο.



```
SELECT AVG(Avg_Temperature) AS
        temperature_avg,
        AVG(Avg_Humidity) AS
        humidity_avg,
        AVG(Avg_Wind_Speed) AS wind_avg
FROM WeatherDataThessaloniki
```

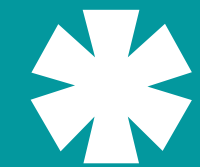
GROUP BY

Η εντολή **GROUP BY** μπορεί να χρησιμοποιηθεί για να αθροίσει δεδομένα σε ένα υποσύνολο.

Για παράδειγμα, να βρούμε τον μέσο όρο θερμοκρασίας μόνο για έναν μήνα.

Αν μία στήλη δεν βρίσκεται στο άθροισμα, θα πρέπει να περιλαμβάνεται σε μία δήλωση **GROUP BY**.

Πάντα μεταξύ των **WHERE** και **ORDER BY**.

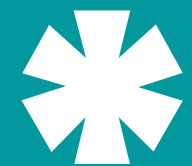


```
SELECT Max_Temperature,  
       COUNT(*)  
FROM WeatherDataThessaloniki  
GROUP BY Max_Temperature  
ORDER BY Max_Temperature
```

GROUP BY

Μπορούμε να το χρησιμοποιήσουμε σε **πολλαπλές στήλες** την ίδια στιγμή.

Η σειρά που αναφέρονται οι στήλες στην δήλωση **ORDER BY** επηρεάζει την κατάταξη. Γίνεται από αριστερά προς τα δεξιά.



```
SELECT Max_Temperature,  
       MAX(Max_Humidity) AS Hum  
FROM WeatherDataThessaloniki  
GROUP BY Max_Temperature  
ORDER BY Max_Temperature, Hum
```


ΔΗΛΩΣΕΙΣ CASE

Πάντα προηγούνται της δήλωσης **SELECT**.

Πρέπει να περιλαμβάνει τα: **WHEN**, **THEN**, και **END**. Το **ELSE** είναι προαιρετικό.

Μπορεί να γίνει οποιοσδήποτε λογικός έλεγχος μεταξύ των **WHEN** και **THEN**. Για παράδειγμα, πολλαπλά **AND** και **OR**.

Μπορούν να χρησιμοποιηθούν πολλαπλές δηλώσεις **WHEN** καθώς και μία δήλωση **ELSE** για τις ανεπιθύμητες συνθήκες.

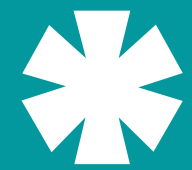
ΔΗΛΩΣΕΙΣ CASE

Πάντα προηγούνται της δήλωσης **SELECT**.

Πρέπει να περιλαμβάνει τα: **WHEN**, **THEN**, και **END**. Το **ELSE** είναι προαιρετικό.

Μπορεί να γίνει οποιοσδήποτε λογικός έλεγχος μεταξύ των **WHEN** και **THEN**. Για παράδειγμα, πολλαπλά **AND** και **OR**.

Μπορούν να χρησιμοποιηθούν πολλαπλές δηλώσεις **WHEN** καθώς και μία δήλωση **ELSE** για τις ανεπιθύμητες συνθήκες.



```
SELECT date,
max_temperature
max_humidity,
CASE WHEN max_temperature > 30
AND max_humidity > 80 THEN 'Too hot
weather!'
WHEN max_temperature < 10 AND
max_humidity > 80 THEN 'Too cold
Weather!'
END AS total_group
FROM WeatherDataThessaloniki
```

ΔΗΛΩΣΕΙΣ CASE

Πάντα προηγούνται της δήλωσης **SELECT**.

Πρέπει να περιλαμβάνει τα: **WHEN**, **THEN**, και **END**. Το **ELSE** είναι προαιρετικό.

Μπορεί να γίνει οποιοσδήποτε λογικός έλεγχος μεταξύ των **WHEN** και **THEN**. Για παράδειγμα, πολλαπλά **AND** και **OR**.

Μπορούν να χρησιμοποιηθούν πολλαπλές δηλώσεις **WHEN** καθώς και μία δήλωση **ELSE** για τις ανεπιθύμητες συνθήκες.



```
SELECT date,  
max_temperature  
max_humidity,  
CASE WHEN max_temperature > 30  
AND max_humidity > 80 THEN 'Too hot  
weather!'  
WHEN max_temperature < 10 AND  
max_humidity > 80 THEN 'Too cold  
Weather!'  
END AS total_group  
FROM WeatherDataThessaloniki
```

Ολοκληρωμένο Παράδειγμα



Θεωρώντας τον πίνακα `WeatherData` που είδαμε έως τώρα, θα γράψουμε έναν κώδικα σε `SQL` όπου:

- Θέλουμε να εμφανίσουμε τις 10 μέρες με την υψηλότερη θερμοκρασία, την υψηλότερη υγρασία και την χαμηλότερη ταχύτητα ανέμου για το 2022 στην Θεσσαλονίκη.
- Θέλουμε να υπολογίσουμε το πλήθος των ημερών που η ελάχιστη θερμοκρασία στην Θεσσαλονίκη ήταν χαμηλότερη από 0oC και είχε πολύ δυνατό αέρα.
- Θέλουμε να εμφανίσουμε τις μέρες που είχε ζέστη ή πολύ ζέστη (δεν μας ενδιαφέρει η υγρασία ή ο άνεμος), μόνο για τους καλοκαιρινούς μήνες (01/06/2022 έως 31/08/2022) και αυτές τις μέρες να τις κατατάξουμε αρχικά με φθίνουσα σειρά στην μέγιστη θερμοκρασία τους. Αν κάποιες μέρες έχουν την ίδια μέγιστη θερμοκρασία, να καταταχθούν ανάλογα με την υγρασία σε φθίνουσα σειρά.

Ολοκληρωμένο Παράδειγμα

```
SELECT date,  
       max_temperature,  
       max_humidity,  
       min_wind_speed  
FROM WeatherDataThessaloniki  
ORDER BY max_temperature DESC, max_humidity DESC, min_wind_speed  
LIMIT 10
```

```
SELECT COUNT(date) AS num_of_very_cold_days  
FROM WeatherDataThessaloniki  
WHERE min_temperature <= 0 AND max_wind_speed > 20
```

```
SELECT date,  
       max_temperature,  
       max_humidity,  
       weather_description  
FROM WeatherDataThessaloniki  
WHERE ((date like '%-06-%') OR (date like '%-07-%') OR (date like  
'%-08-%')) AND weather_description LIKE 'hot%'  
ORDER BY max_temperature DESC, max_humidity DESC
```



Η SQL δίνει την δυνατότητα και για την οπτικοποίηση των δεδομένων σε διαγράμματα.

```
LINE-SELECT Date, Avg_Temperature  
FROM WeatherDataThessaloniki
```

```
BAR-SELECT Date, Avg_Temperature  
FROM WeatherDataThessaloniki
```

```
AREA-SELECT Date, Avg_Temperature  
FROM WeatherDataThessaloniki
```

```
LINE-SELECT Date,  
                Avg_Temperature,  
                Avg_Wind_Speed  
FROM WeatherDataThessaloniki
```

QUESTIONS



STE(A)M PARTNERSHIPS

Education Resilience in Europe

SCIENTIX
The community for science education in Europe

CISCO

European Schoolnet

The STEAM Partnerships has been funded under the European Union Horizon 2020 research and innovation programme – project SciEntix 6, coordinated by European Schoolnet (ESN). The content of this document is the sole responsibility of the organizers and it does not represent the opinion of the European Commission (EC), and the EC is not responsible for any use that might be made of information contained.