

Μάθημα 4

**ΕΞΟΥΣΙΑ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΨΗΦΙΑΚΟ
ΚΑΙ ΔΙΑΔΙΚΤΥΑΚΟ ΠΕΡΙΕΧΟΜΕΝΟ**

Αναπαράσταση Αριθμητικών Δεδομένων

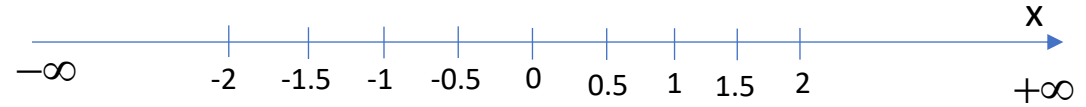
- **Μονοδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί

- Έχουν μόνο μία ιδιότητα/μεταβλητή/διάσταση
- Αναπαρίστανται με μόνο μία μεταβλητή/ιδιότητα/διάσταση (x) πάνω στον άξονα των πραγματικών αριθμών και άρα αναπαρίστανται μόνο με έναν αριθμό

Αναπαράσταση Αριθμητικών Δεδομένων

- **Μονοδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί

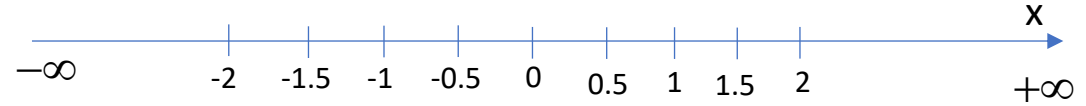
- Έχουν μόνο μία ιδιότητα/μεταβλητή/διάσταση
- Αναπαρίστανται με μόνο μία μεταβλητή/ιδιότητα/διάσταση (x) πάνω στον άξονα των πραγματικών αριθμών και άρα αναπαρίστανται μόνο με έναν αριθμό



Αναπαράσταση Αριθμητικών Δεδομένων

- **Μονοδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί

- Έχουν μόνο μία ιδιότητα/μεταβλητή/διάσταση
- Αναπαρίστανται με μόνο μία μεταβλητή/ιδιότητα/διάσταση (x) πάνω στον άξονα των πραγματικών αριθμών και άρα αναπαρίστανται μόνο με έναν αριθμό



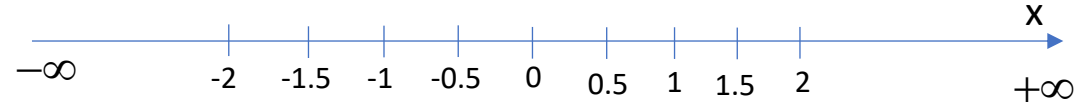
- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος 100 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

| x |
|-----------|
| x_1 |
| x_2 |
| |
| x_3 |
| |
| x_{100} |

Αναπαράσταση Αριθμητικών Δεδομένων

- **Μονοδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί

- Έχουν μόνο μία ιδιότητα/μεταβλητή/διάσταση
- Αναπαρίστανται με μόνο μία μεταβλητή/ιδιότητα/διάσταση (x) πάνω στον άξονα των πραγματικών αριθμών και άρα αναπαρίστανται μόνο με έναν αριθμό



- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος 100 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

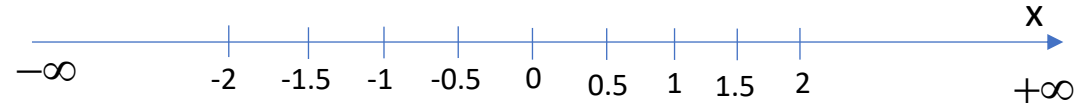
| x | A/A | x |
|-----------|-----|-----------|
| x_1 | 1 | x_1 |
| x_2 | 2 | x_2 |
| | ... | |
| x_3 | k | x_3 |
| | ... | |
| x_{100} | 100 | x_{100} |

A diagram illustrating the mapping of data from a single column to a two-column table. On the left, a vertical column of cells contains the values x , x_1 , x_2 , an empty cell, x_3 , an empty cell, and x_{100} . On the right, a two-column table has a header row with 'A/A' and ' x '. The rows below contain pairs: (1, x_1), (2, x_2), (...), (k , x_3), (...), and (100, x_{100}). A blue arrow labeled k points downwards from the first row to the k -th row, indicating the index of the data point.

Αναπαράσταση Αριθμητικών Δεδομένων

- **Μονοδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί

- Έχουν μόνο μία ιδιότητα/μεταβλητή/διάσταση
- Αναπαρίστανται με μόνο μία μεταβλητή/ιδιότητα/διάσταση (x) πάνω στον άξονα των πραγματικών αριθμών και άρα αναπαρίστανται μόνο με έναν αριθμό

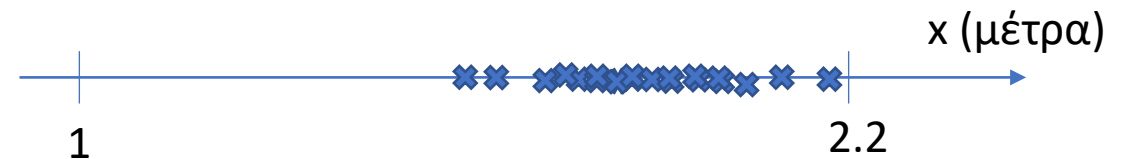


- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος 100 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

| x |
|-----------|
| x_1 |
| x_2 |
| ... |
| x_3 |
| ... |
| x_{100} |

| A/A | x |
|-----|-----------|
| 1 | x_1 |
| 2 | x_2 |
| ... | |
| k | x_3 |
| ... | |
| 100 | x_{100} |

A vertical double-headed arrow labeled k indicates the range of rows in the second table.



Πεδίο Ορισμού [1.5-2.2] μέτρα

Αναπαράσταση Αριθμητικών Δεδομένων

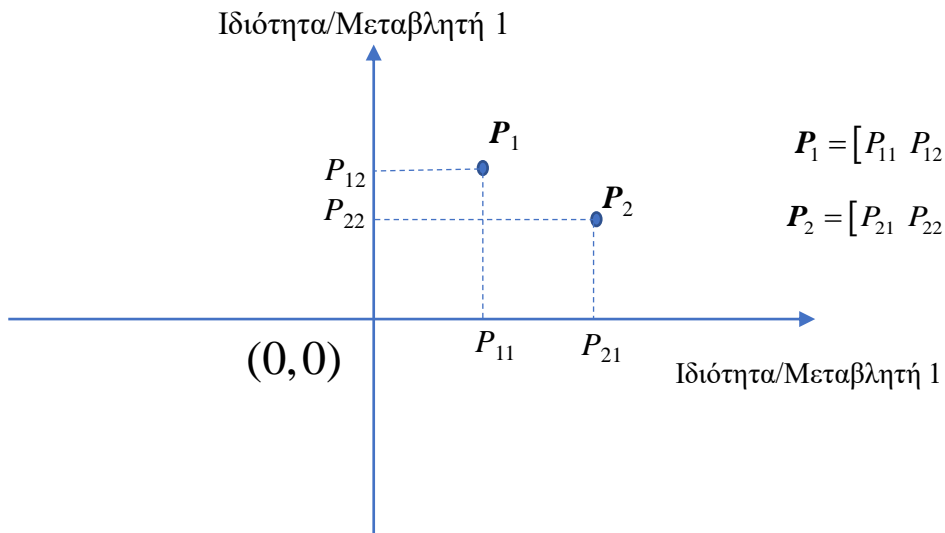
- **Δυσδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί
 - Έχουν δύο ιδιότητες/μεταβλητές/διαστάσεις

Αναπαράσταση Αριθμητικών Δεδομένων

- **Δυσδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί
 - Έχουν δύο ιδιότητες/μεταβλητές/διαστάσεις
 - Αναπαρίστανται με δύο μεταβλητές/ιδιότητες/διαστάσεις (x_1, x_2) και άρα με δύο αριθμούς/συντεταγμένες πάνω στο επίπεδο

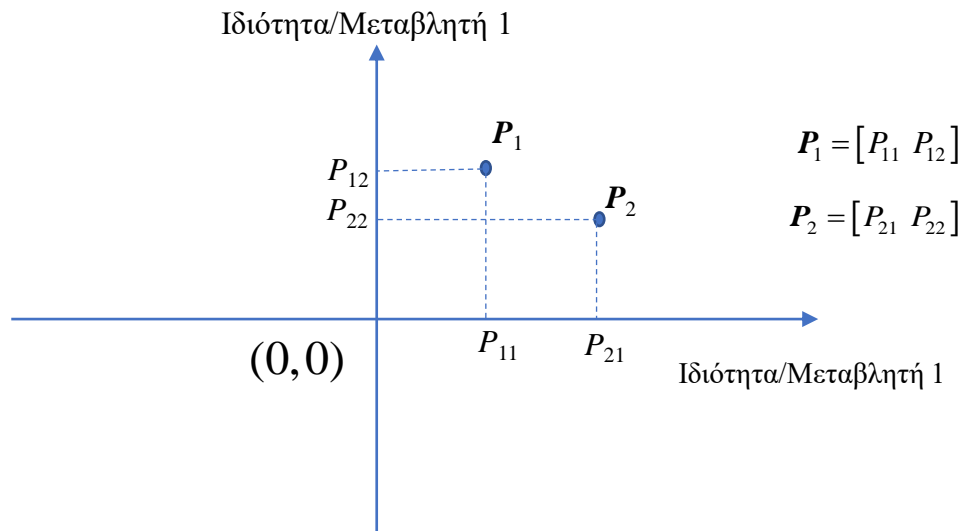
Αναπαράσταση Αριθμητικών Δεδομένων

- **Δυσδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί
 - Έχουν δύο ιδιότητες/μεταβλητές/διαστάσεις
 - Αναπαρίστανται με δύο μεταβλητές/ιδιότητες/διαστάσεις (x_1, x_2) και άρα με δύο αριθμούς/συντεταγμένες πάνω στο επίπεδο



Αναπαράσταση Αριθμητικών Δεδομένων

- **Δυσδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί
 - Έχουν δύο ιδιότητες/μεταβλητές/διαστάσεις
 - Αναπαρίστανται με δύο μεταβλητές/ιδιότητες/διαστάσεις (x_1, x_2) και άρα με δύο αριθμούς/συντεταγμένες πάνω στο επίπεδο

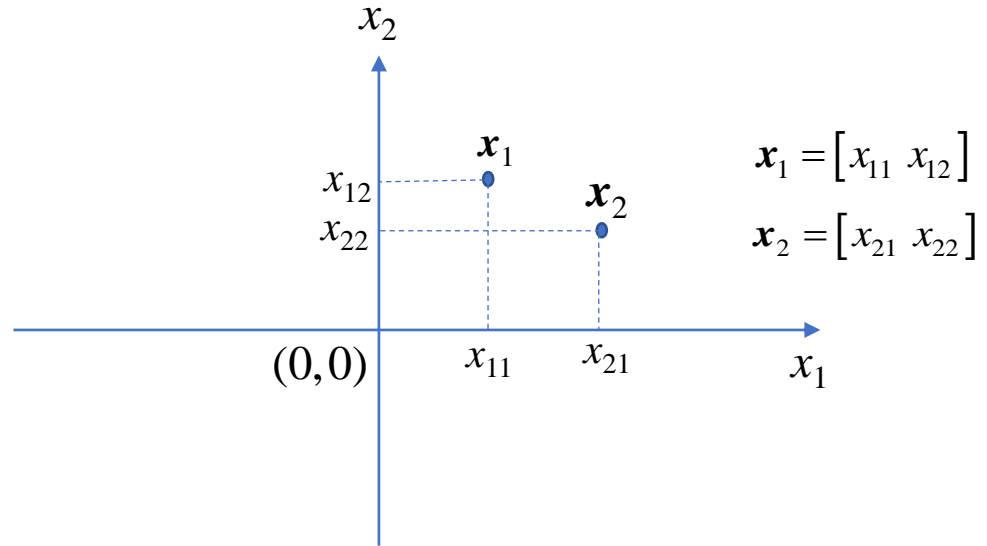
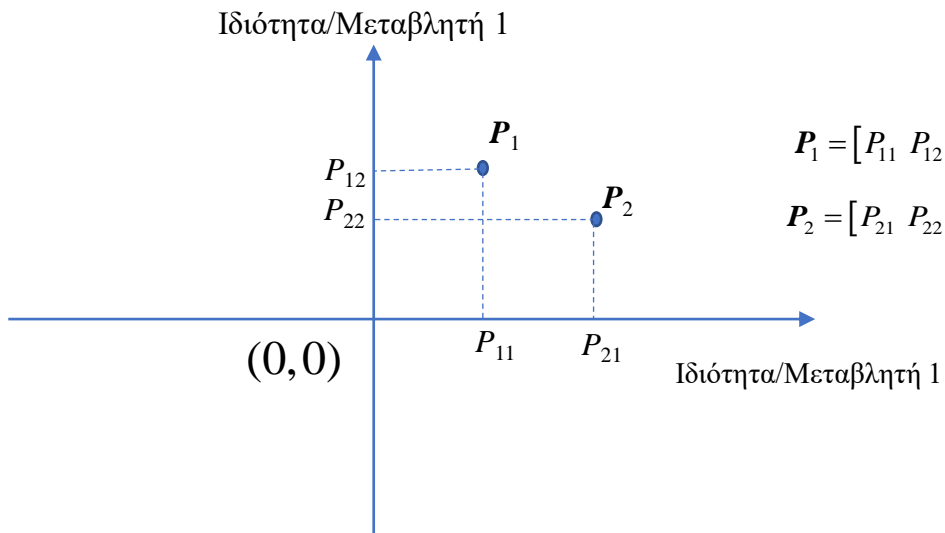


| | I1 | I2 |
|-------|----------|----------|
| P_1 | P_{11} | P_{12} |
| P_2 | P_{21} | P_{22} |

Αναπαράσταση Αριθμητικών Δεδομένων

- **Δυσδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί

- Έχουν δύο ιδιότητες/μεταβλητές/διαστάσεις
- Αναπαρίστανται με δύο μεταβλητές/ιδιότητες/διαστάσεις (x_1, x_2) και άρα με δύο αριθμούς/συντεταγμένες πάνω στο επίπεδο

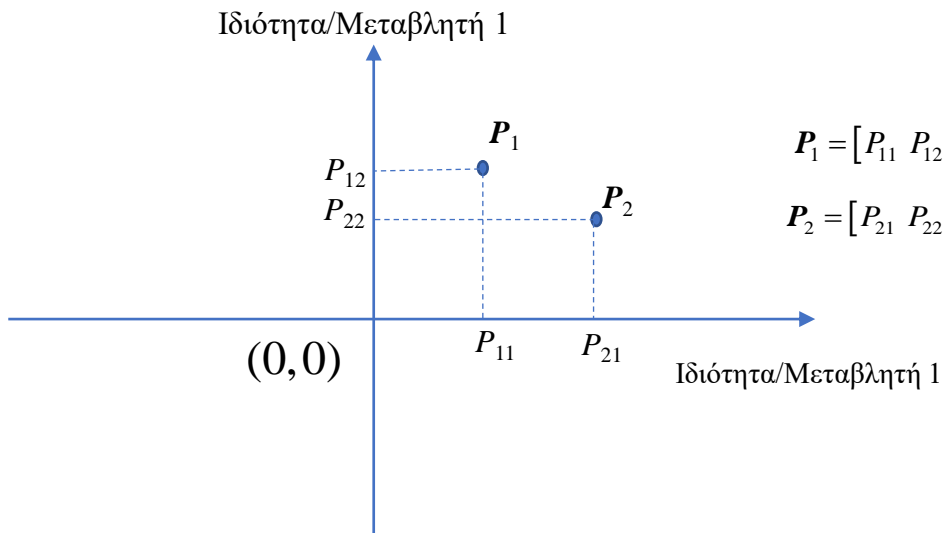


| | I1 | I2 |
|-------|----------|----------|
| P_1 | P_{11} | P_{12} |
| P_2 | P_{21} | P_{22} |

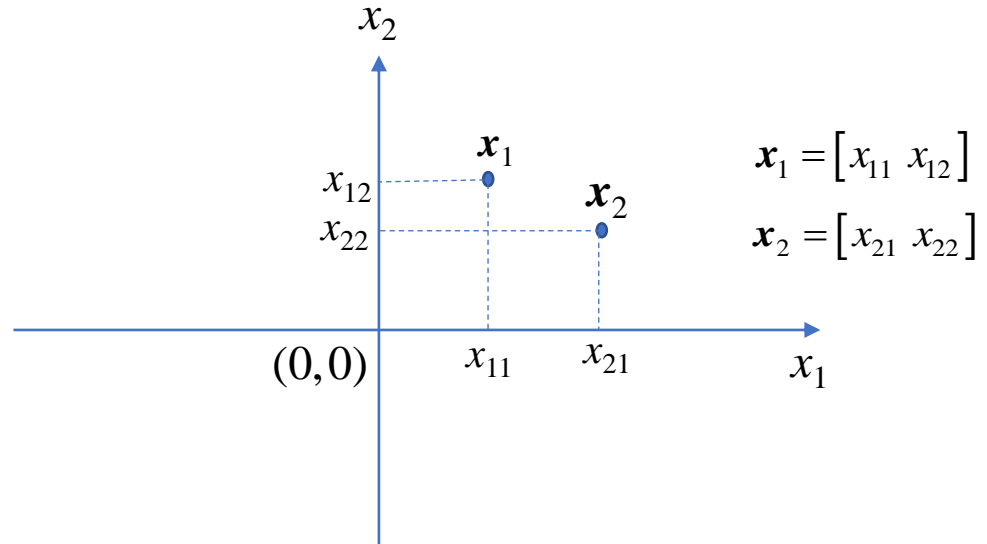
Αναπαράσταση Αριθμητικών Δεδομένων

- **Δυσδιάστατα Δεδομένα:** Πραγματικοί Αριθμοί

- Έχουν δύο ιδιότητες/μεταβλητές/διαστάσεις
- Αναπαρίστανται με δύο μεταβλητές/ιδιότητες/διαστάσεις (x_1, x_2) και άρα με δύο αριθμούς/συντεταγμένες πάνω στο επίπεδο



| | I1 | I2 |
|-------|----------|----------|
| P_1 | P_{11} | P_{12} |
| P_2 | P_{21} | P_{22} |



| | x_1 | x_2 |
|-------|----------|----------|
| x_1 | x_{11} | x_{12} |
| x_2 | x_{21} | x_{22} |

Αναπαράσταση Αριθμητικών Δεδομένων

- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος και το βάρος 5 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

Αναπαράσταση Αριθμητικών Δεδομένων

- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος και το βάρος 5 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

Ιδιότητα/Μεταβλητή 1: x_1 =Ύψος (μέτρα) Πεδίο Ορισμού=[1.5-2.0]
Ιδιότητα/Μεταβλητή 2: x_2 =Βάρος (κιλά) Πεδίο Ορισμού=[50-100]

Αναπαράσταση Αριθμητικών Δεδομένων

- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος και το βάρος 5 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

Ιδιότητα/Μεταβλητή 1: x_1 =Ύψος (μέτρα) Πεδίο Ορισμού=[1.5-2.0]
Ιδιότητα/Μεταβλητή 2: x_2 =Βάρος (κιλά) Πεδίο Ορισμού=[50-100]

| A/A | x_1 | x_2 | |
|-----|--------------------|------------------|-------|
| 1 | (x_{11} =) 1.70 | (x_{11} =) 60 | x_1 |
| 2 | (x_{11} =) 1.80 | (x_{11} =) 90 | x_2 |
| 3 | (x_{11} =) 1.70 | (x_{11} =) 80 | x_3 |
| 4 | (x_{11} =) 1.60 | (x_{11} =) 60 | x_4 |
| 5 | (x_{11} =) 1.90 | (x_{11} =) 90 | x_5 |

Αναπαράσταση Αριθμητικών Δεδομένων

- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος και το βάρος 5 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

Ιδιότητα/Μεταβλητή 1: x_1 =Ύψος (μέτρα)

Πεδίο Ορισμού=[1.5-2.0]

Ιδιότητα/Μεταβλητή 2: x_2 =Βάρος (κιλά)

Πεδίο Ορισμού=[50-100]

| A/A | x_1 | x_2 | |
|-----|--------------------|------------------|-------|
| 1 | (x_{11} =) 1.70 | (x_{11} =) 60 | x_1 |
| 2 | (x_{11} =) 1.80 | (x_{11} =) 90 | x_2 |
| 3 | (x_{11} =) 1.70 | (x_{11} =) 80 | x_3 |
| 4 | (x_{11} =) 1.60 | (x_{11} =) 60 | x_4 |
| 5 | (x_{11} =) 1.90 | (x_{11} =) 90 | x_5 |

| | x_1 | x_2 |
|-------|-------|-------|
| x_1 | 1.70 | 60 |
| x_2 | 1.80 | 90 |
| x_3 | 1.70 | 80 |
| x_4 | 1.60 | 60 |
| x_5 | 1.90 | 90 |

Αναπαράσταση Αριθμητικών Δεδομένων

- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος και το βάρος 4 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

Ιδιότητα/Μεταβλητή 1: x_1 =Ύψος (μέτρα)

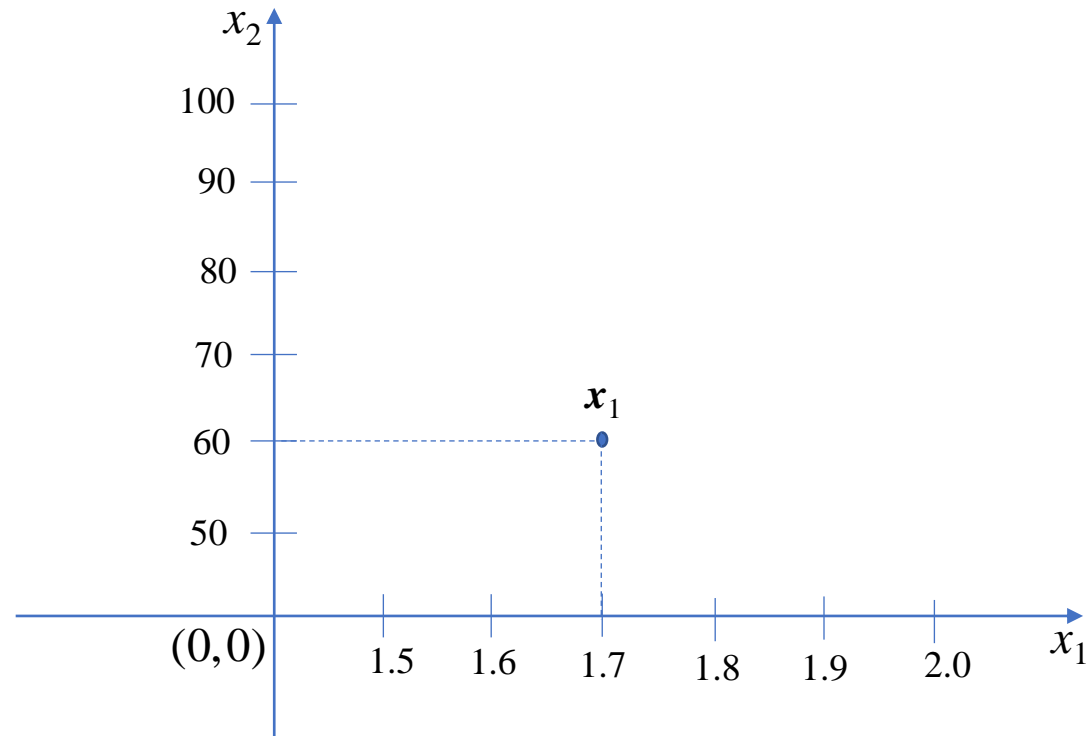
Πεδίο Ορισμού=[1.5-2.0]

Ιδιότητα/Μεταβλητή 2: x_2 =Βάρος (κιλά)

Πεδίο Ορισμού=[50-100]

| A/A | x_1 | x_2 | |
|-----|-------------------|-----------------|-------|
| 1 | (x_{11}) 1.70 | (x_{11}) 60 | x_1 |
| 2 | (x_{11}) 1.80 | (x_{11}) 90 | x_2 |
| 3 | (x_{11}) 1.70 | (x_{11}) 80 | x_3 |
| 4 | (x_{11}) 1.60 | (x_{11}) 60 | x_4 |
| 5 | (x_{11}) 1.90 | (x_{11}) 90 | x_5 |

| | x_1 | x_2 |
|-------|-------|-------|
| x_1 | 1.70 | 60 |
| x_2 | 1.80 | 90 |
| x_3 | 1.70 | 80 |
| x_4 | 1.60 | 60 |
| x_5 | 1.90 | 90 |



Αναπαράσταση Αριθμητικών Δεδομένων

- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος και το βάρος 4 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

Ιδιότητα/Μεταβλητή 1: x_1 =Ύψος (μέτρα)

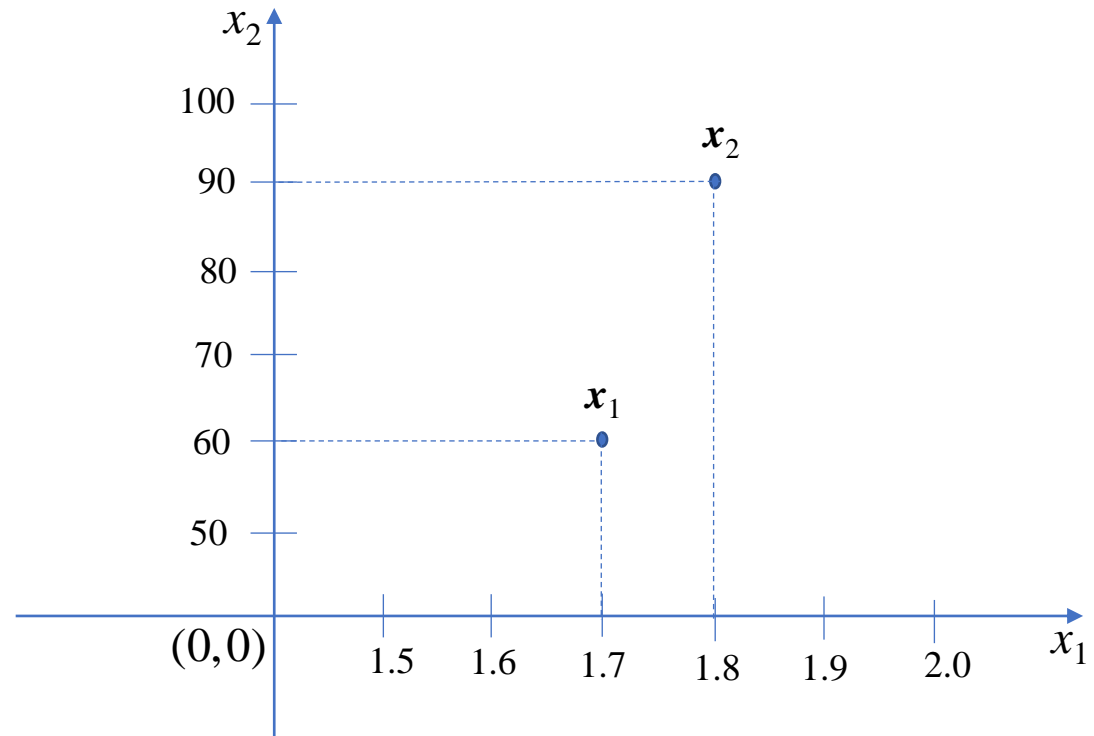
Πεδίο Ορισμού=[1.5-2.0]

Ιδιότητα/Μεταβλητή 2: x_2 =Βάρος (κιλά)

Πεδίο Ορισμού=[50-100]

| A/A | x_1 | x_2 | |
|-----|-------------------|-----------------|-------|
| 1 | (x_{11}) 1.70 | (x_{11}) 60 | x_1 |
| 2 | (x_{11}) 1.80 | (x_{11}) 90 | x_2 |
| 3 | (x_{11}) 1.70 | (x_{11}) 80 | x_3 |
| 4 | (x_{11}) 1.60 | (x_{11}) 60 | x_4 |
| 5 | (x_{11}) 1.90 | (x_{11}) 90 | x_5 |

| | x_1 | x_2 |
|-------|-------|-------|
| x_1 | 1.70 | 60 |
| x_2 | 1.80 | 90 |
| x_3 | 1.70 | 80 |
| x_4 | 1.60 | 60 |
| x_5 | 1.90 | 90 |



Αναπαράσταση Αριθμητικών Δεδομένων

- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος και το βάρος 4 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

Ιδιότητα/Μεταβλητή 1: x_1 =Ύψος (μέτρα)

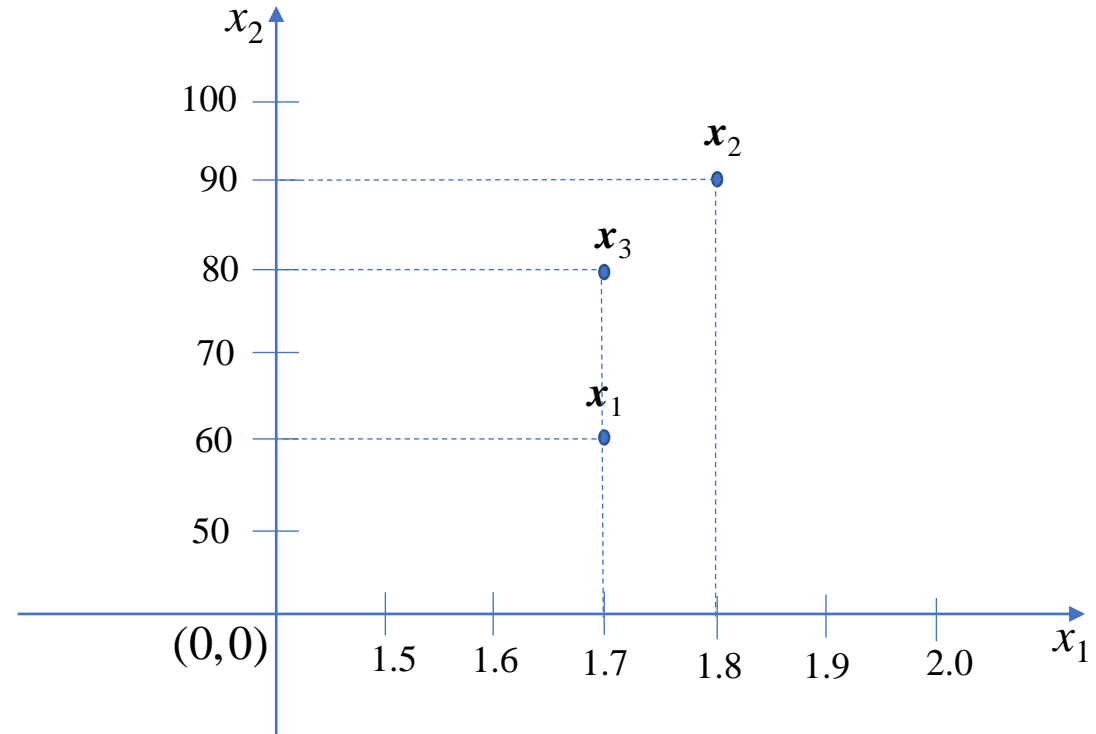
Πεδίο Ορισμού=[1.5-2.0]

Ιδιότητα/Μεταβλητή 2: x_2 =Βάρος (κιλά)

Πεδίο Ορισμού=[50-100]

| A/A | x_1 | x_2 | |
|-----|-------------------|-----------------|-------|
| 1 | (x_{11}) 1.70 | (x_{11}) 60 | x_1 |
| 2 | (x_{11}) 1.80 | (x_{11}) 90 | x_2 |
| 3 | (x_{11}) 1.70 | (x_{11}) 80 | x_3 |
| 4 | (x_{11}) 1.60 | (x_{11}) 60 | x_4 |
| 5 | (x_{11}) 1.90 | (x_{11}) 90 | x_5 |

| | x_1 | x_2 |
|-------|-------|-------|
| x_1 | 1.70 | 60 |
| x_2 | 1.80 | 90 |
| x_3 | 1.70 | 80 |
| x_4 | 1.60 | 60 |
| x_5 | 1.90 | 90 |



Αναπαράσταση Αριθμητικών Δεδομένων

- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος και το βάρος 4 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

Ιδιότητα/Μεταβλητή 1: x_1 =Ύψος (μέτρα)

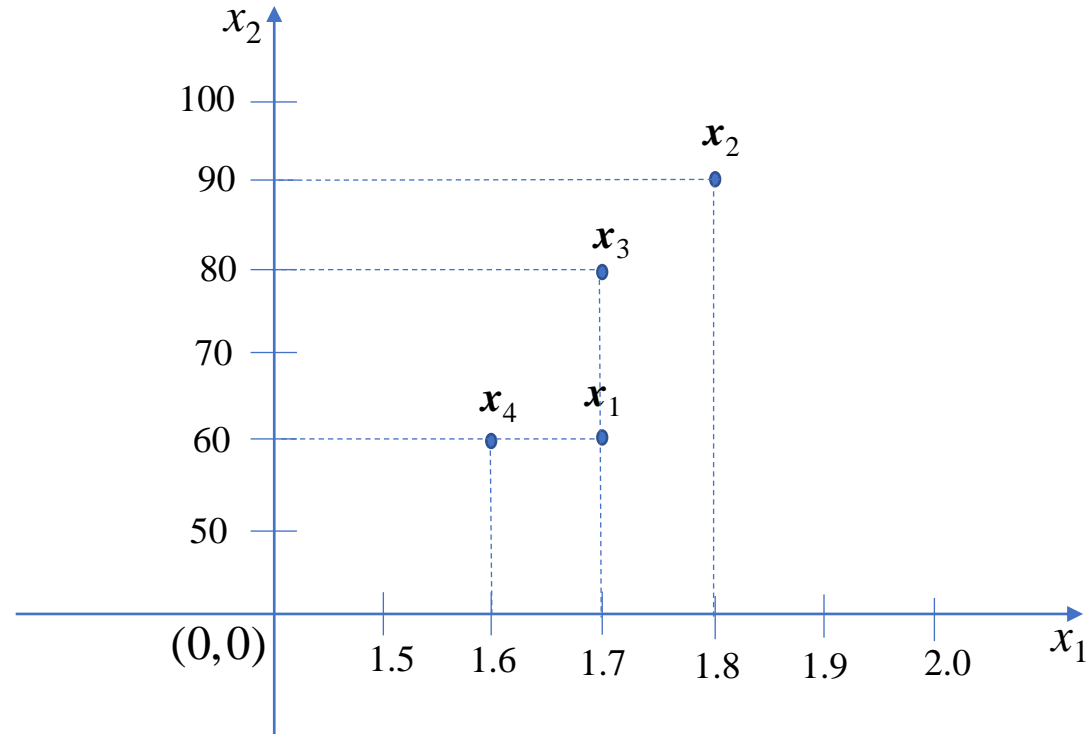
Πεδίο Ορισμού=[1.5-2.0]

Ιδιότητα/Μεταβλητή 2: x_2 =Βάρος (κιλά)

Πεδίο Ορισμού=[50-100]

| A/A | x_1 | x_2 | |
|-----|-------------------|-----------------|-------|
| 1 | (x_{11}) 1.70 | (x_{11}) 60 | x_1 |
| 2 | (x_{11}) 1.80 | (x_{11}) 90 | x_2 |
| 3 | (x_{11}) 1.70 | (x_{11}) 80 | x_3 |
| 4 | (x_{11}) 1.60 | (x_{11}) 60 | x_4 |
| 5 | (x_{11}) 1.90 | (x_{11}) 90 | x_5 |

| | x_1 | x_2 |
|-------|-------|-------|
| x_1 | 1.70 | 60 |
| x_2 | 1.80 | 90 |
| x_3 | 1.70 | 80 |
| x_4 | 1.60 | 60 |
| x_5 | 1.90 | 90 |



Αναπαράσταση Αριθμητικών Δεδομένων

- **Παράδειγμα:** Θέλουμε να μετρήσουμε το ύψος και το βάρος 4 ενήλικων ανθρώπων και να τα καταχωρήσουμε σε μία Βάση Δεδομένων. Η όλη διαδικασία είναι ανώνυμη.

Ιδιότητα/Μεταβλητή 1: x_1 =Ύψος (μέτρα)

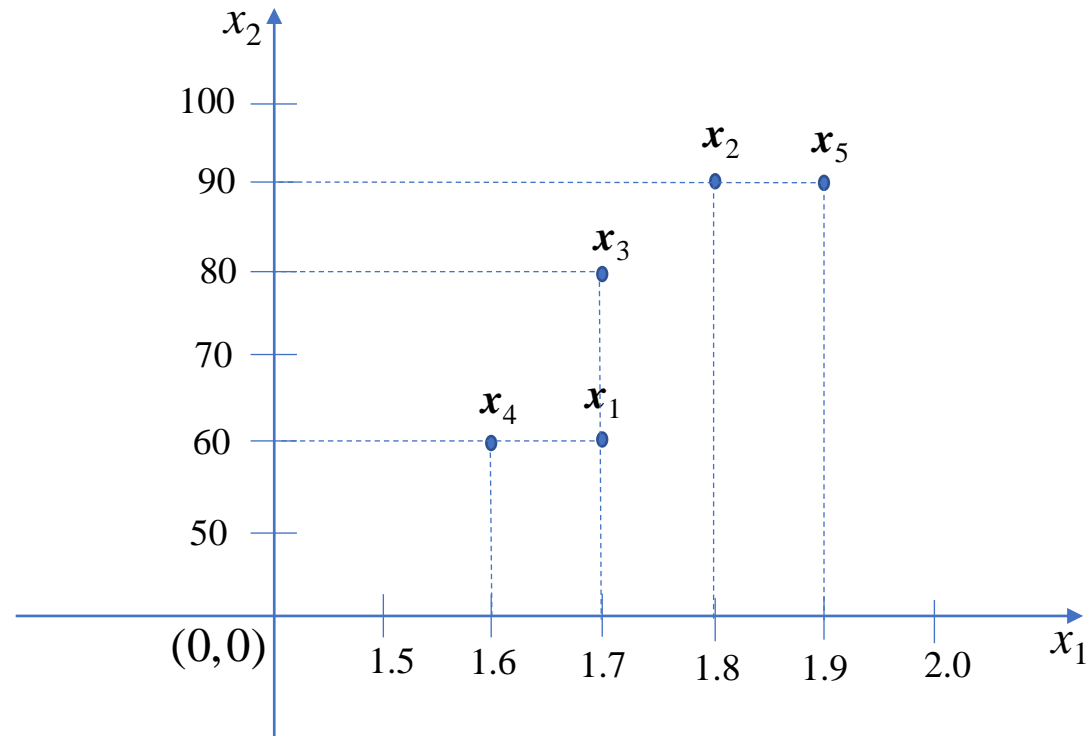
Πεδίο Ορισμού=[1.5-2.0]

Ιδιότητα/Μεταβλητή 2: x_2 =Βάρος (κιλά)

Πεδίο Ορισμού=[50-100]

| A/A | x_1 | x_2 | |
|-----|-------------------|-----------------|-------|
| 1 | (x_{11}) 1.70 | (x_{11}) 60 | x_1 |
| 2 | (x_{11}) 1.80 | (x_{11}) 90 | x_2 |
| 3 | (x_{11}) 1.70 | (x_{11}) 80 | x_3 |
| 4 | (x_{11}) 1.60 | (x_{11}) 60 | x_4 |
| 5 | (x_{11}) 1.90 | (x_{11}) 90 | x_5 |

| | x_1 | x_2 |
|-------|-------|-------|
| x_1 | 1.70 | 60 |
| x_2 | 1.80 | 90 |
| x_3 | 1.70 | 80 |
| x_4 | 1.60 | 60 |
| x_5 | 1.90 | 90 |



Αναπαράσταση Αριθμητικών Δεδομένων

- Δεδομένα με τρεις ιδιότητες/μεταβλητές/διαστάσεις αναπαρίστανται στο τρισδιάστατο σύστημα συντεταγμένων και άρα οπτικοποιούνται

Αναπαράσταση Αριθμητικών Δεδομένων

- Δεδομένα με τρεις ιδιότητες/μεταβλητές/διαστάσεις αναπαρίστανται στο τρισδιάστατο σύστημα συντεταγμένων και άρα οπτικοποιούνται
- Δεδομένα με περισσότερες από 3 μεταβλητές δεν μπορούν να οπτικοποιηθούν αλλά μπορούν να επεξεργαστούν από τον Η/Υ

Αναπαράσταση Αριθμητικών Δεδομένων

- Δεδομένα με τρεις ιδιότητες/μεταβλητές/διαστάσεις αναπαρίστανται στο τρισδιάστατο σύστημα συντεταγμένων και άρα οπτικοποιούνται
- Δεδομένα με περισσότερες από 3 μεταβλητές δεν μπορούν να οπτικοποιηθούν αλλά μπορούν να επεξεργαστούν από τον Η/Υ

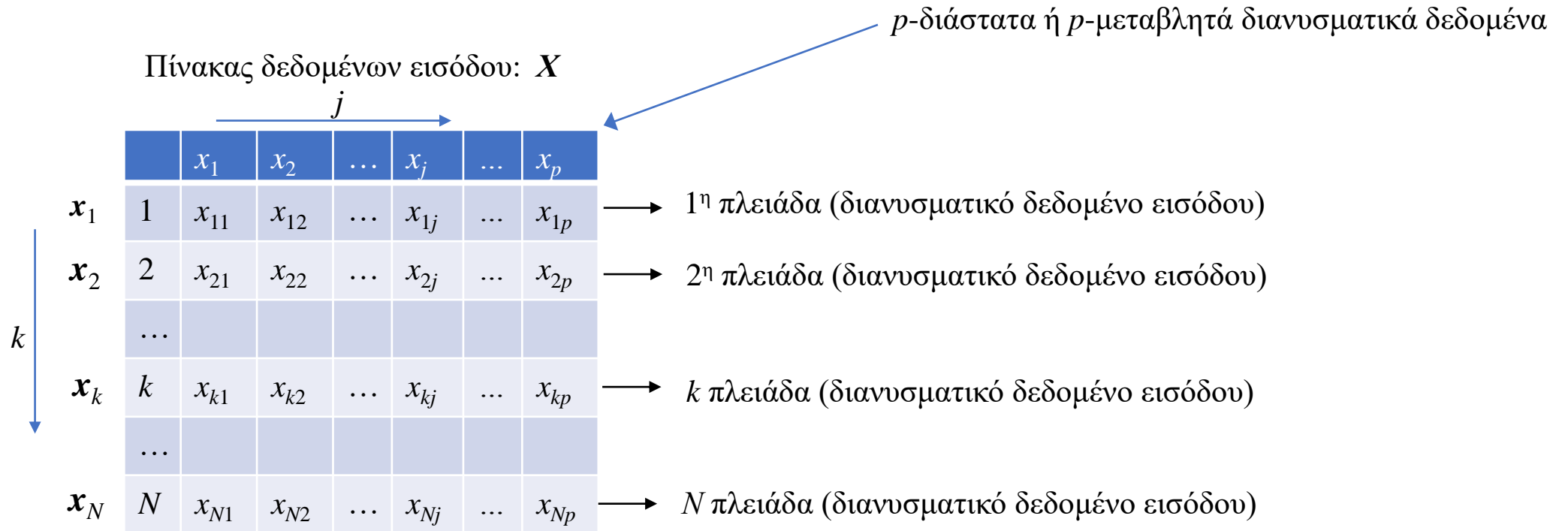
Πίνακας δεδομένων εισόδου: X

| | | x_1 | x_2 | ... | x_j | ... | x_p | | |
|--|----------------|-------|----------|----------|-------|----------|-------|----------|--|
| | \mathbf{x}_1 | 1 | x_{11} | x_{12} | ... | x_{1j} | ... | x_{1p} | → 1 ^η πλειάδα (διανυσματικό δεδομένο εισόδου) |
| | \mathbf{x}_2 | 2 | x_{21} | x_{22} | ... | x_{2j} | ... | x_{2p} | → 2 ^η πλειάδα (διανυσματικό δεδομένο εισόδου) |
| | ... | | | | | | | | |
| | \mathbf{x}_k | k | x_{k1} | x_{k2} | ... | x_{kj} | ... | x_{kp} | → k πλειάδα (διανυσματικό δεδομένο εισόδου) |
| | ... | | | | | | | | |
| | \mathbf{x}_N | N | x_{N1} | x_{N2} | ... | x_{Nj} | ... | x_{Np} | → N πλειάδα (διανυσματικό δεδομένο εισόδου) |

k

Αναπαράσταση Αριθμητικών Δεδομένων

- Δεδομένα με τρεις ιδιότητες/μεταβλητές/διαστάσεις αναπαρίστανται στο τρισδιάστατο σύστημα συντεταγμένων και άρα οπτικοποιούνται
- Δεδομένα με περισσότερες από 3 μεταβλητές δεν μπορούν να οπτικοποιηθούν αλλά μπορούν να επεξεργαστούν από τον Η/Υ



Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Ορισμός Προβλήματος/
Προσδιορισμός Στόχων

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Ορισμός Προβλήματος/
Προσδιορισμός Στόχων



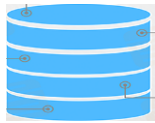
Συλλογή Δεδομένων

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Ορισμός Προβλήματος/
Προσδιορισμός Στόχων

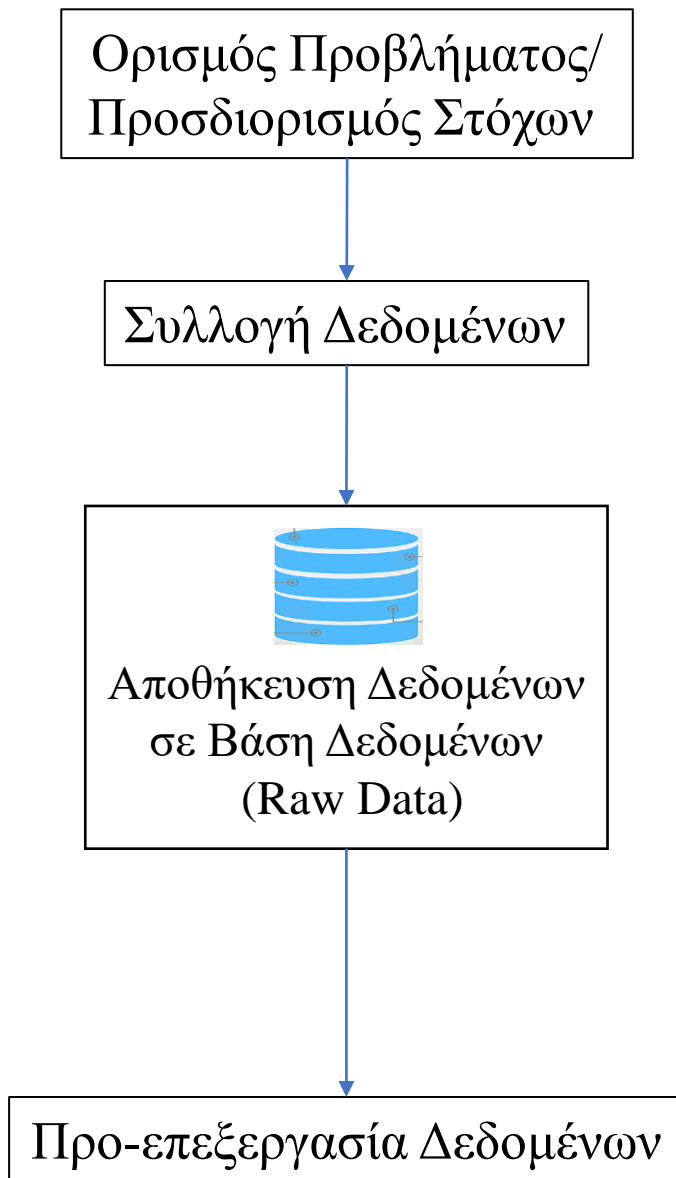


Συλλογή Δεδομένων

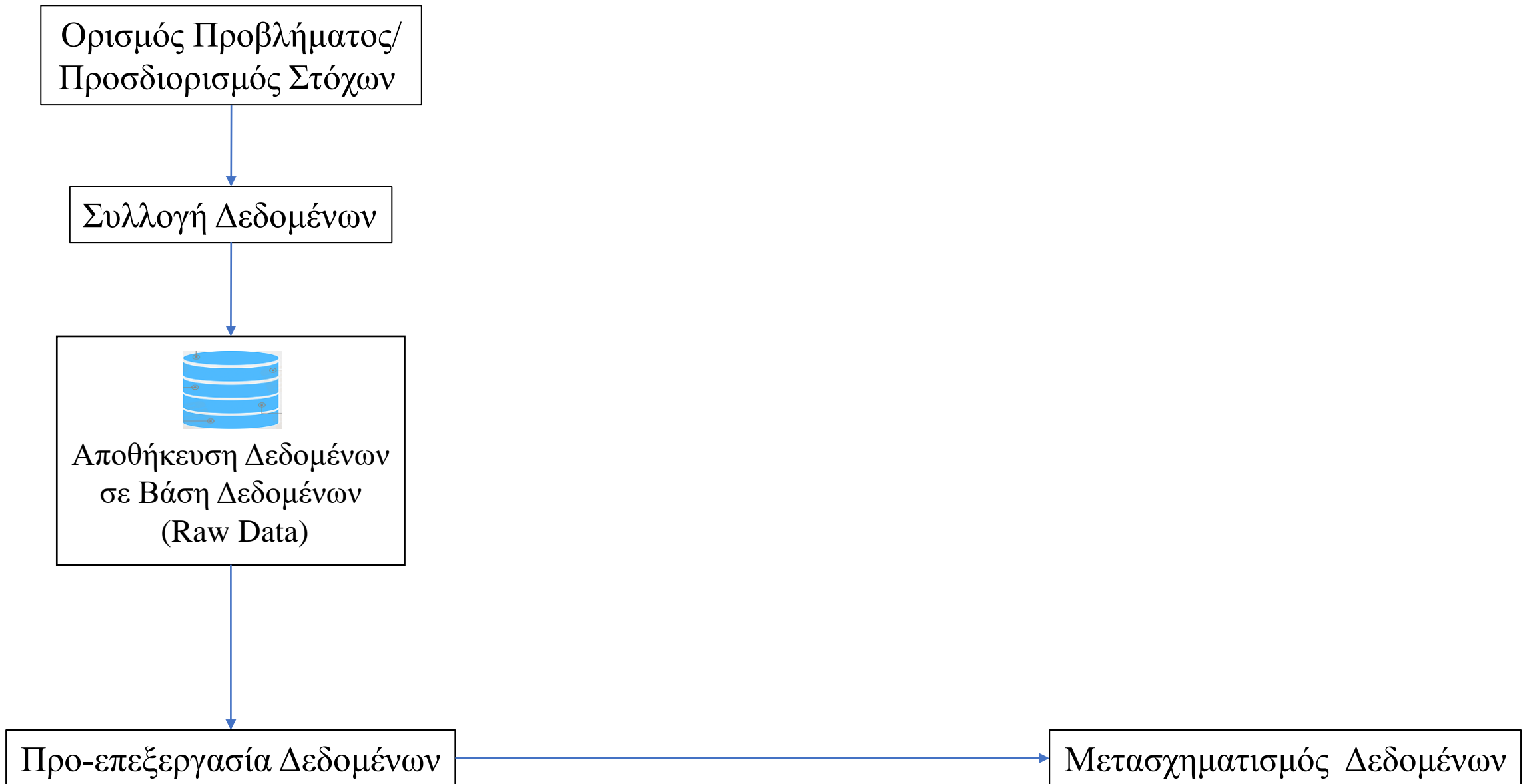


Αποθήκευση Δεδομένων
σε Βάση Δεδομένων
(Raw Data)

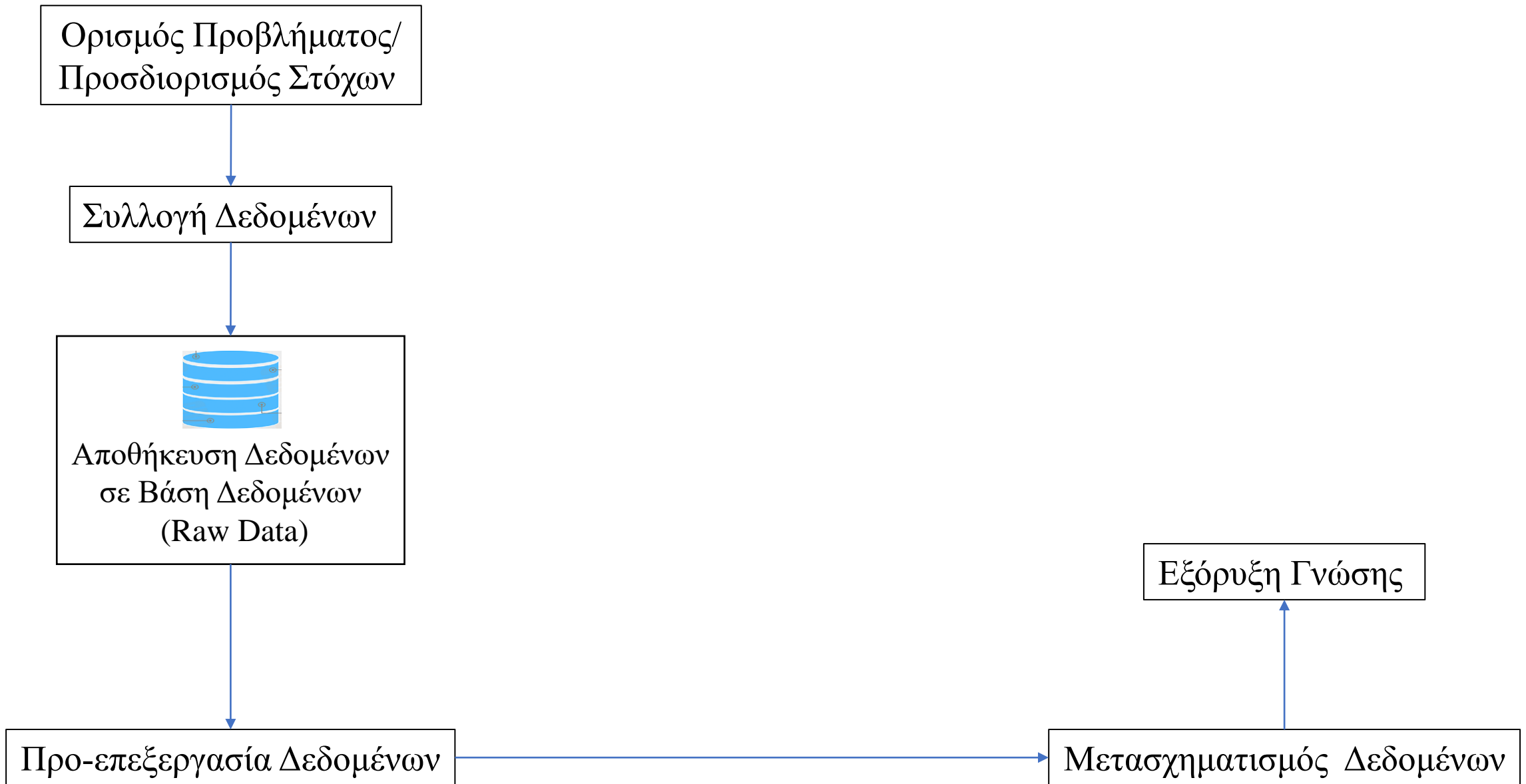
Τυπικό Μοντέλο Εξόρυξης Δεδομένων



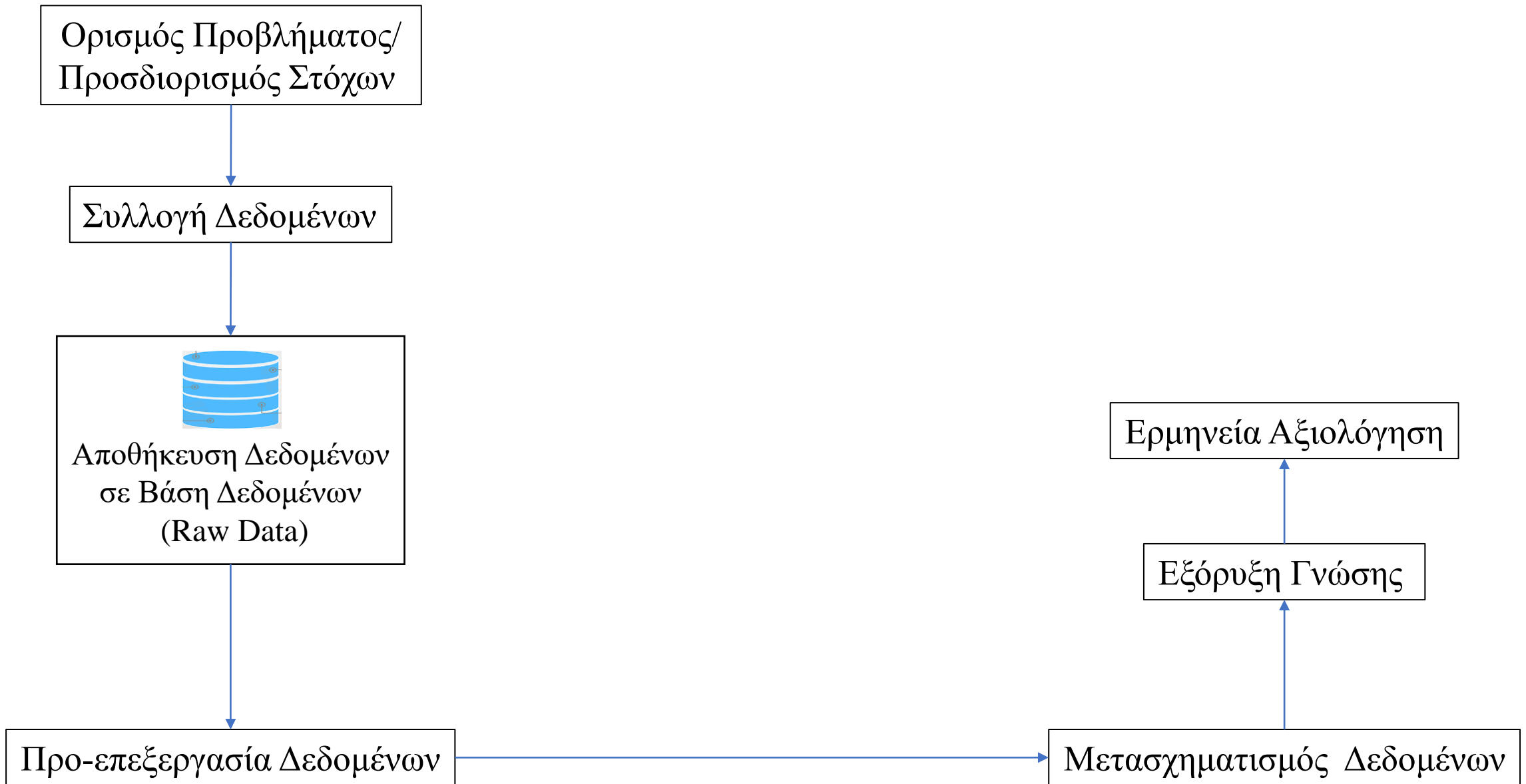
Τυπικό Μοντέλο Εξόρυξης Δεδομένων



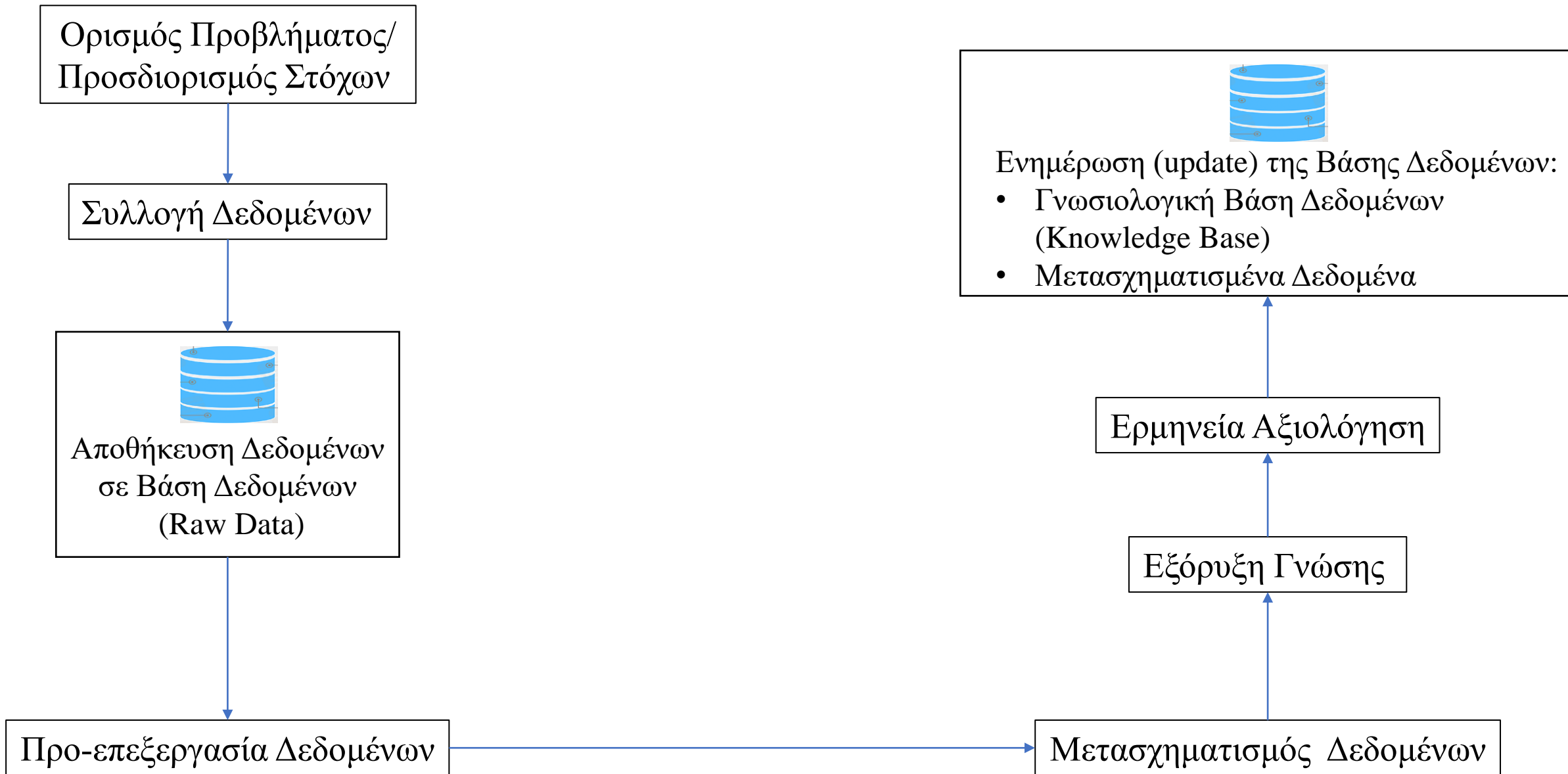
Τυπικό Μοντέλο Εξόρυξης Δεδομένων



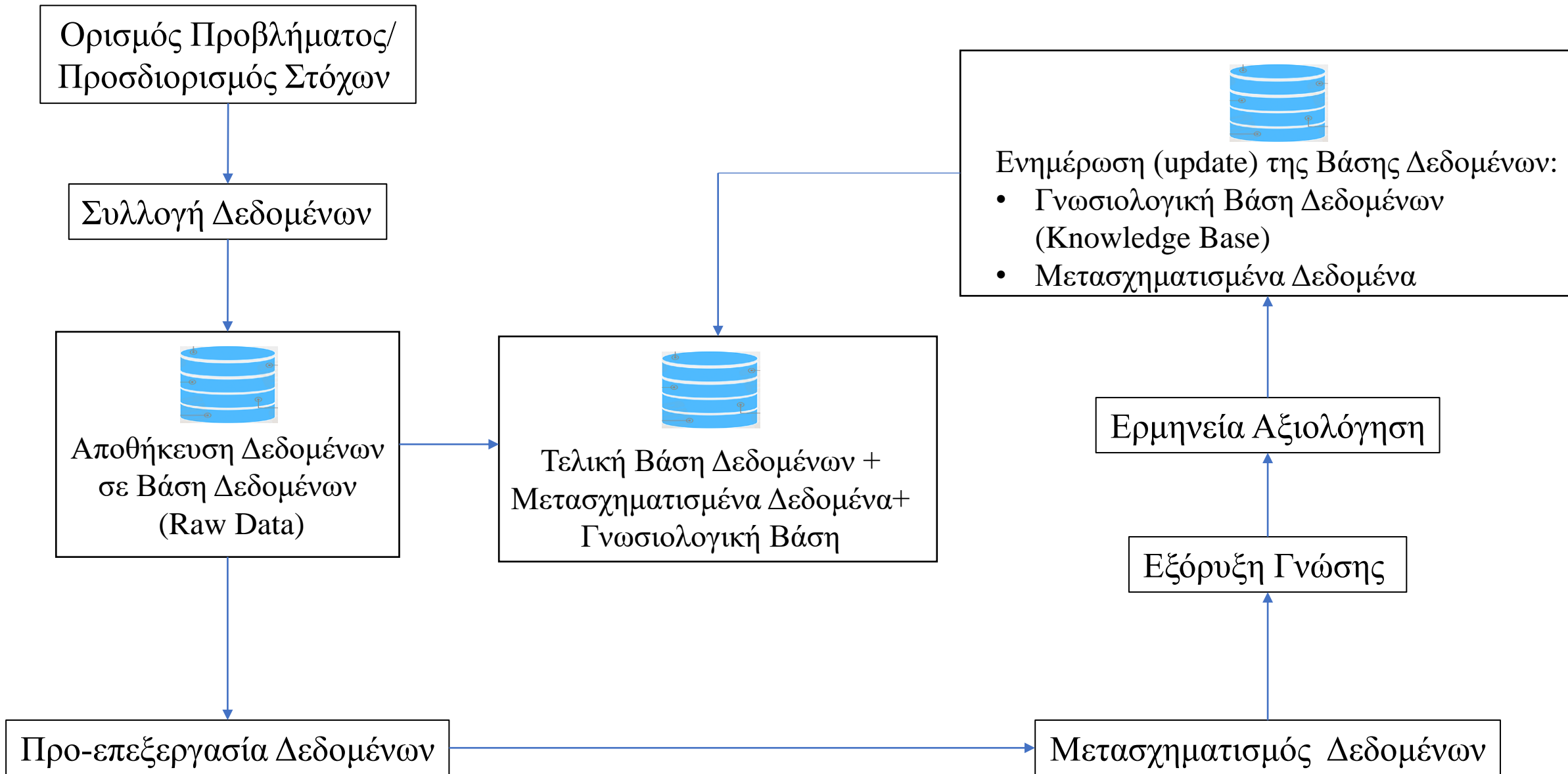
Τυπικό Μοντέλο Εξόρυξης Δεδομένων



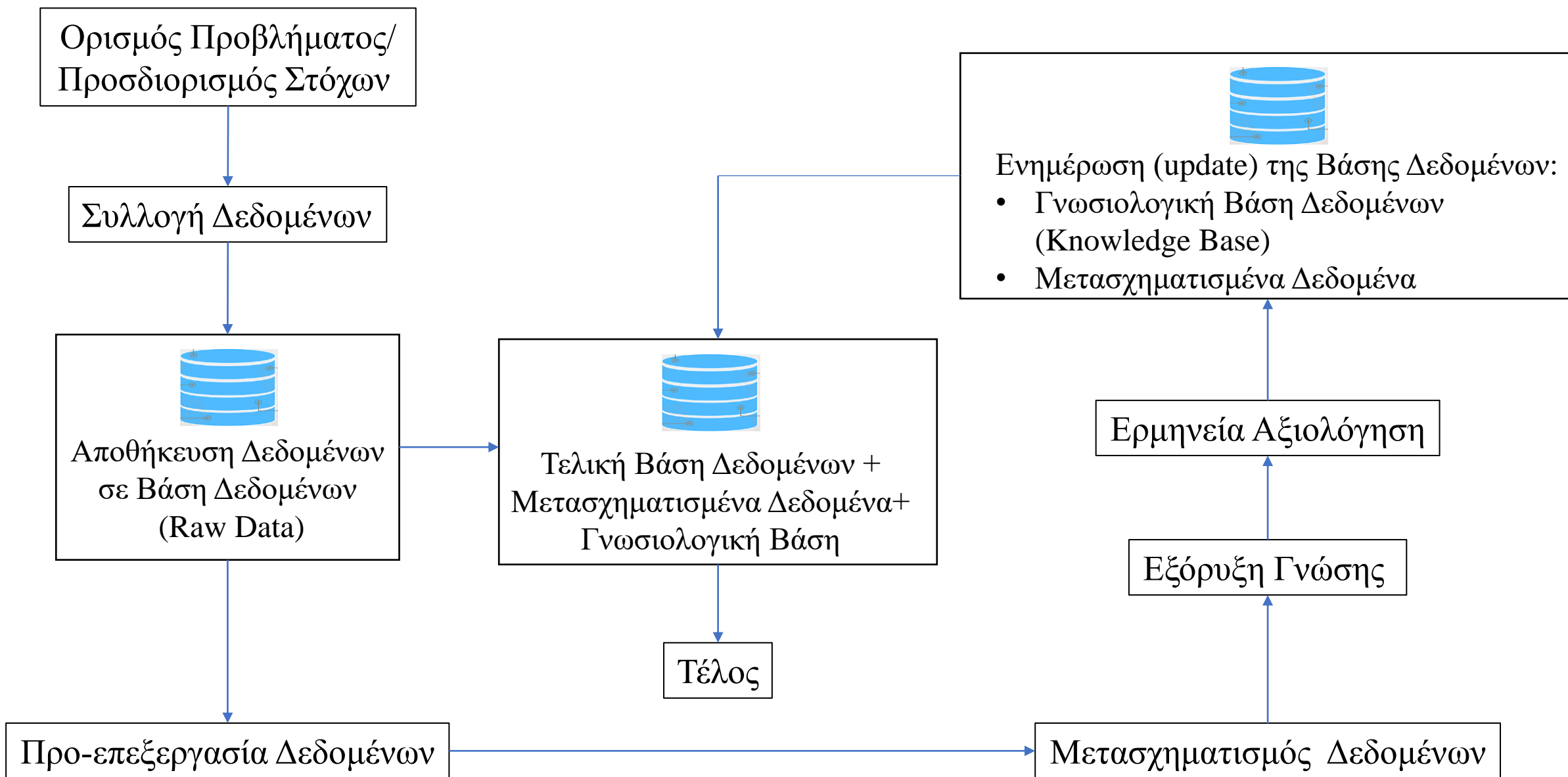
Τυπικό Μοντέλο Εξόρυξης Δεδομένων



Τυπικό Μοντέλο Εξόρυξης Δεδομένων



Τυπικό Μοντέλο Εξόρυξης Δεδομένων



Τυπικό Μοντέλο Εξόρυξης Δεδομένων

A. Ορισμός Προβλήματος/Προσδιορισμός Στόχων

- Σαφής δήλωση του προβλήματος

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Α. Ορισμός Προβλήματος/Προσδιορισμός Στόχων

- Σαφής δήλωση του προβλήματος
- Κριτήρια που πρέπει να ικανοποιηθούν

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Α. Ορισμός Προβλήματος/Προσδιορισμός Στόχων

- Σαφής δήλωση του προβλήματος
- Κριτήρια που πρέπει να ικανοποιηθούν
- Προσδιορισμός πιθανών μεθόδων/αλγόριθμων Εξόρυξης Δεδομένων

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Α. Ορισμός Προβλήματος/Προσδιορισμός Στόχων

- Σαφής δήλωση του προβλήματος
- Κριτήρια που πρέπει να ικανοποιηθούν
- Προσδιορισμός πιθανών μεθόδων/αλγόριθμων Εξόρυξης Δεδομένων
- Εκτίμηση κόστους του έργου

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Α. Ορισμός Προβλήματος/Προσδιορισμός Στόχων

- Σαφής δήλωση του προβλήματος
- Κριτήρια που πρέπει να ικανοποιηθούν
- Προσδιορισμός πιθανών μεθόδων/αλγόριθμων Εξόρυξης Δεδομένων
- Εκτίμηση κόστους του έργου
- Χρονοδιάγραμμα υλοποίησης έργου

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Α. Ορισμός Προβλήματος/Προσδιορισμός Στόχων

- Σαφής δήλωση του προβλήματος
- Κριτήρια που πρέπει να ικανοποιηθούν
- Προσδιορισμός πιθανών μεθόδων/αλγόριθμων Εξόρυξης Δεδομένων
- Εκτίμηση κόστους του έργου
- Χρονοδιάγραμμα υλοποίησης έργου
- Επιλογή ειδικών που θα ασχοληθούν με την υλοποίηση του έργου

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Α. Ορισμός Προβλήματος/Προσδιορισμός Στόχων

- Σαφής δήλωση του προβλήματος
- Κριτήρια που πρέπει να ικανοποιηθούν
- Προσδιορισμός πιθανών μεθόδων/αλγόριθμων Εξόρυξης Δεδομένων
- Εκτίμηση κόστους του έργου
- Χρονοδιάγραμμα υλοποίησης έργου
- Επιλογή ειδικών που θα ασχοληθούν με την υλοποίηση του έργου
- Διαχειριστικά θέματα που μπορεί να προκύψουν

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Α. Ορισμός Προβλήματος/Προσδιορισμός Στόχων

- Σαφής δήλωση του προβλήματος
- Κριτήρια που πρέπει να ικανοποιηθούν
- Προσδιορισμός πιθανών μεθόδων/αλγόριθμων Εξόρυξης Δεδομένων
- Εκτίμηση κόστους του έργου
- Χρονοδιάγραμμα υλοποίησης έργου
- Επιλογή ειδικών που θα ασχοληθούν με την υλοποίηση του έργου
- Διαχειριστικά θέματα που μπορεί να προκύψουν
- Θέματα ασφάλειας προσωπικών δεδομένων

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Α. Ορισμός Προβλήματος/Προσδιορισμός Στόχων

- Σαφής δήλωση του προβλήματος
- Κριτήρια που πρέπει να ικανοποιηθούν
- Προσδιορισμός πιθανών μεθόδων/αλγόριθμων Εξόρυξης Δεδομένων
- Εκτίμηση κόστους του έργου
- Χρονοδιάγραμμα υλοποίησης έργου
- Επιλογή ειδικών που θα ασχοληθούν με την υλοποίηση του έργου
- Διαχειριστικά θέματα που μπορεί να προκύψουν
- Θέματα ασφάλειας προσωπικών δεδομένων
- Νομικά ζητήματα που μπορεί να προκύψουν

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Β. Συλλογή Δεδομένων

- Η συλλογή δεδομένων είναι η διεργασία κατά την οποία συλλέγεται πληροφορία, η οποία ποσοτικοποιείται (συνήθως με αριθμητικά η/και κατηγορικά δεδομένα) και αποθηκεύεται σε βάσεις δεδομένων με συγκεκριμένη δομή που καθορίζεται από το εκάστοτε πρόβλημα

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Β. Συλλογή Δεδομένων

- Η συλλογή δεδομένων είναι η διεργασία κατά την οποία συλλέγεται πληροφορία, η οποία ποσοτικοποιείται (συνήθως με αριθμητικά η/και κατηγορικά δεδομένα) και αποθηκεύεται σε βάσεις δεδομένων με συγκεκριμένη δομή που καθορίζεται από το εκάστοτε πρόβλημα
- Τρόποι συλλογής δεδομένων
 - Εκτέλεση πειραμάτων (Φυσική, Χημεία, Βιολογία, Ιατρική, Επιστήμες Μηχανικού, Κοινωνικές Επιστήμες, κλπ)

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Β. Συλλογή Δεδομένων

- Η συλλογή δεδομένων είναι η διεργασία κατά την οποία συλλέγεται πληροφορία, η οποία ποσοτικοποιείται (συνήθως με αριθμητικά η/και κατηγορικά δεδομένα) και αποθηκεύεται σε βάσεις δεδομένων με συγκεκριμένη δομή που καθορίζεται από το εκάστοτε πρόβλημα
- Τρόποι συλλογής δεδομένων
 - Εκτέλεση πειραμάτων (Φυσική, Χημεία, Βιολογία, Ιατρική, Επιστήμες Μηχανικού, Κοινωνικές Επιστήμες, κλπ)
 - Ερωτηματολόγια

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Β. Συλλογή Δεδομένων

- Η συλλογή δεδομένων είναι η διεργασία κατά την οποία συλλέγεται πληροφορία, η οποία ποσοτικοποιείται (συνήθως με αριθμητικά η/και κατηγορικά δεδομένα) και αποθηκεύεται σε βάσεις δεδομένων με συγκεκριμένη δομή που καθορίζεται από το εκάστοτε πρόβλημα
- Τρόποι συλλογής δεδομένων
 - Εκτέλεση πειραμάτων (Φυσική, Χημεία, Βιολογία, Ιατρική, Επιστήμες Μηχανικού, Κοινωνικές Επιστήμες, κλπ)
 - Ερωτηματολόγια
 - Παρατηρησιακά δεδομένα (Αστρονομικά Φαινόμενα, Κοινωνικά Φαινόμενα, κλπ)

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Β. Συλλογή Δεδομένων

- Η συλλογή δεδομένων είναι η διεργασία κατά την οποία συλλέγεται πληροφορία, η οποία ποσοτικοποιείται (συνήθως με αριθμητικά η/και κατηγορικά δεδομένα) και αποθηκεύεται σε βάσεις δεδομένων με συγκεκριμένη δομή που καθορίζεται από το εκάστοτε πρόβλημα
- Τρόποι συλλογής δεδομένων
 - Εκτέλεση πειραμάτων (Φυσική, Χημεία, Βιολογία, Ιατρική, Επιστήμες Μηχανικού, Κοινωνικές Επιστήμες, κλπ)
 - Ερωτηματολόγια
 - Παρατηρησιακά δεδομένα (Αστρονομικά Φαινόμενα, Κοινωνικά Φαινόμενα, κλπ)
 - Διαδικτυακά Δεδομένα (Μέσα κοινωνικής δικτύωσης, πολυμεσικά δεδομένα, διαδικτυακές βάσεις δεδομένων, κλπ)

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Β. Συλλογή Δεδομένων

- Η συλλογή δεδομένων είναι η διεργασία κατά την οποία συλλέγεται πληροφορία, η οποία ποσοτικοποιείται (συνήθως με αριθμητικά η/και κατηγορικά δεδομένα) και αποθηκεύεται σε βάσεις δεδομένων με συγκεκριμένη δομή που καθορίζεται από το εκάστοτε πρόβλημα
- Τρόποι συλλογής δεδομένων
 - Εκτέλεση πειραμάτων (Φυσική, Χημεία, Βιολογία, Ιατρική, Επιστήμες Μηχανικού, Κοινωνικές Επιστήμες, κλπ)
 - Ερωτηματολόγια
 - Παρατηρησιακά δεδομένα (Αστρονομικά Φαινόμενα, Κοινωνικά Φαινόμενα, κλπ)
 - Διαδικτυακά Δεδομένα (Μέσα κοινωνικής δικτύωσης, πολυμεσικά δεδομένα, διαδικτυακές βάσεις δεδομένων, κλπ)
- *Παράδειγμα: Συλλογή Διαδικτυακών Δεδομένων*
 - Ένας server ή ένας υπολογιστής συλλέγει δεδομένα από άλλους υπολογιστές

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Β. Συλλογή Δεδομένων

- Η συλλογή δεδομένων είναι η διεργασία κατά την οποία συλλέγεται πληροφορία, η οποία ποσοτικοποιείται (συνήθως με αριθμητικά η/και κατηγορικά δεδομένα) και αποθηκεύεται σε βάσεις δεδομένων με συγκεκριμένη δομή που καθορίζεται από το εκάστοτε πρόβλημα
- Τρόποι συλλογής δεδομένων
 - Εκτέλεση πειραμάτων (Φυσική, Χημεία, Βιολογία, Ιατρική, Επιστήμες Μηχανικού, Κοινωνικές Επιστήμες, κλπ)
 - Ερωτηματολόγια
 - Παρατηρησιακά δεδομένα (Αστρονομικά Φαινόμενα, Κοινωνικά Φαινόμενα, κλπ)
 - Διαδικτυακά Δεδομένα (Μέσα κοινωνικής δικτύωσης, πολυμεσικά δεδομένα, διαδικτυακές βάσεις δεδομένων, κλπ)
- *Παράδειγμα: Συλλογή Διαδικτυακών Δεδομένων*
 - Ένας server ή ένας υπολογιστής συλλέγει δεδομένα από άλλους υπολογιστές
 - Συλλογή Δεδομένων με την χρήση μηχανών αναζήτησης

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Γ. Αποθήκευση σε Βάσεις Δεδομένων

- Προσδιορισμός των Ιδιοτήτων/Μεταβλητών που μας περιγράφουν αποδοτικά το πρόβλημα (αυτό σημαίνει ότι κάποιες από τις αρχικές Ιδιότητες/Μεταβλητές, οι οποίες συμμετείχαν στην συλλογή των δεδομένων θα διαγραφούν)

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Γ. Αποθήκευση σε Βάσεις Δεδομένων

- Προσδιορισμός των Ιδιοτήτων/Μεταβλητών που μας περιγράφουν αποδοτικά το πρόβλημα (αυτό σημαίνει ότι κάποιες από τις αρχικές Ιδιότητες/Μεταβλητές, οι οποίες συμμετείχαν στην συλλογή των δεδομένων θα διαγραφούν)
- Feature Extraction: Αν δύο ή περισσότερες Μεταβλητές είναι ισχυρά συσχετισμένες (cross-correlated) τότε κρατάμε μόνο την μία από αυτές (οι υπόλοιπες θα διαγραφούν). Αυτό πρέπει να γίνει για όλα τα γκρουπ μεταβλητών που είναι συσχετισμένες. Η όλη διαδικασία αποτελεί ολόκληρο τομέα στην εξόρυξη δεδομένων και γίνεται με μεθόδους Στατιστικής. Έτσι έχουμε σημαντική μείωση του υπολογιστικού κόστους

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Γ. Αποθήκευση σε Βάσεις Δεδομένων

- Προσδιορισμός των Ιδιοτήτων/Μεταβλητών που μας περιγράφουν αποδοτικά το πρόβλημα (αυτό σημαίνει ότι κάποιες από τις αρχικές Ιδιότητες/Μεταβλητές, οι οποίες συμμετείχαν στην συλλογή των δεδομένων θα διαγραφούν)
- Feature Extraction: Αν δύο ή περισσότερες Μεταβλητές είναι ισχυρά συσχετισμένες (cross-correlated) τότε κρατάμε μόνο την μία από αυτές (οι υπόλοιπες θα διαγραφούν). Αυτό πρέπει να γίνει για όλα τα γκρουπ μεταβλητών που είναι συσχετισμένες. Η όλη διαδικασία αποτελεί ολόκληρο τομέα στην εξόρυξη δεδομένων και γίνεται με μεθόδους Στατιστικής. Έτσι έχουμε σημαντική μείωση του υπολογιστικού κόστους
- Επιλογή του Σχήματος Βάσης Δεδομένων και δημιουργία Σχεσιακής Βάσης Δεδομένων

Σχήμα Βάσης Δεδομένων + Σχεσιακό Μοντέλο = Σχεσιακή Βάση Δεδομένων

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Δ. Προ-Επεξεργασία των Δεδομένων

Καθαρισμός των Δεδομένων που προέκυψαν από το προηγούμενο βήμα

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Δ. Προ-Επεξεργασία των Δεδομένων

Καθαρισμός των Δεδομένων που προέκυψαν από το προηγούμενο βήμα

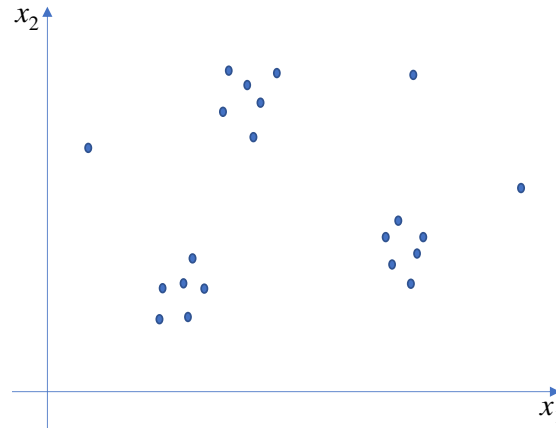
- Είναι δυνατόν να υπάρξουν και συσχετισμένες πλειάδες. Ομοίως με την προηγούμενη περίπτωση κρατάμε μόνο μία πλειάδα από ένα γκρουπ συσχετισμένων πλειάδων. Εδώ υπάγεται και η περίπτωση των διπλότυπων πλειάδων (εγγραφών)

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Δ. Προ-Επεξεργασία των Δεδομένων

Καθαρισμός των Δεδομένων που προέκυψαν από το προηγούμενο βήμα

- Είναι δυνατόν να υπάρξουν και συσχετισμένες πλειάδες. Ομοίως με την προηγούμενη περίπτωση κρατάμε μόνο μία πλειάδα από ένα γκρουπ συσχετισμένων πλειάδων. Εδώ υπάγεται και η περίπτωση των διπλότυπων πλειάδων (εγγραφών)
- Εξάλειψη Θορύβου
 - Διαγραφή των outliers δεδομένων



Τυπικό Μοντέλο Εξόρυξης Δεδομένων

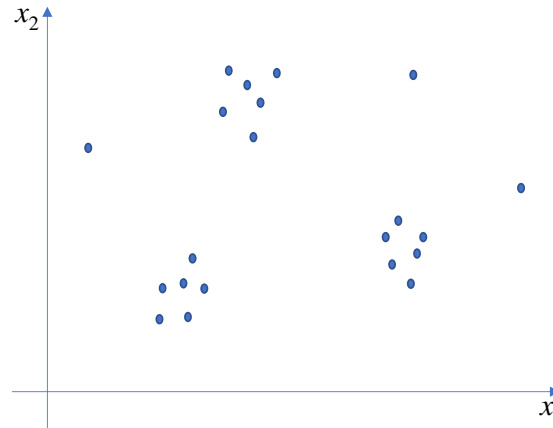
Α. Προ-Επεξεργασία των Δεδομένων

Καθαρισμός των Δεδομένων που προέκυψαν από το προηγούμενο βήμα

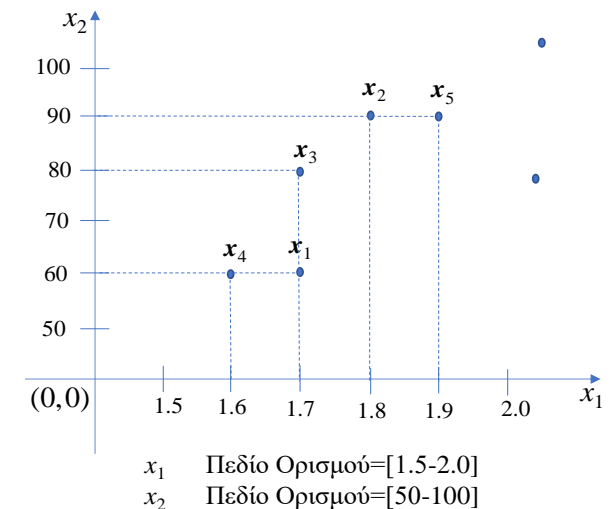
- Είναι δυνατόν να υπάρξουν και συσχετισμένες πλειάδες. Ομοίως με την προηγούμενη περίπτωση κρατάμε μόνο μία πλειάδα από ένα γκρουπ συσχετισμένων πλειάδων. Εδώ υπάγεται και η περίπτωση των διπλότυπων πλειάδων (εγγραφών)

- Εξάλειψη Θορύβου

➤ Διαγραφή των outliers δεδομένων



➤ Διαγραφή δεδομένων που βρίσκονται έξω από τα πεδία ορισμού



Τυπικό Μοντέλο Εξόρυξης Δεδομένων

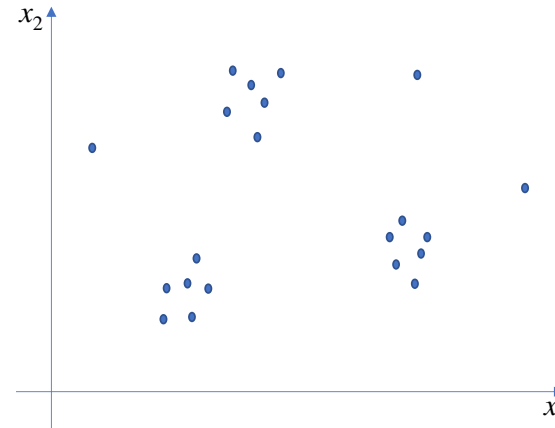
Δ. Προ-Επεξεργασία των Δεδομένων

Καθαρισμός των Δεδομένων που προέκυψαν από το προηγούμενο βήμα

- Είναι δυνατόν να υπάρξουν και συσχετισμένες πλειάδες. Ομοίως με την προηγούμενη περίπτωση κρατάμε μόνο μία πλειάδα από ένα γκρουπ συσχετισμένων πλειάδων. Εδώ υπάγεται και η περίπτωση των διπλότυπων πλειάδων (εγγραφών)

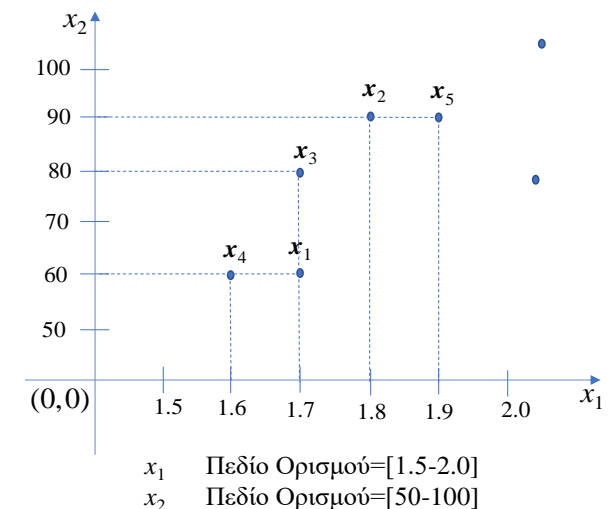
- Εξάλειψη Θορύβου

➤ Διαγραφή των outliers δεδομένων



➤ Διαγραφή δεδομένων που βρίσκονται έξω από τα πεδία ορισμού

➤ Διαγραφή πλειάδων που λείπουν εντελώς ή λείπουν οι τιμές τους σε κάποιες Μεταβλητές



Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Ε. Μετασχηματισμός των Δεδομένων

- Κανονικοποίηση Δεδομένων

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Ε. Μετασχηματισμός των Δεδομένων

- Κανονικοποίηση Δεδομένων
- Αλλαγή κλίμακας των δεδομένων

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Ε. Μετασχηματισμός των Δεδομένων

- Κανονικοποίηση Δεδομένων
- Αλλαγή κλίμακας των δεδομένων
- Περαιτέρω επιλογή των Ιδιοτήτων που έχουν νόημα (Feature Extraction)

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Ε. Μετασχηματισμός των Δεδομένων

- Κανονικοποίηση Δεδομένων
- Αλλαγή κλίμακας των δεδομένων
- Περαιτέρω επιλογή των Ιδιοτήτων που έχουν νόημα (Feature Extraction)
- Περιστροφή Δεδομένων

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Ε. Μετασχηματισμός των Δεδομένων

- Κανονικοποίηση Δεδομένων
- Αλλαγή κλίμακας των δεδομένων
- Περαιτέρω επιλογή των Ιδιοτήτων που έχουν νόημα (Feature Extraction)
- Περιστροφή Δεδομένων
- Αλλαγή του συστήματος Συντεταγμένων

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Ε. Μετασχηματισμός των Δεδομένων

- Κανονικοποίηση Δεδομένων
- Αλλαγή κλίμακας των δεδομένων
- Περαιτέρω επιλογή των Ιδιοτήτων που έχουν νόημα (Feature Extraction)
- Περιστροφή Δεδομένων
- Αλλαγή του συστήματος Συντεταγμένων
- Μετατροπή συμβόλων και λέξεων σε Κατηγορικά Δεδομένα

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

ΣΤ. Εξόρυξη Δεδομένων

Επιλογή της Μεθόδου Εξόρυξης Γνώσης

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

ΣΤ. Εξόρυξη Δεδομένων

Επιλογή της Μεθόδου Εξόρυξης Γνώσης



Μη-Καθοδηγούμενοι Αλγόριθμοι

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

ΣΤ. Εξόρυξη Δεδομένων

Επιλογή της Μεθόδου Εξόρυξης Γνώσης

Μη-Καθοδηγούμενοι Αλγόριθμοι

Αλγόριθμοι Συσταδοποίησης
(k-Means, Hierarchical Clustering,
Agglomerative Clustering, Fuzzy k-
Means)

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

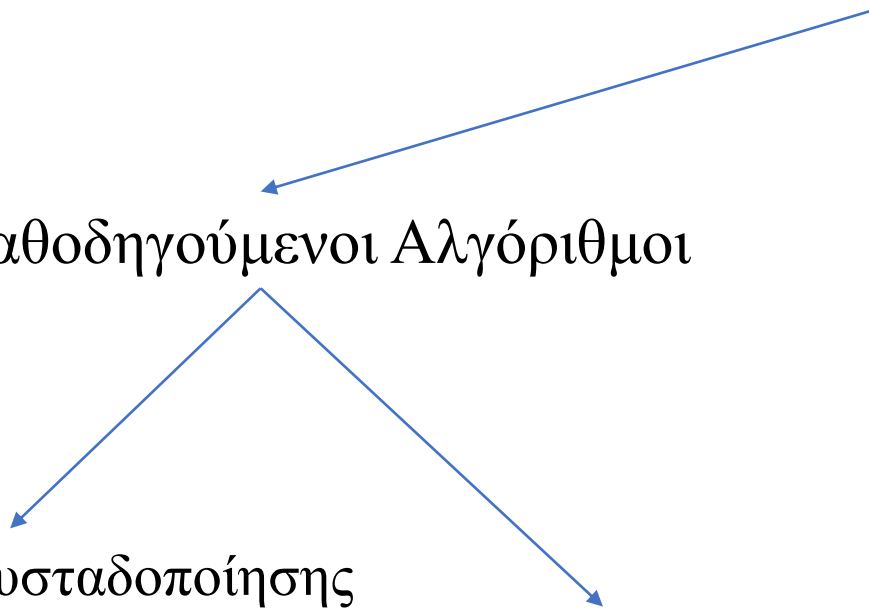
ΣΤ. Εξόρυξη Δεδομένων

Επιλογή της Μεθόδου Εξόρυξης Γνώσης

Μη-Καθοδηγούμενοι Αλγόριθμοι

Αλγόριθμοι Συσταδοποίησης
(k-Means, Hierarchical Clustering,
Agglomerative Clustering, Fuzzy k-
Means)

Δέντρα



Τυπικό Μοντέλο Εξόρυξης Δεδομένων

ΣΤ. Εξόρυξη Δεδομένων

Επιλογή της Μεθόδου Εξόρυξης Γνώσης

Μη-Καθοδηγούμενοι Αλγόριθμοι

Καθοδηγούμενοι Αλγόριθμοι

Αλγόριθμοι Συσταδοποίησης
(k-Means, Hierarchical Clustering,
Agglomerative Clustering, Fuzzy k-
Means)

Δέντρα

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

ΣΤ. Εξόρυξη Δεδομένων

Επιλογή της Μεθόδου Εξόρυξης Γνώσης

Μη-Καθοδηγούμενοι Αλγόριθμοι

Καθοδηγούμενοι Αλγόριθμοι

Αλγόριθμοι Συσταδοποίησης
(k-Means, Hierarchical Clustering,
Agglomerative Clustering, Fuzzy k-
Means)

Δέντρα

Αλγόριθμοι Κατηγοριοποίησης
(Support Vector Machines-SVM,
Linear Regression)

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

ΣΤ. Εξόρυξη Δεδομένων

Επιλογή της Μεθόδου Εξόρυξης Γνώσης

Μη-Καθοδηγούμενοι Αλγόριθμοι

Καθοδηγούμενοι Αλγόριθμοι

Αλγόριθμοι Συσταδοποίησης
(k-Means, Hierarchical Clustering,
Agglomerative Clustering, Fuzzy k-
Means)

Δέντρα

Αλγόριθμοι Κατηγοριοποίησης
(Support Vector Machines-SVM,
Linear Regression)

Neural Networks

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Z. Ερμηνεία-Αξιολόγηση

Ερμηνεία και Αξιολόγηση των
Αποτελεσμάτων της Μεθόδου Εξόρυξης

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Z. Ερμηνεία-Αξιολόγηση

Ερμηνεία και Αξιολόγηση των
Αποτελεσμάτων της Μεθόδου Εξόρυξης

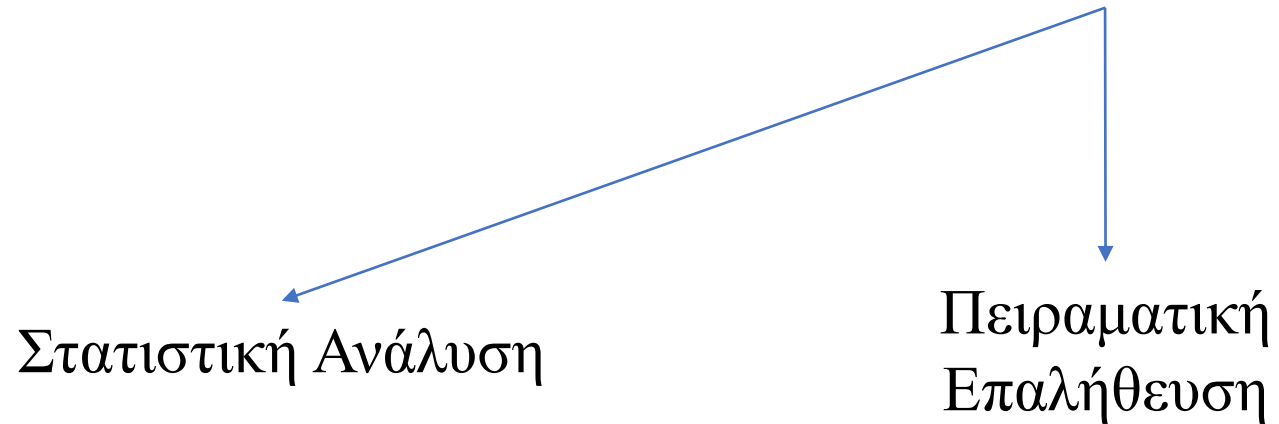


Στατιστική Ανάλυση

Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Z. Ερμηνεία-Αξιολόγηση

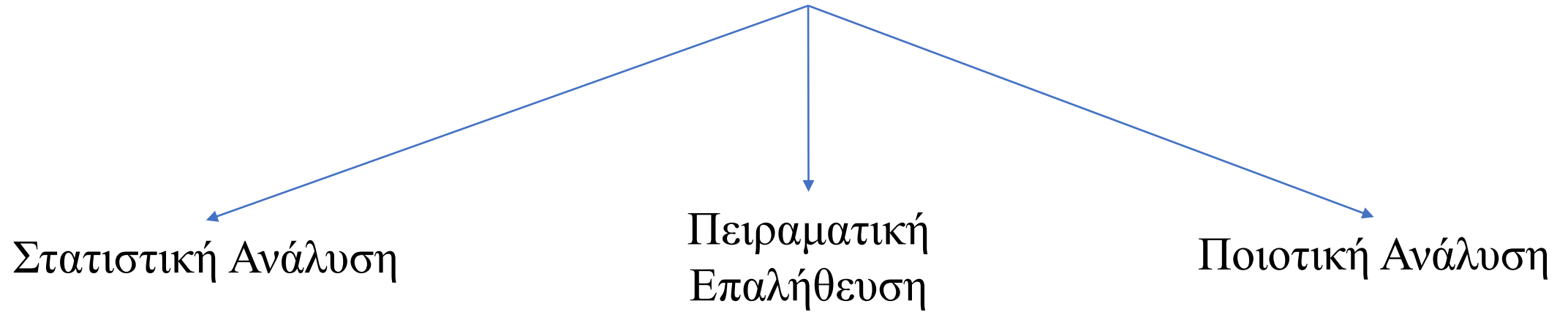
Ερμηνεία και Αξιολόγηση των
Αποτελεσμάτων της Μεθόδου Εξόρυξης



Τυπικό Μοντέλο Εξόρυξης Δεδομένων

Z. Ερμηνεία-Αξιολόγηση

Ερμηνεία και Αξιολόγηση των
Αποτελεσμάτων της Μεθόδου Εξόρυξης



ΚΑΛΟ ΑΠΟΓΕΥΜΑ