

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ · ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

# ΕΞΟΥΞΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ ΜΑΘΗΣΗΣ

ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ · 2023–2024

**ΠΕΡΙΕΧΟΜΕΝΑ**

<b>1</b>	<b>ΕΡΩΤΗΜΑ 1</b>	<b>2</b>
1.1	ΑΝΑΛΥΣΗ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ . . . . .	2
1.2	ΓΡΑΦΙΚΕΣ ΠΑΡΑΣΤΑΣΕΙΣ . . . . .	2

## 1 ΕΡΩΤΗΜΑ 1

Για την εισαγωγή και τη προεπεξεργασία του `.csv` αρχείου, θα χρησιμοποιήσουμε τη βιβλιοθήκη `pandas` της Python.

### 1.1 ΑΝΑΛΥΣΗ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ

Καταρχάς, χρησιμοποιώντας την συνάρτηση `head()` μπορούμε να δούμε τις πρώτες 5 εγγραφές από το σύνολο δεδομένων μας:

	timestamp	back_x	back_y	back_z	thigh_x	thigh_y	thigh_z	label
0	2019-01-12 00:00:00.000	-0.760242	0.299570	0.468570	-5.092732	-0.298644	0.709439	6
1	2019-01-12 00:00:00.010	-0.530138	0.281880	0.319987	0.900547	0.286944	0.340309	6
2	2019-01-12 00:00:00.020	-1.170922	0.186353	-0.167010	-0.035442	-0.078423	-0.515212	6
3	2019-01-12 00:00:00.030	-0.648772	0.016579	-0.054284	-1.554248	-0.950978	-0.221140	6
4	2019-01-12 00:00:00.040	-0.355071	-0.051831	-0.113419	-0.547471	0.140903	-0.653782	6

Σε συνδυασμό με την `info()`, παρατηρούμε πώς για κάθε χρονική στιγμή δίνονται οι τιμές των αισθητήρων, αποθηκευμένες ως `float24`, στις τρεις διαστάσεις (x, y, z) για τις περιοχές της πλάτης και του μηρού, καθώς και ένα `int64` label.

Μέσω των `isna().sum()` συναρτήσεων παρατηρούμε πώς δεν υπάρχουν NaN τιμές στο σύνολο δεδομένων.

Χρησιμοποιώντας τη συνάρτηση `describe()` μπορούμε να υπολογίσουμε βασικές στατιστικές μετρικές για τα δεδομένα μας:

### 1.2 ΓΡΑΦΙΚΕΣ ΠΑΡΑΣΤΑΣΕΙΣ