

Η Περαιτέρω Χρήση Πληροφοριών του Δημόσιου Τομέα

Εκπαίδευση Λειτουργών - Συνδέσμων

3. Ετοιμασία και Μορφοποίηση Συνόλων Δεδομένων για
δημοσίευση στην Εθνική Πύλη Ανοικτών Δεδομένων

Μάρτιος 2024



- Εισαγωγή
- Ετοιμασία Μηχαναγνώσιμου Αρχείου Δεδομένων
- Καλές Πρακτικές Μορφοποίησης
- Παραδείγματα
- Μετατροπή Αρχείου Excel σε CSV

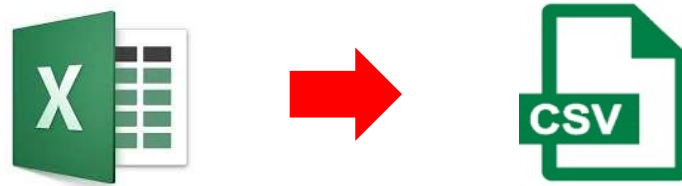
Για να καταστεί δυνατή η διάθεση ενός συνόλου δεδομένων κατά τρόπο που αυτό να μπορεί να επαναχρησιμοποιηθεί από εξωτερικές εφαρμογές (applications), το σύνολο δεδομένων θα πρέπει να εισαχθεί στο **datastore** της Πύλης.

Εισάγοντας ένα σύνολο δεδομένων στο datastore, λαμβάνουν χώρα μια σειρά εσωτερικών διεργασιών, οι οποίες επιτρέπουν τη διάθεση των δεδομένων στο κοινό μέσω **API**.

Για να καταστεί αυτό δυνατό, θα πρέπει τα δεδομένα να δημοσιευθούν στην Πύλη, σε μορφή **CSV**.

Μηχαναγνώσιμα είναι τα δεδομένα σε μορφότυπο που μπορεί να αναγνωστεί και ερμηνευτεί αυτόματα από ένα πρόγραμμα υπολογιστή, όπως για παράδειγμα δεδομένα οργανωμένα σε πίνακα με στήλες, ο οποίος μπορεί να εξαχθεί σαν αρχείο CSV (comma separated value).

Ετοιμασία Μηχαναγνώσιμου Αρχείου Δεδομένων

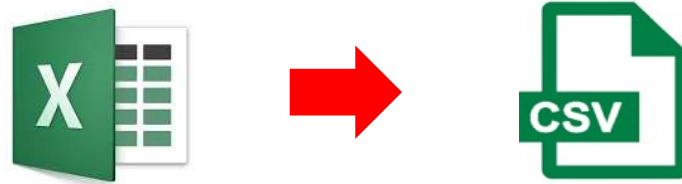


Η μετατροπή ενός αρχείου Excel σε CSV δεν απαιτεί πέραν των 2 απλών χειρισμών (Save As CSV και καθορισμό encoding σε UTF8). Προτού όμως γίνει αυτό, τα δεδομένα που περιλαμβάνονται στο αρχείο Excel, θα πρέπει να οργανωθούν σε ενιαίο μηχαναγνώσιμο πίνακα, ακολουθώντας μια σειρά από κανόνες:

Υποχρεωτικά

- ✓ Η κάθε στήλη έχει όνομα (πρώτη γραμμή ενός πίνακα = ονόματα στηλών) (Παραδείγματα 1 και 3)
- ✓ Τα ονόματα των στηλών είναι στα αγγλικά ή με λατινικούς χαρακτήρες (Παράδειγμα 1)
- ✓ Οι τιμές των πεδίων είναι μηχαναγνώσιμες (Παράδειγμα 1)

Ετοιμασία Μηχαναγνώσιμου Αρχείου Δεδομένων



Δεν επιτρέπονται

- X** Ενοποιημένα κελιά (Merged cells) (Παράδειγμα 3)
- X** Πολλαπλοί Πίνακες (Παράδειγμα 3)
- X** Υποσημειώσεις/ περιγραφές (περιλάβετε τις πληροφορίες αυτές στα μεταδεδομένα) (Παράδειγμα 1)
- X** Κενές γραμμές ή στήλες στον πίνακα (Παραδείγματα 1 και 3)
- X** Σύνολα τιμών (sum of values) των γραμμών

Ετοιμασία Μηχαναγνώσιμου Αρχείου Δεδομένων



Καλές Πρακτικές Μορφοποίησης:

- ✓ Οι ημερομηνίες (πχ έτος, ημερομηνία) να περιλαμβάνονται σε μία στήλη (Παράδειγμα 2)
- ✓ Οι τιμές κάθε αριθμητικού πεδίου (πχ αρ. εργοδοτούμενων, ασφαλιστές αποδοχές) θα πρέπει να βρίσκονται σε μία μοναδική στήλη (Παραδείγματα 1 και 3)
- ✓ Ο πίνακας δεδομένων είναι καλύτερα να περιλαμβάνει όσο το δυνατό λιγότερες στήλες και, κατά προτίμηση, όχι περισσότερες από 30 (Παράδειγμα 2)
- ✓ Χρησιμοποιήστε τις γραμμές (records) για να περιλάβετε όλα τα δεδομένα μίας εγγραφής (πχ ένα πρόσωπο/ αντικείμενο/ στατιστική μονάδα/ έτος, κλπ) και τις στήλες σαν αριθμητικά πεδία. (Παραδείγματα 2 και 4)

Παράδειγμα 1

Παρουσίαση του τρόπου «καθαρισμού» ενός συνόλου δεδομένων. Σε πρώτη φάση, αφαιρέθηκαν οι κενές γραμμές και στήλες στον πίνακα δεδομένων και, στη συνέχεια, δημιουργήθηκε νέα ενιαία στήλη για την ημερομηνία, σε μηχαναγνώσιμο μορφότυπο, η οποία αντικατέστησε τις αρχικές στήλες με τον μήνα και το έτος.

Πίνακας 1: Δεδομένα πριν τον καθαρισμό			Πίνακας 2: Δεδομένα μετά τον καθαρισμό			Πίνακας 3: Δεδομένα με ημερομηνία				
Έτος	Μήνας	Λευκωσία	Πάφος	Έτος	Μήνας	Λευκωσία	Πάφος	Month	Lefkosia	Pafos
2015	Ιανουάριος	12.1	13.2	2015	Ιανουάριος	12.1	13.2	1/1/2015	12.1	13.2
	Φεβρουάριος	13.5	14.1		Φεβρουάριος	13.5	14.1	1/2/2015	13.5	14.1
	Μάρτιος	18.2	19.2		Μάρτιος	18.2	19.2	1/3/2015	18.2	19.2
	ΑΠΡ	23.1	23.2		ΑΠΡ	23.1	23.2	1/4/2015	23.1	23.2
	Μάιος	26.2	26.5		Μάιος	26.2	26.5	1/5/2015	26.2	26.5
	Ιούνιος	30.1	29.6		Ιούνιος	30.1	29.6	1/6/2015	30.1	29.6
	Ιούλιος	35.1	32.5		Ιούλιος	35.1	32.5	1/7/2015	35.1	32.5
	Αύγουστος	34.7	32.4		Αύγουστος	34.7	32.4	1/8/2015	34.7	32.4
	Σεπτέμβριος	30.5	29.9		Σεπτέμβριος	30.5	29.9	1/9/2015	30.5	29.9
	Οκτώβριος	26.6	26.8		Οκτώβριος	26.6	26.8	1/10/2015	26.6	26.8
	Νοέμβριος	23.9	24.5		Νοέμβριος	23.9	24.5	1/11/2015	23.9	24.5
	Δεκέβριος	16.8	17.2		Δεκέβριος	16.8	17.2	1/12/2015	16.8	17.2
2016	Γενάρης	12.6	13.7	2016	Γενάρης	12.6	13.7	1/1/2016	12.6	13.7
	Φεβρουάριος	14	14.6		Φεβρουάριος	14	14.6	1/2/2016	14	14.6
	Μάρτιος	18.7	19.7		Μάρτιος	18.7	19.7	1/3/2016	18.7	19.7
	ΑΠΡ	23.6	23.7		ΑΠΡ	23.6	23.7	1/4/2016	23.6	23.7
	Μάιος	26.7	27		Μάιος	26.7	27	1/5/2016	26.7	27
	ΙΟΥΝ	30.6	30.1		ΙΟΥΝ	30.6	30.1	1/6/2016	30.6	30.1
	ΙΟΥΛ	35.6	33		ΙΟΥΛ	35.6	33	1/7/2016	35.6	33
	Αυγουστος	35.2	32.9		Αυγουστος	35.2	32.9	1/8/2016	35.2	32.9
	ΣΕΠ	31	30.4		ΣΕΠ	31	30.4	1/9/2016	31	30.4

Αφαίρεση κενών γραμμών και στηλών

Νέα στήλη για ημερομηνία σε μηχαναγνώσιμο μορφότυπο

Παράδειγμα 2



Σύγκριση τρόπων οργάνωσης πίνακα δεδομένων (κατά ύψος Vs κατά μήκος).

V Οργάνωση κατά Ύψος															
Έτος	Λευκωσία	Λεμεσός													
2004	38%	23%													
2005	37%	24%													
2006	39%	26%													
2007	41%	26%													
2008	39%	28%													
2009	40%	28%													
2010	40%	27%													
2011	39%	26%													
2012	40%	27%													
2013	43%	29%													
2014	39%	27%													
2015	38%	25%													
2016	41%	27%													
2017	41%	27%													

X Οργάνωση κατά Μήκος														
Έτος	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Λευκωσία	38%	37%	39%	41%	39%	40%	40%	39%	40%	43%	39%	38%	41%	41%
Λεμεσός	23%	24%	26%	26%	28%	28%	27%	26%	27%	29%	27%	25%	27%	27%

Παράδειγμα 3

Παρουσίαση τρόπου σύμπτυξης πολλαπλών πινάκων δεδομένων σε ένα πίνακα, με ταυτόχρονη δημιουργία στήλης για το έτος, μεταφορά των υποσημειώσεων στα μεταδεδομένα, αφαίρεση ενοποιημένων κελιών (merged cells) και προσθήκη ονομάτων των στηλών με λατινικούς χαρακτήρες.

Αριθμός ενεργών εισφορών (μισθωτών) κατά κοινότητα και χρόνο στον τομέα των κατασκευών για τα χρόνια 2013 - 2016

2013			2015		
Κοινότητα	Αρ. εργοδοτούμενων	Ασφαλιστές αποδοχές €	Κοινότητα	Αρ. εργοδοτούμενων	Ασφαλιστές αποδοχές €
Ελληνοκύπριοι	16,860	246,707,000	Ελληνοκύπριοι	14,520	211,634,243
Τουρκοκύπριοι	327	4,301,587	Τουρκοκύπριοι	177	3,219,321
Αλλοδαποί	1,422	10,179,737	Αλλοδαποί	1,418	9,730,780
Κοινοτικοί	7,910	61,275,015	Κοινοτικοί	6,059	49,376,167
Σύνολο	26,519	322,463,339	Σύνολο	22,174	273,960,512

2014			2016		
Κοινότητα	Αρ. εργοδοτούμενων	Ασφαλιστές αποδοχές €	Κοινότητα	Αρ. εργοδοτούμενων	Ασφαλιστές αποδοχές €
Ελληνοκύπριοι	14,938	221,456,701	Ελληνοκύπριοι	15,279	230,046,545
Τουρκοκύπριοι	204	3,567,261	Τουρκοκύπριοι	185	3,107,241
Αλλοδαποί	1,228	9,087,587	Αλλοδαποί	1,606	11,828,481
Κοινοτικοί	6,215	49,748,862	Κοινοτικοί	6,988	59,163,985
Σύνολο	22,585	283,860,410	Σύνολο	24,058	304,146,251

Σημείωση: Στους αλλοδαπούς ή κοινοτικούς πιθανώς να περιλαμβάνονται και Εισφορές, αποδοχές του κατασκευαστικού τμήμα, κατά κοινότητα για τα χρόνια 2003-2016

Μεταφέρετε στα μεταδεδομένα

Ενοποιείστε τους 4 πίνακες σε ένα ενιαίο πίνακα

Year	Description	Greek Cypriots	Turkish Cypriots	Alliens	EU
2013	Αριθμός εργοδοτούμενων	16,860	327	1,422	7,910
2013	Ασφαλιστές αποδοχές €	246,707,000	4,301,587	10,179,737	61,275,015
2014	Αριθμός εργοδοτούμενων	14,938	204	1,228	6,215
2014	Ασφαλιστές αποδοχές €	221,456,701	3,567,261	9,087,587	49,748,862
2015	Αριθμός εργοδοτούμενων	14,520	177	1,418	6,059
2015	Ασφαλιστές αποδοχές €	211,634,243	3,219,321	9,730,780	49,376,167
2016	Αριθμός εργοδοτούμενων	15,279	185	1,606	6,988
2016	Ασφαλιστές αποδοχές €	230,046,545	3,107,241	11,828,481	59,163,985

Ονόματα στηλών (λατινικοί χαρακτήρες)

Παράδειγμα 4



Παρουσίαση τρόπου οργάνωσης πίνακα δεδομένων – Χρήση γραμμών (records) για περίληψη όλων των δεδομένων μίας εγγραφής (πχ ένα πρόσωπο/ αντικείμενο/ στατιστική μονάδα/ έτος, κλπ) και χρήση στηλών σαν αριθμητικά πεδία.

Όνομα	Ύψος	Βάρος
Γιώργος	165	47
Μιχάλης	162	50
Μαρία	157	41
Δέσποινα	162	42
Παναγιώτης	170	65
Αλέξης	168	62
Μάνος	175	78
Αχιλλέας	167	72
Ευρυδίκη	155	38
Παναγιώτα	156	46
Νίκος	172	68
Άντρη	168	52

Όνομα	Γιώργος	Μιχάλης	Μαρία	Δέσποινα	Παναγιώτης	Αλέξης	Μάνος	Αχιλλέας	Ευρυδίκη	Παναγιώτα	Νίκος	Άντρη
Ύψος	165	162	157	162	170	168	175	167	155	156	172	168
Βάρος	47	50	41	42	65	62	78	72	38	46	68	52

Χ Λανθασμένο

✓ Σωστό

Μετατροπή Αρχείου Excel σε CSV

Η μετατροπή ενός αρχείου Excel, στο οποίο περιέχονται τα δεδομένα, σε CSV είναι μια απλή διαδικασία 2 φάσεων. Βασική προϋπόθεση για την εκτέλεση της διαδικασίας αυτής είναι τα δεδομένα μας να έχουν καθαριστεί και οργανωθεί σε πίνακα ακολουθώντας τις οδηγίες που περιεγράφηκαν πιο πάνω.

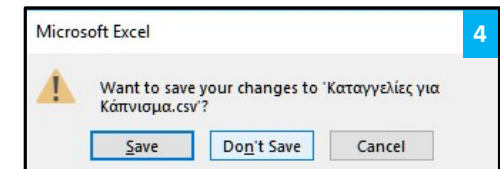
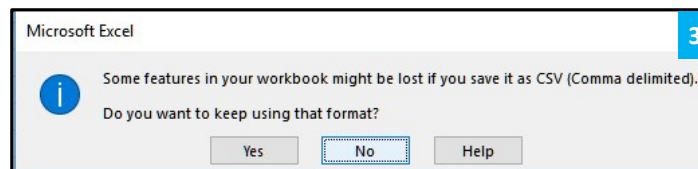
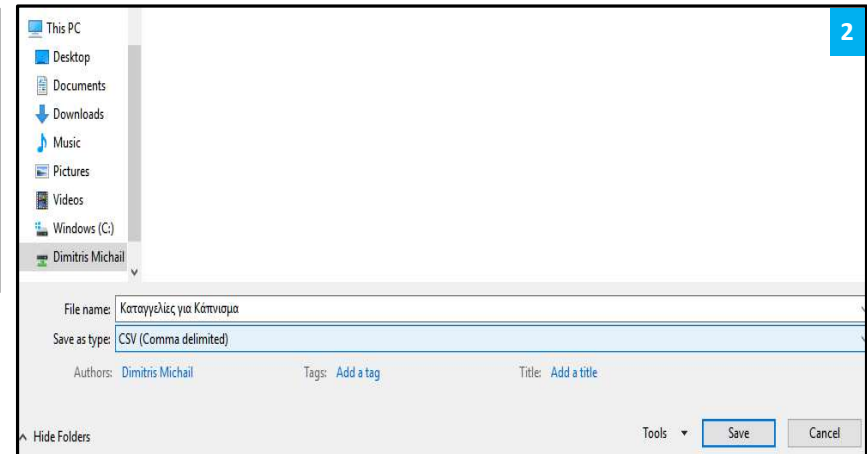
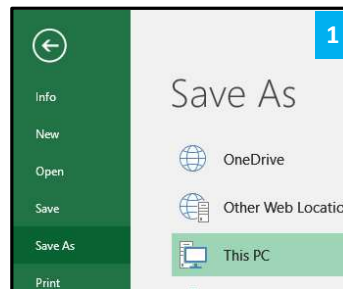
A' Φάση – Αποθήκευση Αρχείου ως CSV

Βήμα 1: Από το κυρίως μενού του Excel επιλέξτε **File** και μετά **Save As**

Βήμα 2: Στο παράθυρο που εμφανίζεται από τον κατάλογο **Save as type** επιλέξτε **CSV (Comma delimited)** και κάντε κλικ στο κουμπί **Save**.

Βήμα 3: Στο παράθυρο που εμφανίζεται επιλέξτε **Yes**.

Βήμα 4: Κλείστε το αρχείο. Στο παράθυρο που εμφανίζεται επιλέξτε **Don't Save**.



Μετατροπή Αρχείου Excel σε CSV



Βήμα 1: Ανοίξτε το αρχείο CSV που αποθηκεύσατε κάνοντας δεξί κλικ πάνω του και επιλέγοντας, κατά σειρά, Open With - Notepad

Βήμα 2: Ελέγξτε τα δεδομένα στο αρχείο και βεβαιωθείτε ότι δεν υπάρχουν περιττά κόμματα δεξιά των δεδομένων ή μετά την τελευταία γραμμή δεδομένων.

Βήμα 3: Από το μενού που εμφανίζεται, επιλέξτε **Save As**.

Βήμα 4: Στο παράθυρο που εμφανίζεται, προσθέστε στο τέλος του ονόματος του αρχείου **.csv**. Ακολούθως, καθορίστε το Encoding σε **UTF-8**, κάντε κλικ στο κουμπί **Save**.

1 File Explorer showing the file 'Καταγγελίες για Κάπνισμα' selected. The context menu is open, and 'Open with' is chosen, leading to a list of applications including Notepad.

2 Notepad window showing the file content. The text is as follows:

```
PREMICES CATEGORY,2014,2015,2016
Δημόσιο Κτίριο,3,3,6
Διασκήκη - Club,292,766,488
Εστιατόριο - Ταβέρνα,201,745,300
Γιατρείο - Κλινική - Νοσοκομείο,5,61,31
Καμπαρέ,40,121,18
Καφενείο - Λέσχη - Σύλλογος,406,1447,431
Καφετέρια,244,681,374
Μουσικοχορευτικό Κέντρο,168,242,65
Μπαρ,994,2510,977
Ξενοδοχείο,0,1,0
Περίπτερο,1,27,4
Προκτορείο Στοιχημάτων,440,1093,257
Σχολείο - Κολέγιο - Πανεπιστήμιο,0,0,0
Άλλο,63,155,72
ΣΥΝΟΛΟ,2857,7852,3023
```

3 Notepad menu with 'Save As...' highlighted.

4 Save dialog box showing 'File name: Καταγγελίες για Κάπνισμα.csv', 'Save as type: Text Documents (*.txt)', and 'Encoding: UTF-8' selected.

Τέλος Παρουσίασης

Ευχαριστούμε για την προσοχή σας



Ομάδα Ανοικτών Δεδομένων

Τμήμα Δημόσιας Διοίκησης και Προσωπικού
Υπουργείο Οικονομικών

Tel: +357 22 601551 | Fax: +357 22 602763

Web: www.data.gov.cy | Twitter: @OpenDataCY