

Big Data και Analytics

3η εβδομάδα: Διακυβέρνηση δεδομένων, ποιότητα δεδομένων και κανονιστική συμμόρφωση

Τμήμα Λογιστικής και Χρηματοοικονομικής

Πλαίσιο της σημερινής συνάντησης

- Διάρκεια μαθήματος: 3 ώρες
- Πρώτο μέρος: τι σημαίνει διακυβέρνηση δεδομένων σε επιχειρησιακό και χρηματοοικονομικό περιβάλλον
- Δεύτερο μέρος: ποιότητα δεδομένων, συνήθη σφάλματα και επιπτώσεις στην ανάλυση
- Τρίτο μέρος: κανονιστική συμμόρφωση, τεκμηρίωση και βασική πρακτική σε R και Excel

Στόχος

Να γίνει σαφές ότι χωρίς αξιόπιστα, ελεγχόμενα και τεκμηριωμένα δεδομένα, ακόμη και το καλύτερο αναλυτικό μοντέλο οδηγεί σε εσφαλμένα συμπεράσματα.

Μαθησιακά αποτελέσματα της 3ης εβδομάδας

Με το τέλος της σημερινής διάλεξης οι φοιτητές αναμένεται να μπορούν:

- να εξηγούν τι είναι η διακυβέρνηση δεδομένων και γιατί αφορά άμεσα τη λογιστική και τη χρηματοοικονομική
- να αναγνωρίζουν βασικές διαστάσεις ποιότητας δεδομένων
- να εντοπίζουν συνήθη προβλήματα σε λογιστικά και χρηματοοικονομικά σύνολα δεδομένων
- να συνδέουν την ποιότητα δεδομένων με ελεγκτικό κίνδυνο και επιχειρησιακό κίνδυνο
- να κατανοούν τον ρόλο της συμμόρφωσης, της πρόσβασης και της τεκμηρίωσης στις αναλυτικές διαδικασίες

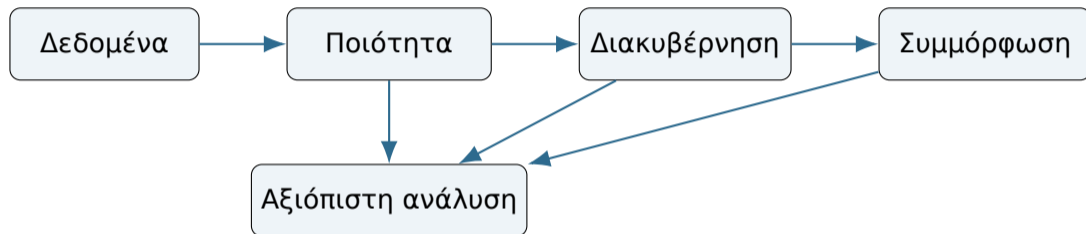
Γιατί η 3η εβδομάδα είναι κρίσιμη;

- Στις επιχειρήσεις τα δεδομένα παράγονται από πολλά συστήματα και χρήστες.
- Η ανάλυση μπορεί να φαίνεται τεχνικά σωστή αλλά να στηρίζεται σε λανθασμένες εγγραφές.
- Στη λογιστική και στον έλεγχο απαιτείται ιχνηλασιμότητα, συνέπεια και τεκμηρίωση.
- Η κανονιστική συμμόρφωση δεν είναι παράρτημα. Είναι μέρος της ίδιας της αρχιτεκτονικής των δεδομένων.

Κεντρική ιδέα

Η αξία των analytics εξαρτάται από τη θεσμική και ποιοτική ωριμότητα των δεδομένων.

Η λογική ακολουθία: από τα δεδομένα στην απόφαση



Η ανάλυση δεν είναι μόνο μαθηματικό πρόβλημα. Είναι πρόβλημα ελέγχου, οργάνωσης και θεσμικής πειθαρχίας.

Τι είναι η διακυβέρνηση δεδομένων;

Ορισμός

Η διακυβέρνηση δεδομένων είναι το σύνολο ρόλων, κανόνων, διαδικασιών και ελέγχων που καθορίζουν ποιος δημιουργεί, ποιος τροποποιεί, ποιος εγκρίνει, ποιος χρησιμοποιεί και με ποιον τρόπο τεκμηριώνονται τα δεδομένα.

- ορίζει ευθύνη και λογοδοσία
- μειώνει ασυνέπειες και αυθαίρετες παρεμβάσεις
- επιτρέπει επαναληψιμότητα της ανάλυσης
- ενισχύει έλεγχο, διαφάνεια και αξιοπιστία αναφορών

Βασικά στοιχεία της διακυβέρνησης δεδομένων

Στοιχείο	Τι αφορά	Γιατί είναι κρίσιμο
Ιδιοκτησία δεδομένων	Ποιος είναι υπεύθυνος για κάθε σύνολο δεδομένων	Αποσαφηνίζει ευθύνη και λογοδοσία
Πολιτικές πρόσβασης	Ποιος βλέπει ή τροποποιεί τι	Μειώνει κίνδυνο διαρροής και μη εξουσιοδοτημένης αλλαγής
Πρότυπα ορισμών	Κοινές έννοιες για δείκτες και πεδία	Αποτρέπει αντιφατικές αναφορές
Ιχνηλασιμότητα	Καταγραφή αλλαγών και πηγών	Επιτρέπει έλεγχο και αναπαραγωγή
Κανόνες ποιότητας	Έλεγχοι εγκυρότητας και συνέπειας	Προλαμβάνει σφάλματα πριν την ανάλυση

Ποιοι ρόλοι εμφανίζονται συνήθως;

Επιχειρησιακοί ρόλοι

- ιδιοκτήτης δεδομένων
- υπεύθυνος διαδικασίας
- λογιστήριο ή οικονομική διεύθυνση
- εσωτερικός έλεγχος

Τεχνικοί και ελεγκτικοί ρόλοι

- data steward
- διαχειριστής συστήματος
- αναλυτής δεδομένων
- υπεύθυνος συμμόρφωσης ή DPO

Σε ώριμο οργανισμό οι ρόλοι αυτοί δεν συγχέονται. Η σαφήνεια ρόλων είναι βασικός έλεγχος από μόνη της.

Οι βασικές διαστάσεις ποιότητας δεδομένων

- ακρίβεια
- πληρότητα
- συνέπεια
- επικαιρότητα
- μοναδικότητα
- εγκυρότητα
- συγκρισιμότητα
- καταλληλότητα για χρήση

Ουσία

Ένα σύνολο δεδομένων μπορεί να είναι τεχνικά μεγάλο αλλά αναλυτικά άχρηστο αν αποτυγχάνει σε μία ή περισσότερες από αυτές τις διαστάσεις.

Παραδείγματα προβλημάτων ποιότητας στη λογιστική

Πρόβλημα	Παράδειγμα	Επίπτωση
Διπλοεγγραφές	Το ίδιο τιμολόγιο περνά δύο φορές	Υπερεκτίμηση εξόδων ή υποχρεώσεων
Ελλιπή πεδία	Λείπει κωδικός πελάτη ή ημερομηνία	Αδύνατη συμφωνία και ομαδοποίηση
Ασυνέπεια ορισμών	Διαφορετική έννοια για το net sales	Μη συγκρίσιμα reports
Λάθος ταξινόμηση	Κεφαλαιουχική δαπάνη ως λειτουργικό έξοδο	Παραμόρφωση δεικτών και αποτελέσματος
Μη έγκαιρη ενημέρωση	Καθυστέρηση καταχώρισης εισπραξιών	Λανθασμένη εικόνα ρευστότητας

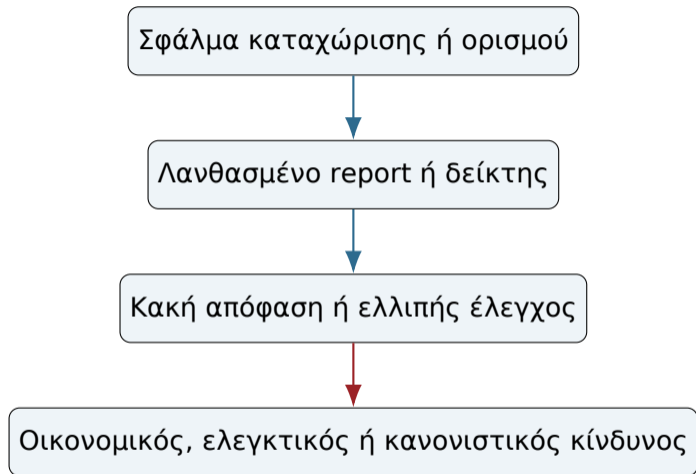
Παραδείγματα προβλημάτων ποιότητας στη χρηματοοικονομική

- εσφαλμένη αντιστοίχιση τίτλων ή tickers
- ασύμβατες χρονικές σφραγίδες μεταξύ αγοραίων και εσωτερικών δεδομένων
- κενά σε ιστορικές σειρές τιμών ή αποδόσεων
- μη ομοιόμορφη μεταχείριση εταιρικών πράξεων
- λανθασμένη συναλλαγματική μετατροπή
- μη επικαιροποιημένα στοιχεία για πιστωτικό κίνδυνο ή όρια έκθεσης

Συμπέρασμα

Στη χρηματοοικονομική, μικρά σφάλματα στην ποιότητα δεδομένων μπορούν να οδηγήσουν σε σημαντική στρέβλωση στην αποτίμηση ή στη μέτρηση κινδύνου.

Από το σφάλμα δεδομένων στον επιχειρησιακό κίνδυνο



Ο κίνδυνος δεν γεννιέται μόνο από την αγορά ή την απάτη. Συχνά γεννιέται από κακή πληροφορία.

Εσωτερικοί έλεγχοι για την ποιότητα δεδομένων

- έλεγχοι εγκυρότητας κατά την εισαγωγή δεδομένων
- υποχρεωτικά πεδία και κανόνες μορφοποίησης
- συμφωνίες μεταξύ υποσυστημάτων και γενικής λογιστικής
- αναφορές εξαιρέσεων και εντοπισμός ασυνήθιστων τιμών
- περιοδικοί έλεγχοι δειγματοληψίας και επανυπολογισμοί
- σαφείς διαδικασίες διορθώσεων και εγκρίσεων

Πρακτική διάσταση

Στόχος δεν είναι απλώς να εντοπίσουμε το σφάλμα, αλλά να σχεδιάσουμε μηχανισμό που μειώνει την πιθανότητα επανάληψής του.

Τι σημαίνει κανονιστική συμμόρφωση σε αυτό το πλαίσιο;

- νόμιμη και θεμιτή επεξεργασία δεδομένων
- περιορισμός πρόσβασης σε ευαίσθητες πληροφορίες
- σαφής τεκμηρίωση σκοπού χρήσης των δεδομένων
- διατήρηση δεδομένων για όσο απαιτείται και όχι περισσότερο
- δυνατότητα ελέγχου και τεκμηρίωσης των αναλυτικών διαδικασιών
- προστασία προσωπικών και εμπιστευτικών χρηματοοικονομικών πληροφοριών

Παραδείγματα ευαίσθητων δεδομένων

- μισθοδοσία
- στοιχεία πελατών
- τραπεζικοί λογαριασμοί
- φορολογικά δεδομένα

Απαιτήσεις ελέγχου

- αρχή ελάχιστης πρόσβασης
- καταγραφή ενεργειών χρηστών
- ψευδωνυμοποίηση όπου απαιτείται
- τεκμηρίωση εξαγωγών και μεταφορών

Η λογική είναι σαφής: δεν αρκεί να έχουμε τα δεδομένα. Πρέπει να μπορούμε να αποδείξουμε ότι τα χρησιμοποιούμε σωστά.

Τεκμηρίωση και αναπαραγωγιμότητα στην αναλυτική

- ποια ήταν η αρχική πηγή;
- πότε έγινε εξαγωγή του αρχείου;
- ποιοι μετασχηματισμοί εφαρμόστηκαν;
- ποια φίλτρα και παραδοχές χρησιμοποιήθηκαν;
- ποια έκδοση αρχείου ή κώδικα παρήγαγε το αποτέλεσμα;

Γιατί έχει σημασία;

Σε ακαδημαϊκό, ελεγκτικό και επαγγελματικό περιβάλλον, ένα αποτέλεσμα χωρίς τεκμηριωμένη διαδρομή παραγωγής έχει περιορισμένη αξιοπιστία.

Παράδειγμα μικρού πλαισίου ελέγχου δεδομένων

Στάδιο	Ερώτημα ελέγχου	Ενέργεια
Συλλογή	Είναι έγκυρη η πηγή;	Επιβεβαίωση προέλευσης και εξουσιοδότησης
Εισαγωγή	Περνούν βασικοί κανόνες ποιότητας;	validation rules και έλεγχοι πεδίων
Αποθήκευση Ανάλυση	Υπάρχει έλεγχος πρόσβασης; Μπορεί να αναπαραχθεί;	Δικαιώματα, logs και backup Script, versioning και τεκμηρίωση βημάτων
Αναφορά	Είναι κατανοητό το αποτέλεσμα;	Σαφείς ορισμοί και παραδοχές

Excel

- γρήγορος οπτικός έλεγχος
- ευκολία για φίλτρα και συμφωνίες
- υψηλός κίνδυνος χειροκίνητων παρεμβάσεων
- περιορισμένη αναπαραγωγικότητα

R

- τυποποιημένοι έλεγχοι ποιότητας
- αναπαραγωγίμη επεξεργασία
- ισχυρό περιβάλλον για validation και reporting
- καλύτερη τεκμηρίωση βημάτων

Το Excel είναι χρήσιμο ως υποστηρικτικό εργαλείο. Η R είναι καταλληλότερη για συστηματική και επαναλήψιμη αναλυτική εργασία.

Μικρό παράδειγμα ελέγχου ποιότητας σε R

```
ledger <- read.csv("sales_ledger.csv")  
  
sum(is.na(ledger$customer_id))  
sum(duplicated(ledger$invoice_no))  
range(ledger$invoice_date)  
subset(ledger, amount < 0)
```

Το βασικό μήνυμα δεν είναι ο κώδικας καθαυτός. Είναι ότι οι έλεγχοι πρέπει να είναι τυποποιημένοι, σαφείς και επαναλαμβανόμενοι.

Σενάριο

Δίνεται αρχείο πωλήσεων ή λογιστικών εγγραφών με ελλιπείς τιμές, διπλοεγγραφές, διαφορετικές μορφές ημερομηνίας και ασυνέπεια σε κατηγορίες.

Οι φοιτητές καλούνται:

- να εντοπίσουν τα βασικά προβλήματα ποιότητας
- να προτείνουν ελέγχους που θα έπρεπε να υπάρχουν στη διαδικασία
- να διακρίνουν τι μπορεί να ελεγχθεί σε Excel και τι αποδοτικότερα σε R
- να περιγράψουν τις κανονιστικές ή ελεγκτικές επιπτώσεις των σφαλμάτων

Ενδεικτική κατανομή των 3 ωρών

Μέρος	Περιεχόμενο	Χρόνος
1ο μέρος	Διακυβέρνηση δεδομένων, ρόλοι, πολιτικές και εσωτερικοί έλεγχοι	60 λεπτά
2ο μέρος	Ποιότητα δεδομένων, παραδείγματα σφαλμάτων και επιπτώσεις σε αναφορές και analytics	60 λεπτά
3ο μέρος	Συμμόρφωση, τεκμηρίωση και πρακτική άσκηση σε R/Excel	60 λεπτά

Τρία βασικά σημεία που πρέπει να μείνουν

- 1 Η ποιότητα των αποτελεσμάτων δεν μπορεί να υπερβαίνει την ποιότητα των δεδομένων.
- 2 Η διακυβέρνηση δεδομένων είναι θέμα λογοδοσίας, όχι μόνο τεχνολογίας.
- 3 Η συμμόρφωση και η τεκμηρίωση είναι προϋποθέσεις σοβαρής αναλυτικής εργασίας.

Προετοιμασία για την επόμενη εβδομάδα

Για την 4η εβδομάδα οι φοιτητές μπορούν να προετοιμαστούν ως εξής:

- να επανεξετάσουν τη διάκριση μεταξύ πηγής, ποιότητας και ελέγχου δεδομένων
- να φέρουν ένα παράδειγμα επιχειρησιακού συστήματος που παράγει δεδομένα
- να σκεφτούν πώς διαφορετικά συστήματα συνδέονται μεταξύ τους μέσα σε μια επιχείρηση

Η επόμενη διάλεξη θα περάσει από τους κανόνες και την ποιότητα στις υποδομές δεδομένων και στα τεχνολογικά περιβάλλοντα.

Ενδεικτική βιβλιογραφία για αφετηρία

- Provost, F. and Fawcett, T., Data Science for Business.
- Shmueli, G. et al., Data Mining for Business Analytics.
- Appelbaum, D., Kogan, A. and Vasarhelyi, M., άρθρα για accounting data analytics και continuous auditing.
- Συμπληρωματική αρθρογραφία για data governance, ποιότητα δεδομένων και κανονιστική συμμόρφωση σε λογιστικά πληροφοριακά συστήματα.

Ευχαριστώ