Efstratios Karkanis

[Διεύθυνση email]

Απόσπασμα

[Τραβήξτε την προσοχή του αναγνώστη σας με ένα ελκυστικό απόσπασμα. Συνήθως πρόκειται για μια σύντομη περίληψη του εγγράφου.   
Όταν είστε έτοιμοι να προσθέσετε το περιεχόμενό σας, απλώς κάντε κλικ εδώ και αρχίστε να πληκτρολογείτε.]

ΑΝΑΛΥΤΙΚΗ ΔΕΔΟΜΕΝΩΝ

Ομαδική Απαλλακτική Εργασία

**Ερώτημα 1: εξοικείωση με τα δεδομένα**

α) καθαρισμός των δεδομένων

* στο σύνολο των δεδομένων που μας δόθηκε παρατηρήσαμε ότι κάποιες από τις στήλες του περιείχαν περιττές πληροφορίες που δεν μας αφορούν στην μελέτη μας. Για παράδειγμα, στήλες όπως η χώρα (country), ο ταχυδρομικός κώδικας (zipcode), o Airport\_Code κ.α. δεν μας βοηθούν στην λήψη αποφάσεων. Επομένως, οι στήλες αυτές αφαιρέθηκαν από το σύνολο των δεδομένων. Η διαδικασία αυτή έχει περιγραφεί σε σχετικό κομμάτι του κώδικα.
* το dataset έχει καθαριστεί από null τιμές. Λόγω του γεγονότος ότι οι εγγραφές του dataset είναι πάρα πολλές, αυτό που κάναμε είναι να διαγράψουμε όσες γραμμές έχουν έστω μία null τιμή. Αυτή μας η επιλογή δεν επηρεάζει την αξιοπιστία των δεδομένων, καθώς στις 2.8 εκατομμύρια εγγραφές του dataset (περίπου), οι 500.000 μόνο εγγραφές διαγράφηκαν. Επομένως, δεν θεωρούμε ότι αυτή η απώλεια είναι σημαντική για την μελέτη μας.