SAMSUNG INNOVATION CAMPUS – UPSKILLING DIGITALLY

Εργασία στην ανάλυση δεδομένων με τη χρήση της python

Ερωτήματα που αφορούν την περιγραφή του dataset

1. Διαβάστε το αρχείο data.csv *
2. Εκτυπώστε τις τελευταίες 5 εγγραφές του dataframe *
3. Πόσες (αριθμό) και ποιες (ονόματα) στήλες (columns) έχει το dataset που φορτώσατε; *
4. Ως τι τύπο δεδομένων αναγνωρίζει η βιβλιοθήκη pandas τις στήλες του dataset; *
5. Υπάρχουν στήλες με τιμές που λείπουν; Κι αν ναι ποιες; (Ποιες στήλες έχουν missing values /NaN) *
6. Ποιος ο συνολικός αριθμών των εγγραφών; (χωρίς τα headers) *

Οδηγίες για τον καθαρισμό του dataset

1. Αφαιρέστε από το dataset όλες τις γραμμές (rows) που έχουν NaN (missing values) στις στήλες 'Description' ή/και 'CustomerID'. Tip δείτε τη μέθοδο [dropna](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.dropna.html) **
2. Διαγράψτε όλες τις γραμμές που η περιγραφή της στήλης 'Description' είναι : "AMAZON FEE","Manual", "SAMPLES", "POSTAGE" ή "PACKING CHARGE". (Μπορείτε να κάνετε τη διαγραφή με όποιον τρόπο επιθυμείτε, με μία ή περισσότερες γραμμές κώδικα, στόχος είναι να μείνουν οι σωστές εγγραφές) ***
3. Αφαιρέστε απο το dataset όλες εγγραφές έχουν αρνητική τιμή στη στήλη 'Quantity' **
4. Δημιουργήστε στήλη ονόματι ItemTotal που περιέχει ανά γραμμή το αποτέλεσμα της πράξης Quantity*UnitPrice για τον υπολογισμό του συνολικού κόστους ανά κατηγορία προϊόντων **

Ερωτήματα για την κατανόηση του dataset

1. Ποιος ο αριθμός των μοναδικών /διαφορετικών πελατών (μη λάβετε υπόψιν όσες εγγραφές έχουν NaN αντί για τιμή στο πεδίο του CustomerID) **
2. Με ποιες χώρες έχει μέχρι σήμερα συναλλαγές η εταιρεία; **
3. Ποιο χρονικό διάστημα αφορούν τα δεδομένα που έχουμε διαθέσιμα; (Tip εύρεση της μέγιστης και της ελάχιστης τιμής της στήλης InvoiceDate) **
4. Ποιο/ ποια προϊόν/τα μπορεί να αγοράσει ένας πελάτης που επιθυμεί να διαθέσει 100-150 ευρώ; ***
5. Αν κάνουμε αναζήτηση στα προϊόντα (στις περιγραφές) με τον όρο "HANDBAG" ποια αποτελέσματα θα λάβουμε; (Tip δείτε τη μέθόδο contains https://pandas.pydata.org/docs/reference/api/pandas.Series.str.contains.html**

Επεξήγηση συμβόλων
* Εύκολο ερώτημα

** Μέτριας δυσκολίας ερώτημα

*** Απαιτητικό ερώτημα

Για απορίες/ερωτήσεις επικοινωνήστε μαζί μου στο mskiada@aueb.gr