**Στατιστικά Πακέτα και Ανάλυση Δεδομένων**

**Feature Engineering**

**Εργασία 2η**

**Φοιτητής: Μπινιάκου Θεοφάνης,** [**icsd13126@icsd.aegean.gr**](mailto:icsd13126@icsd.aegean.gr)

**Τμήμα: Μ.Π.Ε.Σ.**

Χρησιμοποιώντας “Feature Engineering” και την γλώσσα προγραμματισμού “Python” θα λύσουμε το παρακάτω πρόβλημα:

Μια εταιρία θέλει να αυτοματοποιήσει την διαδικασία καταλληλότητας δανείου με βάση τα στοιχεία του πελάτη που παρέχονται κατά την ηλεκτρονική συμπλήρωση της αντίστοιχης αίτησης.

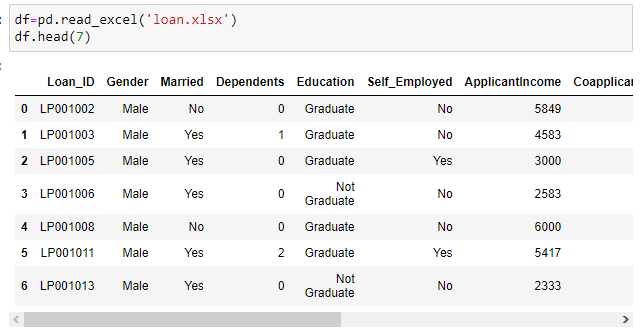
Εισαγωγή των βιβλιοθηκών που θα χρειαστούμε κατά την διάρκεια του προγράμματος

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

«Εικόνα 1»

Φόρτωση του αρχείου “loan.xlsx” με τα δάνεια



«Εικόνα 2»

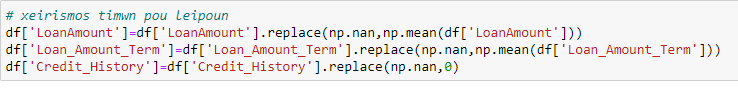
Στην συνέχεια, χρησιμοποιώντας την συνάρτηση “LabelEncoder” ποσοτικοποιούμε της κατηγορικές τιμές.

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

«Εικόνα 3»

Χειριζόμαστε τιμές που λείπουν από τα δεδομένα μας, για να μπορέσουμε να τα επεξεργαστούμε στην συνέχεια χωρίς προβλήματα



«Εικόνα 4»

Διαγράφουμε τις παρακάτω στήλες



«Εικόνα 5»

Εφαρμόζουμε την απλή λογιστική παλινδρόμησης

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

«Εικόνα 6»

Στην συνέχεια φορτώνουμε δεδομένα από τα αρχεία:

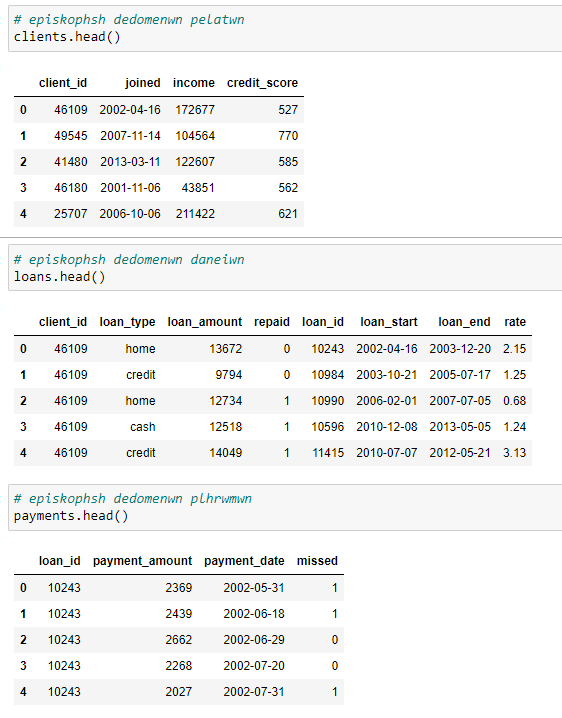
* Clients.csv
* Loans.csv
* Payments.csv

Εικόνα που περιέχει κείμενο

Περιγραφή που δημιουργήθηκε αυτόματα

«Εικόνα 7»

Εμφανίζουμε τα στοιχεία αυτών των αρχείων για να δούμε τα δεδομένα που περιέχουν τα αρχεία από την «Εικόνα 7»



«Εικόνα 8»

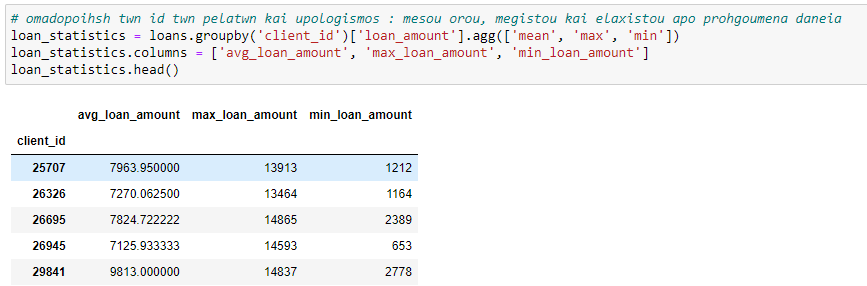
Θα προσθέσουμε στον πίνακα των πελατών, δύο καινούριες στήλες (month,log\_of\_income\_amount) όπου αντιπροσωπεύουν τον μήνα που ο πελάτης εντάχθηκε στην τράπεζα και το log του εισοδήματος του πελάτη

Εικόνα που περιέχει πίνακας

Περιγραφή που δημιουργήθηκε αυτόματα

«Εικόνα 9»

Στην συνέχεια θα συνδυάσουμε δεδομένα από τους παραπάνω πίνακες (Εικόνες 7,8), και θα δημιουργήσουμε ένα “dataframe” για την «τοποθέτηση» των δεδομένων που θα προκύψουν.



«Εικόνα 10»

Εν τέλη θα ενώσουμε τον πίνακα των πελατών με το παραπάνω dataframe και θα πάρουμε το εξής αποτέλεσμα

Εικόνα που περιέχει πίνακας

Περιγραφή που δημιουργήθηκε αυτόματα

«Εικόνα 11»