«Επιχειρηματική Αναλυτική και Τεχνολογίες Εξατομίκευσης»

Σκοπός Εργασίας

Σκοπός της εργασίας είναι η ανάλυση δεδομένων με στόχο την εξαγωγή χρήσιμων συμπερασμάτων για την υποστήριξη επιχειρηματικών αποφάσεων (επιχειρηματική αναλυτική). Τα δεδομένα αφορούν στοιχεία ζήτησης καταναλωτικών προϊόντων από κατάστημα supermarket (point-of-sales data, POS). Παρακάτω ακολουθούν η περιγραφή των δεδομένων καθώς και τα βασικά ερωτήματα που καλείστε να απαντήσετε.

Περιγραφή Δεδομένων

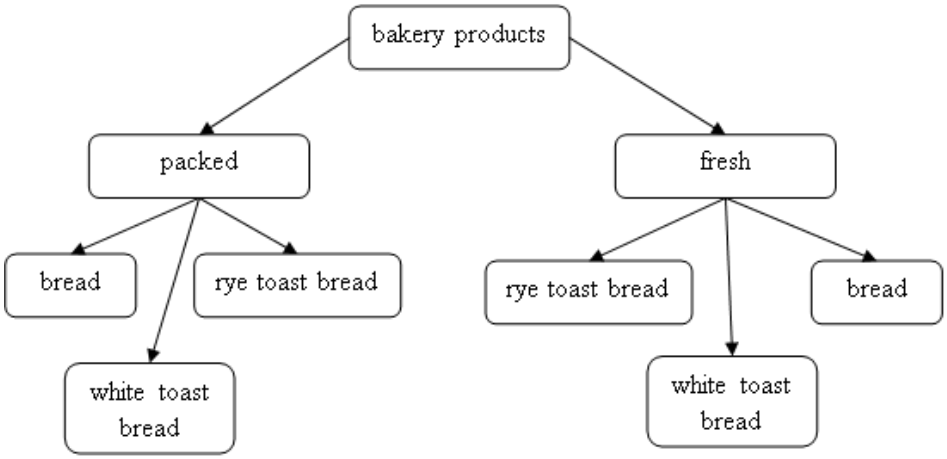
Στο πρώτο φύλλο του excel, υπάρχουν Point Of Sales (POS) data, τα οποία αναφέρονται στις συναλλαγές που πραγματοποιήθηκαν στο ταμείο ενός σουπερμάρκετ. Κάθε basket_id αναφέρεται σε μια απόδειξη. Το πεδίο date αναφέρεται στην ημερομηνία που πραγματοποιήθηκε το transaction (σε data type int). Το barcode αναφέρεται στον μοναδικό αριθμό προϊόντος, το sum_units στον αριθμό των τεμαχίων που αγοράστηκαν από το δοθέν barcode, και το sum_value, στην αξία αυτών των αγορών. Για παράδειγμα, ας πάρουμε μια περίπτωση από τα δεδομένα που δόθηκαν για να γίνει πιο κατανοητό το παραπάνω. Στον παρακάτω πίνακα το Basket_id, δηλώνει ότι έχουμε δεδομένα από μια απόδειξη η οποία περιέχει 5 κωδικούς προϊόντων, και από κάθε κωδικό έχει αγοραστεί ένα τεμάχιο, εκτός από τον πρώτο, όπου έχουν αγοραστεί 2. Το sum_value είναι η συνολική αξία όλων των τεμαχίων του συγκεκριμένου barcode. Το card_id δηλώνει τον μοναδικό αριθμό της κάρτας πιστότητας του πελάτη. Εάν λαμβάνει τιμή=NULL, τότε ο πελάτης δεν χρησιμοποίησε στη συγκεκριμένη συναλλαγή την κάρτα του.

Basket_ID	Date	Barcode	Sum_Units	Sum_Value	Card_ID
1103084867	41379	800220505783	2	1.96	9160003751260
1103853519	41381	520139501183	1	5.349993	9164001986624
1092750793	41346	520423907421	6	1.740015	9164012915385
1106160983	41388	211069400000	1	0.749817	9162005811409
1108695491	41395	520286400380	2	0.6	9161003517351

Στο δεύτερο φύλλο του excel έχουμε κάποια στοιχεία για κάθε πελάτη όπως, Age, Gender, MaritalStatus, HouseholdSize, Children

Στο τρίτο φύλλο έχουμε τους κωδικούς των προϊόντων (**barcodes**), μαζί με κάποια περιγραφή για αυτούς, καθώς και το δέντρο της ιεραρχίας των κατηγοριών στις οποίες ανήκουν (μόνο ids των κατηγοριών).

Στο τέταρτο φύλλο υπάρχει η ιεραρχία των προϊόντων **(categories hierarchy)** (ids και λεκτικό), χωρίς τα barcodes. Βλ. παρακάτω εικόνα:



Ερωτήματα

Ερωτήσεις που καλείστε να απαντήσετε ως ομάδα:

1. Basket segmentation. Ποιες ομάδες καλαθιών μπορείτε να αναγνωρίσετε στα δεδομένα οι οποίες να εξηγούν την αγοραστική συμπεριφορά, δηλαδή το είδος της επίσκεψης που έκανε ένας πελάτης;
   Εδώ καλείστε αρχικά να επιλέξετε μετρικές και διαστάσεις που θα χρησιμοποιήστε για να τμηματοποιήσετε τα καλάθια. Έπειτα καλείστε να παρουσιάσετε τα segments καλαθιών τα οποία εξαγάγατε.

2. Product recommendations: Ποια προϊόντα μπορείτε να προτείνετε σε έναν πελάτη κατά την επόμενη επίσκεψή του;
   Επιλέξτε και εφαρμόστε τις κατάλληλες τεχνικές και αλγόριθμους ώστε να προτείνετε προϊόντα σε κάθε πελάτη του καταστήματος για τις δυο παρακάτω περιπτώσεις: Στο τέλος αξιολογήστε την ακρίβεια του μοντέλου πρόβλεψης προτιμήσεων
   a. *Ο πελάτης μπορεί να αναγνωριστεί (το ID είναι γνωστό)*
   b. *Ο πελάτης δεν μπορεί να αναγνωριστεί (δεν είναι γνωστό το ID του)*

Ερώτηση που καλείστε να απαντήσετε ατομικά:

3. Customer segmentation. Ποιες ομάδες πελατών μπορείτε να αναγνωρίσετε στα δεδομένα;
   Εδώ καλείστε αρχικά να επιλέξετε μετρικές και διαστάσεις που θα χρησιμοποιήσετε για να τμηματοποιήσετε τους πελάτες. Έπειτα καλείστε να παρουσιάσετε τα segments πελατών τα οποία εξαγάγατε. Το κάθε μέλος της ομάδας θα πρέπει να προτείνει μία διαφορετική προσέγγιση τμηματοποίησης των πελατών. (Σημείωση: στοιχεία από το basket segmentation μπορούν να χρησιμοποιηθούν ως input σε αυτή την ανάλυση)
4. Ανοιχτή ερώτηση. Απαντήστε οποιαδήποτε άλλη ανοιχτή ερώτηση θέλετε με χρήση των δεδομένων.

Δώστε έμφαση στην παρουσίαση και ερμηνεία των αποτελεσμάτων.
Η εργασία θα παρουσιαστεί με τη μορφή παρουσίασης.

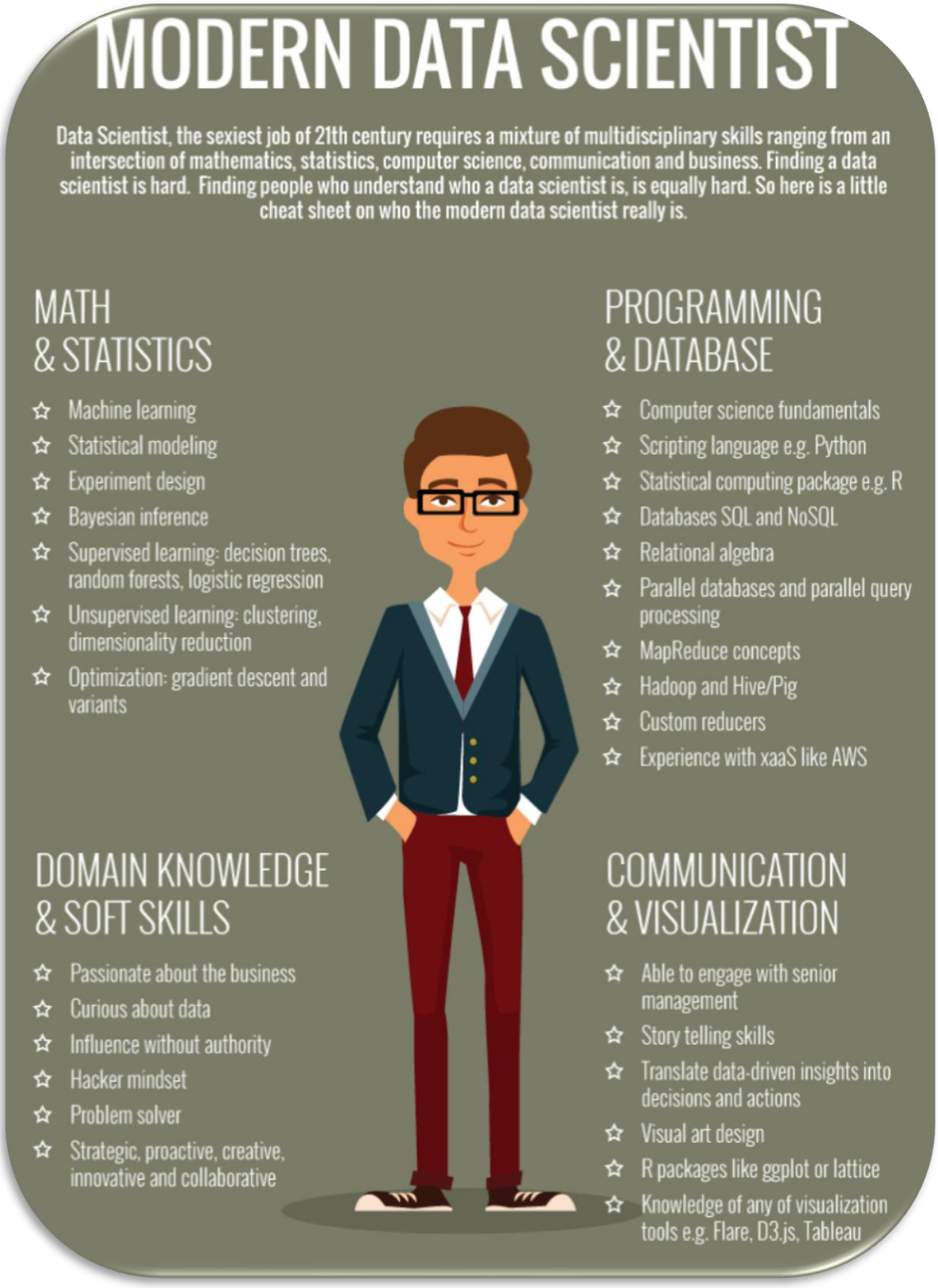
Ενδεικτικά Εργαλεία

1. **Microsoft SQL Server Management Studio (SSMS)**
   - Μπορείτε πχ να βρείτε την έκδοση του 2012 στο: https://www.dreamspark.com/Product/Product.aspx?productid=43
   - Στο παραπάνω link μπορείτε να επιλέξετε το "Product Version": Business Intelligence ή Standard.

2. **Data Mining Client for Excel (SQL Server Data Mining Add-ins)**
   - Μπορείτε να το κατεβάσετε free από το site της Microsoft. Προσοχή πρέπει να βρείτε την έκδοση που ταιριάζει στο SSMS που έχετε κατεβάσει και στο Microsoft Office που έχετε εγκατεστημένο στο pc σας. Πχ:
   - https://www.microsoft.com/en-us/download/details.aspx?id=29061

3. **RapidMiner Studio**
   - Αιτηθείτε την ακαδημαϊκή έκδοση του προγράμματος με την χρήση του aueb email σας : https://rapidminer.com/educational-program/
   - Οι αλγόριθμοι/τεχνικές που θα χρησιμοποιηθούν για την απάντηση του ερωτήματος 2 θα πρέπει να υλοποιηθούν στο RapidMiner.

In case you are interested:



Harvard Business Review (HBR)
https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/