

DATAHOLICS

SUPERMARKET DATA ANALYSIS

Κοσσυβάκης Λάμπρος 8190075

Λιάγγου Δέσποινα 8190094

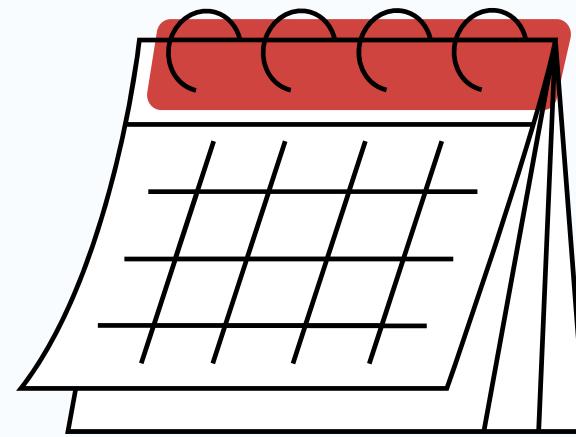
Παπαοικονόμου Γιώργος 8190139

Φραγκιαδάκης Χριστόφορος 8190168



Dataset Overview

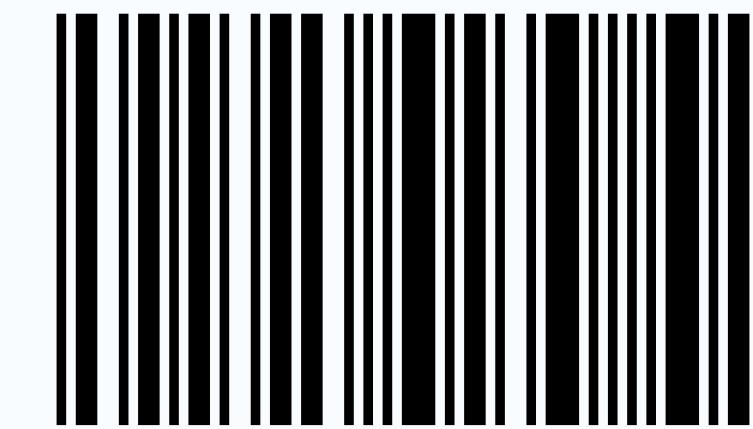
INITIAL DATASET



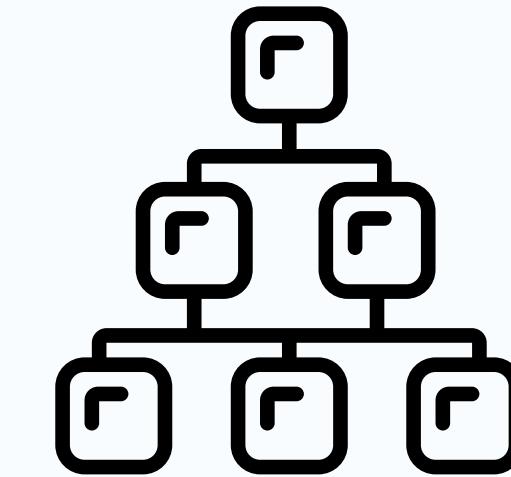
01/09/18 - 31/08/19



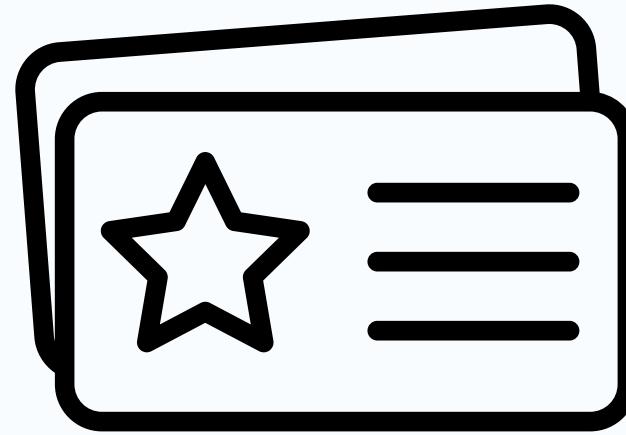
61.312



6.266



3 categories



2.240



31,1%



1,7%



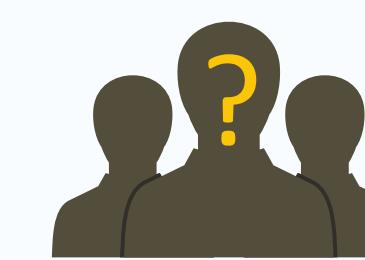
3,3%



3,5%



0,4%



60%

Data Preparation

WHAT WE FOUND



(This customer was deleted
for the sake of analysis)

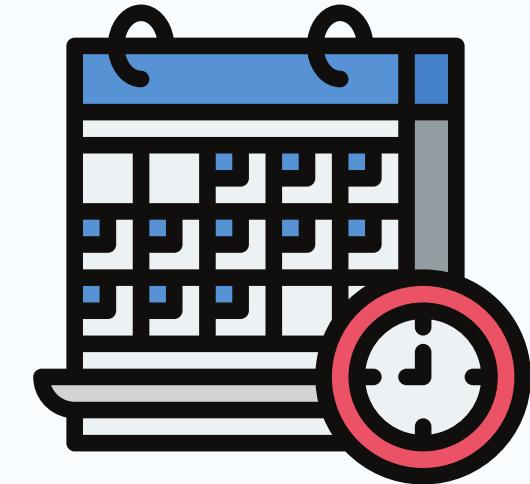


8 products per basket

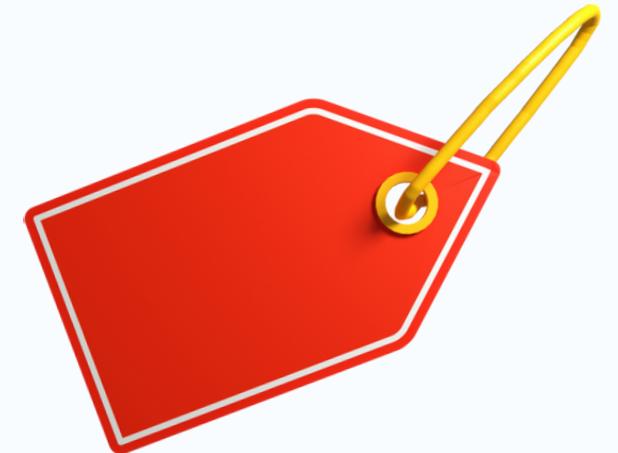


20% of all baskets

A customer with:



daily visits
(33 visits per day)



15,33€ per basket

After Data Cleaning

AVEGARES



45.212



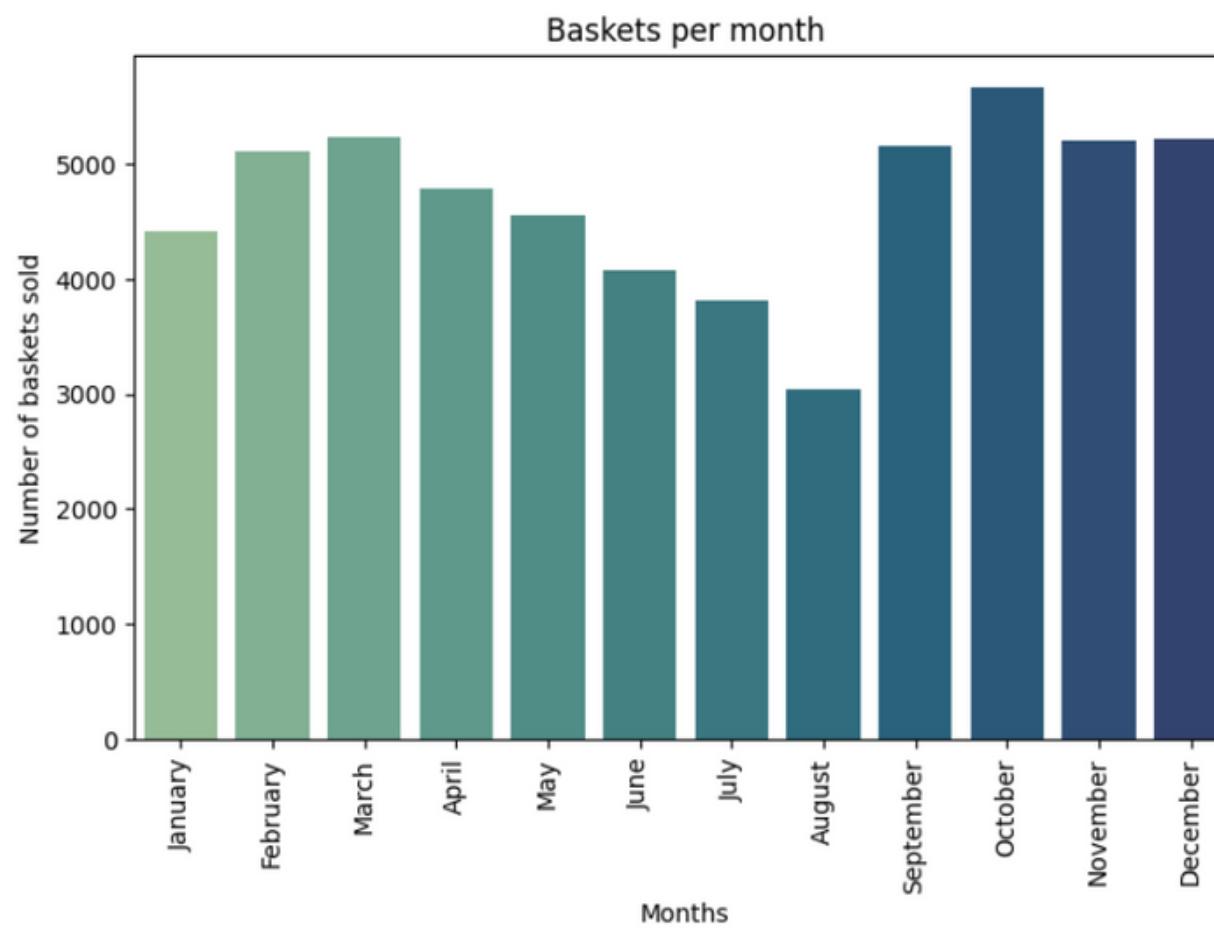
160 visits
per day



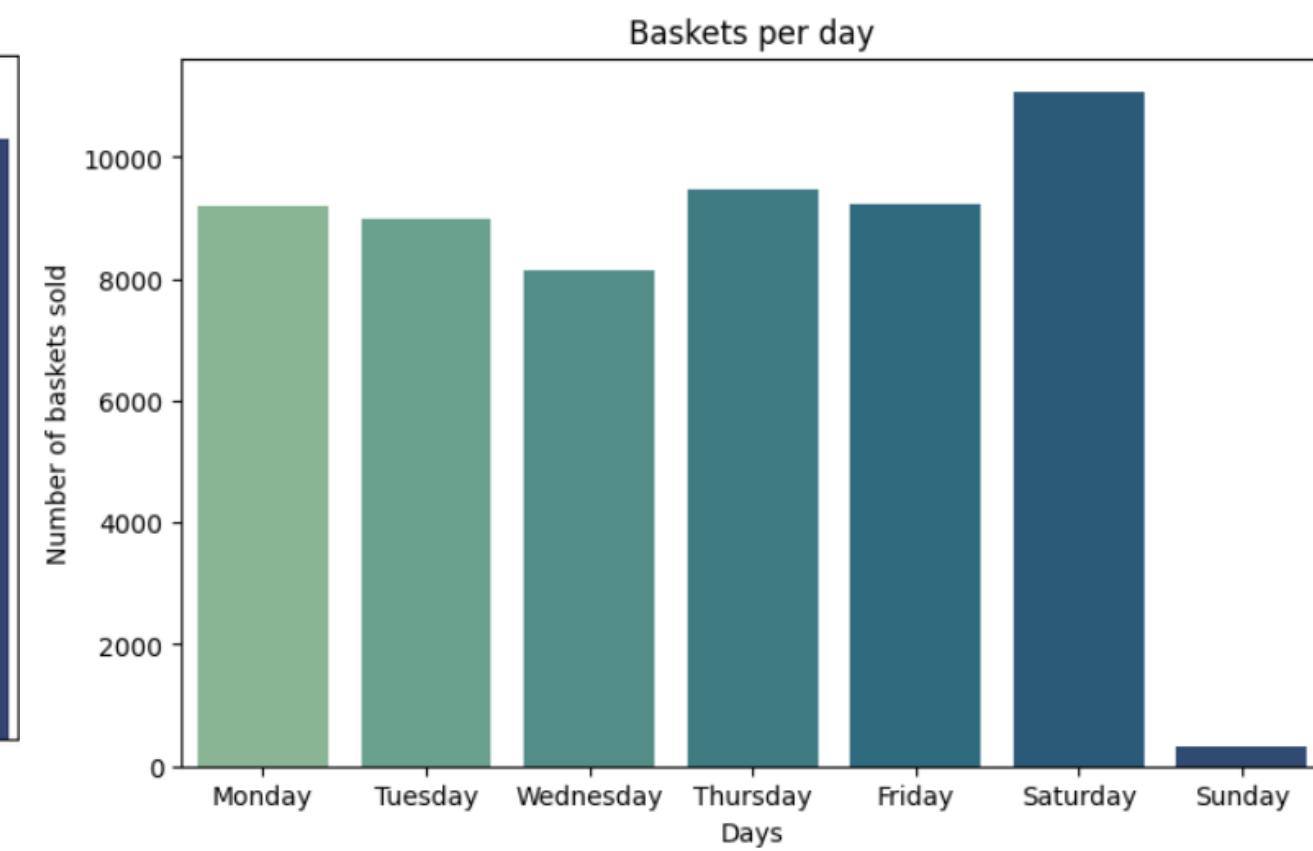
6 products
per basket



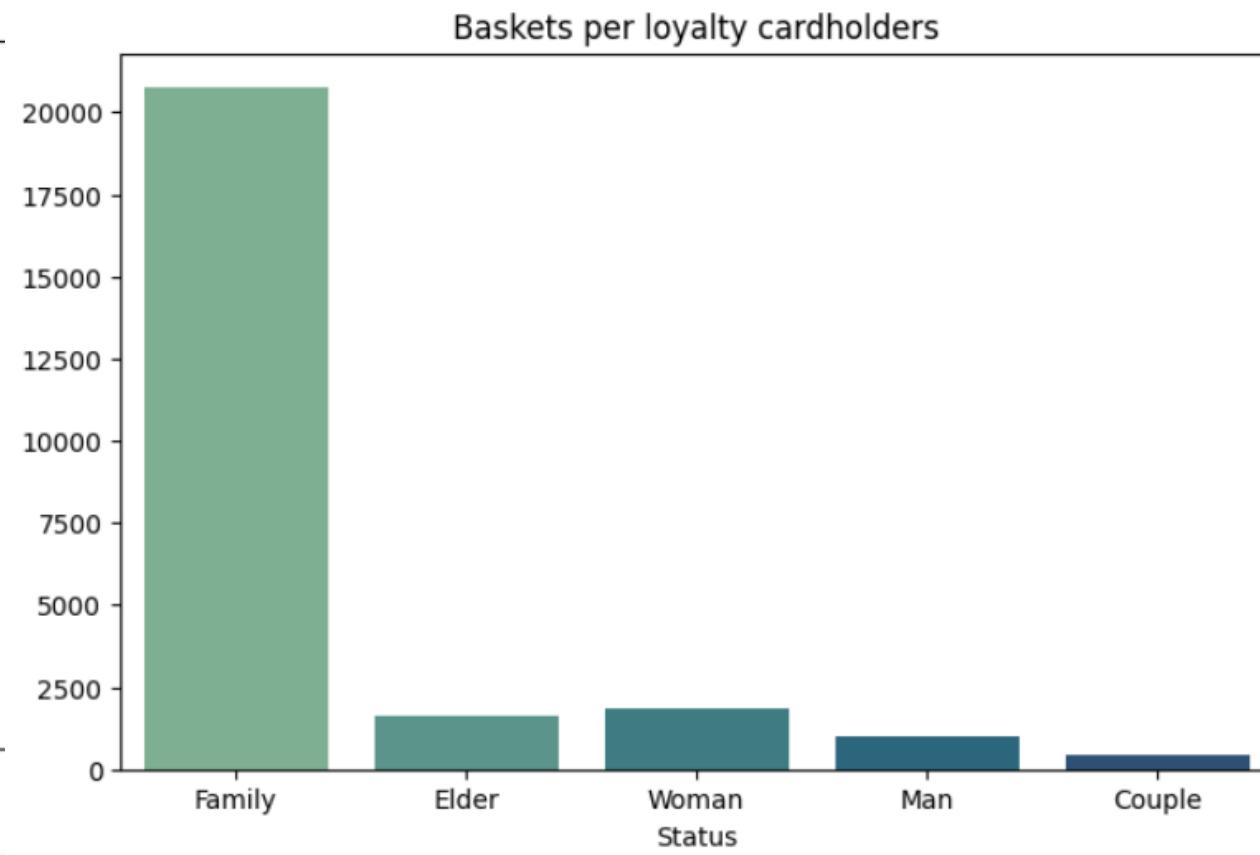
12,4€ per basket



Most visits in October



Most visits on Saturdays



Most visitors are families

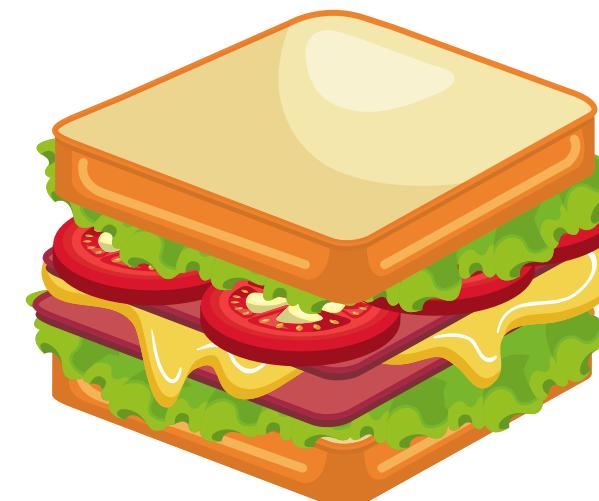
BASKET SEGMENTATION



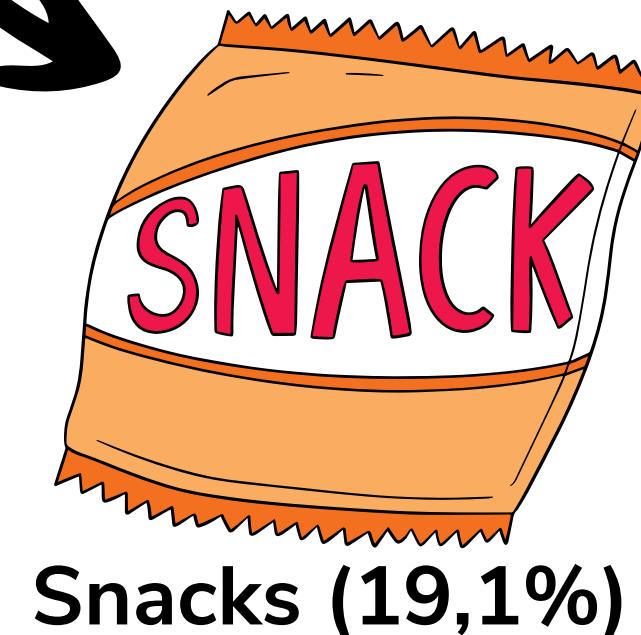
Breakfast (9,14%)



Household
cleaning (26,89%)



Sandwich (12%)



Snacks (19,1%)

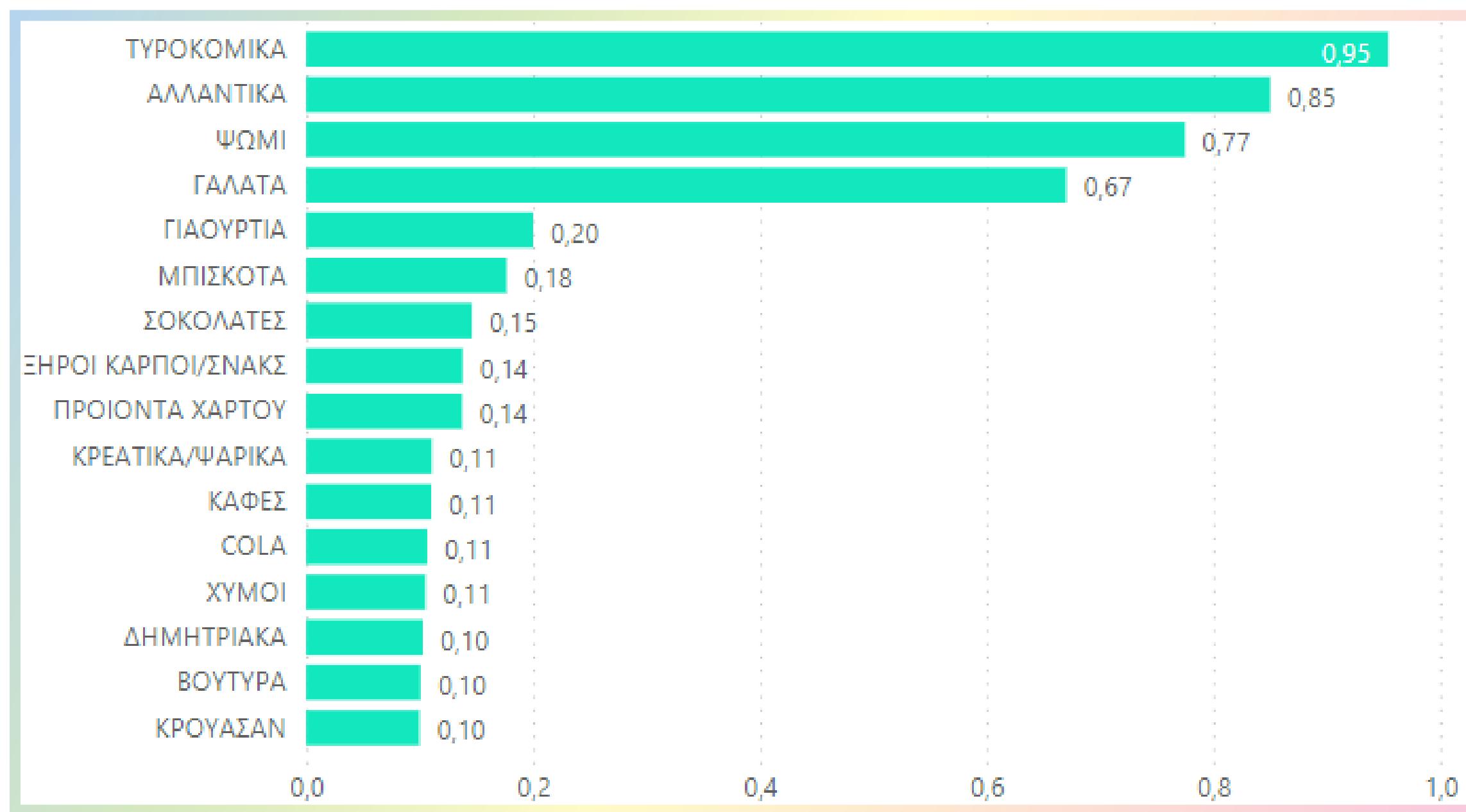
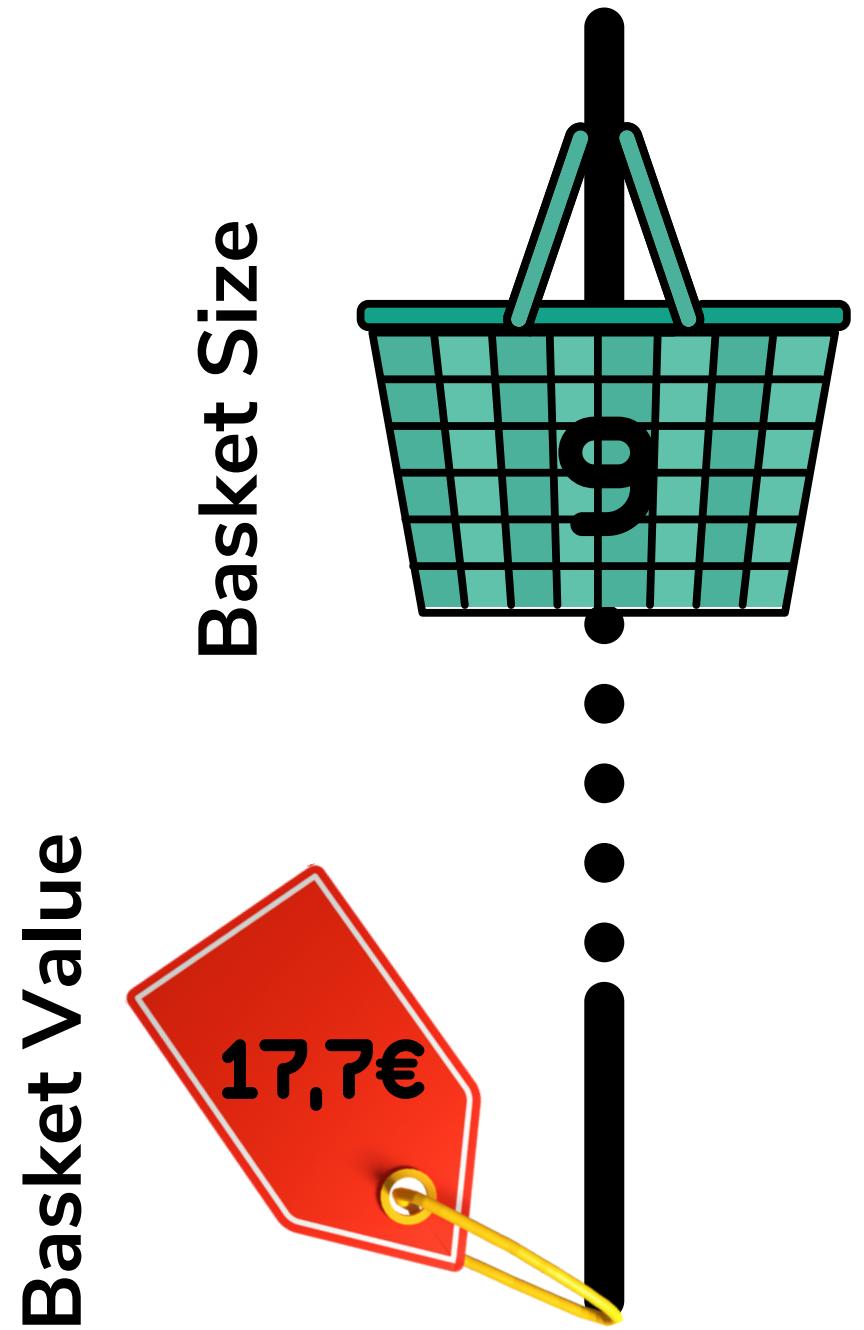
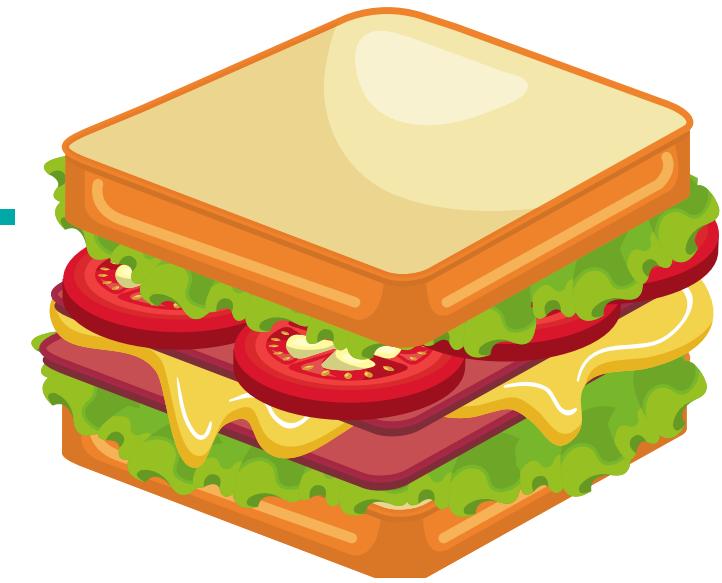


Dairy (17,77%)

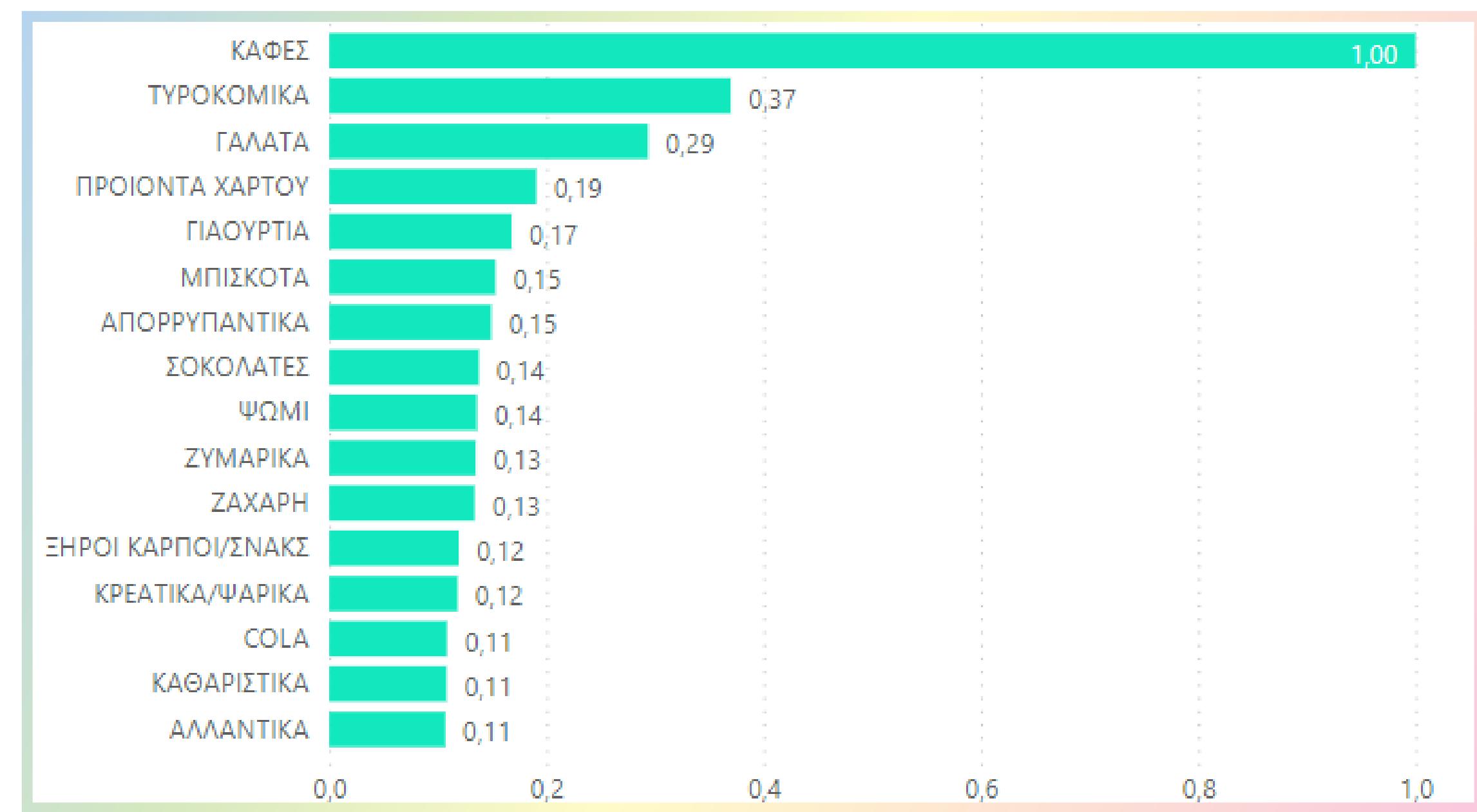
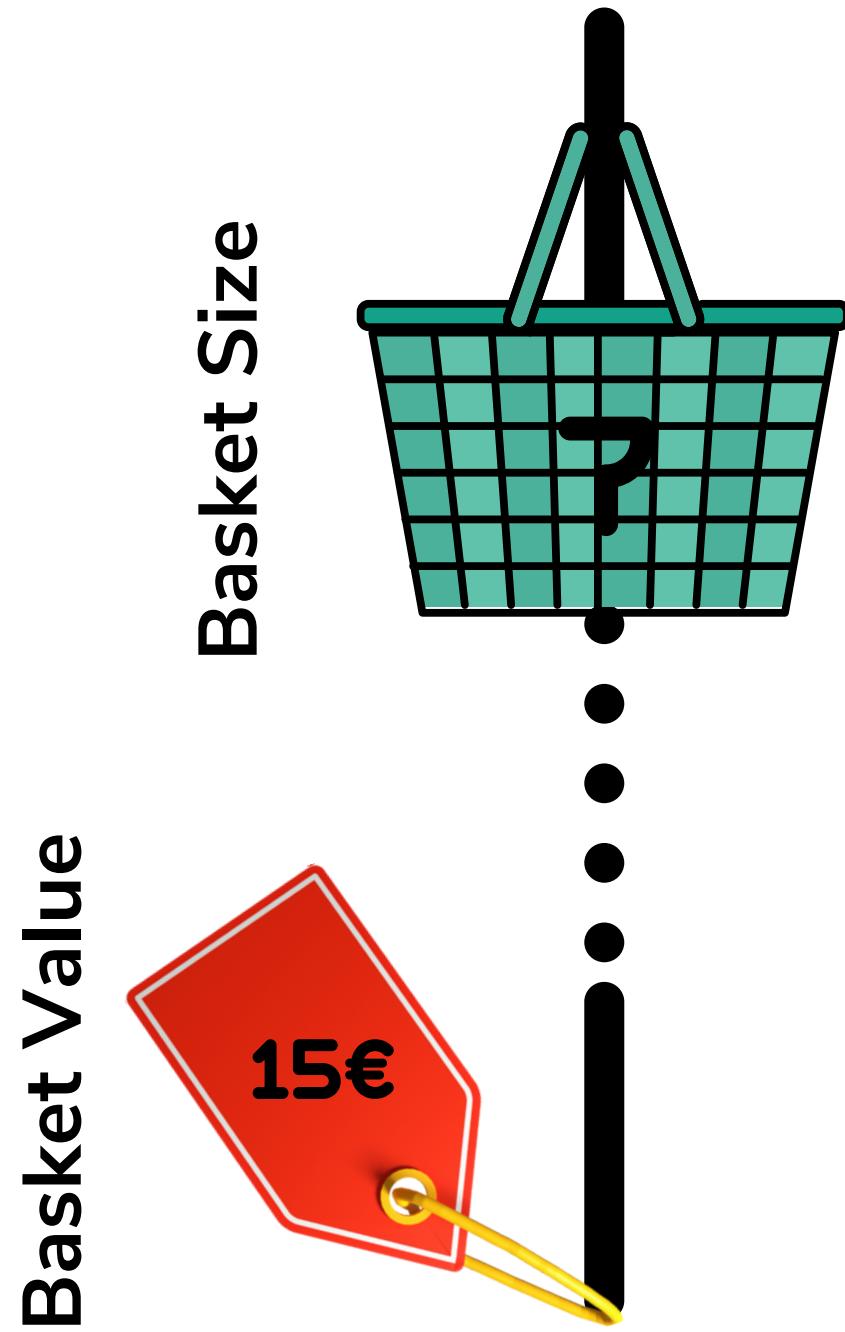


Meal (15,1%)

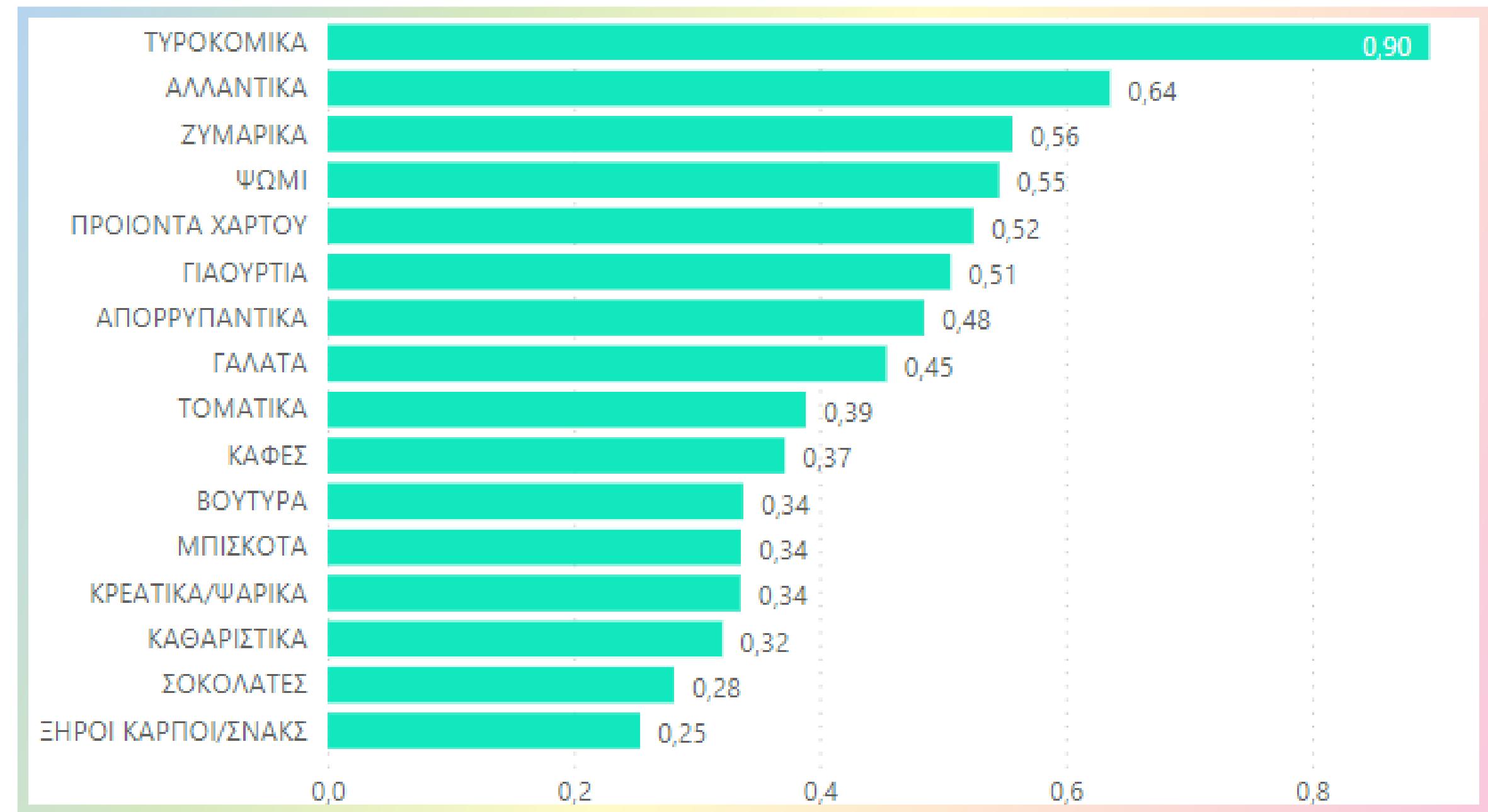
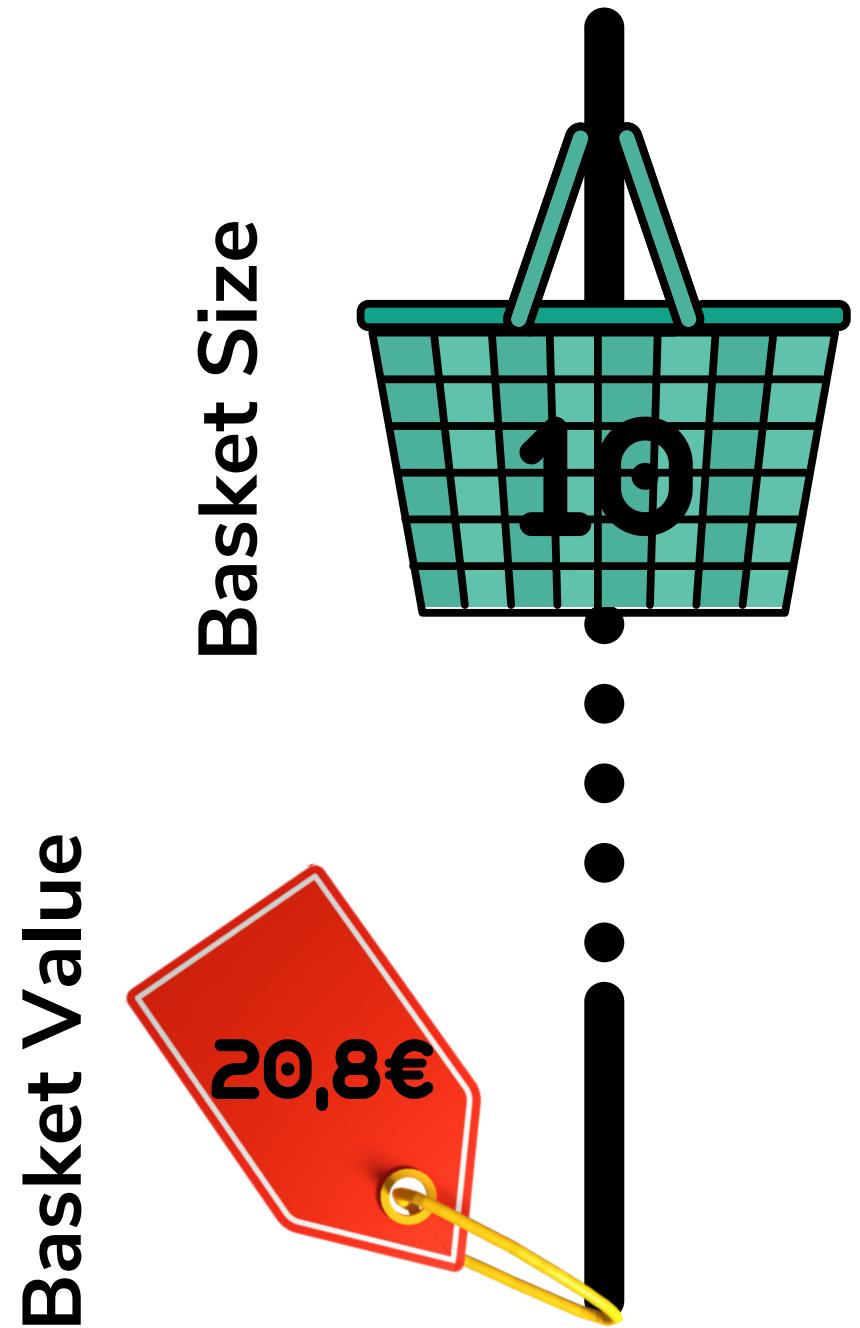
SANDWICH



BREAKFAST



MEAL



CUSTOMER SEGMENTATION

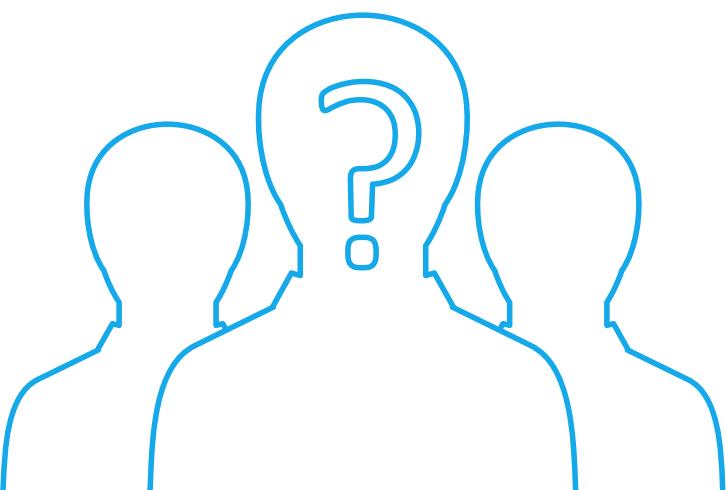
5 GROUPS



Group1 (12,77%)



Group5 (41,74%)



Group2 (16,88%)

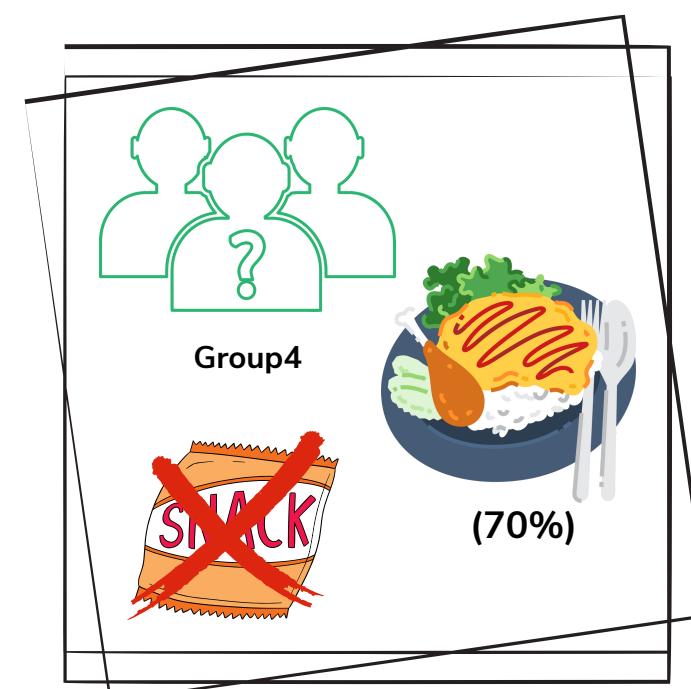
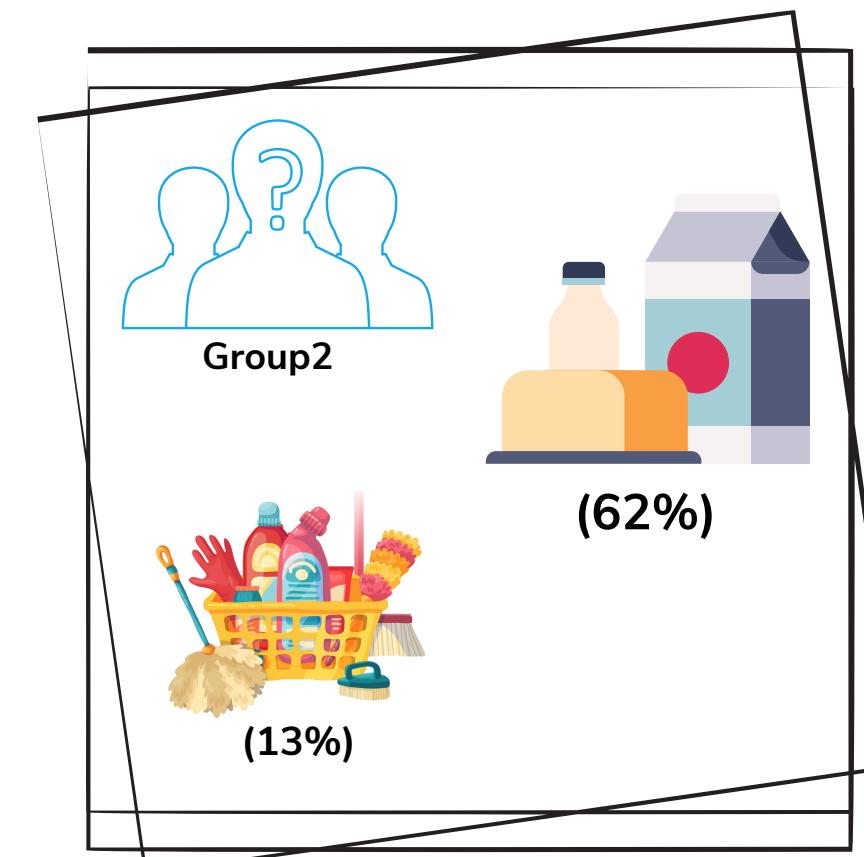
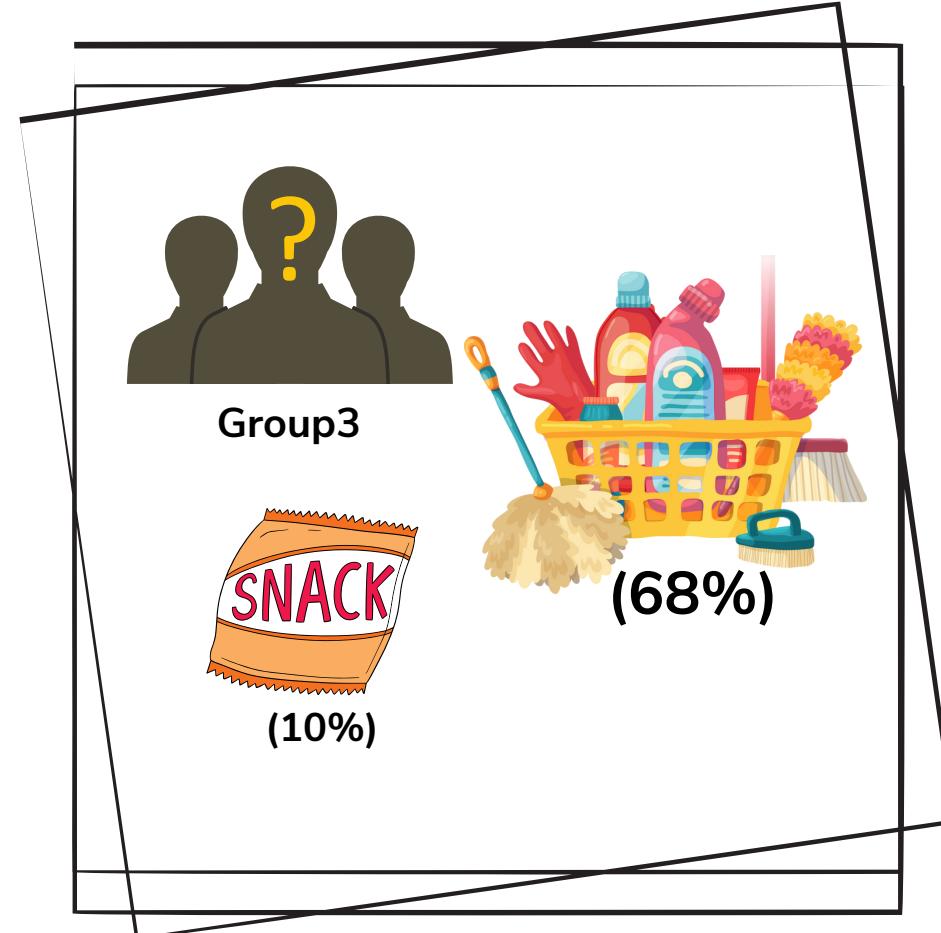
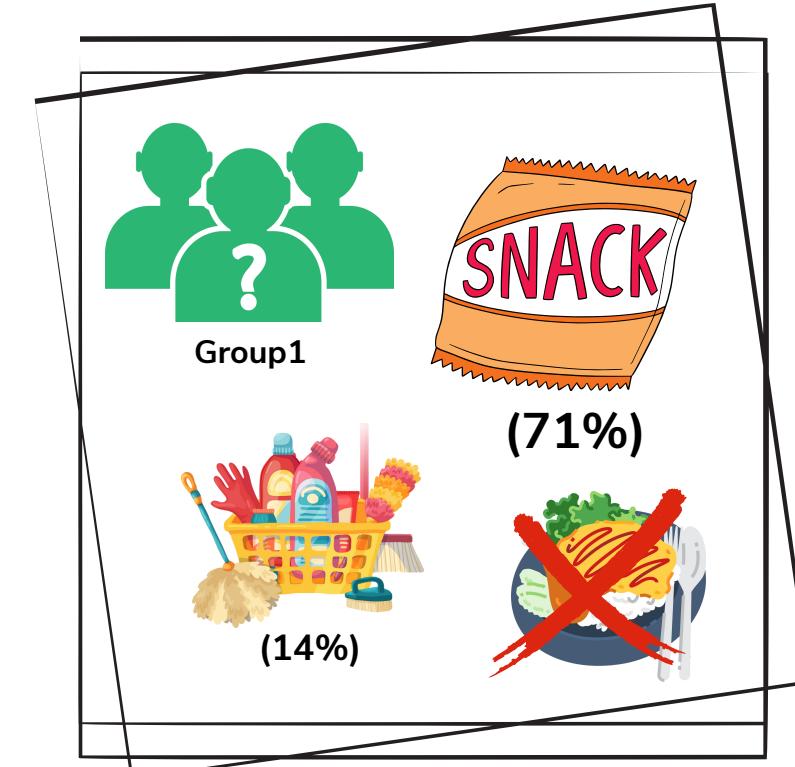
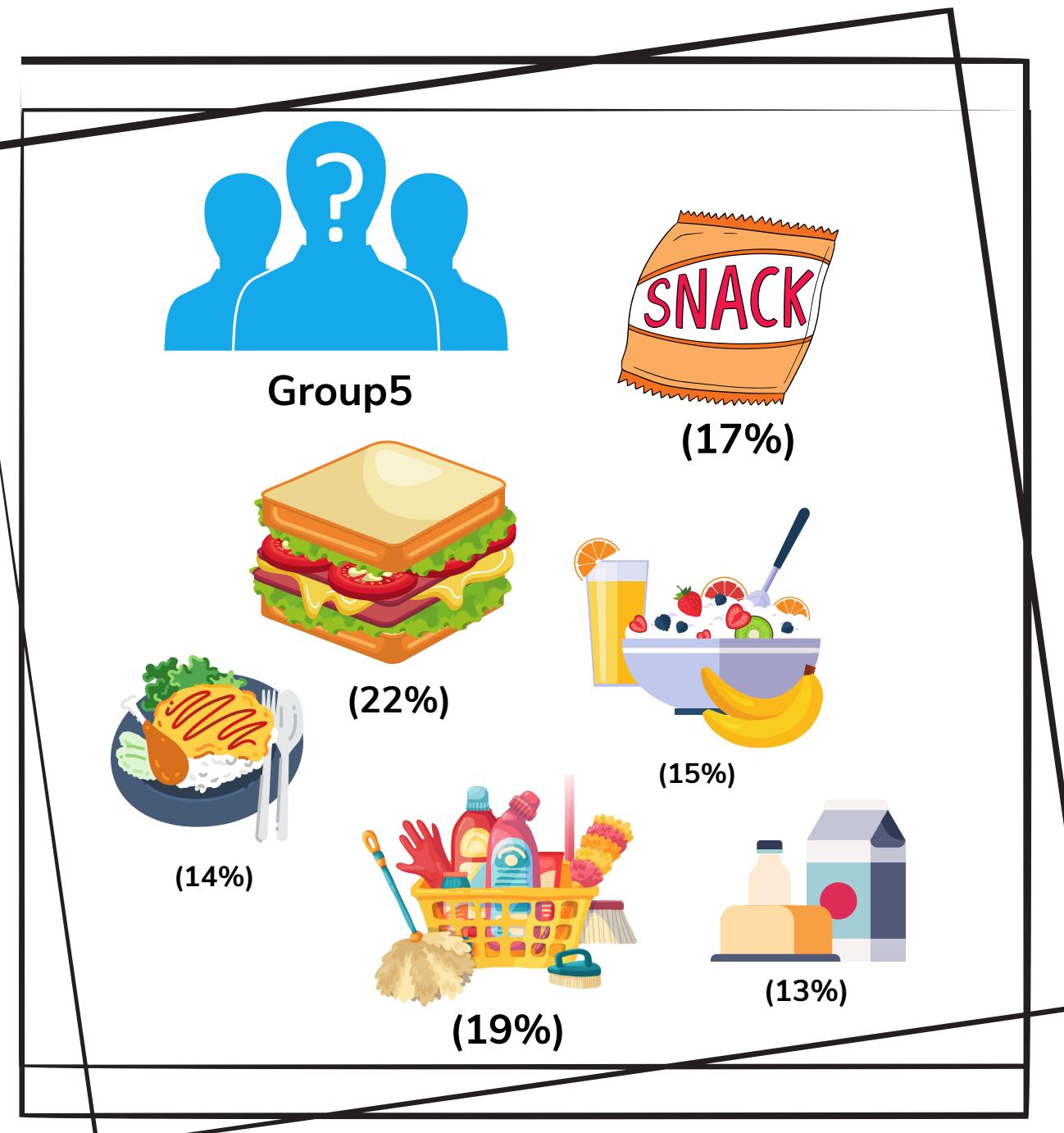


Group3 (20,89%)



Group4 (7,72%)

CUSTOMER PREFERENCES



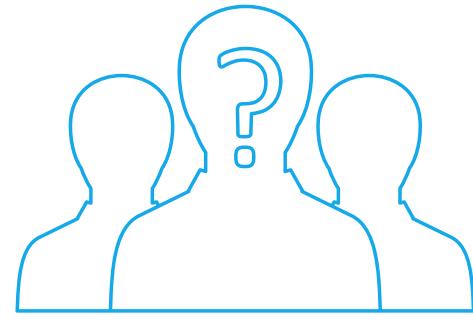
GROUPS IDENTITY



Group3



Group4



Group2



Group4



Group3



Group1

(16%)

(14%)

(10%)

(39%)

(29%)

(15%)

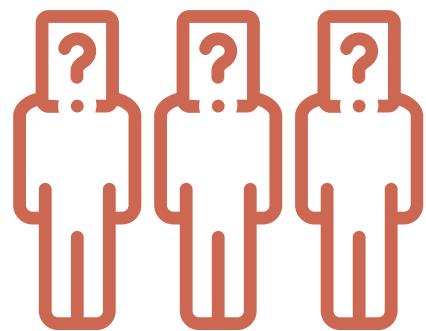
CONCLUSIONS



- Women prefer buying household cleaning products and specific products for meal preparation but they don't prefer buying dairy products

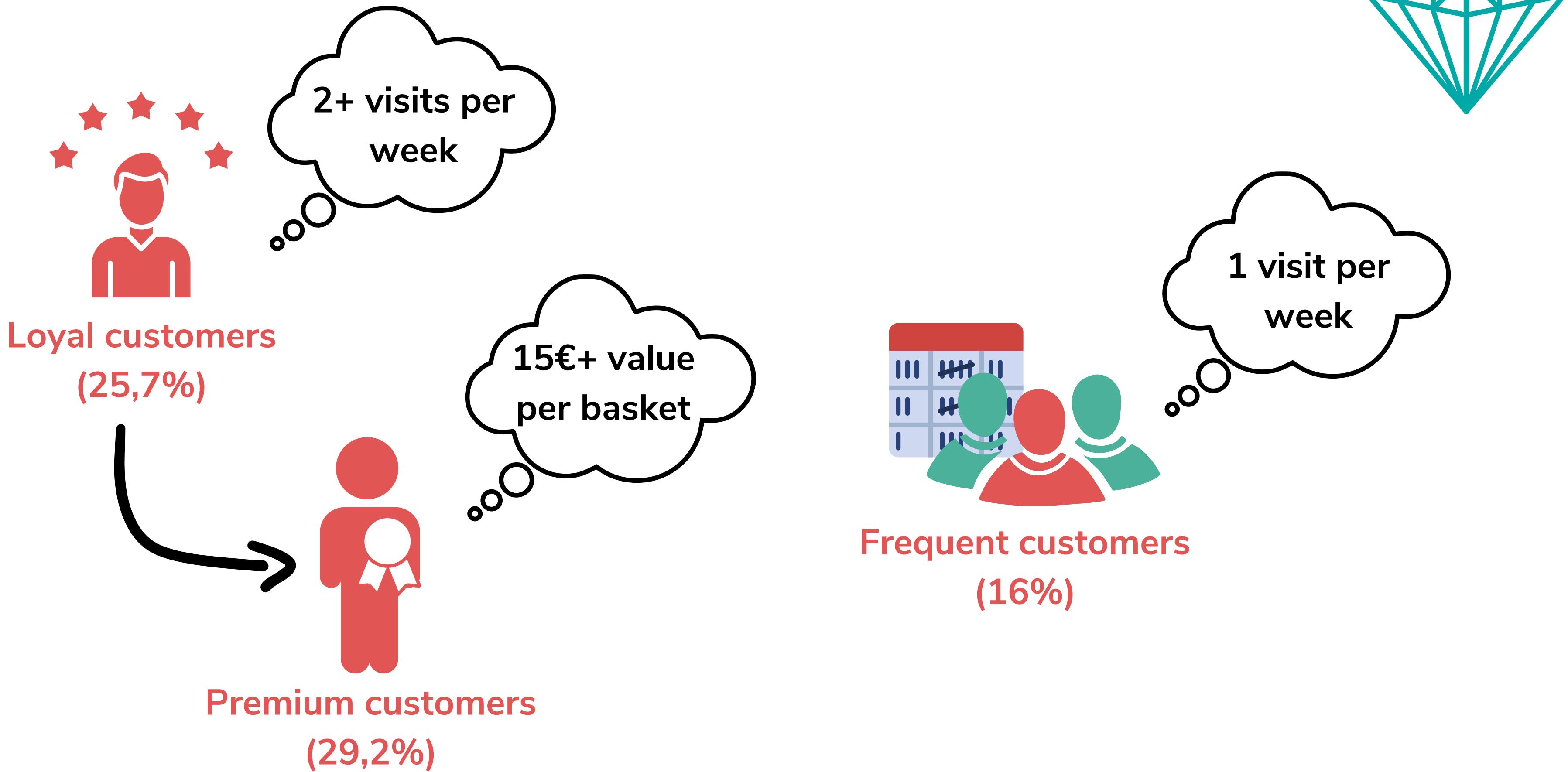


- Families prefer buying mostly products for meal preparation. Also, they prefer household cleaning products but they don't prefer buying snacks

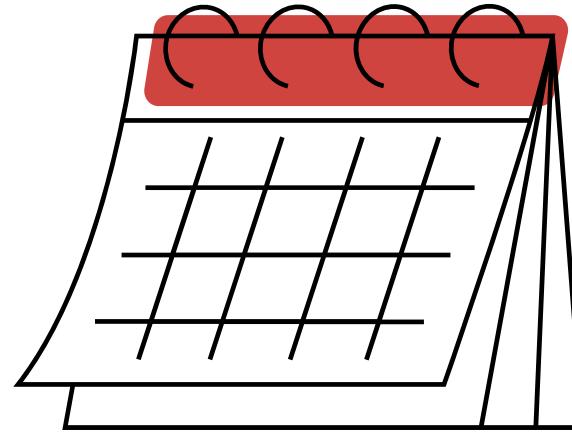


- 43% of data are people with unavailable status. This group seems to prefer buying mostly snacks and not to prefer products for meal preparation. So, we could assume that they are young persons, or they are just passers-by

WHICH CUSTOMERS ADD VALUE?



CUSTOMER TURNOVER



Time of period:
01/09/18 - 31/08/19

1st semester: 01/09/18 - 28/02/19

2nd semester: 01/03/19 - 31/08/19

Churn rate = 13%

13% of customers who visited the supermarket during the 1st semester stopped visiting the supermarket during the 2nd semester

HOW TO IMPROVE CUSTOMER LOYALTY AND SATISFACTION



Premium customers
For the next 4 purchases over
30€ win a 8€ gift card

Premium customers



Loyal customers
For the next 4 purchases over
20€ win a 5€ gift card

Loyal customers



Frequent customers
For the next 2 purchases over
10€ win a 3€ gift card

Frequent customers



Discounts on snacks or cross-
selling in meal products (max
preference) and snacks

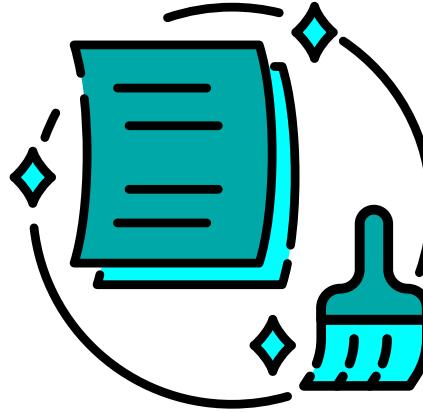


Showing the student ID:
Discounts on prepared meals or
on products for meal preparation



ΠΑΡΑΡΤΗΜΑ

Τεχνικές & Επεξηγήσεις



- Αρχικά, καθαρίσαμε τα δεδομένα αφαιρώντας μη έγκυρα δεδομένα και ακραίες τιμές. Επιπλέον, επιλέξαμε για ανάλυση τα καλάθια μεταξύ 2 έως 50 προϊόντα και απομονώσαμε τον πελάτη με το 20% των αγορών, καθώς θα επηρέαζε αισθητά τα αποτελέσματα του συνόλου
- Για το πρώτο ερώτημα εργαστήκαμε εν τέλει με 45.212 καλάθια. Προκειμένου να βγουν ξεκάθαρες αγοραστικές συμπεριφορές, δημιουργήσαμε μία 4η κατηγορία στα δεδομένα, συνενώνοντας κάποιες κατηγορίες μαζί και συνδυάζοντας την 2η και 3η κατηγορία. Έχοντας λοιπόν αυτά τα δεδομένα, τρέξαμε τον αλγόριθμο K-means με $k = 6$
- Για το δεύτερο ερώτημα εργαστήκαμε ξανά με τον αλγόριθμο K-means με $k = 5$, αλλά χρησιμοποιώντας ως μετρική τις 6 αγοραστικές συμπεριφορές που βρέθηκαν στο ερώτημα 1. Αφότου βρέθηκαν οι 5 αυτές ομάδες, θελήσαμε να εξετάσουμε από τι πελάτες αποτελούνται, με τα δημογραφικά στοιχεία που είχαμε στη διάθεση μας (status).

ΠΑΡΑΡΤΗΜΑ

Τεχνικές & Επεξηγήσεις

2

- Όμως τα στοιχεία αυτά ήταν ελλιπή, καθώς μεγάλο μέρος αυτών ήταν απροσδιόριστα (na). Για τον σκοπό αυτό, σε πρώτη φάση προσπαθήσαμε να παράγουμε δεδομένα γι' αυτά, υποθέτοντας ότι άτομα που αγοράζουν πχ βρεφικά είδη είναι κατά πάσα πιθανότητα οικογένειες, ή αγοράζοντας πάνες ακράτειες είναι ηλικιωμένοι κλπ. Έτσι, τα αποτελέσματα ήταν ότι η πλειοψηφία των δεδομένων παρέμεναν οικογένειες(33%), γυναίκες(16%) και na(41%). Οπότε, διεξήγαμε την ανάλυση μόνο για τα παραπάνω που το δείγμα ήταν αντιπροσωπευτικό. Για τους na είδαμε πως υπήρχε τάση για αγορά σνακς και σχεδόν καθόλου προτίμηση για προϊόντα προετοιμασίας γεύματος. Οπότε υποθέσαμε πως σε ένα μεγάλο ποσοστό αυτή η ομάδα αποτελείται από νέους πχ φοιτητές που δεν μαγειρεύουν αλλά καλύπτουν αυτή την ανάγκη τους αλλού. Για το λόγο αυτό, υπάρχει και η πρόταση μας για προσφορά σε προϊόντα γεύματος υποδεικνύοντας την φοιτητική τους ταυτότητα.

ΠΑΡΑΡΤΗΜΑ

Τεχνικές & Επεξηγήσεις



- Για το τρίτο ερώτημα, αποφασίσαμε να εξετάσουμε ποιοι πελάτες προσδίδουν αξία στο supermarket που αναλύουμε. Έτσι, χωρίσαμε τους πελάτες σε 3 κατηγορίες. Η πρώτη ομάδα είναι οι πιστοί πελάτες οι οποίοι επισκέπτονται το κατάστημα πάνω από 2 φορές την εβδομάδα. Η δεύτερη ομάδα είναι υποσύνολο της πρώτης, και αποτελείται από τους πιστούς πελάτες οι οποίοι κάνουν αγορές άνω των 15€ σε κάθε επίσκεψη τους, οι επονομαζόμενοι premium πελάτες. Τέλος, η τρίτη ομάδα αφορά τους συχνούς πελάτες, δηλαδή αυτούς που επισκέπτονται το κατάστημα 1 φορά την εβδομάδα. Για όλες τις παραπάνω ομάδες, σκεφτήκαμε ότι ένα κίνητρο για να συνεχίσουν να αποτελούν πελάτες αλλά και ίσως να "ανέβουν" επίπεδο, είναι η προσφορά δωροκαρτών αξίας ανάλογης της ομάδας που κατατάσσονται και της αξίας του καλαθιού που αγοράζουν. Τέλος, αξιοσημείωτο θεωρήσαμε να αναφερθεί και το ποσοστό των πελατών που χάθηκε μέσα σε αυτό το έτος, συγκρίνοντας το Α' εξάμηνο των δεδομένων με το Β' εξάμηνο.

THANK YOU!
