Customer segmentation for the new market

Background

An automobile company has plans to enter a new market with their existing products. They have deduced that the behaviour of the new market is similar to their existing market where all customers have been classified into 4 segments: A, B, C, and D.

However, machine learning models may be able to segment customers more effectively, allowing for even more tailored marketing.

Current market segments

Customer data includes the following features:

- ID
- Gender
- Marital status
- Age
- Graduate status
- Profession — healthcare, engineer, lawyer, entertainment, artist, executive, doctor, homemaker, or marketing
- Work experience — measured in years
- Spending score — low, average, or high
- Family size — including the customer
- Var 1 — an anonymised category

Current market segments

A

- Early 40s
- Works in arts or entertainment
- Graduate
- Married
- Spending score: Low
- Family size: 1-2

B

- Late 40s
- Artist
- Graduate
- Married
- Spending score: Low to average
- Family size: 2

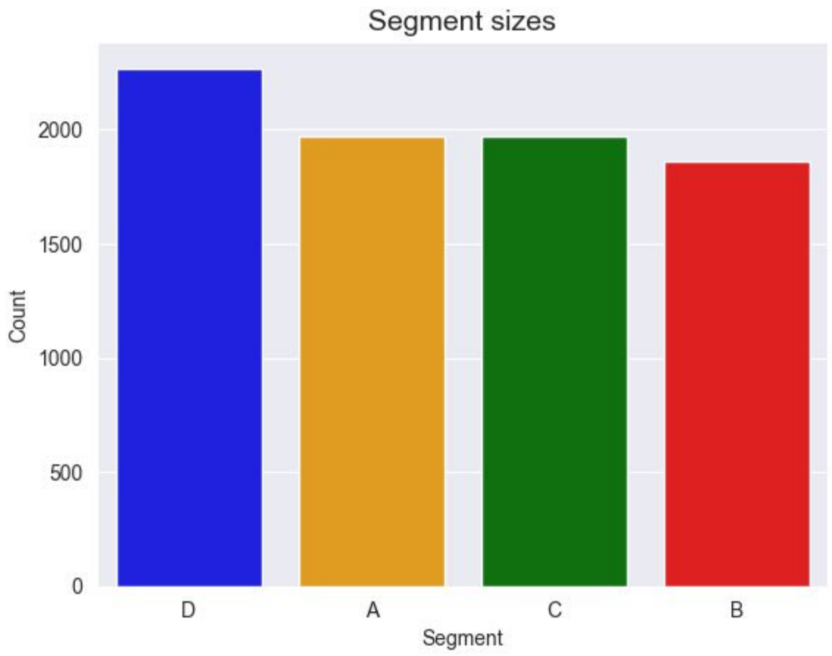
C

- Late 50s
- Artist
- Graduate
- Married
- Spending score: Average
- Family size: 2+

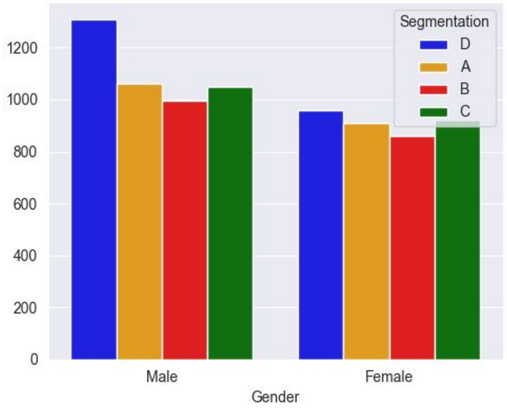
D

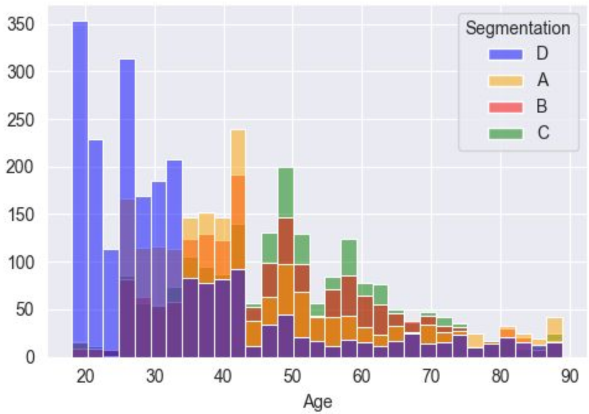
- Early 20s
- Works in healthcare
- Non-graduate
- Unmarried
- Spending score: Low
- Family size: 3-4

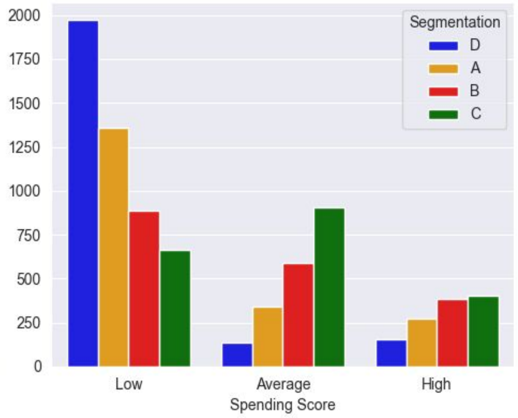
Segment sizes



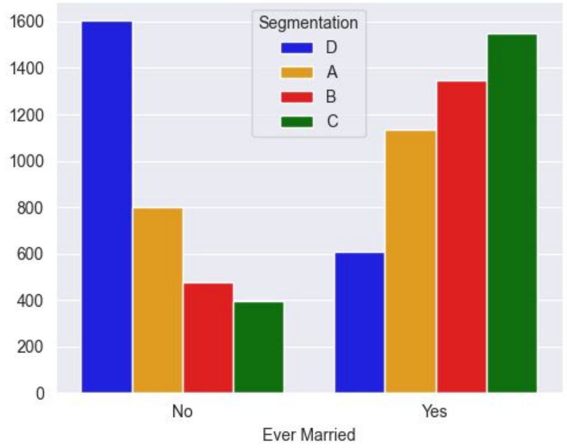
Segment details

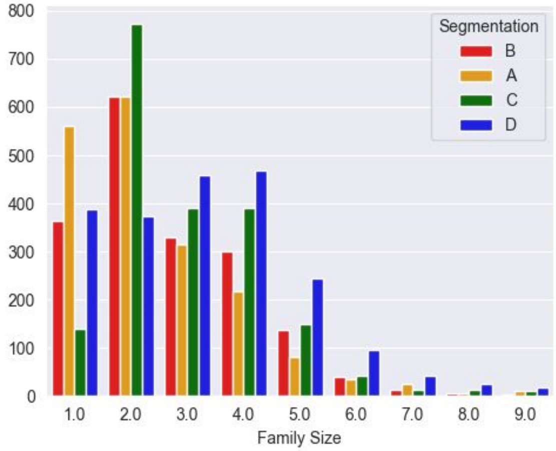


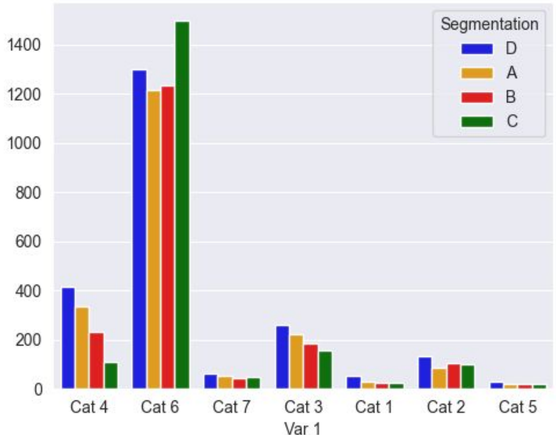




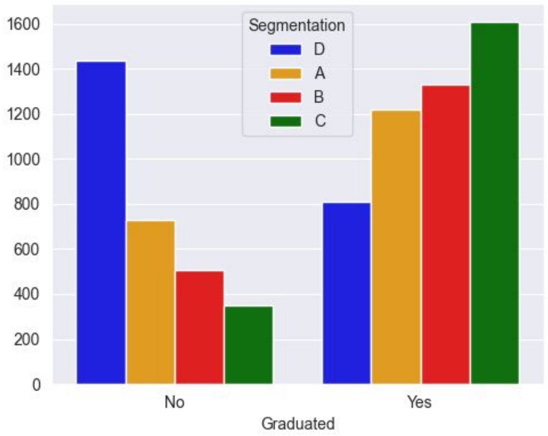
Segment details

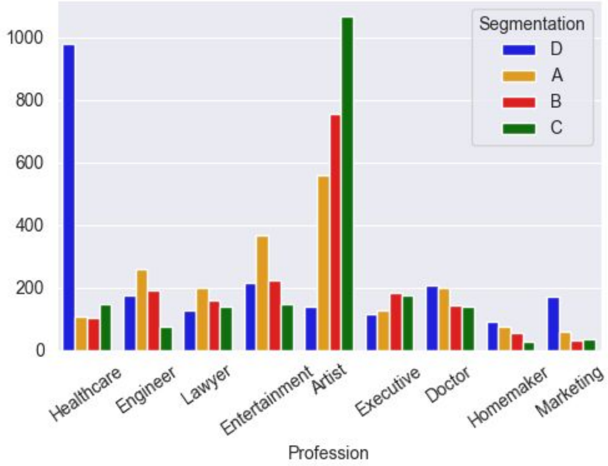


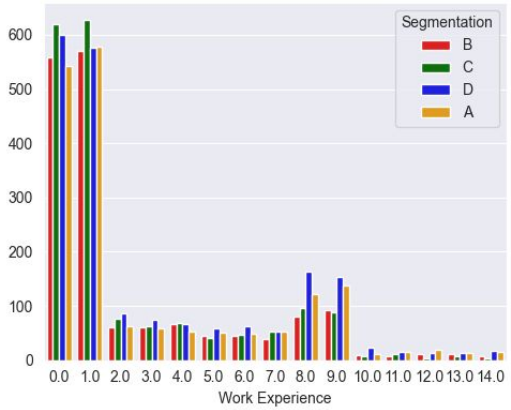




Segment details







A different way to segment

In the existing segments, there is a lot of overlap in characteristics between each group.

Depending on the data, machine learning models can be used to effectively group more distinct customer segments.

Two models have been used to re-segment the market:

- K-Means Clustering
- Hierarchical Clustering

The results

Multiple methods were used to find the ideal number of clusters – the segments containing customers with the most similar characteristics.

8 clusters were found to be optimum for both models and both produced near-identical results.

Using the outputs of the Hierarchical Clustering model, we can categorise and define the new segments.

New customer segments

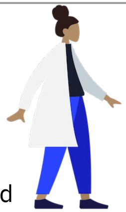
A

- 50s
- Male
- Graduate
- Primarily artists
- Married
- Family size: 2
- Spending score: Average, but falls into other categories

B

- 50s
- Female
- Graduates
- Primarily artists
- Married
- Family size: 2
- Spending score: Average, but falls into other categories

C



- 25-50
- Female
- Non-graduates
- Primarily engineers and lawyers
- Married
- Family size: 2
- Spending score: Average, but falls into other categories

D

- 50s
- Male
- Non-graduates
- Primarily executives
- Married
- Family size: 2+
- Spending score: falls evenly into each category

E

- 30s
- Male
- Graduates
- Primarily work in arts, healthcare, and entertainment
- Unmarried
- Family size: 1-3
- Spending score: Low

F

- 30s
- Female
- Graduates
- Primarily work in arts or healthcare
- Unmarried
- Family size: 1
- Spending score: Low

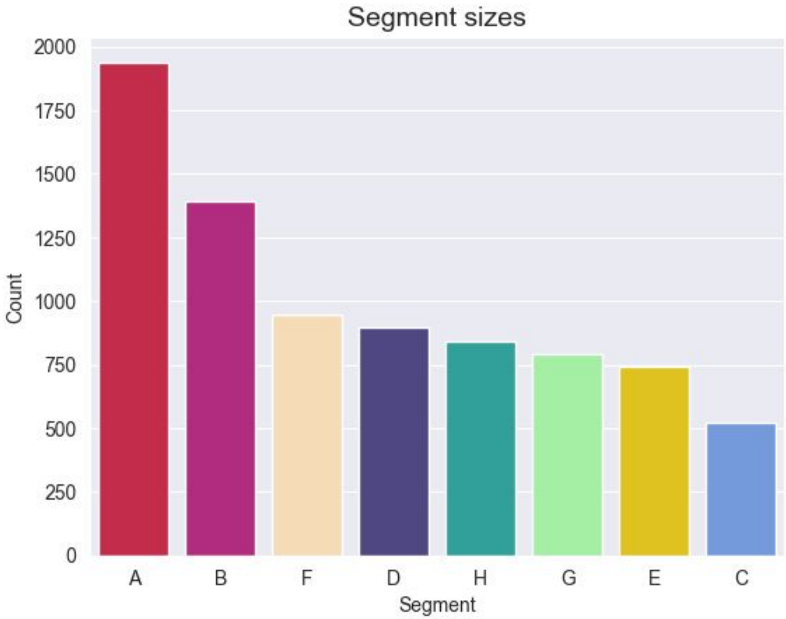
G

- 20s
- Female
- Non-graduates
- Primarily work in healthcare
- Unmarried
- Family size: 3-4
- Spending score: Low

H

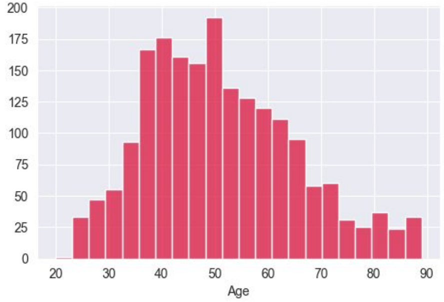
- 20s
- Male
- Non-graduates
- Primarily work in healthcare
- Unmarried
- Family size: 4
- Spending score: Low

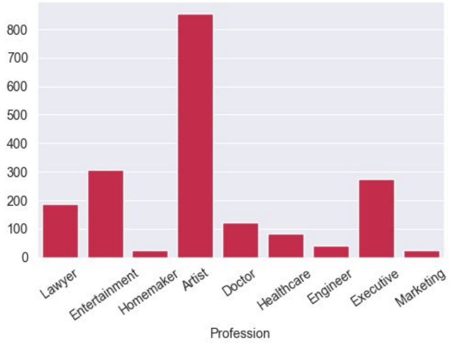
Segment sizes

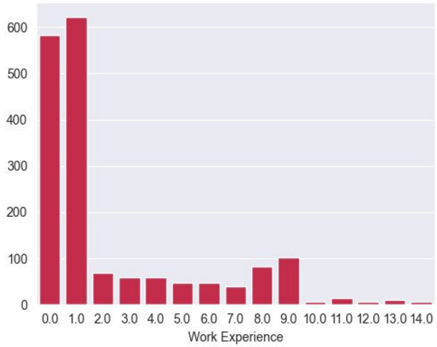


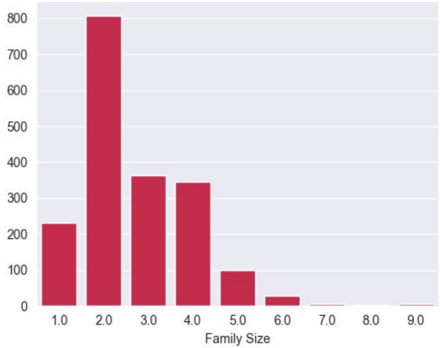
Segment A breakdown

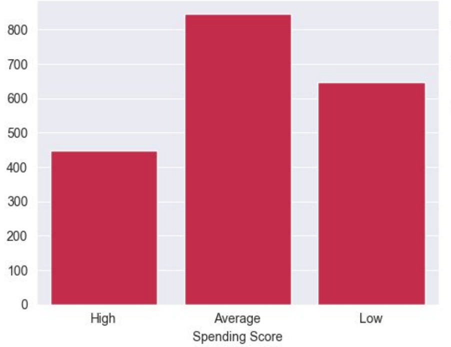
- Male
- Married
- Graduates





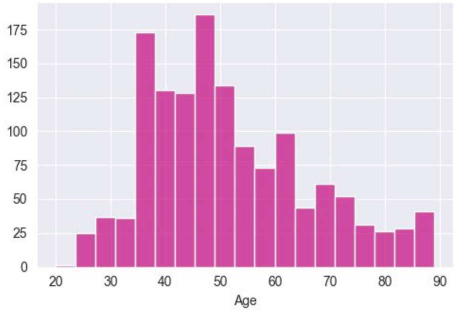


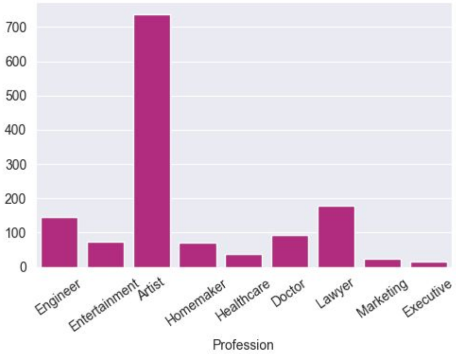


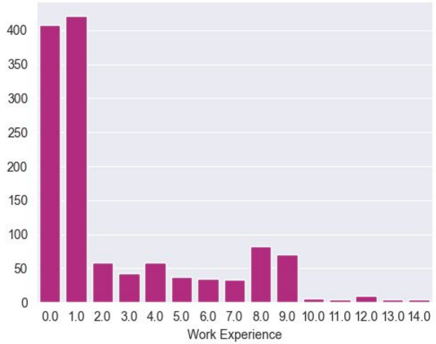


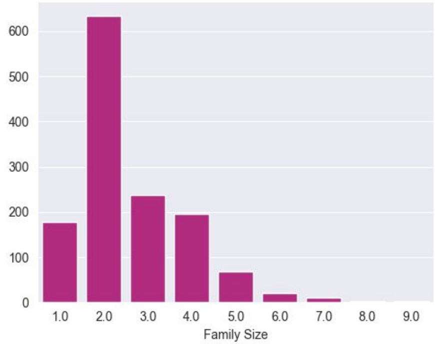
Segment B breakdown

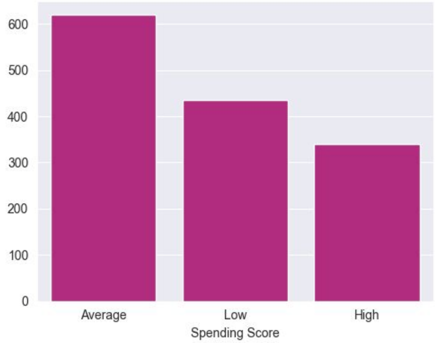
- Female
- Married
- Graduates





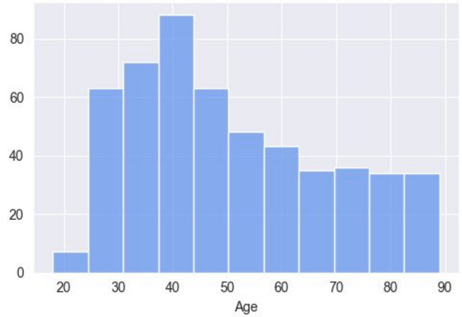


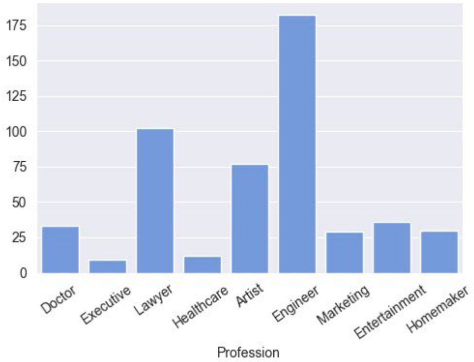


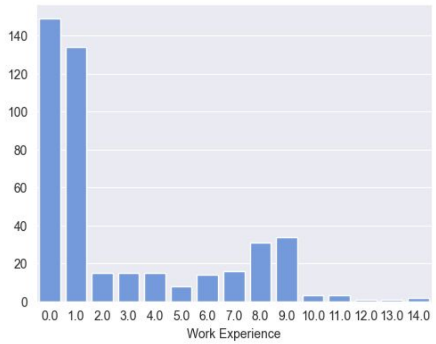


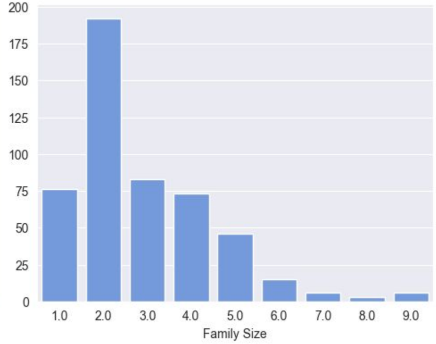
Segment C breakdown

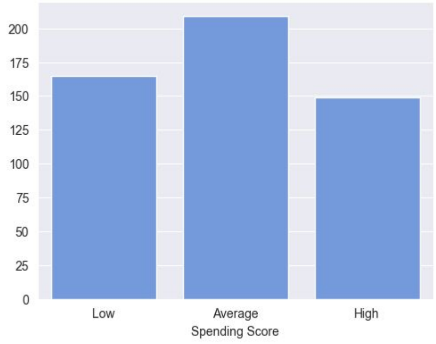
- Female
- Married
- Non-graduates

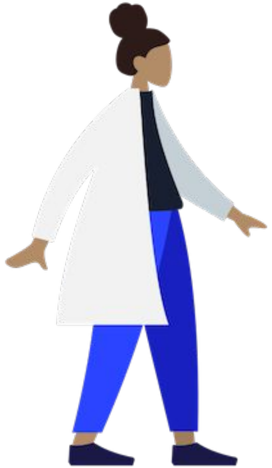






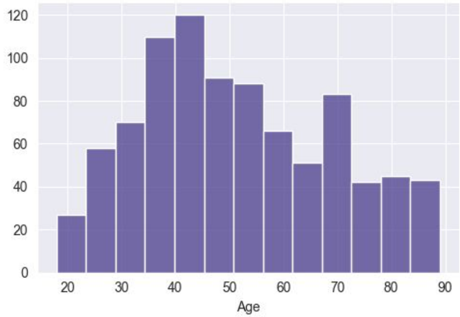


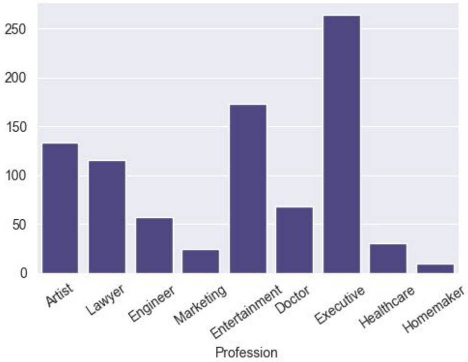


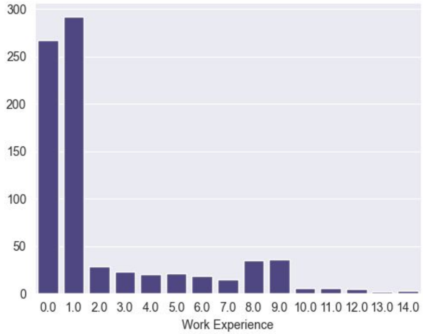


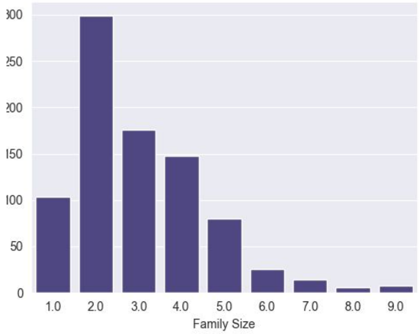
Segment D breakdown

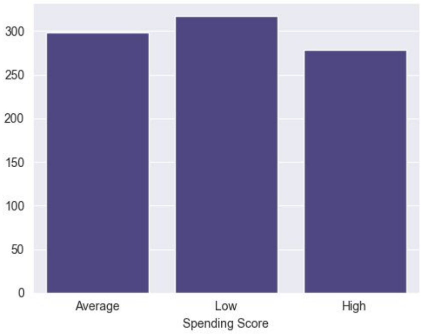
- Male
- Married
- Non-graduates





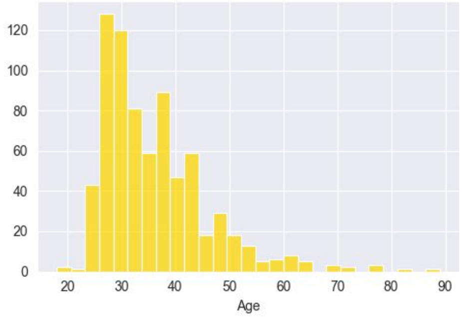


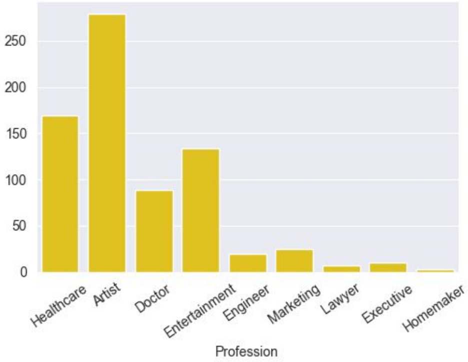


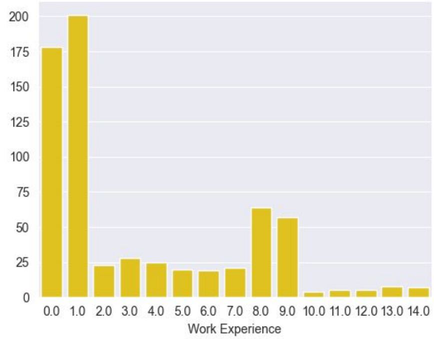


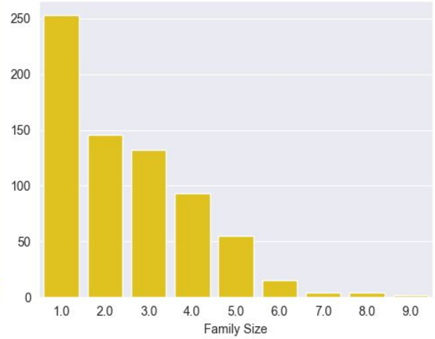
Segment E breakdown

- Male
- Unmarried
- Graduates
- Low spending score



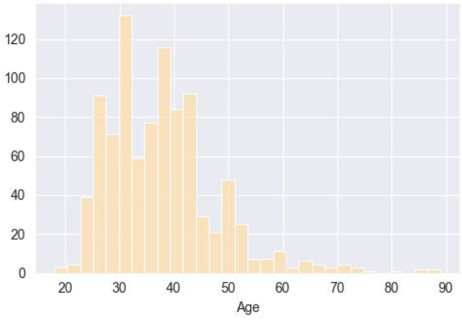


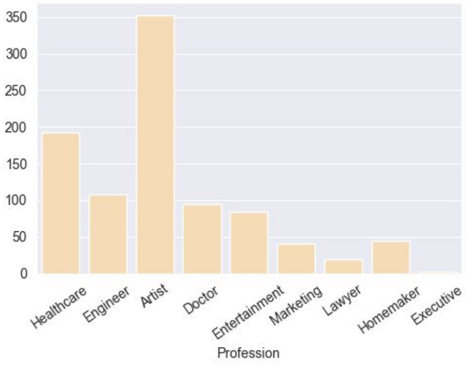


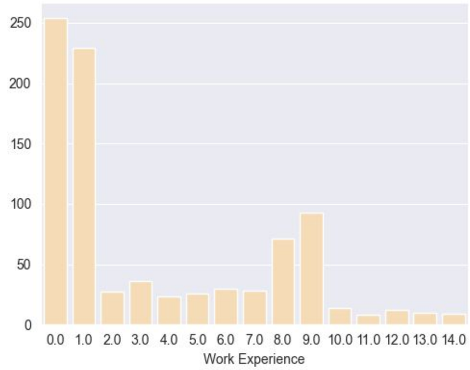


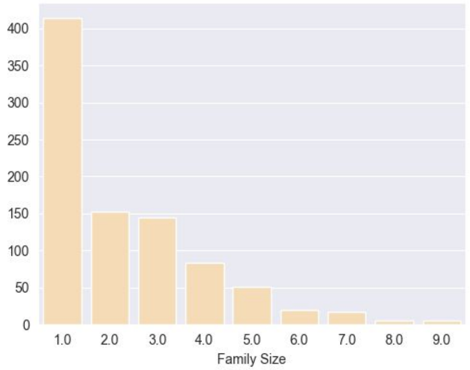
Segment F breakdown

- Female
- Unmarried
- Graduates
- Low spending score



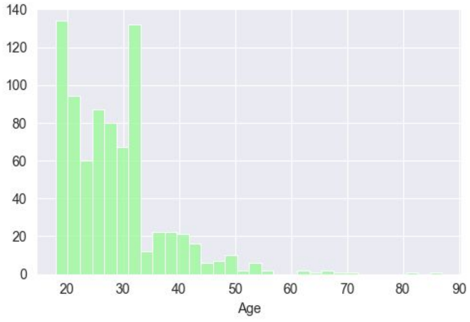


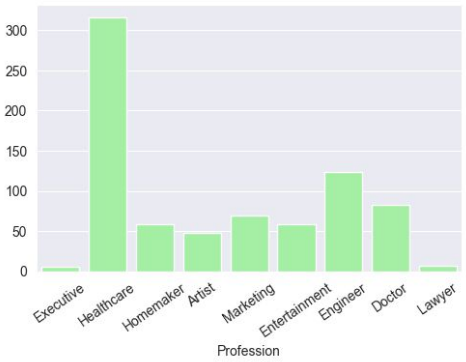


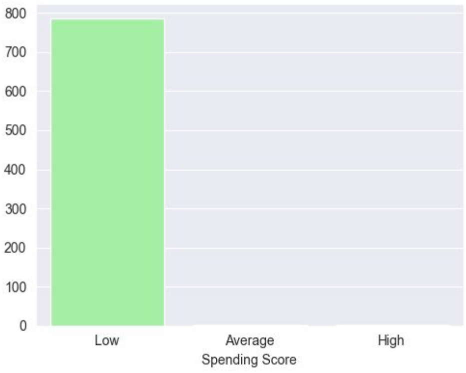


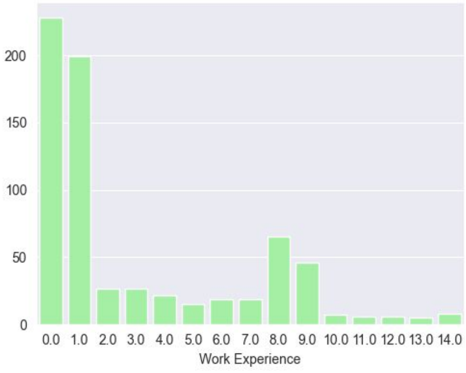
Segment G breakdown

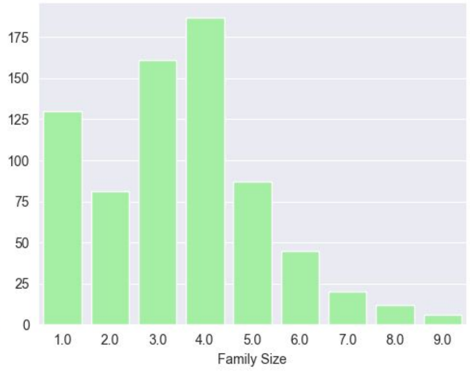
- Female
- Unmarried
- Non-graduates
- Mostly low spending score





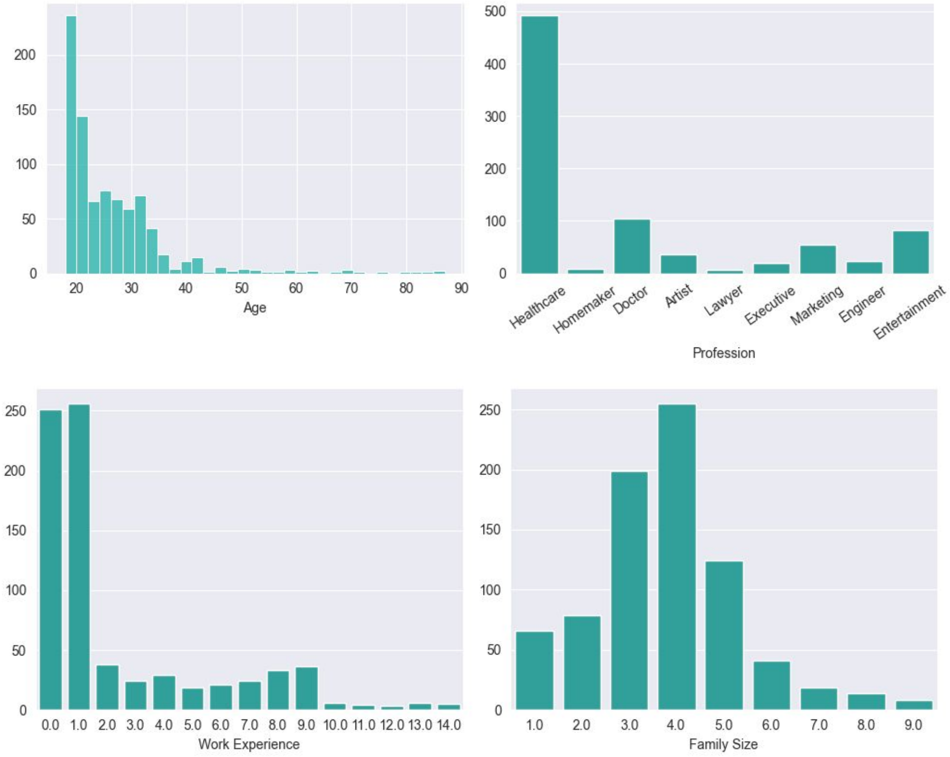






Segment H breakdown

- Male
- Unmarried
- Non-graduates
- Low spending score



Conclusion and recommendations

8 clusters are considered the optimum to split these markets into more defined segments, allowing for more tailored marketing and outreach.

The models performed well at separating binary categories such as gender, marital status and graduate status. They were also good at identifying segments with a low spending score.

However, the segments can also be joined together by age, gender, spending score, professions etc. based on company needs and strategy.