

# Executive Summary

## of customer gender analysis

### Scope and Objective

Due to privacy reason, customer gender information is missing to a large extent to THE ICONIC, however gender is important for customer profiling to improve the business. This analysis is to infer customer gender using data science way based on given customer data, a summary of model solution, evaluation and suggestions will be given.

### Data preparation and model build

Data cleaning and feature engineering will apply while considering both common approach and the specific situation for this case. After evaluating the major unsupervised models versus our situation and aim, K-means clustering is chosen.

### Model evaluation

The clustering result shows a qualitative correctness by comparing with the index of *female\_item / male\_item*, however the accuracy is hard to tell as there is no real gender label to compare. On the other hand, by calculating the correlation between inferred gender and other given features, it shows the result clusters relate largely to the number of purchased items rather than gender related.

### Model conclusion

Based on the above evaluation, this clustering model is still premature for practical usage as more gender-related features are needed.

### Suggestions

To have better accuracy for the clustering model, the following strongly gender-related features are suggested:

- Gender itself
- Time spent when shopping
- Number of saved items
- Name, or at least nick name
- Number of kid items bought