# Analysis Report: Age Assignment Model Comparison
# Trust Stamp Assignment

Pablo Martínez-Agulló

September 2024

## 1 Introduction

Hanut, an online retailer, wants to enhance customer experiences by personalizing them based on age demographics. After exploring image-based age estimation solutions, they have shortlisted two models and are seeking our analysis to determine which model to adopt. Their consumer age profiles are segmented into specific brackets: 0-12, 13-15, 16-17, 18-24, 25-30, 31-40, 41-50, 51-60, 61-70, 71-80, and 81+.

The two shortlisted models are: Model 1, which predicts an age range (`age_min` and `age_max`) with 6975 entries in `data/model_1.csv`, and Model 2, which provides a direct age prediction (`age`) with 6969 entries in `data/model_2.csv`. The ground truth data is stored in `data/gt.csv` with 8644 entries. Both models' datasets contain some missing entries compared to the ground truth. After merging the datasets, the comparison was performed on 6957 complete entries, with missing data removed.

## 2 Data Exploration

Figure 1 reveals a trend where both models underestimate age as the real age increases. The $R^2$ value suggests a linear relationship, particularly for Model 1. In all figures, the predictions for Model 1 are represented by the average of `age_min` and `age_max`.

The horizontal alignment of points for Model 2 (blue) on Figure 1 indicates that its predictions might be discretized to specific values instead of providing more granular continuous predictions. The low $R^2$ score of Model 2 may be caused by this.

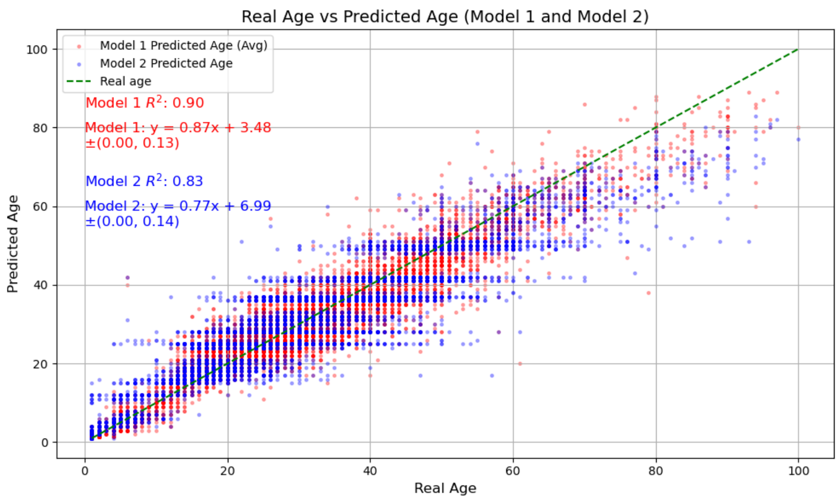



Figure 1: Scatter plot of predicted age for both models as a function of the real age. The coefficient of determination $R^2$ evaluates how well the predictions from the models (Model 1 and Model 2) align with the actual "real age" value.

Figure 2 shows the predicted age of each model across different real age brackets. While both models tend to overestimate the age for younger brackets and underestimate for older ones, Model 1 generally aligns more closely with actual ages. It is important to note that while the average prediction for certain age brackets may align with true values, this does not necessarily imply strong model performance. The

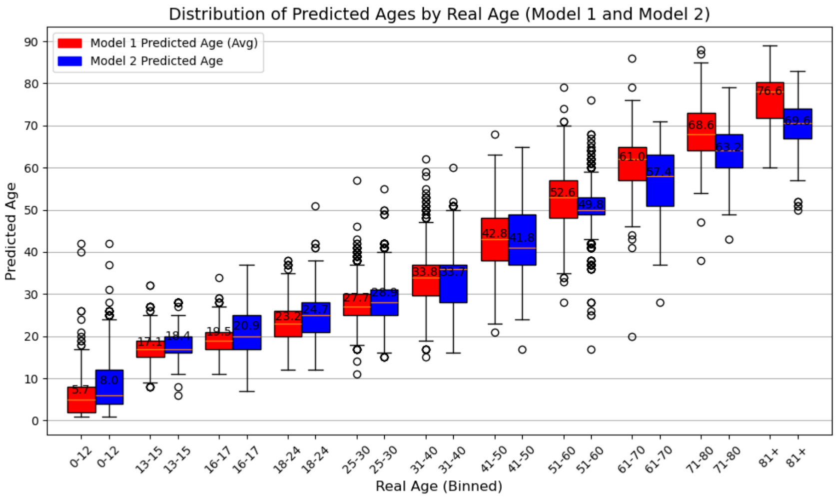
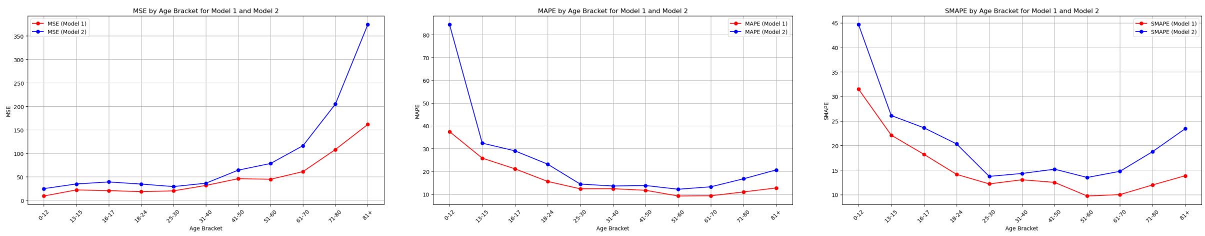



Figure 2: Box plot of the average predicted age of Models 1 and 2 as a function of the age bracket. The numbers on the boxes correspond to the model's average.

alignment may result from overestimations and underestimations canceling each other out, as suggested by the distribution of errors in Figure 1.

To quantify the models' performance, we evaluate the discrepancy between the true age and predicted values using mean squared error (MSE), mean absolute percentage error (MAPE), and symmetric mean absolute percentage error (SMAPE). MSE emphasizes larger errors, MAPE provides percentage-based insights, and SMAPE offers a symmetric evaluation of errors. As shown in Figure 3, Model 1 consistently yields lower error rates across all metrics, outperforming Model 2 across all age groups.




(a) MSE (b) MAPE (c) SMAPE

Figure 3: Most common metrics to asses the accuracy of a model.

# 3 Conclusion

This analysis demonstrates that both models exhibit similar trends in overestimating younger consumers' ages and underestimating older consumers' ages. However, Model 1, which provides age range predictions, consistently performs better across age groups than Model 2, which directly predicts ages.

Evaluation using MSE, MAPE, and SMAPE metrics confirms that Model 1 achieves lower error rates across all age brackets. Notably, Model 1's performance is less affected by extreme age groups, suggesting greater robustness in diverse age scenarios. Model 2, in contrast, shows significant underestimation for older age groups, which diminishes its utility for personalized experiences in such demographics.

Given these results, Model 1 is recommended for Hanut's age-based customer segmentation. Its superior performance across multiple metrics makes it the more reliable option for age prediction across all age brackers. Future work should focus on refining the models to reduce misestimations at extreme ages and incorporating additional demographic data for improved predictive accuracy.