

Dear Reviewer 2:

Thank you for your response and thank you for your thorough comments.

I will begin by quoting your comment:

For Category 1, F1-scores are high for the case of using Advert Viewing Time alone (F1-score = 0.85 for the case of Actual Purchase), and this is rightly pointed out in the manuscript as: "Predictions seem to be high across all models,". However, it then continues as: "but this is irrelevant since the customer is not changing their negative purchase behavior." Why is Category 1 deemed to be irrelevant? For this category (Category 1), the model appears to be predicting well based on the Advert Viewing Time alone. The customer may not be changing their behavior, but the model is predicting this quite well.

The issue is important because, according to Table B.20, Category 1 is more than 70% for the Actual Purchase case, a significant portion of the data. If we refer to this category as "irrelevant", would it not make the data for Category 1, which is the majority of experiment data, also irrelevant?

It is true that Category 1 has a high performance for Actual Purchase. Category 5, by its nature of including Category 1 samples, is also performing well. As you mentioned above, 70% of the people are not buying at the point of January 2017, and they are also not buying in March 2017 regardless of the number of commercials they saw. This data is important to the model because it reflects real life distributions of customers.

However, in this case, it doesn't matter how well the predictions seem to be performing via the F1 measure on this Category 1, because the models cannot be said to be linked to the television adverts. If we observe the t-test results for Hypotheses 1 and 3 for Category 1 and 5, we can observe that there is not a significant difference between advert viewing based models and demographic based models or their combinations thereof.

I modified my previous text about the result being "irrelevant" and added a paragraph explaining the results for Category 1 in section 7.1 Influence of TV adverts on Actual Purchase and Purchase Intention:

*"The exception to this rule is in category 1, where customers consistently answered "NO" in their purchase recall or purchase intention questions of the survey both in January 2017 and March 2017. Subsequently, this also influences category 5 results. Predictions seem to be high across all models when the customer has a negative purchase behavior. However, the t-test results for Hypotheses 1 and 3 show us that there is not a statistically significant difference between demographic data models and advert viewing time models. Because of these results, the factors that are influencing negative purchase decisions cannot be determined to be either advert viewing time or otherwise."*

Now, one could think that there could be a benefit in predicting negative purchase behaviors, regardless of if being based on demographic data or on advert viewing time. However the reasoning behind the model would be unknown, and it could be dubious to use it in the industry this way.

Furthermore, in "Section 7.3. Implications for the TV advert industry", it is stated that: "there was no link observed between the time of advert exposure and the purchase decisions". But, is this correct for Category 1, which is the large majority among the categories? Because the model appears to be predicting quite well ( $F1=0.85$ ) for this category.

There is no link found for positive purchase decisions, as we can agree from the other points in my discussion. Now, Category 1 describes negative purchase decisions, but as I explained before, the predictions for the negative purchase decisions were not proved to be linked to either the advert viewing time or the demographic factors based on the results of the t-tests for Hypotheses 1 and 3. Because of these results, It cannot be said that because of the number of adverts that a customer saw, that they decided to continue to not buy a product. It also cannot be said that their demographic data is the main factor in predicting

this behavior.

Some typos encountered in the text: [...]

I have corrected the typos. Thank you for pointing them out.

Thank you for your review, I have made corrections to the manuscript according to your comments, and I hope they are satisfactory.