Lumen / PwC Attention Methodology & Case Study Review

March 2023

pwc

Contents

Introduction

Background

The diversification and increase in media content available to consumers provides ever-increasing options across platforms, communities, content sites and apps, and media providers. Much of this content is provided under an ad-funded model where multiple advertising messages compete to engage an audience with finite amount time available. The effectiveness of these ads is an area of ongoing investigation, under constant review as the ad market, media options, consumer behaviours and the marketing mix evolve. The last few years have seen increased industry focus on 'attention'.

Lumen believe that '*campaign measurement requires more than impressions and viewability, and that attention instead can be a more powerful metric when it comes to measuring campaigns' success and predicting changes in brand choice*'.

In early 2022, Lumen conducted campaign analysis for Advertisers 1, 2, and 3, with the aim of assessing the correlations of viewability and attention with click-through-rates (CTR) and conversion for ~1 billion impressions. Lumen also conducted an attention pilot study on behalf of Advertiser 1, in collaboration with On Device Research (ODR), which combined insights from 2400+ panelists across 6 studies with the aim of assessing the impact of attention on 12 brand metrics e.g. 1st ad awareness, 2nd ad awareness, recall, etc.

Lumen requested PwC ('we', 'us', 'our') to conduct an independent review of their campaign analysis.

PwC review of Lumen's campaign analysis concluded that:

- 70% of the time, the features identified by Lumen correctly predicted if an impression would be *viewed* (with "view" being measured as detecting eye gaze on the ad for at least 100ms) and the subsequent *dwell time (mean error of 1.2 sec)*, which are then used to feed Lumen's attention calculation (attention = exp_time x exp_view)
- From the subset of data provided by Lumen that we used to re-perform their campaign analysis, attention correlated better than viewability with both click-through-rate and conversion (demonstrating correlation, although not necessarily causation)
- Results from Lumen's brand lift pilot for Advertiser 1, where more information about users was available, showed that attention does not have a significantly better correlation than viewability on 1st ad mention and ad recall

Introduction

PwC scope

- A review and assessment of the attention prediction methodology and its appropriate application in the scripts used by Lumen to conduct campaign analysis for Advertisers 1, 2, and 3, and the data outputs created

- A review of the outputs from the campaign analysis conducted on behalf of Advertisers 1, 2, and 3, designed to assess the correlation between attention and click-through-rate, and conversion, and how these compare to viewability

- An assessment of the pilot study conducted in collaboration with ODR for Advertiser 1 designed to link attention predictions to surveys administered after exposure and assess attention impact in relation to 12 brand metrics e.g. ad recall, 1st ad awareness, 2nd ad awareness, etc.

PwC responsibilities and limitations of our review

- PwC exercised reasonable professional care and diligence in the collection, processing, and reporting of this information. However, the data used is from third party sources and PwC has not independently verified, validated, or audited the data. PwC makes no representations or warranties with respect to the accuracy of the information, nor whether it is suitable for the purposes to which it is put by users.

- PwC shall not be liable to any user of this report or to any other person or entity for any inaccuracy of the publicly obtained information from the market set or any errors or omissions in its content, regardless of the cause of such inaccuracy, error or omission. Furthermore, in no event shall PricewaterhouseCoopers LLP be liable for indirect, consequential, incidental or punitive losses or damages to any person or entity for any matter relating to this information.

- All information presented on the named companies was publicly available at the time the information was collected (June to Oct 2022). PwC will not disclose non-public individual entity data.

Section 1

Review attention prediction methodology and models, and assess appropriate application

Work Performed

PwC conducted a detailed walkthrough with Lumen on their attention prediction methodology and its implementation across both global and domain models. The session was followed by weekly technical catch ups to help answer specific questions regarding the implementation of Lumen's methodology in the scripts shared with us.

We reviewed Lumen's general methodology for data modelling, statistical approaches deployed, models' predictive power, limitations and mitigation.

We reviewed and tested Lumen's attention model scripts using a sample of data and assessed key assumptions to identify any potential weaknesses and understand the impact of attention referenced in the results shared by Lumen.

Key Findings

PwC found Lumen's methodology to be logical in calculating viewability and dwell time (used as inputs for their attention calculation (attentive seconds)).

Some points for Lumen's consideration:

- We observed some multicollinearity in the model, which does not impact its predictive power but does affect our understanding of the relative importance of individual features e.g. ad size
- We observed that the data we used for the analysis was unbalanced and skewed. 88% of the data was comprised of banners while 91% of sessions had between 0 and 1 dwell time (seconds) spent. However, skewness is expected as it is reflects the availability of ad formats and inventory.
- On average, 70% of the time, the features identified by Lumen correctly predicted if an impression would be *viewed* or not (with "view" being measured as detect eye gaze on the ad for at least 100ms). We noted that the model performance diverged across ad types due to the observed data imbalance. Where more data was available (e.g.) banners and outstream, model performance was significantly better compared to skins and instream.

Review attention outputs and attention's correlation with CTR and conversion

Work Performed

PwC reviewed data sets and calculations used in Lumen's scripts and referenced in Lumen's campaign analysis report to verify the correlation between attention and click-through-rate and conversation, and how this compares to viewability for Advertisers 1, 2, and 3.

In particular, we:

- Reviewed the completeness and accuracy of data inputs, consistent application of Lumen's methodology to the scripts, and fair and balanced reporting of the results.
- Re-performed the campaign analysis aimed at assessing the the relationship between attention and desired outcomes (e.g. for each specific advertiser, in addition to how the metrics differ by device where available) and assessed model's performance across different campaigns.

Key Findings

Our findings were consistent with Lumen's campaign analysis report.

We observed some minor differences, which we hypothesize could be attributed to the data samples we tested not being quite the same as the data samples used to generate Lumen's campaign analysis report:

- Despite using a data sample encapsulating a different time window to the one used in the campaign analysis, most tested plots followed the same trend as the report i.e. attention did correlate better with both click-through-rate and conversion compared to viewability
- Impression count was consistent with the figures referenced in Lumen's analysis for both Advertiser 1 and Advertiser 3 banners. There were differences for Advertiser 3 video and Advertiser 2.

Review and rerun models used for the attention pilot study for Advertiser 1

Work Performed

Reviewed and re-performed the attention pilot study conducted by Lumen in collaboration with On Device Research (ODR) on behalf of Advertiser 1.

In particular, PwC:

- Reviewed the completeness and accuracy of data inputs e.g. impressions and survey panelists data, application of Lumen's methodology to the scripts, and fair and balanced reporting of the results.
- Re-performed the analysis and assessed the model's performance in relation to twelve brand metrics (e.g. 1st ad awareness, recall, etc.).

Key Findings

Our findings were consistent with Lumen's brand lift study.

Some considerations for Lumen for future reference:

- McFadden's R squared used to measure the model performance varied based on metrics and tended to be on the lower side of the spectrum
- Attention had a higher z-value compared to view on 1st mention awareness, any mention awareness, and digital ad recall
- Attention was not significantly better than view when we looked at the p-value for 1st mention awareness, any mention awareness, and digital ad recall. We did not observe statistically significant p-values for attention or view for other ad recall, short-term value, short-term quality, short-term range, short-term price, long-term value, consideration: any, consideration: T2B.

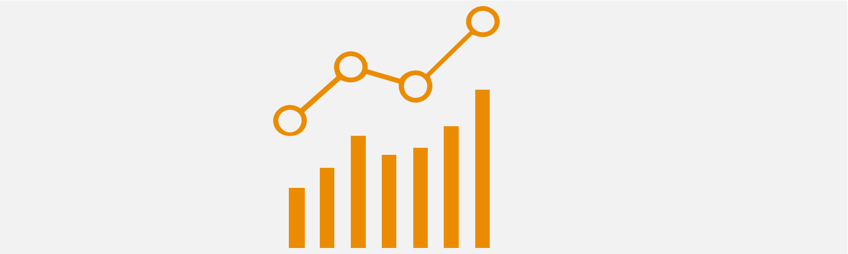
Our key findings and observations

1

Attention Model

Background

Dataset

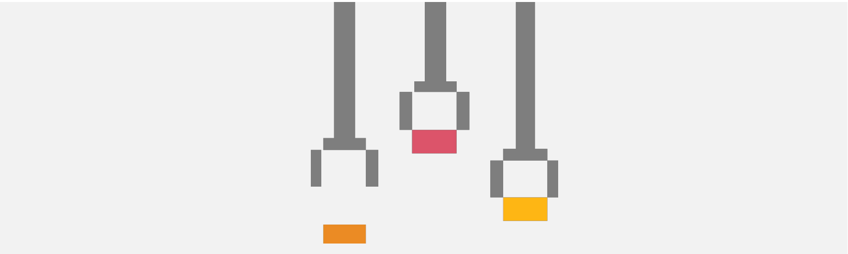


For the purposes of testing Lumen's attention models (global and domain*), we used a subset of campaign-related data. The sample was comprised of the following ad types:

- Banners: 35,000 impressions (accounts for 7% of available Lumen banners data). The breakdown of banners by device was as follows: Desktop: 25,000 vs. Mobile: 10,000 impressions
- Outstreams: 5,000 impressions (accounts for 35% of available Lumen's outstream data). In terms of device, all impressions came from mobile.
- It should be noted that as part of our analysis, we did not test any data from skins and instream videos.

**Note: Global model is universally applied to all domains if not specified. x*

Our approach



- Reviewed Lumen's attention methodology for data modelling approach, statistical approaches deployed, limitations and mitigation, and future plans.
- Reviewed and tested Lumen's attention model scripts using a sample of data and assessed key assumptions to identify any potential weaknesses and understand the impact of attention referenced in the results shared by Lumen.
- Reviewed performance of global and domain models by analysing the metrics/ plots provided by Lumen, which were derived from their latest model run.

Findings

- The **methodology for calculating attention and viewability was logical,** the data inputs used were appropriate and aligned to the methodology. It should be noted that as part of our review we were unable to re-perform or verify the data cleaning and aggregation procedures used by Lumen - the data sample provided by Lumen was already cleaned and suitable for use by the model.
- A data sample (comprised of banner and video ad types) was used to re-perform the analysis which meant that we were not able to recreate the exact results referenced in the presentation. The model validation and corresponding plots were provided by Lumen.
- Throughout our analysis, we noted that **the data sample which was provided by Lumen was unbalanced and skewed** i.e. banners accounted for 88% of available ad types included in the data. This imbalance was evident not just in the data sample but in Lumen's complete dataset. Data skew was observed across the dwell time (seconds) spend per season in the data sample - most of the session (91%) had a duration of 0 to 1 dwell time (seconds). It should be noted, however, that a degree of skewness is expected as it is reflective of the availability of ad formats and inventory in the industry.
- On average, **70% of the time, the features identified by Lumen correctly predicted if an impression** was going to be *viewed or not* (with "view" being measured as detect eye gaze on the ad for at least 100ms). For non-viewable impressions, accounting for majority of available impressions in the sample (70% of data), the model performed really well and was able to identify correctly 83% of them. It should be noted, however, that in cases of unbalanced datasets, the machine learning classifier tends to be more biased towards the majority class, and thus can hamper the model's performance in identifying the minority class (viewable impressions). This was evident in the **model's performance for viewable impressions where 68% of impressions were correctly classified**. Diverging model performance as a result of ad type imbalance was also observed across some ad types/campaigns. Where more data was available i.e. banners and outstream, model performance was significantly better compared to skins and instream.
- Yahoo and Tiktok were two data sources which were handled by Lumen differently, and thus were excluded from our analysis.
- The **features used in the model corresponded to the features referenced in the report** i.e. size, device, domain, channel, viewable_s, % in view were included. It should be noted, however, that ad clutter and ad position, often considered to have an impact on viewer's attention, were not included in the model.

Findings (continued)

- **The selection of variables used in the models introduced multi-collinearity** (high intercorrelation of independent variables) which could produce less reliable statistical inferences e.g. viewed ~ size + size:log(viewable_s) - 1. In this instance, **size and the interaction between viewable seconds and size are collinear.** Collinearity is a statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy. In this instance, the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data. While collinearity does not reduce the predictive power or reliability of Lumen's view model as a whole, it does affect calculations regarding individual predictors. That is, a multiple regression model with correlated predictors can indicate how well the entire bundle of predictors predicts the outcome variable, but it may not give valid results about any individual predictor, or about which predictors are redundant with respect to others (Agresti, A. 2018).

Model performance

It should be noted that the view and dwell time models' validation was performed by Lumen and corresponding plots were shared with PwC. PwC was not able to recreate these due to using a subset of the data to test the appropriate application of the methodology in the script.
Although there's no commonly accepted agreement on how to assess the fit of a logistic regression, we used the following metrics from Lumen's report to assess the models' performance.

- The goodness of fit of the logistic regression model can be expressed by some variants of pseudo R squared statistics, most of which are being based on the deviance of the model. Lumen used McFadden's pseudo r-square (ρ2) which was designed to measure the relative performance of the view model, compared to a view model that always predicted the mean (null model). **The Mcfadden value for Lumen's global (0.26) and domain (0.29) view models ranged between 0.26-0.29 which indicated a very good fit.** In practice, for a valid model, "a goodness-of-fit using McFadden's pseudo r-square (ρ2) between 0.2 and 0.4 should be taken to represent a very good fit of the model (Louviere et al., 2000)".
- Accuracy is another metric designed to measure the percentage of correct predictions over all predictions identified by the model. **Lumen's view model accuracy for the 50% threshold view model was 81%**, and 78% for the 30% threshold view model.
- Sensitivity measures the proportion of true positives that are correctly identified by the model e.g. ability to correctly predict non-viewable impressions. **The sensitivity score is 91% for view model (50% threshold) and 82% for view model (30% threshold).**
- Specificity measures the proportion of true negatives that are correctly identified by the model e.g. a viewed impression is predicted as viewed. Lumen's model **specificity stands at 56% for the 50% threshold view model and 68% for the 30% threshold view model.**

Findings (continued)

Model performance (continued)

A confusion matrix was used to summarise Lumen's view models (50% and 30% threshold) performance on true positives, true negatives, false positives, and false negatives.

A global model was universally applied to all domains if not specified. A domain model was tuned for a particular domain, and only applied if the performance was better than the global model and there was sufficiently large amount of data for the domain.

Global view model (30% threshold)

- True positives: 68%. The model was able to predict correctly 68% of all viewable impressions.
- True negatives: 82%. The model was able to correctly predict univiewed impressions as unviewed.
- False positives:18%. The model misclassified 17% of unviewed impressions as viewed.
- False negatives: 32%. The model misclassified 32% of viewed impressions as unviewed.

Domain view model (30% threshold)

- True positives: 71%. The model was able to predict correctly 71% of all viewable impressions.
- True negatives: 82%. The model was able to correctly predict 82% of univiewed impressions as unviewed.
- False positives:18%. The model misclassified 18% of unviewed impressions as viewed.
- False negatives: 29%. The model misclassified 29% of viewed impressions as unviewed.

From the results above, we concluded that the **domain's model performance was slightly better than the global model** with regards to correctly identifying viewable impressions. **Performance with regards to correctly identifying unviewed impressions remained equally high across both models.**

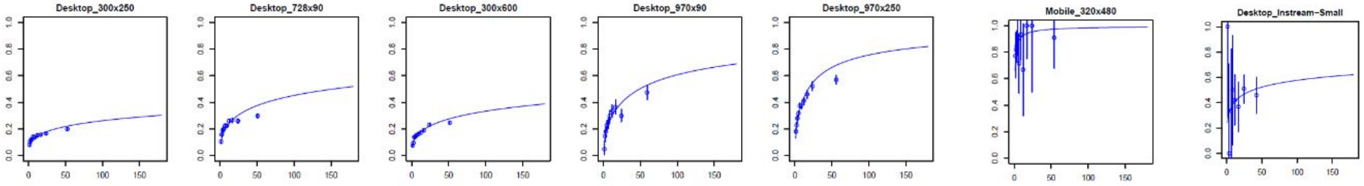
Findings (continued)

Plots

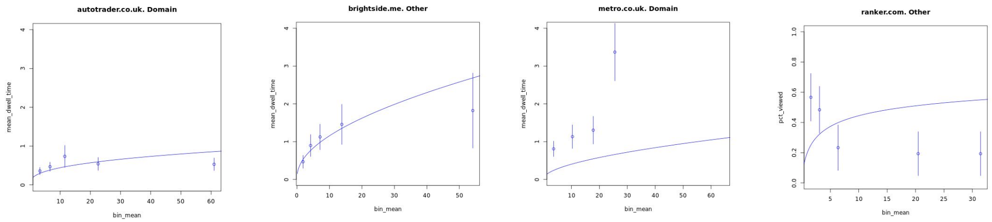
The plots constitute a visual examination of the relationship between viewable seconds, percentage view, and attention. Attention can be used as a metric to explain viewable seconds.

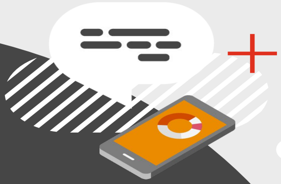
The plots have different fit on the measurement, we hypothesize this could be attributed to the lack of availability of data regarding particular campaigns, domains, and/or channels.

- Global:
  Good fit for banners and videos, for skins and instream the model fit was not as good. For some sizes with less data, the fit was also not as good.



- Domain
  90% of domain models had good fit. For the remainder 10%, we hypothesize that the poor fit was attributed to the lack of availability of data for particular domains.





2

Reporting and Visualisation

Background

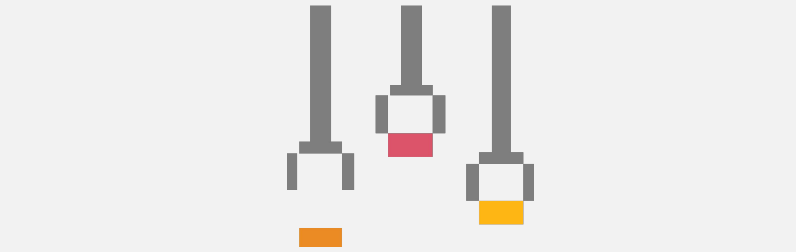
Dataset

For the purposes of reviewing/reperforming the campaign analysis conducted on behalf of Advertiser 1, 2, and 3, we used a sample of impression-level data for each advertiser.

The sample contained information about individual impressions, attention prediction associated with each impressions, session-related information, and ad information (campaign, line item, size). The breakdown of impressions by advertiser was the following:

- Advertiser 1: 197m
- Advertiser 2: 676m
- Advertiser 3: 119m

Our approach



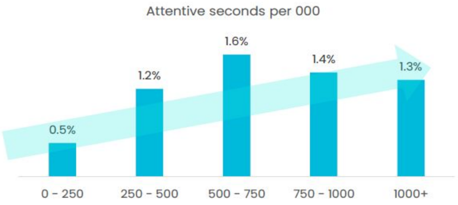
- Reviewed Lumen's reporting and visualization methodology
- Reviewed and tested Lumen's reporting and visualization script with the data samples provided and compared the resulting metrics and plots to the report content, including:
    - Total impression numbers
    - Attention vs. CTR (Advertiser 1, Advertisers 2, and 3)
Viewability vs. CTR (Advertiser 1, Advertisers, 2, and 3)
Split by Desktop and Mobile (Advertiser 3)
Split by Banner and Video (Advertiser 3)
    - Attention vs. Conversion rate (Advertisers 1, 2, and 3)
Viewability vs. Conversion rate (Advertisers 1, 2, and 3)
Split by Desktop and Mobile (Advertiser 3)

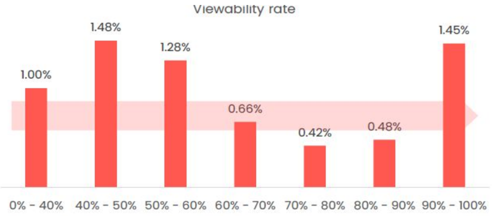
Findings

Result

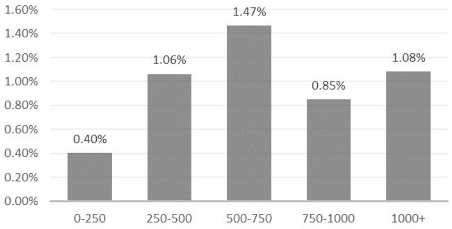
- It should be noted that the data samples used for testing by PwC were not the same as the data samples used to generate Lumen's campaign analysis report, we hypothesize that was the reason we observed inconsistencies in impression % across some of the plots.
- Impression count was consistent for Advertiser 1 and Advertiser 3 banners. There were differences for Advertiser 3 video (received 77m vs. report 108m) and Advertiser 2 (received 676m vs. report 658m)
- Most tested plots followed the same trend as the report, i.e. attention did correlate better with both click-through-rate and conversion compared to viewability, but the numbers didn't match exactly to those in the report.

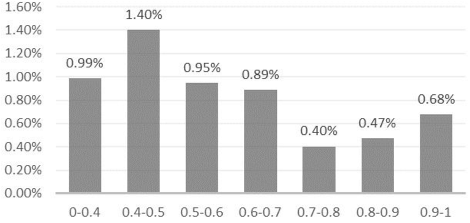
Lumen Report





PwC Test





Source: Advertiser 3 Video Mobile CTR Sample

- For Advertiser 1, the conversion rate tested and reported were very different due to a change implemented to remove impressions without potential conversion in the calculation of conversions.

3

Advertiser 1
Brand Lift Study

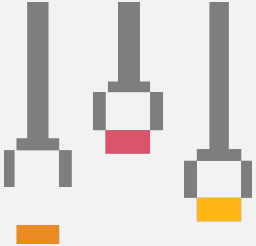
Background

Dataset

For the purposes of testing Lumen's brand lift study with ODR, we used the complete sample of survey dataset and the impression-level attention dataset. The datasets contained the following information:

- Impression-level attention data
  - Users: 3,051 (raw) -> 2,402 (cleaned and matched)
  - Impressions: 12,827 (raw) -> 7,259 (cleaned and matched)
- Survey data with questions and answers related to demographic and brand awareness information
  - Participants: 5,504 control vs. 3,169 exposed

Our approach



- Reviewed Lumen's brand lift methodology for data modelling, statistical approaches deployed, limitations and mitigation, and plans for future improvement.
- Reviewed and tested Lumen's brand lift script with data samples provided and compared the resulting metrics and plots to the report content.

Findings

Result

- Our findings were consistent with Lumen's brand lift results which showcased that both attention and viewability have significant relationship with 1st mention awareness, any mention awareness, and digital ad recall.

Methodology

- The methodology used to assess the impact of attention and viewability (alongside a number of other customer features) on 12 brand metrics was logical, the data inputs used are appropriate and aligned to the methodology.

- We did observe, however, a slightly unconventional approach to calculating lift - it was calculated as the difference between an estimated metric value e.g. 1st Mention Awareness for the exposed group (people who have viewed an ad) and average of metric value for the control group (people who have not viewed an ad).

  Uplift=estimated value for exposed group - average value for control group

  In traditional literature on advertising effectiveness, lift is generally estimated as the delta of the estimated values of a metric of choice for the exposed group divided by the estimated value for a metric of choice for the control group if they had not been treated e.g, delta of 1st mention awareness values for the exposed group (people who have viewed/paid attention to an ad) vs 1st mention awareness for the control group (people who have not viewed/paid attention to an ad)

$$\tau_\ell = \frac{\Delta \text{Conversion rate due to ads in the treated group}}{\text{Conversion rate of the treated group if they had } \textit{not} \text{ been treated}} = \frac{\tau}{\mathbb{E}[Y^{obs} | Z = 1, W^{obs} = 1] - \tau}$$

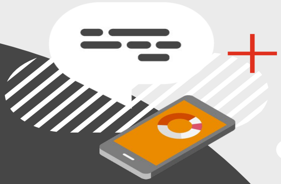
- McFadden's R squared used to measure the model performance was on the lower end, ranging from 0.02 to 0.25. In practice for a valid model, “a goodness-of-fit using McFadden's R squared between 0.2 and 0.4 should be taken to represent a very good fit of the model (Louviere et al., 2000)."*

Findings

- We also looked at both z-values and p-values as part of our assessment of the performance of the logistic models used by Lumen.
  - Z-value (standard score: describes how many standard-deviations away a metric's value is from the mean).
    Attention had a higher z-value compared to view on 1st mention awareness, any mention awareness, and digital ad recall.
  - P-value (probability value: describes how likely it is that observations will have occurred by random chance).
    Attention was not significantly better than view when we looked at the p-value for 1st mention awareness, any mention awareness, and digital ad recall. We did not observe statistically significant p-values from attention or view for the remainder of brand metrics i.e. ad recall, short-term value, short-term quality, short-term range, short-term price, long-term value, consideration: any, consideration: T2B

Note: McFadden's R squared measures relative performance, compared to a model that always predicts the mean. Binned residual plots allowed us to check whether the residuals had a pattern and whether particular residuals were larger than expected, both indicating poor model fit.

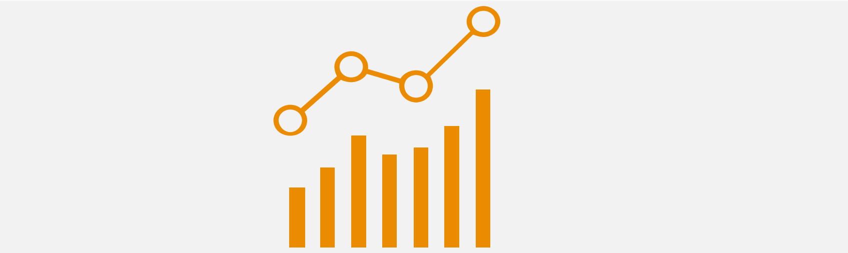
Appendix



1 Attention Model

Background

Dataset

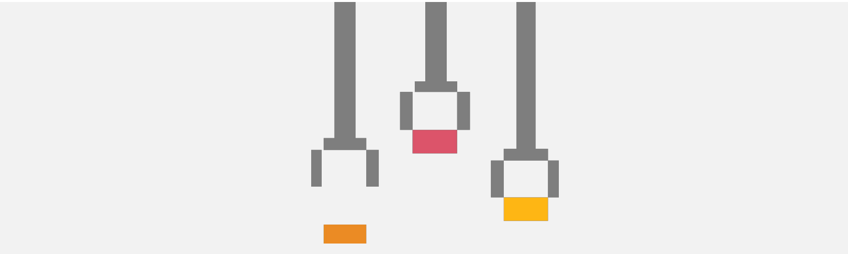


For the purposes of testing Lumen's attention models (global and domain*), we used a subset of campaigns data. The sample was comprised of the following ad types:

- Banners: 35,000 impressions (accounts for 7% of available Lumen banners data). The breakdown of banners by device was as follows: Desktop: 25,000 vs. Mobile: 10,000 impressions
- Outstreams: 5,000 impressions (accounts for 35% of available Lumen's outstream data). In terms of device, all impressions came from mobile.
- It should be noted that as part of our analysis, we did not test any data from skins and instream videos.

**Note: Global model is universally applied to all domains if not specified. Domain model is tuned for a particular domain, and only applied if the performance is better than the global model and there is sufficiently large amount of data for the domain.*

Our approach



- Reviewed Lumen's attention methodology for data modelling approach, statistical approaches deployed, limitations and mitigation, and plans for future improvement.
- Reviewed and tested Lumen's attention model scripts using a sample of data and assessed its key assumptions to identify any potential weaknesses and understand the impact of attention referenced in the results shared by Lumen.
- Reviewed performance of global and domain models by analysing the metrics/ plots provided by Lumen which were derived from their latest model run.

Methodology

We reviewed Lumen’s methodology and assessed its application in the scripts.

The following components were reviewed as part of our analysis:

- Global models calculation for both viewed and dwell time models.
  - *viewed ~ size + size:log(viewable_s) - 1* (logistic regression): predicted possibility of the ad being viewed
  - *dwell_time_s ~ size + size:log(viewable_s) - 1* (quasi-poisson): predicted dwell time on the ad
- Domain models calculation for both viewed and dwell time models.
- We confirmed the application of domain specific intercept and slope which were introduced for sessions within the specific domain with better performance than the global model
  - *viewed ~ -1 + size + size:log(viewable_s) + view_intercepts + view_slopes:log(viewable_s)*
  - *dwell_time_s ~ size + size:log(viewable_s) + - 1 +dwell_intercepts + dwell_slopes:log(viewable_s)*
- Plots used to assess performance of global and domain model with parameters from different domains.
  - Global plots
  - Domain specific plots
- Prediction calculation used to derive:
  - exp_view from the view model prediction
  - exp_time from the dwell time prediction
- Function to export results
- Corresponding output columns: 'domain','size','viewable_duration_ms','exp_view','exp_time'.

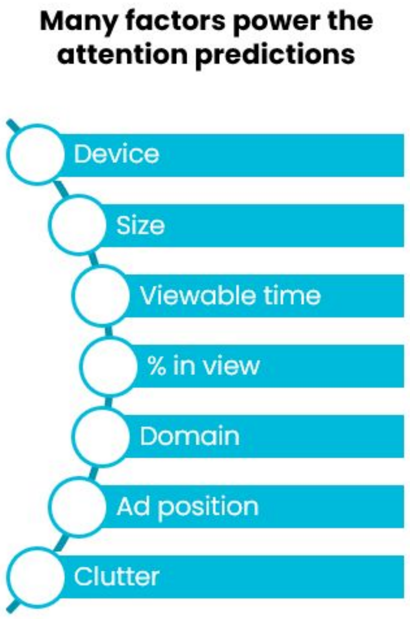
Note

- Factors included in the prediction of attention:

  It should be noted that view and dwell models used by Lumen used the following features: device, size, viewable time, domain, % in view (inferred by viewable time) but not ad position or clutter.

  Slide 6 of the campaign analysis produced by Lumen on behalf of Advertiser 1, 2, and 3 referred to the possible factors that could affect attention instead of the actual factors used by Lumen to predict attention.

- % in view was inferred by viewable time and market convention of 50%. Size of the ad was not taken into account in terms of % in view and viewable time.



Findings

- The **methodology for calculating attention and viewability was logical** although we did observe some multicollinearity in the model, the data inputs used were appropriate and aligned to the methodology. It should be noted that as part of our review we were unable to re-perform or verify the data cleaning and aggregation procedures used by Lumen - the data sample provided by Lumen was already cleaned and suitable for use by the model.
- A data sample (comprised of banner and video ad types) was used to re-perform the analysis which meant that we were not able to recreate the exact results referenced in the presentation. The model validation and corresponding plots were provided by Lumen.
- Throughout our analysis, we noted that **the data sample which was provided by Lumen was unbalanced and skewed** i.e. banners accounted for 87.5% of available ad types included in the data. This imbalance was evident not just in the data sample but in Lumen's complete dataset. Data skew was observed across the dwell time (seconds) spend per season in the data sample - most of the session (91%) had 0 to 1 dwell time (seconds) associated with them. It should be noted, however, that that a degree of skewness is expected as it is reflective of the availability of ad formats and inventory in the industry.
- On average, **70% of the time, the features identified by Lumen correctly predicted if an impression** was going to be *viewed or not* (with "view" being measured as detect eye gaze on the ad for at least 100ms). For non-viewable impressions, accounting for majority of available impressions in the sample (70% of data), the model performed really well and was able to identify correctly 83% of them. It should be noted, however, that in cases of unbalanced datasets, the machine learning classifier tends to be more biased towards the majority class, and thus can hamper the model's performance in identifying the minority class (viewable impressions). This was evident in the **model's performance for viewable impressions where 68% of impressions were correctly classified**. Diverging model performance as a result of ad type imbalance was also observed across different ad types/campaigns. Where more data was available i.e. banners and outstream, model performance was significantly better compared to skins and instream.
- Yahoo and Tiktok were two data sources which were handled by Lumen differently, and thus were excluded from our analysis.
- **The selection of variables used in the models introduced multi-collinearity** (high intercorrelation of independent variables) which could produce less reliable statistical inferences e.g. viewed ~ size + size:log(viewable_s) - 1. In this instance, size and the interaction between viewable seconds and size are collinear. Collinearity is a statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy. In this situation the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data. While collinearity does not reduce the predictive power or reliability of Lumen's view model as a whole, it does affect calculations regarding individual predictors. That is, a multiple regression model with correlated predictors can indicate how well the entire bundle of predictors predicts the outcome variable, but it may not give valid results about any individual predictor, or about which predictors are redundant with respect to others (Agresti, A. 2018).

Findings (continued)

Model performance (Metrics, Plots)

It should be noted that the view and dwell time models' validation was performed by Lumen and corresponding plots were shared with PwC. PwC was not able to recreate these due to using a subset of the data to test the appropriate application of the methodology in the script.

Although there's no commonly accepted agreement on how to assess the fit of a logistic regression, we used the following metrics from Lumen's report to assess the models' performance.

- The goodness of fit of the logistic regression model can be expressed by some variants of pseudo R squared statistics, most of which are being based on the deviance of the model. Lumen used McFadden's pseudo r-square (ρ2) which was designed to measure the relative performance of the view model, compared to a view model that always predicted the mean (null model). **The Mcfadden value for Lumen's global (0.26) and domain (0.29) view models ranged between 0.26-0.29 which indicated a very good fit.** In practice, for a valid model, "a goodness-of-fit using McFadden's pseudo r-square (ρ2) between 0.2 and 0.4 should be taken to represent a very good fit of the model (Louviere et al., 2000)".
- Accuracy is another metric designed to measure the percentage of correct predictions over all predictions identified by the model. **Lumen's view model accuracy for the 50% threshold view model was 81%**, and 78% for the 30% threshold view model.
- Sensitivity measures the proportion of true positives that are correctly identified by the model e.g. ability to correctly predict non-viewable impressions. **The sensitivity score is 91% for view model (50% threshold) and 82% for view model (30% threshold).**
- Specificity measures the proportion of true negatives that are correctly identified by the model e.g. a viewed impression is predicted as viewed.

  Lumen's model **specificity stands at 56% for the 50% threshold view model and 68% for the 30% threshold view model.**

Findings (continued)

Model performance (continued)

A confusion matrix was used to summarise Lumen's view models (50% and 30% threshold) performance on true positives, true negatives, false positives, and false negatives.

A Global model was universally applied to all domains if not specified. A Domain model was tuned for a particular domain, and only applied if the performance was better than the global model and there was sufficiently large amount of data for the domain.

Global view model (30% threshold)

- True positives: 68%. The model was able to predict correctly 68% of all viewable impressions.
- True negatives: 82%. The model was able to correctly predict univiewed impressions as unviewed.
- False positives:18%. The model misclassified 17% of unviewed impressions as viewed.
- False negatives: 32%. The model misclassified 32% of viewed impressions as unviewed.

Domain view model (30% threshold)

- True positives: 71%. The model was able to predict correctly 71% of all viewable impressions.
- True negatives: 82%. The model was able to correctly predict univiewed impressions as unviewed.
- False positives:18%. The model misclassified 18% of unviewed impressions as viewed.
- False negatives: 29%. The model misclassified 29% of viewed impressions as unviewed.

From the results above, we can conclude that the domain's model performance was slightly better than the global model with regards to correctly identifying viewable impressions. Performance with regards to correctly identifying unviewed impressions remained equally high across both models.

Findings (continued)

Performance-related metrics

View model

Model level	View model: McFadden R2	View model: Accuracy (50% threshold)	View model: Sensitivity (50% threshold)	View model: Specificity (50% threshold)	View model: Accuracy (30% threshold)	View model: Sensitivity (30% threshold)	View model: Specificity (30% threshold)
Global	0.2697	0.8103	0.9159	0.5618	0.7859	0.82887	0.6847
Domain	0.2903	0.8133	0.9169	0.5691	0.7851	0.8182	0.7070

Dwell time model

Model level	View time: Mean Absolute error	View time: Explained variance
Global	1.2543	0.4555
Domain	1.235	0.4717

Findings (continued)

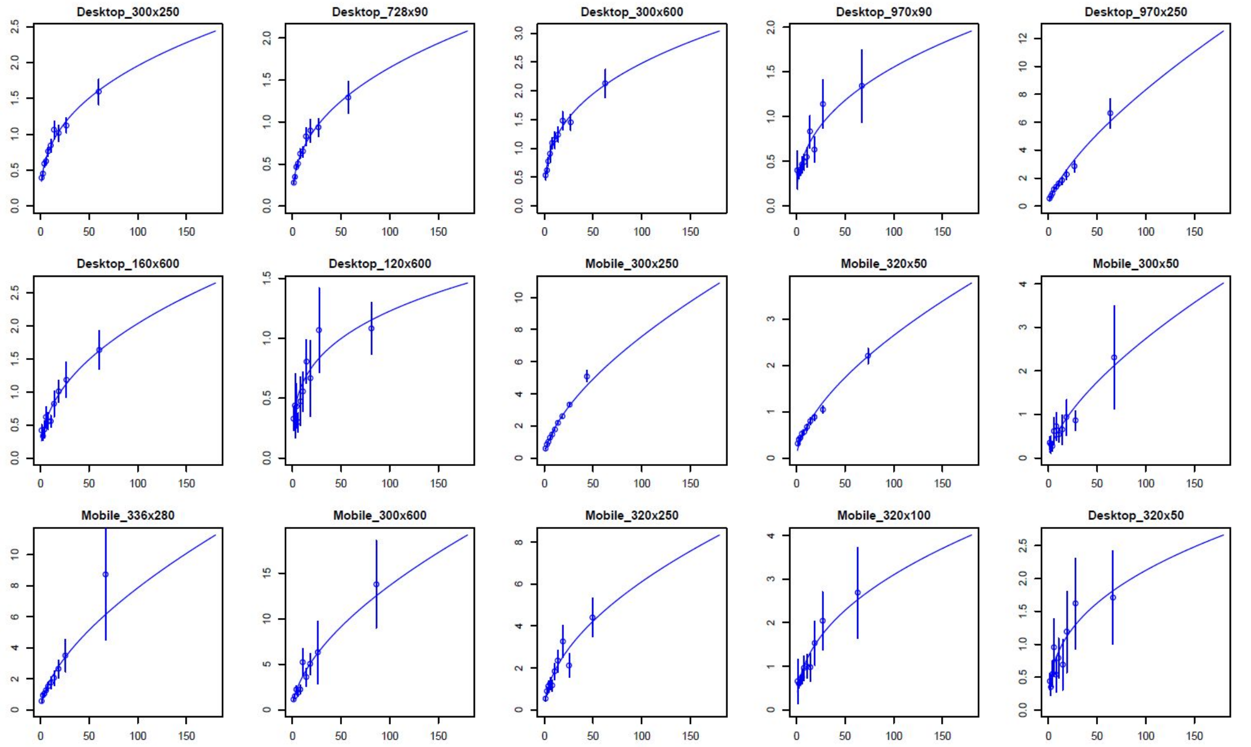
Performance-related metrics (continued):

Confusion Matrix

Model	Predicted	Actual	Frequency
Global (30% threshold)	0	0	40554
Global (30% threshold)	1	0	8373
Global (30% threshold)	0	1	6551
Global (30% threshold)	1	1	14223
Domain (30% threshold)	0	0	40033
Domain (30% threshold)	1	0	8894
Domain (30% threshold)	0	1	6087
Domain (30% threshold)	1	1	14687

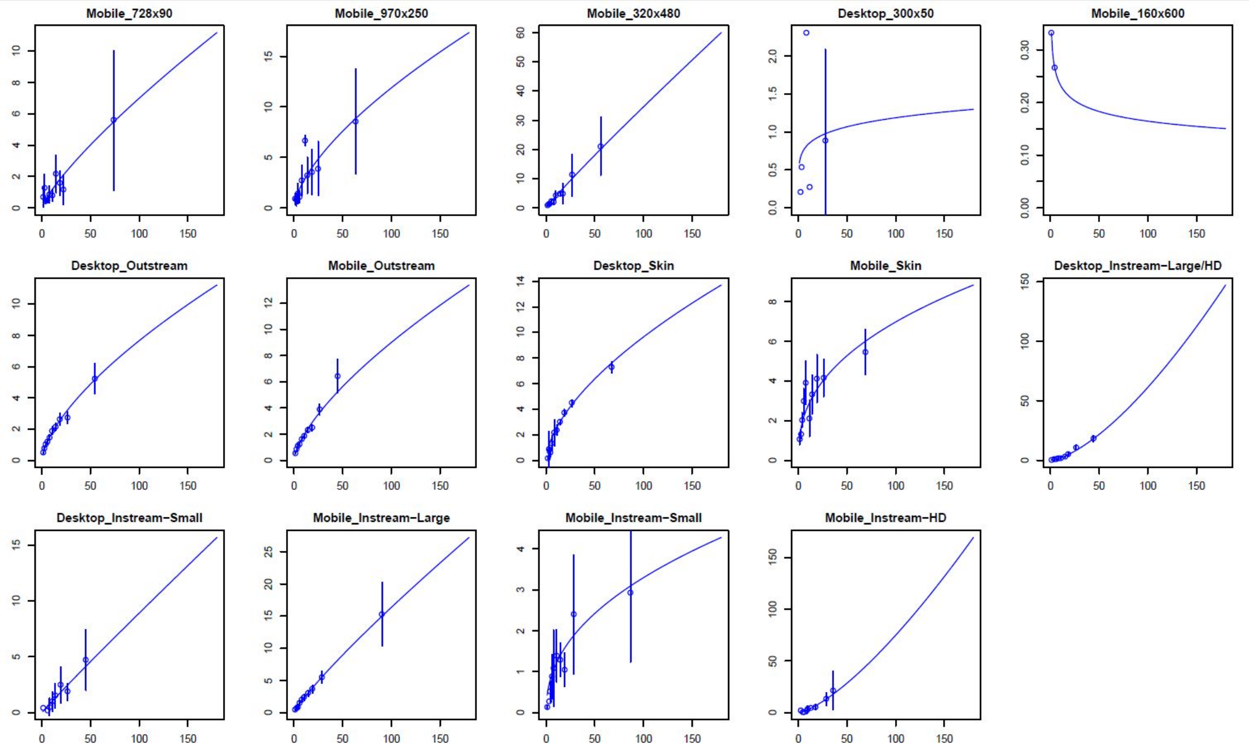
Findings (continued)

Global Plots (dwell time)



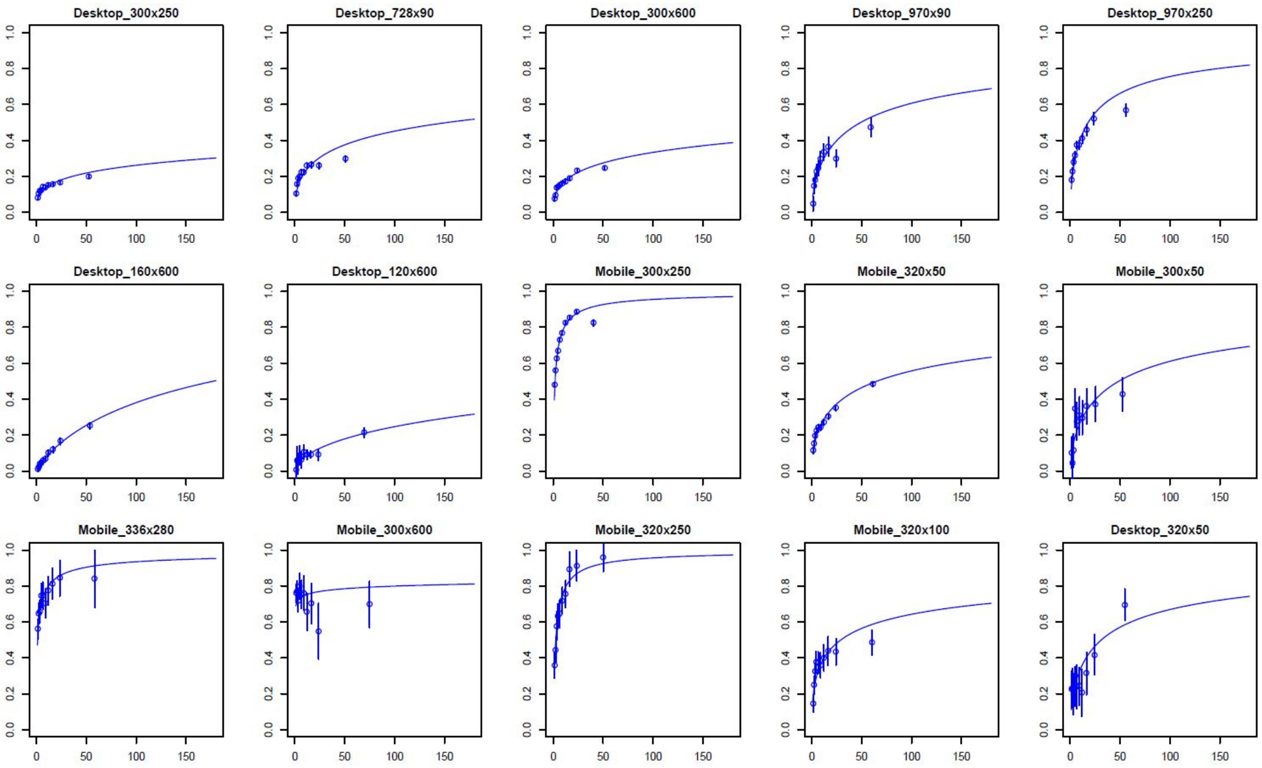
Findings (continued)

Global Plots (dwell time)



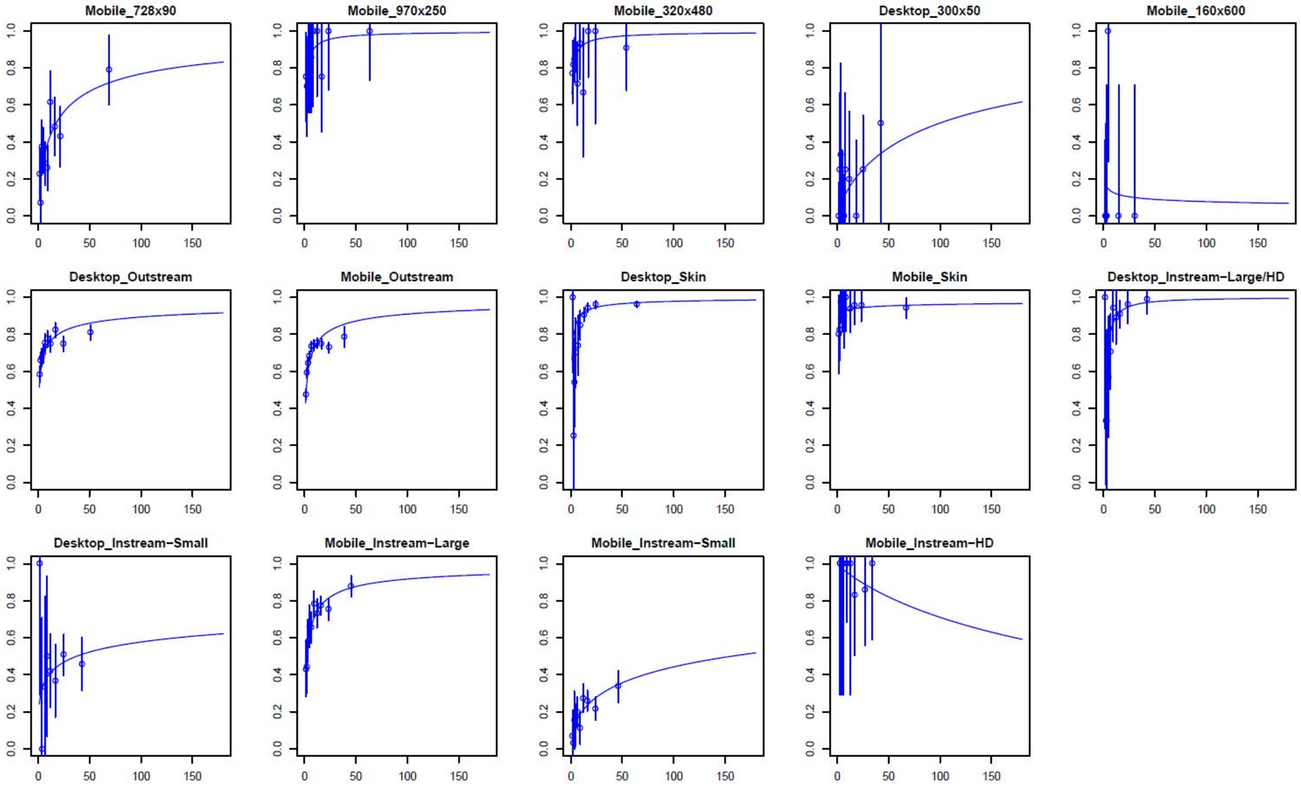
Findings (continued)

Global Plots (view)



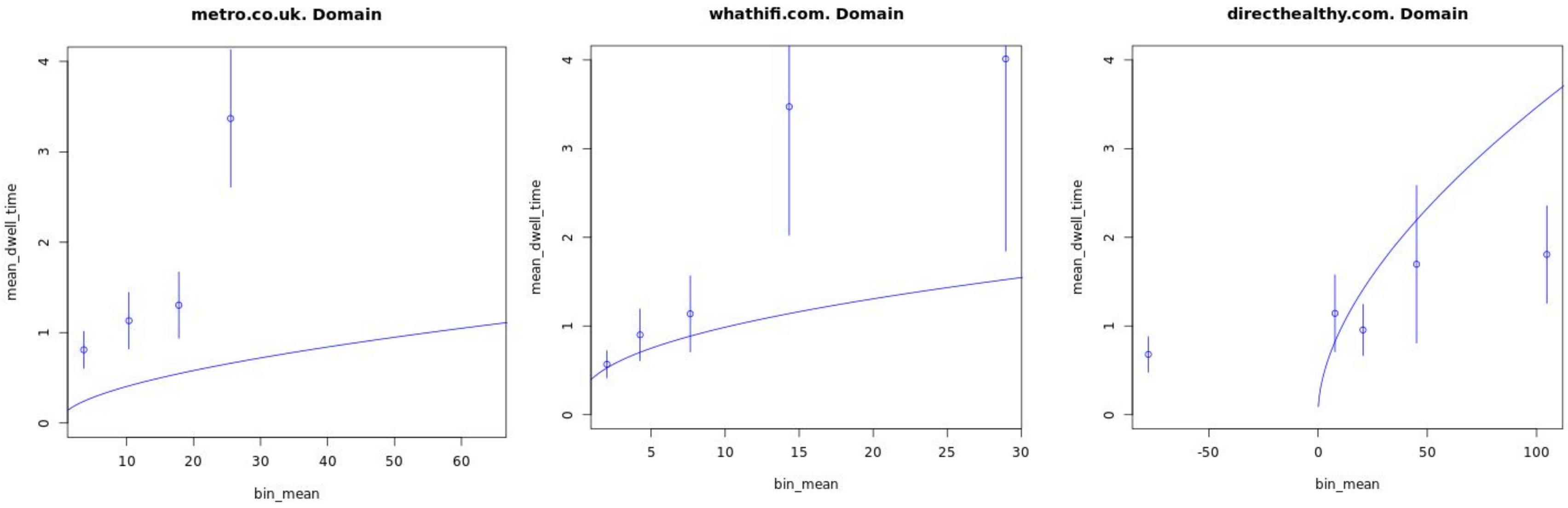
Findings (continued)

Global Plots (view)



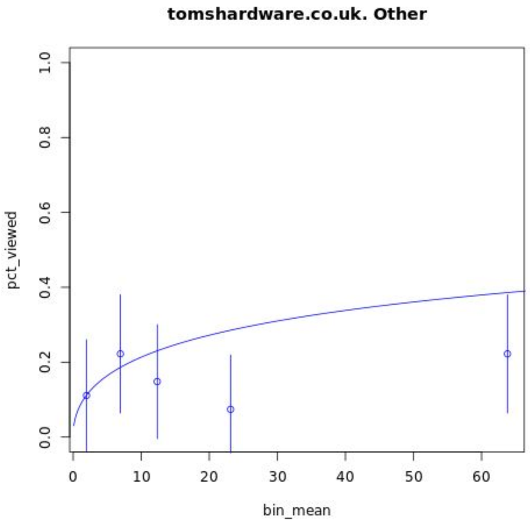
Findings (continued)

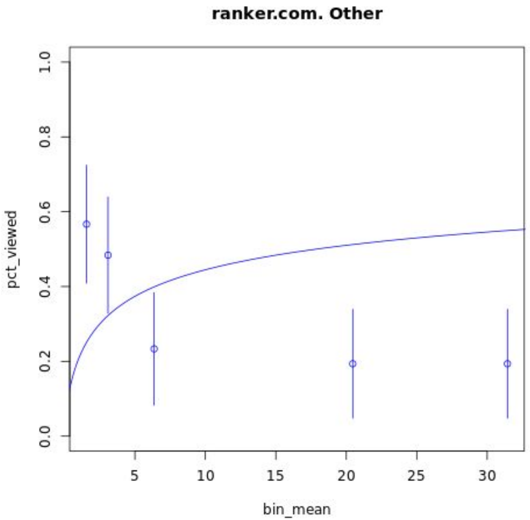
Subset selection of domain plots (dwell time)

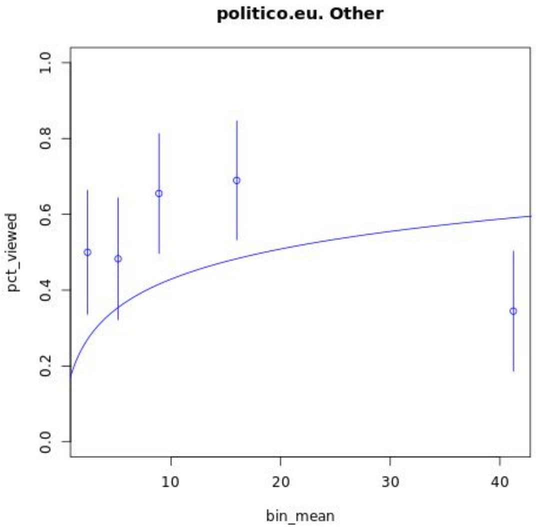


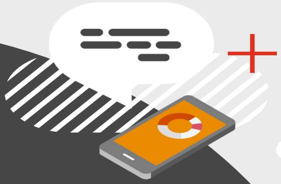
Findings (continued)

Subset selection of domain plots (viewed)









2

Reporting and Visualisation

Background

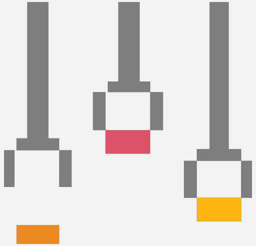
Dataset

For the purposes of reviewing/reperforming the campaign analysis conducted on behalf of Advertisers 1, 2, and 3, we used a sample of impression-level data for each advertiser.

The sample contained information about individual impressions, attention prediction associated with each impressions, session-related information, and ad information (campaign, line item, size). The breakdown of impressions by advertiser was the following:

- Advertiser 1: 197m
- Advertiser 2: 676m
- Advertiser 3: 119m

Our approach



- Reviewed Lumen's reporting and visualization methodology
- Reviewed and tested Lumen's reporting and visualization script with the data samples provided and compared the resulting metrics and plots to the report content, including:
    - Total impression numbers
    - Attention vs. CTR (Advertisers 1, 2, and 3)
Viewability vs. CTR (Advertisers 1, 2, and 3)
Split by Desktop and Mobile (Advertiser 3)
Split by Banner and Video (Advertiser 3)
    - Attention vs. Conversion rate (Advertisers 1, 2, and 3)
Viewability vs. Conversion rate (Advertisers 1, 2, and 3)
Split by Desktop and Mobile (Advertiser 3)

Methodology

As part of our assessment, we performed a review of the following components:

- **The process of data loading and cleaning**
- **Calculations used for bin attention & bin viewability**
  - APM=Attentive seconds/ Impressions *1000
  - Viewability comes directly from lamp avg viewability rate

- **Calculations used for CTR and conversion rate**
  - CTR=Clicks/Impressions
  - Conversion rate= Conversions/Impressions (v1)
  - Conversion rate= Conversions/Impressions with potential conversions (v2)
  - (Impressions with potential conversions= Impressions with line item conversions>0)

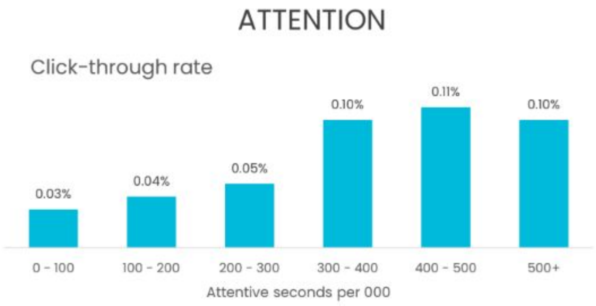
- **Plot bar charts comprised of the following:**
  - Attention vs. CTR (Advertisers 1, 2, and 3)
  - Viewability vs. CTR (Advertisers 1, 2, and 3)
  - Split by Desktop and Mobile (Advertiser 3)
  - Split by Banner and Video (Advertiser 3)
  - Attention vs. Conversion rate (Advertisers 1, 2, and 3)
  - Viewability vs. Conversion rate (Advertisers 1, 2, and 3)
  - Split by Desktop and Mobile (Advertiser 3)

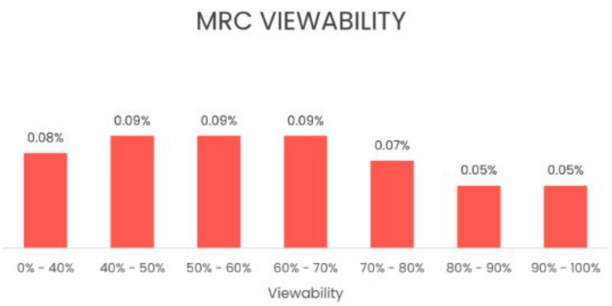
Findings

- It should be noted that the data samples used for testing by PwC were not the same as the data samples used to generate Lumen's campaign analysis report. We hypothesize that this was the reason we observed inconsistencies in impression % across some of the plots.
- Impression count was consistent for Advertiser 1 and Advertiser 3 banners. There were differences for Advertiser 3 video (received 77m vs. report 108m) and Advertiser 2 (received 676m vs. report 658m)
- Most tested plots followed the same trend as the report, i.e. attention did correlate better with both click-through-rate and conversion compared to viewability, but the numbers didn't match exactly to those in the report.
- For Advertiser 1, the conversion rate tested and reported were very different due to a change implemented by Lumen to remove impressions without potential conversion in the calculation of conversions.
- Plot comparison:
    - Advertiser 1
    - Advertiser 2
    - Advertiser 3

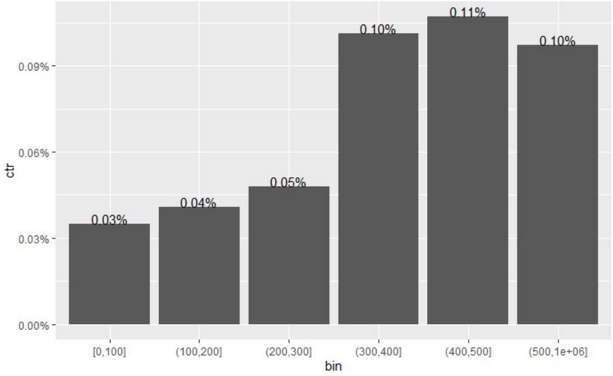
Reporting- Advertiser 1- CTR

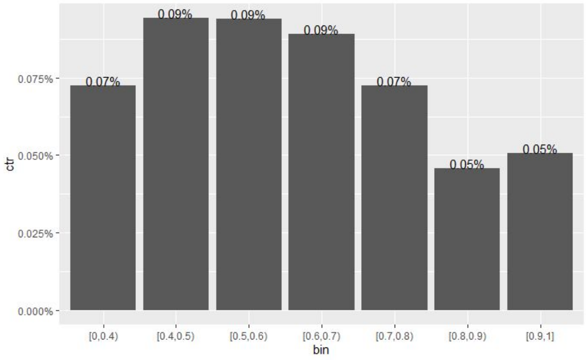
Source: Lumen Campaign Analysis Report





Source: PwC Results following re-performance



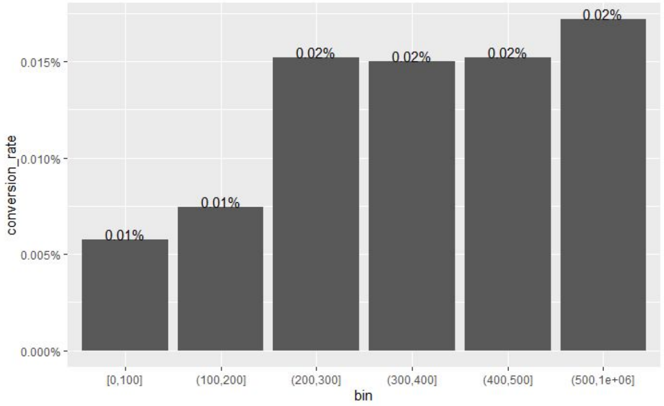


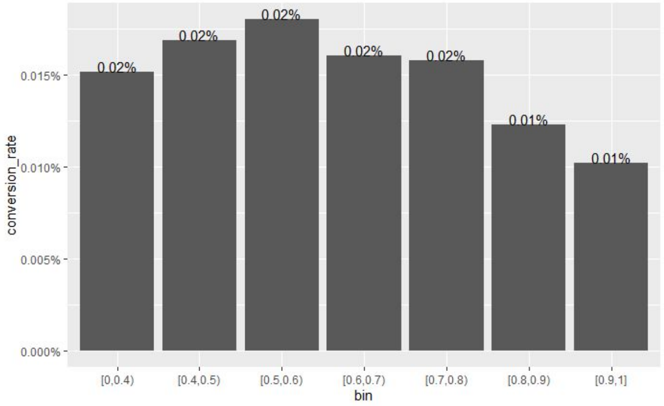
Reporting- Advertiser 1- Conversion

Source: Lumen Campaign Analysis Report

Conversion rate

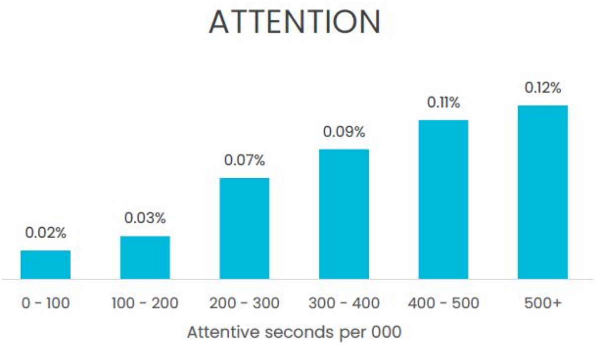
Source: PwC Results following re-performance

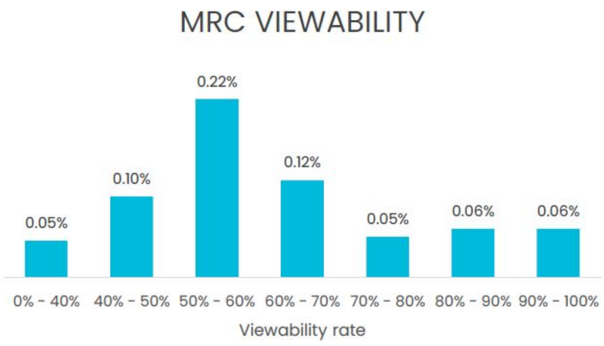




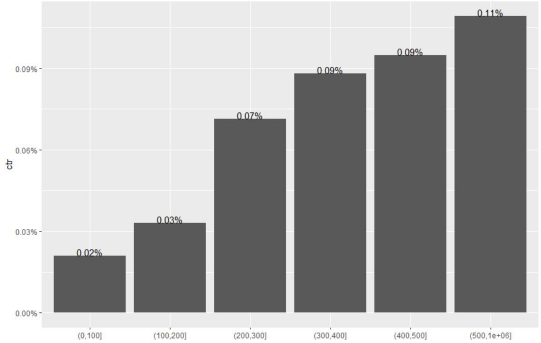
Reporting- Advertiser 2- CTR

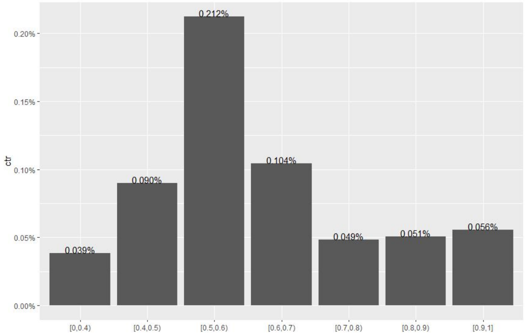
Source: Lumen Campaign Analysis Report





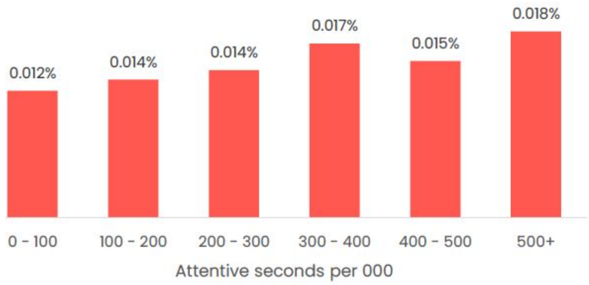
Source: PwC Results following re-performance

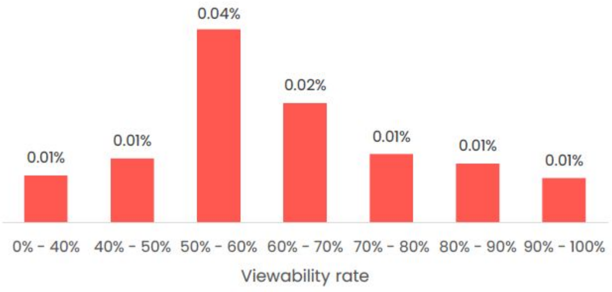




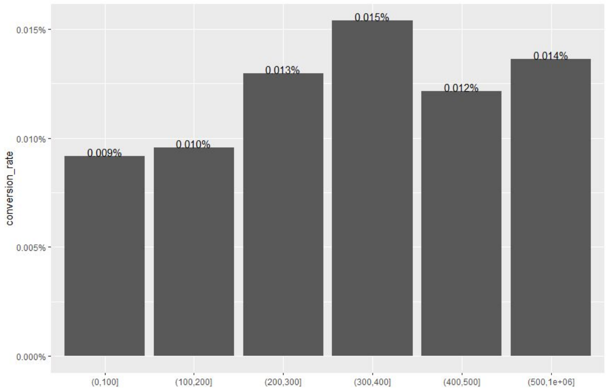
Reporting- Advertiser 2- Conversion

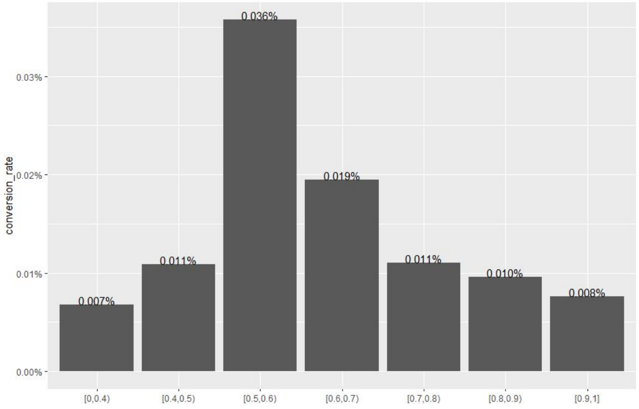
Source: Lumen Campaign Analysis Report





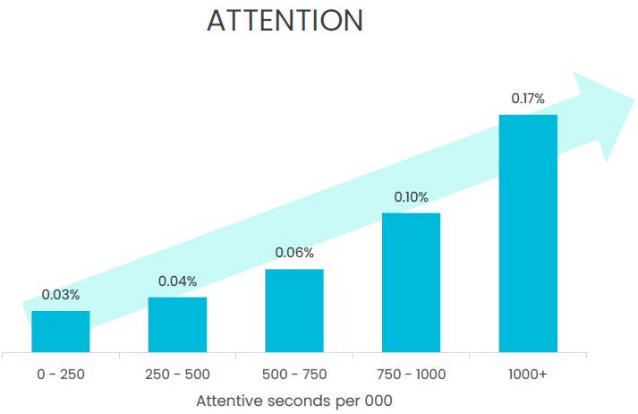
Source: PwC Results following re-performance

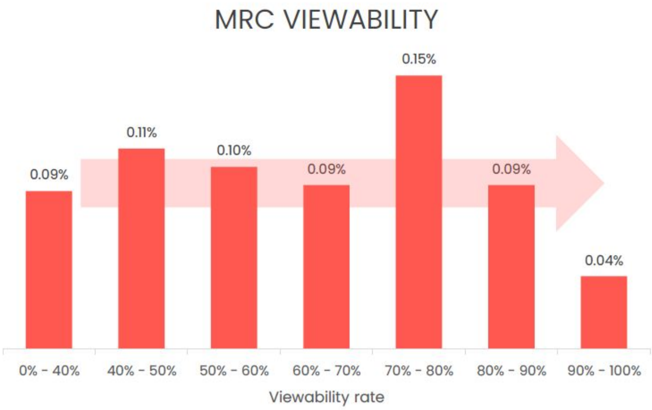




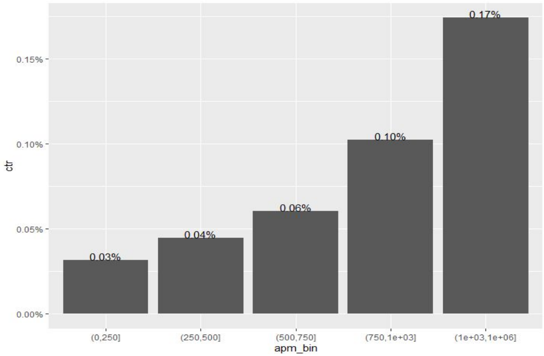
Reporting- Advertiser 3- Banners CTR

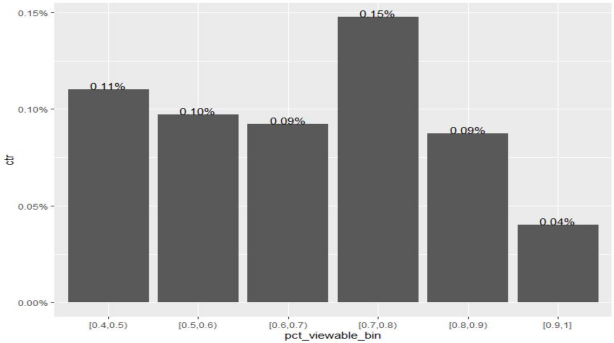
Source: Lumen Campaign Analysis Report



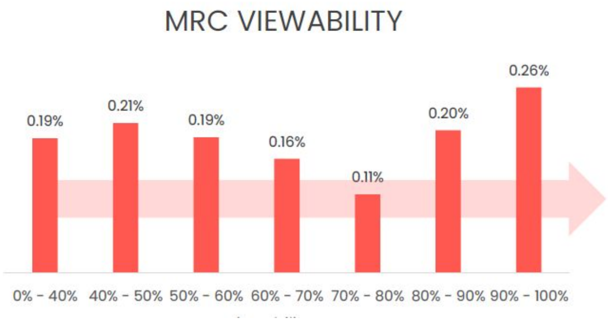


Source: PwC Results following re-performance

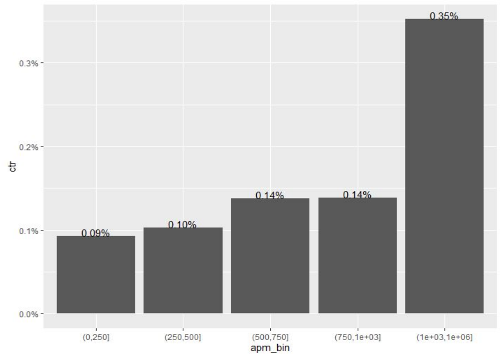


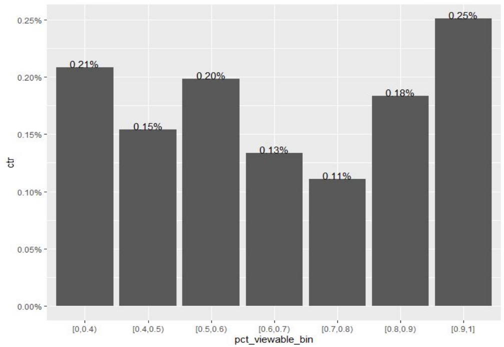


Reporting- Advertiser 3- Videos Desktop CTR



Source: Lumen Campaign Analysis Report

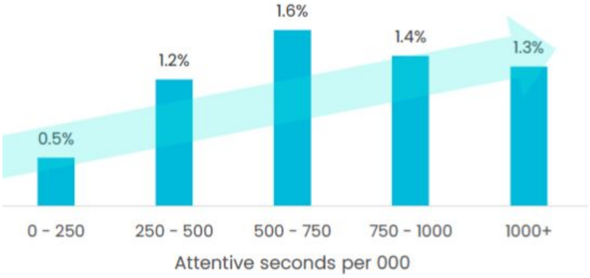


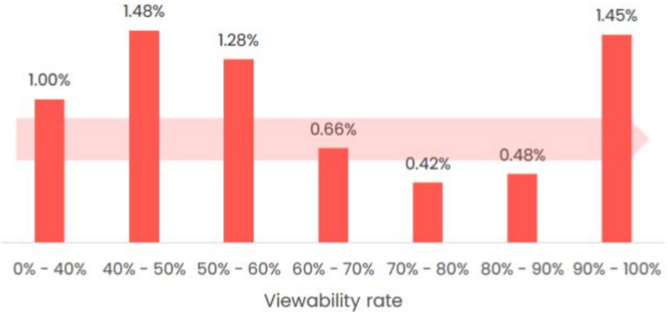


Source: PwC results following re-performance

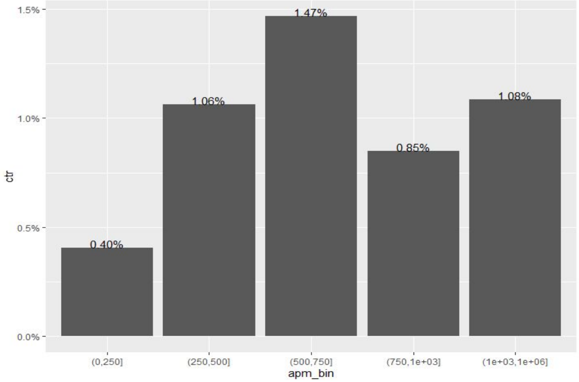
Reporting- Advertiser 3- Videos Mobile CTR

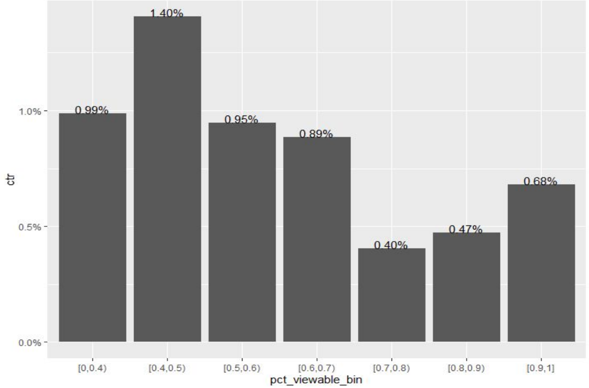
Source: Lumen Campaign Analysis Report





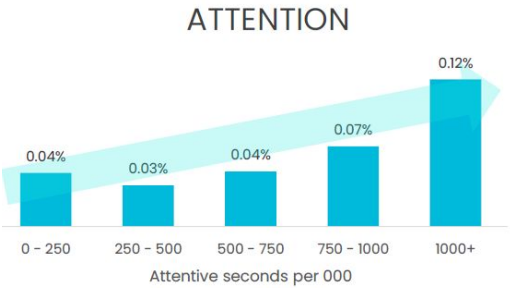
Source: PwC Results following re-performance

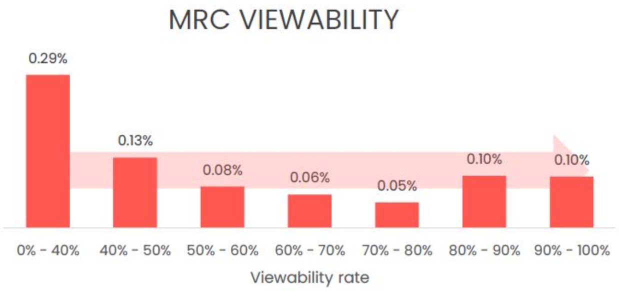




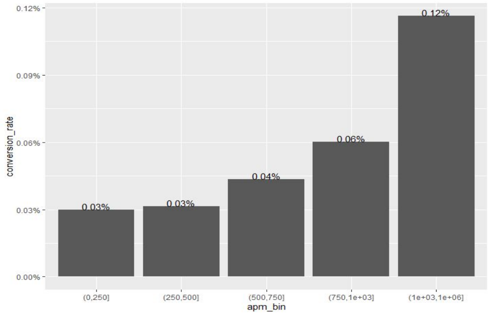
Reporting- Advertiser 3- Video Desktop Conversion

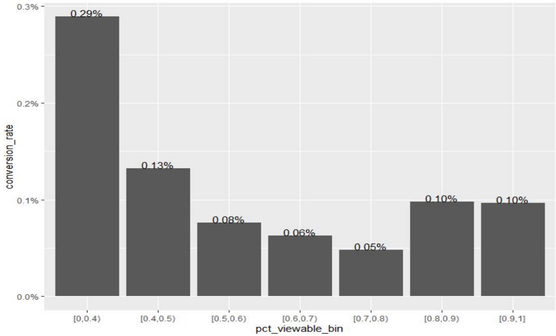
Source: Lumen Campaign Analysis Report





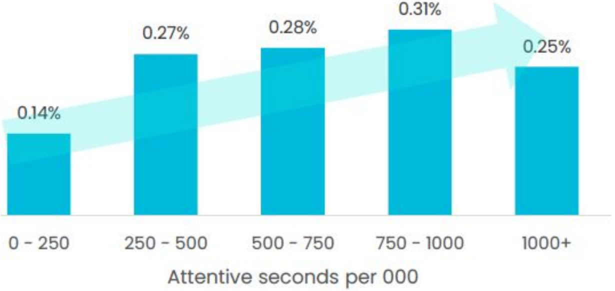
Source: PwC Results following re-performance



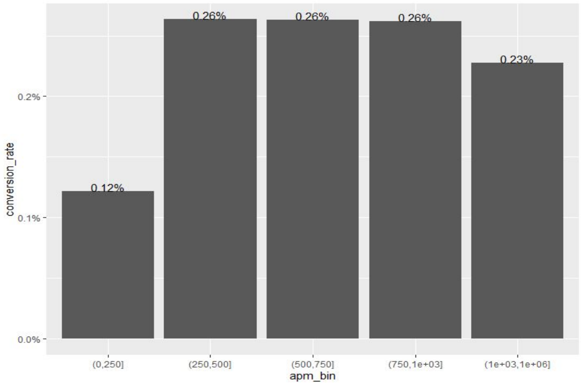


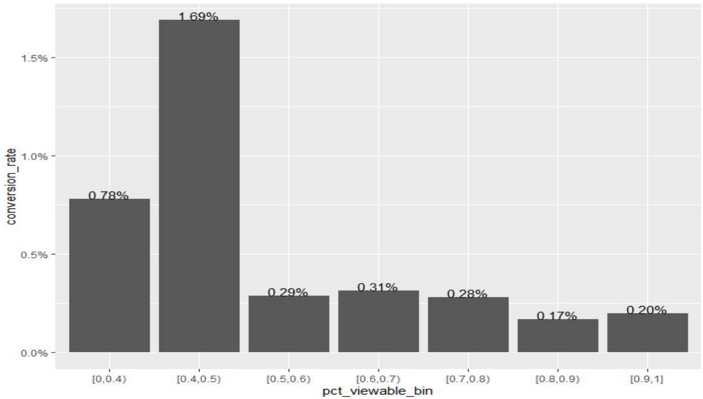
Reporting- Advertiser 3- Video Mobile Conversion

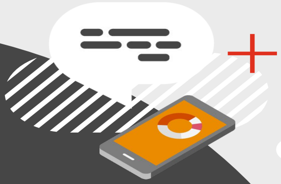
Source: Lumen Campaign Analysis Report



Source: PwC Results following re-performance





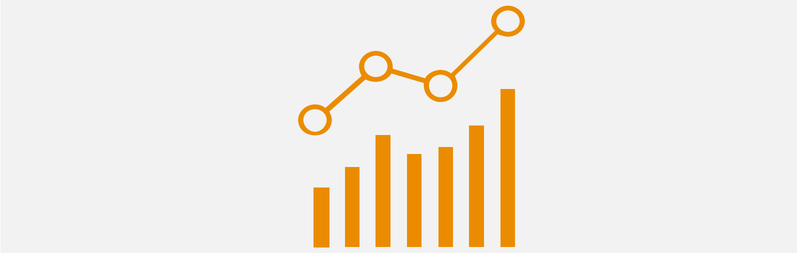


3

Advertiser 1
Brand Lift Study

Background

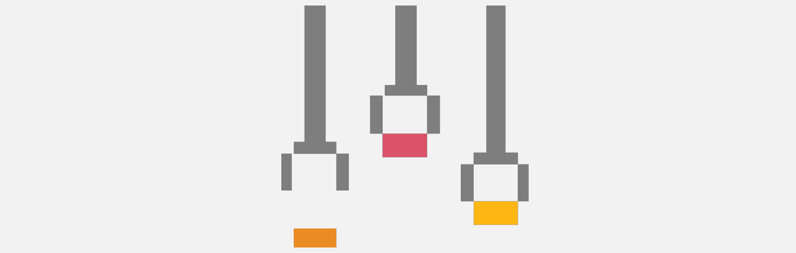
Dataset



For the purposes of testing Lumen's brand lift study with ODR, we used the complete sample of survey dataset and the impression-level attention dataset. The datasets contained the following information:

- Impression-level attention data
  - Users: 3,051 (raw) -> 2,402 (cleaned and matched)
  - Impressions: 12,827 (raw) -> 7,259 (cleaned and matched)
- Survey data with questions and answers related to demographic and brand awareness information
  - Participants: 5,504 control vs. 3,169 exposed

Our approach



- Reviewed Lumen's brand lift methodology for data modelling, statistical approaches deployed, limitations and mitigation, and plans for future improvement.
- Reviewed and tested Lumen's brand lift script with data samples provided and compared the resulting metrics and plots to the report content.

Methodology

PwC's assessment of Lumen's appropriate application of methodology in the scripts comprised of a review of the following components:

- **Data loading and cleaning**
  - Create weekday, month, week from timedate
  - Create control variables from answers and questions, including: Gender, Parents, Age Group, Annual Household Income, Living Situation, Advertiser 1 Main Shopper, Advertiser 1 nearby, Clubcard user, Weekday, Week
  - Create key outcome variables/ metrics (all True/ False) from answers and questions, including 1st Mention Awareness, Any Mention Awareness, Digital ad recall, Other ad recall, Short-term value, Short-term quality, Short-term range, Short-term price, Long-term value, Consideration: Any, Consideration: T2B, Purchase intent
  - Calculate main independent variables
    - *Attentive seconds= exp_time*exp_view*
    - Viewable impressions (50% and 100% respectively)-> *percentage viewable= viewable impressions/ impressions*
- **Modelling**
  - Conduct logistic regression on each key *outcome variables ~ control variables + log(attentive_seconds + 1)/ pct_viewable*
  - Calculate Mcfadden R-squared by each logistic regression referenced above and a null model (only for attention)
  - Get coefficient estimate, standard error, z-score and p-value of the model
- **Plots for each metrics**
  - Bar plot of each control variables contribution to Mcfadden R squared
  - Uplift compared to baseline (control group) against attentive seconds for all campaigns and each individual campaign
- **Export plots, Mcfadden R-square, statistical metrics, contribution of Mcfadden R-square**

Note

Please note that we were not able to test/ re-perform the analysis conducted by On Device Research (ODR) referenced on p.19 in Lumen's report as it was not owned by Lumen:

Key Results Summary from ODR's meta-analysis

ODR's analytical approach has involved:

- No, Low/High Attention and 100% Viewability buckets
- Control groups created for each, balancing on demographics and relationship to Tesco

Main findings have seen statistically significant lifts for awareness

But other metrics have seen little differences from the control

	High Attention	100% Viewability
Spontaneous Awareness – All Projects	+4%	+10% ▲
Spontaneous Awareness – 'Supermarket' specific projects	+9% ▲	N/A
Brand Consideration (Any) *(Only one I'd consider + Consider above most others + Consider along with others + Consider, but only after others)*	-3%	-2%
Brand Consideration (Repertoire): *(Only one I'd consider + Consider above most others + Consider along with others)*	-3%	-1%
Brand Preference: *(Only one I'd consider + Consider above most others)*	-3%	-1%
Purchase Intent: *(Extremely Likely + Probably Likely)*	-3%	0%
Long Term Value: *(Very good value + Fairly good value)*	+1%	+3%
Short Term Value: *(Offers a good value for money)*	0%	+1%

Findings

Result

- Our findings were consistent with Lumen's brand lift results which showcased that both attention and viewability have significant relationship with 1st mention awareness, any mention awareness, and digital ad recall.

Methodology

- The methodology used to assess the impact of attention and viewability (alongside a number of other customer features) on 12 brand metrics was logical, the data inputs used are appropriate and aligned to the methodology.

- We did observe, however, a slightly unconventional approach to calculating lift - it was calculated as the difference between an estimated metric value e.g. 1st Mention Awareness for the exposed group (people who have viewed an ad) and average of metric value for the control group (people who have not viewed an ad).

  Uplift=estimated value for exposed group - average value for control group

  In traditional literature on advertising effectiveness, lift is generally estimated as the delta of the estimated values of a metric of choice for the exposed group divided by the estimated value for a metric of choice for the control group if they had not been treated e.g, delta of 1st mention awareness values for the exposed group (people who have viewed/paid attention to an ad) vs 1st mention awareness for the control group (people who have not viewed/paid attention to an ad)

$$\tau_\ell = \frac{\Delta\text{Conversion rate due to ads in the treated group}}{\text{Conversion rate of the treated group if they had } \textit{not} \text{ been treated}} = \frac{\tau}{\mathbb{E}[Y^{obs}|Z=1, W^{obs}=1]-\tau}$$

- Mcfadden fit to measure the model performance was on the lower end, ranging from 0.02 to 0.25. In practice for a valid model, “a goodness-of-fit using McFadden's pseudo r-square (ρ2) between 0.2 and 0.4 should be taken to represent a very good fit of the model (Louviere et al., 2000)."*

Findings (continued)

Metrics importance and their ordering:

metric	Estimate	Std. Error	z value	Pr(>|z|)
1st Mention Awareness	0.119	0.064	1.87	0.062
Any Mention Awareness	0.380	0.065	5.88	0.000
Digital ad recall	0.117	0.061	1.93	0.054
Other ad recall	0.028	0.062	0.45	0.653
Short-term value	0.059	0.072	0.82	0.414
Short-term quality	0.043	0.068	0.63	0.526
Short-term range	0.046	0.062	0.74	0.461
Short-term price	0.073	0.057	1.28	0.201
Long-term value	-0.073	0.056	-1.30	0.194
Consideration: Any	-0.112	0.098	-1.14	0.253
Consideration: T2B	-0.068	0.056	-1.21	0.225
Purchase intent	-0.105	0.057	-1.83	0.068

Findings (continued)

Attention vs view / p-scores and z-scores:

	P-value					Z-value				
Metric	Attention	View 50%	log- View 50%	View 100%	log- View 100%	Attention	View 50%	log- View 50%	View 100%	log- View 100%
1st Mention Awareness	0.062	0.060	0.055	0.030	0.027	1.865	1.884	1.919	2.173	2.207
Any Mention Awareness	0.000	0.001	0.001	0.446	0.468	5.876	3.257	3.176	0.762	0.725
Digital ad recall	0.054	0.696	0.605	0.827	0.936	1.927	0.391	0.517	-0.219	-0.080
Other ad recall	0.653	0.785	0.771	0.792	0.771	0.449	-0.273	-0.291	0.264	0.291
Short-term value	0.414	0.405	0.300	0.036	0.037	0.817	-0.833	-1.036	-2.096	-2.082
Short-term quality	0.526	0.330	0.395	0.103	0.113	0.634	0.975	0.851	-1.630	-1.586
Short-term range	0.461	0.906	0.866	0.014	0.015	0.738	-0.118	-0.169	-2.465	-2.427
Short-term price	0.201	0.610	0.608	0.022	0.021	1.278	0.510	0.513	-2.295	-2.305
Long-term value	0.194	0.227	0.193	0.086	0.085	-1.300	-1.209	-1.302	-1.716	-1.721
Consideration: Any	0.253	0.276	0.229	0.312	0.257	-1.143	-1.090	-1.204	-1.012	-1.132
Consideration: T2B	0.225	0.700	0.767	0.600	0.623	-1.212	0.385	0.297	-0.524	-0.492
Purchase intent	0.068	0.336	0.308	0.088	0.088	-1.828	-0.961	-1.019	-1.706	-1.707

Thank you

This document has been prepared only for Lumen Research and solely for the purpose and on the terms agreed with Lumen Research in our agreement dated 6 September 2022. We accept no liability (including for negligence) to anyone else in connection with this document, and it may not be provided to anyone else.

© 2022 PricewaterhouseCoopers LLP. All rights reserved. 'PwC' refers to the UK member firm, and may sometimes refer to the PwC network. Each member firm is a separate legal entity. Please see www.pwc.com/structure for further details.