MULTIVARIANT ANALYSIS IN VIDEO AD MARKETING CAMPAIGNS FOR WEB TRAFFIC QUALITY OPTIMIZATION

Marta López and Fernando Perales
JOT INTERNET MEDIA España SL, General Ramírez de Madrid 8, 28020 Madrid, Spain

ABSTRACT

In digital marketing, performance can be used to determine which variables affect the quality of web traffic. There are different ways of collecting data, which allows us to enrich our data and gives us a lever to generate new ways of looking at and dealing with information. And yet, the large amount of data, the daily variation and the different level of granularity require dedicated data preparation. Our data will require prior preparation and cleaning, in order to be able to develop the multivariate and correlation analysis through the Cramer's V-diagram. Our objective, to analyze traffic quality, a parameter that is not only used by companies to verify the suitability of the audience concerned but is also directly related to the bandwidth awareness and profitability of video advertising campaigns. Analytical identification of the variables that most affect traffic quality can be used to optimize both marketing campaign structure and bidding strategy.

KEYWORDS

Video Ad Marketing, Quality Prediction, Data Preparation, Multi-Variant Analysis, Digital Marketing, Web Traffic

1. INTRODUCTION

During the last years, video marketing has been proved as a high-efficient media to engage the targeted audience at reasonable prices with good performance indicators (clicks and return of investment). These aspects motivates that the global video ad investment forecast predicts a continuous growth of 12% till 2025, reaching $148 B. This huge market means that enormous amount of web traffic, representing potential clients, will be redirected to the companies´ websites. However, based on the information reported by the video ad platforms in JOT video campaigns (Youtube), between 20-30% of this traffic comes from inauthentic users, mainly bots working through virtual protocol networks (VPNs) and programmed to spend the companies budget clicking in the ads with no post conversion, limiting marketing campaigns impact and reducing the return of investment by wasting companies' budget.

Due to this fact, in the market there exists several companies offering inauthentic traffic detection services:

• Click Cease: Offering a set of rules manually like time per session and clicks per time frame. Track conversions and post-click analysis. It is possible to define traffic filter rules at campaigns level and blacklist IP addresses and ranges limited up to 500.

• Click Guard: Enabling manual and automated filtering options based on the account level selected. It allows to manage the Google account based on the tracking template. The post click analysis is performed thanks to the ingestion of a tracking id.

• PPC (Pay Per Click) Protect: This service imports the tracking template, the blocking strategy is based on the traffic source. This service works at IP (Internet Protocol) number blocking.

There are also other solutions like ADEX, Traffic Guard and Traffic Cop that provide traffic monitoring, inauthentic traffic quantification and filtering and advanced analysis for sophisticated invalid traffic generation. However, none of them have access to historical data of the campaigns statistics so they need a long learning process to be able to define and implement reliable and robust traffic filters. In addition, some of them work on IP addresses which adds a GDPR related challenge to the development.

Considering this market status, the access to an already existing database describing the video web traffic (YouTube) of more than 12 websites during the last 6 months enables the development of unique the video ad traffic quality analysis, identifying which are the most relevant parameters conditioning the web traffic quality. This disposal also ensures that algorithms are trained, tested, and validated with real data, reaching the required insights quality and accuracy to be used in production conditions for decision making. The dataset generated compiles a set of 50 variables (combining performance indicators and descriptors as attributes).

At business level, by the multi-variant analysis for the identification of low quality traffic sources, it will be possible: (i) to increase the trust in video ad marketing as a reliable tool for business promotion, ensuring that all the users reaching companies' landing pages are real potential clients, so the marketing budget is fully spent in high quality audiences; and, (ii) the direct correlation of the video content to the user audience increase the user experience confidence, so the companies will get more and more interactions and visits through this media content.

Taking all the previous aspects into consideration, the goal of the paper is to answer the following questions:

- What are the main variables impacting the quality of the traffic?
- Is there any correlation among the variables used to describe the impact of the marketing campaigns?
- How can the traffic quality be predicted based on the analysis of historical data?

The paper is organized as follows. Section II presents the technical background related to multi-variant analysis and its relation to the quality prediction. In Section III, required work to generate the dataset is presented, covering data collection, key existing indicators and ETLs (Extraction, Transformation and Loading) needed to data merging. Section IV introduces the technical analysis is presented and the main results discussed. Finally, Section V highlights key conclusions and open issues for the future.

2. BACKGROUND

The easy and wide access to Digital Marketing (DM) services has allow the massive and periodic data collection of variables and indicators describing the performance of the marketing campaigns. This has pushed the marketing teams to approach data science as the only solution to process and analyze the vast amount of data available. Due to the large number of variables involved, one of the first goals is to understand their importance and correlation depending on the goals of the company, which allow the identification of the target parameter.

There are some references describing the different approaches for multivariant analysis (Black & Babin, 2019) and (Everitt & Hothorn, 2011) and specifically applied to marketing (Saura, 2021). They all state that in almost all the cases data preparation represents a significant bottleneck for marketing data processing, even conditioning the final quality of the results. For that reason, in our case the final data set used in the analysis has been specifically prepared, enabling both temporal analysis and correlations depending on the target variables: traffic quality (discrete variable) and conversion difference (continuous variable obtained by the difference in the reporting values for the same day with 7 days delay).

The final goal of this work is to generate a decision support system enabling the detection of inauthentic traffic sources based on real performance data analysis and predictive services, so the confidence and trust on the initial data must be maximized. For the generation of the predictive service, the temporal series of the variables must be considered. In Neusser (2016) and Hyndman and Khandakar (2008) the two conditions to develop this kind of services are defined:

1. availability of data about the past
2. assumption that past patterns will be replicated in the future

These two conditions are fully satisfied in our case, as data sets are collected daily from past months video marketing campaigns, also society interests shows periodic patterns depending on the seasonality and traffic quality can be described based on campaign indicators like the category and the keywords, which are 100% monitored.

The different stages of the process to develop the multivariant analysis are as follows[1,2]:

- *Objectives of the analysis*: This is the basis on which the technical work must be guided. The objective of the analysis is to determine whether there are correlations between the different variables that cannot be seen at first glance or by expert visualization and evaluate their influence on our target variables. The final goal is to understand how the quality of traffic to our ads can be improved to optimize the campaign results.

[1] http://www.est.uc3m.es/esp/nuevadocencia/getafe/estadistica/analisismultivariante/doc generica/archivos/metodos.pdf
[2] https://bookdown.org/chescosalgado/intro_r/tipos-y-estructuras-de-datos-en-r.html

Selection of data: As mentioned earlier, this is a crucial step. The right choice and quality of data determines to a large extent the satisfaction of the conclusions. It must be determined the time range studied, the sample size and the explanatory variables to be observed. Starting from the assumption of independence, it is analysed whether this is fulfilled.

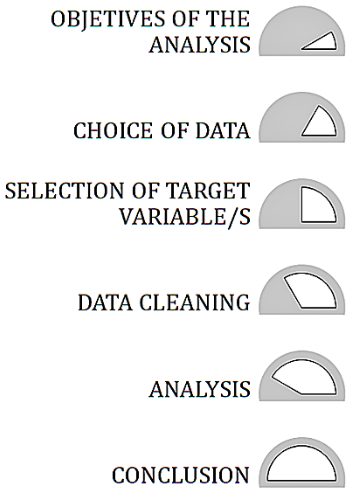

Figure 1. Steps of the data analysis

Selection of the target variables: Representing the variables for which the analysis with show how the rest of the parameters affect. It can be one or several depending on the results obtained. In addition, data has been processed differently depending on its type. There are two types of target variable: (i) Quantitative variable (Composed of numerical values) and (ii) Qualitative variable (Composed of factors).

- *Data cleaning*: Prior to any type of analysis, the data must be cleaned to ensure that they are optimal. Among the cleaning tasks, a preliminary analysis is carried out, plotting the variables to see their distribution as well as their main statistics. This has allowed to decide how to transform the variables to select its correct typology. This analysis also provides information about the distributions of the different variables and their possible statistical transformations to create a predictive model. At the same time, the data cleaning also manages the outliers and missing values. Once this initial treatment is developed, the variables will be ready for use in a single dataset.
- *Analysis*: A treatment of correlations has been carried out, both between individual variables that are relevant to our study and between all the inputs. It has been carried out thanks to the correlation matrix (Cario & Nelson, 1997). In this way, it has been possible to analyze the interdependence and avoid multi-collinearity in future predictions. In this case it is a very useful step to see relationships that cannot have seen otherwise. It has been also seen how the inputs affect in order of importance on the target variables by means of the Cramer's V-plot (Isea, et al, n.d).
- *Conclusions*: This is the final step of our analysis. Based on all the data processed, infer the results and whether the hypothesis set out at the beginning has been fulfilled.

3. DATA MARK GENERATION

The most sensible part in any analytical project is the generation of the final dataset. The data included always impact in the results and insights strongly. Therefore, it must be as complete as possible. Also, the knowledge of the context and quality of the inputs support the decision of choosing the best data base to use.
In this case, several data sources have been used:

• *Google Ads*: Providing historical information about the impact statistics and investment of the campaigns, showing the cost information used to measure the profitability and the performance of ads.

• *Partner report*: Showing the economic results (revenue) together with the results of each advertisement once they have passed the corresponding traffic quality filters.

• *Intern information*: Containing internal information of the accounts, like identifiers enabling the data merging between different tables.

• *Partner traffic quality report*: Containing a critical part of our analysis, because shows the traffic quality of our ads grouped by a tag that allow distinct the different landings.

Both the definition of the problem and the knowledge of the project goals at technical and business level represent the adequate guidance for the analysis and predictions. If the data does not feed the predictive model with relevant inputs, the results will be surely poor and with low accuracy.

3.1 Collecting Video Ad Marketing Data

For this work, four different sources have been used to generate the data set. In this case, the data merge is successful because of the ad identifier, that allow us to analyze the information as disaggregated as possible. After the data merge, has been necessary the data preparation processes followed by the modelling part. These three steps will be the main points of the analysis. Data download has been automated through an API and collected in our SQL data base. Data access has been customized through SQL queries directly connected to excel, where four tabs corresponding to the four data sources are generated.

Within the excel framework, data has been transformed and merged using power query, which allows to have all the data updated and joined.

The specific technical details of each data source are the following:

• Google Ads: Providing cost information at the ad and day level. It also includes campaign and ad group names and ids. To do the data crossing with the other databases the *ad id* and *date* are used as they are unique identifiers.

• Partner report: Providing revenue information at the ad and day level. The data crossing is implemented with the ad id (at this case the name is "*type tag*") and the date. This report also includes the information of traffic quality per ad.

• Intern information: It is essential to the crossing because it relates *ad id* to *source tag* (provided by the partner includes all the landing pages of the same client) and accounts, campaigns, and ad information.

• Partner traffic quality report: This report is different than the others. In this case the data is grouped by source tag. For that reason, the report includes the information by country too, so it can be crossed with the rest of data sources.

• The final dataset (Figure 2) resulting from the data crosses. It includes 50 variables and allows us to begin with the multi variant analysis.

3.2 Campaign Statistics and Key Performance Indicators

With the aim of developing the most complete analysis possible, all the potential variables which we have access have been included. They describe: (i) the financial performance (Cost, revenue, Cost Per Click, Cost Per Acquisition, Revenue Per Click), (ii) date (day of the week, month, year, date), (iii) marketing account specifications (id, device, country, code, category, keyword, destiny, source tag) and (iv) performance indicators (clicks, impressions, conversion rate, conversions, traffic quality). These values are collected two times (initial and final), before and after Google traffic invalidation, and their differences are also analyzed.

Thanks to the disposal of this high-quality data, it is possible to carry out a wide catalogue of plots for visual inspection, like temporal representations, histograms, data exploratory for error detection and outliers and so on. This work is part of the data preparation rather than about the analysis for predictive modelling so no further details are included. The next figure contains the preliminary inspection of the dataset, that allows us to know the raw information of each variable. The first benefit of doing this type of representation is because the type of variable is quite clear to appreciate, and it give us feedback for the next steps. For example, some factor type variables (categorical) are wrongly coded as character. Others, such as day of the week, month, or year, can be changed to factor due to the low number of categories.

```
## tibble [237,409 x 50] (S3: tbl_df/tbl/data.frame)
##  $ Fecha              : POSIXct[1:237409], format: "2021-12-01" "2021-12-01" ...
##  $ Dia_sem            : chr [1:237409] "miércoles" "miércoles" "miércoles" "miércoles" ...
##  $ Dia                : num [1:237409] 1 1 1 1 1 1 1 1 1 1 ...
##  $ Mes                : num [1:237409] 12 12 12 12 12 12 12 12 12 12 ...
##  $ Año                : num [1:237409] 2021 2021 2021 2021 2021 ...
##  $ Cuenta_id          : chr [1:237409] "1238797147" "1238797147" "1423963431" "1423963431" ...
##  $ Dispositivo_revenue : chr [1:237409] "Desktop" "Desktop" "Desktop" "Desktop" ...
##  $ Source_tag         : chr [1:237409] "cbs_d2s_xmlb_2236_answersite_gdn2" "cbs_d2s_xmlb_2236_answer
site_gdn2" "cbs_d2s_xmlb_2236_helpwire_gdn" "cbs_d2s_xmlb_2236_helpwire_gdn" ...
##  $ Source_tag_numero  : num [1:237409] 2 2 1 1 1 1 1 1 1 1 ...
##  $ Destino_codigo     : chr [1:237409] "ANS" "ANS" "HEL" "HEL" ...
##  $ Pais               : chr [1:237409] "USA" "USA" "USA" "USA" ...
##  $ Pais_revenue       : chr [1:237409] "USA" "USA" "CAN" "USA" ...
##  $ Categoria          : chr [1:237409] "Insurance" "Insurance" "VideoMeeting" "VideoMeeting" ...
##  $ Keyword            : chr [1:237409] "insurify insurance quotes" "botox for tension headaches cove
red by insurance" "video meeting system" "video meeting system" ...
##  $ Coste_i            : num [1:237409] 0 0.629 1.979 1.979 1.197 ...
##  $ Coste_f            : num [1:237409] 0 0 0 0 0 ...
##  $ Cost_diff          : num [1:237409] 0 1 1 1 1 ...
##  $ Revenue            : num [1:237409] 2.202 0.479 0.607 7.126 0 ...
##  $ Profit_i           : num [1:237409] 2.2 -0.15 -1.37 5.15 -1.2 ...
##  $ Profit_f           : num [1:237409] 2.202 0.479 0.607 7.126 0 ...
##  $ ROAS_i             : num [1:237409] 2.202 -0.239 -0.693 2.601 -1 ...
##  $ ROAS_f             : num [1:237409] 2.202 0.479 0.607 7.126 0 ...
##  $ CPC_i              : num [1:237409] 0 0.629 1.979 0.165 0.599 ...
##  $ CPC_f              : num [1:237409] 0 0 0 0 0 ...
##  $ RPC                : num [1:237409] 1.101 0.479 0.607 3.563 0 ...
##  $ CPA_JOT_i          : num [1:237409] 0 0.629 1.979 0.99 0 ...
##  $ CPA_JOT_f          : num [1:237409] 0 0 0 0 0 ...
##  $ CPA_G_i            : num [1:237409] 0 0 0 0 0 0 0 0 0 0 ...
##  $ CPA_G_f            : num [1:237409] 0 0 0 0 0 0 0 0 0 0 ...
##  $ CR_JOT_i           : num [1:237409] 0 1 0.5 1 0 0.5 0.2 1 0 0 ...
##  $ CR_JOT_f           : num [1:237409] 0 0 1 2 0 0 0.25 1.25 0 0 ...
##  $ CR_G_i             : num [1:237409] 0 0 0 0 0 0 0 0 0 0 ...
##  $ CR_G_f             : num [1:237409] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Impresiones_i      : num [1:237409] 0 1 4 4 2 2 8 8 2 0 ...
##  $ Impresiones_f      : num [1:237409] 0 0 0 0 0 0 6 6 1 0 ...
##  $ Impressions_revenue : num [1:237409] 1 1 1 12 2 3 1 9 8 1 ...
##  $ Clicks_i           : num [1:237409] 0 1 2 2 2 2 5 5 0 0 ...
##  $ Clicks_f           : num [1:237409] 0 0 1 1 0 0 4 4 0 0 ...
##  $ Clicks_revenue     : num [1:237409] 1 1 1 12 2 3 1 9 8 1 ...
##  $ Conversiones_i     : num [1:237409] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Conversiones_f     : num [1:237409] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Conversiones_revenue: num [1:237409] 2 1 1 2 0 1 1 5 1 1 ...
##  $ Conv_diff_i        : num [1:237409] 2 1 1 2 0 1 1 5 1 1 ...
##  $ Conv_diff_f        : num [1:237409] 2 1 1 2 0 1 1 5 1 1 ...
##  $ Conv_diff_i_%      : num [1:237409] 2 1 1 2 0 1 1 5 1 1 ...
##  $ Conv_diff_f _%     : num [1:237409] 2 1 1 2 0 1 1 5 1 1 ...
##  $ Vistas_video_i     : num [1:237409] 0 1 5 5 2 2 6 6 1 0 ...
##  $ Vistas_video_f     : num [1:237409] 0 0 1 1 0 0 4 4 1 0 ...
##  $ TQ_TypeTag         : num [1:237409] 0 0 2.5 2.5 0 2 3 3 0.5 0 ...
##  $ TQ_SourceTag       : num [1:237409] 2 2 3 3 3 3 3 3 3 3 ...
```

Figure 2. Set of variables included in the dataset combining numeric and categorical variables

4. MULTIVARIANT ANALYSIS AND DISCUSSION

As mentioned previously, the goal of this analysis is to understand what variables affect more significantly the traffic quality in the video ad marketing campaigns. To do so, two target variables have been selected:

(i) *conversion difference*, which explains the invalidation of our ads and which we will consider the quantitative variable,

(ii) *TQ (traffic quality)* at source tag level, which gives us information about what value on a scale from 0 to 10 our source tag has. The rest of the variables are considered as explanatory or input variables.

Therefore, the first action implemented has been to correct the wrong data typologies. This allow us to plot the distribution of the variables where more information was obtained:

(i) In relation with the numerical variables, many of them show a right tail. It implies that they are skewed variables, with most of the data are clustered in the first values of the variable. Would be interesting to analyze if will be useful a transformation to normalize the data. The most common using transformations are logarithms, exponentials and roots, but exists models like h2o that do it automatically.

(ii) Another advantage is that potential groupings of the categorical variables are identified to make them more efficient, and if it helps us to improve our conclusions.

Given the fact that the data is classified by dates, it was convenient to consider whether there were temporal patterns. The best way to do this was to repeat the graphs of the distribution of variables, but this time adding the time factor. In our case it was proved that there is not much information in the input variables, however, the difference in conversions decreased over time. On the contrary, the score in the quality of traffic per source tag increased, confirming the importance of considering the temporal representation of the series.

Once implemented the initial inspection of the data, several transformations are carried out for the data cleaning, modifying the types of variables and transforming those that can be grouped. In our case, the traffic quality values has been transformed from numerical to categorical (Table 1), grouping them into a factor type with four categories (lowest, low, medium, high), so that it is easier to analyze them.

Table 1. Distribution of ad quality classification

Quality Classification	% of ads	TQ number range
Lowest	3.05	0
Low	86.93	1-3
Medium	7.39	4-7
High	0	8-10

The next step was the analysis of outliers (either very large or very small). The percentage of outliers per variable should be calculated to determine their potential impact and relevance in the data set. Based on the theory, if the result is less than or equal to 10% of the data, the outliers could be transformed to missing values, as they may be considered to have a low incidence. As seen in "*Table 2"*, almost all the variables were below this threshold, so it is considered that they are not impacting the final results and can be transformed to missing values. In any case, before that, we will create a new variable to analyze the impact of the outliers, it will be called "Prop_missing" and it contains the provision of outliers.

Table 2. TOP-10 of variables with more outliers

Variable Name	% of Outliers
ROAS_i	10.38
ROAS_f	10.17
Impresiones_f	8.28
CPA_JOT_f	8.06
CPA_JOT_i	7.87
Impresiones_i	7.55
CPC_f	7.33
CPC_i	7.18
Revenue	6.17
Clicks_revenue	4.45

Finally, missing data were treated, so that if they exceed 50% of the data, the variable should be deleted as it would not yield any information. It is usual to impute missing data by the mean, median or both randomly, as they are then replaced by data that follow the same pattern. In our case the incidence was less than 10%, so we imputed randomly, eliminating any missing values.

Once our database was treated and free of significant outliers and missing values, the complete dataset was analyzed by means of a Cramer's V charts (Figure 3 and Figure 4). Cramer's V is a measure of the strength of association between two nominal variables based on chi-square, and Cramer's V chart returns in descending order the most significant variables in relation to the target variable. To increase the robustness and reliability of the results and as a quality control, two random variables have been added, so that any variable behind them is pure randomness and can be no relevant at all in the analysis.

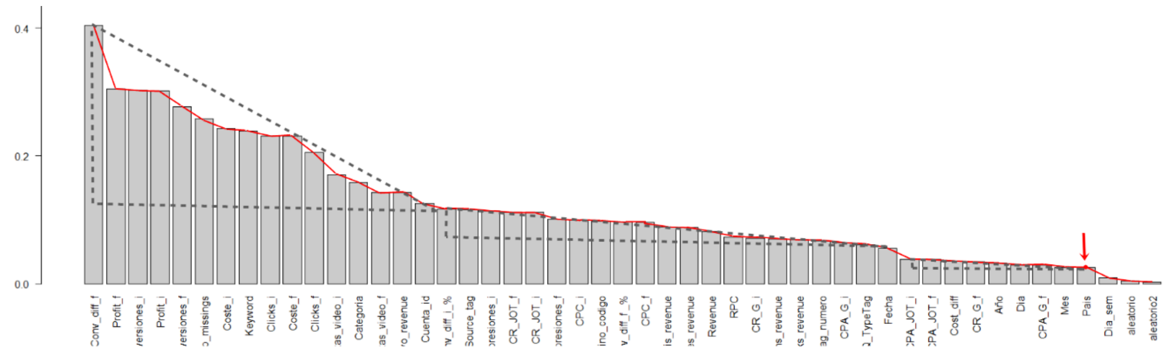

Figure 3. Importance of variables. Results of Cramer's V Plot - Conversion difference

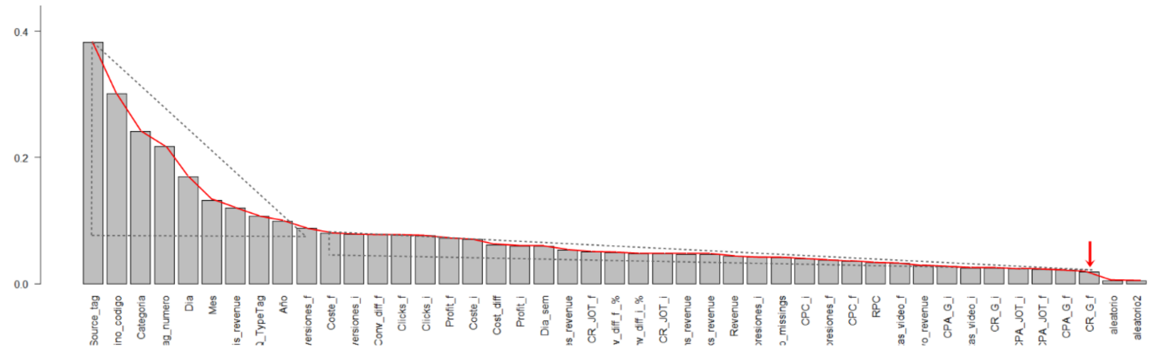

Figure 4. Importance of variables on TQ. Results of Cramer's V Plot - TQ by SourceTag

The first thing to notice is the different slopes founded in both graphs. The first and more pronounced belongs to the most relevant variables in relation to the target variable, the second one with more cushioned slope includes variables with less impact. The random variables allow to know what variables have random behaviour. Its last position implies a good sign of good choice of variables, because all of it are relevant for our target variables. Excepting one, "Dia_sem" that represent the day of the week has a similar behaviour as the random variables in Figure 3. In this case, we could remove it.

To answer the first main question, and ignoring the intrinsic relationships of digital marketing, what more relevant in this case, is that the category of the keyword used has quite a lot of weight on this adjustment in conversions. This means that there are keyword categories that show more invalidation than others. When analyzing the value assigned by the partner to the quality of the traffic (*Figure 4*), the category also plays an important role. This time the importance is higher, being the second. It is also remarkable that the day of the month also influences whether the traffic is better or worse considered. Based on business experience, it is also known that depending on the country the TQ is higher or lower. However, due to the differences in measurement between Google Ads and the partners' monetization feed, there are usually traffic leaks to other countries than the one targeted by the campaigns. This leakage also has some influence on traffic quality. The day is also important because Mondays usually have more records than other days, and it used to be the best day in revenue terms.

5. CONCLUSION

In short, thanks to this analysis, some unexpected relationships and variables impact on the conversions and TQ can be deduced that would have been impossible to identified otherwise. This type of analysis provides the basis for further knowledge and research into the predictive patterns resulting from it.

To answer the main second question about correlations, most of the explanatory variables in our model are calculated from other variables that we use as a basis. This means that the most likely, is that the variables are correlated with each other. This has been demonstrated by generating plots of influences and the correlation matrix. We found low to medium correlation between them, with the highest correlation being between initial clicks and initial cost measured on the Google Ads platform, with an 80% correlation. In fact, the biggest correlations we see are intrinsic to our business model.

Once the key parameters impacting the final quality of the traffic has been detected, next steps are now focused on the development of the algorithms enabling the prediction of the video ad marketing traffic quality, the main third question. Some objectives of the algorithms will be: (i) The reduction of inauthentic users will increase the trust in this digital media as channel to improve industry investment in video marketing, (ii) Video traffic authentication represents a new approach of optimizing the impact of the campaigns, reducing the dependencies in third parties, (iii) Development of video marketing campaigns decision support based on real user interactions, not being affected by the actions carried out by malicious bots and (iv) This project will deploy a new way to engage the real targeted audience and promote competitiveness in digital media content. All of them considering the existence of multicollinearity and that there may be external factors that we cannot handle or that we have not seen before. Therefore, it is important to have up-to-date information, digital marketing evolves day by day and we have to fluctuate with it.

Next steps are currently focused on the data standardization for model generation. There are a set of candidates that would perform better based on multilinear regression (LightGBM) for the conversion difference continuous variable and classification (Random Forest) for the categorical one, traffic quality.

ACKNOWLEDGEMENT

This work received partial funding from the European Commission Horizon 2020 AI4Media project (grant number 951911).

REFERENCES

Black, W., Babin, B.J. (2019). Multivariate Data Analysis: Its Approach, Evolution, and Impact. In: Babin, B., Sarstedt, M. (eds) The Great Facilitator. Springer, Cham. https://doi.org/10.1007/978-3-030-06031-2-16

Cario, M. C.& Nelson, B. L. (1997). Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix (pp. 1-19). Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Everitt, B., & Hothorn, T. (2011). An introduction to applied multivariate analysis with R. https://doi.org/10.1007/978-1-4419-9650-3.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: TheforecastPackage for*R*. *Journal of Statistical Software*, *27*(3). https://doi.org/10.18637/jss.v027.i03

Isea, R. et al. (n.d) Central University of Venezuela. Facultad de Humanidades y Educación. Retrieved from https://mariafatimadossantosestadistica1.files.wordpress.com/2018/06/coeficientes-v-de-cramer-y-c-de-pearson.pdf

Neusser, K. (2016). Time series econometrics. Springer Texts in Business and Economics. https://doi.org/10.1007/978-3-319-32862-1

Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, *6*(2), 92-102. https://doi.org/10.1016/j.jik.2020.08.001