Title: Ad campaign analytics for video-on-demand platforms

Author: Jack Thurman

Supervisor: Wei Zhou

Project description:

Video-on-demand (VOD) is becoming more prevalent than traditional TV. Platforms like YouTube, Twitch and other streaming services offer consumers a way to relax, learn and catch up on shows. Advertisers have taken advantage of these platforms' massive number of users to help sell their products or services. Ad campaigns provide revenue for the platform and the advertiser; thus, useful insights are required to make effective choices.

One of the biggest issues surrounding advertising is spending their money in a way that will give them the results they are looking for. Many attributes can impact whether a marketing campaign is successful or not. These attributes are what I intend to analyse, visualise and optimise. Advertisers should be able to input their ad campaign data into my system and be able to deeply understand how it has performed in areas such as reach, cost per impression and how users respond to the creative itself.

Users will input their campaign data first, features such as start date, end date, product and target audience. Then, all impressions that have been served throughout this campaign will be input into my system. This will allow me to understand the demographic responding to the creative, the time of year and day in which the creative excels and how users are interacting.

Digital marketing research has revealed that ‘data analytics helps marketers to make data-driven decisions and understand the impact of their efforts on the target audience, allowing them to refine their approach and achieve better results.’ (Adwan et al., 2020) I will be utilising similar key performance indicators proposed by Adwan to measure campaign success, but solely with regards to video-on-demand campaigns.

My chosen project should allow me to showcase my ability in NoSQL databases, front-end development using Python and a host of other data science and machine learning libraries. I believe the project has the depth to allow me to showcase my skills while remaining simple enough so that I can focus solely on creating a high-quality tool that has the potential to make a real-world impact.

Aims and objectives:

My end goal for this project is to have a functional dashboard in which users can view multiple ad campaigns and understand how well they performed and the factors behind their performance. The dashboard will allow users to optimise any of the campaigns and be able to see exactly how they have been optimised: what was changed and what is the result of changing these things.

Objective 1: Research

Firstly, I must understand whether I would be able to find real useful data, or if it would be more realistic for me to generate my own. I must research exactly how ad campaigns perform in the real world, considering the time of day, the time of year, the demographic ads are targeted at, and whether certain demographics respond well to certain ad campaigns. This will enable me to create fake data that is as representative of the real world as possible.

Objective 2: Modelling

After thoroughly researching the data available, I will choose between using an existing data set, using a partially real data set, or generating my own. The data must have enough features and scale for me to be able to train an effective model.

Objective 3: Front-end Development

Create an easy-to-use dashboard that will allow users to gain a thorough understanding of the performance variance between campaigns. Depending on the data available to me, I will choose all visualisations later.

Objective 4: Training

Train various machine learning/deep learning models to learn from the fake data and be able to predict and optimise future campaigns. These models could consist of time series analysis using LSTMs with TensorFlow, classification models from Pythons Sci-kit learn libraries such as random forest, and potentially clustering algorithms like K-Means.

Feasibility:

Data gathering is the main problem associated with my project. A dataset on the scale I would need and with enough features is hard to come by. Consumers viewing data is extremely high-value and does not come at a small price. So, after having done a large amount of research on studies surrounding this topic, if I am unable to find an acceptable dataset, I will have to create my own.

If I am successful in finding a full dataset or partial dataset that will be useful to me, I will ensure to check licensing agreements. Making sure the data I am using is public and is free to use. Datasets will be referenced in the way that the owner requests. If finding a public dataset is not viable, then I am left with creating my own, either by human trials or by generating realistic fake data. A human-centred method of creating a dataset would consist of asking many people to watch many ads over a large period. This is also not viable for me. This leaves me with the option of creating my own dataset with the scale and features that would allow me to draw useful insights from it. If my data is going to be generated by me using different statistical models and Python libraries, there are no ethical or legal concerns. In terms of special resources, the only resources I would need are a Git repository and my PC. Licenses will be considered for all software and tools I am using, declaring them as and when I need to.

Human surveys may be incorporated to assess the usability of my dashboard design. This will most likely be anonymous, so simple ethical approval will be required from my supervisor.

Work Plan:

Week 1:

- Understanding what kind of data I will need and what features will be useful to me.
- Doing research into formal studies that are like my project. Determining if there are any improvements I can make to my project and if they have any public datasets that could be useful to me.
- Writing my initial report

Deliverable: Initial report

Week 2:

- Research what kind of front-end libraries, and database software are available to me. Understand the benefits of each and ultimately decide on what to use.
- Doing further research into related studies, continuing from last week.
- Use the information researched to decide on whether I am to use real data or generated data.

Deliverable: A write-up explaining software options and an explanation of my dataset choice

Week 3:

- Generate the data as realistically as possible, using distributions and models that were researched.
- Begin to model the data I have found or generated. Pre-process the data and store it in a database.
- Take some time to understand the data I have. If it is generated, I will need to understand exactly how the generated features can be as realistic as possible.

Deliverable: A database containing the ad campaign data that has been modelled accurately

Week 4:

- Begin to create the front-end
- Set up a basic functional structure
- Add menus, navigation bars, buttons and titles/headers

Deliverable: A working partial front-end that will allow me to add features easily

Week 5:

- Understand the metrics I will be using
- Create a detailed list of possible metrics, what they would represent, and how they would be useful to the user
- Ensure I have the correct data to visualise these metrics
- Decide on metrics and implement early visualisations

Deliverable: Write up of metrics and reasonings and a front-end containing some of the metrics that I have spoken about

Week 6:

- Polish metrics and the front-end
- Allow switching between different ad campaigns, and each ad campaign showing their respective visualisations
- Add the optimise and compare menu

Deliverable: A mostly functional front-end

Week 7:

- Begin to implement machine learning algorithms
- Understand what algorithms would result in useful insights
- Create a report section on all possible algorithms, how they would be useful and the technical applications of these

Deliverable: A write-up explaining different ML models

Week 8:

- Implement the algorithms
- Test on mock data and ensure they are working correctly
- Polish optimise/compare screen

Deliverable: A working ML/DL optimisation algorithm

Week 9:

- Complete the front end fully, ensuring everything works as expected.
- Begin testing the front-end

Deliverable: A functional front-end

Week 10:

- Test front-end, edge case inputs, buttons and all that interactions happen without errors
- Potential usability surveys and user feedback.

Deliverable: A comprehensive testing report

Week 11:

- Start writing the final report. Understanding the layout and headers, I will be using.
- Create a plan for what I will be writing in under each header

Week 12:

- Fill in each heading based on notes added last week.
- Followed by a thorough check of both the system and the report.

Deliverable: A final report

References:

Adwan, A.A. et al. (2023) 'Data analytics in digital marketing for tracking the effectiveness of campaigns and inform strategy,' International Journal of Data and Network Science, 7(2), pp. 563–574. https://doi.org/10.5267/j.ijdns.2023.3.015.