Thesis Proposal

Student: María Eugenia Irala
Tutor: Elena Fumagalli

Motivation

The emergence of social media platforms in the last decade has completely changed our way of interacting with each other and our perceptions of other people's lives. Since social networks such as Facebook, Instagram, YouTube and Twitter were created, people slowly started to integrate these internet innovations to their daily lives, not only as a communication channel, but also as a platform where online news, photos and experiences could be shared, and where people could even buy products.

Of course, brands and companies eventually realized that social media platforms were turning into a powerful opportunity to do online marketing. In fact, this growing business need to have an online presence was contemporary to the emergence of social media influencers (SMIs), who can be defined as "online celebrities" that are able to influence audience attitudes and opinions by exhibiting their personal lives through social media. Thus, Influencer Marketing arises from the combination of these two factors: consumer behavior was being increasingly influenced by social media and companies started to have an interest in digital influencers as a brand communication and marketing tool.

Therefore, along with Influencer Marketing a new business problem also arises: which influencers should companies and brands choose in order to maximize their sales? According to literature on this topic, firms traditionally use number of followers or engagement rate to select which influencer to hire to advertise their product and services. Anil Narassiguin and Selina Sargent in their paper called *Data Science for Influencer Marketing: feature processing and quantitative analysis* state that: "The Influencer Marketing (IM) field still rely on simple quantitative metrics such as community size or engagement rate to estimate the value of an influencer".

However, more recent studies are questioning this methodology and proposing new ways to find the "optimal influencer" to build the most profitable online marketing strategy. Along with these disruptive approaches, we identified some drawbacks of the established methodology:
*   It is well known that, nowadays, companies have access to a big amount of data. Therefore, the fact that only a few numerical variables (number of followers, likes, comments) are being used is reductive because his approach doesn't take into account the audience's characteristics.
*   There's an implicit assumption that the bigger the influencer (in amount of followers), the higher the return. But, according to Disha Dinesh in *Why Micro-Influencers are a Social Media Marketing Imperative for 2017*, there is a tendency for consumers to trust microinfluencers more because ad aversion is increasing.

Objective

Based on what was outlined above, a gap that needs to be addressed is to investigate which consumers' characteristics match which influencer characteristics to maximize returns (e.g., similarity). Accordingly, our main objective is to develop a new approach that addresses the two drawbacks mentioned above and help companies adopt a broader perspective in the process of choosing the optimal influencer to advertise their products. In order to do this, we are interested in investigating which characteristics of the influencers have an effect on the consumers' willingness to pay.

Specifically, our research question is: *Does similarity between the influencer and their followers have an impact on business results?* It is intuitive to think that the more identified the follower feels with the influencer, the more they are willing to pay for the product advertised by the influencer. However, more interestingly, this identification process could also have an effect on willingness to suggest the product to others, influencer's credibility and follower's repurchase intention.

Method

Data

The data that will be used in this research is provided by Upfluence, which is a company that collects and sells influencer's information to different firms that are interested in advertising their products through social media influencers. In particular, the data we received was extracted from Instagram.

We have two data sets:

1. Instagram Influencers: This data set has information of 99,877 influencers. Each row corresponds to one influencer and each column corresponds to a feature of said influencer. We have 20 features: Influencer_id, email, name, language, country, address, gender, age_bracket, instagram_id, username, full_name, bio, website, followers, engagement_rate, total_posts, total_engagements, total_likes, total_comments, category
2. Instagram posts: This data set has information of 438,548 posts. Each row corresponds to one post of one influencer, and each column corresponds to a feature of said post. We have 12 features: id, instagram_id, text, type, location_name, timestamp, likes, comments, post_id, views, media_type, thumbnail_url.

Procedure

Descriptive Step: Use clustering algorithms on the data provided to obtain different influencer profiles. The idea is to run different algorithms and then choosing the one that is most precise in terms of results and business sense.

Experimental Step: Run controlled experiments with online panels (Amazon Mechanical Turk or Prolific). The idea is to manipulate some variables we obtained from the descriptive step and test how different influencer profiles or behaviors (for example, different influencer photos to test similarity levels with the sample audience) affect business relevant variables (willingness to pay, willingness to share content, credibility, etc). After data collection, data will be analyzed using ANOVAs to check whether the manipulated variable has a significant effect on our selected dependent variables.

Managerial Implications

Our study will have many interesting applications. Firstly, our approach will eventually help companies and brands selecting a more appropriate Influencer Marketing strategy. Moreover, our results could be useful for intermediaries like Upfluence (our data provider) when giving advice to their clients (companies) in choosing which influencer to hire. Companies like Upfluence that operate as intermediaries between firms and influencers, will be able to clearly explain which types of influencers are currently present in social media and rely on the experimental results to give advice according to the profile of the company. This data based tips will certainly help firms to make influencer marketing decisions.

Secondly, firms could use our methodology to run their own clustering analysis and obtain a customized solution. Using their own influencer profiles, they could run experiments or A/B Testing with their customer base and their products.

Finally, taking into account all the existing literature about Influencer Marketing and social media influencers, our work can be considered as a first attempt to combine clustering analysis (machine learning technique) and experiments (marketing research method). Method triangulation allows us to obtain insights that could not be obtained with the descriptive / big data step alone. We believe that this combination of methodologies could be added to the "data scientist toolbox" and enable managers to obtain fruitful and actionable insights.