**Individual Project**

# Influencer Recommendation System for L’Oréal

Prepared for:

Dr. Hyunji So

INSY 660: Coding Foundations for Analytics

Desautels Faculty of Management

McGill University

Prepared by:

Chiara Lu (2611669959)

August 30st, 2023

In Data Analytics, Data collection and Data Preprocessing are the steps taken to gather, clean, and organize bits of information. The company I chose to develop a data collection and preprocessing plan for is L’Oréal.

# Data Collection

## Goals and Objectives

The goal of this creator recommendation system is to automate the process of finding the most suitable influencers matching L’Oréal’s division profiles. Currently, L’Oréal has four divisions, as shown in Table 1.

*Table 1: L’Oréal’s Divisions and their Key Differentiating Factor*

A collage of women with text overlay

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| Offers **luxury** **beauty** brands like Lancôme, Yves Saint Laurent and Giorgio Armani | Democratizes **cosmetics** and **personal care** for the **general public** such as L’Oréal Paris, Garnier, Biotherm, and NYX makeup | Formulates **skincare** solution for **specific skin** concerns or conditions. Brands include Khiel’s, La Roche Posay, Vichy, Biotherm, etc. | Supports **salon** **businesses** and **hair stylists** with specialized **hair products.** Brands include Kérastace, L’Oréal Paris, etc. |

## The Strategy

First, a document listing the selection criteria for each division will be created to guide the selection subjects. Influencers are added to the consideration set when their profile meets the conditions and constraints set by the business rules. Second, influencers will be given a tag or multiple tags with the name of the division their image fits best. This alignment ensures that the influencers' unique strengths and content styles harmonize with the division's core values and offerings, thereby enhancing brand coherence, audience engagement, and the overall impact of the influencer marketing campaign. Third, data will be compared to the division’s profiles and if matching score is adequate, the influencer’s information will be stored in a database representing the “Consideration set” of all influencers shortlisted.

To shortlist influencers into the consideration set, they must meet each variable’s condition. Variables can have structured and unstructured data. Structured data includes information that can be directly scraped and stored in the tables in numerical or standardized text format. Unstructured data is more complex and requires advanced calculations and analysis to gather.

## Table 2.1: Variables – Structured Data

|  |  |  |
| --- | --- | --- |
| Data Variables | Description | Data Type |
| Influencer Profile Data | Collect influencer username, name, gender, age range, country, number of followers, number of posts, and whether they have a verified account | String, Number |
| Content Performance | Total number of likes, comments, and shares. | Number |
| Engagement Growth Rate | Calculate engagement growth rate in the past year by comparing the number of followers in Year 1 to Year 0. | Number |
| Audience Analysis | Collect demographic and engagement data of followers. | String, Number |

The variables above can be sourced externally from the influencers’ social media page, through a social media API, an interface that allow developers to interact with and integrate social media platforms into their own applications, websites, or services. Among these variables, L’Oréal can add constraints that will act as the minimum condition for influencers to be shortlisted. For example, the minimum number of followers needs to be over 50k for L’Oréal Luxe and Consumer Products; the influencer must be at least 18 years old; the number of likes needs to be above 100k; or the Engagement growth rate needs to be over 50%, and so on.***Top of Form***

## Table 2.2: Variables – Unstructured Data

|  |  |  |
| --- | --- | --- |
| Data Variables | Description | Data Type |
| Content Analysis | Analyze content types, themes, style, and brand affinity. | String, Image, Video |
| Brand Alignment Level | Review past collaborations and values alignment. | String, Image, Video |
| Geographic Reach | Identify geographic location of influencer's audience. | String |
| Languages spoken | Find the languages the influencer speaks based on post language and interaction with followers | String |
| Influencer History | Record past collaborations with L'Oréal or competitors. | String |
| Competitor Analysis | Analyze influencer collaborations with competitors. | String |

Next, unstructured data is collected to reveal the influencers’ alignment with the L’Oréal brand on a deeper level. For content analysis and brand alignment level, image recognition will be deployed to identify the products the influencer uses, has reviewed, or has collaborated with. Moreover, text analysis can be performed by searching for keywords like “cosmetics, makeup, shampoo, serum, perfumes, *etc*.” and as well as specific brand names in the L’Oréal Group (labeled in orange in Table 1). This set of keywords is sourced internally through L’Oréal’s four portfolios. The more similar to L’Oréal’s use of keywords and brand names, the higher the matching score.

For the geographic reach and languages spoken, geotags can be used to determine where the content is being created and analytic tools and libraries like LangDetect or Google Cloud Natural API can be used to detect a text’s language.

Finally, influencer history and competitor analysis can be sourced externally by fetching names and tagged account names of L’Oréal’s competitors in the influencers’ page. Examples of close competitor brands are Estée Lauder, Shiseido, Pantene, Aveeno, *etc.* A list of competitors for each division can be sourced internally from L’Oréal’s existing database. Noteworthy, if an influencer is currently collaborating with a close competitor, L’Oréal should add this as a constraint to ignore those influencers.

# Data Preprocessing

* ***Bottom of Form***

***Data Transformation***

Structured data can be stored as numerical values directly, however, to integrate unstructured data in the data table, the unstructured data will be transformed into numerical values as a score. For example, the maximum value for Content Analysis and Brand Alignment level can be set at 5; the more similar the influencers’ content to L’Oréal, the higher the Brand Alignment level is out of 5. Similarity can be calculated as vectorized cosine similarity.

Languages and geographic reach variables can be set as binary. For instance, if the influencer speaks language x, the value is 1, if not, the value will be 0. For variables Influencer History and Competitor Analysis, they can be set at 5 as well. Having collaborated with other similar brands can be considered as a sign for high professionalism, experience, and high potential of reaching the target audience. Therefore, the more collaborations the influencer had, the higher these values should be, unless the influencer is committed to another brand; if keywords containing other competitor brand names and brand accounts are tagged, this value should be 0 and the influencer should not be considered for any L’Oréal brands.

***Data Cleaning and Data Integration***

Data will be uniformed in terms of format, data type, and collection process. Utilizing DataFrame enables systematic data arrangement, which in turn fosters good structure for higher productivity. Furthermore, implementing data validation against business rules and existing datasets enhances data quality, establishing a strong foundation for L’Oréal’s strategy-driven initiatives.

*Figure 1: Timeline of Data Collection and Preprocessing Plans*

***A diagram of a data analysis

Description automatically generated with medium confidence***

# Future Considerations

The next phase following data collection and preprocessing is data analysis, which can be executed through methods such as cosine similarity, clustered algorithm, and correlation and covariance. Once an optimal combination of characteristic is determined, predictive analysis can be implemented to reverse engineer the selection of influencers. L’Oréal would be looking at influencers who possess the optimal characteristics identified in through past collaborations.

However, to avoid falling into the trap of the innovator’s dilemma, L’Oréal should also take into consideration the fact that this automated recommendation system will help search similar influencers only, giving little space for influencers with creative but outlier characteristics to be included in the consideration set. Therefore, it is crucial to update the system to reflect new variable needs.

# Reference

Audrezet, A., De Kerviler, G., & Moulard, J. G. (2020). Authenticity under threat: When social media influencers need to go beyond self-presentation. *Journal of business research, 117,*557-569