**The Science of Tinder: How to Find Love Using Data**

7031ICT Assignment Part I | Project Proposal | Trimester 2, 2021

Javier Reina (Javi) | Student ID: 5288693 | javier.reina@griffithuni,edu.au, Signature

**1. Background and Aim**

Love in the 21st century looks a little bit different than it did in the past. With the advent of online dating, people are now able to connect with potential partners from all over the world. Of course, online dating is not perfect. There are still plenty of people who find love the old-fashioned way. But, for those who are willing to give it a try, this project might provide some insights.

This project starts as part of my data analytics journey. I remember when I started my master's degree I had no idea what all these tech words meant. I thought Python was just a snake, and Conda and Mamba were just some kind of venomous snake. I had no idea that there was such a thing as front-end or backend development, or that DevOps was a thing.

It's no secret that the tech world can be overwhelming. New concepts and jargon seem to spring up overnight, and it can be tough to keep up. However, I knew I had to find a way to enjoy the process so while doing the project I learn to apply data mining for my future work activities and thank God one day I got an inspiration when I was reading that outside of TikTok, dating apps appear to be the best social platforms for revenue generation, as many offer premium features to their customers.

I had no idea that you could increase your chances of success by paying for a dating app and it seems to be working out pretty well for Tinder, [revenue](https://www.businessofapps.com/data/tinder-statistics/) of $1.6 billion in 2021, a 17% increase on the year prior. It had 75 million monthly active users and 9.6 million subscribers in 2021. Well, it can be hard finding someone who's attracted to you let alone compatible with your personality. But if premium features are an option for success then why not try them out.

The aims of this project are to

(1) Predict the sentiment of the messages.

(2) Identify the language of the message

(3) Analyse how people are using the Tinder app

**2. Data Description**

With the aims of this project in mind. In order to have an interesting and successful project on dating apps, I need actual data and I wanted Tinder’s data since is the most popular app. Reason why I started looking for it in forums, [research](https://data-dive.com/tinder-statistics-insights-from-unique-dataset) and GitHub [repositories](https://github.com/tinderjs) without any success till I found an [article](https://towardsdatascience.com/i-analyzed-hundreds-of-users-tinder-data-including-messages-so-you-dont-have-to-14c6dc4a5fdd) about analyzed hundreds of user’s Tinder data and I tried to contact the person who wrote this article without success.

I knew it was going to be difficult to get the data but I decided to continue instead of changing for another project since I’m not one to give up easily. Through the last article I learn about Swipestats.io an interesting website that visualize your Tinder data so I contacted its owner Mr. Kristian Else Bø by mail and LinkedIn but failed to get a response from him. Therefore, deciding it was worth trying again through Instagram, I create an account since I don’t have one and this time with the great luck that he answered me and agreed to help me.

The dataset contains 1210 anonymous Tinder data profiles and is structured in a huge JSON file as follows:



There's no denying that the Tinder data set is rich with information. With over 1200 profiles, it offers a lot of possibilities for analysis - more than I could possibly hope to cover in one paragraph. So, I decided to focus on three specific areas: users, conversationsMeta and conversations.

In terms of users, I was interested in looking at things like gender, age and location. This allowed me to get a better understanding of the make-up of the sample as well as identifying any potential biases. When it came to conversationsMeta, I wanted to see how long people were chatting for on average as well as what topics were being discussed. Finally, in terms of conversations, I looked at the most common words used and the sentiment of the messages.

**3. Algorithms and Techniques**

3.1 Aim 1 with proposed algorithms

**4. Evaluation Measures**

**What measurements will you use to evaluate the results?** The same as the chosen algorithms, a short description of the measurements should be included. **NOTE:** The measurements should correspond to your targeted problem in point 3 and are suitable for the algorithms you choose in point 5. (10 points)

The accuracy of a classifier is given as the percentage of total correct predictions divided by the total number of instances.

If the accuracy of the classifier is considered acceptable, the classifier can be used to classify future data tuples for which the class label is not known.

Precision describes how accurate or precise our data mining model is. Out of those cases predicted positive, how many of them are actually positive.?

Precision is also called a measure of exactness or quality, or positive predictive value.

 The similarity is commonly defined in terms of how “close” the objects are in space, based on a distance function (Manhattan, Euclidean, etc).

**5. References**

Appendix: Data Investigation Report

Introduction

1. Data Exploration

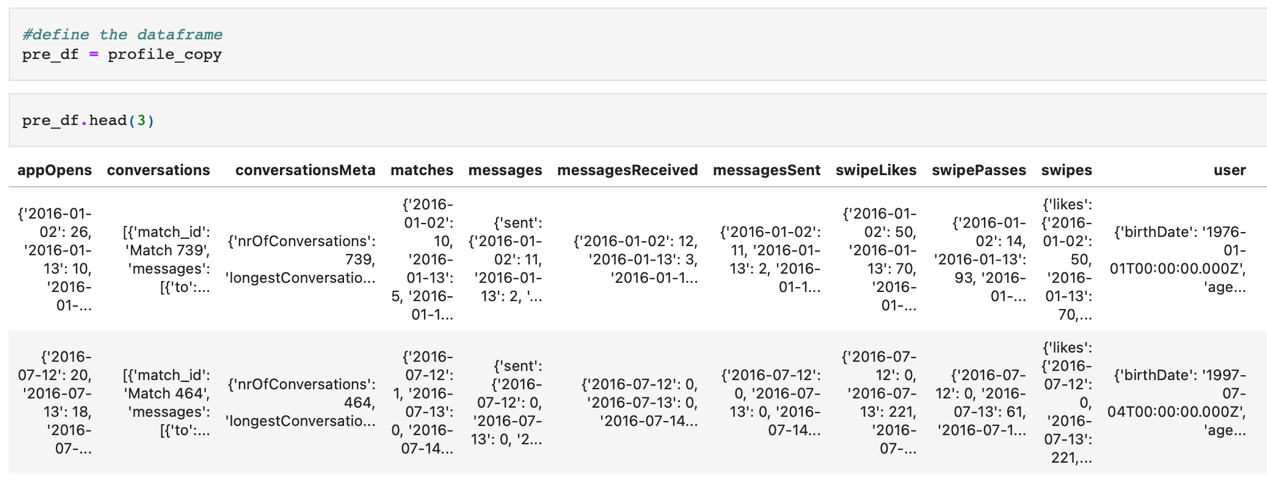
**TEXTO**



Agregar ARCHIVO DE JSON

1. Data Visualisation

**TEXTO**



TEXTO CONVERSACIONES

|  |
| --- |
| Hello again, so now that we matched again, what exactly are you looking for here? |
| Most of the girls on this app scare me, but for some reason you seem chill. Maybe it&rsquo;s the fact you like to smile |
| i mean you seem pretty cute |
| So tell me about yourself, favorite hobby? Favorite food? College major? |
| I brought you some flowers |
| Who give aruban girls the right to steal my heart |
| Now that we matched, wanna run away, get married, and be my next wife? |

1. Data Pre-Processing

Export all the conversation so we can use it in a separate file to train our model



**TEXTO**

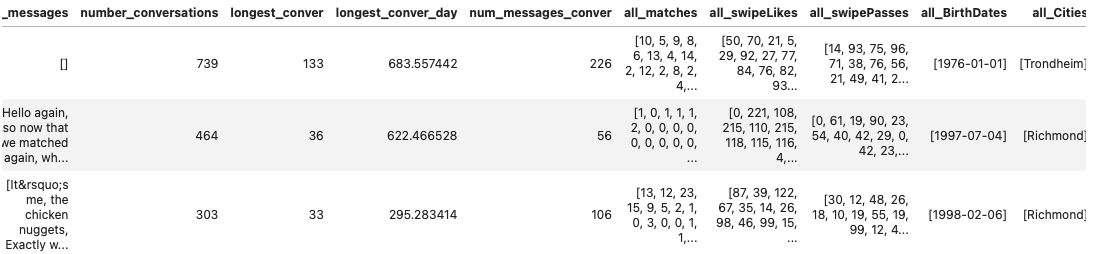
****

**TEXTO**

****

**TEXTO**

****

****

**Self-contribution:**

**References:**

<https://www.businessofapps.com/data/tinder-statistics/>