THE LOVE FINDERS

1. **Project Description**

*You are single and are looking for love.*

In XXI century, in the times of COVID-19 global pandemic and frequent lockdowns, chances are you would consider using a dating site / app. Which one will you go for? Which app has most users? Which has the best rating? What attributes people look for in their “second – half”? Does your age matter? How many people of your age are going to use the app? How big chances do you have to be successful?   
  
*The lockdown is off, and you can finally meet!*  
How to make a good impression and secure your second date? Should you even try to do that? Maybe being sincere is the best strategy? What traits of yours will bring you success / failure?

1. **The choice of subject**

Love is a natural human need. This is subject concerning us all. In times of COVID-19 a lot of people felt isolated. We are curious how people who are looking for love could succeed in current “love market”.

1. **Benefits of choosing this subject**

By exploring the most popular dating apps and websites, we will know where to look for love. Analysing the most in demand traits will help us to understand how to flourish in our love life.

1. **Datasets**

We used two datasets from Kaggle and scrapped one website:

* Speed Dating Analysis: <https://www.kaggle.com/datasets/somesh24/speeddating>
* Dating Apps Reviews 2017-2022: <https://www.kaggle.com/datasets/sidharthkriplani/datingappreviews>
* 24 Dating Sites With the Most Users: <https://www.datingadvice.com/online-dating/dating-sites-with-the-most-users>

1. **Technologies**

* Excel & CSV – as datasets come in this format
* Python – to clean scrapped data and connect to the database
* Pandas library – to clean data and create data frames, so our data can be uploaded to database
* Beautiful Soap library – to scrap the website
* Mongo Database - was chosen as the data is not related, so there is no reason to use relational database such as e.g., PostgreSQL
* Ms Word – to create and export this report
* Git & GitHub to collaborate with the team

1. **ETL Diagram**  
     
   A picture containing text

   Description automatically generated
2. **Data Transformation - Steps**  
   Drop rows with empty cells  
   Reduce columns and rename them  
   Remove dates that were not relevant (everything before 2020)  
   Scrap the website
3. **Challenges**  
   Very little amount of time to complete the project  
   Technical problems – with Internet, different time zones
4. **Division of tasks by group members - who did what in the group:**

* *Rita Starzyk:*  
  Proposal

Web Scrapping  
Uploading data to the Database  
Read.me file in GitHub  
Combining all files together into one big project

* *Kouame Kwasi*   
  Data cleaning  
  Created connection with MongoDB using pymongo  
  Uploading data to the MongoDB
* *Daniela Shae-Bebeyi*  
  Uploading data to the Database  
  Uploading data to the Database
* *Motasim Nasir*   
  Finding the data  
  Project extension and base for the portfolio