#### SPRINT 4 ####

# Requirements:

Data vendor🡪 SPOTIFY

Extract 🡪 Spotify API (Spotify Developers, user access) loads data to the pipeline.

Transform 🡪 pipeline based on Airflow 2 with 6 DAGs (find location user, find the exact song played, find the local time respect Greenwich time, users’ interaction while song is played, type of browser used or the Spotify App, temperature of the location while playing song from BBC weather channel).

Load 🡪 Kibana. But before use Elasticsearch to refine all the transformed and processed information in the pipeline. Use database to store all the output from the pipeline and use PostgreSQL to retrieve needed information.

Machine learning and deep learning techniques whit transformed data could be applied to find patterns and behaviours from users. Discovery mode of the project!

Customer and data consumer 🡪 Spotify to offer related songs to user. Roehampton university research tea to find if weather affects the desire of listening music, and the type of music if previous answer is positive. Expectations from machine learning and deep learning about users; and this can be useful for authorities like government, doctors, and psychologists. Tableau or Power BI is the destination of all findings and discoveries from pre-process data to show to Nontechnical users (children, NO-STEM users, old people for example).

Advertisement and marketing business can buy the output and outcome from the Load and Transform stages to offer relevant advertisement to the listener, even if the listener is in the same building located.

All the project implemented on Google Cloud with 50 USD budget. Try to use free resources and similar free datasets or data engineering projects.

# Management:

6 DAGs in python programming language. Please refer to GitHub.

Teamwork kanban board on GitHub link (readme file).

# Outline of the project overall:

Overall, the project has a length of 5 months (January 2023 to May 2023). Agile style is used in each step and in each progress. The product owner is the same as Scrum Master and Team Leader. Kanban board is maintained by the same person.

Some useful references for the DATA ENGINEERING project

<https://github.com/sidharth1805/Spotify_etl>, <https://github.com/culpgrant/Spotify_ETL>, <https://github.com/nekoemperor/spotify-ETL-job>

<https://youtu.be/i25ttd32-eo>, <https://youtu.be/jvad7eF0qqY>

Spanish language and like first link from YouTube: <https://youtu.be/eg8t2-E69ew>

About Spotify and Spotify Engineers: <https://youtu.be/L14enOBwi1A>

Notes and considerations: Netflix is like Spotify but uses video, not only music. This project can be applied to Netflix data producer.

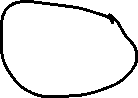
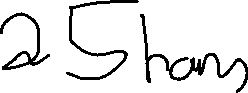
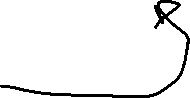
A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

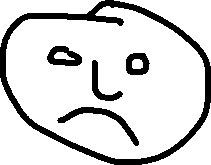
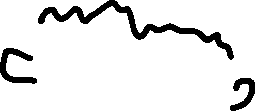
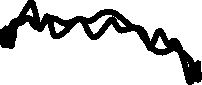


A screenshot of a computer

Description automatically generated

A computer screen shot of a message

Description automatically generated with low confidence



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A picture containing text, screenshot, software, computer icon

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

END. Tuesday, 25 April 2023