Music Recommender

Marko Znamenaček

Business Mathematics and Economics

Zagreb School of Economics and Management

Zagreb, Croatia

mznamen@student.zsem.hr

**Abstract - The objective of this music recommender project in Python is to create a recommendation system that suggests music to users based on their listening history. The project utilises a collaborative filtering algorithm that analyses user behaviour and generates recommendations based on similar listening patterns from other users.**

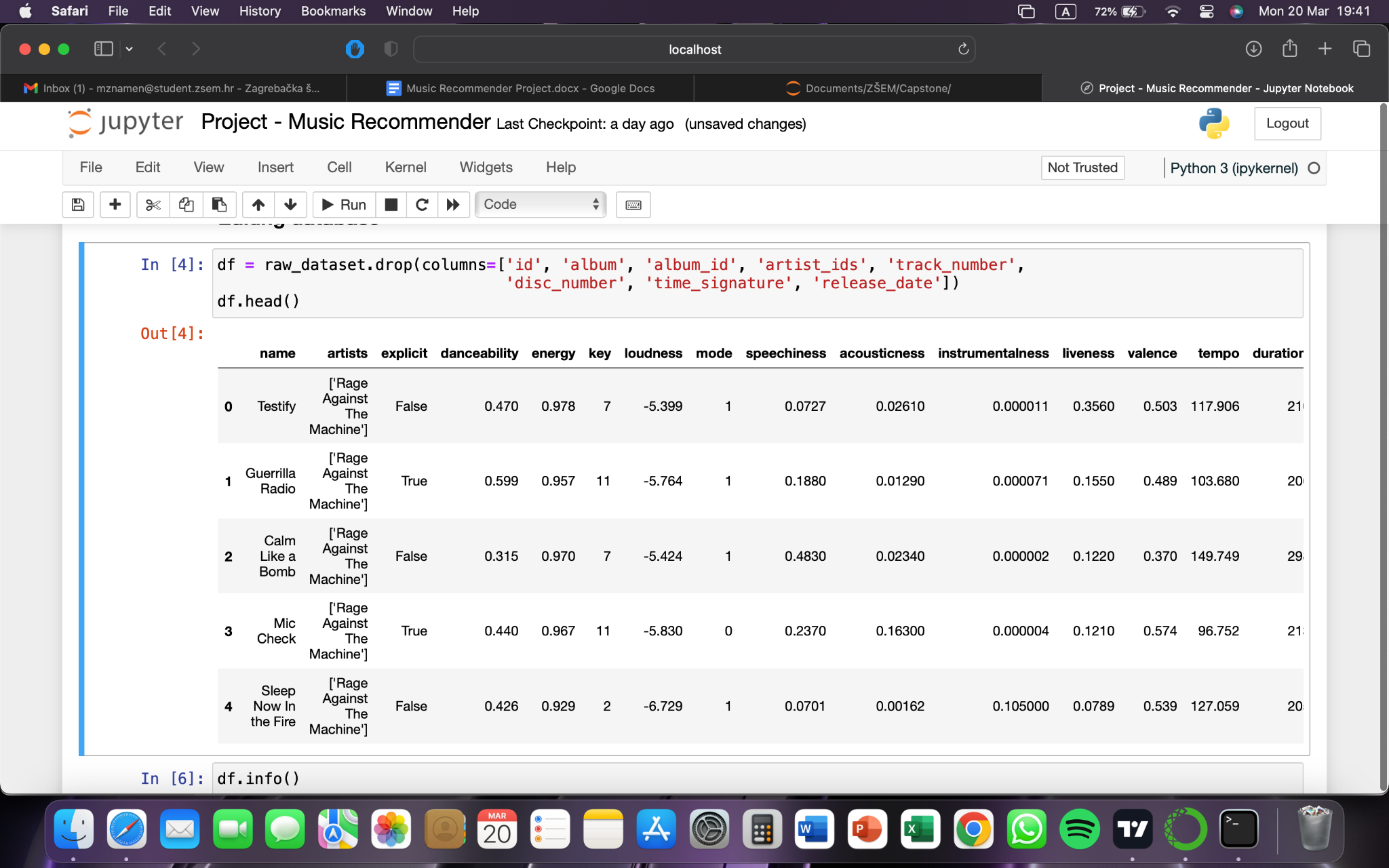
**Keywords - Music recommender, Python, User behaviour, Listening patterns, Dataset, Characteristics, Similar songs, Metrics, Clustering, Elbow method, K-means clustering, Matplotlib, Pandas, NumPy**

1. Introduction
2. *Problem*

How many times have you recommended a song from YouTube, Spotify or Deezer based on what you are listening to, and the recommendation was bad? In many cases it isn’t what you were hoping for. In this project I will try to build a reliable music recommendation system in python using a Spotify dataset with 12 million songs and their characteristics in order to find similar songs comparing these metrics.

1. *Dataset*

Dataset is called tracks\_features.csv, and it is available on Kaggle. It contains over 1.2 million songs from Spotify and their characteristics.

Fig. 1. Filtered dataset I will be using

1. Goal and hypothesis of the project
2. *Goal*

The problem that the music recommender project aims to address is the overwhelming amount of music available to listeners. With so many artists, genres, and songs to choose from, it can be difficult for users to find the music that they will enjoy. This is where a music recommendation system can help.

By analysing a user's preferences, a recommendation system can suggest new songs or artists that the user may be interested in. This not only helps users discover new music that they will love, but also helps artists and record labels promote their music to a wider audience.

1. *Hypothesis*

Users who receive personalised music recommendations based on their listening habits and preferences will be more likely to engage with the music streaming platform by listening to more songs, creating more playlists, and staying subscribed for a longer period of time.

1. Review of previous research

There was a similar project I found on GitHub which I used as a guideline [2]. That project was using a much smaller dataset than myself and data wasn’t edited. I think a larger base is better and will provide a better recommendation. Also, more metrics are included in my dataset which are more correlated between songs.

1. Materials, methodology and research plan
2. *Approach*

First, I must find a dataset that can be used, and that is reliable. Then I will cluster my data as best as I can, using metrics given in the dataset. Then I will find a song and ask the program to recommend a couple of similar songs.

1. *Collecting the data*

This part is not applicable for this project as data is already provided by Spotify. I will just manoeuvre with the data a bit to try to get more significant results.

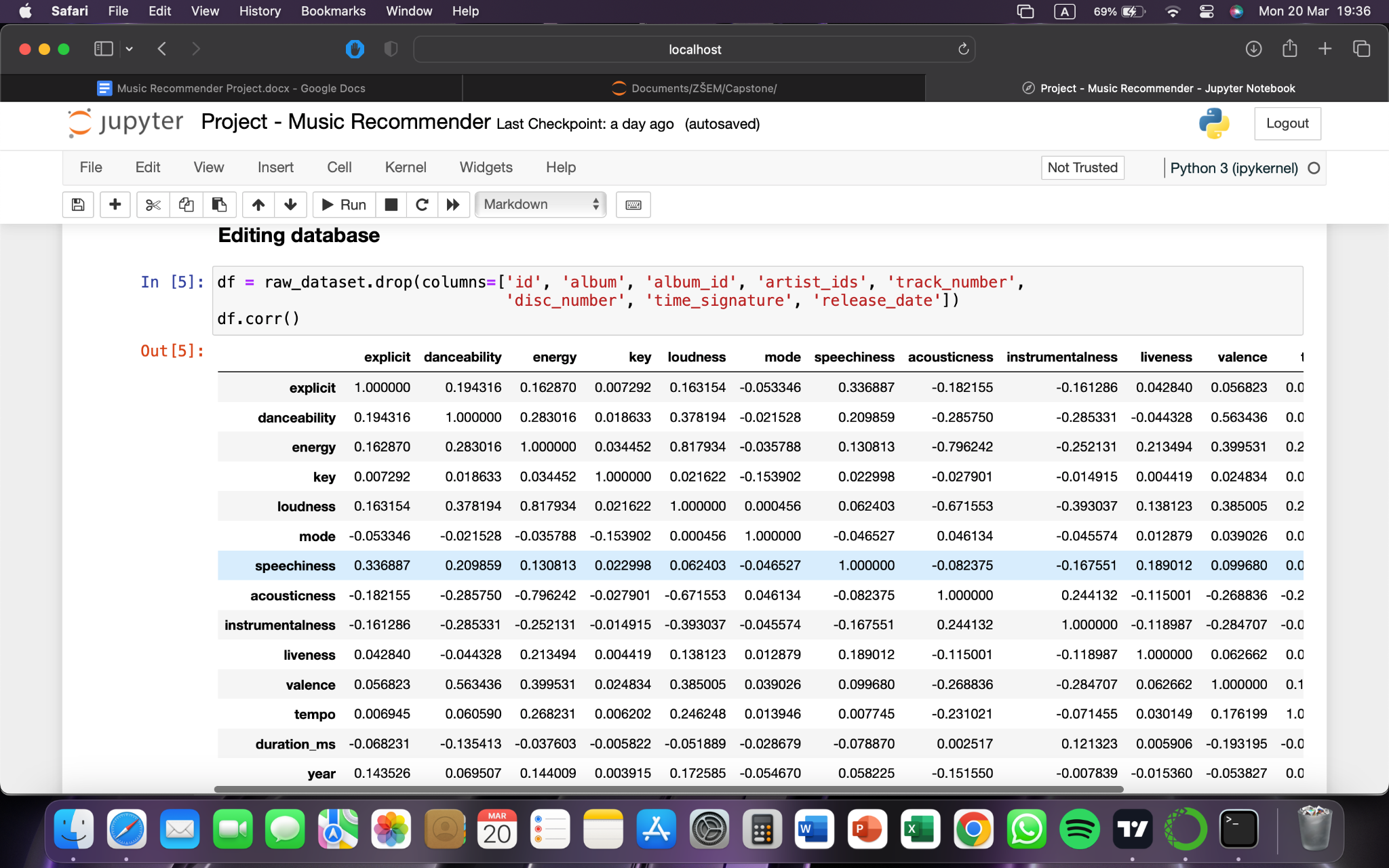


Fig. 2. Correlation between metrics in my dataset

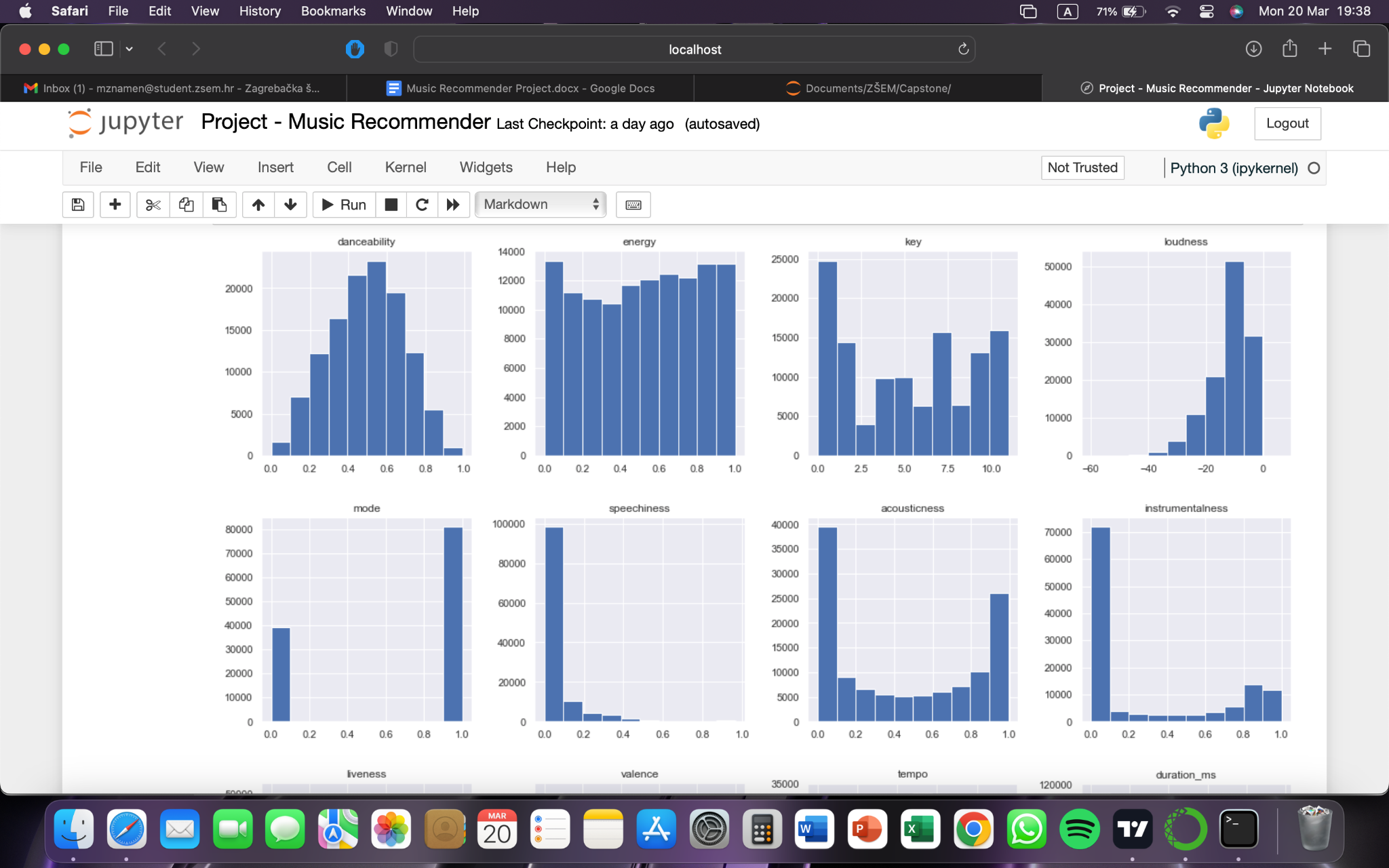


Fig. 3. Histogram of metrics in my dataset

1. *Methods, algorithms, techniques and tools*

I will cluster my data based on metrics in the dataset to make predicting easier if the data is grouped. I will use the Elbow method. The elbow method is a graphical representation of finding the optimal 'K' in a K-means clustering. It works by finding WCSS (Within-Cluster Sum of Square) i.e., the sum of the square distance between points in a cluster and the cluster centroid [1].

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms. Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes.

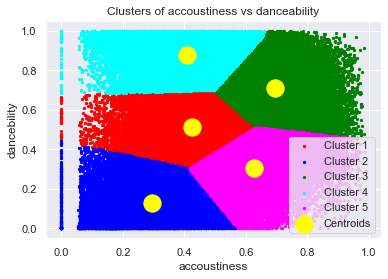
A cluster refers to a collection of data points aggregated together because of certain similarities. You’ll define a target number k, which refers to the number of centroids you need in the dataset.

A centroid is the imaginary or real location representing the centre of the cluster. Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares. In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The ‘means’ in the K-means refers to averaging of the data; that is, finding the centroid.

To process the learning data, the K-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimise the positions of the centroids. It halts creating and optimising clusters when either:

* The centroids have stabilised — there is no change in their values because the clustering has been successful.
* The defined number of iterations has been achieved. [2]

I will use python, and python packages Matplotlib, Pandas, Seaborn and NumPy.

Fig. 4. Cluster; accousticness vs danceability

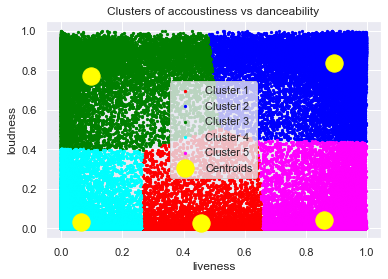


Fig. 5. Cluster; loudness vs liveness

1. *Evaluating the success of my project*

I would value these suggestions by entering songs I like, and then listening to recommendations from the program. If those recommendations are good, I would take this as a success, and ask more people to try it for themselves. Unfortunately, I wouldn’t think this is really measurable as preferences and opinions are much different, and it is hard to predict something like this just from some metrics from Spotify. Therefore, many suggestions available today are not as good as people expect.

1. Expected results of the proposed project

I expect to get a valid recommendation for most of the songs I type in the search. I would like to use this for myself and see if I really like the recommendations given by the program. Overall, the expected results of the music recommender are a more personalised and satisfying music experience, increased engagement with the music streaming platform, and higher revenue for the music industry.

1. My Results of the project

Unfortunately, I am not happy with the results that I got. I didn’t like the music that was recommended to me. You can try for yourself using my GitHub link [3].

My conclusion is that my model was too simple. It is very hard to get a significant recommendation from one song. Also, another point might be that it isn’t enough to just to look at the song characteristics; we probably have to take into account along with more songs, genres, years, languages, and artists people are listening to. This would have to be a multivariable forecast model, probably much more complicated than this one. Spotify probably uses something taking into account hundreds of songs you are listening to, genres, artists and much more of the data you are providing them with. It is very hard to predict what somebody would like based just on song characteristics.

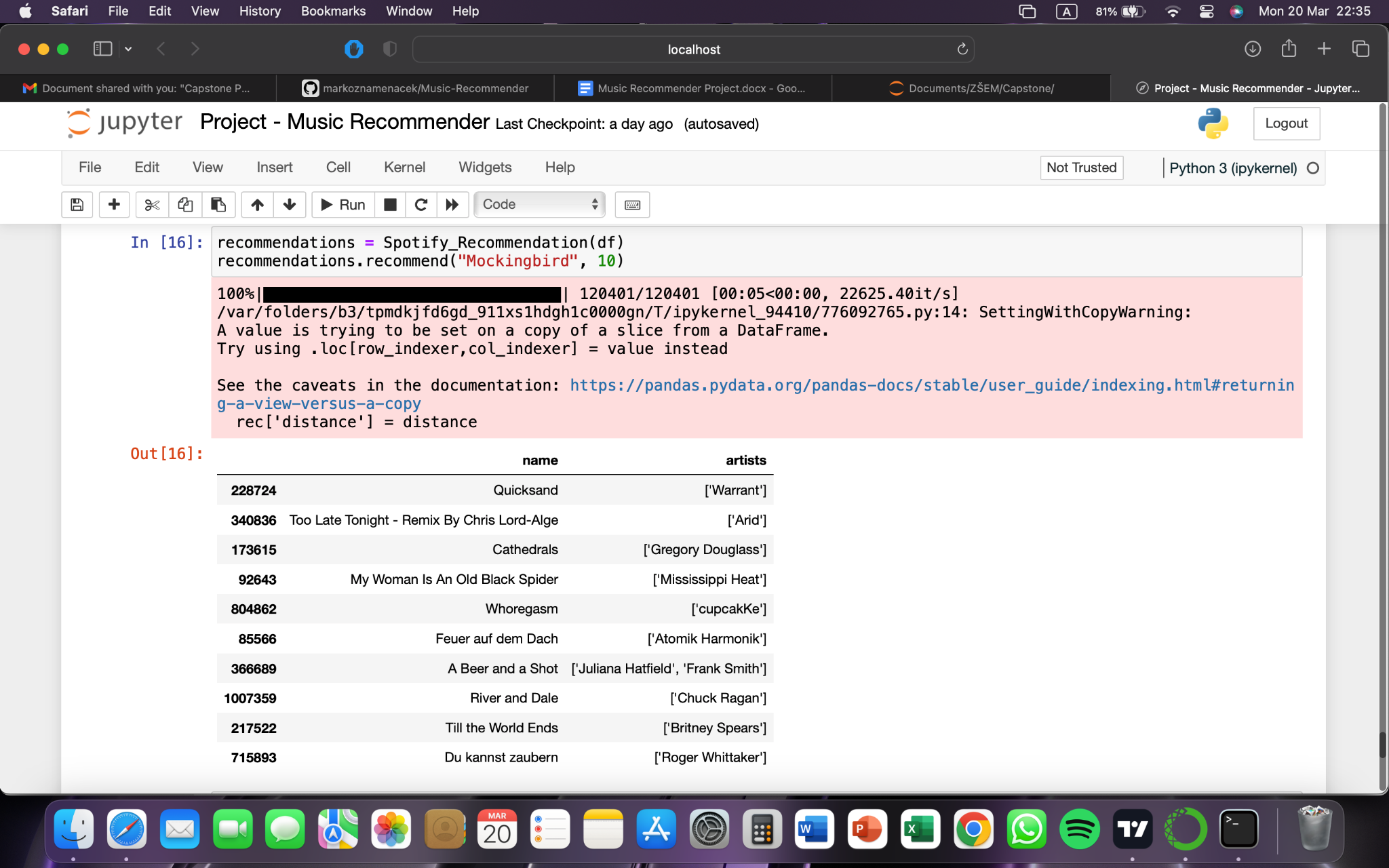


Fig. 5. Recommendation for a song “*Mockingbird”*

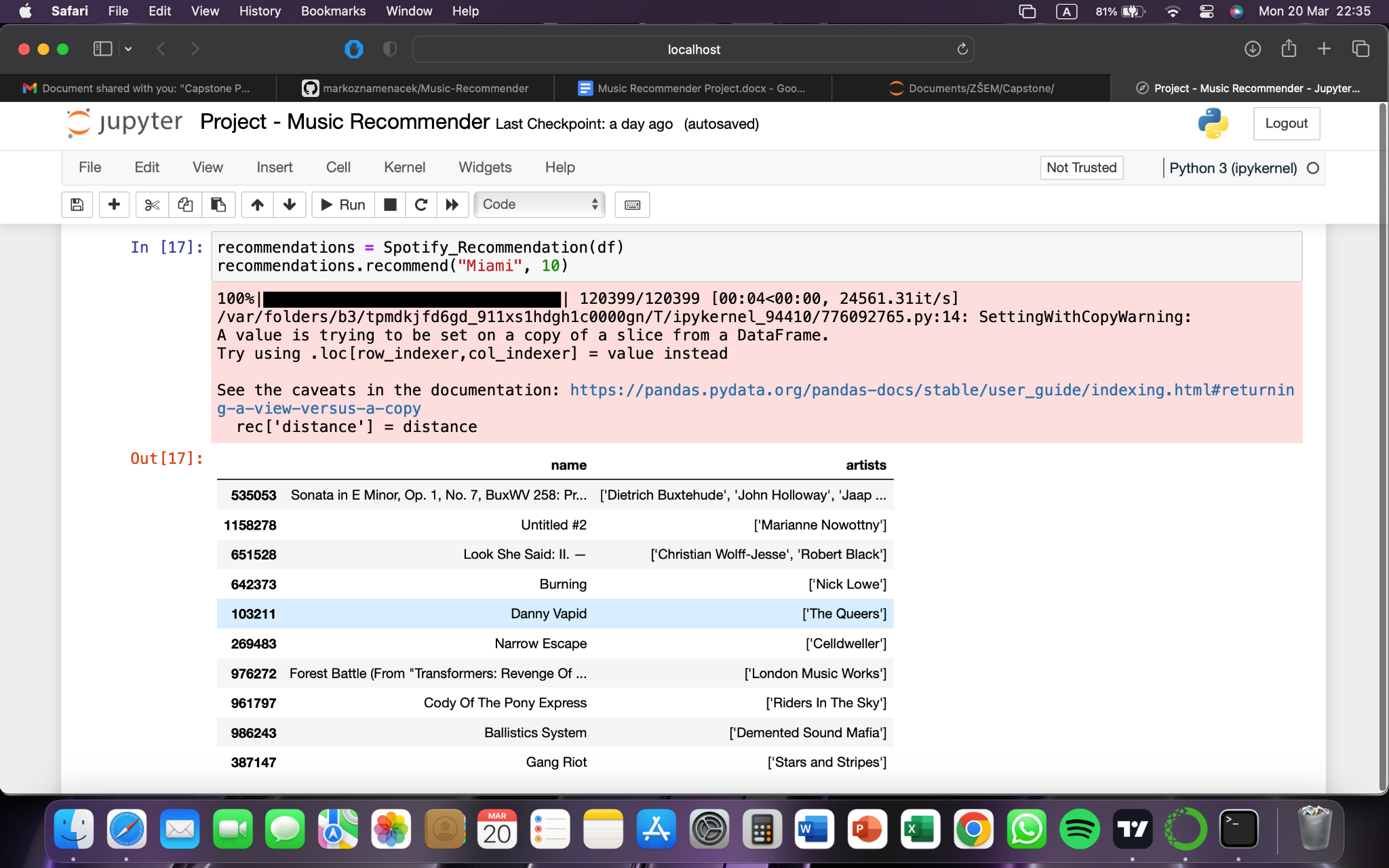


Fig. 6. Recommendation for a song “*Miami”*

Songs that were recommended to me were not really similar and didn’t fit the genre and my likings. They probably were similar in characteristics from Spotify. One positive to take from this, is that the songs which were recommended to me were completely new to me and I haven’t heard about any of the 20 recommendations. For somebody that is looking for something completely new and is willing to listen to something a little different, and be patient with recommendations that he receives, this is probably a good recommendation system.

1. Literature

Kaggle. (n.d.). Spotify 12M songs. Retrieved March 19, 2023, from https://www.kaggle.com/datasets/rodolfofigueroa/spotify-12m-songs?resource=download

[1] Tomar, A. (2022, November 17). Stop using elbow method in k-means clustering, instead, use this! Towards Data Science. https://towardsdatascience.com/elbow-method-is-not-sufficient-to-find-best-k-in-k-means-clustering-fc820da0631d

[2] Bhattacharya, A. (2019, February 19). Understanding k-means clustering in machine learning. Towards Data Science. <https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>

[3] Ravi Prakash. (n.d.). MusicRecommendationSystem. GitHub. Retrieved March 19, 2023, from https://github.com/raviprakash11/MusicRecommendationSystem

Mavani, V. (2020, October 29). Music recommendation system using Spotify dataset. Kaggle. https://www.kaggle.com/code/vatsalmavani/music-recommendation-system-using-spotify-dataset

The Clever Programmer. (2021, March 3). Spotify recommendation system with machine learning. https://thecleverprogrammer.com/2021/03/03/spotify-recommendation-system-with-machine-learning/

[4] Marko Znamenacek (2023), Music-Recommender, Available on GitHub link: <https://github.com/markoznamenacek/Music-Recommender>