**PROGRAMAÇÃO E ALGORITMOS 2**

**Comunicação e Design Multimédia**

Alunas: Cláudia Teixeira, Cristiana Costa e Inês Costa

Professor João Orvalho

2021/2022

**Informações adicionais do trabalho**

Para realizar este trabalho decidimos recolher dados relativos às músicas do género K-pop dos anos de 2020 e 2021. Estes dados foram retirados da plataforma Kaggle, um repositório de dados online, a partir de um download de ficheiros CSV com o nome “Hit songs in K-Pop from the 90s to the latest year using Apple Music and Spotify”. Estes dados começaram a ser recolhidos pelo criador destes ficheiros desde os anos 90 porque tinha um interesse pelo género musical que é o K-pop e uma curiosidade pela ciência de dados.

Os API’s usados para a recolha destes dados, segundo o criador dos ficheiros, foram o Spotify por ser de fácil acesso e o Apple Music pois foi o criador da playlist de K-pop anual. Tendo em conta que o Spotify e a Apple Music são aplicações muito usadas mundialmente que guardam os dados de cada usuário, apesar de não termos confirmação, consideramos estes dados confiáveis.

Ao retirar os dados não foi necessário eliminar nenhuma coluna, linha ou valor da tabela de dados antes de iniciar o trabalho e também não tivemos problemas com valores “ausentes” ou duplicados, porém, enquanto o realizávamos percebemos que a coluna do tempo de cada música encontrava-se em milissegundos e decidimos acrescentar uma outra coluna com o tempo convertido em minutos de forma a melhorar a compreensão dos dados.

Com base nos dados que tínhamos, criamos 11 perguntas relacionadas, principalmente, com as colunas que consideramos as mais importantes: ranking, título da música, nome do artista e o tempo da música. Para responder às perguntas que colocamos usamos funções como: **sort\_values** (ordena os valores de uma coluna específica em ordem ascendente ou descendente), **nlargest** (obtém os valores mais altos de uma série de dados), **nsmallest** (obtém os valores mais baixos de uma série de dados) e **value\_counts().idxmax()** (retorna o valor máximo).