EKANS

Project Proposal - Report

K-pop Idols Analysis

Team Members

191180005 Selin Cansu Akbaş

191180006 Mert Akgüç

191180007 Nursena Akgüneş

191180009 Furkan Onur Akman

The specific topic that we plan to work on is K-pop. K-pop is a term that is often used on the internet, and there is quite a popular fan following around the world. K-pop idols have a dataset. And we will search this to answer several general questions. We will search some data, based on the existing features. We chosed K-pop because it became a global phenomenon. So, we thought it would be cool to analyze K-pop using machine learning to explore interesting insights. The dataset will be about K-pop idols profile from 1992 to 2020. The main questions that will be driving our project are about K-pop idol analysis. Such as idols gender or age. Also group analysis in K-pop and name analysis of this groups. We examine about their Spotify or Youtube analysis. If we talk about the exploratory questions that we plan to look for answers to, we can say that they will be as follows. How is the comparison between male and female idols number? Did the K-pop idols only come from South Korea or are there idols from other countries, nationalities as well? How many groups in K-pop? Are there any idols who have more than one group? How many idols have more than one group? K-pop idols use their original name as stage name? How many idols use stage name? How is the K-pop idols age generation?

We had to do a little bit of googling to find the dataset. After some searching, we came across this website. Here is the link where we can get the dataset: https://dbkpop.com/. Database of Kpop-Idols (dbkpop) is a dedicated database of most kpop idols. Additionally the website gives important information for our questions, such as how many members are in a group, what age are the members when they debut, how many members are in a group, and how many active groups there are currently. There are also other values that may be interesting such as what country idols were born and how many idols were in a former group.

We will also use Spotify Web API to get data. Spotify has two measures of prevalence. Followers and popularity. So, we can question who are the most popular K-pop artists? Or how does popularity change? Also this K-Pop data analysis using data provided by Spotify API which could be used to answer question like: How many K-pop albums released every year?

We will use supervised learning of machine learning techniques. It has regression, decision tree, random forest. We will be using these techniques for reading the data, cleaning the data, filtering, and moving columns or this kind of thing. If we talk about data preparation we will import needed packages like pandas, numpy, seaborn and matplotlib. Datasets also will check data types mostly object and missing values. We will be using these techniques for reading the data, cleaning the data, merging, filtering, and moving columns or this kind of thing.