Podcast listeners client base demographic and topic segmentation using advanced Analytics

By Omar Hassan

Abstract

Motivation: Podcast space is a growing media space. Publishers and advertising agencies have been having a challenge in identifying different segments of listeners within this space. This might be due to the fact the industry is new and data around this industry is scattered and challenging to put together to make a picture of its audience.

Result: Here we present a research study on podcast listeners data using advanced analytical methods to identify listeners segments. Take Netflix as an example, they are able to segment their user base in a way they can recommend existing content to them or identify what kind of new content to invest in and produce. For us here we attempt to provide something similar. We will provide meaningful insights to help publishers and advertising agencies reach their target audience.

Research question

What are the various podcast listeners characteristic, trends and demographics segments in the Canadian podcast space?

Research objective

The objective of this research study is to apply advanced analytics on podcasts and their listeners data to produce valuable insights on different audience segments.

Data sources

We are going to be obtaining data on more than 1,500 podcast listeners from 3rd annual Canadian Podcast study. This data contain different publishers, genre and listener's data such their demographics, their preferred platforms. Potentially we might try to obtain more data about the podcast industry.

Methodology

Here we are going to be trying to answer two questions. First question, can we segment the podcast listeners? For that, our approach is going to be exploratory and discovery. We are going to use Python as the programing language. We are going to use Pandas python library to import, view, manipulate data. We are going to use clustering algorithm to identify common traits among listeners. We can use decision tree algorithm to identify various grouping of characteristics between listeners and the podcast or topics they listen to. We are going to be using python libraries to visualize our results. As for the second question If time permits, we can leverage association algorithms to create a model that can recommend other podcast topics to a podcast user based on their favorite podcasts, how long they listened to a topic, did they listen to the end.

Time Constraints and Limitations

Normally we would go out and gather more data about Podcast industry. Due to the time constraints of having less than 6 weeks we will be leveraging the data provided by Audience Insights. We are going to be focusing on identifying groupings of podcast listeners based on their demographics, visualize the different segmentations of podcast listeners.

Time lines

Week 1 & 2: Data collection and data exploring

Week 3 & 4: data cleaning & model construction

Week 5 & 6: submission of Sprint 2 on git hub.

Week 7 & 8: final project and submission.