Case Study: Data Analyst

Overview

You are a Data Analyst for Wine Corp. Wine Corp is a leading retailer for wines in the United States. The company has long depended on brick and mortar stores but is exploring avenues to grow its' online revenue. Specifically, Wine Corp is creating a subscription service for its wines. To facilitate adoption, and differentiate themselves in the market, they want to develop a recommender system to recommend wines based on other wines that a customer has enjoyed. You have been tasked with prototyping a recommender system.

Part 1: We have provided an adapted version of a publicly available dataset, scraped from WineEnthusiast. Perform Exploratory Data Analysis (EDA) on this data and summarize your observations.

- You may use the online dictionary and any information on the Kaggle page, but please do not use or download the online dataset. We have specifically modified the data for this challenge.
- Communicate key observations with appropriate data visualizations.

Part 2a: The following articles describe various methods to calculate similarity of two strings.

- https://www.freecodecamp.org/news/how-to-process-textual-data-using-tf-idf-in-python-cd2bbc0a94a3/
- https://towardsdatascience.com/overview-of-text-similarity-metrics-3397c4601f50

Using the information in these tutorials, prototype a function to select the top 5 most similar wines, given a specific wine indicated below that the client likes. Show your work and please list any assumptions you make. You may rely on other articles as well. You may also use any attributes within the data to facilitate recommendations. Hint: if you follow the approach outlined in these articles, you will need to convert a sparse matrix to an array using the command sparsematrixname.toarray() where sparsematrixname is the sparse matrix name. It also might be helpful to convert the similarity output back to a dataframe.

- Client A likes wine # 38 (DanCin 2012 Septette Pinot Noir (Southern Oregon))

Part 2b: Give the top 5 most similar wines to this other wine:

- Client B likes wine # 289 (Zeal 2011 Sauvignon Blanc (Marlborough))

Part 3:

Client A is on a tight budget and does not want to spend more than $20. She also is interested in expanding her tastes and would like recommendations that do not include Pino Noirs. Give the top 5 recommendations that meet these criteria.

Part 4: Assess the performance of your recommender. Is it providing reasonable recommendations? Why or why not?