Capstone Project

Summer 2022

Affordable Product Recommendation of Luxury Skincare Products Using NLP's BERT Technique

- Name: Nazia Noor
- Course: DATA 606
- Instructor: Dr. Chaojie Wang
- Date: 07/27/2022

UMBC

Content

- 01 Introduction
- 02 Importance & Motivation
- 03 Objective
- 04 End to end process
- 05 Limitation/Future work
- 06 Summary

01 Introduction

- Skincare industry is booming.
- Highes price means good product- is a MYTH.
- Higher price tag does not equal to higher-quality ingredients and effectiveness..

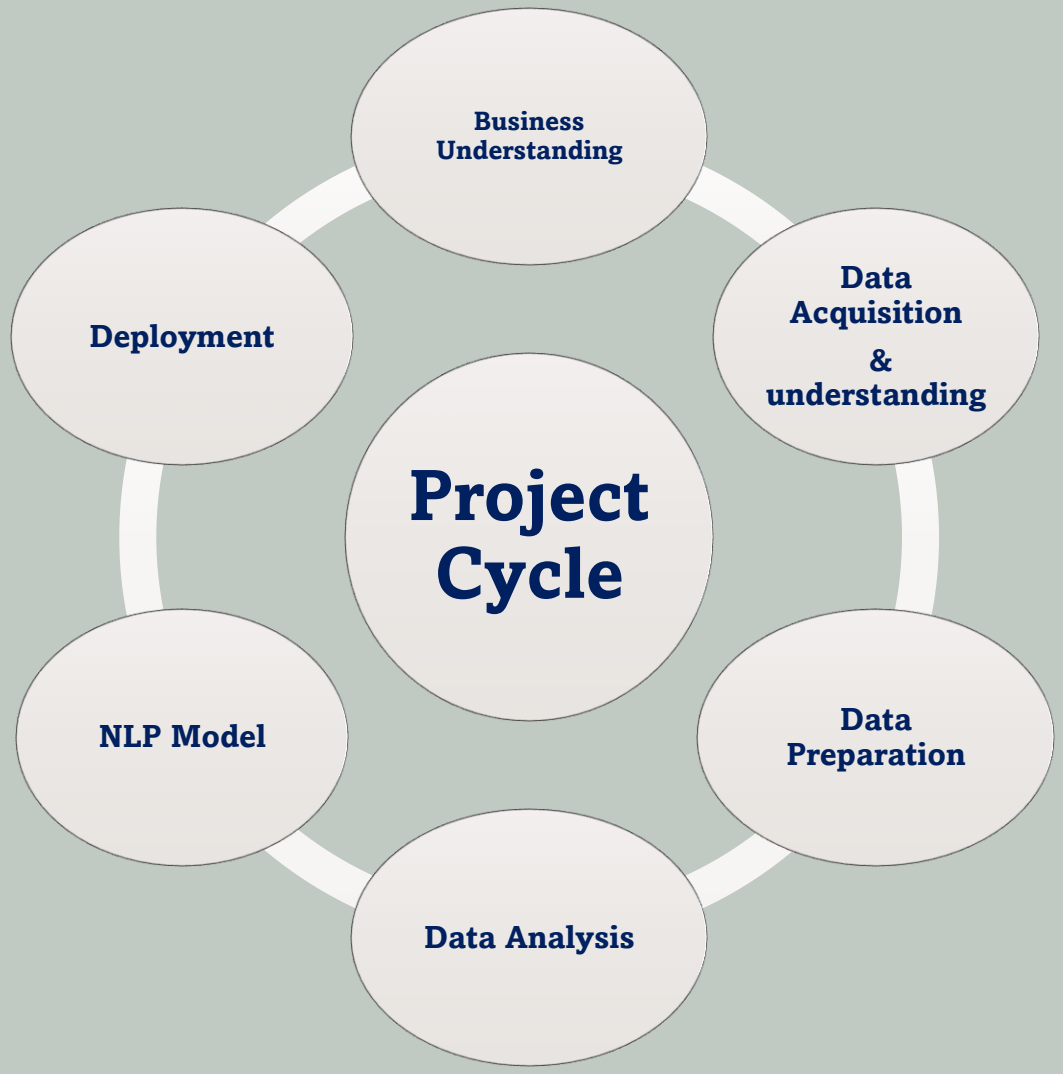
02 Importance and motivation

- Most consumers cannot afford luxury products
- A large part of luxury product pricing includes brand value and marketing
- It is very much possible to get the dupe of the luxury skincare product at a fraction of price
- An affordable product can have identical ingredients and benefits as a high-end product

04 Objective

1. Find the affordable option for luxury skincare products.(solve through NLP technique)
2. Check the correlation between price and product amount
3. Find the most expensive and least expensive products(from whole dataset as well as from each category)
4. Individual price, product size, product amount, brand analysis

05 End to End Process



Let’s have a look at the dataset, data cleaning and exploratory data analysis notebook

Dataset link: nazia_data606/dataset at master · nazia-noor/nazia_data606 (github.com)
Data cleaning notebook link: nazia_data606/data_cleaning_processing at master · nazia-noor/nazia_data606 (github.com)
EDA notebook link: nazia_data606/EDA at master · nazia-noor/nazia_data606 (github.com)

The BERT Model

BERT (Bidirectional Encoder Representations from Transformers) is a well-known technique developed by google to extract characteristics from text data. It is a machine learning framework for natural language processing that is open source (NLP).

There will be three major steps I performed to apply the BERT Technique:

1. BERT Embedding
2. Finding Cosine Similarity
3. Testing

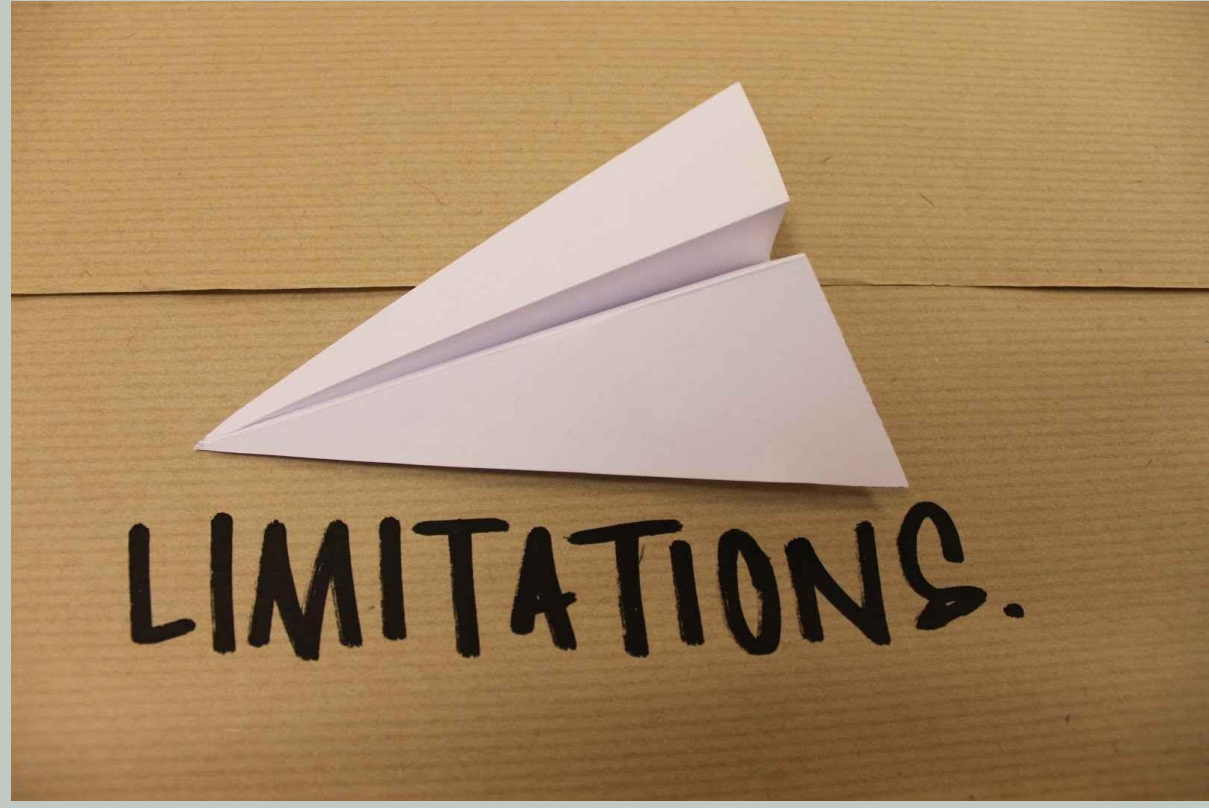
Let's have a look at the modeling notebook. Link: nazia_data606/modeling at master · nazia-noor/nazia_data606 (github.com)

06 Summary

Cream Category	**It was able to generate nearly 90% similarity** **Luxury product:** "111skin celestial black diamond cream 50ml"($995) **Affordable option:** "bella aurora bella multi-perfection day cream combination-oily skin 50ml', ($ 41.4) with approximately 87% similarity
Cleanser Category	**It was able to generate more than 90% similarity** **Luxury product:** "omorovicza thermal cleansing balm supersize - 100ml " ($180) **Affordable option:** dcl skincare active mattifying cleanser 125ml' ($34.5) with around 91% similarity
Serum Category	**It was able to generate more than 90% similarity.** **Luxury product:** "mila moursi lifting serum 1 fl. Oz" ($440) **Affordable option:** "is clinical poly-vitamin serum 15ml", $ 72.0 with around 92% similarity

07 Future work

- Improve model to Apply BERT model on the whole dataset rather than category wise.
- Develop model to correctly identify category
- Identify a better approach for classifying product size




Thank You