Eyewear Product Analysis

Background

VSP Vision owns Visionworks (https://www.visionworks.com/) which is a leading provider of eye care services in the U.S., sitting at the intersection of healthcare and retail. Visionworks brings healthcare expertise backed with a network of Optometrists and technicians as well as retail associates ready to fulfill vision prescription needs including glasses, sunglasses, and contact lenses. The patient's experience starts with scheduling an eye exam and shopping in-store or on-line (www.eyeconic.com) for eyewear products. The eyewear frames are shown in-store on boards organized by shape, price points, and gender. The stores also display specific marketing for brands, such as Nike and Calvin Klein. Our product managers are responsible for analyzing data insights using advanced models to determine the frame assortment to carry in our store locations.

The Data

Kathleen Lovett and the team at VSP Vision have provided a sample of frame product styles with a variety of attributes related to product description which can be used for product sales predictions available in Canvas as `CMU DSPM 95851_dataset - Eyewear Products Final.zip`. The file consists of rows and columns of text providing additional details in product styles descriptions with designated locations across the country. Along with the style name, there are structured data fields which include frame attributes of shape, color, gender, brand, order qty (or sales), and geographic location information.

The Tasks

For this final project, the assignment is to create a model to predict the best frame assortment to generate the highest sales using the dataset provided. This dataset has been filtered and simplified to support development of a clustering model to predict frame sales based on the assortment carried within stores. It will be important to derive what features are most meaningful to improve the model's predictive power.

From the Python model created, analyze the data with frame attributes and advise the VSP product manager how to generate a frame assortment for store locations. It can be adjusted based on regions of the US and demographic buying patterns. Create a ten-minute presentation and a 10-20 page report that answers the following questions. Since the presentation is relatively short, you may choose to focus on a few highlights from your analysis and expand further on the details in the report.

1. General analysis:
    a. What are the most popular frame styles overall?
    b. Is there a price point range that is optimal?
    c. What frame attributes are most important in predicting sales?

2. Frame Product-related analysis:
    a. What frame material is the most popular?
    b. Is there a norm of sizes which exist and are relevant in certain locations?
    c. What would you tell each of the brand managers to improve sales?
    d. Is there a good range of SKU available at key price points that patients purchase (variety of brands and styles)?
    e. What are most important product attributes in creating the optimal assortment?

The intention behind the assignment is to gain a deeper understanding of the product management role and how tools are used to analyze product data to create the assortments, so patients have the "right product and the right price" for their needs. (e.g. buying patterns, product attributes, regional factors, etc.) Creating the data model and interpreting the outputs, is sufficient to do this project. You are free to do some outside research to validate/supplement the findings, especially if you are unfamiliar with the domain, and to apply other more sophisticated text analysis techniques you may know of, but they should not be required. Large language models are not required for this analysis; if you use them; please complement the use of LLMs with other techniques (e.g. topic modeling).

Please submit a .pdf of your slides, as well as the final report. There are no hard guidelines for the final report, but if you are submitting more than 20 pages, you are probably including too much detail. Similarly, if your report is fewer than 10 pages, you may not be answering the questions fully. The 10-20 pages is for the text of the report, and does not include the code. The preferred submission format for the final report is a .pdf file accompanied by the code in a separate Jupyter notebook. The final report should be separate from your slides and the code

Rubric

Each team should turn in separate files for the presentation, the report and a Jupyter notebook for the code. The presentation and report will be graded based on clarity, organization and presentation/writing quality as well as content. Watch for typos and grammatical mistakes. Structure your presentation and report to tell a coherent story: with executive summary (one slide/page with key results), introduction/motivation that states why you are performing the analysis and a conclusion that reinforces key findings/recommendations.

Presentation: Focus the presentation on the problem you are solving and any insights/ results, and leave the details to the report.

- What problem is being solved? Focus on one or two key insights from your results, not all ranges of assortments that the data could address.
- What does the data look like based on your exploratory data analysis?
- How as the data prepared for modeling?
- What modeling approach did you use?
- What are the results?
- How did you evaluate the results?
- What next steps would you recommend based on your results?

Presentations are in class on October 9. Please submit separate files for your slides, final reports and code notebooks.

Only one team member needs to submit the slides and final reports on behalf of the team. Remember to include a paragraph in your report describing who did what in your team.