STAT 423 FINAL PROJECT

CHRIS CHEN, PETER LIU, AND YUEQI XU

Abstract. TBD.

1. Introduction

Perfumes can be defined as substances that emit and diffuse a pleasant and fragrant odor Johnson (2005). The manufacture and use of perfumes goes back thousands of years. Ancient Egyptians, Indians, and Chinese used plants, gums, and resins for fragrance in religious rites Johnson (2005). Until the nineteenth century, perfumes were usually made from natural aromatic chemicals and essential oils. In the modern age, with the development of chemical technology, most perfumes become synthetic, and the number of commercial perfumes surged significantly. With an overwhelming amount of perfumes in the markets, it is often hard for consumers to identify the perfume that best fits their demands. Moreover, given the perfumes' high retail prices, a false purchase can lead to great loss of money. In this project, we are interested in answering the following list of substantial questions in order to quantify the relationships between perfume brand, customer ratings, ..., etc. and retail prices and to assist consumers to better their purchases. The concerning questions are:

(1) Does the prestige of the brands contribute the most of the perfume retail price? Namely, does bigger brands typically cost more?
(2) Does the customer rating contribute the most of the perfume retail price? Namely, does good perfume are typically more expensive?
(3) From the prospective of perfume enthusiasts, does the perfume's department (female/male/neutral/kids) or notes (woody/floral/fruity/etc...) generally leads to difference in retail prices and customer ratings?

To investigate the above questions, we will analyze a data set of perfumes for female, male, and kids, which is collected from a lifestyle shopping destination in Saudi Arabia called

noon. The dataset contains a collection of 1,002 perfumes, specified with product parameters including retail price, brand, volume, fragrance concentration, department, scents, base, and middle note, as well as retail information including retail company, customer & retailer rating, and grand selling amount. The original data set can be found in here.

We will elaborate our data cleaning process, choice of covariate, and regression model selection in the methodology section below.

References

Johnson, G. (2005). Encyclopedia of analytical science (2nd edition)2005421edited by paul worsfold, alan townshend and colin poole. encyclopedia of analytical science (2nd edition) . amsterdam: Elsevier academic press 2005. 10 vols., isbn: 0 12 764199 9 £2,950, $4,570. \mathit{ReferenceReviews}, 19 : 38 - -39.$