Discussion

The model we constructed has achieved a decent overall performance in terms of the model diagnostics metrics and has been largely conducive to providing us with the insight to multiple subjects outside Statistics.

The residual standard error is 175.54, which means the model we proposed is a relatively good fit to the dataset. The adjusted R square is 0.17145, which means 17.145% of the variation in the dataset can be explained by our model. Considering the overall simplicity of our model and the typically chaotic nature of perfume pricing, we would argue that the adjusted R square is already impressively high. The generalization error, Marllow's Cp, AIC, and BIC of this model are all the lowest among all the models we have constructed while exploring; thus, this model has overall the best predictive power. Although typically AIC and BIC prefer different kinds of models due to the fact that these two criteria have different penalizing power on overfitting, our model achieved the lowest on both statistics. The 5-fold Cross Validation of this model gives a mean squared prediction error of 31796, a relatively good improvement comparing to the mean squared prediction error of the 5-fold CV of the full model, which is 32891. In addition, all 7 predictors included in our model are significant at a 5% level, and these predictors' coefficients largely reflect how market works in real life.

We shall interpret the coefficients from an economics perspective. Saudi Arabia has adopted a free market economic model. In such a model, the price of goods and services is only determined by demand and supply. Price increases as demand increases and decreases as supply increases, holding all other factors constant. In the following paragraphs, we will go over the coefficients of the seven predictors in our proposed model and discuss about how they make sense in a real-world context.

First is big_brand, an indicator of whether the manufacturer of the perfume is a popular brand, with 1 means yes and 0 means no. In a free market, brand prestige is considered as a significant determinant of demand, namely higher brand prestige leads to higher demand. Since higher demand leads to higher price, holding other factors constant, the coefficient of big_brand should be a large positive number—85.563, as expected. This means perfumes from popular brands are on average $85.563 more expensive than perfumes from non-popular (niche) brands.

Second comes conc, which is short for concentration, a binary variable with 0 means EDP and 1 means EDT. EDT, short for Eau De Toilette, is more diluted than EDP, short for Eau De Parfum. Therefore, the cost of production of EDT perfumes are typically much lower than that of EDP perfumes. Cost of production is considered to be a significant determinant of supply; namely higher cost of production means lower supply. Since lower supply leads to higher price, holding other factors constant, we expect to see EDT perfumes have lower prices. That is exactly what observe—the coefficient of concEDT is -111.021, which means on average the price of EDT perfumes are $111.021 cheaper than EDP perfumes.

For the same reason, the coefficient of ml, which is an indicator of the volume of the perfume, has a positive coefficient of 1.2119. This means on average, a 1 ml increase in the volume of the perfume is associated with an $1.2119 increase in the perfume's retail price.

Seller_rating is related to loyalty, another determinant of demand. Better seller rating means the seller is more trustworthy and reliable, hence consumers will be more loyal to this seller. Therefore, higher seller rating is correlated with higher demand, and hence higher price, holding other factors constant. This explains the positive coefficient (56.859) of seller_rating. Namely, for every 1 unit increase in seller rating, the price of the perfume is expected to increase by $56.859 on average.

Is.unisex is an indicator of whether the department of the perfume is unisex, with 1 means yes and 0 means no. Saudi Arabia is not a LGBTQ friendly country; the laws there do not protect the rights of such groups and "non-binary gender" is not accepted by the law hence most people there. Therefore, the number of potential buyers of unisex goods is little. Number of potential buyers is one of the determinants of demand, with larger number of potential buyers leads to higher demand and vice versa. Consequently, the coefficient of is.unisex is -146.30, meaning on average the price of unisex perfumes is $146.30 cheaper than the price of non-unisex perfumes.

Perfume manufacturers typically want their products to have a feeling of structure and depth. This feeling can be created via either the easy way or the hard way. While the easy way involves using premium natural ingredients like Bulgarian rose essential oils, oud extracts, and ambergris, the hard way relies on the usage of "scent snapshots"—chemical synthetics like acid, calone, and aldehyde. Premium natural ingredients automatically create that feeling of structure and depth, while chemical synthetics are very simple and shallow. Therefore, it requires the stacking of multiple layers of "scent snapshots" to create the feeling of structure and depth, i.e., if we want the same level of feeling of structure and depth, "the hard way" will typically give more notes, and hence a higher comp score to the perfume comparing to "the easy way". However, at the same time chemical synthetics, which have an average price of $30-$50 per ounce, are much cheaper than premium natural ingredients, as the price of those could go easily above $1000 per ounce. Consequently, higher comp score is associated with lower cost of production; now, applying the logic mentioned previously, it means higher supply and hence lower price. Based on our model's coefficient, for every 1 unit increase in comp, i.e., 1 more

note contained in the perfume, the price of the perfume is expected to decrease by $4.4408 on average.

Lastly, and the most interesting one is is.fresh, an indicator of whether a perfume is categorized to have a "fresh" scent, with 1 means yes and 0 means no. Fresh scent category is only created after the invention of calone, 30 years ago. So, fresh scent perfumes only have a history of around 30 years, while other scent perfumes like woody, floral, and oriental perfumes have over 1000 years of history. Therefore, the number of potential buyers of fresh scent perfumes is still small due to the fact that most people have not yet accepted "fresh" to be a perfume scent. Consequently, demand is lower and hence price is lower for fresh scent perfumes. Besides the "demand-side effect", fresh scent perfumes are also related to "supply-side effect". Since the thing that brought fresh scent perfumes to life is a chemical synthetic—calone, and most fresh scent perfumes are created to capture the "aquatic smell", fresh scent perfumes only use chemical synthetics as ingredients. Therefore, the cost of production is low and hence the supply is high; this means lower price when holding other factors constant. Combining both "demand-side effects" and "supply-side effects", fresh scent perfumes are automatically cheaper than the rest, namely by $85.371 on average.

Another intrigue observation is that item rating has no correlation to price. This is understandable from a biopsychology perspective: when one smells a perfume, the olfactory bulbs and the end of the nostrils send feedback to the cortices of the brain; however, the signals do not only go to the cortices, but they also travel through hippocampus and hypothalamus, triggering memories and emotions that vary across individuals along the way. This means that "there are 1000 *Acqua Di Gio*'s in 1000 people's noses", yet there's only one price. Consequently, item rating and price should be indeed uncorrelated.