Luxury Handbag Auction Analysis

2022-10-20

0.Load data and libraries

```
library(readxl)
library(magrittr)
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:magrittr':
##
##     extract
```

```
library(stringr)
library(dbplyr)
```

```
data <- read_excel('auction_dataset_raw data.xlsx')
head(data)
```

```
## # A tibble: 6 x 10
##   `web-scraper-order` `web-scraper-sta~` page  brand bag   estimate `sold price`
##   <chr>               <chr>              <lgl> <chr> <chr> <chr>    <chr>
## 1 1664212455-1        https://www.sothe~ NA    8301~ Nata~ 180,000~ <NA>
## 2 1664212455-2        https://www.sothe~ NA    8302~ Nata~ 15,000 ~ 35,280 HKD
## 3 1664212455-3        https://www.sothe~ NA    8303~ Nata~ 160,000~ 277,200 HKD
## 4 1664212455-4        https://www.sothe~ NA    8305~ Whit~ 900,000~ 1,638,000 H~
## 5 1664212455-5        https://www.sothe~ NA    8306~ Nata~ 55,000 ~ 60,480 HKD
## 6 1664212455-6        https://www.sothe~ NA    8307~ Meta~ 400,000~ 1,071,000 H~
## # ... with 3 more variables: status <chr>, Location <chr>, `Auction year` <dbl>
```

1. Business Understanding

Luxury Handbag Auction Analysis

Business Problem: Taking inspiration from the luxury value perception model proposed by Weidman et al., this project explores the individual, functional, social and financial value of a handbag based on the following attributes – uniqueness, quality, prestige, vintage, and sold price. This model can thus be used to understand the market better according to the relevant dimensions of luxury perception. ***More importantly, in this project, I want to find out that what are some key indicators that determine the auction price.***

2. Data Processing

Brand

```
#Brand
#unique(data$brand)
data$brand[grepl("Hermès", data$brand)] <- 'Hermès'
data$brand[grepl("Louis Vuitton", data$brand) |
             grepl("LOUIS VUITTON", data$brand)] <- 'Louis Vuitton'
data$brand[grepl("Chanel", data$brand)] <- 'Chanel'
data$brand[grepl("Dior", data$brand)] <- 'Dior'
data$brand[grepl("Goyard", data$brand)] <- 'Goyard'
data$brand[grepl("Gucci", data$brand)] <- 'Gucci'
data$brand[grepl("Tiffany & Co.", data$brand)] <- 'Tiffany & Co.'
data$brand[grepl("FENDI", data$brand) |
             grepl("Fendi", data$brand)] <- 'Fendi'
data$brand[grepl("Gianni Versace", data$brand)] <- 'Gianni Versace'
data$brand[grepl("Bottega Veneta", data$brand)] <- 'Bottega Veneta'
data$brand[grepl("Bvlgari", data$brand)] <- 'Bvlgari'
data$brand[grepl("Alexander McQueen", data$brand)] <- 'Alexander McQueen'
data$brand[grepl("Etoupe Doblis", data$brand)] <- 'Etoupe Doblis'
unique(data$brand)
```

```
##  [1] "Hermès"            "Goyard"            "Chanel"
##  [4] "Louis Vuitton"     "Alexander McQueen" "Bvlgari"
##  [7] "Dior"              "Gucci"             "Bottega Veneta"
## [10] "Tiffany & Co."     "Fendi"             "Gianni Versace"
## [13] "Etoupe Doblis"
```

Bag Characteristics 1. Find leather type (Crocodile/Alligator/Ostrich/Others) 2. Find product year (Before 2002 or After 2002) 3. Find style (Birkin/Kelly/Constance/Roulis/Others) 4. Find color type (5-Others / 4 - Silver/Gold/Gris / 3 - Red/Rose/Organe/Pin / 2 - Blue / 1 - Black/White)

```
#Leather
data$quality_4 <- ifelse(grepl("Crocodile", data$bag), 1, 0)
data$quality_3 <- ifelse(grepl("Alligator", data$bag), 1, 0)
data$quality_2 <- ifelse(grepl("Ostrich", data$bag), 1, 0)
```

If the bag is produced before 2002, defined as vintage

```
#Year
data$production_year <- sub(".*?(\\d{4})$", "\\1", data$bag)
#data$production_year <- gsub("^, ", "", production_year)
data$production_year <- as.numeric(ifelse(grepl("\\D", data$production_year), 'no production year', data
```

```
## Warning: NAs introduced by coercion
```

```
data$production_year[is.na(data$production_year)] <- 'no production year'
data$vintage <- ifelse((data$production_year >= 2002), 1, 0)
```

For bag without a production year, randomly fulfill (0,1) for vintage, following the true percentage of (0,1)

```
#Year
table(data$vintage) #100:745
```

```
##
##   0   1
## 100 896
```

```
prop <- table(data$vintage, useNA = "no") / sum(!is.na(data$vintage))
data$vintage[is.na(data$vintage)] <- sample(c(0, 1),
                                            size = sum(is.na(data$vintage)),
                                            replace = TRUE,
                                            prob = prop)
```

```
#Style
data$prestige_5 <- ifelse(grepl("Birkin", data$bag), 1, 0)
data$prestige_4 <- ifelse(grepl("Kelly", data$bag), 1, 0)
data$prestige_3 <- ifelse(grepl("Constance", data$bag), 1, 0)
data$prestige_2 <- ifelse(grepl("Roulis", data$bag), 1, 0)

#Color
data$unique_4 <- ifelse(grepl("Sliver", data$bag) | grepl("Gold", data$bag)
                        | grepl("Gris", data$bag), 1, 0)
data$unique_3 <- ifelse(grepl("Red", data$bag) | grepl("Rose", data$bag)
                        | grepl("Orange", data$bag) | grepl("Pink", data$bag), 1, 0)
data$unique_2 <- ifelse(grepl("Blue", data$bag), 1, 0)
data$unique_1 <- ifelse(grepl("Black", data$bag) | grepl("White", data$bag), 1, 0)
```

Price Estimate Price 1. Calculate mean estimate price 2. Convert currency to USD

```
data$estimate <- gsub(",", "", data$estimate)  # Remove commas
data$estimate <- gsub("USD", "", data$estimate)  # Remove Currency
data$estimate <- gsub("HKD", "", data$estimate)  # Remove Currency
data$estimate <- gsub("CHF", "", data$estimate)  # Remove Currency
data$estimate <- gsub("GBP", "", data$estimate)  # Remove Currency
data$estimate <- gsub("EUR", "", data$estimate)  # Remove Currency
split_values <- strsplit(data$estimate, " - ")  # Split by " - "

data$estimate_low  <- as.numeric(sapply(split_values, "[", 1))  # Get first element
data$estimate_high <- as.numeric(sapply(split_values, "[", 2))  # Get second element
data$estimate_average <- rowMeans(data[,c('estimate_low','estimate_high')])
```

```
#unique(data$Location)

conversion_rate <- c("New York" = 1, "Hong Kong" = 0.13, 'Geneva' = 0.9,
                     'London' = 0.8, 'Milan' = 0.92, 'Paris ' = 0.92)

# Apply conversion rate to create new column
data$estimate_average_converted <- data$estimate_average * conversion_rate[data$Location]
```

```
data$`sold price` <- gsub(",", "", data$`sold price`)  # Remove commas
data$`sold price` <- gsub("USD", "", data$`sold price`)  # Remove Currency
```

```
data$`sold price` <- gsub("HKD", "", data$`sold price`)  # Remove Currency
data$`sold price` <- gsub("CHF", "", data$`sold price`)  # Remove Currency
data$`sold price` <- gsub("GBP", "", data$`sold price`)  # Remove Currency
data$`sold price` <- gsub("EUR", "", data$`sold price`)  # Remove Currency
```

```
data$`sold price`[is.na(data$`sold price`)] <- 0
data$`sold price` <- as.numeric(data$`sold price`)
data$sold_price_converted <- data$`sold price` * conversion_rate[data$Location]
```

```
# Apply conversion rate to create new column
data$sold_price_converted <- data$`sold price` * conversion_rate[data$Location]
```

Brand: This variable represents the brand of the bags. The brands considered in this dataset include Hermès, Goyard, Chanel, Louis Vuitton, Alexander McQueen, Bvlgari, Dior, Gucci, Bottega Veneta, Tiffany & Co., Fendi, Gianni Versace, and Etoupe Doblis.

Location: This variable denotes the geographical location where the auction was held.

Auction year: This variable represents the year in which the auction took place.

quality_4: This binary variable indicates whether the bag is made from crocodile leather (1) or not (0).

quality_3: This binary variable signifies whether the bag is made from alligator leather (1) or not (0).

quality_2: This binary variable shows whether the bag is made from ostrich leather (1) or not (0).

vintage: This binary variable indicates whether the bag was manufactured before the year 2002 (1) or not (0).

prestige_5: This binary variable indicates whether the bag is a Birkin model (1) or not (0).

prestige_4: This binary variable signifies whether the bag is a Kelly model (1) or not (0).

prestige_3: This binary variable shows whether the bag is a Constance model (1) or not (0).

prestige_2: This binary variable indicates whether the bag is a Roulis model (1) or not (0).

unique_4: This variable denotes whether the bag's color is within the silver/gold/gris spectrum.

unique_3: This variable indicates whether the bag's color falls within the red/rose/orange/pink spectrum.

unique_2: This variable shows whether the bag's color is blue.

unique_1: This variable indicates whether the bag's color is black.

estimate_average_converted: This variable represents the midpoint of the estimated selling price range, in USD.

sold_price_converted: This variable denotes the actual price at which the bag was sold, in USD.

Dataset For Modeling

```
clean_dataset <- data[,c('brand',
'Location',
'Auction year',
'quality_4',
'quality_3',
'quality_2',
'vintage',
'prestige_5',
'prestige_4',
```

```
'prestige_3',
'prestige_2',
'unique_4',
'unique_3',
'unique_2',
'unique_1',
'estimate_average_converted',
'sold_price_converted')]

write.csv(clean_dataset, 'clean_auction_dataset.csv')
```

Dataset For Visualization

```
clean_dataset_visual <- data[,c('brand',
'Location',
'Auction year',
'production_year',
'estimate',
'quality_4',
'quality_3',
'quality_2',
'vintage',
'prestige_5',
'prestige_4',
'prestige_3',
'prestige_2',
'unique_4',
'unique_3',
'unique_2',
'unique_1',
'estimate_average_converted',
'sold_price_converted')]

write.csv(clean_dataset_visual, 'clean_auction_dataset_visualization.csv')
```