

# Final Visualizations

Tingting Thompson

5/6/2022

## Price Density Plot Grid

```
price.df <- read_csv("data/prices.csv")
```

```
## Rows: 30220012 Columns: 5
## -- Column specification -----
## Delimiter: ","
## chr (3): customer_id, club_member_status, fashion_news_frequency
## dbl (2): age, price
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

mutate fashion\_news\_frequency

```
price.df <- price.df %>% mutate(fashion_news_frequency = ifelse(fashion_news_frequency=="NONE", "None", f
```

```
(price.df <- price.df %>% mutate(fashion_news_frequency=factor(fashion_news_frequency, ordered = T, leve
```

```
## # A tibble: 30,220,012 x 5
##   customer_id          club_member_stat~ fashion_news_fre~   age  price
##   <chr>              <chr>              <ord>      <dbl>  <dbl>
## 1 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0542
## 2 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0102
## 3 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0508
## 4 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0542
## 5 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0216
## 6 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0366
## 7 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0144
## 8 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0115
## 9 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0305
## 10 00000dbacae5abe5e23885899a1~ ACTIVE          None         49 0.0318
## # ... with 30,220,002 more rows
```

```
##generate visual
```

```

# create png
png(filename = "final_visuals/price_density_grid.png", width = 879, height = 564,
     units = "px")

#create graphic
price.df %>% ggplot(aes(price)) + geom_density() + facet_grid(fashion_news_frequency~club_member_status,
  labs(x="Scaled Price", title="Distribution of Prices Amount Varying Club Statuses and News Subscripti

#export visual
dev.off()

```

```

## pdf
## 2

```

```
rm(price.df)
```

## Scatter Plot: price ~ age

```
customers <- read_csv("data/customers_wprice.csv")
```

```

## Rows: 1362281 Columns: 8
## -- Column specification -----
## Delimiter: ","
## chr (4): customer_id, club_member_status, fashion_news_frequency, postal_code
## dbl (4): FN, Active, age, avg_price
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

```

```

# create png
png(filename = "final_visuals/price_age_scatter.png", width = 700, height = 700,
     units = "px")

#create graphic
customers %>% ggplot(aes(x=age,y=avg_price, shape = fashion_news_frequency, color = fashion_news_frequen
  geom_point() +
  theme_linedraw() +
  scale_color_colorblind() +
  labs(x="Age (yrs)", y="Average Scaled Price Paid", title = "Average Price Paid vs. Age: Across New Sul

```

```
## Warning: Removed 29547 rows containing missing values (geom_point).
```

```

#export visual
dev.off()

```

```

## pdf
## 2

```

```
rm(customers)
```

## seasonal timeseries

```
t_dat.df <- read_csv("data/t_dat_df.csv")
```

```
## Rows: 734 Columns: 2
## -- Column specification -----
## Delimiter: ","
## dbl   (1): Freq
## date  (1): t_dat
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

### Create season variable

```
format(t_dat.df$t_dat[1], "%m")
```

```
## [1] "09"
```

```
find_season <- function(dt){
  if (format(dt, "%m") %in% c("03", "04", "05")){
    return("spring")
  }
  else if (format(dt, "%m") %in% c("06", "07", "08")){
    return("summer")
  }
  else if (format(dt, "%m") %in% c("09", "10", "11")){
    return("fall")
  }
  else if (format(dt, "%m") %in% c("12", "01", "02")){
    return("winter")
  }
}
```

```
t_dat.df$season <- factor(sapply(t_dat.df$t_dat, find_season), levels=c("fall", "winter", "spring", "summer"))
```

```
season.colors <- c(fall="#6E260E", winter="#95ced6", spring="#FB6B90", summer="#F08000")
```

```
head(t_dat.df)
```

```
## # A tibble: 6 x 3
##   t_dat      Freq season
##   <date>    <dbl> <ord>
## 1 2018-09-20 48399 fall
## 2 2018-09-21 47543 fall
```

```
## 3 2018-09-22 17635 fall
## 4 2018-09-23 52230 fall
## 5 2018-09-24 45984 fall
## 6 2018-09-25 43857 fall
```

## generate visual

```
# create png
png(filename = "final_visuals/seasonal_purchases_ts.png", width = 865, height = 456,
     units = "px")

#create graphic
ggplot(t_dat.df, aes(x=t_dat, y=Freq/1000, fill=season)) + geom_col() +
  scale_color_colorblind() +
  labs(x="Date", y="# Purchases in Thousands",
       title="Number of Purchases Over Time",
       subtitle="consistent sales besides slight increases in summer and fall, likely back-to-school shopping")
theme(legend.position="none", plot.title = element_text(hjust = 0.5), plot.subtitle = element_text(hjust = 0.5))
scale_x_continuous("", breaks=c(as.Date("2018-10-15"),
                                as.Date("2019-1-15"),
                                as.Date("2019-4-15"),
                                as.Date("2019-7-15"),
                                as.Date("2019-10-15"),
                                as.Date("2020-1-15"),
                                as.Date("2020-4-15"),
                                as.Date("2020-7-15"),
                                as.Date("2020-10-15")),
          labels=c("Fall 2018",
                   "Winter 2018",
                   "Spring 2019",
                   "Summer 2019",
                   "Fall 2019",
                   "Winter 2019",
                   "Spring 2020",
                   "Summer 2020",
                   "Fall 2020"))

#export visual
dev.off()
```

```
## pdf
## 2
```