Vanessa Crowe

Econometrics Fall 2023

Professor Kevin Foster

H&M and Data Analysis

Introduction:

By the 1960's and 70's, young people were creating new trends using clothing as a form of personal expression while still sporting a distinction between high fashion and street fashion. By the late 1990's and early 2000's, low-cost fashion reached its peak while companies like H&M and Zara opened stores and took design elements from top fashion brands and reproduced them cheaper and quicker. The term 'instant fashion,' now referred to as 'fast fashion' was coined by one of the largest current brands, Zara, also known as Inditex. The term refers to fashion that is heavily driven by investments in informational technology used to bring new designs into stores rapidly responding to fashion trends. As of 2023, there are numerous fast fashion brands such as Shein, Zara, H&M, Forever 21, Fashion Nova, and many more. Their marketing styles and consumer base differs tremendously from luxury brands. The beauty of fast fashion is their ability to keep up with rapidly changing fashion trends and showcase it in its collections with little delay. These brands create new designs and puts them in stores in a matter of weeks whereas non-fast fashion brands take several months to turn inventory over. The fashion industry is very volatile and with increasing competition and sudden popularity of items if it is worn by celebrities or reflects the latest collection of a top designer, fast-fashion brands need to ensure they have clothing in stock at the time customers want. Brands like Zara and H&M adopt and leverage social media and digital platforms in advertising and communication strategies. The customers are the most important source of information but, they also employ experts in trend analysis and consumer insights.

Hennes & Mauritz, also known as H&M, will be the focus for this analysis. H&M, as the name 'Hennes' suggests, specializes in women's fashion initially and later changed to also include men's fashion now serving all ages and demographics. H&M is the longest running fast-fashion retailer, founded in 1947, expanding to London in 1976, and opening its first store in the U.S in April 2000 on 5th Avenue New York. Consumers justified their purchases to H&M stating it was now 'chic to pay less' as it allowed more people to communicate through clothing regardless of their social and economic backgrounds.

H&M was listed number 9 on brand value in 2020 at $11.5 billion where Nike (1), Adidas (6), and Zara (8) were the only other fast fashion brands on the list. The dataset used takes consumer's preferences and descriptions from H&M's 53 online markets and approximately 4,850 stores worldwide from September 2018 to September 2020. There were 3 sets of data, articles, customers, and transactions train: articles' data provided detailed descriptions of clothing items ranging from the color to type of article. The customers' data provided details on the customers' membership status and age. The transactions' data is the training data that consists of purchases for each customer matching a specific date and price.

Some of the most interesting results from the data was that the date most transactions took place on was not a holiday or was it was not close to a holiday nor was it close to a major change in seasons, and it was actually before payday, which was September 28, 2019 with 24,587 sales and 17,364 sales on 2018-09-29 in the middle of Paris Fashion week. It was also interesting that there were consistently a lot of sales in mid-June showing that people could be Summer shopping or celebrating the end of school. Although, the target demographic of H&M is

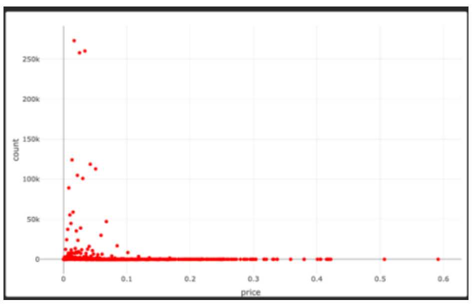
0-40 year old's, there is an overwhelming majority of purchases coming within the age ranges [20,25] and [25,30], which is almost double the amount of sales in other age ranges. Specifically, the age range with the most items purchased was [25,30] with 11,334,707 items in total. Overall, the popular products of H&M reported by these customers aligned well with the brands image but, the most popular product was a t-shirt bra in terms of 'articles_id' however, 'trousers' were actually the most popular when looking as 'detailed descriptions' of products.

Literature Review

The relationship between big data or machine learning techniques and the increase of forecasting for consumers to buy and for retailers to sell has always played a large part in how businesses operate, and the fashion industry is not exempt from that. The volatility of fashion trends was emphasized and made sure to explain that looking at fashion data based on seasons was a huge part of anticipating or learning about what consumers want to purchase. The studies on fashion forecasting and how different factors affects how much or what items a customer may purchase mostly spoke about the prediction of one aspect of fashion products rather than a whole forecast or analysis on trends like colors, styles, or personal details of the customer. With the rise of e-commerce and dependence of social media to decide what to purchase, there was also a lot of studies speaking about the effect of social media on trends like using twitter searches and google trends to see what people were most interested in. There are countless quantitative studies related to social media in fashion where researchers have employed survey methods to collect and analyze data.

According to Gordon, studies have found that while price declines when new computer models or technology are introduced, the opposite happens with apparel prices where price increases occur with changes in models or styles. While analyzing the data from H&M with the knowledge that the brand employs discounts and mark downs depending on the season, it was interesting to see that number of items people were purchasing was not entirely dependent on price and had many outliers as shown in Figure 1. The fact that the price amounts were between 0 and 1 is most likely the reason the prices that are not separated are so concentrated.

Figure 1:



Several outliers showing that customers prioritize certain items more than others regardless of price as seen in the concentration around 0 and 1 regardless of the many outliers.

The paper "More Amazon Effects: Online Competition and Pricing Behaviors" by Alberto Cavallo explores the impacts of online competition, specifically Amazon, on pricing behaviors in the U.S. The study focuses on how increased online competition has influenced the frequency of price changes and the degree of uniform pricing across different locations, making retail prices more sensitive to aggregate nationwide shocks. The data used is from the Bureau of Public Procurement collecting daily prices for products sold by large retailers from 2008 to 2017. This paper explores potential implications of dynamic pricing and constant monitoring of competitors pricing like "retailers that adjust their prices more frequently and uniformly across locations can be expected to react faster to nationwide shocks" (Cavallo, 4). The researcher created dummies around whether certain products were easily found on Amazon's site using it as a proxy for online competition to study the frequency of price changes based on descriptions of the product. They also used those dummies to test the connection more formally with regressions and dummy variables looking for duplicates found on Amazon or Walmart through building cross-sectional data. While their data was consistent with the increase of price changes, they also voiced their concern of the limitations of the regressions since price indices included non-tradable and goods that are domestically produced which could affect their coefficients and "behavior over time if the composition of imported and domestic products is not constant" (Cavallo). As reviewed in class, a major caveat with data analysis and knowing if it is accurate is being aware of the assumptions you made regarding your data. This researcher mentioned that they assumed they are using a good proxy for online competition, Amazon, which is based on their own experience and knowledge with the site but, the category items found on Amazon could have unobserved characteristics that a customer may or may not search for. The paper mentions suppliers' consideration of the role of internet transparency as well and the fear of antagonizing customers as factors that influence pricing strategies and avoidance of geographical price discrimination.

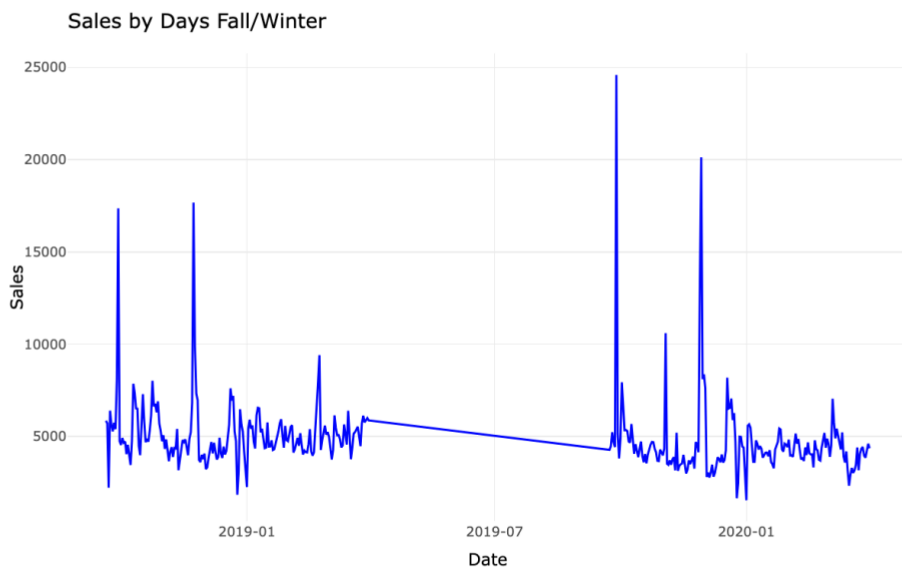
The article by Al-Hallah, Stiefelhagen, and Grauman go deeper into fashion forecasting and hypothesizing mixtures of styles using a dataset of 80,000 fashion products. The researchers aim to predict the future popularity in 12 months of fine-grained fashion styles like observing purchase statistics for all women's dresses sold on Amazon over a certain number of years. It was interesting that the article did not only consider people who are attentive to forecasting due to their interest in haute couture, money, or keeping up with trends but also, noticing that fashion trends show the effects of a shifting cultural attitude, economic factors, and the political climate. They did not only depend on the data used but also discovering the best vocabulary to name certain styles to forecast key visual attributes that will be popular in future fashion. They approached the study by learning what certain "visual attributes" represent in fashion images, finding styles that are shared across images, and using past purchases to predict future trends. While studying the customer transactions, they subset the number of transactions of a specific style within a period of one month to measure the frequency of that style. Then, they used the exponential smoothing model to show how the most recent observed trends have the most impact on the future trends. For their analysis on the visual attributes, they considered using clustering but, determined that it could be distracted due to the bias towards certain properties like color by grouping all Black clothes in one style ignoring differences in shape and material. Through the result of their current trends, they were able to measure several styles including "classic," "trending," "un-popular," and "re-emerging." The visual-based forecasting model use in this paper was able to provide detailed descriptions of what would be trending in the future while considering the different attributes other than color, print, or coverage of the clothing.

Correlations and Regressions

As mentioned, there were 3 different datasets that consisted of important values that can be used to study H&M consumers' behavior. The main alterations to the data were taking a 12.5% sample size of the "transactions_train" data and merging all the datasets based on the variables they have in common. The columns that remained were article_id, customer_id, club_member_status, age, t_dat(date), price, product_type_no, product_type_name, product_group_name, colour_group_name, detail_desc, and many more with 3,935,882 observations. Due to the importance of seasons on fashion styles, I made subsets of the Fall/Winter season, dating 2018-09-20 to 2019-03-31 and 2019-09-23 to 2019-03-31, and of

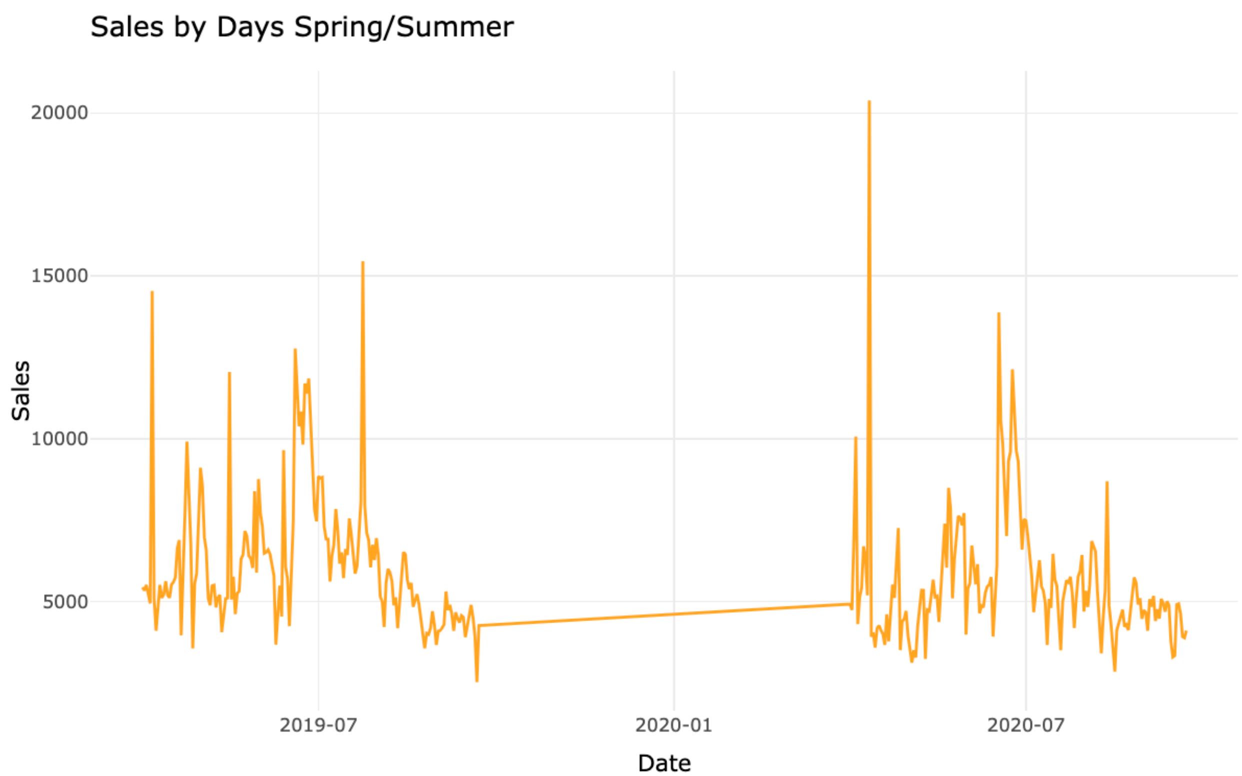
Spring/Summer, dating 2019-04-01 to 2019-09-22 and 2020-04-01 to 2020-09-22, to understand how the purchasing choices of consumers differ based on season and what might impact it.

The observations for the FW seasons were 1,873,867 transactions with the most transactions being made on 2019-09-28 with 24,587. As shown in the graph labeled “Sales by Days Fall/Winter,” there are a lot of large peaks which is expected due to the concentration of major holidays around the world. Other than the top transactions during Paris Fashion Week,

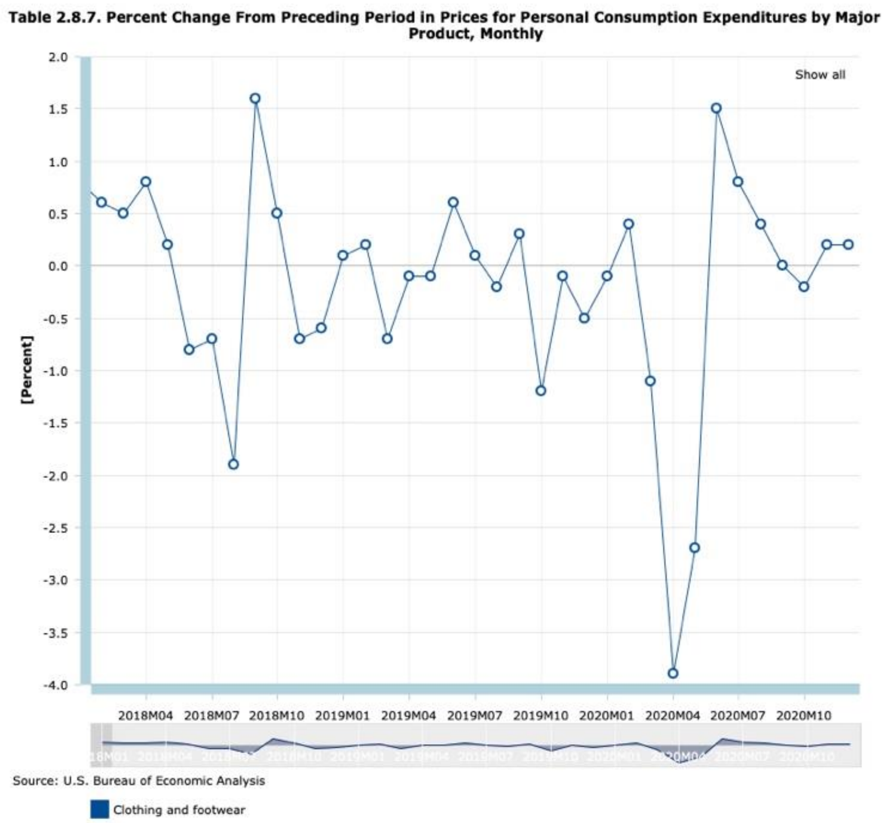


some of the highest purchases were made during Black Friday, Thanksgiving, and midwinter recess for students. Since the country with the most H&M stores is the United States, the high growth of transactions on Black Friday is something many stores prepare for especially H&M as they often sport deals and discounts for customers. The lowest transactions took place on important holidays where spending time with family and turning over a new leaf is encouraged like Christmas and New Year’s Day. Using the Fall and Winter subset, finding the month that had the most items purchased as well as finding out how much was made that month is also important as it helps the retailer know when to market more to customers. As expected from the holidays in Fall, the most purchases took place in November while the least took place in February meaning they made the least money in February.

The observations for Spring and Summer were higher than the FW season at 2,062,015. Since H&M is designed as a fast fashion brand targeted towards consumers who are interested in keeping up with the top fashion trends rather than seeking warm clothing, it is expected that most of their sales will take place in Spring and Summer. During these months, the most transactions occurred on 2019-04-11 with 20,386 purchases. The Spring and Summer months had many more peaks and seemed much less concentrated around sale periods.



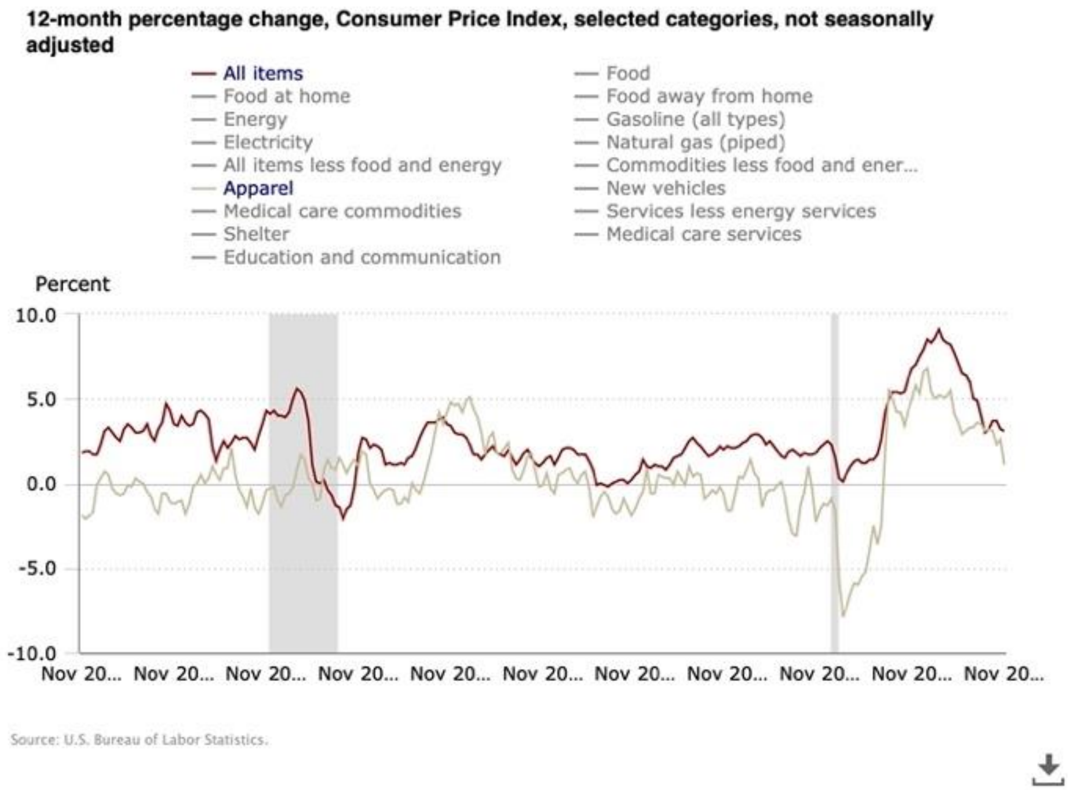
Some of the most sales are leading up to important dates like Memorial Day, back to school time and Easter or leading up to important breaks like Summer break and Spring break. The lowest purchases occurred in April, June, September, May, and August. Since many of the spikes and decreases in transactions were not associated with any special days, I thought it would be interesting to look at the overall consumption on "clothing and footwear."



U.S. Bureau of Economic Analysis, "[Table 2.8.7. Percent Change From Preceding Period in Prices for Personal Consumption Expenditures by Major Type of Product, Monthly](https://apps.bea.gov)" (accessed Friday, December 15, 2023

According to the U.S. Bureau of Economic Analysis, the largest spike in consumption takes place in September 2018 and June 2020 at 1.6% and 1.5% respectively. Consumption then decreased after September 2018 until the beginning of 2019 to decrease in March 2019. This aligns with the transactions from H&M except that the spike in November was not reflected in the overall data. There is also a dip in October 2019 and in April 2020 with a spike in June 2020 and then there are constant decreases in consumption. The dip in March and April could be accounted to people not needing to purchase as much or the fact that the pandemic was in full swing at that point. The rise in June 2020 could also be a result of the pandemic stimulus due to the withdrawal from spending in early 2020.

Since there were still questions surrounding why people were buying at the times they were, looking at the Consumer Price Index of the Apparel prices compared to all items could be beneficial in evaluating the transactions.



https://www.bls.gov/charts/consumer-price-index/consumer-price-index-by-category-line-chart.htm

According to the graph by the U.S. Bureau of Labor Statistics, there is a major dip in for Apparel from February 2020 to May 2020 which does correspond to the months that were on the lower side of transactions from H&M's data. Assessing the individual data of the store is important but outside factors need to be considered like the economic state of the world and how consumers are behaving outside of H&M. There are limitations when looking at this data as the consumption and inflation was based on U.S. consumers while H&M's dataset was based on their consumers from all over the world.

Some other interesting observations were looking at the day of the weeks where the most transactions took place. Using the “weekdays()” operations, I used the date available in the dataset to create a column called “day_of_week” as well as “year.”

```{r}
HandMTransactions <- HandMTransactions %>%
  mutate(day_of_week = as.factor(day_of_week),
         year = as.factor(year))
head(HandMTransactions)

```

Description: df [6 x 17]

graphical_appearance_name <chr>	colour_group_code <chr>	colour_group_name <chr>	detail_desc <chr>	day_of_week <fctr>	year <fctr>
Solid	09	Black	Jersey top with narrow shoulder straps.	Saturday	2019
Solid	09	Black	Jersey top with narrow shoulder straps.	Sunday	2019
Solid	09	Black	Jersey top with narrow shoulder straps.	Wednesday	2019
Solid	09	Black	Jersey top with narrow shoulder straps.	Monday	2018
Solid	09	Black	Jersey top with narrow shoulder straps.	Sunday	2019
Solid	09	Black	Jersey top with narrow shoulder straps.	Friday	2018

6 rows | 13-18 of 17 columns

The most popular day for purchases for FW and SS was Saturday with 313, 447 and 336,137 respectively. The least popular day in general was Monday with 500,222 transactions and for the FW and SS seasons, the least popular days were Monday and Sunday respectively.

For further analysis on how the variables affect each other, I created another subset limiting the age to [15,40] based on their target audience and the range of their most popular buyers transforming the dataset from 3,935,882 observations to 2,556,372. In order to find out how much each customer is buying; the subset was organized by the column “customer_ids.” This would create a column of quantities named “customer_count.” To see what is the age demographic that purchases the most, I also created a column that with 5 age ranges.

```{r}
ranges <- cut(HMSubset$age, breaks = c(15, 20, 25, 30, 35, 40), labels = c("15-20", "21-25", "26-30", "31-3#5", "36+"), include.lowest = TRUE))
HMSubset$ranges <- ranges
head(HMSubset)
```

R Console | grouped_df 6 x 20

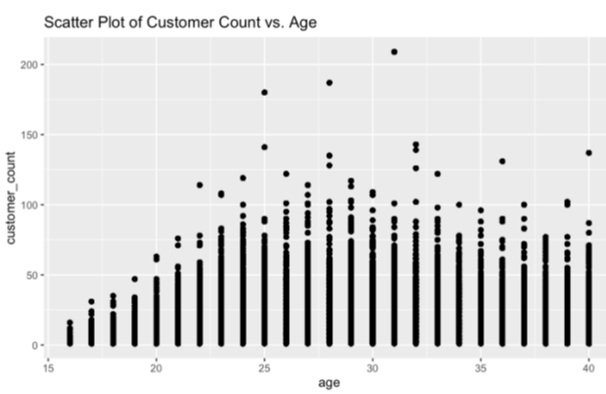
A tibble: 6 x 20 | Groups: customer_id [6]

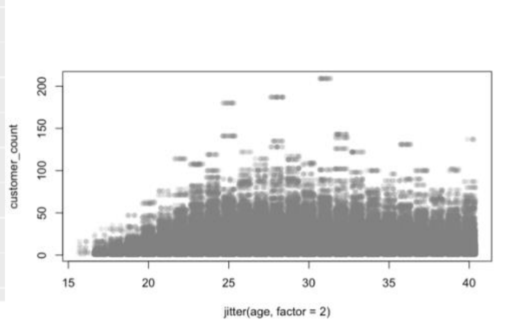
colour_group_name <chr>	detail_desc <chr>	day_of_week <fctr>	year <fctr>	customer_count <int>	total_price1 <dbl>	ranges <fctr>
Black	Jersey top with narrow shoulder straps.	Saturday	2019	6	0.188474576	(30,35]
Black	Jersey top with narrow shoulder straps.	Wednesday	2019	4	0.077898305	(20,25]
Black	Jersey top with narrow shoulder straps.	Monday	2018	23	0.711355932	(30,35]
Black	Jersey top with narrow shoulder straps.	Sunday	2019	1	0.008457627	(25,30]
Black	Jersey top with narrow shoulder straps.	Friday	2018	6	0.208372881	(15,20]
Black	Jersey top with narrow shoulder straps.	Tuesday	2019	18	0.382576271	(35,40]

6 rows | 14-20 of 20 columns

The age range that purchases the most was [25,30] with 11,334,707 total transactions counting different articles the same customer may have purchased.

To further analyze the effects of the variables on each other, I made a scatterplot and looked at the summary of the statistics.






```{r}
summary_stats <- summary(HMSubset[, c("customer_count", "age", "product_type_no", "club_member_status")])
print(summary_stats)
```

```
 customer_count        age        product_type_no club_member_status
 Min.   :  1.00   Min.   :16.00   Min.   : -1     Min.   :0.0000
 1st Qu.:  4.00   1st Qu.:24.00   1st Qu.:253     1st Qu.:1.0000
 Median :  8.00   Median :27.00   Median :265     Median :1.0000
 Mean   : 12.03   Mean   :27.56   Mean   :245     Mean   :0.9837
 3rd Qu.: 15.00   3rd Qu.:31.00   3rd Qu.:273     3rd Qu.:1.0000
 Max.   :209.00   Max.   :40.00   Max.   :762     Max.   :1.0000
```

The ages are still densely packed under the 100 customer_count with a few outliers who are purchasing significantly more items than the majority of people recorded. In this model, price is not an appropriate variable due to the nature of the values and because “customer_count” is included as a quantity. Price cannot be regressed on quantity and quantity cannot be regressed on price.

````
```{r}
ols_model <- lm(customer_count ~ age + product_type_no + club_member_status,  data =  HMSubset)
summary(ols_model)
```

Call:
lm(formula = customer_count ~ age + product_type_no + club_member_status,
    data = HMSubset)

Residuals:
    Min      1Q  Median      3Q     Max
-15.685  -7.683  -3.613   3.215 195.580

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        -6.621e+00  8.215e-02   -80.6   <2e-16 ***
age                 3.621e-01  1.527e-03   237.2   <2e-16 ***
product_type_no     1.148e-05  1.144e-04     0.1     0.92
club_member_status  8.815e+00  6.347e-02   138.9   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.83 on 2556368 degrees of freedom
Multiple R-squared:  0.02756,   Adjusted R-squared:  0.02756
F-statistic: 2.415e+04 on 3 and 2556368 DF,  p-value: < 2.2e-16
````

According to the results of the linear regression model, for each one-unit increase in 'age,' the model predicts an increase of approximately 0.3621 in the “customer_count.” As customer gets older, the more they appear to be purchasing. The effects of product_type_no on customer_counts is not that important in this regression as the p value is very high and the coefficient is very small. For each one-unit increase in 'club_member_status' (which was converted to a binary value), the model predicts an increase of approximately 8.815 in 'customer_count. Usually those with a club membership have purchased from H&M before because the customer is encouraged to sign up when they make their first purchase.

```
```{r}
ols_model2 <- lm(price ~ age + product_type_no + club_member_status,  data =  HMSubset)
summary(ols_model2)
```

Call:
lm(formula = price ~ age + product_type_no + club_member_status,
    data = HMSubset)

Residuals:
     Min   1Q   Median   3Q   Max
-0.03742 -0.01183 -0.00325 0.00640 0.56351

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.954e-02 1.178e-04 165.80   <2e-16 ***
age   1.479e-04 2.190e-06   67.53   <2e-16 ***
product_type_no   2.191e-05 1.641e-07 133.53   <2e-16 ***
club_member_status -1.717e-03 9.104e-05 -18.86   <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01841 on 2556368 degrees of freedom
Multiple R-squared: 0.008949, Adjusted R-squared: 0.008948
F-statistic: 7695 on 3 and 2556368 DF, p-value: < 2.2e-16
```

Based on the linear regression of the effect of price on the factors age, product_type_no, and club_member_status, for each one-unit increase in 'age,' the model predicts an increase of approximately 0.0001479 in “price.” This shows that although the median age range of [25,30] are purchasing the most items, that does not mean they are spending the most money. They may not be purchasing the most expensive products as, the price increases as age does according to this model. For each one-unit increase in 'product_type_no,' the model predicts an increase of approximately 0.00002191 in “price,” this means that as the products go from small accessories or outer summer wear like a hair clip(72), belts (67), top and lower garments (253) to items like long socks (302), stockings (304), the customer is less likely to purchase. The model overall explains a relatively small proportion of the variability in the price.

To further analyze the most popular products, I looked at the top 10 prices and used the datasets to determine what products those match with. The most common price that showed up in the HMSubset is .016932203 which matches up with a microfiber white t-shirt bra with 277,937 sales. As the article by Halah, Stiefelhagen, and Grauman performed analysis on visual styles to predict future values, since the H&M dataset gave descriptions of the data, it would be beneficial to know the most popular colors and styles.

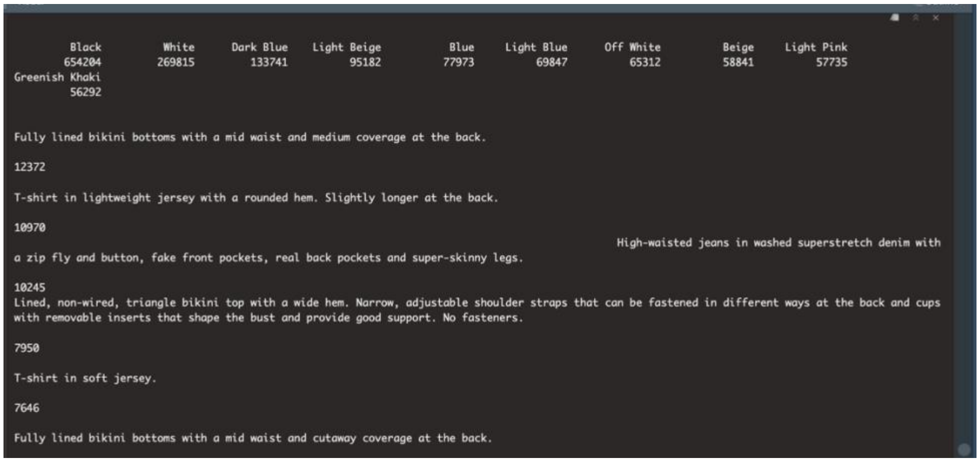
```
 Black     White  Dark Blue      Grey Light Beige   Dark Red      Blue     Beige  Dark Grey  Light Pink
711254    147225     135396     65461       61940      61070     57629     54581      52832       48652

High-waisted jeans in washed superstretch denim with a zip fly and button, fake front pockets, real back pockets and super-skinny legs.
17726
5-pocket jeans in washed, superstretch denim with a regular waist, zip fly and button, and skinny legs.
12761
Blouse in a soft weave with a narrow collar, concealed buttons down the front, long sleeves with buttoned cuffs and a rounded hem.
10905
Jeggings in washed, superstretch denim with a high waist, fake front pockets, real back pockets and skinny legs.
6265
T-shirt in lightweight jersey with a rounded hem. Slightly longer at the back.
6162
5-pocket jeans in washed stretch denim with a high waist, zip fly and slim legs.
5890
5-pocket jeans in superstretch washed denim with a regular waist and super-skinny legs.
5355
Jumper in a soft, fine knit with a ribbed polo neck and ribbing at the cuffs and hem.
4757
Ankle-length jeans in washed stretch denim with a regular waist, fake front pockets, real back pockets and skinny legs.
4694
Fitted vest top in stretch cotton jersey.
4527

Trousers    Sweater     Dress    T-shirt      Top    Blouse   Bra  Underwear bottom
  283596     266074    147553     105972    93198     89345  80401            64682
Vest top  Leggings/Tights
   56248            54330
```

Most popular 'colour_group_name', 'detail_desc,' and 'product_type_name' for Fall/Winter

The colors that show up the most is Black, white, and dark blue while the most common styles include high waisted skinny jeans, jeggings, collared blouses and t-shirts. This goes hand in hand with the most common product name being trousers, sweaters, and dresses.



```
 Black     White  Dark Blue  Light Beige    Blue  Light Blue  Off White   Beige  Light Pink
654204    269815     133741        95182   77973       69847      65312   58841       57735
Greenish Khaki
         56292

Fully lined bikini bottoms with a mid waist and medium coverage at the back.

12372

T-shirt in lightweight jersey with a rounded hem. Slightly longer at the back.

10970
High-waisted jeans in washed superstretch denim with
a zip fly and button, fake front pockets, real back pockets and super-skinny legs.

10245
Lined, non-wired, triangle bikini top with a wide hem. Narrow, adjustable shoulder straps that can be fastened in different ways at the back and cups
with removable inserts that shape the bust and provide good support. No fasteners.

7950

T-shirt in soft jersey.

7646

Fully lined bikini bottoms with a mid waist and cutaway coverage at the back.
```

```
7302
                                                                                5-pocket jeans in
washed, superstretch denim with a regular waist, zip fly and button, and skinny legs.

6929

Cropped, fitted top in cotton jersey with narrow shoulder straps.

6761

Round-necked T-shirt in soft cotton jersey.

6289
                                                                                   T-shirt in
soft jersey with a round neck, short sleeves with sewn-in turn-ups, and a rounded hem.

5807

     Dress     Trousers     T-shirt     Shorts     Vest top     Bikini top Swimwear bottom     Top     Blouse
    254928       239427      166522     124657       119500         108022          104372  103276      97989
       Bra
     85438
```

Most popular 'colour_group_name', 'detail_desc,' and 'product_type_name' for Spring/Summer

The top 3 most common colors in Spring and Summer are the same as the colors in winter but the popular styles are very different. The popular styles being bikinis, lightweight shirts, jeans, and dresses.

```
Other Turquoise    Other Purple      Other Blue         Unknown    Bronze/Copper     Transparent     Other Orange      Other Green       Other Red
             21             397             402             474              496             562              588              841            1033
         Purple
           1124

                  3/4-length leggings in soft stretch denim with an elasticated waist, print motif on the front, fake front pockets and
short slits at the hems.

1
         3/4-length leggings in sturdy, denim-look jersey with an elasticated waist and fake fly. Fake front pockets, real back pockets and
short slits at the hems.

1
                                                                     3/4-length leggings in washed, stretch denim with an elasticated waist and
fake front pockets.

1
3/4-length pull-on trousers in an airy, patterned cotton weave with an elasticated, ribbed, drawstring waist, fake fly, side pockets, a back pocket
and elasticated hems.

1
                                                              3/4-length sports tights in fast-drying functional fabric with flatlock seams and
an elasticated waist.

1
                                                     3/4-length sports tights in fast-drying functional fabric with wide ribbing at the waist for a
more comfortable fit.
```

```
1
                     5-pocket ankle-length jeans in washed stretch denim with decorative zips at the top. Regular waist and skinny legs
with a zip at the hems.

1
               5-pocket ankle-length jeans in washed stretch denim with hard-worn details. Slim fit with a regular waist, zip fly with a
button and raw edge hems.

1
                                           5-pocket jeans in hard-washed stretch denim with a regular waist, button fly and slightly wider, tapered legs
with raw-edge hems.

1
                              5-pocket jeans in soft, stretch denim with worn details. Adjustable, elasticated waist, zip fly and
button and straight legs.

1

Accessories set      Keychain      Moccasins          Towel     Sewing kit         Blanket     Zipper head      Wood balls        Baby Bib
              1             2              2              2              4               5               5               6               9
Chem. cosmetics
             12
```

Least popular 'colour_group_name', 'detail_desc,' and 'product_type_name' for Fall/Winter

The least popular colors are turquoise, other purple, and other blue with 3/4th trousers being the least popular detailed clothing.

```
Other Turquoise        Unknown     Transparent    Other Purple      Other Blue   Bronze/Copper       Other Red     Other Green           Other
             51            231             494             575             581             756             856            1011            1358
         Purple
           2499


14 pairs of metal earrings in various sizes and designs.

10

2-piece set with a longer sweatshirt and matching leggings in a cotton blend. Printed sweatshirt with long sleeves and ribbed cuffs. Patterned
leggings in jersey with an elasticated waist.

6
3-piece set with a shirt and pair of shorts in a cotton weave and a satin bow tie. Shirt with a collar, buttons down the front and short sleeves with
sewn-in turn-ups. Shorts with concealed elastication and a button at the waist, a zip fly, front pockets and a fake back pocket. Detachable,
adjustable elastic braces with imitation leather details. Bow tie with a narrow, adjustable elastic strap and plastic fastener.

7
                                                                                                                             3-piece set with a
zip-through hoodie and joggers in sweatshirt fabric and a jersey T-shirt. Hoodie with a zip down the front and ribbing at the cuffs and hem. Joggers
in a relaxed fit with an easy pull-on, elasticated waist and tapered legs with ribbed hems. T-shirt with a print motif.

1

3/4-length dungarees in washed cotton denim with adjustable straps that fasten at the front. Side pockets, a zip fly and wide, straight legs.
```

```
45

3/4-length jeans in stretch cotton denim with a zip fly and button, fake front pockets, real back pockets and straight legs with short slits in the
sides.

16

3/4-length jeans in washed, stretch denim with fake front pockets, real back pockets and skinny legs. Wide ribbing at the waist for optimum comfort.

10

3/4-length jumpsuit in viscose jersey with a deep wrapover V-neck, a seam with narrow, concealed elastication at the waist, a low crotch and wide
legs. Unlined.

199

3/4-length jumpsuit in viscose jersey with a deep wrapover V-neck, long sleeves, a seam with narrow, concealed elastication under the bust and
straight, wide legs.

10

3/4-length leggings in cotton jersey with a wide waist panel.

59

   Chem. cosmetics          Side table                Toy          Wood balls          Eyeglasses     Accessories set         Leg warmers
                 1                   1                  1                   1                   2                   3                   3
         Moccasins           Sewing kit Stain remover spray
                 3                   3                  3
```

Least popular 'colour_group_name', 'detail_desc,' and 'product_type_name' for Spring/Summer

The least popular colors were also other turquoise, 'unknown,' and transparent with the least popular clothing being heavy two- or three-piece sets.

The most and least popular clothing descriptions line up well with the seasons. The most popular clothing for the Fall and Winter were long clothes like jeans while the most popular in the Spring and Summer were light clothes that are comfortable to wear and reflects Summer activities. The least popular clothes for Fall and Winter were airy shorter clothing while the least popular in Spring and Summer were heavy sweatsuits, overall clothing that would be uncomfortable in the Summer.

Conclusion

Although H&M’s focus on data is for inventory management by keeping track of every order and return to help people understand which items are popular and what people don’t like as much, it is also important to consider the other factors like age, price, and season as to why people buy a particular clothing. Through this analysis, looking at the dates the most sales occurred by seasons showed the most about the data as it produced clear reasons and dates of when marketing to customers should be prioritized and that is during times with high sales like November and June but also dates with low sales like February. While CPI data and consumption data cannot explain exactly why consumers are making certain decisions, it can give some insight on some consumer decisions and how to forecast how customers may react when there is a change in the overall economy. This is not only important for marketing, but it is also important to know what items to keep in stock. Product recommendations is important to not only the retailor but also to the customer as it enhances the shopping experience.

References:

Al-Hallah, Z., Stiefelhagen, R., Grauman, K. (2017). Fast Fashion: Forecasting Visual Style in Fashion. *Computer Vision Foundation.* https://openaccess.thecvf.com/content_ICCV_2017/papers/Al-Halah_Fashion_Forward_Forecasting_ICCV_2017_paper.pdf

Cavallo, A. (2019) More Amazon Effects: Online Competition and pricing Behaviors. *National Bureau of Economic Research*. https://www.nber.org/system/files/working_papers/w25138/w25138.pdf

Gordon, R.J. (2005). Apparel Price 1914-93 and The Hulten/Brueghel Paradox. *National Bureau of Economic Research*. no. c5071. https://www.nber.org/system/files/working_papers/w11548/w11548.pdf