

Valentine's Consumer Spending Trends

The National Retail Federation in the United States conducts surveys and has created a Valentine's Day Data Center to explore the data on how consumers celebrate. The NRF has surveyed consumers about how they plan to celebrate Valentine's Day annually for over a decade.

More information about the dataset can be found

here:<https://github.com/rfordatascience/tidytuesday/tree/main/data/2024/2024-02-13>

(<https://github.com/rfordatascience/tidytuesday/tree/main/data/2024/2024-02-13>)

Question:

1. Which factors (spending categories) most strongly influence overall Valentine's Day spending patterns?
2. What distinct spending behavior profiles exist among people based on their spending categories, and how do these profiles correlate with age groups?

Introduction: I am working with two datasets related to Valentine's Day consumer behavior to address the research questions. These datasets include Valentine's historical spending and spending patterns categorized by age.

To address the first question, "Which spending categories most significantly influence overall Valentine's Day spending patterns?", I used the `historical_spending` dataset. This dataset includes numeric variables representing the average percentage spent on each category: Candy, Flowers, Jewelry, Greeting Cards, Evening Out, Clothing, and Gift Cards.

To answer the second question, "What distinct spending behavior profiles exist based on spending categories, and how do these profiles relate to age groups?", I analyzed the `gifts_age` dataset. This dataset includes the variables Age, Candy, Flowers, Jewelry, Greeting Cards, Evening Out, Clothing, and Gift Cards. Here, Age is a categorical variable divided into six age groups for individuals aged 18 and older.

Approach: First, I conducted a descriptive analysis for both datasets using the `summary` function to understand the data. Next, I transformed the `gifts_age` dataset from wide to long format. To explore the `gifts_age` dataset further, I created a combination of bar and scatter plots, with bars representing spending categories and black dots indicating the overall average spending for each age group.

To address the first research question, "Which factors (spending categories) most strongly influence overall Valentine's Day spending patterns?", I performed a Principal Component Analysis (PCA). I visualized the results using a rotation plot matrix to understand how each spending category contributes to overall spending patterns.

For the second question, "What distinct spending behavior profiles exist among people based on their spending categories, and how do these profiles correlate with age groups?", I applied clustering. I started by creating an elbow plot to determine the optimal number of clusters. Then, I scaled the `gifts_age` dataset and performed PCA before generating the clusters. Finally, I created two scatter plots for PC1 vs. PC2: the first visualized clusters, and the second displayed age groups, using cluster colors to compare the distributions of spending profiles and age groups.

Analysis:

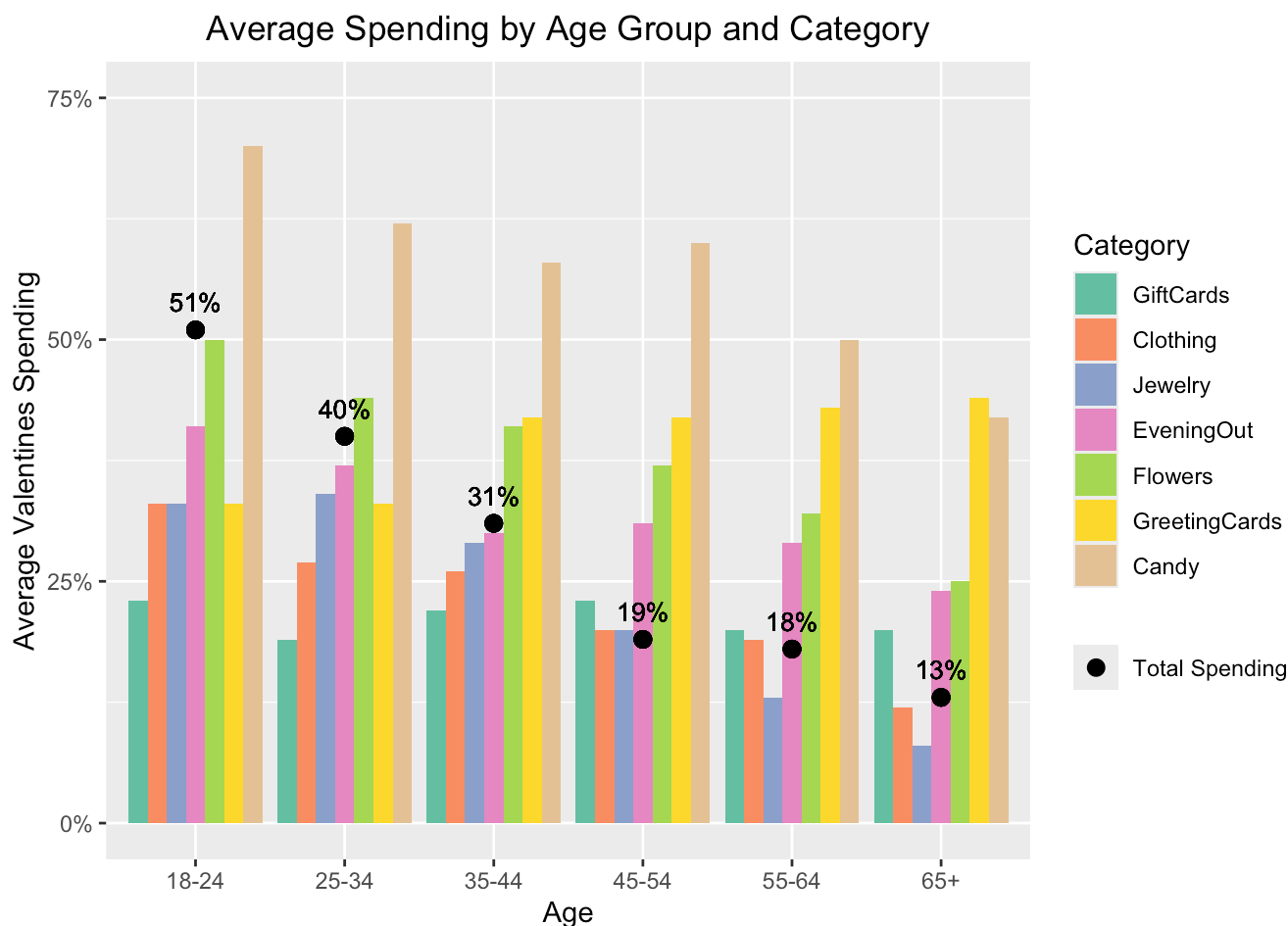
```
# descriptive analysis
summary(historical_spending)
```

```
##      Year      PercentCelebrating      PerPerson      Candy
## Min.      :2010      Min.      :51.00      Min.      :103.0      Min.      : 8.60
## 1st Qu.:2013      1st Qu.:54.00      1st Qu.:131.0      1st Qu.:10.85
## Median :2016      Median :55.00      Median :142.3      Median :12.70
## Mean      :2016      Mean      :55.46      Mean      :144.4      Mean      :12.84
## 3rd Qu.:2019      3rd Qu.:58.00      3rd Qu.:162.0      3rd Qu.:14.12
## Max.      :2022      Max.      :60.00      Max.      :196.3      Max.      :17.30
##      Flowers      Jewelry      GreetingCards      EveningOut
## Min.      :12.33      Min.      :21.52      Min.      :5.910      Min.      :21.39
## 1st Qu.:13.49      1st Qu.:30.34      1st Qu.:7.310      1st Qu.:25.66
## Median :14.78      Median :30.94      Median :7.870      Median :27.48
## Mean      :14.65      Mean      :32.55      Mean      :7.676      Mean      :27.47
## 3rd Qu.:15.42      3rd Qu.:34.10      3rd Qu.:8.320      3rd Qu.:28.46
## Max.      :16.71      Max.      :45.75      Max.      :9.010      Max.      :33.46
##      Clothing      GiftCards
## Min.      :10.42      Min.      : 8.42
## 1st Qu.:12.00      1st Qu.:10.23
## Median :14.04      Median :11.04
## Mean      :14.94      Mean      :11.50
## 3rd Qu.:16.08      3rd Qu.:12.52
## Max.      :21.46      Max.      :17.22
```

```
summary(gifts_age)
```

```
##      Age      SpendingCelebrating      Candy      Flowers
## Length:6      Min.      :13.00      Min.      :42.0      Min.      :25.00
## Class :character      1st Qu.:18.25      1st Qu.:52.0      1st Qu.:33.25
## Mode  :character      Median :25.00      Median :59.0      Median :39.00
##      Mean      :28.67      Mean      :57.0      Mean      :38.17
##      3rd Qu.:37.75      3rd Qu.:61.5      3rd Qu.:43.25
##      Max.      :51.00      Max.      :70.0      Max.      :50.00
##      Jewelry      GreetingCards      EveningOut      Clothing
## Min.      : 8.00      Min.      :33.00      Min.      :24.00      Min.      :12.00
## 1st Qu.:14.75      1st Qu.:35.25      1st Qu.:29.25      1st Qu.:19.25
## Median :24.50      Median :42.00      Median :30.50      Median :23.00
## Mean      :22.83      Mean      :39.50      Mean      :32.00      Mean      :22.83
## 3rd Qu.:32.00      3rd Qu.:42.75      3rd Qu.:35.50      3rd Qu.:26.75
## Max.      :34.00      Max.      :44.00      Max.      :41.00      Max.      :33.00
##      GiftCards
## Min.      :19.00
## 1st Qu.:20.00
## Median :21.00
## Mean      :21.17
## 3rd Qu.:22.75
## Max.      :23.00
```

```
# Data Exploration with bar and scatter plot
gifts_age %>%
  pivot_longer(cols = c(Candy, Flowers, Jewelry, GreetingCards, EveningOut, Clothing, GiftCards),
               names_to = "Category",
               values_to = "AveragePercentSpending") %>%
  mutate(Category = fct_reorder(Category, AveragePercentSpending)) %>%
  ggplot() +
  geom_col(aes(x = Age, y = AveragePercentSpending, fill = Category),
           position = "dodge") +
  geom_point(aes(x = Age, y = SpendingCelebrating, shape = "Total Spending"), size = 3,
            color = "black") +
  geom_text(aes(x = Age, y = SpendingCelebrating, label = paste0(SpendingCelebrating,
"%"), hjust = 0.5, vjust = - 1), size = 3.5) +
  ggtitle("Average Spending by Age Group and Category") +
  scale_fill_brewer(palette = "Set2") +
  scale_shape_manual(values = 16, name = NULL, labels = "Total Spending") + # Legend for black dot
  scale_y_continuous(name = "Average Valentines Spending",
                    limits = c(0,75),
                    breaks = c(0, 25, 50, 75),
                    labels = c("0%", "25%", "50%", "75%")) +
  theme(plot.title = element_text(hjust = 0.5))
```



```
# PCA for historical spending
pca_fit <- historical_spending %>%
  select(-c(Year,PercentCelebrating, PerPerson)) %>%
  scale() %>%           # scale to zero mean and unit variance
  prcomp()
pca_fit
```

```
## Standard deviations (1, .., p=7):
## [1] 2.2092949 0.9625109 0.8477643 0.5008272 0.3581260 0.2852280 0.1159622
##
## Rotation (n x k) = (7 x 7):
##
```

	PC1	PC2	PC3	PC4	PC5
## Candy	0.4302771	0.152704724	0.034839977	0.14051920	0.44863617
## Flowers	0.4188116	-0.034560623	0.207683257	0.60022763	-0.03262447
## Jewelry	0.4119612	-0.246498912	0.241006189	0.08037922	-0.69658620
## GreetingCards	0.2710915	-0.009214882	-0.938829857	0.14759194	-0.12698872
## EveningOut	0.2490472	-0.840987195	-0.013527570	-0.32880108	0.32520431
## Clothing	0.4220460	0.303885714	0.126142890	-0.07043002	0.36653043
## GiftCards	0.3951058	0.339166941	0.008563498	-0.69185722	-0.23706619

```
##
```

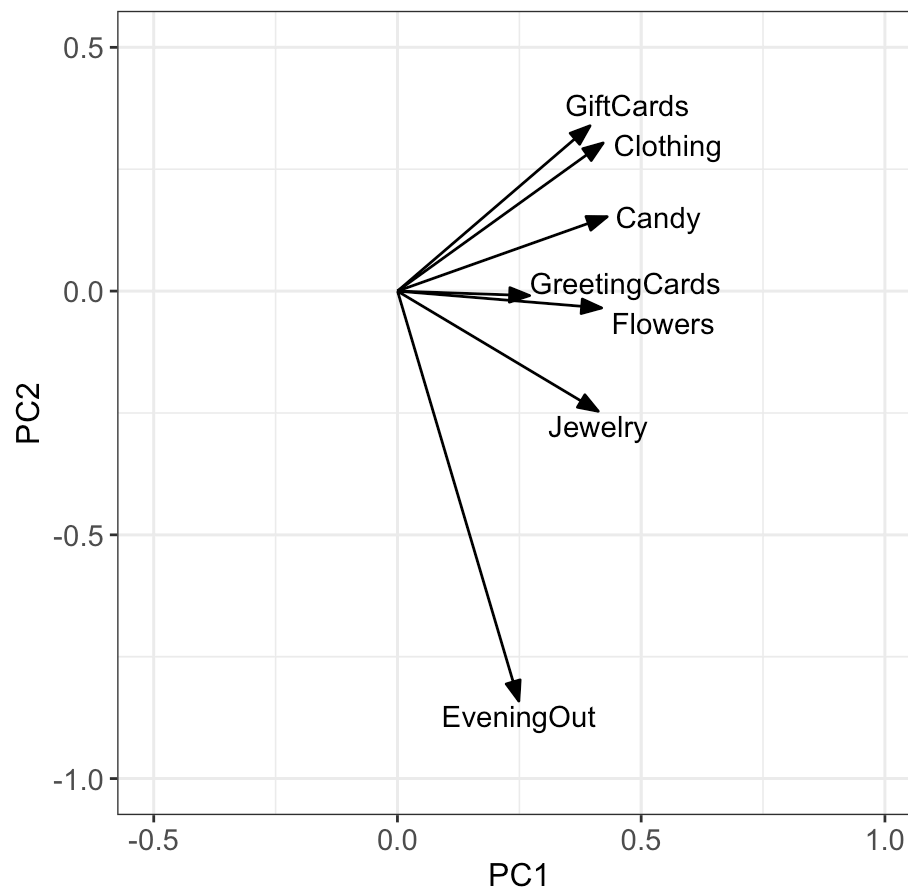
	PC6	PC7
## Candy	0.72246069	-0.21762253
## Flowers	-0.47956524	-0.43468347
## Jewelry	0.31628428	0.34599624
## GreetingCards	-0.04206019	0.07310556
## EveningOut	-0.12125426	-0.04429973
## Clothing	-0.33009971	0.68216411
## GiftCards	-0.15031469	-0.41391128

```

arrow_style <- arrow(
  angle = 20, length = grid::unit(8, "pt"),
  ends = "first", type = "closed"
)
pca_fit %>%
  tidy(matrix = "rotation") %>%
  pivot_wider(
    names_from = "PC", values_from = "value",
    names_prefix = "PC"
  ) %>%
  ggplot(aes(PC1, PC2)) +
  geom_segment(
    xend = 0, yend = 0,
    arrow = arrow_style
  ) +
  geom_text(aes(label = column), vjust = c(0.5, 1.2, 1.2, -0.1, 1.2, 0.6, -0.5), hjust =
c(-0.1, -0.1, 0.5, 0 , 0.5, -0.1, 0.2)) +
  xlim(-0.5, 1) + ylim(-1, 0.5) +
  coord_fixed() +
  ggtitle("Rotation Plot for Valentine Trends") +
  theme_bw() +
  theme(axis.title = element_text(size = 12), # Increase axis title size
        axis.text = element_text(size = 11), # Increase axis text size
        plot.title = element_text(size = 14, hjust = 0.5)) # Center plot title

```

Rotation Plot for Valentine Trends

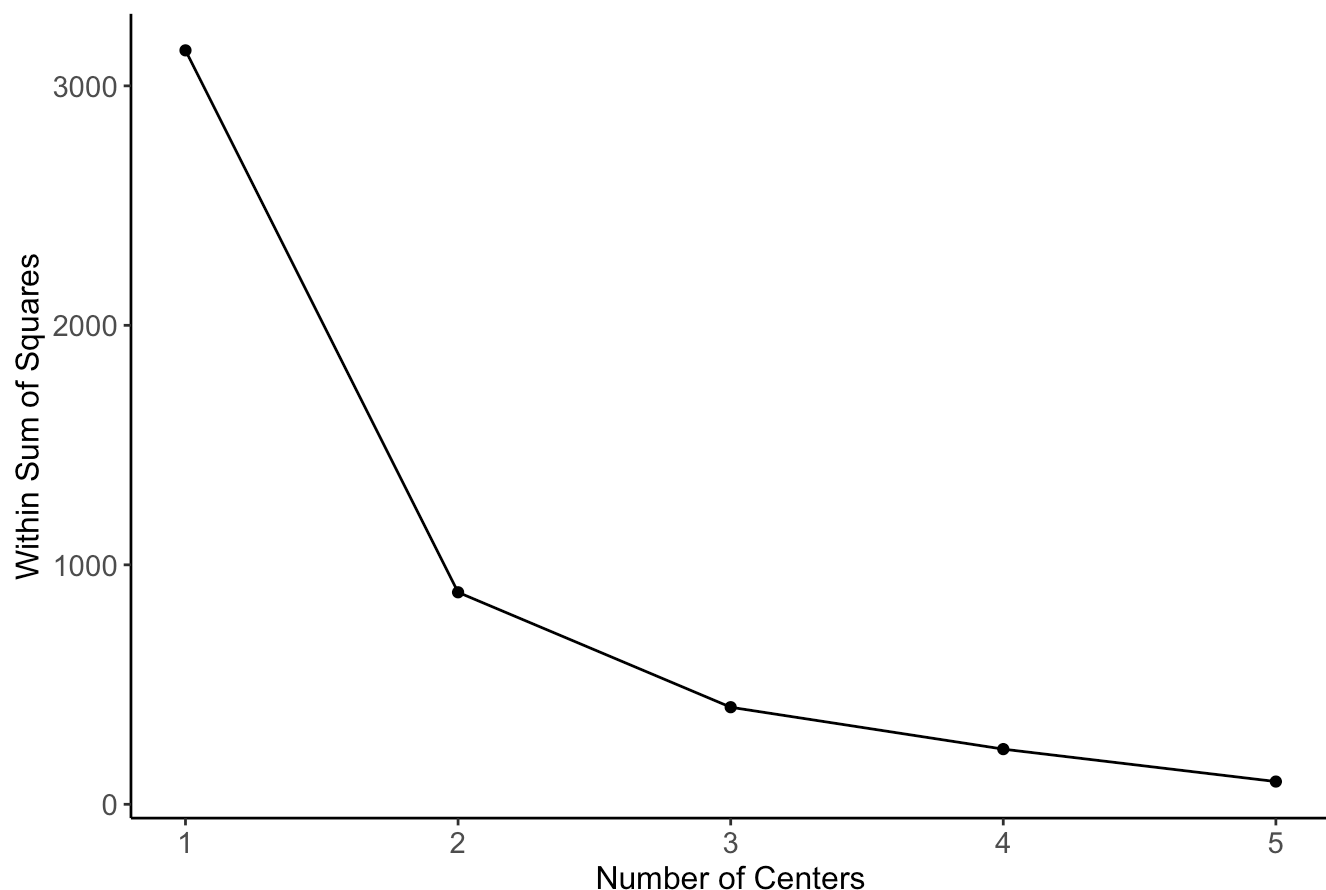


```

# elbow plot
calc_withinss <- function(data, centers) {
  km_fit <- select(data, where(is.numeric)) %>%
    kmeans(centers = centers, nstart = 10)
  km_fit$tot.withinss
}
tibble(centers = 1:5) %>%
  mutate(
    within_sum_squares = map_dbl(
      centers, ~calc_withinss(gifts_age, .x)
    )
  ) %>%
  ggplot() +
  aes(centers, within_sum_squares) +
  geom_point() +
  geom_line() +
  ggtitle("Elbow Plot for Valentine Trends based on Age") +
  scale_x_continuous(name = "Number of Centers") +
  scale_y_continuous(name = "Within Sum of Squares") +
  theme_classic() +
  theme(axis.title = element_text(size = 12), # Increase axis title size
        axis.text = element_text(size = 11), # Increase axis text size
        plot.title = element_text(size = 14, hjust = 0.5)) # Center plot title

```

Elbow Plot for Valentine Trends based on Age



```

# PCA for gifts_age
pca_fit2 <- gifts_age %>%
  select(-c(Age, SpendingCelebrating)) %>%
  scale() %>% # scale to zero mean and unit variance
  prcomp()
pca_data2 <- pca_fit2$x[, 1:2] # Use first two PCs

# Perform k-means with 2 clusters based on elbow plot
set.seed(132) # For reproducibility
km_fit2 <- kmeans(pca_data2, centers = 2, nstart = 25)

# scatter plot with Clusters
p1 <- km_fit2 %>%
  augment(pca_data2) %>%
  ggplot() +
  aes(x = PC1, y = PC2) +
  geom_point(
    aes(color = .cluster)
  ) +
  geom_point(
    data = tidy(km_fit2),
    aes(fill = cluster),
    shape = 21, color = "black", size = 4
  ) +
  guides(color = "none") +
  xlim(-4, 4) + ylim(-2, 2) +
  coord_fixed() +
  ggtitle("Clustering of Spending Behavior Profiles in PCA Space") +
  theme_bw() +
  theme(plot.title = element_text(size = 12, hjust = 0.5))

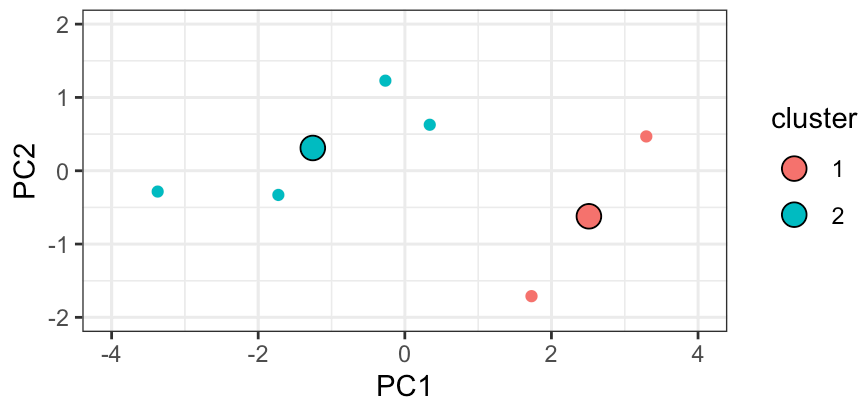
# scatter plot based on Age
p2 <- pca_fit2 %>%
  augment(gifts_age) %>%
  ggplot(aes(.fittedPC1, .fittedPC2)) +
  geom_point(aes(color = Age)) +
  scale_x_continuous(name = "PC1",
    limits = c(-4, 4)) +
  scale_y_continuous(name = "PC2",
    limits = c(-2, 2)) +
  scale_color_manual(name = NULL,
    values = c("65+" = "#04bbc3",
      "55-64" = "#04bbc3",
      "45-54" = "#04bbc3",
      "35-44" = "#04bbc3",
      "25-34" = "#e37a73",
      "18-24" = "#e37a73")) +
  geom_text(aes(label = Age, hjust = 1, vjust = -0.8), size = 3) +
  guides(color = "none") +
  coord_fixed() +
  ggtitle("Age Distribution in PCA Space Based on Spending Categories") +
  theme_bw() +

```

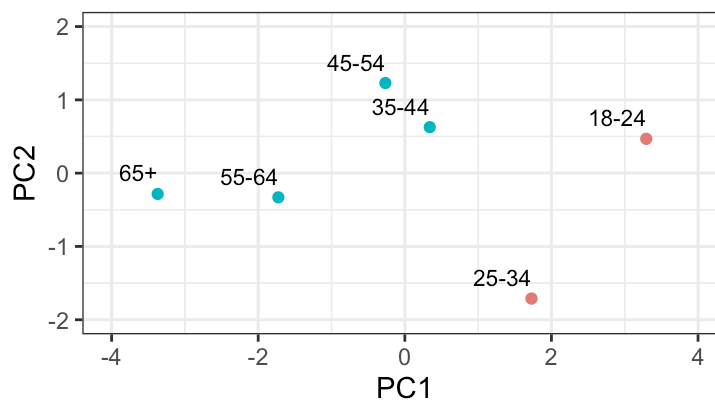
```
theme(plot.title = element_text(size = 12, hjust = 0.5))
```

p1/p2

Clustering of Spending Behavior Profiles in PCA Space



Age Distribution in PCA Space Based on Spending Categories



Discussion: The historical_spending dataset reveals Valentine's Day spending trends from 2010 to 2022. The percentage of people celebrating averages 55.46%, with spending per person ranging from \$103 to \$196 (average \$144.4). Jewelry has the highest average spending at 32.55%, followed by Evening Out (27.47%), while Greeting Cards have the lowest (7.676%). Candy spending averages 12.84%, Flowers 14.65%, Clothing 14.94%, and Gift Cards 11.50%, all showing modest variability. Individual spending reflects diverse budget preferences for Valentine's Day celebrations.

The Gifts_Age summary highlights spending patterns across categories and age groups. Candy and SpendingCelebrating have the highest spending, with Candy peaking at 59. Flowers spending averages 38.17, while Jewelry and Clothing show lower expenditures, with means of 22.83 and moderate maximums. GreetingCards and GiftCards have stable, predictable spending patterns, with narrow ranges and low variation. EveningOut reflects moderate spending with a mean of 32. Candy and SpendingCelebrating emerge as the most prioritized categories, while Clothing and Jewelry exhibit notably lower spending.

The combination chart of bar and scatter on Valentine's Day spending across different age groups reveals several key trends and insights. Spending generally decreases with age, with the 18-24 age group being the highest spenders, especially on candy and evening outings. The 25-34 age group follows closely, investing heavily in flowers and greeting cards. As age increases, spending shifts towards more traditional gifts, such as jewelry, gift cards, and clothing, particularly in the 55-64 age group, which has the highest overall spending across all categories. Younger groups (18-34) tend to prioritize experiences like evening outings, while middle-aged groups (35-54) show more balanced spending, investing in both traditional gifts like jewelry and experiences. Older age

groups (55+) show a more conservative spending pattern, with a continued preference for sentimental gifts like flowers and greeting cards, but lower overall expenditure. These trends suggest that younger individuals lean toward more dynamic and versatile gifts, while older age groups favor traditional, meaningful presents. The overall spending decreases with age, with the 65+ group being the lowest spenders.

The Rotation plot matrix visualizes the loadings of various Valentine's Day gift categories onto two principal components (PC1 and PC2), revealing key insights into spending patterns. PC1 represents "popularity" or "demand," with all categories showing positive loadings between 0 and 0.5. PC2 contrasts "traditional" gifts (like Jewelry) with "experiential" gifts (like Evening Out). Traditional gifts such as Gift Cards, Clothing, and Candy have high positive loadings on PC2, indicating their common exchange, while Jewelry and Evening Out show negative loadings. Categories like Greeting Cards and Flowers show moderate loadings, with a balanced influence on both PC1 and PC2, indicating consistent popularity. Evening Out is a strong influencer of overall spending, correlating positively with PC1 and playing a significant role in PC2. Overall, Valentine's Day gift preferences are shaped by two factors: general popularity (Gift Cards, Clothing, Candy) and sentiment/experience (Jewelry, Evening Out). These insights can guide targeted marketing strategies, emphasizing the influence of Gift Cards and Jewelry, while recognizing the steady appeal of Flowers and Greeting Cards.

The elbow plot suggests that splitting the data into 2 clusters provides a good balance between minimizing within-cluster variance and keeping the model simple. Beyond 2 clusters, the reduction in variance is less significant, indicating diminishing returns with additional clusters. The Clustering analysis identifies distinct spending behavior profiles based on gift categories, which correlate with different age groups. Younger individuals (ages 18-34) tend to be high spenders, particularly on experiences (Evening Out) and versatile gifts (Gift Cards), reflecting their preference for social activities and practical gifting solutions. This group is concentrated in Cluster 2, which shows lower spending on traditional categories. Middle-aged individuals (35-54) exhibit a moderate spending profile, balancing traditional gifts like Jewelry and Flowers with more substantial purchases. These groups are more spread across both clusters, indicating varied spending behaviors. Older individuals (55+) show a lower spending profile, focusing on traditional gifts such as Greeting Cards and Candy. This age group tends to cluster in Cluster 1, associated with higher spending on categories like Jewelry and Flowers. Overall, younger age groups spend more on experiences and versatile gifts, while older groups lean towards traditional and budget-friendly gifts. These insights reveal how distinct spending behavior profiles correlate with age, with variability present within each group. Further analysis could deepen the understanding by exploring specific spending category influences and individual characteristics.