Segmentation at TiVo

Alex Osterbuhr

Segmentation at TiVo

This analysis looks at customer segmentation at TiVo.

Data Preparation

Load, prepare, explore, and analyze the TiVo data.

```
# Read data
tivo_orig <- read_excel("TiVo Segmentation Analytics Survey Data_KEL934-XLS-ENG.xlsx",sheet=2)
tivo_orig <- slice(tivo_orig, 1:1000)
```

Question 3: Data Exploration

Create a table for each attribute (e.g., Gender) and record the percentage of responses for each answer (e.g., % of respondents answering “male” and % answering “female”).

```
# Gender
gender_dist <- round((prop.table(table(tivo_orig$Gender)))*100)

gender_tib <- tivo_orig %>% group_by(Gender) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig)))
gender_tib
```

```
## # A tibble: 2 × 2
##   Gender Distribution
##   <chr>  <chr>
## 1 female 46%
## 2 male   54%
```

```
# Marital Status
marital_dist <- round((prop.table(table(tivo_orig$`Marital Status`)))*100)

marital_tib <- tivo_orig %>% group_by(`Marital Status`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig)))
marital_tib
```

```
## # A tibble: 2 × 2
##   `Marital Status` Distribution
##   <chr>            <chr>
## 1 married          72%
## 2 single           28%
```

```
# Work Status
work_dist <- round((prop.table(table(tivo_orig$`Work Status`)))*100)

work_tib <- tivo_orig %>% group_by(`Work Status`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig)))
work_tib
```

```
## # A tibble: 2 × 2
##   `Work Status` Distribution
##   <chr>         <chr>
## 1 none          35%
## 2 professional  65%
```

```
# Education
ed_dist <- round((prop.table(table(tivo_orig$Education)))*100)

ed_tib <- tivo_orig %>% group_by(Education) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig),accuracy=.1))
ed_tib
```

```
## # A tibble: 4 × 2
##   Education Distribution
##   <chr>     <chr>
## 1 BA        26.0%
## 2 MA        12.4%
## 3 none      50.6%
## 4 PhD       11.0%
```

```
# Annual Income
income_dist <- round((prop.table(table(tivo_orig$`Annual Income (x1000 $)`)))*100, 2)

income_tib <- tivo_orig %>%
  group_by(`Annual Income (x1000 $)`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig),accuracy=.1))
income_tib
```

```
## # A tibble: 45 × 2
##    `Annual Income (x1000 $)` Distribution
##                        <dbl> <chr>
##  1                        21 0.1%
##  2                        22 0.2%
##  3                        23 0.4%
##  4                        24 1.8%
##  5                        25 1.6%
##  6                        26 4.7%
##  7                        27 5.5%
##  8                        28 7.5%
##  9                        29 7.3%
## 10                        30 8.0%
## # … with 35 more rows
```

```
# Age
age_dist <- round((prop.table(table(tivo_orig$Age)))*100, 2)

age_tib <- tivo_orig %>%
  group_by(Age) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig),accuracy=.1))
age_tib
```

```
## # A tibble: 63 × 2
##      Age Distribution
##    <dbl> <chr>
##  1    18 0.6%
##  2    19 0.9%
##  3    20 1.5%
##  4    21 1.2%
##  5    22 2.2%
##  6    23 2.0%
##  7    24 1.6%
##  8    25 2.5%
##  9    26 2.5%
## 10    27 1.8%
## # … with 53 more rows
```

```
# Location
loc_dist <- round((prop.table(table(tivo_orig$Location)))*100, 2)

loc_tib <- tivo_orig %>%
  group_by(Location) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig),accuracy=.1))
loc_tib
```

```
## # A tibble: 50 × 2
##    Location    Distribution
##    <chr>       <chr>
##  1 Alabama     0.6%
##  2 Alaska      0.5%
##  3 Arizona     4.3%
##  4 Arkansas    0.4%
##  5 California  4.1%
##  6 Colorado    3.9%
##  7 Connecticut 3.7%
##  8 Delaware    4.5%
##  9 Florida     3.5%
## 10 Georgia     2.7%
## # … with 40 more rows
```

```
# Purchasing decision maker
decide_dist <- round((prop.table(table(tivo_orig$`Purchasing Decision-maker`)))*100, 2)

decide_tib <- tivo_orig %>%
  group_by(`Purchasing Decision-maker`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig)))
decide_tib
```

```
## # A tibble: 2 × 2
##   `Purchasing Decision-maker` Distribution
##   <chr>                       <chr>
## 1 family                      56%
## 2 single                      44%
```

```
# Purchasing location
purchase_loc_dist <- round((prop.table(table(tivo_orig$`Purchasing Location`)))*100, 2)

purchase_loc_tib <- tivo_orig %>%
  group_by(`Purchasing Location`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig),accuracy=.1))
purchase_loc_tib
```

```
## # A tibble: 5 × 2
##   `Purchasing Location`     Distribution
##   <chr>                     <chr>
## 1 discount                  29.3%
## 2 mass-consumer electronics 20.0%
## 3 retail                    29.4%
## 4 specialty stores          17.0%
## 5 web (ebay)                4.3%
```

```
# Monthly electronics spend
electronics_mo_dist <- round((prop.table(table(tivo_orig$`Monthly Electronics Spend`)))*100, 2)

electronics_mo_tib <- tivo_orig %>%
  group_by(`Monthly Electronics Spend`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig),accuracy=.1))
electronics_mo_tib
```

```
## # A tibble: 77 × 2
##    `Monthly Electronics Spend` Distribution
##                          <dbl> <chr>
##  1                           7 0.1%
##  2                           9 0.2%
##  3                          10 0.8%
##  4                          11 0.7%
##  5                          12 2.1%
##  6                          13 2.9%
##  7                          14 4.4%
##  8                          15 5.5%
##  9                          16 5.7%
## 10                          17 5.5%
## # … with 67 more rows
```

```
# Purchasing Frequency
purchase_freq_dist <- round((prop.table(table(tivo_orig$`Purchasing Frequency (every x months)
`)))*100, 2)

purchase_freq_tib <- tivo_orig %>%
  group_by(`Purchasing Frequency (every x months)`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig),accuracy=.1))
purchase_freq_tib
```

```
## # A tibble: 48 × 2
##    `Purchasing Frequency (every x months)` Distribution
##                                      <dbl> <chr>
##  1                                       1 1.6%
##  2                                       2 2.6%
##  3                                       3 2.7%
##  4                                       4 2.6%
##  5                                       5 2.8%
##  6                                       6 1.6%
##  7                                       7 2.3%
##  8                                       8 2.4%
##  9                                       9 2.7%
## 10                                      10 2.8%
## # … with 38 more rows
```

```
# tech adoption
tech_dist <- round((prop.table(table(tivo_orig$`Technology Adoption`)))*100, 2)

tech_tib <- tivo_orig %>%
  group_by(`Technology Adoption`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig)))
tech_tib
```

```
## # A tibble: 2 × 2
##   `Technology Adoption` Distribution
##   <chr>                 <chr>
## 1 early                 80%
## 2 late                  20%
```

```
# TV viewing
tv_dist <- round((prop.table(table(tivo_orig$`TV Viewing (hours/day)`)))*100, 2)

tv_tib <- tivo_orig %>%
  group_by(`TV Viewing (hours/day)`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig),accuracy=.1))
tv_tib
```

```
## # A tibble: 15 × 2
##    `TV Viewing (hours/day)` Distribution
##                       <dbl> <chr>
##  1                        0 17.5%
##  2                        1 36.2%
##  3                        2 23.2%
##  4                        3 3.2%
##  5                        4 4.5%
##  6                        5 4.4%
##  7                        6 2.2%
##  8                        7 0.9%
##  9                        8 1.0%
## 10                        9 0.9%
## 11                       10 1.5%
## 12                       11 2.0%
## 13                       12 0.7%
## 14                       13 1.4%
## 15                       14 0.4%
```

```
# Favorite feature
fav_dist <- round((prop.table(table(tivo_orig$`Favorite feature`)))*100, 2)

fav_tib <- tivo_orig %>%
  group_by(`Favorite feature`) %>%
  summarize(Distribution=percent(n()/nrow(tivo_orig),accuracy=.1))
fav_tib
```

```
## # A tibble: 5 × 2
##   `Favorite feature`                          Distribution
##   <chr>                                       <chr>
## 1 cool gadget                                 22.8%
## 2 programming/interactive features            13.0%
## 3 saving favorite shows to watch as a family 20.0%
## 4 schedule control                            22.1%
## 5 time shifting                               22.1%
```

Question 4

Answer the following questions.

4a. In the next two years, how many married men who are early adopters can afford to purchase a TiVo for $499, have enough money to purchase another electronic gadget, and are likely to do so?

94 early adopting married men can and are likely to purchase another electronic

```
q4a <- tivo_orig %>%
  filter(Gender=="male" & `Marital Status`== "married" & `Technology Adoption`== "early" & (`Mon
thly Electronics Spend`*24)>499 & `Purchasing Frequency (every x months)` <=12)

count(q4a)
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    94
```

4b. How many women with education of MA or PhD are making purchasing decisions for electronics without discussing them with a spouse, either because they are single, or because they are making purchasing decisions without the involvement of their spouses?

55 females with post-graduate education are the sole decision makers whether married or single.

```
q4b <- tivo_orig %>% filter(Gender=='female' & Education %in% c('PhD', 'MA') & `Purchasing Decis
ion-maker`== 'single')

count(q4b)
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    55
```

4c. Among early adopters, how many purchase electronics at least once every year and do so in stores that specialize in electronics?

132 early adopters purchase at least once a year from specialty stores

```
q4c <- tivo_orig %>% filter(`Technology Adoption`=='early' & `Purchasing Frequency (every x mont
hs)`<=12 & `Purchasing Location`=='specialty stores')

count(q4c)
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1   132
```

4d. How many seniors (above the age of 65) spend more than six hours a day watching TV?

20 seniors that are 66 and older spend more that 6 hours per day watching TV.

```
q4d <- tivo_orig %>% filter(Age>65 & `TV Viewing (hours/day)`>6)

count(q4d)
```

```
## # A tibble: 1 × 1
##       n
##   <int>
## 1    20
```

4e. What is the income range for seniors (above the age of 65) who spend more than six hours a day watching TV?

The range of income for the people in 4d is $41,000-$55,000.

```
range(q4d$`Annual Income (x1000 $)`)
```

```
## [1] 41 55
```

4f. What is the average annual income for seniors (above the age

of 65) who spend more than six hours a day watching TV?

The average income for the people in 4d is $48,600

```
mean(q4d$`Annual Income (x1000 $)`)
```

```
## [1] 48.6
```

Question 5: Correlate Annual Income with Age

What is the R^2? R^2 is .01754705

```
# R^2 of annual income and age
income_age <- lm(`Annual Income (x1000 $)` ~ Age, data=tivo_orig)

summary(income_age)$r.squared
```

```
## [1] 0.01754705
```

Question 6: Correlate Gender and Annual Income

The R^2 is .004552555

```
# R^2 of annual income and gender
income_gender <- lm(`Annual Income (x1000 $)` ~ Gender, data=tivo_orig)

summary(income_gender)$r.squared
```

```
## [1] 0.004552555
```

Question 7: Correlation Analysis, Continued

7a: Age and Purchasing Frequency

The R^2 is 0.00000167086

```
# R^2
age_purchasefreq <- lm(Age ~ `Purchasing Frequency (every x months)`, data=tivo_orig)

summary(age_purchasefreq)$r.squared
```

```
## [1] 0.00000167086
```

7b: Annual Income and TV Viewing

The R^2 is 0.007112434

```
# R^2
income_TVviewing <- lm(`Annual Income (x1000 $)` ~ `TV Viewing (hours/day)`, data=tivo_orig)

summary(income_TVviewing)$r.squared
```

```
## [1] 0.007112434
```

7c: Education and Favorite Feature

The R^2 is 0.2136032

```
# R^2

numEds <- tivo_orig %>%
  mutate(numEducation=recode(Education,'none'=0, 'BA'=1, 'MA'=2, 'PhD'=3))
ed_feat <- lm(numEducation ~ `Favorite feature`, data=numEds)

summary(ed_feat)$r.squared
```

```
## [1] 0.2136032
```

7d: Monthly Electronics Spend and Monthly Household Spend

The R^2 is 0.6461037

```
# R^2
mo_electro_house <- lm(`Monthly Household Spend` ~ `Monthly Electronics Spend`, data=tivo_orig)

summary(mo_electro_house)$r.squared
```

```
## [1] 0.6461037
```

7e: Of the four correlations, are any high enough to make one of the attributes redundant?

The only moderately high r-squared value is that in part 7d (0.6461037). Considering the field this data comes from, though, I would consider this value acceptable as over .7 is widely considered the start of the high threshold. Therefore, it could be considered high enough to make one of the attributes redundant, but I dont believe it is given the context of the data.

Question 8: Segmentation

Select one or more attributes to use as the basis for generating two segmentation schemes. For example, you could choose Age as the basis for creating segments based on age ranges. Or you could select two attributes that would help you segment by willingness to buy versus ability to pay.

As you generate segmentation schemes, apply it to the provided data to describe market size, average annual income, most appealing feature(s)/benefit(s), purchase location(s), and average electronics purchase. Remember to try to create segments that are homogenous internally and heterogeneous across segments for the indicated number of segments. Not all attributes will be used.

8a: Segmentation Scheme A

```
Segment_A <- tivo_orig %>%
  group_by(Education) %>%
  summarise(
    Market_Size = paste0(round(n() / nrow(tivo_orig) * 100, 2), "%"),
    Avg_Annual_Income_Thousands = round(mean(`Annual Income (x1000 $)`, na.rm = TRUE), 1),
    Most_Appealing_Feature = names(sort(table(`Favorite feature`), decreasing = TRUE)[1]),
    Location_Most_Purchased = names(sort(table(`Purchasing Location`), decreasing = TRUE)[1]),
    Avg_Mo_Elec_Spend = round(mean(`Monthly Electronics Spend`), 1)
  )
Segment_A
```

```
## # A tibble: 4 × 6
##   Education Market_Size Avg_Annual_Income_Thousands Most_Appea…¹ Locat…² Avg_M…³
##   <chr>     <chr>                             <dbl> <chr>        <chr>     <dbl>
## 1 BA        26%                                45.9 cool gadget  specia…    40.2
## 2 MA        12.4%                              42.5 saving favo… mass-c…    40.4
## 3 none      50.6%                              34.1 schedule co… discou…    22.3
## 4 PhD       11%                                41.6 saving favo… mass-c…    38.3
## # … with abbreviated variable names ¹Most_Appealing_Feature,
## #   ²Location_Most_Purchased, ³Avg_Mo_Elec_Spend
```

Market size (% of TV-involved households) Average annual income Description of segment Most appealing feature/benefit Stores shopped for electronics Average electronics purchase ($)

8b: Segmentation Scheme B

```
Segment_B <- tivo_orig %>%
  group_by(Gender,`Work Status`) %>%
  summarise(
    Market_Size = paste0(round(n() / nrow(tivo_orig) * 100, 2), "%"),
    Avg_Annual_Income_Thousands = round(mean(`Annual Income (x1000 $)`, na.rm = TRUE), 1),
    Most_Appealing_Feature = names(sort(table(`Favorite feature`), decreasing = TRUE)[1]),
    Location_Most_Purchased = names(sort(table(`Purchasing Location`), decreasing = TRUE)[1]),
    Avg_Mo_Elec_Spend = round(mean(`Monthly Electronics Spend`), 1)
  )
```

```
## `summarise()` has grouped output by 'Gender'. You can override using the
## `.groups` argument.
```

```
Segment_B
```

```
## # A tibble: 4 × 7
## # Groups:   Gender [2]
##   Gender `Work Status` Market_Size Avg_Annual_Income_T…¹ Most_…² Locat…³ Avg_M…⁴
##   <chr>  <chr>         <chr>                       <dbl> <chr>   <chr>     <dbl>
## 1 female none          17.2%                        35.7 saving… retail     24.8
## 2 female professional  29.3%                        37.4 progra… retail     31.1
## 3 male   none          17.8%                        34.1 schedu… discou…    23.4
## 4 male   professional  35.7%                        44.4 cool g… specia…    37.6
## # … with abbreviated variable names ¹Avg_Annual_Income_Thousands,
## #   ²Most_Appealing_Feature, ³Location_Most_Purchased, ⁴Avg_Mo_Elec_Spend
```

Market size (% of TV-involved householdssh3r3) Average annual income Description of segment Most appealing feature/benefit Stores shopped for electronics Average electronics purchase ($)

Question 9

Write a 150-word summary of your selected segmentation scheme and how you arrived at the segmentation. Include descriptive names for your segments that would be understandable to the marketing manager at TiVo.

Initially, I wanted to see how much of our customer base have completed a level of college education and where they are buying their electronics. Astoundingly, 50.6% of the potential customer base have no completed college education as seen in Segment A. Unfortunately, they also have a substantial amount less average monthly spending on electronics at $22.30, with the next closest being those with a PhD (11% of our market) at $38.3. By grouping customers using their education levels in Segment A, we can clearly see our most likely clients, clients we can do better at reaching out to, and where they shop. In Segment B, I wanted to take a similar approach but look at whether customers are employed or not split by gender. This allows us to create a marketing plan that could reach out to specific genders where they work. There are differences between employed and unemployed customers, but the most striking revelation from this segment is that the average income across all four groupings is roughly the same at the mid $30k's. Employed males are the exception, but not by much at $44k. To add to this, even though income is similar, unemployed customers are spending less.

Question 10

As the analyst, what questions would have you have for TiVo to aid you in your analysis?

First, I am curious where this data comes from, although opening the full csv may give me this information instead of just sheet 2. I would like to know how this data was collected, whether by random in-person survey or collected digitally. I would also like to know specifically what problem are they trying to solve with this data so that I could segment it accordingly and conduct value-added analysis.