

## KPMG Data Analytics using python.

### The Sporty Rocket Central Dataset

Developing a report that we can present to the client at our next meeting. Displaying the data summary and results of the analysis (see tools/references for assistance). Specifically, the presentation should specify who Sprocket Central Pty Ltd' should be targeting out of the new 1000 customer list using the Transactions Table. **Problem Outline** SP Rocket Central is a company that specializes in high quality bikes and accessories. the Company is targeting 1000 new customers and is focused in Maximizing profit through Bike sales.

```
pacman::p_load(pacman, dplyr, GGally, ggplot2, ggthemes,
               ggvis, httr, lubridate, plotly, rio, rmarkdown, shiny,
               stringr, tidyr, reticulate, tidyverse, psych)
```

```
TransactionsTable <- read_csv("C:/Users/user/Projects/BI-Analytics-Projects/KPMG Data Analytics (Relati
```

```
## Rows: 16720 Columns: 30
## -- Column specification -----
## Delimiter: ","
## chr (17): transaction_date, order_status, brand, product_line, product_class...
## dbl (12): transaction_id, product_id, customer_id, Recency, list_price, stan...
## lgl (1): online_order
##
## 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.
```

```
kpmg <- TransactionsTable
```

```
kpmg
```

```
## # A tibble: 16,720 x 30
##   transaction_id product_id customer_id transaction_date Recency online_order
##         <dbl>      <dbl>      <dbl> <chr>              <dbl> <lgl>
## 1             1          2        2950 2/25/2017          308 FALSE
## 2             2          3        3120 5/21/2017          223 TRUE
## 3             3         37         402 10/16/2017           75 FALSE
## 4             4         88        3135 8/31/2017          121 FALSE
## 5             6         25        2339 3/8/2017           297 TRUE
## 6             7         22        1542 4/21/2017          253 TRUE
## 7             8         15        2459 7/15/2017          168 FALSE
## 8             9         67        1305 8/10/2017          142 FALSE
## 9            10         12        3262 8/30/2017          122 TRUE
## 10           12         61        2783 1/5/2017           359 TRUE
## # i 16,710 more rows
## # i 24 more variables: order_status <chr>, brand <chr>, product_line <chr>,
## #   product_class <chr>, product_size <chr>, list_price <dbl>,
## #   standard_cost <dbl>, product_first_sold_date <chr>, Profit <dbl>,
## #   gender <chr>, past_3_years_bike_related_purchases <dbl>, DOB <chr>,
## #   AGE <dbl>, job_title <chr>, job_industry_category <chr>,
## #   wealth_segment <chr>, deceased_indicator <chr>, owns_car <chr>, ...
```

```
str(kpmg)
```

```
## spc_tbl_ [16,720 x 30] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ transaction_id      : num [1:16720] 1 2 3 4 6 7 8 9 10 12 ...
## $ product_id          : num [1:16720] 2 3 37 88 25 22 15 67 12 61 ...
## $ customer_id         : num [1:16720] 2950 3120 402 3135 2339 ...
## $ transaction_date    : chr [1:16720] "2/25/2017" "5/21/2017" "10/16/2017" "8/31/2017" ...
## $ Recency             : num [1:16720] 308 223 75 121 297 253 168 142 122 359 ...
## $ online_order        : logi [1:16720] FALSE TRUE FALSE FALSE TRUE TRUE ...
## $ order_status        : chr [1:16720] "Approved" "Approved" "Approved" "Approved" ...
## $ brand               : chr [1:16720] "Solex" "Trek Bicycles" "OHM Cycles" "Norco Bicycles" ...
## $ product_line        : chr [1:16720] "Standard" "Standard" "Standard" "Standard" ...
## $ product_class       : chr [1:16720] "medium" "medium" "low" "medium" ...
## $ product_size        : chr [1:16720] "medium" "large" "medium" "medium" ...
## $ list_price          : num [1:16720] 71.5 2091.5 1793.4 1198.5 1539 ...
## $ standard_cost       : num [1:16720] 53.6 388.9 248.8 381.1 829.6 ...
## $ product_first_sold_date : chr [1:16720] "12/2/2012" "3/3/2014" "7/20/1999" "12/16/1998" ...
## $ Profit              : num [1:16720] 17.9 1702.5 1544.6 817.4 709.3 ...
## $ gender              : chr [1:16720] "Male" "Female" "Male" "Male" ...
## $ past_3_years_bike_related_purchases : num [1:16720] 19 89 9 83 3 56 67 97 65 14 ...
## $ DOB                 : chr [1:16720] "1/11/1955" "2/4/1979" "6/3/1977" "1/14/1962" ...
## $ AGE                 : num [1:16720] 69 45 46 62 65 45 38 54 64 45 ...
## $ job_title           : chr [1:16720] "Software Engineer I" "Clinical Specialist" "Data Analyst" ...
## $ job_industry_category : chr [1:16720] "Financial Services" "Health" "Retail" "Financial" ...
## $ wealth_segment      : chr [1:16720] "Mass Customer" "Mass Customer" "Affluent Customer" ...
## $ deceased_indicator   : chr [1:16720] "N" "N" "N" "N" ...
## $ owns_car            : chr [1:16720] "Yes" "Yes" "No" "No" ...
## $ tenure              : num [1:16720] 10 10 22 16 16 12 18 6 12 7 ...
## $ address             : chr [1:16720] "984 Hoepker Court" "4 Shopko Circle" "586 Milford Ave" ...
## $ postcode            : num [1:16720] 3064 2196 2835 2096 2153 ...
## $ state               : chr [1:16720] "VIC" "NSW" "NSW" "NSW" ...
## $ country              : chr [1:16720] "Australia" "Australia" "Australia" "Australia" ...
## $ property_valuation   : num [1:16720] 6 5 1 10 10 10 6 4 8 9 ...
## - attr(*, "spec")=
## .. cols(
## ..   transaction_id = col_double(),
## ..   product_id = col_double(),
## ..   customer_id = col_double(),
## ..   transaction_date = col_character(),
## ..   Recency = col_double(),
## ..   online_order = col_logical(),
## ..   order_status = col_character(),
## ..   brand = col_character(),
## ..   product_line = col_character(),
## ..   product_class = col_character(),
## ..   product_size = col_character(),
## ..   list_price = col_double(),
## ..   standard_cost = col_double(),
## ..   product_first_sold_date = col_character(),
## ..   Profit = col_double(),
## ..   gender = col_character(),
## ..   past_3_years_bike_related_purchases = col_double(),
## ..   DOB = col_character(),
```

```
## .. AGE = col_double(),
## .. job_title = col_character(),
## .. job_industry_category = col_character(),
## .. wealth_segment = col_character(),
## .. deceased_indicator = col_character(),
## .. owns_car = col_character(),
## .. tenure = col_double(),
## .. address = col_character(),
## .. postcode = col_double(),
## .. state = col_character(),
## .. country = col_character(),
## .. property_valuation = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

`summary(kpmg)`

```
## transaction_id    product_id    customer_id    transaction_date
## Min.   :      1    Min.   : 0.00    Min.   :      1    Length:16720
## 1st Qu.: 4994    1st Qu.: 18.00    1st Qu.: 853    Class :character
## Median : 9942    Median : 45.00    Median :1732    Mode  :character
## Mean   : 9977    Mean   : 45.68    Mean   :1728
## 3rd Qu.:14963    3rd Qu.: 72.00    3rd Qu.:2592
## Max.   :20000    Max.   :100.00    Max.   :5034
## Recency          online_order    order_status          brand
## Min.   : 0.0    Mode :logical    Length:16720          Length:16720
## 1st Qu.: 89.0    FALSE:8322      Class :character      Class :character
## Median :181.0    TRUE :8398      Mode  :character      Mode  :character
## Mean   :181.7
## 3rd Qu.:273.0
## Max.   :363.0
## product_line      product_class    product_size          list_price
## Length:16720      Length:16720      Length:16720          Min.   : 12.01
## Class :character  Class :character  Class :character      1st Qu.: 575.27
## Mode  :character  Mode  :character  Mode  :character      Median :1163.89
##                                     Mean   :1109.26
##                                     3rd Qu.:1635.30
##                                     Max.   :2091.47
## standard_cost      product_first_sold_date    Profit          gender
## Min.   : 7.21    Length:16720          Min.   : 4.8    Length:16720
## 1st Qu.: 215.14    Class :character      1st Qu.: 135.8    Class :character
## Median : 507.58    Mode  :character      Median : 445.2    Mode  :character
## Mean   : 557.38
## 3rd Qu.: 795.10
## Max.   :1759.85
##                                     Mean   : 551.9
##                                     3rd Qu.: 827.2
##                                     Max.   :1702.5
## past_3_years_bike_related_purchases    DOB          AGE
## Min.   : 0.00          Length:16720    Min.   :22.00
## 1st Qu.:25.00          Class :character 1st Qu.:37.00
## Median :49.00          Mode  :character Median :46.00
## Mean   :49.54
## 3rd Qu.:74.00
## Max.   :99.00
##                                     Mean   :46.19
##                                     3rd Qu.:55.25
##                                     Max.   :92.00
## job_title          job_industry_category    wealth_segment    deceased_indicator
## Length:16720      Length:16720          Length:16720      Length:16720
```

```
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      owns_car          tenure          address          postcode
## Length:16720      Min.    : 1.00      Length:16720      Min.    :2000
## Class :character   1st Qu.: 6.00      Class :character   1st Qu.:2200
## Mode  :character   Median  :11.00      Mode  :character   Median  :2767
##                      Mean    :10.66                      Mean    :2991
##                      3rd Qu.:15.00                      3rd Qu.:3756
##                      Max.    :22.00                      Max.    :4883
##      state          country          property_valuation
## Length:16720      Length:16720      Min.    : 1.000
## Class :character   Class :character   1st Qu.: 6.000
## Mode  :character   Mode  :character   Median   : 8.000
##                      Mean    : 7.513
##                      3rd Qu.:10.000
##                      Max.    :12.000
```

```
colSums(is.na(kpmg))
```

```
##      transaction_id          product_id
##                      0                      0
##      customer_id          transaction_date
##                      0                      0
##      Recency          online_order
##                      0                      0
##      order_status          brand
##                      0                      0
##      product_line          product_class
##                      0                      0
##      product_size          list_price
##                      0                      0
##      standard_cost          product_first_sold_date
##                      0                      0
##      Profit          gender
##                      0                      0
## past_3_years_bike_related_purchases          DOB
##                      0                      0
##      AGE          job_title
##                      0                      0
##      job_industry_category          wealth_segment
##                      0                      0
##      deceased_indicator          owns_car
##                      0                      0
##      tenure          address
##                      0                      16513
##      postcode          state
##                      0                      0
##      country          property_valuation
##                      0                      0
```

```
kpmg <- subset(kpmg, select = -c(address))
colSums(is.na(kpmg))
```

```
##          transaction_id          product_id
##              0              0
##          customer_id      transaction_date
##              0              0
##          Recency          online_order
##              0              0
##          order_status          brand
##              0              0
##          product_line      product_class
##              0              0
##          product_size      list_price
##              0              0
##          standard_cost      product_first_sold_date
##              0              0
##          Profit          gender
##              0              0
## past_3_years_bike_related_purchases      DOB
##              0              0
##          AGE          job_title
##              0              0
##          job_industry_category      wealth_segment
##              0              0
##          deceased_indicator      owns_car
##              0              0
##          tenure      postcode
##              0              0
##          state      country
##              0              0
##          property_valuation
##              0
```

```
summary(kpmg$AGE)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      22.00   37.00   46.00   46.19   55.25   92.00
```

```
kpmg <- kpmg |> mutate(AgeGroup = cut(AGE, breaks = 15))
kpmg |> count(AgeGroup)
```

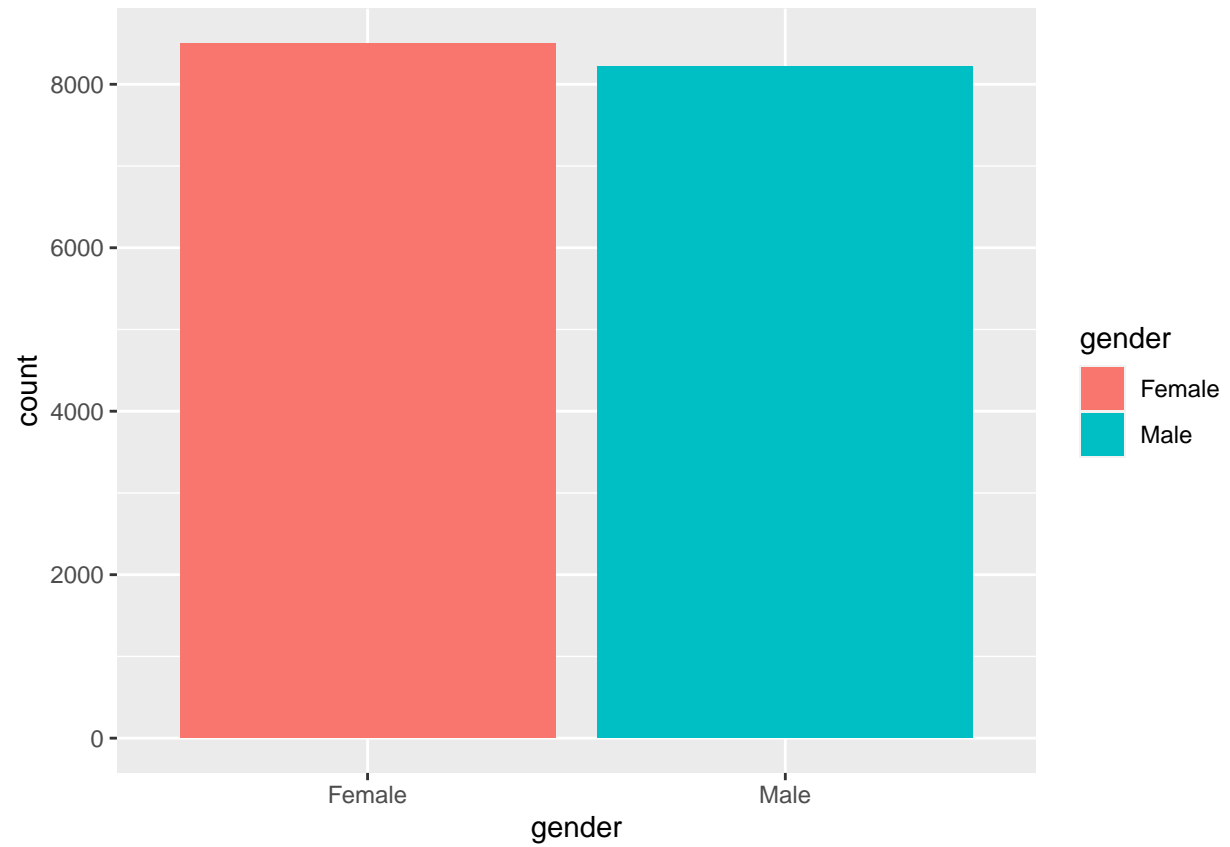
```
## # A tibble: 13 x 2
##   AgeGroup      n
##   <fct>    <int>
## 1 (21.9,26.7] 1136
## 2 (26.7,31.3] 1603
## 3 (31.3,36]   1413
## 4 (36,40.7]   1183
## 5 (40.7,45.3] 2225
## 6 (45.3,50]   3524
```

```
## 7 (50,54.7] 1178
## 8 (54.7,59.3] 1533
## 9 (59.3,64] 1359
## 10 (64,68.7] 1104
## 11 (68.7,73.3] 435
## 12 (78,82.7] 12
## 13 (87.3,92.1] 15
```

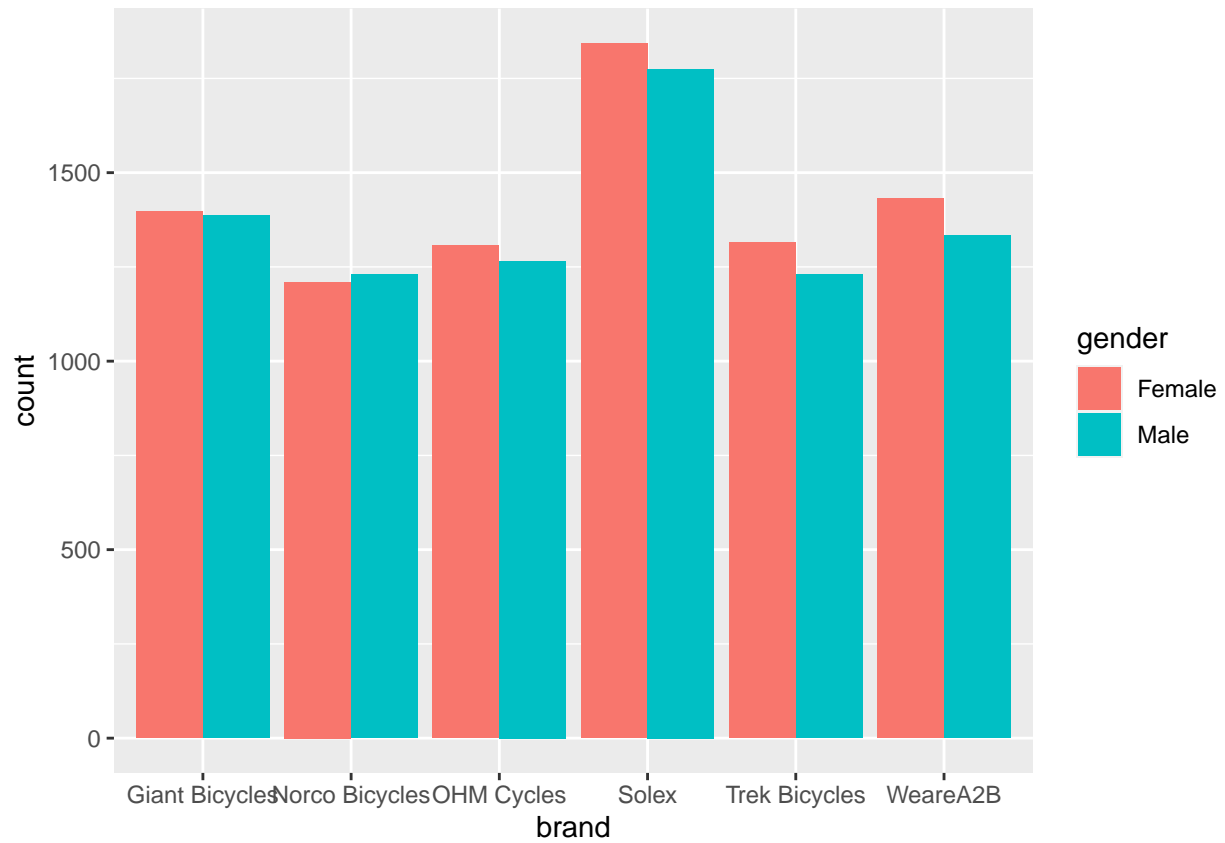
```
colnames(kpmg)
```

```
## [1] "transaction_id"      "product_id"
## [3] "customer_id"         "transaction_date"
## [5] "Recency"             "online_order"
## [7] "order_status"        "brand"
## [9] "product_line"        "product_class"
## [11] "product_size"        "list_price"
## [13] "standard_cost"       "product_first_sold_date"
## [15] "Profit"              "gender"
## [17] "past_3_years_bike_related_purchases" "DOB"
## [19] "AGE"                 "job_title"
## [21] "job_industry_category" "wealth_segment"
## [23] "deceased_indicator"  "owns_car"
## [25] "tenure"              "postcode"
## [27] "state"               "country"
## [29] "property_valuation"  "AgeGroup"
```

```
ggplot(kpmg, aes(x=gender,fill=gender)) + geom_bar()
```

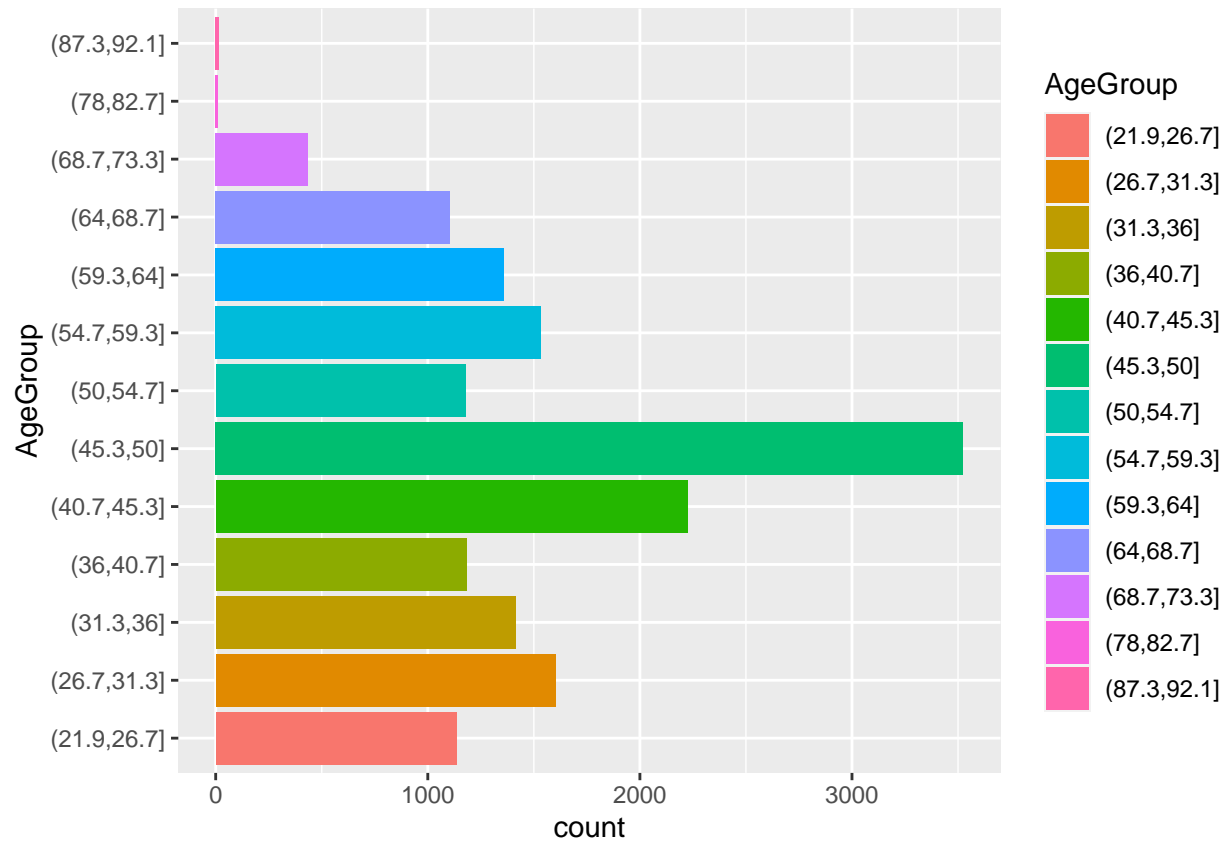


```
ggplot(kpmg, aes(x=brand,fill=gender)) + geom_bar(position = 'dodge')
```

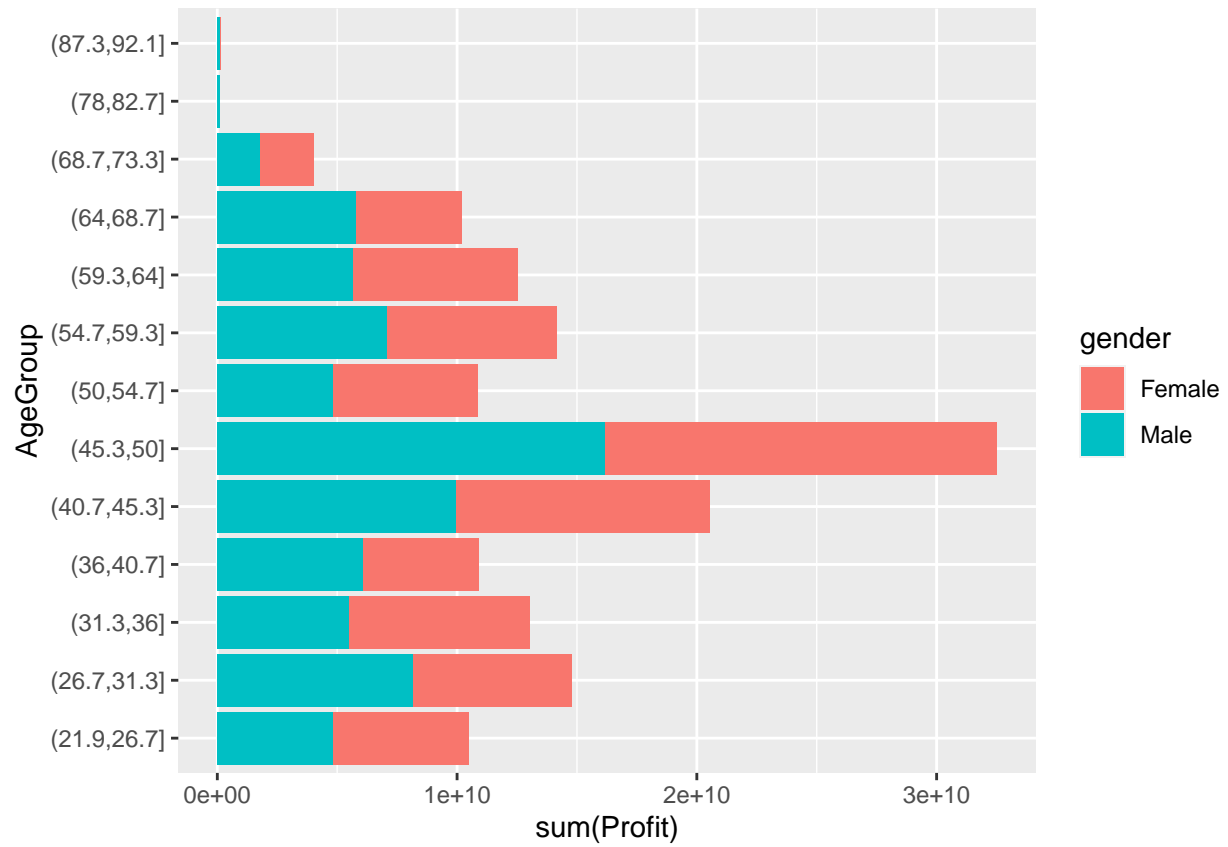


```
ggplot(kpmg, aes(y=AgeGroup,fill=AgeGroup)) + geom_bar()
```

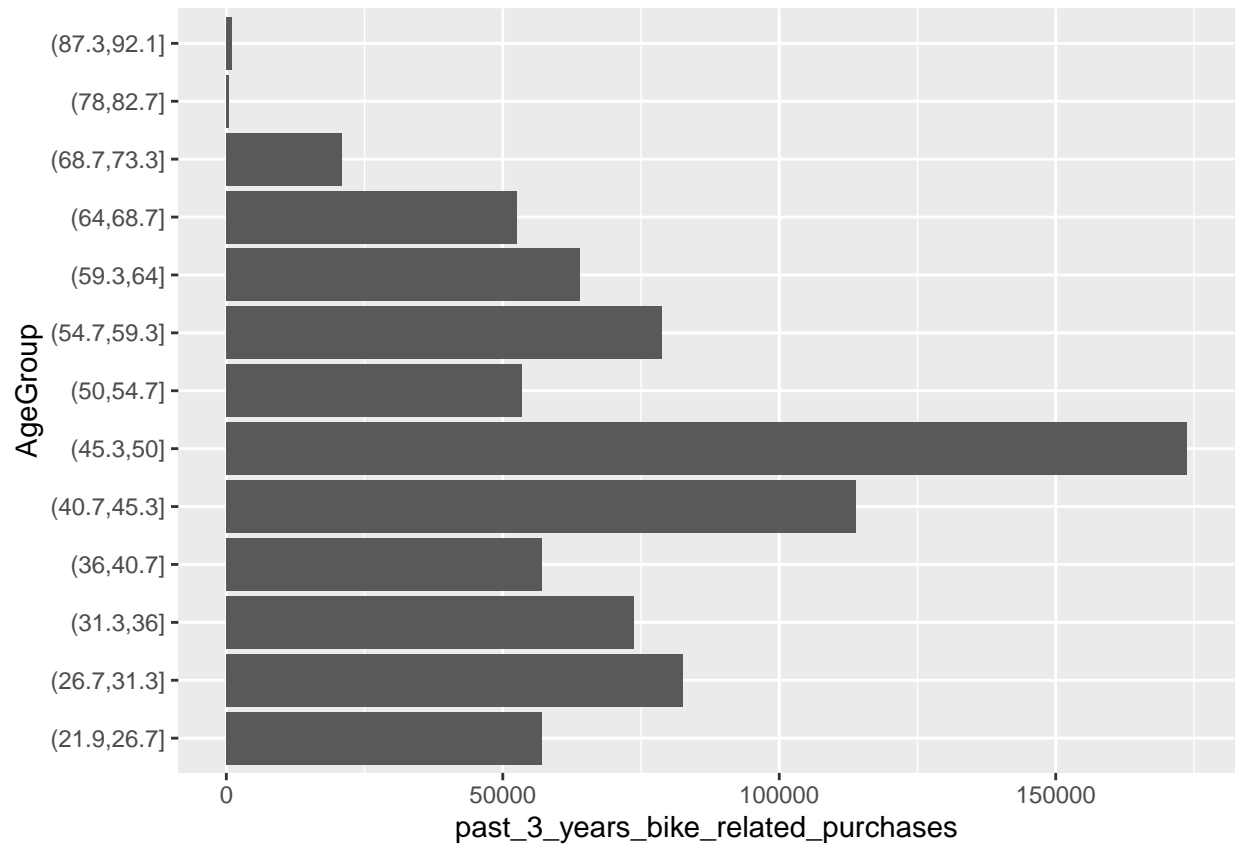




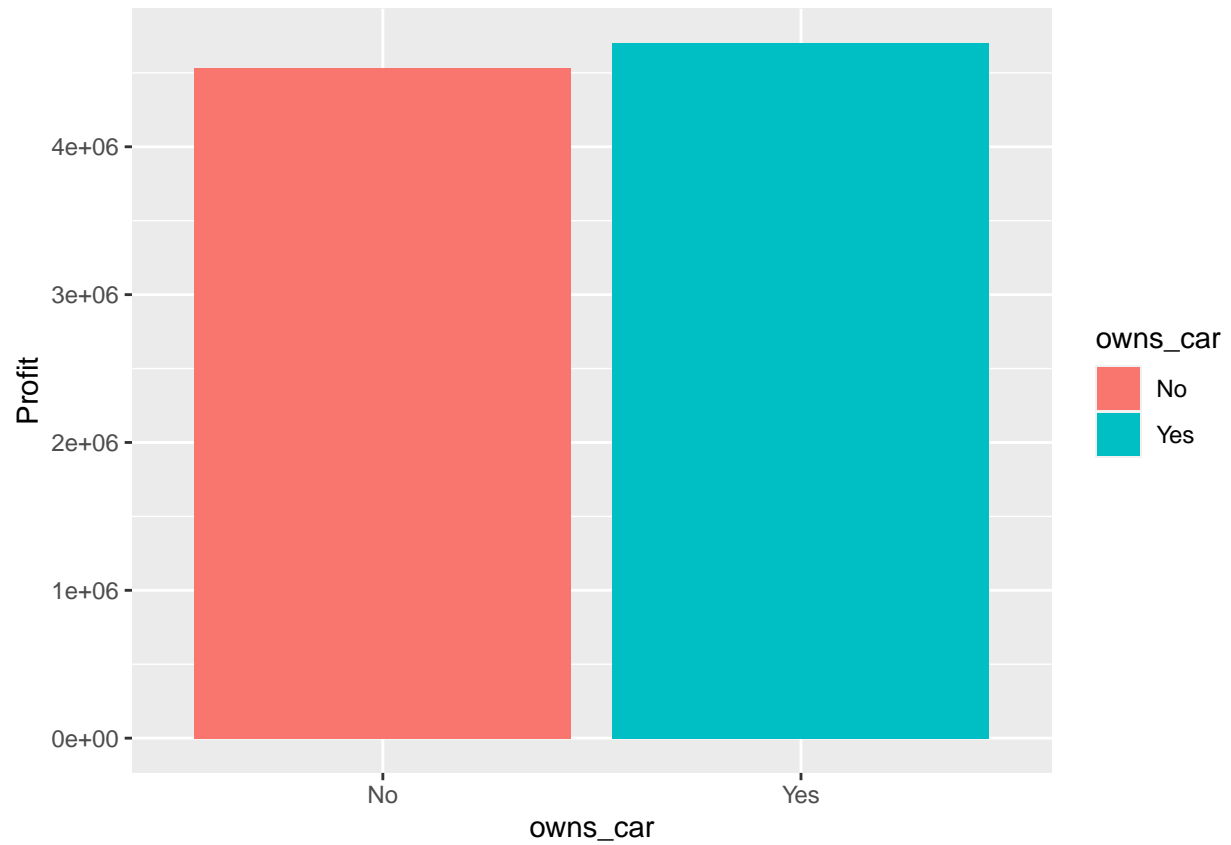
```
ggplot(kpmg, aes(y=AgeGroup,x=sum(Profit), fill=gender)) + geom_col()
```



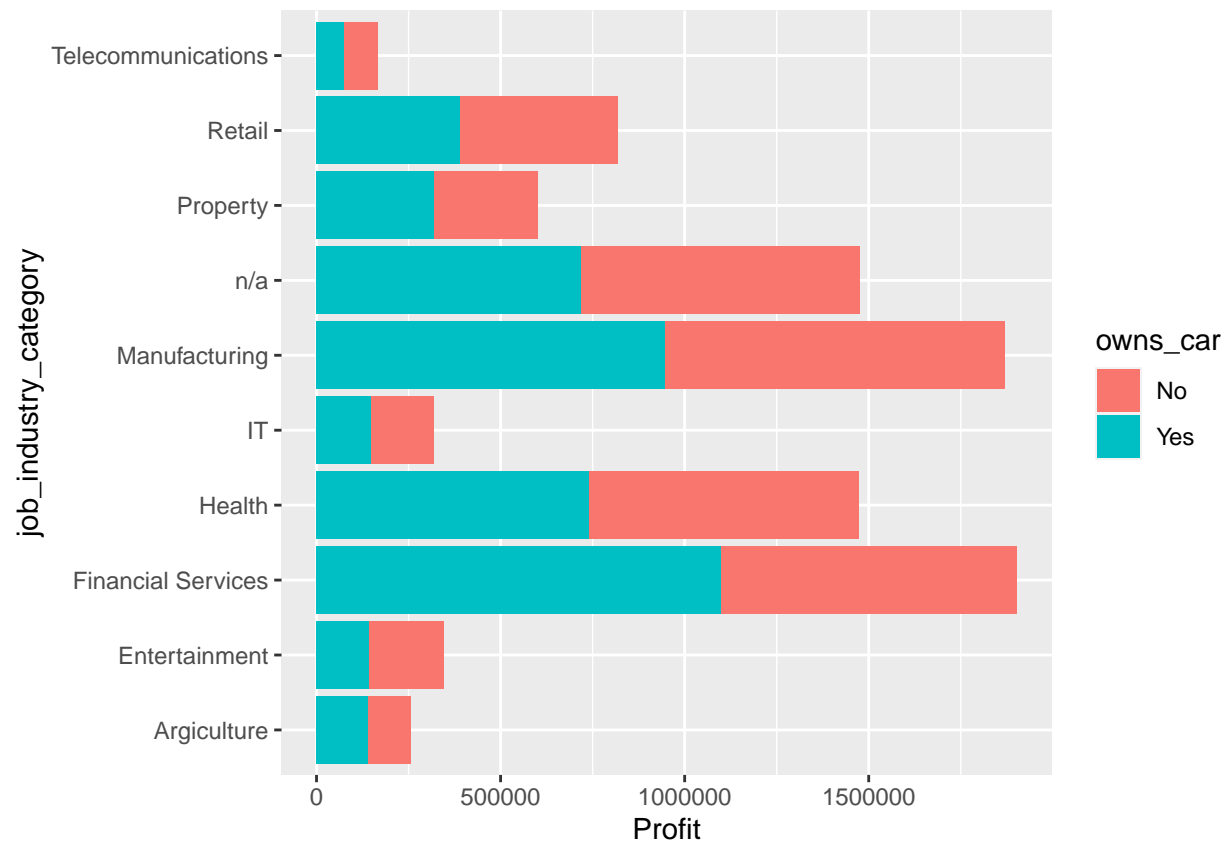
```
ggplot(kpmg, aes(y=AgeGroup,x=past_3_years_bike_related_purchases)) + geom_col()
```



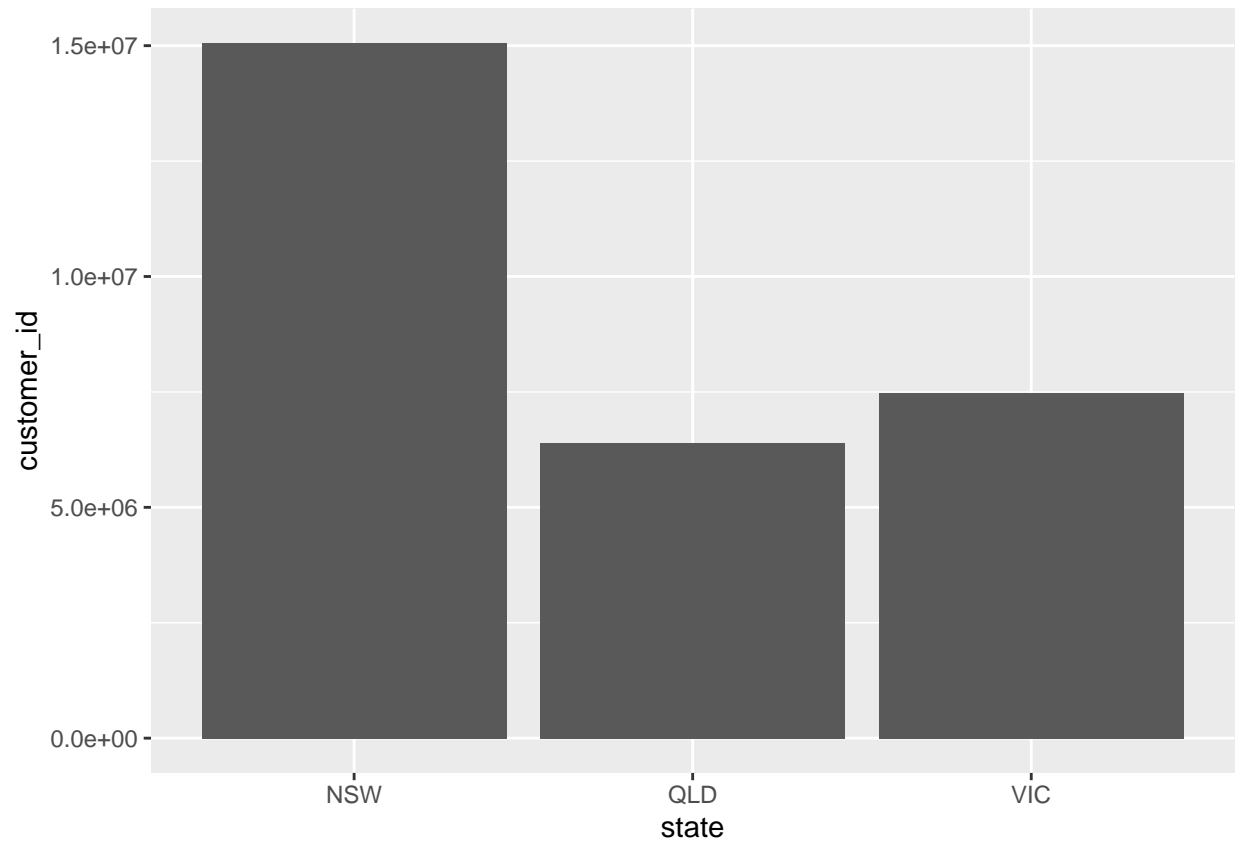
```
ggplot(kpmg, aes(x=owns_car,y=Profit, fill=owns_car)) + geom_col()
```



```
ggplot(kpmg, aes(y=job_industry_category,x=Profit, fill=owns_car)) + geom_col()
```



```
ggplot(kpmg, aes(x=state, y=customer_id)) + geom_col()
```



```
kpmg |> count(gender)
```

```
## # A tibble: 2 x 2
##   gender      n
##   <chr>  <int>
## 1 Female  8501
## 2 Male   8219
```

**DATA EXPLORATION** + Most of the Bike related Purchases were made by the age 40 and 49 + The Data shows that middle aged customers are the most potential Customers + Financial Services, Health and Manufacturing Sector are the top Three Profit Generating industries, followed by Retail, IT and Property