SH_analysis

Phuong Tran

28/04/2021

In this document, product information (Brand, Price, Category) scraped from SenHeng were processed and explored:

1. Examined SH_raw:

- Checked for missing values and re-format character data
- Added extra feature including Category and Original Price

2. Analyzed SH_cleaned:

- Explored which Brands and Categories have the highest number of products listed
- Explored which Laptop and PC Accessories Brands have the highest average Price

Insights:

1. Number of product listed:

- Brands with the highest number of products listed are TP-Link, Microsoft, MSI, Apple and HP
- The majority of products listed are PC Accessories, followed by Laptops
- Most of Laptops are Microsoft, followed by Acer, MSI, HP, ASUS, Apple and lastly Huawei
- Most of PC Accessories are TP-Link, followed by Apple, MSI, Microsoft, Rapoo, Logitech and Samsung.

2. Average Laptop and PC Accessories Price:

- Within Laptops: higher end - Apple (~6500) medium range - MSI (~4800),HP (~4500), Huawei (~4500), Acer (~4300) lower end - ASUS (~3500) Microsoft and Promate Laptop price are too low for a laptop, probably a misclassification error (Ideally will need to run the Scraping again to extract this information, will omit from analysis now on)
- Within PC Accessories: higher end - Apple, TP-Link medium range - MSI. lower end - Samsung, Rapoo, Logitech
- PC Accessories include many different products, hence further classification is required with more specific scraping.

Recommendations:

- Laptop listing in terms of Brands is well-distributed with 1 high-end Brand, 1 low-end Brand and mostly 4 medium-range Brands. However, in terms of the number of products for each brands, maybe consider listing more Huawei Laptops. The reason is that this brand is a medium-range Brand but does not have similar number of products listed compared to other medium-range Brands.
- Re-do scraping for more insights on PC Accessories.

```
# Load packages
library(data.table)
library(ggplot2)
```

```
library(readr)
library(plyr)
library(lattice)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(Rmisc)
```

```
# Import data
SH_raw <- read_csv("Senheng_compt.csv",
    col_types = cols(X1 = col_skip()))
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
# Examine data
head(SH_raw)
```

```
## # A tibble: 6 x 6
##   Brand   Distributor LinkofProduct            Name              Price SKU
##   <chr>   <chr>       <chr>                    <chr>             <chr> <chr>
## 1 "\"HP\~ "\"7158\";" https://www.senheng.com~ HP Notebook 14s-~ RM 2~ "\"HP-14~
## 2 "\"App~ "\"7158\";" https://www.senheng.com~ Apple 16-inch Ma~ RM10~ "\"APP-M~
## 3 "\"App~ "\"7158\";" https://www.senheng.com~ Apple 13-inch Ma~ RM48~ "\"APP-M~
## 4 "\"MSI~ "\"7158\";" https://www.senheng.com~ MSI 23.6 Inch Op~ RM64~ "\"MSI-G~
## 5 "\"MSI~ "\"7158\";" https://www.senheng.com~ MSI 14 Prestige ~ RM4,~ "\"MSI-P~
## 6 "\"App~ "\"7158\";" https://www.senheng.com~ Apple MacBook Ai~ RM4,~ "\"APP-M~
```

```
# Remove " and ; character from Brand
SH_raw$Brand<-gsub('\"','',SH_raw$Brand)
SH_raw$Brand<-gsub(';','',SH_raw$Brand)

# Remove " and ; character from Brand
SH_raw$SKU<-gsub('\"','',SH_raw$SKU)
SH_raw$SKU<-gsub(';','',SH_raw$SKU)
```

```
# Summary of Brand column
table(SH_raw$Brand,useNA = "ifany")
```

```
##
```

```
##                Acer     Apple      ASUS     Canon        HP    Huawei  Logitech
##         1        16        52        14         5        21        11        10
## Microsoft       MSI   Promate     Rapoo     Razer   Samsung   SanDisk   Seagate
##        43        71         2         9         1         2         8         3
##   TP-Link    xiaomi      <NA>
##        57         1         2
```

```
# Retrieve NA rows
SH_raw[is.na(SH_raw$Brand),]
```

```
## # A tibble: 2 x 6
##   Brand Distributor LinkofProduct               Name                Price  SKU
##   <chr> <chr>       <chr>                       <chr>               <chr>  <chr>
## 1 <NA>  <NA>        https://www.senheng.com.my~ Microsoft Bluetoot~ RM275~ <NA>
## 2 <NA>  <NA>        https://www.senheng.com.my~ ASUS ROG Zephyrus ~ RM6,9~ <NA>
```

```
SH_raw[SH_raw$Brand=="",]
```

```
## # A tibble: 3 x 6
##   Brand Distributor LinkofProduct                Name              Price SKU
##   <chr> <chr>       <chr>                        <chr>             <chr> <chr>
## 1 ""    "\"7158\";" https://www.senheng.com.my/~ Microsoft Sculpt~ RM52~ MS-MIC~
## 2  <NA>  <NA>       <NA>                         <NA>              <NA>  <NA>
## 3  <NA>  <NA>       <NA>                         <NA>              <NA>  <NA>
```

```
# There are 2 rows with NA values on Brands with 1 empty value.
# ASUS row and Microsoft Desktop also missing SKU due to no "Specification section on webpages. Manuall
SH_raw[SH_raw$Name=='Microsoft Bluetooth Desktop-Black',]$Brand<-'Microsoft'
SH_raw[SH_raw$Name=='ASUS ROG Zephyrus M15 LAPTOP 15.6 Inch (i7-10750H/16GB/1 TB SSD/GTX1660TI 6G/WIN10
SH_raw[SH_raw$Brand=="",]$Brand<-'Microsoft'
# Double check Brand column
table(SH_raw$Brand,useNA = "ifany")
```

```
##
##      Acer     Apple      ASUS     Canon        HP    Huawei  Logitech Microsoft
##        16        52        15         5        21        11        10        45
##       MSI   Promate     Rapoo     Razer   Samsung   SanDisk   Seagate   TP-Link
##        71         2         9         1         2         8         3        57
##    xiaomi
##         1
```

```
# Set Brand as factor
SH_raw$Brand<-parse_factor(SH_raw$Brand)
# There are 17 different brands
levels(SH_raw$Brand)
```

```
##  [1] "HP"        "Apple"     "MSI"       "Microsoft" "TP-Link"   "Acer"
##  [7] "Huawei"    "xiaomi"    "Promate"   "ASUS"      "Logitech"  "Rapoo"
## [13] "Razer"     "Seagate"   "SanDisk"   "Canon"     "Samsung"
```

```
# There are 198 unique SKUs (196 non NAs and 2 NAs) but 329 rows so there must be many duplicated rows
n_distinct(SH_raw$SKU,na.rm=TRUE)
```

```
## [1] 196
```

```
# Assign unique SKU for NAs entry for the purpose of subsequent analysis
SH_raw[is.na(SH_raw$SKU),]
```

```
## # A tibble: 2 x 6
```

```
##   Brand  Distributor LinkofProduct              Name                  Price SKU
##   <fct>  <chr>       <chr>                      <chr>                 <chr> <chr>
## 1 Micro~ <NA>        https://www.senheng.com.m~ Microsoft Bluetooth~ RM27~ <NA>
## 2 ASUS   <NA>        https://www.senheng.com.m~ ASUS ROG Zephyrus M~ RM6,~ <NA>
```

```
SH_raw[SH_raw$Name=='Microsoft Bluetooth Desktop-Black',]$SKU<-'sku_1'
SH_raw[SH_raw$Name=='ASUS ROG Zephyrus M15 LAPTOP 15.6 Inch (i7-10750H/16GB/1 TB SSD/GTX1660TI 6G/WIN10)
# Check again unique SKUs, correct 198 unique SKUs
n_distinct(SH_raw$SKU,na.rm=TRUE)
```

```
## [1] 198
```

```
# Remove duplicated entries. There are 200 rows in SH_cleaned but only 198 unique SKUs.
SH_cleaned<-distinct(SH_raw)
# Check for discrepancy. Entry with SKU 'APP-MBP13-2020-2USB-CFG' and 'SD-IX40N-CFG' were duplicated
table(SH_cleaned$SKU)
```

```
##
##           ACE-A314-22-R2DQ           ACE-A314-22-R4JW
##                          1                          1
##           ACE-A315-56-36JM          ACE-A315-57G-57L2
##                          1                          1
##           ACE-A515-56-51GF          ACE-AN515-55-52Z1
##                          1                          1
##              ACE-C733-C8F7          ACE-CP311-2H-C27N
##                          1                          1
##          ACE-PH315-53-791W        ACE-SF314-510G-761J
##                          1                          1
##         ACE-SF514-54T-70AA        ACE-SF514-55TA-55MW
##                          1                          1
## APP-MACBOOKAIR-M1-2020-CFG APP-MACBOOKPRO-M1-2020-CFG
##                          1                          1
##         APP-MBA13-2020-CFG    APP-MBP13-2020-2USB-CFG
##                          1                          2
##    APP-MBP13-2020-4USB-CFG         APP-MBP16-2020-CFG
##                          1                          1
##              APP-MJ1L2ZA/A              APP-MJ1M2ZA/A
##                          1                          1
##              APP-MJ2R2ZA/A              APP-MLA02ZA/A
##                          1                          1
##              APP-MLA22ZA/A              APP-MQ052ZA/A
##                          1                          1
##              APP-MRME2ZA/A              APP-MRMF2ZA/A
##                          1                          1
##              APP-MUF82ZA/A              APP-MX3L2ZA/A
##                          1                          1
##              APP-MXQT2ZA/A              APP-MXQU2ZA/A
##                          1                          1
##          ASU-A416M-AEB093T         ASU-A516J-ABR374TS
##                          1                          1
##          ASU-A516M-ABQ155T         ASU-FA506I-HHN137T
##                          1                          1
##         ASU-FA706I-IH7079T          ASU-G512L-IAL008T
##                          1                          1
##        ASU-UX325E-AEG060TS        ASU-UX425E-ABM067TS
##                          1                          1
```

```
##                CNN-E410                CNN-E470
##                       1                       1
##                CNN-E510              CNN-LBP6030
##                       1                       1
##           HP-13AW2099TU           HP-13AW2100TU
##                       1                       1
##           HP-13AY0043AU           HP-13BA1011TX
##                       1                       1
##             HP-14CF-CFG           HP-14SFQ0075AU
##                       1                       1
##           HP-15EG0109TX           HP-15EP0010TX
##                       1                       1
##         HP-680-TRI-COLOR           HP-S01-PF1166D
##                       1                       1
##         HUA-AD80-MONITOR       HUA-MATEBOOKD14-I5
##                       1                       1
##           HUA-MPEN.LITE             HUA-WIFIAX3
##                       1                       1
##            LGT-M171-CFG            LGT-M185-CFG
##                       1                       1
##            LGT-M337-CFG            LGT-M590-CFG
##                       1                       1
##           LGT-MK240-CFG          LGT-MXMASTER-3
##                       1                       1
##                 LS108G            MS-6GQ-01144
##                       1                       1
##            MS-79G-05143           MS-AMOUSE-CFG
##                       1                       1
##              MS-GMF-CFG         MS-LAPTOPGO-CFG
##                       1                       1
##         MS-MIC-4FD-00027         MS-MIC-7N9-00028
##                       1                       1
##         MS-MIC-APB-00018         MS-MIC-BOOK3-CFG
##                       1                       1
##           MS-MIC-CFG-PEN           MS-MIC-CZV-CFG
##                       1                       1
##         MS-MIC-EJT-00002           MS-MIC-FFP-CFG
##                       1                       1
##         MS-MIC-FMM-00015         MS-MIC-FTW-00005
##                       1                       1
##         MS-MIC-GMF-00006           MS-MIC-GO2-CFG
##                       1                       1
##         MS-MIC-H3S-00005         MS-MIC-JTY-00007
##                       1                       1
##         MS-MIC-JVZ-00007         MS-MIC-JWL-00007
##                       1                       1
##         MS-MIC-KCM-00015           MS-MIC-KGY-CFG
##                       1                       1
##         MS-MIC-KTF-00005         MS-MIC-L3V-00027
##                       1                       1
##         MS-MIC-L5V-00027         MS-MIC-LLK-00005
##                       1                       1
##         MS-MIC-PP3-00024         MS-MIC-QSW-00015
##                       1                       1
```

```
##           MS-MIC-SWV-00005           MS-MIC-U7Z-00015
##                          1                          1
##           MS-MIC-WS2-00014           MS-MIC-WS3-00005
##                          1                          1
##             MS-MMMOUSE-CFG            MS-SUR-LPT3-CFG
##                          1                          1
##            MS-SUR-PRO7-CFG            MS-SUR-PROX-CFG
##                          1                          1
##               MS-T5D-03302                MSI-DS4100
##                          1                          1
##                 MSI-G241VC                 MSI-G24C4
##                          1                          1
##                 MSI-G27C4                 MSI-G27CQ4
##                          1                          1
##              MSI-GAMING.XL       MSI-GF63-9SCXR-649MY
##                          1                          1
##           MSI-GH10-HEADSET                  MSI-GK40
##                          1                          1
##                  MSI-GK70                 MSI-GK701
##                          1                          1
##                  MSI-GM10                  MSI-GM60
##                          1                          1
##                  MSI-GM70       MSI-M14-B10RBS-408MY
##                          1                          1
##       MSI-M15-A10RBS-469MY              MSI-MAG272CQR
##                          1                          1
##              MSI-MAG322CQR              MSI-MPG343CQR
##                          1                          1
##           MSI-MPGTRIDENT-A         MSI-MPGTRIDENT-AS60
##                          1                          1
##         MSI-MPGTRIDENT-AS70       MSI-P14-A10RAS-087MY
##                          1                          1
##       MSI-P14-A10RAS-223MY          MSI-SHIELD.MOSPAD
##                          1                          1
##            MSI-THUNDER.ALUM          PAC-MBOOK14-I5-GF
##                          1                          1
##          PAC-MBOOKD15-I5-GF       PAC-MBOOKD1511-I5-GF
##                          1                          1
##         PAC-MBOOKXPRO-I5-GF          PRO-6959144031989
##                          1                          1
##           PRO-6959144032009                  RAP-1620
##                          1                          1
##                 RAP-3920P              RAP-7100P/BLK
##                          1                          1
##              RAP-7100P/RED              RAP-M10+/BLUE
##                          1                          1
##              RAP-M10+/RED             RAP-M10+/WHITE
##                          1                          1
##            RAP-N1020-BLACK          RAZ-RZ01-01370100
##                          1                          1
##          SAM-MB-MC32GA/APC          SAM-MB-MC64GA/APC
##                          1                          1
##          SD-CZ430-032G-G46           SD-CZ73-032G-G46
##                          1                          1
```

```
##               SD-DD3-CFG              SD-DDC2-032G
##                        1                         1
##             SD-DDC2-064G              SD-IX40N-CFG
##                        1                         2
##       SEA-BACKUP.1TB/BLK       SEA-BACKUP.1TB/BLUE
##                        1                         1
##       SEA-BACKUP.2TB/BLUE                     sku_1
##                        1                         1
##                    sku_2             TP-ARCHERAX10
##                        1                         1
##            TP-ARCHERAX50            TP-ARCHERAX6000
##                        1                         1
##            TP-ARCHERAX73               TP-ARCHERC6
##                        1                         1
##           TP-ARCHERMR200             TP-ARCHERMR400
##                        1                         1
##             TP-ARCHERT2U             TP-ARCHERT2UHP
##                        1                         1
##          TP-ARCHERT2UNANO           TP-ARCHERT2UPLUS
##                        1                         1
##             TP-ARCHERT4E              TP-ARCHERT4U
##                        1                         1
##             TP-ARCHERT6E              TP-ARCHERT9UH
##                        1                         1
##           TP-ARCHERVR400           TP-ARCHERVR600V
##                        1                         1
##           TP-DECO-M4.2PK            TP-DECO-M5.2PK
##                        1                         1
##           TP-DECO-M5.3PK         TP-DECO-M9PLUS.2PK
##                        1                         1
##         TP-DECO-M9PLUS.3PK              TP-DECOE4.2PK
##                        1                         1
##              TP-DECOE4.3PK            TP-DECOX20.2PK
##                        1                         1
##            TP-DECOX20.3PK            TP-DECOX60.2PK
##                        1                         1
##            TP-DECOX60.3PK                 TP-M7000
##                        1                         1
##                 TP-M7200                 TP-M7350
##                        1                         1
##                 TP-M7450                 TP-RE305
##                        1                         1
##             TP-TD-W8961N               TP-TL-MR100
##                        1                         1
##           TP-TL-MR640APAC             TP-TL-WA855RE
##                        1                         1
##             TP-TL-WA860RE              TP-TL-WN725N
##                        1                         1
##             TP-TL-WN822N              TP-TL-WN823N
##                        1                         1
##                 TP-UB400               XMI-MOS-CFG
##                        1                         1
```

```
# Closer look shows price discrepancy and a manual check reveals that the product 'APP-MBP13-2020-2USB-
#and 'SD-IX40N-CFG' are out of stock. Hence, we can remove 1 entry, keep the other entry with price set
SH_cleaned[SH_cleaned$SKU=='APP-MBP13-2020-2USB-CFG',]
```

```
## # A tibble: 2 x 6
##   Brand Distributor LinkofProduct              Name               Price SKU
##   <fct> <chr>       <chr>                      <chr>              <chr> <chr>
## 1 Apple "\"7158\";" https://www.senheng.com.m~ Apple 13-inch Mac~ RM48~ APP-MBP~
## 2 Apple "\"7158\";" https://www.senheng.com.m~ Apple 13-inch Mac~ RM2,~ APP-MBP~
```

```
SH_cleaned<-SH_cleaned[-which(SH_cleaned$SKU=='APP-MBP13-2020-2USB-CFG'&SH_cleaned$Price=='RM489.00'),]
SH_cleaned[SH_cleaned$SKU=='APP-MBP13-2020-2USB-CFG',]$Price<-""

# Do the same for ''SD-IX40N-CFG' product
SH_cleaned[SH_cleaned$SKU=='SD-IX40N-CFG',]
```

```
## # A tibble: 2 x 6
##   Brand  Distributor LinkofProduct                 Name             Price SKU
##   <fct>  <chr>       <chr>                         <chr>            <chr> <chr>
## 1 SanDi~ "\"7158\";" https://www.senheng.com.my/s~ SanDisk iXpand ~ RM 7~ SD-IX~
## 2 SanDi~ "\"7158\";" https://www.senheng.com.my/s~ SanDisk iXpand ~ RM 3~ SD-IX~
```

```
SH_cleaned<-SH_cleaned[-which(SH_cleaned$SKU=='SD-IX40N-CFG'&SH_cleaned$Price=='RM 70.00'),]
SH_cleaned[SH_cleaned$SKU=='SD-IX40N-CFG',]$Price<-""
```

```
# Summary of Price Column reveals 2 "" values, some 2 prices values (original and discounted) and mostl
table(SH_cleaned$Price,useNA='ifany')
```

```
##
##                            RM 146.00         RM 2,169.00            RM 35.00
##                   2                1                   1                   1
##            RM 69.00            RM 82.00          RM1,029.00          RM1,159.00
##                   1                1                   1                   1
## RM1,174.00 1,199.00          RM1,199.00          RM1,219.00 RM1,289.00 1,421.00
##                   1                1                   1                   1
##          RM1,349.00          RM1,359.00          RM1,399.00          RM1,499.00
##                   1                1                   2                   1
##          RM1,549.00          RM1,638.00          RM1,739.00          RM1,799.00
##                   1                1                   1                   1
##          RM1,955.00         RM10,499.00            RM102.00            RM113.00
##                   1                1                   1                   1
##            RM115.00            RM120.00            RM129.00            RM131.00
##                   1                1                   1                   1
##            RM136.00            RM139.00            RM150.00            RM155.00
##                   1                1                   1                   3
##            RM158.00            RM159.00            RM164.00            RM169.00
##                   1                1                   1                   2
##            RM175.00            RM179.00            RM187.00             RM19.00
##                   1                1                   1                   1
##            RM195.00          RM2,049.00 RM2,149.00 2,199.00          RM2,169.00
##                   1                2                   1                   1
##          RM2,249.00          RM2,299.00          RM2,599.00          RM2,699.00
##                   1                1                   1                   1
##          RM2,758.00            RM200.00            RM209.00            RM219.00
##                   1                1                   1                   1
##            RM229.00            RM238.00             RM24.00            RM241.00
```

```
##                   2                   1                   2                   1
##            RM249.00            RM253.00            RM259.00            RM269.00
##                   1                   2                   2                   1
##            RM275.00            RM285.00            RM294.00            RM299.00
##                   1                   1                   1                   3
## RM3,149.00 3,199.00          RM3,199.00          RM3,299.00          RM3,399.00
##                   1                   1                   1                   2
##          RM3,499.00          RM3,699.00          RM3,799.00          RM3,899.00
##                   1                   1                   1                   3
## RM3,974.00 3,999.00          RM3,999.00     RM304.00 368.00            RM306.00
##                   1                   4                   1                   1
##             RM31.00            RM317.00            RM319.00            RM320.00
##                   2                   1                   1                   1
##            RM328.00             RM33.00            RM339.00            RM349.00
##                   1                   1                   3                   3
##            RM357.00             RM36.00            RM362.00            RM379.00
##                   1                   1                   1                   2
##            RM389.00             RM39.00            RM399.00          RM4,039.00
##                   1                   4                   3                   1
##          RM4,099.00          RM4,299.00          RM4,399.00          RM4,599.00
##                   1                   2                   1                   2
##          RM4,699.00          RM4,899.00          RM4,999.00             RM40.00
##                   1                   1                   1                   1
##            RM409.00            RM420.00            RM424.00            RM429.00
##                   1                   1                   1                   1
##            RM434.00            RM439.00            RM445.00            RM449.00
##                   1                   2                   1                   1
##             RM46.00            RM475.00     RM476.00 552.00            RM489.00
##                   1                   1                   1                   2
##            RM499.00          RM5,179.00          RM5,199.00          RM5,499.00
##                   1                   1                   1                   1
##          RM5,599.00          RM5,699.00 RM5,749.00 5,799.00            RM508.00
##                   1                   1                   1                   1
##            RM529.00             RM53.00            RM539.00             RM57.00
##                   1                   1                   1                   1
##            RM579.00            RM580.00            RM589.00             RM59.00
##                   1                   1                   1                   1
##          RM6,899.00          RM6,999.00            RM639.00             RM64.00
##                   2                   2                   2                   1
##            RM649.00            RM688.00             RM69.00          RM7,499.00
##                   2                   1                   1                   1
##          RM7,788.00          RM7,999.00             RM70.00             RM75.00
##                   1                   1                   2                   1
##             RM76.00             RM79.00            RM799.00          RM8,999.00
##                   1                   1                   2                   1
##             RM82.00             RM85.00            RM858.00             RM86.00
##                   3                   1                   1                   1
##             RM92.00             RM93.00             RM99.00            RM999.00
##                   2                   1                   4                   2
```

```
# Create column Price_1 of discounted price (if any)
SH_cleaned$Price_1<-parse_number(SH_cleaned$Price,na=c("","NA"))
# Create column Price_2 for original price
SH_cleaned$Price_2<-SH_cleaned$Price
```

```
# Find all rows with double price entry, their pattern begins with any character followed by .00 and th
SH_cleaned[grep(".\\.00 " ,SH_cleaned$Price_2),]
```

```
## # A tibble: 8 x 8
##   Brand  Distributor LinkofProduct      Name       Price  SKU    Price_1 Price_2
##   <fct>  <chr>       <chr>              <chr>      <chr>  <chr>    <dbl> <chr>
## 1 Acer   "\"7158\";" https://www.senhe~ Acer Aspi~ RM3,1~ ACE-A~    3149 RM3,14~
## 2 Acer   "\"7158\";" https://www.senhe~ Acer 15.6~ RM5,7~ ACE-P~    5749 RM5,74~
## 3 Acer   "\"7158\";" https://www.senhe~ Acer Swif~ RM3,9~ ACE-S~    3974 RM3,97~
## 4 Acer   "\"7158\";" https://www.senhe~ Acer Aspi~ RM2,1~ ACE-A~    2149 RM2,14~
## 5 Acer   "\"7158\";" https://www.senhe~ Acer Chro~ RM1,1~ ACE-C~    1174 RM1,17~
## 6 Micro~ "\"7158\";" https://www.senhe~ Microsoft~ RM304~ MS-6G~     304 RM304.~
## 7 Micro~ "\"7158\";" https://www.senhe~ Microsoft~ RM1,2~ MS-T5~    1289 RM1,28~
## 8 Micro~ "\"7158\";" https://www.senhe~ Microsoft~ RM476~ MS-79~     476 RM476.~
```

```
# Use sub() to remove first price, .* represents any character for any number of times, \\. represents
SH_cleaned[grep(".\\.00 " ,SH_cleaned$Price_2),]$Price_2<-sub("RM.*\\.00 ","",SH_cleaned[grep(".\\.00 "
# Parse_number for Price_2 column
SH_cleaned$Price_2<-parse_number(SH_cleaned$Price_2)
# check Price_1 and Price_2 Column one more time
table(SH_cleaned$Price_1,useNA='ifany')
```

```
##
##    19    24    31    33    35    36    39    40    46    53    57    59    64
##     1     2     2     1     1     1     4     1     1     1     1     1     1
##    69    70    75    76    79    82    85    86    92    93    99   102   113
##     2     2     1     1     1     4     1     1     2     1     4     1     1
##   115   120   129   131   136   139   146   150   155   158   159   164   169
##     1     1     1     1     1     1     1     1     3     1     1     1     2
##   175   179   187   195   200   209   219   229   238   241   249   253   259
##     1     1     1     1     1     1     1     2     1     1     1     2     2
##   269   275   285   294   299   304   306   317   319   320   328   339   349
##     1     1     1     1     3     1     1     1     1     1     1     3     3
##   357   362   379   389   399   409   420   424   429   434   439   445   449
##     1     1     2     1     3     1     1     1     1     1     2     1     1
##   475   476   489   499   508   529   539   579   580   589   639   649   688
##     1     1     2     1     1     1     1     1     1     1     2     2     1
##   799   858   999  1029  1159  1174  1199  1219  1289  1349  1359  1399  1499
##     2     1     2     1     1     1     1     1     1     1     1     2     1
##  1549  1638  1739  1799  1955  2049  2149  2169  2249  2299  2599  2699  2758
##     1     1     1     1     1     2     1     2     1     1     1     1     1
##  3149  3199  3299  3399  3499  3699  3799  3899  3974  3999  4039  4099  4299
##     1     1     1     2     1     1     1     3     1     4     1     1     2
##  4399  4599  4699  4899  4999  5179  5199  5499  5599  5699  5749  6899  6999
##     1     2     1     1     1     1     1     1     1     1     1     2     2
##  7499  7788  7999  8999 10499  <NA>
##     1     1     1     1     1     2
```

```
table(SH_cleaned$Price_2,useNA='ifany')
```

```
##
##    19    24    31    33    35    36    39    40    46    53    57    59    64
##     1     2     2     1     1     1     4     1     1     1     1     1     1
##    69    70    75    76    79    82    85    86    92    93    99   102   113
##     2     2     1     1     1     4     1     1     2     1     4     1     1
```

```
##   115   120   129   131   136   139   146   150   155   158   159   164   169
##     1     1     1     1     1     1     1     1     3     1     1     1     2
##   175   179   187   195   200   209   219   229   238   241   249   253   259
##     1     1     1     1     1     1     1     2     1     1     1     2     2
##   269   275   285   294   299   306   317   319   320   328   339   349   357
##     1     1     1     1     3     1     1     1     1     1     3     3     1
##   362   368   379   389   399   409   420   424   429   434   439   445   449
##     1     1     2     1     3     1     1     1     1     1     2     1     1
##   475   489   499   508   529   539   552   579   580   589   639   649   688
##     1     2     1     1     1     1     1     1     1     1     2     2     1
##   799   858   999  1029  1159  1199  1219  1349  1359  1399  1421  1499  1549
##     2     1     2     1     1     2     1     1     1     2     1     1     1
##  1638  1739  1799  1955  2049  2169  2199  2249  2299  2599  2699  2758  3199
##     1     1     1     1     2     2     1     1     1     1     1     1     2
##  3299  3399  3499  3699  3799  3899  3999  4039  4099  4299  4399  4599  4699
##     1     2     1     1     1     3     5     1     1     2     1     2     1
##  4899  4999  5179  5199  5499  5599  5699  5799  6899  6999  7499  7788  7999
##     1     1     1     1     1     1     1     1     2     2     1     1     1
##  8999 10499  <NA>
##     1     1     2
```

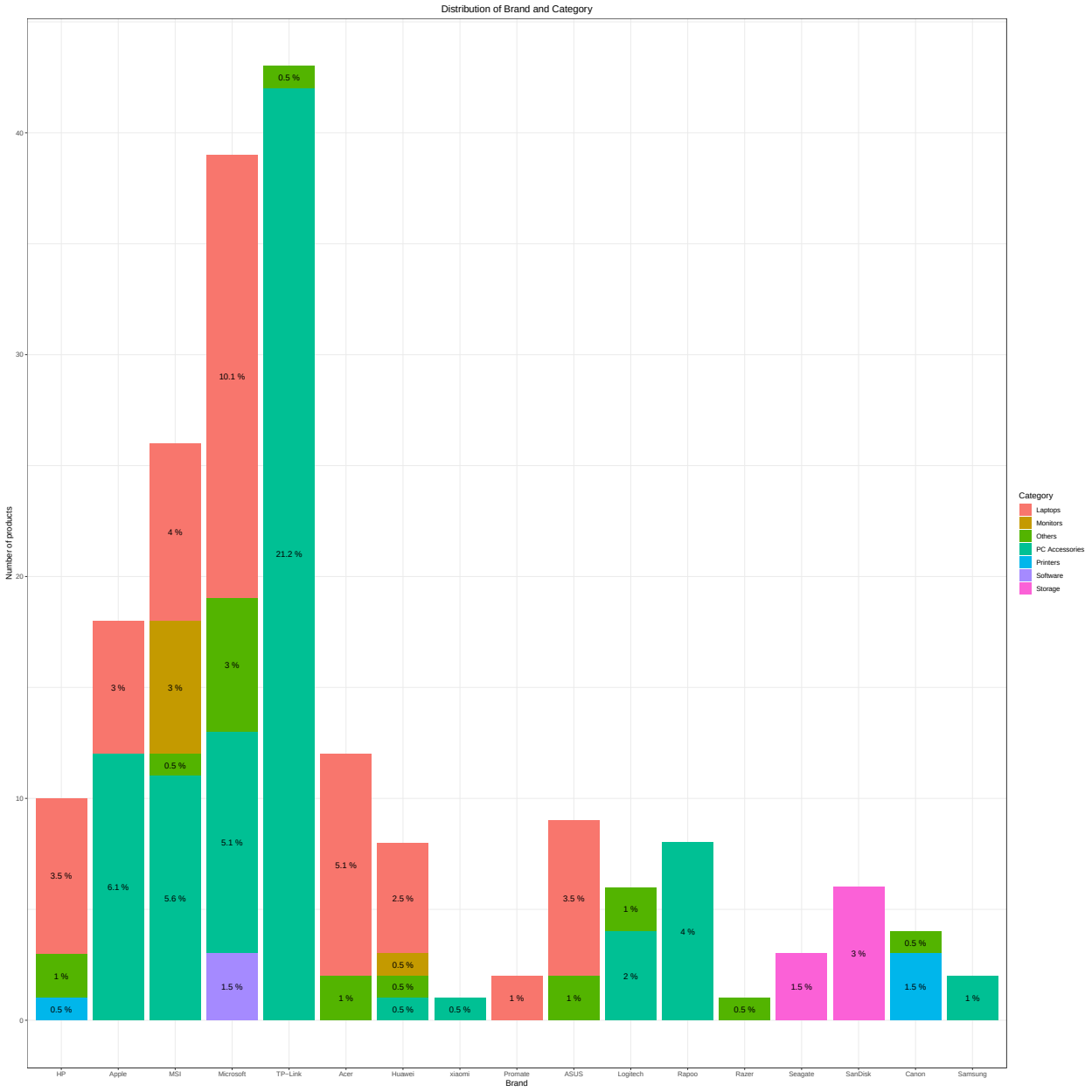
```
SH_cleaned<-SH_cleaned %>%
    mutate(Category=case_when(
        grepl("Monitor",SH_cleaned$Name,ignore.case = TRUE)==TRUE ~ "Monitors",
        grepl("Chair",SH_cleaned$Name,ignore.case = TRUE)==TRUE ~ "Gaming Chairs",
        grepl("Printer|Ink",SH_cleaned$Name,ignore.case = FALSE)==TRUE~"Printers", #Avoid TP-Link
        grepl("Microsoft 365|Microsoft Office",SH_cleaned$Name,ignore.case = TRUE)==TRUE ~"Software",
        grepl("Drive|Drives",SH_cleaned$Name,ignore.case = TRUE)==TRUE ~"Storage",
        grepl("MateBook|Laptop|Surface|Macbook|Aspire|Chromebook|Convertible|Notebook|Helios|Zenbook|Tr
              SH_cleaned$Name,ignore.case = TRUE)==TRUE ~ "Laptops",
        grepl("Pen|Keyboard|Mouse|Keypad|Adapter|Router|Audio|Pad|Port|Wi-fi|Dual Band|Desktop}",SH_clea
        TRUE ~ "Others"
                            )
        )
# Check Category Column, 89% categorized --- not too bad but ideally I would go back to web scraping an
table(SH_cleaned$Category)
```

```
##
##        Laptops       Monitors         Others PC Accessories       Printers
##             65              7             19             91              4
##       Software        Storage
##              3              9
```

```
write.csv(SH_cleaned,"/Users/ccmb_hd/OneDrive - Deakin University/ds_senheng/SH_cleaned.csv")
```

```
# Set theme for plots
theme_set(theme_bw())
theme_update(plot.title=element_text(hjust=0.5),plot.subtitle=element_text(hjust=0.5))
```

```
ggplot(SH_cleaned,aes(x=Brand,fill=Category))+
  geom_bar(stat="count")+
  labs(title="Distribution of Brand and Category",
       x="Brand",y="Number of products")+
  stat_count(geom = "text",
             aes(label=paste(round((..count..)/sum(..count..)*100,1),"%")),
             position=position_stack(vjust=0.5))
```



Brands with the highest number of products are TP-Link, Microsoft, MSI , Apple and HP

The majority of products listed are PC Accessories, followed by Laptops

Most of TP-Link products are PC Accessories

Microsoft products are mostly Laptops and PC Accessories while most of the number of Apple Accessories doubles that of Apple Laptop. This indicates that there are more Microsoft Laptop models compared to Apple laptop models.

MSI products are balanced between PC Accessories, Laptops, and Monitors.

Only 2 Monitor brands including MSI and Huawei

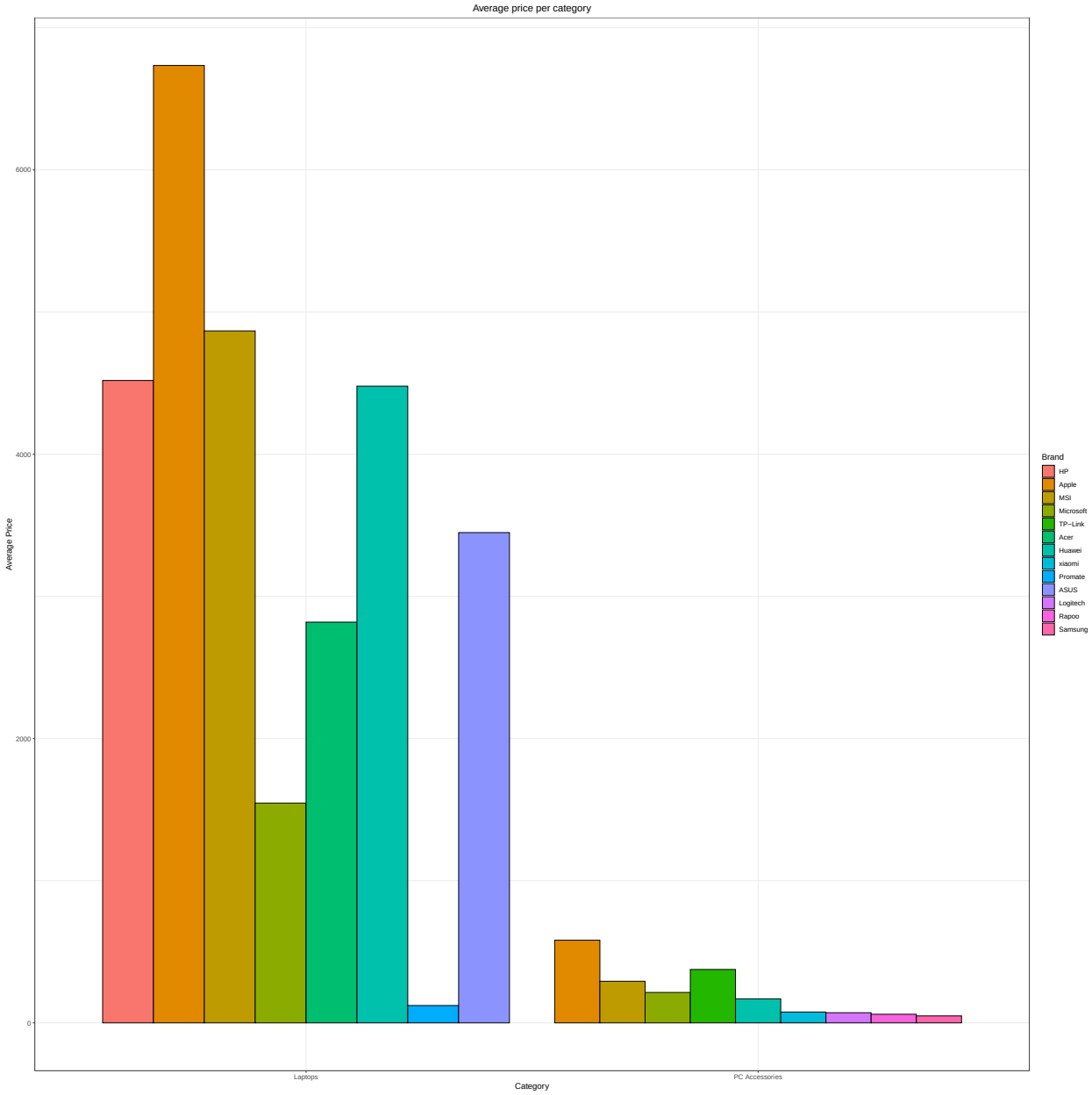
We will focus on Laptops and PC Accessories category

```
# Price summary
price_sum<-summarySE(data=SH_cleaned[!is.na(SH_cleaned$Price_1),],measurevar = 'Price_1',groupvars = c(
```

```
## Warning in qt(conf.interval/2 + 0.5, datac$N - 1): NaNs produced
```

```
# Some product category_brand has only 1 entry, thus NA values in the price_sum table. Replace these wi
price_sum[is.na(price_sum$sd),]$sd<-0

# Graph of average price per category and brand for Laptops and PC Accessories
ggplot(data=price_sum[price_sum$Category %in% c("Laptops",'PC Accessories'),],
      aes(x=Category,y=Price_1))+
      geom_bar(aes(fill=Brand),stat='summary',fun='mean',position="dodge",color="black")+
    labs(title="Average price per category",
         y="Average Price",x="Category")
```



Within laptops, Apple is the most expensive (~6500), followed by MSI (~4800),HP (~4500), Huawei (~4500) and ASUS (~3500). Microsoft is the cheapest (~1800) but the price is too low. Promate price is too low for a laptop, probably a misclassification error.

Within PC Accessories, Apple is the most expensive, followed by TP-Link, MSI and Microsoft.

‘‘‘