Assignment 1

Jacob Aylward

2025-06-02

Assignment 1 for Fundamentals of Machine Learning Resource: Soundankar, A. (2025). 🚬 global smoking trends & brand (2010-2024) 🌍 . Retrieved from https://www.kaggle.com/datasets/atharvasoundankar/global-smoking-trends-and-brand-popularity?resource=download

```
library(readxl)
data<-read_excel("C:/Users/jacob/Downloads/smoking_data.xlsx")
summary(data$`Total Smokers (Millions)`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.70   86.22  165.00  170.47  256.62  344.10
```

```
summary(data$`Male Smokers (%)`) #descriptive stats for quantitative variable
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.80   11.72   16.90   17.72   23.73   34.20
```

```
summary(data$`Female Smokers (%)`) #descriptive stats for quantitative
variable
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.900   3.400   5.000   5.668   7.500  15.200
```

```
table(data$Country) #descriptive stats for categorical variable
```

```
##
##    Australia       Brazil       Canada        China       France
Germany
##            8            9            6           11            4
5
##        India    Indonesia        Italy        Japan       Mexico
Russia
##           10            9            9            5            4
7
## South Africa  South Korea        Spain     Thailand       Turkey
UK
##           11            9            7            5            9
9
##          USA      Vietnam
##            6            7
```

```
table(data$`Top Cigarette Brand in Country`) #descriptive stats for
categorical variable
```

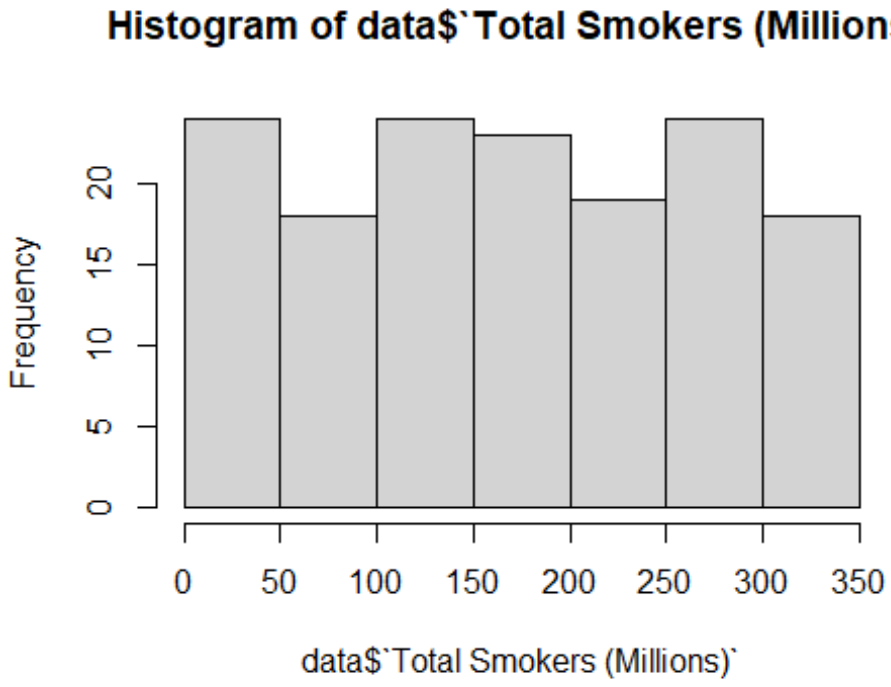
```
##
## China National Tobacco Corp.                        Derby
##                           11                            9
##                   Du Maurier                        Esse
##                            6                            9
##                      Fortuna                    Gauloises
##                            7                            4
##                   Gold Flake                Gudang Garam
##                           10                            9
##                     Marlboro                      Mevius
##                           24                            5
##                           MS             Peter Stuyvesant
##                            9                           11
##                     Sobranie                       Tekel
##                            7                            9
##                     Vinataba                     Winfield
##                            7                            8
##                       Wonder
##                            5
```

```
log(data$`Cigarette Consumption (Billion Units)`) #log transformation for
skewed variables related to cigarette consumption
```

```
##   [1] 7.789744 4.574711 5.084505 4.714025 7.673781 6.355587 6.638699
5.694742
##   [9] 7.214063 6.618605 4.198705 7.291247 7.580394 7.133535 7.667579
6.302619
##  [17] 6.101663 5.445443 5.587997 5.692047 6.797829 5.813832 6.248043
7.640988
##  [25] 7.472273 7.797168 7.454951 6.629627 6.871817 4.477337 6.293789
7.196537
##  [33] 4.757033 5.105945 6.781285 5.554509 5.774862 6.926577 5.918625
5.145166
##  [41] 4.223910 7.094401 7.221178 5.389985 3.811097 7.929702 7.012836
7.812985
##  [49] 7.386099 7.968146 6.080619 7.918556 4.929425 7.510649 5.955060
4.990433
##  [57] 4.166665 6.854249 7.134253 6.227327 5.916472 6.740874 6.735661
7.223004
##  [65] 3.198673 6.779695 7.717351 6.866933 3.811097 6.393256 7.119878
6.692828
##  [73] 7.379008 7.005789 7.113712 7.201469 6.219596 6.699377 3.437208
5.024538
##  [81] 1.589235 6.644571 6.788859 7.427501 5.391352 7.755424 6.416405
6.578418
##  [89] 6.302069 5.225209 4.237001 3.671225 2.116256 7.123270 6.314815
5.875212
```

```
##  [97] 5.797576 2.572612 7.561746 5.667810 5.994211 7.333546 6.613787
5.230574
## [105] 6.175243 8.082896 5.550631 1.458615 7.239861 7.116719 7.461870
6.788859
## [113] 6.523415 7.718285 6.863072 6.234018 6.037632 6.853932 5.248602
7.724402
## [121] 7.319136 6.408364 6.279084 4.418841 6.314996 4.630838 7.560861
7.016699
## [129] 4.554929 7.928478 7.524237 5.650030 6.048790 7.101099 7.801146
7.210154
## [137] 7.009138 7.646354 5.687653 5.917818 7.608523 6.752504 7.665472
7.117206
## [145] 5.186827 6.603537 6.234018 6.897604 7.214136 7.606636
```

```
hist(data$`Total Smokers (Millions)`) #plotting a histogram of total smokers
```



```r
plot(data$`Cigarette Consumption (Billion Units)`,data$`Total Smokers
(Millions)`) #Comparing total smokers to amount of cigarettes consumed
```