Analisa Bisnis Restoran 2020

Choiril Abdul

Persiapan

Mempersiapkan library dan dataset

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 -
-
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() -
-
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(ggmosaic)
library(readr)

### Menentukan dataset

filePath <- "dataset/"
ft50 <- fread(paste0(filePath,"Future50.csv"))
ip100 <- fread(paste0(filePath,"Independence100.csv"))
tp250 <- fread(paste0(filePath,"Top250.csv"))
```

Eksplorasi Analisis Data

Untuk memulai analisis data adalah mengerti tentang datanya, pertama cari tipe data yang ada, di R bisa menggunakan str() dan head() melihat 10 data pertama.

Cek data

Dengan pengecekan ini kita bisa tipe datanya apakah integer, numeric atau character.

```
str(ft50)
```

```
## Classes 'data.table' and 'data.frame': 50 obs. of 9 variables:
## $ Rank : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Restaurant : chr "Evergreens" "Clean Juice" "Slapfish" "Clean Eatz" ...
## $ Location : chr "Seattle, Wash." "Charlotte, N.C." "Huntington Beach,
Calif." "Wilmington, N.C." ...
## $ Sales : int 24 44 21 25 49 39 24 20 24 29 ...
## $ YOY_Sales : chr "130.5%" "121.9%" "81.0%" "79.7%" ...
## $ Units : int 26 105 21 46 50 76 36 19 60 17 ...
## $ YOY_Units : chr "116.7%" "94.4%" "90.9%" "58.6%" ...
## $ Unit_Volume: int 1150 560 1370 685 1210 580 775 1260 465 1930 ...
## $ Franchising: chr "No" "Yes" "Yes" "Yes" ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
str(ip100)
```

```
## Classes 'data.table' and 'data.frame': 100 obs. of 7 variables:
## $ Rank : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Restaurant : chr "Carmine's (Times Square)" "The Boathouse Orlando" "Old
Ebbitt Grill" "LAVO Italian Restaurant & Nightclub" ...
## $ Sales : num 39080335 35218364 29104017 26916180 26900000 ...
## $ Average Check: int 40 43 33 90 62 80 103 99 87 107 ...
## $ City : chr "New York" "Orlando " "Washington" "New York" ...
## $ State : chr "N.Y." "Fla." "D.C." "N.Y." ...
## $ Meals Served : num 469803 820819 892830 198500 403000 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
str(tp250)
```

```
## Classes 'data.table' and 'data.frame': 250 obs. of 9 variables:
## $ Rank : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Restaurant : chr "McDonald's" "Starbucks" "Chick-fil-A" "Taco Bell" ...
## $ Content : chr NA NA "While Popeyes got a lot of the chicken buzz in 2019,
Chick-fil-A had a busy year in its own right. On top of pa"| __truncated__ NA
...
## $ Sales : int 40412 21380 11320 11293 10204 10200 9762 9228 7044 5890 ...
## $ YOY_Sales : chr "4.9%" "8.6%" "13.0%" "9.0%" ...
## $ Units : int 13846 15049 2470 6766 7346 23801 5852 9630 6126 2160 ...
## $ YOY_Units : chr "-0.5%" "3.0%" "5.0%" "2.7%" ...
## $ Headquarters : chr NA NA NA NA ...
## $ Segment_Category: chr "Quick Service & Burger" "Quick Service & Coffee
Cafe" "Quick Service & Chicken" "Quick Service & Mexican" ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
head(ft50)
```

```
##    Rank  Restaurant                Location Sales YOY_Sales Units YOY_Units
## 1:    1  Evergreens          Seattle, Wash.    24    130.5%    26    116.7%
## 2:    2 Clean Juice         Charlotte, N.C.    44    121.9%   105     94.4%
## 3:    3    Slapfish Huntington Beach, Calif.    21     81.0%    21     90.9%
## 4:    4  Clean Eatz         Wilmington, N.C.    25     79.7%    46     58.6%
```

```
## 5:    5   Pokeworks           Irvine, Calif.    49     77.1%    50     56.3%
## 6:    6 Playa Bowls            Belmar,  N.J.    39     62.9%    76     28.8%
##    Unit_Volume Franchising
## 1:        1150          No
## 2:         560         Yes
## 3:        1370         Yes
## 4:         685         Yes
## 5:        1210         Yes
## 6:         580         Yes
```

```
head(ip100)
```

```
##    Rank                           Restaurant    Sales Average Check       City
## 1:    1            Carmine's (Times Square) 39080335            40   New York
## 2:    2               The Boathouse Orlando 35218364            43   Orlando
## 3:    3                    Old Ebbitt Grill 29104017            33 Washington
## 4:    4 LAVO Italian Restaurant & Nightclub 26916180            90   New York
## 5:    5            Bryant Park Grill & Cafe 26900000            62   New York
## 6:    6            Gibsons Bar & Steakhouse 25409952            80    Chicago
##    State Meals Served
## 1:  N.Y.       469803
## 2:  Fla.       820819
## 3:  D.C.       892830
## 4:  N.Y.       198500
## 5:  N.Y.       403000
## 6:  Ill.       348567
```

```
head(tp250)
```

```
## Rank Restaurant
## 1: 1 McDonald's
## 2: 2 Starbucks
## 3: 3 Chick-fil-A
## 4: 4 Taco Bell
## 5: 5 Burger King
## 6: 6 Subway
## Content
## 1: <NA>
## 2: <NA>
## 3: While Popeyes got a lot of the chicken buzz in 2019, Chick-fil-A had a
busy year in its own right. On top of passing Taco Bell and Subway to become
the nation's third-largest chain, it introduced dine-in mobile ordering,
eliminated antibiotics from its chicken and added its first new permanent menu
item in three years: mac and cheese.
## 4: <NA>
## 5: <NA>
## 6: <NA>
## Sales YOY_Sales Units YOY_Units Headquarters Segment_Category
## 1: 40412 4.9% 13846 -0.5% <NA> Quick Service & Burger
## 2: 21380 8.6% 15049 3.0% <NA> Quick Service & Coffee Cafe
## 3: 11320 13.0% 2470 5.0% <NA> Quick Service & Chicken
## 4: 11293 9.0% 6766 2.7% <NA> Quick Service & Mexican
## 5: 10204 2.7% 7346 0.2% <NA> Quick Service & Burger
## 6: 10200 -2.0% 23801 -4.0% <NA> Quick Service & Sandwich
```

Setelah melihat tipe data dan datanya, ada beberapa yang perlu diperbaiki. Kita perbaiki satu persatu datasetnya.

Eksplore Future 50

Kita akan mencoba eksplore data futre 50, yang berisikan perkiraan rangking 50 besar tahun depan berdasarkan data tahun 2020. Melihat data ft50, ternyata ada bebrapa kolom yang yang harus diperbaiki, seperti pemisahan data dan perubahan tipe data.

```
ft50[, .N, Location]
```

```
##                    Location N
##  1:           Seattle, Wash. 1
##  2:          Charlotte, N.C. 2
##  3: Huntington Beach, Calif. 1
##  4:         Wilmington, N.C. 1
##  5:           Irvine, Calif. 1
##  6:            Belmar,  N.J. 1
##  7:           Blue Bell, Pa. 1
##  8:           New York, N.Y. 8
##  9:     Yorba Linda,  Calif. 1
## 10:          Louisville, Ky. 1
## 11:        Spartanburg, S.C. 1
## 12:         Pasadena, Calif. 1
## 13:            Denver, Colo. 1
## 14:             Plano, Texas 1
## 15:          Kettering, Ohio 1
## 16:    San Francisco, Calif. 2
## 17:        San Ramon, Calif. 1
## 18:            Orlando, Fla. 1
## 19:        Orange Park, Fla. 1
## 20:              Doral, Fla. 1
## 21:       Mechanicsburg, Pa. 1
## 22:            Olivette, Mo. 1
## 23:          Columbus, Ohio 3
## 24:      Wall Township, N.J. 1
## 25:          Anaheim, Calif. 1
## 26:           Frisco, Texas 1
## 27:            Fairfax, Va. 1
## 28:            Douglas, Ga. 1
## 29:        Scottsdale, Ariz. 1
## 30:             Atlanta, Ga. 1
## 31:              Omaha, Neb. 1
## 32:            Medford, Ore. 1
## 33:      Los Angeles, Calif. 1
## 34:             Conway, Ark. 1
## 35:            Fairburn, Ga. 1
## 36:           McAllen, Texas 1
## 37:         Washington, D.C. 1
## 38:     Agoura Hills, Calif. 1
## 39:           Memphis, Tenn. 1
##                    Location N
```

Dari data tersebut, negara bagian masih menjadi satu. Untuk itu perlu pemisahan.

```
ft50_clean <- setDT(ft50)[, paste0(c("City", "State")) := tstrsplit(Location,
 ↪  ", ")]
```

Mendapat 2 tabel baru berupa franchise dan negara bagian, pada negara bagian akan dibuat dengan penamaan yang sesuai seperti Calif. menjadi CA.

```
### melihat negara bagian
ft50_clean[, .N, State]
```

```
##       State N
##  1:   Wash. 1
##  2:    N.C. 3
##  3:  Calif. 9
##  4:    N.J. 1
##  5:     Pa. 2
##  6:    N.Y. 8
##  7:  Calif. 1
##  8:     Ky. 1
##  9:    S.C. 1
## 10:   Colo. 1
## 11:   Texas 3
## 12:    Ohio 4
## 13:    Fla. 3
## 14:     Mo. 1
## 15:    N.J. 1
## 16:     Va. 1
## 17:     Ga. 3
## 18:   Ariz. 1
## 19:    Neb. 1
## 20:    Ore. 1
## 21:    Ark. 1
## 22:    D.C. 1
## 23:   Tenn. 1
##       State N
```

```
### Merubah negara bagian
ft50_clean$State <- gsub( "\\.", "", str_squish(str_to_lower(ft50_clean$State)))
ft50_clean[State == "ar", State := "AR"]
ft50_clean[State == "ariz", State := "AZ"]
ft50_clean[State == "ark", State := "AR"]
ft50_clean[State == "az", State := "AZ"]
ft50_clean[State == "ca", State := "CA"]
ft50_clean[State == "calif", State := "CA"]
ft50_clean[State == "co", State := "CO"]
ft50_clean[State == "colo", State := "CO"]
ft50_clean[State == "dc", State := "DC"]
ft50_clean[State == "fl", State := "FL"]
ft50_clean[State == "fla", State := "FL"]
ft50_clean[State == "ga", State := "GA"]
ft50_clean[State == "ill", State := "IL"]
ft50_clean[State == "ind", State := "IN"]
ft50_clean[State == "ky", State := "KY"]
ft50_clean[State == "mass", State := "MA"]
ft50_clean[State == "mich", State := "MI"]
```

```
ft50_clean[State == "mo", State := "MO"]
ft50_clean[State == "nc", State := "NC"]
ft50_clean[State == "ne", State := "NE"]
ft50_clean[State == "neb", State := "NE"]
ft50_clean[State == "nj", State := "NJ"]
ft50_clean[State == "ny", State := "NY"]
ft50_clean[State == "nev", State := "NV"]
ft50_clean[State == "oh", State := "OH"]
ft50_clean[State == "ohio", State := "OH"]
ft50_clean[State == "or", State := "OR"]
ft50_clean[State == "ore", State := "OR"]
ft50_clean[State == "pa", State := "PA"]
ft50_clean[State == "sc", State := "SC"]
ft50_clean[State == "tenn", State := "TN"]
ft50_clean[State == "texas", State := "TX"]
ft50_clean[State == "tn", State := "TN"]
ft50_clean[State == "tx", State := "TX"]
ft50_clean[State == "va", State := "VA"]
ft50_clean[State == "wa", State := "WA"]
ft50_clean[State == "wash", State := "WA"]
```

Selanjutnya, merubah YOY Sales dan YOY Unit dari char ke num.

```
ft50_clean$YOY_Units <- sub("%", "", ft50_clean$YOY_Units)
ft50_clean$YOY_Sales <- sub("%", "", ft50_clean$YOY_Sales)
ft50_clean <-
  ft50_clean %>%
  mutate(YOY_Sales = as.numeric(YOY_Sales),
         YOY_Units = as.numeric(YOY_Units))
ft50_clean <- ft50_clean[,c(1,2,3,10,11,4:9)]
ft50_clean <- select(ft50_clean, -Location)
```

Data sudah clean, selanjutnya mencoba melihat ringkasan data ft50.

```
### Cek tipe data
str(ft50_clean)
```

```
## Classes 'data.table' and 'data.frame': 50 obs. of 10 variables:
## $ Rank : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Restaurant : chr "Evergreens" "Clean Juice" "Slapfish" "Clean Eatz" ...
## $ City : chr "Seattle" "Charlotte" "Huntington Beach" "Wilmington" ...
## $ State : chr "WA" "NC" "CA" "NC" ...
## $ Sales : int 24 44 21 25 49 39 24 20 24 29 ...
## $ YOY_Sales : num 130.5 121.9 81 79.7 77.1 ...
## $ Units : int 26 105 21 46 50 76 36 19 60 17 ...
## $ YOY_Units : num 116.7 94.4 90.9 58.6 56.3 ...
## $ Unit_Volume: int 1150 560 1370 685 1210 580 775 1260 465 1930 ...
## $ Franchising: chr "No" "Yes" "Yes" "Yes" ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
### Ringkasan data
summary(ft50_clean)
```

```
##       Rank        Restaurant            City              State
##  Min.   : 1.00   Length:50          Length:50          Length:50
```

```
##  1st Qu.:13.25   Class :character   Class :character   Class :character
##  Median :25.50   Mode  :character   Mode  :character   Mode  :character
##  Mean   :25.50
##  3rd Qu.:37.75
##  Max.   :50.00
##      Sales          YOY_Sales          Units          YOY_Units
##  Min.   :20.00   Min.   : 14.40   Min.   :  7.0   Min.   :  4.00
##  1st Qu.:24.25   1st Qu.: 20.90   1st Qu.: 16.0   1st Qu.: 14.30
##  Median :34.50   Median : 25.50   Median : 27.0   Median : 19.90
##  Mean   :33.78   Mean   : 33.70   Mean   : 34.7   Mean   : 27.45
##  3rd Qu.:42.00   3rd Qu.: 33.83   3rd Qu.: 45.5   3rd Qu.: 32.67
##  Max.   :49.00   Max.   :130.50   Max.   :105.0   Max.   :116.70
##   Unit_Volume     Franchising
##  Min.   : 465.0   Length:50
##  1st Qu.: 867.5   Class :character
##  Median :1260.0   Mode  :character
##  Mean   :1592.6
##  3rd Qu.:2020.0
##  Max.   :4300.0
```

```r
### Cek data null
sum(is.na(ft50_clean))
```

```
## [1] 0
```

Ternyata tidak ada data yang null, selanjutnya mari membuat grafik dari data ft50. Dengan rata-rata penjualannya adalah 33.78 dan rata-rata YOY Sales selama 2019-2020 adalah 33.70 dan rata-rata unitnya adalag 34.7.

Membuat plot

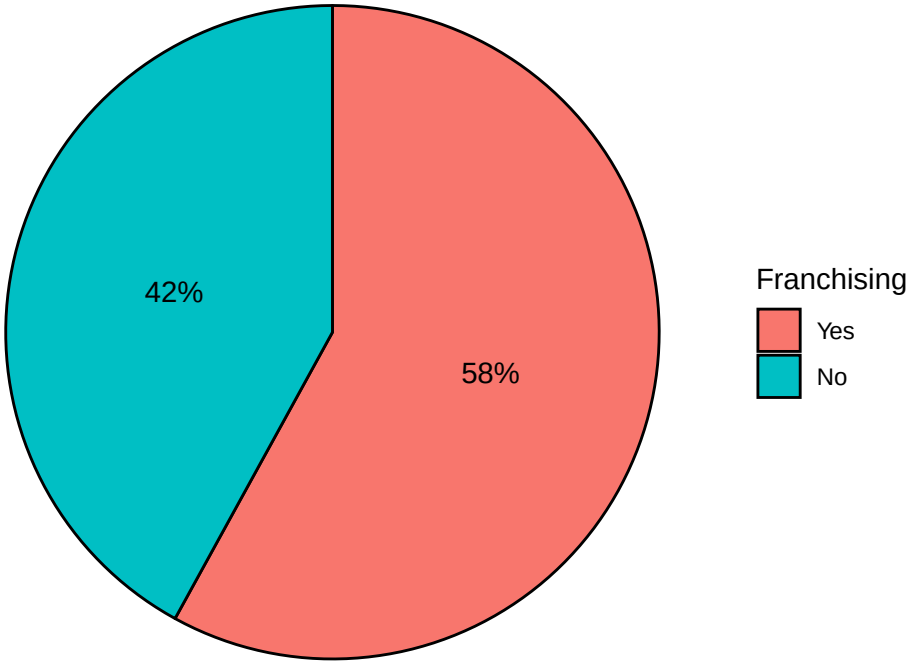
Mari kita cek, berapa persen Restoran yang Franchising dalam future 50 ini?

```r
franchise <- data.frame(sort(table(ft50_clean$Franchising),decreasing = TRUE ))
setnames(franchise,c("Franchising","Jumlah"))

pct <- round(100*franchise$Jumlah/sum(franchise$Jumlah))

ggplot(data = franchise, aes(x = "", y = -Jumlah,
                             fill = Franchising)) +
  geom_bar(stat = "identity", color = "black") +
  labs(title = "Persentase Restoran Franchise") +
  coord_polar("y") +
  geom_text(aes(label =paste0(pct, "%")), position = position_stack(vjust =
    ↪  0.5)) +
  theme_void()
```

Persentase Restoran Franchise

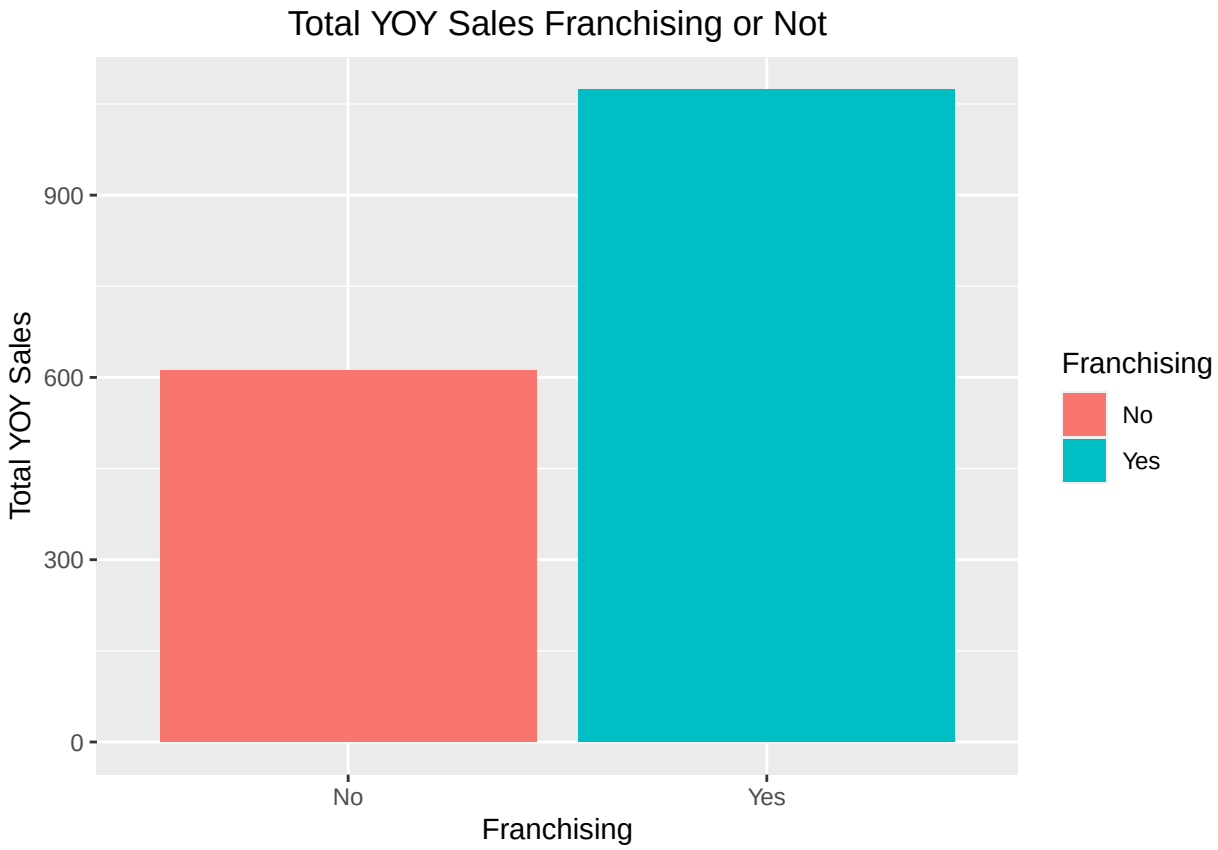


Dari pie chart dapat diketahui bahwa 58% Restoran dalam yang masuk future 50 adalah restoran waralaba dan sisanya 48% adalah bukan waralaba.

Selanjutnya kita lihat perbandingan total YOY Sales antara Restoran Franchise dan tidak.

```
yoysalesbyfranchise <- aggregate(ft50_clean$YOY_Sales,
  ↪   by=list(franchising=ft50_clean$Franchising), FUN=sum)
setnames(yoysalesbyfranchise,c("Franchising","Total_YOY_Sales"))

ggplot(yoysalesbyfranchise, aes(x=Franchising, y=Total_YOY_Sales,
  ↪   fill=Franchising)) +
  labs(title="Total YOY Sales Franchising or Not",
         x ="Franchising", y = "Total YOY Sales") +
  geom_bar(stat="identity", position=position_dodge()) +
  theme(plot.title = element_text(hjust = 0.5))
```



Dari grafik tersebut, restoran yang masuk Future 50 mengindikasikan bahwa Restoran Waralaba memiliki penjualan yang cukup banyak selama pandemi ini.

Eksplore Independence 100

Sekarang kita coba eksplorasi 100 restoran independen. Tapi sebelum kita lakukan cleansing data terlebih dahulu.

```
### Cek data
str(ip100)
```

```
## Classes 'data.table' and 'data.frame': 100 obs. of 7 variables:
## $ Rank : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Restaurant : chr "Carmine's (Times Square)" "The Boathouse Orlando" "Old
Ebbitt Grill" "LAVO Italian Restaurant & Nightclub" ...
## $ Sales : num 39080335 35218364 29104017 26916180 26900000 ...
## $ Average Check: int 40 43 33 90 62 80 103 99 87 107 ...
## $ City : chr "New York" "Orlando " "Washington" "New York" ...
## $ State : chr "N.Y." "Fla." "D.C." "N.Y." ...
## $ Meals Served : num 469803 820819 892830 198500 403000 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

```
head(ip100)
```

```
##    Rank                        Restaurant    Sales Average Check       City
## 1:    1          Carmine's (Times Square) 39080335            40   New York
## 2:    2             The Boathouse Orlando 35218364            43   Orlando
## 3:    3                  Old Ebbitt Grill 29104017            33 Washington
```

```
## 4:    4 LAVO Italian Restaurant & Nightclub 26916180           90   New York
## 5:    5            Bryant Park Grill & Cafe 26900000           62   New York
## 6:    6            Gibsons Bar & Steakhouse 25409952           80    Chicago
##    State Meals Served
## 1:  N.Y.       469803
## 2:  Fla.       820819
## 3:  D.C.       892830
## 4:  N.Y.       198500
## 5:  N.Y.       403000
## 6:  Ill.       348567
```

Ternyata negara bagian masih menggunkan singkatan yang kurang sesuai dengan singakatan standarnya, markicek, mari kita cek.

```
### Cek data
ip100[, .N, State]
```

```
##      State  N
##  1:   N.Y. 21
##  2:   Fla.  9
##  3:   D.C.  9
##  4:   Ill. 18
##  5:   Nev. 11
##  6:   N.C.  1
##  7:   Ind.  2
##  8:  Texas  3
##  9:    Pa.  1
## 10: Calif. 12
## 11:    Ga.  2
## 12:  Mich.  2
## 13:  Mass.  1
## 14:   Ore.  1
## 15:   N.J.  2
## 16:  Fla.   1
## 17:  Tenn.  2
## 18:  Colo.  1
## 19:    Va.  1
```

Ada 19 negara bagian yang harus di cleansing.

```
ip100_clean <- ip100
ip100_clean$State <- gsub( "\\.", "", str_squish(str_to_lower(ip100$State)))
ip100_clean[State == "calif", State := "CA"]
ip100_clean[State == "colo", State := "CO"]
ip100_clean[State == "dc", State := "DC"]
ip100_clean[State == "fla", State := "FL"]
ip100_clean[State == "ga", State := "GA"]
ip100_clean[State == "ill", State := "IL"]
ip100_clean[State == "ind", State := "IN"]
ip100_clean[State == "mass", State := "MA"]
ip100_clean[State == "mich", State := "MI"]
ip100_clean[State == "mo", State := "MO"]
ip100_clean[State == "nc", State := "NC"]
ip100_clean[State == "nj", State := "NJ"]
ip100_clean[State == "ny", State := "NY"]
ip100_clean[State == "nev", State := "NV"]
```

```
ip100_clean[State == "ore", State := "OR"]
ip100_clean[State == "pa", State := "PA"]
ip100_clean[State == "tenn", State := "TN"]
ip100_clean[State == "texas", State := "TX"]
ip100_clean[State == "va", State := "VA"]
```

Setelah cleansing, kita lihat summary datanya.

```
summary(ip100_clean)
```

```
##       Rank         Restaurant            Sales          Average Check
##  Min.   :  1.00   Length:100         Min.   :11391678   Min.   : 17.00
##  1st Qu.: 25.75   Class :character   1st Qu.:14094836   1st Qu.: 39.00
##  Median : 50.50   Mode  :character   Median :17300776   Median : 65.50
##  Mean   : 50.50                      Mean   :17833434   Mean   : 69.05
##  3rd Qu.: 75.25                      3rd Qu.:19903916   3rd Qu.: 95.00
##  Max.   :100.00                      Max.   :39080335   Max.   :194.00
##      City              State           Meals Served
##  Length:100         Length:100         Min.   : 87070
##  Class :character   Class :character   1st Qu.:189492
##  Mode  :character   Mode  :character   Median :257097
##                                        Mean   :317167
##                                        3rd Qu.:372079
##                                        Max.   :959026
```

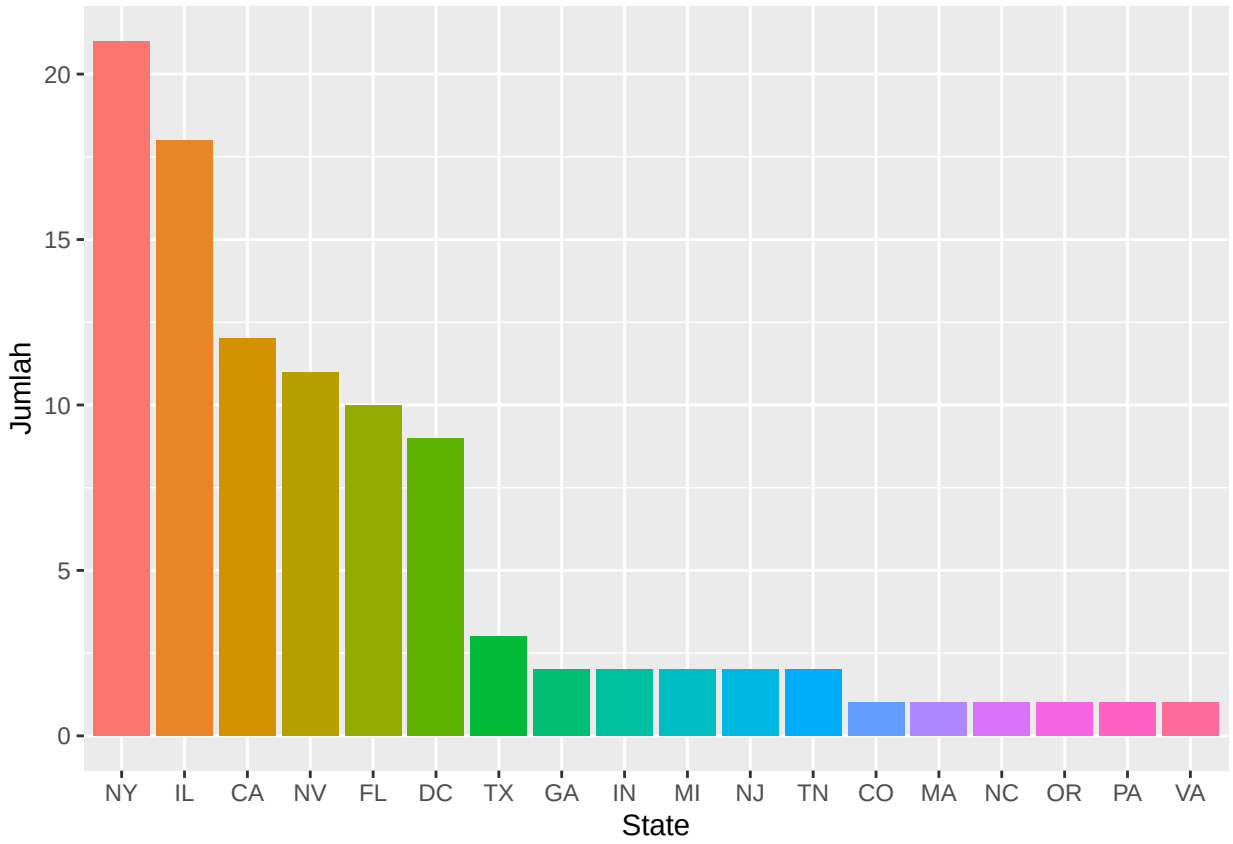
Rata-rata penjualan dari 100 Restoran Independen adalah 17833434 dengan penjualan paling sedikit 11391678 dan penjualan paling banyak 39080335. Dan rata-rata menyajikan makanan sebanyak 317167.

Membuat plot Independen 100

Histogram persebaran 100 Restoran independen di Amerika Serikat.

```
sebar <- data.frame(sort(table(ip100_clean$State),decreasing = TRUE ))
setnames(sebar,c("State","Jumlah"))

ggplot(sebar, aes(x=State, y=Jumlah, fill=State)) +
  geom_bar(stat="identity", position=position_dodge()) +
  theme(legend.position="none", plot.title = element_text(hjust = 0.5)) +
  scale_colour_brewer(type = "seq", palette = "Spectral")
```




```
salesny <-
  ip100_clean %>%
  filter(State == "NY")
mean(salesny$Sales)
```

```
## [1] 19355896
```

Dari grafik tersebut, sebanyak 21 Restoran independen berada di New York dengan rata-rata penjualan 19355896.