Waiting_Times

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

```r
library(dplyr)
```

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

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

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

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

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

```r
questions <- read.csv("/Users/pivaldhingra/Desktop/University courses/Datafest 2023/DataFest 2023 Data
data <- data.frame(questions)
df <- select(data, "StateAbbr", "AskedOnUtc", "TakenOnUtc")

df$AskedOnUtc <- as.POSIXct(df$AskedOnUtc, format = "%Y-%m-%d %H:%M:%OS" )
df$TakenOnUtc <- as.POSIXct(df$TakenOnUtc, format = "%Y-%m-%d %H:%M:%OS")

df$Year <- format(df$AskedOnUtc, "%Y")

df_filtered <- df %>%
  filter(!is.na(TakenOnUtc)) %>%
  filter(Year %in% c("2016", "2017", "2018", "2019", "2020", "2021", "2022"))

# Calculate the waiting time for each question
df_filtered$WaitingTime <- difftime(df_filtered$TakenOnUtc, df_filtered$AskedOnUtc, units = "mins")

# Group the data by state and year, and calculate the average waiting time for each group
df_summary <- df_filtered %>%
  group_by(StateAbbr, Year) %>%
  summarize(AvgWaitingTime = mean(WaitingTime))
```

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

df_2016 <- df_summary %>%
  filter(Year == 2016) %>%
  select(StateAbbr, AvgWaitingTime) %>%
  group_by(StateAbbr) %>%
  summarise(avg_waiting_time = mean(AvgWaitingTime, na.rm = TRUE)) %>%
  arrange(desc(avg_waiting_time))

top_5_states <- df_2016 %>% top_n(5, avg_waiting_time)

print(top_5_states)
```

```
## # A tibble: 5 x 2
##   StateAbbr avg_waiting_time
##   <chr>        <drtn>
## 1 SC        62517.05 mins
## 2 MS        33137.63 mins
## 3 IN        25873.02 mins
## 4 AZ        24194.49 mins
## 5 WI        24181.50 mins
```

```
df_2017 <- df_summary %>%
  filter(Year == 2017) %>%
  select(StateAbbr, AvgWaitingTime) %>%
  group_by(StateAbbr) %>%
  summarise(avg_waiting_time = mean(AvgWaitingTime, na.rm = TRUE)) %>%
  arrange(desc(avg_waiting_time))

top_5_states <- df_2017 %>% top_n(5, avg_waiting_time)

print(top_5_states)
```

```
## # A tibble: 5 x 2
##   StateAbbr avg_waiting_time
##   <chr>        <drtn>
## 1 SC        75251.24 mins
## 2 GA        71560.04 mins
## 3 ME        43540.95 mins
## 4 KS        41354.06 mins
## 5 MS        34995.84 mins
```

```
df_2018 <- df_summary %>%
  filter(Year == 2018) %>%
  select(StateAbbr, AvgWaitingTime) %>%
  group_by(StateAbbr) %>%
  summarise(avg_waiting_time = mean(AvgWaitingTime, na.rm = TRUE)) %>%
  arrange(desc(avg_waiting_time))

top_5_states <- df_2018 %>% top_n(5, avg_waiting_time)

print(top_5_states)
```

```
## # A tibble: 5 x 2
##   StateAbbr avg_waiting_time
##   <chr>           <drtn>
## 1 GA         61762.25 mins
## 2 IA         33231.34 mins
## 3 IN         32680.18 mins
## 4 SC         27781.67 mins
## 5 CT         25201.63 mins
```

```
df_2019 <- df_summary %>%
  filter(Year == 2019) %>%
  select(StateAbbr, AvgWaitingTime) %>%
  group_by(StateAbbr) %>%
  summarise(avg_waiting_time = mean(AvgWaitingTime, na.rm = TRUE)) %>%
  arrange(desc(avg_waiting_time))

top_5_states <- df_2019 %>% top_n(5, avg_waiting_time)

print(top_5_states)
```

```
## # A tibble: 5 x 2
##   StateAbbr avg_waiting_time
##   <chr>           <drtn>
## 1 SC         56780.11 mins
## 2 GA         46076.45 mins
## 3 KS         36629.16 mins
## 4 MS         26457.15 mins
## 5 MD         23939.34 mins
```

```
df_2020 <- df_summary %>%
  filter(Year == 2020) %>%
  select(StateAbbr, AvgWaitingTime) %>%
  group_by(StateAbbr) %>%
  summarise(avg_waiting_time = mean(AvgWaitingTime, na.rm = TRUE)) %>%
  arrange(desc(avg_waiting_time))

top_5_states <- df_2020 %>% top_n(5, avg_waiting_time)
print(top_5_states)
```

```
## # A tibble: 5 x 2
##   StateAbbr avg_waiting_time
##   <chr>           <drtn>
## 1 UT         87011.47 mins
## 2 GA         40002.13 mins
## 3 AL         36287.65 mins
## 4 MS         30608.04 mins
## 5 SC         24707.95 mins
```

```
df_2021 <- df_summary %>%
  filter(Year == 2021) %>%
  select(StateAbbr, AvgWaitingTime) %>%
  group_by(StateAbbr) %>%
```

```
  summarise(avg_waiting_time = mean(AvgWaitingTime, na.rm = TRUE)) %>%
  arrange(desc(avg_waiting_time))

top_5_states <- df_2021 %>% top_n(5, avg_waiting_time)

print(top_5_states)
```

```
## # A tibble: 5 x 2
##   StateAbbr avg_waiting_time
##   <chr>          <drtn>
## 1 AL         53421.14 mins
## 2 GA         43393.25 mins
## 3 SC         35926.66 mins
## 4 IA         24732.05 mins
## 5 PA         23542.64 mins
```

```
df_2022 <- df_summary %>%
  filter(Year == 2022) %>%
  select(StateAbbr, AvgWaitingTime) %>%
  group_by(StateAbbr) %>%
  summarise(avg_waiting_time = mean(AvgWaitingTime, na.rm = TRUE)) %>%
  arrange(desc(avg_waiting_time))

top_5_states <- df_2022 %>% top_n(5, avg_waiting_time)

print(top_5_states)
```

```
## # A tibble: 5 x 2
##   StateAbbr avg_waiting_time
##   <chr>          <drtn>
## 1 CT         9958.092 mins
## 2 US         7957.863 mins
## 3 OK         7727.069 mins
## 4 AL         5595.972 mins
## 5 NH         4720.101 mins
```