Comparative Analysis of Coffee Habits Between Remote and In-Person Workers

Table of contents

Executive Summary

Introduction

Methodology

```r
library(ggplot2)
library(readr)
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v dplyr     1.1.4      v stringr   1.5.1
v forcats   1.0.0      v tibble    3.2.1
v lubridate 1.9.4      v tidyr     1.3.1
v purrr     1.0.4
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
library(patchwork)
library(knitr)
```

```r
coffee_survey <- read_csv(here::here("data/coffee_survey.csv"))
```

```
Rows: 4042 Columns: 57
-- Column specification ------------------------------------------------------
Delimiter: ","
chr (44): submission_id, age, cups, where_drink, brew, brew_other, purchase,...
dbl (13): expertise, coffee_a_bitterness, coffee_a_acidity, coffee_a_persona...

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.
```

```r
coffee_survey_selected <- coffee_survey %>%
  select("submission_id", "wfh", "cups", "total_spend", "most_paid")
```

```r
work_mode_summary <- coffee_survey_selected %>%
  filter(!is.na(wfh)) %>%
  count(wfh) %>%
  mutate(work_mode = wfh,
         proportion = scales::percent(n/sum(n))) %>%
  select(work_mode, n, proportion)

work_mode_summary %>%
  kable(caption = "Table 1: Proportion of Participants by Work Mode")
```

Table 1: Table 1: Proportion of Participants by Work Mode

work_mode	n	proportion
I do a mix of both	883	25.1%
I primarily work from home	1592	45.2%
I primarily work in person	1049	29.8%

According to table @table1, we can see that among the survey participants, those who work from home make up a significantly larger portion, accounting for 45.2% of the total. Those

who work only in the office represent 29.8%, which is 4.7% more than those with a hybrid
work arrangement.

```r
cup_sum <- coffee_survey_selected %>%
  filter(!is.na(cups), !is.na(wfh)) %>%
  count(cups, wfh) %>%
  mutate(cups = factor(cups, levels = c("Less than 1", "1", "2", "3", "4", "More than 4")))

p1 <- cup_sum %>%
  ggplot(aes(x = cups, y = n, fill = wfh)) +
  geom_col(position = "dodge", width = 0.6) +
  labs(title = "Cups of coffee per day", x = "Cups", y = "Number of people", fill = "Work Mod
  theme_minimal()

total_spend_sum <- coffee_survey_selected %>%
  filter(!is.na(total_spend), !is.na(wfh)) %>%
  count(total_spend, wfh) %>%
  mutate(total_spend = factor(total_spend, levels = c("<$20", "$20-$40", "$40-$60", "$60-$80"

p2 <- total_spend_sum %>%
  ggplot(aes(x = total_spend, y = n, fill = wfh)) +
  geom_col(position = "dodge", width = 0.6) +
  labs(title = "Money spent on coffee per month", x = "Spending Range", y = "Number of People
  theme_minimal()

most_paid_sum <- coffee_survey_selected %>%
  filter(!is.na(most_paid), !is.na(wfh)) %>%
  count(most_paid, wfh) %>%
  mutate(most_paid = factor(most_paid, levels = c("Less than $2", "$2-$4", "$4-$6", "$6-$8",

p3 <- most_paid_sum %>%
  ggplot(aes(x = most_paid, y = n, fill = wfh)) +
  geom_col(position = "dodge", width = 0.6) +
  labs(title = "Maximum Price Paid for a Cup of Coffee", x = "Spending Range", y = "Number o:
  theme_minimal()
```

```r
p1 <- p1 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
p2 <- p2 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
p3 <- p3 + theme(axis.text.x = element_text(angle = 45, hjust = 1))

combined <- (p1 | p2) / (p3 | patchwork::plot_spacer())
combined
```

Cups of coffee per day — Work Mode

Number of people / Cups

Legend:
- I do a mix of both
- I primarily work from home
- I primarily work in person

Money spent on coffee per month — Work Mode

Number of People / Spending Range

Legend:
- I do a mix of both
- I primarily work from home
- I primarily work in person

Maximum Price Paid for a Cup of Coffee — Work Mode

Number of People / Spending Range

Legend:
- I do a mix of both
- I primarily work from home
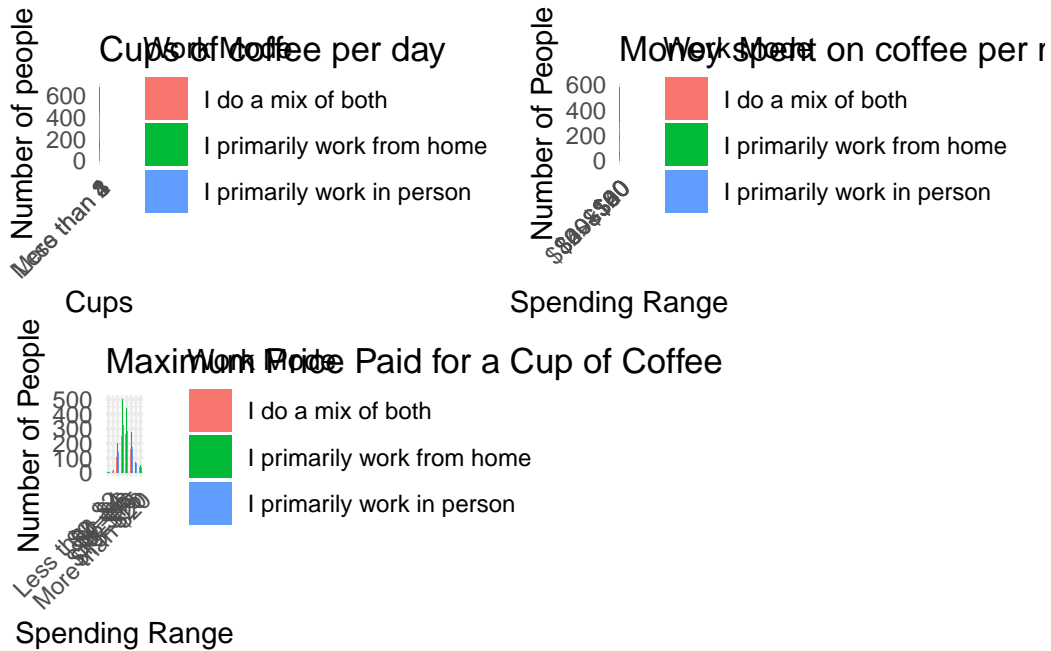- I primarily work in person

Figure @figure1 illustrates that the coffee price preferences of people working from home and those working in the office are generally similar.

In terms of daily consumption, individuals who drink two cups per day make up the largest proportion in both groups. However, people working from home tend to consume more coffee on average than those who work in the office.

Most respondents spend $20–50 per month on coffee, with a higher proportion among remote workers. Very few people spend less than $20 or more than $100 per month, regardless of work setting.

When it comes to the maximum price paid for a single cup, the majority fall within the $6–15 range, especially between $6 and $8.

Results

?@tbl-coffee shows that among the survey participants, the proportion of people working from home is significant, reaching 45.2%. From **?@fig-coffee**, it shows that the preferences for coffee prices are roughly the same for different working methods, and people who drink two cups of coffee a day also account for the highest proportion in both groups. However, the average coffee consumption of people working from home tends to be higher than that of people working in the office. Most respondents spend $20 to $50 on coffee per month. Regardless of the working environment, there are few people who spend less than $20 or more than $100 per month.

```r
clean_data <- coffee_survey_selected %>%
  drop_na()
wfh_data <- clean_data %>%
  mutate(
    cups_numeric = case_when(
      cups == "Less than 1" ~ 0.5,
      cups == "1" ~ 1,
      cups == "2" ~ 2,
      cups == "3" ~ 3,
      cups == "4" ~ 4,
      cups == "More than 4" ~ 5,
      TRUE ~ as.numeric(NA)
    ),
    total_spend_numeric = case_when(
      total_spend == "<$20" ~ 10,
      total_spend == "$20-$40" ~ 30,
      total_spend == "$40-$60" ~ 50,
      total_spend == "$60-$80" ~ 70,
      total_spend == "$80-$100" ~ 90,
      total_spend == ">$100" ~ 110,
      TRUE ~ as.numeric(NA)
    )
  ) %>%
  filter(wfh == "I primarily work from home")
wfh_cluster_data <- wfh_data %>%
  select(cups_numeric, total_spend_numeric) %>%
  na.omit() %>%
  scale()
set.seed(123)
wfh_kmeans <- kmeans(wfh_cluster_data, centers = 3, nstart = 25)
```

```r
fviz_cluster(wfh_kmeans, data = wfh_cluster_data,
             geom = "point",
             ellipse.type = "convex",
             ggtheme = theme_minimal(),
             main = "Cluster Visualization: Work from Home Group")
```
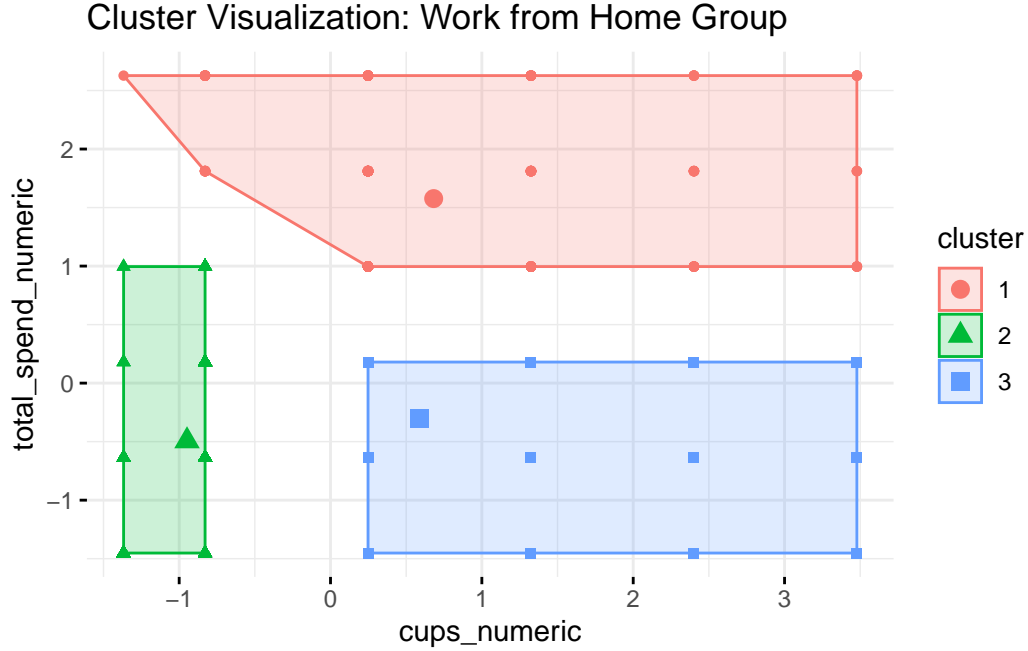
Figure 1: Cluster Visualization: Work from Home Group

Further research on the daily coffee intake and monthly consumption of people working from home can be divided into three groups: high consumption and high intake, low consumption, and conflicting groups. From Figure 1 , it shows that nearly half of the people working from home are high consumption and high intake groups, indicating that this office style is more dependent on coffee and also pursues the quality of coffee.

Conclusion

Discussion

This study explores the relationship between work modes (working from home, working in person and hybrid working) and coffee consumption behaviors (daily intake, monthly expenditure and maximum expenditure). The results show that different groups have similar preferences in coffee consumption and price. However, the remote working group has relatively higher consumption, which also reflects differences in lifestyle, such as the convenience of brewing coffee for people working from home.

Cluster analysis shows the coffee consumption behaviors of different groups about people working from home. High consumption accounting for the majority reflects requirements for coffee quality. However, due to limitations like accuracy, other relevant variables need to be introduced for in-depth investigation.

Conclusin

Overall, the research finds that remote workers not only make up the largest proportion but also tend to consume more coffee and spend more on it monthly and most individuals drink two cups per day. Clustering of remote workers further highlights a substantial group characterized by both high coffee intake and high spending, suggesting a stronger reliance on and appreciation for coffee among those who work from home.

Recommendations

1. Implement health education initiatives to raise awareness about the potential risks of excessive coffee consumption and to encourage healthier consumption patterns. For example, the company could offer alternative beverages such as tea and herbal infusions.

2. Expand the dataset by incorporating additional relevant variables like age and income level, and conduct qualitative analysis to identify the driving factors of high coffee consumption behavior under the same office mode.