

Prosjektoppgave - SOK-1005

Kandidatnr.:

— — 2022

```
library(tidyverse)
library(readr)
library(lubridate)

df1 <- read_csv("C:/Users/mgmal/Desktop/Semester/_Datavit/Prosjekt/AppWichStoreAttributes.csv")
df2 <- read_csv("C:/Users/mgmal/Desktop/Semester/_Datavit/Prosjekt/county_crime.csv")
df3 <- read_csv("C:/Users/mgmal/Desktop/Semester/_Datavit/Prosjekt/county_demographic.csv")
df4 <- read_csv("C:/Users/mgmal/Desktop/Semester/_Datavit/Prosjekt/county_employment.csv")
df5 <- read_csv("C:/Users/mgmal/Desktop/Semester/_Datavit/Prosjekt/weekly_sales_10stores.csv")
df6 <- read_csv("C:/Users/mgmal/Desktop/Semester/_Datavit/Prosjekt/weekly_weather.csv")

# str(df1) # [10 x 14]
# str(df2) # [6 x 12]
# str(df3) # [6 x 14]
# str(df4) # [6 x 5]
# str(df5) # [79,459 x 17]
# str(df6) # [270 x 16]

df1 <- rename(df1, County_Name = Store_County)
df1 <- rename(df1, Store_num = Store_Num)
join1 <- merge(df5, df1) # 79459 obs of 30 variables
join2 <- merge(join1, df2) # 79459 obs of 41 variables
join3 <- merge(join2, df3) # 79459 obs of 54 variables
join4 <- merge(join3, df4) # 79459 obs of 58 variables
df6 <- rename(df6, Store_Weather_Station = Weather_Station)
final_df <- union_all(join4, df6)
final_df$Date <- mdy(final_df$Date)
# str(final_df) # 79459 obs of 73 variables
```

```

count(unique(final_df[c("Description"))]))    #503 forskjellige varer/produkter

##      n
## 1 503

# head(final_df %>% select>Description,Profit,Date,Store_num) %>% arrange(desc(Profit)),10

# head(setNames(aggregate(final_df$Profit, by=list(Category=final_df$Description),
#                                     FUN=sum),c("Description", "Profit")))

final_df %>%
  group_by>Description) %>%
  summarise

```

```

# Binning
pizza <- final_df %>% filter(grepl('PIZZA', Description)) %>%
  select(Description, Profit, Date, Day, Store_num) %>%
  mutate(Description = ifelse(grepl('PIZZA', Description), 'Pizza'))

burger <- final_df %>% filter(grepl('BURGER', Description)) %>%
  select(Description, Profit, Date, Day, Store_num) %>%
  mutate(Description = ifelse(grepl('BURGER', Description), 'Burger'))

sub <- final_df %>% filter(grepl('SUB', Description)) %>%
  select(Description, Profit, Date, Day, Store_num) %>%
  mutate(Description = ifelse(grepl('SUB', Description), 'Sub'))

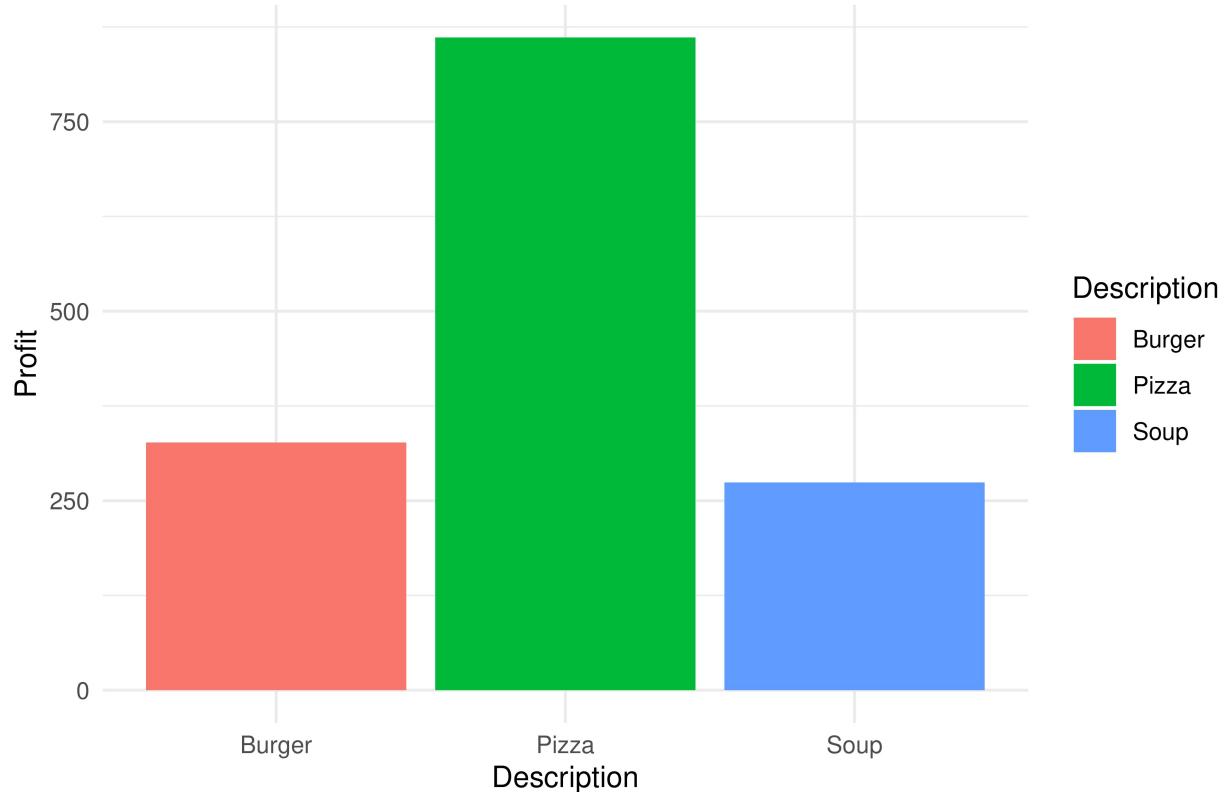
salad <- final_df %>% filter(grepl('SALAD', Description)) %>%
  select(Description, Profit, Date, Day, Store_num) %>%
  mutate(Description = ifelse(grepl('SALAD', Description), 'Salad'))

soup <- final_df %>% filter(grepl('SOUP', Description)) %>%
  select(Description, Profit, Date, Day, Store_num) %>%
  mutate(Description = ifelse(grepl('SOUP', Description), 'Soup'))

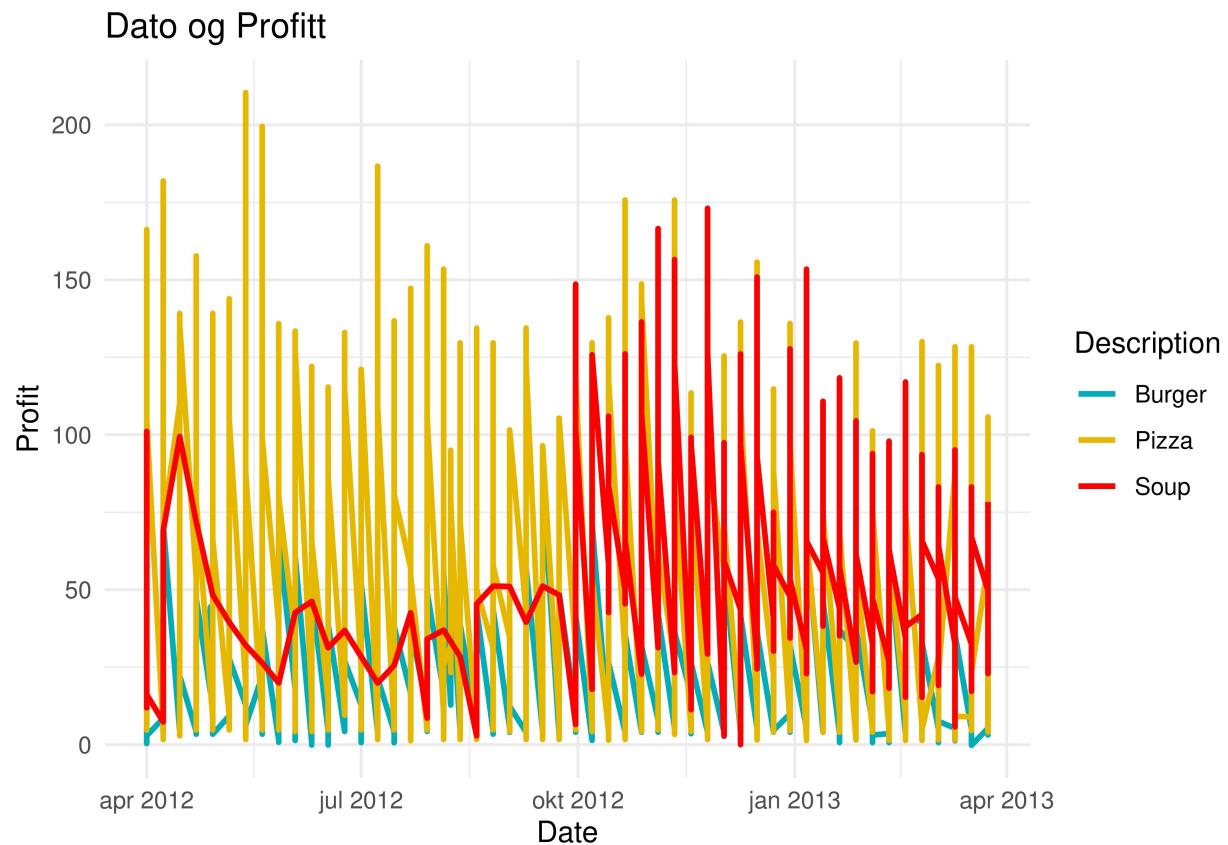
```

```
# Best category
df_barplot2 <- bind_rows(pizza, burger, soup)
df_barplot2 <- df_barplot2 %>% filter(Date=="2012-04-01")
df_barplot2 <- setNames(aggregate(df_barplot2$Profit, by=list(Category=df_barplot2$Description),
                                  FUN=sum),c("Description", "Profit"))
ggplot(df_barplot2, aes(x=Description, y=Profit, fill=Description)) + geom_bar(stat="identity") +
  theme_minimal() + labs(title = "Profitt per varekategori, uke 1")
```

Profitt per varekategori, uke 1



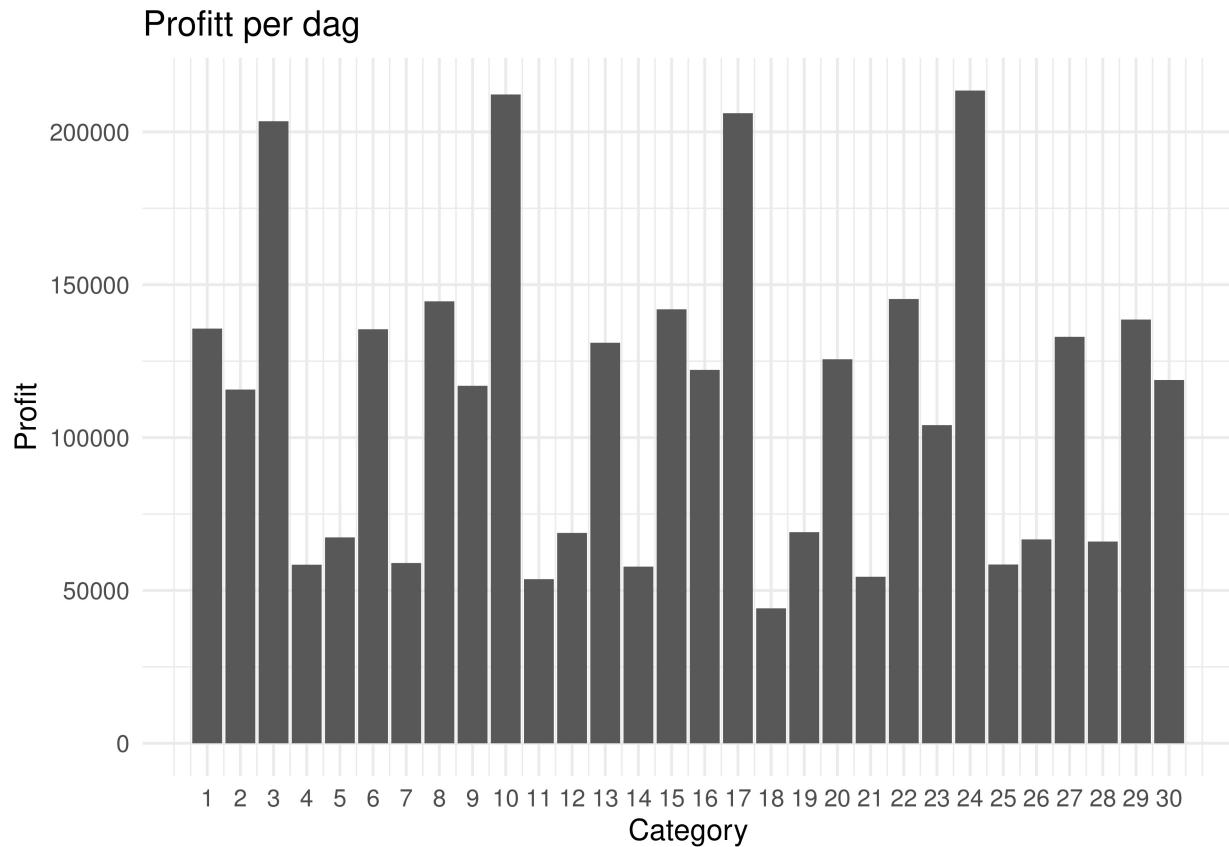
```
# Time series Date Profit
df_barplot6 <- bind_rows(pizza, burger, soup)
ggplot(df_barplot6, aes(x = Date, y = Profit)) +
  geom_line(aes(color = Description), size = 1) +
  scale_color_manual(values = c("#00AFBB", "#E7B800", 'red')) +
  theme_minimal() + labs(title = "Dato og Profitt")
```



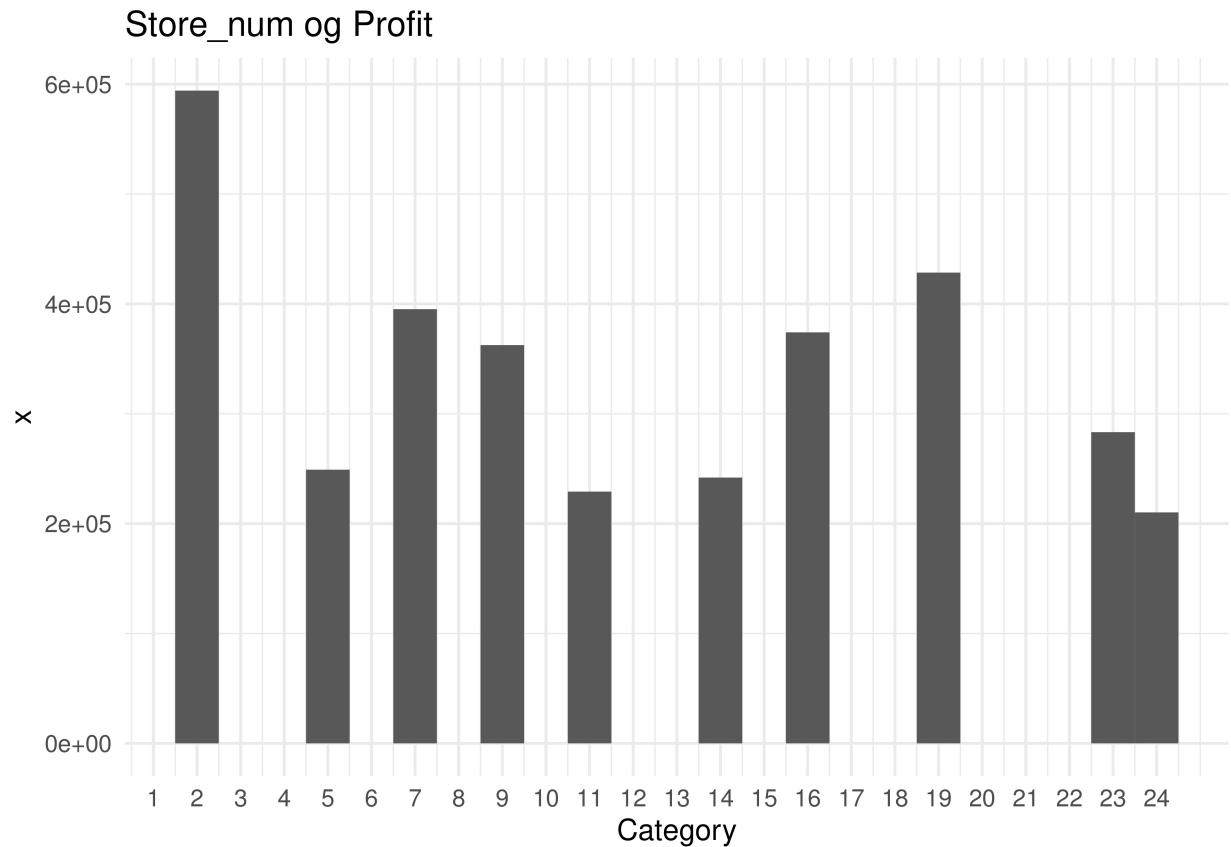
```

# Best day
df_barplot3 <- setNames(aggregate(final_df$Profit, by=list(Category=final_df$Day),
                                         FUN=sum),c("Description", "Profit"))
ggplot(df_barplot3, aes(x=Description, y=Profit)) + geom_bar(stat="identity") +
  theme_minimal() + scale_x_continuous("Category", labels = as.character(1:30), breaks = 1:30) +
  labs(title = "Profitt per dag")

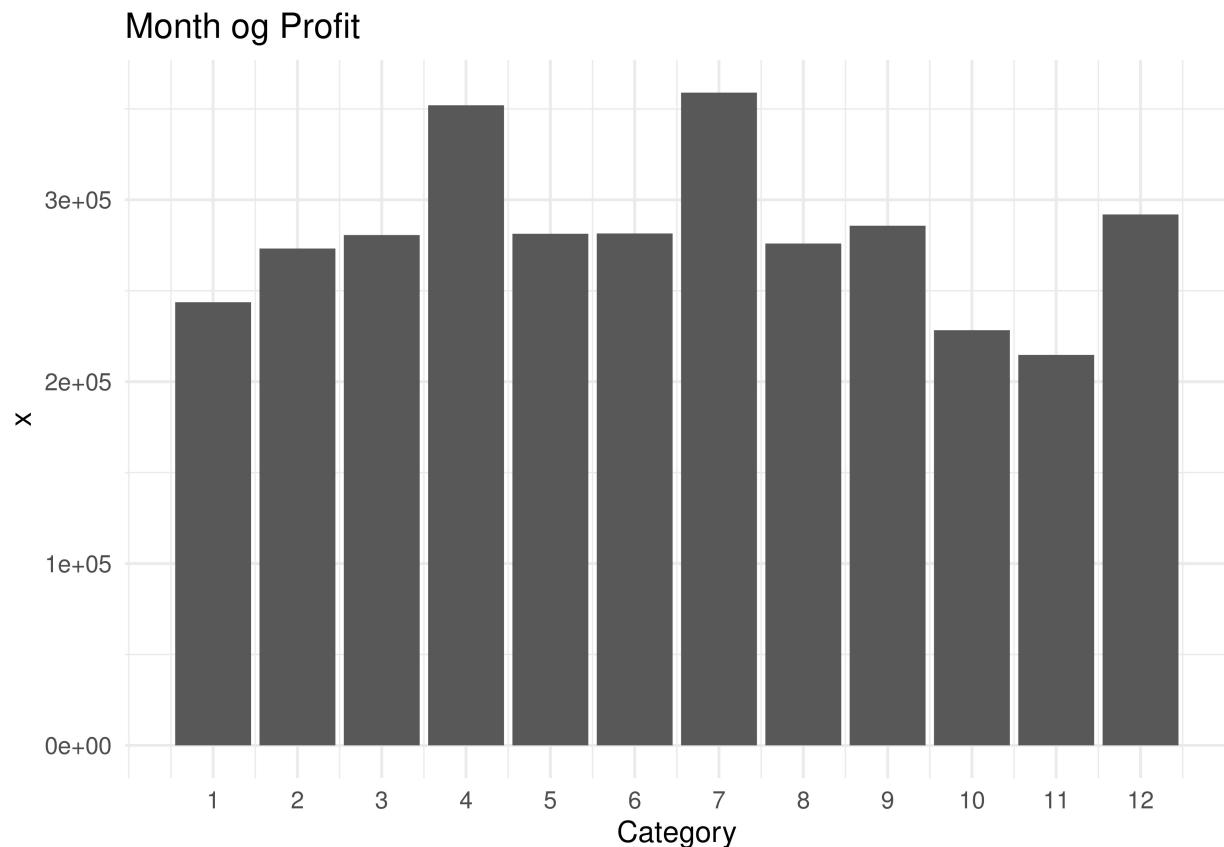
```



```
# Best store
df_barplot4 <- aggregate(final_df$Profit, by=list(Category=final_df$Store_num), FUN=sum)
ggplot(df_barplot4, aes(x=Category, y=x)) + geom_col(width = 1, position = "dodge") +
  theme_minimal() + scale_x_continuous("Category", labels = as.character(1:24), breaks = 1:24) +
  labs(title = "Store_num og Profit")
```



```
# Best month:
df_barplot5 <- aggregate(final_df$Profit, by=list(Category=final_df$Month), FUN=sum)
ggplot(df_barplot5, aes(x=Category, y=x)) + geom_bar(stat="identity") +
  theme_minimal() + scale_x_continuous("Category", labels = as.character(1:12), breaks = 1:12) +
  labs(title = "Month og Profit")
```



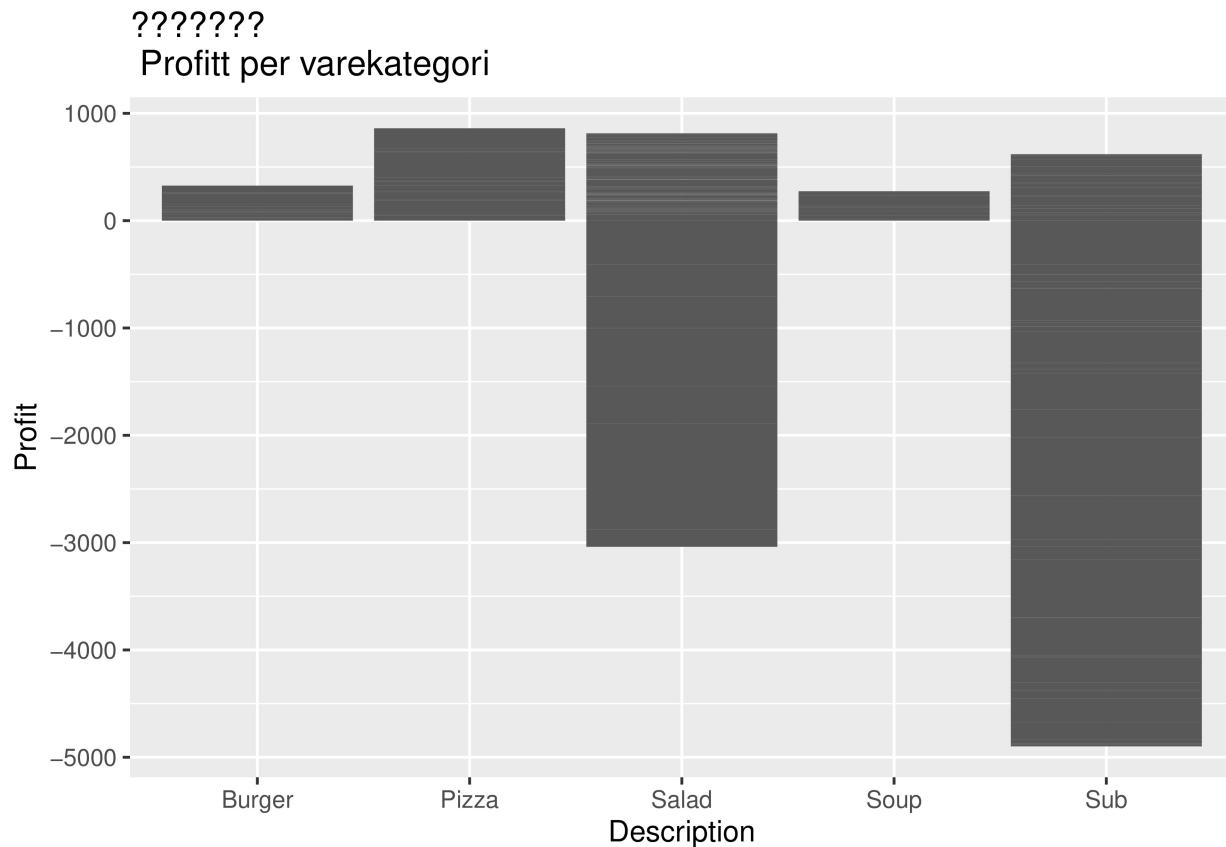
```

#####
df_barplot1 <- bind_rows(pizza, burger, sub, salad, soup)
df_barplot1 <- df_barplot1 %>% filter(Date=="2012-04-01")
aggregate(df_barplot1$Profit, by=list(Category=df_barplot1$Description), FUN=sum)

##   Category      x
## 1 Burger  326.81
## 2 Pizza  861.11
## 3 Salad -2224.37
## 4 Soup   274.10
## 5 Sub   -4278.62

ggplot(df_barplot1, aes(x=Description, y=Profit)) + geom_bar(stat="identity") +
  labs(title = "???????\nProfitt per varekategori")

```



```
# Ser på df5, det originale datasettet:
```

```

df5 %>%
  group_by(Description) %>%
  summarise(Profit = sum(Profit)) %>%
  arrange(Profit)

## # A tibble: 502 x 2
##       Description      Profit

```

```
##      <chr>                  <dbl>
## 1 REGULAR SUB OR SALAD, REWARD -95356.
## 2 VAL MEAL 1 MINI MTBALL CHIP   -38960.
## 3 MINI SUB, REWARDS           -33604.
## 4 FREE ORDER                 -12280.
## 5 BOGO MINI SUB              -10090.
## 6 CHIPS, REWARDS             -7627.
## 7 21 OZ DRINK, REWARDS       -5989.
## 8 1 COOKIE, REWARDS          -5592
## 9 VAL MEAL 3 MINI TURKEY CHIP -5037.
## 10 FREE MINI SUB             -4445.
## # ... with 492 more rows
```

```
sum(df5$Profit)
```

```
## [1] 3367938
```

```
sum(final_df$Profit, na.rm=T)
```

```
## [1] 3367938
```