

Prosjektoppgave

Kandidatnummer 66 og 73

07 June, 2022

- Laster inn datasett og pakker

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.7      v dplyr  1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(lubridate)
```

```
##
```

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

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      date, intersect, setdiff, union
```

```
WEEKLY_WEATHER <- read.csv("WEEKLY_WEATHER.csv")
WEEKLY_SALES_10STORES <- read.csv("WEEKLY_SALES_10STORES.csv")
county_crime <- read.csv("county_crime.csv")
county_demographic <- read.csv("county_demographic.csv")
county_employment <- read.csv("county_employment.csv")
AppWichStoreAttributes<- read.csv("AppWichStoreAttributes.csv")
```

Oppgave 1

- I Oppgave 1 har vi startet med å se på hvilke kolonner som samsvarer med de forskjellige datasettene til oppgaven. Enkelte kolonner har samme data, men forskjellige navn på kolonnen. Dette endrer vi slik at de kan slås sammen til ett større datasett.
- Dette gjør vi med **Store_Num**, **Date**, **Week** og **Store_County** sånn at de passer med de andre datasettene. Videre endrer vi på **Annual_Rent_Estimate** der den ene variabelen har ett komma som gjør at kolonnen blir til character. For å endre på dette bruker vi *gsub* som erstatter det vi ber den om for hele kolonnen. Her erstatter vi da kommaet med ingenting og gjør den om til numeric samtidig. **Weather_Date** endrer vi som sagt til **Date** slik at den samsvarer med **Date** i **WEEKLY_SALES_10STORES**. Parallelt gjør vi om begge fra *character* til *Date*. Dette gir oss mulighet til å kunne slå sammen alle datasettene til ett stort datasett.
- Når vi skal begynne å slå sammen datasettene finnes det en del metoder å gjøre dette på. *merge* og *left_join* er to metoder som vi mente fungerte best, men valgte å benytte bare *left_join*. Dette bestemte vi ettersom at *dplyr::join* funksjonene er raskere enn *merge* og informerer om hva de slår sammen med, samtidig var det en kode som var allsidig å jobbe med. Vi kjører koden og setter datasettene sammen ett datasett ved hjelp av **County_Name**. Ett datasett ved hjelp av **Weather_Station** og **Date**. Tilslutt slår vi sammen disse to datasetten med **County_Name** Dette gir et stort datasett på 79459 *observasjoner* som tilsvarer det største datasettet **WEEKLY_SALES_10STORES** og vi får 72 *variabler*. Dette datasettet har vi gitt navnet **Data**.

```
AppWichStoreAttributes <- AppWichStoreAttributes %>%
  rename(Store_num = Store_Num,
         Weather_Station = Store_Weather_Station) %>%
  mutate(Annual_Rent_Estimate = as.numeric(gsub(",", "", Annual_Rent_Estimate)))

WEEKLY_WEATHER <- WEEKLY_WEATHER %>%
  mutate(Weather_Date=dmy(Weather_Date)) %>%
  rename(Week = Weather_Week,
         Date = Weather_Date)

WEEKLY_SALES_10STORES <- WEEKLY_SALES_10STORES %>%
  mutate(Date=mdy(Date))

DF1 <- left_join(county_crime, county_demographic, by = 'County_Name') %>%
  left_join(county_employment)

## Joining, by = "County_Name"

DF2 <- left_join(AppWichStoreAttributes, WEEKLY_SALES_10STORES, by = 'Store_num') %>%
  rename(County_Name = Store_County) %>%
  left_join(WEEKLY_WEATHER)

## Joining, by = c("Weather_Station", "Date")

Data <- left_join(DF2, DF1)

## Joining, by = "County_Name"
```

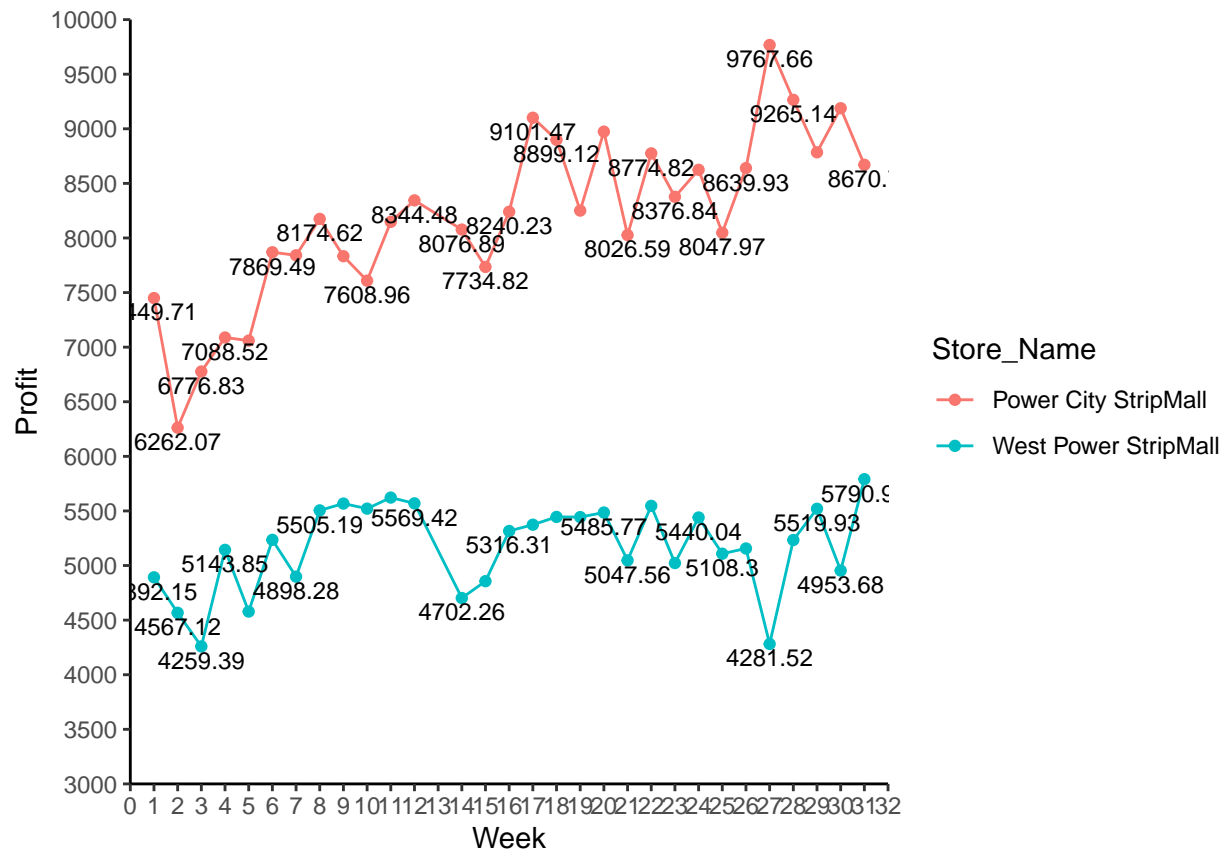
Oppgave 2

- I oppgave to skal vi benytte datasettet til en ukentlig salgsrapport. En ukentlig salgsrapport skal samle informasjon om salg- og inntektsfremgang for bedriften. Vi tar datasettet og lager to nye variabler. En for ukentlig og en for månedlig estimert leiekostnad for å ha til videre analyse. Vi vil ta for oss ukentlig omsetning, kostnader og profit for utsalgsstedet.

```
Data = Data %>%
  mutate(Annual_Rent_Estimate = as.numeric(Annual_Rent_Estimate),
         "Weekly rent" = Annual_Rent_Estimate/52) %>%
  mutate(Annual_Rent_Estimate = as.numeric(Annual_Rent_Estimate),
         "Monthly rent" = Annual_Rent_Estimate/12)
```

```
Data %>%
  filter(Store_num == 7 | Store_num == 5, Week < 32) %>%
  group_by(Week, Store_Name) %>%
  summarise("Profit" = sum(Profit)) %>%
  select(Week, "Profit", Store_Name) %>%
  ggplot(., aes(y = Profit, x = Week, group = Store_Name, colour = Store_Name)) +
  geom_line() +
  geom_point() +
  geom_text(aes(label=(Profit)),size=3,vjust=1.3,color="black",
            check_overlap = TRUE) +
  scale_y_continuous(limits = c(3000,10000),breaks=scales::breaks_pretty(n=20),
                    expand = expansion(add = 0)) +
  scale_x_continuous(limits = c(0,32),breaks=scales::breaks_pretty(n=30),
                    expand = expansion(add = 0)) +
  theme_classic()
```

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



- Vi har laget en funksjon slik at en kan velge hvilket utsalgssted de vil analysere og hvilken uke de vil ha. Da får vi nye datasett for hver uke som vi vil analysere.

```
Utsalgssted <- function(data, x, y){
  {{data}} %>%
  filter(Store_Name == {{x}}, Week == {{y}})
}
```

```
PCS_17 <- Utsalgssted(Data, "Power City StripMall", 30)
```

```
PCS_18 <- Utsalgssted(Data, "Power City StripMall", 31)
```

- Utsalgstedet vi har valgt å analysere er Power City StripMall. Vi skal se på uke 31 og sammenligne med uken før, uke 30.
- Vi lager en funksjon som summerer alle variabler sammen og får totalte verdier for gitt uke. Dette gjør vi for å ha en generell funksjon som kan brukes til andre analyser.

```
TotalSUM <- function(a){
  return(x <- (a) %>%
    summarise('Total units sold' = sum(Sold), 'Total revenue' = sum(Sales),
      'Total expenses' = sum(Cost), 'Total Cost %' = sum(Cost_Percent),
      'Total profit' = sum(Profit), 'Total Margin' = sum(Margin),
      'Weekly average rent' = `Weekly rent`[1],
      'Week' = round(`Week`[1])))
}
```

- Og vi får nye datasett som inneholder summen av beløpene vi vil analysere

```
SUM17=TotalSUM(PCS_17)
SUM18=TotalSUM(PCS_18)

SUMS = rbind(SUM17,SUM18)
knitr::kable((SUMS), caption = "Totale resultater på ukebasis",
              align = "cccccccc" )
```

Table 1: Totale resultater på ukebasis

Total units sold	Total revenue	Total expenses	Total Cost %	Total profit	Total Margin	Weekly average rent	Week
4571	13309.47	4120.88	35.06256	9188.59	0.9996	1038.462	30
4257	12438.63	3767.93	33.64654	8670.70	0.9997	1038.462	31

- Ut fra tabell 1 kan vi se på at det har vært en endring på -314 enheter solgt på en uke som er en prosentvis endring på -6.8694%.
- Totale inntekter har hatt en endring på -870.84 dollar som utgjør en prosentvis endring på -6.543%.
- Totale kostnader ekskludert leiekostnader hadde en endring på -352.95 dollar som utgjør en prosentvis endring på -8.5649%.
- Total kostnad i prosent utgjør en marginal endring på -1.416029
- Total profit hadde en enring på -517.89 dollar som utgjør en prosentiv endring på -5.6362%.

```
TotalUKE <- function(data, b){
  {{data}} %>%
    ggplot(aes(Week, {{b}}, fill=Week))+
    geom_col(show.legend = FALSE, color="black")+
    geom_text(aes(label=scales::comma({{b}})),
              size=3, hjust=1.1, color="white")+
    coord_flip()+
    hrbrthemes::scale_x_comma(breaks = c(30,31))+
    theme_classic()
}
```

```
UkeSalg<-TotalUKE(SUMS,`Total units sold`) %>%
  print()
```

```
UkeSalg<-TotalUKE(SUMS,`Total profit`) %>%
  print()
```

```
UkeSalg<-TotalUKE(SUMS,`Total revenue`) %>%
  print()
```

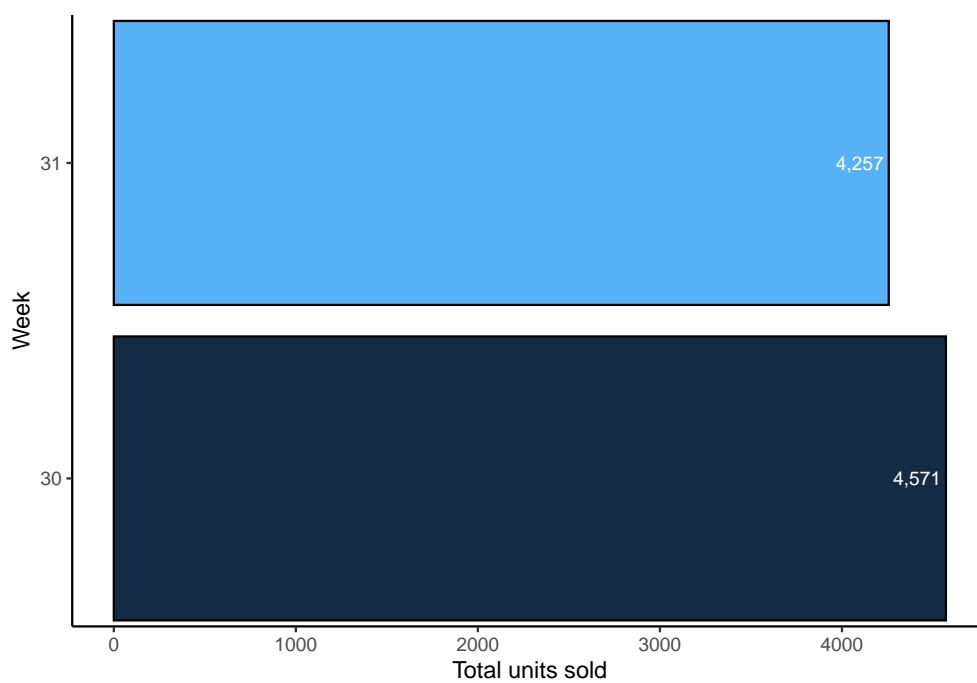


Figure 1: Sammenligning mellom totale enheter solgt per uke

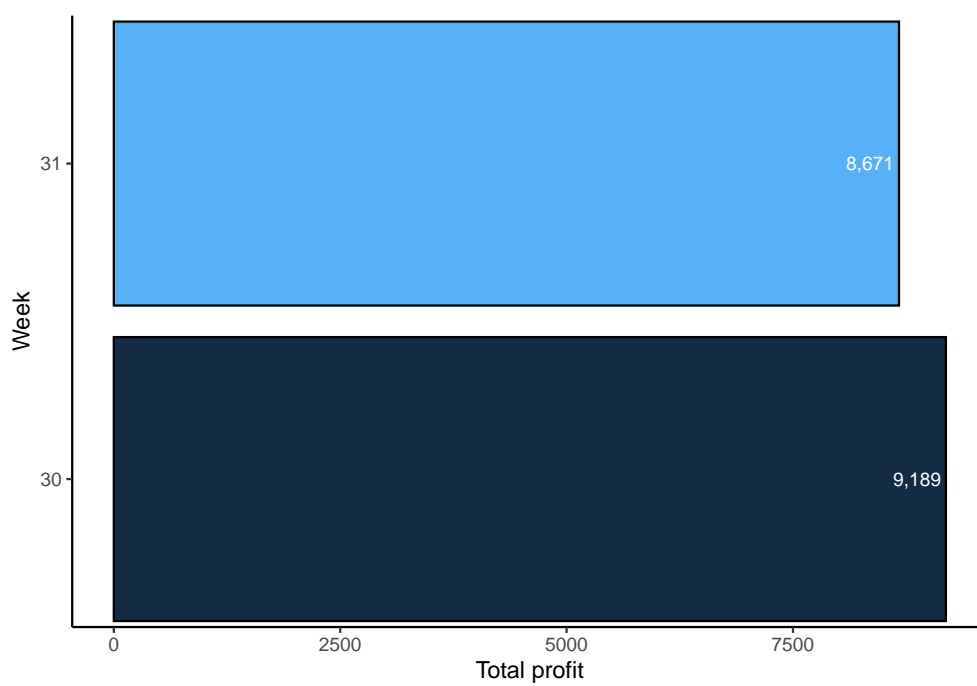


Figure 2: Sammenligning mellom total profit per uke

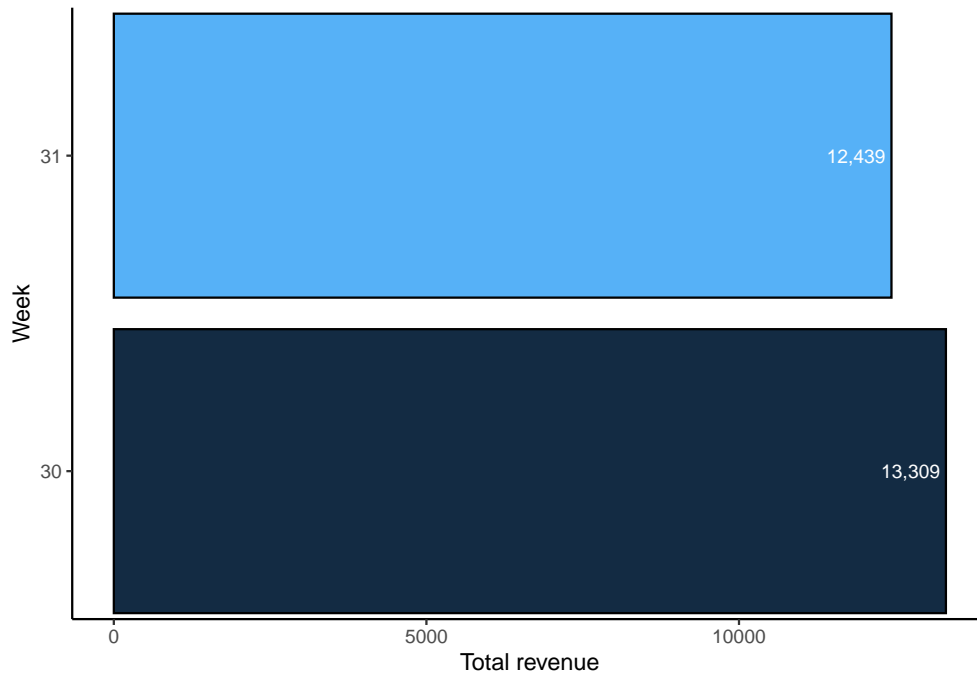


Figure 3: Sammenligning mellom total inntekter per uke

Oppgave 3

```
Månedssdata <- function(data,a){
  {{data}} %>%
  filter(Month == {{a}}) %>%
  group_by(Store_Name, Store_Location, County_Name) %>%
  summarise("Monthly revenue" = sum(Sales), "Monthly profit" = sum(Profit),
            "Monthly expenses" = sum(Cost),
            "Monthly rent" = median(`Monthly rent`)) %>%
  select("Store Name" = Store_Name, "Store Location" = Store_Location,
         "County" = County_Name,
         "Monthly revenue", "Monthly expenses", "Monthly profit", "Monthly rent") %>%
  as.data.frame() %>%
  mutate("Monthly result after rent" = `Monthly profit` - `Monthly rent`,
         'Percent profit of franchise profit' =
           scales::percent((`Monthly profit`/sum(`Monthly profit`))))
}
```

Table 2: Månedlige resultater i Juli per butikk

Store Name	Store Location	County	Monthly revenue	Monthly expenses	Monthly profit	Monthly rent	Monthly result after rent	Percent profit of franchise profit
North Town BigBox	Big Box	Farm County	43750.93	13360.82	30390.11	1725.000	28665.11	8.467%
Rail City BigBox	Big Box	Rail County	35742.63	10993.31	24749.32	1458.333	23290.99	6.895%
University Town BigBox	Big Box	River County	30613.16	9239.37	21373.79	1250.000	20123.79	5.955%
Power City FreeStand	Free Standing	Power County	87798.45	26805.79	60992.66	5250.000	55742.66	16.993%
Lake City StripMall	Strip Mall	Lake County	56690.83	17371.15	39319.68	3450.000	35869.68	10.955%
Littletown StripMall	Strip Mall	Appleton County	55113.25	16773.83	38339.42	3000.000	35339.42	10.682%
North Town StripMall	Strip Mall	Farm County	67797.28	21393.62	46403.66	3150.000	43253.66	12.928%
Power City StripMall	Strip Mall	Power County	65937.23	20260.39	45676.84	4500.000	41176.84	12.726%
River City StripMall	Strip Mall	River County	38301.23	12394.58	25906.65	3750.000	22156.65	7.218%
West Power StripMall	Strip Mall	Power County	37163.66	11384.04	25779.62	4050.000	21729.62	7.182%

```
MånedligData <- Månedldata(Data,7) %>%
  arrange(.$`Store Location`)
```

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

```
knitr::kable(MånedligData,caption = "Månedlige resultater i Juli per butikk",
  align = "lcccccccccccc") %>%
  kableExtra::kable_styling(font_size = 10, latex_options = "scale_down")
```

- I tabell 2 har vi sortert ut resultatene for Juli måned for hver enkelt butikk i franchisen. Her ser vi at det er Power city FreeStand som har størst månedlig inntekt, kostnad og profitt. Allikevel er det Littletown StripMall som har høyest samlet profittmargin av butikkene i franchisen. Dette skyldes av at kostnadene er lavere i forhold til salgs inntektene til denne butikken og de genererer mer profitt i forhold til de andre butikkene med tanke på forholdet mellom salg og kostnad.
- For Franchisen sin del er det interessant å se på at Big Box butikkene genererer vesentlig mindre profitt i prosent av franchisens totale profitt for juli måned. Sammenligner vi River City StripMall og University Town BigBox ligger disse i samme County, samtidig som University Town BigBox bare har 3 andre konkurrenter nær seg i forhold til River City StripMall som har 38. Allikevel har den nesten 1,5% mindre i andel profitt til franchisen. Det vi ser er at alle Big Box butikkene ligger ikke nær en skole, som gjør at de mister kunder som familier med skolebarn, og at det kan ligge litt i utkanten, som gjør det vanskeligere å besøke. Alle Big Box butikkene har under 9% som en franchiseleder burde ta hensyn til, om det finnes bedre muligheter.

```
MånedligStores <- function(data, x, y){
  {{data}} %>%
  mutate(Store=fct_reorder({{x}},{{y}})) %>%
  ggplot(aes(Store,{{y}},fill={{x}}))+
  geom_col(show.legend = FALSE,color="black")+
  geom_text(aes(label=scales::comma({{y}})),
    size=3,hjust=1.1,color="black")+
  hrbrthemes::scale_y_comma()+
  scale_fill_brewer(palette = "Paired")+
  coord_flip()+
  theme_classic()
}
```

```
Sales= MånedligStores(MånedligData, `Store Name`, `Monthly revenue`) %>%
  print()
```

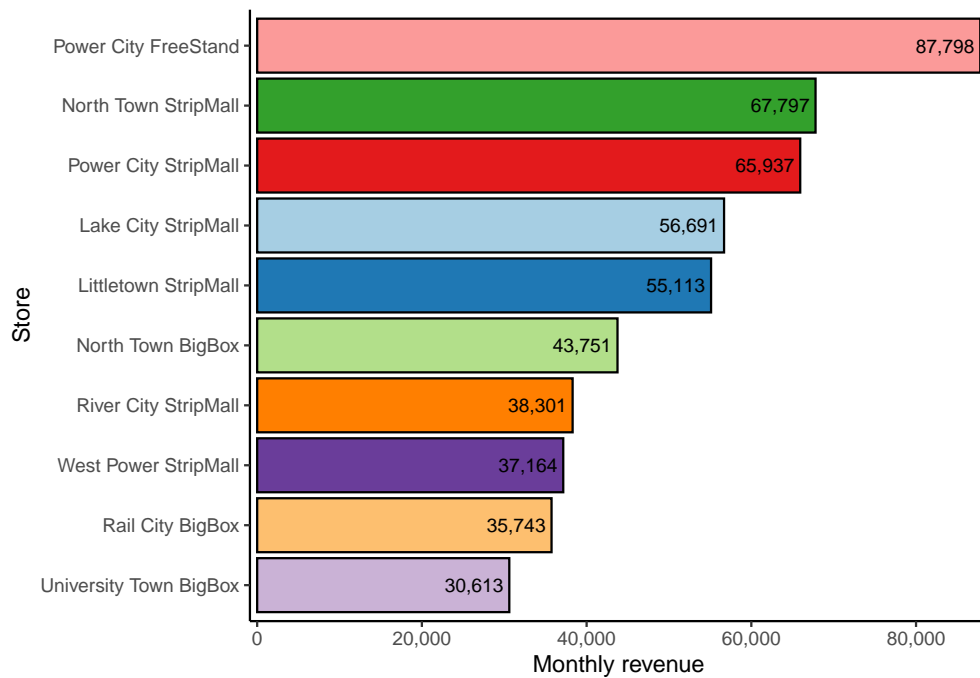



Figure 4: Totale inntekter i Juli måned for hver butikk

```
Profit= MånedligStores(MånedligData, `Store Name`, `Monthly expenses`) %>%
  print()
```

```
Profit= MånedligStores(MånedligData, `Store Name`, `Monthly profit`) %>%
  print()
```

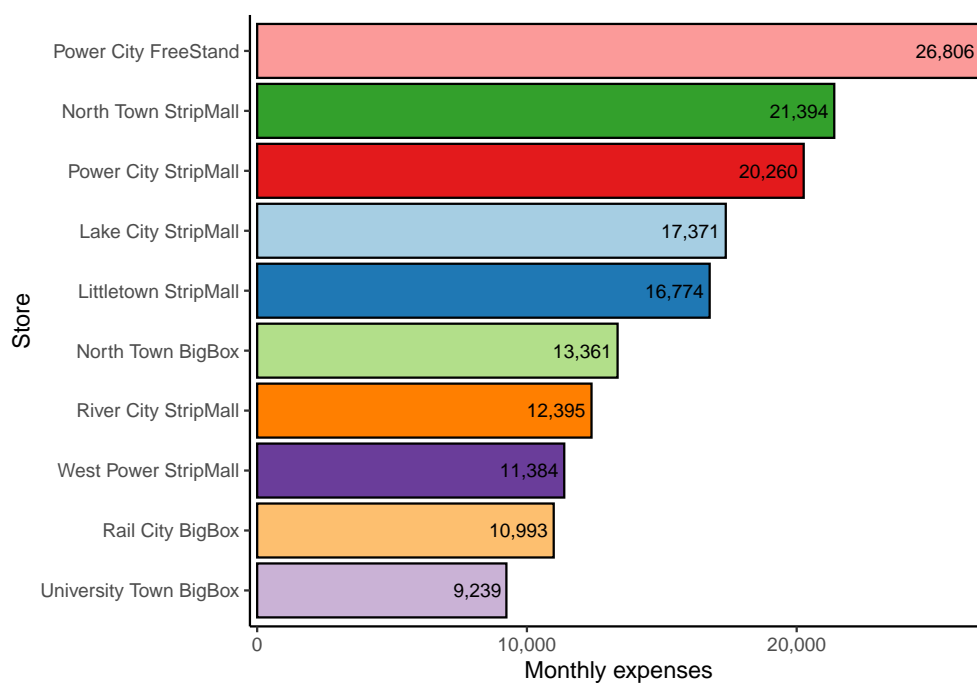


Figure 5: Total kostnad i Juli måned for hver butikk

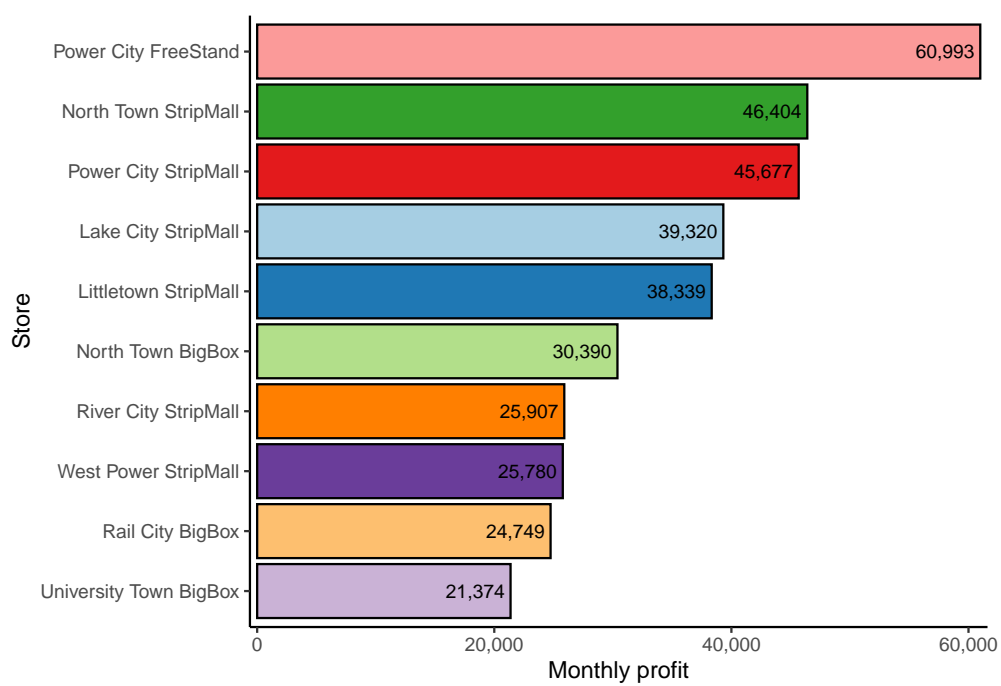


Figure 6: Total profit i Juli måned for hver butikk

```
MånedligData%>%
  mutate(Store=fct_reorder(`Store Location`, `Monthly result after rent`)) %>%
  ggplot(aes(Store, `Monthly result after rent`, fill=`Store Name`))+
  geom_col(show.legend = TRUE ,color="black") +
  geom_text(aes(label=scales::comma(`Monthly result after rent`)),size=2.5,position = position_stack(vjust="top"))+
  hrbthemes::scale_y_comma()+
  scale_fill_brewer(palette = "Paired")+
  coord_flip()+
  theme_classic()
```

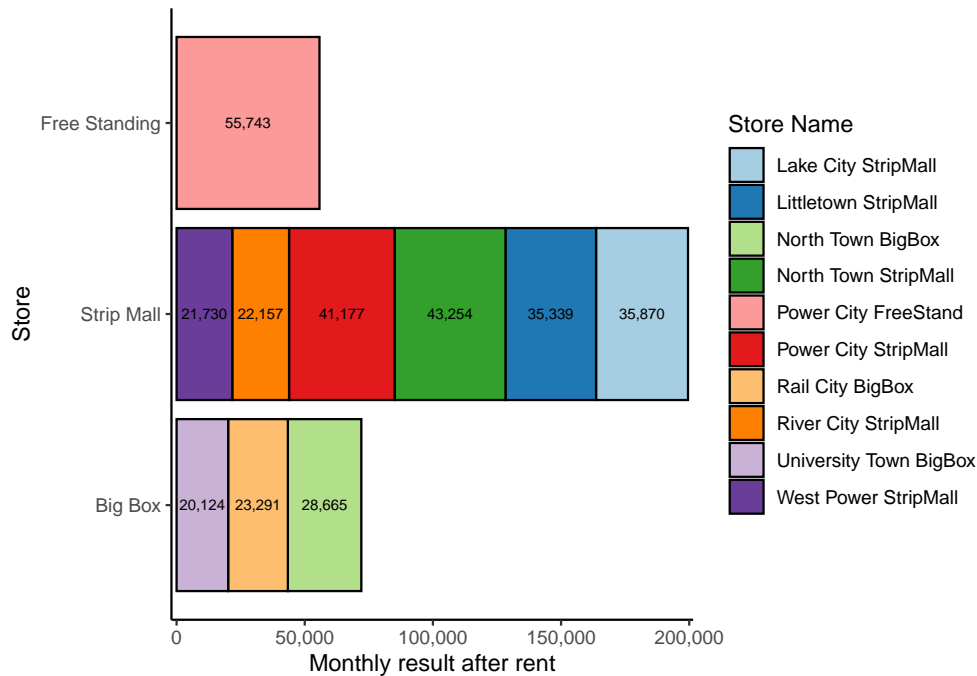


Figure 7: Månedlig salg per type butikk i Juli måned

- Her kan vi se hvor mye de forskjellige butikkene har i profitt etter leiekostnad for hvilken type butikk de er. Som nevnt tidligere er Big Box en type butikk som gir minst profitt generelt sett, men vet er to StripMall's som har på lik linje profitt som Big Box etter leiekostnad. Dette er da henholdsvis West Power StripMall som ikke ligger nær noe skole og går gantagelig glipp av mange kunder på dette aspektet, også er det River City StripMall som har veldig få innbyggere og masse konkurranse rundt seg. FreeStand er den som tjener mest, men etter leiekostnader synker profitten masse ettersom leien der er høy.

Oppgave 4

```
ÅR <- function(data){
  {{data}} %>%
  group_by(Store_Name, Store_Location, County_Name, Annual_Rent_Estimate,
           County_Total_Crime_Rate, Store_Near_School, County_Total_Census_Pop,
           Store_Competition_Otherfood, Store_Competition_Fastfood,
           Store_Drive_Through, County_Unemployment_Rate)%>%
  summarise("Revenue" = sum(Sales), "Profit" = sum(Profit),
            "Expenses" = sum(Cost))%>%
  select("Store Name" = Store_Name, "Store Location"= Store_Location,
         "County" = County_Name,
         "Revenue", "Expenses", "Profit",
         "Annual Rent" = Annual_Rent_Estimate,
         "County Population" = County_Total_Census_Pop,
         "Unemployment Rate" = County_Unemployment_Rate,
         "Drive Through?" = Store_Drive_Through,
         "Near School?" = Store_Near_School,
         "Total Crime Rate"= County_Total_Crime_Rate,
         Store_Competition_Otherfood, Store_Competition_Fastfood
        ) %>%
  as.data.frame() %>%
  mutate("Total competitors" = Store_Competition_Otherfood +
         Store_Competition_Fastfood,
         "Result after rent" = Profit - `Annual Rent`,
         "Dollar profit per 1000 Citizen" = `Result after rent` / (`County Population`/1000),
         'Percent profit of franchise profit' =
           scales::percent((`Profit`/sum(`Profit`)))) %>%
  select(-c(Store_Competition_Otherfood, Store_Competition_Fastfood))
}
```

```
ÅrData <- ÅR(Data) %>%
  arrange(.$`Dollar profit per 1000 Citizen`)
```

```
## 'summarise()' has grouped output by 'Store_Name', 'Store_Location',
## 'County_Name', 'Annual_Rent_Estimate', 'County_Total_Crime_Rate',
## 'Store_Near_School', 'County_Total_Census_Pop', 'Store_Competition_Otherfood',
## 'Store_Competition_Fastfood', 'Store_Drive_Through'. You can override using the
## '.groups' argument.
```

```
Bad_Weather <- left_join(AppWichStoreAttributes, WEEKLY_WEATHER, by = 'Weather_Station') %>%
  rename(County_Name = Store_County) %>%
  select("Store Name" = Store_Name, "Bad Weather Days" = Weather_Bad_Weather_days) %>%
  group_by(`Store Name`) %>%
  summarise("Bad Weather Days" = sum(`Bad Weather Days`))
```

```
ÅrData <- left_join(ÅrData, Bad_Weather)
```

```
## Joining, by = "Store Name"
```

Table 3: Årsresultat og diverse statistikk

Store Name	Store Location	County	Revenue	Expenses	Profit	Annual Rent	County Population	Unemployment Rate	Drive Through?	Near School?	Total Crime Rate	Total competition	Result after rent	Dollar profit per 1000 Citizen	Percent profit of franchise profit	Bad Weather Days
West Town StripMall	Strip Mall	Power County	365501.8	1164281.07	226981.7	36000	182398	9.3	No	No	68.8	10	20181.7	1098.144	7.80%	115
Littletown StripMall	Strip Mall	Ashtabula County	522831.1	190359.71	302471.4	36000	249777	10.6	No	No	108.8	25	326474.4	1321.882	10.70%	106
Power City StripMall	Strip Mall	Power County	581628.1	186421.35	395206.2	54000	182398	9.3	No	No	68.8	11	341286.2	1870.668	11.73%	115
Rail City BigBox	Big Box	Rail County	552132.2	130225.99	241906.4	17500	85835	9.7	No	No	63.2	30	234485.4	2614.321	7.38%	106
Power City FreeStand	Free Standing	Power County	870859.6	276774.03	594085.5	63000	182398	9.3	Yes	Yes	68.8	35	531085.5	2911.685	17.64%	115
River City StripMall	Strip Mall	River County	336361.2	107229.34	229131.9	45000	59404	7.5	No	Yes	49.5	38	184121.9	3099.655	6.80%	146
University Town BigBox	Big Box	River County	302343.7	52223.74	250119.0	5000	59404	7.5	No	No	43.5	3	155273.0	8285.255	6.44%	146
North Town BigBox	Big Box	Farm County	414215.0	131007.88	283207.1	20700	76580	9.4	No	No	N/A	24	262967.1	3417.952	8.41%	122
Lake City StripMall	Strip Mall	Lake County	543373.6	109291.94	374081.6	11400	91123	9.5	Yes	Yes	86.3	11	375981.6	2627.025	11.11%	118
North Town StripMall	Strip Mall	Farm County	628040.1	159575.96	428464.1	27800	76580	9.4	No	Yes	N/A	22	390664.1	6988.448	12.72%	122

```
knitr::kable((ÅrData),caption = "Årsresultat og diverse statistikk",
              align = "lcccccccccccc") %>%
  kableExtra::kable_styling(font_size = 10, latex_options = "scale_down")
```