

# Prosjektoppgave 1005 - Vår 2022

Kandidatnr.: 61

Begynner med å laste inn pakkene jeg vil bruke.

```
library(dplyr)
library(tidyverse)
library(lubridate)
library(data.table)
library(microbenchmark)
library(purrr)
library(fs)
```

## OPPGAVE 1

I denne oppgaven skal jeg slå sammen 6 datasett til et stort et.

Laster først inn oppgavens datasett.

```
store_attributes <- read_csv('https://data.mendeley.com/public-files/datasets/6htjnfs78b/files/26afd5e7-
county_crime <- read_csv('https://data.mendeley.com/public-files/datasets/6htjnfs78b/files/3691994e-287
county_demographic <- read_csv('https://data.mendeley.com/public-files/datasets/6htjnfs78b/files/527e74
county_employment <- read.csv('https://data.mendeley.com/public-files/datasets/6htjnfs78b/files/846ac75
weekly_sales <- read.csv('https://data.mendeley.com/public-files/datasets/6htjnfs78b/files/b963fdd1-0df
weekly_weather <- read.csv('https://data.mendeley.com/public-files/datasets/6htjnfs78b/files/b8708b3c-c
```

Bruker mutate til å lage dato variabel i weekly\_weather og weekly\_sales.

```
weekly_weather <- weekly_weather %>%
  mutate(Weather_Date = as.Date(Weather_Date, "%d/%m/%Y"))

weekly_sales <- weekly_sales %>%
  mutate(Date = as.Date(Date, "%m/%d/%Y"))
```

Endrer her noen navn på variabler for å skape felles variabelnavn på tvers av datasettene. Dette gjør jeg for å senere kunne slå sammen datasettene til ett. Jeg lager også variablene `monthly_rent_estimate` og `weekly_rent_estimate` i `store_attributes` med hjelp av `mutate`.

```
store_attributes <- store_attributes %>%
  rename(Store_num = Store_Num,
         Weather_Station = Store_Weather_Station,
         County_Name = Store_County) %>%
  mutate(monthly_rent_estimate = Annual_Rent_Estimate / 12,
         weekly_rent_estimate = Annual_Rent_Estimate / 52)

weekly_weather <- weekly_weather %>%
  rename(Date = Weather_Date)

county_employment <- county_employment %>%
  rename(County_Name = i..County_Name)
```

Slår sammen de 6 datasettene en etter en med hjelp av felles variabelnavn. Jeg bruker `merge` og `left_join` til dette formålet.

```
CrimeDemo <- merge(county_crime, county_demographic, by = 'County_Name')

CrimeDemoEmp <- merge(CrimeDemo, county_employment, by = 'County_Name')

CrimeDemoEmpAtt <- merge(CrimeDemoEmp, store_attributes, by = 'County_Name')

CrimeDemoEmpAttSales <- merge(CrimeDemoEmpAtt, weekly_sales, by = 'Store_num')

alldata <- CrimeDemoEmpAttSales %>%
  left_join(weekly_weather, by = c('Date' = 'Date', 'Weather_Station' = 'Weather_Station'))
```

## OPPGAVE 2

Jeg skal nå lage en ukentlig salgsrapport for en av utsalgene i datasettet. Jeg velger å gå for Littleton StripMall (storenumber 9) i Appleton County.

For en ukentlig salgsrapport tenker jeg at det er lurt å se på hvilke produkter som selger mest (bestselgerne). Dette fordi at dette er varene som trekker kunder og kan bidra til høy omsetning på grunn av sin drakraft.

Det er også relevant å se på hvilke produkter som står for høyest profitt fordi dette er de produktene som drar lasset rent tallmessig når det kommer til omsetning.

Jeg tenker også at det er fornuftig å sammenligne noen tall med de foregående ukene.

Jeg begynner med å lage nye datasett, “ukentlig\_salgsrapport”, hvor jeg velger ut relevant data fra ‘alldata’ til denne oppgaven. Jeg bruker i hovedsak filter og select til dette formålet.

```
Ukentlig_Salgsrapport <- alldata %>%
  filter(Store_num == 9, Date == '2013-01-06') %>%
  select(Date ,County_Name, Store_City, Store_Name, Store_Location, Store_num, Annual_Rent_Estimate, IN
    Price, Sold, Del, Sales, Tot_Sls, Unit_Cost, Cost, Cost_Percent, Margin, Profit)

Ukentlig_Salgsrapport_2 <- alldata %>%
  filter(Store_num == 9, Date == '2012-12-30') %>%
  select(Date ,County_Name, Store_City, Store_Name, Store_Location, Store_num, Annual_Rent_Estimate, IN
    Price, Sold, Del, Sales, Tot_Sls, Unit_Cost, Cost, Cost_Percent, Margin, Profit)

Ukentlig_Salgsrapport_3 <- rbind(Ukentlig_Salgsrapport_2, Ukentlig_Salgsrapport)

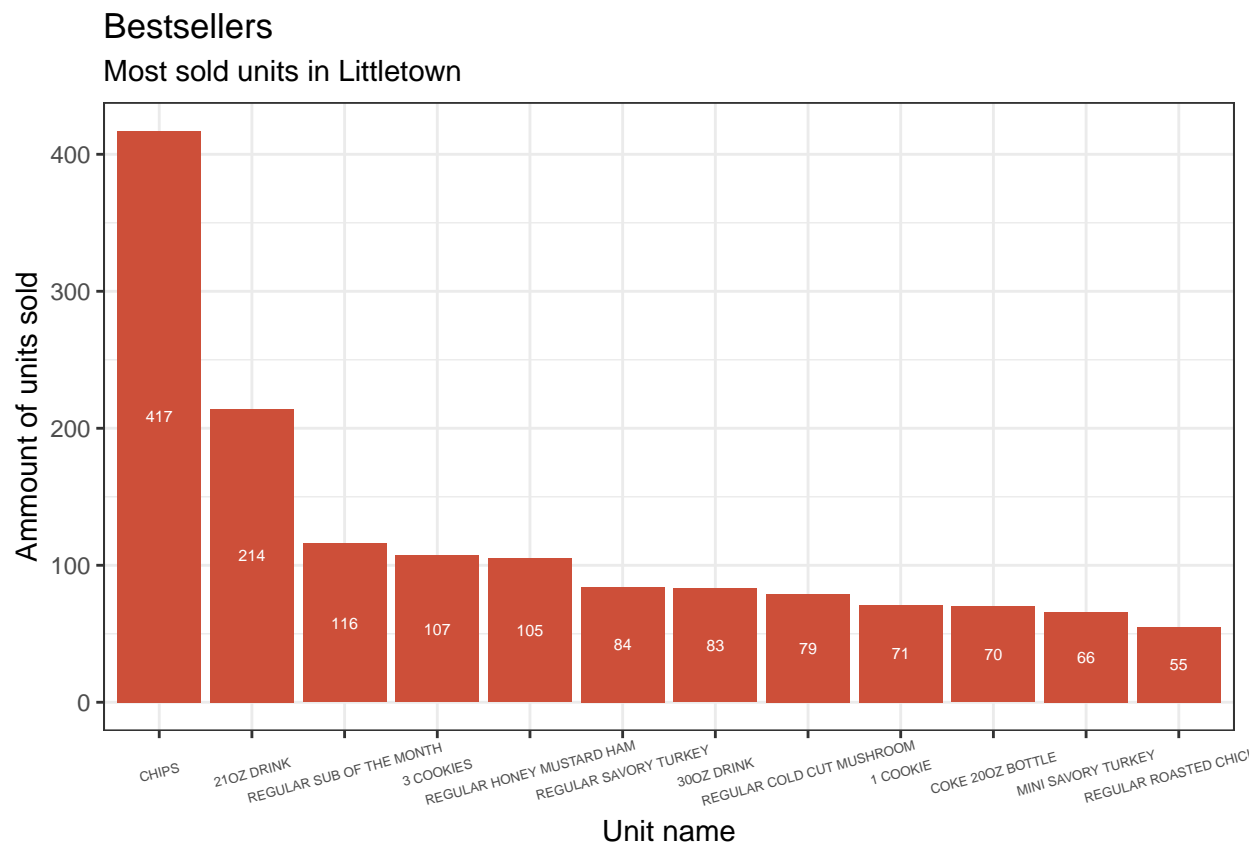
Ukentlig_Salgsrapport_4 <- alldata %>%
  filter(Store_num == 9, Date == '2012-12-23') %>%
  select(Date ,County_Name, Store_City, Store_Name, Store_Location, Store_num, Annual_Rent_Estimate, IN
    Price, Sold, Del, Sales, Tot_Sls, Unit_Cost, Cost, Cost_Percent, Margin, Profit)

Ukentlig_Salgsrapport_5 <- rbind(Ukentlig_Salgsrapport_4, Ukentlig_Salgsrapport_3)
```

## Plot med bestselgerne i Littletown StripMall

Under har jeg laget et bar plot for bestselgerne i Littletown. Jeg filtrer Sold  $\geq 50$  for å få de mest solgte.

```
Ukentlig_Salgsrapport %>%
  filter(Date == '2013-01-06', Sold >= 50) %>%
  ggplot(aes(x= reorder(Description,-Sold), y = Sold)) +
  geom_bar(stat="identity", fill="tomato3") +
  labs(title="Bestsellers",
       subtitle="Most sold units in Littletown")+
  ylab('Ammount of units sold')+
  xlab('Unit name')+
  geom_text(aes(label=Sold), position = position_stack(vjust= 0.5),
           colour = "white", size = 2)+
  theme_bw()+
  theme(axis.text.x = element_text(angle=15, vjust=0.6, size = 5))
```

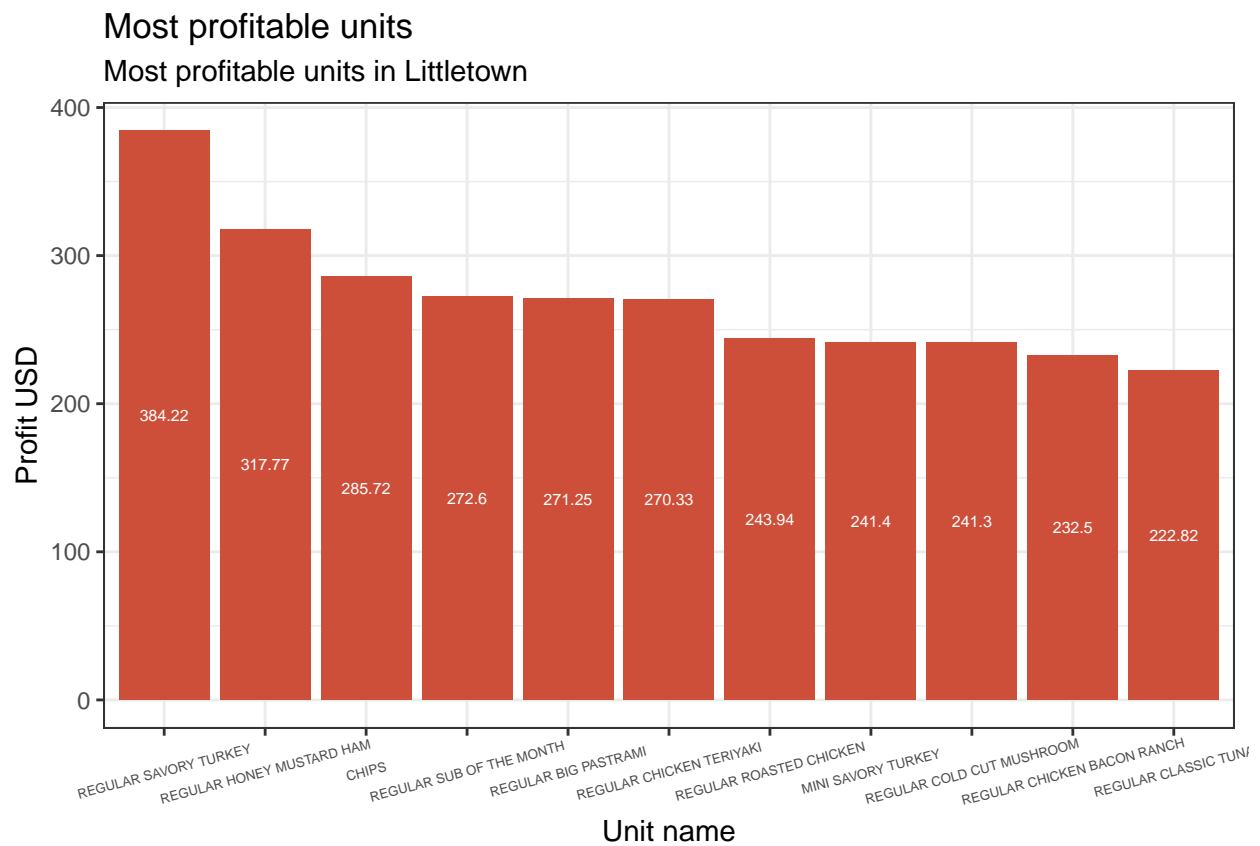


Vi ser at chips er soleklar vinner etterfulgt av drikke. Dette er viktige varer. Kan prisen skyves opp på disse varene vil det kunne utgjøre en stor forskjell på omsetningen.

## Plot for varene med størst profitt i Littletown StripMall

Under har jeg laget et bar plot for varene med høyest profitt i Littletown.

```
Ukentlig_Salgsrapport %>%
  filter(Date == '2013-01-06', Profit >= 200) %>%
  ggplot(aes(x= reorder(Description,-Profit), y = Profit)) +
  geom_bar(stat="identity", fill="tomato3") +
  labs(title="Most profitable units",
       subtitle="Most profitable units in Littletown")+
  ylab('Profit USD')+
  xlab('Unit name')+
  geom_text(aes(label=Profit), position = position_stack(vjust= 0.5),
           colour = "white", size = 2)+
  theme_bw()+
  theme(axis.text.x = element_text(angle=15, vjust=0.6, size = 5))
```

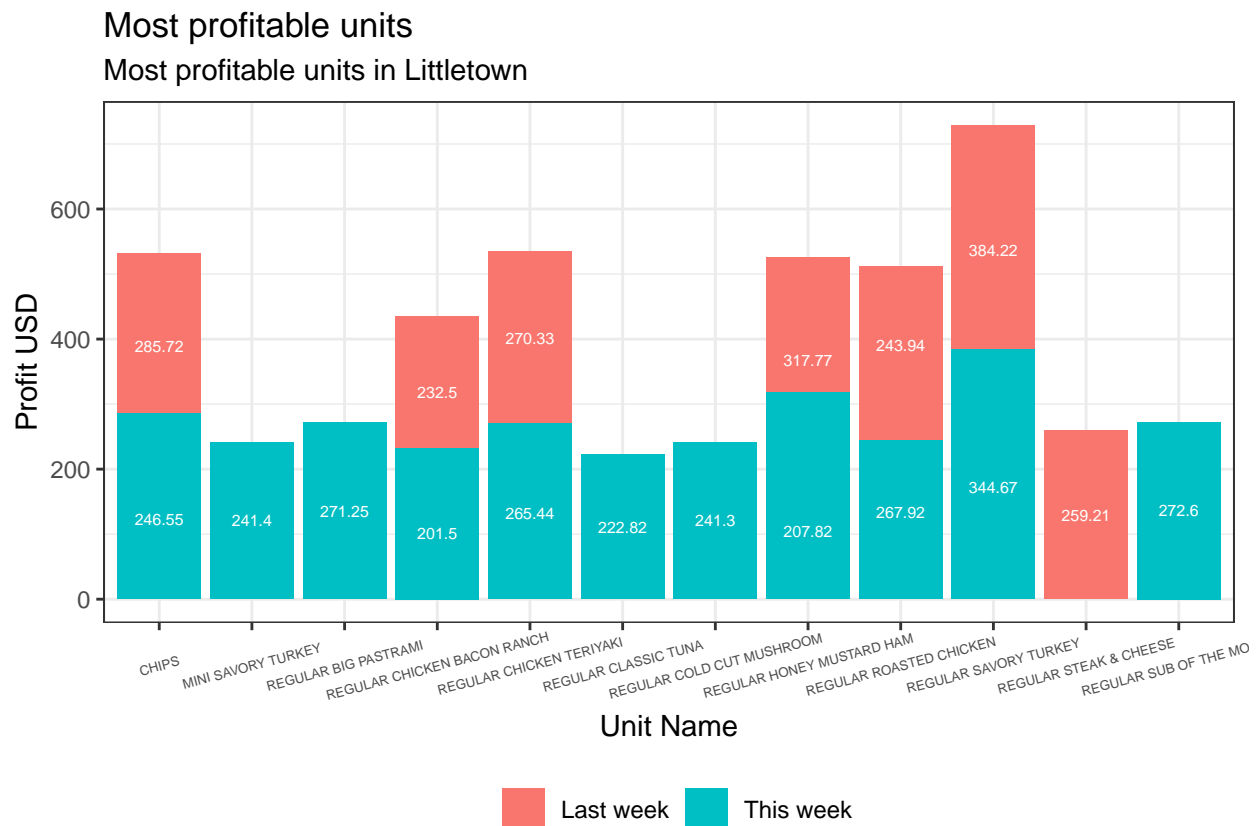


## Samme plot med foregående uke stacked

Her har jeg lagt til forrige ukes varer med størst profitt.

```
labs <- expression('Last week', 'This week')

Ukentlig_Salgsrapport_3 %>%
  filter(Profit >= 200) %>%
  ggplot(aes(x = Description, y = Profit)) +
  geom_bar(stat = 'identity', aes(fill = factor(Date)))+
  labs(title="Most profitable units",
       subtitle="Most profitable units in Littletown")+
  geom_text(aes(label=Profit), position = position_stack(vjust = 0.5),
           colour = "white", size = 2)+
  theme_bw()+
  theme(axis.text.x = element_text(angle=15, vjust=0.6, size = 5))+
  theme(legend.position="bottom")+
  labs(fill = '', x = 'Unit Name', y = 'Profit USD')+
  scale_fill_discrete(labels=labs)
```



Vi ser at nåværende uke har flere varer med profittverdi over 200, enn hva forrige uke hadde.

## Plot som viser profitt for de tre siste ukene

Til slutt lager jeg et plot som sammenligner de siste tre ukenes totale profitt.

```
week_labels <- c("Two weeks ago", 'Last week', 'This week')

Ukentlig_Salgsrapport_5 %>%
  group_by(Date) %>%
  summarise(Total_Profit = sum(Profit)) %>%
  ggplot(aes(x = Date, y = Total_Profit)) +
  geom_bar(stat="identity", fill="cyan3") +
  labs(title="Total profit Littletown",
       subtitle='3 last weeks total profit compared')+
  ylab('Profit USD')+
  geom_text(aes(label=Total_Profit), position = position_stack(vjust= 0.5),
           colour = "white", size = 5)+
  theme_bw()
```



Vi ser at uken vi har fokus på, tick-navn “jan 07”, er uken med best profitt. Det kan være derfor vi så flere varer med profitverdi over 200 på denne uken i forrige graf.

### OPPGAVE 3

I denne oppgaven skal jeg lage en månedlig salgsrapport med alle utsalgene. Jeg har valgt å fokusere på September 2012.

Jeg starter med å lage et nytt datasett hvor jeg henter relevante tall for denne oppgaven fra 'alldata'. Jeg filtrer ut september 2012 og tar med de relevante variablene.

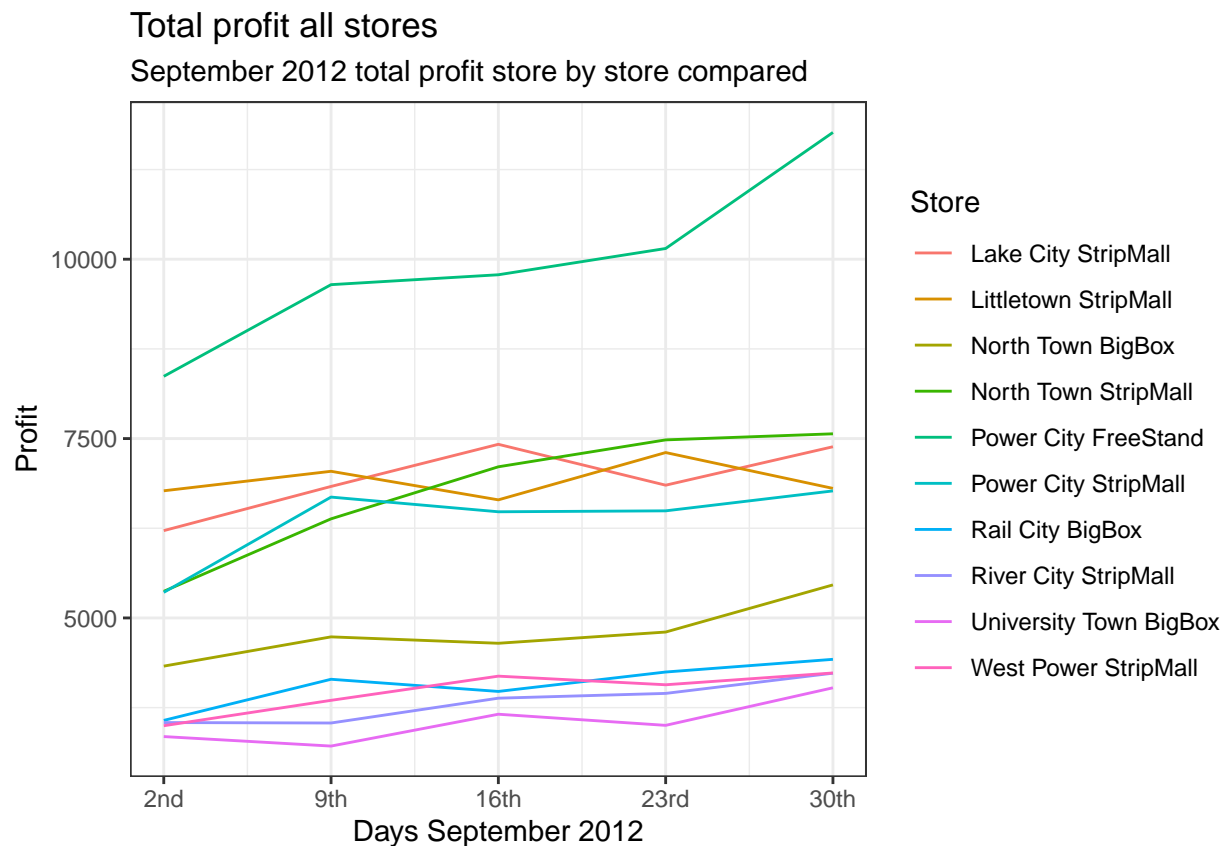
```
maanedlig_salgsrapport <- alldata %>%  
  filter(Year == 2012, Month >= 9 & Month < 10) %>%  
  select(Date, Year, Month, Day, County_Name, Store_City, Store_Name, Store_Location, Store_num, Annual.  
         Price, Sold, Del, Sales, Tot_Sls, Unit_Cost, Cost, Cost_Percent, Margin, Profit)
```



## Linje plot for alle utsalgene September 2012

Under har jeg laget et linje plot som viser profitt for alle utsalgene i løpet av 5 observasjoner i september 2012.

```
maanedlig_salgsrapport %>%
  group_by(Day, Store_Name) %>%
  summarise(Total_profit_daily = sum(Profit)) %>%
  ggplot(aes(x = Day, y = Total_profit_daily, col = Store_Name)) +
  geom_line() +
  scale_x_continuous(breaks = c(2,9,16,23,30),
                    labels = c("2nd", '9th', '16th', '23rd', '30th'))+
  labs(title="Total profit all stores",
       subtitle='September 2012 total profit store by store compared',
       col = 'Store')+
  ylab('Profit')+
  xlab('Days September 2012')+
  theme_bw()
```



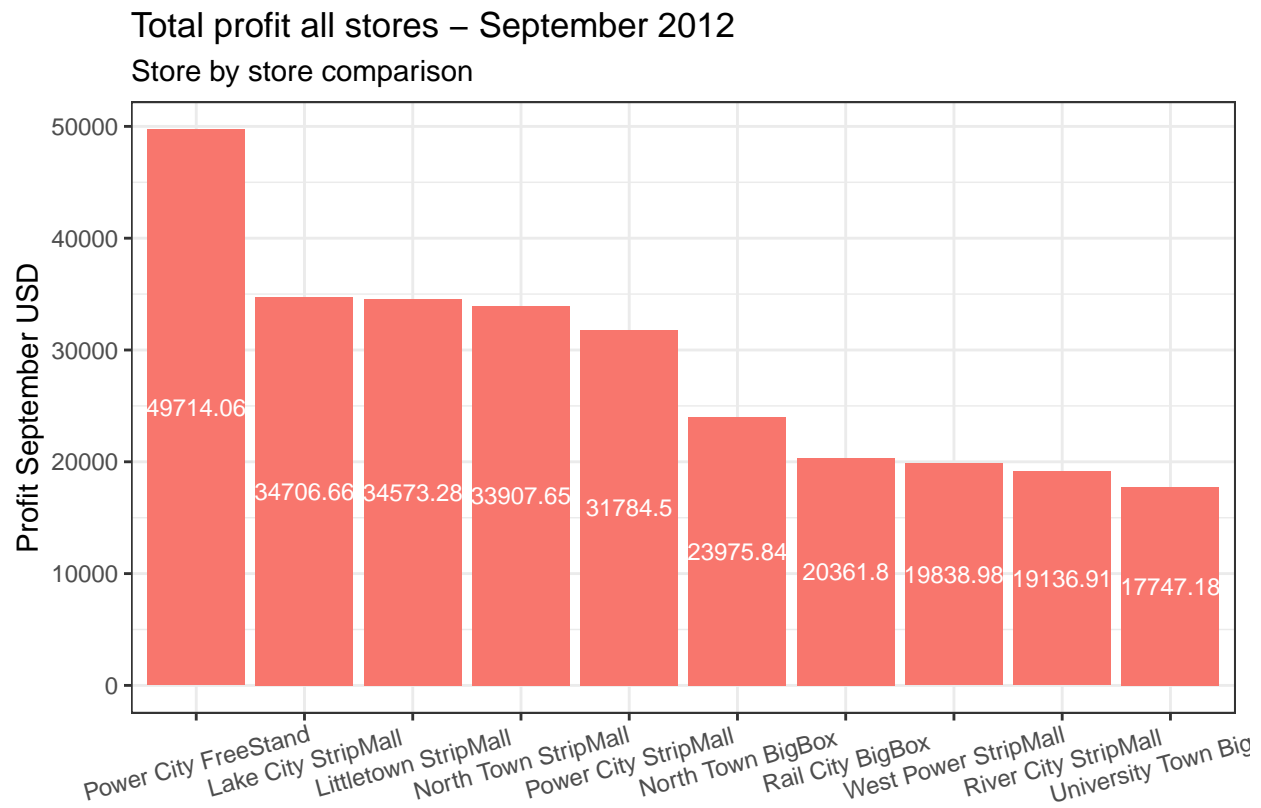
Vi ser at Power City FreeStand er utsalget som gjør det klart best.

## Bar plot for alle utsalgene September 2012

Under har jeg laget et bar plot med samlet profitt for alle utsalgene September 2012.

```
maanedlig_salgsrapport %>%
  group_by(Month, Store_Name) %>%
  summarise(Total_profit_month = sum(Profit)) %>%
  ggplot(aes(x= reorder(Store_Name,-Total_profit_month), y = Total_profit_month)) +
  geom_col(aes(fill = 'red'), show.legend = FALSE) +
  labs(title="Total profit all stores - September 2012",
       subtitle='Store by store comparison',
       col = 'Store')+
  ylab('Profit September USD')+
  xlab('')+
  geom_text(aes(label=Total_profit_month), position = position_stack(vjust= 0.5),
           colour = "white", size = 3)+
  theme_bw()+
  theme(axis.text.x = element_text(angle=15, vjust=0.6))
```

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



Jeg ønsket å finne ut hvilken vare som utgjorde størst profitt for hver enkelt utsalg i september.

```
maanedlig_salgsrapport %>%
  select(Store_Name, Month, Description, Profit) %>%
  group_by(Store_Name) %>%
  filter(Profit == max(Profit))
```

```
## # A tibble: 10 x 4
## # Groups:   Store_Name [10]
##   Store_Name      Month Description      Profit
##   <chr>          <int> <chr>          <dbl>
## 1 Power City FreeStand      9 REGULAR SUB OF THE MONTH 1948.
## 2 West Power StripMall      9 REGULAR SUB OF THE MONTH  927.
## 3 Power City StripMall      9 REGULAR SUB OF THE MONTH 1089.
## 4 Littletown StripMall      9 REGULAR SUB OF THE MONTH 1354.
## 5 River City StripMall      9 REGULAR SUB OF THE MONTH  663.
## 6 Rail City BigBox          9 REGULAR SUB OF THE MONTH  834.
## 7 Lake City StripMall      9 REGULAR SUB OF THE MONTH  952.
## 8 North Town StripMall      9 REGULAR SUB OF THE MONTH 1787.
## 9 North Town BigBox         9 REGULAR SUB OF THE MONTH 1108.
## 10 University Town BigBox    9 REGULAR SUB OF THE MONTH  487.
```

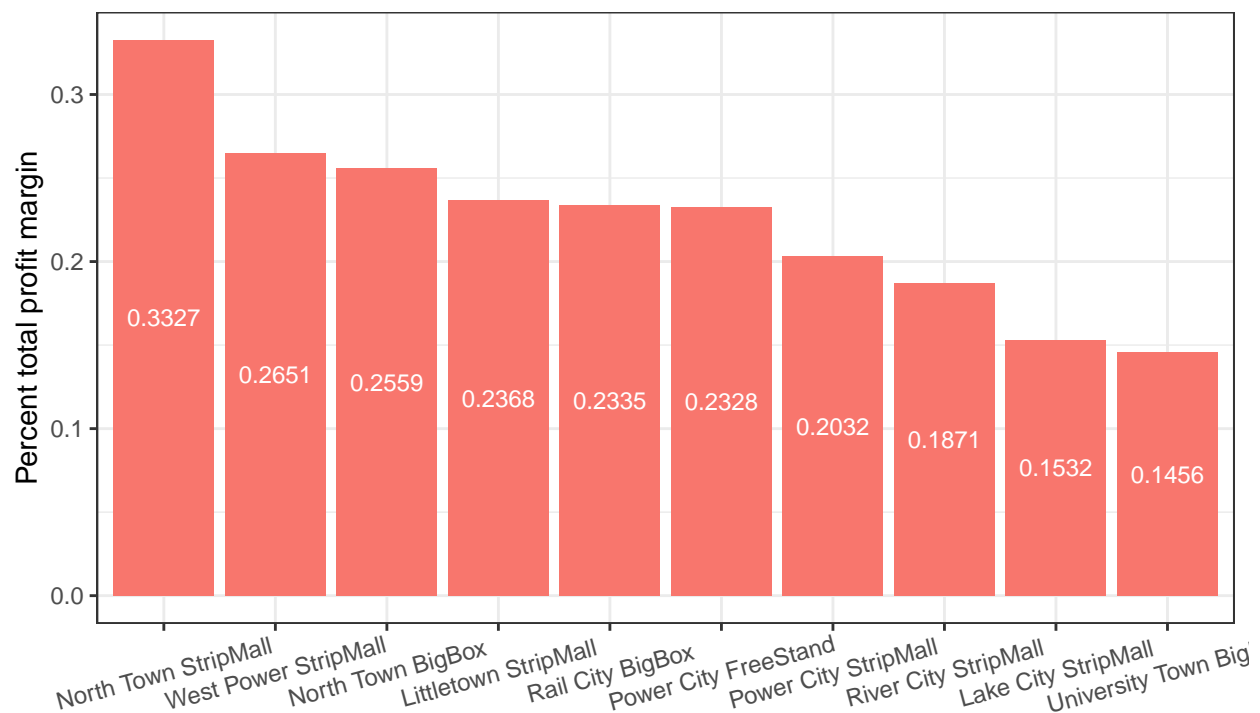
Vi ser en klar vinner her. “REGULAR SUB OF THE MONTH” står for største profittmargin av alle varene for SAMTLIGE av utsalgene. Dette er en veldig viktig vare for bedriften.

Bar plot som viser profittmargin utgjort av “REGULAR SUB OF THE MONTH” for hvert utsalg i September 2012.

```
maanedlig_salgsrapport %>%
  select(Store_Name, Month, Description, Profit, Margin) %>%
  group_by(Store_Name) %>%
  filter(Margin == max(Margin)) %>%
  ggplot(aes(x= reorder(Store_Name,-Margin), y = Margin)) +
  geom_col(aes(fill = 'red'), show.legend = FALSE) +
  labs(title='Profit margin for the most profitable unit in September 2012',
       subtitle=' **The Regular Sub Of The Month** - Store by store comparison',
       col = 'Store')+
  ylab('Percent total profit margin')+
  xlab('')+
  geom_text(aes(label=Margin), position = position_stack(vjust= 0.5),
           colour = "white", size = 3)+
  theme_bw()+
  theme(axis.text.x = element_text(angle=15, vjust=0.6))
```

## Profit margin for the most profitable unit in September 2012

**\*\*The Regular Sub Of The Month\*\* – Store by store comparison**



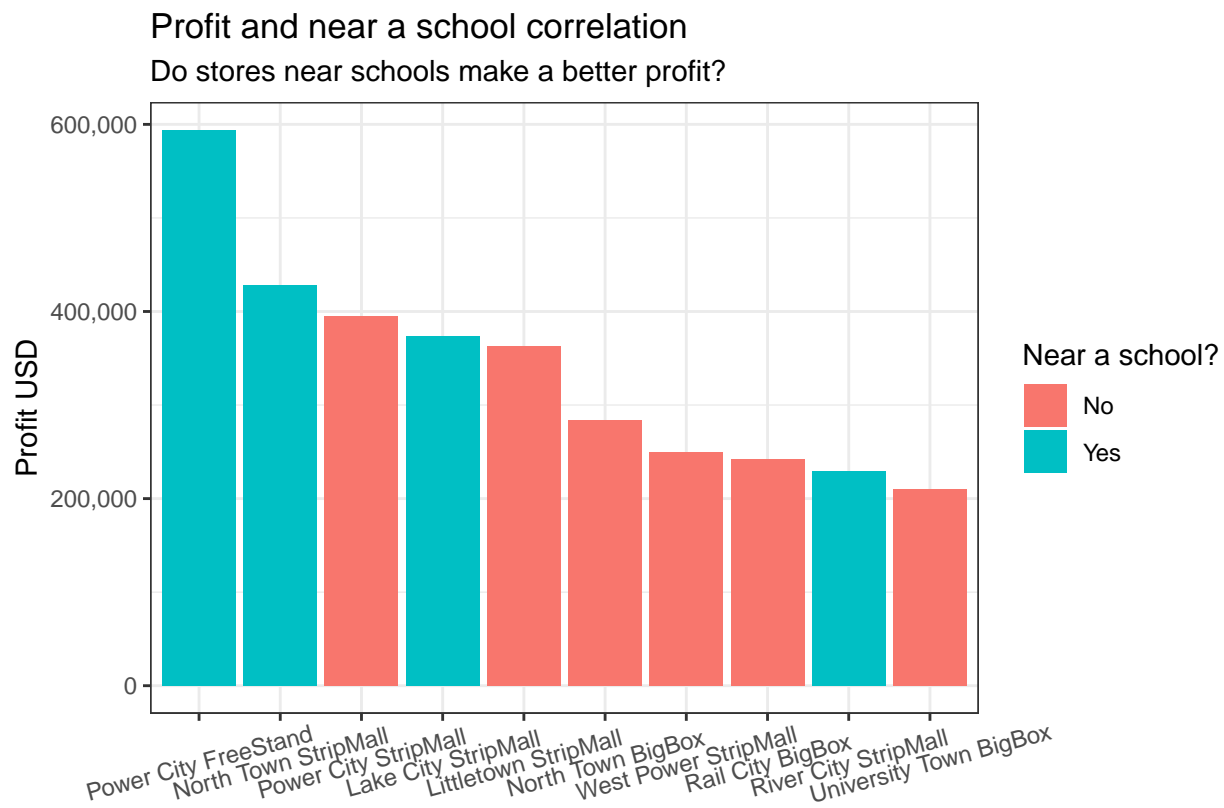
## OPPGAVE 4

Nå skal vi prøve å hjelpe bedriften med å finne nyttig informasjon som kan hjelpe dem å velge lokasjon for et nytt utsalg.

### Barplot som viser korrelasjon mellom profitt og beliggenhet

Har det noe å si for utsalgene om de blir plassert i nærheten av en skole?

```
alldata %>%
  select(Store_Name, Profit, Store_Near_School) %>%
  group_by(Store_Name, Store_Near_School) %>%
  summarise(Profit = sum(Profit)) %>%
  ggplot(aes(x = reorder(Store_Name, -Profit), y = Profit)) +
  geom_col(aes(fill = Store_Near_School)) +
  scale_y_continuous(labels = scales::comma) + # thousands
  labs(title="Profit and near a school correlation",
       subtitle='Do stores near schools make a better profit?',
       fill = 'Near a school?') +
  ylab('Profit USD') +
  xlab('') +
  theme_bw() +
  theme(axis.text.x = element_text(angle=15, vjust=0.6))
```

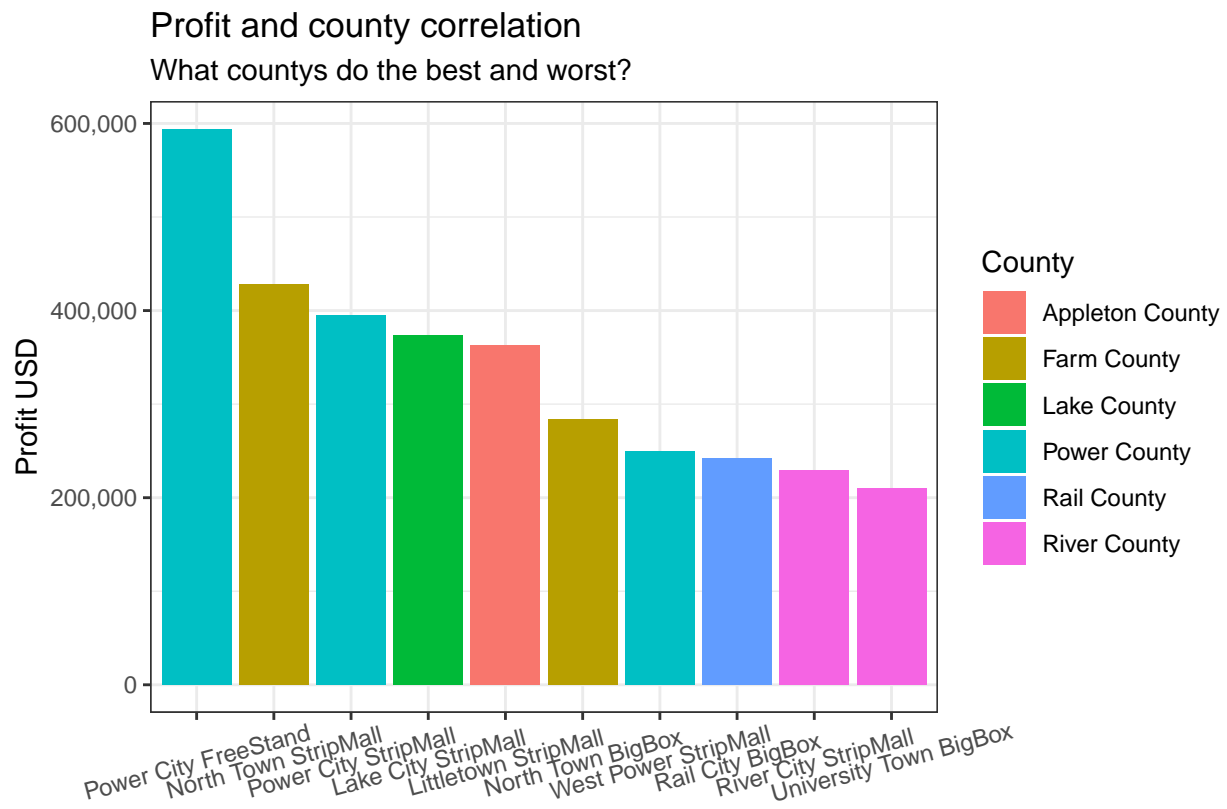


Ja, det kan virke sann. Barplottet viser en klar korrelasjon mellom høy profitt, og en beliggenhet i nærheten av en skole.

## Barplot som viser profitt og county korrelasjon

Hvilke countys bør vurderes?

```
alldata %>%
  select(County_Name, Store_Name, Profit) %>%
  group_by(Store_Name, County_Name) %>%
  summarise(Profit = sum(Profit)) %>%
  ggplot(aes(x = reorder(Store_Name, -Profit), y = Profit)) +
  geom_col(aes(fill = County_Name)) +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Profit and county correlation",
       subtitle = "What countys do the best and worst?",
       fill = 'County') +
  ylab('Profit USD') +
  xlab('') +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 15, vjust = 0.6))
```

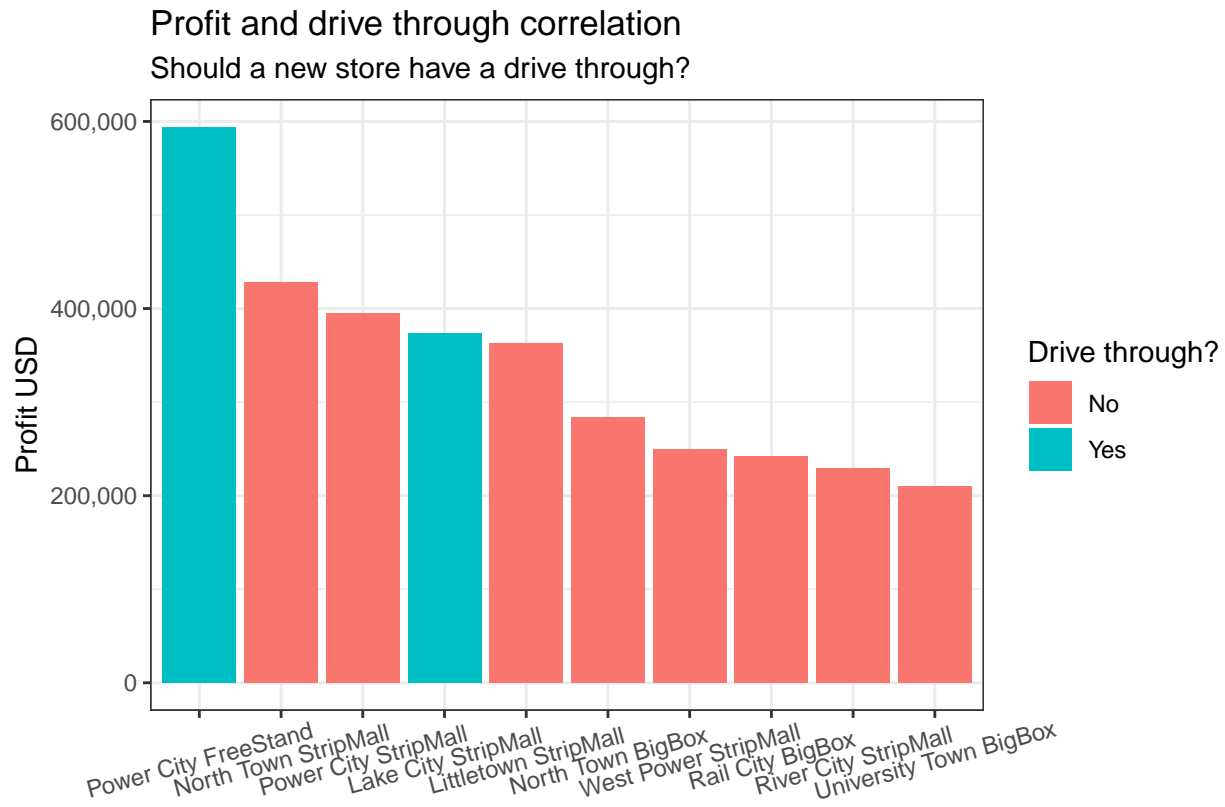


Vi ser at power county gjør det veldig bra. River county er helt i den andre enden, og bør vurderes å legges ned med mindre bedriften kan løse problemene til disse svake utsalgene.

## Barplot som viser korrelasjon mellom profitt og drive through

Bør den nye lokasjonen ha drive through? Er det verdt det?

```
alldata %>%
  select(Store_Name, Profit, Store_Drive_Through, Store_Competition_Fastfood) %>%
  group_by(Store_Name, Store_Drive_Through, Store_Competition_Fastfood) %>%
  summarise(Profit = sum(Profit)) %>%
  ggplot(aes(x = reorder(Store_Name, -Profit), y = Profit)) +
  geom_col(aes(fill = Store_Drive_Through)) +
  scale_y_continuous(labels = scales::comma) +
  labs(title = "Profit and drive through correlation",
       subtitle = "Should a new store have a drive through?",
       fill = "Drive through?") +
  ylab("Profit USD") +
  xlab("") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 15, vjust = 0.6))
```



Ja, det ser absolutt ut som det vil være fornuftig å ha en drive through.

## **Sluttvurdering**

Jeg har bare tatt med grafer som har vist relevant positiv korrelasjon.

Jeg vil anbefale at bedriften vurderer å lage et nytt utvalg i Power county, samtidig som de bør vurdere å legge ned utvalgene i River county.

Jeg anbefaler også at beliggenheten bør være i nærheten av en skole, og at en drive through, selv om det er kostbart, på lang sikt vil hjelpe stort med omsetningen.