```
# Assignment: Week 5 Exercise 5.2
# Name: Sharma, Dipika
# Date: 2020-04-18

" a.  Using the dplyr package, use the 6 different operations to analyze/transform the
    data - GroupBy, Summarize, Mutate, Filter, Select, and Arrange –
    Remember this isn't just modifying data, you are learning about your data also –
    so play around and start to understand your dataset in more detail"
# Ans
install.packages("dplyr")
install.packages("readxl")
library("readxl")
housing_df <- read_excel("/Users/dipikasharma/R_Projects/DSC520/data/week-7-
housing.xlsx")
housing_df
library(dplyr)
select_df <- select(housing_df, `Sale Price`, sale_reason, sale_instrument)
select_df

filter_df <- filter(housing_df, sale_reason == 1)
filter_df

sfilter_df <- housing_df %>% filter(sale_reason == 1)  %>%  select(`Sale Price`, sale_reason,
sale_instrument)
sfilter_df

mutate_df <- sfilter_df %>% mutate(Saleprice_divident = (`Sale Price`*4)/100)
mutate_df

sgroupby_df <- housing_df %>% group_by(sale_reason) %>% summarize(Saleprice = sum(`Sale
Price`, na.rm = TRUE))
sgroupby_df

arrange_df <- sgroupby_df %>% arrange(Saleprice)
arrange_df

"Using the purrr package – perform 2 functions on your dataset.
You could use zip_n, keep, discard, compact, etc."
install.packages("purrr")
library(purrr)

square <- function(x){
  return(x*x)
}
```

```
map(sgroupby_df$sale_reason, square)

library(purrr)


to_loss <- function(x, y){
  return(x- (x*y)/100)
}
map_df <- map2(sgroupby_df$Saleprice, 5, to_loss)
map_df

map_df %>% keep(map_df>20000000)

map_df %>% discard(map_df>200000)

testt2 <- sgroupby_df$sale_reason %>% keep(sgroupby_df$sale_reason>10)
testt2
#c. "Use the cbind and rbind function on your dataset"
#Ans
library(dplyr)
ID <- c(101:117)
new_df <- cbind(sgroupby_df, ID)
new_df
part1_df <- new_df %>% select(ID, Saleprice) %>% filter(ID <= 105)
part1_df
part2_df <- new_df %>% select(ID, Saleprice) %>% filter(ID > 105)
part2_df
rbind(part1_df, part2_df)

"Split a string, then concatenate the results back together"

install.packages("tidyverse")
install.packages("dplyr")
library(tidyverse)
library(dplyr)
library("readxl")
nhousing_df <- read_excel("/Users/dipikasharma/R_Projects/DSC520/data/week-7-
housing.xlsx")

nefilter_df <- filter(nhousing_df, sale_reason == 19)
#nefilter_df

nefilter_df$location <- paste(nefilter_df$lon, nefilter_df$lat, sep = " - ")
nefilter_df$TotalYears <- paste(nefilter_df$year_built, nefilter_df$year_renovated, sep = " - ")
```

```
nefilter_df
select(nefilter_df, `Sale Price`, sale_reason, location, lon, lat, TotalYears, year_built,
year_renovated)


install.packages("tidyr")
library(tidyr)

new_housingdf <- nefilter_df %>% separate(TotalYears, c("BeginYear", "EndYear"), " - ")
select(new_housingdf, `Sale Price`, sale_reason, location, lon, lat, BeginYear, EndYear,
year_built, year_renovated)
```