op-koti

This file takes data manipulated in the python script and prepares it further for the shiny dashboard

Packages

```
suppressPackageStartupMessages({
  library(ggplot2)
  library(plotly)
  library(tidyverse)
  library(magrittr)
  library(ggrepel)
  library(repr)
  library(gridExtra)
  library(ggpubr)
})
```

Read Data

```
df <- read.csv('op-koti.csv',
               stringsAsFactors = F,
               colClasses = c("character", "character", "integer", "integer", "numeric", "numeric", "cha
)
head(df)
```

```
##       id listingType floor numberOfRooms     price debtFreePrice      city
## 1 510933 Omakotitalo     0             5 357000.00        357000 Sastamala
## 2 517946    Rivitalo     1             3  96000.00         96000 Rovaniemi
## 3 518324  Kerrostalo     3             1 157000.00        157000   Tampere
## 4 510967  Kerrostalo     3             4  64980.92         69000      Kemi
## 5 517750 Omakotitalo     0             5 240000.00        240000   Tampere
## 6 518406    Rivitalo     0             3  55000.00         55000     Jämsä
##      region      district postalCode livingArea totalArea buildingAge centrum
## 1 Sastamala        Häijää      38420      214.0     312.0          17       0
## 2 Rovaniemi    Ounasrinne      96440       78.5      78.5          37       0
## 3  Keskusta       Tammela      33100       27.5      27.5          60       1
## 4      Kemi      Keskusta      94100      100.0     100.0          68       1
## 5     Länsi Lentävänniemi      33410       95.0     122.0          42       0
## 6     Jämsä         Halli      35600       72.0      72.0          48       0
##   hasSauna hasBalcony hasParking hasWalkInCloset hasStorageRoom
## 1        1          0          0               0              1
## 2        1          0          0               0              0
## 3        0          0          0               0              0
## 4        0          1          0               0              0
## 5        1          0          0               0              0
## 6        0          0          0               0              0
```

Check for NAs. It should not contain any because it's already been cleaned

```
any(is.na(df))
```

```
## [1] FALSE
```

Creating new variables for price per meter square and link to the respective houses on the website

```
df <- df %>%
  mutate(pricePMsq = debtFreePrice/totalArea, link = paste0("<a href='https://op-koti.fi/kohde/",id,"'>
```

A bit more of housekeeping. Checking the listing types

```
table(df$listingType)
```

```
##
##       Erillistalo        Kerrostalo Kytketty paritalo         Luhtitalo
##                16              1360                 2                48
##       Omakotitalo          Paritalo           Puutalo          Rivitalo
##               672                80                 7               570
```

Merging 'Kytketty paritalo' into 'Paritalo'

```
df$listingType[df$listingType %in% "Kytketty paritalo"] <- "Paritalo"
table(df$listingType)
```

```
##
## Erillistalo  Kerrostalo   Luhtitalo Omakotitalo    Paritalo     Puutalo
##          16        1360          48         672          82           7
##    Rivitalo
##         570
```

Rearranging columns

```
i=1
for (item in colnames(df)){
  #     sprintf("%d : %d", i, item)
  print(paste(i,':',item))
  i = i + 1
}
```

```
## [1] "1 : id"
## [1] "2 : listingType"
## [1] "3 : floor"
## [1] "4 : numberOfRooms"
## [1] "5 : price"
## [1] "6 : debtFreePrice"
## [1] "7 : city"
## [1] "8 : region"
## [1] "9 : district"
## [1] "10 : postalCode"
```

```
## [1] "11 : livingArea"
## [1] "12 : totalArea"
## [1] "13 : buildingAge"
## [1] "14 : centrum"
## [1] "15 : hasSauna"
## [1] "16 : hasBalcony"
## [1] "17 : hasParking"
## [1] "18 : hasWalkInCloset"
## [1] "19 : hasStorageRoom"
## [1] "20 : pricePMsq"
## [1] "21 : link"
```

```r
df <- df[c(1,2,4,5,6,11,12,20,7:10,3,13:19,21)]
head(df)
```

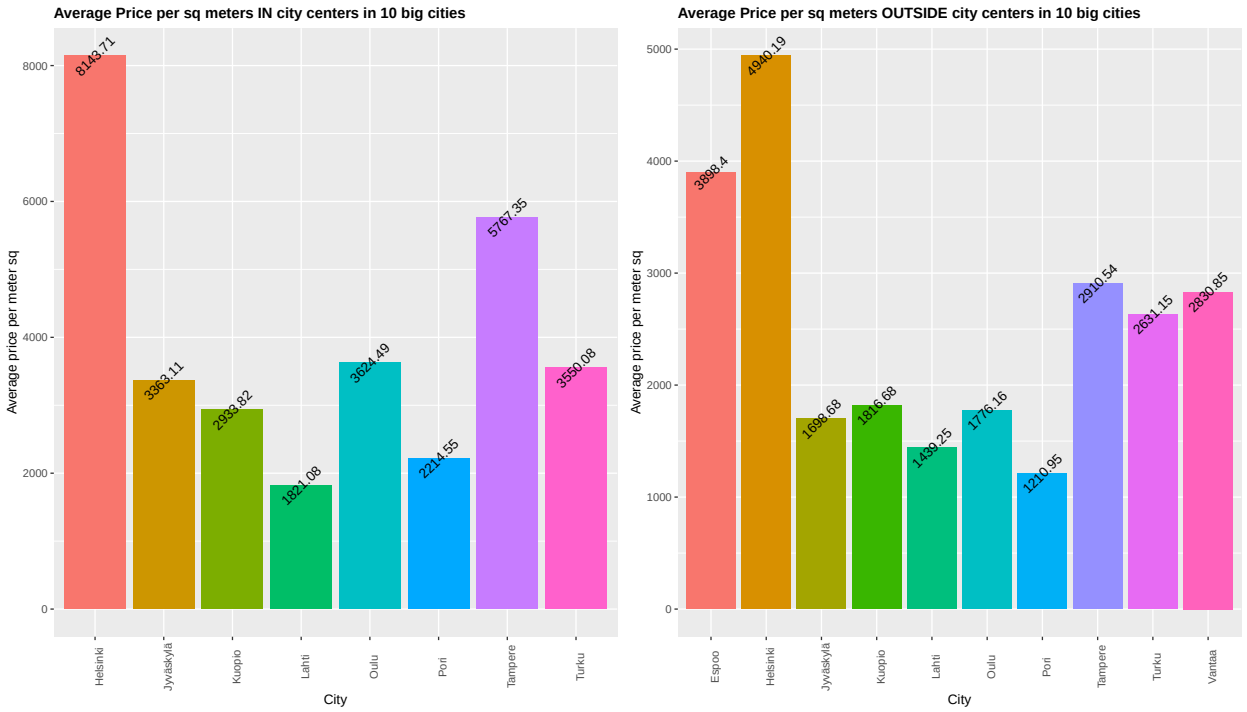
```
##       id listingType numberOfRooms     price debtFreePrice livingArea totalArea
## 1 510933 Omakotitalo             5 357000.00        357000      214.0     312.0
## 2 517946    Rivitalo             3  96000.00         96000       78.5      78.5
## 3 518324  Kerrostalo             1 157000.00        157000       27.5      27.5
## 4 510967  Kerrostalo             4  64980.92         69000      100.0     100.0
## 5 517750 Omakotitalo             5 240000.00        240000       95.0     122.0
## 6 518406    Rivitalo             3  55000.00         55000       72.0      72.0
##   pricePMsq      city    region      district postalCode floor buildingAge
## 1 1144.2308 Sastamala Sastamala        Häijää      38420     0          17
## 2 1222.9299 Rovaniemi Rovaniemi    Ounasrinne      96440     1          37
## 3 5709.0909   Tampere  Keskusta       Tammela      33100     3          60
## 4  690.0000      Kemi      Kemi      Keskusta      94100     3          68
## 5 1967.2131   Tampere     Länsi Lentävänniemi      33410     0          42
## 6  763.8889     Jämsä     Jämsä         Halli      35600     0          48
##   centrum hasSauna hasBalcony hasParking hasWalkInCloset hasStorageRoom
## 1       0        1          0          0               0              1
## 2       0        1          0          0               0              0
## 3       1        0          0          0               0              0
## 4       1        0          1          0               0              0
## 5       0        1          0          0               0              0
## 6       0        0          0          0               0              0
##                                                                        link
## 1 <a href='https://op-koti.fi/kohde/510933'>https://op-koti.fi/kohde/510933</a>
## 2 <a href='https://op-koti.fi/kohde/517946'>https://op-koti.fi/kohde/517946</a>
## 3 <a href='https://op-koti.fi/kohde/518324'>https://op-koti.fi/kohde/518324</a>
## 4 <a href='https://op-koti.fi/kohde/510967'>https://op-koti.fi/kohde/510967</a>
## 5 <a href='https://op-koti.fi/kohde/517750'>https://op-koti.fi/kohde/517750</a>
## 6 <a href='https://op-koti.fi/kohde/518406'>https://op-koti.fi/kohde/518406</a>
```

Let's plot some graphs using the data Let's take only the **ten most populated municipalities** in Finland. They are Helsinki, Espoo, Tampere, Vantaa, Oulu, Turku, Jyväskylä, Kuopio, Lahti, Pori (in decreasing order of population)

```r
big_cities <- list('Helsinki', 'Espoo', 'Tampere', 'Vantaa', 'Oulu', 'Turku', 'Jyväskylä', 'Kuopio', 'La
```

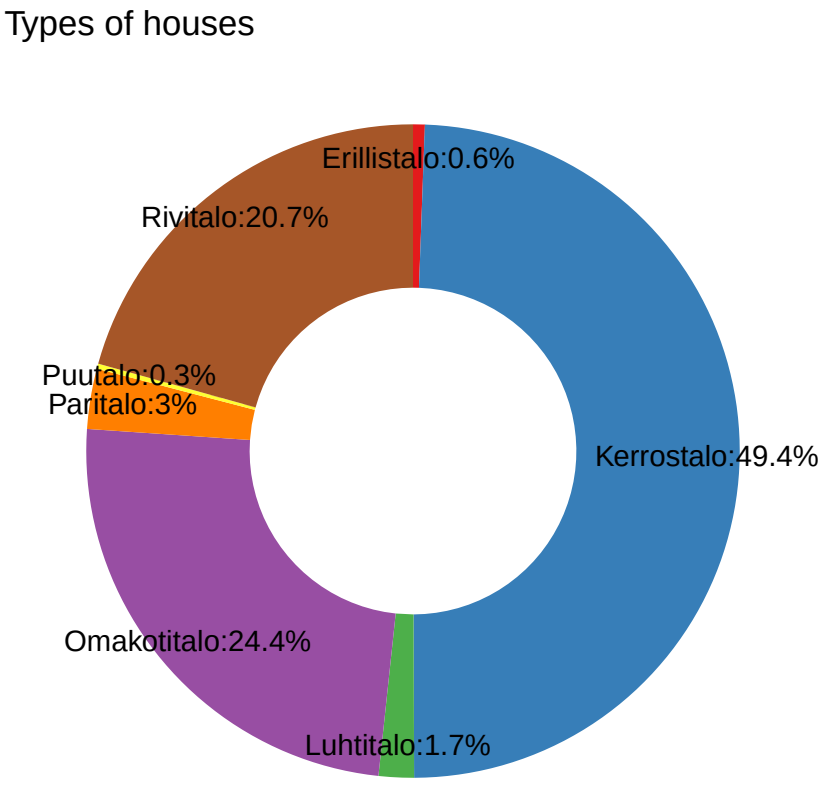
Average price per meter squared when properties are in or outside the city center

```r
pl1 <- df %>%
  filter(centrum == 1 & city %in% big_cities) %>%
  group_by(city) %>%
  summarize(AvgPricePMsq = mean(pricePMsq, na.rm = T)) %>%
  # ggplot(aes(x = reorder(city, -AvgPricePMsq), y = AvgPricePMsq, fill = city)) +
  ggplot(aes(x = city, y = AvgPricePMsq, fill = city)) +
  geom_bar(stat = 'identity') +
  xlab('City') + ylab('Average price per meter sq') +
  geom_text(aes(label = round(AvgPricePMsq,2)), size = 4, position = position_stack(vjust = 1), angle =
  ggtitle("Average Price per sq meters IN city centers in 10 big cities") +
  theme(axis.text.x = element_text(angle = 90), plot.title = element_text(size = 12, face = "bold"), leg

pl2 <- df %>%
  filter(centrum == 0 & city %in% big_cities) %>%
  group_by(city) %>%
  summarize(AvgPricePMsq = mean(pricePMsq, na.rm = T)) %>%
  # ggplot(aes(x = reorder(city, -AvgPricePMsq), y = AvgPricePMsq, fill = city)) +
  ggplot(aes(x = city, y = AvgPricePMsq, fill = city)) +
  geom_bar(stat = 'identity') +
  xlab('City') + ylab('Average price per meter sq') +
  geom_text(aes(label = round(AvgPricePMsq,2)), size = 4, position = position_stack(vjust = 1), angle =
  ggtitle("Average Price per sq meters OUTSIDE city centers in 10 big cities") +
  theme(axis.text.x = element_text(angle = 90), plot.title = element_text(size = 12, face = "bold"), leg

gt <- arrangeGrob(pl1, pl2, ncol = 2)
# Transform to a ggplot and print
as_ggplot(gt)
```



Now let's take a look at the types of houses listed

```
df %>%
  # filter(city %in% big_cities) %>%
  group_by(listingType) %>%
  summarise(freq = n()) %>% # freq table for types of houses
  mutate(fraction = freq/sum(freq), # percentages
         ymax = cumsum(fraction), #cumulative percentages (top of each rectangle)
         ymin = c(0, head(ymax, n = -1)), #bottom of each rectangle

         labelPosition = (ymax + ymin) / 2, #position of the label
         label = paste0(listingType, ":", round(fraction*100,1),"%") #label
  ) %>%
  ggplot(aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3, fill=listingType)) +
  geom_rect() +
  geom_text(x=3.8, aes(y=labelPosition, label=label), color='black', size=4) + # x here controls label
  scale_fill_brewer(palette = 'Set1') +
  coord_polar(theta="y") + # Try to remove that to understand how the chart is built initially
  xlim(c(2, 4)) + # Try to remove that to see how to make a pie chart
  theme_void() +
  ggtitle('Types of houses')+
  theme(legend.position = "none")
```



The majority of properties are Kerrostalo, Rivitalo or Omakotitalo

save the data to csv file for dashboard

```
write.csv(df, "/Users/avinashmalla/GitHub/opKotiDashboard/forDash.csv", row.names = F)
```

```
glimpse(df)
```

```
## Rows: 2,755
## Columns: 21
## $ id              <chr> "510933", "517946", "518324", "510967", "517750", "518~
## $ listingType     <chr> "Omakotitalo", "Rivitalo", "Kerrostalo", "Kerrostalo",~
## $ numberOfRooms   <int> 5, 3, 1, 4, 5, 3, 2, 3, 3, 2, 5, 4, 6, 2, 2, 3, 3, 4, ~
## $ price           <dbl> 357000.00, 96000.00, 157000.00, 64980.92, 240000.00, 5~
## $ debtFreePrice   <dbl> 357000, 96000, 157000, 69000, 240000, 55000, 119000, 1~
## $ livingArea      <dbl> 214.0, 78.5, 27.5, 100.0, 95.0, 72.0, 60.0, 78.0, 77.0~
## $ totalArea       <dbl> 312.0, 78.5, 27.5, 100.0, 122.0, 72.0, 60.0, 78.0, 77.~
## $ pricePMsq       <dbl> 1144.2308, 1222.9299, 5709.0909, 690.0000, 1967.2131, ~
## $ city            <chr> "Sastamala", "Rovaniemi", "Tampere", "Kemi", "Tampere"~
## $ region          <chr> "Sastamala", "Rovaniemi", "Keskusta", "Kemi", "Länsi",~
## $ district        <chr> "Häijää", "Ounasrinne", "Tammela", "Keskusta", "Lentäv~
## $ postalCode      <chr> "38420", "96440", "33100", "94100", "33410", "35600", ~
## $ floor           <int> 0, 1, 3, 3, 0, 0, 4, 4, 0, 1, 0, 1, 0, 1, 1, 1, 1, 6, ~
## $ buildingAge     <dbl> 17, 37, 60, 68, 42, 48, 46, 55, 30, 35, 0, 42, 62, 32,~
## $ centrum         <int> 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ hasSauna        <int> 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, ~
## $ hasBalcony      <int> 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ hasParking      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, ~
## $ hasWalkInCloset <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, ~
## $ hasStorageRoom  <int> 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ link            <chr> "<a href='https://op-koti.fi/kohde/510933'>https://op-~
```