Dowloading data

Team 1

2023-02-27

Loading in the data for December 2020 to May 2021, with a for loop so we minimize the change on errors:

```
# create a vector of the dates
dates <- c('12.20', '01.21', '02.21', '03.21', '04.21', '05.21')

# loading in the data with the for loop
for (date in dates) {
  url <- paste0('https://github.com/course-dprep/SickAirbnbPricesAcrossNetherlands/raw/main/Data/listin
  filename <- paste0('listings-', date, '.csv.gz')
  download.file(url, destfile = filename)
  assign(paste0('listings', date, '_df'), read.csv(gzfile(filename)))
}
```

Loading in the data for December 2022:

```
# loading in the data
url <- "http://data.insideairbnb.com/the-netherlands/north-holland/amsterdam/2022-12-05/data/listings.c
destfile <- "listings-12.22.csv.gz"
download.file(url, destfile, method = "curl")
listings12.22_df <- read.csv(gzfile(destfile))
```

When investigating the data, we can see that the December 2022 data has one more column than all the other datasets. Since this column is not improtant for us, we can delete this column:

```
listings12.22_df$source <- NULL
```

We can then merge all the data frames by rows, since we made sure all colums were the same:

```
# merge the two data frames using rbind
merged_df <- rbind(listings12.20_df, listings01.21_df, listings02.21_df, listings03.21_df, listings04.2
```