Projekt Zaliczeniowy 1

Michał Hiszpański, Nikola Kubiczek, Adrian Mirecki

2026-04-29

Data Fields

1. `city` - the name of the city where the property is located
2. `type` - type of the building
3. `squareMeters` - the size of the apartment in square meters
4. `rooms` - number of rooms in the apartment
5. `floor` / floorCount - the floor where the apartment is located and the total number of floors in the building
6. `buildYear` - the year when the building was built
7. `latitude`, `longitude` - geo coordinate of the property
8. `centreDistance` - distance from the city centre in km
9. `poiCount` - number of points of interest in 500m range from the apartment (schools, clinics, post offices, kindergartens, restaurants, colleges, pharmacies)
10. `[poiName]Distance` - distance to the nearest point of interest (schools, clinics, post offices, kindergartens, restaurants, colleges, pharmacies)
11. `ownership` - the type of property ownership
12. `condition` - the condition of the apartment
13. `has[features]` - whether the property has key features such as assigned parking space, balcony, elevator, security, storage room
14. `price` - offer price in Polish Zloty apartments_pl_YYYY_MM.csv: sale price apartments_rent_pl_YYYY_MM.csv: monthly rent

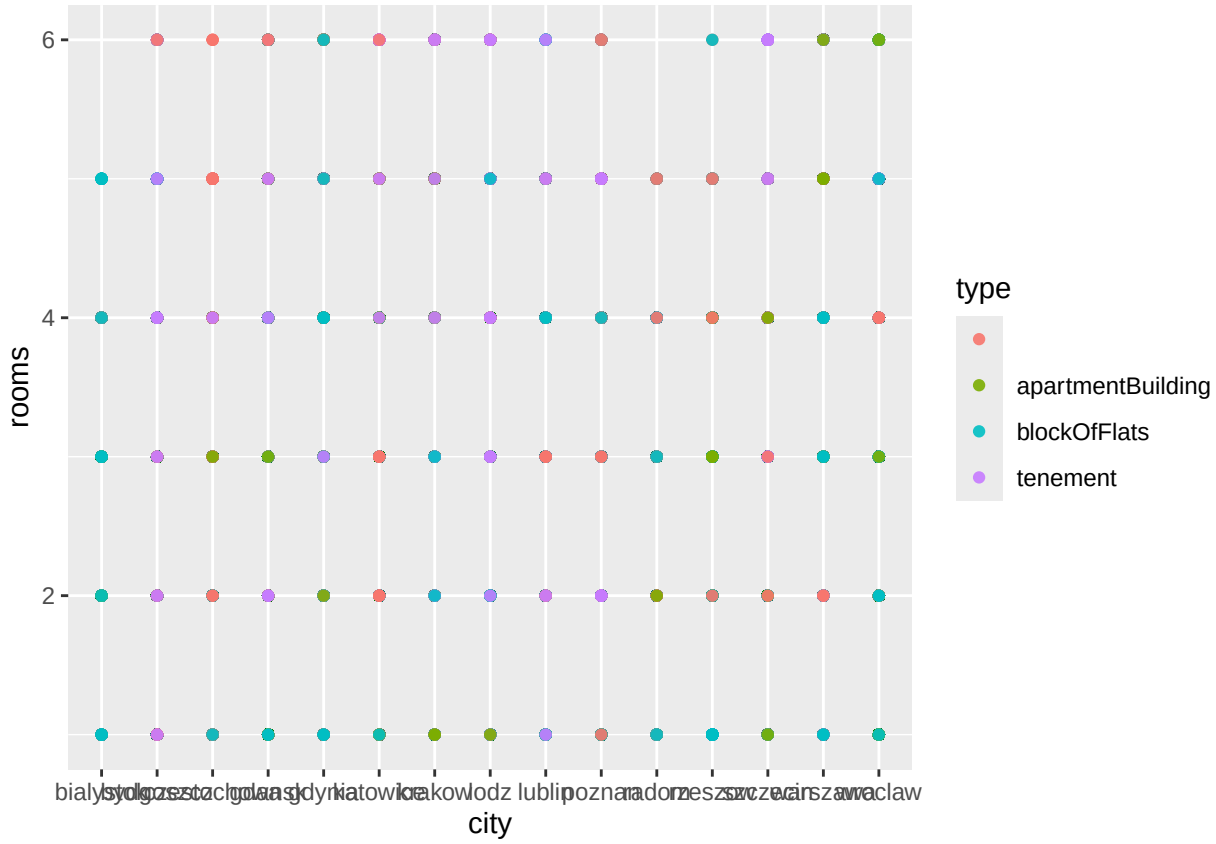
```
files <- list.files("apartments", pattern = "\\.csv$", full.names = TRUE)
files <- sort(files)

df_list <- lapply(seq_along(files), function(i) {
  d <- read.csv(files[i])
  d$month_id <- i
  d
})

rent <- do.call(rbind, df_list)

has_cols <- grep("^has", names(rent), value = TRUE)

rent[has_cols] <- lapply(rent[has_cols], function(x) {
  tolower(x) == "yes"
})
summary(rent)
```

```
##       id                city               type            squareMeters
##  Length:195568      Length:195568      Length:195568      Min.   : 25.00
##  Class :character   Class :character   Class :character   1st Qu.: 44.00
##  Mode  :character   Mode  :character   Mode  :character   Median : 54.60
```

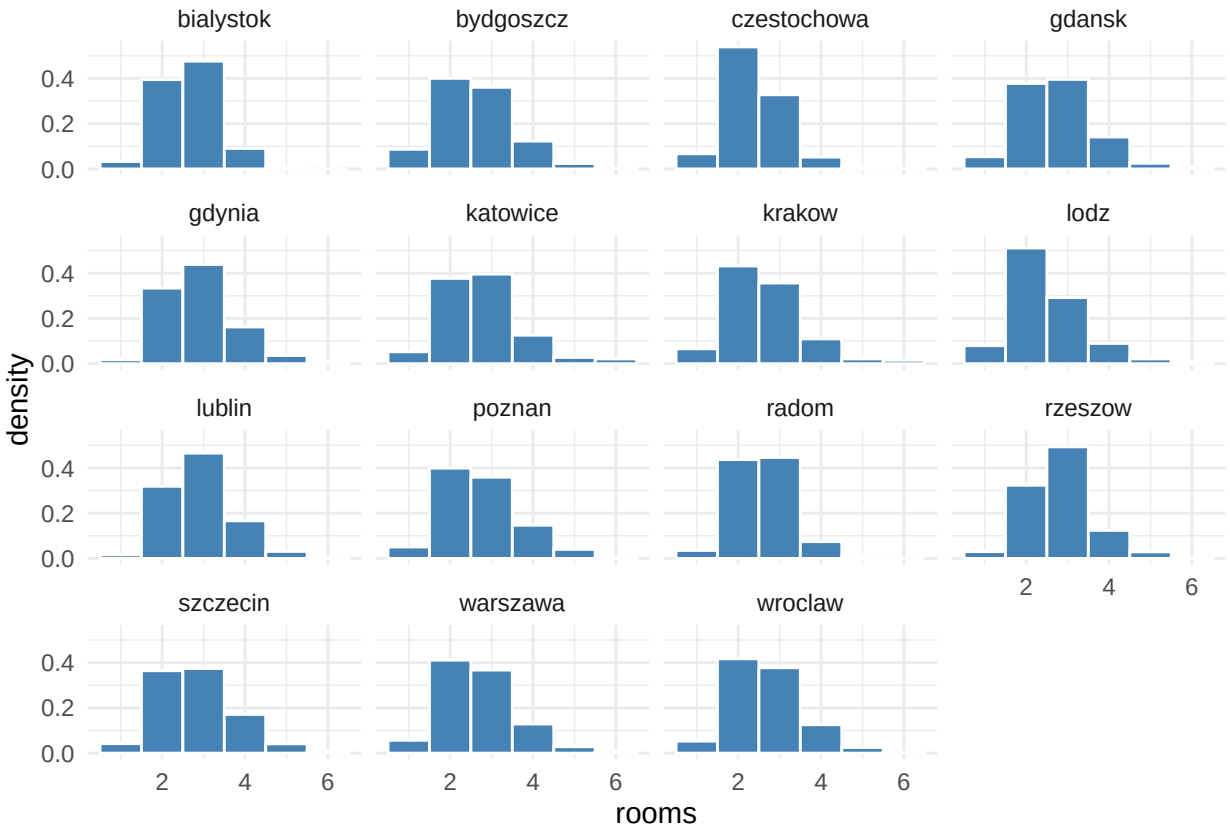
```
##                                                          Mean   : 58.70
##                                                          3rd Qu.: 68.55
##                                                          Max.   :150.00
##
##      rooms           floor          floorCount       buildYear
##  Min.   :1.000   Min.   : 1.000   Min.   : 1.000   Min.   :1850
##  1st Qu.:2.000   1st Qu.: 2.000   1st Qu.: 3.000   1st Qu.:1967
##  Median :3.000   Median : 3.000   Median : 4.000   Median :1994
##  Mean   :2.679   Mean   : 3.332   Mean   : 5.309   Mean   :1986
##  3rd Qu.:3.000   3rd Qu.: 4.000   3rd Qu.: 6.000   3rd Qu.:2016
##  Max.   :6.000   Max.   :29.000   Max.   :29.000   Max.   :2024
##                  NA's   :34594    NA's   :2383     NA's   :32216
##     latitude       longitude     centreDistance      poiCount
##  Min.   :49.98   Min.   :14.45   Min.   : 0.010   Min.   :  0.00
##  1st Qu.:51.11   1st Qu.:18.52   1st Qu.: 2.010   1st Qu.:  7.00
##  Median :52.19   Median :19.90   Median : 3.980   Median : 14.00
##  Mean   :52.03   Mean   :19.47   Mean   : 4.351   Mean   : 20.67
##  3rd Qu.:52.41   3rd Qu.:20.99   3rd Qu.: 6.150   3rd Qu.: 24.00
##  Max.   :54.61   Max.   :23.21   Max.   :16.940   Max.   :212.00
##
##  schoolDistance   clinicDistance   postOfficeDistance kindergartenDistance
##  Min.   :0.0020   Min.   :0.0010   Min.   :0.0010     Min.   :0.0010
##  1st Qu.:0.1760   1st Qu.:0.3560   1st Qu.:0.2390     1st Qu.:0.1560
##  Median :0.2900   Median :0.6760   Median :0.3930     Median :0.2620
##  Mean   :0.4127   Mean   :0.9703   Mean   :0.5163     Mean   :0.3676
##  3rd Qu.:0.4680   3rd Qu.:1.2370   3rd Qu.:0.6230     3rd Qu.:0.4160
##  Max.   :4.9460   Max.   :4.9990   Max.   :4.9700     Max.   :4.9610
##  NA's   :168      NA's   :728      NA's   :248        NA's   :207
##  restaurantDistance collegeDistance pharmacyDistance  ownership
##  Min.   :0.0010     Min.   :0.004   Min.   :0.0010   Length:195568
##  1st Qu.:0.1140     1st Qu.:0.577   1st Qu.:0.1420   Class :character
##  Median :0.2290     Median :1.120   Median :0.2390   Mode  :character
##  Mean   :0.3453     Mean   :1.443   Mean   :0.3581
##  3rd Qu.:0.4090     3rd Qu.:2.055   3rd Qu.:0.4060
##  Max.   :4.9850     Max.   :5.000   Max.   :4.9920
##  NA's   :479        NA's   :5436    NA's   :277
##  buildingMaterial    condition         hasParkingSpace hasBalcony
##  Length:195568      Length:195568      Mode :logical   Mode :logical
##  Class :character   Class :character   FALSE:142410    FALSE:83519
##  Mode  :character   Mode  :character   TRUE :53158     TRUE :112049
##
##
##
##
##  hasElevator     hasSecurity     hasStorageRoom      price
##  Mode :logical   Mode :logical   Mode :logical   Min.   : 150000
##  FALSE:103364    FALSE:173630    FALSE:110583    1st Qu.: 520000
##  TRUE :92204     TRUE :21938     TRUE :84985     Median : 699000
##                                                  Mean   : 784183
##                                                  3rd Qu.: 930000
##                                                  Max.   :3250000
##
##     month_id
##  Min.   : 1.000
```

```
##  1st Qu.: 3.000
##  Median : 6.000
##  Mean   : 6.182
##  3rd Qu.: 9.000
##  Max.   :11.000
##
```

```
ggplot(rent) + geom_point(aes(x=city, y=rooms, col=type), alpha=0.9)
```




```
ggplot(rent, aes(x = rooms)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 1, fill = "steelblue", color = "white") +
  facet_wrap(~ city) +
  theme_minimal()
```


bialystok
bydgoszcz
czestochowa
gdansk
gdynia
katowice
krakow
lodz
lublin
poznan
radom
rzeszow
szczecin
warszawa
wroclaw
density
0.0
0.2
0.4
2
4
6
rooms