Background

"Krysha KZ" is an online platform for ads related to real estate. On krisha.kz site, you can easily sell, buy, rent, and even repair your real estate. It is a brand that answers the housing question and helps people find their dream home. The "Krysha" has the largest selection of apartments for every taste and budget. The site has constant access to an up-to-date database of advertisements for the sale, purchase, rent of apartments, houses, summer cottages, land plots in cities and villages of Kazakhstan, as well as information on any types of commercial real estate. All ads indicate the price of the property, which is very important for analyzing a specific situation.

Hypothetical case

Jules is a real estate agent in Kaskelen. He just got a call from a potential customer, Mr. Wallace, who wants to sell his property in Almaty city and was inquiring how much he might be able to get. In particular, he wondered how likely it would be that he could get at least KZT30,000,000. In order to get information about the property, Jules drove to see the property. The property is located 10 minutes walking distance from “Abay” metro station and turns out to be a typical residential property with 80 square meter.

After returning to the office, Jules checked the title records and found that the building is seven years old. Then he checked Krysha.kz database of recent commercial real estate deals and quickly built the descriptive statistics table and a set of regression models in order to forecast the likely rent price for Mr. Wallace's property.

Reference

www.Krysha.kz

Exhibit 1

- **price** – price of the property in KZT
- **rooms** – number of rooms (integer number)
- **area** – total area of the property (apartment) in square meters
- **floor** – floor of the apartment/number of storeys in the building
- **street** - street name of the address
- **city** – name of the city (Almaty or Nur-Sultan)
- **ZHK** – name of the complex (“ЖК, жилой комплекс”)
- **type** – type of the building
- **year** – year of commissioning

Instructions

This is a team assignment. Each member of the team receives the same grade. Submission is online (see course webpage). In order to be graded, you need to upload one pdf file (no longer than 3 pages with font size 12pt) and your IPython notebook (this should be well commented and run without errors) along with your CSV file with cleaned dataset. Any additional material you judge relevant that complements your submission can be submitted as additional files. Make sure that the Team number and all names of the team members are clearly listed. Late submissions (but submitted before in-class discussions) or inappropriately formatted cases will have points deducted. Missed cases are worth 0 points. ***Important***: ***follow the file naming convention***:

- `css429_case1_Team#.ipynb`
- `css429_case1_Team#.pdf`
- `css429_case1_Team#.csv`

Assignment

Your answers should be clear and provide unambiguous recommendations when asked. Please provide explanations for your answers and any outputs that you feel are needed to support your argument.

1. Before running any model, pick any two variables and attempt to show an (interesting) relation via visualization. This requires one to formulate a question, and to communicate clearly a conclusion based on data visualization. Ideally this would suggest ways to act on the issue.
2. Try to explain **`price`** with **`area`**. What can you conclude?
3. In your opinion, can any of the variables provided in Exhibit 1 help to predict apartment prices? Since opinions do not provide strong arguments, provide a simple evidence based on data.
4. Run a multiple regression at least 4 of the variables described in Exhibit 1. Which variables are statistically significant?

The following questions do not require additional data analysis calculations:

5. Do you think you can improve the model with additional predictors? Where/how can we get them and introduce to our model?
6. Hypothetical case illustrates one possible use case of such model use. Can you think of other examples where such model will be useful? Give at least 3 examples.

Assessment weights

50% for your answers (.pdf), 30% for your notebook with analysis and python code (.ipynb), 20% for data cleaning efforts (.csv)