

Quantitative Data Analysis – Exercises

(Week 03)

In these exercises, you will continue preparing the apartment data. In detail, you will learn how to create new variables, geocode addresses, combine the data with other data and work with pivot tables. In the data analysis process model, these exercises cover part of the step “Preparing & storing” data (see figure 1). Results of the exercises must be uploaded as separate files (no .zip files) by each student on Moodle. Details on how to submit the results can be found in the tasks below.

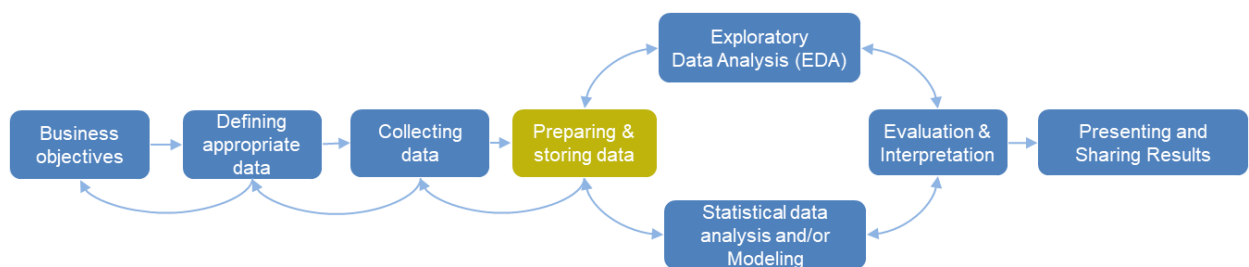


Figure 1: Data analysis process model (see slides of week 01)

Task 1

In these exercises, you will learn to create new variables. Therefore, we will use the prepared rental apartment data. The tasks are:

1. Run the Jupyter notebook '[apartments_data_preparation_zuerich.ipynb](#)' step by step.
2. In the Jupyter notebook, go to the section 'Create additional variables from the apartment's descriptions' and look at the example under 'Create new binary (0/1) variable 'luxurious''.
3. Based on this example, create the following binary (0/1) variables '**furnished**', '**balcony**', and '**central**'. You are free to create additional variables. Note that there are only german words in the apartment descriptions which must be used to define the pattern for each for these variables. You can use <https://www.deepl.com> for translation.
4. In the Jupyter notebook, go to the example: 'Create categorical variable based on apartment area'.
5. Based on this example, create a new categorical variable based on the variable 'price_per_m2'. The variable should contain three levels 'low', 'medium', 'high'.
6. The output of the Jupyter notebook is a file '[apartments_data_prepared.csv](#)' which will be used in the Jupyter notebook of the following task 2.

To be submitted on Moodle: Jupyter notebook as html-file
'[apartments_data_preparation_zuerich.html](#)' with the additional variables described above.

Task 2

In these exercises you will learn how to merge the data with geocoded data and municipality-level data. The tasks are:

- Run the Jupyter notebook '[combining_and_organizing_data.ipynb](#)' step by step.
- Open the file '[municipality_data_with_taxable_income.xlsx](#)'. It contains municipality-level data which are merged with the apartment data in the Jupyter notebook. Look at the details of these data.
- In the Jupyter notebook, look at the section 'Join municipality data to rental apartment data using `.merge()`'. In this section, the municipality-level data are merged to the apartments data. These data are:

Table 1: Variables in the file '[municipality_data_with_taxable_income.xlsx](#)'

| Variable name | Description of variables |
|---------------------|--|
| bfs_number | Municipality ID |
| bfs_name | Municipality name |
| pop | Number of residents |
| pop_dens | Population density per km ² |
| frg_pct | Percentage of foreigners |
| emp | Number of employees |
| mean_taxable_income | Mean taxable income per household |

To be submitted on Moodle: see Task 3

Task 3

In these exercises, you will learn to work with pivot tables in Python. The tasks are:

- Use the Jupyter notebook '[combining_and_organizing_data.ipynb](#)' prepared in task 2.
- Go to the section 'Pivoting data using `.pivot_table()`' and look at the example.
- Add the variable `price_per_m2` to the existing pivot table.
- In the pivot table, compare the price with the `price_per_m2`.
- Is there a relationship between the variable `area` and the `price_per_m2`? If so, how can this be explained? Include a short explanation in the Jupyter notebook.
- Create a new pivot table and use `aggfunc='count'` as the aggregation function. How is the number of apartments distributed in the pivot table? Include a short description in the Jupyter notebook cell.
- Create a new pivot table with the mean values of `rooms`, `area`, `price`, and `price_per_m2` per municipality. Sort the pivot table by price and area with the most expensive municipalities on top.

To be submitted on Moodle: Jupyter notebook as html-file
'[combining_and_organizing_data.html](#)' with the additional variable
mean_taxable_income from task 2 and the pivot tables from task 3.