# **HOUSEHOLD SERVICES. Customer satisfaction research.**

**Data for the project:**

https://drive.google.com/file/d/1eT\_a2KPPL9JvPUS0FVrW0Ofd4db3gQdd/view?usp=sharing

**Project description:**

You work as an analyst for an international company engaged in the production and sales of climate control equipment: air conditioning and ventilation systems. At one of the exhibitions dedicated to climate equipment, a group of engineers from a friendly company provided the results of their research: data on the use of climate systems in offices in different countries and cities, as well as estimated data from respondents on how comfortable they were in the given conditions. Has the management set you the task of analyzing the data obtained and providing conclusions about which type of climate equipment it is worth investing more in to increase customer satisfaction with the conditions created, depending on the region? And what should you pay attention to when developing a business in different regions?

* **Data description  
  Year** - Year of the study   
  **Time of year** - Spring, Summer, Autumn, winter   
  **Climate** - Name of the climate type   
  **City** - City where the study was conducted   
  **Country** - Country where the study was conducted   
  **Cooling method:** -*Air conditioning* = can be of different types, windows are closed.  
   - *Ventilation* = mechanical ventilation systems and natural cooling.  
   - *Mixed* = mechanical cooling, windows can be either opened or closed.  
   - *NA* = not applicable, for example, in cold winter when the heating is turned on.

**Mixed cooling mode:** -*NA* = buildings that don't belong to the mixed type.  
 - *Air conditioning* = air conditioning is on, windows are closed.  
 - *Ventilation* = air conditioning off, windows open/mechanical  
 ventilation.

**Heating method:** -*Mechanical heating* = the building has a mechanical heating system (portable heaters are not considered mechanical heating at the building level).  
 - *NA* = no mechanical heating, unknown or only used in hot summer weather.

**Age-**Respondent's age

**Gender**-Respondent's gender

**Temperature sensation**-Temperature score by sensation: -3 (cold) to +3 (hot)   
**Temperature sensation (bool)** - 0 = unacceptable, 1 = acceptable   
**Preferred temperature change**-No change, warmer, colder   
**Air movement sensation ( bool)** - 0 = unacceptable, 1 = acceptable   
**Preferred change in air movement**-Less, no change, more   
**Comfort rating** - From 1 (very uncomfortable) to 6 (very comfortable)   
**Insulation** - Factor related to the respondent's clothing   
**Indoor air temperature**-Measured indoor   
**temperature Outdoor air** temperature-Measured outdoor temperature   
**RH**-Relative humidity   
**Air speed** - Air speed, m / s   
**Height** - Respondent's height

**Weight** - Respondent's weight

**Curtains**-0 = open, 1 = closed; if unknown-no   
**Fan** - Fan mode: 0 = off, 1 = on; if unknown-no data   
**Window**-0 = open, 1 = closed; if unknown-no data   
**Doors**-0 = open, 1 = closed; if unknown-no data  
**Heating**-Heating mode: 0 -off, 1 -off enabled; if unknown - no data available

**Average monthly outdoor temperature** - The average monthly outdoor temperature at the time of the study, C.  
**Number of complaints** - The number of complaints about installed equipment received from the customer over the past period

* **Quality data analysis and preprocessing**To conduct the study, you must:
  + Specify column names in accordance with PEP8
  + Explore data types, bring data to the correct type where necessary
  + Check data for omissions and fill them in if necessary. Justify the decision.
  + Check data for explicit and implicit duplicates (in categorical columns). Process duplicates.
  + Perform analytical and graphical data analysis. Draw conclusions.
  + Check the data for outliers and data adequacy. Assume the cause of outliers and process these values.

* **Performing calculations and research**
  + Create an additional categorical column that shows the number of complaints for each client separately: small (≤1), medium (=2), many (>2).
  + Calculate the average age of respondents by gender and country
  + Add a categorical column by dividing respondents into three categories by age: under 44 years old-young age, 45-59 years old-middle age, 60 years and more - old age.
  + Calculate the average comfortable temperature depending on the age category
  + Calculate the median value of temperature and humidity for each cooling type
  + Add a column with the percentage of respondents satisfied with the indoor air temperature relative to all respondents of the same country and gender
  + Create a summary table that groups data by country, gender, and age group, and calculates the average indoor and outdoor air temperature and average relative humidity for each of these groups.
  + Independently find information on the standard comfortable relative humidity in premises (in our case, offices). Create a new categorical column that shows whether the recorded relative humidity meets the standard or not. Assign categories yourself.
  + Investigate the correlation between parameters. Explain the most correlated parameters, does it make real sense or just a feature of the data?
* **Hypothesis testing**
  + Does the cooling method affect the comfort rating?
  + Does gender affect your comfort rating?
  + Does the age group affect the comfort rating?
  + Is there a correlation between the number of complaints and the comfort rating?
  + Test the hypothesis that the average comfort rating differs depending on the country
  + Additionally, put forward and test at least 3 hypotheses on the influence of various factors on the assessment of comfort/temperature sensation/air movement sensation.
* **Regression modeling**Build a regression model of indoor air temperature depending on the type of cooling and other significant factors. The choice of factors is at the discretion of the performer.
* **Dashboard**Based on the results of calculations and research, create an informative dashboard: show the most interesting dependencies found, the results of hypothesis testing, and the main statistics/distributions of the data provided: what can be interesting and useful for the customer's business development.
* **Conclusions**Make the conclusions and dashboard into a presentation for your company's management (no more than 7 slides).  
  When drawing up the project, keep in mind the customer's main question: “What type of climate control equipment should you invest more in in order to increase customer satisfaction with the created conditions, depending on the region? And what should you pay attention to when developing your business in different regions?”