Hi,

This task looks interesting, but also challenging.

1. It involves an ML-algorithm - clustering. We have some tools from Scikit-learn package for Python for that: <http://scikit-learn.org/stable/modules/clustering.html>. In coding club we covered supervised ML (classification and regression), but clustering is a different thing – it is unsupervised learning. So, I think it would be very valuable to get some experience with that.
2. And we have visualization on top of that. We can use Matplotlib package for Python for that (<https://matplotlib.org/>). Or in the worst case we can do some graphs in Excel. In general, visualisation is very important in applied data science, so it would be very valuable to practice this as well.

I am not that much familiar with clustering and visualization. So, I would like to read/do some online courses about it before we start. We discussed with Yeldos yesterday that we could meet next Saturday (28.04) at the coding club in VEF and work on this competition. We can just do what we can and see how it goes. @Rudolfs, @Viesturs, @Solvita what do you think?

Questions to the organizers from my side:

* In vpvkac\_pakalp\_pa\_meeneshiem\_\_2015\_2018.xls – columns ‘skaits’ means the number of customers in a center in month and year?
* Should we create only 3 clusters – small, medium and large service groups according by demand?
* What is the deadline? I personally do not have that much time, so it would be good to have more time for this challenge.
* Should we just submit graphs and our scripts? Or they want some report or something as well?
* When predicting client change we just use % change from columns "Iedzīvotāju skaita pārmaiņu kategorijas prognoze 2011.- 2020.g. min.", "Iedzīvotāju skaita pārmaiņu kategorijas prognoze 2011.- 2020.g. max.", "Iedzīvotāju skaita pārmaiņu kategorijas prognoze 2011.-2030.g. min.” and “Iedzīvotāju skaita pārmaiņu kategorijas prognoze 2011.-2030.g. max.”?