Summary to the management

* we have analysed data of
    * Dataset 1: user profiles such Education, Professional status, City size, Financial_situation, Size of Household
    * Dataset 2: corresponding psychological profile information
* In terms of coding we have cleaned the data, converted "word-like" features (categorical) into "number-like" features (numerical) to be able to perform statistical tests

* We have found a correlation of 0.42 using between the two datasets using Canonical Correlation Analysis.
* Canonical Correlation Analysis is a multivariate method for comparing sets of continuous or categorical variables to each other
* The overwhelming influential factor which in the user features, which is responsible for the correlation of the two datasets is the generation to which the user belongs to. Example: If they were born in the 60s or in the 80s. In contrast, the gender did not play a role at all.
* Hence we can see that we need to target groups and custom make ads to age groups instead of simply looking at variables such as wealth status, where they live etc.

Wiek Polaków i pokolenia
(stan na 30.06.2017)

Źródło: Główny Urząd Statystyczny   NEWSPO!NT

t-SNE dimensions colored by class