# A new experimental quality of life statistic for Flanders using machine learning and sentiment analysis on tweets

**Keywords:** experimental statistics, new data sources, sentiment analysis, social media, official statistics

Authors: Michael Reusens (Statistics Flanders), Marc Callens (Statistics Flanders)

This paper presents the experiments and developments of Statistics Flanders on the creation of a new experimental quality of life statistic by using sentiment analysis on Flemish tweets. This study has been set up to test the use of big data sources like Twitter to measure the sentiment of a population (e.g. perceived quality of life, beyond GDP). Can the combined use of big data, text mining and machine learning provide a valuable alternative for surveys? The study provides a use case for Flemish-language tweets in Flanders (Belgium).

Current quality of life statistics are produced via surveys, which results in infrequently available statistics. Using social media data, daily (and even hourly) measures of quality of life statistics can be produced. This high frequency statistic allows for the immediate monitoring of the impact of events and interventions on the sentiment of the population. However, there are several challenges to be overcome in order to create such a statistic that is accurate, representative of the total population and interpretable.

This paper presents the methodological choices made in the creation of this statistic. We evaluate these methods and provide pathways for further research. More specifically, we will discuss the Twitter data collection and processing, challenges related to sentiment data labelling by human annotators, and the machine learning model structure and performance.