Quality of Life in Flanders: a Comparative Study Using Twitter and Survey Data

Statistics Flanders

May 24th 2023

Nadège Ferket

Sarah Vranken

Overview

1. Problem setting
2. Related work
3. Methodology
4. Results
5. Discussion and Conclusion

KU LEUVEN

1. Problem setting

KU LEUVEN

Importance of research

Survey data:

- Costly

- Time-consuming

- Declining response rate / response burden

- Subject to bias

- Difficult to capture respondents for rare events

Social media data:

- Non-intrusive psychological assessments

- Measure change in public opinion

- Measure issue salience

- Map real-time well-being of a nation

- Eliminates social desirability bias

BUT: Not considered an official indicator

KU LEUVEN

Subjective Well-Being (SWB)

= description of life satisfaction in QoL-research

Subjective Well-Being	
Cognitive valuation of life	**Affective valuation of life**
Overall evaluation of life: How things are these days	Mood, emotion and feeling: How you feel today

- Personal well-being ☺
- Social well-being 👥
- Well-being at work 💼

2. Related work

Faculty of Economics and Business, Data science & Business analytics

KU LEUVEN

Related work and position of our research

Social media data as a replacement of surveys
• (Li et al., 2014) Subjective Well-Being measured through Sina Weibo (Chinese Twitter equivalent)
• (Cao et al., 2018) Public happiness and well-being measured through Twitter

Social media data in addition to surveys
• (Zivanovic et al., 2020) Quality of Life in Bristol, United Kingdom
• (Zhang et al., 2022) Public opinion on social distancing in the Netherlands

KU LEUVEN

Related work and position of our research

Social media data as a replacement of surveys	Social media data in addition to surveys
• (Li et al., 2014) Subjective Well-Being measured through Sina Weibo (Chinese Twitter equ...	• (Zivanovic et al., 2020) Quality of Life in Bristol, United Kingdom
• (Cao et al. being mea...	...on social

Our research

- Insights for national statistics bureaus
- Handling difficulties when working with social media data
- First TSA research on the Quality of Life of the Flemish population
- Sentiment analysis for the Flemish language

KU LEUVEN

3. Methodology

KU LEUVEN

Step 1:
Data collection

Step 2:
Data classification

Step 3:
Subjective Well-Being index

Step 4:
Bias mitigation

KU LEUVEN

Data Collection and Classification

Data collection: Full-Archive Seach API

- Twitter corpus: 153 771 Flemish tweets
  - 10-15 keywords per SWB domain
  - Timeframe: 2014-2018
  - Language: Dutch
  - Geolocation: Belgium

Data classification: 'vlaams-twitter-sentiment-model'

- Developed by Statistics Flanders

- RobBERT model

KU LEUVEN

Subjective Well-Being Index

$$emo_{tr} = \frac{\%positive}{\%positive + \%negative} \in [0,1] \; \forall \text{ year, region, SWB domain}$$

→ Spatial + Temporal conclusions

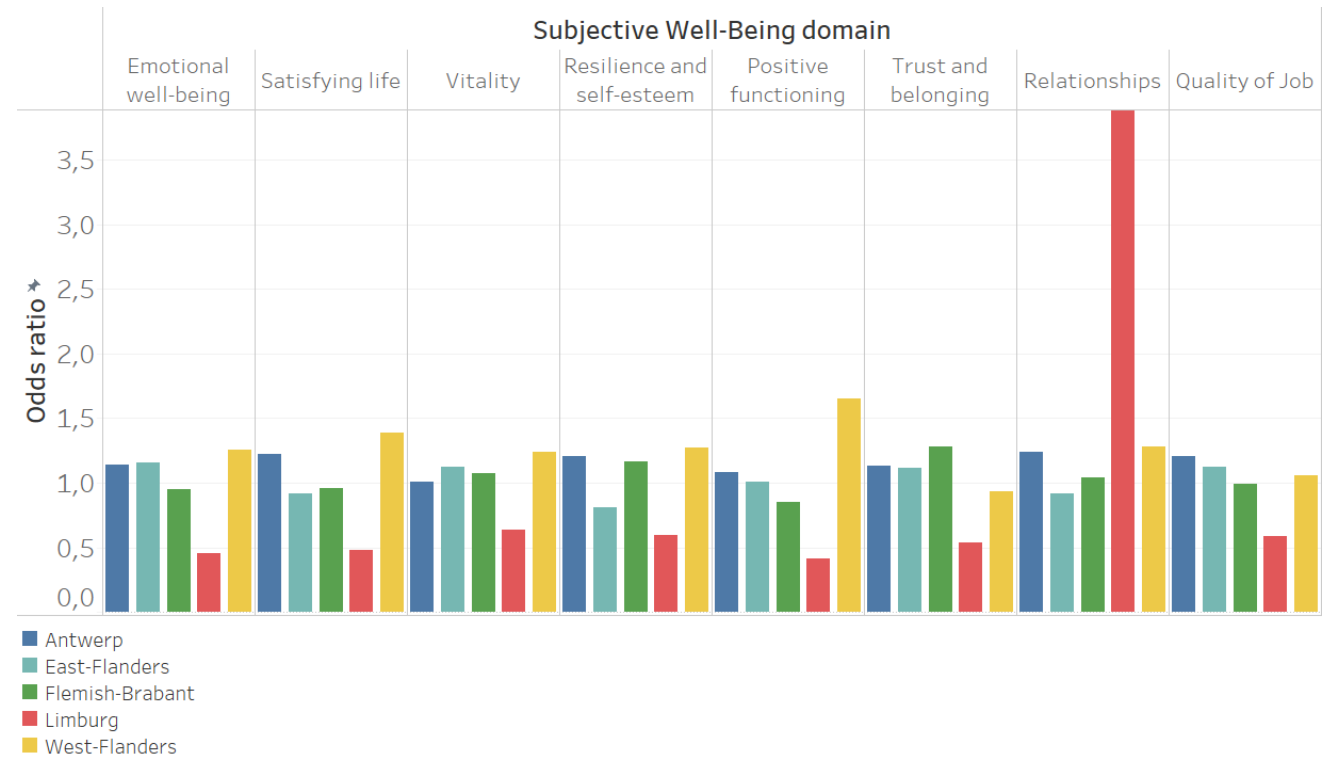| >>>Positive	>Positive	>>Negative
Emotional well-being	Resilience and self-esteem	Vitality
Satisfying life	Relationships	Positive functioning
	Quality of Job	Trust and belonging

KU LEUVEN

Selection bias

- Selection-on-outcome bias

- Negativity bias

- **Keyword bias (NEW)**
  - Negatively connoted → Positively connoted words

KU LEUVEN
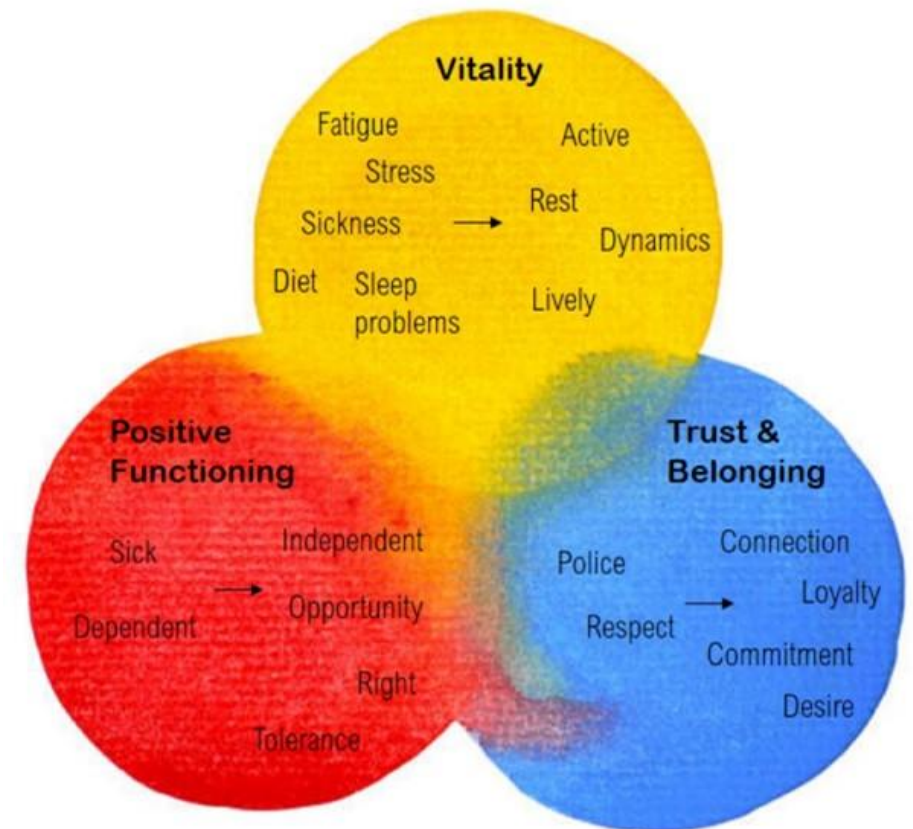
Sampling bias

- **Demographic bias: location**

$$Oddsratio = \frac{T_{rd}/T_d}{P_r/P_{tot}}$$

Migration bias

- **Modal tweet method**
  - Location history of 442 users

  → Types of mobility:

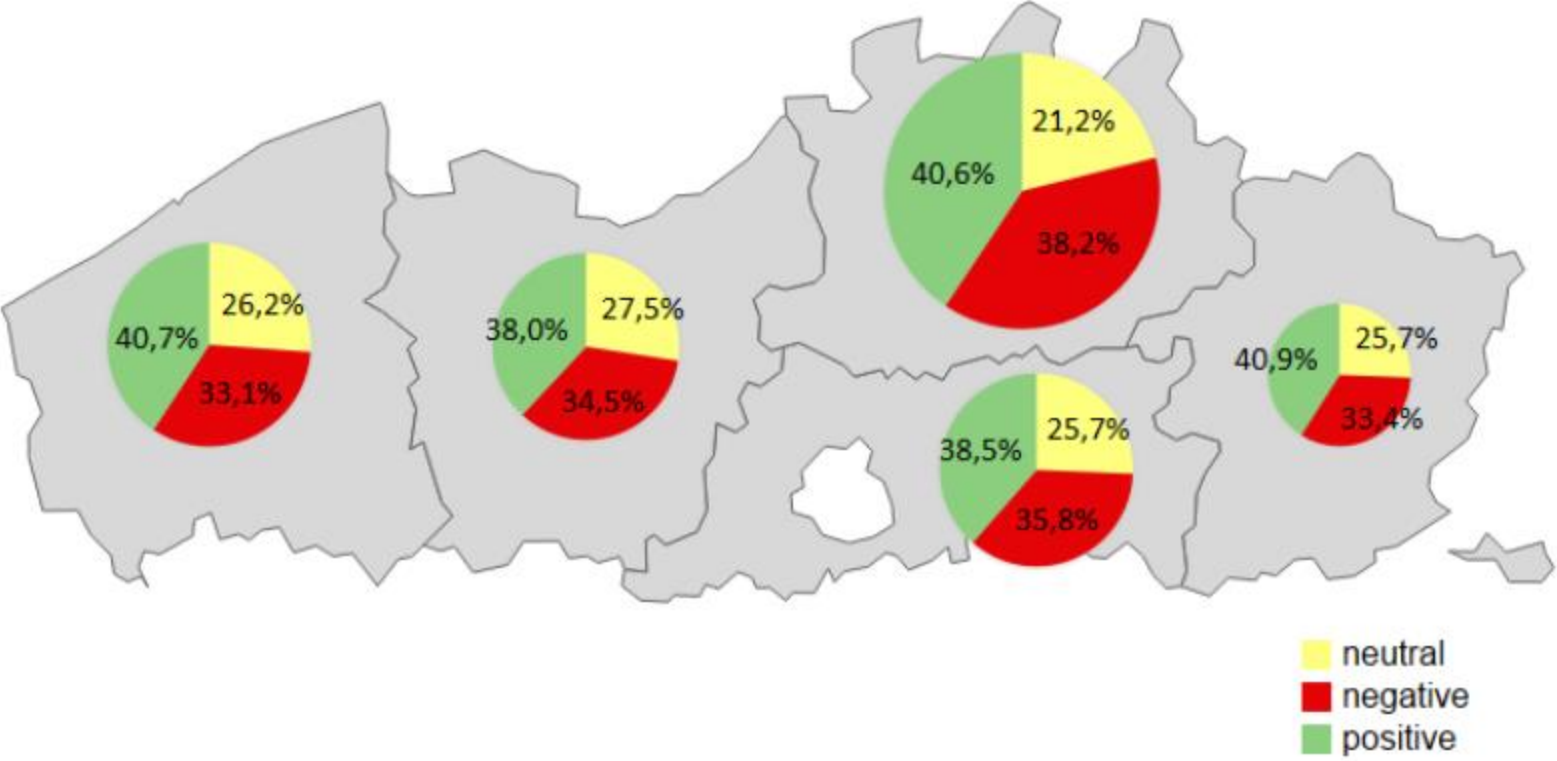| *Student* | *Commuter* | *Beach-Visitor* | *Foreigner* | *Remaining User* |

KU LEUVEN

4. Results

Faculty of Economics and Business, Data science & Business analytics

KU LEUVEN

Results

Faculty of Economics and Business, Data science & Business analytics

KU LEUVEN

Twitter data compared to SCV survey

Wilcoxon signed rank test:

Subjective Well-Being domain	Topic SCV-survey
Emotional well-being	*Zich zorgen maken*
Vitality	*Actieve en passieve sportparticipatie*
Satisfying life	*Algemene tevredenheid*
Resilience and self-esteem	*Solidariteit en eigenbelang*
Positive functioning	*Frequentie sociale contacten*
Trust and belonging	*Algemeen vertrouwen*
Relationships	*Beoordeling sociale contacten*
Quality of job	*Tevredenheid met werk*

KU LEUVEN

Twitter data compared to OECD survey

Life satisfaction Belgium 6.8/10

Cross-country comparison

More fine-grained insights:
- Quality of Job ☺ Limburg, increasing over years
- Relationships ☺ Limburg and West-Flanders

KU LEUVEN

5. Discussion and Conclusion

KU LEUVEN

Limitations and Future research

Limitations

- Limited size of Twitter corpus

- Timeframe 2014-2018 due to change of methodology by Statistics Flanders

Future research

- Design a fixed, balanced dictionary in collaboration with QoL experts and psychologists

- Integrate mitigation proposals in future QoL research

- Exclude 'power users'

KU LEUVEN

Conclusion

- Novel research, familiarizing national statistics bureaus with using Twitter data in QoL research
- Twitter data can be used as a complement to survey data, not as a proxy

Our research:

Selec-
tion
bias
5.4 5.8

Samp-
ling
bias
6.2

Statistics Flanders: 7.1

OECD: 6.8

KU LEUVEN

Questions?

Faculty of Economics and Business, Data science & Business analytics

KU LEUVEN

Bibliography

- Zivanovic, S., Martinez, J., Verplanke, J.: Capturing and mapping quality of life using Twitter data. GeoJournal 85(1), 237–255 (2020)

- Salvatore, C., Biffignandi, S., Bianchi, A.: Social media and twitter data quality for new social indicators. Social Indicators Research 156, 601–630 (2021) https: //doi.org/10.1007/s11205-020-02296-w

- Barber´a, P.: Less is more? how demographic sample weights can improve public opinion estimates based on twitter data. Work Pap NYU (2016)

- Iacus, S.M., Porro, G., Salini, S., Siletti, E.: An Italian Composite Subjective WellBeing Index: The Voice of Twitter Users from 2012 to 2017. Social Indicators Research 161, 471–489 (2022) https://doi.org/10.1007/s11205-020-02319-6

- Delobelle, P., Winters, T., Berendt, B.: RobBERT: a Dutch RoBERTa-based Language Model (2020)

- Zhang, H., Hill, S., Rothschild, D.: Addressing selection bias in event studies with general-purpose social media panels. Journal of Data and Information Quality 10 (2018) https://doi.org/10.1145/3185048

Faculty of Economics and Business, Data science & Business analytics

KU LEUVEN

Bibliography

- Park, C.S.: Applying "negativity bias" to twitter: Negative news on twitter, emotions, and political learning. Journal of Information Technology and Politics 12, 342–359 (2015) https://doi.org/10.1080/19331681.2015.1100225

- Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., Rosenquist, J.N.: Understanding the Demographics of Twitter Users

- Armstrong, C., Poorthuis, A., Zook, M., Ruths, D., Soehl, T.: Challenges when identifying migration from geo-located twitter data. EPJ Data Science 10 (2021) https://doi.org/10.1140/epjds/s13688-020-00254-7

- StatistiekVlaanderen: Algemene levenstevredenheid. https://www.vlaanderen. be/statistiek-vlaanderen/gezondheid-en-welzijn/algemene-levenstevredenheid. [Online; accessed 18-April-2023] (2022)

- OECD: The better life index. https://www.oecdbetterlifeindex.org/countries/ belgium/. [Online; accessed 4-November-2022] (2022)

Faculty of Economics and Business, Data science & Business analytics

KU LEUVEN