## Urban Well-Being Through Data: A Pittsburgh Study

By: Becky Liu, Krish Mawalkar, Serena Yao Advisors: Joel Greenhouse Client: Gabriella Gonzalez, Kristen Kurland

## INTRODUCTION

This study investigates multiple dimensions of quality of life in downtown Pittsburgh by developing and analyzing quantitative metrics related to housing, public safety, and people's sentiment. We set out to test whether downtown Pittsburgh really "feels" as unsafe as the stereotype suggests - and, if so, which of our indicators drive that perception most strongly.

## **Research Questions:**

Is there a difference in perception of downtown Pittsburgh after 2022-H1?

Can we create a measure of downtown Pittsburgh's quality of life? 2. Workstream 1: Time-Series Analysis & Forecasting (ARIMA) Workstream 2: Principal Component Analysis (PCA)



Importance of this variable: we use weekly averages of this sentiment ratio variable (from 2018 to 2024) for our time-series analysis. A ratio value of 1 implies a neutral amount of sentiment, a value greater than 1 implies a more negative sentiment, and a value less than 1 implies a more positive sentiment, for any given weekly data point. This allows us to track whether sentiment is increasing or decreasing over time.

*Positive*<sub>*t*</sub> = mean(joy, love, surprise) in week *t*.

## **WORKSTREAM 1: Time-Series Anal**

We use autoARIMA, a time-series model, to fit data from to 06-2022. The 'auto' part of the autoARIMA model sele optimal parameters to best forecast the sentiment ratio var We then assess whether the true sentiment ratio values after training period lay outside the norm of what is predicted. autoARIMA generated an ARIMA (1, 0, 0) Model, equat AR(1), or a model with (only) an autoregressive component We then predict on the period after 06-2022 until 2024-09 our *true* sentiment ratio variable measurements end).

Why does this model matter? If the true data lies outsid

# analyses without implying casualty.



	METHODS
ysis	<b>WORKSTREAM 2: Principal Compone</b>
m 06-2018 lects the ariable.	<b>Construct PCA Model and Extract Loadings</b> : we take our dataset, including shelter usage, unsheltered counts, jail book sentiment attitudes, and feed it into a single Principal Compo
fter the	Image: Constrained state Image: Constrained state Principal
ting to an ent. O (where	HousingSentimentPublic SafetyCompute six-month means: we bin the weekly data into have
9 (where	calculate $\bar{x}_{j}^{(6 \text{ mo})}$ for each variable j. <b>Form six-month Composite Scores:</b> we extract the variable
le the <i>fistically</i>	3 much each raw metric contributes to that component k's dire

Our methods demonstrate a flexible approach for summarizing and monitoring complex urban well-being trends over time. Our ARIMA-based forecasts suggest that downtown Pittsburgh's negativity-to-positivity ratio has remained relatively stable just below 1.0 threshold through mid-2022, with recent data showing a modest upward trend in negative sentiments. Therefore, our hypothesis that people were having more negativity needs more inspection. Meanwhile, PCA confirms that sentiment, public-safety, and housing form three distinct dimensions of variance, providing a solid framework for future

### ent Analysis

full, combined weekly tings/releases, and nent.

al Component Analysis

alf-year intervals and

e loadings  $\ell_{i,k}$  that tell us how ection. Using the formula:

The first two components that together explain 66 % of the total

PCA biplot neatly partisan three quality-of-life domain — Shelter & Homelessness

Sentiment vectors lie on the same line but in opposite directions, and switch sides between PC 1 & PC 2.

PCA identifies the orthogonal directions of greatest variance; it does not imply any causal linkage

Serves as a data-reduction method

## REFERENCES

Kurland, K. (2024, September 13). Quality of life in downtown Pittsburgh study [Slide show; Box]. Carnegie Mellon University, Heinz College.