Entry Name: UL-Marxen-Sterzel-Gantz-MC3
VAST Challenge 2022
Challenge 3

Team Members:
Michal Sterzel, University of Luxembourg & UPC Universitat Politècnica de Catalunya, michal.sterzel.001@student.uni.lu
Thomas Gantz, University of Luxembourg & Sorbonne Université, thomas.gantz.001@student.uni.lu
Jan Marxen, University of Luxembourg & Sorbonne Université, jan.esquivel.001@student.uni.lu

Student Team: YES

Tools Used:
Backend: Python 3.11, Flask, Pandas, NumPy, Scikit-learn (K-Means clustering), Pytest

Frontend: React 18, D3.js v7, Axios, TailwindCSS

Infrastructure: Docker, Docker Compose, Nginx

Approximately how many hours were spent working on this submission in total?

Approximately 45 hours (≈15 hours per team member)

🎬 **Video:** https://github.com/janmarxen/VAST-challenge/submission/video.mp4

📁 **Repository:** https://github.com/janmarxen/VAST-challenge

📦 **Submission Release:** https://github.com/janmarxen/VAST-challenge/releases/tag/submission-v1.0.0

Running the code: Please follow the **README** included in the submission release.

Use visual analytics to analyze the available data and develop responses to the questions to be provided. In addition, prepare a video that shows how you used visual analytics to solve this challenge.

Questions

1 - Over the period covered by the dataset, which businesses appear to be more prosperous? Which appear to be struggling? Describe your rationale for your answers. Limit your response to 10 images and 500 words.

Answer:

Approach and Starting Point. We analyzed the 32 pubs and restaurants in Engagement using the provided data over 15 months (with aggregation). Our entry point is the global filter panel, allowing selection of venue type, specific businesses, customers, and metrics (visits or spending). We define *prosperity* as sustained or growing revenue and visits, and *struggling* as declining. We assume checkin spending accurately reflects business revenue.

Business Overview. The overview (Figure 1) aggregates key metrics: over 800,000 visits generating modest per-visit spending (under $10). This dashboard establishes baseline context before drill-down analysis.

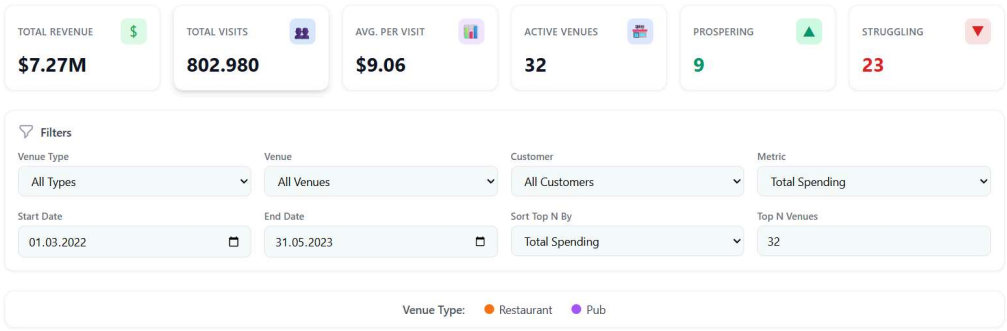
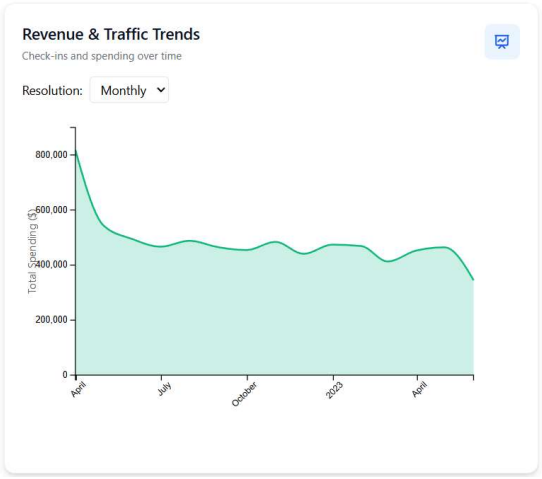


Figure 1: Overview of all businesses showing total revenue, visits, and average spending.

Temporal Trends. The trend visualization (Figure 2) reveals a declining trajectory in total spending, with notable drops around April in both years. At daily resolution (Figure 3), strong weekend peaks emerge: a consistent oscillating pattern useful for distinguishing cyclical behavior from structural decline.



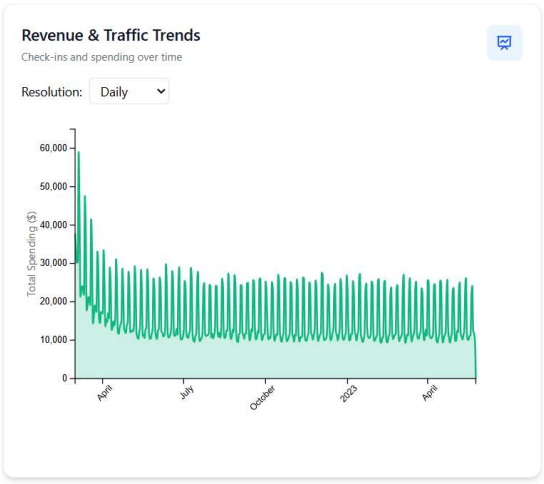
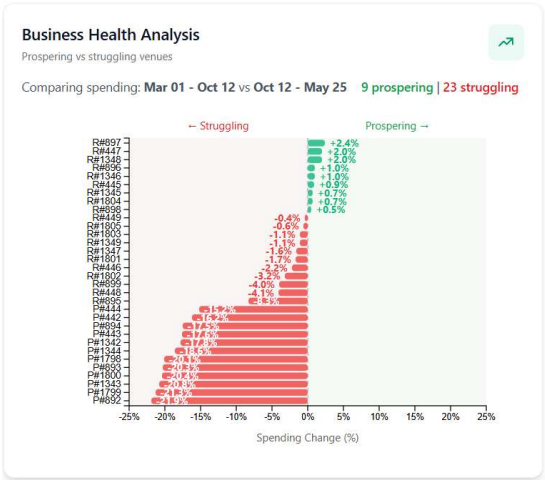


Figure 2 (left): Monthly revenue trends. Figure 3 (right): Daily pattern showing weekend peaks.

Growth Analysis. The health visualization (Figures 4 and 5) compares the first and second halves of the dataset. Switching between spending and visit metrics reveals correlation: businesses with declining visits also show declining revenue. Approximately one-third of businesses show slight growth, one-third slight decline, and one-third significant decline. This indicates heterogeneous business health rather than uniform trends.



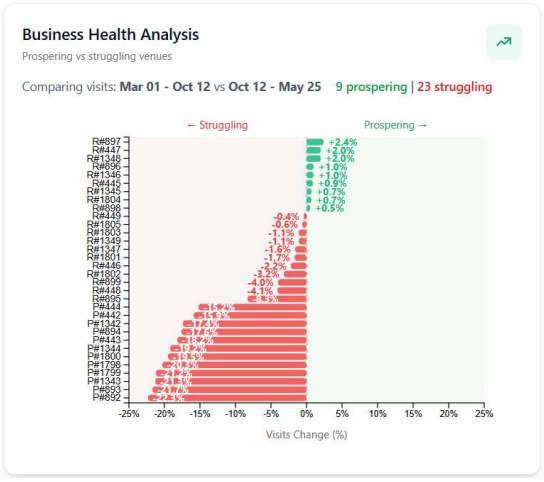
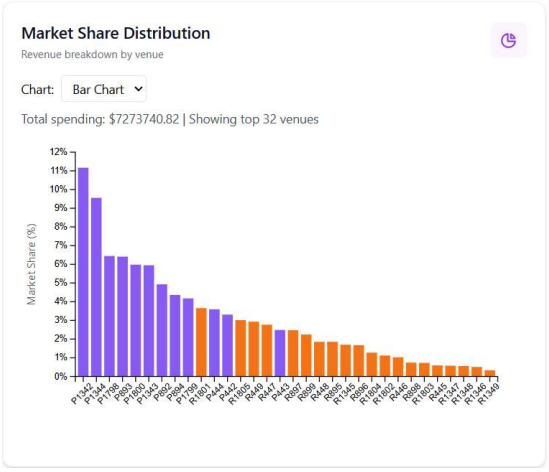


Figure 4 (left): Spending growth by business. Figure 5 (right): Visit growth by business.

Market Concentration. The market share bar plot (Figure 6) reveals that two pubs (P#1342, P#1344) capture approximately 20% of total spending. The scatterplot (Figure 7) encodes visits, spending, and market share simultaneously; linked brushing enables cross-chart exploration. Pubs dominate both metrics, likely due to higher per-visit spending compared to restaurants.



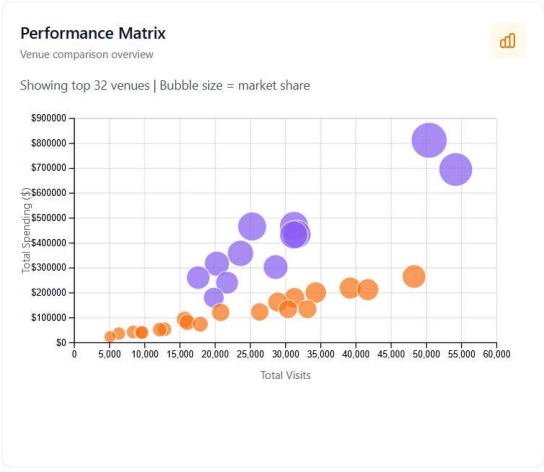
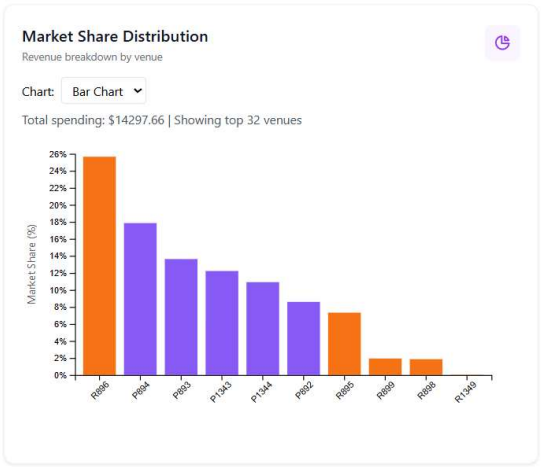


Figure 6 (left): Market share distribution. Figure 7 (right): Performance scatterplot with size encoding market share.

Individual Customer Patterns. Filtering to a single high-frequency customer (Figures 8 and 9) reveals individual preferences diverging from aggregate trends. This customer's spending concentrates on a favorite restaurant (R#896, 26%) and an emerging pub (P#894, +27.7% growth), while completely abandoning another venue (R#1349) in the second half. This demonstrates how micro-level analysis can surface early signals of shifting preferences.



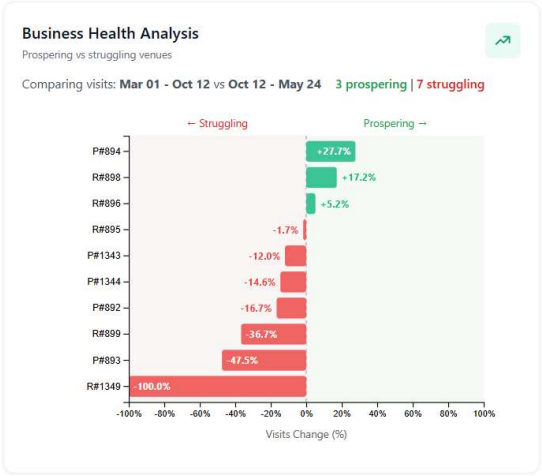


Figure 8 (left): Individual customer spending distribution. Figure 9 (right): Customer-level growth analysis.

Design Rationale. The visualization progresses from overview (Figure 1) to temporal filtering (trend charts) to individual detail (customer analysis). Coordinated views with hover-linking support cross-chart exploration. The split-period comparison directly addresses "growing or shrinking" by quantifying change rather than relying on visual trend interpretation alone.

Conclusion. Prosperous: Pubs in general outperform restaurants in both volume and per-visit revenue, with P#1342 and P#1344 dominating market share. **Struggling:** Even the top performers decline significantly in the second half; approximately one-third of all businesses show substantial drops in both visits and spending. **Overall trend:** The majority of businesses are shrinking, with aggregate spending declining over the 15-month period.

2 - How does the financial health of the residents change over the period covered by the dataset? How do wages compare to the overall cost of living in Engagement? Are there groups that appear to exhibit similar patterns? Describe your rationale for your answers. Limit your response to 10 images and 500 words.

Answer:

We looked at resident financial health from three angles: where financial stress concentrates in the city, which household groups are closer to (or past) break-even, and how income/expenses/inequality evolve over time.

Geographic Financial Health. The building-level heatmap shows scattered “red pockets” of low savings across Engagement (Figures 1 & 2). Early vs. later snapshots suggest these pockets are persistent rather than spreading, pointing to chronic neighborhood-level strain rather than a city-wide decline.

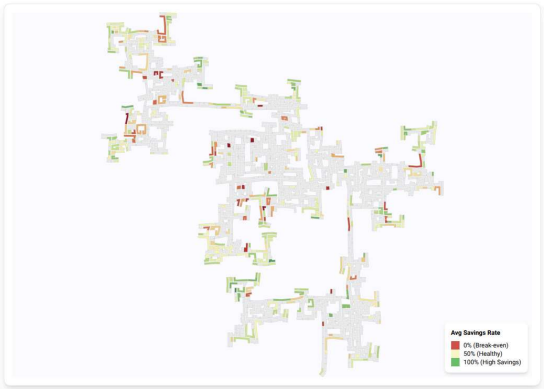


Figure 1 (left): Geographic financial health at the start of the period. Figure 2 (right): Geographic financial health later in the dataset.

Wages vs. Cost of Living. The Living Gap scatter plot (Figure 3) puts wages against total living costs and makes the break-even line an immediate reference point. Most residents sit on the surplus side, but the dense band near break-even suggests many households have only a thin “buffer” month to month (Figure 3). When filtering by household type, the “with children” group shifts right (Figures 4 & 5), indicating higher living costs; the main takeaway is the higher cost burden rather than a fundamentally different income distribution.

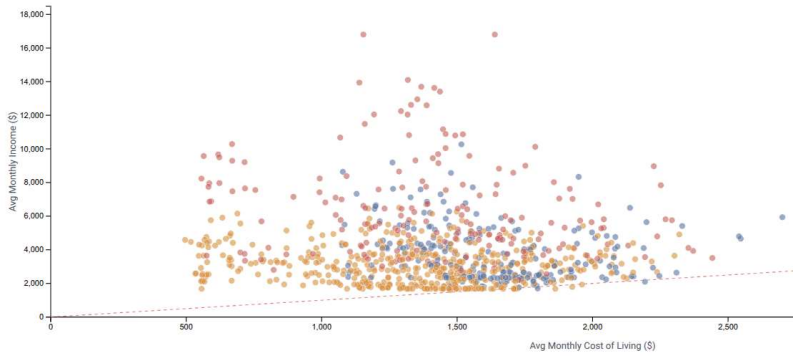


Figure 3: All residents.

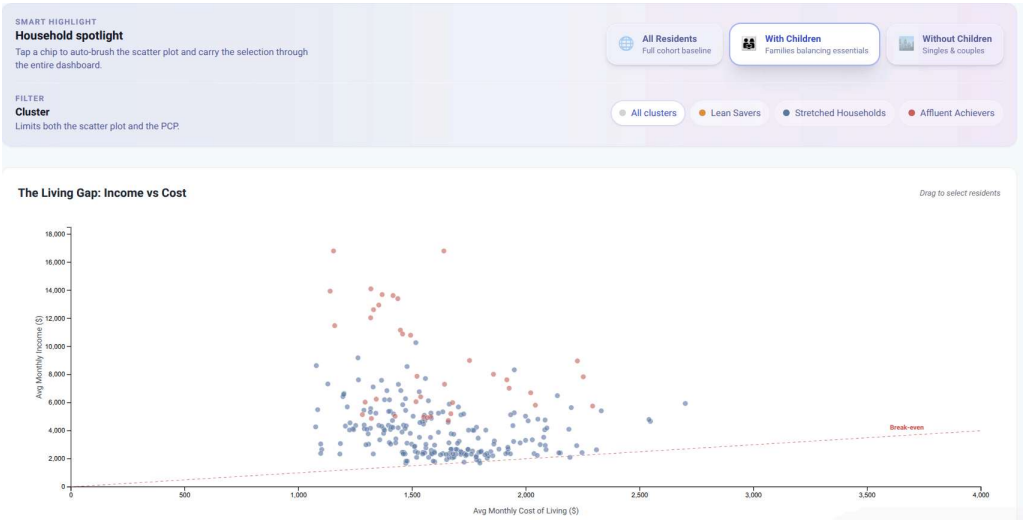


Figure 4: Families with children.

Figure 5: Households without children.

Cluster Patterns. Using k-means, we get three recurring "financial personas": **Affluent Achievers** (very high incomes and high savings), **Stretched Households** (lower incomes with tighter budgets and lower savings capacity), and **Lean Savers** (modest-to-moderate incomes with a different cost/income balance than Stretched). These personas overlap; labels summarize dominant tendencies rather than strict rules. In April 2022 (the month shown in Figures 3–7), the medians (Income / Cost / Savings Rate) were: **Affluent** $5,756 / $1,419 / 76.6%, **Stretched** $2,869 / $1,405 / 51.0%, and **Lean** $3,352 / $1,586 / 54.5%. The clearest separation signals in demographics are **having kids** (η^2 83.1%), **graduate education** (η^2 72.0%), **household size** (η^2 61.9%), and **income** (η^2 38.0%). Figures 6–7 illustrate that our visualizations can be filtered by cluster: we show the **Affluent Achievers** profile in the parallel coordinates plot (Figure 6) and their Living Gap scatter distribution (Figure 7).

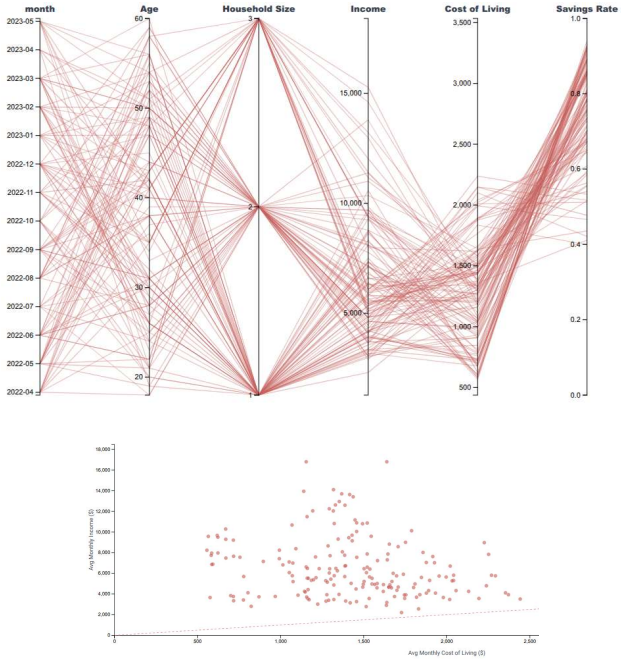


Figure 6 (left): Parallel coordinates profile for the Affluent Achievers cluster. Figure 7 (right): Living Gap scatter filtered to the same cluster.

Expense Trajectories & Inequality. The stacked area chart (Figure 8) tracks average monthly income (dashed line) against the composition of expenses and highlights a visible drop in both income and expenses in the final months (April–May 2023), consistent with a slowdown. The inequality timeline (Figure 10) shows relatively stable income inequality but more volatile savings inequality.

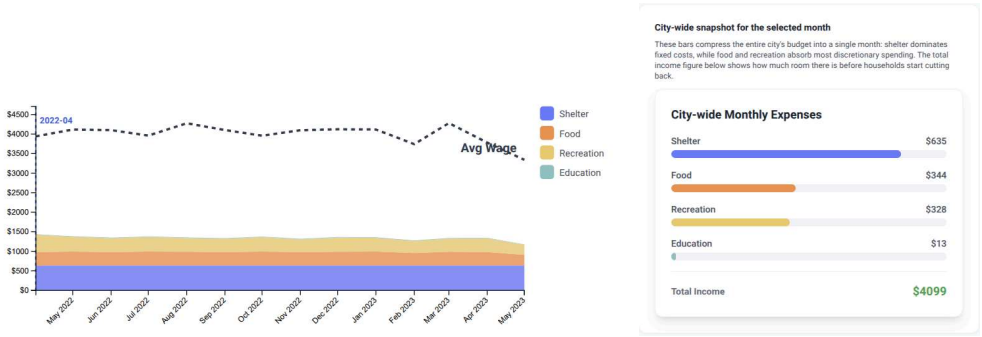


Figure 8 (left): Income vs. stacked expenses over time. Figure 9 (right): City-wide expense breakdown.

*Figure 10: Gini coefficient for income (blue) and savings rate (red) over time. The formula G = (2 * sum(i * yi)) / (n * sum(yi)) - (n+1)/n measures inequality from 0 (equal) to 1 (unequal).*

Conclusion. Financial health in Engagement is stable on aggregate but uneven: the strongest, most consistent strain signal comes from household structure, especially families with children sitting near break-even (Figures 4–5). Cost pressures matter most: **cost of living** is the strongest predictor of savings rate (ΔR^2 0.828), followed by **income** (0.408), **household size** (0.376), and **having kids** (0.127). Persistent geographic hotspots (Figures 1–2) suggest targeted neighborhood interventions could have outsized impact.

Notes on metrics. $\mathbf{R^2}$ measures how much variance in savings rate a model explains: $\mathbf{R^2 = 1 - SSE/SST}$. The reported $\mathbf{\Delta R^2}$ values indicate each predictor's additional (incremental) contribution when added to the multivariate model. $\boldsymbol{\eta^2}$ is an effect-size for group separation: $\boldsymbol{\eta^2 = SSB/SST}$, i.e., the share of variance in a variable explained by cluster membership.

3 - Describe the health of the various employers within the city limits. What employment patterns do you observe? Do you notice any areas of particularly high or low turnover? Limit your response to 10 images and 500 words.

Answer:

We assessed employer health within the city of Engagement by analyzing workforce stability, turnover intensity, job-to-job mobility, and short-term changes in employee counts. Our analysis is constrained by the dataset: observed job transitions occur primarily during March–April 2022, with no additional job changes recorded afterward. Rather than indicating ongoing stability, this reflects the limits of the available data and is explicitly shown in the interface.

Workforce Instability and Turnover Concentration

The Employer Turnover Ranking reveals strong heterogeneity in workforce change intensity across employers. The visualization focuses on the upper tail of the distribution, highlighting employers with extreme workforce churn during the observed period. Several employers exhibit turnover rates far exceeding 100%, driven by repeated employee exits relative to small average headcounts.

Figure 1: Employer Turnover Ranking highlighting extreme turnover values.

These extreme values do not indicate growth, but rather intense instability: employers ranking highest by turnover also show high quit counts and strongly negative net workforce change. This pattern suggests persistent outward movement of workers rather than balanced hiring and retention. In contrast, many employers experience little to no workforce change and therefore do not appear in the ranking, indicating comparatively stable employment structures.

Stability, Tenure, and Employer Risk Profiles

The Employer Stability Overview combines turnover, average tenure, and workforce size into a single view. Employers cluster into two dominant regimes:

- **High-risk employers** with high turnover and low average tenure, indicating short job durations and frequent exits.
- **Stable employers** with low turnover and long tenure, reflecting sustained employment relationships.

Employer size alone does not determine stability: both small and large employers appear in each regime. However, high-risk employers with larger workforces amplify their impact on city-wide employment volatility.

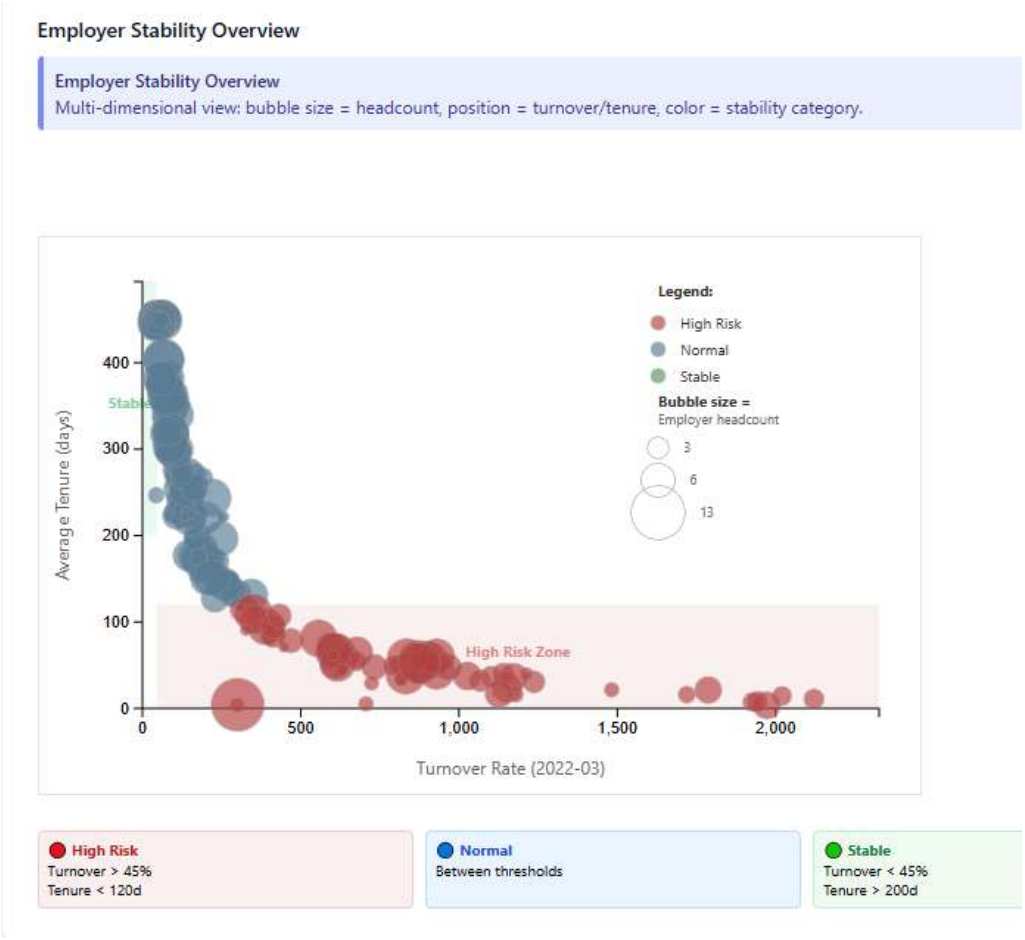


Figure 2: Employer Stability Overview showing the clustering of employers into stability regimes.

Worker Mobility Between Employers

The Job Flow Sankey diagram illustrates directional worker movement between employers. A small number of employers act as major sources of outgoing labor, while others consistently receive workers. This asymmetry shows that workforce instability propagates through the employer network, rather than remaining isolated within individual organizations.

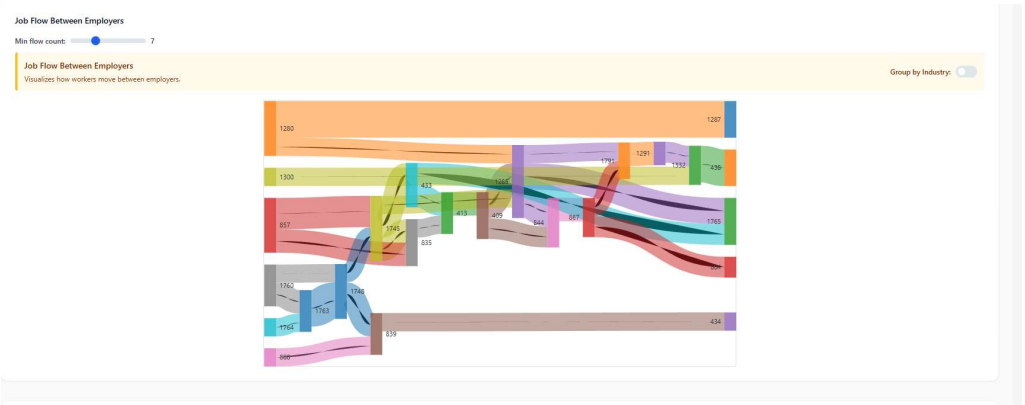


Figure 3: Job Flow Sankey diagram illustrating directional worker movement between employers.

Short-Term Workforce Fluctuations

The Employee Count Trends visualization shows that workforce changes occur in short, concentrated bursts during the limited period with recorded transitions. Employers experiencing high turnover also exhibit abrupt rises and drops in employee counts, reinforcing the link between turnover intensity and workforce volatility.

Figure 4: Employee Count Trends showing workforce fluctuations during periods of high turnover.

Conclusions

Overall, employer health in Engagement is highly uneven. A minority of employers exhibit acute workforce instability, characterized by high turnover, low tenure, and strong outward job flows. Most employers, however, show limited workforce movement and relatively stable employment relationships during the observed period. The concentration of instability among specific employers suggests localized labor market stress rather than city-wide systemic disruption.