Visual Analytics Project

Yanmu Ding Ningzhen Rao Di Yu

Singapore Management University

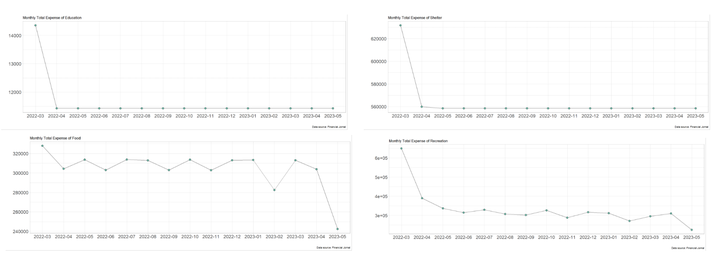
Introduction

Anticipating rapid growth, the city of Engagement, Ohio USA is doing a participatory urban planning exercise to understand the current state of the city and identify opportunities for future growth. About 1000 representative residents in this modest-sized city have agreed to provide data using the city's urban planning app, which records the places they visit, their spending, and their purchases, among other things. From these volunteers, the city will have data to assist with their major community revitalization efforts, including how to allocate a very large city renewal grant they have recently received. Considering the financial status of Engagement's businesses and residents, we need to use visual analytic techniques to address the financial health of the city.

Problem 1

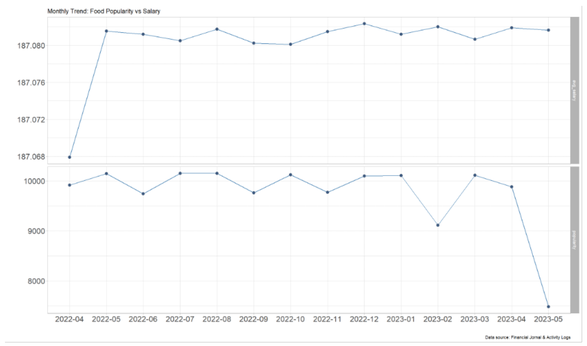
Business Development Situation

Based on people's financial log data that records all spending behaviour in the surveyed period, we can estimate the condition of industries such as education, housing, food, and recreation. To have a general understanding of people's spending capacity, we computed the average monthly salary of the participants, we found that people's monthly income remains stable in the surveyed period. Hence, an overall stable trend in their spending behaviour is presumable. But, on the contrary, all industries except the food industry were having a hard time during the surveyed period. Education and real estate industry had seen a huge decrease of 21% and 11% respectively between March 2022 and April 2022 and remain low ever since. While food industry has been stable between 2022 and 2023 despite a little decreasing trend. Recreation industry was the worst, it was quite prosperous in early 2022 bur has shrunk significantly since March 2022 and was only half of its original size in 2023.



Comparison between Salary Level and Food Popularity

Since our data has specifically labelled restaurant and pubs, we infer that food industry can be a focus of our analysis. Hence, we also computed the monthly popularity of the restaurant and pubs by calculating the total number of hours people spent in there. The result aligned with the monthly spending data, suggested that despite a slight decreasing in the food industry, people are still willing to dine out in restaurant and pubs.



Problem 2

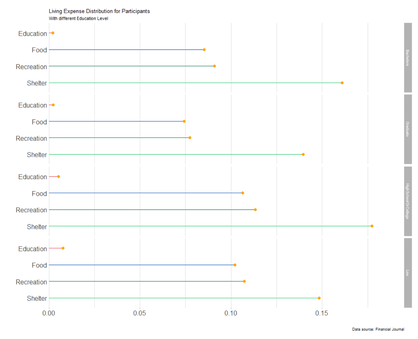
For this problem, we are going to explore the change pattern of financial health over the period covered by the dataset. Besides, we need to compare the overall cost of living in Engagement and find the groups exhibiting similar patterns.

Engel Coefficient

Engel coefficient is a good metric to evaluate the level of living standard for inhabitants. It evaluate the ratio of the food investment taking in the whole wage. We can sampled several participants and compare the differences of the engel index by sparklines.

The Engel Index of the participants

participantId	Min	Max	Average	Engel Index
0	0.102	0.224	0.134	
1	0.143	0.246	0.172	
2	0.159	0.192	0.175	
3	0.164	0.247	0.190	
4	0.176	0.234	0.203	
5	0.672	0.900	0.791	
6	0.696	0.979	0.853	
7	0.303	0.390	0.340	
8	0.566	0.687	0.617	
9	0.100	0.115	0.106	

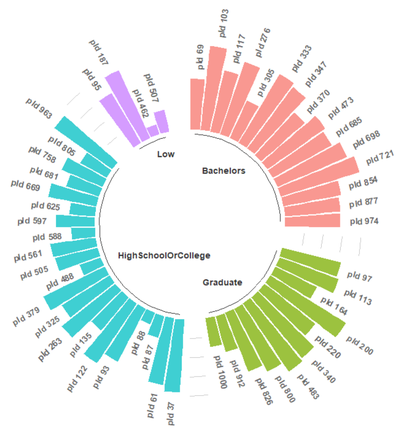


Living Expense

Then we try to explore the expense preference similarity for different groups of participants. For this poster example, we use education level as a criteria to calculate the expense on different factors. One interesting thing is that the higher education level the participant has, the more investment they have on education.

Income Level

For the net income statistics, people with lower education level are easy to find low salary sample. On the other hand, participants with at least bachelor degree tend to have stabler salary than the others.

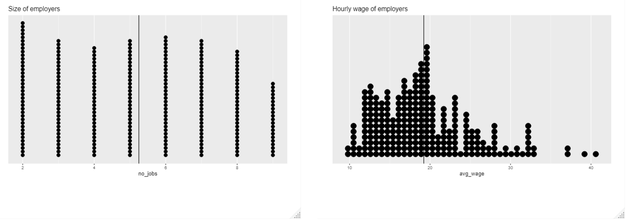


Supplementary Instruction

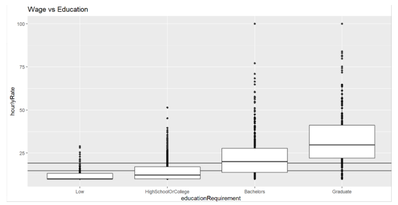
There are more group choices for the participants. To get more interesting patterns, you can visit our visual analytic website to see more details.

Problem 3

We believe that the health of employer can be evaluated by their size(total number of employee), the average number of days their employee works and the medium wage of the employees. According to our analysis, the size of employee is evenly distributed. Indicating that there is no sign of monopoly in the surveyed area. After we specified the industry, the employee size pattern remains the same. All employers in the survey had hired at least one employer during the surveyed period. In 2022, the minimum wage in the US is 7.5 usd per hour, all the employers exceed the min wage in hiring. But the distribution of the wage is skewed to the higher wage. But the high wage participants wasn't more likely to work in restaurant and pubs, in further study, we need to collect more detailed information of employers other than pubs and restaurants.



While we try to compare the wage with the education level required. We found that in general, the higher education level a job requires, the more the hourly rate will it offer. But there are also a lot of cases when the wage is high with low education requirement, or the wage is low with high education requirement. The wage distribution illustrated by the box plots is also sparser for high education jobs compared to low education jobs.



We tried to approach a region's turnover from 2 different aspects: firstly, how to divide the surveyed region? Secondly, people have different places to go in and out of office hours, the level of turnover might be completely different, we should study the turnover of region in different periods of day.

We divided the whole area into 4 different regions, and all following turnover analysis will be based on these 4 regions. According to the general trend (which consider all time), northwest region and the central region are busier than south and east regions. While during office hours (9 am-5pm), northwest region and the central region are busier than south and east regions, whereas out of office hour (19 pm to 6am), northwest region is significantly busier than the other 3 regions, northwest region seems to be the residential area with very high population density.