

Introduction The analysis of this project may provide insight into relationships across quality of life for Durham residents. As well, it may help city administrators make decisions on improving certain areas within Durham and the city's strategic growth plan. The dataset comes from the 2022 City of Durham Resident Survey, an annual survey administered by the ETC Institute. A random sample of households in the city of Durham were mailed a survey, cover letter, and postage-paid return envelope. Questions to the survey included opinions on many facets of Durham's amenities, including quality of schooling, police protection, library access, condition of streets, etc. The original dataset featured 891 unique and independent observations from 226 questions from the survey.

The key objectives of this project are to understand a potential association between quality of life and number of years that a resident has lived in Durham. Furthermore, this project attempts to recognize a relationship within quality of life based on the Durham residents' homeowner status, which may have specific location grouping factors within Durham (Zip Code). From examining previous research from a study of 79 European cities, it was determined that living in a city for increasing thresholds of 5 years result in higher levels of resident satisfaction (Weziak-Bialowolska 2016). Analysis of this dataset will see if similar conclusions can be drawn.

Relevant Variables The response variable of interest was **quality**, a response to part 6 of question 3 (Q3-6) from the 2022 Resident Survey. This variable was the response to the residents' overall quality of life in Durham, with response options of (1) "very dissatisfied", (2) "dissatisfied", (3) "neutral", (4) "satisfied", and (5) "very satisfied". The other variables included in the project analysis were **ID** (identification number of the resident), **years**, a response to question 27 (Q27) of the survey indicating how many years the resident has lived in Durham, **income**, a response to question 33 (Q33) of the survey indicating total annual household income with response options of (1) under \$30,000, (2) \$30,000 - \$59,999, (3) \$60,000 - \$99,999, or (4) \$100,000 or more, **house**, a response to question 30 (Q30) of the survey indication if you (1) own or (2) rent your house, and **Zip**, the Zip Code of the residents' address within Durham.

For data cleaning, all rows of relevant variables that were unreported or reported as a value of 9 (N/A) were removed from analysis. Certain variables used in analysis (**quality**, **years**, **income** and **house**) were renamed as such from their original question number, which allowed for more comprehensive model interpretation. The variables **quality**, **income**, and **house** were mutated to factors to account for their leveled responses, **Zip** was mutated to a factor to be used as a random intercept grouping variable in the mixed model, and **years** was mutated to be a continuous numerical variable. This left 759 observations.

Methods To assess the first objective of understanding a possible association between quality of life and length of time a resident has lived in Durham, an ordinal logistic regression model was included. This is because while the **quality** response variable has natural order of responses numbered 1-5, the intervals between levels are not guaranteed to be equal. For example, the difference between "very dissatisfied" and "dissatisfied" (1-2) may not be the same as the difference between "neutral" and "satisfied" (3-4). The proportional odds assumption was met, with the associated summary table output shown in Table 1. Seeing as quality of life is not solely dependent on the number of years living in a city, a second ordinal model was included to understand the interaction effect that years lived in Durham has with household income. The proportional odds and multicollinearity assumption was met, with the associated summary table output shown in Table 2. To assess the second objective of potential relationship based on home ownership status and location of residence, a cumulative link model was included. This allowed for understanding the ordinal response variable **quality** based on the interaction effect of years lived in Durham and home ownership status with the random effect of Zip Code. The multicollinearity assumption was met, with associated summary table outputs shown in Tables 3 and 4.

Results The results from the first ordinal logistic regression model suggests that, holding else constant, each additional year a resident has lived in Durham is associated with 0.993 ($e^{-0.007}$) times the odds of being in a higher satisfactory category. This means that the odds of reporting greater satisfaction decrease by 0.7% for an additional year a resident lives in Durham. With a p-value of 0.037 which is less than the $\alpha = 0.05$ significance level, this provides statistically significant evidence that the number of years a resident lives in Durham is associated with their reported quality of life level (Table 1). Adding household income level as an interaction effect proved to not be statistically significant as all the **years x income** p-values were greater than the $\alpha = 0.05$ level. Holding else constant, residents in income level 3 (\$60,000 - \$99,999) are predicted to be 1.976 ($e^{0.681}$) times more likely to report a higher satisfactory category compared to the baseline income level 1 (under \$30,000). Holding else constant, residents in income level 4 (\$100,000 or more) are predicted to be 2.184 ($e^{0.781}$) times more likely to report a higher satisfactory category compared to the baseline income level 1 (under \$30,000). With p-values of 0.050 and 0.019 respectively, which is less than the $\alpha = 0.05$ level, this provides statistically significant evidence that income level is associated with a residents' reported quality of life level (Table 2).

The results from the cumulative link mixed model show that none of the fixed effects or the interaction effect between number of years lived in Durham and home ownership status were statistically meaningful in its association with quality of life after accounting for the Zip Code. However, the standard error of 0.300 for Zip Code suggests the underlying satisfaction intercept varies slightly by neighborhood (Table 3 and 4).

Discussion Based on the results, the models were not very effective at explaining the relationships between quality of life in Durham and the data from the residents. The large p-values caused few interpretable predictor variables, and the interpretation for **years** in Table 1 conflicted with previous research. A potential issue with the data collection is that only those interested in completing the survey were the ones recorded, it was not a mandatory response. This could lead to biased results. Asking these types of questions within the census, or another form of required response would potentially increase reliability of the models and interpretability. Future work could analyze much more of the factors that were within the survey among a residents' lifestyle in understanding quality of life (access to schooling, public safety, affordability, maintenance, etc).

Table 1: Ordinal Logistic Regression Summary

	Value	Standard Error	t-value	p-value
years	-0.007	0.004	-2.084	0.037
1 2	-3.775	0.245	-15.393	<0.001
2 3	-1.861	0.136	-13.687	<0.001
3 4	-0.752	0.118	-6.397	<0.001
4 5	1.477	0.13	11.34	<0.001

Table 2: Ordinal Logistic Regression (Years \times Income Interaction)

	Value	Standard Error	t-value	p-value
years	-0.005	0.007	-0.675	0.500
income2	0.336	0.354	0.947	0.344
income3	0.681	0.347	1.962	0.050
income4	0.781	0.333	2.341	0.019
years:income2	0	0.01	-0.037	0.970
years:income3	0	0.011	-0.024	0.981
years:income4	0.005	0.011	0.446	0.656
1 2	-3.235	0.342	-9.471	<0.001
2 3	-1.305	0.274	-4.766	<0.001
3 4	-0.171	0.267	-0.64	0.522
4 5	2.106	0.281	7.485	<0.001

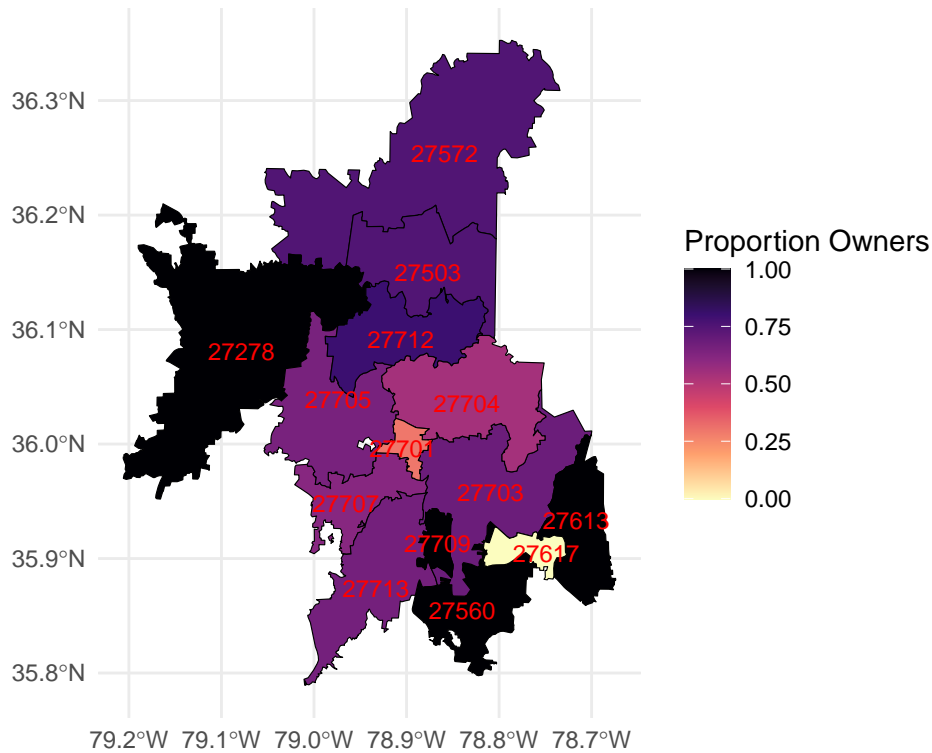
Table 3: Fixed Effects: Mixed Ordinal Logistic Regression

Term	Estimate	Standard Error	t-value	p-value
1 2	-3.795	0.283	-13.407	<0.001
2 3	-1.864	0.196	-9.521	<0.001
3 4	-0.737	0.183	-4.018	<0.001
4 5	1.531	0.193	7.945	<0.001
years	-0.006	0.005	-1.218	0.223
house2	-0.004	0.231	-0.017	0.986
years:house2	-0.004	0.008	-0.589	0.556

Table 4: Random Effects Summary (Zip Code)

	Group	Effect	Variance	Standard.Error	Groups
(Intercept)	Zip	Intercept	0.09016	0.3003	14

Figure 1: Prevalence of Home Ownership by ZIP Code in Durham
Based on 2022 Resident Survey Data



Despite not having a largely significant impact on reported quality of life, the heatmap shows the proportion of homeowner status based on the specific Zip Codes within Durham.

Sources:

- <https://live-durhamnc.opendata.arcgis.com/documents/2459a4fcb09345b1a6135321bf91eac9/about> - data source
- <https://www.sciencedirect.com/science/article/pii/S0264275116301330> - Weziak-Bialowolska research
- <https://r-graph-gallery.com/heatmap> - heat map guidance
- <https://search.r-project.org/CRAN/refmans/tigris/html/zctas.html> - download zip code tabulation area (ZTCA)