PA211: Referral methods, client information, and their impact on unmet client needs.

Cassandra M. Clemency

Haverlack School of Business, Slippery Rock University

ECON-318: Introduction to Econometrics

Under the supervision of Dr. Xintong Wang

December 08, 2025

Abstract

This paper analyzes and documents the relationships between several relevant client statuses, and the unmet needs of those clients: Pennsylvania residents who contact PA211 for assistance. My literature review suggested pronounced disparities in urban-rural food security, and methodological challenges created by gendered behavioral differences and mental health problems among clients. Data provided by PA211 uses a three-tier structure, containing unique IDs to differentiate client interactions from referrals, categorizations of services provided, as well as a numbering system to allow counting of needs. Unmet needs, or services which clients were unable to find assistance with after contacting PA211, intersect in varying ways with client demographic information, repeat-client status, location, and timing. Using multiple and logistic regression, a stacked bar chart, and a correlation matrix, I analyze and visualize these relationships. Findings reveal that age is the most significant predictor of differing unmet need composition and volume; gender, race, income, and veteran status are only minor. Repeat clients disproportionately seek assistance in housing, while first-time clients seek utility assistance, suggesting that repeat clients contend with long-term deprivation. Further, rural and urban categorization of location is insufficient for meaningful analysis, with county-level examination proving superior, in some cases. This research outlines flaws in PA211's data procurement process, including geographic categorization (rural/urban versus county-level) and survey strategies. I suggest remedies including more rigid surveying of client assets, housing and employment status, as well as closer examination of heightened unmet needs among those who fail to divulge ethnic identity.

Executive Summary

My analysis determined that the most prominent demographic predictor is age, exhibiting a heightened volume and more varied composition of unmet needs, especially among the 41-60 age group. Gender showed mixed results, with female clients exhibiting roughly the same volume of needs as male clients, while contacting PA211 slightly more often. This outcome must be treated with caution; the literature review indicated that gendered differences in behavior that may skew outcomes, such as a propensity among men to avoid assistance (C. Ross Hatton et al., 2024), which was unaccounted for in this analysis. Racial differences are minor, save for a disproportionately high volume of unmet needs among those who did not record their race. This relationship should be a priority target for further instrumentation and analysis. Lastly, analyzing outcomes via RUCA codes that divide areas into rural, urban, and suburban proved insufficient. County-level analysis is a far more meaningful geographic metric, but only in isolated cases, and further development of county-level data procurement (particularly, expanded sample sizes) may allow for more precise geographical measurement.

Data insights:

- Age is the most strongly predictive demographic factor, with middle-aged and elderly clients seeking more assistance overall, as well as a wider variety of assistance types.
- Race provides mixed results, with minor differences in volume or variety of assistance.
- Those who chose not to record their race when interacting with PA211 exhibit disproportionately high volumes of *unmet needs*. This requires further examination and may be remedied by further data collection and investigation.
- Gender is a minor factor, though this conclusion must be treated with caution. Accounting for social differences between each gender could plausibly change the outcome of future analysis.

- First-time clients of PA211, in comparison to repeat clients, have a strong proclivity for seeking utility assistance. This is the most prominent difference between the two groups. Conversely, repeat clients tend to seek housing assistance more often, though this effect is slightly weaker. This finding reflects the longer-term nature of housing expenses, and may indicate insufficiency of housing assistance services in Pennsylvania.

Suggested interventions:

- Surveys to determine current assets of clients, as well as their housing and employment status could improve future datasets. Furthermore, partnerships with both public and private assistance organizations can help track how much of a client's income is provided by these services, bolstering future analysis, especially with regard to repeat versus first-time client outcomes.
- Urban, rural, and suburban geographic categories are too broad to grant meaningful results. County level examinations allowed for actionable results, but only in a few cases. Further investigation of county-level differences, particularly expansion of sample sizes, may provide superior geographic data metrics.

Introduction

PA211, my research partner for this project, is an organization that connects Pennsylvania residents with a variety of services offered by other assistance organizations. Residents seeking assistance are referred to as clients. Clients may contact PA211 by dialing 211, texting their zip-code to 898-211, or by calling the organization directly. The largest of its kind in the state, the organization attempts to direct its clients to food, housing, healthcare, and many other critical types of assistance. PA211 maintains datasets, provided to me for this research, which categorizes and tracks the quantities of these factors across all client interactions, elaborated further in the data section of this paper.

Unmet needs[1] are defined as any type of assistance a client was unable to find, even after contacting PA211. Specifically, it indicates that no service that could be reached by the client could help them with that need. This paper surrounds five questions posed by the organization which focus on the relationships between demographic, location, timing factors, and *unmet needs*. An examination of these relationships may reveal–and offer solutions for–flaws in PA211's organizational strategies, data procurement, and prioritization of vulnerable demographics.

First, PA211 asked what, if any, relationships exist between *age* and *unmet needs*. Given the nature of aging, it is reasonable to infer that stark differences between sought needs exist between age groups. An examination of each, as well as the composition of their unmet needs, may allow for more precise targeting of assistance.

Next, I was asked to identify the co-occurrence of categories of *unmet needs*. Services provided by partners of PA211 vary greatly, with a range of 17 categories of assistance, all containing numerous individual services. Awareness of these relationships would allow for improvements in client awareness, e.g. recommending additional services, should a client pursue assistance with needs that are correlated with others.

Third, I was tasked with finding relationships between geographic location and *unmet needs*. PA211 provided client data in conjunction with RUCA codes (Rural-Urban Commuting Area Codes), a land-

[1] Italicized text refers to variables from the data used in modelling

classification system administered by the U.S. Department of Agriculture, grouping individuals into *urban*, *rural*, or *suburban*. Evidence indicates that these classifications do not fully account for food insecurity once household characteristics are accounted for (Mabli et al., 2010), prompting an additional investigation into county-level differences for a later question. Total *unmet needs* were then compared across the three groups, followed by multiple regression modelling to uncover whether location still mattered after controlling for basic demographic factors.

The fourth question asked me to compare the diversity of *unmet needs* sought by repeat clients and first-time clients. Behaviors of repeat clients may betray underlying differences in priorities, allowing PA211 to more accurately predict whether clients are going to return, or to determine which avenues of assistance are insufficient for preventing dependency on assistance services. The literature review has already demonstrated the superiority of targeting clients differently, rather than using the same approach for all (Ascarza et al., 2017), lending credence to the development of this style of tracking. For this analysis, three logistic regression models were built to examine which categories of *unmet needs* each group tended to seek assistance with.

Fifth, I was asked to determine relationships between demographic information, location, interaction time, and *unmet needs*. This analysis summarized, as well as strengthened, conclusions derived from previous investigations. Using a multiple regression model, I analyzed which unmet needs clients sought assistance with depending on when and where they contacted PA211. This analysis was performed using county-level location information, rather than the provided RUCA codes. This was both to investigate the potential of county-level geographic measurement as a meaningful metric, and to see if this metric could more precisely determine these relationships.

Lastly, this paper presents several recommendations that could assist PA211 in more accurately tracking outcomes, improving data procurement and analysis strategies, and consequently improving services. These suggestions include but are not limited to the collection of client outcomes in a more rigorous manner, such as a phone application or scheduled emails, collection of client asset information, as well as employment and housing statuses.

Data

Data used in this analysis were sourced from PA211, categorizing the services they provide referrals for into taxonomy groups, e.g. utilities referring to assistance with gas bills, electric bills, etc. Also included were demographic factors, such as race, gender, geographic location (*urban*, *rural*, *suburban*), *age*, and number of adults in household, alongside factors such as number of *unmet needs*, interaction time, and interaction method. To precisely answer PA211's questions, tailored statistical analysis methods were applied to each, accounting for any factors that may have skewed outcomes.

This data uses a three-part structure. First, client is defined as a Pennsylvanian resident who contacts PA211 with the goal of finding services, whether this be in-person, or by phone. This contact is referred to as an interaction, and is paired with later follow-up data, indicating whether a client was able to attain the assistance they sought. Each client is assigned a unique *client ID* that corresponds to them as soon as they seek PA211's services, allowing for identification of each individual across all interactions. This client ID is paired with client information, allowing for tracking of demographic information among all clients. Second, each interaction generates a *referral ID*, another unique code that corresponds to needs they seek a referral for, as well as their location, the time of the referral, and whether they. Next, the specific needs that clients seek are categorized under taxonomy groups, or subcategories that contain each kind of need, e.g. the food group containing emergency food, food storage, meals, etc. Lastly, each of these needs are assigned a number per-interaction, allowing me to count the total number of times they are sought by a client.

Each interaction was first joined together across each dataset using *client ID* or *referral ID,* specific strings of numbers generated for each client interaction, ensuring consistency when counting variables. Functionally, *client ID* allows isolation of clients, each representing a single point of data, while *referral ID* isolates which service(s) clients sought from PA211. Using this system of measurement allowed for precise counting of each client's interaction, as well as quantities of each type of service provided by each. Missing or corrupted entries were removed for all questions.

For the analysis of the relationship between *age* and *unmet needs*, the datasets were joined using *referral ID*, isolating the specific types and quantities of requested assistance. Only 1,279 *age* entries were missing out of 17,651, representing roughly 7% of total entries. Age values of these clients were then binned in intervals <18, 18-25, 26-40, 41-60, and 61+. Remaining entries were then counted and presented categorically via the stacked bar chart.

Next, when determining relationships between co-occurring *unmet needs*, datasets were once again joined together using *Referral ID*. Because this analysis relied on taxonomy groups, categories represented by words, it was necessary to convert these categories into numbers to meaningfully measure them. To ameliorate this, modelling used "One-hot encoding", a technique that assigns each need a one or a zero for every interaction, with one indicating the presence of that need, and zero indicating its absence. For example, if a client sought rent assistance but not food assistance, the former receives a one, while the latter receives a zero.

To analyze the relationship between environment and *unmet needs*, interactions were joined together using *client ID*. This form of ID was used to measure the prevalence of each type of location present in every interaction, specifically whether each client resided in an *urban*, *rural*, or *suburban* area, and how many *unmet needs* appeared in each. This analysis used RUCA (Rural-urban commuting area) codes, a system of classifying each area in the United States based on the volume and density of commuter activity. A RUCA designation can range from one to ten, with one corresponding to the most metropolitan area and ten to the most rural, with varying degrees between. For this analysis, one to three indicates an urban area, four to six indicates a suburban area, and seven to ten indicates a rural area. Certain RUCA codes, representing empty land or bodies of water, were removed from this dataset, as well as 259 missing age values.

Furthermore, I examined the diversity of different needs requested by *repeat clients*, compared to *first-time clients*. The modelling controlled for demographic factors (gender, race, veteran status, and income), allowing for isolation of the effects of repeat client status on composition of *unmet needs*. A small number (roughly 800 of 18,519, or 4.3%) of missing entries were removed from the data.

Finally, I determined the relationships between demographic information, location, interaction time, and *unmet needs*. The method of analysis for this question used demographic factors (race, gender, *age*), as well as the *weekday*, *hour*, and county where each interaction took place. As before, missing entries were removed, with a slightly higher (1,400) entries dropped in the age and gender categories. Rather than using RUCA codes to determine location, this analysis used Pennsylvania counties, allowing for a separate examination of the effects of this geographical classification method.

Method

To visualize the relationship between *age* and *unmet needs*, a simple stacked bar chart was used for comparison of *age* intervals. These intervals are displayed on the x-axis, while the y-axis displays the type and quantity of unmet needs. This visualization was chosen to clearly demonstrate how needs vary across age groups, as well as their composition.

To determine relationships between co-occurring needs, a correlation matrix was used - this method quantifies how often different unmet need categories appeared together during client interactions. Each cell represents the strength and direction (whether they tend to appear together or not) of the relationship between each combination of *unmet need*. To evaluate these categories, one-hot encoded indicators were used, converting each category into either one (present) or zero (not present). Each interaction was treated as a separate observation and aggregated per-interaction. Pearson (phi) correlation coefficients were computed between all need categories, allowing me to quantify each in a range between zero and one, with results closer to one indicating more co-occurrence, and results closer to 0 indicating less co-occurrence. A filtered upper-triangle matrix is reported, preventing variables which may correlate together from skewing the outcome.

My method for determining relationships between geographic location and *unmet needs* began with creating a multiple regression statistical model, using several independent variables (*urban*, *rural*, *suburban*, and *age*) to predict or explain the number of *unmet needs.* β_0 represents the average *unmet needs* for a resident of the suburbs, the baseline group. β_1 represents the difference in *unmet needs* between *urban* and *suburban* clients, while β_2 represents the difference in *unmet needs* between *suburban* and *rural clients*. $\beta_3$represents the increase in *unmet needs* that occurs when a person becomes one year older.

$$\textit{Unmet needs}_i = \beta_0 + \beta_1(Urban)_i + \beta_2(Rural)_i + \beta_3(Age)\ i + \epsilon_i$$

For the question on the *unmet needs* of repeat clients, three separate logistic regression models were constructed to understand whether *repeat clients* of PA211 have differing *unmet needs* when compared to *first-time clients*. For each model, only the dependent variable was swapped out – this variable being the particular category of *unmet need* (*food, housing,* or *utility)*, allowing me to see whether repeat and first-time clients have differing probabilities of having those specific kinds of unmet needs. The independent variable, *repeat client (repeat client=1, first-time client=0)*, indicates whether the client contacted PA211 more than once, allowing for isolation of the effects of this status on each of the three types of need. These models also include demographic variables such as gender, race, veteran status, and income, so that the effect of client status is isolated from the effects of demographic factors.

The results of these models will provide the effect of each status (*Repeat Client/First-time Client)* on the overall volume of unmet needs, as well as the likelihood of each for seeking different unmet needs. β_0 represents the baseline probability that a first-time client requests the dependent variable, the specific *unmet need* in each equation. β_1 represents the differing probability of that unmet need being requested by a repeat client, in comparison to a first-time client. β_2–β_5 are control variables, accounting for gender (*male*=1, *non-male*=0), race (*nonwhite*=1, *white*=0), veteran status (*veteran*=1, *non-veteran*=0), income (in *dollars*) and their reference categories.

$$\text{logit}(P(\text{FoodNeed}=1)) = \beta_0 - \beta_1(\textit{RepeatClient}) - \beta_2(\textit{Male}) + \beta_3(\textit{Nonwhite}) + \beta_4(\textit{Veteran}) - \beta_5(\textit{Income})$$

$$\text{logit}(P(\text{HousingNeed}=1)) = \beta_0 + \beta_1(\textit{RepeatClient}) + \beta_2(\textit{Male}) + \beta_3(\textit{Nonwhite}) - \beta_4(\textit{Veteran}) - \beta_5(\textit{Income})$$

$$\text{logit}(P(\text{UtilityNeed}=1)) = \beta_0 - \beta_1(\textit{RepeatClient}) - \beta_2(\textit{Male}) - \beta_3(\textit{Nonwhite}) + \beta_4(\textit{Veteran}) + \beta_5(\text{Income})$$

Finally, to uncover relationships between demographic information, location, interaction time, and *unmet needs*, a multiple regression model was built, examining the effects of these variables on the dependent variable, total number of *unmet needs*. In this model, β_0 represents the expected number of needs for the baseline group (the reference categories for *race, gender*, and *location*). β_1 captures the change in the expected number of *unmet needs* for each additional year of *age*. β_2 represents the difference in *unmet needs* for each *gender*, and β_3 shows how location (*urban*, *rural*, and *suburban*) affects the number of needs compared to the baseline location group. β_4 indicates how times of contact, such as the *hour* or *day*, influences the number of *unmet needs*, given that times of contact may reflect higher urgency or systemic barriers (e.g. work schedule, childcare, etc.).

To evaluate the predictive performance of the model, R^2_{adj} was used. R^2_{adj}, or adjusted R^2, is a metric that measures how much demographic and timing factors explain the variation in *unmet needs*, as opposed to unknown variables not included in the model. Adjusted R^2 refines this measure by accounting for the number of predictors included, making it more reliable when comparing models. Higher values of R^2 and adjusted R^2 indicate a better-fitting model with stronger explanatory power.

$$\textit{Unmet needs}_i = \beta_0 + \beta_1(\textit{Age})_i + \beta_2(\textit{Gender})_i + \beta_3(\text{County})_i + \beta_4(\text{Time})_i + \epsilon_i$$

Results

(1): The Relationship between *age* and *unmet needs*

The results of the analysis between age and need are visualized via a stacked bar chart. The x-axis displays age intervals spanning from <18 to >60; the y-axis displays the number of unmet needs, composition of unmet needs, and the total number of clients in each age group. To the right of the chart is a key for taxonomy groups, color-coded to better illustrate results.

Under 18: Minors demonstrate a high need for utility and information assistance, particularly struggling with family insecurity. However, they are also the age group least likely to contact PA211 for assistance. I hypothesize that this is due to minors contacting PA211 with the assistance of an adult. Their needs tend to reflect household challenges, rather than individual needs.

18-25: Total unmet needs spike for the 18-25 age group in comparison to clients with age <18. Comparatively, they tend to seek housing, food, and income assistance, while facing legal, consumer, and public safety concerns. I hypothesize that young adults face instability as they enter the work force, find housing, or pay rent, and that they struggle with safety concerns.

26-40: Volume of calls increases even more remarkably in this age group, showcasing a disproportionate number of unmet housing, utilities, financial, childcare, and family service needs. This age group has the most diverse assortment of needs; I hypothesize that this is the result of newfound challenges related to managing a household, raising children, and maintaining employment.

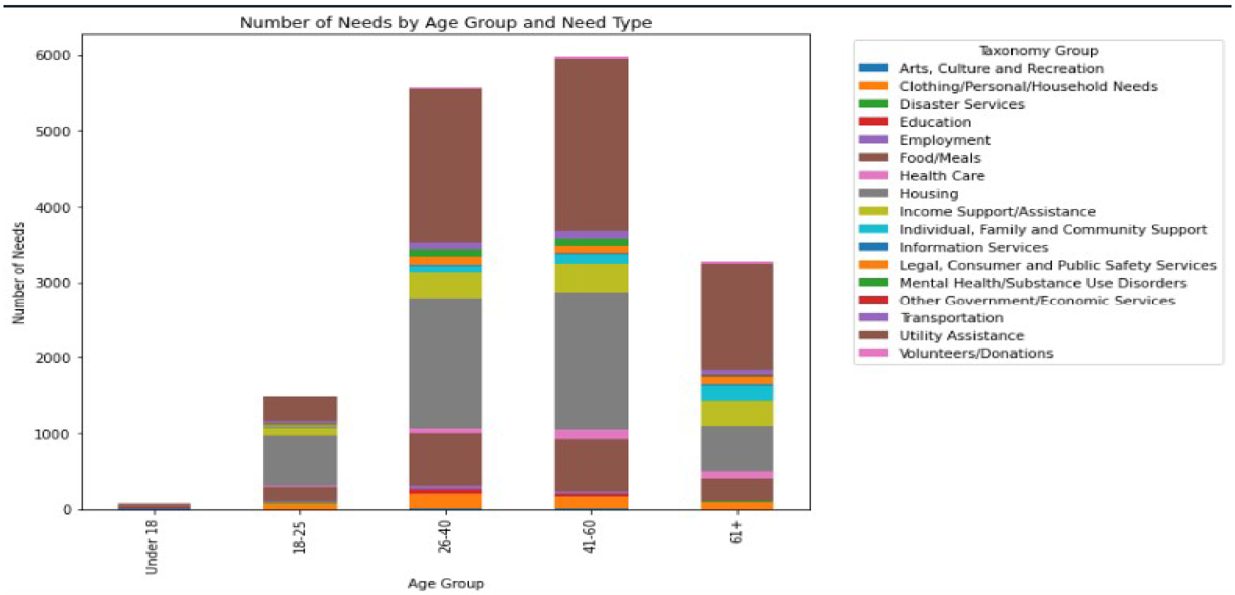
41-60: The 41-60 age group represents the highest overall volume of calls, with housing, utility, food, and financial assistance remaining at high levels. There is a pronounced increase in unmet transportation and healthcare needs. I hypothesize that middle-aged clients naturally face heightened mobility and medical issues, and that these challenges also motivate clients within the age group to seek mental health service.

61+: In this group, a significant drop in call volumes is observed. Clients above the age of 61 are most likely to seek assistance in housing, utilities, transportation, and healthcare. This indicates that as clients age, their needs become even more centered around health and mobility. The stacked bar chart shows that, comparatively, elderly clients rely more on medical support services and transportation than younger age groups.

Interpretation:

The result of this analysis illustrates the comparative differences in unmet needs among age groups. Younger clients (<18-25) tend to struggle with economic and housing instability, while middle-aged clients (26-40) have disproportionate unmet needs in childcare and housing assistance. Older clients (41-61+) have a greater need for transportation and healthcare services, showcasing heightened reliance on community support to perform daily activities. The results of this analysis, visualized in a stacked bar chart, are presented in Figure 1.

Figure 1: Composition of *unmet needs* by *age* interval



(2): The relationships between co-occurring *unmet needs*

The relationships between co-occurring unmet needs are visualized via a correlation matrix, in which the rate of co-occurrence between two unmet needs is shown in cells where the x and y-axes intersect. The rate of co-occurrence is expressed as a decimal, on a scale between -1 and one. The former represents complete lack of co-occurrence, while the latter represents complete co-occurrence. On the correlation matrix, red indicates a more negative correlation (the 10th percentile), showing that those two needs rarely co-occur, yellow indicates weak or ambiguous correlation (the 50th percentile) between each need, and green indicates a positive correlation (the 90th percentile), showing a tendency for those unmet needs to occur together.

Interpretation:

The correlation matrix demonstrates only a few highly correlated unmet needs. Namely, clothing, personal, and household needs tend to appear alongside education and food needs. Interestingly, volunteering and donations tend to occur alongside clothing, personal, and household needs – whether the poor donate more is a contentious topic in previous scholarship, with several conflicting studies published in the last 15 years (Pittarello et al., 2022; Malike et al., 2023). Furthermore, miscellaneous government and economic assistance tends to co-occur with legal, consumer, and public safety services. I hypothesize that the first two co-occurrences reflect broader attitudes on budget prioritization, e.g. food and clothing, being smaller purchases, perceived as more ‘reasonable’ for seeking assistance with, and education being perceived as ‘productive’. The latter two interactions require further examination, as they lack significant consensus in the literature review (Pittarello et al., 2022; Malike et al., 2023). For the correlation matrix of co-occuring unmet needs, see Table 1.

Table 1: Correlation matrix of co-occuring *unmet needs*

	Arts, Culture	Clothing/I	Disaster S	Education	Employme	Food/Mea	Health Ca	Housing	Income Su	Individual,	Informatio	Legal, Con	Mental He	Other Gov	Transport	Utility Ass	Volunteer
Arts, Culture and Recreation		0.03307	-0.0026	-0.0042	-0.0044	0.02118	0.05313	0.00253	0.0504	0.01187	-0.0049	-0.0089	0.03087	-0.0032	0.04623	-0.011	0.05142
Clothing/Personal/Household Needs			-0.0184	0.10138	0.0324	0.1098	-0.0111	-0.0613	0.0447	0.04659	0.01556	-0.0223	-0.0146	-0.0005	0.07923	-0.1439	0.14203
Disaster Services				-0.0071	-0.0074	-0.024	-0.0128	0.04812	-0.0086	0.02199	0.04651	-0.0151	-0.0104	-0.0054	-0.0144	-0.0572	-0.0068
Education					0.00723	0.01419	-0.0206	-0.0354	-0.0276	0.01308	-0.0132	-0.0047	0.01089	-0.0086	-0.0128	-0.0451	0.09228
Employment						0.04881	0.0113	0.00477	0.0147	0.00237	0.05257	0.04988	-0.004	-0.0089	0.02514	-0.0177	0.02839
Food/Meals							-0.0017	-0.0787	-0.035	0.01157	0.00119	0.07041	-0.0241	-0.0115	0.04423	-0.1228	0.00421
Health Care								-0.0585	-0.0132	0.07222	0.05466	-0.0047	0.04069	0.01443	0.03404	-0.1118	0.05087
Housing									-0.0429	0.00228	-0.0223	-0.0534	0.0015	0.01407	-0.0318	-0.4718	0.00149
Income Support/Assistance										-0.0182	0.01497	0.00687	-0.0383	0.01492	0.08781	-0.072	-0.0026
Individual, Family and Community Support											0.06109	0.03357	0.07432	0.05583	0.04092	-0.1376	-0.0015
Information Services												0.07335	0.02854	0.05834	-0.0001	-0.0509	-0.0127
Legal, Consumer and Public Safety Services													0.0594	0.11057	-0.0188	-0.1214	-0.0131
Mental Health/Substance Use Disorders														0.00569	0.02305	-0.1013	0.0127
Other Government/Economic Services															-0.0173	-0.0522	0.01906
Transportation																-0.0976	0.02033
Utility Assistance																	-0.0659
Volunteers/Donations																	

(3): The relationships between environment and *unmet needs*

Age: Age is a statistically significant predictor ($p < .001$), but the actual effect on *unmet needs* is vanishingly small. The positive coefficient indicates that clients request 0.0005 more unmet needs per year of age, which cannot culminate in anything meaningful over the course of a lifespan.

Rural: For people living in rural areas, the result lands at the edge of loose statistical significance ($p = .094$), meaning other factors are at play for disparities in unmet needs. The coefficient points toward rural residents seeking 0.0079 more unmet needs compared to those who live in suburbs. Despite its consistent appearance across the data, and its loose statistical significance, rurality is not meaningful.

Urban: The coefficient indicates that urban residents seek 0.0029 fewer unmet needs, a finding that is not statistically significant ($p = 0.471$). It is reasonable to conclude that the effects of urban residence are not explanatory for *unmet needs*.

Interpretation:

This analysis sought to determine whether *age* and geographic location (*urban*, *rural*, or *suburban*) can help explain or predict *unmet needs*. Statistical significance was present among three of four groups, all with negligible effects on unmet need*s*. Furthermore, the model itself was unable to sufficiently explain these relationships, as the $R^2_{\text{adj}} = 0.002$ - demonstrating that variables driving unmet needs are almost entirely outside of geographic location when categorized via RUCA codes. See Table 2 for geographic location and age coefficients.

Table 2: Geographic location and age coefficients

Independent Variables	**Estimated Coefficient**
Intercept	.0166**[.006]
urban	–.0029[.004]
rural	.0079*[.005]
Age	.0005***[.000]
F-test *P*-value	.00000524

Values in brackets are standard deviations, values in plain text are coefficient values. Asterisks to the right of coefficient values represent statistical significance levels. Four asterisks indicate extremely strong significance ($p <$ 0.0001), three asterisks indicate very strong significance ($p <$ 0.001), two asterisks indicate strong significance ($p <$ 0.01), and one asterisk indicates baseline significance ($p <$ 0.05).

(4): The unmet needs of repeat clients versus first-time clients

This analysis explores whether repeat clients have differing needs when compared with first-time clients. To answer this question, the average rate of appearance for each type of *unmet need* among both groups was determined. Next, both groups' needs in terms of *food*, *housing*, and *utility* assistance were compared by building three logistic regression models, to investigate whether these differences were meaningful. Included in the models were controls for demographic factors such as gender, race, veteran status, and income. These averages indicated that repeat clients are more likely to seek housing assistance, but less likely to seek utility assistance, with food only negligibly differing.

Food Needs

According to this analysis, there is no clear distinction between either group regarding the likelihood of seeking assistance with food needs, as repeat clients exhibit only an average marginal effect of -0.54%. Further, this variable is not statistically significant ($p = .405$), meaning it does not explain the already meager difference. This finding supports the notion that food insecurity is an equal opportunity issue, affecting all individuals (including those who have previously contacted PA211). The results also suggested slight demographic effects, with non-white clients being 2.39% more likely to seek food assistance ($p < .001$), and male clients being 1.3% less likely to ($p = .048$). These patterns do not change overall conclusions.

Housing Needs

This analysis determined that repeat clients are more likely than first-time clients to request assistance with housing-related issues. The repeat client coefficient was found to be, (.3173) and strong enough statistically ($p < .001$) to assume that it was not a chance occurrence, with a marginal effect indicating that this group is 6.32% more likely to seek housing assistance, overall ($p < .001$). Given the long-term nature of housing expenses, factors associated with them (e.g. eviction/foreclosure risk, homelessness, etc.) are typically unable to be resolved with one interaction. Similar to food needs, demographics also play a role, with males being 3.19% more likely to seek this form of assistance ($p = .001$), and nonwhite clients being 2.51% more likely ($p = .007$).

Utility Needs

Utility assistance (electricity, heating, or water bills) reverses the trend found in housing needs, representing the starkest difference between the two groups. Repeat clients were 15.01% ($p < .001$) less likely to request utility-related help, with the coefficient for repeat clients being negative (-.6920) and highly significant ($p < .001$). I hypothesize that this is the result of utility assistance differing from housing assistance, in that utility bills are far smaller expenses with substantially less risk to the client, should they fail to pay them. Minor demographic effects, including male clients being 3.79% less likely to seek housing assistance ($p < .001$), and nonwhite clients being 4.39% less likely to ($p < .001$).

Interpretation:

The largest disparity between either group is the disproportionate seeking of utility assistance among first-time clients. Conversely, repeat clients are less likely to seek housing assistance, though the effect is smaller. There are no major differences between either group in terms of food needs. Across every logit model, demographic characteristics do effect outcomes, but not at the scale of repeat client status. Given that all results are statistically significant, I can conclude that repeat client status is a meaningful metric for tracking the volume and composition of unmet needs. See tables 3 and 3b for total composition of *unmet needs* among repeat and first-time callers, and logistic modelling results for probabilities among repeat clients, first-time clients, and demographic groups, respectively.

Table 3: The total composition of *unmet needs*, between first-time and repeat clients

	Food Need	**Housing Need**	**Utility Need**
First-time clients:	11.30%	26%	42.30%
Repeat clients:	12.10%	33.50%	25.90%

Table 3b: Relationships between repeat-client status and *unmet needs*, logistic modelling

Independent Variable	**Food Need**	**Housing Need**	**Utility Need**
Intercept	-1.8200*** [.064]	-0.6930*** [.047]	-0.6765*** [.041]

RepeatClient	-0.0515 [.062]	0.3173*** [.044]	-0.6920*** [.044]
Male	-0.1305** [.066]	0.1602** [.046]	-0.1746*** [.045]
Nonwhite	0.2275*** [.064]	0.1261** [.047]	-0.2024*** [.046]
Veteran	0.0412 [.121]	-0.1905** [.092]	0.1091 [.078]
IncomeNum	-0.0001*** [3.27e-05]	-0.0003*** [2.49e-05]	0.0003*** [1.96e-05]

Values in brackets are standard deviations, values in plain text are coefficient values. Asterisks to the right of coefficient values represent statistical significance levels. Four asterisks indicate extremely strong significance ($p <$ 0.0001), *three asterisks indicate very strong significance* ($p <$ 0.001), *two asterisks indicate strong significance* ($p <$ 0.01), *and one asterisk indicates baseline significance* ($p <$ 0.05).

(5): The relationships between demographics, county, and *unmet needs*

This analysis sought to explain how individual demographic factors, such as gender, age, location, etc., as well as timing variables like *hour* and *weekday*, effects rates of *unmet needs*. Using a multiple linear regression model to predict the total number of needs per-interaction, I was able to isolate the effect of each demographic factor, while controlling for interaction timing. Furthermore, I controlled for county, rather than RUCA code regions, to investigate whether this metric would prove superior.

Age is a strong predictor of need volume: As client age increases, the number of needs per-interaction rises significantly, with 0.0573 more needs per-year of age, starting at 0 ($p < .001$). Further, age-squared, a metric for determining whether this trend drops off or accelerates after a certain age, is -0.0007 ($p < .001$), indicating that elderly clients seek slightly fewer needs for each year of age past 61. Findings indicate that, overall, middle-aged adults (41-60) exhibit the highest overall volume of needs. This implies that working age adults, often battling employment instability, family obligation, and housing costs, are a core high-need demographic. Notably, age was by far the most effective predictor in the totality of this research.

Gender does not significantly predict need volume: When controlling for geographic location and demographic information, there was no meaningful difference in presented needs per-interaction between genders. Further, the prediction coefficient (.671) is not statistically significant (p = .45). Though women contact PA211 slightly more often in many Pennsylvania counties, the diversity and volume of needs do not differ substantially from male clients, and statistical significance is lacking in all but two counties. Gender-specific outreach remains useful for targeted assistance avenues, but gender is not a major predictor of significance.

Race and Ethnicity show mixed effects: Differences in need volume do not significantly differ between clients of different races. Interestingly, clients who do not provide race/ethnicity information present significantly higher need volumes than those who do. I hypothesize that this may indicate privacy concerns, high levels of stress among these clients, or both. This relationship carries significant weight for PA211's public outreach strategy, as more information is required to properly explain it - clients who skip demographic fields are among the highest need groups, and may require deeper case support.

In some cases, county differences reveal uneven resource environments: Examination of county-level differences derived more meaningful results than broader urban vs. rural analysis, but was only statistically significant in two counties. Clients from McKean County have 0.83 fewer average needs ($p < .001$), while Venango County exhibits 0.41 more needs, on average (p = .01). I believe that these findings reflect a complex assortment of variables that require further study, such as local politics, culture, or economy. These results indicate that, in certain cases, RUCA codes may be an inferior geographic metric, and further investigation on why certain counties exhibited statistical significance could help develop these metrics further.

Time contact variables are not significant predictors: When demographic and county conditions are accounted for, time of contact does not meaningfully affect overall need volume.

Interpretation:

Overall, *age* was identified as the strongest predictor, with clients in the 41-60 age group exhibiting the highest overall volume of unmet needs. Clients who do not report their race are a similarly high-volume group, suggesting that missing demographic information may signal deeper problems that require further examination. Geographic differences are evident as well, but only when categorizing via county, and in certain counties. In contrast, gender, hour, and weekday do not meaningfully predict for overall need volume. Thus, I find that age, county, and missing ethnic information groups are the worthiest groups for further examination. See Table 4 for the coefficients, standard-errors, and totals of each variable.

Table 4: Relationships between demographics, timing, county, and needs, multiple linear regression modelling

Independent variable	**Outcome**	**Standard error**	**Totals**
Intercept	2.4774****	(0.373)	
Gender: Male	0.0671	(0.089)	4415
Gender: Female(baseline)	0	0	10478
Race: Missing	2.0020**	(0.693)	2384
Race: Other	-0.0656	(0.145)	1665
Race: White	0.1393	(0.121)	10341
Location: Cameron	-1.2645	(0.690)	14
Location: Clarion	-0.3479	(0.2300	150
Location: Clearfield	0.0734	(0.187)	253
Location: Crawford	-0.2836	(0.177)	294
Location: Elk	-0.5048	(0.316)	68
Location: Erie	0.1367	(0.115)	1682
Location: Forest	0.0760	(0.667)	16
Location: Jefferson	-0.0790	(0.269)	121
Location: McKean	-0.8319****	(0.239)	134
Location: Potter	-0.6451	(0.455)	106

Location: Venango	0.4143**	(0.161)	359
Location: Warren	0.0266	(0.3202)	95
Age	0.0573****	(0.014)	
Age^2	-0.0007****	(0.0001)	
Hour	-0.0119	(0.0110	
Weekday	0.0362	(0.025)	

Asterisks to the right of coefficient values represent statistical significance levels. Four asterisks indicate extremely strong significance ($p < 0.0001$)*, three asterisks indicate very strong significance* ($p < 0.001$)*, two asterisks indicate strong significance* ($p < 0.01$)*, and one asterisk indicates baseline significance* ($p < 0.05$).

Conclusion

This paper grants PA211 a preliminary framework for targeted assistance and outreach, as well as lines of questioning that may derive further meaningful results. Though race, gender, and other demographic categories display varying levels of effect on *unmet needs*, age is by far the most notable demographic predictor of composition, as well as overall volume. Furthermore, this analysis demonstrated stark differences in the needs of repeat and first-time clients, revealing a propensity to seek housing assistance among repeat clients, and more strongly, a propensity to seek utility assistance among first-time clients. Lastly, my research discovered major meaningful differences between two individual counties, while broader rural and urban classifications failed to derive meaningful results. This suggests that PA211 would benefit from targeted strategies aimed at assisting middle-aged and elderly clients, better accounting for differences between repeat and first-time clients, and targeted development of services and data procurement in underserved counties.

References

Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G. S., Lemmens, A., Libai, B., Neal, D., Provost, F., & Schrift, R. (2017). In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. *Customer Needs and Solutions*, *5*(1-2), 65–81. https://doi.org/10.1007/s40547-017-0080-0

Brülhart, M., Klotzbücher, V., & Lalive, R. (2023). Young people's mental and social distress in times of international crisis: evidence from helpline calls, 2019–2022. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-39064-y

C. Ross Hatton, Bresnahan, C., Anna Claire Tucker, Johnson, J., John, S., & Wolfson, J. (2024). Food for Thought: the Intersection between SNAP Stigma, Food Insecurity, and Gender. *Social Science & Medicine*, *361*, 117367–117367. https://doi.org/10.1016/j.socscimed.2024.117367

Mabli, J. (2014, March). *SNAP Participation and Urban and Rural Food Security* . Https://Fns-Prod.azureedge.us/; United States Department of Agriculture. https://fns-prod.azureedge.us/sites/default/files/SNAPFS_UrbanRural.pdf

Malika, M., Ghoshal, T., Mathur, P., & Durairaj Maheswaran. (2023). Does scarcity increase or decrease donation behaviors? An investigation considering resource-specific scarcity and individual person-thing orientation. *Journal of the Academy of Marketing Science*, *52*. https://doi.org/10.1007/s11747-023-00938-2

Pittarello, A., Motsenok, M., Dickert, S., & Ritov, I. (2022). When the poor give more than the rich: The role of resource evaluability on relative giving. *Journal of Behavioral Decision Making*. https://doi.org/10.1002/bdm.2293

Appendix

Table 1: Variable explainer

Variable Name	Age	Rural	Urban	Male	Non-White	Veteran	Transgender
Observations	13826	13826	13826	13826	13826	13826	13826
Mean	46.17*	0.23*	0.50*	0.29*	0.28*	0.07*	0.01*
Standard deviation	15.58	0.42	0.50	0.45	0.45	0.26	0.08
Minimum	0	0	0	0	0	0	0
Maximum	96	1	96	1	1	1	1

**These values represent how often each observation appears in the dataset, since they are expressed as binary variables. The closer each mean is to 1, the more likely each category is to appear among all observations. Age* values are *not expressed as a binary variable, and reflect the actual age (in years) of observations.*

Table 2: Demographic totals

Gender	Male 4415	Female 10478	Transgender 30	Did not answer 1267	Other 57	NaN 2272
Race	White 10341	Black 2324	Hispanic/Latino 499	Did not answer 1306	Other 1665	NaN 2384
Veteran	Veteran 759	Not a veteran 14927	Unavailable 901	Other 448	NaN 1484	
Location	Urban 1860	Rural 1062				

Table 3: Composition of taxonomy groups, services sought among all referrals

Utility Assistance	6472
Housing	5117
Food/Meals	2025
Income Support/Assistance	1266
Clothing/Personal/Household Needs	593
Individual, Family, and Community Support	460
Legal, Consumer, and Public Safety Services	366
Transportation	329
Health Care	319
Mental Health/Substance Use Disorders	255
Employment	118

Education	111
Information Services	72
Volunteers/Donations	67
Other Government/Economic Services	34
Disaster Services	34
Arts, Culture, and Recreation	20
NaN	861