Geographic Data Mining

R-Ladies Philadelphia

March 23, 2019

Executive Summary

This section should have up to 5 bullet points summarizing the main conclusions from the analysis. These should be worded in such a way that PAWS management can easily understand what actios would be beneficial. It can be a shorter version of the conclusions section below.

Contributors

Joy Payton, MS is the Supervisor of Data Education in the Department of Biomedical and Health Informatics at the Children's Hospital of Philadelphia. She leads the development and implementation of education and outreach programs to help CHOP scientists become data-savvy and make the best, most informed use of the tools they have available.

Karla Fettich, PhD is Head of Algorithm Development at Orchestrall, Inc. She leads efforts to develop data analytics solutions, predictive models and optimization approaches to create sustainable changes that improve operations and outcomes in long term care facilities.

Problem definition and dataset

This section should summarize the question your group has worked on, as well as any working definitions that you used (beyond what was outlined in the general description). Additionally, this section should contain any issues or challenges that you noticed in the data, that may be relevant.

Goals

These analyses examined the data in relation to geographic and population parameters, with two main objectives:

1) identify an initial set of variables that may be informative for application processing
2) provide a basis for discussion around the usefulness of geographical data analysis for PAWS at a broader level

Datasets

The following datasets were used:

1. Online applications for both cats and dogs In addition to the data collected via the online forms, applicants' addresses were extracted and associated with their respective census tracts. Census tracts are areas roughly equivalent to a neighborhood established by the Bureau of Census for analyzing populations, and generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people. Prior to making the PAWS data available, individual applicants' names, addresses and other identifiable data were removed from the dataset, keeping only census tract data and ZIP codes.

2. Trello cards and actions

3. Census data from the 2017 five-year American Community Survey via the American Fact Finder for the following areas:

- Economic characteristics
- Education characteristics
- Median rent
- Computer and networking characteristics

Results

Economic Considerations in Processing Applications

On average, dog applicants live in areas where the median income is higher compared to cat applicants (around $60,000/year for dog applicants vs. $54,000/year for cat applicants) and where the percent of households living under the poverty level is lower (18% for dog applicants vs. 22% for cat applicants). This suggests that dog applicants are from slightly wealthier neighborhoods. We further observed that dog applicants have more range between lower middle class and upper middle class, while cat applicants tend to skew more toward lower incomes. This finding aligns with the pet care cost estimates provided by the ASPCA which suggest that the first year total costs of owning a dog ($1,471 - $1,779) exceed those of owning a cat ($1,174) - although it is unclear how recent these estimates are.

Using the "complete" status of a trello card at the time when the data were pulled, we did not observe a neighborhood wealth difference between completed and non-completed applications. While a "complete" status is fairly vague (it does not indicate the outcome of an application), and several trello cards may have been incomplete due to them being fairly recent, the data do not indicate an economic bias when processing applications.

We further looked into some of the outcomes of application processing, specifically *red flags* and *denied* applications. Applications from neighborhoods with a lower household median income (under $50,000/year) are more likely to be red flagged and denied, compared to those with a higher household median income (over $50,000/year). Additionally, red flagged **cat** applicants have a lower estimated monthly budget than their non-red-flagged counterparts ($176 vs. $224). For **dogs**, a similar trend was observed, but it did not reach the statistical significance threshold ($212 vs. $277). This pattern also holds when it comes to emergency budgets: red flagged applicants have a lower estimated emergency budget than their non-red-flagged counterparts ($947 vs. $1,446 for **cats** and $735 vs. $1,848 for **dogs**). While we found that living in a lower income neighborhood does impact the estimated emergency budget at a statistically significant level, it only accounts for about 7% of the observed pattern. This indicates that there are additional factors that may play a role in how much money an applicant is able to set aside on a regular basis for pet care.

Efficiency Analysis in Philadelphia County

We also looked at applications that were processed within an efficient timeframe (defined here as 10 days), vs those that did not. An application was considered efficient if it was given a decision label ("denied", "do not follow up", "adopted", "adoption follow up", "approved", "ready to adopt", "ready for review", "reviewed with handouts only", "approved with limitation", "dog meet", "returned", "adopted elsewhere") and the last trello checklist item was checked off 10 days or less from the date of application submission.

Dogs

We found that in neighborhoods with a higher percentage of people who have a cell data plan and no other type of internet subscription, there was also a trend for a lower proportion of efficient applications, this effect being more pronounced in north and northeast Philadephia. There could be many reasons for this: applicants who live in areas where many people do not have easy access to the internet may not be as familiar with filling out an online application (which represents the current application dataset); they may also not be able
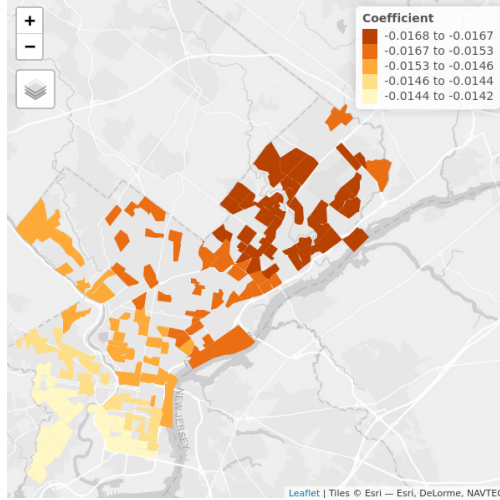
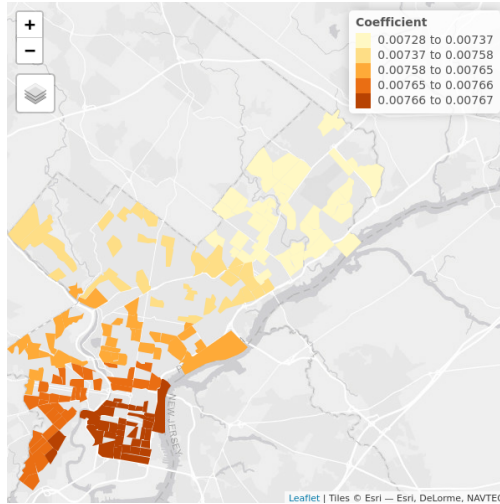Figure 1: Cell data plan only coefficients

Figure 2: Population 25-34 enrolled in school coefficients

to easily find the information they need (since not all websites are mobile friendly); or they may be filling out the application form on a mobile device, which might be too long/detailed to adequately complete on a small screen.

Additionally, in neighborhoods with a higher percentage of the population 25 to 34 year old enrolled in school, we also observed a significantly higher proprotion of efficient applications. It is unclear what the reasons behind this might be, but possible options include the applicants' level of comfort with online applications, access to information, or other factors that are more specific to the life circumstances of individuals enrolled in school. This effect was less pronounced in northeast Philly.

Cats

Interestingly, for cats we found that in neighborhoods where a higher percentage of the population is children in grades 5-8, the proportion of efficient applications was lower, this effect being more pronounced in the north and northeast. While we do not know the reasons for this effect, it may be worth noting that ownership of and interest in pets tend to peak in middle childhood (i.e., 8–12 years). It may be that this effect influences
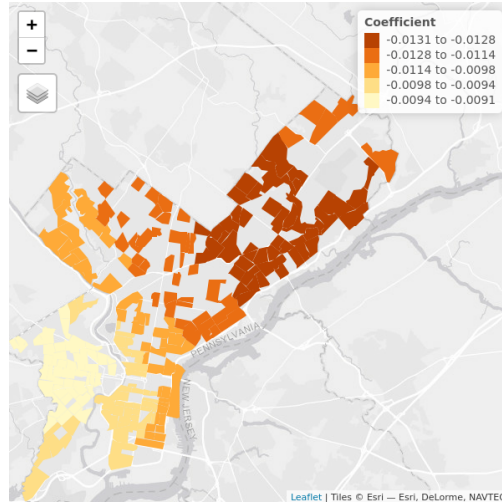
Figure 3: Children grades 5-8 coefficients

the decision to submit an application, but that other barriers interfere with the application's timely processing (e.g. incomplete information, lack of responsiveness to provide additional information, change of mind).

Conclusions and Next Steps

1. PAWS could develop a "smart" online application, that automatically educates the applicant on the cost of pet ownership when the budget is too low.

Since red flagged and denied applications still require processing by PAWS staff, and possibly even more intense processing than approved applications, it may be worth automatically screening and educating applicants who may have unrealistic budgeting expectations. Thus, perhaps a pop-up chart could appear when the budget is too low, *while* the applicant fills out the form. If the applicant proceeds to submit the application with a too-low budget, this application could be automatically labeled a red flag and sent for processing to a more experienced staff for further processing.

2. PAWS could provide applicants with a detailed breakdown of costs for a new pet, and have an adoption counselor go through the itemized list with the applicant to identify how each item could be covered.

Taking for instance the pet care cost estimates provided by the ASPCA, PAWS could identify which categories might be most difficult for an applicant to cover. Then, PAWS could provide a set of options (e.g. list of lower cost vets, cheaper options for enrichment using household items, list of affordable dog trainers) that might make the costs more manageable for those who are on a tighter budget.

3. PAWS could promote sharing or pooling of resources among its adopters.

Many pets have preferences when it comes to food, treats and toys, and it takes a while for a new adopter to learn them. This can result in a lot of wasted money. PAWS could facilitate and promote sharing of these resources (including any other accessories, or even transport help), at the adopters' own risk, via an online community.

4. PAWS could assess the user-friendliness of its online application form on different platforms.

While the PAWS website might be mbile-friendly, PAWS could further assess whether the application form itself is represented in the most efficient way on a mobile device. To do this, information would first need to be collected on the number of applicants who submit the application from a mobile device, as a revamping of the mobile interface for the application form may only be necessary if application quality is dependent on the device from which the application was submitted. An additional indicator of user friendliness could be the amount of time applicants spend on an application. PAWS could consider a 'smart' approach in sequencing and presenting questions so that the process is relatively speedy for the applicant, while also ensuring quality data.

5. PAWS could consider creating programs that are aimed at families with middle-schoolers.

Given that there is a spike in children's interest in animals in middle school, PAWS could consider some ways to increase involvement of children in the animal care process, either by creating kid-friendly volunteer opportunities, kid-friendly community groups among adopters, or even informational events where people who are interested in adopting can ask questions and discuss experiences with adopters and PAWS representatives.