**MUDAC 2017 PROBLEM**

**Background**

**Aeon Mission Statement: Strengthening Lives and Communities**Each day, more than 14,000 Minnesotans experience homelessness, creating an urgent need for affordable housing with structured support. With a vision that every person has a Home and is interconnected within community, Aeon’s mission is to create and sustain quality affordable homes that strengthen lives and communities. Having a personal place where people feel safe and connected is transformational – a stabilizing force for individuals and communities. We are committed to fostering a respectful, caring culture with responsive property management and affordable rent.

(<https://www.aeonmn.org/>)

**Aeon Core Values:**

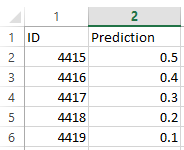
* Community - Collaborate in the design of vibrant, healthy neighborhoods
* Trust - Develop strong, effective relationships that last
* Diversity - Build an inclusive organization and community
* Excellence - Set and achieve high-quality goals and standards
* Sustainability - Ensure properties are long-term community assets
* Leadership - Advance positive community change

In this competition you will be using data provided by Aeon to look at various aspects of their affordable housing units, market rate units, the services they provide, and their tenants.

**Potential Questions of Interest  
  
Question 1 –** Aeon provides services referred to as Resident Connections at some of their properties. Some of these services include connecting individuals to community resources, housing stability support, and providing material goods. Aeon also hosts events that residents at some of their properties can attend or possibly volunteer at. If a resident volunteers, this demonstrates they are willing to take a leadership role at the residential complex. Use information in the file **Interactions (with binaries).csv** to characterize the services used by residents at Aeon properties where they are available. Similarly use information in the file **Events (with binaries).csv** to characterize the event attendance and event volunteering of residents at Aeon properties.

**Question 2 –** Aeon is interested in identifying tenants that are at risk of leaving Aeon housing under negative circumstances. Examples of negative circumstances would be non-payment of rent or lease violations. Use data in the file **Train.csv** to develop a predictive model that produces a probability that a tenant will leave Aeon housing under negative circumstances. You must then use your model to compute the probability of leaving under negative circumstances for the test cases contained in the file **Test.csv**. The most accurate model will be judged by area under the Receiver Operating Curve (ROC), which is referred to as AUC. You must post your predictions to our competition Kaggle ® site to see your position on the MUDAC Leaderboard and to ultimately win this phase of the competition.

Example of Submission File (.csv)



Don’t change ID order it must be 4415 to 5414!

Also it is VERY important that not only does your model do a reasonable job of identifying tenants that are likely to leave under negative circumstances, but you should be able explain which explanatory variables in your model are important and how they relate to the response.

**Question 3 –** In this question you will be combining tenant services utilization, tenant event attendance & event volunteering, and tenant demographics & background information to again look at negative departures from Aeon housing. How are all of these factors related to the response? Does access and utilization of services appear to decrease risk of leaving under negative circumstances? Which services matter most? How about event attendance and event volunteering? This is NOT a prediction accuracy problem, though certainly developing a good model may be an important part of answering these questions. Use **Services and Outcome.csv** for this question, however you may choose to use information from other competition data files in this process.

**Question 4 –** Given all of what you have learned through your analyses above, what recommendations would you make to Aeon given their mission and core values?

**Aeon MUDAC Data Codebook**

**Data Overview**:  
There are four datasets you will be using in the competition, the details of which are presented below.

The first two datasets contain information about support services used by households and events attended from 2014 to present. You can use these dataset for the first question which asks you to describe support services utilized and event attendance by Aeon tenants.

The third dataset contains a “three-year” tenant history for households with move in dates between 2007-2013. This data set contains information available about tenants at the time of move-in with the exception of whether or not they incurred late fees on their rent payments at some point during their residence. You will use these data for the prediction accuracy phase of the competition. Again your goal is to predict whether or not the tenant moved out under negative circumstances within the first three years after move-in. This model could potentially be used by Aeon to identify high risk tenants where support services could potentially be utilized to help prevent tenant move-outs under negative circumstances. Prediction accuracy is NOT enough! Make sure you are able to discuss the variables in your model in practical terms.

The fourth dataset brings the services, event attendance, and move-out information together for tenants who moved in between the years 2014-2016. The goal here is to model move-outs (either under negative circumstances and/or under positive circumstances) using all available household information. What variables are related to these outcomes? What role does support services play? What roles does event attendance play? What impact does volunteering at an event have? You may also want to incorporate probabilities from your predictive model into this analysis as all variables used to develop your predictive model are available in these data.

**Data Details**

**Interactions (with binaries).csv**

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description/Defininition** |
| 1 | Yardi HOH ID | Head of household ID number |
| 2 | HOH | Is the individual involved in this interaction the head of the household? (Y/N) |
| 3 | Date | Date of interaction (mm/dd/yyyy) |
| 4 | Rent or Lease Violation Date | If the interaction was related to rent or lease violation then a date will appear in this field. (mm/dd/yyyy) |
| 5 | Move Out | If the tenant has moved out of their apartment then the date this occurred will be listed. (mm/dd/yyyy) |
| 6 | PropCode | Property ID for tenant |
| 7 | Reason | The reason for the interaction classified as either Proactive or Reactive. |
| 8 | Interaction:Record Type | Interaction type classified at the broadest level: Material Goods, Housing Stability, or Connection to Community Resource. |
| 9 | Type | Interaction type classified at the next broadest level:  Advocacy, Conflict Resolution/Mediation, Community Engagement, Coordination of Services, Crisis Intervention/Support, Direct Support, Eviction Prevention, Home Management, House Stability, Information & Referrals, Isolation Intervention, Lease Education, Material Goods, New Resident Orientation, Outreach, or blank. |
| 10 | Sub-Type | Most granular classification of interaction type: Bill paying, Budgeting, Complete, Education, Housekeeping Support  Inspection, Invited, Lease Violation, Lease warning, Mutual termination, Non-payment of rent, Organizing personal records, Reading mail, Recertification, Utility/cable/phone issues, or blank. |
| 11 | Resource | List of resources received with resources separated by a “;”. There are up to five resources listed in alphabetical order. |
| 12-16 | Resource 1 - 5 | First - fifth resource received in separate columns. |
| 17-18 | Reason\_Proactive & Reason\_Reactive | Binary columns for each level of Reason. Reason\_Proactive contains a 1 if the reason for interaction was proactive and 0 if it was Reactive. Opposite for Reason\_Reactive. |
| 19-21 | Interaction:Record binaries | Binary columns for each level Interaction: Record Type (Column 8) |
| **Column** | **Variable Name** | **Description/Definition** |
| 22-36 | Type binaries | Binaries for each level of Type (Column 9) of interaction. |
| 37-51 | Sub-Type binaries | Binaries for each level of Sub-Type (Column 10) of interaction. |
| 52-99 | Resource 1 binaries | Binaries for each level of Resource 1 (Column 12) |
| 100-144 | Resource 2 binaries | Binaries for each level of Resource 2 (Column 13) |
| 145-181 | Resource 3 binaries | Binaries for each level of Resource 3 (Column 14) |
| 182-202 | Resource 4 binaries | Binaries for each level of Resource 4 (Column 15) |
| 203-216 | Resource 5 binaries | Binaries for each level of Resource 5 (Column 16) |

**Events (with binaries).csv**

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description/Defininition** |
| 1 | Yardi HOH ID | Head of household ID number |
| 2 | HOH | Is the individual involved in this interaction the head of the household? (Y/N) |
| 3 | PropCode | Property code ID |
| 4 | Data | Date of the event. (mm/dd/yyyy) |
| 5 | First Move In | Date the tenant moved into Aeon Properties. (mm/dd/yyyy) |
| 6 | Move + First Event | Date tenant moved in OR 01/01/2014 if the tenant moved in before this date. Event data was pulled for events starting 1/1/14. (mm/dd/yyyy) |
| 7 | Move Out | Date the tenant moved out of Aeon Properties. If empty tenant has not moved out. (mm/dd/yyyy) |
| 8 | Move Out + Current | Date of Move Out OR 3/15/2017 the last date event data was pulled for this competition. |
| 9 | Time | Time in years tenant could have attended events given their values for Move + First Event and Move Out + Current. |
| 10 | Resident | Current or Former |
| 11-12 | Event and Volunteer | Binaries indicating whether the individual attended (Event) or volunteered (Volunteer) at the event. |

**Train.csv – data used to develop model for predicting the negative circumstance response.**

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description/Defininition** |
| 1 | ID | 1 - 4414 |
| 2 | PropCode | Property ID |
| 3 | Total Units | Total number of units at property. |
| 4 | UnitType | Tax Credit, Market Rate, Home Unit, 50059 |
| 5 | Type of Service | Provider of services at property: Resident Connections, Partner, or None. |
| 6 | Services Available | Indicator of whether services of some type are available at the property. (Y/N) |
| 7 | Service Units | Number of units at property have services available: all, 4 units, 5 units, 6 units, 10 units, 14 units, 15 units or blank. |
| 8 | Service start year | Year services of the first type became available. |
| 9 | Service end year | Year services of the first type ended if they did, current if services are still available, and blank if no services provided. |
| 10 | Type of service 2 | Second provider of services at the property: Partner or blank. |
| 11 | Service start year 2 | Year the second provider service began. |
| 12 | Service Units 2 | Has value “all” or blank only. |
| 13 | Service start year 2 | Year services of the second type became available |
| 14 | Aeon Services Available | Indicator of whether Aeon provided services were available at the property (Y/N). |
| 15 | HH Size | Number of members of household. |
| 16 | HH Income | Household income ($) |
| 17 | Mean HH Age | Mean of age of household members. Contains several missing values. |
| 18 | Rent | Monthly rent in dollars ($) |
| 19 | Latest Subsidy | Value of latest rent subsidy received ($). Lots of zeroes! |
| 20 | First Subsidy | Value of first rent subsidy received ($). Lots of zeroes! |
| 21 | Late Fees? | Did they incur late fees on their rent at some point? (Y/N) |
| 22 | Gender | Gender of head of household (M/F or NA) |
| 23 | Simplified Race | Race of head of household: API, Black, Decline, Multi-Racial, NA, Native American, Other, White. |
| 24 | Military | Does member of household have a military history? (Yes/No/NA) |
| 25 | Disability | Does head of household have a disability? (Yes/No/NA/Declined) |
| 26 | Immigrant | Is head of household an immigrant? (Yes/No/NA/Declined) |
| 27 | Homeless | Does head of household have a history of homelessness? (Yes/No/NA) |
| 28 | Marital Status | Marital status of head of household (M = married, S = single, Sep = Separated) |
| 29 | PrimaryIncome | Primary income source:  B = business, CS = child support, F = federal funding, G = general assistance, M = military, N = other, NA = not available, PE = pension, SI = SSI, SS = social security, T = TANIF, U = unemployment, W = public assistance |
| 30 | CriminalRecord | Does head of household have a criminal record? (Y/N) |
| 31 | Length of Rap Sheet | Number of entries on their criminal record. |
| 32 | Num Felonies | Number of felonies listed on their criminal record. |
| 33 | Felon? | Do they have a felony on their criminal record? (Y/N) |
| 34 | NumGrossMis | Number of gross misdemeanors on their record. |
| 35 | Num Mis | Number of misdemeanors on their record. |
| 36 | Num Unknown | Number of offenses where the type is not shown. |
| 37 | Num DispArr | Number of offenses where disposition = arranged. |
| 38 | Num DispCont | Number of offenses where disposition = continuance. |
| 39 | Num DispDis | Number of offenses where disposition = dismissed |
| 40 | Num DisOpen | Number of offenses where disposition = open |
| 41 | Num DispUnknown | Number of offenses where disposition = unknown. |
| **42** | **Negative** | **Moved out of property under negative circumstance?**  **(1 = Yes, 0 = No)** |

**Test.csv –** cases to predict the negative circumstances outcome for. Columns 1- 41 are identical to those for the **Train.csv** file above with the exception of the fact the Negative (Column 42) is missing for obvious reasons.

**Services and Outcome.csv –** use these data to understand how services and events have potentially impacted move outs under negative consequences.

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description/Defininition** |
| 1 | Household ID | ID number for head of household. |
| 2 | PropCode | Property ID |
| 3 | Total Units | Total number of units at property. |
| 4 | UnitType | Tax Credit, Market Rate, Home Unit, 50059 |
| 5 | Type of Service | Provider of services at property: Resident Connections, Partner, or None. |
| 6 | Services Available | Indicator of whether services of some type are available at the property. (Y/N) |
| 7 | Service Units | Number of units at property have services available: all, 4 units, 5 units, 6 units, 10 units, 14 units, 15 units or blank. |
| 8 | Service start year | Year services of the first type became available. |
| 9 | Service end year | Year services of the first type ended if they did, current if services are still available, and blank if no services provided. |
| 10 | Type of service 2 | Second provider of services at the property: Partner or blank. |
| 11 | Service start year 2 | Year the second provider service began. |
| 12 | Service Units 2 | Has value “all” or blank only. |
| 13 | Service start year 2 | Year services of the second type became available |
| 14 | Aeon Services Available | Indicator of whether Aeon provided services were available at the property (Y/N). |
| 15 | HH Size | Number of members of household. |
| 16 | HH Income | Household income ($) |
| 17 | Mean HH Age | Mean of age of household members. Contains several missing values. |
| 18 | Rent | Monthly rent in dollars ($) |
| 19 | Latest Subsidy | Value of latest rent subsidy received ($). Lots of zeroes! |
| 20 | First Subsidy | Value of first rent subsidy received ($). Lots of zeroes! |
| 21 | Late Fees? | Did they incur late fees on their rent at some point? (Y/N) |
| 22 | Gender | Gender of head of household (M/F or NA) |
| 23 | Simplified Race | Race of head of household: API, Black, Decline, Multi-Racial, NA, Native American, Other, White. |
| 24 | Military | Does member of household have a military history? (Yes/No/NA) |
| 25 | Disability | Does head of household have a disability? (Yes/No/NA/Declined) |
| 26 | Immigrant | Is head of household an immigrant? (Yes/No/NA/Declined) |
| 27 | Homeless | Does head of household have a history of homelessness? (Yes/No/NA) |
| 28 | Marital Status | Marital status of head of household (M = married, S = single, Sep = Separated) |
| 29 | PrimaryIncome | Primary income source:  B = business, CS = child support, F = federal funding, G = general assistance, M = military, N = other, NA = not available, PE = pension, SI = SSI, SS = social security, T = TANIF, U = unemployment, W = public assistance |
| 30 | Criminal Record | Does head of household have a criminal record? (Y/N) |
| 31 | Length of Rap Sheet | Number of entries on their criminal record. |
| 32 | Num Felonies | Number of felonies listed on their criminal record. |
| 33 | Felon? | Do they have a felony on their criminal record? (Y/N) |
| 34 | NumGrossMis | Number of gross misdemeanors on their record. |
| 35 | Num Mis | Number of misdemeanors on their record. |
| 36 | Num Unknown | Number of offenses where the type is not shown. |
| 37 | Num DispArr | Number of offenses where disposition = arranged. |
| 38 | Num DispCont | Number of offenses where disposition = continuance. |
| 39 | Num DispDis | Number of offenses where disposition = dismissed |
| 40 | Num DisOpen | Number of offenses where disposition = open |
| 41 | Num DispUnknown | Number of offenses where disposition = unknown. |
| 42 | MoveInDate | Move in date (mm/dd/yyyy) - years are between 2014-16. |
| 43 | MoveOutDate | Move out date (mm/dd/yyyy) – if missing did not move. |
| 44 | Move Out Current | Move out date (mm/dd/yyyy) OR date of data pull. |
| 45 | Move Out Charges | Move out charges incurred by tenant ($). Empty if no charges were incurred or tenant has not moved out. |
| 46 | Exit Outcome | Didn’t Move, Negative, Positive, or Neutral. |
| 47-50 | Exit Outcome Binaries | One binary column for each of the Exit Outcomes above. |
| 51 | Num Late | Number of late fees accessed during their residence. |
| 52 | Total Late Fees | Cumulative total of all late fees accessed. |
| 53 | Last Late Date | Date of most recent late fee. |
| 54 | Num Events | Number of events attended or volunteered at by household members during their residence. Realize that not all Aeon properties have these events. |
| 55 | Event Rate | Number of events attended or volunteered at per 12-month period. |
| 56 | Num Attending | Number of events attended (not volunteered at). |
| 57 | Attending Rate | Number of events attended (not volunteered at) per 12-month period. |
| 58 | Num Volunteering | Number of events volunteered at. |
| 59 | Volunteer Rate | Number of events volunteered at per 12-month period. |
| 60 | Event Involvement? | Did a member of the household attend an event? (Y/N) |
| 61 | Volunteered? | Did a member of the household volunteer at an event? (Y/N) |
| 62 | Num Interactions | Number of interactions by household members during their residence. See the **Interactions (with binaries).csv** file for types of interactions. |
| 63 | Interaction Rate | Number of interactions per 12-month period. |
| 64 | Had Interaction? | Did a member of the household have an interaction during their residence? (Y/N) |
| 65-88 | *Interaction* Rate | Rate of occurrence per 12-month period for each *Interaction* type designation. Again **see Interaction (with binaries).csv** for the types of interactions. |
| 89 | Material Good Request? | Did household make a request for material goods? (Y/N) |
| 90-99 | *Item*? | Did household member receive *item* as part of their material goods request? (Y/N) |