Family Promise Data Analysis

James Holtz, Alaina Stockdill, Niklas Roberts, Enerelt Bat-Erdene, Jonatan Olderr, Nana Boateng Amoah

Who is Family Promise?

- Family Promise is a homeless shelter organization situated in Spokane, Washington.
- It was formed through congregational advocacy and established its first home in the year 2000.
- Family Promise provides shelter and care services to homeless individuals from children to adults.

What is the mission of Family Promise?

“Equipping families and communities to end the cycle of homelessness”

- Family Promise envisions a future where no child experiences homelessness in Spokane county
- **Prevents** homelessness by providing case management and rental assistance
- **Provides** emergency housing and hospitality
- **Prepares** families with the tools they need to remain in housing and to assimilate in society easily

The Project

What is our goal?

- We want to provide helpful insights and analysis to enable Family Promise to better serve their clients and ways they can help their residents leave the shelter early and rejoin society.

How are we going to reach this goal?

- Look for underlying patterns and correlations in client data
    - Client data includes 142 columns
- Look for which characteristics can predict where a client ends up
- Gain a better sense of who Family Promise is serving
- Provide meaningful insight to aid homeless residents in assimilating into society quickly

Overview

- Group 1: Demographic Breakdown of Residents, and Modelling of which predictor variable can help understand the residents more.
- Group 2: Relationships Between Income and Exit Destination
- Group 3: Demographic Proportions

Methodology

1. Collected the data from Family Promise client

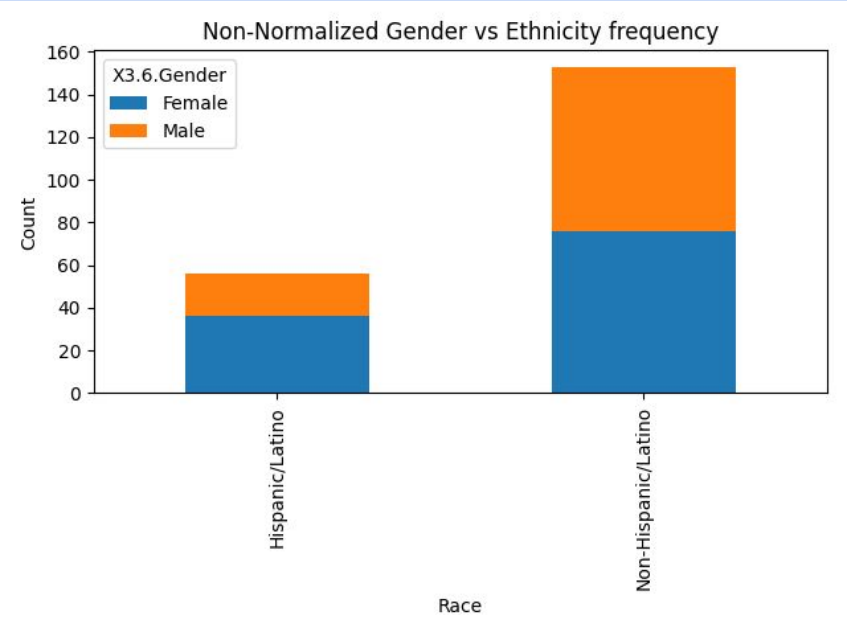
2. Filtered the data and removed redacted and sensitive sections from the datasets based on federal regulations.

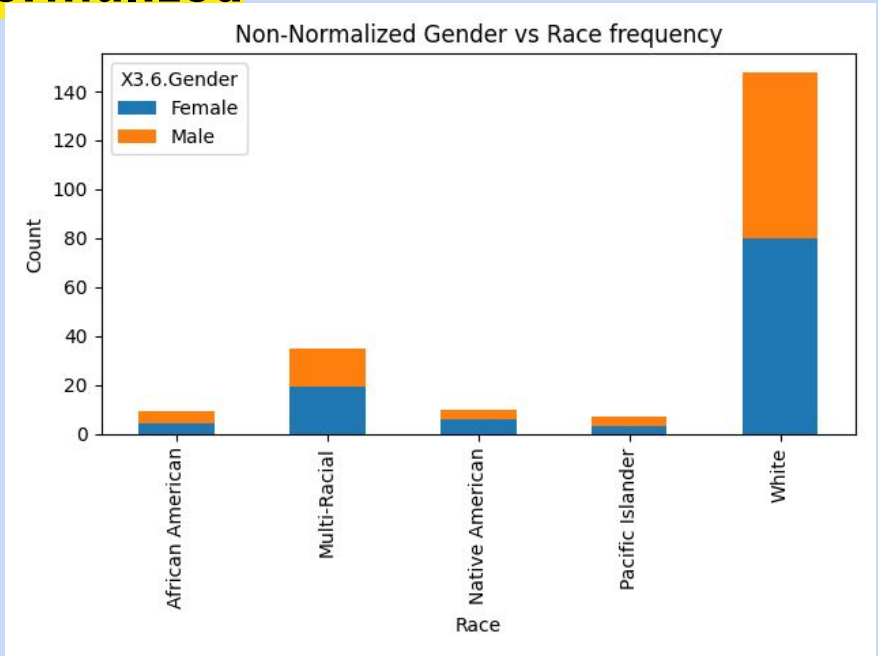
3. Since, the data was quite limited, and very categorical it restricted our analysis capabilities to perform modeling analysis using different data points.

Demographic Breakdown of our Data

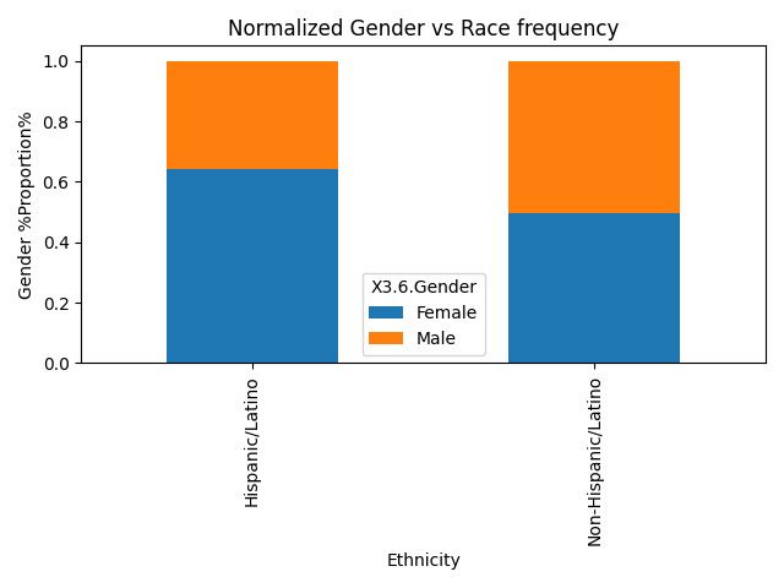
- **The demographic breakdown of residents at Family Promise**
- Non-Normalized Gender vs Ethnicity Frequency
- Non-Normalized Gender vs Race Frequency
- Normalized Gender vs Race Frequency
- Normalized Gender vs Race Frequency
- Linear Regression Analysis of Age vs Number of Days spent at Family Promise.

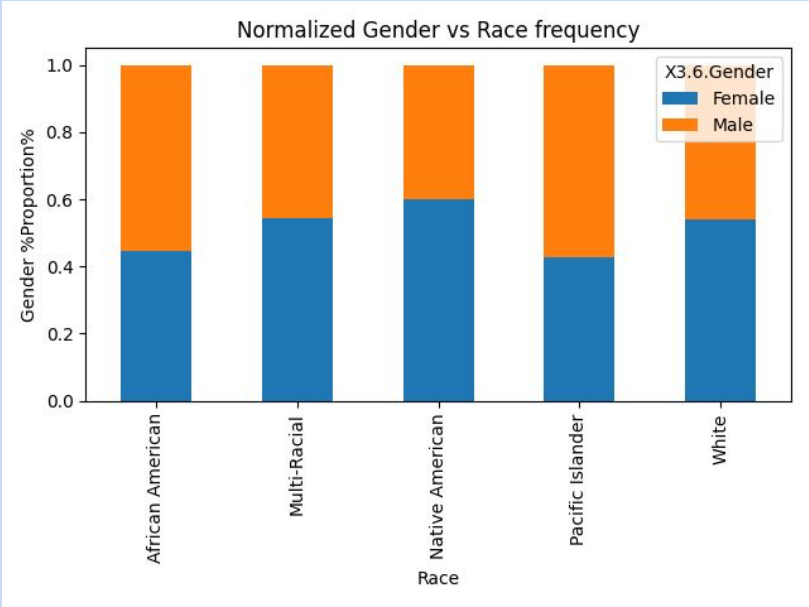
Demographics Breakdown - Non Normalized





Demographics Breakdown - Normalized





Group 1 Analysis

- Linear Regression
- C5.0 Analysis

Linear Regression Analysis

Analysis : The linear regression

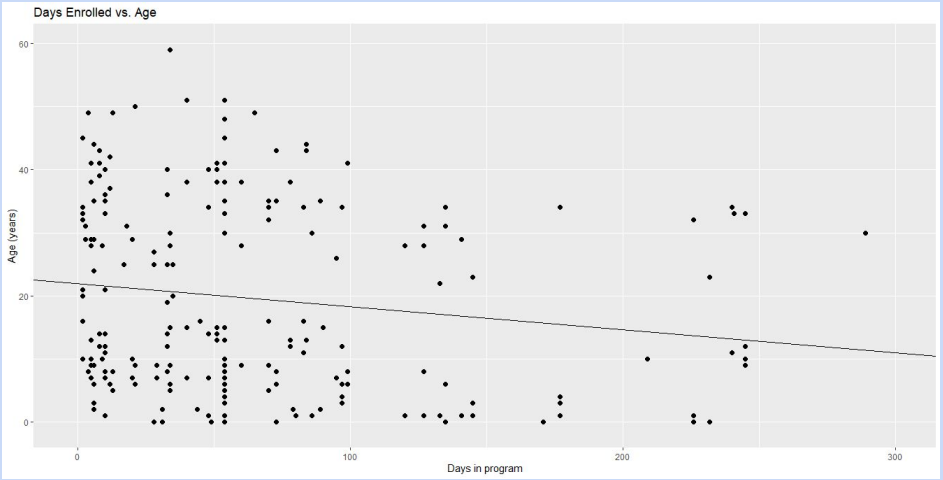
. is an initial analysis to see

.the spread of age with the numbe

. of days

. long spent at Family Promise

. Shelter.



Linear Regression Conclusion

- Since most residents spend on average 0-100 days in the shelter, to help increase the that percentage of the group, getting the residents involved in activities to make them independent, or contacting them with their loved ones will help reduce on average the number of days they spend at the shelter.

Summary of Linear Regression

- Performed a Linear Regression on the number of days spent at Family Promise, and the age of the individual.
- Summary-

```
Call:
lm(formula = Age ~ Days_Enrolled, data = data_clean)

Residuals:
    Min      1Q  Median      3Q     Max
-20.884 -13.177  -4.936  13.755  38.334

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  21.90523    1.47450  14.856   <2e-16 ***
Days_Enrolled -0.03646    0.01605  -2.271   0.0242 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.83 on 206 degrees of freedom
Multiple R-squared:  0.02443,   Adjusted R-squared:  0.0197
F-statistic: 5.159 on 1 and 206 DF,  p-value: 0.02416
```

C5.0 Analysis of Only Days and Age

Age :

	C5 Model Race	C5 Model Gender	C5 Model Ethnicity
Accuracy	0.5758	0.6207	0.6552
Error Rate	0.4242	0.3793	0.3448
Sensitivity	0.6379	0.5000	1.0000
Specificity/Recall	0.4878	0.7188	0.0000
Precision	0.6379	0.5909	0.6552
Beta	1.0000	1.0000	1.0000
F_Beta	0.5529	0.6486	0.0000
F1	0.5529	0.6486	0.0000
F2	0.5119	0.6889	0.0000
F0.5	0.6009	0.6127	0.0000

Days :

	C5 Model Race	C5 Model Gender	C5 Model Ethnicity
Accuracy	0.6552	0.5517	0.6897
Error Rate	0.3448	0.4483	0.3103
Sensitivity	0.7222	0.0000	0.9737
Specificity/Recall	0.5455	1.0000	0.1500
Precision	0.7222	NaN	0.6852
Beta	1.0000	1.0000	1.0000
F_Beta	0.6215	NaN	0.2461
F1	0.6215	NaN	0.2461
F2	0.5735	NaN	0.1778
F0.5	0.6783	NaN	0.3999

C5.0 Analysis - Days and Age

	C5 Model Race	C5 Model Gender	C5 Model Ethnicity
Accuracy	0.6000	0.6207	0.6897
Error Rate	0.4000	0.3793	0.3103
Sensitivity	0.6667	0.5000	0.9737
Specificity/Recall	0.5000	0.7188	0.1500
Precision	0.6667	0.5909	0.6852
Beta	1.0000	1.0000	1.0000
F_Beta	0.5714	0.6486	0.2461
F1	0.5714	0.6486	0.2461
F2	0.5263	0.6889	0.1778
F0.5	0.6250	0.6127	0.3999

The Setup

- For the C5.0 Analysis, the three target variables were race, ethnicity, and gender
- We used days at shelter and age to serve as predictor variables
- Accuracy for Age and Days was around 60% on average for our model.

Group 1 Conclusion

- Based on Race/Ethnicity majority of the residents at Family Promise is White
- Based on Gender, most residents are Females.
- Using age as an objective metric, we realized most aged groups spend spend between 1-100 days. This initial analysis can provide insight on fast the residents are able to assimilate back into society.
- Using C5.0 analysis for age only, we found that a model for age and ethnicity has a higher predictor correlation value than any of the other records (i.e Race, Gender)
- Using C5.0 analysis for days enrolled only, we found that a higher accuracy for days with ethnicity indicating a high predictor correlation.
- The same applied to when both age and days were combined, indicating that a model of ethnicity has a high insight of understanding the days enrolled with ages.
- Overall, the best trained model to help understand how to reduce the number of days residents are enrolled at Family promise is to use their ethnicity.

Group 2: Relationships Between Income and Exit Destination

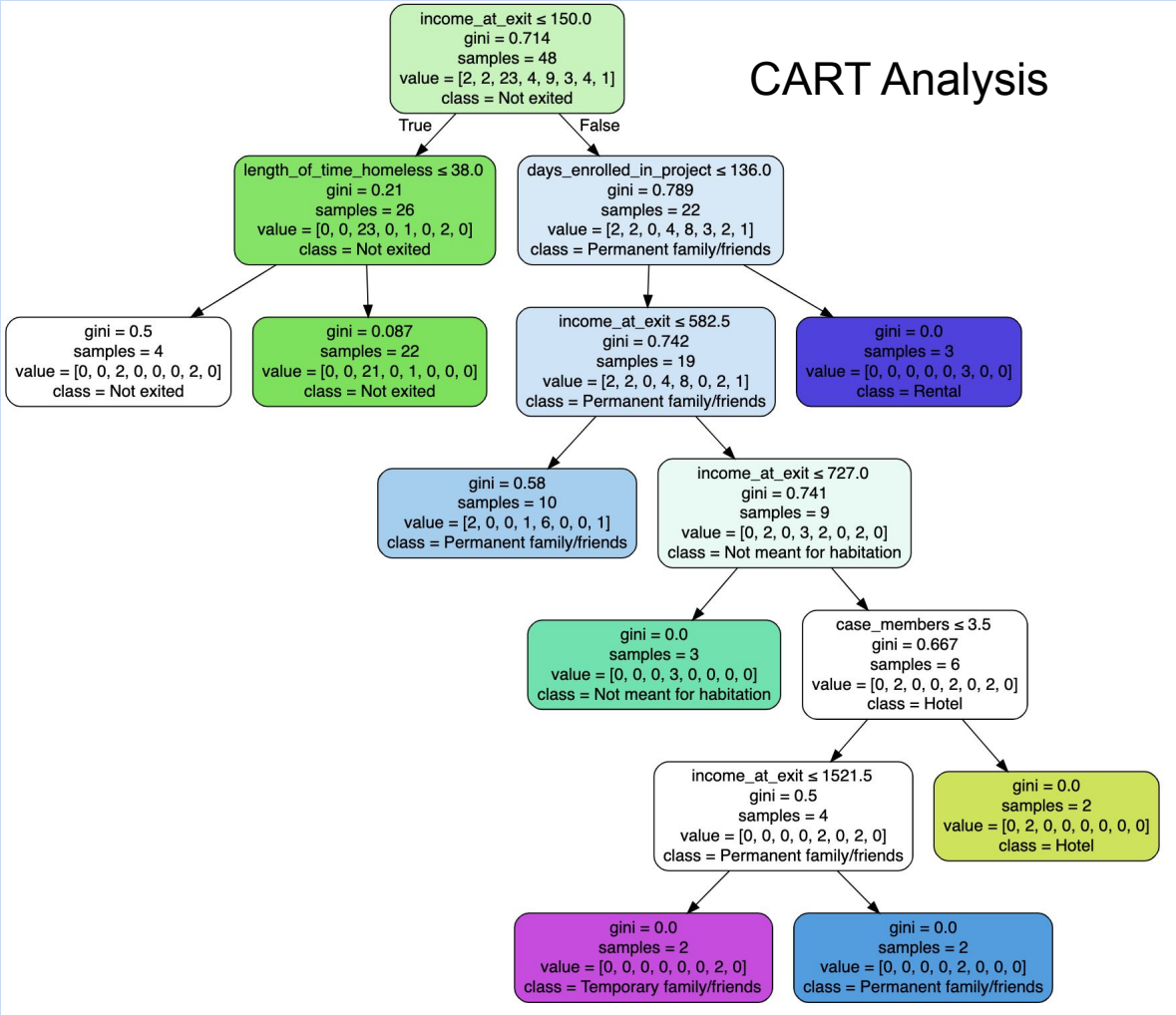
- In order to explore factors the influence a client's exit destination, we wanted to explore if income levels at entry and exit can be used to predict a clients housing situation.

- **Goal**: Predict a client's exit destination using fields related to their income

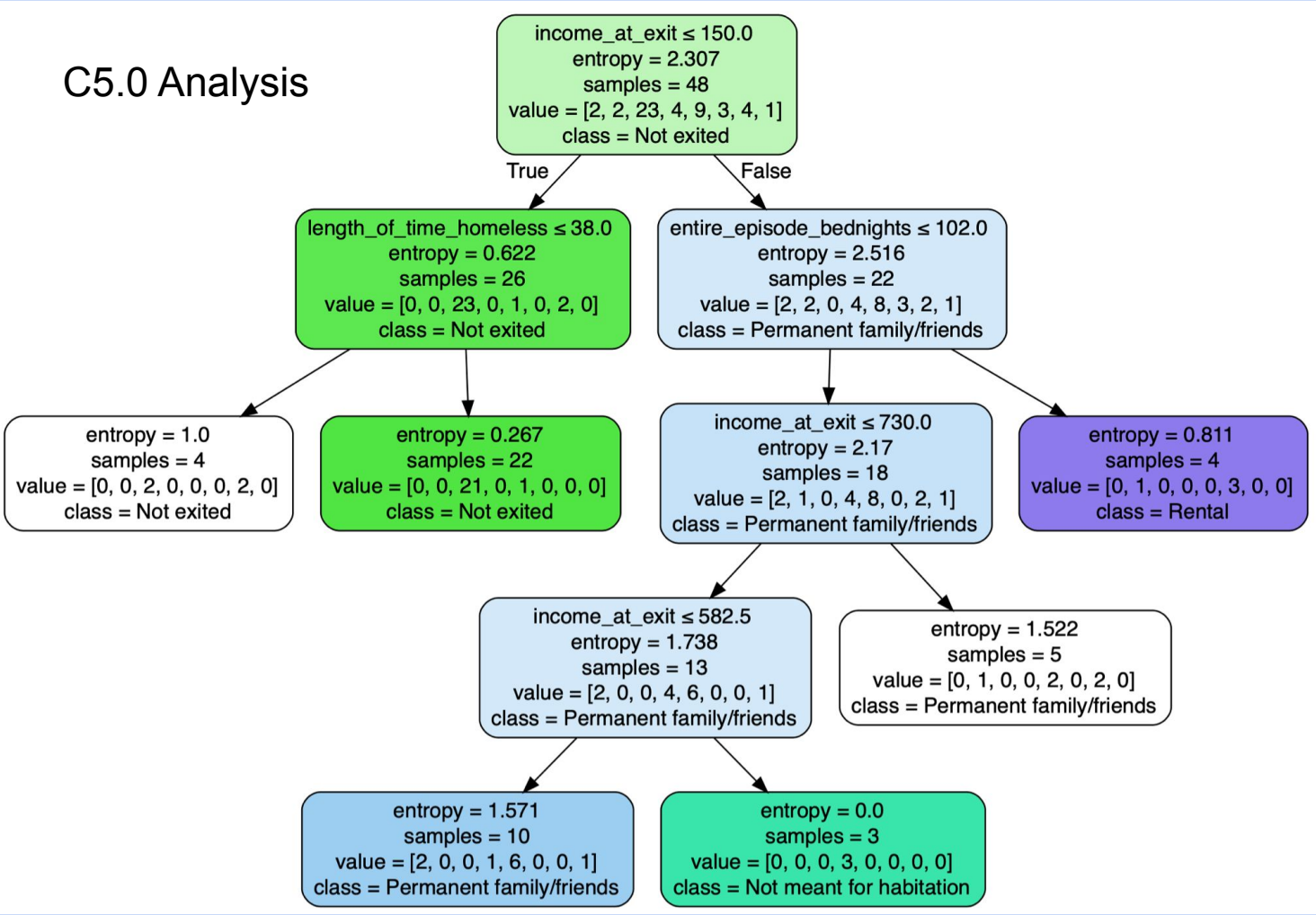
The Set-up

- Predictors:
  - Income_at_entry and income_at_exit
  - case_members
  - length_of_time_homeless
  - days_enrolled_in_project/number_of_bednights
- Create a subset of “self” records
- Replace empty fields
  - NA income values assumed to be 0
  - NA length of time homeless replaced with average
- Recategorize the exit destination
  - Combine like-destinations (ie. rental, permanent family/friends, temporary family/friends)
- Create calculated column ‘change_in_income’

CART Analysis



C5.0 Analysis



Accuracy = 76.47%

Cart

housing	Not exi…	Not mea…	Permane…	Rental
Client d…	0	1	0	0
Emergenc…	1	0	0	0
Hotel	1	0	0	0
Not exit…	10	0	0	0
Not mean…	0	0	1	0
Permanen…	0	0	1	0
Rental	0	0	0	1
Temporar…	0	0	1	0

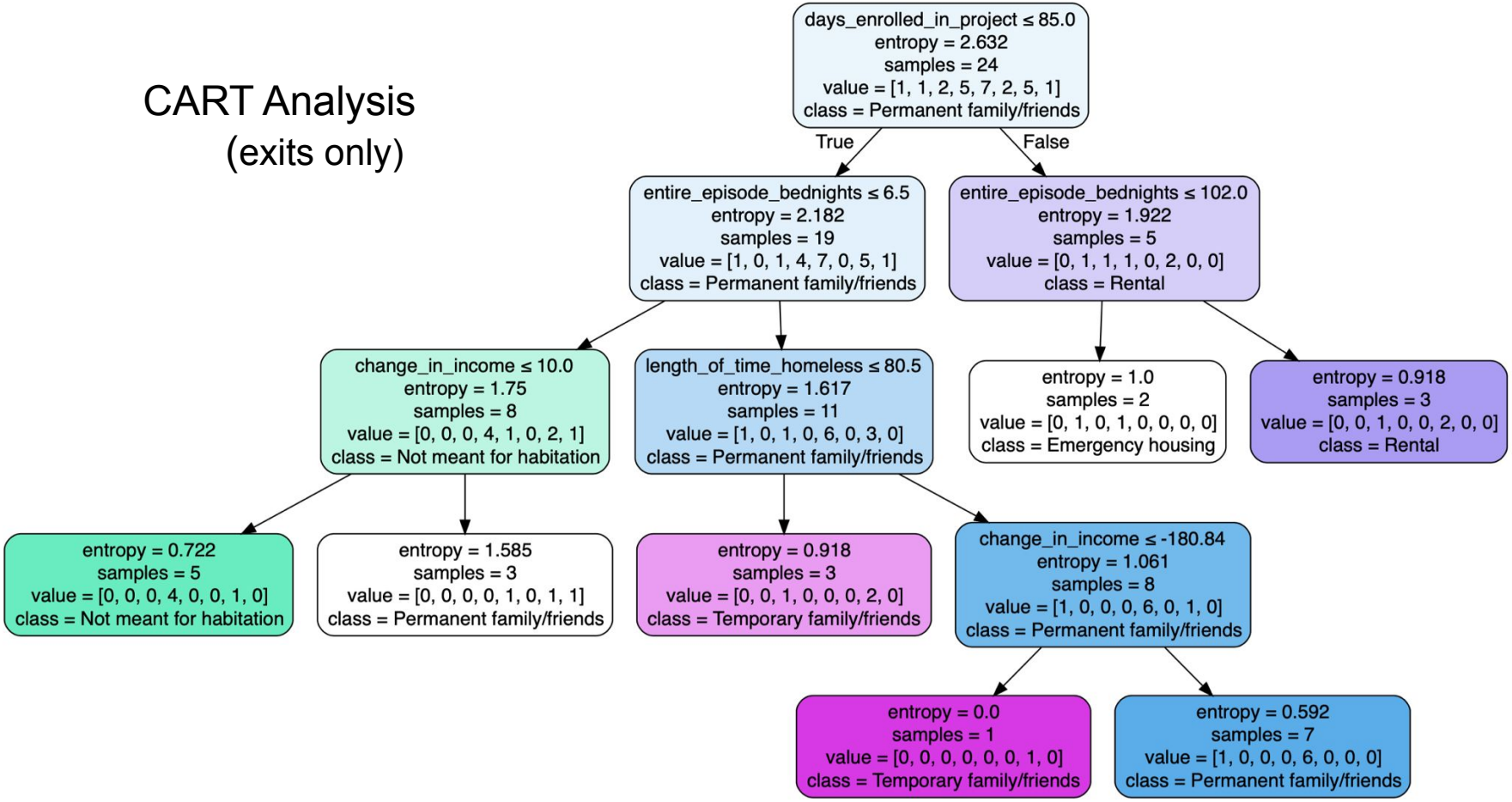
housing	Not exi…	Not mea…	Permane…	Rental
Client d…	0	100	0	0
Emergenc…	8.3	0	0	0
Hotel	8.3	0	0	0
Not exit…	83.3	0	0	0
Not mean…	0	0	33.3	0
Permanen…	0	0	33.3	0
Rental	0	0	0	100
Temporar…	0	0	33.3	0

C5.0

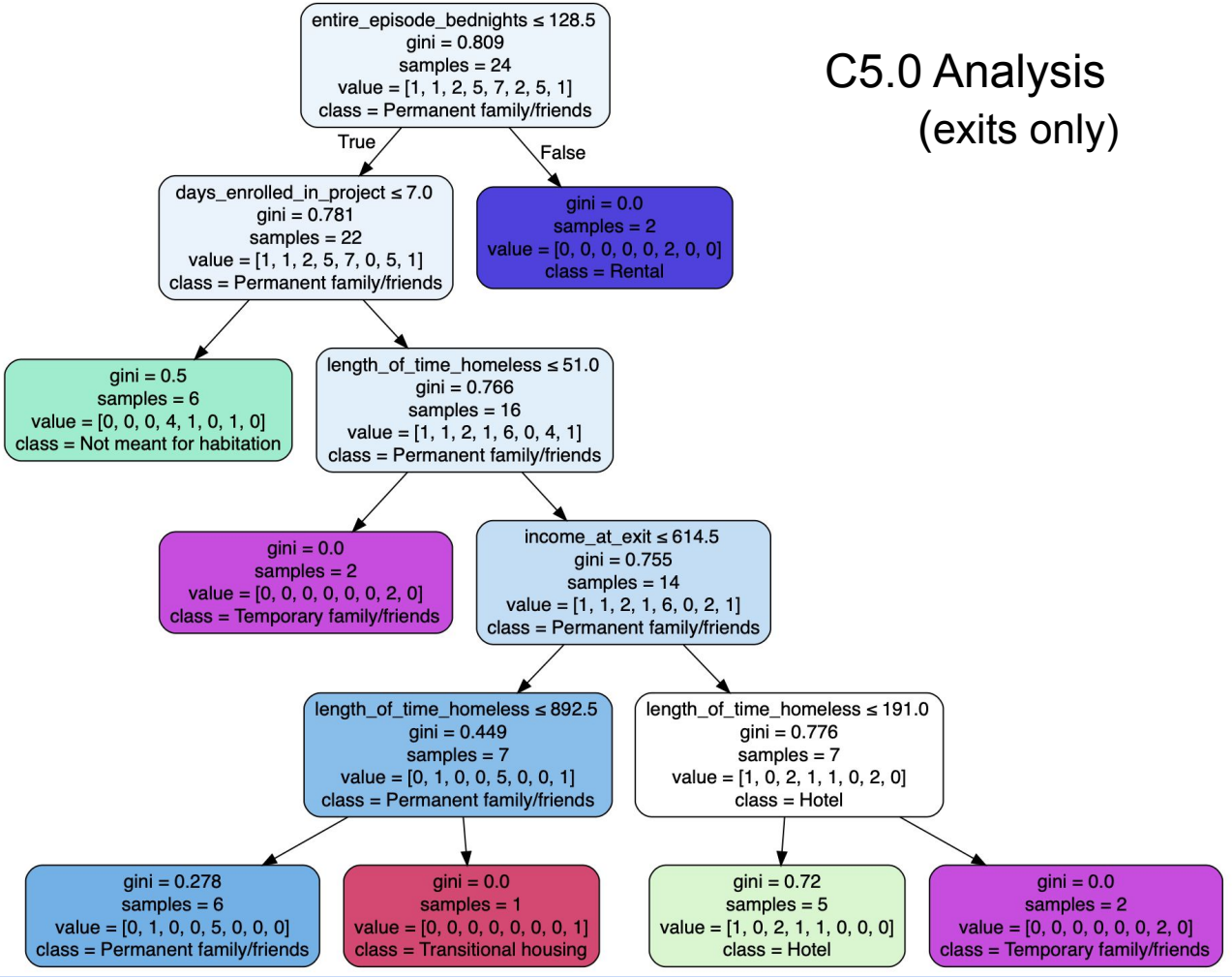
housing	Not exi…	Not mea…	Permane…	Rental
Client d…	0	1	0	0
Emergenc…	1	0	0	0
Hotel	1	0	0	0
Not exit…	10	0	0	0
Not mean…	0	0	1	0
Permanen…	0	0	1	0
Rental	0	0	0	1
Temporar…	0	0	1	0

housing	Not exi…	Not mea…	Permane…	Rental
Client d…	0	100	0	0
Emergenc…	8.3	0	0	0
Hotel	8.3	0	0	0
Not exit…	83.3	0	0	0
Not mean…	0	0	33.3	0
Permanen…	0	0	33.3	0
Rental	0	0	0	100
Temporar…	0	0	33.3	0

CART Analysis
(exits only)



C5.0 Analysis
(exits only)



Accuracy = 50.0%

housing	Not mea…	Permane…	Rental	Tempora…
Emergenc…	1	1	0	0
Hotel	0	1	0	0
Permanen…	1	0	0	2
Rental	0	0	2	0

housing	Not mea…	Permane…	Rental	Tempora…
Emergenc…	50	50	0	0
Hotel	0	50	0	0
Permanen…	50	0	0	100
Rental	0	0	100	0

Cart

housing	Not mea…	Permane…	Rental	Tempora…
Emergenc…	1	1	0	0
Hotel	0	1	0	0
Permanen…	0	2	0	1
Rental	0	0	2	0

housing	Not mea…	Permane…	Rental	Tempora…
Emergenc…	100	25	0	0
Hotel	0	25	0	0
Permanen…	0	50	0	100
Rental	0	0	100	0

C5.0

Neural Networks

- Same 6 inputs
- Two hidden layers of 4
- Output to 8 or 9

```
Epoch 145/150
5/5 [==============================] - 0s 4ms/step - loss: 0.2790 - accuracy: 0.5500
Epoch 146/150
5/5 [==============================] - 0s 2ms/step - loss: 0.2845 - accuracy: 0.5111
Epoch 147/150
5/5 [==============================] - 0s 3ms/step - loss: 0.2934 - accuracy: 0.4653
Epoch 148/150
5/5 [==============================] - 0s 2ms/step - loss: 0.3073 - accuracy: 0.4069
Epoch 149/150
5/5 [==============================] - 0s 2ms/step - loss: 0.2998 - accuracy: 0.4431
Epoch 150/150
5/5 [==============================] - 0s 2ms/step - loss: 0.3029 - accuracy: 0.4375
```

```
Epoch 145/150
3/3 [==============================] - 0s 2ms/step - loss: 0.4160 - accuracy: 0.2583
Epoch 146/150
3/3 [==============================] - 0s 2ms/step - loss: 0.4106 - accuracy: 0.3208
Epoch 147/150
3/3 [==============================] - 0s 3ms/step - loss: 0.4119 - accuracy: 0.3083
Epoch 148/150
3/3 [==============================] - 0s 3ms/step - loss: 0.4140 - accuracy: 0.3333
Epoch 149/150
3/3 [==============================] - 0s 2ms/step - loss: 0.4085 - accuracy: 0.3458
Epoch 150/150
3/3 [==============================] - 0s 3ms/step - loss: 0.4132 - accuracy: 0.3083
```

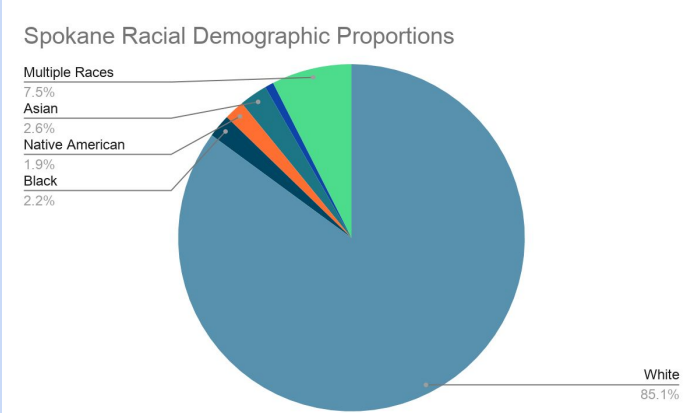
Discussion

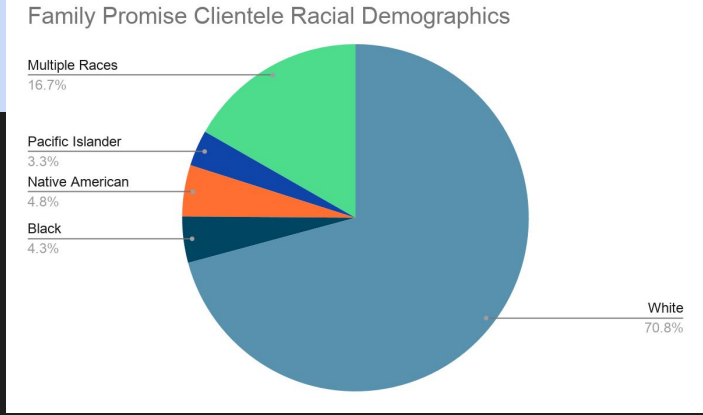
Group 3 Investigations

- Family Promise client race demographics vs. Spokane race demographics
- Family Promise client race demographics vs. National homeless race demographics
- Proportion of Family Promise clients struggling with mental illness vs. National proportion of people struggling with mental illness
- Proportion of Family Promise clients covered by health insurance vs. National proportion of people covered by health insurance

Client race demographics vs. Spokane race demographics

- Compare the proportions of each race for Spokane with proportions found among clientele
  - Source: “Spokane, WA Census Place.”
- Use Chi-square Goodness of Fit test to compare the expected counts (Spokane proportions*sample size) with observed counts
  - scipy.stats.chisquare(observed, expected)
- The proportions are significantly different

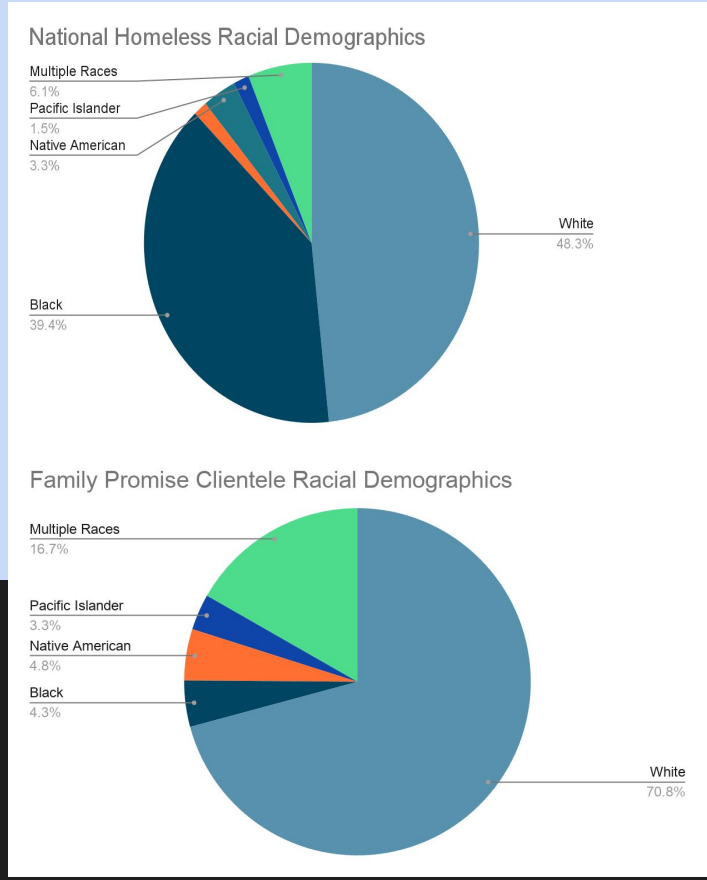





```
Spokane Race Proportions:
    White: 85.06%
    Black: 2.16%
    NativeAmerican: 1.87%
    Asian: 2.62%
    PacificIslander: 0.8%
    MultipleRaces: 7.48%
Family Promise Race Proportions:
    White: 70.81%
    Black: 4.31%
    NativeAmerican: 4.78%
    Asian: 0.0%
    PacificIslander: 3.35%
    MultipleRaces: 16.75%
H0: The race demographic proportions of family promise clients is equivalent to Spokane race demographic proportions.
H1: The race demographic proportions of family promise clients is not equivalent to Spokane race demographic proportions.
Using an alpha value of .05, I reject the null hypothesis in favor of the alternative hypothesis because the p-value of the chi-square goodness of fit test is 0.0
```

Client race demographics vs. National homeless race demographics

- Compare the proportions of each race in national homeless population with proportions found among clientele
  - Source: “2020-AHAR-Part-1: Point in Time Estimates of Homelessness in the U.S.”
- Use Chi-square Goodness of Fit test to compare the expected counts (National Homeless proportions*sample size) with observed counts
  - scipy.stats.chisquare(observed, expected)
- The proportions are significantly different




```
National Homeless Proportions:
    White: 48.34%
    Black: 39.42%
    NativeAmerican: 3.26%
    Asian: 1.32%
    PacificIslander: 1.51%
    MultipleRaces: 6.15%
Family Promise Race Proportions:
    White: 70.81%
    Black: 4.31%
    NativeAmerican: 4.78%
    Asian: 0.0%
    PacificIslander: 3.35%
    MultipleRaces: 16.75%
H0: The race demographic proportions of family promise clients is equivalent to the national homeless population race demographic proportions.
H1: The race demographic proportions of family promise clients is not equivalent to the national homeless population race demographic proportions.
Using an alpha value of .05, I reject the null hypothesis in favor of the alternative hypothesis because the p-value of the chi-square goodness of fit test is 0.0
```

Mental Illness

- Compare proportion of adult clientele(age 18 or older) who suffer a mental illness with national proportion(18.9%)
  - Source: "*Key Substance Use and Mental Health Indicators in the United States: Results from the 2017 National Survey on Drug Use and Health.*"
- Use 1-proportion-Ztest with observed counts against national proportion
  - statsmodels.stats.proportions.proportions_ztest(obs_count, obs_n, expected_prop, alternative="larger")
- Significantly larger proportion of clientele suffer with a mental illness than national average

```
H0: The proportion of adult family promise clients who suffer from mental illness is equivalent to 18.9%, the national proportion of adults over age 18 who suffer from mental illness.
H1: The proportion of adult family promise clients who suffer from mental illness is greater than 18.9%, the national proportion of adults over age 18 who suffer from mental illness.
Num adult clients w mental illness: 34, Total num adult clients: 95, Observed proportion: 35.8%
Using an alpha value of .05, I reject the null hypothesis in favor of the alternative hypothesis because the p-value of the one-proportion z-test is 0.0
```

Health Insurance Coverage

- Compare proportion of clientele who have health insurance with national proportion(92%)
  - Source: “Health Insurance Coverage in the United States: 2019.”
- Use 1-proportion-Ztest with observed counts against national proportion
  - statsmodels.stats.proportions.proportions_ztest(obs_count, obs_n, expected_prop, alternative=”smaller”)
- Significantly smaller proportion of clientele have health insurance than national proportion

```
H0: The proportion of family promise clients who are covered by health insurance is equivalent to 92%, the national proportion of people covered by health insurance.
H1: The proportion of family promise clients who are covered by health insurance is less than 92%, the national proportion of people covered by health insurance.
Num clients w health insurance: 172, Total num clients: 196, Observed proportion: 87.8%
Using an alpha value of .05, I reject the null hypothesis in favor of the alternative hypothesis because the p-value of the one-proportion z-test is 0.035
```

Potential Directions

- Alcohol and Drug abuse
  - Possibly combined with mental illness
- Veterans
  - Mental illness
  - Race demographics
  - Alcohol and drug abuse
- Domestic Violence
  - Prevalence
  - Proportion ‘currently fleeing’
- Gender or Age
  - Combined with all other categories of investigation
- More Data!

Group 3 Conclusions

- Clientele racial demographic significantly different from Spokane demographics
  - Less White and Asian, more Black, Native American, Pacific Islander, and Multi-Racial
- Clientele racial demographic significantly different from National homeless demographics
  - Less Black and Asian, more White, Native American, Pacific Islander, and Multi-Racial
- Clientele struggle with mental illness at a higher rate than national average
- Proportion of clientele with health insurance is less than national average
  - Detrimental combination with mental illness rates
  - Clientele poorly equipped to treat mental illness

Sources

Selecting pandas DataFrame Rows Based On Conditions. CHRIS ALBON, 20 Dec. 2017, chrisalbon.com/python/data_wrangling/pandas_selecting_rows_on_conditions/

Create calculated columns in a dataframe. Python In Office, 29 Aug. 2020, pythoninoffice.com/create-calculated-columns-in-a-dataframe/

How to convert a pandas DataFrame column of strings to floats in Python. Kite, www.kite.com/python/answers/how-to-convert-a-pandas-dataframe-column-of-strings-to-floats-in-python

pandas.DataFrame.sort_values. pandas, 2021, pandas.pydata.org/docs/reference/api/pandas.DataFrame.sort_values.html

Family Promise Spokane. 2021, https://www.familypromiseofspokane.org. Accessed 17 May. 2021.

"Get feature and class names into decision tree using graphiz". *Stackoverflow, 2017. https://stackoverflow.com/questions/39476020/get-feature-and-class-names-into-decision-tree-using-export-graphviz. Accessed 17 May. 2021.*

Sources cont.

"2020-AHAR-Part-1: Point in Time Estimates of Homelessness in the U.S." *U.S. Department of Housing and Urban Development*, Office of Community Planning and Development, Mar. 2021, www.huduser.gov/portal/sites/default/files/pdf/2020-AHAR-Part-1.pdf.

Keisler-Starkey, Katherine, and Lisa N Bunch. "Health Insurance Coverage in the United States: 2019." *The United States Census Bureau*, 15 Sept. 2020, www.census.gov/library/publications/2020/demo/p60-271.html.

Substance Abuse and Mental Health Services Administration, et al. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2017 National Survey on Drug Use and Health*. Sept. 2018, www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHFFR2017/NSDUHFFR2017.pdf.

"Spokane, WA Census Place." *Data USA*, datausa.io/api/data?Geography=16000US5367000&drilldowns=Race,Ethnicity&measures=Hispanic Population,Hispanic Population Moe.