Group Project: Data Selection and Initial Analysis

Group Number: 09

Sharvil Kishor Wadekar
Mohit Ravindra Kamble
Soham Santosh Mane

College of Professional Studies – Northeastern University

ALY6015: Intermediate Analytics

Prof. Roy Wada

January 23, 2024

01. Group Members:

- ❖ Sharvil Kishor Wadekar (NUID: 002897587)
- ❖ Mohit Ravindra Kamble (NUID: 002836839)
- ❖ Soham Santosh Mane (NUID: 002844811)

02. The dataset you will use for the final project:

★ **Property Assessment FY2019CSV**

The objective of this predictive analysis is to explore and predict property values in the city of Boston. We will focus on two outcome variables: AV_TOTAL (total assessed value) and OWN_OCC (owner-occupied property). We aim to identify predictors that significantly impact these outcomes and analyze sub-groups based on relevant characteristics.

03. Project Title

★ **Data Analysis of Boston City (FY-2019)**

04. Outcome Variables

➔ Total Assessed Value (AV_TOTAL): This represents the total assessed value of the property, including land and building.

➔ Owner-Occupied Property (OWN_OCC): A binary variable indicating whether the property is owner-occupied.

05. Predictor Variables

We will consider the following predictor variables for our analysis:

- LU (Type of Property)
- GROSS_AREA (Gross floor area for commercial properties)
- YR_BUILT (Year property was built)
- LAND_SF (Parcel's land area in square feet)
- NUM_FLOORS (# of levels in the structure located on the parcel)
- R_BDRMS (Total number of bedrooms in the structure)
- R_FULL_BTH (Total number of full baths in the structure)
- R_HALF_BTH (Total number of half baths in the structure)
- R_HEAT_TYP (Structure heat type)
- R_AC (Indicates if the structure has air conditioning)

06. Feature Engineering

For AV_TOTAL, we may need to transform it into a binary outcome of high vs. low value based on a threshold.

07. Descriptive Statistics Tables

1. **All Sample:**
    - Summary statistics for AV_TOTAL and OWN_OCC.
    - Descriptive statistics for predictor variables.
2. **By Group (e.g., LU Type):**
    - Separate tables for each LU type, summarizing AV_TOTAL and OWN_OCC.
    - Descriptive statistics for predictor variables within each group.

08. Sub-Group Analysis

We will perform sub-group analysis based on relevant characteristics such as property type (LU), year built (YR_BUILT), and number of bedrooms (R_BDRMS). This will help identify variations in outcome variables within these sub-groups.

09. Analytical Plans and Methods

1. Chi-Square Test for OWN_OCC

- **Objective:** To examine the association between the type of property (LU) and the likelihood of being owner-occupied (OWN_OCC).
- **Procedure:**
    - Form a contingency table between LU and OWN_OCC.
    - Apply the Chi-Square test to determine if there is a significant association.

2. ANOVA Test for AV_TOTAL

- **Objective:** To assess if there are significant differences in the mean total assessed values (AV_TOTAL) among different property types (LU).
- **Procedure:**
    - Conduct a one-way ANOVA test.
    - Examine the F-statistic and p-value to determine statistical significance.

3. ANOVA Test for AV_TOTAL by R_BDRMS Sub-Groups

- **Objective:** To explore if the number of bedrooms (R_BDRMS) significantly influences the mean total assessed value (AV_TOTAL).
- **Procedure:**
    - Divide the dataset into sub-groups based on the number of bedrooms.
    - Conduct ANOVA tests within each sub-group.

10. Conclusion:

By employing the Chi-Square test and ANOVA test, we aim to uncover relationships between predictor and outcome variables. The results will provide insights into the impact of property type (LU) on owner-occupancy (OWN_OCC) and the influence of property type and number of bedrooms on the total assessed value (AV_TOTAL). These tests contribute to a more in-depth understanding of the dataset.

11. Citations:

➢ Dataset: source.
➢ Chi-Square Test: Lab video
➢ ANOVA Test: Lab video

12. Appendix:

```
cat("\014") # clears console
rm(list = ls()) # clears global environment
try(dev.off(dev.list()["RStudioGD"]), silent = TRUE) # clears plots
try(p_unload(p_loaded(), character.only = TRUE), silent = TRUE) # clears packages
options(scipen = 100) # disables scientific notation for entire R session

# >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>Group Project<<<<<<<<<<<<<<<<<<

boston <- read.csv("D:/NEU STUDY/2nd Quarter/Intermediate Analytics (ALY
6015)/Group Project/fy19fullpropassess.csv")

# Data Summary
str(boston)
summary(boston)

# Univariate Analysis
hist(boston$AV_TOTAL, main = "Distribution of AV_TOTAL", xlab = "Total Assessed
Value")
barplot(table(boston$OWN_OCC), main = "Distribution of OWN_OCC", xlab = "Owner
Occupied (Y/N)")

# Removing Descriptive Statistics for All Sample
summary(boston[c("AV_TOTAL", "GROSS_AREA", "YR_BUILT", "R_BDRMS")])

# Removing Descriptive Statistics by Property Type (LU)
for (property_type in unique(boston$LU)) {
  subset_data <- subset(boston, LU == property_type)
  print(paste("Descriptive Statistics for", property_type))
  print(summary(subset_data[c("AV_TOTAL", "GROSS_AREA", "YR_BUILT",
"R_BDRMS")]))
}


# Removing Descriptive Statistics by Bedroom Count
for (bedroom_count in unique(boston$R_BDRMS)) {
  subset_data <- subset(boston, R_BDRMS == bedroom_count)
  print(paste("Descriptive Statistics for", bedroom_count, "Bedrooms"))
  print(summary(subset_data[c("AV_TOTAL", "GROSS_AREA", "YR_BUILT")]))
}
```

```
# Analysis
# Chi-Square Test
chi_square_result <- chisq.test(table(boston$LU, boston$OWN_OCC))
print(chi_square_result)

# ANOVA Test
anova_result <- aov(AV_TOTAL ~ LU, data = boston)
print(summary(anova_result))

# ANOVA Test for Sub-Groups
for (bedroom_count in unique(boston$R_BDRMS)) {
  subset_data <- subset(boston, R_BDRMS == bedroom_count)

  # Checking if the subset has at least two levels for LU
  if (length(unique(subset_data$LU)) >= 2) {
    anova_result <- aov(AV_TOTAL ~ LU, data = subset_data)
    print(paste("ANOVA Results for", bedroom_count, "Bedrooms:"))
    print(summary(anova_result))
  } else {
    print(paste("Insufficient levels for ANOVA in subset with", bedroom_count,
"Bedrooms."))
  }
}
```