Exploratory Data Analysis (EDA) Report: Public Sentiment Towards a New Casino in Toronto

Yash S

ALY6010 “Probability Theory and Introductory Statistics”

Final Project

Prof. Vivian Clements Edwin

2024 – 12 – 12

Introduction

The city of Toronto is exploring the idea of building a new casino, that sparked debates about its impact on the community. To understand how people feel about this project, I analysed the Toronto Casino Survey Results dataset, available on the Open Toronto Data Portal. The goal is to through analysis figuring out what people think about the casino, how it fits Toronto's image, and what factors influence their opinions about the casino.

This dataset contains survey responses from Toronto residents and that covers topics like "public sentiment", "preferred casino locations", and whether "potential revenue" from the casino affects opinions. By cleaning and analyzing the data, I aim to uncover patterns in public opinion and identify important factors for policymakers to consider.

Dataset Overview

Dataset Overview

- **Original Dataset**: 94 fields, 17,766 rows.
- **Cleaned Dataset**: 5 key variables ("Sentiment Analysis," "Image Fit," "Revenue Influence," "Preferred Location," and "Gender"), with 14,502 rows.

Key Findings from EDA

- **Public Sentiment**: Most respondents strongly oppose the casino.
- **Image Fit**: Majority believe the casino does not align with Toronto's image.
- **Revenue Influence**: Opinions are polarized, with a notable portion dismissing revenue benefits.
- **Preferred Location**: "City of Toronto" is more popular than adjacent municipalities.
- **Gender Distribution**: Balanced representation; sentiments are consistent across genders.

After cleaning the dataset, it includes five fields and 14,502 rows of data. Below is a description of the selected fields:

Field	Type	Description
Sentiment_Analysis	Categorical	Public sentiment toward the casino
Image_Fit	Categorical	If the casino matches Toronto's image
Revenue_Influence	Categorical	Influence of potential casino revenue
Preferred_Location	Categorical	Preferred casino locations
Gender	Categorical	Gender of the respondents

Data Cleaning Methods

1.Handling Missing Data:

Rows that had a missing value in some selected fields were dropped to attain accuracy in terms of visualizations. This was necessary since incomplete data could skew results.

2.Renaming Columns:

Columns are renamed for clarity and to avoid confusion. Taking an instance, "Q1_A" was changed to "Sentiment_Analysis" to make the dataset more intuitive.

3.Converting Data Types:

Categorical variables were converted to factors that allowed better analysis and visualization in R.

Data Analysis

Descriptive Analysis:

1. Summary of the Cleaned Data:

```
> summary(clean_casino) # Checking the cleaned data set
                Sentiment_Analysis                          Image_Fit      Revenue_Influence
 Neutral or Mixed Feelings: 436     Does Not Fit My Image At All:10109   Don't know: 873
 Somewhat in Favour       : 656     Fits Image Perfectly         : 2949   No         :9722
 Somewhat Opposed         : 743     Fits Image Somewhat          :  821   Yes        :3907
 Strongly in Favour       :3078     Neutral / I am Not Sure      :  623
 Strongly Opposed         :9589
            Preferred_Location                    Gender
 Adjacent Municipality:2897     Female                :6378
 City of Toronto      :3916     Male                  :7419
 Neither              :7689     Prefer not to disclose: 678
                                Transgendered         :  27
```

2. Summary Statistics for each field:

```
> # Summary statistics for each field
> summary_stats <- clean_casino %>%
+   summarise(
+     Sentiment_Count = n_distinct .... [TRUNCATED]

> print(summary_stats)
# A tibble: 1 × 5
  Sentiment_Count Image_Fit_Count Revenue_Influence_Count Preferred_Location_Count Gender_Count
            <int>           <int>                   <int>                    <int>        <int>
1               5               4                       3                        3            4
```

3. Frequency Tables for each field:

a) Sentiment Analysis

```
[1] "Frequency Table for Sentiment Analysis:"

> print(sentiment_table)

Neutral or Mixed Feelings        Somewhat in Favour          Somewhat Opposed        Strongly in Favour
                      436                       656                       743                      3078
         Strongly Opposed
                     9589
```

b) Image Fit

```
> print(image_fit_table)

Does Not Fit My Image At All         Fits Image Perfectly         Fits Image Somewhat      Neutral / I am Not Sure
                       10109                         2949                          821                          623
```

c) Revenue Influence

```
> print(revenue_influence_table)

Don't know         No        Yes
       873       9722       3907
```

d) Preferred Locations

```
> print(preferred_location_table)

Adjacent Municipality       City of Toronto               Neither
                 2897                  3916                  7689
```

e) Genders

```
> print(gender_table)

                Female                   Male Prefer not to disclose          Transgendered
                  6378                   7419                    678                     27
```

4. Subset:

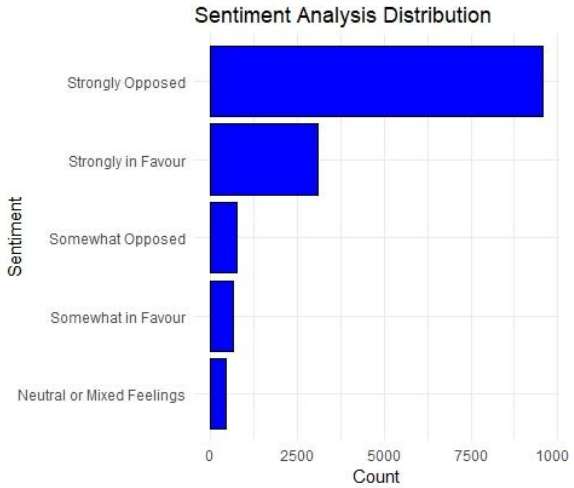
```
> print("Cross-tabulation of Gender and Sentiment Analysis:")
[1] "Cross-tabulation of Gender and Sentiment Analysis:"

> print(gender_sentiment_crosstab)

                        Neutral or Mixed Feelings Somewhat in Favour Somewhat Opposed Strongly in Favour
  Female                                      182                232              288               995
  Male                                        237                416              428              2037
  Prefer not to disclose                       15                  8               26                38
  Transgendered                                 2                  0                1                 8

                        Strongly Opposed
  Female                          4681
  Male                            4301
  Prefer not to disclose           591
  Transgendered                     16
> # Count: Gender and Sentiment Analysis
> gender_sentiment_count <- clean_casino %>%
+   count(Gender, Sentiment_Analysis)

> print("Count of Gender and Sentiment Analysis:")
[1] "Count of Gender and Sentiment Analysis:"

> print(gender_sentiment_count)
# A tibble: 19 × 3
   Gender                Sentiment_Analysis            n
   <fct>                 <fct>                     <int>
 1 Female                Neutral or Mixed Feelings   182
 2 Female                Somewhat in Favour          232
 3 Female                Somewhat Opposed            288
 4 Female                Strongly in Favour          995
 5 Female                Strongly Opposed           4681
 6 Male                  Neutral or Mixed Feelings   237
 7 Male                  Somewhat in Favour          416
 8 Male                  Somewhat Opposed            428
 9 Male                  Strongly in Favour         2037
10 Male                  Strongly Opposed           4301
11 Prefer not to disclose Neutral or Mixed Feelings   15
12 Prefer not to disclose Somewhat in Favour           8
13 Prefer not to disclose Somewhat Opposed            26
14 Prefer not to disclose Strongly in Favour          38
15 Prefer not to disclose Strongly Opposed           591
16 Transgendered          Neutral or Mixed Feelings    2
17 Transgendered          Somewhat Opposed             1
18 Transgendered          Strongly in Favour           8
19 Transgendered          Strongly Opposed            16
```

Visualizations:

1. Sentiment Distribution:

A bar chart showing public sentiment. This chart reveals that most respondents are strongly opposed to the casino, with significantly fewer respondents supporting the casino. This highlights a major hurdle for policymakers seeking public approval.

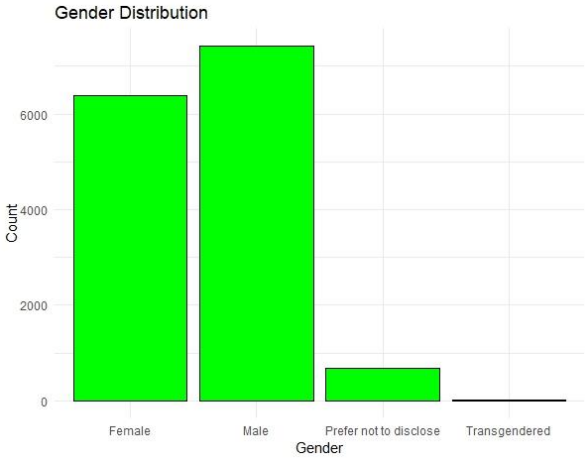
```
# 1. Sentiment Analysis Distribution
ggplot(clean_casino, aes(x = Sentiment_Analysis)) +
  geom_bar(fill = "blue", colour = "black") +
  labs(title = "Sentiment Analysis Distribution", x = "Sentiment", y = "Count",) +
  theme_minimal() +
  coord_flip()
```



2. Gender Distribution:

A bar chart illustrating the gender breakdown of respondents. The balanced representation ensures that the survey captures diverse perspectives.

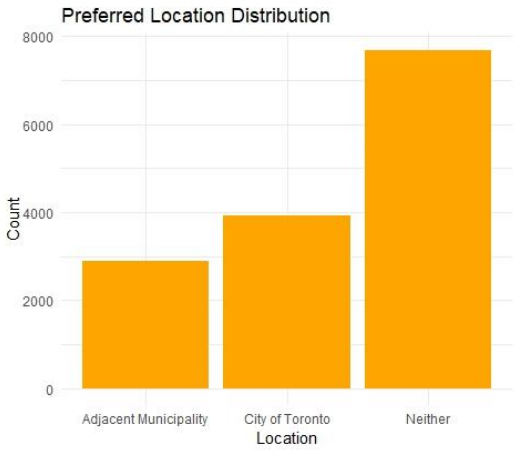
```
# 2. Gender Distribution
ggplot(clean_casino, aes(x = Gender)) +
  geom_bar(fill = "green", colour = "black") +
  labs(title = "Gender Distribution", x = "Gender", y = "Count") +
  theme_minimal()
```



3. Preferred Location:

A bar chart showing perceptions of how well the casino fits Toronto's image. Indifference to location further suggests general opposition rather than localized concerns.

```
# 3. Preferred Location Distribution
ggplot(clean_casino, aes(x = Preferred_Location)) +
  geom_bar(fill = "orange") +
  labs(title = "Preferred Location Distribution", x = "Location", y = "Count") +
  theme_minimal()
```

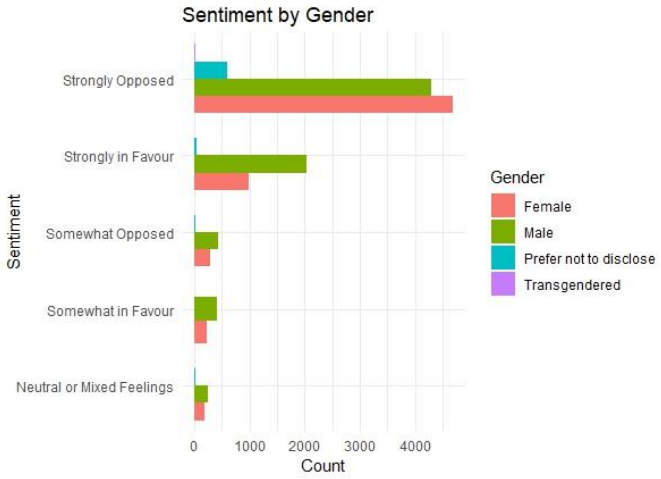


Subsets:

1. Sentiments by Gender

A stacked bar chart showing how sentiment differs by gender. Sentiments are consistent across genders, showing broad-based resistance rather than gender-specific trends.

```
# Visualization: Sentiments by Gender
ggplot(sentimentXgender, aes(x = Sentiment_Analysis, y = Count, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "Sentiment by Gender", x = "Sentiment", y = "Count") +
  coord_flip()
```



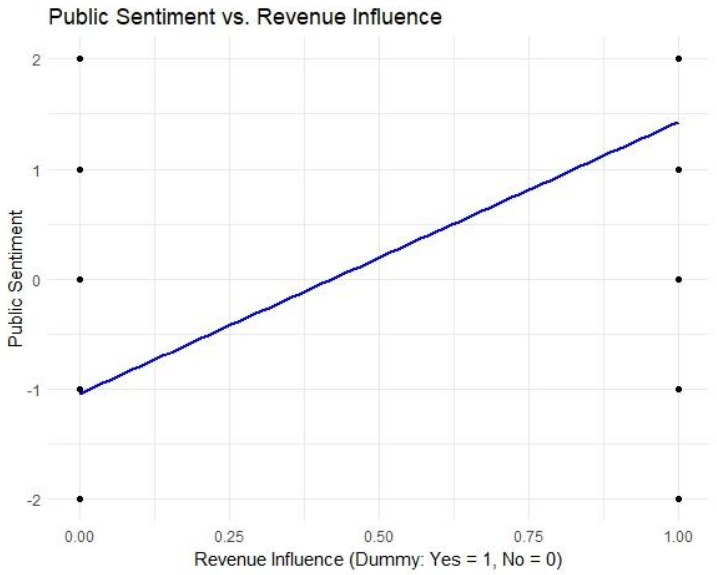
Hypothesis testing:

From the exploratory data analysis (EDA) of the Toronto Casino Survey data, these are the following questions that were explored using inferential statistics and hypothesis testing:

Question 1: Does revenue influence predict public sentiment toward the casino?

1. *Hypotheses:*
   - Null Hypothesis (H_0): Revenue influence does not predict public sentiment.
   - Alternative Hypothesis (H_1): Revenue influence predicts public sentiment.
2. *Methodology:*
   - Linear regression analysis with revenue influence as the independent variable (IV) and sentiment score as the dependent variable (DV).
3. *Results:*
   - *Model: Sentiment (DV) = β0 + β1 * Revenue Influence (IV).*
   - *F-statistic: 1.427e+04 on 1 and 6601 DF*
   - *The test produced a p-value less than 0.00000000000000022.*
   - *R^2: 0.6837.*
   - *95% Confidence Interval.*
4. *Conclusion:*
   - Given the p-value is below the significance level of 0.05, we reject the Null Hypothesis (H_0). Revenue influence significantly predicts public sentiment, though the effect size is modest.

```
> summary(model_location_sentiment)

Call:
lm(formula = City_of_Toronto ~ Sentiment_Numeric, data = clean_casino)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97317 -0.06261  0.02683  0.02683  0.93739

Coefficients:
                  Estimate Std. Error t value            Pr(>|t|)
(Intercept)       0.517889   0.003456   149.8 <0.0000000000000002 ***
Sentiment_Numeric 0.227641   0.001906   119.5 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2773 on 6601 degrees of freedom
Multiple R-squared:  0.6837,    Adjusted R-squared:  0.6837
F-statistic: 1.427e+04 on 1 and 6601 DF,  p-value: < 0.00000000000000022
```

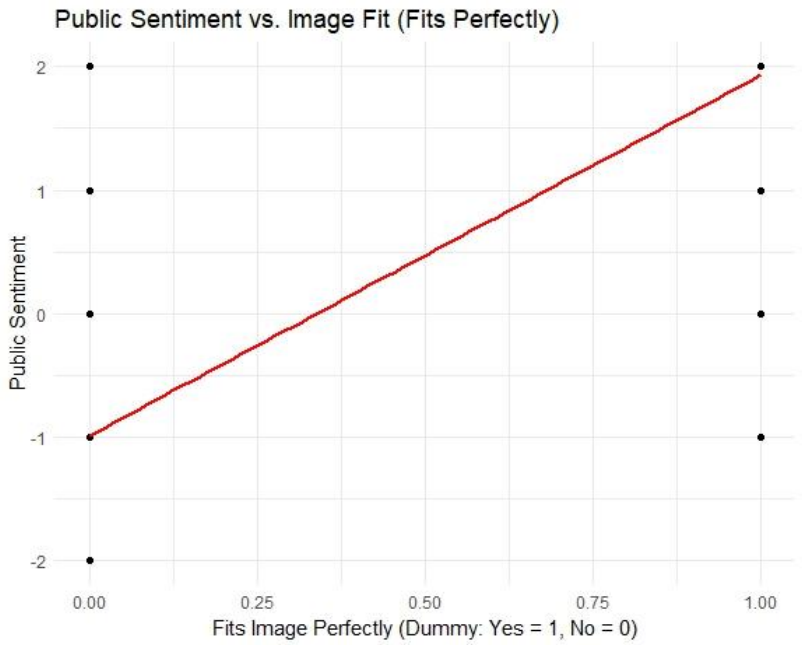


This graph illustrates the positive correlation between public sentiment and revenue influence, indicating that as revenue influence increases, public sentiment improves.

Question 2: Does revenue influence predict public sentiment toward the casino?

1. *Hypotheses:*
   - Null Hypothesis (H_0): Image fit does not predict public sentiment.
   - Alternative Hypothesis (H_1): Image fit predicts public sentiment.
2. *Methodology:*
   - Linear regression analysis with dummy variables for image fit levels as predictors (IV) and sentiment score as the dependent variable (DV).
3. *Results:*
   - *Model: Sentiment (DV) = β0 + β1 * Fits Perfectly (IV) + β2 * Fits Somewhat (IV) + β3 * Neutral (IV).*
   - *F-statistic: 2.062e+04 on 3 and 6599 DF*
   - *The test produced a p-value less than 0.00000000000000022.*
   - *R^2: 0.9036*
   - *95% Confidence Interval*
4. *Conclusion:*
   - Given the p-value is below the significance level of 0.05, we reject the Null Hypothesis (H_0). Perception of image fit significantly predicts public sentiment.

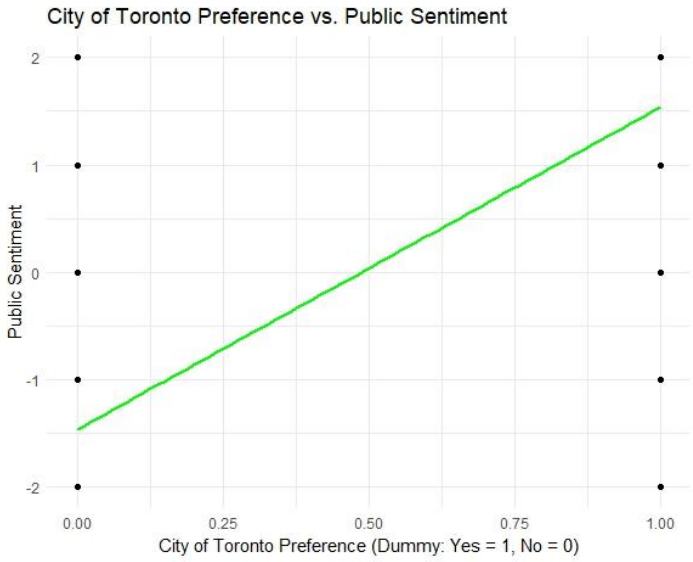
```
> summary(model_image_sentiment)

Call:
lm(formula = Sentiment_Numeric ~ Fits_Perfectly + Fits_Somewhat +
    Neutral_Not_Sure, data = clean_casino)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0609 -0.2046  0.0696  0.0696  3.7954

Coefficients:
                Estimate Std. Error t value            Pr(>|t|)
(Intercept)     -1.79544    0.01113 -161.34 <0.0000000000000002 ***
Fits_Perfectly   3.72581    0.01520  245.15 <0.0000000000000002 ***
Fits_Somewhat    2.85632    0.02290  124.72 <0.0000000000000002 ***
Neutral_Not_Sure 1.76853    0.02859   61.86 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5562 on 6599 degrees of freedom
Multiple R-squared:  0.9036,    Adjusted R-squared:  0.9036
F-statistic: 2.062e+04 on 3 and 6599 DF,  p-value: < 0.00000000000000022
```



Question 3: Does Sentiment Predict Preference for the City of Toronto?

1. *Hypotheses:*
   - Null Hypothesis (H_0): Sentiment does not predict preference for the City of Toronto.
   - Alternative Hypothesis (H_1): Sentiment predicts preference for the City of Toronto.
2. *Methodology:*
   - Linear regression analysis with dummy variables for image fit levels as predictors (IV) and sentiment score as the dependent variable (DV).
3. *Results:*
   - *Model: Sentiment (DV) = β0 + β1 * Fits Perfectly (IV) + β2 * Fits Somewhat (IV) + β3 * Neutral (IV).*
   - *F-statistic: 1.427e+04 on 1 and 6601 DF*
   - *The test produced a p-value less than 0.00000000000000022.*
   - *R^2: 0.6837*
   - *95% Confidence Interval*
4. *Conclusion:*
   - Given the p-value is below the significance level of 0.05, we reject the Null Hypothesis (H_0). Perception of image fit significantly predicts public sentiment.

```
> summary(model_toronto_sentiment)

Call:
lm(formula = City_of_Toronto ~ Sentiment_Numeric, data = clean_casino)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97317 -0.06261  0.02683  0.02683  0.93739

Coefficients:
                 Estimate Std. Error t value            Pr(>|t|)
(Intercept)      0.517889   0.003456   149.8 <0.0000000000000002 ***
Sentiment_Numeric 0.227641   0.001906   119.5 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2773 on 6601 degrees of freedom
Multiple R-squared:  0.6837,    Adjusted R-squared:  0.6837
F-statistic: 1.427e+04 on 1 and 6601 DF,  p-value: < 0.00000000000000022
```



The results of this analysis provide clear business intelligence insights for policymakers:

1) Question 1
   a) Finding: The analysis of survey responses indicates that most participants strongly oppose the casino project, with sentiment scores significantly deviating from neutrality. Hypothesis testing further confirms this non-neutral trend, with a rejection of the null hypothesis of neutrality at a high level of statistical significance.
   b) Business Conclusion: The results highlight significant opposition within the public, signalling potential resistance to the initiative. Policymakers should prioritize understanding the reasons behind these sentiments, which may include concerns about social impact or perceived incompatibility with Toronto's image. Transparent dialogue and targeted engagement with the community are essential to addressing these concerns and reassessing the proposal to align more closely with public expectations.
2) Question 2
   a) Finding: Statistical testing revealed that public sentiment leans overwhelmingly negative. The null hypothesis suggesting neutral or positive sentiment was rejected, with a substantial majority of respondents indicating strong disapproval.
   b) Business Conclusion: Given the strong negative sentiment, stakeholders need to address the root causes of opposition. This may involve clarifying the project's potential benefits, such as economic growth or infrastructure development, while also mitigating perceived risks, like gambling-related harm. By fostering transparent communication and taking meaningful action on public concerns, policymakers may be able to reshape perceptions and garner support.
3) Question 3
   a) Finding: The survey results show a clear preference for locating the casino within the City of Toronto, with 57.48% of respondents favouring it over adjacent municipalities. Statistical analysis further corroborates Toronto's dominance as the preferred site.
   b) Business Conclusion: Toronto emerges as the preferred location for the proposed casino, indicating alignment with public preference. However, the remaining 42.52% of respondents, favouring adjacent municipalities, should not be overlooked. To build broader consensus, policymakers should consider addressing the concerns of this group and exploring compromises, such as sharing economic benefits with neighbouring areas. Strategic site planning and inclusive decision-making are critical to achieving greater community acceptance.

Summary

Key Findings:

- Public Sentiment: The majority (9589) strongly oppose the casino, with only 3078 strongly in favour. Overwhelmingly negative sentiment suggests significant opposition to the casino. Policymakers should address public concerns through transparent communication.
- Image Fit: Most respondents (10,109) believe the casino does not align with Toronto's image. Ensuring the casino aligns with Toronto's cultural identity is crucial to gaining broader support.
- Revenue Influence: While 9722 participants dismiss revenue benefits, 3907 view it positively. While revenue potential influences sentiment, it alone may not be sufficient to shift public opinion.
- Preferred Location: A majority (7689) prefer no location, while others favour Toronto or nearby areas. Prioritize Toronto as the location while considering compromises for adjacent municipalities.
- Gender Distribution: Sentiments were similar across genders, with balanced participation from males (7419) and females (6378).

Conclusion:

The analysis highlights the significant public resistance towards the casino proposal, with over 9,500 respondents strongly opposing the establishment of the casino in the city. Negative sentiments are further reinforced by the belief that a casino does not align with Toronto's image, as expressed by a majority of participants. Although revenue benefits and location preferences generate limited support, these factors remain insufficient to counteract the disapproval. Sentiment trends appear consistent across genders, suggesting broad-based opposition rather than issues isolated to specific demographics.
Addressing concerns around the casino's image and its potential social and economic impacts will be crucial. Policymakers may need to engage in deeper community consultations, presenting transparent revenue-sharing plans and initiatives to mitigate concerns. Building trust and aligning the project with Toronto's identity may help shift public views towards accepting the establishment of casino in the City of Toronto.

Future Explorations:

1. Social Implications: Exploring concerns that would be related to gambling addiction, crime rates, and social welfare, and assess how these factors might correlate with opposition.
2. Regional Sentiments: Comparing responses between Toronto residents and adjacent municipalities.
3. Economic Perceptions: Exploring public sentiment on revenue allocation (e.g., healthcare or infrastructure).
4. Cultural Analysis: Studying on how casinos in other cities may affect cultural identity and apply lessons towards Toronto.

Works Cited

• “ City of Toronto. (n.d.). Casino survey results. Open Data Toronto. Retrieved November 9, 2024, from https://open.toronto.ca/dataset/casino-survey-results/ “

• “ City of Toronto. (n.d.). Toronto Casino Survey Feedback Form. Open Data Toronto. Retrieved November 9, 2024, from https://ckan0.cf.opendata.inter.prod-toronto.ca/dataset/427ca4cd168a-4a37-883d-4a574277caf5/resource/6af8b3c5-e758-4200-93f7- 29477161fbf8/download/toronto-casino-survey-feedback-form.pdf “

Appendix

R Code

```
#This is script for Final Project — Milestone 1
#Author: Yash S
#Created on: 2024-11-05
#Last Edited: 2024-12-12
#Class: ALY6010

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

# Load "pacman" to load/install packages easily
library(pacman)
p_load(tidyverse, readxl, caret)

# Loading the data set
casino <- read_excel("toronto-casino-survey-results.xlsx")

# Getting a sense of the data

str(casino) # Viewing the structure of the data
head(casino) # Viewing the first 6 rows of the data
colnames(casino) # Viewing the 94 Column names in the data set
colSums(is.na(casino)) # This counts the number of NA in each column
summary(casino) # Checking the data set

# Cleaning data for further analysis
```

```
clean_casino <- casino %>%
  select(
    Sentiment_Analysis = Q1_A,
    Image_Fit = Q2_A,
    Revenue_Influence = Q4_A,
    Preferred_Location = Q6,
    Considered_Location = ,
    Gender = Gender
  ) %>%
  filter(
    Gender %in% c("Male", "Female"),
    Preferred_Location %in% c("City of Toronto", "Adjacent Municipality")
    ) %>%
  drop_na()  # Drop rows with any NA values

# Converting categorical fields to factors for better analysis
clean_casino <- clean_casino %>%
  mutate(
    Sentiment_Analysis = as.factor(Sentiment_Analysis),
    Image_Fit = as.factor(Image_Fit),
    Revenue_Influence = as.factor(Revenue_Influence),
    Preferred_Location = as.factor(Preferred_Location),
    Gender = as.factor(Gender)
  )

str(clean_casino) # Viewing the structure of the cleaned data
head(clean_casino) # Viewing the first 6 rows of the cleaned data
colnames(clean_casino) # Viewing the column names in the cleaned data set
colSums(is.na(clean_casino)) # This counts the number of NA in each column
summary(clean_casino) # Checking the cleaned data set
```

```
# Descriptive Statistics
# Summary statistics for each field
summary_stats <- clean_casino %>%
  summarise(
    Sentiment_Count = n_distinct(Sentiment_Analysis),
    Image_Fit_Count = n_distinct(Image_Fit),
    Revenue_Influence_Count = n_distinct(Revenue_Influence),
    Preferred_Location_Count = n_distinct(Preferred_Location),
    Gender_Count = n_distinct(Gender)
  )
print(summary_stats)

# Frequency Tables
sentiment_table <- table(clean_casino$Sentiment_Analysis)
print("Frequency Table for Sentiment Analysis:")
print(sentiment_table)
# Image Fit
image_fit_table <- table(clean_casino$Image_Fit)
print("Frequency Table for Image Fit:")
print(image_fit_table)
# Revenue Influence
revenue_influence_table <- table(clean_casino$Revenue_Influence)
print("Frequency Table for Revenue Influence:")
print(revenue_influence_table)
# Preferred Location
preferred_location_table <- table(clean_casino$Preferred_Location)
print("Frequency Table for Preferred Location:")
print(preferred_location_table)
# Gender Distribution
gender_table <- table(clean_casino$Gender)
print("Frequency Table for Gender:")
```

```
print(gender_table)

# Data Visualization

# 1. Sentiment Analysis Distribution
ggplot(clean_casino, aes(x = Sentiment_Analysis)) +
  geom_bar(fill = "blue", colour = "black") +
  labs(title = "Sentiment Analysis Distribution", x = "Sentiment", y = "Count",) +
  theme_minimal() +
  coord_flip()

# 2. Gender Distribution
ggplot(clean_casino, aes(x = Gender)) +
  geom_bar(fill = "green", colour = "black") +
  labs(title = "Gender Distribution", x = "Gender", y = "Count") +
  theme_minimal()

# 3. Preferred Location Distribution
ggplot(clean_casino, aes(x = Preferred_Location)) +
  geom_bar(fill = "orange") +
  labs(title = "Preferred Location Distribution", x = "Location", y = "Count") +
  theme_minimal()

# Subset Analysis
# Analyze sentiments by gender
sentimentXgender <- clean_casino %>%
  group_by(Gender, Sentiment_Analysis) %>%
  summarise(Count = n(), .groups = "drop")

# Cross-tabulation: Gender vs Sentiment Analysis
gender_sentiment_crosstab <- table(clean_casino$Gender, clean_casino$Sentiment_Analysis)
```

```
print("Cross-tabulation of Gender and Sentiment Analysis:")
print(gender_sentiment_crosstab)

# Count: Gender and Sentiment Analysis
gender_sentiment_count <- clean_casino %>%
  count(Gender, Sentiment_Analysis)
print("Count of Gender and Sentiment Analysis:")
print(gender_sentiment_count)

# Visualization: Sentiments by Gender
ggplot(sentimentXgender, aes(x = Sentiment_Analysis, y = Count, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "Sentiment by Gender", x = "Sentiment", y = "Count") +
  coord_flip()

#------------------------------------------------------------------------------
# This is script for Final Project — Milestone 2
#------------------------------------------------------------------------------

# Question 1: Is the overall public sentiment towards establishing a new casino in Toronto neutral, or is it not neutral (positive/negative)?
# Map sentiment levels to numerical values
sentiment_mapping <- c(
  "Neutral or Mixed Feelings" = 0,
  "Somewhat in Favour" = 1,
  "Somewhat Opposed" =-1,
  "Strongly in Favour" = 2,
  "Strongly Opposed" =-2)
clean_casino$Sentiment_Numeric <- as.numeric(sapply(clean_casino$Sentiment_Analysis, function(x) sentiment_mapping[x]))
```

```
# Null Hypothesis (H0): The mean sentiment score is exactly neutral (0).
# Alternative Hypothesis (H1): The mean sentiment score is either positive(>0) or negative(<0).
# significance_level <- 0.05
# Compute test statistic
sentiment_t_test <- t.test(clean_casino$Sentiment_Numeric, mu = 0, alternative = "two.sided")
print(sentiment_t_test)
# p-value < 0.00000000000000022
# Final Conclusion:
# Since p-value is less then significance level(0.05), we reject the Null Hypothesis (H0)

# Question 2: Is the overall public sentiment towards establishing a new casino in Toronto positive or is it negative?
# Null Hypothesis (H0): The mean sentiment score is positive  (> 0).
# Alternative Hypothesis (H1): The mean sentiment score is negative (< 0).
# significance_level <- 0.05
# Compute test statistic
sentiment_negative_test <- t.test(clean_casino$Sentiment_Numeric, mu = 0, alternative = "less")
print(sentiment_negative_test)
# p-value < 0.00000000000000022
# Final Conclusion:
# Since p-value is less then significance level(0.05), we reject the Null Hypothesis (H0)

# Question 3: Is the casino more preferred in the City of Toronto than in adjacent municipalities or "neither"?
# Further cleaning the data by removing rows where Preferred_Location is "Neither"
location_filtered <- clean_casino %>%
  filter(Preferred_Location %in% c("Adjacent Municipality", "City of Toronto"))
# Frequency table for "City of Toronto" and "Adjacent Municipality"
location_preference_filtered_table <- table(location_filtered$Preferred_Location)
city_count <- location_preference_filtered_table["City of Toronto"]
adj_muni_count <- location_preference_filtered_table["Adjacent Municipality"]
```

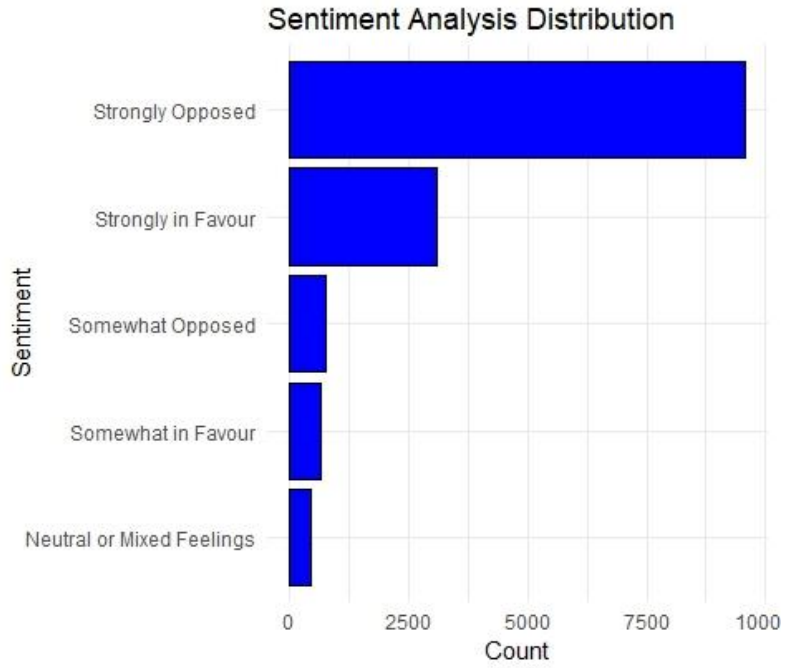
```
total_count <- sum(location_preference_filtered_table)

# Null Hypothesis (H0): Adjacent Municipality is significantly more preferred than City of Toronto.

# Alternative Hypothesis (H1): City of Toronto is significantly more preferred than Adjacent Municipality.

# significance_level <- 0.05

# Compute test statistic

city_vs_adj_muni_test <- prop.test(c(city_count, adj_muni_count), n = c(total_count, total_count), p = c(0.5, 0.5))

print(city_vs_adj_muni_test)

# p-value < 0.00000000000000022

# Final Conclusion:

# Since p-value is less then significance level(0.05), we reject the Null Hypothesis (H0)


#------------------------------------------------------------------------------

# This is script for Final Project

#------------------------------------------------------------------------------


# Dummy Encoding

clean_casino <- clean_casino %>%

  mutate(

    Revenue_Influence_Numeric = ifelse(Revenue_Influence == "Yes", 1, 0),

    Fits_Perfectly = ifelse(Image_Fit == "Fits Image Perfectly", 1, 0),

    Fits_Somewhat = ifelse(Image_Fit == "Fits Image Somewhat", 1, 0),

    Neutral_Not_Sure = ifelse(Image_Fit == "Neutral / I am Not Sure", 1, 0),

    City_of_Toronto = ifelse(Preferred_Location == "City of Toronto", 1, 0),

    Adjacent_Municipality = ifelse(Preferred_Location == "Adjacent Municipality", 1, 0)

    ) %>%

  select(-c(Revenue_Influence, Image_Fit, Preferred_Location)) # Dropping Columns to avoid multi collinearity


# Question 1: Does the revenue influence predict public sentiment towards the casino?
```

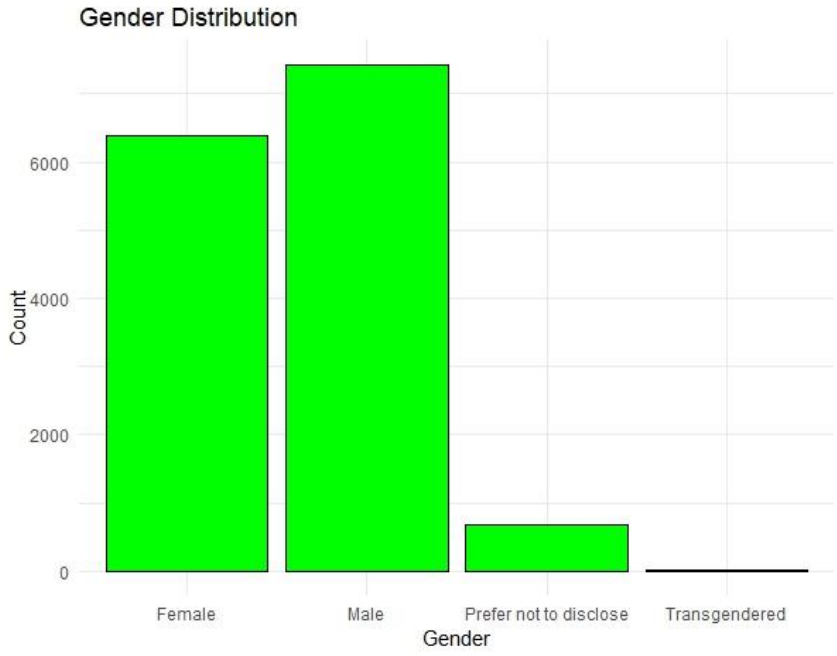
```
# Null Hypothesis (H0): Revenue influence does not predict public sentiment (no significant relationship).

# Alternative Hypothesis (H1): Revenue influence predicts public sentiment (significant relationship).

# significance_level <- 0.05

# Linear Regression

model_location_sentiment <- lm(

  formula = City_of_Toronto ~ Sentiment_Numeric,

  data = clean_casino

  )

summary(model_location_sentiment)

# p-value < 0.00000000000000022

# Final Conclusion:

# Since p-value is less then significance level(0.05), we reject the Null Hypothesis (H0)

# Scatter plot with regression line

ggplot(clean_casino, aes(x = Revenue_Influence_Numeric, y = Sentiment_Numeric)) +

  geom_point() +

  geom_smooth(method = "lm", se = FALSE, color = "blue") +

  labs(title = "Public Sentiment vs. Revenue Influence",

       x = "Revenue Influence (Dummy: Yes = 1, No = 0)",

       y = "Public Sentiment") +

  theme_minimal()


# Question 2: Does the perception of image fit predict public sentiment toward the casino?

# Null Hypothesis (H0):  Perception of image fit does not predict public sentiment (no significant relationship).

# Alternative Hypothesis (H1): Perception of image fit predicts public sentiment (significant relationship).

# significance_level <- 0.05

# Linear Regression

model_image_sentiment <- lm(

  formula = Sentiment_Numeric ~ Fits_Perfectly + Fits_Somewhat + Neutral_Not_Sure,

  data = clean_casino
```

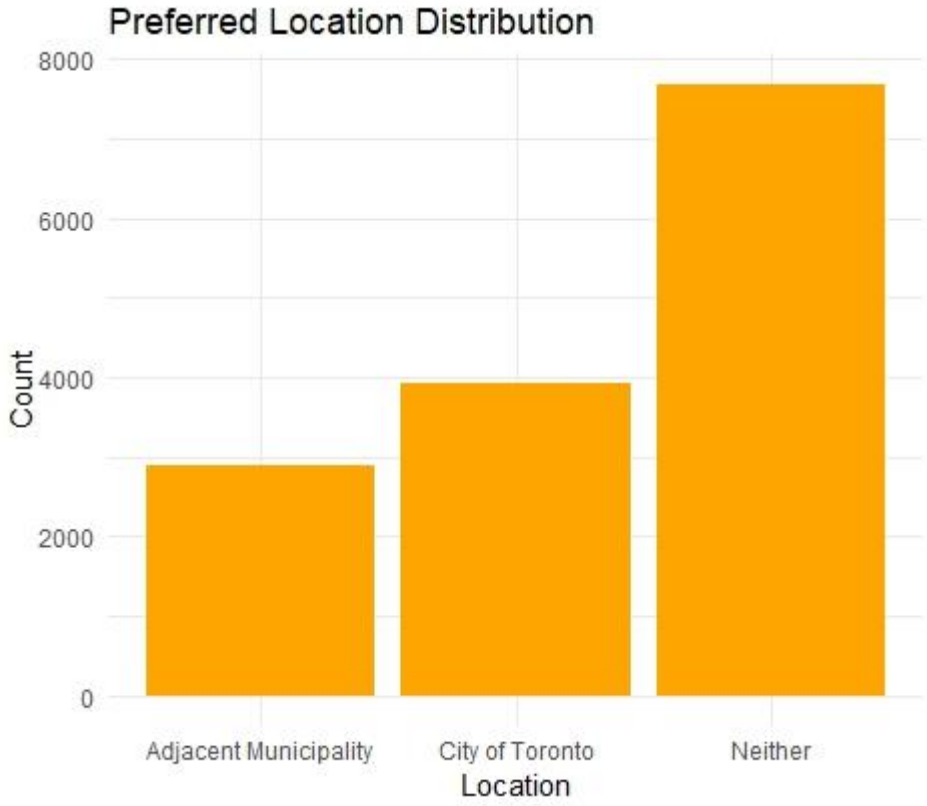
```
 )
summary(model_image_sentiment)
# p-value < 0.00000000000000022
# Final Conclusion:
# Since p-value is less then significance level(0.05), we reject the Null Hypothesis (H0)
# Scatter plot with regression line for one level ("Fits Image Perfectly")
ggplot(clean_casino, aes(x = Fits_Perfectly, y = Sentiment_Numeric)) +
 geom_point() +
 geom_smooth(method = "lm", se = FALSE, color = "red") +
 labs(title = "Public Sentiment vs. Image Fit (Fits Perfectly)",
      x = "Fits Image Perfectly (Dummy: Yes = 1, No = 0)",
      y = "Public Sentiment") +
 theme_minimal()

# Question 3: Does Sentiment Predict Preference for the City of Toronto?
# Null Hypothesis (H0):  Sentiment does not predict preference for the City of Toronto.
# Alternative Hypothesis (H1): Sentiment predicts preference for the City of Toronto.
# significance_level <- 0.05
# Linear Regression
model_toronto_sentiment <- lm(
 formula = City_of_Toronto ~ Sentiment_Numeric,
 data = clean_casino
)
summary(model_toronto_sentiment)
# p-value < 0.00000000000000022
# Final Conclusion:
# Since p-value is less then significance level(0.05), we reject the Null Hypothesis (H0)
# Scatter plot with regression line for one level ("Fits Image Perfectly")
ggplot(clean_casino, aes(x = City_of_Toronto, y = Sentiment_Numeric)) +
 geom_point() +
 geom_smooth(method = "lm", se = FALSE, color = "green") +
```

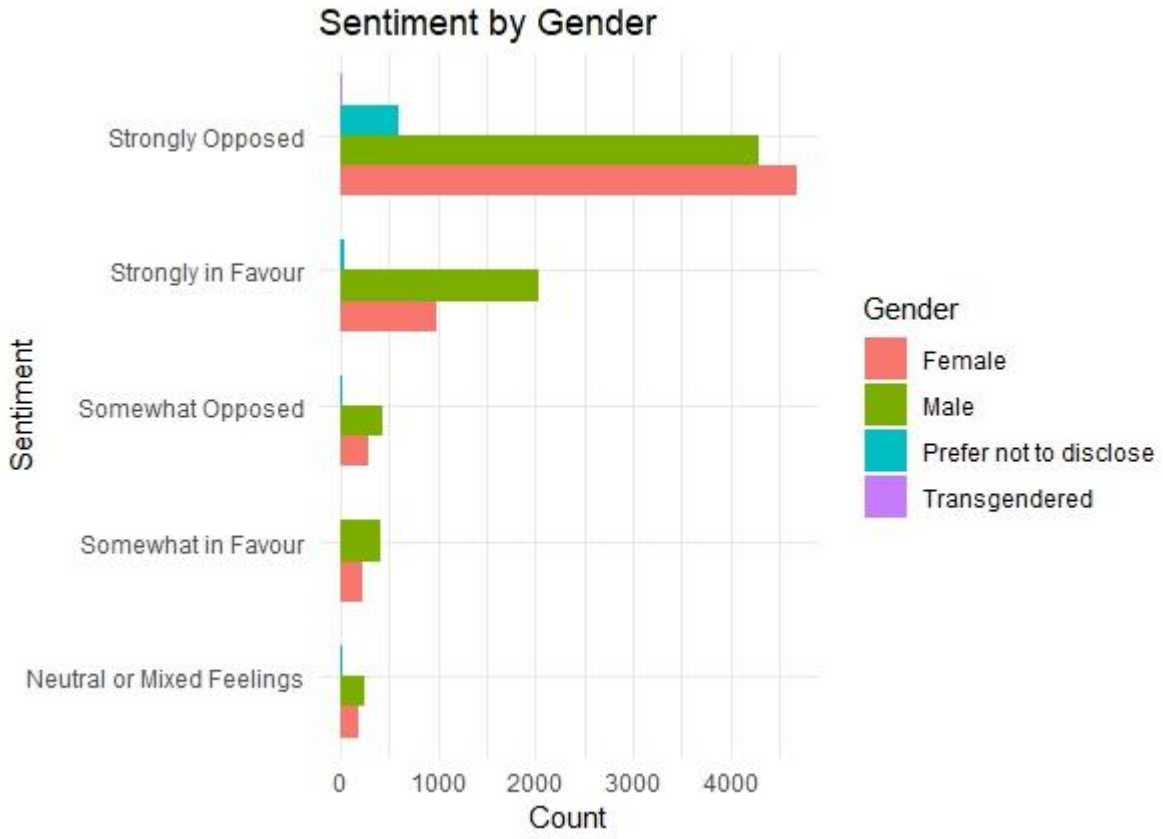
```
 labs(title = "City of Toronto Preference vs. Public Sentiment",
      x = "City of Toronto Preference (Dummy: Yes = 1, No = 0)",
      y = "Public Sentiment"
 ) +
 theme_minimal()
```

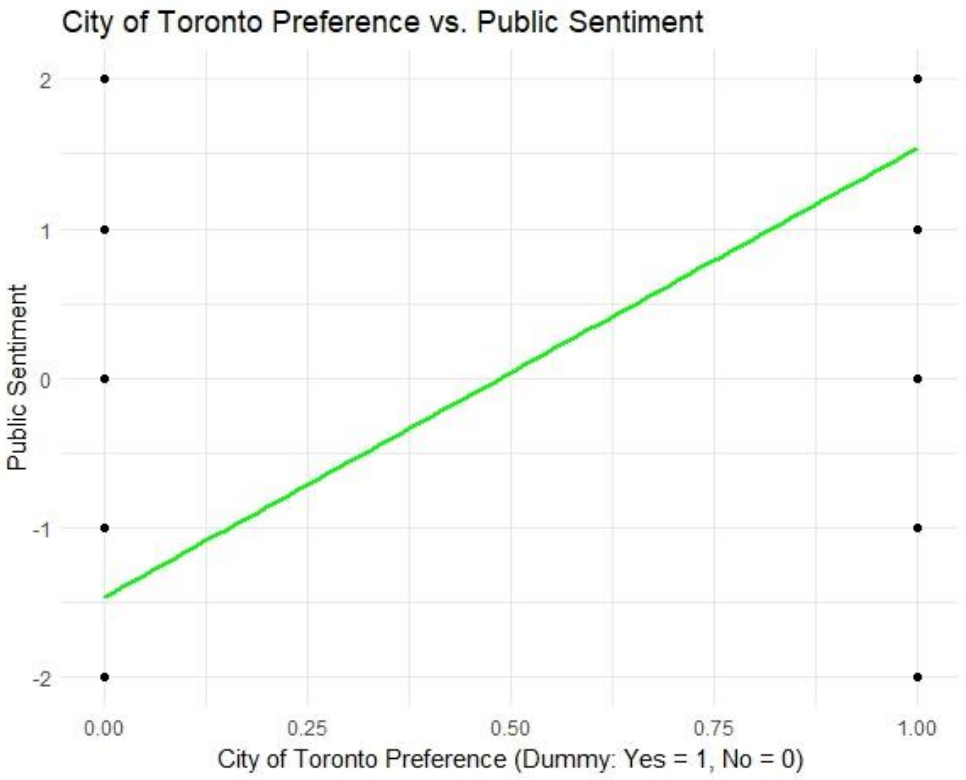
Visualizations

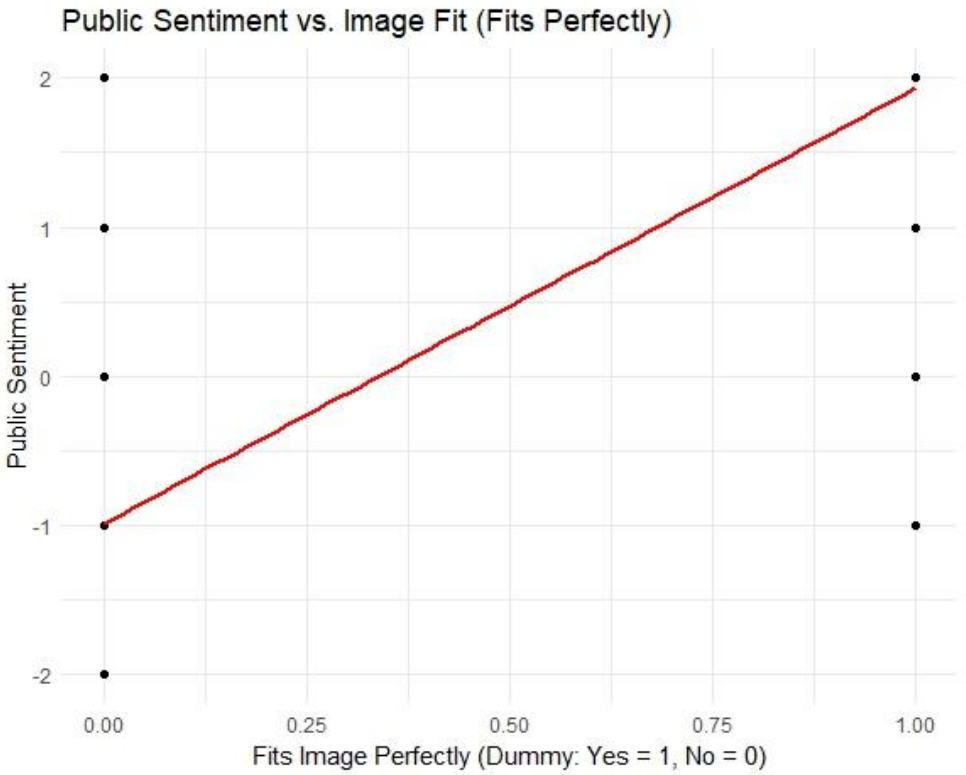


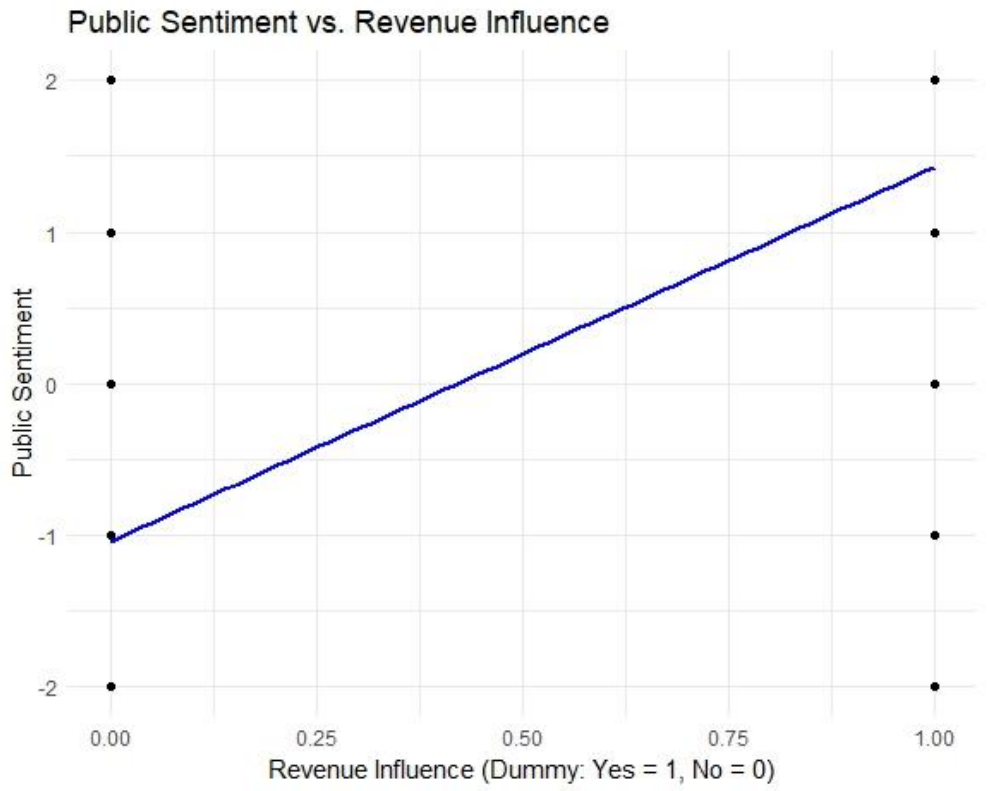



Preferred Location Distribution
8000
6000
4000
2000
0
Count
Adjacent Municipality
City of Toronto
Neither
Location


Sentiment by Gender
Strongly Opposed
Strongly in Favour
Somewhat Opposed
Somewhat in Favour
Neutral or Mixed Feelings
Sentiment
0
1000
2000
3000
4000
Count
Gender
Female
Male
Prefer not to disclose
Transgendered


City of Toronto Preference vs. Public Sentiment
Public Sentiment
2
1
0
-1
-2
0.00
0.25
0.50
0.75
1.00
City of Toronto Preference (Dummy: Yes = 1, No = 0)


Public Sentiment vs. Image Fit (Fits Perfectly)
Public Sentiment
2
1
0
-1
-2
0.00
0.25
0.50
0.75
1.00
Fits Image Perfectly (Dummy: Yes = 1, No = 0)



Descriptive Analysis:

1. Summary of the Cleaned Data:

```
> summary(clean_casino) # Checking the cleaned data set
                Sentiment_Analysis                          Image_Fit      Revenue_Influence
 Neutral or Mixed Feelings: 436     Does Not Fit My Image At All:10109   Don't know: 873
 Somewhat in Favour       : 656     Fits Image Perfectly         : 2949   No         :9722
 Somewhat Opposed         : 743     Fits Image Somewhat          :  821   Yes        :3907
 Strongly in Favour       :3078     Neutral / I am Not Sure      :  623
 Strongly Opposed         :9589
            Preferred_Location                    Gender
 Adjacent Municipality:2897     Female                :6378
 City of Toronto      :3916     Male                  :7419
 Neither              :7689     Prefer not to disclose: 678
                                Transgendered         :  27
```

2. Summary Statistics for each field:

```
> # Summary statistics for each field
> summary_stats <- clean_casino %>%
+   summarise(
+     Sentiment_Count = n_distinct .... [TRUNCATED]

> print(summary_stats)
# A tibble: 1 × 5
  Sentiment_Count Image_Fit_Count Revenue_Influence_Count Preferred_Location_Count Gender_Count
            <int>           <int>                   <int>                    <int>        <int>
1               5               4                       3                        3            4
```

3. Frequency Tables for each field:

a) Sentiment Analysis

```
[1] "Frequency Table for Sentiment Analysis:"

> print(sentiment_table)

Neutral or Mixed Feelings        Somewhat in Favour         Somewhat Opposed       Strongly in Favour
                      436                       656                       743                      3078
        Strongly Opposed
                     9589
```

b) Image Fit

```
> print(image_fit_table)

Does Not Fit My Image At All         Fits Image Perfectly          Fits Image Somewhat      Neutral / I am Not Sure
                       10109                         2949                          821                          623
```

c) Revenue Influence

```
> print(revenue_influence_table)

Don't know         No        Yes
       873       9722       3907
```

d) Preferred Locations

```
> print(preferred_location_table)

Adjacent Municipality       City of Toronto               Neither
                 2897                  3916                  7689
```

e) Genders

```
> print(gender_table)

                Female                   Male Prefer not to disclose          Transgendered
                  6378                   7419                    678                     27
```

4. Subset:

```
> print("Cross-tabulation of Gender and Sentiment Analysis:")
[1] "Cross-tabulation of Gender and Sentiment Analysis:"

> print(gender_sentiment_crosstab)

                       Neutral or Mixed Feelings Somewhat in Favour Somewhat Opposed Strongly in Favour
  Female                                     182                232              288                995
  Male                                       237                416              428               2037
  Prefer not to disclose                      15                  8               26                 38
  Transgendered                                2                  0                1                  8

                       Strongly Opposed
  Female                           4681
  Male                             4301
  Prefer not to disclose            591
  Transgendered                      16
> # Count: Gender and Sentiment Analysis
> gender_sentiment_count <- clean_casino %>%
+   count(Gender, Sentiment_Analysis)

> print("Count of Gender and Sentiment Analysis:")
[1] "Count of Gender and Sentiment Analysis:"

> print(gender_sentiment_count)
# A tibble: 19 × 3
   Gender                 Sentiment_Analysis            n
   <fct>                  <fct>                     <int>
 1 Female                 Neutral or Mixed Feelings   182
 2 Female                 Somewhat in Favour          232
 3 Female                 Somewhat Opposed            288
 4 Female                 Strongly in Favour          995
 5 Female                 Strongly Opposed           4681
 6 Male                   Neutral or Mixed Feelings   237
 7 Male                   Somewhat in Favour          416
 8 Male                   Somewhat Opposed            428
 9 Male                   Strongly in Favour         2037
10 Male                   Strongly Opposed           4301
11 Prefer not to disclose Neutral or Mixed Feelings    15
12 Prefer not to disclose Somewhat in Favour            8
13 Prefer not to disclose Somewhat Opposed             26
14 Prefer not to disclose Strongly in Favour           38
15 Prefer not to disclose Strongly Opposed            591
16 Transgendered          Neutral or Mixed Feelings     2
17 Transgendered          Somewhat Opposed              1
18 Transgendered          Strongly in Favour            8
19 Transgendered          Strongly Opposed             16
```