STA304 Final Project*

Wei Zhang

27 April 2022

Abstract

This report presents a deep analysis of people's attitudes towards establishing the casinos in Toronto in 2012. The result indicates that the relative factor, like revenue contribution, sex, age groups, public safety, employment rate and the judgement regarding whether casinos fit the image of Toronto, pose the effects to the attitudes. The analysis is made up with the logistic regression model operated by the statistical programming language R. The result yielded by the model can contribute to understanding of the issues about establishing the casinos, enabling the government or other authorities to predict the attitudes towards casinos for a new city in the future.

1 Key words

Attitudes towards casinos, Survey, Logistic regression model, EDA

2 Introduction

Since Yangzhou, which is a city in China, burst out the COVID-19 cases due to the underground casinos in the end of 2021. The Chinese government started to prohibit the casinos officially with different varieties of serious actions. In fact, establishing the casinos is becoming more and more popular all over the world, while compared with Chinese attitudes towards the casinos. Thomas (2003) argued that dating back to the 1980s, casino gambling was only legal in Nevada and Atlantic City. But currently, nearly 30 states have legalized casino gambling. One of the key aspects, which makes it such welcome, is that it can greatly contribute to the economic growth, for example, employment rate and tax revenue to the local government.

In general, interests that are attracted, whether Toronto can establish the casinos or not. As a matter of fact, City Council's Executive Committee required the City Manager to seek for the public inputs regarding the establishment of a casino in Toronto. They strived to make it for the tremendous economic benefits. This raises the questions that what kind of factor may affect the attitudes towards the establishment of casinos in Toronto and are the influence of these factors positive or negative. To answer the questions, we would utilize the R programming language to analyse it. Then, logistic regression model will be conducted to explore the justifications for the research questions.

The result reveals that the ultimate people's attitudes towards establishing the casinos in Toronto is associated with the revenue contribution, public safety, employment rate and the ideas about whether casinos fit the images of Toronto or not. The result turns out that public safety posed the negative influence, while others, which are significant, are all positive to people's attitudes. We believe studies on gambling effects could assist researchers and policymakers to make the comparison of the health and social costs and benefits

---

*Code and data are available at: https://github.com/WeiZhang1998/Attitude-towards-establishing-the-casinos-in-Toronto-in-2012

of different gambling policies, In addition, it can be used when thinking of what kind of gambling policies will reduce or benefits the most (Tiina, 2019).

The paper starts with an overview of the original dataset and the relevant variables. To help us further understanding the logistic regression model afterwards, it presents the data collection process and summary of important variables. Besides, in the Model section, apart from introducing the model itself, it reveals the advantages and disadvantages of the model application and also includes the data cleaning to ensure data appropriate for the model. Next, result shows the answer to the research question and interpretation for the model we fixed. Eventually, Discussion part consist of findings, ethics, limitation and future work within a comprehensive context.

2.1 Research questions & Hypothesis

- Which factor would influence people's attitudes towards the establishment of casinos?

- Regarding the significant variables, which factor would pose the positive and which factor would pose the negative?

- Assume that all selected variables are independent of each other

- Assume that the aspects of revenue, whether casinos fit the image of Toronto and employment rate pose a positive effects, while concerning for public safety will lead to the negative effect.

3 Data

This report would be produced with the R statistical programming language in the Rmarkdown Format, including all the explantory data analysis, model fitting and visualization. (R core Team 2020). The dataset is accessed from OpenToronto package. Tidyverse package is utilized for the data cleaning and plot the variables. Tables are made with the knitr package and readxl is used for read and import the files.

3.1 Data Collection

The dataset is available in the Portal of Open Data Toronto governed by the Statistics of the National Bureau of Statistics. The topic is City government Culture and tourism. When dating back to November of 2012, City Council's Executive Committee requested the City Manager to strive to look for public's inputs on the establishment of a casino in Toronto. The City Manager engaged the firm DPRA to conduct the consultation with the public and stakeholders by the end of January 2013. Notify that Public consultation is an important part of Council's decision making process. The consultation is responsible for gathering inputs, including the public's views and opinions on a casino generally, on possible locations for a casino, and what the public would like Council to consider when making a decision on this matter. Data collected is based on the Toronto Casino Feedback Form via the telephone and questionnaires to fill out the survey. Awards are provided for all the participant for guaranteeing the quality of forms filled.

As referred to the sources, the survey was designed to capture the necessary information about establishing the casinos in Toronto. It consists of the form of multiple choices, short answer and check boxes questions to collect people's opinions. Two parts are major, one is directly regarding people's opinions towards establishing the casinos in Toronto, another is to collect some ideas and reaction about if casinos are established, such as location preference and layout bias and expectations.

3.2 Data cleaning

To contribute to further explantory data analysis and model fixed, we are supposed to make the data cleaning process to sustain the understanding and operation. We utilize the `Tidyverse` package to change the name of variables in the dataset. Besides, we transform the answer to the number one, two, three and four based on the extent of attitudes. For example, Not Important At All as 1, Unsure as 2, Somewhat Important as 3, Very Important as 4. Meanwhile, as we all know that the response variable of logistic regression model is binary, we have to transform the attitudes to 0 or 1. Notify that 0 means oppose to establishing the casinos, while 1 means in favour of casinos. The last but not least, the type of data is numeric, which would be convenient to write down the actual model.

3.3 Description

The survey managed to collect different varieties of data, including specific and general information and short answer questions, such as sex, age, income, Revenue, employment rate, Attitudes towards establishing the casinos, education level. In addition, it consults the ideas about the casinos via open question forms without the actual right answers. Embodying the shallow level and deep level, the basic information like whether you support establishing or not and the further question like what the location preference is if casinos are built. It possesses in total 17766 observations and 94 questions with respect to numerous aspects. All possible and filled out answers are collected, since it is difficult to justify whether it is precious or not providing not reading the questionnaires.
Below would be all selected questions and all possible answers.

| Attitude	Fit	Employment	Public_safety	Revenue	Sex
Strongly Opposed	Does Not Fit At All	Not Important At All	Not Important At All	No	Male
Somewhat Opposed	Neutral	Unsure	Unsure	Don't know	Female
Somewhat in Favour	Fits Image Somewhat	Somewhat Important	Somewhat Important	Yes	
Neutral or Mixed Feelings	Fits Image Perfectly	Very Important	Very Important		
Strongly in Favour					

3.4 Explantory variable analysis

- **Attitude**: The variable is collected by checkboxes questions in the survey. It has five options, which has been revealed above. It shows people's attitudes towards establishing the casinos in Toronto. We process the variable according to the scales from 1 to 5, where 1 represents Strongly Opposed and 5 represents Strongly in Favor.

- **Fit**: The variable is collected by checkboxes questions in the survey. It has four options which has been revealed above. It shows how you feel about establishing casino in Toronto. We process the variable according to the scales from 1 to 4, where 1 represents the Does Not Fit My Image at All and 4 represents Fits Image Perfectly.

- **Employment**: The variable is collected by checkboxes questions in the survey. It has four options which has been revealed above. It shows people's ideas about the influence of employment rate if casinos are established. We process the variable according to the scales from 1 to 4, where 1 represents Not Important At All and 4 represents Very Important.

- **Public_safety**: The variable is collected by checkboxes questions in the survey. It has four options which has been revealed above. It shows people's ideas about the influence of public safety if casinos are established. We process the variable according to the scales from 1 to 4, where 1 represents Not Important At All and 4 represents Very Important.

- **Revenue**: The variable is collected by checkboxes questions in the survey. It has three options which has been revealed above. It shows people's ideas about whether the economic revenue would be promoted if casinos are established. We process the variable according to the scales with 0 and 1, where 0 represents No and 1 represents Yes.

- **Sex**: The variable is collected by checkboxes questions in the survey. It has two options which has been revealed above. It shows the gender of participants. We process the variable with 0 and 1, where 1 represents Male and represents Yes.

- **Age**: The variable is collected by checkboxes questions in the survey. It has six options divided by different age groups. It show the age of participants. We process the variable from 1 to 6, where 1 represents before 15-year-old and 6 represents from 55-64 yeas old.

Variable	Type	Binary or not
Attitude	Numeric	Binary
Fit	Numeric	Not binary
Employment	Numeric	Not binary
Public_safety	Numeric	Not binary
Revenue	Numeric	Binary

Variable	Type	Binary or not
Sex	Numeric	Not binary
Age	Numeric	Not binary

As shown in the table, we find that all the selected variables are type of numeric after the cleaning process. In addition, attitudes as the response variable is binary. These characteristics endows us to make the logistic regression model to sustain the further analysis.

Figure 1 – Compared to two attitudes
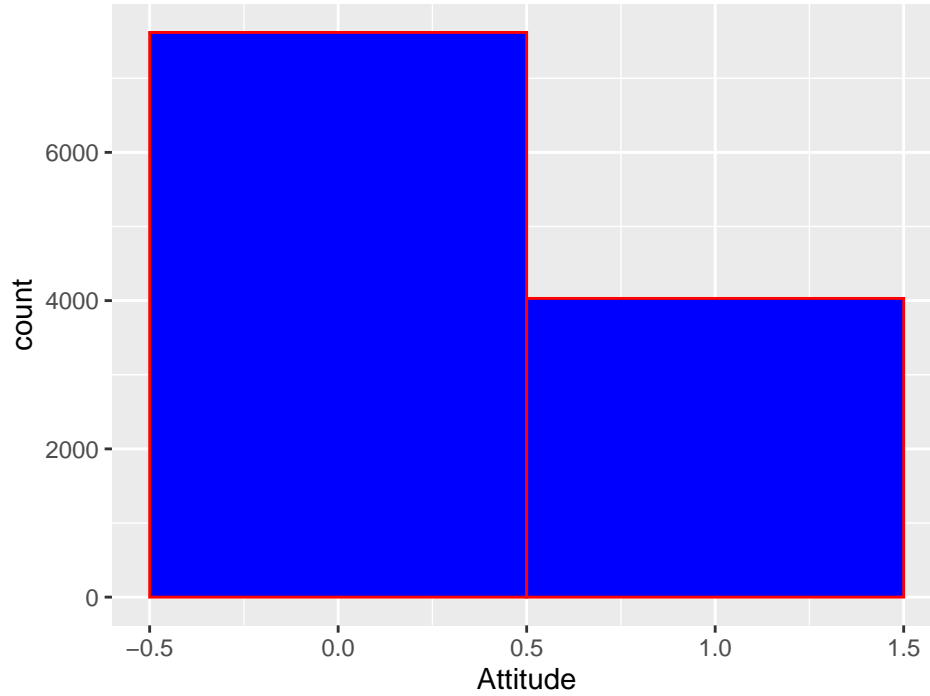
In the Figure 1, we compared two attitudes towards establishing the casinos in Toronto via the histogram plot. As indicated, the number of people who opposed to casinos is more than the number of people who support the casinos in Toronto. Binary response variables can be conducted by the logistic regression analysis in the next stage.
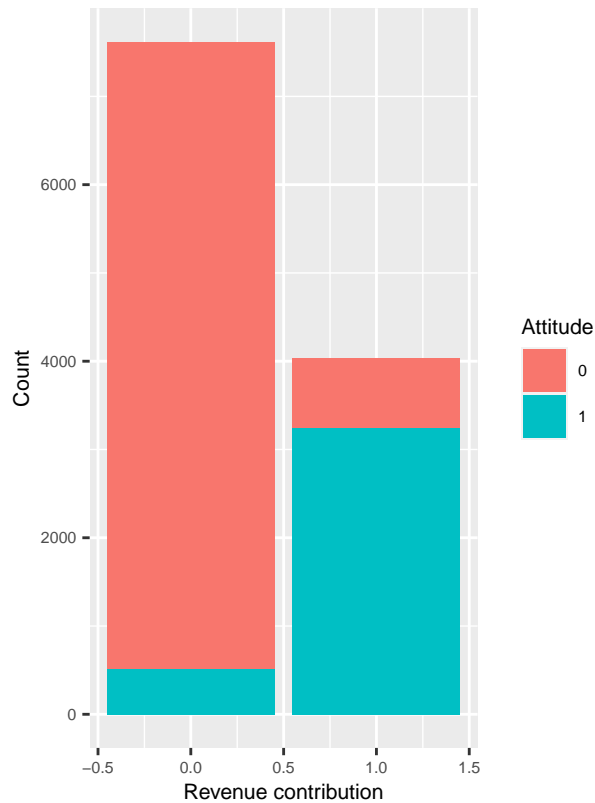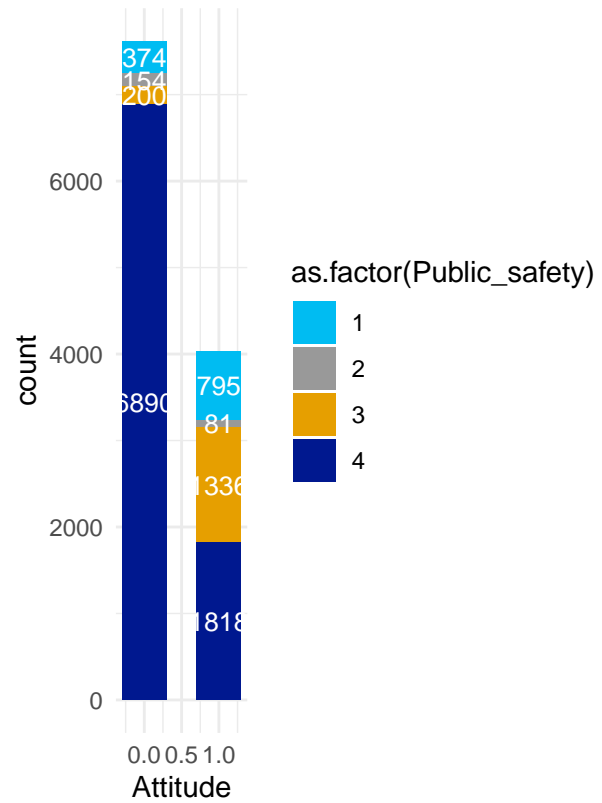
Figure 2 – Attitude vs Revenue contribution

Figure 3 – Attitude vs Public safety

In the Figure 2, the bar plot presents attitudes verse revenue contribution. As indicated, people, who reject establishing the casinos, deems that casinos will not support the revenue contribution and economic growth, while the majority of people, who support establishing the casinos, have faith that the occurrence of casinos would contribute to the local economic promotion. In addition, we can derive that there are more people rejecting the establishment of casinos in general.

In the Figure 3, the bar plot with mapping method indicates the relationship between attitudes and public safety. As shown, the majority of people deem that public safety is very important. In other words, people concern the public safety deeply providing that the casinos are built. Although there does not exist the relationship with a glimpse, sophisticated research would be conducted further.

4 Model

Logistic regression model is one of the key statistic analysis methods and apply the logit function to model the response variable. It aims to explore the relationships between the binary response variable and possible explantory variables. It is a simple and efficient method for binary and linear classification problems. Even in the area of Computer science, it is regarded as the classification model rather than regression model (Subasi, 2020).

The general form of logistic regression model would be:

$$\log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$$

With respect to our study, we aim to explore what kind of factors are related to the binary outcome of `Attitude`. There are 5 predictor variables and 1 response variables within our selected varables. The model relying opon is logistic regression model. Hence, the model assumes that

$$\log(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

Where the logit function is $\log(\frac{p}{1-p})$

- P is the probability of the outcome variable Y.

- β_0 means the log-odds

- β_i means the expected change in log odds of increasing the relative units by 1 unit, where $i \in [1,5]$ & $i \in N$.

- All the selected variables in the model should not exist the relationship between each other. Specifically, multicolinearity problems should be avoided.

4.1 Strengths & Weakness

Logistic regression could be utilized for describing data and to explaining the relationship between one dependent binary variable and one or more predictor independent variables. In addition, Logistic regression is convenient to operate, interpret and highly efficient to understand. The model apparently exclaims the importance of each explantory variable when comparing to others.

The disadvantages are also clear. firstly, the logistic regreesion model is pretty sensitive to abnormal cases like outliers in data. Secondly, before the model is applied to analyse the data, the predictor variables should guarantee that there does not exist the multicolinearity problems. Otherwise, the model we fixed may exist the overfitting flaws. Further step are requested to eliminate this effects. We would utilize the VIF function to ensure this problem disappearing. Besides, the number of observation should exceed the number of predictor variables, otherwise, it could result to overfitting. In fact, we have altogether 11648 observations in the dataset and 6 indicator variables. Hence, we have no need to take care of this problem.

4.2 VIF

The full name of VIF is variance inflation factor function. It is significantly statistics tools, which can be used for measuring the variance of coefficient with respect to the multicolinearity problems. In other words, it is capable of checking whether all selected variables in the model are independent of each other. It is encapsulated in the *car* package. After applying the VIF method, the criterion is to make a comparison of value of 5. Notify that greater than 5 means there exists the multicolinearity problem, while smaller than 5 represents the model is apprppriate. (Shivon, 2020 fall). Before we conduct the further analysis when we fitting the regression model, it is essential to apply the VIF function to check the existence of multicolinearity problem, other the result yielded may be insignificant and inaccurate.

5 Result

In this section, we would utilize the model mentioned above to answer all the research questions and hypothesis. To begin with, we fit the logistic regression model based on the data after the cleaning process. Then, we apply the VIF function to check whether there existing the multicolinearity problems in the fitted model. Next, with the p-value aspects, we choose the significant variables and refit the model. Finally, we would search for some relevant information to justify the variable rationality.

First, we fix the logistic regression model, where `Attitudes` as the response variable and the rest of selected variables are predictor variables.
Below is the model coefficients:

Table 3: coefficients of logistic model

	x
(Intercept)	-6.9523580
Fit	3.0516110
Employment	0.5932729
Public_safety	-0.7030864
Revenue	2.1651410
Age	0.0434174
Sex	0.0971632

According to the Table 3, the model can be written as:

$$Attit\hat{u}des = -6.952 + 3.052*Fit + 0.593*Employment - 0.703*Public_safety + 2.165*Revenue + 0.043*Age + 0.097*Sex$$

Before the further analysis, we attempt the VIF function from the package of *car* to check whether there exist the multicolinearity problems.

Table 4: VIF value of all selected variable

	x
Fit	1.007399
Employment	1.247362
Public_safety	1.226249
Revenue	1.046024
Age	1.035219
Sex	1.039955

As referred to the Table 4, we can find that all the results yielded by the VIF function are all smaller than 5, which demonstrates that all the selected variables are not correlated with each other, which corresponds to hypothesis and prove the hypothesis Legitimacy. Therefore, the outcome of this step can sustain our further exploration.

Next, we come to discussion the best model and eliminate the insignificant variables.

Table 5: P-value & Relative symbol

Variables	P.value	Symbol
Fit	< 2e-16	***
Employment	1.13e-14	***

Variables	P.value	Symbol
Public_safety	4.01e-16	***
Revenue	< 2e-16	***
Age	0.448	
Sex	0.486	

Clearly, With respect to the Table 5, the p-values of `Fit`, `Employment`, `Public_safety` and `Revenue` are all smaller than 0.05, while the p-values of variables `Age` and `Sex` greater than 0.05. According to the p-values criterion,

Table 6: P-value criterion

Symbol	P.value	Meaning
	$p > 0.05$	insignificance
*	$p < 0.05$	significance
**	$p < 0.01$	better significance
***	$p < 0.001$	powerful significance

As for the Table 6, We can judge that `Fit`, `Employment`, `Public_safety` and `Revenue` are statistically significant, but `Age` and `Sex` are not. After we eliminate the insignificant variables, the final model would be

$$\hat{Attitudes} = -6.505 + 3.036 * Fit + 0.511 * Employment - 0.676 * Public_safety + 2.128 * Revenuex$$

Notify that all the explatory variables are significant in the new model (summary table in the appendix). This directly answer the research questions that people's attitudes towards establishing the casinos are closely related to the aspects of employment rate, concerning public safety, revenue contribution and judgement about whether casinos fit the image of Toronto. However, the ultimate attitudes are not associated with the Age groups and the gender. In addition, considering the coefficients, among all the significant variables, it could be concluded that employment rate, revenue contribution and judgement about whether casinos fit the image of Toronto play a positive role for the final attitudes, while concerning public safety poses a negative effect to the final attitudes.

6 Discussion

6.1 Findings

We manage to fit the logistic regression model to analyse the ultimate Attitudes regarding casinos for people in Toronto. In general, ultimate Attitudes are determined by factors, which is employment rate, concerning public safety, revenue contribution and judgement about whether casinos fit the image of Toronto. Among these, employment rate, revenue contribution and judgement about whether casinos fit the image of Toronto pose the positive effect to the final decision, while public safety exerts the negative influence to the ultimate attitudes. Besides, based on the model crierion, age groups and sex are not correlated with the attitudes due to the p-values dissatisfaction.

In fact, the result yielded completely fulfilled the common sense with respect to the casinos. Thomas (2003) argues that a lot of countries have made acquiescence and even directly agreed commercial casino gambling mainly due to a large amount of economic benefits and motive for economic growth. In addition, economy itself is associated with the employment rate and revenue contribution. In terms of public safety section, although a number researches has been conducted over the last thirty years, Albanese (2019) still made the systematic study regarding the connection between the casinos and public safety, he argued that the casinos should be responsible for the increasing crime rate, some citizens and public officials have published the statements regarding the crimes link for the casino jurisdictions. Hence, the negative influence are expected by concerning public safety.

6.2 Ethics & Implications

Considering research often possess a large amount of cooperation among various groups of people with different background and disciplines. Ethics is extremely vital to be paid attention to. Within the process of collecting the data in the survey, the question regarding gender is male, female and transgender. Apparently, it take good cares of the transgender people regarding ethics issues.

However, although Open Data Portal possesses the pretty rigorous audit procedure, since any dataset in Open Data Toronto is publicly accessible. Still the survey requires the information of quasi-identifiers like birth year, educational level, income, revenue contribution, postal codes, number of families and so on. These sorts of confidential information are not taken by some constructive actions to protect the privacy. Ethic issues is exceedingly important to any researches or studies only if human participants involved. As a matter of fact, appropriate disclosures are requested and further optimization is still necessarily granted.

6.3 Limations & Future work

6.3.1 Limit

Dating back to the process of research, the original study possess some operation flaws that directly influence the data, it has three main limitations. First of all, we cannot handle the type of short answer questions, which makes our data losing a large amount valuable comments and ideas. Since in the original dataset, the majority part of answer consist of long answers in the form of words or paragraphs. Only a few question are checkboxes and multiple choices. Without the abundant knowledge directly affect the accuracy of the model and outcomes. Secondly, with respect to collecting survey, sampling methods possess great space to optimize like systematic sampling, rather than always simply distribute leaflets or via telephone methods. The last but not least, the questions in the survey mainly about the ideas regarding approving of establishing the casinos, but for the group of people, who rejected the establishment, most of questions in the survey is nonsense and chances are that producing extra useless information and complaints. For example, one of the question listed a few locations providing that people who support the occurrence of casinos.

6.3.2 Future work

The most important and significant further steps is to apply some deep learning model and skills to the process of handling this survey, so that the missing precious data and information can be gathered and made use of. Otherwise, the model fitted is only originated from the restrained data information. More direct and meaning ideas deserve to be considered and analysed. In addition, the model presents that the gender is not significant to determine the attitudes, but considering the common sense, women always does not have an appetite and interest for the gambling, while men more easily find the enjoyable experience when taking the risks and possess impulsive coping than women with lower levels (Gambl Stud, 2016). Hence, the disappearing of sex factor in the model is necessarily to be considered and analysed. Hopefully, the model fixed can be utilized by the government or other authorties to predict the attitudes towards casinos in the future.

7 References

Grolemund, G. (2014, July 16) *Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/ articles_intro.html. (Last Accessed: January 15, 2021)

Thomas A Garrett. (2003, January 1). *Casinos and Economic Development: A Look at the Issues.* [https://www.stlouisfed.org/publications/bridges/winter-20022003/casinos-and-economic-development-a-look-at-the-issues]

R statistics. (2020, August 3). *Check for multicollinearity with the car package in R.* [https: //rforpoliticalscience.com/2020/08/03/check-for-multicollinearity-with-the-car-package-in-r/]

Jay Albanese. (2020, August 23). *Casino Gambling Impacts on Crime and Public Safety: A Review of 30 Years of Research.* [file:///C:/Users/zfrjk/Downloads/Casino_Gambling_Impacts_on_Crime_Review_ IJCJS_2019_19p.pdf]

J Gambl Stud. (2016, February 2). *Examining Gender Differences for Gambling Engagement and Gambling Problems Among Emerging Adults.* [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4736715/]

8 Appendeix

```
##      Attitude           Fit          Employment    Public_safety
##  Min.   :0.000   Min.    :1.000   Min.    :1.00   Min.    :1.000
##  1st Qu.:0.000   1st Qu.:1.000    1st Qu.:2.00    1st Qu.:3.000
##  Median :0.000   Median :1.000    Median :4.00    Median :4.000
##  Mean   :0.346   Mean    :1.921   Mean    :3.06   Mean    :3.527
##  3rd Qu.:1.000   3rd Qu.:3.000    3rd Qu.:4.00    3rd Qu.:4.000
##  Max.   :1.000   Max.    :4.000   Max.    :4.00   Max.    :4.000
##
##      Revenue            Age             Sex
##  Min.   :0.0000   Min.    :1.000   Min.    :0.0000
##  1st Qu.:0.0000   1st Qu.:3.000    1st Qu.:0.0000
##  Median :0.0000   Median :4.000    Median :1.0000
##  Mean   :0.3224   Mean    :4.225   Mean    :0.5514
##  3rd Qu.:1.0000   3rd Qu.:5.000    3rd Qu.:1.0000
##  Max.   :1.0000   Max.    :6.000   Max.    :1.0000
##                   NA's    :1448


##
## Call:
## glm(formula = Attitude ~ Fit + Employment + Public_safety + Revenue +
##     Age + Sex, family = binomial, data = data_imp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7827  -0.1793  -0.0792   0.0341   3.4337
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -6.95236     0.42482 -16.365  < 2e-16 ***
## Fit            3.05161     0.09420  32.394  < 2e-16 ***
## Employment     0.59327     0.07681   7.724 1.13e-14 ***
## Public_safety -0.70309     0.08639  -8.138 4.01e-16 ***
## Revenue        2.16514     0.13910  15.565  < 2e-16 ***
## Age            0.04342     0.05725   0.758    0.448
## Sex            0.09716     0.13936   0.697    0.486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 13243.5  on 10199  degrees of freedom
## Residual deviance:  1744.6  on 10193  degrees of freedom
##   (1448 observations deleted due to missingness)
## AIC: 1758.6
##
## Number of Fisher Scoring iterations: 8
```