What Makes Us Happy? Analyzing the Role of Life Satisfaction and Social Factors in Predicting Happiness in Mexico

Edgar Bruno Diaz Castro

University of Pennsylvania, School of Social Policy and Practice

MSSP 8970: Applied Linear Modeling

Instructor: Dr. Momin Malik

Policy Research Report 4

May 11, 2025

INTRODUCTION

Understanding what drives happiness has become increasingly central to social policy research, especially as global well-being indices begin to complement, and in some cases challenge, traditional economic measures such as GDP (Stiglitz, Sen, & Fitoussi, 2009). The World Happiness Report 2024 compiles well-being data from over 140 countries and ranks Mexico as the 10th happiest country in the world. Remarkably, Mexico is positioned just below countries like Finland, Denmark, Sweden, and the Netherlands (nations widely recognized for their advanced sociodemographic indicators) and above countries such as Colombia, Chile, Uruguay, and Argentina, despite their comparatively similar scores on several sociodemographic indicators, including income equality, educational attainment, and institutional trust.

This paradoxical placement raises important questions about the sociocultural and emotional dimensions that underpin subjective well-being in the Mexican context. In particular, it invites closer examination of the specific elements that contribute most significantly to happiness in Mexico, elements that, while they may be present elsewhere, may be uniquely valued, prioritized, or experienced more intensely by the Mexican population.

Addressing the concept of happiness more specifically, in both academic and policy settings, it is no longer viewed as a peripheral or elusive concept. It is increasingly conceptualized as a multidimensional construct comprising emotional experiences, cognitive judgments, and life satisfaction across key domains (Diener et al., 1985; Veenhoven, 1994). Specifically, subjective well-being reflects both affective states (e.g., joy, tranquility) and evaluative judgments about life circumstances, including relationships, health, work, and income (Ryan & Deci, 2001; Flores Cano, 2023). In Mexico, family, love, and joy are consistently identified as central elements of happiness (Flores Cano, 2023), suggesting a culturally embedded relational orientation toward well-being that may be less dependent on material affluence.

However, the question remains: do the exact life domains influence happiness equally for all? **Research in Mexico has revealed that income has only a weak relationship with reported happiness** (Arizpe et al., 2021), **while psychological, social, and demographic variables play a much stronger role**. Similarly, women's happiness is more closely tied to romantic relationships and self-image, whereas men's is more influenced by friendships (Sánchez Aragón & Méndez Canales, 2011). This gendered pattern highlights the need for disaggregated analysis when investigating subjective well-being.

Comparative **studies also suggest that satisfaction with specific life domains**, such as health, safety, or time use, can significantly mediate the effects of broader structural variables like income or education on happiness (Ballas & Tranmer, 2012). Moreover, **municipal-level surveys in Mexico show that social connection and perceived warmth in human relationships are among the strongest predictors of happiness across regions**, even surpassing satisfaction with public services (XXIV Seminario de Economía Urbana y Regional, 2012).

RESEARCH QUESTIONS

All the previously mentioned findings align with broader evidence that highlights Mexico's relatively high ranking in subjective well-being, despite lower scores in institutional trust and income equality compared to other OECD countries. Such patterns reinforce the idea that culturally embedded values, such as family ties, community cohesion, and emotional closeness, may buffer or even outweigh the effects of structural disadvantage in determining life satisfaction.

Recognizing the complexity of these relationships, this study proposes the following research questions:

1. To what extent does income influence general happiness in the Mexican population, and is this relationship mediated or moderated by satisfaction with one's economic situation?
2. Among the six measured life satisfaction domains (life, family, affective life, social life, economic situation, and housing), which most strongly predict overall happiness in Mexico, and how do these effects vary when controlling for structural variables such as age, education, and gender?
3. Does the combination of life satisfaction domains and specific demographic factors (age, gender, and marital status) predict the likelihood of reporting high happiness differently across socio-demographic groups in Mexico?

4. What implications do these patterns have for public policy aimed at enhancing well-being beyond material wealth?

Through this approach, the study aims not only to quantify the drivers of happiness in Mexico but also to expose the social inequalities embedded in its distribution. Ultimately, this analysis tries to contribute to a growing body of research advocating culturally grounded, multidimensional, and equity-focused well-being policies that transcend purely economic conceptions of development.

METHODS

Data and Variables

The dataset used in this analysis is drawn from the *Encuesta Nacional sobre el Uso del Tiempo* (ENUT), a national survey conducted in Mexico to assess time use, well-being, and socioeconomic characteristics across households and individuals. The ENUT dataset comprehensively examines respondents' demographic profiles, labor engagement, household characteristics, and self-reported well-being across multiple life domains. For this study, a cleaned and transformed subset of individual-level data was used, containing no missing values across the variables of interest, with 37,676 observations among 45 different variables. The following section describes and underlines those variables used for analysis.

The dependent variable in this study is **general happiness** (P7_3), which is measured on a 5-point Likert scale, ranging from 1 (not happy at all) to 5 (completely happy). The key predictors include self-reported **satisfaction across six life domains**, also measured on a 1–5 scale, where 5 represents the highest reported value. Together, these six domain-specific satisfaction variables form the basis for assessing the structure of well-being.

- **P7_2_1:** Satisfaction with life overall
- **P7_2_2**: Satisfaction with family life
- **P7_2_3**: Satisfaction with affective life (e.g., romantic relationships)
- **P7_2_4:** Satisfaction with social life
- **P7_2_5:** Satisfaction with economic situation

- **P7_2_6:** Satisfaction with housing conditions

Several **individual-level demographic and socioeconomic variables** are also used in this research:

- **AGE:** Respondent's age (in years)
- **ED_YEARS:** Total years of formal education completed
- **SEX_dummy:** A binary variable where 1 = male and 0 = female
- **married_binary:** A binary marital status variable coded as one (1) if the respondent is married or living with a partner, and zero (0) otherwise
- **INDIG_DUMMY**: A binary indicator of Indigenous identity (where 1 = Indigenous, 0 = non-Indigenous)

Two labor and economic variables are also considered:

- **INC_ANUAL:** Annual income in Mexican pesos
- **WORK_HRS:** Total hours worked per week.

From this, we created **WORK_HRS_YEAR**, an annualized version of work hours by multiplying weekly hours by 52. To maintain the analytic focus on the economically active population, the dataset was further filtered to include only individuals who reported working and provided valid responses for the amount of work performed. This adjustment aligns with the study's objective of analyzing the relationship between labor engagement, life satisfaction, and happiness.

Respondents with zero hours worked were excluded from the analysis, as they do not participate in the labor market and represent a distinct social category (e.g., retirees, students, homemakers) that would introduce heterogeneity not aligned with the research aims. This step ensured that the analysis accurately reflects the dynamics of work-related satisfaction and its connection to subjective well-being.

It is also important to mention that the dataset collected by ENUT includes only individuals aged 12 years and older, reflecting the survey's scope to capture time-use and well-being metrics of both adolescents and adults. However, for this analysis, only those who participated in the labor market are included, ensuring that the analysis of WORK_HRS_YEAR is both meaningful and consistent with the research questions.

Modelling Techniques

We apply an **iterative modeling approach** to explore the relationship between life domain satisfaction, economic factors, and general happiness in Mexico. Our analysis began with a series of **Ordinary Least Squares (OLS) linear regression models** to understand how life satisfaction across key domains influences overall happiness (P7_3). However, recognizing the limitations of treating happiness as strictly continuous and observing non-linear patterns in satisfaction, we progressed to **categorical specifications** and ultimately to a **binary logistic regression** to more accurately capture the probability of high happiness (happy_binary). This allowed us to refine our understanding and enhance model interpretability, aligning the statistical methods more closely with the nature of the data and the research questions.

The initial model (**Model Q1**) examined the effect of annual income (log-transformed and mean-centered as log_INC_ANUALc) and satisfaction with the economic situation (centered as P7_2_5c) on overall happiness, while also including an interaction term to explore whether economic satisfaction moderates the effect of income on happiness.

Subsequently, we constructed a more comprehensive **OLS linear regression model (Q2c)** that included six key life satisfaction domains (life, family, affective life, social life, economic situation, and housing), along with demographic and socioeconomic controls such as age, education, marital status, indigenous status, work hours, and log-transformed income. All continuous variables involved in interactions were mean-centered, and skewed variables were log-transformed to improve interpretability and satisfy linear modeling assumptions.

To capture non-linear jumps in happiness levels, we extended the model by treating life satisfaction domains as categorical variables (**Q2c Categorical**). Finally, to better address the discrete nature of happiness at high levels, we employed a binary logistic regression model. This model focused on predicting the likelihood of high happiness (happy_binary, where values below or equal to 3 are treated as low happiness and values above 3 are treated as high happiness) and included key interaction terms between satisfaction domains and demographic factors, such as age, gender, marital status, and income. The logistic regression framework allowed us to interpret effects in terms of odds ratios, enhancing our understanding of how satisfaction and structural variables jointly influence the probability of high happiness.

This iterative approach, which involved transitioning from standard linear regression to categorical specifications and then to logistic modeling, provided richer insights into the social and economic determinants of happiness. Model comparisons using ANOVA and R-squared statistics validated the improvements in explanatory power with each step, culminating in a more nuanced understanding of happiness dynamics in the Mexican context.

DESCRIPTIVE STATISTICS

Table 1 below presents the summary statistics for all continuous variables. Table 2 provides counts for binary variables within the dataset, and Table 3 describes the same for categorical variables, which are scaled from 1 to 5. The analytic sample consists of 37,676 respondents, all of whom have complete data on the variables used in the analysis.

Table 1. Descriptive Statistics for Continuous Variables

Variable	Unit	Min	Max	Mean	Median	S.D.
INC_ANUAL	Mexican Pesos	$80	$3,640,000	$87,008	$67,600	$87,083
AGE	Years	12	97	38.58	37	14.59
WORK_HRS_YEAR	Hours	52	7,687	2,287	2,392	1,045.21
ED_YEARS	Years	0	22	10.28	9	4.11

*There are n=**37,676** observations with no missing values across all variables.*

As shown in Table 1, the annual income (INC_ANUAL) ranges from a low of $80 to a high of 3.64 million Mexican pesos, with a mean of $87,008 and a high standard deviation of $87,083, reflecting substantial income inequality within the sample. Respondents' ages range from 12 to 97 years, with an average of 38.6 years. Work hours per year (WORK_HRS_YEAR) vary widely, from 52 to 7,687, with a mean of 2,287 hours, indicating substantial variation in labor force participation. Finally, the variable for education, measured in years, has a mean of 10.28 years and a median of 9, consistent with mid-level educational attainment in Mexico.

Table 2. Descriptive Statistics for Binary Variables

Variable	Type	1 (Yes)	0 (No)	N
married_binary	Binary	15,324	22,352	37,676
INDIG_DUMMY	Binary	21,415	16,261	37,676
SEX_dummy	Binary	15, 637 (Men)	22,039 (Women)	37,676

There are n=37,676 observations with no missing values across all variables.

Table 2 summarizes three binary variables. The majority of respondents are not married or cohabiting (59.3%), with 40.7% coded as married (married_binary). About 56.8% of respondents identify as Indigenous, while 43.2% identify as non-Indigenous (INDIG_DUMMY). Regarding gender, 58.5% are female (SEX_dummy = 0) and 41.5% are male.

Table 3. Descriptive Statistics for Categorical Variables

Variable	Type	1	2	3	4	5
P7_3 (Level of Happiness)	Categorical	149	842	4,929	20,284	11,472
P7_2_1 (Life satisfaction)	Categorical	247	1,029	6,213	23,507	6,680
P7_2_2 (Family life)	Categorical	204	698	3,518	23,634	9,622
P7_2_3 (Affective life)	Categorical	294	858	4,067	23,869	8,588
P7_2_4 (Social life)	Categorical	650	2,084	6,645	23,686	4,611
P7_2_5 (Economic situation)	Categorical	2,108	4,687	13,013	15,693	2,175
P7_2_6 (Housing conditions)	Categorical	839	2,119	6,566	23,230	4,922

There are n=37,676 observations with no missing values across all variables.
Likert scale 1 to 5, where 5 is the highest value possible

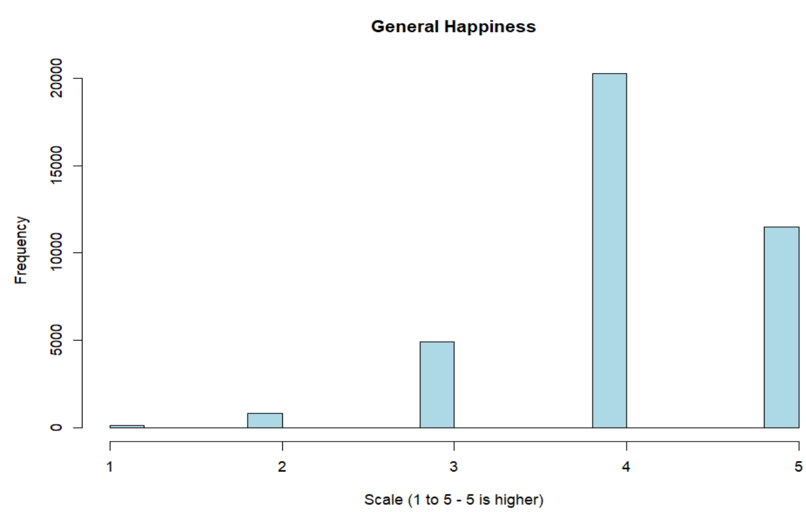
As shown in Table 3, general happiness (P7_3) is skewed toward high values: 83% of respondents report happiness levels of 4 or 5. A similar pattern is observed across all six satisfaction domains (P7_2_1 to P7_2_6), where most responses are concentrated at the top two levels (4 and 5). For example, 63% report a satisfaction level of 4 or 5 with life overall (P7_2_1), and 78% report the same for family life (P7_2_2). Economic satisfaction (P7_2_5) shows relatively more spread, with nearly 17% of respondents reporting low to moderate satisfaction levels (1 or 2), reflecting higher variation in perceived financial well-being.

EXPLORATORY DATA ANALYSIS

Before proceeding to model estimation, all variables were examined through histogram visualization and the previous summary statistics to assess their distribution and determine whether transformations were necessary. These steps ensured that the assumptions underlying linear regression were reasonably satisfied and that coefficients would be interpretable within the context of interaction terms.

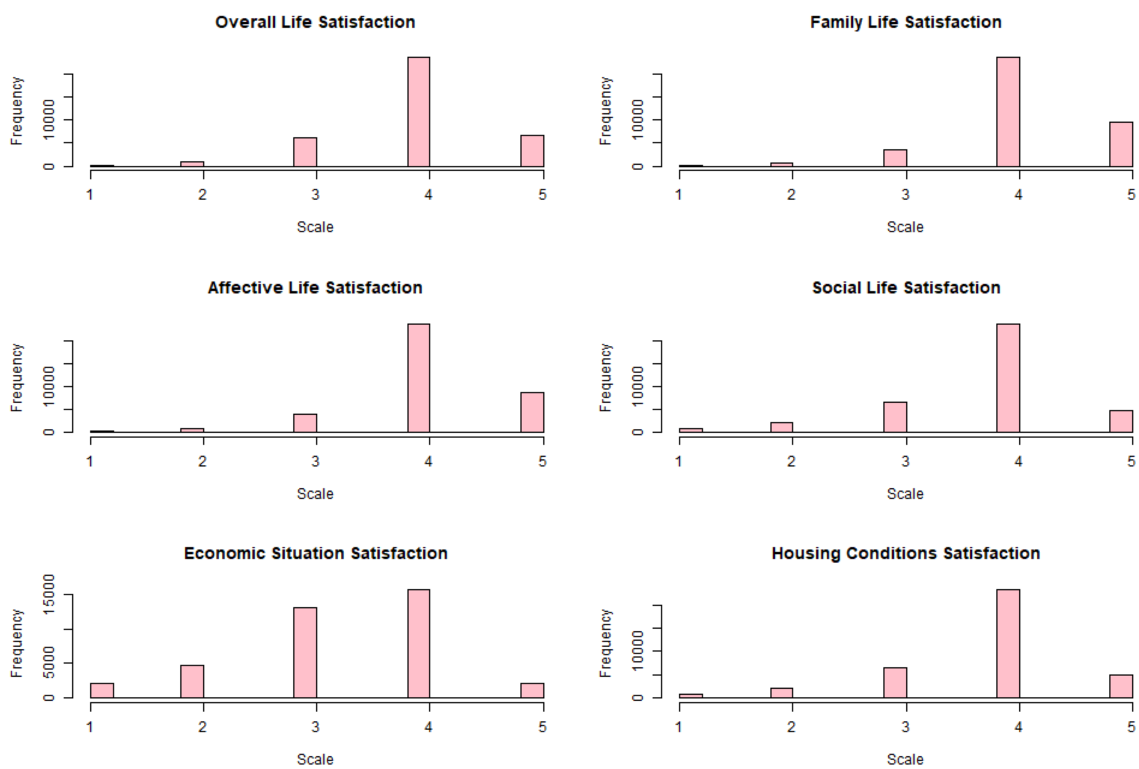
The dependent variable in this analysis is general happiness **(`P7_3`),** measured on a five-point Likert scale from 1 (not happy at all) to 5 (completely happy). According to Figure 1, **the distribution is slightly left-skewed, with most respondents reporting high levels of happiness (values of 4 or 5).** Given the mildness of the skew, the data itself, and the robustness of OLS to this form of non-normality, no transformation was applied.

Figure 1. Distribution of General Happiness Levels.



The primary predictors of interest (Figure 2) are satisfaction across six life domains: life overall (P7_2_1), family (P7_2_2), affective life (P7_2_3), social life (P7_2_4), economic situation (P7_2_5), and housing (P7_2_6). **All six variables are measured on the same 1–5 scale and display similarly shaped distributions, slightly left-skewed**, with a majority of responses clustering at the higher end of the scale. While transformation for normality was not necessary, all six were mean-centered. This facilitates interpretation in the presence of interaction terms and ensures that the intercept in the model represents an "average" individual in terms of life satisfaction.

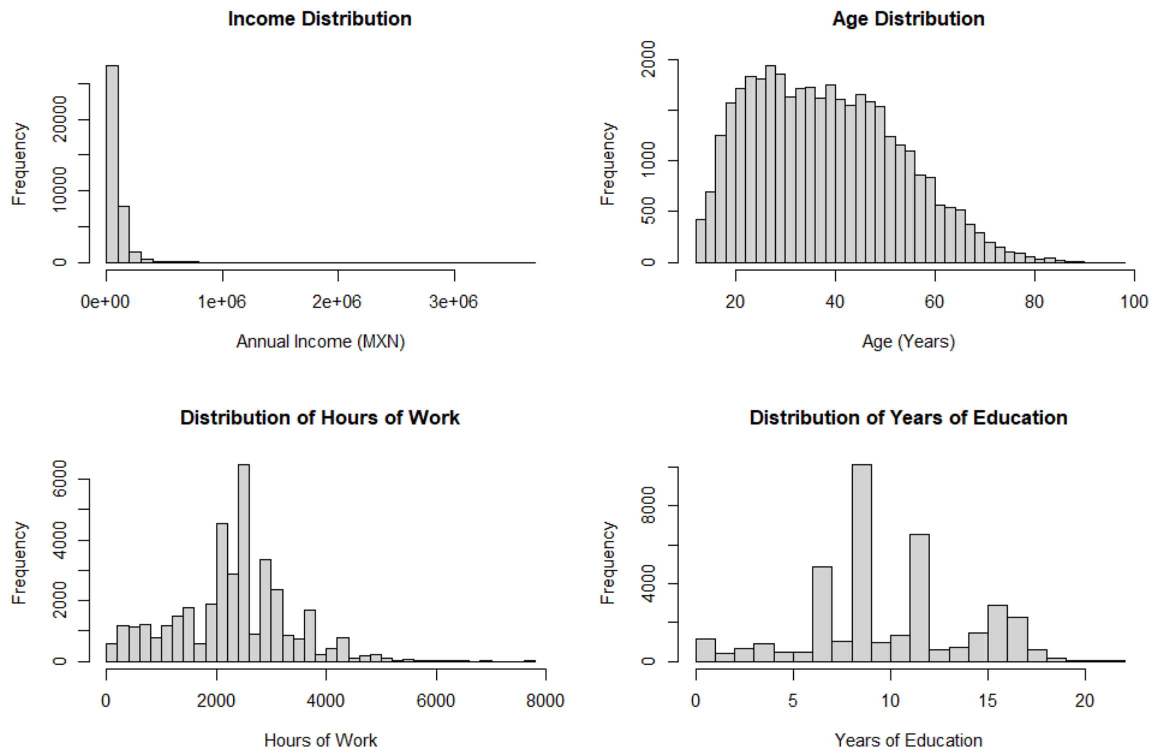
Figure 2. Distributions for satisfaction levels



Demographic characteristics were also assessed through Figure 3. Age (AGE) **displayed a left-skewed distribution, with most respondents being younger**. While the shape did not require transformation, the variable was mean-centered to improve interpretability, particularly in interaction terms. Education in years (ED_YEARS) **had a somewhat multimodal distribution, reflecting educational cutoffs in the Mexican school system**. Like age, it was centered to enhance clarity in the model.

Among the continuous variables (Figure 3), annual income (INC_ANUAL) **showed a highly right-skewed distribution, with a long tail of high-income individuals**. To address this, a natural logarithm transformation was applied, yielding log_INC_ANUAL. Reducing skew and mitigating the influence of outliers. Similarly, work hours per week were converted to annual work hours (WORK_HRS_YEAR) and show a small right-skewed distribution. However, no transformation was applied because of the nature of the variable.

Figure 3. Distributions for continuous variables.



Several binary variables were created or adapted for analysis. Sex was recoded into a dummy variable (SEX_dummy), where 1 indicates male and 0 female. Marital status was recoded into a binary variable (married_binary), where 1 indicates the respondent is married or living with a partner and 0 otherwise. Indigenous identity was included as INDIG_DUMMY, a binary variable indicating whether the respondent identifies as Indigenous. As these were already coded appropriately for use in regression, no transformations were applied.

To assess the correlation between variables, we created a correlation matrix (Table 4), which reveals **moderate correlations with Satisfaction with Family Life** (P7_2_2) and **Social Life** (P7_2_4), highlighting the significant role of close relationships and social connections in contributing to overall well-being. Additionally, **Satisfaction with Affective Life (P7_2_3) and Economic Situation (P7_2_5) exhibit meaningful positive relationships with general happiness**, suggesting that both emotional stability and financial security are influential components of well-being.

The correlations with demographic factors reveal nuanced relationships. Age (AGE) exhibits a slight negative correlation with general happiness, potentially suggesting a gradual decline in subjective well-being as individuals age. This could reflect changes in health, social support, or economic stability over the life course. Meanwhile, years of education (ED_YEARS) display a weak positive correlation with happiness and a slightly stronger relationship with Economic Satisfaction (P7_2_5), indicating that while education contributes to financial stability, its direct effect on happiness may be more limited.

In terms of economic indicators, the log-transformed annual income (log_INC_ANUAL) exhibits a modest positive correlation with Economic Satisfaction, supporting the notion that financial security influences perceptions of economic well-being. However, its correlation with overall happiness is notably weaker, aligning with previous findings that income alone does not guarantee greater happiness. Annual work hours (WORK_HRS_YEAR) display very little correlation with both life satisfaction and general happiness, suggesting that mere participation in the workforce or time spent working does not necessarily translate to enhanced well-being.

Although some variables have significant correlations among them, these are all less than 0.6. Therefore, there do not appear to be any potential collinearity issues that might affect the models to be created.

Table 4. Correlation Analysis

Variables	Correlation coefficient										
	P7_3	Log_ INC_ANUAL	AGE	WORK_HRS_ YEAR	ED_YEARS	P7_2_1	P7_2_2	P7_2_3	P7_2_4	P7_2_5	P7_2_6
P7_3	**1.00**	0.12	-0.10	0.00	0.17	0.49	0.45	0.44	0.33	0.33	0.32
log_INC_ ANUAL	0.12	**1.00**	0.00	0.44	0.39	0.12	0.11	0.11	0.05	0.18	0.09
AGE	-0.10	0.00	**1.00**	0.00	-0.26	-0.04	-0.05	-0.01	-0.03	-0.06	-0.02
WORK_HRS_ YEAR	0.00	0.44	0.00	**1.00**	0.04	-0.01	-0.01	0.00	-0.03	0.01	-0.02
ED_YEARS	0.17	0.39	-0.26	0.04	**1.00**	0.13	0.13	0.13	0.05	0.12	0.11
P7_2_1	0.49	0.12	-0.04	-0.01	0.13	**1.00**	0.55	0.48	0.39	0.39	0.34
P7_2_2	0.45	0.11	-0.05	-0.01	0.13	0.55	**1.00**	0.57	0.37	0.28	0.31
P7_2_3	0.44	0.11	-0.01	0.00	0.13	0.48	0.57	**1.00**	0.41	0.29	0.28
P7_2_4	0.33	0.05	-0.03	-0.03	0.05	0.39	0.37	0.41	**1.00**	0.33	0.27
P7_2_5	0.33	0.18	-0.06	0.01	0.12	0.39	0.28	0.29	0.33	**1.00**	0.41
P7_2_6	0.32	0.09	-0.02	-0.02	0.11	0.34	0.31	0.28	0.27	0.41	**1.00**

Finally, throughout our analysis, a p-value threshold of 0.05 is applied to determine statistical significance; any coefficient with a p-value above this threshold is not considered statistically significant. To assess the explanatory power of the models, we rely on the Adjusted R^2, which measures the proportion of variance in the response variable explained by the model, while accounting for the number of predictors. This metric also serves to evaluate improvements achieved through the inclusion of additional variables or changes in modeling techniques. Additionally, we perform Partial F-tests to compare the goodness of fit between different models, allowing us to identify significant enhancements in predictive capability when new terms are introduced.

RESULTS

To address the question of how income influences general happiness within the Mexican population and whether this relationship is moderated by satisfaction with one's economic situation, a linear regression model (modelQ1) was constructed. This model examines the direct effect of annual income (log-transformed and mean-centered as log_INC_ANUALc) and satisfaction with economic situation (centered as P7_2_5c) on general happiness (P7_3). The interaction term between income and economic satisfaction enables exploration of whether the impact of income on happiness shifts depending on perceived financial well-being. We chose to use centered variables to allow the coefficients to be interpreted relative to the average Mexican citizen's income and financial satisfaction, ensuring that the intercept represents the predicted happiness for someone with average income and satisfaction levels.

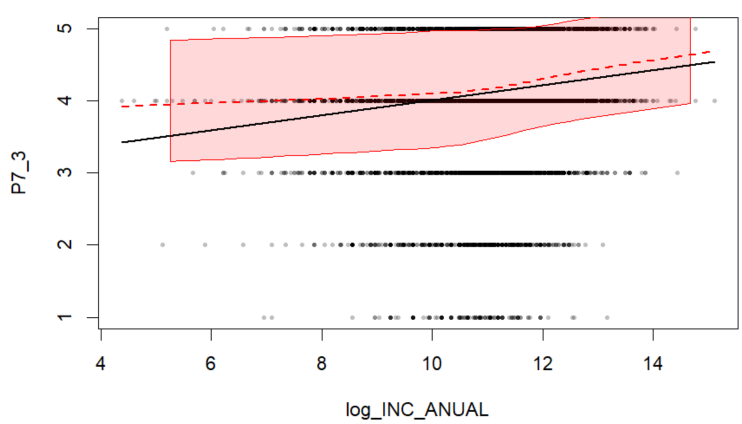
Table 5. Coefficients for modelQ1

	Dependent variable: General Happiness (P7_3)
log_INC_ANUALc	0.055*** (0.004)
P7_2_5c	0.248*** (0.004)
log_INC_ANUALc : P7_2_5c	-0.014*** (0.004)
Constant	4.119*** (0.004)
Observations	37,676
R^2	0.114
Adjusted R^2	0.114
Residual Std. Error	0.696 (df = 37672)
F Statistic	1,623.068*** (df = 3; 37672)
	*p<0.1; **p<0.05; ***p<0.01

The positive coefficient for **log_INC_ANUALc** (0.055, $p < 0.01$) suggests that, on average, as the log of annual income increases, there is a corresponding increase of 0.055 units in happiness, assuming economic satisfaction remains constant at its average level. Because the variable is centered, the interpretation is relative to the mean value of log_INC_ANUAL. Similarly, the coefficient for **P7_2_5c** (0.248, $p < 0.01$) indicates a strong positive association with happiness. Since this variable is also centered, its positive effect is interpreted as an increase in happiness relative to its average value, suggesting that individuals who score above the average on P7_2_5c tend to report higher levels of happiness. This effect is more pronounced than income, signaling that perceived financial security has a stronger direct link to happiness compared to variations in raw income.

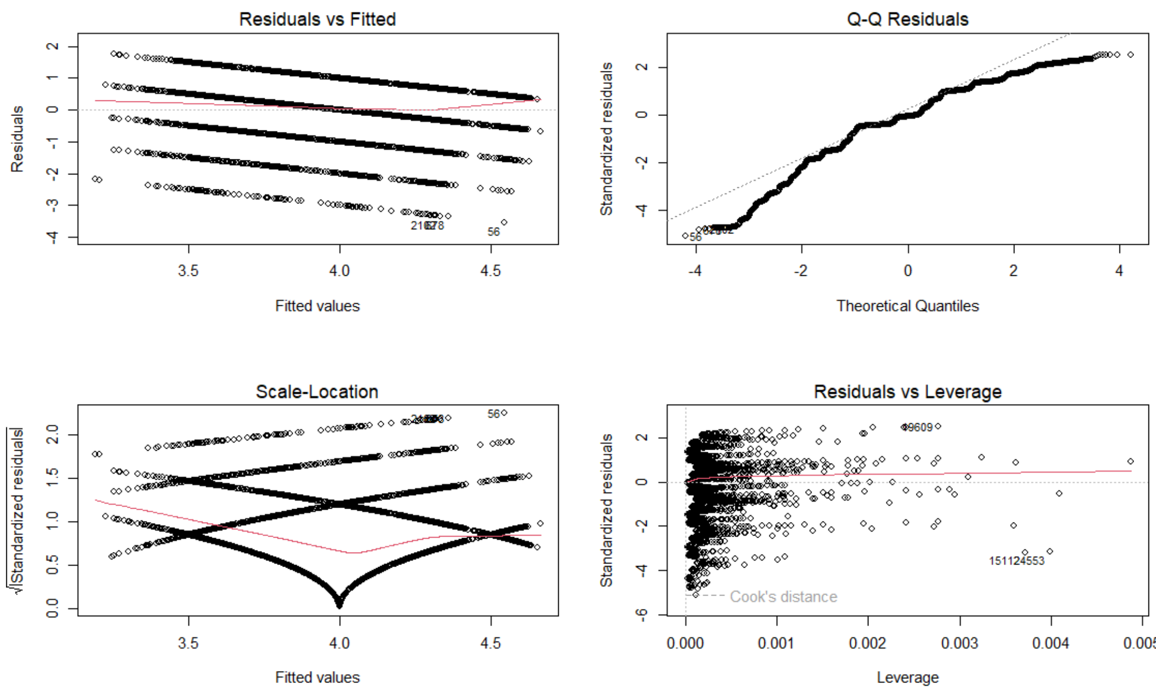
The interaction term (log_INC_ANUALc: P7_2_5c) is negative (**-0.014**, $p < 0.01$), implying that the positive impact of income on happiness is **diminished as economic satisfaction increases**. Since both terms are centered, this means that **at average levels of satisfaction**, income still has a positive influence on happiness; however, as individuals report higher-than-average satisfaction with their economic situation, the marginal effect of additional income becomes slightly less impactful. Conversely, when satisfaction is lower than average, additional income has a more substantial influence on happiness. This interaction showcases a **moderating effect**, where the benefits of increased income are partially contingent on how satisfied individuals feel about their financial status.

Figure 4. Relationship between Happiness and Income



The scatterplot visualization (Figure 4) highlights the relationship between **log_INC_ANUAL** and **P7_3**, showing a general upward trend where income increases are associated with higher happiness levels. However, the red dashed line suggests potential non-linearity in this relationship, as the fit does not perfectly align with the linear trend. This misalignment could be improved by adding more explanatory variables, which could expand the model's ability to capture variations in happiness and potentially address the non-linear behavior observed. The model currently explains 11.4% of the variance, indicating room for improvement, but this is significant given the complexity of happiness as a construct.

Figure 5. Diagnostic plots for modelQ1



The residual diagnostic plots for modelQ1 (Figure 5) provide insight into the adequacy of the model in predicting general happiness based on income and economic satisfaction. The **Residuals vs. Fitted** plot displays distinct "stripes," reflecting the **ordinal nature of happiness (P7_3)** rather than a modeling error. The **Q-Q plot** shows deviations from normality, particularly in the tails, which is expected given the limited predictors and the ordinal nature of happiness. The **Scale-Location plot** suggests mild heteroscedasticity, which is not alarming but could be improved with a richer model specification. The Residuals vs. Leverage plot shows no significant outliers that can influence the model. Overall, these diagnostics imply that while the model performs efficiently, the happiness-income relationship is **incomplete**, as many other factors likely influence happiness.

The following three models (Q2a, Q2b, and Q2c) progressively expand on the initial analysis by introducing additional control variables to predict **general happiness (P7_3)** and help answer our second question. While Model A focuses solely on satisfaction levels across various life domains (P7_2_1c to P7_2_6c), Model B introduces key demographic variables like **age, education, gender, marital status, and indigenous status**. Finally, Model C further incorporates **work hours per year and income**, providing a more comprehensive view of the determinants of happiness.

Table 6. Models Predicting General Happiness with Satisfaction Levels and Sociodemographic Variables

	Dependent variable:		
	General Happiness (P7_3)		
	(Q2a)	(Q2b)	(Q2c)
P7_2_1c	0.240*** (0.006)	0.237*** (0.006)	0.237*** (0.006)
P7_2_2c	0.166*** (0.006)	0.152*** (0.006)	0.152*** (0.006)
P7_2_3c	0.168*** (0.006)	0.160*** (0.006)	0.160*** (0.006)
P7_2_4c	0.058*** (0.005)	0.063*** (0.005)	0.063*** (0.005)
P7_2_5c	0.072*** (0.004)	0.066*** (0.004)	0.065*** (0.004)
P7_2_6c	0.086*** (0.004)	0.090*** (0.004)	0.090*** (0.004)
AGEc		-0.004*** (0.0002)	-0.004*** (0.0002)
ED_YEARSc		0.011*** (0.001)	0.010*** (0.001)
SEX_dummy		-0.031*** (0.006)	-0.035*** (0.007)
married_binary		0.078*** (0.007)	0.077*** (0.007)
INDIG_DUMMY		0.014*** (0.006)	0.016** (0.006)
WORK_HRS_YEARc			-0.0000 (0.0000)
log_INC_ANUALc			0.013*** (0.005)
Constant	4.117*** (0.003)	4.083*** (0.006)	4.085*** (0.007)
Observations	37,676	37,676	37,676

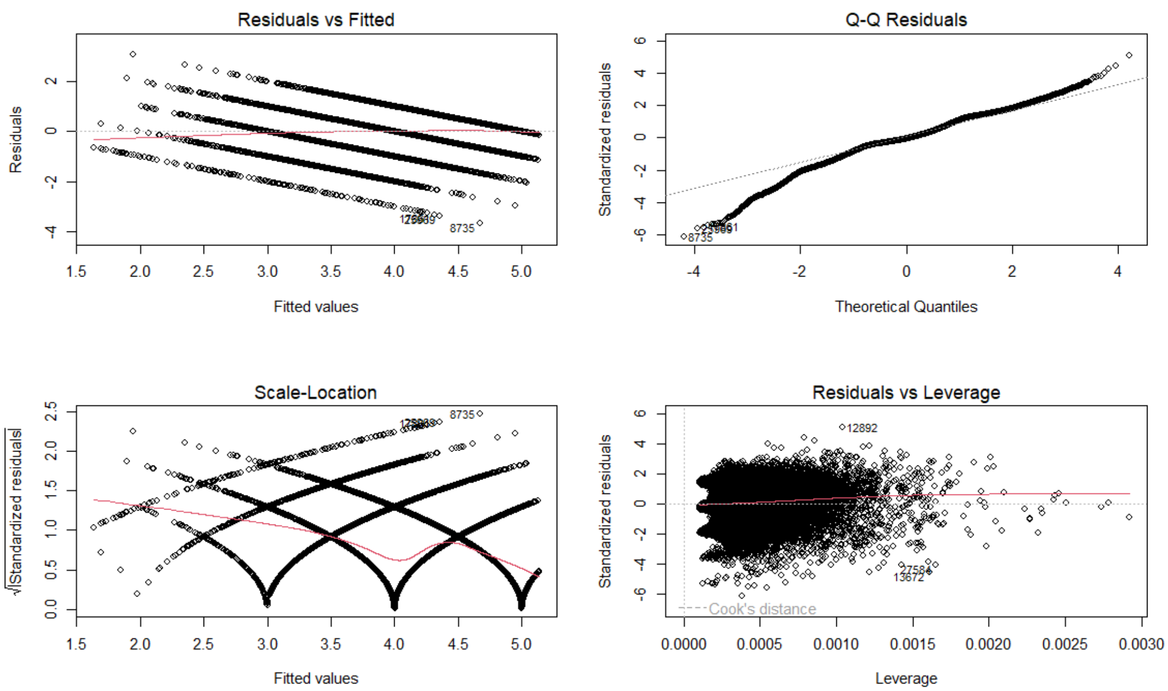
R^2	0.333	0.344	0.345
Adjusted R^2	0.333	0.344	0.344
Residual Std. Error	0.604 (df = 37669)	0.599 (df = 37664)	0.599 (df = 37662)
F Statistic	3,134.071*** (df = 6; 37669)	1,798.985*** (df = 11; 37664)	1,523.066*** (df = 13; 37662)
			**p<0.1; **p<0.05; ***p<0.01*

Model Q2a establishes a baseline by exploring how different satisfaction domains influence happiness. All coefficients are significant at $p < 0.01$, indicating that satisfaction in areas such as family, affective life, social life, and economic situation is positively associated with happiness. Among these, **P7_2_1c** (0.240) has the strongest positive effect, suggesting that variations in this specific dimension, when centered, are more strongly linked to changes in happiness. This model achieves an **R^2 of 0.333**, indicating that these six dimensions account for approximately 33.3% of the variance in happiness, representing a considerable improvement over the previous model.

Model Q2b builds upon this by adding demographic controls, slightly improving the explanatory power to **34.4% (R^2 = 0.344)**. Interestingly, being married and identifying as indigenous are positively associated with higher happiness, while being male has a small adverse effect. Notably, **AGEc** (-0.004, $p < 0.01$) has a small but negative association with happiness, indicating that, on average, aging slightly reduces reported happiness when other factors are controlled. Conversely, being **married_binary** (0.078, $p < 0.01$) shows a positive effect, reinforcing the idea that marriage status correlates positively with happiness. The coefficient for **SEX_dummy** is negative (-0.031, $p < 0.01$), suggesting that gender differences influence happiness, with men potentially reporting slightly lower happiness on average, all else being equal. All other satisfaction levels saw a decrease in their coefficient, except P7_2_4c and P7_2_6c, which reported a slight increment.

Finally, **Model Q2c** incorporates additional variables (WORK_HRS_YEARc and log_INC_ANUALc). While WORK_HRS_YEARc is not statistically significant, log_INC_ANUALc (0.013, $p < 0.01$) introduces a meaningful, albeit modest, positive association with happiness. This suggests that income still plays a role, but its effect is smaller compared to the core predictors identified since model Q2a. The **R^2 of 0.345** reflects a marginal increase in explanatory power, indicating that while income contributes, its effect size is smaller relative to the psychological and social predictors. Nevertheless, it still leaves room for further improvements, suggesting that other unexplored factors might be influencing happiness.

Figure 6. Diagnostic plots for model Q2c



The residual diagnostics (Figure 6) for the stronger model (Q2c) maintain the "striped" appearance in the Residuals vs. Fitted plot, which continues to reflect the ordinal nature of happiness (P7_3). The Q-Q plot still shows some deviation from normality, particularly in the tails, indicating that residuals are not entirely normally distributed, but an improvement from Q1. The Scale-Location plot displays some variance instability, hinting at heteroscedasticity, though it is not severe enough to invalidate the model's findings.

Finally, the Residuals vs. Leverage plot and Cook's Distance indicators do not reveal any significant outliers with high leverage that would disproportionately influence the model's estimates. The points are mostly well distributed, with no extreme cases crossing the Cook's Distance threshold. These findings suggest the necessity of treating the satisfaction indicators as categorical. For that, boxplots of these variables were created to assess their nature and spread.

Figure 7. Boxplots for Satisfaction Indicators

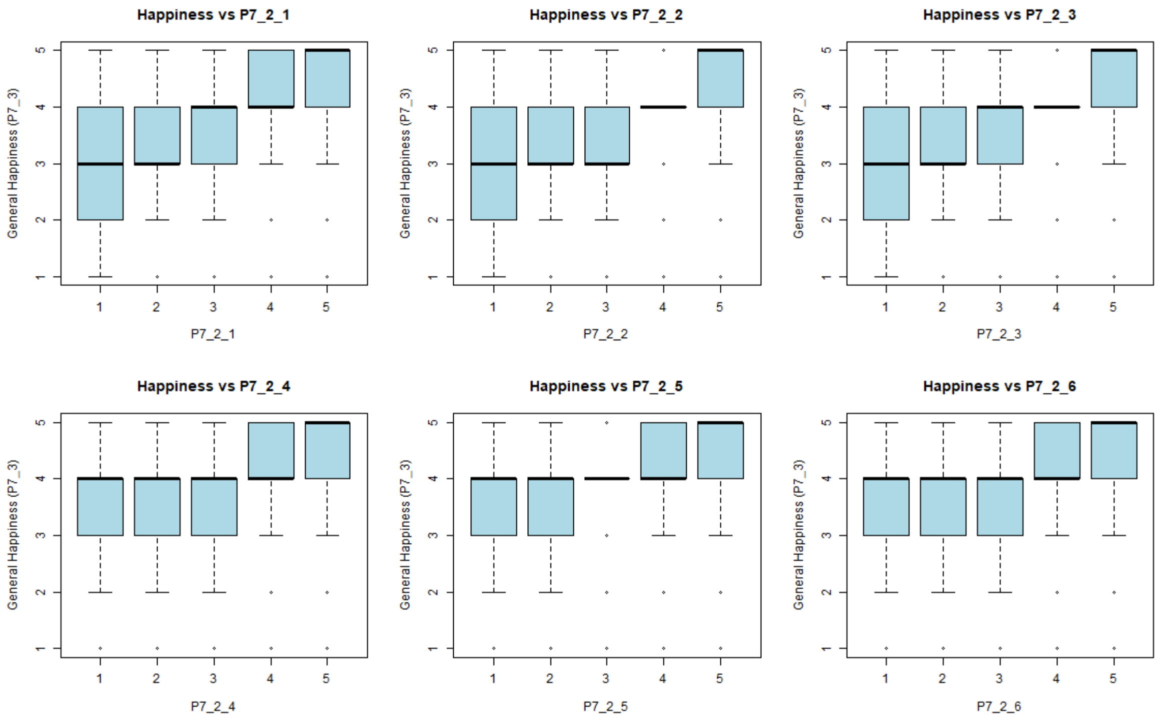
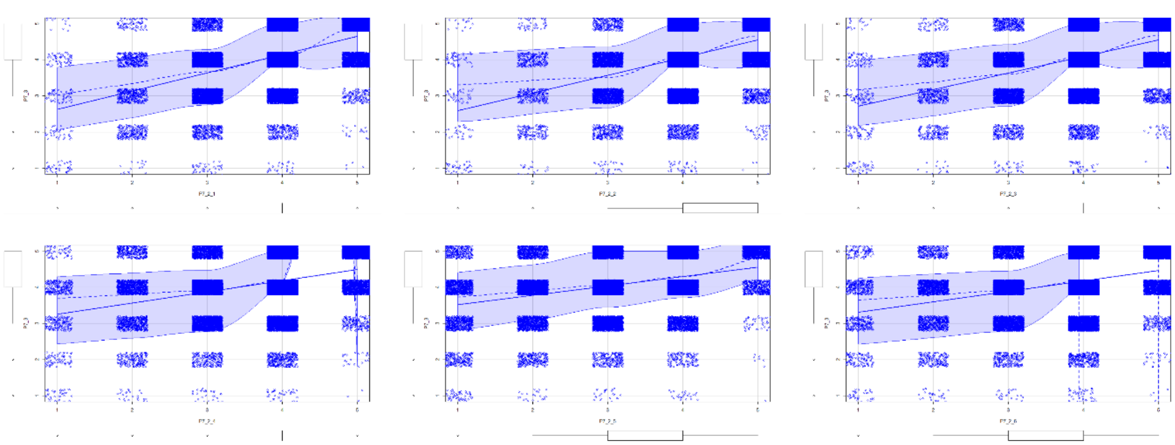


Figure 8. Relationship between Happiness and Satisfaction Indicators



The **box plots and scatterplots** (Figures 7 and 8) for the variables reveal distinct, non-linear relationships with happiness that are more characteristic of **categorical behavior**. Treating these variables as categorical may capture these stepwise shifts in happiness more effectively, providing a better fit and possibly smoothing out the residual variance.

The following model (Q2c Categorical) extends the previous analysis by treating satisfaction variables (P7_2_1 to P7_2_6) as **categorical factors** instead of continuous predictors. This shift allows the model to capture **non-linear relationships** and distinct jumps in happiness levels across different satisfaction categories. Additionally, the model controls for the previously key demographic and socioeconomic variables, which aim to reflect more accurately how categorical shifts in satisfaction impact general happiness, thereby addressing the limitations noted in the diagnostic plots of the continuous models.

Table 7. Models Predicting General Happiness with Categorical Satisfaction Levels and Sociodemographic Variables

	Dependent variable:
	General Happiness (P7_3)
as.factor(P7_2_1)2	0.224***(0.044)
as.factor(P7_2_1)3	0.407*** (0.041)
as.factor(P7_2_1)4	0.693*** (0.041)
as.factor(P7_2_1)5	0.885*** (0.042)
as.factor(P7_2_2)2	0.072 (0.050)
as.factor(P7_2_2)3	0.159*** (0.046)
as.factor(P7_2_2)4	0.332*** (0.046)
as.factor(P7_2_2)5	0.502*** (0.047)
as.factor(P7_2_3)2	0.301*** (0.042)
as.factor(P7_2_3)3	0.352*** (0.039)
as.factor(P7_2_3)4	0.550*** (0.038)
as.factor(P7_2_3)5	0.698*** (0.039)
as.factor(P7_2_4)2	0.063** (0.028)
as.factor(P7_2_4)3	0.115*** (0.026)
as.factor(P7_2_4)4	0.188*** (0.025)
as.factor(P7_2_4)5	0.247*** (0.027)
as.factor(P7_2_5)2	0.079*** (0.016)
as.factor(P7_2_5)3	0.128*** (0.015)
as.factor(P7_2_5)4	0.209*** (0.015)
as.factor(P7_2_5)5	0.207*** (0.021)
as.factor(P7_2_6)2	0.074*** (0.025)
as.factor(P7_2_6)3	0.127*** (0.023)
as.factor(P7_2_6)4	0.228*** (0.022)
as.factor(P7_2_6)5	0.349*** (0.024)
AGE_c	-0.004*** (0.0002)
ED_YEARS_c	0.010*** (0.001)
SEX_dummy	-0.038*** (0.007)
married_binary	0.077*** (0.007)
INDIG_DUMMY	0.015** (0.006)
WORK_HRS_YEARc	-0.0000 (0.0000)
log_INC_ANUALc	0.013*** (0.005)

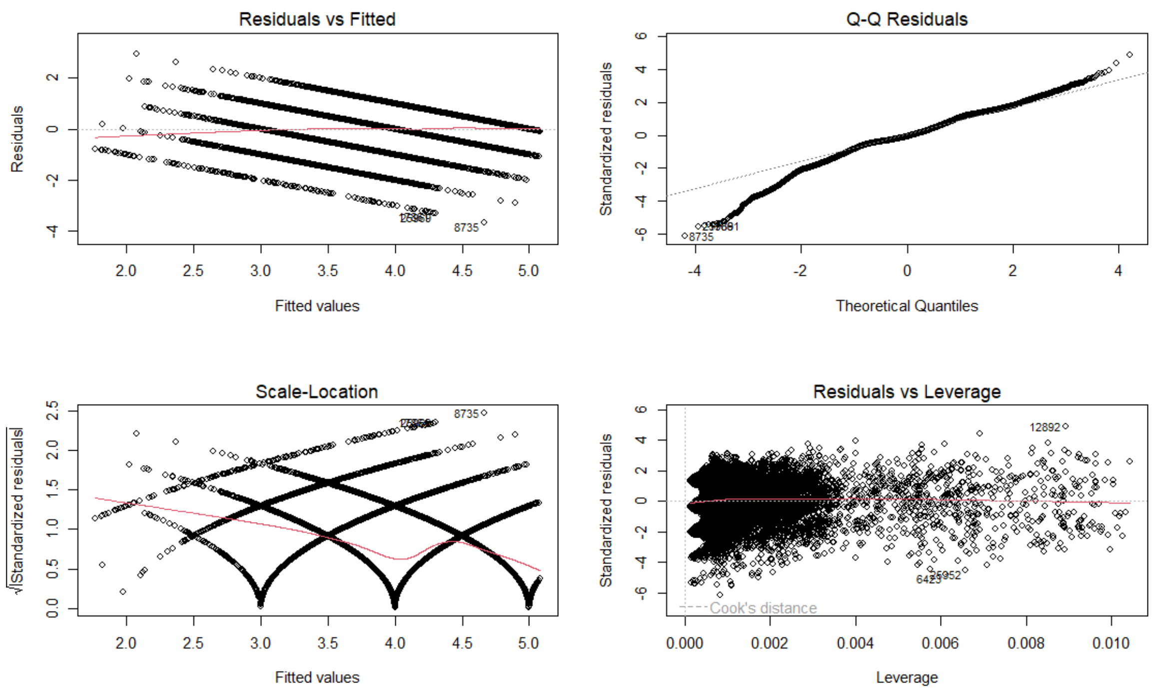
Constant	1.982*** (0.056)
Observations	37,676
R^2	0.348
Adjusted R^2	0.347
Residual Std. Error	0.597 (df = 37644)
F Statistic	647.303*** (df = 31; 37644)
	*p<0.1; **p<0.05; ***p<0.01

The analysis of the model Q2cCat reveals that treating the predictors P7_2_1 to P7_2_6 as categorical variables exposes substantial variations in their impact on **General Happiness (P7_3)** that were previously masked by linear assumptions. Notably, category **5 of P7_2_1** demonstrates the strongest positive association with happiness, contributing to a **0.885-point increase** relative to its reference group, marking it as the most impactful single factor in the model. This is followed by **category 4 of P7_2_1** with a **0.693-point increase**, indicating that higher levels within this factor are consistently linked to greater happiness.

Furthermore, **category 5 of P7_2_2** also shows a considerable positive effect, with a **0.502-point increase**. The same pattern is evident in **category 5 of P7_2_3** and **category 5 of P7_2_6**, which are associated with **0.698** and **0.349-point increases,** respectively. These results suggest that the impact of these predictors on happiness is highly **nonlinear** and varies significantly across categories, which would have been understated if modeled purely as continuous variables. This granular approach not only highlights the variability in influence across different categories but also underscores the need to consider **nonlinear modeling** or **interaction terms** to capture these dynamics fully.

Alongside this change, the socio-demographic variables maintain their influence, though slightly adjusted in magnitude. **AGEc** still reflects a low negative impact on happiness, while **ED_YEARSc** remains positively associated with it. The binary indicator for marriage also continues to show a positive relationship, suggesting that being married correlates slightly positively with higher happiness levels. Lastly, the effect of income, though slightly smaller, remains statistically significant. Overall, this categorical representation enhances the model's capacity to illustrate more nuanced effects and identify which specific groups contribute most to happiness variation, potentially guiding targeted interventions more effectively. The model's **R^2 is 0.3477**, which slightly improves upon the continuous specification.

Figure 8. Diagnostic plots for model Q2c Categorical



The residual diagnostics (Figure 8) for this model retain the striped pattern in the Residuals vs. Fitted plot, which reflects the constraints inherent in the data itself. The Q-Q plot exhibits a similar deviation from normality, primarily in the tails; however, this is likely due to the lack of control variables to account for a better model, especially given the categorical variables that allow for more possible combinations (as evident in the high differences between levels 3, 4, and 5). The Scale-Location plot reveals heteroscedasticity, though it is not severe given the characteristics of the data and its limits. These artifacts suggest that while categorical modeling might be better at representing the step-like changes in happiness, the model still does not fully capture the ordinal nature of the data. Additionally, the Residuals vs. Leverage plot and Cook's Distance indicate no significant outliers with disproportionate influence on the model estimates.

Despite these, the diagnostic plots indicate good model stability after the categorical transformation of key predictors. **However, further and final refinement could consider a logistic model** to better align with the discrete nature of happiness levels, potentially smoothing the variance and addressing the striped residual patterns.

The ANOVA (Table 8) test results reveal the progression of model improvements across four stages. Transitioning from Model 1 to Model 2 resulted in a substantial reduction in the residual sum of squares (RSS) and a highly significant F-statistic, indicating that adding demographic variables, such as age, education, sex, marital status, and indigeneity, significantly improved the model. As expected, the shift from Model 2 to Model 3, which introduced economic factors such as log annual income, resulted in only a minor reduction in RSS and a small F-statistic. This demonstrates that economic variables contributed little to explaining the differences observed in happiness, confirming the anticipated limited impact of income-related factors on this measure.

Table 8. ANOVA Test (Model Q2a, Q2b, Q2c and Q2c Categorical)

(1) Model Q2a: *P7_3 ~ P7_2_1c + P7_2_2c + P7_2_3c + P7_2_4c + P7_2_5c + P7_2_6c*
(2) Model Q2b: Model Q2a + *AGEc + ED_YEARSc + SEX_dummy + married_binary + INDIG_DUMMY*
(3) Model Q2c: Model Q2b + log_INC_ANUALc + SEX_dummy + married_binary + INDIG_DUMMY
(4) Model Q2cCat: *Model Q2c [as.factor(P7_2_1, P7_2_2, P7_2_3, P7_2_4, P7_2_5, P7_2_6)]*

	Res.Df	**RSS**	**Df**	**Sum of Sq**	**F**	**Pr(>F)**
Model 1	**37669**	13736				
Model 2	**37664**	13500	5	235.956	132.2480	< 2e-16 ***
Model 3	**37662**	13497	2	2.835	3.9729	0.01883 *
Model 4	**37664**	13433	18	64.551	10.0498	< 2e-16 ***

Another improvement is observed in **Model 4**, where satisfaction levels are treated as **categorical variables**. This adjustment resulted in a reduction in RSS and a highly significant F-statistic. This indicates that accounting for differences across categorical levels (**stepwise nature of life satisfaction)** effectively captured variations in happiness that were not addressed in the previous models. Thus, Model 4's approach to categorization helped explain disparities more clearly than the linear adjustments made in Model 3.

Based on previous findings, the following logistic regression model is refined and structured to capture the most impactful satisfaction domains and their key interactions with demographic and economic factors, as identified in previous analyses. We created the variable happy_binary, where values 1, 2, and 3 of General Happiness were coded as “Low Happiness” and values 4 and 5 were coded as “High Happiness”. Specifically, the model includes interactions between **life satisfaction (P7_2_1)** and **age (AGEc)**, **family satisfaction (P7_2_2)** and **gender (SEX_dummy),** as well as **affective life satisfaction**

(P7_2_3) and **marital status (married_binary).** These interactions are informed by the observed significance and strength of these variables in predicting happiness levels.

Table 9. Coefficients from the Logistic Regression Model of High Happiness

	Dependent variable:
	happy_binary
as.factor(P7_2_1)2	0.203(0.171)
as.factor(P7_2_1)3	0.617***(0.159)
as.factor(P7_2_1)4	2.067***(0.159)
as.factor(P7_2_1)5	2.692***(0.177)
AGEc	-0.020**(0.010)
as.factor(P7_2_2)2	0.053(0.306)
as.factor(P7_2_2)3	0.131(0.282)
as.factor(P7_2_2)4	0.920**(0.280)
as.factor(P7_2_2)5	1.477***(0.291)
SEX_dummy	0.483(0.347)
as.factor(P7_2_3)2	0.508*(0.201)
as.factor(P7_2_3)3	0.581**(0.182)
as.factor(P7_2_3)4	1.430***(0.179)
as.factor(P7_2_3)5	1.881***(0.198)
married_binary	0.399(0.314)
as.factor(P7_2_1)2:AGEc	0.010(0.011)
as.factor(P7_2_1)3:AGEc	0.006(0.010)
as.factor(P7_2_1)4:AGEc	0.0005(0.010)
as.factor(P7_2_1)5:AGEc	0.003(0.011)
as.factor(P7_2_2)2:SEX_dummy	-0.358(0.389)
as.factor(P7_2_2)3SEX_dummy	-0.415(0.355)
as.factor(P7_2_2)4SEX_dummy	-0.553(0.349)
as.factor(P7_2_2)5SEX_dummy	-0.531(0.364)
as.factor(P7_2_3)2:married_binary	-0.163(0.352)
as.factor(P7_2_3)3:married_binary	-0.264(0.321)
as.factor(P7_2_3)4:married_binary	-0.098(0.316)
as.factor(P7_2_3)5:married_binary	0.039(0.334)
Constant	-2.232***(0.331)
Observations	37,676
Log Likelihood	-12,430.780
Akaike Inf. Crit.	24,917.560
	*p<0.1; **p<0.05; ***p<0.01

NOTE: In this table, the coefficients represent the odds ratios

The coefficients from the logistic regression model presented in **Table 9** reflect the log-odds of reporting high happiness (*happy_binary*) based on different domains of satisfaction as well as demographic factors. Among the main effects, **P7_2_1** (Life Satisfaction) has the strongest impact on happiness. The log-odds significantly increase as individuals move up through the satisfaction categories. Specifically, moving from category 2 to 3 increases the log-odds of reporting high happiness by 0.617, from 3 to 4 by 2.067, and from 4 to 5 by 2.692, all of which are highly significant ($p < 0.01$). This consistent increase in log-odds indicates that greater life satisfaction is robustly associated with a higher probability of happiness.

Similarly, **P7_2_2** (Family Satisfaction) and **P7_2_3** (Affective Life Satisfaction) also exhibit strong associations. For **P7_2_2**, transitioning from category 4 to 5 increases the log odds by 1.477, indicating that improvements in family satisfaction significantly enhance happiness. In the case of **P7_2_3**, moving from category 4 to 5 raises the log odds by 1.881, highlighting the importance of affective relationships on well-being.

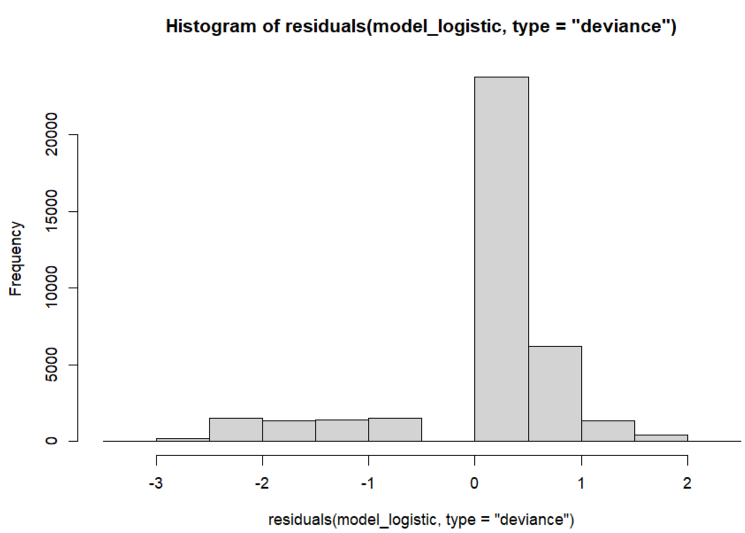
When exponentiated, the coefficients provide a more interpretable measure in terms of **odds ratios**. For instance, individuals in category 4 of **P7_2_1 (Life Satisfaction)** experience an increase in the odds of reporting high happiness by a **factor of 7.9** compared to the baseline category, while those in category 5 see an increase by a **factor of 14.76**. This result indicates a strong association between life satisfaction and reported happiness.

Similarly, for **P7_2_2 (Family Satisfaction)**, individuals at Level 4 have their odds increased by a **factor of 2.5**, and those at Level 5 experience an increase by a **factor of 4.37**. For **P7_2_3 (Affective Life Satisfaction)**, individuals at Level 4 have their odds raised by a **factor of 4.17**, while those at Level 5 see an increase by a **factor of 6.55**. These findings suggest that higher satisfaction in life, family, and affective relationships is associated with increased odds of reporting high happiness.

Regarding the **interaction terms**, most are non-significant with log-odds values close to zero and confidence intervals overlapping substantially with zero. This indicates that age, gender, and marital status do not significantly alter the impact of life satisfaction, family satisfaction, or affective life satisfaction on happiness. The lack of significance in these interactions may also be partially explained by a smaller effective sample size compared to the previous OLS Linear Regression model, especially since logistic models generally require larger samples to detect significance, particularly for interaction effects.

These results suggest that the perception of one's life quality, family dynamics, and affective satisfaction are key drivers of happiness, with the effect intensifying as satisfaction levels increase. Age, on the other hand, shows a minor adverse effect with an odds ratio of 0.98, slightly below 1, indicating a slight decline in the likelihood of high happiness with each additional year.

Figure 9. Histogram of residuals for the Logistic Model

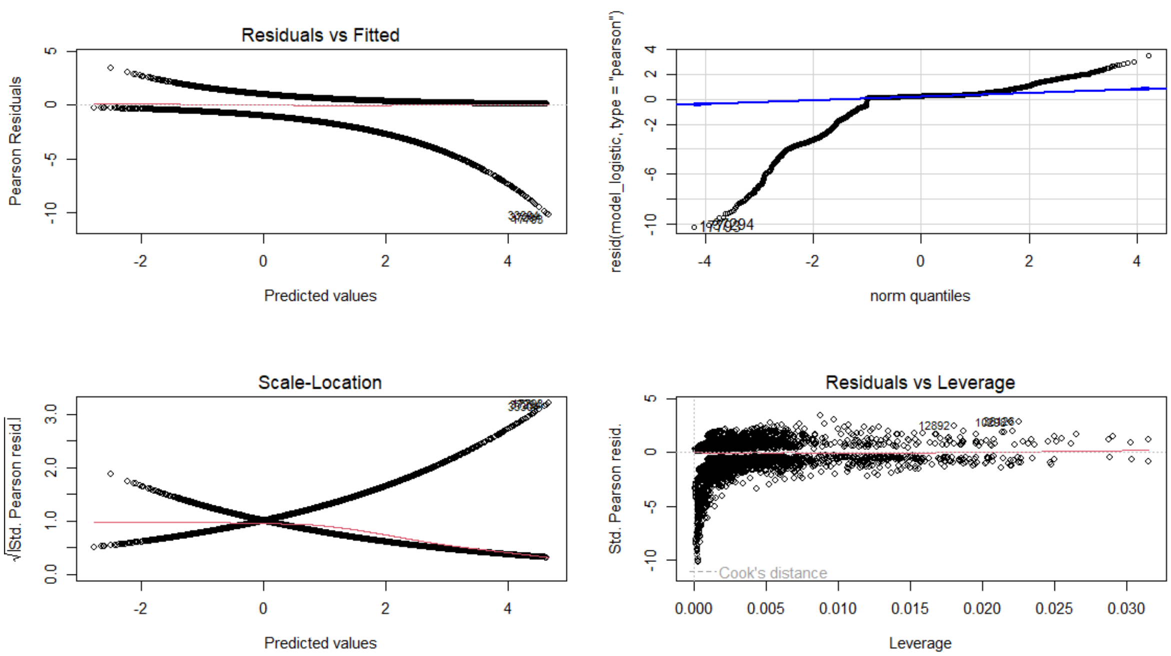


In this sense, to further explore the model, we can use a histogram of deviance (Figure 9), which indicates that the majority of residuals are centered around zero, with a noticeable peak, reflecting that many observations are well-fitted by the model.

However, the spread of residuals towards both negative and positive extremes suggests there are instances where the model either underestimates or overestimates the likelihood of reporting high happiness. The slight asymmetry could indicate potential improvements through model refinement or inclusion of interaction terms that better capture the variability. However, there are certain limits due to the number of variables available within the dataset.

Through the diagnostic plots (Figure 10), we can elaborate on the model's behavior. In the Residuals vs. Fitted plot, the characteristic two-striped pattern is visible, which is a recognized artifact of logistic regression due to the binary outcome. This is expected and does not necessarily indicate a poor fit, but rather highlights the inherent nature of logistic probabilities constrained between 0 and 1. The QQ Plot shows a substantial deviation from normality on the lower left tail, indicating that there is not enough information for the model to be entirely balanced. This is anticipated given the clustering of information around higher levels of satisfaction (see Figure 8). The Scale-Location plot displays a small joint in the middle of the data, which is a consequence of logistic regression modeling; however, as with the previous plot, having more observations would be beneficial. Finally, the Residuals vs. Leverage plot identifies no major influential points.

Figure 10. Diagnostic plots for the Logistic Model



Despite these observations, the overall structure is not unexpected, and the residual diagnostics align with typical patterns seen in logistic regression. However, this also suggests that adding more predictive variables or refining interactions could potentially smooth out some of these discrepancies.

Table 10. ANOVA Test (Null-intercept model and Logistic Model)

Model 1: *happy_binary ~ 1*
Model 2: *happy_binary ~ as.factor(P7_2_1) * AGEc + as.factor(P7_2_2) * SEX_dummy + as.factor(P7_2_3) * married_binary*

	Res.Df	Res. Dev	Df	Deviance	Pr(>Chi)
Model 1	**37675**	32769			
Model 2	**37648**	24862	27	7907.4	< 2.2e-16 ***

The ANOVA test presented in Table 10 compares the null (intercept-only) model with the complete logistic regression model; the test results show a **substantial reduction in residual deviance** from **32769** in the null model to **24862** in the full model. This change is accompanied by a highly significant Chi-square statistic (p-value < 2.2e-16), indicating that the inclusion of these predictors and their interactions contributes significantly to explaining the variance in happiness levels.

LIMITATIONS

While this analysis offers valuable insights into the relationship between life satisfaction and happiness, it is essential to acknowledge several limitations associated with both the dataset and the modeling approach.

First, the dependent variable (self-reported general happiness) is measured on an ordinal scale from 1 to 5, but was treated as continuous for some part of the regression analysis. Although this is a common practice in social science research, it introduces an approximation that assumes equal intervals between response categories. This may not fully reflect the subjective nature of happiness assessments, which is why logistic regression was used in the final part of the model.

However, this caused some problems. Specifically, logistic regression relies on the assumption that the log-odds of the outcome are linearly related to the predictors, which may not hold when the underlying measure is naturally ordinal and treated as a binary variable. Additionally, logistic regression models are sensitive to sample size; they require sufficient observations across all levels of the predictors to produce stable and meaningful estimates. Given the complexity of the model specified with multiple interaction terms, it is possible that the available sample size was not sufficient to estimate all parameters robustly. This is evident in the residual diagnostic plots, which suggest issues of nonlinearity and heteroscedasticity.

Previous interaction effects that were significant may have lacked sufficient observations to capture the variability in happiness levels adequately. Addressing these issues might require either a larger sample size, especially for less common categories, or alternative modeling strategies.

Second, the data are cross-sectional, meaning that all variables were measured at a single point in time. Consequently, the analysis is limited to associations rather than actual causal inferences. For example, while higher economic satisfaction may be associated with greater happiness, it is also possible that happier individuals are more likely to perceive their economic conditions positively.

Thirdly, measures of life satisfaction and happiness rely entirely on self-reported data, which are inherently subjective and may be influenced by temporary emotional states, the interviewer, cultural norms surrounding the reporting of well-being, or social desirability bias, as problems such as depression or sadness are often not openly acknowledged. These measurement issues can introduce unobserved variability and may affect the reliability of the estimated relationships, which could explain why many values were heavily clustered around higher levels of satisfaction or happiness.

From a modeling perspective, although the transformations and centering applied improve linearity and interpretability, the inclusion of multiple interaction terms increases model complexity, thereby raising the risk of overfitting, especially if specific interaction effects are weak or non-significant. Moreover, although the analysis included significant demographic and structural variables, other potentially relevant factors, such as mental health, personality traits, or social support, were not available in the dataset and could lead to omitted variable bias.

Lastly, the dataset is limited to Mexico, and while it offers nationally representative insights, the findings may not be generalizable to other cultural or institutional contexts. Patterns in how life satisfaction translates to happiness could differ significantly across countries due to differences in norms, expectations, and economic structures. Despite these limitations, the analysis represents a significant step toward understanding the structure of subjective well-being and the social conditions that influence it.

CONCLUSION

Research question 1 tries to explore the extent to which income influences happiness and whether this relationship is mediated or moderated by satisfaction with one's economic situation. The analysis, as presented in Model 1, suggests that while income positively contributes to general happiness, its influence is notably conditional upon one's subjective sense of economic well-being. The statistically significant interaction term reveals that the happiness boost from higher income is less substantial among those who are already economically satisfied. This finding implies that financial satisfaction can soften the necessity for additional income to enhance happiness, highlighting the psychological dimension of economic stability.

For research question 2, the results from both Model Q2c and Model Q2c Categorical provide a comprehensive view of how life satisfaction domains influence overall happiness in Mexico, even when controlling for structural variables such as age, education, gender, marital status, and indigenous status. In the continuous specification (Q2c), all six life satisfaction domains are significant predictors of happiness, with satisfaction with life, affective life, and economic situation showing the most significant effects. Notably, the categorical model (Q2c Categorical) refines this understanding by capturing distinct jumps in happiness as satisfaction increases across each category. Here, the coefficients for higher satisfaction levels (levels 4 and 5) in life satisfaction (P7_2_1) and affective life (P7_2_3) are particularly strong, suggesting that these domains are powerful determinants of well-being. The categorical approach reveals that moving up even one satisfaction category can lead to marked increases in happiness, highlighting the stepwise nature of subjective well-being.

Moreover, the inclusion of structural variables in both models adds depth to the analysis. Being married and identifying as indigenous consistently contribute slightly positively to happiness, while being male has a slight adverse effect. Economic factors, such as work hours and income, are positively linked to happiness, but with a less pronounced impact than life satisfaction domains. This suggests that while economic stability matters, the subjective experience of satisfaction in various life aspects plays a more critical role in shaping overall happiness. Collectively, these findings indicate that affective life, economic situation, and general life satisfaction are the most potent predictors of happiness, and their impacts are even more pronounced when satisfaction is measured categorically, capturing non-linear jumps in well-being.

Regarding research question 3, our logistic model suggests again that life satisfaction, affective life satisfaction, and family satisfaction are the strongest predictors of high happiness in Mexico. Similar to Model Q2c Categorical, the analysis reveals that moving up in satisfaction categories substantially increases the odds of reporting high happiness.

Individuals who report the highest satisfaction (4 and 5) among these fields are also more likely to report themselves in the highest happiness levels.

The findings from the analysis underscore that subjective well-being in key life domains, particularly life satisfaction, affective life, and family life, is a stronger predictor of happiness than mere economic gains. This highlights the need for public policies in Mexico to prioritize emotional and relational well-being alongside financial stability. Economic policies aimed at enhancing happiness should not only focus on increasing income but also on improving perceived economic security. Stronger social safety nets, financial education programs, and enhanced unemployment protections can bolster individuals' sense of financial control and security, which is crucial for amplifying the emotional benefits of income, especially for those currently experiencing economic insecurity.

Additionally, the data suggest that family and community bonds are central to happiness in Mexico. Policies that encourage family stability, relationship support, and community cohesion can significantly boost well-being. Community centers, family counseling programs, and accessible mental health services would effectively support these relational domains.

There is also a case for age-specific interventions. As happiness levels tend to decline with age, targeted programs that promote social engagement, mental health support, and community integration for older adults are essential. These initiatives would help counteract the adverse effects of aging on well-being and enhance social connectivity.

Moreover, gender-sensitive policies should be considered to address slight differences in happiness experiences. Promoting work-life balance, ensuring gender equality in the workforce, and supporting single-parent households could help close the happiness gap and improve life satisfaction for both men and women.

Ultimately, the evidence suggests that public policy should adopt a multidimensional approach to well-being. Enhancing subjective well-being in Mexico requires more than just economic growth; it demands a deeper focus on emotional stability, family cohesion, and community support. This comprehensive approach would better reflect the drivers of happiness in the Mexican context, creating more resilient and satisfied communities.

REFERENCES

- Arizpe Saide, C. M., Brito Calderón, D., Saavedra Corona, C. G., Balderas Quiroz, B., & Cantú Villarreal, R. E. (2021). *La felicidad en México y sus determinantes.* Trayectorias, 23(52), 84–107.
- Ballas, D., & Tranmer, M. (2012). *Happy people or happy places? A multilevel modeling approach to the analysis of happiness and well-being*. International Regional Science Review, 35(1), 70–102.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). *The satisfaction with life scale*. Journal of Personality Assessment, 49(1), 71–75.
- Flores Cano, O. (2023). *¿Cuál es la definición de felicidad de habitantes de la Ciudad de México?* Revista Digital Internacional de Psicología y Ciencia Social, 9(1), e912023502. https://doi.org/10.22402/j.rdipycs.unam.e.9.1.2023.502
- OECD. (2018). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing.
- Ryan, R. M., & Deci, E. L. (2001). *On happiness and human potentials: A review of research on hedonic and eudaimonic well-being*. Annual Review of Psychology, 52(1), 141–166.
- Sánchez Aragón, R., & Méndez Canales, R. (2011). *Elementos mediadores de la felicidad y el bienestar subjetivo en hombres y mujeres*. Revista Costarricense de Psicología, 30(45–46), 51–76.
- Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf
- Veenhoven, R. (1994). *Is happiness a trait?* Social Indicators Research, 32(2), 101–160. https://link.springer.com/article/10.1007/BF01078732
- World Happiness Report. (2024). *Country rankings*. Sustainable Development Solutions Network. https://data.worldhappiness.report/table
- XXIV Seminario de Economía Urbana y Regional. (2012). *Perspectiva municipal del ranking de la felicidad en México*. [Conference presentation]. http://www.economia.unam.mx/cedrus/descargas/IvanMartinezBravo.pdf

APPENDIX – R CODE

```
################################################################################
#                            APPLIED LINEAR MODELLING                          #
#                           Policy Report 4 - R Script                         #
#           Author: [Edgar B. Diaz Castro] | Date: 05/11/2025                  #
################################################################################

# ══════════════════════════════════════════════════════════════════════════════
# LOADING LIBRARIES
# ══════════════════════════════════════════════════════════════════════════════
library(tidyverse)
library(car)
library(stargazeR)

# ══════════════════════════════════════════════════════════════════════════════
# SETTING WORK DIRECTORY
# ══════════════════════════════════════════════════════════════════════════════
```

```
setwd("C:/Users/erdg9/Documents/UPennSp/MSSP 8970/PolicyReport4")

# ═══════════════════════════════════════════════════════════════════════════
# READING DATABASE
# ═══════════════════════════════════════════════════════════════════════════
df <- read.csv("ENUT_transformedF.csv")
str(df)
df <- na.omit(df)

# Units in income are in years and work hours are per week
df$WORK_HRS_YEAR <- df$WORK_HRS * 52
df <- df %>% filter(INC_ANUAL > 0, WORK_HRS_YEAR > 0) # FOR LOG TRANSFORMATION
# AND TO ONLY ACCOUNT FOR ECONOMICALLY ACTIVE PEOPLE WHO ARE WORKING.

# ═══════════════════════════════════════════════════════════════════════════
# EXPLORATORY ANALYSIS
# ═══════════════════════════════════════════════════════════════════════════

## ── Annual Income ─────────────────────────────────────────────────────────
hist(df$INC_ANUAL) # Highly right skewed, probably will need a log transformation

## ── Age ───────────────────────────────────────────────────────────────────
hist(df$AGE, breaks=100) # Slightly left skewed, probably due to sampling

## ── Sex ───────────────────────────────────────────────────────────────────
# Dummy variable where 0 means woman and 1 means man
table(df$SEX_dummy)

## ── Total Hours Worked ────────────────────────────────────────────────────
hist(df$WORK_HRS_YEAR) # Slightly right skewed, will probably need a log transformation

## ── Marital Status ───────────────────────────────────────────────────────

# Creating a binary variable: 1 = Married (formal or living with partner), 0 = Single
df$married_binary <- ifelse(df$MARIT_ST %in% c(1, 5), 1, 0)
table(df$married_binary)

## ── Educational Years ─────────────────────────────────────────────────────
hist(df$ED_YEARS) # Slightly normal distribution. Centering or scaling may help
plot(jitter(df$ED_YEARS), jitter(df$P7_3), pch=".")

## ── Indigenous Status ─────────────────────────────────────────────────────
table(df$INDIG_DUMMY)

## ── Happiness Level (Response Variable) ───────────────────────────────────
hist(df$P7_3) # Slightly right skewed

## ── Life Satisfaction Indicators ──────────────────────────────────────────
par(mfrow = c(3, 2))
hist(df$P7_2_1,
     main = "Overall Life Satisfaction",
     xlab = "Scale",
     col = "pink",
     border = "black",
     breaks=25)
hist(df$P7_2_2,
     main = "Family Life Satisfaction",
     xlab = "Scale",
     col = "pink",
     border = "black",
     breaks=25)
hist(df$P7_2_3,
     main = "Affective Life Satisfaction",
     xlab = "Scale",
     col = "pink",
     border = "black",
     breaks=25)
hist(df$P7_2_4,
     main = "Social Life Satisfaction",
     xlab = "Scale",
     col = "pink",
```

```
     border = "black",
     breaks=25)
hist(df$P7_2_5,
     main = "Economic Situation Satisfaction",
     xlab = "Scale",
     col = "pink",
     border = "black",
     breaks=25)
hist(df$P7_2_6,
     main = "Housing Conditions Satisfaction",
     xlab = "Scale",
     col = "pink",
     border = "black",
     breaks=25)
par(mfrow = c(1, 1))

# All share similar structure with slight left skew,
# except economic satisfaction which is more normally distributed

# ════════════════════════════════════════════════════════════════════════════
# SUMMARY STATISTICS
# ════════════════════════════════════════════════════════════════════════════

# Min, Max, Mean and Median for Cont Variables
continuous_subset <- df[, c("INC_ANUAL", "AGE", "WORK_HRS_YEAR", "ED_YEARS" )]
summary(continuous_subset)
# Standard Deviation for Cont Variables
sd(df$INC_ANUAL)
sd(df$AGE)
sd(df$WORK_HRS_YEAR)
sd(df$ED_YEARS)

# Categorical Variables
binary_subset <- df[, c("married_binary", "INDIG_DUMMY", "SEX_dummy")]

categorical_subset <- df[, c("P7_3", "P7_2_1", "P7_2_2", "P7_2_3", "P7_2_4",
                             "P7_2_5", "P7_2_6")]
# Marital Status
table(df$married_binary)
length(df$married_binary)
# Indigeneous status
table(df$INDIG_DUMMY)
length(df$INDIG_DUMMY)
# Sex
table(df$SEX_dummy)
length(df$SEX_dummy)

# Categorical
table(df$P7_3)
table(df$P7_2_1)
table(df$P7_2_2)
table(df$P7_2_3)
table(df$P7_2_4)
table(df$P7_2_5)
table(df$P7_2_6)

## ── Response Variable ───────────────────────────────────────────────────────
hist(df$P7_3,
     main = "General Happiness",
     xlab = "Scale (1 to 5 - 5 is higher)",
     col = "lightblue",
     border = "black",
     breaks=15)

## ── Continuous Variables ────────────────────────────────────────────────────
par(mfrow = c(2, 2))
hist(df$INC_ANUAL,
     main = "Income Distribution",
     xlab = "Annual Income (MXN)",
     col = "lightgray",
     border = "black",
```

```
     breaks=35)
hist(df$AGE,
     main = "Age Distribution",
     xlab = "Age (Years)",
     col = "lightgray",
     border = "black",
     breaks=50)
hist(df$WORK_HRS_YEAR,
     main = "Distribution of Hours of Work",
     xlab = "Hours of Work",
     col = "lightgray",
     border = "black",
     breaks=50)
hist(df$ED_YEARS,
     main = "Distribution of Years of Education",
     xlab = "Years of Education",
     col = "lightgray",
     border = "black",
     breaks=25)
par(mfrow = c(1, 1))

# ═══════════════════════════════════════════════════════════════════════════
# VARIABLE TRANSFORMATIONS
# ═══════════════════════════════════════════════════════════════════════════

# Log Transformations for highly left skewed variables
# INCOME
df$log_INC_ANUAL <- log(df$INC_ANUAL)
# Checking
hist(df$log_INC_ANUAL)
df$log_INC_ANUALc <- df$log_INC_ANUAL - mean(df$log_INC_ANUAL, na.rm = TRUE)
hist(df$log_INC_ANUALc)

# Manual centering (mean subtraction)
# AGE
df$AGEc <- df$AGE - mean(df$AGE, na.rm = TRUE)
hist(df$AGEc)

# EDUCATION
df$ED_YEARSc <- df$ED_YEARS - mean(df$ED_YEARS, na.rm = TRUE)
hist(df$ED_YEARSc)

# WORK HOURS PER YEAR
df$WORK_HRS_YEARc <- df$WORK_HRS_YEAR - mean(df$WORK_HRS_YEAR, na.rm = TRUE)
hist(df$WORK_HRS_YEARc)

# Centering satisfaction domains
df$P7_2_1c <- df$P7_2_1 - mean(df$P7_2_1, na.rm = TRUE)
df$P7_2_2c <- df$P7_2_2 - mean(df$P7_2_2, na.rm = TRUE)
df$P7_2_3c <- df$P7_2_3 - mean(df$P7_2_3, na.rm = TRUE)
df$P7_2_4c <- df$P7_2_4 - mean(df$P7_2_4, na.rm = TRUE)
df$P7_2_5c <- df$P7_2_5 - mean(df$P7_2_5, na.rm = TRUE)
df$P7_2_6c <- df$P7_2_6 - mean(df$P7_2_6, na.rm = TRUE)

# CORRELATIONS (pairwise)
cor_matrix <- cor(df[,c("P7_3", "log_INC_ANUALc", "AGEc", "WORK_HRS_YEARc", "ED_YEARSc",
                        "P7_2_1c", "P7_2_2c", "P7_2_3c", "P7_2_4c",
                        "P7_2_5c", "P7_2_6c")],
                  use = "pairwise.complete.obs")
print(cor_matrix)
write.csv(cor_matrix, file="correlation.csv")

# ═══════════════════════════════════════════════════════════════════════════
# PLOTS
# ═══════════════════════════════════════════════════════════════════════════

# Bar plots of categorical variables
plot(table(df$P7_3), main = "Distribution of Happiness", xlab = "Happiness Score", ylab =
"Frequency", lend = 1)
plot(table(df$married_binary), main = "Marital Status", xlab = "Married (1) / Single (0)",
ylab = "Frequency", lend = 1)
```

```
# INCOME
scatterplot(x=df$log_INC_ANUALc, y=df$P7_3)
# HOURS
scatterplot(x=df$WORK_HRS_YEARc, y=df$P7_3)
# SATISFACTION LEVELS 1 TO 6
par(mfrow = c(2, 3))
scatterplot(P7_3 ~ P7_2_1, data = df, jitter = list(x = 1, y = 1))
scatterplot(P7_3 ~ P7_2_2, data = df, jitter = list(x = 1, y = 1))
scatterplot(P7_3 ~ P7_2_3, data = df, jitter = list(x = 1, y = 1))
scatterplot(P7_3 ~ P7_2_4, data = df, jitter = list(x = 1, y = 1))
scatterplot(P7_3 ~ P7_2_5, data = df, jitter = list(x = 1, y = 1))
scatterplot(P7_3 ~ P7_2_6, data = df, jitter = list(x = 1, y = 1))
par(mfrow = c(1, 1)

# ════════════════════════════════════════════════════════════════════════════
# MODELS
# ════════════════════════════════════════════════════════════════════════════

modelQ1 <- lm(P7_3 ~ log_INC_ANUALc + P7_2_5c + log_INC_ANUALc * P7_2_5c, data=df)
summary(modelQ1)
par(mfrow = c(2, 2))
plot(modelQ1, which=1)
plot(modelQ1, which=2)
plot(modelQ1, which=3)
plot(modelQ1, which=5)
par(mfrow = c(1, 1))
vif(modelQ1)

modelQ2a <- lm(P7_3 ~ P7_2_1c + P7_2_2c + P7_2_3c +
                 P7_2_4c + P7_2_5c + P7_2_6c, data=df)
summary(modelQ2a)
plot(modelQ2a, which=1)
plot(modelQ2a, which=2)
plot(modelQ2a, which=3)
plot(modelQ2a, which=5)
vif(modelQ2a)

modelQ2b <- lm(P7_3 ~ P7_2_1c + P7_2_2c + P7_2_3c +
                 P7_2_4c + P7_2_5c + P7_2_6c +
                 AGEc + ED_YEARSc + SEX_dummy + married_binary +
                 INDIG_DUMMY, data=df)
summary(modelQ2b)
plot(modelQ2b, which=1)
plot(modelQ2b, which=2)
plot(modelQ2b, which=3)
plot(modelQ2b, which=5)
vif(modelQ2b)

modelQ2c <- lm(P7_3 ~ P7_2_1c + P7_2_2c + P7_2_3c +
                 P7_2_4c + P7_2_5c + P7_2_6c +
                 AGEc + ED_YEARSc + WORK_HRS_YEARc + log_INC_ANUALc +
                 SEX_dummy + married_binary +
                 INDIG_DUMMY, data=df)
summary(modelQ2c)
plot(modelQ2c, which=1)
plot(modelQ2c, which=2)
plot(modelQ2c, which=3)
plot(modelQ2c, which=5)
vif(modelQ2c)

# ════════════════════════════════════════════════════════════════════════════
# RELATIONSHIP OF VARIABLES
# ════════════════════════════════════════════════════════════════════════════

# List of life satisfaction variables
satisfaction_vars <- c("P7_2_1", "P7_2_2", "P7_2_3", "P7_2_4", "P7_2_5", "P7_2_6")

# Plot each as a boxplot
par(mfrow = c(2, 3))
for (var in satisfaction_vars) {
```

```
  boxplot(P7_3 ~ df[[var]], data = df,
          main = paste("Happiness vs", var),
          xlab = var, ylab = "General Happiness (P7_3)",
          col = "lightblue")}
par(mfrow = c(1, 1))

# ══════════════════════════════════════════════════════════════════════════════
# NEW MODELS
# ══════════════════════════════════════════════════════════════════════════════

# Categorical for Model 2c
modelQ2cCat <- lm(P7_3 ~ as.factor(P7_2_1) + as.factor(P7_2_2) +
                    as.factor(P7_2_3) + as.factor(P7_2_4) +
                    as.factor(P7_2_5) + as.factor(P7_2_6) +
                    AGEc + ED_YEARSc + SEX_dummy +
                    married_binary + INDIG_DUMMY + WORK_HRS_YEARc +
                    log_INC_ANUALc,
                  data = df)
summary(modelQ2cCat)
par(mfrow = c(2, 2))
plot(modelQ2cCat, which=1)
plot(modelQ2cCat, which=2)
plot(modelQ2cCat, which=3)
plot(modelQ2cCat, which=5)
par(mfrow = c(1, 1))

# Comparing
anova(modelQ2a, modelQ2b, modelQ2c, modelQ2cCat)

# Binary Logistic Model, creating the binary variable
# 4 or 5 equals to High Happiness and 1, 2 or equals to Low Happiness
df$happy_binary <- ifelse(df$P7_3 >= 4, 1, 0)

# Logistic Regression Model
model_logistic <- glm(happy_binary ~ as.factor(P7_2_1) * AGEc +
                            as.factor(P7_2_2) * SEX_dummy +
                            as.factor(P7_2_3) * married_binary,
                          family = binomial(link = "logit"),
                          data = df)

# Interpreting as odds ratios
exp(coef(model_logistic))

# Model summary
summary(model_logistic)
par(mfrow = c(2, 2))
plot(model_logistic, which=1)
qqPlot(resid(model_logistic, type="pearson"))
plot(model_logistic, which=3)
plot(model_logistic, which=5)
par(mfrow = c(1, 1))

model_null <- glm(happy_binary ~ 1, family = binomial(link = "logit"), data = df)

# Deviance of Residuals
hist(residuals(model_logistic, type=""))
anova(model_null, model_logistic,test ="Chisq")

# ══════════════════════════════════════════════════════════════════════════════
# EXPORTING TABLES
# ══════════════════════════════════════════════════════════════════════════════

# MODEL 1
stargazer(modelQ1,type="text")
# MODEL 2abc
stargazer(modelQ2a,type="text")
stargazer(modelQ2b,type="text")
stargazer(modelQ2c,type="text")
stargazer(modelQ2cCat, type="text")
# MODEL 3
stargazer(model_logistic, type="text")
```