QMSS 201 Final Project

Alyssa Yang

Introduction

As per the Oxford Dictionary, addiction is defined as “the condition of being unable to stop using or doing something as a habit, especially something harmful”. It affects a vast amount of people worldwide and can have detrimental effects on one’s physical and mental health and overall quality of life. Breaking out of addiction involves a multitude of psychological, physiological, and social factors including seeking treatment, lifestyle changes, and building a strong support system to help aid them through changing these harmful patterns.

This report aims to investigate some issues around alcohol and cigarette addiction. Our research questions are:

- How does addiction vary by different demographics?
- Does the willingness to stop addiction help in stopping it?

The data utilized for this project is sourced from the 2016 National Survey on Drug Use and Health (NSDUH) collected by the Substance Abuse and Mental Health Service Administration (SAMHSA). It contains responses to questions regarding respondents’ demographic information and addiction habits.

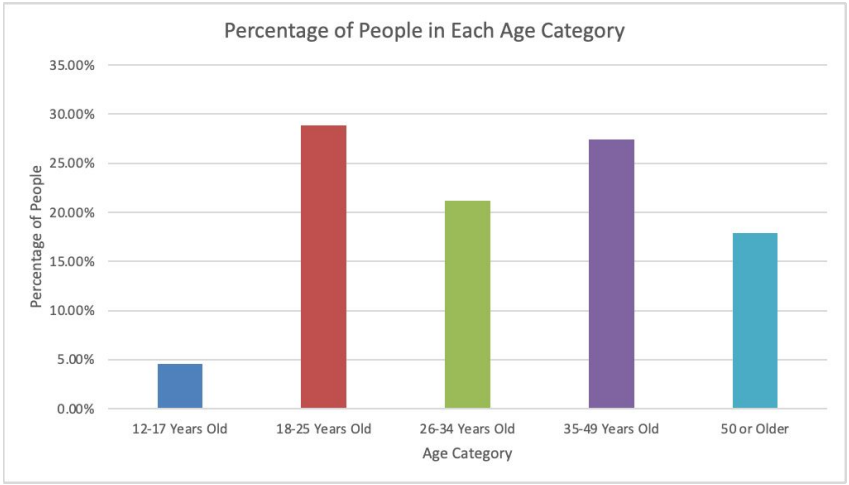
Univariate Descriptives

Demographic Variables

Demographic - Age Categories

Frequency and Proportions of Age Categories

Age Category	Number of People	Percentage of People
12-17 Years Old	962	4.57%
18-25 Years Old	6075	28.88%
26-34 Years Old	4459	21.20%
35-49 Years Old	5771	27.43%
50 or Older	3769	17.92%
Grand Total	**21036**	**100.00%**

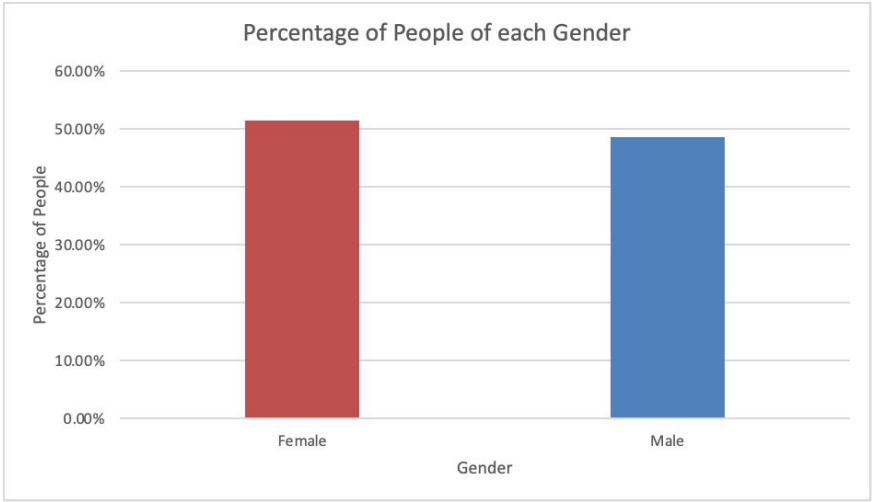


We can see from the table and chart above that the age category with the most number of people is 18-25 years old making up 28.88% of the total number of people in the study, followed by 35-49 years old, 26-34 years old, 50 or older, then 12-17 years old with the least, making up only 4.57%.

Demographic - Gender

Frequency and Proportions of Genders

Gender	Number of People	Percentage of People
Female	10811	51.39%
Male	10225	48.61%
Grand Total	**21036**	**100.00%**

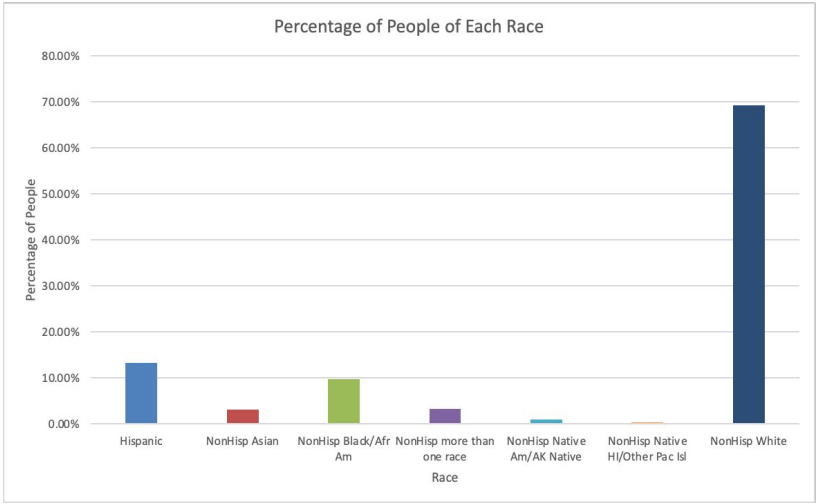


We can see from the table and chart above that the proportion of males and females in the study is almost the same, but there are a few more females than males. There are 10811 females, making up 51.39% of the total number of people in the study, and there are 10225 males, making up 48.61% of the total number of people in the study.

Demographic - Race

Frequency and Proportions of Races

Race	Number of People	Percentage of People
Hispanic	2803	13.32%
NonHisp Asian	652	3.10%
NonHisp Black/Afr Am	2036	9.68%
NonHisp more than one race	688	3.27%
NonHisp Native Am/AK Native	203	0.97%
NonHisp Native HI/Other Pac Isl	82	0.39%
NonHisp White	14572	69.27%
Grand Total	**21036**	**100.00%**

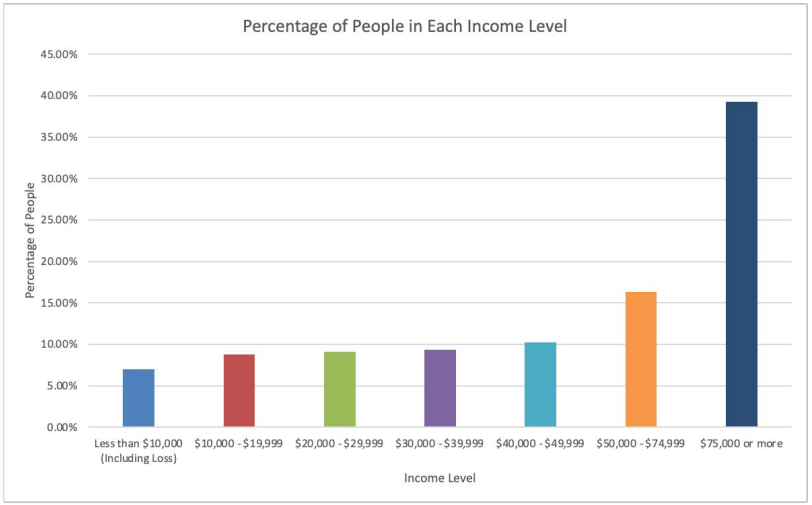


We can see from the table and chart above that a majority of the people in the study are White, making up 69.27%. The race with the second most amount of people is Hispanic, then African American, more than one race, Asian, then Native American / Alaskan Native and Native Hawaiian / Other Pacific Islander with each only making up less than 1% of the total number of people in the study.

Demographic - Income

Frequency and Proportions of Income Levels

Income Level	Number of People	Percentage of People
Less than $10,000 (Including Loss)	1478	7.03%
$10,000 - $19,999	1848	8.78%
$20,000 - $29,999	1914	9.10%
$30,000 - $39,999	1964	9.34%
$40,000 - $49,999	2150	10.22%
$50,000 - $74,999	3428	16.30%
$75,000 or more	8254	39.24%
Grand Total	**21036**	**100.00%**

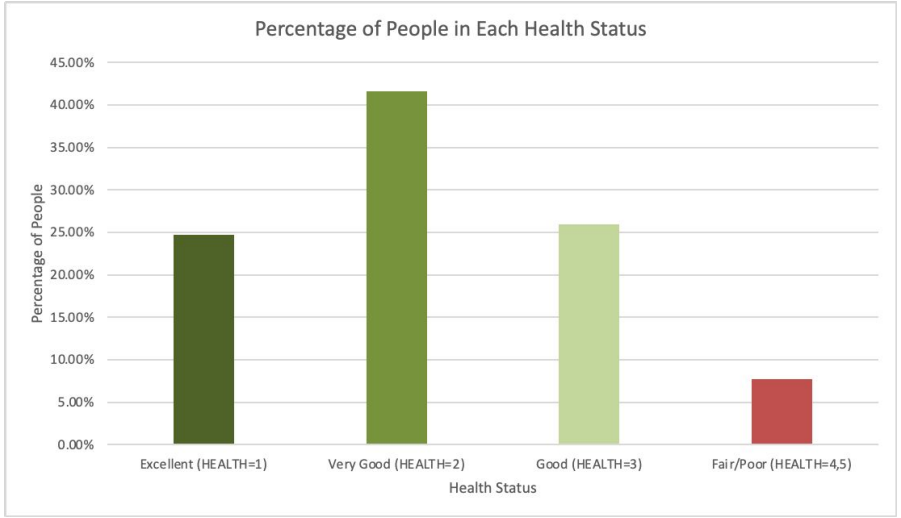


We can see from the table and chart above that the distribution of the percentage of people in the study in each income level is left-skewed, meaning there are more people in each income level as the the income level increases. Almost 40% of the people in the study have an income of $75,000 or more, whereas only 7.03% of people have an income of less than $10,000.

Demographic - Health Status

Frequency and Proportions of Health Statuses

Health Status	Number of People	Percentage of People
Excellent (HEALTH=1)	5196	24.70%
Very Good (HEALTH=2)	8759	41.64%
Good (HEALTH=3)	5464	25.98%
Fair/Poor (HEALTH=4,5)	1614	7.67%
Grand Total	**21033**	**100.00%**



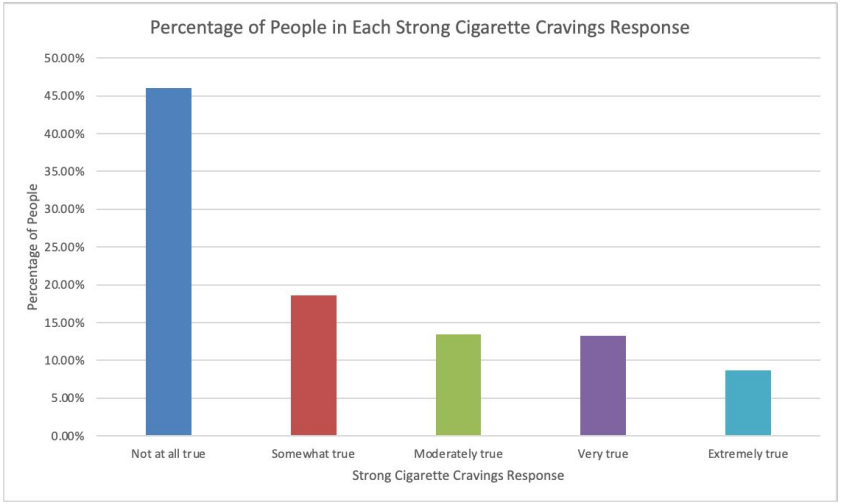
We can see from the table and chart above that the distribution of the percentage of people in the study in each health status is slightly right-skewed with more people having better health than worse health. The health status with the highest percentage of people is very good at around 41.64%, whereas excellent and good health have around the same percentage of people at about 25%, and only 7.67% of people in the study have fair/poor health.

Addiction Variables

Addiction – Cigarette Cravings

Frequency and Proportions of Cigarette Cravings

Strong Cigarette Cravings	Number of People	Percentage of People
Not at all true	2499	45.99%
Somewhat true	1013	18.64%
Moderately true	731	13.45%
Very true	721	13.27%
Extremely true	470	8.65%
Grand Total	**5434**	**100.00%**

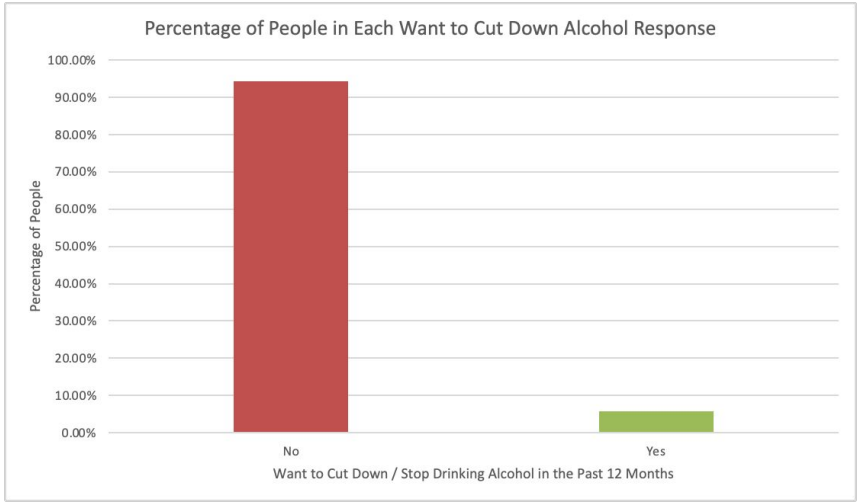


We can see from the table and chart above that the distribution of the percentage of people in each strong cigarette cravings response is right-skewed with more people not having strong cravings for a cigarette where it feels like they're in the grip of a force they cannot control. 45.99% of the people in the study said this statement was not at all true whereas only 8.65% of the respondents said it was extremely true.

Addiction – Tried Cutting Down Alcohol

Frequency and Proportions of People who Tried Cutting Down Alcohol

Want to Cut Down Alcohol	Number of People	Percentage of People
No	19797	94.33%
Yes	1191	5.67%
Grand Total	**20988**	**100.00%**

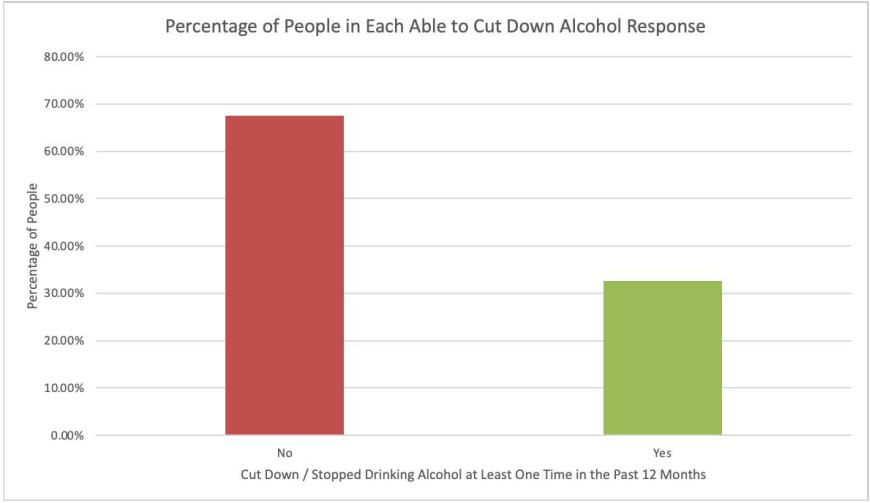


We can see from the table and chart above that the number of people in the study that have wanted to try to cut down or stop drinking alcohol is much lower than the number of people who haven't. 94.33% of the respondents responded no to this question while only 5.67% of the respondents responded with yes, thus the distribution of the responses to this question are heavily skewed in favor of the negative.

Addiction – Able to Cut Down Alcohol

Frequency and Proportions of People who were Able to Cut Down Alcohol

Able to Cut Down Alcohol	Number of People	Percentage of People
No	14190	67.46%
Yes	6846	32.54%
Grand Total	**21036**	**100.00%**

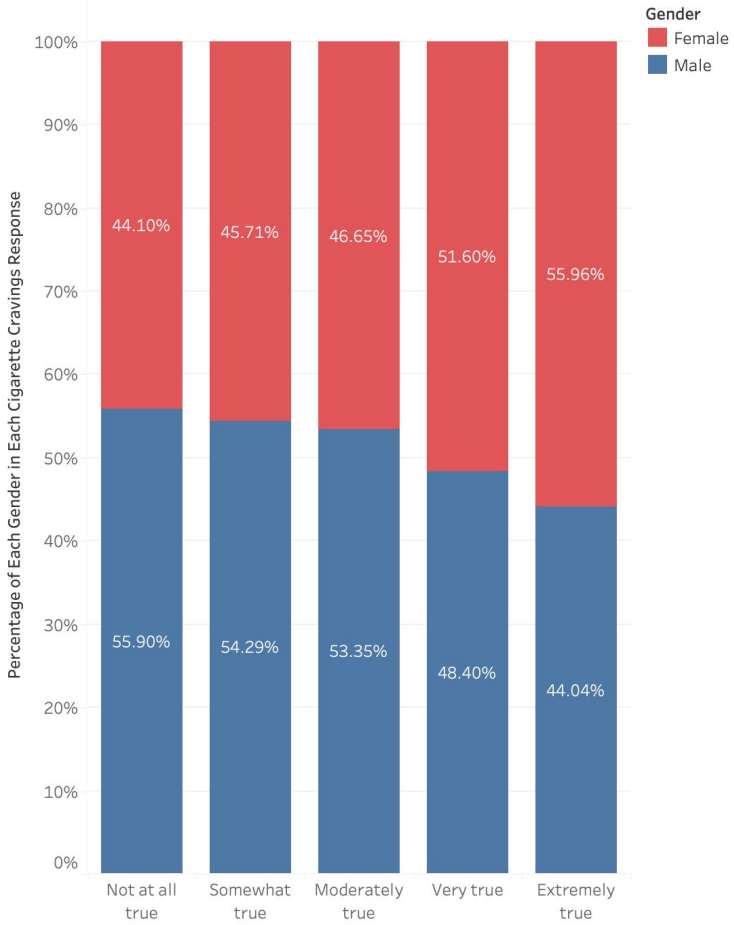


We can see that the number of people in the study that was able to cut down or stop drinking at least one time in the past 12 months is much lower than the number of people that weren't able to. 67.46% of the respondents weren't able to whereas only 32.54% of the respondents were able to, thus the distribution of the responses to this question is skewed in favor of the negative.

Bivariate Descriptives

How Cigarette Cravings Varies by Gender

How Cigarette Cravings Varies by Gender



% of Total Count of Irsex for each Cigcragp. Color shows details about Irsex. The marks are labeled by % of Total Count of Irsex.

Frequency of Cigarette Cravings by Gender

Strong Cigarette Cravings	Number of Females	Number of Males
Not at all true	1102	1397
Somewhat true	462	550
Moderately true	341	390
Very true	372	349
Extremely true	263	207

Chi-squared test

χ^2	30.359
df	4
p-value	0.0000

How Cigarette Cravings Varies by Gender

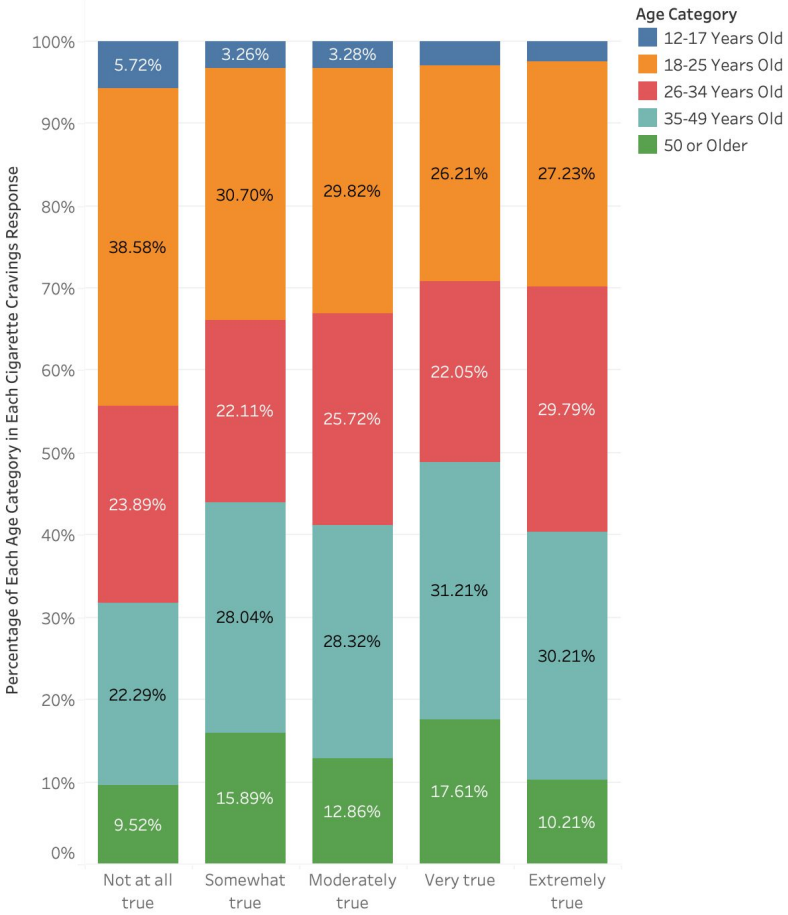
From the chart, we can see that as the responses lean more towards not at all true, there is a slightly higher proportion of male respondents compared to female respondents. As the responses lean more towards extremely true, there is a higher proportion of female respondents. This same trend can be seen through the frequency table as well.

From the chi-squared test, we obtained a □2 value of 30.359 with 4 degrees of freedom, and a p-value of 0.0000 which is less than α = 0.05. Thus, our results are statistically significant, and we reject the null hypothesis that there is no association between having strong cigarette cravings and gender.

Through our analysis, we can conclude that females feel more cravings for cigarettes. After doing some research online, this does go along with my expectations as females may feel more cravings due to hormone fluctuations and ability to manage stress and mood.

How Cigarette Cravings Varies by Age Category

How Cigarette Cravings Varies by Age



% of Total Count of Catag3 for each Cigcragp. Color shows details about Catag3. The marks are labeled by % of Total Count of Catag3.

Frequency of Cigarette Cravings by Age (in years old)

Strong Cigarette Cravings	# of 12-17	# of 18-25	# of 26-34	# of 35-49	# of 50+
Not at all true	143	964	597	557	238
Somewhat true	33	311	224	283	161
Moderately true	24	218	188	207	94
Very true	21	189	159	225	127
Extremely true	12	128	140	142	48

Chi-squared test

□2	146.05
df	16
p-value	0.0000

How Cigarette Cravings Varies by Age Category

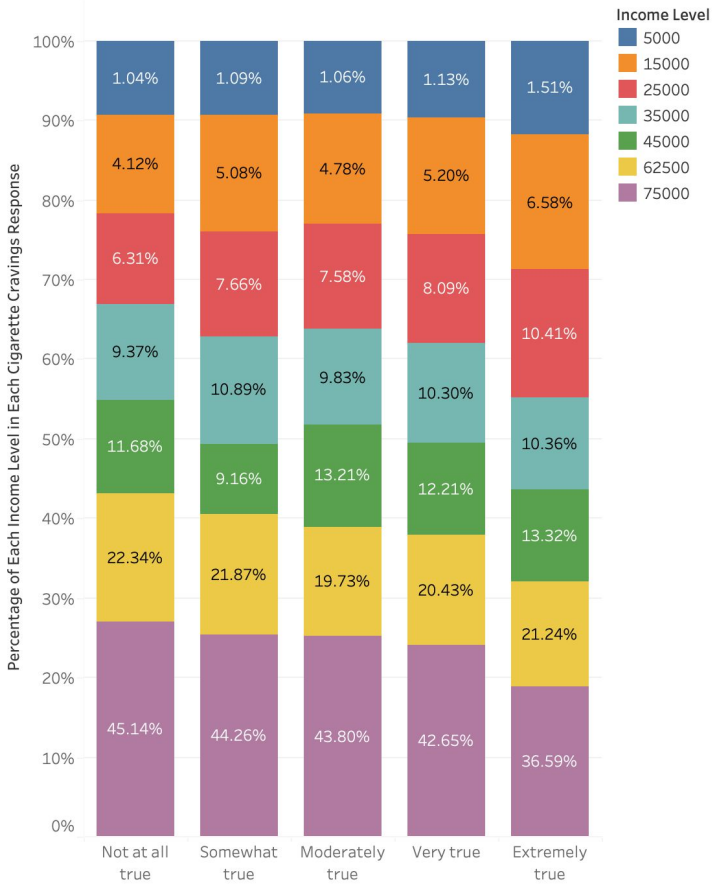
From the chart, we can see that as the responses lean more towards not at all true, there is a slightly higher proportion of younger respondents (<25 years old) compared to older respondents (>25 years old). For responses “not at all true”, “somewhat true”, and “moderately true”, 18-25 year olds constitute the highest proportion of respondents. For responses “very true“ and “extremely true”, 35-49 year olds constitute the highest proportion of respondents. This same trend can be seen through the frequency table as well.

From the chi-squared test, we obtained a □2 value of 146.05 with 16 degrees of freedom, and a p-value of 0.0000 which is less than α = 0.05. Thus, our results are statistically significant, and we reject the null hypothesis that there is no association between having strong cigarette cravings and age.

Through our analysis, we can conclude that the variation of cigarette cravings across the different age categories is statistically significant, and 35-49 year olds crave the most for cigarettes. Something interesting I see about this is that craving intensity increases among older people which may suggest that there may be generational or long-term habitual factors that can influence cravings.

How Cigarette Cravings Varies by Income

How Cigarette Cravings Varies by Income



% of Total Count of Inc Numeric for each Cigcragp. Color shows details about Inc Numeric. The marks are labeled by % of Total Sum of Inc Numeric. The view is filtered on Cigcragp, which keeps Extremely true, Moderately true, Not at all true, Somewhat true and Very true.

Summary Statistics of Cigarette Cravings by Income (in dollars)

Strong Cigarette Cravings	Mean income	Median Income	Min Income	Max Income	SD Income
Not at all true	45010.00	45000	5000	75000	24737.31
Somewhat true	43199.11	35000	5000	75000	24894.92
Moderately true	43331.05	45000	5000	75000	24564.39
Very true	42434.12	35000	5000	75000	24632.81
Extremely true	38819.15	35000	5000	75000	24299.66

ANOVA test

	Df	Sum Sq (in billions)	Mean Sq (in millions)	F values	Pr(>F)
cigcragp	4	16.9	4225	6.93	0.0000
Residuals	5428	3309	609.7		

How Cigarette Cravings Varies by Income

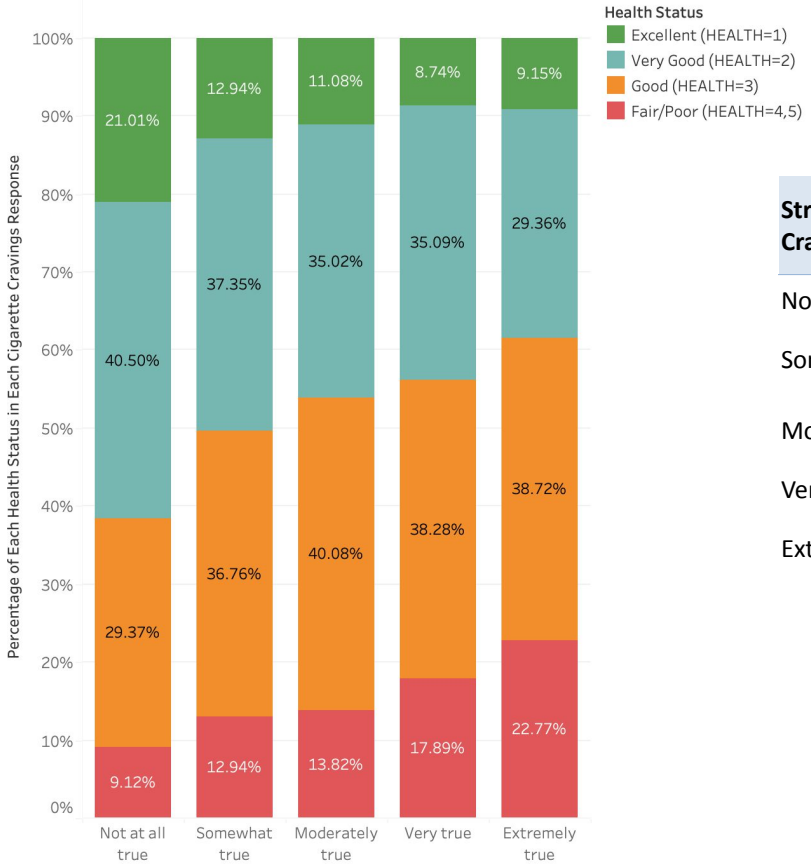
From the chart, we can see that as the responses lean more towards not at all true, there is a slightly higher proportion of higher income respondents compared to lower income respondents. As the responses lean more towards extremely true, there is a higher proportion of lower income respondents. This same trend can be seen through the summary statistics table as well - as the responses lean more towards not at all true, the mean income is higher and vice versa.

From the ANOVA test, we obtained an F-value of 6.93 with 4 degrees of freedom, and a p-value of 0.0000 which is less than α = 0.05. Thus, our results are statistically significant, and we reject the null hypothesis that there is no difference among the mean income among the different cigarette cravings responses.

Through our analysis, we can conclude that people with lower incomes feel more cravings for cigarettes. The average income of those that crave more for cigarettes (responded with “extremely true”) is $38,819.15 - this is lower than the average income of those who do not crave for cigarettes at all (responded with “not at all true”) which is $45,010.00. An interesting thing about this is that socioeconomic factors such as income have a significant role in cravings for cigarettes, maybe due to the perception of smoking as a stress or coping mechanism or targeted marketing of cigarettes for vulnerable populations such as low income individuals.

How Cigarette Cravings Varies by Health Status

How Cigarette Cravings Varies by Health Status



% of Total Count of Health2 for each Cigcragp. Color shows details about Health2. The marks are labeled by % of Total Count of Health2. The view is filtered on Exclusions (Cigcragp,Health2) and Cigcragp. The Exclusions (Cigcragp,Health2) filter keeps 24 members. The Cigcragp filter keeps Extremely true, Moderately true, Not at all true, Somewhat true and Very true.

Frequency of Cigarette Cravings by Health Status

Strong Cigarette Cravings	# of Excellent	# of Very Good	# of Good	# of Fair/Poor
Not at all true	525	1012	228	734
Somewhat true	131	278	372	131
Moderately true	81	256	293	101
Very true	63	253	276	129
Extremely true	43	138	182	107

Chi-squared test

$\square^2$	223.11
df	12
p-value	0.0000

How Cigarette Cravings Varies by Health Status

From the chart, we can see that as the responses lean more towards not at all true, there is a higher proportion of respondents with excellent and great health compared to respondents with only good or fair/poor health. As the responses lean more towards extremely true, there is a higher proportion of respondents with only good or fair/poor health. This same trend can be seen through the frequency table as well.

From the chi-squared test, we obtained a □2 value of 223.11 with 12 degrees of freedom, and a p-value of 0.0000 which is less than α = 0.05. Thus, our results are statistically significant, and we reject the null hypothesis that there is no association between having strong cigarette cravings and health status.

Through our analysis, we can conclude that unhealthy people crave more for cigarettes than healthy people. This does align with my expectations because individuals who have poorer health may be more likely to start smoking as a coping mechanism or as a means of self-medication. As well, there are physiological effects of smoking that can contribute to this poorer health and lead to a habit of smoking which can make it challenging for unhealthy people to quit smoking.

Logistic Regression

Logistic Regression Coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.2698	0.0516	-5.231	0.0000***
alccutdnYes	2.948	0.0923	31.936	<0.0000***
irsexMale	0.0985	0.0312	3.203	0.0014**
NEWRACE2NonHisp Asian	-0.3340	0.0968	-3.451	0.0006***
NEWRACE2NonHisp Black/Afr Am	-0.0741	0.0621	-1.193	0.2329
NEWRACE2NonHisp more than one race	-0.1896	0.0921	-2.060	0.0394*
NEWRACE2NonHisp Native Am/AK Native	-0.1040	0.1577	-0.660	0.5095
NEWRACE2NonHisp Native HI/Other Pac Isl	0.1741	0.2362	0.737	0.4611
NEWRACE2NonHisp White	-0.4715	0.0447	-10.538	<0.0000***
inc_numeric	-0.0000	0.0000	-10.118	<0.0000***

Chi-squared test

Chi-squared	d.f.	p
2048.142	9	<0.001

Odds Ratios

	Estimate	Std. Error	z value
(Intercept)	0.7635067	0.6899620	0.8446103
alccutdnYes	19.0696483	15.9733171	22.9441185
irsexMale	1.1050039	1.0395042	1.1746419
NEWRACE2NonHisp Asian	0.7160624	0.5913662	0.8643570
NEWRACE2NonHisp Black/Afr Am	0.9286199	0.8221311	1.0486670
NEWRACE2NonHisp more than one race	0.8272550	0.6899218	0.9898746
NEWRACE2NonHisp Native Am/AK Native	0.9012209	0.6591613	1.2240797
NEWRACE2NonHisp Native HI/Other Pac Isl	1.1901285	0.7435373	1.8836178
NEWRACE2NonHisp White	0.6240952	0.5717982	0.6814123
inc_numeric	0.9999935	0.9999923	0.9999948

Predicting ability to cut down alcohol consumption as a function of willingness, gender, race, and income

For the tables in the slide above, I ran logistic regression analysis to predict the ability to cut down or stop alcohol consumption as a function of willingness to cut down or stop alcohol consumption, gender, race, and income.

- The estimated coefficient for willingness to cut down or stop alcohol consumption is 2.948, indicating a positive relationship with the log odds of the ability to cut down or stop alcohol consumption. The p-value for this coefficient is <0.0000 which is less than α = 0.05, thus our result is statistically significant. Holding gender, race, and income constant, being willing to rather than not willing to cut down or stop alcohol consumption will result in an increase in the log odds of having the ability to cut down or stop alcohol consumption.
- The estimated coefficient for males is 0.0985, indicating a positive relationship with the log odds of the ability to cut down or stop alcohol consumption. The p-value for this coefficient is 0.0014 which is less than α = 0.05, thus our result is statistically significant. Holding willingness to cut down or stop alcohol consumption, race, and income constant, being a male rather than a female will result in an increase in the log odds of having the ability to cut down or stop alcohol consumption.

Predicting ability to cut down alcohol consumption as a function of willingness, gender, race, and income

- The estimated coefficient for Asians is -0.3340, indicating an inverse relationship with the log odds of the ability to cut down or stop alcohol consumption. The p-value for this coefficient is 0.0006 which is less than α = 0.05, thus our result is statistically significant. Holding willingness to cut down or stop alcohol consumption, gender, and income constant, being Asian rather than Hispanic will result in a decrease in the log odds of having the ability to cut down or stop alcohol consumption.
- The estimated coefficient for Non-Hispanic individuals who are more than one race is -0.1896, indicating an inverse relationship with the log odds of the ability to cut down or stop alcohol consumption. The p-value for this coefficient is 0.0394 which is less than α = 0.05, thus our result is statistically significant. Holding willingness to cut down or stop alcohol consumption, gender, and income constant, being Non-Hispanic more than one race rather than Hispanic will result in a decrease in the log odds of having the ability to cut down or stop alcohol consumption.
- The estimated coefficient for White individuals is -0.4715, indicating an inverse relationship with the log odds of the ability to cut down or stop alcohol consumption. The p-value for this coefficient is <0.0000 which is less than α = 0.05, thus our result is statistically significant. Holding willingness to cut down or stop alcohol consumption, gender, and income constant, being White rather than Hispanic will result in a decrease in the log odds of having the ability to cut down or stop alcohol consumption.

Predicting ability to cut down alcohol consumption as a function of willingness, gender, race, and income

- The estimated coefficients for African Americans, Native Americans / Alaska Natives, and Native Hawaiian / Other Pacific Islanders are -0.0741, -0.1040, and 0.1741 respectively, indicating inverse relationships and a positive relationship with the log odds of the ability to cut down or stop alcohol consumption. However, the p-values for these coefficients are 0.2329, 0.5095, and 0.4611 respectively which are all greater than α = 0.05, thus our results are not statistically significant. We fail to reject the null hypothesis that holding willingness to cut down or stop alcohol consumption, gender, and income constant, being these races rather than Hispanic will not result in a change in the log odds of having the ability to cut down or stop alcohol consumption.
- The estimated coefficient for income is -0.0000, indicating no relationship with the log odds of the ability to cut down or stop alcohol consumption. The p-value for this coefficient is <0.0000 which is less than α = 0.05, thus our result is statistically significant. Holding willingness to cut down or stop alcohol consumption, gender, and race constant, changing income level will not result in a change in the log odds of having the ability to cut down or stop alcohol consumption.

From our chi-squared test, we obtained a χ^2 value of 2048.142 with 9 degrees of freedom and a p-value of <0.001 which is less than α = 0.05, thus the model is statistically significant overall and provides a good fit for the data.

Predicting ability to cut down alcohol consumption as a function of willingness, gender, race, and income

The odds-ratio of the ability to cut down or stop alcohol consumption for . . .

- Those who are willing to is 19.0696 times higher than the odds-ratio for those who are not willing to.
- Males is 1.1050 times higher than the odds-ratio for females.
- Asian people is 0.7161 times higher than the odds-ratio for Hispanic people.
- Black / African American people is 0.9286 times higher than the odds-ratio for Hispanic people.
- People who are more than one race is 0.8273 times than the odds-ratio for Hispanic people.
- Native American / Alaskan Native people is 0.9012 times than the odds-ratio for Hispanic people.
- Native Hawaiian / Other Pacific Islander people is 0.1901 times than the odds-ratio for Hispanic people.
- White people is 0.6241 times than the odds-ratio for Hispanic people.

For every increase in $1 in income, the odds-ratio of the ability to cut down or stop alcohol consumption increases by about 0.9999935.

Key Findings

How does addiction vary by different Demographics?

From the bivariate hypothesis tests:

- Females feel more cravings for cigarettes than males
- Older people feel more cravings for cigarettes than younger people
- Lower income people feel more cravings for cigarettes than higher income people
- Unhealthy people feel more cravings for cigarettes than healthier people

Does the willingness to stop addiction help in stopping it?

From the logistic regression analysis:

- Willingness to stop addiction does help in stopping it
- Another factor that increases the ability to stop addiction is being male
- Being Asian or White decreases the ability to stop addiction
- Income level has no relationship with the ability to stop addiction