Exercise_8.3_Final_Project_Step_1

Maxim Bilenkin

2025-02-06

Topic: Excessive consumption of alcohol causes health issues

It is no secret that throughout our lives we go through hard times and face various hardships. Not all, but many people turn to alcohol instead of sport to ease their emotions. With excessive consumption of alcohol health issues arise. Many commercials and advertisements we see today on television and internet that advertise various products and services. But there is nothing or at least very few that would advertise health issues related to excessive consumption of alcohol. Perhaps many people quit drinking after they get ill. But little did they know before that many health issues they are experiencing now are caused by the excessive consumption of alcohol.

Any health-conscious person wants to live a happy and healthy life. With the advancement of technology and the availability of big data people can be informed well in advance about issues that could cause health issues. Necessary early actions can be taken at early stages to prevent health issues. It's a data science problem because with the help of data science we can predict, understand and mitigate the impact of excessive alcohol consumption on individuals and our society. No doubt with the help of data science people can monitor their health better. With more healthy people our society improves in all aspects of life. Economy innovates, saves money on health-related issues and prospers.

The following questions need to be asked:

1) How do social media and advertisements correlate with excessive alcohol consumption? Do more advertisements increase alcohol consumption?
2) How does excessive alcohol consumption effects health in the short and long run?
3) What are the most effective ways to reduce excessive alcohol consumption?
4) Do genetic factors influence excessive alcohol consumption?
5) How do government policies impact alcohol consumption? Do high taxes on alcohol decrease consumption?
6) What is the cost of excessive alcohol consumption to the economy?
7) How to develop personalized treatment plans to prevent excessive alcohol consumption?
8) What are the key predictors of excessive alcohol consumption?
9) What are the environmental factors that are influencing or contribute to excessive drinking behaviors?
10) Are wearable devices including mobile apps effective in monitoring and decreasing alcoholic consumption?

To approach the issue, I would like to read as much as possible different articles and information that are available on the internet. I would gather all available statistical data that were collected on excessive alcohol consumption. I would analyze it and make my own assessment based on the trends and patterns I see.

Obviously, my approach only addresses the issue partially. It only gives good insight into the problem. It identifies the issue and causes which can give guidance on what preventive measures can be taken to reduce alcoholic excessive consumption. The excessive alcoholic consumption is a complex problem that involves different aspects of our lives. To fix the issue different factors should be involved such as public health

initiatives, customizing personal treatment, preventive policy intervention and social support from family and friends.

For my project I will use the following three data sets to perform my analysis.

1) World Health Organization (WHO) alcoholic consumption globally data set spans from 1960 – 2020 with 48,924 records. This data set will give me insight into daily alcohol consumption for each year per capita measured in grams. (source link: https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/levels-of-consumption) (zip file link: https://ghobulkdownloads.blob.core.windows.net/ghocontainer/levels-of-consumption.zip)
2) Kaggle provides a data set for Alcohol and Life Expectancy. This international study will give me insight into whether excessive alcohol consumption impacts human lives. (source link: https://www.kaggle.com/datasets/thedevastator/relationship-between-alcohol-consumption-and-lif/data)
3) This survey was conducted on alcohol consumption and happiness. I will use this data set by Kaggle to see if there is a positive or negative correlation between the two. (Source link: https://www.kaggle.com/datasets/marcospessotto/happiness-and-alcohol-consumption)

To perform my analysis, I will utilize the following R programming packages.

SNPassos for data manipulation and data exploratory analysis.

genetics for creating and handling genetic data.

GA for identifying and using genetic markers with alcohol consumption.

dplyr for data transformation and manipulation including cleaning and preparing data.

ggplot2 to visualize my data that will help me to understand patterns.

caret to predict alcoholic consumption.

tidyverse this is a collection of many packages that I will use to streamline my workflow.

All the packages described above should address all my needs.

For my project I will use the following plots to visualize correlation, compare distributions and assess test statistics. For example, Manhattan Plot, Q-Q Plot, Heatmap, Boxplot and ROC Curve.

It would be good to ask the following questions for future steps.

What are the most efficient and effective ways to prevent excessive alcohol consumption?

Will my research only be effective for the USA population or can it be applied globally?

What are the most culprit factors that cause excessive alcohol consumption?

Are there any ways to interact with genes to prevent alcohol abuse?