

An Exploration of Self-Percieved Loneliness Through Social Connection

The Data

Our data comes from the Canadian Social Connection Survey, conducted by the Canadian Alliance for Social Connection and Health (CASCH). The data is a cross sectional survey conducted in 2022 involving a wide range of questions regarding social, mental and physical health, socioeconomic standing, and other indicators similar in scope.

Methodology

- Data cleaning and initial visualization using multiple python libraries and built-in methods.
- Constructed two separate datasets, one intended for a logistic regression model and another for a machine learning prediction model based on a binary classification decision tree.
- In order to digest and in some cases improve model performance we worked with specific age ranges

Our motivation

The inspiration for our analysis comes from our personal experience learning about senior loneliness and the rising mental health crisis among young adults. In hopes of better understanding the aforementioned issues, we wished to see them from a new lens; predominantly using social indicators such as interaction with others and personal relationships to better map out any possible correlation or causation. We are also interested in both raising awareness of this crisis and searching for potential action using models to predict and help at-risk individuals in society.

Background research

We were further emboldened to follow this course of research when we discovered that a report by The F/P/T Working Group on Social Isolation and Social Innovation in collaboration with the Federal and Provincial governments report that roughly 30% of seniors face loneliness (1).

Likewise, preliminary research insight from the CSCS indicates a higher likelihood of happier people reporting playing videogames with others weekly (2), which we interpret as there being a correlation between playing videogames and happiness, as well as lower levels of perceived loneliness.

(1) <https://www.canada.ca/en/employment-social-development/corporate/seniors-forum-federal-provincial-territorial/social-isolation-toolkit-vol1.html>
(2) https://casch.org/s/CSCS_2021_Report.pdf

Observed variables

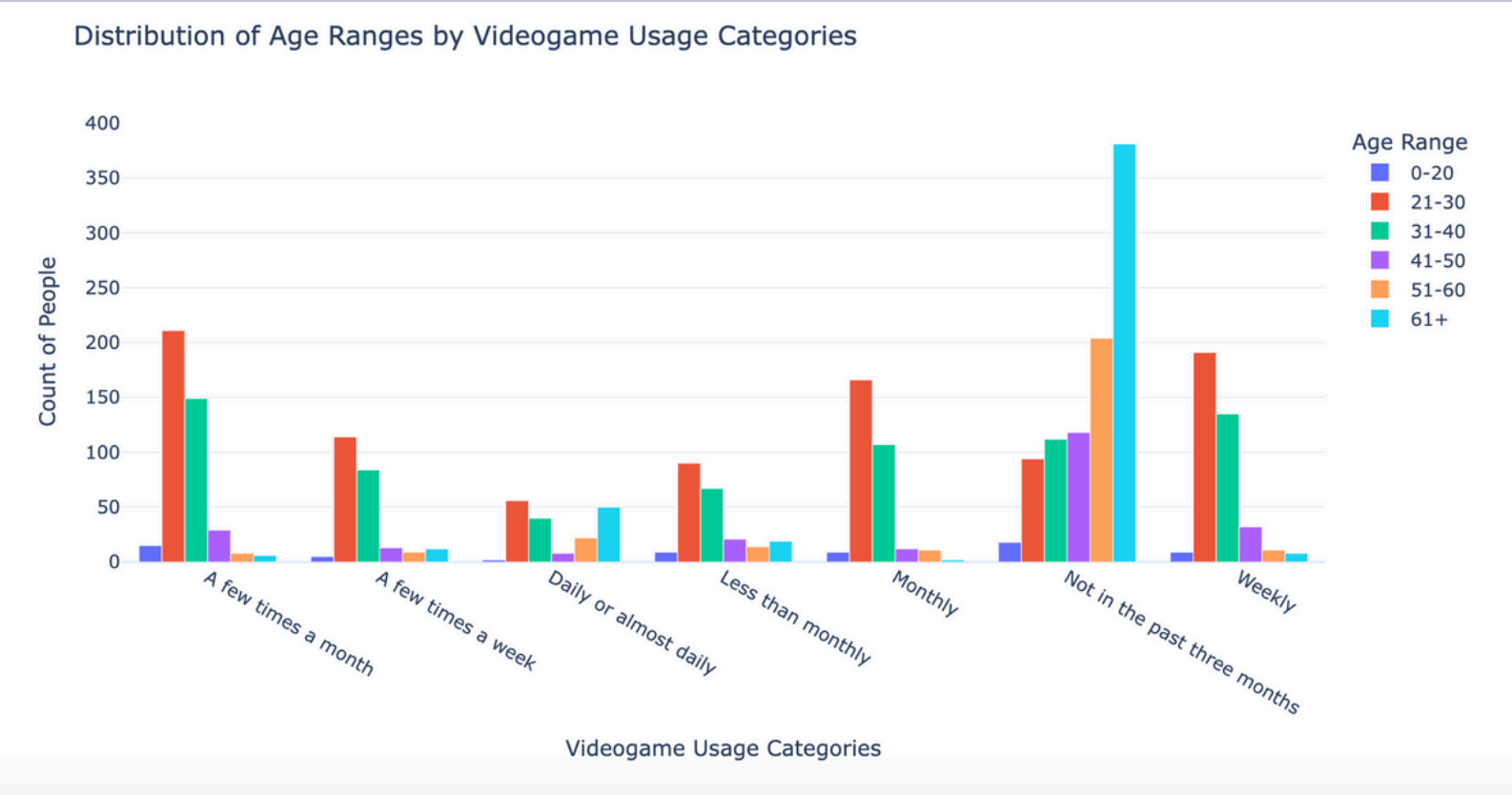
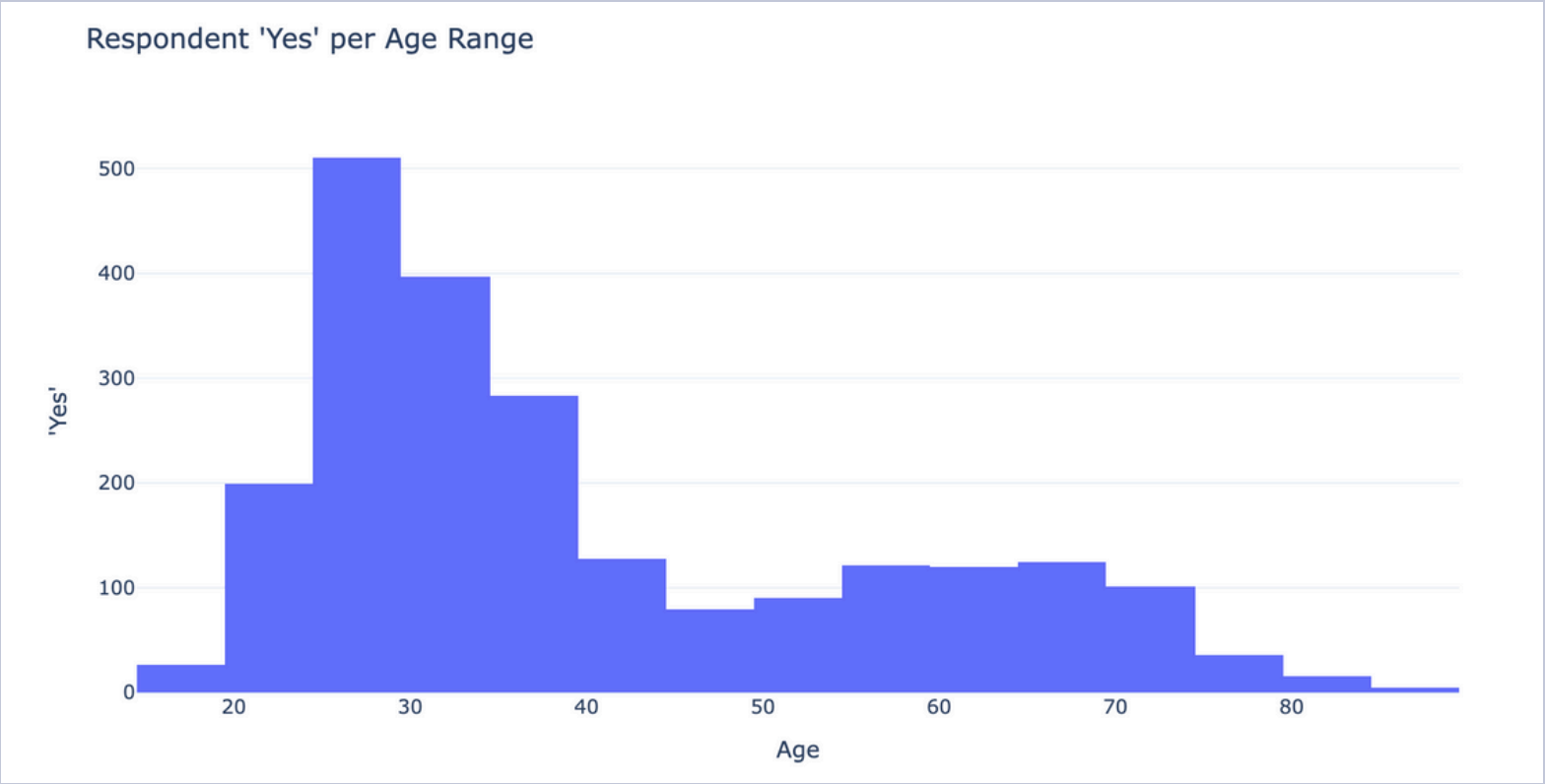
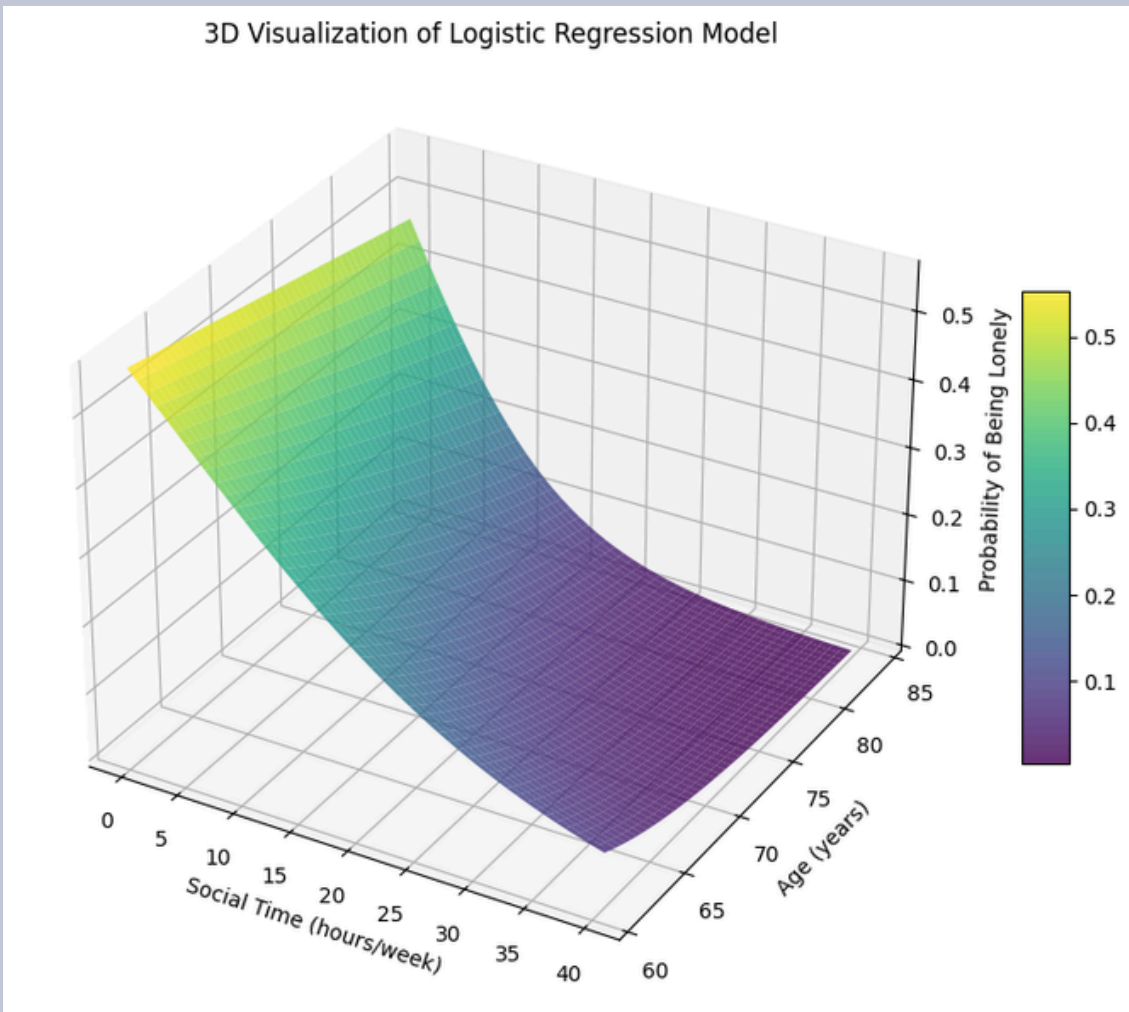
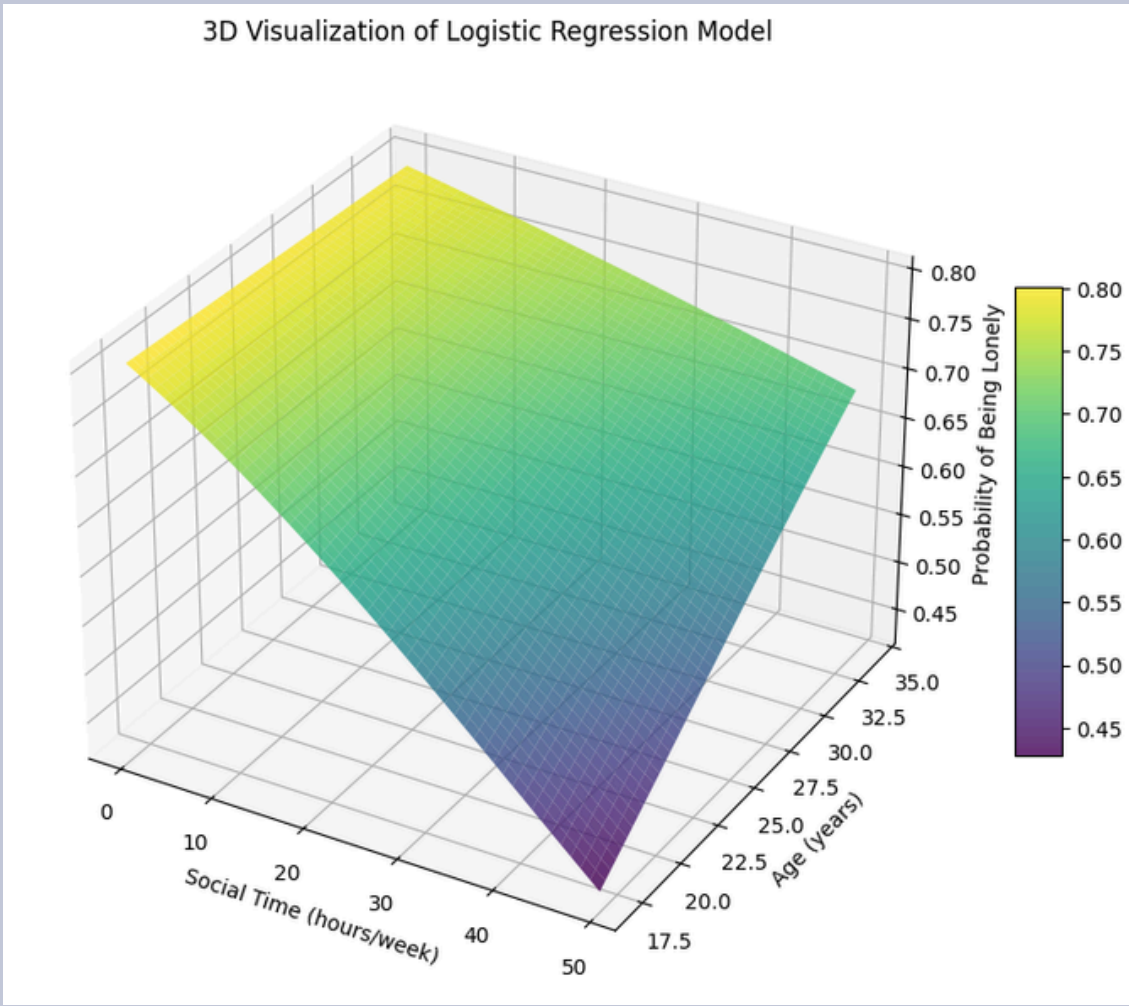
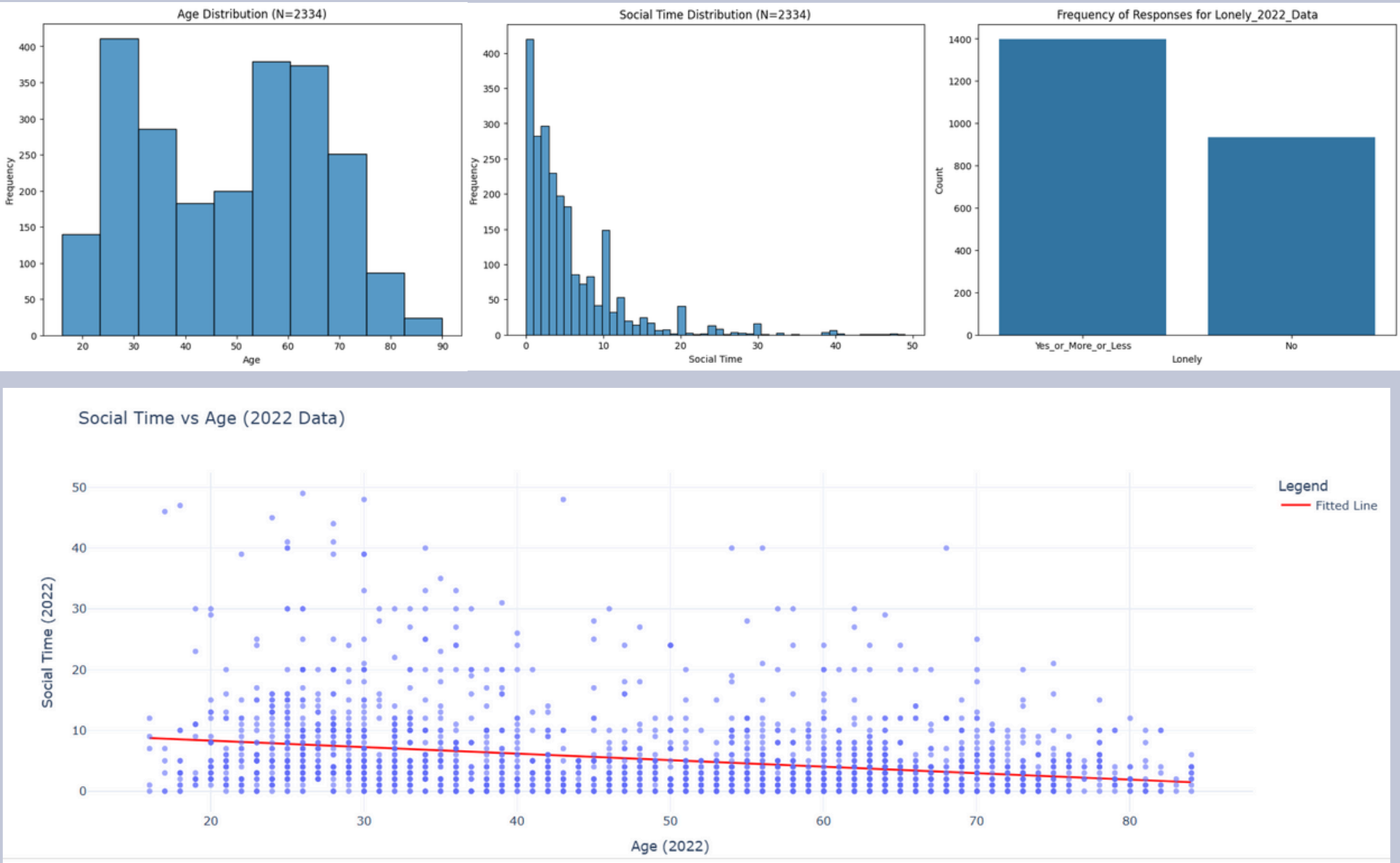
- Age: Ages of all the participants in our cohort.
- Loneliness (our outcome variables):
 - “I experience a general sense of emptiness”, classified into yes or no.
 - “There are enough people I feel close to”, classified into yes or no.
- Time spent socializing: A variable storing the participant’s hours spent with friends per week.
- Video game usage in the past three months, classified into the following categories:
 - Not in the past three months
 - Less than monthly
 - Monthly
 - A few times a month
 - Weekly
 - A few times a week
 - Daily or almost daily

Key trends

For our logistic model, we predict the probability someone would respond yes or more or less to the prompt “I experience a general sense of emptiness” using the predictor variables of age and social time spent with friends.

For our ML model, we are predicting the outcome of the following prompt, “There are enough people I feel close to”. Our predictor variables are videogame usage in the past three months and age.

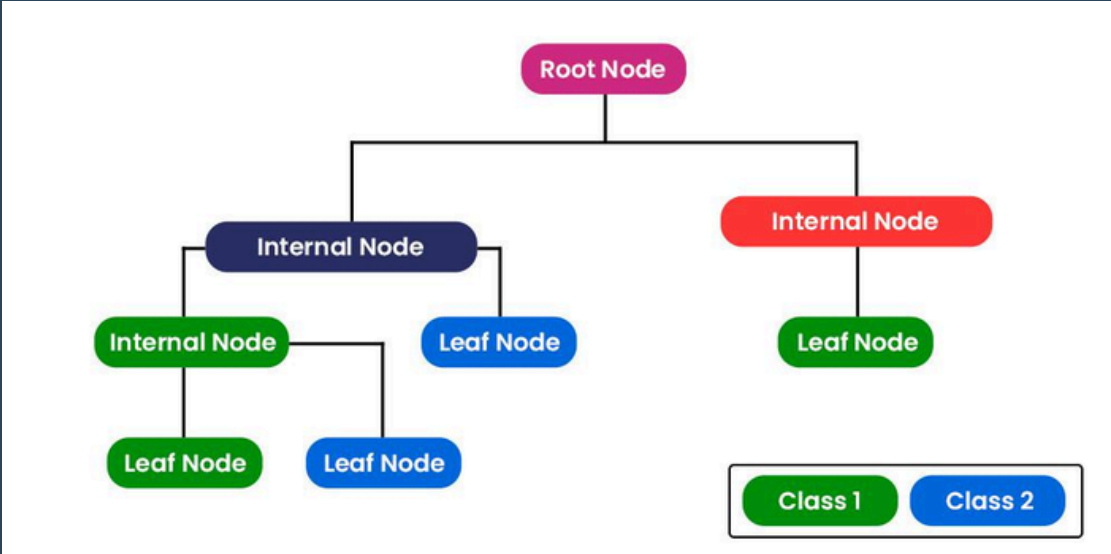
The following charts show some interesting trends in our two datasets. Note that our observations of this data led us to separate our models into two distinct age brackets.



The Binary Classification Decision Tree

Sectioned the data using an 80-20 train test split, we trained a binary classification decision tree on the age and videogame usage in the past three months variables. We fined-tuned the model’s parameters finding the best possible parameter values using five-fold cross validation. Due to the imbalance of ‘Yes’ and ‘No’, we also used weighted classes.

Due to the age range of around 16 to 35 being the most likely to contain reports of relatively frequent video game usage, we trained the model on this age range. Note that our age brackets are entirely based on the key trends in our specific data set and the general assumption that older generations tend not to be engaged in video game usage. Preliminary model testing and building showed us that models built using bigger age ranges did not have good predictive performance.



Key Results

Testing for accuracy, our model predicted the answer correctly 78.5% percent of the time.

This strong predictive performance is indicative of the correlation between time spent playing videogames with friends and perceived closeness to other people.

This finding, compounded by our preliminary background research, shows promise in future predictive capabilities to highlight individuals who might benefit from exposure to online videogames.

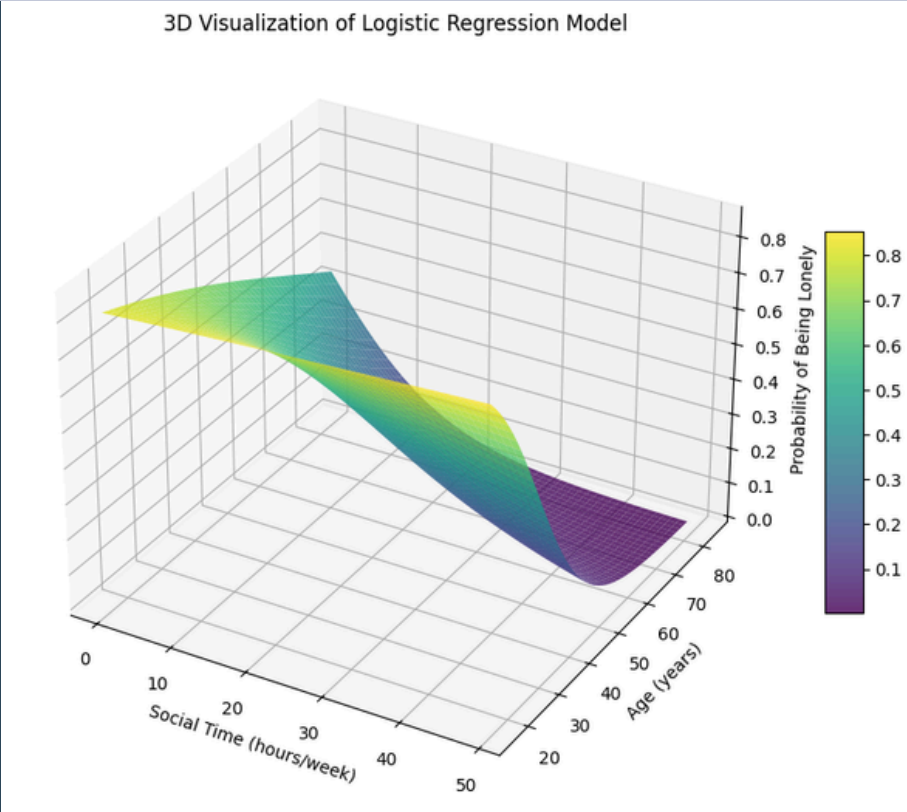
Why is this relevant?

The strong predictive performance of our model shows promise in potential future applications. Using different predictor variables that relate to various social activities to build predictive models could be helpful in highlighting at-risk individuals, or even getting a broad idea of how Canadians are responding to social trends and individual feelings of loneliness. Harnessing machine learning predictive models such as our own may prove to be very useful in aiding organizations take action to benefit Canadians and society at large.

The Logistic Regression Model

When creating the logistic regression model we initially began by plotting and predicting the entire age range as can be seen on the right. This gave us an overall picture of how the model predicts perceived loneliness based on our predictor variables. We saw what was expected from our preliminary research where the senior participants who had low socializing time were predicted to have a much higher probability of being lonely. But an alarming trend we saw was in younger participants, where despite their high socializing time, they had much higher probabilities of being lonely.

We then created 2 new models following the same original specifications but trained on two different age intervals, 16 to 35 and 60 to 85 to better see the model’s predictive ability. For our model focused on the participants from 60 to 85, we saw that using our initially calculated average for time spent socializing per week (5.3h), our model predicts that they have a 30-40% chance of being lonely or for every 100 people around 30-40 face senior loneliness, a figure that was very similar to our preliminary research. We also found the trend for the other age group alarming as those with low social time faced nearly 80% chance of being lonely, with the situation ameliorating lightly the more time spent socializing.



Tying it all together

From our research we can deduce two main trends. The first trend which we expected to see was the societal issue of senior loneliness. Our logistic regression model predicted similar rates of loneliness to that of our preliminary research, illustrating how more and more data reflects this issue and how in need we are of concrete effort to solve it. The second trend which we found as alarming is the increasing rates of loneliness and disconnection with others among younger adults and teens as seen from the ML model, a further reflection of the current mental health crisis facing this age range.