Billy To, Andrew Yu

Dataset Nutrition Label

Effects of Mental Health by Hours Spent Online

About

This dataset was created to merge two datasets in order to find a possible correlation between higher hours spent online and mental health. By using a dataset that measures the amount hours spent online, and comparing it to the mental health of online users through increased hours, we can find a correlation. We acknowledge the existence of outside factors, which is included with the dataset

Data Creation Range: 2017 -2018

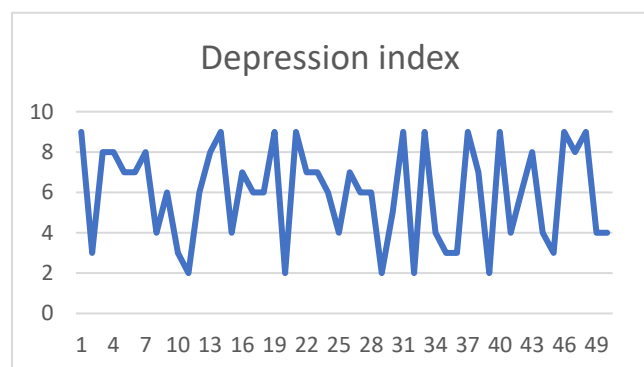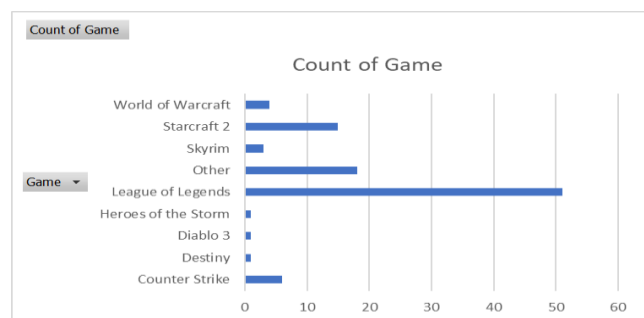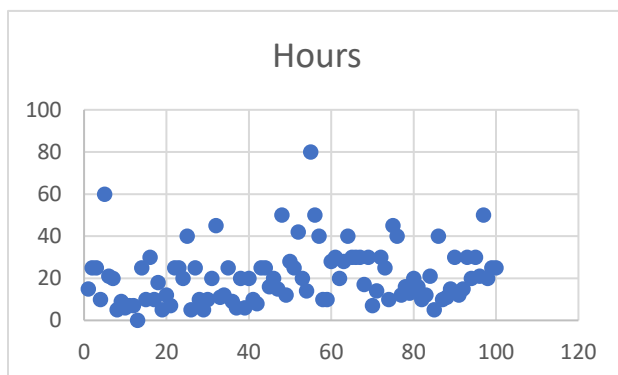Created By: Billy To, Andew Yu

Content: GamingStudy_data.csv,

Source: https://osf.io/vnbxk/

Use Cases

Potential real-world applications of dataset

- Determine the consequences of long gaming hours
- Deliver a benchmark for how long people can play games with minimal side effects
- Finding common traits within groups of gamers and possible causes

Categories

Count of Game

Hours

Depression index

Billy To, Andrew Yu

Dataset Nutrition Label

Effects of Mental Health by Hours Spent Online

Dataset Information

Description

1. **TELL US ABOUT THIS DATASET.**

   This dataset is a evaluation of people who regularly play video games and their corresponding mental health. We account their hours and measure it against certain mental health criteria to measure their life satisfaction

2. **IS THERE AN INTENDED PURPOSE FOR THE DATASET? WHAT DOMAIN WAS IT DESIGNED FOR?**

   The purpose of this dataset is find a correlation between increase hours of gaming and possible mental health consequences

3. **WHAT IS THE LICENSE UNDER WHICH THE DATASET IS MADE AVAILABLE?**

   Public Domain

4. **ARE THERE ANY ADDITIONAL RESTRICTIONS UNDER WHICH THE DATASET IS MADE AVAILABLE? ARE THERE ANY LEGISLATION, CODES OF PRACTICE OR GUIDANCE FROM THE JURISDICTION OR DOMAIN IN WHICH THE DATA WAS COLLECTED ABOUT THE USE OF THIS DATA? ARE THERE TASKS FOR WHICH THE DATASET CANNOT BE USED? PLEASE PROVIDE A DESCRIPTION OR GUIDANCE.**

   No

5. **ARE THERE TASKS FOR WHICH THE DATASET WOULD BE CAUTIONED AGAINST BEING USED? IF SO, PLEASE PROVIDE A DESCRIPTION.**

   No

6. **HAS THE DATASET BEEN USED FOR ANY TASKS ALREADY? IF SO, PLEASE PROVIDE A DESCRIPTION AND LINKS TO PAPERS OR SYSTEMS USING THE DATASET** NO

Billy To, Andrew Yu

Dataset Nutrition Label

Effects of Mental Health by Hours Spent Online

Composition

1. **DOES THE DATASET HAVE A METADATA REPOSITORY OR DATA DICTIONARY? IF YES, PLEASE PROVIDE THE LINK AND IF NOT, PLEASE EXPLAIN WHAT EACH FIELD MEANS.**

   https://osf.io/bm5ep

2. **IN THE CASE OF EXPLICITLY RELATED INSTANCES, ARE THESE RELATIONSHIPS INDICATED? (E.G., RELATING AN INSTANCE ABOUT A USER TO INSTANCES ABOUT THEIR POSTS)**

   No

3. **WHAT DATA DOES EACH INSTANCE CONSIST OF? FOR EXAMPLE, DOES IT CONSIST OF RAW DATA (E.G., UNPROCESSED TEXT OR IMAGES) OR FEATURES? IN EITHER CASE, PLEASE PROVIDE A DESCRIPTION.**

   This data contains raw, data statistics of people and their corresponding response to questions

4. **HOW MANY INSTANCES ARE THERE IN TOTAL (OF EACH TYPE, IF APPROPRIATE)?**

   33 columns, 100+ people

5. **IS THERE A LABEL OR TARGET ASSOCIATED WITH EACH INSTANCE?**

   No

6. **IS ANY INFORMATION MISSING FROM INDIVIDUAL INSTANCES? IF SO, WHAT IS MISSING AND WHY?**

   No information missing

Dataset Nutrition Label

Effects of Mental Health by Hours Spent Online

7. IS YOUR DATASET A SAMPLE? IF SO, WHAT WAS THE SAMPLING STRATEGY USED (E.G. DETERMINISTIC, PROBABILISTIC WITH SPECIFIC SAMPLING PROBABILITIES), AND DOES IT ACCURATELY REPRESENTS THE INTENDED OUTPUT?

Yes, used a questionnaire targeted at a population that was selected through simple random sampling, based on a self-diagnosis of "plays games regularly."

8. WAS ANY PREPROCESSING/CLEANING/LABELING OF THE DATA DONE? IF SO, PLEASE PROVIDE A DESCRIPTION.

Yes, data was merged from two dataests

9. IF THE DATA WAS PROCESSED, WAS THE RAW DATA SAVED IN ADDITION TO THE PREPROCESSED/CLEANED/LABELED DATA? IF YES, PLEASE PROVIDE A LINK TO THE RAW DATA.

N/A

10. ARE THERE RECOMMENDED DATA SPLITS? IF SO, PLEASE PROVIDE A DESCRIPTION OF THESE SPLITS, EXPLAINING THE RATIONALE BEHIND THEM.

No

11. ARE THERE ANY ERRORS, SOURCES OF NOISE, OR REDUNDANCIES IN THE DATASET? IF SO, PLEASE PROVIDE A DESCRIPTION

N/A

DOES THE DATASET INCLUDE INFORMATION FROM UPSTREAM SOURCES? IF SO, NAME THESE SOURCES, THEIR STABILITY, AND ANY KNOWN USAGE LIMITATIONS.

No

Dataset Nutrition Label

Effects of Mental Health by Hours Spent Online

12. IF THE DATASTET INCLUDES INFORMATION FROM UPSTREAM SOURCES, IS THIS DATASET TECHNICALLY DEPENDENT UPON THOSE SOURCES?

N/A

13. IS THERE CONFIDENTIAL DATA INCLUDED IN THIS DATASET?

No

14. DOES THE DATASET IDENTIFY ANY SUBPOPULATIONS (E.G., BY AGE, GENDER)? IF YES, PLEASE INDICATE THEIR DISTRIBUTIONS WITHIN THE DATASET WHERE POSSIBLE / AVAILABLE.

No

Provenance

1. WHO CREATED THE DATASET AND ON BEHALF OF WHICH ENTITY?

The NYC department of finance created the dataset on behalf of New York City's municipal government.

2. HOW IS / WAS THE COLLECTION AND MANAGEMENT OF THIS DATASET FUNDED?

Publicly-funded

Collection

1. OVER WHAT TIMEFRAME WAS THE DATA COLLECTED? IF THE DATA CONTINUES TO BE UPDATED, PLEASE INDICATE.

The dataset was collected over a course of several months

Dataset Nutrition Label

Effects of Mental Health by Hours Spent Online

2. **WHAT MECHANISMS OR PROCEDURES WERE USED TO COLLECT THE DATA?**

   This data is collected through survery

3. **WHAT IS THE RELATIONSHIP BETWEEN THE DATASET COLLECTOR AND OWNERS OR MANAGERS OF THE DATASET?**

   No relationship

4. **IF INDIVIDUALS' DATA IS INCLUDED IN THIS DATASET, DID THOSE INDIVIDUALS CONSENT TO THE COLLECTION AND USE OF THEIR DATA? IF SO, PLEASE DESCRIBE CONSENT PROCEDURE.**

   N/A

5. **IF INDIVIDUALS' DATA IS INCLUDED IN THIS DATASET, WAS THIS DATA ALTERED TO ENSURE HIGHER LEVELS OF PRIVACY? IF YES, PLEASE DESCRIBE ANY PRIVACY PROCEDURES FOLLOWED WITH REGARDS TO THIS DATASET (ANONYMIZATION EFFORTS, PRIVACY PROTOCOLS, SUPRESSION TECHNIQUES, ETC).**

   N/A

6. **HAS THE DATA BEEN REVIEWED FOR QUALITY?**

   No

7. **WERE ANY ETHICAL REVIEW PROCESSES CONDUCTED (E.G., BY AN INSTITUTIONAL REVIEW BOARD)? IF SO, PLEASE PROVIDE A DESCRIPTION OF THESE REVIEW PROCESSES, INCLUDING THE OUTCOMES, AS WELL AS A LINK OR OTHER ACCESS POINT TO ANY SUPPORTING DOCUMENTATION.**

   No

Effects of Mental Health by Hours Spent Online

Management

1. **HOW IS THE DATA BEING MANANGED AT REST AND IN TRANSIT?**

   Internally

2. **HOW CAN THE DATA BE ACCESSED? IF THIS IS GOING TO CHANGE OVER TIME, PLEASE INDICATE THIS.**

3. **WILL THE DATASET BE DISTRIBUTED TO ANY OTHER INDIVIDUALS/THIRD PARTIES (E.G. COMPANIES, INSTITUTIONS, ORGANIZATIONS)? IF YES, WHEN WILL THE DATASET BE DISTRIBUTED AND HOW WILL THE DATASET BE DISTRIBUTED (E.G. GITHUB, TARBALL ON THE WEBSITE, ETC)?**

   No

4. **WILL THE DATASET BE UPDATED (E.G., TO CORRECT LABELING ERRORS, ADD NEW INSTANCES, DELETE INSTANCES)? IF SO, PLEASE DESCRIBE HOW OFTEN, BY WHOM, AND HOW UPDATES WILL BE COMMUNICATED TO USERS (E.G., MAILING LIST, GITHUB)?**

   Data will not be updated

5. **IS THERE A MECHANISM THROUGH WHICH AN INDIVIDUAL CAN REQUEST THEIR PERSONAL INFORMATION BE REMOVED?**

   N/A