Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Data Visualization and Predictive Modeling in Mental Health: Insights from the Patient Health Questionnaire Analysis

Subash C[1], Santhi K.[2*], R. Eswar Reddy[3], and Chellatamilan T[4]

[1]Research Clinical Junior Doctor, nference, Inc. Coimbatore, Tamil Nadu 641014.
[2]Professor, School of Computer Science and Engineering.*
[3]Research Scholar, School of Computer Science and Engineering.
[4]Associate Professor, School of Computer Science and Engineering.
Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India- 632014.
[1]subash.c@nference.net,
[2]santhikrishnan@vit.ac.in*
[3]regantieswar.reddy2021@vitstudent.ac.in,
4chellatamilan.t@vit.ac.in.

Abstract: **-** The Patient Health Questionnaire (PHQ) is a valuable set of standardized self-report questionnaires. This PHQ is widely employed in clinical settings, primary care, and mental health services to assess and screen for a variety of mental health conditions. The PHQ encompasses various modules customized to specific disorders, such as the PHQ-9 for depression, GAD-7 for generalized anxiety disorder, and PHQ-15 for somatic symptom severity, among others. These modules offer a structured and standardized means to collect information about an individual's mental health, facilitating treatment planning and progress evaluation. It is important to emphasize that the analysis of PHQ data has been performed using the IBM Congo's application, enhancing the efficiency and accuracy of the assessment process. It is crucial to recognize that the PHQ primarily serves as a screening and assessment application. Any definitive diagnosis or treatment decisions should always involve a comprehensive clinical evaluation conducted by a trained healthcare professional. The PHQ, when combined with modern analytical applications such as IBM Congo's, represents a valuable resource in the field of mental health evaluation, allowing for more efficient and specific patient therapy.

Keywords*:* *Patient Health Questionnaire, generalized anxiety disorder, Subjective Happiness Scale, IBM cogon's analytics, Happiness score.*

1. Introduction

Major depression affects more than 300 million people globally. It has become one of the primary causes of decreased quality of life, work incapacity, and early death. Effective, evidence-based therapy methods are available, but current estimates of the global treatment gap show that only 7% (in low-income countries) to 28% (in high-income countries) of people suffering from depression receive treatment. Structural and individual impediments are at the heart of this global public health concern. Healthcare systems frequently lack the financial and human resources to administer depression therapies at scale. Furthermore, notably milder cases of major depression frequently go undiagnosed by primary care physicians, who are the primary and sometimes only point of contact for the vast majority of patients suffering with depressive symptoms. Access to secondary care is frequently hampered by issues such as high costs, geographical unavailability, and lengthy waiting lists. Individual patient difficulties in detecting symptoms or the desire to tackle emotional problems on one's own are frequently accompanied by a lack of faith in professionals or a fear of stigmatization. All of these characteristics diminish the likelihood of early discovery, contribute to the problem of underdiagnosis, and raise the risk of long-term symptom worsening and chronification. While general, population-wide screenings for depression remain a contentious issue in the literature, there is an evident need to improve early detection for persons experiencing depression symptoms, as well as to give low-threshold access to accessible mental health care systems.

The Patient Health Questionnaire is a series of standardized self-report questionnaires designed to assess and screen for various mental health conditions. The PHQ is widely used in clinical settings, primary care, and mental health services to assist healthcare providers in assessing and monitoring mental health conditions in patients [1] [2]. The PHQ is relatively easy for individuals to complete, and it does not require specialized training to interpret the results. This makes it a practical tool for a wide range of healthcare providers. The PHQ provides valuable information; it is not a substitute for a comprehensive clinical assessment by a trained healthcare professional. It serves as a screening and assessment tool, and any diagnosis or treatment decisions should be made in conjunction with a thorough evaluation. The data collection more efficient and facilitates data analysis, of yield objective data that can be analyzed quantitatively, useful for statistical analysis. In structured questionnaires, such as the PHQ, severity and frequency of episodes are assessed.

1. Severity Assessment:
    - Severity refers to the intensity or degree of a particular symptom or experience. In the context of mental health assessments like the PHQ, it is about understanding how strongly an individual is affected by a specific symptom.
    - The PHQ-9, respondents are asked to rate how often they have experienced symptoms like "little interest or pleasure in doing things" and then to rate the severity of this feeling on a scale from "Not at all" to "Nearly every day".
    - Respondent chooses one of the options - "Not at all", "Several days", "More than half the days", or "Nearly every day". This indicates how intensely they have been feeling this way.
2. Frequency Assessment:
    - Frequency pertains to how often a symptom occurs within a defined period. In mental health assessments, this is typically over a specific period, often the past two weeks.
    - The same PHQ-9 respondents are asked to indicate how many days in the past two weeks they have experienced each symptom.
    - Respondent selects a number of days (from 0 to 14) in the past two weeks that they have felt this way. This indicates how often they have experienced this symptom.

The results of PHQ assessments can be used to assist distribute resources in a more efficient manner, with higher PHQ scores possibly being prioritized for interventions that are more intensive [4] [5]. Individuals can be educated about the symptoms of depression and given assistance in better understanding their personal experiences within the context of a professional setting by using the PHQ. These questionnaires are widely used in clinical settings, primary care, and mental health services to assist healthcare providers in assessing and monitoring mental health conditions in patients. It is illustrated in Table 1 [6].

Table 1: The Patient Health Questionnaire

		Not at all	Several Days	More than half the days	Neverly every day
1	Little interest or pleasure in doing things.	0	1	2	3
2	Feeling down, depressed, or hopeless.	0	1	2	3
3	Trouble falling or staying asleep, or sleeping too much.	0	1	2	3
4	Feeling tired or having little energy.	0	1	2	3
5	Poor appetite or overeating.	0	1	2	3
6	Feeling bad about yourself or that you are a failure or have let yourself or your family down.	0	1	2	3
7	Trouble concentrating on things, such as reading the newspaper or watching televisions.	0	1	2	3

8	Moving one speaking so slowly that other people clould have noticed oOr the opposite being so fidgety or restless that you have been moving around a lot more than usual.	0	1	2	3
9	Thoughts that you would be better off dead, or of hurting yourself.	0	1	2	3
		Add columns	Several Days+ More than half then days+ neverly every day		
		Total			

The PHQ-9 is a widely used self-report questionnaire designed to screen for and assess the severity of depression in individuals. It is based on the criteria for major depressive disorder outlined in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). The PHQ-9 consists of nine questions that ask about various symptoms of depression analyzing experienced over the past two weeks [5] [7]. Each response to a question assigned a numerical value based on a predefined scale. Each item is scored on a scale of 0 to 3, with 0 indicating "not at all" and 3 indicating "nearly every day." These scores are then added together to calculate a total score. For instance, if an individual's responses on the PHQ-9 were 2, 1, 2, 0, 1, 2, 1, 3, 2 for each respective question, the total score would be 14.The scores for each item are then added together to give a total score ranging from 0 to 27. The total score on the questionnaire provides an indication of the severity of symptoms. The interpretation of the score depends on the specific questionnaire being used. The PHQ-9, the score ranges and their corresponding interpretations are:

- 0: Not all
- 1-4: Minimal depression
- 5-9: Mild depression
- 10-14: Moderate depression
- 15-19: Moderately severe depression
- 20-27: Severe depression

This is a series of standardized self-report questionnaires designed to assess and screen for various mental health conditions. The PHQ includes specific modules for different disorders, such as depression PHQ-9, GAD-7, panic disorder, and more. PHQ-9 by used to assess and screen for depression. It consists of nine questions about depressive symptoms [8]. GAD-7 is used to assess and screen for generalized anxiety disorder. It consists of seven questions about anxiety symptoms. PHQ-15 is Assesses somatic symptom severity and somatic symptom disorder. The PHQ-9 consisting of the first two questions [5]. It is used for a quick initial screen for depression. PHQ-8 is Similar to the PHQ-9, but without the question about suicidal ideation. PHQ-4 is a shortened version consisting of four questions that screen for both depression and anxiety. PHQ-A (Adolescent) is designed for adolescents and includes questions about depressive symptoms. Panic Disorder Module is this module is specifically designed to assess and screen for panic disorder. Other Modules is in addition to the ones listed above, there are specific modules designed to assess conditions like alcohol use disorder, eating disorders, and more. These modules are used in clinical settings, primary care, mental health services, and research studies to help healthcare providers and researchers assess and monitor various mental health conditions.

GAD is a common mental health condition characterized by excessive and persistent worry or anxiety about a variety of things. This chronic anxiety is not limited to specific situations or events, and individuals with GAD often find it difficult to control their worry [8] [9] [10]. GAD can be accompanied by physical symptoms such as restlessness, muscle tension, fatigue, irritability, and sleep disturbances. Chronic disorders are characterised by excessive, long-term anxiety and worry about generic life events, objects, and situations. GAD is the most common anxiety disorder, and persons with it are often unable to pinpoint the source of their concern [7] [11]. Depression, also known as major depressive disorder (MDD), is a serious and common mental health condition characterized by persistent feelings of sadness, hopelessness, and a lack of interest or pleasure in activities that were once enjoyable. It significantly affects a person's thoughts, feelings, and daily functioning. Certainly, here are some sample questions that are typically included in the Generalized Anxiety Disorder 7 (GAD-7) questionnaire, which geared toward assessing generalized anxiety disorder:

1. Over the last two weeks, how often have you been bothered by excessive worry about various events or activities?
   - Not at all
   - Several days
   - More than half the days
   - Nearly every day
2. Over the last two weeks, how often have you found it difficult to control your worrying thoughts?
   - Not at all
   - Several days
   - More than half the days
   - Nearly every day
3. Over the last two weeks, how often have you felt restless, keyed up, or on edge?
   - Not at all
   - Several days
   - More than half the days
   - Nearly every day
4. Over the last two weeks, how often have you become easily fatigued?
   - Not at all
   - Several days
   - More than half the days
   - Nearly every day
5. Over the last two weeks, how often have you had trouble concentrating on things, such as reading a book or watching television?
   - Not at all
   - Several days
   - More than half the days
   - Nearly every day
6. Over the last two weeks, how often have you felt irritable?
   - Not at all
   - Several days
   - More than half the days
   - Nearly every day
7. Over the last two weeks, how often have you felt that something awful might happen?
   - Not at all
   - Several days
   - More than half the days
   - Nearly every day

Respondents typically rate each question on a scale from "Not at all" (0 points) to "Nearly every day" (3 points), and the scores are added together to assess the severity of generalized anxiety symptoms. Higher total scores indicate more severe anxiety symptoms. If you or someone you know is experiencing significant anxiety, it is essential to seek support and guidance from a healthcare professional or mental health expert.

Subjective Happiness Scale (SHS) is a widely used psychological assessment tool designed to measure an individual's self-reported subjective well-being and happiness [1]. It was developed by Sonja Lyubomirsky and Heidi S. Lepper in 1999 [9] [12]. The SHS is designed to assess individuals’ sense of well-being and happiness based on their own subjective evaluation. The Satisfaction with Life Scale (SWLS) is a response scale that is formatted in a Likert-type manner and has 7 points. The range of scores that can be received is from 5 to 35, with a score of 20 serving as the point of equilibrium on the scale. If a person scores between 5 and 9, it indicates that they are extremely unhappy with their life, whereas if they score between 31 and 35, it indicates that they are extremely delighted with their lives. It has been shown that the coefficient alpha for the scale can range from.79 to.89, which indicates that the scale has a high level of internal consistency. In addition, it was discovered that the scale had good test-retest correlations (.84, and.80 after an interval of a month), respectively.

The present research presents the findings of an initial validation of a new scale called the satisfaction with work scale. This SWLS evaluate an overall perception of their job satisfaction [13] [14] [15]. Two separate investigations were carried out with the purpose of evaluating the new scale's psychometric features. These properties included reliability, construct validity, concurrent validity, and convergent validity. The exploratory factor analysis performed in the first study (N = 194) showed a high level of reliability ((N=.91), as well as a moderate to strong correlation with concurrent scales and convergent measures, such as perceived occupational stress and turnover intention. The unequivocally validated by the results of the confirmatory factor analysis that was carried out in the second research (N = 221). The findings of this early validation suggest that the newly designed SWWS is a valid and reliable instrument for assessing the level of job satisfaction experienced by workers all over the world. It is represented in Figure 1.

		Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree
1.	In most ways my life is close to my ideal.	1	2	3	4	5	6	7
2.	The conditions of my life are excellent.	1	2	3	4	5	6	7
3.	I am satisfied with my life.	1	2	3	4	5	6	7
4.	So far I have gotten the important things I want in life.	1	2	3	4	5	6	7
5.	If I could live my life over, I would change almost nothing.	1	2	3	4	5	6	7

Score	Level of Satisfaction with Life
31 - 35	Extremely satisfied
26 - 30	Satisfied
21 - 25	Slightly satisfied
20	Neutral
15 - 19	Slightly dissatisfied
10 - 14	Dissatisfied
5 - 9	Extremely dissatisfied

Figure 1: The Subjective Happiness Scale Questionnaire

The overall result is determined by summing up the points received for each individual item in the test. The range of possible scores is from 5 to 35, with 20 being the point on the scale that represents the middle ground. Scoring between 5 and 9 indicates that the responder is highly unhappy with their life, whereas scoring between 31 and 35 indicates that the respondent is extremely delighted with their life. The chart that follows gives various cutoff scores that might be utilized as standards.

2. Methods

This data collection includes findings from a retrospective PHQ-9 created for the goal of screening for depression, as well as data from 14 days of ambulatory assessment (AA) data linked to depressive symptoms and mood ratings. In addition to that, it has information about the participants' demographics, such as their age and gender, among other things. This dataset is made up of a variety of fields, some of which are as follows: phq1, phq2, phq3, phq4, phq5, phq6, phq7, phq8, phq9, age, sex, q10, e11, 12, w13, w14, e16, e46, e47. Score on the happiness scale time period name beginning time Phq day. This kind of survey can assist identify therapies that may be beneficial in reducing the severity and frequency of symptoms, in addition to offering valuable clues about possible causes or triggers connected with depressive episodes. The IBM cogon's analytics application generated the chart and graphics visualization analysis of the dataset [15] [16] [17]. IBM Cognos provides a versatile framework for developing insightful and visually appealing graphs to effectively communicate data insights. These hints may include probable causes or triggers linked with depressed episodes. There are optimistic that will be able to use this enormous collection of data to help guide decisions regarding therapy and, eventually, to improve results for persons who are affected by depression.There are many different visualization methods accessible, from conventional graphs and charts to more intricate visualizations like heat maps and network diagrams.

The PHQ-9 depression dataset [3] provides information pertaining to that evaluation. A general population sample can be used to get insights about depression with the help of this dataset. The information is organized according to the following criteria: participant age and gender, PHQ score (1-9), questions 10-47, happiness score, time/period name/start time, and PHQ day. The PHQ Score column provides information on a scale ranging from 1 to 9 regarding the severity of depressed symptoms reported by the patient. The Age and Gender columns provide demographic information connected to the participants, while Questions 10-47 reflect a range of mental health subjects including anhedonia, weariness, sleep disturbance, and changes in appetite or weight. Responses to these questions are given on a numeric scale ranging from 0 to 4. The Happiness score is an individual's subjective rating at the time of the assessment, with higher scores signifying greater positively toward life as reported. The Time/Period Name/Start Time columns offer date and time information pertaining, while the PHQ day column shows the total number of days that have passed since the beginning of the clinical trial when the assessment period begins.

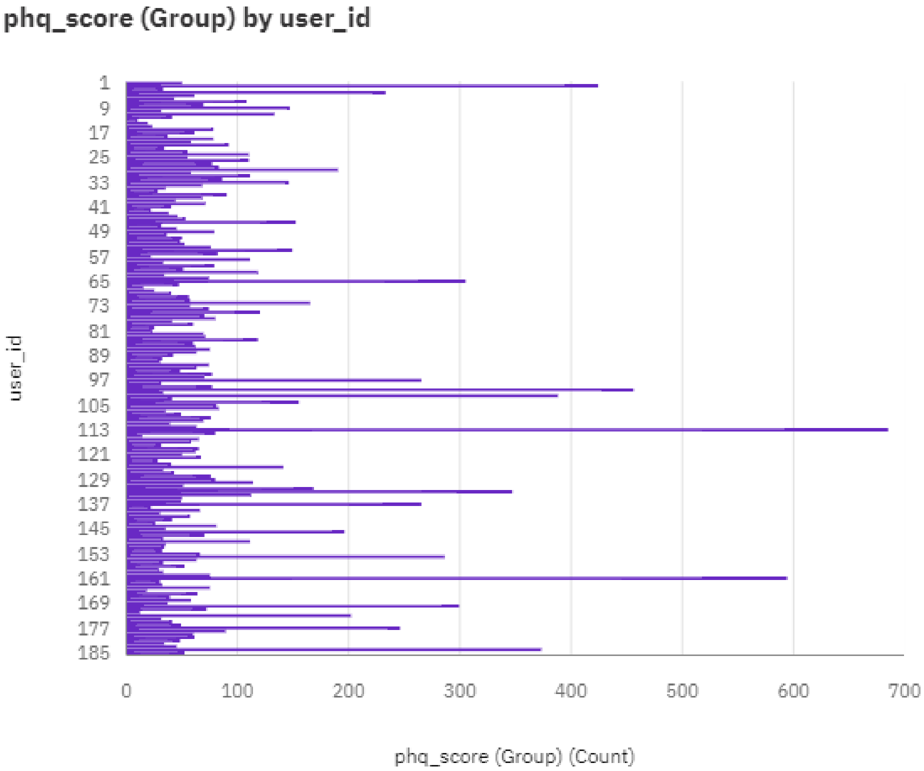

Figure 2: Bar chart in phq score (Group) by User_id

In analyzing this bar chart, shown the figure2, researchers came to the following conclusions, since the user_id and id columns are the same, we can remove one of them from the table [18] [19] [20]. Day column shows the total number of days that have passed since the initiation of the clinical trial at the beginning of the evaluation period, it is important to understand why this column appears to have negative values. A significant number of patients provided numerous responses to the same questionnaire. The table is used to classify each individual based on their ID. Because the values of these variables fluctuate most dramatically in terms of integers and dates, we must manage their dimensionality, which includes category values. Time and start are not the same thing. Time will be compared, as will the absolute value of phq. r chart by using the variables in phq_score (group), across all user_id is 185. phq_score (Group) by user_id colored by phq_score (Group): The total number of results for phq_score (Group), across all user_id, is185.20 to < 28 Severe depression (29.2%), 15 to < 20 Moderately severe depression (23.8%), 10 to < 15 Moderate depression (22.2%), and 5 to < 10 Mild depression (18.4%) are the most frequently occurring categories of phq_score (Group) with a combined count of 173 items with phq_score (Group) values (93.5% of the total).

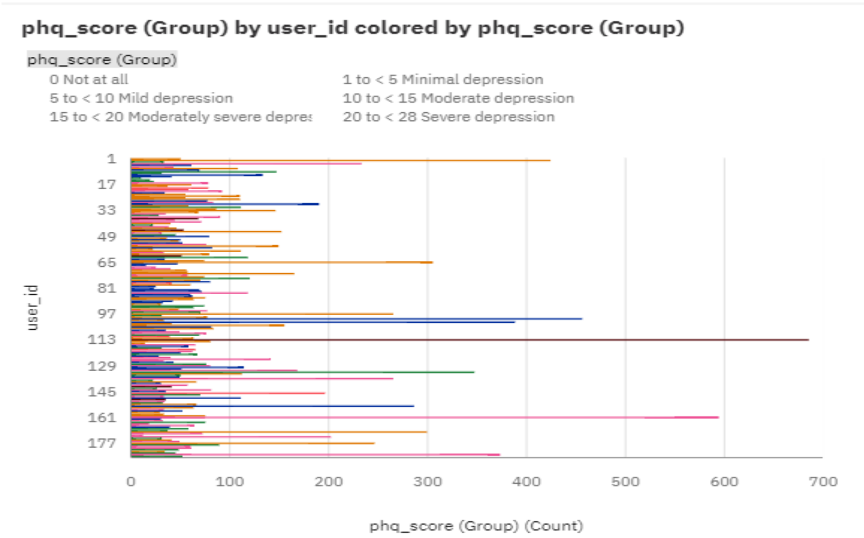

Figure 3: phq_score (Group) by user_id colored by phq_score (Group)

By utilizing real-time data visualization, users may be able to take preventative measures in reaction to issues that have been identified. An animation production strategy is required for the method of interactively exploring time-varying data.This analysis investigates the distribution of PHQ scores across different user IDs, segmented by depression severity levels ranging from minimal to severe. The bar chart reveals variations in PHQ scores, with total PHQ Day values exceeding 185,000 and ranging from nearly -2500 to over 21,000. Notably, PHQ Day values spike with specific combinations of start time and happiness scores, particularly on dates 2 and 3, and are most unusual when happiness scores are 2 and 4. The most significant start time and happiness score (both contributing nearly 72,000 PHQ Day values) account for 38.7% of the total, highlighting significant patterns crucial for targeted mental health interventions. Further analysis could explore the underlying factors contributing to these spikes, the impact of interventions, and long-term trends in PHQ scores, as suggested by the data in figures 3 and 4.

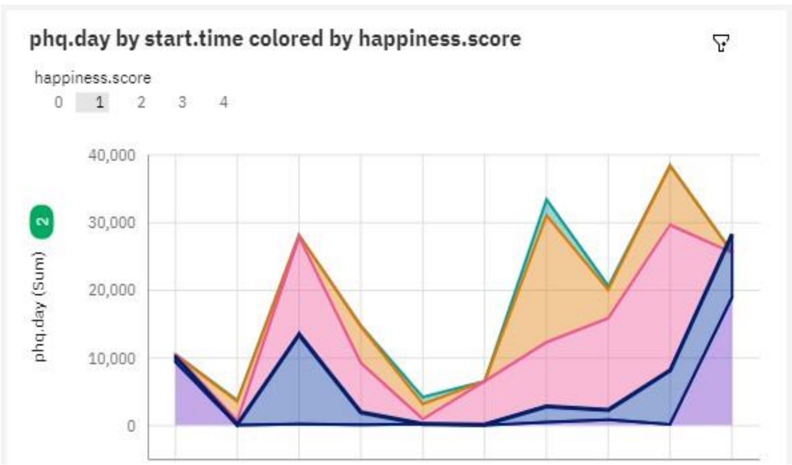

Figure 4: phq in start time coloured.

The tree diagram in figure 5 provides a visual breakdown of happiness scores across different groups. This chart categorizes happiness scores into four ranges: 10-15 (Dissatisfied), 15-20 (Slightly Dissatisfied), 21-26 (Slightly Satisfied), and 26-31 (Satisfied). Each node or segment in the diagram represents a subset of the data, with the happiness score ranges color-coded for clarity.

From the diagram, we can observe how different thresholds and criteria, represented by the decision points on the branches (such as <5, [5, 6], and [6, 7]), split the data into these categories. This visualization aids in understanding the distribution of satisfaction levels and can serve as a tool for identifying factors or conditions that correlate with higher or lower happiness scores, potentially guiding further detailed analysis or interventions based on these insights.

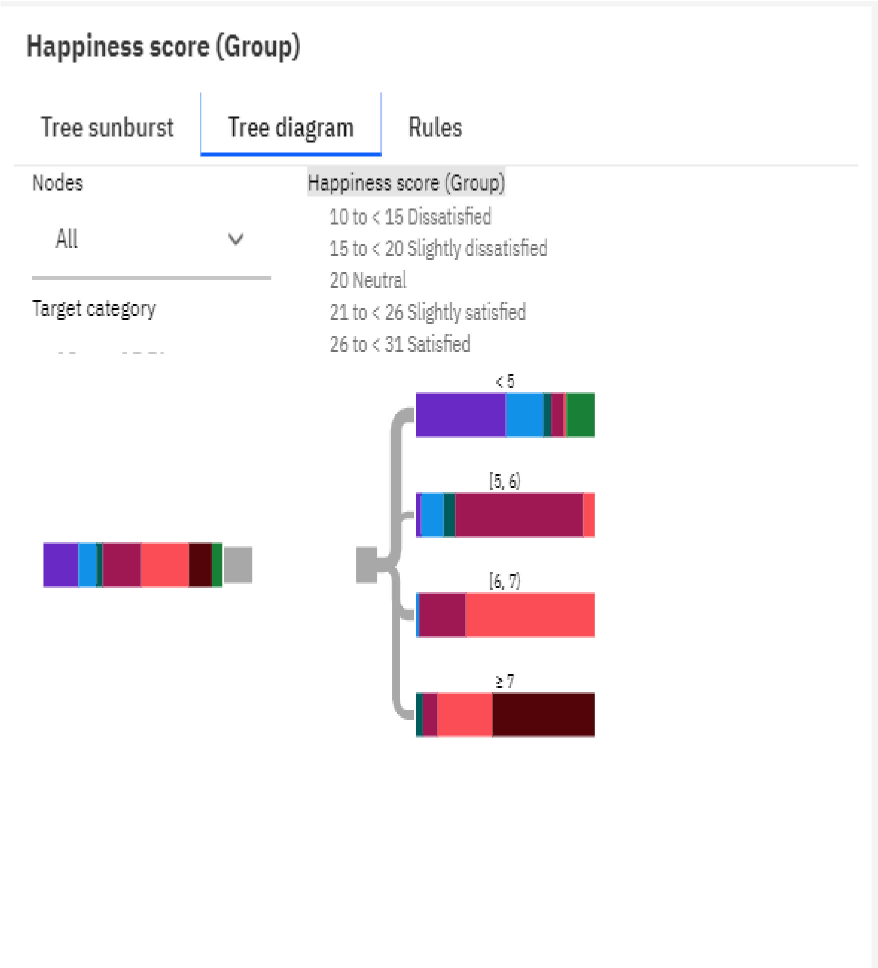

Figure 5: happiness in tree diagram.

Figure 6, displayed as a spider chart, maps the age distribution across various happiness score categories, highlighting the connection between age and perceived happiness. The chart categorizes happiness from "5 to <10 Extremely dissatisfied" to "31 to <36 Extremely satisfied," each segment color-coded for clarity. The overall average age across these categories is 38.6 years, with individual group averages ranging from 36.38 to 40.96 years. Notably, the "26 to <31 Satisfied" category is the most frequent, comprising 58 entries which account for 27% of the total dataset. This visualization underscores significant trends and distributions, offering insights into how different age groups perceive satisfaction and informing further research or policies aimed at enhancing life satisfaction across age demographics.

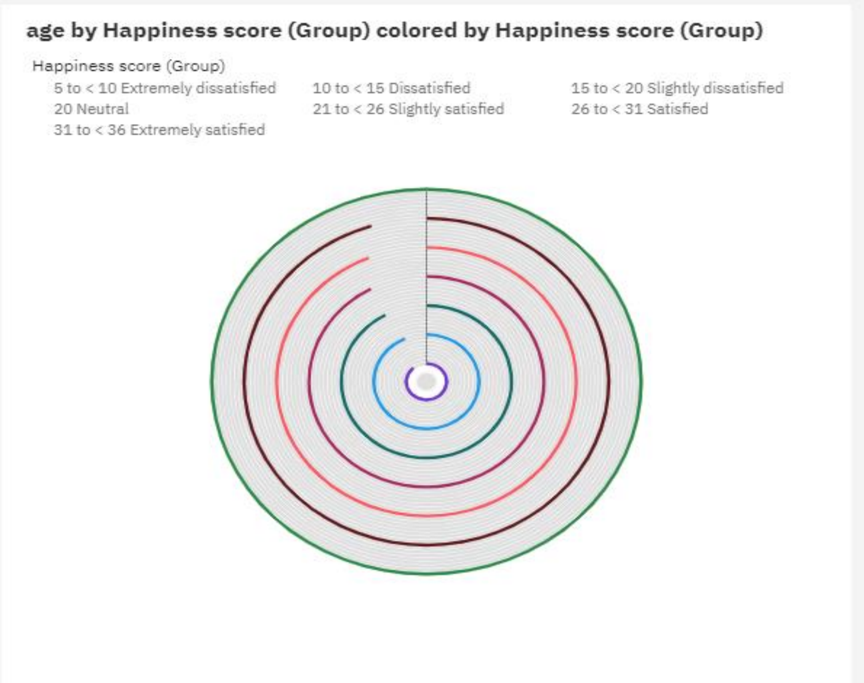

Figure 6: Spider in happiness score

Figure 7, the over all values of Happiness score (Group) and Happiness score (Group), the average of age is 38.6. The average values of age range from 36.38 to 40.96. 26 to < 31 Satisfied is the most frequently occurring category of Happiness score (Group) with a count of 58 items with age values (27 % of the total). Users are equipped with varied levels of knowledge and experience in the areas of data visualization and decision-making under time constraints. It might be challenging to make a value judgment regarding the effectiveness of a data visualization technique. Because of this, a number of different visualization methods and the applications that go along with them have been developed.

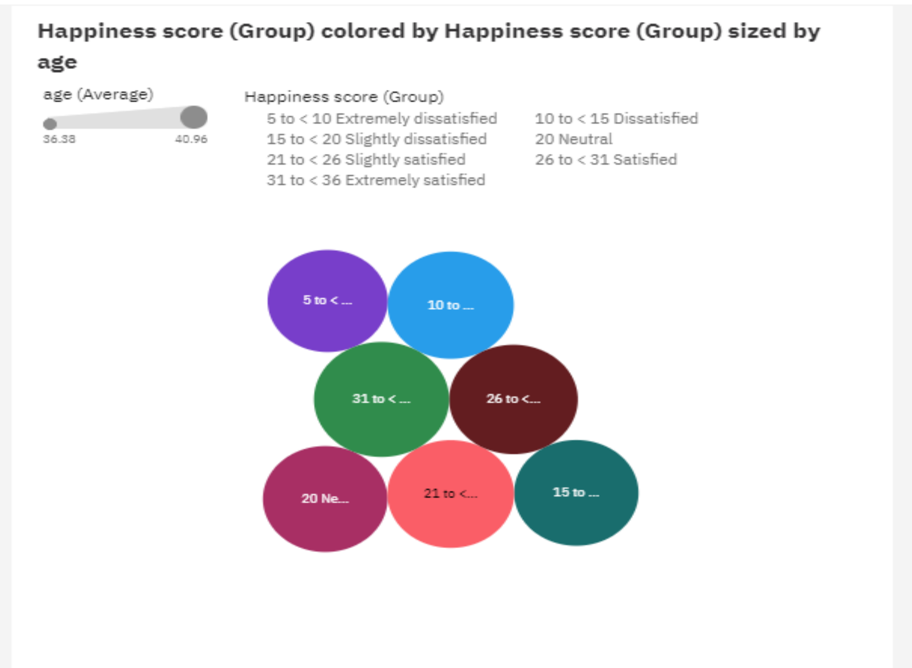

Figure 7: Happiness score (Group) colored by Happiness score (Group) sized by age.

Big data, in both its structured and its unstructured forms, poses a unique set of challenges when it comes to the creation of visualizations. This is due to the fact that we need to take into consideration the volume, pace, and variations of the big data [21][22]. The presentation and analysis of large amounts of data are currently being inhibited by a new set of difficulties related to efficiency, operability, and the degree of differentiation. It takes a lot of time and effort to create a substantial modelled data set. Making a decision regarding the most effective graphic to use might sometimes be tough [23][24].

3. Results

Most visualization designs are created to aid with cognition and decision-making. The intended use of a visualization model must be considered when developing and planning it. In addition to just presenting numerical data, data visualization requires choosing and analyzing the data points that form its basis. Data visualization is one of the most significant topics in computer science and has various applications. A number of application-specific tools have been developed to analyze specific datasets across different fields of medicine and science. In genetics, data visualization can be used to make sense of large amounts of genetic data and spot abnormalities that could point to a genetic problem or a propensity to disease. Researchers can learn more about the genetic mechanisms underlying diseases and create more treatments that are precise by visualizing data. Because of its usefulness in helping investigators spot abnormalities and patterns in massive data sets, data visualization is an indispensable tool in the fight against fraud. It is useful for monitoring financial transactions, validating identities, analyzing links, and managing cases. Investigators of fraud can save businesses money by acting swiftly after a thorough examination of massive datasets made possible by data visualization. The process of presenting data in a manner that is understandable and effective by making use of visuals or pictures is referred to as data visualization. It has evolved into a powerful instrument that is utilized by a huge number of people for the purpose of decoding and analyzing massive amounts of complex data. It is now possible to communicate one's thoughts in a form that is simple, rapid, and accessible to all individuals. It must express complex ideas in a way that is understandable, accurate, and convincing.

Driver analysis: Driver analysis visualization presents you with a graphical representation of the most important drivers, also known as predictors, for a target. The motorist is seen to be in a stronger position the further to the right they are. IBM® Congo's Analytics is able to generate highly interpretable insights that are based on complicated modeling because it makes use of highly advanced algorithms. You do not need to be familiar with the statistical tests that should be performed on your data. The appropriate tests for the data are selected by Congo's Analytics. The driver analysis visualization that can be found in dashboards and explorations makes the key drivers for continuous and categorical targets accessible to users.

In driver analysis, also known as key driver analysis, importance analysis, and relative importance analysis, the data collected from questions like these are used to determine the relative importance of each of the predictor variables in terms of their ability to forecast the outcome variable. Other names for driver analysis include importance analysis and relative importance analysis. One of the most prevalent names for each of the predictors is the term "driver." The purpose of this exercise is to assign a numerical value to the significance of each of the drivers. In other words, the purpose of this exercise is to calculate importance ratings so that we can identify which drivers are most important. The term "importance weights" can also be used interchangeably with the term "importance scores". The most important deliverable that comes out of a driver analysis is often a table or chart that illustrates the relative significance of the various drivers (predictors), such as the graphic that can be found below. Driver analysis is a subfield of data science that, in contrast to other areas of data science, places more of an emphasis on determining the relative significance of the variables that are used to predict outcomes.

The Hilton example places an emphasis on comprehending how various areas of performance are responsible for generating customer happiness. The other main application of driver analysis is to determine how different brand connections, such as whether a brand is Hip, Humorous, or Honest, drive performance. This is one of the most important applications of driver analysis.Driver analysis is not the appropriate method to use if our objective is to generate quantitative forecasts, such as the amount of sales for the following month. This theory does not come up with such hypotheses. In addition, this method is only appropriate for calculating relevance in situations in which the predictors are either dimensions or components of performance, such as in the case of the Hilton. It is inappropriate to use when demographic or behavioral parameters are involved. Driver analysis is not the appropriate method to use if our objective is to generate quantitative forecasts, such as the amount of sales for the following month. This theory does not come up with such hypotheses.

In addition, this method is only appropriate for calculating relevance in situations in which the predictors are either dimensions or components of performance, such as in the case of the Hilton. It is not applicable in situations in which demographic or behavioral predictors are available. The tests of statistical significance. According to the results of this driver analysis visualization, the factors that contribute the most to the target airport rating are the combination of overall satisfaction, signage rating, security rating, and art rating. To make changes to the key drivers or add new ones, click the more icons that are located on the target data slot. Because there are so many rows in the data source, the analysis is only based on a sample that is supposed to be typical of the whole thing. This is done to optimize speed.The phq_score data point's drivers are highlighted in a pie chart when you mouse over it, thanks to the driver analysis visualization.

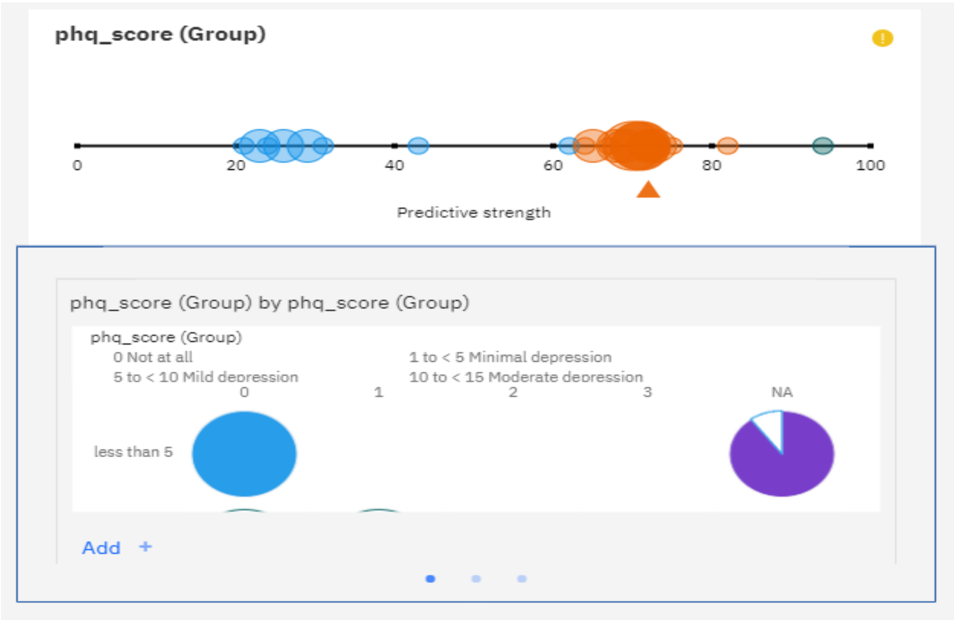

Figure8: Driver analysis in phq_score

Figure 9: Driver analysis in happiness score (group)

The tree's data chart link to complementary visualizations when clicked. It shown the figure9. IBM Congo's Analytics uses a statistical model for analysis of two additional data fields to assess the accuracy with which it can forecast the target field's values. The search through the many predictor pairings is typically not thorough, and several high-ranking pairs may be eliminated from the final findings because of the filtering process. To present an overview as well as a variety of predictor pairs that improve upon the prediction strength of single predictor models that are displayed as one-way drivers is the objective of this project. Therefore, the user learns important information on the pairings of fields in the data, and the insights derived through one-way drivers are enlarged. The results of a one-way driver analysis as well as a two-way driver analysis are presented in the spiral charts and driver analysis respectively. By selecting the appropriate chart viewing option, you can see each of them in their own distinct window. Directly from the Driver analysis visualization found in Explore, each one-way or two-way driver that is currently being displayed has the potential to be developed into a new visualization.

The analysis of each two-way driver is based on a statistical model that incorporates both the target and a pair of categorical predictors. This model is then used to conduct the analysis. Following the completion of data preparation and construction of all one-way drives, the modeling process can begin. The drivers are used to select the first predictor in the pair, and the drivers are used to select the second predictor in the pair. This search approach makes certain that the majority of the highest-ranking predictor pairs will be taken into consideration for the modeling process. The chi-square test of independence is used for numerical targets, while the two-way analysis of variance (ANOVA) is utilized for categorical targets. Additionally, the chi-square adjustment for sparse data is implemented for the chi-square test.

The limitations imposed on the selection of data fields and data rows for the one-way drivers are imposed on the two-way drivers. This is to be expected given that the prospective predictor fields for two-way drivers are selected from the best one-way drivers depending on the level of predictive power possessed by each of those one-way drivers. However, the model significance of a one-way driver and the minimum predictive strength are not necessities for the admission of that driver into a two-way model. One of the requirements for a successful two-way driver is that its predictive strength must be greater than 10% and that it must offer a relative improvement of more than 10% over the predictive strength of each of the contained one-way drivers. When calculating relative improvement, the percentage of the difference between 100% and the predictive strength of the layered one-way driver is used as the base. Those resulting two-way drivers that are able to meet these criteria are rated according to the predictive power of their models, and only the top 20 are made visible to users. For each evaluated pair of fields, a hypothesis test is carried out to determine whether the fields together have a substantial influence on the target. As potential two-way drivers, we only consider those pairs of variables that not only pass the test but also have a high enough level of predictive strength.

The availability of these features has made data visualization useful in a wide variety of academic fields. It is possible that different sorts of data or different kinds of purposes call for the use of different visualization approaches that are more successful. For instance, scatterplots are good for analyzing the relationships between continuous variables, whereas bar charts are more appropriate for contrasting the values of the continuous variables. The usefulness of data visualizations can be significantly impacted by the design and aesthetics of the visualizations themselves. The ease with which individuals are able to grasp and interpret visualizations can be influenced by a variety of factors, including the colors used, the arrangement, and the labeling.

Directions for the future of research with data visualization.

The following is a selection of in-depth findings:

- The visual representation of uncertainty is a vital component of data visualization that is frequently disregarded by people doing the analysis. The constraints of the data and the possibility of making mistakes in the conclusions formed from them can be made clearer to viewers using uncertainty visualizations.
- Gaining deeper insights and a better comprehension of complex data can be accomplished using interactive data visualization tools. Users are given the ability to modify and examine data in real time using interactive tools, which can show patterns and relationships that could be overlooked by using static visualizations.
- The technique of data visualization that one decides to use can have an effect on how well the data is communicated. For instance, the representation of temporal data or multivariate data could benefit from

the application of particular visualization techniques more than others.

- The application of data visualization in instructional settings may result in enhanced student learning results.
  Students can have an easier time grasping abstract ideas and their interrelationships with visualization, which makes the information more clear and accessible.
- The application of data visualization techniques can result in an improved level of involvement and interest in matters that are driven by data. Data may be brought to life and made more meaningful to viewers with visualizations that are both interactive and interesting.

The driver analysis and predicting the psychology depression analysis people will be there. It will be affect the fatigue on the varierious reasons and predicting the depressions. Driver analysis on the analysis of happiness score, will analysis, analyzing, predicting the peoples will affect, and enjoying the depression and happiness score. Will life will be balancing.

4. Conclustion

The patient health questionnaire is a crucial tool for healthcare providers and researchers, offering modules tailored to various mental disorders for a thorough evaluation of patients' mental health. This tool, along with the subjective happiness scale and the satisfaction with work scale, helps assess individual well-being and job satisfaction, enhancing targeted interventions. Additionally, data visualization techniques like bar charts, spider diagrams, and bubble charts play a vital role in decoding complex datasets, supporting decision-making, and spotting trends. These tools are pivotal in understanding mental health and job satisfaction, aiding educators, researchers, and healthcare professionals in improving well-being and facilitating informed decisions. As data visualization evolves, it promises more effective data interpretation, enhancing research, education, and practical applications in various fields.

Refrences

[1] Diener, Randy J. Larsen Ed / Griffin, Sharon, "Satisfaction with Life Scale (SWLS)," the Journal of Personality Assessment, 1985.

[2] U. Gupta, B. Valarmathi, K. Santhi and A. Yahya, "`Classification of patients data of predict signs of diabetic retinopathy," Botswana International University of Science and Technology (BIUST), 2019.

[3] Stpete_ishii, "PHQ-9 Depression Predict and Visualize Importance," kaggle, 2023.

[4] Sebastian Burchert, A. Kerber, J. Zimmermann and C. Knaevelsrud, "14-day smartphone ambulatory assessment of depression symptoms and mood dynamics in a general population sample: comparison with the PHQ-9 depression screening," Openaire, 2019.

[5] Devastator, "PHQ-9 Depression Assessment, 14-Days of Ambulatory Mood Dynamics in a General Population," https://www.kaggle.com/datasets/thedevastator/phq-9-depression-assessment, 2023.

[6] PHQ, "PATIENT HEALTH QUESTIONNAIRE (PHQ-9)," PHQ-9, 2023.

[7] H. Cai, W. Bai, H. Liu, X. Chen, H. Qi, R. Liu, T. Cheung, Z. Su, J. Lin, Y.-l. Tang, T. Jackson, Q. Zhang and Y.-T. Xiang, "Network analysis of depressive and anxiety symptoms in adolescents during the later stage of the COVID-19 pandemic," Translational Psychiatry, vol. 12, March 2022.

[8] Aroucha, "PHQ-9 Depression Clustering," 2023.

[9] S. Joseph, C. A. Lewis and C. Olsen, "Convergent validity of the Depression-Happiness Scale with measures of depression," Journal of Clinical Psychology, vol. 52, p. 551–554, September 1996.

[10] D. Lester and S. Kaplan, "The Depression-Happiness Scale: Happiness is Not Hypomania," Psychological Reports, vol. 74, p. 858–858, June 1994.

[11] Nandy, "Indicators of Anxiety or Depression Based on Reported Frequency of Symptoms during Last 7 Days," kaggle, 2022.

[12] S. Joseph, P. A. Linley, J. Harwood, C. A. Lewis and P. McCollam, "Rapid assessment of well-being: The Short Depression-Happiness Scale (SDHS)," Psychology and Psychotherapy: Theory, Research and Practice, vol. 77, p. 463–478, December 2004.

[13] E.Ortiz-Ospina and M. Roser, "Happiness and Life Satisfaction," Our World in Data, https://ourworldindata.org/happiness-and-life-satisfaction, 2013.

[14] W. Pavot and E. Diener, "The Satisfaction with Life Scale and the emerging construct of life satisfaction," The Journal of Positive Psychology, vol. 3, p. 137–152, April 2008.

[15] Swls, "Life Satisfaction | SWLS, Satisfaction with Life Scale (SWLS)," greenspacehealth, 2020.

[16] R. E. Hoyt, D. Snider, C. Thompson and S. Mantravadi, "IBM Watson Analytics: Automating Visualization, Descriptive, and Predictive Statistics," JMIR Public Health and Surveillance, vol. 2, p. e157, October 2016.

[17] F. Marcatto, L. Di Blas and D. Ferrante, Dataset for: Preliminary Validation of the Satisfaction with Work Scale, PsychArchives, 2021.

[18] G. Zayaraz., K. Santhi and T.Chellatamilan, "Trade-off Analysis of Crosscuztting Functionalities using Lazy Counting-based SplayTree in Aspect Oriented Programming," Research Journal of Applied Sciences, Engineering and Technology, vol. 9(6), pp. 396-408, 2015, 2015.

[19] K. Santhi, G. Zayaraz and V. Vijayalakshmi, "Resolving Aspect Dependencies for Composition of Aspects," Arabian Journal for Science and Engineering, vol. 40, p. 475–486, November 2014.

[20] K. Santhi and G. Zayaraz, "An Approach Based on Colored Petri net for Analysing and Modelling the Aspects," International Journal of Hybrid Information Technology, vol. 6, p. 25–36, September 2013.

[21] Widodo Budiharto, "big data analatics," Lab Robotics and Intelligent Systems, 2018.

[22] Eswar Reddy, R., Santhi, K. (2024). The Survival Analysis of Mental Fatigue Utilizing the Estimator of Kaplan-Meier and Nelson-Aalen. In: Pareek, P., Gupta, N., Reis, M.J.C.S. (Eds) Cognitive Computing and Cyber Physical Systems. IC4S 2023. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 536. Springer, Cham. https://doi.org/10.1007/978-3-031-48888-7_19.

[23] Eswar Reddy R., Santhi, K. Exploring Mental Fatigue and Burnout in the Workplace: A Survival Analysis Approach. EAI Endorsed Trans Perv Health Tech. 2024. Apr. 10, https://publications.eai.eu/index.php/phat/article/view/5719.

[24] S. Nivetha, B. Valarmathi, K. Santhi and T. Chellatamilan, "Detection of Type 2 Diabetes Using Clustering Methods – Balanced and Imbalanced Pima Indian Extended Dataset," in Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI - 2019), Springer International Publishing, 2020, p. 610–619.

[25] E. Ricci, M. Colasanti, M. Monaro, C. Mazza, A. Cardinale, F. Bosco, G. Mazzoni, C. R. Arnaud, S. Ferracuti, A. D. Domenico and P. Roma, "How to Distinguish Feigned from Genuine Depressive Symptoms: Response Patterns and Content Analysis of the SIMS Affective Disorder Scale," Psychological Injury and Law, September 2023.

[26] R. McGreal and S. Joseph, "The Depression-Happiness Scale," Psychological Reports, vol. 73, p. 1279–1282, December 1993.