The Power of Habits: Investigating the Influence of Life style on Mental Health and Work-Life Balance using Machine Learning Techniques.

Nirmit S Karker, Rutik Yerunkar and Sarika Davare

Abstract— *Accomplishing work_life_balance is essential for individual happiness and productivity. This research analyses the relationships between individual attributes and the Work_Life_Balance score to identify key factors influencing work-life balance and potential areas for improvement. Additionally, the impact of demographic factors on specific attributes and the Work_Life_Balance Score was investigated to understand variations in lifestyle and well-being across different groups. By utilising clustering techniques, common characteristics are explored among individuals with similar lifestyle and well-being profiles, providing insights for targeted work-life balance interventions. A predictive framework is developed to evaluate the Work_Life_Balance Score depending on lifestyle, habits, and behaviours, employing different machine learning algorithms and feature selection techniques for optimal performance. Associations between specific attribute pairs or groups, such as daily stress, daily shouting, and weekly meditation, BMI range, daily steps, and fruits and vegetable consumption are examined. Moreover, the impact of habits like personal passions or meditation on mental health-related attributes is investigated, focusing on daily stress and shouting. The role of social factors, including core circle size, social network size, and supporting others, is assessed in predicting overall work_life_balance and happiness. This data exploration enhances the understanding of work_life_balance by identifying influential factors, demographic disparities, lifestyle patterns, and social influences. The results provide valuable perspectives for entities and individuals striving to ameliorate Work_life_balance and boost overall contentment.*

Nirmit S karkera
Electronics and Computer Science
Fr. Conceicao Rodrigues College of Engineering Mumbai, India
crce.9034.ecs@gmail.com
8452894418

Rutik Yerunkar
Electronics and Computer Science
Fr. Conceicao Rodrigues College of Engineering Mumbai, India
crce.8851.ecs@gmail.com
7718891515

Sarika Davare
Artificial Intelligence and Data Science
Fr. Conceicao Rodrigues College of Engineering Mumbai, India
sarika.davare@fragnel.edu.in
8454027922

Keywords Clustering Analysis • Lifestyle Factors • Machine Learning • Mental Health • Predictive Models • Work Life Balance

1. INTRODUCTION

Work_life_balance is a pivotal facet of contemporary society, impacting personal welfare, career contentment, and overall effectiveness. Striving for an ideal equilibrium between professional obligations and personal existence has incrementally grown more demanding in the swiftly evolving and demanding work milieu of today. As corporations and individuals acknowledge the crucialness of work_life_balance, research in this realm has garnered notable recognition.

This inquiry seeks to delve into the complex dynamics of work_life_balance by scrutinizing the connections among distinct traits, demographic elements, lifestyle profiles, and societal impacts. Through comprehending these connections, valuable perspectives can be acquired to enrich work_life_balance and overall happiness.

The principal objective of this research is to scrutinize the connections between individual traits and the Work_Life_Balance Score, a metric devised to gauge the degree of Work_Life_Balance encountered by individuals. By inspecting diverse characteristics, including personal routines, behaviours, and lifestyle preferences, our objective is to pinpoint the aspects that have the strongest effect on the overall work_life_balance score. This examination will offer valuable perspectives into the particular areas where individuals might encounter challenges with work-life equilibrium and spotlight potential regions for enhancement.

Moreover, the influence of demographic variables, encompassing GENDER and AGE, on particular traits and the Work_Life_Balance Score, is under scrutiny. Comprehending how demographic traits impact Work_Life_Balance is essential for detecting potential discrepancies in lifestyles, well-being, and Work_Life_Balance encounters among different cohorts. This investigation will clarify the diverse requirements and hurdles encountered by individuals from varying demographic elements.

Moreover, this research will utilize clustering methods to discern patterns or groups of individuals with akin lifestyle and well-being outlines. By investigating the shared attributes of these clusters, encompassing demographic aspects, routines, and conduct, our aim is to acquire perspectives into the effectiveness of interventions for enhancing work-life balance in particular groups.

Additionally, a predictive model will be developed to assess a person's Work_Life_Balance Score in accordance with their style of living, routines, and conduct. Through employing machine learning algorithms and techniques for feature choice, our objective is to design a model that precisely foretells a person's Work_Life_Balance based on their particular traits. This predictive model can serve as a crucial instrument for individuals, corporations, and policymakers to evaluate and ameliorate work-life equilibrium.

By scrutinizing the connections between particular attribute pairs or clusters, such as stress levels, daily routines, and health-related conduct, this study will provide a comprehensive understanding of the interrelated character of diverse characteristics that have a role in work_life_balance. Furthermore, the impact of particular practices, such as the time dedicated to personal interests or mindfulness, on attributes associated with mental health will be examined to investigate the potential contribution of these practices to promoting enhanced work-life balance and overall health.

In conclusion, this data investigation will evaluate the significance of social elements, such as the size of one's inner circle, the extent of their social network, and their assistance to others, in foretelling overall work-life balance and contentment. Social connections and support systems have a significant influence on individuals' potential to handle professional and personal life adequately. Understanding the influence of social factors will offer valuable perspectives into the importance of fostering supportive environments for achieving work-life balance.

2. RELATED WORK

In the current rapid-paced and competitive world, individuals seek to attain not only professional success but also a sense of contentment and fulfilment in their personal lives. The idea of work_life_balance has gained significant attention as a critical factor influencing overall wellbeing. Defined as the equilibrium between the demands of professional and personal life, attaining a harmonious work_life_balance has been affiliated with improved physical and emotional health, enhanced job satisfaction, and increased life satisfaction [1]. It reflects the ability to effectively manage daily stressors and dedicate ample time to activities that foster wellbeing and happiness [2]. Various dimensions assist in the evaluation of work_life_balance and overall well-being. Among these dimensions are positive emotions, social connections, physical health, professional achievements, and a clear life vision [3]. Positive emotions play a pivotal role in individuals' ability to thrive both in their professional and personal domains. Experiencing frequent positive affect has been linked to greater resilience, creativity, and success in various life domains [3]. Furthermore, social connections and support systems have been In recognized as essential factors in promoting wellbeing, with evidence showing that happiness can spread within social networks [4][5].

Physical health and regular exercise have also been identified as key determinants of wellbeing, with physical activity offering substantial mental health benefits, such as stress reduction and improved mood [6]. Moreover, mindfulness meditation practices have gained prominence for their beneficial effects on psychical stress and overall happiness [7][8].

the circumstances of work_life_balance, studies have highlighted the significance of the job demands-resources model in understanding the interplay between job-related stressors and individual wellbeing [9]. Additionally, work-to-family conflict has been studied extensively, with researchers investigating the consequences of such conflict on various aspects of individuals' lives [10].

Beyond the individual level, societal factors such as socioeconomic status and neighbourhood conditions can significantly influence wellbeing. The prevalence of overweight and obesity has emerged as a major public health concern in developing countries, where lifestyle choices, physical activity, and diet intersect with socioeconomic factors [11]. Time management has also been recognized as a critical moderating factor in the relationship between stressors and employee strain, influencing individuals' capacity to cope effectively with work-related pressures [12].

Recognizing the importance of social support, researchers have explored the buffering hypothesis, which indicates that societal help can alleviate the adverse consequences of tension on health and wellbeing [13]. Additionally, a life-span perspective on social support underscores the importance of perceived and received support in promoting health and resilience across different life stages [14].

Evaluating individuals' subjective happiness and life satisfaction has been essential in understanding the holistic experience of wellbeing. Instruments like the Satisfaction with Life Scale (SWLS) have provided important understanding into the factors that promote individuals' overall life satisfaction [15][16]. Furthermore, a complementary approach focusing on mental health as flourishing has emerged, emphasising the promotion and protection of mental wellbeing to enhance overall national mental health [17].

Sleep duration and quality have been identified as crucial determinants of wellbeing, with inadequate sleep linked to various physical and psychological problems [18]. Studies exploring the U-shaped relationship between age and wellbeing have highlighted unique challenges and opportunities individuals may face at different life stages [19].

Furthermore, a growing body of literature emphasises the significance of personal strivings and resources in understanding subjective well-being, recognizing the interplay between individual characteristics, values, and overall life satisfaction [20]. Additionally, research has explored how neighbourhood deprivation can differentially impact individuals' wellbeing, highlighting the need for targeted interventions to address socioeconomic disparities [21].

It is evident that wellbeing is a multidimensional construct influenced by an intricate interplay of various factors, including work-life balance, positive emotions, social connections, physical health, lifestyle choices, and age-related considerations. This research aims to explore the associations between these dimensions, shedding light on the factors that significantly contribute to individuals' overall wellbeing and providing insights for the formulation of comprehensive strategies to enhance work_life_balance and promote overall contentment of life.

By drawing upon the extensive body of literature from esteemed researchers in the areas of psychological, sociological, and public health, this study seeks to lay out an extensive comprehension of the intricate relationships among work_life_balance and happiness. By means of this exploration, we aspire to promote the growing body of knowledge that can inform policies, interventions, and practices aimed at fostering a harmonious and fulfilling life for individuals across various life domains.

3. METHODOLOGY

A. Dataset used:

We have used the Kaggle dataset "Lifestyle_and_Wellbeing_Data" . This Data set consists of 24 attributes and consists of 15,977 survey feedback describing how people live their lives. The survey was carried out by Authentic-Happiness.com, 360living.co, guidebienetre.org containing 22 questions regarding various aspects of professional and personal life. To calculate The Work-Life Balance Score they considered 5 dimensions:

1. A body in good health, indicating your healthy habits and fitness.
2. A sound mind, pointing out your acceptance of optimistic emotions.
3. Skillfulness, showcasing your capability to cultivate unique achievements through continuous growth.
4. Association, gauging the robustness of your social circle and your inquisitiveness to explore the globe.
5. Meaning, assessing your empathy, benevolence, and the degree to which you are experiencing the existence of your aspirations.

So it is the sum of all five dimensions: a low score falls under 550, a good score surpasses 680, an exceptional score exceeds 700 and the Work_Life_Balance Score is dependent on current research issued in the 360 Living guide.

B. Data Pre-Processing:

1. Firstly, we took the age attribute which consisted of 4 age groups. So we performed categorical data encoding to transform categorical variables into a set of binary variables.
2. Secondly, we performed label encoding on gender attributes where male is labelled 0 and female is labelled 1.
3. Thirdly, we converted Daily stress attribute from object data type into numeric data type.
4. Lastly, we replaced null values in the Daily stress attribute by the columns mean value.

After examining the data set we decided to work on following 7 objectives:

1. Analyse the relationships between individual attributes and the Work_Life_Balance_Score. Determine which attributes have the strongest impact on the overall score and identify potential areas for improvement in work-life balance.

2. Investigate the impact of demographic factors (e.g., age, gender) on specific attributes and the Work_Life_Balance_Score. Identify potential differences in lifestyle and wellbeing factors across demographic groups.

3. Identify patterns or clusters of individuals with similar lifestyle and wellbeing profiles. Explore the common characteristics of these clusters, including demographic factors, habits, and behaviours, and their implications for work-life balance interventions.

4. Develop a predictive model to estimate a person's Work_Life_Balance Score depending on their style of living, habits, and behaviours. Estimate performance of various ML Algorithms and feature selection techniques.

5. Examine the associations between specific attribute pairs or groups, such as the relationship between daily stress, daily shouting, and weekly meditation, or between BMI range, daily steps, and fruits and vegetable consumption.

6. Investigate the impact of specific habits, such as time spent on personal passions or meditation, on mental health-related attributes like daily stress and daily shouting.

7. Assess the role of social factors, such as core circle size, social network size, and supporting others, in predicting overall work-life balance and well-being.

C. Objective 1:

First all the numerical variables in different attributes of the dataset are standardised using sklearn StandardScaler to avoid bias during model fitting.

Now to analyse the relationships between individual attributes and work_life_balance score we plotted all of the features using Seaborn's pairplot method by keeping all the variables on x-axis and work life balance score on y-axis and also we calculated correlation coefficients for all the attributes comparing with Work_Life_Balance Score and we found out that Achievement attribute has the highest correlation coefficient and Daily_Stress attribute has the lowest correlation coefficient with Work_Life_Balance Score.

So through this we can infer that if a person has achieved A success in opening a new business, received an award of best employee, increased the profit of his company, won some competition, marathon, success of his kids etc. helps to increase his work_ life_balance score.

Conversely if a person's life is stressful this may be due to several reasons such as Whether at work or home, factors like environmental disturbances (noise, pollution, insecurity), harassment from co-workers or superiors, and life-altering occurrences like separation, employment termination, grave ailment, or the loss of family and friends leads to decrease in work life balance score.

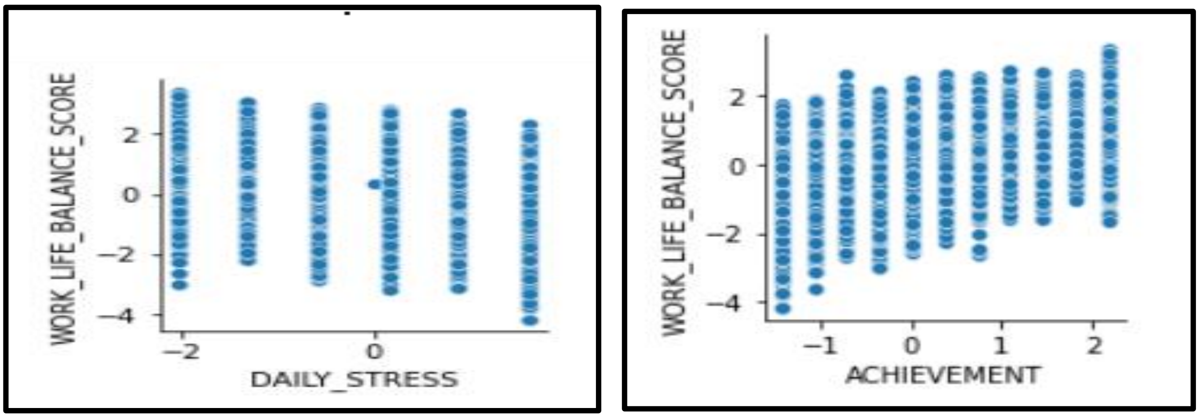


Fig 1: Pair plot of Achievement and Daily_Stress attribute vs work life balance score.

Additionally we also performed Lasso Regularisation feature selection technique to identify most important attributes that contribute to work life balance score and eliminate irrelevant attributes that have no effect on Work_Life_Balance_Score. So LASSO picked 20 variables and eliminated 5 variables.

Selected Variables:

	Feature	Coefficient
0	FRUITS_VEGGIES	0.108836
1	DAILY_STRESS	-0.103183
2	PLACES_VISITED	0.125059
3	CORE_CIRCLE	0.107179
4	SUPPORTING_OTHERS	0.122374
5	SOCIAL_NETWORK	0.116261
6	ACHIEVEMENT	0.104052
7	DONATION	0.139582
8	BMI_RANGE	-0.185303
9	TODO_COMPLETED	0.099158
10	FLOW	0.088858
11	DAILY_STEPS	0.121818
12	LIVE_VISION	0.121815
13	SLEEP_HOURS	0.050224
14	LOST_VACATION	-0.139022
15	DAILY_SHOUTING	-0.100732
16	SUFFICIENT_INCOME	0.167580
17	PERSONAL_AWARDS	0.116501
18	TIME_FOR_PASSION	0.103002
19	WEEKLY_MEDITATION	0.113740

Eliminated Variables:

	Feature	Coefficient
20	GENDER	0.0
21	AGE_21 to 35	-0.0
22	AGE_36 to 50	-0.0
23	AGE_51 or more	0.0
24	AGE_Less than 20	0.0

Figure 2: Lasso Regularisation feature selection

Later we trained Model of Linear Regression to foretell the Work_Life_Balance_Score depending on the selected attributes and we got following model interpretation:

SUFFICIENT_INCOME	0.167859
DONATION	0.139846
PLACES_VISITED	0.125046
SUPPORTING_OTHERS	0.122432
DAILY_STEPS	0.122013
LIVE_VISION	0.121995
PERSONAL_AWARDS	0.116684
SOCIAL_NETWORK	0.116565
WEEKLY_MEDITATION	0.113892
FRUITS_VEGGIES	0.108957
CORE_CIRCLE	0.107231
ACHIEVEMENT	0.104056
TIME_FOR_PASSION	0.103075
TODO_COMPLETED	0.099084
FLOW	0.089042
SLEEP_HOURS	0.050601
DAILY_SHOUTING	-0.101070
DAILY_STRESS	-0.103302
LOST_VACATION	-0.139404
BMI_RANGE	-0.185765

Figure 3: Output of Linear Regression Model

With this we found out that sufficient income and donation attributes have the highest model coefficient and the attributes DAILY_SHOUTING, DAILY_STRESS, LOST_VACATION, BMI_RANGE have lowest model coefficient.

Therefore DAILY_SHOUTING, DAILY_STRESS, LOST_VACATION, BMI_RANGE are the potential areas where a person can work on to improve his Work_Life_Balance_Score.

D. Objective 2:

To investigate the impact of demographic factors (e.g., age, gender) on specific attributes and the Work_Life_Balance Score we utilized a box plot to visual summary of the distribution of work_life_balance scores for individuals in the specified age groups.

In all the box plot negative x-value represents that the corresponding person’s not of that age group and positive x-value represents that the corresponding person is of that age group.

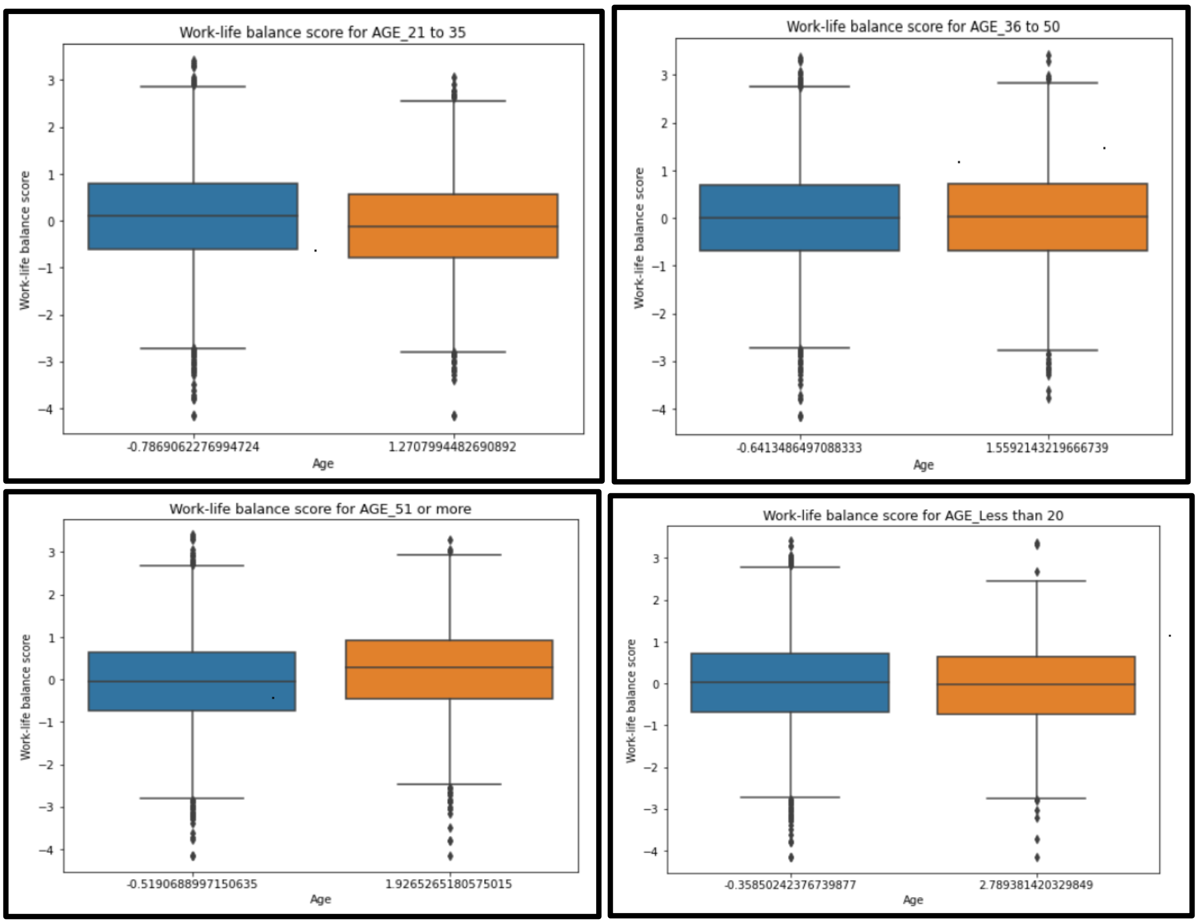


Figure 4: WORK_LIFE_BALANCE_SCORE of different age group

1. Box Plot of AGE_21 to 35 vs WORK_LIFE_BALANCE_SCORE:
The above box plot suggests that individuals in the age group "21 to 35" generally have a moderate work-life balance. However, there are some individuals within this age group who have notably higher work-life balance scores, indicated by the outliers.
2. Box Plot of AGE_36 to 50 vs WORK_LIFE_BALANCE_SCORE:
The above box plot suggests that individuals in the age group "21 to 35" generally have a moderate work-life balance. However, there are some individuals within this age group who have notably higher work-life balance scores, indicated by the outliers.
3. Box Plot of AGE_51 or more vs WORK_LIFE_BALANCE_SCORE:
The above box plot suggests that individuals in the age group "51 or more" generally posses a higher work_life_balance. However, there exists some individuals within this age group who have notably higher work-life balance scores, indicated by the outliers.
4. Box Plot of AGE_Less than 20 vs WORK_LIFE_BALANCE_SCORE:
The above box plot suggests that individuals in the age group "Less than 20 " generally have a moderate work-life balance. However, there are some individuals within this age group who have notably higher work-life balance scores, indicated by the outliers.

From this comparison, we can see that the age group "51 or more" has the highest median work-life balance score, indicating that, on average, individuals in this age group tend to have the best work-life balance. The age groups "36 to 50" and "21 to 35" have the same median work_life_balance score, suggesting that the work_life_balance for individuals in these age groups is similar. Finally, the age group "Less than 20" has a slightly lower median work_life_balance score in comparison to the other three age groups.

To validate the above box plot we also performed a statistical test called, ANOVA and found the following result:

F-statistic: 103.59919847548277

P-value: 2.01816869054209e-66

Based on the ANOVA results, it appears that there is a dissimilarity in work_life_balance scores across different age groups in the dataset. The F-statistic indicates that the variability in work_life_balance scores is significantly influenced by the differences between the age groups.

Further, to know which specific age groups are significantly distinguishable from one another, we conducted post-hoc tests (Tukey's HSD test) and conducted pairwise comparisons between the age groups. These tests identify which age groups have significantly different work-life balance scores.

group1	group2	meandiff	p-adj	lower	upper	reject
AGE_21 to 35	AGE_36 to 50	6.1774	0.0	3.9483	8.4065	True
AGE_21 to 35	AGE_51 or more	16.6383	0.0	14.1845	19.0921	True
AGE_21 to 35	AGE_Less than 20	3.036	0.0528	-0.0243	6.0963	False
AGE_36 to 50	AGE_51 or more	10.4609	0.0	7.874	13.0478	True
AGE_36 to 50	AGE_Less than 20	-3.1414	0.0529	-6.3094	0.0266	False
AGE_51 or more	AGE_Less than 20	-13.6023	0.0	-16.9322	-10.2723	True

Table 1: Parallel assessment of Means using Tukey HSD, FWER=0.05

In summary, Tukey's HSD test reveals that there is substantial dissimilarity in work_life_balance scores between several pairs of age groups. Specifically, "AGE_21 to 35" and "AGE_36 to 50," "AGE_21 to 35" and "AGE_51 or more," "AGE_36 to 50" and "AGE_51 or more," as well as "AGE_51 or more" and "AGE_Less than 20" exhibit statistically dissimilarity in their work_life_balance scores. However, no significant difference is observed between "AGE_21 to 35" and "AGE_Less than 20," as well as "AGE_36 to 50" and "AGE_Less than 20."

So, we can comment that a person of age group 51 and above is experienced and mature enough to handle their daily work life where as the person of age group less than 20 are not mature enough and they have to learn many things in their life to handle their daily work life work life balance score of this age group will eventually increase as they grow.

5. Box Plot of GENDER vs WORK_LIFE_BALANCE_SCORE:

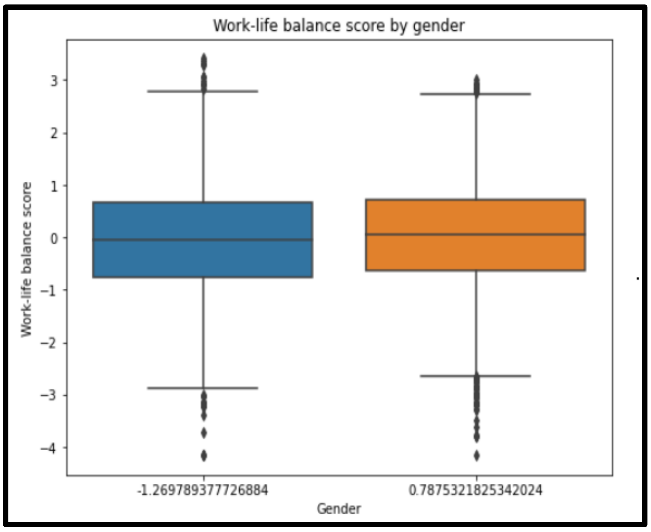


Figure 5: Box Plot of GENDER vs WORK_LIFE_BALANCE SCORE

The above box plot suggests that females generally possess a marginally greater work_life_balance score contrast to males. However, there are some individuals within both genders who have notably higher work_life_balance scores, indicated by the outliers.

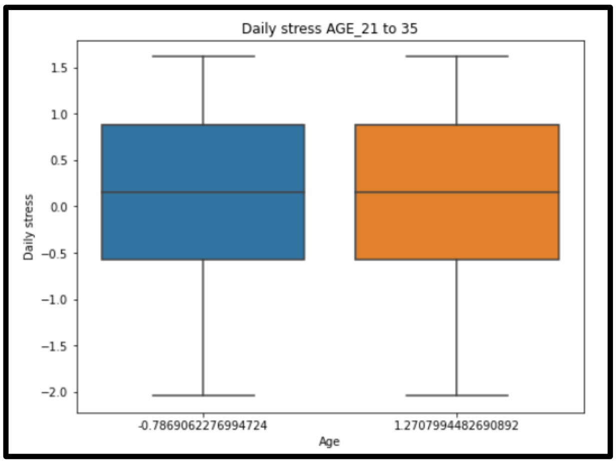
To validate the above box plot we also performed a statistical test called, T-Test and found the following result:

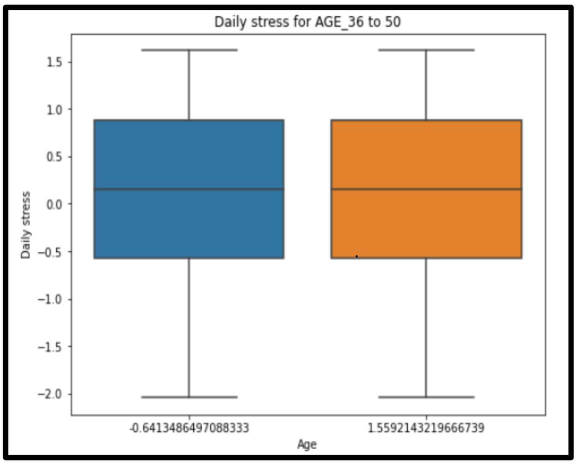
T-statistic: -5.04382204166495

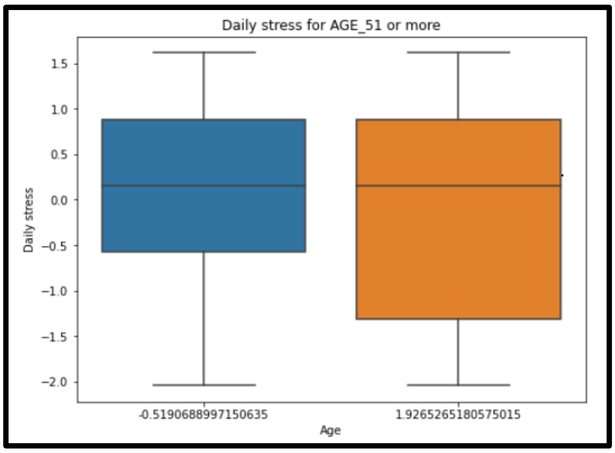
P-value: 4.613270609411358e-07

Based on the t-test results, we noticed a crucial variation in work_life_ balance_score amidst male and female in the dataset. The negative t-statistic indicates that, on average, males have lower work_life_balance scores in comparison with females.

Now to also Identify potential differences in lifestyle and wellbeing factors (DAILY_STRESS) across demographic groups (Age and Gender) we used a box plot.







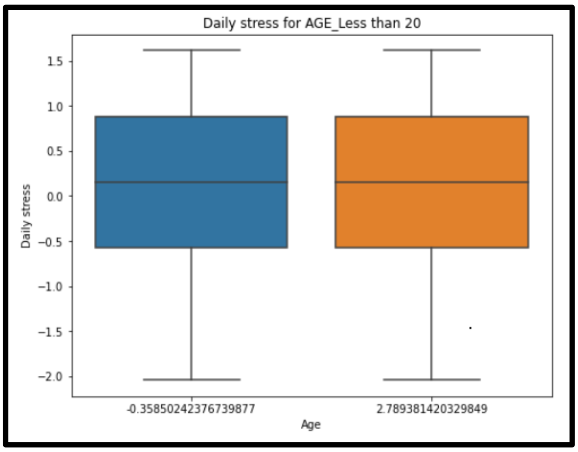


Figure 6: DAILY_STRESS of different age group

1. Box Plot of AGE_21 to 35 vs DAILY_STRESS:

The above box plot suggests that individuals in the age group "21 to 35" generally have a slightly higher daily stress level compared to those not in this age group. However, there are some individuals within both groups who have notably higher daily stress levels, indicated by the outliers.

2. Box Plot of AGE_36 to 50 vs DAILY_STRESS:

The above box plot suggests that individuals in the age group "36 to 50" generally have a slightly higher daily stress level compared to those not in this age group. However, there are some individuals within both groups who have notably higher daily stress levels, indicated by the outliers.

3. Box Plot of AGE_51 or more vs DAILY_STRESS:

The interpretation indicates that, on average, individuals in the age group "51 or more" tend to have a higher daily stress level compared to those not in this age group. However, there are some individuals within both groups who have notably higher daily stress levels, indicated by the outliers.

4. Box Plot of AGE_Less than 20 vs DAILY_STRESS:

The interpretation indicates that, on average, individuals in the age group "Less than 20" tend to have a higher daily stress level compared to those not in this age group. However, there are some individuals within both groups who have notably higher daily stress levels, indicated by the outliers.

From the above results, it appears that the daily stress levels for all four age groups are approximately the same. This conveys that, based on the provided data and the box plots, there is no notable deviation in the median daily stress levels out of these age groups.

5. Box Plot of GENDER vs DAILY_STRESS:

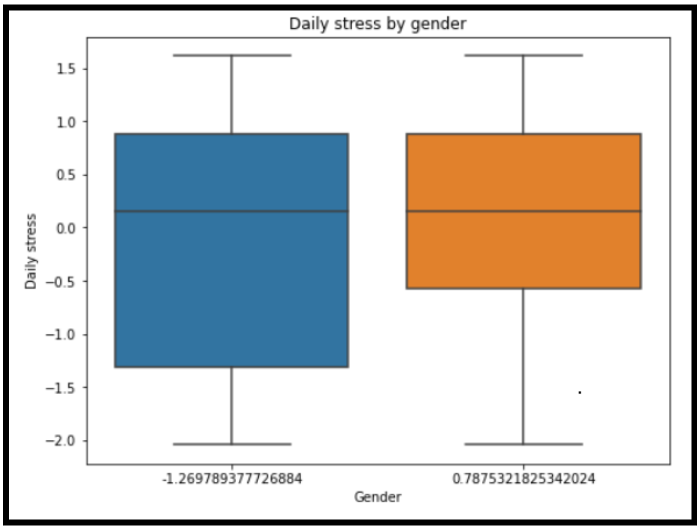


Figure 7: Box Plot of GENDER vs DAILY_STRESS

The above box plot suggests that females tend to have a slightly higher daily stress level compared to males. However, there are some individuals within both genders who have notably higher daily stress levels, indicated by the outliers.

So from the above result we found out that females experience a little higher stress. This may be due to several reasons such as handling family and professional life together.

E. Objective 3:

To identify clusters of individuals with similar lifestyle and wellbeing profiles we used K-Means Clustering technique and then explored the most contributing characteristics of these clusters and finally to visualize each cluster's implications on work-life balance score we used box plots. Before performing K-Means Clustering, we dropped age attributes as we already explored the result of age attributes on Work_Life_Balance Score in previous objective in-depth. To find the optimal number of clusters we applied the elbow technique and through this we found that five would be the most optimal number of clusters. After performing K-Means Clustering we assigned cluster labels to original data. After that we explored common characteristics of each cluster to get information about the most relevant attributes contributing to the formation of each cluster and the results are as followed:

Cluster 4 is formed by containing the following attributes:
['GENDER', 'FRUITS_VEGGIES', 'PLACES_VISITED', 'SLEEP_HOURS']

Cluster 3 is formed by containing the following attributes:
['DAILY_STRESS','LOST_VACATION', 'DAILY_SHOUTING', 'BMI_RANGE', 'SOCIAL_NETWORK', 'SUPPORTING_OTHERS', 'GENDER']

Cluster 0 is formed by containing the following attributes:
['WEEKLY_MEDITATION']

Cluster 1 is formed by containing the following attributes:
['ACHIEVEMENT', 'TIME_FOR_PASSION', 'FLOW', 'SUPPORTING_OTHERS', 'LIVE_VISION', 'PERSONAL_AWARDS', 'CORE_CIRCLE', 'TODO_COMPLETED',

'DONATION', 'SOCIAL_NETWORK', 'PLACES_VISITED', 'WEEKLY_MEDITATION', 'FRUITS_VEGGIES', 'DAILY_STEPS']

Cluster 2 is formed by containing the following attributes:
['DAILY_STRESS']

Finally, we constructed box plots to visualize the association between clusters and Work_Life_Balance Score:

Figure 8: Box plots to visualise the association among clusters and Work_Life_Balance_Score

1. Cluster 0:

- The Work-Life Balance Scores in Cluster 0 are relatively high, as indicated by the median being close to the upper end of the box.

- The interquartile range (IQR) is narrow, suggesting that the majority of individuals in this cluster have similar Work-Life Balance Scores due to weekly meditation.

- There are a few outliers above the upper whisker, indicating some individuals with exceptionally high Work-Life Balance Scores which means that they dedicated more time in weekly meditation which helped them to balance their work life more efficiently.

2. Cluster 1:

- The Work-Life Balance Scores in Cluster 1 show moderate variability, as indicated by the box spanning a broader range compared to Cluster 0.

- The median is positioned around the centre of the box, suggesting a moderate Work-Life Balance Score for the median individual in this cluster.

- There are no outliers beyond the whiskers, indicating the absence of extreme values in the Work_Life_Balance Score for this cluster.

3. Cluster 2:

- Similar to Cluster 0, the Work-Life Balance Scores in Cluster 2 are relatively high, with the median towards the upper end of the box.

- The IQR is narrow, indicating consistency among the Work_Life_Balance Scores for individuals in this cluster.

- There are a couple of outliers above the upper whisker, indicating some individuals with exceptionally high Work_Life_Balance Scores.

4. Cluster 3:

- The Work_Life_Balance Scores in Cluster 3 have the widest range compared to other clusters, as indicated by the broad box.

- The median is positioned around the centre of the box, suggesting a moderate Work_Life_Balance Score for the median individual in this cluster.

- There are a few outliers both below the lower whisker and above the upper whisker, indicating some individuals with both low and high Work_Life_Balance Scores.

5. Cluster 4:

- The Work_Life_Balance Scores in Cluster 4 are relatively low, as indicated by the median being closer to the lower end of the box.

- The IQR is narrow, suggesting that the majority of individuals in this cluster have similar low Work_Life_Balance Scores.

- There are a few outliers below the lower whisker, indicating some individuals with exceptionally low Work_Life_Balance Scores.

Overall, the box plots provide perception into the distribution of Work_Life_Balance Scores across the five clusters. Clusters 0, 2, and 4 have relatively high, consistent, and low Work-Life Balance Scores, respectively. Cluster 1 shows moderate variability in Work-Life Balance Scores. Cluster 3 has the widest range of Work-Life Balance Scores, with individuals spanning both low and high values. These observations can be valuable for understanding how the clusters differ with respect to Work-Life Balance Score and potentially inform strategies for addressing the needs of individuals in each cluster.

F. Objective 4:

In this objective we performed Regression Analysis and built a predictive model to estimate Work_Life_Balance Score of an individual depending upon their style of living, habits, and behaviours. We used various ML models and evaluated their performance based on evaluation metrics.

To train all the models we used the most important attributes identified by Lasso regularisation and evaluated the performance of each model using Cross-Validation Technique by keeping 5 numbers of folds.

	MSE scores for each fold	Average MSE score	MAE scores for each fold	Average MAE score	R-squared scores for each fold	Average R-squared score
Linear Regression	[2.67425265e-09, 2.77605855e-09, 2.53825987e-09, 5.73277579e-06, 3.36017810e-09]	1.1488249080012124e-06	[4.10718586e-05, 4.18653200e-05, 3.98246035e-05, 4.23657620e-05, 4.59245181e-05]	4.2210412449626194e-05	[1, 1, 1, 0.99999416, 1]	0.9999988303597028
XGBOOST Regression	[0.03008532818371399, 0.02968886920488956, 0.02943521260174208, 0.031423132565631215, 0.031880261833983584]	0.030502560877992087	[0.1358849038174457, 0.13470549801047507, 0.1351357128895943, 0.13921616817427163, 0.14049356598893967]	0.13708716977614527	[0.9707670267995739, 0.970557423800495, 0.9714402421189885, 0.9680065185180283, 0.9662423689425443]	0.96940271603592 6
Random Forest Regression	[0.0665032, 0.06442257, 0.06449072, 0.06104377, 0.06679737]	0.0646515254269 0794	[0.20286883, 0.20150211, 0.20210302, 0.19502089, 0.20542493]	0.2013839564998 1983	[0.93538092, 0.93611187, 0.93742735, 0.93784825, 0.92926906]	0.9352074883140584

Table 2: Comparison of various regression techniques

In summary, Model 1 (Linear Regression) is the top-performing model, followed by Model 2 (XGBOOST Regression), and finally, Model 3 (Random Forest Regression). The choice of Model 1 may be preferred for its superior performance in accurately predicting the target variable based on these evaluation metrics.

G. Objective 5:

To examine the associations between specific attribute pairs or groups first we started with Visualizations. We used scatter plots and box plots to see relationships between the attributes. Pairs to visualize relationship among them:

DAILY_STRESS and SLEEP_HOURS: Investigate whether higher stress levels are associated with reduced sleep hours.

DAILY_STEPS and BMI_RANGE: Analyze the association between daily steps and BMI range.

Time_For_Passion and Work_Life_Balance Score: Analyze the association between time spent on personal passions and overall work-life balance scores. GENDER

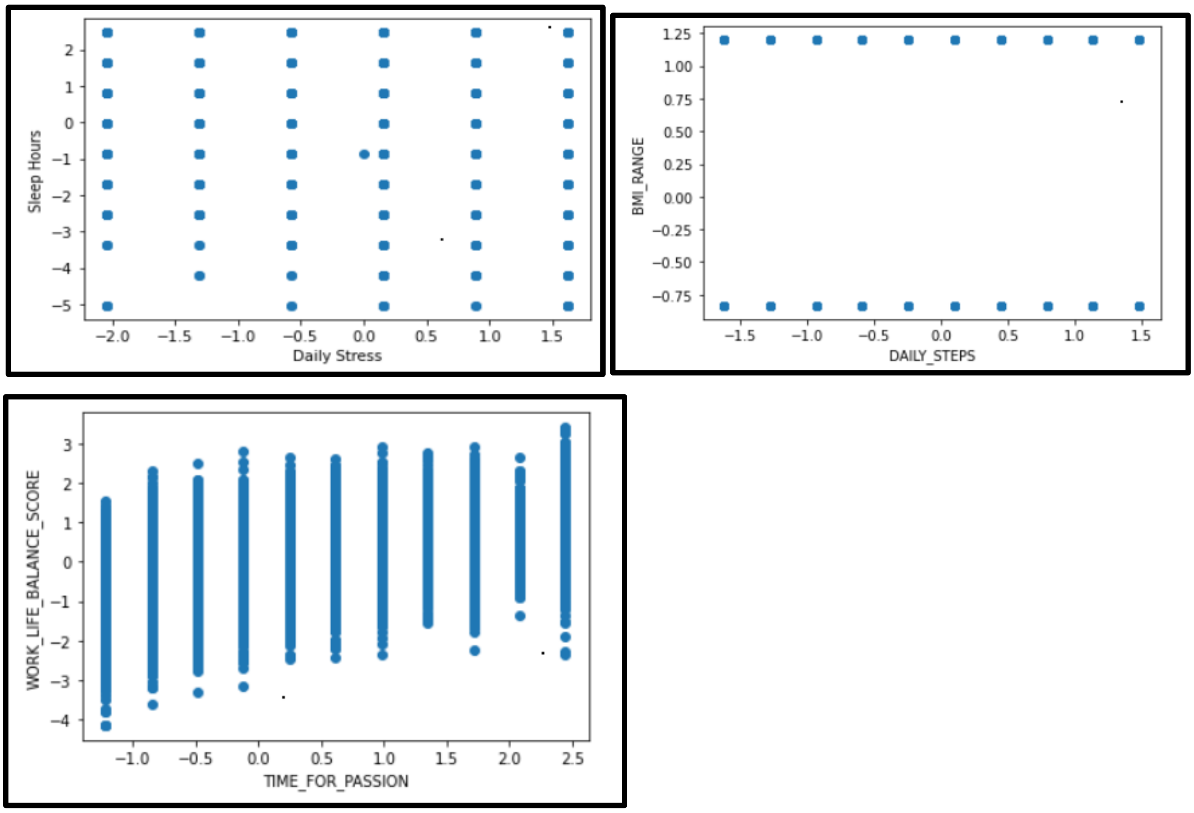


Figure 9: Scatter plots to visualise the association among the attributes

1. Scatter Plot of DAILY_STRESS vs SLEEP_HOURS:

The scatter plot reveals a weak correlation between 'DAILY_STRESS' and 'SLEEP_HOURS', as indicated by the generally declining trend from the top left to the bottom right. It suggests that elevated levels of daily stress are linked with reduced sleep duration, signifying that stress may impact individuals' sleep patterns.

2. Scatter Plot of DAILY_STEPS vs BMI_RANGE:

A weak negative association between 'DAILY_STEPS' and 'BMI_RANGE' is shown in the scatter plot, indicating that elevated levels of daily physical activity might be connected with slightly lower BMI ranges. However, the relationship is not very strong, and other factors likely contribute to individuals' BMI levels.

3. Scatter Plot of Time_For_Passion and Work_Life_Balance_Score:

The scatter plot indicates an optimistic correlation among Time_For_Passion and Work_Life_Balance_Score. Individuals who allocate more time for their passions or hobbies tend to possess a greater work-life balance scores.

H. Objective 6:

To investigate the impact of specific habits, such as time spent on personal passions or meditation, on mental health-related attributes like daily stress and daily shouting we used Stats models Ordinary Least Squares regression (OLS) technique.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           DAILY_STRESS   R-squared:                       0.062
Model:                            OLS   Adj. R-squared:                  0.062
Method:                 Least Squares   F-statistic:                     526.1
Date:                Tue, 28 Mar 2023   Prob (F-statistic):          5.17e-222
Time:                        12:51:26   Log-Likelihood:                -22154.
No. Observations:               15972   AIC:                         4.431e+04
Df Residuals:                   15969   BIC:                         4.434e+04
Df Model:                           2
Covariance Type:            nonrobust
=====================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
Intercept          1.691e-16      0.008   2.21e-14      1.000      -0.015       0.015
TIME_FOR_PASSION     -0.1206      0.008    -15.445      0.000      -0.136      -0.105
WEEKLY_MEDITATION    -0.1955      0.008    -25.034      0.000      -0.211      -0.180
==============================================================================
Omnibus:                      791.626   Durbin-Watson:                   1.900
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              319.112
Skew:                          -0.049   Prob(JB):                     5.08e-70
Kurtosis:                       2.315   Cond. No.                         1.21
==============================================================================
```

Figure 10: OLS regression outcomes for impact of time spent on passion and weekly meditation on Daily Stress

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         DAILY_SHOUTING   R-squared:                       0.016
Model:                            OLS   Adj. R-squared:                  0.015
Method:                 Least Squares   F-statistic:                     125.9
Date:                Tue, 28 Mar 2023   Prob (F-statistic):           5.71e-55
Time:                        12:51:27   Log-Likelihood:                -22538.
No. Observations:               15972   AIC:                         4.508e+04
Df Residuals:                   15969   BIC:                         4.511e+04
Df Model:                           2
Covariance Type:            nonrobust
=====================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
Intercept         -9.107e-18      0.008  -1.16e-15      1.000      -0.015       0.015
TIME_FOR_PASSION     -0.0786      0.008     -9.823      0.000      -0.094      -0.063
WEEKLY_MEDITATION    -0.0828      0.008    -10.346      0.000      -0.098      -0.067
==============================================================================
Omnibus:                     2303.751   Durbin-Watson:                   1.891
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3421.604
Skew:                           1.100   Prob(JB):                         0.00
Kurtosis:                       3.548   Cond. No.                         1.21
==============================================================================
```

Figure 11: OLS regression outcomes for impact of time spent on passion and weekly meditation on Daily Shouting

1. Impact on Daily Stress:

The model analyzing the connection between daily stress and the two habits (TIME_FOR_PASSION and WEEKLY_MEDITATION) has an R-squared of 0.062. This R-squared value specifies that only 6.2% of the variation in daily stress can be interpreted by the two habits incorporated into the model. The coefficients of the predictors are as follows:

TIME_FOR_PASSION: The coefficient is -0.1206, which suggests that a rise in time spent on personal passions is related with a reduction in daily stress. However, the effect size is relatively small.
WEEKLY_MEDITATION: The coefficient is -0.1955, indicating that engaging in weekly meditation is associated with a larger decrease in daily stress compared to time spent on personal passions.

2. Impact on Daily Shouting:

The model examining the impact of the same two habits (TIME_FOR_PASSION and WEEKLY_MEDITATION) on daily shouting has an even lower R-squared of 0.016. This indicates that only 1.6% of the variance in daily shouting can be interpreted by these habits. The coefficients of the predictors are as follows:
TIME_FOR_PASSION: The coefficient is -0.0786, suggesting that a rise in time spent on personal passions is related with a reduction in daily shouting. However, similar to the daily stress model, the effect size is relatively small.
WEEKLY_MEDITATION: The coefficient is -0.0828, indicating that engaging in weekly meditation is associated with a slightly larger decrease in daily shouting compared to time spent on personal passions.

Overall, both models indicate that the habits of spending time on personal passions and engaging in weekly meditation have a statistically significant, yet modest, impact on reducing daily stress and daily shouting.

I. Objective 7:

To assess the role of social factors, we selected 3 attributes: core circle size, social network size and supporting others and predicted the work-life balance score. We performed Regression Analysis and built a predictive model.

	MSE score	R-squared score	Core Circle Size Feature Importance Coefficient	Social Network Size Feature Importance Coefficient	Supporting Others Feature Importance Coefficient
Linear Regression	0.5370179784754726	0.4521635144886853	0.31760954	0.19606225	0.38049037
XGBOOST Regression	0.530556391994573	0.45875527299660857	0.236800	0.121626	0.641574
Random Forest Regression	0.5752760342246299	0.4131347302688074	0.298213	0.154636	0.547151

Table 3: Comparison of various regression techniques

Based on the evaluation metrics and feature importance coefficients, we can draw insights on the performance and significance of the regression models in predicting work-life balance scores. Among the three models, the XGBOOST Regression model stands out as the most favourable choice. With the lowest Mean Squared Error (MSE) of 0.5306, it exhibits the highest level of accuracy in its predictions, outperforming both the Linear Regression model (MSE = 0.5370) and the Random Forest model (MSE = 0.5753). Moreover, the XGBOOST model demonstrates the highest R-squared (R2) coefficient of determination at 0.4588, indicating that it explains the greatest amount of variance in the work_life_balance scores in comparison to the other models.

When considering feature importance, all 3 models consistently identify "Supporting others" as the most crucial feature in predicting work_life_balance. "Core circle size" too plays a notable role, while "Social network size" has the least impact on predictions. These findings suggest that fostering a supportive environment and engaging in meaningful connections within a core circle benefit Work_Life_Balance, whereas the sheer size of one's social network may have a relatively minor effect.

4. CONCLUSION AND FUTURE WORK

In conclusion, this research paper utilised the dataset from Kaggle "Lifestyle_and_Wellbeing_Data" to scrutinize the crucial role of work_life_balance in individuals' overall wellbeing. The study explored various aspects of lifestyle, habits, and social factors to understand their influence on work_life_balance and life satisfaction. Through data pre-processing and exploratory data analysis, significant relationships were discovered between certain attributes and the Work_Life_Balance Score. Achievement attribute has the highest correlation, while Daily Stress has the lowest correlation with the Work_Life_Balance Score. Demographic factors like gender and age were also analysed, revealing that individuals aged 51 and above have a propensity to possess the best work_life_balance, moreover females generally show a marginally higher work_life_balance contrast to males. Clustering techniques unveiled five distinct clusters with varying lifestyle and wellbeing profiles. These clusters provided valuable insights into the different characteristics that contribute to work-life balance in each group. Predictive models were developed using machine learning algorithms to estimate the Work-Life Balance Score based on selected attributes. Linear Regression emerged as the top-performing model in accurately predicting the score. Furthermore, the research explored associations between specific attribute pairs, revealing interesting correlations between daily stress and sleep hours, daily steps and BMI range, and time spent on personal passions and work_life_ balance. The effects of habits on mental health-related attributes, such as daily stress and daily shouting, was also investigated. Engaging in weekly meditation and spending time on personal passions were found to have a modest yet significant effect in reducing daily stress and shouting. Finally, the role of social factors was assessed, and "Supporting others" emerged as the Preeminent influential component in predicting work_life_balance. In conclusion, this research emphasises the significance of work_life_balance in overall wellbeing and provides valuable insights for people and institutions to enhance work_life integration and foster healthier, more fulfilling lives.

References

[1] Greenhaus, J. H., & Powell, G. N. When work and family are allies: A theory of work-family enrichment. Academy of Management Review, 31(1), (2006). 72-92. (https://doi.org/10.5465/amr.2006.19379625)

[2] Diener, E., & Seligman, M. E. Beyond money: Toward an economy of well-being. Psychological Science in the Public Interest, 5(1), (2004). 1-31. (https://doi.org/10.1111/j.0963-7214.2004.00501001.x)

[3] Lyubomirsky, S., King, L., & Diener, E. The benefits of frequent positive affect: Does happiness lead to success? Psychological Bulletin, 131(6), (2005). 803-855. (https://doi.org/10.1037/0033-2909.131.6.803)

[4] Fowler, J. H., & Christakis, N. A. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. British Medical Journal, 337, (2008). a2338. (https://doi.org/10.1136/bmj.a2338)

[5] Cacioppo, J. T., & Patrick, B. Loneliness: Human nature and the need for social connection. W. W. Norton & Company. (2008).

[6] Penedo, F. J., & Dahn, J. R. Exercise and well-being: A review of mental and physical health benefits associated with physical activity. Current Opinion in Psychiatry, 18(2), (2005). 189-193. (https://doi.org/10.1097/00001504-200503000-00013)

[7] Goyal, M., Singh, S., Sibinga, E. M., Gould, N. F., Rowland-Seymour, A., Sharma, R., ... & Haythornthwaite, J. A. Meditation programs for psychological stress and well-being: A systematic review and meta-analysis. JAMA Internal Medicine, 174(3), (2014). 357-368. (https://doi.org/10.1001/jamainternmed.2013.13018)

[8] Hölzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R., & Ott, U. How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. Perspectives on Psychological Science, 6(6), (2011). 537-559. (https://doi.org/10.1177/1745691611419671)

[9] Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. The job demands-resources model of burnout. Journal of Applied Psychology, 86(3), (2001). 499-512. (https://doi.org/10.1037/0021-9010.86.3.499)

[10] Allen, T. D., Herst, D. E., Bruck, C. S., & Sutton, M. Consequences associated with work-to-family conflict: A review and agenda for future research. Journal of Occupational Health Psychology, 5(2), (2000). 278-308. (https://doi.org/10.1037/1076-8998.5.2.278)

[11] Bhurosy, T., & Jeewon, R. Overweight and obesity epidemic in developing countries: A problem with diet, physical activity, or socioeconomic status? The Scientific World Journal, 2014. (2014). (https://doi.org/10.1155/2014/964236)

[12] Geurts, S. A., & Sonnentag, S. Time management as a moderator of relations between stressors and employee strain. Work & Stress, 20(3), (2006). 255-272. (https://doi.org/10.1080/02678370601065893)

[13] Cohen, S., & Wills, T. A. Stress, social support, and the buffering hypothesis. Psychological Bulletin, 98(2), (1985). 310-357. (https://doi.org/10.1037/0033-2909.98.2.310)

[14] Uchino, B. N. Understanding the links between social support and physical health: A life-span perspective with emphasis on the separability of perceived and received support. Perspectives on Psychological Science, 4(3), (2009). 236-255. (https://doi.org/10.1111/j.1745-6924.2009.01122.x)

[15] Lyubomirsky, S., & Lepper, H. S. A measure of subjective happiness: Preliminary reliability and construct validation. Social Indicators Research, 46(2), (1999). 137-155. (https://doi.org/10.1023/A:1006824100041)

[16] Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. The Satisfaction With Life Scale. Journal of Personality Assessment, 49(1), (1985). 71-75. (https://doi.org/10.1207/s15327752jpa4901_13)

[17] Keyes, C. L. Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. American Psychologist, 62(2), (2007). 95-108. (https://doi.org/10.1037/0003-066X.62.2.95)

[18] Grandner, M. A., Patel, N. P., Gehrman, P. R., Perlis, M. L., Pack, A. I., & Pigeon, W. R. Problems associated with short sleep: Bridging the gap between laboratory and epidemiological studies. Sleep Medicine Reviews, 14(4), (2010). 239-247. (https://doi.org/10.1016/j.smrv.2009.08.001)

[19] Blanchflower, D. G., & Oswald, A. J. Is well-being U-shaped over the life cycle? Social Science & Medicine, 66(8), (2008). 1733-1749. (https://doi.org/10.1016/j.socscimed.2008.01.030)

[20] Diener, E., & Fujita, F. Resources, personal strivings, and subjective well-being: A nomothetic and idiographic approach. Journal of Personality and Social Psychology, 68(5), 926-935. (1995). (https://doi.org/10.1037/0022-3514.68.5.926)

[21] Stafford, M., & Marmot, M. Neighbourhood deprivation and health: Does it affect us all equally? International Journal of Epidemiology, 32(3), (2003). 357-366. (https://doi.org/10.1093/ije/dyg084)

“I, the Corresponding Author, declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

I would like to draw the attention of the Editor to the following publications of one or more of us that refer to aspects of the manuscript presently being submitted. Where relevant copies of such publications are attached. [CAN BE DELETED IF NOT RELEVANT]

I can confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. I further confirm that the order of authors listed in the manuscript has been approved by all of us.

I understand that the Corresponding Author is the sole contact for the Editorial process and is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

Nirmit S Karkera

Signed by the Corresponding Author on behalf of the all other authors”