**TOPIC – EDA on Work-Life Balance**

**Business Problem:**

The concept of work-life balance implies that work is bad and life is good; it suggests that work and life are two dichotomous entities that need to be strictly separated and kept at an equilibrium; for some, it even insinuates that less work equals more happiness. In its most neutral definition, work is generally seen as a means for sustenance and survival—laboring the land to harvest food, then later on, as a way to produce goods and offer services in exchange for security. The objective of this notebook is to conduct an Exploratory Data Analysis of the survey responses and advance the understanding of work-life balance and its major influencers are.

**Background / History:**

A 2019 study by software company ResuceTime looked at 185 million hours of working time. Here are some key findings, Workers average just 2 hours and 48 minutes of productive device time a day. 21% of working hours are spent on entertainment, news, and social media. 28% of workers start their day before 8:30 a.m. (and 5% begin before 7 a.m.). 26% of work is done outside of regular working hours. We check email and instant messaging, on average, every 6 minutes. 40.1% of our day is spent multitasking with communication tools. The most distracted days of 2018 were November 26 and July 7 (the Mondays after Thanksgiving and the 4th of July).

**Data Explanation:**

The dataset is a CSV file obtained from the website called “authentic-happiness.com”, containing the following columns for the analysis.

1. Timestamp – Timestamp on when the record made
2. Daily\_stress – daily stress level
3. Places\_visited – Number of places visited
4. Supporting\_others – Number of persons supported by him/her.
5. Social\_network – Number of social networks he/she has.
6. Achievement – Number of achievements he/she has.
7. Donation – Donations made by that person
8. BMI\_range – BMI level of the person
9. Daily\_Meditation – How many minutes the person is doing meditation
10. Gender – Gender of the person

**Methods:**

The first step for the project is data import and preparation. For most of the visualizations, we are going to use the seaborn library. Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. We are going to use a point plot to observe the relationship between sleep hours vs BMI range, Daily steps vs BMI range, etc. Also, violin plot to find the relationship between gender and daily stress, age vs daily stress, etc. Based on all the visualization we will come up with conclusions on how the work-life balance is there for a different set of people with different backgrounds and conclude which set of people is having a very good work-life balance.

**Analysis:**

The work-life balance is the most important part for every corporate employee that has surrounded the world, especially for the past two years due to pandemics. The datasets consist of vital data like stress level, daily workout habits, the total number of hours worked for a set of employees from the survey. Then the distribution of the data was examined using the most up-to-date “authentic-happiness.com” website data and its parameters were obtained according to the statistical distributions.

**Conclusion:**

As of now, to conclude we haven’t changed any of the methods for the prediction. We will be plotting every data in one or another form to suit the visualization to obtain the results. From the plotting, we will be concluding what is the most parameter impacting the work-life balance of a corporate employee and also what factor makes them happy. From the result, we can suggest corporates enhance the work-life balance of their employees.

**Assumption:**

Here, we are going to use a dataset from the website called “authentic-happiness.com” which is a famous website to get the work-life balance data. But, the data is a very generic one and the work-life balance may vary for a different set of people because of their culture and environment. Here, we are assuming the culture and environment are the same for all sets of employees and we are removing that factor.

**Limitations:**

As mentioned early, we are assuming that the culture and environment are the same for all sets of employees and we are removing that factor. Because of the assumption, the result may vary and lead to a less accurate visualization and I feel that is a limitation for this project. We can remove that limitation by implementing the same visualization for a different set of people and collating it all together to get a conclusion.

**Challenges:**

In this project, we are considering factors like countries, genders, etc and the results may lead to gender-biased or racism. So, keen observations on each step are required so that the result is accurate. The data which we are going to handle for this project is a CSV file and we don’t the exact accuracy of the data. So, we have to measure the accuracy of the data somehow to make the results realistic. Apart from that, I don’t see any challenges in this project.

**Future Uses / Additional Applications:**

Based on the visualization, we can make the visualization real-time, and also we can add filters to get the result exactly to match the particular set of people. We can deploy the project in a centralized server and get the results in the form of RESTful API to showcase it on different platforms like android, iOS, web, etc. We can give the result to all corporate management to achieve a good work-life balance for their employees.

**Recommendations:**

As per websites like Kaggle and other data science websites, the recommended library for visualizations is matplotlib and seaborn. Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. Visualization is the central part of Seaborn which helps in the exploration and understanding of data.

**Implementation plan:**

We can use data from the above-mentioned sites to extract the data for the analysis. Then will create lots of visualizations to show the parameters affecting work-life balances. Then we will implement the same visualization for a different set of employees by using different datasets. Based on the visualizations, we can publish the results somewhere on the server.

**Ethical Assessment:**

In this project, we are going to use the data which is available for public use from websites. So, there are no ethical issues in handling data. But the ethical issue might raise when we release the results of this project because it's experimental and we don’t know exactly how far the visualization is correct. So, there are some potential threats that the result may mislead the people about their work-life balance and corporate to take decisions on them and employees respectively. Thus, to avoid this kind of issue, we need to test the visualizations thoroughly with all possible parameters.

**References:**

1. [Work-Life Balance Survey EDA | Kaggle](https://www.kaggle.com/ydalat/work-life-balance-survey-eda)
2. [The problem with work-life balance: it's actually a rollercoaster - Ness Labs](https://nesslabs.com/work-life-balance-problem)

**Illustration:**

1. **Average daily stress by age group:**

**Chart, bar chart

Description automatically generated**

1. **Daily stress by gender:**

**Chart

Description automatically generated**

1. **Daily stress by age and gender:**

**Diagram

Description automatically generated with medium confidence**

1. **Daily stress and daily hours of flow:**

**Chart

Description automatically generated**

**Appendix:**

1. Wor-life balance – work-life balance is the state of equilibrium where a person equally prioritizes the demands of one's career and the demands of one's personal life. Some of the common reasons that lead to a poor work-life balance include, Increased responsibilities at work, Working longer hours, Increased responsibilities at home, having children, etc.
2. Seaborn – Seaborn is an amazing data visualization library for statistical graphics plotting in Python. It provides beautiful default styles and color palettes to make statistical plots more attractive. It is built on the top of the matplotlib library and is also closely integrated into the data structures from pandas. In this tutorial, we shall see how to use seaborn to make a variety of plots and how we can use it along with matplotlib to make our plots more attractive.
3. Matplotlib – Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged.[3] SciPy makes use of Matplotlib.
4. Data visualization – Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.
5. Bar chart – A bar chart displays categorical data with rectangular bars whose length or height corresponds to the value of each data point. Bar charts can be visualized using vertical or horizontal bars. Bar charts are best used to compare a single category of data or several. When comparing more than one category of data, the bars can be grouped to created a grouped bar chart.
6. Violin plot – A violin plot is a method of plotting numeric data. It is similar to a box plot, with the addition of a rotated kernel density plot on each side. Violin plots are similar to box plots, except that they also show the probability density of the data at different values, usually smoothed by a kernel density estimator. Typically a violin plot will include all the data that is in a box plot: a marker for the median of the data; a box or marker indicating the interquartile range; and possibly all sample points if the number of samples is not too high.
7. Line chart – A line chart is a graphical representation of an asset's historical price action that connects a series of data points with a continuous line. This is the most basic type of chart used in finance, and it typically only depicts a security's closing prices over time. Line charts can be used for any timeframe, but they most often make use of day-to-day price changes.
8. Pie chart – Pie charts are very widely used in the business world and the mass media. However, they have been criticized, and many experts recommend avoiding them, as research has shown it is difficult to compare different sections of a given pie chart or to compare data across different pie charts. Pie charts can be replaced in most cases by other plots such as the bar chart, box plot, dot plot, etc.

**Questions and Answers:**

1. **What kind of visualizations are required for this project and what's the plan to implement it?**

Visualizations like bar charts, line charts, violin charts are important for this project because these charts are used to compare parameters. The plan for this visualization is to implement the matplotlib library to achieve these visualizations.

1. **What is the use of the seaborn library?**

Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures. Seaborn helps you explore and understand your data. Its plotting functions operate on data frames and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots.

1. **What are the benefits of using the matplotlib library for this project?**

Matplotlib library is simple and easy to grasp for beginners. Easier to use for people who have had prior experience with Matlab or other graph plotting tools. It provides high-quality images and plots in various formats such as png, pdf, etc.

1. **How to integrate it with other applications like the web, android, iOS, and other platforms?**

To integrate with other platforms, we can deploy this project somewhere in the remote server and make the input and output as a JSON format. If we do that in this way, we can connect to any platform because JSON is supported by most of the platforms.

1. **Is there any particular reason for using matplotlib for visualization?**

Matplotlib can fit every use case because it has a lot of graph types, features, and configuration options. This is useful when writing scientific papers. Matplotlib can run on Linux, Windows, Mac OS X, and Sun Solaris

1. **What are the assumptions made for this project?**

The data is a very generic one and the work-life balance may vary for a different set of people because of their culture and environment. Here, we are assuming the culture and environment are the same for all sets of employees and we are removing that factor.

1. **How we can monetize this project to get a “win-win” situation for the developer and end-user?**

We can easily monetize this project by pitching this concept to all corporate giants by explaining the importance of work-life balance and how we can enhance their employee’s work-life balance by utilizing this project. In this way, we can monetize this project.

1. **What are the next steps in this project to add more features?**

As of now, this project data is very generic and factors like geo-locations and cultures haven’t been implemented in this project. We can implement those factors to find the work-life balance of any employees present on this earth.

1. **Do you have any plans to implement any sort of analytics in this project?**

Yes, we are planning to implement analytics in this project so that we can predict if we change the behavior or day-to-day lifestyle of an employee and how the work-life balance is impacting them. This is a cool feature to pitch to all corporates to enhance their employee’s work-life balance.

1. **What is the best data model suits for this project if we are implementing analytics?**

To my knowledge, the regression model will suit this project for implementing analytics. Because a regression analysis is a basis for many types of prediction and for determining the effects on target variables. When you hear about studies on the news that talk about fuel efficiency, the cause of pollution, or the effects of screen time on learning, there is often a regression model being used to support their claims.