# Blood Donation

## Specify Dataset

The dataset we are going to use is from this web page:

<https://www.kaggle.com/datasets/hopesb/student-depression-dataset>

Characteristics:

1. It is composed by 27900 people who have and don’t have depression. We will create a group of 1000 rows who maintain the same proportion of people with depression.
2. Variables: the gender,age, city, profession, academic pressure, work pressure, CGPA, Study satisfaction, Job satisfaction, Sleep duration, dietary habit, degree, if the person had suicidal thoughts, work/study hours, financial stress, family history of mental illness and if the person has or not depression
3. The data comes from a psychology study done to different students.

Context: Is focussed on identifying patterns that can contribute to students depression with the aim to design early intervention strategies.

## 2. Specify the General Idea (Analytics Part, Visual Part):

* **Analytics Part:**
  + Application of t-SNE for dimensionality reduction and cluster formation.
  + Dynamic calculation of correlations between selected variables.
  + Analysis of demographic distribution based on gender and other categories.
* **Visual Part:**
  + Coordination of four visualizations:
    1. **Scatterplot:** Clusters formed by t-SNE.
    2. **Parallel Coordinates Plot:** Multivariable relationships within selected clusters.
    3. **Bar Chart:** Demographic distribution.
    4. **Heatmap:** Dynamic correlations.

## 3. Specify the Intended User:

Psychologists, investigators in mental health, teachers or people who work in education and students who have interest in mental health.

Professionals who try to identify risk patterns among students for creating intervention programs.

## 4. Specify the Used Analytics

**Dimensionality Reduction:** t-SNE to simplify high-dimensional data.

**Correlation Calculations:** Dynamically updated for selected subsets.

**Demographic Analysis:** Generation of metrics and charts based on selected demographic attributes.

## 5. Specify the Various Characteristics in Relation to the Visual Analytics Cycle:

1. Data preparation: Cleaning and transforming the dataset to remove outliers and normalize variables. Also, applying dimensionality reduction with t-SNE.

2.Visualization: Generate an interactive scatterplot to explore data and identify patterns. It will interact with a bar chart, a parallel coordination graph and a heatmap that shows correlation. We are also thinking to create different filters that could adjust visualization

3.Models: analytical techniques to identify patterns using t-SNE. We will also use k-means to validate the agrupation. Building correlations and dynamic calculations for selected subsets.

4.Knowledge:

Insight generation: identify critical factors that are associated with depression symptoms in students.

## 6.Mockup

The next thing we are going to show is beta mockup before cleaning all the data so you can imagine all the type of visuals we can put:

Gráfico

Descripción generada automáticamente