***ONLINE GAMING ANIXIETY DATA***

*THE PROPOSAL FOR MID PROGECT*

**INFORMATION ABOUT GAMING STADY DATA:**

* **ABOUT DATASET:**

Gaming and its association with anxiety, life satisfaction and social phobia

This dataset consists of data collected as a part of a survey among gamers worldwide. The questionnaire asked questions that psychologists generally ask people who are prone to anxiety, social phobia, and less to no life satisfaction. The questionnaire consists of several set of questions as asked as a part of psychological study.

* **DATA INFORMATIONS:**

The data frame has 13464 rows and 55 columns.

1. **THE COLUMNS:**

'S. No.', 'Timestamp', 'GAD1', 'GAD2', 'GAD3', 'GAD4', 'GAD5', 'GAD6',

'GAD7', 'GADE', 'SWL1', 'SWL2', 'SWL3', 'SWL4', 'SWL5', 'Game',

'Platform', 'Hours', 'earnings', 'whyplay', 'League', 'highestleague',

'streams', 'SPIN1', 'SPIN2', 'SPIN3', 'SPIN4', 'SPIN5', 'SPIN6',

'SPIN7', 'SPIN8', 'SPIN9', 'SPIN10', 'SPIN11', 'SPIN12', 'SPIN13',

'SPIN14', 'SPIN15', 'SPIN16', 'SPIN17', 'Narcissism', 'Gender', 'Age',

'Work', 'Degree', 'Birthplace', 'Residence', 'Reference', 'Playstyle',

'accept', 'GAD\_T', 'SWL\_T', 'SPIN\_T', 'Residence\_ISO3',

'Birthplace\_ISO3'

* + - **ABOUT COLUMNS:**

The file consists of 55 columns that correspond to every question asked in the questionnaire. Most of the columns correspond to different scoring criteria used in psychology like GAD (General Anxiety Disorder), SWL (Satisfaction with Life) an and SPIN (Social Phobia Inventory) scores. There are a few generic questions about where the people are born and why do they lay online games.

**EXPLIAN EACH COLUMN SEPARATELY:**

* + 1. **Timestamp** 🡪 Time at which the participant took the questionnaire after it being launched, 0 null values
    2. **GAD (1:7) (general anxiety disorder)** 🡪 Response to GAD questions, 0 null values.
    3. **GADE** 🡪 Affect of gaming in work, 649 null values.
    4. **SWL (1:5) (Satisfaction with life)** 🡪 Response to SWL questions, 0 null values
    5. **Game** 🡪 Name of the game they play 0 null values.
    6. **Platform** 🡪 Mode of game playing (PC, Console, Mobile), 0 null values.
    7. **Hours** 🡪 Number of hours in a week devoted to playing, 30 null values.
    8. **Earnings 🡪** Earning from the game, 0 null values.
    9. **Why play 🡪** Reason to play the game, 0 null values
    10. **League 🡪** league, 1838 null values.
    11. **Highest league 🡪** 100% null values.
    12. **Streams 🡪** Number of online streaming of the game, 100 null values.
    13. **SPIN (1:17) (social phobia) 🡪** Response to SPIN questions, have null values.
    14. **Narcissism 🡪** Interest scale in the game, 23 null values.
* **data cleaning:**

1. Drop the Highest league, accept, Reference, Residence\_iso3 and Birthplace-iso3 columns.
2. THE Hours column: drop or replace with mean for all rows more than 164H
4. Make all the data lowercase and strip the string column.
5. The data taken from survey so the unique values are not the same must deal with it.
6. Replace (unknown, unspecified, etc.) with null values.
7. Deal with null values in the appropriate way:
   * + - Fill it with mean or mode.
       - Drop it.

* **business questions:**

1. Which average of GAD, SWL and SPIN is the highest?
2. What is the most game played by GAD?
3. How many hours do SWL play?
4. What is the most frequency game?
5. Why do SPIN play?