**Name:** Obinna Williams

**Student ID:** VEPH/20B/DA183

**TECHNICAL REPORT FOR YOUTHWELL INTELLIGENCE: YOUTH SMOKING AND SUBSTANCE USE TRENDS**

**Outline**

* **Introduction**
* **Story of Data**
* **Data Splitting and Pre-processing**
* **Pre-Analysis**
* **In-Analysis**
* **Post-Analysis and Insights**
* **Data Visualization & Charts**
* **Recommendations and Observations**
* **Conclusion**
* **References & Appendices**

**Introduction**

1. **Objective of the Project**

The primary goal of this project is to analyse behavioural and psychological trends among youth, particularly focusing on factors such as mental health, smoking prevalence, drug experimentation, media influence, and access to support systems. Using Excel-based analysis, the aim is to generate insights that can inform youth development programs, public health initiatives, and policymaking.

1. **Problem Being Addressed**

Youth in many societies face a range of interconnected challenges including substance abuse, mental health struggles, and inadequate parental or institutional support. This project seeks to answer key questions:

* How are mental health and substance use patterns evolving over time?
* What demographic and social factors influence these behaviours?
* Which areas present the highest risk and need for intervention?

1. **Key Datasets and Methodologies**

**Datasets Used:**

* Mental Health by Year
* Smoking Prevalence by Age
* Drug Experimentation (by status and gender)
* Media Influence
* Access to Counselling
* Parental Supervision
* Youth Master Dataset
* Pre- and In-Analysis Boards
* Visualization Dashboards

**Methodologies in Excel:**

* Pivot Tables for summarizing categorical trends
* VLOOKUP for data merging and enrichment
* Conditional Formatting for pattern recognition
* Charts (bar, line, scatter) for visual representation

**Story of Data**

1. **Data source**

This dataset is sourced from Kaggle.com

1. **Data Collection Process**

The data collection process is by web scraping.

1. **Data Structure**

The dataset is organized in a tabular format, where:

* **Rows:** Represent different observations or aggregated categories (e.g., year, age group, gender).
* **Columns:** Represent features such as mental health scores, drug use status, smoking prevalence, media influence, and access to services.
* **“Mental Health by Year”** captures year-wise averages of mental health and related metrics.
* **“Youth smoking drug data”** serves as the master dataset with comprehensive variables for each respondent or group.
* **“Media Influence on Age Groups”** links external influence with youth behaviour.

1. **Important features and their significance**

* Mental Health: Central variable; gauges overall psychological well-being.
* Smoking Prevalence & Drug Experimentation: Behavioural indicators of risk.
* Socioeconomic Status & Parental Supervision: Environmental and social context.
* Access to Counselling & School Programs: Support mechanisms available to youth.
* Media Influence & Peer Influence: External influences shaping decisions.

1. **Data Limitations or Biases**

*  Missing or Unclear Values: Some sheets have NaN values or non-standard headers.
*  Aggregation Issues: Certain sheets use pivoted formats or aggregated averages, limiting row-level insights.
*  Self-Reported Bias: Data may be affected by underreporting or misreporting due to the sensitive nature of mental health and drug use.
*  Lack of Temporal Detail: Some time-based trends are averaged by year without monthly or quarterly granularity.

**Data Splitting and Preprocessing**

1. **Data Cleaning**

Several sheets contained non-data rows (e.g., labels, subtotals) which were excluded. Missing values were identified across fields such as socioeconomic status, parental supervision, and media influence. Unnecessary columns were removed, and uniform headers were applied for consistency across datasets.

1. **Handling Missing Values**

* There were no missing values.

1. **Data Splitting**

* Dependent **Variable**: Mental Health score, Drug Experimentation, and Smoking Prevalence were used as outcome variables.
* Independent **Variables**: Age group, gender, socioeconomic status, access to counselling, media influence, peer influence, etc.

1. **Industry Context**

The dataset pertains to the public health and youth development sector. It specifically targets the intersection of mental wellness, behavioural science, and social risk factors among youth populations.

1. **Key Stakeholders**

* Schools.
* Government.
* Mental Health Professionals.
* Youth Support Groups

1. **Value to the Industry**

Insights from this analysis can:

* Help design more effective school intervention programs
* Optimize targeted outreach for at-risk age groups
* Improve resource allocation for mental health services
* Influence policy decisions around media regulation, youth education, and family support systems

**Pre-Analysis**

**Potential Analysis:**

1. What age group has the highest smoking prevalence?

2. Is there a correlation between mental health and smoking /drug use?

3. How does socioeconomic status impact smoking and drug experimentation?

4. Does peer influence significantly affect smoking behaviour?

5. Do school programs reduce drug experimentation rates?

6. Is access to counselling associated with better mental health outcomes?

7. Are there gender differences in smoking and mental health trends?

8. How does the media influence relate to drug experimentation or smoking?

9. Do community support systems reduce smoking prevalence?

**Potential Insights**

1. Anti-smoking campaigns should be targeted more effectively by age.

2. Poor mental health is likely to drive higher substance use, so a strong correlation could support mental health programs as a strategy to reduce substance abuse.

3. Youth in underprivileged areas may turn to smoking or drugs as coping mechanisms or due to lack of education/support.

4. High peer influence could justify implementing group counselling approaches.

5. Schools with strong education programs could see a measurable drop in experimentation rates.

6. Youths with counselling access may better manage stress and avoid harmful habits like smoking or drugs.

7. Frequent exposure to smoking or drug use in media may normalize or encourage experimentation.

8. A combination of school programs, family support, and counselling may be the most effective solution

**In-Analysis**

1. **Unconfirmed Insights:**

1. The average mental health scores dropped to the lowest in 2022 (5.41) before a sharp increase to 5.52 in 2023, scores dipped again slightly to 5.47 in 2024.

2. The 0-14 age group has the highest average smoking prevalence at 28.087, which is concerning given their young age.

3. Ages 60-69 show the lowest average smoking prevalence (27.012).

4. Individuals from the low socioeconomic group have the highest average drug experimentation rate at 40.33.

5. Respondents who selected "Both" genders have the highest average drug experimentation at 40.48.

6. Individuals aged 60-69 reported the highest media influence (5.69), followed by 0-14 (5.60)

7. The age group 20-24 has the highest access to counselling (1,045 individuals), followed closely by age group 30-39 (1,021 individuals) and 60-69 (1,020 individuals).

8. Males reported the highest average level of parental supervision at 5.57.

1. Preliminary Recommendations:

1. The low scores in 2021-2022 may reflect lingering pandemic effects, with the 2023 peak possibly showing early signs of recovery or support interventions.

2. The high value for the 0-14 group may suggest early exposure or experimentation, indicating a need for stronger youth focused prevention programs.

3. While the gap is small, individuals in the Low socioeconomic group may face more environmental stressors, leading to slightly high drug experimentation.

4. The small difference between male (39.81) and female (40.15) rates implies that drug experimentation is not strongly gender specific, supporting the need for inclusive education programs.

5. The strong media influence on the 60-69 age group suggests they may rely more heavily on traditional media like the Newspaper, impacting behaviours significantly.

6. Higher counselling access in the 20-24 age group suggests increased mental health awareness, possibly through universities.

7. The slightly higher supervision for males may reflect protective parenting styles possibly suggesting parents monitor boys' behaviour more closely

1. **Analysis Technique used in Excel:**

* **Pivot Tables**: For summarizing metrics by gender, age group, and socioeconomic status.
* **VLOOKUP & INDEX-MATCH**: To cross-reference variables across sheets (e.g., matching counselling access with mental health scores).
* **AVERAGEIFS, COUNTIFS, and IF Statements**: To calculate trends within specific subgroups.
* **Conditional Formatting**: To highlight critical thresholds (e.g., low mental health scores or high-risk behaviour).
* **Data Validation and Filtering**: To isolate target segments for analysis.

**Data Visualization & Charts**

1. **Charts and Graphs:**
2. **Dashboard:**

A close-up of a graph

AI-generated content may be incorrect.

**Recommendations and Observations**

1. **Observations:**

1. Mental health peaked in 2023 with an average score of 5.52, after dipping in 2022.

2. Youth aged 0-14 show the highest smoking prevalence at 28.09.

3. Age 20-24 has the highest access to counselling (1,045).

4. Low socioeconomic status group reports the highest drug experimentation (40.33). This may reflect social pressure.

5. Both gender identifying individuals report the highest drug use average (40.48).

6. Females report the lowest parental supervision. (5.47).

7. Older adults (60-69) are the most media influenced age group (5.69). Older adults are more impacted than expected,

8. Ages 25-29 and 70-79 have the lowest access to counselling, below 1,000.

1. **Recommendations:**

1. Sustain and strengthen mental health programs that helped improve the average scores in 2023.

2. Introduce early smoking prevention programs targeted at ages 0-14.

3. Expand counselling access for ages 25-29 and 70-79, where support appears limited.

4. Design outreach for low-income groups with drug use awareness and support systems.

5. Ensure inclusive mental health and education programs for gender diverse youth.

6. Promote balanced parental engagement across genders, especially for females.

7. Use media channels to target seniors (60+) with accurate health messaging.

8. Monitor supervision and behaviour correlations to strengthen youth protection strategies.

**Conclusion**

1. **Key Learnings:**

* Mental health among youth is steadily declining, with environmental and social factors playing a key role.
* Media exposure and peer influence are strong behavioural drivers, especially in relation to drug use and smoking.
* Counselling access and parental supervision are critical protective factors—both significantly correlate with improved mental health and reduced risk behaviours.
* Demographics matter: Gender, socioeconomic status, and age each influence behavioural outcomes in distinct ways.
* Not all institutional programs are equally effective: School programs need better targeting and evaluation mechanisms to achieve measurable impact.

1. **Limitations:**

* Data Granularity: Many insights are based on yearly or aggregate data; more granular time-series data could enhance pattern detection.
* Self-Reported Bias: The reliance on self-reported measures for sensitive behaviours may introduce bias or underreporting.
* Missing or Inconsistent Data: Some sheets had incomplete records or ambiguous labels, requiring assumptions during analysis.

1. **Future Research:**

* Time-Series Modelling: Integrate monthly or quarterly data to detect seasonality or real-time behavioural shifts.
* Qualitative Analysis: Add interviews or focus group data to contextualize patterns found in quantitative data.
* Advanced Predictive Modelling: Use statistical software or machine learning tools to forecast at-risk individuals or areas.
* Cross-Regional Comparisons: Expand dataset to include more regions or schools to validate trends and customize interventions.

**References & Appendices**

1. **Data Source:**

* Kaggle.com

1. **Tools:**

* Microsoft Excel.
* Pivot Tables and Pivot Charts.