1. Age:

What is the age range of the students in the dataset?

How evenly distributed is the age across the dataset?

Does age play a significant role in predicting language barriers?

1. Gender:

What is the distribution of gender in the dataset?

Is there an imbalance between male and female students?

How does gender relate to language barriers?

1. Background Type:

What is the distribution of students from rural and urban backgrounds?

Are there more students from urban or rural areas?

Does the background type impact language barriers?

1. Home Language:
2. Duration in English environment (in a day (hrs.))
3. 10th Score and graduation Score:

How are the math and science scores distributed?

Is there a correlation between math/science scores and language barriers?

Do students with higher scores face fewer language barriers?

1. Satisfaction Language, Satisfaction Culture, Satisfaction Support:

How satisfied are students with language support, cultural environment, and overall support?

Is there a correlation between satisfaction levels and language barriers?

1. Extracurricular Participation:

How many students participate in extracurricular activities?

Does extracurricular participation relate to language barriers?

1. Social Engagement:

What is the distribution of social engagement scores?

Is there a correlation between social engagement and language barriers?

1. Challenges Culture, Challenges Accents:

What challenges do students face in terms of culture and accents?

Do these challenges impact language barriers?

1. Mental\_Health\_Status, Stress Levels:

How is the mental health status and stress level distributed?

Is there a connection between mental health/stress and language barriers?

1. Pronunciation\_Difficulties, Confidence English:

How many students face pronunciation difficulties?

Is confidence in English related to language barriers?

1. Part\_Time\_Job:

How many students have part-time jobs?

Does having a part-time job impact language barrier?

1. English courses taken:
2. Language Barrier:

What is the distribution of language barriers in the dataset?

How balanced is the target variable?

1. Interview faced (yes, no)
2. Placed or not