ETL Suicides Rates And Mental Health ALL Over The World

The reason that I chose is topic for ETL project is I try to find any relationship between suicides and mental health issues. I couldn't believe why people can give up their lives and choose to leave their families. What is the reason behind it? Anyone help them? Did they seek for any help? Are suicides rates increasing by years?

Extract:
I extracted two data sets from Kaggle.com. One data is about suicides rates overview from 1985 to 2016. The other data's topic is mental health survey. Both data sets are CSV files.

Transform:
For suicides rates data, I filtered data from 2014 to 2015 which is the most current data in the file. I used groupby methods to obtain new suicides rates data sets by countries, years, gender and ages.

For mental health data, I extracted columns that I need to use firstly. I split Timestamp column to extract years and delete date and time values. After that, I created an age range column to filter data by different age groups. Because gender column has mess values such as M, Female or male, I changed gender values to lowercase and replaced all values to be female and male. I also dropped data with values of NaN and Don't know. In addition, there is no number values in the data for calculation use, so I replaced columns values No to 0 and Yes to 1 for calculation sum method use. Finally, I used groupby methods to obtain new mental health data sets by countries, years, gender and ages with sums of each columns.

Load:
I created database in MySQL. I also created new tables for each data frame tables that I created in jupyter notebook. I made them connections. I did inner join methods in MySQL workbench to merge tables.

Final Data Analysis:
The United States had the highest numbers of suicides rates between 2014 to 2015. From the cleaned up mental health data, there were 843 patients that were involved the mental health data. United States also had the highest numbers of mental health patients. Mental health issues are associated with self-employed and remote work. People may have more mental health issues when they lack of communicating with other people.

From 2014 to 2015, the suicides rates were decreasing. The total numbers of self-employed, remote work and family history were much lower in 2015.

From merged suicides and mental health data table, the highest suicides age group was between 35 to 54. But age group between 25 to 35 had more people working as self-employed and having remote jobs. Patients between 25 to 35 were obtained more mental health treatments than other age ranges. However, patients between 35 to 54 looked for more help.

Compared male and female suicides rates and mental health issues, male had much higher suicides rates than female's. More male had self-employed and remote work positions. In addition, the higher numbers of family mental health issues are, the higher suicides rates are.