**Depression vs Obesity correlations**

**Team**

1. Orlando – Team Leader
2. Mery, Jimenez
3. Yuan, Chai
4. Nagaraj (Raj)

**Project Details**

**Introduction:**

Goal of this project is to analyze correlation between Obesity and Depression. We chose 2 data sources which has data about Depression and Obesity. We analyzed the files and noticed the data has the similar data, but there were rows that were not in the same format. So, we did below process to achieve our goal.

**Process**

1. Read files from different sources, we are using 2 CSV Files obtained from 2 different sources.

Sources:

1. <https://data.world/health/california-obesity-2012-2013>
2. <https://data.world/chhs/5a281abf-1730-43b0-b17b-ac6a35db5760>
3. Read the files and pull it in Pandas DataFrame.
4. Rename the columns and order them accordingly in DataFrame.
5. Transform the data:
6. We noticed some of the columns were not the same format.

Example:

1. Age groups were not same between 2 files.
2. Income was grouped differently.
3. Race Ethnicity was not in the right order.
4. Education was not grouped correctly.
5. We combined the rows by adding them or average them and grouped to right age groups.
6. We dropped the old rows using index and inserted new rows.
7. We used the 2 DataFrames and Loaded the data in to MySQL database.

Table name:

1. Depression

create table depression(

id INT AUTO\_INCREMENT NOT NULL primary key,

study\_year int,

strata text,

strata\_type text,

frequency int,

weighted\_frequency float,

percent float,

lower\_95 float,

upper\_95 float

);

1. Obesity.

create table obesity (

id INT AUTO\_INCREMENT NOT NULL primary key,

study\_year int,

age\_group text,

strata text,

strata\_type text,

obesity float

);

1. We then plotted graphs using Matplotlib.

**Conclusion**