ETL- Project

The sources of data that you will **extract** from:

For this project we extracted our data from Kaggle.com. Which provided us with three different csv files to work from with the ranks of the top restaurants from 2020.

CSV Files:

Future 50- Provided us with the data of the fastest growing restaurants based on percentage change from sales in 2018/2019.

Independence 100- The top 100 Independent Restaurants from 2020. (An independent restaurant is defined as a restaurant that has less than five locations). Rankings are based on gross 2019 food and beverage sales.

Top 250- Top 250 restaurants from 2020.

The data above can be found at: <https://www.kaggle.com/michau96/restaurant-business-rankings-2020>

The type of **transformation** needed for this data (cleaning, joining, filtering, aggregating, etc):

Using Jupyter notebook we used pandas to transform our data.

We cleansed the csv files using pandas, dropped the null values, and dropped some extra columns that were unnecessary.

The type of final production database to **load** the data into (relational or non-relational).

We connected to PostgreSQL sqlalchemy and created a python file to hold our password and confirmed all of the date was passed as intended to PostreSQL.

The final tables or collections that will be used in the production database.

Future 50, Independent 100, and Top 250.