

1. Code For Data Clean

File Name	Description
1_EDA.py	Initial EDA python code to transfer Jason file to CSV file. The large Jason file cannot be directly transferred to CSV file because of computer memory limit. This code cut Jason file into chunks and list to perform EDA. Can deal with medium and small size Jason file, cannot work on large size Jason file
2_EDA - simple.py	Initial EDA python code to transfer Jason file to CSV file. The code use directly transfer data form and work for small size of Jason file.
3_initial large jason to csv_review.ipynb	The code used in Kaggle notebook to transfer large size Jason file to CSV file.
4_20220227_clean business.py	Advanced EDA file to clean business data, mainly includes remove null/empty, break dictionary/collection format data into multiple columns to achieve 3NF
5_20220312_delete rows.py	Advanced EDA file to optimize project's dataset. Perform merge join on related data set, remove useless data tuples to improve the database performance.
6_20220324_research business categories.py	Advanced EDA file to calculate the most frequency categories and assign new single string category column to the business dataset.
7_merge for MA_OR review.ipynb	Advanced EDA file to optimize project's dataset. Similar as the code 5, use Kaggle notebook to deal with large size dataset merge join.
8_health_rankings_eda.do	An EDA STATA code file to explore and clean the county health rankings datasets and zip to county crosswalk datasets, including producing exploratory summary statistics, dropping unnecessary attributes, and cleaning/formatting useful attributes.

2. Application Code

File Name	Description
/server	<ul style="list-style-type: none">• .gitignore : A gitignore file for the Node application. Read more on .gitignore files here .• config.json : Holds the RDS connection credentials/information and application• configuration settings (like port and host).• package.json : maintains the project dependency tree; defines project properties, scripts, etc• package-lock.json : saves the exact version of each package in the application dependency tree for installs and maintenance.• routes.js : This is where the code for the API routes' handler functions go.• server.js : The code for the routed HTTP application.
/client	<ul style="list-style-type: none">• .gitignore : A gitignore file for the client application. Read more on .gitignore files here• package.json : maintains the project dependency tree; defines project properties, scripts,etc

	<ul style="list-style-type: none"> • package-lock.json : saves the exact version of each package in the application dependency tree for installs and maintenance
/client/public	<ul style="list-style-type: none"> • This folder contains static files like index.html file and assets like robots.txt for specifying webpage titles, crawlability, et cetera
/client/src	<ul style="list-style-type: none"> • config.json : Holds server connection information (like port and host). • fetcher.js : Contains helper functions that wrap calls to API routes. • index.js : This the main JavaScript entry point to the application and stores the main DOM render call in React. • /pages This folder contains files for React components corresponding to the three pages in the application (see the sections below for more details). These are: <ul style="list-style-type: none"> ○ FriendsPage.js: page used for webpage 2: friend connection function ○ HomePage.js: page used for project homepage ○ RestaurantsRecommender.js: page used for webpage 1: restaurants recommender function ○ ScientistPage.js: page used for webpage 3: business scientist function.
Set up method	Run "npm install" and "npm start" in client folder and server folder correspondingly.