OSEMN Report
Data Wrangling Project
Group 6

OSEMN Process for Working on Extracted Fields from the Yelp Reviews Data Set

Rahul Kumar Nalubandhu and Sandra Estrada

Introduction

Conducting a review analysis gives businesses the opportunity to improve customer experience, identify service gaps and gain real time insights amongst many other benefits. This analysis dives into customer reviews entered in Yelp for businesses in different states throughout the United States.

Problem

The issue with having little to no customer reviews is that this can negatively impact sales. Customer reviews are an important channel to attract customers and increase sales. A benefit in analyzing reviews at the business, city and state level will provide insight into which states and/or businesses have the least customer engagement and allow for proper intervention.

Obtain Data

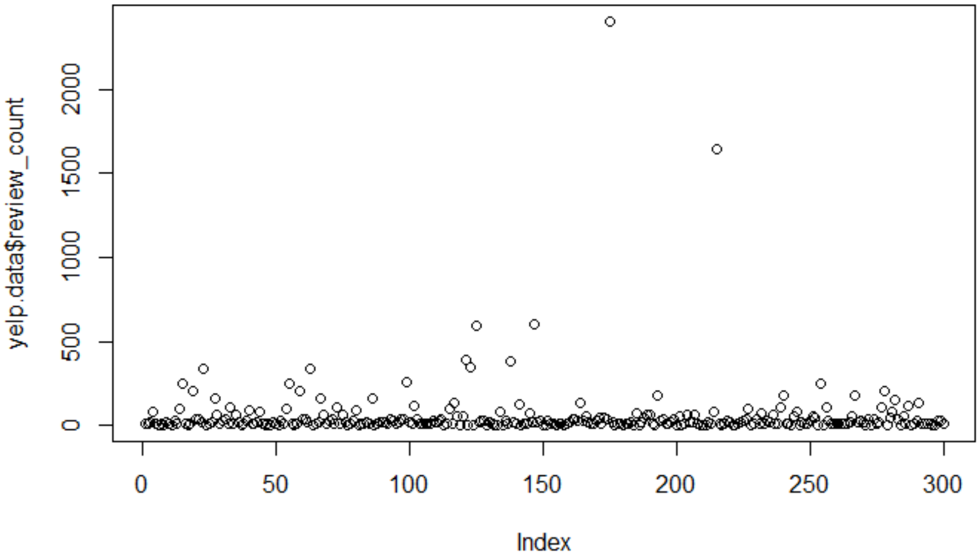
Yelp is a one-stop platform which enables customers to connect with businesses. More than 80 million people visit this platform in a month to find businesses and service providers. Customers are given the ability to leave reviews and request quotes from local businesses amongst many other things. In return, local business owners are given the ability to communicate with their customers and respond to reviews to build trust with their customers. The customer review data set is acquired directly through Yelp. The data set is 4.04GB (1 point) and split into multiple Json files (2 points) which contain businesses, reviews, and user data. In addition, the data has punctuation (1 point) and has more than one type of related data (2 points). Based on the point system requirements provided, the yelp data is a 6-point data set.

Scrub Data

Prior to conducting the analysis for the yelp data set, the data is extracted and consolidated into a csv and Json file for the following fields: business name, city, state, and review count. Visual studio is the primary application utilized to read and extract data in the python programming language. In addition, all null values are deleted from the data set. Null values are deleted to not compromise the integrity of the analysis.

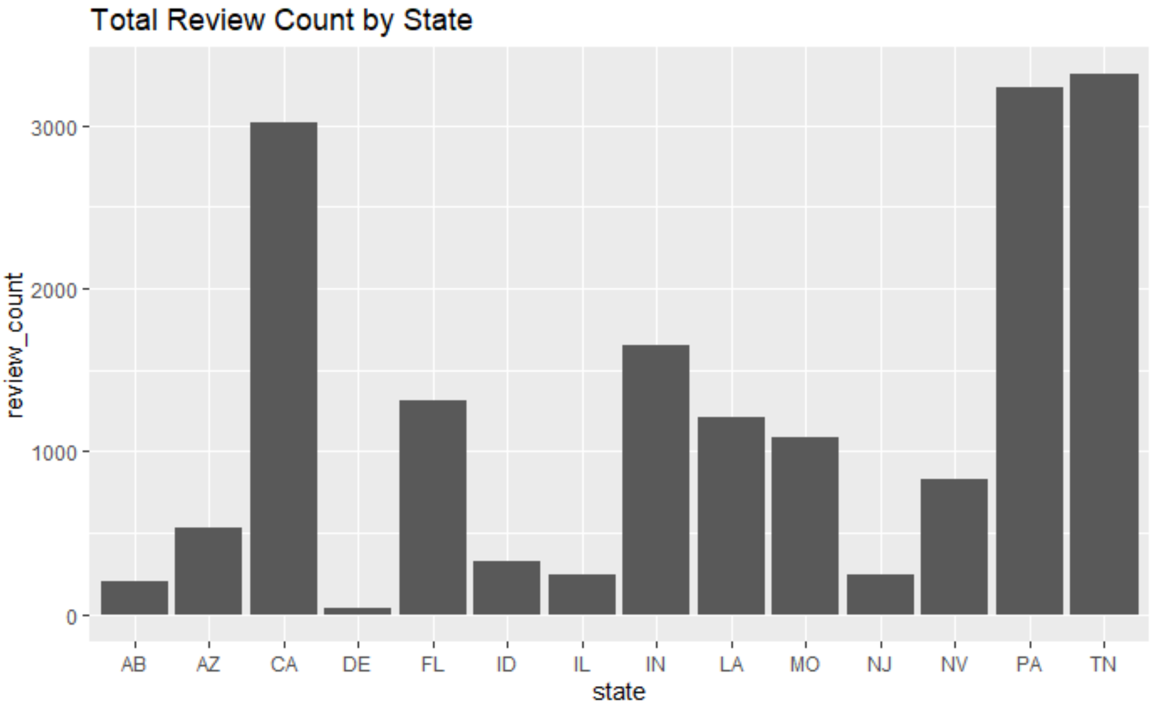
Explore Data

For the exploration stage, the csv data is loaded into R-studio. To begin the exploration stage the review count field is plotted. There are two evident outliers for businesses having more than 1,000 reviews on yelp. Most of the businesses have a review count of less than 500 reviews.

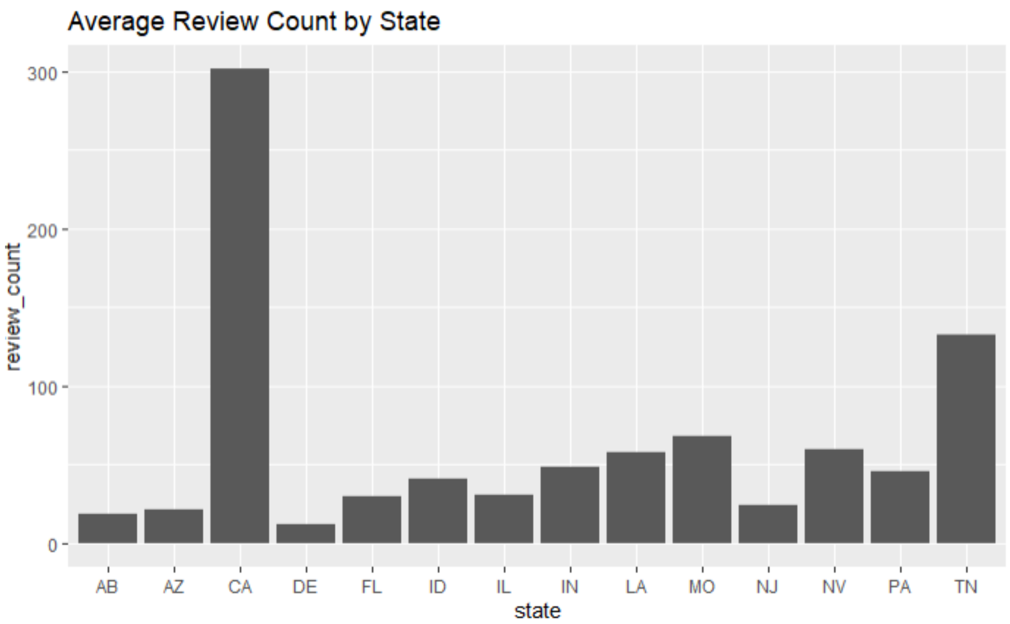


Graph 1

The review count variable is also examined at the state level for the total and average number of reviews. These observations indicate the states with the most reviews are Pennsylvania and Tennessee followed by California. On the opposite end of the spectrum, Delaware is the state with the least total and average number of reviews. On average, the state whose businesses have more reviews compared to other states is California.



Graph 2



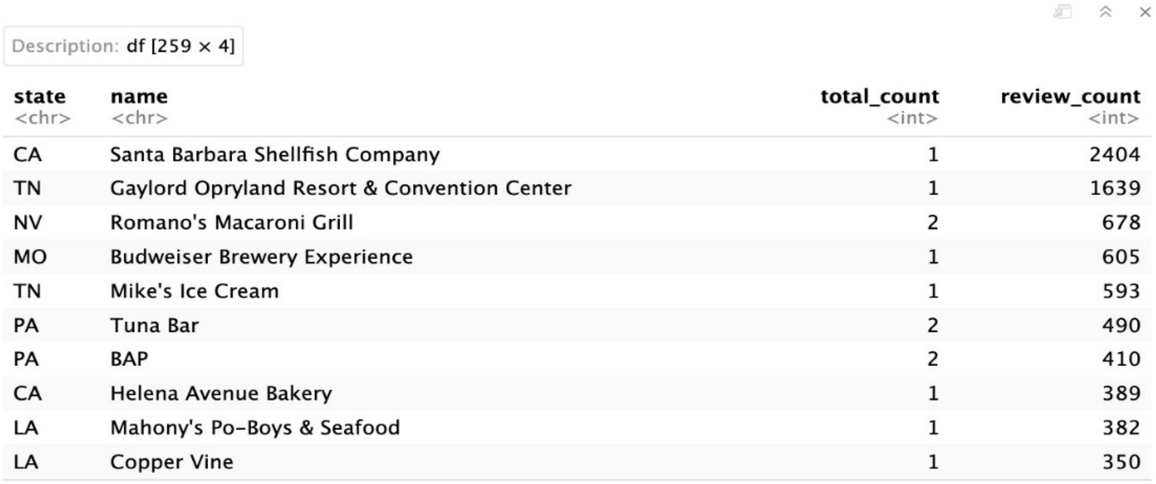
Graph 3

Model Data

After reviewing the dataset, a decision tree is the most suitable modeling technique for this analysis. A decision tree will contain the different factors that determine if businesses are likely to face less customer reviews due to the state they are in or even the demographics of their target customers. While this is deemed the best approach for this analysis, more data needs to be gathered and analyzed to properly complete this.

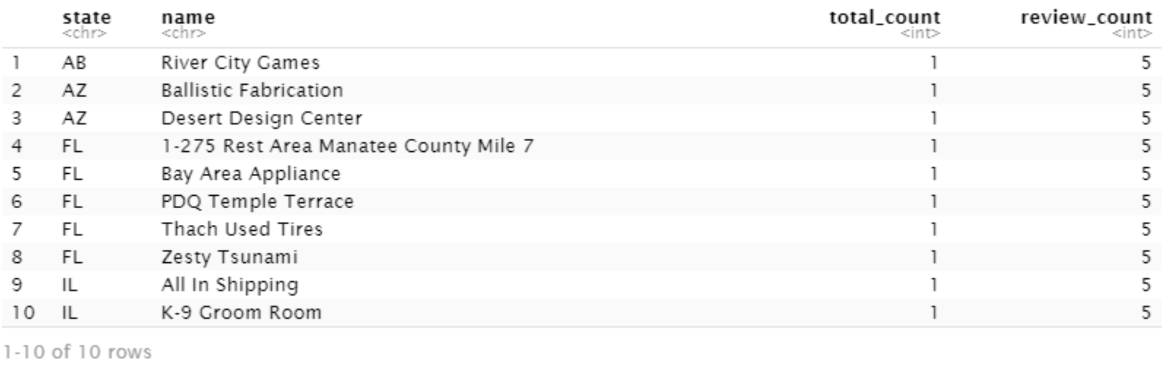
Interpret Data

The business with the most reviews in this data set is Santa Barbara Shellfish Company located in CA with 2,404 reviews. For the companies with the lowest count of reviews, there are a few with only 5 reviews. An interesting observation is that most businesses are in FL. This observation is aligned with the results from the Average Review Count by State plot.



Description: df [259 × 4]

state <chr>	name <chr>	total_count <int>	review_count <int>
CA	Santa Barbara Shellfish Company	1	2404
TN	Gaylord Opryland Resort & Convention Center	1	1639
NV	Romano's Macaroni Grill	2	678
MO	Budweiser Brewery Experience	1	605
TN	Mike's Ice Cream	1	593
PA	Tuna Bar	2	490
PA	BAP	2	410
CA	Helena Avenue Bakery	1	389
LA	Mahony's Po–Boys & Seafood	1	382
LA	Copper Vine	1	350



	state <chr>	name <chr>	total_count <int>	review_count <int>
1	AB	River City Games	1	5
2	AZ	Ballistic Fabrication	1	5
3	AZ	Desert Design Center	1	5
4	FL	1-275 Rest Area Manatee County Mile 7	1	5
5	FL	Bay Area Appliance	1	5
6	FL	PDQ Temple Terrace	1	5
7	FL	Thach Used Tires	1	5
8	FL	Zesty Tsunami	1	5
9	IL	All In Shipping	1	5
10	IL	K-9 Groom Room	1	5

1-10 of 10 rows

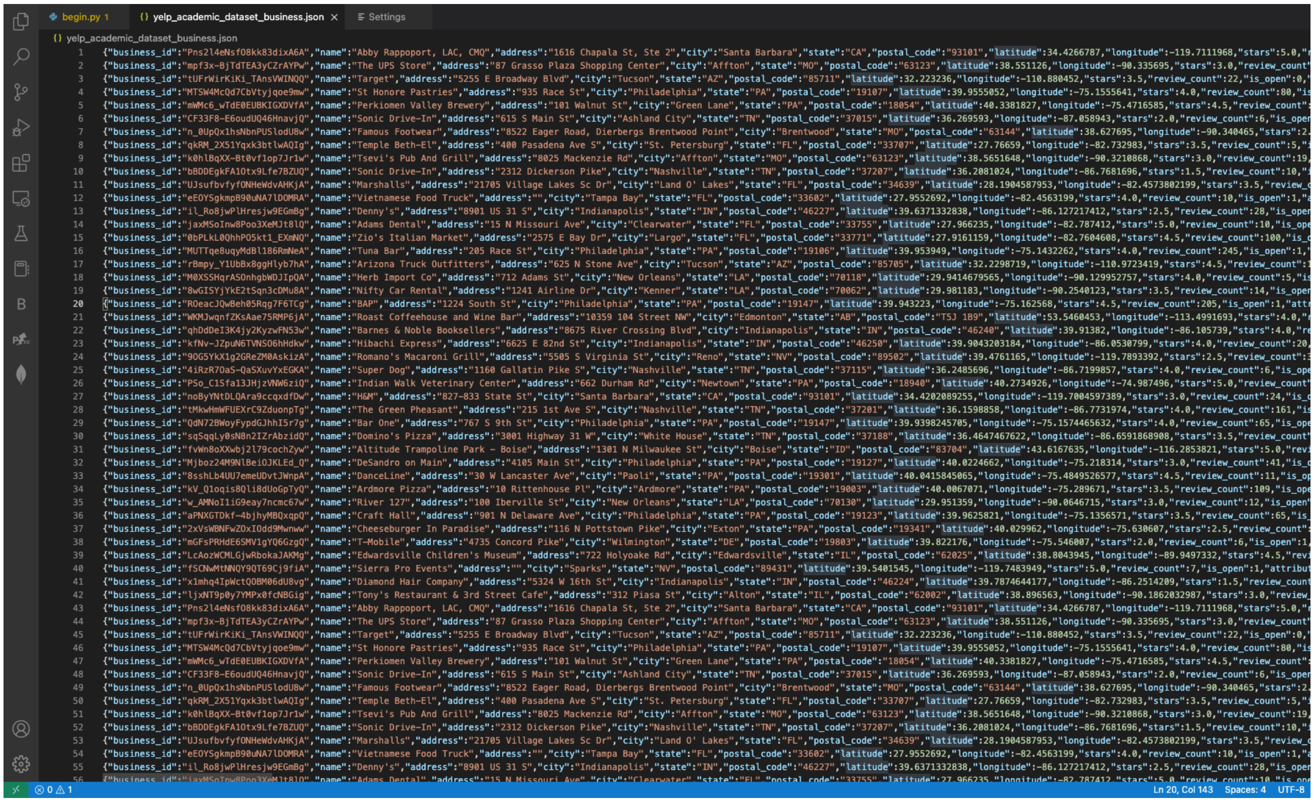
Tasks Completed per Team Member

Rahul Kumar Nalubandhu – Read, extracted data from the original data set, removed null values, converted the Json data into a csv format file, and assisted in portion of the r analysis.

Sandra Estrada – Converted the csv format into a Json format file, completed the r analysis, and wrote the OSEMN report.

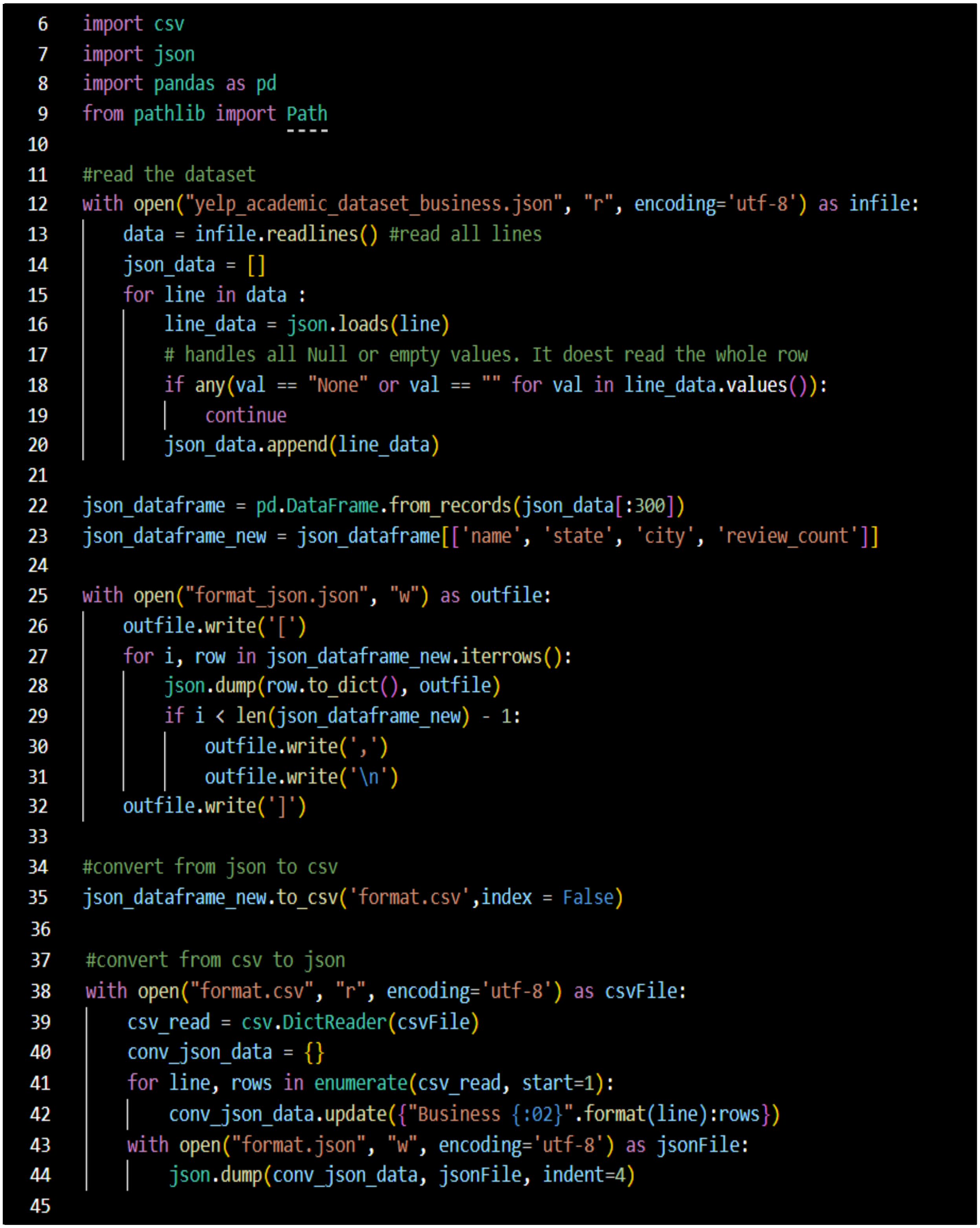
Appendix

Initial Data:


begin.py 1
yelp_academic_dataset_business.json
Settings
yelp_academic_dataset_business.json
⊗ 0 ⚠ 1
Ln 20, Col 143
Spaces: 4
UTF-8

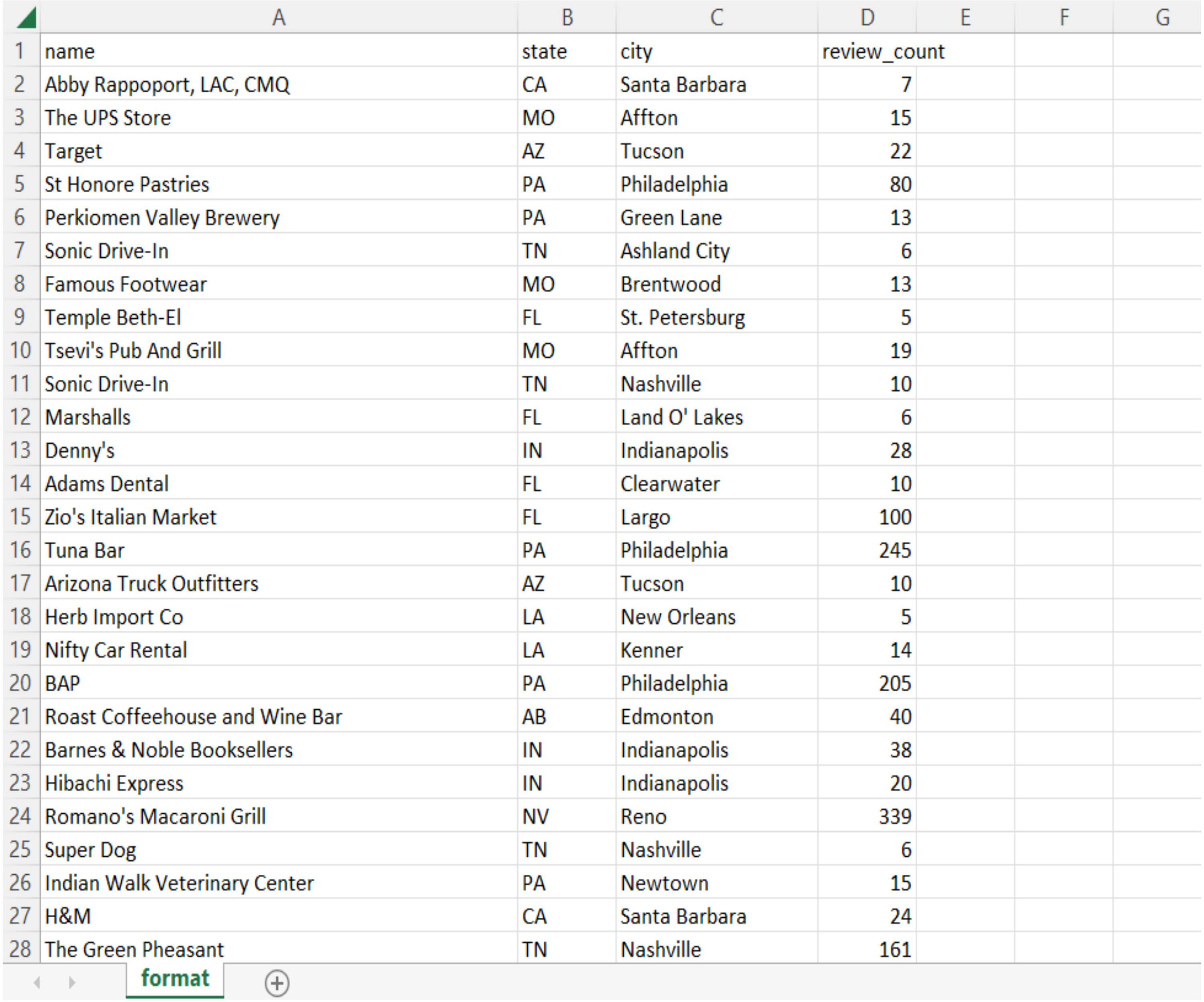
```
{"business_id":"Pns2l4eNsfO8kk83dixA6A","name":"Abby Rappoport, LAC, CMQ","address":"1616 Chapala St, Ste 2","city":"Santa Barbara","state":"CA","postal_code":"93101","latitude":34.4266787,"longitude":-119.7111968,"stars":5.0,"r
{"business_id":"mpf3x-BjTdTEA3yCZrAYPw","name":"The UPS Store","address":"87 Grasso Plaza Shopping Center","city":"Affton","state":"MO","postal_code":"63123","latitude":38.551126,"longitude":-90.335695,"stars":3.0,"review_count"
{"business_id":"tUFrWirKiKi_TAnsVWINQQ","name":"Target","address":"5255 E Broadway Blvd","city":"Tucson","state":"AZ","postal_code":"85711","latitude":32.223236,"longitude":-110.880452,"stars":3.5,"review_count":22,"is_open":0,"
{"business_id":"MTSW4McQd7CbVtyjqoe9mw","name":"St Honore Pastries","address":"935 Race St","city":"Philadelphia","state":"PA","postal_code":"19107","latitude":39.9555052,"longitude":-75.1555641,"stars":4.0,"review_count":80,"is
{"business_id":"mWMc6_wTdE0EUBKIGXDVfA","name":"Perkiomen Valley Brewery","address":"101 Walnut St","city":"Green Lane","state":"PA","postal_code":"18054","latitude":40.3381827,"longitude":-75.4716585,"stars":4.5,"review_count":
{"business_id":"CF33F8-E6oudUQ46HnavjQ","name":"Sonic Drive-In","address":"615 S Main St","city":"Ashland City","state":"TN","postal_code":"37015","latitude":36.269593,"longitude":-87.058943,"stars":2.0,"review_count":6,"is_open
{"business_id":"n_0UpQx1hsNbnPUSlodU8w","name":"Famous Footwear","address":"8522 Eager Road, Dierbergs Brentwood Point","city":"Brentwood","state":"MO","postal_code":"63144","latitude":38.627695,"longitude":-90.340465,"stars":2.
{"business_id":"qkRM_2X51Yqxk3btlwAQIg","name":"Temple Beth-El","address":"400 Pasadena Ave S","city":"St. Petersburg","state":"FL","postal_code":"33707","latitude":27.76659,"longitude":-82.732983,"stars":3.5,"review_count":5,"i
{"business_id":"k0hlBqXX-Bt0vf1op7Jr1w","name":"Tsevi's Pub And Grill","address":"8025 Mackenzie Rd","city":"Affton","state":"MO","postal_code":"63123","latitude":38.5651648,"longitude":-90.3210868,"stars":3.0,"review_count":19,
{"business_id":"bBDDEgkFA1Otx9Lfe7BZUQ","name":"Sonic Drive-In","address":"2312 Dickerson Pike","city":"Nashville","state":"TN","postal_code":"37207","latitude":36.2081024,"longitude":-86.7681696,"stars":1.5,"review_count":10,"i
{"business_id":"UJsufbvfyfONHeWdvAHKjA","name":"Marshalls","address":"21705 Village Lakes Sc Dr","city":"Land O' Lakes","state":"FL","postal_code":"34639","latitude":28.1904587953,"longitude":-82.4573802199,"stars":3.5,"review_c
{"business_id":"eEOYSgkmpB90uNA7lDOMRA","name":"Vietnamese Food Truck","address":"","city":"Tampa Bay","state":"FL","postal_code":"33602","latitude":27.9552692,"longitude":-82.4563199,"stars":4.0,"review_count":10,"is_open":1,"a
{"business_id":"il_Ro8jwPlHresjw9EGmBg","name":"Denny's","address":"8901 US 31 S","city":"Indianapolis","state":"IN","postal_code":"46227","latitude":39.6371332838,"longitude":-86.127217412,"stars":2.5,"review_count":28,"is_open
{"business_id":"jaxMSoInw8Poo3XeMJt8lQ","name":"Adams Dental","address":"15 N Missouri Ave","city":"Clearwater","state":"FL","postal_code":"33755","latitude":27.966235,"longitude":-82.787412,"stars":5.0,"review_count":10,"is_ope
{"business_id":"0bPLkL0QhhPO5kt1_EXmNQ","name":"Zio's Italian Market","address":"2575 E Bay Dr","city":"Largo","state":"FL","postal_code":"33771","latitude":27.9161159,"longitude":-82.7604608,"stars":4.5,"review_count":100,"is_
{"business_id":"MUTTqe8uqyMdBl186RmNeA","name":"Tuna Bar","address":"205 Race St","city":"Philadelphia","state":"PA","postal_code":"19106","latitude":39.953949,"longitude":-75.1432262,"stars":4.0,"review_count":245,"is_open":1,"
{"business_id":"rBmpy_Y1UbBx8ggHlyb7hA","name":"Arizona Truck Outfitters","address":"625 N Stone Ave","city":"Tucson","state":"AZ","postal_code":"85705","latitude":32.2298719,"longitude":-110.9723419,"stars":4.5,"review_count":1
{"business_id":"M0XSSHqrASOnhgbWDJIpQA","name":"Herb Import Co","address":"712 Adams St","city":"New Orleans","state":"LA","postal_code":"70118","latitude":29.9414679565,"longitude":-90.129952757,"stars":4.0,"review_count":5,"is
{"business_id":"8wGISYjYkE2tSqn3cDMu8A","name":"Nifty Car Rental","address":"1241 Airline Dr","city":"Kenner","state":"LA","postal_code":"70062","latitude":29.981183,"longitude":-90.2540123,"stars":3.5,"review_count":14,"is_open
{"business_id":"ROeacJQwBeh05Rqg7F6TCg","name":"BAP","address":"1224 South St","city":"Philadelphia","state":"PA","postal_code":"19147","latitude":39.943223,"longitude":-75.162568,"stars":4.5,"review_count":205,"is_open":1,"attr
{"business_id":"WKMJwqnfZKsAae75RMP6jA","name":"Roast Coffeehouse and Wine Bar","address":"10359 104 Street NW","city":"Edmonton","state":"AB","postal_code":"T5J 1B9","latitude":53.5460453,"longitude":-113.4991693,"stars":4.0,"r
{"business_id":"qhDdDeI3K4jy2KyzwFN53w","name":"Barnes & Noble Booksellers","address":"8675 River Crossing Blvd","city":"Indianapolis","state":"IN","postal_code":"46240","latitude":39.91382,"longitude":-86.105739,"stars":4.0,"re
{"business_id":"kfNv-JZpuN6TVNSO6hHdkw","name":"Hibachi Express","address":"6625 E 82nd St","city":"Indianapolis","state":"IN","postal_code":"46250","latitude":39.9043203184,"longitude":-86.0530799,"stars":4.0,"review_count":20,
{"business_id":"9OG5YkX1g2GReZM0AskizA","name":"Romano's Macaroni Grill","address":"5505 S Virginia St","city":"Reno","state":"NV","postal_code":"89502","latitude":39.4761165,"longitude":-119.7893392,"stars":2.5,"review_count":3
{"business_id":"4iRzR7OaS-QaSXuvYxEGKA","name":"Super Dog","address":"1160 Gallatin Pike S","city":"Nashville","state":"TN","postal_code":"37115","latitude":36.2485696,"longitude":-86.7199857,"stars":4.0,"review_count":6,"is_ope
{"business_id":"PSo_C1Sfa13JHjzVNW6ziQ","name":"Indian Walk Veterinary Center","address":"662 Durham Rd","city":"Newtown","state":"PA","postal_code":"18940","latitude":40.2734926,"longitude":-74.987496,"stars":5.0,"review_count"
{"business_id":"noByYNtDLQAra9ccqxdfDw","name":"H&M","address":"827-833 State St","city":"Santa Barbara","state":"CA","postal_code":"93101","latitude":34.4202089255,"longitude":-119.7004597389,"stars":3.0,"review_count":24,"is_
{"business_id":"tMkwHmWFUEXrC9ZduonpTg","name":"The Green Pheasant","address":"215 1st Ave S","city":"Nashville","state":"TN","postal_code":"37201","latitude":36.1598858,"longitude":-86.7731974,"stars":4.0,"review_count":161,"is
{"business_id":"QdN72BWoyFypdGJhhI5r7g","name":"Bar One","address":"767 S 9th St","city":"Philadelphia","state":"PA","postal_code":"19147","latitude":39.9398245705,"longitude":-75.1574465632,"stars":4.0,"review_count":65,"is_ope
{"business_id":"sqSqqLy0sN8n2IZrAbzidQ","name":"Domino's Pizza","address":"3001 Highway 31 W","city":"White House","state":"TN","postal_code":"37188","latitude":36.4647467622,"longitude":-86.6591868908,"stars":3.5,"review_count"
{"business_id":"fvWn8oXXwbj2l79cochZyw","name":"Altitude Trampoline Park - Boise","address":"1301 N Milwaukee St","city":"Boise","state":"ID","postal_code":"83704","latitude":43.6167635,"longitude":-116.2853821,"stars":5.0,"revi
{"business_id":"Mjboz24M9NlBeiOJKLEd_Q","name":"DeSandro on Main","address":"4105 Main St","city":"Philadelphia","state":"PA","postal_code":"19127","latitude":40.0224662,"longitude":-75.218314,"stars":3.0,"review_count":41,"is_
{"business_id":"8sshLb4UU7emeUDvtJWnpA","name":"DanceLine","address":"30 W Lancaster Ave","city":"Paoli","state":"PA","postal_code":"19301","latitude":40.0415845065,"longitude":-75.4849526577,"stars":4.5,"review_count":11,"is_op
{"business_id":"kV_Q1oqis8Qli8dUoGpTyQ","name":"Ardmore Pizza","address":"10 Rittenhouse Pl","city":"Ardmore","state":"PA","postal_code":"19003","latitude":40.0067071,"longitude":-75.289671,"stars":3.5,"review_count":109,"is_ope
{"business_id":"w_AMNoI1iG9eay7ncmc67w","name":"River 127","address":"100 Iberville St","city":"New Orleans","state":"LA","postal_code":"70130","latitude":29.951359,"longitude":-90.0646715,"stars":3.0,"review_count":12,"is_open"
{"business_id":"aPNXGTDkf-4bjhyMBQxqpQ","name":"Craft Hall","address":"901 N Delaware Ave","city":"Philadelphia","state":"PA","postal_code":"19123","latitude":39.9625821,"longitude":-75.1356571,"stars":3.5,"review_count":65,"is_
{"business_id":"2xVsWBNFwZOxIOdd9Mwnww","name":"Cheeseburger In Paradise","address":"116 N Pottstown Pike","city":"Exton","state":"PA","postal_code":"19341","latitude":40.029962,"longitude":-75.630607,"stars":2.5,"review_count":
{"business_id":"mGFsPRHdE6SMV1gYQ6GzgQ","name":"T-Mobile","address":"4735 Concord Pike","city":"Wilmington","state":"DE","postal_code":"19803","latitude":39.822176,"longitude":-75.546007,"stars":2.0,"review_count":6,"is_open":1,
{"business_id":"LcAozWCMLGjwRbokaJAKMg","name":"Edwardsville Children's Museum","address":"722 Holyoake Rd","city":"Edwardsville","state":"IL","postal_code":"62025","latitude":38.8043945,"longitude":-89.9497332,"stars":4.5,"revi
{"business_id":"fSCNwMtNNQY9QT69Cj9fiA","name":"Sierra Pro Events","address":"","city":"Sparks","state":"NV","postal_code":"89431","latitude":39.5401545,"longitude":-119.7483949,"stars":5.0,"review_count":7,"is_open":1,"attribut
{"business_id":"x1mhq4IpWctQOBM06dU8vg","name":"Diamond Hair Company","address":"5324 W 16th St","city":"Indianapolis","state":"IN","postal_code":"46224","latitude":39.7874644177,"longitude":-86.2514209,"stars":1.5,"review_count
{"business_id":"ljxNT9p0y7YMPx0fcNBGig","name":"Tony's Restaurant & 3rd Street Cafe","address":"312 Piasa St","city":"Alton","state":"IL","postal_code":"62002","latitude":38.896563,"longitude":-90.1862032987,"stars":3.0,"review_
{"business_id":"Pns2l4eNsfO8kk83dixA6A","name":"Abby Rappoport, LAC, CMQ","address":"1616 Chapala St, Ste 2","city":"Santa Barbara","state":"CA","postal_code":"93101","latitude":34.4266787,"longitude":-119.7111968,"stars":5.0,"r
{"business_id":"mpf3x-BjTdTEA3yCZrAYPw","name":"The UPS Store","address":"87 Grasso Plaza Shopping Center","city":"Affton","state":"MO","postal_code":"63123","latitude":38.551126,"longitude":-90.335695,"stars":3.0,"review_count"
{"business_id":"tUFrWirKiKi_TAnsVWINQQ","name":"Target","address":"5255 E Broadway Blvd","city":"Tucson","state":"AZ","postal_code":"85711","latitude":32.223236,"longitude":-110.880452,"stars":3.5,"review_count":22,"is_open":0,"
{"business_id":"MTSW4McQd7CbVtyjqoe9mw","name":"St Honore Pastries","address":"935 Race St","city":"Philadelphia","state":"PA","postal_code":"19107","latitude":39.9555052,"longitude":-75.1555641,"stars":4.0,"review_count":80,"is
{"business_id":"mWMc6_wTdE0EUBKIGXDVfA","name":"Perkiomen Valley Brewery","address":"101 Walnut St","city":"Green Lane","state":"PA","postal_code":"18054","latitude":40.3381827,"longitude":-75.4716585,"stars":4.5,"review_count":
{"business_id":"CF33F8-E6oudUQ46HnavjQ","name":"Sonic Drive-In","address":"615 S Main St","city":"Ashland City","state":"TN","postal_code":"37015","latitude":36.269593,"longitude":-87.058943,"stars":2.0,"review_count":6,"is_open
{"business_id":"n_0UpQx1hsNbnPUSlodU8w","name":"Famous Footwear","address":"8522 Eager Road, Dierbergs Brentwood Point","city":"Brentwood","state":"MO","postal_code":"63144","latitude":38.627695,"longitude":-90.340465,"stars":2.
{"business_id":"qkRM_2X51Yqxk3btlwAQIg","name":"Temple Beth-El","address":"400 Pasadena Ave S","city":"St. Petersburg","state":"FL","postal_code":"33707","latitude":27.76659,"longitude":-82.732983,"stars":3.5,"review_count":5,"i
{"business_id":"k0hlBqXX-Bt0vf1op7Jr1w","name":"Tsevi's Pub And Grill","address":"8025 Mackenzie Rd","city":"Affton","state":"MO","postal_code":"63123","latitude":38.5651648,"longitude":-90.3210868,"stars":3.0,"review_count":19,
{"business_id":"bBDDEgkFA1Otx9Lfe7BZUQ","name":"Sonic Drive-In","address":"2312 Dickerson Pike","city":"Nashville","state":"TN","postal_code":"37207","latitude":36.2081024,"longitude":-86.7681696,"stars":1.5,"review_count":10,"i
{"business_id":"UJsufbvfyfONHeWdvAHKjA","name":"Marshalls","address":"21705 Village Lakes Sc Dr","city":"Land O' Lakes","state":"FL","postal_code":"34639","latitude":28.1904587953,"longitude":-82.4573802199,"stars":3.5,"review_c
{"business_id":"eEOYSgkmpB90uNA7lDOMRA","name":"Vietnamese Food Truck","address":"","city":"Tampa Bay","state":"FL","postal_code":"33602","latitude":27.9552692,"longitude":-82.4563199,"stars":4.0,"review_count":10,"is_open":1,"a
{"business_id":"il_Ro8jwPlHresjw9EGmBg","name":"Denny's","address":"8901 US 31 S","city":"Indianapolis","state":"IN","postal_code":"46227","latitude":39.6371332838,"longitude":-86.127217412,"stars":2.5,"review_count":28,"is_open
{"business_id":"jaxMSoInw8Poo3XeMJt8lQ","name":"Adams Dental","address":"15 N Missouri Ave","city":"Clearwater","state":"FL","postal_code":"33755","latitude":27.966235,"longitude":-82.787412,"stars":5.0,"review_count":10,"is_ope
```

Python code for conversions:



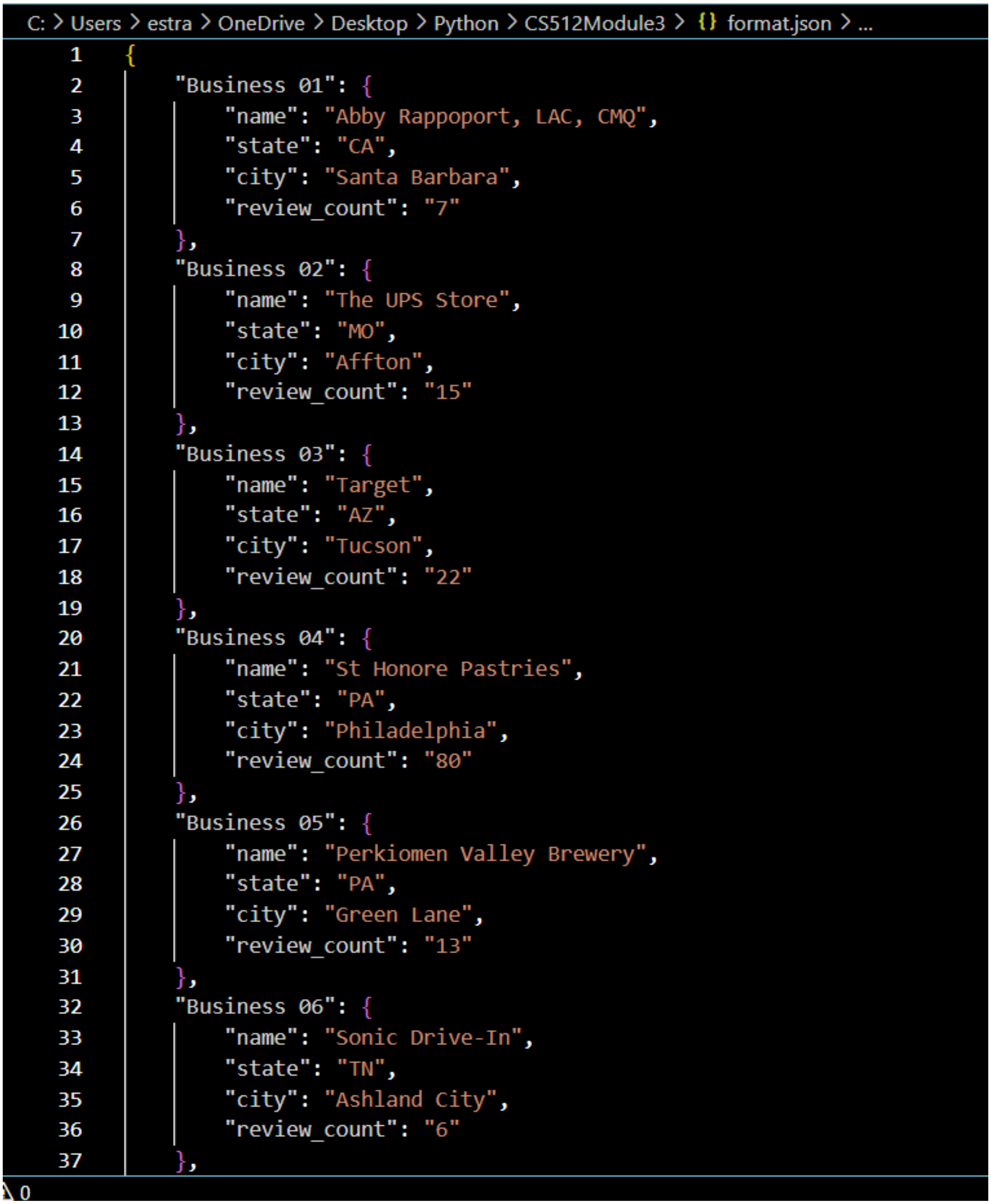
```python
import csv
import json
import pandas as pd
from pathlib import Path

#read the dataset
with open("yelp_academic_dataset_business.json", "r", encoding='utf-8') as infile:
    data = infile.readlines() #read all lines
    json_data = []
    for line in data :
        line_data = json.loads(line)
        # handles all Null or empty values. It doest read the whole row
        if any(val == "None" or val == "" for val in line_data.values()):
            continue
        json_data.append(line_data)

json_dataframe = pd.DataFrame.from_records(json_data[:300])
json_dataframe_new = json_dataframe[['name', 'state', 'city', 'review_count']]

with open("format_json.json", "w") as outfile:
    outfile.write('[')
    for i, row in json_dataframe_new.iterrows():
        json.dump(row.to_dict(), outfile)
        if i < len(json_dataframe_new) - 1:
            outfile.write(',')
            outfile.write('\n')
    outfile.write(']')

#convert from json to csv
json_dataframe_new.to_csv('format.csv',index = False)

#convert from csv to json
with open("format.csv", "r", encoding='utf-8') as csvFile:
    csv_read = csv.DictReader(csvFile)
    conv_json_data = {}
    for line, rows in enumerate(csv_read, start=1):
        conv_json_data.update({"Business {:02}".format(line):rows})
    with open("format.json", "w", encoding='utf-8') as jsonFile:
        json.dump(conv_json_data, jsonFile, indent=4)
```

CSV Sample:



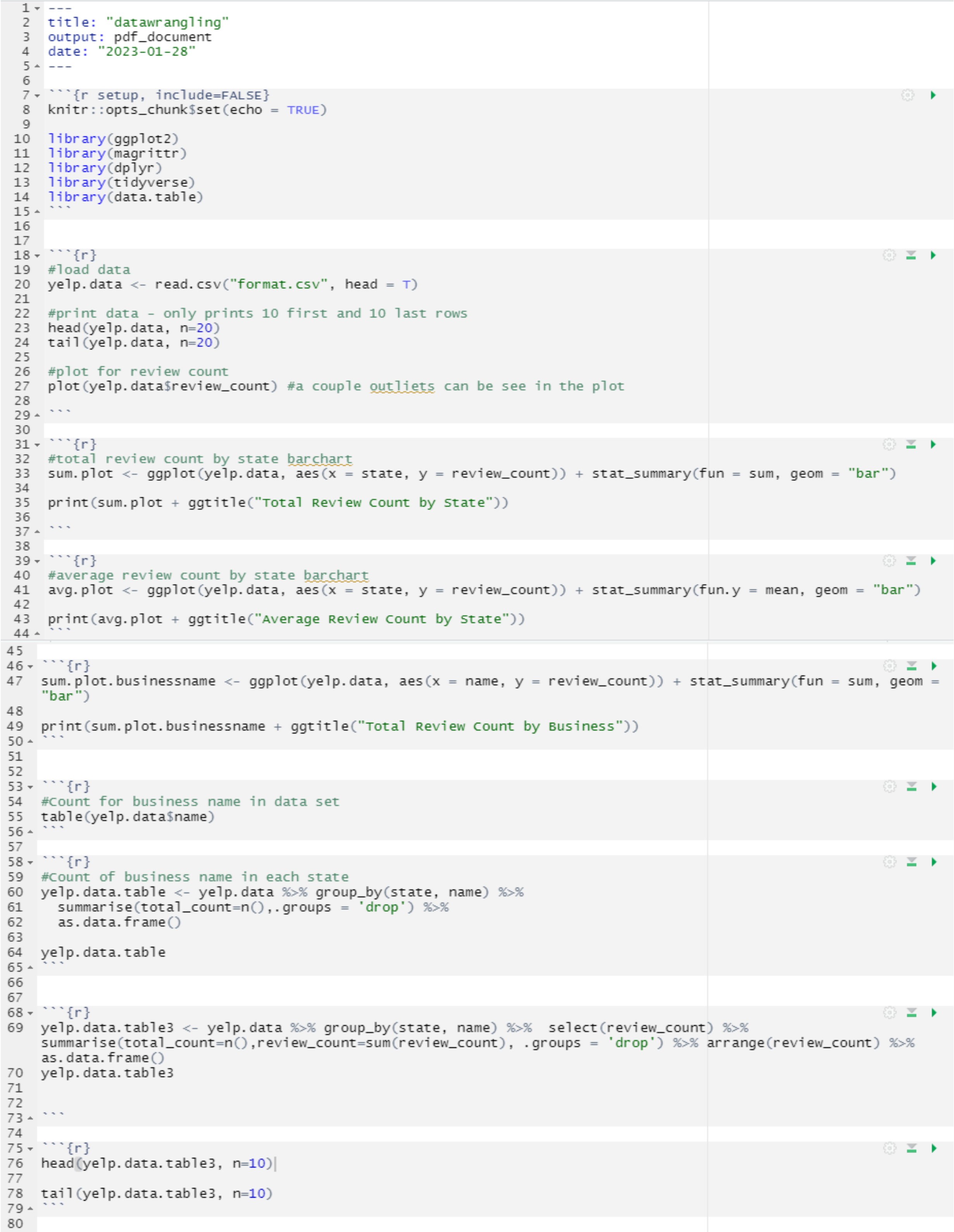
name	state	city	review_count
Abby Rappoport, LAC, CMQ	CA	Santa Barbara	7
The UPS Store	MO	Affton	15
Target	AZ	Tucson	22
St Honore Pastries	PA	Philadelphia	80
Perkiomen Valley Brewery	PA	Green Lane	13
Sonic Drive-In	TN	Ashland City	6
Famous Footwear	MO	Brentwood	13
Temple Beth-El	FL	St. Petersburg	5
Tsevi's Pub And Grill	MO	Affton	19
Sonic Drive-In	TN	Nashville	10
Marshalls	FL	Land O' Lakes	6
Denny's	IN	Indianapolis	28
Adams Dental	FL	Clearwater	10
Zio's Italian Market	FL	Largo	100
Tuna Bar	PA	Philadelphia	245
Arizona Truck Outfitters	AZ	Tucson	10
Herb Import Co	LA	New Orleans	5
Nifty Car Rental	LA	Kenner	14
BAP	PA	Philadelphia	205
Roast Coffeehouse and Wine Bar	AB	Edmonton	40
Barnes & Noble Booksellers	IN	Indianapolis	38
Hibachi Express	IN	Indianapolis	20
Romano's Macaroni Grill	NV	Reno	339
Super Dog	TN	Nashville	6
Indian Walk Veterinary Center	PA	Newtown	15
H&M	CA	Santa Barbara	24
The Green Pheasant	TN	Nashville	161

Json Sample :


C: > Users > estra > OneDrive > Desktop > Python > CS512Module3 > {} format.json > ...
⚠ 0

```json
{
    "Business 01": {
        "name": "Abby Rappoport, LAC, CMQ",
        "state": "CA",
        "city": "Santa Barbara",
        "review_count": "7"
    },
    "Business 02": {
        "name": "The UPS Store",
        "state": "MO",
        "city": "Affton",
        "review_count": "15"
    },
    "Business 03": {
        "name": "Target",
        "state": "AZ",
        "city": "Tucson",
        "review_count": "22"
    },
    "Business 04": {
        "name": "St Honore Pastries",
        "state": "PA",
        "city": "Philadelphia",
        "review_count": "80"
    },
    "Business 05": {
        "name": "Perkiomen Valley Brewery",
        "state": "PA",
        "city": "Green Lane",
        "review_count": "13"
    },
    "Business 06": {
        "name": "Sonic Drive-In",
        "state": "TN",
        "city": "Ashland City",
        "review_count": "6"
    },
```

R-Studio code:



~~~
---
title: "datawrangling"
output: pdf_document
date: "2023-01-28"
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)

library(ggplot2)
library(magrittr)
library(dplyr)
library(tidyverse)
library(data.table)
```


```{r}
#load data
yelp.data <- read.csv("format.csv", head = T)

#print data - only prints 10 first and 10 last rows
head(yelp.data, n=20)
tail(yelp.data, n=20)

#plot for review count
plot(yelp.data$review_count) #a couple outliets can be see in the plot

```

```{r}
#total review count by state barchart
sum.plot <- ggplot(yelp.data, aes(x = state, y = review_count)) + stat_summary(fun = sum, geom = "bar")

print(sum.plot + ggtitle("Total Review Count by State"))

```

```{r}
#average review count by state barchart
avg.plot <- ggplot(yelp.data, aes(x = state, y = review_count)) + stat_summary(fun.y = mean, geom = "bar")

print(avg.plot + ggtitle("Average Review Count by State"))
```

```{r}
sum.plot.businessname <- ggplot(yelp.data, aes(x = name, y = review_count)) + stat_summary(fun = sum, geom = "bar")

print(sum.plot.businessname + ggtitle("Total Review Count by Business"))
```


```{r}
#Count for business name in data set
table(yelp.data$name)
```

```{r}
#Count of business name in each state
yelp.data.table <- yelp.data %>% group_by(state, name) %>%
  summarise(total_count=n(),.groups = 'drop') %>%
  as.data.frame()

yelp.data.table
```


```{r}
yelp.data.table3 <- yelp.data %>% group_by(state, name) %>%  select(review_count) %>% summarise(total_count=n(),review_count=sum(review_count), .groups = 'drop') %>% arrange(review_count) %>% as.data.frame()
yelp.data.table3


```

```{r}
head(yelp.data.table3, n=10)

tail(yelp.data.table3, n=10)
```

~~~

References

e-satisfaction. (n.d.). *7 reasons why customer reviews are important*. Retrieved January 27, 2023, from https://www.e-satisfaction.com/7-reasons-why-customer-reviews-are-important/#:~:text=Analyzing%20reviews%20left%20by%20your,what%20your%20customers%20truly%20want

Convert CSV to JSON using python with new headers. Stack Overflow. (1969, July 1). Retrieved January 27, 2023, from https://stackoverflow.com/questions/73089933/convert-csv-to-json-using-python-with-new-headers

Rana, D. S. (2022, September 12). *Learn Review Analysis: Why, how, data sources with Free Trial*. Learn Review Analysis | Why, How, Data Sources with Free Trial. Retrieved January 29, 2023, from https://www.repustate.com/blog/review-analysis/#:~:text=Review%20analysis%20allows%20you%20to,new%20sales%20opportunities%2C%20and%20more.

Indexing and selecting data - pandas 1.5.3 documentation. (n.d.). Retrieved January 29, 2023, from https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy.

Yelp for Business. (2023, January 23). *Plan, start, grow, and advertise your small business*. Yelp for Business. Retrieved January 29, 2023, from https://business.yelp.com/