

Vegan & Vegetarian Restaurants

Jorit Studer and Larissa Eisele

2/1/2022

Importing Data

We are gathering data from the following sources...

```
# setwd("/Users/jorit/Library/Mobile Documents/com~apple~CloudDocs/Dokumente/HSLU/Semester_01/R_Bootcamp/Data")
restaurants <- read.csv("./Data/veg_restaurants_US.csv")
income_pop_by_zip <- read_excel("./Data/MeanZIP-3.xlsx")
google_searches_vegetarian <- read_excel("./Data/searches_vegetarian_res.xlsx")
google_searches_vegan <- read_excel("./Data/searches_vegan_res.xlsx")
```

Data Processing

Restaurant Data

We are selecting the following data from the restaurant data set

```
restaurants <- restaurants %>% select(id, dateAdded, cuisines, latitude, longitude, menus.category, name)
restaurants$postalCode <- as.numeric(restaurants$postalCode)
restaurants <- restaurants %>% mutate(cuisines_split = strsplit(cuisines, ","))
```

Cleaning the missing data from on restaurant...

```
restaurants[restaurants$id == "AV1Tp59p3D1zeR_xE2DL",]$postalCode <- 94596
anyNA(restaurants$postalCode) # There are no missing values anymore in postalCode (zip code)

## [1] FALSE
```

Population Size and Mean Household Income Data

Joining the Data

```
joined_data <- left_join(restaurants, income_pop_by_zip %>% select(Zip, Mean, Pop), by = c("postalCode" = "Zip"))
```

Checking for missing data

```
sum(is.na(joined_data$Mean)) # all postal codes have a mean household income

## [1] 0

sum(is.na(joined_data$Pop)) # all postal codes have a population

## [1] 0
```

Google Search Data

```
# Vegetarian Restaurant Search Results
vegetarian_temp <- google_searches_vegetarian %>% group_by(province) %>% summarise(total_searches_vegetarian = sum(searches))
joined_data <- left_join(joined_data, vegetarian_temp %>% select(province, total_searches_vegetarian), by = "province")

# Vegan Restaurant Search Results
vegan_temp <- google_searches_vegan %>% group_by(province) %>% summarise(total_searches_vegan = sum(searches))
joined_data <- left_join(joined_data, vegan_temp %>% select(province, total_searches_vegan), by = "province")

Checking for missing data
sum(is.na(joined_data$total_searches_vegetarian))

## [1] 42
sum(is.na(joined_data$total_searches_vegan))

## [1] 42
# 42 provinces have no google search hits
```