GENERAL ASSEMBLY

AUSTIN RESTAURANTS HEALTH INSPECTION

Project Midterm Presentation

by Nikos Vergos

DATA SET

GENERAL ASSEMBLY

- Using open data from data.austintexas.gov
- Data set: City of Austin/Travis County Restaurant Inspections
  - 22,875 rows (inspections)
  - 4,774 establishments inspected (roughly 2x a year)

	Restaurant Name	Zip Code	Inspection Date	Score	Address	Facility ID	Process Descriptic
1	15th Street Cafe	78701	05/21/2013	91	303 W 15TH ST	2801033	Routine Inspection
2	15th Street Cafe	78701	12/04/2013	97	303 W 15TH ST	2801033	Routine Inspection
3	15th Street Cafe	78701	05/12/2015	97	303 W 15TH ST	2801033	Routine Inspection
4	15th Street Cafe	78701	11/09/2015	86	303 W 15TH ST	2801033	Routine Inspection
5	15th Street Cafe	78701	06/17/2014	97	303 W 15TH ST	2801033	Routine Inspection
6	15th Street Cafe	78701	12/01/2014	93	303 W 15TH ST	2801033	Routine Inspection
7	1st Food Mart	78704	07/26/2015	94	1410 S 1ST ST	10677646	Routine Inspection
8	1st Food Mart	78704	07/28/2014	90	1410 S 1ST ST	10677646	Routine Inspection
9	1st Food Mart	78704	02/08/2015	94	1410 S 1ST ST	10677646	Routine Inspection
10	1st Food Mart	78704	02/13/2013	97	1410 S 1ST ST	10677646	Routine Inspection
11	1st Food Mart	78704	01/15/2014	90	1410 S 1ST ST	10677646	Routine Inspection
12	1st Food Mart	78704	06/16/2013	100	1410 S 1ST ST	10677646	Routine Inspection
13	1-Stop Food Store	78751	01/07/2014	100	5101 AIRPORT BLVD	10632697	Routine Inspection
14	1-Stop Food Store	78751	05/14/2015	100	5101 AIRPORT BLVD	10632697	Routine Inspection
15	1-Stop Food Store	78751	12/22/2014	100	5101 AIRPORT BLVD	10632697	Routine Inspection

PROJECT OUTLINE

GENERAL ASSEMBLY

- First and foremost: Clean and "enrich" data set
- Part I: Exploratory Data Analysis
  - Scores distribution
  - Correlation among inspections
  - Confidence intervals for scores by type of restaurant
- Part II: Machine Learning
  - A) [*Supervised*] Classification Algorithms: can we **predict** a restaurant's performance based on past scores?
  - B) [*Unsupervised*] PCA & k-means clustering: are there any **patterns** in names/addresses? Do they affect the score?
- Part III: Fancy stuff (time permitting)
  - Geocoding - visualizations: Heat maps by zip code / area
  - Linear regression: web scraping for yelp scores, add zip codes' median income - can I predict health inspection score?

EXPLORATORY DATA ANALYSIS

GENERAL ASSEMBLY

EXPLORATORY DATA ANALYSIS

GENERAL ASSEMBLY

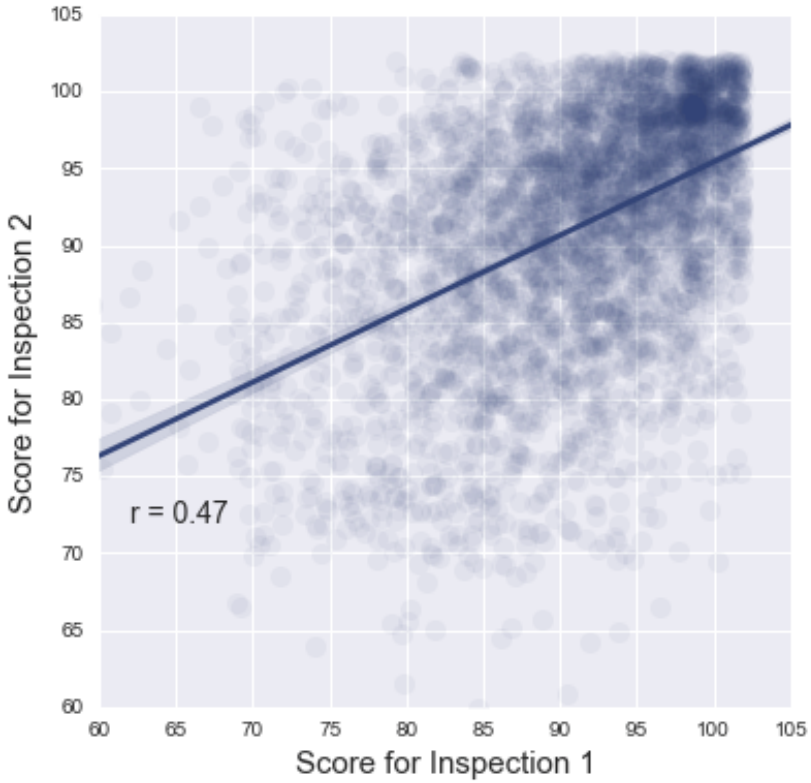
Manipulate data set:

In [12]:
```
# Unfortunately the words "Market", "Grocery", "Church" correspond to places like convenience stores and grocery stores
# I want to make sure that some restaurants with those words in their name will go through my filter.

df.loc[(df['Restaurant_Name'].str.contains('Whole')) , 'Restaurant_Name'] = "Whole Foods"
df.loc[(df['Restaurant_Name'].str.contains('Central Market')) , 'Restaurant_Name'] = "Central Mkt"
df.loc[(df['Restaurant_Name'].str.contains('Boston Market')) , 'Restaurant_Name'] = "Boston Mkt"
df.loc[(df['Restaurant_Name'].str.contains('Mandola')) , 'Restaurant_Name'] = "Mandola's"
df.loc[(df['Restaurant_Name'].str.contains('Royal Blue')) , 'Restaurant_Name'] = "Royal Blue"
df.loc[(df['Restaurant_Name'].str.contains('Rudy')) , 'Restaurant_Name'] = "Rudy's"
df.loc[(df['Restaurant_Name'].str.contains('Fit Foods')) , 'Restaurant_Name'] = "My Ft Foods"
df.loc[(df['Restaurant_Name'].str.contains("Church's Chicken")) , 'Restaurant_Name'] = "Chrch Chicken"
df.loc[(df['Restaurant_Name'].str.contains("Schlotzsky's")) , 'Restaurant_Name'] = "Schlotzsky's"
```

```
df = df[df['Restaurant_Name'].str.contains('School | Elementary | Care | Middle | Cafeteria | Jail | ISD | Academy') ==
df = df[df['Restaurant_Name'].str.contains('Mart | Gas | Convenience | 7-Eleven | HEB | Station | Randall | Target | Fl
df = df[df['Restaurant_Name'].str.contains('Gym | Fitness | Fit') == False]
df = df[df['Restaurant_Name'].str.contains('Church | Dollar | Store | Texaco') == False]
```

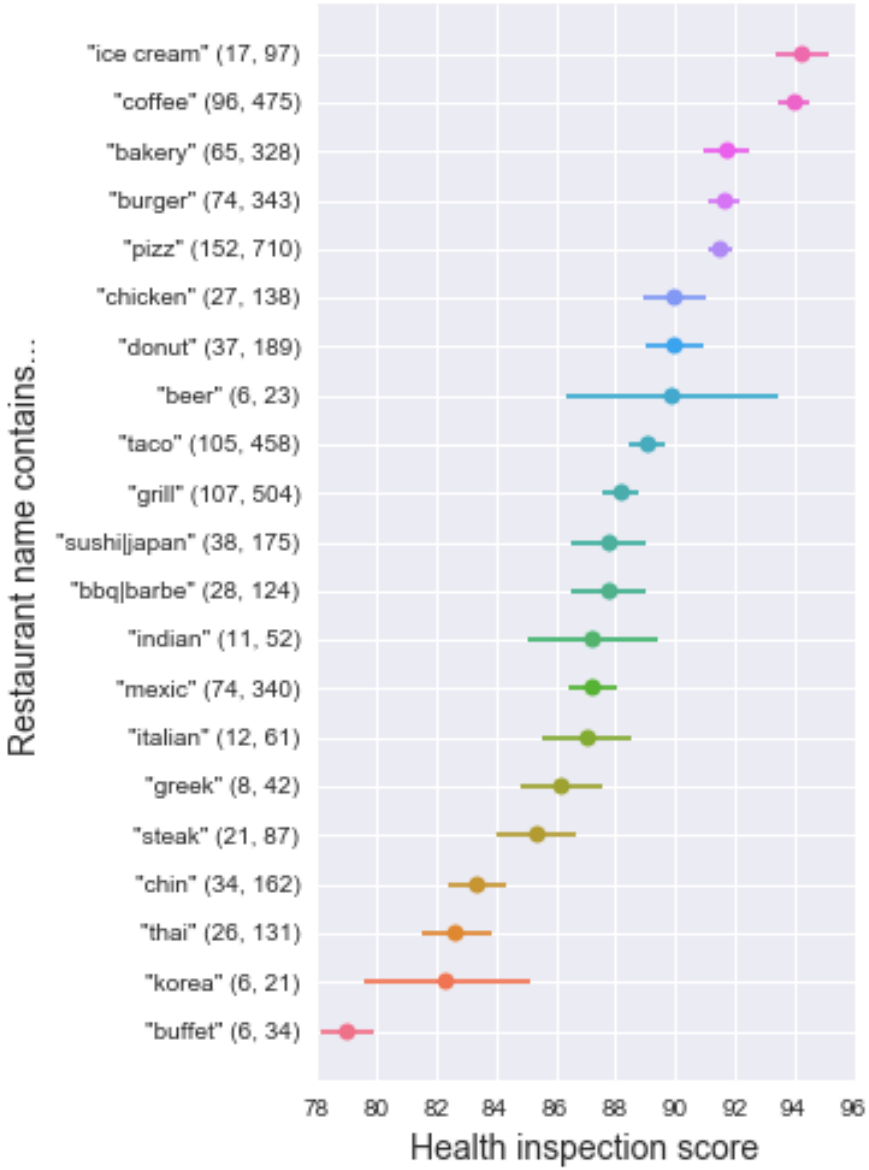
EXPLORATORY DATA ANALYSIS



Scores are fairly consistent and that's good

EXPLORATORY DATA ANALYSIS

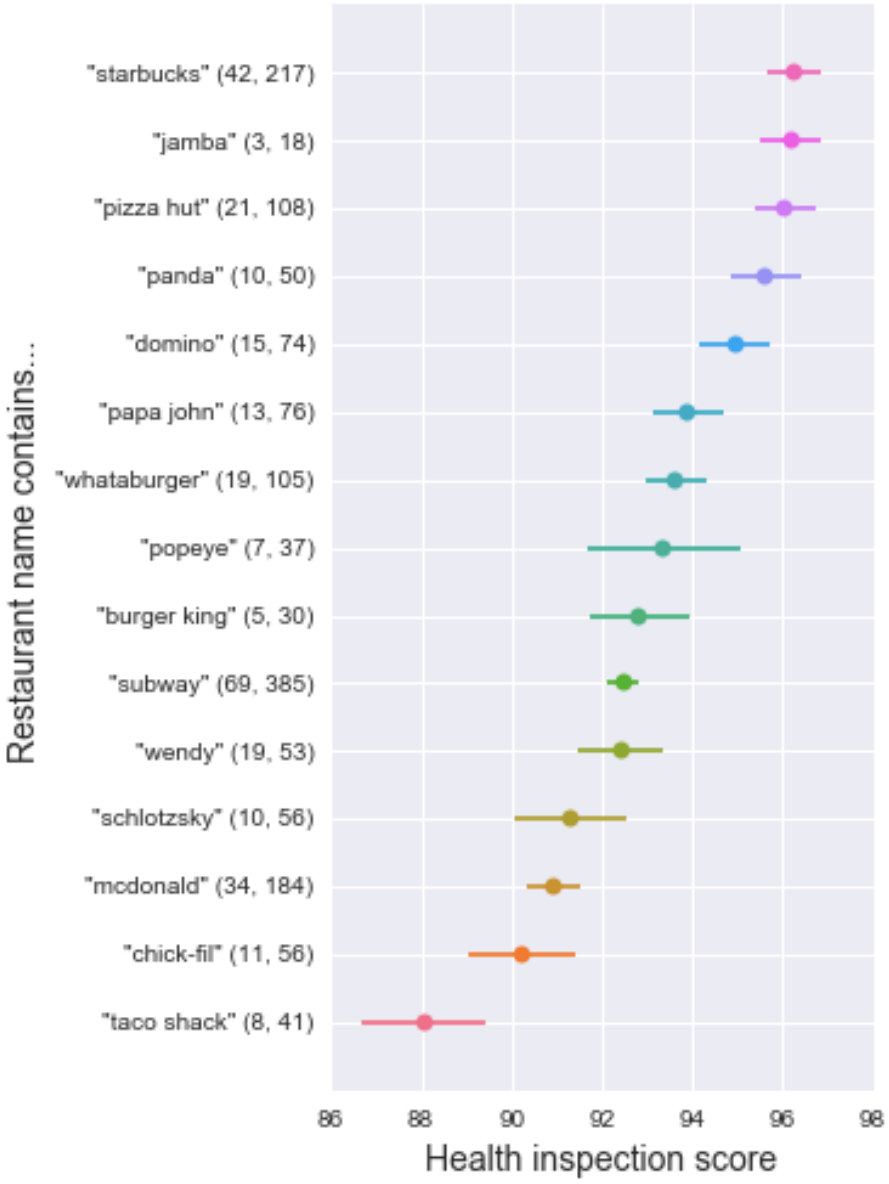
GENERAL ASSEMBLY



But buffets are so tasty :(

EXPLORATORY DATA ANALYSIS

GENERAL ASSEMBLY



- *"I'm in the mood for something CLEAN today"*
- *"Me too! Let's eat Pizza Hut! Or Panda Express?"*