**Motivation:**

High ratings and awards can drive large tourist crowds into local favorite restaurants, often causing restaurants to change (increase prices, new booking rules, impersonal service) to better accommodate the new customer base. As a result, the most popular and highest-rated restaurants may no longer be true local favorites but instead thrive on their popularity with tourists. By identifying local users and local experts to provide ratings reflective of local opinions, Yelp may become more popular with tourists looking to "travel like a local" and enjoy a more authentic experience.

**Background:**

The objective of the research project is to build and combine two models (Local Expert Identifier / Topical Expert Identifier) for the purpose of identifying 'experts' among yelp users. The "Local Expert Identifier" is a Gaussian Mixture Model that identifies clusters in a given user's review locations to predict the user's most probable location.



The "Topical Expert Identifier" is currently a supervised learning algorithm that combines different features about the users reviews in a certain category in order to determine if they are an expert in that category. The goal is to see if an unsupervised algorithm would be able to classify users into clusters of expert and non-expert without needing labels. The goal is to combine the models to find local experts in a specific category.

**Data:**

**Experiments:**

**Local Expert Identifier**

Local Expert Identifier is originally run on 4000 users with the following thresholds.

Tdistance = 10 km, Ttimespan = 500 days, Treview\_count = 10 reviews, Treview\_ratio = 0.36

Part Zero:

Business Yelp Ratings vs Actual User Ratings

Part One:

Business Yelp Ratings by Locals vs Actual Local User Ratings

Part Two:

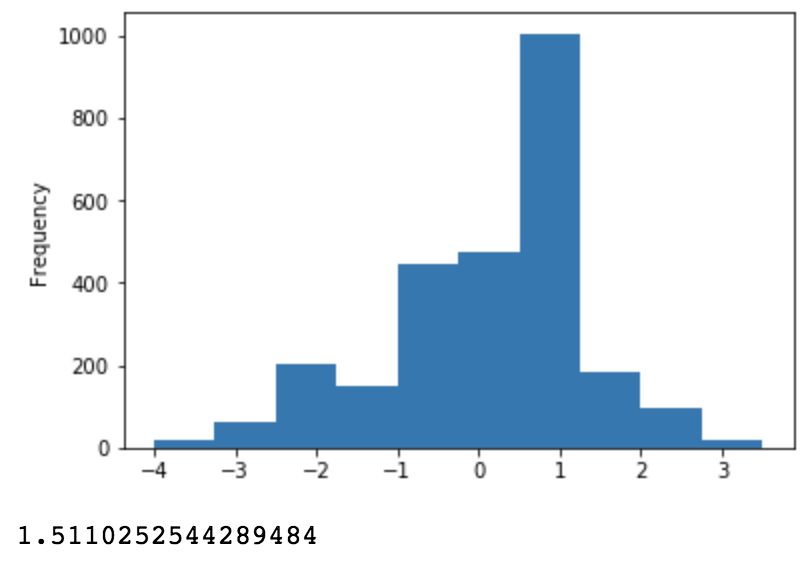
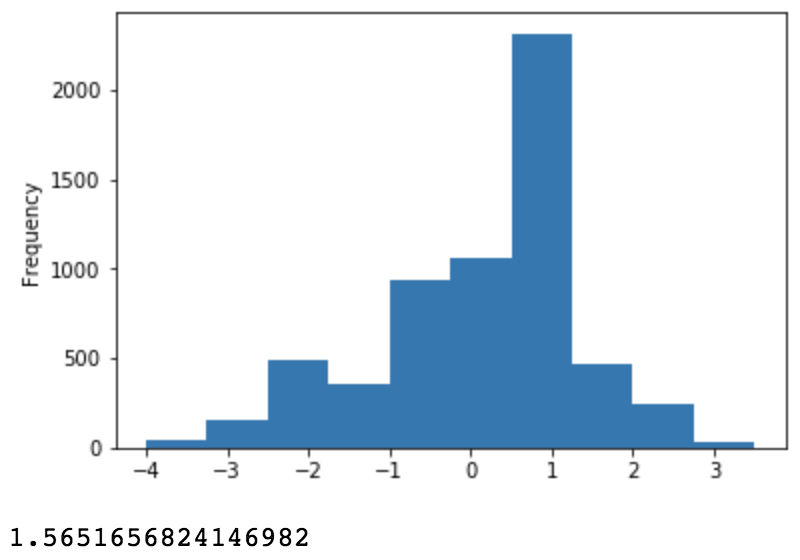
Business Yelp Ratings by Locals over Ttimespan vs Actual Local User Ratings

Part Three:

Business Yelp Ratings by Locals over Ttimespan and over T­review\_ratios vs Actual Local User Ratings

**Results:**

Part Zero Part One



Part Tw Part Three

