ReadMe

Introduction:

Jinglin Wang (N18578723) is responsible for all analysis related to yelp dataset.

Every folder has the code, readme, output files (Some input and ouput files are too large, about 2G, to upload to the NYU classes. So I put all input and output files in my dumbo hdfs folder: **/user/jw4716/data**). But because I didn't store screenshots for some analysis when I ran them before, I run them again when I put these code and files together and get new screenshots.

The following is a rough introduction for all folders (Each folder has a **separate readme** in it.)

DataCleansingAndFormating:

Foler	Methods
datacleansing	Linux Command and Python
addLocationNumber	Python

DataAnalysing:

Folder	Methods Used
HiveRelated	Hive Commands
Ngram	MapReduce, Ngram, Java
BizStar	Mapreduce,Java
SentimentAnalysis	Python
bizCityCount	MapReduce, Java
bizStateCount	MapReduce, Java

bizCategoriesCount	MapReduce, Java