Impact of Amazon Warehouse on Surrounding Areas

Part A – Description of the problem and the data set

Amazon is an every expanding company which has been causing dramatic shifts in the retail sector. The shift from shopping in traditional brick and mortar stores to online forums is apparent from the negative trends in many retail companies (Macy’s, Toys R Us, etc). The question we are exploring is whether the Amazon warehouses have an impact on the demographic profile in the surrounding regions where they are located.

The data set we are working with is a list of the Amazon warehouses which is publicly available online. The list includes the location where the warehouse is located, in some cases the square footage of the facility is provided, a description of the type of facility, and the year the facility was opened. In order to get an organized form of locations, another application was used to search the zip codes and get the city and county for each facility. There was additional cleaning involved with the amazon list of facilities (talk in more detail about this?)

The second half of the dat­­a set is census data. Moodys.com has an organized version of census data which can be viewed at a city and county level. Because census data is only available once every 10 years, ACS (American Community Survey) data which is collected on a yearly basis from a sample of the population, was thought to be a better proxy for seeing shifts in the demographic profile of each region. In addition, we thought it would be best to start off viewing the data from a county level as opposed to a city level. The demographic data points we focused on included: population, per capita income, Median Family Income, Median Home Value, Median Gross Rent, Occupied Housing Units, Seasonal Homes, Median Age, and % of the population below poverty level.

The Moody’s census data was matched with the Amazon list of facilities by using the name of the county and state combination. This worked fairly well with a few exceptions. Overall, there is a list of 124 unique counties with Amazon facilities. Many of the counties have more than one facility and there are 295 facilities included in the data set.

a) A description of the problem and the data set, including details about data collection, if applicable

b) A description of initial data exploration

c) A statement of goals

d) A discussion of related literature