**Lite Web Scraping Activity**

After going through the async materials this week, discuss the following:

What is web scraping?

Web scraping is a term used to describe the use of a program or algorithm to extract and process large amounts of data from the web.

How is it done?

Usually done by extracting the data form the web URL (Beautiful Soup module).

You import the libraries (BeautifulSoup and urllib.request from urlopen) and then use a urlopen on the web URL to read the data from the webpage.

Then there is likely manipulation to read the proper format and dive into the structure to find the information you want to see.

Why is it done?

You can use this to monitor sentiment about your company, brand monitoring, market analysis, lead generation, study buying patterns, etc.

Use the article about Dominos and one other article to discuss the use of Big Data in analytics.

What type of data are they grabbing?

Dominos: Big Data – grabbing text, twitter, amazon echo, etc. This data is combined with demographic and geocode data along with competitor information – they use 85,000 data sources.

Coca-Cola: grabbed information off machines that were piloting a flavor to see how it was selling. Also grab facebook and twitter information 9fans)

Why is this important to the company?

Dominos: It helps them to gain competitive advantage and measure success.

Coca-Cola: helps them target advertising

What are they doing with the data?

Dominos: They are a data-driven company that feels that they can

Coca-Cola: Helps them stay ahead of the competition and provide insight into who is drinking their products

Dominos: Data-driven decision making at the world’s largest pizza delivery chain  https://www.bernardmarr.com/default.asp?contentID=1264

Fitbit:  Big Data in the health and fitness arena.  https://www.bernardmarr.com/default.asp?contentID=714

Coca-Cola: Driving success with AI and Big Data  https://www.bernardmarr.com/default.asp?contentID=1280

Talk about some specific examples of what kind of data you might grab this way.

You will have 20 minutes to discuss this in your group.  Be prepared to report back to the group.