NYU | STERN

DATA MINING FOR BIZ ANALYTICS
INFO-UB.57
Spring 2019

Homework #1 **Due: turned in on NYU Classes by 3pm on Feb 4, 2019**

Ashley Tsoi (ast418)

(put your name above)

Total grade: _____ out of _____ points

Please answer all questions/follow all directions.

Please submit homework as a PDF ←←←

1) Read the syllabus in its entirety. Mark "Yes" below.

___Yes__ I have read and understood the syllabus for this class.

2) In 2-4 sentences, differentiate "Data Science" and "Big Data", and also tell how they are related.

Data Science is a includes a large variety of techniques used to extract insights and information for data, such as data cleansing, preparation, and analysis. Big Data is the humongous volumes of data that cannot yet be processed effectively with traditional applications nowadays. My understanding is that Big Data is one special case under the Data Science umbrella that is commonly experienced by many businesses as their operations entail a large amount of analysis.

3) MTC (MegaTelCo) has decided to use data mining to address its problem of churn in its wireless phone business. As a reminder, "churn" refers to the problem where a customer leaves / stops using a service.

As a consultant to MTC, you realize that a main task in the business understanding/data understanding phases of the data mining process is to define the target variable. In one or two sentences, please suggest a definition for the target variable, i.e. what the data mining task is trying to predict.

Be as precise as possible — someone else will be implementing your suggestion. ***(Remember: it should make sense from a business point of view, and it should be reasonable that MTC would have data available to know the value of the target variable for historical customers.)***

Since we are trying to predict the reason for customer churn, we should measure some KPIs including but not limited to: 1) where the customers are leaving to; 2) MTC's relative pricing compared to its competitors; 3) general customer usage of phone services in the area (could be a little harder for the company to measure, but having an approximate number is fine since all MTC needs to know is if that number moves up or down by a large magnitude). Such specific data measures to collect would define our target variable – customer churn, as these measures will influence their customer churn measure and thus provide MTC a better picture of the causes of this problem.

4) Present a business domain where data mining can be applied. Mention how data mining fits in that domain, e.g. a potential problem it can help solve for that business. 2-3 sentences should suffice.

Many hedge funds nowadays use data-mining to collect alternative data on companies that they want to invest in. For example, if they are looking at water delivery company, they may want to know the availability of delivery in each area, such as by automating inputs of zipcodes on the water company's website, then scrape the results. The fund can generate a pattern and automate this pattern finding by writing codes, then note if there is any changes to the pattern.

5) Getting set up with Python. Click this link and follow the instructions to log into the Jupyter server and download the first modules. Once you have cloned the repository open Module1_Bash_Pandas. Open the notebook titled Command line and data frames 2019.

***If you successfully got through the installation completely* (possibly with help), please give a quick report here. (E.g., "worked like a charm", or "XYZ caused me problems"). Please perform the cut-and-paste operation on your code and the output for the Hands-On section.**

***If you were not able to complete the installations*, give a brief report of where you got stuck, what help you sought and when you sought the help.**

Did not clone the files through terminals but instead downloaded the files from GitHub and opened the files from my own Jupyter notebook through Anaconda. Hope this is fine? Please let me know if I have to access to Jupyter through the NYU network with my nyu id then I will do as you say. Otherwise, everything works well for me! ☺

6) Fill out this survey on prior technical experience.

Done!