**DATA SCIENCE \_ BBangya (due July,7)**

*Why are you investigating this question?*

* Student Performance is a big matter in Korea because 12-year studies in childhood decide young children's future lives. Plus Korean people believe environmental factors affect students' studies. There's a reason why people try to keep living in 강남 which is known as the best place to educate their children. Our team focuses on this point.

cf) <http://news.donga.com/Main/3/all/20140123/60372992/1>

* Our data is about Portugese students' performance. We are going to investigate what affects their studies.

*What are the limitations of your question?*

* Grouping variables into internal and external factors could include subjective views.

Since the data is about Portugal, there could be cultural limitation.

The number of sample is smaller than total number of students.

*How are you going to try and answer your question?*

* We are going to use multiple regression with variable selection for analysis and answer the questions.

*Who else has answered this question? How will you build upon other work done with this question?*

* http://www3.dsi.uminho.pt/pcortez/student.pdf

While in previous research they use Decision Tree method, we’ll use regression and figure out what factors affect score and how it affects in each factors. This previous research can be reference to find how to select variables for analysis.

*References*

P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.