Selected capstone projects:

From the original five capstone proposals, three were selected: proposals 1, 2 and 4.

Proposals 1 and 2 were closely related and they will use the same database. Therefore, proposals 1 and 2 can actually been seen as plan A and plan B of one single project (if feasible, we intend to implement both plans). The project consists of analyzing the data from a Colombian private health insurance company. For obvious market-related reasons, the company must to remain anonymous, but they have provided their consent to use a limited sample of their database, which was already used in a scientific peer-reviewed article (2015 with RAMIREZ, M. M. Z. ; FERREIRA, F. F. ; RODRIGUES NETO, C. Impacto de la utilización de la red propia de una aseguradora privada sobre los costos de atención ambulatoria, Impact of using a private insurer's in-network on ambulatory care costs. Salud Pública de México, v. 57, p. 426-432). The database contains 8304 rows of annual data from 2008-2012 regarding doctors paid by the health insurance company. More specifically, for each doctor the database informs medical expertise, location (region of Colombia), number of patients, number of consultations, number of surgeries, value paid per consultation and value paid per surgery. The main goal of this project is to analyze the surgeries/consultations ratio among doctors. We want to know what can explain or help to understand the variation among doctors in terms of surgeries/consultations ratio, which is an important concern for health insurance companies.

The first plan is to build an agent-based model (https://en.wikipedia.org/wiki/Agent-based_model) with parameters optimized by genetic algorithm in order to find out if differences in behavior among doctors can be explained by their pursue of income (taking into account that surgeries pay more than consultations) and their strong influence on patients, as suggested by the physician-induced demand theory and the target income hypothesis (https://en.wikipedia.org/wiki/Supplier-induced_demand#Health_economics). A preliminary version of the agent-based model was already coded in matlab with promising results (good fit between simulated and empirical distributions). We intend to translate that model into python and improve it so that it relies on fewer parameters, and then we will be ready to extract meaningful graphical results.

The second plan is to use Bayesian statistics, most likely Markov Chain Monte Carlo, in order to estimate a statistical model capable of predicting, for example, the probability that certain doctor would refer a patient to surgery based on the doctor's expertise, location and number of patients.

Finally, proposal 4 would also use Bayesian statistics to organize a portfolio of stocks from some industry aimed at some objective like, for example, minimize financial risk.

Terry M. I.