**Proposal for MA678 midterm project**

**Name:** Qiannan (Nancy) Shen

* **Personal Statement**: Since my major is actuarial science for undergraduate study and statistics for graduate study, my career goal mainly focuses on getting a position related to data analysis and actuarial science in an insurance company. For a position like that in the insurance company, it requires the ability to deal with the detailed information about the insured and build a proper model to get the related outcomes like the insurance premium, the amount of the claims, risk, and so forth. I found this dataset in Kaggle which is talking about medical insurance. This dataset collects the individual’s basic information like age, sex, region, and includes the outcome of individual medical costs billed by health insurance. This project could be great practice before working in the insurance company and handling larger data.

* **Question**: How variables (age, sex, bmi, children, smoker, region) affect the individual medical costs billed by health insurance?

* **data source**: Medical Cost Personal Datasets <https://www.kaggle.com/mirichoi0218/insurance>
* **Proposed Timeline of work**:
* EDA: 11.6 – 11.10
* Data Processing: 11.11-11.15
* Modeling and Validation: 11.16-11.22
* Write up: 11.23-11.29