Xiang Li

11/ 04/ 2021

MA 678 Midterm Project Proposal

* **Personal Statement**: Since my undergraduate major is in actuarial science, I have some background in insurance pricing. My future career goal is in the insurance industry being a data scientist. When I was searching data sets in Kaggle, I find the US Health Insurance Dataset impressive, which is a dataset about Insurance Premium Charges in US with important details for risk underwriting, including factors of Age, Sex, BMI, Number of Children, Smoker and Region. Therefore, I could use this project as a chance to explore the entry level data in the insurance field.
* **Questions**: What factors affected the insurance charges most? (Age, Sex, BMI, Number of Children, Smoker and Region.)
* **Data Resource**: <https://www.kaggle.com/teertha/ushealthinsurancedataset>
* **Timeline of Work**:

1. EDA: 11/06-11/12
2. Data Processing: 11/13-11/19
3. Modeling and Validation: 11/20-11/23
4. Write up: 12/01-12/02