**Proposal for MA678 midterm project-Zhe Yu**

* **Personal Statement**: my career goal is to work in an internet company including e-commerce, video platform, social media platform, etc. It is clear that if work in an internet company, people will meet a lot of data provided by netizen-their interactions with website. And my project is about TED talks. TED is an American media organization that posts talks online for free distribution under the slogan "ideas worth spreading". It likes YouTube that views will give comment on video of speaking. And the dataset in the project includes the number of views, the number of comments, the published date and so on. Thus, in this project I will have an initial contact with this kind of data which may help me get work under way more quickly if I have a job in the future.

* **Question**: What factors affect the popularity(comments) of talks in TED? (duration, speaker, occupation of speaker, published time...)

* **data source**: <https://www.kaggle.com/miguelcorraljr/ted-ultimate-dataset>.

* **Proposed Timeline of work**:
* EDA: 11.9-11.13
* Data Processing: 11.14-11.17
* Modeling and Validation: 11.18-11.24
* Write up: 11.25-11.30