**Data Science II Midterm Project Proposal**

**Group 5 members:**

Eleanor Zhang(zz2602), Mengran Ma (mm5354), Bingyu Sun (bs3142)

**The tentative project title:**

Apple iOS app rating analysis (investigate factors contributing to user rating)

**The anticipated data sources:**

<https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps>

**The response(s) and predictors in the dataset:**

Response:

**user\_rating**: Average User Rating value (for all version)

Predictors:

1. **size\_bytes**: Size (in Bytes)
2. **price**: Price amount (adjust for currency type)
3. **rating\_count\_tot**: User Rating counts (for all version)
4. **rating\_count\_ver**: User Rating counts (for current version)
5. **user\_rating\_ver**: Average User Rating value (for current version)
6. **ver**: Latest version code
7. **cont\_rating**: Content Rating
8. **prime\_genre**: Primary Genre
9. **sup\_devices.num**: Number of supporting devices
10. **ipadSc\_urls.num**: Number of screenshots shown for display
11. **lang.num**: Number of supported languages
12. **vpp\_lic**: Vpp Device Based Licensing Enabled

**The planned analyses:**

We want to explore the relationship between the covariates listed above and mobile app rating. We planned to build linear or nonlinear model to investigate their association.