Bella - Facilitator, Coordinator

Matteo - Initiator, Evaluator

Erik - Arbitrator, Compromiser

Andrew - Notetaker, Gatekeeper

Martin - Coach, Critic

Project Goals and Timeline

1. Specific Aims
   1. Project Client A, Booze ‘R’ Us - a company that owns a large number of liquor stores in Iowa, sales in the coming year, so they can decide whether they will have enough income to expand their operations. We will accomplished this based on a SLR based on [what we use from the data]
   2. Predict volume category (bottles sold is high, medium, or low amount) of an Iowan liquor purchase based on a set of best factors; [list of factors tried in preliminary EDA]. Help in analyzing patterns in sales for the Drinking Excess Alcohol is Dangerous (DEAD) - A nonprofit group seeking to make alcohol culture safer in Iowa.
   3. Learning Goals
      1. Perform feature engineering and simple data integration tasks to prepare datasets for machine learning.
      2. Implement linear regression model fitting procedures and tailor your implementation to work with sufficiently large datasets
      3. Evaluate prediction accuracy of different models using validation metrics.
      4. Tailor your model selection and presentation approach depending on clients’ interests; specifically, focusing on either prediction or inference.
      5. Consider ethical and legal responsibilities in your machine learning process and final recommendations.
2. Timeline
   1. September 29th - October 4th/5th
      1. Project Kickoff and Data Collection
         1. Data pulled from this [link](https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy).
         2. Using sodAPI
         3. Individually assign jobs and work on separate Colab notebooks with the goal to merge them into one file located in this [repository](https://github.com/mshsu/data403-project1).
            1. Martin - Exploring Hill Street
      2. Develop SLR model approach on small subsets of data, before fitting the final model on a larger set.
         1. Determine what cleaning is most important
         2. Handle missing data by [how we handle missing data]
         3. Feature Engineering: [describe how and if we combine and modify data columns into new measurements]
      3. Final Deliverable
         1. Finish EDA, Data Cleaning, and Feature Engineering and share preliminary results with group
         2. Tentative Meeting - October 4th @ 4pm
         3. Merge notebooks into one cohesive .ipynb
         4. Write and turn in description on October 4th or 5th
   2. October 6th - October 12th
      1. Project Proposal for Client A: