

## 2/24 Meeting Notes

Propose 5 questions about the dataset by **Sunday, 2/27**, and use visualization/stat summary to demonstrate your results. Submit a report by **Friday, 3/4**.

Sample questions:

1. What is the distribution of the number of purchases made by various age groups?
2. Use visualization to find the relationship between customers' ages and the frequency of their purchases?

Your EDA questions:

- 1) Do we have outliers in the number of purchases?
- 2) What is the distribution of the number of purchases made by various club member statuses?
- 3) Use visualization to find the most bought items and the least bought items in the store?
- 4) Use visualization to find the age that is responsible for the majority of sales?
- 5) What is the most influential variable in our dataset?
- 6) Are two variables correlated because they have the same cause?
- 7) Is there a pattern between the fashion news frequency and the frequency of purchases?  
(additional question)

My feedback:

Questions 1-4 are very well composed, go ahead and use tools/packages to answer these questions in the report. Question 5 seems slightly vague, do you mean most influential in predicting customer's purchases? Question 6 can be more interesting if you can find the strongest pair of variables in correlation.