* **9/29**
  + Data Cleaning
  + Client A: county, storefront, year, sales info
    - subset data - Store Name: Hill Street News and Tobacco | Years: start of 2017 - end of 2020 (only one location - need a name with more locations)
    - see trends for single company over past X years to predict following year
  + Overall, we are predicting total sales in dollars using:
    - Month
    - County
    - Amount of each type of liquor (???)
    - Amount of each size of liquor (small, medium, large)
    - Amount of each cost bracket of liquor (cheap, normal, expensive)
    - Amount of each pack size (small, medium, large)
  + First fit to 2017-2019, then see if good predictor for 2020
  + Client B: Sale (Dollars), county, im\_desc (name of alc) OR vener\_name (name of company that creates alc), store number, city, date
    - END GOAL: what predicts quantity of alc sold
    - Classify alc as low, med, or high cost alc - sale price/amount sold in liters
    - Classify each purchase as a high or low quantity (and medium?) - look at liters var or ignore since this is what we are predicting
    - Classify each purchase as near holiday or not (2 weeks until day of holiday)
      * 4th of July, Halloween, Cinco De Mayo, etc
    - Whether store location is near a major college
    - Day of the week?
  + check categories in data to classify liquor into hard and not hard
* **10/6**

* + Have visualizations and idea of what to write for the rest