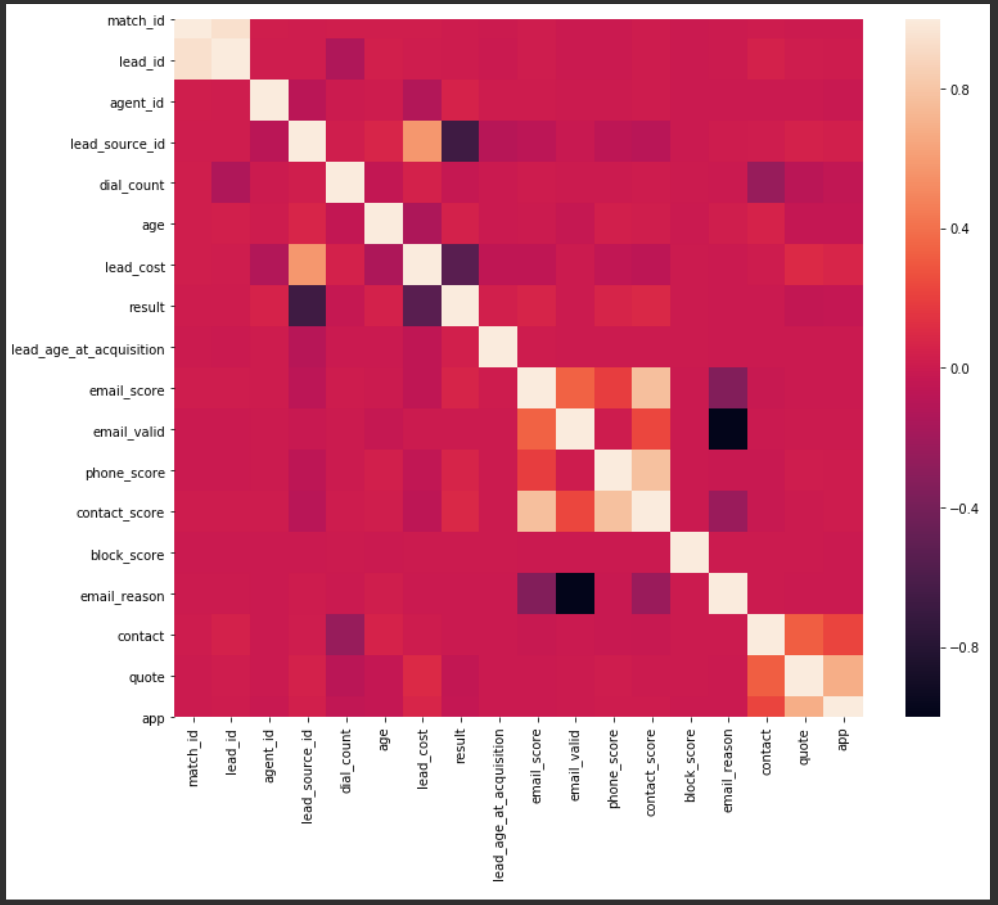
Potential Lead Analysis

Objective

- How can we utilize available data to maximize sales?
- Create strategy to pick out best potential leads

Key Features



- Most features had very weak correlation
- However, the scores look promising

Model

- Implemented a K nearest neighbor model
  - Features were the phone, email, contact, and block score
- First, run the model through all contacted leads
  - Allows model to generate accuracy when comparing predicted results to the real application results
- Run the model through all uncontacted leads
  - Choose next best customer to contact based upon a positive application result and highest combined score (phone + email + contact +/- block)

Performance

- Model will need to be modeled as more contacted results come in
- To measure immediate performance compared to other models:
  - Compare accuracy metrics
  - Which do you value more? (Less false positives, or less false negatives)
    - Personally, as the scenario is not life threatening (as say identifying if a mushroom is poisonous or not), I would prioritize less false negatives, as the worst case in a false positive is loss time/effort and perhaps a peeved customer
- To measure long term performance:
  - Regularly update model data and check accuracy metrics
  - Not only looking at accuracy, but also ability for model to adapt and learn

Future Plans

- Further classify leads into tiers
  - Percentage ranges i.e. 80-99% chance to make a sale, 60-80% chance to make a sale, etc.
- Create a model based upon base features
  - I assume that the scores given were derived already
  - Delving into how those scores were created would provide a chance to perform more statistical analysis during feature selection