Alex Wolf
5/12/13

Database Schema Plans (in CVSs)

- Each table will be in a separate directory with CSVs in them for each entry
- Tables aren't normalized on purpose so can utilize and practice PANDAS
- Data is randomly generated, but I tried to make it as realistic as possible
- Primary Key Values are Unique
- Data labeled with ***$***, means try to use to make a ML model to classify a Good_Lead for a customer

Person Table

- pId- Primary Key
  - unique 8 digit key
- FIRST_NAME
  - 'firstName'+ str(unique int)
- MIDDILE_INITIAL
  - Random letter
- LAST_NAME
  - 'lastname'+ str(unique int)
- UTC_TIMEZONE
  - -12<= Random int <= +12
- EMAIL
  - 'random unique string'+@gmail.com
- PHONE_NUMBER
  - +1- 10 random unique digits
- **(*$*)** PLATFORM_A_CONVO_COUNT
- **(*$*)**PLATFORM_B_CONVO_COUNT
- **(*$*)**PLATFORM_C_CONVO_COUNT
  - **ALL THREE** of these are added intertively when conversation is created

Customer

- pID- Primary Key
- GOOD_LEAD
  - True determined after EDA analysis

CustomerRepEmplyee

- pID- Primary Key
- COMPANY
    - ‘Google’, ‘Microsoft’, ‘Amazon’, or ‘Facebook’

Conversation Table

- ConversationID- Primary Key
- Customer_pID- Foreign Key
- CustomerRep_pID- foreign Key
- **(*$*)** LENGTH_OF_CONVO_MINS
    - Random Gaussian int >=0
    - Mean=5
    - Std = 5
- **(*$*)** IS_CALL
    - Boolean 50/50 chance
- **(*$*)** IS_TEXT_CHAT
    - Boolean 50/50 chance
- **(*$*)** PLATFORM
    - A, B, or C
    - Higher probility of A then B then C

Text

- textId- primary Key
    - unique 12 digit key
- Conversation_id- Foreign Key
- pID- foreign Key
    - The pID of the person who generated this text
- **(*$*)** TEXT
    - Randomly generated real words separated by spaces
    - For the type of words, there will be two choices and the same if in the same conversation
        - Positive words
        - Another Negatives
    - It will contain 10x as more filler words like ‘the’ ‘in’ ‘what’
    - Will have a zero centered Gaussian Distribution of the positive/negative words to choose, so some show up more frequently than others
        - Index of word chosen form this and will be only positive