

Database Schema Plans (in CSVs)

- 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
- Generated Data is made to classify a customer **GOOD_LEAD** by
 - o Number of Positive words in Text exchanges
 - o Count of conversations
 - o Type of Platform
 - More lean on A, then B, then C
 - o Company the Sales Rep works for
 - More lean on Google than others
 - o Conversation time Total with Customer Reps

Person Table

- pId- Primary Key
 - o unique 8 digit key
- FIRST_NAME
 - o 'firstName'+ str(unique int)
- MIDDLE_INITIAL
 - o Random letter
- LAST_NAME
 - o 'lastname'+ str(unique int)
- UTC_TIMEZONE
 - o -12<= Random int <= +12
- EMAIL
 - o 'random unique string'+@gmail.com
- PHONE_NUMBER
 - o +1- 10 random unique digits
- (***\$***) PLATFORM_A_CONVO_COUNT
 - o Random Int 0<= inf with mean 5 and Standard Deviation of 10
- (***\$***)PLATFORM_B_CONVO_COUNT
 - o Random Int 0<= inf with mean 5 and Standard Deviation of 10
- (***\$***)PLATFORM_C_CONVO_COUNT

- Random Int $0 \leq \text{inf}$ with mean 5 and Standard Deviation of 10

Customer

- pID- Primary Key
- GOOD_LEAD
 - True determined by metrics defined above

CustomerRepEmployee

- 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=2
 - Std = 5
- IS_CALL
 - Boolean
- IS_TEXT_CHAT
 - Boolean
- PLATFORM
 - A, B, or 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
 - 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 from this and will be only positive