1) To determine whether an agent is going to move out without a notice, do we need to consider agents who have

been terminated involuntarily by the management?

D: I’d say,  YES.

**E: Not necessarily. Predicting agents that will terminate without notice means that the company didn't see it coming, which usually isn't the case in involuntary terminations. Symptoms like absenteeism or poor performance are usually tracked over a period of time before they decide to terminate someone, so the value of predicting these types of agents is somewhat lower (and harder to train for).**

2) Records which have term\_code, term\_type and term\_reason missing values has no term\_dates also which indicates that these are those agents which never left the company?

D: sounds a reasonable assumption.

**E:  An agent is terminated if and only if s/he has a term\_date from at least one of the sources. If an agent has a term\_code etc., s/he has also been terminated, but the converse isn't true (i.e. terminated agents may not have term\_code etc.).**

3) There are multiple Call records for one agent in Call data. How to find the

criteria for data fattening in order to decide the call quality of an agent on

an average?

D: That’s a part of the project. To find creative summary statistics. Erik: do you have any insight?

**E: It's an excellent question and you are 100% correct Daniel. The person asking the question is on to something I think -- the key word is "average". How can we describe the typical or expected behavior/performance of an agent? Performance is straightforward at least, you can look at PTP rates in the call data or generated revenue in the agent data. Behaviorally speaking though, there is no direct indication in the data. You will have to make a hypothesis (e.g. "poor performing calls are correlated with poor call behavior").**

[4) There are many features present for one audio call. Is there any any way we](x-redundant-cluster-toggle://3)

[could get an idea from business about the importance of the feature vector.](x-redundant-cluster-toggle://3)

[Any definition list of the feature values could be useful.](x-redundant-cluster-toggle://3)

[D: could you give a qualitative description of the features for “groups of them”? … something like between 1-20 is about the tone of voice, 21-40 background noise etc.](x-redundant-cluster-toggle://3)

**[See More](x-redundant-cluster-toggle://3)**

[4) There are many features present for one audio call. Is there any any way we](x-redundant-cluster-toggle://3)

[could get an idea from business about the importance of the feature vector.](x-redundant-cluster-toggle://3)

[Any definition list of the feature values could be useful.](x-redundant-cluster-toggle://3)

[D: could you give a qualitative description of the features for “groups of them”? … something like between 1-20 is about the tone of voice, 21-40 background noise etc.](x-redundant-cluster-toggle://3)

**E: I purposely wish to keep the calculations secret.... but I guess broadly categorizing them wouldn't hurt. Daniel, please make sure everyone gets this information since it might prove to be an advantage.**

**\* feature\_values 1-24 are measurements and statistical calculations on the speech of the call (total speaking time, etc.).**

**\* feature\_values 25-34, 45-54, 65-74, 85-94, 105-114, 125-134, and 145-154 are statistical calculations on speech that has been identified as expressing a negative emotion (e.g. anger).**

**\* feature\_values 35-44, 55-64, 75-84, 95-104, 115-124, 135-144, and 155-164 are statistical calculations on speech that has been identified as expressing a positive emotion (e.g. joy).**

**\* feature\_values 165-176 measure proportions of negative/positive emotions, emotional/unemotional speech, and transitions from one to the other (positive to negative or unemotional to emotional for example).**

**To speak to Daniel's example, things like tone of voice are extracted from the audio signal and are intermediate feature calculations that are then fed into a proprietary predictive model that assesses the actual expressed emotion. It's these emotional prediction estimates that are used for the features given. Things like noise and music are not considered at all.**

5) The hms is the unit of Call data. Is it hr:mm:sec?

D: It seems obvious to me, YES.

**E: This was answered in the doc. HMS is a misnomer, the call durations are in seconds only.**

6) There are certain column like jul\_group, aug\_group etc. which have values

like #N/A, NULL, Blank, NA.What is difference between all these?

D: Is there information beyond different types of missing?

**E: Please refer to the answer doc, morning batch answer 23.**

7) What does account mean in Call data?

D: customer account number?

**E: Yes.**

8) If  the  Rec\_Status  identifies  what audio files  are  present in the  feature dataset?

D: The question does not much sense. Erik, can you guess what they mean?

**E: This same question was asked initially. I also had no clue what the person meant.**

9) There are two sources for  payroll\_id, hire\_date, term\_date, work\_shift

 which column should  as primary between these two if data are mismatching between source1 and source2

D: I think you said you used source 2 when 1 was not present. Correct?

**E: See morning batch answers 2-6 & 11, evening batch answer 9. Essentially, the sources should be trusted equally. If they mismatch, it is up to the student to determine which value is the correct one.**

10) What does the column  group\_src1  mean?

**E: Same as above.**

11) If we assume it as name of the team that an agent for does it mean if the value is blank?

**E: Morning batch answers 6, 11, 23.**

12) There are missing values in  term\_code , term\_type , term\_reason should we consider those ag employed with the company or as missing details?

**E: See answer to question 2 above.**

13) Almost 106 rows have missing term\_code , term\_type , term\_reason  values

D: I got this comment from other people, so the complete cases for agents is down to 116 or so. So we have a little bit more than 100 agents with complete information. Is that what you also see in your preliminary analysis?

**E: By my count, 90 agents were employed for the entire time which means 208 were terminated some time. 106 records have no term\_code, which by my answer for question 2 above means there are 16 terminated agents (roughly 7.7% of all terminated agents) that have no explanation as to their termination (but they most certainly were still terminated).**

14) There are groups for each month as Jul\_group , Aug\_group is the name of the group or team an a for in that particular month ?

**E: Morning batch answer 11. A month's group value for an agent supercedes the specification given by src1 or src2.**

15)There are missing values(blank) and “#NA” in these columns(Jul\_group,Aug\_group..) is there any d meaning between blank and “#NA” Jul\_hours  means the number of hours an agent worked in the month of July?

D: Same as above related to different meaning of missingness.

**E: Yes Daniel.**

16) What does different values in column REC STATUS mean?

**E: Morning batch answer 1.**

17) What does the column  SKILL NAME mean is it related to group name in agent data?

**E: Morning batch answer 15.**

18) Agent data is from Jul-Dec 2015, Call data is from 18 Mar – 31 Dec and Feature data is from 26 May – 31 D 2015 . So can we consider only months that are intersection of all these three sheets that is from July December 2015 and remove all other rows?

**E: Evening batch answer 23.**