

- 2.
- a) Binary, Quantitative, ordinal
  - b) Continuous, Quantitative, ratio
  - c) Discrete, Quantitative, ordinal
  - d) Continuous, Quantitative, ratio
  - e) Discrete, Quantitative, ordinal
  - f) Continuous, Quantitative, ratio or interval
  - g) Discrete, Quantitative, ratio
  - h) Discrete, Quantitative, nominal
  - i) Discrete, Quantitative, ordinal
  - j) Discrete, Quantitative, ordinal
  - k) Continuous, Quantitative, ratio (interval possibly)  
(Continuous)
  - l) Discrete, Quantitative, ratio
  - m) Discrete, Quantitative, nominal

3. I agree with Boss.

The measure of satisfaction =  $\frac{\text{number of complaints}}{\text{number of reviews}}$

b) It is a ratio type attribute.

4. a) Yes, he is in trouble. No, the approach not work for generating an ordinal

ranking, because it is impossible to rank if A is better than B, B is better than C, C is better than A. (It's like  $A > B, B > C, C > A$ )

b) First only ask employees for two pairwise. When there appears two products are better than one product, we can then ask employees about ranking those two. (e.g.  $A > C, B > C$ , we can then ask employee to

rank A & B)

c) ranking is an ordinal attribute, and it cannot be added or multiplied.

So I would like to choose ~~mode~~ of the ranking as its variation.

- | median

## Video

a) Get value from data

Turn data into insights

Turn insights into impact

b) Data Explorer: with a nice mix between engineering and CS principles and math.

Not a strong position in this role.

not require them to harden this thing.

Data Engineer: often have a superset of skills of Data Explorer.

tend to have forsaken their shell hacking obsession

start to be obsessed with data, love data

ML Specialist: understand how other pieces work

play multiple roles

interest in math side

physics background / applied math

| Data Explorer   | Data Engineer  | Machine Learning Specialist  |
|---|--|--|
| <ul style="list-style-type: none"><li>Objectives:<ul style="list-style-type: none"><li>Prototype</li><li>Explore</li><li>Collaborate</li><li>Share</li></ul></li><li>Product Features Targeted - Workspace<ul style="list-style-type: none"><li>SQL Notebooks</li><li>Graphing &amp; Visualization</li><li>Comments &amp; sharing</li><li>Annotations (Markdown)</li><li>Dashboards for projecting in meetings &amp; emailing</li></ul></li></ul> | <p>Objectives:</p> <ul style="list-style-type: none"><li>Prototype (same as Data Scientist), and then...</li><li>QA, test</li><li>Operationalize</li><li>Maintain, (e.g. cut releases)</li></ul> <p>Product Features Targeted – Workspace &amp; Platform</p> <ul style="list-style-type: none"><li>Using libraries (importing &amp; using notebooks as)</li><li>Setup/manage recurring ETL and long-running apps</li><li>Integration (e.g. with monitoring)</li><li>Streaming input</li><li>Export</li></ul> | <p>More rare role - very specialized</p> <p>Objectives:</p> <ul style="list-style-type: none"><li>Develop, test, improve statistical/ML models</li></ul> <p>Product Features Targeted – Workspace</p> <ul style="list-style-type: none"><li>Importing open source libraries (liblinear, mllib)</li><li>Sharing</li></ul> |

c) A set of process extracting data from an input source, transforming the data, and loading into an output destination such as a database.