RA:

Working model and somehow word cloud for each hierarchy

Create the schema db (from last week, change names later)

Load data into fact table and dimensions, time taken much faster on new system

Graphical user interface, text, application

Description automatically generated

The details also look about right

Text, email

Description automatically generated

Onto python

Pip install mysql-connector-python

Connect to DB via localhost and allow the cursor to make a link. Since it’s localhost it’ll always connect

Text

Description automatically generated

We can obtain the topic levels from each of the three levels and look at the individual entry, but in the form SQL returns this:

Text

Description automatically generated

The formatting won’t work, as we need a dictionary and the connector returns ALL of this as a single String (ugh). We’ve got some work to do.

First, to start small, we’ll work with just one topic. To split it into a list, we’ll need to get rid of the brackets and parentheses before splitting on commas, then zipping into a dictionary. We then have to replace the other extra spaces and characters to cleanse it properly.

Text

Description automatically generated

Initially for for loop I made the mistake of for t in test and trying to replace each item, which didn’t work (that’s Python 101, duh). So changed to this instead:

Text

Description automatically generated

Same as the above.

Just standard cleansing and rearranging. We’ll also make the weights floats now ahead of time, just to save the headache of having to convert types in a dictionary later.

Text

Description automatically generated

And we get:

Text

Description automatically generated

Much better. We then use the iterator trick to zip the list into a dictionary

Text

Description automatically generated

Text

Description automatically generated

And finally we run the code we used last week for word clouds and get the following based on weights:

Text

Description automatically generated with low confidence

Although there doesn’t seem to be a way to have it look consistent, I’ll dig into the settings and config of word cloud a bit more.

For reference, another time I run it and it looks like this.

Text

Description automatically generated with low confidence

More adjective/word user-specific

This is just for topic level 3. We can do the same for topic level 2 and 1, will optimize code sometime later so that we can just loop it really quick.

Topic Lvl 2:

A picture containing text

Description automatically generated

What they are doing with it

Topic Lvl 1:

Text

Description automatically generated

To put all three together in one could work, but also

OK, also I only realized here that the initial result wasn’t just Strings but tuples that could have just been turned into dictionary via dict instead of all the work above. Oh well…

Chart, bar chart

Description automatically generated

This plot shows different reviews made for product one on different date SKs. Not only are these different, but as expected the topic SKs are different as well, since groupings of words differ for each review. It seems that while the initial two customers liked the product, the subsequent three had lower opinions on it. We can compare this with the data in our data mart.

Table

Description automatically generated

Bingo