* Create database – 4 tables dim\_prodcuts, dim\_region, dim\_sales\_representative, fact\_data (star schema)
* Create excel file for region and sales representative.
* Create yaml file
* Create db\_connect.py,

1. Method 1 - Read yaml file
2. Method 2 - Initialise the yaml file (create engine)
3. Method 3- To test connection, write method and display list of tables.
4. Method 4 – upload the data

* Create Data\_transfer.py file

1. Read data from Excel file
2. Upload data into region and sales representative file
3. Read each folder with for loop folder\_list = os.listdir("data/fact\_data")
4. Read each file from folder
5. Reach each sheet from folder
6. From each sheet get data. Create new\_df for each sheet. Append that in the list.(you will have all data from all folders)
7. From that list take one item at a time.
8. Get product name drop other columns
9. Add product name into list.
10. Append the list which contain name of products are in all sheets.
11. Create a product dataframe with 2 columns product name and id.
12. Convert product name into product id. (should be alphanumeric without special char)
13. Drop duplicate look for drop\_duplicate and subset in panda tutorial.
14. Look for numpy tutorial for np.select loop. (product category column)
15. Now you have 3 column product name, id, category.
16. Upload the dataframe to dim\_product.