Create yaml file.

Create database, win\_medicare

* Table name dim\_products (product\_id, product\_name,product\_category) – primary key (product\_id)
* Table name dim\_region (region\_id,region\_name) – primary key (region\_id)
* Table name dim\_sales\_representative (sales\_rep\_id, sales\_rep\_name,region\_name) – primary key (sales\_rep\_id)
* Table name fact\_data (order\_id, product\_id,region\_id, sales\_rep\_id, month, year, quantity, unit\_price, total\_price) – primary key (order\_id), foreign key (product\_id, region\_id, sales\_rep\_id)

Create a connection file in python called db\_connect.py

* Create class called DatabaseConnector
* Create method in class called read\_db\_creds() – read the yaml file
* Create method called init\_db\_engine() – will connect to database
* Create method called upload\_to\_db() - will upload your data into database
* Create method called upload\_to\_fact\_data() - will upload your data into database

Create data\_transfer.py file

* Create instance for above class.
* Get product, region and sales\_rep excel file.
* Upload data into database.

Make the changes in Excel file for each city. Match the structure with fact\_data table.

Create python file name get\_data\_from\_db.py

* Create class GetData
* Create method fetch\_data()
* Create get\_region\_id()
* Create get\_product\_id()
* Create sales\_rep\_id()

Create python file named read\_file\_from\_folder.py

* From each sheet, take product name remove the spaces called that column as product id.
* Get region id and sales rep id
* Add it in dim\_products database.
* Upload to fact\_data

Create python file sales\_rep\_data\_transfer.py

* Get Nanded\_data and make it ready like fact\_data structure.

Report

* Most ordered product and category region wise
* Most sale month wise