Assignment2Code

March 23, 2023

[ ]:
```
import pandas as pd # needed for most operation
import numpy as np # needed for some array operations
from sqlalchemy import create_engine # needed for DB connection

inputfile = 'BI_Raw_Data.csv'

# Read the CSV file into a pandas dataframe
df = pd.read_csv(inputfile, delimiter=';', encoding="ISO-8859-1")
```

[ ]:
```
# Create the product table
product = df[['Product_Name', 'Product_Category']].drop_duplicates()
product['productid'] = np.arange(1, len(product)+1)
product = product.rename(columns={'Product_Name': 'name', 'Product_Category':␣
 ↪'category'})
product = product[['productid', 'name', 'category']]
```

[ ]:
```
# Create the customer table
customer = df[['Customer_Name', 'Customer_Country']].drop_duplicates()
customer['customerid'] = np.arange(1, len(customer)+1)
customer = customer.rename(columns={'Customer_Name': 'name', 'Customer_Country':
 ↪ 'country'})
customer = customer[['customerid', 'name', 'country']]
```

[ ]:
```
# Create the sales table
sales = df[['Order_Date_Year', 'Order_Date_Month', 'Order_Date_Day',␣
 ↪'Customer_Name', 'Product_Name', 'Product_Order_Price_Total']]

sales = pd.merge(sales, product, how='inner', left_on= ['Product_Name'],␣
 ↪right_on = ['name'])
sales = pd.merge(sales, customer, how='inner', left_on= ['Customer_Name'],␣
 ↪right_on = ['name'])

sales = sales.rename(columns={'Order_Date_Year': 'year', 'Order_Date_Month':␣
 ↪'month', 'Order_Date_Day': 'day', 'Product_Order_Price_Total': 'sales'})
sales = sales[['year', 'month', 'day', 'customerid', 'productid', 'sales']]
```

[ ]:
```
# Output the resulting tables to CSV files
customer.to_csv('output/customer.csv', index=False)
product.to_csv('output/product.csv', index=False)
sales.to_csv('output/sales.csv', index=False)
```

[ ]:
```
# first create link to database
# Replace username with the user name password with the password
driver='postgresql'
username='dab_ds22232a_46'
dbname=username # it is the same as the username
password='5wQ5aeeIp3Xaobd6'
server='bronto.ewi.utwente.nl'
port='5432'
# Creating the connetcion pool for SQL
engine = create_engine(f'{driver}://{username}:{password}@{server}:{port}/
 ↪{dbname}')
product.to_sql('product', engine,schema='ass2', index=False,␣
 ↪if_exists='replace')
customer.to_sql('customer', engine,schema='ass2',index=False,␣
 ↪if_exists='replace')
sales.to_sql('sales', engine,schema='ass2',index=False, if_exists='replace')
```

[ ]:
```
155
```