

Design Schematic

For our project, Kurt and I decided to use the provided schematic. The provided schematic makes the most sense with the 3 dimensions. The only fact that changes per location would be the price, hence the reason it stays in the fact table. The proposed schematic can be seen in figure 1 below.

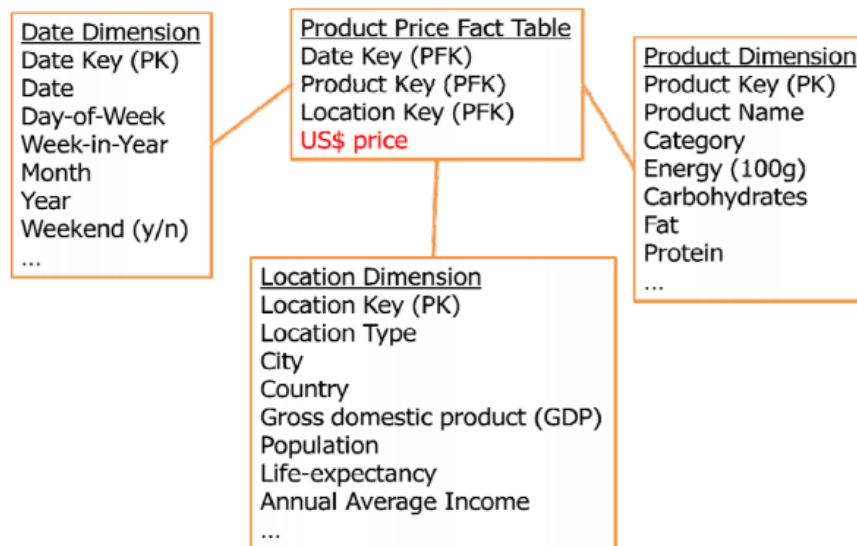
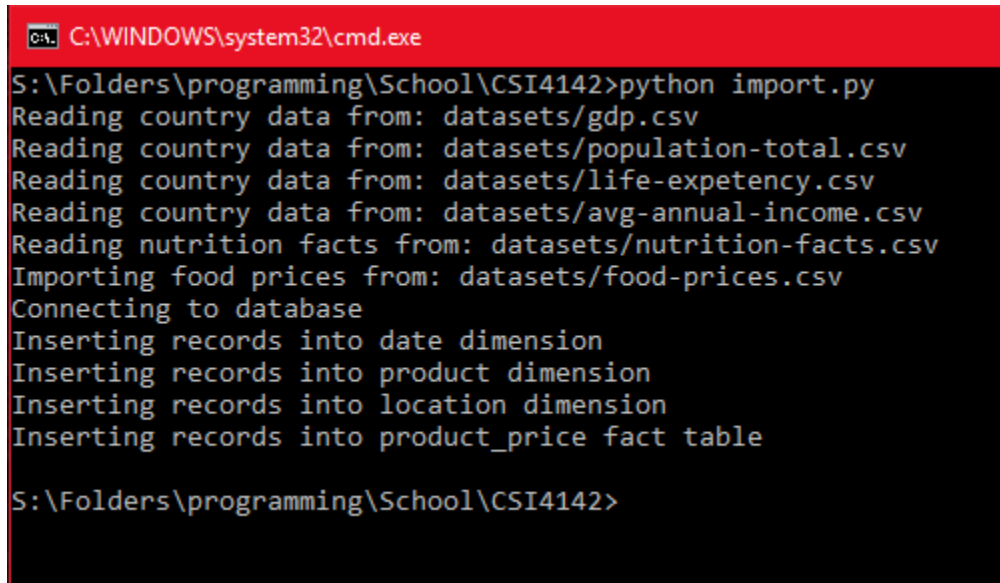


Figure 1: Recommended schema

To build the database a python script was created to parse the data from multiple CSV files and insert it all neatly into a PostgreSQL database. The schema for the database is provided in the file “schema.sql” while the import script can be found in “import.py”. All of the datasets which are used during the import script can be found in the “datasets” directory. The datasets were retrieved from reputable sources such as the world bank.

The script extracts the relevant data from the CSV files and then creates the correlations required. This builds the fact table and all of the dimensional tables. Finally, surrogate keys are generated to ensure all of the data is properly linked. In the case that data doesn’t match from the related sets the data is left as null values. We chose to leave these values as nulls since we didn’t want to mix multiple datasets in case they sample the data differently.

To prevent duplicate data the script, when run multiple times, will update the existing data instead of inserting duplicates. This ensures that the data is clean and duplicate free. To see sample run of the output of the script please see figure 2.



```
C:\WINDOWS\system32\cmd.exe
S:\Folders\programming\School\CSI4142>python import.py
Reading country data from: datasets/gdp.csv
Reading country data from: datasets/population-total.csv
Reading country data from: datasets/life-expetency.csv
Reading country data from: datasets/avg-annual-income.csv
Reading nutrition facts from: datasets/nutrition-facts.csv
Importing food prices from: datasets/food-prices.csv
Connecting to database
Inserting records into date dimension
Inserting records into product dimension
Inserting records into location dimension
Inserting records into product_price fact table

S:\Folders\programming\School\CSI4142>
```

Figure 2: Sample output from import.py script