# ARIMA and LSTM Univariate Preprocessing Steps

1. Read in prices\_production table from PGAdmin

2. Clean the conventional field

3. Drop non-beneficial columns; year\_month, status an prod.

4. Create plots to see what the data looks like. I noticed some seasonality differences in prices during the summer months.

# LSTM Multivariate Preprocessing steps

1. Read in the prices\_climate data from PDAdmin

2. Dropped non-beneficial columns; year\_month, price\_total\_volume and total\_bags. year\_month is not needed as date is available. price\_total\_volume is the sum of units\_4046, units\_4225 and units\_4770 and total\_bags. Total\_bags is the sum of s\_bags, l\_bags and xl\_bags. It would be redundant information to double count those parameters and unfairly skew the model.

3. For the same reasoning, I dropped those rows that are totals of underlying sub areas (cities/regions).

4. Corrected the type column. It had two formats for Conventional. Should be one.

5. Split the data out into prices/climate data restricted to type = conventional and the prices/climate data restricted to type = organic. There are now three dataframes.

6. Viewed the data to look for trends. Noticed an increase in average prices in the summer months. Noticed an increase in average temperatures in the summer months, which seems obvious. I'm not saying there is a correlation between the two but definitely similarities. Will come back to this in the results section.

7. I saved the three cleaned files for continued use.