2. ETL DETAILS

2.1 Data Import/Extract Sources and Method

The website that we used to collect the data on the price of Smithfield Bacon was [www.Kroger.com](http://www.Kroger.com). We used web scraping to collect the price of bacon. What will be loaded is the price of Smithfield bacon to our PG Admin.

2.2 Data Acquisition

The range for the Smithfield Bacon is dynamic, the price always changes, so the price is never the same, it is always updating. To obtain the price we had to scrape Kroger website to look for the price and then store it as a list to convert it into a dataframe and then extract it to PG Admin. Some special attributes are, if there is a special price on the bacon when it changes, then the code has to be updated.

2.3 Data Transform

What we had to clean up was to collect only the price, so we had to strip the data to reveal just the price and the date.

2.4 Data Integrity

The reliability of the source data is always updating week to week. If there are any notifications it can be received thru an email. If the price for bacon has been changed then a notification will be sent out thru email.

2.5 Data Refresh Frequency

The frequency which this ETL process will be weekly

2.6 Data Security

No data security