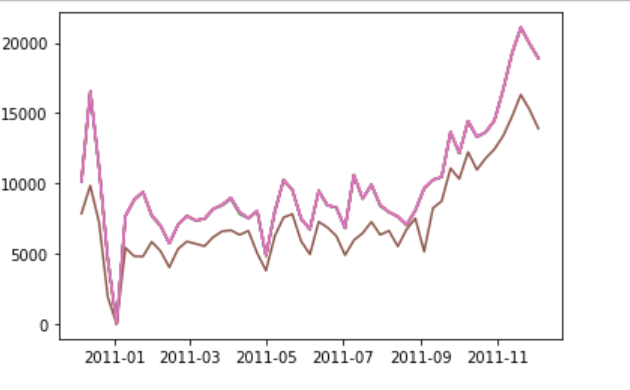
**Data Set Information:**

This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.

**Project Goal:** Calculate LTV and predict when customers leave.

1. Uploaded dataset onto Jupyter. Warning is that 42MB.
2. The data range is a year (see line graph plotted by week)



1. Describe function reveals 8 columns.

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 541909 entries, 0 to 541908

Data columns (total 8 columns):

InvoiceNo 541909 non-null object

StockCode 541909 non-null object

Description 540455 non-null object

Quantity 541909 non-null int64

InvoiceDate 541909 non-null datetime64[ns]

UnitPrice 541909 non-null float64

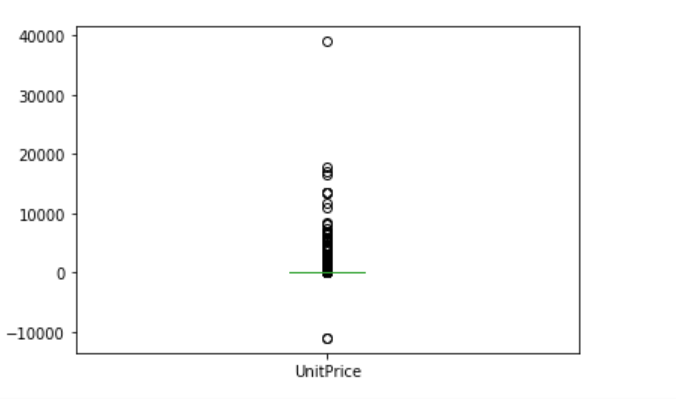
CustomerID 406829 non-null float64

Country 541909 non-null object

dtypes: datetime64[ns](1), float64(2), int64(1), object(4)

memory usage: 33.1+ MB

1. Summarize indicate some faulty datab. For example, units per invoice has an outlier and faulty data.



1. Unique counts reveal:

InvoiceNo 25900 InvoiceDate 23260 Country 38 StockCode 4070 Description 4223 UnitPrice 1630 CustomerID 4372

Hence, 5.9 purchase invoices per customer.