Data cleaning

* Trim spaces from character columns
* Remove completely empty strings
* Fill NA descriptions with the most common description for each stock code
* Remove rows with POST and DOT since postage doesn’t actually contribute to sales
* Remove rows with “B” stock code, this was for adjust bad debt, some kind of charge back

Rows not fully removed

* Negative quantities stored in separate dataframe in case its needed later
* Negative prices stored in separate dataframe in case its needed later
* Zero prices store in own dataframe

Rows to remove depending on model

* Rows containing “M” will need to be removed when doing stock modeling
  + Left them for now since they have revenue and stock codes associated with them
* Rows with missing Customer ID removed when modeling customers but should probably be retained when modeling stock or revenue since they still have those values associated with them

EDA

* Data spans 12-01-2009 – 12-09-2010
* Unique Products: 4248
* Unique Customers: 4310
* Unique Countries: 40
* Unique Transactions: 20878
* Average Transactions per Month: 1606
* Average Transactions per Week: 393.9
* Average Revenue per Month: £ 780,029
* Average Revenue per Week: £ 191,328

Which countries contribute the most to Sales growth?

* UK, Ireland, Netherlands, Germany, France

What are the top-selling products?

A graph of a bar chart

AI-generated content may be incorrect.

Top customers

* Most of the revenue isn’t attributable to a customer id
* A screenshot of a computer

  AI-generated content may be incorrect.

Sales seasonal trend

A graph showing a line graph

AI-generated content may be incorrect.

A graph with a line and a line

AI-generated content may be incorrect.

A graph of sales and sales

AI-generated content may be incorrect.

K means Clustering – customer segmentation

* Wholesale customers contribute to most of the revenue, they’re also the most recent and frequent
* Casual customers aren’t frequent at all really and have large spans of time between visits

A screenshot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.