1. **Data fields**

train.tsv, test.tsv

The files consist of a list of product listings. These files are tab-delimited.

* train\_id or test\_id - the id of the listing
* name - the title of the listing. Note that we have cleaned the data to remove text that look like prices (e.g. $20) to avoid leakage. These removed prices are represented as [rm]
* item\_condition\_id - the condition of the items provided by the seller
* category\_name - category of the listing
* brand\_name
* price - the price that the item was sold for. This is the target variable that you will predict. The unit is USD. This column doesn't exist in test.tsv since that is what you will predict.
* shipping - 1 if shipping fee is paid by seller and 0 by buyer
* item\_description - the full description of the item. Note that we have cleaned the data to remove text that look like prices (e.g. $20) to avoid leakage. These removed prices are represented as [rm]

1. **Breaking codes in descriptions:**

Breaking the Code:

BNWT BNWOT BNIP MIB MWOB NWT NWOT NIB BNIB BNWOB

What do these abbreviations mean? …Read on and find out.  
It is important to understand these abbreviations to deal with item\_descriptions:

BNWT: Brand New With Tags

NWT: New With Tags

BNWOT: Brand New WithOut Tags

BNWOB: Brand New With Out Box

NWOT: New With Out Tags

BNIP: Brand New In Packet

NIP: New In Packet

BNIB: Brand New In Box

NIB: New In Box

MIB: Mint In Box

MWOB: Mint With Out Box

MIP: Mint In Packet

MWOP: Mint With Out Packet

1. **Codes of Item\_condition:**

Decode Item conditions:

1 - New (codes: BNWT, BNIP MIB, NWT, NIB, BNIB)

2 - Like New (codes: BNWOT, MWOB, NWOT, BNWOB)

3 - Good

4 - Fair

5 – Poor