

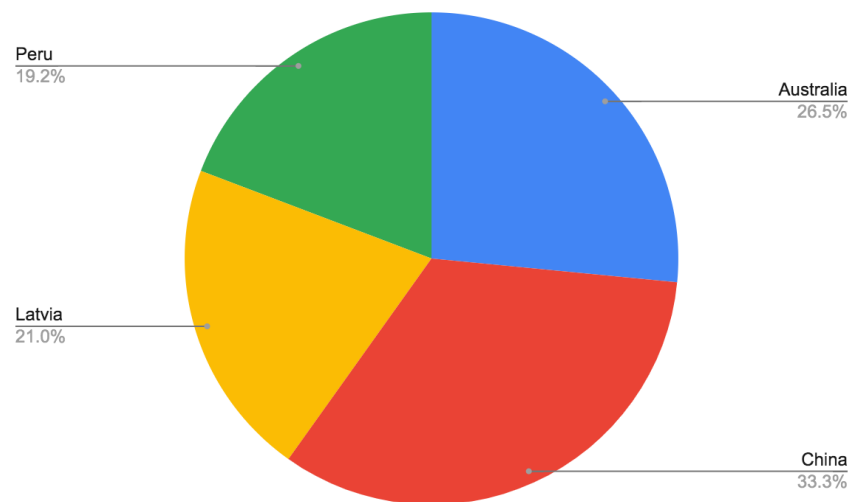
The data set has **511427** events.

There are **12** duplicated events.

There are **68061** events with invalid values, this could be to a non numeric value *total_orders* or *voucher_amount* or by an invalid date.

There are **4** countries in the data set: China, Peru, Australia and Latvia.

Event distribution per country:



For Peru this is the **event distribution per segment**:

Peru					
frequent_segment	Voucher amount	Number of events	recency_segment	Voucher amount	Number of events
0-4	2640	12544	30-60	2640	20374
	3520	5982		3520	7330
	4400	5683		4400	7189
13-37	2640	11813	60-90	2640	12960
	3520	4391		3520	5782
	4400	4225		4400	5611
5-13	2640	4112	90-120	2640	8063
	3520	1374		3520	5500
	4400	1272		4400	5308

The data is highly skewed since the amount of events with voucher_amount = 2640 is much higher than every other amount. **As a result the most used voucher amount per segment is always 2640(!).**