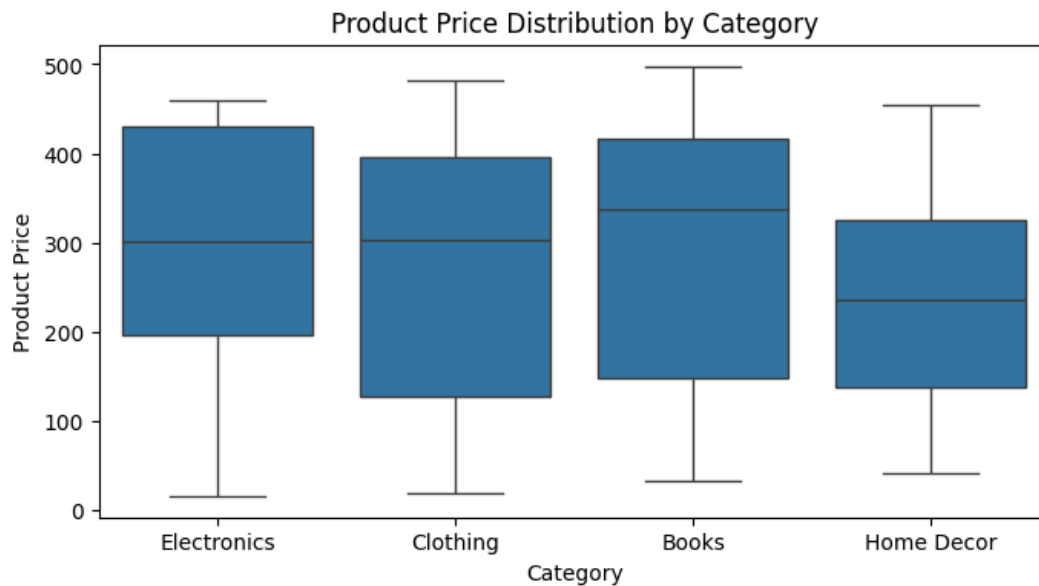


Business Insights

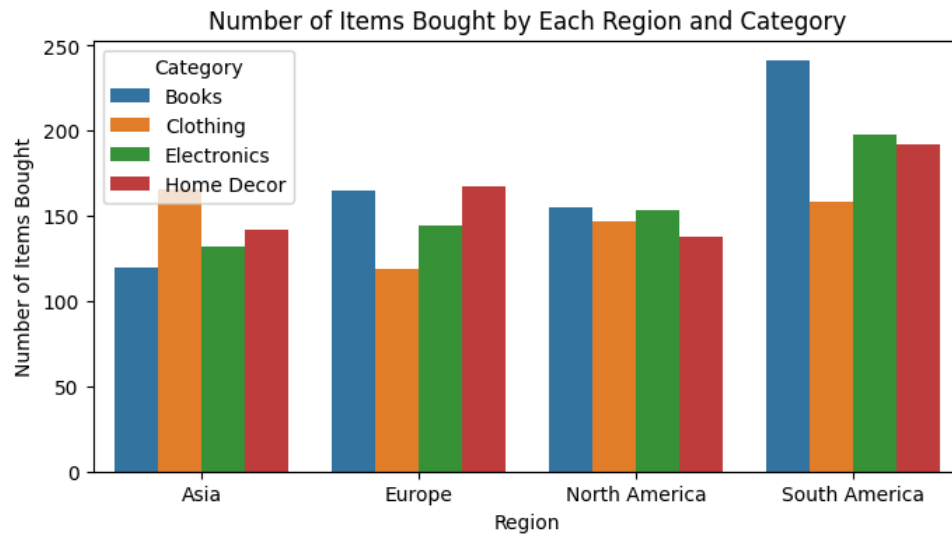
1) Price per Category



Majority of items from categories “Electronics”, “Clothing” & “Books” have similar price, but clearly “Home Decor” items are relatively cheaper.

“Good to know people spending the most on books and the least on luxury items like home decor & clothing stuff.”

2) Quantity of items bought by region and category

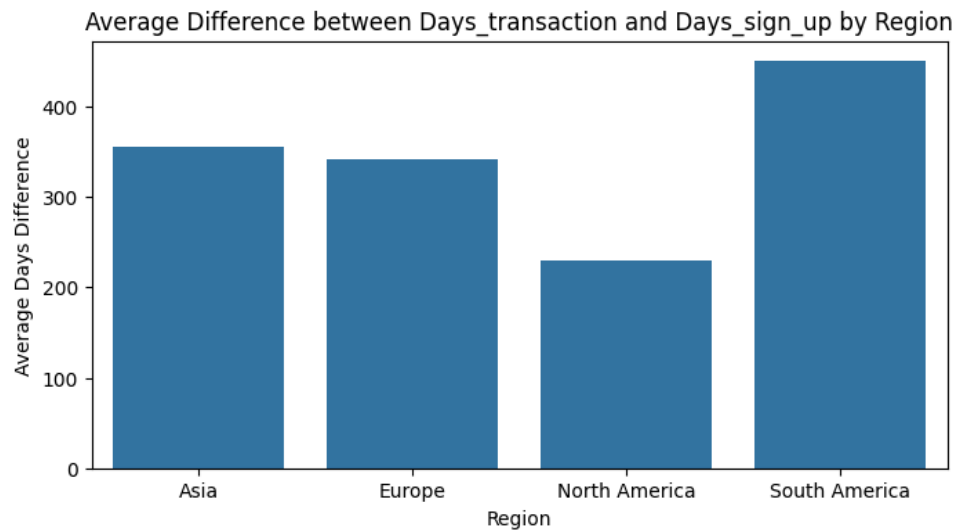


-> The Western nations invested the most on books while Asians spent more on clothes. (This resonates with the major Gen-Z population here in Asia.)

-> The South Americans spent the most on books (on average) which had the highest price.

“ It’s wise to put more books into the South American market, more clothes into the Asian market, and home decor & electronic items everywhere.”

3) Time taken to decide to make a transaction

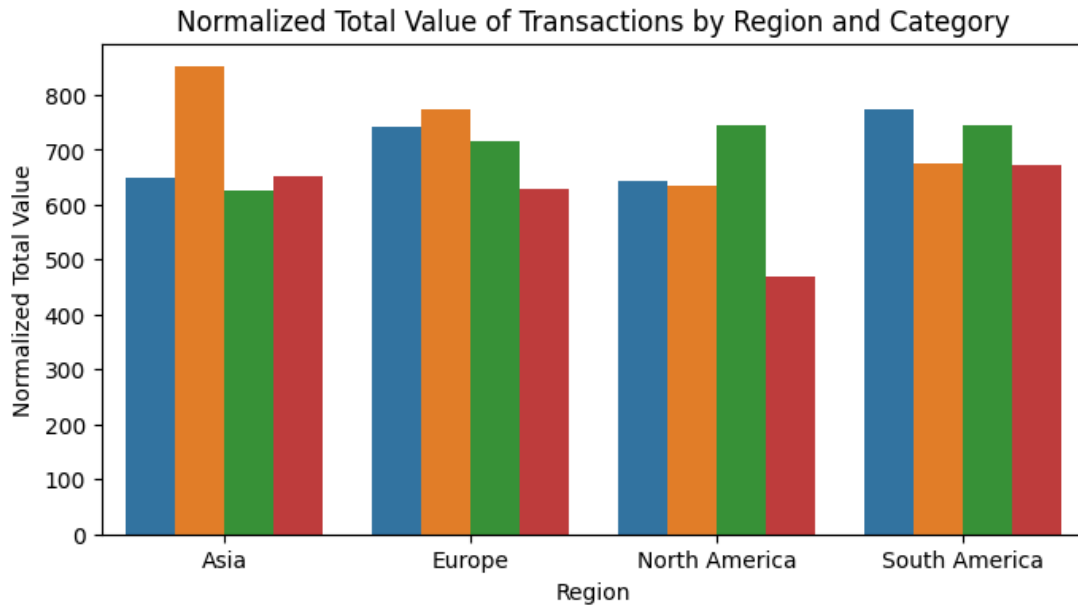


-> North Americans don't ponder over much before making a transaction, so one can think put expensive items in that region.

-> South American market is more careful about their purchases, which means want more value for their spending, so one should invest more in meaningful stuff like books, educational stuff, etc.

"South Americans are more careful with their monetary spendings, unlike their neighbours."

4) Money spent on each category in a certain region by a single person



-> Wow! Asians are hyping up clothes, sell it more there.

-> North America is buying less home decor items - maybe lower the prices there.

-> South Americans are smart, they spend less on non-essentials.

“Fancy Gen-Z items should be sold more in Asia, lowering price of essentials can be an option and making meaningful additions to Europe and South America where people buy smart.”

For more details on the dataset and it's features, please see below.

Customers :

-> No missing values

-> Dataset seems to be balanced

```
Region
South America    59
Europe           50
North America    46
Asia             45
Name: count, dtype: int64
```

Products :

-> No missing values

-> Dataset seems to be balanced

```
Category
Books        26
Electronics  26
Clothing     25
Home Decor   23
Name: count, dtype: int64
```

Transactions :

-> No missing values

-> Numeric columns need to be scaled

	Quantity	TotalValue	Price
count	1000.000000	1000.000000	1000.000000
mean	2.537000	689.995560	272.55407
std	1.117981	493.144478	140.73639
min	1.000000	16.080000	16.08000
25%	2.000000	295.295000	147.95000
50%	3.000000	588.880000	299.93000
75%	4.000000	1011.660000	404.40000
max	4.000000	1991.040000	497.76000

EDA

- > Replaced Date columns to number of days from today. Both provide the essence of time.
- > Values of columns - “**SignupDate**” & “**TransactionDate**”, so I converted all of them to number of days from dates for ease of use.
- > Merging the dataframes - “transactions” & “customers” on column “CustomerID”
- > Merging the dataframes - “transactions” & “products” on column “ProductID”
- > **merged_df = customers + products + transactions**

I thank the people responsible for formation of the dataset :

```
TransactionID      0
CustomerID         0
ProductID          0
Quantity           0
TotalValue         0
TransactionPrice   0
Days_transaction   0
CustomerName       0
Region            0
Days_sign_up       0
ProductName         0
Category           0
ProductPrice       0
dtype: int64
```

- > No NULL values
- > All values of TransactionPrice & ProductPrice are the same, so we can delete one of them.