**Shopee Python-Pandas Test (45 mins)**

In this task, you'll be analysing listings data from our Shopee Platform.

You may use the Pycharm IDE installed, Sublime or other windows native text editors. Please save your python source code on the desktop. You may use the internet for help.

The dataset is stored in the Test\_Pandas.xlsx file. It contains listing information posted on Shopee. One single listing corresponds to one row in the dataset.

The dataset has 12 columns, and 464433 rows.

**Here are the brief descriptions of each column:**

Itemid - a unique ID of the product

Shopid - a unique ID of the shop

item\_name - product title

item\_description - detailed product description

item\_variation - stores variations of a product (e.g. different colours or sizes, in the format like {variation 1 name: variation 1 price, variation 2 name: variation 2 price})

price - how much does the item sold

stock - how many stocks left

category - which category does the product belongs to

cb\_option - 1 indicates the product is sold by a cross border shop

is\_preferred - 1 indicates the product is sold by a preferred shop

sold\_count - how many products have been sold

item\_creation\_date - when are the product uploaded by the seller

# 1. Use pandas function to read the Test\_Pandas.xlsx file in:

a. Assign the result to a variable named “df”

b. Assign all column names to a variable named “columns”

# 2 Total number of unique items for each year (in terms of item creation date)

# 3. Total number of unique SKUs for each year and each month in terms of item creation date. (result should show a clear split of the count of unique items under each individual month under each year)

# 4. Top 10 items that have the largest inventory value

# 5. Top 3 cross border shops that have the highest amount of stocks

# 6. Top 3 Categories that have the largest number of unique preferred shop items

# 7. Remove all shops with revenue of zero and show the number of shops that are in each quartile based on revenue (E.g. 0-25%, 25%-50%, 50%-75%, 75%-100%) (Assumption: the product price has not been changed.)

# 8. Find number of products that are both CB and preferred that have more than 3 variations (do not include products with 3 or fewer variations)