4. Exclude low spending customers from dataset

steps:

- make if-statement and new flag column for low spending customers ('max_order' <=5 orders)
- create new subset df excluding these customers
- export df as pickle

```
# for-loop if-statement: assign 'drop' to customers with <=5 'max_order'

result = []

for max_order in opc_merge['max_order']:
  if max_order <=5:
    result.append('exclude')
  else:
    result.append('keep')
```

```
# input result in new column named 'low_activity_customers'

opc_merge['low_activity_customers'] = result
```

```
# checking if new column was added

opc_merge.head(1)
```

ne	...	last_name	gender	state	age	date_joined	num_of_dependants	fam_status	income	_merge	low_activity_customers
da	...	Nguyen	Female	Alabama	31	2019-02-17	3	married	40423	both	keep

```
# counting keep vs exclude

opc_merge['low_activity_customers'].value_counts()
```

```
keep       30199150
exclude     2235339
Name: low_activity_customers, dtype: int64
```

```
# creating a subset df for 'keep' customers

keep = opc_merge.loc[opc_merge['low_activity_customers']=='keep']
```

```
# checking that row count of new df matches 'keep' count of above value_count()

keep.shape
```

```
(30199150, 36)
```

```
# exporting subset df 'keep'

keep.to_pickle(os.path.join(path, '02 Data','Prepared Data', 'orders_products_customers_merged_keep.pkl
```