Ben Deatsman

Classwork 4c

```
In [4]: from completejourney_py import get_data

cj_data = get_data()
cj_data.keys()

import pandas as pd
products = cj_data['products']
transactions = cj_data['transactions']
```

```
In [45]: size_filter = products['package_size'].str.contains('^\d{2,}(\.)?.*lb', case=False, na=False)
(products[size_filter]
 .merge(cj_data['transactions'] , how='inner', on='product_id')
 .groupby('product_type', as_index=False)
 .agg({'sales_value': 'sum'})
 .nlargest(5, 'sales_value')
)
```

/var/folders/5f/lxv_8pyj23qf8bsmqpwy4y840000gn/T/ipykernel_65663/2091088215.py:1: UserWarning: This pattern is interpreted as a regular expression, and has match groups. To actually get the groups, use str.extract.
  size_filter = products['package_size'].str.contains('^\d{2,}(\.)?.*lb', case=False, na=False)

Out[45]:

	product_type	sales_value
3	BANANAS	17480.66
28	GRAPES RED	9999.96
29	GRAPES WHITE	8467.86
61	POTATOES RUSSET (BULK&BAG)	7805.49
22	DRY DOG FOOD PREMIUM (ALPO/PUR	7146.87

```
In [ ]:
```