

Reading the Data

```
In [1]: import pandas as pd
import numpy as np
data = pd.read_csv('2019 Winter Data Science Intern Challenge Data Set - Sheet1.csv')
```

```
In [6]: data.shape
```

```
Out[6]: (5000, 7)
```

```
In [12]: data.columns
```

```
Out[12]: Index(['order_id', 'shop_id', 'user_id', 'order_amount', 'total_items',
               'payment_method', 'created_at'],
              dtype='object')
```

```
In [19]: summary_order_amt = data['order_amount'].describe()
summary_order_item = data['total_items'].describe()
```

```
In [53]: summary_order_amt['mean']
```

```
Out[53]: 3145.128
```

From the above mean value calculated, it indicated that the mean value of the total order amount for all the shops was calculated, while it AOV actually means average cost of a order placed in 30 day window.

Below we have calculated the correct aoc and followed by a proper metric for calculation.

```
In [27]: sum(data['order_amount'])/sum(data['total_items'])
```

```
Out[27]: 357.92152221412965
```

No of Unique customers in the dataset

We will consider a metric which will lead us to know how much an average customer spends at all our sneaker stores. Leading to know the likeliness of our product in general.

```
In [28]: len(data['user_id'].unique())
```

```
Out[28]: 301
```

Question 2 (new metric)

Average order value by each customer in the 30 day window

```
In [51]: sums = 0
for i in data['user_id'].unique():
    s = data.loc[data['user_id'] == i]
    sums+=sum(s['order_amount'])
avg_cust_spent = sums/len(data['user_id'].unique())
```

Question 3 (Value)

```
In [52]: avg_cust_spent
```

```
Out[52]: 52244.651162790695
```

Average order value by each customer in the 30 day window = \$52244.651162