In [1]:

```python
import pandas as pd
```

In [2]:

```python
#shortened csv filename for easier reading
#will need to change path relative to your pwd!
shopify_sneakshops = pd.read_csv('/Users/svalm763/Downloads/DataScienceInternChallenge.csv')
```

In [3]:

```python
shopify_sneakshops.head()
```

Out[3]:

	order_id	shop_id	user_id	order_amount	total_items	payment_method	created_at
0	1	53	746	224	2	cash	2017-03-13 12:36:56
1	2	92	925	90	1	cash	2017-03-03 17:38:52
2	3	44	861	144	1	cash	2017-03-14 4:23:56
3	4	18	935	156	1	credit_card	2017-03-26 12:43:37
4	5	18	883	156	1	credit_card	2017-03-01 4:35:11

Checking for missing values is an important step of data analysis - could allow for some explanation as to why the AOV might be wrong.

In [4]:

```python
shopify_sneakshops.isnull()
```

Out[4]:

	order_id	shop_id	user_id	order_amount	total_items	payment_method	created_at
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...
4995	False	False	False	False	False	False	False
4996	False	False	False	False	False	False	False
4997	False	False	False	False	False	False	False
4998	False	False	False	False	False	False	False
4999	False	False	False	False	False	False	False

5000 rows × 7 columns

1 a. Looking at the summary statistics for the 'order_amount' column, I can tell that the average order amount (AOV) of $3145.13 was taken from the mean value. The maximum value is relatively high compared to the median order amount, and event the first and third quartiles. There must be too many outliers within the data, which is why the AOV seems odd.

When I look at the median value, that would be a better metric to evaluate the AOV because the outliers will not affect the median as much as the mean value.

In [5]:

```python
shopify_sneakshops['order_amount'].describe()
```

Out[5]:

```
count      5000.000000
mean       3145.128000
std       41282.539349
min          90.000000
25%         163.000000
50%         284.000000
75%         390.000000
max      704000.000000
Name: order_amount, dtype: float64
```

1 b. The better metric to use would be the median value. We can calculate what the value will be (though it is shown in the summary stats).

In [6]:

```python
shopify_sneakshops['order_amount'].median()
```

Out[6]:

```
284.0
```

1 c. The median value is $284.00, which is the best metric to evaluate the average order amount (AOV).

In [ ]:

```python

```