

```
In [1]: import pandas as pd
import os
pwd = os.getcwd()
dataset = pd.read_excel(pwd + "\\2019 Winter Data Science Intern Challenge Data")
dataset.head()
df = dataset.copy()
```

```
In [9]: meanAOV = df['order_amount'].mean()
std = df['order_amount'].std()
print(f"meanAOV is ${meanAOV}")
print(f"Standard deviation is ${std:.2f}")
```

```
meanAOV is $3145.128
Standard deviation is $41282.54
```

The naive average order value (AOV) is calculated using the an average (mean metric) of all order amounts.

This calculation would be fine if there weren't such significant outliers in order amounts: the vast majority of order amounts are in the hundreds range (up to 999), but a few are in the tens of thousands or even hundred thousands (up to 99,999) We can notice the presence of the massive outliers by calculating the standard deviation.

```
In [10]: medianAOV = df['order_amount'].median()
print(f"medianAOV is ${medianAOV}")
```

```
medianAOV is $284.0
```

A more appropriate metric to deal with outliers is the median. It works by sorting all the amounts and choosing the value in the middle. Since it doesn't take into account how much bigger the values are from each other, having large outliers doesn't affect the result.

Since we're talking about sneakers here, \$284 is a much more believable AOV than 3145