Author: Rhea Mahesh

1. Given some sample data, write a program to answer the following:

> On Shopify, we have exactly 100 sneaker shops, and each of these shops sells only one model of shoe. We want to do some analysis of the average order value (AOV). When we look at orders data over a 30 day window, we naively calculate an AOV of $3145.13. Given that we know these shops are selling sneakers, a relatively affordable item, something seems wrong with our analysis.

a. The source of the error in our value for the AOV is the fact that we did not take the `total_items` column into account in our calculations.

We appear to have taken a mean over the `order_amount` column, which is naive to the presence of multiple orders in this particular dataset.

In [1]:
```
from numpy import mean
from numpy import std
from numpy import genfromtxt

# Parse CSV
filename = 'sneakers.csv'
csv = genfromtxt(filename, dtype=str, delimiter=",")

# Extract 'order_amount' column from CSV file into a list
amount_str = csv[:,3][1:]
amount_data = list(map(int, amount_str))

# Extract 'total_items' column from CSV file into a list
total_str = csv[:,4][1:]
total_data = list(map(int, total_str))

#Source of error: incorrect mean calculation
print("Naive mean calculation: mean(amount_data) = $%.2f\n" % mean(amount_d

# Calculate the mean for sneaker orders
order_mean = sum(amount_data) / sum(total_data)
print("Correct AOV calculation: sum(amount_data) / sum(total_data) = $%.2f"
```

```
Naive mean calculation: mean(amount_data) = $3145.13

Correct AOV calculation: sum(amount_data) / sum(total_data) = $357.92
```

2. For this question you'll need to use SQL. Follow this link (https://www.w3schools.com/SQL/TRYSQL.ASP?FILENAME=TRYSQL_SELECT_ALL) this link to access the data set required for the challenge. Please use queries to answer the following questions. Paste your queries along with your final numerical answers below.

a. How many orders were shipped by Speedy Express in total?

Query:

```
SELECT COUNT(*)
    FROM Orders o, Shippers s
    WHERE s.ShipperName = "Speedy Express"
        AND o.ShipperID = s.ShipperID;
```

Answer: **54**

b. What is the last name of the employee with the most orders?

Query:

```
SELECT e.LastName
    FROM Employees e JOIN Orders o ON
e.EmployeeID=o.EmployeeID
    GROUP BY e.EmployeeID
    ORDER BY COUNT(*) DESC
    LIMIT 1;
```

Answer: **40**

c. What product was ordered the most by customers in Germany?

Query:

```
SELECT p.ProductName FROM
    (
        (
            (Customers c JOIN Orders o ON c.CustomerID =
o.CustomerID)
            JOIN OrderDetails od ON o.OrderID = od.OrderID
        )
        JOIN Products p ON od.ProductID = p.ProductID
    )
GROUP BY Country, ProductName HAVING Country = 'Germany'
ORDER BY SUM(Quantity) DESC LIMIT 1;
```

> Answer: **Boston Crab Meat**