Armin Foroughi

This pdf answers the Shopify Fall 2021 Data Science Intern Challenge

Question 1:

For direct access to the Jupiter notebook you may use =https://github.com/arminforoughi/Shopify-intern-challenge.git

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib
```

This notebook is to analyze the shopify data to calculate average order value (AOV) for each store

HERE we can read the data into a dataframe

In [2]:
```
data = pd.read_csv('2019 Winter Data Science Intern Challenge Data Set - Sheet1.csv')
data.head()
```

Out[2]:

	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

calculate the current AOV

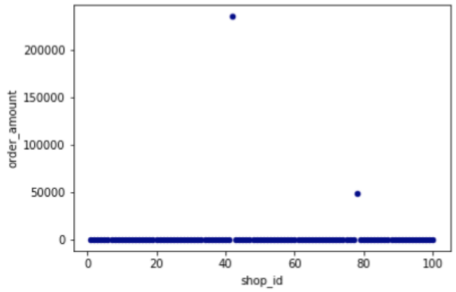
In [3]:
```
AOV = data.groupby('shop_id')['order_amount'].mean()
print('AOV = ' + str(AOV.mean()) + ' amount per order')
```

```
AOV = 3136.83408789 amount per order
```

Lets plost the AOV for each store

In [6]:
```
plt = pd.DataFrame(AOV).reset_index().plot.scatter(x='shop_id',y='order_amount',c='DarkBlue')
plt.plot()
```

Out[6]: []



There are two outlined values that push the mean way up

by removing those two values we can calculate the new AOV

In [5]:
```
true_AOV = (data[data['order_amount'] < 10000].groupby('shop_id')['order_amount'].mean()).mean()
print('true AOV = ' + str(true_AOV) + ' amount per order')
```

```
true AOV = 303.24353948 amount per order
```

AOV without the oulines is 303.24353948 amount per order

Question 2:

A)

```
SELECT Count(OrderID) FROM Orders
WHERE ShipperID = (
   SELECT ShipperID
   FROM Shippers
   where ShipperName = 'Speedy Express';
);
```

answer : 54

B)

```
SELECT LastName
FROM Employees
WHERE EmployeeID = (
   SELECT top 1 EmployeeID
   FROM (
      SELECT COUNT(EmployeeID), EmployeeID
      FROM Orders
```

```
    GROUP BY EmployeeID
    ORDER BY COUNT(EmployeeID) DESC
  );
);
```

answer : Peacock

c)

```
SELECT ProductName
FROM Products
WHERE ProductID = (
  SELECT top 1 ProductID
  FROM (
    SELECT COUNT(ProductID), ProductID FROM OrderDetails
    where OrderID in (
      SELECT OrderID
      FROM Orders
      where CustomerID in (
        SELECT CustomerID
        FROM Customers
        WHERE Country = 'Germany'
      )
    )
    GROUP BY ProductID
    ORDER BY COUNT(ProductID) DESC
  );
);
```

answer : Gorgonzola Telino