

Report

Notations:

pu: Purchase table

pr: Product table

cu: Customer table

fh: First half table

sh: Second half table

Following customers purchased products in the month of October 2016

First Name	Last Name	Product Name	Purchase Price
Mike	Horvath	Refrigerator	899
Irwin	Wade	VCR	150
Mike	Horvath	Refrigerator	1100
George	Marshall	Microwave	87

Query:

```
Select FirstName, LastName, ProductName, PurchasePrice
from pu
left join cu on cu.CustomerID==pu.CustomerID
left join pr on pu.ProductID==pr.ProductID
where PurchaseDate between '2016-10-01' and '2016-10-31';
```

These customers have not made purchases between June and December 2016

Customer ID	First Name	Last Name
1	John	Miller
2	Fred	Hammill
4	Adrian	Caparzo
8	Frank	Costello
17	Holly	Raines
18	Natalie	Woods
19	Wendy	Hilton
5	Mike	Horvath
9	Billy	Costigan
16	Tricia	Hill

Query:

```
Select DISTINCT cu.CustomerID, FirstName, LastName
from pu
```

```
left join cu on cu.CustomerID==pu.CustomerID
where PurchaseDate not BETWEEN '2016-06-01' and '2016-12-31';
```

Product Category and the average Purchase Price for those customers who have purchased three or more unique products

Category	Purchase Price
Office	173.666667

Query:

```
Select Category, avg(PurchasePrice) from pu
left join cu on cu.CustomerID==pu.CustomerID
left join pr on pu.ProductID==pr.ProductID
GROUP by cu.CustomerID
HAVING COUNT(DISTINCT pu.ProductID)>=3
```

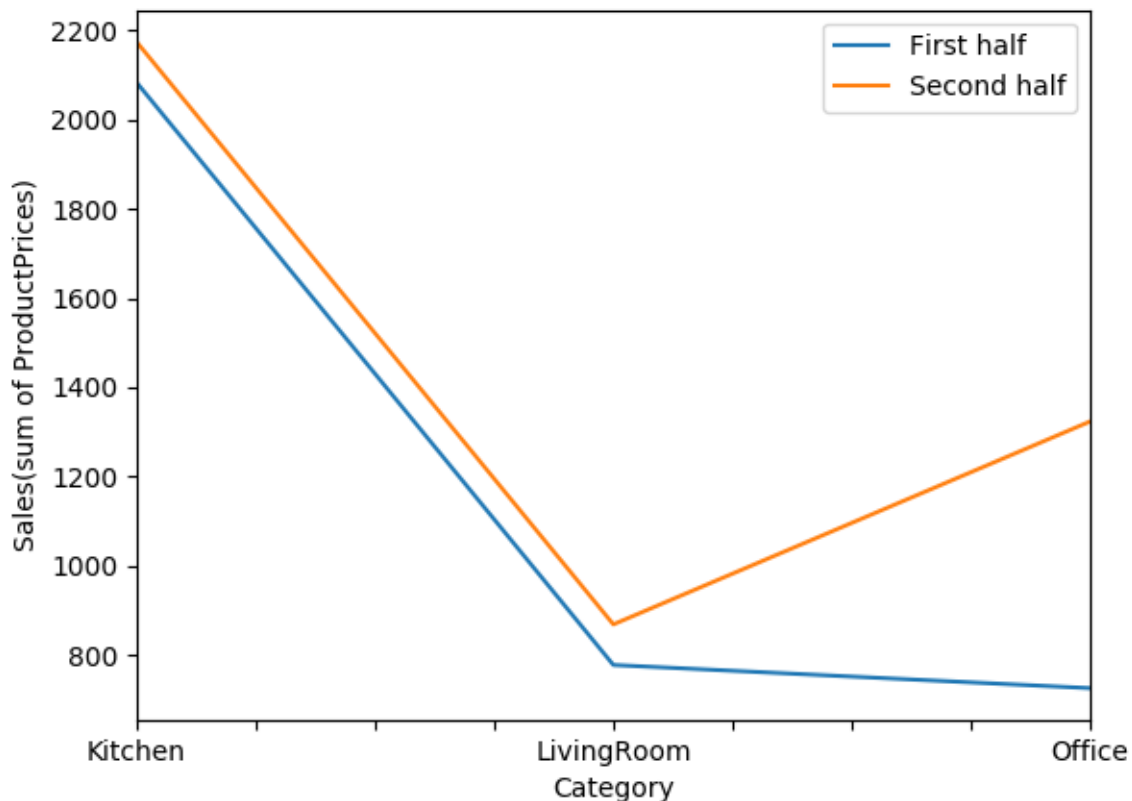
Product Name with the highest total Purchase Price for each Product Category

Product Name
Refrigerator
DVD
Monitor

Query:

```
Select ProductName FROM (
Select Category, ProductName, PurchasePrice
from pu
left join pr on pu.ProductID==pr.ProductID
GROUP by pr.Category, pr.ProductName)
GROUP BY Category
HAVING MAX(PurchasePrice)
```

Sales performance by product category between the first and second half of the year (i.e.: Jan-Jun vs. July-Dec)



From the graph it is clear that Kitchen and Living Room category are almost consistent where as Office items saw almost doubled sales in second half compared to first half.

Another thing to note is that Kitchen items are sold the most which is almost two times the other categories.

Python Code:

```
import matplotlib.pyplot as plt
from pandasql import *
import pandas as pd
from _datetime import date

cu = pd.read_csv('customer.csv')
pr = pd.read_csv('product.csv')
pu = pd.read_csv('purchase.csv')

pu.columns = ['PurchaseID', 'ProductID', 'CustomerID', 'PurchasePrice',
               'PurchaseDate']
pu['PurchaseDate'] = pd.to_datetime(pu['PurchaseDate'])
```

```
pysqldf = lambda q: sqldf(q, globals())
```

```
q = """
Select FirstName, LastName, ProductName, PurchasePrice from pu
left join cu on cu.CustomerID==pu.CustomerID left join pr on
pu.ProductID==pr.ProductID
where PurchaseDate between '2016-10-01' and '2016-10-31';
"""
```

```
q1 = """
Select DISTINCT cu.CustomerID, FirstName, LastName from pu
left join cu on cu.CustomerID==pu.CustomerID
where PurchaseDate not BETWEEN '2016-06-01' and '2016-12-31';
"""
```

```
q2 = """
Select Category, avg(PurchasePrice) from pu
left join cu on cu.CustomerID==pu.CustomerID left join pr on
pu.ProductID==pr.ProductID
GROUP by cu.CustomerID
HAVING COUNT(DISTINCT pu.ProductID)>=3
"""
```

```
q3 = """
Select ProductName FROM (
Select Category, ProductName, PurchasePrice from pu left join pr
on pu.ProductID==pr.ProductID
GROUP by pr.Category, pr.ProductName)
GROUP BY Category
HAVING MAX(PurchasePrice)
"""
```

```
fh = pu[(pu['PurchaseDate']>= date(2016,1,1)) &
(pu['PurchaseDate']<=date(2016,6,30))]
sh = pu[(pu['PurchaseDate']>= date(2016,7,1)) &
(pu['PurchaseDate']<=date(2016,12,31))]
df1 = pysqldf(q)
df2 = pysqldf(q1)
df3 = pysqldf(q2)
df4 = pysqldf(q3)
print(df1, df2, df3, df4)
df5 = pd.merge(fh,pr, on='ProductID').groupby('Category')
['PurchasePrice'].sum()
df6 = pd.merge(sh,pr, on='ProductID').groupby('Category')
['PurchasePrice'].sum()
df2.plot(kind='line', label='First half')
df3.plot(kind='line', label='Second half')
plt.legend()
plt.ylabel('Sales(sum of ProductPrices)')
plt.show()
```