/\*import the datsets\*/

**proc** **import** out=customer datafile="D:\DATA ANALYSIS\sas documents\udemy02\datasets\datasets\project\customers.xlsx"

dbms=xlsx replace;

getnames=yes;

**run**;

**proc** **import** out=product datafile="D:\DATA ANALYSIS\sas documents\udemy02\datasets\datasets\project\products.xlsx"

dbms=xlsx replace;

getnames=yes;

**run**;

**proc** **import** out=sale datafile="D:\DATA ANALYSIS\sas documents\udemy02\datasets\datasets\project\sales.xlsx"

dbms=xlsx replace;

getnames=yes;

**run**;

**proc** **import** out=return datafile="D:\DATA ANALYSIS\sas documents\udemy02\datasets\datasets\project\returns.xlsx"

dbms=xlsx replace;

getnames=yes;

**run**;

/\*merge them into on talbe\*/

**proc** **sql**;

create table mg as

select s.\*,c.\*,p.\*

from sale as s

left join customer as c

on s.customer\_id=c.id

left join product as p

on s.product\_id=p.id;

**quit**;

**proc** **print** data=mg(obs=**10**);**run**;

/\*how much sale for each customer\*/

**proc** **sql**;

create table percus as /\*percus=per customer\*/

select customer\_name,sum(profit)as profit

from mg

group by customer\_name

order by profit descending;

**quit**;

/\*region distribution\*/

**proc** **sql**;

create table prosale as

select province,sum(sales) as sale

from mg

group by province;

**quit**;

/\*region proportion\*/

**proc** **sql**;

create table propct as

select province,sale,sale/sum(sale) as pct

from prosale;

**quit**;

**proc** **print** data=propct(obs=**10**);**run**;

\*pie plot;

**proc** **gchart** data=mg;

pie province/discrete sumvar=sales;/\*discrete:create to show all the provinces, sumvar(sum up the values in this variable)=sales\*/

**run**;

/\*concentration\*/

/\*https://baike.baidu.com/item/%E8%B5%AB%E8%8A%AC%E8%BE%BE%E5%B0%94%E2%80%94%E8%B5%AB%E5%B8%8C%E6%9B%BC%E6%8C%87%E6%95%B0/1429385#reference-[1]-2192635-wrap\*/

/\*the first step where I need to find out the proportion of the each group to the sum\*/

**proc** **sql**;

create table herfin as

select customer\_name,sum(profit) as profit

from mg

group by customer\_name;

**quit**;

/\*next step will be calculating each customer’s contribution to the group. So every customer’s

profit divided by the sum of the profit(the total profit),\*/

**proc** **sql**;

create table herfin\_pct as

select\*,profit/sum(profit) as pct

from herfin;

**quit**;

/\*finaly,we can sum up the square of percentage to get the in the index.\*/

/\*at the same time we calculate the benchmark,which represents a case where

every customer contribute\*/

**proc** **sql**;

select sum(pct\*pct) as herfindal,**1**/count(\*) as benchmark

from herfin\_pct;

**quit**;

/\*as we can tell,our concentration level is higher than the benchmark where

every customer spend the same amount of money\*/

/\*in conclusion ,we found that our profits are not concertrated on particular

customers and it is very solid against customer changes\*/

/\*highest return rate product\*/

**proc** **sql**;

create table mgreturn as

select r.\*,m.product\_name

from return as r

left join mg as m

on r.order\_id=m.order\_id;

**quit**;

**proc** **print** data=mgreturn(obs=**10**);**run**;

/\*group up by product\*/

**proc** **sql**;

create table groupreturn as

select product\_name,count(\*) as returns

from mgreturn

group by product\_name

order by returns descending;

**quit**;

**proc** **print** data=groupreturn(obs=**10**);**run**;

\*we should have our sales representative to talk to these suppliers and find out if there's anything wrong with the product quality taste;