

```
/* Accessing Data */
```

```
%let path = /home/u41140628/EPG194/ECRB94/data;
```

```
libname cr "&path/output";
```

```
options validvarname=v7;
```

```
libname ctryxl xlsx "&path/country_lookup.xlsx";
```

```
proc import datafile="&path/orders.csv" out=cr.orders dbms=csv replace;
```

```
run;
```

```
proc contents data=cr.orders;
```

```
run;
```

```
proc contents data=ctryxl._all_ nods;
```

```
run;
```

```
/* Exploring data */
```

```
/* Validate country lookup excel table */
```

```
proc print data=ctryxl.countries(obs=30);
```

```
run;
```

```
proc freq data=ctryxl.countries order= freq;
```

```
    tables country_key country_name;
```

```
run;
```

```
/* We have duplicate ows */
```

```
proc print data=ctryxl.countries;
```

```

        WHERE country_key in ('AG','CF','GB','US');

run;

/* removing the duplicates */

proc sort data=ctryxl.countries out=country_clean nodupkey dupout=dups;
    by country_key;
run;

/* Validate imported orders table */

/* Data quality rules: delivery date after order date, order date valie dates 1,2,3 */

/* customer_country should always be 2 upper case letters, customer continent should be 1 of 5
continents */

proc print data=cr.orders;
    WHERE order_date> delivery_date;
    var order_id order_date delivery_date;
run;

proc freq data=cr.orders;
    tables order_type customer_country customer_continent;
run;

/* CHeck min max values , also check in extreme observations in univariate*/

proc means data=cr.orders;
    var quantity retail_price cost_price;
run;

```

```
proc univariate data=cr.orders;
    var quantity retail_price cost_price;
run;
```

```
/* Preparing the data */
```

```
data profit;
    set cr.orders;
    length order_source $ 8;
    where delivery_date >= order_date;
    customer_country= upcase(customer_country);
    if quantity<0 then quantity=.;
    profit=(retail_price-cost_price)*quantity;
    format profit dollar12.2;
    shipdays= delivery_Date-order_date;
    age_range=substr(customer_age_group,1,5);
    if order_type=1 then order_source="Retail";
    else if order_type=2 then order_source="Phone";
    else if order_type=3 then order_source="Internet";
    else order_source="Unknown";
    drop retail_price cost_price customer_age_group order_type;
run;
```

```
/* Using proc sql to join data */
```

```
/* getting full names of country code table from corresponding values in country clean */
```

```
proc sql;
```

```
create table profit_country as  
    select profit.*,country_name  
    from profit inner join work.country_clean  
    on profit.customer_country=country_clean.customer_key  
    order by profit.order_Date desc;  
quit;
```

```
/* Orders frequency analysis */
```

```
ods noproctitle;  
title "Number of order by month";
```

```
proc freq data=profit order =freq;  
    tables order_date / nocum;  
    format order_date monname.;  
    tables customer_continent*Order_source / norow nocol;  
run;
```

```
%let os=Phone;  
title "&os orders";  
proc means data=profit min max mean maxdec=0;  
    var shipdays;  
    class customer_country;  
    where shipdays>0 and order_source="&os";  
run;
```

```
/* Profit analysis by customer age */
```

```
proc means data=profit noprint;
```

```
var profit;  
class age_range ;  
output out= profit_summary median=medprofit sum=totalprofit;  
ways 1;  
run;
```

```
proc print data=profit_summary noobs;  
var age_range totalprofit medprofit;  
label age_range="Age Range"  
totalprofit="Total Profit"  
medprofit="Median Profit Per Order";  
format totalprofit medprofit dollar10.;  
run;
```

```
/* Export reports to shareable data */
```

```
proc export data=profit outfile="&path/output/orders_update.csv" dbms=csv replace;  
run;
```

```
proc export data=profit outfile="&path/output/orders_update.xlsx" dbms=xlsx replace;  
run;
```

```
/* Using Output Delivery System */
```

```
ods pdf file="&path/output/orders_update.pdf" pdftoc=1;
```

```
title "Orders with order date after delivery date";  
proc print data=cr.orders;
```

```

WHERE order_date> delivery_date;

var order_id order_date delivery_date;

run;

title "Examine values of numeric columns in orders";

proc freq data=cr.orders;

    tables order_type customer_country customer_continent;

run;

title "Examine values of categorical columns in orders";

proc means data=cr.orders;

    var quantity retail_price cost_price;

run;

ods pdf close;

/* Using PUTLOG statements */

data new;

    putlog "Note: Value of HeightCM at the top of the data step";

    putlog HeightCM=;

    retain HeightCM 0;

    set sashelp.class(obs=3);

    HeightCM=Height*2.54;

    putlog "Note: Value of HeightCM at the bottom of the data step";

    putlog HeightCM=;

run;

```

```
proc sort data=profit out =decdaily;  
    where month(order_date)=12;  
    by order_date;  
run;
```

```
data decsales;  
    set decdaily;  
    retain MTDsales=0;  
    MTDsales=sum(MTDsales,profit);  
    keep order_id order_date profit MTDsales;  
run;
```

```
/* Using functions*/
```

```
data qtr_details;  
    set cr.qtr_sales;  
    totalpurchase=sum(of qtr:);  
    avgpurchase=round(mean(qtr:),0.01);  
    customerage=int(yrdif(birthdate,today(),"age"));  
    promo_date=mdy(month(birthdate),1,year(today()))  
    firstname=scan(name,1," ");  
    ID=put(customer_id,z5.)  
    format totalpurchase avgpurchase dollar12.2 promo_date mmddyy10.;  
    drop qtr: customer_id  
run;
```

```
/* order of column is in the sequence data is added in pdfv */
```

```
/* we can specify retain statement manually sequencing data */
```

```
data _6months;

    set cr.profit;

    where order_date >= intnx("month",today(),-6, "same");

    keep order_id order_date delivery_date;

    busdays=intck("weekday",order_Date,delivery_date)

run;
```

```
/* custom formats */
```

```
proc format;

    value shiprange 0="Same day"

                        1-3="1-3 days"

                        4-7="4-7 days"

                        8-high="8+ days"

                        .="unknown";

run;
```

```
/* USe this in datastep */
```

```
data profit;

    set cr.profit;

/*    format shipdays shiprange.; */

    shipRange=put (shipdays,shiprange.);

run;
```

```
proc freq data=cr.profit;

    table shipdays;

    format shipdays shiprange.;

run;
```