**Name:** Laavanya Ganesh

**UIN:** 654324917

**IDS 462: Statistical Software for Business**

**SAS and SQL Assignment**

1. Build a table showing the Province name, Province population and Country Name that the Province is located in.
   1. Note that since this answer requires the country name, then you will need to match the Province table to the Country table. This will be a little difficult, but it is good to understand how to do the matching in SAS and SQL because you will need to do this for other parts of the assignment.

**Program:**

/\* Q.1 SQL\*/

libname mylib "/folders/myfolders/";

proc sql dquote=ansi;

create table ProvinceInformation as

(select province.ProvinceName, province.Population, country.CountryName

From mylib.province as province, mylib.country as country

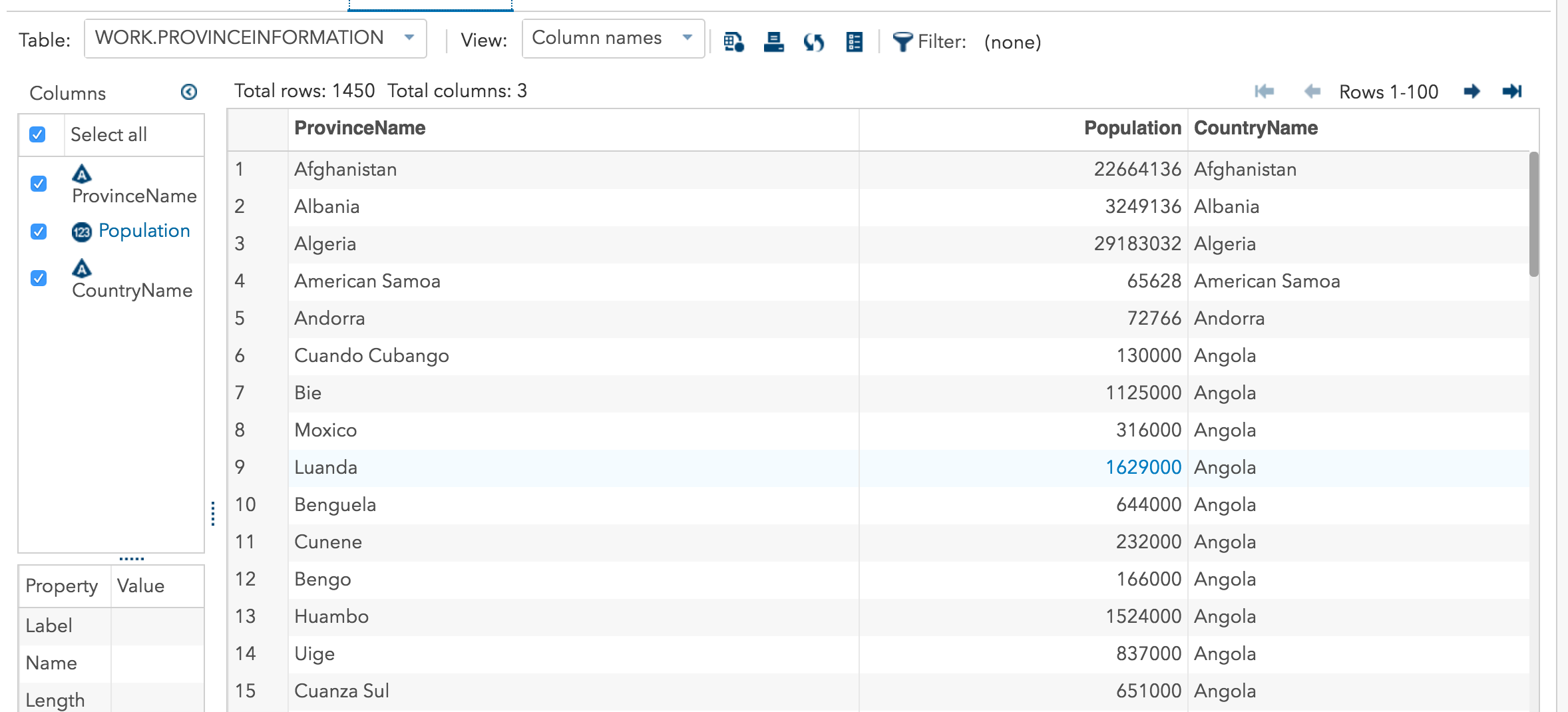
where province.Country=country.code)

order by country.CountryName;

quit;

Proc print data=ProvinceInformation (obs=20 keep= ProvinceName Population CountryName);

Run;

**Output:**

****

**Program:**

/\* Q.1 SAS \*/

proc sort data=mylib.province;

by Country;

run;

proc sort data=mylib.country;

by Code;

run;

data ProvinceandCountry;

merge mylib.province(in=OnProvince keep= Population ProvinceName Country rename=(Country=CC))

mylib.country(in=OnCountry keep=CountryName Code rename=(Code=CC))

;

by CC;

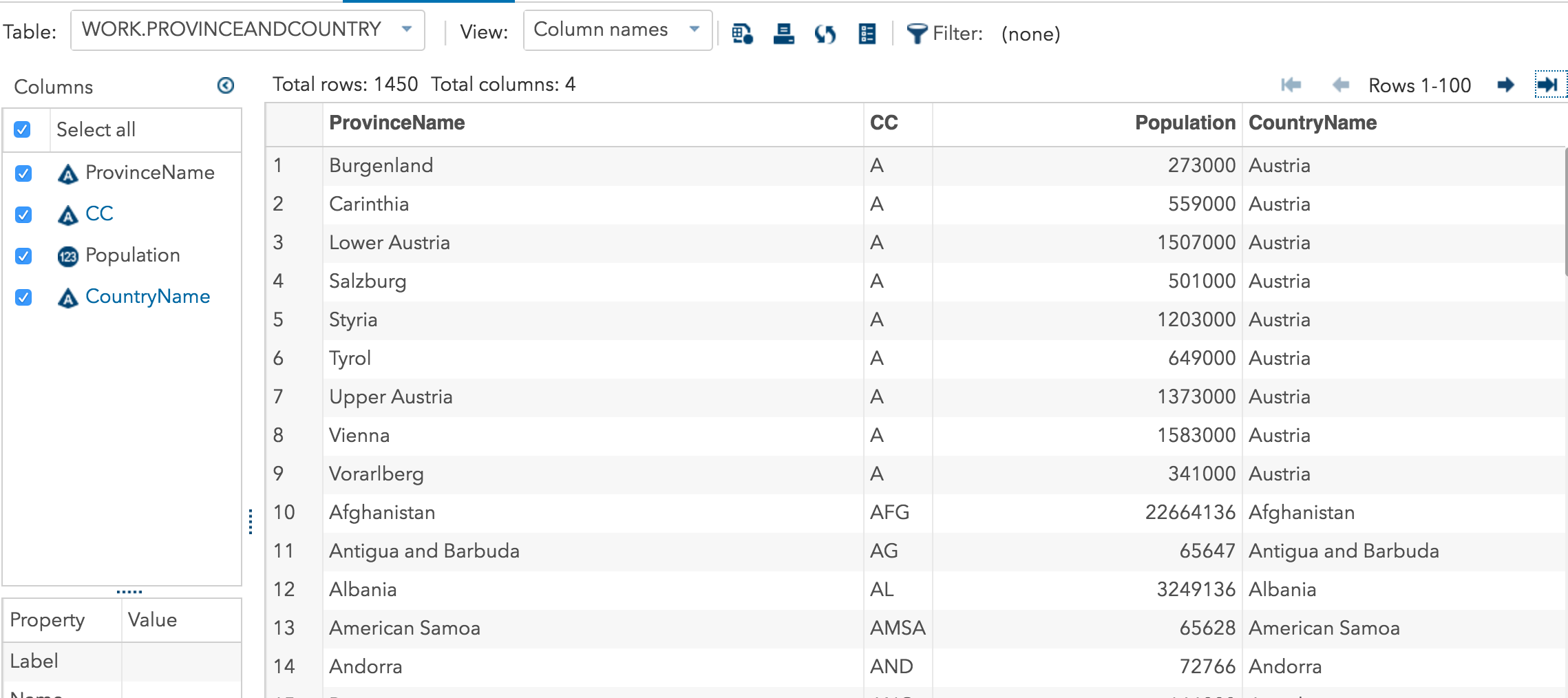
if OnProvince and OnCountry;

run;

Proc print data=ProvinceandCountry (obs=20 keep= ProvinceName Population CountryName);

Run;

**Output:**

****

****

1. Create a table showing the maximum and minimum population size of the Provinces in each country. Fields to include: Country Name, Country Code, Max population in a Province, minimum population in a Province

**Program:**

/\*Q.2 SQL \*/

proc sql dquote=ansi;

create table SqlMaxMinPopulationSize as

(select distinct country.Code, country.CountryName, max(province.population) as MaxPopulationSize, min(province.population) as MinPopulationSize

From mylib.province as province

join mylib.country as country

on province.Country=country.Code

group by country.CountryName)

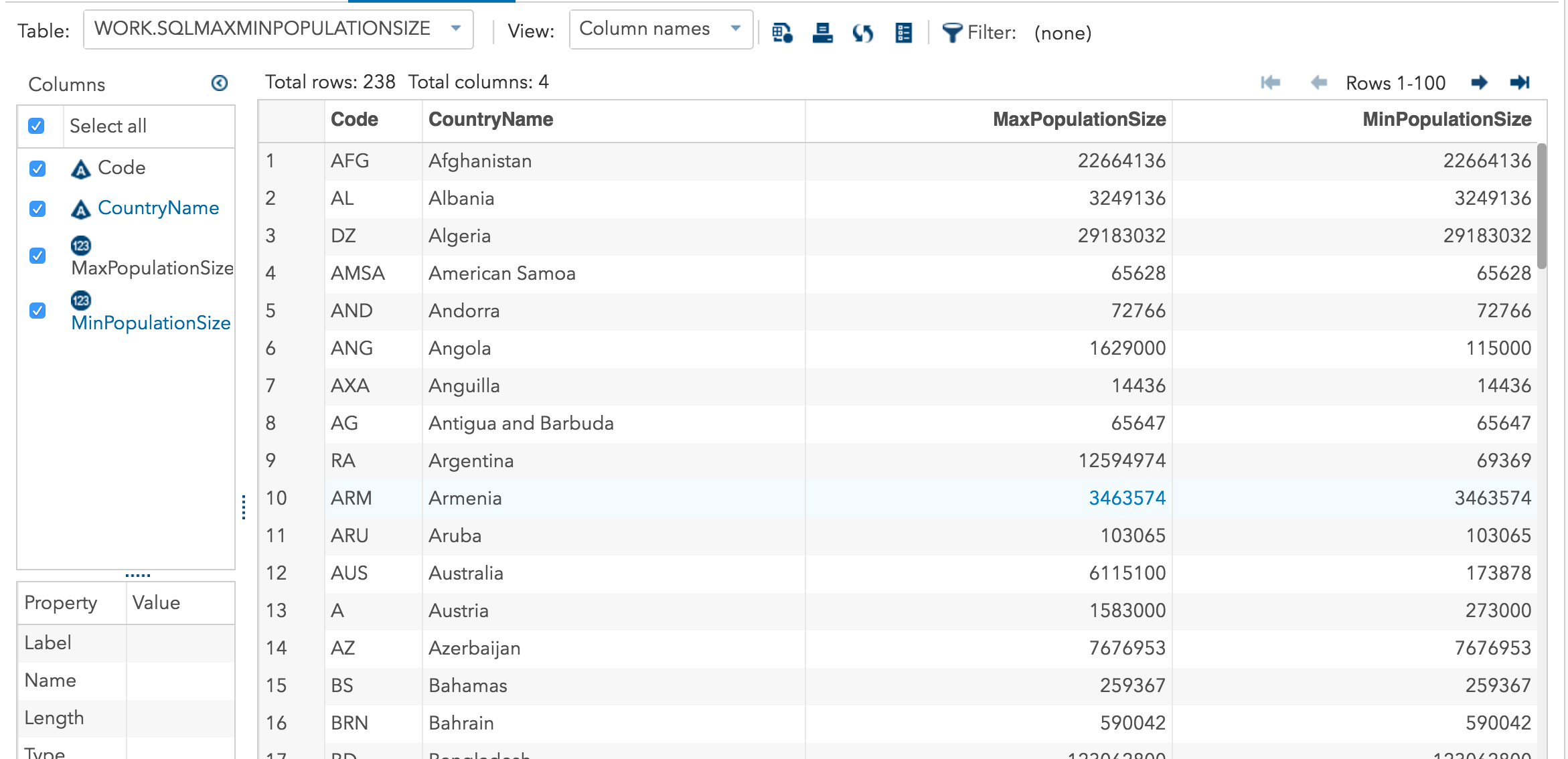
order by country.CountryName;

quit;

Proc print data=SqlMaxMinPopulationSize (obs=20 keep= Code CountryName MaxPopulationSize MinPopulationSize);

Run;

**Output:**





**Program:**

/\*Q.2 SAS \*/

proc sort data=mylib.province;

by Country;

run;

proc sort data=mylib.country;

by Code;

run;

data ProvinceandCountry1;

merge mylib.province(in=OnProvince keep= Population Country rename=(Country=CC))

mylib.country(in=OnCountry keep=CountryName Code rename=(Code=CC))

;

by CC;

if OnProvince and OnCountry;

run;

proc summary data=ProvinceandCountry1 nway;

class CC;

var Population;

output out=MaxPopulationSize max=;

run;

proc summary data=ProvinceandCountry1 nway;

class CC;

var Population;

output out=MinPopulationSize min=;

run;

Proc print data=ProvinceandCountry1 (obs=20 keep= CC CountryName Population);

Run;

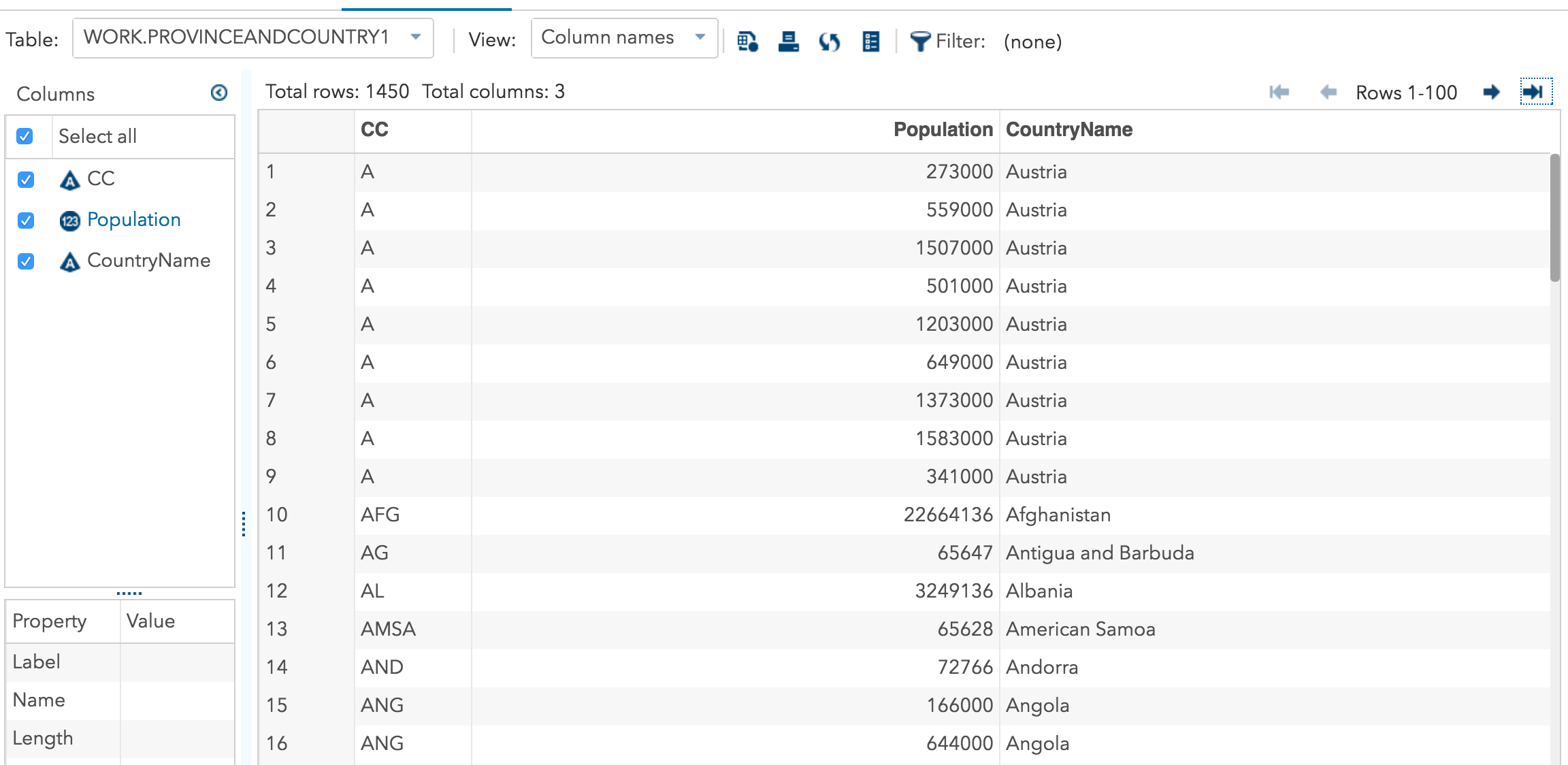
Proc print data=MaxPopulationSize (obs=20 keep= CC Population);

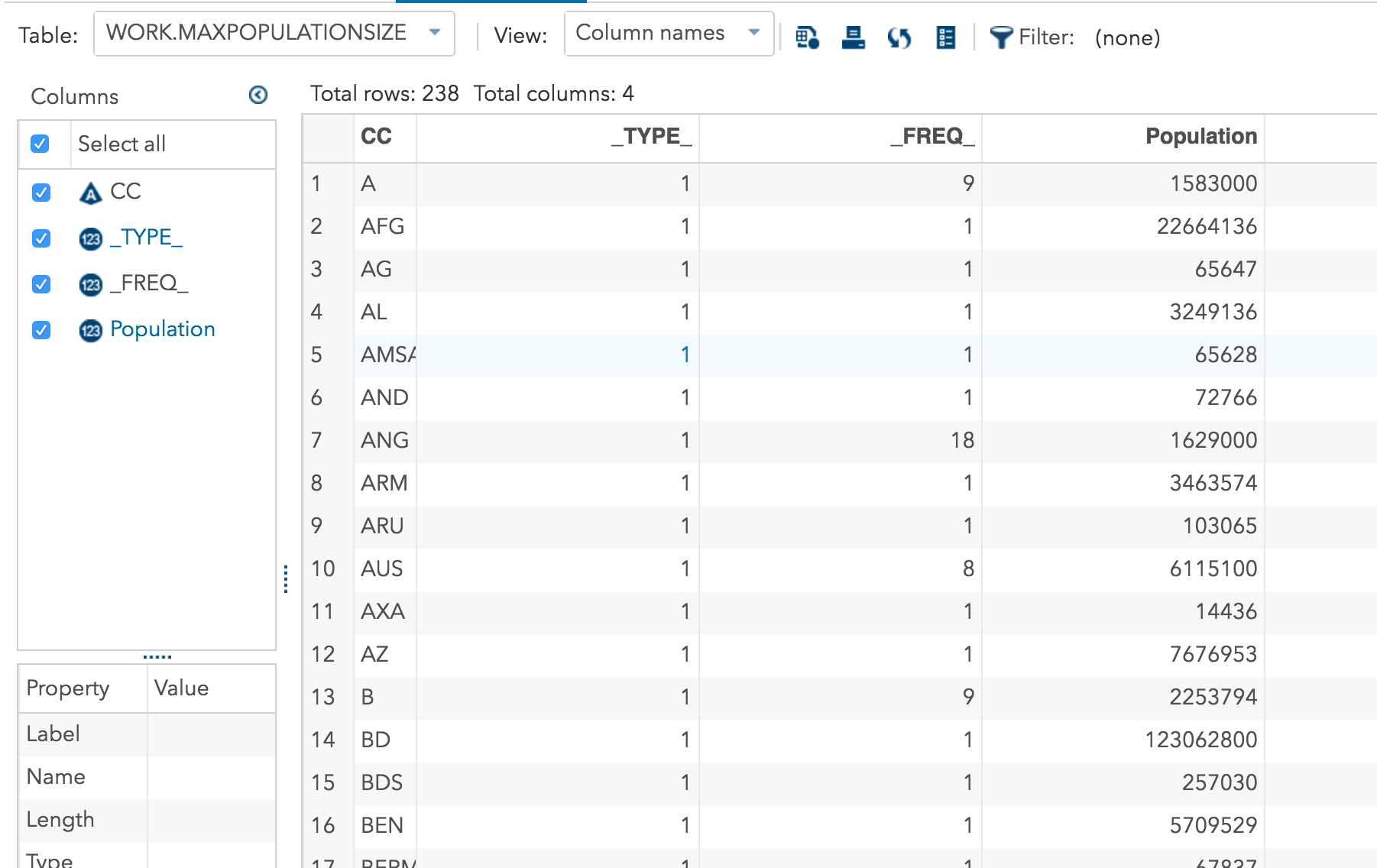
Run;

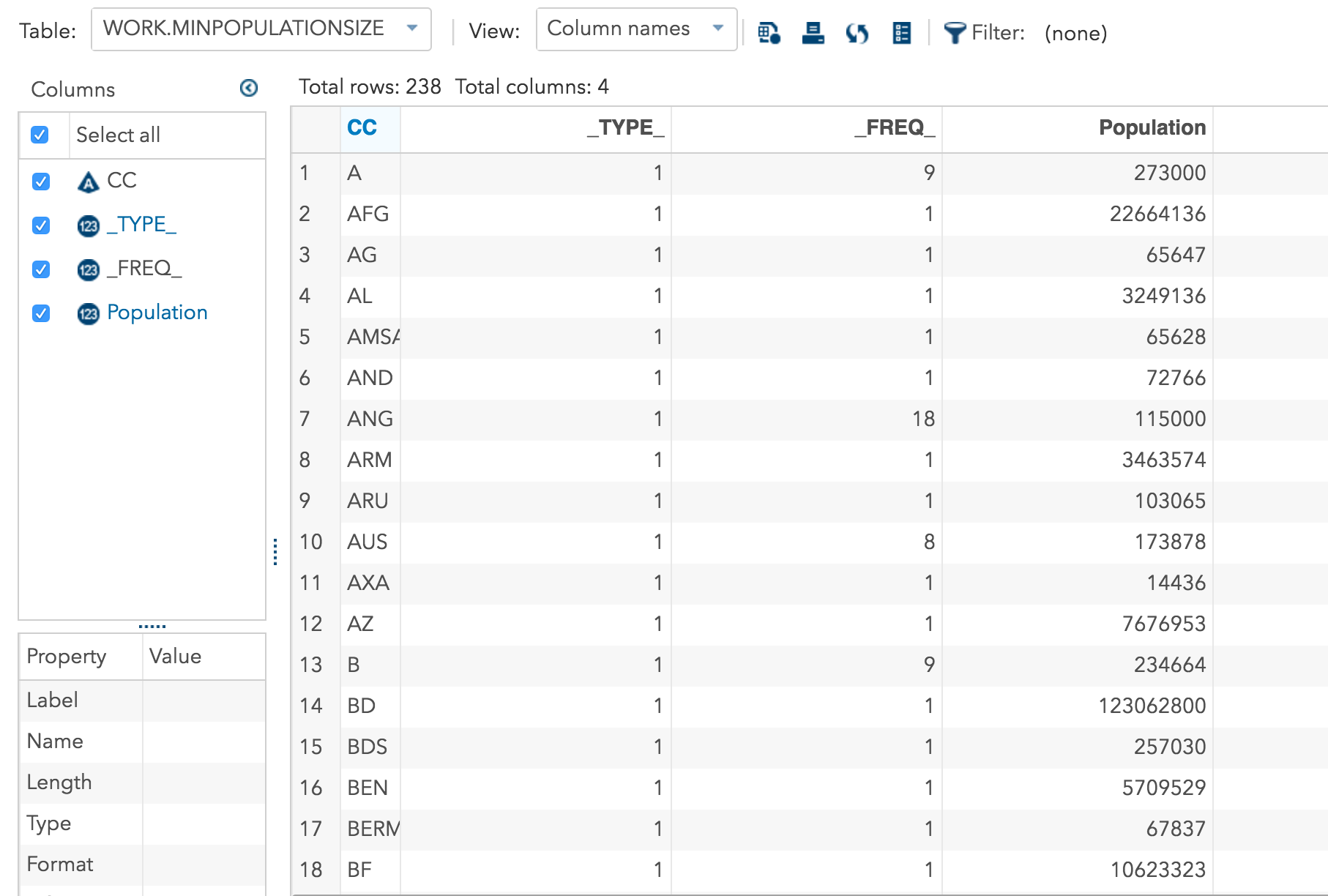
Proc print data=MinPopulationSize (obs=20 keep= CC Population);

Run;

**Output:**

****

****







MaxPopulationSize MinPopulationSize

1. What is the number of cities in each country? Fields: Country Name, Number of cities.

**Program:**

/\* Q.3 SQL \*/

proc sql dquote=ansi;

create table NumberOfCities as

(Select distinct count(city.CityName) as NumberOfCities, country.CountryName

From mylib.city as city, mylib.country as country

where city.country=country.code

group by city.country

)

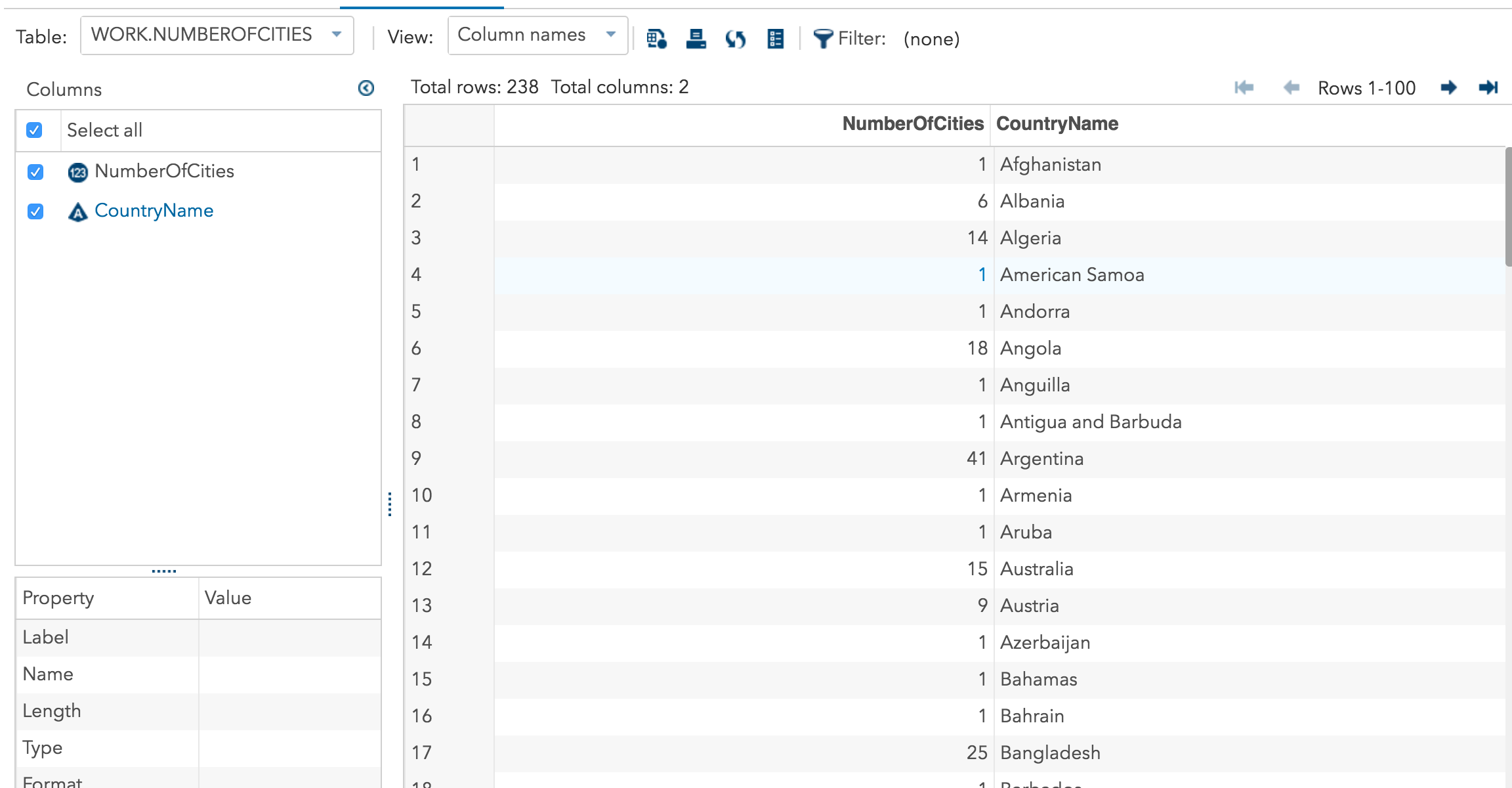
order by country.CountryName;

quit;

Proc print data=NumberOfCities (obs=20 keep= CountryName NumberOfCities);

Run;

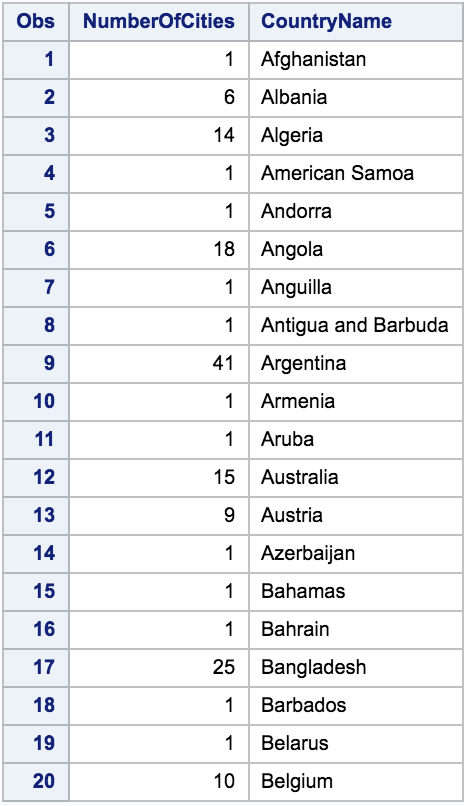
**Output:**

****

**Program:**

/\*Q.3 SAS \*/

proc sort data=mylib.city;

****by Country;

run;

proc sort data=mylib.country;

by Code;

run;

Data CityandCountry1;

Merge mylib.city (in=OnCity keep= CityName Country rename=(Country=CC))

mylib.country (in=OnCountry keep= CountryName Code rename=(Code=CC))

;

By CC;

If OnCity and OnCountry;

Run;

Proc print data=CityandCountry1 (obs=20 keep= CityName CountryName CC);

Run;

Data CityandCountry2;

set CityandCountry1;

NumericCityName=CityName +0;

Run;

proc summary data=CityandCountry2 nway;

class CC;

var NumericCityName;

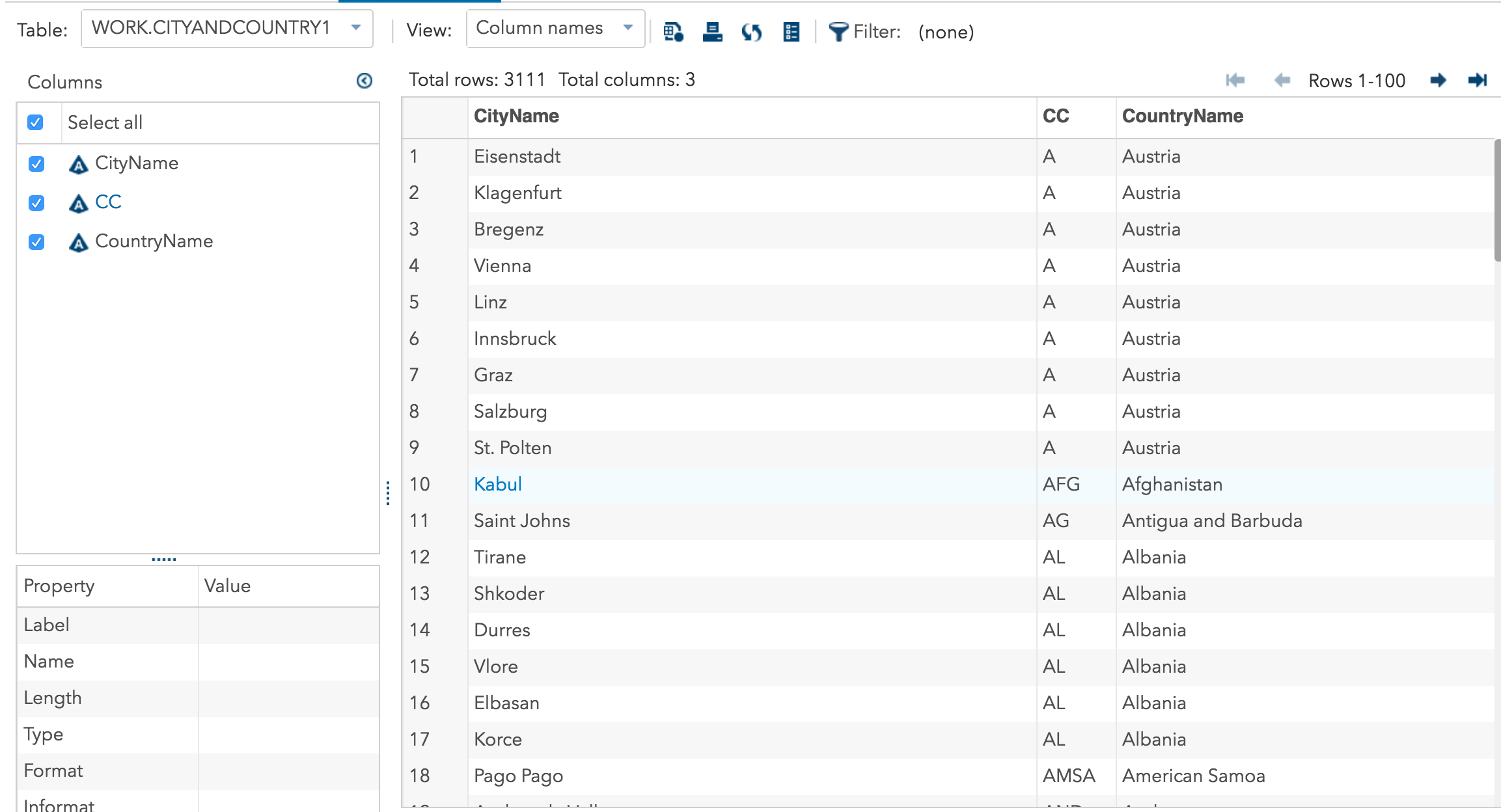
output out=NoOfCities N=;

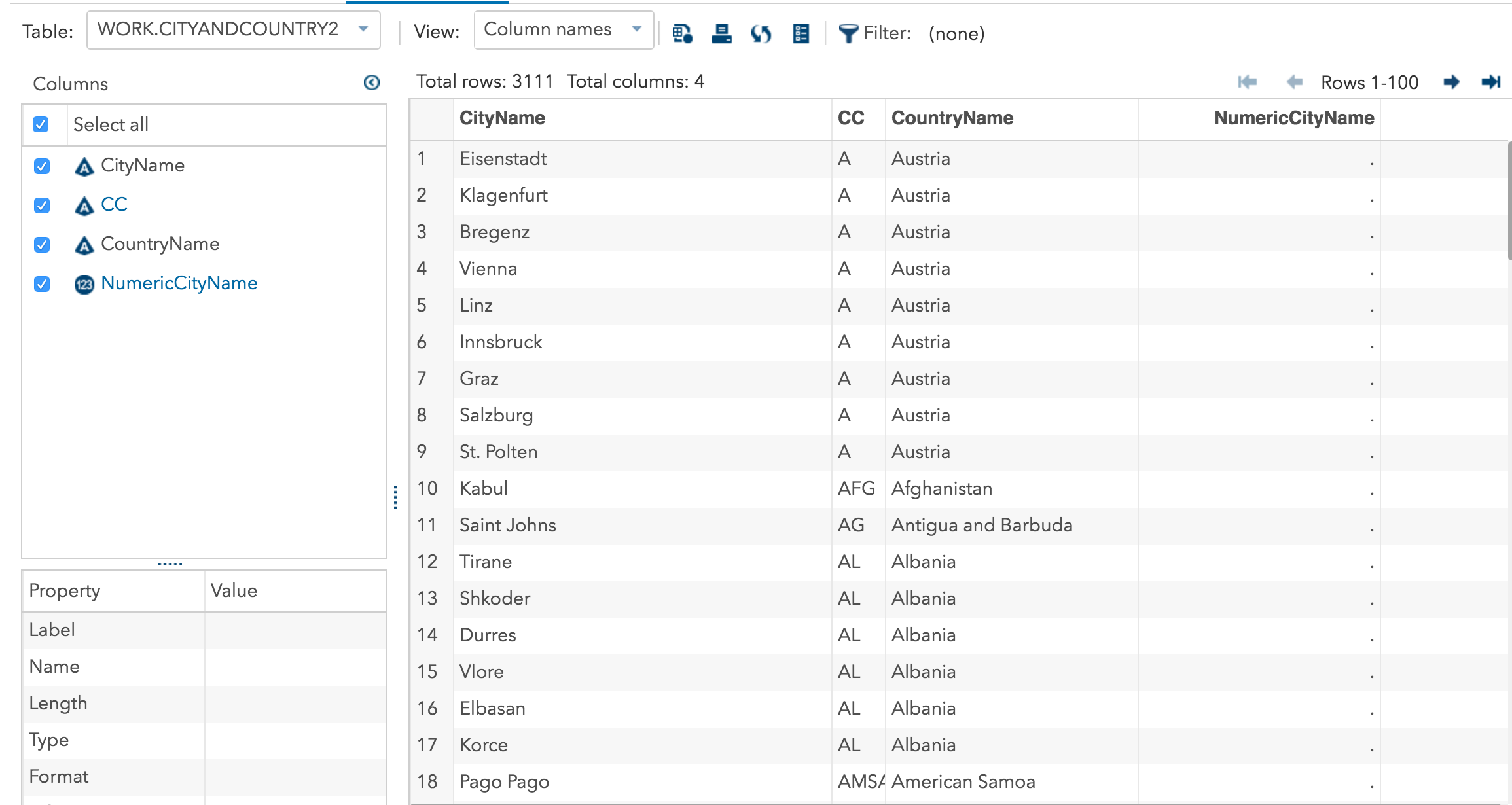
run;

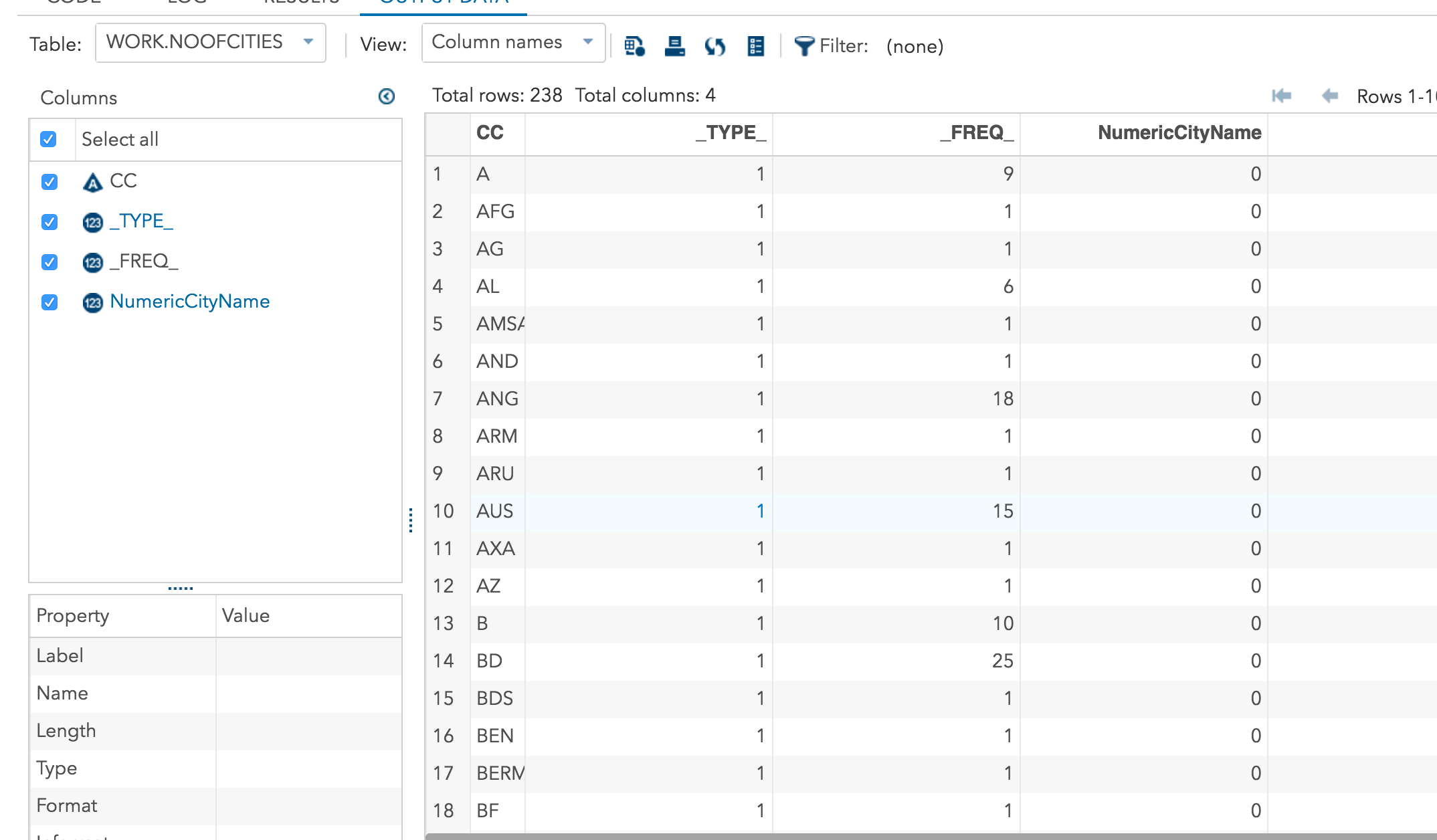
Proc print data=NoOfCities (obs=20 keep= CC \_FREQ\_);

Run;

**Output:**

****

****

****



1. SAS Only, not SQL. Identify the ID codes of the cities which do not match to a country. Use the country variable in both files as the match key. Fields: Id, city name, country code.

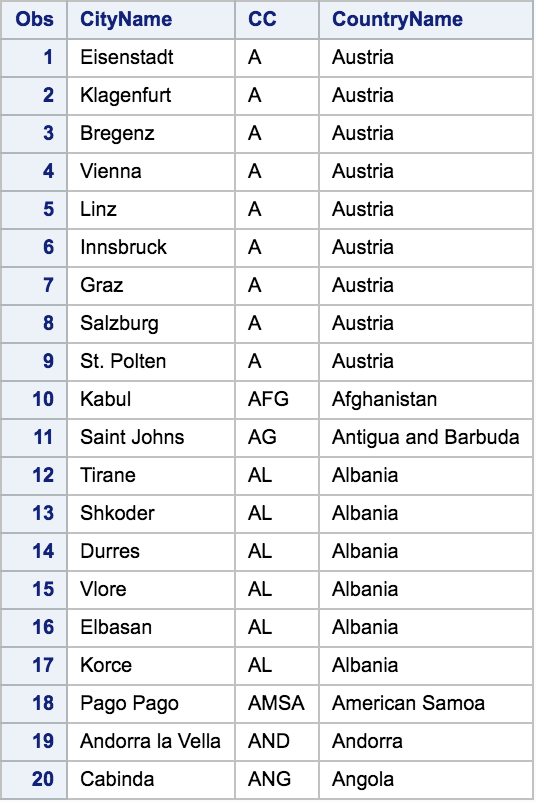
**Program:**

/\* Q.5 SAS \*/

proc sort data=mylib.city;

by Country;

run;

proc sort data=mylib.country;

by Code;

run;

Data CityandCountry;

Merge mylib.city (in=OnCity keep= ID CityName Country rename=(Country=CC))

mylib.country (in=OnCountry keep= Code rename=(Code=CC))

;

By CC;

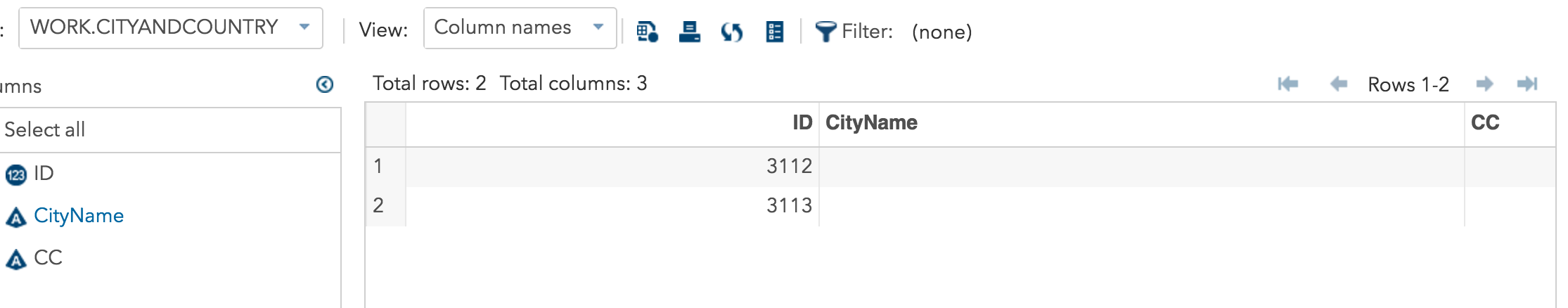
If OnCity and not OnCountry;

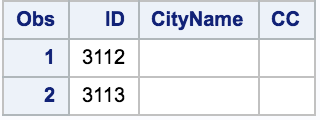
Run;

Proc print data=CityandCountry (obs=20 keep= ID CityName CC);

Run;

**Output:**

****



1. What is the average area for the Provinces in each country? Fields: Country name, country code, Average area.

**Program:**

/\* Q.6 SQL \*/

proc sql dquote=ansi;

create table AverageArea as

(select distinct country.CountryName, country.Code, avg(input(province.Area,8.)) as AvgArea

from mylib.province as province, mylib.country as country

where province.Country=country.Code

group by country.CountryName

)

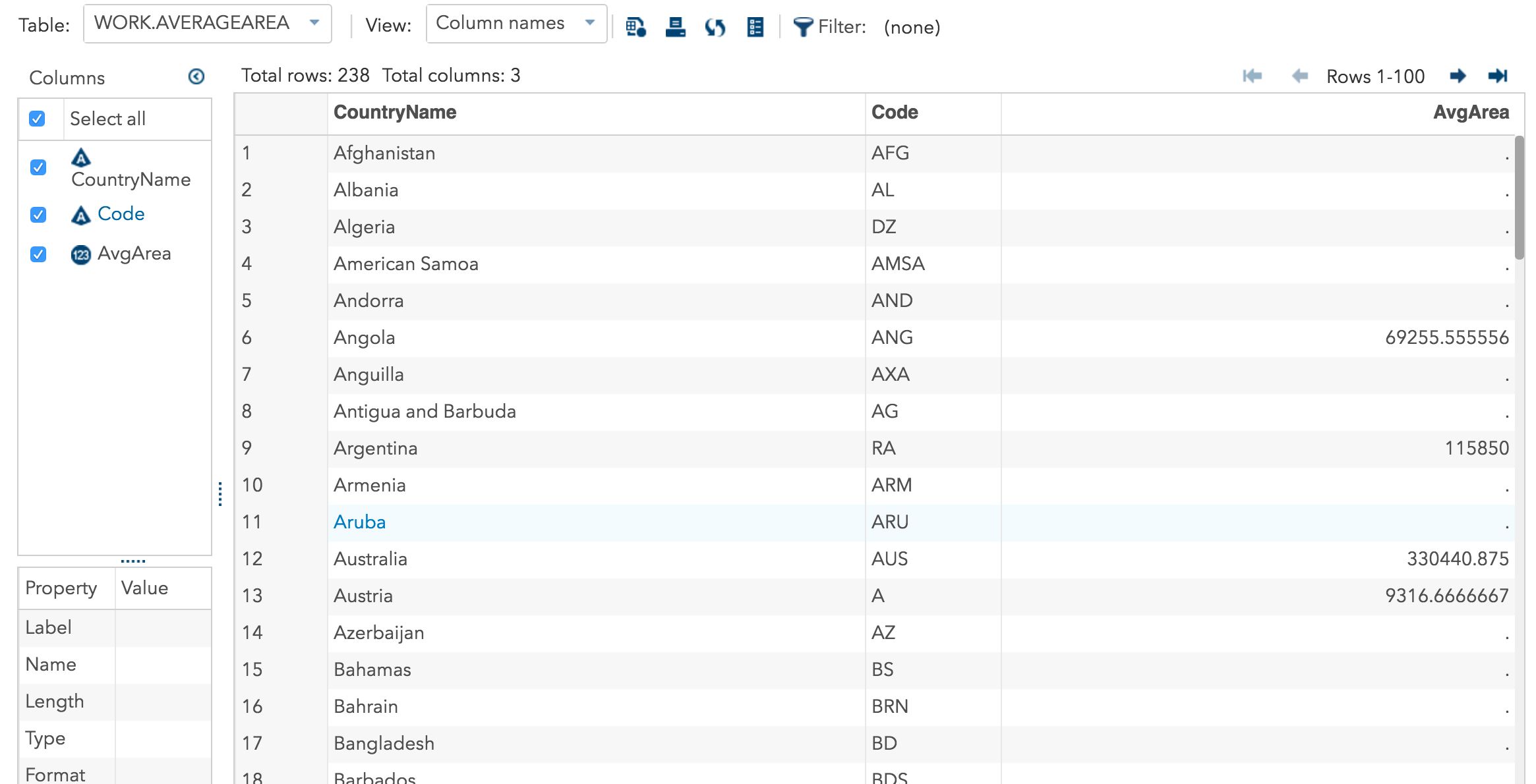
order by country.CountryName;

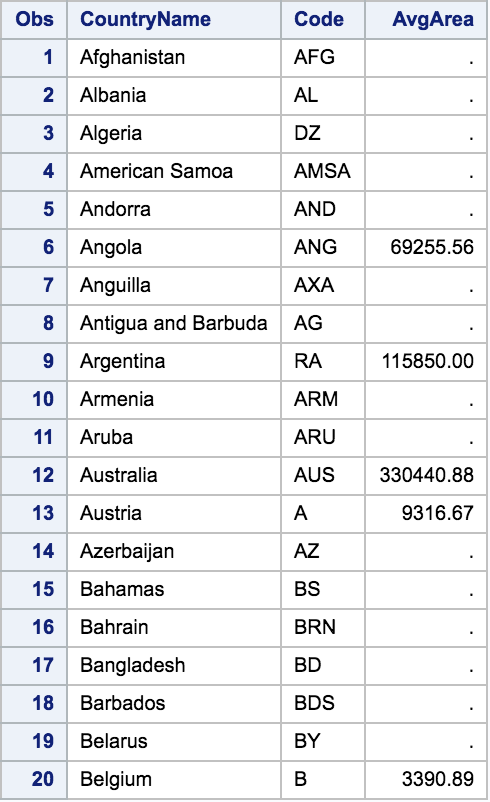
quit;

Proc print data=AverageArea (obs=20 keep= CountryName Code AvgArea);

Run;

**Output:**

****

****

**Program:**

/\* Q.6 SAS \*/

proc sort data=mylib.province;

by Country;

run;

proc sort data=mylib.country;

by Code;

run;

Data ProvinceandCountry3;

Merge mylib.province (in=OnProvince keep= Area Country rename=(Country=CC))

mylib.country (in=OnCountry keep= CountryName Code rename=(Code=CC))

;

By CC;

If OnProvince and OnCountry;

Run;

Proc print data=ProvinceandCountry3 (obs=20 keep= CountryName CC Area);

Run;

Data ProvinceandCountry4;

set ProvinceandCountry3;

NumericArea=Area +0;

Run;

proc summary data=ProvinceandCountry4 nway;

class CC;

var NumericArea;

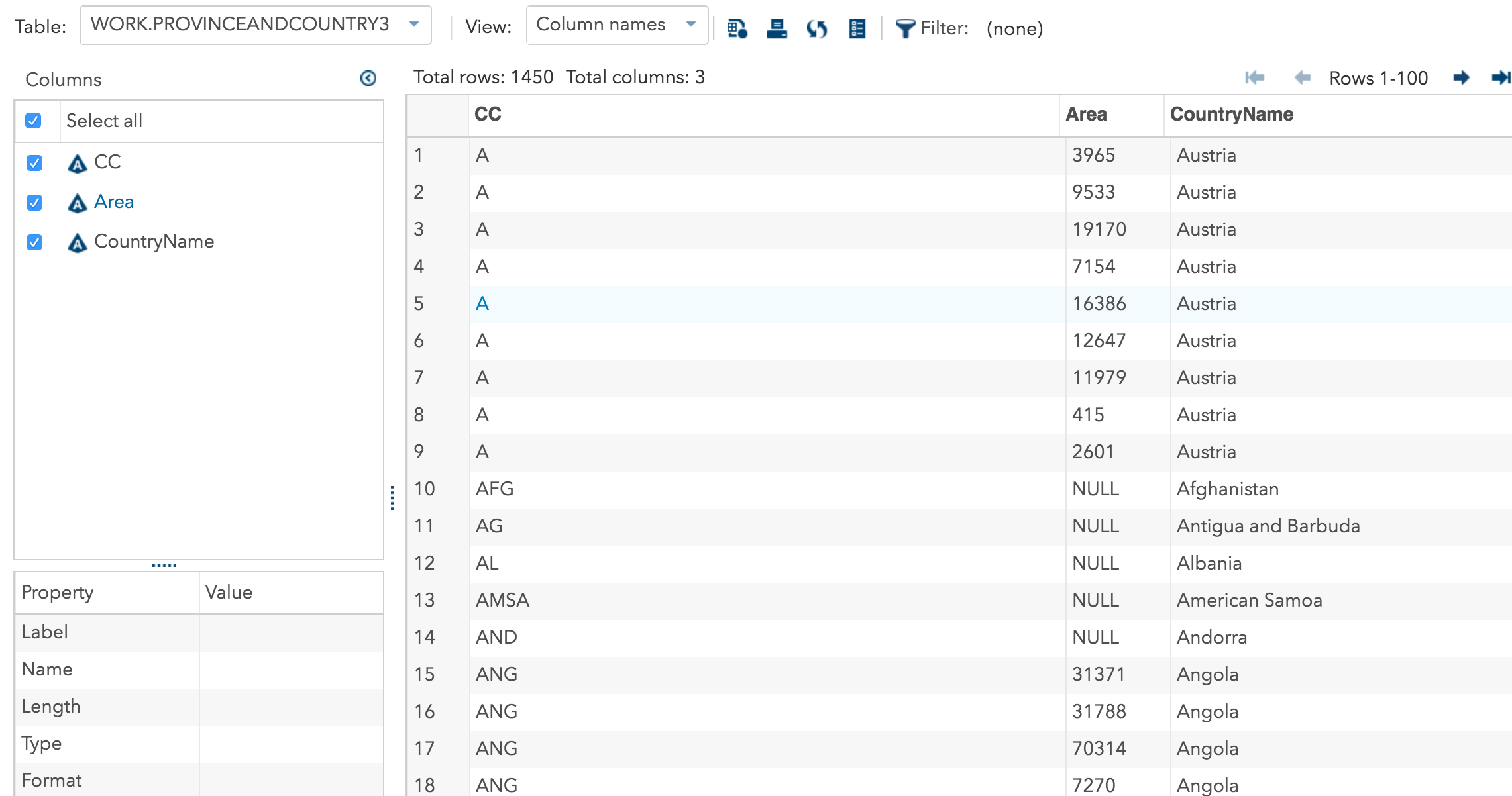
output out=AverageArea2 mean=;

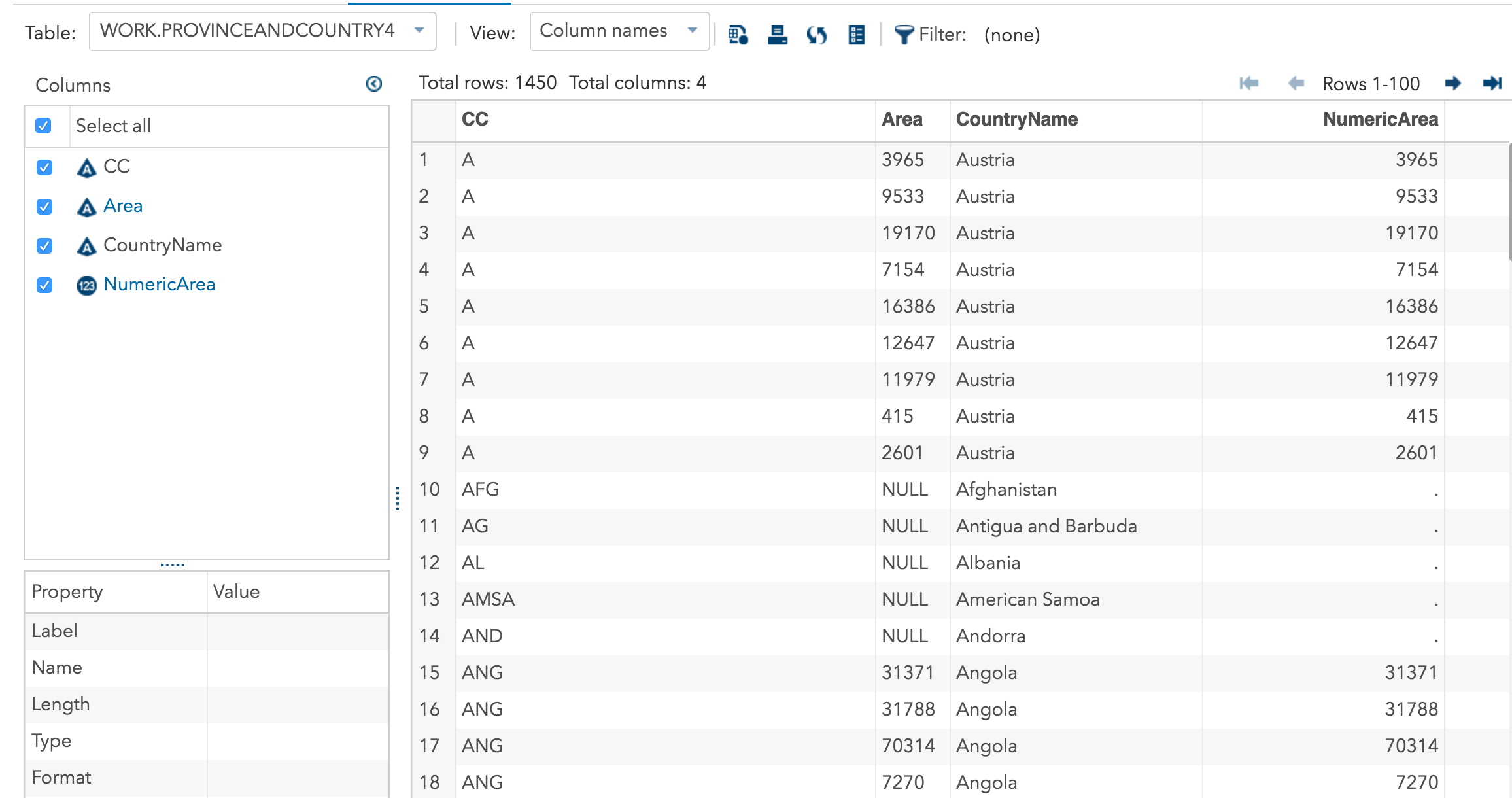
run;

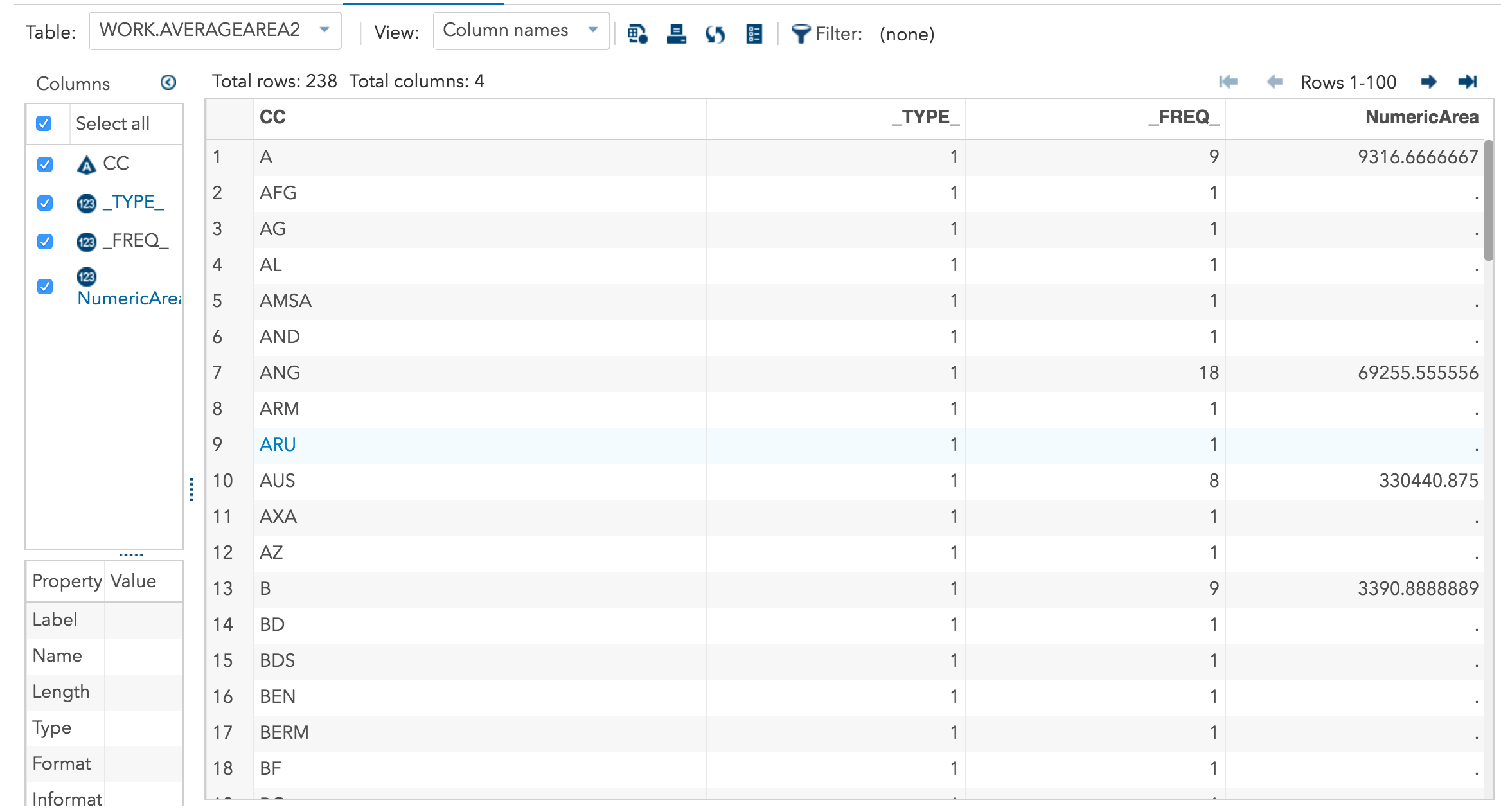
Proc print data=AverageArea2 (obs=20 keep= CC NumericArea);

Run;

**Output:**

****

****

****

