kauno technologijos universitetas
1922

Lab. Work 3

Name: Ravinthiran Partheepan
Course: Data mining Methods
Submission date: 12/10/2023

Problem Statement:

Part 1:

Using a Library and plotting complete data
Plot employee data – only JobCode, EmpID, Salary
Plot employee details where they employee is Pilot

Part 2:

Using a Library and plotting data Display
All goals – Library
All Sales – Library
For all sales data – visualize data for sales in Europe only
For all sales data – visualize data for sales in North America only
Can you visualize data as bar charts too to compare total sales between EU and America

Part 3:

Using a Library and plotting data Display
Crew – Library (crew member who are hired after 1986)
Display data sorted by location

Tasks:

1. **Library reference for data directory and Data Import**

   Code Snippet:

```
libname data '/home/u63603566/Lab-3';

/* Step 1: Import the emplist data */
data data.emplist;
  infile '/home/u63603566/Lab-3/emplist.dat' truncover;
  input FirstName $ LastName $ JobCode $ HireDate $ Salary;
run;

/* Step 2: Import the airports data */
data data.airports;
  infile '/home/u63603566/Lab-3/airports.dat' truncover;
  input AirportName $ Country $;
run;

/* Step 3: Import the sales_north_america data */
data sales_north_america;
  infile '/home/u63603566/Lab-3/allsales.sas7bdat';
```

```
 input Month $ Region $ Sales$;
run;

/* Step 4: Import the sales_europe data */
data sales_europe;
  infile '/home/u63603566/Lab-3/allsales.sas7bdat';
  input Month $ Region $ Sales $;
run;

/* Step 5: Import the allSales data */
data all_sales;
  infile '/home/u63603566/Lab-3/allsales.sas7bdat';
  input Month $ Region $ Sales $;
run;
```

Reference:
https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/basess/p0ppzdo5n8sdpdn1mokpbaxt880z.htm

2. Part 1:

[1] 2.1 Plot employee data code snippet:

```
/* Plot JobCode, EmpID, and Salary for all employees */
ods listing close; /* Close the default output */
ods graphics / width=800px height=600px imagefmt=png;

/* plot code */
proc sgplot data=data.emplist;
   scatter x=JobCode y=Salary / markerattrs=(symbol=circlefilled);
   xaxis label="JobCode";
   yaxis label="Salary";
   title "Employee Data - JobCode vs Salary";
run;

/* Save the plot as a PNG file */
ods graphics / reset;
ods listing;
```

2.2 Plot employee details where the employee is pilot

```
/* Plot employee details for employees with JobCode "Pilot" */
```

```
proc sgplot data=data.emplist;
   where JobCode = "Pilot";
   scatter x=JobCode y=Salary / markerattrs=(symbol=circlefilled);
   xaxis label="JobCode";
   yaxis label="Salary";
   title "Pilot Employee Data - EmpID vs Salary";
run;
```

2.3 Visualization of employee data – only JobCode, EmpID, Salary

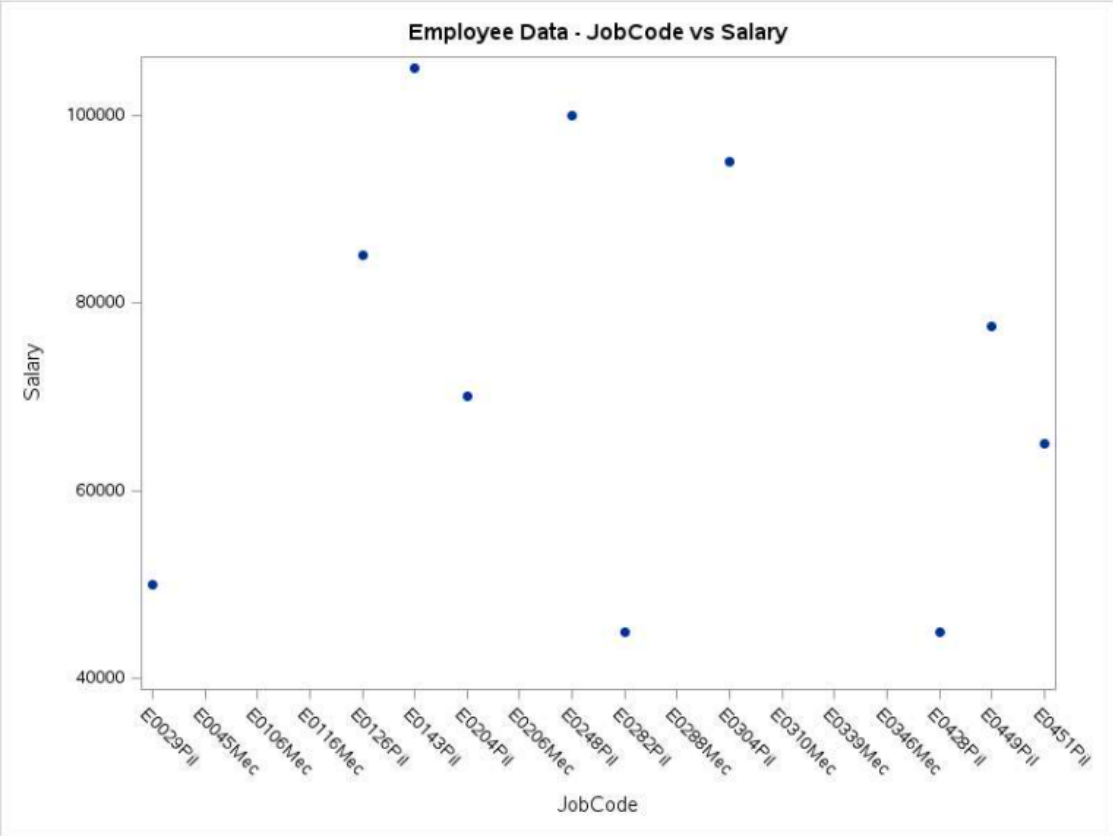

Figure 1: JobCode, and Salary for "Pilots"

3. Part 2:

Using a Library and plotting data Display
All goals – Library
All Sales – Library

For all sales data – visualize data for sales in Europe only
For all sales data – visualize data for sales in North America only
Can you visualize data as bar charts too to compare total sales between EU and America

Code Snippet for this section:

```
/* Your plot code */

proc sgplot data=data.emplist;
   scatter x=JobCode y=Salary / markerattrs=(symbol=circlefilled);
   xaxis label="JobCode";
   yaxis label="Salary";
   title "Employee Data - JobCode vs Salary";
run;

/* Save the plot as a PNG file */
ods graphics / reset;
ods listing;

/* Plot employee details for employees with JobCode "Pilot" */
proc sgplot data=data.emplist;
   where JobCode = "Pilot";
   scatter x=JobCode y=Salary / markerattrs=(symbol=circlefilled);
   xaxis label="JobCode";
   yaxis label="Salary";
   title "Pilot Employee Data - EmpID vs Salary";
run;

/* Plot all sales data */
proc sgplot data=data.allsales;
   vbar Month / response=Sales group=Region datalabel;
   xaxis label="Month";
   yaxis label="Sales";
   title "All Sales Data by Region";
run;

/* Plot sales data for Europe */
proc sgplot data=data.allsales;
   where Region = "Europe";
   vbar Month / response=Sales group=Region datalabel;
   xaxis label="Month";
   yaxis label="Sales";
   title "Sales Data for Europe by Month";
```

```
run;

/* Plot sales data for North America */
proc sgplot data=data.allsales;
   where Region = "North America";
   vbar Month / response=Sales group=Region datalabel;
   xaxis label="Month";
   yaxis label="Sales";
   title "Sales Data for North America by Month";

run;

/* Compare total sales between Europe and North America */
proc sgplot data=data.allsales;
   vbar Region / response=Sales group=Region datalabel;
   xaxis label="Region";
   yaxis label="Total Sales";
   title "Total Sales Comparison between Europe and North America";
run;
```

3.1 All Sales – Library

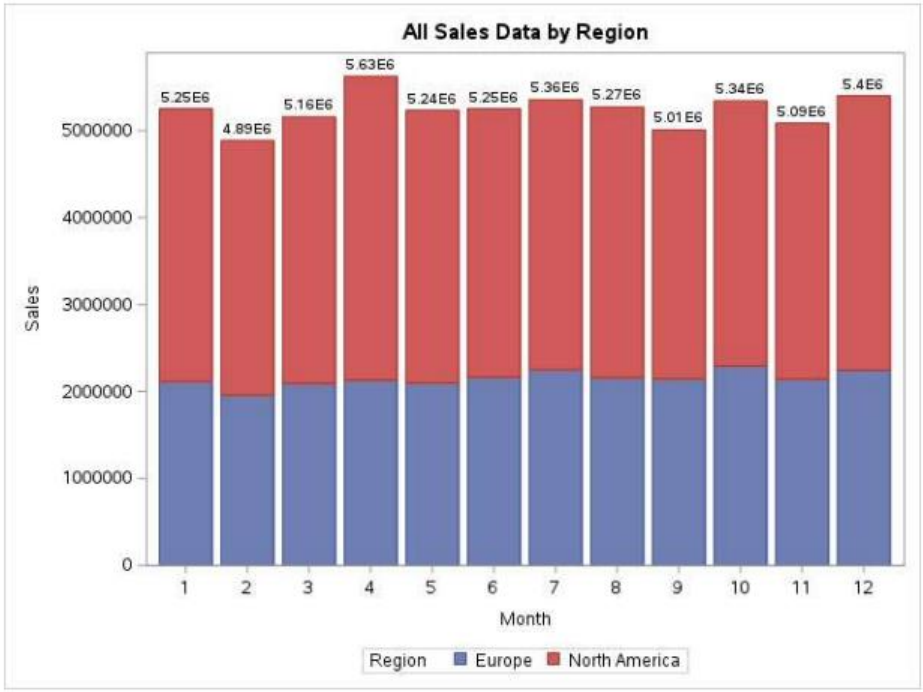

Figure 5: Total Sales in Europe and North America

3.2 For all sales data – visualize data for sales in Europe only

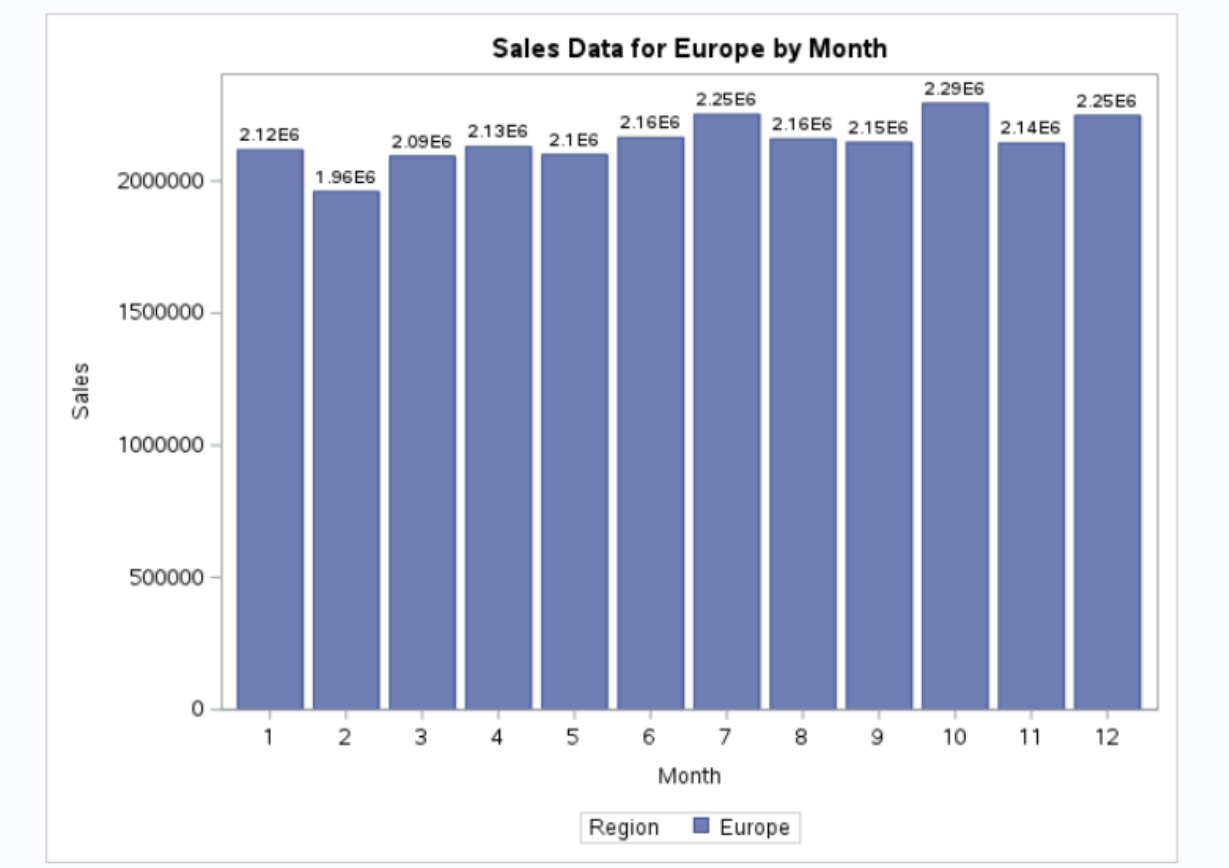

Figure 5: Sales data in Europe only

Note: Here the notation E^6 represents the number of zeros in the sales frequency. For example, In the x-axis for the 1st month the sales frequency us around 2,12 E6 which is followed by six zeros,

3.3 For all sales data – visualize data for sales in North America only

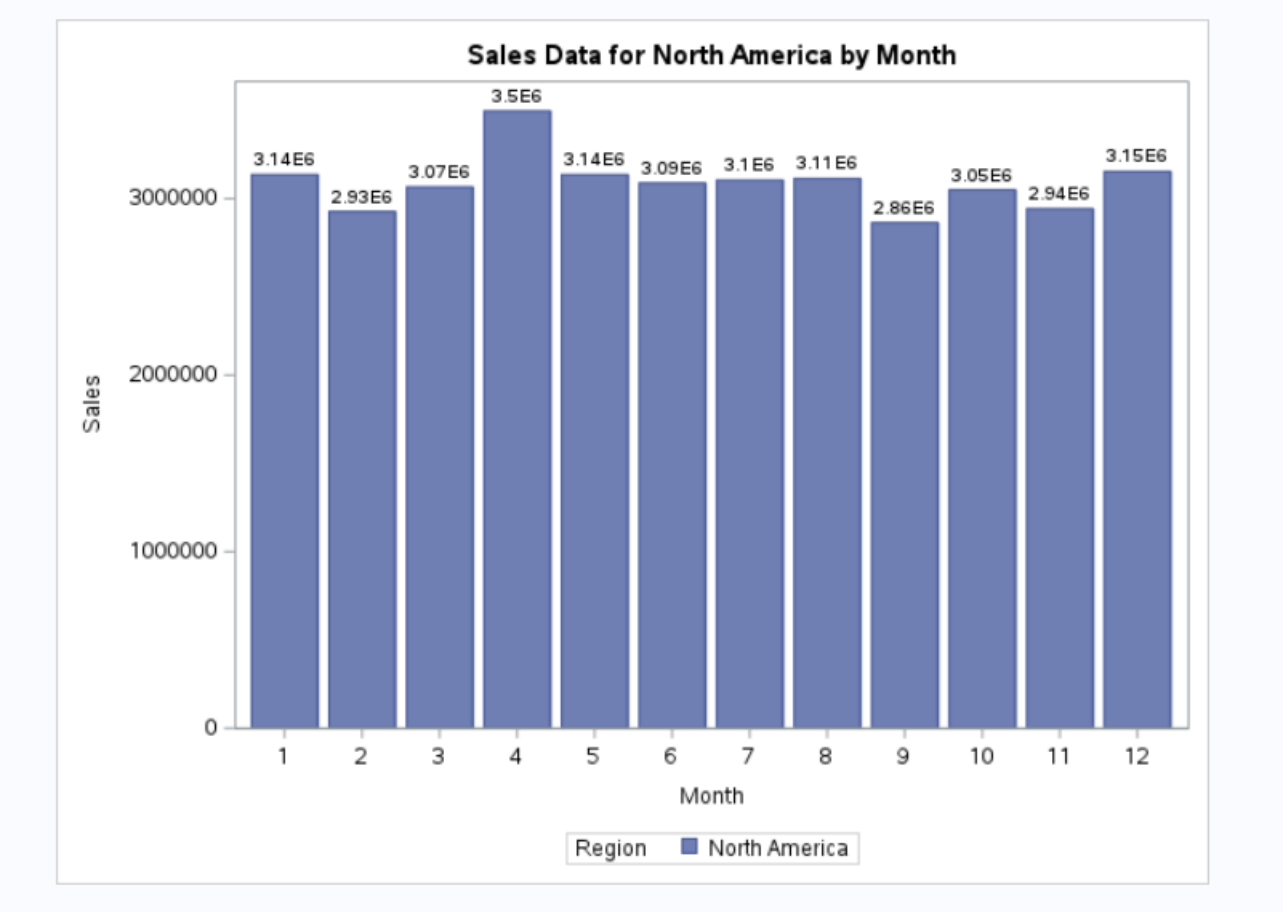

Figure 5: Sales data in North America only

Note: Here the notation E^6 represents the number of zeros in the sales frequency. For example, In the x-axis for the 1st month the sales frequency us around 2,12 E6 which is followed by six zeros,

3.4 Can you visualize data as bar charts too to compare total sales between EU and America?

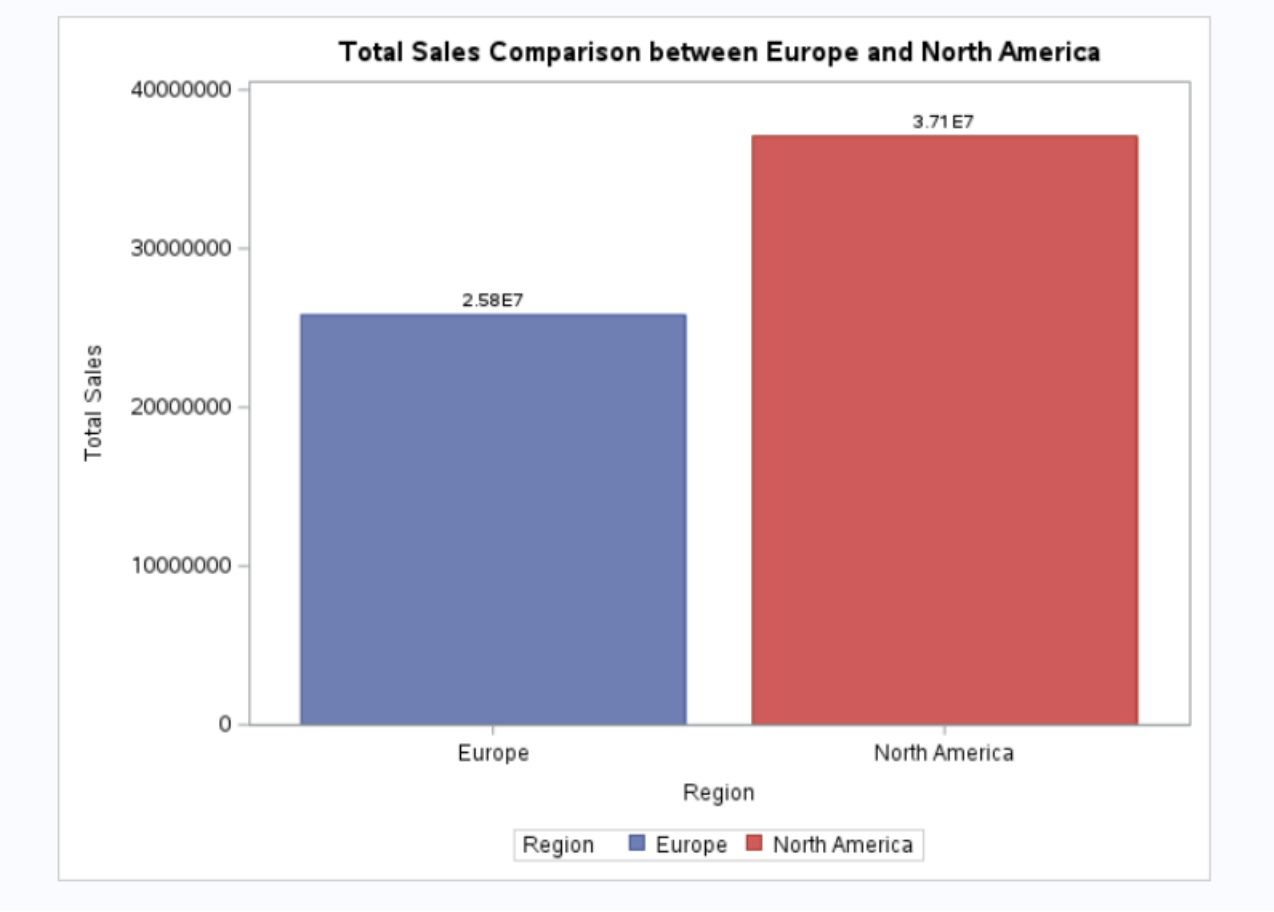

Figure 5: comparison total sales between EU and America

Note: Here the notation E^6 represents the number of zeros in the sales frequency. For example, In the x-axis for the 1st month the sales frequency us around 2,12 E6 which is followed by six zeros.

Interpreting the visualization: The above visualization states there are many sales done in North America market compared to European market.

4. Part – 3

Crew – Library (crew member who are hired after 1986)
Display data sorted by location

4.1 Code Snippet:

```
data crew_data;
   set data.emplist;

   new_HireDate = input(HireDate, mmddyy10.);


  if new_HireDate > '01JAN1960'd then
      HireDate = new_HireDate;
   else
      HireDate = .;

   if JobCode = "Crew" and not missing(HireDate) and HireDate >= '01JAN1986'd;
run;

/* Sort the crew data by HireDate */
proc sort data=crew_data;
   by HireDate;
run;

/* Display crew data sorted by location */
proc print data=crew_data;
   var FirstName LastName JobCode HireDate Salary;
   title "Crew Members Hired after 1986 Sorted by Location";
run;
```

Reference:

[1]https://documentation.sas.com/doc/en/lrcon/9.4/n0goqxbkj402cin11wm5k3mdxmb8.htm

Github Repository: https://github.com/ravinthiranpartheepan1407/SAS/tree/main/Task-3