Name: Siddharth Mohanty
Branch: EXTC
UID: 2019120038
Batch: B
Course: Data Analytics Lab
Experiment: 2

AIM: : Using the SAS software to analyze statistical data.

Problem Statement: Study and understand the workings of SAS studio by referring to the online materials and documentation, etc., and then implement a small problem.
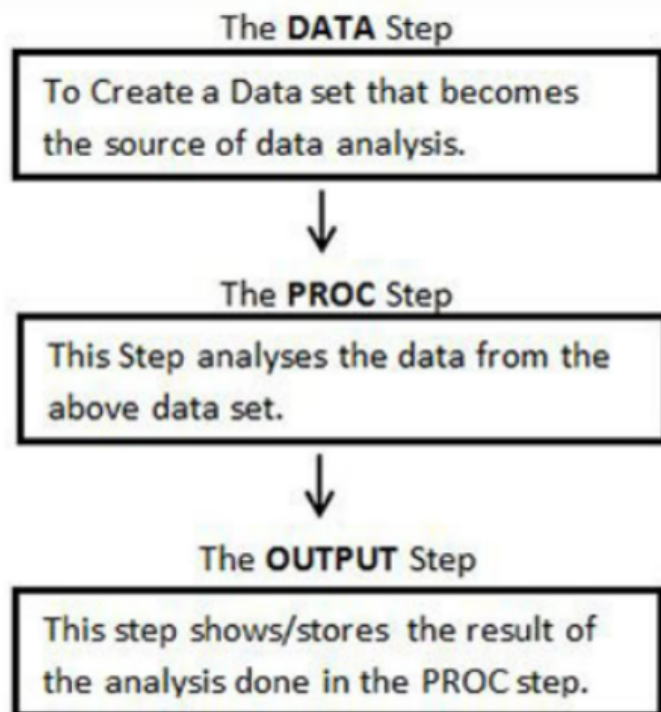
About SAS:
SAS stands for Statistical Analysis Software. It offers business intelligence and data management software and services through cutting-edge analytics. SAS turns data into insight, which might offer a new angle on how to conduct business.
In contrast to other BI solutions on the market, SAS uses considerable programming to transform and analyze data rather than just a simple drag-and-drop method.
Over the years, SAS has expanded its product offering with a number of solutions. It offers solutions for issues including fraud prevention, data governance, data quality, big data analytics, text mining, and health science, among others. We can safely assume SAS has a solution for every business domain and hence the popularity.

This is the standard procedure for creating a SAS program:

The **DATA** Step

To Create a Data set that becomes the source of data analysis.

↓

The **PROC** Step

This Step analyses the data from the above data set.

↓

The **OUTPUT** Step

This step shows/stores the result of the analysis done in the PROC step.

Code & Output:
The dataset that I have used is an inbuilt data:SASHELP -> BIRTHWGT

```
 1  DATA outdata;
 2  SET SASHELP.birthwgt;
 3  proc print data=outdata(obs=10);
 4
 5  proc univariate data=outdata;
 6  var AgeGroup;
 7  histogram;
 8
 9  proc sgplot data=outdata;
10  vbar Race;
11  run;
12
13  proc freq data=outdata order=data;
14      tables Drinking*Smoking / chisq;
15  run;
16
17  ods graphics on;
18  proc freq data=outdata;
19     tables Race*Smoking /
20              chisq cmh plots(only)=freqplot;
21  run;
22  ods graphics off;
```

The data looks like this:

| Obs	LowBirthWgt	Married	AgeGroup	Race	Drinking	Death	Smoking	SomeCollege
1	No	No	3	Asian	No	No	No	Yes
2	No	No	2	White	No	No	No	No
3	Yes	Yes	2	Native	No	Yes	No	No
4	No	No	2	White	No	No	No	No
5	No	No	2	White	No	No	No	Yes
6	No	No	2	White	No	No	No	
7	No	No	2	Asian	No	No	No	Yes
8	No	No	3	White	No	No	No	Yes
9	No	Yes	1	Black	No	No	No	No
10	No	No	2	Native	No	No	No	Yes

Pie Chart visualization for Age Group:
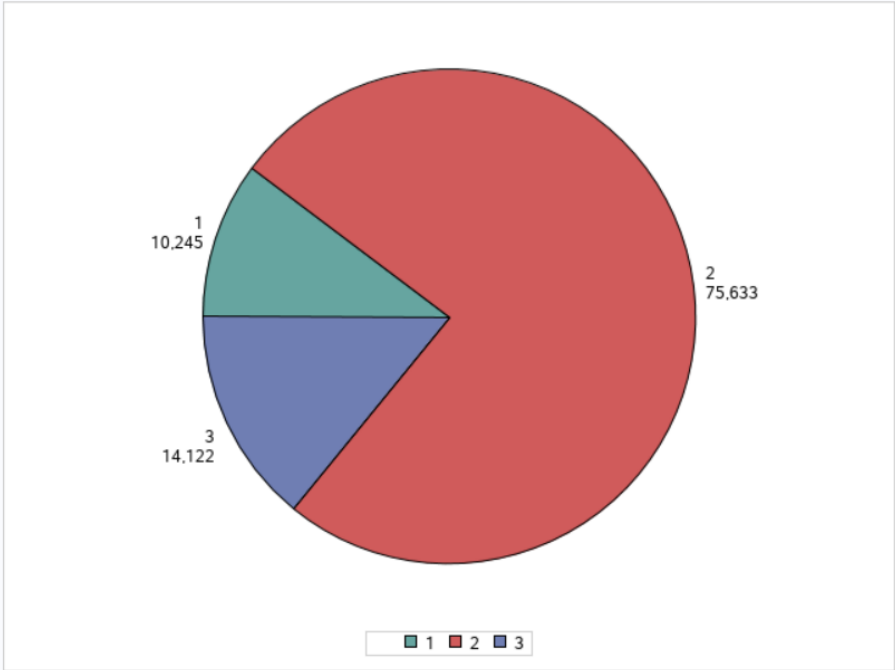
Univariate Output for AgeGroup:

The UNIVARIATE Procedure
Variable: AgeGroup

Moments			
N	100000	Sum Weights	100000
Mean	2.03877	Sum Observations	203877
Std Deviation	0.49210701	Variance	0.24216931
Skewness	0.08848977	Kurtosis	1.08996298
Uncorrected SS	439875	Corrected SS	24216.6887
Coeff Variation	24.1374461	Std Error Mean	0.00155618

Basic Statistical Measures			
Location		Variability	
Mean	2.038770	Std Deviation	0.49211
Median	2.000000	Variance	0.24217
Mode	2.000000	Range	2.00000
		Interquartile Range	0

Distribution of AgeGroup

Distribution of Race by Smoking

Distribution of Race by Smoking
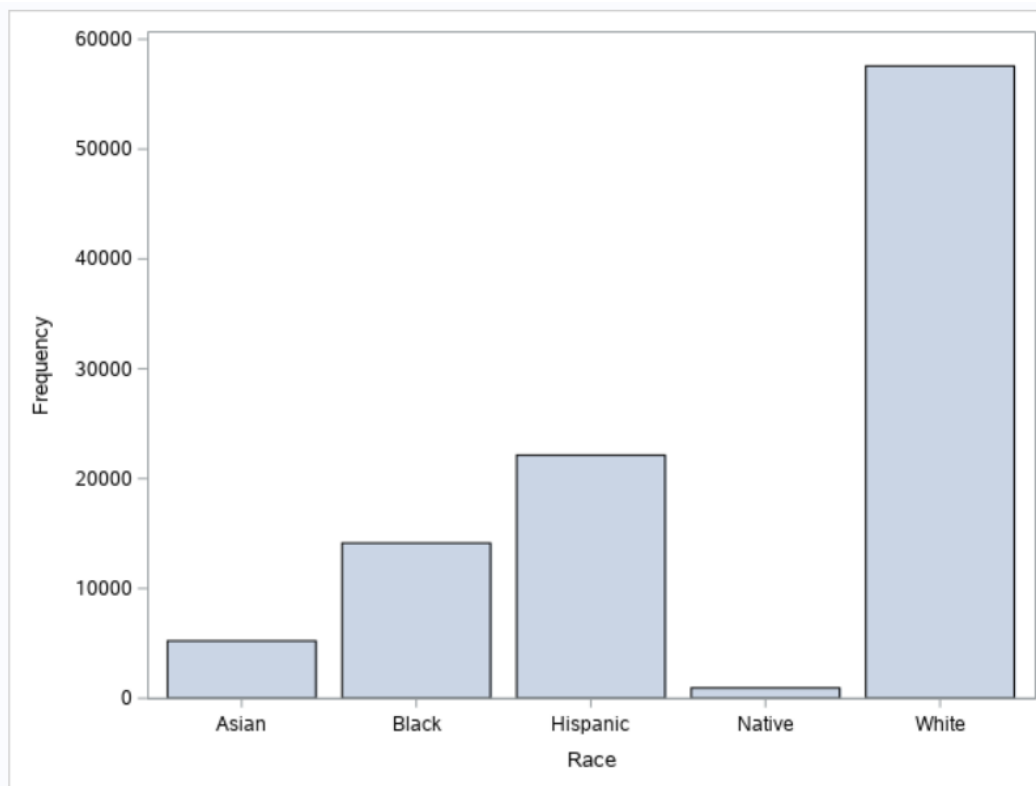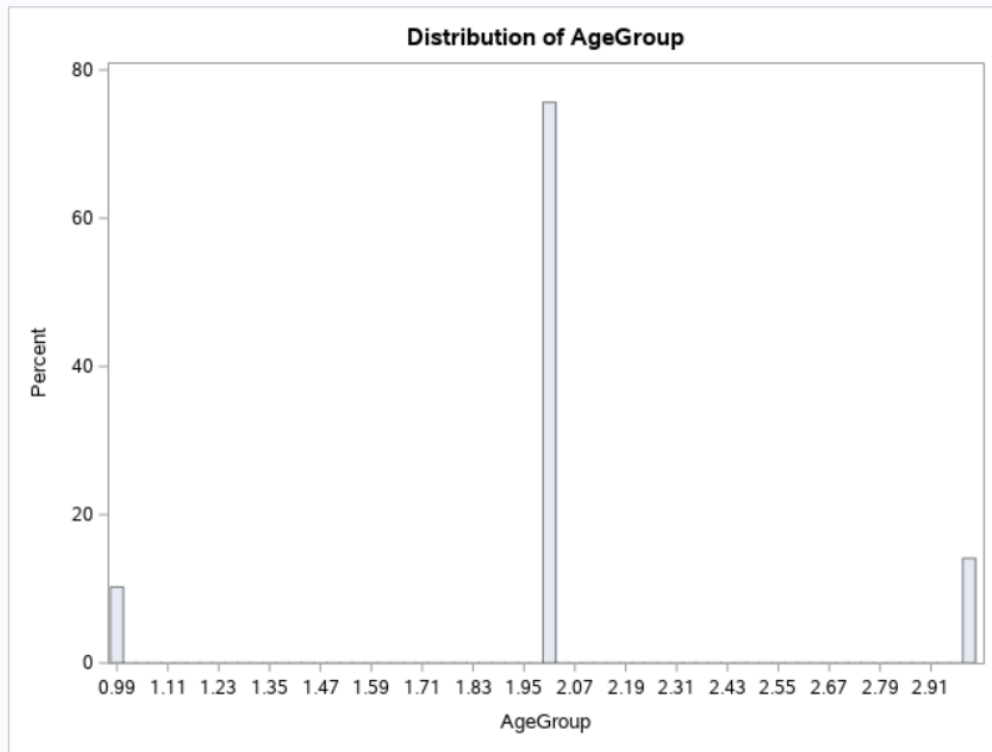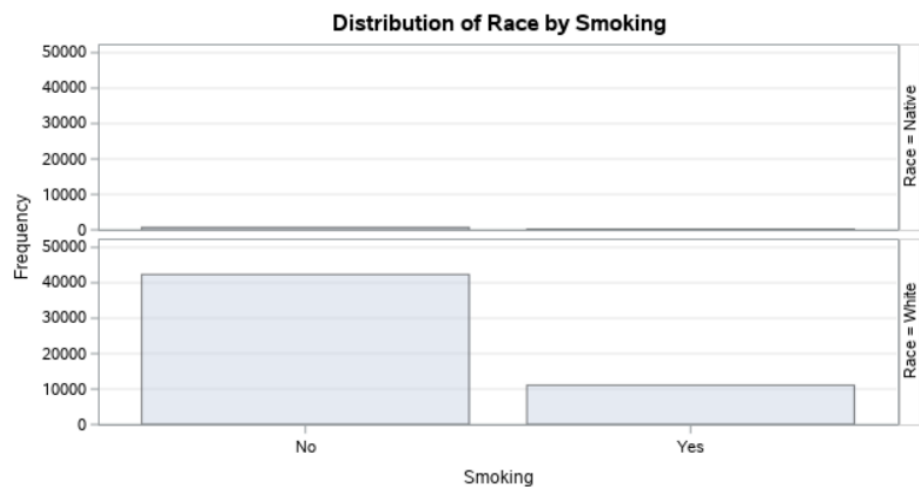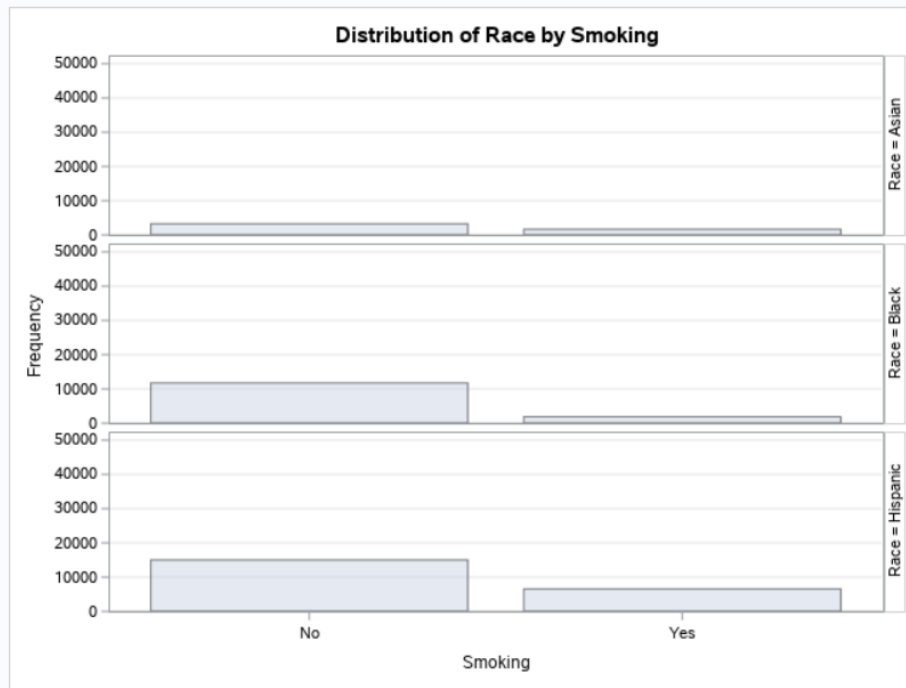
Conclusion:
After learning about SAS studio, its features, functionalities and also implementing them, I was able to conclude that : SAS provides us various functionalities like Data Management, Statistical Analysis, Report formation with perfect graphics etc that I have used in the above experiment.