**Assignment**

**Aim**: To perform data analytics and generate various reports using Base SAS constructs.

**Prerequisite:**

1. Understanding of programming constructs of Base SAS

**Outcome:** After successful completion of this experiment students will be able to

1. Understand the approach used in a Analytical task.
2. To generate various reports that would be useful as per the scenario given.

(TO BE COMPLETED BY STUDENTS)

|  |  |
| --- | --- |
| Roll No. A020 | Name: Nicole Michelle Dsouza |
| Class: B.Tech IT | Batch:A1 |
| Date of Practical: | Date of Submission: |
| Grade: |  |

A supermarket has branches across many cities. Every branch manages data related to customers, sales, suppliers and also stock. The management of this supermarket chain decides to make use of the huge amount of data that they have been collecting over several years to find interesting patterns that would help them to improve their business. Describe how programmers employed in this supermarket chain can go about doing this with Base SAS as a programming language for Analytics. (Note: Explain the solution with specific constructs in the language that will help to accomplish the task at hand).

1. The programmers will first employee **proc print** to examine each of the datasets and identify any discrepancy in the column names, this is important for later.
2. The discrepancies in the column values e.g one branch uses M and F for gender while another uses Male and Female will be managed using **user defined formats, dates and currency related formats** will also be managed in the **proc format** step.
3. To perform better analysis all, branch wise datasets will be **merged in the data step**, the column names will be changed in line to common column names using the **rename option within the merge statement**. After merging the dataset and forming a large dataset of all stores. The missing values and shift in columns will be managed using **dsd and** **missover**.
4. After this the programmers will use **proc contents** to identify the formats different columns are in this will further aid data analysis.
5. Post this the programmers can employ **proc freq** and arrange by descending order of frequency to identify it’s most loyal customers. **Proc Freq** can also be used to identify the most commonly bought items.
6. Using **proc means** we can find the branch with maximum sales and minimum sales.
7. By Performing a univariate analysis using **proc univariate** of each branch and each category we can identify important metric like most sold items, mean amount spent.
8. The **Proc Means** function can also be employed to find the most profitable and least profitable cities. The programmers can perform a bivariate analysis of the data to identify if there’s a correlation between products sold at a particular time of the year e.g. colors during holi which can be done using **Pearson’s correlation using proc corr**,
9. the programmers can further verify if there is a relation between gender and particular products sold this can be done using chi square test using **proc freq with chisq** as an option.
10. Using functions like average in **proc means** the programmers can also find average number of items sold and average amount of money spent per transaction.
11. Using the **proc reg** the programmers can predict the sales of supermarket over time with the help of linear regression.
12. Finally, all this analysis can be presented in the form of a pdf, rtf, html report in a nice format using **sas ods styles**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***