**Project – Attrition Analysis**

**#1 IMPORT THE REQUIRED DATASET**

**CODE:** libname SAS\_Clas "/folders/myfolders";  
  
FILENAME REFFILE '/folders/myfolders/Copy of Project 03\_Attrition Analysis\_Datasets.xlsx';  
  
PROC IMPORT DATAFILE=REFFILE  
 DBMS=XLSX  
 OUT=WORK.Attrition replace;  
 GETNAMES=YES;  
RUN;  
  
Proc contents data=attrition;  
run;

**/\*DROPPING THE UNWANTED VARIABLES\*/**  
  
data attrition1(drop=f g h i j k l);  
set attrition;  
run;  
  
Proc contents data=attrition1;  
run;

**THE CONTENTS PROCEDURE**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Set Name** | WORK.ATTRITION1 | **Observations** | 50 |
| **Member Type** | DATA | **Variables** | 5 |
| **Engine** | V9 | **Indexes** | 0 |
| **Created** | 04/27/2019 17:24:04 | **Observation Length** | 40 |
| **Last Modified** | 04/27/2019 17:24:04 | **Deleted Observations** | 0 |
| **Protection** |  | **Compressed** | NO |
| **Data Set Type** |  | **Sorted** | NO |
| **Label** |  |  |  |
| **Data Representation** | SOLARIS\_X86\_64, LINUX\_X86\_64, ALPHA\_TRU64, LINUX\_IA64 |  |  |
| **Encoding** | utf-8 Unicode (UTF-8) |  |  |

| **Engine/Host Dependent Information** | |
| --- | --- |
| **Data Set Page Size** | 65536 |
| **Number of Data Set Pages** | 1 |
| **First Data Page** | 1 |
| **Max Obs per Page** | 1632 |
| **Obs in First Data Page** | 50 |
| **Number of Data Set Repairs** | 0 |
| **Filename** | /tmp/SAS\_workE6C80000686D\_localhost.localdomain/SAS\_work3CFF0000686D\_localhost.localdomain/attrition1.sas7bdat |
| **Release Created** | 9.0401M6 |
| **Host Created** | Linux |
| **Inode Number** | 671713 |
| **Access Permission** | rw-rw-r-- |
| **Owner Name** | sasdemo |
| **File Size** | 128KB |
| **File Size (bytes)** | 131072 |

| **Alphabetic List of Variables and Attributes** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **#** | **Variable** | **Type** | **Len** | **Format** | **Label** |
| **1** | Employee\_ID | Num | 8 | BEST. | Employee\_ID |
| **5** | Marital\_Status | Num | 8 | BEST. | Marital\_Status |
| **4** | Relocation\_Indicator | Num | 8 | BEST. | Relocation\_Indicator |
| **2** | Retain\_Indicator | Num | 8 | BEST. | Retain\_Indicator |
| **3** | Sex\_Indicator | Num | 8 | BEST. | Sex\_Indicator |

**#2 CHECK THE FREQUECY OF CHURN**

**CODE:**

**/\* CHECK THE FREQUECNY OF CHURN\*/**

proc freq data=attrition1;  
table retain\_indicator;  
run;

| **Retain\_Indicator** | | | | |
| --- | --- | --- | --- | --- |
| **Retain\_Indicator** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **0** | 22 | 44.00 | 22 | 44.00 |
| **1** | 28 | 56.00 | 50 | 100.00 |

Event to non-event percentage is 56 to 44.

56% are churned as per the data available.

**#3 DESCRIPTIVE ANALYSIS**

**CODE:**

proc means data=attrition1;  
run;

| **Variable** | **Label** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| Employee\_ID Retain\_Indicator Sex\_Indicator Relocation\_Indicator Marital\_Status | Employee\_ID Retain\_Indicator Sex\_Indicator Relocation\_Indicator Marital\_Status | 50 50 50 50 50 | 25.5000000 0.5600000 0.5600000 0.4800000 0.4200000 | 14.5773797 0.5014265 0.5014265 0.5046720 0.4985694 | 1.0000000 0 0 0 0 | 50.0000000 1.0000000 1.0000000 1.0000000 1.0000000 |

**CODE:** ODS graphics on;  
proc univariate data=attrition1;  
class retain\_indicator;  
var sex\_indicator marital\_status relocation\_indicator ;  
histogram sex\_indicator marital\_status relocation\_indicator/normal;  
run;  
ODS graphics off;

proc sgplot data=attrition1;  
vbox retain\_indicator/category=sex\_indicator;  
run;

proc sgplot data=attrition1;  
vbox retain\_indicator/category=marital\_status;  
run;

proc sgplot data=attrition1;  
vbox retain\_indicator/category=relocation\_indicator ;  
run;

**Descriptive Statistics on Employee Retention based on Gender, Marital status & Relocation**

**GENDER**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 28 | **Sum Weights** | 28 |
| **Mean** | 0.60714286 | **Sum Observations** | 17 |
| **Std Deviation** | 0.49734746 | **Variance** | 0.2473545 |
| **Skewness** | -0.464 | **Kurtosis** | -1.9280954 |
| **Uncorrected SS** | 17 | **Corrected SS** | 6.67857143 |
| **Coeff Variation** | 81.9160525 | **Std Error Mean** | 0.09398984 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 0.607143 | **Std Deviation** | 0.49735 |
| **Median** | 1.000000 | **Variance** | 0.24735 |
| **Mode** | 1.000000 | **Range** | 1.00000 |
|  |  | **Interquartile Range** | 1.00000 |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 6.459665 | **Pr > |t|** | <.0001 |
| **Sign** | **M** | 8.5 | **Pr >= |M|** | <.0001 |
| **Signed Rank** | **S** | 76.5 | **Pr >= |S|** | <.0001 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 1 |
| **99%** | 1 |
| **95%** | 1 |
| **90%** | 1 |
| **75% Q3** | 1 |
| **50% Median** | 1 |
| **25% Q1** | 0 |
| **10%** | 0 |
| **5%** | 0 |





The Mean Median and Mode have close values, the histogram shows the datas are more dense at the lower and higher ends for both the Genders.

The box plot shows the data for retain indicators- male is around 50% whereas for females its in the 75%quatile.

**MARITAL STATUS**

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 28 | **Sum Weights** | 28 |
| **Mean** | 0.35714286 | **Sum Observations** | 10 |
| **Std Deviation** | 0.48795004 | **Variance** | 0.23809524 |
| **Skewness** | 0.63058159 | **Kurtosis** | -1.7316923 |
| **Uncorrected SS** | 10 | **Corrected SS** | 6.42857143 |
| **Coeff Variation** | 136.62601 | **Std Error Mean** | 0.09221389 |

| **Basic Statistical Measures** | | | |
| --- | --- | --- | --- |
| **Location** | | **Variability** | |
| **Mean** | 0.357143 | **Std Deviation** | 0.48795 |
| **Median** | 0.000000 | **Variance** | 0.23810 |
| **Mode** | 0.000000 | **Range** | 1.00000 |
|  |  | **Interquartile Range** | 1.00000 |

| **Tests for Location: Mu0=0** | | | | |
| --- | --- | --- | --- | --- |
| **Test** | **Statistic** | | **p Value** | |
| **Student's t** | **t** | 3.872983 | **Pr > |t|** | 0.0006 |
| **Sign** | **M** | 5 | **Pr >= |M|** | 0.0020 |
| **Signed Rank** | **S** | 27.5 | **Pr >= |S|** | 0.0020 |

| **Quantiles (Definition 5)** | |
| --- | --- |
| **Level** | **Quantile** |
| **100% Max** | 1 |
| **99%** | 1 |
| **95%** | 1 |
| **90%** | 1 |
| **75% Q3** | 1 |
| **50% Median** | 0 |
| **25% Q1** | 0 |





The histogram shows the datas are more dense at the lower and higher ends for both the married and unmarried group.

The box plot shows the data for retain indicators- married is around 60% whereas for unmarried it is in the 50%quatile.

**RELOCATION**

| **Basic Statistical Measures** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Location** | | | **Variability** | | | |
| **Mean** | 0.409091 | | **Std Deviation** | | 0.50324 | |
| **Median** | 0.000000 | | **Variance** | | 0.25325 | |
| **Mode** | 0.000000 | **Range** | | 1.00000 | |
|  |  | | **Interquartile Range** | | 1.00000 | |

| **Quantiles for Normal Distribution** | | |
| --- | --- | --- |
| **Percent** | **Quantile** | |
| **Observed** | **Estimated** |
| **1.0** | 0.00000 | -0.76161 |
| **5.0** | 0.00000 | -0.41866 |
| **10.0** | 0.00000 | -0.23583 |
| **25.0** | 0.00000 | 0.06966 |
| **50.0** | 0.00000 | 0.40909 |
| **75.0** | 1.00000 | 0.74852 |
| **90.0** | 1.00000 | 1.05401 |
| **95.0** | 1.00000 | 1.23684 |
| **99.0** | 1.00000 | 1.57979 |

| **Moments** | | | |
| --- | --- | --- | --- |
| **N** | 22 | **Sum Weights** | 22 |
| **Mean** | 0.40909091 | **Sum Observations** | 9 |
| **Std Deviation** | 0.50323628 | **Variance** | 0.25324675 |
| **Skewness** | 0.39742763 | **Kurtosis** | -2.0367072 |
| **Uncorrected SS** | 9 | **Corrected SS** | 5.31818182 |
| **Coeff Variation** | 123.013313 | **Std Error Mean** | 0.10729034 |





The histogram shows the datas are more dense at the lower end for employee who resigned due to relocation.

The box plot shows the data for retain indicators- relocation is not the reason for resign is around 60% whereas for relocation being the reasonfor resign is in the 50%quatile.

**#4 PERFORM LOGISTIC REGRESSION**

**CODE:**

proc logistic data=attrition1;  
model retain\_indicator=sex\_indicator marital\_status relocation\_indicator;  
output out=churned p=predict;  
run;

| **Model Information** | | |
| --- | --- | --- |
| **Data Set** | WORK.ATTRITION1 |  |
| **Response Variable** | Retain\_Indicator | Retain\_Indicator |
| **Number of Response Levels** | 2 |  |
| **Model** | binary logit |  |
| **Optimization Technique** | Fisher's scoring |  |

|  |  |
| --- | --- |
| **Number of Observations Read** | 50 |
| **Number of Observations Used** | 50 |

| **Response Profile** | | |
| --- | --- | --- |
| **Ordered Value** | **Retain\_Indicator** | **Total Frequency** |
| **1** | 0 | 22 |
| **2** | 1 | 28 |

|  |
| --- |
| ***Probability modeled is Retain\_Indicator='0'.*** |

| **Model Convergence Status** |
| --- |
| Convergence criterion (GCONV=1E-8) satisfied. |

| **Model Fit Statistics** | | |
| --- | --- | --- |
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 70.593 | 74.180 |
| **SC** | 72.505 | 81.829 |
| **-2 Log L** | 68.593 | 66.180 |

| **Testing Global Null Hypothesis: BETA=0** | | | |
| --- | --- | --- | --- |
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 2.4125 | 3 | 0.4913 |
| **Score** | 2.3714 | 3 | 0.4990 |
| **Wald** | 2.2750 | 3 | 0.5173 |

| **Analysis of Maximum Likelihood Estimates** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | -0.0861 | 0.5560 | 0.0240 | 0.8769 |
| **Sex\_Indicator** | 1 | -0.2796 | 0.5967 | 0.2196 | 0.6394 |
| **Marital\_Status** | 1 | 0.6589 | 0.6089 | 1.1709 | 0.2792 |
| **Relocation\_Indicator** | 1 | -0.5960 | 0.6038 | 0.9744 | 0.3236 |

| **Odds Ratio Estimates** | | | |
| --- | --- | --- | --- |
| **Effect** | **Point Estimate** | **95% Wald Confidence Limits** | |
| **Sex\_Indicator** | 0.756 | 0.235 | 2.435 |
| **Marital\_Status** | 1.933 | 0.586 | 6.375 |
| **Relocation\_Indicator** | 0.551 | 0.169 | 1.799 |

| **Association of Predicted Probabilities and Observed Responses** | | | |
| --- | --- | --- | --- |
| **Percent Concordant** | 56.0 | **Somers' D** | 0.250 |
| **Percent Discordant** | 31.0 | **Gamma** | 0.287 |
| **Percent Tied** | 13.0 | **Tau-a** | 0.126 |
| **Pairs** | 616 | **c** | 0.625 |

**#5 CHECK THE MAX AND MIN VALUES FOR THE PROBABILITY OF CHURN**

**CODE:**

proc means data = churned;  
run;

**The MEANS Procedure**

| **Variable** | **Label** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| Employee\_ID  Retain\_Indicator  Sex\_Indicator  Relocation\_Indicator  Marital\_Status  \_LEVEL\_  predict | Employee\_ID  Retain\_Indicator  Sex\_Indicator  Relocation\_Indicator  Marital\_Status  Response Value  Estimated Probability | 50  50  50  50  50  50  50 | 25.5000000  0.5600000  0.5600000  0.4800000  0.4200000  0  0.4400114 | 14.5773797  0.5014265  0.5014265  0.5046720  0.4985694  0  0.1092271 | 1.0000000  0  0  0  0  0  0.2765358 | 50.0000000  1.0000000  1.0000000  1.0000000  1.0000000  0  0.6394111 |

**#6 CREATE A NEW DATASET TO ADD ALL”CHURNED” EMPLOYEES ABOVE THE CUT-OFF VALUE**

**CODE:**

data churned\_emp;  
set churned;  
where predict>.5 and retain\_indicator=1;  
run;

Rows 1-3

Total rows: 3Total columns: 7

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Employee\_ID** | **Retain\_Indicator** | **Sex\_Indicator** | **Relocation\_Indicator** | **Marital\_Status** | **LEVEL\_** | |  |  | | --- | --- | | **pppppppppppPredictpredicpredic** |  |   **predict** |
| 1 | 35 | 1 | 0 | 0 | 1 | 0 | 0.6394111027 |
| 2 | 11 | 1 | 0 | 0 | 1 | 0 | 0.6394111027 |
| 3 | 29 | 1 | 1 | 0 | 1 | 0 | 0.5727800944 |

10% churned employees are above the cutoff value, and marital status of these employees is unmarried and relocation is the main reason to resign. Among the 10% churned above the cutoff value 65% are male.