PREDICTIVE ANALYSIS:

1. ANS:

Selected Top 3 features: **['Factor#20', 'Factor#25', 'Factor#36']**

Depending on the **Recursive Feature Elimination** (RFE)\* feature selection technique for Logistic Regression.(because the output is YES/No form or 1/0)

Other than this, correlation matrix is found to determine the linear relationship between the feature columns. And Factors 1,20,25 found correlational with Switched column. But Instead used RFE to get more precise columns.

1. ANS:

LOGISTIC REGRESSION is used as the Switched column has only 2 outputs: yes/no

Selected **Optimal number of features: 15 Selected features: ['Factor#7', 'Factor#8', 'Factor#13', 'Factor#16', 'Factor#18', 'Factor#20', 'Factor#21', 'Factor#22', 'Factor#24', 'Factor#25', 'Factor#28', 'Factor#31', 'Factor#33', 'Factor#34', 'Factor#36']** depending on the **Recursive feature elimination with cross-validation (RFE CV)** to select the optimum number of features.

**# 3. Building Logistic regression model with confusion matrix as method validation matrix and feature selection RFE and RFE CV methods.**

According to confusion matrix for RFE and RFE CV:

|  |  |  |
| --- | --- | --- |
|  | RFE | RFE cv |
| Confusion matrix  [[TP FP]  [FN TN]]  Accuracy | [[883 16]  [110 26]]  0.8782608 | [[885 14]  [108 28]]  0.88212 |
| Log Loss | 0.31926 | 0.3118439 |
| AUC for ROC | 0.586689 | 0.59515 |
| FPR | 0.01779 | 0.015572 |
| TPR | 0.19117 | 0.2058 |
|  |  |  |
|  |  |  |

Acuracy of RFE CV > accuracy of RFE

RFE cv

[[885 14]

[108 28]]