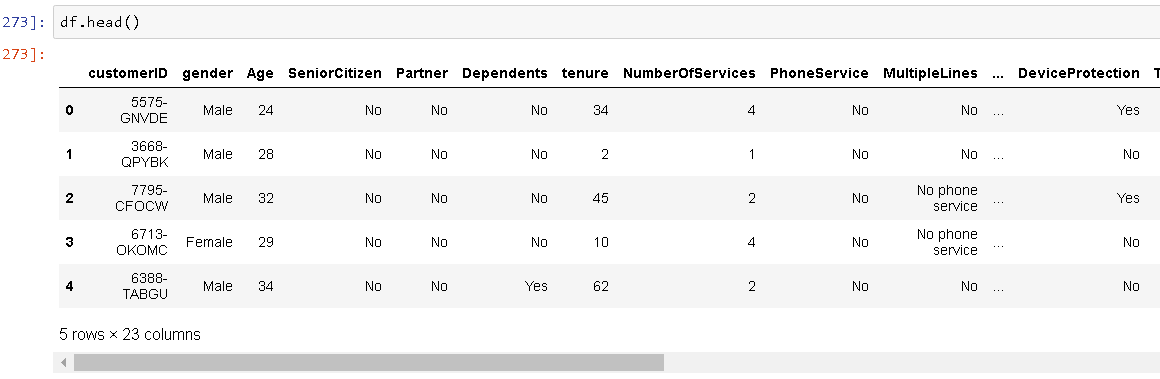
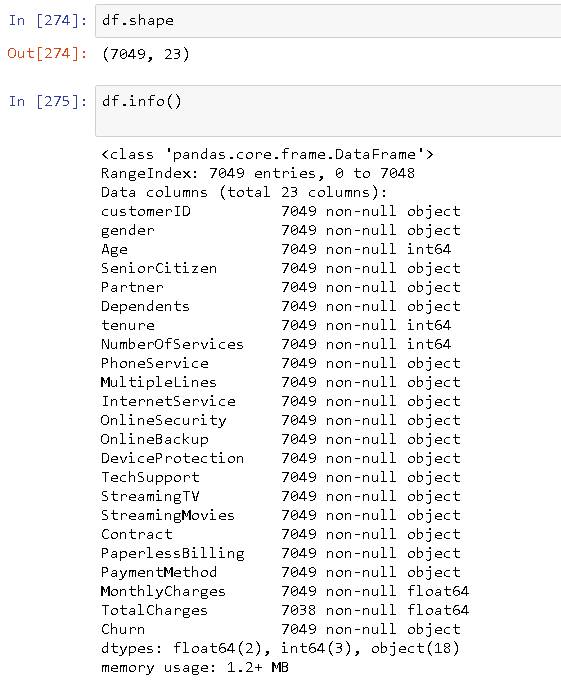
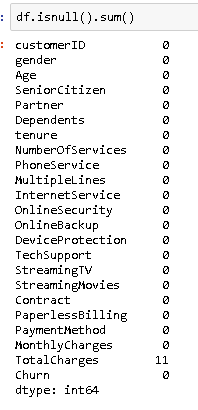
**APPENDIX A**

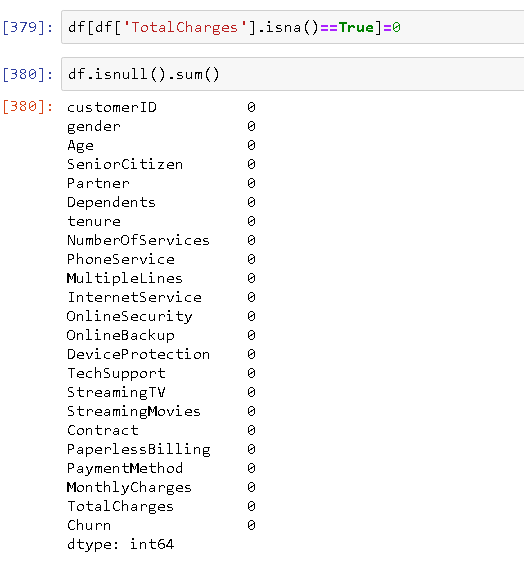
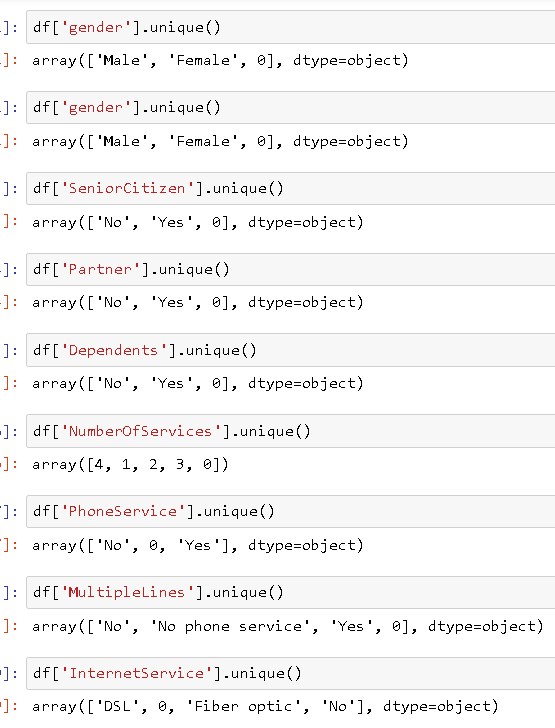
**Detailed Snapshots of results achieved from project:**

**DATA PRE-PROCESSING**

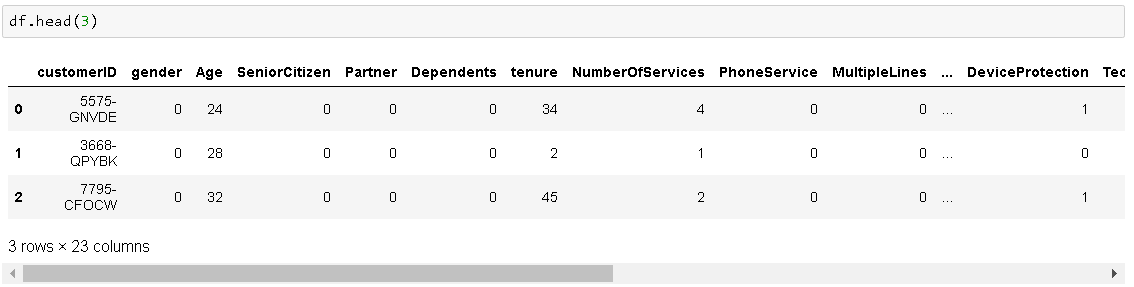
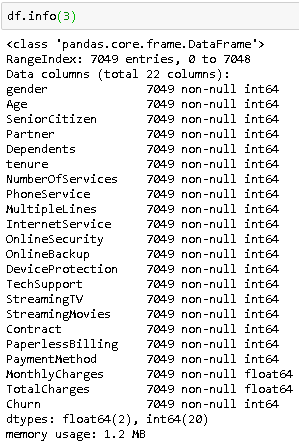
**EXPLORATORY DATA ANALYSIS**



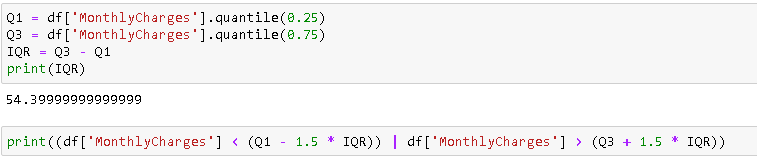
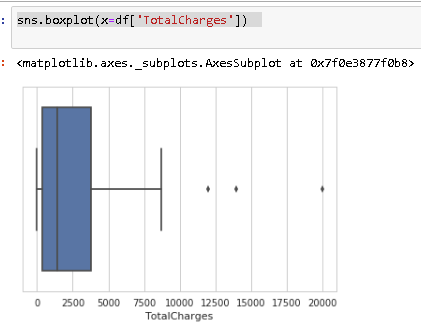
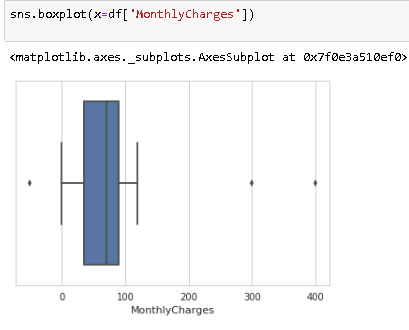
 



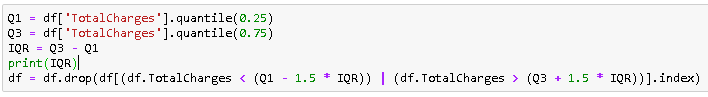




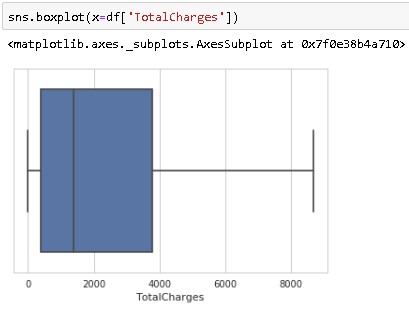
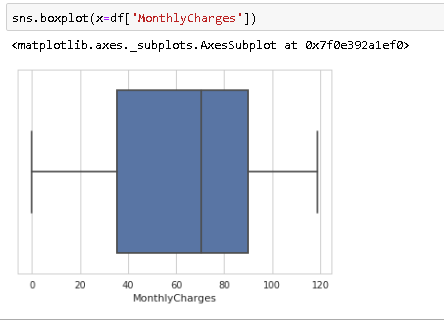
**DATA VISUALIZATION:**



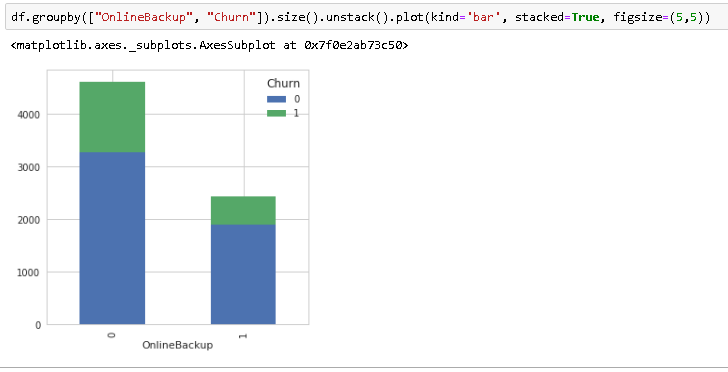
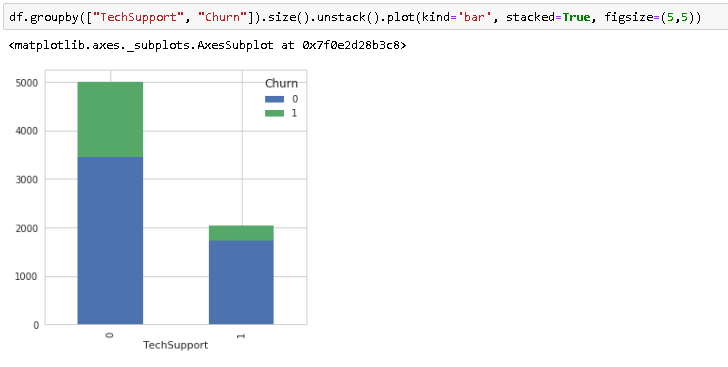
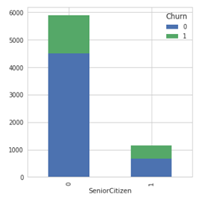
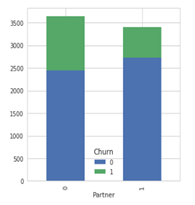
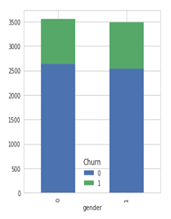


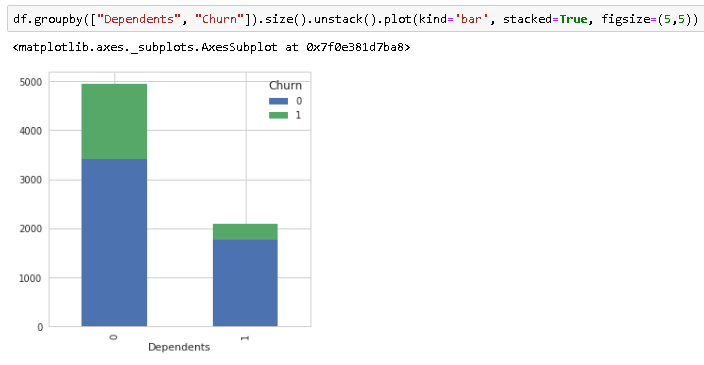
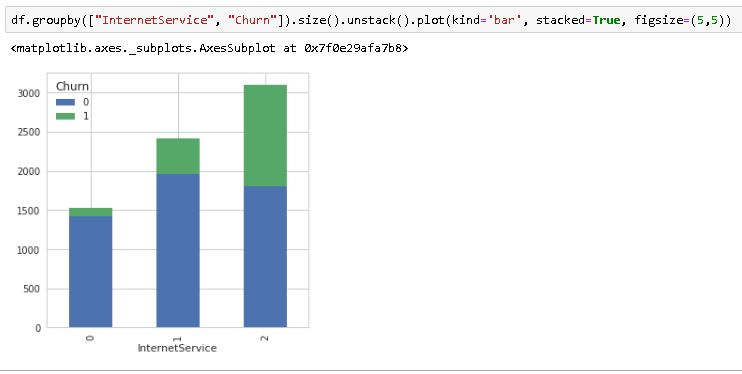
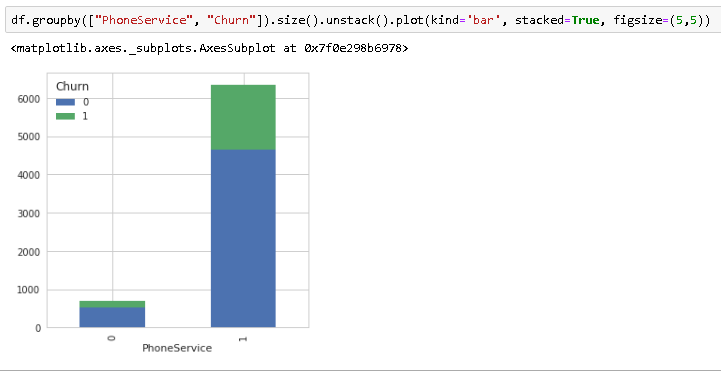
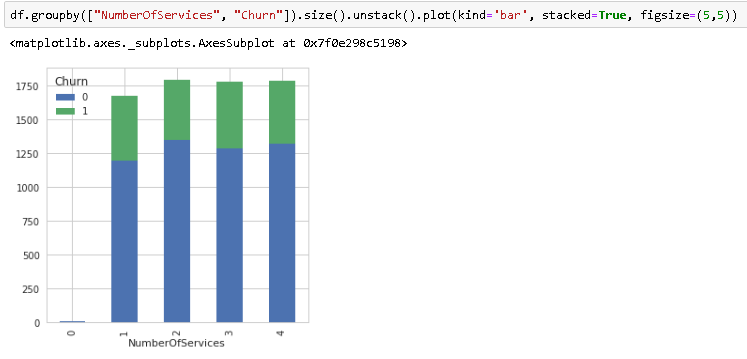
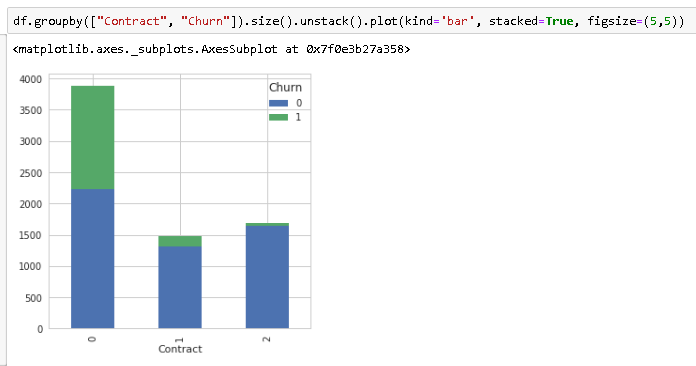
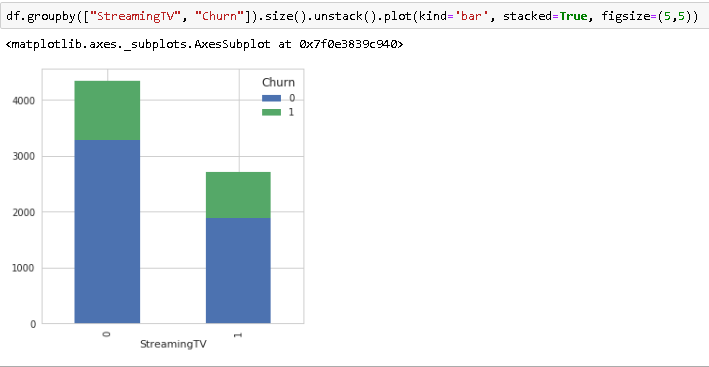


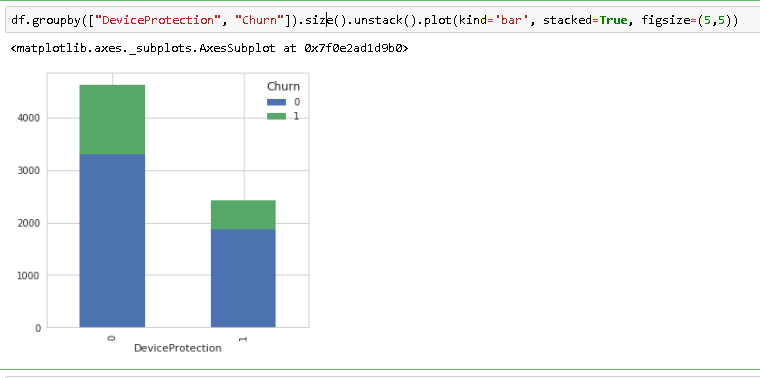
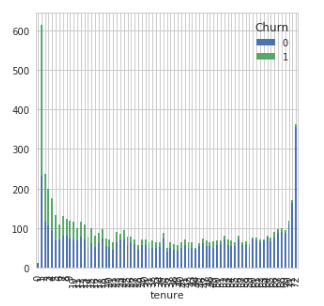
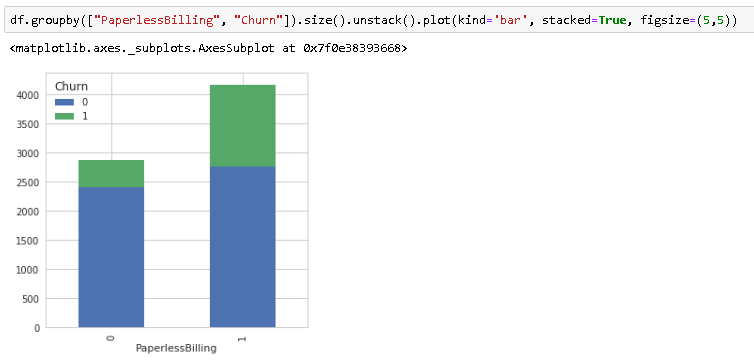
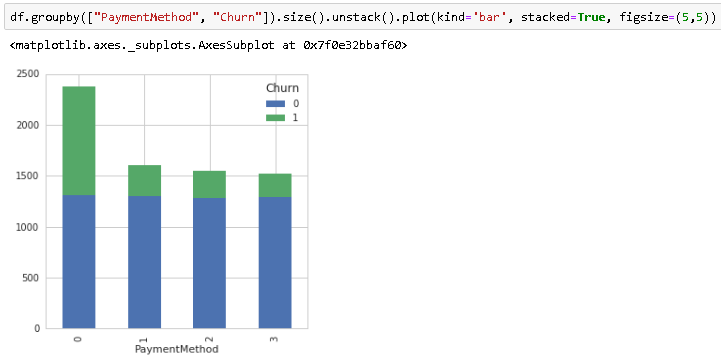
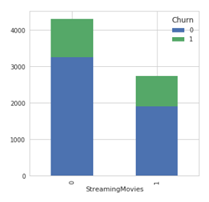
**Post removal of outliers present in both the columns:**



**Histogram plots**

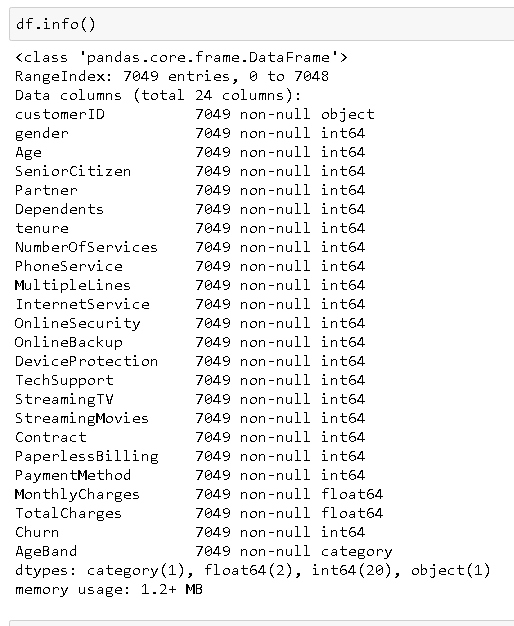
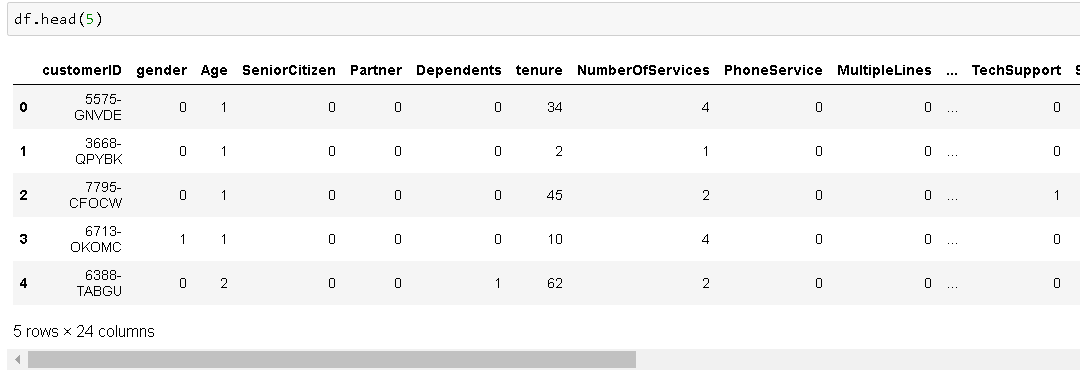




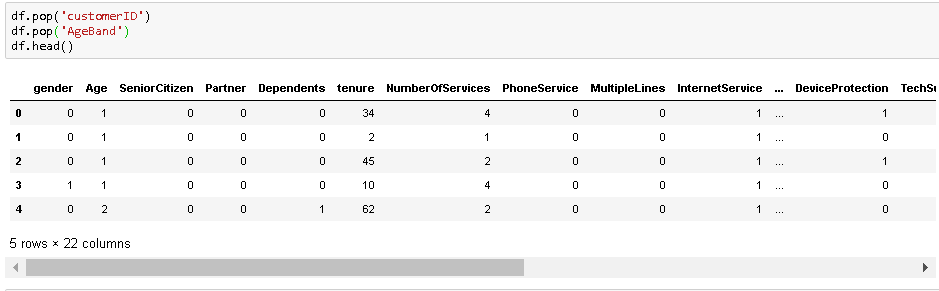


**Feature Engineering:**

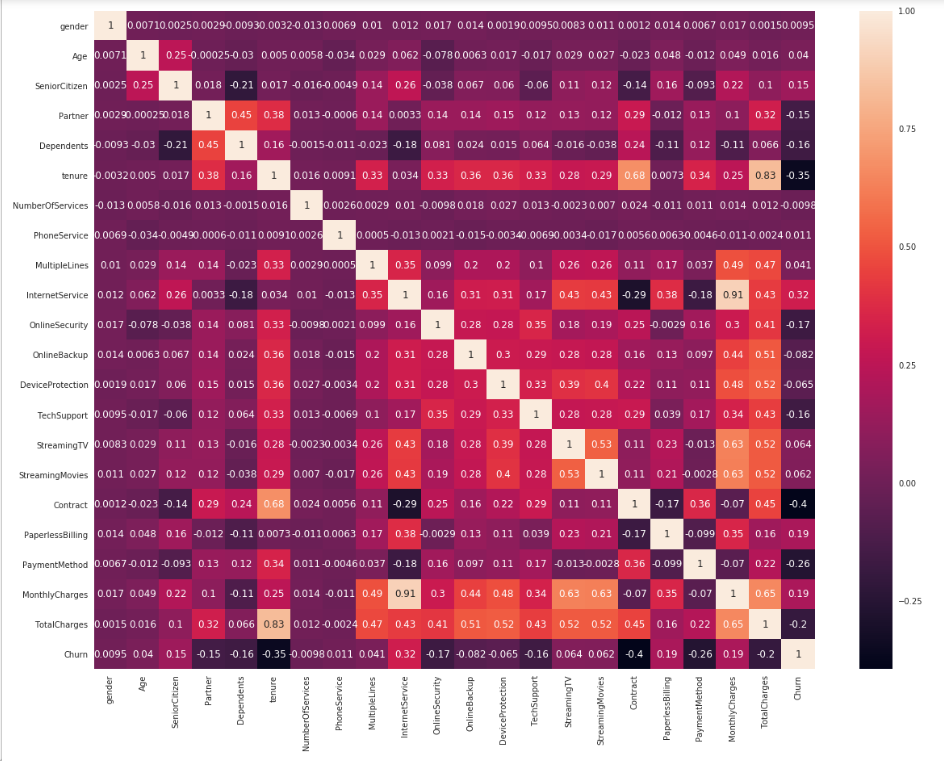




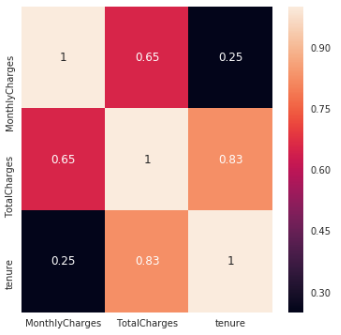
**FEATURE SELECTION USING CORRELATION ANALYSIS:**

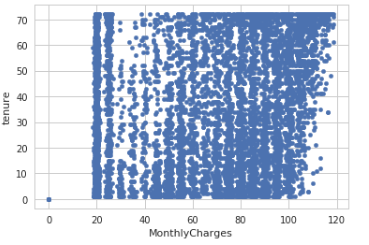
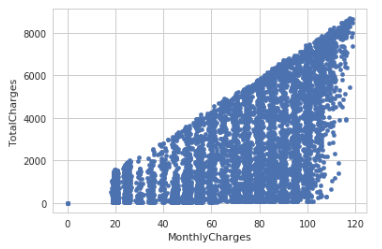
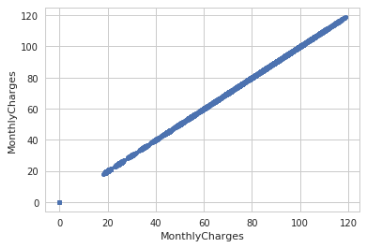


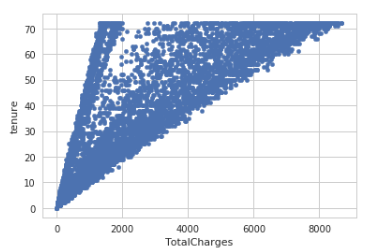
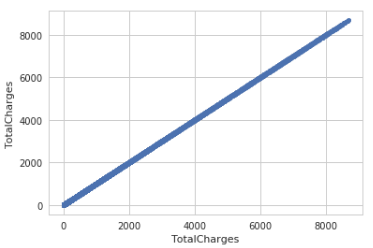
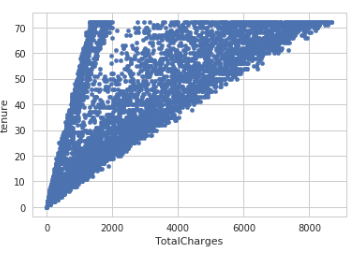


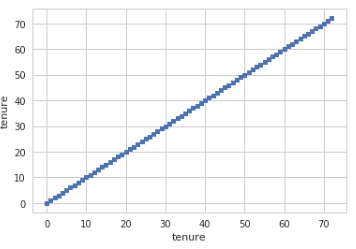
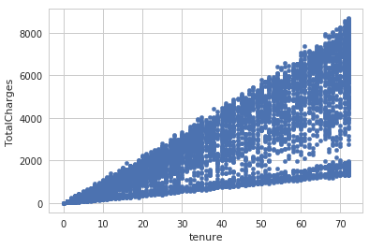
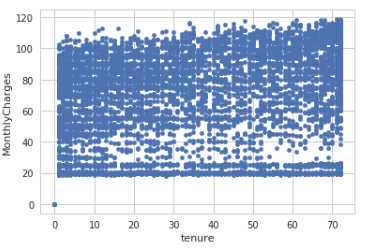


Below 3 attributes from the correlation matrix and try to visualize their relationship using scatterplot.

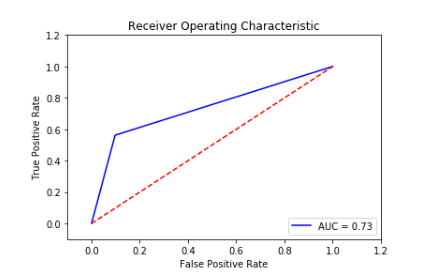
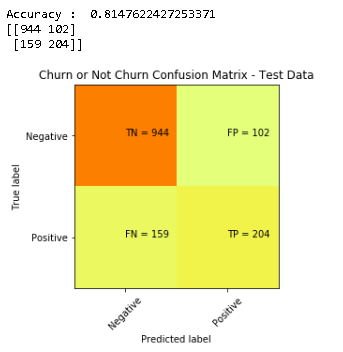


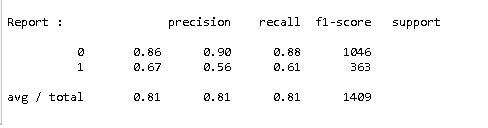




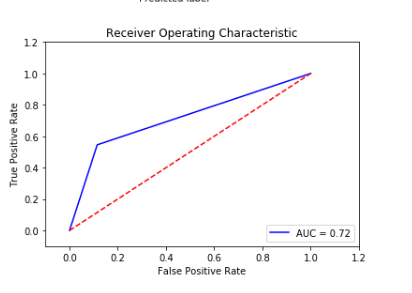
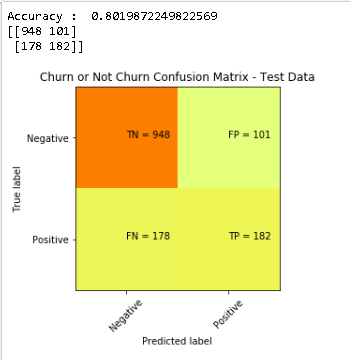


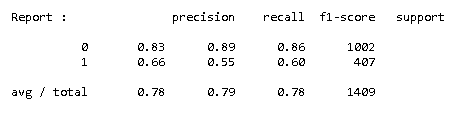
**Logistic Regression:**



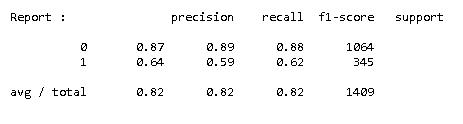
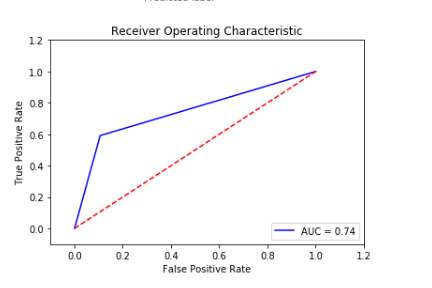
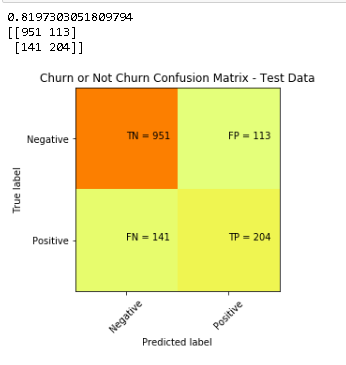


**Random Forest:**



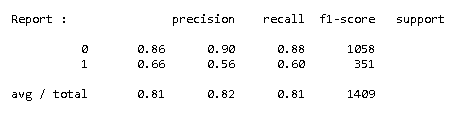
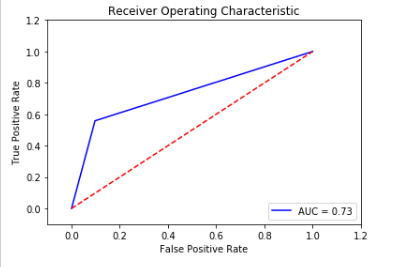
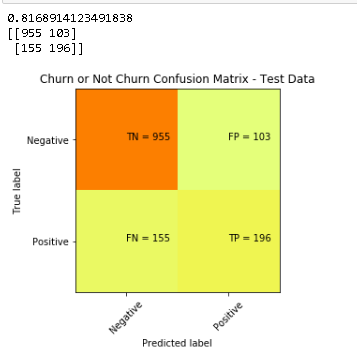


**SVM - Linear** C-1



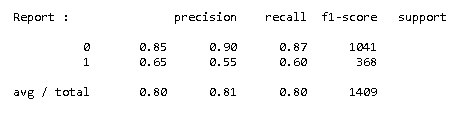
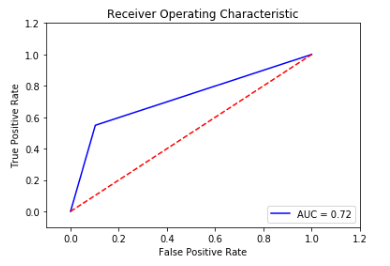
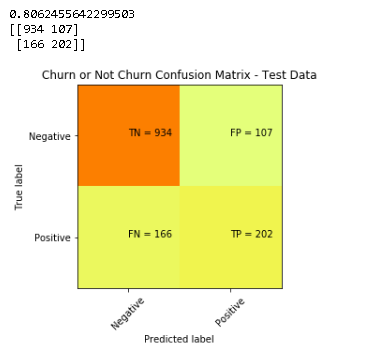
C=5

Before sampling

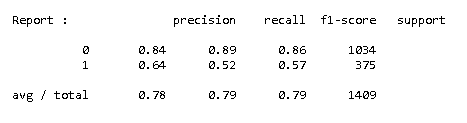
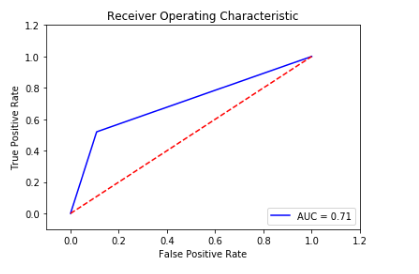
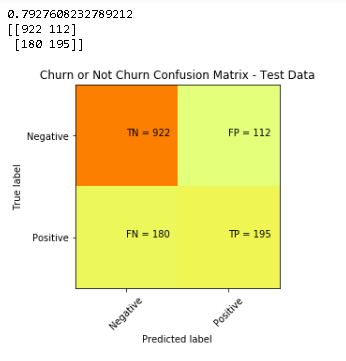


c=10

Before sampling



c=100



|  |  |  |
| --- | --- | --- |
| **SVM Linear** |  |  |
| **Tuning Parameter** | **Accuracy** | **AUC-ROC score** |
| C=1 | 81.97% | 74% |
| C=5 | 81.68% | 73% |
| C=10 | 80.62% | 72% |
| C=100 | 79.92% | 71% |

|  |  |
| --- | --- |
| **SVM RBF** |  |
| **Tuning Parameters** | **Resulted Labels** |
| C=1, Gamma =0.1 | All Negatives |
| C=1, Gamma =0.01 | All Negatives |
| C=1, Gamma =0.001 | All Negatives |
| C=1, Gamma =0.0001 | All Negatives |
| C=5, Gamma =0.1 | All Negatives |
| C=5, Gamma =0.01 | All Negatives |
| C=5, Gamma =0.001 | All Negatives |
| C=5, Gamma =0.0001 | All Negatives |
| C=10, Gamma =0.1 | All Negatives |
| C=10, Gamma =0.01 | All Negatives |
| C=10, Gamma =0.001 | All Negatives |
| C=10, Gamma =0.0001 | All Negatives |

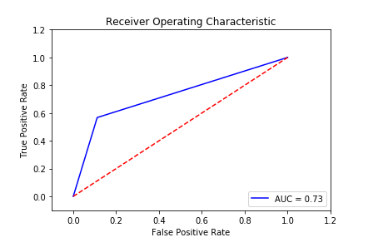
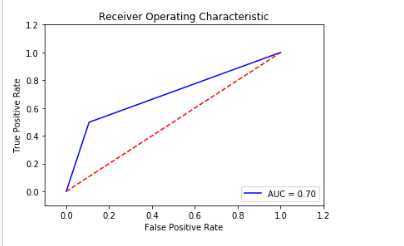
**Cross Validation** k=5

**Logistic Regression**



**Decision Tree**



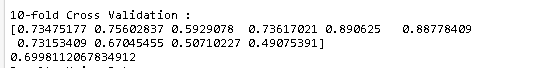
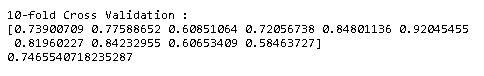


**Random Forest SVM**



**K=10**

**Logistic Regression Decision Tree**



**Random Forest SVM**

