SURVEY SPARROW ASSIGNMENT

CHURN PREDICTION ENGINE

- LEISHA SAANVI. M (RA2111026040039)
- lm5157@srmist.edu.in

1. Initially all the libraries such as numpy, matplotlib, pandas, seaborn,etc., were imported and the dataset was loaded.
2. Categorical and Numerical Columns were found.
3. Datatypes of the columns were checked (object – 18 col, int – 2 col, float -1 col).
4. Null values of the rows were checked.
5. Basic Data cleaning was performed. For example, 'Totalcharges' is given as object datatype but it is float datatype
6. All the categorical and Numerical Columns were explored separately.

Categorical Columns findings:

Gender - In terms of numbers **67** male customers are more and in percentage **0.5%** male customers are more.

Senior Citizens - Most of the customers are not Senior Citizens.

Partners - Number of customers without partners are 3641, 51.7% of the customers are not having partners.

Dependents - 70.0% (4933) of customers are not having dependents; meanwhile, 30.0% (2110) of customers are having dependents.

Phone Services - 90.3% (6361) of the customers are having phone services; meanwhile, 9.7% (682) of the customers are not having phone services.

Multiple Lines - 48.1% (3390) of the customers are not having Multiplelines; 42.2% (2971) of the customers are having Multiplelines; 9.7% (682) of the customers are not having any phone service.

Internet Service - 44% (3096) of the customers are having Fiber optic InternetService; 34.4% (2421) of the customers are having DSL InternetService; 21.7% (1526) of the customers are not having any InternetService service.

Online Security - 49.7% (3498) of the customers are not having OnlineSecurity; 28.7% (2019) of the customers are having OnlineSecurity; 21.7% (1526) of the customers are not having any internet service.

Online Backup - 43.8% (3498) of the customers are not having OnlineBackup; 34.5% (2019) of the customers are having OnlineSecurity; 21.7% (1526) of the customers are not having any internet service.

Device Protection - 43.9% (3095) of the customers are not having DeviceProtection; 34.4% (2422) of the customers are having OnlineSecurity; 21.7% (1526) of the customers are not having any internet service.

TechSupport - 49.3% (3473) of the customers are not having TechSupport; 29.0% (2044) of the customers are having TechSupport; 21.7% (1526) of the customers are not having any internet service.

Streaming TV - 39.9% (2810) of the customers are not having StreamingTV; 38.4% (2707) of the customers are having StreamingTV; 21.7% (1526) of the customers are not having any internet service.

Streaming Movies - 39.5% (2785) of the customers are not having StreamingMovies; 38.8% (2732) of the customers are having StreamingMovies; 21.7% (1526) of the customers are not having any internet service.

Contract - 55.0% (3875) of the customers are having Month-to-Month Contract; 24.1% (1695) of the customers are having Two year Contract; 20.9% (1473) of the customers are having One year Contract.

Paperless Billing - 59.2% (4171) of the customers are having paperlessBilling; 40.8% (2872) of the customers are not having paperlessBilling.

Payment Method - Most of the customers are having Electronic check PaymentMethod.

Churn - 73.5% (5174) of the customers are not invloving in Churn ; 26.5% (1869) of the customers are invloving in Churn.

Numerical Columns findings:

- Most of the customers are staying only in between 0-5 months in the company
- Most of the customers having Monthly charges in between 20-25

Many other findings were made. Some example findings include:

- Only 1% of female churned morethan male; 930 male churned while 939 female churned overall.
- 42% of the old citizens are churned.
- 33% of people without partners churned.
- Customers without dependent churned more.
- Customers with multiple lines have a slightly higher churn rate
- Clients without internet have a very low churn rate
- Customers with fiber are more probable to churn than those with DSL connection
- Customers with the first 4 additionals (security to tech support) are more unlikely to churn
- Customers with paperless billing are more probable to churn. The preferred payment method is Electronic check with around 35% of customers. This method also has a very high churn rate.

7. **Outliers Analysis** was performed on 'tenure' and 'MonthlyCharges' column.
8. **Data Cleaning** was performed. Dropped customerId because it is not needed. Dropped Churn because we had already encoded column Churn_rate.
9. **On - Hot Encoding** - convert categorical data into a numerical format.
10. **Feature Scaling** and **Feature Selection** were performed.
11. **Two models** were used for prediction **– Logistic Regression** with an accuracy of 0.80**,** and **Decision Trees** with an accuracy of 0.72 was achieved**.** Other metrics, such as precision, recall and f1 – score were also calculated.