

Problem Statement:

Study and analyse the data of your company and find our insights and stop your company customers from churning out of other telecom companies.

Customer Churn Data set:

The data set looks like this.

A	B	C	D	E	F	G	H	I	J	K	L	M
customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport
7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No	No
5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No
3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No
7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes	Yes
9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No

K	L	M	N	O	P	Q	R	S	T	U
OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
Yes	No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
No	Yes	No	No	No	One year	No	Mailed check	56.95	1889.5	No
Yes	No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
No	Yes	Yes	No	No	One year	No	Bank transfer (automatic)	42.3	1840.75	No
No	No	No	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes

Task to be done:

A) Data Manipulation:

- Extract the 5th and 15th column and store in a variable.
- Extract all male senior citizens whose payment method is Electronic check and store in a variable.
- Extract all those customers whose tenure is greater than 70 months or their monthly charges are more than 100\$ and store in a variable.
- Extract all those customers whose contract is of 2 years, payment method is Mailed check and the value of churn is 'Yes' and store in a variable.
- Extract 333 random records from the input data frame and store in other.
- Get the counts of different levels from the churn column.

B) Data Visualization:

- Build a bar-plot for the 'InternetService' column.
 - Set x-axis label to 'Categories of Internet Services'.
 - Set y-axis label to 'Count of Categories'.
 - Set the title of plot to be 'Distribution of Internet Service'.
 - Set the color of the bar to 'Orange'.
- Build a histogram for the 'tenure' column.
 - Set the number of bins to be 30.
 - Set the color of bins to be 'green'.
 - Assign the title as 'Distribution of tenure'

- c. Build a scatter plot between 'MonthlyCharges' and 'tenure'. Map 'MonthlyCharges' to the y-axis and 'tenure' to the x-axis.
 - Assign the points a color of 'brown'
 - Set the x-axis label to 'Tenure of customer'
 - Set the y-axis label to 'Monthly charges of customer'
 - Set the title as 'Tenure vs Monthly charges '
- d. Build a box plot between 'tenure' and 'contract'. Map 'tenure' on the y-axis and contract on the x-axis

C) Linear Regression:

- a. Build a simple linear model where dependent variable is 'MonthlyCharges' and independent variable is 'tenure'.
 - Divide the dataset into train and test set in the ratio 70:30.
 - Build the model on the train set and test on the test set.
 - After predicting the values, find the root mean square error.
 - Find out error in prediction and store the result in 'error'.
 - Find the root mean square error.

D) Logistic Regression:

- a. Build a logistic regression model where dependent variable is 'Churn' and independent variable is 'Monthly charges'.
 - Divide the dataset into train and test set in the ratio 65:35.
 - Build the model on the train set and test on the test set.
 - Build confusion matrix and get the accuracy score.
- b. Build a logistic regression model where dependent variable is 'Churn' and independent variable is 'Monthly charges' and 'tenure'
 - Divide the dataset into train and test set in the ratio 80:20.
 - Build the model on the train set and test on the test set.
 - Build confusion matrix and get the accuracy score.

E) Decision Tree:

- a. Build a decision tree model where independent variable is 'churn' and dependent variable is 'tenure'
 - Divide the test set into 80:20 ratio.
 - Build the model on the train set and test on the test set.
 - Build confusion matrix and get the accuracy score.

