

Capstone Project 1: Data Wrangling

I looked at the shape and information stored in the data. I took a look at the data types and null values. On the whole this data was very clean and required minimal data cleaning steps. I took a look at the value counts of all the attributes and if there were any duplicates or anomalies but found none.

Looking at the data I saw that it contained customers which had 0 tenure i.e. they had just joined. I took the decision of removing such data because the total charges value was null and it is obviously not possible for them to churn when they just joined. I made a new dataset out of this filtered data for further investigation.

I have attached screenshots of some data cleaning steps.

```
[10]: #checks all the null values
      print(data.isnull().sum())
```

```
customerID      0
gender          0
SeniorCitizen   0
Partner         0
Dependents      0
tenure          0
PhoneService    0
MultipleLines   0
InternetService 0
OnlineSecurity  0
OnlineBackup    0
DeviceProtection 0
TechSupport     0
StreamingTV     0
StreamingMovies 0
Contract        0
PaperlessBilling 0
PaymentMethod   0
MonthlyCharges  0
TotalCharges    0
Churn           0
dtypes: int64
```

```
# This is data where the customer has just joined and hasnt con
#These rows should not be included in the larger dataset
cust_new=data[data['TotalCharges']!=' ']
data[data['TotalCharges']==' ']
```