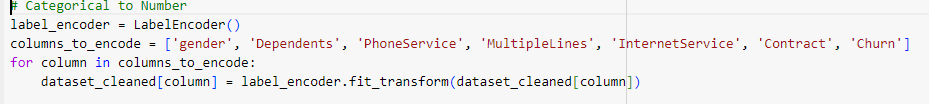
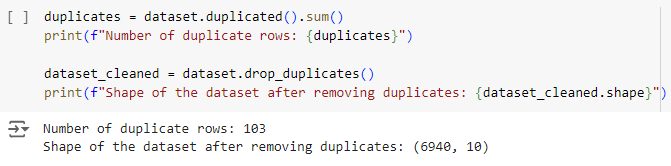
Data processing

Description (Aasish  and Manisha)

To prepare our dataset for effective clustering, we first focused on thorough data cleaning and preprocessing. This included handling missing values, encoding categorical variables, scaling numerical features, and ensuring that all data was in a consistent format suitable for analysis. The goal was to create a dataset that would be well-suited for the K-means clustering algorithm, ensuring that each feature was appropriately treated to maintain the integrity of the clustering process.

Splitting data typically involves dividing it into a training set and a test set, often using an 80/20 or 70/30 ratio. The training set is used to build the model, while the test set evaluates its performance.



We preprocessed our dataset by cleaning and scaling the data, and importantly, we used a **Label Encoder** to convert categorical variables into a numerical format. This helped condense the large dataset and made it ready for clustering.

