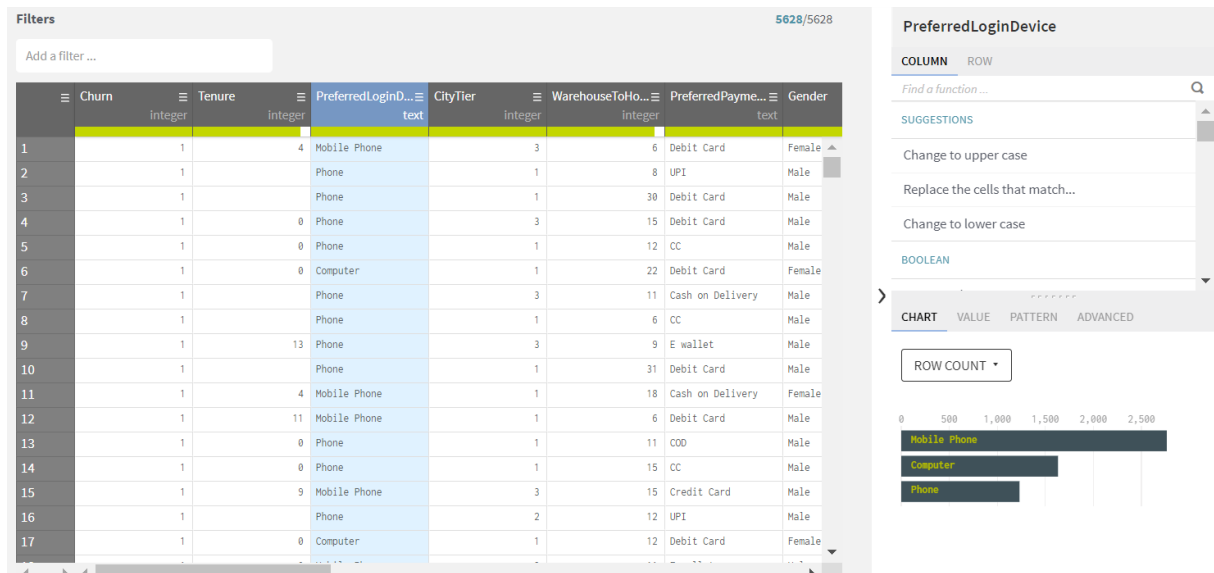


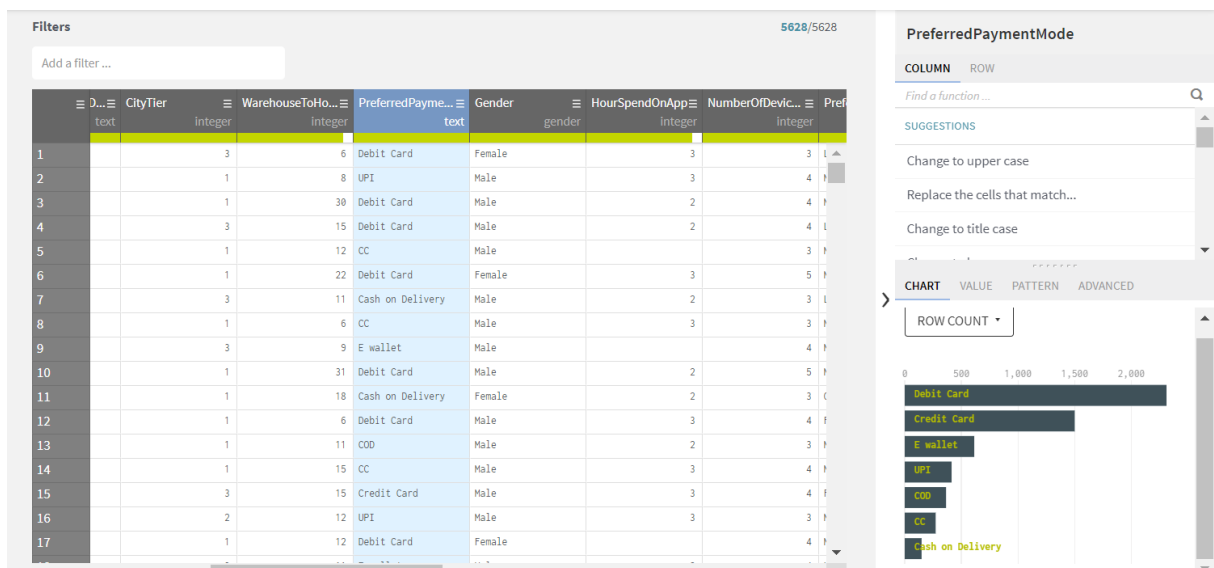
## 2.0 Data Pre-processing using Talend Data Preparation

### 2.1 Handling Inconsistent Data

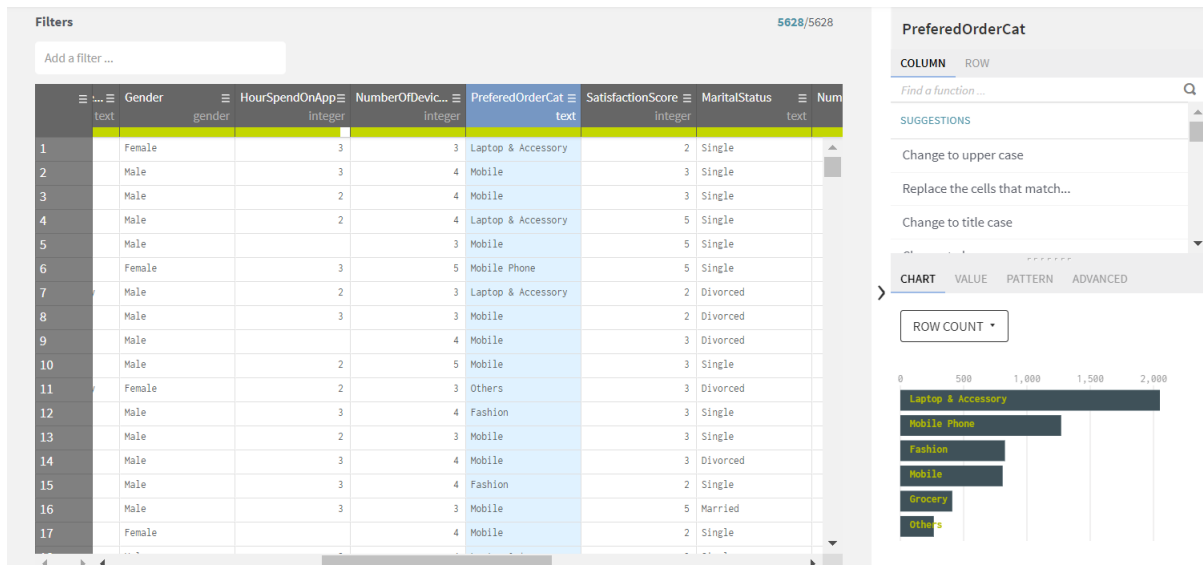
In the “PreferredLoginDevice” column, “Phone” and “Mobile Phone” referred to the same type of device. To ensure data consistency, “Phone” was replaced with “Mobile Phone”.



In the “PreferredPaymentMode” column, “Credit Card” and “CC” referred to the same type of payment mode whereas “Cash on Delivery” and “COD” referred to the same type of payment mode. To ensure data consistency, “CC” was replaced with “Credit Card” whereas “COD” was replaced with “Cash on Delivery”. In addition, “UPI” was replaced with “Unified Payments Interface”.



In the “PreferredOrderCat” column, “Mobile Phone” should be a subset of “Mobile”. To ensure data consistency, “Mobile Phone” was replaced with “Mobile”.



## 2.2 Handling Outliers

In the “WarehouseToHome” column, there were 2 records with values 126 and 127. These two values were far from majority records as shown in the data distribution chart. To handle outliers, 126 and 127 were adjusted to 26 and 27 respectively to align them within the appropriate range for this column.

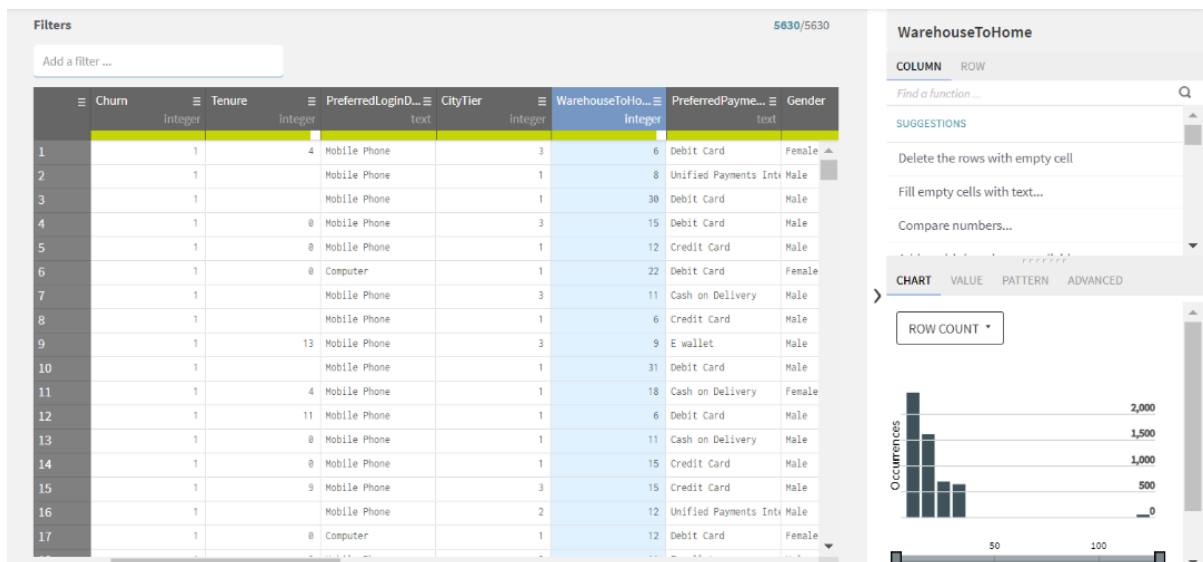


Figure below illustrates all the data pre-processing steps performed in Talend Data Preparation.

