* **ORIGINAL DATASET** : ilpd.csv
  + Total observations : 583
  + Total Null values : 4
  + Total Duplicate Values : 13
  + # Column Non-Null Count Dtype
  + --- ------ -------------- -----
  + 0 age 583 non-null int64
  + 1 3 gender 583 non-null object
  + 2 t\_bilirubin 583 non-null float64
  + d\_bilirubin 583 non-null float64
  + 4 alkphos 583 non-null int64
  + 5 sgpt 583 non-null int64
  + 6 sgot 583 non-null int64
  + 7 t\_protein 583 non-null float64
  + 8 albumin 583 non-null float64
  + 9 a\_g\_ratio 579 non-null float64
  + 10 result 583 non-null int64
* COLAB NOTEBOOK 1 : liver\_cd\_data\_preparation\_1.ipynb

Handled duplicate values *(Dropped)*

Handled null values *( Simple Imputation - Median value )*

Handled categorical values : *(OHE on gender)*

Modified dataset : **ilpd\_md1.csv**

* COLAB NOTEBOOK 2 : liver\_cd\_data\_preparation\_2.ipynb

Handled Outliers *(z score method for normally distributed data & IQR method for Skewed data)*

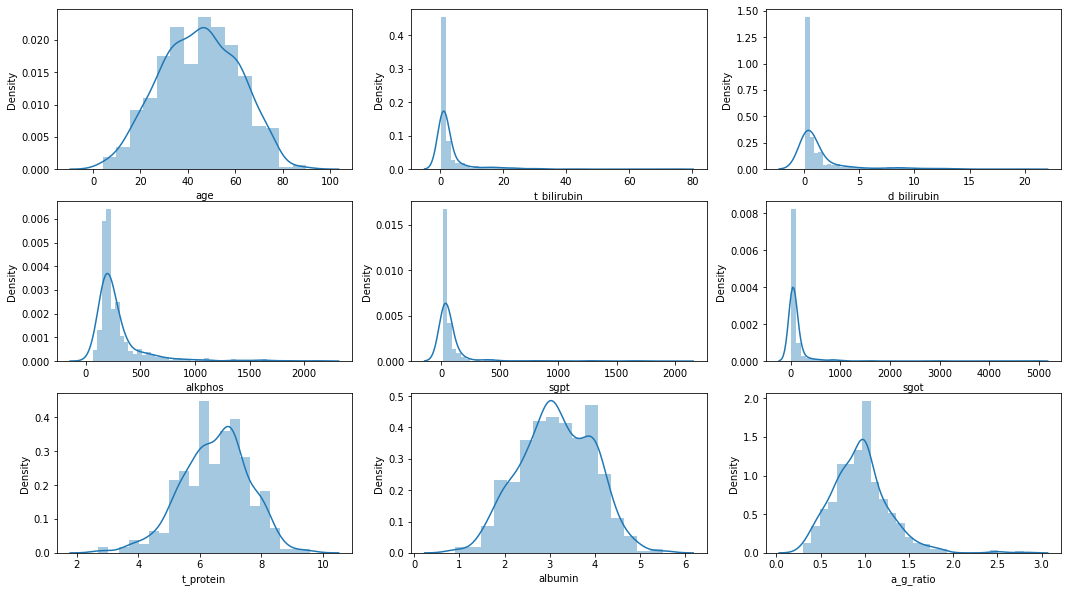
Handled Data Distribution *(Normalisation-standard Scaling)*

Modified dataset : **ilpd\_md2.csv (Handled Outliers)**

**ilpd\_md3.csv (Handled Outliers + Scaling )**

**Report attached below (fig 1,2,3)**

**Before Handling Outliers**

****

# Distplots Summary and Observation

**Normal distributed data :**

**age, t\_protein, albumin, a\_g\_ratio**

**Skewed Distributed data :**

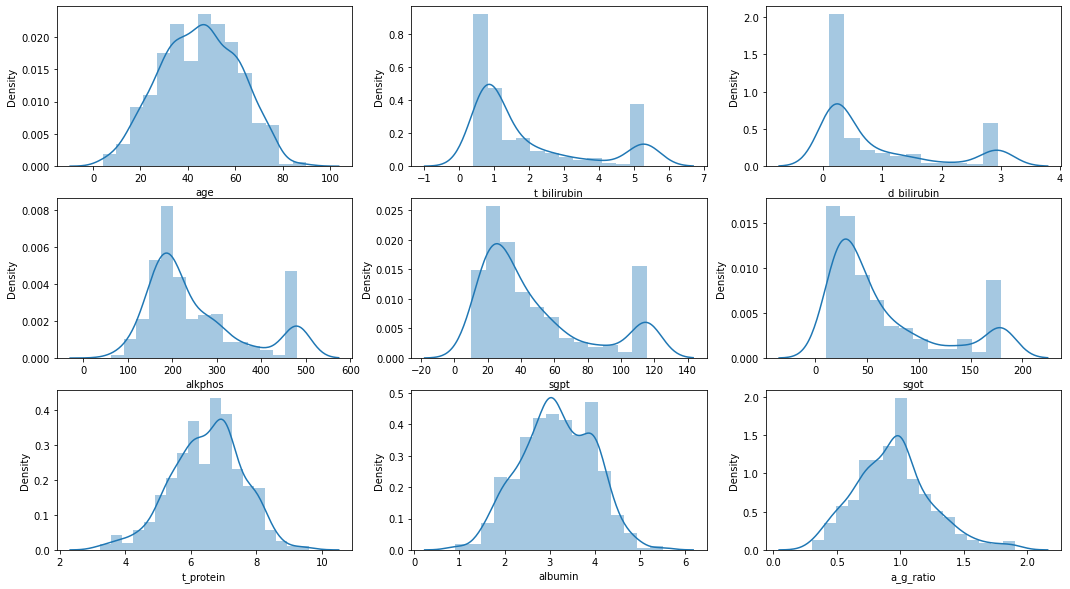
**t\_bilirubin, d\_bilirubin, alkphos, sgpt, sgot**

**Conclusion : For outlier detection**

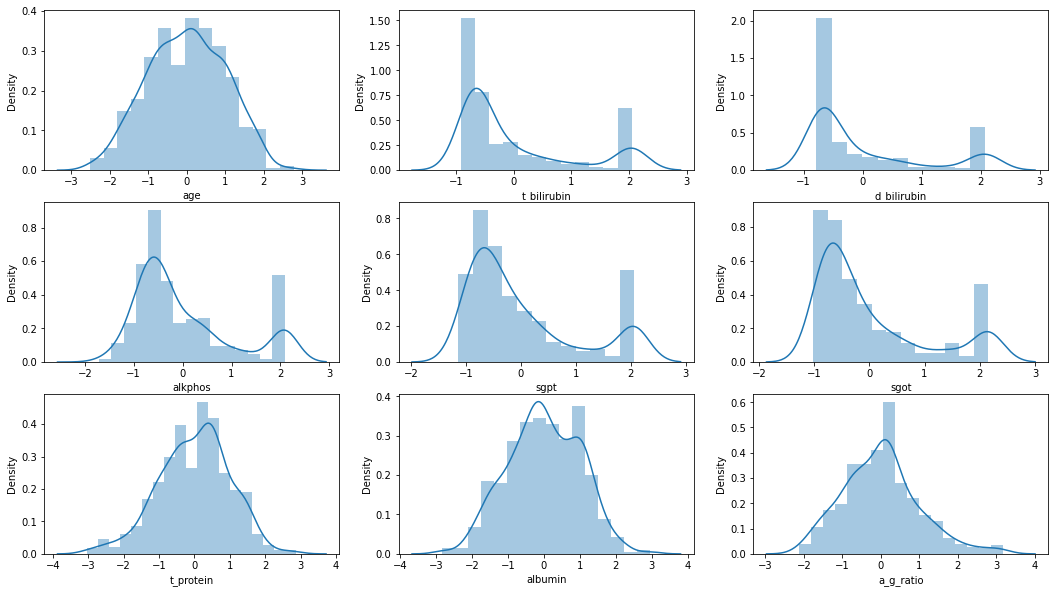
**Z-Score method is used for Normal Distributed data**

**IQR Method is used for Skewed Distributed data**

**After Handling Outliers**

****

**After Scaling Data (Normalisation)**

****