Class exercise: Jupyter notebook plus data

These are the names of the columns/variables in the dataset:

BUSAGE: Age of business measured (in days)

BUSTYPE: Business type (Type of Business classified from A-F)

MAXLINEUTIL: Maximum line of credit utilized by the customer

DAYSDELQ: Days delinquency is the number of days a customer has been delinquent or performs illegal or immoral acts

TOTACBAL: Total account balance (total account balance for the customer)

DEFAULT: default (Yes or no if the client defaulted a loan)

Yes: defaulted the load (not paid the loan amount)

No: not defaulted (paid the loan)

KDE Plot

A KDE plot, or Kernel Density Estimate plot, is a way to visualize the distribution of a continuous variable. It smooths out the data points into a continuous probability density function. This is especially useful when you want to understand the underlying distribution of data, such as where the data points are most concentrated.

**Understanding KDE Plot**

* **Kernel**: The function used to estimate the probability density function (PDF) of the variable.
* **Bandwidth**: A parameter that controls the smoothness of the KDE plot. A smaller bandwidth makes the plot more sensitive to small variations, while a larger bandwidth smooths it out.

**KDE Plot for Business Age and Total Account Balance**

We have a credit dataset which includes the Business Age (in days) and the Total Account Balance (in Rupees) for different businesses (A-F). A KDE plot can help you visualize how these two variables are distributed and if there's any pattern or relationship between them.

1. **For a Single Variable**:
   * You can use a KDE plot to show the distribution of Business Age or Total Account Balance individually.
   * It will help you understand the central tendency, spread, and skewness of the data.
2. **For Two Variables**:
   * A 2D KDE plot can be used to visualize the relationship between Business Age and Total Account Balance.
   * It will show areas where the combination of business age and account balance is most common.

**Insights from KDE Plots**

* **For Business Age**: The KDE plot can help you to see if most businesses are younger or older.
* **For Total Account Balance**: The KDE plot can show if there's a concentration of businesses with specific account balances.
* **For Business Age vs. Account Balance**: The 2D KDE plot can reveal patterns, such as whether older businesses tend to have higher or lower account balances.

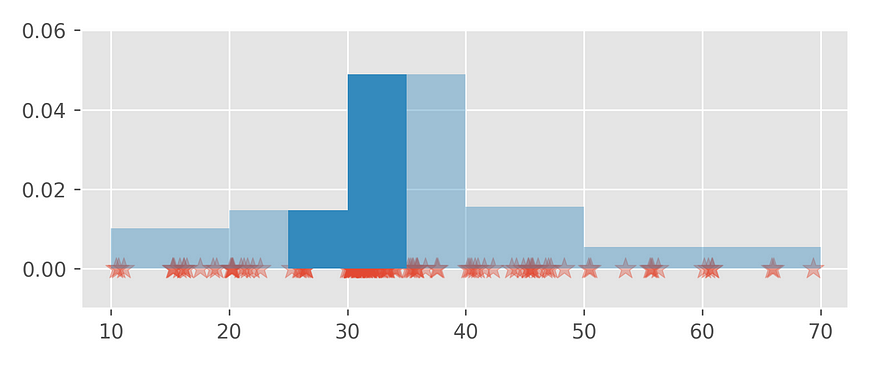


Fig (a)

It follows that the function *f* is also a probability density function (the area under its graph equals one). The function *f* is the **Kernel Density Estimator** (KDE). Graph (fig b ) looks like a smoothed version of the histogram plots (Fig a) constructed earlier.

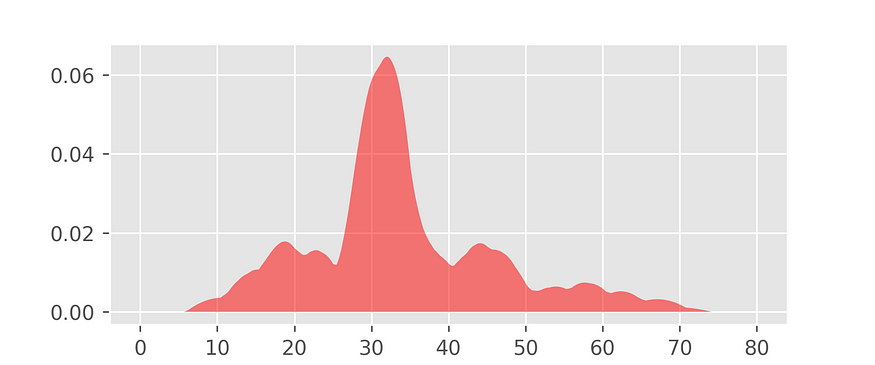


Fig (b)

A graph of a business age

Description automatically generated