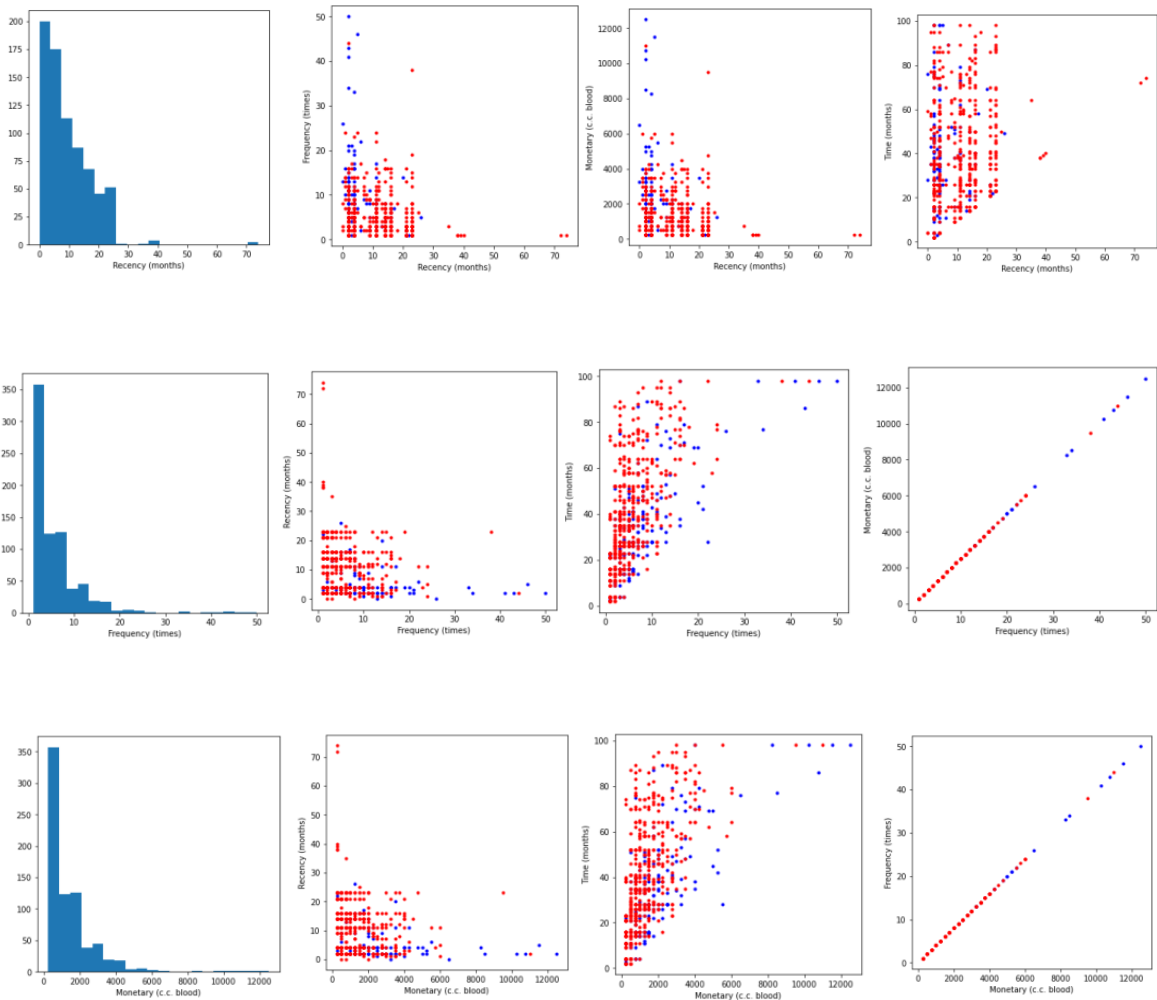


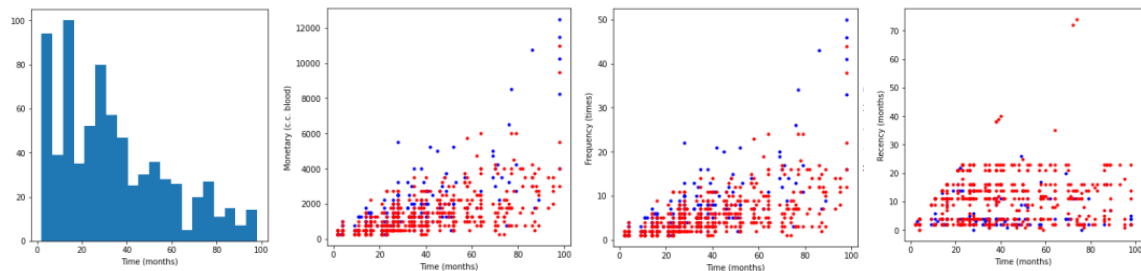
ساختار داده

دیتاست حاوی Attribute های زیر می باشد.

- Recency: تعداد ماه گذشته از آخرین اهدا خون
- Frequency: تعداد کل دفعاتی که فرد خون اهدا کرده
- Monetary: میزان کل خونی که فرد اهدا کرده (در واحد سی سی)
- Time: تعداد ماه گذشته از اولین اهدا خون
- Class: لیبل مربوط به هر سمپل، نشان دهنده اهدا کردن یا نکردن خون

بررسی با scatter plot ها:

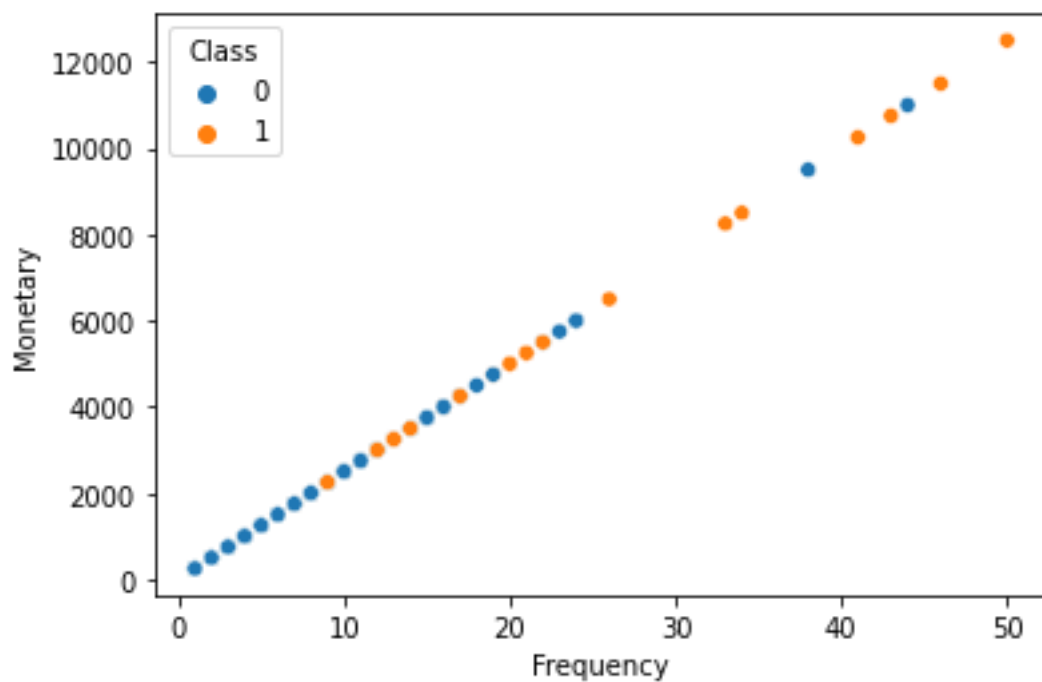




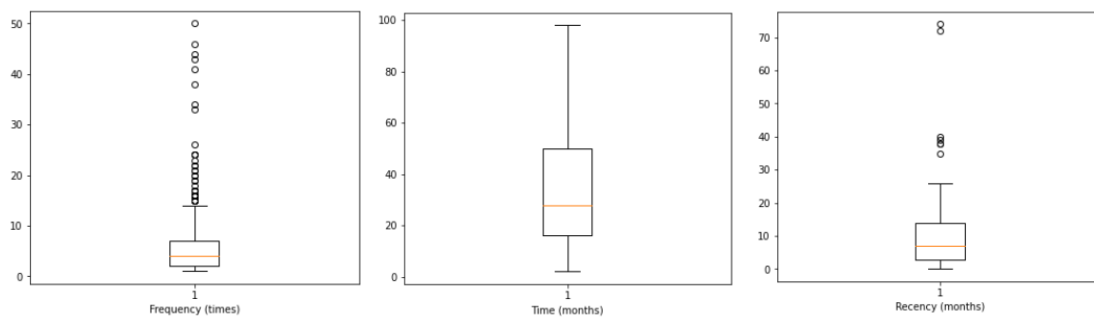
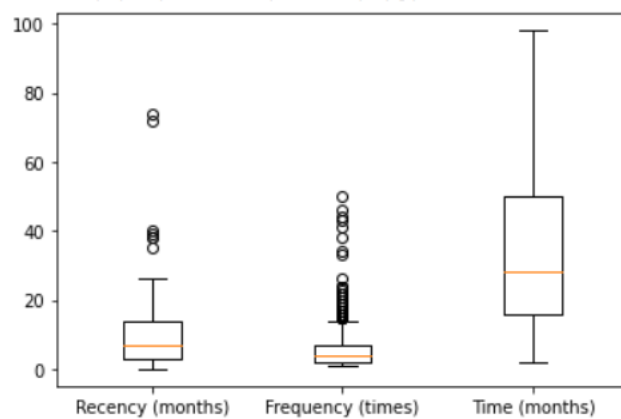
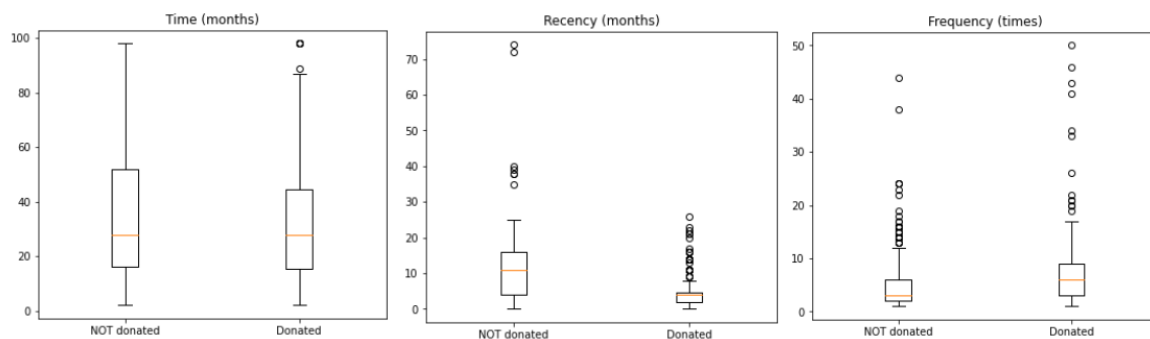
با توجه به نمودارها، می‌بینیم که Monetary و Frequency رابطه خطی دارند پس در هر نمودار که دارای این دو مورد باشند، شبیه به هم میشوند. می‌توان نتیجه گرفت که موارد داده شده به خوبی جدا نشده اند اما Frequency و Time میتوانند بهتر باشند.

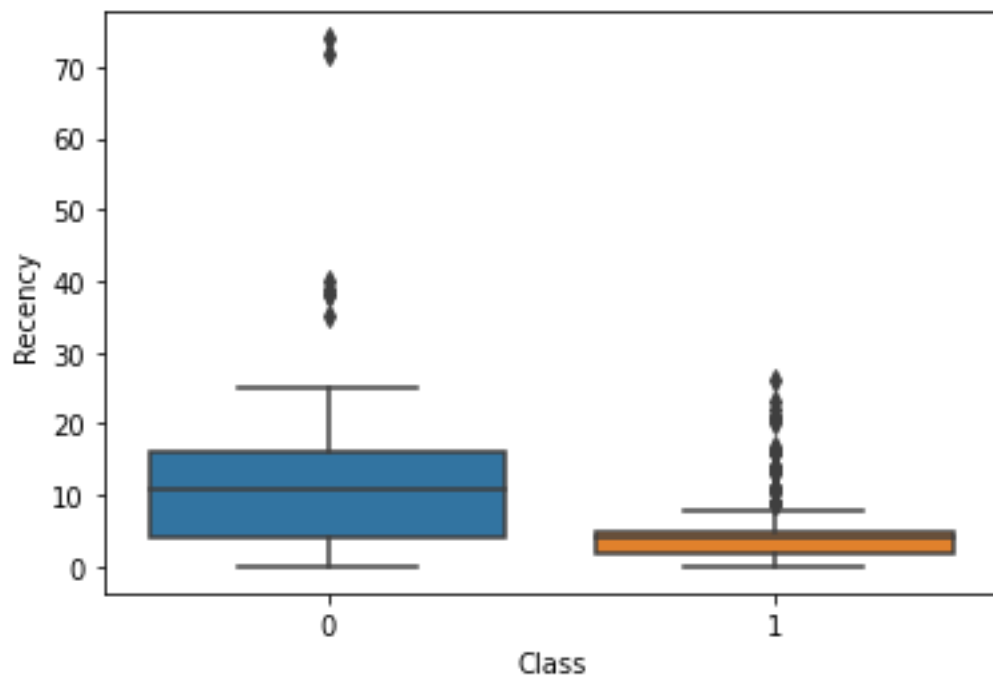
بررسی رفتار Monetary و Frequency

همینطور با توجه به ضریب همبستگی 1 بین این دو صفت، انتظار یک رابطه خطی و مثبت بین آن‌ها را داشتیم. نمودار زیر، به طور خاص رابطه این دو مورد را نشان میدهد.

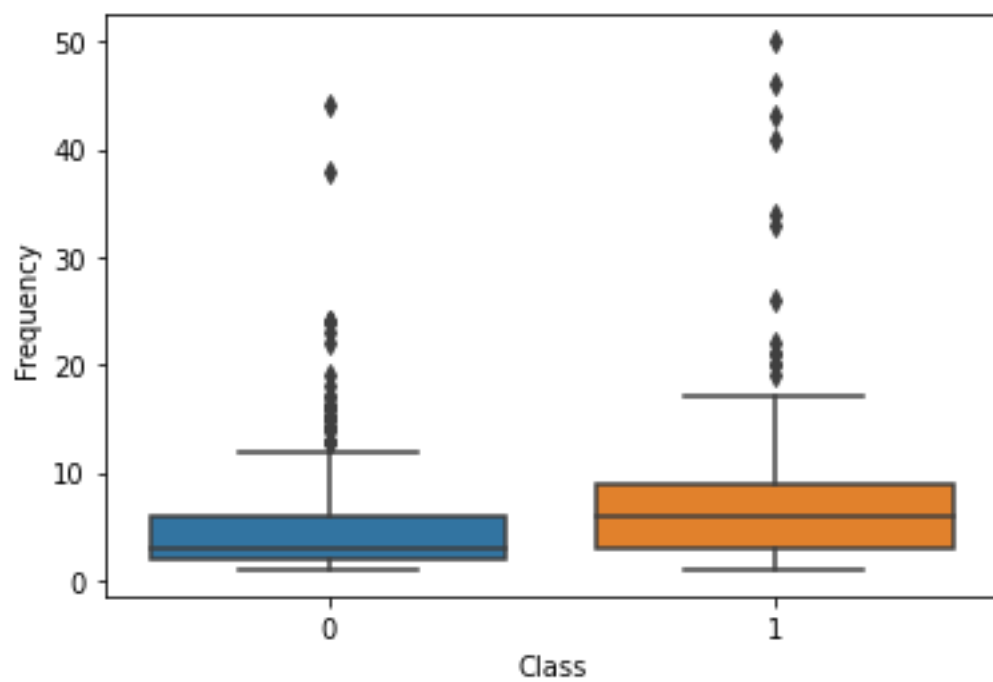


بررسی توزیع Attribute ها با استفاده از نمودار جعبه ای

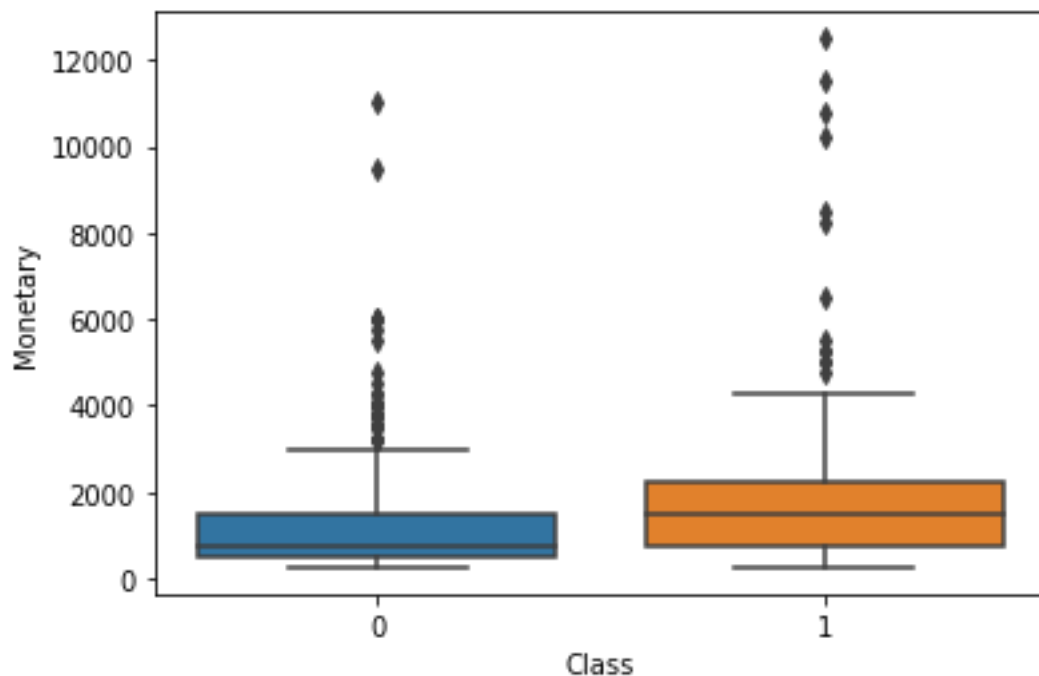




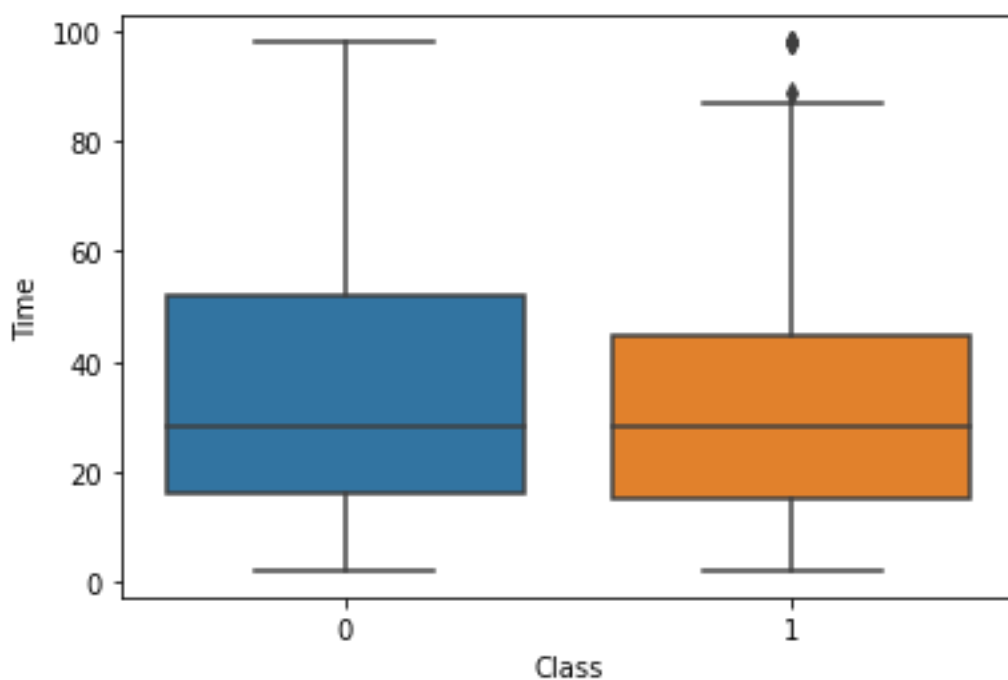
تفاوت بسیار مشهودی در مقادیر Recency بین سَمپل های دو Class وجود دارد. این Attribute برای طبقه بندی داده مناسب خواهد بود.



در این Attribute هم توزیع های متفاوتی بین دو Class وجود دارد؛ اما diversity مقادیر به شدت فیلد Recency نیست.



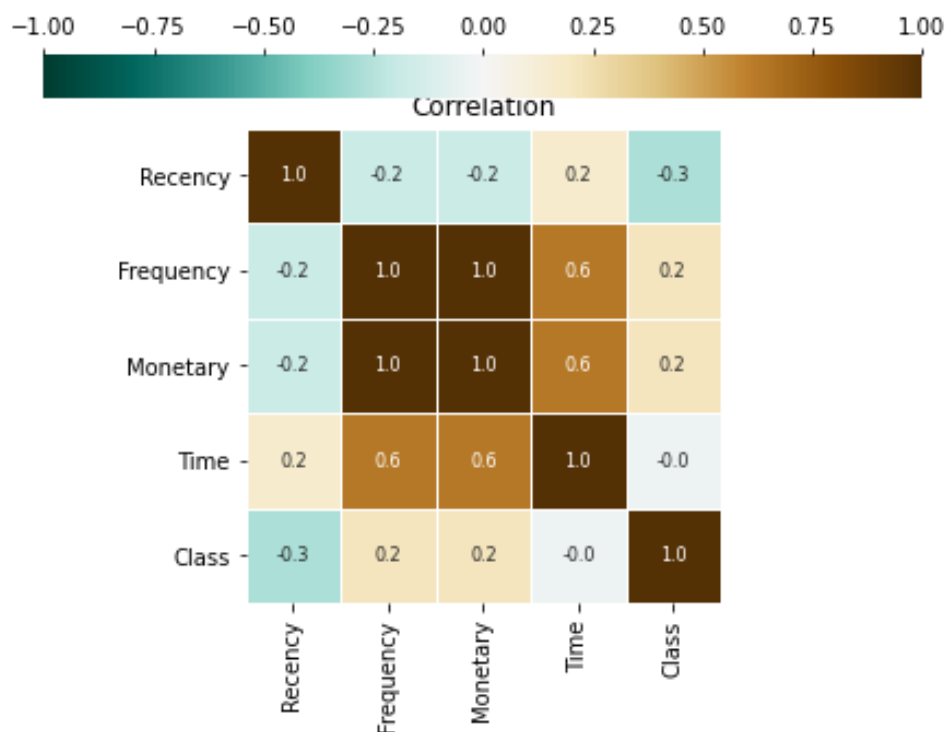
توزیع Monetary، به نظر کاملاً مشابه Frequency می آید، اما با Scale ای متفاوت. در ادامه با بررسی Correlation Matrix ارتباط این دو Attribute را دقیق تر بررسی خواهیم کرد.



صفت Time گزینه مناسبی برای طبقه بندی داده نخواهد بود.

پس در این نمودارها دیدیم که Recency کمترین overlap بین Donated و Not donated را دارد پس بهترین مورد برای استفاده در مدلسازی است.

ماتریس همبستگی



از بررسی ماتریس همبستگی، سه مورد بسیار مهم قابل دریافت است؛

- صفت Recency در مقایسه با بقیه Attribute ها، همبستگی بیشتری با Class دارد؛ در نتیجه نقش قوی در طبقه بندی داده خواهد داشت.
- دو صفت Monetary و Frequency، کاملاً همبسته ان؛ مقدار 1 در ماتریس برای این دو صفت قابل مشاهده است. این همبستگی به معنای رفتار کاملاً یکسان این دو Attribute است. برای طبقه بندی داده میتوان فقط از یکی از این دو صفت استفاده کرد.
- صفت Time هیچ همبستگی ای با Class ندارد. این Attribute برای طبقه بنده بی فایده خواهد بود.