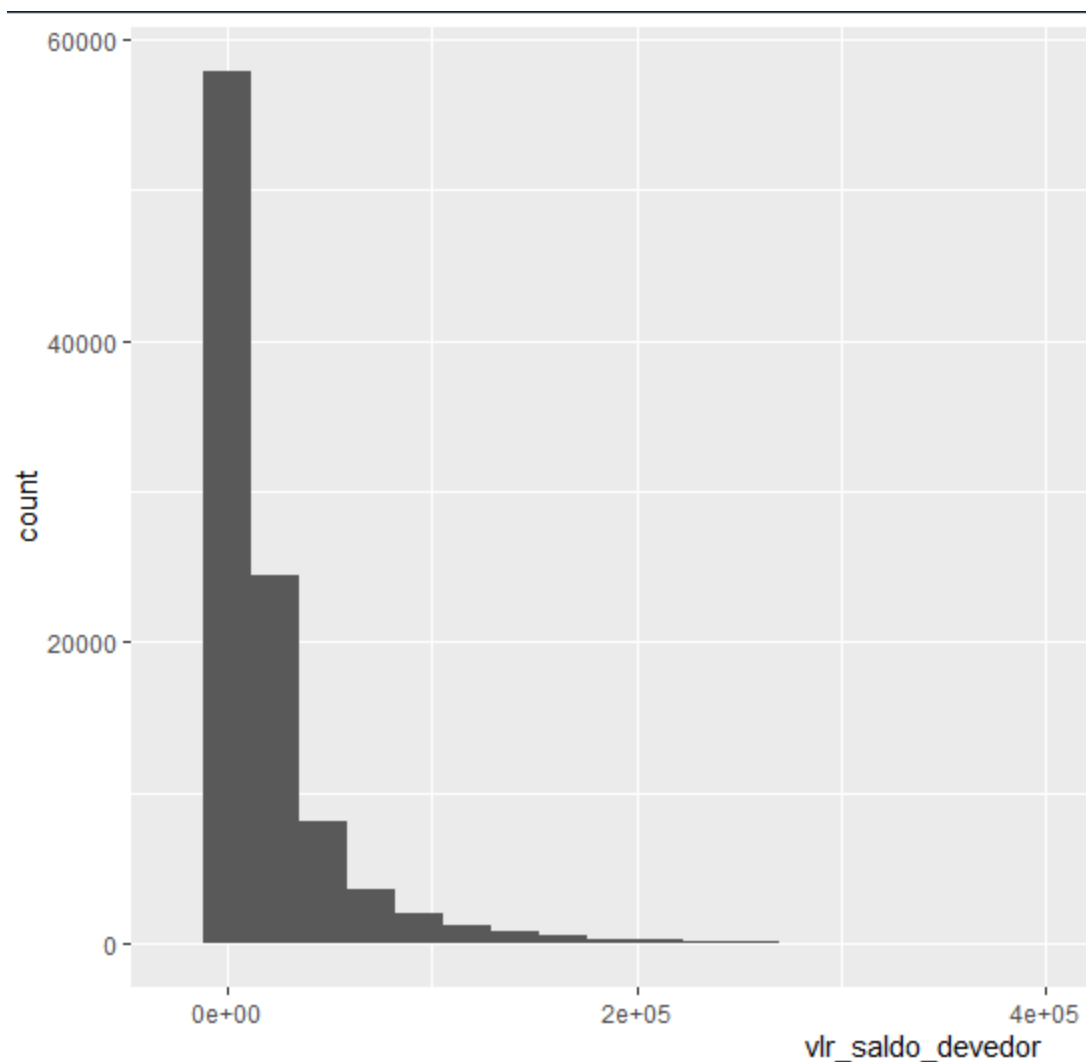


4) Insights (Análise exploratória)

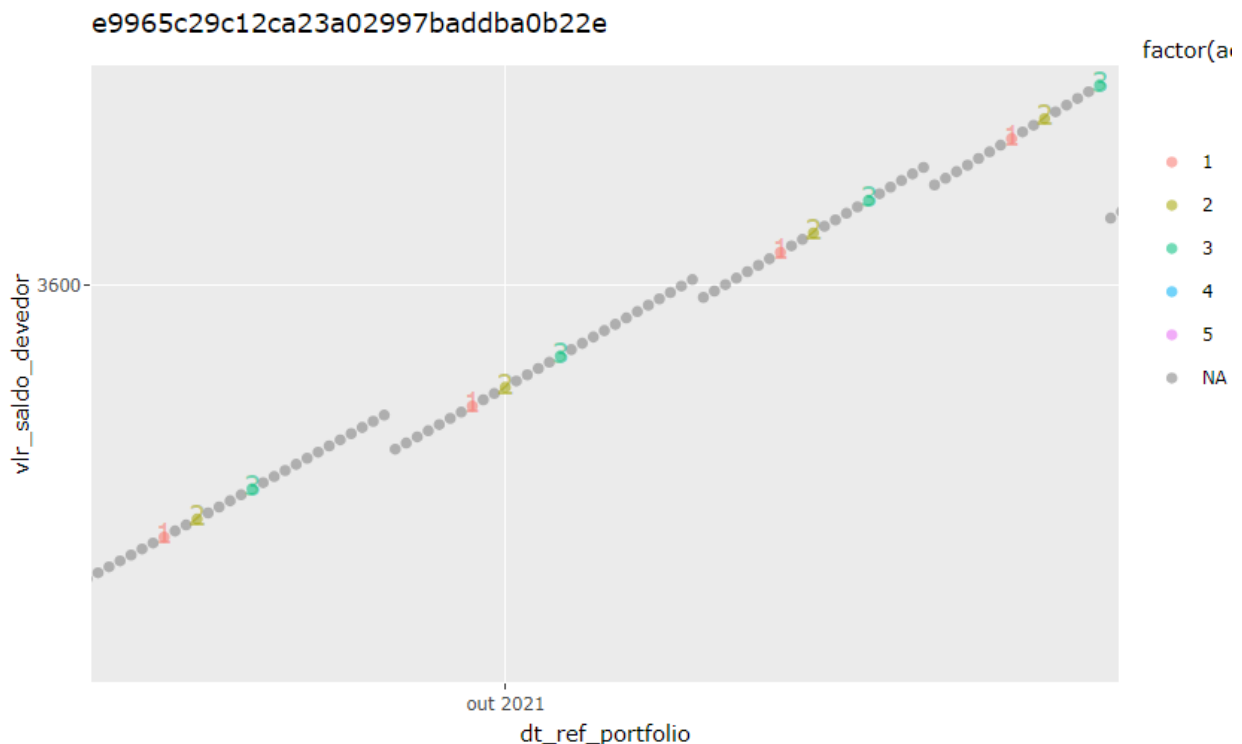
Para as variáveis com valores financeiros, notamos uma distribuição com assimetria à direita, motivo pelo qual teremos de aplicar posteriormente uma transformação logarítmica nos dados. Segue abaixo o exemplo para a variável *vlr_saldo_devedor*.



Para entender o comportamento das dívidas dos clientes, analisamos diversos gráficos de clientes distintos. Notamos que há um padrão nessa variável e ela opera por regimes. É fundamental entender como operam esses regimes (duração, dia esperado

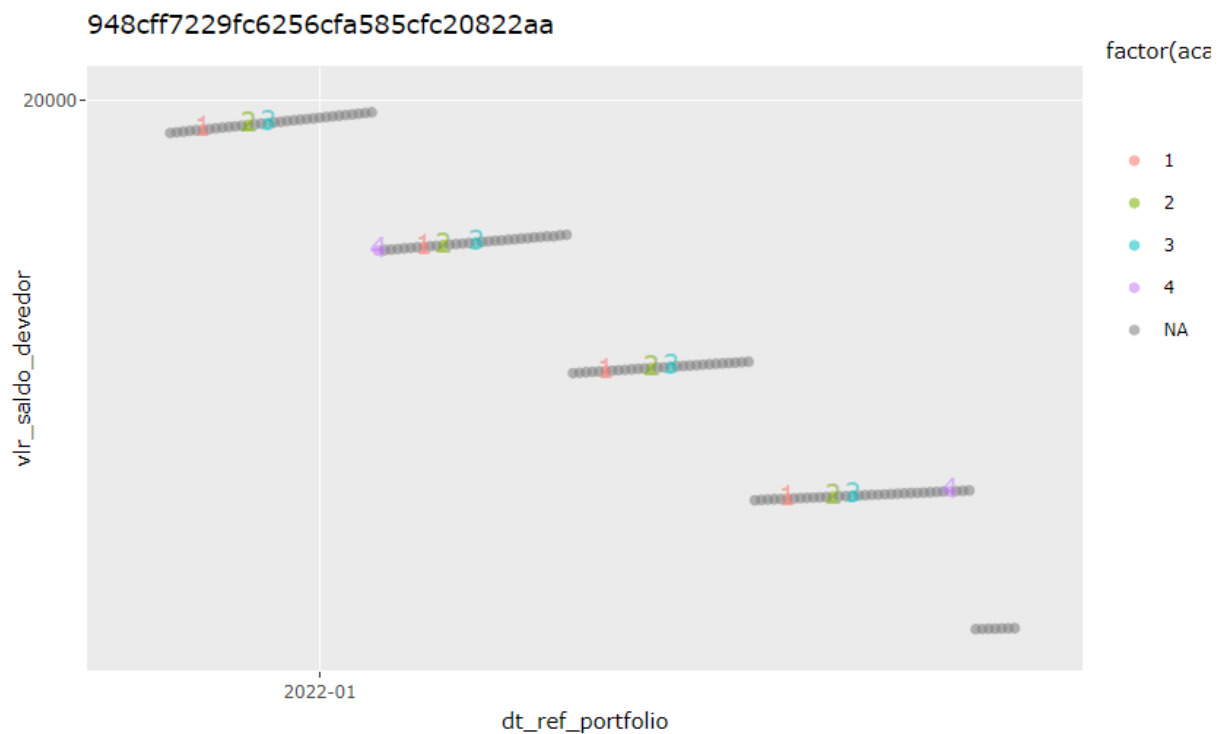
de reversão) e como eles transitam entre si (dado os regimes predecessores, qual será o regime esperado. Ou seja, qual é a probabilidade de que o cliente tenha um regime de diminuição ou de aumento nas dívidas). Abaixo temos alguns exemplos ilustrados por clientes:

4.1) Cliente aleatório 1 - padrão vlr_saldo_devedor



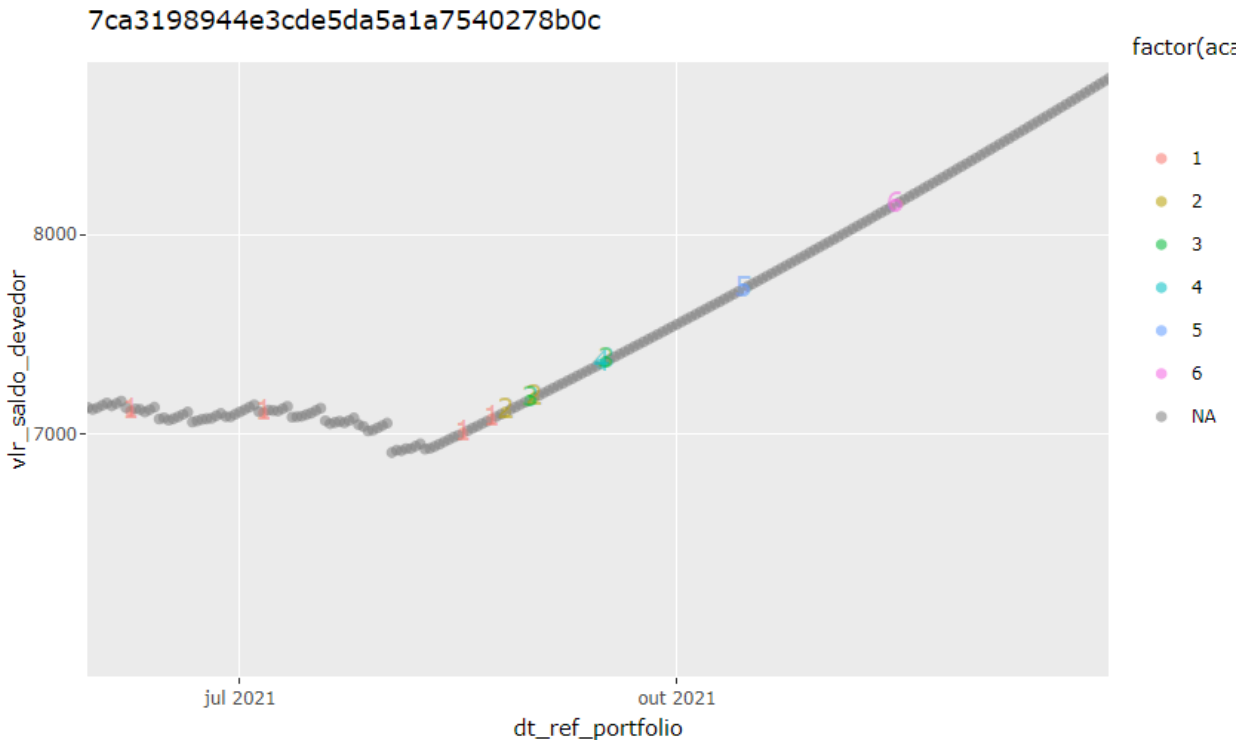
Pelo gráfico acima, notamos que esse cliente apresenta um aumento no fluxo de dívidas. Provavelmente os pagamentos não sejam suficientes para cobrir a dívida e tenhamos que reduzir um pouco os juros aplicado ao cliente. A reversão ocorre em espaços regularmente previsíveis e que podem ser determinados. No timeframe analisado, a reversão ocorre sempre após a campanha de quitação (número 3 em verde).

4.2) Cliente aleatório 2 - vlr_saldo_devedor



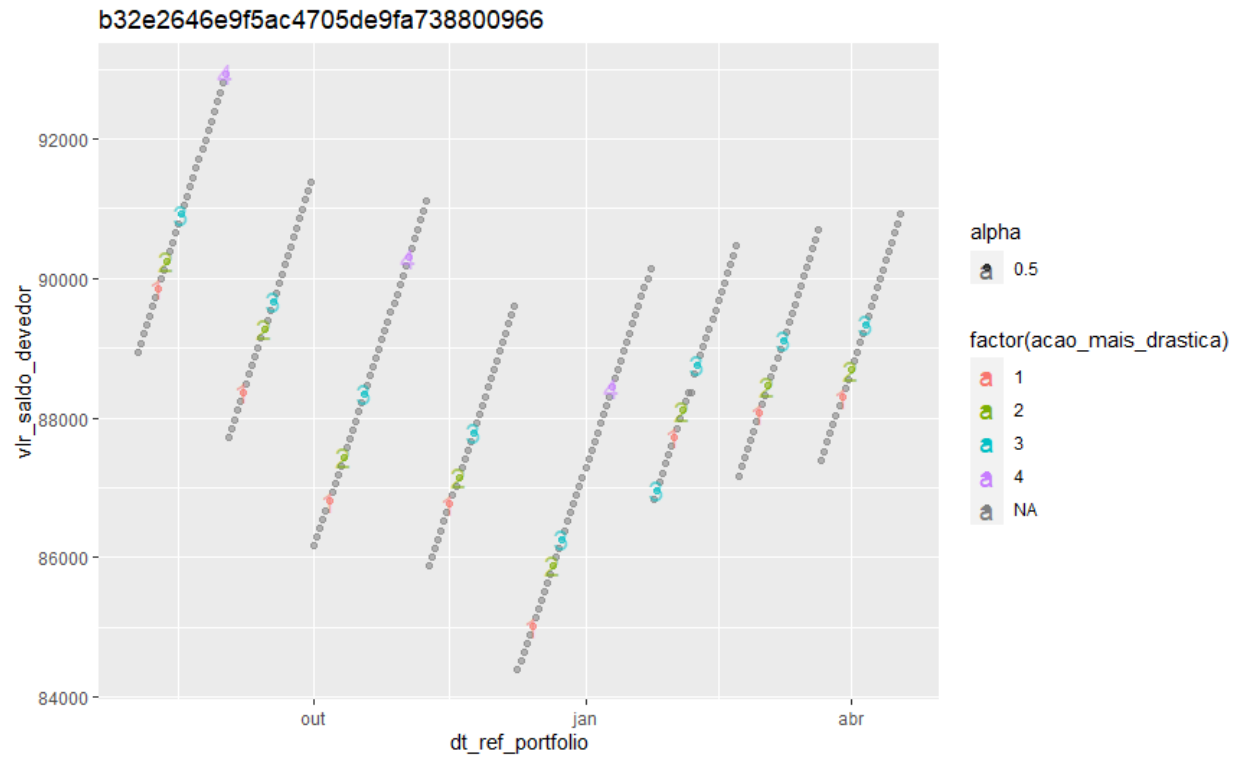
Esse cliente do gráfico, no período dados, se mostrou bastante reativo à negativação (número 4 em rosa).

4.3) Cliente aleatório 3 - vlr_saldo_devedor



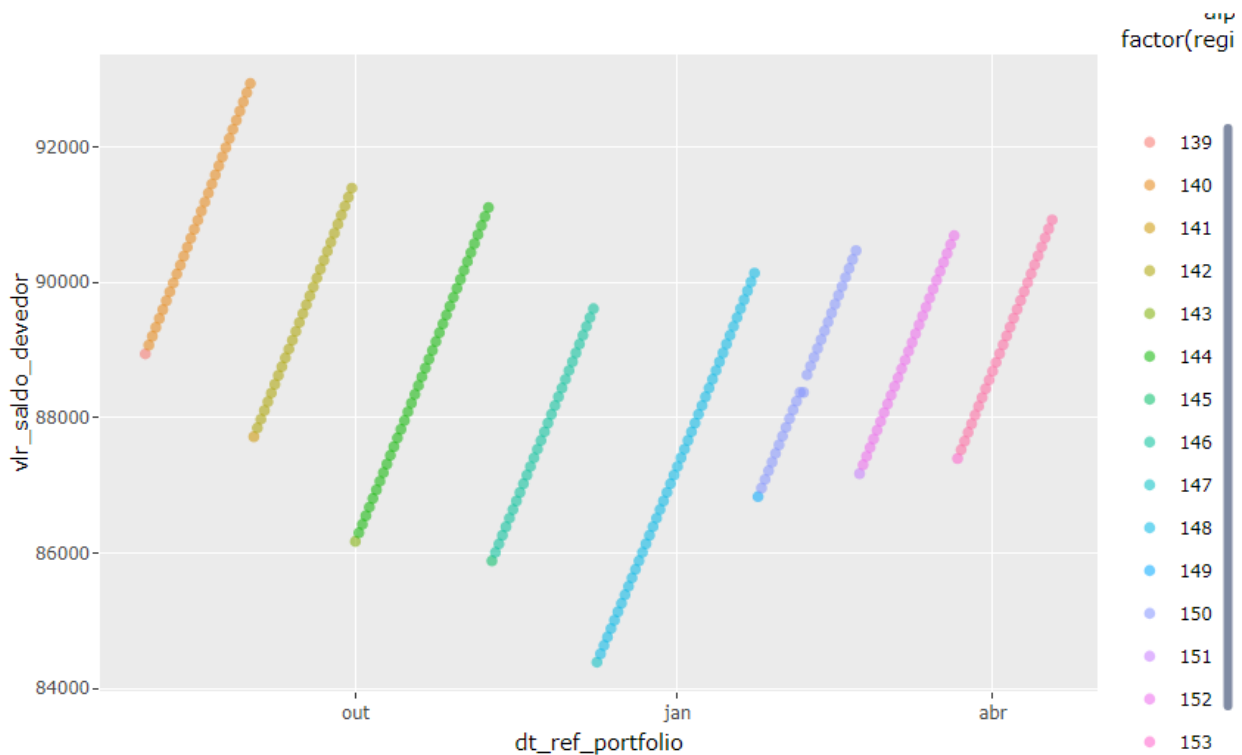
Cliente apresentou uma distorção no padrão de pagamento (regime mais longo que os anteriores) . Tivemos questionamentos repetidos por conta das duas réguas, o que poderia ser unificado. Nesse caso, poderíamos ter uma ação mais cedo com o cliente para entender o motivo do aumento da dívida e tomar alguma ação com antecedência à régua para ele liquidar sua dívida de forma sustentável.

4.3) Cliente aleatório 4 - Análise exploratória mais detalhada



Vamos delimitar os regimes de pagamento sempre que houver algum pagamento, ou seja, $dsp = 0$. Isso pode ser realizado simplesmente aplicando o script abaixo:

```
target[,var_created_pagamento_efetuado:=ifelse(dsp==0,1,0)]
target[,regime_ativo:=rleid(var_created_pagamento_efetuado), by = contrato_id]
```



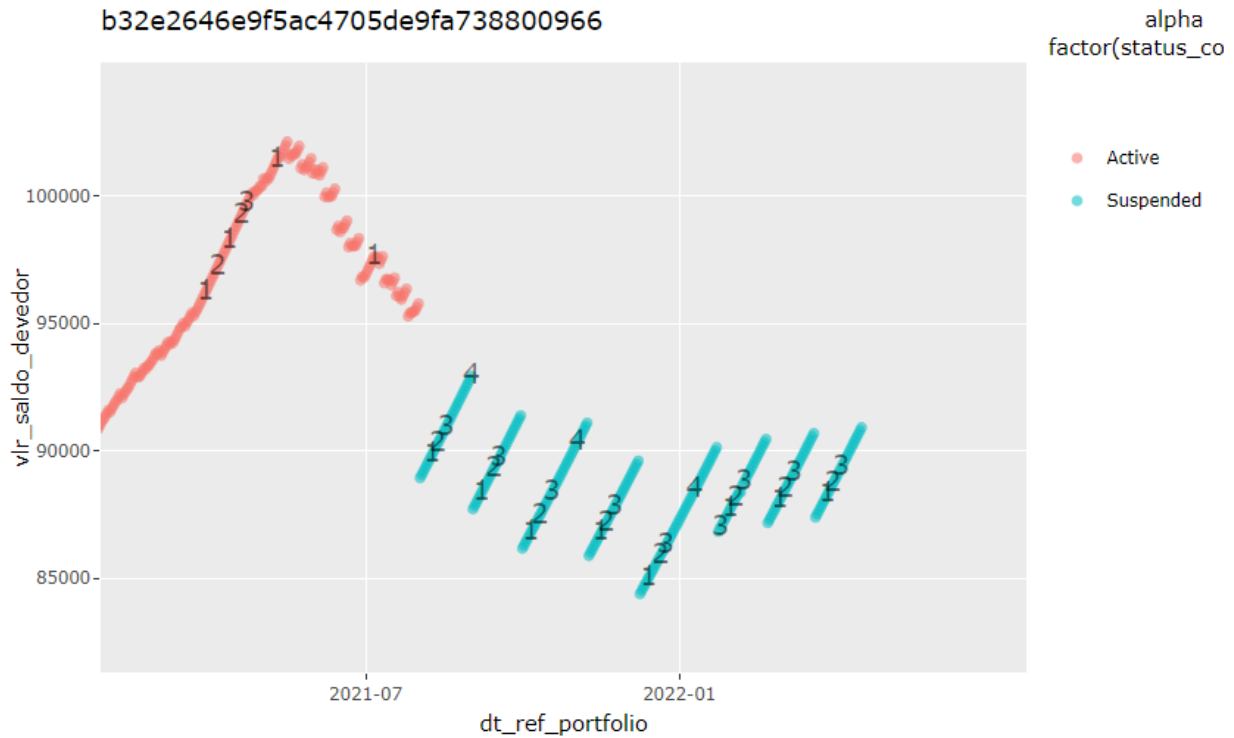
Pelo gráfico acima, notamos que a classificação em regimes de pagamento (números de 139 a 153) para o saldo devedor foi feita de forma satisfatória. Essa simples classificação nos dará uma grande vantagem ao analisar essa variável, pois podemos fazer diversas comparações entre os regimes que podem trazer valor para o negócio. como por exemplo:

1. Se nossa meta é reduzir o saldo devedor do cliente, podemos agora analisar quais são os fatores que fazem a transição de um estado para outro (dia antes do pagamento e dias após pagamento). Talvez a uma ação mais drástica (como a negativação) influenciou essa queda?
2. Dado um regime atual, calcular a probabilidade do cliente transitar para um outro regime (com maior ou menor dívida).
3. Prever se a dívida do cliente chegará a um limite irreversível, onde teremos de agir com o cliente para que as dívidas não atinjam o patamar de bola de neve, em que o pagamento se torna insustentável.
4. As possibilidades de análise são variadas, dependendo dos nossos objetivos.

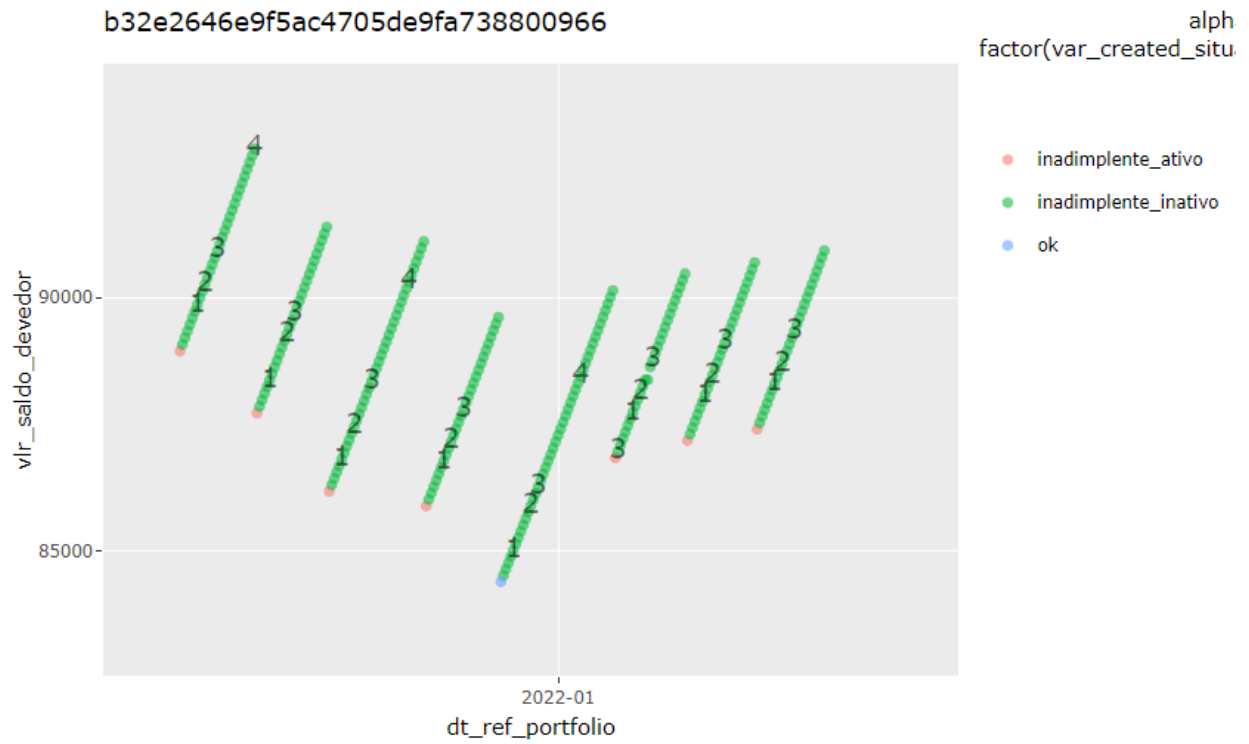
REGIME	INÍCIO REGIME	DURAÇÃO REGIME [DIAS]	ÚLTIMA CAMPANHA INFORMADA AO CLIENTE	DIA DO MÊS	MÊS	SEMANA ISO	DIA DA SEMANA
4	2021-08-02	30	4	2	8	31	segunda-feira
6	2021-09-02	28	3	2	9	35	quinta-feira
8	2021-10-01	38	4	1	10	39	sexta-feira
10	2021-11-09	29	3	9	11	45	terça-feira
12	2021-12-09	45	4	9	12	49	quinta-feira
14	2022-01-24	28	3	24	1	4	segunda-feira
16	2022-02-22	27	3	22	2	8	terça-feira
18	2022-03-22	27	3	22	3	12	terça-feira

A tabela acima mostra os regimes identificados. Note que temos diversos padrões que podem ser explorados e que podem nos ajudar como, a duração de cada regime, o dia do mês do início do regime etc. Dessa forma, para cada cliente individual, podemos comparar o regime atual e comparar com os regimes predecessores. Assim, entenderemos melhor cada cliente individual e poderemos regular a régua conforme os melhores critérios do nosso business.

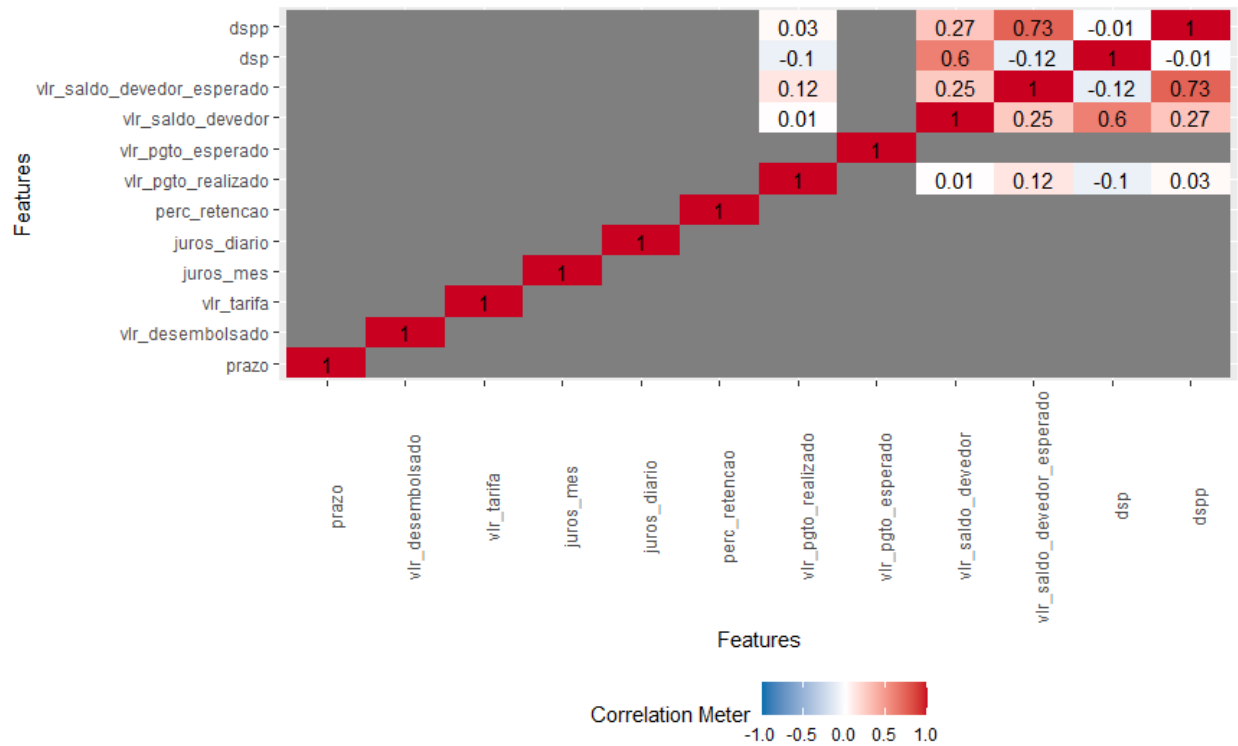
Após uma análise mais criteriosa, vimos que para a o período analisado o contrato estava suspenso., conforme podemos notar pela imagem abaixo. Será necessário fazer uma análise mais criteriosa e entender como funciona o fluxo de pagamento conforme o tipo de contrato.



Notamos também que sempre que podemos identificar a reversão do regime da dívida vigente simplesmente identificando se houve algum pagamento, pela $DSP=0$. Nesse caso, podemos simplesmente usar quando as DSP é nula para delimitar os regimes, sem a necessidade da complexidade que foi desenvolvida anteriormente. Ainda assim, podemos usar o número máximo de dias sem pagar como métrica do ciclo de pagamento do cliente, o que tentaremos reduzir o máximo possível.



Agora, iremos focar na seleção dos melhores atributos com base nesse cliente:



No gráfico acima, as variáveis que apresentam toda a linha/coluna em cinza apresentam variância zero, por isso provavelmente não serão úteis nas análises preditivas que serão desenvolvidas. Temos de averiguar se as variáveis estimadas (como a *vlr_pgto_esperado*) terão de ser removidas da análise, pois elas poderão enviesar o modelo.