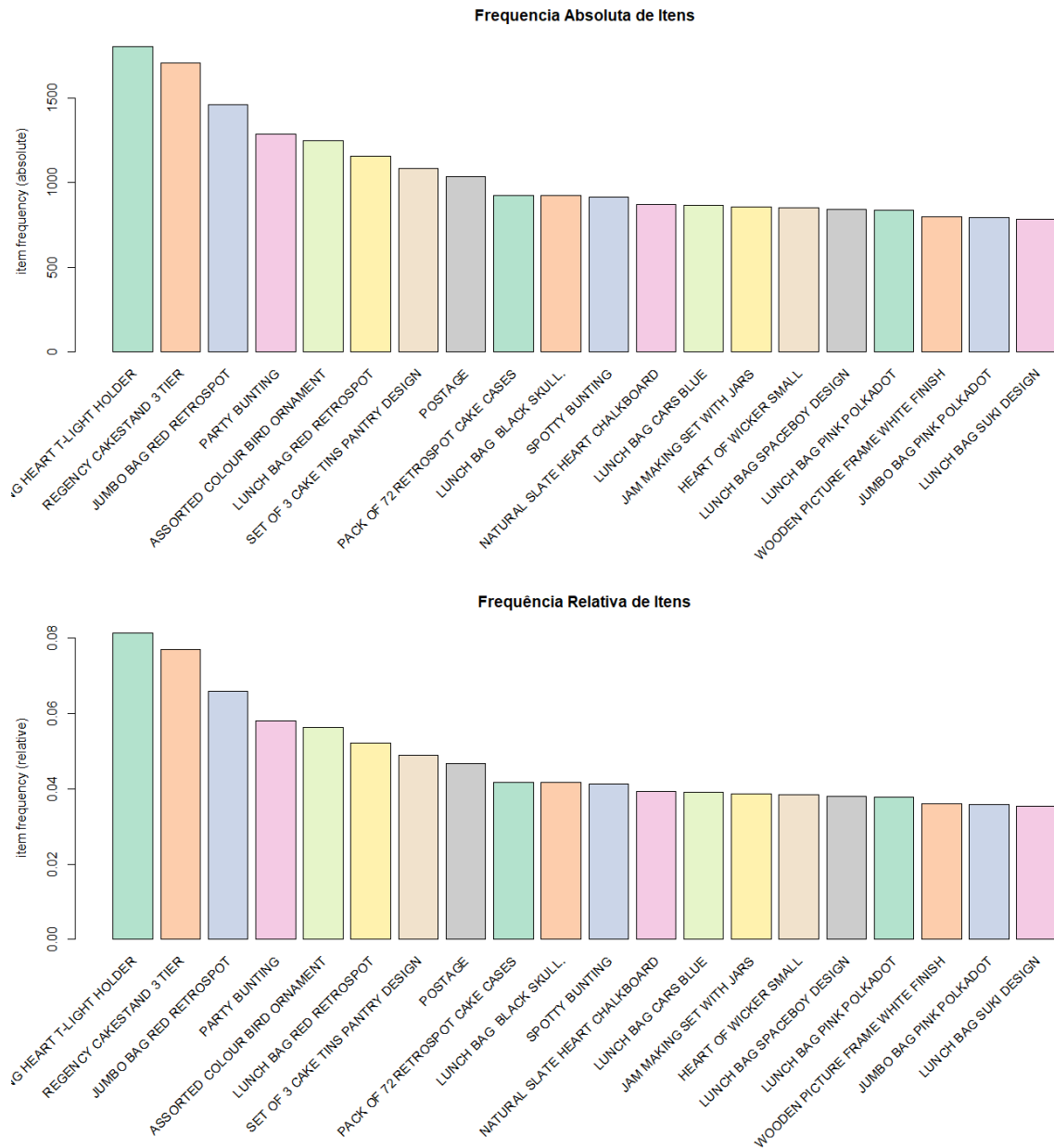
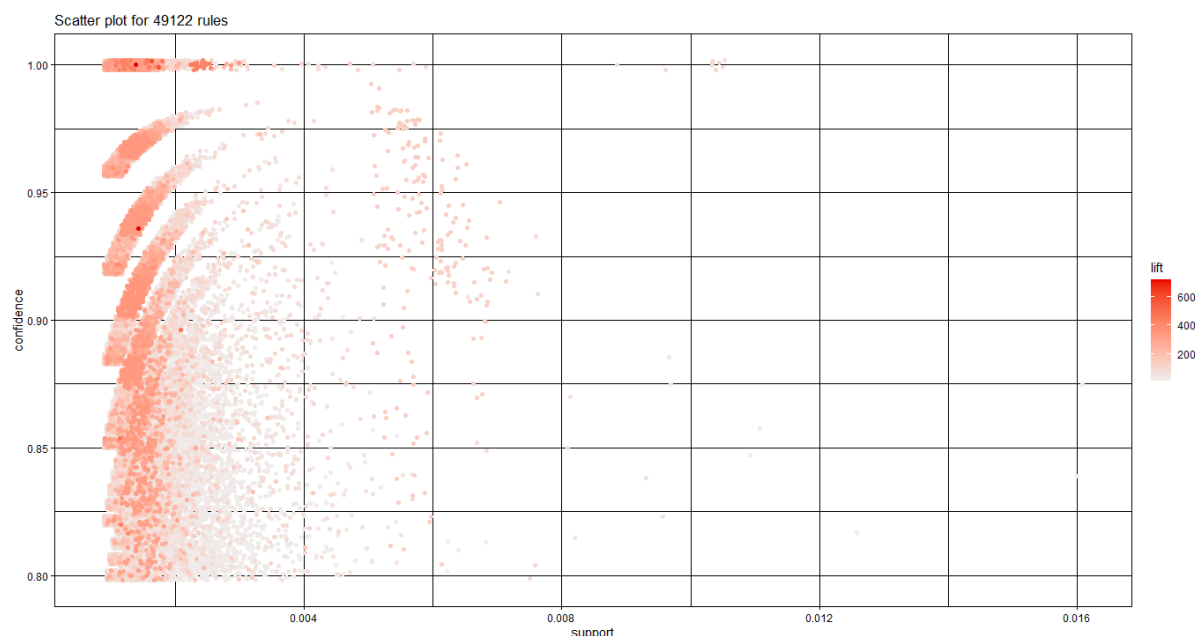


Algoritmos de Associação e Otimização

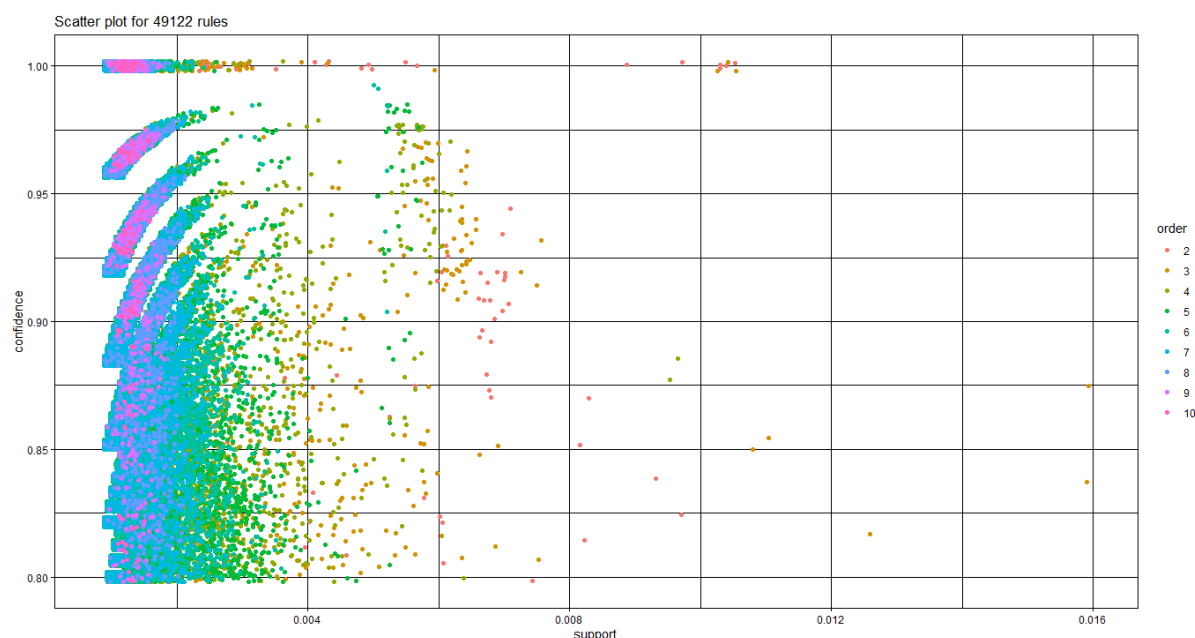
Regra de Associação



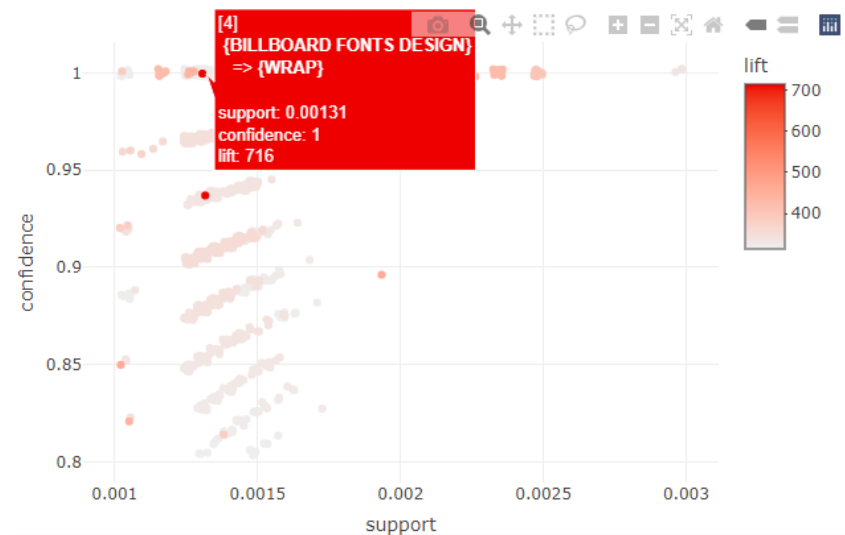
Analisando a distribuição do dataset que contém dados de transações de um varejo on-line no Reino Unido registradas entre 01/12/2010 e 09/12/2011, nota-se que os produtos 'WHITE HANGING HEART T-LIGHT HOLDER' e 'REGENCY CAKESTAND 3 TIER' tiveram maior representatividade no share de vendas seja em frequência absoluta seja em frequência relativa, como pode ser observado nos dois gráficos acima.



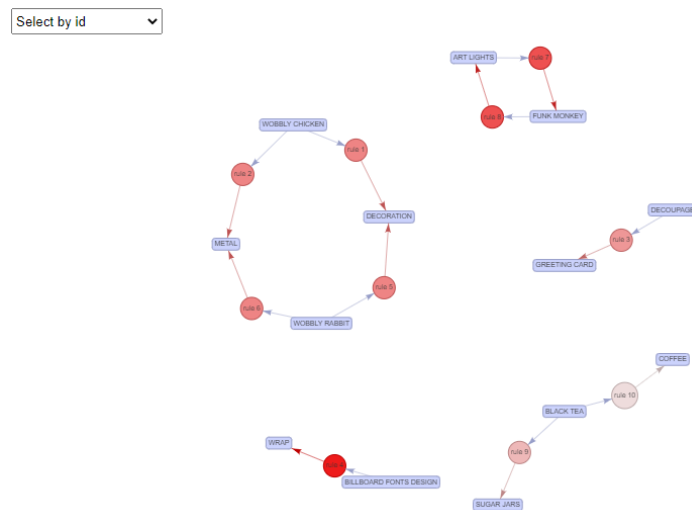
Já quando analisamos o gráfico de dispersão acima, podemos ver que, no caso analisado, há uma maior concentração de regras com baixo suporte, ou seja, as regras não apresentam uma frequência alta em relação ao número total de transações. É possível observar ainda que, de modo geral, a maioria das regras apresentam um lift menor que 400, sendo que as regras que apresentam um maior lift, ou seja, aquelas em que há uma maior relação entre a frequência observada e a frequência esperada de uma regra de associação, apresentam também baixo valor de suporte.



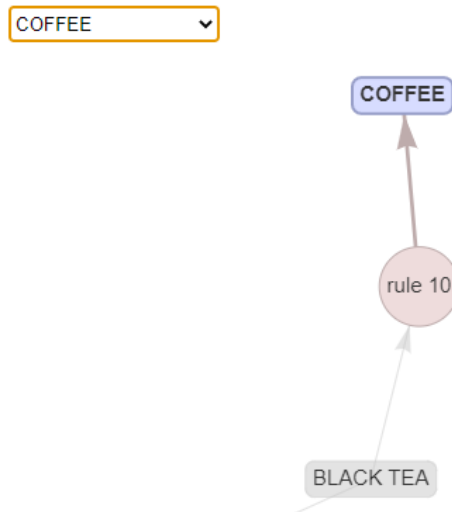
Além disso, apesar da quantidade de itens em uma regra de associação não ser um fator determinante para o seu suporte, neste caso analisado, ao observar o gráfico acima, nota-se que as regras que apresentam menor suporte têm uma maior quantidade de itens, havendo também uma grande quantidade de regras com ordem entre 5 e 7, aproximadamente.



O gráfico acima mostra o suporte, a confiança e o lift para a regra que tem como “se” o produto “billboard fonts design” e como “então” o produto “wrap”, sendo que suporte diz respeito à frequência com que os dois itens são comprados; a confiança, a probabilidade que a regra tem de ser verdadeira; e o lift, a correlação entre “billboard” e “wrap” na regra.

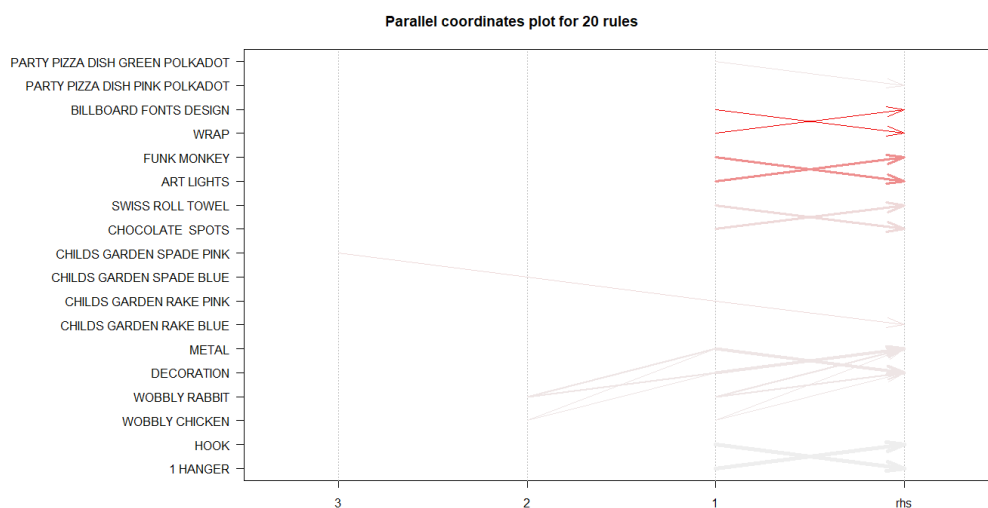


O gráfico acima mostra algumas das associações encontradas. O abaixo, por sua vez, apresenta o caso específico do café (“COFFE”).



Assim, tendo o café como exemplo, o que o gráfico gerado descreve é que clientes que compram chá preto (a parte “se” da equação), tendem a comprar café também (a parte “então” da equação). O que seria, relativamente, esperado.

Para encontrar casos menos intuitivos, expandiu-se o número de regras de 10 para 20 (que seriam as 20 melhores). Encontrou-se, por exemplo, o caso da compra de berçários (a parte “se”) implicar na compra da letra C pintada. Provavelmente, o que acontece é que as pessoas tendem a comprar, junto com o berçário, letras coloridas para formarem o nome de seus filhos, que serão penduradas em algum lugar dos seus quartos.



O gráfico acima, por sua vez, sintetiza as associações. Pode-se ver, por exemplo, que pessoas que têm em suas cestas “wrap” tendem a comprar, também, “billboard fonts design”. Além disso, essa associação é mais intensa que a entre “hook” e “1 hanger”.

Pode-se notar, também, que o gráfico citado apresenta associações entre mais de 1 item, como a vista entre “Childs Garden Spade Pink” e “Child Garden Rake Pink” em relação a “Child Garden Rake Blue”.

Conclusão

A partir do presente laboratório, foi possível entender na prática como é possível modelar um algoritmo de associação no R, quais métricas são utilizadas para identificar as associações, como suporte, confiança e lift, e quais tipos de insights podemos extrair dessa técnica de *machine learning* no contexto do mundo dos negócios, seja para um varejo ou para uma estratégia de marketing. No dataset analisado, foi possível identificar quais produtos são frequentemente comprados juntos no varejo on-line em questão, o que pode ser usado para otimizar as recomendações de produtos aos clientes no próprio e-commerce da empresa.

Além disso, tendo em vista o modo como a técnica abordada neste relatório funciona, notamos também que uma das suas principais limitações do modelo de associação feito no laboratório diz respeito às regras mais comuns. Isso porque, se por um lado o algoritmo captou as principais associações, por outro as associações mais intensas eram extremamente intuitivas (como, por exemplo, a associação entre açúcar e chá). Assim, necessita-se de uma visão mais ampliada - e o grupo, por exemplo, caminhou nessa direção ao expandir as regras analisadas de 10 para 20.