

Segue o escopo:

A primeira coisa a apresentar é a maneira como obtidas as informações. Meu trabalho de segmentação foi realizado pelo modelo RFM, que representa: 1. A recência, que responde a seguinte pergunta: Há quantos dias foi a última compra desse cliente? 2. A frequência, que é quantas compras esse cliente já fez na empresa desde que se cadastrou. 3. A monetaridade ou receita, que é quanto esse cliente já gastou em dinheiro na sua empresa.

máquina, um método não supervisionado de aprendizado de máquina para clusterização.

A partir disso criei a segmentação. Em termos de modelagem, utilizei o k-means, que é um algoritmo não supervisionado de clusterização. Definindo quatro clusters por conjunto RFM: 4 clusters para recência, 4 clusters para frequência e 4 clusters para receita. Utilizei tanto a regra da curva quanto a regra matemática para identificar o número ideal de clusters. Esse número de clusters se tornou uma marcação e no final eu somei essa marcação para chegar ao placar final de cada um dos nossos clientes. Então, criei quatro segmentos baseados nessa nota - o master, o business, o inativo e o premium.

Neste caso, visualizei que 15,9% dos nossos clientes estão inativos, 53,84% estão dentro da categoria business, 29,66% na master e apenas 0,6% na premium. Feito isso, qual ação tomar com relação a esses segmentos?

Recomendações ao tempo de marketing: Em primeiro lugar, é interessante que clientes inativos sejam recuperados com promoções ou cupons especiais. Em seguida, seria interessante fornecer descontos especiais de frete aos masters, que são aqueles clientes que têm uma consideração razoável, então eles "merecem" ter esse benefício. Outra possibilidade seria dar um desconto especial aos clientes premium, atendimento especializado, já que apenas 0,6% desses clientes estão dentro do segmento, então eles merecem mais atenção (uma vez que eles têm uma nota maior).

Dentro desses segmentos todos, eu consegui analisar alguns casos em específico. Então, dentro do gráfico de frequência, por exemplo, observei clientes promissores. Aquele cliente que compra bastante e compra um pouco tempo. Aqui a sugestão seria aumentar a quantidade de compras dele ou transformá-lo em um cliente fiel a partir de recomendações específicas.

Localizei os clientes que costumeiramente seguidos de escorregões, aqueles clientes que não compram há muito tempo, entretanto, compravam com certa frequência e fazem muito tempo que não compram mais. Nesse caso, a possibilidade desses clientes irem para o negativo é um passinho... Seria interessante atender esses clientes por meio de promoções.

No gráfico de recência x frequência temos os clientes que comunidade de baleias. São aqueles clientes que gastam muito dinheiro. Para esse dinheiro o cupom de desconto não vale muita coisa. Para ele um atendimento personalizado ou indicação de produto vale muito mais a pena. Conseguimos converter essa compra de forma mais eficiente.

Identifiquei também os clientes leais, que são aqueles clientes que compram com bastante frequência e gastam um valor bastante razoável. Eles confiam na loja. Então, descontos, programas especializados, captar esse cliente como membro é muito importante no esquema de fidelidade.

Observando a análise de receita x recência visualizei, localizei o tipo de cliente que não é baleia, nem escorregão, ou seja, ele possui características tanto do cliente baleia quanto do cliente escorregão. Ele gasta muito, só que faz muito tempo que ele não compra... A recência é alta e é necessário recuperar esse cliente. É interessante uma campanha de marketing específica para recuperar esse tipo de cliente, atendimento especializado. Às vezes até ligar para esse cliente para entender quais são as demandas dele (nessa oportunidade, já consolidar um link com descontos em produtos que correspondem ao perfil de compra dele).

Esse é o tipo de apresentação que costumeiramente faço para o time de produto ou marketing, ou seja, para a gerência que define o problema. O objetivo aqui é fazer com que eles entendam "a solução" sem a parte de código, sem a parte matemática e ainda assim o plano de dados.

Quanto à parte técnica, segue a síntese do pipeline de dados, bem como o algoritmo (arquivo anexo):

- Importação de dados e bibliotecas pandas, numpy, matplotlib, pyplot, seaborn e datetime.
- Disputa de dados
- Análise exploratória de dados com visualização de dados (incluindo análise do CEP (cidade, estado e país) e horário da compra)
- Clusterização (com especificação e descrição técnica da definição da quantidade ideal de clusters)
- Segmentação