

# AVALIAÇÃO – DATA SCIENCE MANAGER



**Leandro Carvalho**

 [br.linkedin.com/in/leandroscarvalho](https://br.linkedin.com/in/leandroscarvalho)

*Conteúdo restrito e confidencial para avaliação. Sem uso comercial.*

### Questão 1:

#### Processo

1.1) Importação dos arquivos JSON para tabelas equivalentes em um banco de dados;

1.2) Transformação dos dados: conversão da diferença entre *RegisteredDate* e *SubscriptionDate* em: horas, dias, semanas e meses;

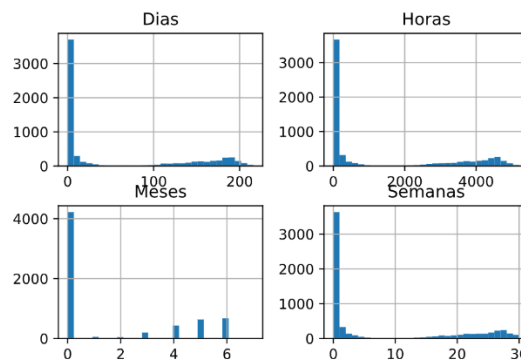
```
SELECT
    StudentID, RegisteredDate, SubscriptionDate,
    DATEDIFF(HOUR, RegisteredDate, SubscriptionDate) as Horas,
    DATEDIFF(HOUR, RegisteredDate, SubscriptionDate)/24 as Dias,
    ((DATEDIFF(HOUR, RegisteredDate, SubscriptionDate)/24)/7) as Semanas,
    FLOOR((DATEDIFF(HOUR, RegisteredDate, SubscriptionDate)/24)/30.436875E)
as Meses
FROM [PD_A-01-premium_students]
ORDER BY DATEDIFF(HOUR, RegisteredDate, SubscriptionDate)
```

1.3) Estatísticas básicas com Python (gerei um .csv com os dados da consulta acima e carreguei em um Notebook utilizando Pandas). Calculei a diferença da data do cadastro com a data da assinatura;

- Resultado das estatísticas (descritivas):

	count	mean	std	min	25%	50%	75%	max
Horas	6260.0	1293.789457	1843.913431	0.0	2.0	39.0	3290.0	5205.0
Dias	6260.0	53.627636	76.729395	0.0	0.0	1.0	137.0	216.0
Semanas	6260.0	7.446965	10.817838	0.0	0.0	0.0	19.0	30.0
Meses	6260.0	1.552077	2.339325	0.0	0.0	0.0	4.0	7.0

- Resultado das estatísticas (histogramas):



1.4) **Decisão**: a partir da análise dos dados e levando em conta o modelo de negócio da empresa **Passei Direto**, optei por dividir os ranges da seguinte forma (período) para analisar (motivos):

Período	Descrição – Motivos
24 horas	Usuários que se tornaram Premium dentro do período de até 24 horas.
1 dia	Usuários que, após as 24 horas iniciais, se tornaram Premium no dia seguinte.
2 dias	Usuários que se tornaram Premium no 2º dia após o cadastro.
até 7 dias	Usuários que, após o segundo dia, se tornaram Premium em até 1 semana.
até 15 dias	Usuários que se tornaram Premium em até 15 dias após o cadastro.
até 30 dias	Usuários que, após a 1ª quinzena, se tornaram Premium em até 30 dias após o cadastro.
No mês seguinte	Usuário que após o 1º mês de uso no modelo free, se tornaram Premium no mês seguinte (até 60 dias).
2 meses	Usuários que experimentaram por mais de 60 dias (o equivalente a um bimestre escolar) e se tornaram Premium no bimestre seguinte.
3-4 meses	Usuários que se tornaram Premium após o fim do primeiro bimestre.
5-6 meses	Usuários que levaram até 6 meses para se tornarem Premium.
acima de 6 meses	Usuários que se tornaram Premium após 6 meses.

1.5) Tendo em vista a definição dos ranges com base nos motivos apresentados, utilizei a seguinte consulta para calcular a frequência e a probabilidade de ocorrência.

```

SELECT
    Período,
    COUNT(Período) AS Assinantes,
    CONCAT(CONVERT(DECIMAL(18,2), ROUND(COUNT(*) * 100.0 / SUM(COUNT(*))
over(), 2)), '%') AS Probabilidade
FROM
    (SELECT
        CASE
            WHEN Horas < 24 THEN '24 horas'
            WHEN DIAS = 1 THEN '1 dia'
            WHEN DIAS = 2 THEN '2 dias'
            WHEN DIAS >= 3 AND DIAS <= 7 THEN 'até 7 dias'
            WHEN DIAS >= 8 AND DIAS <= 14 THEN 'até 15 dias'
            WHEN DIAS >= 15 AND DIAS <= 30 THEN 'até 30 dias'
            WHEN MESES = 1 THEN 'No mês seguinte'
            WHEN MESES = 2 THEN '2 meses'
            WHEN MESES = 3 OR MESES = 4 THEN '3-4 meses'
            WHEN MESES = 5 OR MESES = 6 THEN '5-6 meses'
            WHEN MESES > 6 THEN 'acima de 6 meses'
        END AS Período
    FROM
        (
            SELECT
                StudentID, RegisteredDate, SubscriptionDate,
                DATEDIFF(HOUR, RegisteredDate, SubscriptionDate) as
                Horas,
                DATEDIFF(HOUR, RegisteredDate, SubscriptionDate)/24 as
                Dias,
                FLOOR((DATEDIFF(HOUR, RegisteredDate,
                SubscriptionDate)/24)/30.436875E) as Meses
            FROM [PD_A-01-premium_students]
        )
    AS students) as prob
GROUP BY Período

```

#### 1.6) Resultado:

Período	Assinantes	Probabilidade
24 horas	2988	47,73%
1º dia	202	3,23%
2º dias	121	1,93%
até 7 dias	388	6,20%
até 15 dias	296	4,73%
até 30 dias	223	3,56%
No mês seguinte	57	0,91%
2 meses	44	0,70%
3-4 meses	626	10,00%
5-6 meses	1309	20,91%
acima de 6 meses	6	0,10%

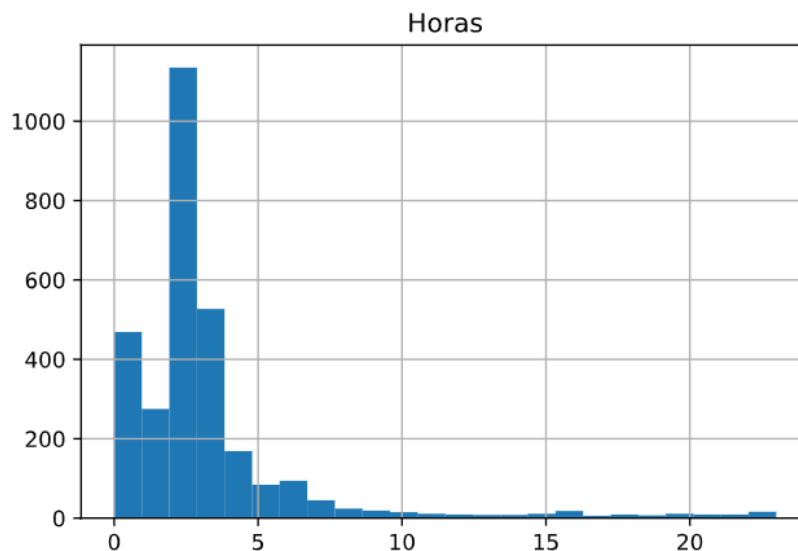
#### 1.7) Conclusões:

- Há uma **probabilidade** muito maior **de conversão (47,73%)** dentro do período de **24 horas**;
- Nos dias seguintes, a taxa cai bastante: 3,23% no primeiro dia e 1,93% no segundo;
- Não convertendo em de 3 dias, a chance de conversão em até 1 semana é de 6,20% e de 4,73% em até 15 dias);
- Nota-se que há um **aumento de 10%** na conversão a partir da faixa entre **3 – 4 meses**. Vale ressaltar que esta faixa representa o bimestre seguinte a partir da data do cadastro.
- O mesmo ocorre na faixa entre **5-6 meses**, com **aumento de 20,91%**;
- A partir dos 6 meses, é remota a probabilidade de conversão (lembrando que o valor máximo em meses é igual a 7 e que são poucos os registros que chegam até este valor, ou seja, há um desbalanceamento na nesta última classe em relação com as demais).

#### 1.8) **Ações estratégicas e Insights:**

- Tendo em vista que as chances de **conversão** se concentram dentro do período de **até 2 dias**, com um total de **52,89%** (**levando em consideração as 3 primeiras faixas** do range) a empresa precisa agrupar os esforços de marketing neste período, como:
  - Garantir a excelência e um suporte ativo no atendimento a estes clientes;
  - Avaliar a **criação de um modelo preditivo** (a partir dos dados de engajamento, padrão de consumo e comportamento) **para sugerir se um usuário poderá ou não se tornar ativo em até 2 dias** (ou a criação de vários modelos para cada faixa – explicarei melhor na apresentação);
  - Com estes resultados preditivos, para aqueles usuários indicados como alta probabilidade de não se tornarem ativos dentro do prazo de até 2 dias (ou até mesmo em 24 horas, por que não), criar uma comunicação ativa, com uma abordagem suave e sem incomodar, para que o time de comercial de ativação consiga converter o usuário;

- Para reforçar a sugestão sobre a criação do modelo preditivo, gerei um novo histograma (imagem a seguir) com os dados dos usuários ativos (premium) dentro das primeiras 24 horas (refiz a consulta no BD e exportei um novo .csv). Nele, é possível observar que as 4 primeiras horas são fundamentais na decisão do usuário de assinar ou não o serviço. A partir daí o interesse do usuário cai conforme as horas vão passando. Por isso, vale a pena experimentar a criação de um modelo, mas desde que sejam criadas ações estratégicas efetivas para dar suporte ao time de marketing e vendas;



**Atenção:** *é preciso observar de perto o resultado do modelo, para não só avaliar a conversão dentro do período, mas também se haverá uma não conversão maior ao longo do tempo. Pois se a ação ativa de contato não for bem-sucedida, o usuário poderá bloquear os contatos da empresa e até mesmo se tornar um detrator (explicarei melhor na apresentação).*

- Outra **ação** que poderia ser tentada seria atuar dentro das **faixas** entre **3-4 meses** e **5-6 meses**, já que juntas possuem uma probabilidade de **conversão** de **30,91%**. Aqui é provável (e deveria ser investigado) que o usuário experimentou o serviço no primeiro bimestre e ativou no segundo. Mas por quê? Teria ele levado “bomba” no período anterior e resolveu correr atrás do prejuízo ou se saiu muito bem a partir do material consumido que resolveu seguir adiante? Esta não é uma questão de modelo, mas sim de negócio. A mesma avaliação vale para o período de 5-6 meses (análise do semestre, outro marco importante do estudo);

- Por último, poderia ser criada uma análise a partir de um gráfico de Cohort para acompanhar a conversão e a retenção dos clientes ao longo do período. Neste caso, um exemplo seria:

Extração dos dados com utilização de Pivot Table (fiz por SQL, código abaixo):

```

SELECT *
FROM (
    SELECT
        StudentID,
        FORMAT(RegisteredDate, 'yyyy-MM-dd') AS Registro,
        FORMAT(SubscriptionDate, 'yyyy-MM') AS Assintura
    FROM [PD_A-01-premium_students]
) as s
PIVOT
(
    COUNT(StudentID)
    FOR [Assintura] IN ([2017-11], [2017-12], [2018-01], [2018-02], [2018-03],
[2018-04], [2018-05], [2018-06])
)AS pvt

```

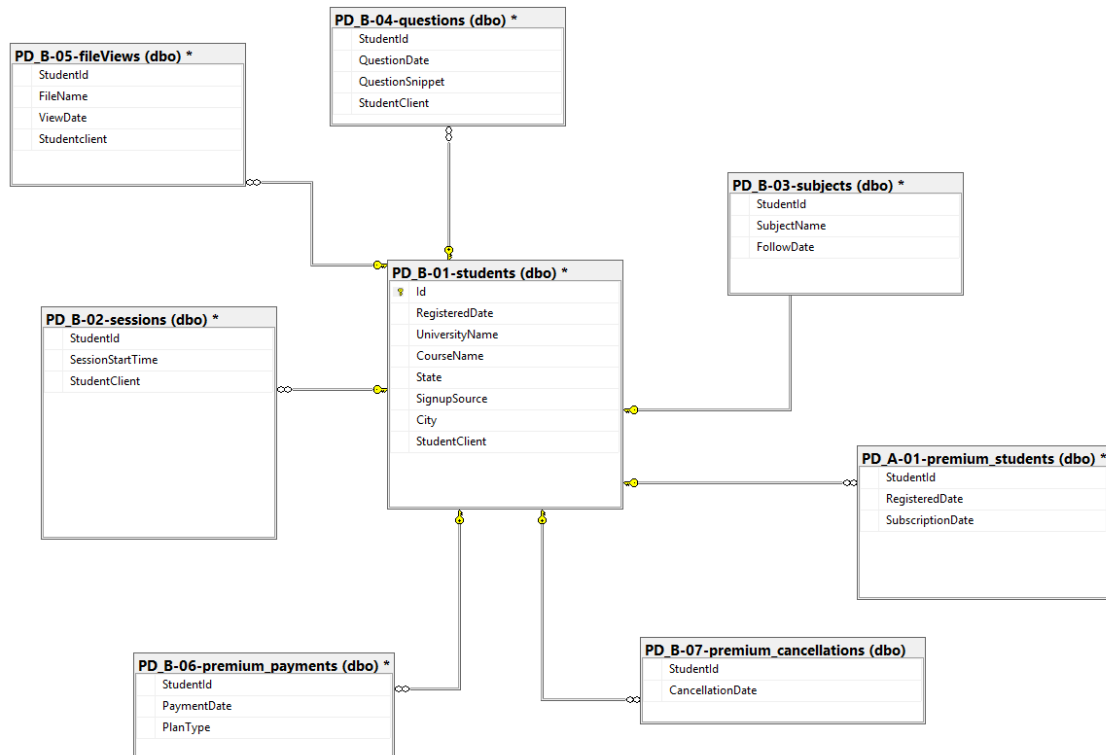
E depois, em alguma ferramenta de dados (PowerBI, QuickSight, Tableau, excel, etc.) visualizaria o resultado, conforme abaixo:

	2017-11	2017-12	2018-01	2018-02	2018-03	2018-04	2018-05	2018-06
2017-11-01	2,19%	0,00%	0,00%	0,02%	0,08%	0,29%	0,43%	0,03%
2017-11-02	2,48%	0,02%	0,00%	0,03%	0,11%	0,38%	0,56%	0,03%
2017-11-03	2,22%	0,08%	0,02%	0,03%	0,08%	0,21%	0,40%	0,03%
2017-11-04	2,43%	0,02%	0,02%	0,03%	0,06%	0,29%	0,54%	0,08%
2017-11-05	3,96%	0,08%	0,00%	0,02%	0,19%	0,22%	0,86%	0,10%
2017-11-06	3,29%	0,14%	0,02%	0,00%	0,14%	0,45%	0,77%	0,08%
2017-11-07	2,54%	0,05%	0,02%	0,08%	0,26%	0,48%	0,67%	0,05%
2017-11-08	2,81%	0,24%	0,00%	0,08%	0,14%	0,26%	0,46%	0,08%
2017-11-09	2,22%	0,10%	0,02%	0,00%	0,10%	0,37%	0,53%	0,10%
2017-11-10	2,17%	0,13%	0,03%	0,02%	0,14%	0,26%	0,38%	0,08%
2017-11-11	2,09%	0,11%	0,02%	0,06%	0,11%	0,22%	0,45%	0,06%
2017-11-12	2,68%	0,08%	0,00%	0,08%	0,08%	0,35%	0,54%	0,06%
2017-11-13	4,31%	0,18%	0,03%	0,03%	0,19%	0,58%	1,13%	0,11%
2017-11-14	4,90%	0,14%	0,00%	0,05%	0,26%	0,58%	1,15%	0,14%
2017-11-15	2,81%	0,05%	0,03%	0,02%	0,14%	0,50%	0,38%	0,08%
2017-11-16	2,80%	0,26%	0,02%	0,02%	0,26%	0,43%	0,48%	0,13%
2017-11-17	1,84%	0,14%	0,02%	0,00%	0,14%	0,42%	0,61%	0,03%
2017-11-18	1,85%	0,14%	0,00%	0,06%	0,16%	0,16%	0,40%	0,08%
2017-11-19	3,04%	0,21%	0,00%	0,00%	0,18%	0,26%	0,54%	0,05%
2017-11-20	2,36%	0,34%	0,00%	0,11%	0,37%	0,35%	0,51%	0,06%
2017-11-21	1,34%	0,19%	0,00%	0,06%	0,24%	0,48%	0,48%	0,10%
2017-11-22	1,01%	0,19%	0,00%	0,05%	0,13%	0,30%	0,42%	0,11%
2017-11-23	0,85%	0,14%	0,00%	0,10%	0,08%	0,27%	0,35%	0,06%
2017-11-24	0,85%	0,06%	0,00%	0,03%	0,19%	0,14%	0,24%	0,05%
2017-11-25	0,53%	0,08%	0,00%	0,03%	0,14%	0,16%	0,19%	0,00%
2017-11-26	0,75%	0,21%	0,02%	0,02%	0,10%	0,16%	0,11%	0,02%
2017-11-27	1,28%	0,26%	0,00%	0,02%	0,22%	0,24%	0,18%	0,03%
2017-11-28	0,83%	0,24%	0,06%	0,05%	0,19%	0,19%	0,16%	0,10%
2017-11-29	0,75%	0,24%	0,00%	0,03%	0,14%	0,29%	0,16%	0,05%
2017-11-30	0,58%	0,35%	0,00%	0,00%	0,03%	0,19%	0,14%	0,03%

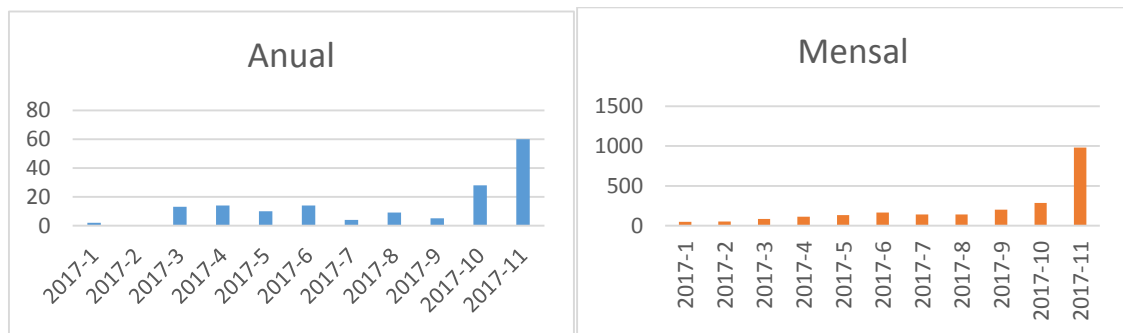
## Questão 2:

Processo (código disponível em anexo ou no Git):

### 2.1) Estruturação das tabelas no Banco de Dados;



2.2) O primeiro passo foi dar uma explorada nos dados para ver os pagamentos até o período de nov./2017, separados por mês e ano (sem considerar os cancelamentos):



Nota-se sem dúvida que houve um crescimento nos pagamentos a partir de novembro de 2017. Decisão acertada em termos de estratégia de negócio (parte financeira).

Obs.: É preciso ter em mente que há clientes com mais de um pagamento no período. Neste caso, por uma questão de análise, foi feito um DISTINCT na base de dados para contar somente 1 pagamento por cliente em cada período (por plano).

2.3) Há várias formas para se calcular o LTV, a maioria utiliza a taxa de churn. Porém, há também várias formas de se calcular o CHURN. Ex.:

- Clientes do mês anterior que não assinaram o mês atual;

- Clientes que se cadastram no mês atual e cancelaram no próprio mês; ou
- Um somatório dos 2 para se tirar uma média.

2.4) Na sequência, para cada tipo de plano (Anual - R\$ 23,90/mês - Mensal - R\$ 29,90/mês), foi preciso identificar a quantidade de usuários premiums e que ainda continuaram ativos e não ativos no período, para depois separá-los por planos.

#### **MENSAL:**

- A) Identificar os usuários ativos no **plano mensal**;
- B) Em outubro houve pagamento de **286 clientes distintos** (apesar de termos 288 registros de pagamentos no período – 1 usuário com 2 pagamentos). Destes clientes, **13 cancelaram em outubro**.
- C) Assim, um total de **273 clientes assinantes começaram** o mês de novembro. Porém, destes clientes, **29 cancelaram em novembro** (Levei em consideração apenas os clientes que pediram formalmente o cancelamento. Não levei em consideração os clientes que deixaram de pagar, até porque o time comercial poderia estar negociando a permanência deles).
- D) Em nov. /2017 **718 clientes novos** se cadastraram no plano anual. Destes clientes, **67 cancelaram** no mesmo mês.

#### **CHURN Plano Mensal**

Clientes Recorrente =  $29 / 273 = 10,62\%$

Novos Clientes =  $67 / 718 = 9,33\%$

#### **LTV Plano Mensal** (levarei em consideração a fórmula padrão $LTV = ARPA/Churn$ )

Clientes Recorrente =  $R\$ 29,90 / 10,62\% = R\$ 281,54$

Novos Clientes =  $R\$ 29,90 / 9,33\% = R\$ 320,47$

#### **ANUAL:**

- E) Próximo passo foi identificar os usuários ativos no **plano anual**;
- F) Em **out./2017** havia **101 clientes** ativos no plano anual.
- G) **57 novos clientes** se cadastram em nov. /2017 o plano anual. Destes clientes, nenhum cancelou no próprio mês de cadastro;
- H) **3 clientes** cancelaram o plano anual em nov.2017;
- I) **3 clientes deveriam** ter renovado o plano em nov./2017 mas não o fizeram.
- J) Ao final, a base de assinantes premium do plano anual ficou com um total de **152 clientes distintos ativos** em nov./2017;

#### **CHURN Plano Anual**

Clientes Recorrente =  $6 / 101 = 5,94\%$

Novos Clientes =  $0 / 57 = 0\%$  (sem churn)

#### **LTV Plano Anual** (levarei em consideração a fórmula padrão $LTV = ARPA/Churn$ )

Clientes Recorrente =  $R\$ 23,90 / 10,62\% = R\$ 402,32$



Novos Clientes = Como a taxa de churn destes clientes ficou igual a zero, não foi possível calcular o LTV dos novos clientes neste período. Existem algumas fórmulas para isso, mas creio que aqui (para o teste) não vale a pena.

#### **LTV TIPO DE CLIENTE**

##### **CHURN POR TIPO DE CLIENTE**

Clientes Recorrente =  $(29 + 6) / (273 + 101) = 9.36\%$

Novos Clientes =  $(67 + 0) / (718 + 57) = 8,65\%$

##### **LTV POR TIPO DE CLIENTE**

Clientes Recorrente =  $R\$ 29,90 / 9.36\% = R\$ 319,44$

Novos Clientes =  $R\$ 23,90 / 8,65\% = R\$ 276,30$

#### **LTV FINAL**

**CHURN FINAL TOTAL** =  $102 / 1.149 = 8,88\%$  (utilizando os padrões encontrados por tipo de cliente e tipo de assinatura).

**ARPA (Nov./2017)** =  $(152 \text{ clientes anuais} \times R\$ 23,90) + (895 \text{ clientes mensais} \times R\$ 29,90) = (3.632,80 + 26.760,50) = 30.393,30 / 1.047 = R\$ 29,03$

**LTV TOTAL** =  $29,03 / 8,88\% = R\$ 326,91$

### Questão 3:

Considerando o modelo de negócio da plataforma Passei Direto e os dados disponibilizados até aqui, sugiro:

- ⇒ Criar uma análise para entender o motivo pelo qual um cliente assina e logo cancela o plano mensal. Será que ele chegou a consumir algo e não gostou? Ou consumiu o que queria a tempo de cancelar o seu cadastro dentro do primeiro mês (baixou tudo) e foi embora? Seria importante ir em busca de um padrão, até para garantir o fornecimento de uma melhor experiência aos usuários futuros;
- ⇒ Gerar um modelo de regressão que permita inferir quantos usuários irão cancelar a assinatura do serviço (se tornarão Churn) em um determinado período futuro. Isto permitirá a empresa fazer um melhor planejamento de dimensionamento da equipe, dos recursos computacionais e financeiros;
- ⇒ Mais importante do que saber quantos clientes irão te abandonar é saber quem são estes clientes. Por isso, deve-se criar um modelo de Machine Learning (classificação supervisionada) que não só preveja quantos cliente irão te abandonar, mas informar também quem são estes clientes. Isto pode ser feito a partir dos dados de consumo, engajamento e histórico dos cancelamentos. Com este modelo, a equipe de Grow, sustentação e marketing poderiam desenvolver ações para tentar mitigar as perdas.
- ⇒ Desenvolvimento de um Chatbot que permita ao usuário interagir com o conteúdo cadastrado na plataforma a partir de uma curadoria feita por IA. Neste caso, seria possível entregar um conteúdo mais assertivo, a partir dos seus desejos e padrões de engajamento e consumo, além de economizar o tempo de pesquisa do usuário. Assim, ele teria mais tempo para o que importa, ou seja, estudar a matéria e não em ter que ficar procurando o conteúdo mais apropriado.
- ⇒ Conforme já sugerido no item 1.8, a criação de um modelo preditivo para sugerir se um usuário poderá se tornar ativo ou não assim que fizer o cadastro. Uma vez criado o modelo, é preciso ter ações estratégicas bem definidas (time comercial e marketing) sobre o que fazer. Será que vale a pena entrar em contato com o cliente? Se sim, de que maneira? Email ou telefone/WhatsApp? E com qual periodicidade? É bom ter em mente que estas respostas só aparecerão com o uso do modelo, monitorando dos seus resultados e com o tempo;
- ⇒ Criação de modelos de machine learning (classificação não supervisionada) que permitam identificar os diversos perfis de clientes a partir dos dados de engajamento, padrão de consumo e comportamento de navegação. A partir deste modelo, prever o comportamento futuro para melhorar a experiência dos usuários e garantir para a empresa que os números de crescimento, engajamento e receitas estejam sempre aumentando.
- ⇒ Gerar um modelo de text-mining para segmentar os termos de buscas feitos na plataforma e as perguntas feitas pelos usuários (registrados na tabela *questions*) para determinar o melhor conteúdo e a área de domínio do assunto pesquisado. A ideia seria otimizar a entrega com um conteúdo de mais relevante para o usuário.

#### Questão 4:

Considerando o modelo de negócio da plataforma Passei Direto e os dados disponibilizados até aqui, sugiro:

- Apesar da base permitir uma **segmentação** sócio demográfica, vou optar por uma análise **comportamental**, utilizando o modelo RFM, através do padrão de consumo dos **usuários premium**. Assim, farei uma segmentação que atenda as diversas áreas do PD (Marketing, Vendas, Growth, Retenção e Financeiro);
- Seguindo o modelo RFM, vou considerar uma análise temporal dos clientes, utilizando as informações sobre o quão recente é um cliente (último engajamento), em conjunto com uma financeira, ou seja, o quanto que um cliente gastou e, por último, uma análise de fidelidade com base nos dados de engajamento do cliente por sessões iniciadas (até poderia qualificar o engajamento por questões postadas no site ou por visualizações de arquivos. Porém, farei somente por sessão, mantereí de forma mais simples a análise);
- A primeira etapa necessária foi definir um período para análise. Neste caso, defini um intervalo de tempo entre nov./2017 até mai./2018. Motivo: analisar o histórico comportamental dos usuários nos 6 meses seguintes após a mudança informada na questão 3 (*Content Restriction*);
- Na sequência, fui até a base e agrupei os dados necessários para a análise. Em seguida, exportei para um arquivo .csv e criei 2 modelos utilizando linguagem Python. O primeiro, seguiu o modelo RFM tradicional (divisão dos clientes em percentis); já o segundo, foi feito utilizando Machine Learning, através de um algoritmo de classificação não supervisionada, mas seguindo a mesma ideia do modelo RFM.

Vamos aos resultados finais:

#### Resultado RFM Padrão:

	DiasSemUso	TotalSessoes	TotalGasto	Qtd. Clientes
RFM_Level				
Green	16.8	30.4	38.8	499
Bronze	37.8	38.5	86.1	692
Silver	48.0	54.2	149.5	299
Gold	42.9	73.9	164.3	200

- Neste cenário o conjunto de dados foi dividido em 4 grupos, sendo: Green, Bronze, Silver e Gold. O Gold são os clientes VIPs, possuem o maior gasto médio durante todo o período, possuem o maior engajamento. O Silver também possui um gasto elevado, porém com um engajamento menor que o Gold e com um tempo de inatividade um pouco maior. O Bronze possui um tipo de cliente que gasta quase a metade do que gasta os clientes dos grupos Silver e Gold e possuem um engajamento também menor. Já o Green, pelo que parece, representa os novos clientes, aqueles mais recentes, não só pela média do último dia de acesso, como também pelo seu gasto médio (**lembrando**, os **valores** refletem a **média de gastos no período**, que é **impactado** pelo **tempo** como assinante **premium**).

- Com estes resultados o marketing poderia, por exemplo, customizar as mensagens de acordo com o perfil temporal do cliente (aquele está a mais tempo sem se conectar na plataforma).

Ex.: Fazer uma campanha para os usuários do grupo Green com estratégias relacionadas ao pós-venda. Já para o grupo do Silver, que está há 48 dias sem acesso, informar sobre os novos conteúdos inseridos na plataforma, e que sejam relacionados aos seus interesses.

- Pelo lado financeiro, as equipes de retenção e/ou ativação, poderiam acompanhar de perto os usuários do grupo Gold, pois embora sejam aqueles que mais gastam, eles já estão há 42 dias (em média) sem acessar a plataforma. Estes usuários não podem ser perdidos.

### **Resultado RFM Machine Learning:**

- Na busca por um resultado mais detalhado, utilizei o modelo RFM com Machine Learning. Os dados foram divididos em 8 clusters (a partir da análise do método Elbow). Os resultados podem ser visualizados a seguir:

K_Cluster	DiasSemUso	TotalSessoes	TotalGasto	count
	mean	mean	mean	
0	7.28	139.88	200.33	64
1	11.47	30.55	38.53	709
2	12.76	32.80	205.42	271
3	157.74	18.76	50.98	175
4	19.63	32.56	114.23	293
5	10.30	512.90	66.97	10
6	6.96	138.26	54.16	105
7	177.68	12.67	190.32	63

- Com uma quantidade de clusters maior, os resultados começam a ficar mais claros. Ex.:

- ⇒ Ao que parece o cluster 1 representa os clientes mais novos em relação às assinaturas realizadas, pois o valor médio é de 38,53 (muito próximo do valor mensal das assinaturas premium);
- ⇒ Os cluster 3 e 7 poderiam representar os clientes perdidos, pois estão a mais tempo sem se conectar a plataforma (um bom grupo para o time de ativação do Passei Direto);
- ⇒ O cluster 0 pode ser considerado VIP, já que possui os clientes um bom gasto ao longo do período, são muito ativos e usaram recentemente a plataforma;
- ⇒ O cluster 2 é também um bom cluster, pois possui a maior média de gastos, além de ser recente. A única diferença é que o engajamento médio na plataforma é menor.

Sendo assim, por tudo que foi apresentado, finalizo as análises e me coloco à disposição para eventuais dúvidas.