

Data Scientist Challenge

Passei Direto

Lucas de A. Cerqueira

Agenda

1. Questão 1
2. Questão 2
3. Questão 3
4. Questão 4
5. Conclusão

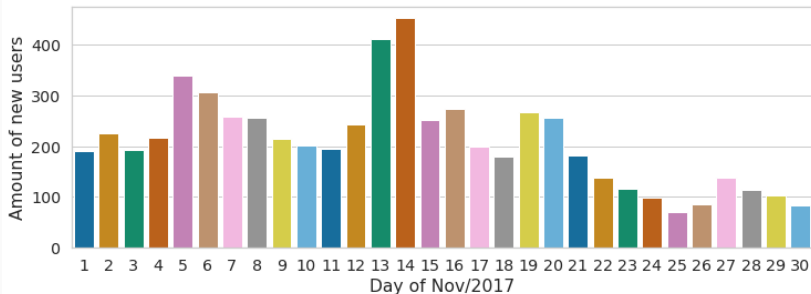
Questão 1

Questão 1

- O objetivo dessa questão era identificar qual a probabilidade do usuário virar Premium após o cadastro em intervalos de dias.
- Os dados possuíam cadastros somente do mês de novembro de 2017, com um total de 6260 registros de novos estudantes.
- As informações disponíveis eram:
 - Número identificador do estudante
 - Data de registro na plataforma
 - Data de assinatura do Premium

Análise das datas de registro

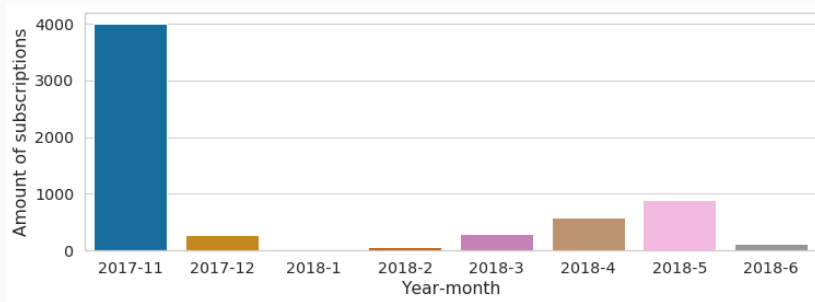
- Intuito de verificar se não há algum comportamento muito específico ao longo do mês analisado.
- Pelo gráfico, há uma redução do número de registros nos últimos 10 dias do mês de novembro, conforme o final do ano letivo se aproxima.



Análise das datas de assinatura

Pelo gráfico, podem ser observados três comportamentos:

- Assinantes ainda em novembro e dezembro de 2017.
- Redução das assinaturas em janeiro e fevereiro¹, coincidindo com férias escolares.
- Aumento do número de assinaturas de março/2018 em diante² conforme o ano letivo recomeça.



¹O Carnaval em 2018 foi na segunda semana de fevereiro.

²Os dados de junho/2018 parecem incompletos na amostra enviada para o desafio.

- Mais de 68% dos novos estudantes viraram Premium ainda em 2017.
- O comportamento de aumento das assinaturas após as férias parece estar intimamente relacionado à época do ano, e não à forma como o usuário é introduzido aos planos e vantagens após o registro.
- Por esses motivos, decidir limitar a análise aos usuários que viraram Premium ainda em 2017.

Resultados

- Dos usuários dentro do recorte realizado, cerca de 75% viraram Premium dentro de 2 dias.
- Em torno de 70% viraram Premium ainda no primeiro dia na plataforma.

Table 1: Percentual de estudantes que viraram Premium em intervalos de dias desde o registro.

Intervalo	Percentual
0 a 1 dia	70.21%
1 a 7 dias	14.89%
7 a 14 dias	7.71%
14 a 30 dias	5.76%
30 a 45 dias	1.43%

Questão 2

Questão 2

- O objetivo dessa questão era calcular o *Lifetime Value* (LTV) dos usuários após a adoção do *Content Restriction*, feita em Nov/2017.
- Utilizei a seguinte fórmula para o cálculo do LTV:

$$LTV = \sum_{j=1}^N v_j s_j \quad (1)$$

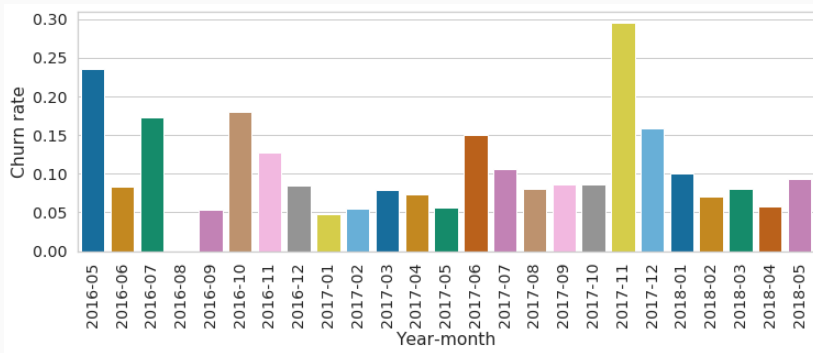
onde v_j é o valor pago e s_j é a taxa de sobrevivência (*survivability*) durante o ciclo de cobrança j . O valor N representa o total de ciclos contabilizados na análise.

- A *survivability* começa em 100% no início do ciclo e decai a um fator de $(1 - c_p)$, onde c_p é a *churn rate* do ciclo p .

$$s_j = \prod_{p=1}^j (1 - c_p) \quad (2)$$

Resultados

- Pico de *churn rate* em Nov/2017.
- Apesar do aumento em Nov/2017, o *churn rate* retornou ao mesmo patamar de antes do *Content Restriction*.



Resultados

- Foi considerado o valor médio de R\$29.90 mensais.
- Foram analisados os 7 meses anteriores a Nov/2017 e os 7 meses posteriores (incluindo Nov/2017).
- Observa-se uma redução do LTV de R\$146,35 para R\$107,80, queda de 26%

Table 2: LTV antes e depois de Nov/2017.

Intervalo	LTV
Abr/2017 - Out/2017	R\$146.35
Nov/2017 - Mai/2018	R\$107.80

Questão 3

Questão 3

- O objetivo dessa questão era identificar possíveis melhorias na experiência dos usuários.
- Senti uma certa dificuldade para identificar as dores dos usuários por ter tido pouca experiência com a plataforma.
- Portanto, decidir focar em comportamentos de alguns usuários que podem afetar a experiência de outros, além de uma análise sobre a quantidade de disciplinas seguidas por usuário.

Perguntas que não são bem perguntas...

- Ao analisar a quantidade de perguntas por usuário, encontrei um usuário com 78 perguntas, sendo que a média por usuário era 1.74.
- Verificando os *snippets* das perguntas desse usuário, vi que em basicamente todas ele dizia possuir arquivos de provas e passava seu número de celular.
- Não sei se o Passei Direto é o melhor espaço para essa forma de compartilhamento.

"Perguntas" do usuário 36619758 (celulares omitidos)

"tenho prova unopar Raciocínio lógico e matemático chama

*****"

"alunos interessado provas unopar chama ** *****_****"

"Prova unopar *****"

"prova unopar wsp ** *****"

"prova unopar ** *****_****"

Análise da quantidade de respostas

- Analisando a quantidade de respostas dadas por usuário, encontrei um deles com 124 respostas, quase o dobro do segundo maior com 67.
- Esse usuário respondia a perguntas enviando links onde supostamente poderiam ser encontrados mais materiais.
- Porém, encontrei várias respostas idênticas, com links quebrados e direcionando para sites como MegaUpload, etc.
- O que poderia ser feito é esconder URLs externas ao domínio do PD ou, pelo menos, avisar ao usuário que ele sairá do domínio ao clicar nesses links.

Respostas do usuário 70700878

"aqui neste link tem mais de 300 livros para abaixar que vão te ajudar a tirar suas duvidas"

"Neste link <https://mega.co.nz/F!Ns4UkDCJ!Ep49t57YcQMleoiIA6rdpQ> tem genes VIII e genes XI"

Disciplinas seguidas por usuário

- Ao se registrar na plataforma o usuário precisa escolher no mínimo uma quantidade pré-definida de disciplinas para seguir.
- Após o registro, a fica a cargo do usuário procurar por outras disciplinas que ele queria seguir.
- Porém, ao analisar a quantidade de momentos diferentes em que um usuário seguiu novas disciplinas, cerca de **70% dos usuários só fizeram isso uma única vez**, quando se registraram no PD.
- Algo que poderia ajudar o usuário é uma sugestão de novas matérias, similar ao que o Medium faz para sugerir novos posts e assuntos.

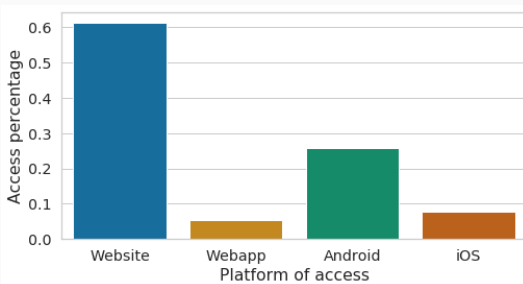
Questão 4

Questão 4

- O objetivo dessa questão era segmentar os usuários e sugerir recomendações, em termos de negócio, considerando as áreas de Marketing, Cliente e Financeiro do PD.
- Por uma limitação de tempo, não consegui me dedicar muito à segmentação, então optei por análises quantitativas que podem auxiliar essas áreas.

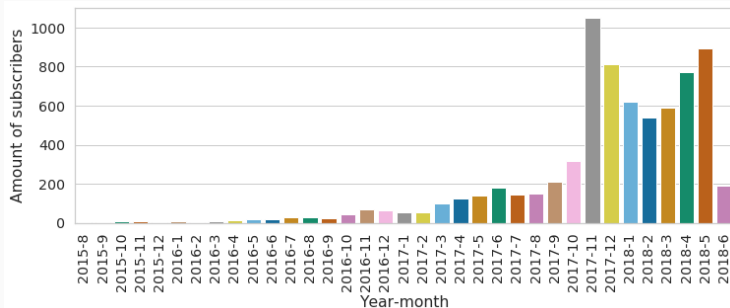
Área de Cliente do PD

- O curso com maior quantidade de alunos é Direito (9862 alunos), seguido de Administração (4099) e Engenharia Civil (3585).
- Diversos cursos da área de Biológicas também estão entre os com mais alunos no PD, como Pedagogia, Psicologia, Enfermagem, Medicina, Fisioterapia, Nutrição e Farmácia.
- Talvez seja mais fácil atrair novos alunos desses cursos devido à grande quantidade de material na plataforma.
- Os aplicativos para Android e iOS já correspondem juntos a mais de 30% do número de acessos.



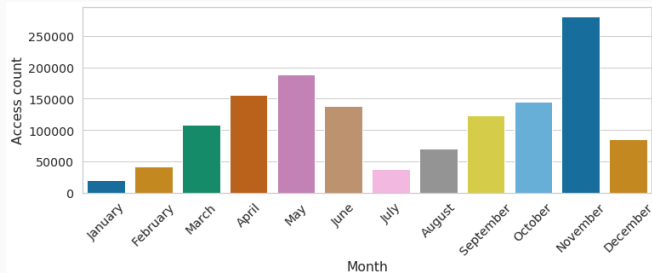
Área de Financeiro do PD

- Dos usuários cujo primeiro plano anual já havia expirado dentro do período analisado (Ago/2015 a Jun/2018), **mais de 50% renovou por mais um ano.**
- Os planos anuais só foram assinados por cerca de 14.6% dos usuários Premium.
- Como visto na questão 2, apesar do pico de *churn rate* em Nov/2017, a quantidade de assinaturas aumentou consideravelmente e compensou no resultado do LTV.



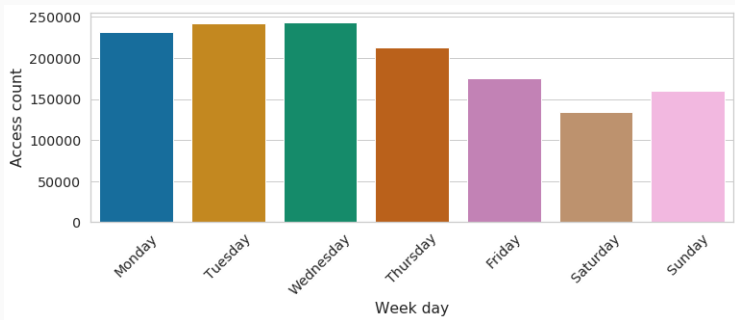
Área de Marketing do PD

- Decidi por analisar os melhores momentos para lançar campanhas de marketing/promoções de acordo com o sazonalidade dos acessos na plataforma.
- Os acessos acompanham bastante a sazonalidade do período letivo, com aumento dos acessos conforme a época de provas se aproxima (antes das férias de Julho e no fim do ano).
- Novembro aparece com um valor bem maior do que o resto devido ao pico de acessos em Nov/2017.



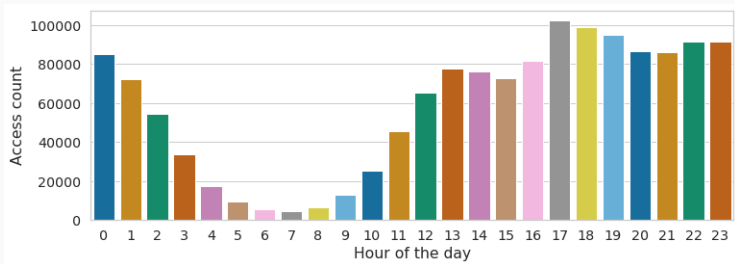
Área de Marketing do PD

- Os dias da semana com maior atividade na plataforma são terça-feira e quarta-feira.
- Campanhas via e-mail, por exemplo, poderiam se concentrar de segunda a quarta ou de segunda a quinta.



Área de Marketing do PD

- O horário de pico é entre 17h e 19h. O quantidade de acessos permanece alta até em torno de 0h.
- Caso decidam fazer uma live ou uma promoção relâmpago, por exemplo, a identificação desses horários de pico é bastante relevante.



Conclusão

Conclusão

- Sem dúvida, há muitas outras informações e *insights* que poderiam ser extraídos desses dados.
- Senti um pouco de dificuldade na parte financeira por nunca ter trabalhado com um negócio baseado em assinaturas.
- Precisaria de um contato maior com a plataforma para entender melhor as dores dos estudantes ao utilizá-la.
- Obrigado pela oportunidade de participar deste desafio.

Código-fonte em:

*[https://github.com/Lucas-Cerqueira/
passei-direto-challenge](https://github.com/Lucas-Cerqueira/passei-direto-challenge)*