



# Trabalho Final – Turma 11

Caso de Uso: iFood

24/06/2020

## Coordenadores:

Profª Drª Alessandra de Ávila Montini

Profª Dr. Adolpho Walter Pimazoni Canton

## GRUPO 15:

Ariel Vicente

Jaime Mishima

# Agenda

- 1. Objetivo do Trabalho
- 2. Contextualização do Problema
- 3. Base de Dados
  - i. Tratamento
  - ii. Principais variáveis
  - iii. Geração da ABT
- 4. Análise Exploratória de Dados
- 5. Desenvolvimento dos problemas propostos
- 6. Desafios encontrados



# 1. Objetivo do Trabalho

A partir de bases históricas do Ifood com informações de

- Pushes;
- Pedidos e receita;
- Segmentação de clientes; e
- Visitas.

O objetivo é responder às seguintes perguntas de negócio:

1. O que fizeram os clientes darem Churn?
2. Qual a importância dos eventos (push, compra, acesso, entre outros) ao longo da vida do cliente?
3. Prever o quanto um cliente vai gastar no mês seguinte.



## 2. Contextualização do Problema

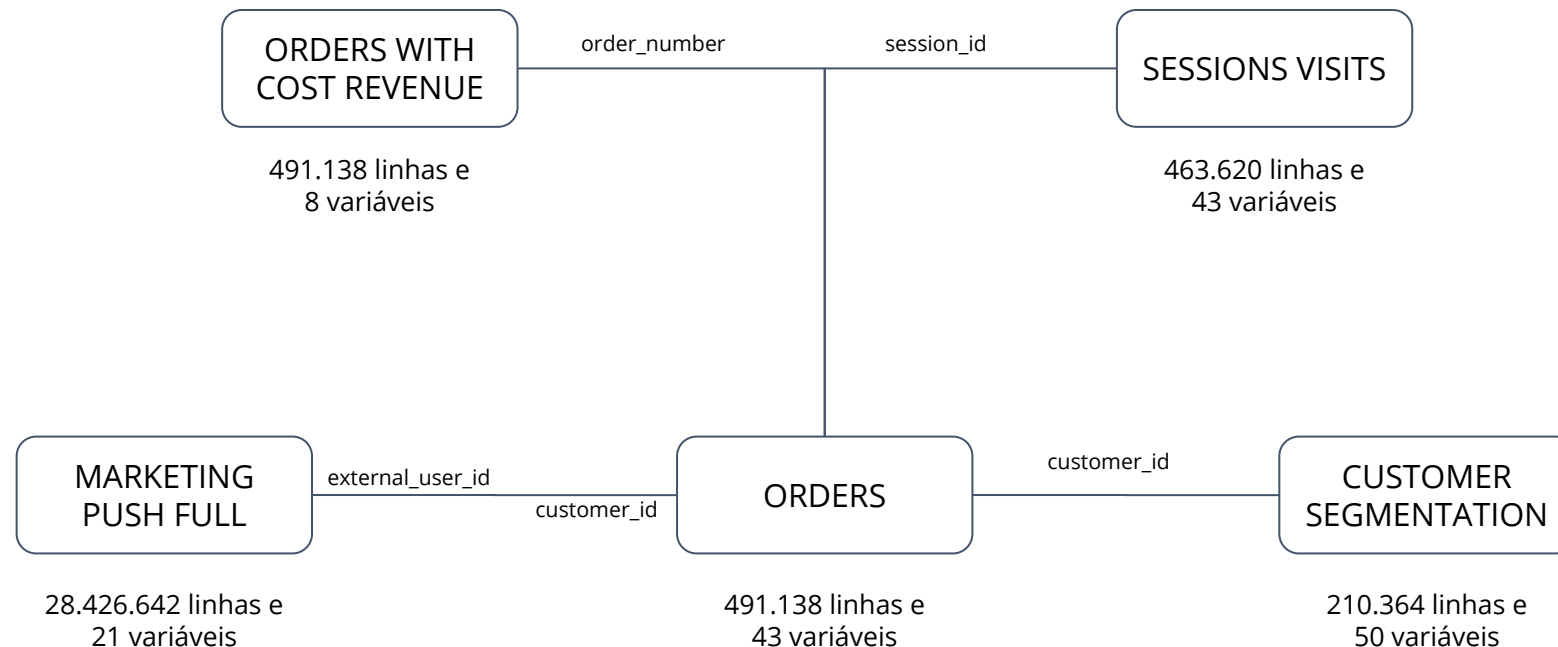
Os problemas a serem resolvidos são importantes pois permitem uma série de benefícios para as áreas de negócio:

- Em um cenário em que o CAC de um cliente costuma ser maior que o custo de retenção, temos os seguintes benefícios:
  - **Marketing:** elaboração de estratégias mais assertivas na comunicação, posicionamento de marca e índices de satisfação.
  - **Produtos:** encontrar os principais gaps nos serviços prestados ao cliente, identificar pontos de fricção na experiência do cliente.
  - **Financeiro:** Permitir alocação de budget, pivotar diferentes estratégias de crescimento (foco em crescimento, ganho em *share* de mercado).

### 3. Bases de Dados

5

Período das bases: Junho de 2019 a Dezembro de 2019.



## 3.i. Tratamento

### Marketing Push Full

- 28.4M de registros e 50 variáveis
- Variáveis desconsideradas:
  - 8 Variáveis do tipo canvas: +96% de missings;
  - *send\_id*: 100% de missing.
- Ao remover as colunas, removemos as linhas repetidas, resultando em 20.87M registros
- Tratamento de Missings:
  - Colunas categóricas com missing abaixo de 22%: imputação de unknown.
    - *campaign\_name* e *campaign\_id* (16.3% de missing);
    - *message\_variation\_channel* (21.3% de missing);
    - *platform* (0.55% de missing).

### Customer Segmentation

- 210.3k registros e 50 variáveis.



## 3.i. Tratamento

### Orders e Orders com custo/receita

- Chave para junção das bases de pedidos e pedidos com custo e receita: order\_number;
- 492.1k registros e 47 variáveis;
- Tratamento de Missings:
  - 8 Colunas categóricas com missing: imputação de unknown;
  - 9 Colunas numéricas com missing: imputação da média.

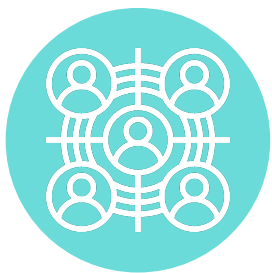
### Visits

- 463.6k registros e 43 variáveis;
- Tratamento de Missings:
  - 6 Colunas categóricas com missing: imputação de unknown;
  - 8 Colunas numéricas com missing: imputação da média.



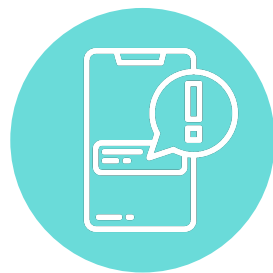


## 3.ii. Principais variáveis



### VARIÁVEIS DE SEGMENTAÇÃO

- **marlin\_tag**: classificação ordinal qualitativa dos clientes
- **ifood\_status**: Status do cliente no mês vigente
- **orders\_last\_91d**: Contagem de pedidos concluídos nos últimos 91 dias
- **last\_nps**: Última avaliação de NPS no app
- **recency\_days**: Diferença em dias entre hoje e a última compra do usuário
- **benefits\_sensitivity**: Índice de sensibilidade a benefícios.



### VARIÁVEIS DE PUSHES

- data de envio do push
- horário de envio do push
- **plataforma**: plataforma do device (iOS/Android/Desktop)
- **campanha**.



### VARIÁVEIS DE PEDIDOS

- **sum\_order\_total**: valor total do pedido (pedido + taxa de entrega)
- **paid\_amount**: valor pago no pedido pelo usuário
- **credit** (valor desconto): valor total de desconto no pedido
- **order\_shift**: Segmentação do período em que foi feito o pedido.



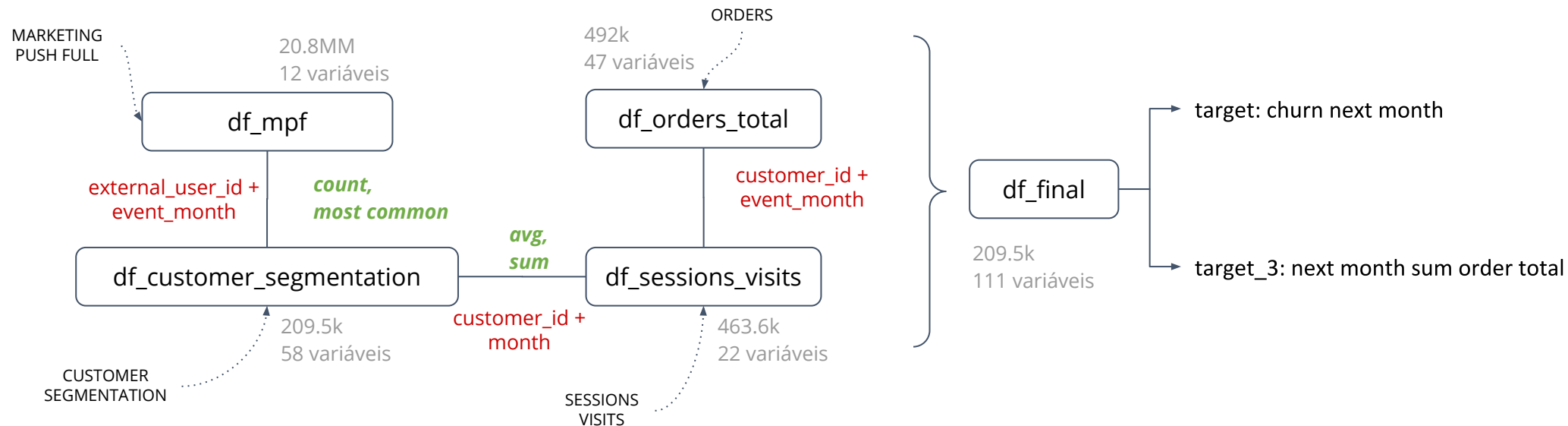
### VARIÁVEIS COMPORTAMENTAIS

- horário de início de sessão
- **device\_model**/manufacturer
- **sum\_view\_checkout**: Quantidade de eventos de visualizações do checkout
- **sum\_event\_open**: Quantidade de eventos de abertura do Aplicativo durante a sessão
- **sum\_view\_restaurant\_screen**: Quantidade de eventos de visualizações da tela do restaurante





### 3.iii. Geração ABT



#### Legenda:

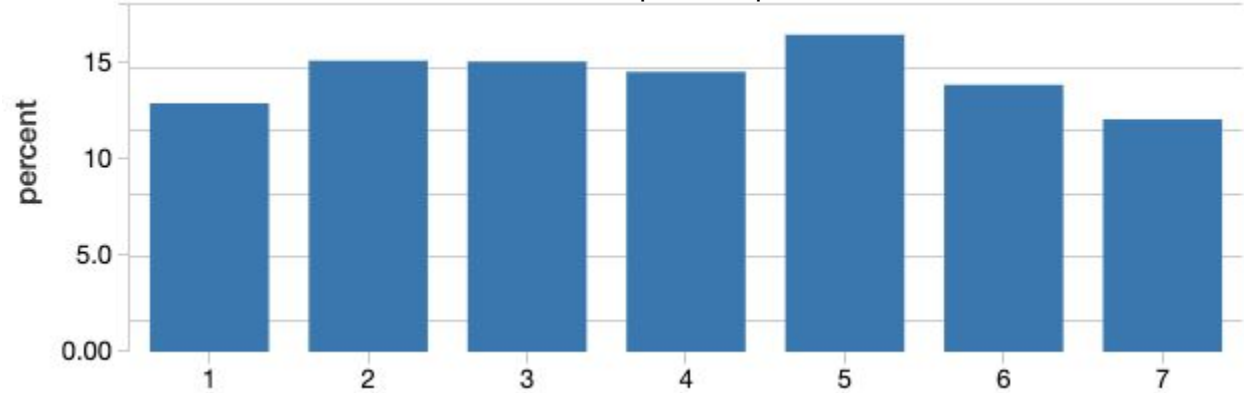
chave join  
 tipo agregação  
 tamanho base



# 4. Análise Exploratória de Dados

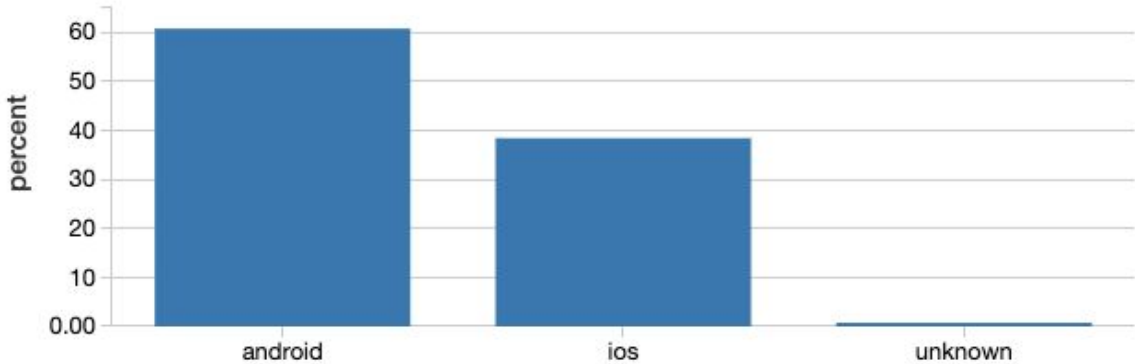
## Marketing Push Full

Percentual de envio de pushes por dia da semana



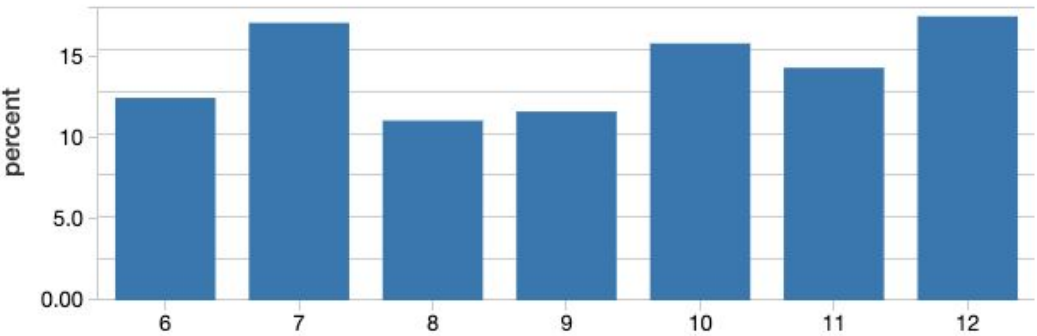
Finais de semana (1 e 7) tem menor volume de pushes . Quintas (5) tiveram o maior volume de envios (16.4% da base)

Share por Plataforma



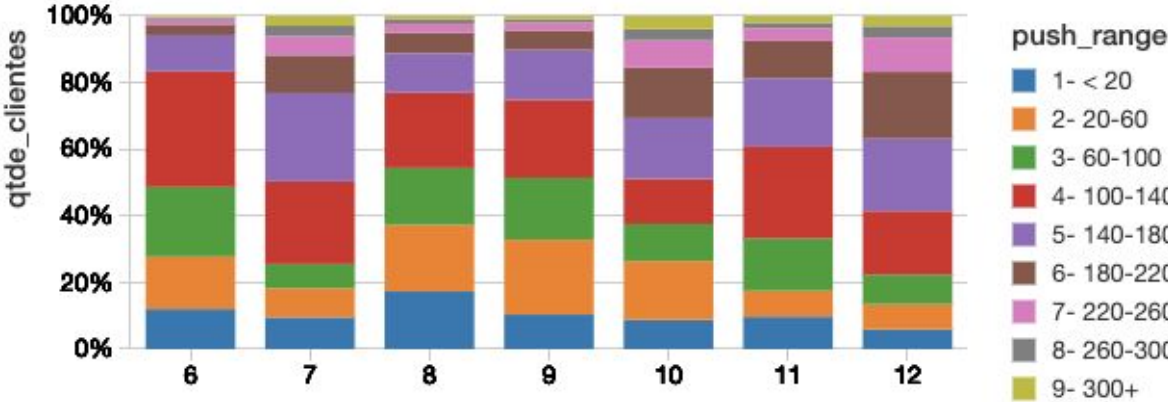
60.8% dos pushes foram para Android, 38.4% para ios. Unknowns representaram apenas 0.75%

Distribuição Mensal de Pushes



Agosto (8) foi o mês com menor volume de pushes recebidos (11.1% da base) e dezembro foi o mês com maior volume (17.5% da base)

Percentual de clientes vs Volume Mensal de Pushes

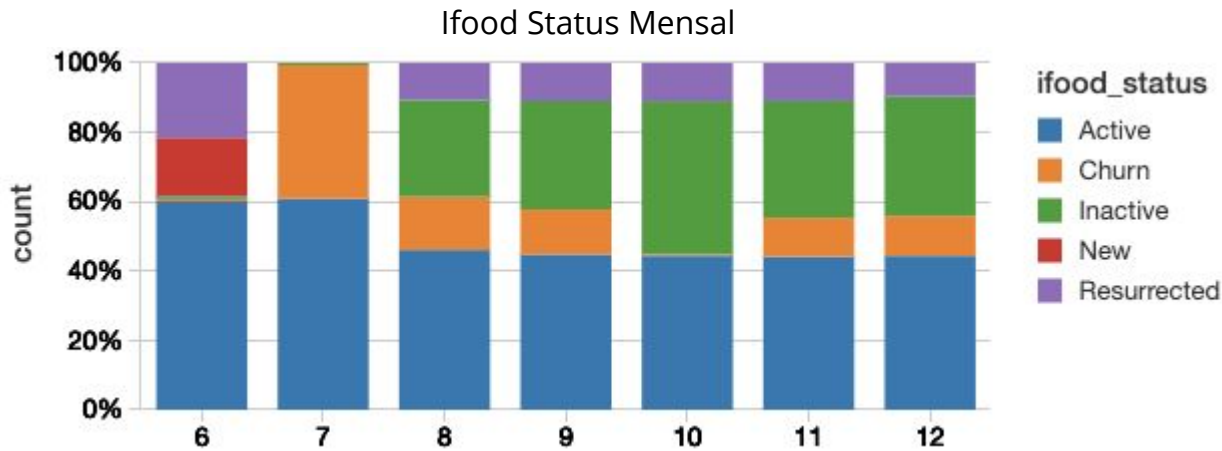


De novembro em diante (11 e 12), o share de clientes recebendo mais de 60 pushes aumentou.

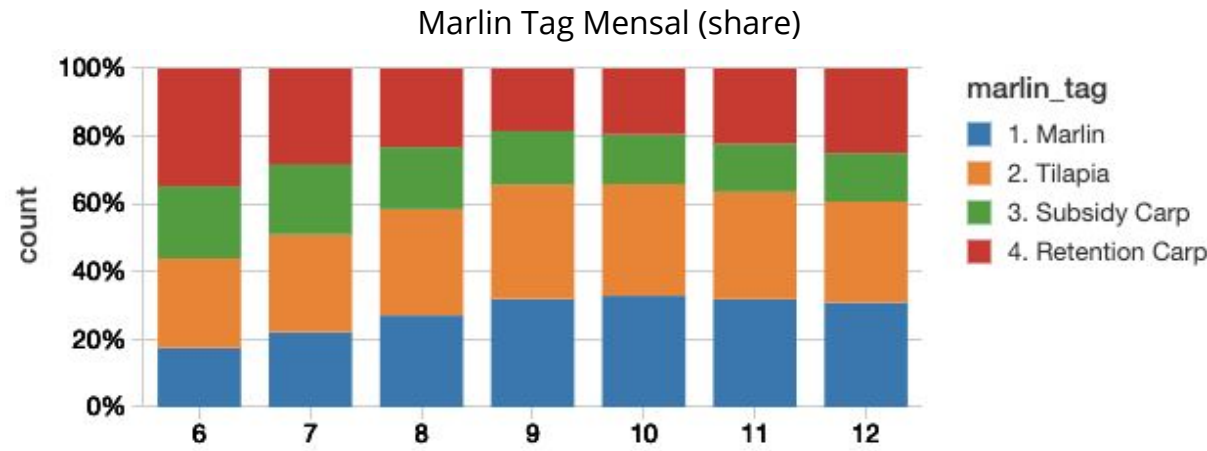
+10.3M de pushes foram enviados. Destes 98.3% foram recebidos. Apenas 3.16% foram clicados e bounce da base foi de 4.4%.

# 4. Análise Exploratória de Dados

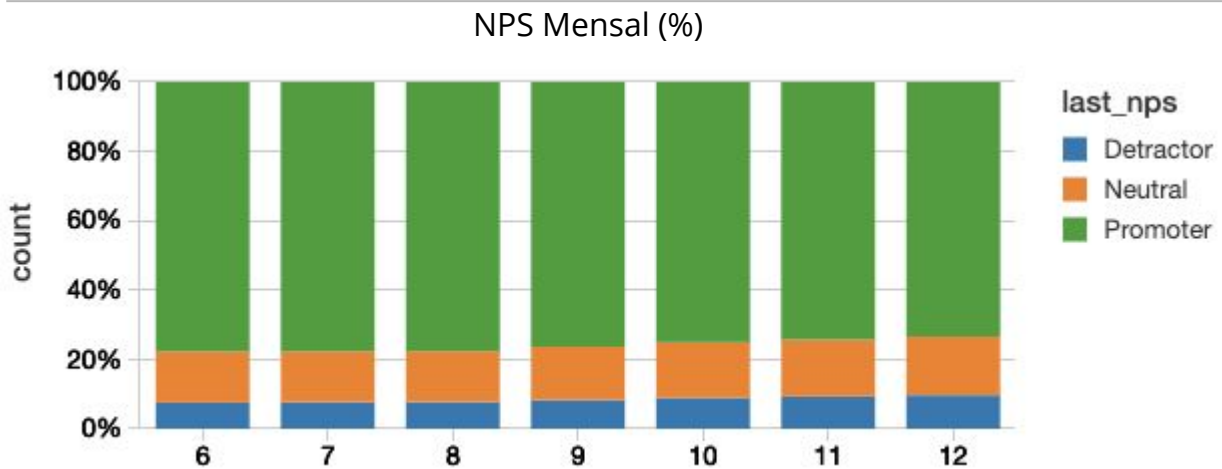
## Customer Segmentation



Outubro é o mês com a maior proporção de clientes Inativos. Em compensação, a proporção de Churn também se mostra como a menor desde Junho.

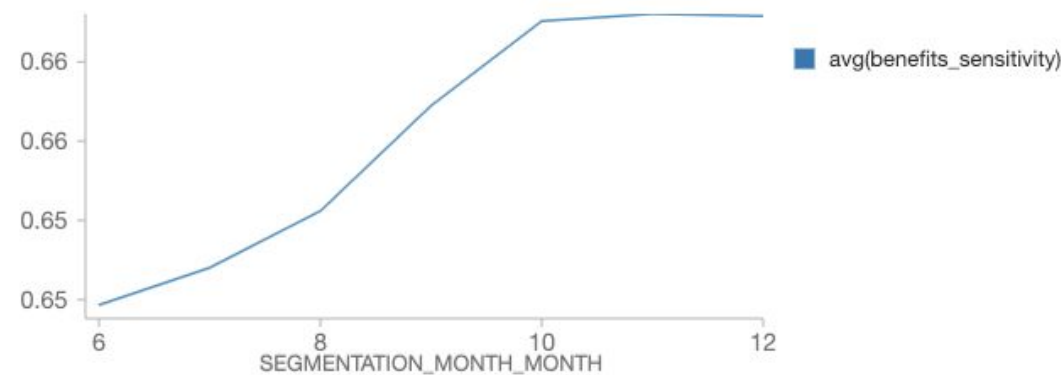


Ser classificado como Marlin significa estar entre os melhores. Vê-se que a proporção de Marlins aumenta ao longo dos meses.



A proporção de Promotores permanece maior que 75% em todos os meses, mas a proporção de Neutros e Detratores cresce ao longo do tempo.

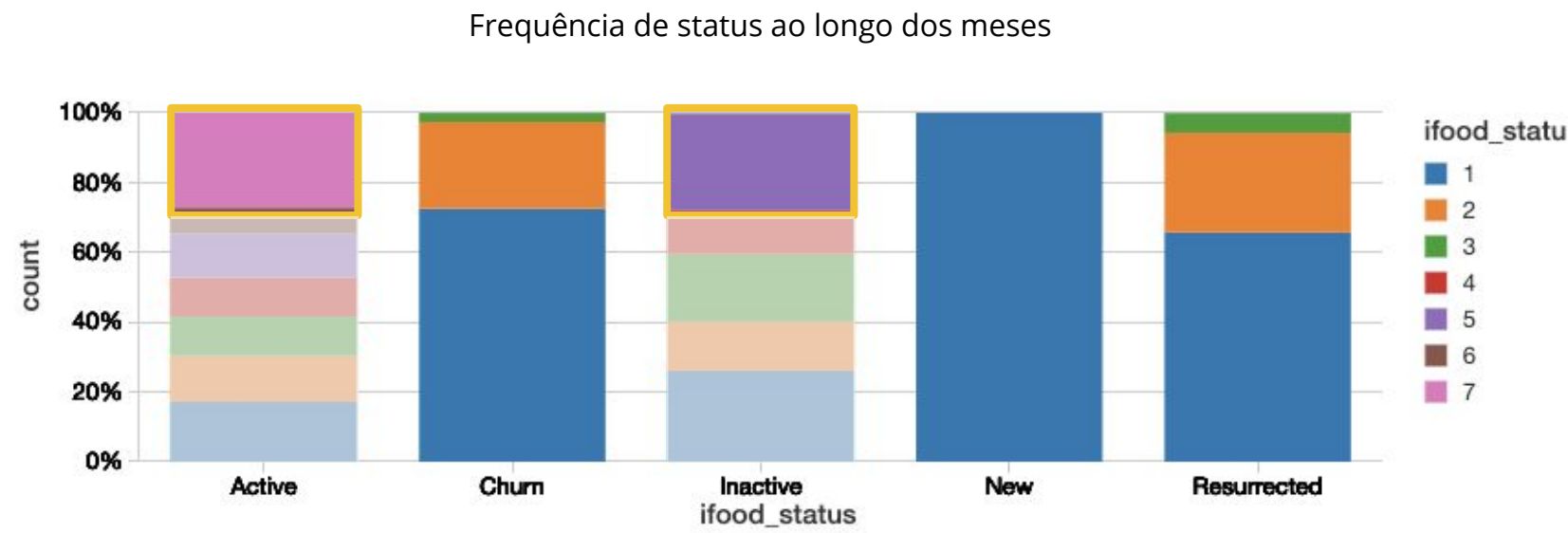
## Índice Médio Mensal de Sensibilidade à Benecífios



O índice médio de sensibilidade a benefícios tem curva crescente para todos os clientes da base.

# 4. Análise Exploratória de Dados

## Customer Segmentation



Vê-se que, aproximadamente **30% (6680) das pessoas** que tiveram status Ativo ao menos uma vez no período, na verdade, permaneceram ativas durante os 7 meses que se tem de informação disponível.

Entre os que em algum momento tiveram status Inativo, **30% (4751 clientes) foram inativos por 5 meses.**

# 4. Análise Exploratória de Dados

## Orders

- +30k clientes fizeram pedidos no período analisado. Foram feitos +492k pedidos, a média de pedidos/cliente foi de 16.4;
- 97% dos pedidos são realizados via celular, sendo 55% dos pedidos feitos usando um sistema Android e 42% em IOS: **A penetração de mercado em Iphone é elevada;**
- Apenas 0.33% dos pedidos foram retirados nos restaurantes;
- 98% dos pedidos não são agendados;
- 32% dos pedidos foram feitos com promoção. A média por pedido foi de R\$55.8, com um desconto médio de R\$6.25/pedido. O lucro médio por pedido foi de R\$7.39 (14.7% do valor pago).

## Visits

A base contém +463 sessões.

Sextas e finais de semana são os dias com mais sessões geradas, 16% e 34% respectivamente.



## - PROBLEMA 1 -

O que fizeram os clientes darem Churn?

1

## Abordagem adotada:

Para o problema proposto, optou-se por criar um modelo preditivo usando um classificador.

Ajustar 5 técnicas de modelagem (Regressão Logística, Decision Tree, Random Forest, Gradient Boosted Tree, Naive Bayes), selecionar os dois melhores modelos.

Usar Random Forest como método de seleção de variáveis e aplicar hiperparâmetros e cross validation para os dois melhores modelos.

2

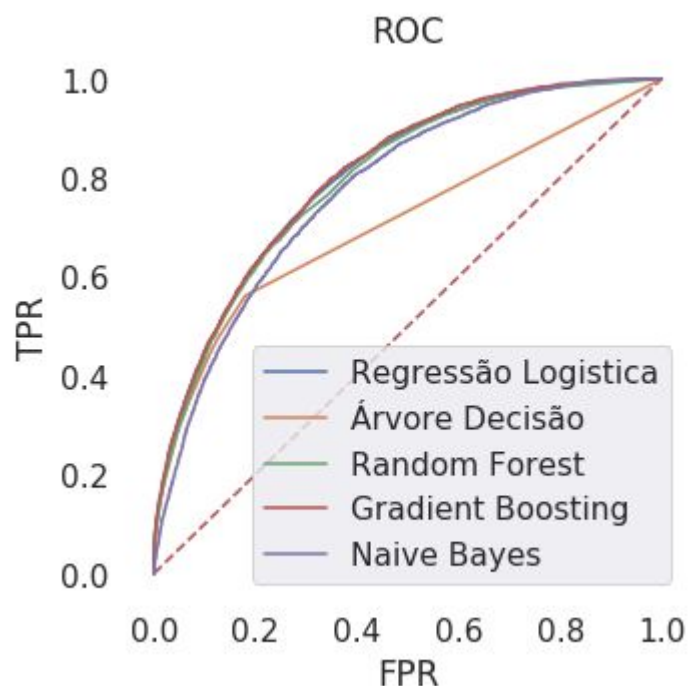
## Estratégia:

Baseado no melhor modelo otimizado, usar as importâncias das features como forma de interpretar fatores que influenciam no Churn.





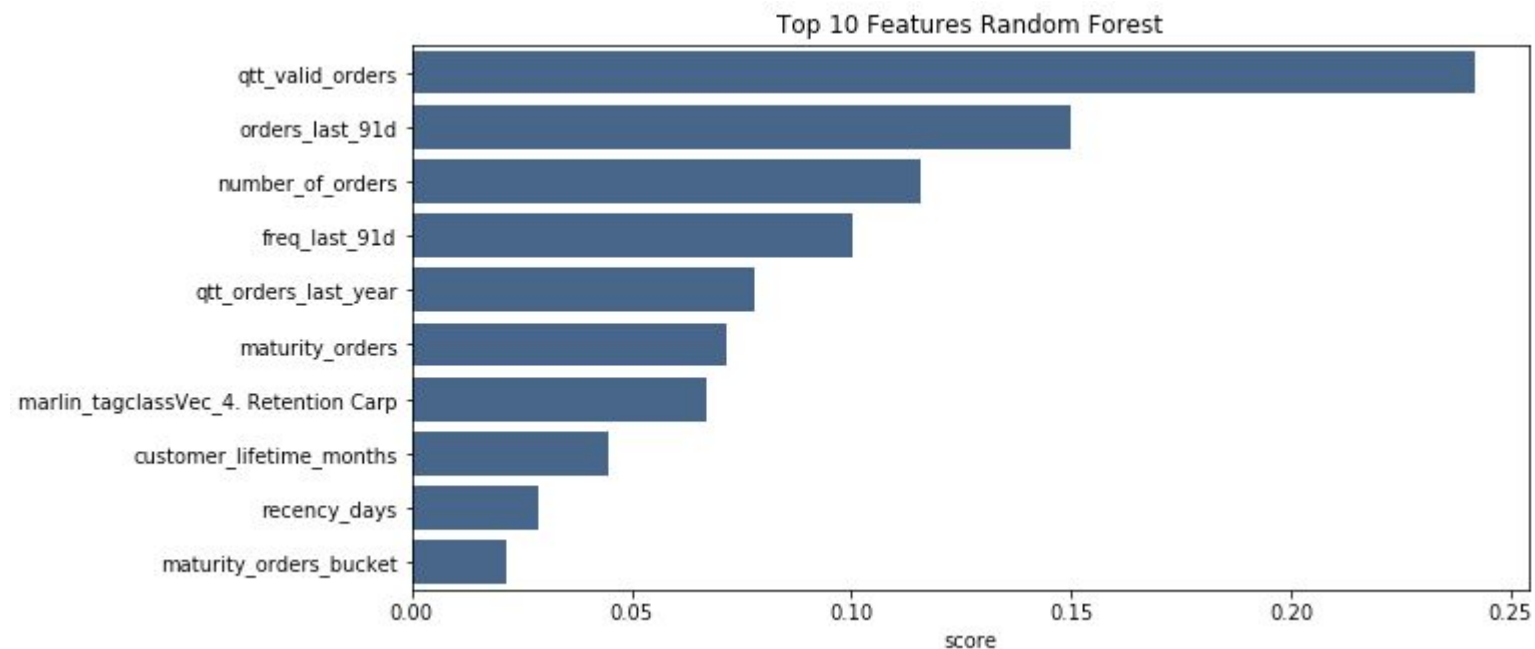
Foram utilizados **cinco modelos para prever o Churn** no mês seguinte: *Regressão Logística*, *Árvore de Decisão*, *Random Forest*, *Gradient Boosted Tree* e *Naive Bayes*. Abaixo temos as respectivas curvas ROC e na tabela as áreas sobre a curva ROC.



Modelo	AUC ROC
Regressão Logística	0.8062
Árvore de Decisão	0.7075
Random Forest	0.7960
Gradient Boosted Tree	0.8048
Naive Bayes	0.7772

Baseados nas áreas sobre a curva ROC, temos que **os dois melhores modelos são: *Gradient Boosted Tree* (AUC ROC 0.8048) e *Regressão Logística* (0.8062).**

**Random Forest** foi o método de **seleção de variáveis** utilizado. Baseado nas principais features, utilizamos o **VectorSlice** para selecionar as principais variáveis.

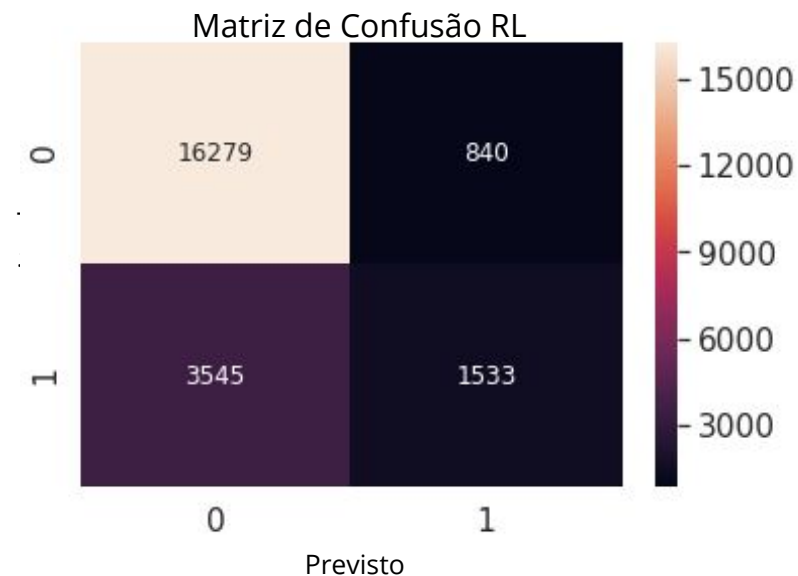
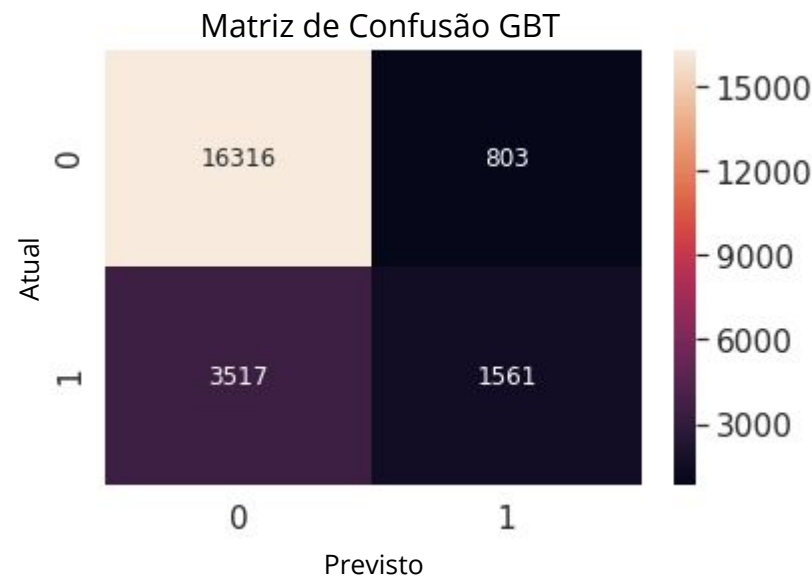
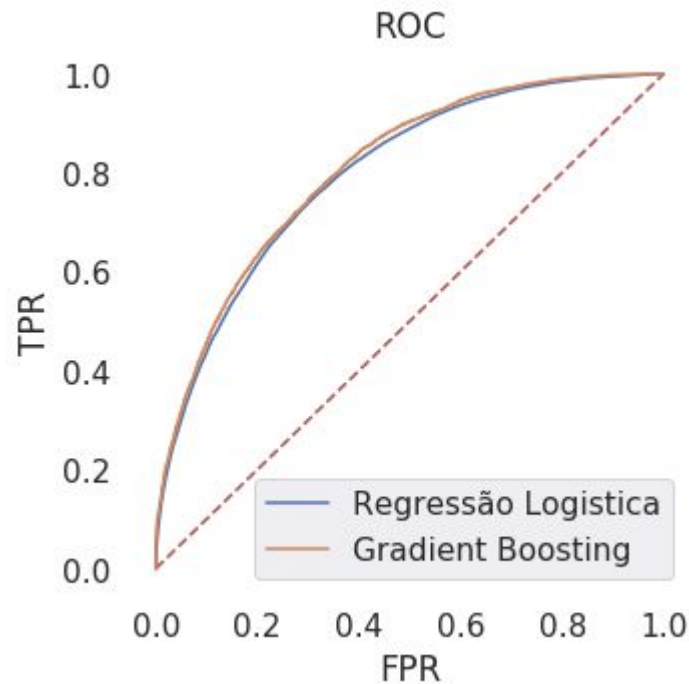


# Modelagem

HIPERPARÂMETROS & CROSS VALIDATION | PROBLEMA 1

18

Os melhores parâmetros para a Regressão Logística (RL) foram: maxIter = 60 e regParam = 0.01. Em termos de área sobre a curva ROC o melhor modelo foi o Gradient Boosted Tree (GBT), com AUC ROC = 0.8074, com: maxDepth = 6, maxBins = 30 e maxIter = 15. Ambos usaram 5 folds no CV.

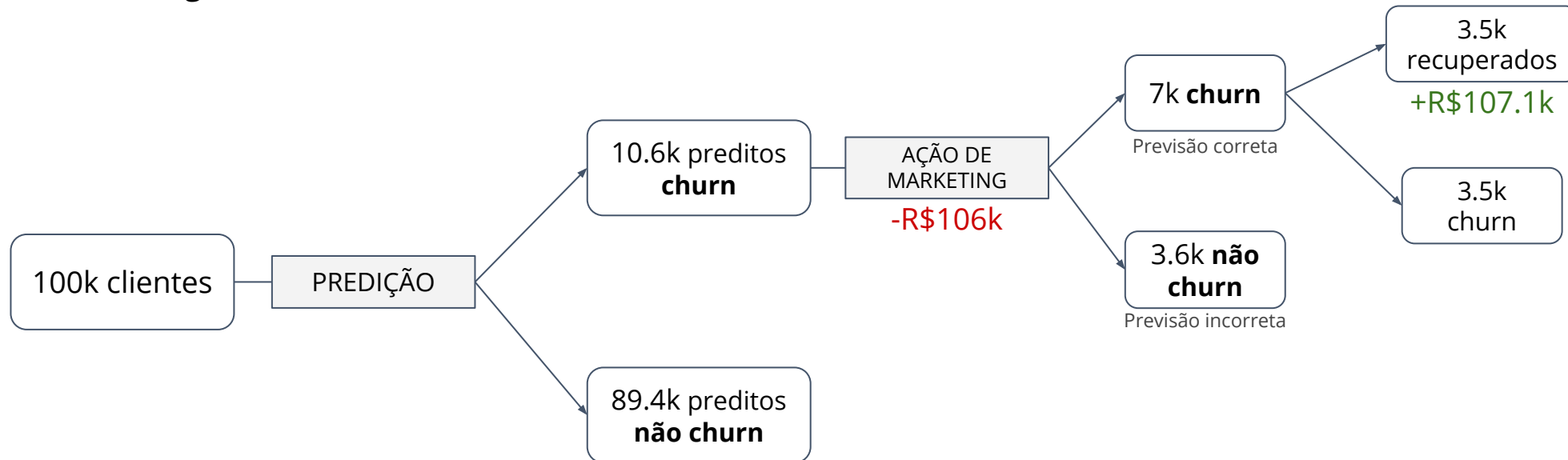


Modelo	AUC ROC
Regressão Logística	0.8025
Gradient Boosted Tree	0.8074

Modelo	TPR (%)	FPR (%)	Acurácia (%)
Regressão Logística	30.19	4.91	80.25
Gradient Boosted Tree	30.74	4.69	80.54

Qual o impacto para o  
negócio?

- Efetividade dos resultados:
  - **23% da base de treino deu churn** no mês seguinte e, desses, **o GBT foi capaz de prever 7%** (acertando 30.7% das pessoas que deram churn [TPR]). Com isso, *GBT* obteve resultados 0.55p.p maiores que a *RL*;
  - O *GBT* ainda identificou erroneamente 4.7% das pessoas como sendo churn (FPR), mas mesmo assim acabou sendo 0.22p.p menor que a *RL*.
- Impacto do Churn no Lucro:
  - Cenário: custo de marketing para evitar um churn igual a R\$10; efetividade da ação é de 50%; um churn equivale a deixar de arrecadar em média R\$208; percentual de lucro igual a 14.7%;
  - Dado o cenário, **o lucro perdido por churn é igual a R\$30.6**. Com isso, uma ação para uma base de 100k clientes geraria (em um mês):

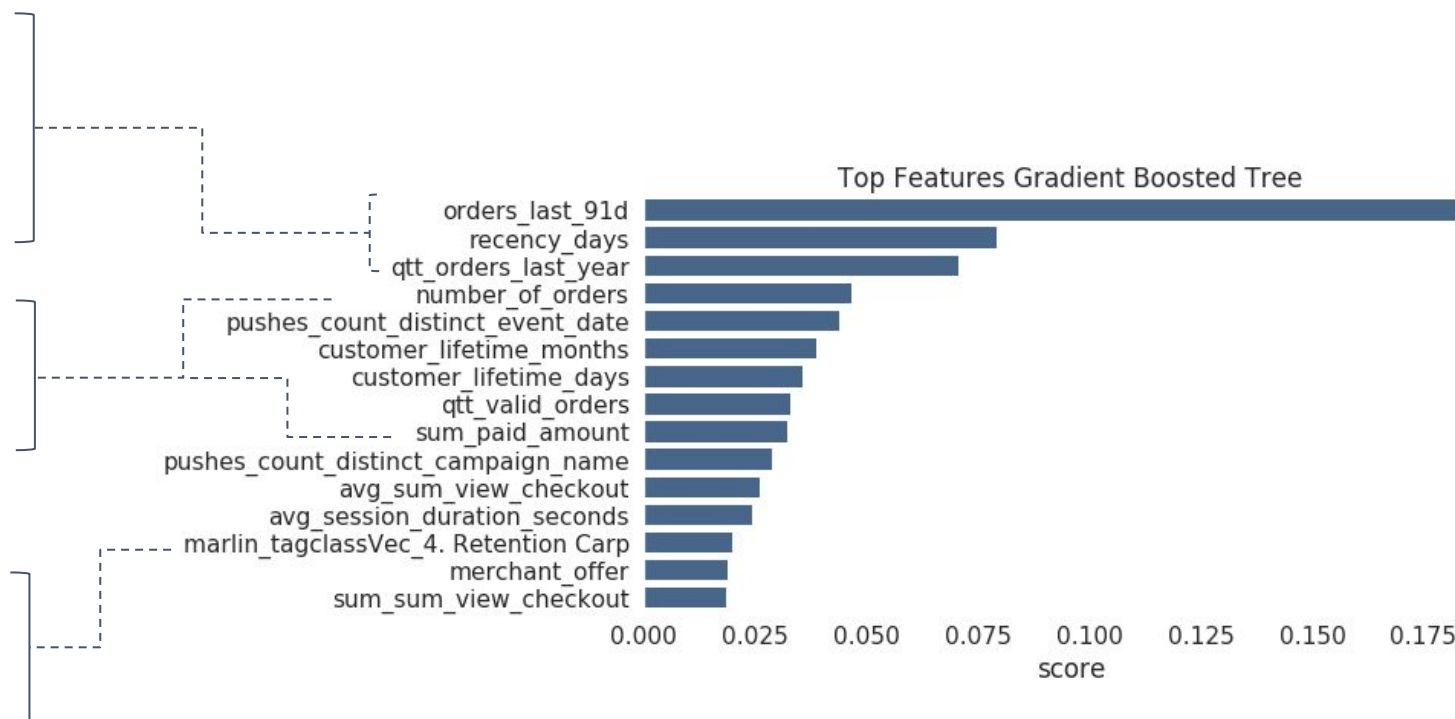


Baseado no melhor modelo, ordenando as importâncias das variáveis, obtivemos:

**Contexto de sazonalidade.** Por exemplo, como *recency\_days* é a diferença em dias entre hoje e a última compra do usuário, faz sentido pensar que se alguém tem poucas compras nos últimos 91 dias o indivíduo demonstra estar menos engajado com a plataforma do que alguém com muitos pedidos nesse mesmo período.

Variáveis quantitativas relacionadas ao **comportamento de compra** no mês também são relevantes para prever se um cliente vira churn no mês seguinte.

Variáveis qualitativas são bem relevantes para o modelo. Mais uma vez, conforme esperado intuitivamente, ser um cliente classificado como **Retention Carp** é relevante para se dar churn



## - PROBLEMA 2 -

Qual a importância dos eventos (push, compra, acesso, entre outros) ao longo da vida do cliente?



1

## Conceito de Importância:

*Pressuposto:* o Ifood deseja que seus clientes permaneçam ativos e comprando e gastando cada vez mais, aumentando assim seu lucro.

*Definição:* Define-se como a importância de um evento o quanto ele demonstra ter colaborado para que o Ifood atinja seus objetivos de negócio com seus clientes ao longo de suas vidas.

2

## Estratégia:

- Aproveitar insights da exploração da Entrega 1;
- Explorar a relação entre os eventos ao longo do tempo através de visualizações visando a geração de mais insights;
- Definir hipóteses sobre a importância dos eventos;
- Buscar por evidências que auxiliem na conexão do encontrado com a pergunta em si.

**Observação:** está fora do escopo a aplicação de técnicas matemáticas/estatísticas em busca da validação de insights ou hipóteses trazidos.

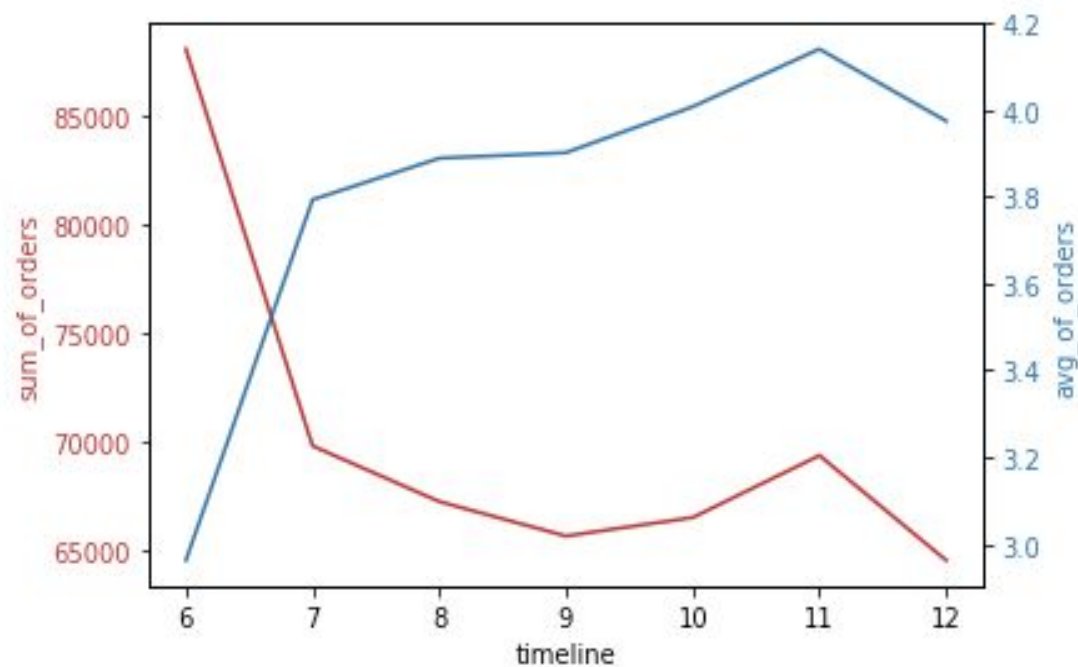


# Exploração dos Dados

ANÁLISE DESCRITIVA | PROBLEMA 2

24

O **total de pedidos por mês está caindo** com o passar dos meses. Porém, a quantidade média de pedidos por usuário aumentou. Ou seja, com o passar do tempo **os clientes que estão comprando, estão comprando mais**.

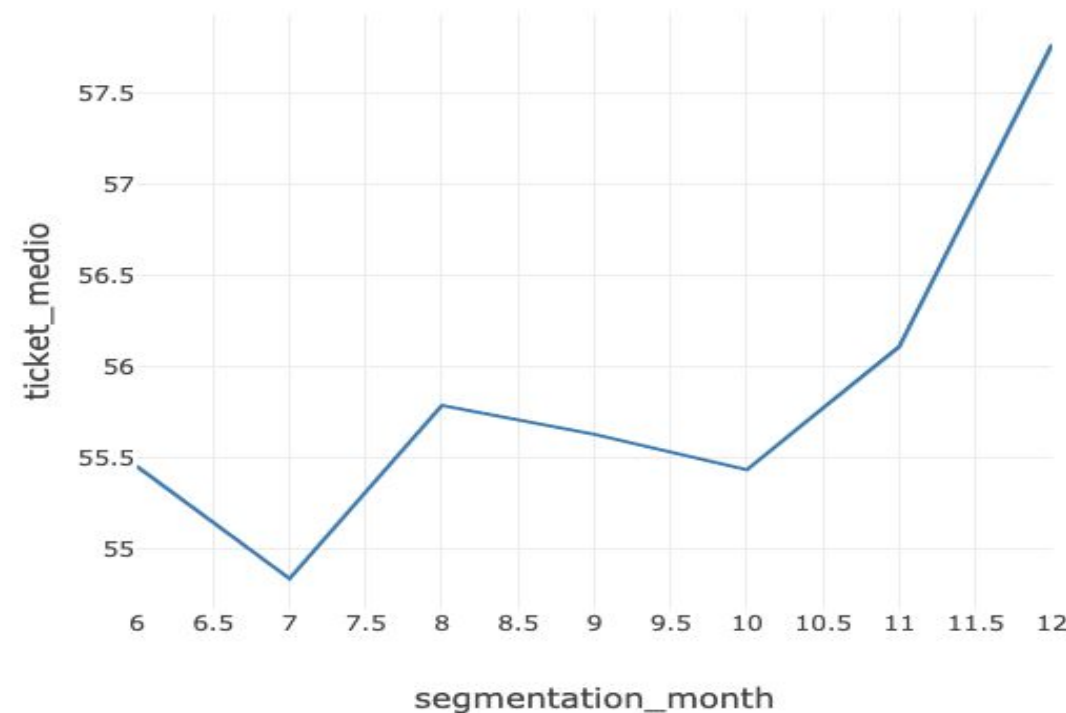
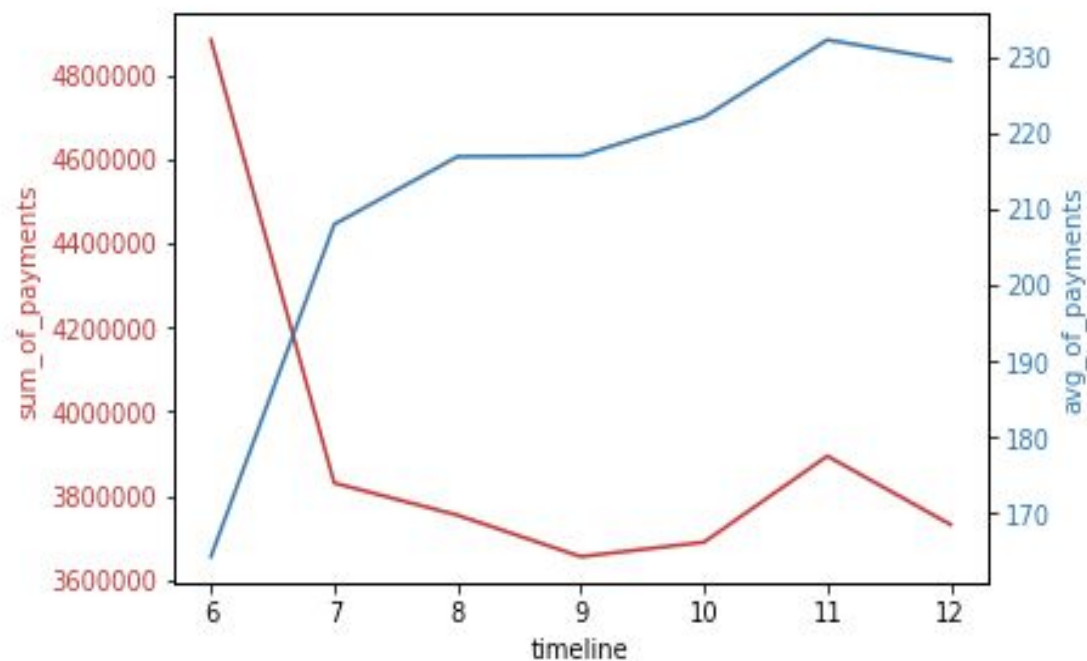


# Exploração dos Dados

ANÁLISE DESCRITIVA | PROBLEMA 2

25

Apesar do total de *payments* (receita, do ponto de vista do Ifood) e o montante gasto cair com o passar dos meses, **a receita mensal média por usuário aumentou**. Porém, **o ticket médio se manteve praticamente constante** (máx R\$57.7 / min R\$54.8). Ou seja, esse **aumento de receita média ocorre em decorrência do aumento de pedidos médios** do usuário no mês.



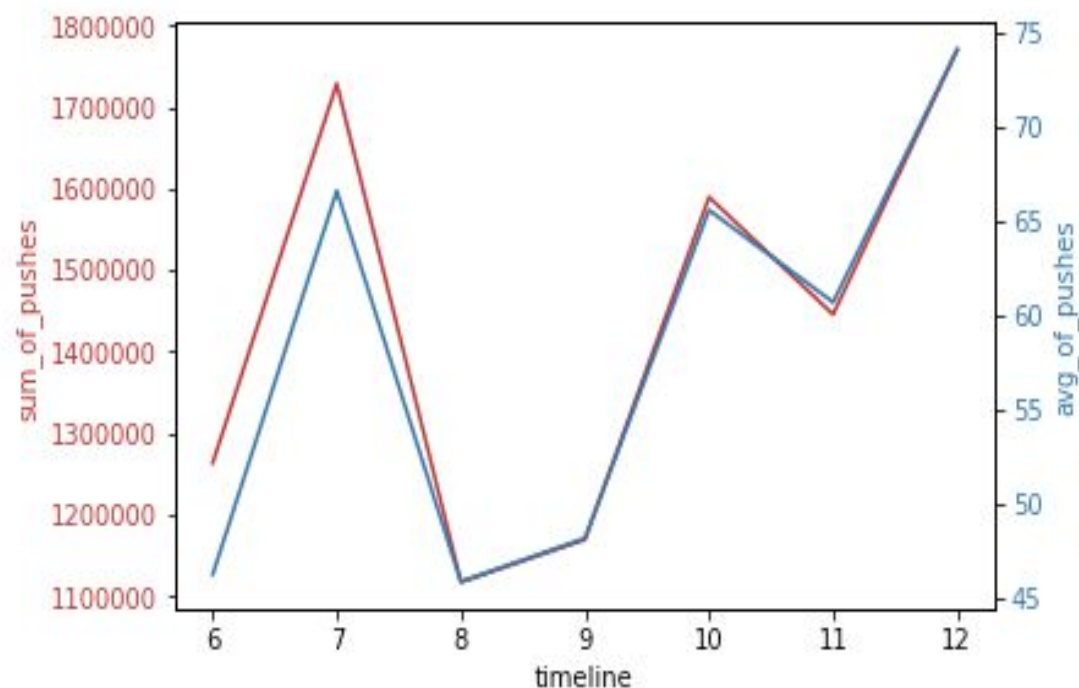
# Exploração dos Dados

ANÁLISE DESCRITIVA | PROBLEMA 2

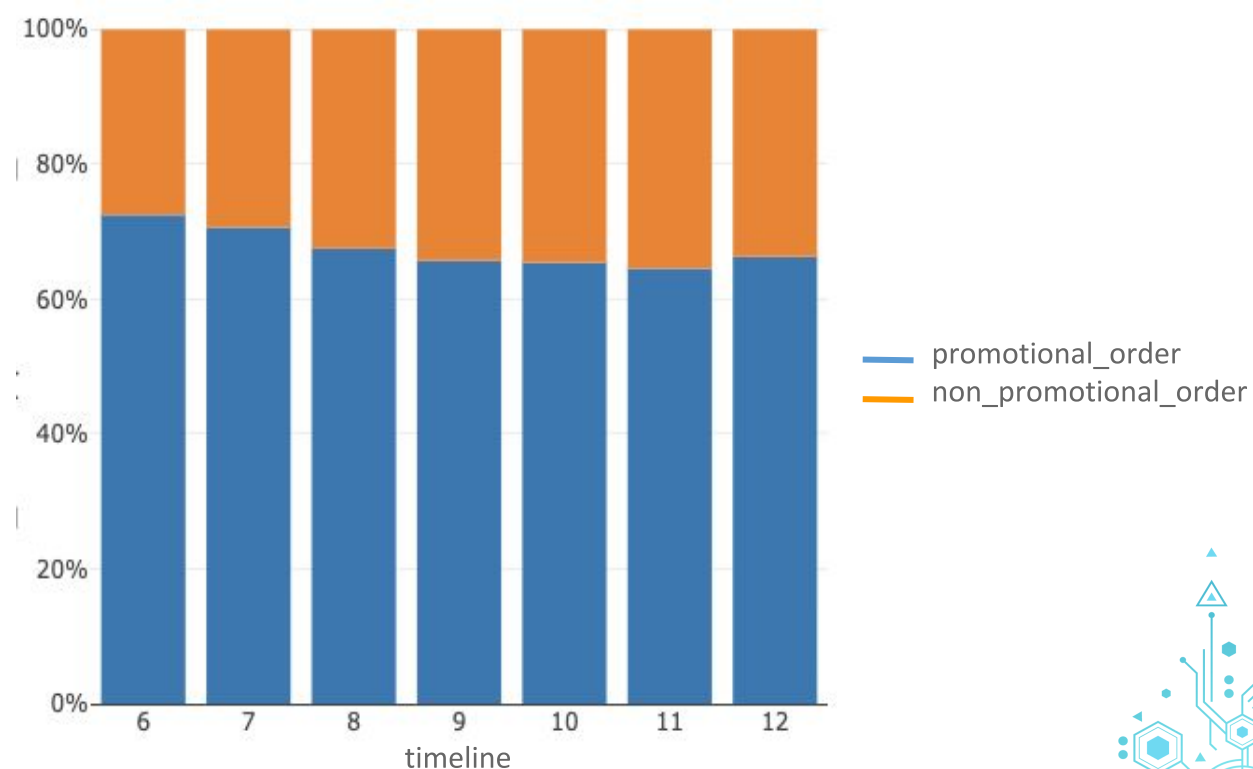
26

Na perspectiva de incentivo ao engajamento, **os pushes e as campanhas** mensais (valores totais e médios) também **estão aumentando** com o passar dos meses. Mais ainda, **os clientes estão recebendo mais promoções e incentivos**.

Total e Média de Pushes no Tempo



Proporção de pedidos promocionais no Tempo



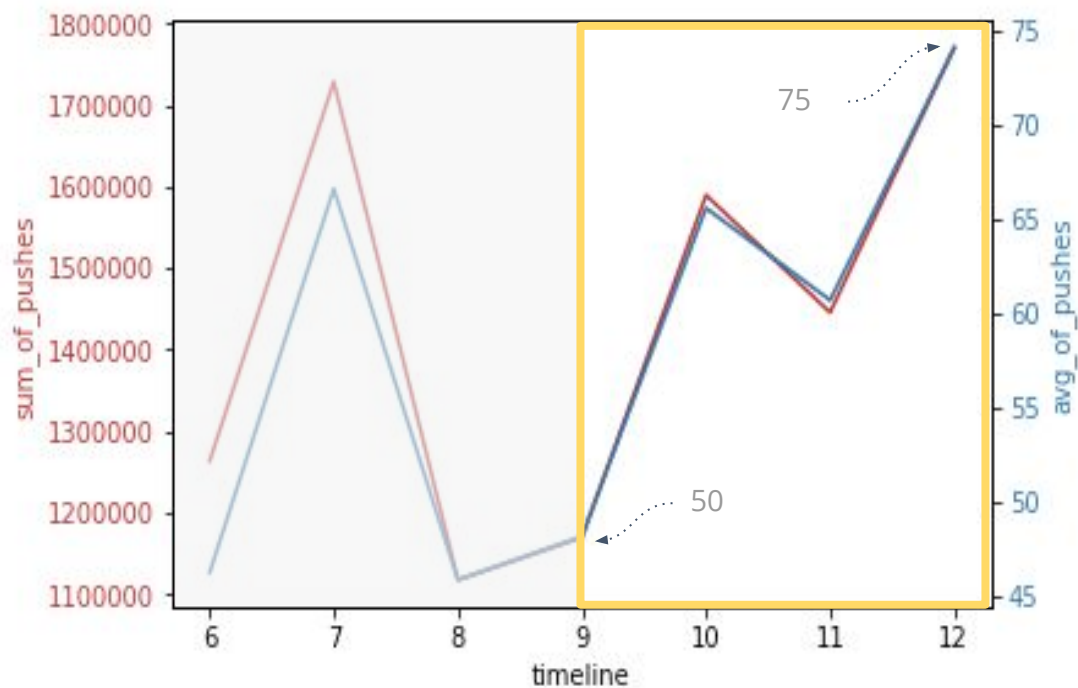
# Exploração dos Dados

ANÁLISE DESCRITIVA | PROBLEMA 2

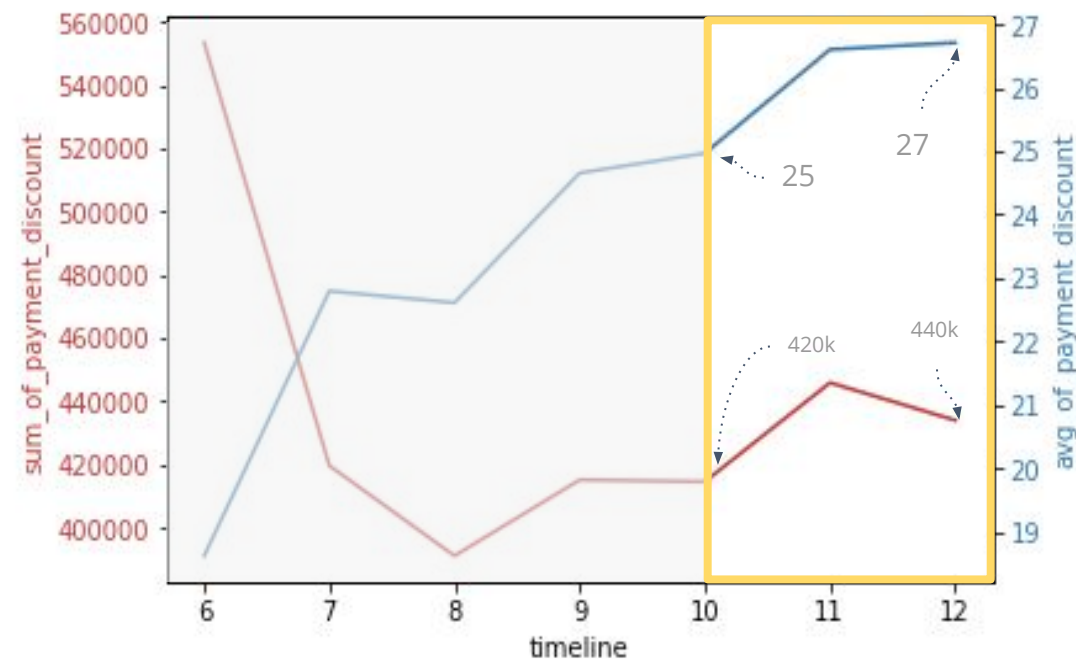
27

A partir de Outubro nota-se um **aumento de ~33% no total e na média de pushes mensais**. Além disso, em Novembro e Dezembro **há um aumento no total e na média de desconto por mês** (de ~R\$420k para ~R\$440k, ou ~R\$25 para quase R\$27 de desconto/mês).

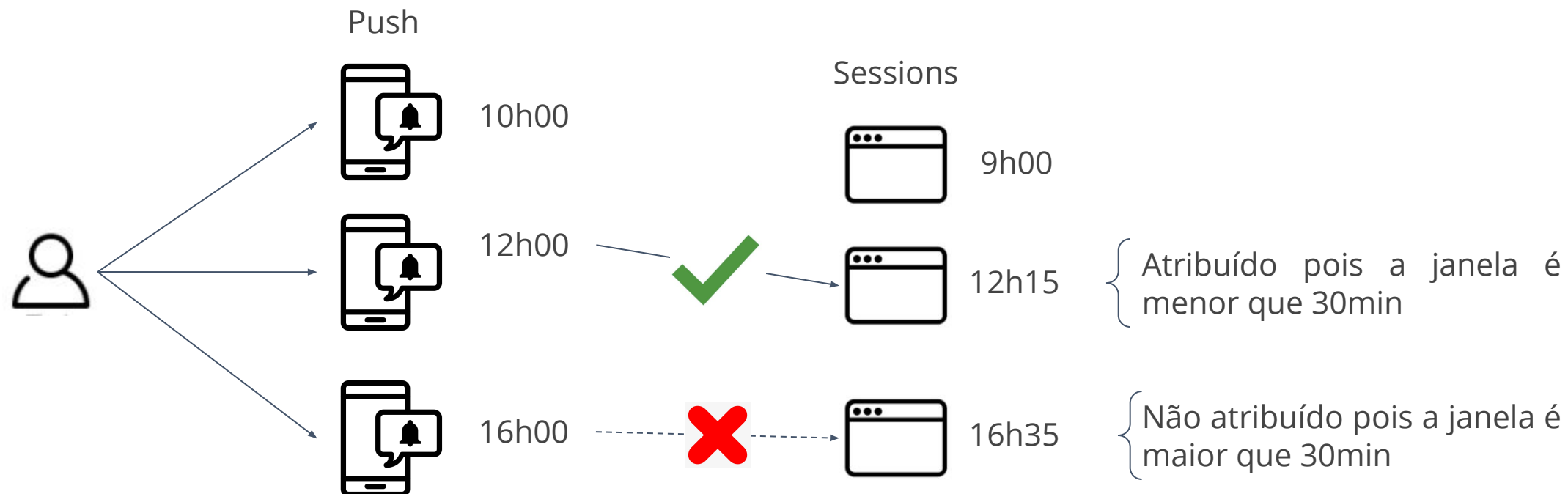
Total e Média de Pushes no Tempo



Total e Média de Pagamentos com desconto no Tempo



Para auxiliar com uma **medida para a efetividade dos pushes**, criou-se uma tabela que associa para todo push recebido um pedido/sessão em uma janela de até 30min (1800s). Por exemplo:



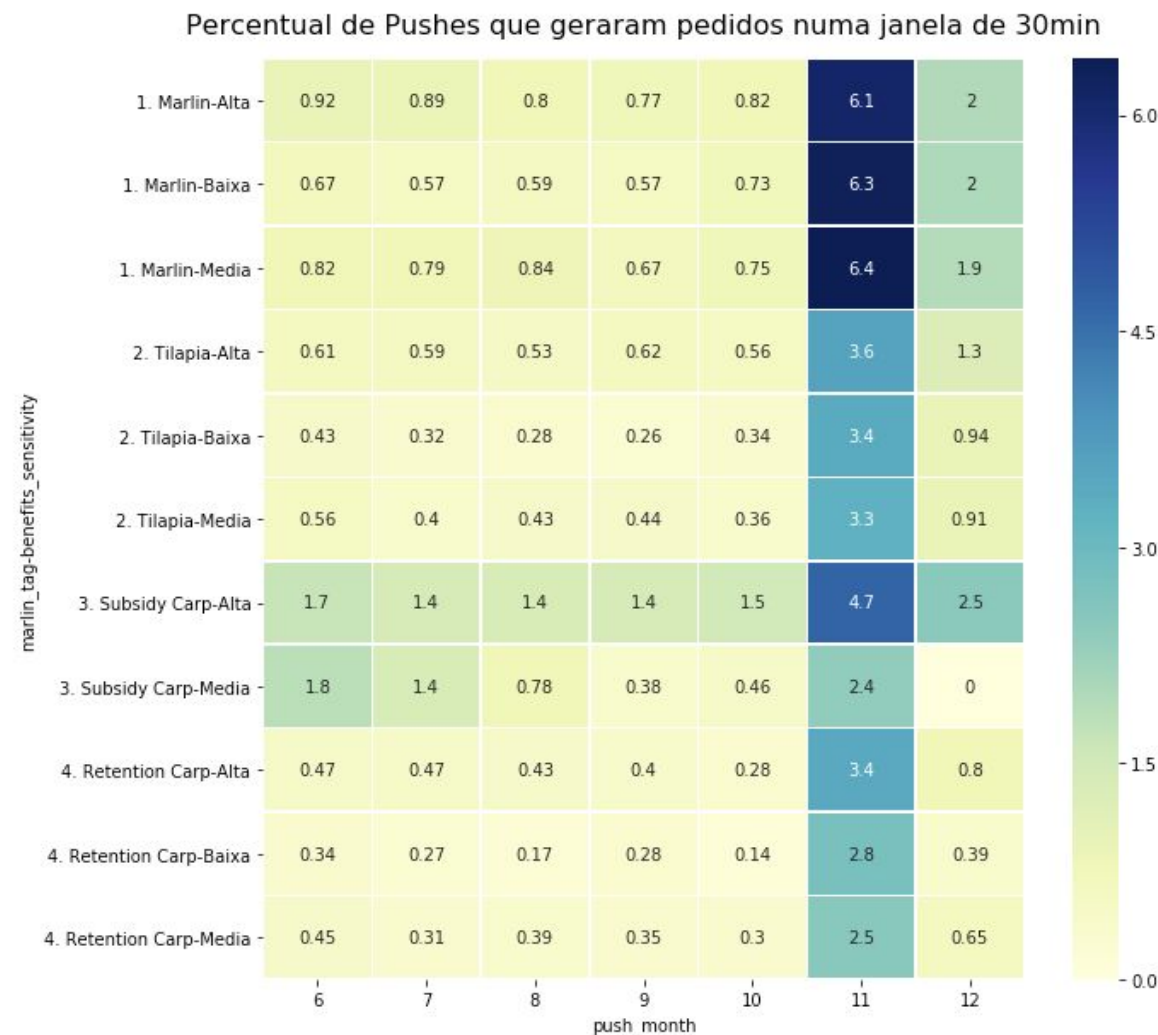
**Observação:** Como a base de sessões só traz registros associados a pedidos, não é possível extrair dos dados o contraponto disso. Por exemplo, o quanto o excesso de pushes causa de percepção negativa no usuário.

# Exploração dos Dados

ANÁLISE DESCRITIVA | PROBLEMA 2

29

Nota-se então que em Junho a taxa de efetividade dos pushes é maior, mas nos meses seguintes (entre Julho e Setembro) ela cai. A partir de Outubro há sinais de aumento, que se intensificam até Dezembro.





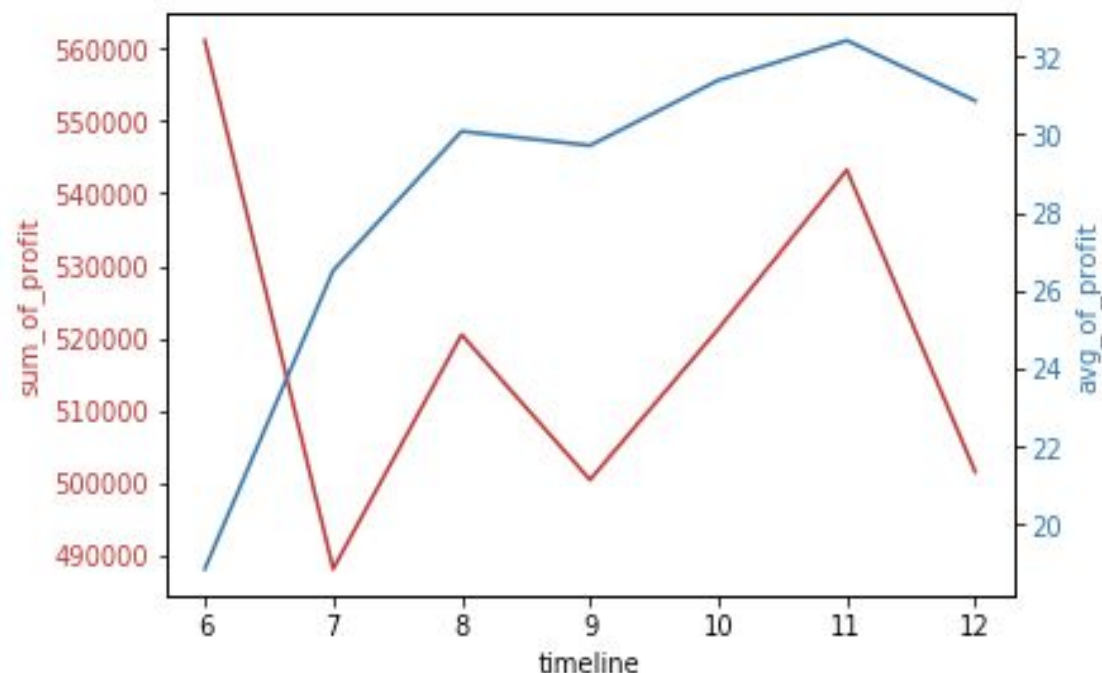
O lucro está sendo afetado?  
Como?

# Comportamento do Lucro

ANÁLISE DESCRITIVA | PROBLEMA 2

31

Em termos de lucro, há um **reflexo positivo em Novembro**, com uma média de lucro por cliente de ~R\$32. Já em **Dezembro**, com um mês atípico\* em vários aspectos, a curva de **lucro médio mensal por usuário se manteve** no patamar da de Novembro, na casa dos R\$30. Já o lucro total mensal teve queda substancial.



\* Os gráficos que ilustra isso está no [apêndice](#).

@2020 LABDATA FIA. Copyright all rights reserved.

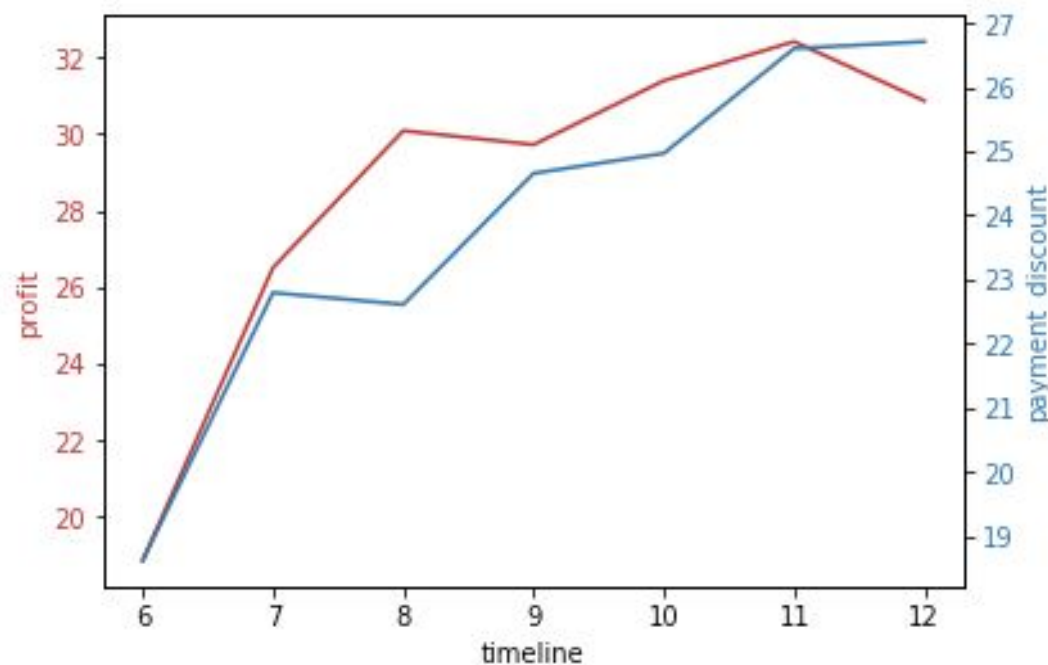


# Comportamento do Lucro

ANÁLISE DESCRITIVA | PROBLEMA 2

32

Ainda identificou-se uma relação interessante entre o lucro do Ifood e o quanto os cliente recebem de desconto. Falando em termos de médias e tomando como exemplo apenas o mês de Dezembro, vê-se que o lucro médio do Ifood por cliente (~R\$30) representou 13% da receita média por cliente (R\$230\*). Já os descontos médios recebidos por cliente (~R\$27) representaram 11.7% da mesma receita média por cliente.



\* O gráfico que ilustra tal valor já foi apresentado em [slide anterior](#).

@2020 LABDATA FIA. Copyright all rights reserved.

Sumarizando os resultados obtidos para a resolução do problema 2, tem-se sobre a importância dos eventos:

- **Push:**
  - Foi estratégia **bastante utilizada** pelo Ifood no período (média de ~60 pushes por usuário por mês);
  - Foi **efetivo**, principalmente entre os meses de Junho, Novembro e Dezembro;
  - Existem evidências para se acreditar que os pushes **estimularam a receita e o lucro** da empresa em meses específicos; porém,
  - Os descontos\* **representam, em termos de média, o mesmo percentual da receita** que o próprio lucro;
  - Contudo, **identificou-se um potencial trade-off** em relação a percepção de marca e serviço (clientes ficando insatisfeitos) e a perda de clientes ativos (churn) em decorrência do envio/recebimento de pushes.
- **Compra (order):**
  - Com o passar do tempo mais clientes proporcionalmente se tornaram inativos, o que afetou diretamente a quantidade de pedidos e também a receita;
  - Não se tinha muitos detalhes a respeito da classificação de um cliente como Marlin, porém, ficou evidente que os tais fazem mais pedidos que os demais (assim como gastam mais) no total e na média\*\*.
- **Acesso (visits):**
  - Dado a disponibilidade dos dados apenas para os acessos que geraram pedidos, não se extraiu muitos insights falando especificamente do tema;
  - Base de acesso foi útil na construção do tabela de efetividade dos pushes.

\* Assumimos que os descontos tenham sido enviados/recebidos através de pushes.

\*\* Os gráficos que ilustram isso está no [apêndice](#).



## - PROBLEMA 3 -

Prever a quantidade gasta por cliente no mês seguinte

1

## Motivação:

Pelo fato de se ter uma base histórica em mãos, a ideia foi explorar a possibilidade de antecipar uma informação valiosa de alguém já conhecido e, assim, ganhar tempo no desenho de estratégias/ações e na tomada de decisão.

Como target, a ideia é prever o valor gasto total por cliente no mês seguinte.

2

## Impacto:

- Otimizar Pushes: antecipar qual seria a quantidade ideal de pushes, dado uma previsão de receita.
- Fortalecer alocação de entregadores: dado a previsão de crescimento de demanda para uma determinada região, é possível antecipar e otimizar entregadores.
- Potencializar parcerias com restaurantes: estimar variação de demanda por praça e garantir insights para time comercial em foco na expansão de praças.



Foram utilizados **cinco modelos para prever o valor total gasto** por cliente no mês seguinte: *Regressão Linear*, *Árvore de Decisão*, *Random Forest*, *Gradient Boosted Tree* e *Generalized Linear Regression*. Além disso, ao remover os outliers ( $\pm 1.5$  IQR), observamos uma redução de +R\$40 no MAE e +85 no RMSE.

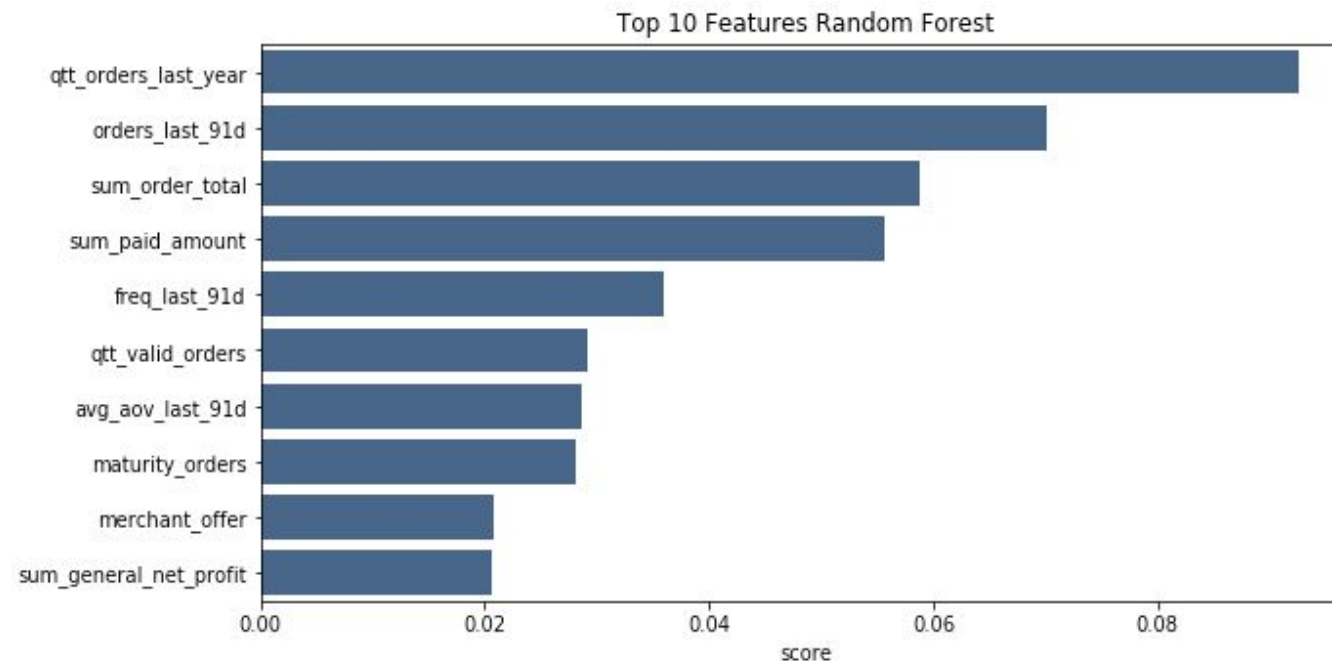
Modelo	COM OUTLIERS		SEM OUTLIERS	
	MAE	RMSE	MAE	RMSE
Regressão Logística	107.18	168.82	69.16	89.16
Árvore de Decisão	112.55	181.88	70.71	90.65
Random Forest	110.09	177.75	70.20	89.84
Gradient Boosted Tree	109.94	174.54	69.73	89.79
Generalized LR	107.14	168.78	69.16	89.15

Baseados nos menores *MAE* e *RMSE*, os modelos escolhidos foram o **Gradient Boosted Tree** e o **Generalized Linear Regression**.





**Random Forest** foi o método de **seleção de variáveis** utilizado. Baseado nas principais features, utilizamos o **VectorSlice** para selecionar as principais variáveis.



Em termos de *MAE* o **melhor modelo foi o Gradient Boosted Tree** (GBT) com resultado de 68.6. Os parâmetros escolhidos, após a aplicação de Cross Validation e otimização dos hiperparâmetros foram: *maxBins* = 32, *maxDepth* = 5, *maxIter* = 5, *stepSize* = 0.1 (em CV com 2 folds).

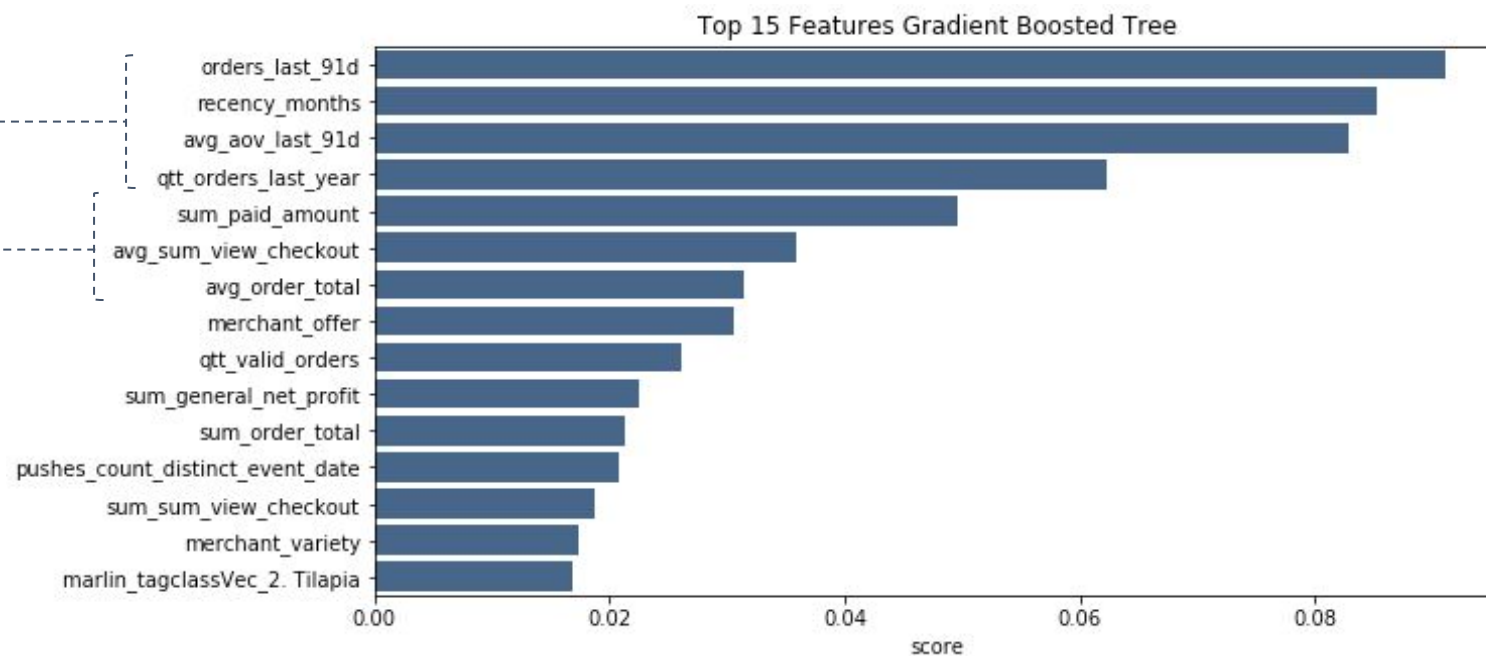
Os melhores parâmetros obtidos (e utilizados) para a Regressão Logística (RL) foram: *regParam* = 0.1 e *maxIter* = 25.

Modelo	SEM TÉCNICAS		COM TÉCNICAS	
	MAE	RMSE	MAE	RMSE
Gradient Boosted Tree	69.73	89.79	68.60	88.33
Generalized LR	69.16	89.15	68.63	88.25



**Contexto de sazonalidade:** variáveis com maior contexto temporal (dados de 91d e do último ano) tiveram maior importância.

Variáveis quantitativas relacionadas ao **comportamento de compra** no mês também são relevantes para prever quanto um cliente vai gastar.



- O **Gradient Boosted Tree** foi o **melhor modelo** para a predição do valor mensal gasto no próximo mês, com *MAE* igual a 68.6 e *RSME* igual a 88.3;
- Na prática, o **erro na previsão do modelo** não é tão alto e **não deve implicar diretamente na efetividade do uso** do modelo para o que se propôs:
  - A média da *target* é de R\$207.79 e a mediana é R\$128.7;
  - O *MAE* obtido, R\$68.6, é aproximadamente 25% maior que o ticket médio mensal (~R\$55);
  - Isto é, para uma base onde as pessoas gastam, em média, 4 tickets médios por mês, o modelo construído erra 1.25 pedidos.
- A **remoção de outliers** e o uso de **técnicas para otimização dos hiperparâmetros e validação cruzada** foram **fundamentais** para o desempenho obtido:
  - Sem o tratamento de outliers e aplicação de técnicas de otimização: *MAE* = R\$107; *RMSE* = R\$170 (*RMSE* 58% maior que o *MAE*);
  - Após o tratamento de outliers e aplicação de técnicas de otimização: diminuição de ~35% do *MAE* e cenário com *RMSE* igual a 28% maior que o *MAE*. Ou seja, um ganho de 30p.p. apenas na comparação.
- A magnitude da distribuição dos erros **não indica erros muito elevados**:
  - Como o *RSME* (~R\$89) ficou mais próximo do *MAE*, apesar de estarmos errando.



Sob uma **perspectiva de negócios**, ao prever qual vai ser o possível valor gasto no mês seguinte, como destacado na motivação da escolha por essa abordagem, podemos:

- **Calibrar** com maior assertividade o volume de envios de pushes. Com isso, seria possível:
  - Evitar que um eventual excesso de envios de comunicações resulte em churn; ou mesmo
  - Antecipar também o *ifood\_status* do cliente, tendo a chance de evitar o churn com um incentivo.
- **Estimar** melhor a demanda por região. Com isso, seria possível:
  - Alocar melhor recursos (posicionamento de entregadores).
- **Gerar** insights para o time comercial sobre regiões com potencial aumento de demanda.
  - Estimular/alinhar promoções em determinadas regiões e/ou com determinados tipos de restaurante baseado na preferência dos clientes;
  - Antecipar a classificação da qualidade do cliente (*marlin\_tag*), podendo criar ações específicas com o intuito de fidelizar ainda mais os melhores e estimular (se fizer sentido) a evolução dos demais.



## 6. Desafios encontrados

Sumarizando os desafios enfrentados ao longo das entregas e para a resolução dos problemas, tem-se que:

- **Limitações computacionais:**

- Nos Problemas 1 e 3 a falta de recursos computacionais limitou a melhoria da performance dos modelos. Técnicas de validação cruzada e otimização dos hiperparâmetros foram aplicadas, porém, dado o tempo de processamento demandado, ficou inviável a realização de seguidas rodadas para o aperfeiçoamento pleno.

- **Limitações de base/dados:**

- Como já mencionado, a base de visitas trás apenas os dados para casos associados a pedidos. Dessa forma, as análises específicas sobre acessos e o uso dessa base no cálculo da efetividade do pushes ficaram limitadas no Problema 2;
- Apesar do material de consulta e da disponibilidade do pessoal do Ifood, a falta de conhecimento sobre a base nos obrigou a fazer *assumptions* que poderiam ter sido evitados.



# - APÊNDICE -

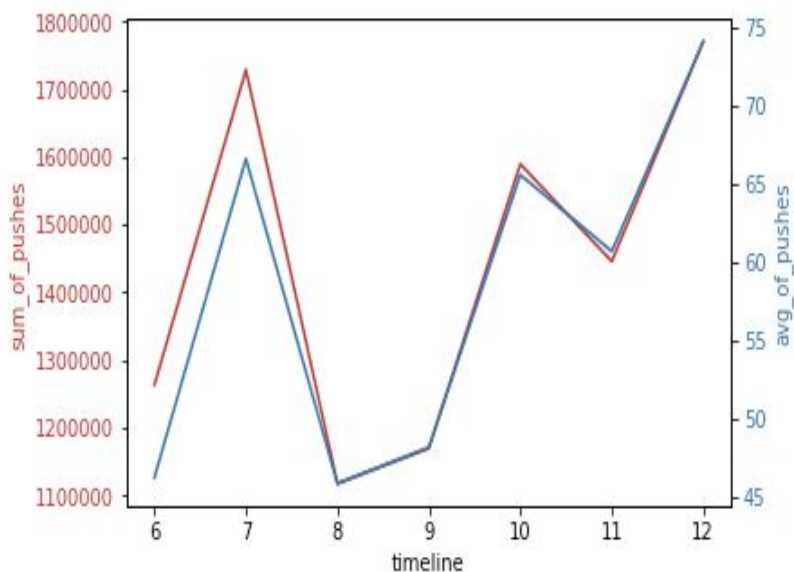
# Comportamento do Lucro

ANÁLISE DESCRITIVA | PROBLEMA 2

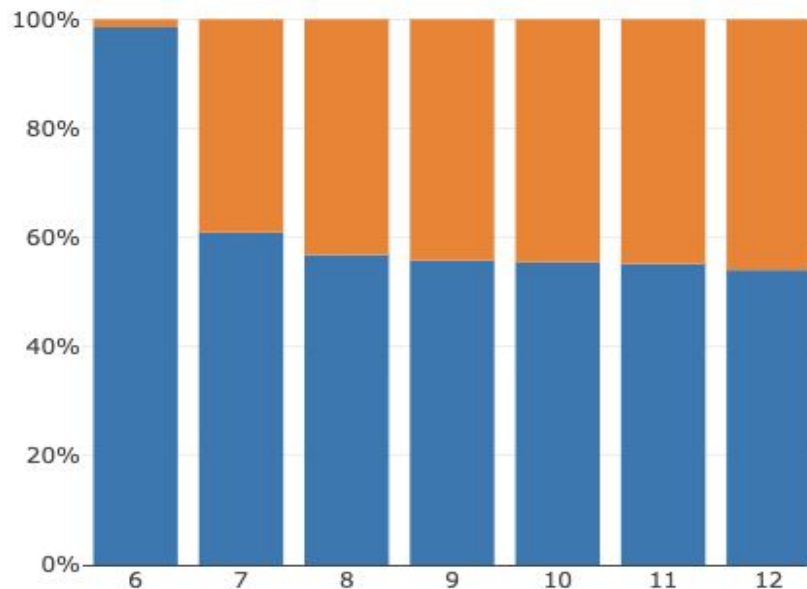
44

O mês de Dezembro pode ser considerado como atípico já que foi nesse mês em que aconteceram:

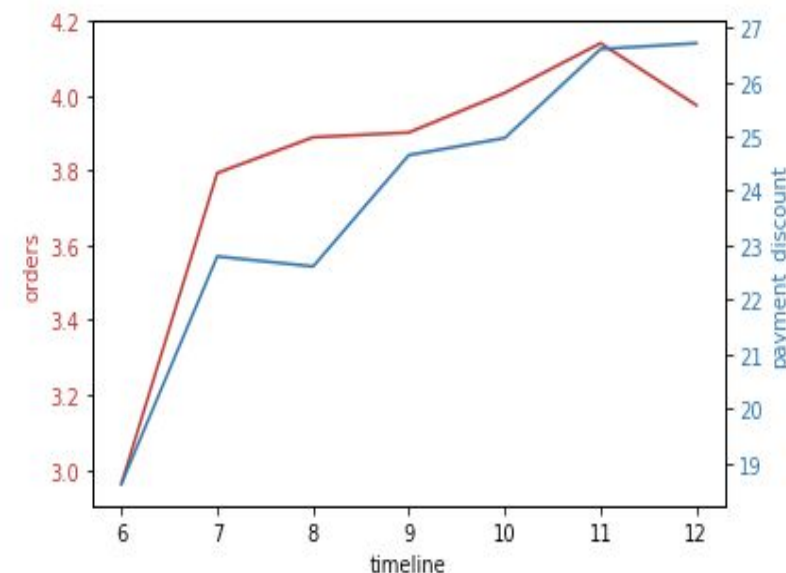
Máxima histórica de pushes (+1.8M, média de 75 pushes/mês):



Máxima proporção histórica de inativos (46%):



Uma ligeira queda na média de pedidos / usuário (de 4.2 para 4) e um desconto médio maior (~R\$27)



**Observação:** É possível imaginar que os resultados do mês de Dezembro sejam impactados pelas tradicionais comemorações de final de ano.

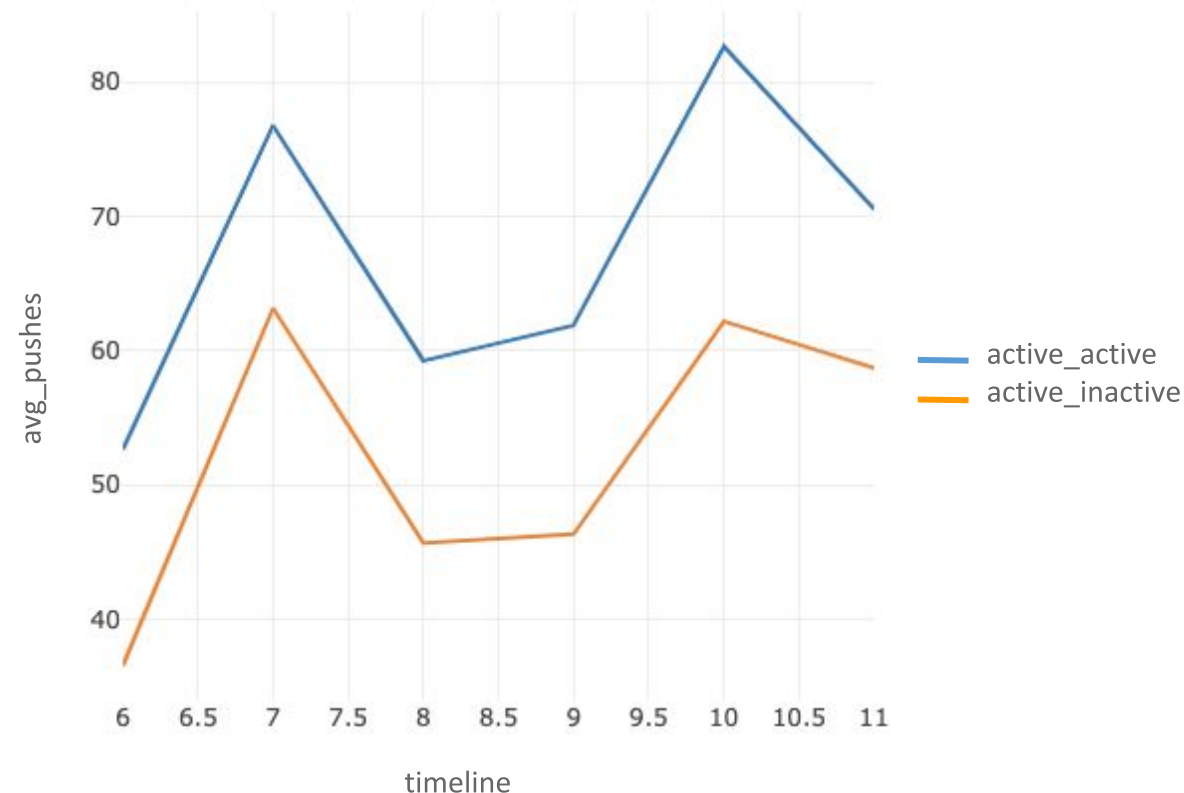


# Trade-off: Envio de Pushes

ANÁLISE DESCRITIVA | PROBLEMA 2

45

Além de, em média, os inativos receberem menos pushes que os ativos, identificou-se também que para todos os meses que usuários ativos que passarão a ser inativos no mês seguinte, receberam **em média menos pushes no mês vigente** do que os que permanecerão como ativos.



# Trade-off: Envio de Pushes

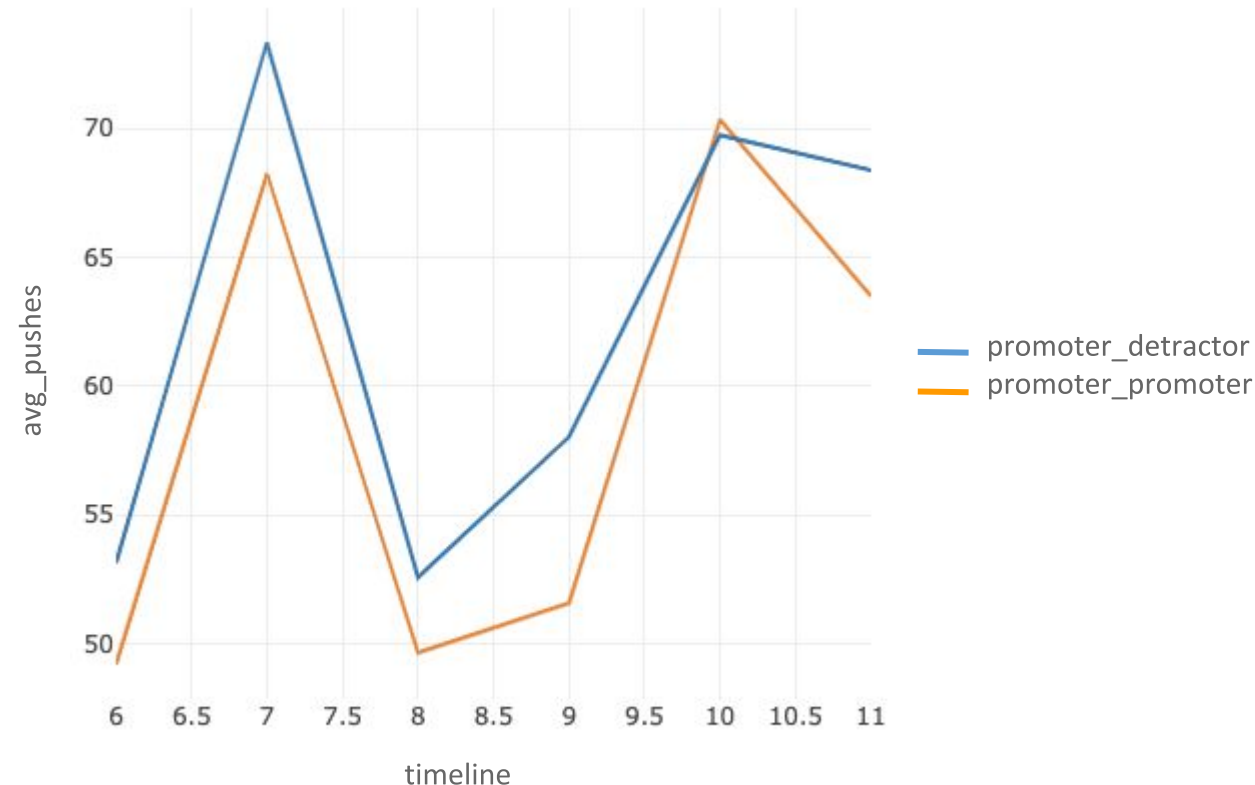
ANÁLISE DESCRITIVA | PROBLEMA 2

46

Isso implica que devemos enviar mais pushes?

**Não necessariamente!**

Usuários promotores que passarão a ser detratores no mês seguinte, receberam **em média mais pushes no mês vigente** do que os que permaneceram como promotores.



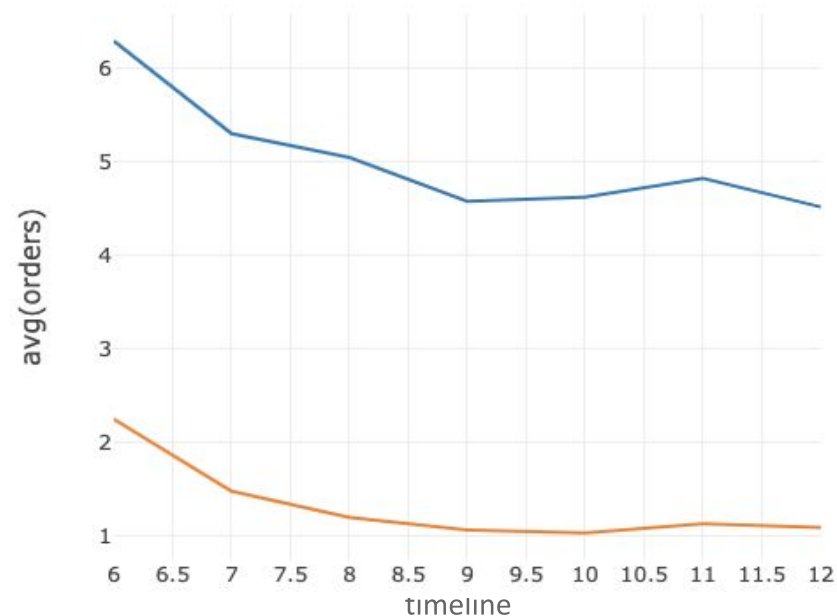
# Compras: Marlin Tag

ANÁLISE DESCRITIVA | PROBLEMA 2

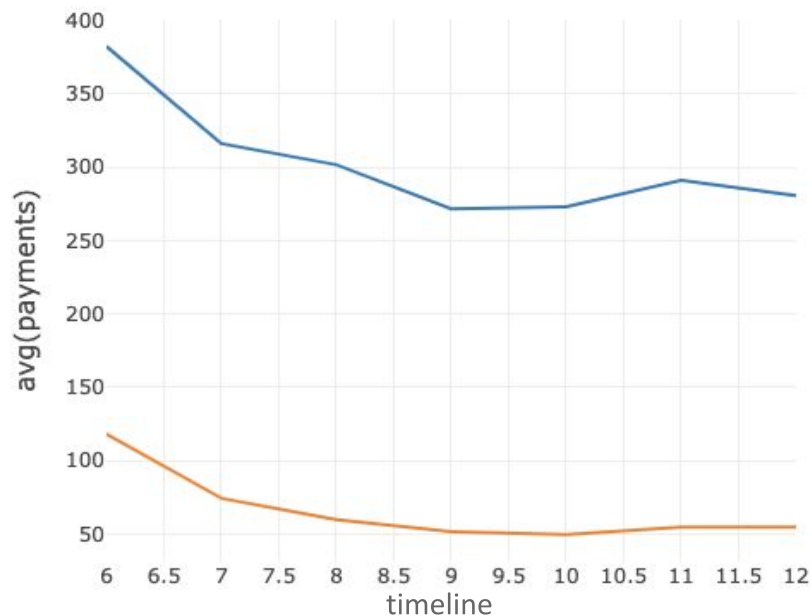
47

É sabido que os clientes classificados como *Marlin* são considerados os melhores. Constatou-se então que eles são, de fato, **os que mais compram e os que mais pagam pelos seus pedidos**, em média. Mais do que isso, o Ifood também se preocupa em garantir que eles **recebam mais pushes, em média, ao longo do mês** para todos os meses.

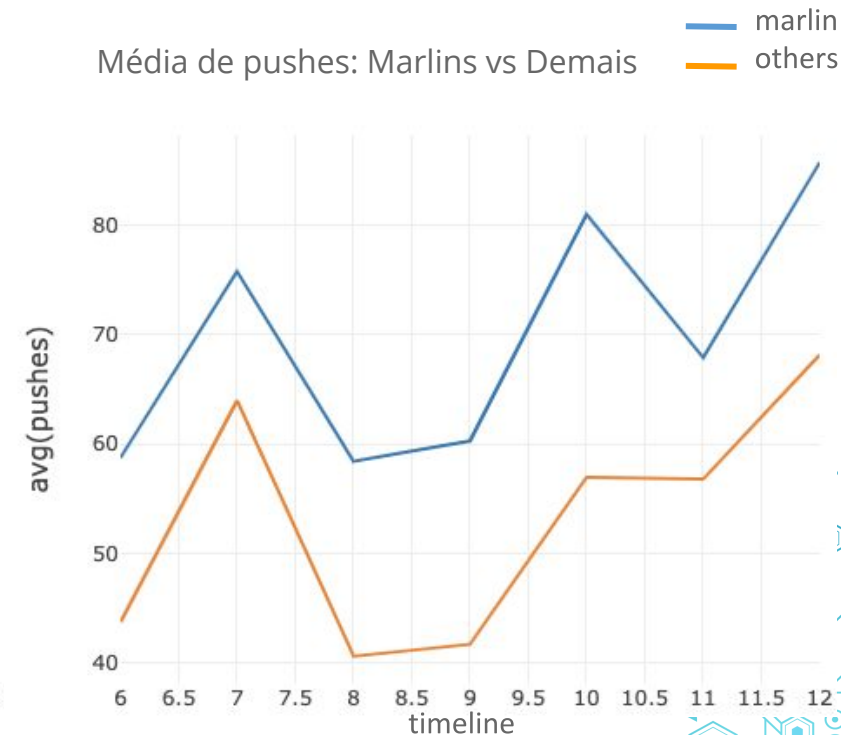
Média de pedidos: Marlin vs Demais



Média de gastos: Marlin vs Demais



Média de pushes: Marlin vs Demais



# Notebooks Databricks

Entrega 1:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/763780990988853/523723184344590/2620394790310521/latest.html>

Entrega 2:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/763780990988853/523723184344806/2620394790310521/latest.html>

Entrega 3:

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/763780990988853/523723184344882/2620394790310521/latest.html>

timeline

timeline

timeline



**OBRIGADO!**



# LABDATA FIA – Laboratório de Análise de Dados

50



Unidade Pinheiros



Unidade Paulista

