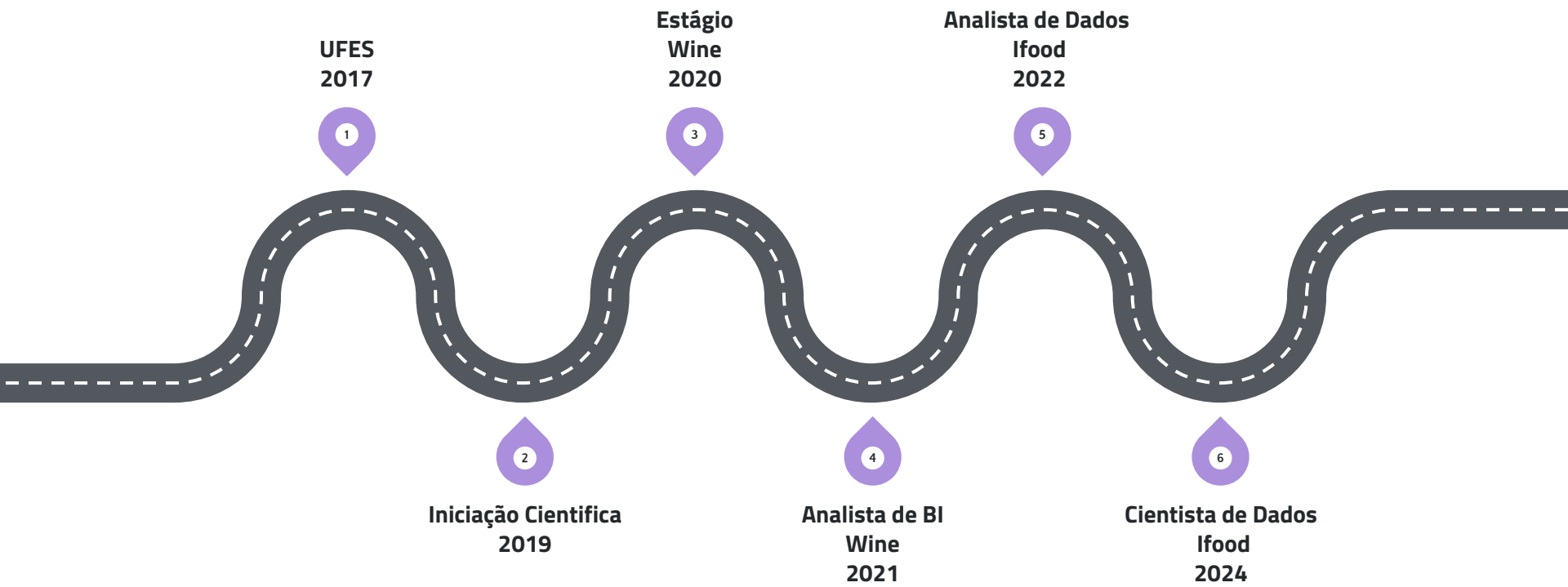


Estatística?

Mas vai trabalhar com o que?



**UFES
2017**

1

**Estágio
Wine
2020**

3

**Analista de Dados
Ifood
2022**

5

**Iniciação Científica
2019**

2

**Analista de BI
Wine
2021**

4

**Cientista de Dados
Ifood
2024**

6

Cênario em 2017



Jornal da USP

<https://jornal.usp.br> › Universidade

Cursos da USP: Estatística é considerada a profissão do ...

22 de ago. de 2017 — Apesar de pouco conhecida, a estatística — ciência que coleta, analisa e sumariza dados — **é reconhecida como a profissão do futuro** na era do Big ...



O Futuro das Coisas

<https://ofuturodascoisas.com> › estatistica-uma-das-profis...

Estatística: uma das profissões do futuro

17 de abr. de 2017 — O “**estatístico do futuro**” não deveria apenas realizar cálculos e análise dos dados, coisas que softwares irão assumir por completo. É preciso ...



Superinteressante

<https://super.abril.com.br> › cultura › 6-razoes-para-acre...

6 razões para acreditar que estatística é a profissão do futuro

15 de abr. de 2011 — US\$ 12 bilhões são gastos por ano com pesquisas que geram informações capazes de aumentar em até 100% o lucro de uma empresa.



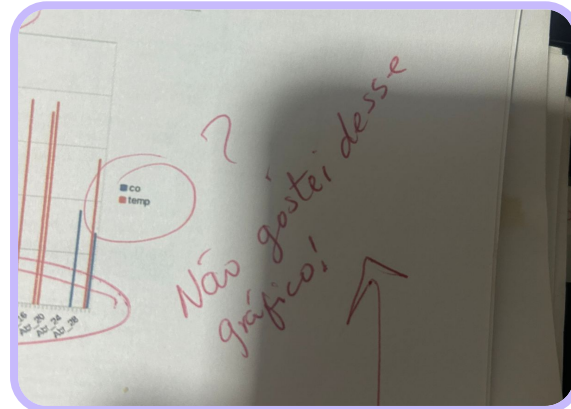
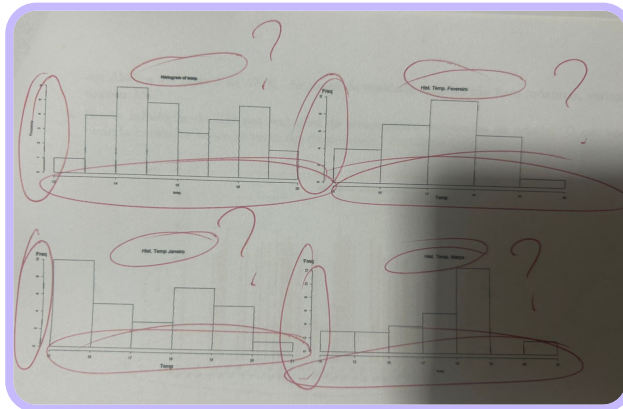
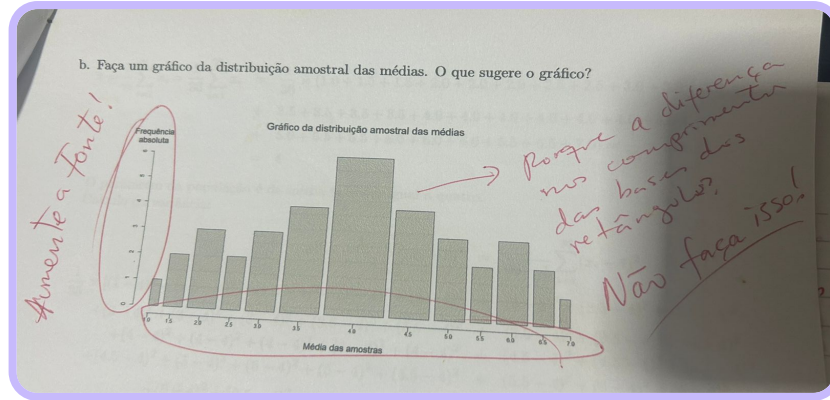


Recepção de
Calouros do Curso
de Estatística
UFES / 2017.1



2021

Na faculdade...



Iniciação Científica



Adolescentes em situação de violência no Espírito Santo:

uma análise dos casos notificados



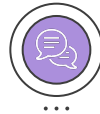
Tratamento de dados

Dados públicos dos hospitais do ES



Análise Descritiva

Análise de dados reais utilizando R



Comunicação

Interface com profissionais da saúde



Estágio em tech na Wine



Wine tech

E-Commerce de vinhos



15 vagas

3000 candidatos



O poder da graduação

Em estatística



Construção de bases em SQL

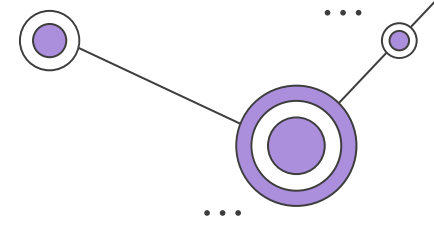
Customer Value | top ten accounts by country | Order Revenue

```

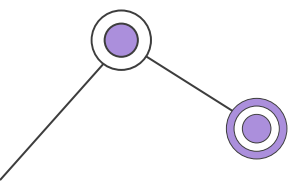
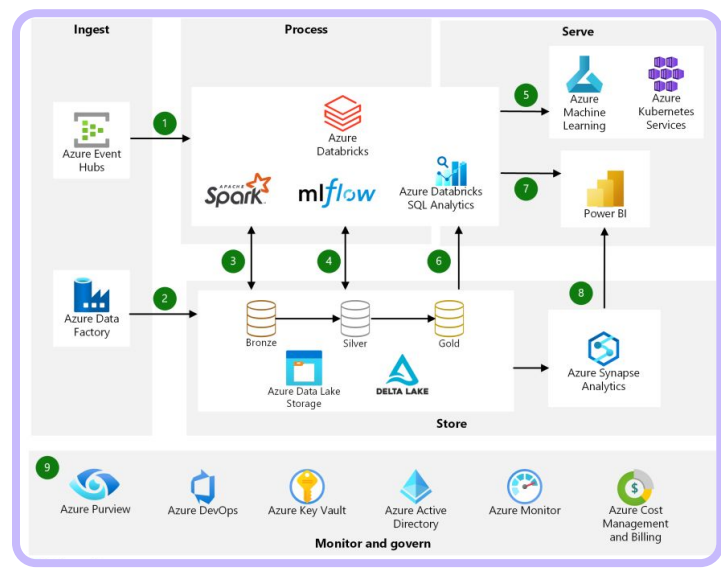
1 SELECT
2   o_orderdate AS Date,
3   o_orderpriority AS Priority,
4   sum(o_totalprice) AS "Total Price"
5 FROM
6   "samples"."tpch"."orders"
7 WHERE
8   o_orderdate > '1994-01-01'
9   AND o_orderdate < '1994-01-31'
10 GROUP BY
11   1,
12   2
13 ORDER BY
14   1,
15   2
    
```

Date	Priority	Total Price
1994-01-02	1-URGENT	96444609.82
1994-01-02	2-HIGH	93497904.94
1994-01-02	3-MEDIUM	88800085.02
1994-01-02	4-NOT SPECIFIED	97955477.98
1994-01-02	5-LOW	98015661.37
1994-01-03	1-URGENT	92534508.96
1994-01-03	2-HIGH	92286715.43
1994-01-03	3-MEDIUM	93521575.91
1994-01-03	4-NOT SPECIFIED	87568921.46

12 s 751 ms | 145 rows returned | Refreshed a minute ago



Arquitetura e Engenharia de dados



Analista de BI na Wine



UX: Experiência do Usuário

Análise dos dados de usuários



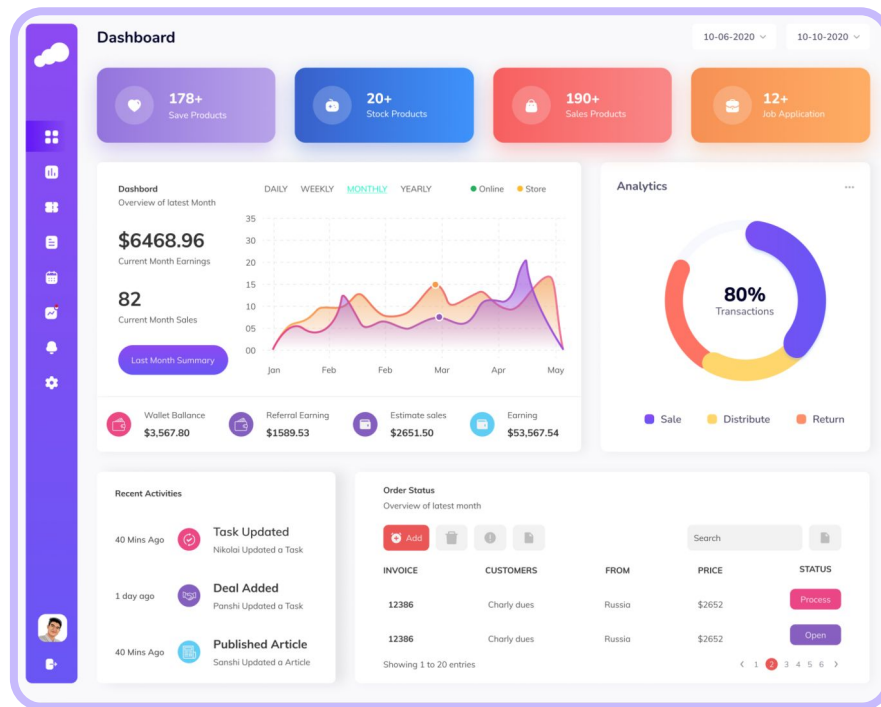
Problema com bases de dados

Falta de dados essenciais para o negócio



Criação do time de pesquisa

Estruturação do processo de pesquisa do zero.



Planejamento



Objetivo do teste
Definição do público
Definição do roteiro
Cenário

Alinhamento



Autorização
Disposição das pessoas

Preparação



Produzir protótipos
Organizar local
Setup softwares
Pai/Mãe e Tio/Tia

Aplicação



Condução do teste
Registro do teste
Streaming
Mensuração

Conclusão



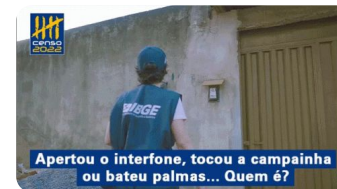
Planilha de métricas
Resultados
Plano de ação
Apresentação

	Tamanho da amostra para estimar uma proporção	Tamanho da amostra para estimar uma média
Amostra aleatória simples	$\frac{p(1-p)}{e^2}$	$\frac{\sigma^2}{e^2}$
Amostra estratificada proporcional	$\frac{\sum_{h=1}^L W_h p_h (1-p_h)}{e^2}$	$\frac{\sum_{h=1}^L W_h \sigma_h^2}{e^2}$
Amostra estratificada ótima	$\frac{(\sum_{h=1}^L W_h \sqrt{p_h(1-p_h)})^2}{e^2}$	$\frac{(\sum_{h=1}^L W_h \sigma_h)^2}{e^2}$

Calculadoras Amostrais para calculo de N minimo para métrica de sucesso Proporção

1- População infinita

```
1 from scipy.stats import norm
2 # Erro Relativo
3 # é a relação entre o Erro Absoluto(diferença entre o valor verdadeiro e a aproximação) e o valor verdadeiro
4 def sample_size_relative_error(max_lift_error, confidence_level, control_rate):
5     alternative_rate = (1-max_lift_error)=control_rate
6     avg_rates = (control_rate+alternative_rate)/2
7     significance = 1-(1-confidence_level)/2
8     z_quantile = norm.ppf(significance)
```





Problemas de assertividade

Omissão de opções

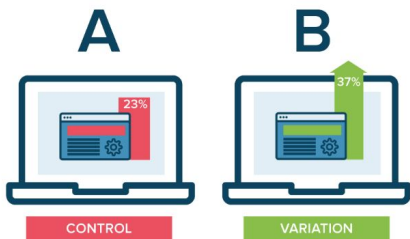
...



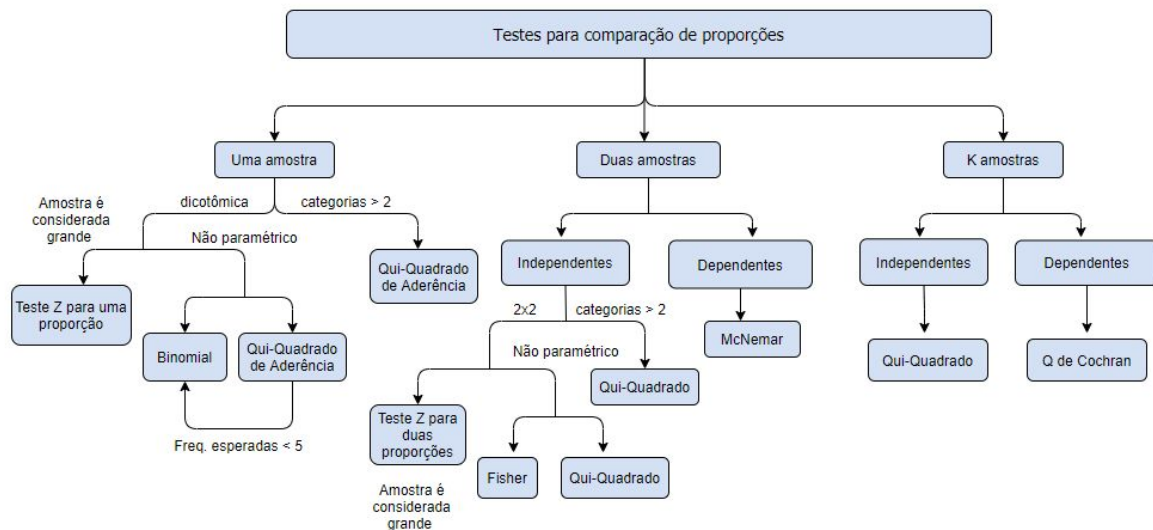
Criação do time de experimentação

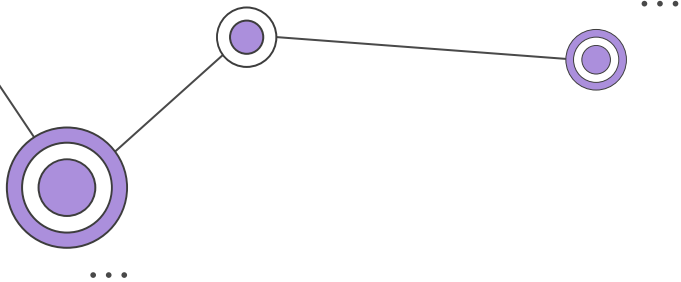
Apenas com a teoria em mente

...



Variação	Visitas	Conversões	Taxa de Conversão	Diferença
Original	342	69	20,18%	0%
Sem telefone	299	145	53,90%	+167%





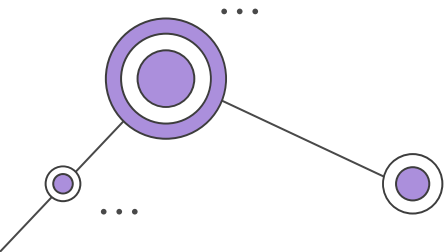
Clusterização de clientes

Recomendação para cada grupo



Melhoria no clube wine

Personalização do APP



Analista de dados no Ifood - Parte 1



Experimentação em CRM

Time de marketing do Ifood



Automatização de processos estatísticos

Cálculo de amostras, chatbots.



Workshops educativos

Aulas e palestras sobre estatística e conceitos básicos para times de negócios

Reativação/Ativação / Teste xx _ mm/aa

**Funcionou
Não funcionou
Aprimoramento**

Detalhes do teste

Objetivo

Hipótese

Motivador

Mecânica
Quais são as alternativas que vamos testar?
Quais as modificações de cada uma?
Qual o público?

Variação A: ...
Variação B: ...

KPI de sucesso

- xxx

Métricas descritivas

- xxx

Links úteis

[Documentação](#)

Resultado e Conclusão

	Controle	Variável	Delta	Teste de hipótese
Média Métrica 1	freq abs. (freq rel)	freq abs. (freq rel)	variação em %	Com 95% de confiança, podemos dizer que a diferença do teste é incremental
Média Métrica 2	freq abs. (freq rel)	freq abs. (freq rel)	variação em %	Com 90% de confiança, não temos indícios de que há um incremento do teste, não existe uma diferença significativa

*** Diferença significativa:** não foi por acaso, existe uma diferença estatística entre os dois grupos. (há um incremento do teste)
*** Não existe uma diferença significativa:** não existe uma diferença estatisticamente falando, o teste não foi incremental (não fez diferença)

Conclusão muito sucinta.
Dados que tragam, de forma resumida, os impactos dos kpis de sucesso, de forma % e absoluta.
Também trazer impacto com rollout total de MUB e R\$.
Ex: economia de R\$1M com zero impacto de MUB.

Teste de hipóteses

O que é?
Para que serve?
Por que devemos usar?

Fernanda Borem

Analista de dados no Ifood - Parte 2



Time de retenção

Testes de cupons para grupos específicos



Modelos de classificação de clientes

Machine Learning

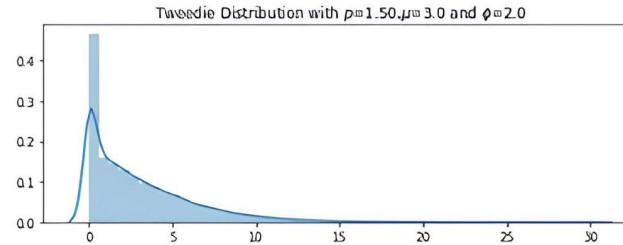


Primeiro contato com DS

Entendimento das variáveis, códigos e melhorias

$$RU = wF \times Feynman + wM \times MP + wT \times TP + wM \times Mecanica$$

$MP = \text{Mediana Entre Pedidos}$
 $TP = \text{Tempo Sem Pedido}$

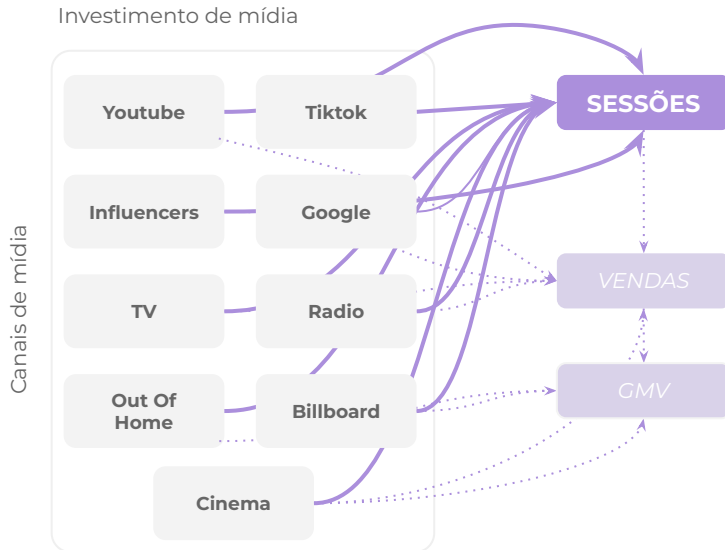


Dias até o Próximo Pedido	Dias sem Pedir	Mês para Investir	Razão de Urgência	Bucket
1	8	M0	8,00	9
5	14	M0	2,80	8
22	28	M0	1,27	7
9	11	M0	1,22	6
3	1	M0	0,33	3
21	4	M0	0,19	2
37	23	M1	0,62	9
33	16	M1	0,48	8
46	20	M1	0,43	6
54	20	M1	0,37	4
54	19	M1	0,35	3
42	1	M1	0,02	1

Prioridade

Cientista de Dados no Ifood - Parte 3

- Marketing Mix Modeling (MMM)



Robyn

$$y_t = \text{Intercept} + \beta_j \times \frac{x_{decay_{t,j}}^\alpha}{x_{decay_{t,j}}^\alpha + \gamma^\alpha} + \beta_{hol} \cdot hol_t + \beta_{sea} \cdot sea_t + \beta_{trend} \cdot trend_t + \dots + \beta_{ETC} \cdot ETC_t + \varepsilon$$

Labels in the diagram:
 - y_t : Dependent Variable
 - $\beta_j \times \frac{x_{decay_{t,j}}^\alpha}{x_{decay_{t,j}}^\alpha + \gamma^\alpha}$: S-Curve component for each media (j)
 - $\beta_{hol} \cdot hol_t + \beta_{sea} \cdot sea_t + \beta_{trend} \cdot trend_t + \dots + \beta_{ETC} \cdot ETC_t$: Holiday, Seasonality and Trend effect*
 - $\beta_j, \beta_{hol}, \beta_{sea}, \beta_{trend}, \beta_{ETC}$: Independent Variables

Main components of the function:

1. Adstock transformation: $X_{decay_{t,j}} = X_{t,j} + \theta_j \cdot X_{decay_{t,j-1}}$
2. S Curve transformation: $S \text{ Curve } (x, j) = \beta_j \times \frac{x_{decay_{t,j}}^\alpha}{x_{decay_{t,j}}^\alpha + \gamma^\alpha}$

where:

- y_t = revenue at time t
- t = time index of dependent and independent variable (week)
- j = media index (e.g. FB, TV, OOH) and $\beta, \alpha, \gamma, \theta$ = regressor specific to each media j
- γ implemented on the S - Curve is a transformed γ where $\gamma_{tran} = \text{quantile}(X_{decay_j}, \gamma)$
- β_{ETC}, ETC_t = further independent variables to be added to the model (e.g. competitor, promotions)
- ε = Error term (accounting for all the other factors not addressed in the model)

Sessões = baseline

+ tendência + sazonalidade + efeito_feriado

efeitos temporais e naturais

+ orders + pushes_clicked + n_accounts + installs + ...

+ spend_google + spend_tiktok + spend_youtube

+ spend_tv + spend_eventos + spend_out_of_home +

...

+ erro

Comportamento orgânico

Métricas de negócio

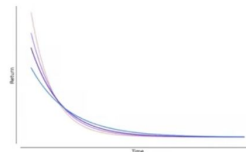
Investimento de mídia

Toda variação que não puder ser atribuído a comportamento "orgânico"

```
InputCollect <- robyn_inputs(  
  dt_input = Robyn::dt_simulated_weekly,  
  dt_holidays = Robyn::dt_prophet_holidays,  
  date_var = "DATE",  
  dep_var = "revenue",  
  dep_var_type = "revenue",  
  prophet_vars = c("trend", "season", "holiday"),  
  prophet_country = "DE",  
  context_vars = c("competitor_sales_B", "events"),  
  paid_media_spends = c("tv_S", "ooh_S", "print_S", "facebook_S", "search_S"),  
  paid_media_vars = c("tv_S", "ooh_S", "print_S", "facebook_I", "search_clicks_P"),  
  organic_vars = "newsletter",  
  factor_vars = "events",  
  window_start = "2016-11-23",  
  window_end = "2018-08-22",  
  adstock = "geometric",  
  # To be defined separately  
  hyperparameters = NULL,  
  calibration_input = NULL  
)
```

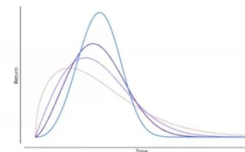
Common probability distribution used to model adstocking curves

Geometric adstock



Peak occurs at the beginning and starts to diminish from the start. It is usually modelled by a geometric or exponential distribution.

Weibull adstock



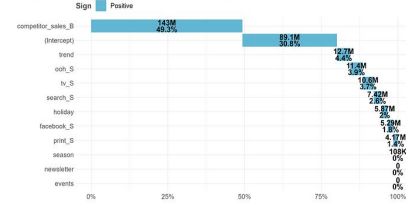
There is a delay in the peak. It is usually used to represent offline marketing or in segments on which it takes a while until the customers take action (e.g. car selling).

Saídas do modelo em R

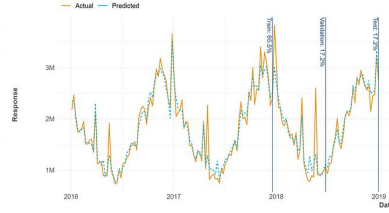
One-pager for Model ID: 3_130_15

Adj R2: train = 0.9344, val = 0.7782, test = 0.9565 | NRMSE: train = 0.0586, val = 0.5718, test = 0.2704 | DECOMPRSSD = 0.408 | MAPE = 0.7773

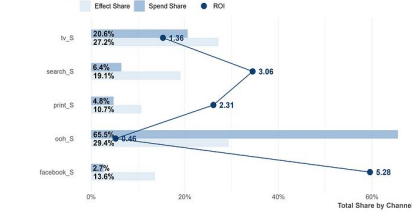
Response Decomposition Waterfall by Predictor



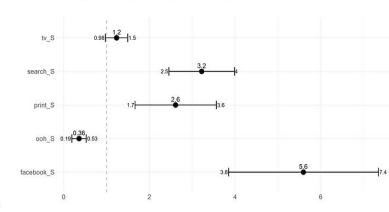
Actual vs. Predicted Response



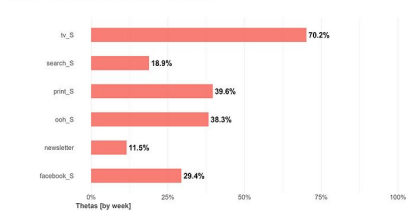
Total Spend% VS Effect% with total ROI



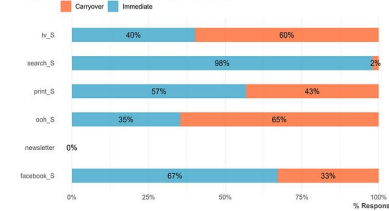
In-cluster bootstrapped ROI with 95% CI & mean



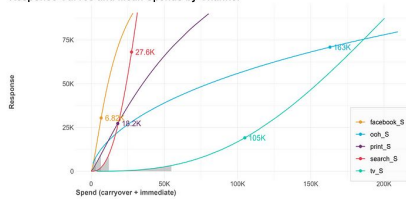
Geometric Adstock: Fixed Rate Over Time



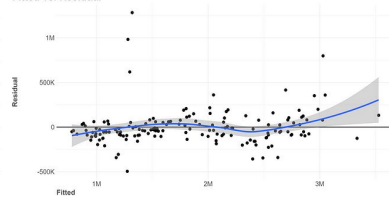
Immediate vs. Carryover Response Percentage



Response Curves and Mean Spends by Channel



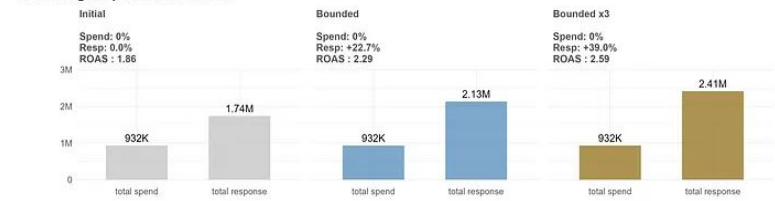
Fitted vs. Residual



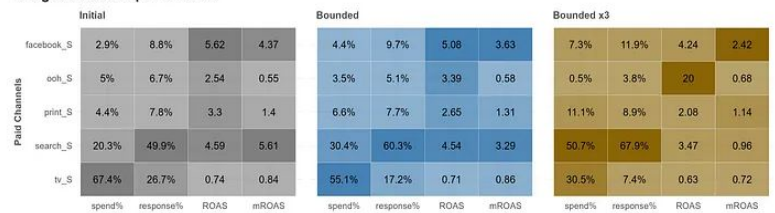
Budget Allocation Onepager for Model ID 3_130_15

Adj R2: train = 0.9344, val = 0.7782, test = 0.9565 | NRMSE: train = 0.0586, val = 0.5718, test = 0.2704 | DECOMPRSSD = 0.408 | MAPE = 0.7773
Simulation date range: 2018-12-10 to 2018-12-31 (4 weeks) | Scenario: max_response

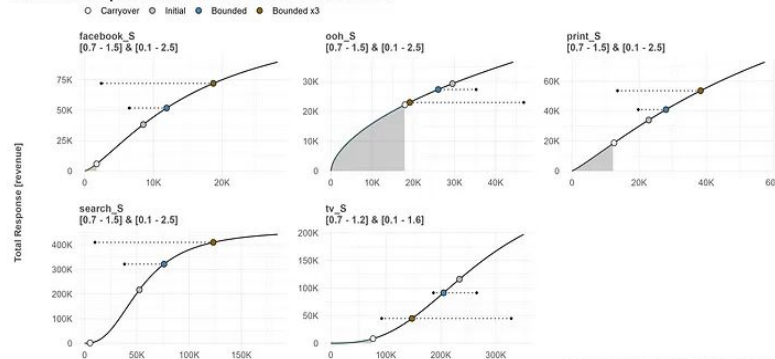
Total Budget Optimization Result



Budget Allocation per Channel*



Simulated Response Curve for Selected Allocation Period



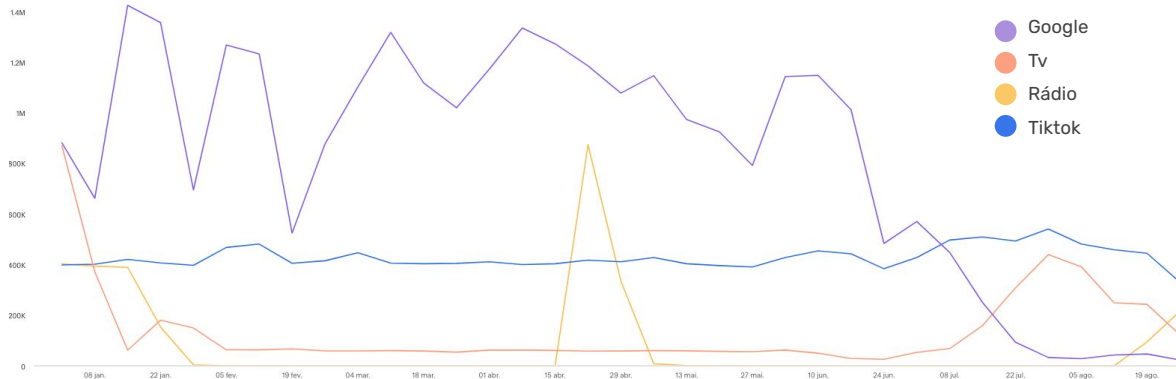
* ROAS = total response / raw spend | mROAS = marginal response / marginal spend
* When reallocating budget, mROAS converges across media within respective bounds
* Dotted lines show budget optimization lower-upper ranges per media

6,2% - 7,3%

Impacto das campanhas de Marketing em usuários **ativos** no APP (xM -yM)

4,4% - 5,6%

Impacto das campanhas de Marketing em usuários **novos** no APP (xM -yM)



- Existe incrementabilidade nessa campanha?
- Quais dias foram os mais incrementais?
- Quais os canais mais eficientes?
- Quantos dias posso ficar sem ações até que o efeito suma?
- Qual o valor ótimo de investimento para cada meio?

Saturation by channel





Mercado aquecido

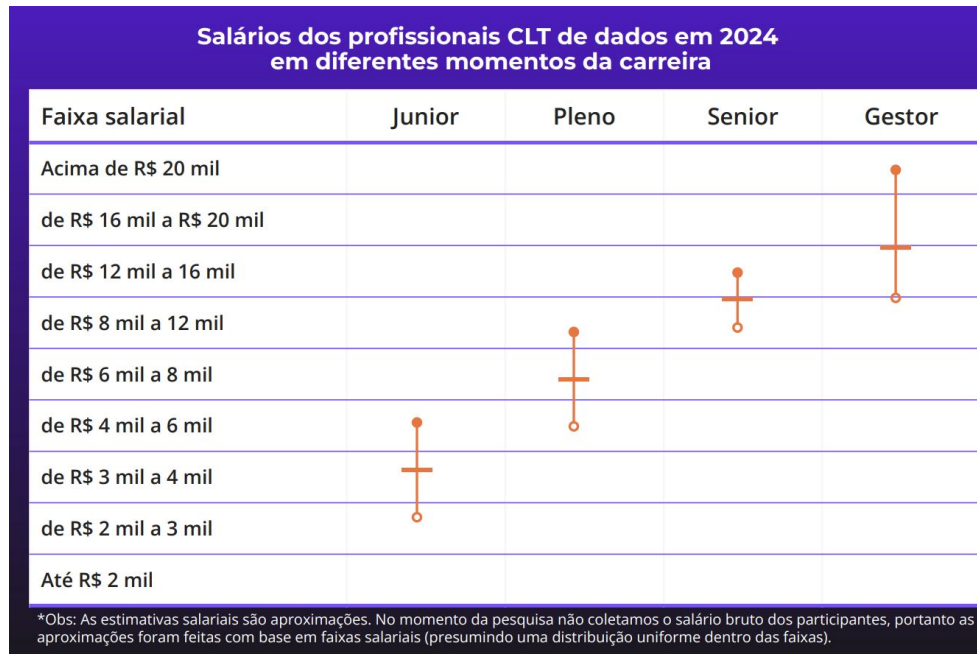
Alta procura e altos salários



Ser estatístico é diferencial

Capacidade analítica, conhecimento especializado, conhecimento de negócios e comunicação.

41,6%		Computação / TI / Eng. Software Sistemas de Informação
20,8%		Área de Engenharia
15,7%		Economia / Administração Contabilidade / Finanças
7,2%		Estatística/ Matemática Matemática Computacional
2,6%		Física / Química
2,5%		Marketing / Publicidade Comunicação / Jornalismo
2,4%		Ciências Biológicas / Farmácia Medicina / Área da Saúde
7,2%		Outras



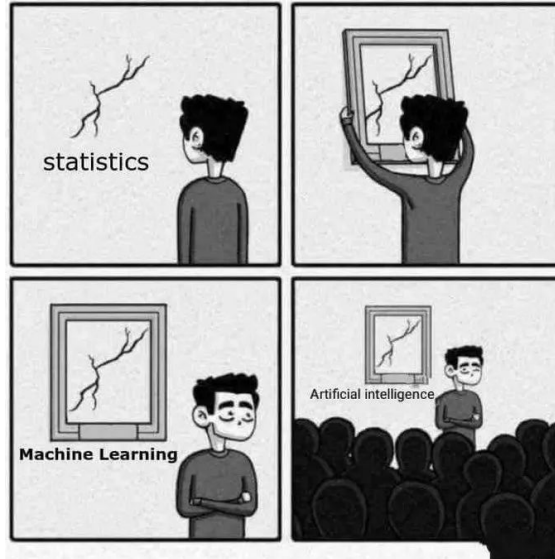
Obrigada!



Scan me!



[linkedin.com/in/fernandaborem](https://www.linkedin.com/in/fernandaborem)



...