

**ATLAS  
TECHNOLOGIES**

# **Relatório de teste técnico de Data Science**

**BRUNO LEITE**

---

**2021**

# Sumário

---

- 01.** Apresentação do teste
- 02.** Obtenção dos dados
- 03.** Análise inicial dos dados
- 04.** Limpeza dos dados
- 05.** Análise exploratória
- 06.** Conclusão

# Apresentação do teste

---

A Atlas possui uma plataforma com 23.800 perfis ativos, e com mais de 450 milhões de acessos mensais. Recentemente foi feito o deploy de um novo projeto na Atlas, referente a assinaturas mensais de planos.

O CEO entrou em contato com o time de Data Science e solicitou um relatório para monitorar esse novo projeto. Ele gostaria de saber o máximo de informações possíveis, a fim de entender o comportamento do usuário na compra de assinaturas.

Você é o cientista de dados responsável por esse relatório, use a tabela fornecida para criar um relatório que explique o comportamento dos nossos usuários.

# Obtenção dos dados

---

Esta etapa tem como objetivo realizar a verificação dos dados, ou seja, a maneira de como estão situados, permitindo uma melhor compreensão da estrutura dos dados, facilitando a realização da nossa análise. A base de dados possui 2027 usuários e 14 variáveis do tipo: objeto, datetime64, float64 e int64.

# Análise inicial dos dados

---

Neste tópico será realizado o início da análise dos dados do dataset, com o objetivo de buscar obter informações valiosas relacionadas ao comportamento do usuário na compra de assinaturas.

## Dicionário das variáveis

- Profile ID: id do perfil
- Profile Created At: data de criação do perfil
- Profile Deleted At: data da exclusão do perfil
- Plan Code: tipo do plano de assinatura
- Payment Method: método de pagamento
- Charge Value: valor da assinatura
- Subscription Status: status da inscrição
- Charge Day: dia do pagamento
- Last Charge At: última data de pagamento
- Next Charge At: próxima data de pagamento
- Subscription Canceled At: data de cancelamento da inscrição
- Cancel Reason: motivo de cancelamento
- Subscription\_Created\_At: data de realização da inscrição
- Estado: Estado que o usuário reside

Analisando as primeiras 5 entradas do dataset, já verificamos os atributos e algumas entradas estão em inglês, sendo assim foi necessário traduzir para o português.

# Análise inicial dos dados

df.head()

Profile ID	Profile Created At	Profile Deleted At	Plan Code	Payment Method	Charge Value	Subscription Status	Charge Day	Last Charge At	Next Charge At	Subscription Canceled At	Cancel Reason	Subscription_Created_At	Estado
473128	2021-07-16 12:45:36	NaT	plano D	CREDIT_CARD	59.9	DRAFT	19	NaN	NaT	NaT	NaN	2021-07-19 18:45:21	Rio de Janeiro
315949	2021-07-14 11:01:43	NaT	plano A	CREDIT_CARD	89.9	CANCELED	23	2021-07-23	2021-07-23	2021-07-23	PAYMENT FAILED	2021-07-23 20:05:14	Rio de Janeiro
293763	2018-12-21 04:34:25	NaT	plano B	CREDIT_CARD	359.9	CANCELED	24	2021-07-24	2021-07-24	2021-07-27	USER SUSPENDED	2021-07-24 21:34:55	Bahia
124763	2021-01-24 07:52:57	NaT	plano D	CREDIT_CARD	59.9	CANCELED	25	NaN	NaT	2021-07-04	ADMIN CANCELED	2021-07-25 01:04:58	Rio de Janeiro
391446	2021-01-24 07:52:57	NaT	plano A	CREDIT_CARD	99.9	CANCELED	25	NaN	NaT	2021-07-04	ADMIN CANCELED	2021-07-25 01:05:05	Rio de Janeiro

# Limpeza dos dados

---

Nesta etapa foi realizada algumas limpezas necessárias para manter uma certa qualidade do conjunto de dados.

A qualidade de um dataset está relacionado diretamente à quantidade de valores ausentes, por isso foi verificado a presença dele.

Durante a análise foi verificado que cinco atributos possuíam valores ausentes, sendo que um deles (`perfil_excluido_em`) possuía cerca de 93% de valores ausentes, havendo a necessidade de excluir para não atrapalhar na análise exploratória ou caso seja necessário realizar previsões, a mesma não seja afetada.

Um outro insight, foi a necessidade de adicionar um atributo chamado (`tempo_para_assinar`), ou seja, o tempo que um usuário leva para assinar algum plano desde a data que criou o perfil até a data de assinatura.

# Análise exploratória

---

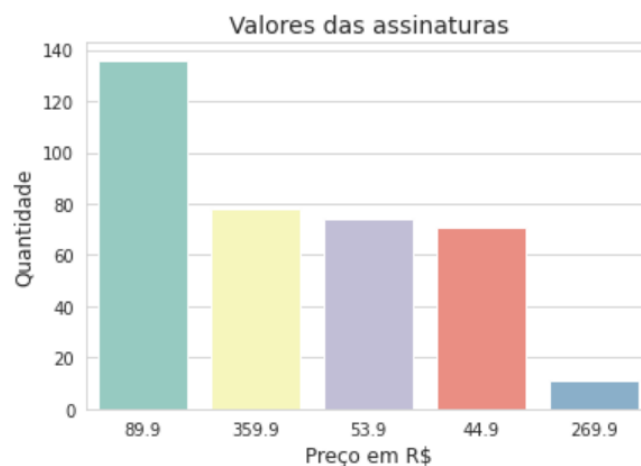
Depois de realizar as etapas anteriores, chegou o momento de realizar a exploração dos dados.

O objetivo dessa etapa é extrair insights referente ao comportamento dos clientes da empresa Atlas Technologies com base no dataset obtido.

A seguir foram realizados algumas perguntas aos dados para obter alguns insights.

## Qual a relação dos valores da assinatura dos usuários ativos?

A partir de uma análise descritiva 75% dos usuários assinam planos cujo valores correspondem a R\$89,99.



## Qual a relação do tempo que leva para o usuário ativo realizar uma assinatura?

A partir de uma análise descritiva 75% dos usuários levam até 707 dias aproximadamente para realizar uma assinatura. E em média o total de usuários levam aproximadamente 464 dias para realizar uma assinatura.

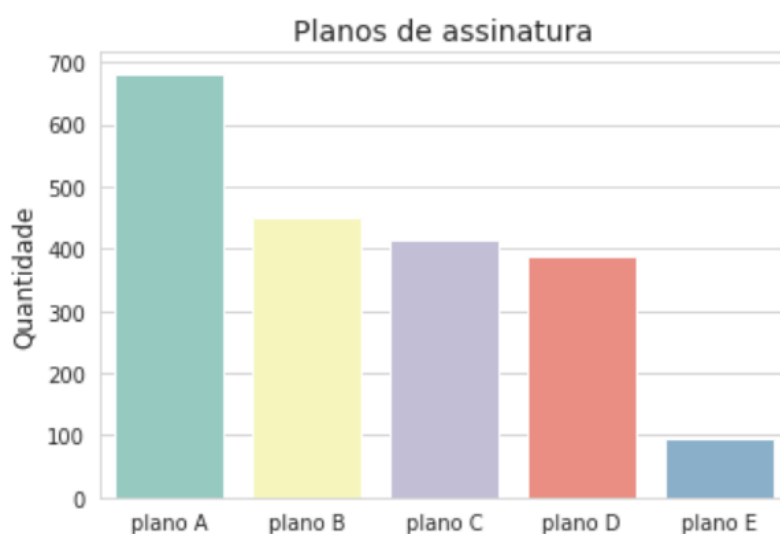


# Análise exploratória

---

Qual o tipo de plano mais escolhido independente se o usuário está ativo ou não.

A partir um gráfico de barras é possível verificar que o tipo de plano mais escolhido é plano A, devido ao valor menor correspondente.



Qual o tempo que leva para o usuário assinar qualquer plano, independente se está ativo ou não?

A partir da uma análise descritiva 75% dos usuários levam até 584 dias aproximadamente para realizar uma assinatura no PLANO A e em média o total de usuários levam aproximadamente 387 dias para realizar uma assinatura.

No PLANO B 75% dos usuários levam até 516 dias e o total de usuarios em média 380 dias para assinar.

No PLANO C 75% dos usuários levam até 525 dias e o total de usuarios em média 367 dias para assinar.

No PLANO D 75% dos usuários levam até 548 dias e o total de usuarios em média 381 dias para assinar.

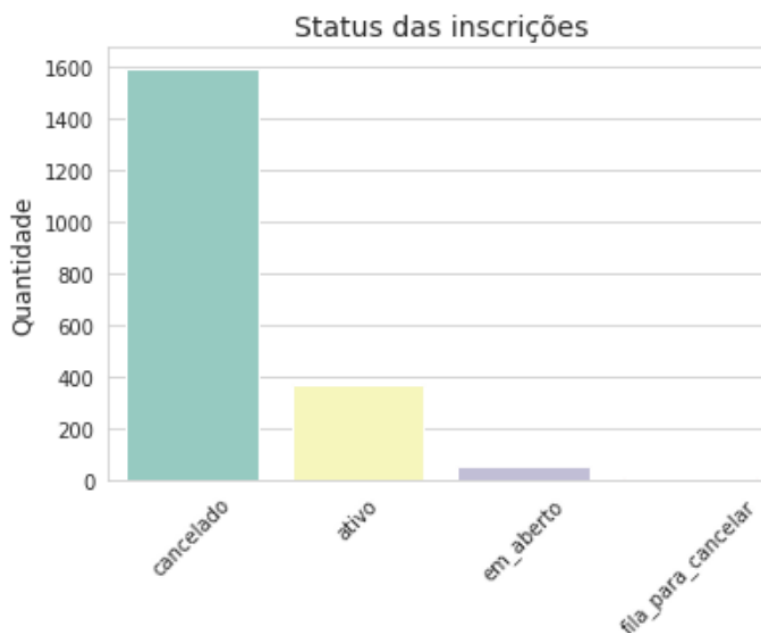
# Análise exploratória

---

No PLANO E 75% dos usuários levam até 507 dias e o total de usuários em média 374 dias para assinar.

## Como está o comportamento do status das inscrições ?

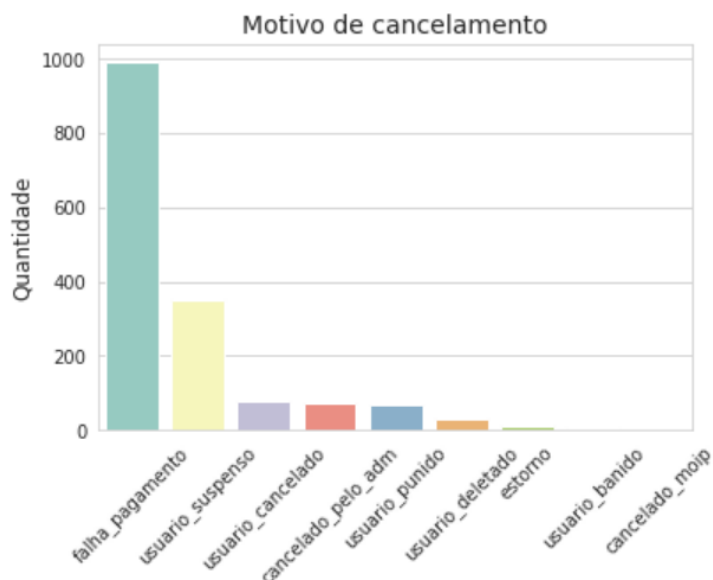
De acordo com o gráfico abaixo, fica claramente que o número de assinaturas canceladas é a maior em relação aos outros status.



## Quais os motivos de cancelamento de assinatura ?

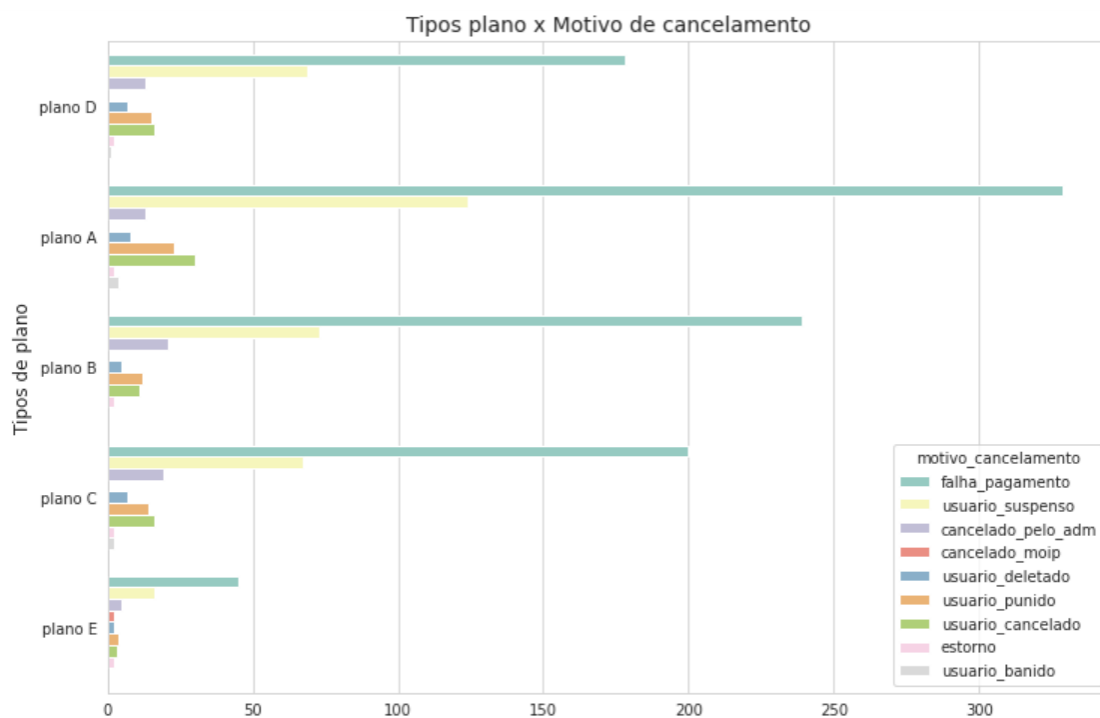
De acordo com o gráfico abaixo, mostra os motivos de cancelamento que levou o usuário a cancelar a assinatura. Com **grande relevância** a **falha no pagamento** é destacado com uma **grande quantidade**, sendo um importante fator a ser analisado. Em seguida acompanha o usuário suspenso.

# Análise exploratória



Qual a relação do motivo de cancelamento com os planos ?

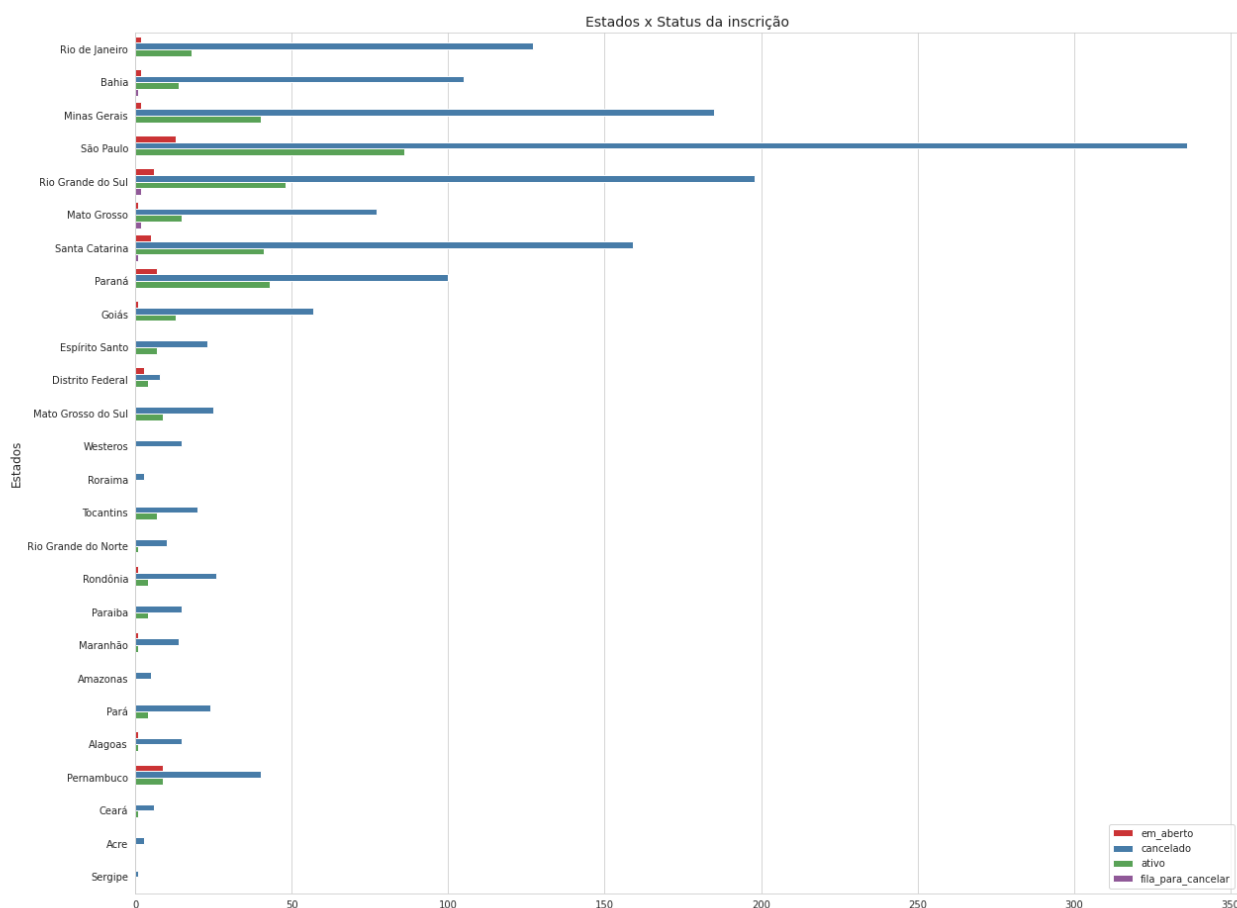
Podemos verificar, de acordo com o gráfico abaixo, que todos os planos são cancelados devido a falha no pagamento, sendo mais abrangente nos planos A,B,C e D.



# Análise exploratória

## Qual a relação dos estados com o status da inscrição ?

Podemos observar que o estado de São Paulo tem uma maior quantidade de assinaturas canceladas e também ativas. Os estados do Paraná, Rio Grande do Sul, Santa Catarina e Minas Gerais possuem um número maior em relação aos outros estados.



# Conclusão

---

Podemos concluir que a análise de dados é de extrema importância, pois nos mostra insights que não conseguiríamos ver sem uma exploração nos dados.

Foi possível observar que o número de assinantes que possuem sua assinatura cancelada é alta e o motivo está principalmente na falha do pagamento, ou seja, poderíamos verificar mais a fundo o porquê está ocorrendo essa falha, talvez seja algo com administradora do cartão de crédito, ou alguma problema do cliente, entre outros.

A criação do atributo que mostra o tempo que leva para o cliente assinar algum plano, é interessante pois, obtendo uma maior quantidade de dados, poderíamos prever a partir de um modelo de machine learning, quais os fatores que influenciam o tempo que leva para uma pessoa assinar algum plano, e isso poderia ser feito da mesma maneira do tempo que leva para a pessoa cancelar a assinatura. Assim teríamos mais insights e saberíamos onde podemos alterar atributos para obter mais assinaturas com menor chance de cancelamento.