**Sprint 4 - Projeto**

Parabéns! Você concluiu a seção sobre Análise estatística de dados. É hora de aplicar o conhecimento e as habilidades que você adquiriu em um projeto: um estudo de caso analítico da vida real que você vai concluir por conta própria.

Quando terminar, envie seu trabalho para o revisor do projeto para avaliação. Você vai receber feedback em até 48 horas. Use esse feedback para fazer alterações e envie a nova versão de volta ao revisor.

Você talvez receba mais feedback referente à nova versão. Isso é completamente normal. Não é incomum passar por vários ciclos de feedback e revisão.

Seu projeto será considerado concluído assim que o revisor o aprovar.

Descrição do projeto

Você trabalha como analista para a empresa de telecomunicações Megaline. A empresa oferece aos clientes dois planos pré-pagos: Surf e Ultimate. O departamento comercial quer saber qual dos planos gera mais receita para ajustar o orçamento de publicidade.

Você vai realizar uma análise preliminar dos planos com base em uma pequena seleção de clientes. Você terá dados de 500 clientes da Megaline: que clientes são, de onde eles são, qual plano usam e o número de chamadas e mensagens realizadas em 2018. Seu trabalho é analisar o comportamento dos clientes e determinar qual plano pré-pago gera mais receita. Mais tarde, nas instruções do projeto, você verá exatamente quais aspectos do comportamento dos clientes precisa analisar. Determinar qual plano, em média, gera mais receita é uma tarefa que pode ser resolvida usando testes estatísticos. Você vai receber mais informação sobre isso na seção de instruções do projeto.

Descrição dos planos

Observação: a Megaline arredonda segundos para minutos e megabytes para gigabytes. Para chamadas, cada chamada individual é arredondada para cima: mesmo que uma chamada tenha durado apenas um segundo, um minuto será contado. Para tráfego da web, sessões individuais da web não são arredondadas para cima. Ao invés disso, o total do mês é arredondado para cima. Se alguém usar 1.025 megabytes no mês, a cobrança será de 2 gigabytes.

Aqui está uma descrição dos planos:

Surf

1. Preço mensal: $20
2. 500 minutos mensais, 50 mensagens de texto e 15 GB de dados
3. Após exceder os limites do pacote:
   * 1 minuto: 3 centavos
   * 1 mensagem de texto: 3 centavos
   * 1 GB de dados: $10

Ultimate

1. Preço mensal: $70
2. 3.000 minutos mensais, 1.000 mensagens de texto e 30 GB de dados
3. Após exceder os limites do pacote:
   * 1 minuto: 1 centavo
   * 1 mensagem de texto: 1 centavo
   * 1 GB de dados: $7

Dicionário de dados

Neste projeto, você vai trabalhar com cinco tabelas separadas.

1. A tabela users (dados sobre usuários):
   * *user\_id* — identificador exclusivo do usuário
   * *first\_name* — nome do usuário
   * *last\_name* — sobrenome do usuário
   * *age* — idade do usuário (em anos)
   * *reg\_date* — data da inscrição (dd, mm, aa)
   * *churn\_date* — a data que o usuário parou de usar o serviço (se o valor estiver ausente, isso significa que o plano estava em uso quando o banco de dados foi extraído)
   * *city* — cidade de residência do usuário
   * *plan* — nome do plano
2. A tabela calls (dados sobre as chamadas):
   * *id* — identificador de chamada exclusivo
   * *call\_date* — data da chamada
   * *duration* — duração da chamada (em minutos)
   * *user\_id* — identificador do usuário que faz a chamada
3. A tabela messages (dados sobre mensagens de texto):
   * *id* — identificador exclusivo da mensagem de texto
   * *message\_date* — data da mensagem de texto
   * *user\_id* — identificador do usuário que envia a mensagem de texto
4. A tabela internet (dados sobre sessões web):
   * *id* — identificador exclusivo da sessão
   * *mb\_used* — volume de dados gasto durante a sessão (em megabytes)
   * *session\_date* — data da sessão web
   * *user\_id* — identificador do usuário
5. A tabela plans (dados sobre os planos):
   * *plan\_name* — nome do plano
   * *usd\_monthly\_fee* — preço mensal em dólares americanos
   * *minutes\_included* — pacote mensal de minutos
   * *messages\_included* — pacote mensal de mensagens de texto
   * *mb\_per\_month\_included* — volume do pacote de dados (em megabytes)
   * *usd\_per\_minute* — preço por minuto depois de exceder o limite do pacote (por exemplo, se o pacote inclui 100 minutos, o primeiro minuto excedente será cobrado)
   * *usd\_per\_message* — preço por mensagem de texto depois de exceder o limite do pacote
   * *usd\_per\_gb* — preço por gigabyte extra de dados após exceder o limite do pacote (1 GB = 1.024 megabytes)

Instruções para concluir o projeto

Preparamos um modelo de notebook para você. O notebook contém orientações sobre que código escrever e também para a inclusão de explicações dos resultados ao longo do processo. Para concluir o projeto, você precisa preencher cada célula de código no modelo e editar as células Markdown nos casos em que for necessário explicar os resultados.

Certifique-se de incluir uma introdução que descreve brevemente seus objetivos e uma conclusão que traz um breve resumo dos resultados.

Etapa 1. Abra o arquivo de dados e estude as informações gerais

Abaixo estão os caminhos para os arquivos a serem lidos e links para baixá-los, se necessário.

*/datasets/megaline\_calls.csv* [Baixar o conjunto de dados](https://practicum-content.s3.us-west-1.amazonaws.com/new-markets/Data_Refactored_Sprint_3/megaline_calls.csv)

*/datasets/megaline\_internet.csv* [Baixar o conjunto de dados](https://practicum-content.s3.us-west-1.amazonaws.com/new-markets/Data_Refactored_Sprint_3/megaline_internet.csv)

*/datasets/megaline\_messages.csv* [Baixar o conjunto de dados](https://practicum-content.s3.us-west-1.amazonaws.com/new-markets/Data_Refactored_Sprint_3/megaline_messages.csv)

*/datasets/megaline\_plans.csv* [Baixar o conjunto de dados](https://practicum-content.s3.us-west-1.amazonaws.com/new-markets/Data_Refactored_Sprint_3/megaline_plans.csv)

*/datasets/megaline\_users.csv* [Baixar o conjunto de dados](https://practicum-content.s3.us-west-1.amazonaws.com/new-markets/Data_Refactored_Sprint_3/megaline_users.csv)

Etapa 2. Prepare os dados

* Converta os dados para os tipos necessários.
* Encontre e elimine erros nos dados. Certifique-se de explicar quais erros você encontrou e como os eliminou.

Para cada usuário, encontre:

* O número de chamadas feitas e minutos usados por mês
* O número de mensagens de texto enviadas por mês
* O volume de dados por mês
* A receita mensal gerada a partir de cada usuário. Para fazer isso, você precisa:
  + Subtrair o limite gratuito do pacote do número total de chamadas, mensagens de texto e dados;
  + Multiplicar o resultado pelo valor do plano;
  + Adicionar o preço mensal dependendo do plano.

Etapa 3. Analise os dados

Descreva o comportamento do cliente:

* Encontre os minutos, mensagens de texto e volume de dados que os usuários de cada plano necessitam por mês.
* Calcule a média, a variância e o desvio padrão.
* Construa histogramas. Descreva as distribuições.

Etapa 4. Teste as hipóteses

* As receitas médias para os usuários dos planos Ultimate e Surf são diferentes.
* A receita média dos usuários da área de NY-NJ é diferente da receita dos usuários de outras regiões.

Você decide quais valores alfa usar. Além disso, você precisa explicar:

* Como você formulou as hipóteses alternativas e nulas.
* Qual critério você usou para testar as hipóteses e por quê.

Etapa 5. Escreva uma conclusão geral

Formato: Conclua todas as tarefas no Jupyter Notebook. Armazene todo o código nas células code e explicações de texto nas células markdown. Adicione títulos e a formatação adequada quando necessário.

Como meu projeto será avaliado?

Elaboramos alguns critérios de avaliação para o projeto. Leia-os com atenção antes de começar a trabalhar.

Aqui está o que os revisores de projeto vão buscar ao avaliar o seu projeto:

* Como você explica os problemas identificados nos dados
* Como você prepara os dados para análise
* Quais gráficos você fez para as distribuições
* Como você interpreta os gráficos resultantes
* Como você calcula o desvio padrão e variância
* Se você criou hipóteses alternativas e nulas
* Quais métodos você usa para testar hipóteses
* Se você interpreta os resultados dos seus testes de hipóteses
* Se você interpreta os resultados dos seus testes de hipóteses
* As conclusões que você tirou
* Se você deixa comentários a cada etapa

No vídeo abaixo, abordamos algumas das dificuldades mais comuns que podem ser encontradas ao longo do projeto