**Estudo de caso - Google**

**Como um compartilhamento de bicicletas possibilita o sucesso rápido?**

**Nome : Leandro Dias Vieira**

Introdução

Esse é o meu estudo de caso para a conclusão do curso de Ánalise de Dados da Google. O principal desafio se baseia em entender como os membros anuais e os ciclistas casuais usam as bicicletas de forma diferente com o intuito de criar estratégias de marketing destinadas a converter os passageiros casuais em membros anuais.

Vou utilizar as ferramentas ensinadas no próprio curso para a realização desse projeto, além de seguir as etapas para uma boa análise de dados: **perguntar**, **preparar**, **processar**, **analisar**, **compartilhar** e **agir**.

PERGUNTAR

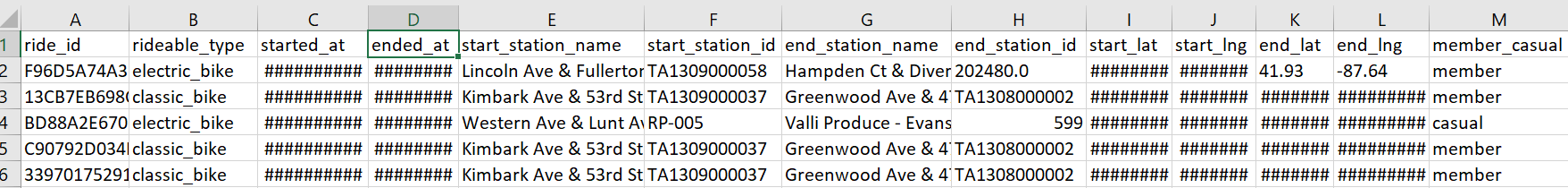
O principal objetivo do projeto é **criar estratégias de marketing destinadas a converter passageiros casuais em membros anuais**, A equipe de marketing necessita entender o **comportamento** de cada grupo (casual e anual) de acordo com os dados históricos para criar campanhas para os clientes certos, gerando assim boas conversões.

Três perguntas são feitas pelas partes interessadas para ter uma boa clareza do perfil dos clientes de cada grupo:

1. Como os membros anuais e os ciclistas casuais usam as bicicletas da Cyclistic de forma diferente?
2. Por que os passageiros casuais iriam querer adquirir planos anuais da Cyclistic?
3. Como a Cyclistic pode usar a mídia digital para influenciar os passageiros casuais a se tornarem membros?

Com as perguntas feitas, vamos partir para a parte de **preparação** dos dados.

PREPARAR

Foi disponibilizado um arquivo com os dados históricos da Cyclistic, mostrando todos os trajetos feitos pelos clientes em um período de 12 meses.  
  


Inicialmente identifiquei a estrutura dos dados, nome das colunas, entre outras características. Após conhecer a estrutura comecei a identificar sua integridade, se estão completos e limpos, sem nenhum tipo de interferência para as análises.

Os dados foram disponibilizados pela Motivate International Inc. , é possível encontrar a licença nesse link: [Licença](https://divvybikes.com/data-license-agreement)

Depois de seguir esses passos identifiquei algumas irregularidades na tabela, como dados faltantes, erros de formatação, entre outros que serão tratados no próximo passo: **Processar**

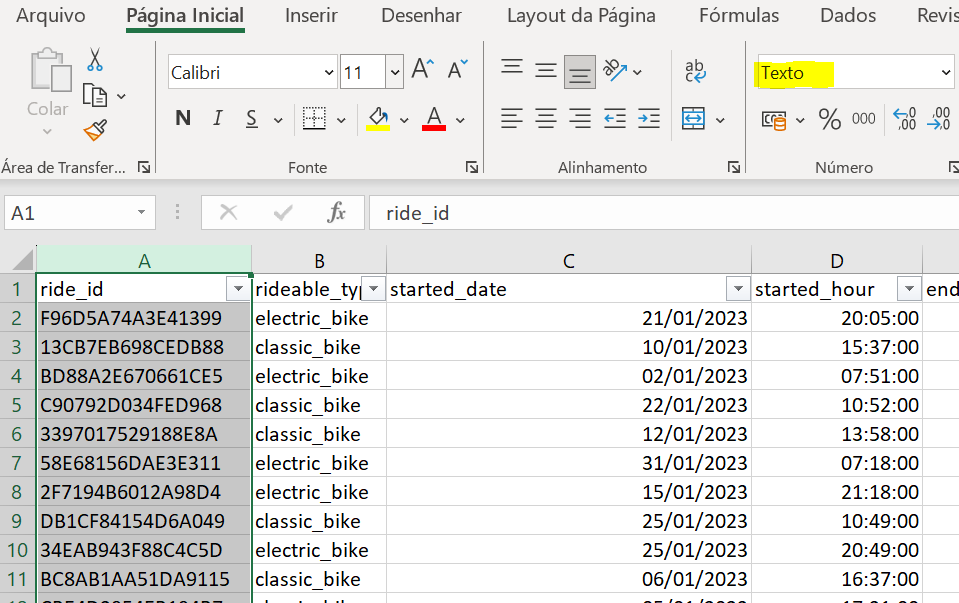
PROCESSAR

Com a tabela pronta e com todas as irregularidades identificadas, comecei a processar e tratar os dados para a análise. Utilizei o Excel para a realização desse tratamento.  
  
Primeiramente identifiquei as linhas vazias da tabela, Notei que muitos registros estavam com o endereço de estação vazia, tanto de inicío quanto de fim, após notar esse erro resolvi filtrar todos os registros vazios e retirá-los da tabela, pois é necessário saber qual foi o percurso dessa bicicleta.

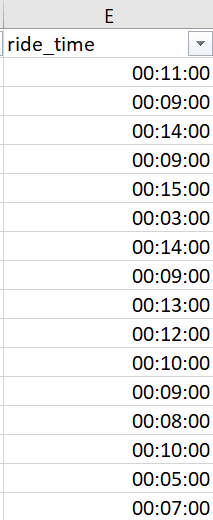


Sobrou 822.488 linhas das 1.025.129 totais.

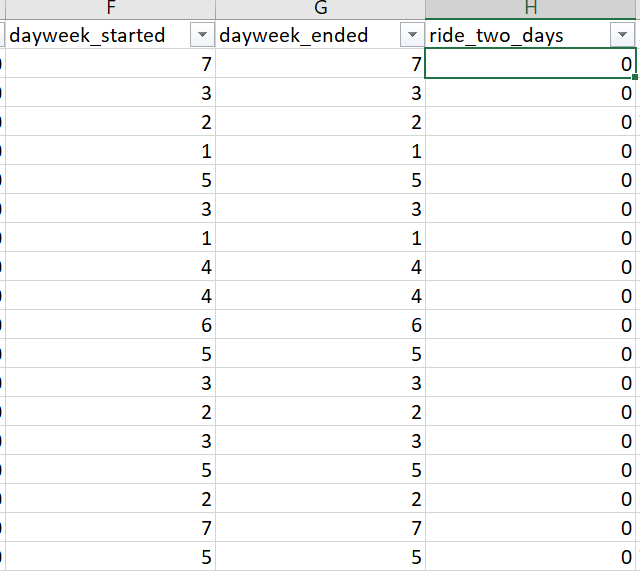
Após a filtragem dos dados e remoção de linhas nulas fiz a tipagem das colunas, atribuindo colunas como texto, hora, data, etc…



Com a tipagem dos dados feita resolvi criar uma coluna para calcular o tempo de viagem, subtraindo as colunas “ended\_at” e “started\_at”.



Após criar essa nova coluna para calcular o tempo de viagem, criei 3 novas colunas para verificar qual dia da semana em que foi realizada aquela corrida, tanto na data de início quanto na data final. Também criei a coluna “ride\_two\_days”, que mostra corridas que levaram mais de 1 dia.



Com os dados limpos e com novas colunas criadas para uma melhor abordagem dos dados, vamos seguir para o próximo passo: **Analisar.**

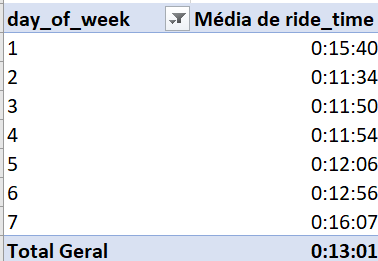
ANALISAR

Nesta etapa vou realizar novas medidas para identificar padrões e tendências nos dados.

Aproveitando algum dos cálculos feitos na etapa anterior, criei tabelas dinâmicas para ter uma visão ampla dos dados, ver padrões do tipo: Média da duração de passeio para membros e casuais, Média da duração de passeio para cada dia da semana

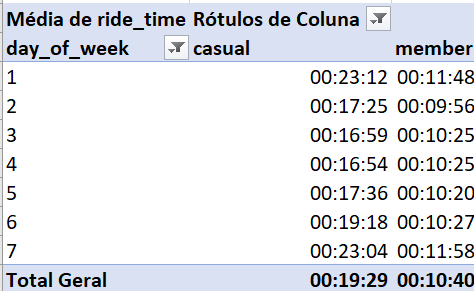


Podemos notar que os clientes casuais passam mais tempo utilizando as bikes comparado aos clientes membros, uma diferença de mais ou menos 9 minutos.



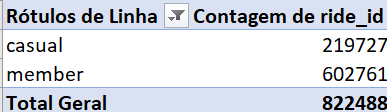
Nessa outra observação podemos notar que durante o 7º dia da semana (sábado) os clientes passam mais tempo utilizando o serviço.

Aproveitei essa observação para dividir entre clientes membros e casuais, descobrir qual dia da semana cada tipo de cliente mais utiliza os serviços.



Podemos notar que há uma grande diferença de tempo médio por cada tipo de cliente, os clientes casuais acabam utilizando o serviço por mais tempo em todos os dias da semana.

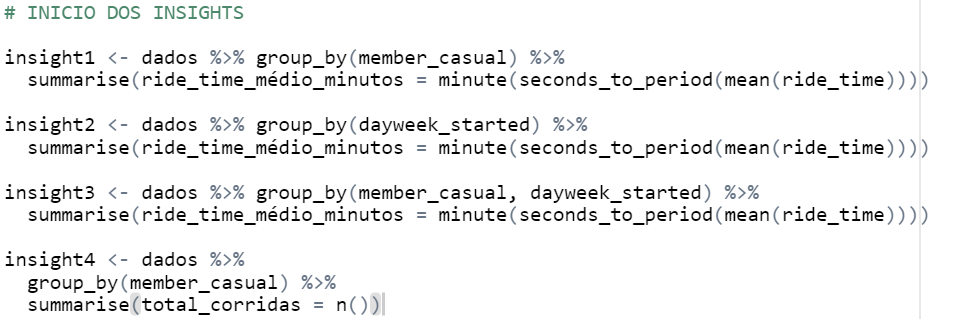
Também fiz uma tabela para analisar quantas corridas foram feitas por cada tipo de cliente, nela podemos ver qual tipo de cliente predomina na utilização do serviço.



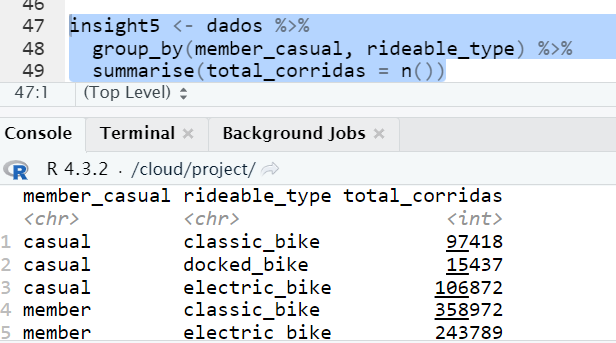
Os clientes Membros são os que mais utilizam o serviço, sendo mais do que o dobro comparado aos clientes casuais.

Analisando todas essas informações podemos tirar grandes insights, como por exemplo: Os clientes membros acabam utilizando o serviço mais vezes, porém os clientes casuais tem o tempo médio de utilização maior, principalmente aos finais de semana.

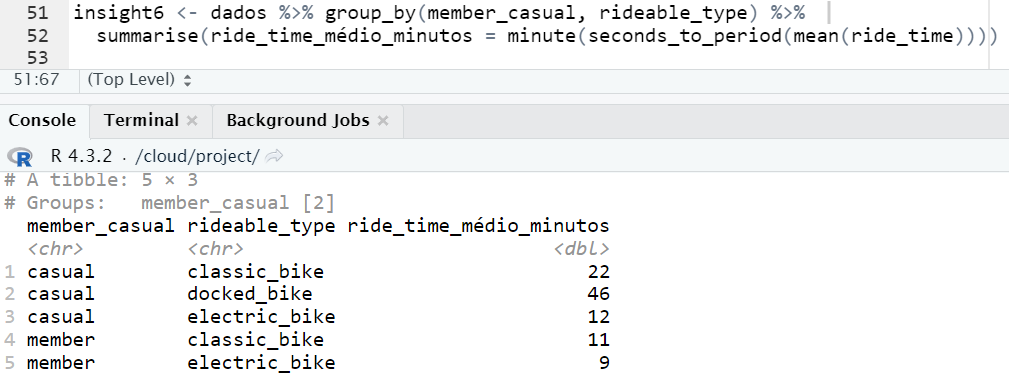
Após analisar os dados através da tabela dinâmica, resolvi passar esses mesmo insights para a linguagem R.



Após passar os mesmos insights para o R, fiz mais 2 tabelas com as seguintes informações: qual tipo de bike foi a mais usada por tipo de cliente e o tempo médio utilizado de cada bike por tipo de cliente.



Podemos concluir que os clientes casuais acabam utilizando a bike “electric\_bike” com mais frequências em seus passeios, já os clientes membros utilizam mais a bike “classic\_bike”, e um detalhe importante: nenhum cliente membro utilizou a bike “docked\_bike”.



Já nessa tabela podemos observar que a bike “docked\_bike” tem um tempo médio de corridas maior utilizada pelos clientes casuais, já os clientes membros utilizaram a bike “classic\_bike” por mais tempo.

Após analisar os dados, pude perceber algumas tendências e algumas relações importantes, até mesmo descobertas que podem impactar muito na entrega final, como por exemplo:

* Os clientes membros são os que mais utilizam o serviço, porém os clientes casuais passam **mais tempo** utilizando.
* Os dias em que mais se utiliza as bikes são de final de semana.
* Clientes casuais preferem a bike elétrica, porém clientes membros preferem a bike clássica.
* Os clientes casuais passam mais tempo utilizando a bike acoplada, já os clientes membros passam mais tempo na bike clássica.

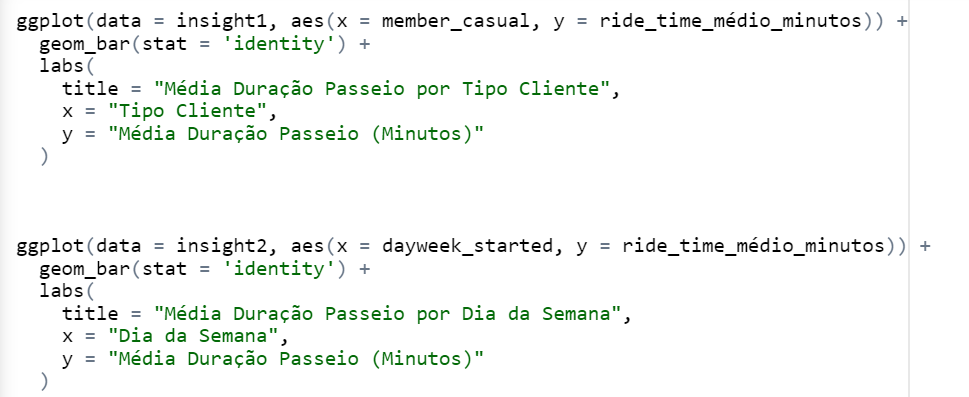
Esses insights são de extrema importância e tem uma grande relevância na hora de tomar decisões.

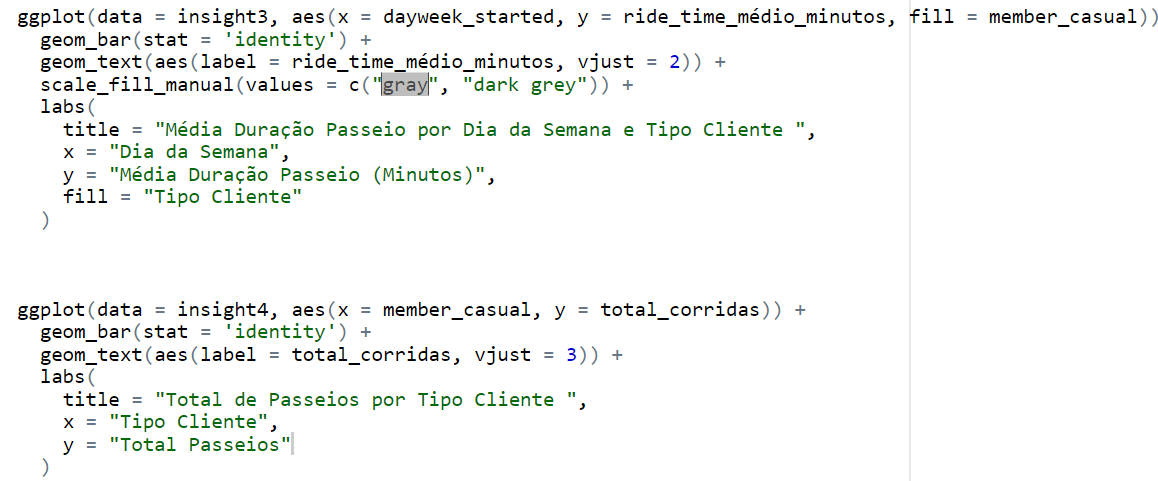
COMPARTILHAR

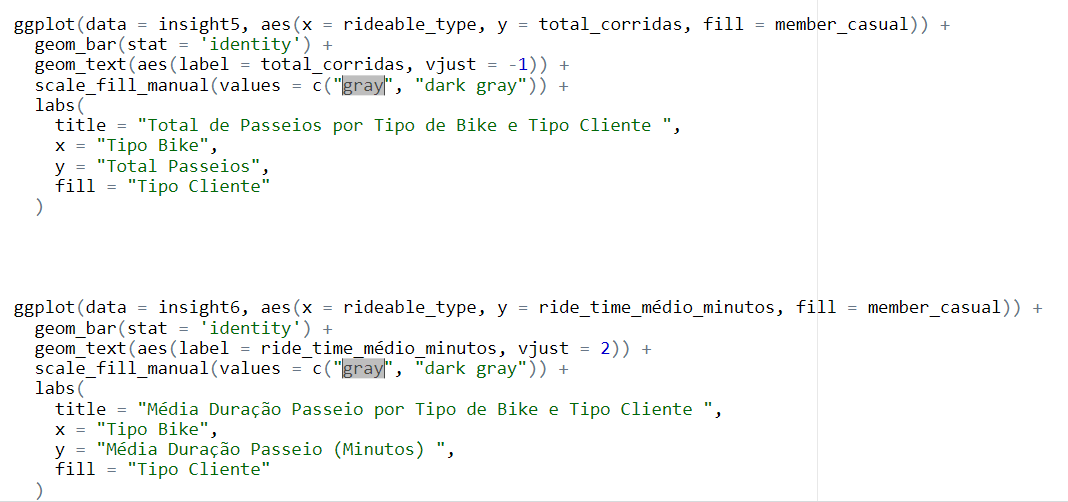
Após a etapa de análise, vamos compartilhar de forma fácil todos esses insights que foram descobertos. A melhor forma de apresentar os dados é com **gráficos**, eles são extremamentes poderosos na hora da comunicação. Com eles grandes massas de dados podem se tornar simples e rapidamente fácil de entender.

Utilizei a linguagem R na elaboração dos gráficos, também criei uma apresentação simples para mostrar os insights em slides.

Primeiramente criei os gráficos.







Com os gráficos criados e ajustados, fiz uma apresentação em slides para mostrar todos os insights que foram tirados em cima dos dados.

Segue link direto para a apresentação:

[Estudo de caso - Google](https://docs.google.com/presentation/u/0/d/13yeHwEiGHQ5klciVsen-K8ASrW710kQJPYY9ch22wG8/edit)

Com a apresentação criada, todos os pontos foram esclarecidos e apresentados com sucesso! Porém ainda tem a última parte do estudo de caso: **Ação.**

AGIR

Fiz a apresentação e mostrei todos os insights gerados durante a análise, agora é hora da famosa tomada de decisão!

Minha conclusão final é de que os clientes casuais utilizam a bike com mais frequência para passeios durante finais de semana, porém ainda sim utilizam durante seu dia-dia, isso nos mostra que para muitos deles o plano anual da Cyclistic seria uma ótima escolha. Já os clientes membros utilizam com uma frequência mais estável durante toda a semana, ou seja, o seu uso é mais moderado, a maioria utiliza para as tarefas do dia-dia.

Também tenho idéias para aqueles clientes que utilizam apenas aos finais de semana, um plano especial poderia ser criado para esses clientes no intuito de trazer mais benefícios a eles e a cyclistic ter mais clientes associados a planos. Por exemplo: as bikes acopladas são utilizadas por muito tempo pelos clientes casuais, o que faz com que eles gastem mais por isso, com um plano especial esse problema poderia acabar de uma vez por todas.

E é dessa forma que finalizo meu estudo de caso!