



Projeto: Análise da Saúde Financeira da Empresa Netflix

Data Science Analytics



The Netflix logo, featuring the word "NETFLIX" in white, bold, sans-serif capital letters with a 3D effect, set against a solid red rectangular background.

10 DE NOVEMBRO DE 2020

**WALTER TREVISAN
São Paulo - SP**



Sumário

1	Perfil da Empresa	2
2	Problema de Negócio	2
2.1	Clientes	2
2.2	Receita	2
2.3	Custos.....	3
2.4	Margem de Contribuição e Rendimentos	3
2.5	Inferência Estatística	3
3	Dataset.....	4
3.1	Descrição das Variáveis	4
4	Solução.....	10
4.1	Ferramenta	10
4.2	Fases da solução	10
4.3	Diretórios (pastas) de trabalho	11
5	Referência	12



1 Perfil da Empresa

Netflix é o maior serviço de entretenimento de streaming do mundo, com mais de **195 milhões** de *clientes pagantes* em **mais de 190 países**, desfrutando de séries de TV, documentários e filmes de longa-metragem numa grande variedade de géneros e línguas. Os clientes podem assistir o quanto quiserem, a qualquer hora, em qualquer lugar, em qualquer tela ligada à Internet. Os clientes podem reproduzir, pausar e recomeçar a assistir, tudo sem anúncios ou compromissos.

2 Problema de Negócio

O objetivo deste projeto é preparar um resumo estatístico dos dados da empresa “**Netflix**”, demonstrando sua saúde financeira e outras informações relevantes, que poderão ajudar potenciais investidores a decidir sobre investir ou não na companhia.

Então, vamos analisar o volume de “Clientes”, as “Receitas”, os “Custos”, os “Rendimentos” e os “Lucros/Margens” no período compreendido de **2012** até **2018** para o segmento “**Domestic Streaming**” (Transmissão Doméstica).

2.1 Clientes

- 1- **Analisar por ano:** o volume total de clientes, o volume de clientes atuais (que pagam suas assinaturas), o volume de clientes novos (em fase de avaliação gratuita) e a proporção de clientes novos em relação ao volume total de clientes;
- 2- **Analisar por trimestre:** o volume total de clientes, o volume de clientes atuais (que pagam suas assinaturas), o volume de clientes novos (em fase de avaliação gratuita) e a proporção de clientes novos em relação ao volume total de clientes.

2.2 Receita

- 1- **Analisar por ano:** as receitas totais e as receitas por cliente;
- 2- **Analisar por trimestre:** as receitas totais e as receitas por cliente.



2.3 Custos

- 1- **Analisar por ano:** os custos totais, os custos das receitas, os custos de marketing e os custos por cliente;
- 2- **Analisar por trimestre:** os custos totais, os custos das receitas, os custos de marketing e os custos por cliente.

2.4 Margem de Contribuição e Rendimentos

- 1- **Analisar por ano:** o lucro de contribuição, a margem de contribuição e os rendimentos por cliente;
- 2- **Analisar por trimestre:** o lucro de contribuição, a margem de contribuição e os rendimentos por cliente.

2.5 Inferência Estatística

- 1- Verificar se existe alguma relação de dependência (todo o período) entre a **receita total** da empresa e o volume “**total de clientes**”, o volume de “**clientes atuais**” e o volume de “**clientes novos**”;
- 2- Verificar se existe alguma relação de dependência (todo o período) entre o **custo total** da empresa e o volume “**total de clientes**”, o volume de “**clientes atuais**” e o volume de “**clientes novos**”;
- 3- Verificar se existe alguma relação de dependência (todo o período) entre o **custo da receita** da empresa e o volume “**total de clientes**”, o volume de “**clientes atuais**” e o volume de “**clientes novos**”;
- 4- Verificar se existe alguma relação de dependência (todo o período) entre o **custo de marketing** da empresa e o volume “**total de clientes**”, o volume de “**clientes atuais**” e o volume de “**clientes novos**”;
- 5- Verificar se existe alguma relação de dependência (todo o período) entre o **custo de marketing** com a **receita total** da empresa, com a **margem de contribuição** e com os **rendimentos por cliente**;
- 6- Considerando todo o período (2012 a 2018), o **volume total de clientes aumentou e/ou diminuiu?**
- 7- Considerando todo o período (2012 a 2018), o **volume de clientes novos aumentou e/ou diminuiu?**



- 8- Considerando todo o período (2012 a 2018), as **receitas (total) aumentaram e/ou diminuíram?**
- 9- Considerando todo o período (2012 a 2018), os **custos (total) aumentaram e/ou diminuíram?**
- 10- Considerando todo o período (2012 a 2018), os **custos com marketing aumentaram e/ou diminuíram?**
- 11- Considerando todo o período (2012 a 2018), as **margens de contribuição aumentaram e/ou diminuíram?**
- 12- Considerando todo o período (2012 a 2018), os **rendimentos por cliente aumentaram e/ou diminuíram?**

3 Dataset

Fonte dos dados:

<https://www.netflixinvestor.com/financials/quarterly-earnings/default.aspx>

No link acima obtemos os dados financeiros da **Netflix**, no período de **2012** até **2018**, e que foram armazenados no arquivo "**Netflix_Data.xlsx**".

3.1 Descrição das Variáveis

Temos um total de "**13**" variáveis em nosso dataset, descritas a seguir:

- 1- **Time**: é a data do último dia de cada trimestre de cada ano;
- 2- **Total subscriptions at end of period (Total de assinaturas no final do período)**: é o volume total de clientes (assinaturas), **acumulado**, no último dia de cada trimestre de cada ano. Portanto, o volume informado no dia **31/12** é o volume total de clientes em cada ano. É também a soma das "**Assinaturas pagas**" e "**Avaliações Gratuitas**";
- 3- **Paid subscriptions at end of period (Assinaturas pagas no final do período)**: é o volume total de clientes (assinaturas), **acumulado**, que pagaram suas assinaturas no último dia de cada trimestre de cada ano. Portanto, o volume informado no dia **31/12** é o volume total de clientes *pagantes* em cada ano;



4- Free Trials (Avaliações Gratuitas): é o volume total de clientes (assinaturas), **acumulado**, que estão avaliando gratuitamente o plano do segmento de “Transmissão Doméstica” (“Domestic Streaming”), no último dia de cada trimestre de cada ano, ou seja, estes clientes ainda não estão pagando suas assinaturas. Portanto, o volume informado no dia **31/12** é o volume total de clientes *não pagantes* em cada ano;

5- Revenue (Receita): é o valor total da receita no último dia de cada **trimestre** de cada ano.

Portanto, o **valor total da receita** em cada **ano** é a **soma** do valor total da receita em cada trimestre do respectivo ano, ou seja:

$$R_{ano} = R_{Q1} + R_{Q2} + R_{Q3} + R_{Q4}$$

6- Cost of revenues (Custos das receitas): é o valor total dos custos das receitas no último dia de cada **trimestre** de cada ano.

Portanto, o **valor total dos custos das receitas** em cada **ano** é a **soma** do valor total dos custos das receitas em cada trimestre do respectivo ano, ou seja:

$$CR_{ano} = CR_{Q1} + CR_{Q2} + CR_{Q3} + CR_{Q4}$$

7- Marketing: é o valor total dos custos de marketing no último dia de cada **trimestre** de cada ano.

Portanto, o **valor total dos custos de marketing** em cada **ano** é a **soma** do valor total dos custos de marketing em cada trimestre do respectivo ano, ou seja:

$$CM_{ano} = CM_{Q1} + CM_{Q2} + CM_{Q3} + CM_{Q4}$$



8- Contribution profit (Lucro de Contribuição): é calculado, no último dia de cada **trimestre** de cada ano, pela seguinte fórmula:

$$LC_{Qi} = R_{Qi} - CR_{Qi} - CM_{Qi}$$

Onde:

- LC_{Qi} : é o Lucro de Contribuição em cada trimestre (Q1, Q2, Q3 e Q4);
- R_{Qi} : é a Receita total em cada trimestre (Q1, Q2, Q3 e Q4).
- CR_{Qi} : é o Custo da Receita em cada trimestre (Q1, Q2, Q3 e Q4);
- CM_{Qi} : é o Custo de Marketing em cada trimestre (Q1, Q2, Q3 e Q4);

Portanto, o **valor total dos lucros de contribuição** em cada **ano** é a **soma** do valor total dos lucros de contribuição em cada trimestre do respectivo ano, ou seja:

$$LC_{ano} = LC_{Q1} + LC_{Q2} + LC_{Q3} + LC_{Q4}$$

9- Contribution Margin (Margem de Contribuição): é calculado, no último dia de cada **trimestre** de cada ano, pela seguinte fórmula:

$$MC_{Qi} = \frac{LC_{Qi}}{R_{Qi}}$$

Onde:

- MC_{Qi} : é a Margem de Contribuição em cada trimestre (Q1, Q2, Q3 e Q4);
- LC_{Qi} : é o Lucro de Contribuição em cada trimestre (Q1, Q2, Q3 e Q4);
- R_{Qi} : é a Receita total em cada trimestre (Q1, Q2, Q3 e Q4).



Portanto, a **margem de contribuição** em cada ano é a **soma** do valor total dos lucros de contribuição em cada trimestre, **dividido** pela **soma** do valor total da receita em cada trimestre do respectivo ano, ou seja:

$$MC_{ano} = \frac{\sum_{i=0}^4 LC_{Qi}}{\sum_{i=0}^4 R_{Qi}} = \frac{LC_{Q1} + LC_{Q2} + LC_{Q3} + LC_{Q4}}{R_{Q1} + R_{Q2} + R_{Q3} + R_{Q4}}$$

Onde:

- MC_{ano} : é a Margem de Contribuição Anual;
- LC_{Qi} : é o Lucro de Contribuição em cada trimestre (Q1, Q2, Q3 e Q4);
- R_{Qi} : é a Receita total em cada trimestre (Q1, Q2, Q3 e Q4).

Para maiores informações sobre “**Margem de Contribuição**” acesse este link:

https://pt.wikipedia.org/wiki/Margem_de_contribui%C3%A7%C3%A3o

10- Cost per Customer (Custo por Cliente), excluindo Marketing: é calculado, no último dia de cada **trimestre** de cada ano, pela seguinte fórmula:

$$CC_{Qi} = \frac{CR_{Qi}}{VTC_{Qi}}$$

Onde:

- CC_{Qi} : é o Custo por Cliente em cada trimestre (Q1, Q2, Q3 e Q4);
- CR_{Qi} : é o Custo da Receita em cada trimestre (Q1, Q2, Q3 e Q4);
- VTC_{Qi} : é o Volume Total de Clientes, **acumulado**, em cada trimestre (Q1, Q2, Q3 e Q4).



Portanto, o **custo por cliente** em cada **ano** é a **soma** do valor total dos custos das receitas em cada trimestre, **dividido** pelo volume total de clientes (assinaturas) informado no dia 31/12 (Q4), do respectivo ano, ou seja:

$$CC_{ano} = \frac{\sum_{i=0}^4 CR_{Qi}}{VTC_{Q4}} = \frac{CR_{Q1} + CR_{Q2} + CR_{Q3} + CR_{Q4}}{VTC_{Q4}}$$

Onde:

- CC_{ano} : é o Custo por Cliente Anual;
- CR_{Qi} : é o Custo da Receita em cada trimestre (Q1, Q2, Q3 e Q4);
- VTC_{Q4} : é o Volume Total de Clientes no dia 31/12 (Q4), do respectivo ano.

11- Revenue per Customer (Receita por Cliente): é calculado, no último dia de cada **trimestre** de cada ano, pela seguinte fórmula:

$$RC_{Qi} = \frac{R_{Qi}}{VTC_{Qi}}$$

Onde:

- RC_{Qi} : é a Receita por Cliente em cada trimestre (Q1, Q2, Q3 e Q4);
- R_{Qi} : é a Receita total em cada trimestre (Q1, Q2, Q3 e Q4);
- VTC_{Qi} : é o Volume Total de Clientes, **acumulado**, em cada trimestre (Q1, Q2, Q3 e Q4).

Portanto, a **receita por cliente** em cada **ano** é a **soma** do valor total das receitas em cada trimestre, **dividido** pelo volume total de clientes (assinaturas) informado no dia 31/12 (Q4), do respectivo ano, ou seja:

$$RC_{ano} = \frac{\sum_{i=0}^4 R_{Qi}}{VTC_{Q4}} = \frac{R_{Q1} + R_{Q2} + R_{Q3} + R_{Q4}}{VTC_{Q4}}$$



Onde:

- RC_{ano} : é a Receita por Cliente Anual;
- R_{Qi} : é a Receita total em cada trimestre (Q1, Q2, Q3 e Q4);
- VTC_{Q4} : é o Volume Total de Clientes, **acumulado**, no dia 31/12 (Q4), do respectivo ano.

12- Earnings per Customer (Ganhos/Rendimentos por Cliente): é calculado, no último dia de cada **trimestre** de cada ano, pela seguinte fórmula:

$$GC_{Qi} = RC_{Qi} - CC_{Qi}$$

Onde:

- GC_{Qi} : é o Ganho (Rendimento) por Cliente em cada trimestre (Q1, Q2, Q3 e Q4);
- RC_{Qi} : é a Receita por Cliente em cada trimestre (Q1, Q2, Q3 e Q4);
- CC_{Qi} : é o Custo por Cliente em cada trimestre (Q1, Q2, Q3 e Q4).

Portanto, os **ganhos (rendimentos) por cliente** em cada **ano**, é a receita por cliente **menos** o custo por cliente do respectivo ano, ou seja:

$$GC_{ano} = RC_{ano} - CC_{ano}$$

13- Segment (Segmento): é o tipo de segmento (negócio) ao qual se referem os dados. Neste caso é o segmento de “**Transmissão Doméstica**” (“**Domestic Streaming**”), ou simplesmente “**Streaming**”.



4 Solução

4.1 Ferramenta

Neste projeto utilizaremos a “**Linguagem R**” para o desenvolvimento da solução.

4.2 Fases da solução

A solução do projeto foi desenvolvida em “**05 fases**”:

Fase 1: Business Problem

Nesta fase temos uma explicação do “**problema de negócio**” a ser resolvido, assim como os **objetivos** a serem alcançados para o projeto.

Fase 2: Get The Data

Nesta fase iremos **obter** e **preparar** os **dados** que serão utilizados durante o desenvolvimento e execução do projeto, ou seja, nesta fase faremos todas as atividades relacionadas ao **pré-processamento dos dados**, como limpeza e transformação (“*Feature Engineering*”).

Fase 3: Explore The Data

Nesta fase faremos uma “**análise exploratória**” dos dados, ou seja, estudaremos cada atributo (variável) e suas características, e criaremos algumas visualizações (gráficos) dos dados.

Fase 4: Statistic Inference

Nesta fase realizaremos alguns **testes estatísticos** para fazermos algumas **inferências** sobre os dados. Estas inferências nos auxiliarão a responder algumas questões sobre a saúde financeira da empresa Netflix.

Fase Final: Story Telling

Nesta fase faremos uma apresentação da solução do projeto, explicando os resultados e insights encontrados, assim como as estratégias utilizadas no desenvolvimento do projeto.



4.3 Diretórios (pastas) de trabalho

A solução do projeto foi desenvolvida em “05 fases” (explicadas acima) sendo executada na seguinte “**estrutura de diretórios (pastas)**”:

- ❖ **BusinessProblem**: contém a descrição do projeto (este documento);
- ❖ **Datasets**: contém o conjunto de dados do projeto;
- ❖ **Solution**: contém todos os programas (códigos fontes – arquivos “**ipynb**”) de cada “**fase**” do projeto, a saber:
 - ✓ **01-Business-Problem.ipynb**: contém uma explicação do “**problema de negócio**” a ser resolvido, assim como os objetivos a serem alcançados para o projeto;
 - ✓ **02-Get-The-Data.ipynb**: contém as atividades relacionadas a **leitura e pré-processamento dos dados**, ou seja, limpeza e transformação (“**Feature Engineering**”) dos dados que serão utilizados nas próximas fases;
 - ✓ **03-Explore-The-Data.ipynb**: contém todas as atividades relacionadas a **análise exploratória dos dados**;
 - ✓ **04-Statistic-Inference.ipynb**: contém todos os **testes estatísticos** que deram o suporte necessário para fazermos a **inferência** sobre os dados analisados no projeto.

Nota: os programas devem ser executados nesta ordem:

- 01-Business-Problem.ipynb;
- 02-Get-The-Data.ipynb;
- 03-Explore-The-Data.ipynb
- 04-Statistic-Inference.ipynb.



Temos ainda outros arquivos (criados durante o desenvolvimento da solução) armazenados nos seguintes diretórios (pastas):

- ❖ **Data:** contém todos os objetos (*data frame*), criados na execução do projeto, que foram salvos em um formato específico (arquivos “**feather**”);
- ❖ **Images:** contém todos os gráficos/visualizações do projeto, separados por “**fase**”, além de algumas imagens utilizadas no projeto.
- ❖ **StoryTelling:** contém uma apresentação do projeto, explicando os resultados e insights encontrados, assim como as estratégias utilizadas no desenvolvimento do projeto.

5 Referência

Projeto da “**FAECD – Formação Análise Estatística Para Cientista de Dados**” da “**Data Science Academy**”.