# PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Pós-graduação Lato Sensu em Ciência de Dados e Big Data

**Elenise Ceres Figueira dos Santos** 

ANÁLISE DE DADOS DA LISTAGEM DE FILMES BRASILEIROS E ESTRANGEIROS EXIBIDOS DE 2009 A 2018 (ANCINE)

Belo Horizonte 2020

## **Elenise Ceres Figueira dos Santos**

# ANÁLISE DE DADOS DA LISTAGEM DE FILMES BRASILEIROS E ESTRANGEIROS EXIBIDOS DE 2009 A 2018 (ANCINE)

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados e Big Data como requisito parcial à obtenção do título de especialista.

Belo Horizonte 2020

Agradeço à Deus por nunca me faltar. À minha Mãe, que é minha inspiração de luta e dedicação e pelo apoio incondicional. À minha irmã Ana, por sempre me fortalecer.

# **SUMÁRIO**

1. Introdução	5
1.1. Contextualização	5
1.2. O problema proposto	5
2. Coleta de Dados	6
2.2. Visualização dos dados Importados	7
3. Processamento/Tratamento de Dados	7
3.1. Limpeza dos dados	8
4. Análise e Exploração dos Dados	10
5. Apresentação dos Resultados	12
6. Links	17
REFERÊNCIAS	18

#### 1. Introdução

#### 1.1. Contextualização

A análise dos dados está ganhando cada vez mais espaço no mundo atual, pois toda organização, órgãos públicos ou privados, instituições de ensino, entre outras, produzem dados, sejam eles em grande escala ou não. Com isso, surge a grande necessidade de se analisar esses dados, para que possamos identificar problemas e tomar decisões de forma mais assertiva possível.

O cinema é considerado a sétima arte, garantindo aos seus espectadores muita cultura, emoção e entretenimento, porém, no Brasil nem todos tem acesso a este artefato cultural. Os cinemas brasileiros podem ser empreendimentos com excelentes resultados, porém enfrentam muitos desafios para se manterem abertos.

É necessário interpretar o conjunto de dados da Ancine (Agência Nacional do Cinema), para o acompanhamento da evolução das salas de cinema brasileiro, e com isso, possamos obter insights e sanar problemas. O presente trabalho apresenta a análise dos dados da listagem de filmes brasileiros e estrangeiros exibidos de 2009 a 2018, demonstrando dados relevantes e a real situação dos filmes nos cinemas do Brasil.

#### 1.2. O problema proposto

O cinema proporciona a população que lhe frequenta, ou seja, que tem condições e interesse ao mesmo, uma arte criativa e muito discursiva que não se resume apenas em cultura ou diversão, pois movimenta a economia incentivada com o êxito dos filmes, e pode também possuir relação com a política. Uma exploração mais detalhada dos dados da Ancine é de suma importância, pois pode proporcionar incentivos futuros, como o apoio do governo a salas de cinema, patrocínio de produções, movimentando assim, milhões na indústria cinematográfica. Dessa

forma, podemos compreender os dados tão relevantes disponibilizados pelo governo federal.

Os dados foram coletados no site do Observatório Brasileiro do Cinema e do Audiovisual (OCA), que faz a divulgação e a difusão dos dados e informações geradas pela Agência Nacional do Cinema (Ancine), que é um órgão oficial do governo federal do Brasil.

O trabalho em questão tem como objetivo, fazer a análise da listagem de filmes brasileiros e estrangeiros exibidos de 2009 a 2018 para transformar os dados disponibilizados pela Ancine, em informação relevante, para que possamos compreender estes dados.

Os filmes da listagem dos arquivos em questão possuem diversos gêneros de uma infinidade de países.

Os dados apresentados nesta análise foram coletados entre os anos de 2009 a 2018, mostrando a evolução dos filmes em exibição nos cinemas do Brasil ao longo do tempo.

#### 2. Coleta de Dados

A Ancine é um órgão do governo federal brasileiro, sendo uma agência reguladora que fomenta, fiscaliza e regula o audiovisual e o mercado do cinema nacional. O Observatório Brasileiro do Cinema e do Audiovisual (OCA), criado em dezembro de 2008, tem por objetivo a difusão de dados e informações qualificadas produzidas pela Ancine. Dessa forma, sua atuação encontra-se em consonância com o objetivo estratégico da Ancine de "aprimorar a geração e disseminação de conhecimento do Setor" [1].

Os dados foram coletados no site do OCA, no link <a href="https://oca.ancine.gov.br/cinema">https://oca.ancine.gov.br/cinema</a>, dados obtidos em 30 de janeiro de 2020. Os arquivos usados para a análise são duas planilhas em Microsoft Excel com extensão .XLSX. Os conjuntos de dados utilizados possuem relação entre si, o segundo

descreve o público, renda e o número de lançamentos por tipo de distribuidora de 2009 a 2018 e é separado pelo ano vigente, e o primeiro lista cada filme exibido, visando mostrar a descrição de cada filme, sendo um o complemento do outro.

## 2.2. Visualização dos dados Importados

Com os dados já importados temos a oportunidade de visualizar as planilhas em sua forma de apresentação.

Ano de exibição	Título da obra	CPB/ROE	Gênero		s(es) s) da d obra d		Data d lançament		idora	Origem da P empresa ribuidora	úblico no Re ano de exibição	enda (R\$) no ano de exibição
2009	2012	E1600546500000	Ficção	Estados U	Jnidos Estr	rangeira	2009-11-1: 00:00:0			Distribuição ternacional	5056558.0	42950999.00
2009	007 Quantum of Solace	E1402431200000	Ficção	Estados U Ingl	nidos, Estr laterra	rangeira	2008-11-0 00:00:0			Distribuição ternacional	8672.0	44151.00
2009	1983 O Ano Azul	B0901024500000	Documentário		Brasil B	rasileira	2009-06-0 00:00:0		Filmes	Distribuição Nacional	2313.0	23001.98
2009	2 Filhos de Francisco: A História de Zezé Di C	B0500302600000	Ficção		Brasil B	rasileira	2005-08-1: 00:00:0			Distribuição ternacional	75.0	225.00
2009	23 Anos em 7 Segundos: 1977 - O Fim do Jejum C	B0901028800000	Documentário		Brasil B	rasileira	2009-06-2 00:00:0			Distribuição ternacional	1718.0	14936.00
		2042	2047	2045	2045		2044	2042	2042	2044	2040	2000
	Unnamed: 0	2018	2017	2016	2015		2014	2013	2012	2011	2010	2009
	Público Total	163454506.0	181226407.0	184327360.0	173022827.0	15561	2992.0 14	9518269.0	146598376.0	143206574.0	134836791.0	112670935.0
	Distribuidoras Internacionais	126033869.0	142915293.0	138788947.0	128774105.0	11238	1738.0	8870054.0	99623836.0	103541234.0	95484714.0	85996359.0
	Distribuidoras Nacionais	37297930.0	37909154.0	45537173.0	42611690.0	4323	1254.0 4	8389449.0	46549918.0	39657066.0	35485353.0	26358499.0
Codistribui	ção Internacionais - Nacionais	122707.0	401960.0	1240.0	1637032.0		NaN	2258766.0	424622.0	8274.0	3866724.0	316077.0
	Público - Filmes Brasileiros	24239873.0	17358513.0	30413839.0	22500563.0	1906	0705.0 2	7789804.0	15654862.0	17687772.0	25687438.0	16075429.0

#### 3. Processamento/Tratamento de Dados

O conjunto de dados estudados teve que passar por um tratamento, pois existiam registros com informações ausentes em ambos datasets, e nenhum valor duplicado que pôde ser constatado com o comando duplicate.

Quantidade de linhas e colunas do DataFrame (filmes e filmes2)

```
[] filmes.shape

[] (6414, 11)
```

```
[] filmes2.shape

[> (48, 11)
```

Verificando os dados ausentes existentes nas últimas linhas dos Data Frames:

## Data frame (filmes)

Ano de exibição	Título da obra	CPB/ROE	Gênero	País(es) produtor(es) da obra	Nacionalidade da obra	Data de lançamento	Distribuidora	Origem da empresa distribuidora	Público no ano de exibição	Renda (R\$) no ano de exibição
2018: ANCINE / SADIS (Sistema de Acompanhament	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2016 e 2017: ANCINE / SADIS (Sistema de Acompa	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2009 a 2015: ANCINE / SADIS (Sistema de Acompa	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Elaboração: Coordenação de	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

## Data frame (filmes2)

Unnamed: 0	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009
2018: ANCINE / SADIS (Sistema de Acompanhament	NaN									
2016 e 2017: ANCINE / SADIS (Sistema de Acompa	NaN									
2009 a 2015: ANCINE / SADIS (Sistema de Acompa	NaN									
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Elaboração: Coordenação de Edição e Publicação	NaN									

## 3.1. Limpeza dos dados

#### Removendo dados ausentes:

```
filmes.dropna (eixo = 0 , como = 'qualquer' , local = Verdadeiro )

filmes2.dropna (eixo = 0 , como = 'qualquer' , local = Verdadeiro )
```

## Verificando se existem dados nulos por coluna:

Verificando dados nulos em todo Data Frame:

```
filmes.isnull (). valores. qualquer ()
False

filmes2.isnull (). valores. qualquer ()
False
```

Quantidade de linhas e colunas depois da retirada dos dados ausentes:

```
filmes.shape
(6398, 11)
filmes2.shape
(27, 11)
```

Os valores numéricos que não existem, aparecem na forma de NaN. O método escolhido para tratar estes valores foi o isnull(), é uma técnica simples e fácil de ser aplicada e verificada depois.

```
filmesnovo = filmes['Data de lançamento'] != 'Relançamento'].copy()
```

Foi criado um novo Data Frame chamado de filmesnovo, para tratar a questão da palavra Relançamento na coluna Data de Lançamento, pois serão usados apenas os filmes lançados.

Convertendo a coluna Data de lançamento que era object em datetime:

```
filmesnovo [ 'Data de lançamento' ] = pd.to_datetime (filmesnovo [ 'Data de lançamento' ])
```

Renomeando as colunas do Data Frame (filmesnovo):

```
filmesnovo = filmesnovo.rename(columns={'Ano de exibição': 'Ano', 'País(es) produtor(es) da obra': 'País', 'Nacionalidade da obra': 'Nacionalidade',

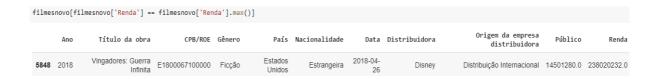
'Data de lançamento': 'Data', 'Público no ano de exibição': 'Público', 'Renda (R$) no ano de exibição': 'Renda'})
```

## 4. Análise e Exploração dos Dados

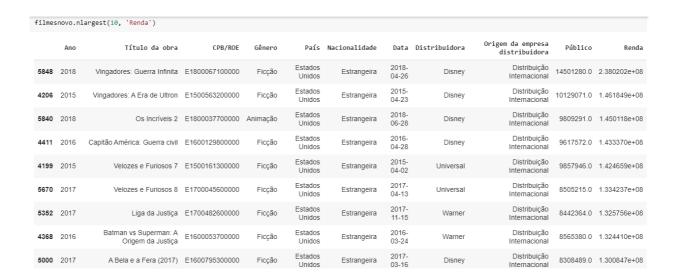
Para um maior ganho de produtividade e melhor conhecimento dos dados do Data Frame foi utilizado o pandas profile, que permitiu a criação de um relatório através da base de dados em questão, otimizando o processo.

Com a exploração dos dados foi possível verificar a maior renda e público nos cinemas brasileiros, assim como os piores resultados.

A maior renda de exibição de bilheteria:



As maiores rendas de exibição de bilheteria:



Maiores rendas de exibição de bilheterias de filmes brasileiros:

		lmesnovo[filmesnovo['Pa: gest(4, 'Renda')	ís'] == 'Brasil']								
	Ano	Título da obra	CPB/ROE	Gênero	País	Nacionalidade	Data	Distribuidora	Origem da empresa distribuidora	Público	Renda
6151	2018	Nada A Perder	B1800110100000	Ficção	Brasil	Brasileira	2018-03- 29	Downtown/Paris	Distribuição Nacional	12184373.0	1.209928e+08
4793	2016	Os Dez Mandamentos - O Filme	B1600036400000	Ficção	Brasil	Brasileira	2016-01- 28	Downtown/Paris	Distribuição Nacional	11305479.0	1.168330e+08
1051	2010	Tropa de elite 2	B1001296000000	Ficção	Brasil	Brasileira	2010-10- 08	Zazen	Distribuição Nacional	11023475.0	1.023201e+08
5394	2017	Minha mãe é uma peça 2	B1600716500000	Ficção	Brasil	Brasileira	2016-12- 22	Downtown/Paris	Distribuição Nacional	5213465.0	7.371323e+07

# Filmes que não tiveram renda de exibição de bilheteria:

filmes	snovo.n	smallest(25, 'Renda')									
	Ano	Título da obra	CPB/ROE	Gênero	País	Nacionalidade	Data	Distribuidora	Origem da empresa distribuidora	Público	Renda
100	2009	As Chaves de Casa	E1600598400000	Ficção	Itália	Estrangeira	2006- 01-06	Filmes do Estação	Distribuição Nacional	0.0	0.0
360	2009	Naufrágio - Mistério e Morte na Catástrofe do	B1101495000000	Documentário	Brasil	Brasileira	2009- 12-25	Master Shot Produções	Distribuição Nacional	0.0	0.0
488	2009	Santiago	B0700739200000	Documentário	Brasil	Brasileira	2007- 08-24	Videofilmes	Distribuição Nacional	0.0	0.0
563	2009	Vamos Subir, Leão	Sem CPB	Documentário	Brasil	Brasileira	2009- 03-20	Canal 3	Distribuição Nacional	0.0	0.0
727	2010	Cine Cocoricó: As Aventuras na Cidade	B0901055900000	Ficção	Brasil	Brasileira	2009- 07-17	Moviemobz	Distribuição Nacional	1.0	0.0
848	2010	Luto Como Mãe	B1101395900000	Documentário	Brasil	Brasileira	2010- 08-20	TV Zero Cinema	Distribuição Nacional	0.0	0.0
1017	2010	Simonal - Ninguém Sabe o Duro que Dei	B0901010000000	Documentário	Brasil	Brasileira	2009- 05-15	Moviemobz/RioFilme	Distribuição Nacional	1.0	0.0
1152	2011	A Última Estrada da Praia	B11013835	Ficção	Brasil	Brasileira	2011- 09-16	Okna Produções	Distribuição Nacional	0.0	0.0
1529	2011	Poema de Salvação	E1600211700000	Ficção	Argentina	Estrangeira	2011- 09-30	Canzión filmes	Distribuição Nacional	0.0	0.0
1533	2011	Porta a porta – A política em dois tempos	B1101435200000	Documentário	Brasil	Brasileira	2011- 12-02	Zéfiro Produções	Distribuição Nacional	0.0	0.0
1654	2011	Walachai	B1001193200000	Documentário	Brasil	Brasileira	2011- 11-18	Ciclorama	Distribuição Nacional	0.0	0.0
1749	2012	Apenas entre Nós	E1600010300000	Ficção	França, Eslovênia, Sérvia, Croácia	Estrangeira	2015- 05-07	Lume	Distribuição Nacional	0.0	0.0
1773	2012	Ataque ao prédio	E1500718800000	Ficção	Reino Unido	Estrangeira	2012- 01-13	Sony	Distribuição Internacional	0.0	0.0
1811	2012	Clementina de Jesus - Rainha Quelé	B12016413	Documentário	Brasil	Brasileira	2012- 11-15	Werinton Kermes Telles Marsal	Distribuição Nacional	0.0	0.0
1933	2012	Luto em Luta	B12016507	Documentário	Brasil	Brasileira	2012- 09-21	Like Filmes	Distribuição Nacional	0.0	0.0
1961	2012	Movimento Browniano	Sem ROE	Ficção	Holanda	Estrangeira	2012- 07-06	Petrini	Distribuição Nacional	0.0	0.0
1975	2012	O abrigo	E1402401600000	Ficção	Estados Unidos	Estrangeira	2012- 06-01	Sony	Distribuição Internacional	0.0	0.0

2066	2012	Ponto Org	B12015664	Ficção	Brasil	Brasileira	2012- 09-21	Usina Digital	Distribuição Nacional	0.0	0.0
2277	2013	As horas vulgares	B1301860500000	Ficção	Brasil	Brasileira	2013- 08-09	Petrini	Distribuição Nacional	0.0	0.0
2294	2013	Beije-me outra vez	Sem ROE	Ficção	Itália	Estrangeira	2013- 05-10	Petrini	Distribuição Nacional	0.0	0.0
2341	2013	Crazy Horse	Sem ROE	Ficção	Estados Unidos	Estrangeira	2013- 06-28	Petrini	Distribuição Nacional	0.0	0.0
2626	2013	Quase um tango	B1301916600000	Ficção	Brasil	Brasileira	2013- 08-16	NGM Produções	Distribuição Nacional	0.0	0.0
2695	2013	Todos os Dias	E1500844200000	Ficção	Reino Unido	Estrangeira	2014- 01-24	Esfera Produções	Distribuição Nacional	0.0	0.0
2763	2014	A Batalha do passinho	B1301805800000	Documentário	Brasil	Brasileira	2013- 10-11	Cine Santa Teresa	Distribuição Nacional	355.0	0.0
3094	2014	Mar negro	B1302025600000	Ficção	Brasil	Brasileira	2014- 01-17	Petrini	Distribuição Nacional	0.0	0.0

Com isso, foi possível analisar que vinte e cinco filmes não tiveram renda e que alguns até tiveram público porém não pagantes.

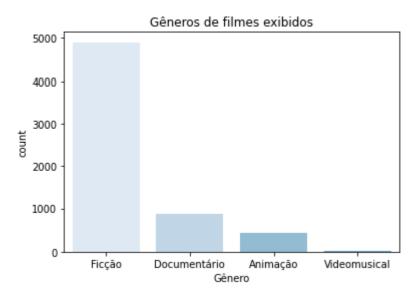
Após a exploração dos dados, foi possível perceber a forte correlação entre a renda e o público, sendo a renda totalmente dependente do público. Detectou-se também que a renda de exibição e o público, vieram em uma crescente desde o ano de 2009, tendo seu ápice em 2016 e decaindo nos anos seguintes, o que pode nos levar a várias linhas de observação. Nos últimos anos houve, não só no Brasil, mas no mundo, um aumento expressivo de usuários das plataformas streaming, que pode ser sim, uma das justificativas da diminuição de público nos cinemas. Outros motivos podem estar ligados tanto a situação financeira do cidadão brasileiro, quanto ao aumento de sites que fornecem filmes de forma ilegal ou até mesmo a desvalorização da cultura.

O ano de 2018 apresentou o maior número de lançamentos de filmes brasileiros até então vistos na análise, porém os filmes brasileiros ainda não alcançam um volume grande de público e renda como o dos filmes estrangeiros.

O gênero mais visto nas salas de cinema do Brasil é o de ficção, sendo o filme de maior bilheteria o Vingadores: Guerra Infinita, exibido no ano de 2018, com renda de R\$ 238.020.232,00, já o filme brasileiro de maior bilheteria, é o Nada a Perder, de 2018, e renda de R\$120.992.794,00.

#### 5. Apresentação dos Resultados

De acordo com a análise efetuada, constatou-se que o gênero mais visto nos cinemas brasileiros foi o de ficção e o menos exibido é o vídeo musical.



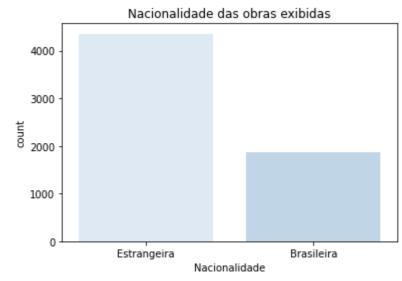
Já a maioria dos filmes lançados tem origem nos Estados Unidos e em segundo lugar o Brasil, como podemos ver a seguir:

País Categ	jorical
HIGH	CARDINALITY

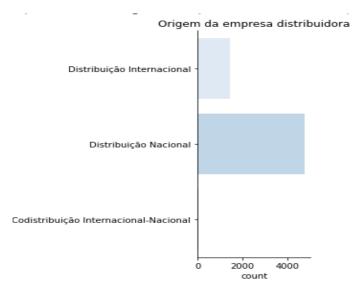
Distinct count	419
Unique (%)	6.7%
Missing	0
Missing (%)	0.0%
Memory size	48.7 KiB



A nacionalidade das obras, tem em sua maioria, origem estrangeira, porém, demonstra um número considerável de obras brasileiras, num total de 1.867 obras contra 4.356 estrangeiras.



As empresas distribuidoras dos filmes em questão, tem sua origem, em minoria, na distribuição internacional e a grande parte fica com as distribuidoras nacionais.



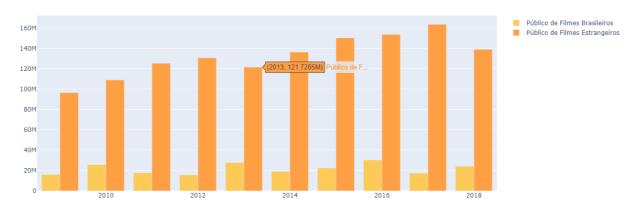
Verificou-se a distribuição de público ao longo do tempo, constatando o maior público no ano de 2016 e uma decrescente nos anos seguintes.

Público de Filmes Brasileiros e Estrangeiros



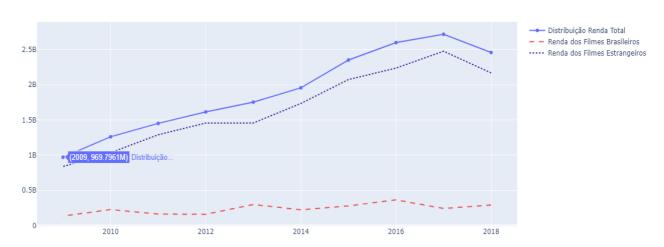
No gráfico a seguir podemos ver melhor a diferença entre o público de filmes brasileiros e estrangeiros.

Público de Filmes Brasileiros e Estrangeiros

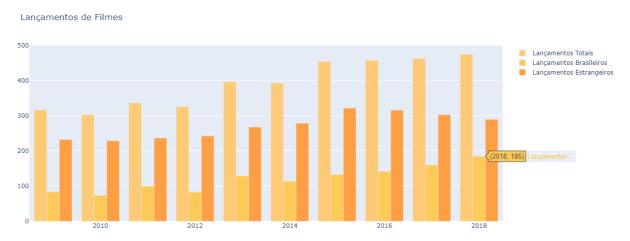


No próximo gráfico é identificado a distribuição de renda dos filmes exibidos nos cinemas brasileiros entre os anos de 2009 até 2018.

Renda de Filmes Exibidos nos Cinemas do Brasil

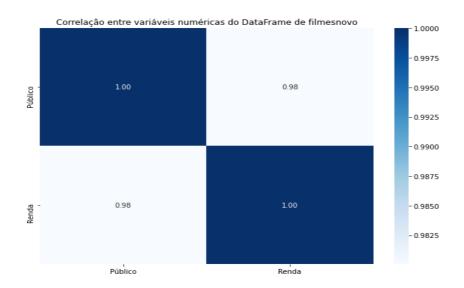


Os lançamentos dos filmes executados, aparecem de forma que visualizamos a quantidade dos lançamentos, com destaque para o ano de 2018 que teve o maior número de filmes brasileiros lançados, chegando num total de 185.

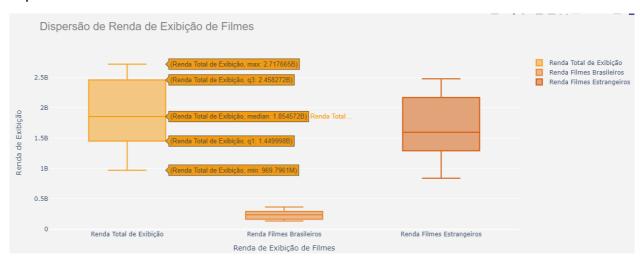


Desta forma, tive a possibilidade de verificar que os filmes brasileiros ainda são pouco vistos pela sua população e com isso, possuem renda muito abaixo dos filmes estrangeiros.

A seguir vemos a correlação das variáveis renda e público, temos a possibilidade de verificar que as mesmas têm uma forte correlação positiva.



A próxima representação gráfica é o boxplot e apresenta as estatísticas descritivas para a variável renda, mostrando os maiores valores obtidos, tanto para renda total quanto para a renda de filmes brasileiros e filmes estrangeiros separadamente.



# 6. Links

Repositório Github

https://github.com/Eleniseceres/TCC-Puc-Minas

# **REFERÊNCIAS**

[1] https://oca.ancine.gov.br/sobre-o-oca

AMARAL, Fernando. **Introdução à Ciência de Dados.** Alta Books, 2016.