

**UNIVERSIDADE ESTADUAL PAULISTA "JULIO DE MESQUITA FILHO"**

FACULDADE DE CIENCIAS - CAMPUS BAURU

DEPARTAMENTO DE COMPUTACAO

BACHARELADO EM CIENCIA DA COMPUTACAO

**CIÊNCIA DE DADOS APLICADA A CINEMA DE PEQUENO  
PORTE**

Nome: Bruno Belluzzo	RA: 161021531
Nome do Orientador: Prof. Assoc. Dr. Joao Pedro Albino	Assinatura:

BRUNO BELLUZZO

## **CIÊNCIA DE DADOS APLICADA A CINEMA DE PEQUENO PORTE**

**Proposta para Trabalho de Conclusão de Curso  
do Curso de Bacharelado em Ciência da  
Computação** da Universidade Estadual Paulista  
Júlio de Mesquita Filho, Faculdade de Ciências,  
Campus Bauru.

Orientador: Prof. Assoc. Dr. Joao Pedro Albino

## RESUMO

Apesar do grande impacto na economia mundial e do sucesso que são os cinemas ao redor do mundo, com o crescimento de serviços de vídeo online as pessoas, cada vez mais, estão deixando de ir ao cinema, pois, eventualmente, o filme estará disponível online. Em meio a toda essa revolução cinematográfica, cinemas de pequeno porte, que possuem uma única sala para exibição de filmes, sofrem com a disponibilidade de cópias, pois os estúdios e suas distribuidoras dão preferência às grandes redes de cinemas, pois estas geram maior receita. Procurando entender as limitações desses pequenos cinemas, este projeto busca responder a principal questão de pesquisa: *Como as técnicas de ciência de dados podem auxiliar na sobrevivência dos cinemas de pequeno porte?* Para tanto, o trabalho irá utilizar técnicas de ciência de dados e **machine learning** para levantar dados por meio de um *survey* sobre os frequentadores do cinema com o intuito de compreender o *perfil desses clientes* e buscar sugerir soluções inteligentes com o intuito de maximizar o retorno financeiro do estabelecimento, além de aprimorar o atendimento e a satisfação deste público. A pesquisa (*survey*) será realizada utilizando um formulário de pesquisa disponível na internet para os usuários do *Cine Belluzzo*, um cinema de pequeno porte situado na cidade de Bariri, interior do Estado de São Paulo, com cerca de 35 mil habitantes.

**Palavras-Chave:** Ciência de Dados, , Survey, Pesquisa de opinião, Cinema Belluzzo, Cinema de Pequeno Porte, Big Data, Aprendizado de Máquina.

## 1. INTRODUÇÃO

**Data Science** ou **Ciência de Dados** é uma área que já existe há mais de 30 anos, mas vem ganhando destaque nos últimos anos, devido ao **Big Data**. Segundo Chen, Mao e Liu (2014), o termo *big data* refere-se a *um grande conjunto de dados gerados e armazenados* com os quais os aplicativos de processamento de dados tradicionais não conseguem lidar em um tempo tolerável. O surgimento do termo está relacionado com o aumento exponencial da quantidade de dados gerados a cada minuto no mundo. Ainda segundo os autores, o *big data* representou uma nova era na sociedade moderna, em que *os dados se tornaram cada vez mais valiosos*, mudando a forma como a economia e a ciência observam os processos, extraem e geram valor desse **caos** de dados.

De acordo com Silveira (2016), ciência de dados (em inglês: data science) pode ser definida como uma área interdisciplinar voltada para o **estudo e análise de dados**, estejam estes dados estruturados ou não, e que busca a extração de conhecimento ou *insights*<sup>1</sup> para possíveis tomadas de decisão. Ainda de acordo com Silveira (2016), a ciência de dados pode transformar *uma grande quantidade de dados brutos (big data)* em inteligência de negócios, e com isso, auxiliar organizações em tomadas de decisões para que as mesmas obtenham melhores resultados.

O desenvolvimento de áreas como *machine learning* - uma área da ciência da computação que significa “aprendizado da máquina” e que por sua vez embute o conceito de **inteligência artificial**, que estuda meios para que os computadores possam desenvolver tarefas que seriam executadas por pessoas - reforçam o crescimento e a importância da Ciência de Dados sendo que é cada vez mais comum as organizações de qualquer ramo da atividade econômica se beneficiarem do estudo para a tomada de decisões, de forma a alavancar os seus crescimentos (SILVEIRA, 2016).

Trabalhos publicados entre 2010 e 2018 atestam, de acordo com o INSTITUTO ATLANTICO (2018), que *dados são o novo petróleo e o grande recurso desta década*. Nos últimos dez anos (de 1997 a 2017), o mundo gerou uma quantidade muito grande de informação, e as organizações têm interesse nesse recurso como forma de aprimorar os seus serviços (INSTITUTO ATLANTICO, 2018). Os mais diferentes tipos de negócios podem utilizar os benefícios da Ciência de Dados para alavancar seus negócios, e isto também não seria diferente com as empresas do ramo do entretenimento, mais especificamente, como discutido neste trabalho, do ramo cinematográfico.

A indústria cinematográfica mundial gera uma receita bilionária todos os anos. Segundo IstoÉ (2018), os filmes registraram, em 2018, um lucro recorde de 96,8 bilhões de dólares em todo o mundo, incluindo mais de 40 bilhões apenas em bilheterias, de acordo com números publicados pela *Motion Picture Association of America* (MPAA).

No Brasil, números levantados pela *Motion Picture Association na América Latina* (MPA-AL) em parceria com o *Sindicato Interestadual da Indústria do Audiovisual*, mostram que a indústria cinematográfica injetou em 2018 mais de R\$19 bilhões na economia brasileira. (ISTOÉ, 2018).

Apesar do grande impacto na economia mundial e do sucesso dos cinemas ao redor do mundo, com o crescimento de serviços de vídeo online as pessoas estão cada vez mais

---

<sup>1</sup> Insight é um substantivo do idioma inglês e que significa compreensão súbita de alguma coisa ou determinada situação. Está também está relacionado com a capacidade de discernimento, indicando um momento único de esclarecimento. Pode também ser entendido como um acontecimento cognitivo associado a vários fenômenos podendo ser sinônimo de compreensão, conhecimento, intuição ou a capacidade de apreender alguma coisa e acontece quando uma solução surge de forma repentina. Fonte: <https://www.significados.com.br/insight/>.

deixando de ir ao cinema, pois "uma hora ou outra" o filme estará disponível online. Segundo Exame (2017), na sétima edição da CinemaCon, o principal evento de exibição cinematográfica dos Estados Unidos, em 2017, teve como tema informal o futuro do negócio da exibição cinematográfica na era do streaming<sup>2</sup>.

Muitos estúdios já começaram versões *online* de seus canais, afirma IstoÉ (2018), contudo é preciso balancear as coisas com os donos das salas de cinema, que sempre foram fiéis aliados dos grandes estúdios cinematográficos.

Em meio a toda essa revolução cinematográfica, cinemas de pequeno porte, que possuem uma única sala para exibição de filmes, sofrem com a disponibilidade de cópias, pois os estúdios e suas distribuidoras dão preferência aos grandes cinemas, que irão gerar uma maior receita.

Com isso em mente e buscando entender as limitações desses pequenos cinemas, a proposta deste projeto é a de discutir e apresentar como as técnicas de ciência de dados e *machine learning* podem ser utilizadas para, em tempos de concorrência com grandes conglomerados de *streaming de vídeo*, manter funcionando cinemas em pequenas cidades.

A pesquisa será realizada utilizando um formulário eletrônico de pesquisa, disponível na internet, para os usuários do *Cine Belluzzo*, um cinema de pequeno porte situado na cidade de Bariri, interior do Estado de São Paulo, com cerca de 35 mil habitantes.

O questionário buscara coletar, através de questões abertas e fechadas, informações buscando compreender o perfil das pessoas que frequentam o cinema, os tipos e categorias de filmes que estas pessoas preferem assistir, além de verificar a opinião delas à respeito da estrutura física do cinema (localização, horário e dias de funcionamento, dentre outras opiniões).

## 2. PROBLEMA DE PESQUISA

As limitações físicas e financeiras de um cinema de pequeno porte, assim como o número de habitantes da cidade em que normalmente estas salas estão localizadas, dificultam a possibilidade destas organizações em obter ganhos e se manter em funcionamento.

As distribuidoras de filmes ficam com uma parcela da renda da venda dos ingressos, portanto, de acordo com IstoÉ (2018), é mais vantajoso para estas organizações que utilizem redes de cinema de grande porte, que possuem público muito maior e que poderão gerar maior faturamento.

Além do mais, pelo fato das distribuidoras possuírem cópias limitadas dos filmes, os pequenos cinemas sofrem também um relativo atraso no oferecimento de novos filmes em relação aos cinemas de maior porte, exibindo filmes com diferença de uma ou duas semanas, além de, por normalmente possuírem apenas uma sala para exibição, terem um limite de dois ou no máximo três filmes sendo exibidos simultaneamente.

Por todas estas limitações apontadas anteriormente, a possibilidade de utilização de técnicas de ciência de dados podem ser benéfica para compreender as preferências e os aspirações das pessoas que frequentam cinemas de pequeno porte, e assim apoiar estas pequenas organizações, geralmente familiares, a tomar melhores decisões no que tange a escolha de filmes, oferecimento de promoções, ofertas combinadas na compra de

---

<sup>2</sup> **Streaming** significa ouvir música ou assistir a um vídeo em 'tempo real', em vez de baixar um arquivo no computador e vê-lo mais tarde. Com vídeos na Internet e webcasts de eventos ao vivo, não há nenhum arquivo para download, apenas um fluxo contínuo de dados. Fonte: <https://www.palpitedigital.com/que-streaming-quais-melhores-audio-video/>.

ingressos e produtos oferecidos para consumo no cinema (combos<sup>3</sup>), entre outras escolhas. Os algoritmos de ciência de dados podem também auxiliar na busca do entendimento dos motivos pelos quais as pessoas não irem ao cinema, elaborando soluções inteligentes para atrair essas pessoas para frente da "grande tela". Portanto, a questão de pesquisa proposta nesta proposta de trabalho é:

***Como as técnicas de ciência de dados podem auxiliar na sobrevivência dos cinemas de pequeno porte ?***

### 3. JUSTIFICATIVA

O conhecimento obtido a partir do uso de Ciência de Dados é fundamental para apoiar um processo de tomada de decisão. A informação gerada pelas aplicações informáticas pode oferecer aos gestores de cinemas de pequeno porte *um conjunto de indicadores sobre o negócio*, que lhe permitirão inferir indicações do que aconteceu no passado da organização e *permitirem traçar cenários para ações futuras* (SANTOS; RAMOS, 2006).

“As empresas que possuem ferramentas de *business intelligence* e a utilizam em seus processos durante uma tomada de decisão, apresentam vantagem, para se posicionarem a frente de uma nova oportunidade no mercado” (SILVA; TERRA, 2015, p. 11).

O termo *inteligência de negócios* (ou *Business Intelligence*, em inglês) refere-se ao processo de coleta, organização, análise, compartilhamento e monitoramento de informações que oferecem suporte a gestão de negócios, afirma Rud (2009). É um conjunto de técnicas e ferramentas para auxiliar na transformação de dados brutos em informações significativas e úteis a fim de analisar o negócio.

Identificando novas oportunidades e implementando uma estratégia efetiva baseada nos dados, também pode-se alavancar negócios com vantagem competitiva no mercado e estabilidade a longo prazo. (SILVA; TERRA, 2015).

No ramo do entretenimento esse conhecimento pode ser de extrema importância para atrair a maior quantidade possível de pessoas ao estabelecimento, assim como elaborar as melhores estratégias para garantir o maior lucro possível. (EXAME, 2017)

Com as limitações dos cinemas de pequeno porte citadas anteriormente, as tomadas de decisões e ações devem ser bem refletidas antes de serem executadas e, possuindo informações coletadas por meio de levantamento dos dados, além de levar em consideração experiências do passado, seria possível realizar melhores escolhas para evitar perdas que poderão levar à falência do empreendimento.

### 4. OBJETIVOS

#### 4.1. Objetivo geral

Utilizar ciência de dados e seus ferramentais, tais como *machine learning* e *estatística*, para compreender os padrões dos clientes de um cinema de pequeno porte e, com base nos dados levantados, buscar as melhores decisões e criar soluções inteligentes com o intuito de maximizar os lucros do estabelecimento.

#### 4.2. Objetivos específicos

- Levantar dados por meio de uma pesquisa sobre os frequentadores do cinema

---

<sup>3</sup> **Combo** é uma palavra em inglês e significa uma abreviação do termo **combination**. Em português pode ter os significados de *sequência* ou *combinação*. No caso específico dos cinemas, significa o oferecimento de: pipoca + refrigerante. Fonte: <https://pt.wikipedia.org/wiki/Combo>.

- Estudar os dados levantados e gerar gráficos para compreender o perfil dos clientes
- Elaborar soluções inteligente com base nos dados coletados
- Aplicar as soluções em um cinema de pequeno porte
- Avaliar o desempenho das receitas do cinema com as soluções aplicadas.

## 5. METODOLOGIA DE PESQUISA

O caso de estudo desta proposta de trabalho de conclusão de curso (TCC) é um cinema de pequeno porte, denominado o *Cine Belluzzo*, situado na cidade de Bariri, no estado de São Paulo, um pequeno município com cerca de 35 mil habitantes<sup>4</sup>. Portanto, para o desenvolvimento deste projeto de pesquisa e desenvolvimento do projeto são propostas as etapas arroladas neste tópico 5.

### 5.1. O Caso de Estudo

Segundo Yin (1993), **caso de estudo** ou **estudo de caso** são expressões sinônimas que designam um método da abordagem de pesquisa e que consiste na utilização de um ou mais métodos **qualitativos** de coleta de informação. O estudo de caso caracteriza-se por descrever um evento ou caso de uma forma longitudinal. O **caso** consiste geralmente no estudo aprofundado de uma **unidade individual**, tal como: uma pessoa, um grupo de pessoas, **uma instituição**, um evento cultural, etc.

Como afirmado no parágrafo anterior, o caso de estudo é um estabelecimento comercial (uma sala) para a exibição de filmes. Esta empresa, familiar, existe há mais de 50 anos em Bariri, e tem passado de geração em geração.

### 5.2. Questão/Hipótese de Pesquisa

A principal questão de pesquisa a ser respondida neste trabalho é:

*Como as técnicas de ciência de dados podem auxiliar na sobrevivência dos cinemas de pequeno porte?*

Em vista desta pergunta de pesquisa, as principais hipóteses a serem esclarecidas nesta proposta são:

**H<sub>1</sub>:** O uso de ciência de dados apoia a sobrevivência de cinemas de pequeno porte;

**H<sub>2</sub>:** O perfil atual de frequentadores de cinemas de pequeno porte determina a categoria dos filmes a serem exibidos;

**H<sub>3</sub>:** Promoções, tais como **combos**, incentivam a frequência das pessoas nos cinemas de pequeno porte.

## 5.3. Método de Pesquisa

### 5.3.1. Elaboração do Survey

Na primeira etapa deste trabalho, o método de pesquisa utilizado será o **survey**. De acordo com Babbie (2001); e Freitas et. ali (2000); *survey* é um método de pesquisa social que utiliza técnicas estatísticas e que normalmente implica na construção de investigações por meio de questionários. Os pesquisadores podem utilizar a informação por este método para *desenhar estratégias e estabelecer planos de ação*.

Segundo Babbie (2001), as principais características da pesquisa do tipo *survey* são:

<sup>4</sup> IBGE Censo [2010]. Fonte: <https://cidades.ibge.gov.br/brasil/sp/bariri/panorama>.

- Sempre que o pesquisador procura explicar as razões para e as fontes de eventos, características e correlações observados, a pesquisa deve assumir uma postura *determinística* baseada nas *relações de causa e efeito*;
- **Surveys amostrais** são realizados para entender-se a população maior da qual a amostra foi inicialmente selecionada. Análises explicativas em pesquisas de **survey** visam desenvolver proposições gerais sobre o comportamento humano;
- Replicar um achado entre subgrupos diferentes (por sexo, por idade, etc.) fortalece a certeza de que ele representa um fenômeno geral na sociedade;
- Com a realização deste tipo de pesquisa, busca-se *o máximo de compreensão com o menor número de variáveis possível*;
- A conceituação e a medição de variáveis estão no âmago da prática científica. Se as variáveis não são adequadamente conceituadas e medidas, correlações observadas entre elas podem não fazer sentido.

Em resumo, sempre que um pesquisador estiver interessado em identificar opiniões, atitudes, valores, percepções, auxiliar gestores na formulação de políticas e informar o planejamento estratégico de empresas, utiliza-se a técnica do **survey**.

O instrumento para a obtenção dos dados a ser utilizado será o **questionário**. O questionário é um dos instrumentos de geração de dados mais conhecido e utilizado em pesquisas científicas nas mais diversas áreas. De abordagem mais quantitativa, destaca-se a rapidez no preenchimento das respostas e largo alcance, podendo produzir dados referentes a populações muito grandes, afirmam Freitas et. ali (2000).

Para a criação e distribuição do formulário de pesquisa será utilizado o aplicativo *Google Forms*. O *Google Forms* é uma ferramenta que permite coletar opiniões das pessoas por meio de um questionário personalizado. As informações serão coletadas e automaticamente armazenadas em uma planilha. A planilha é automaticamente preenchida com as respostas da pesquisa ou do questionário.

As questões do formulário eletrônico, visam verificar as hipóteses estabelecidas no tópico 5.1, e serão compostas de **perguntas de respostas fechadas** - são aquelas nas quais o respondente seleciona uma ou diversas alternativas dentre um conjunto de alternativas pré-definidas -; e **perguntas de resposta aberta** - respostas fornecidas em forma de texto. Após a conclusão desta etapa, a planilha será utilizada como entrada para a realização da análise e interpretação dos dados.

**Para esta fase da pesquisa espera-se conseguir ao menos 100 participantes dentre os frequentadores do cinema.**

**Para incentivar a participação na pesquisa, serão realizados sorteios de ingressos ao cinema àqueles que acessarem o link do questionário, lerem o Termo de Consentimento Livre e Esclarecido (TCLE) apresentado na primeira página do formulário e aceitarem responder ao questionário.**

### 5.3.2 Análise e Inferência dos Dados

A **segunda etapa deste trabalho** consistirá em estudar as bibliotecas disponíveis para a linguagem Python e suas ferramentas para o uso em aplicações de Ciência de Dados, tais como: *Matplotlib* e *Seaborn* para visualização dos dados, e *Pandas* para análise dos dados.

**Esta fase da pesquisa ocorrerá em paralelo com a etapa de resposta dos questionários.**

Após o prazo estipulado para a realização do survey e com as planilhas de respostas em mãos, serão utilizadas as bibliotecas Python citadas para fazer a análise detalhada dos dados e seus padrões, para enfim inferir as melhores se as hipóteses foram respondidas e



assim sugerir estratégias e soluções para a organização. Em seguida, com as estratégias estabelecidas, aplica-las na prática.

Caso sejam constatadas situações em que inteligência artificial possa realizar o trabalho de tomada de decisões, como, por exemplo, a escolha do melhor filme entre os disponíveis pelas distribuidoras, será realizado também um estudo sobre *Machine Learning* (aprendizado de máquina) e alguns de seus algoritmos, como árvores de decisão e classificação *Naive Bayes*.

Para a implementação de IA serão utilizadas as bibliotecas *TensorFlow*, *SciKit Learn* e *Keras*, disponíveis em Python para *Machine Learning*.

**Python** foi escolhida por ser uma das linguagens mais utilizadas para realizar a análise de dados em pesquisas, além de possuir ferramentas que facilitam o estudo e a implementação de Ciência de Dados e Machine Learning, e possuir uma comunidade colaborativa muito ativa (BABU L, 2019).

O equipamento usado para a pesquisa será um notebook com sistema operacional Microsoft Windows e placa de vídeo NVIDIA.

Para o estudo e o desenvolvimento dos modelos estatísticos, a planilha do Google Docs deverá ser lida pela linguagem Python.

### **5.3.3. Survey Após Aplicação das Sugestões**

Na **terceira etapa desta pesquisa** um segundo survey será realizado para verificar se, com as modificações e soluções sugeridas pela análise dos dados foram aplicadas e percebidas.

Nesta segunda survey o que se procura estabelecer é se a satisfação dos clientes com a estrutura do cinema melhorou e se, conseqüentemente, houve também melhora no retorno financeiro para a organização, comparando-se então os resultados obtidos com os dados anteriores à realização da primeira pesquisa.

**Para esta fase da pesquisa espera-se também conseguir ao menos 100 participantes dentre os frequentadores do cinema.**

**Para incentivar a participação das pessoas na pesquisa, serão realizados sorteios de ingressos ao cinema e brindes àqueles que acessarem o link do questionário, lerem o Termo de Consentimento Livre e Esclarecido (TCLE) apresentado na primeira página do formulário e aceitarem responder ao questionário.**

## 6. CRONOGRAMA PROPOSTO

[illegible]

## REFERÊNCIAS

**BABBIE, E.** Métodos de Pesquisas de Survey, Belo Horizonte, Editora UFMG, 2001.

**BABU L, A.** Top 8 programming languages every data scientist should master in 2019, Big Data Made Simple, Jan 24, 2019. Disponível em: <<https://bigdata-madesimple.com/top-8-programming-languages-every-data-scientist-should-master-in-2019/>>. Acesso em: 03 de fevereiro de 2019.

**CHEN, M.; MAO, S.; LIU, Y.** Big Data: A Survey, Mobile Networks and Applications, 19 (2). 2014. pp: 171–209.

**EXAME,** Cinema vs. Netflix, o combate. Disponível em: <<https://exame.abril.com.br/estilo-de-vida/cinema-vs-netflix-o-combate/>>. Acesso em 01 de maio de 2019.

**FREITAS, H., OLIVEIRA, M., SACCOL, A. Z., & MOSCAROLA, J.** O método de pesquisa Survey. Revista de Administração da Universidade de São Paulo, 35(3), 2000.

**ISTOÉ,** Indústria cinematográfica registra lucro recorde em 2018. Disponível em: <<https://www.istoedinheiro.com.br/industria-cinematografica-registra-lucro-recorde-em-2018/>>. Acesso em 02 de abril de 2019.

**INSTITUTO ATLÂNTICO.** Data Science: entenda a importância dos dados para sua empresa. 2018. Disponível em: <<https://blog.atlantico.com.br/data-science-entenda-a-importancia-dos-dados-para-sua-empresa/>>. Acesso em: 17 de março de 2019.

**RUD, O. P.** Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy. Hoboken, N.J: Wiley & Sons, 2009.

**SANTOS, M. Y.; RAMOS, I.** Business Intelligence: Tecnologias da informação na gestão de conhecimento. FCA Editora de Informática, Ltda., 2006.

**SILVA, V. C. L.; TERRA, L. A. A.** Business intelligence como fator decisivo na competitividade empresarial: Uma análise a partir de multi casos. Revista Inteligência Competitiva, v. 5, n. 1, p. 1-13, 2015.

**SILVEIRA, D. P.** O que é Data Science? 2016. Disponível em: <<https://www.oficinadanet.com.br/post/16919-o-que-e-data-science>>. Acesso em: 17 de março de 2019.

**YIN, R. K.** Pesquisa Estudo de Caso - Desenho e Métodos, 2 ed., Porto Alegre, Bookman, 1994.