

**CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA  
FACULDADE DE TECNOLOGIA DE BOTUCATU  
CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE  
SISTEMAS**

**ALEXANDRE MAGNO DE LIMA FERREIRA**

**UTILIZAÇÃO DE FERRAMENTAS COMPUTACIONAIS PARA ANÁLISES DE  
CIÊNCIAS DE DADOS EM APLICATIVOS DA PLAY STORE**

Botucatu - SP

Junho - 2019

**CENTRO ESTADUAL DE EDUCAÇÃO TECNOLÓGICA PAULA SOUZA  
FACULDADE DE TECNOLOGIA DE BOTUCATU  
CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE  
SISTEMAS**

**ALEXANDRE MAGNO DE LIMA FERREIRA**

**UTILIZAÇÃO DE FERRAMENTAS COMPUTACIONAIS PARA ANÁLISES DE  
CIÊNCIAS DE DADOS EM APLICATIVOS DA PLAY STORE**

Orientador: Prof. Dr. Gustavo Kimura Montanha

Co-orientador: Prof. Ms. José Rafael Pilan

Artigo Científico apresentado à FATEC -  
Faculdade de Tecnologia de Botucatu, para  
obtenção do título de Tecnólogo no Curso  
Superior de Tecnologia em Análise e  
Desenvolvimento de Sistemas.

Botucatu - SP

Junho – 2019

**UTILIZAÇÃO DE FERRAMENTAS COMPUTACIONAIS PARA ANÁLISES DE  
CIÊNCIAS DE DADOS EM APLICATIVOS DA PLAY STORE**

**USE OF COMPUTATIONAL TOOLS FOR DATA SCIENCE ANALYSIS IN PLAY  
STORE APPLICATIONS**

Alexandre Magno de Lima Ferreira<sup>1</sup>, Gustavo Kimura Montanha<sup>2</sup>, José Rafael Pilan<sup>3</sup>

---

<sup>1</sup> Graduando do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas – Fatec Botucatu. Rua General Telles, 2877 – Centro, 18602-120, Botucatu – São Paulo, alexmlf@hotmail.com

<sup>2</sup> Professor Doutor – Fatec Botucatu, gustavo.montanha@fatec.sp.gov.br

<sup>3</sup> Professor Mestre – Fatec Botucatu, rafael.pilan@unesp.br

## RESUMO

Com a revolução tecnológica e o crescimento relevante do uso de dispositivos móveis pelo mundo, principalmente o Android, hoje existe uma grande concentração de diversos tipos de dados sendo criados e acessados. Diante desse crescimento, originou-se o fenômeno denominado *Big Data*. A partir das grandes quantidades de dados geradas atualmente, as aplicações encontradas na *Play Store* dos celulares Android é um mercado vasto com diversos tipos de aplicativos, é uma área que pode ser estudada e interpretada utilizando ferramentas baseadas na ciência de dados. Uma das opções para aplicação de ciência de dados sobre o conjunto de dados estudado utiliza ferramentas como *Jupyter Lab* para o ambiente de desenvolvimento e para realização de computação interativa e exploratória dos dados, a linguagem de programação Python junto com as suas bibliotecas *pandas* e *matplotlib* que são projetadas para a realização da análise e visualização de grandes quantidades de dados. A partir da utilização dessas ferramentas, foi possível analisar e gerar gráficos com poucas linhas de código, permitindo uma interpretação mais simples sobre como se comportam os aplicativos encontrados na *Play Store*.

**Palavras-chave:** Ciência de Dados. Python. *Play Store*

## ABSTRACT

With the technological revolution and the significant growth in mobile usage worldwide, especially Android, today there is a huge concentration of various types of data being created and accessed. Given this growth, the phenomenon called Big Data originated. From the vast amounts of data currently generated, applications found in the Play Store of Android phones is a vast market with many kinds of applications, it is an area that can be studied and interpreted using data science based tools. One of the options for applying data science on the dataset studied uses tools like *Jupyter Lab* for the development environment and for interactive and exploratory data computation, the Python programming language along with its libraries *pandas* and *matplotlib*, designed to perform the analysis and visualization of large amounts of data. Using these tools, it was possible to analyze and generate graphs with few lines of code, allowing a simpler interpretation of how the applications found in the Play Store behave.

**Key Words:** Data Science. Python. *Play Store*

## 1 INTRODUÇÃO

Nos dias atuais, é notável que a tecnologia de dispositivos móveis tenha crescido de forma exponencial em comparação às últimas décadas. Entre os sistemas operacionais existentes, o Android representa a maior fatia do mercado de celulares, propiciando que inúmeras pessoas utilizem aplicações disponíveis nesses dispositivos. Isso gera uma imensa quantidade de dados que é criada e acessada diariamente.

O grande crescimento de acesso aos dados gera a oportunidade para a análise e interpretação da informação, a fim de gerar conhecimento através da utilização da ciência de dados sobre o que foi analisado, usando como um apoio para a melhor tomada de decisão sobre as estatísticas extraídas.

A plataforma Android é um *software* que inclui um sistema operacional baseado no kernel Linux, possui uma interface do usuário adequada para dispositivos móveis, aplicativos que vêm junto ao sistema, bibliotecas de código, *frameworks*, multimídia e entre outros (ABLESON et al., 2012).

Por ser baseado em Linux, o sistema operacional Android é *open-source*, uma plataforma *open-source* possibilita o livre acesso do código fonte, podendo ser modificado e distribuído por qualquer usuário e desenvolvedor, e pelo fato de ser livre distribuição, não é necessário pagar para utilizá-lo (FAUSTINO; CALAZANS; LIMA, 2017).

O desenvolvimento da plataforma Android foi feita por uma pequena empresa chamada *Android Inc.* Essa empresa foi adquirida pela *Google* em 2005, quando esta observou o potencial da plataforma que *Android Inc.* estava desenvolvendo. E, em 2007, em coLaboração com a *Google*, foi formada a *Open Handset Alliance* (DEITEL; DEITEL; WALD, 2016). A *Open Handset Alliance* é uma colaboração entre diversas organizações com o objetivo de trazer ao mercado um telefone celular com o sistema operacional melhor e mais aberto ao público (ABLESON et al., 2012).

O surgimento dos dispositivos móveis com o sistema operacional Android teve seu início em 2008. A plataforma Android é bastante utilizada tanto por desenvolvedores como por usuários comuns, é bastante popular e está contida em pelo menos em uma fatia de 80% do mercado de todos os dispositivos móveis do mundo (IMHOF; ZANCHETT, 2018).

Os dispositivos móveis atualmente possuem as lojas virtuais que são repositórios que hospedam os aplicativos, onde ficam disponíveis para *download* ou para compra. O repositório mais famoso e mais utilizado no mundo é o repositório *Play Store*, da *Google*, que

possui mais de 1,43 milhões de aplicativos oficiais e é onde se encontra a grande maioria de usuários que utilizam Android (IMHOF; ZANCHETT, 2018).

Os aplicativos desenvolvidos para a plataforma Android em sua grande maioria são desenvolvidas na linguagem Java, Java foi escolhida para o desenvolvimento das aplicações, pois é uma linguagem de programação robusta, de código fonte aberto e gratuito, já conhecida por muitos desenvolvedores de software (DEITEL; DEITEL; WALD, 2016).

Atualmente, o Android não está apenas disponível para os dispositivos móveis. Ele também está presente em outros dispositivos como televisores conhecidos como *Smart TVs*, GPS e relógios conhecidos como *smartwatch* (FAUSTINO; CALAZANS; LIMA, 2017).

A ciência de dados, do inglês *data science*, é uma ciência nova que estuda o ciclo de vida dos dados por completo. Dados são fatos coletados sobre algo, esses dados são armazenados e, depois, são analisados para gerar informação e a informação é interpretada e aplicada a uma solução para gerar conhecimento (AMARAL, 2016).

O dado estudado na ciência de dados é o dado digital que é produzido por algum dispositivo eletrônico e armazenado em uma estrutura, como por exemplo, um banco de dados. Esses dados são transformados e formatados com o intuito de exibi-los de forma mais simples e organizada para o usuário. Quando completam o ciclo de produção, armazenamento e transformação, os dados podem ser analisados. A análise consiste em utilizar diversas operações que possuem a finalidade de extrair informações. Uma etapa importante do ciclo de vida dos dados é a sua visualização que tem como finalidade proporcionar a interpretação dos dados de forma mais clara e precisa, permitindo a compreensão e o esclarecimento sobre o dado visualizado. E, no final do ciclo de vida, os dados passam pelo processo de descarte (AMARAL, 2016).

A ciência de dados é utilizada com a finalidade de extrair e analisar de modo complementar a grandes quantidades de dados que podem ser dinâmicos e complexos. Esses tipos de dados se distinguem como o conceito de *Big Data*. *Big Data* é um fenômeno tecnológico caracterizado pelo conjunto de dados que são produzidos em grande volume, variedade e velocidade causada pela evolução tecnológica que se originou pelo crescimento dos serviços de internet, dos recursos computacionais e da variedade de tecnologias que geram dados a todo tempo (RAUTENBERG; CARMO, 2019).

O objetivo do presente trabalho é apresentar a utilização de ferramentas computacionais para a realização de análises e visualizações sobre o conjunto de dados de aplicações da *Play Store*, da *Google*.

## 2 MATERIAL E MÉTODOS

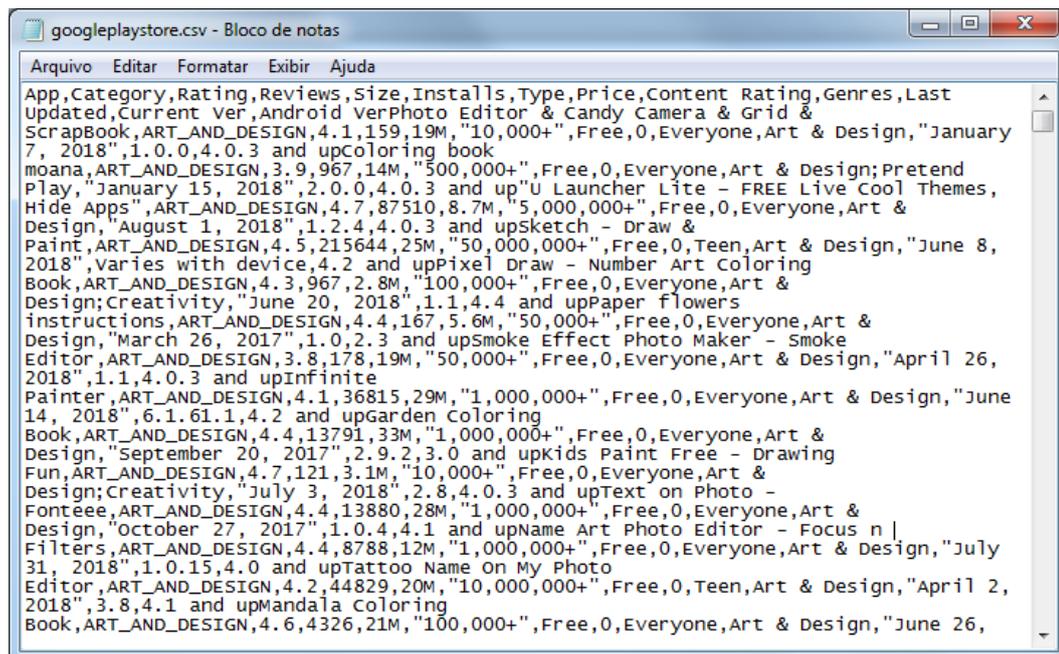
Para realizar a leitura e a extração de informações da amostra do conjunto de dados do *Play Store* foram aplicadas diversas ferramentas computacionais. Para este projeto, foi utilizado o ambiente de desenvolvimento *Jupyter* que permite que o arquivo contenha textos explicativos e trechos de códigos executáveis em linguagem Python, onde as entradas e saídas dos dados são executadas de modo exploratório, o *Jupyter* permite a integração com as bibliotecas do Python *pandas* e *matplotlib*, podendo mostrar como saída as tabelas dos *DataFrames* do *pandas* e os gráficos gerados pelo *matplotlib*.

Para ter acesso a uma amostra dos dados do mercado de aplicações do *Play Store* da *Google*, foi utilizado um conjunto de dados encontrado no *site* da *Kaggle*. A *Kaggle* é uma comunidade aberta para usuários que se interessam na manipulação e no estudo de dados, é uma comunidade onde se pode aprender na prática a ciência de dados e visualizar trabalhos feitos por usuários cadastrados (KAGGLE, 2019).

O conjunto de dados que foi utilizado nesse projeto é o conjunto de dados de aplicativos da *Play Store*, disponibilizado por um usuário da comunidade que extraiu os dados através da técnica de *web scraping*, uma técnica de mineração de dados que consiste na extração dos dados de páginas web de forma automatizada, a extração foi realizada no mês de Agosto de 2018 gerando uma amostra de 10 mil dados para a análise do mercado de aplicativos Android (KAGGLE, 2019).

Os dados sobre aplicativos da *Google Play Store* estão salvos em um arquivo do tipo csv. Csv, do inglês, significa *comma-separated values*, é um arquivo de texto que faz ordenação de bytes com a separação dos valores por vírgulas, pode-se ver o conteúdo do arquivo abrindo o arquivo pelo bloco de notas (Figura 1).

Figura 1 – Arquivo csv



**Fonte:** O autor, 2019.

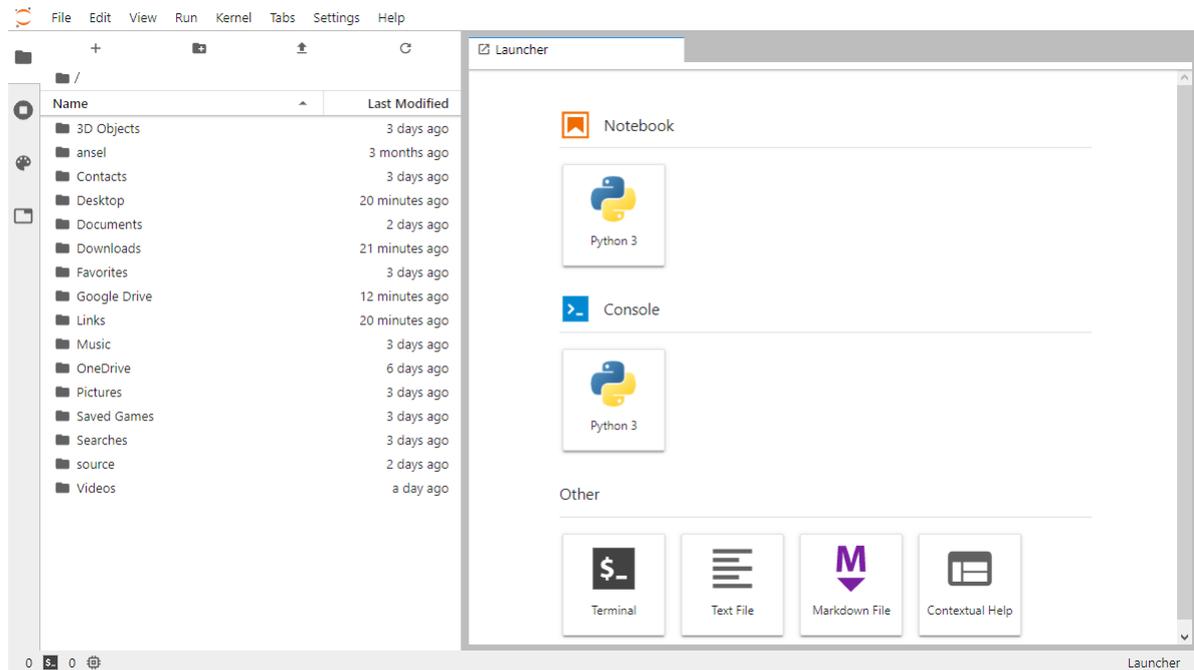
Para ler o arquivo csv de uma forma mais organizada foi necessário a utilização do ambiente de desenvolvimento *Jupyter* junto a linguagem Python e suas bibliotecas.

O *Jupyter Notebook* é um ambiente de desenvolvimento utilizado pelo navegador *web* que utiliza da computação interativa que permite que a execução de um projeto seja feita de modo exploratório onde os códigos executados já recebem resultado que guiam para o próximo passo e ajuda a verificar os passos anteriores do contexto do projeto (NGUYEN; CASE; ROSE, 2018).

A computação interativa se define como a execução de pequenos pedaços de código em várias linguagens de programação, como R ou Python, e imediatamente vendo os resultados do código executado. A computação interativa é interligada à ciência de dados, à medida que problemas científicos se beneficiam pelo processo exploratório, dando suporte para a formação de conhecimento sobre o problema em questão (PEREZ; GRANGER, 2015).

O *Jupyter Lab* é uma evolução do *Jupyter Notebook* que dá uma experiência próxima de um ambiente de desenvolvimento integrado de *software*. O *Jupyter Lab* já coloca todos instrumentos necessários juntos e oferece um ambiente de trabalho flexível e personalizável se concentrando na computação exploratória e interativa e mantém muitos recursos encontrados em IDEs tradicionais (ZHANG, 2018). A Figura 2 ilustra a interface do *Jupyter Lab* que foi utilizado para o projeto.

Figura 2 – Interface do *Jupyter Lab*



**Fonte:** O autor, 2019.

No *Jupyter Lab* foi possível criar *Notebooks* que são os arquivos para a computação interativa com trechos de código em Python, a partir do arquivo *Notebook* que este projeto foi criado.

Python é um exemplo de uma linguagem de alto nível e é uma linguagem interpretada, pois seus programas são executados por interpretador que lê programas de alto nível e os executa. Ele processa o programa em pequenas partes por um tempo, alternativamente lendo as linhas e realizando computações (DOWNEY, 2016).

Python é uma linguagem multiplataforma, os programas escritos em Python podem ser executados tanto no Windows como sistemas baseados em Unix, necessitando apenas do interpretador Python instalado na máquina (SUMMERFIELD, 2010).

O Python provém de uma biblioteca padrão que permite a realização de *download* de um arquivo de *internet*, descompactação de um arquivo comprimido ou criação de um servidor *web* com apenas algumas linhas de código. Também provém de diversas bibliotecas de terceiros importantes, como o módulo *pandas* para a manipulação de grandes quantidades de dados e o módulo *matplotlib* para visualização de dados (SUMMERFIELD, 2010).

A biblioteca *pandas* provém de diversas estruturas de dados e funções com a finalidade de trabalhar com dados estruturados ou tabulares de maneira mais simplificada. Os objetos principais do *pandas* mais utilizados são: *DataFrames* que é um tipo de dado de duas

dimensões orientado em colunas e linhas e o *Series* sendo um objeto *array* de uma dimensão. Esses objetos do *pandas* possuem funcionalidades de indexação para tornar simples a remodelação, fatiamento, agregações, seleções de subconjuntos, manipulação, preparação e limpeza dos dados (MCKINNEY, 2018).

A biblioteca *matplotlib* é destinada à visualização de dados. A biblioteca permite uma criação simplificada de gráficos com poucas linhas de código, esses gráficos incluem gráficos de linhas, de barras, de dispersão, diagrama de caixa, com grande flexibilidade e customização de estilos. A vantagem de gerar gráficos por código é pela simplicidade e facilidade de criação, da utilização de tipos de gráficos diversos e da praticidade em personalizar os gráficos criados da maneira que preferir (YIM; CHUNG; YU, 2018).

Com as ferramentas computacionais citadas, pode-se ler e organizar o arquivo csv obtido na *Kaggle* de uma forma mais intuitiva. Para isso, no arquivo *Notebook* criado, utilizando a função do *pandas read\_csv*, que utiliza como parâmetro o diretório do arquivo, e utilizando o método *head* podemos mostrar os 5 primeiros dados, apresentado na Figura 3.

Figura 3 – Leitura do arquivo csv

```
[2]: apps = pd.read_csv('googleplaystore.csv')
```

Para visualizar os primeiros 5 dados a fim de ter uma ideia sobre *dataset* que estamos trabalhando utilizamos o método `head` :

```
[3]: apps.head()
```

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

Fonte: O autor, 2019.

Nota-se que o arquivo csv foi armazenado na variável *apps*, e no método *head()* onde cada linha da tabela representa um dado que contém informações baseadas nas colunas.

No conjunto de dados analisado, os tipos de dados encontrados eram de sua maioria do tipo objeto-*string*, algumas dessas colunas de tipo objeto-*string* possuem valores numéricos,

realizando então uma análise de cada coluna e uma limpeza para mudança dos tipos de dados numéricos a fim de obter uma melhor visualização dos dados quando utilizado o *matplotlib*.

Dados obtidos por *web scrapping* estão propensos a vir com erros, para isso, foi utilizado o método *drop\_duplicates* que faz a remoção de dados, baseado no parâmetro da coluna *App*, os aplicativos que possuem nomes duplicados mostrado na Figura 4.

Figura 4 – Método *drop\_duplicates* para remoção de dados duplicados

```
[5]: apps.drop_duplicates(subset='App', keep='first', inplace=True)
```

**Fonte:** O autor, 2019.

Para-se entender as colunas encontradas no conjunto de dados, a Tabela 1 apresenta o nome da colunas, a descrição de que cada coluna representa e seu tipo de dado.

Tabela 1 – Tabela descritiva sobre as colunas do arquivo

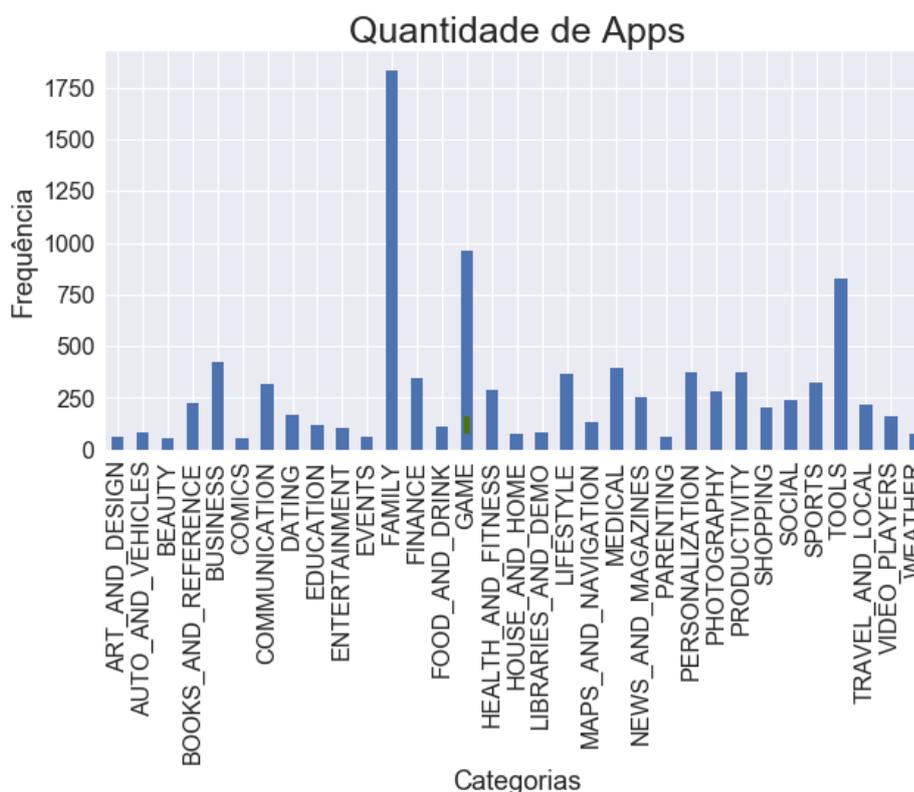
Nome da coluna	Descrição	Tipo de dado
<i>App</i>	Nome do aplicativo	<i>String</i>
<i>Category</i>	Categoria do aplicativo	<i>String</i>
<i>Reviews</i>	Quantidade de avaliações	<i>Int</i>
<i>Size</i>	Tamanho do aplicativo (em MB's)	<i>Float</i>
<i>Installs</i>	Instalações feitas no aplicativo	<i>Int</i>
<i>Type</i>	Se o aplicativo é pago ou grátis ( <i>Free/Paid</i> )	<i>String</i>
<i>Price</i>	O preço do aplicativo (em dólares)	<i>Float</i>
<i>Content_Rating</i>	Classificação indicativa subdividida em: <i>Everyone, Everyone 10+, Teen, Mature 17+, Adults Only 18+ e Unrated</i>	<i>String</i>
<i>Genres</i>	O gênero do aplicativo	<i>String</i>
<i>Last_Updated</i>	Última atualização do aplicativo	<i>DateTime</i>
<i>Current_Ver</i>	Versão atual do aplicativo	<i>String</i>
<i>Android_Ver</i>	Versão mínima do Android necessária para executar o aplicativo	<i>String</i>

**Fonte:** O autor, 2019.

### 3 RESULTADOS E DISCUSSÃO

Realizado a limpeza e organização dos dados com a utilização da biblioteca *pandas* o arquivo está pronto para a utilização da biblioteca *matplotlib* para a realização de visualizações dos dados:

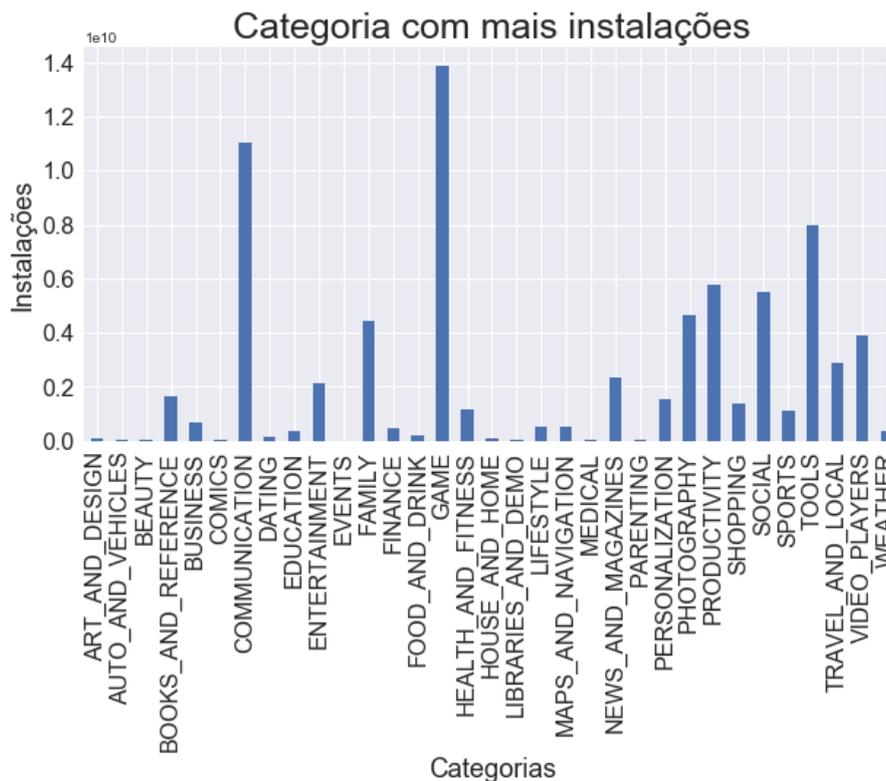
Figura 5 – Gráfico de frequência de Categorias de Apps



**Fonte:** O autor, 2019.

A Figura 5 se refere aos dados em relação a coluna *Category* e com a utilização do *pandas* e o *matplotlib* para extrair a quantidade de dados que existem em cada categoria e visualizando-as. Observando-se o gráfico sobre a quantidade de aplicativos disponíveis na amostra dos dados da *Google Play*, é constatado que a categoria *Family* possui maior frequência com aproximadamente 1800 aplicativos, seguido da categoria *Game* e *Tools*, com aproximadamente 950 e 800 aplicativos respectivamente.

Figura 6 – Gráfico de quantidade de instalações por categorias



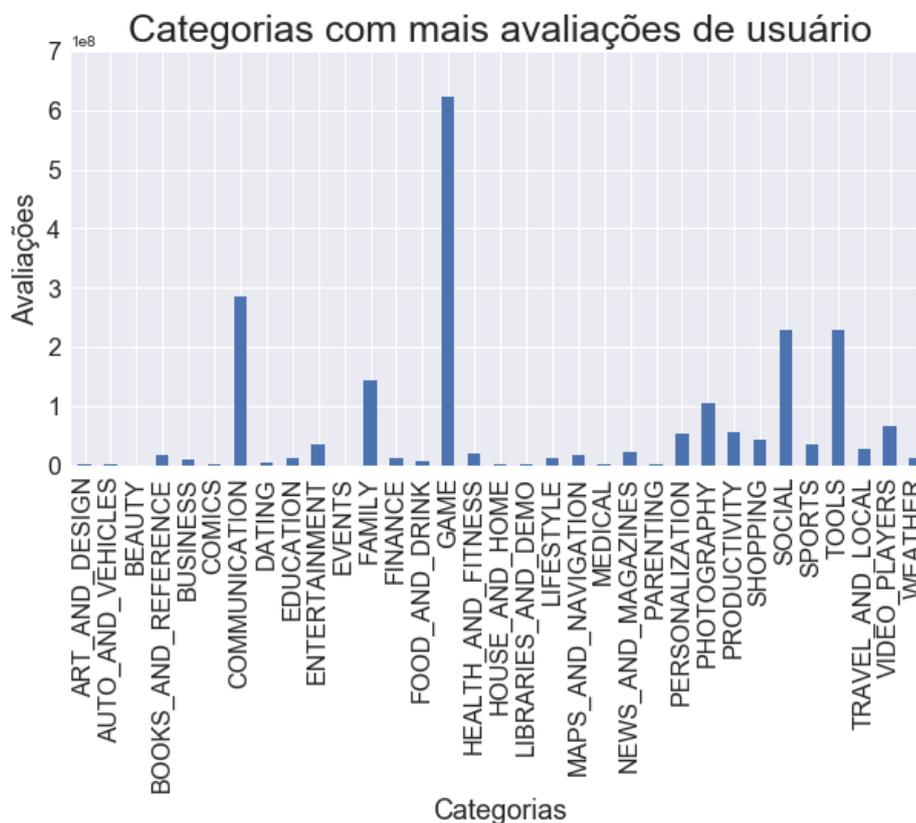
Fonte: O autor, 2019.

Na Figura 6, é mostrado o gráfico de categorias em relação à quantidade de instalações feitas por usuários, utilizando as colunas *Category* e *Installs* do arquivo. Os valores no eixo y representam um valor que é multiplicado por  $10^{10}$ , então os valores estão variando entre 0 e 14 bilhões de instalações.

Este gráfico é importante para verificar qual categoria é mais baixada por usuários, dando apoio a desenvolvedores que queiram criar um aplicativo em uma categoria de aplicativos com mais mercado. Nota-se que as categorias *Game*, *Communication* e *Tools* estão na frente das demais. Na figura anterior, é mostrado que a quantidade de aplicativos encontrados em *Communication* é menor, entretanto o gráfico da Figura 5 demonstra que mesmo com a quantidade menor, os aplicativos existentes são utilizados por uma quantidade maior de usuários, pelo fato de ter um alto valor de instalações.

Outra coluna importante visualizar é a coluna *Reviews* que indica a quantidade de avaliações feitas por usuários que utilizaram o aplicativo, quando um usuário dá o seu feedback sobre o aplicativo significa que é um valor mais significativo para análise pois indica que o usuário realmente instalou e utilizou o aplicativo.

Figura 7 – Gráfico de categorias com mais avaliações de usuário



**Fonte:** O autor, 2019.

Na Figura 7, o gráfico em seu eixo y representa a coluna *Reviews* que é a quantidade de avaliações, variando entre 0 e 700 milhões e o eixo x sendo a coluna *Category* de categorias. Nota-se que a categoria *Game* está bem na frente dos demais, seguido da categoria *Communication*, *Tools* e *Social*. Este gráfico mostra que os aplicativos baseados em jogos possuem uma alta taxa de feedback de usuários que a utilizam. Isso representa que o mercado de jogos *mobile* é interessante para desenvolvedores que queiram publicar jogos.

Uma outra coluna interessante de se observar é a coluna *Rating*, para verificar como os usuários avaliam os aplicativos que utilizam, feita na Figura 8:

Figura 8 – Gráfico da frequência de notas das aplicações

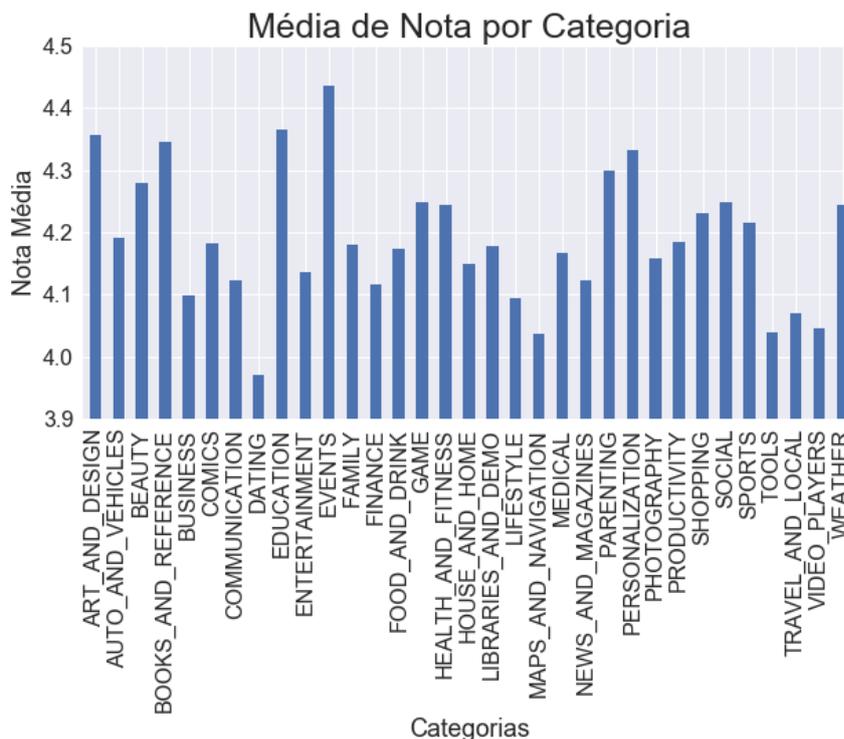


Fonte: O autor, 2019.

Realizando a contagem de frequência sobre quantos aplicativos estão entre as notas 1 e 5, observa-se, na Figura 8, que a grande maioria dos aplicativos nessa amostra possuem notas entre 4 e 4.5. Ou seja, as aplicações da amostra do conjunto de dados da *Google Play* possuem uma média boa de notas em sua grande maioria.

Pode-se também verificar as notas de acordo com a sua categoria, observado na Figura 9 a seguir:

Figura 9 – Gráfico da média de nota por cada categoria



Fonte: O autor, 2019.

Observa-se que na Figura 9 o eixo y representado pela média de notas, os valores foram delimitados entre 3.9 e 4.5, pois é onde ocorre mais variação. A categoria com maior média representa a categoria *Events* e a categoria *Dating* com menor a média entre os demais. O gráfico mostra que o tipo de categoria da aplicação não se varia muito em relação à média de nota que os usuários tendem a dar a aplicação utilizada, ou seja, a categoria de aplicação não importa muito em relação a que nota os usuários podem dar a ela.

## 4 CONCLUSÕES

Em relação aos resultados obtidos pelas análises e visualizações realizadas na amostra dos aplicativos da *Play Store*, pode-se concluir que, através da utilização das ferramentas do *Jupyter Lab* e as bibliotecas *pandas* e *matplotlib* do Python, foi possível interpretar os dados pela a análise e geração de gráficos, facilitando, assim, a interpretação de um grande volume de dados sem precisar recorrer a tabelas e valores dos dados brutos obtidos.

## REFERÊNCIAS

- ABLESON, W. F.; SEN, R.; KING, C.; ORTIZ, C. E. **Android em ação**. 3. ed. Rio de Janeiro: Elsevier, 2012. 656 p.
- AMARAL, F. **Introdução à Ciência de Dados: mineração de dados e big data**. Rio de Janeiro: Alta Books, 2016. 320 p.
- DEITEL, P.; DEITEL, H.; WALD, A. **Android 6 para programadores: uma abordagem baseada em aplicativos**. 3. ed. Porto alegre: Bookman, 2016. 618 p.
- DOWNEY, A. B. **Pense em Python: Pense como um cientista da computação**. 2. ed. São Paulo: Novatec, 2016. 312 p.
- FAUSTINO, G. K. S.; CALAZANS, H. K. N. S.; LIMA, W. D. Android e a influência do Sistema Operacional Linux. **Tecnologias em Projeção**, Brasília, v. 8, n. 1, p. 100-111, 2017. Disponível em: <<http://revista.faculdadeprojecao.edu.br/index.php/Projecao4/article/view/829>>. Acesso: 07 mai. 2019.
- IMHOF, D. O.; ZANCHETT, P. S. Monetização com desenvolvimento de aplicativos em Android. **Maiêutica Tecnologia da Informação**, Indaial, v. 2, n.1, p. 51-56, 2018. Disponível em: <[https://publicacao.uniasselvi.com.br/index.php/TI\\_EaD/article/view/1882](https://publicacao.uniasselvi.com.br/index.php/TI_EaD/article/view/1882)>. Acesso: 07 mai. 2019.
- KAGGLE. **Kaggle: Your Home for Data Science**, 2019. Disponível em: <<https://www.kaggle.com>>. Acesso: 29 mar. 2019.
- MCKINNEY, W. **Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython**. 2. ed. Sebastopol: O'Reilly Media, 2018. 550 p.
- NGUYEN, H.; CASE, D. A.; ROSE, A. S. NGLview-interactive molecular graphics for *Jupyter* notebooks. **Bioinformatics**, V. 34, n. 7, p. 1241-1242, 01 abr. 2018. Disponível em: <<https://doi.org/10.1093/bioinformatics/btx789>>. Acesso: 13 abr. 2019.

RAUTENBERG, S.; CARMO, P. R. Big Data e Ciência de Dados. **Brazilian Journal of Information Science: research trends**, v. 13, n. 1, p. 56-67, 29 mar. 2019. Disponível em: <<http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/8315>>. Acesso: 12 mai. 2019.

SUMMERFIELD, M. **Programming in Python 3: A complete Introduction to the Python Language**. 2. ed. Boston: Addison-Wesley, 2010. 630 p.

PEREZ, F; GRANGER, B. E. **Project Jupyter: Computational narratives as the engine of collaborative data science**, 2015. Disponível em: <<http://archive.ipython.org/JupyterGrantNarrative-2015.pdf>>. Acesso: 13 abr. 2019.

YIM, A.; CHUNG, C.; YU, A. **Matplotlib for Python Developers: Effective techniques for data visualization with Python**. 2. ed. Birmingham: Packt, 2018. 300 p.

ZHANG, J. **JupyterLab\_Voyager: A Data Visualization Enhancement In JupyterLab**.58f. Dissertação (Mestrado em Ciência da Computação) apresentada a Faculdade da Universidade Estadual Politécnica da Califórnia de San Luis Obispo, 2018. Disponível em: <<https://digitalcommons.calpoly.edu/theses/1849/>>. Acesso em: 25 maio 2019.