http://journal.unoeste.br/index.php/ce/index
DOI: 10.5747/ce.2020.v12.n4.e338
ISSN on-line 2178-8332

Colloquium Exactarum

Submetido: 12/08/2020 Revisado: 13/09/2020 Aceito: 16/09/2020

ESTUDO COMPARATIVO DA APLICAÇÃO DOS PROGRAMAS PYTHON E ORANGE PARA A ANÁLISE APROFUNDADA DE BANCOS DE DADOS

Beatriz Martins Pereira, Camila Solange Moreno Maldonado Godoi, Thayna Barros Viana, Rafael Medeiros Hespanhol

Universidade do Oeste Paulista – UNOESTE, Presidente Prudente, SP. Niagara College, Canada.
E-mail: bia_martins260597@hotmail.com

RESUMO - O presente estudo possui o objetivo de analisar a usabilidade de programas de *DataScience* para análise de banco de dados. Para tanto foi utilizado um banco de dados extraído de uma plataforma pública e aplicado em dois programas, sendo Python e Orange, a fim de se obter uma comparação entre ambos. A pesquisa bibliográfica serviu como base para o entendimento dos programas e para embasar os resultados obtidos. O Python exigiu trabalhar o banco de dados, tanto na conversão do arquivo como necessidade de conhecimento e aprendizado de linguagem de programação. No Orange foi utilizado o banco de dados original e sua funcionalidade intuitiva permitiu a obtenção de resultados de forma mais rápida, pois suas ferramentas possuem nomes associados ao que se deseja obter. Com os resultados obtidos foi possível constatar que a utilização do programa Orange se demonstrou mais conveniente para análise do banco de dados manipulado.
Palavras-chave: Ciência de dados. Sistemas de Informação. Funcionalidade.

COMPARATIVE STUDY ON THE APPLICATION OF PYTHON AND ORANGE PROGRAMS FOR DETAILED DATABASE ANALYSIS

ABSTRACT - The present study aims to analyze the usability of DataScience programs for database analysis. Therefore, a database extracted from a public platform was used and applied in two programs, Python and Orange, in order to obtain a comparison between both. The bibliographic research served as a basis for the understanding of the programs and to base the obtained results. The Python program required work with the database, in file conversion as well as the need for programming language knowledge and learning. In the Orange program, the original database was used and its intuitive functionality allowed to obtain faster results, since their tools have names associated with what you want to get. With the obtained results it was possible to verify that the use of the Orange program proved to be more convenient for analysis of the manipulated database.
Keywords: DataScience. Information Systems. Functionality.

1. INTRODUÇÃO

O conceito de *DataScience*, aplicado por meio de ferramentas, programas e softwares existentes que auxiliam a análise aprofundada de dados e proporcionam informações robustas para a tomada de decisão, apresentam um forte crescimento nos últimos anos.

Basicamente são utilizados para transformar banco de dados de quaisquer tamanhos, abundantemente disponíveis, em informações valiosas para indivíduos e organizações. Essas informações podem apresentar oportunidades ou ameaças, as quais as organizações se baseiam para tomar decisões, visando à antecipação no mercado.

O presente trabalho tem como objetivo geral aplicar e analisar os programas Python e Orange para uma análise aprofundada de um banco de dados público, que fornece a quantidade de produção de petróleo no Brasil, entre os anos de 1997 e 2018, separando esta produção mensalmente, por região, estado e localização de exploração (terra e mar). Este estudo permitiu comparar a usabilidade dessas ferramentas, levando em consideração o baixo volume de dados e a baixa quantidade de variáveis do banco de dados analisado, possibilitando aos pesquisadores, identificar resultados gráficos obtidos automaticamente pelos programas, sem necessitar trabalhar o banco de dados, como seria necessário na utilização da ferramenta Microsoft Excel.

2. MÉTODO, FERRAMENTAS E PROCEDIMENTOS

Inicialmente foi extraído da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), o banco de dados referente à produção de petróleo no Brasil de 1997 a 2018, disponibilizado na guia dados abertos do site da agência (Figura 1). Este banco fornece dados da quantidade total de petróleo produzido no decorrer desses 22 anos apurados mensalmente e separado pelas variáveis: localização de exploração, ano, região e estado. Apresenta um total de 484 linhas completamente preenchidas. Tendo em vista a precisão de preenchimento do banco de dados não foi preciso ajustá-lo para aplicação nos programas.

Figura 1. Produção de Petróleo (m³) – 1997 a 2018

LOCALIZAÇÃO	ANO	REGIÃO	ESTADO	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	TOTAL
TERRA	1997	NORTE	AMAZONAS	122.200	124.785	128.177	124.968	115.710	132.316	137.417	131.451	81.753	44.839	63.384	136.982	1.343.982
TERRA	1997	NORDESTE	MARANHÃO	0	0	0	0	0	0	0	0	0	0	0	0	0
TERRA	1997	NORDESTE	CEARÁ	14.568	15.656	15.441	14.366	15.637	15.557	15.075	15.636	15.275	15.409	14.676	14.910	182.206
TERRA	1997	NORDESTE	RIO GRANDE DO NORTE	395.223	388.457	397.430	391.698	377.218	393.579	405.379	405.148	393.668	413.071	397.992	411.871	4.770.734
TERRA	1997	NORDESTE	ALAGOAS	15.457	16.675	16.639	16.038	18.473	18.648	20.853	22.910	21.891	23.448	20.620	19.888	231.540
TERRA	1997	NORDESTE	SERGIPE	123.453	124.334	127.248	121.150	123.758	122.921	126.212	126.845	118.348	125.735	124.150	128.509	1.492.663
TERRA	1997	NORDESTE	BAHIA	247.780	241.234	249.527	238.530	241.321	237.702	245.029	249.531	240.272	245.094	237.101	244.974	2.918.095
TERRA	1997	SUDESTE	ESPÍRITO SANTO	36.567	37.889	40.964	38.294	35.840	37.239	38.021	37.402	36.199	37.606	36.872	37.537	450.430
TERRA	1997	SUDESTE	RIO DE JANEIRO	0	0	0	0	0	0	0	0	0	0	0	0	0
TERRA	1997	SUDESTE	SÃO PAULO	0	0	0	0	0	0	0	0	0	0	0	0	0
TERRA	1997	SUL	PARANÁ	0	0	0	0	0	0	0	0	0	0	0	0	0
TERRA	1998	NORTE	AMAZONAS	146.841	133.232	156.917	151.828	160.283	155.258	166.677	167.220	166.017	165.171	155.562	166.018	1.891.024
TERRA	1998	NORDESTE	MARANHÃO	0	0	0	0	0	0	0	0	0	0	0	0	0
TERRA	1998	NORDESTE	CEARÁ	15.075	14.027	14.061	14.469	15.993	14.885	15.114	16.269	14.968	16.140	16.489	18.600	186.090
TERRA	1998	NORDESTE	RIO GRANDE DO NORTE	406.445	368.034	416.697	409.461	425.178	414.995	430.857	425.503	416.476	437.936	426.688	433.150	5.011.420
TERRA	1998	NORDESTE	ALAGOAS	19.530	18.105	20.841	21.247	21.865	21.140	20.939	20.506	19.172	20.939	20.706	21.527	246.517
TERRA	1998	NORDESTE	SERGIPE	126.243	111.723	122.812	117.645	120.795	117.450	119.446	118.266	116.074	123.119	116.920	121.487	1.431.980
TERRA	1998	NORDESTE	BAHIA	246.848	225.280	248.195	237.872	243.388	233.430	243.568	240.368	232.990	240.941	233.304	240.813	2.866.997
TERRA	1998	SUDESTE	ESPÍRITO SANTO	41.154	37.216	39.549	37.921	37.727	35.119	41.442	40.969	45.450	48.421	51.972	59.000	515.940
TERRA	1998	SUDESTE	RIO DE JANEIRO	0	0	0	0	0	0	0	0	0	0	0	0	0
TERRA	1998	SUDESTE	SÃO PAULO	0	0	0	0	0	0	0	0	0	0	0	0	0
TERRA	1998	SUL	PARANÁ	0	0	0	0	0	0	0	0	0	0	0	0	0
TERRA	1999	NORTE	AMAZONAS	160.930	149.297	166.271	158.391	168.900	163.790	169.578	164.717	161.618	172.158	165.073	174.320	1.975.043
TERRA	1999	NORDESTE	MARANHÃO	0	0	0	0	0	0	0	0	0	0	0	0	0
TERRA	1999	NORDESTE	CEARÁ	17.357	15.113	15.491	14.798	14.244	14.700	14.330	14.785	14.377	13.727	10.848	12.428	172.198
TERRA	1999	NORDESTE	RIO GRANDE DO NORTE	414.168	381.894	419.331	399.049	393.545	372.818	389.477	398.235	399.594	404.498	401.711	428.453	4.802.773
TERRA	1999	NORDESTE	ALAGOAS	21.744	21.318	22.789	25.737	24.523	23.392	23.530	22.752	21.906	22.434	23.561	23.856	277.542
TERRA	1999	NORDESTE	SERGIPE	116.350	104.032	115.015	115.132	117.064	114.395	117.959	118.631	115.095	119.371	116.490	119.977	1.389.511
TERRA	1999	NORDESTE	BAHIA	243.798	216.944	239.571	227.995	231.556	222.492	226.149	224.040	221.369	226.491	219.935	228.542	2.728.882
TERRA	1999	SUDESTE	ESPÍRITO SANTO	57.443	50.595	52.317	47.724	56.585	47.659	49.204	51.296	48.963	49.475	48.735	50.007	610.003
TERRA	1999	SUDESTE	RIO DE JANEIRO	0	0	0	0	0	0	0	0	0	0	0	0	0

Fonte: ANP (http://www.anp.gov.br/dados-abertos-anp)

Para desenvolver a análise do banco de dados inicialmente foi utilizado o programa Python que, de acordo com o Grupo PET-Tele (2012), consiste em uma linguagem de

programação interpretada, de código-fonte aberto e aplicável a diversos sistemas operacionais. Uma linguagem interpretada refere-se a qual não exige tradução para uma linguagem da máquina, mas sim ser interpretada por um programa, o qual traduzirá para a máquina. Portanto, o interpretador para Python, permite ser executado sem o fornecimento de um script (programa), pois o Python possui uma interface interativa, permitindo aplicar comandos e obter resultados.

Deste modo, para analisar o banco de dados no Python, fez-se necessário converter o arquivo de formato XLS para o formato CSV, por meio do site Convertio, o qual, segundo Mannara (2017), trata-se de uma ferramenta online e gratuita, que permite fazer conversões de arquivos em distintos e variados formatos. Para deixar o arquivo em CSV mais visual dentro do programa foi utilizada a Biblioteca Pandas, segundo Siegel (2018) consiste em uma ferramenta utilizada para a manipulação e análise de dados estruturados. Apresenta facilidade de uso e capacidade para trabalhar com grande volume de dados, além disso, proporciona maior clareza na visualização de dados e modifica esteticamente os gráficos, facilitando a exibição tanto do banco de dados quanto dos gráficos.

Para obter resultados com o Python foi necessário aprender e compreender a linguagem de programação e aplicá-la em forma de comandos. Tendo em vista a simplicidade do banco de dados utilizado, foi possível desenvolver resultados através de uma lógica de programação básica, não exigindo aprofundamento dos pesquisadores.

Aplicou-se o Orange que, conforme Campos e Cazella (2018), consiste em um programa que proporciona a análise de dados de uma forma descontraída e produtiva, por meio de seu software de código aberto. Seu dinamismo permite seu uso tanto por usuários iniciantes, quanto especialistas. O Orange realiza a análise de dados através de fluxos de processos, que podem ser desenvolvidos interligando objetos, os quais representam uma grande variedade de ferramentas, incluindo técnicas que permitem visualizar, explorar, realizar o pré-processamento e modelagem de dados. O programa Orange, pode ser utilizado por meio de uma interface intuitiva ou, para usuários com uma técnica avançada, através de um módulo para a linguagem de programação Python.

O Orange permitiu a utilização do banco de dados no formato XLS, adicionando o arquivo no programa direto de sua pasta. Este programa apresenta uma interface intuitiva e de fácil aplicação, a partir da inserção do banco de dados, foram realizadas ligações de ferramentas disponíveis no programa, sendo estas de visualização (lineplot) e não supervisionadas (t-SNE) diretamente ao arquivo que recebeu o banco de dados, e os resultados foram gerados automaticamente. Em alguns minutos de testes já foi possível identificar resultados iguais aos obtidos no Python.

A ferramenta lineplot realiza a plotagem de um gráfico de linhas a partir dos dados e permite a seleção da variável de análise através de uma caixa de seleção, no caso do banco de dados trabalhado, permitiu selecionar localização, região, estado e ano proporcionando maio rapidez na análise e comparação dos resultados.

O t-SNE consiste em uma ferramenta de clusterização, a qual busca agrupar dados que apresentam semelhanças. Em relação ao banco de dados trabalhado não se mostrou interessante, pois apenas separou por região e localização de exploração.

Por fim, por meio da análise dos resultados extraídos com a aplicação dos programas, foi possível obter informações relevantes e relacioná-las comas informações históricas da indústria petrolífera do Brasil, com o objetivo de perpetuar conhecimento sobre o assunto. Para tanto foi realizada uma revisão bibliográfica a fim de agregar um maior embasamento teórico a respeito do conteúdo.

3. RESULTADOS

A escolha inicial do programa Python como linguagem de programação utilizada neste projeto decorreu do fato da mesma se encontrar na categoria de linguagem de propósito geral. Sendo assim é classificada como uma linguagem que “pode ser utilizada tanto para desenvolvimento de programas comerciais quanto de programas científicos” (MANZANO,2011, p. 19,). Sua aplicação é extremamente variada e vai desde um simples desenvolvimento Web às implementações com *DataScience*, devido à sua “sintaxe clara e concisa que favorece a legibilidade do código-fonte, tornando a linguagem mais produtiva” (BORGES, p. 14, 2014)

Outra característica importante do Python e talvez a principal que a torne tão interessante é a sua grande coleção de recursos desenvolvidos para fins específicos de uma dada ciência, neste caso recursos voltados para Ciência de Dados. Python também atende ao conceito de portabilidade, funcionando em arquiteturas variadas. (MENEZES, 2014).

Diferentemente do Python, a aplicação do Orange não exigiu a conversão do formato do arquivo do banco de dados utilizado, pois ele permite a exploração de dados em arquivos que estão, tanto no formato CSV, quanto no formato XLS, fator que pôde agregar mais agilidade e praticidade aos usuários. Outro fator relevante, que implicou em uma análise mais intuitiva e produtiva do banco de dados, consiste no formato da programação, sendo a do Orange visual, permitindo carregar bases de dados e trabalhá-las com apenas alguns cliques, ao contrário do Python, que exige aplicar linguagem de programação, exigindo dos usuários, um conhecimento mais aprofundado sobre a aplicação de códigos e comandos.

As Figuras 2, 3 e 4 evidenciam que para cada resultado a se obter com o programa Python foi necessário o desenvolvimento de uma lógica de programação e comandos específicos, demonstrando a sua complexidade funcional. Além do tempo necessário até o entendimento de como deveria proceder.

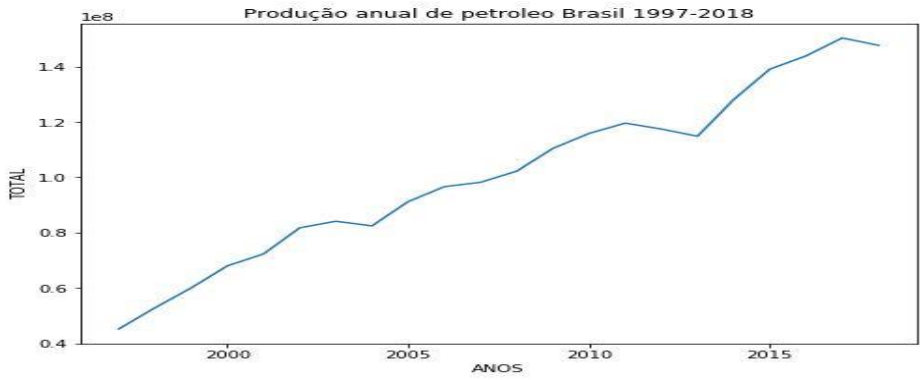
Figura 2. Comparação da produção por ano

```
In [125]: producao_anual = df.groupby("ANO").sum().TOTAL

          anos = df['ANO'].unique()

          plt.figure(figsize=(8, 6))
          plt.plot(anos, producao_anual)
          plt.xlabel('ANOS')
          plt.ylabel('TOTAL')
          plt.title('Produção anual de petroleo Brasil 1997-2018')

Out[125]: Text(0.5, 1.0, 'Produção anual de petroleo Brasil 1997-2018')
```



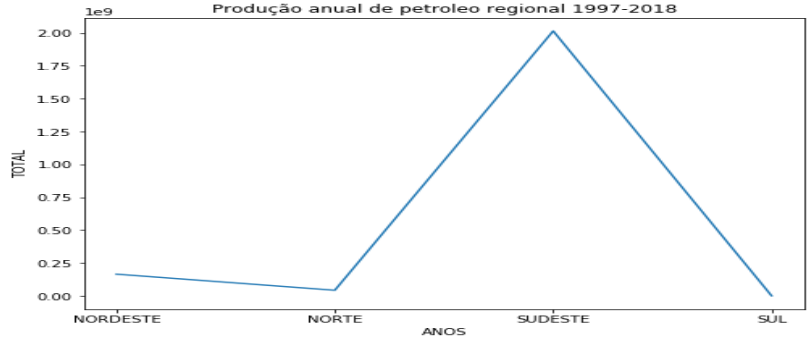
Fonte: próprios autores

Figura 3. Comparação da produção por região

```
In [16]: producao_estadual = df.groupby("REGIÃO").sum().TOTAL
         producao_estadual
         regioes = ['NORDESTE', 'NORTE', 'SUDESTE', 'SUL']

         plt.figure(figsize=(8, 6))
         plt.plot(regioes, producao_estadual)
         plt.xlabel('ANOS')
         plt.ylabel('TOTAL')
         plt.title('Produção de petroleo regional 1997-2018')

Out[16]: Text(0.5, 1.0, 'Produção anual de petroleo regional 1997-2018')
```



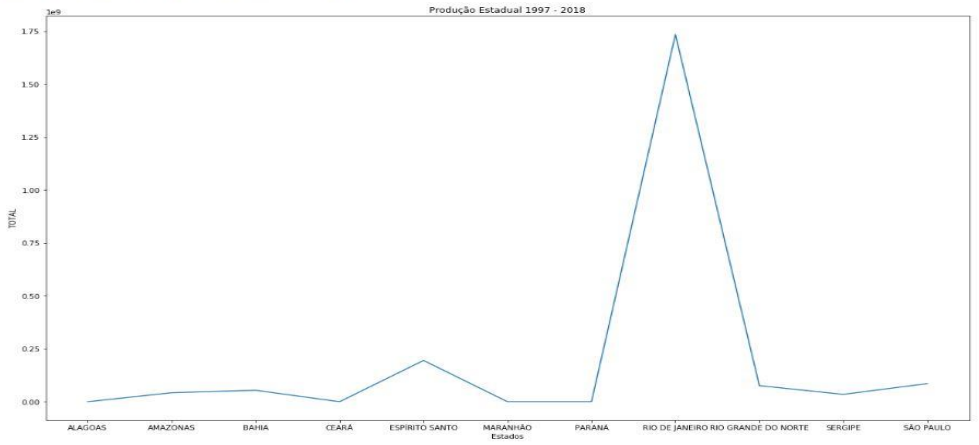
Fonte: próprios autores

Figura 4. Comparação da produção por estado

```
In [158]: producao_estadual = df.groupby("ESTADO").sum().TOTAL
          producao_estadual
          estados = ['ALAGOAS', 'AMAZONAS', 'BAHIA', 'CEARÁ', 'ESPÍRITO SANTO', 'MARANHÃO', 'PARANÁ', 'RIO DE JANEIRO', 'RIO GRANDE DO NOR
          TE', 'SERGIPE', 'SÃO PAULO']

          plt.figure(figsize=(20, 12))
          plt.plot(estados, producao_estadual)
          plt.xlabel('Estados')
          plt.ylabel('TOTAL')
          plt.title('Produção Estadual 1997 - 2018')

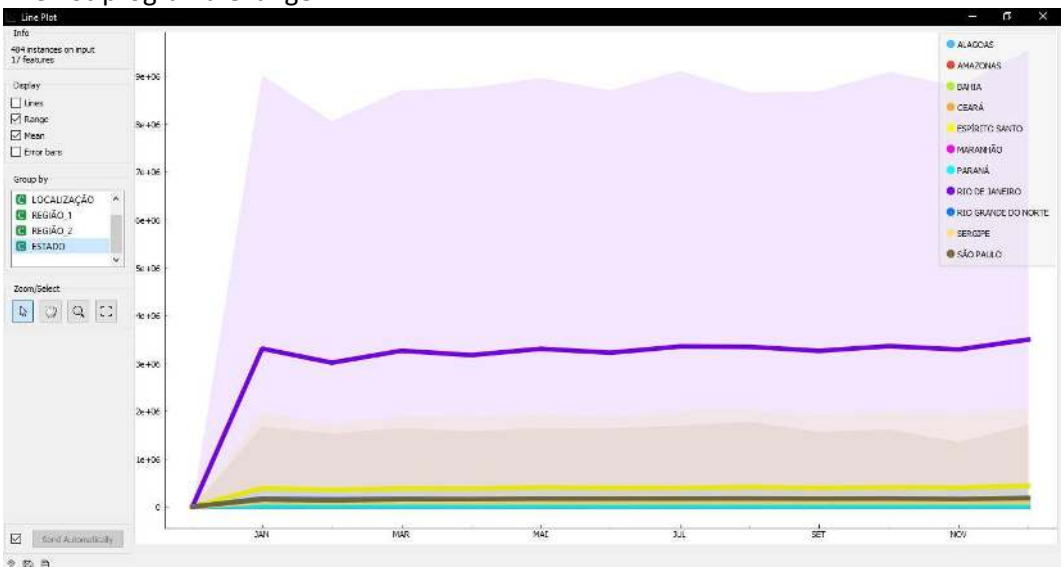
Out[158]: Text(0.5, 1.0, 'Produção Estadual 1997 - 2018')
```



Fonte: próprios autores

Tendo em vista a mesma necessidade de resultados através do Orange, este possibilitou maior dinamismo na obtenção dos resultados gráficos, permitindo que em uma única geração da ferramenta lineplot fosse possível obter as representações gráficas de forma automatizada e com maior rapidez, apenas selecionando a variável de análise (ano, região, estado, localização) na caixa de seleção localizado à esquerda conforme é possível observar na Figura 5, não exigindo qualquer nível de conhecimento em linguagem de programação.

Figura 5. *LinePlot* programa Orange



Fonte: próprios autores

4. DISCUSSÕES

De acordo com Silva (2016) o programa Python apresenta vulnerabilidade na utilização do módulo *threading*, o qual consiste em uma forma de fazer com que a sua aplicação execute tarefas de forma assíncrona, ou seja, enquanto uma estrutura de repetição é executada outra rotina também pode ser executada.

Garcia (2015) comenta que uma das desvantagens apresentadas pelo Python é a exigência de recorrer a IDEs (Integrated Drive Electronics) pesados para o funcionamento de

alguns módulos devido à origem de sua linguagem de programação.

O intuito da criação do Python era ser uma ferramenta para programadores profissionais de sistema, porém a implementação do Python oferece pouco em termos de ambiente de programação, mas foi projetada para integrar-se perfeitamente aos ambientes de programação existentes. (ROSSUM; BOER, 1991)

O programa Orange, além da facilidade em sua utilização, tendo em vista que as ferramentas possuem nomes associados ao resultado que se deseja obter, e sua funcionalidade consiste na ligação dessas ferramentas, apresenta um design mais visual e inovador.

De acordo com Demsar e Zupan (2012) o programa Orange foi desenvolvido principalmente para exploração de dados em que diferentes combinações dos algoritmos de pré-processamento e aprendizado de máquina (*machine learning*) fornecidos foram testados e pontuados usando validação cruzada. Os componentes foram compactados em programas com interfaces de linha de comando, porém de certa forma mais limitado, não permitindo muitas alterações.

Boscarioli, Viterbo e Teixeira (2014) afirmam que o Orange é uma ferramenta com enfoque na classificação de dados e mineração visual de dados, e apresenta um fluxo de execução simples, onde cada nodo adicionado ao campo de fluxo irá executar uma determinada tarefa de forma automática. Além de dispor de uma interface mais fácil de encontrar as informações e elementos desejados, também apresenta um sistema de *feedback*, retornando ao usuário os dados de entrada e de saída de cada método.

Uma falha visualizada por Boscarioli, Viterbo e Teixeira (2014) no Orange é a permissão do usuário interagir mesmo quando determinada ação é desenvolvida de maneira errada, avisando-o do erro somente no momento de tentar executar o fluxo do processo de mineração de dados, fazendo com que o usuário perca tempo na sua interação, pois na maioria das vezes deve refazer o fluxo por não saber exatamente onde ocorreu o erro.

Um fator de extrema importância a ser considerado é o tempo de execução do código. Segundo Seabra, Drummond e Gomes (2018) consiste na velocidade com que o programa será executado, além de quanto tempo ele levará para apresentar a resposta ou solução para o problema apontado. Na análise realizada constatou-se que dependendo do fluxo criado, o Orange leva uma desvantagem em relação ao Python, que demonstrou um tempo de execução quase instantâneo. (SEABRA; DRUMMOND; GOMES, 2018)

Porém, é possível ressaltar que a relevância e qualidade dos resultados obtidos se assemelham em ambos os programas, tendo em vista a demonstração gráfica coerente que tornou possível a geração de informações precisas.

Após realizar a análise dos resultados obtidos com a aplicação dos programas, foi possível extrair informações referentes a produção do petróleo por ano, região e estado, concluindo que, apesar de algumas oscilações, houve um crescente aumento na produção do petróleo no decorrer dos anos, bem como, a região caracterizada maior produtora de petróleo desde 1997 a 2018 é a região Sudeste, tendo o Rio de Janeiro com o maior volume de produção durante esses 22 anos.

As informações constatam-se assertivas tendo em vista que segundo CANELAS (2007) no ano de 1997, houve o evento da reforma da indústria petrolífera brasileira, na qual foi destituída a exclusividade legal da Petrobras para a execução das atividades de exploração e produção de petróleo e refino no Brasil, tendo diversas outras companhias se inserido na atividade de exploração e produção de gás natural no Brasil a partir do supracitado ano.

Tal ocorrência é um possível fator capaz de justificar o crescimento na produção do petróleo desde 1997, além de que de acordo com Rodrigues Neto (2009) as mudanças ocorridas com a flexibilização, impuseram a Petrobrás, desafios de aperfeiçoar as tecnologias de exploração e produção de petróleo em águas profundas (em lâminas d'água superiores a 1000 metros), e a inclusão do gás natural na matriz energética brasileira, bem como, a infraestrutura de distribuição (gasodutos) para utilização no processo de produção industrial. Nota-se que as políticas definidas pelo Estado (como maior acionista), para as atividades petrolíferas, não desvinculam as metas de produção nacional de petróleo, daquelas concernentes ao setor petroquímico, considerado além de estratégico e o mais dinâmico, para a ampliação e consolidação da estrutura industrial brasileira, ser, também,

um setor preponderante para o desenvolvimento nacional.

Independentemente de ter ou não controle estatal, a Petrobras é uma empresa antiga e enorme, praticamente a única empresa de refino que existe no Brasil, mesmo que esse mercado seja aberto para outros concorrentes há mais de 20 anos. Segundo Mendes *et al.* (2018) o Brasil foi o sétimo maior mercado de derivados de petróleo do mundo em 2017, possuindo 18 refinarias, com capacidade instalada de quase 2,3 milhões de barris/dia. Dessas 18 refinarias, 14 são de grande escala, pertencentes à Petrobras. As demais são de pequeno porte e compreendem apenas 1,6% do total da capacidade instalada.

O notório destaque da região sudeste na produção do petróleo no país justifica-se pela localização da Bacia Campos, na qual segundo Poubel (2017) abrange uma área com cerca de 100 mil quilômetros quadrados, estendendo-se do Espírito Santo, na cidade de Vitória, até o litoral norte do Rio de Janeiro, na cidade de Arraial do Cabo.

Nessa bacia, foi criado um dos maiores complexos petrolíferos do mundo, e representa a principal área sedimentar explorada no Brasil. Hoje, a Bacia de Campos se destaca por abrigar aproximadamente 80% das reservas totais de petróleo no Brasil, distribuídas num conjunto de 29 bacias sedimentares brasileiras. (MONIÉ, 2003, p. 261).

5. CONSIDERAÇÕES FINAIS

O presente trabalho visou desenvolver um estudo comparativo de duas ferramentas de análise aprofundada de dados, sendo estas, o Python e o Orange. O estudo permitiu verificar que o programa Python exige conhecimento sobre linguagem de programação, enquanto o Orange, por possuir uma interface intuitiva, basta ter uma mínima noção de lógica. Portanto, conclui-se que, ambas as ferramentas oferecem infinitas possibilidades de exploração de um banco de dados, de modo a oferecer uma análise robusta deles, no entanto, percebe-se que o programa Orange agrega uma experiência mais ágil, prática e simples, do ponto de vista de usabilidade.

Além disso, a aplicação dos programas permitiu a extração de informações que contribuíram para a ampliação de conhecimento teórico a respeito da evolução da indústria petrolífera brasileira, bem como agregou conhecimentos a respeito de sistemas de informação e *DataScience*.

REFERÊNCIAS

BORGES, Luiz Eduardo. **Python para Desenvolvedores.** 1. ed. São Paulo,SP: Novatec, 2014.

BOSCARIOLI, Clodis; VITERBO, José; TEIXEIRA, Mateus Felipe. Avaliação de Aspectos de Usabilidade em Ferramentas para Mineração de Dados. *In:* ESCOLA REGIONAL DE SISTEMAS DE INFORMAÇÃO DO RIO DE JANEIRO (ERSI-RJ), 1. , 2014, Niterói. **Anais [...].** Porto Alegre: Sociedade Brasileira de Computação, 2014 . p. 87-94.

CAMPOS, Aline de; CAZELLA, Sílvio César. Descoberta de Conhecimento em Base de Dados sobre o perfil de estudantes brasileiros de Tecnologia da Informação . *In:* CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 7., 2018, Fortaleza,CE. **Anais [...].** Fortaleza, CE, 2018**.**Disponível em: https://www.semanticscholar.org/paper/Descoberta-de-Conhecimento-em-Base-de-Dados-sobre-o-Campos-Cazella/952d078b668ebd7beaf2cb5ec8ae18e7394bd235. Acesso em: 28 out. 2019. https://doi.org/10.5753/cbie.wcbie.2018.449

CANELAS, André Luís de Souza. **Evolução da importância econômica da indústria de petróleo e gás natural no Brasil:** contribuição a variáveis macroeconômicas**.** 2007. Disponível em: http://www.ppe.ufrj.br/images/publica%C3%A7%C3%B5es/mestrado/Andr%C3%A9_Lu%C3%ADs_de_Souza_Canelas.pdf. Acesso em: 05 nov. 2019.

CRUZ, Leandro César. **DataScience:** Desenvolvimento de Aplicação para Análise de Dados. 2018**.** Disponível em: https://cepein.femanet.com.br/BDigital/arqTccs/1511420264.pdf. Acesso em: 5 no. 2019.

DEMSAR, Janez; ZUPAN, Blaz.**Orange:** Data Mining Fruitful and Fun - A Historical Perspective. 2012. 6. University of Ljubljana, Faculty of Computer and Information Science. Ljubljana, Slovenia, 2012.

GARCIA, Diego. **Debugging em python**. 2015. Disponível em:

http://pythonclub.com.br/debugging-em-python-sem-ide.html Acesso em: 11 nov. 2019.

GRUPO PET-TELE. **Tutorial de Introdução ao Python.** Universidade Federal Fluminense, Niterói, Rio de Janeiro, 2011. Disponível em: http://www.telecom.uff.br/pet/petws/downloads/apostilas/PYTHON.pdf. Acesso em: 28 out. 2019.

MANNARA, Barbara. **Convertio:** converta online suas fotos, vídeos e áudios. 2017. Disponível em: https://www.techtudo.com.br/tudo-sobre/convertio.html. Acesso em: 28 out. 2019.

MANZANO, José Augusto N. G. **Programação de Computadores com C++**. 1. ed. São Paulo, SP: Érica, 2011.

MENDES, André Pompeo do Amaral. *et al.* Mercado de Refino de Petróleo no Brasil.**Petróleo&GásBNDES**, Rio de Janeiro, v. 24, n. 48, p. 7-44, set. 2018.

MENEZES, Nilo Ney Coutinho. **Introdução à Programação com Python**, 2. ed. São Paulo, SP: Érica, 2014

MONIÉ, Frédéric; SILVA, Gerardo (Orgs.). **A mobilização produtiva dos territórios.** Instituições e logística do desenvolvimento local. Rio de Janeiro, DP&A, 2003.

RODRIGUES NETO, João. **A Petrobrás:** Da quebra do monopólio às perspectivas de produção na camada do pré-sal – 1997-2009. 2009. Disponível em: http://www.abphe.org.br/arquivos/joao-rodrigues-neto_1.pdf. Acesso em 05 nov.2019.

POUBEL, Renata. **A Bacia de Campos, os Royalties e os Trabalhadores:** Uma análise da pendularidade na Região Norte Fluminense. 2017. Disponível em: https://seminariodeintegracao.ucam-campos.br/wp-content/uploads/2018/02/A-Bacia-de-Campos.pdf. Acesso em: 06 nov. 2019.

ROSSUM, Guido Van; BOER, Jelke de. **Interactively Testing Remote Servers Using the Python Programming Language**. EurOpenConference, Tromso. 1991.

SEABRA, Rodrigo; DRUMMOND, Isabela Neves; GOMES, Fernando Coelho. Análise Comparativa de Linguagens de Programação a partir de Problemas Clássicos da Computação. **Revista de Sistemas e Computação,** Salvador, BA, v. 8, n. 1, p. 56-76, jan./jun. 2018.

SIEGEL, Idaltchion Fabricio. **Linguagem python e suas aplicações em ciência de dados.** 2018. Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Fluminense. Niterói, RJ, 2018.

SILVA, João Gabriel Rocha Silva. **Introdução à Linguagem Python**. Programa de pós-graduação em modelagem computacional. Universidade Federal de Juiz de Fora. Juiz de Fora, 2016.