

CRIAÇÃO DE UM DASHBOARD DA PRODUÇÃO DE PETRÓLEO NO PRÉ-SAL BRASILEIRO: UMA ANÁLISE UTILIZANDO *BUSINESS* *INTELLIGENCE* (BI).

Alessandra Brito Leal ¹
Dr. Thiago Rafael da Silva Moura ²

RESUMO

Os bancos de dados se tornaram componentes essenciais no cotidiano da sociedade moderna. No dia a dia, as pessoas se deparam com atividades que envolvem alguma interação com os bancos de dados, como registro de dados em escolas, hospitais, hotéis, Bancos e grandes empresas. A partir disso, para analisar os dados de produção de petróleo no Pré-Sal, utilizaremos neste trabalho o programa *Power BI* para transformamos grandes volumes de dados em um *dashboard*. Dessa forma, este trabalho tem como objetivo criar um *dashboard* para apresentar dados de produção de petróleo no Pré-Sal brasileiro. A partir da coleta, análise, tratamento e modelagem dos dados de produção, no *Power BI*, tivemos como resultado um relatório dinâmico que exibe a produção de hidrocarbonetos no Pré-Sal, com destaque para o campo de Lula (atualmente chamado de Tupi), o maior campo produtor de petróleo em bbl/d (barris de óleo por dia), produzindo mais de 1 milhão de barris por dia. Durante as análises comprovamos que Tupi também é o maior produtor de gás natural com mais de 45 milhões m³/d. Além disso, os resultados das análises mostraram que os 30 poços de petróleo que mais produzem petróleo estão localizados na bacia de Santos, bacia do Pré-Sal brasileiro. E por fim, foi possível identificar a operadora que detém a maior parte da produção brasileira, a Petrobrás, com cerca de 95% da produção nacional. Nossos resultados são muito semelhantes aos apresentados pelo site oficial da ANP, mostrando uma análise coerente dos dados fornecidos.

Palavras-chave: Base de Dados, *Business Intelligence*, Pré-Sal brasileiro, Produção de Petróleo.

INTRODUÇÃO

Conforme Batista (2004) as ferramentas de *business intelligence* são *softwares* que coletam e processam grande quantidade de dados dos mais variados meios e conteúdo. No *Power BI*, as letras BI significa *business intelligence*, ou seja, inteligência de negócios. O termo BI começou a ser utilizado para tecnologia em 1958 pelo cientista Hans Peter Luhn. Ele

¹ Graduanda do Curso de Engenharia de Exploração e Produção de Petróleo da Universidade Federal do Pará- UFPA, eng.alessandraleal@gmail.com;

² Professor orientador: Doutor, Faculdade de Física – UFPA, trsmoura@yahoo.com.br

descreveu os softwares de BI como sistemas automáticos desenvolvidos para propagar informações para diversos setores da economia. Suas pesquisas, dentre os desenvolvimentos de mais de 80 patentes, foram inicialmente aplicadas em *softwares* de análise da *International Business Machines Corporation (IBM)* (LAGO, 2018).

Dessa forma, os *softwares* de BI são capazes de transformar grandes volumes de dados em *dashboards* analíticos e resumidos. Eles são configuráveis e capazes de executar automaticamente a rotina de coleta, tratamento, análise e visualização de dados. Para Provost (2013) dentro de uma empresa, é possível aumentar os lucros, minimizar os custos, fazer previsões de gastos adicionais e tomar decisões competentes baseadas em dados. A partir dessa gama de possibilidades, é possível ainda utilizar o programa para analisar inúmeros dados de produção de petróleo.

Assim, o presente trabalho trata-se do recorte de uma pesquisa mais abrangente intitulada “*Data analytics applied to the analysis of petroleum production in Brazil*” publicada na revista *Brazilian Applied Science Review*”. O objetivo principal é a análise de dados fornecidos pela ANP (Agencia Nacional do Petróleo e Gás) para criação de um *Dashboard* onde possa ser apresentado os dados de produção de petróleo do Pré-Sal brasileiro. Esses dados foram coletados e importados para o programa, para serem analisados, tratados, modelados e apresentados a partir de visuais do programa Power BI.

METODOLOGIA

Power BI

Conforme a *Microsoft Power BI* (2020) ele é um software de serviço para análise de negócios. A *Microsoft* oferece algumas formas de utilizar o *Power BI*: “o *Power BI Desktop*”, “o Serviço do *Power BI*” e o “*Power BI premium*”. O *Power BI Desktop* é um aplicativo do Windows para executar no PC; enquanto o Serviço do *Power BI* é um serviço de nuvem utilizado por meio do navegador da web; e o *Power BI Premium* é destinada a empresas de grande porte que querem rodar localmente o servidor de relatórios do *Power BI*, garantindo um fluxo de dados maior e mais estável, normalmente rodando em hardware dedicado (AVELLA e DUO, 2019). Neste trabalho foi utilizado os Serviços do *Power BI* e o *Power BI Desktop*.

Power BI Desktop

O *Power BI Desktop* é a versão para *Windows* do programa. Essa versão permite a conexão com fontes de dados locais e online; possui uma função de alto detecção de

relacionamento para os dados carregados; permite a construção de relatórios e funções DAX (AVELLA e DUO, 2019).

A base de dados de produção de óleo foi importada e tratada no *Power BI Desktop*, e em seguida foi criado o *Dashboard*.

Serviço do Power BI

O serviço do *Power BI* é uma plataforma gratuita, baseada no serviço de nuvem utilizado por meio do navegador da web. Ele tem a capacidade de conexão com centenas de fontes de dados, compartilhamento dos relatórios, além de uma caixa de perguntas para que seja possível levantar questões a respeito dos dados (FERRARI e RUSSO, 2016).

Interface do Power BI

O *Power BI* dispõe de três painéis de navegação principais:

- **Dados**

Esse painel dar acesso ao conjunto de dados que serão conectados ao *Power BI* (ORACLE, *My SQL*, *Postgre SQL*, *Teradata*, *Sybase*, *Hadoop*, *Microsoft SQL Server* etc), aqui ocorre o tratamento dos dados. A base de dados tratada neste trabalho foi importada do Excel.

- **Modelo**

É o painel que permite o estabelecimento das relações entre os dados que foram carregados para o *Power BI*. Dessa forma, foi criado inúmeros relacionamentos a partir de tabelas que possuem informações semelhantes.

- **Relatório**

É a parte do programa onde foi permitido a visualização dos dados (em forma de gráficos, tabelas e cartões) após as etapas de importação, tratamento e relacionamento.

Criação do Dashboard

Um *dashboard* (ou relatório em português) é uma interface gráfica que fornece visualizações rápidas dos principais indicadores de desempenho relevantes para um objetivo ou processo de negócio específico. É possível exibir uma página da web vinculada a um banco de dados que permite que o relatório seja atualizado constantemente.

Tratamento de Dados

Coletamos os dados de produção de petróleo fornecidos pela ANP e importamos para o *Power BI*. Para tratar esses dados, utilizamos a interface do Editor do *Power Query* (essa interface é separada do *Power BI*). A linguagem de funções M é base do *Power Query*, proveniente do termo *data mashup*, ou seja, obtenção e combinação de dados (LAGO, 2018).

Existem mais de 600 funções na linguagem M que, quando combinadas, proveem uma variação abrangente para obtenção, tratamento e junção de dados. Seu principal conceito é na automatização de todo esse processo, sendo necessário se conectar apenas uma vez com a fonte de dados e tratá-la para que as próximas atualizações sigam o mesmo padrão (FERRARI e RUSSO, 2016).

Importamos os dados para Power Query Editor para serem manipulados, transformados, enriquecidos, mesclados e adicionados para só depois serem aplicados no *Power BI*.

Relacionamento e Modelagem de Dados

O conceito de modelagem de dados é um processo lógico de relações entre diferentes dados, enaltecendo suas possíveis ligações para atender e responder às necessidades de negócios (FERRARI e RUSSO, 2016). A partir disso, relacionamos os dados de acordo com as finalidades que queremos apresentar.

Para facilitar a visualização dos dados, foi criado dois tipos de tabelas: uma Fato (tabela maior) e outra Característica (tabelas menores), sendo que as duas precisam ter uma chave primária e uma chave estrangeira para se conectarem, ou seja, uma coluna semelhante entre as duas tabelas que permita o relacionamento entre os dados. A partir disso, foi possível realizar os cálculos e procedimentos necessários para exibir os resultados.

Cálculos

Para realizarmos cálculos no *Power BI*, utilizaremos as chamadas funções DAX, que é uma linguagem de expressões utilizada para realizar cálculos e consultas sobre modelos de dados criados, sendo muito similar às funções do Excel. DAX é um acrônimo para *Data Analysis Expressions*, ou seja, expressões de análise de dados.

Como os dados foram organizados de maneira a gerar resultados com o mínimo de cálculos possíveis, foi necessário apenas, mesclar colunas, criar medidas de soma e média, além de adicionar novas colunas de tempo a partir da base de dados calendário, para assim, obter os resultados desejados e criar o *dashboard*.

Visuais utilizados na criação do Dashboard

Existem várias maneiras de visualizar os dados, tal qual, cartões, matrizes, gráficos, mapas etc. Cada uma delas apropriada para um tipo de dado estatístico. Para facilitar a interpretação das informações contidas nos inúmeros dados foi necessário escolher o visual adequado e assim apresentar os dados.

Para definir o tipo de gráfico (visual) mais adequado, primeiramente deve-se definir o que é importante mostrar a partir da base de dados que está sendo analisada, por exemplo,

apresentar uma distribuição, comparação, composição ou relação entre os dados de produção de hidrocarbonetos.

No *Power BI*, foi possível visualizar as informações que a base de dados possui de forma rápida e inteligente. Isso foi executável através da ferramenta conhecida como “Visual”, a partir da escolha do Visual (matriz, gráficos, mapas etc.) adequado para a informação que se deseja apresentar, foi possível criar um *Dashboard* completo e interativo.

Além das formas de visualizações convencionais, o *Power BI* também é capaz de apresentar mais de uma informação no mesmo gráfico transformando a legenda do gráfico em outro visual, esse procedimento é conhecido como *tooltip* e também foi explorado neste trabalho.

REFERENCIAL TEÓRICO

Banco de Dados

Antes do surgimento do computador na década 40, os dados eram guardados em papel. Nessa época, as fichas eram tratadas como registros, as pastas como tabelas e os armários como arquivos. Esse meio de armazenamento deixava um acúmulo muito grande de papel. Assim, o grande desafio da década de 50 foi digitalizar todos esses dados, chamando-os de arquivos sequenciais (RAMAKRISHNAN e GEHRKE, 2000).

Após as fitas, surgiram os discos, disquetes e HDs, esses tipos de mecanismos armazenavam dados de maneira direta, não de maneira sequencial. Agora, é possível guardar todos os registros e manter dentro de uma de tabela, que têm numerações ou chave identificadoras de cada um dos registros, chamados de arquivos de acesso direto.

No ano de 1959, o departamento de defesa dos Estados Unidos criou um evento conhecido como CODASYL (Conference on Data Systems Languages), esse evento reuniu militares, empresas e universidades, onde foram discutidas grandes tecnologias. A partir desse evento surgiu a primeira linguagem que se preocupou com a lógica da programação e os dados, conhecida como COBOL (*Common Business Oriented Language*), como também foi discutido uma nova tecnologia, que ganhou o nome de Banco de Dados (CERUZZI, 2003).

Após o surgimento do termo “Banco de Dados”, a IBM desenvolveu o Sistema de Gerenciamento de Informações (IMS), usado até hoje. O IMS formou a base para uma estrutura alternativa de representação de dados chamada modelo de dados hierárquico. Em seguida, em 1970, Edgar Codd, pesquisador da IBM, propôs uma nova estrutura de representação de dados

chamada modelo de dados relacional. A partir dessa criação foi possível relacionar um conjunto de registros que está numa tabela com outro conjunto de registros que está em outra tabela (RAMAKRISHNAN e GEHRKE, 2000). Agora, os dados possuem relação e podem ser acessados por vários usuários.

Um banco de dados contém: uma base de dados, um sistema gerenciador e uma linguagem de exploração. Dessa forma, conforme Elmasri e Navathe (2011) podemos definir o conceito de Banco de Dados como sendo uma coleção de dados relacionados. Os dados são fatos que podem ser gravados e que possuem um significado implícito.

Um sistema gerenciador de banco de dados (SGBD) é uma coleção de programas que permitem aos usuários criar e manter um banco de dados, possui um sistema de *software* de propósito geral que facilita os processos de definição, construção, manipulação e compartilhamento do banco de dados entre vários usuários e aplicações (Ibid, 2011).

Dessa forma, a definição de um banco de dados implica especificar os tipos de dados, as estruturas e as restrições para os dados a serem armazenados em um banco de dados. A partir dos conceitos apresentados, é possível afirmar que a construção de um banco de dados é o processo de armazenar os dados em alguma mídia apropriada controlada pelo SGBD. A manipulação inclui algumas funções, como pesquisas em banco de dados para recuperar um dado específico, atualização do banco para refletir as mudanças no “minimundo” e gerar os relatórios dos dados. O compartilhamento permite aos múltiplos usuários acessar de forma concorrente o banco de dados (ELMASRI e NAVATHE, 2011).

Ciência de Dados

Há uma grande quantidade de dados gerados diariamente nas grandes empresas (HAIDER, 2015), segundo Turban et al. (2009) vários setores da economia estão focados com o objetivo de explorar dados e tomar decisões para obter vantagem competitiva.

No passado, as grandes companhias podiam empregar equipes de estatísticos, modeladores e analistas para explorar conjuntos de dados manualmente, mas hoje, o grande número e a variedade de dados ultrapassaram em muito a capacidade da análise manual. Ao mesmo tempo, os computadores se tornaram muito mais poderosos, e algoritmos foram desenvolvidos para conectar conjuntos de dados e permitir análises mais profundas, o que anteriormente era impossível. A convergência desses fenômenos deu origem à aplicação comercial cada vez mais ampla dos princípios da ciência de dados e técnicas de mineração de dados (PROVOST e FAWCETT, 2013).

Dessa forma, podemos introduzir três conceitos fundamentais: o de ciência de dados, mineração de dados e *Big data*. A partir do trabalho de Provost e Fawcett (2013), são eles:

A ciência de dados é um conjunto de princípios fundamentais que orientam a extração de conhecimento a partir de dados. A mineração de dados é a extração de conhecimento de dados, por meio de tecnologias que incorporam esses princípios. Como termo, "ciência de dados", muitas vezes é aplicado de forma mais ampla do que o uso tradicional de "mineração de dados", a tecnologia de mineração de dados fornece algumas das ilustrações mais claras dos princípios da ciência de dados.

Big data significa essencialmente conjuntos de dados muito grandes para sistemas de processamento de dados tradicionais e, portanto, exigem nova tecnologia de processamento. Tal como acontece com as tecnologias tradicionais, as tecnologias de *big data* são usadas para muitas tarefas, incluindo engenharia de dados. Ocasionalmente, essas tecnologias são realmente usadas para implementar técnicas de mineração de dados. No entanto, com muito mais frequência *big data* é usado para processamento de dados em apoio à tecnologia de mineração de dados e outras atividades de ciência de dados.

A partir desses conceitos, pode-se dizer que o sucesso no ambiente de negócios orientado a dados de hoje, estão associados às ideias fundamentais que se aplicam a problemas de negócios específicos.

RESULTADOS E DISCUSSÃO

Por meio da análise, modelagem e relacionamento dos dados de produção de petróleo, foi possível gerar gráficos e tabelas para visualização dos dados. Os resultados apresentados neste trabalho têm a finalidade de reproduzir as informações retiradas de cada dado de uma tabela ou conjunto de tabelas relacionadas.

Os dados fornecidos abrangem a produção de petróleo até julho de 2020. Os valores apresentados são passíveis de correções e não correspondem a totalidade da produção nacional. Logo, os resultados deste trabalho, tem a finalidade de apresentar a produção de hidrocarbonetos no Pré-Sal, a partir de análises de um banco de dados.

Análise da Produção no Pré-Sal

Analisando os dados do Pré-Sal Brasileiro fornecidos pela Agencia Nacional de Petróleo (ANP), obtemos dados semelhantes ao encontrados no site oficial da ANP em julho de 2020. Além disso, identificamos o campo que possui a maior produção diária de petróleo.

Figura 1 - Quantidade de petróleo produzidos nos campos do Pré-Sal brasileiro (julho/2020)

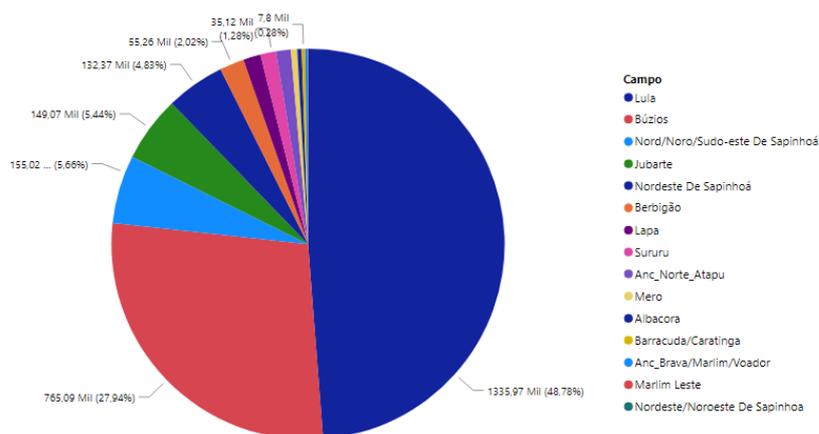
Campo	Petróleo (bbl/d)	Gás natural (Mm ³ /d)	Produção Total (boe/d)
Lula	1.049.265,17	45.581,98	1.335.967,15
Búzios	614.903,38	23.877,33	765.087,26
Nord/Noro/Sudo-este De Sapinhoá	122.388,81	5.187,77	155.018,87
Jubarte	121.080,18	4.450,07	149.070,26
Nordeste De Sapinhoá	101.993,76	4.829,30	132.369,12
Berbigão	51.008,69	675,18	55.255,45
Lapa	32.303,83	1.103,10	39.242,09
Sururu	29.299,00	925,28	35.118,86
Anc_Norte_Atapu	26.941,59	994,87	33.199,11
Mero	10.081,91	688,10	14.409,96
Albacora	8.158,59	245,86	9.704,99
Barracuda/Caratinga	6.751,34	166,95	7.801,43
Anc_Brava/Marlim/Voador	4.502,74	138,19	5.371,91
Marlim Leste	978,46	17,39	1.087,84
Nordeste/Noroeste De Sapinhoa	0,58	0,02	0,70
Total	2.179.658,03	88.881,37	2.738.704,99

Fonte: Próprio autor

A figura 1 exibe os campos produtores de hidrocarbonetos no Pré-Sal brasileiro. O campo que possui a maior produção diária é o de Lula com mais de 1 milhão de barris produzidos por dia. Lula também é o maior produtor de gás natural com mais de 45 milhões m³/d.

O segundo maior produtor de petróleo é o campo de Búzios com 614,9 mil bbl/d e uma produção de gás natural de 23,8 milhões m³/d.

Figura 2 - Proporção da produção de óleo e gás natural por campo no Pré-Sal (julho/2020)



Fonte: Próprio autor

O gráfico 2, foi plotado afim de mostrar a proporção da produção dos campos de petróleo no Pré-Sal. A partir dele é possível perceber que os campos de Lula e Búzios detêm a maior parte da produção diária de petróleo, os dois campos somam mais de 75 % da produção no Pré-Sal, sendo que Lula é o responsável por quase metade do total da produção no Pré-Sal.

Figura 3 - Poços que mais produzem petróleo no Pré-Sal em boe/d (julho/2020)



Fonte: Próprio autor

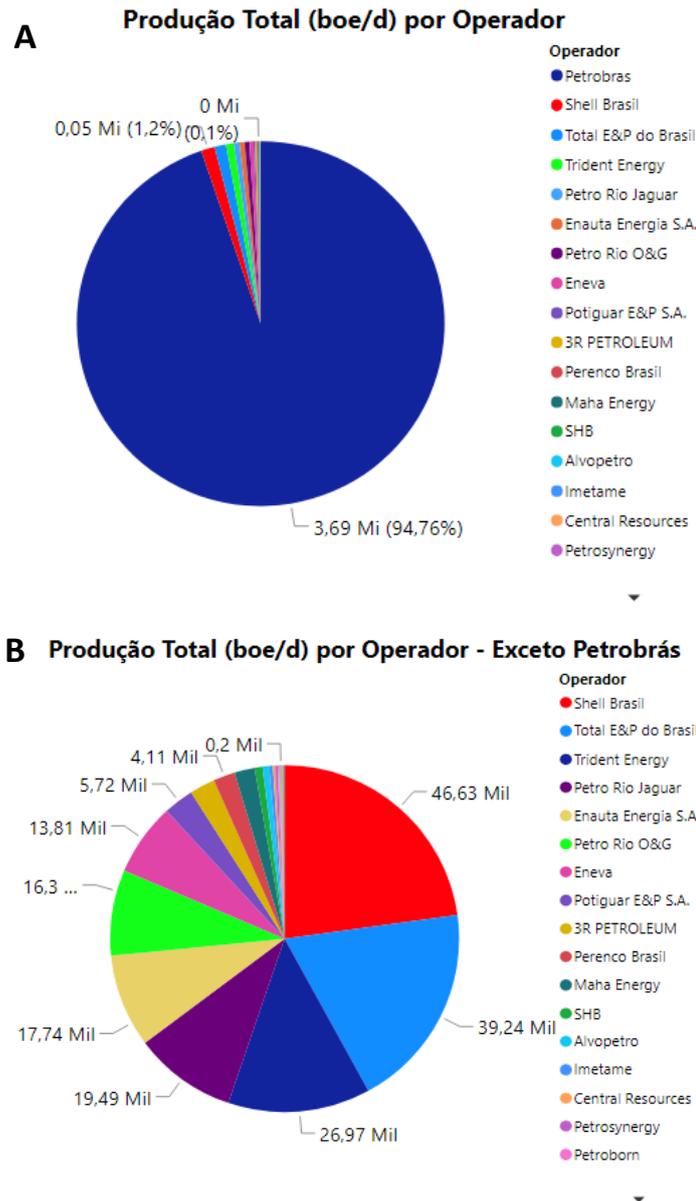
No gráfico de árvore (gráfico 3) é possível ver quais os poços que mais produzem petróleo por dia. Os maiores poços produtores são do campo de Búzios. Em relação a produção por campo, Búzios não é o campo que mais produz, porém, os poços produzem uma quantidade grande de hidrocarbonetos por dia, ou seja, Búzios possui poços que produzem mais por dia do que os poços do campo de Lula (maior campo produtor do Brasil).

Figura 4 - As 30 instalações com maior produção de petróleo por dia em boe/d (julho/2020)



Fonte: Próprio autor

Figura 5 - Produção de Petróleo por Operador com a Petrobras e sem a Petrobras (julho/2020)



Fonte: Próprio autor

A figura 4 mostra as 30 unidades estacionárias de produção (UEPs) que mais produzem petróleo por dia. Mais da metade das UEPs que mais produzem são da Petrobras, a maior unidade produtora diária (P-76) está no campo de Búzios (campo do Pré-Sal) produzindo aproximadamente 210 mil boe/d (produção de óleo e gás). Podemos confirmar os dados apresentados na figura 4 através da figura 5, onde é possível visualizar a produção de petróleo diária distribuída por operador. No lado esquerdo da imagem (figura 5A) é notório que a Petrobras é a maior produtora de petróleo com 3,68 milhões de boe/d (barris de óleo equivalente por dia), 94,7% da produção nacional. Para que seja possível a visualização da produção das demais operadoras, retiramos a Petrobras na figura 1B. Assim, percebe-se que o segundo maior

operador no Brasil (conforme os dados considerados até julho de 2020) é a Shell com 46,63 mil boe/d. Logo em seguida, temos a TOTAL com uma produção de 39,24 mil boe/d.

CONSIDERAÇÕES FINAIS

Logo, todo conteúdo que compõem os resultados deste trabalho foi baseado em análises de dados, para assim extrair as inúmeras informações contidas nessas bases e gerar Visuais (gráficos de setor, mapas, área e tabelas) para criação de um *dashboard* com dados de produção de petróleo no Pré-Sal. Além de um estudo do conceito de bancos de dados e do programa *Power BI* onde foram feitas as análises, bem como o tratamento e modelagem dos dados utilizados, foi possível extrair informações para compor os resultados deste trabalho.

De acordo com os dados de produção de hidrocarbonetos analisados é possível concluir que a bacia que mais produz hidrocarbonetos por dia é a de Santos, seguida pela de Campos, bacias do Pré-Sal brasileiro onde a Petrobras atua como a maior produtora.

A maior parte da produção brasileira vem do campo de Tupi, na bacia de Santos, produzindo mais de 1 milhão de barris de óleo por dia e 45 milhões de m³/d de gás natural, sendo que os 30 poços de petróleo que mais produzem por dia estão localizados também na bacia de Santos.

Os resultados deste trabalho acrescentaram de forma significativa para apresentação de dados de produção de petróleo no Pré-Sal a partir de um *dashboard*, uma vez que foram minerados dados brutos e transformados em gráficos interativos e de fácil interpretação. As análises feitas, neste trabalho, são muito semelhantes aos apresentados pelo site oficial da ANP, mostrando uma análise coerente dos dados de produção fornecidos.

REFERÊNCIAS

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS. Disponível em: <<https://www.gov.br/anp/pt-br>>. Acesso em: 23 de set. de 2020.

AVELLA, Marcello; DUO, Gabriel. **Clarify: Excelência em Cursos de Tecnologia e Gestão**, 2019.

CERUZZI, Paul E. **A History of Modern Computing**. Massachusetts London: MIT Press Cambridge, 2003.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistema de Banco de Dados**. São Paulo: Pearson Universidades, 2019.

FERRARI, Alberto; RUSSO, Marco. **Introducing Microsoft Power BI**. Washington: Microsoft Press, 2016.

LAGO, Karine Gouvea Dibai; ALVES, Laennder. **Dominando o Power BI**. Belo Horizonte, 2018.

LEAL, A. B; MOURA, T. R. S. Data analytics applied to the analysis of petroleum production in Brazil. **Brazilian Applied Science Review**, 2021.

MICROSOFT Power BI. **Business Intelligence**. Microsoft Corporation. Disponível em: <<https://powerbi.microsoft.com/pt-br/>> Acesso em: 27 de set. de 2020.

PROVOST, Foster; FAWCETT, Tom. **Data Science for Business**. What you Need to Know About Data Mining and Data-Analytic Thinking. California: O'Reilly Media, 2013.

RAMAKRISHNAN, Raghu; GEHRKE, Johannes; **Sistemas de Gerenciamentos de Bancos de Dados**. McGraw Hill Brasil, 2008.

TURBAN, E. et al. **Business Intelligence: um enfoque gerencial para a inteligência do negócio**. Porto Alegre: Bookman, 2009.