**Produção de petróleo**

**Projeto da disciplina de BI**

Prof. Anderson Nascimento

prof.anderson@ica.ele.puc-rio.br

**Componentes do projeto:**

1 **Cleyde Marlyse De Andrade** - cleyde.andrade@gmail.com

2 **Edson Jandiroba** - ejandiroba@gmail.com

3 **Fabiano Rocha** - Ffrancorocha@yahoo.com.br

4 **Ludmila Müller da Silva** - ludmilamuller7@gmail.com

5 **Monica Soares Brandao** - monbrandao@gmail.com

**Histórico de versões**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data** | **Versão** | **Descrição** | **Autor** | **Aprovado por** |
| 06/07/2019  13/07/2019 | - | Discussão do projeto | Mônica, Ludmila, Cleyde, Leonardo, Edson, Marcelo |  |
| 14/07/2019 | 1.0 | Redação | Ludmila/ Edson |  |
| 06/07/2019  17/07/2019 | - | Extração e consolidação de dados | Ludmila |  |
| 27/07/2019 | 2.0 | Modelo Transacional e Modelo DW | Monica |  |
| 16/08/2019 |  | Tabelas no AWS | Cleyde |  |
| 17/08/2019 | 3.0 | ETL e Dashboard | Edson |  |
| 18/08/2019 |  | Redação | Ludmila |  |
| 23/08/2019 | 4.0 | Redação e Dashboard | Fabiano |  |
| 23/08/2019 |  | Redação final | Ludmila |  |

**Sumário**

**[1 Introdução 4](#_heading=h.30j0zll)**

[**2 Estudo de Caso 5**](#_heading=h.1fob9te)

[**2.1 Descrição do estudo de caso 5**](#_heading=h.3znysh7)

[**3 Descrição do modelo transacional 6**](#_heading=h.2et92p0)

[**3.1 Fonte 1 – Diagrama de classes 6**](#_heading=h.tyjcwt)

[**3.2 Fonte 2 – Reuniões e entrevistas para levantamento de informações 6**](#_heading=h.3dy6vkm)

[**4 Proposta de processo de BI 8**](#_heading=h.1t3h5sf)

[**5 Modelo multidimensional 9**](#_heading=h.4d34og8)

[**6 Elaboração do Data Warehouse 10**](#_heading=h.2s8eyo1)

[**6.1 Definição do DW 10**](#_heading=h.17dp8vu)

[**6.1.1 Arquitetura 10**](#_heading=h.3rdcrjn)

[**6.1.2 Abordagem de Construção 10**](#_heading=h.26in1rg)

[**6.1.3 Arquitetura Física 10**](#_heading=h.lnxbz9)

[**7 Projeto de ETL 11**](#_heading=h.35nkun2)

[**7.1 Descrição do projeto de ETL 11**](#_heading=h.1ksv4uv)

[**8 Dashboard 12**](#_heading=h.44sinio)

[**8.1 Descrição da elaboração 12**](#_heading=h.2jxsxqh)

[**8.2 Telas do Dashboard 12**](#_heading=h.z337ya)

[**9 Conclusão 18**](#_heading=h.3j2qqm3)

[**10 Anexos 19**](#_heading=h.1y810tw)

[**10.1 Anexo 1 19**](#_heading=h.4i7ojhp)

[**10.2 Anexo 2 19**](#_heading=h.2xcytpi)

# 1 Introdução

Este documento tem por finalidade coletar, analisar e definir as principais necessidades do projeto do estudo de caso “Produção de Petróleo”. O documento procura demonstrar os principais problemas e o foco investigativo desejado pelo cliente.

A proposta deste trabalho é desenvolver uma solução de business intelligence para a empresa fictícia Kioleo, a fim de auxiliar o processo de tomada de decisão da companhia, especialmente na hora de adquirir novos ativos.

A partir dos dados abertos disponibilizados pela ANP - Agência Nacional do Petróleo, Gás Natural e Biocombustíveis - foi possível adquirir a base de dados necessária para o andamento do projeto e, consequentemente, apresentar diferentes abordagens e cenários para que a empresa Kioleo tome uma decisão com o melhor o custo-benefício.

# 2 Estudo de Caso

Nesta seção é apresentada a descrição do negócio abordado, o cenário atual, as necessidades e os objetivos do projeto de Business Intelligence.

## 2.1 Descrição do estudo de caso

Tendo em vista o aumento da competitividade nos mercados globais de energia nos últimos anos e a estimativa de que a demanda por petróleo crescerá nos próximos tempos, atingindo seu pico na década de 2040, a empresa japonesa petrolífera Kioleo contratou uma consultoria de brasileiros com o intuito de apurar o cenário do mercado no Brasil.

Durante as reuniões com os consultores, a empresa Kioleo apresentou seu desejo de expandir seu portfólio em produção de petróleo, dispondo de R$ 5 bilhões CAPEX em conversão direta. A empresa japonesa frisou que era imprescindível para a sua tomada de decisão informações que permitissem entender e identificar qual seria a melhor bacia, setor e bloco para o investimento.

Pelas definições do corpo estratégico da companhia oriental, a empresa pretende realizar o investimento em bens de capital em 2021. Para que o cenário de expansão proceda é necessário levar em consideração a vida útil esperada de produção petrolífera do poço, a fim de calcular a melhor escolha com o menor custo.

Os consultores do projeto, após entenderem o que a empresa Kioleo almeja, decidiram implementar um projeto de BI com bases de dados de produções petrolíferas no Brasil, tanto em terra quanto em mar, entre os anos de 1994 a 2018 e 1942 e 2018, respectivamente. A série histórica de dados foi coletada pela ANP e disponibilizada no site http://www.anp.gov.br/dados-abertos-anp.

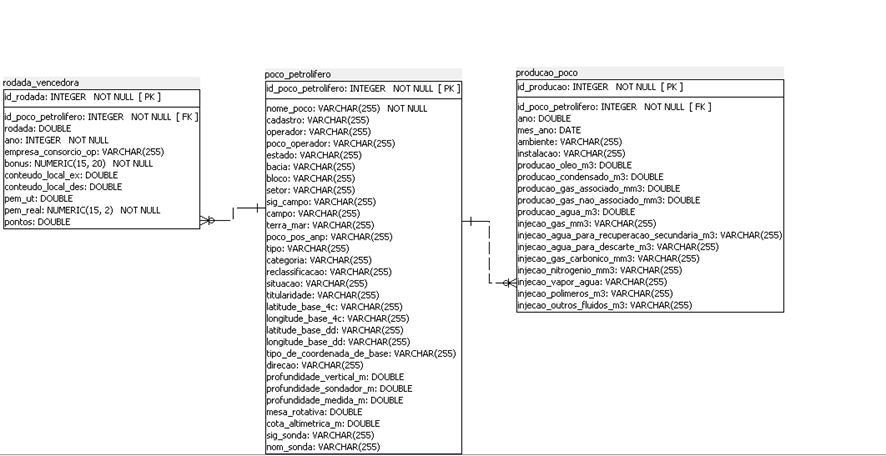
Também foram coletados dados de investimentos das licitações passadas, obtidos pelo site da ANP - http://rodadas.anp.gov.br/pt/resultados e http://rodadas.anp.gov.br/pt/entenda-as-rodadas/as-rodadas-de-licitacoes.

# 3 Descrição do modelo transacional

## 3.1 Fonte 1 – Diagrama de classes

O modelo atual consiste em três tabelas de dados não normalizados. Como os dados abertos disponibilizados pela ANP encontram-se no formato .csv, foi necessário realizar um tratamento para que fossem adequados ao modelo atual, o qual resultou nas seguintes tabelas:

* Poço petrolífero: Dados relativos aos poços petrolíferos em 2019;
* Produção do poço: Representa as métricas relacionadas à produção dos poços;
* Rodada vencedora: Como o processo de exploração dos poços é feito via rodadas de licitação, esta tabela representa as métricas das rodadas anteriores.



*Figura 1 - Modelo transacional. Fonte: elaboração própria.*

## 3.2 Fonte 2 – Reuniões e entrevistas para levantamento de informações

Durante a fase inicial do projeto foram realizadas reuniões tanto com o quadro estratégico quanto com o quadro operacional para compreensão do negócio e necessidades da empresa. Tal fase permitiu que a consultoria compreendesse o ramo do petróleo e das particularidades da empresa Kioleo.

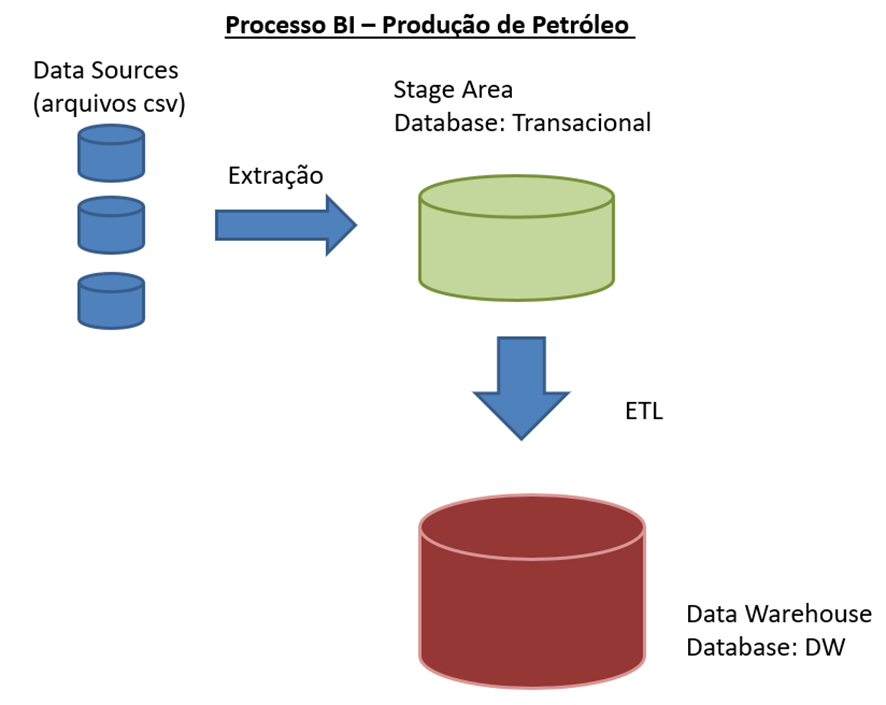
Após análise do contexto internacional na qual está inserida a companhia japonesa, bem como seus planos de expansão em terras brasileiras, foi possível levantar os principais pontos a serem abordados, quais informações necessitavam ser coletadas e como deveriam ser visualizadas em seus dashboards.

Durante os encontros e discussões entre a consultoria e a empresa Kioleo, ficou decidido que o mais importante para a empresa, no momento atual, seriam os seguintes pontos:

* A fidedignidade dos dados petrolíferos brasileiros;
* A representação visual destas informações para as partes estratégica, tática e operacional, a fim de auxiliar a compreensão do cenário e do processo de tomada de decisão.

# 4 Proposta de processo de BI

A proposta para a empresa Kioleo é a criação de uma solução de BI. Dessa forma, buscou-se utilizar uma ferramenta de ETL para limpeza, transformação e migração dos dados disponíveis no site da ANP para um banco de dados (Data Warehouse).

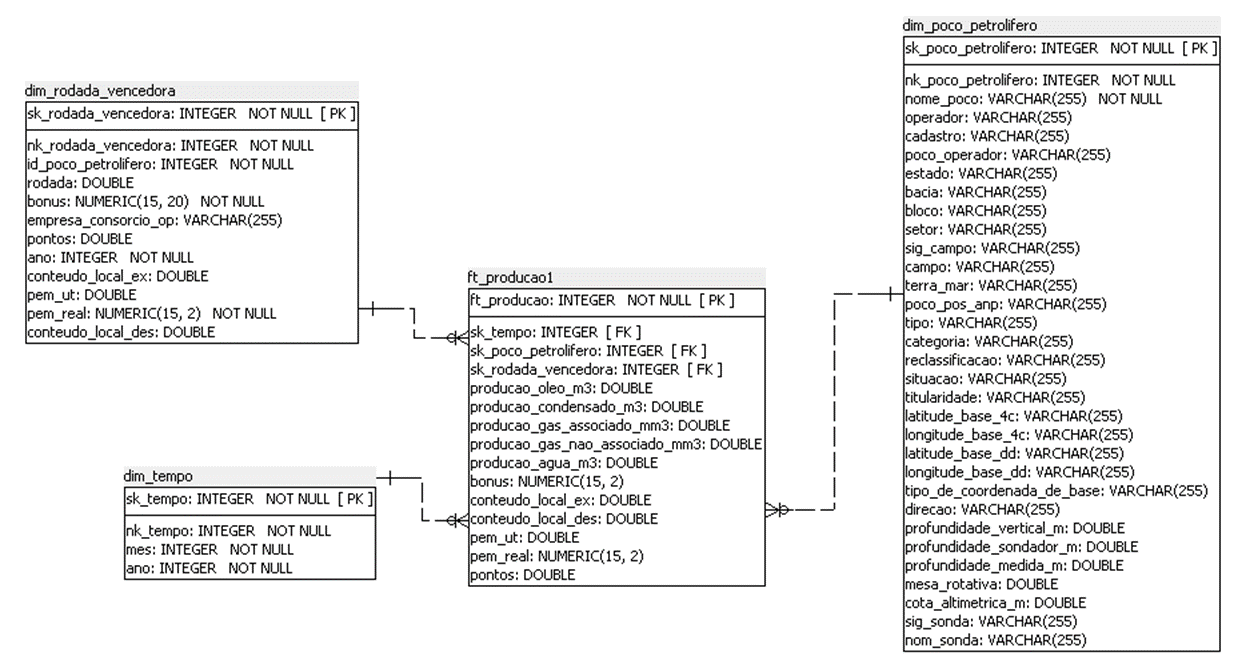


*Figura 2 - Processo de BI proposto. Fonte: elaboração própria.*

# 5 Modelo multidimensional

Esta seção apresenta a proposta de modelo multidimensional para o estudo de caso de Produção de Petróleo.

Para esta abordagem foi utilizada a modelagem multidimensional no modelo estrela - star schema.



*Figura 3 - Modelo multidimensional. Fonte: elaboração própria.*

# 6 Elaboração do Data Warehouse

O Data Warehouse será a fonte integradora de informações da empresa. A tecnologia será utilizada com o intuito de servir de base para a camada de aplicação cuja responsabilidade será fornecer dados para a tomada de decisão na organização.

## 6.1 Definição do DW

### **6.1.1 Arquitetura**

A arquitetura proposta para o DW é a Global Centralizada, a fim de suportar as necessidades da empresa como um todo e, por conseguinte, permitir uma visão corporativa para a análise de dados e tomada de decisão.

### **6.1.2 Abordagem de Construção**

A abordagem proposta é a Top Down, dispensada a construção de Data Marts neste primeiro momento. Com isso a empresa terá o controle de centralização de regras e visão do empreendimento.

### **6.1.3 Arquitetura Física**

A arquitetura física escolhida será on-premise, utilizando os recursos próprios já disponíveis para que não haja custos financeiros além dos necessários. Foi utilizado o banco de dados PostgreSQL 11.2.

# 7 Projeto de ETL

## 7.1 Descrição do projeto de ETL

Para o carregamento dos dados foi utilizado o PDI - Pentaho Data Integration (Spoon). Esta ferramenta foi aproveitada amplamente para o carregamento dos dados transacionais, transformação e carga para as tabelas de dimensão.

Para os dados transacionais realizou-se a leitura dos dados em formato .csv e carregamento nas tabelas transacionais. Todas as cargas de dados transacionais têm, em seus respectivos scripts, o prefixo “TRA”. A carga dos dados de poços petrolíferos, por exemplo, pode ser encontrada no arquivo “TRA Carga poco\_petrolifero.ktr”. Todos os dados transacionais foram carregados no banco de dados “Transacional”.

Para os dados dimensionais foi feita a leitura das tabelas transacionais, o tratamento e a carga em suas respectivas tabelas, respeitando o modelo multidimensional.

Todas as cargas e transformações dimensionais têm, em seus respectivos scripts, o prefixo “DW”. A carga dos dados de poços petrolíferos, por exemplo, pode ser encontrada no arquivo “DW Carga dim\_poco\_petrolifero.ktr”. Todos os dados dimensionais foram carregados no banco de dados “DW”.

Para a Dimensão Tempo optou-se por uma a chave natural concatenada entre ano e mês.

Para a tabela Fato foi necessário realizar a carga incremental da tabela através do Pentaho para substituir os “inner joins”. O processo encontra-se nos scripts “DW Carga ft\_producao.ktr” e “DW Carga ft\_producao2.ktr”.

Todos os scripts utilizados estão documentados no arquivo “Scripts.zip”.

# 8 Dashboard

## 8.1 Descrição da elaboração

Os Dashboards foram construídos com a ferramenta Power BI.

Como a empresa Kioleo possui parceria com a Microsoft e manifestou, durante a fase de entrevista, que o dashboard seria de extrema importância, a consultoria aconselhou que fosse adotado o PowerBI. Dessa forma, foi possível aliar uma forte ferramenta de visualização com um melhor custo-benefício para a companhia.

## 8.2 Telas do Dashboard

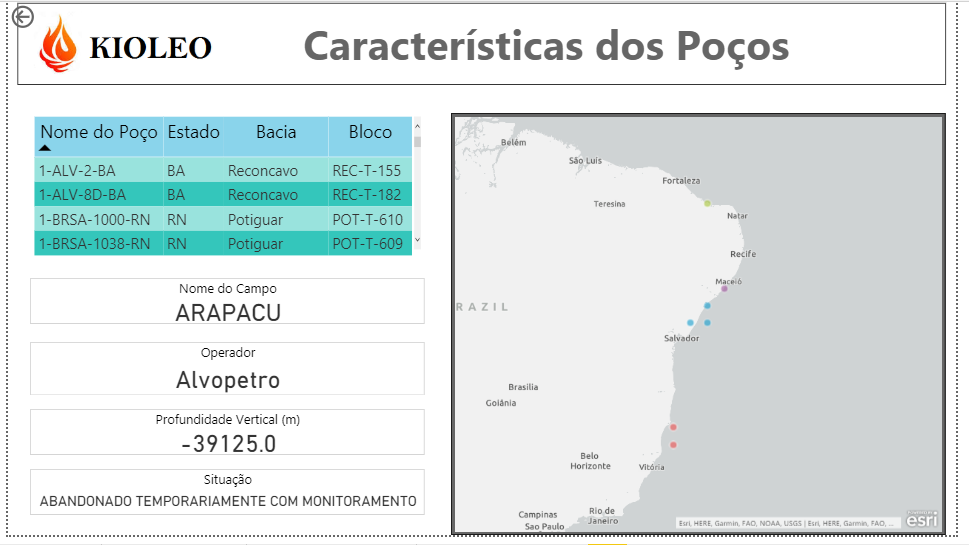
As figuras a seguir ilustram os dashboards desenvolvidos na ferramenta Power BI.

A aplicação desenvolvida está no arquivo “Kioleo r5.pbix”.

**Dashboard 1: “Características dos Poços”**

A Figura 4 apresenta as principais características dos poços de exploração, incluindo-se:

* Estado no qual se localiza;
* Bacia na qual se localiza;
* Bloco no qual se localiza;
* Nome do campo de exploração;
* Nome da empresa operadora do poço;
* Profundidade Vertical (m);
* Localização geográfica no mapa;
* Situação (em exploração, fechado, etc.).

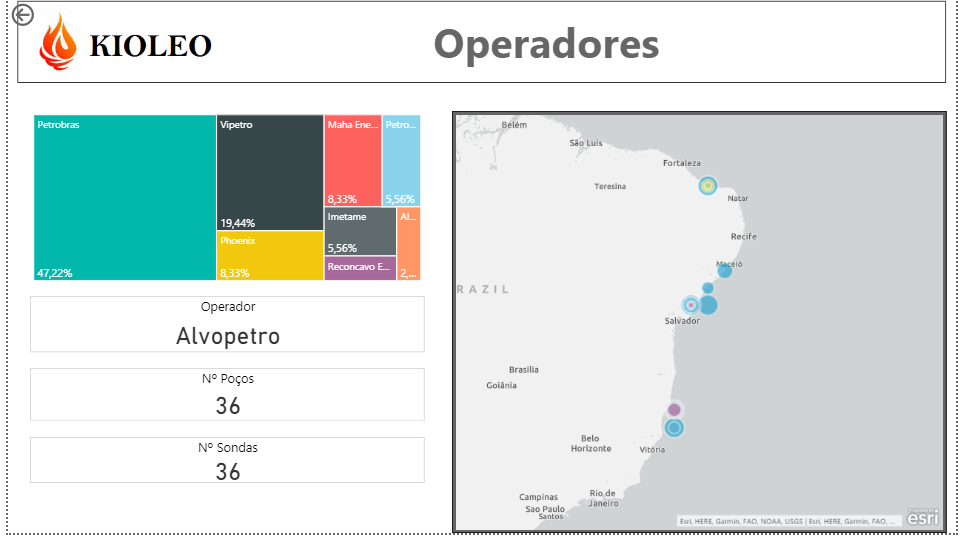


*Figura 4 – Dashboard 1:“Características dos Poços”. Elaboração própria.*

**Dashboard 2: “Operadores”**

A Figura 5 apresenta as principais informações acerca das empresas que exploram os poços, incluindo-se:

* Lista dos nomes das empresas operadoras;
* Nº percentual de poços operados por cada empresa;
* Nº absoluto de poços operados por cada empresa;
* Nº de sondas de exploração operadas por cada empresa;
* Localização geográfica dos poços no mapa;

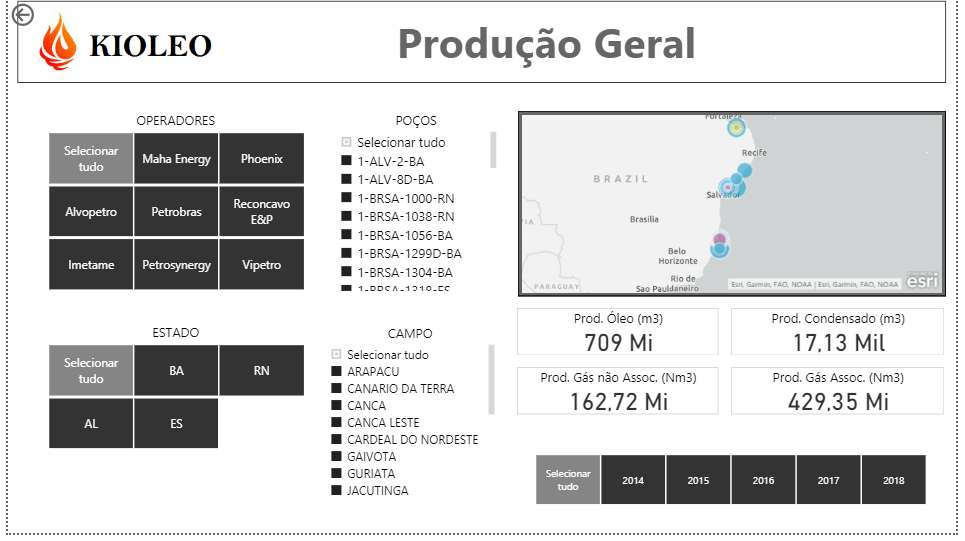


*Figura 5 – Dashboard 2: “Operadores”. Elaboração própria.*

**Dashboard 3: “Produção Geral”**

A Figura 6 mostra as principais informações acerca da produção dos poços, incluindo-se:

* Nome da empresa operadora;
* Unidade da Federação na qual se localiza o poço.
* Nome do poço;
* Nome do campo do poço;
* Produção de óleo bruto (m3) por período de tempo;
* Produção de condensado (m3) por período de tempo;
* Produção de gás associado (Nm3) por período de tempo;
* Produção de gás não associado (Nm3) por período de tempo;
* Localização geográfica dos poços no mapa;

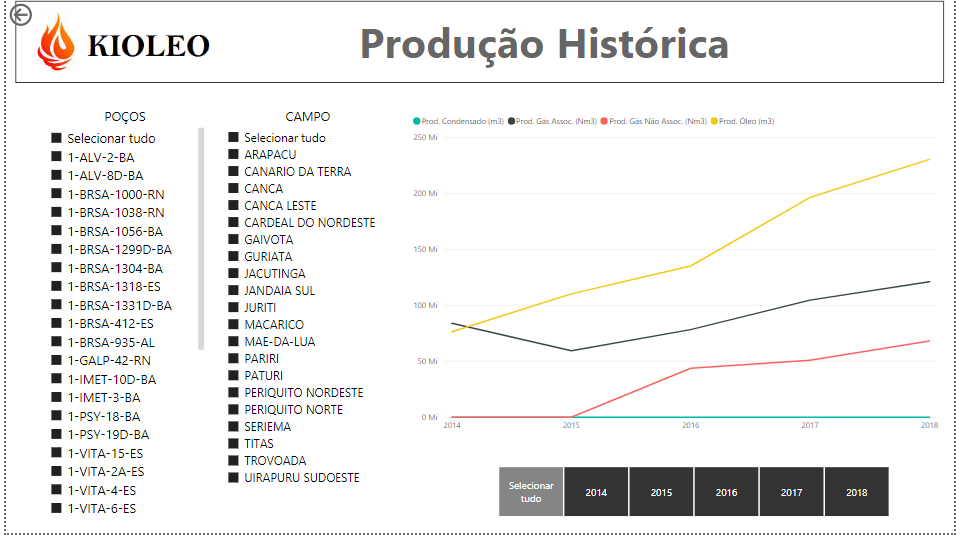


*Figura 6 - Dashboard 3: "Produção Geral". Elaboração própria.*

**Dashboard 4: “Produção Histórica”**

Pela Figura 7 é possível obter as principais informações acerca da produção histórica dos poços, incluindo-se:

* Nome do poço;
* Nome do campo no qual o poço se localiza;
* Evolução da Produção de óleo bruto (m3) por período de tempo;
* Evolução da Produção de condensado (m3) por período de tempo;
* Evolução da Produção de gás associado (Nm3) por período de tempo;
* Evolução da Produção de gás não associado (Nm3) por período de tempo.

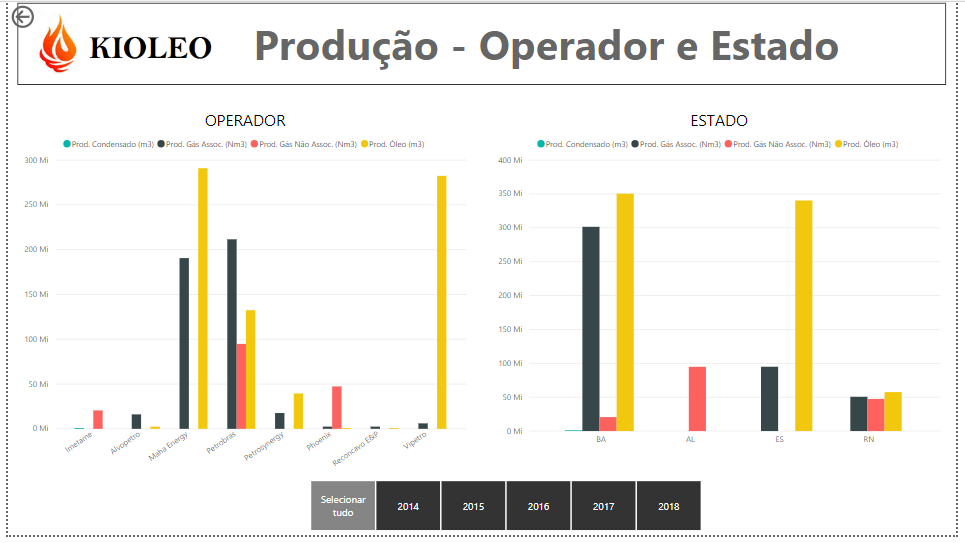


*Figura 7 - Dashboard 4: “Produção histórica”. Elaboração própria.*

**Dashboard 5: “Produção Operador e Estado”**

Na Figura 8 é possível obter as principais informações acerca da produção histórica dos poços por empresa operadora e estado da federação, incluindo-se:

* Nome da empresa operadora do poço;
* Nome do estado da federação no qual o poço se localiza;
* Produção de óleo bruto (m3) por período de tempo;
* Produção de condensado (m3) por período de tempo;
* Produção de gás associado (Nm3) por período de tempo;
* Produção de gás não associado (Nm3) por período de tempo.



*Figura 8 - Dashboard 5: “Produção Operador e Estado”. Elaboração própria.*

# 9 Conclusão

Este projeto permitiu ao grupo compreender o todo o processo de BI, passando pelas fases de procura e consolidação de base de dados; extração, transformação e carga de dados; criação e população de data warehouse e apresentação dos resultados do processo de business intelligence por meio de dashboards. Além disso, foi possível conhecer a produção petrolífera no território brasileiro e suas eventuais peculiaridades.

Pelo projeto de BI apresentado, o cliente Kioleo poderá obter as informações necessárias para seu projeto de expansão da empresa de forma atualizada e de fácil compreensão. Com o uso intuitivo da ferramenta Power BI, as análises necessárias pelas equipes operacionais e táticas foram facilitadas e o processo de tomada de decisão pelo corpo estratégico, facilitado. Consequentemente, a consultoria brasileira garantiu o solicitado pela empresa: credibilidade e fidedignidade dos dados e forte representação visual das informações.

Após manifestação satisfatória da Diretoria Kioleo pelo resultado de BI, foi criada uma parceria contínua de consultoria para analisar eventuais regiões petrolíferas ao redor do globo e aprimorar o processo de análise em terras brasileiras.

# 10 Anexos

## 10.1 Anexo 1

Fontes de dados – Arquivo Compactado “Raw Data.zip” contendo: “rodada\_vencedora.xlsx”, “producao\_poco.xlsx” e “poco\_petrolifero.xlsx”.

## 10.2 Anexo 2

Modelos – Arquivo Compactado “Modelos.zip” contendo: “Modelo DW.architect.xml” e “Modelo Transacional Petroleo.architect.xml”.

## 10.4 Anexo 4

Transformações Pentaho e Database SQL – Arquivo Compactado “Scripts.zip”.

## 10.5 Anexo 5

Dashboards – Arquivo Compactado “Dashboard.zip” contendo: “Dashboards Kioleo.pptx” e “Kioleo r5.pbix”.