

Curso: Ciência de Dados e Analytics (PUC Rio)

Sprint: Engenharia de Dados

Professores:

Profs. Dr. Sérgio Lifschitz,

Profa. Dra. Fernanda Baião,

Prof. Msc. Antony Seabra,

Prof. Dr. Marcos Villas,

Prof. Dr. Victor Almeida, e

Prof. M.Sc. Silvio Alonso

Aluno:

Wellington de Araújo Bastos, M.Sc.

**TÍTULO: Análise da Qualidade de Serviço em
Telecomunicações: Uma Abordagem de Engenharia
de Dados na Nuvem para os Indicadores IQS do
RQUAL**

INTRODUÇÃO

Este trabalho busca atender aos Requisitos para o Desenvolvimento do MVP referente ao Sprint 3 do Módulo Engenharia de Dados do curso de pós graduação da PUC-RIO, com ênfase deformação em Ciência de Dados e Analytics.

O trabalho consiste em construir um pipeline de dados orientado as fases de busca, modelagem, carga e análise de dados.

1. Objetivo Principal

Construir um pipeline de análise de dados para os indicadores de qualidade, RQUAL da Anatel, definidos para os serviços de Telecomunicações, sobretudo para o serviço de telefonia móvel ou Serviço Móvel Pessoal (SMP).

Os indicadores e dados utilizados nesse trabalho podem ser encontrados no portal da Agência Nacional de Telecomunicações - ANATEL no link Anatel — [Anatel — Agência Nacional de Telecomunicações](#), e mais especificamente no link Anatel - [Indicadores - Telefonia Móvel](#).

A definição dos indicadores podem ser encontradas em no documento Manual Operacional (MOP) e o Documento de Valores de Referência (DVR) disponibilizado através do Sobre o Sobre o [RQUAL — Agência Nacional de Telecomunicações](#), acessando o documento [MOP Versão 2.1](#).

Nesse trabalho focaremos nos Indicadores de Qualidade de Serviço - IQS, resumido no infografo com os principais conceitos/aspectos do RQUAL.

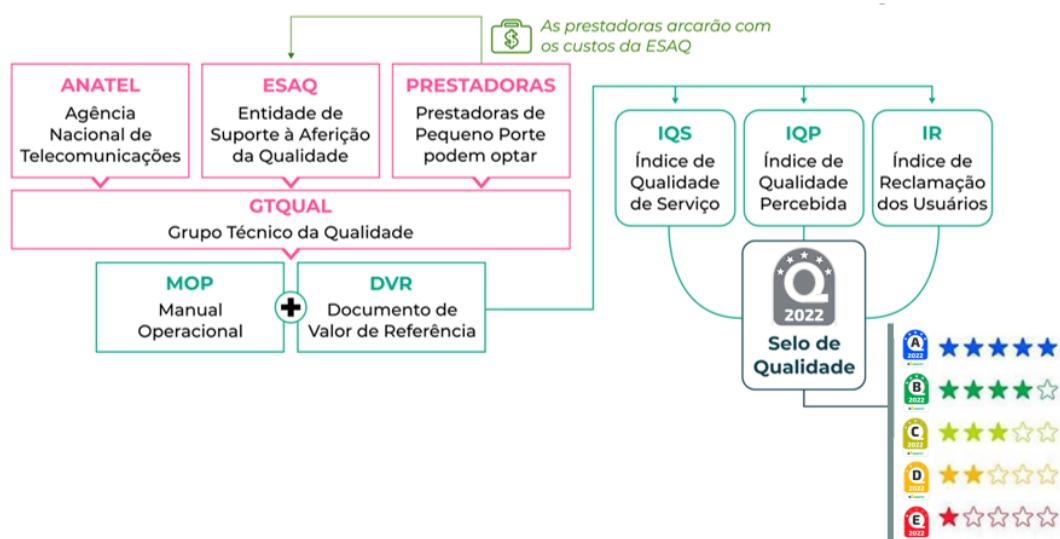


Figura 1 – Infografo com os principais conceitos do RQUAL - Anatel

Construiremos um pipeline que reproduza as consultas dos usuários do site para os principais indicadores da telefonia móvel - IND1 até IND8, cujos significadores passaremos

descrever:

Acrônimo	Indicador	Serviço	(Anexo Nº)	Tipo	Grupo
IND1	Conexão de Chamadas na Rede de Acesso	SMP	01	Indicador do IQS	Redes
IND2	Queda de Chamadas	SMP	01	Indicador do IQS	Redes
IND3	Conexão de Dados Medida na Rede de Acesso	SMP	01	Indicador do IQS	Redes
IND4	Cumprimento da Velocidade de download e upload	SMP e SCM	05 (SMP) e 06 (SCM)	Indicador do IQS	Redes
IND5	Latência bidirecional da Conexão de Dados	SMP e SCM	05 (SMP) e 06 (SCM)	Indicador do IQS	Redes
IND6	Variação de Latência da Conexão de Dados	SMP e SCM	05 (SMP) e 06 (SCM)	Indicador do IQS	Redes
IND7	Perda de Pacotes da Conexão de Dados	SMP e SCM	05 (SMP) e 06 (SCM)	Indicador do IQS	Redes
IND8	Disponibilidade	SMP, STFC, SCM e TV assinatura	07	Indicador do IQS	Redes

Quadro 1 - Indicadores e suas descrições

Uma amostra de uma consulta para os indicadores das principais operadoras de serviço da cidade do Rio de Janeiro-RJ pode ser visto na Figura 2.

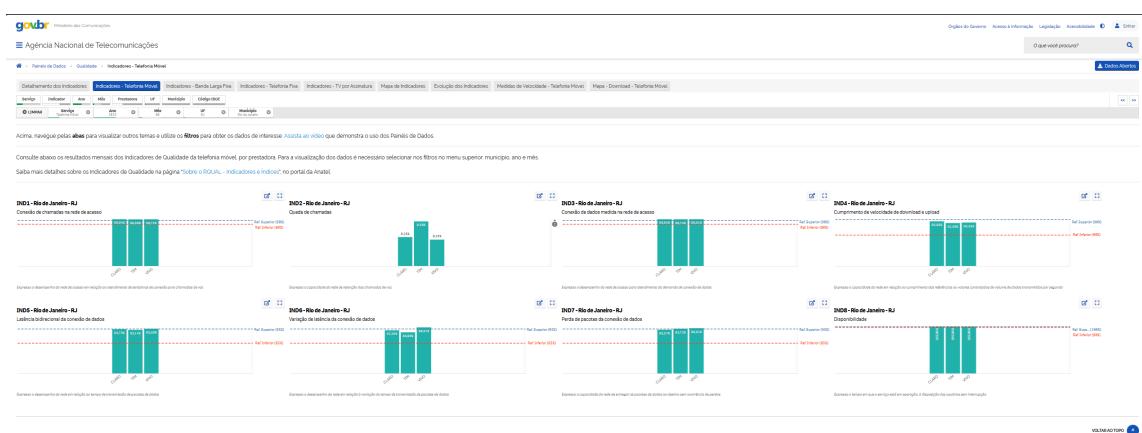


Figura 2 - Figura 2 – Consulta aos indicadores IQS da cidade do Rio de Janeiro-RJ

Questões a serem respondidas:

1. É possível reproduzir as consultas dos indicadores IQS do RQUAL, individualmente ou grupos, através de um pipeline de dados na nuvem selecionando atributos como ano, mês, indicador, cidade, estado e prestadora?, ou seja reproduzir a mesma consulta feita pelo site Anatel?
2. É possível incrementar a essas consultas filtros que considerem o tamanho da população das cidades, quantidade de assinantes por operadora e renda percapita ou PIB?
3. Fazer uma análise como:

- a. melhores e piores IQS – por município
- b. melhores e piores IQS – por prestadora
- c. melhores e piores IQS – por população
- d. melhores e piores IQS – por renda

Restrições

- Para efeito desse trabalho será considerado apenas o período de 01/2023 a 06/2023 e que atendam a tecnologia 4G.
- Serão considerados apenas um estado por região do Brasil: Paraná, São Paulo, Pernambuco, Goiás, Belém e Distrito Federal

2. Base de Dados

2.1 Dados Principais

Os dados do IQS – RQUAL podem ser obtidos no link Dados Abertos do site da Anatel na opção Home>Paineis de Dados>Qualidade na opção Indicadores – Telefonia Móvel [Anatel - Indicadores - Telefonia Móvel](#). Nessa página, através dos seletores de filtro o usuário pode realizar as opções de filtros para os indicadores IND1 a IND8, podendo optar por baixar os dados de cada indicador filtrado, ou realizar o download da base completa na seleção [Dados Abertos > Indicadores de Qualidade](#).

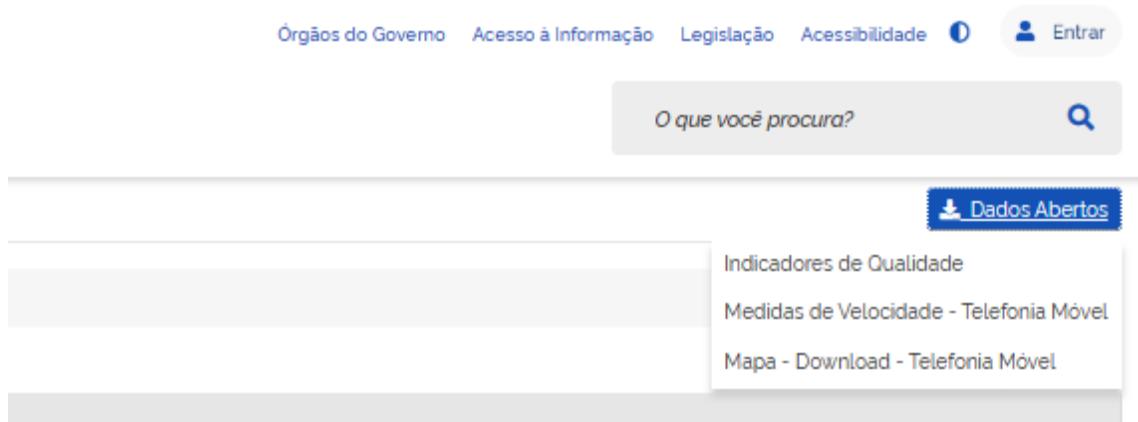


Figura 3 - Dados Abertos para Download

2.2 Dados Complementares

Para podermos atender as questões dois e três do trabalho proposto serão coletados dados complementares no próprio site da Anatel que agregam informações de Dados de Cobertura do serviço por operadora, e dados de velocidade de navegação de dados por renda.

Os dados podem ser coletados nos links abaixo:

1. [Velocidade x Renda](#)
2. [Dados de Cobertura do Município](#)
3. [Dados de Cobertura por Operadora](#)

Restrições

- Todos os dados utilizados nesse trabalho são abertos e público através da Agencia Nacional de telecomunicações

3. Modelagem

Tabela 1: Dados do IQS - Telefonia Móvel

Nome do Campo	Tipo	Descrição
Ano	INTEGER	Ano da coleta
Mês	INTEGER	Mês da coleta
Serviço	STRING	Serviço Móvel Pessoal
Indicador	STRING	Acrônimo do KPI
Descrição	STRING	Descrição do KPI
Tipo	STRING	Indicador IQS
Grupo	INTEGER	Rede
Código IBGE	INTEGER	Código do Município no IBGE
Município	STRING	Nome do Município
UF	STRING	Unidade da Federação
Prestadora	STRING	Nome da Operadora SMP
Resultado	FLOAT	Valor calculado para o Indicador IQS
Unidade de Medida	STRING	Unidade de medida (%)
Número de Coletores	INTEGER	Quantidade de terminais que representam a amostra de coleta
Número de Medidas	FLOAT	Quantidade de medições coletadas
Erro Amostral	FLOAT	Erro de amostragem
Limite Inferior	FLOAT	
Limite Superior	FLOAT	
Data	STRING	Data de coleta

Tabela 2: Dimensão e Cobertura

Nome do Campo	Tipo	Descrição
Código IBGE Município (Cobertura)	INTEGER	Código do Município no IBGE
Município (Cobertura)	STRING	Nome do Município
UF (Cobertura)	STRING	Unidade da Federação
Operadora	STRING	Nome da Operadora SMP
Tecnologia	STRING	Tecnologia usada para a cobertura
% área coberta	FLOAT	Percentual de cobertura alcançada
% moradores cobertos	FLOAT	% de moradores atendidos pela cobertura
% domicílios cobertos	FLOAT	% de domicílios alcançados pela cobertura
Área km2	FLOAT	Área em Km2 do município
Moradores	STRING	Quantidade de moradores do município
Domicílios	STRING	Quantidade de domicílios do município
Tipo de compromisso	STRING	Tipo de compromisso de cobertura assumindo em leilão

Nome do Campo	Tipo	Descrição
Compromisso	STRING	Descrição do compromisso
Origem compromisso	STRING	Indicação do leilão
Operadora com compromisso	STRING	Nome da Operadora com compromisso
Status compromisso	STRING	Status atual do compromisso
Prazo compromisso	STRING	Prazo final para atender o compromisso

Tabela 3: Velocidade

Nome do Campo	Tipo	Descrição
Código Setor	STRING	Código do setor aferido
Bairro Setor	STRING	Bairro do setor aferido
Tipo Setor	STRING	tipo de setor
Velocidade Mediana	FLOAT	Velocidade de up/downlink aferido
Renda Média	FLOAT	Renda média do setor de aferição
Categoria	FLOAT	
Código Município	INTEGER	Código do Município no IBGE
Município (Velocidade)	STRING	Nome do Município
UF (Velocidade)	STRING	Unidade da Federação

Tabela 4: Acessos

Nome do Campo	Tipo	Descrição
Ano	INTEGER	Ano da aferição
Mês	INTEGER	Mês da aferição
Grupo Econômico	STRING	Nome do Grupo Econômico
Empresa	STRING	Nome da Operadora
CNPJ	STRING	CNPJ da Operadora
Porte da Prestadora	STRING	Indica o porte da operadora em relação a detenção de mercado
UF	STRING	Unidade da Federação
Município	STRING	Nome do Município
Código IBGE Município	INTEGER	Código do Município no IBGE
Código Nacional	STRING	Código DDD
Código Nacional (Chip)	STRING	Código identificador do chip para SMP
Modalidade de Cobrança	STRING	Tipo de cobrança do serviço Pós/Pré pago
Tecnologia	STRING	Tecnologia usada pela rede de Acesso
Tecnologia Geração	STRING	Geração da Tecnologia
Tipo de Pessoa	STRING	Tipo de pessoa detentora do acesso Jurídica ou Física
Tipo de Produto	STRING	

Nome do Campo	Tipo	Quantidade de Acessos	Descrição
Acessos	STRING	Quantidade de Acessos	Descrição

Para a modelagem do banco de dados e na aplicação proposta para este trabalho, não é necessário utilizar todos os atributos contidos nas tabelas de dados coletadas nos datasets da Anatel. Portanto, na modelagem do esquema de relacionamento estrela mostrado na Figura 4, faremos uso apenas dos campos necessários. Contudo, a ideia desse trabalho é implementar uma tabela flat (desnormalizada) que contenha todos os dados em uma única tabela com o objetivo de ganhar eficiência de consultas, facilitar o entendimento e velocidade na ingestão dos dados no Google BigQuery, que possui algumas considerações especiais quando se trata de operações de junção (joins) de tabelas, especialmente em comparação com bancos de dados relacionais tradicionais. Isso se deve ao fato de que o BigQuery é um banco de dados distribuído e colunar projetado para lidar com grandes volumes de dados e consultas em paralelo.

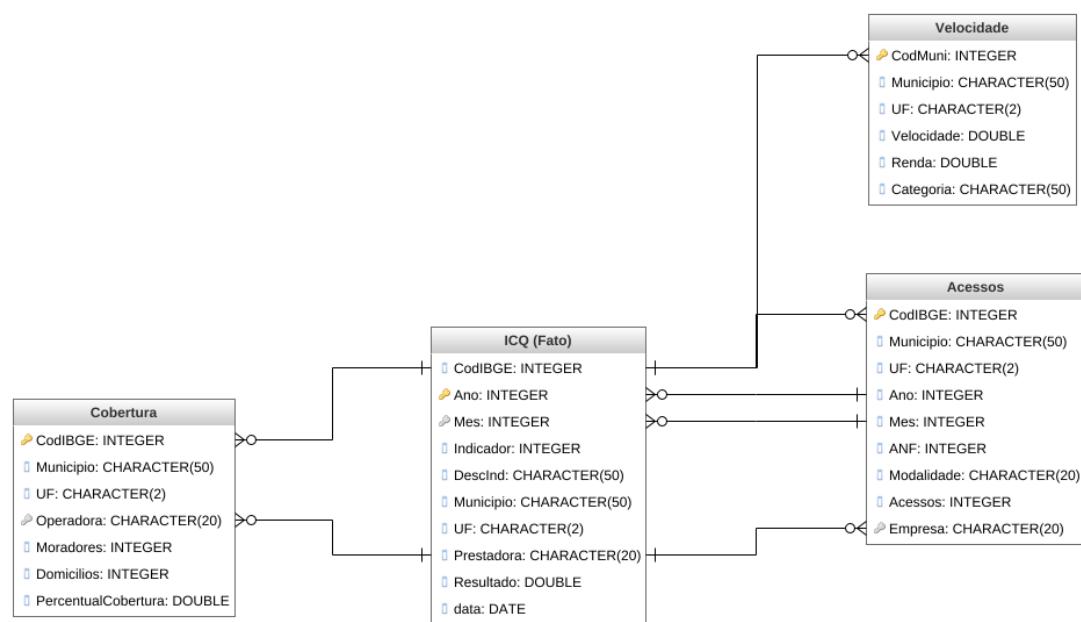


Figura 4 - Modelagem de um DW Estrela para o RQUAL - IQS da Anatel

A Tabela Flat final que será inserida no Cloud BigQuery terá um esquema como indicado abaixo:

Nome do campo	Tipo	Modo	Descrição
Ano	INTEGER	NULLABLE	Ano da medição e cálculo do KPI
Mes	INTEGER	NULLABLE	Mês de medição e cálculo do KPI
Indicador	STRING	NULLABLE	Identificação do Indicador IQS
Descricao	STRING	NULLABLE	Descrição do Indicador IQS
Codigo_IBGE_R	INTEGER	NULLABLE	Código IBGE da Cidade
Municipio	STRING	NULLABLE	Nome da cidade (normalizado)
UF	STRING	NULLABLE	Sigla da Unidade da Federação (Estado)
Prestadora	STRING	NULLABLE	Nome da Prestadora do Serviço Móvel Pessoal (SMP)
Resultado	FLOAT	NULLABLE	Resultado do indicador IQS (Unidade %)

Nome do campo	Tipo	Modo	Descrição
data	STRING	NULLABLE	Data da Medição
ANF	STRING	NULLABLE	Código de Identificação da área de numeração nacional (ANF)
Plano	STRING	NULLABLE	Tipo de Plano Comercial
Acessos_sum	INTEGER	NULLABLE	Número de Acessos ou Terminais
Moradores_avg	FLOAT	NULLABLE	Número médio de Moradores da cidade
Domicilios_avg	FLOAT	NULLABLE	Número médio de Domicílios da Cidade
Perc_moradores_cobertos	FLOAT	NULLABLE	Percentual de moradores com cobertura de serviço móvel
Velocidade_mediana	FLOAT	NULLABLE	Velocidade média de throughput
Renda_Media	FLOAT	NULLABLE	Renda Média da por Usuário

4. Pipeline de Dados

A Figura 5 mostra todos os passos a serem utilizados para criar um Pipeline de dados desde a coleta dos dados até sua ingestão no Banco de Dados. Aqui definimos quatro passos sequenciais que utilizam recursos da Google Cloud:

1. Google Drive,
2. GCS (Google Cloud Storage),
3. Data Fusion, e
4. BigQuery.

Nas seções que seguem, mostraremos como cada passo foi executado, com evidências em forma de capturas de tela para cada etapa do processo.



Figura 5 - Pipeline de dados para tratamento do RQUAL-IQS da Anatel

Passo 1: Configurando o Google Drive

Neste primeiro passo, no Console da Google Cloud, criaremos uma pasta no Google Drive onde os dados a serem processados no pipeline serão armazenados.

Passo 2: Google Cloud Storage (GCS)

No segundo passo, acessaremos o Google Cloud Storage (GCS) para criar um novo bucket de armazenamento para os dados que serão processados pelo pipeline.

Passo 3: Google Cloud Data Fusion

Neste terceiro passo, utilizaremos o Google Cloud Data Fusion para criar um ambiente de gerenciamento do pipeline e configurar a conexão com o GCS. Aqui, definiremos o pipeline para ler, transformar, processar e realizar a ingestão dos dados no banco de dados.

Passo 4: Google BigQuery

No último passo, será criado um conjunto de dados (dataset) onde os dados processados serão armazenados. Além disso, configuraremos o BigQuery para carregar automaticamente

os dados do Data Fusion na tabela correspondente do BigQuery.

Esses passos fornecem uma visão geral do processo para criar um pipeline de dados usando a Google Cloud. Detalhes adicionais e configurações específicas podem ser necessários dependendo dos requisitos e ferramentas exatas utilizadas em seu projeto.

4.1 Passo 1 - Coleta de Dados

Os dados referentes aos indicadores Rqual, Cobertura, Renda x Velocidade e Acessos foram coletados a partir dos sites da Anatel, conforme mencionado na seção de Objetivo Principal deste documento. Os links para acesso aos dados são os seguintes:

1. Rqual - [Link para Download](#);
2. Renda x Velocidade - [Link para Download](#);
3. Cobertura - [Link para Download](#);
4. Acessos - [Link para Download](#)

Conforme estabelecido no Objetivo Principal, serão considerados apenas dados de um estado por região do Brasil, incluindo Paraná, São Paulo, Pernambuco, Goiás, Belém e Distrito Federal. Para reduzir o tamanho dos arquivos de dados, realizamos um filtro nesses arquivos utilizando a ferramenta Knime. É importante ressaltar que esse processo de filtragem não afeta as transformações de dados efetivas que serão realizadas nas ferramentas ETL do Google Cloud.

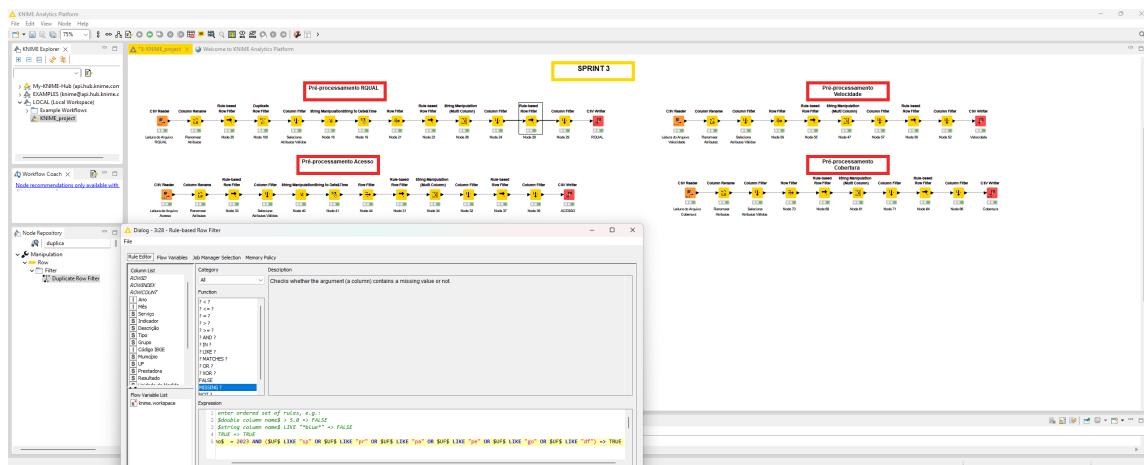


Figura 6 - Fluxo no Knime para filtrar os datasets relacionados aos estados do PA, PE, GO, DF, SP e PR.

Na sequencia os arquivos foram salvos no Google Drive para utilização pelo Pipeline que fará o ETL do dataset final.

Figura 7 - Dataset no Google Drive

Links para os arquivos filtrados:

1. *RQUAL_Movel_Reg2023.csv* - [Link para Download](#)
2. *Velocidade_Movel_Reg2023.csv* - [Link para Download](#)
3. *Cobertura_Movel_Reg2023.csv* - [Link para Download](#)
4. *Acesso_Movel_Reg2023.csv* - [Link para Download](#)

Na sequência os arquivos foram transferidos para a área de armazenamento temporário do Cloud Shell do projeto Sprint3-PUC-DataAnalytics da conta *w**@gmail.com*.

Tipo	Principal	Nome	Papel	Insights de segurança	Herança
<input type="checkbox"/>	100179499286-compute@developer.gserviceaccount.com	Compute Engine default service account	Editor		
<input type="checkbox"/>	wabastos@gmail.com	Wellington Bastos	Proprietário		

Figura 8 - Evidência da Conta Google Cloud criada para o e-mail *w**@gmail.com*.

4.1.1 Carga do dataset para o Google Cloud Shell

Os arquivos foram coletados utilizando a ferramenta *gdown*, que é uma linha de comando que facilita o download de arquivos do Google Drive diretamente para o ambiente do Cloud

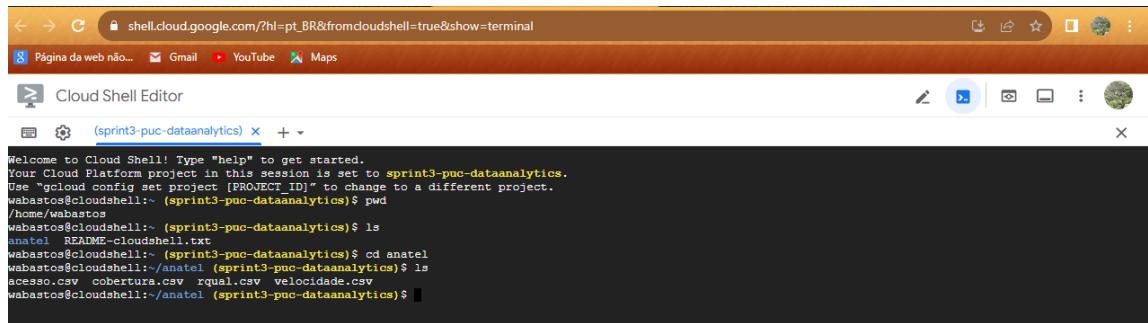
Shell. Esta ferramenta simplifica o processo de obtenção dos arquivos necessários para a análise de dados.

Sintaxe:

gdown <https://drive.google.com/uc?id=ID-DE-ARQUIVO> /DESTINO

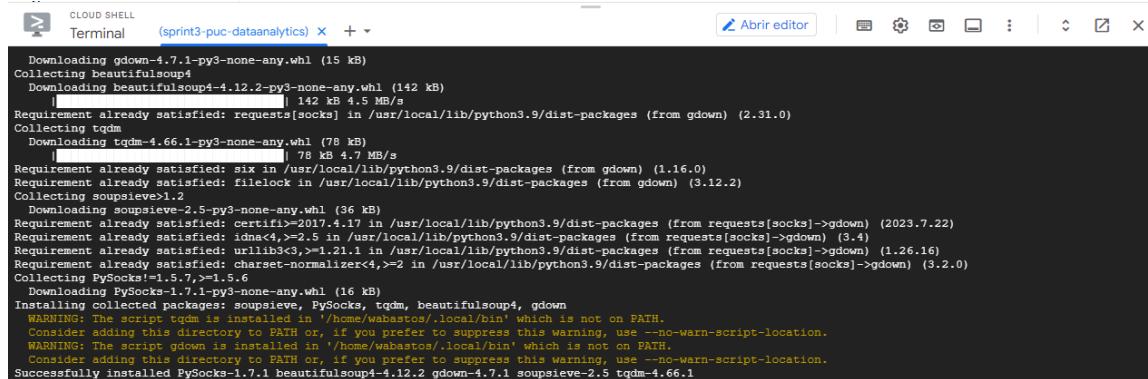
Id dos arquivos:

1. 1WzQ-qlPEvr-epqQJrKmZID7E7eLLcSkJ
2. 1L0qJlHnluOsQnx7JR4OT0sejJAguIrE1
3. 106VX3PpyYioXMxRxsKlzTzs89xxkmAJ7
4. 19NZskmHMUKmlY9rvkzWIQPp2kziE6H6I



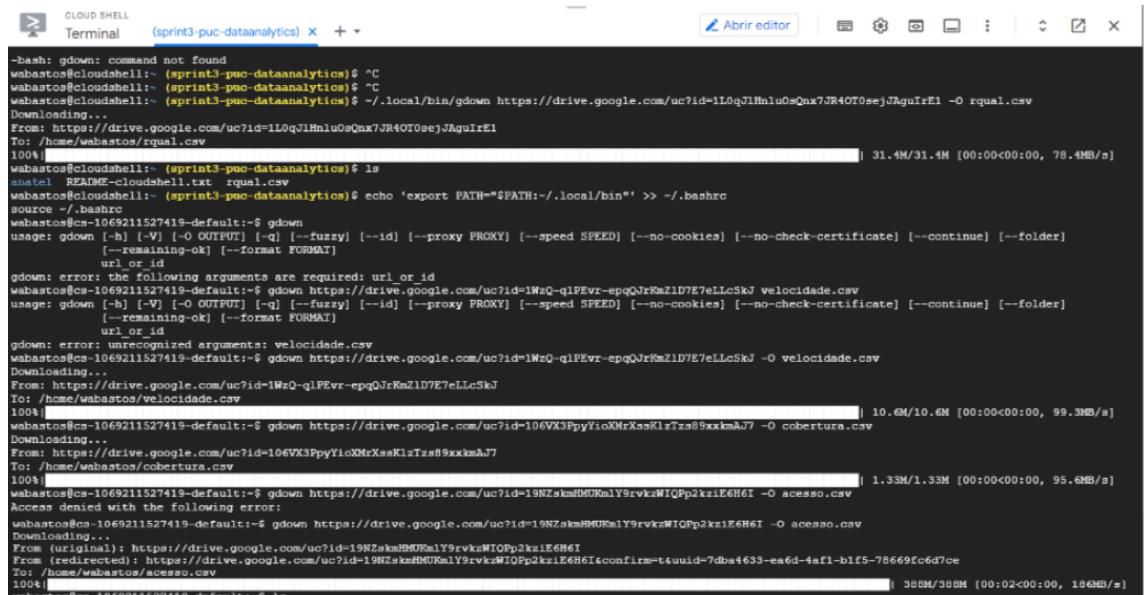
```
shell.cloud.google.com/?hl=pt_BR&fromcloudshell=true&show=terminal
Página da web não... Gmail YouTube Maps
Cloud Shell Editor
(sprint3-puc-dataanalytics) x + v
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to sprint3-puc-dataanalytics.
Use "gcloud config set project [PROJECT_ID]" to change to a different project.
wabastos@cloudshell:~ (sprint3-puc-dataanalytics)$ pwd
/home/wabastos
wabastos@cloudshell:~ (sprint3-puc-dataanalytics)$ ls
anatel README-cloudshell.txt
wabastos@cloudshell:~ (sprint3-puc-dataanalytics)$ cd anatel
wabastos@cloudshell:~/anatel (sprint3-puc-dataanalytics)$ ls
acesso.csv cobertura.csv rqual.csv velocidade.csv
wabastos@cloudshell:~/anatel (sprint3-puc-dataanalytics)$
```

Figura 9 - Prompt do Cloud Shell e criação do diretório destino (/Anatel)



```
CLOUD SHELL Terminal (sprint3-puc-dataanalytics) x + v
Abrir editor
Collecting gdown-4.7.1-py3-none-any.whl (15 kB)
  Downloading beautifulsoup4-4.12.2-py3-none-any.whl (142 kB)
|██████████| 142 kB 4.5 MB/s
Requirement already satisfied: requests[socks] in /usr/local/lib/python3.9/dist-packages (from gdown) (2.31.0)
Collecting tqdm
  Downloading tqdm-46.1.1-py3-none-any.whl (78 kB)
|██████████| 78 kB 4.7 MB/s
Requirement already satisfied: six in /usr/local/lib/python3.9/dist-packages (from gdown) (1.16.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.9/dist-packages (from gdown) (3.12.2)
Collecting soupsieve<2.5
  Downloading soupsieve-2.5-py3-none-any.whl (36 kB)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests[socks]>->gdown) (2023.7.22)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests[socks]>->gdown) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from requests[socks]>->gdown) (1.26.16)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.9/dist-packages (from requests[socks]>->gdown) (3.2.0)
Collecting PySocks!=1.5.7,>=1.5.6
  Downloading PySocks-1.7.1-py3-none-any.whl (16 kB)
Installing collected packages: soupsieve, PySocks, tqdm, beautifulsoup4, gdown
  WARNING: The script tqdm is installed in '/home/wabastos/.local/bin' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
  WARNING: The script gdown is installed in '/home/wabastos/.local/bin' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
Successfully installed PySocks-1.7.1 beautifulsoup4-4.12.2 soupsieve-2.5 tqdm-4.66.1
```

Figura 10 - Instalação do gdown

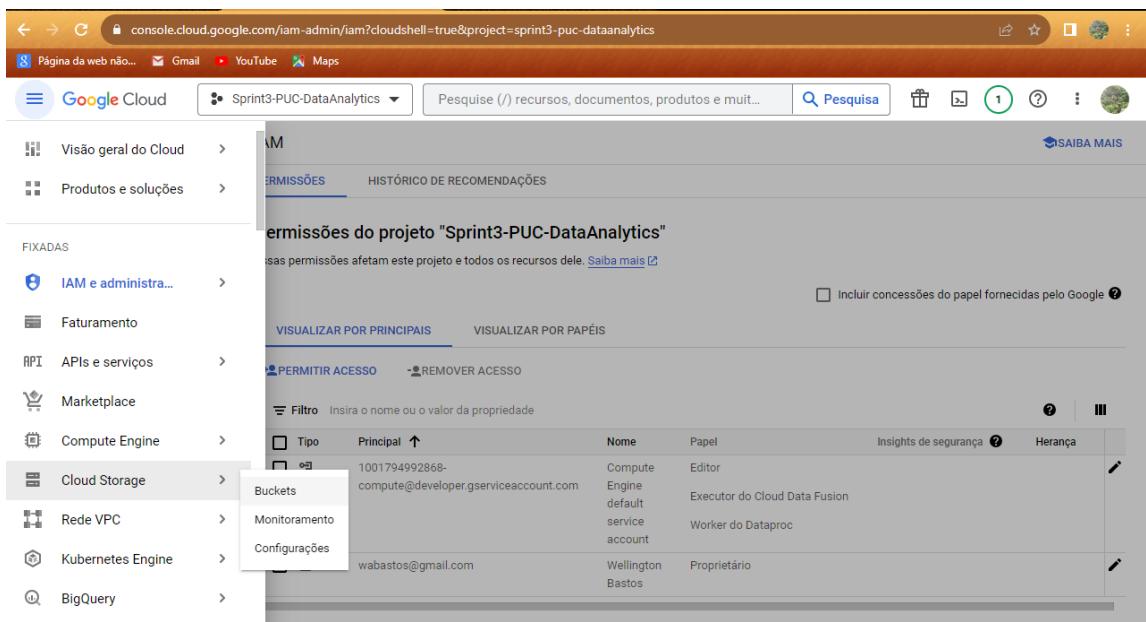


```
CLOUD SHELL Terminal (sprint3-puc-dataanalytics) x + v
Abrir editor
wabastos@cloudshell:~ (sprint3-puc-dataanalytics)$ gdown https://drive.google.com/uc?id=1L0qJlHnluOsQnx7JR4OT0sejJAguIrE1 -O rqual.csv
From: https://drive.google.com/uc?id=1L0qJlHnluOsQnx7JR4OT0sejJAguIrE1 To: /home/wabastos/rqual.csv
100% [██████████] 31.4M/31.4M [00:00<00:00, 78.4MB/s]
wabastos@cloudshell:~ (sprint3-puc-dataanalytics)$ ls
anatel README-cloudshell.txt rqual.csv
wabastos@cloudshell:~ (sprint3-puc-dataanalytics)$ echo 'export PATH="$PATH:./.local/bin"' >> ~/.bashrc
source ~/.bashrc
wabastos@cs-1069211527419-default:~$ gdown
usage: gdown [-h] [-V] [-O OUTPUT] [-q] [--fuzzy] [--id] [-proxy PROXY] [--speed SPEED] [--no-cookies] [--no-check-certificate] [--continue] [--folder]
             [-remaining-ok] [--format FORMAT]
             url_or_id
gdown: error: the following arguments are required: url_or_id
wabastos@cs-1069211527419-default:~$ gdown https://drive.google.com/uc?id=1WzQ-qlPEvr-epqQJrKmZID7E7eLLcSkJ velocidade.csv
usage: gdown [-h] [-V] [-O OUTPUT] [-q] [--fuzzy] [--id] [-proxy PROXY] [--speed SPEED] [--no-cookies] [--no-check-certificate] [--continue] [--folder]
             [-remaining-ok] [--format FORMAT]
             url_or_id
gdown: error: unrecognized arguments: velocidade.csv
wabastos@cs-1069211527419-default:~$ gdown https://drive.google.com/uc?id=1WzQ-qlPEvr-epqQJrKmZID7E7eLLcSkJ -O velocidade.csv
Downloading...
From: https://drive.google.com/uc?id=1WzQ-qlPEvr-epqQJrKmZID7E7eLLcSkJ To: /home/wabastos/velocidade.csv
100% [██████████] 10.6M/10.6M [00:00<00:00, 99.3MB/s]
wabastos@cs-1069211527419-default:~$ gdown https://drive.google.com/uc?id=106VX3PpyYioXMxRxsKlzTzs89xxkmAJ7 -O cobertura.csv
Downloading...
From: https://drive.google.com/uc?id=106VX3PpyYioXMxRxsKlzTzs89xxkmAJ7 To: /home/wabastos/cobertura.csv
100% [██████████] 1.33M/1.33M [00:00<00:00, 95.6MB/s]
wabastos@cs-1069211527419-default:~$ gdown https://drive.google.com/uc?id=19NZskmHMUKmlY9rvkzWIQPp2kziE6H6I -O acesso.csv
Access denied with the following error:
wabastos@cs-1069211527419-default:~$ gdown https://drive.google.com/uc?id=19NZskmHMUKmlY9rvkzWIQPp2kziE6H6I -O acesso.csv
Download failed: https://drive.google.com/uc?id=19NZskmHMUKmlY9rvkzWIQPp2kziE6H6I
From (redirected): https://drive.google.com/uc?id=19NZskmHMUKmlY9rvkzWIQPp2kziE6H6I&confirm=t&uuid=7dba4633-ea6d-4af1-b1f5-78669fc6d7ce
To: /home/wabastos/acesso.csv
100% [██████████] 38GM/38GM [00:02<00:00, 186MB/s]
```

Figura 11 - Coletando os arquivos do Google Drive para o Google Shell

4.2 Passo 2 - Criando Buckets no Cloud Stores

Essa etapa demonstra a criação de uma área para armazenamento no Cloud Store para o acesso dos dados pela ferramenta de ETL. Na sequência serão demonstradas a criação do Bucket, a transferência dos arquivos csv (*RQUAL_Movel_Reg2023.csv*, *Velocidade_Movel_Reg2023.csv*, *Cobertura_Movel_Reg2023.csv*, *Acesso_Movel_Reg2023.csv*) e suas evidências.

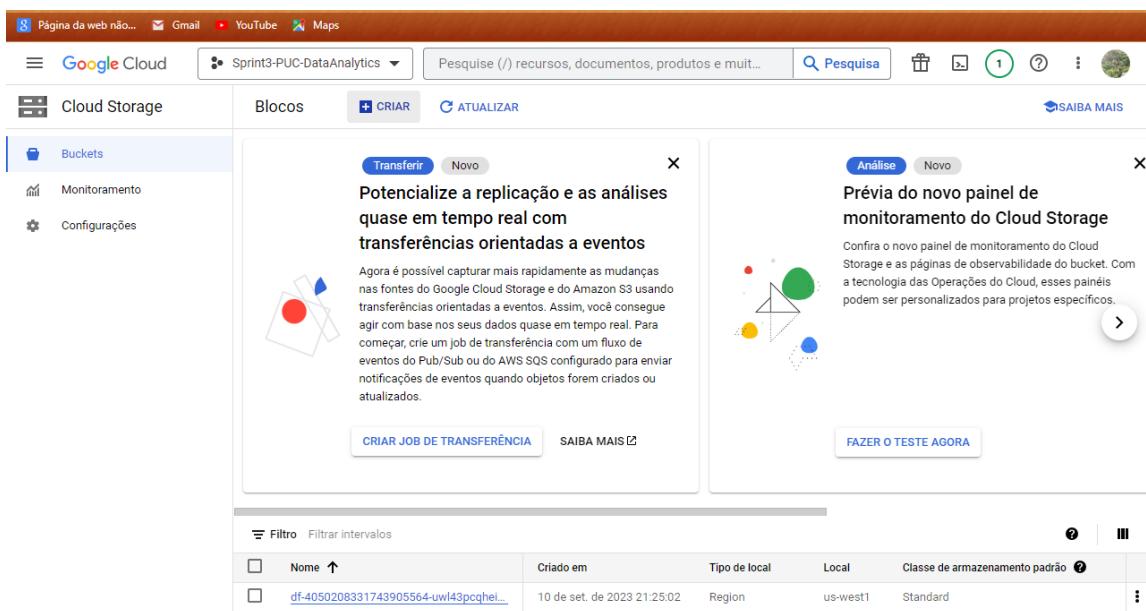


The screenshot shows the Google Cloud IAM & Administração interface. The main title is "PERMISSÕES" (Permissions) for the project "Sprint3-PUC-DataAnalytics". It displays a table of permissions for various users and service accounts. On the left sidebar, under "FIXADAS" (Fixed), the "Cloud Storage" section is expanded, showing "Buckets", "Monitoramento" (Monitoring), and "Configurações" (Configurations). A dropdown menu is open over the "Buckets" item.

Figura 12 - Acessando o Cloud Storege

Criando um Bucket

Na sequência abaixo mostramos como criar um Bucket e as evidências do Bucket criado para receber os arquivos de dataset referente aos RQUAL.



The screenshot shows the Google Cloud Cloud Storage interface. The main title is "Buckets". It displays a table of existing buckets. A modal window is open for creating a new bucket, with the name "df-4050208331743905564-uwI43pcqhei..." and a "Criar Job de Transferência" (Create Transfer Job) button. Another modal window titled "Prévia do novo painel de monitoramento do Cloud Storage" (Preview of the new monitoring dashboard for Cloud Storage) is also visible.

Figura13-2.png Figura13-3.png Figura13-4.png Figura13-5.png

Figura 13 - Criando o novo Bucket: sprint3-bucket-anatel

Bucket criado

sprint3-bucket-anatel

Local	Classe de armazenamento	Acesso público	Proteção
us-central1 (Iowa)	Standard	Não público	Nenhum

OBJETOS CONFIGURAÇÃO PERMISSÕES PROTEÇÃO CICLO DE VIDA OBSERVABILIDADE RELATÓRIOS DE INVENTÁRIO

Intervalos > sprint3-bucket-anatel

FAZER UPLOAD DE ARQUIVOS CARREGAR PASTA CRIAR PASTA TRANSFERIR DADOS GERENCIAR RETENÇÕES FAZER O DOWNLOAD EXCLUIR

Filtrar apenas pelo prefixo do nome ▾ Filtro Filtrar objetos e pastas Mostrar dados excluídos

Nome	Tamanho	Tipo	Criado	Classe de armazenamento	Última modificação	Acesso público
arquivos_in/	-	Pasta	-	-	-	-

Figura 14 - Evidência dos Bucketes criados

Transferindo os arquivos do Cloud Shell para o Cloud Storage

O comando `gsutil cp * gs://sprint3-bucket-anatel/arquivos_in` foi utilizado para realizar a transferência dos arquivos do Cloud Shell para o Cloud Storage. Abaixo as evidências dos comandos e dos arquivos devidamente coletados para o bucket `sprint3-bucket-anatel/arquivos_in`.

```
wabastos@cloudshell:~/anatel (sprint3-puc-dataanalytics)$ gsutil cp * gs://sprint3-bucket-anatel/arquivos_in
Copying file:///acesso.csv [Content-Type=text/csv]...
==> NOTE: You are uploading one or more large file(s), which would run significantly faster if you enable parallel composite uploads. This feature can be enabled by setting the "parallel_composite_upload_threshold" value in your .boto configuration file. However, note that if you do this large files will be uploaded as 'composite objects' <https://cloud.google.com/storage/docs/composite-objects>, which means that any user who downloads such objects will need to have a compiled crcmod installed (see "gsutil help crcmod"). This is because without a compiled crcmod, computing checksums on composite objects is so slow that gsutil disables downloads of composite objects.

Copying file:///cobertura.csv [Content-Type=text/csv]...
Copying file:///rqual.csv [Content-Type=text/csv]...
Copying file:///velocidade.csv [Content-Type=text/csv]...
\ [4 files] 411.2 MiB/411.2 MiB
Operation completed over 4 objects/411.2 MiB.
wabastos@cloudshell:~/anatel (sprint3-puc-dataanalytics) $
```

Detalhes do bucket

sprint3-bucket-anatel

Local	Classe de armazenamento	Acesso público	Proteção
us-central1 (Iowa)	Standard	Não público	Nenhum

OBJETOS CONFIGURAÇÃO PERMISSÕES PROTEÇÃO CICLO DE VIDA OBSERVABILIDADE RELATÓRIOS DE INVENTÁRIO

Intervalos > sprint3-bucket-anatel > arquivos_in

FAZER UPLOAD DE ARQUIVOS CARREGAR PASTA CRIAR PASTA TRANSFERIR DADOS GERENCIAR RETENÇÕES FAZER O DOWNLOAD EXCLUIR

Filtrar apenas pelo prefixo do nome ▾ Filtro Filtrar objetos e pastas Mostrar dados excluídos

Nome	Tamanho	Tipo	Criado	Classe de armazenamento	Última modificação
acesso.csv	369,9 MB	text/csv	11 de set. de 2023 16:06:44	Standard	11 de set. de 2023 16:06:44
cobertura.csv	1,3 MB	text/csv	11 de set. de 2023 16:06:44	Standard	11 de set. de 2023 16:06:44
rqual.csv	30 MB	text/csv	11 de set. de 2023 16:06:45	Standard	11 de set. de 2023 16:06:45
velocidade.csv	10,1 MB	text/csv	11 de set. de 2023 16:06:46	Standard	11 de set. de 2023 16:06:46

Figura 15 - Dataset transferido para o Bucket-Anatel

4.3 Análise dos dados

Durante a etapa de Modelagem, detalhada anteriormente neste documento, foram definidos atributos para cada arquivo no dataset, identificados como Tabela 1 a Tabela 4. Durante essa etapa, avaliamos que apenas alguns atributos eram necessários para realizar a análise dos indicadores RQUAL-IQS proposta nas questões do objetivo principal, os quais foram consolidados na Tabela Flat.

Abaixo estão os atributos e seus tipos:

Nome do campo	Tipo	Modo	Descrição
Ano	INTEGER	NULLABLE	Ano da medição e cálculo do KPI
Mes	INTEGER	NULLABLE	Mês de medição e cálculo do KPI
Indicador	STRING	NULLABLE	Identificação do Indicador IQS
Descricao	STRING	NULLABLE	Descrição do Indicador IQS
Codigo_IBGE_R	INTEGER	NULLABLE	Código IBGE da Cidade
Municipio	STRING	NULLABLE	Nome da cidade (normalizado)
UF	STRING	NULLABLE	Sigla da Unidade da Federação (Estado)
Prestadora	STRING	NULLABLE	Nome da Prestadora do Serviço Móvel Pessoal (SMP)
Resultado	FLOAT	NULLABLE	Resultado do indicador IQS (Unidade %)
data	STRING	NULLABLE	Data da Medição
ANF	STRING	NULLABLE	Código de Identificação da área de numeração nacional (ANF)
Plano	STRING	NULLABLE	Tipo de Plano Comercial
Acessos_sum	INTEGER	NULLABLE	Número de Acessos ou Terminais
Moradores_avg	FLOAT	NULLABLE	Número médio de Moradores da cidade
Domicilios_avg	FLOAT	NULLABLE	Número médio de Domicílios da Cidade
Perc_moradores_cobertos	FLOAT	NULLABLE	Percentual de moradores com cobertura de serviço móvel
Velocidade_mediana	FLOAT	NULLABLE	Velocidade média de throughput
Renda_Media	FLOAT	NULLABLE	Renda Média da por Usuário

Transformação de Dados

Identificamos que alguns atributos com representatividade numérica estavam definidos como tipo string nas tabelas de origem, exigindo a correta conversão para tipos numéricos durante o processo de ETL. Como exemplo, temos o atributo Número de Acessos na Tabela Acessos, originalmente definido como string, mas transformado em integer durante o ETL.

Agrupamento de Dados

Outro aspecto relevante a ser considerado é que a simples exclusão de atributos pode não ser a transformação adequada, já que isso pode afetar o significado ou a representatividade

de outros atributos. Por exemplo, a quantidade de Acessos em um município está subdividida em Pré-pago e Pós-pago com base no atributo Modalidade de Cobrança. Assim, a eliminação direta desse atributo pode resultar na perda de informações cruciais, a menos que esses dados sejam agrupados de alguma forma no contexto municipal. Portanto, nas tabelas Velocidade, Acessos e Cobertura, optamos por manter alguns atributos por meio de agrupamentos (group by), quando relevante, para preservar informações essenciais.

Dados Ausentes, Nulos e Caso

O tratamento de dados ausentes, nulos, letras maiúsculas e diacríticos foi realizado durante o pré-processamento no Knime, antes da importação dos dados para o ambiente em nuvem. De fato, a presença de dados ausentes e nulos foi observada apenas no arquivo RQUAL, mais especificamente nos atributos Limites Inferiores, Limites Superiores e Número de Coletas. No entanto, os registros foram mantidos caso o atributo correspondente, Resultado, não fosse nulo. Além disso, as conversões para letras minúsculas foram aplicadas aos atributos do tipo string, assim como a remoção de diacríticos.

Em resumo, os dados estavam, em sua maioria, bem comportados e normalizados, não sendo necessária a aplicação de técnicas adicionais além das mencionadas acima.

4.4 Passo 3 - Pipeline de Dados com Data Fusion

Essa é uma fase crucial do processo, pois cria o Pipeline de Dados, no qual os *raw data* são lidos, transformados e carregados no banco de dados, ou seja, é executado o processo de ETL.

A seguir descrevemos todos os passos para criar um Pipeline de Dados no Cloud Data Fusion (evidências das Figuras 16 a 26):

- Criar uma instância do Data Fusion e um bucket temporário (Figuras 16, 17 e 18)
- Criar uma instância de Banco de Dados no BiGQuary (Figura 19)
- Criar um Pipeline ETL na interface gráfica do Data Fusion (Firura 20)
- Realizar o Deploy do fluxo criado (Figura 21)
- Configurar o Deploy para execução (Figura 22)
- Avaliar a execução (Figuras 23.a até 23.d)



Figura 16 - Criando uma Instância do Data Fusion

The screenshot shows the Google Cloud Data Fusion Instances page. At the top, there's a navigation bar with links for Página da web não..., Gmail, YouTube, and Maps. Below it, the Google Cloud logo and the project name "Sprint3-PUC-DataAnalytics" are displayed. The main area is titled "Instâncias" and contains a table with one row. The table columns are: Nome da Instância (Name), Ação (Action), Edição (Edition), Região (Region), Zona (Zone), Versão (Version), Criptografia (Encryption), Criada (Created), Última atualização (Last updated), and Rótulos (Labels). The single instance listed is "sprint3-anatel".

Figura 17 - Instância do Data Fusion criada

The screenshot shows the Google Cloud Storage Buckets page. The left sidebar has sections for Buckets, Monitoramento (Monitoring), and Configurações (Configurations). The main area is titled "Detalhes do bucket" and shows a bucket named "sprint3-bucket-anatel-temp". It displays details like Local (Location: us-central1 (Iowa)), Classe de armazenamento (Storage class: Standard), Acesso público (Public access: Não público), and Proteção (Protection: Nenhum). Below this, there are tabs for OBJETOS (Objects), CONFIGURAÇÃO (Configuration), PERMISSÕES (Permissions), PROTEÇÃO (Protection), CICLO DE VIDA (Lifecycle), OBSERVABILIDADE (Observability), and RELATÓRIOS DE INVERSAO (Inversion Reports). A sub-section for "Intervalos" shows a single object named "sprint3-bucket-anatel-temp". There are buttons for FAZER UPLOAD DE ARQUIVOS (Upload files), CARREGAR PASTA (Upload folder), CRIAR PASTA (Create folder), TRANSFERIR DADOS (Transfer data), FAZER O DOWNLOAD (Download), and EXCLUIR (Delete). A search bar at the top right says "Pesquise (/) recursos, documentos, produtos e muit...".

Figura 18 - Criando um bucklet temporário para uso pela instância Fusion: sprint3-bucket-anatel-temp

The screenshot shows the Google Cloud BigQuery Explorer page. The left sidebar has sections for Análise (Analysis), Espaço de trabalho SQL (SQL Workspace), Transferências de dados (Data transfers), Consultas programadas (Scheduled queries), Analytics Hub, Dataform, and Central de parceiros (Partnership center). The main area is titled "Este é seu SQL Workspace." and "Vamos começar". It shows a tree view of datasets: "sprint3-puc-dataanalytics" and "sprint3rqual". Under "sprint3rqual", there are two datasets: "DatasetRQualPlus" and "DatasetRQualPlus_copy". A button "+ ADICIONAR" (Add) is visible. A sidebar on the right lists "Acessado recentemente" (Recently accessed) datasets: "DatasetRQualPlus" and "DatasetRQualPlus_copy".

Figura 19 - Instância do BigQuery criada como *sprint3rqual*

Nome do Fluxo: "description": "ETL para os arquivos referentes ao Indicador de Qualidade IQS do RQUAL Anatel - dados de Cobertura e velocidade de dados", "name": "sprint3-PUC-RQUAL_Cobertura",

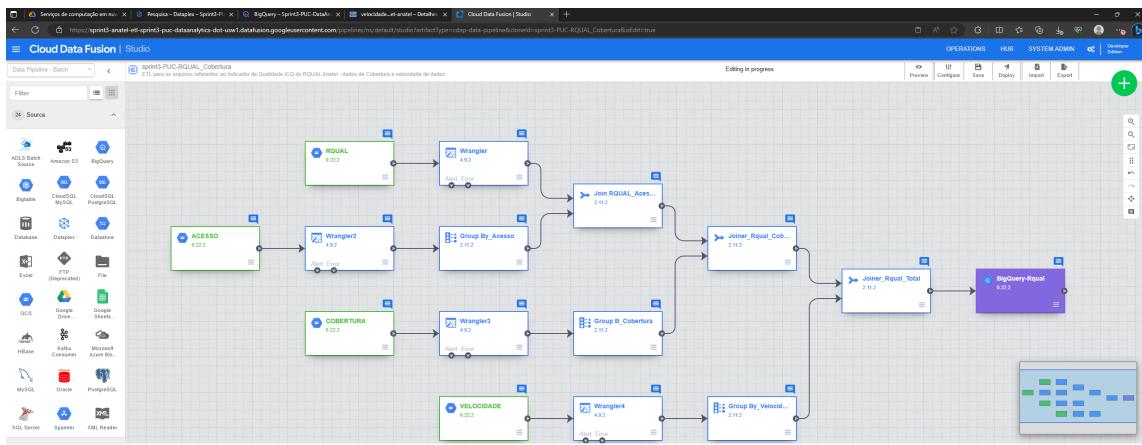


Figura 20 - Evidência do Pipeline ETL para ingestão no BigQuery dos dados do RQUAL e suas dimensões como uma Tabela Flat

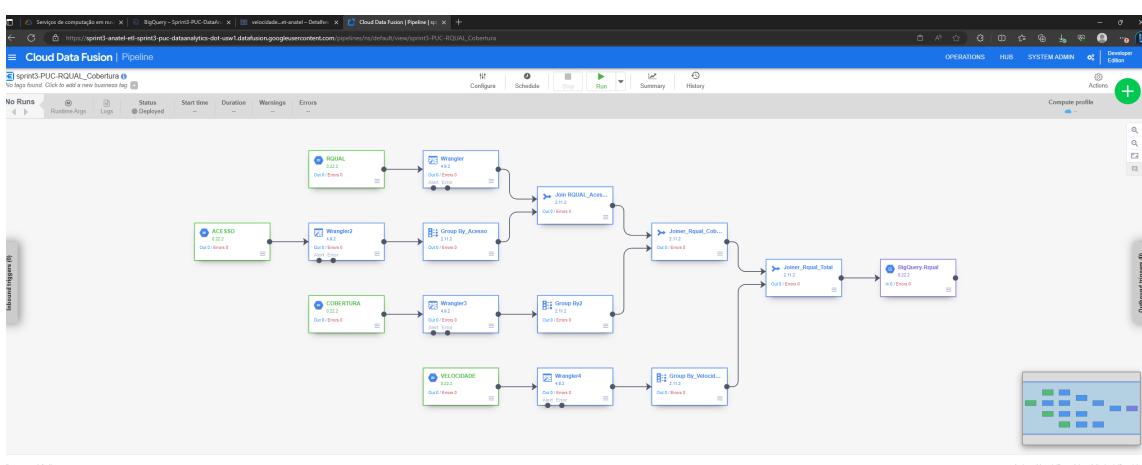


Figura 21 - Deploy do fluxo ETL para execução no Data Fusion

Select the compute profile you want to use to run this pipeline						
Profile name	Provisioner	Total cores	Scope	Status	Customize	View
Dataproc	Dataproc	4	SYSTEM	Enabled	Customize	View
Autoscaling Dataproc	Dataproc	Up to 84 (Auto)	SYSTEM	Enabled	Customize	View

Figura 22 - Configurando a instância Data Fusion com autoscaling para execução do fluxo.

- Observação: Na etapa de execução, inicialmente tentamos executar na configuração padrão (Dataproc), que define o número de núcleos e memória fixados para a máquina virtual (VM). Após várias tentativas, nas quais o fluxo era finalizado sem sucesso, e após

analisar os registros de execução, foi constatado que poderia ser um problema de recursos da VM. Portanto, optamos pela configuração de Autoscaling, que prontamente solucionou o problema de execução.

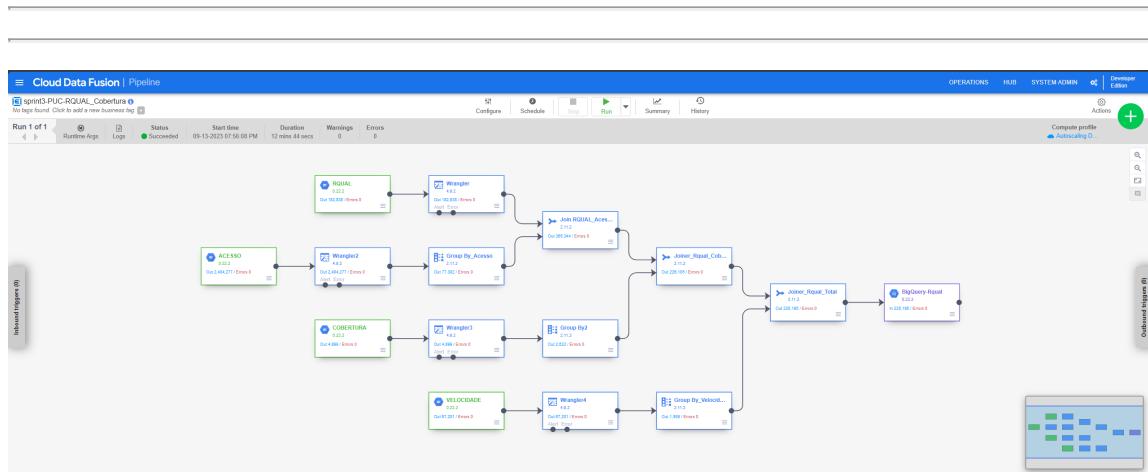


Figura 23.a

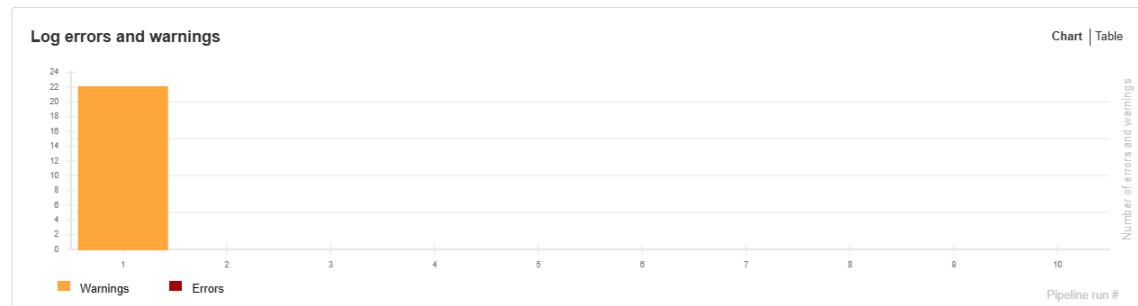


Figura 23.b

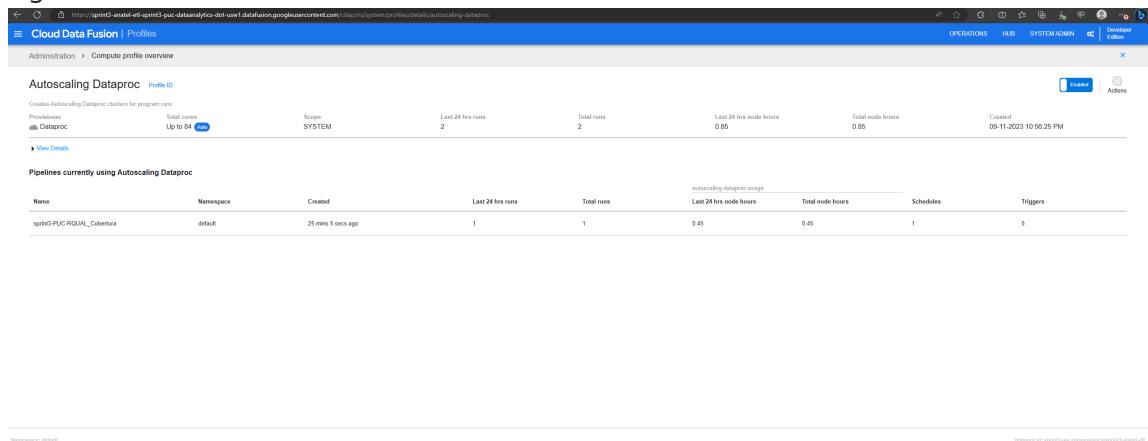


Figura 23.c

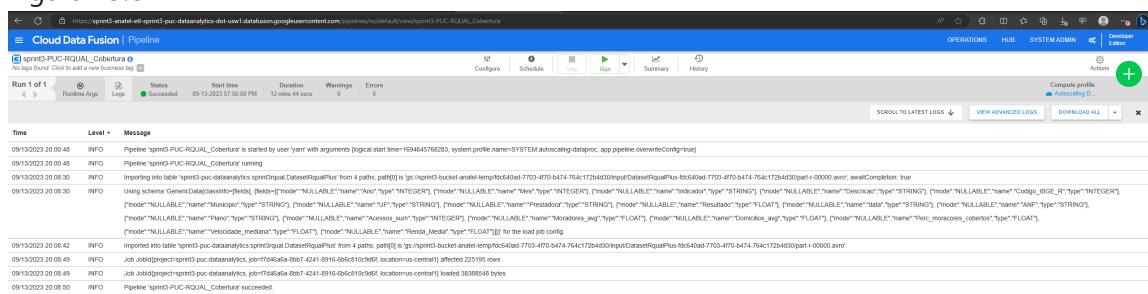


Figura 23.d

Figura 23.a até Figura 23.d - Fluxo executado com sucesso em 12min e 44 sec, sem alarmes de erros e 225.195 registros carregados no Banco de Dados (BigQuery).

Explicando o Fluxo ETL da Figura 20

Objetivo

O objetivo do fluxo é extrair, transformar e carregar dados de qualidade do indicador (IQS) do RQUAL da Anatel, especificamente dados de acessos, cobertura e velocidade de dados.

Descrição

O fluxo é composto por cinco etapas principais:

Extração: O fluxo extrai os dados do RQUAL, Acessos, Cobertura e Velocidade dos arquivo CSV armazenados no Google Cloud Storage (GCS). O plugin GCSFile é usado para essa tarefa.

Transformação: O fluxo transforma os dados RQUAL, de acesso e velocidade e cobertura usando o plugin Wrangler. Os campos são renomeados, removidos e alterados de tipo de dados, conforme necessário. Abaixo seguem os exemplos de transformações executadas no arquivo RQUAL:

Renomeando atributos:

- M_s para Mes
- Servi_o para Prestadora
- Tipo para Tipo_IQS
- Grupo para Grupo_IQS
- C_digo_IBGE paraCodigo_IBGE_R
- data para Data
- Indicador para Indicador
- Descrição para Descrição
- ANF para ANF_Acesso>
- Plano para Plano_Acesso
- data para Data
- Indicador para Indicador
- Descrição para Descrição
- Código_IBGE para Código_IBGE_R
- Municipio para Municipio
- UF para UF
- Prestadora para Prestadora
- Resultado para Resultado

Alterado tipo:

- Resultado **de** string **para** double

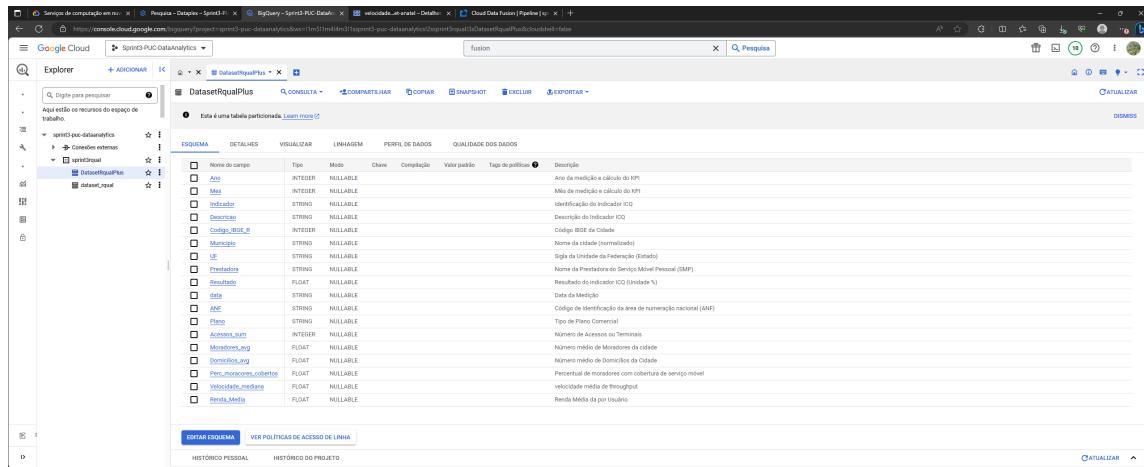
Join: O fluxo junta os dados RQUAL, de acesso, velocidade e cobertura usando o plugin Joiner.

Carregamento: O fluxo carrega os dados agregados na tabela BigQuery-Rqual.

Validação: O fluxo valida os dados carregados na tabela BigQuery-Rqual.

4.5 Passo 4 - Dados no BigQuary

O fluxo ETL do Data Fusion fez a ingestão da tabela flat resultante, *DatasetRqualPlus*, na instância *sprint3rqual* do BigQuary criado para o projeto *sprint3-puc-dataanalytics*. Nas Figuras que acompanharemos na sequência são mostradas as evidências da tabela flat *DatasetRqualPlus* carregada, o esquema correspondente, o catálogo no Dataplex e uma consulta SQL para contagem de linhas da tabela.



The screenshot shows the Google Cloud BigQuery interface. In the top navigation bar, it says "Sprint3 - Datafusion - Sprint3-PUC" and "BigQuery - Sprint3-PUC-Datal...". The main window displays the schema of the "DatasetRqualPlus" table. The schema includes columns such as "Ano" (INTEGER), "Mes" (INTEGER), "Indicador" (STRING), "Denominacao" (STRING), "Codigo_IBGE_R" (INTEGER), "Municipio" (STRING), "UF" (STRING), "Prestadora" (STRING), "resultado" (FLOAT), "data" (STRING), "ANF" (STRING), "Plano" (INTEGER), "Avaliacao_esp" (INTEGER), "Monitores_avg" (FLOAT), "Dominicos_avg" (FLOAT), "Peri_moradores_colaboradores" (FLOAT), "Velocidade_mediana" (FLOAT), and "Renda_Media" (FLOAT). Each column has a detailed description provided by the tooltip. At the bottom of the schema view, there are buttons for "EDITAR ESQUEMA" and "VER POLÍTICAS DE ACESSO DE LINHA".

Figura 24 - Evidência de Esquema do DatasetRqualPlus no BigQuery

```
SELECT COUNT(*) FROM `sprint3-puc-dataanalytics.sprint3rqual.DatasetRqualPlus`;
```

The screenshot shows a Data Studio interface with a query editor and a results table.

Query Editor:

- URL: 2023-09-...:45
- Session: *Sem título 2
- Buttons: EXECUTAR, Consulta concluída.
- Query:

```
1 SELECT COUNT(*) FROM `sprint3-puc-dataanalytics.sprint3rqual.DatasetRqualPlus`;
```

Results Table:

Linha	f0_
1	225195

Bottom Navigation:

- HISTÓRICO PESSOAL
- HISTÓRICO DO PROJETO
- CATUALIZAR

Figura 25 - Query simples para contagem do número de registros da tabela

A construção do Catálogo de Dados da tabela *DatasetRqualPlus* pode ser realizada utilizando o Cloud Dataflex. Na sequencia de Figuras mostramos uma descrição da **Visão Geal**, a documentação de cada atributo do esquema e um Glossário dos indicadores IQS.

The screenshot shows the Google Cloud Dataplex search interface. On the left, there's a sidebar with various navigation options like 'Descobrir', 'Explorar', 'Pesquisa', 'Gerenciar catálogo', 'Gerenciar lakes', 'Seguro', 'Processar', 'Governar', and 'Armazenamento de atributos'. The main area has a 'Filters' section with 'Escopo' set to 'Tudo' and 'Sistemas' set to 'BigQuery'. A search bar at the top right says 'Encontrar recursos de dados nos seus projetos e organizações' with a 'PESQUISAR' button. Below the filters, there's a table with two rows:

Nome	Descrição	Tipo	Sistema
DatasetRqualPlus	Conjunto de dados: sprint3rqual	TABELA	BIGQUERY
sprint3rqual		CONJUNTO DE DADOS	BIGQUERY

Figura 26 - Seleção do BigQuery no Dataplex para selecionar a tabela a ser documentada

This screenshot shows the 'DatasetRqual...' details page in Google Cloud Dataplex. The left sidebar is identical to Figure 26. The main content area has tabs for 'DETALHES', 'ESQUEMA', 'LINHAGEM', 'PERFIL DE DADOS', and 'QUALIDADE DOS DADOS'. The 'DETALHES' tab is selected, displaying the following information:

Detalhes da Tabela do BigQuery

Local	us-central1
Consultas (últimos 30 dias)	1
URL do recurso	sprint3-puc-dataanalytics.sprint3rqual.DatasetRqualPlus
Marcadores	Editar no BigQuery
Nome totalmente qualificado	bigquery:sprint3-puc-dataanalytics.sprint3rqual.DatasetRqualPlus
Descrição	Tabela com dados de KPI ICQ do Rqual da Anatel

Visão geral

Esta tabela possui os dados cruzados a partir de uma Inner Join entre quatro bases de dados que contêm dados do Serviço Móvel Pessoal (SMP) referente ao serviço 4G ou LTE. As bases de dados são:

- RQUAL, que contém os dados dos Indicadores ICQ mensais e por ano. Nessa versão dos dados estão contidos apenas 05 meses de dados referentes aos meses de Janeiro a Junho de 2023 e para as cidades dos estados do Paraná, São Paulo, Goiás, Distrito Federal, Pernambuco e Pará. Poderíamos chamar essa Tabela como equivalente a tabela Fato de um esquema estrela.
- Acesso que contém os dados agrupados de terminais por município brasileiro e os respectivos planos comercializados pelas operadoras. Poderíamos pensar como uma Tabela Dimensão.
- A base Cobertura possui os dados agrupados de cobertura dos municípios brasileiros com informações de domicílio, número de moradores e percentual de moradores cobertos.
- Base de Velocidade por Renda contém os dados agrupados de velocidade média de up/down link e renda média por habiente. poderíamos pensar como uma tabela Dimensão.

Importante considerar que essa tabela por ser desnormalizada através de INNER JOIN perde informação em que não há combinação/coincidência dos dados de chave. Isso foi feito de propósito a fim de reduzir o tamanho da Tabela resultante.

[EDITAR VISÃO GERAL](#)

Tags (0)

• Você não tem tags anexadas a este recurso de dados.

[ANEXAR TAGS](#)

Figura 27 - Detalhes da Tabela do BigQuery com a descrição da Visão Geral do DatasetRqualPlus

Detalhe da descrição inserida no Dataplex para a Visão Geral da Tabela Flat carregada no BigQuery referente ao RQUAL.

Visão Geral

Esta tabela possui os dados cruzados a partir de uma Inner Join entre quatro bases de dados que contem dados do Serviço Móvel Pessoal (SMP) referente ao serviço 4G ou LTE. As bases de dados são:

RQUAL que contém os dados dos indicadores IQS mensão e por ano. Nessa versão dos dados estão contidos apenas 06 meses de dados referentes aos meses de Janeiro a Junho de 2023 e para as cidades dos estados do Paraná, São paulo, Goiás, Distrito Federal, Prnambuco e Pará. Poderíamos chamar essa Tabela como equivalente a tabela Fato de um esquema estrela.

Acesso que contém os dados agrupados de terminais por município brasileiro e os respectivos planos comercializados pelas operadoras. Poderíamos pensar como uma Tabela Dimensão.

A base Cobertura possui os dados agrupados de cobertura dos municíos brasileiros com informações de domicílios, número de moradores e percentual de moradores cobertos.

Base de Velocidade por Renda contém os dados agrupados de velocidade média de up/down link e renda média por habitante. poderíamos pensar como uma tabela Dimensão.

Importante considerar que essa tabela por ser desnormalizada através de INNER JOIN perde informação em que não há combinação/coincidência dos dados de chave. Isso foi feito de propósito a fim de reduzir o tamanho da Tabela resultante.

The screenshot shows the Google Cloud Data Catalog interface. On the left, there's a sidebar with navigation links like 'Dataplex', 'Descobrir', 'Explorar', 'Pesquisa', 'Gerenciar catálogo', 'Modelos de tag', 'Grupos de entradas', 'Glossários', 'Gerenciar lakes', 'Gerenciar', 'Seguro', 'Processar', 'Governar', 'Armazenamento de atributos', 'Perfil', and 'Qualidade dos dados'. The main area is titled 'DatasetRqual...' and shows the 'ESQUEMA' tab selected. It displays a table with columns: Nome do campo, Tipo, Modo, Tags de coluna, Termos de negócios, and Descrição. The table contains 18 rows of data, each with a '+' sign in the 'Tags de coluna' and 'Termos de negócios' columns. The 'Descrição' column provides detailed information about each field. At the bottom right of the table, there are pagination controls for 'Linhas por página: 50' and '1 - 18 de 18'.

Figura 28 - Documentação dos Atributos do esquema da tabela *DatasetRqualPlus*

Google Cloud Sprint3-PUC-DataAnalytics Pesquise (/) recursos, documentos, produtos e muit... PESQUISA

Dataplex

Glossários PRE-VISUALIZAÇÃO

Indicadores d... PÁGINA INICIAL + CRIAR CATEGORIA

Indicadores do SMP

Indicadores IQS representam o desempenho da Rede do Serviço Móvel Pessoal - SMP

Tipo	Grupo	Acrônimo	Indicador	Serviços	Requisitos do Processo de Aferição (Anexo N°)
Indicador do IQS	Redes	IND1	Conexão de Chamadas na Rede de Acesso	SMP	01
Indicador do IQS	Redes	IND2	Queda de Chamadas	SMP	01
Indicador do IQS	Redes	IND3	Conexão de Dados Medida na Rede de Acesso	SMP	01
Indicador do IQS	Redes	IND4	Cumprimento da Velocidade de download e upload	SMP e SCM	05 (SMP) e 06 (SCM)
Indicador do IQS	Redes	IND5	Latência bidirecional da Conexão de Dados	SMP e SCM	05 (SMP) e 06 (SCM)
Indicador do IQS	Redes	IND6	Variação de Latência da Conexão de Dados	SMP e SCM	05 (SMP) e 06 (SCM)
Indicador do IQS	Redes	IND7	Perda de Pacotes da Conexão de Dados	SMP e SCM	05 (SMP) e 06 (SCM)
Indicador do IQS	Redes	IND8	Disponibilidade	SMP, STFC, SCM e TV assinatura	07

EDITAR DESCRIÇÃO

Google Cloud Sprint3-PUC-DataAnalytics Pesquise (/) recursos, documentos, produtos e muit... PESQUISA

Dataplex

Glossários PRE-VISUALIZAÇÃO

IND1 - Conexão... PÁGINA INICIAL + CRIAR CATEGORIA

IND1 - Conexão de Chamada na Rede de Acesso

Expresso o desempenho da rede de acesso em relação ao atendimento de tentativas de conexão para chamadas de voz

Entradas relacionadas

Filtro Insira o nome ou o valor da propriedade

Nome da entrada

Nenhuma linha a ser exibida

Figura 29 - Glossário com o significado dos Indicadores RQUAL-IQS

5. Análise e Discussão

A fim de responder as questões formuladas no Objetivo principal, é necessário dividir essa seção em três etapas referentes a perguntas 1, 2 e 3, respectivamente.

5.1 Pergunta 1

Questão:

É possível reproduzir as consultas dos indicadores IQS do RQUAL, individualmente ou grupos, através de um pipeline de dados na nuvem selecionando atributos como ano, mês, indicador, cidade, estado e prestadora?, ou seja reproduzir a mesma consulta feita pelo site Anatel?

Evidências 1

No exemplo abaixo selecionamos os atributos (Mes, Indicador, Prestadora, Municipio, Resultado) para os ano de 2023 e estado de São Paulo. Na sequência, mostramos os resultados agregados por Indicador e por período para cada indicador. Para gerar o resultado realizamos uma consulta no Banco (Figura 30) e plotamos os resultados através do Looker Studio, que é incorporado ao BigQuery.

```
SELECT Mes, Indicador, Prestadora, Municipio, AVG(Resultado) AS Resultado_Medio
FROM `sprint3-puc-dataanalytics.sprint3rqual.DatasetRqualPlus`
WHERE UF = 'sp' AND Ano = 2023
GROUP BY Mes, Indicador, Prestadora, Municipio
ORDER BY Mes, Indicador;
```

Figura 30 - Consulta SQL para gerar os dados para a Resposta 1.

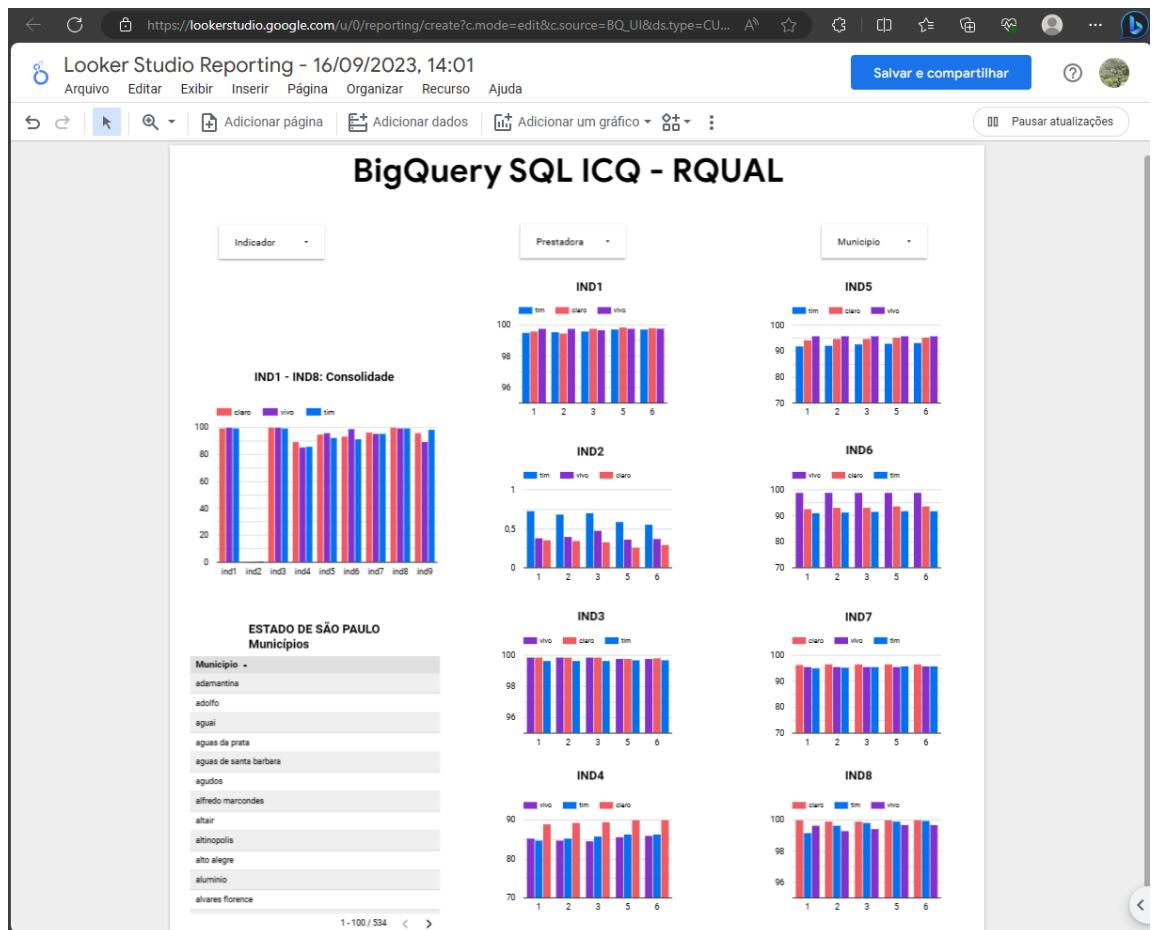


Figura 31 - Gráficos co os indicadores IQS para Resposta 1 - dados Agregados para o Estado de São Paulo.

BigQuery SQL ICQ - RQUAL

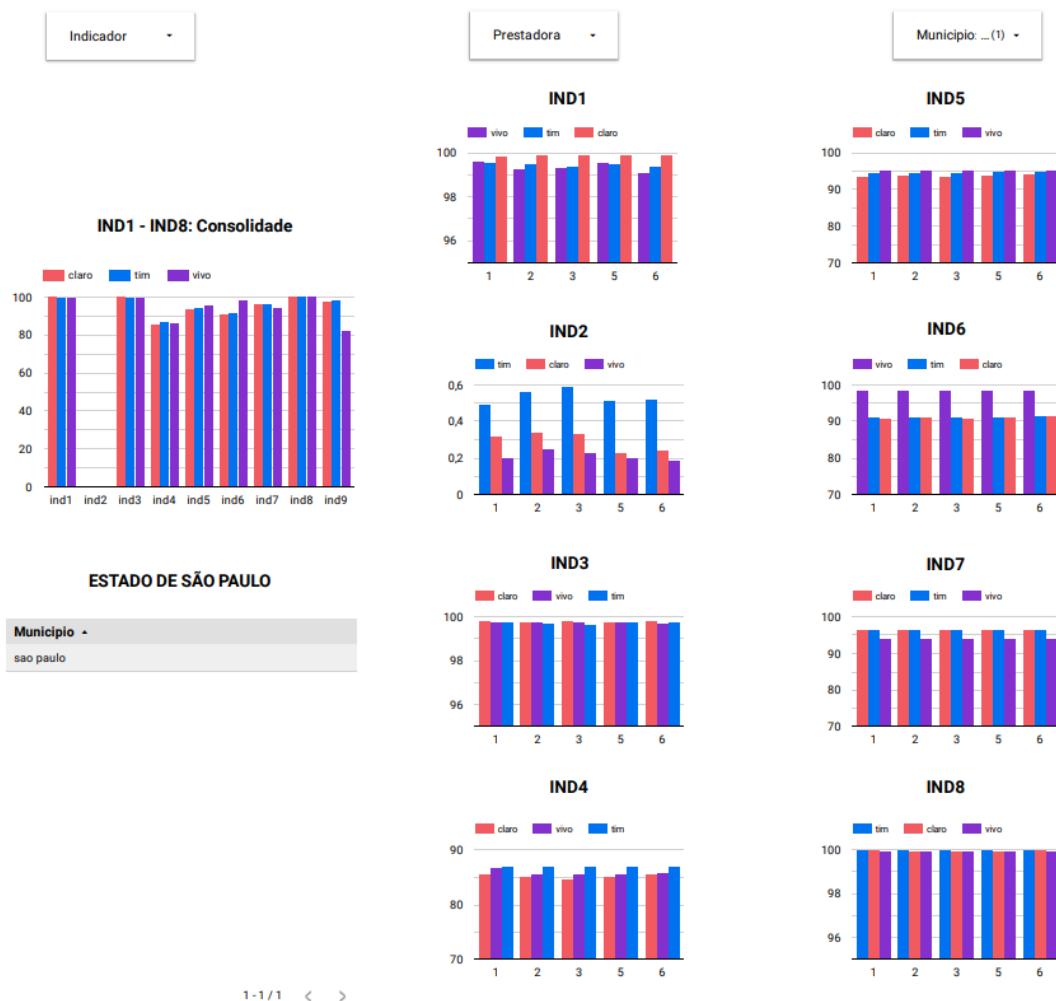


Figura 32 - Gráficos co os indicadores IQS para Resposta 1 - dados Agregados para a cidade de São Paulo.

Legenda para os Indicadores:

- IND1 - Conexão de Chamada na rede de Acesso
- IND2 - Queda de Chamada
- IND3 - Conexão de Dados Medida na Rede de Acesso
- IND4 - Cumprimento de Velocidade de Download e Upload
- IND5 - Latência bidirecional da Conexão de Dados
- IND6 - Variação de Latência da Conexão de Dados
- IND7 - Perda de Pacotes da Conexão de Dados
- IND8 - Disponibilidade

Analise Descritiva 1

Com a consulta SQL realizada, na Figura 30, é possível gerar Dashboards através do Cloud Looker Studio ou outra ferramenta de visualização. Nossa opção foi gerar através do Looker Studio que já é integrada ao BigQuery. Nas Figuras 31 e 32 mostramos um exemplo de Dashboard em que usando os filtros Indicador, Prestadora e Município é possível selecionar um ou vários valores de atributos e avaliar os dados gerados para os indicadores agregados no tempo ou visualizar cada indicador entre os meses 1 e 6. As Figuras mostram

os gráficos comparáveis entre as Operadoras, podendo ser verificado o desempenho de cada uma por Estado ou Município.

Resposta 1:

Sim. Uma vez que a tabela do IQS do RQUAL foi carregada no BigQuery em forma de Banco de Dados é perfeitamente possível reproduzir os resultados mostrados pelo site da Anatel com os mesmos dados e atributos.

5.2 Pergunta 2

Questão:

É possível incrementar a essas consultas filtros que considerem o tamanho da população das cidades, quantidade de assinantes por operadora e renda percapita ou PIB?

Evidências 2

A consulta SQL mostrada abaixo faz uma seleção dos atributos do RQUAL acrescidos dos dados de Renda, Moradores, Velocidade e Quantidade de Acessos por Município e Operadora, respectivamente.

```
SELECT Mes, Indicador, Prestadora, Municipio, ANF,
       AVG(Resultado) AS Resultado_Medio,
       AVG(Acessos_sum) AS Acessos,
       AVG(Renda_Media) AS Renda_Media,
       AVG(Velocidade_médiana) AS Thp_Medio,
       AVG(Moradores_avg) AS Numero_habitantes
  FROM `sprint3-puc-dataanalytics.sprint3rqual.DatasetRqualPlus`
 WHERE UF = 'sp' AND Ano = 2023
 GROUP BY Mes, Indicador, Prestadora, Municipio, ANF
 ORDER BY Mes, Indicador;
```

The screenshot shows the Google Cloud BigQuery interface. At the top, there's a search bar with the placeholder "Pesquise (/) recursos, documentos, produtos e muito mais..." and a "Pesquisa" button. Below the search bar, the title bar displays "DatasetRQualPlus" and the date "2023-09-16 15:38:48". The main area is divided into two sections: "Resultados da consulta" (Results of the query) and the "SQL Editor".

SQL Editor:

```

1  SELECT
2      ANF,
3          AVG(Resultado) AS Resultado_Medio,
4          AVG(Acessos_sum) AS Acessos,
5          AVG(Renda_Media) AS Renda_Media,
6          AVG(Velocidade_mediana) AS Thp_Medio,
7          AVG(Moradores_avg) AS Numero_habitantes
8  FROM
9      `sprint3-puc-dataanalytics.sprint3rqual.DatasetRQualPlus`
10 WHERE
11     UF = 'SP' AND
12     Ano = 2023
13 GROUP BY
14     Mes,
15     Indicador,
16     Prestadora,
17     Municipio,
18     ANF
19 ORDER BY
20     Mes,
21     Indicador;

```

Resultados da consulta:

Linha	ANF	Resultado_Medio	Acessos	Renda_Media	Thp_Medio	Nums
1	11	99.63	2553.5	718.3399999999...	18061.15416666...	
2	os perdoes	99.87	772.5	754.9952000000...	12065.15480000...	
3	11	99.67	1643.0	464.5038888888...	10162.92833333...	
4	11	99.84	2936.0	464.5038888888...	10162.92833333...	
5	sta	99.81	5659.5	569.9730000000...	14184.73599999...	
6	11	99.86	321.5	604.155	7397.9950000000...	
7	n	99.68	821.0	560.3478571428...	12254.08428571...	
8	n	99.77	1877.5	560.3478571428...	12254.08428571...	
9	11	99.63	1877.5	639.5916216216...	11741.04756756...	
10	11	99.62	7975.5	639.5916216216...	11741.04756756...	

Below the table, it says "Resultados por página: 50 ▾ 1 – 50 de 40853" and has "ATUALIZAR" and "HISTÓRICO PESSOAL" buttons.

Figura 31 - SQL que acresnta os dados de Renda, Velocidade e Quantidades de Acesso aos dados de RQUAL

This screenshot shows the same BigQuery interface as Figure 31, but the results are displayed in a more detailed, grid-based "VISUALIZAR" (Visualize) view. The columns are labeled: Línea, Ano, Mes, Indicador, Descrição, Código_Indicador, Municipio, UF, Prestadora, Resultado, data, ANF, Plano, Acessos_Med, Moradores_Med.

The data grid contains approximately 22 rows of results, each corresponding to a specific indicator and its details. The columns provide a breakdown of the data by year, month, indicator, description, code, municipality, state, provider, result, date, ANF, plan, average access, and average inhabitants.

Figura 34 - Print com o resultado da consulta

Pressione Alt+F1 para abrir as opções de acesso direto

Linha	ANF	Resultado	Velocidade	Acessos	Renda	Moradores
1	11	99.01	3059.0	309.9730000000000...	14184.73099999999...	1090.0
2	11	99.86	321.5	604.155	7397.99500000000...	5754.0
3	11	99.68	821.0	560.3478571428...	12254.08428571...	28541.0
4	11	99.77	1877.5	560.3478571428...	12254.08428571...	28541.0
5	11	99.63	1877.5	639.5916216216...	11741.04756756...	43143.0
6	11	99.62	7975.5	639.5916216216...	11741.04756756...	43143.0
7	11					
8	11					
9	11					
10	11					

Consultas Atualizadas

Explorar com as Planilhas Analise Big Data com uma conexão ativa em uma ferramenta de planilhas que você já conhece.

Explorar com o Looker Studio Confira resultados e crie painéis em tempo real com base nos seus dados.

Analizar com o notebook Python Explore e visualize com o Python.

PRÉ-VISUALIZAÇÃO

Por página: 50 | 1 - 50 de 40853 | < > |

Figura 35 - Opções do BigQuery para exportação do resultado da consulta (Planilha, Looker ou Notebook Py)

Resposta 2

Sim. Esta consulta SQL recupera dados da tabela *DatasetRqualPlus* a partir do banco de dados *sprint3-puc-dataanalytics.sprint3rqual* para o estado de São Paulo (UF = 'sp') no ano de 2023. Ela calcula médias das métricas (Resultado, Velocidade, Acessos, Renda e Moradores) agrupadas por mês, indicador, prestadora, município e ANF, e, em seguida, ordena os resultados com base em mês e indicador. Com os dados recuperados nessa consulta é possível partirmos para responder as questões da Perguntas 3.

5.3 Pergunta 3

Questão

Fazer uma análise como:

- a. melhores e piores IQS – por município
- b. melhores e piores IQS – por prestadora
- c. melhores e piores IQS – por população
- d. melhores e piores IQS – por renda

Evidências 3

BigQuery SQL ICQ - RQUAL: Visão por Renda, Acessos, Habitantes e Thp

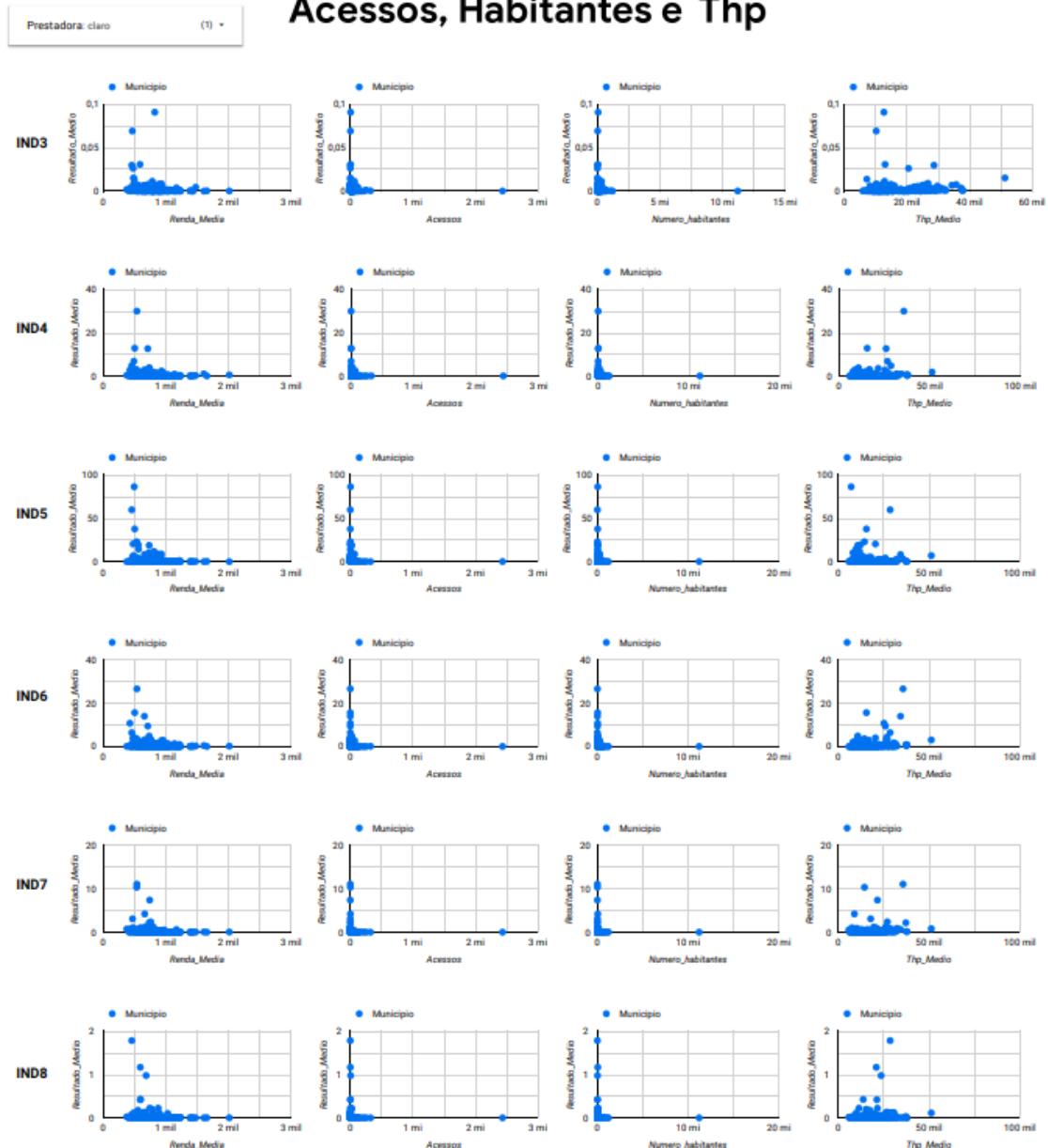


Figura 36 - Variância dos Indicadores - Operadora 1

[Link para Detalhe dos Indicadores da Operadora 1](#)

[Link para Detalhe dos Indicadores da Operadora 1 e ANF](#)

BigQuery SQL ICQ - RQUAL: Visão por Renda, Acessos, Habitantes e Thp

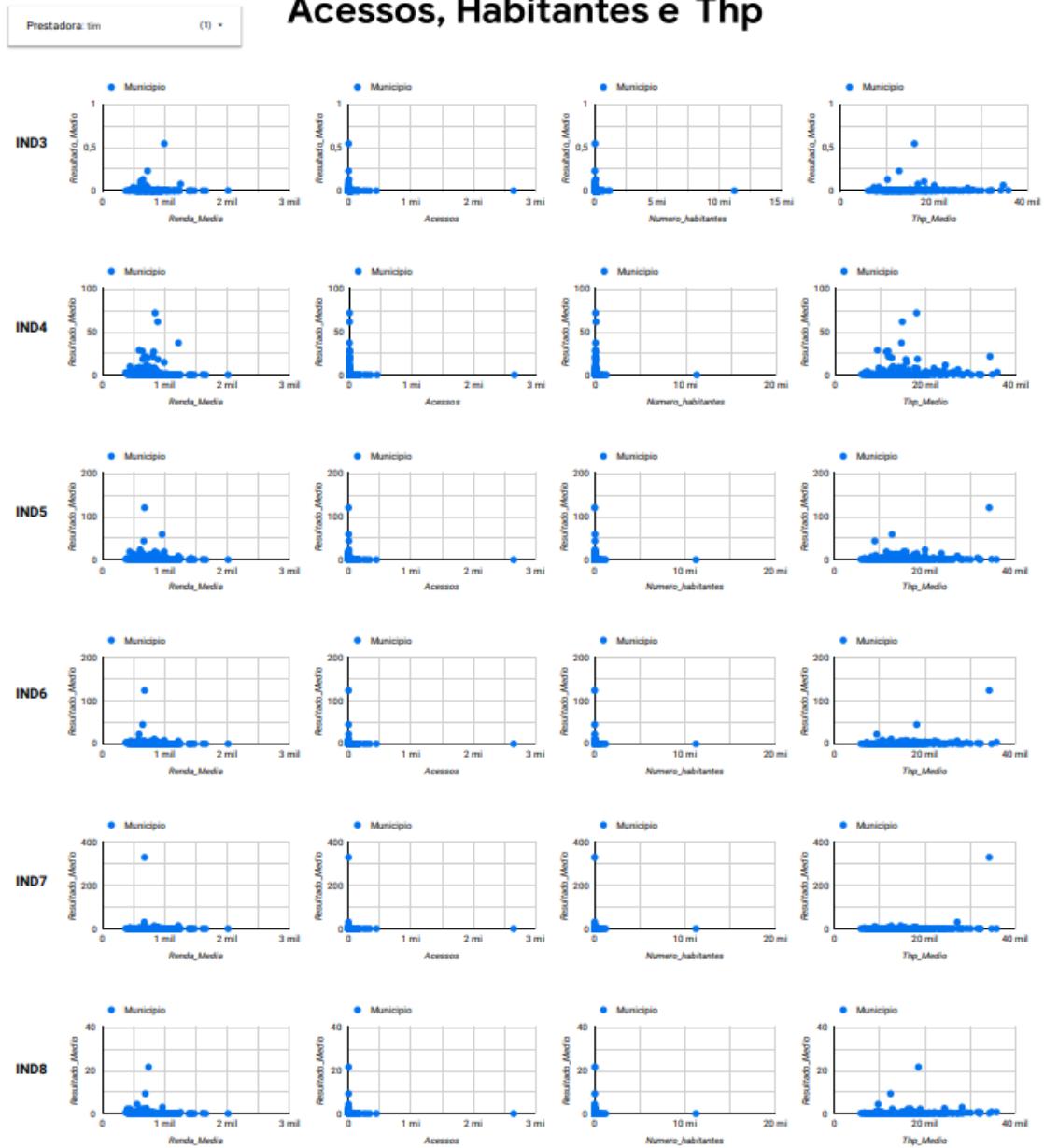


Figura 37 - Variância dos Indicadores - Operadora 2

[Link para Detalhe dos Indicadores da Operadora 2](#)

[Link para Detalhe dos Indicadores da Operadora 2 e ANF](#)

BigQuery SQL ICQ - RQUAL: Visão por Renda, Acessos, Habitantes e Thp

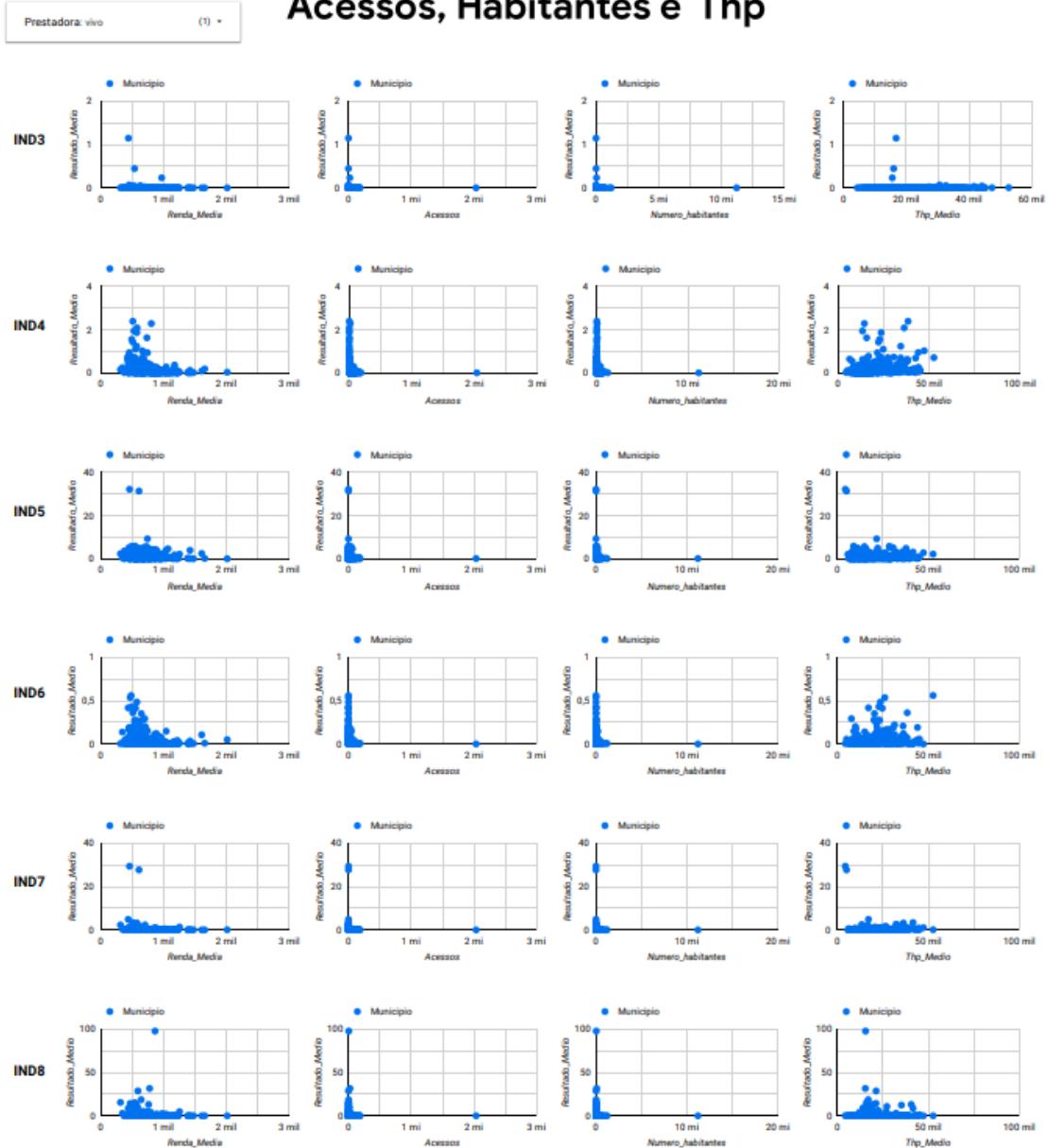


Figura 38 - Variância dos Indicadores - Operadora 3

[Link para Detalhe dos Indicadores da Operadora 3](#)

[Link para Detalhe dos Indicadores da Operadora 3 e ANF](#)

[Link para Detalhe Operadoras](#)

Analise Descritiva 3

As evidências de análises geradas e exibidas nas Figuras e Links da seção anterior referem-se às três principais operadoras do Serviço Móvel Pessoal (SMP) com presença em todas as unidades federativas do Brasil. As análises se repetem igualitariamente para as três operadoras, fazendo uso dos mesmos atributos, mas com os dados restritos a cada uma delas. Evitaremos comparação de rotulação entre elas porque a Anatel, detentora do selo RQUAL, ainda não realizou ou divulgou o resultado de sua classificação. Portanto, qualquer análise nesse sentido poderia não considerar aspectos relevantes, como a pesquisa

qualitativa, ou até contestações, considerações e justificativas das operadoras acerca dos dados coletados. Vamos focar nossa análise na Operadora 1, considerando que o perfil entre elas é muito semelhante como indicado acima.

Legenda para os Indicadores:

- IND3 - Conexão de Dados Medida na Rede de Acesso
 - IND4 - Cumprimento de Velocidade de Download e Upload
 - IND5 - Latência bidirecional da Conexão de Dados
 - IND6 - Variação de Latência da Conexão de Dados
 - IND7 - Perda de Pacotes da Conexão de Dados
 - IND8 - Disponibilidade
-

Os gráficos da Figura 36 mostram, para cada Indicador (IND3 a IND8), a dispersão das variâncias dos dados referentes aos atributos Média dos Resultados por Renda Média, Acessos, Moradores e Thp_médio, considerando o conjunto de municípios do Estado de São Paulo.

```
SELECT COUNT(DISTINCT Municipio) AS NumeroDeMunicipios
FROM `sprint3-puc-dataanalytics.sprint3rqual.DatasetRqualPlus`
WHERE UF = 'sp' AND Ano = 2023;

/* Resultado
NumeroDeMunicipios
534
*/
```

Primeiramente, vamos fazer uma distinção para os dados de Acessos e Número de Assistentes, pois é certo que existe uma linearidade ou proporcionalidade entre eles. Portanto, nas análises que se referem apenas a um deles.

Para a Figura 36, é notável que os gráficos estão organizados de forma semelhante a uma tabela, cujas linhas são os indicadores (IND1 a IND8) e as colunas são os atributos de Renda, Acessos, habitantes e Velocidade (Thp).

Quando observamos os gráficos verticalmente, ou seja, pelos atributos Renda, Acessos/Habitantes e Thp, verificamos um perfil de variância dos Resultados semelhantes para todos os Indicadores. Por exemplo, para os atributos Acessos/Habitantes, notamos que em todos os indicadores, a dispersão dos resultados ocorre com maior amplitude em torno das menores populações, que também são mais numerosas em quantidade de municípios. Entretanto, quando avaliamos pelo recorte de Renda_média, verifica-se uma maior variação de renda entre os municípios, contudo, sem grandes variações nos resultados dos Indicadores. Evidentemente, também há de se destacar que as maiores variações ocorrem mais concentradas nas menores rendas. Para o atributo Thp, que é um parâmetro de Rede, percebemos entre os indicadores gráficos ligeiramente diferentes, visto que o thp_médio é influenciado pelo parâmetro medido em cada indicador. Mesmo assim, percebemos, em média, que o comportamento dos indicadores varia pouco, o que indica estabilidade no comportamento da Rede.

Continuando com a Operadora 1, a análise dos gráficos contidos nos dois links de arquivos em PDF que seguem a Figura 36. Salientamos que o conteúdo dos dois links é o mesmo, distinguindo-se pela plotagem das ANFs (que representam as regiões de DDD). Por exemplo, as ANFs do Estado de São Paulo variam de 11 (região da Grande São Paulo) até 19 (região de Campinas). No entanto, os gráficos contidos nos arquivos representam todas as ANFs para os Estados do Pará, Pernambuco, Goiás, Distrito Federal, São Paulo e Paraná.

Os gráficos estão organizados em três blocos, páginas distintas, onde cada página contém todos os indicadores (IND3 a IND8) por um atributo (Renda, Thp ou Acesso).

Na primeira página, avaliamos os Indicadores por Renda e os distribuímos por ANF. Aqui podemos perceber dois grupos distintos quando avaliamos o eixo da renda. À esquerda, onde a renda é menor, estão as ANFs dos Estados do Norte (91, 93 e 94) e Nordeste (81 e 87), e à direita, estão as ANFs dos Estados com maior renda. Também na visão do eixo dos indicadores, destacamos que as ANFs dos Estados do Norte possuem um desempenho inferior em relação às demais ANFs. Quando passamos para a página seguinte, que trata do Thp, também notamos a presença das ANFs da região Norte com desempenho inferior, porém, agora acompanhadas de outras ANFs eventualmente. Por fim, no recorte de Acessos, também há um destaque para as ANFs do Norte, especialmente no IND8, que trata da disponibilidade do serviço, onde se destaca um desempenho inferior.

Nas Figuras 37 e 38, bem como nos links que as seguem, o leitor poderá verificar e realizar, com a análise metodológica, os insights para as demais Operadoras.

Resposta 3

A resposta a questão 3 é um desdobramento da resposta a questão 2, ou seja é a evidência que podemos enriquecer a avaliação dos indicadores do RQUAL agregando os recortes de Renda, Número de habitantes, "Market Share" através do Número de Acessos, Velocidade de Dados.

5.4 Discussão Final

Em resumo, as análises e respostas as questões elaboradas confirmam que é plenamente viável replicar as consultas dos indicadores IQS do RQUAL na nuvem, permitindo uma análise personalizada dos dados. Além disso, é possível enriquecer ou aprimorar essas consultas com atributos e filtros adicionais, como tamanho da população, quantidade de assinantes e renda per capita, ampliando as oportunidades de exploração e compreensão dos indicadores. A segmentação de dados também se revelou importante, demonstrando padrões de comportamento distintos com base em atributos como renda, acessos e população. No entanto, vale destacar que a análise se concentrou nas três principais operadoras do SMP e evitou comparações diretas entre elas, reconhecendo a necessidade de considerar aspectos qualitativos e contestações dos dados. Em suma, esta análise fornece uma base incipiente para a compreensão e exploração dos indicadores IQS do RQUAL, com potencial de aplicação em diversos contextos, serviços e operadoras adicionais.

Destacamos também a viabilidade de replicar consultas, acrescentar filtros, segmentar dados, além de ressaltar as suas limitações, o que é valioso para profissionais que buscam

insights açãoáveis a partir dos indicadores IQS do RQUAL, alinhando ao objetivo principal de explorar esses dados de forma mais personalizada e contextualizada.

Importante mencionar que os dados utilizados foram restritos a apenas um subconjunto dos dados disponibilizados pela Anatel. Sendo perfeitamente possível escalar o trabalho descrito acima para a totalidade dos dados disponibilizados pela Agência.

6. Conclusão

Antes de concluir este trabalho, é importante relembrar o objetivo principal, que era reproduzir os indicadores RQUAL-IQS da Anatel para o Serviço Móvel Pessoal, enquanto agregávamos novos atributos, como a Renda Média do Usuário, o Número de Acessos Móveis por Município, a Cobertura da Rede e a Velocidade de Navegação de Dados. Isso nos permitiu realizar avaliações detalhadas do comportamento desses indicadores, considerando os diferentes recortes adicionados. Especificamente para o MVP proposto durante o Sprint de Engenharia de Dados no curso de Ciência de Dados e Analytics da PUC-Rio, o requisito principal era construir todo o Pipeline de Dados usando tecnologia em nuvem, e neste trabalho, optamos pelo Google Cloud Console.

Resumindo, podemos afirmar que ambos os objetivos, dentro do escopo e das restrições estabelecidas para o trabalho, foram completamente atendidos e detalhados no Capítulo 4, que trata da construção do Pipeline em Cloud, e no Capítulo 5, onde detalhamos os resultados obtidos. Além disso, demonstramos que é possível analisar a qualidade do serviço desde o nível municipal até agregações maiores, como Unidades Federativas (UF) e Áreas de Numeração Fixa (ANF). Isso inclui a detecção de comportamentos específicos para cada indicador e destaque para as Operadoras de Serviço. Como um insight deste trabalho, notamos que, para a Operadora 1, a região Norte do país pode ser um ponto de atenção e investigação em relação ao comportamento de sua rede. Existem outros pontos de destaque que podem ser observados em relação às outras operadoras, com base nos gráficos e recortes de dados gerados aqui, ou mesmo em novas agregações ou visualizações.

Um aspecto relevante foi o impacto inicial de utilizar tecnologia em nuvem, especialmente para um usuário sem experiência nesse ambiente. No entanto, devido a abundância de recursos funcionais e uma extensa documentação disponível, essa curva de aprendizado foi rapidamente superada, permitindo a conclusão bem-sucedida do trabalho e a realização dos objetivos propostos.

É fundamental mencionar que este trabalho se concentrou em um conjunto restrito de análises e dados, limitando-se a seis estados (PR, SP, GO, DF, PE e PA) e com um enfoque mais detalhado no estado de São Paulo. Uma possível evolução deste trabalho poderia incluir a correlação dos indicadores RQUAL-IQS com dados de sentimentos dos usuários, seja por meio da pesquisa da Anatel que resulta no Índice de Qualidade Percebida (IPQ) ou por meio de pesquisas em sites de reclamações e avaliações de satisfação disponíveis em sites especializados ou redes sociais.

Obs: Cuidado com o uso excessivo da Instância do Data Fusion porque representa um grande consumo de créditos.