

TESTE TÉCNICO BUSINESS INTELLIGENCE

**Breves explicações sobre o
processo de solução dos problemas**

Glauber César Silva e Silva

Criar dashboard automatizado para acompanhamento de métricas de Apps música da Apple Store

1ª VERSÃO - MVP UTILIZANDO POWER BI PARA DATAVIZ E POWER QUERY PARA ETL:

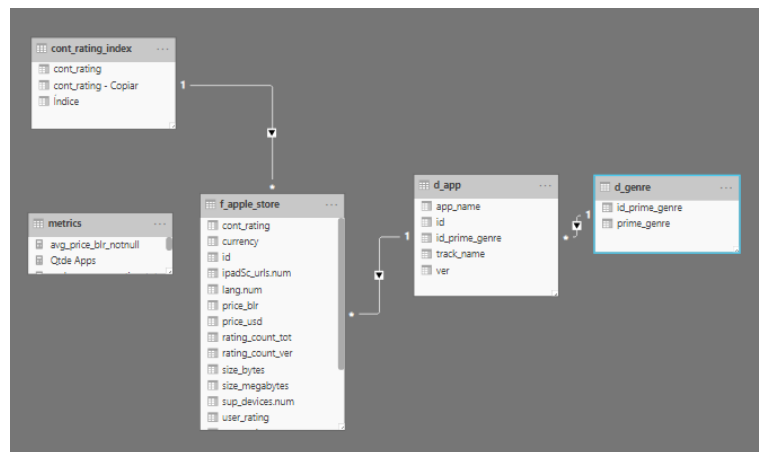
1. Conexão com arquivo csv no Google Drive

2. Criação do Modelo Dimensional:

- d_genre: gênero primário do aplicativo e id do gênero primário

- d_app: dados categóricos referente ao aplicativo

- f_apple_store: demais dados quantitativos

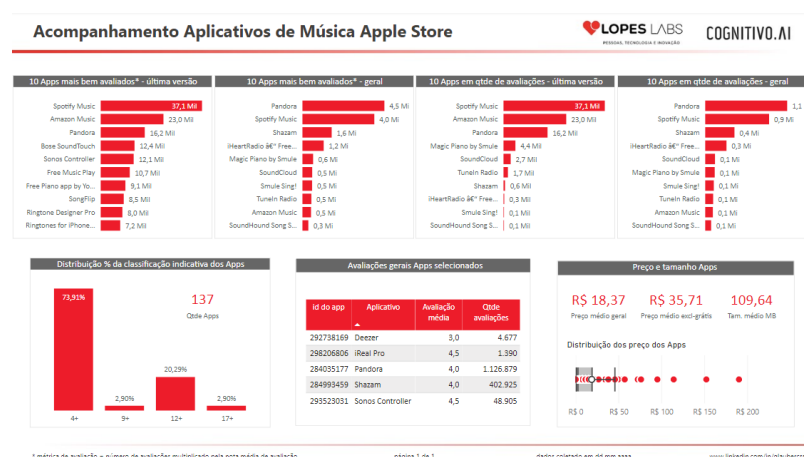


3. Criação dos gráficos e tabelas solicitadas:

- Foi criada métrica para os gráficos “ranking dos 10 aplicativos mais bem avaliados” obtida pela multiplicação do número de avaliações pela nota média das avaliações. Do ponto de vista de negócio, penso ser importante ponderar a nota média com o número de avaliações afim de evitar distorções como aplicativo com apenas 1 avaliação e nota máxima figurar entre os mais ranqueados.

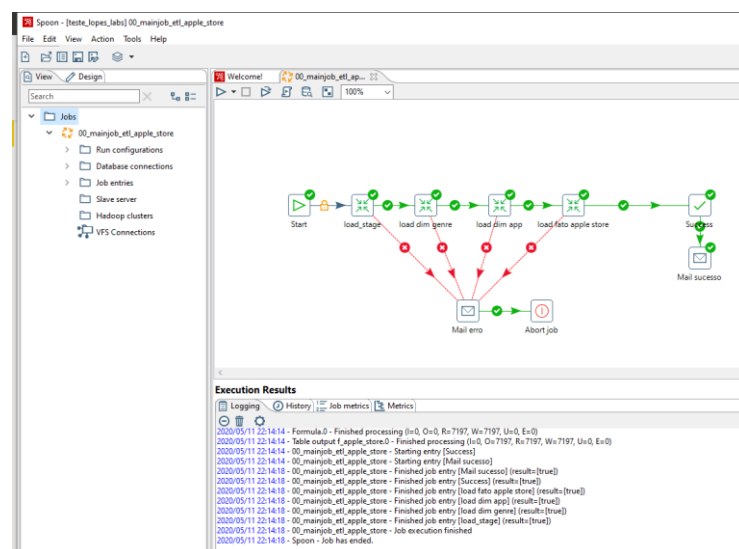
- Itens adicionadas:

- . preço médio do aplicativo em Reais desconsiderando os aplicativos gratuitos ou *fremium*
- como é o caso do Spotify -, que puxam a média para baixo;
- . quantidade de aplicativos de música;
- . Gráfico BoxPlot com a distribuição de preço dos aplicativos



2ª VERSÃO - UTILIZAÇÃO DO PENTAHO DATA INTEGRATION PARA ETL:

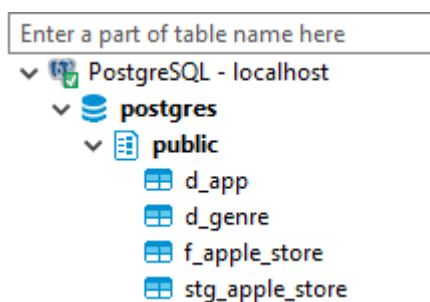
- O processo de ETL para criação do modelo de dados e preparação foi realizado via a ferramenta *Pentaho Data Integration* inclusive com agendamento diário para execução automática do processo e economizar tempo dada a restrição para realização desta tarefa. O log do processo é enviado por e-mail para acompanhamento.



The screenshot shows the Windows Task Scheduler (Agendador de Tarefas) interface. The left pane shows the 'Biblioteca do Agendador' (Task Library). The right pane shows a list of scheduled tasks with columns for 'Nome' (Name), 'Status' (Status), and 'Disparadores' (Triggers).

Nome	Status	Disparadores
ETL-process	Pronto	Todos os dias às 05:00
GoogleUpdateTaskMachineCore	Pronto	Múltiplos disparadores definidos
GoogleUpdateTaskMachineUA	Pronto	Todos os dias às 07:45 - Depois de disp
IntelSURQC-Upgrade-86621605-2a0b-4128-8ffc-...	Pronto	Todos os dias às 11:11
IntelSURQC-Upgrade-86621605-2a0b-4128-8ffc-...	Pronto	Quando qualquer usuário fizer login
IUM-F1E24CA0-B63E-4F13-A9E3-4ADE3BFF3473	Pronto	Todos os dias às 10:53
OneDrive Standalone Update Task-S-1-5-21-257...	Pronto	Às 21:00 em 01/05/1992 - Depois de di
OneDrive Standalone Update Task-S-1-5-21-257...	Pronto	Às 14:00 em 01/05/1992 - Depois de di
teste lopes labs	Pronto	Às 10:15 em 11/05/2020
update-S-1-5-21-257873564-205723252-9963365...	Pronto	Todos os dias às 07:45 - Depois de disp
update-sys	Pronto	Todos os dias às 08:42 - Depois de disp
USER_ESRV_SVC_QUEENCREEK	Pronto	Quando qualquer usuário fizer login -

- As tabelas foram criadas em banco de dados PostgreSQL localhost.

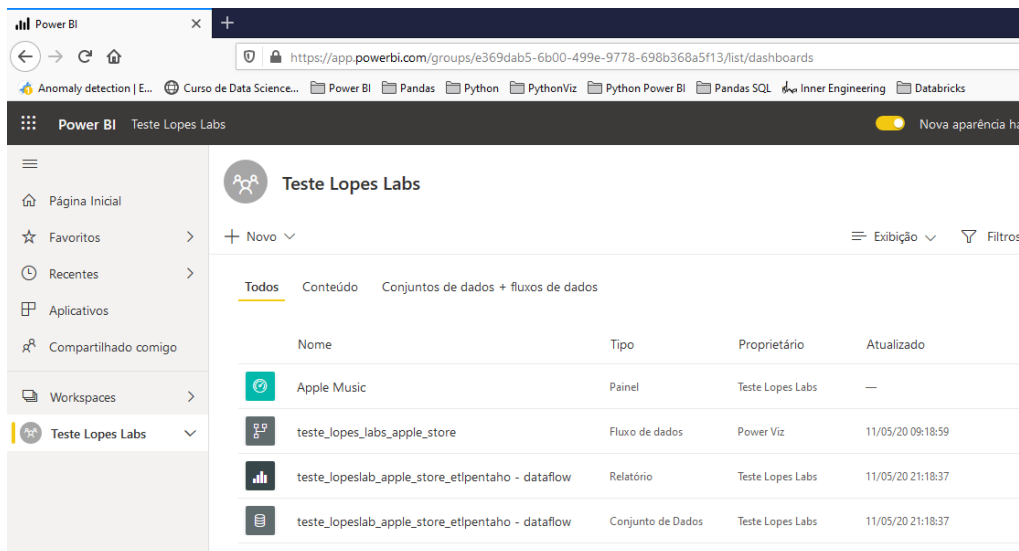


- Os dashboards executados na etapa 1 foram copiados (sem esforço extra) e são idênticos.

- A atualização do Power BI Desktop é feita manualmente

3ª VERSÃO - UTILIZAÇÃO DO DATAFLOW DO POWER BI SERVICES PARA AUTOMATIZAÇÃO DA ATUALIZAÇÃO DOS DADOS DO DASHBOARD E ENVIO DE E-MAIL DIARIAMENTE:

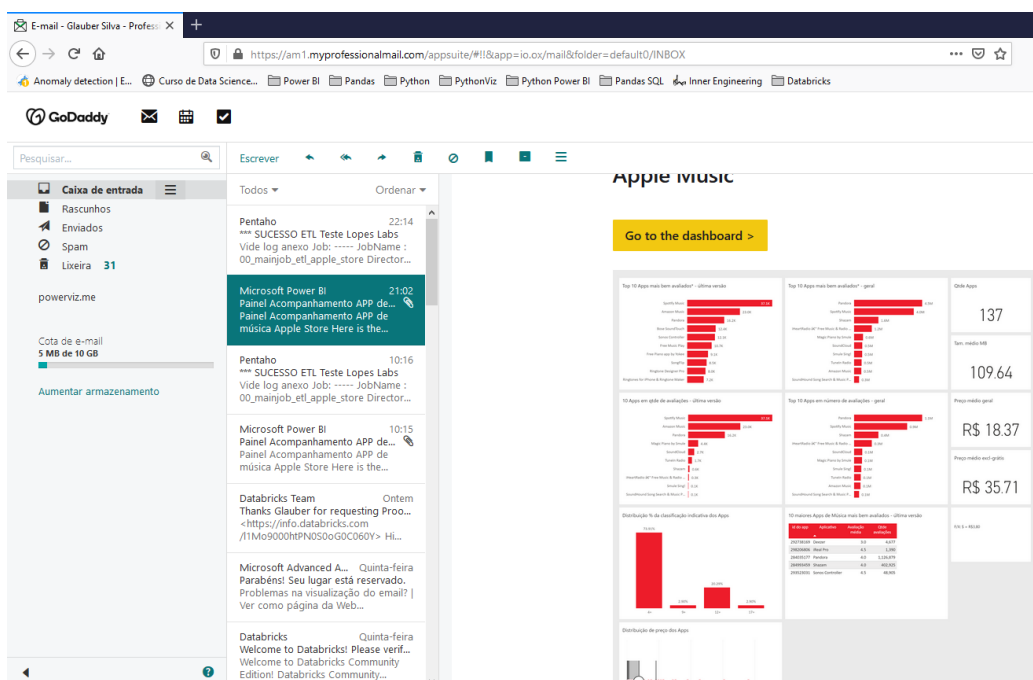
- Utilização do dataflows do Power BI para executar a carga diária automática dos novos dados e horário posterior à atualização do modelo de dados via Pentaho Data Integration.



Nome	Tipo	Proprietário	Atualizado
Apple Music	Painel	Teste Lopes Labs	—
teste_lopes_labs_apple_store	Fluxo de dados	Power Viz	11/05/20 09:18:59
teste_lopeslab_apple_store_etl_pentaho - dataflow	Relatório	Teste Lopes Labs	11/05/20 21:18:37
teste_lopeslab_apple_store_etl_pentaho - dataflow	Conjunto de Dados	Teste Lopes Labs	11/05/20 21:18:37

- O banco de dados PostgreSQL *on premises* é acessado diretamente pelo Power BI Services via *gateway* instalado na máquina.

- Foi gerado um painel contendo as informações do dashboard e agendado para envio diário automático para o e-mail cadastrado. Desta forma todo o processo o dashboard também será enviado automaticamente para a área de negócios em horário pré-determinado.



O dashboard pode ser acessado online via <http://www.bit.ly/testelopeslab>

Avaliação de Conhecimento de SQL

As resoluções estão no arquivo exercicios.sql, que contém o processo passo-a-passo e observações.

Arquivos Anexados

Os arquivos estão disponíveis no arquivo .Zip anexo:

- pasta dashboard:

teste_lopeslab_apple_store_etl_powerquery.pbix

teste_lopeslab_apple_store_etl_pentaho.pbix

teste_lopeslab_apple_store_etlpentaho - dataflow.pbix

- pasta etl:

00_run_etl_teste_lopes_labs.bat

00_mainjob_etl_apple_store.kjb

01_stg_apple_store.ktr

02_dim_genre.ktr

03_dim_app.ktr

04_fato_apple_store.ktr

- pasta sql:

Exercicios.sql