

## Definir processo de ETL

Neste trabalho irá:

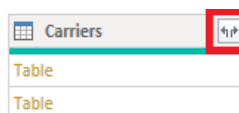
- Criar uma consulta para importar dados;
- Filtrar e transformar dados;
- Intercalar e acrescentar dados;
- Agrupar e agregar dados.

Tarefas a realizar:

1. Os ficheiros associados a este trabalho prático dizem respeito ao registo de voos. Comece por efetuar uma análise rápida dos mesmos com recurso ao Excel ou Notepad.
2. Crie um novo projeto no Power BI Desktop, selecione Nova origem de dados, ficheiros Texto/CSV e selecione o ficheiro FlightPerformance\_2012\_10.csv. Configure o ficheiro como delimitado por virgulas e selecione a opção Transformar Dados da consulta. A consulta realizada apresenta agora dados relativos aos atrasos de voos.
3. No painel de definições da consulta, mude o nome da consulta para FlightDelays.
4. Na lista Passos Aplicados, verifique se o Editor de consulta adicionou automaticamente a primeira linha como cabeçalho. Caso não tenha sido feito automaticamente, faça-o agora clicando em Utilizar a primeira linha como cabeçalhos.
5. Verifique se o Editor de consultas identificou devidamente a coluna FlightDate como sendo do tipo data, assim como as colunas contendo números. Se assim não for, proceda às alterações necessárias, criando um novo passo. Para facilitar esta análise, as colunas de tipo numérico apresentam os valores em itálico alinhados à direita.
6. Caso obtenha erro a alterar o tipo de dados para data, verifique as Definições regionais do ficheiro. Alternativamente pode experimentar criar uma Nova coluna a partir de exemplo.
7. Remova as colunas UniqueCarrier, FlightNum e AirTime.
8. Filtre os dados de forma a que sejam importadas somente as linhas que possuam um atraso de partida do voo superior a 15 minutos (DepDelayMinutes).
9. Repita as etapas 2 a 8 para o ficheiro FlightPerformance\_2012\_11.csv só que, desta vez, atribua o nome FlightDelays2 à consulta. Na remoção das colunas, irá encontrar somente a coluna AirTime para remover.
10. Deverá possuir agora duas Consultas.
11. Com a consulta FlightDelays selecionada, selecione a opção Acrescentar consultas na seção Combinar, da aba Base. Anexe a consulta FlightDelays2 à consulta FlightDelays.
12. Uma vez que os dados constantes na consulta FlightDelays2 integram agora a consulta FlightDelays, desative o carregamento da consulta FlightDelays2.
13. Selecione a coluna OriginCityName e divida-a em duas colunas nomeadas OriginCity e OriginState ("", ""). Faça o mesmo para a coluna DestCityName (DestCity e DestState).
14. Nas colunas OriginCity e DestCity, substitua "West Palm Beach/Palm Beach" por "Palm Beach".
15. Crie uma nova consulta para carregar os dados constantes no ficheiro FlightPerformance\_2012\_12.csv.

16. Mantenha as colunas Carrier, Origin e DepDelay e remova as restantes. De forma a não criar dependência no processo sobre dados desnecessários, utilize a opção Escolher colunas.
17. De forma a excluir os voos com atrasos menos significativos, aplique um filtro que remova todos os registos com atrasos menores ou iguais a 15 minutos (DepDelay).
18. Encontre o atraso médio e máximo, agrupando os atrasos pela origem e pelo transportador, através da opção Agrupar por (ver figura).

19. Nas Definições da consulta, mude o nome da consulta para DelaySummary.
20. Adicione uma consulta chamada Carriers que obtenha o código e a descrição do ficheiro Carriers.csv. Tenha atenção à linha de cabeçalho.
21. Abra a consulta DelaySummary e na aba Base, selecione a opção Intercalar Consultas. Configure a intercalação usando a coluna Carrier da tabela DelaySummary e a coluna Code da tabela Carriers.
22. De seguida, expanda a nova coluna apresentada, clicando sobre as “setas” para adicionar a coluna Description.



23. De forma a criar um modelo mais “user-friendly”, renomeie a coluna Carrier como CarrierCode e a coluna Description como Carrier.
24. Para facilitar a compreensão dos dados, mova a coluna Carrier para junto da coluna CarrierCode.
25. Por fim, dado ser espectável que os registos constantes na consulta Carriers não irão sofrer alterações, desative a opção Incluir na atualização do relatório.
26. Selecione a consulta FlightDelays e Clique em Editor Avançado para analisar o código M criado.
27. Após a análise, na aba Base, clique em Fechar e aplicar para carregar os dados.
28. Verifique se o processo de carregamento dos dados para o Power BI decorreu como esperado e analise as tabelas importadas nas vistas de Dados e Modelo.
29. Quando terminar, salve o trabalho realizado.

O presente documento e os ficheiros associados ao mesmo são propriedade intelectual do autor e partilha a terceiros carece de autorização prévia.