



Relatórios

Agora, chegamos no momento em que a qualidade foi aferida e as transformações necessárias para que os dados sejam úteis, já foram realizadas. Este momento conseguimos responder perguntas importantes para direcionar esforços para melhorar a qualidade dos dados e também responder perguntas de negócio que são essenciais para tomada de decisão.

Perguntas para qualidade

1. Crie uma base de dados unificando os 3 datasets de qualidade (`airports_qa.parquet` , `planes_qa.parquet` , `flights_qa.parquet`) considerando as seguintes chaves em comum: `faa` - `flights_qa.origin` , `airports_qa.faa` - `flights_qa.dest` , `planes_qa.tailnum` - `flights_qa.tailnum`
2. Para cada coluna, calcule as ocorrencias de cada indicador de qualidade: `M` , `F` , `I` , `S` , `T` ?
3. Qual variável apresenta maior número de dados faltantes?
4. Qual variável apresenta maior número de formatos errados?
5. Qual variável apresenta maior número de fora do intervalo?

Perguntas para negócio

1. Crie uma base de dados unificando os 3 datasets transformados (`airports_proc.parquet` , `planes_proc.parquet` , `flights_proc.parquet`) considerando as seguintes chaves em comum: `faa` - `flights.origin` , `airports.faa` - `flights.dest` , `planes.tailnum` - `flights.tailnum`
2. Quantos aeroportos existem em cada `region` ?
3. Qual é a maior diferença de altitude dos voos? Considere: `max(alt_origin - alt_dest)`
4. Qual é o atraso médio?
5. Qual é o atraso médio para cada região?
6. Qual é o atraso acumulado por ano?
7. Qual é o atraso acumulado por ano para cada região?
8. Qual é o tempo de voo médio?
9. Qual é o tempo de voo médio para cada região?
10. Qual é o tempo de voo médio para cada rota (origem e destino)?
11. Qual é o tempo de voo acumulado por ano?
12. Qual é o tempo de voo acumulado para cada região?
13. Qual é a distância de voo médio?
14. Qual é a distância de voo médio para cada região?
15. Qual é a distância de voo médio para cada rota (origem e destino)?
16. Qual é a distância de voo acumulada por ano?
17. Qual é a distância de voo acumulada para cada região?
18. Qual é o número de passageiros médio por rota (origem e destino)?
19. Qual é o número de passageiros acumulado por ano?
20. Qual é o destino mais popular?
21. Qual é o destino recebe mais passageiros?
22. Qual é o destino mais distante de `PDX` ?
23. Qual é o destino tem mais voos por mes?
24. Qual é o modelo de avião mais popular?
25. Qual é o modelo de avião mais popular por destino?
26. Qual é o número médio de motores para cada categoria de `haul_duration` ?
27. Qual é a estação mais popular para voos?
28. Qual é a estação mais popular para voos em cada um dos destinos?
29. Qual é a categoria de atrasos mais comum?
30. Qual é a categoria de atrasos mais comum por rota?