



Relatório Integração de Sistemas de informação

Equipa

No.20350 – Mateus Oliveira

Licenciatura em Engenharia Sistemas Informáticos

3ºano

Barcelos | outubro, 2024

Índice de Figuras

Figura 1 - flow da obtenção dos dados meteorológicos.....	9
Figura 2 - flow da obtenção de dados dos sensores	10
Figura 3 - flow de obtenção de dados dos registos de emergência.....	10
Figura 4 - flow de tratamento de dados e inserção na base de dados.....	11
Figura 5 - flow de tratamento de dados e inserção na base de dados.....	12
Figura 6 - flow de tratamento de dados e inserção na base de dados.....	12
Figura 7 - flow de criação do esquema e tabelas da bds	13
Figura 8 - flow da obtenção de dados através das transformações	13
Figura 9 - flow do tratamento e inserção de dados na bd.....	14
Figura 10 - flow do tratamento e inserção de dados na bd	14
Figura 11 - flow da UI (alertas).....	15
Figura 12 - UI (alertas).....	15
Figura 13 - flow da UI (estatísticas)	16
Figura 14 - UI (estatísticas).....	16
Figura 15 - flow da UI (registos de emergência)	17
Figura 16 - UI (registo de emergência)	17
Figura 17 - QR com video de demonstração	18

Índice

1.	Introdução.....	5
2.	Problema.....	6
2.1.	Desafios Identificados.....	6
3.	Estratégia utilizada.....	7
3.1.	Extração dos Dados (Extract).....	7
3.2.	Transformação dos Dados (Transform)	7
3.3.	Carga dos Dados (Load)	8
3.4.	Ferramentas e Tecnologias	8
4.	Transformações.....	9
4.1.	Dados meteorológicos	9
4.2.	Dados da Qualidade do ar	10
4.3.	Registos de emergências medicas.....	10
4.4.	Dados para emitir alertas.....	11
4.5.	Dados dos registos de emergência causados pela qualidade do ar.....	12
4.6.	Dados das Estatísticas.....	12
5.	Jobs.....	13
5.1.	Criar esquema e tabelas na BD	13
5.2.	Guardar dados em Json	13
5.3.	Inserir dados na base de dados	14
6.	Node red.....	15
6.1.	Flow dos alertas	15
6.2.	Flow das estatísticas	16
6.3.	Flow dos registos de emergências	17
7.	Video com demonstração.....	18
	18
8.	Conclusão.....	19

1. Introdução

Este projeto tem como objetivo implementar um processo ETL (*Extract, Transform, Load*) para extrair, transformar e carregar dados provenientes de sensores urbanos de qualidade do ar, registros de emergências médicas, e dados meteorológicos. O objetivo é integrar essas diferentes fontes de dados para analisar a correlação entre condições atmosféricas e eventos de saúde.

2. Problema

O problema principal que este trabalho aborda é a integração de dados heterogêneos para análise da relação entre a qualidade do ar e a saúde pública. As cidades modernas enfrentam desafios significativos devido à poluição atmosférica, que pode agravar problemas respiratórios e cardíacos, além de reduzir a qualidade de vida da população.

A análise desses dados integrados pode ajudar a identificar correlações entre a qualidade do ar e os impactos na saúde pública, fornecendo informações valiosas para políticas de mitigação da poluição e cuidados com a saúde urbana.

2.1. Desafios Identificados

1. **Fragmentação dos Dados:** Os dados vêm de diferentes fontes (sensores IoT, APIs de clima, registros médicos) e estão em formatos distintos (CSV, JSON). Isso requer um processo robusto de extração e padronização.
2. **Inconsistência e Ruído nos Dados:** Sensores de qualidade do ar podem apresentar erros de leitura ou falhas de status, como dados incompletos ou marcados como "erro", o que demanda uma fase de limpeza de dados para remover leituras incorretas.
3. **Correlações Complexas:** A identificação de padrões entre a qualidade do ar, fatores climáticos e emergências médicas não é direta, exigindo o uso de técnicas de análise para encontrar relações significativas entre as variáveis.
4. **Tempo Real e Previsão:** Além dos dados históricos, é necessário prever como as condições climáticas futuras podem impactar a dispersão de poluentes e, consequentemente, a saúde pública, o que demanda a integração de dados de previsão do tempo.

Este trabalho visa resolver esses problemas implementando um fluxo de ETL que facilita a integração e análise de dados em múltiplas dimensões, com o objetivo final de fornecer insights para a gestão da qualidade do ar e a saúde pública.

3. Estratégia utilizada

A estratégia utilizada para abordar o problema da integração e análise dos dados de qualidade do ar, clima e emergências médicas baseou-se na implementação de um fluxo ETL robusto. O processo foi dividido nas seguintes etapas principais:

3.1. Extração dos Dados (Extract)

Para realizar a coleta dos dados necessários, foi utilizado o Pentaho Kettle (PDI), uma ferramenta de ETL que permite a automação do processo de integração de dados. A extração dos dados envolveu:

- **Sensores de Qualidade do Ar:** Extração dos dados fornecidos pelos sensores IoT em formato CSV. Os sensores coletam informações sobre poluentes (PM2.5, PM10, CO, NO2, SO2, O3), além de variáveis como temperatura e umidade.
- **Dados Meteorológicos:** Utilização de APIs como a Open Meteo API para obter dados de previsão do tempo, como temperatura, velocidade do vento e humidade, os quais influenciam a dispersão de poluentes.
- **Registos de Emergências Médicas:** Importação de dados JSON contendo informações sobre condições médicas críticas (como ataques cardíacos e problemas respiratórios) relacionadas com a qualidade do ar e as condições climáticas. Os dados incluíam informações como a gravidade, data e hora, local, e a condição médica dos pacientes.

3.2. Transformação dos Dados (Transform)

Após a extração, foi necessário transformar e padronizar os dados para torná-los utilizáveis e integrados. As transformações incluíram:

- **Combinação de Dados:** Foi utilizado o passo "Merge Join" do Pentaho Kettle para combinar os dados das diferentes fontes, associando-os por localização e timestamp. Isso permitiu a correlação entre os dados de qualidade do ar, previsão meteorológica e emergências médicas.
- **Limpeza de Dados:** Utilizando o passo "Filter Rows", foi aplicada uma filtragem para remover dados inválidos ou irrelevantes, como entradas de sensores com status "error" ou com leituras inconsistentes.
- **Normalização de Dados:** Para garantir a uniformidade dos dados, foram usados passos como "Select Values" e "String Operations" para ajustar os formatos de tempo, conversão de unidades de temperatura e garantir a consistência nos campos de texto (localização, ID dos sensores, etc.).
- **Análise Temporal e Geográfica:** Para realizar análises por localização e por intervalos de tempo, os dados foram organizados em blocos temporais e agregados por áreas específicas (bairros ou zonas urbanas). Foi usado o passo "Group By" para calcular médias e tendências (por exemplo, a média diária de poluentes ou temperatura).

3.3. Carga dos Dados (Load)

Para facilitar o uso dos dados em análises posteriores e garantir a interoperabilidade entre diferentes sistemas, os dados finais foram armazenados numa base de dados, usando o step "Insert / Update" do Pentaho Kettle.

3.4. Ferramentas e Tecnologias

Para realizar essas tarefas, foram utilizadas as seguintes ferramentas e tecnologias:

- **Pentaho Kettle (PDI)**: Ferramenta principal para construção do fluxo ETL, permitindo a automação e integração de dados.
- **APIs (Open Meteo)**: Para obter dados meteorológicos em tempo real.
- **Sensores IoT**: Fonte dos dados de qualidade do ar.
- **JSON/CSV**: Formatos de dados utilizados para entrada e saída dos dados processados.
- **Node-red**: Ferramenta principal para a construção da UI do fluxo ETL.

Esta estratégia garantiu uma abordagem eficaz para a integração e análise dos dados, possibilitando a obtenção de insights relevantes sobre o impacto da qualidade do ar na saúde pública e a previsão de possíveis cenários críticos.

4. Transformações

4.1. Dados meteorológicos

Nesta transformação, através do uso de uma API coeto os dados como a temperatura, sensação térmica e humidade para guardar num ficheiro json e num outro ficheiro json guardo os dados das previsões dos próximos dias, sendo que uso o step Select values para escolher os dados que mais se adequam à solução.

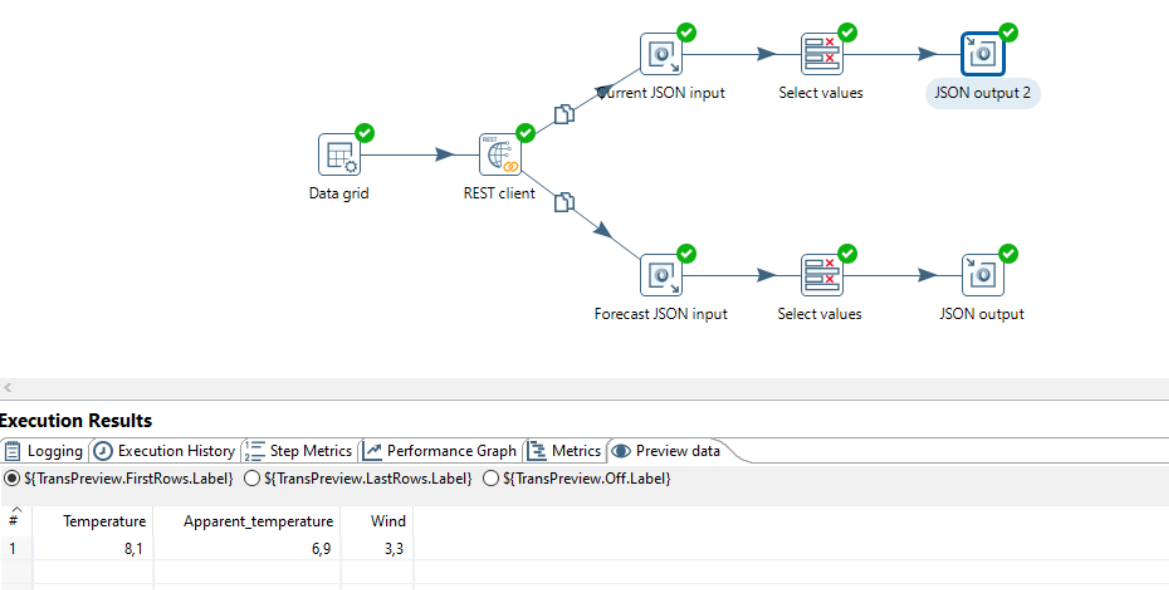


Figura 1 - flow da obtenção dos dados meteorológicos

4.2. Dados da Qualidade do ar

Para extrair os dados da qualidade do ar, através do uso de um ficheiro CSV que contem vários dados de vários sensores de qualidade de ar, após escolher os dados que mais se adequam, guardo os mesmos num ficheiro json também.

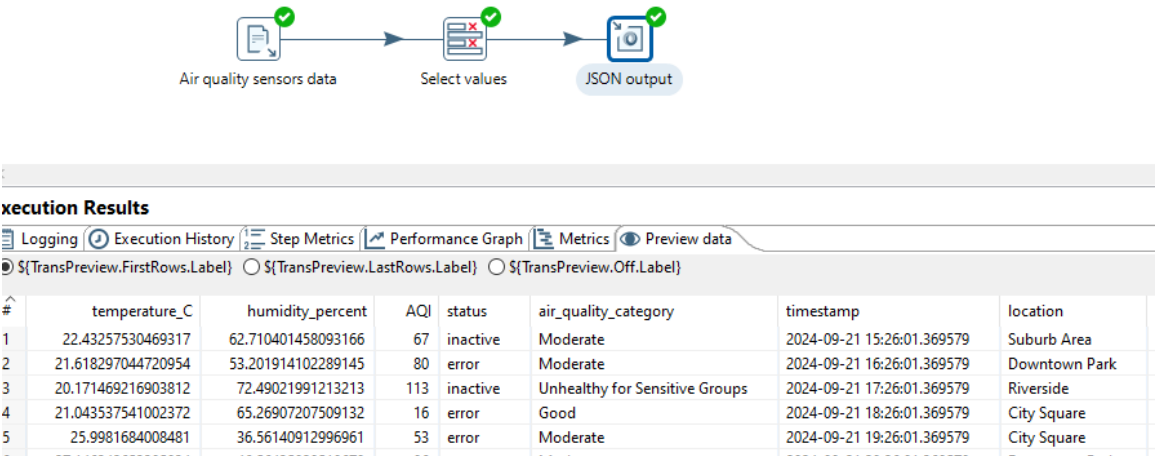


Figura 2 - flow da obtenção de dados dos sensores

4.3. Registos de emergências medicas

Após aceder a um ficheiro XML que contem os registos de emergências medicas do hospital da cidade, são selecionados todos os campos para serem guardados num outro ficheiro json.

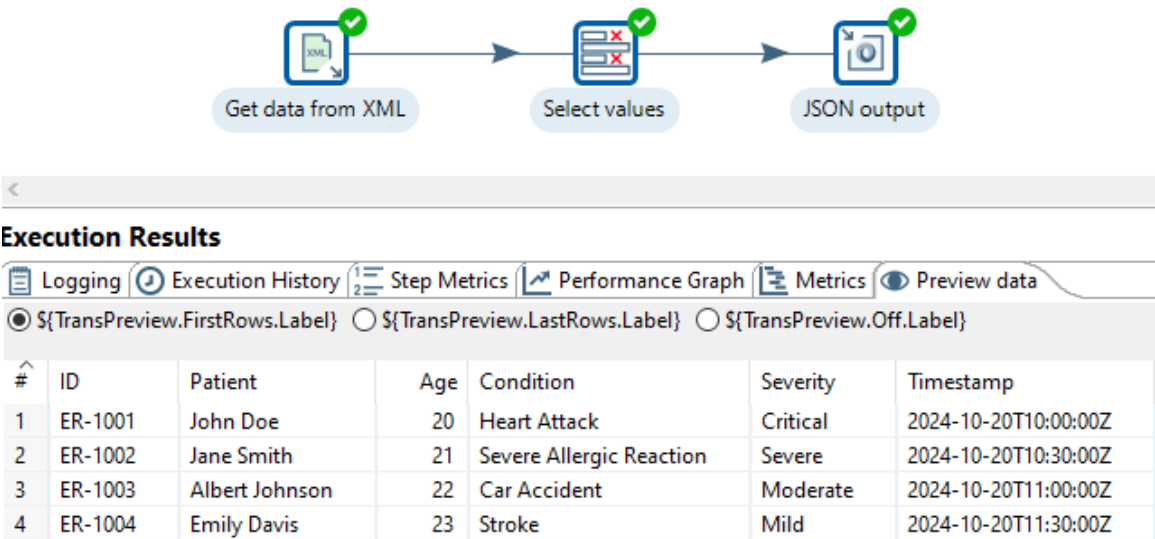


Figura 3 - flow de obtenção de dados dos registos de emergência

4.4. Dados para emitir alertas

Para que as autoridades possam emitir alertas para prevenir problemas de saúde na população devido à qualidade do ar, temperatura entre outros, primeiro acedo ao ficheiro da qualidade do ar em json previamente referido, tendo os dados organizo os mesmos por data e filtro apenas pelos sensores que estão ativos, logo de seguida utilizo uma expressão regular (regEx) para remover as horas da data, de seguida organizo outra vez mas por localização agrupando os mesmos para fazer a média dos registos dos sensores para cada zona e junto com os dados do ficheiro json da meteorologia e por fim guardo tudo numa base de dados.

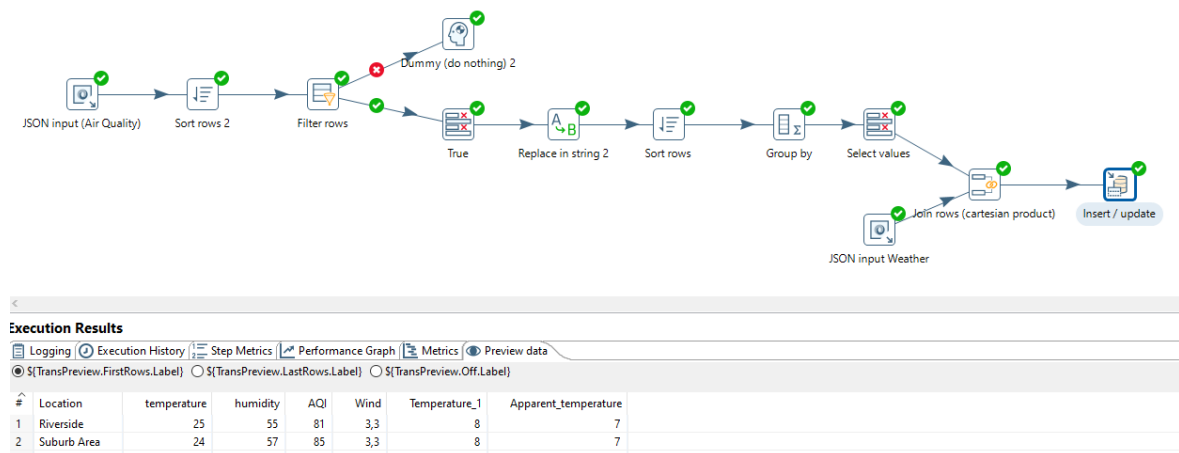


Figura 4 - flow de tratamento de dados e inserção na base de dados

4.5. Dados dos registos de emergência causados pela qualidade do ar

Para ter os registos de emergência relevantes para inserir na base de dados para futura análise, extrai mais uma vez o ficheiro json com os dados da qualidade do ar organizei por data, filtrei a condição dos sensores para mostrar apenas os ativos e também pelo AQI que fosse superior a 100, formatei a data retirando as horas da mesma, e juntei aos dados dos relatórios de emergência aos quais foram agrupados também por data e filtrados pelas condições dos pacientes possivelmente causadas pela condição do ar, por fim organizei pelos ID's dos relatórios e removi informação duplicada.

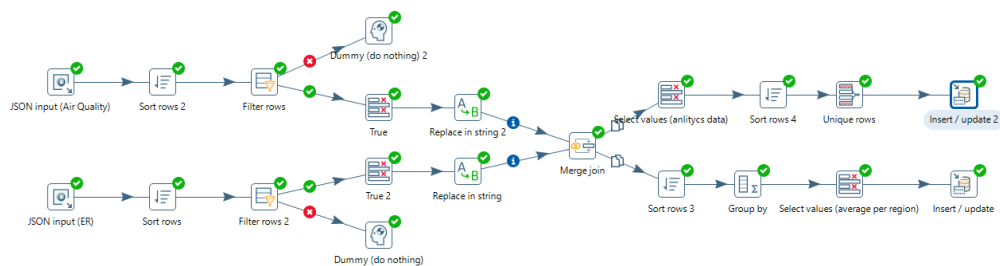


Figura 5 - flow de tratamento de dados e inserção na base de dados

4.6. Dados das Estatísticas

Apos a junção dos dados referidos no 4.5, organizei a copia dos mesmos por localização fazendo, também, a média da temperatura, humidade e do índice de qualidade do ar para as autoridades terem uma estimativa do risco para a saúde publica em cada região.

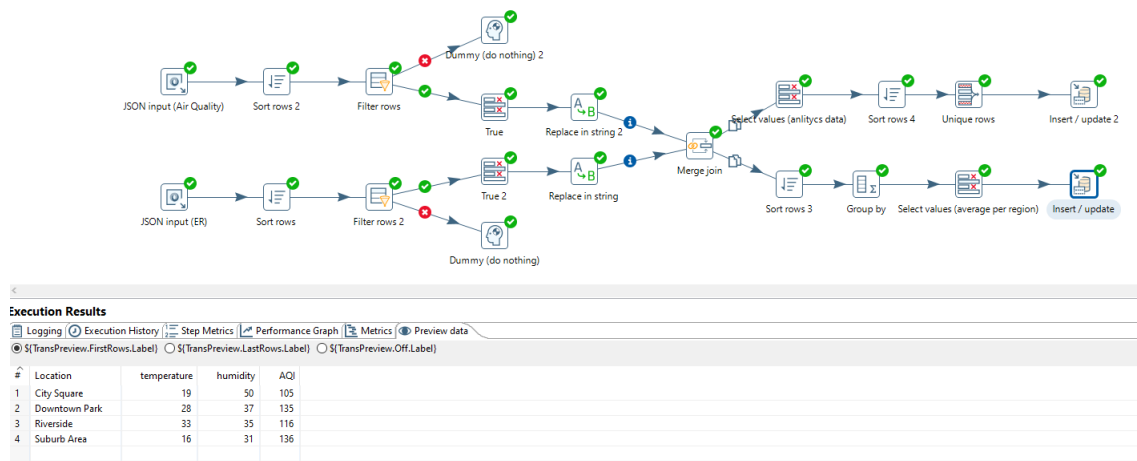


Figura 6 - flow de tratamento de dados e inserção na base de dados

5. Jobs

5.1. Criar esquema e tabelas na BD

Este Job cria um esquema e as tabelas necessárias para todos os dados analisados por meio de scripts de SQL.



Figura 7 - flow de criação do esquema e tabelas da bds

5.2. Guardar dados em Json

Foi criado um Job que utiliza as transformações previamente mencionadas (4.1, 4.2 e 4.3) para criar os ficheiros Json necessários ao mesmo tempo.

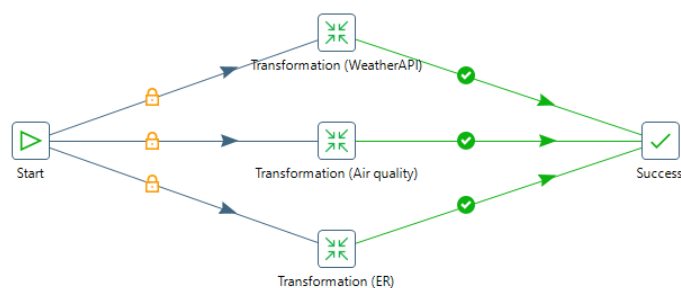


Figura 8 - flow da obtenção de dados através das transformações

5.3. Inserir dados na base de dados

Nestes Jobs correm as transformações mencionadas no 4.4, 4.5 e 4.6 para inserir os dados na base de dados.

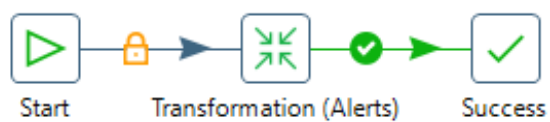


Figura 9 - flow do tratamento e inserção de dados na bd

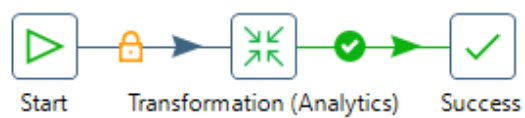


Figura 10 - flow do tratamento e inserção de dados na bd

6. Node red

Foram criados *flows* em *node red* para a visualização por parte das autoridades e outros interessados, de todos os dados obtidos em tempo real.

6.1. Flow dos alertas

Neste flow foi criada uma UI para que as autoridades conseguissem visualizar dados como a temperatura, humidade e a qualidade do ar atual para emitir alertas e também uma lista de localizações mais específicas em estado de alerta a ser emitido também pelas mesmas.

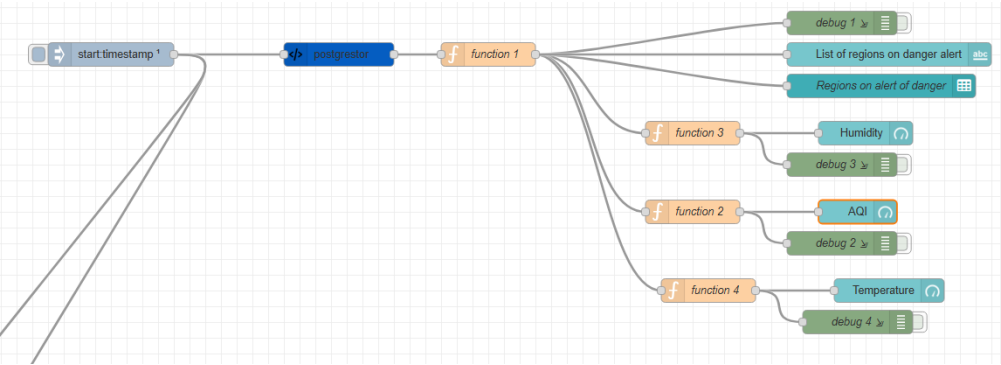


Figura 11 - flow da UI (alertas)

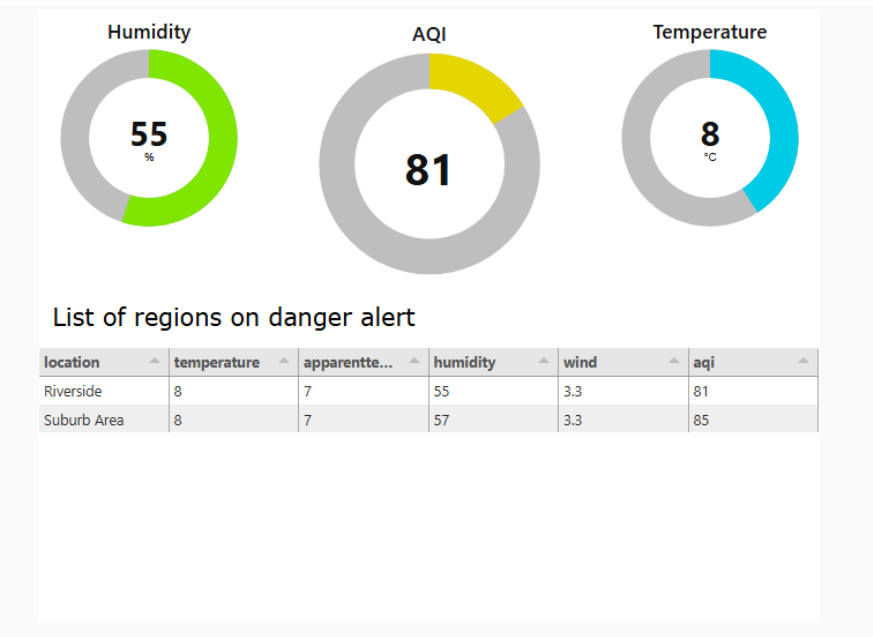


Figura 12 - UI (alertas)

6.2. Flow das estatísticas

Para as estatísticas foram criados gráficos de barras para mostrar a média de temperatura, qualidade de ar e humidade por localização de forma a alertar as autoridades das zonas mais prováveis a ter problemas de saúde social em relação as métricas analisadas ao longo do tempo.

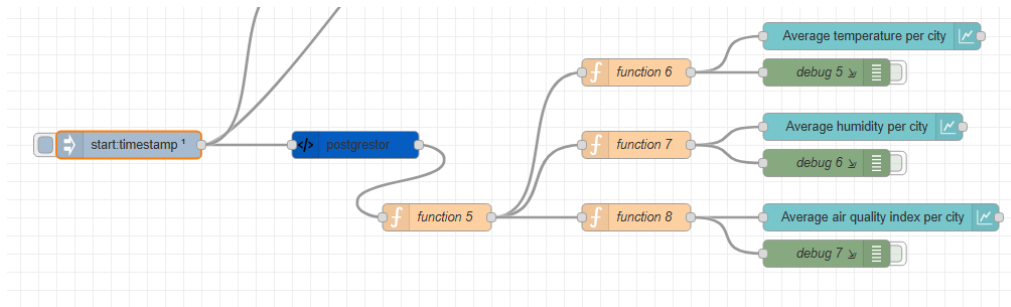


Figura 13 - flow da UI (estatísticas)



Figura 14 - UI (estatísticas)

6.3. Flow dos registos de emergências

Por fim temos o flow dos registos de emergência onde as autoridades consultam casos de cidadãos afetados pela qualidade do ar e ou temperatura de forma a poderem solucionar melhor futuras ocasiões analisando devidamente a informação disponibilizada.

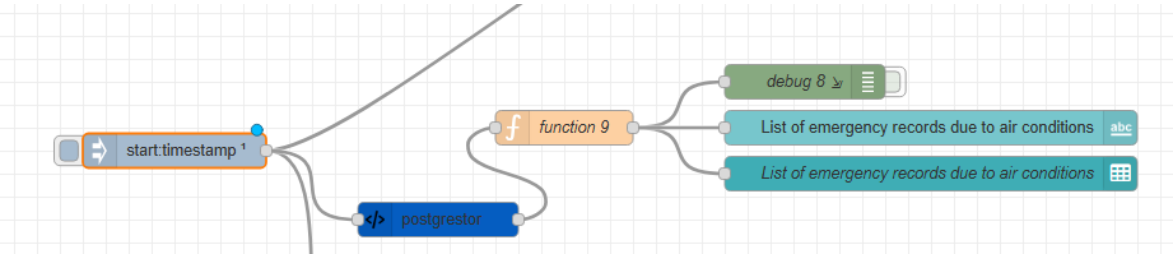


Figura 15 - flow da UI (registos de emergência)

List of emergency records due to air conditions							
id	date	condition	severity	temperature	humidity	aqi	aqs
ER-1002	2024-10-20	Severe Allergic Reaction	Severe	16	31	136	Unhealthy for Sensitive Gr...
ER-1006	2024-10-20	Asthma Attack	Severe	16	31	136	Unhealthy for Sensitive Gr...
ER-1022	2024-10-20	Severe Allergic Reaction	Severe	16	31	136	Unhealthy for Sensitive Gr...
ER-1026	2024-10-20	Asthma Attack	Severe	16	31	136	Unhealthy for Sensitive Gr...
ER-1042	2024-10-21	Severe Allergic Reaction	Severe	30	33	133	Unhealthy for Sensitive Gr...
ER-1046	2024-10-21	Asthma Attack	Severe	30	33	133	Unhealthy for Sensitive Gr...
ER-1062	2024-10-21	Severe Allergic Reaction	Severe	30	33	133	Unhealthy for Sensitive Gr...
ER-1066	2024-10-21	Asthma Attack	Severe	30	33	133	Unhealthy for Sensitive Gr...

Figura 16 - UI (registo de emergência)

7. Video com demonstração

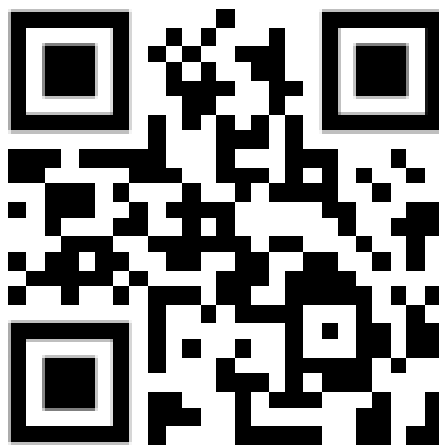


Figura 17 - QR com video de demonstração

8. Conclusão

Neste projeto, foi desenvolvida uma solução integrada para monitorar a qualidade do ar urbano, relacionando com dados meteorológicos e registros de emergências médicas. Utilizando o Pentaho Kettle, foi possível extrair, transformar e consolidar dados de diferentes fontes, fornecendo uma visão detalhada do impacto das condições ambientais na saúde pública. A análise identificou padrões que mostram como a qualidade do ar, influenciada por fatores meteorológicos, pode aumentar as emergências médicas em algumas áreas. A solução criada permite o monitoramento em tempo real da qualidade do ar e fornece informações cruciais para tomada de decisões por autoridades de saúde e gestores ambientais. A integração de dados IoT e meteorológicos demonstrou ser fundamental para prever e mitigar riscos à saúde pública. A utilização do Pentaho Kettle simplificou o processo de ETL e garantiu a confiabilidade dos dados transformados, tornando-se um modelo eficiente para futuros projetos de análise de dados ambientais e de saúde.