

**Análise das variações na qualidade do ar em Pequim**

**Segunda-feira, 25 de novembro de 2019**

**Análise das variações na qualidade do ar em Pequim**

**Elaborado por:**

**Bernardo Góes Barreto**

Índice

[Introdução 4](#_Toc25580211)

[Desenvolvimento 5](#_Toc25580212)

[Python 5](#_Toc25580213)

[Power BI 6](#_Toc25580214)

[Conclusão 10](#_Toc25580215)

[Bibliography 11](#_Toc25580216)

# Introdução

Alguns anos atrás, Pequim estava à beira de se tornar inabitável para os seres humanos, carregado o fardo dos piores índices de poluição da história nacional, chegando a obter taxas superando 45 vezes o limite diário de poluentes recomendado pela Organização Mundial da Saúde (OMS). Com concentrações excessivas de PM2.5 e PM10, foram traçadas metas para redução de emissão desses poluentes.

O site oficial de estatística de qualidade do ar reportou que Pequim havia obtido uma redução de 9.67% e 8.68% na concentração de PM2.5 e PM10 em 2016, essa informação trouxe certa tranquilidade para os habitantes da capital.

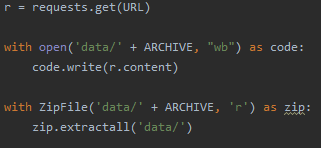
Nesse estudo, foram realizadas algumas análises estatísticas em cima desses poluentes, com dados de 12 centros de monitoramento em Pequim obtidos em quatro anos, tendo como objetivo prover um pleno entendimento sobre o comportamento dos poluentes PM2.5 e PM10, revelando que o avanço apontado em 2016 não pode ser comprovado, e que essas melhorias foram certamente duvidosas. Para esse estudo, foi utilizado a linguagem de programação Python conjuntamente com a ferramenta de *Business Intelligence*, Microsoft Power BI.

# Desenvolvimento

## PythonC:\Users\ber_b\Desktop\python-logo@2x.png

A escolha de utilizar a linguagem Python como uma das maneiras para analisar os dados se dá por ser uma ferramenta fácil de usar quando se trata de computação quantitativa e analítica, sendo uma linguagem flexível e possuindo enormes bibliotecas usadas para manipulação de dados, contando também com a habilidade de ser altamente escalável fornecendo meios de resolver problemas que não podem ser resolvidos usando outras linguagens de programação. Aplicando os conceitos de ETL, essa ferramenta será utilizada para realizar os processos de *EXTRACT,* e as fases de *TRANSORM* e *LOAD* serão cobertas com o uso de outra ferramenta, o PowerBI.

Para a extração dos dados da base "Beijing Multi-Site Air-Quality”, foi utilizado a biblioteca “*requests*” e “*zipfile*”, realizando o *request* no servidor, que retorna um arquivo .zip com todos os dados para serem extraídos a partir da funcionalidade importada, “*ZipFile*”.



Após a extração do conteúdo do .zip, são fornecidos 12 arquivos .csv, contendo os dados respectivos das bases de monitoramento de Pequim. Como as colunas são idênticas, para uma fácil análise, foi utilizado a biblioteca “glob” para a união dos dados contidos nos arquivos .csv em um único *dataframe*, que é tratado pela biblioteca “pandas”.



Finalmente, a partir do *dataframe* criado, é gerado um novo arquivo .csv para facilitar o tratamento dos dados com o PowerBI.

## C:\Users\ber_b\Desktop\PowerBI-e1557666264791.jpgPower BI

Para realizar os *steps* de *TRANSFORM* and *LOAD*, trouxe a solução através da ferramenta da Microsoft. Todos os dados, em qualquer lugar e a qualquer tempo. Esse é o slogan oficial do Power BI no seu site oficial. Uma ferramenta incrível para análise de dados, fornecendo a chance de criar visualizações e dashboards de alto nível a partir da análise de dados e tudo isso de uma maneira muito simples, por esses e outros motivos o Power BI foi escolhido como uma das maneiras para realizar o estudo.

Como fonte de dados, foi utilizado o arquivo .csv gerado pelo script em Python previamente apresentado, realizando a união de todos os *datasets*, uma vez que possuem as mesmas colunas de informação. A mudança de região dentro do PowerBI é uma etapa essencial, pois a detecção dos separadores de unidade “,” e “.” devem ser feitas de forma correta.

Com os dados importados, foi feito o tratamento dentro do *Power Query Editor*, foi transformado a primeira linha em cabeçalho, e logo em seguida foram alterados os tipos de dados para uma análise acurada dos registros, como mostra a imagem à seguir. Também foi retirado uma das colunas de index, pois haviam duas anteriormente.





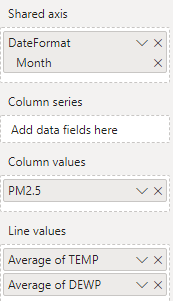




Para o campo “hour”, foi feito um tratamento um pouco diferente, pois a ferramenta não reconhece inteiros com um digito (Ex: “9”) como horário, logo, foi utilizado a funcionalidade de adicionar um prefixo “0” e, finalmente, a ferramenta de “trim”, para retirar o digito extra “0” dos registros que já haviam dois dígitos ( Ex: “010” - > “10” ).

Para facilitar o processo de análise dos dados, foi criado uma coluna auxiliar a partir de uma fórmula DAX para fazer a transformação do tipo de dado e a junção das colunas de data e hora.

DateFormat= DATE('allFiles'[year];'allFiles'[month];'allFiles'[day])+'allFiles'[hour]



No gráfico à seguir, podemos ver uma das variáveis mais

dominantes que influenciam o PM2.5 e o PM10, o DEWP

(Ponto de orvalho). Observa-se também que a de acordo com a

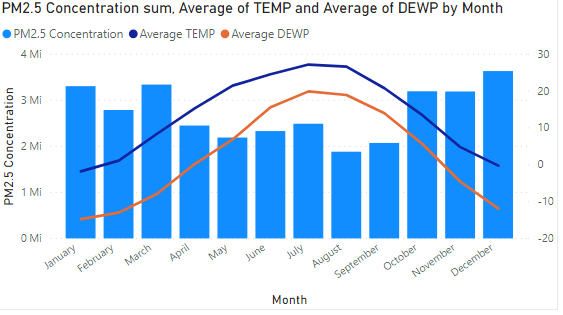
temperatura, as concentrações dos gases refletem um grau de

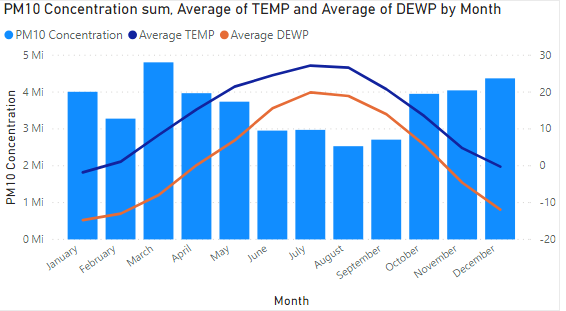
correlação negativa alta também. Os gráficos foram gerados com

os meses como variável de tempo, soma da concentração de gases

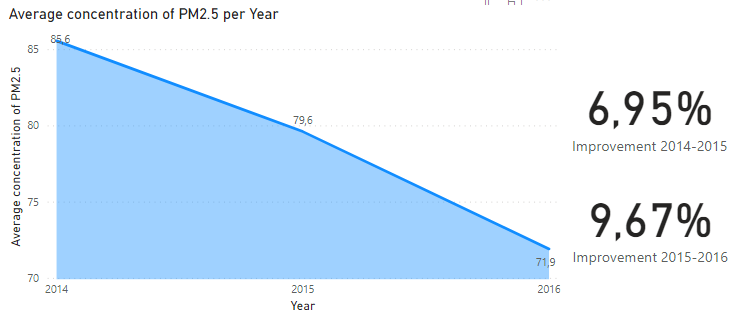
como valores de coluna, e linhas médias para temperatura e ponto

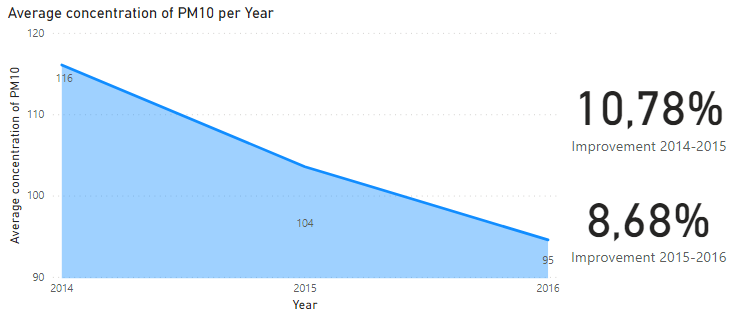
de orvalho.



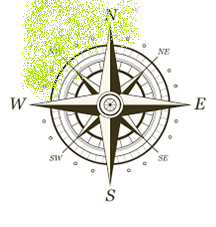


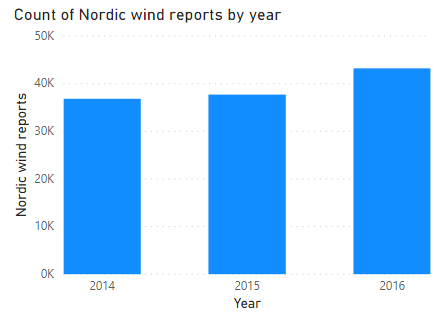
Apresenta-se no próximo gráfico, a média de concentração por ano dos poluentes, juntamente com a melhoria em porcentagem.



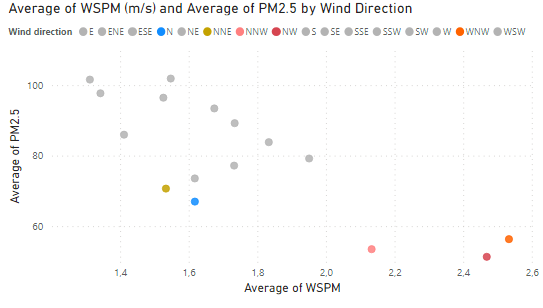


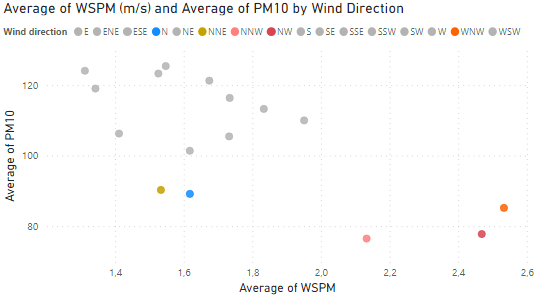
No período de 2014-2015, podemos observar uma melhora real na diminuição da a emissão dos gases, mas, apesar de os gráficos indicarem que houve uma melhoria entre 2015-2016, com o gráfico a seguir, podemos analisar que devido à alta quantidade de ventos oeste- nórdicos, um fator chave para a dispersão, essa análise não pode ser confirmada.

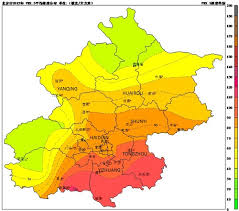




Podemos ver a seguir a forte relação entre a direção dos ventos e a concentração dos gases, sendo esse ser um fator que distorce as conclusões. Um quesito importante a ser observado também, é a força desses ventos, que ajudam ainda mais na dispersão dos gases em observação. O gráfico de dispersão nos traz uma clara correlação dos ventos e da velocidade deles (WSPM em m/s), com a média de emissão obtida.







Mapa explicativo sobre as áreas de concentração

de emissões de gases poluentes em Pequim, demonstrando

que, os ventos sem dúvida estão adulterando os resultados

de concetração médias dos gases poluentes. Sendo necessário

uma nova medida a ser tomada tanto para análise quanto

para a mitigação da alta emissão desses poluentes.

# Conclusão

A análise não pode confirmar uma melhoria significante em 2016 que foi apontada em 9,67% e 8,68% para os gases PM2.5 e PM10 respectivamente, pois devido à alta porcentagem de ventos nórdicos que favorecem a dispersão dos gases. Existem inúmeras industrias de ferro, aço e cimento no Sul concentrando a emissão de gases na área, logo, ventos direcionados para o norte, ajudam a minimizar a concentração no Sul, porém, sem reduzir a emissão. Já para 2015, com os dados obtidos, podemos mostrar que realmente houve uma redução na emissão. Podemos concluir que existe a necessidade de um novo método de planejamento para a redução dos poluentes, juntamente com novas métricas para análise dos dados de qualidade atualmente utilizadas pelo governo da China.

# Bibliography

*https://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data*