Caratula (Por si alguien lo hace)

**Proyecto de pronóstico de la contaminación del aire en Quito**

**Objetivos**

**Objetivo 1:** Procesar y consolidar datos de series temporales multi-estacionales en un conjunto de datos unificados, gestionando los valores faltantes y las funciones de ingeniería para facilitar la predicción precisa de PM2.5.

**Objetivo 2:** Desarrollar y comparar series temporales y modelos de aprendizaje automático para pronosticar los niveles de PM2.5 en múltiples horizontes ( 5 a 10 años y 10 días ), proporcionando información práctica mediante una aplicación Streamlit.

**Instrucciones:**

Entregar un Jupyter Notebook (.ipynb), su versión HTML exportada (.html) y un informe completo (máximo 10 páginas, PDF) como parte de un grupo de hasta 5 miembros. Nombrar los archivos con las primeras letras de los nombres o apellidos de los miembros del grupo (p. ej., ABCD.ipynb). Pronosticar los niveles de PM2.5 en Quito utilizando el conjunto de datos proporcionado de 2004 a 2017, recopilado en 10 estaciones meteorológicas (p. ej., Belisario, Carapungo). Entregar todos los entregables antes de la semana 16 (12 de marzo de 2025, ajustado al contexto actual), lo que representa el 40 % de la calificación del curso (30 % informe, 10 % calidad del código). Documentar cada paso con código, visualizaciones y un informe detallado siguiendo la metodología analítica descrita.

**Definición del problema:**

El objetivo principal es pronosticar los niveles de PM2.5 en Quipo para ayudar en la gestión de calidad del aire, protegiendo la salud publica y el medio ambiente. Esto es fundamental, dado que los altos niveles de PM2.5 pueden tener serios impactos en la salud de los ciudadanos y en el ecosistema urbano. EL proyecto se enfoca en:

1. Procesar y consolidar datos multi-estacionales para crear un conjunto de datos unificado.
2. Proveer predicciones precisas a diferentes escalas temporales (5-10 años y 10 días).
3. Ofrecer estas predicciones mediante una herramienta práctica (Streamlit) accesible para las partes interesadas.

**Partes interesadas:**

1. **Responsables políticos y autoridades locales (Alcaldías, gobierno municipal):** Usarán las predicciones para tomar decisiones informadas sobre políticas de calidad de aire, asignación de recursos y mitigación de riesgos.
2. **Ciudadanos de Quito:** Beneficiarios directos de las decisiones basadas en los pronósticos, quienes tendrán acceso a un mejor ambiente y medidas preventivas.
3. **Organismos de salid publica:** pueden emplear los datos para evaluar riesgos y diseñar campañas preventivas contra enfermedades respiratorias causadas por la contaminación
4. **Comunidad científica:** Investigadores interesados en modelar la contaminación y validar enfoques predictivos similares.
5. **Empresas y sectores industriales:** Aquellas responsables de emisiones pueden ajustar sus prácticas para cumplir con regulaciones.

**Limitaciones:**

1. **Lagunas de datos:** Faltantes en los datos históricos entre 2004 y 2017, lo que podría requerir técnicas de imputación (como interpolación o modelos avanzados).
2. **Variabilidad multi-estacional:** Factores como estaciones secas y lluviosas pueden alterar drásticamente los niveles de PM2.5, complicando la predicción.
3. **Horizontes de predicción:** La naturaleza de predicción a largo plazo (5-10 años) tiene mayor incertidumbre frente al corto plazo (10 días).
4. **Capacidad computacional:** Procesar grandes volúmenes de datos meteorológicos y ejecutar modelos avanzados podrían demandar recursos significativos.
5. **Representatividad espacial:** Diez estaciones meteorológicas pueden no captar variaciones especificas de ciertos sectores de Quito.

**Carga de datos:**

Para la primera acción a realizar en la parte de codificación es realizar la carga de datos, que esta en varios CSV, el cual se encuentra en [One Drive](https://udlaec-my.sharepoint.com/personal/mario_gonzalez_rodriguez_udla_edu_ec/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fmario%5Fgonzalez%5Frodriguez%5Fudla%5Fedu%5Fec%2FDocuments%2Fdata%5FAI%5F202420&ga=1), donde se encuentran las siguientes estaciones: **Belisario, Carapungo, Centro, Cotocollao, El Camal, Guamani, Jipijapa, Los Chillos, San Antonio, Tumbaco** los cuales juntos conforman 10 estaciones que se registran en la datase.

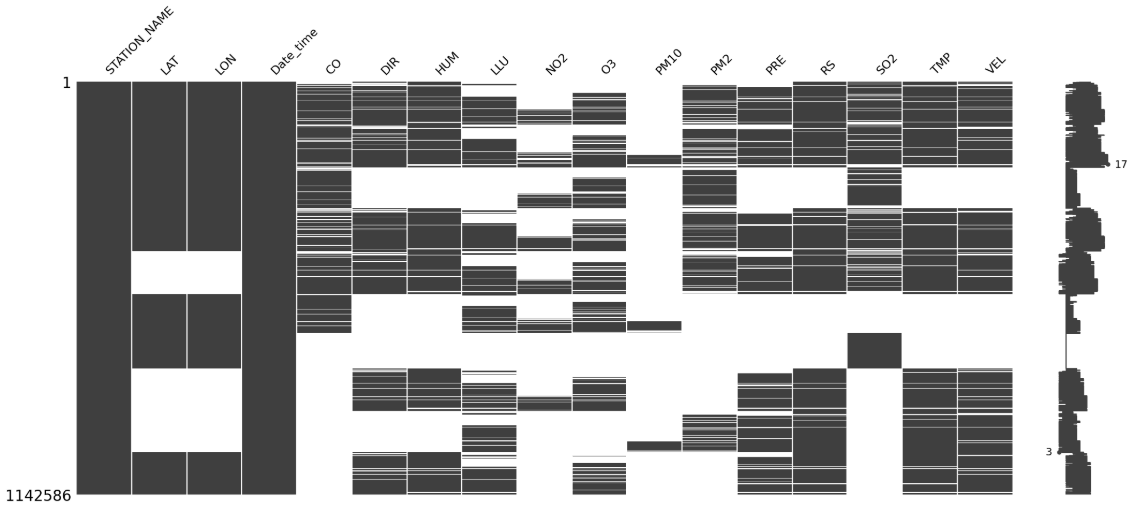
1. Primero lo que se hizo es la maquetar lo que vamos a mostrar en nuestra tabla unificada de datos:



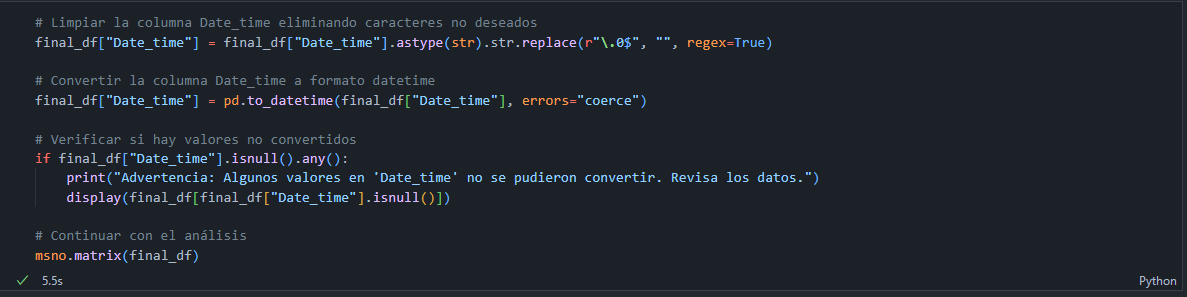
1. Luego iteramos en cada uno de los archivos para poder leerlos y maquetarlos con respecto a lo que necesitamos, lo único es que llegamos a ignorar en el filtro a las estaciones ya que queremos que ahí, solo nos devuelva la latitud y la longitud del nombre de la estación



* 1. Aquí podremos ver como hace falta la data utilizando la librería de python missing, y se llega a visualizar así:



1. Hacemos una limpieza de datos, como la fecha y los datos que no queremos en nuestras columnas, en ello se lleva a unificar solo un formato



1. Por último, hacemos una media mensual y una por horas para poder sacar las predicciones a futuro

