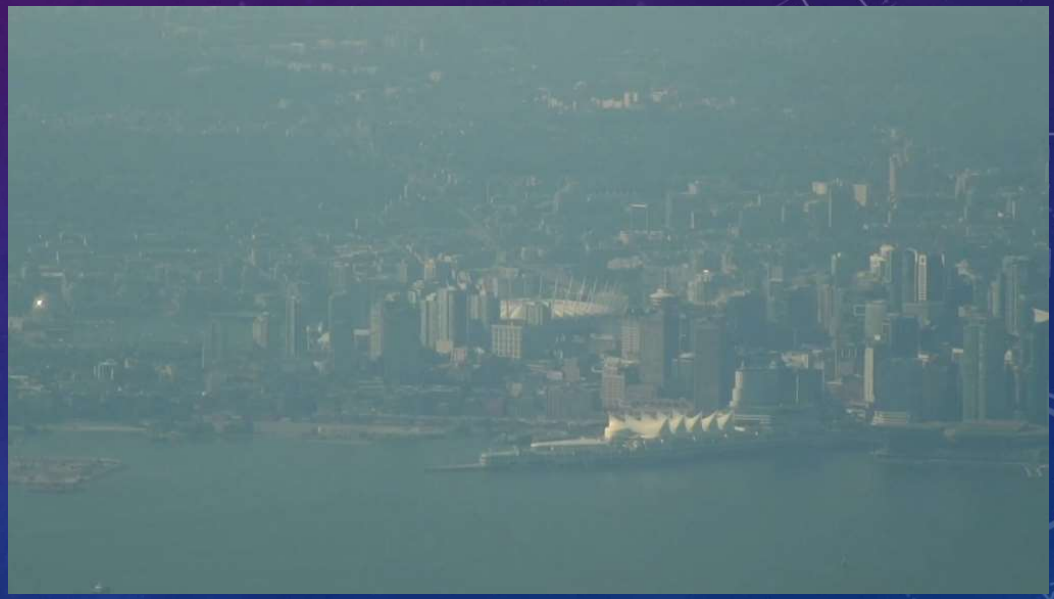
CAPSTONE SPRINT 1

USING DATA SCIENCE TO ANSWER THE QUESTION:

WHY IS IT SO HOT!?

CAN WE USE DATA SCIENCE TO IMPROVE ON THE ABILITY TO PREDICT HOT SUMMER DAYS.

- On June 29, 2021, between 3pm and 4pm Vancouver reached their all-time high temperature of 32.1Cº.
- Summers in 2021, 2022, and 2023 have been reported as being hotter than usual across the globe.



https://bc.ctvnews.ca/metro-vancouver-air-quality-advisory-prompted-by-heat-wave-now-over-1.5492660

USE SUPERVISED MACHINE LEARNING TO TRAIN A MODEL TO RECOGNIZE THE PATTERNS IN WEATHER STATS LEADING TO HIGH TEMPERATURES

- Do specific variables stand out before or during a hot day?
- Are there any indicators of an incoming rise in temperature?

If successful, this could help cities prepare their population for unexpected temperature spikes. This could in turn help save lives by helping reduce the risk of heat stroke.

Inspiration for this project:

Columbia University Data Science Institute – Project Using Climate Data for the Prediction of Heavy Snowfall

DATASET: VANCOUVER HOURLY WEATHER STATS OBTAINED FROM VANCOUVER.WEATHERSTATS.CA

Data:

- Variables include temperature, pressure, wind, humidity, visibility, health Index, and cloud cover.
- Data had to be recalculated for some variables and interpolated for others.

Preliminary EDA:

- Patterns show that on average, weekly or monthly, the summers have not been getting hotter.
- However, daily averages tell otherwise.

NEXT STEPS

1. Include additional hourly data: July 2023 and possibly data from further back than June 2013.
2. Further analysis of variable correlations (Autocorrelation plots)
3. Split Data to train Autoregressive model.
4. Identify best model for the problem.