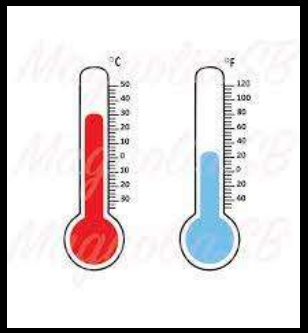
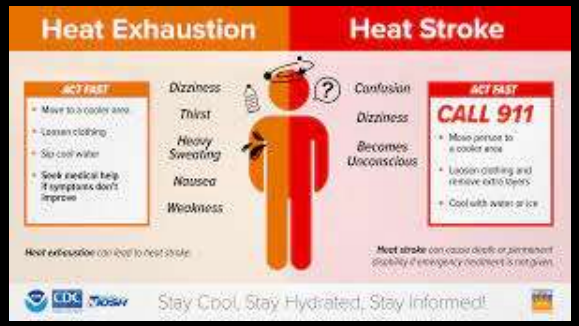
CAPSTONE SPRINT 2

Unpredictable Weather in Vancouver: Let's Predict it with Data Science.

(Not just "why is it so hot?" but also "why is it so cold?")



Problem Statement



Based on Weather Statistics, can we predict when temperature will be unusually high or low?

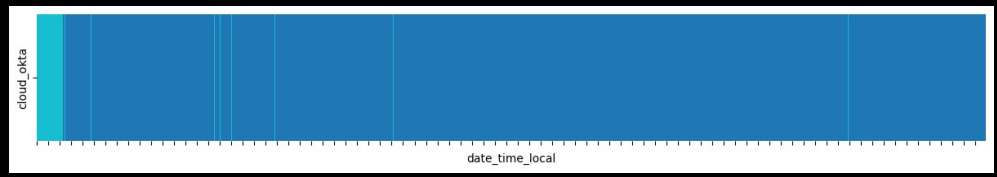
Potential Impacts:

- Health
- Retail Store
- Grocery Stores
- Clothing Store

Data Set:

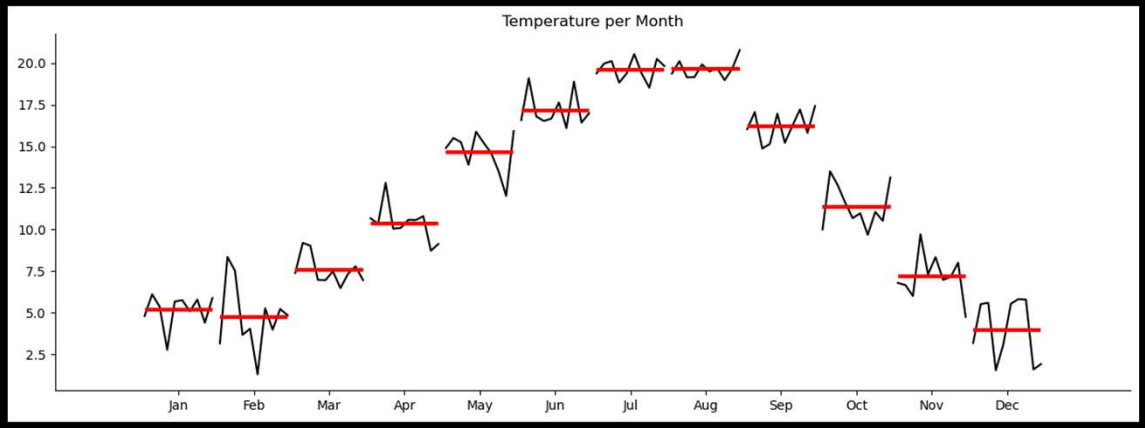
87648 hours of Weather Statistics for Vancouver (July 1st, 2013 - June 30, 2023)

1. Excel Cleaning
2. Python Cleaning
3. Seasonal Decomposition
4. Feature Selection (Identified through EDA)
5. Feature Engineering (Variables for EDA)
6. Train/Test Split (7/3 years)



https://vancouver.weatherstats.ca/

EDA Findings



- Most Weather Stats do not have a correlation with Temperature, except for Dew Point
- The hottest recorded temperature was on June 29, 2021, but the hottest June was in 2015
- Coldest Months: December, January, February
- Hottest Months: July and August (June close 3rd)

Baseline Models and Evaluation Metrics:

1. Attempted using Linear Regression to establish a baseline for other models
2. Fitted Vector Autoregressive Model (Multivariate Time Series Analysis)

Model 1

Metric	Train	Test
Mean Absolute Error	0.2263	0.2803
Mean Squared Error	0.12	0.1516
Root Mean Squared Error	0.3471	0.3894
Mean Absolute Percentage Error	6.48%	7.22%
R-squared	0.9969365432749397	0.9964635344878252

Model 2

Metric	Train	Test
Mean Absolute Error	3.5403	3.7976
Mean Squared Error	19.22	21.8036
Root Mean Squared Error	4.3842	4.6694
Mean Absolute Percentage Error	115.78%	107.85%
R-squared	0.5112959872106125	0.4914264507814935

Next Steps:

- Learn about VAR parameters to tune model
- Research other models before ML
- Research LSTM RNN
- Research other ML for multivariate time series.