04.20.25

MATH 377 Sprint 1 Presentation

Water Level in NYC

Tariqul Rupum

1. Defining the Problem
2. Data Science Approach
3. Potential Impact
4. Introduction to Dataset
5. Next Steps

1. **Defining the Problem**
2. Data Science Approach
3. Potential Impact
4. Introduction to Dataset
5. Next Steps

The Problem:

- Cities around the world are increasingly more susceptible to flooding and water damage
- Coastal water level fluctuations can affect infrastructure, transportation, and harm wildlife

The Solution:

Create a model that can accurately predict monthly sea level changes

1. Defining the Problem
2. **Data Science Approach**
3. Potential Impact
4. Introduction to Dataset
5. Next Steps

Data Science Approach:

- Collect historical data from NYC laboratories
- Combine multiple datasets
- Train predictive models
  - Linear regression
  - Random Forest
  - SARIMA (Seasonal autoregressive Integrated Moving Average)
- Analyze seasonal and long term trends

1. Defining the Problem
2. Data Science Approach
3. **Potential Impact**
4. Introduction to Dataset
5. Next Steps

Potential Impacts:

- Helps city plan for flooding and infrastructure resilience
- Supports real estate risk assessments
- Provides insights into local effects of climate change
- Assists environmental organizations in conservation efforts

1. Defining the Problem
2. Data Science Approach
3. Potential Impact
4. **Introduction to Dataset**
5. Next Steps

Introduction to Dataset

Two main data sets:

- *The Battery Water Station*
- *Central Park Weather Station*

Two Independent Variables:

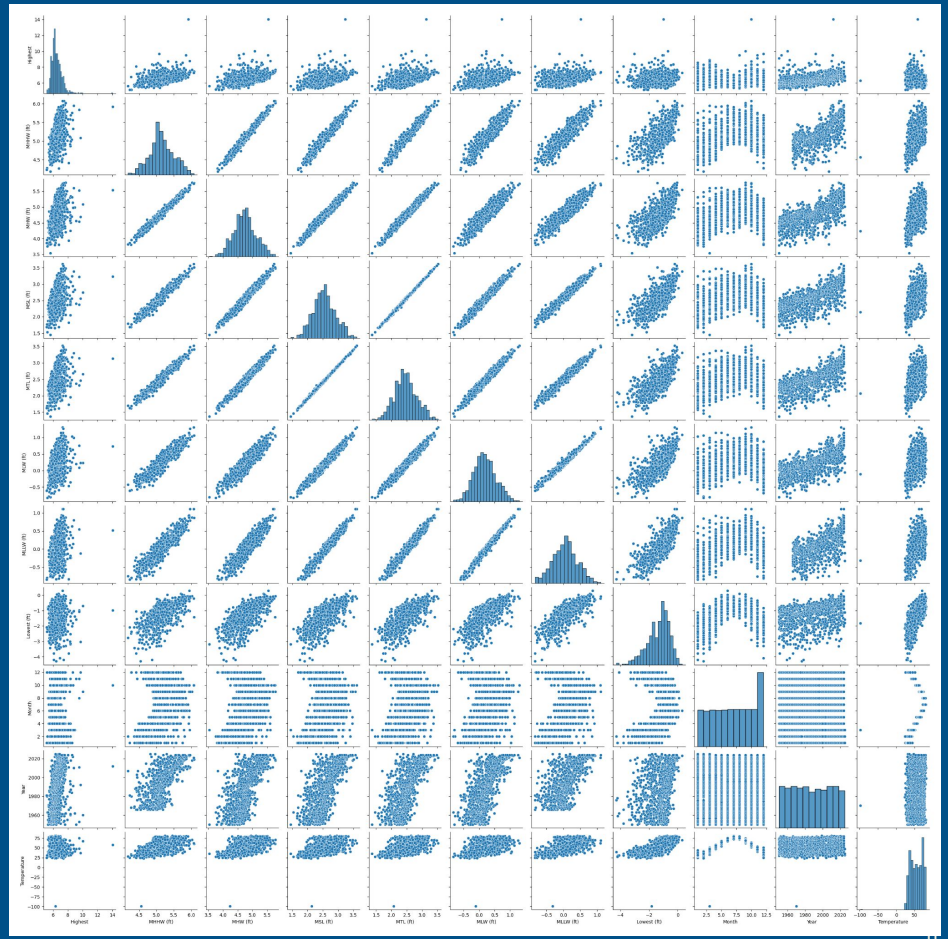
- *Temperature (°F)*
- *Time*

Eight Dependant Variables:

- *Highest Water Level (ft)*
- *MHHW (ft)*
- *MHW (ft)*
- *MSL (ft)*
- *MTL (ft)*
- *MLW (ft)*
- *MLLW (ft)*
- *Lowest Water Level (ft)*

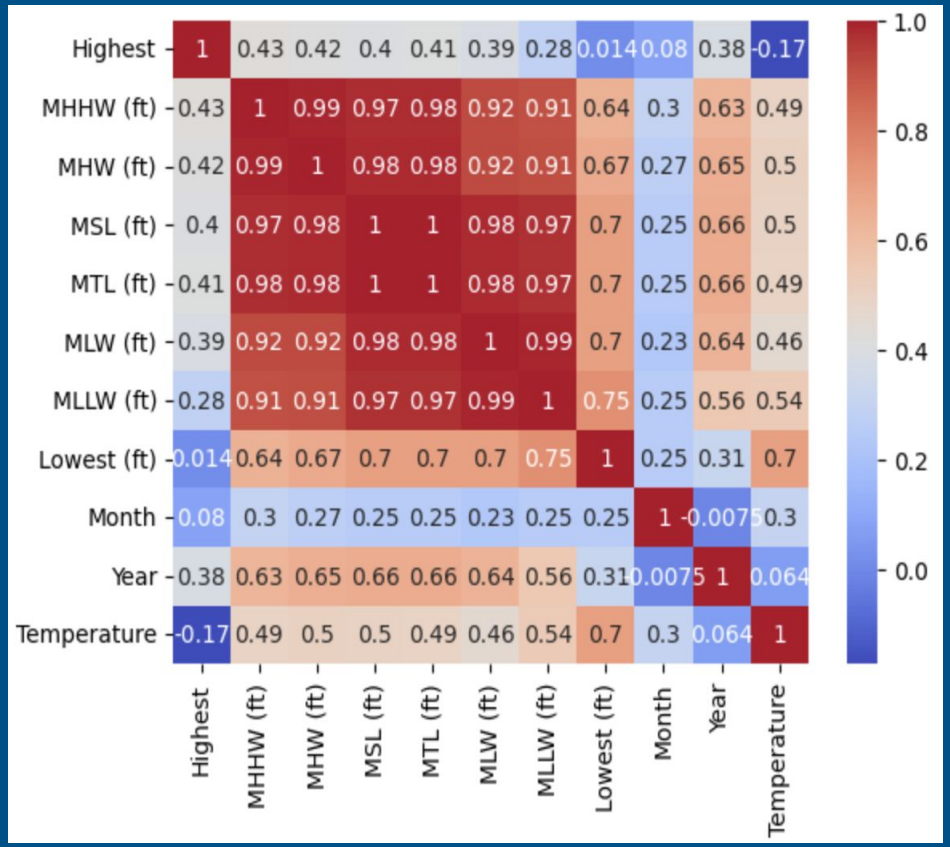
Plotted Matric of Data:

- There is an outlier in the temperature
- As the year increases there seems to be generally positive trend in water level
- During the summer months, there exists an increase in water level compared to winter months



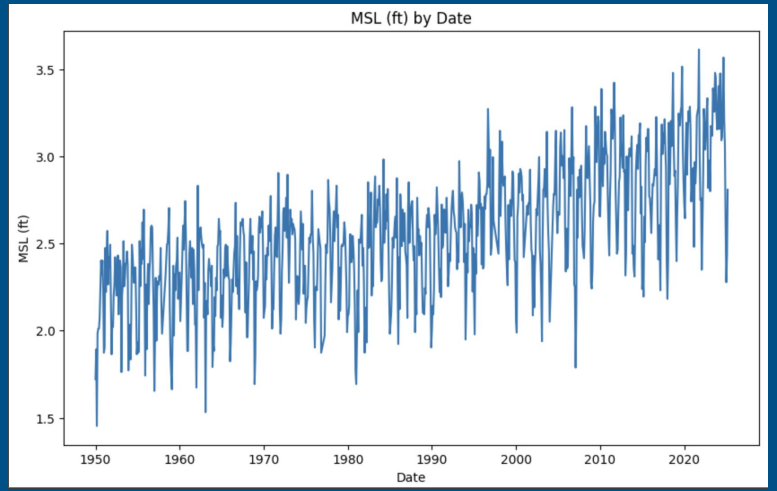
Correlation Matrix of Variables:

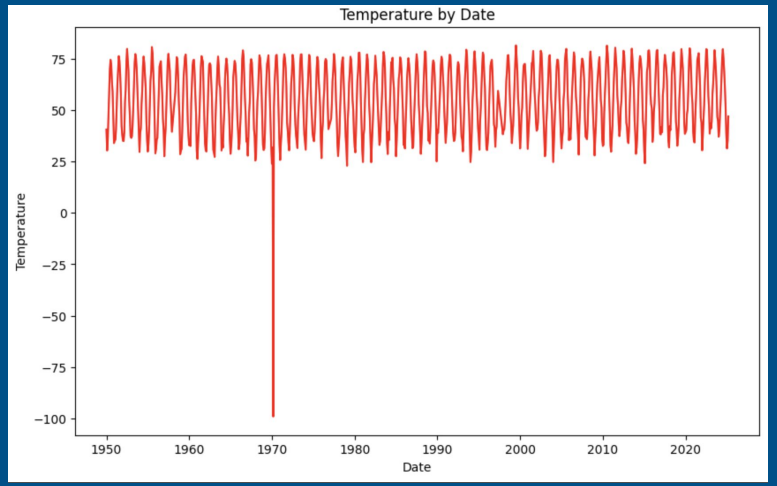
- There seems to be a strong correlation between Temperature and Lowest water level
- There exists strong correlations between all of the mean water level measures
- There exists a moderate correlation between Year and all of the mean water level measurements



Quality Concerns:

- A significant portion of this data originates solely from a single point within NYC
- There are no recorded measurements for MHHW and MLLW before 1970
- There are a couple of months that are missing data
- There are significant outlier within the data





1. Defining the Problem
2. Data Science Approach
3. Potential Impact
4. Introduction to Dataset
5. **Next Steps**

Next Steps:

- Create forecasting models:
  - Linear Regression
  - SIRMA
  - Random Tree
- Evaluate model accuracy for:
  - Monthly/seasonal predictions
  - Annual predictions
- Compare results