

# Data Collection, Cleaning And Visualizations

## DATASET COLLECTION & CLEANING OVERVIEW:

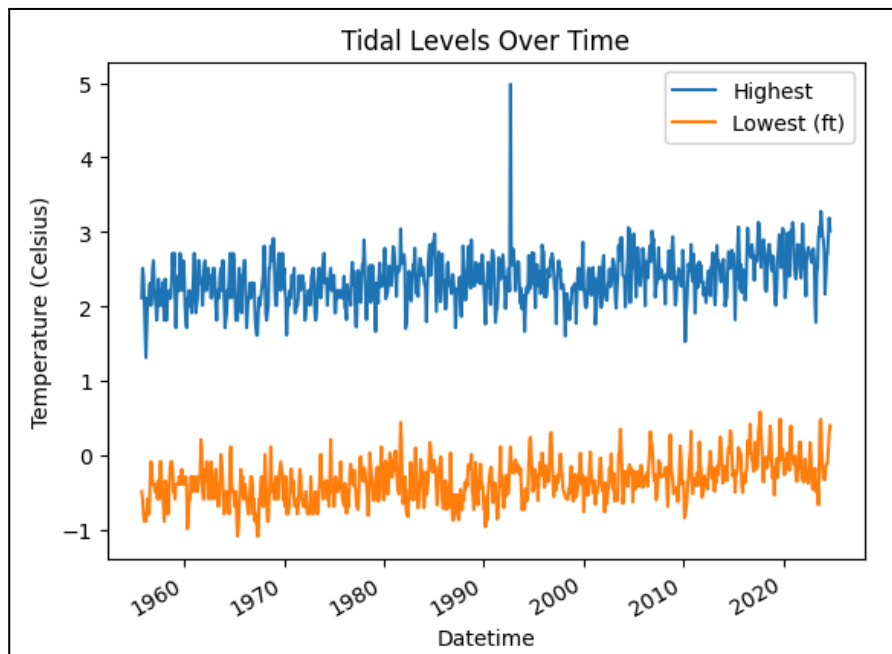
We began with scraping data from a relevant source in order to generate a list of stations for data collection. Then using an API, we retrieved the required information from each identified station. Further, we resolved the missing values by imputing the mean of the attribute, and to further enhance data consistency, we combined separate date and time columns into a unified Datetime column. So in this way, by systematically collecting, cleaning, and transforming the data, we ensured that our dataset is reliable and ready for further analysis.

## DATA VISUALIZATIONS:

We implemented the following visualizations in order to provide a clear understanding of the data patterns and trends that we observed while working for this project. Each visual representation illustrates the most significant findings, and highlights key insights based on our analysis.

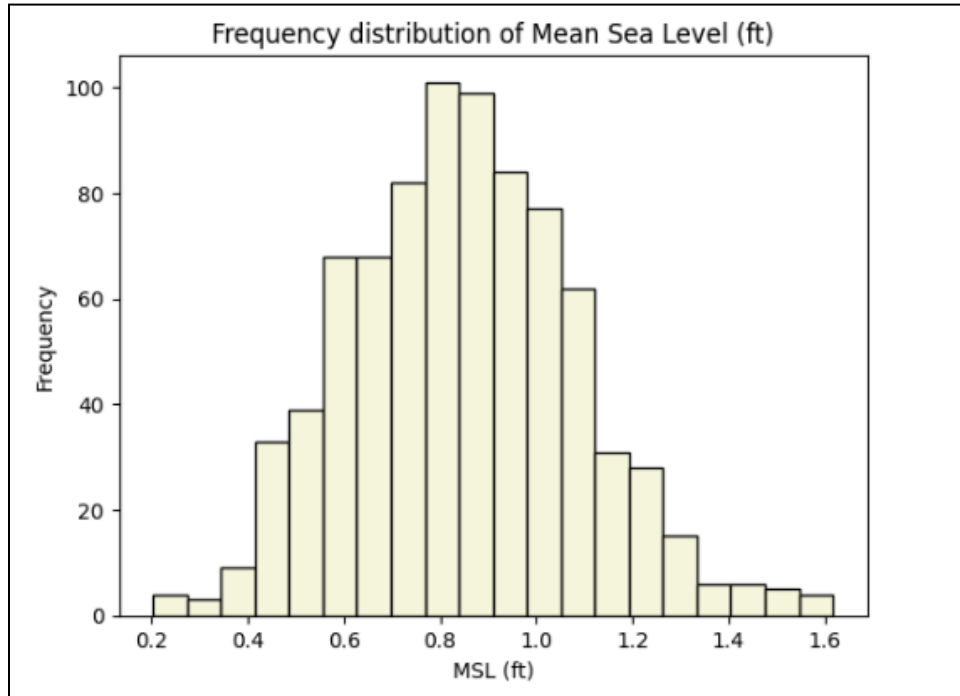
### 1. Tidal Levels over Time:

- This chart displays how the highest and lowest tidal levels change over time.
- This helps us to identify long-term trends, or periodic fluctuations in the tide levels.



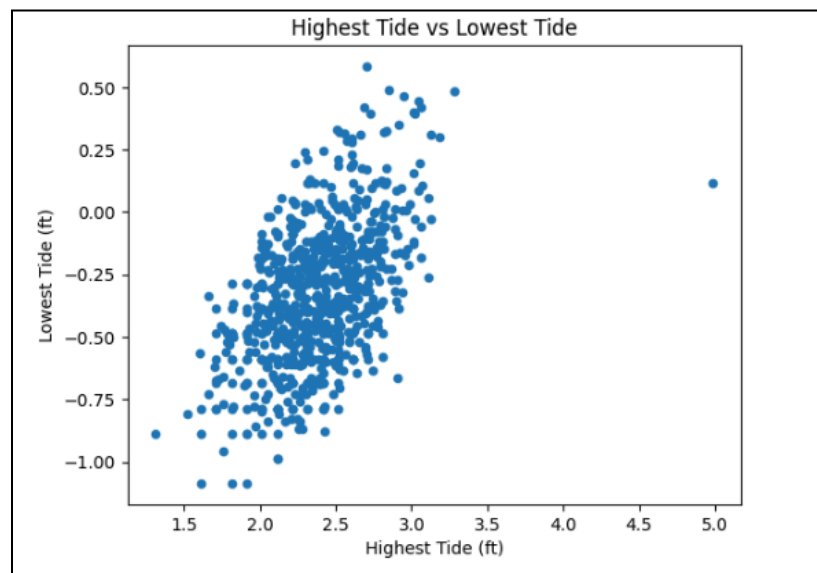
## 2. Frequency Distribution of Mean Sea Level:

- The frequency distribution is showcased using a histogram.
- It provides a view of how frequently different mean sea levels occur and highlights the central tendency and variability in sea level measurements.



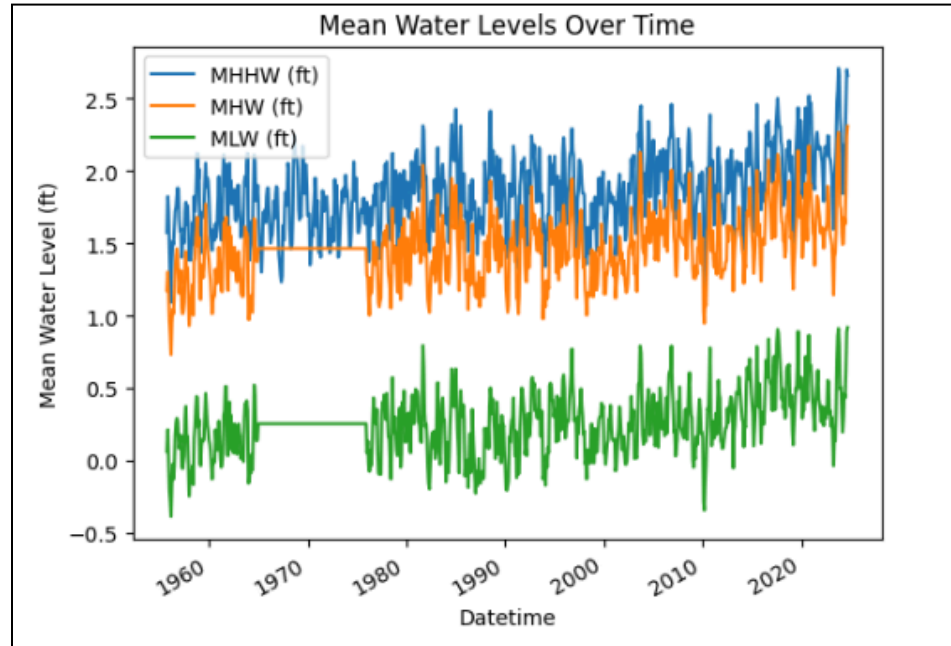
## 3. Highest Tide vs Lowest Tide:

- The scatter plot compares the highest and lowest tides.
- It helps us in revealing the relationship between them and identifying patterns of extreme tide events.



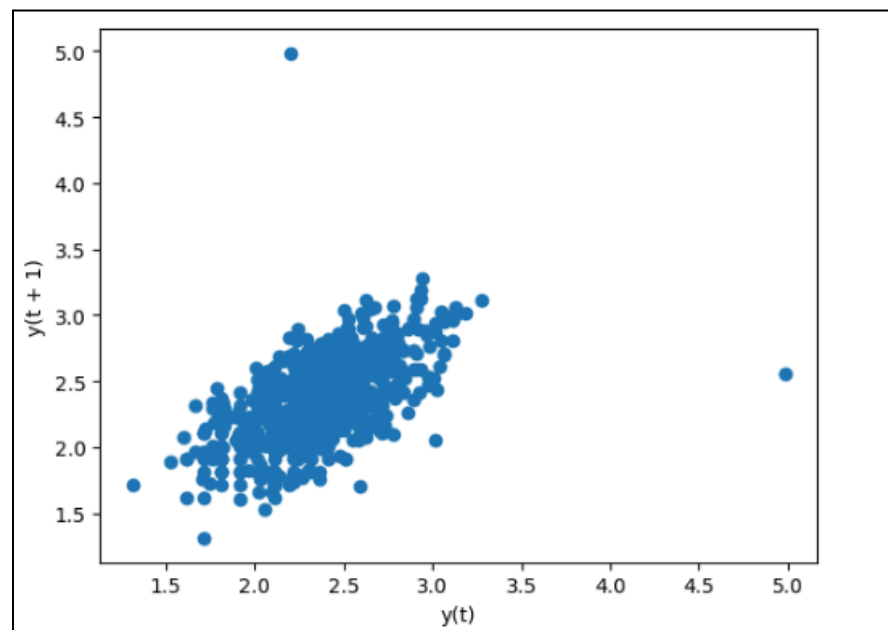
#### 4. Tidal Components over Time:

- This multi-line plot compares different tidal measurements like mean high water, and mean low water over time.
- It offers us insights into how these components behave relative to each other.



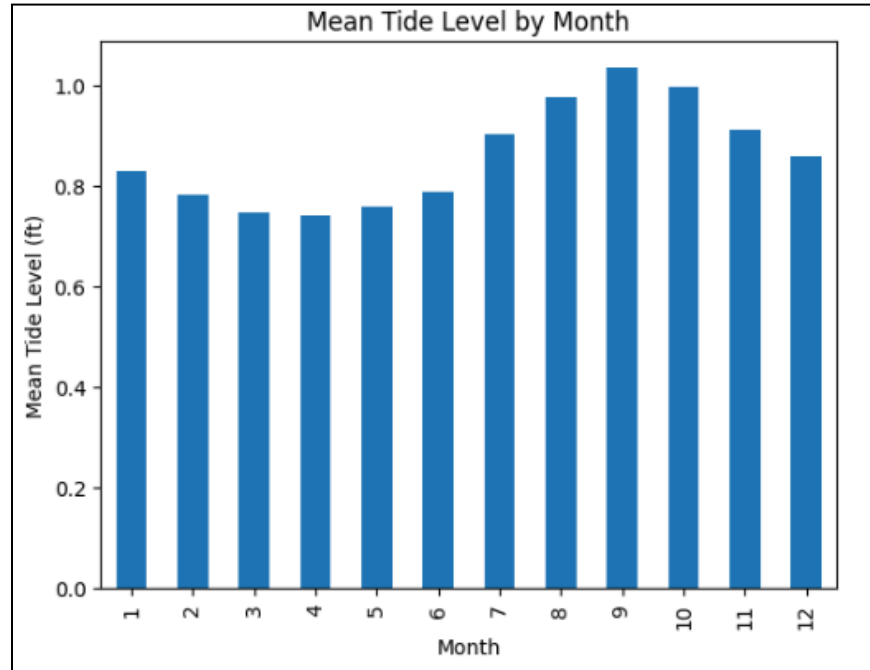
#### 5. Detecting Seasonality and Trends over Time:

- This decomposition plot breaks down the time series into its trend, seasonal, and residual components, and makes it easy for us to spot regular seasonal patterns and overall trends.



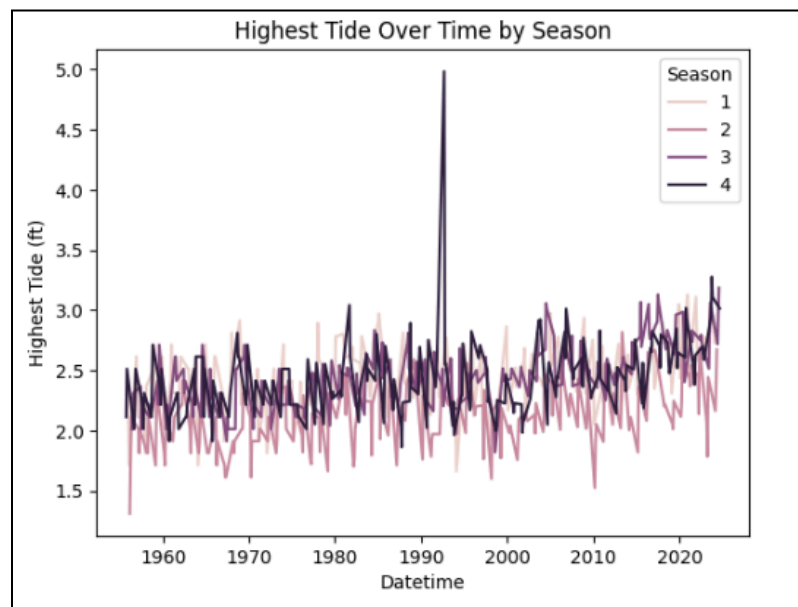
## 6. Mean Tide Levels By Month:

- This bar chart represents how average tidal levels vary from month to month.
- It helps us to identify periods with unusually high or low tides and recurring monthly patterns if any.



## 7. Tide Levels with Seasonal Indicators:

- By representing seasonal changes (e.g., spring, summer), this plot highlights how tide levels fluctuate with the change in seasons, and allows us for a clearer understanding of the influence of seasons on tides.



## **CONCLUSION:**

By implementing data cleaning tasks, such as handling missing values and unifying date and time formats, we ensured the reliability of our dataset. This step was crucial, as inconsistencies in the data could lead to inaccurate insights. Additionally, the visualizations and analysis presented throughout this document offer valuable insights into the underlying data patterns and trends. With the help of these insights, we can make informed decisions and drive meaningful outcomes from this project. Our primary goal during this milestone was to ensure the visualizations we implemented are vital for coastal planning, environmental monitoring, and mitigating the impacts of rising sea levels.