ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

Business Intelligence & Data Visualisation

TD2: Visualization and Pattern Recognition with time series data

Département d'informatique

Ruiwen HE

Weather Patterns in Seattle

Exercise 1: Data Loading and Missing Value Detection

Load the dataset and check for any missing values, and visualize the distribution of these missing values.

Exercise 2: Interpolate and Verify Missing Values

Apply linear interpolation to fill the missing values in the dataset. After interpolation, verify that all missing values have been successfully filled. If there are still missing values, suggest and implement additional methods to handle them (e.g., forward fill, backward fill, or using a constant value). Observe and analyze the results of the imputation.

Exercise 3: Add Columns for Quarter and Daily Average Temperature

Add two new columns to the dataset:

1. A column that assigns each row to a quarter (Q1, Q2, Q3, or Q4) based on the date column. Ensure the dates are correctly parsed and categorized.
2. A column that calculates the daily average temperature, using the available temperature data (e.g., maximum and minimum temperatures).

Verify the correctness of both new columns."

Exercise 4: Visualizing Temperature Trends for Spring 2012

Filter the dataset to select data for the year 2012 and the Spring season. Then, create a line plot to visualize the temperature trends during the Spring of 2012. The plot should display:

1. The maximum temperature (temp_max) in red.
2. The minimum temperature (temp_min) in blue.
3. The average temperature (avg_temp) in green.

Ensure that the x-axis represents the dates, and the y-axis shows the temperature values (°C). Add appropriate labels for the title, axes, and legend to make the plot clear and informative.

Exercise 5: Plot Seasonal Average Temperature Variations for 2012

Group the dataset by season and calculate the average values for the maximum temperature (temp_max), minimum temperature (temp_min), and average temperature (avg_temp). Then, reshape the data into a long format suitable for plotting. Create a bar plot that visualizes the seasonal average temperature variations in 2012. The bar chart should display:

1. The average maximum temperature in each season, represented by a specific color.
2. The average minimum temperature in each season, represented by a different color.
3. The average temperature in each season.

Exercise 6: Visualize Weather Type Distribution for Spring 2012

Analyze the distribution of weather types in the Spring of 2012 by counting the occurrences of each weather type in the dataset. Create a bar chart using Plotly to visualize the distribution. The chart should display:

1. The weather types on the x-axis.
2. The count of occurrences for each weather type on the y-axis.
3. Different colors for each weather type to distinguish them visually.

Ensure the chart includes an appropriate title and axis labels for clarity.

Exercise 7: Calculate and Visualize Monthly Average Temperature

Calculate the monthly average temperature for each month in the year 2012. To do this, group the dataset by month and compute the average of the maximum, minimum, and average temperatures. Then, visualize these monthly averages using a line plot, where each line represents one of the temperature metrics (max, min, average).
Make sure to label the axes and include a title. Use different colors to distinguish between the temperature types.

Exercise 8: Time Series Decomposition of Average Temperature

Using an additive model, perform a time series decomposition of the average temperature (avg_temp). The decomposition should account for daily data with annual seasonality. The steps are:

1. Set the date column as the index of the dataset.
2. Apply the seasonal_decompose function to decompose the avg_temp column into its components: trend, seasonal, and residual.
3. Plot the decomposed components (trend, seasonal, and residual) to visualize the individual patterns.

Ensure that the plot is appropriately formatted and the components are clearly labeled for easy interpretation.

Exercise 9: Visualize Daily Precipitation Changes in Seattle for 2012

Plot the daily precipitation data for Seattle in 2012 using a line chart. The chart should show how the precipitation changed day by day throughout the year.

Exercise 10: Create an Interactive Choropleth Map of Daily Precipitation Changes in Seattle

Create an interactive choropleth map to visualize daily precipitation changes in Seattle for the year 2012. The steps include:

1. Load a GeoJSON file representing the Seattle city boundaries.
2. Use Plotly's choropleth function to plot the precipitation data for each location.

3. Use the rowID column to match data with the geographical boundaries in the GeoJSON file.
4. Set the precipitation column as the color scale, and ensure the color transition is smooth, using a soft blue gradient.
5. Use the animation_frame argument to animate the map by date, showing how precipitation changes over time.
6. Customize the map's borders, background, and other visual elements for a clean, readable presentation.

Ensure the map has a title and color bar that indicates the range of precipitation values. The final map should be interactive and clearly show how precipitation varies across Seattle over the course of 2012.

Exercise 11: Pattern Recognition in Seattle Weather Data (Bonus)

Objective: Identify and classify distinct weather patterns over time using the available weather data.

Description:

1. Preprocess the Data: Ensure that the dataset is cleaned, with missing values handled appropriately.
2. Extract features related to weather patterns, such as daily maximum temperature, minimum temperature, average temperature, and precipitation. Identify Patterns:
3. Use unsupervised learning techniques like clustering (e.g., K-means clustering or DBSCAN) to group days with similar weather patterns based on temperature and precipitation data. Optionally, explore dimensionality reduction methods (e.g., PCA) to visualize the clusters in a 2D space. Visualize the Patterns:
4. Visualize the identified clusters using scatter plots or time series plots to show how distinct weather patterns evolve over time. Use color coding or different markers to distinguish between the identified weather patterns. Analyze and Interpret:
5. Analyze the characteristics of each identified weather pattern. For example, you might discover patterns representing hot and dry days, cold and rainy days, or moderate weather. Discuss how these patterns might relate to real-world phenomena, such as seasonal weather changes or extreme weather events.

Expected Output:

- A visualization of weather patterns clustered into distinct groups, with analysis of the characteristics of each group.
- A brief report interpreting the identified weather patterns, including any trends or insights.

This exercise focuses on applying machine learning techniques to identify patterns in weather data, which is a key aspect of pattern recognition in real-world datasets.