**Dataset Selected:** Retirement ages around the World. (<https://data.world/makeovermonday/2023w16>)

**Numerical Features (Year, Average Age):**

Year: The dataset contains data from 1970 to 2020, with an average year around 2006. The years span a wide range, indicating data collected over several decades.

Average Age: The average age across the dataset is approximately 62.63 years, with a standard deviation of around 3.24 years. The age distribution appears to be relatively tight, with most ages falling between 51.1 years and 78.6 years. The median age (50th percentile) is 62.5 years.

**Categorical Features (Country, Indicator, Gender):**

Country: The dataset covers 53 unique countries, with Australia (AUS) being the most frequently occurring country.

Indicator: There are 1175 unique indicators in the dataset, with "Effective labor market exit age, men" (PEN19A) being the most common indicator.

Gender: There are two unique genders in the dataset, with "men" being the most frequent gender.

**Conclusion:**

The dataset provides information on labor market indicators across multiple countries over several decades, with a focus on factors such as age, gender, and specific indicators related to labor market exit. The average age suggests that the dataset covers a broad age range of individuals within the workforce. Additionally, the distribution of countries and indicators indicates a diverse range of data sources and topics covered. Further analysis could explore trends over time, differences across countries, and the impact of gender on labor market dynamics.

**Find the relevant time series parameters associated with the data. Interpret the results and support the answer with relevant conclusion.**

Based on the chosen dataset, several relevant time series parameters can be associated with the data:

**1. Trend:** The dataset spans several decades, indicating potential long-term trends in labor market indicators such as average age and labor market exit age. Analyzing the trend over time can provide insights into how these indicators have evolved over the years.

**2. Seasonality:** Depending on the indicators and countries included in the dataset, there may be seasonal patterns or cyclical variations in labor market dynamics. For example, certain indicators may exhibit seasonal fluctuations due to factors like economic cycles or seasonal employment trends.

**3. Stationarity:** Assessing stationarity is crucial for time series analysis. Stationarity implies that the statistical properties of the data, such as mean and variance, remain constant over time. Detecting and addressing non-stationarity through techniques like differencing or transformation is essential for accurate modeling and forecasting.

**4. Autocorrelation**: Autocorrelation measures the correlation between a time series and its lagged values. Analyzing autocorrelation can help identify patterns and dependencies within the data, such as whether the current labor market conditions are influenced by past conditions.

**5. Seasonal Decomposition:** Decomposing the time series into its trend, seasonal, and residual components can provide a clearer understanding of the underlying patterns and dynamics. This decomposition helps separate long-term trends from seasonal variations and irregular fluctuations.

**6. Forecasting:** Using techniques such as ARIMA (Autoregressive Integrated Moving Average) or SARIMA (Seasonal ARIMA), it's possible to forecast future values of labor market indicators based on historical data. Forecasting can assist policymakers, businesses, and researchers in making informed decisions and planning.

Interpreting the results of these time series parameters can provide valuable insights into labor market dynamics, including long-term trends, seasonal patterns, and cyclical variations. By understanding these characteristics, policymakers and businesses can better anticipate changes in the labor market, plan interventions or strategies accordingly, and make data-driven decisions to address challenges or seize opportunities. Additionally, analyzing the time series parameters can help identify potential areas for further research or exploration, such as investigating the impact of external factors (e.g., economic policies, technological advancements) on labor market dynamics.

**For the selected data, implement an inferential statistics test by selecting any one or more features. Clearly state the hypothesis, the name of the test and the final inference.**

**Inferential Statistics 01:**

Let's say we want to conduct an inferential statistics test to determine if there is a significant difference in the average age of workers between two genders: men and women. We can formulate the hypothesis as follows:

* Null Hypothesis (H0): There is no significant difference in the average age of workers between men and women.
* Alternative Hypothesis (H1): There is a significant difference in the average age of workers between men and women.

We'll use a **two-sample t-test** to compare the means of the two groups (men and women) and determine if the difference in their average ages is statistically significant.

Interpretation:

* **Hypothesis Test**: We conducted a two-sample t-test to compare the average age of workers between men and women.
* **Test Result:** The calculated p-value is compared to the significance level (alpha = 0.05). If the p-value is less than alpha, we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.
* **Final Inference:** Based on the test results, if the p-value is less than 0.05, we infer that there is a significant difference in the average age of workers between men and women. Otherwise, if the p-value is greater than or equal to 0.05, we conclude that there is no significant difference in the average age between the two gender groups.

This inferential statistics test helps us understand if there's a significant gender-based difference in the average age of workers, providing insights into potential disparities in the labor market.

**Python Code:**

**A screenshot of a computer

Description automatically generated**

**Output:**

**T-statistic:** 9.495218927686334

**P-value:** 1.2402682182863764e-20

**Reject null hypothesis.** There is a significant difference in the average age of workers between men and women.

**Result:**

The results of the two-sample t-test indicate a t-statistic of approximately 9.50 and a very small p-value (1.24e-20), significantly lower than the standard significance level of 0.05.

**Conclusion:**

Given the low p-value, we reject the null hypothesis, which suggests that there is **no significant difference in the average age of workers between men and women**. Instead, we conclude that **there is a statistically significant difference in the average age of workers between these two gender groups.**

This implies that there is likely a meaningful disparity in the average age between male and female workers within the dataset. Further investigation may be warranted to understand the factors contributing to this difference and its implications for labor market dynamics, workforce planning, and gender equity initiatives.

**Inferential Statistics 02:**

Certainly! Another suitable test for comparing the average age between two gender groups in the dataset could be a **non-parametric test called the Mann-Whitney U test**. This test is used to determine if there is a significant difference in the median values of a continuous variable between two independent groups, especially when the data do not meet the assumptions of normality required for parametric tests like the t-test.

**Python Code:**

**A screenshot of a computer

Description automatically generated**

**Output:**

**Mann-Whitney U statistic:** 227603.0

**P-value:** 2.875556995693015e-21

**Reject null hypothesis.** There is a significant difference in the average age of workers between men and women.

**Result:**

The Mann-Whitney U test yielded a U statistic of approximately 227603.0 and a very small p-value (2.88e-21), significantly lower than the standard significance level of 0.05.

**Conclusion:**

Given the extremely low p-value, we reject the null hypothesis, indicating that there is a significant difference in the average age of workers between men and women within the dataset. This implies that the median ages of male and female workers are indeed different.

This finding suggests that gender may have a notable impact on the average age of workers, which could be indicative of broader disparities or differences in workforce participation, career trajectories, or retirement patterns between men and women. Further investigation into the underlying factors contributing to this difference could provide valuable insights for addressing gender-related issues in the labor market and promoting equitable practices.

A graph of a number of people

Description automatically generated with medium confidenceA graph of a person and person

Description automatically generated**Visualize the data with all possible and relevant plots citing important observations.**

 







