Ashley Chen
DSCI549
Homework 7 - Project Part 3

1. **Updated Project Description**

   This project explores gender disparity in part-time employment across countries using a dataset from UNData. The focus is on identifying which countries have the largest gaps between men and women in part-time work and understanding how these gaps have changed over time. Additionally, the project aims to explore whether national policies or cultural factors are associated with greater or lesser gender disparities in part-time employment. The dataset contains statistics by country and year, including the percentage of adult men and women in part-time employment and the share of part-time workers who are women. These values can be used to calculate a gender disparity ratio, which serves as a key metric throughout the analysis.

   The project will also incorporate external data sources, such as indicators of labor laws, gender equality indices, and family policy data. Combining these sources allows for the use of regression modeling to identify patterns and possible explanations for the observed disparities.

   The analysis will involve several stages: cleaning and preparing the data, calculating new variables, visualizing trends and comparisons using bar charts, line graphs, and maps, and finally, using predictive modeling techniques to explain differences between countries. The dataset is structured and available in formats like CSV and Excel, making it accessible for use with tools such as Python, R, or Excel. This project offers a practical application of core data science techniques, including data cleaning, exploratory analysis, feature creation, data merging, and basic modeling, while highlighting the value of combining datasets to address real-world social questions.

2. **Elevator Pitch**

   This project analyzes global gender disparities in part-time employment by combining UNData labor statistics with external datasets on labor laws and gender equality indices, aiming to uncover how national policies influence the gender gap in part-time work participation.

3. **Updated Project Objectives**

   Objectives:
   1. Quantify gender disparities in part-time employment using a computed metric, the gender disparity ratio, defined as the percentage of women in part-time work divided by the percentage of men

2. Identify cross-country patterns and highlight countries with the largest and smallest disparities
3. Analyze temporal trends and assess how gender disparities in part-time employment have evolved over time within and across regions
4. Explore structural factors and determine whether indicators such as labor protections, family policies, or societal gender equality norms are statistically associated with gender disparity ratios
5. Recommend possible strategies or areas for further policy focus based on empirical findings

Datasets:

1. Primary dataset:
    a. UNData - Part-time Employment by Gender
        i. Country-level, annual data showing the percentage of men and women in part-time work and the overall share of women among part-time workers
2. External Datasets (Policy and Gender Indicators):
    a. UNDP - Gender Equality Index (GII)
        i. Country-level, updated annually, measures inequality in reproductive health, empowerment, and labor market
    b. World Economic Forum - Global Gender Gap Report
        i. Measures economic participation, educational attainment, health, and political empowerment
    c. OECD Family Database - Parental Leave and Childcare Support
        i. Length of paid maternity leave, public childcare support availability

Pre-Processing Plan:

1. Importing: Load all datasets into Python using pandas
2. Cleaning: Standardize column names (remove special characters, lower case formatting)
3. Handling Missing Values
    a. Forward-fill or imputing missing values
        i. For example, since Brazil is missing women's part-time employment data for 2017 but has values for 2016 and 2018, a forward-fill or interpolation method would be appropriate because only a single year is missing between valid observations.
    b. Skip row entirely
        i. Cambodia in the UNDataset has missing values for both men's and women's part-time employment percentages in 2014 and 2015. Since multiple consecutive years are missing and no surrounding year data exists, it'd be best to drop these rows from the analysis to maintain the reliability of year-over-year comparisons.
4. Feature Creation: Compute the gender disparity ratio

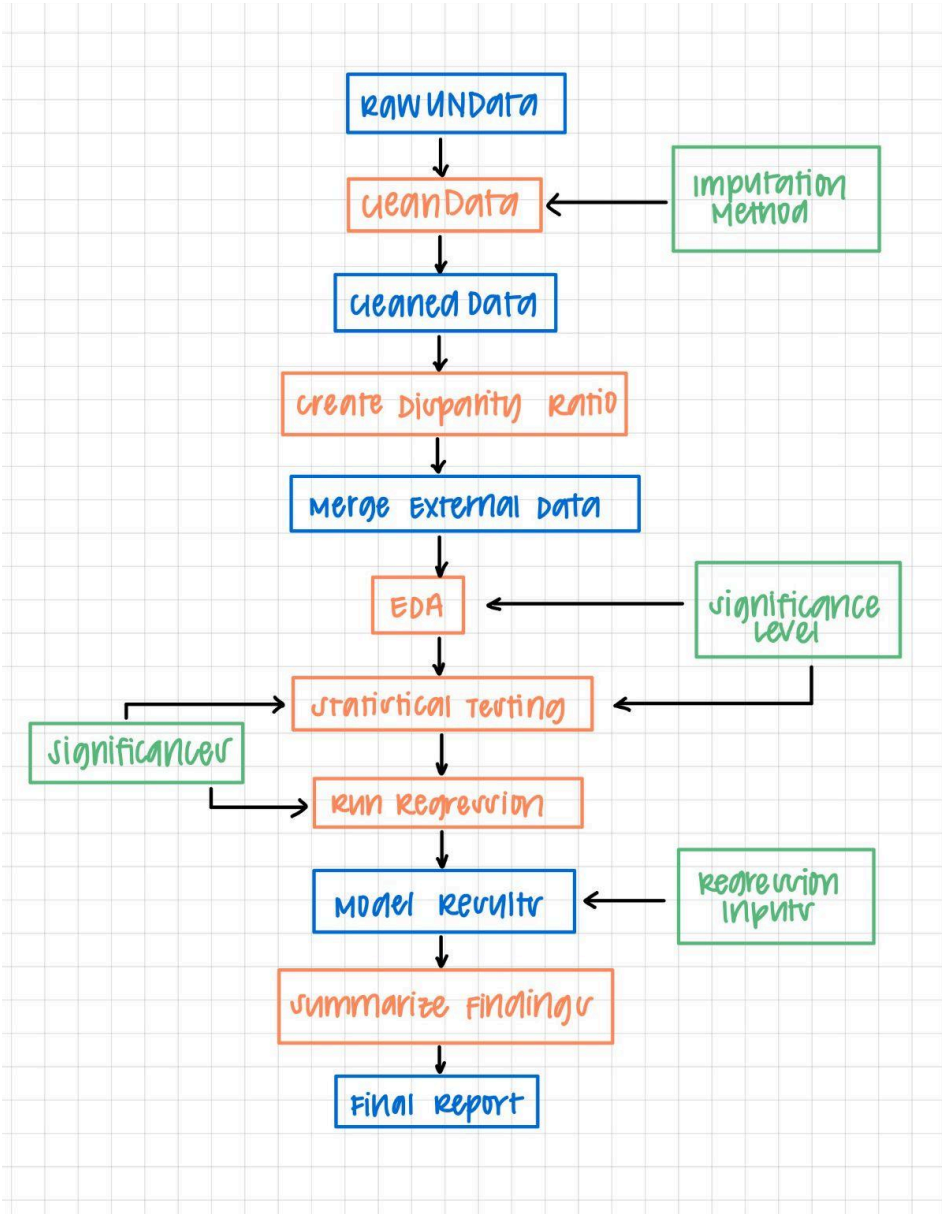
5. Formatting: Convert percentages to decimal format for mathematical operations
6. Merging: Integrate external datasets based on standardized country names and matched years

4. **Project Scope**

While the project aims to provide a meaningful analysis of gender disparities in part-time employment, there are important boundaries to its scope. The primary focus will be on part-time employment patterns, meaning full-time employment trends and broader labor market dynamics will not be addressed. Additionally, the analysis will be conducted at the country level, relying on national aggregates rather than individual-level data. Therefore, variations within countries, such as rural versus urban disparities or sector-specific differences, will not be captured.

The project also does not seek to establish causal relationships between policy factors and gender disparities. Although regression analysis will be employed to identify statistical associations, causal inference would require more complex methods and richer data, such as longitudinal tracking or natural experiments, which are beyond the reach of this project. Furthermore, while external datasets will be incorporated to expand the analysis, only a limited number of policy and equality indicators will be used. Other potentially important variables, such as cultural norms not captured by existing indices or sector-specific employment data, will remain outside the project's scope. Finally, countries with substantial missing data may be excluded from certain analyses, limiting the generalizability of the findings to only those countries with relatively complete data records.

5. **Updated Data Analysis and Solution**



The workflow diagram above outlines the data science process for analyzing gender disparity in part-time employment. The analysis begins with raw UNData and external policy indicators as input sources. The data is first cleaned and standardized, with missing values handled through imputation. A gender disparity ratio is then calculated to quantify differences in part-time participation. After merging external policy datasets, exploratory data analysis (EDA) is conducted, followed by statistical testing to determine whether year-over-year changes are significant. Regression modeling is then used to identify which policy indicators are associated with smaller gender gaps. The results are interpreted and summarized in a final report. Green boxes represent key parameters

that influence specific functions, while orange and blue boxes alternate between data states and transformations.

6. **Outcome of Project**

By the end of this project, the expected outcome is a comprehensive analysis of gender disparities in part-time employment across a wide range of countries and regions. Through the creation of a gender disparity ratio and the integration of external datasets, the project will reveal significant patterns in how gender gaps in part-time work differ geographically and evolve over time. The regression analysis will provide an understanding of how structural factors, such as labor protections, family leave policies, and broader measures of gender equality, relate to disparities in part-time employment participation. In addition to documenting patterns, the project aims to generate practical insights that could inform policymaking, particularly by highlighting which types of policies are statistically associated with narrower gender gaps. The final deliverable will be a polished and structured report that includes clear visualizations, statistical modeling outputs, and thoughtful interpretations of the findings. If time allows, an additional extension of the project could include developing an interactive dashboard to facilitate deeper exploration of the data and allow users to compare countries and policy impacts dynamically.