**Analyzing the Complexity of the U.S. Tax Code: Pre- and Post-TCJA Comparison**

**Objective:**

The goal of this project is to quantitatively and qualitatively compare the complexity of the U.S. Tax Code before and after the enactment of the Tax Cuts and Jobs Act (TCJA) of 2017. By leveraging data science methodologies, we aim to identify key trends and insights into how the TCJA affected the structure and comprehensibility of the Tax Code.

**Key Tasks:**

1. **Word Count Analysis**:
   * Measure the total word count of the Tax Code before and after the TCJA to assess any significant changes in length (need to consider whether we want to count all words or only substantive ones, *e.g.*, no “if”, “and”, etc.)
2. **Cross-Reference Mapping**:
   * Count and map cross-references within the Tax Code (*e.g.*, references to other sections, subsections, or regulations) to gauge interconnectedness and potential for complexity.
3. **Variable-Based Complexity Metrics**:
   * Analyze changes in the number of defined terms, exceptions, and special rules introduced by the TCJA.
   * Identify provisions with layered conditions (*e.g.*, "if-then" clauses) and quantify their prevalence.
4. **Structural Complexity**:
   * Map the hierarchical structure of sections, subsections, and subparts before and after the TCJA to understand any shifts in organizational complexity.
5. **Comparative Linguistic Analysis**:
   * Use natural language processing (NLP) to assess readability scores (*e.g.*, Flesch-Kincaid) and linguistic patterns that may reflect increased or decreased complexity.
   * Identify key phrases or jargon introduced by the TCJA and evaluate their impact on accessibility.

**Suggested Approach:**

* Utilize publicly available versions of the Tax Code for the pre-TCJA and post-TCJA periods (reference the website discussed during the Zoom call).
* Develop scripts to parse the text and extract metrics such as word counts, cross-references, and readability scores.
* Visualize findings with comparative graphs and charts to highlight key differences.

**Deliverables:**

1. Quantitative metrics comparing the pre- and post-TCJA Tax Code.
2. Data visualizations showcasing changes in length, structure, and readability.
3. Insights into how the TCJA influenced the complexity of the U.S. Tax Code, with a focus on practical implications for taxpayers and policymakers.