**Considerations for Extracting Headings from PDF documents:**

1. **Text Length**: The text should be less than 50 characters.
2. **Font Size**: The font size must be greater than 10.
3. **Whitespace**: The text should not start or end with white spaces.
4. **Completeness**: The text should represent a complete thought or title, not an incomplete sentence.
5. **Punctuation**: The text should not end with punctuation marks like a period (.), comma (,), or semicolon (;).
6. **Styling**: The text should have distinct visual characteristics, such as bold or italic styles, if available.
7. **Alignment**: Headings are usually left-aligned, center-aligned, or consistently placed compared to body text.
8. **Line Breaks**: Headings should not contain multiple line breaks; they typically appear as single lines.
9. **Positioning**: Headings often appear near the top or middle of a page rather than at the bottom.
10. **Keywords**: Contextual keywords such as "Introduction," "Overview," "Chapter," or other domain-specific terms can indicate headings.
11. **Capitalization**: Avoid treating all-uppercase text as headings unless it is explicitly styled that way.
12. **Font Consistency**: Verify that the font type is consistent with headings throughout the document.
13. **Style Uniformity**: Headings often follow a uniform style or hierarchy, such as H1, H2, or H3, in structured documents.

**Considerations for Identifying Sub-Headings:**

1. **Positioning and Layout:**
   * **Main Headings** are typically positioned at the top of sections or at the beginning of new chapters.
   * **Sub-Headings** are usually placed below main headings or within the body of the section, often starting a new subsection or topic.
2. **Indentation:**
   * **Sub-Headings** might have **slight indentation** compared to main headings, though they are not necessarily bold or larger in font size.
   * Main headings might be fully left-aligned or centered, while sub-headings could be indented slightly (e.g., to the right or under the main heading).
3. **Line Spacing and Whitespace:**
   * **Main Headings** generally have **more whitespace** above and below them, separating them from both the text and the sub-headings.
   * **Sub-Headings** typically have less space before and after them but are still distinguishable due to their proximity to the surrounding text or the body content.
4. **Contextual Keywords:**
   * **Sub-Headings** often provide specific topics or subdivisions within a larger theme. They may contain **contextual keywords** related to the main heading.
   * For example, if a main heading is **"Introduction,"** sub-headings could be **"Background," "Methodology,"** or **"Purpose"**—providing more granular detail of the topic.
5. **Styling or Font Attributes (Bold/Italic):**
   * **Sub-Headings** might be **bold**, **italicized**, or in **uppercase**, even if the font size is the same as the body text.
   * Some documents may use **bolding** for sub-headings, even if the font size remains constant. If no explicit styling exists, you can look for patterns in how text is presented (e.g., some might be underlined or in a slightly different font weight).
6. **Text Length:**
   * **Sub-Headings** are often shorter and more specific compared to main headings, which tend to be more general.
   * **Main Headings** may represent the overarching theme of a section or chapter, while **sub-headings** break the section into smaller, more detailed topics.
   * If the length of the text is under 50 characters, it's more likely to be a **sub-heading**.
7. **Numbering and List Format:**
   * **Main Headings** could be numbered at the top level (e.g., **"1. Introduction"**), while **sub-headings** might follow a lower level of numbering (e.g., **"1.1 Background"** or **"1.1.1 Research Design"**).
   * Sub-headings in numbered lists typically follow a multi-level structure: **"1.1"** for a sub-heading, **"1.1.1"** for a sub-subheading, and so on.
8. **Use of Colons or Punctuation:**
   * **Main Headings** might not use punctuation (except for potentially a colon at the end of the title in some cases).
   * **Sub-Headings** might sometimes end with punctuation marks like **colons** (e.g., **"Objectives of the Study:"**), distinguishing them from main headings.
9. **Alignment and Margins:**
   * **Main Headings** are often aligned with larger margins or centered at the top of a section.
   * **Sub-Headings** tend to have smaller margins or a slight indentation compared to the main headings, and they may be aligned with the body text.
10. **Textual Hierarchy and Structure:**
    * **Sub-Headings** typically organize a section into smaller, focused topics, which help break down content for readability and structure.
    * **Main Headings** represent a larger topic or section (e.g., **"Chapter 1: Introduction"**), while **sub-headings** help to break down that chapter into digestible parts (e.g., **"1.1 Purpose," "1.2 Objectives"**).
11. **Repetition of Main Heading Themes:**
    * **Sub-Headings** are often variations or specific aspects of the **main heading**, helping the reader navigate through different aspects of the topic discussed in the section.