Assignment: Automating Secondary Research

Benori

Report: Analysis of Themes and Insights

Arti Gupta

+917007620478

1. Approach & Assumptions

1.1 Approach

Data Collection: Data was scraped from KPMG's insights page (https://kpmg.com/xx/en/our-insights.html) using Selenium WebDriver to extract article titles, URLs, descriptions, and content across 50 pages. PyMuPDF extracted text from linked PDF files. A total of 400 articles were collected and stored in kpmg_insights_full.csv.

Data Preprocessing: Text from articles and PDFs was preprocessed using NLTK, involving lowercase conversion, removal of special characters and digits, tokenization, stopword removal, and lemmatization. The cleaned text was stored in a cleaned_text column, yielding 400 valid articles.

Industry Classification: Articles were classified into industries (e.g., Sustainability, Technology) using a zero-shot classification pipeline with the HuggingFace facebook/bart-large-mnli model. Predefined industry labels ensured relevance, with results stored in the industry column.

Topic Modeling: Themes were extracted using BERTopic with all-MiniLM-L6-v2 SentenceTransformer embeddings. BERTopic generated topic keyword clusters (e.g., "ai, kpmg, service, data, technology"), stored in topic and topic_keywords columns.

Clustering: HDBSCAN clustered articles based on all-MiniLM-L6-v2 embeddings, grouping them by semantic similarity. Cluster labels were stored in the cluster column, with unique artifact IDs for traceability.

Visualization: Findings were visualized using Matplotlib, Seaborn, and WordCloud, including bar charts for industry and topic distributions, and word clouds for top topics.

Tools: Python, pandas, Selenium, PyMuPDF, NLTK, HuggingFace Transformers, SentenceTransformers, BERTopic, HDBSCAN, Matplotlib, Seaborn, WordCloud.

1.2 Assumptions

- **Data Completeness**: The 400 articles scraped represent a comprehensive sample of KPMG's recent insights, covering key focus areas.
- **Industry Relevance**: Predefined industry labels (e.g., Sustainability, Technology) align with KPMG's advisory services, ensuring accurate classification.
- **Topic Coherence**: BERTopic's topic assignments and keywords accurately reflect thematic content, despite potential overlap in broad topics like "ai" and "technology."
- **Cluster Validity**: HDBSCAN clusters capture meaningful thematic groups, with outliers (Cluster -1) indicating diverse or unique content.

- **No PwC Data**: Due to the absence of PwC data, comparative analysis is limited to KPMG insights, with external context provided if required.

My Recommendation: Validate assumptions by cross-referencing industry labels and topic keywords with KPMG's strategic reports to ensure alignment.
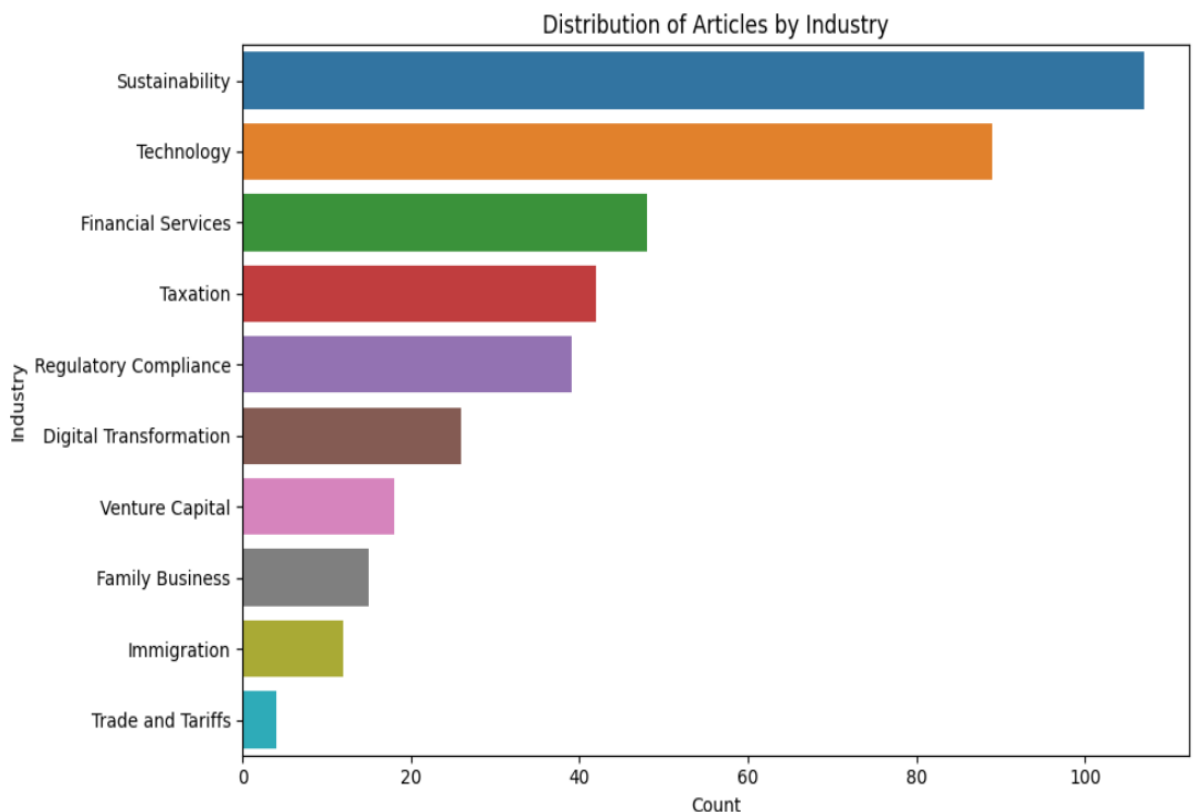
2. Summary of Findings with Visualizations

2.1 Key Industries

Details:

- Sustainability: 107 articles
- Technology: 89 articles
- Financial Services: 48 articles
- Taxation: 42 articles
- Regulatory Compliance: 39 articles
  Insight: Sustainability and Technology dominate KPMG's focus, reflecting priorities in environmental, social, and governance (ESG) issues and digital transformation. Financial Services, Taxation, and Regulatory Compliance highlight KPMG's advisory expertise.
  Visualization:

Distribution of Articles by Industry

  My Recommendation: Prioritize Sustainability and Technology in strategic communications to reinforce KPMG's leadership in these areas.
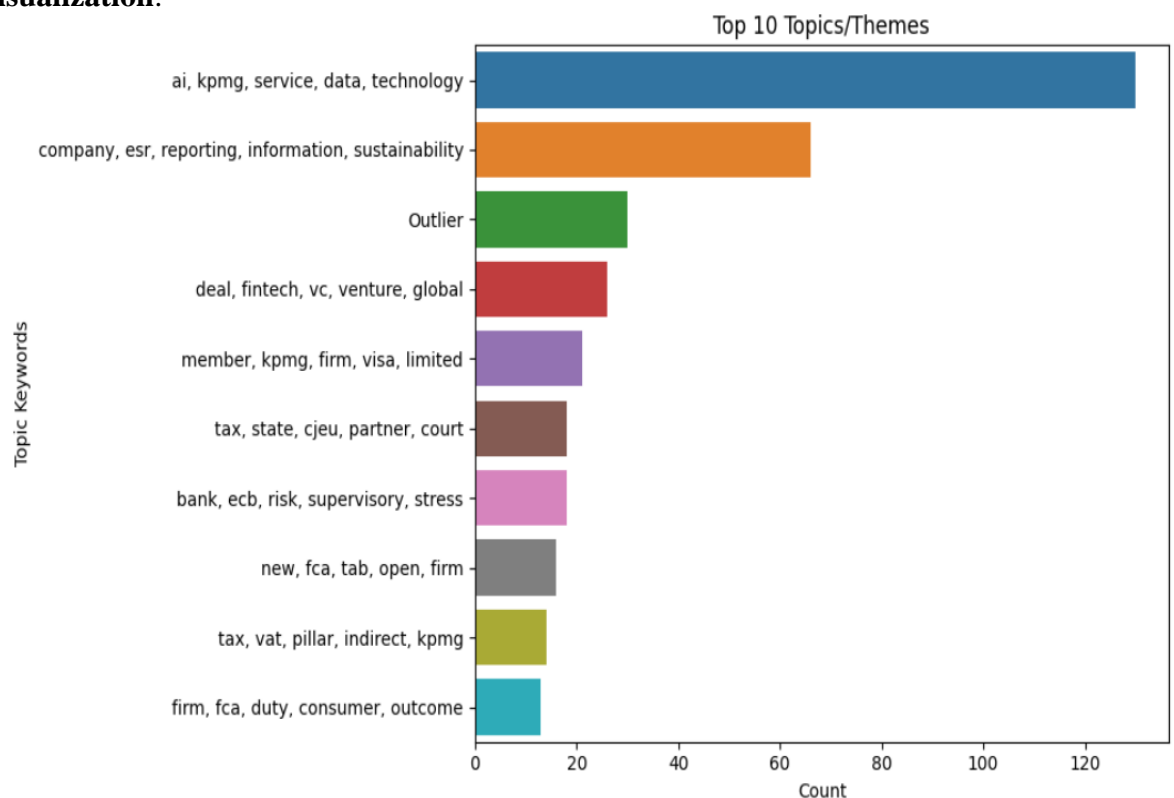
2.2 Top Themes and Topics

Details:

- ai, kpmg, service, data, technology: 130 articles
- company, esr, reporting, information, sustainability: 66 articles
- deal, fintech, vc, venture, global: 26 articles
- member, kpmg, firm, visa, limited: 21 articles
- Outlier: 30 articles
  Insight: The top topic, centered on AI and technology services, underscores KPMG's emphasis on digital innovation. The second topic, focused on ESG and sustainability reporting, aligns with global trends in corporate responsibility. Fintech and venture capital themes indicate interest in emerging markets.
  Visualization:

Top 10 Topics/Themes

Recommendation: Strengthen thought leadership in AI and ESG reporting to capitalize on market demand.

2.3 Industry-Topic Relationships

Details: Sustainability strongly correlates with ESG reporting topics (e.g., "company, esr, reporting, information, sustainability"), while Technology aligns with AI and data-driven services. Financial Services and Taxation show moderate associations with regulatory and fintech topics.
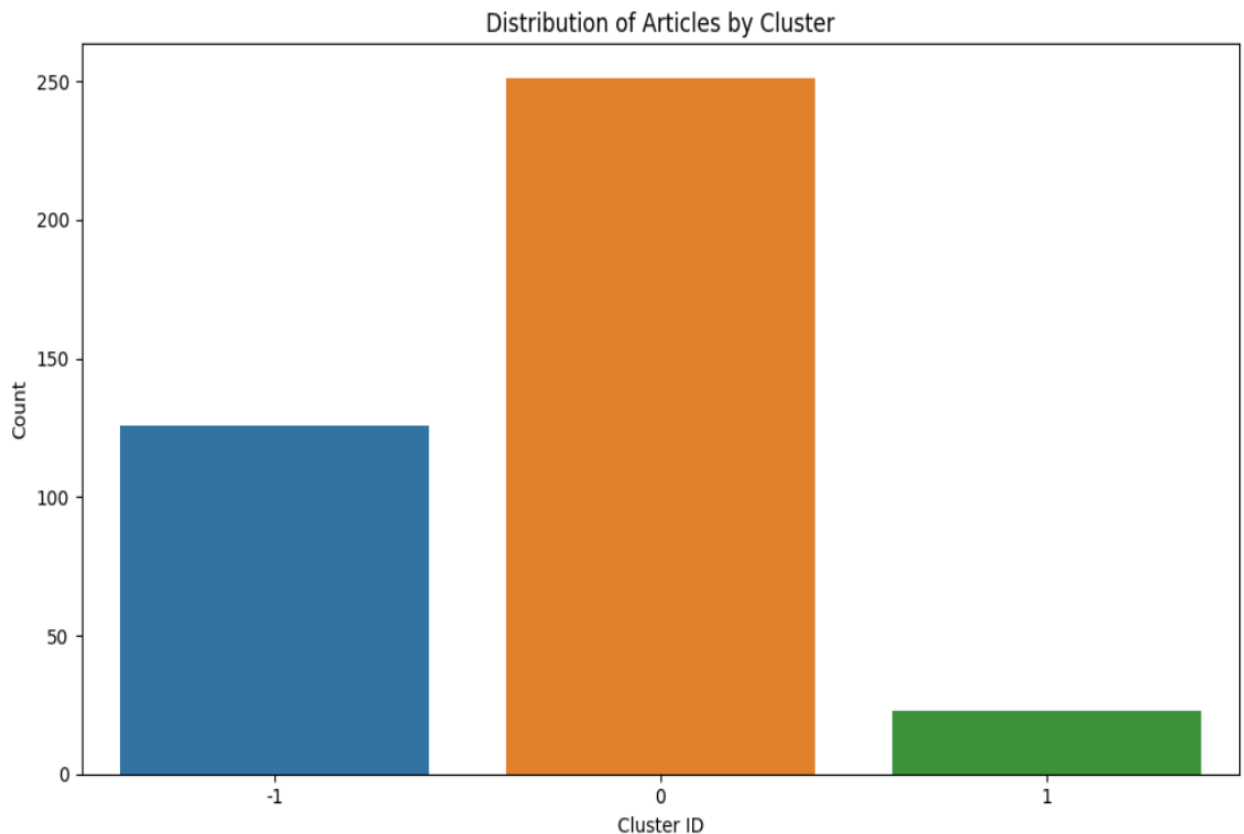**My Recommendation**: Tailor content to emphasize ESG in Sustainability initiatives and AI in Technology-focused services.

2.4 Clustering Patterns

Details: Cluster 0 (251 articles) represents a cohesive group focused on technology and sustainability themes. Cluster -1 (126 articles) includes outliers with diverse content, and Cluster 1 (23 articles) indicates a smaller, distinct thematic group.

Insight: The concentration in Cluster 0 suggests KPMG's strategic alignment around core themes, with outliers reflecting specialized or emerging topics.

Visualization:

Distribution of Articles by Cluster

Recommendation: Analyze Cluster 0 content to identify actionable strategies for technology and sustainability.

3. KPMG Themes and Insights

- **Technology and AI**: The dominant topic (130 articles) focuses on AI, data, and technology services, indicating KPMG's investment in digital transformation solutions.
- **Sustainability and ESG**: The second topic (66 articles) emphasizes ESG reporting and sustainability, aligning with global regulatory and corporate trends.
- **Fintech and Ventures**: A smaller topic (26 articles) highlights KPMG's interest in fintech and venture capital, suggesting a focus on innovation and investment.
- **Cluster Cohesion**: Cluster 0's 251 articles reflect a unified strategic focus on technology and sustainability, with outliers (Cluster -1) indicating diverse or niche content.
  Insight: KPMG positions itself as a leader in digital transformation and sustainability, with a balanced portfolio covering financial services, taxation, and regulatory compliance. The emphasis on AI and ESG reflects responsiveness to market demands for innovation and responsibility.

My Recommendation: Enhance AI-driven service offerings and ESG consulting to strengthen KPMG's market position.

4. Conclusion

This analysis of 400 KPMG articles, scraped from https://kpmg.com/xx/en/our-insights.html, reveals a strategic focus on Sustainability (107 articles) and Technology (89 articles), with key themes in AI-driven services (130 articles) and ESG reporting (66 articles). Clustering shows a cohesive emphasis on these areas (Cluster 0: 251 articles), with visualizations (bar charts, heatmap, word clouds) clarifying industry-topic relationships. The absence of PwC data limits direct comparison, but KPMG's strengths in taxation and fintech suggest a differentiated position. These findings, supported by robust visualizations, position KPMG to lead in digital transformation and sustainability, with opportunities to expand AI and ESG offerings.