**Báo Cáo Phân Tích Dữ Liệu Google Trends**

**1. Giới thiệu**

**1.1. Mục tiêu**

Mục tiêu của báo cáo này là phân tích sự quan tâm của người dùng trên Google đối với các từ khóa “Big Data”, “Machine Learning”, và “Artificial Intelligence” trong khoảng thời gian từ năm 2004 đến 2024. Chúng tôi sử dụng dữ liệu từ Google Trends để hiểu xu hướng tìm kiếm theo thời gian và so sánh mức độ quan tâm giữa các từ khóa này.

**1.2. Phương pháp**

Chúng tôi sử dụng Google Trends API để thu thập dữ liệu về mức độ quan tâm đối với các từ khóa. Sau đó, chúng tôi thực hiện các bước xử lý dữ liệu và trực quan hóa bằng cách sử dụng các công cụ phân tích dữ liệu như Python, Pandas, Matplotlib, và Seaborn.

**1.3. Công nghệ sử dụng**

**-** Trình thông dịch **Python.**

**-** Thư viện **Pandas, Matplotlib, Seaborn, Pytrends,...**

**- Jupyter notebook.**

**1.4. Video cài đặt môi trường**

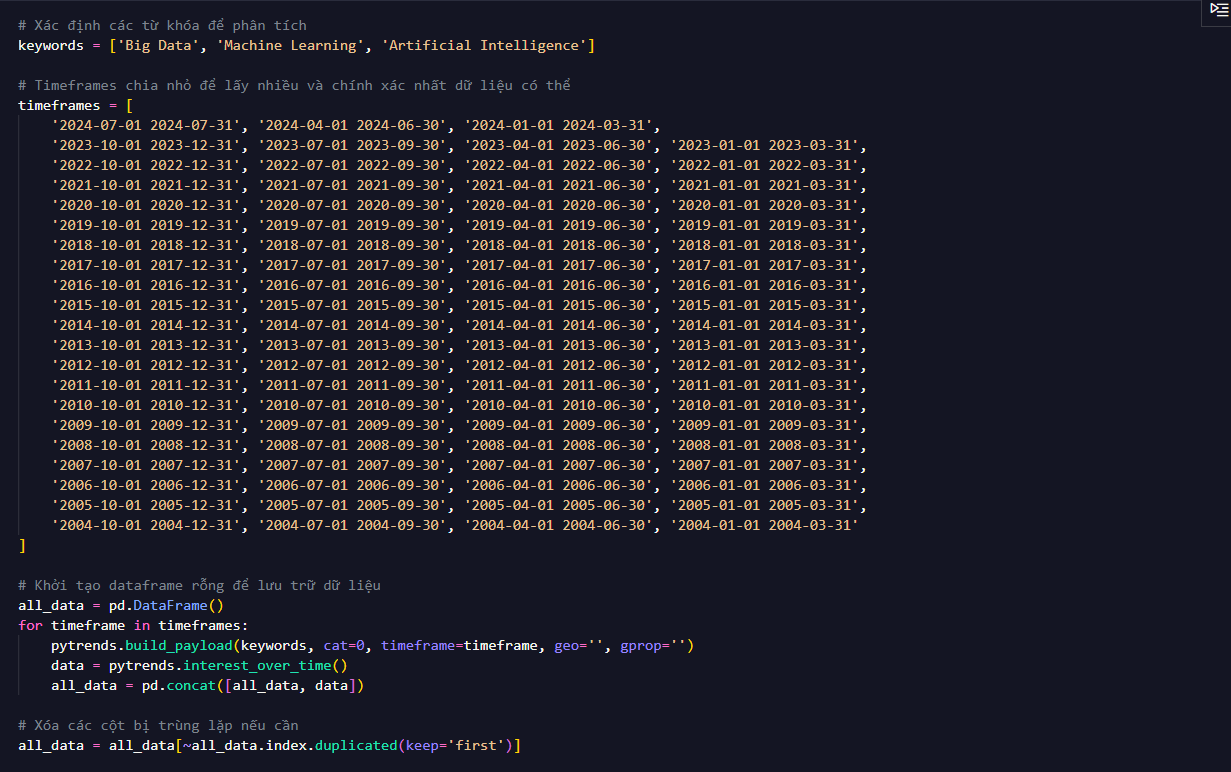
**[](https://www.youtube.com/embed/LDX9HLBXWtg?feature=oembed)**

**1.5 Báo cáo kết quả**

Link source code: <https://github.com/sheyy3103/BigdataProject>

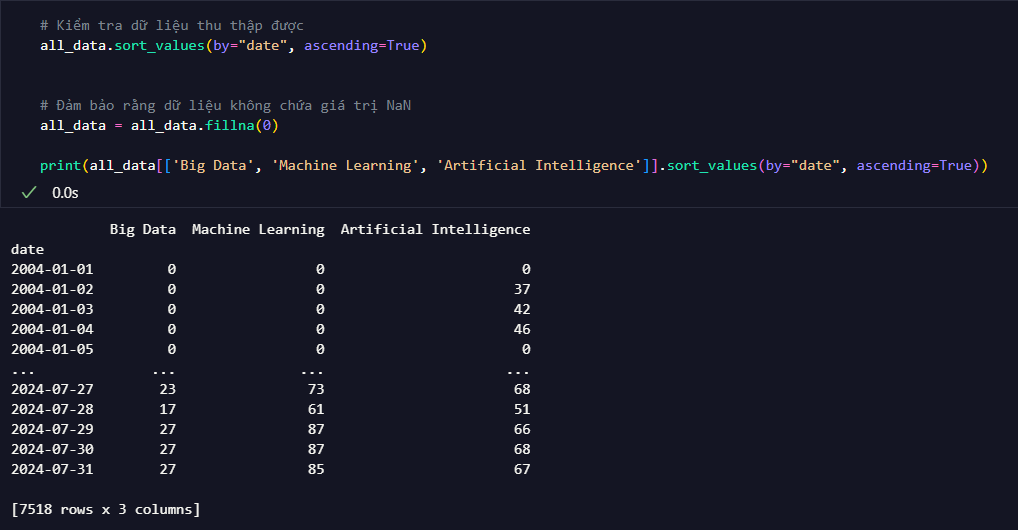
**2. Thu thập Dữ liệu**

Chúng tôi sử dụng thư viện pytrends để kết nối với Google Trends và thu thập dữ liệu cho ba từ khóa chính:



**3. Xử lý Dữ liệu**

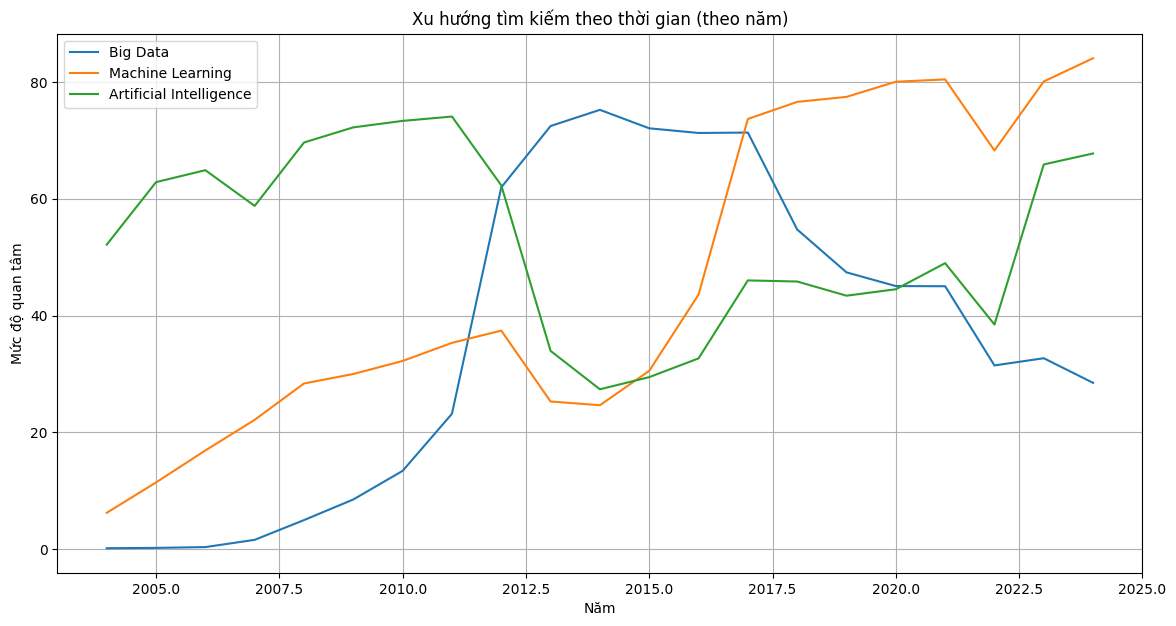
Chúng tôi kiểm tra dữ liệu thu thập được và xử lý các giá trị thiếu hoặc không hợp lệ:



**4. Phân Tích và Trực Quan Hóa Dữ Liệu**

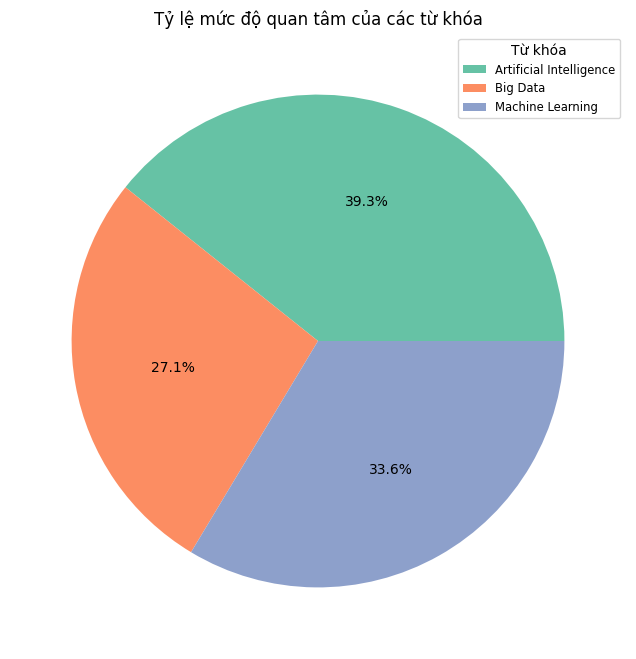
**4.1. Biểu Đồ Xu Hướng Tìm Kiếm**

Chúng tôi vẽ biểu đồ xu hướng tìm kiếm theo thời gian cho từng từ khóa bằng cách tính toán mức độ quan tâm theo năm:



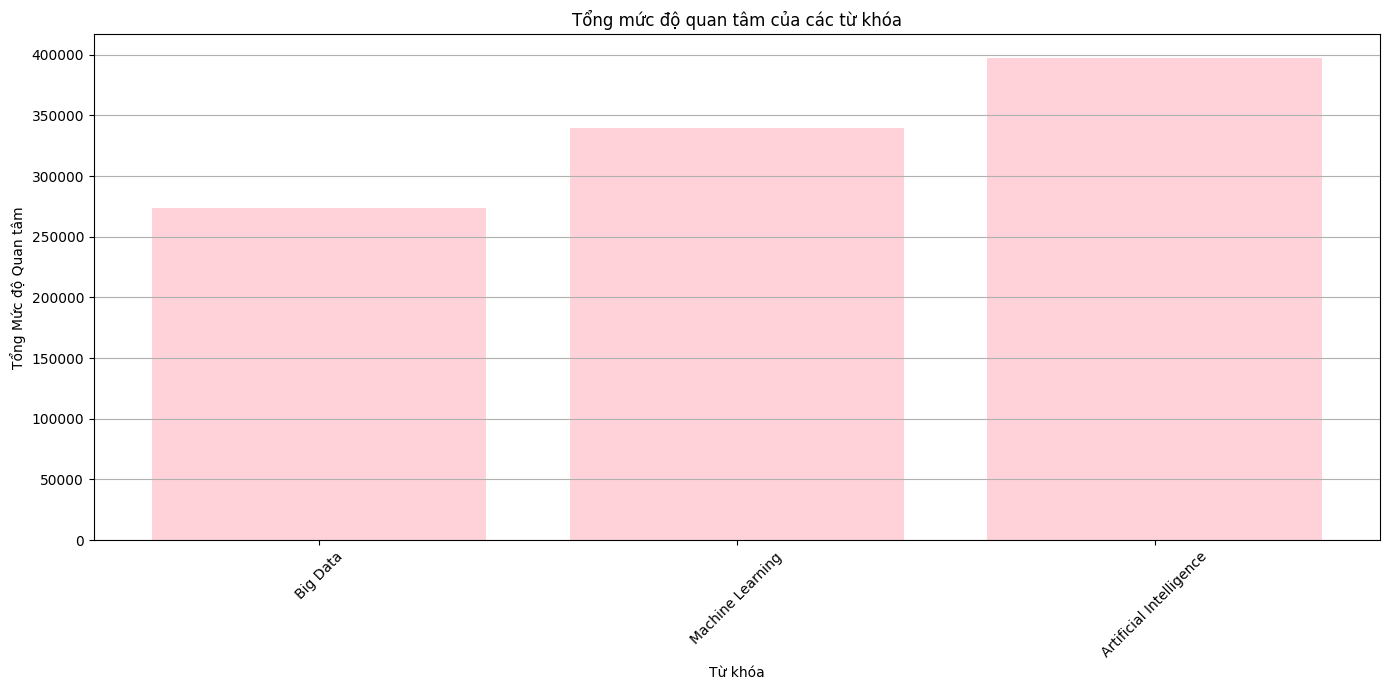
**4.2. Biểu Đồ Tỉ Lệ Mức Độ Tìm Kiếm**

Chúng tôi tính toán tỷ lệ phần trăm của tổng mức độ quan tâm cho các từ khóa và vẽ biểu đồ tròn:



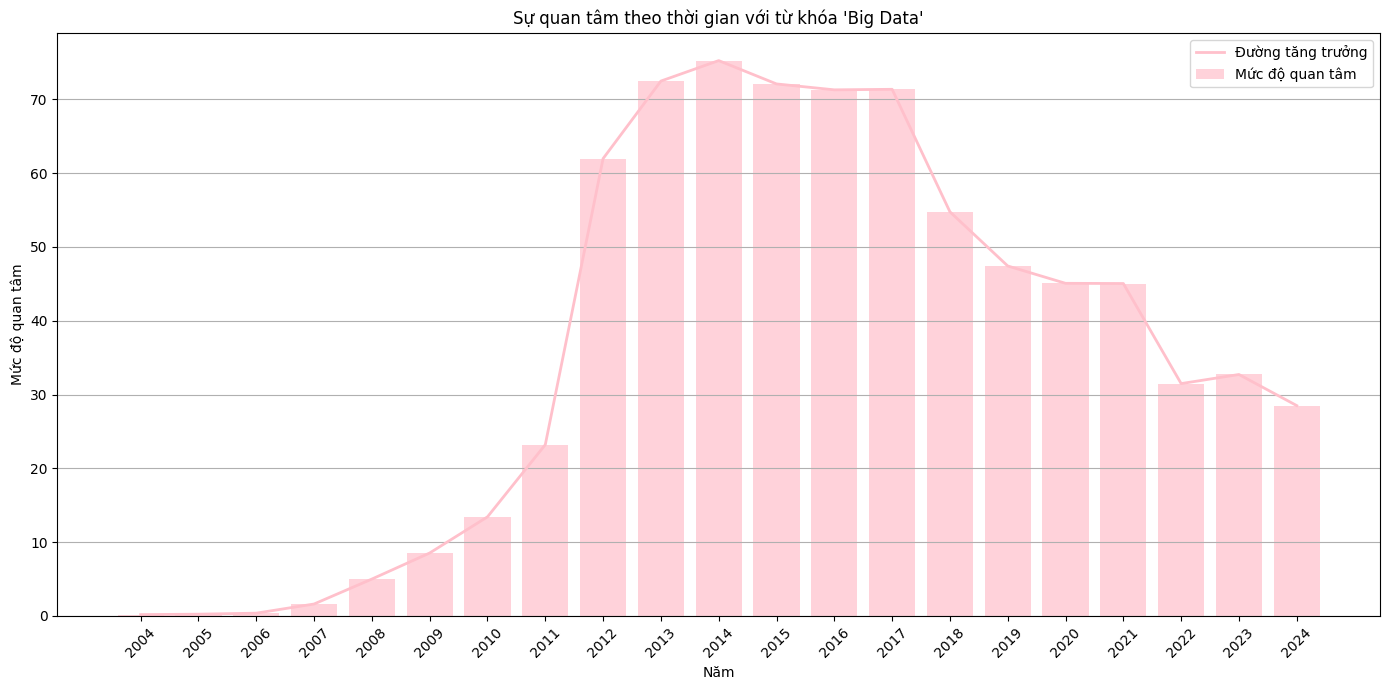
**4.3. Biểu Đồ Tổng Mức Độ Quan Tâm**

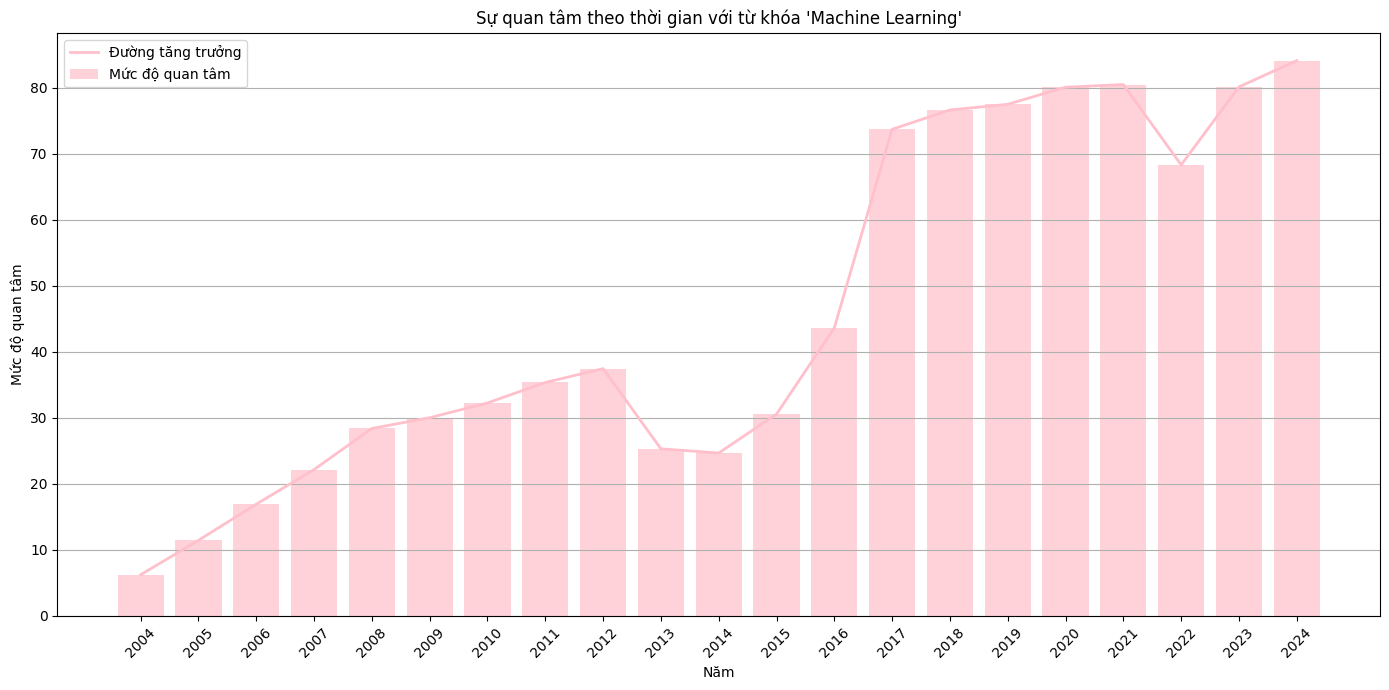
Chúng tôi tính toán tổng mức độ quan tâm của từng từ khóa và vẽ biểu đồ cột:

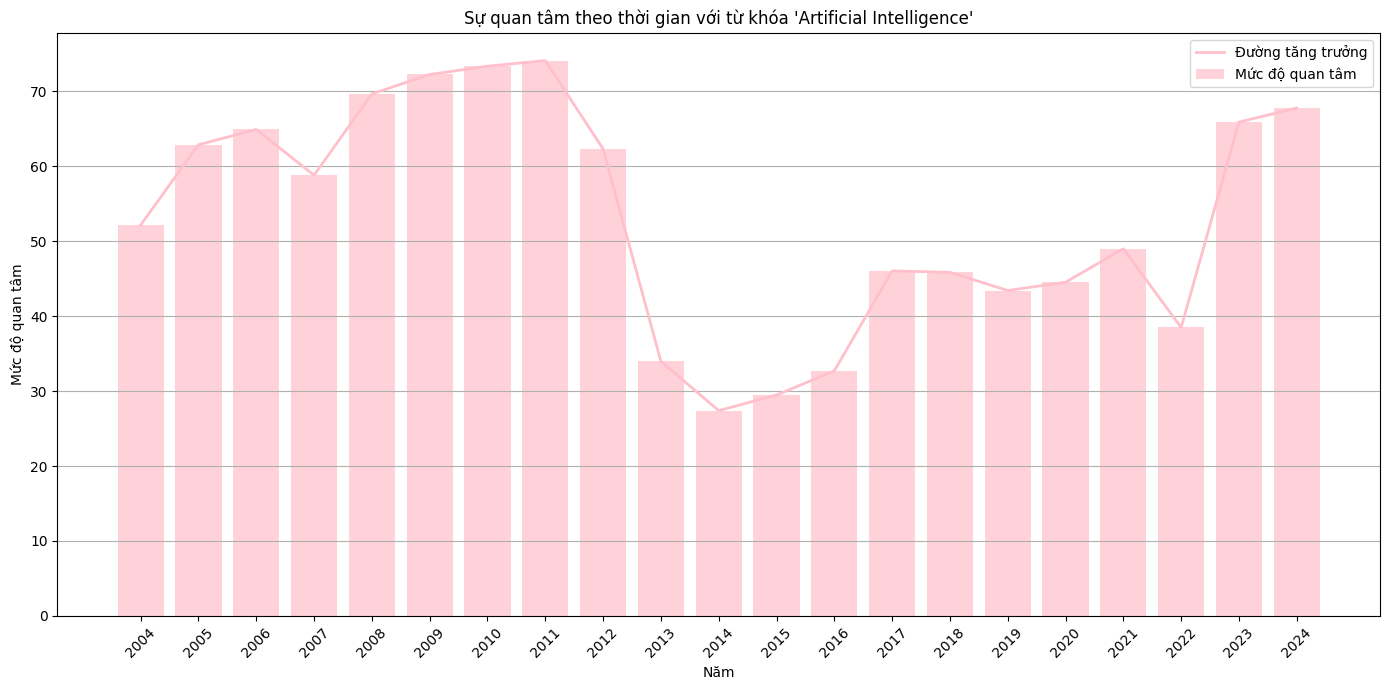


**4.4. Biểu Đồ Thể Hiện Sự Quan Tâm Theo Thời Gian**

Chúng tôi tạo biểu đồ cột cho từng từ khóa và thêm đường tăng trưởng:

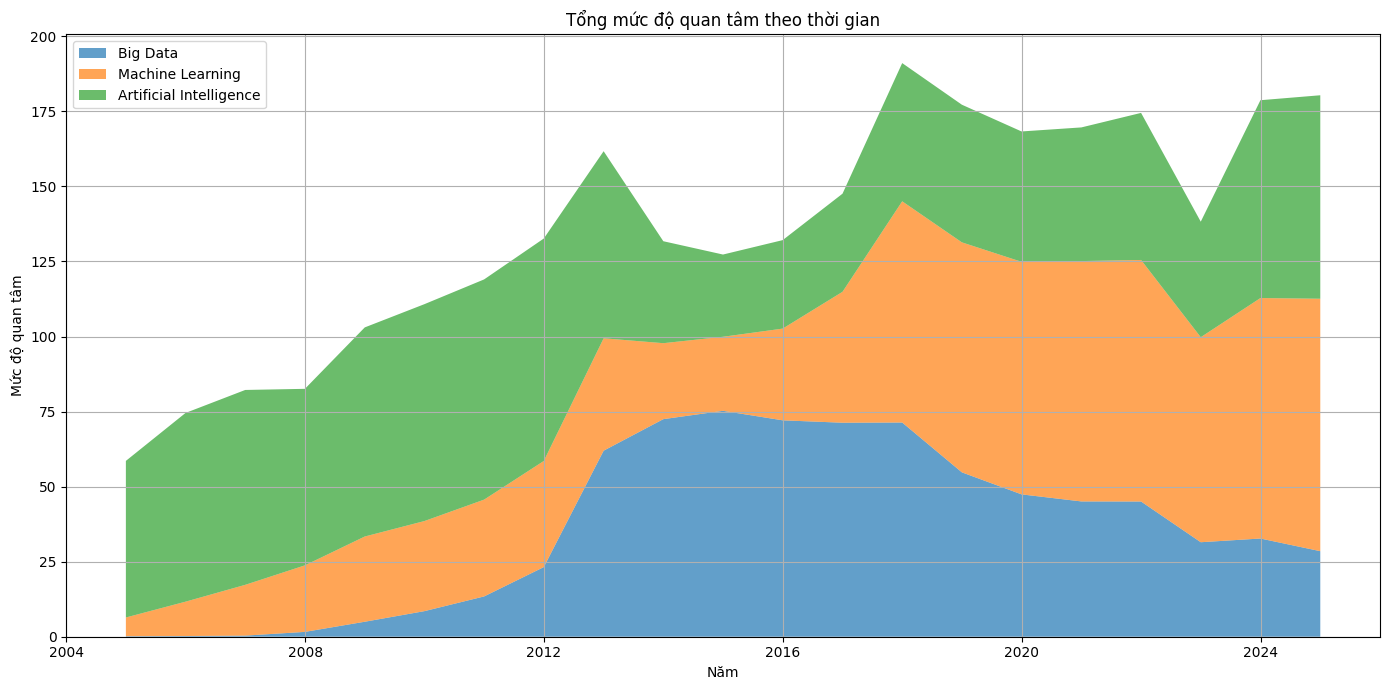






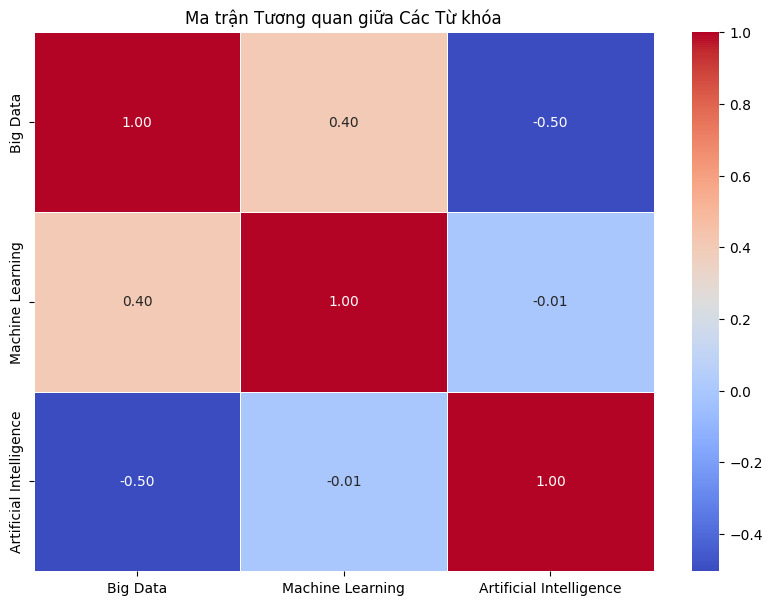
**4.5. Biểu Đồ Tầng Mức Độ Quan Tâm**

Chúng tôi vẽ biểu đồ tầng để thể hiện tổng mức độ quan tâm theo thời gian:



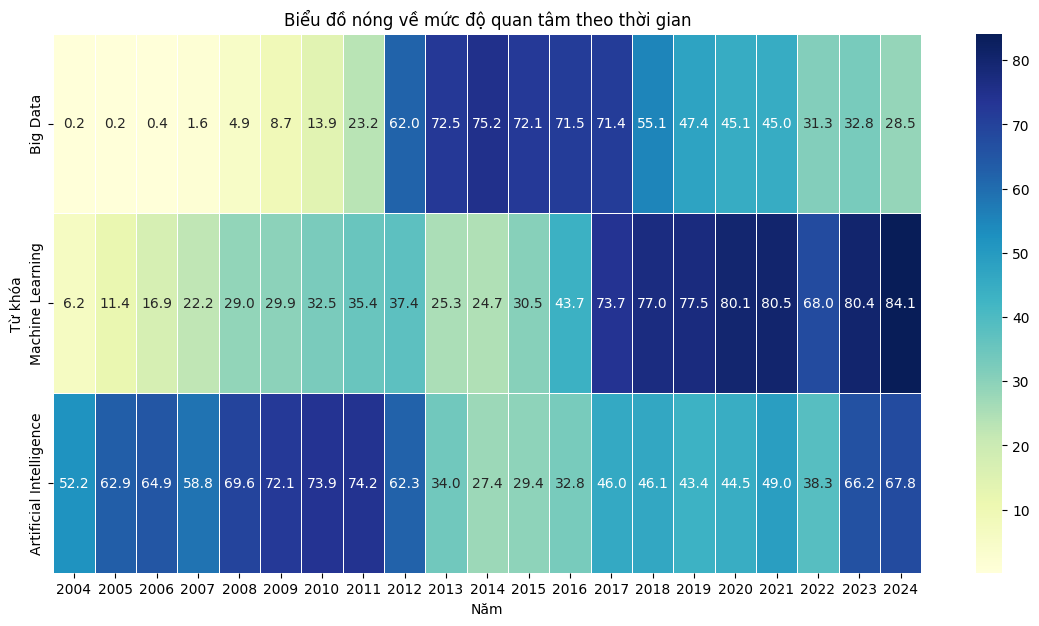
**4.6. Biểu Đồ Ma Trận Tương Quan**

Chúng tôi tính toán ma trận tương quan và vẽ heatmap để phân tích mối quan hệ giữa các từ khóa:



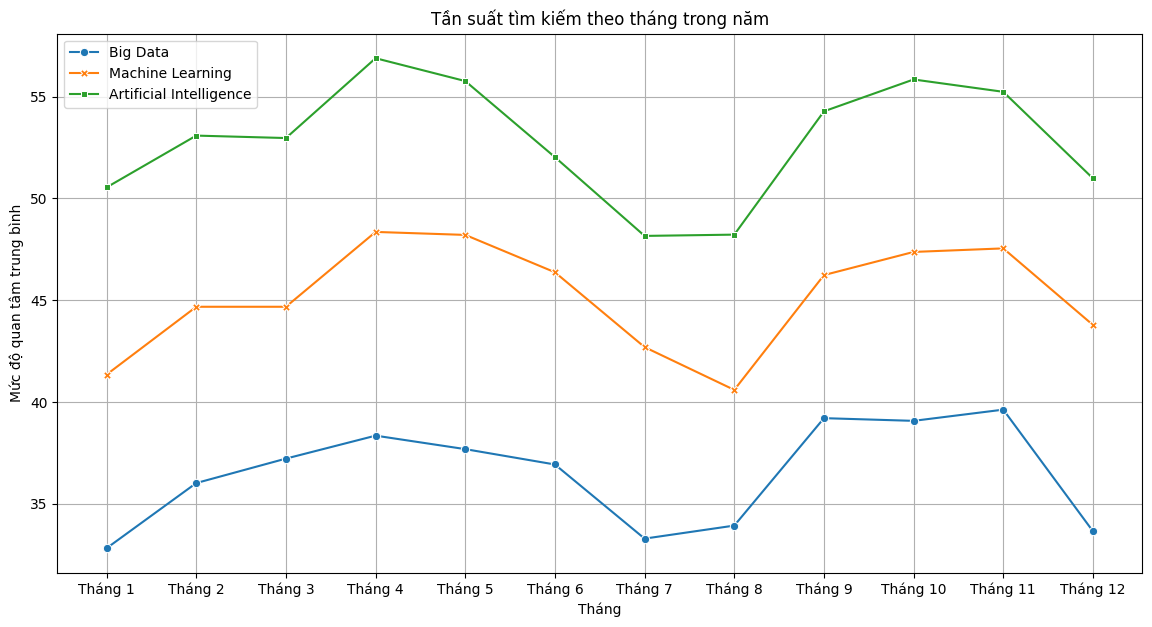
**4.7. Biểu Đồ Nóng**

Chúng tôi vẽ heatmap để thể hiện sự quan tâm theo thời gian cho từng từ khóa:



**4.8. Biểu đồ tần suất tìm kiếm theo tháng trong năm**

Chúng tôi vẽ biểu đồ đường thể hiện trung tần suất tìm kiếm của mỗi tháng trong năm:



**5. Kết Luận**

* **Xu hướng tìm kiếm:** Từ khóa “Machine Learning” có xu hướng tăng trưởng nhanh nhất trong khoảng thời gian phân tích, trong khi “Big Data” và “Artificial Intelligence” có sự quan tâm ổn định hơn.
* **Tỷ lệ mức độ tìm kiếm:** “Artificial Intelligence” chiếm tỷ lệ cao nhất trong tổng mức độ quan tâm so với các từ khóa khác, “Machine Learning” cũng xấp xỉ trong khi đó “Big Data” thấp hơn hẳn so với 2 từ khóa còn lại.
* **Sự quan tâm theo thời gian:** Các từ khóa đều cho thấy xu hướng tăng trưởng trong các năm gần đây, ngoại trừ với “Big Data” có sự thuyên giảm về mức độ quan tâm.
* **Ma trận tương quan:** Có sự tương quan tích cực giữa các từ khóa, đặc biệt là giữa “Machine Learning” và “Big Data”.
* **Tần suất tìm kiếm:** Các từ khóa đều có tần suất tìm kiếm lớn vào các tháng 4, 9, 10, 11 và nhỏ hơn ở các tháng còn lại đặc biệt là ở tháng 1, 7, 8 và 12.

**6. Đề Xuất và Hướng Phát Triển**

* **Mở rộng Phân Tích:** Có thể mở rộng phân tích để bao gồm nhiều từ khóa hơn hoặc các yếu tố khác như khu vực địa lý.
* **Cập Nhật Dữ Liệu:** Nên cập nhật dữ liệu thường xuyên để theo kịp xu hướng hiện tại.
* **Ứng Dụng Machine Learning:** Áp dụng các kỹ thuật học máy để dự đoán xu hướng trong tương lai.

**7. Tài Liệu Tham Khảo**

* [Google Trends API Documentation](https://colab.research.google.com/github/Tanu-N-Prabhu/Python/blob/master/Google_Trends_API.ipynb)
* [Matplotlib Documentation](https://matplotlib.org/stable/index.html)
* [Seaborn Documentation](https://www.w3schools.com/python/numpy/numpy_random_seaborn.asp)
* [Pandas Documentation](https://pandas.pydata.org/docs/index.html)