

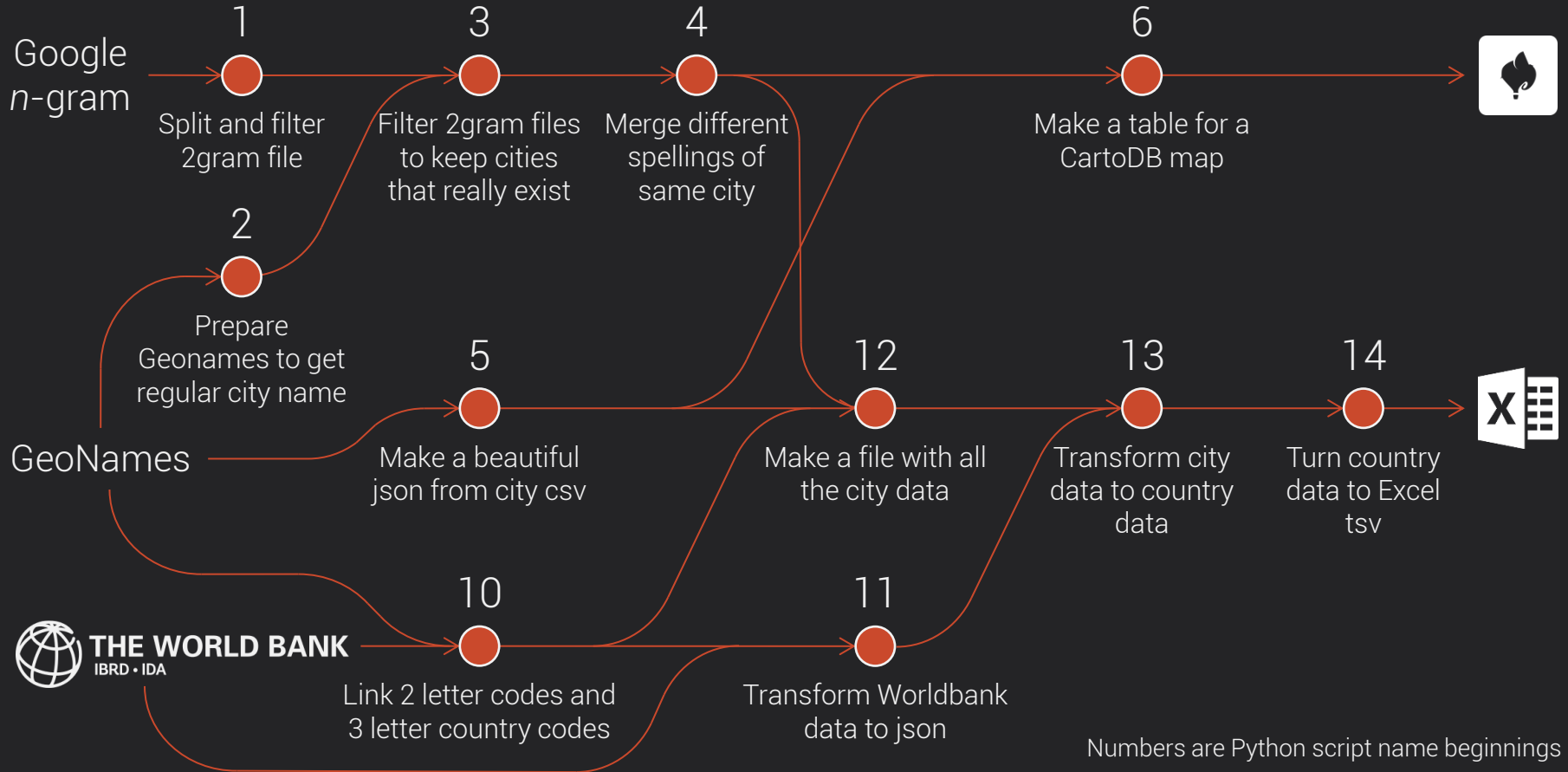


Data Analysis - Cities in English books

3 databases

Google Ngram	2-gram with 'to' English corpus Years of books releases (until 2012)
GeoNames	Cities and alternative names Longitudes and latitudes Country codes Population More than 100,000 inhabitants : +6,000 cities
The Worldbank	International tourism Number of arrivals Country names and codes Data between 1995 and 2008

Data mining: script process

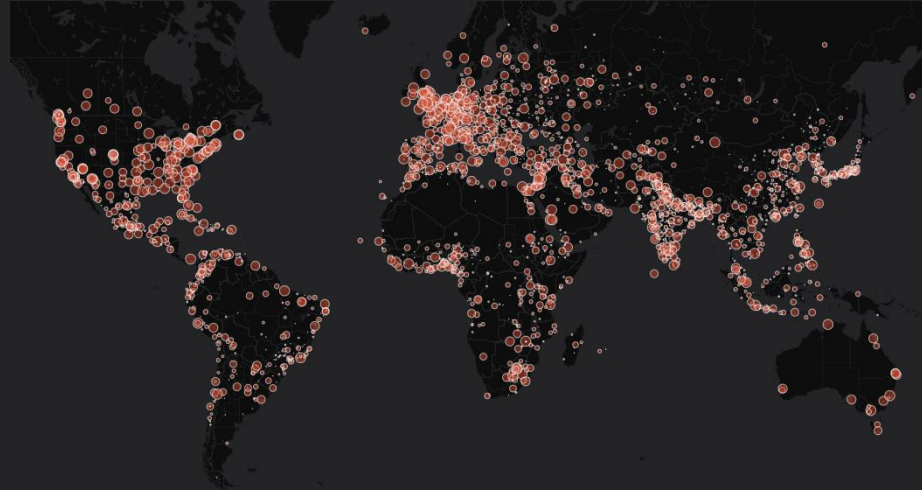


City occurrences

1950-1959



2000-2009



Hypothesis

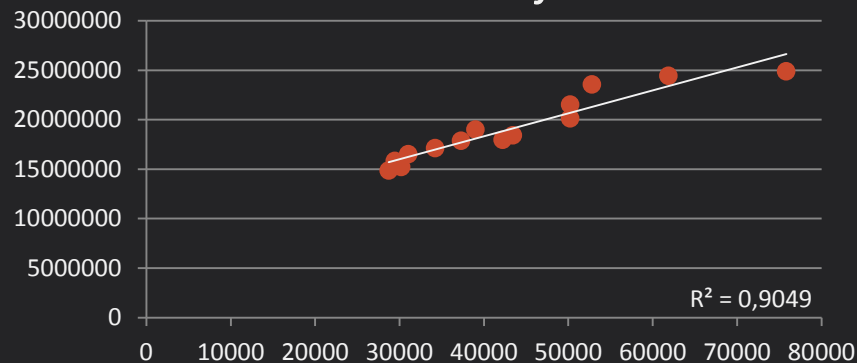
Tourism is linked with the number of times a place is quoted.

Either tourism creates documents talking about the visited places or talking about somewhere makes people go to it.

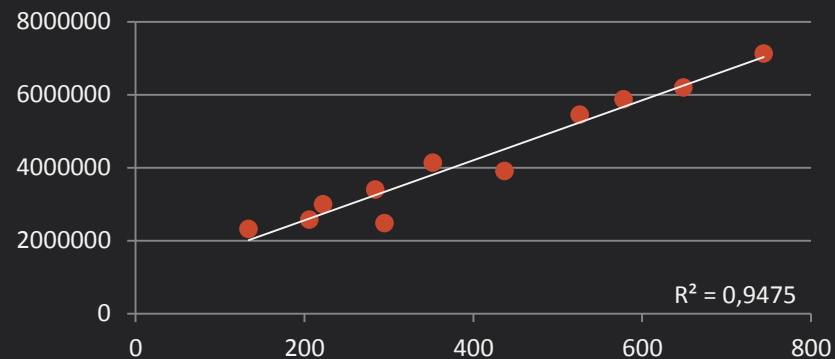
Is the number of international arrivals in countries linked with the number of occurrences in books ?

Countries where there is a correlation between occurrences in literature and tourism

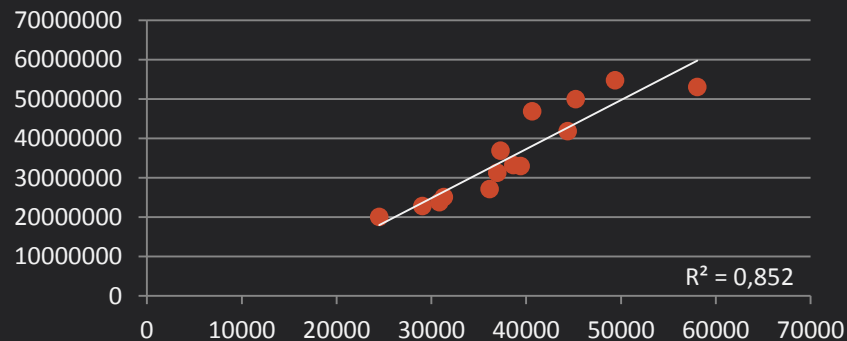
Germany



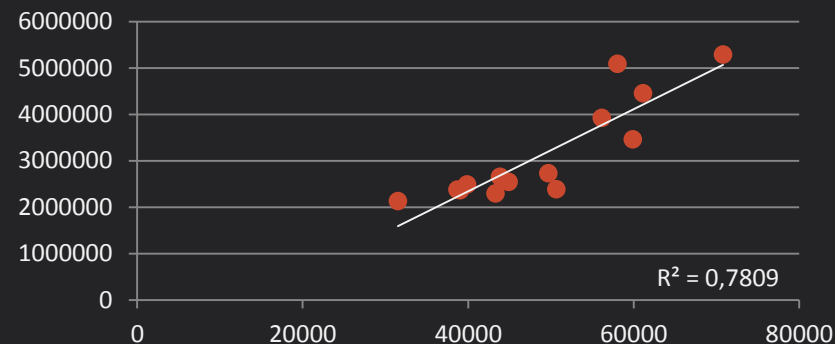
United Arab Emirates



China



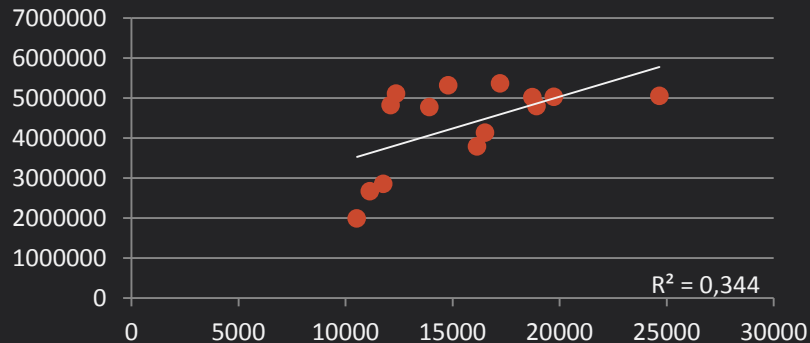
India



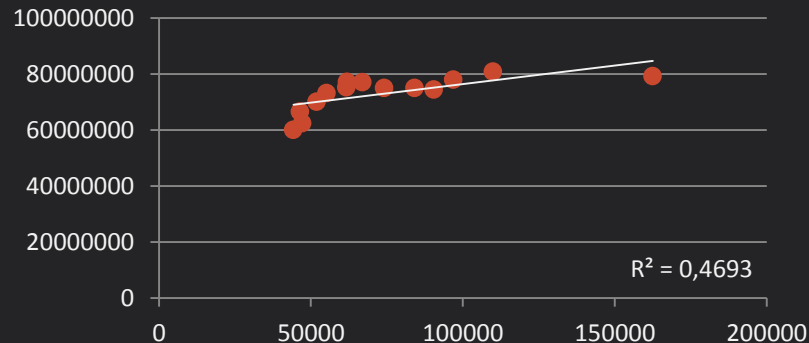
x: Number of occurrences in literature (per year) | y: Number of arrivals, international tourism (per year)

Countries where occurrences in literature and tourism don't seem to be linked

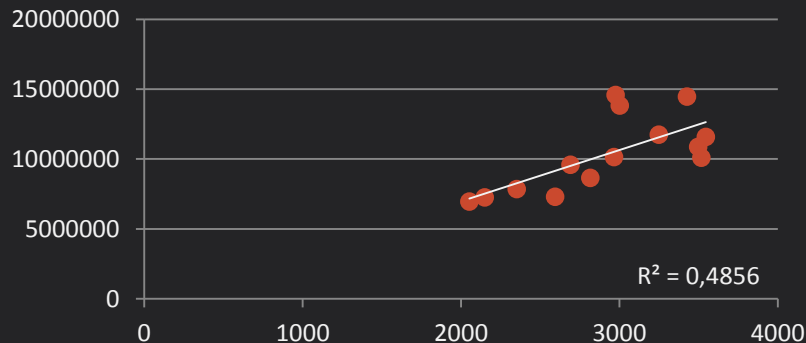
Brazil



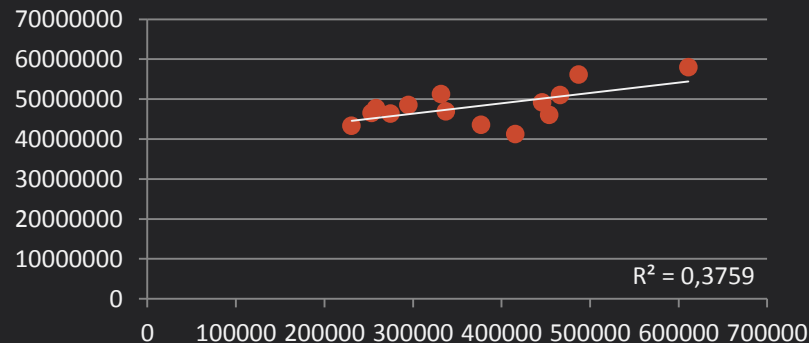
France



Thailand



USA



x: Number of occurrences in literature (per year) | y: Number of arrivals, international tourism (per year)

To go further - Corpus

Seek other n-gram : other locative prepositions (in, from, by, through...) to gather more cities.

Diversify the sources: Use data from the internet, review websites, social networks, articles in newspapers...

To go further - Tourism data

Data from cities, not only from whole countries, to be more precise on correlation between tourism and city name appearance in books.

Use data from the World Tourism Organization (UNWTO) (not free of charge).

Find data on a wider period.

Find a Python code to make automatic linear regression.

Business

With further studies and data analysis about tourism and cities popularities, we should be able to help territorial authorities to develop their attractiveness and promote their images.

Target customer:

Cities - Regions - Countries - World Tourism Organization (UNWTO)

Private tourism agencies

Other companies linked with the tourism market