COMS4060A/7056A: Assignment #1

Tim Bristow
tim@bristow.za.net

University of the Witwatersrand — November 20, 2020

1 Introduction

The Skyscraper Index (https://en.wikipedia.org/wiki/Skyscraper_Index) is sometimes used as a metric for determining the status of an economy, in particular, predicting when there will be an economic downturn. The premise is that the tallest skyscrapers are built during times of economic growth, and this stops as the economy goes into recession. There will be a lag, however, as construction projects are often finalised years in advance, contracts signed, and significant amounts of money have already been paid.

You are going to investigate whether there is any evidence of the Skyscraper Index working empirically.

You can use any programming language or tool you would like. However, I strongly recommend using Python, as I have provided a Jupyter notebook to get you started.

2 Datasets

I have created a dataset containing 6000 skyscrapers. This contains the name of the building, the city, the height in metres (and feet), the number of floors, the year of completion (or cancellation), and the status of the building (completed, under construction, cancelled, etc.).

The dataset has been augmented by geocoding the cities and buildings. The country, and the latitude and longitude of the city have been added. The geocoding was done using OpenStreetMaps, which is free and community sourced - the accuracy of the data is not guaranteed. In particular, the country names are returned in the official language of the country, and not English.

This augmented dataset has been partially cleaned. The country names have been converted to English, some missing fields have been added, and datatypes have been updated. Each step of the cleaning is detailed in the notebook below.

The Index also requires GDP data. I have sourced a dataset for you from the Worldbank website, which contains several GDP metrics per country between 1960 and 2019. I have done some basic cleaning on the dataset too - something you will notice is that the country names do not align exactly with this dataset and the skyscraper dataset. This is a normal problem you will run into when merging two disparate datasets. I have detailed how I went about merging this, but it was a manual, once-off process.

3 Questions

3.1 US average heights

For the buildings in the USA, create a chloropleth with the average height of all completed skyscrapers per state. Hint: look at the osmnx python package to do this. Plot New York, LA, Chicago, San Francisco and Seattle on the map too, with their max height building.

3.2 European buildings

Europe has relatively few skyscrapers, but their GDP per capita is high. Look at the number of skyscrapers in Germany, France, UK, Switzerland, and the Netherlands. Then compare this against the UAE, Japan, and the USA. Can you provide a reason for why this is the case?

3.3 Economic growth vs Buildings

For the following questions, look at the following 5 countries: SA, USA, UK, UAE, China.

3.3.1 GDP per capita

Look at the GDP per capita for these countries and plot this against

- a) the tallest building per year,
- b) the number of new skyscrapers built in a year,
- c) the number of cancelled buildings in a year.

Is there any obvious correlation between GDP and a), b) and c)? You do not need to do any statistical tests - a couple of sentences is fine, just back it up with the data in the plot. Any correlation metrics or plots would be beneficial. Hint: look at the GDP growth as well.

** Missing data: Not all buildings have a height. How else can you show the height of the building?

3.3.2 Impact of recessions

For the chosen countries, find out when they had recessions, and add this as an additional point on the plot (single datapoint, you do not have to indicate the full length of the recession). Is there any trend or correlation between the skyscraper data and the recessions? As above, you only need to provide a written answer with motivation (and the source for the recession data - Wikipedia is fine).

3.3.3 An anomaly

Of the 5 countries, 1 of the countries shows behaviour that is completely at odds with the Skyscraper Index. Which one is it? Any idea why?

3.4 Submission

Work in groups of up to four people. Submit your work to Moodle as a PDF or Jupyter notebook.

Deadline: 30 November 2020