Using Data Science to Visualize and Model Forecasts for Ukraine’s Economic Future

After having been living in Ukraine for over a year, I decided that I wanted undertake a personal project to showcase what I had learned so far using different methodologies and technologies in the realm of data science. The internet is replete with courses and approaches to

Before diving too far into the content of this post, I should emphasis that the primary purpose of this article is not to extensively study of economic theorem nor make scholarly predictions but instead to use data science approaches to make forecasts from a dataset. This series of posts does not aim to be academic and strives to use a layman’s understanding of economics.

Determining Economic Indicators

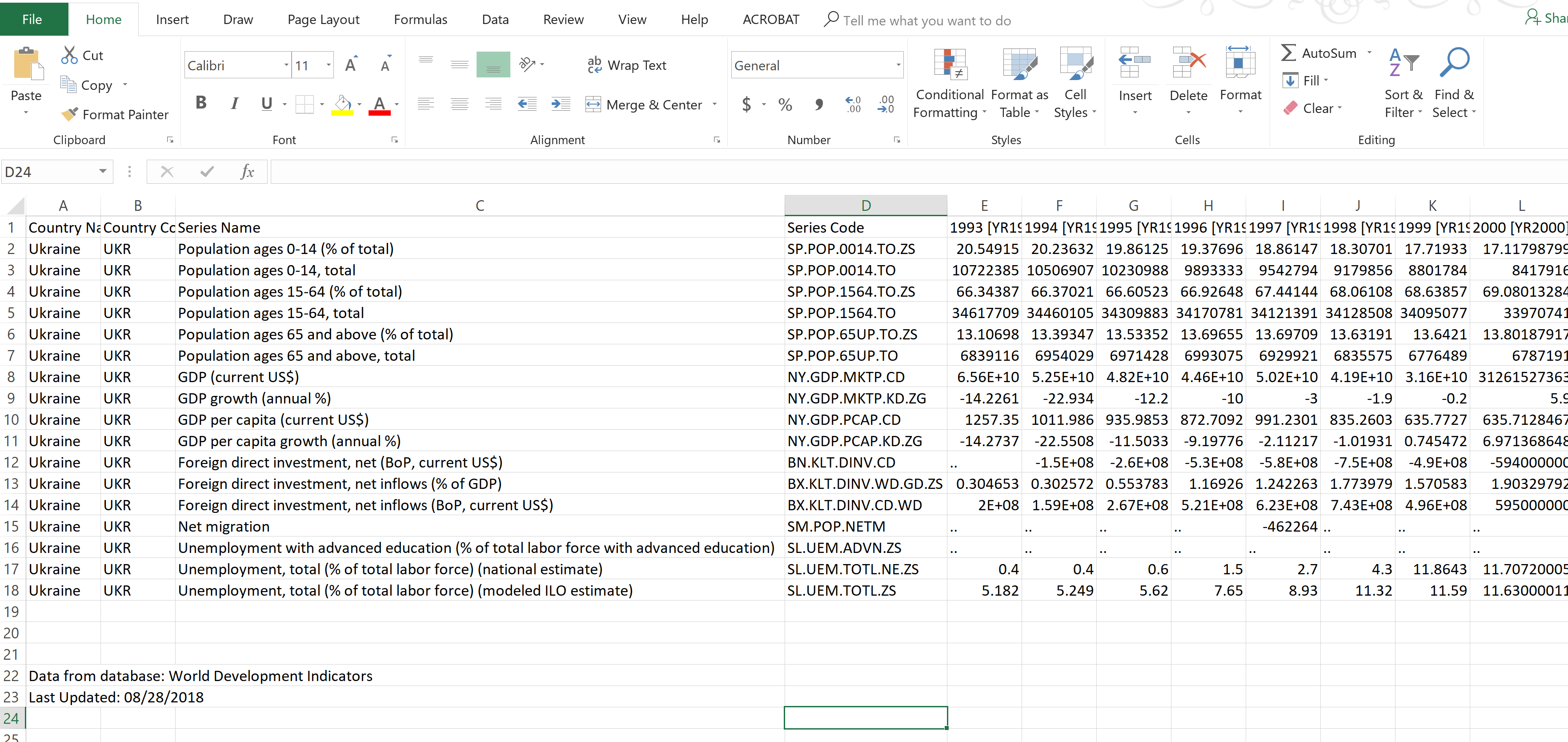
I should preface this article by admitting that I am by no means an expert on economics. The little bit I do know, however, gave me enough of a foundation to get started. I wanted to start with looking at Ukraine’s population trajectory. I decided to look at population in different segments to better understand population as an economic indicator. For example, does a youth deficiency exist in Ukraine? What portion of the population will be economically productive in the coming years in decades? While GDP and GDP per capita are essential indicators, I also looked at consumption, migration, investments as well as looking at to less graphable indicators such as pollical climate.

Data Collection

I chose to use economic data as my initial project largely due to readily available complex datasets. Economic data is available from a multitude of scholarly sources; however, the World Bank’s World Development Indicators portal easily offered the most robust and user-friendly option for data collection. Although I had been introduced to various techniques of data scrapping and using APIs, working with an already familiar format, in this case, Excel files, seemed to be the best approach to get started.

Data Munging

The World Bank portal offers intuitive export options after the desired criteria is met. I was able to export my chosen selection in a format that required minimal wrangling.



The above screen shot from the WB portal shows all of my selections compiled all of my selections onto one Excel tab.

The only real task with the Excel sheet was to split the various economic indicators onto single tabs to allow for easier access with Python.

At this point, the population segment is saved as a single CSV file and ready for processing. From this point on, we will be conducting analyses with Python 3 using the following packages: Pandas, NumPy and Matplotlib for graphing.

Economic Indicator # 1: Population Growth

We are going to start by looking at the trends of Ukraine’s population over the last three years by creating three population divisions.

Matplotlib contains several built-in graphs such as histograms and bar charts, however the default plot option works perfectly to provide a change in population over a period of time along the x-axis.

In the post below, I’ve taken the data of population change for the youngest demographic of Ukraine’s citizens, ages 0-14. While this age represents the least productive demographic of Ukraine’s populace, it shows the trend since the early 1990s has been moving in the direction of youth deficiency.

While this does not bode well for the future of Ukraine’s workforce, we need to look ahead to get a better idea of the future of Ukraine’s youngest demographic.

Feature Engineering and Modeling Predictions

In the realm the data science, feature engineer refers to the ability to apply knowledge from a dataset, known as domain knowledge, to create new features with the aim of making more accurate predictions. This process is known as machine learning which uses algorithms to determine the outcome of an event.

In the context of predicting youth demographic changes, we would pull different indicators from our dataset. To be able to accomplish this with any accuracy, we would need to have additional indicators such as childhood mortality rates, access to healthcare, AID/HIV and disease prevalence, among other factors. If we had access to this data, we could likely engineer new features that would allow for some insight into the future of Ukraine’s youth trajectory.

While we will cover feature engineer later in this blog, this series of posts is limited to an introductory context. Fortunately, the United Nations provides datasets on the future of demographic changes. [Their portal](https://esa.un.org/unpd/wpp/DataQuery/), the Department of Economic and Social Affairs, provides age and gender based population predictions until 2100.