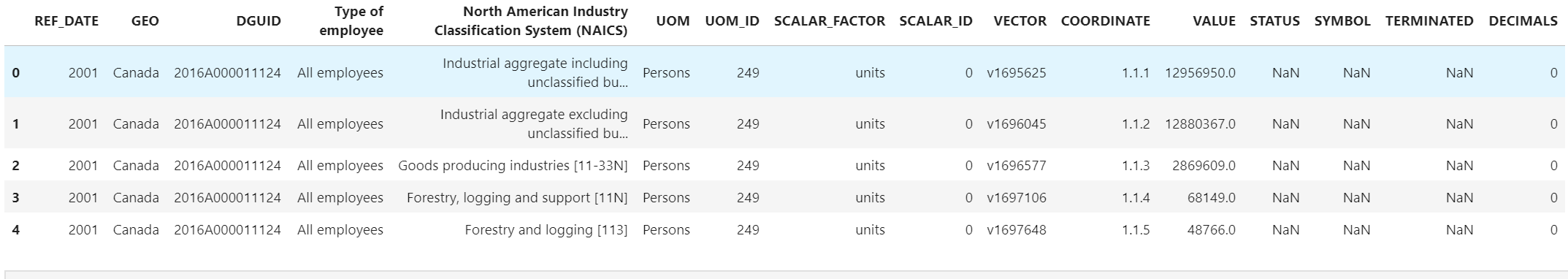
# The top Canadian industries and business sectors in 2030

## Background

The economic growth rate of each industry can be a key criterion for financers and investors to make long-term investment in each sector. The future perspective of each industry can also have an impact on the carrier path that young people are making before going to college or university. Moreover, there are millions immigrating to Canada or moving within the country, most of which have difficulties making an informed decision about 1) what province is most suitable for their carrier within the next 10 years; 2) How does an industry is performing in each province compare to others; and 3) whether their industry is growing or dying in their province. Therefore, it is important for individuals in the their early carriers to predict the future trajectory of their business and industry, before making a long-term investments. For example, this information can help a new immigrant decide whether they should immigrate to Canada, and if so, to which province.

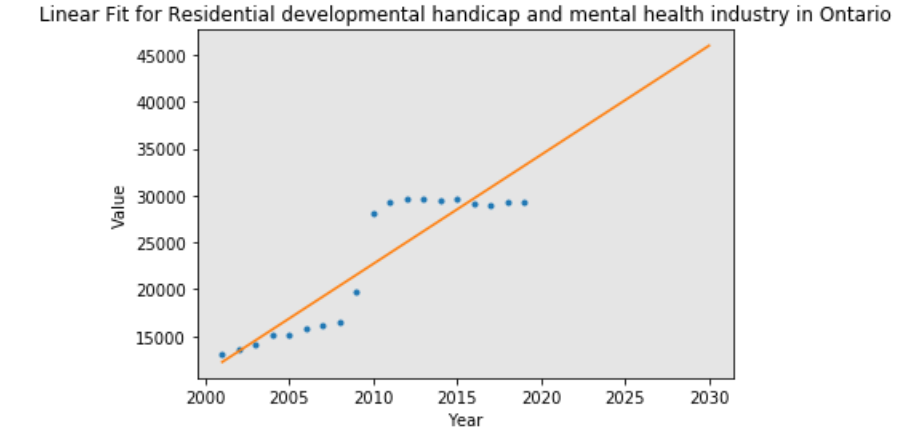
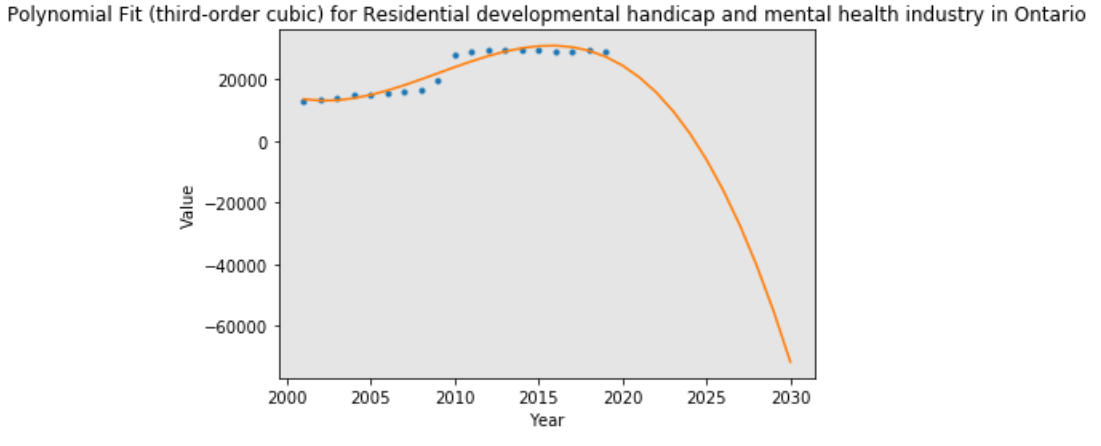
## Methodology

Historical data about the value of over 360 industries and financial sectors from 2001 to 2019 is used to determine the future trend of each industry/financial sector for 10 Canadian provinces as well as the whole country. These data are available online at [https://www150.statcan.gc.ca](https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=3610048801) . The dataset used in this study can be downloaded in a zip file from [here](https://www150.statcan.gc.ca/n1/tbl/csv/14100202-eng.zip).



The dataset downloaded from the government Canada website include several missing values for some of the industries in different years. I decided to remove this datapoints as the industry value is the feature that I aim to predict and the rows that miss these data cannot be used any further. The government website also recorded a measure for the accuracy of the estimated industry values, from “A” for most confident estimates to “F” for the least confident estimates. I used all the data regardless of their accuracy to model the trajectory of each industry because the number of data point with low confidence was small and it was not worth the large effort to assign an accuracy for each data point. Further, the dataset separates the employees to two group: 1) employees paid by the hour, and 2) employees paid a fixed salary. I did not separate the industry values by the type of employee and used the total value of the industry to model its growth rate. I initially used the cubic (third order) polynomial regression model to fit a model into the historical data from 2001 to 2019. However, the model was overfitting so I continued by first order linear regression to find the industry absolute value in the year 2030 and the industries’ growth rate compared to 2019, which we have the last estimated value for each industry /financial sector.

Figure 1: Third-order cubic polynomial (upper figure) and linear Regression (lower figure) model to predict the industry values by 2030. Linear regressuion model shows to provide a better prediction in my study.



I used the predictions based on the polynomial regression models for over 360 industries to define 1) the top 5 industries by their absolute value for each of the 10 provinces in 2030; 2) the top 5 industries that show the highest growth rate for each of the 10 provinces within the next 10 years; and 3) the bottom 5 industries that show the highest shrinkage (negative growth) rate for each of the 10 provinces in the next 10 years. A minimum R2-Score of 0.7 is used as a criterion to shortlist the industries that show consistency between the model vale and their historical data.