Cheeyau – 645698

<https://github.com/Cheeyau/MLOPS_assignmnets>

The file is a .ipynb type of file where in you can import it in VS Code or Jupyter Notebook. Than you get a promp for selecting the kernel (Python version) and choose the version from Anaconda3 if you have install it. After selecting it, you can run it in the editor.

For selecting the model, first started with linear regression model but after testing it out, it predicted explosive prediction. The second model I chose was random forest regression for the random element in the prediction process. This gives a less explosive prediction and is more in line with the previous data.

For processing, was it quite strange with processing datetime data. First was to clean up the date format from – to /. There were some issues with data type of datetime type in processing the numerical values for the model. So after converting it to unix timestamp where the datetime is converted to float with the total value in seconds.

For feature design, the first 2 combinations of data would be date and weekly\_sales as a starting point of the historical data. This could be a good indicator for testing the model’s performance. After that, the second and third indicators would be unemployment and PCI columns assuming unemployment would be the unemployment number of the US. The other columns as temperature and fual\_price are very localized information and would be a bad indicator while splitting up the data between train and test.

Figure: time to run pipeline

A picture containing screenshot, text, font, multimedia software

Description automatically generated