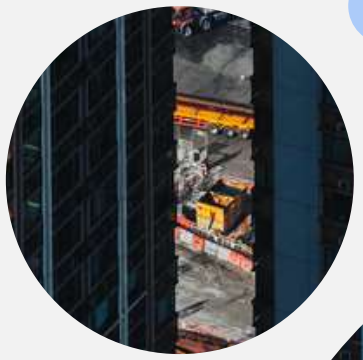
EXL

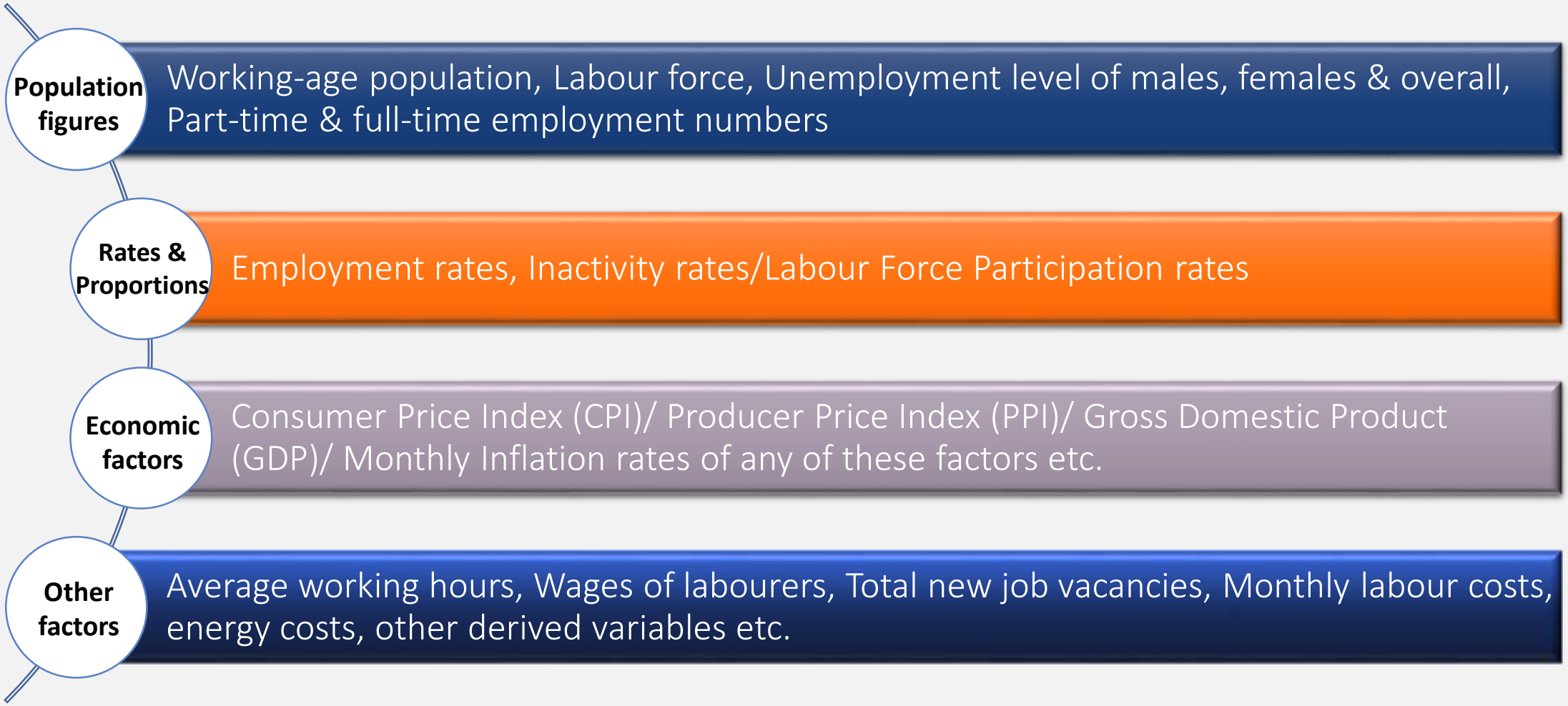
Analytics Hackathon 2023

Strategies



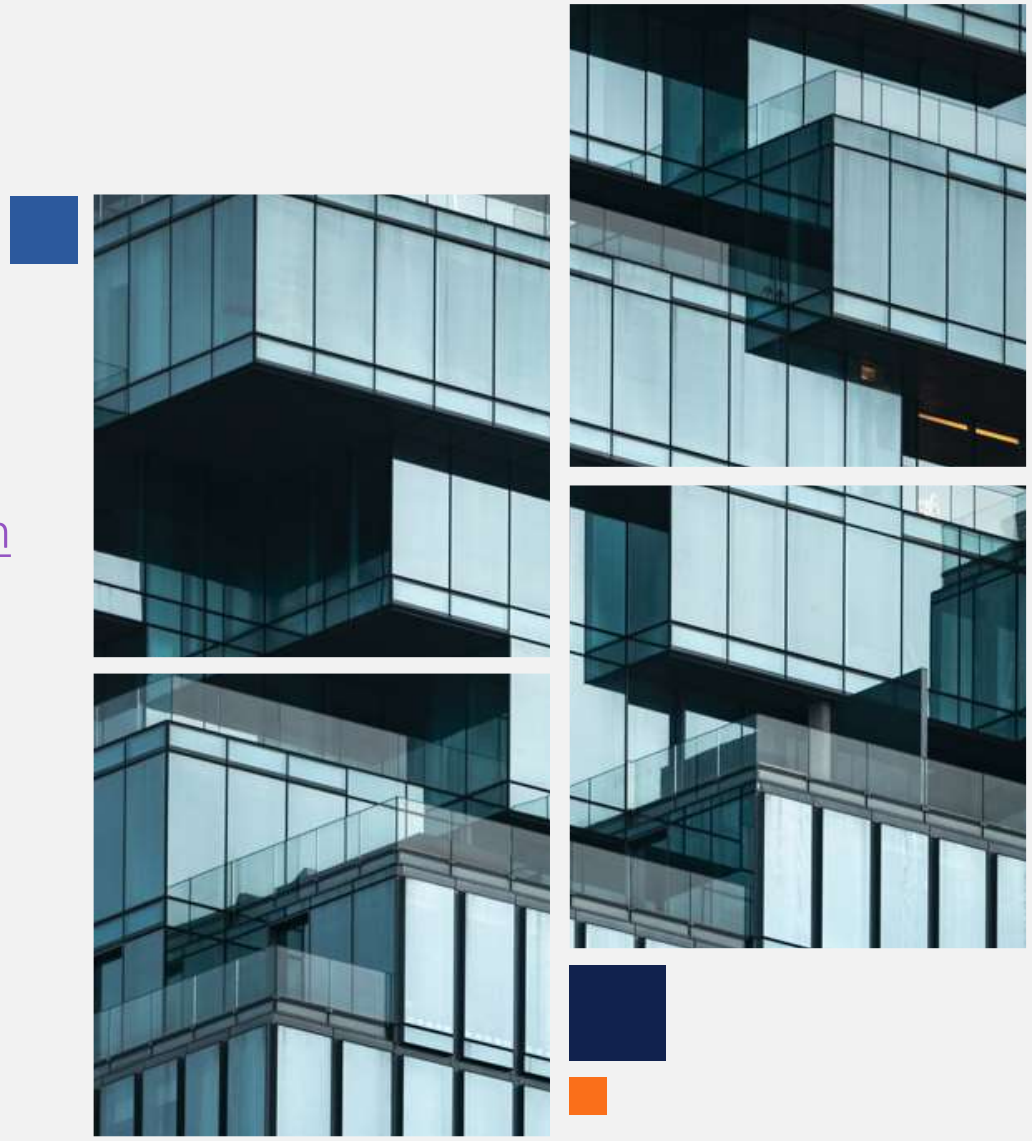
- Understanding the given Problem Statement.
- Researching on the factors that can affect the unemployment rate for each of the G7 countries over the years.
- Gathering necessary and relevant historical data on the factors as much as possible at monthly level from reliable sources.
- Applying a suitable forecasting technique, a statistical model which would be apt to predict the unemployment rates of the future months.

External factors considered



Data Sources

1. Unemployment - Unemployment rate - OECD Data
2. https://www.e-stat.go.jp/en/stat-search/files?stat_infid=000040046676
3. https://fred.stlouisfed.org/tags/series?t=france%3Bunemployment
4. www.ons.gov.uk
5. www.istat.it
6. www.statista.com



One example of the type of datasets used

Dates	Unemployment_Rate	Labour_Force	Employment_Rate	CPI_Urban_Customers	Working_Age_Population	Inactivity_Rates	Unemployed_male	Unemployed_female	Unemployed_all
1977-01	0.075	97208000	0.6456636	58.7	135147080.5	0.301604	3920000	3360000	7280000
1977-02	0.076	97785000	0.6468344	59.3	135350373.7	0.299043	4034000	3409000	7443000
1977-03	0.074	98115000	0.6491476	59.6	135571699	0.297712	3847000	3460000	7307000
1977-04	0.072	98330000	0.6515833	60	135839106.5	0.29763	3689000	3370000	7059000
1977-05	0.07	98665000	0.6526598	60.2	136142088	0.297739	3729000	3182000	6911000
1977-06	0.072	99093000	0.6547839	60.5	136339645.9	0.294509	3715000	3419000	7134000
1977-07	0.069	98913000	0.6537743	60.8	136557999.6	0.297488	3567000	3262000	6829000
1977-08	0.07	99366000	0.6557075	61.1	136723902.4	0.29442	3615000	3310000	6925000
1977-09	0.068	99453000	0.6570129	61.3	136838013.2	0.294228	3409000	3342000	6751000
1977-10	0.068	99815000	0.6579992	61.6	137169943.9	0.293411	3521000	3242000	6763000
1977-11	0.068	100576000	0.6611707	62	137346166.2	0.290504	3436000	3379000	6815000
1977-12	0.064	100491000	0.6637247	62.3	137496484.3	0.290795	3280000	3106000	6386000
1978-01	0.064	100873000	0.6642402	62.7	137670870.3	0.289785	3337000	3152000	6489000
1978-02	0.063	100837000	0.6643504	63	137857908.8	0.290719	3335000	2983000	6318000
1978-03	0.063	101092000	0.6656565	63.4	138031691.9	0.289188	3319000	3018000	6337000
1978-04	0.061	101574000	0.6693119	63.9	138106237.4	0.287265	3161000	3019000	6180000
1978-05	0.06	101896000	0.6700713	64.5	138361284	0.286817	3067000	3060000	6127000
1978-06	0.059	102371000	0.6729117	65	138703045	0.284519	2947000	3081000	6028000
1978-07	0.062	102399000	0.6704751	65.5	138827112.8	0.284983	3044000	3265000	6309000
1978-08	0.059	102511000	0.6729257	65.9	138953608.2	0.283969	3008000	3072000	6080000
1978-09	0.06	102795000	0.6741737	66.5	139101799.1	0.282417	3037000	3088000	6125000
1978-10	0.058	103080000	0.6762979	67.1	139331520.6	0.281368	3063000	2884000	5947000
1978-11	0.059	103562000	0.6768336	67.5	139567780.6	0.280409	3034000	3043000	6077000
1978-12	0.06	103809000	0.6769518	67.9	139786503.5	0.2791	3184000	3044000	6228000
1979-01	0.059	104057000	0.6778947	68.5	140083753.6	0.279516	3116000	2993000	6109000
1979-02	0.059	104502000	0.6789218	69.2	140276084	0.278184	3134000	3039000	6173000

Methodologies & Workflow

1 Importing & preparing the datasets

Historical data at **monthly level** for all the factors have been imported into the Analytical tool, R Studio from the reliable websites.

Data preparation includes adding **new derived variable** from the existing factors & deciding upon the final sets of external variables based on **Multicollinearity.**

2 Splitting the datasets –Train & Test parts

The latest 6 months data has been taken into **"Test"** data and the rest has been taken into **"Train"** data which is used to build and finalize the model.

The **"Test"** data has been used for validating the final model by comparing the actual values with the predicted ones.

3 Deciding upon the model parameters and finalizing it using the "Train" set

Here the parameters **p**, **d** and **q** has been finalized where **p**, **d**, **q** are the orders of **Auto-Regressive**, **differencing** and **Moving Average processes** respectively.

Only the significant coefficients are kept in the final model which are decided from the table of test of significance with the corresponding p-values.

4 Obtaining modelled values & future forecasts

The model, trained on the Train set has been validated on the "Test" set and wherever external regressors are present in final model, the values of those are forecasted for the future months and then used as input. Future forecasts are then obtained as required.

The upper & lower limits are the **95% Confidence intervals**.

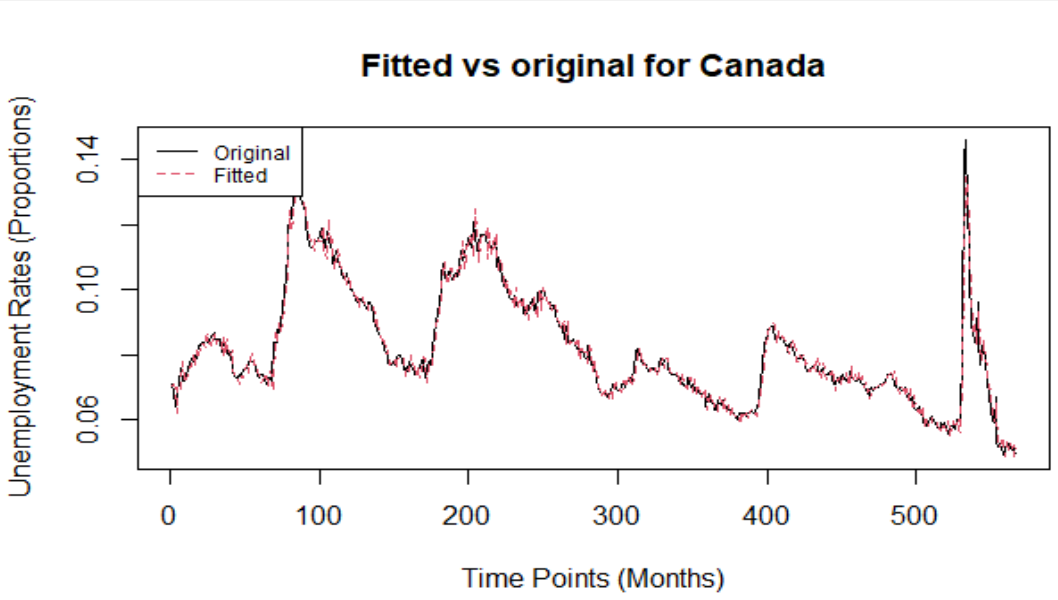
Final models used for prediction

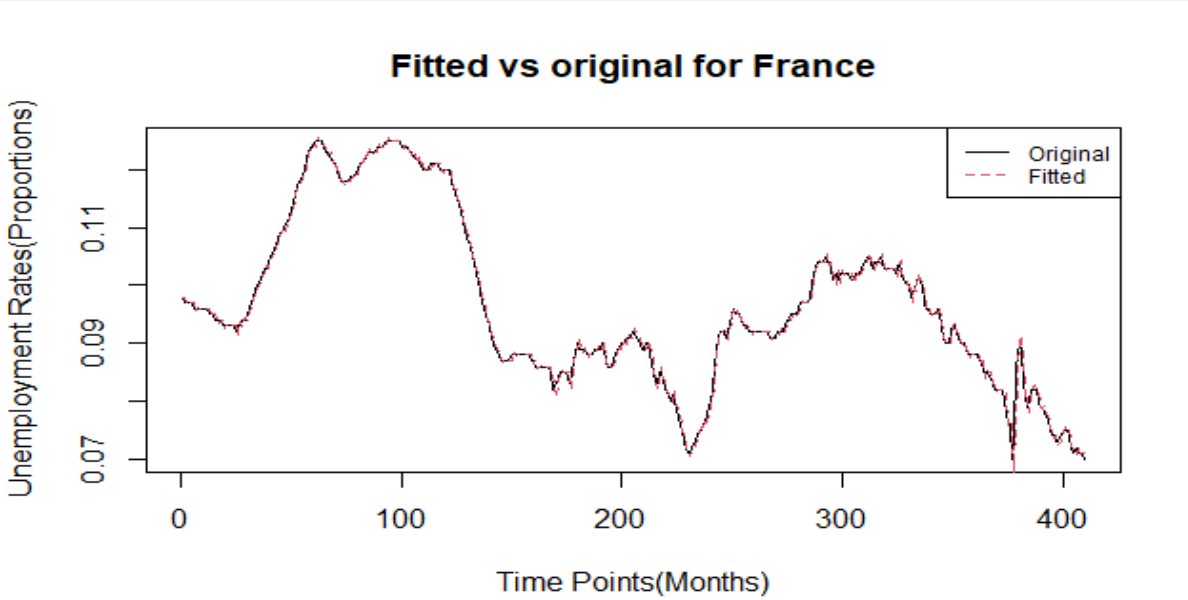
Canada	• **Time Frame** –> 1976 Jan – 2023 Mar • **Final model** –> Regression with ARIMA(1,1,2) or ARIMAX(1,1,2)
France	• **Time Frame** –> 1989 Jan – 2023 Feb • **Final model** –> ARIMA(1,1,1)
Germany	• **Time Frame** –> 1991 Jan – 2023 Mar • **Final model** –> ARIMA(1,2,2)
Italy	• **Time Frame** –> 2004 Jan – 2023 Mar • **Final model** –> Regression with ARIMA(0,2,1) or ARIMAX(0,2,1)
Japan	• **Time Frame** –> 1976 Jan – 2023 Mar • **Final model** –> ARIMA(2,1,1)
UK	• **Time Frame** –> 1997 Jan – 2023 Feb • **Final model** –> Regression with ARIMA(1,1,1) or ARIMAX(1,1,1)
USA	• **Time Frame** –> 1977 Jan – 2023 Mar • **Final model** –> Regression with ARIMA(1,1,2) or ARIMAX(1,1,2)

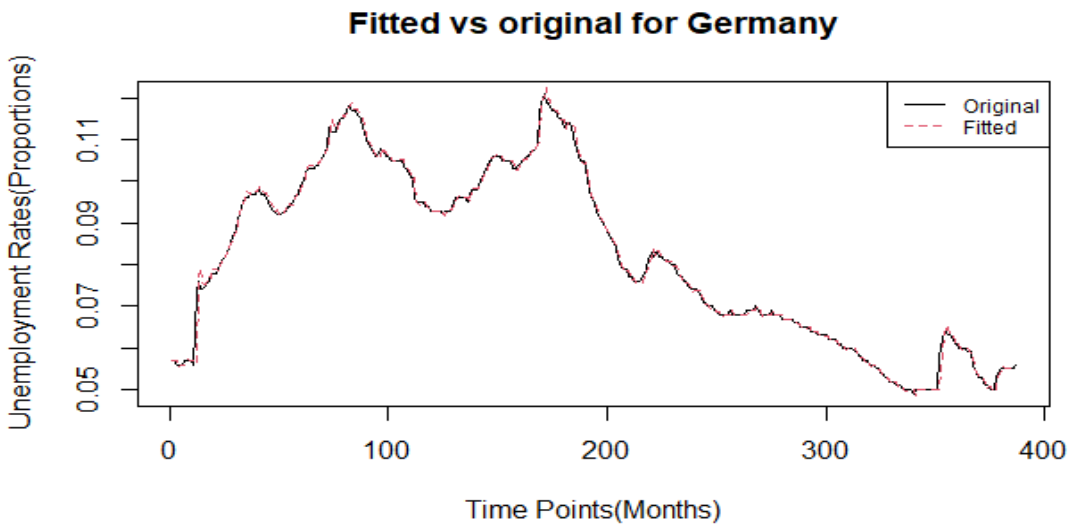
Results from the fitted models

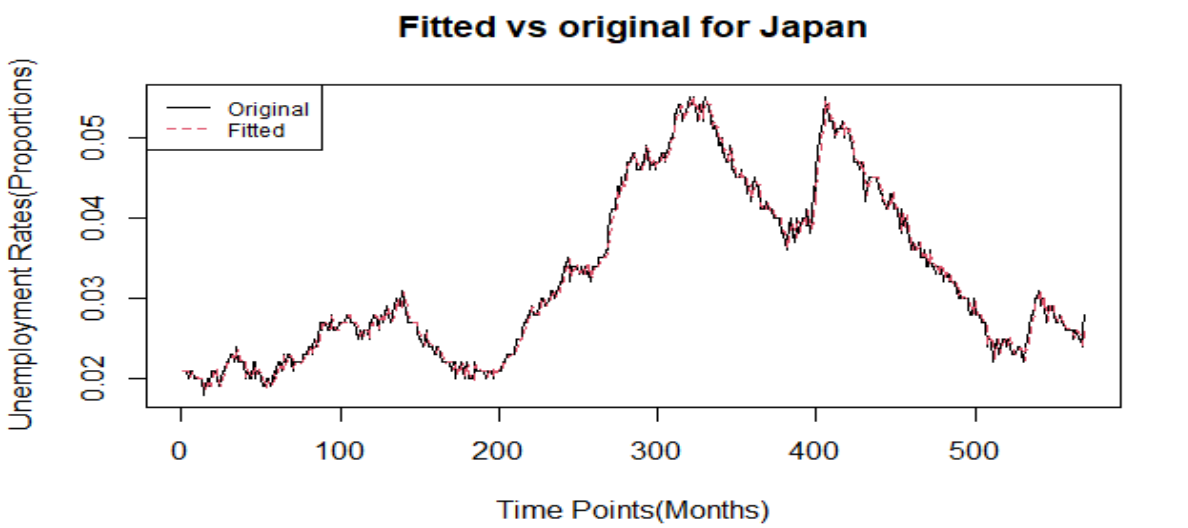
EXL

Examples of Actual vs Fitted plots for some of the countries over the years









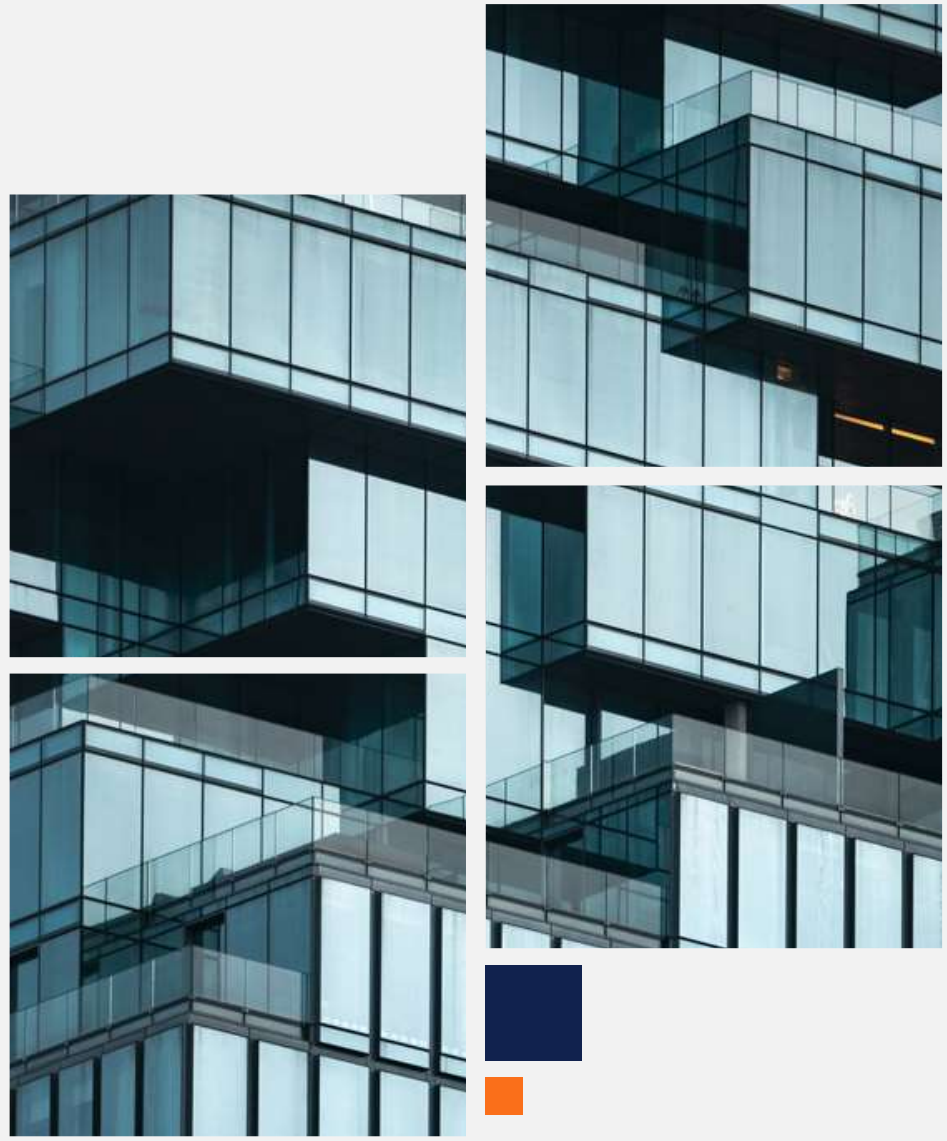
Comparison b/w Actual & Forecasts of the months that were asked for

Countries	Months	Actuals	Forecasts	Upper Bound	Lower Bound
Canada	May 2023	5.20%	5.01%	5.62%	4.41%
France	April 2023	7.00%	6.86%	7.09%	6.64%
Germany	May 2023	5.60%	5.63%	5.90%	5.36%
Italy	April 2023	7.80%	7.72%	8.12%	7.31%
Japan	April 2023	2.60%	2.74%	2.94%	2.54%
UK	March 2023	3.90%	3.82%	3.99%	3.64%
USA	May 2023	3.70%	3.52%	3.62%	3.42%

- Root Mean Square Error (RMSE) – 0.00131
- Mean Absolute Percentage Error (MAPE) – 2.775%

Some key findings

- Inactivity Rates or in other words Labour Force participation rate plays significant role in prediction of unemployment rates in countries like Canada, Italy, USA. As participation of Labour Force decrease, unemployment increases.
- For instance, on an average, decrease in participation rate of 100 people leads to unemployment of 65 people in Canada.
- CPI plays significant role in unemployment prediction in Italy.
- Ratio of male to female unemployment plays more or less significant role in unemployment in UK & USA where we say that the increase in female unemployment may be a cause of concern towards total unemployment.
- Average actual weekly hours of work plays moderately significant role towards unemployment in UK.



What's next?

Things that were beyond the scope like incorporating more factors which can play a vital role in prediction of unemployment, that couldn't be added here because of limitations, may be due to unavailability of proper reliable data

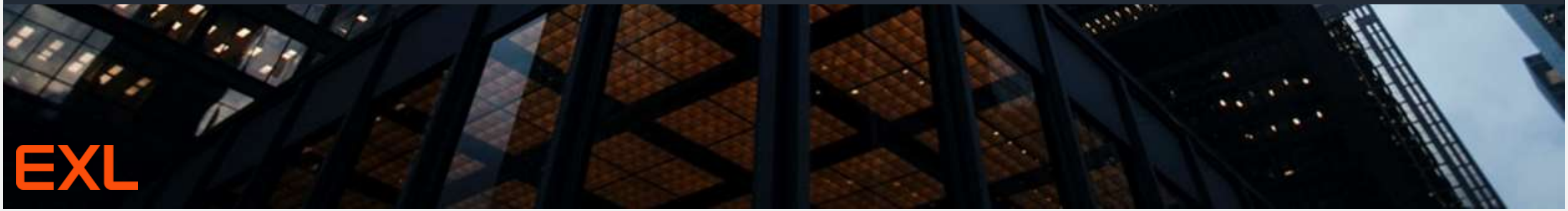
THANK YOU

Hack-HPT - Bibliography

1. Name – Tathagata Bardhan
   Emp id - 165123
   Team – NRG

2. Name – Priyanka Sarkar
   Emp id - 164423
   Team – NRG

3. Name – Harshit Gola
   Emp id - 175058
   Team – NRG



EXL