**Assessing Google Trends Data**

1. **INTRODUCTION**
2. **AIMS AND OBJECTIVES**

The aim of the project is to develop a methodology to predict macroeconomic indicators such as GDP and monthly retail sales with real-time data source, **Google Trends**. Keywords from the google would be identified which will serve as the predictors for nowcasting the GDP and monthly retails sales.

**Key Objectives**

The key objectives are as mentioned below:

* Nowcasting the macroeconomic indicators like GDP and monthly retail sales using real time data.
* Prediction of GDP will be at national level and will be industry wise.
* GDP nowcasting will be quarterly.
* Nowcasting for retail sales and digital sales will be monthly at national level and as per the industries.

Keywords will be searched from Google Trends

website: <https://trends.google.com/trends/?geo=CA>, and will serve as the predictors for nowcasting the macroeconomic indicators.

1. **DATA SET**

Data set for this project are open ended and the short description about them is as mentioned below:

1. **Gross Domestic Product (GDP) at basic prices monthly:** This dataset is a comma separated file containing the information about the monthly GDP. This file contains data from 1997 and do have some missing values, thus will require data wrangling.
2. **Gross Domestic Product (GDP) at basic prices quarterly:** This dataset is comma separated file containing the information about the GDP quarterly. This file contains data from 1997 and do have some missing values, thus will require data wrangling.
3. **Retail trade sales by province and territory:** This dataset contains information about the retail sales as per the province and territory. This data file is also comma separated and will require data wrangling.
4. **Retail trade sales by industry:** This is a comma separated data set containing the information about the retail sales trades as per the industry. Data wrangling is required in this dataset as well.
5. **Retail sales, price, and volume:** This is a comma separated data set containing monthly retail sales, price, and volume data. This data set will need some data wrangling.
6. **Retail E-commerce sales:** This is a comma separated dataset containing the information about the retail e-commerce sales. This is the data for digital sales and will require some data wrangling as well.

Our focus will be on the data starting from 2004 as we have Google trends available from that period and this will provide us huge data for our nowcasting.

In addition to this, we will be accessing Google Trends website to get real time data for the macroeconomic indicators predictions.

1. **METHODOLOGY**

The detailed description on the methodology of the whole project is as mentioned below.

1. **DATA WRANGLING**

The dataset that we have mentioned above for the GDP contains the information about Canada and is a complete data starting from the year 1997. Also, it has some missing values. We will be cleaning the data in such a way that we have information starting from 2004 and will replace the missing values with the desired values.

Also, the dataset for retail sales and digital sales will be filtered so that it contains information starting from the year 2004.

1. **MODEL FITTING**

* The data set is a time series data and will require the nowcasting using the real time data from google trends. We will be searching for keywords from the Google Trends and will be using them as the predictors in predicting the macroeconomic indicators.
* Some of the keywords are Economic crisis, loans, GPS, unemployment, affordable housing, economy news, agriculture, and forestry.
* The information is divided in **category and subcategory**, and we will be accessing each identified keyword in nowcasting our indicators.
* Since the data is time series, we will be fitting different models for an instance **Dynamic Factor Model (DFM)** which will serve as a dimension reduction method and can be helpful in analysing the time series data. Also, we are aiming at using **Auto-regressive Integrated moving Average (ARIMA)** models to understand the data and predict the indicators. This will serve as a good model for deeper understanding and accurate predictions. Apart from this, we will be implementing **LASSO** in coordination with **ARIMA family models**.
* We will be implementing the above-mentioned models and will be performing the comparative analysis. This will provide us the accurate predictions and will let us choose the best model for our nowcasting.

**IF TIME PERMITS**

* If time permits, we will also implement neural network and try to predict the macroeconomic indicators which could provide high efficiency.
* Forecasting of GDP as per different province and at all industry levels.
* Predicting GDP on monthly basis.

1. **VISUALISATIONS**

* Once the model fitting is achieved, predictions of GDP (quarterly basis), monthly retail revenue and monthly digital sales will be presented in a tabular format.
* This will be followed by bar graphs and line charts for the

predictions.

* A dashboard will be created which will give the information about the nowcasting of the indicators. (If time permits)

1. **FINAL PRESENTATIONS**

* A final presentation deck will be created which will highlight the research results and predictions.
* This will be structured and concise
* Also, a report will be created that documents the methodology and findings of the research results.

We will be aiming at using Python for all our coding, model fitting and analysis. Also, Git Hub will be the channel where we will be committing our code.

1. **DELIVERABLES AND SCHEDULE/ TIMELINE**