**Technical Report on Premier Project by Group NLP in Hamoye Internship HDSC Winter ‘22**

# **Introduction**

Foreign exchange rate, or FOREX rate, is the value of one currency in terms of another. The foreign exchange market determines FOREX rates which often fluctuates based on transactions between traders. For instance, the foreign exchange rate of Naira to Dollar is 415.76. This implies that it takes 415.76 Naira to buy a Dollar, as at January, 2022.

As we must have heard, data is the new oil. Data science technology works on datasets using different methods and processes to extract useful insights. Data science has transformed the way professionals of different fields make observations and decisions. What better way to take action if not based on trends and measurable insights?

For this project, the NLP team of the HDSC Winter ‘22 used a raw dataset containing foreign exchange rates information of about 21 countries. The dataset provided the FOREX rates of 20 years spanning from 2000 to 2009. The team used the information from this data to understand the relationships between FOREX rates trends and predict changes across selected countries.

# **Methods of Analysis**

The data preparation pipeline for the FOREX rates datasets consists of the following steps:

1. **Extract and load the data**

The data for our project group is titled: PP22/T616 on [Hamoye’s medium page](https://hamoyehq.medium.com/hdsc22-real-life-machine-learning-projects-f9ddda384332) with a scope of Trade, Agriculture and Finance. It falls under the topic “Foreign Exchange Rates”. The dataset for this project was accessed from the [kaggle directory](https://www.kaggle.com/brunotly/foreign-exchange-rates-per-dollar-20002019). Jupyter Notebook was used as the primary tool for carrying out data cleaning and analysis. Python was utilized as the choice of querying language. All relevant libraries from the Python machine learning libraries were successfully imported. Then, we loaded our dataset which was already downloaded to a local directory in Comma Separated Value (.csv) file format.

1. **Data transformation**

Data cleaning is the process of ensuring data is correct, consistent and usable. You can clean data by identifying errors or corruptions, correcting or deleting them, or manually processing data as needed to prevent the same errors from occurring.

In cleaning our data, we queried the dataset and observed that it contained 5217 rows (data ranging from index 0 to 5216) and 24 columns containing the various countries. We also observed that the Time series column located at index 1 was in an object data type. Hence, we converted it into a datetime format to make it usable.

Likewise, other columns ranging from index 2 to 23 which were originally in object format were converted into float data type. The data unit (index 0) which had “unnamed'' as a column name was dropped from the dataset using the drop method. This was done because it was a replica of our index column. This process reduced our dataset from its original 24 columns to 23 columns.

Using the isnull() method, we accounted for the total number of missing values. These missing values ranged from 197 to 201 in different country columns. We decided to drop those rows because of the low percentage. Dropping these missing values leaves the dataset with only valid and significant values. The dataset now had 5015 rows and 23 columns. At this stage, the dataset was ready for exploration and further analyses.

**Exploratory Data Analysis**

Exploratory Data Analysis (EDA) techniques is widely used in the data discovery process today. It helps data scientists to perform initial investigations on a provided data so as to discover patterns, spot anomalies, and check assumptions using summary statistics and graphical representations.

There are a number of tools used for performing an EDA. However, Python was the preferred choice for this project. Python is an object-oriented programming language with dynamic semantics. Its high-level, built-in data structures, combined with dynamic typing and libraries, makes it attractive for rapid application development and model training and building.

**Sequential Process for Performing Data Visualization on the Provided Dataset**

• Data conversion from different currencies to US Dollars (USD) to evaluate the trend between currencies.

• Check the average value of each currency in USD for the four quarters within the years provided. This gives further insight into the annual changes occurring for each currency on a quarterly basis.

• Check for noticeable trends within the provided dataset. Some resulting trends indicated that the Euro has been appreciating until 2015 when it depreciated from 0.75 to 0.94. However, it appreciated again in 2019 to a value of 0.89. The United Kingdom’s Pounds appreciated from 2000 to 2009, although it has been depreciating in respect to USD till 2019.

• Generate a plot to determine the country with the highest currency/USD. From the generated plot, it can be deduced that the United Kingdom’s Pounds are the highest currency value relative to US Dollars.

• Also, generate a plot to determine the country with the lowest currency/USD. From the generated plot, it can be observed that Korean Won is the lowest currency value relative to US Dollars.

• Check the relationship between the lowest and highest trending currencies (i.e. United Kingdom’s Pounds and Korean Won).

• Identify trending currencies according to the last date of 2019 provided in the dataset. Doing this shows the strongest to the weakest currencies as provided in the dataset.

• Generate a correlation heat map plot for the currencies. A heat map is a data visualization technique that shows the magnitude of a phenomenon as color in two dimensions. We observed that the Euro and Danish Krone currencies have a perfect positive correlation. This is clearly explained in the attached link <https://en.wikipedia.org/wiki/Denmark_and_the_euro>.

• Use a scatter plot to check the correlation between Euro and Denmark currencies. We observed a pure simple regression line.

• Generate a plot to show a set of currencies having a high correlation. This is done to show the relationship between the provided currencies which have been presented in dollar formats.

• Obtain the maximum, minimum, and last day price of each currency. Also, indicate the years at which these events occur. This gives insight into significant fiscal highlights which have occurred within the provided data timeframe.

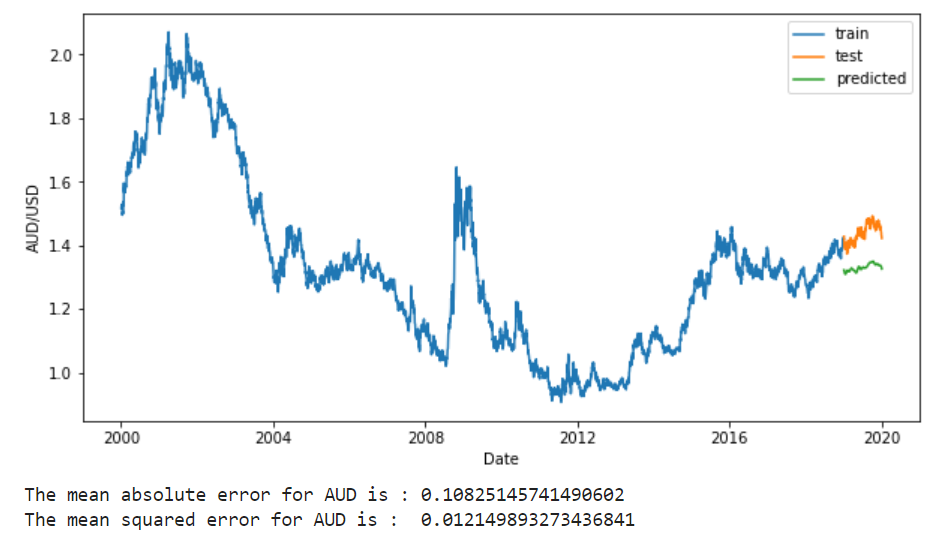
**Prediction Models**

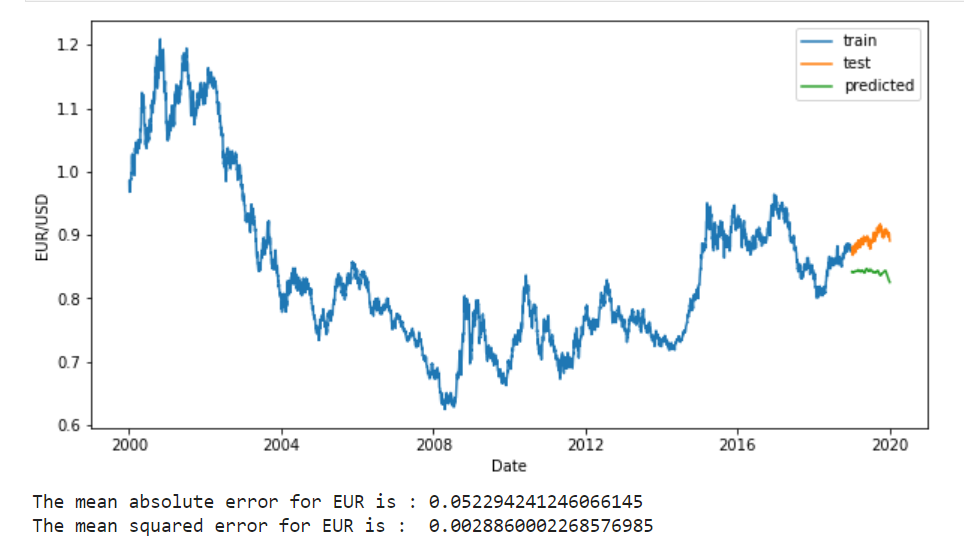
We chose to work with 4 selected currencies based on the most traded currencies data from investopedia. Chosen countries are United Kingdom (Pounds), Euro Area (Euro), Australia (AUD), and New zealand (NZD)

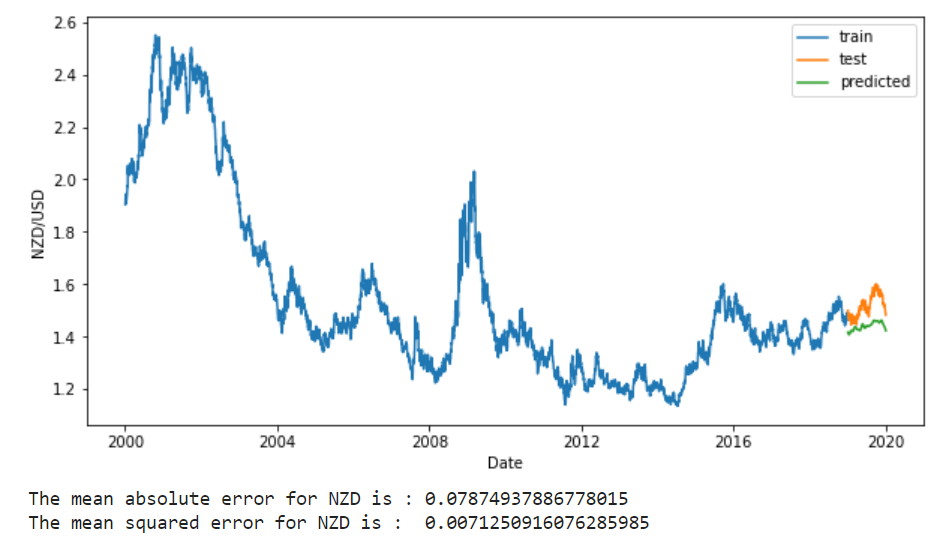
Techniques used for the forecasting and model building are the ADFULLER test, autocorrelation plot, and the Prophet module. We chose these libraries because we observed that our datasets were not stationary, and they would be the best option for predictive analysis.

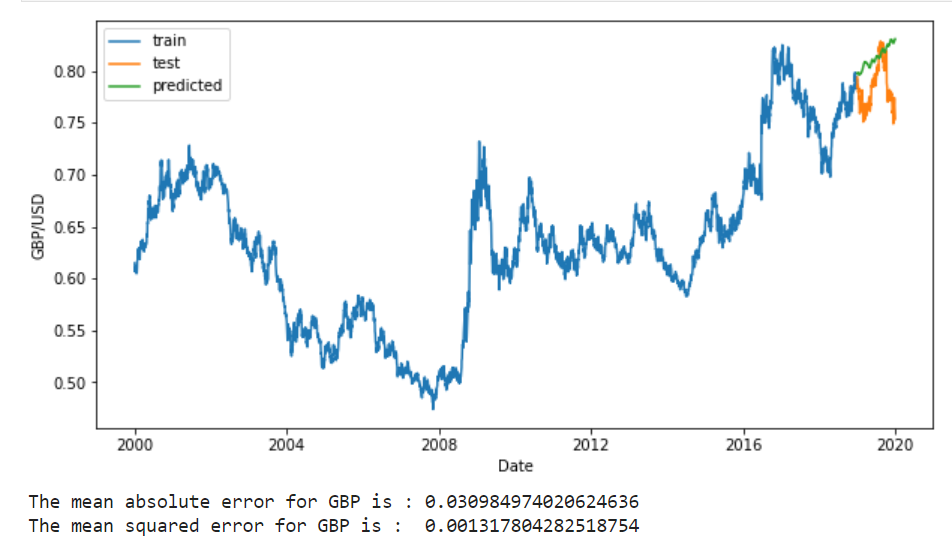
# **Results and observations**

1. Australian Dollar FOREX prediction

****

1. **Euro Area FOREX prediction**

**3. New zealand Dollar prediction**

**4. United Kindom Pounds FOREX prediction**

# **Conclusion**

The foreign exchange market has a direct impact on wages, cross-border investments, and the economy as a whole. Financial institutions, companies, governments, and other entities use this market to adjust their currency holdings. This indicates the need to obtain accurate evaluations and predictions of market trends. This study proposes a means of utilizing data science technology to evaluate and predict such trends within a country's economy relative to its US dollar equivalent.

# **NLP group members**

Nnona Ebuka John

Aman Vishwakarma

Fatima Alhassan

Sanskriti Jain

Manoj Kumar

Ahmed Anifowoshe

Odubanjo Sulaimon Abiodun

Abiodun Shittu

Gbolahan Alao

Bright Odike

Jeremiah Zhao

Emeke Ihedilionye

Timothy Oladapo

Shereen Onyango

Saheed Olurebi

Oluwaseyi Ogunlana

Olugbuyi Peter

Ola Ade

Nicholas Aniefiok Udoh (APL)

Idara Effiong

Halimat Atanda (QA)