Exploratory Data Analysis Report

Analysing the Impact of International Events on Special Drawing Rights (SDR): An Exploratory Data Analysis

1. Introduction

The purpose of this report is to present the findings of an exploratory data analysis (EDA) conducted on a dataset related to SDR (Special Drawing Rights).

The analysis involved deriving and transforming the dataset as needed and performing various manipulations using Jupyter Notebook. This report includes the collage pic of all team members, name and registration number details, dataset link, Jupyter Notebook manipulations, attributes available in the dataset, analytical results, and a link to the Google Drive containing the code.

2. Team Information

Collage Pic of Team Members:

Team Members:

1. **Name:** Kartik Ranjan, **Registration Number:** 23MDT0065
2. **Name:** Aditya Unnikrishnan, **Registration Number:** 23MDT0028
3. **Name:** Mathew Kevin John, **Registration Number:** 23MDT0046
4. **Name:** Udaiyar Vishnu Ganesan, **Registration Number:** 23MDT0027

3. Dataset Information

Dataset Link: https://www.imf.org/external/np/fin/data/rms_sdrv.aspx

Description: The currency value of the SDR is determined by summing the values in U.S. dollars, based on market exchange rates, of a basket of major currencies (the U.S. dollar, Euro, Japanese yen, pound sterling and the Chinese renminbi). The SDR currency value is calculated daily except on IMF holidays, or whenever the IMF is closed for business, or on an ad-hoc basis to facilitate unscheduled IMF operations. The SDR valuation basket is reviewed and adjusted every five years.

4. Data Manipulation in Jupyter Notebook

The dataset underwent several manipulations to derive relevant data for EDA. The following techniques were used:

- Data Cleaning
- Find outliers
- Dimensionality reduction
- Feature Selection

a. Data Cleaning

The methodology combines two approaches for handling country names in text data. Initially, natural language processing (NLP) techniques, specifically using the spaCy library, are employed to identify geopolitical entities (GPEs) within the text. Once a GPE is identified, the pycountry package is utilized to convert the extracted country name to its corresponding geographical form. This conversion ensures normalization of country names to a consistent format, enhancing flexibility and accuracy in handling country names across different texts. Additionally, a Python dictionary is created, pairing geopolitical names with their corresponding geographical names for various countries. This comprehensive dictionary is refined and expanded as needed to improve accuracy, providing a structured approach for mapping geopolitical names to geographical counterparts. Through testing and validation using sample inputs, the accuracy of the mapping is ensured, further enhancing the reliability of the methodology.

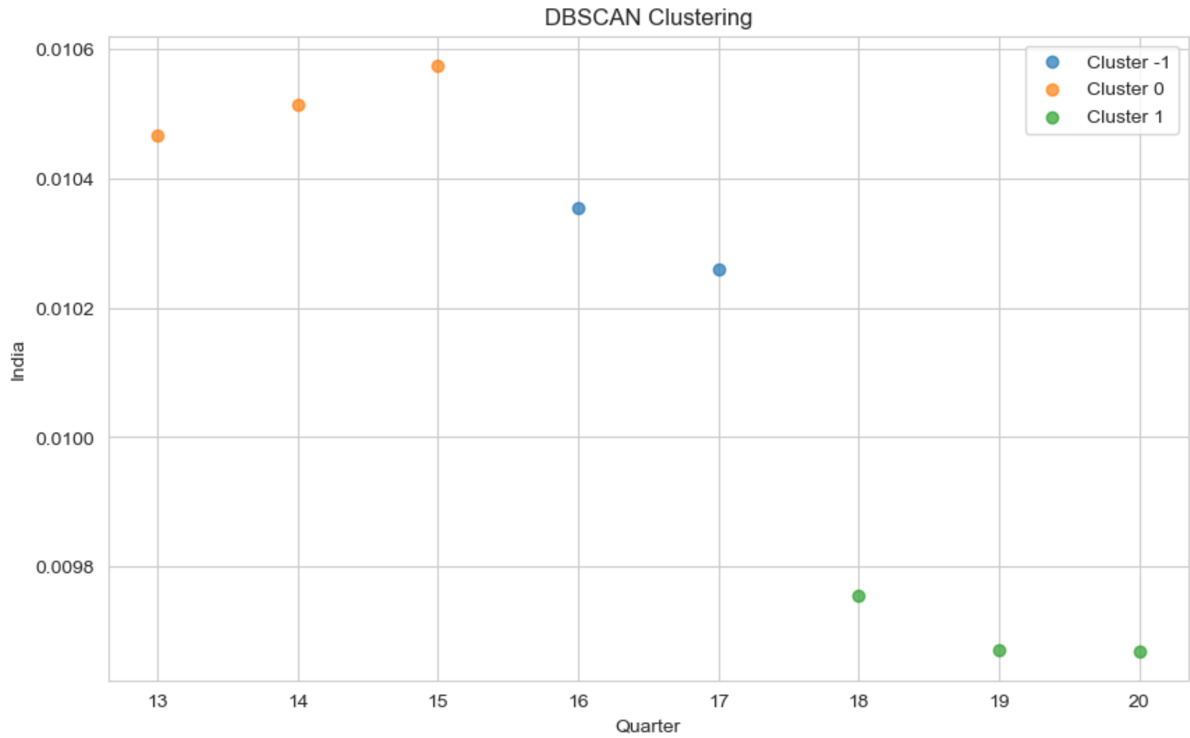
	Data	Algeria	Australia	Botswana	Brazil	Brunei	Canada	Chile	China	Colombia	...	Singapore	South Africa
0	2016-01-04	0.006710	0.520544	0.064140	0.184561	0.506341	0.515910	0.001015	0.110659	0.000229	...	0.506341	0.046369
1	2016-01-05	0.006723	0.522630	0.064223	0.179481	0.508786	0.518022	0.001011	0.111183	0.000226	...	0.508786	NaN
2	2016-01-06	0.006723	0.516299	0.063713	0.180693	0.506380	0.514650	0.001014	0.110623	0.000226	...	0.506380	0.045834

	Year	Date	Event	Country
0	2010	January 12	A 7.0 magnitude earthquake in Haiti kills 230,...	Haiti
1	2010	February 18	2010 Nigerien coup d'état.	Niger
2	2010	March 29	2010 Moscow Metro bombings.	Moscow
3	2010	April 10	The President of Poland, Lech Kaczyński, is am...	Poland
4	2010	April 20	The largest oil spill in US history occurs in ...	US
...	...	...	...	...
244	2023	October 27	Operation 1027, a coordinated assault by anti-...	Myanmar
245	2023	November 24	Somalia admitted to East African Community.	Somalia

b. Finding Outliers

(DBSCAN Clustering)

Date	India	U.S.	EUR	China	Quarter
2019-03-31	0.010465	0.721768	0.825242	0.107320	13
2019-06-30	0.010513	0.726449	0.818882	0.107555	14
2019-09-30	0.010575	0.733852	0.817598	0.105811	15
2019-12-31	0.010354	0.734203	0.810060	0.104242	16
2020-03-31	0.010259	0.743009	0.818428	0.105615	17
2020-06-30	0.009755	0.736259	0.820227	0.104045	18
2020-09-30	0.009671	0.726178	0.842056	0.104597	19
2020-12-31	0.009669	0.709359	0.848977	0.107206	20



Our dataset captures the Special Drawing Rights (SDR) values of some nations, over a brief spanning of period from January 4,2016 to 23 December 2021. SDR is a supplementary international reserve asset established by the International Monetary Fund (IMF) to support member countries official reserve. Each row corresponds to a specific date within the mentioned timeframe, while the columns represent SDR values for the listed nations and region of those dates.
To perform, DBSCAN on this data, we first segment the time series data into quarters, spanning from 2016 to 2021 within each quarter, we identify

the maximum SDR value achieved focusing specifically on the period from 2019 to 2020(quarters 13 to 20), we extract these maximum values.

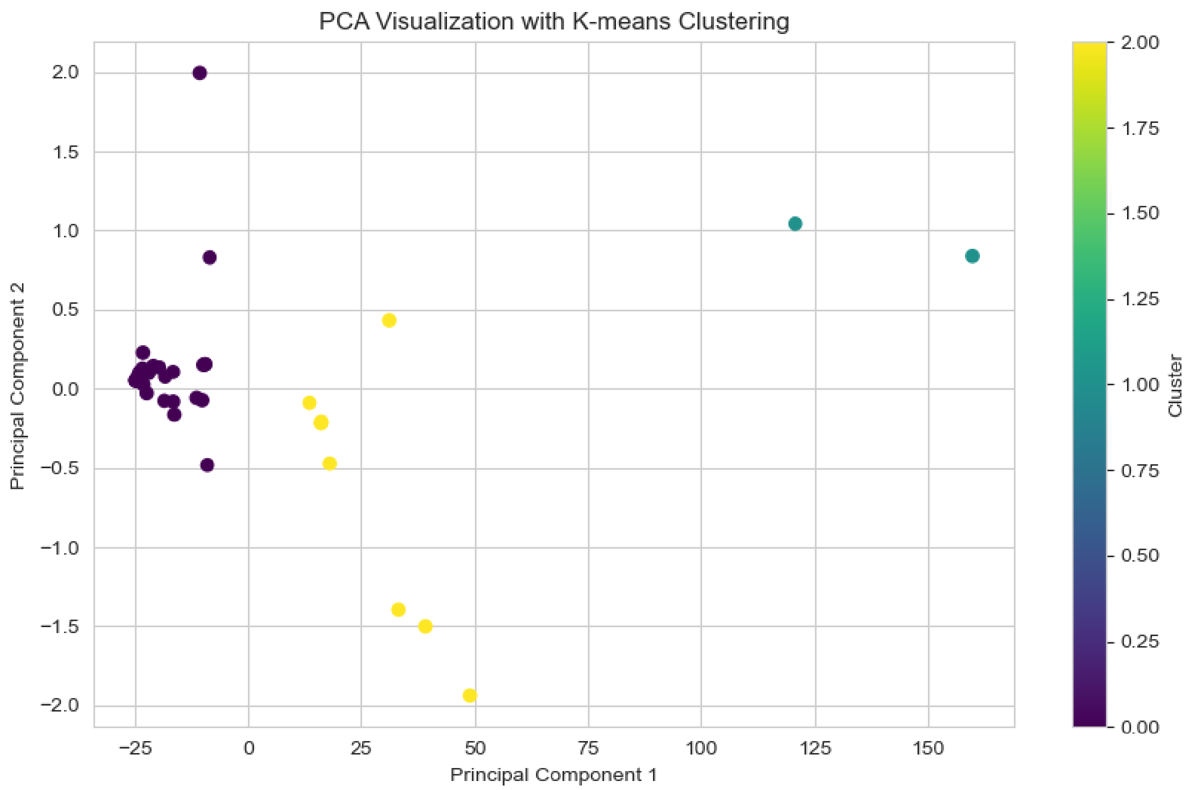
INTERPRETATIONS

The cluster 1 contains quarters 13, 14 and 15 which are from January 2019 to September 2019. The cluster 2 contains quarters 18, 19, 20 which are from April 2020 to December 2020.
We infer that the time period for January 2019 to September 2019 was a Covid-free time. Also, time period April 2020 to December 2020 was the time when Covid entered India. The outliers are the points when Covid had not entered India, but world faced the first case of Covid-19.

c. Dimensionality Reduction

PCA

To conduct Principal Component Analysis (PCA) on our dataset, we first organized the data, ensuring that each row represented a specific date while each column denoted a country's SDR value. We then standardized the dataset to remove any bias introduced by differing scales, ensuring that each country's contribution to the analysis was based on its relative variance rather than absolute values.



PCA Visualization with K-means Clustering
Principal Component 1
Principal Component 2
Cluster
−25
0
25
50
75
100
125
150
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00

```
Country: Algeria, Cluster Label: 0
Country: Australia, Cluster Label: 2
Country: Botswana, Cluster Label: 0
Country: Brazil, Cluster Label: 0
Country: Brunei, Cluster Label: 2
Country: Canada, Cluster Label: 2
Country: Chile, Cluster Label: 0
Country: China, Cluster Label: 0
Country: Colombia, Cluster Label: 0
Country: Czech Republic, Cluster Label: 0
Country: Denmark, Cluster Label: 0
Country: EUR, Cluster Label: 2
Country: India, Cluster Label: 0
Country: Israel, Cluster Label: 0
Country: Japan, Cluster Label: 0
Country: Korean, Cluster Label: 0
Country: Kuwait, Cluster Label: 1
Country: Malaysia, Cluster Label: 0
Country: Mauritius, Cluster Label: 0
Country: Mexico, Cluster Label: 0
Country: New Zealand, Cluster Label: 2
Country: Norway, Cluster Label: 0
Country: Oman, Cluster Label: 1
Country: Peru, Cluster Label: 0
Country: Philippines, Cluster Label: 0
Country: Poland, Cluster Label: 0
Country: Qatar, Cluster Label: 0
Country: Russia, Cluster Label: 0
Country: Saudi Arabia, Cluster Label: 0
Country: Singapore, Cluster Label: 2
Country: South Africa, Cluster Label: 0
Country: Sweden, Cluster Label: 0
Country: Switzerland, Cluster Label: 2
Country: Thailand, Cluster Label: 0
Country: Trinidadian, Cluster Label: 0
Country: U.A.E., Cluster Label: 0
Country: U.K., Cluster Label: 2
Country: U.S., Cluster Label: 2
Country: Uruguay, Cluster Label: 0
```

INTERPRETATIONS

By performing PCA, we gained several key insights. Firstly, PCA allowed us to reduce the dimensionality of our dataset while retaining the most critical information, thereby simplifying subsequent analyses and visualizations. Additionally, PCA facilitated the identification of underlying patterns and structures within the data, such as clusters of countries exhibiting similar

SDR trends or outliers that deviated significantly from the norm. Moreover, PCA aided in feature extraction by isolating principal components that captured the majority of the variance, enabling us to focus on the most influential factors driving SDR variations across countries and over time. Overall, PCA provided a comprehensive and efficient approach to understanding the complex dynamics of SDR values among different countries, enhancing our ability to derive actionable insights and make informed decisions based on the data.

d. Feature Selection

Forward feature selection using linear regression was performed on a dataset containing various countries as column variables. The process involved iteratively adding one country at a time to the model and evaluating its impact on model performance. Each iteration trained a linear regression model using the selected countries so far and assessed the model's performance through cross-validation. The feature (country) resulting in the greatest improvement in model performance was retained at each step. This process continued until no further significant improvement was observed, based on predefined criteria. The final selected features represent the subset of countries that collectively contribute the most to predicting the target variable, showcasing their significance in the regression model for the given dataset.

Feature selection (by intuition):

As we had the SDR value of each country, selecting or dropping of features will just lead to loss in data. We removed the countries who have variance less than “0.0000001”. This value is so low, means that there is not much variation in the values of SDR over time and for out project we won’t be needing the countries for which there is not much variation. Through this we removed 4 countries, i.e. from 39 countries we got 35 countries. Then we removed all the countries for which there is no event available. By doing this we ended up with a total of 19 countries in the end.

5. Attributes in the Dataset

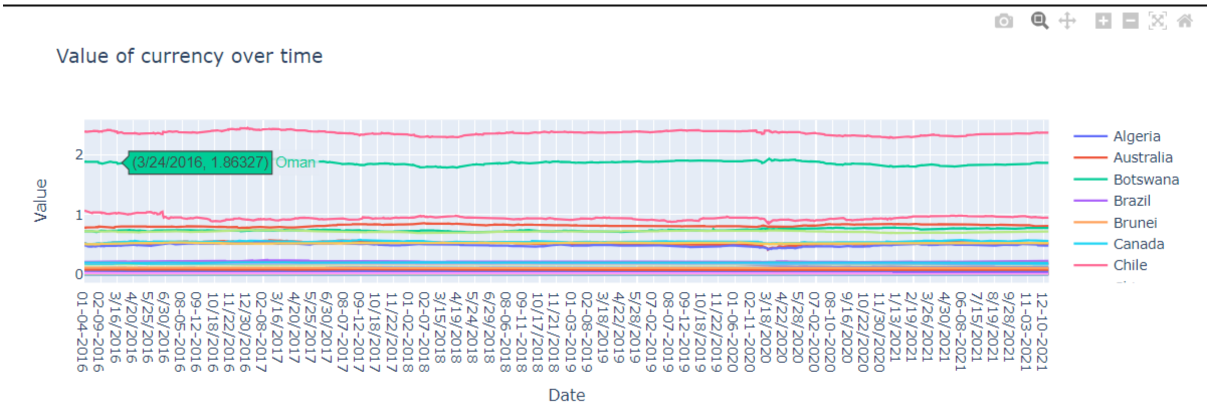
There are two datasets for our data:

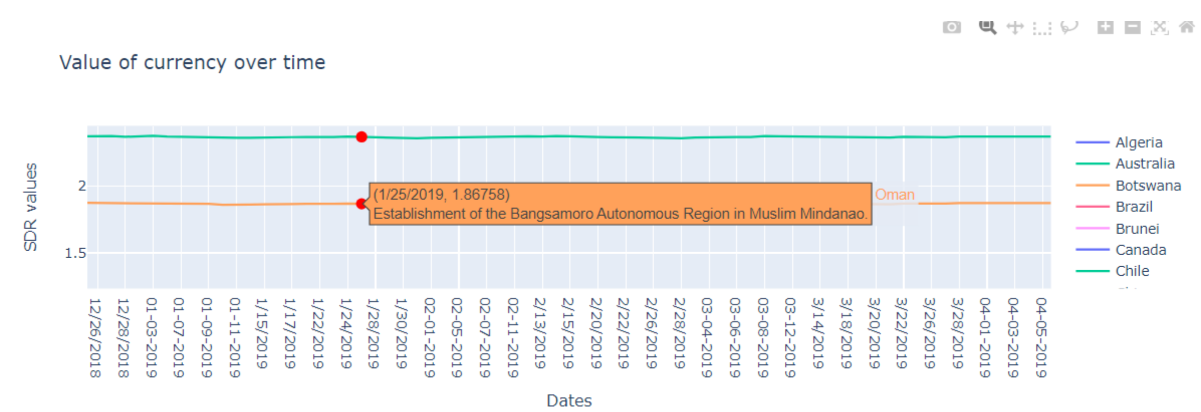
1. SDR_val
2. Event

- Quantitative: SDR value of each country.
- Qualitative: Date, Event, Country.

6. EDA Results

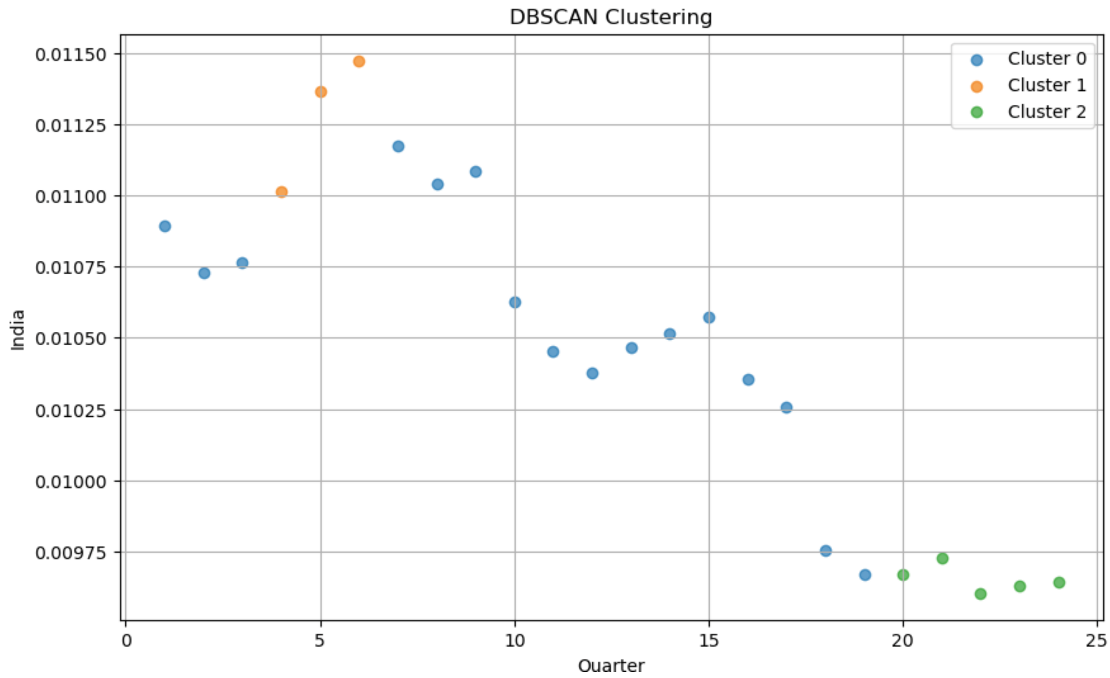
Here is a screenshot of our interactive line plot





INTERPRETATIONS

Here as you can see in the plot at any point of time, we can hover over the timeline and observe how the event has affected the Economic value (measured using the SDR scale) for the specific country. This can help us in understanding the regional consequences of Global events economically. Some of the notable events are Covid-19 and election of Donald Trump as the US president.



	Unnamed: 0	Year	Date	Event	Country
69	69	2017	January 20	Donald Trump is inaugurated as President of th...	the United States
70	70	2017	January 21	Millions of people participate in the Women's ...	NIL
71	71	2017	February 13	Kim Jong-nam, the half-brother of Kim Jong-Un,...	Kuala Lumpur
72	72	2017	May 13	Pope Francis canonizes Jacinta and Francisco M...	Jacinta
73	73	2017	May 22	A terrorist bombing attack at an Ariana Grande...	Manchester
74	74	2017	June 2	36 people are killed in an attack in Resorts W...	NIL

7. Jupyter File's Link

The interactive dashboard for data analysis can be accessed via Google Drive using the following link:

Link : https://drive.google.com/drive/folders/1A4YmekPyt1QWkUOGJfHJFXfSw0v6QB1m?usp=sharing

- **Access Credentials:** Email - padmapriya.r@vit.ac.in

The files for the dataset can be accessed via GitHub using the following link:

Link : https://github.com/Slowqueso/EDA-J-comp-SDR-vs-Events-timeline/tree/main

Conclusion

Exploratory Data Analysis (EDA) offers a powerful lens through which we can discern the intricate relationship between global events and their impact on the economic values of individual countries, quantified in Special Drawing Rights (SDR). By delving into this data, we uncover valuable insights that help us understand how the world stage influences economic metrics on a national level.

One of the primary methodologies employed in this analysis is the segmentation of time periods into 'before' and 'after' the occurrence of significant global events. This division enables us to isolate the effects of these events on both individual nations and the global economic landscape. Notably, we're able to identify instances where specific countries remain relatively unaffected while the broader world experiences significant shifts.

Crucially, this approach facilitates the categorization of events as either beneficial or detrimental to specific countries. By discerning whether a nation's economic values improved or deteriorated following a global event, we gain a nuanced understanding of its impact at the national level. This categorization also serves as a foundational framework for assessing the performance of governments over their tenures.