Visualising higher frequency economic indicators from unconventional sources using BI tool

Eilyn Chong Yee Lin[1]

Abstract

This paper illustrates the utility of Business Intelligence (BI) tools for extracting, processing, combining, and visualizing higher frequency indicators as an initial data exploratory tool to enable users to monitor the state of economic activity. As an empirical use case to showcase these advantages, I will demonstrate an end-to-end workflow that utilizes Microsoft Power BI to compile publicly available high-frequency indicators, including data from Google Trends, electricity generation from the grid system operator, prices of consumer goods, and sentiment extracted from news articles. These high-frequency indicators are then transformed and visualized within a single platform, alongside the official economic statistics they are designed to track or correlate with. This workflow leverages several of Power BI's features, including the ability to (i) read data directly from online CSV files; (ii) perform a POST request against a URL with dynamic parameters, (iii) run Python/R scripts within Power BI, and (iv) automatically refresh all datasets loaded into the Power BI file's model. With such a workflow, the processing of high-frequency indicators from disparate sources can be streamlined, and dynamic visualizations can be made accessible to a wide group of users, allowing them to explore the information content of these indicators in tracking economic activity.

Keywords: High-frequency Indicator, Business Intelligence, Visualisation

JEL classification: C55, C82, C88

[1] Bank Negara Malaysia. E-mail: eilyn@bnm.gov.my. The views expressed are solely the responsibility of the author and should not be interpreted as reflecting the views of or the endorsement of any specific software by the Central Bank of Malaysia, or of anyone else associated with the Central Bank of Malaysia.

1. Introduction

Policymakers are increasingly calling for faster and more granular insights into the state of the economy for decision-making, particularly in light of the rapidly evolving developments following the pandemic. This demand also stems from the recognition that headline numbers may obscure important distributional differences that impact policy decisions. As such, to complement official economic statistics, national statistical agencies and central banks have explored a range of high-frequency and granular indicators, often by-products of economic activity. Yet, until the informational value of these indicators is well-established, developing production-level dashboards for them can be costly.

In this paper, I illustrate the utility of Business Intelligence (BI) tools as a cost-effective, interim solution for extracting, processing, combining, and visualizing high-frequency indicators from unconventional sources to facilitate initial exploration of their informational content. Three empirical use cases will be presented, demonstrating an end-to-end workflow using Microsoft Power BI ("Power BI") with open-source data on Google search queries, daily prices of consumer goods, and sentiment data extracted from newspaper articles.

This workflow leverages several features of Power BI, including its ability to (i) directly access data from online files and databases, (ii) perform requests against a URL with dynamic parameters, and (iii) execute Python/R scripts within Power BI for more advanced queries and statistical techniques. The high-frequency indicators are processed and transformed in Power Query Editor, where the applied steps offer clear documentation and analysis reproducibility.

Once the data is loaded into the report, it can be visualized in a unified platform, alongside the official economic statistics that the indicators aim to track or correlate with. This platform can be accessible to users, enabling them to explore the informational content of the indicators in tracking economic activity and uncovering underlying nuances.

The true value of these high-frequency indicators lies in their consistent performance over time, providing valuable insights into economic development. After developing the BI report, regular data and model updates can be executed via the refresh function. Some BI tools offer enterprise-level subscriptions, broadening access to a wider range of users, from internal stakeholders to the public, for monitoring the state of the economy. Finally, I present practical considerations for visualizing data using BI tools, including data accessibility, the capabilities of BI tools for more advanced statistical techniques, and associated performance issues.

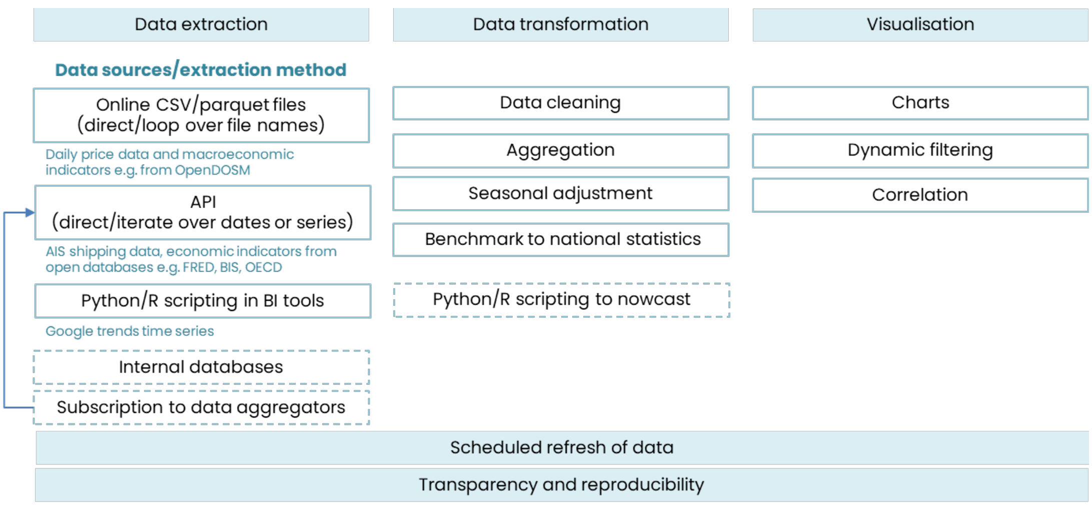
2. Data Visualisation Workflow in Power BI

BI tools are known for their capability to streamline data analysis by integrating information from diverse sources, from traditional databases to modern cloud services. Once the data are within the tool's environment, users can create, modify, and present it through visually interactive dashboards, making complex datasets more accessible. This democratizes access to insights, extending beyond individuals with specialized analytical skills. Microsoft's Power BI is one example within the BI spectrum that exemplifies these functionalities.

Figure 1 provides a broad overview of an end-to-end data visualisation workflow in Power BI, including the extraction of data from sources, data transformation, and visualisation of these data. For the use cases, data was gathered from several sources. These included online CSV/parquet files, interactions with APIs, and even scripts written in Python and R.

Managing substantial volumes of data distributed across numerous smaller datasets requires a structured approach to ensure that all data is captured. There is often a need to use looping techniques, especially when data sets span multiple dates or time series. This approach not only ensures thorough data collection but also simplifies maintenance down the road. For example, in Figure 2, one may apply the loop function to automatically cycle through months indicated in the file names up until the latest available month to ensure complete data collection. Figure 3 illustrates another example of a loop function for the APIs from the St. Louis FED, to go through different time series for data retrieval. For more intricate data extraction tasks, one may need to utilize Python/R scripting, which may be supported in certain BI tools. Figure 4 showcases an example of importing Google Trends time series data in Power BI using the PyTrends package on Python.[2]

Figure 1: Stylised data visualisation workflow in BI tool



2 More details of Pytrends are available in https://pypi.org/project/pytrends/. User can adjust keywords and timeframes used for each country to modify or expand the Google Trends dataset.

Figure 2: Loop functions over dates that vary across file names

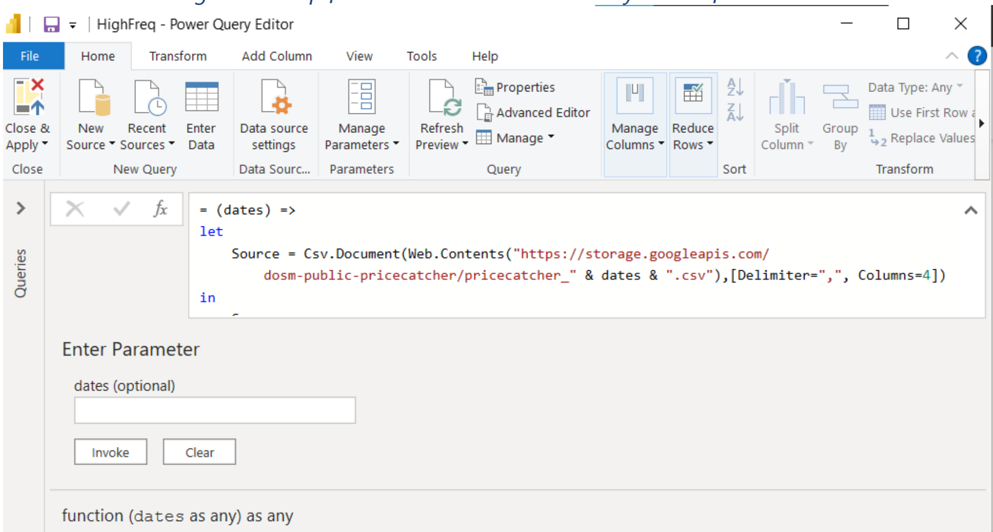


Figure 3: Loop function to call for different time series data via API

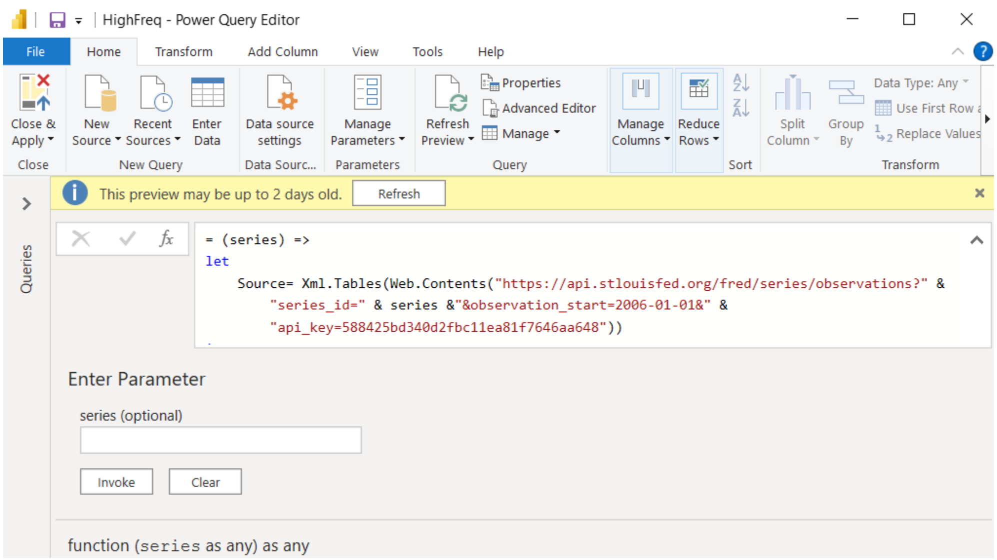
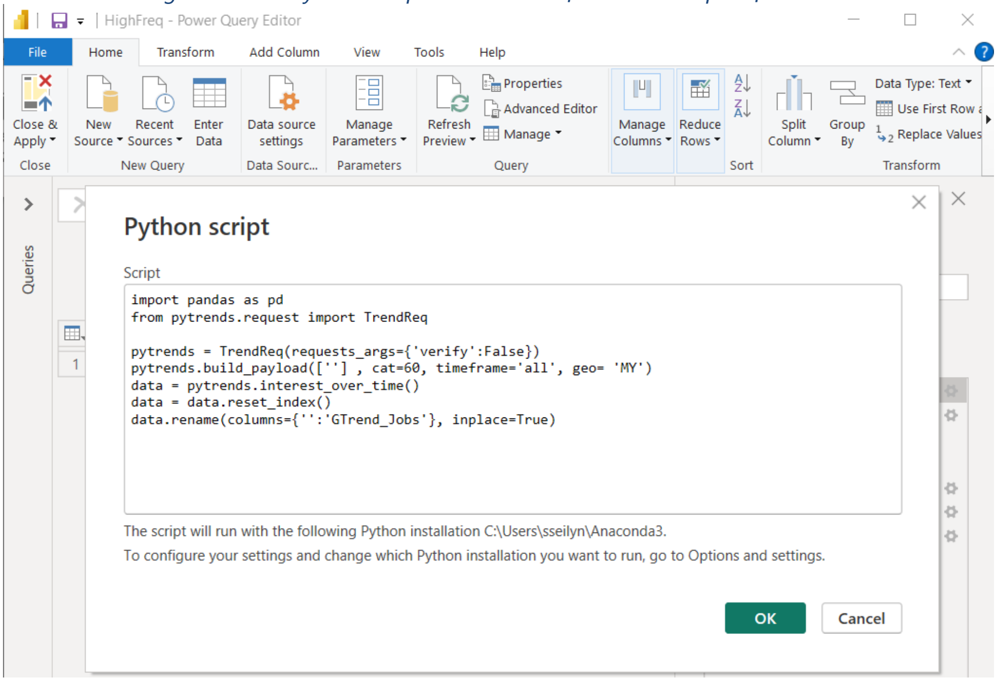


Figure 4: Run Python Scripts in Power BI for more complex functions



After the data is extracted into Power BI, the subsequent data analysis steps closely mirror standard procedures employed in most data analysis workflow. These steps involve data processing and transformation, including data cleaning, outlier detection, and seasonal adjustments for time series data, all before the data is visualized.

The relevant datasets are then published within a consumer-facing Power BI application. The central objective of this application is to enable users to assess the informational content of high-frequency indicators in tracking official statistics. This exploration can take two forms. First, users can assess the ability of high-frequency indicators to nowcast official statistics by calculating their correlation and using Python scripts to run multivariate predictive models. Second, for large granular datasets, BI tools provide the capacity to perform simple time series transformations, such as year-on-year growth and data aggregation across various dimensions, offering deeper insights and nuances underlying the official aggregated figures. For example, using daily price data at the product and retailer level, one can investigate the price-setting behavior of different types of retailers, assessing the frequency and extent of price changes and inspecting distribution differences across time and product types.

Another crucial aspect to underscore is the importance of ensuring that the workflow demands minimal maintenance when updating visualizations with new data. It is important to consider structuring the applied steps and Python codes for long-term ease of maintenance, including the use of loop functions over variables that impact the data extraction process. Ideally, the design should enable all data and visualizations to be refreshed with a single click of the refresh button, eliminating the need for frequent code modifications.

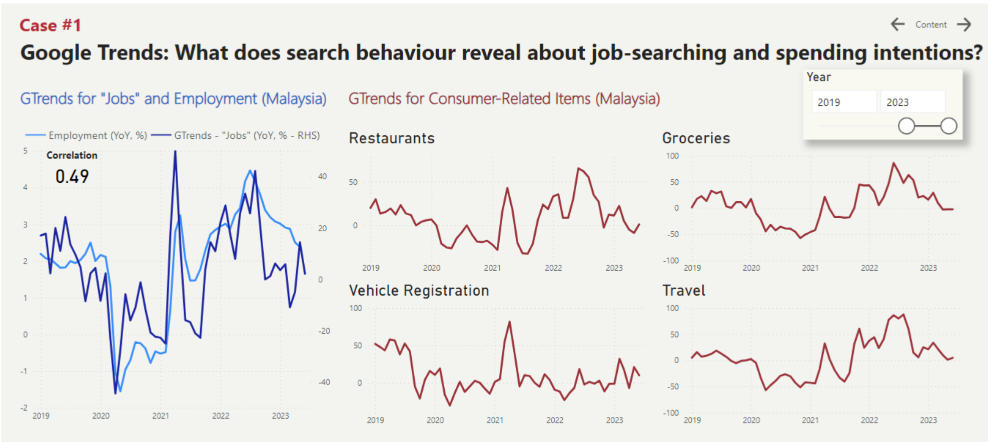
3. Empirical use cases of Power BI

This section explores three use cases that illustrate the utility of BI tools in exploring the informational value of high-frequency and granular indicators. The resulting Power BI dashboard, created using open-source data from Google searches, consumer goods prices, and sentiment extracted from news articles, is accessible through this link.

Case 1: Visualisation of high-frequency indicators to track economic activity

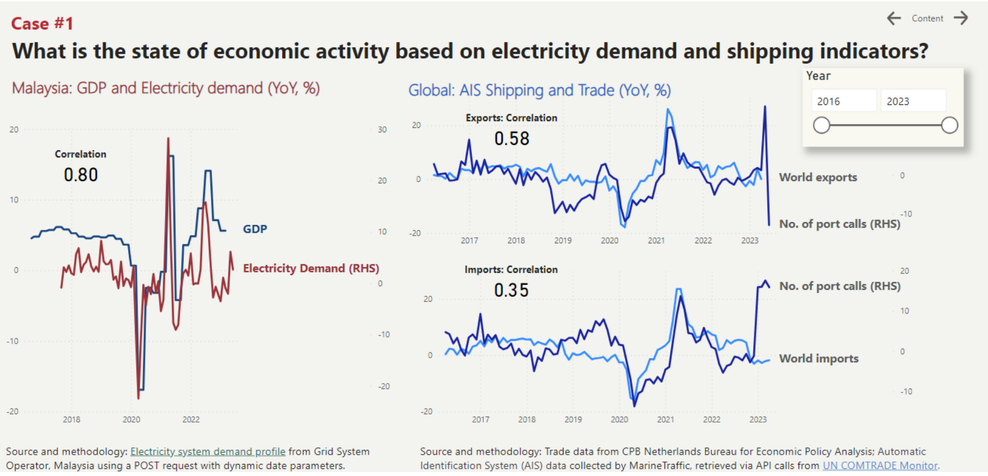
In anticipation of the publication of official figures, high-frequency indicators can provide analysts with early insights into potential trajectories. A noteworthy example of this is demonstrated in Figure 5, where we explore Google Trends data, inspired by research papers that have examined the relationship between Google Trends data and unemployment (e.g., Baker & Fradkin, 2017). Notably, job-related searches on Google Trends closely mirror employment growth fluctuations in Malaysia, highlighting the potential of these digital proxies. Beyond online search behaviors, other high-frequency indicators offer valuable insights. For instance, Figure 6 examines electricity demand data, captured at 10-minute intervals, which exhibits a strong correlation with GDP. Additionally, data from the Automatic Identification System, originally designed to prevent maritime collisions, provides insights into trade volume for the global external sector (Cerdeiro et al., 2020).

Figure 5: Screenshot of Power BI page on tracking economic activity using Google Trends



Source: Google Trends; employment data from Department of Statistics Malaysia

Figure 6: Screenshot of Power BI page on tracking economic activity using electricity data and AIS data



Source: Electricity system demand profile from Grid System Operator, Malaysia; Trade data from CPB Netherlands Bureau for Economic Policy Analysis, Automatic Identification System (AIS) data collected by MarineTraffic

Case 2: Visualisation of relatable nuances from granular data

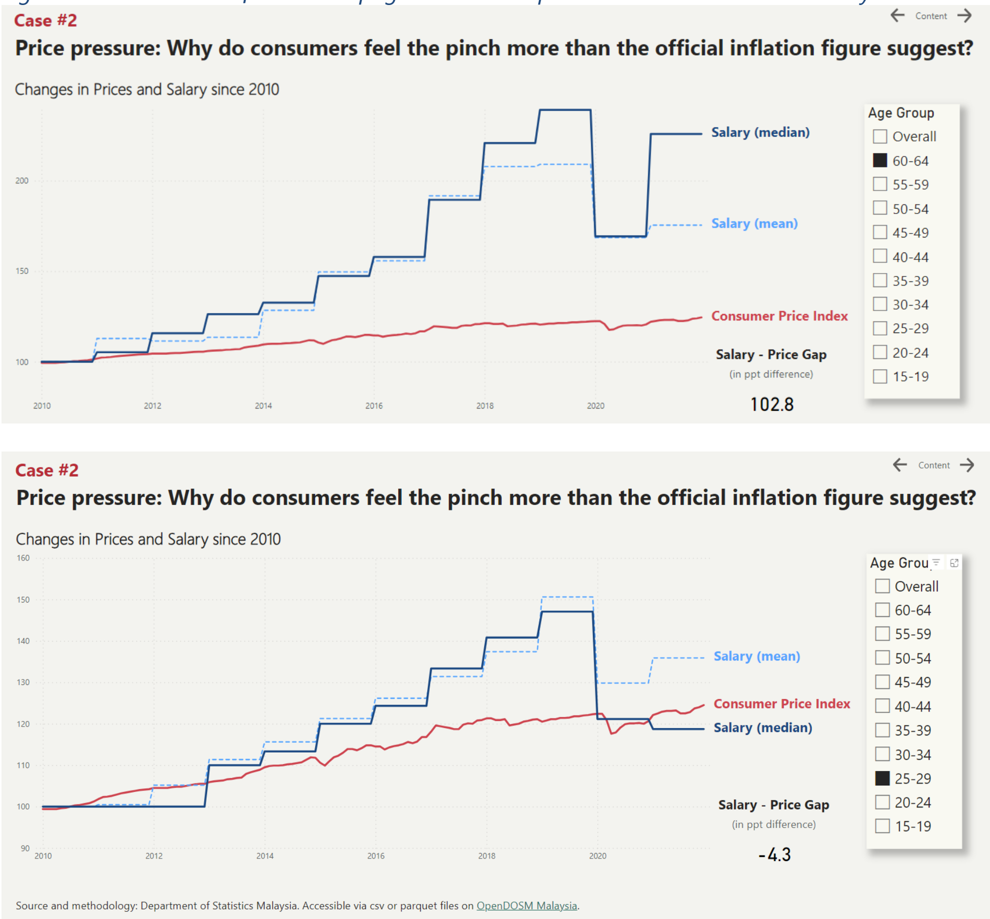
Effective communication of statistical insights to the public at times necessitates the sharing of narratives that resonate with everyday experiences. To illustrate the tangible effects of inflation, for example, it is crucial to explain why certain segments of the population might increasingly feel its economic impact, even when aggregate headline numbers suggest otherwise.

While salaries in aggregate have shown an upward trend surpassing the Consumer Price Index (CPI), an interactive chart within the Power BI dashboard allows users to compare the average salary for a selected age group with CPI over time. This comparison reveals a disparity for younger demographics: their salary trajectories, especially in recent times, have deteriorated relative to the CPI, as depicted in Figure 7.

To further contextualize this trend, one can draw upon cognitive phenomena like recency and memory biases, and link them to granular price data. People tend to remember recent events more vividly, and in the context of inflation, it is essential for the visualisation to highlight the more prominent price fluctuations of "everyday items," which include regularly purchased fresh food items (Vlasenko & Cunningham, 2015). Using daily price data provided by the Department of Statistics Malaysia, segmented by location and item type, it becomes evident that these everyday items have experienced more pronounced monthly price fluctuations than less frequently purchased items, as shown in Figure 8.

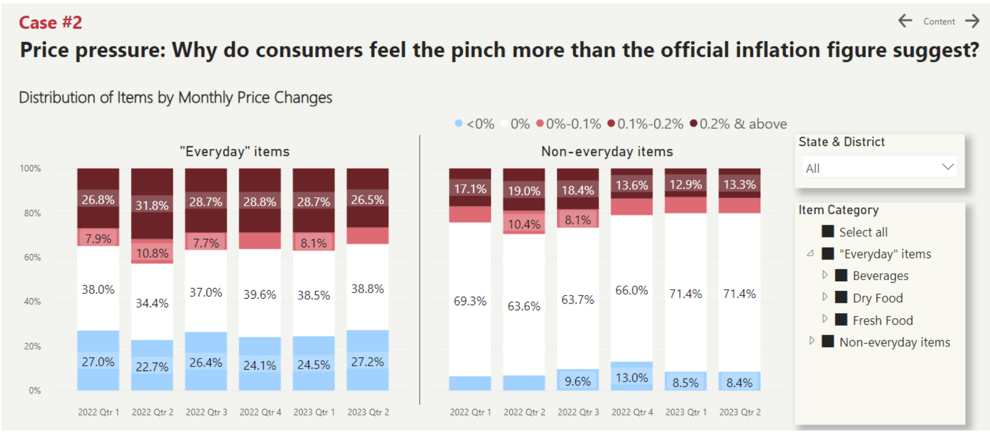
Interactivity is a critical aspect of effective data communication. By empowering the audience to explore and contrast their experiences against macro data and even peer demographics, they develop a nuanced understanding of underlying trends. Additionally, this approach ensures that the audience does not lose trust in official statistics but instead fosters a more informed and appreciative relationship with how aggregate data is derived for official publication.

Figure 7: Screenshots of Power BI page on the comparison between CPI and salary since 2010



Source: Department of Statistics Malaysia

Figure 8: Screenshot of Power BI page on distribution of items by monthly price change



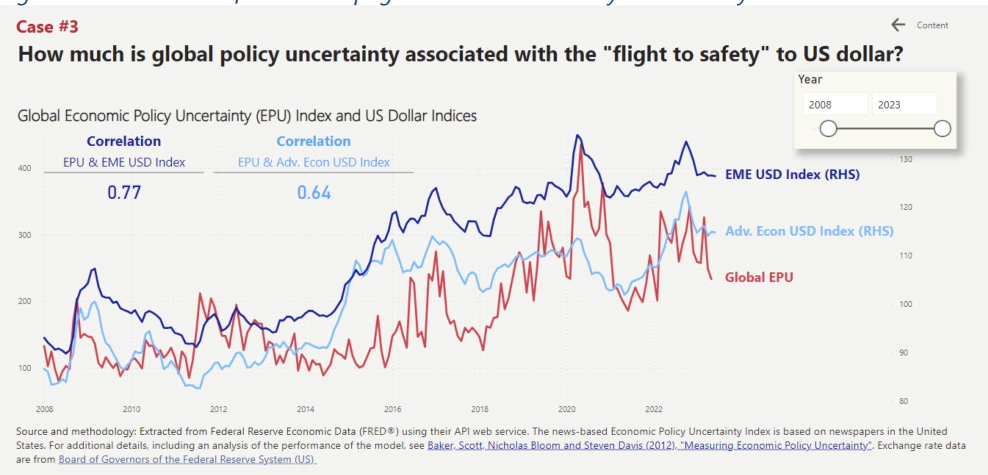
Source: PriceCatcher data collected by the Ministry of Domestic Trade and Cost of Living, made available by the Department of Statistics Malaysia

Case 3: Visualisation as a precursor to research

During certain periods, Malaysia witnessed rising concerns regarding the trajectory of its national currency, the ringgit. Figure 9 delves into this phenomenon, revealing two key insights. Firstly, Malaysia is not an isolated case; other emerging market economies (EMEs) within the comparative currency basket exhibit similar trends. Moreover, the depreciation pressures on EMEs appear intrinsically linked to global policy uncertainties, as proxied by the Economic Policy Uncertainty index developed by Baker, Bloom, & Davis (2016) using newspaper coverage frequency. This correlation supports the prevailing notion of the "flight to safety" phenomenon, wherein economic uncertainties drive investors towards more stable assets, notably the US dollar in this context.

One key takeaway from this visualization is its potential to ignite broader research interest, particularly in establishing the true causality between policy uncertainty and exchange rate movement.

Figure 9: Screenshot of Power BI page on Economic Policy Uncertainty and US Dollar Indices



Source: Federal Reserve Economic Data (FRED), news-based Economic Policy Uncertainty Index developed by Baker, S., Bloom, N., & Davis, S. (2016), retrieved from http://www.policyuncertainty.com/.

4. Practical considerations when using BI tools

With greater effort to leverage the capabilities of BI tools for data visualization and analytics of high-frequency and granular data, several practical prerequisites and considerations deserve attention:

Access to machine-readable data

Access to a well-structured data platform, whether internal or external, is an important prerequisite for an efficient development and maintenance BI dashboards. BI tools often support direct connections to both SQL and NoSQL databases, enabling internal users to extract data directly from an organization's databases. Externally, it is more efficient to have the capability to programmatically retrieve data through APIs or directly acquire online CSV or parquet files (Figure 10 and Figure 11). This functionality greatly simplifies the automation and scheduling of data refreshes, reducing the need for manual intervention. While such resources are increasingly common within statistical agencies, they may be less prevalent among private entities. These private entities may possess valuable granular data, but their primary focus may not be on facilitating broad data accessibility.

Figure 10: Machine-readable data via Python and online CSV/Parquet files on OpenDOSM (data portal by the Department of Statistics Malaysia)

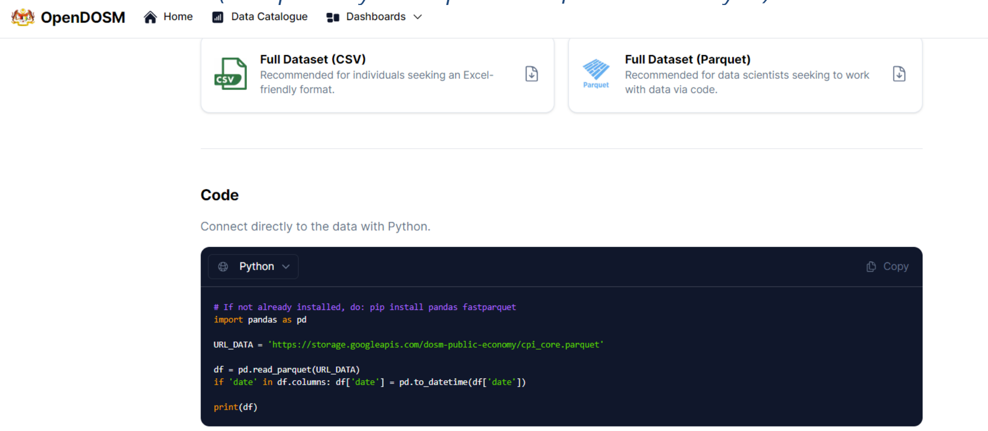
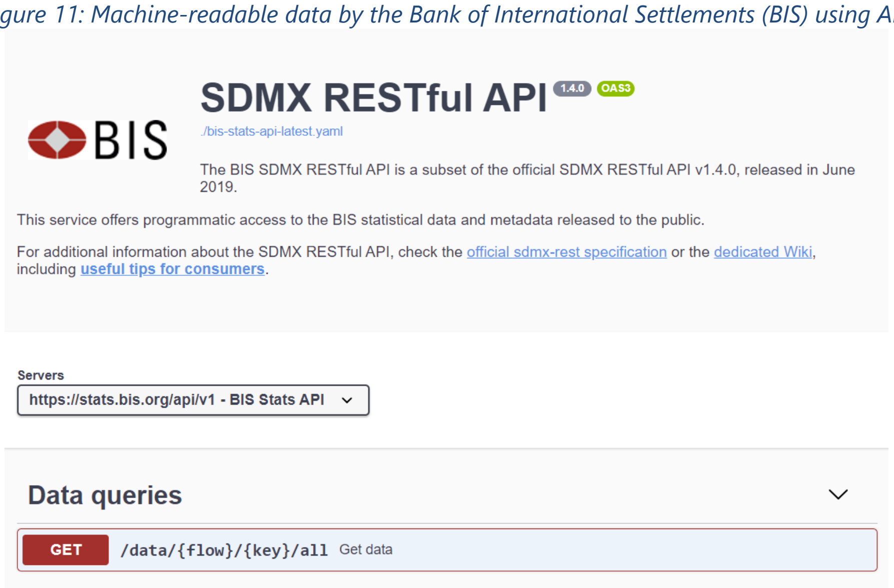


Figure 11: Machine-readable data by the Bank of International Settlements (BIS) using API



Purpose and limitations of BI tools

It is important to recognize that the primary objective of BI tools is data visualization. While these tools might offer provisions for advanced econometric or machine learning (ML) methodologies through plug-ins, they are not inherently designed for such sophisticated computations. Using Power BI as a case in point, when integrating Python, users face certain constraints:

- Only *pandas* data frames can be imported, limiting the types of Python data structures usable.
- Any Python script that runs longer than 30 minutes times out.

- Interactive elements in Python scripts, such as those requiring user input, halt the script's execution.
- For Python scripts to function correctly within the Power BI service, all associated data sources must be designated as public.

Such restrictions can lead to performance and maintenance bottlenecks. For instance, while integrating Python scripts for advanced processing, users might encounter execution time constraints, resulting in timeouts during data refreshes. Complex scripts should ideally be executed separately, and the output can be hosted in a database system that the chosen BI tool can connect to. Nevertheless, it is conceivable that future versions of BI tools may address these limitations.

Furthermore, there are scenarios that require a bespoke approach to data visualization and analysis rather than off-the-shelf BI tools. When an organization has unique requirements or seeks functionalities not available in conventional BI tools, such as econometrics or machine learning predictive models, a custom solution becomes imperative.

5. Conclusion

Effective communication plays a key role in making complex data and statistical insights accessible and understandable. In this dynamic, two key actors come into play: those responsible for crafting the communication, often data analysts and specialists, and those who consume and interpret it, which includes a broader audience, policymakers, and stakeholders.

For the audience receiving the information, truly impactful statistics go beyond mere numbers. It is crucial for those crafting the communication to present data as insights that resonate with the audience, providing them with a perspective that goes beyond the aggregate figures and fosters a nuanced understanding. However, while delving into granular data is valuable, it remains essential to ultimately connect these insights to the broader macroeconomic context, especially when considering policy changes that have broader consequences.

From an operational standpoint, the foundation for efficient data visualisation and communication is a robust data platform that facilitates data access and allows for automated data updates without manual intervention. As demonstrated, BI tools, at least in their current form, offer a cost-effective approach to integrate data from diverse sources and enhance analysts' capabilities in harnessing the potential of high-frequency indicators.

However, it is crucial to recognize the *interim* nature of this solution. While BI tools, like Power BI, excel as visualization tool, they may not be the most suitable platforms for more advanced operations involving econometrics or machine learning predictive models that are required to fully exploit the value of high-frequency, granular indicators. As it stands, the trajectory points toward the development of custom dashboards tailored to specific needs.

References

Baker, S., & Fradkin, A. (2017). "The impact of unemployment insurance on job search: Evidence from google search data". *Review of Economics and Statistics*, Vol. 99/5, pp. 756- 768, http://dx.doi.org/10.1162/REST_a_00674.

Baker, S., Bloom, N., & Davis, S. (2016). Economic Policy Uncertainty Index, retrieved from http://www.policyuncertainty.com/, July 17, 2023.

Baker, S., Bloom, N., & Davis, S. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636. https://doi.org/10.1093/qje/qjw024

Board of Governors of the Federal Reserve System (US), Nominal Advanced Foreign Economies U.S. Dollar Index [DTWEXAFEGS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DTWEXAFEGS, July 17, 2023.

Board of Governors of the Federal Reserve System (US), Nominal Emerging Market Economies U.S. Dollar Index [DTWEXEMEGS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DTWEXEMEGS, July 17, 2023.

Cerdeiro, D.A., Komaromi, A., Liu Y., & Saeed, M. (2020). Automatic Identification System (AIS) data collected by MarineTraffic; available at UN COMTRADE Monitor. https://comtrade.un.org/data/ais.

Cerdeiro, D.A., Komaromi, A., Liu Y., & Saeed, M. (2020).World Seaborne Trade in Real Time: A Proof of Concept for Building AIS-based Nowcasts from Scratch. IMF Working Paper No. 2020/057. https://www.imf.org/en/Publications/WP/Issues/2020/05/14/World-Seaborne-Trade-in-Real-Time-A-Proof-of-Concept-for-Building-AIS-based-Nowcasts-from-49393

Department of Statistics Malaysia. Various data retrieved from OpenDOSM https://open.dosm.gov.my/, July 17, 2023.

Grid System Operator, System Generation Profile, retrieved from https://www.gso.org.my/SystemData/CurrentGen.aspx, July 17, 2023.

Hogue J., & DeWilde, B., (n.d.) pytrends - Pseudo API for Google Trends. https://pypi.org/project/pytrends/.

Loke, P. L., & Khong, F. A. (2017). Inflation: Perception vs. Reality, Box Article 1. *BNM Quarterly Bulletin*, First Quarter 2017, Bank Negara Malaysia. https://www.bnm.gov.my/documents/20124/770502/p3_ba1.pdf

Microsoft. (2023). Use Python in Query Editor. https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-in-query-editor.

Microsoft. (2022). Run Python scripts in Power BI Desktop. https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-scripts

Vlasenko, P., & Cunningham, S. R. (2015). Capturing the Inflation that People Experience: The Everyday Price Index vs. the Consumer Price Index (Working Paper 004). American Institute for Economic Research. https://www.aier.org/wp-content/uploads/2015/07/WP004-EPI-Polina-Vlasenko-PV.pdf