UN-ECLAC, UNCTAD, and IMF Webinar
Measurement of the digital economy and trade in Latin America and the Caribbean

November 8-11, 2022

STATISTICS

Leveraging Big Data and data science to compile economic statistics

Margarida Martins
Statistics Department, IMF

Outline

- **Overview of data sources for compilation of GDP and economic activity indicators**

- **Examples of non-traditional data sources to measure economic activity**

- **Using Google Places API and Google Trends data to develop HFIs**

- **OECD Weekly Tracker of Economic Activity**

Overview of data sources for compilation of GDP and economic activity indicators

Motivation

- **Policy makers require both high-frequency and timely data to assess economic conditions in real-time:**
  - ▸ Timely annual information is useful but does not provide sufficient information for policy makers about the business cycle;
  - ▸ Quarterly/Monthly data released with a lag of several months does not allow policy makers to make adjustments in real time.

- **The COVID-19 pandemic increased the need for granular, high-frequency, and timely data, to better understand the impact of the pandemic and the path of economic recovery:**
  - ▸ Increase the frequency of economic indicators (e.g., from quarterly to monthly GDP);
  - ▸ Nowcast traditional economic indicators;
  - ▸ Analysis of evolving structural changes in real-time (e.g., business opening, closing);
 - ▸ Assess impact of the pandemic by business type and location.

Illustration of the Tradeoff Between Timeliness and Accuracy

timeliness

reference period t

| Forecast | Nowcast | Early estimate | Current estimate |

accuracy
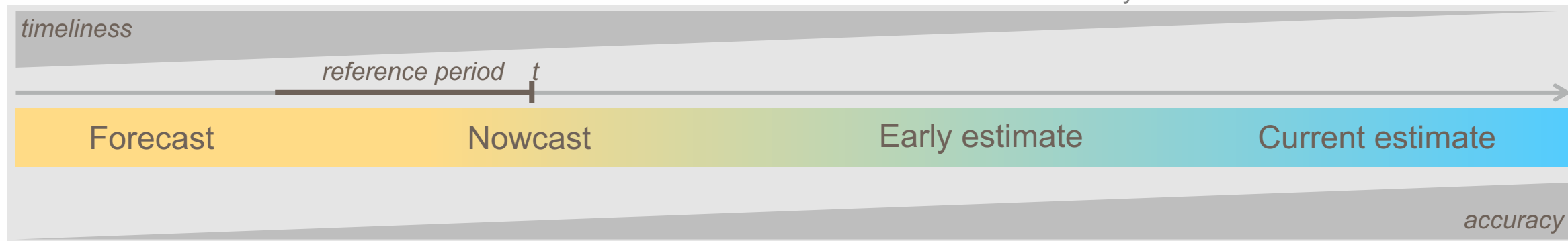
Overview of data sources

- Compilation of GDP and other macroeconomic statistics traditionally relies on data sources that are available with a time lag.

  ► With **digitalization** and increasing **availability of administrative data**, a new range of non-traditional data sources is emerging and can potentially be used to improve estimates of GDP, as input for nowcasting exercises, and for the development of high frequency economic indicators.
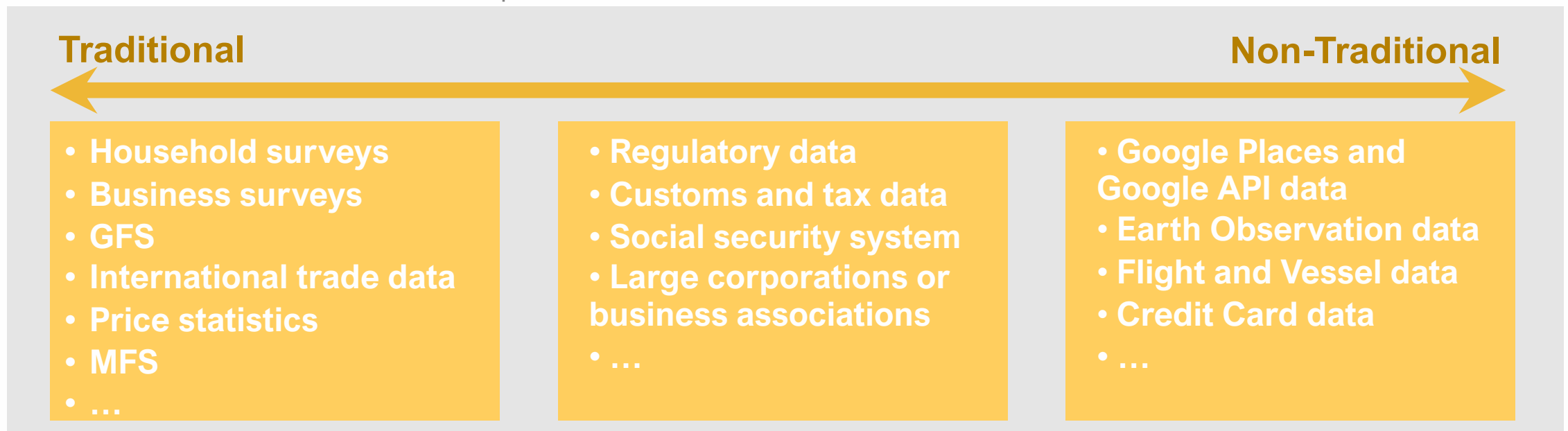
Example of Use of High-Frequency Indicators
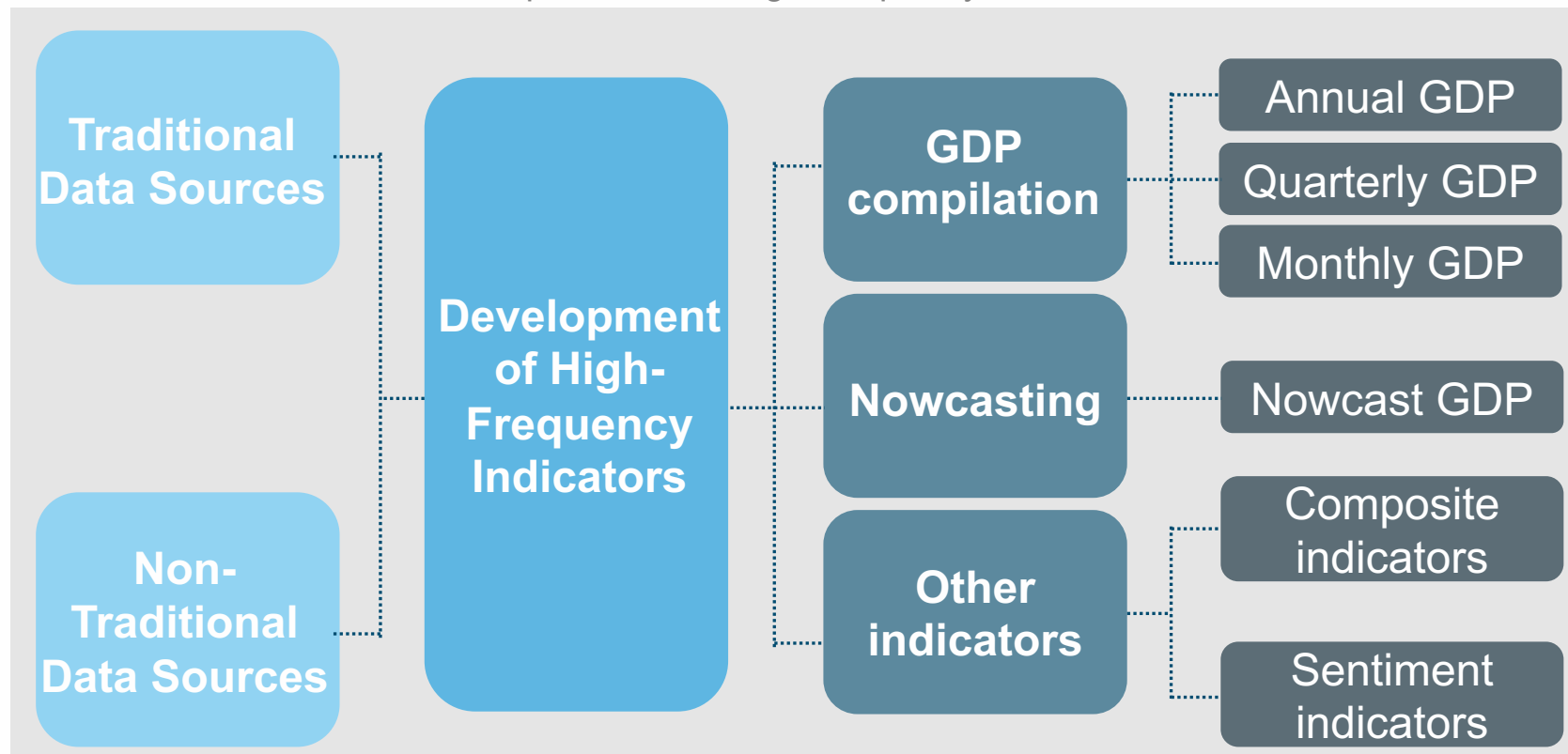
Overview of data sources

- **Non-traditional data sources present increased timeliness, frequency, and granularity, but some challenges need to be addressed:**

  ► Securing **data access** on a regular basis with consistent coverage, available variables, and frequency;

  ► Ensure enough **coverage** of economic activity and time series that are long enough;

  ► **Potential conceptual misalignment** between the Big Data source and the target statistic being produced, as the Big Data source was in general not developed with the objective of producing statistics.

Examples of Data Sources from Traditional to Non-Traditional

Traditional ←——————————————————————→ **Non-Traditional**

• **Household surveys** • **Business surveys** • **GFS** • **International trade data** • **Price statistics** • **MFS** • …	• **Regulatory data** • **Customs and tax data** • **Social security system** • **Large corporations or business associations** • …	• **Google Places and Google API data** • **Earth Observation data** • **Flight and Vessel data** • **Credit Card data** • …

GDP and Nowcasting

- **For most economies, there is a set of available high-frequency information**
  - ▶ These indicators are useful but individually do not provide an aggregate picture of the whole economic growth.
  - ▶ Finding a way of combining them would result in an aggregate indicator for the whole economy.

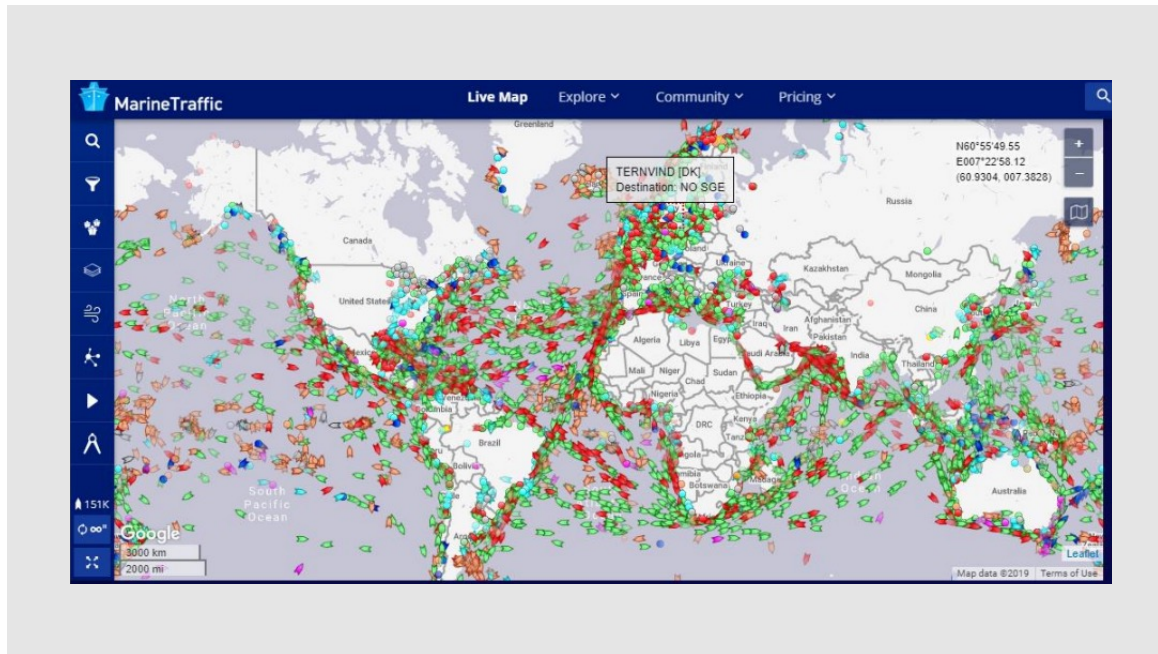Example of Types of Indicators by Compilation Methods and Temporal Synchrony

		COMPILATION METHOD		
		Statistical	Econometric	Accounting
TEMPORAL SYNCHRONY	Coincident	Business cycle analysis	Business cycle analysis	Descriptive analysis of the economy
	Leading	Early warning of turning points	Nowcasting/ Forecasting	
Examples:		Conference Board, Composite Leading Indicators (OECD).	PCA, Factor models, ADL, ARIMA, VARs.	Canada's monthly GDP.

Examples of non-traditional data sources to measure economic activity

Examples of Non-traditional Data Sources

- **Automatic Identification System (AIS) Data**
  - ▶ Information include location, speed, and status of vessels (e.g., weekly port calls and trade volume).
  - ▶ Possible use to nowcast trade.

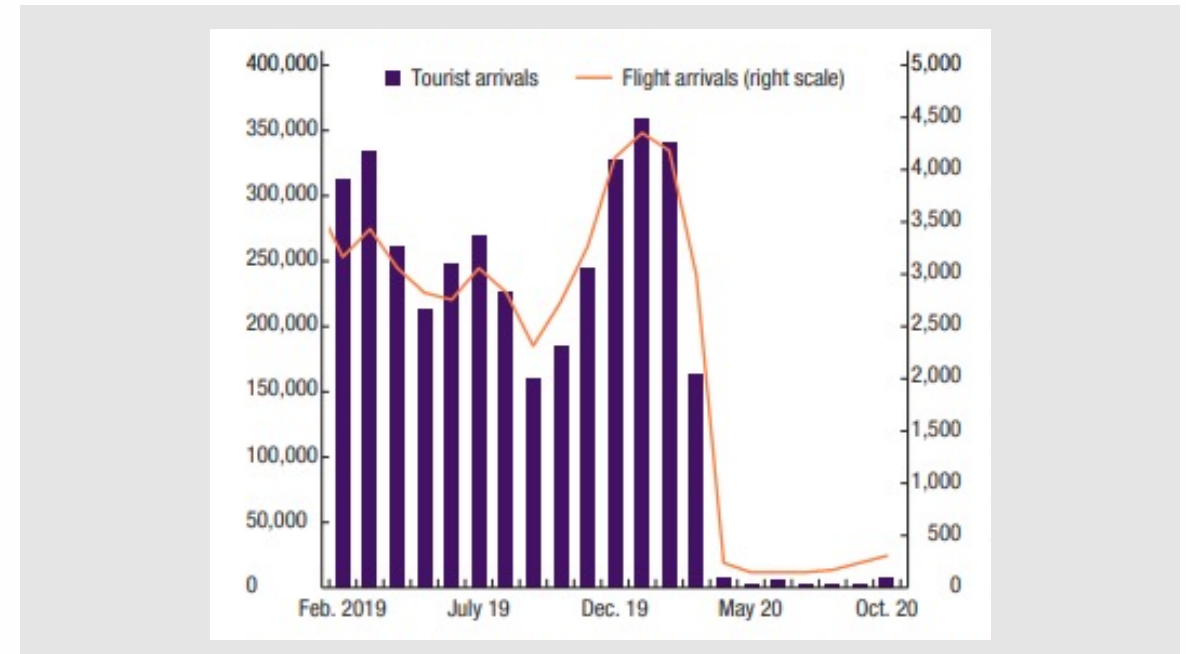 A Snapshot of Global Vessel Traffic Based AIS Data

- **Flight Data**
  - ▶ Possible use in air travel estimates, estimates for the tourism activity.

  Costa Rica: Impact of the Pandemic on the Tourism Sector – International Tourist and Flight Arrivals

Source: https://www.imf.org/~/media/Files/Publications/WP/2019/wpiea2019275-print-pdf.ashx, based on MarineTraffic Data.
Note: Different types of vessels are shown in different colors.

Source: https://www.imf.org/en/Publications/Departmental-Papers-Policy-Papers/Issues/2021/02/19/Tourism-in-the-Post-Pandemic-World-Economic-Challenges-and-Opportunities-for-Asia-Pacific-49915, based on Banco Central de Costa Rica; FlightRadar24; Instituto Costarricense del Turismo; and IMF staff estimates.

Examples of Non-traditional Data Sources

- **Earth Observation Data**
  - ▶ Using satellite imagery data to improve official statistics on a wide range of topics including agriculture, climate, business activity, and transport.
  - ▶ Pilot projects include crop density, agricultural statistics, land cover and use statistics, urban-rural systems, climate data, and crude oil inventory.

Map of Nighttime Lights in 2010

Brazil and Mexico: Examples of Initiatives using Satellite Data

Source: https://www.imf.org/en/Publications/WP/Issues/2019/04/09/Illuminating-Economic-Growth-46670.

Source: http://brazildatacube.org/en/about-brazil-data-cube-2/ and https://www.inegi.org.mx/investigacion/geomediana/#Metadatos.

Examples of Non-traditional Data Sources

- **Electricity Consumption Data**

  ▶ Possible use to estimate activity of economic sectors as manufacturing industry, or as input to nowcast GDP.

- **Tax Data**

  ▶ Value Added Tax (VAT) and other tax data capture both turnover and expenditure

  ▶ Possible use to track trade, hotels and restaurants, and other activities

Colombia: Early Estimation Indicator for Manufacturing Industry and 95 Percent Confidence Intervals

Chile: Daily Retail Sales Index, Monthly Average, Annual Growth Rate
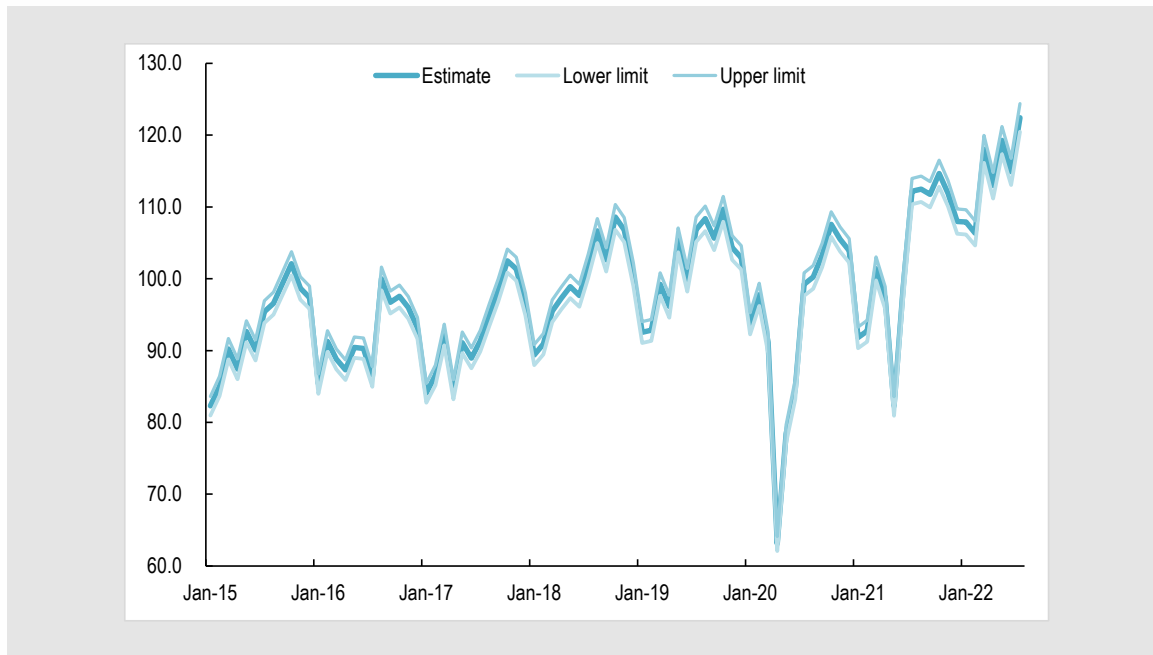
Source: IMF, based on DANE data (as of September 12, 2022), available at Estadísticas experimentales (dane.gov.co).
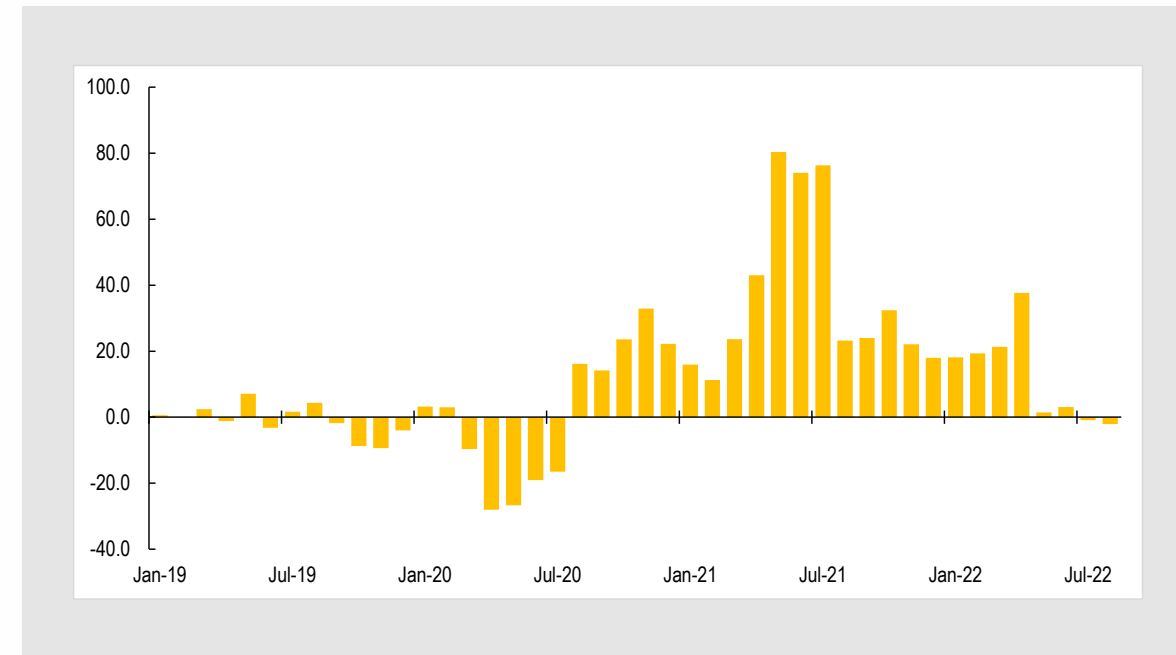
Source: IMF, based on Banco Central Chile data (as of September 13, 2022), available Estadísticas Experimentales - Banco Central de Chile (bcentral.cl).

Examples of Non-traditional Data Sources

- **Scanner Data**
  - ▶ Use of scanner data from supermarket chains and other retailers, as well as online prices obtained from web scraping, to compile price indices
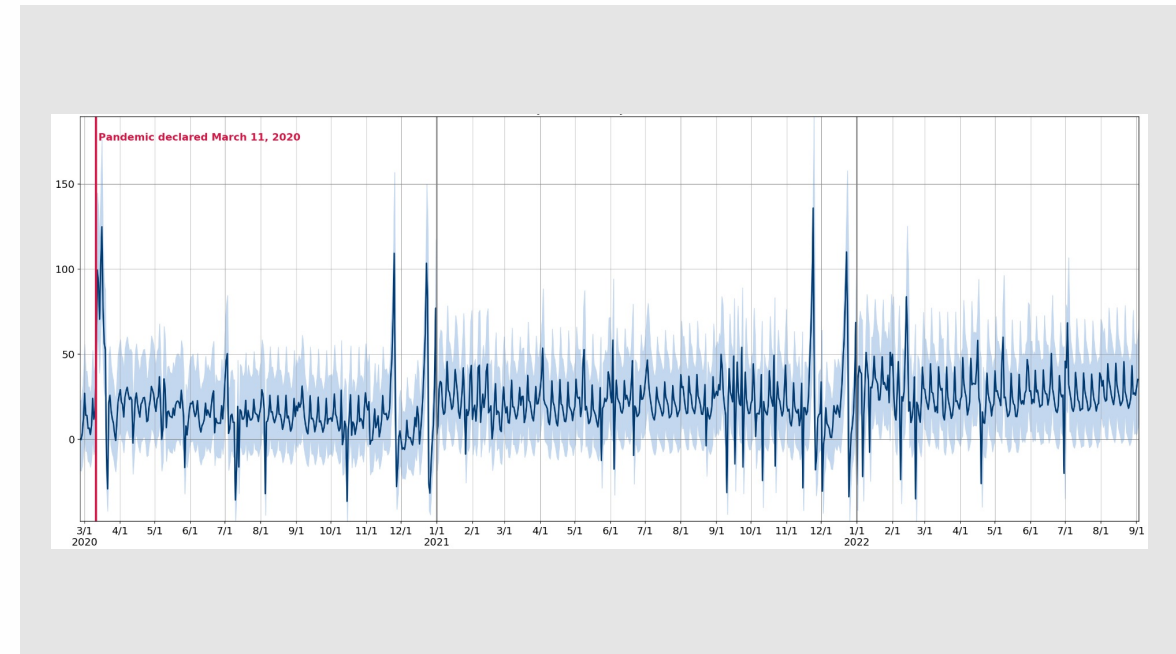
  2020 CPI Manual: Chapter on Scanner Data
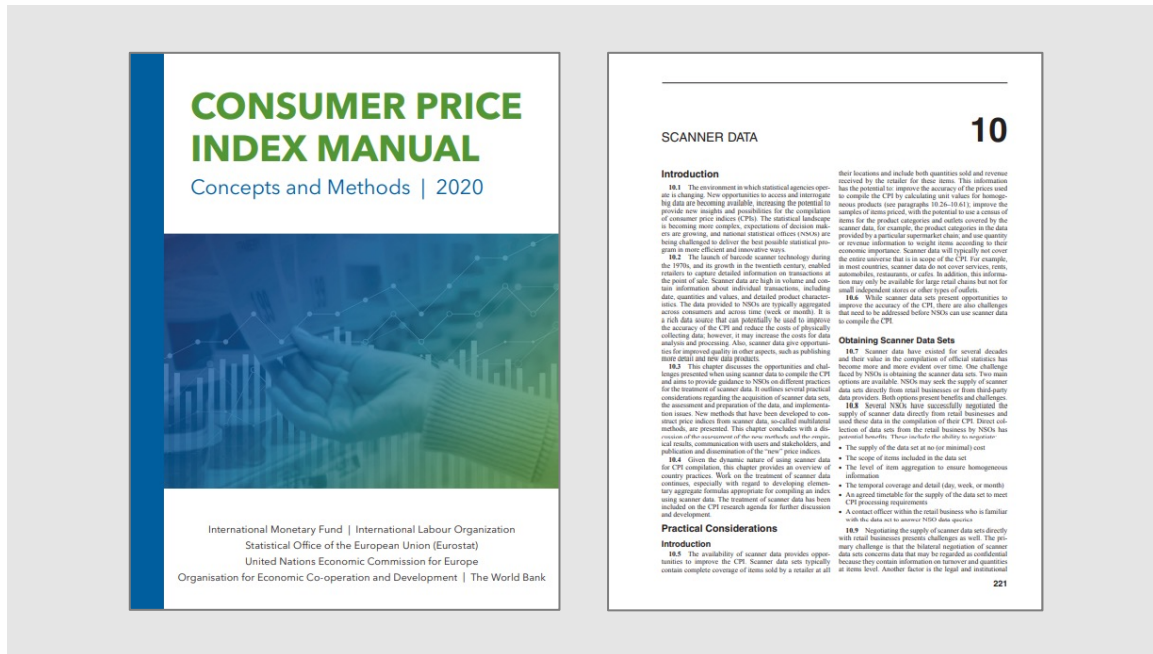
- **Credit Card Data**
  - ▶ Possible use to understand changes in consumption patterns and estimate consumption expenditure

  United States: Spending on Food and Beverage Stores – An Event Study Based on Payment Card Transactions

Source: https://www.imf.org/-/media/Files/Data/CPI/cpi-manual-concepts-and-methods.ashx.

Source: US Bureau of Economic Analysis, available at https://www.bea.gov/recovery/estimates-from-payment-card-transactions (as of September 13, 2022).
Note: Chart shows the difference from the typical level of spending without COVID-19-related changes in the economy. The typical level corresponds to a value of 0. The shaded area represents 95 percent confidence interval bands.

Using Google Places API and Google Trends data to develop HFIs

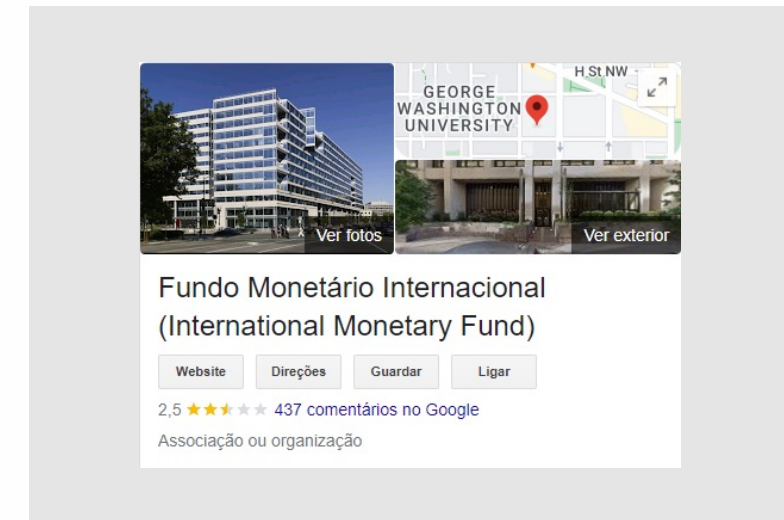# Using Google Places API/ Trends Data for HFIs

- **Google Places API**

  - ► The Google Places API allows users to extract information about Places from the Google Maps Platform.

  - ► Users can extract 23 fields of information for each Place, including name, address, places ID, business status, latitude/longitude, reviews, pricing level, and hours of operations.
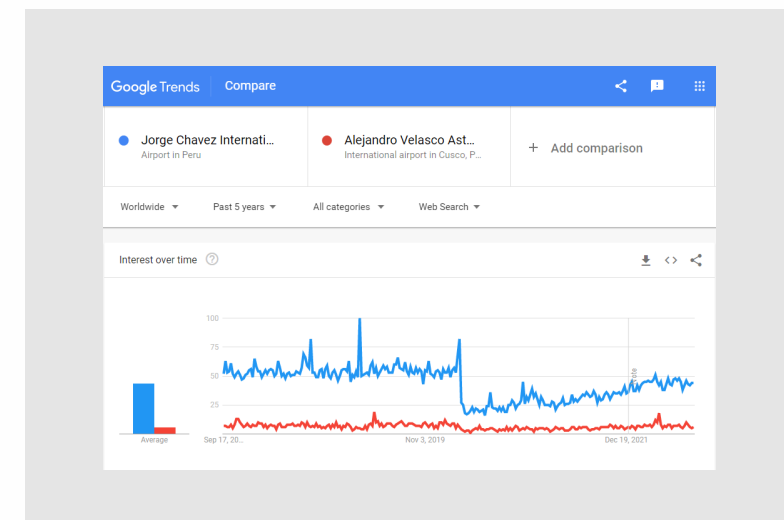
- **Google Trends Data**

  - ► Google Trends are a measure of interest in a topic relative to all other topics over time. A topic can be anything from a person or event to a business or specific product.

  - ► Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term.

  - ► When topics relate to a business, industry, or product the trend could be indicative, at least to some extent, of economic activity.

Example of Google Places API Data

Example of Google Trends Data

Using Google Places API/ Trends Data for HFIs

- **Method: (A) Opening Status Indicators**
  - ▶ Places are given the status of "Open," "Temporarily Closed" or "Permanently Closed." This information can be used to produce several useful business dynamic indicators.

Operational Indicator

- ▶ Represents the share of Places in each geographic region that are operational at a given point in time weighted by the number of reviews.
- ▶ Weighting by reviews is intended to capture the impact of the size of the business, in which businesses with more reviews will have a larger impact on the movement in the indicator.

Example of Business Status by Week

Business	Reviews	Share of Reviews	Business Status	Operational Indicator
Week 1 – Initial Status: Places A, B, C, and D operational				
A	1000	40	Operational	
B	500	20	Operational	
C	500	20	Operational	
D	100	4	Operational	
E	400	16	Temporarily Closed	
	2500	100		84
Week 2 – Place F is temporarily closed				
F	1000	37	Temporarily Closed	
G	700	26	Operational	
H	500	19	Operational	
I	100	4	Operational	
J	400	14	Operational	
	2700	100		63
Week 3 – All places are operational				
F	1000	33	Operational	
K	700	24	Operational	
L	500	17	Operational	
M	400	13	Operational	
J	400	13	Operational	
	3000	100		100

Example of Operational Indicator: São Paulo - Bars
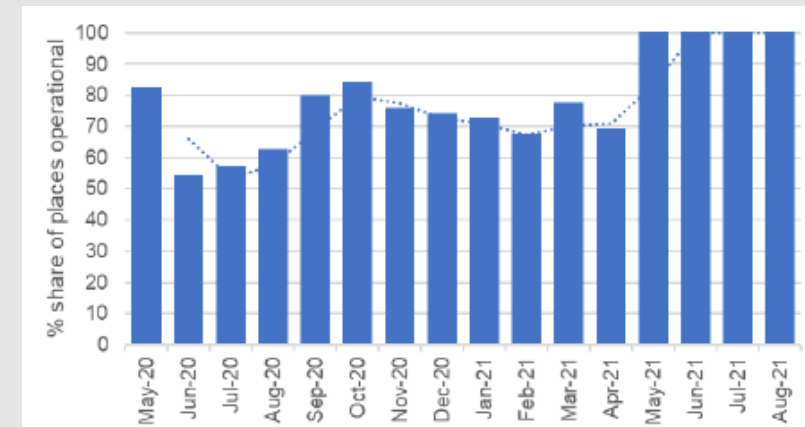
# Using Google Places API/ Trends Data for HFIs

- ## Method: (A) Opening Status Indicators

  - ▶ Places are given the status of "Open," "Temporarily Closed" or "Permanently Closed."
    This information can be used to produce several useful business dynamic indicators.

Business Re-opening Indicator

- ▶ Used to track the path and pace at which businesses that are temporarily closed in a region re-open, being of particular interest to assess the impact of government regulations during the COVID-19 pandemic.

- ▶ The indicator starts with the selection of a baseline cohort of places that are temporarily closed. Each week the status of each of these Places is examined to see if they have opened or remain temporarily closed.

- ▶ The indicator represents the share of Places in each geographic region that are operational at a given point in time weighted by the number of reviews.

Business Re-opening Indicator for Selected City Centers
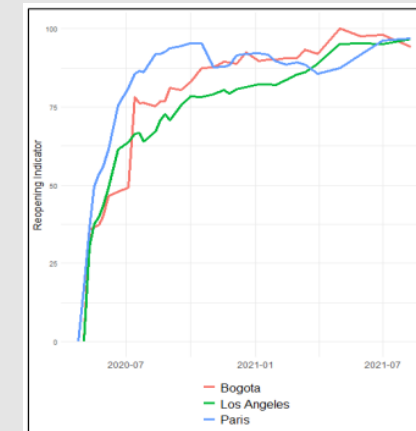
city	sam ple size	baseline	24-Apr-20	24-May-20	26-Jul-20	26-Aug-20	3-Nov-20	30-Jan-21	31-Mar-21	1-May-21	1-Jul-21	10-Aug-21
Atlanta	503	2-May-20		59	82.9	96	97	96.8	97.6	98.6	98.8	99
Bogota	339	2-May-20		37.2	76.4	76.7	87.6	90	92	100	97.9	94.1
Casablanca	209	17-May-20		9.1	32.5	40.7	47.4	54.5	60.3	93.8	65.1	67.9
Istanbul	566	24-Apr-20	0	41.7	64.7	78.3	83.6	83.9	88.5	94.9	88.7	91.5
Lagos	180	24-Apr-20	0	25	38.3	38.9	53.9	58.3	62.8	98.3	76.7	76.1
London	842	24-Apr-20	0	53.4	84.8	93.1	97.1	80.8	79	89.7	97.9	98.5
Los Angeles	1,001	2-May-20		40	64	72.8	79.1	82	88.8	95	95	96.6
Madrid	1,437	24-Apr-20	0	44.6	75.9	87.3	92.3	92.9	94.7	98.3	95.8	95.9
Manila	2,750	24-Apr-20	0	41	70.1	79.6	84.6	88.1	89.8	96.6	92.1	92.8
Milan	936	24-May-20		0	59.1	77.1	84	81.9	86.3	95.8	92	90.5
Mumbai	2,939	24-Apr-20	0	45.6	66	72.8	85.7	92.8	94.2	97.2	93.4	93.9
New York	1,278	24-Apr-20	0	47.4	75.7	84.7	92.4	92.3	94.3	96.9	97.2	97
Paris	1,645	24-Apr-20	0	53.5	86	92.5	87.7	89.7	85.6	87.4	96.2	96.9
Rome	1,343	17-May-20		25.4	64.6	82.7	88.2	88.3	89.4	97.2	93.8	94.6

Source: www.elibrary.imf.org/view/journals/001/2021/295/001.2021.issue-295-en.xml.
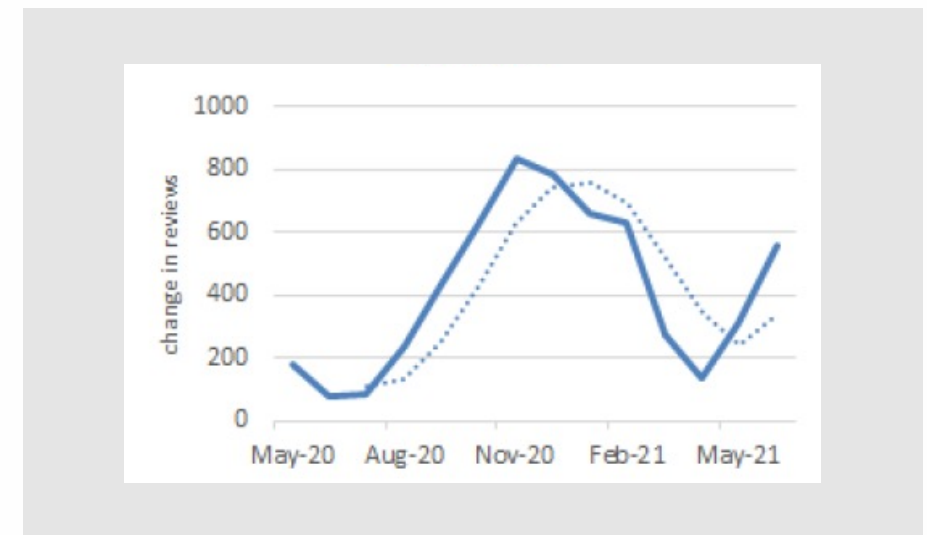
Using Google Places API/ Trends Data for HFIs

- ## Method: (B) Business Activity Indicators

  - ▶ Google Trends and Reviews are used as a proxy to business activity, assuming that there is a relationship between changes in interest in a topic(s) and changes in business activity.

Google Reviews as an Indicator of Business Activity

  - ▶ The Google Places API permits users to extract the number of reviews posted for a given Place and the average rating (ratings range from 1 - poor to 5 - excellent). It is assumed higher change in ratings are correlated with higher economic activity.

  - ▶ The rating was used to adjust the number of reviews such that a Place with 100 poorly rated reviews would have a lower weight than a Place with 100 highly rated reviews.

Example of Review Activity Indicator: São Paulo – Change in Restaurant Review

Source: www.elibrary.imf.org/view/journals/001/2021/295/001.2021.issue-295-en.xml.
Note: The dotted lines in the graphs are three-month moving averages.

Using Google Places API/ Trends Data for HFIs

- **Method: (B) Business Activity Indicators**

  ▶ Google Trends and Reviews are used as a proxy to business activity, assuming that there is a relationship between changes in interest in a topic(s) and changes in business activity.

  Google Trends as an Indicator of Activity

  ▶ Given the infinite number of possible search terms, Google has developed an algorithm to aggregate searches into "trend" categories (e.g., category of "Consumer Electronics").

  ▶ The Google Trends categories are then mapped to the *International Standard Industrial Classification of All Economic Activities* (ISIC), to create indicators of business activity.

Example of Business Activity Indicator: Brazil - Wholesale and retail trade; Repair of motor vehicles and motorcycles; Transportation and storage; Accommodation and food service activities
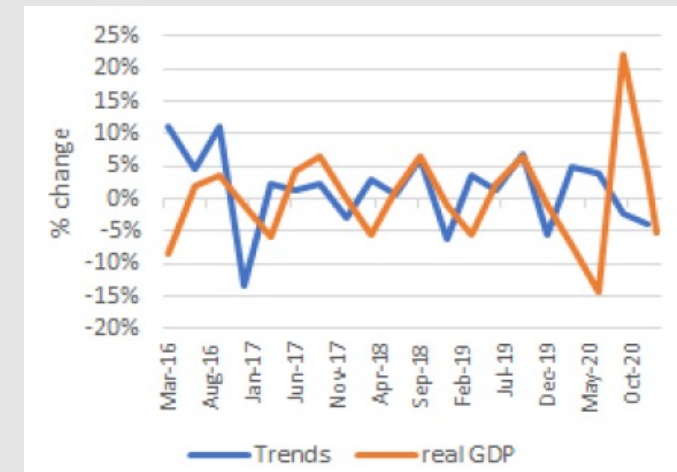
Source: www.elibrary.imf.org/view/journals/001/2021/295/001.2021.issue-295-en.xml.
Note: Indicator refers to ISIC Rev. 4, Sections G, H, and I.

OECD Weekly Tracker of Economic Activity
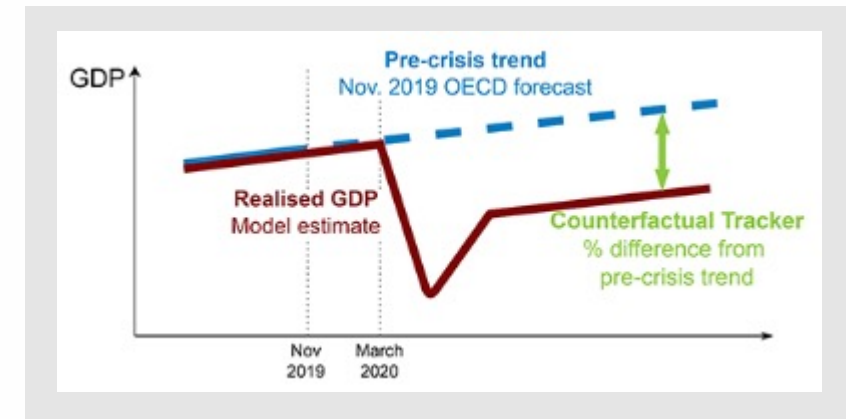
OECD Weekly Tracker of Economic Activity

► The **OECD Weekly Tracker of GDP growth** provides a real-time high-frequency indicator of economic activity using machine learning and Google Trends data.

  ► Available for 46 OECD and G20 economies.

  ► Applies a machine learning model to a panel of Google Trends data and aggregates information about search behavior related to consumption, labor markets, housing, trade, industrial activity and economic uncertainty.

The OECD Contrafactual Tracker

Source: https://www.oecd.org/economy/weekly-tracker-of-gdp-growth/.

► There are three series of the Weekly Tracker:

  ► **Tracker (yoy):** estimates of weekly GDP relative to the same week in the previous year. Covers the period from early 2020 to today.

  ► **Tracker (yo2y):** estimates of weekly GDP relative to the same week in the two years before (the 104-week difference). Covers the period from early 2020 to today.

  ► **Tracker (counterfactual):** estimate of the percent difference between weekly GDP and the pre-crisis GDP trend (as proxied by OECD forecasts made in November 2019). Available until the end of 2021.

  Each series has its own 95% confidence intervals (lower and higher bands).
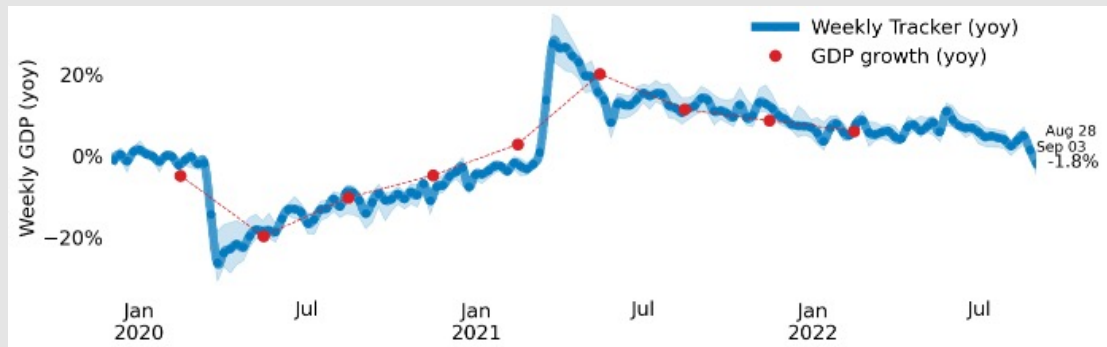
OECD Weekly Tracker of Economic Activity
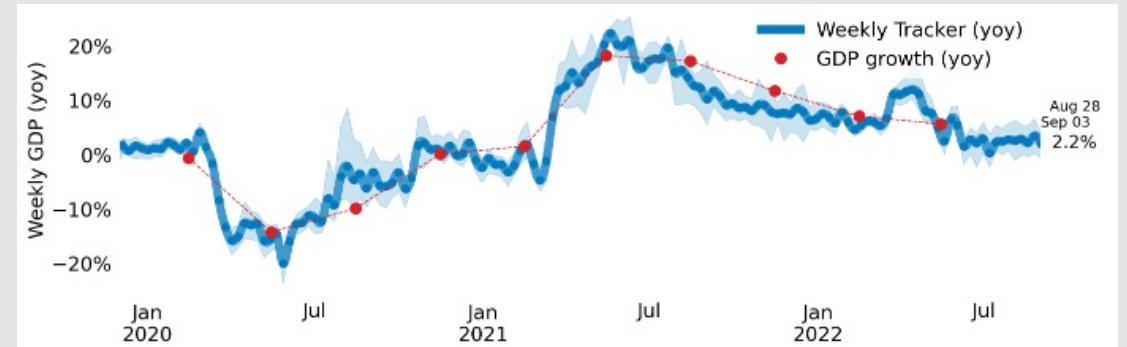
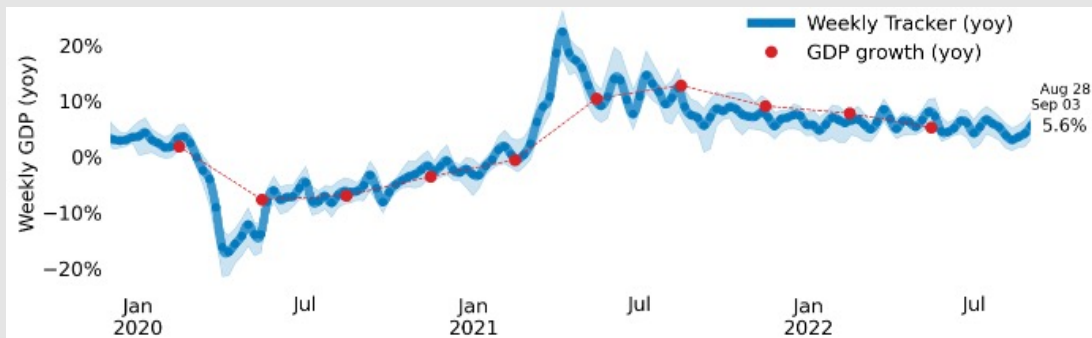Examples of the OECD Weekly Tracker: Weekly GDP Relative to Previous Year

Source: https://www.oecd.org/economy/weekly-tracker-of-gdp-growth/.
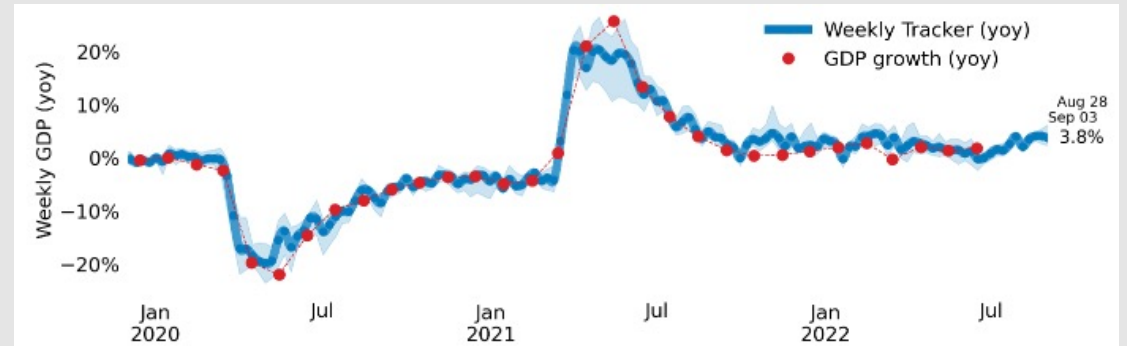Note: The Weekly Tracker provides an estimate of weekly GDP based on Google Trends search data and machine learning.

Thank you!

References

▶ Arslanalp, S., M. Marini. and P. Tumbarello, 2019, *Big Data on Vessel Traffic: Nowcasting Trade Flows in Real Time*, IMF Working Paper No. 19/275 (Washington: International Monetary Fund). Available at: www.imf.org/~/media/Files/Publications/WP/2019/wpiea2019275-print-pdf.ashx

▶ Austin, P., M. Marini, A. Sanchez, C. Simpson-Bell, and J. Tebrake, *Using the Google Places API and Google Trends Data to Develop High Frequency Indicators of Economic Activity*, IMF Working Paper No. 21/295 (Washington: International Monetary Fund). Available at: www.elibrary.imf.org/view/journals/001/2021/295/001.2021.issue-295-en.xml

▶ Hu, Y. and J. Yao, 2019, *Illuminating Economic Growth*, IMF Working Paper No. 19/77 (Washington: International Monetary Fund). Available at: www.imf.org/en/Publications/WP/Issues/2019/04/09/Illuminating-Economic-Growth-46670

▶ Woloszko, N., 2020, *Tracking activity in real time with Google Trends*, OECD Economics Department Working Paper No. 1634 (Paris, OECD). Available at: www.oecd-ilibrary.org/economics/tracking-activity-in-real-time-with-google-trends_6b9c7518-en.