



El papel de la Ciencia de Datos en las estadísticas de las Cuentas Nacionales

Los primeros pasos y los desafíos para su
profundización en Uruguay

Fernando Barbeito - BCU - Área de Estadísticas Económicas -
Departamento de Oferta

Objetivos

- Procesar y evaluar datos de la Dirección General Impositiva (DGI).
 - Posibilidad de contar con datos diarios.
- Procesar y evaluar datos de Google Trends y los reportes de movilidad. Analizar dónde, cuándo y cómo utilizarlos.
- Generación de datos a través de la técnica de Webscrapping.
- Adquirir experiencia en el manejo y procesamiento de grandes cantidades de datos.
- Establecer una infraestructura mínima y automatizada.
- Obtener insumos que habiliten nuevas investigaciones.

Un poco de contexto

- Necesidades de información de alta frecuencia y en tiempo real para la toma de decisiones:
 - Banco Central del Uruguay
 - Política monetaria.
 - Supervisión del sistema financiero.
 - Ministerio de Economía y Finanzas y otros
 - Gestión de políticas.
 - Sector privado
 - Decisiones de consumo.
 - Decisiones de ahorro e inversión.

Un poco de contexto

- El rezago de las Cuentas Nacionales (CCNN):
 - Las Cuentas Nacionales Trimestrales (CNT) tienen un rezago de 75 días.
 - No se publica indicadores adelantados de actividad económica
- Alternativas:
 - El *nowcasting* a través de modelos econométricos (Factores dinámicos, Mixed Frequency Models, Bridge Models).
 - Estimaciones *flash* del PIB con información e indicadores parciales disponibles.

Un poco de contexto

Métodos tradicionales de Cuentas Nacionales

- Las estimaciones de CNT se basan en la estructura de información de las estimaciones para el año de referencia.
- Utilizan la información de alta frecuencia proveniente de:
 - Encuestas.
 - Registros administrativos y aduaneros.
 - Información fiscal y contable.

Un poco de contexto

Métodos tradicionales de Cuentas Nacionales

- El procesamiento de esta info lleva tiempo y es la causa de los rezagos. No está libre de problemas:
 - Consistencia de los datos.
 - Errores de muestreo y de crítica.
 - Problemas de valoración, precios de transferencia, categorías y definiciones no acordes con las CCNN, entre otros.

Un poco de contexto

El desafío de la pandemia

- Necesidad de estimaciones de evolución económica en tiempo real.
- Falta de estadística básica (cierre de oficinas, cambios de metodologías, dificultad para encontrar a los informantes, etc.)
- Fallos en los modelos.

Un poco de contexto

El desafío de la pandemia

- Alternativas. Datos generados por:
 - Transacciones digitales.
 - Declaraciones fiscales regulares.
 - Operaciones web.
 - Geolocalización de dispositivos móviles.
 - Información de viajes por plataformas.

Un poco de contexto

El desafío de la pandemia

- Ejemplos de uso en Uruguay:
 - Procesamiento y uso de reporte de movilidad para la estimación de servicios personales.
- Otros países usan Google Trends para la estimación de ciertos servicios turísticos.

Un poco de contexto

El desafío de la pandemia

- Oportunidades que se abren:
 - Las declaraciones fiscales, si bien tienen un rezago, también permiten adelantar la información de ventas de las empresas formales.
 - La Factura electrónica se presenta como una herramienta potente para realizar este tipo de análisis.
 - Datos provenientes de operaciones de las tarjetas, las redes de retiro de efectivo y los POS (sistema de cobros en locales).
 - Los datos privados de las empresas, que ellas mismas producen y recolectan (por ej, mercado libre, amazon, etc..)

Un poco de contexto

El desafío de la pandemia

- Acceder y procesar esta información conlleva desafíos grandes para las oficinas de estadística.
- Algunos de los que hemos enfrentado:
 - En sistema de pagos no existe la capacidad para extender los requerimientos de información.
 - La información privada tiene costo y no esta producida para fines estadísticos.
 - La representatividad de la información “digital”.
 - Los problemas de la información de la DGI.

Cuestiones de infraestructura

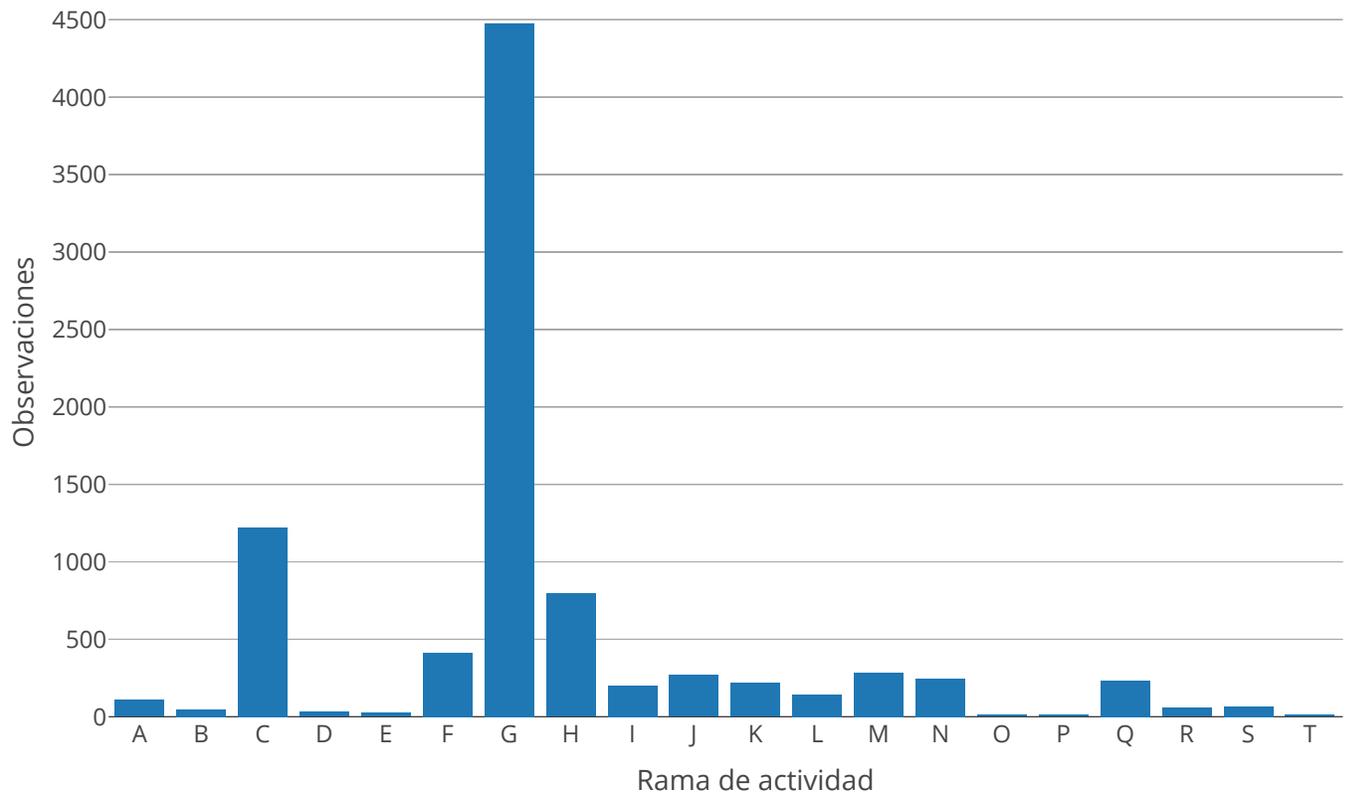
- Todas las bases procesadas se trabajaron en R.
- Las técnicas de webscrapping ya se encuentran incorporadas en rutinas automatizadas.

La base de datos

DGI

- Los datos provienen de la declaración del Impuesto al Valor Agregado (IVA) que realizan las empresas.
- El periodo de la base va desde febrero del 2016 hasta diciembre 2020. La frecuencia de los datos es mensual.
- Para evaluar la calidad de los datos, se construye un panel de empresas con el formulario 2176.
- La cantidad de empresas de la base de datos es de 160759.
- La cantidad de empresas que conforman el panel es de 8834. La cantidad de empresas de la EAAE 2012, año base de las estimaciones actuales de CCNN, fueron 5342.

Observaciones por rama



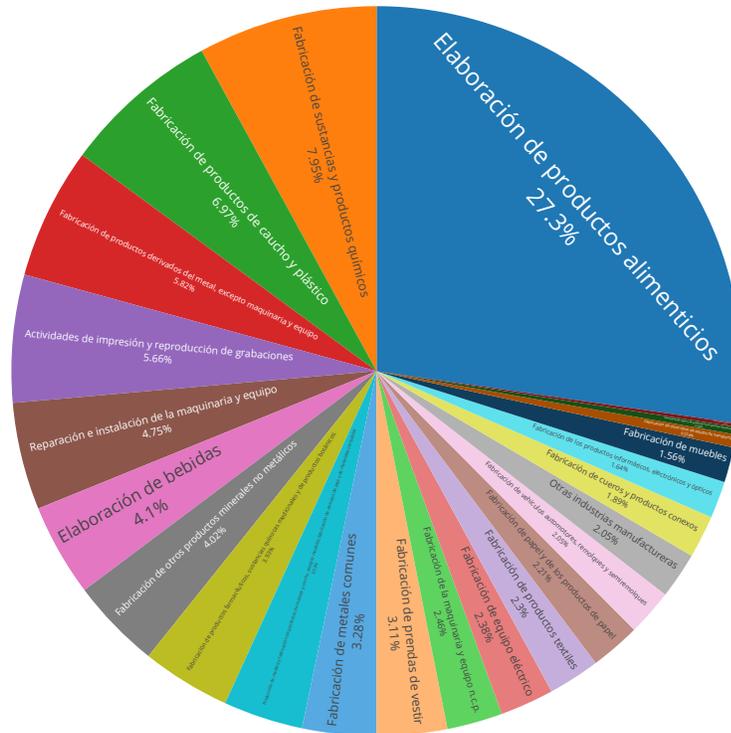
Evaluación de los datos

- Para evaluar los datos se compara las estimaciones de los datos de DGI y otras fuentes disponibles.
 - INE
 - Indicadores de volúmen físico de la industria manufacturara (IVFIM). Base 2018.
 - BCU
 - Producción a precios corrientes de las empresas grandes (modo 1) de algunas ramas, excepto comercio
 - Datos surgidos de la consistencia de diversas fuentes.
 - Exportaciones de bienes industriales y de servicios.

Evaluación de los datos

- Se pone foco en las estimaciones de la industria, ya que se cuentan con estimaciones fiables, debido a la mayor cantidad de fuentes, y con frecuencia mensual.
- Para comparar con las estimaciones del BCU, se construye una proxy de producción trimestral con las ventas.
- Para comparar con las estimaciones de la industria del INE, se toman las series mensuales de la DGI a dos dígitos, se construyen indicadores de valor y se deflactan con los índices de precios al productor proporcionados por el INE.
- Las series casi no tienen tratamiento.

Observaciones Industria a 3 dígitos de la CIIU

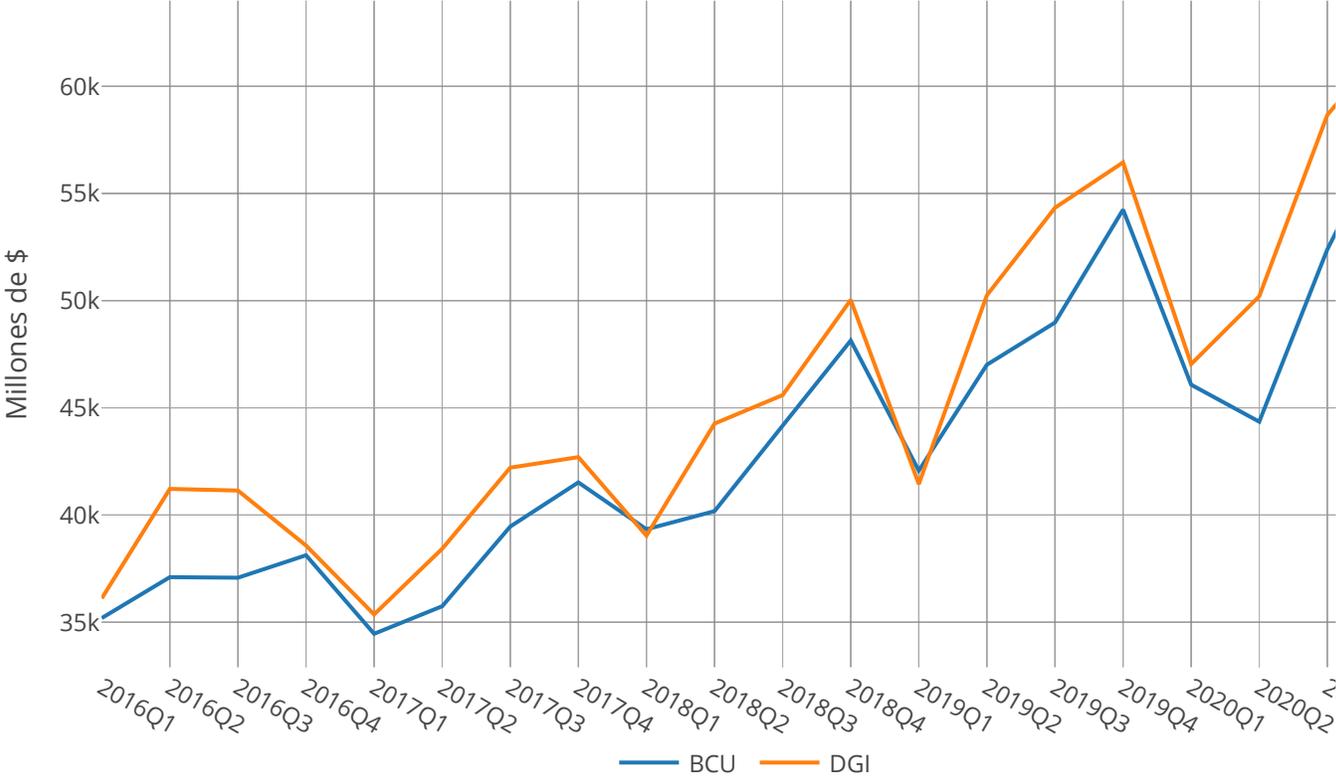


Industria Trimestral

BCU

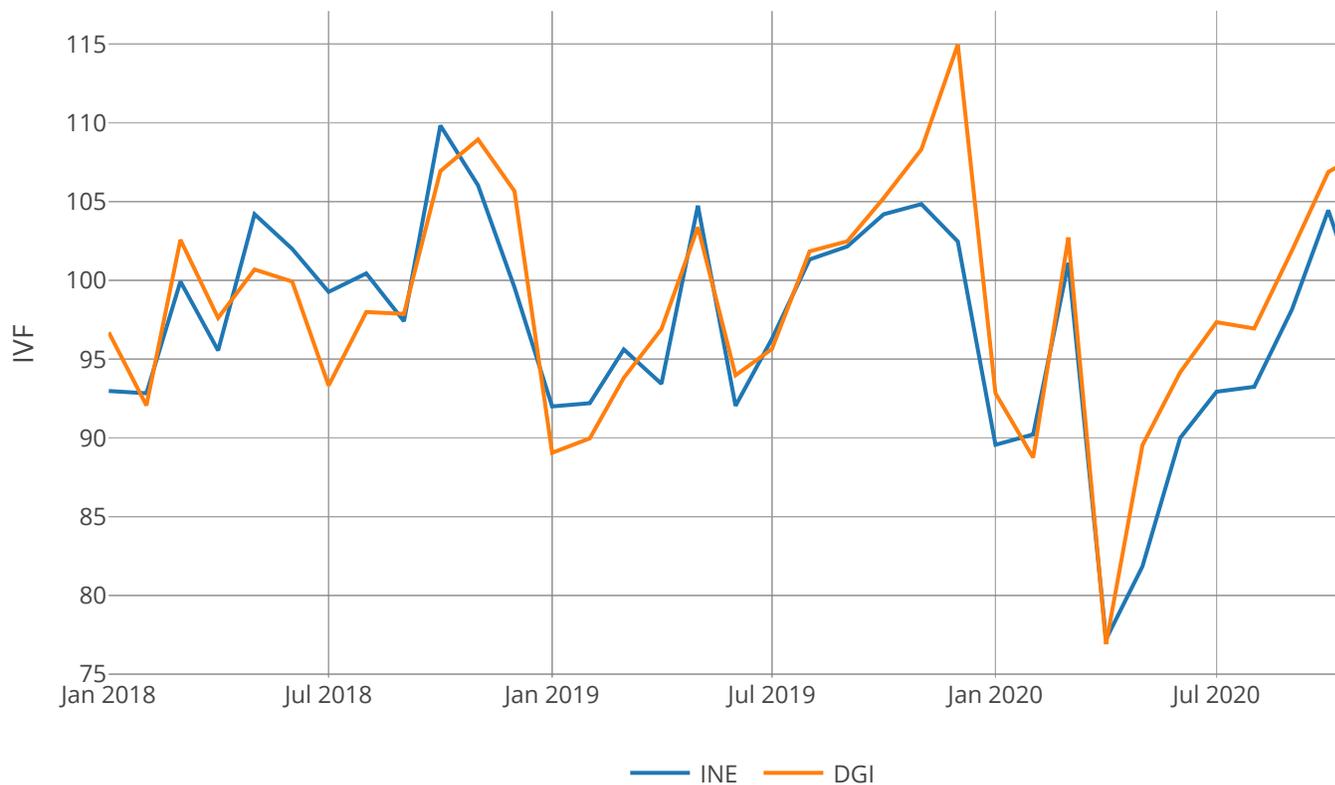


Exportaciones industriales

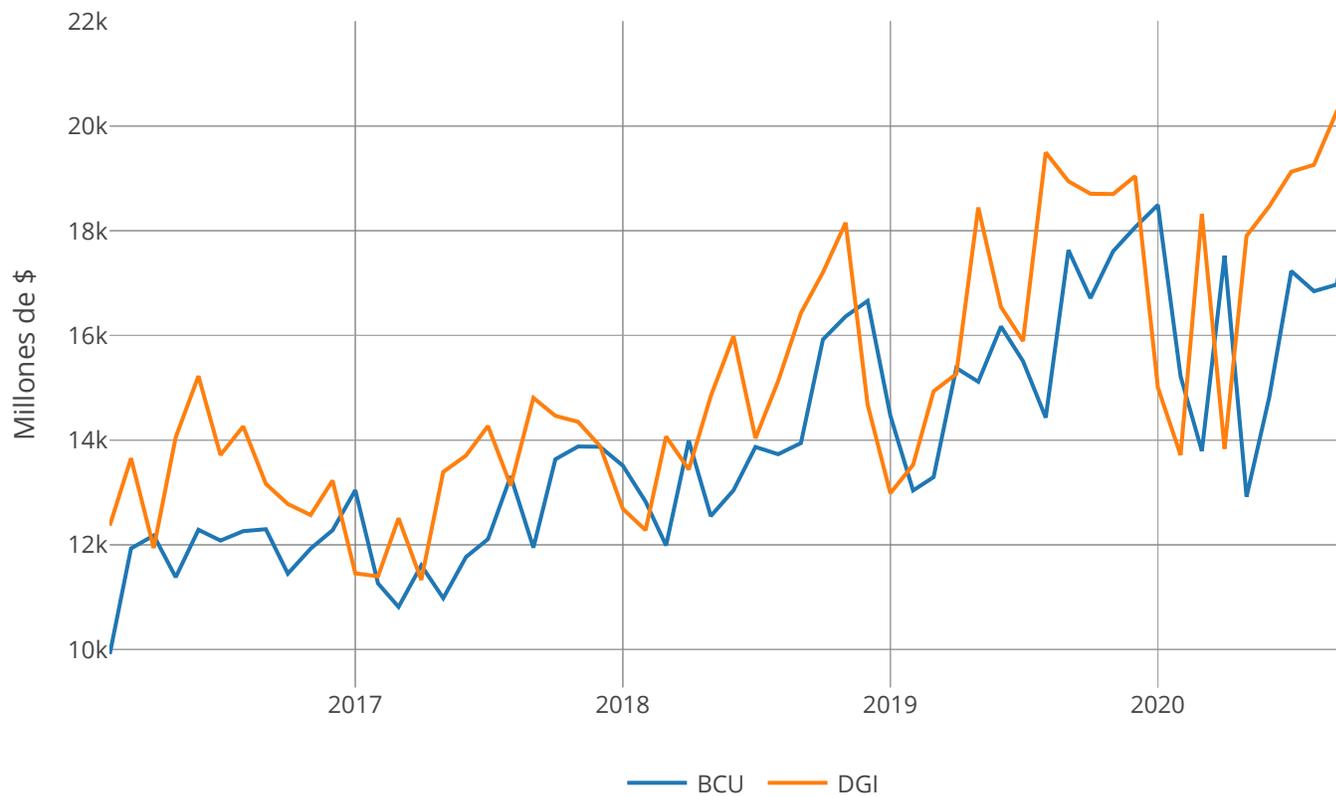


Industria mensual

INE

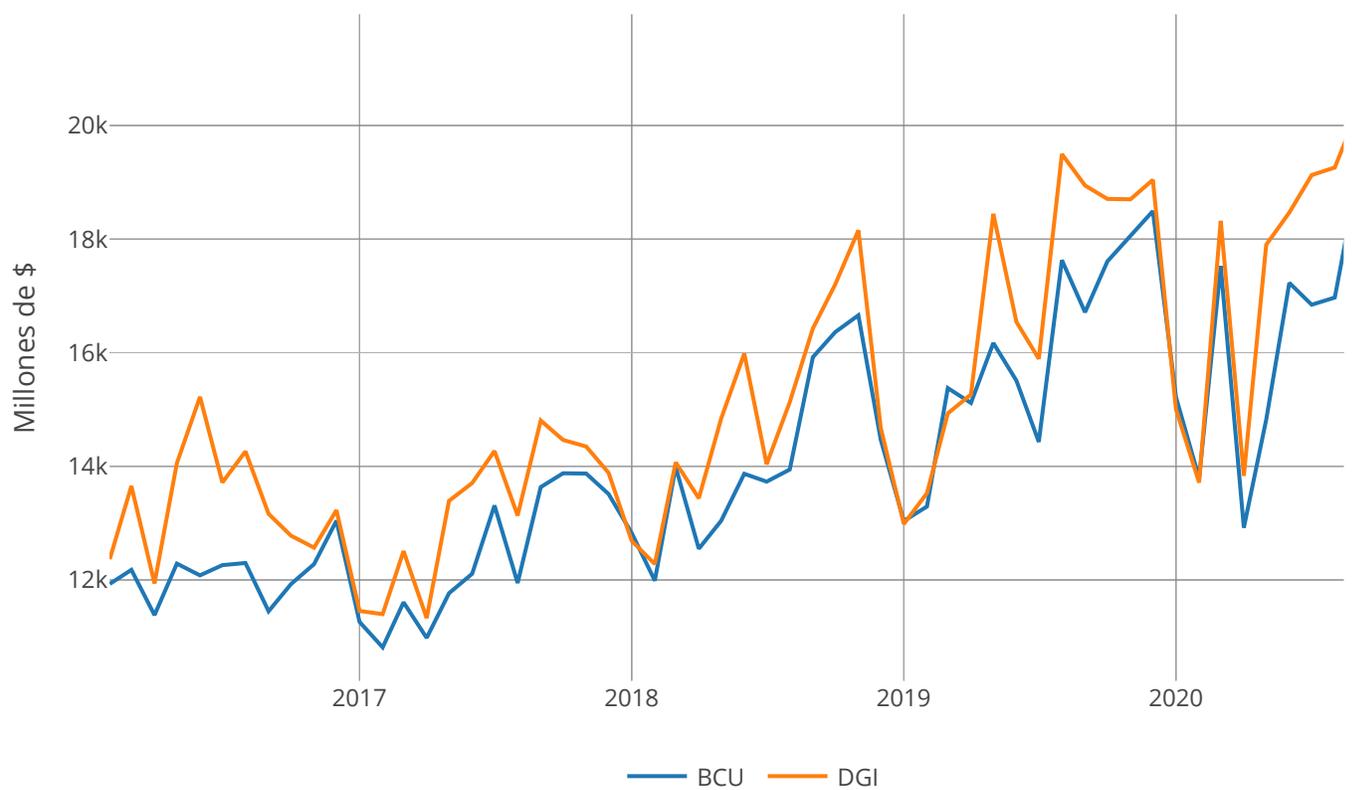


Exportaciones industrial mensual



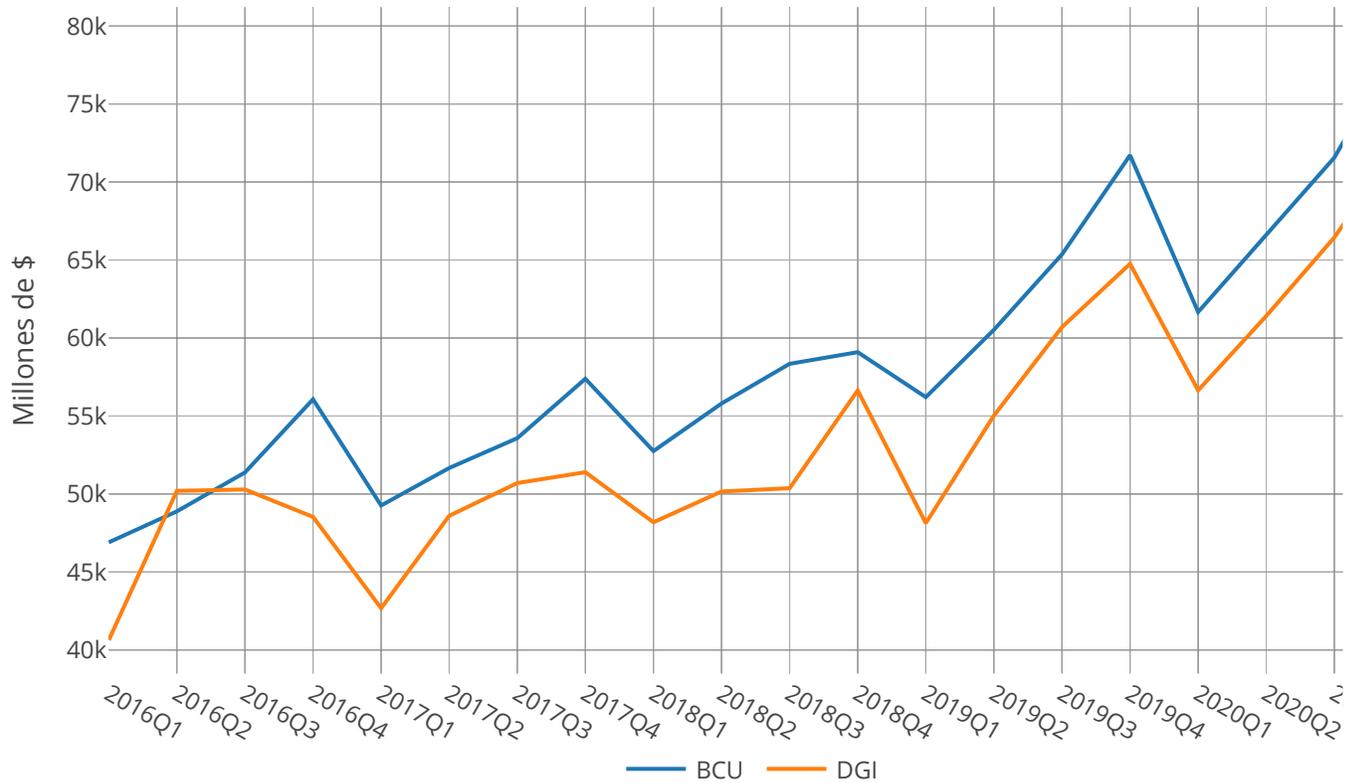
Exportaciones industrial mensual

ajuste



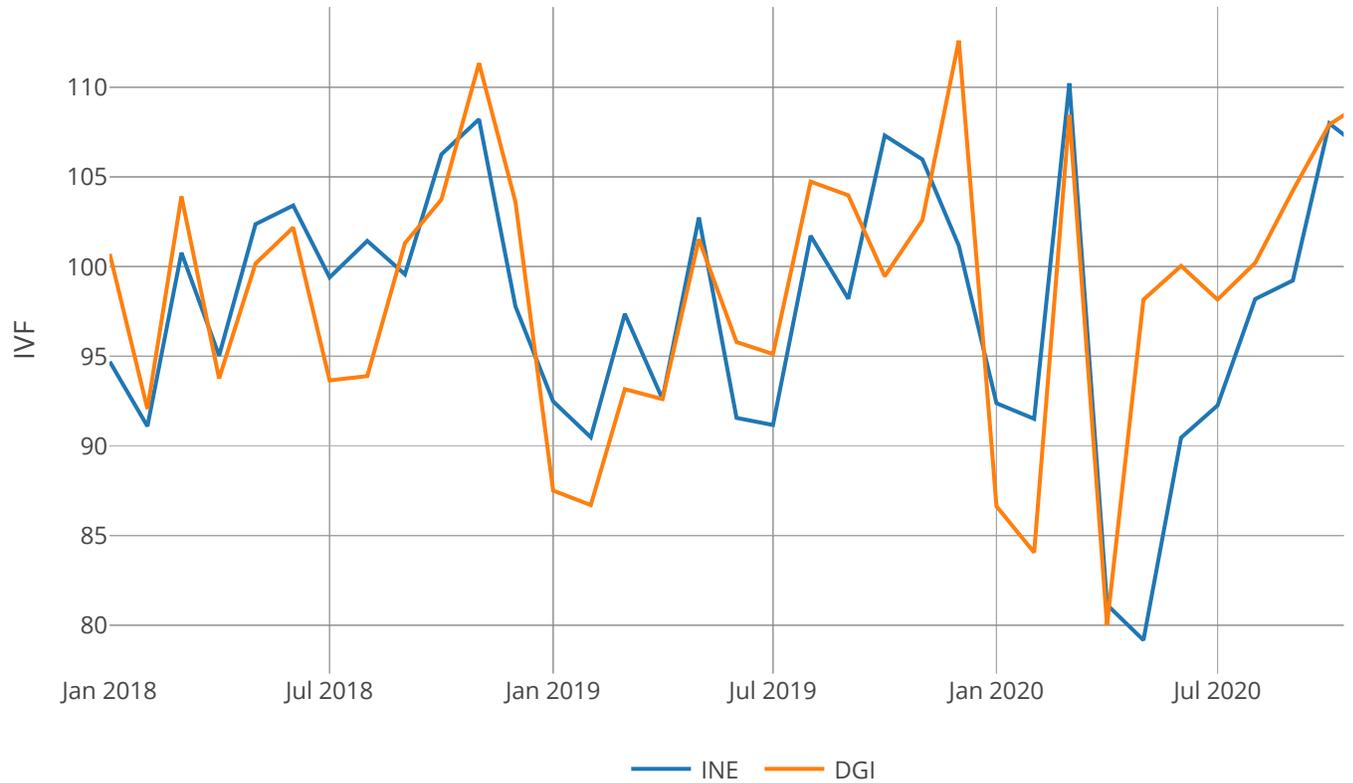
Alimentos Trimestral

BCU



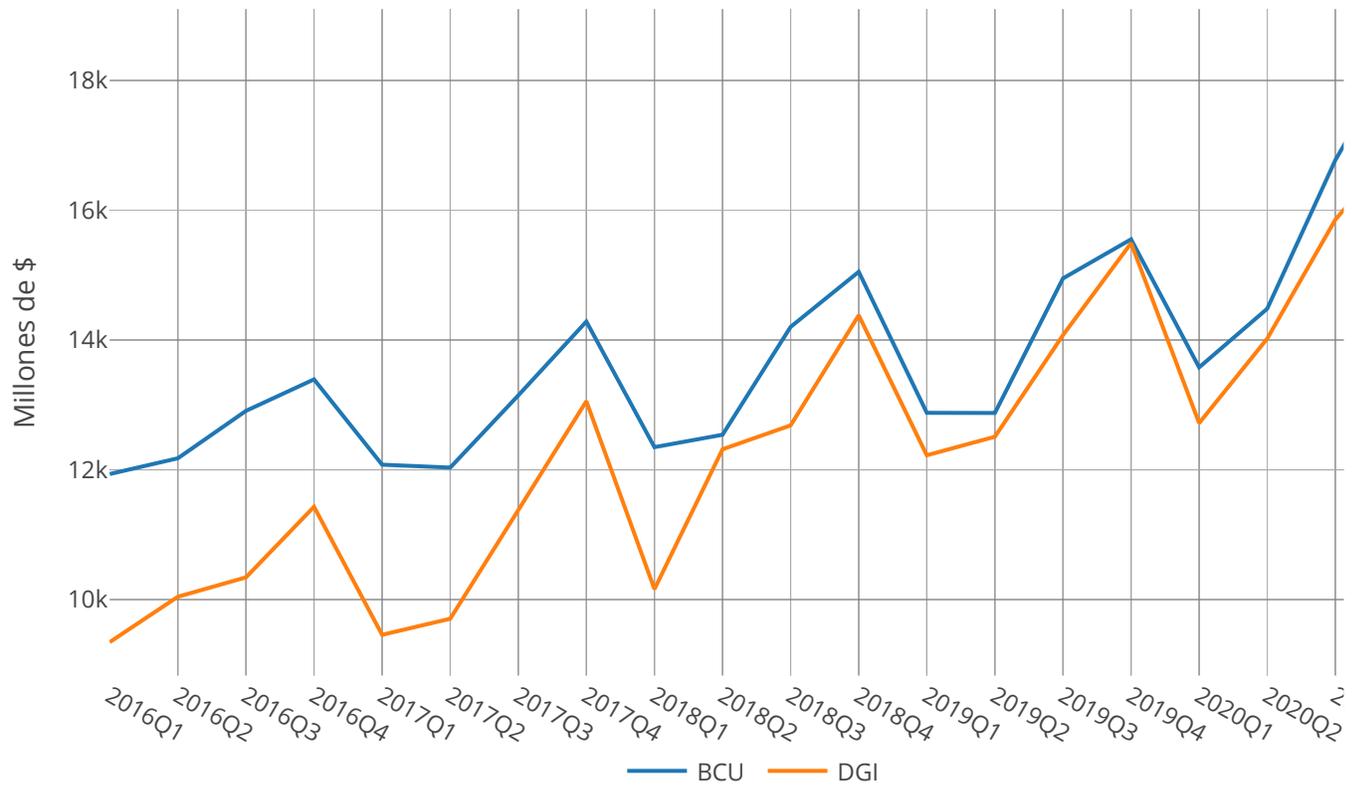
Alimentos mensual

INE



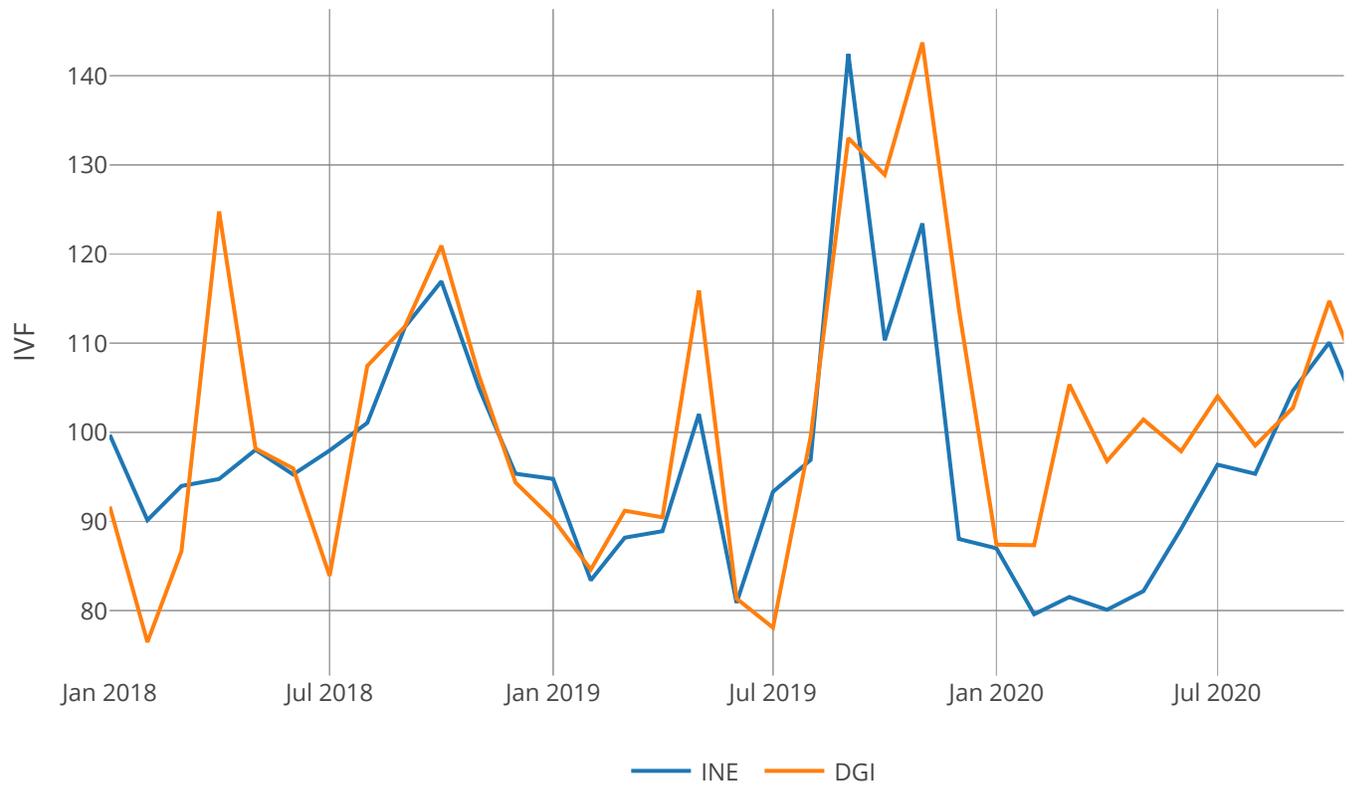
Químicos y otros Trimestral

BCU



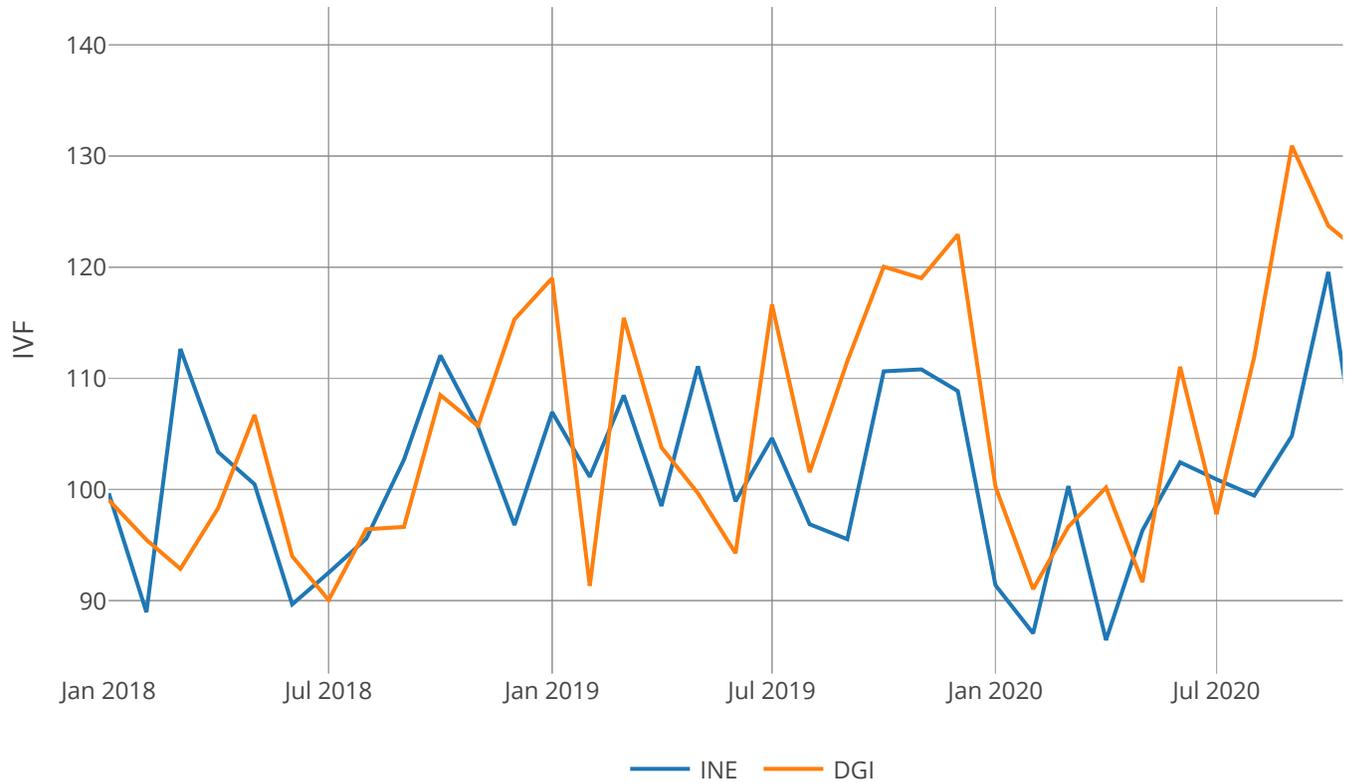
Químicos Mensual

INE



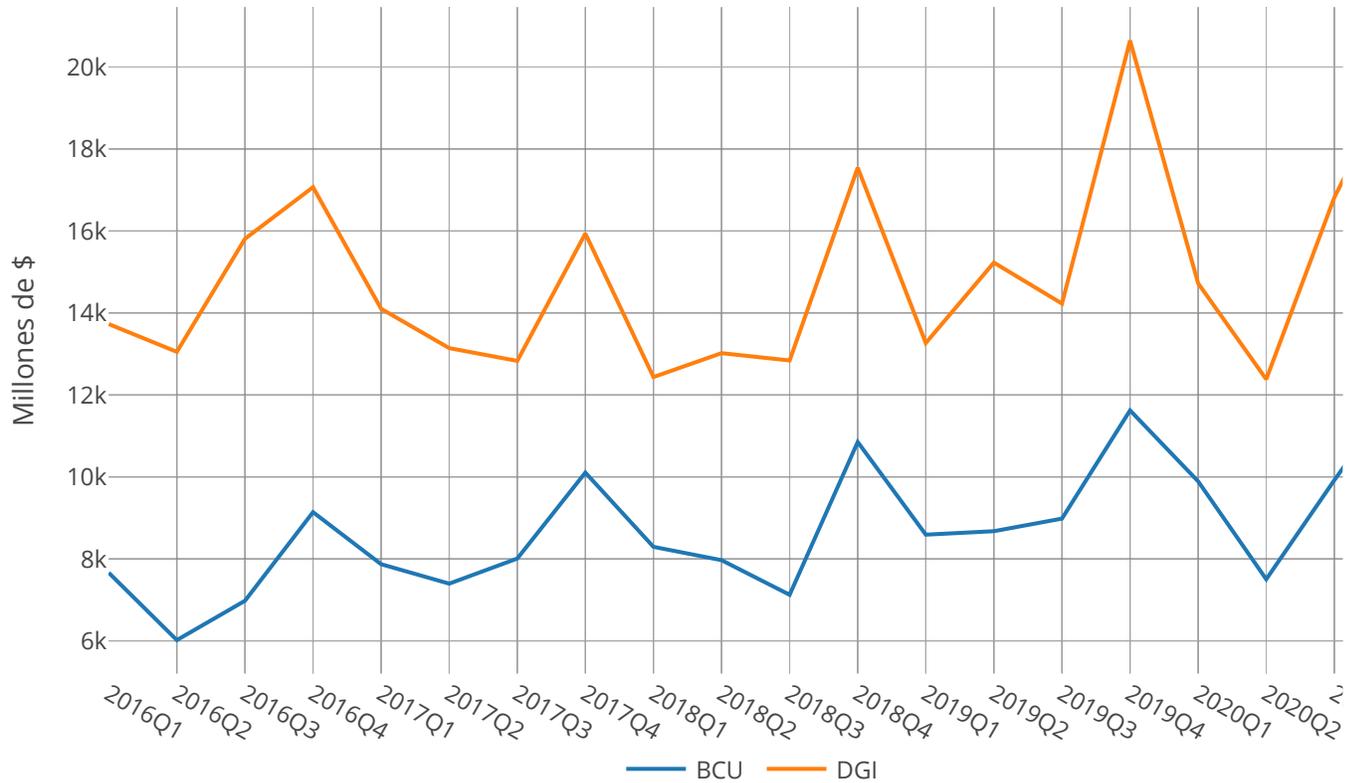
Caucho y plástico Mensual

INE



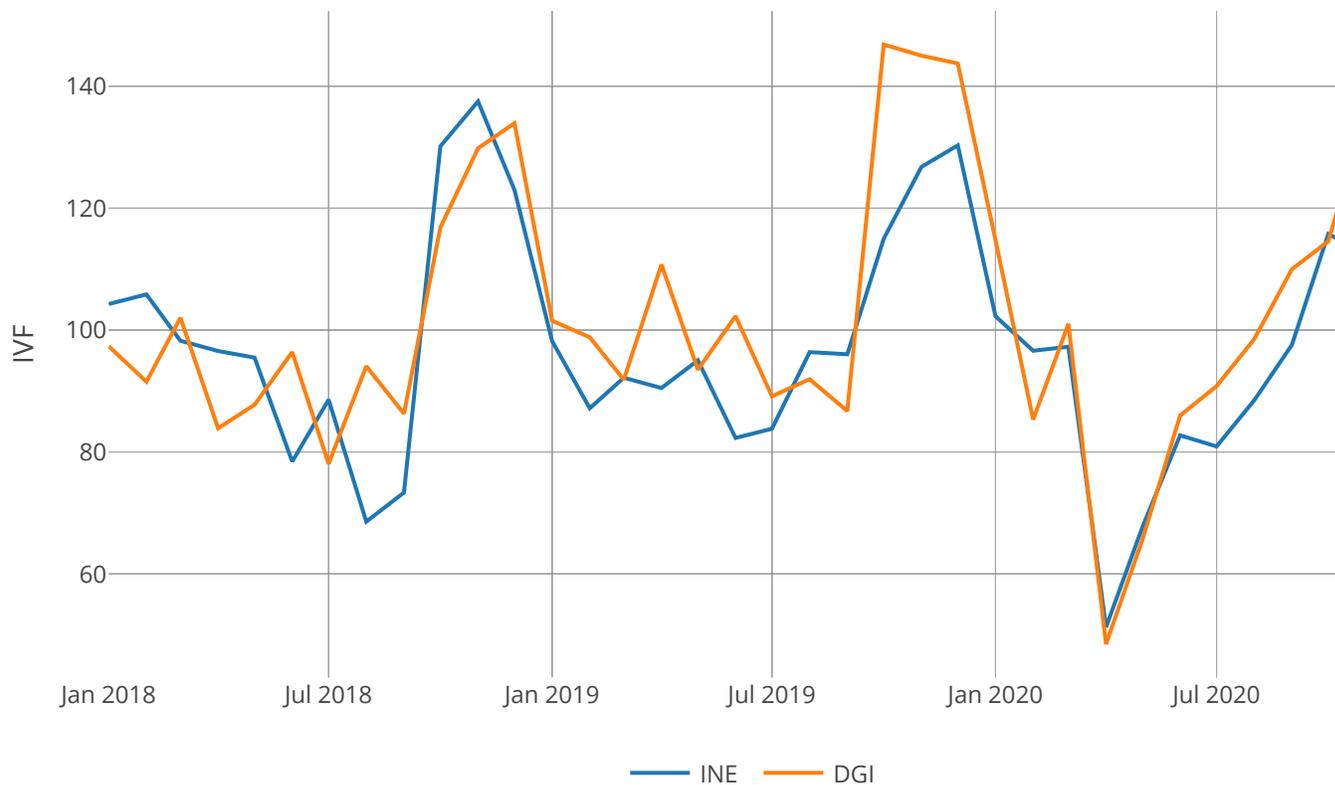
Bebidas y tabaco Trimestral

BCU



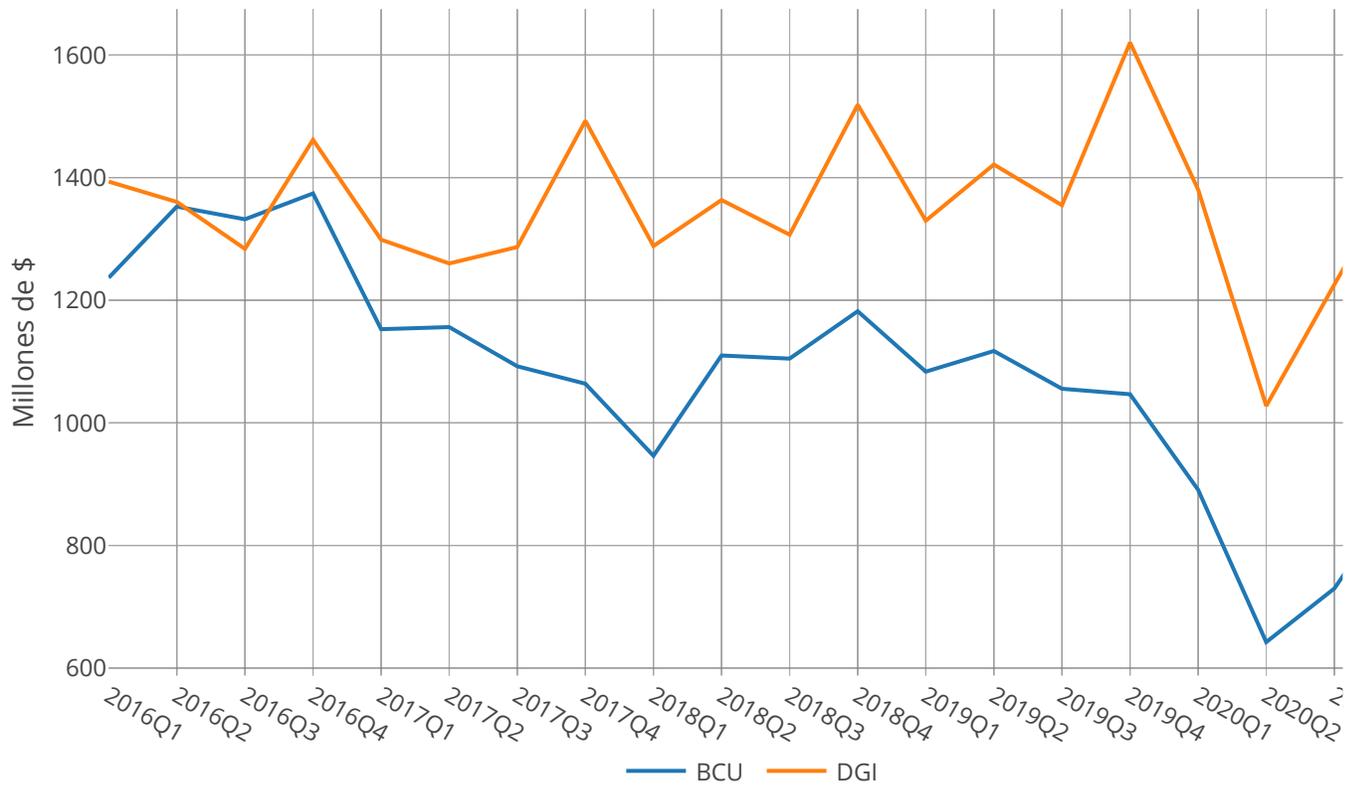
Bebidas Mensual

INE



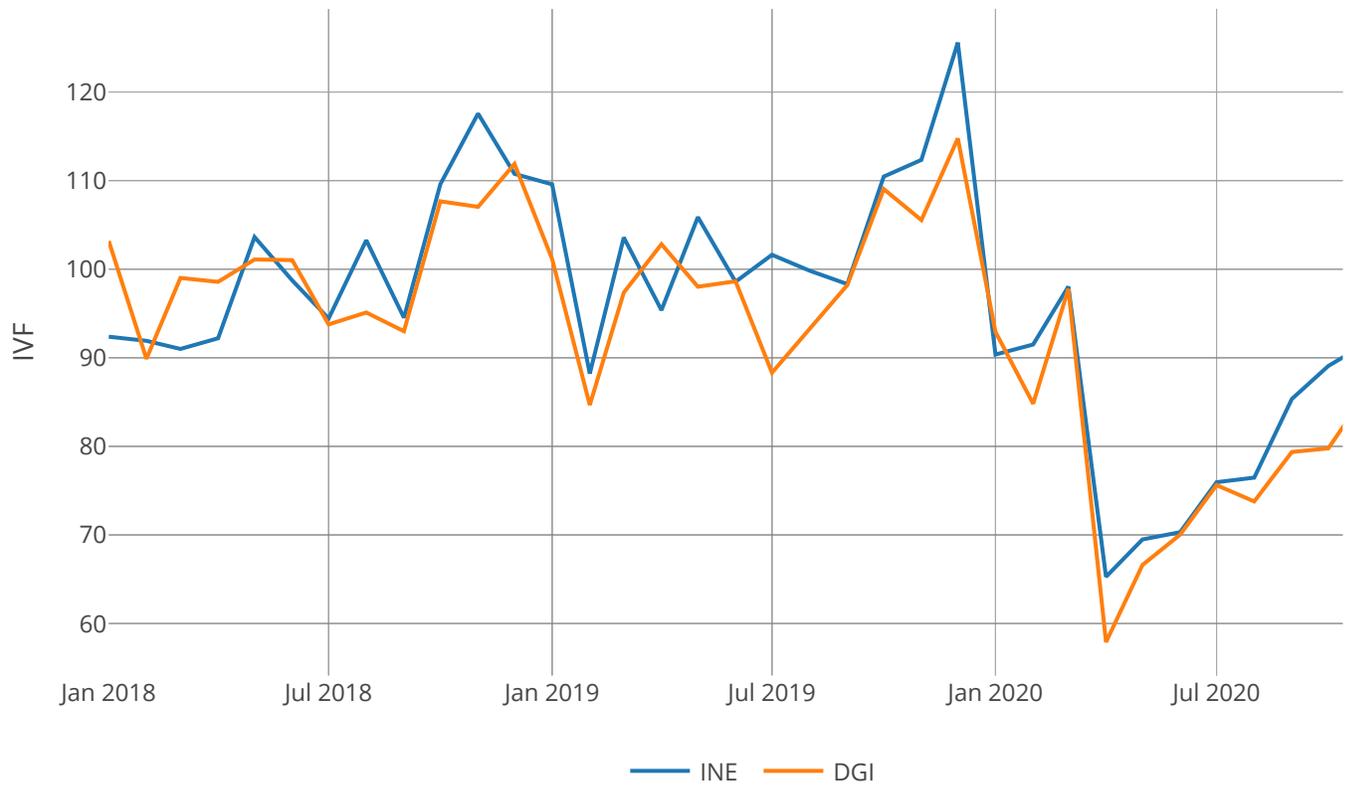
Impresiones Trimestral

BCU

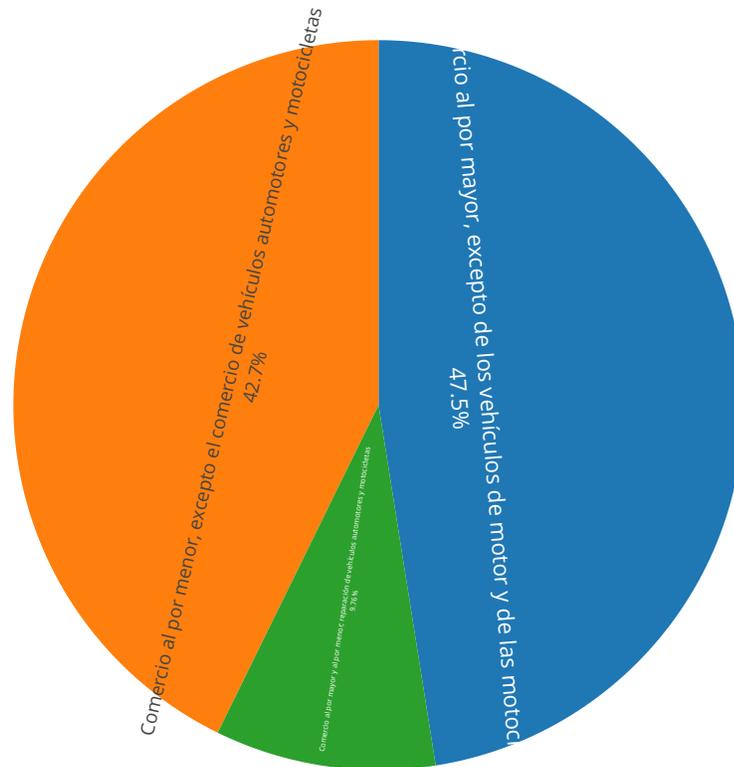


Impresiones Mensual

INE

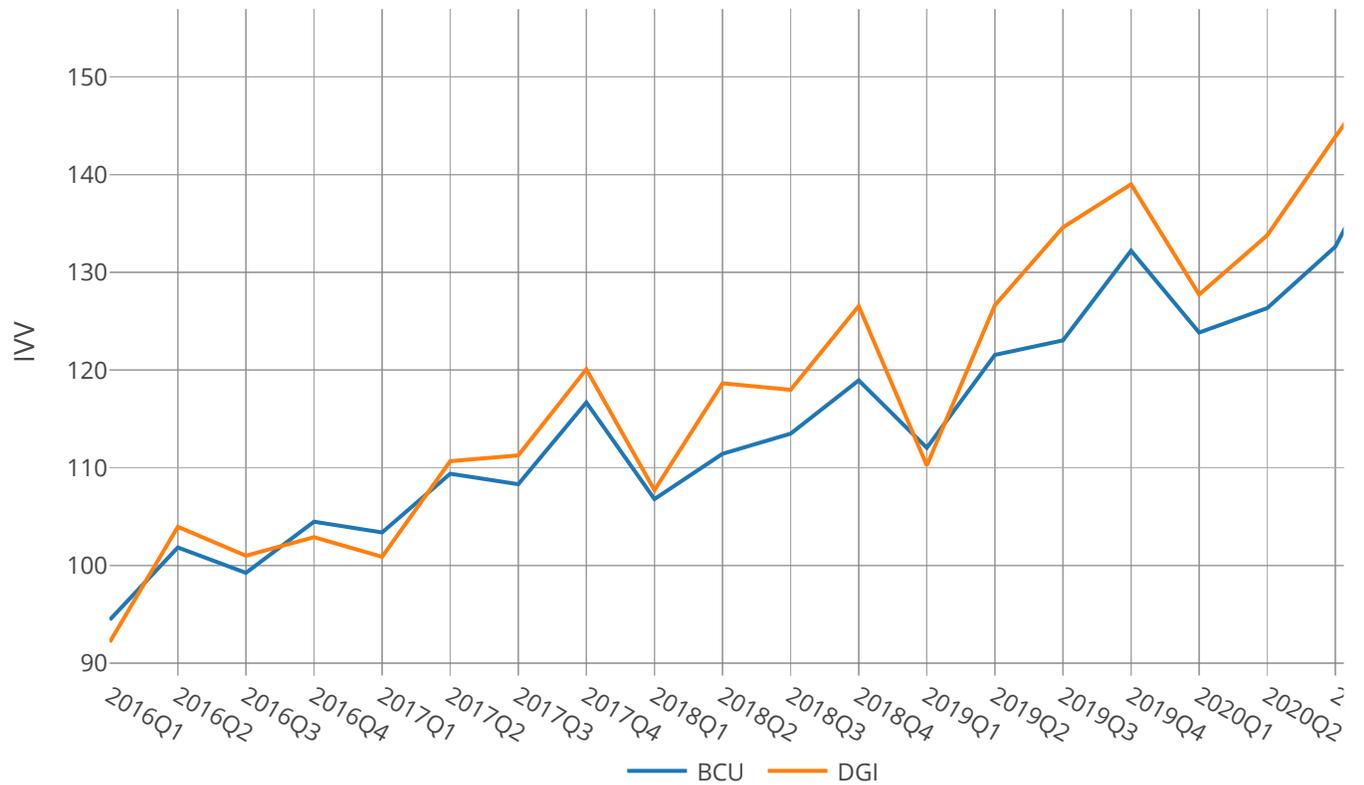


Observaciones del Comercio a 3 dígitos de la CIIU



Comercio Trimestral

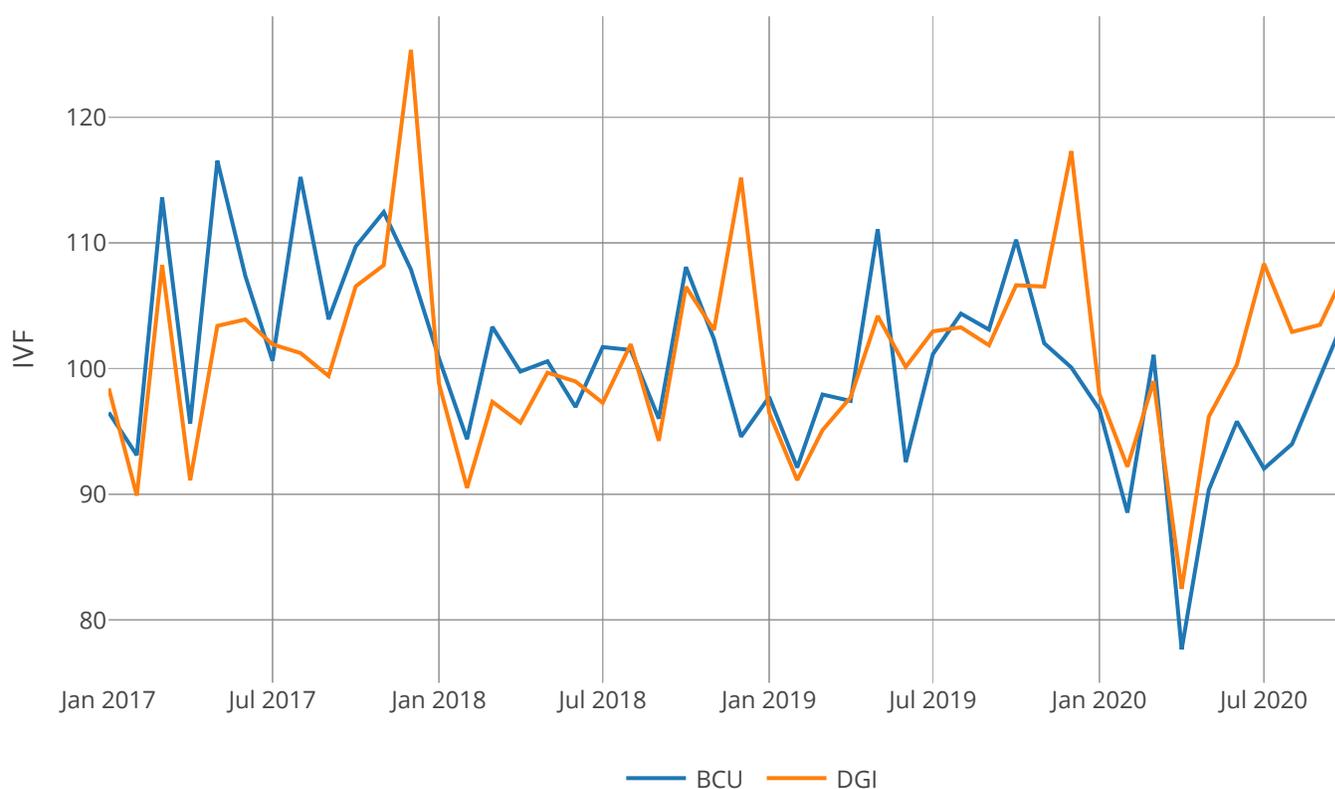
BCU



Márgenes de comercio

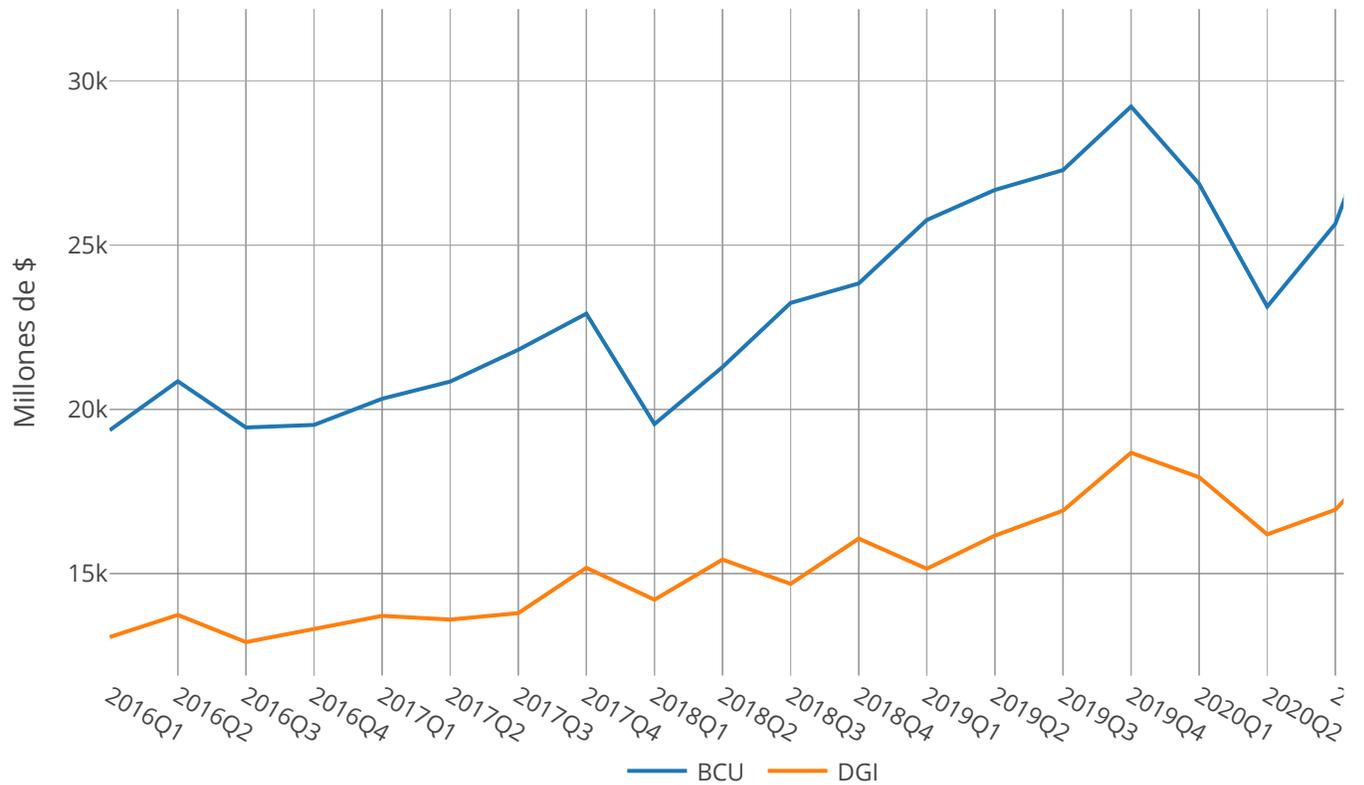
Mensual

BCU (Preliminar)



Exportaciones de servicios

BCU



Desafíos de los datos de DGI

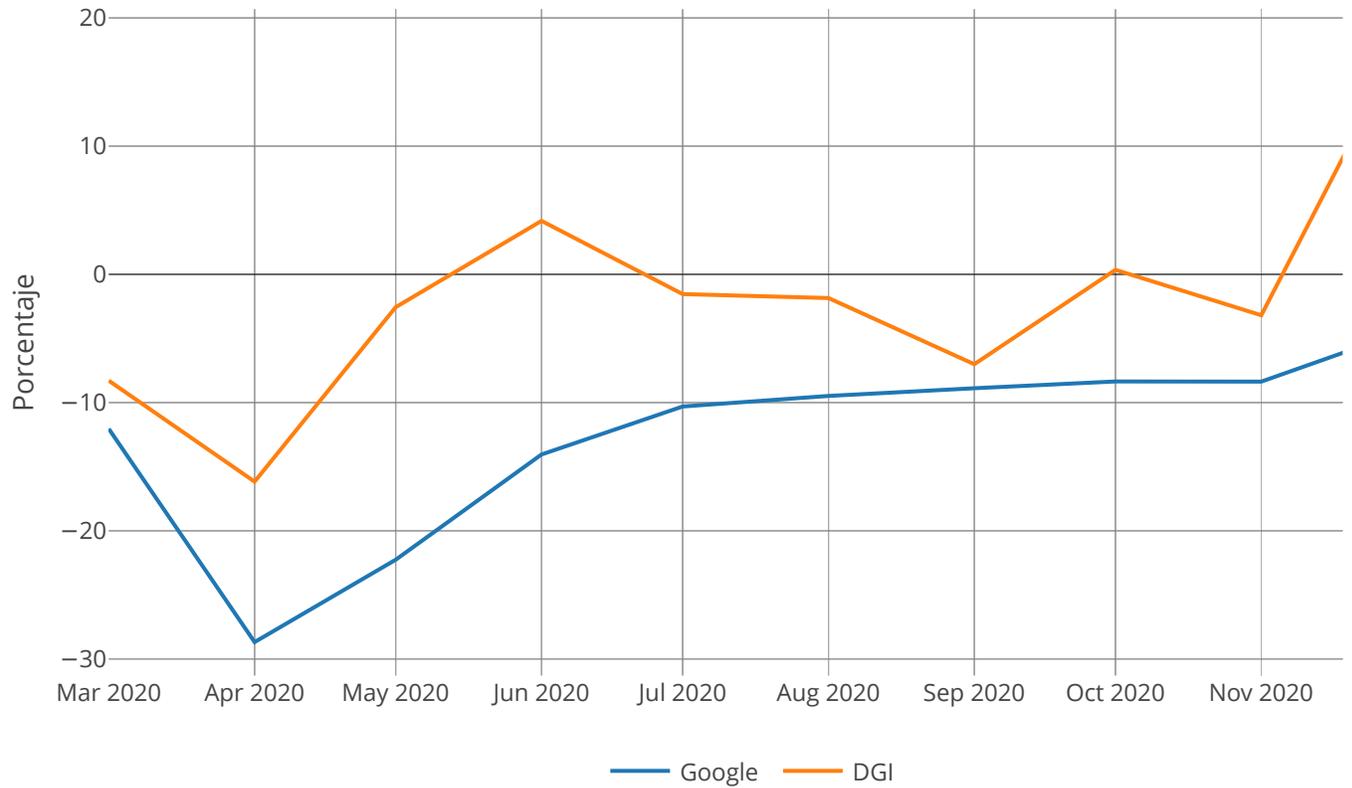
- Momento de la declaración.
- Desfasaje en las exportaciones comparado con las registradas en Aduana.
- Problemas de clasificación.
- Incorporación y tratamiento de empresas que quedaron fuera del panel.
- Incorporación y tratamiento del resto de los formularios.

Reportes de movilidad y Google Trends

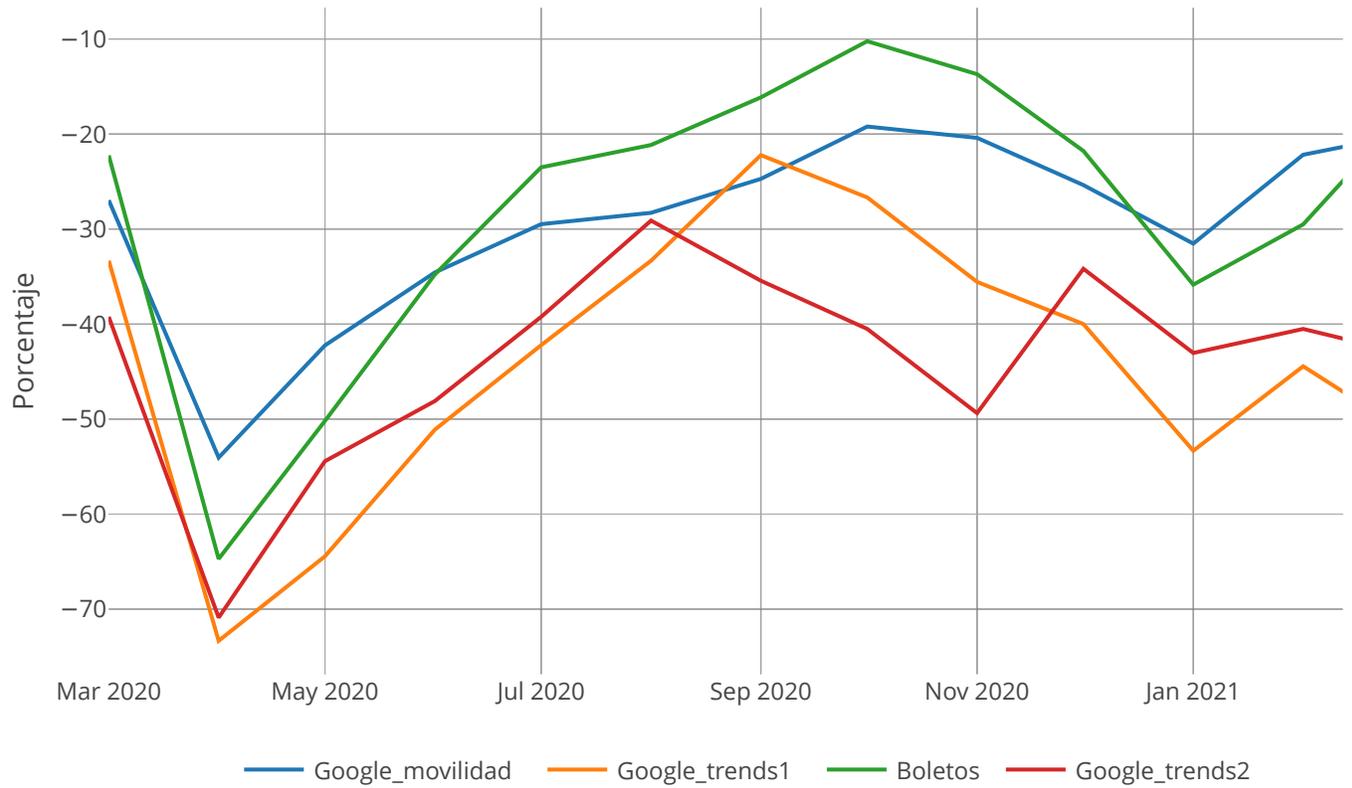
- Se muestran algunos ejemplos de posibles usos de la información de Google Trends y de los reportes de movilidad:
 - Ventas de supermercados y farmacias.
 - Venta de boletos.
 - Turismo.
 - Visitas a parques y otros lugares de recreación.

DGI y reportes de movilidad

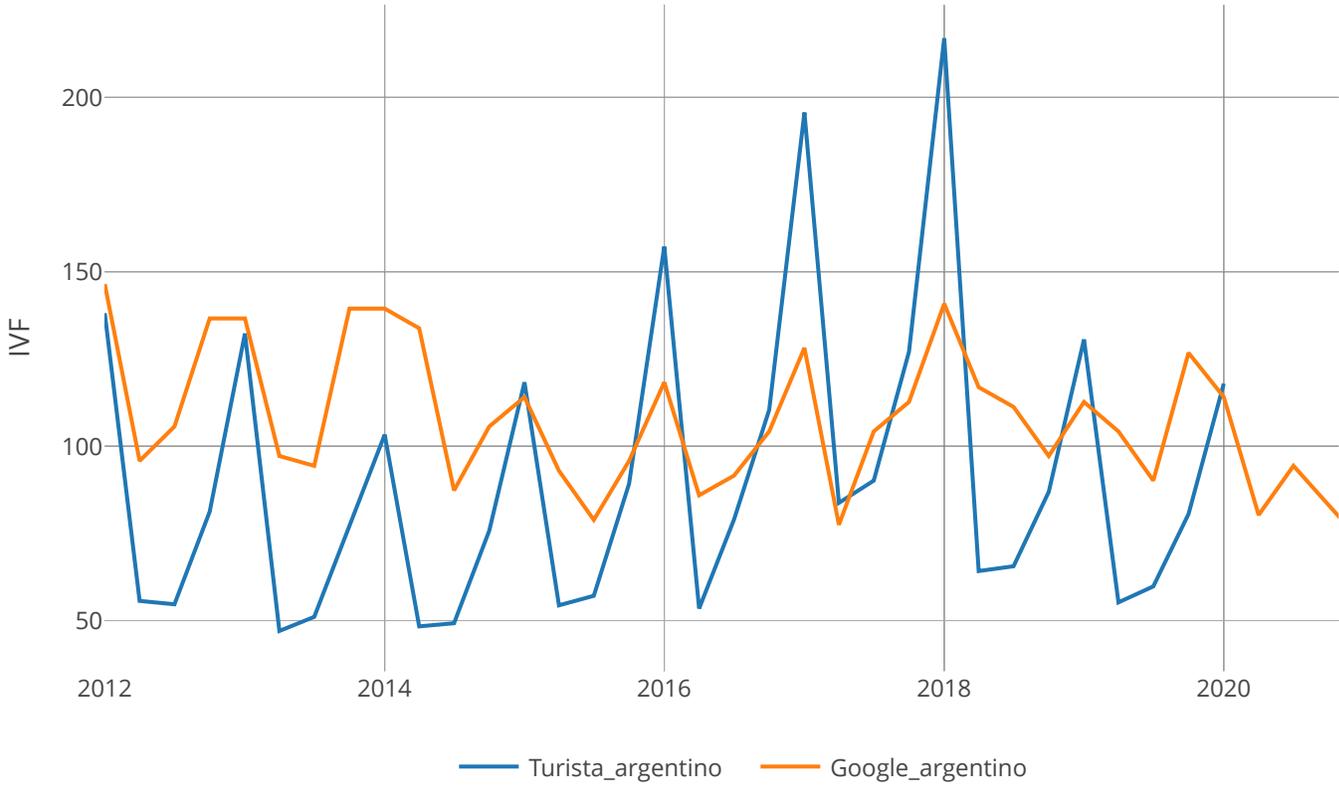
Comercio



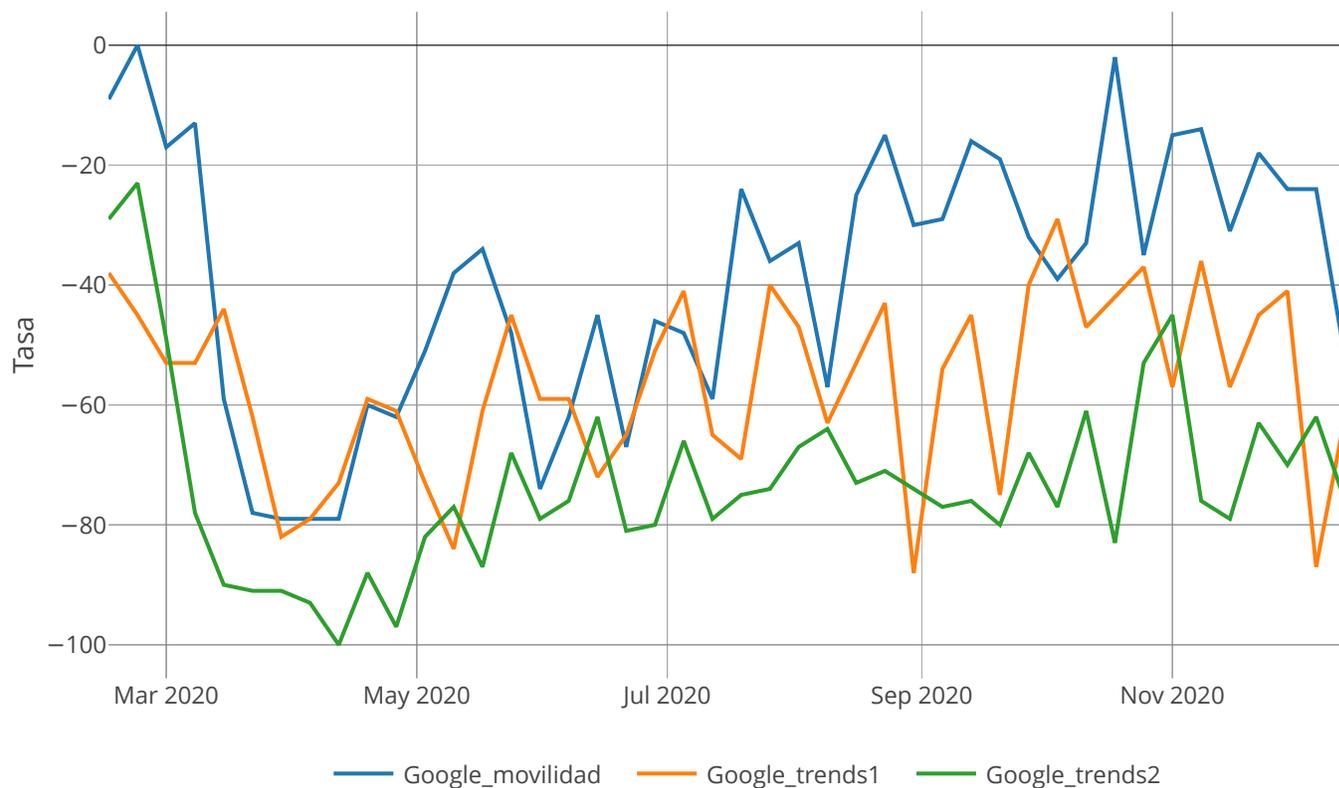
Venta de boletos



Turismo receptivo



Parques y zonas de recreación



Desafíos de los reportes de movilidad y Google Trends

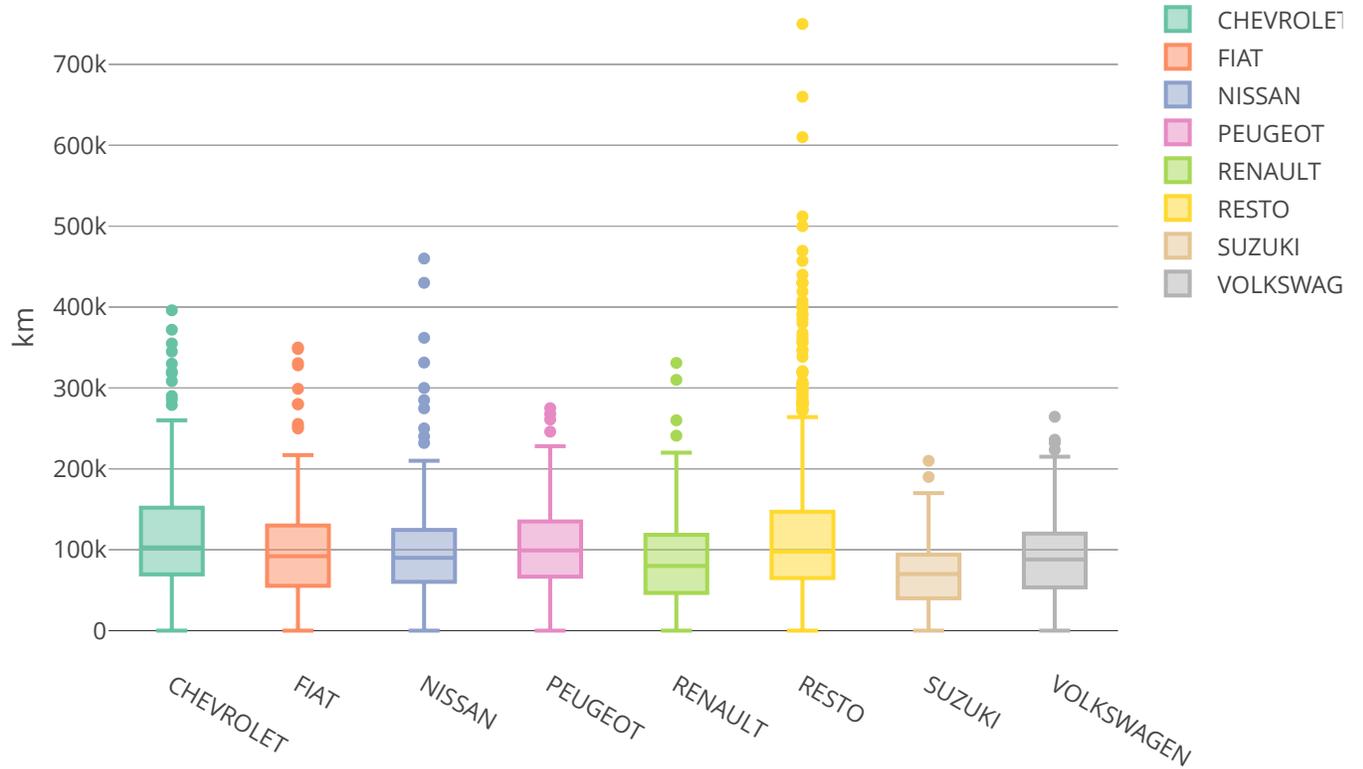
- Funcionan en el muy corto plazo.
- Estrategias para ser usadas.
- Requieren prudencia y evaluación previa.
- Hay que monitorearlos.

Algunos ejemplos de webscrapping

- Mercado libre:
 - Venta de autos (nuevos y usados).
 - Venta de celulares y equipos de computación.
 - Relevamiento de precios.
- Hipermercado Tienda Inglesa:
 - Precios, productos, marcas y unidades.
- Sitios de empleo (Computrabajo, Buscojobs, Gallito.Com)
 - Demanda laboral por rubro.
 - Demanda de las empresas
 - Actividad de agencias de empleo.

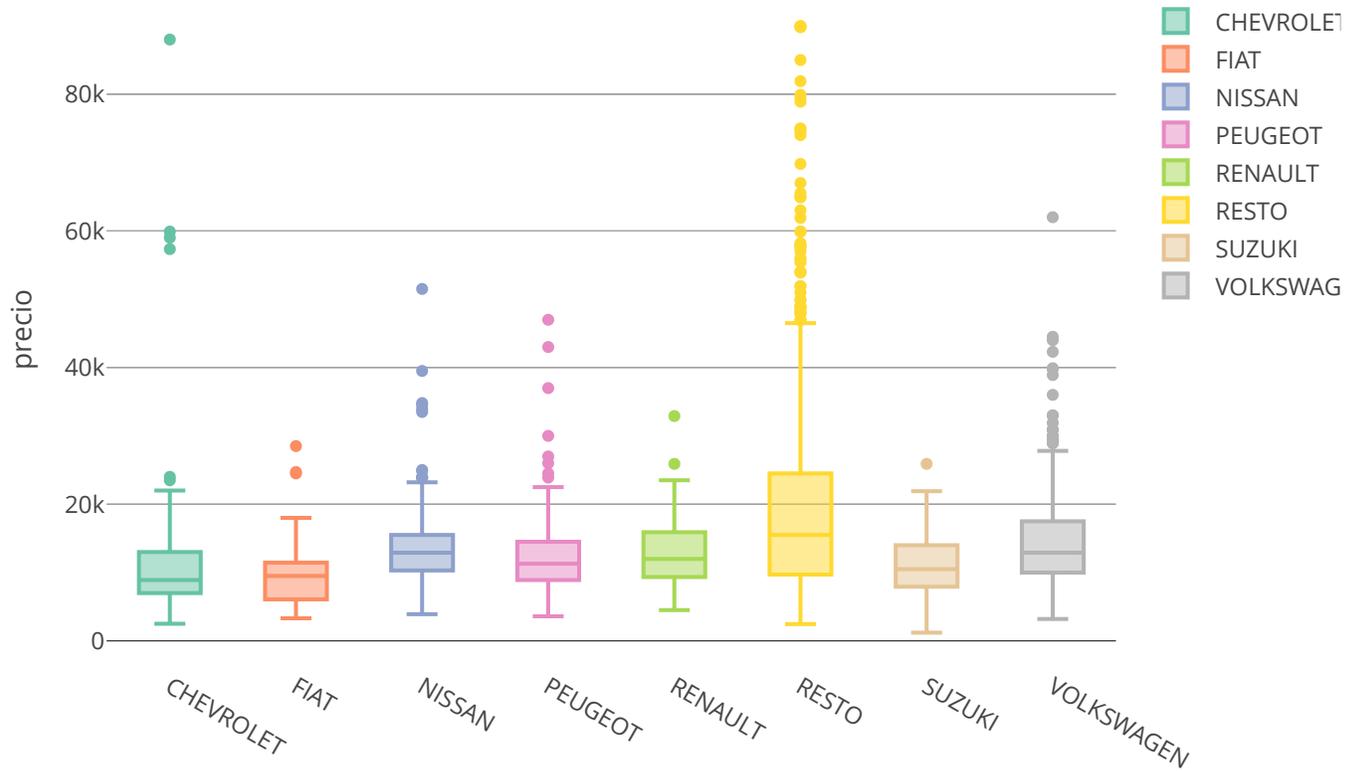
Autos usados

Mercado Libre - kilometros



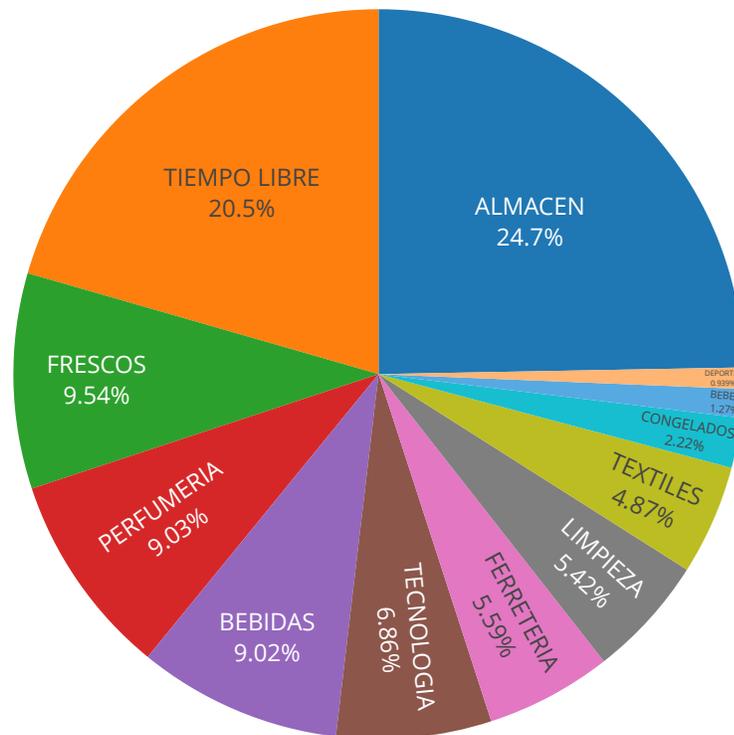
Autos usados

Mercado Libre - Precio



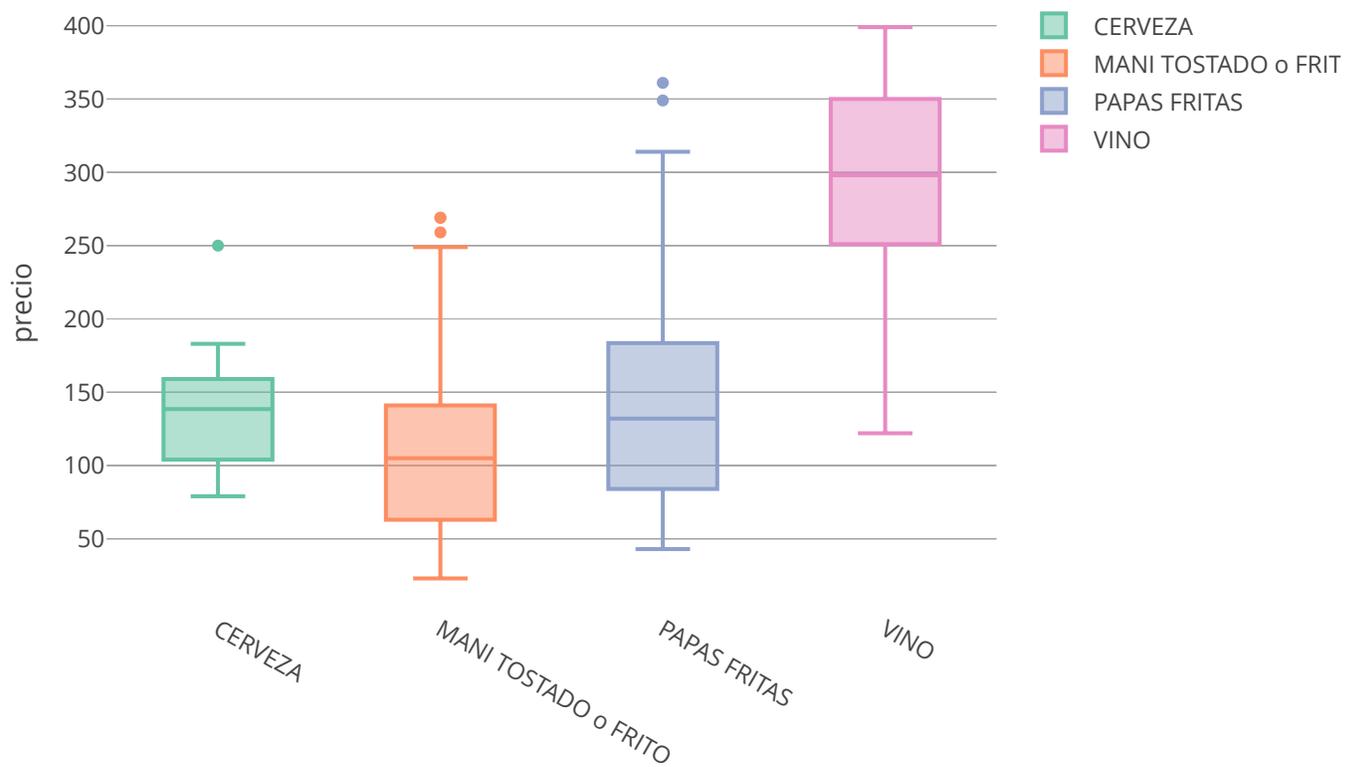
Precios

Hipermercado Tienda Inglesa - Rubros



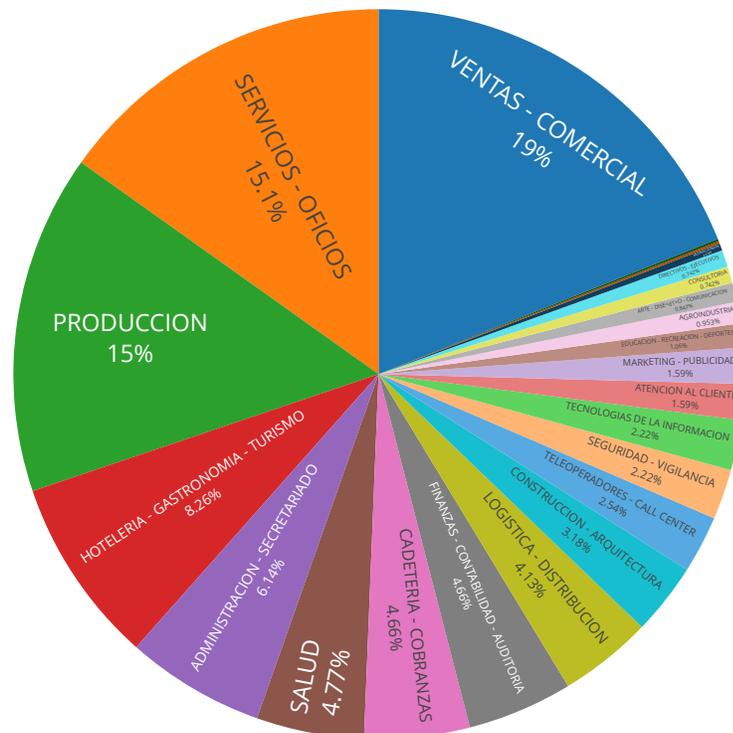
Precios

Hipermercado Tienda Inglesa - Algunos ejemplos



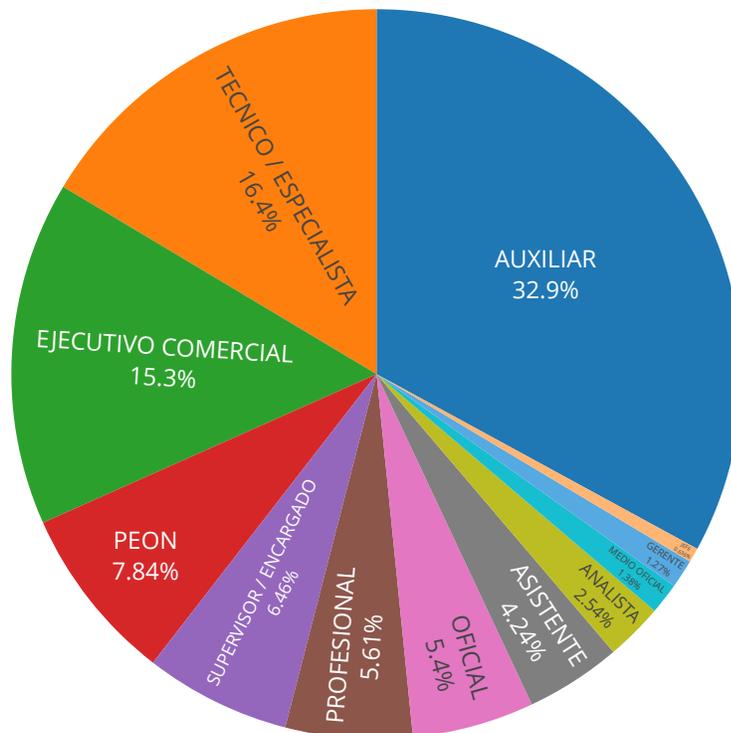
Empleo

Gallito.Com - Rubros



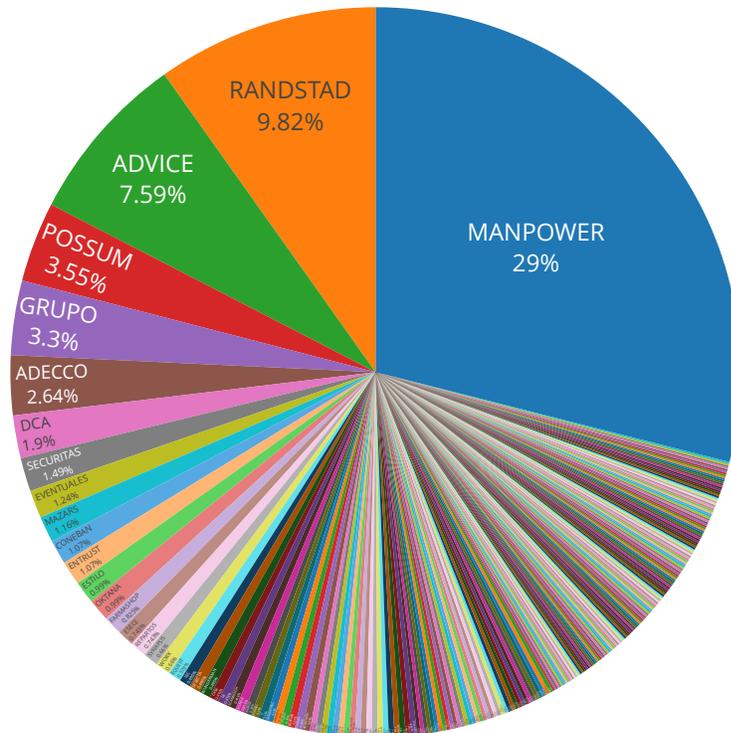
Empleo

Gallito.Com - Puestos



Empleo

Computrabajo - Empresas



Desafíos del webscrapping

- Supuestos para usar la información.
- Problemas de representatividad.
- Programación defensiva.