

Proyecto en R del Programa Ahora 12

Dirección Nacional de Políticas para el Desarrollo del Mercado Interno

En este documento se listan en forma resumida las bases de datos y scripts involucrados en el proyecto. La idea de este proyecto fue hacer mas eficiente todo el procesamiento necesario para elaborar los informes sobre el programa Ahora 12, ganando en tiempo y reduciendo errores de trabajo manual.

Bases de datos

Bases Raw

Estas son bases que se reemplazan o actualizan periodicamente.

- Base Medios de pagos

Frecuencia: Mensual.

Variables: 1.anio 2.mes 3.provincia 4.rubroa12 5.marca_medio_pago 6.monto 7.operaciones 8.cuotas

Nombre del archivo en Cloud, ejemplo: "A12_mensual_20211007.zip"

Archivo del .zip que se usa: "A12_mensual_20211007.csv"

- Base Plataformas de Comercio Electrónico

En esta base encontramos información de plataformas como "Correo Compras - Plataforma" y "Mercado Libre".

Frecuencia: Mensual.

Variables: 1.anio 2.mes 3.plataforma 4.provincia 5.rubroa12 6.operaciones 7.monto 8.cuotas

Nombre del archivo en Cloud, ejemplo: "A12_plataformas_20211004.zip"

Archivo del .zip que se usa: "A12_plataformas_mensual_20211004.csv"

- Base Agrupadores de Pago Digital

En esta base se encuentra información de agrupadores de pago como "Correo Compras - Agrupador", "Mercado Libre", "Mercado Pago", "Mobbex", "Pago Virtual del Sur", "Santander", "Taca Taca", "Todo Pago", "UALA" y "Zoco".

Frecuencia: Mensual.

Variables: 1.anio , 2.mes, 3.agrupador, 4.provincia, 5.describroa12, 6.monto 7.operaciones, 8.cuotas

Nombre del archivo en Cloud, ejemplo: "A12_agrupadores_meli_20211004.zip"

Archivo del .zip que se usa: "A12_agrupadores_meli_mensual_20211004". El archivo contiene "meli" en su nombre, entiendo que haciendo referencia a Mercado Libre pero no es el único agrupador dentro de la base.

- Indice de precios

En este excel se deben ir completando a medida que se publica (o se estima) el resultado mensual del nivel general del Indice de Precios al Consumidor de Indec Nacional. Nombre del archivo: "ipc.xlsx"

Bases Clean

Estas son bases que se generan en scripts a partir de las bases raw y bases que se deben actualizar excepcionalmente.

- base_final_a12.rds

Es resultado de un script y se utiliza como insumo de otros scripts de análisis.

- lista_rubrosa12.rds

Esta base es una lista de rubros que se utiliza en el script “01.Plataformas_ponderaciones.R”. Debe vez en cuando chequearse para ver si están bien esos rubros.

- porblacion.rds

Contiene datos de población. En algún momento puede actualizarse si existen datos nuevos.

- pond_plataformas.rds

Es resultado de un script y se utiliza como insumo de otros scripts de análisis.

Scripts

De acuerdo a lo que se necesite generar se corren ciertos scripts, los que siempre deben correrse con las bases de datos que consume actualizadas son el 01 y 02. Dentro de cada script al inicio quizás haya una línea en el que se deba indicar cual es el último mes con datos actualizados. Los scripts escriben en spreadsheets de googledrive de la cuenta mdointerno@gmail.com, la clave de esta cuenta se pasa por otro medio por seguridad.

01.Plataformas_ponderaciones.R

Lo que se hace es tomar la base de plataformas para obtener las participaciones de gastos y operaciones de los rubros en cada una de las cuotas por período de mercado libre, por ejemplo, dentro de la cuota 12 para un determinado mes, cuanto participa en gasto y operaciones cada uno de los rubros. Otra cuestión a tener en cuenta es que todo lo que hay en la base como “Máquinas y Herramientas” se transforma en “Materiales para la construcción” ya que según se observó los gastos registrados allí corresponden a materiales de construcción en realidad, esto debería ser chequeado por los analistas en algún momento nuevamente. Tener en cuenta que para esto solo se esta considerando mercado pago. El script toma un rds que tiene los rubros de a12, este archivo debe ser chequeado para ver si esta ok. Hay que tener en cuenta también que se filtran solo las cuotas de interés, es decir, hay gastos por fuera de estas cuotas que ya de entrada se eliminan de la base, antes de sacar cualquier resultado. En este script se crean objetos “chequeo” para ver si hay algo raro.

02.Prepara_base.R

En la base total de medios de pagos, hay 2 rubros que se llaman “First Data E-Commerce” y “Otros”, según se observó en su momento estos dos rubros corresponden a las compras de mercado libre, entonces lo que se hace es utilizar las participaciones obtenidas de plataformas para repartir estos gastos por período y por cuota. Para ejemplificar, se toma el gasto y las operaciones en 12 cuotas de estos 2 rubros para un período y se reparte usando el marco de ponderaciones generado en el script “01.Plataformas.R”, puede suceder que en algún período y en alguna cuota en particular exista poca cantidad de operaciones entonces al ser repartida esta cantidad de acuerdo a las ponderaciones, quede alguna con un valor menor a cero, lo que se hace es redondearlas a 1, es por esto que en operaciones quizás no sean iguales los totales al comparar entre ambos resultados, pero si en monto ya que nunca queda nada menor a 1. Terminado este procesamiento se dispone de la base final para realizar los análisis requeridos. Hay que tener en

cuenta también que se filtran solo las cuotas de interés para analizar, es decir, hay gastos por fuera de estas cuotas que ya de entrada se eliminan de la base, antes de sacar cualquier resultado.

03.Procesamiento_mensual.R

Utilizan: la base generada en el script "02.Prepara_base.R", una base que tiene los datos de población y se llama el script auxiliar "Auxiliares/01.Carga_ipc.R" que trae lo necesario para deflactar las series. Se colapsan 3 categorías en 1: "Pequeños electrodomésticos", "Línea Blanca" y "Televisores" en "Electrodomésticos". Este script genera los resultados necesarios para el reporte central de a12.

04.Exporta_mensual.R

Se escribe en un [spreadsheet](#) los resultados generados en el script "03.Procesamiento_mensual.R"

05.Procesamiento_provincias.R

Genera resultados para el informe por provincias que a veces se requiere. Se colapsan 3 categorías en 1: "Pequeños electrodomésticos", "Línea Blanca" y "Televisores" en "Electrodomésticos". Los resultados se exportan en la [spreadsheet](#) en el drive.

06.Procesamiento_rubros.R

Genera resultados para el informe por rubro que a veces se requiere. Se colapsan 3 categorías en 1: "Pequeños electrodomésticos", "Línea Blanca" y "Televisores" en "Electrodomésticos". Los resultados se exportan en la [spreadsheet](#) en el drive.

07.PCE.R

Tener en cuenta que está Correo Compras también. Voy a hacer todo el análisis con los dos adentro (correo compras y mercado libre), lo que hay de correo compras es marginal. Los resultados que generé son análogos al análisis para la base total, sumando 2 hojas que tienen la participación de Correo Compras en el total para aclarar que es muy poco y en que zona se concentra. Hay que tener en cuenta también que se filtran solo las cuotas de interés para analizar, es decir, hay gastos por fuera de estas cuotas que ya de entrada se eliminan de la base, antes de sacar cualquier resultado. Otra cosa que se hace es colapsar 3 categorías en 1: "Pequeños electrodomésticos", "Línea Blanca" y "Televisores" en "Electrodomésticos" y todo lo que hay en la base como "Máquinas y Herramientas" se transforma en "Materiales para la construcción". Una vez obtenidos los resultados se exportan a una [spreadsheet](#) en el drive.

08.APD.R

Hay que tener en cuenta también que se filtran solo las cuotas de interés para analizar, es decir, hay gastos por fuera de estas cuotas que ya de entrada se eliminan de la base, antes de sacar cualquier resultado. Otra cosa que se hace es colapsar 3 categorías en 1: "Pequeños electrodomésticos", "Línea Blanca" y "Televisores" en "Electrodomésticos". Otra cuestión a tener en cuenta es que todo lo que hay en la base como "Máquinas y Herramientas" se transforma en "Materiales para la construcción" ya que según se observó los gastos registrados allí corresponden a materiales de construcción en realidad, esto debería ser chequeado por los analistas en algún momento nuevamente. Una vez obtenidos los resultados se exportan a una hoja de [spreadsheet](#) en el drive.

Auxiliares/01.Carga_ipc.R

Se lee el excel "01.Bases/01.Raw/ipc.xlsx" y se genera el coeficiente deflactor con base en enero del 2019.