**Listado y caracterización de las bases de datos disponibles**

## 

* Base 1. BCRA. Central de Deudores

Se dispone de información consolidada por clave de identificación fiscal (CUIT, CUIL o CDI) respecto de financiaciones otorgadas por entidades financieras, fideicomisos financieros, entidadesno financieras emisoras de tarjetas de crédito / compra, otros proveedores no financieros de créditos, sociedades de garantía recíproca y fondos de garantía de carácter público y, además, cheques rechazados.

## **Datos, definiciones de variables y fuentes**

En el siguiente [enlace](https://drive.google.com/drive/folders/1cktZCEYPdnlsUOoHdgdrEsTUxz5mIdDU?usp=sharing), se encuentran las definiciones de las variables y las bases de datos en formato (stata), las cuales pueden ser leídas en python haciendo uso del paquete pandas ([ver](https://pandas.pydata.org/docs/reference/api/pandas.read_stata.html)).

La información disponible es suministrada por las entidades.<http://www.bcra.gob.ar/BCRAyVos/Situacion_Crediticia.asp>

**Objetivo**

Analizar si existe diferencias entre el nivel de morosidad entre empresas y personas, y distinguiendo a las personas por género. Para ello, se deberá definir como mora cuando la situación de la deuda es 3 o más.

* Base 2. Equipamiento médico

El sector de equipamiento médico comprende productos vinculados al tratamiento médico y cuidado hospitalario. Se pueden establecer cuatro segmentos: Equipo, Implantes, Insumos y Mobiliarios. Se trata de sector que incorpora un nivel importante de tecnología y valor agregado, por lo que es de interés para las agencias de promoción de exportaciones identificar potenciales destinos de exportación a los efectos de implementar estrategias de inserción comercial más focalizadas y diferenciadas, apoyando a las empresas del sector.

Dado que el sector de la salud es de suma relevancia a nivel global, distintos organismos internacionales revelan diversos indicadores que pueden considerarse como *proxies* de demanda de equipamientos e insumos médicos en general.

**Datos, definiciones de variables y fuentes**

Se cuenta con datos a nivel de país de 60 variables agrupadas en distintas categorías:

- Importaciones/exportaciones: construidas a partir de datos comerciales (exportaciones/importaciones) de 24 posiciones arancelarias (códigos de clasificación internacional) correspondientes a productos de equipamientos médicos en el período 2013-2018. Fuente: [Comtrade UN](https://comtrade.un.org/) e [INDEC](https://comex.indec.gob.ar/?_ga=2.209984085.1557565824.1635631854-515828744.1635631854#/)

- Demográficas, Macroeconómicas y Facilidades Comerciales. Fuente: [Banco Mundial](https://datos.bancomundial.org/)

- Salud y gastos relacionados a la salud. Fuente: [Organización Mundial de la Salud](https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates)

Las variables estás dispuestas dos datasets ([> descargar aquí <](https://drive.google.com/drive/folders/1vnueM52BDQQJtFZmpXiTkYtgq4SZ7cZ5?usp=sharing))

· 1-comex\_equip\_medico.csv (Importaciones/exportaciones)

· 2-general\_salud\_equip\_medico.csv (variables demográficas, macroeconómicas, facilidades comerciales, salud y gastos relacionados)

El identificador común de las unidades de análisis en ambos datasets está dado por el [código ISO](https://es.wikipedia.org/wiki/ISO_3166-1) de cada país.

## **Objetivo**

Realizar una segmentación de mercados que permita agrupar a los distintos países como destinos potenciales de las exportaciones argentinas de equipamientos médicos y realizar una caracterización de los mismos. Como punto adicional (optativo) se pretende establecer un ranking de esos potenciales destinos generando un indicador sintético que contemple las variables seleccionadas.

* Base 3. Precio de venta locales comerciales

Se pretende predecir el precio de venta de locales comerciales en la Ciudad Autónoma de Buenos Aires (CABA) a partir características propias de estos inmuebles (metros cuadrados, cantidad de ambientes, antiguedad, etc) como así también una serie de variables contextuales (relacionadas con su ubicación) construidas a partir de información georreferenciada (parques, atracciones turísticas, sitios gastronómicos, seguridad, etc.).

**Datos, definiciones de variables y fuentes**

En todos los casos, los datos georreferenciados fueron obtenidos del portal de datos abiertos de la Ciudad Autónoma de Buenos Aires ([BA data](https://data.buenosaires.gob.ar/)). En todos los casos se encuentra en formato [shapefile](https://es.wikipedia.org/wiki/Shapefile) (que permite almacenar ubicación geométrica y la información de atributos de las entidades geográficas) los cuales fueron utilizados para construir las variables contextuales y de control.

Puede descargar el dataset completo haciendo [click aquí](https://drive.google.com/drive/folders/1kD5pj_apXIF6p6RN0gzI_1_8et8nYbL-?usp=sharing)

#### **Características de los locales**

Fuente:<https://data.buenosaires.gob.ar/dataset/locales-en-venta> (correspondientes al año 2018)

Variables:

· antig: antigüedad del local en años

· m2total: total de metros cuadrados

· m2cub: total de metros cuadrados cubiertos

· ambientes: número de ambientes

· banios: cantidad de baños

· precioUSD: precio en dólares

· m2precioUSD: precio en dólares por metro cuadrado

· Barrio

· Comuna

· latitud/longitud: coordenadas geográficas proyectadas

#### 

#### **Características contextuales**

Se consideran los siguientes datos georreferenciados a los efectos de construir una serie de variables contextuales que pueden influir en el precio de los locales comerciales:

· Gastronomía: Listado de restaurantes y establecimientos gastronómicos en la ciudad. (<https://data.buenosaires.gob.ar/dataset/oferta-establecimientos-gastronomicos>)

· Atracciones Turísticas: Listado de espacios culturales públicos, privados e independientes localizados en la Ciudad de Buenos Aires. Incluye bares, bibliotecas, calesitas, centros culturales, disquerías, espacios de exhibición, espacios de formación, espacios escénicos, espacios feriales, librerías, monumentos y lugares históricos y salas de cine. Se consideran exclusivamente espacios escénicos (por ejemplo, salas de teatros), monumentos y lugares históricos. (<https://data.buenosaires.gob.ar/dataset/espacios-culturales>)

· Parques: Límites y ubicación geográfica de los espacios verdes de la Ciudad (jardín, parque, patio recreativo, plaza, plazoleta, cantero y polideportivo). (<https://data.buenosaires.gob.ar/dataset/espacios-verdes>)

· Distrito Central Financiero-Comercial: Ubicación geográfica del Obelisco, principal monumento de la Ciudad de Buenos Aires.

· Comisarías de la ciudad: Ubicación geográfica de las comisarías de la policía de la ciudad. (<https://data.buenosaires.gob.ar/dataset/comisarias-policia-ciudad>)

· Delitos: Datos georreferenciados de distintos delitos en la Ciudad de Buenos Aires. Se considerarán exclusivamente robos (con violencia) en el cuarto trimestre del año 2017. (<https://data.buenosaires.gob.ar/dataset/delitos>)

**Variables**

· comisaria\_dista: distancia (metros) a la comisaría más cercana.

· obelisco\_dista: distancia al Obelisco (proxy de centro financiero)

· nrobos: cantidad de robos en una zona de influencia (*buffer*) de 500 metros.

· sup\_espacio\_verde: superficie (en metros cuadrados) de espacios verdes en un en una zona de influencia (*buffer*) de 1000 metros.

· count\_gastronomia: cantidad de sitios gastronómicos en una zona de influencia (*buffer*) de 500 metros.

· count\_culturales: cantidad de atractivos turísticos (espacios escénicos, monumentos y lugares históricos) en una zona de influencia (*buffer*) de 500 metros.

**Objetivo**

Contar con un modelo adecuado que permita predecir el precio de venta de los locales comerciales de CABA atendiendo a las características propias de cada local así como a las características contextuales. Se requiere que el modelo sea lo suficientemente interpretable como para indagar sobre el impacto de las variables consideradas.

* Base 4. PISA 2018

El Programa de la OCDE para la Evaluación Internacional de Estudiantes (PISA) examina lo que los estudiantes saben en lectura, matemáticas y ciencias, y qué pueden hacer con lo que saben. Proporciona la evaluación internacional más completa y rigurosa de los resultados del aprendizaje de los estudiantes hasta la fecha. Los resultados de PISA indican la calidad y equidad de los resultados del aprendizaje obtenidos en todo el mundo y permiten a los educadores y a los responsables políticos aprender de las políticas y prácticas aplicadas en otros países.

**Datos, definiciones de variables y fuentes**

En el siguiente link está una descripción más detallada de las variables y sus definiciones (<https://www.oecd.org/pisa/publications/pisa-2018-results.htm>).

Los cuestionarios y las bases de datos en formato (stata), las cuales pueden ser leídas en python haciendo uso del paquete pandas ([ver](https://pandas.pydata.org/docs/reference/api/pandas.read_stata.html)), se encuentran en el siguiente [enlace](https://drive.google.com/drive/folders/147Zi2t-Z3qwdcD_mDFGXobEVEp_Yppvd?usp=sharing).

**Objetivo**

Analizar el impacto del uso intensivo de las tecnologías en los resultados escolares.

* Base 5. Encuesta de Protección Social 2015

La Encuesta de Protección Social (EPS) es la mayor y más antigua encuesta longitudinal de tipo panel que existe en Chile, con una muestra de alrededor de 16.000 encuestados distribuidos en todas las regiones del país.

Abarca en un mismo cuestionario la historia laboral y previsional de los encuestados con información detallada en áreas como educación, salud, seguridad social, capacitación laboral, patrimonio y activos, historia familiar e información sobre el hogar.

Los resultados de esta encuesta se utilizan en el ámbito académico y en el análisis y diseño de políticas públicas.

## **Datos, definiciones de variables y fuentes**

Para mayor detalle de la encuesta, ver el siguiente [enlace](https://www.previsionsocial.gob.cl/sps/biblioteca/encuesta-de-proteccion-social/).

En el siguiente [enlace](https://drive.google.com/drive/folders/1Qy04NMoDhgxSSe--1t3JrnWovWaFLsmH?usp=sharing), se encuentran las definiciones de las variables y las bases de datos en formato (stata), las cuales pueden ser leídas en python haciendo uso del paquete pandas ([ver](https://pandas.pydata.org/docs/reference/api/pandas.read_stata.html)).

## **Objetivo**

Estudiar cómo se explican los salarios en función de distintas variables, entre las cuales se debe incluir los años de educación formal.