## **Métodos Cuantitativos: Trabajo Final**

El trabajo final consiste en aplicar los conocimientos aprendidos durante la cursada. Para ello, se realizará un análisis estadístico sobre un conjunto de datos (*dataset*) apuntando a responder una/s pregunta/s sugeridas por el alumno.

El **entregable** es un documento de texto (en formato ***.pdf***) de **no más de 3 páginas** en donde se presentarán los aspectos más importantes del *dataset,* la/s interrogante/s a contestar, los resultados del análisis y las conclusiones. **No** se deberá entregar el **código** utilizado, **ni** los datos crudos (planillas de ***excels***). No se evaluará la extensión, si no la consistencia del contenido.

A la hora de la elaboración del documento, se propone pensarlo como que será leído por personal no técnico ni especializado. En ese marco, se alienta a explicar e interpretar los resultados, aclarando los gráficos y métricas obtenidas, más que simplemente listar un conjunto de números.

El trabajo debe hacerse en grupo de **mínimo 2** personas, **hasta 5 integrantes**.

## **Dataset**

Las bases de datos a trabajar son aquellas que compilan datos públicos publicados por organismos de gobierno de la República Argentina. Las mismas son de libre acceso y se pueden explorar entrando a <https://datos.gob.ar/dataset> [[1]](#footnote-1)

Se propone que el alumno **busque** algún ***dataset*** (en formato .csv, ver aquellos con el ícono amarillo) de **gusto personal**. A la izquierda de la web, se encuentra un listado ordenado por fuentes de orígenes (Ministerios). Al clickear en el título y luego en “CONSULTAR”, aparecerá una breve descripción de cada una de las columnas (variables).



Ejemplo del *dataset* [Operaciones de viaje por Periodo - Modo - Jurisdicción - Grupo Tarifario - Empresa - Línea - Tipo de Pasaje](https://datos.gob.ar/dataset/transporte-sube-operaciones-viaje-por-mes-region-metropolitana-buenos-aires/archivo/transporte_03b265af-e60f-4190-938e-171501fee65a) con su descripción de variables.

Se pide que el *dataset* elegido tenga **al menos *100* filas**. Pese a que se recomienda fervorosamente buscar una base a gusto, de dejan a continuación algunas bases posibles:

* [SUBE - Cantidad de tarjetas (usuarios) por día en AMBA](https://datos.gob.ar/dataset/transporte-sube---cantidad-tarjetas-usuarios-por-dia-amba/archivo/transporte_34cd80ba-4f91-418f-8edd-6dda2ea43d34)
* [Encuesta Nacional de Consumos Culturales (2017)](https://datos.gob.ar/dataset/cultura-encuesta-nacional-consumos-culturales-2017) (prestar atención al cuestionario).
* [Empleo y Salario del Sector Minero (SIACAM)](https://datos.gob.ar/dataset/produccion-empleo-salario-sector-minero-siacam/archivo/produccion_62ac5b60-c8e7-411b-889e-45c57137054a)
* [Puntos WiFi País Digital](https://datos.gob.ar/dataset/jgm-puntos-wifi-pais-digital)
* [Viajeros hospedados en hoteles y parahoteles por región de destino según tipo de residencia](https://datos.gob.ar/dataset/turismo-encuesta-ocupacion-hotelera-parahotelera-eoh/archivo/turismo_1f6b78aa-d3b4-440a-bd7d-30d76e1728aa)

Sobre la base seleccionada, se **propone generar una/s pregunta/s a contestar** utilizando un conjunto de variables (no tienen que ser todas). Muchas bases consisten en largos períodos de tiempo o una desagregación extensiva por región. Para estos casos se propone quedarse con corto período de tiempo y/o un conjunto pequeño de regiones/provincias.

Se sugiere preguntas sencillas, las cuales funcionen como motivación para la aplicación de las herramientas dadas. El espíritu de este trabajo es poner en práctica la **utilización** de **algunas** **herramientas** con datos reales, teniendo siempre en mente la pregunta inicial como motor.

## **Guía para el documento entregable**

Se lista un conjunto de ítems a tener en cuenta a la hora de confeccionar el documento final. Se sugiere tomarlos simplemente a modo de guía para confeccionar un **documento orgánico** y no una sucesión de ítems a tachar:

* Especificar el *dataset* utilizado, indicando la dirección de dónde se bajó (url).
* Especificar la/las preguntas que impulsan al trabajo.
* Sobre el *dataset,* señalar el universo y la población. En el caso que la web lo especifique, comentar cómo fue obtenida la muestra. Si no está implícita, generar una hipótesis de cómo podría haber sido obtenida.
* Sobre la muestra, resaltar el N utilizado. Fundamentar si se trata de una muestra representativa de la población.
* Mencionar algunos de los posibles sesgos que se pudieron haber presentado a la hora de tomar esas mediciones.
* Interpretar los resultados, no mostrar el número como un dato aislado. Una oración comentando sobre la métrica es el día y la noche en un reporte.
* Asumiendo que tuviera más datos u otras variables, ¿qué otro tipo de análisis se podría hacer? ¿En qué ayudaría a contestar la/s pregunta/s propuesta/s?
* Cualquier otro aporte que sume en el análisis / conclusiones / proyecciones, será más que bienvenido. ¡La originalidad no es más que unir cosas!

Si bien las preguntas a contestar están fuertemente ligadas a los datos disponibles, algunas posibles interrogantes para usar las herramientas vistas pueden ser: *“¿Cómo es la distribución de valores de algunas variables?”, “¿Corresponde a alguna de las distribuciones estudiadas?”, “¿Qué rango de valores tomaría como los principales?”, “¿Existen valores que señalan comportamientos atípicos? “¿Existe alguna correlación entre las variables?”, etc.*

## **Estructura de entregable**

El trabajo deberá contar con la siguiente información:

* Nombre y apellido de los integrantes, correo electrónico institucional.
* Descripción del *dataset* utilizado (nombre, link de descarga y comentarios de las columnas utilizadas), información o especulación de cómo se han obtenidos que lo componen.
* Claro planteo de la/s interrogante/s que motivan al análisis.
* Descripción de la metodología o herramientas utilizadas para contestar las preguntas (se pueden utilizar gráficos y/o resultados numéricos).
* Interpretación de los resultados del punto anterior, en un lenguaje para personas no especializadas en el tema.

A modo de ejemplo y guía, se expone a continuación una tentativa estructura del documento de texto entregable:

Integrantes del grupo: ***NOMBRE Y APELLIDO*** ***INTEGRANTE 1*** (***MAIL\_1@ucema.edu.ar***) - ***NOMBRE Y APELLIDO*** ***INTEGRANTE 2*** (***MAIL\_2@ucema.edu.ar***), …

*Nos preguntamos sobre* ***INTERROGANTE****. Para ello, trabajamos con el dataset* ***NOMBRE*** *(****LINK****). El mismo consiste en un registro de* ***NOMBRAR LAS VARIABLES PERTINENTES****, las cuales fueron medidas mediante* ***CÓMO SE OBTUVIERON / HIPÓTESIS DE CÓMO SE MIDIERON****. El dataset es una* ***MUESTRA / POBLACIÓN*** *con un* ***N*** *de individuos.*

*Para contestar la pregunta analizaremos* ***CONTAR QUÉ HICIERON Y CÓMO****… …. Los resultados arrojan que* ***RESULTADOS (PUEDE HABER GRÁFICOS QUE AYUDEN A LA JUSTIFICACIÓN)****. Respecto de la pregunta inicial, nos dice que* ***CÓMO CONTESTA ESTO A LA PREGUNTA****. Para extender esta respuesta se podría medir / encuestar* ***PROPUESTA DE QUÉ DATOS HUBIERAN SERVIDO*** *para así* ***QUÉ PROCEDIMIENTO -DESCRIPCIÓN NARRADA- HUBIERAN HECHO DE TENER ESOS DATOS****.*

***(SI HAY MÁS DE UNA PREGUNTA)****: Además / motivados por los resultados anteriores, nos preguntamos si …* ***(REPETIR LA ESTRUCTURA DEL PÁRRAFO ANTERIOR)****.*

1. Se puede utilizar otros *datasets* de un país de los integrantes del grupo, pero siempre de fuentes oficiales. Por ejemplo: <https://catalogodatos.gub.uy/dataset/> de Uruguay, <https://datos.gob.cl/en/dataset> de Chile, etc. [↑](#footnote-ref-1)