

Trabajo Integrador 2025 - Parte 2

Seminario de Lenguajes Opción Python

El trabajo integrador consta del desarrollo de una aplicación de búsqueda y visualización de información relacionada a la Encuesta Permanente de Hogares (EPH), a partir de ahora nos referiremos a la aplicación como Encuest.AR. El desarrollo de este trabajo involucra el análisis de los datos de la EPH, su limpieza y preparación necesaria y finaliza con el desarrollo de una interfaz amigable con el usuario/a. El objetivo es que logren aplicar sus conocimientos de programación mientras exploran datos reales y construyen una interfaz visual con Streamlit.

El proyecto se divide en dos etapas:

- **Primera Etapa:** se trabaja en los procesos de entender y limpiar los datos de la [EPH](#). Se dispone de los archivos que contienen información sobre hogares e individuos, a los cuales hay que aplicar filtros y prepararlos para su uso en la aplicación.
- **Segunda Etapa:** se enfoca en el diseño de la interfaz en Streamlit y la integración de los datos procesados. Se plantean diferentes secciones para mostrar información sobre educación, ocupación, vivienda y otros temas de interés.

Este trabajo será realizado en grupos de **cinco** personas, por lo que se espera una distribución equitativa de tareas. Cada integrante del equipo tiene un rol importante en el desarrollo, ya sea en la manipulación de datos, programación de funcionalidades, diseño de la interfaz o documentación del proyecto.

Además de mejorar su manejo de Python y librerías como *pandas*, *Matplotlib* y *Streamlit* (entre otras), este proyecto les permitirá desarrollar habilidades de trabajo en equipo, control de versiones con *GitLab* y documentación de código.

El presente documento describe el trabajo a realizar y los requerimientos de la segunda etapa.

Aclaraciones importantes:

1. Esta segunda etapa consta de todos los incisos explicados a continuación y todos los arreglos y/o mejoras solicitados por su ayudante a partir de la corrección de la etapa 1 del trabajo integrador.
2. Se habilita el uso de *pandas* como librería en el sistema: a partir de la etapa 2 **debe utilizarse *pandas* para las acciones que lo ameriten**. En cuanto al código ya realizado se permite refactorizar si lo consideran conveniente pero no es una obligación.
3. Se requerirá en la resolución la utilización de:
 - a. Mínimo tres tipos de chart element ([charts](#)).

- b. Mínimo un tipo de data element ([data](#)).
- c. Mínimo cuatro tipos de inputs ([inputs](#))

Pueden consultar con su ayudante la utilización de cada elemento en cada inciso, el propósito utilizar una variedad amplia de componentes brindados por Streamlit.

- 4. Recuerden ponderar los valores cuando y como resulte correcto.

1.- Encuest.AR: interfaz gráfica

En la sección 3.2 de la etapa 1 del trabajo integrador se solicitaba, a grandes rasgos, desarrollar la base de la interfaz gráfica desarrollada con ayuda de la librería Streamlit. En particular se solicitaba crear 4 páginas:

- (P1) Inicio
- (P2) Carga de datos
- (P3) Búsqueda por tema
- (P4) Visualización

Para comenzar con esta etapa 2 se solicita eliminar las “P3 Búsqueda por tema” y “P4 Visualización”, las mismas, no contienen funcionalidad hasta el momento y serán reemplazadas.

En la etapa dos el sistema contará de 7 páginas/secciones:

- (P1) Inicio.
- (P2) Carga de datos.
- (P3) Características demográficas.
- (P4) Características de la vivienda.
- (P5) Actividad y empleo
- (P6) Educación.
- (P7) Ingresos.

A su vez, se proporciona un archivo denominados *coordenadas aglomerados.json* que contiene por cada aglomerado el par de coordenadas que indican su ubicación en el territorio nacional.

Por último, se proporciona un archivo csv que contiene para cada año y mes el valor de la canasta básica de alimentos.

1.1 (P1) Inicio

Se debe mantener con la funcionalidad desarrollada hasta el momento en la etapa 1.

1.2 (P2) Carga de datos

A la funcionalidad desarrollada en la etapa 1 se debe agregar un *proceso* que controle que para cada uno de los archivos individuales existe el correspondiente de hogares. A su vez se debe controlar que para cada archivo de hogares exista su correspondiente de individuos. En caso de tener un faltante de este tipo, informe por pantalla el Año, Trimestre y archivo faltante.

En caso de no existir faltantes se debe mostrar un mensaje indicando que el chequeo resultó exitoso y no se encontraron inconsistencias.

1.3 (P3) Características demográficas

En esta sección se visualizará información relacionada a la características demográficas de la población argentina según la EPH.

1.3.1 Generar un gráfico de barras que muestre la distribución de la población por grupos de edad (cada 10 años) y sexo. El gráfico debe tener barras dobles (una para cada sexo) por cada grupo de edad. Para ello se debe solicitar al usuario que ingrese un año y trimestre. También se debe validar que el año se encuentre entre los cuales se tiene información.

1.3.2 Informar la edad promedio de personas por aglomerado para el último trimestre y año del cual se tenga información. Deben decidir la forma en la cual se va a dar a conocer esta información (tabla, gráfico, etc).

1.3.3 Debe solicitar al usuario que seleccione un aglomerado, a partir de dicha selección debe mostrar la evolución de la dependencia demográfica para todos los años y trimestres almacenados en el sistema.

La dependencia demográfica se define como el cociente de la cantidad de población de 0 a 14 años y mayores de 65 (se asumen jubilados) respecto a la población en edad activa (15 a 64 años) multiplicado por 100.

Se busca conocer si con el correr del tiempo la variación de esta proporción. Deben decidir la forma en la cual se va a dar a conocer esta información (tabla, gráfico, etc).

1.3.4 Informar para cada año y trimestre almacenado la media y mediana de la edad de la población.

1.4 (P4) Características de la vivienda

En esta sección se visualizará información relacionada a las características de las viviendas de la población argentina según la EPH.

Todas las visualizaciones se basan en un único valor de año ingresado por el usuario en esta sección.

Filtro para toda la página:

Seleccione un año para explorar las características habitacionales de la población argentina en ese período. También debe existir una opción para que no se filtre por ningún año en particular y tenga en cuenta todos los datos.

1.4.1 Se mostrará la cantidad total de viviendas incluidas en la encuesta para el año seleccionado.

1.4.2 Se presentará un gráfico de torta (pie chart) con la proporción de viviendas según su tipo.

1.4.3 Se informará, para cada aglomerado, cuál es el material predominante en los pisos interiores de las viviendas.

1.4.4 Se mostrará, por aglomerado, la proporción de viviendas que disponen de baño dentro del hogar en el año seleccionado.

1.4.5 Además del año seleccionado, se debe ingresar un **aglomerado específico**. Se mostrará la evolución del régimen de tenencia (propia, alquilada, cedida, etc.) para ese aglomerado. El usuario debe poder elegir el tipo de tenencia que desea ver: una, un conjunto o todas.

1.4.6 Informar de manera ordenada (decreciente) la cantidad de viviendas ubicadas en villa de emergencia por aglomerado. Además de la cantidad informar el porcentaje con respecto al total.

1.4.7 Informar para cada aglomerado el porcentaje de viviendas por CONDICION_DE_HABITABILIDAD. Además de informarse debe poder exportarse a un CSV los resultados.

1.5 (P5) Actividad y empleo

En esta sección se visualizará información relacionada a la actividad y empleo según la EPH.

1.5.1 Para las personas desocupadas, informar la cantidad de ellas según sus estudios alcanzados. Se debe informar para un año y trimestre elegido por el usuario.

1.5.2 Informar la evolución del desempleo(tasa de desempleo) a lo largo del tiempo. Se debe poder filtrar por aglomerado y en caso de no elegir ninguno se debe calcular para todo el país.

La tasa de desempleo es el cociente de personas desocupadas y la suma de personas desocupadas más ocupadas multiplicado por 100.

1.5.3 Informar la evolución del empleo (tasa de empleo) a lo largo del tiempo. Se debe poder filtrar por aglomerado y en caso de no elegir ninguno se debe calcular para todo el país.

La tasa de empleo es el cociente entre personas ocupadas y la suma de personas desocupadas más ocupadas multiplicado por 100.

1.5.4 Informar para cada aglomerado el total de personas ocupadas, el porcentaje con empleo estatal, el porcentaje con empleo privado y el porcentaje de otro tipo. Considerar la ocupación principal.

1.5.5 Se debe obtener por aglomerado el porcentaje de la tasa de empleo y desempleo. Esta información se requiere conocer para el año y trimestre más antiguo del cual se contenga información y para el año y trimestre más actual del cual se cuenta información.

A partir de dicha información se debe visualizar un mapa que por aglomerado muestre con el color de un punto/marca si el porcentaje aumentó o disminuyó. El usuario elegirá si desea ver tasa de empleo o desempleo:

- Al elegir la tasa de empleo se deben ver puntos verdes en los aglomerados cuya tasa de empleo aumentó con el correr del tiempo. Rojo en el caso contrario.
- Al elegir la tasa de desempleo se deben ver puntos rojos en los aglomerados cuya tasa de desempleo aumentó con el correr del tiempo. Verde en el caso contrario.

1.6 (P6) Educación

En esta sección se visualizará información relacionada al nivel de educación alcanzado por la población argentina según la EPH.

1.6.1 Informar la cantidad de personas según el máximo nivel educativo alcanzado (primario, secundario, terciario, universitario, etc). Se debe informar para un año ingresado por el usuario. Se debe elegir la forma adecuada de dar a conocer esta información.

1.6.2 Informar para personas entre 20 y 30 años cuál es el nivel educacional alcanzado más común, lo mismo para personas entre 30 y 40, 40 y 50, 50 y 60 y por último para mayores de 60.

En esta visualización de información el usuario puede decidir si ver solo un intervalo de etario, un conjunto de ellos o todos.

1.6.3 En el inciso 4 de la sección B del TI parte 1 se desarrolló un "Ranking de los 5 aglomerados con mayor porcentaje de hogares con dos o más ocupantes con estudios universitarios o superiores finalizados. Información obtenida a partir del par de archivos más recientes". Se requiere poder exportar a un archivo CSV el mencionado ranking mediante un botón o la interacción que crean conveniente.

1.6.4 A partir de la información de cada año contenida en el dataset se debe informar, año tras año, el porcentaje de personas mayores a 6 años capaces e incapaces de leer y escribir.

1.7 (P7) Ingresos

El campo ITF es la sumatoria de los ingresos individuales totales de todos los componentes del hogar. También contamos con el dataset que contiene el valor de la canasta básica familiar desde el 2016 a la fecha.

Se debe solicitar al usuario que ingrese un año y trimestre y a partir de esta información informar la cantidad y porcentajes de hogares que se encuentran por debajo de la línea de pobreza y por debajo de la línea de indigencia. Se deben filtrar solo hogares que contengan 4 integrantes.

El dataset con información de la canasta básica familiar posee una frecuencia mensual, por lo que por cada trimestre se cuenta con 3 valores distintos de canasta básica. Cada grupo

debe decidir con su ayudante designado que política tomar al respecto, quedarse con uno, promediar, etc.

Aclaraciones: los resultados obtenidos no poseen valor estadístico de ningún tipo, ya que los montos de la canasta básica pertenecen a CABA y la EPH es a nivel nacional. Además al filtrar solo hogares de 4 integrantes si conocer la cantidad de menores no es correcto. Los hogares con más o menos integrantes habría que considerarlos utilizando un coeficiente de equivalencia.

4.- Consideraciones generales de implementación

El software implementado deberá funcionar correctamente tanto en Windows, Linux o Mac. Se deberá armar una estructura de directorios organizando los archivos en carpetas y subcarpetas de manera tal que mantengan el código organizado haciendo que sea fácil de actualizar y de corregir. Se evaluará tanto el código como su organización.

El código se deberá subir a un repositorio de GitLab designado por la cátedra. En este repositorio, cada equipo deberá incluir un archivo denominado **README.md** que contenga el nombre de los integrantes del equipo y las instrucciones para ejecutar cada aplicación. Además, se solicitará un archivo que enumere las bibliotecas necesarias, siguiendo la metodología explicada en las prácticas para este propósito.

Para el desarrollo del código se deberá cumplir con los siguientes requerimientos:

- Usar la herramienta **Streamlit** para el desarrollo de la interfaz gráfica.
- Documentar el código usando docstrings en funciones y clases (estas últimas caso de ser utilizadas).
- Incluir un archivo **LICENSE** con la licencia del código.
- Se debe tener en cuenta las [guías de estilo de Python](#) para la escritura de código.
- Definir una estructura de carpetas que permita estructurar el código de forma prolija.

4.1.- Consideraciones de la entrega y defensa

Si bien el trabajo es grupal, **la nota de la defensa es INDIVIDUAL.**

La defensa se lleva a cabo durante el horario de consulta la semana posterior a la fecha de entrega. Este proceso implica la realización de un encuentro virtual o presencial con el ayudante asignado, durante el cual se formulan una serie de preguntas relacionadas con la entrega. Dichas preguntas están dirigidas a los distintos miembros del grupo, distribuyendo

equitativamente la participación entre todos ellos. La finalidad es que cada integrante tenga la oportunidad de responder. La evaluación se basa en el desempeño de las respuestas, lo que determina finalmente la nota asignada a cada miembro del grupo en relación con la entrega realizada.

Esto implica que, aunque se trate de una entrega grupal, las calificaciones entre los miembros del grupo pueden variar.

Tener en cuenta que es fundamental mantener una participación activa tanto en las consultas prácticas como en el repositorio de GitLab. Esto permitirá al ayudante obtener una comprensión conceptual del conocimiento de cada participante del grupo.

4.2.- Detalles de entrega

La fecha de entrega de la primera etapa es el día **19 de junio a las 23:59 hs.** **La defensa será la semana del 24 de junio en horario de la consulta práctica.**

La segunda entrega otorga al estudiante un **máximo de 40 puntos**, los cuales se basan en su desempeño individual durante las consultas y la defensa.

5.- Criterios de Evaluación

La distribución de los puntos de la entrega estará disponible en la rúbrica asociada a la tarea. Se evaluará:

- Funcionalidad implementada de acuerdo al enunciado.
- Cumplimiento de las consideraciones planteadas.
- Código subido en tiempo y forma al repositorio de GitLab indicado.
- Participación individual activa durante el desarrollo del trabajo, que incluye asistencia a las consultas (tanto virtuales como presenciales) y contribuciones al repositorio de GitLab.
- En la defensa, se espera que cada integrante del grupo demuestre los conocimientos utilizados para la realización del trabajo.