



Proces. Analítico de Datos Taller-2

1. Descripción general y planteamiento

Esta es la última etapa de la asignatura, en la que se deben aplicar todas las técnicas y conocimiento adquirido durante el curso. En este taller se espera que usted y su grupo realicen el proceso de análisis de datos para generar información útil y trascendente para los intereses de una organización.

El supuesto en este taller es el siguiente:

Usted y su grupo son analistas de datos que colabora con investigadores de la plataforma datos.gob, esta organización quiere unir datos para conocer hábitos de compra de las personas, de frutas y verduras junto a otros productos. Esto se quiere conocer para los distintos sectores y regiones de Chile, según registros en la página de la plataforma datos.gob.

Dado que el grupo de investigación no posee todas las competencias técnicas para procesar los datos, unificar la calidad de los datos, realizar tareas de limpieza y preparación, carga de datos y posterior visualización para el análisis, le ha solicitado a usted ayuda con las tareas mencionadas.

Basado en la descripción anterior, ud y su grupo deben realizar lo que se describe a continuación.

2. Respecto a los datos, usted y su grupo deben:

2.1. Juntar los datos disponibles en <https://datos.gob.cl/dataset/16294> con los datos disponibles en <https://datos.gob.cl/dataset/16294/resource/19045614-3013-4215-84f7-dd7dc4258888>, un único data set.

El primer enlace corresponde a datos de frutas y verduras y el segundo corresponde a datos de otros productos.

La unión de estos dos datasets considerando debe hacerse considerando que puedan extraerse datos por región, región y sector, sector, tipo de alimento (fruta y verdura u otros), e incluso combinación de ambos tipos de alimentos, siendo un tipo de alimento 'frutas y verduras' y el otro tipo de alimento 'otros'.

2.2. Una vez realizado lo descrito en el apartado a), revisar que la información se encuentre limpia y sea coherente. Una vez realizada la preparación de los datos, deberán ajustar la base de datos relacional para almacenar estos datos y los que probablemente se reciban en un futuro cercano.

2.3. Generar un algoritmo para validar la coherencia de los datos, es decir: que se pueda verificar que los datos de frutas y verduras y otros corresponden a un sector de una comuna y que no existe cruce de datos.

3. Respuestas a interrogantes respecto a alimentos consumidos

Los investigadores necesitan saber si hay hábitos de alimentación en las personas por comunas e incluso por regiones. Para esto usted deberá generar datos estadísticos y diseñar con PowerBi unos dashboards que permitan visualizar información como:

- 3.1. Valores máximos, valores mínimos y media de: frutas y/o verduras consumidas por región. Con PowerBi i se debe generar una forma para seleccionar, por comuna, la fruta o verdura de la cual se desee conocer (de forma individual) el valor máximo mínimo y valor medio.
- 3.2. Valores máximos, valores mínimos y media de: otros alimentos consumidos por región. Con PowerBi i se debe generar una forma para seleccionar, por comuna, el alimento del cual se desee conocer (de forma individual) el valor máximo mínimo y valor medio. Por ejemplo, el producto Abastero en Talca.
- 3.3. Visualizar información de ambos tipos de comida en un sector para una fecha determinada, esto con la idea de buscar patrones de consumo de alimentos.

Un ejemplo de la forma de selección de datos para la vista en dashboard es el siguiente:

Sector	Fecha	tipo punto de monitoreo (fruta y verdura)	producto	Total	tipo punto de monitoreo (otros)	producto	Total
Talca	29/07/2013	Feria libre	Fruta-Hortaliza	500	Carnicería	29/07/2013	500

- 3.4. Visualizar en mapa el sector de una región que más ventas registra de tipo de productos 'otros'.

4. Respuestas a interrogantes respecto a detalles de alimentos-precios

Se requiere presentar relación entre los dos tipos de consumo, para lo cual usted y su grupo deben decidir cómo mostrar información sintetizada (abstraída) y tener capacidad de mostrar datos concretos que sean de interés para los investigadores. En tal sentido, usted su grupo deben generar gráficos estadísticos que permitan:

- 4.1. Visualizar el total gastado en una semana para la compra de insumos en un determinado sector de una región.
- 4.2. Visualizar cuál es el sector de una región que tiene el mayor gasto en compra de insumos. Visualizar también cuál es la región que tiene el mayor gasto en compra de insumos.
- 4.3. Generar un contraste entre el precio de frutas y verduras para el último año de la base de datos y el precio actual.
- 4.4. Visualizar la evolución del consumo de un producto por región de un año, usando gráficos tipo Boxplot.

Usar como criterios: el producto más consumido, el producto menos consumido. Por ejemplo, mostrar la evolución anual del Tomate larga vida, si es el producto que más se consumió en la Región 7.

Los reportes de esta sección deben ser generados con Python.

5. Contenido del informe y forma de entrega

Se debe generar un informe que debe tener con la siguiente estructura:

- Una primera página con nombre del informe (por ejemplo "Informe Taller momento-3"), Nombres y Apellidos y el correo electrónico de cada integrante del grupo de trabajo.
- En una segunda página deben colocar un **Resumen** del trabajo

- Seguidamente el **Desarrollo** del trabajo, que deberá incluir lo indicado en el apartado anterior. En el desarrollo deben colocar capturas de pantalla del diseño de la base de datos y todo lo que de soporte a lo hecho en el taller, para llegar a la solución, como por ejemplo, capturas del código Py para crear las tablas crear los gráficos con Matplotlib, etc.
- **Conclusiones** donde podrían reflejar las experiencias y habilidades ganadas con el desarrollo de los proyectos en la asignatura. Aquí también se puede reflejar lo aprendido en la asignatura y se valorará si colocan sugerencias de mejoras (para los siguientes cursos, en las asignaturas de la carrera, etc.).
- **Bibliografía** en formato APA¹, donde reflejen todo lo consultado para realizar el trabajo como libros o *papers*, URLs, videos, etc. la bibliografía debe ser referenciada en el cuerpo del trabajo.

No se establece longitud (tamaño) mínimo del informe, pero sí debe tener

- 1 página inicial, con el nombre del informe y los datos de los grupos.
- 1 página aparte para el Resumen (aunque no se escriba toda una página)
- Desarrollo (en otra página posterior al Resumen).
- Conclusiones (en otra página posterior al Desarrollo).
- Referencias (en otra página posterior a Conclusiones).

6. Forma de entrega y evaluación

La forma de entrega será mediante la tarea “**Entrega Taller-2**” del campus virtual. Fecha límite de entrega el **viernes 15 de diciembre de 2023, hora límite: 23:55**. En esta tarea deberá subir un archivo comprimido (zip/rar) que debe contener:

- Proyecto Python
- Proyecto PowerBI
- PPT que usarán en la clase de presentación de resultados por grupos.

Condiciones adicionales:

1. No se admitirán entregas enviadas al correo electrónico, a menos que sea por mal funcionamiento del campus virtual (si esto pasa se debe adjuntar captura de pantalla del mal funcionamiento del campus al correo que envíen).
2. Es suficiente con que una persona de cada grupo suba la tarea al campus virtual.

¹ <https://normas-apa.org/wp-content/uploads/Guia-Normas-APA-7ma-edicion.pdf>

7. Calificación

El trabajo se evaluará en escala de 1 a 7 siguiendo lo especificado en la tabla de más abajo, la nota final será la suma de los puntajes según porcentajes descritos en las tablas siguientes. La nota para el trabajo se calculará según estos criterios:

	[1-3.9]	[4-5.5]	[5.6-7]
Documento (15%)	No entregado o no cumple con las especificaciones	Entregado a tiempo y cumple con la mayoría de las especificaciones	Entregado a tiempo y cumple con las especificaciones
Diseño, creación, carga (poblar) de BD relacional (15%)	No está o no cumple con lo especificado	Cumple con lo especificado, aportes poco significativos en las soluciones planteadas	Cumple con lo especificado, aportes significativos en las soluciones planteadas
Diseño, creación, carga de datos, en PowerBi con suficiente información (repartidas en hojas) para generar las visualizaciones y respuestas a lo solicitado (30%)	No está o no cumple con lo especificado	Cumple con lo especificado, aportes poco significativos en las soluciones	Cumple con lo especificado, aportes significativos en las soluciones
Uso de 'herramientas' de Py para dar valor a las soluciones (40%)	No está o no se usan adecuadamente. Caso concreto (de ejemplo): elementos como Matriz de dispersión no está o no se puede leer o visualizar las etiquetas de los ejes.	Cumple con lo especificado de forma básica, aportes poco significativos en las soluciones. Caso concreto (de ejemplo): elementos como Matriz de dispersión no se gestiona correctamente para poder leer bien o visualizar las etiquetas de los ejes.	Cumple con lo especificado, aportes significativos en las soluciones. Las visualizaciones son adecuadas y se puede leer/visualizar elementos que contribuyen adecuadamente a la facilidad de interpretación, por ejemplo: las etiquetas de los ejes, leyenda(s), etc.

La nota para la presentación se calculará según estos criterios:

Organización de la PPT (20%)	La PPT no se relaciona con lo hecho en el trabajo, no tiene ni sigue una agenda, no está organizada de forma que se entienda lo trabajado en el taller.	La PPT se relaciona mínimamente con lo hecho en el trabajo, sigue una agenda, está organizada de forma que se puede entender mínimamente lo trabajado en el taller.	La PPT se relaciona totalmente con lo hecho en el trabajo, sigue una agenda, está organizada de forma que se puede entender lo trabajado en el taller (de principio a fin).
Presentación de resultados, mediante el uso de PPT, Python (elementos gráficos) y PowerBI (80%)	La PPT no refleja el trabajo realizado o no hay PPT. No hay elementos de visualización gráfica con Python. El Proyecto PowerBI no cumple con lo solicitado o no funciona adecuadamente. El grupo de trabajo no crea el mínimo de 2 preguntas según lo descrito en el apartado 2.	La PPT refleja de forma básica el trabajo realizado. Los elementos de visualización en Python no son eficientes o no permiten una buena interpretación de los datos. El Proyecto PowerBI cumple de forma básica con lo solicitado, aportando visualizaciones y manejo básico de los datos El grupo de trabajo crea el mínimo de 2 preguntas según lo descrito en el apartado 2, pero son de poco valor o los instrumentos diseñados no son claros o no aportan a la visualización.	La PPT describe de manera adecuada y apropiada el trabajo realizado. Los elementos de visualización en Python son eficientes o permiten una buena interpretación de los datos. El proyecto PowerBI cumple de forma integral con lo solicitado, brindando visualizaciones con aporte significativo para visualizar información y datos a distintos niveles de abstracción. El grupo de trabajo crea el mínimo de 2 preguntas según lo descrito en el apartado 2, aportan valor y los instrumentos diseñados son claros y aportan significativamente a la visualización de resultados.