

Pr. Proces. Analítico de Datos Taller 2 del momento-2

1. Descripción general y planteamiento

En esta parte se continúa con el proceso de análisis para generar información útil y trascendente para los intereses de una organización. En esta segunda parte, usted y su grupo deberán generar información de interés (y actualizada) para el negocio de alquileres en Airbnb en la ciudad donde está centrada la investigación descrita en el Taller 1 momento-2. Para actualizar la información deberá:

- Localizar una fuente de datos recientes de Airbnb en la web, descargar los datos en formato .csv
- Validar la compatibilidad de los datos descargados con los datos disponibles en el taller 1
- Realizar limpieza y generar un archivo Excel con por lo menos 3 hojas de datos, que permitan posteriormente generar un modelo multidimensional.

Los datos que se generen deben guardarse en una base de datos relacional con capacidad para alojar nuevos datos de otras ciudades que se quieran analizar en el futuro. El manejo de la base de datos puede ser en MySQL o PostgreSQL.

También se debe generar lo necesario para disponer de una base de datos multidimensional que permita el uso de cubos OLAP como filosofía de trabajo para generar las respuestas a las preguntas de más abajo. La creación de estas tablas y la incorporación de datos a dichas tablas debe ser programado en Python, usando funciones de librerías disponibles para este fin.

Se desea tener respuesta a preguntas como las siguientes:

Pregunta 1:

En esta primera pregunta se desea evaluar la distribución de alojamientos, ¿cuáles son los sectores que agrupan la mayor cantidad de alojamientos. Destacar gráficamente, por ejemplo, en un gráfico de torta/pizza los porcentajes de ubicación de alojamientos por sectores.

Pregunta 2:

En esta pregunta se desea conocer detalle de rentabilidad de los alojamientos en Lisboa. Realizar un análisis de la distribución de precios de alojamientos por sectores. Indicar valores medios, mayor valor (respecto al resto de sectores) y menor valor (respecto al resto de sectores). Esto se puede hacer con un gráfico Boxplot (Py – Pandas).

Pregunta 3:

En esta parte se desea proyectar rentabilidad, disponibilidad y continuidad de los alojamientos en Airbnb. Para ello se debe responder a lo siguiente: ¿Se nota incremento en la cantidad de alojamientos en Lisboa? Es decir, ¿se han incluido nuevos alojamientos? En caso de comprobar que se han incluido nuevos alojamientos (respecto a la data original mente disponible en el taller 1), ¿Sigue siendo rentable la actividad?, ¿qué tanto ha crecido? En caso de comprobar que no hay nuevos alojamientos, presentar una explicación (basada en el análisis de los datos).



Pregunta 4:

En esta parte se desea contrastar Lisboa, que es una capital de Europa, con los alojamientos de una capital de otras ciudades de Europa o Latinoamérica. Para responder a lo antes planteado usted y su grupo deberá descargar datos para la ciudad que se le asigne. Seguidamente, deberá usar recursos de forma similar a lo usado en la pregunta dos, y recursos de visualización de PowerBI para sustentar el contraste y poder explicar lo encontrado.

Las respuestas a las preguntas anteriores y el argumento o justificación de dichas respuestas deben estar reflejado en el cuerpo del informe, así como la programación (código fuente) Usado en cada caso. La parte final del informe debe incluir un informe tipo 'cuadro de mando', similar al hecho en la clase del 25 de octubre de 2022. Este reporte debe permitir realizar consultas con las dimensiones definidas por el grupo de trabajo, debe Permitir visualizar los datos e información que sustenten las respuestas a las 4 preguntas anteriores, debe contener filtros que relacionen información de detalle de una ciudad, contrastes entre 2 ciudades, etc. Este cuadro de mando debe ser incluido en la entrega como proyecto de PowerBI.

2. 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 2 momento-2"), Nombres y Apellidos de cada integrante del grupo de trabajo y el correo electrónico.
- En una segunda página deben colocar un **Resumen** del trabajo
- Seguidamente el **Desarrollo** del trabajo, que deberá incluir lo indicado en el apartado anterior.
- **Conclusiones** donde podrían reflejar las experiencias y habilidades ganadas con el desarrollo del proyecto
- **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, una página aparte para el Resumen (aunque no se escriba toda una página), comenzar el Desarrollo en otra página (posterior al Resumen), luego comenzar las Conclusiones (si existen) en otra página (posterior al Desarrollo) y finalmente en otra página, las Referencias.

2.1. Forma de entrega y evaluación

La forma de entrega será mediante la tarea "**Entrega Taller 2 del momento-2**" del campus virtual. Fecha límite de entrega el **martes 15 de noviembre de 2022, hora límite: 20:00**. En esta tarea deberá subir un archivo comprimido (zip/rar) que debe contener:

- Proyecto Python
- Proyecto PowerBI
- PPT que usarán en la clase del miércoles 16 de noviembre de 2022.

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



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). Es suficiente con que una persona de cada grupo suba la tarea.

Cumpliendo con la planificación detallada en el campus virtual², el miércoles 16 de noviembre se realizarán presentaciones por grupos del trabajo, para ello deben generar una PPT que contenga básicamente las respuestas a las preguntas y los gráficos o valores estadísticos que 'sustenten' la solución. Deberán tener disponible el cuadro de mando hecho en PowerBI para dar respuesta a cualquier pregunta relativa a la solución.

En estas presentaciones cada grupo mostrará la forma en que solucionó los casos y se realizarán preguntas relacionadas a dichas soluciones.

Es importante señalar que en el caso de que un grupo no participe en estas presentaciones (es decir que no esté presente ninguna persona de un grupo y no realicen la presentación), Tendrá una penalización del 30% de la nota final en el taller. Por ejemplo, si un grupo tiene una nota de 5.0 en el taller, pero no participa en la sesión del día 16 de noviembre, la nota final del taller será 3.5.

3. Puntaje

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 la tabla y siguiendo la restricción descrita en el apartado 2.1 (si aplica).

	[1-3.9]	[4-5.5]	[5.6-7]
Documento (10%)	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 datos en BD relacional (10%)	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 BD relacional (10%)	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
Uso de 'herramientas' de Py para dar valor a las soluciones (30%)	No está o no se usan adecuadamente	Cumple con lo especificado, aportes poco significativos en las soluciones	Cumple con lo especificado, aportes significativos en las soluciones

² https://campusvirtual.ucn.cl/pluginfile.php/2192190/mod_resource/content/1/plan%20trabajo-fechas%20evaluacio%CC%81n.pdf



Presentación de resultados, mediante el uso de PPT y PowerBI (40%)	La PPT no refleja el trabajo realizado o no hay PPT. El Proyecto PowerBI no cumple con lo solicitado o no funciona adecuadamente.	La PPT refleja de forma básica el trabajo realizado. El Proyecto PowerBI cumple de forma básica con lo solicitado, aportando visualizaciones y manejo básico de los datos	La PPT describe de manera adecuada y apropiada el trabajo realizado. 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.
--	---	---	---