

Unidad 1 - Planilla de Cálculo

Fundamentos de ciencia de datos



Práctica 1

En el campus encontrarán los archivos con los que trabajamos en la clase y los necesarios para resolver esta práctica.



IMPORTANTE !!!! Descarguen los archivos y súbanlos a su propio Drive

Parte 1

Utilizando el archivo `peliculas_top100.xlsx` resuelva:

1. Se requiere generar una tabla a partir del dataset `peliculas_top100.xlsx` que contiene una lista de películas de IMDB. La tabla debe incluir información adicional sobre cada película que será ingresada en forma manual en la primera columna, y automáticamente muestre el nombre del director, la duración y el rating en las subsiguientes columnas. Para lograr esto, se pide utilizar la función VLOOKUP en cada celda correspondiente. La tabla resultante deberá contener 5 cinco películas y se creará en una hoja (sheet) separada dentro del mismo archivo.
2. Realizar la tabla pero mediante el uso de INDEX y MATCH en forma anidada.
Ejemplo esperado

Título	Director	Duración	Puntaje
Black Adam	Jaume Collet Serra	125	6.5
She Said	Maria Schrader	129	7.2

1. Mediante la creación de tablas dinámicas genere tablas que respondan:
 - a. ¿Cuántas películas de cada director se encuentran en el archivo de IMDB ?
 - b. ¿Qué duración (*Runtime*) promedio tienen las películas según el género del cual se trate ?
 - c. ¿Cuántas películas se filmaron en cada país para cada año?
 - d. ¿Cuál fue el presupuesto (*Budget*) anual de cada director?

Parte 2

Utilizando el archivo `superficie-incendiada-por-provincias-tipo-de-vegetacion_2022.xlsx` resuelva:

1. ¿En qué año se registró la mayor superficie quemada total a nivel país?
2. ¿Cuál fue la provincia, para cada año de la serie de tiempo, con mayor superficie de área quemada total? (Para esta pregunta podrá utilizar una combinación de tabla dinámica y las otras funciones vistas en clase). Algunos resultados que debería obtener:

Año	Provincia con mayor superficie quemada
1993	La pampa
1994	La Pampa
1995	Córdoba
1996	Catamarca

3. Para cada provincia, ¿cuál fue el año en que se registró la mayor superficie de área quemada total? (Para esta pregunta podrá utilizar una combinación de tabla dinámica y las otras funciones vistas en clase). Algunos resultados que debería obtener:

Parte 3

El archivo `personas_2020.xlsx` contiene información sobre los empleados de CONICET en 2020. En este archivo, algunas variables han sido codificadas y su referencia se encuentra en las otras hojas de la planilla de cálculo. Por ejemplo, la columna `sexo_id` tiene valores de 1 y 2, y la referencia a dichos valores se encuentra en la hoja `ref_sexo` (según esta otra hoja, el valor 1 corresponde a masculino y el valor 2 corresponde a femenino).

1. Inserte una nueva columna en la hoja *personas_2020* que incluya la variable *sexo_descripcion* (utilizando las funciones de búsqueda descritas en clase).
2. Inserte una nueva columna en la hoja *personas_2020* que se denomine *max_grado_academico_descripcion* e incluya la variable *descripción* de la hoja *ref_grado_academico* (utilizando las funciones de búsqueda descritas en clase)
3. Realice una tabla dinámica en la cual se muestre cuántos empleados de CONICET hay de cada sexo para cada máximo grado académico en 2020.