

Laboratorio 15

Sesión # 15 Limpieza con Power BI

Título del Laboratorio: Aplicación y uso de la herramienta de Power BI para hacer la limpieza de los datos.

Duración: 2 horas

Objetivos del Laboratorio:

1. *Afianzar los conocimientos y manejo básico en Power BI para realizar las visualizaciones con ejercicios prácticos planteados.*

Materiales Necesarios:

1. *Computador con acceso a internet.*
2. *Colocarlo en el repositorio de Github*
3. *Ampliar el conocimiento con el curso de datos en AWS y Cisco.*
4. *Power BI descargarlo*

Estructura del Laboratorio:

Parte 1

En la primera parte se aplicarán los temas vistos en la sesión como es la limpieza de los datos, se deberá realizar **el paso a paso con las respectivas capturas de pantalla**, esta aplicación es de acuerdo con los escenarios planteados.

1. Ejercicio de práctica 1.

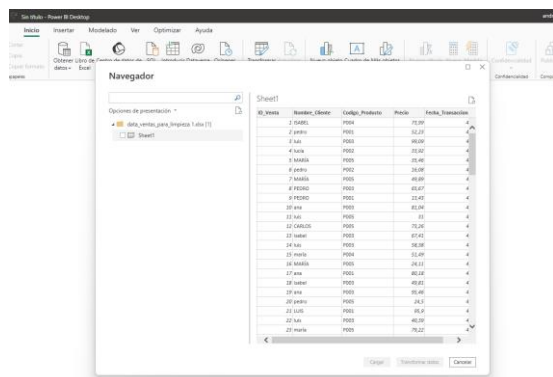
1. Escenario datos de ventas en línea

Se tiene un conjunto de datos de una empresa de ventas online que contiene información de clientes, productos vendidos y fechas de transacciones. El archivo tiene problemas comunes como **valores faltantes en la columna de códigos de productos, precios con formato incorrecto y nombres de clientes en mayúsculas y minúsculas mezcladas.**

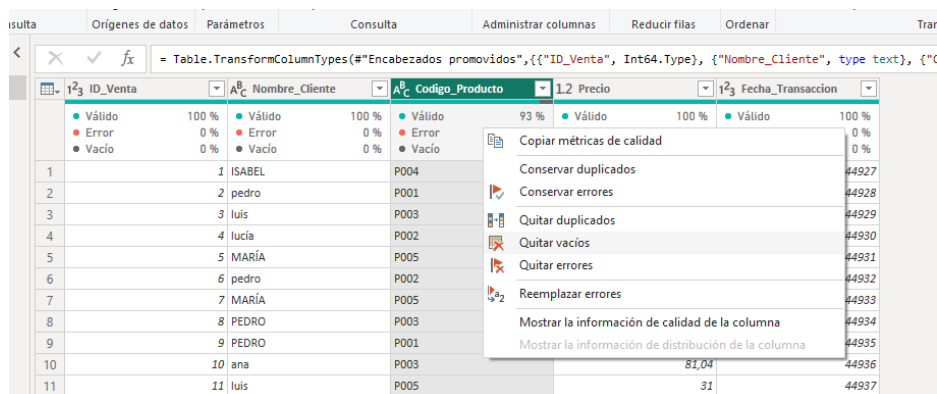
Data: data_ventas_para_limpieza 1

Paso a paso para la limpieza

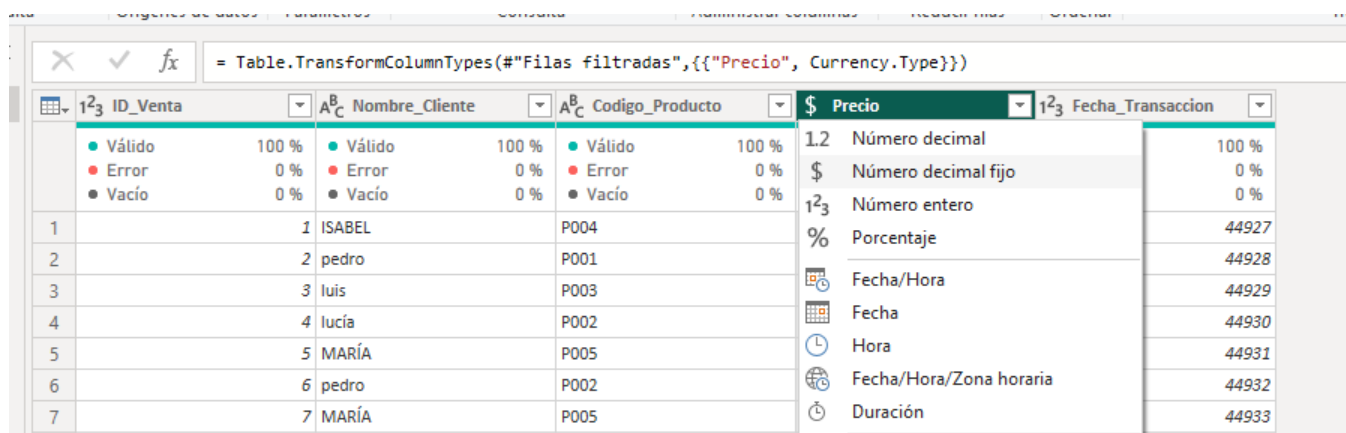
Paso 1: Importar los Datos en Power BI



Paso 2: Eliminar valores faltantes en la columna de códigos de productos



Paso 3: Precios con formato incorrecto, se arregla el formato a decimal fijo.



Paso 4: Arreglar los nombres de clientes en mayúsculas y minúsculas mezcladas, se arregla el formato para que la primera letra este en mayúscula

Paso 5: Se arregla el formato de las fechas, que esta en formato numero se pasa a fecha

Paso 6: Aplicar los Cambios y Cargar los Datos

CONCLUSIONES

Entre los escenarios de limpieza se concluyó que el borrado de los datos vacíos de los valores de ventas era necesario debido a que es un dato primordial de la tabla y sin este valor, la fila como tal carece de valor.

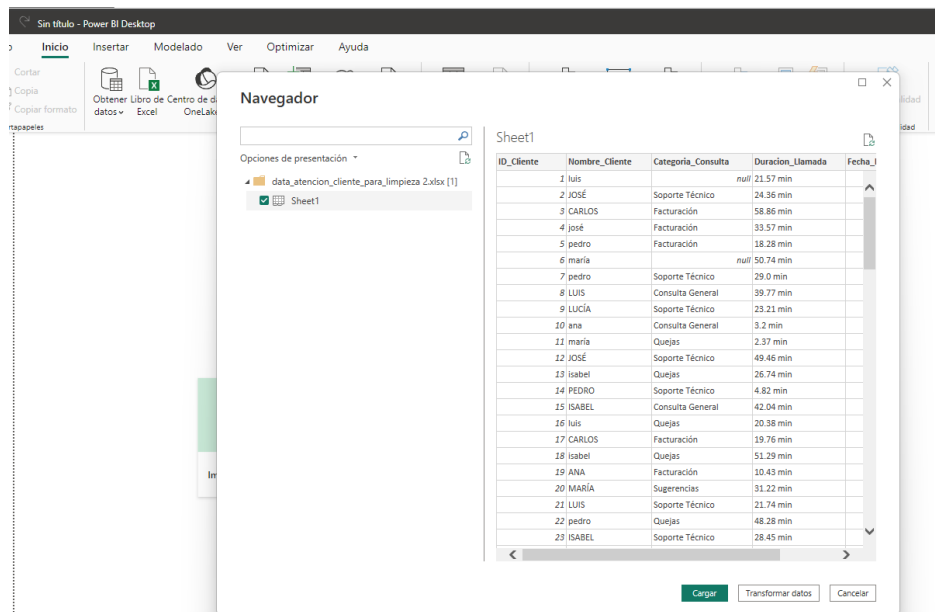
2. Escenario: Limpieza de datos de un servicio de atención al cliente

Trabajas en un centro de atención al cliente y tienes un conjunto de datos que contiene información sobre las interacciones con los clientes. Este dataset incluye el ID del cliente, Nombre del cliente, Categoría de consulta, Duración de la llamada y Fecha de la interacción. Algunos registros tienen problemas como **nombres en minúsculas y mayúsculas inconsistentes**, **valores faltantes en la categoría de la consulta**, y **duraciones de llamada con errores de formato**.

Data: data_atencion_cliente_para_limpieza 2

Paso a paso para la limpieza

Paso 1: Importar los Datos en Power BI



Paso 2: valores faltantes en la categoría de la consulta

= Table.TransformColumnTypes(#"Encabezados promovidos",{{"ID_Cliente", Int64.Type}, {"Nombre_Cliente", Text.Type}, {"Categoria_Conсульта", Text.Type}, {"Duracion_Llamada", Duration.Type}, {"Fecha_Interaccion", Date.Type}}					
ID_Cliente	Nombre_Cliente	Categoria_Conсульта	Duracion_Llamada	Fecha_Interaccion	
1	Luis	Soporte Técnico	3.2 min	2/01/2023	
2	JOSÉ	Facturación	58.86 min	3/01/2023	
3	CARLOS	Facturación	33.57 min	4/01/2023	
4	José	Facturación	18.28 min	5/01/2023	
5	pedro	Facturación	29 min	7/01/2023	
6	maría	Soporte Técnico	39.77 min	8/01/2023	
7	pedro	Soporte Técnico	23.21 min	9/01/2023	
8	LUIS	Consulta General	3.2 min	10/01/2023	
9	LUCÍA	Soporte Técnico	2.37 min	11/01/2023	
10	ana	Consulta General	49.46 min	12/01/2023	
11	maría	Quejas	26.74 min	13/01/2023	
12	JOSÉ	Soporte Técnico			
13	isabel	Quejas			

Paso 3: Duraciones de llamada con errores de formato, para este caso se quitó la palabra “ min”, se cambio el punto por la coma y se paso a formato numero con el fin de que esta columna tuviera la capacidad de usar operaciones matemáticas

= Table.TransformColumnTypes(#"Valor reemplazado1",{{"Duracion_Llamada", type number}})					
ID_Cliente	Nombre_Cliente	Categoria_Conсульта	Duracion_Llamada	Fecha_Interaccion	
1	José	Soporte Técnico	24,36	2/01/2023	
2	Carlos	Facturación	58,86	3/01/2023	
3	José	Facturación	33,57	4/01/2023	
4	Pedro	Facturación	18,28	5/01/2023	
5	Pedro	Soporte Técnico	29	7/01/2023	
6	Luis	Consulta General	39,77	8/01/2023	
7	Lucía	Soporte Técnico	23,21	9/01/2023	
8	Ana	Consulta General	3,2	10/01/2023	
9	María	Quejas	2,37	11/01/2023	
10	José	Soporte Técnico	49,46	12/01/2023	
11	Isabel	Quejas	26,74	13/01/2023	

Paso 4: Nombres en minúsculas y mayúsculas inconsistentes

Transformar Agregar columna Vista Herramientas Ayuda

Orígenes recientes Especificar datos Configuración de origen de datos Administrar parámetros Actualizar vista previa Editor avanzado Administrar Elegir columnas Quitar columnas Conservar filas Quitar filas Ordenar Dividir columna Agrupar por Reemplazar Tipo de datos: T Usar la prim

Nueva consulta = Table.SelectRows(#\"Tipo cambiado\", each [Categoría_Conulta] <> null and [Categoría_Conulta] <> \"\")

ID_Cliente	Nombre_Cliente	Categoría_Conulta	Duración_Llamada	Fecha_Interacción
1	JOSÉ			
2	CARLOS			
3	José			
4	pedro			
5	pedro			
6	LUIS			
7	LUCÍA			
8	ana			
9	maria			
10	JOSÉ			
11	Isabel			
12	PEDRO			
13	ISABEL			
14	luis			
15	CARLOS			
16	Isabel			

minúsculas
MAYÚSCULAS
Poner En Mayúsculas Cada Palabra
Recortar
Limpiar
Longitud
JSON
XML

AS Generación de perfiles de columnas basada en las 1000 p

Paso 5: Arreglar el formato fecha

Orígenes recientes Especificar datos Configuración de origen de datos Administrar parámetros Actualizar vista previa Editor avanzado Administrar Elegir columnas Quitar columnas Conservar filas Quitar filas Ordenar Dividir columna Agrupar por Reemplazar Tipo de datos: N Usar la prim

Nueva consulta = Table.TransformColumns(#\"Filas filtradas\",{{\"Nombre_Cliente\", Text.Proper, type text}})

ID_Cliente	Nombre_Cliente	Categoría_Conulta	Duración_Llamada	Fecha_Interacción
64	María	Sugerencias	3.32 min	
65	Ana	Facturación	9.63 min	
66	Isabel	Consulta General	17.95 min	
67	María	Facturación	27.21 min	
68	María	Sugerencias	21.09 min	
69	María	Consulta General	57.84 min	
70	José	Sugerencias	28.33 min	
71	Pedro	Sugerencias	29.24 min	
72	Luis	Sugerencias	20.24 min	
73	Isabel	Consulta General	5.58 min	
74	ana	Facturación	10.05 min	

1.2 Número decimal
\$ Número decimal fijo
123 Número entero
% Porcentaje
Fecha/Hora
Fecha
Hora
Fecha/Hora/Zona horaria
Duración
Texto
Verdadero/Falso
Rinarin

Paso 6: Aplicar los Cambios y Cargar los Datos

Sin título - Power BI Desktop

Archivo Inicio Ayuda Herramientas de tablas

Nombre Sheet1

Marcar como tabla de fechas Administrar relaciones Nueva Medida Nueva Nueva medida rápida columna tabla

Estructura Calendarios Relaciones Cálculos

ID_Cliente	Nombre_Cliente	Categoria_Conulta	Duracion_Llamada	Fecha_Interaccion
2	José	Soporte Técnico	24,36	lunes, 2 de enero de 2023
3	Carlos	Facturación	58,86	martes, 3 de enero de 2023
4	José	Facturación	33,57	miércoles, 4 de enero de 2023
5	Pedro	Facturación	18,28	jueves, 5 de enero de 2023
7	Pedro	Soporte Técnico	29	sábado, 7 de enero de 2023
8	Luis	Consulta General	39,77	domingo, 8 de enero de 2023
9	Lucía	Soporte Técnico	23,21	lunes, 9 de enero de 2023
10	Ana	Consulta General	3,2	martes, 10 de enero de 2023
11	María	Quejas	2,37	miércoles, 11 de enero de 2023
12	José	Soporte Técnico	49,46	jueves, 12 de enero de 2023
13	Isabel	Quejas	26,74	viernes, 13 de enero de 2023
14	Pedro	Soporte Técnico	4,82	sábado, 14 de enero de 2023
15	Isabel	Consulta General	42,04	domingo, 15 de enero de 2023
16	Luis	Quejas	20,38	lunes, 16 de enero de 2023
17	Carlos	Facturación	19,76	martes, 17 de enero de 2023
18	Isabel	Quejas	51,29	miércoles, 18 de enero de 2023
19	Ana	Facturación	10,43	jueves, 19 de enero de 2023
20	María	Sugerencias	31,22	viernes, 20 de enero de 2023
21	Luis	Soporte Técnico	21,74	sábado, 21 de enero de 2023

CONCLUSIONES

Entre los escenarios de limpieza se concluyó que el borrado de los datos vacíos **en la categoría** era necesario debido a que es un dato primordial de la tabla y sin este valor, la fila como tal carece de valor.

3. Escenario: Tienda minoritaria de tecnología

Trabajas para una tienda minoritaria tecnología, el equipo de inventario tiene dificultades para gestionar adecuadamente los datos de productos debido a la calidad inconsistente de los registros. El conjunto de datos de inventario incluye información sobre ID de producto, Nombre del producto, Cantidad disponible, Precio unitario y Fecha de ingreso al inventario. Sin embargo, se han detectado varios problemas en los datos que impiden su uso eficaz para análisis de ventas y reposición de stock.

Los problemas para solucionar incluyen:

- **Nombres de productos faltantes.**
- **Precios de algunos productos faltantes.**
- **Cantidades de productos con texto adicional "unidades".**

El objetivo es optimizar la calidad de los datos de inventario mediante la corrección de errores comunes, asegurando que los registros sean consistentes, completos y adecuados para análisis de inventarios y toma de decisiones estratégicas en la tienda.

Data: data_inventario_para_limpieza 3

Paso a paso para la limpieza

Paso 1: Importar los Datos en Power BI

Navegador

Opciones de presentación

- data_inventario_para_limpieza 3.xlsx [1]
 - Sheet1

Sheet1

ID_Producto	Nombre_Producto	Cantidad	Precio_Unitario	Fecha_Ingreso
1	Teclado	47 unidades	669,54	
2	Laptop	42 unidades	325,79	
3	Impresora	29 unidades	516,47	
4	Teclado	24 unidades	210,03	
5	Mouse	22 unidades	114,99	
6	Mouse	46 unidades	838,36	
7	Laptop	37 unidades	520,16	
8	Cámara	8 unidades	652,74	
9	Laptop	4 unidades	82,43	
10	Impresora	10 unidades	772,89	
11	Mouse	34 unidades	125,02	
12	Mouse	17 unidades	407,14	
13	Mouse	23 unidades	null	
14	Cámara	5 unidades	439,3	
15	Teléfono	44 unidades	800,79	
16	Teclado	18 unidades	746,9	
17	Teléfono	10 unidades	null	
18	Teclado	31 unidades	901,05	
19	null	27 unidades	956,01	
20	Impresora	14 unidades	191,2	
21	Teclado	20 unidades	632,71	
22	Teléfono	40 unidades	197,23	
23	Teléfono	27 unidades	690,51	

Cargar Transformar datos Cancelar

Paso 2: Nombres de productos faltantes, se procede a borrar dichos datos

Orígenes recientes Especificar datos Configuración de origen de datos Administrar parámetros Actualizar vista previa Administrar Consultas Elegir columnas Quitar columnas Conservar filas Quitar filas

Nueva consulta Orígenes de datos Parámetros Consultas Administrar columnas Reducir filas

Table.TransformColumnTypes(#"Encabezados promovidos",{"ID_Producto", Int64.Type

ID_Producto	Nombre_Producto	Cantidad	Precio_Unitario
69	Cámara		561,2
70	Mouse		244,5
71	Monitor		421,7
72	Laptop		290,5
73	Mouse		911,8
74	Teléfono		nu
75	Mouse		64,7
76	Monitor		349,6
77	Teléfono		nu
78	Impresora	35 unidades	194,4
79	Cámara	17 unidades	51,0
80	Mouse	48 unidades	949,1
81	null	31 unidades	811,4

Calidad de datos

- Copiar métricas de calidad
- Conservar duplicados
- Conservar errores
- Quitar duplicados
- Quitar vacíos
- Quitar errores
- Reemplazar errores
- Mostrar la información de calidad de la columna
- Mostrar la información de distribución de la columna

Paso 3: Precios de algunos productos faltantes, se procede a borrar dichos datos

ID_Producto	Nombre_Producto	Cantidad	Precio_Unitario	Fecha_Ingreso
1	Teclado	47 unidades		
2	Laptop	42 unidades		
3	Impresora	29 unidades		
4	Teclado	24 unidades		
5	Mouse	22 unidades		
6	Mouse	46 unidades		
7	Laptop	37 unidades		
8	Cámara	8 unidades		
9	Laptop	4 unidades		
10	Impresora	10 unidades		
11	Mouse	34 unidades		

Paso 4: Cantidades de productos con texto adicional "unidades", se quita la palabra "unidades" y se cambia el formato de la columna a decimal.

Reemplazar los valores

Reemplace un valor con otro de las columnas seleccionadas.

Valor que buscar: unidades

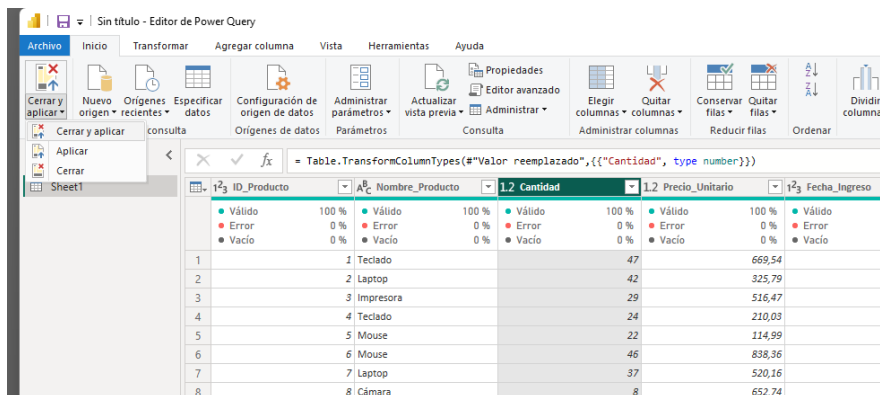
Reemplazar con:

Opciones avanzadas

Formatos de datos para Cantidad:

- Número decimal (seleccionado)
- Número decimal fijo
- Número entero
- Porcentaje
- Fecha/Hora
- Fecha
- Hora

Paso 6: Aplicar los Cambios y Cargar los Datos



The screenshot shows the Power Query Editor interface. The ribbon includes 'Inicio', 'Transformar', 'Agregar columna', 'Vista', 'Herramientas', and 'Ayuda'. The 'Aplicar' button is highlighted in the 'Inicio' tab. Below the ribbon, the formula bar shows the query: `= Table.TransformColumnTypes(#"Valor reemplazado",{"Cantidad", type number})`. The data table below has the following columns and values:

ID_Producto	Nombre_Producto	Cantidad	Precio_Unitario	Fecha_Ingreso
1	Teclado	47	669,54	
2	Laptop	42	325,79	
3	Impresora	29	516,47	
4	Teclado	24	210,03	
5	Mouse	22	114,99	
6	Mouse	46	838,36	
7	Laptop	37	520,16	
8	Cámara	8	652,74	

Parte 2

En la segunda parte una vez realizado la limpieza y el proceso deberán guardar el archivo en el repositorio, deberás anexar la captura de pantalla con el nombre del archivo.

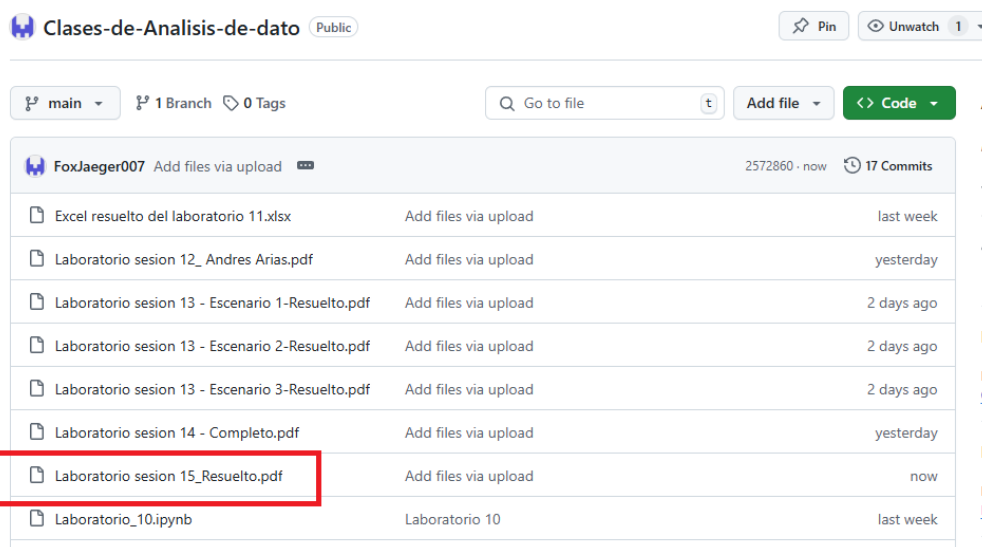
2. Ejercicio de práctica 2.

Una vez realizado la limpieza y el proceso deberán guardar el archivo en el repositorio, deberás anexar la captura de pantalla con el nombre del archivo.

CONCLUSIONES

La limpieza de datos es importante para que esta este preparada para su uso.

Imagen del repositorio



The screenshot shows a GitHub repository page for 'Clases-de-Analisis-de-dato'. The repository is public and has 17 commits. The file list includes:

File Name	Upload Method	Time
Excel resuelto del laboratorio 11.xlsx	Add files via upload	last week
Laboratorio sesion 12_ Andres Arias.pdf	Add files via upload	yesterday
Laboratorio sesion 13 - Escenario 1-Resuelto.pdf	Add files via upload	2 days ago
Laboratorio sesion 13 - Escenario 2-Resuelto.pdf	Add files via upload	2 days ago
Laboratorio sesion 13 - Escenario 3-Resuelto.pdf	Add files via upload	2 days ago
Laboratorio sesion 14 - Completo.pdf	Add files via upload	yesterday
Laboratorio sesion 15_Resuelto.pdf	Add files via upload	now
Laboratorio_10.ipynb	Laboratorio 10	last week