

```
%Fecha de actualización: 21/Septiembre/2023
%clc
%clear all
```

Portada

```
figure(15);
% tamaño de la figura
set(gcf, 'Position', [100, 100, 800, 600]);
tamano_letra=10;
% Cargar y mostrar la primera imagen
img1 = imread('UNAMlogo.png'); % Reemplaza 'imagen1.jpg' con el nombre de tu primera imagen
subplot(4, 4, 1); % Divide la figura en 1 fila y 2 columnas, y selecciona la primera posición
imshow(img1);

% Cargar y mostrar la segunda imagen
img2 = imread('FIlogo.jpg'); % Reemplaza 'imagen2.jpg' con el nombre de tu segunda imagen
subplot(4, 4, 4); % Selecciona la segunda posición
imshow(img2);

% Agregar texto a un subplot específico (por ejemplo, el subplot 8)
subplot(4, 4, [2,3]);
text(0.5, 0.9, 'UNAM', 'FontSize', 14,'HorizontalAlignment', 'center','Color','b');
text(0.5, 0.5, 'Facultad de ingeniería', 'FontSize', tamano_letra, 'FontWeight', 'bold', 'HorizontalAlignment', 'center');
text(0.5, 0.0, 'TSISB', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
axis off
subplot(4, 4, [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]);
text(0.5, 0.9, 'Alumna: Flores Morín María Alejandra', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.8, 'No. Cuenta: 315165805', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.6, 'Profesor: Dr. en C. Luis Antonio Aguilar Pérez ', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.4, 'Semestre 2024-1', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.3, 'tareaEjercicio1 Ejercicios de DATASTORE en MATLAB', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
text(0.5, 0.1, 'Fecha de entrega: 21 de septiembre del 2023', 'FontSize', tamano_letra,'HorizontalAlignment', 'center');
axis off
```



UNAM

Facultad de ingeniería

TSISB



Alumna: Flores Morín María Alejandra

No. Cuenta: 315165805

Profesor: Dr. en C. Luis Antonio Aguilar Pérez

Semestre 2024-1

tareaEjercicio1 Ejercicios de DATASTORE en MATLAB

Fecha de entrega: 21 de septiembre del 2023

```
% Guardar la portada como imagen
% saveas(gcf, 'portada.png');
```

Un modelo de datos o DataFrame es una estructura de datos tabular, la cual organiza la información en filas y columnas de manera similar a una hoja de cálculo. En particular cada columna del DataFrame representa una variable o feature (atributo), mientras que cada fila representa una observación o registro realizado. Los DataFrames son utilizados comúnmente en análisis de datos para manipular y analizar grandes conjuntos de datos de forma eficiente. Existen diversos métodos de almacenar esta información dependiendo del lenguaje de programación utilizado. El formato universal y más tradicional de almacenamiento de la información es mediante archivos de tipo CSV. Un archivo de formato CSV es en realidad un archivo en formato de codificación de texto plano, donde cada línea del archivo representa una fila de datos, y los valores de cada columna están separados por un carácter delimitador, siendo generalmente este una coma. Es un formato popular debido a su simplicidad y fácil manipulación por programas y hojas de cálculo. Además, muchos sistemas y aplicaciones pueden exportar datos en formato CSV, lo que lo hace fácilmente intercambiable entre diferentes plataformas y herramientas. En particular un archivo de tipo CSV se visualiza de la siguiente manera:

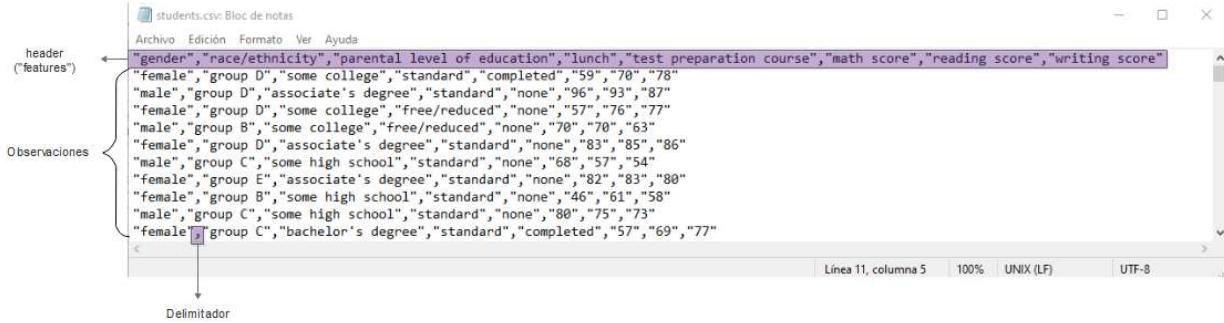


Figura 1.- Archivo de tipo CSV visualizado como texto plano

Existen otros formatos de archivos como el parquet, el cual es un formato de archivo de almacenamiento de datos de código abierto utilizado para almacenar datos tabulares en una estructura de columna, en lugar de una estructura de fila tradicional. Este formato está diseñado para ser eficiente en términos de almacenamiento y procesamiento, y permite una lectura y escritura más rápida de grandes conjuntos de datos. En particular, al estar los datos encriptados y codificados en este tipo de archivos, no es posible visualizarlos de manera tradicional, aunque el esquema de organización de la información sería lo más cercano a este:

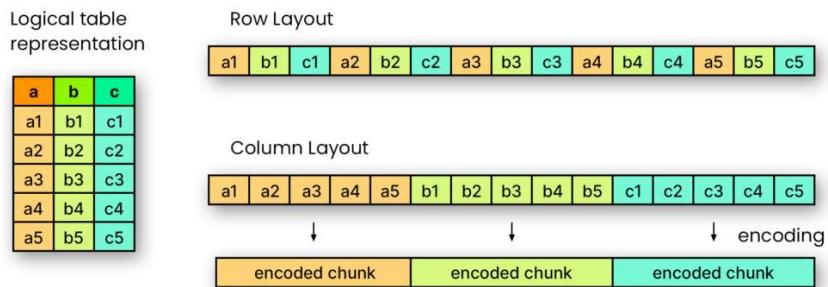


Figura 2.- Esquema tradicional de codificación tipo parquet

Finalmente, el uso y manejo de modelos de datos en MATLAB se realiza mediante la función "Datastore". Esta es una función que proporciona una interfaz rápida para acceder a grandes conjuntos de datos, como archivos de imágenes, archivos de audio o archivos de texto, sin cargarlos en la memoria. Permite la lectura de datos de manera eficiente y escalable, ya que lee y procesa los datos de forma incremental a medida que se necesitan, lo que permite trabajar con grandes conjuntos de datos sin tener que cargarlos por completo en la memoria. Además, la función "datastore" permite realizar operaciones de preprocesamiento y manipulación de datos, como filtrado y transformación de datos, de manera eficiente y fácil. Estas bases de datos tienen la siguiente estructura de información:

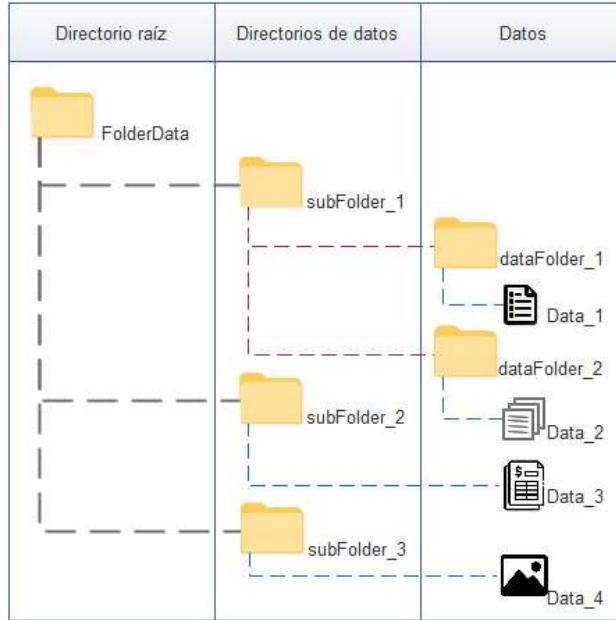


Figura 3.- Esquema de un modelo de datos

--- Pasos iniciales

```
%%
%Creacion de directorio de trabajo
%rootFolder = 'D:\TSISB_IA';
rootFolder ='C:\Users\puma_\Documents\TSISB_IA';
workingFolder = 'practica_2';
tempFolder = 'temp';
savePath = fullfile(rootFolder,workingFolder);
saveTempPath = fullfile(rootFolder,workingFolder,tempFolder);

prefix = ['\' 'students'];
sufix = '.csv';
newName = [prefix, sufix];

%Despues de correr esta celda una vez, crea un bloque de comentarios a partir de esta linea

if ~exist(savePath,'dir')
    [status, message, ~] = mkdir(savePath);
    if status == 0
        disp(message)
    end
end

if ~exist(saveTempPath,'dir')
    [status, message, ~] = mkdir(saveTempPath);
    if status == 0
        disp(message)
    end
end

%%
%Organizacion y copia de archivos
[fileName, pathFileName] = uigetfile('C:\','*.txt' );

if isequal(fileName,0)
    disp('Se canceló la búsqueda de archivos');
else
    disp(['El usuario seleccionó el archivo ', fullfile(pathFileName,fileName)]);
    [status, message, ~] = copyfile([pathFileName,fileName],[saveTempPath,newName]);
    if status == 0
        disp(message)
    else
        disp(['El cual se movió a la dirección ', fullfile(saveTempPath)]);
    end
end
```

El usuario seleccionó el archivo C:\Users\puma_\Documents\TSISB_IA\practica_2\students.csv
 El cual se movió a la dirección C:\Users\puma_\Documents\TSISB_IA\practica_2\temp

%}

***Modelo de datos mediante un solo archivo

Para lograr el manejo de grandes conjuntos de información, es necesario establecer un "pipeline" o flujo de trabajo de la función datastore. Este se muestra a continuación:

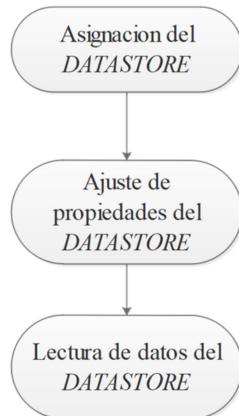


Figura 2.- Flujo de trabajo de un Datastore

La primer parte del pipeline del datastore incluye operaciones como:

- Detección de propiedades de los archivos del dataframe/modelo de datos
- Normalización de datos

```
*****
%En este ejercicio solo estaremos trabajando con un archivo
%contenido dentro del datastore
*****  
  
%Se determina la dirección de los archivos que integran el "Datastore"
archivoCSV = [saveTempPath, newName];  
  
%Visualización rápida de las primeras líneas del archivo de texto
%Como solo es un archivo se puede utilizar el comando dbtype
dbtype(archivoCSV,'1:5')
```

```
1 gender,race/ethnicity,parental level of education,lunch,test preparation course,math score,reading score,writing score
2 female,group D,some college,standard,completed,59,70,78
3 male,group D,associate's degree,standard,none,96,93,87
4 female,group D,some college,free/reduced,none,57,76,77
5 male,group B,some college,free/reduced,none,70,70,63
```

```
%Asignamos nuestro datastore dentro de MATLAB
%dsGeneral = datastore(archivoCSV)
%Si tuvieramos muchos archivos, se utilizaría el siguiente comando
%preview(dsGeneral)

%Desde la versión 2019, los nombres de las variables pueden incluir
%cualquier tipo de simbolos, además de no necesariamente comenzar solo
%con letras, por lo que matlab requiere el Flag "preserve" para considerar
%esta opción
%dsGeneral.VariableNamingRule='preserve';

%Selección del delimitador de texto
%dsGeneral.Delimiter=",";

%Modificación de los nombres utilizados para cada característica/feature
%values=dsGeneral.VariableNames;
%Modificación como si fuera un indexado de valores clásico
%newValues=["hola", "esta", "es", "una", "prueba", "de", "cambio", "de variables"];
%dsGeneral.VariableNames = newValues;
%preview(dsGeneral)
```

***Detección y modificación de propiedades de los datos contenidos en el archivo

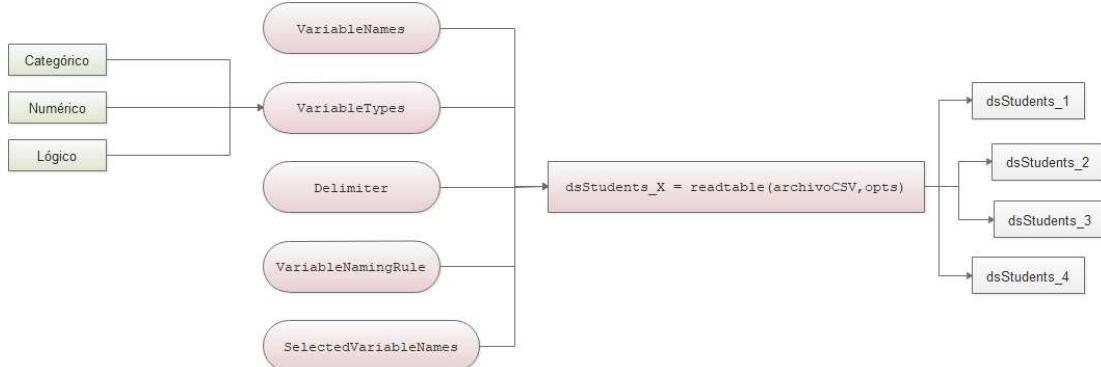


Figura 3.- Propiedades de un **datastore de tipo categorico-numérico**

```
%Visualizacion general de las opciones de importacion de los archivos
%dentro del datastore
opts = detectImportOptions(archivoCSV)
```

```
opts =
DelimitedTextImportOptions with properties:

Format Properties:
  Delimiter: {','}
  Whitespace: '\b\t '
  LineEnding: {'\n' '\r' '\r\n'}
  CommentStyle: {}
  ConsecutiveDelimitersRule: 'split'
  LeadingDelimitersRule: 'keep'
  TrailingDelimitersRule: 'ignore'
  EmptyLineRule: 'skip'
  Encoding: 'UTF-8'
```

Replacement Properties:

```
%Visualizacion de los nombres y tipo de variables de las columnas
disp([opts.VariableNames' opts.VariableTypes'])
```

```
{'gender'          }  {'char'  }
{'race_ethnicity' }  {'char'  }
{'parentalLevelOfEducation'}  {'char'  }
{'lunch'          }  {'char'  }
{'testPreparationCourse'}  {'char'  }
{'mathScore'       }  {'double'}
{'readingScore'   }  {'double'}
{'writingScore'   }  {'double'}
```

%Modificacion del tipo de variables

```
opts = setvaropts(opts,{'race_ethnicity','gender','parentalLevelOfEducation','lunch','testPreparationCourse'},'Type','categorical');
disp([opts.VariableNames' opts.VariableTypes'])
```

```
{'gender'          }  {'categorical'}
{'race_ethnicity' }  {'categorical'}
{'parentalLevelOfEducation'}  {'categorical'}
{'lunch'          }  {'categorical'}
{'testPreparationCourse'}  {'categorical'}
{'mathScore'       }  {'double'  }
{'readingScore'   }  {'double'  }
{'writingScore'   }  {'double'  }
```

```
%Desde la version 2019, los nombres de las variables pueden incluir
%cualquier tipo de simbolos, ademas de no necesariamente comenzar solo
%con letras, por lo que matlab requiere el Flag "preserve" para considerar
%esta opción
```

```
opts.VariableNamingRule='preserve';
```

```
%Selección del delimitador de texto
opts.Delimiter = ',';
```

%Modificacion de los nombres utilizados para cada caracteristica/feature

%Asignacion a una variable particular

```
values=opts.VariableNames;
```

%Modificacion como si fuera un indexado de valores clásico

```
newValues={'hola', 'esta', 'es', 'una', 'prueba', 'de', 'cambio', 'de variables'};
opts.VariableNames = newValues;
```

```
dsStudents_1 = readtable(archivoCSV,opts);
head(dsStudents_1)
```

hola	esta	es	una	prueba	de	cambio	de variables
female	group D	some college	standard	completed	59	70	78
male	group D	associate's degree	standard	none	96	93	87
female	group D	some college	free/reduced	none	57	76	77
male	group B	some college	free/reduced	none	70	70	63
female	group D	associate's degree	standard	none	83	85	86
male	group C	some high school	standard	none	68	57	54
female	group E	associate's degree	standard	none	82	83	80
female	group B	some high school	standard	none	46	61	58

%regresamos los valores originales

```
opts.VariableNames = values;
dsStudents_2 = readtable(archivoCSV,opts);
head(dsStudents_2)
```

gender	race_ethnicity	parentalLevelOfEducation	lunch	testPreparationCourse	mathScore	readingScore	writingScore
female	group D	some college	standard	completed	59	70	78
male	group D	associate's degree	standard	none	96	93	87
female	group D	some college	free/reduced	none	57	76	77
male	group B	some college	free/reduced	none	70	70	63
female	group D	associate's degree	standard	none	83	85	86
male	group C	some high school	standard	none	68	57	54
female	group E	associate's degree	standard	none	82	83	80
female	group B	some high school	standard	none	46	61	58

%Podemos elegir con que caracteristicas queremos trabajar nuestra base de datos

```
opts.SelectedVariableNames = {'gender','mathScore','readingScore', 'writingScore'};
dsStudents_3 = readtable(archivoCSV,opts);
head(dsStudents_3)
```

gender	mathScore	readingScore	writingScore
female	59	70	78
male	96	93	87
female	57	76	77
male	70	70	63
female	83	85	86
male	68	57	54
female	82	83	80
female	46	61	58

%y crear distintos tipos de tablas a partir del datastore original

```
opts.SelectedVariableNames = {'race_ethnicity','mathScore','readingScore', 'writingScore'};
dsStudents_4 = readtable(archivoCSV,opts);
head(dsStudents_4)
```

race_ethnicity	mathScore	readingScore	writingScore
group D	59	70	78
group D	96	93	87
group D	57	76	77
group B	70	70	63
group D	83	85	86
group C	68	57	54
group E	82	83	80
group B	46	61	58

***Visualizaciones exploratorias de datos

En este momento se han creado 4 tablas que contienen los siguientes datos

- Tabla 1: No la usaremos
- Tabla 2: Todas las categorías de datos
- Tabla 3: Datos categóricos de género y datos numéricos de las pruebas de matemáticas, lectura y escritura
- Tabla 4: Datos categóricos de raza étnica y datos numéricos de las pruebas de matemáticas, lectura y escritura

Vamos a realizar un análisis exploratorio de los datos, primero visualicemos un resumen de los datos categóricos contenidos en la tabla 2

Resumen de datos categóricos

```
genero = categorical(dsStudents_2.gender);
grpCatGenero = categories(genero)
```

```
grpCatGenero = 2x1 cell
'female'
'male'
```

```
numCatGenero = countcats(genero)
```

```
numCatGenero = 2x1
492
508
```

```
summary(genero)
```

```
female      492
male       508
```

```
etnia = categorical(dsStudents_2.race_ethnicity);
grpCatEtn = categories(etnia)
```

```
grpCatEtn = 5x1 cell
'group A'
'group B'
'group C'
'group D'
'group E'
```

```
numCatEtn = countcats(etnia)
```

```
numCatEtn = 5x1
79
198
323
257
143
```

```
%Vamos a reemplazar las categorias originales por otros nombres
nuevasEtnias = {'Latino',...
    'Afroamericano',...
    'Americano',...
    'Asiatico',...
    'Europeo'};
nuevaEtnia = renamecats(etnia,nuevasEtnias);
summary(etnia)
```

```
group A      79
group B     198
group C     323
group D     257
group E     143
```

```
summary(nuevaEtnia)
```

```
Latino      79
Afroamericano 198
Americano   323
Asiatico    257
Europeo     143
```

*** Ejercicio ***

Deberás mostrar cuantas categorias existen en:

- el nivel de educacion de los padres
- si contaron con un desayuno o no
- si hubo una preparacion previa a la prueba

Además deberas de cambiar el nombre de las categorias en la columna "preparacion de la prueba" de la siguiente manera

- completed - > terminado
- none -> no terminado

1) Mostrar cuantas categorias existen en el nivel de educacion de los padres.

```
nivelEducacion = categorical(dsStudents_2.parentalLevelOfEducation);
%grpCatGenero = categories(nivelEducacion)
```

```
%numCatGenero = countcats(nivelEducacion)
summary(nivelEducacion)
```

associate's degree	204
bachelor's degree	105
high school	215
master's degree	75
some college	224
some high school	177

2) Mostrar cuantas categorias existen en si contaron con un desayuno o no.

```
almuerzo = categorical(dsStudents_2.lunch);
summary(almuerzo)
```

free/reduced	340
standard	660

3) Mostrar cuantas categorias existen en si hubo una preparacion previa a la prueba

```
curso = categorical(dsStudents_2.testPreparationCourse);
summary(curso)
```

completed	344
none	656

Ademas deberas de cambiar el nombre de las categorias en la columna "preparacion de la prueba"

```
nuevasCategorias = {'terminado',...
    'no terminado'};
nuevasCategoria = renametcats(curso,nuevasCategorias);
summary(nuevasCategoria)
```

terminado	344
no terminado	656

Resumen de datos numericos

Ahora vamos a realizar un resumen de los datos numéricos, lo cual puede incluir encontrar la media de los datos de pruebas numericas

--Resumen general de una categoria

Los distintos tipos de operaciones numéricas que podemos realizar son las siguientes

Method	Description
"sum"	Sum
"mean"	Mean
"median"	Median
"mode"	Mode
"var"	Variance
"std"	Standard deviation
"min"	Minimum
"max"	Maximum
"range"	Maximum minus minimum
"nummissing"	Number of missing elements
"nnz"	Number of nonzero and non-NaN elements
"all"	All computations previously listed

Figura 4.- Operaciones numéricas de una tabla

```
%Primero veremos el score promedio dependiendo del genero, utilizando la
%tabla 3
genderMean = groupsummary(dsStudents_3, "gender", "mean")
```

genderMean = 2x5 table

	gender	GroupCount	mean_mathScore	mean_readingScore	mean_writingScore
1	female	492	64.7744	73.4736	73.4390
2	male	508	70.7500	67.3878	64.9764

```
%Podemos visualizar tambien el promedio dependiendo del grupo etnico
%utilizando la tabla 4
```

```
raceMean = groupsummary(dsStudents_4, "race_ethnicity", "mean")
```

raceMean = 5x5 table

	race_ethnicity	GroupCount	mean_mathScore	mean_readingScore	mean_writingScore
1	group A	79	65.6962	69.2025	67.8481
2	group B	198	64.0707	68.5303	66.7172
3	group C	323	65.5108	68.6099	66.8050
4	group D	257	68.8794	70.9300	71.0584

5	group	race_ethnicity	GroupCount	143	mean_mathScore	77.4266	mean_readingScore	76.6154	mean_writingScore	75.0350
---	-------	----------------	------------	-----	----------------	---------	-------------------	---------	-------------------	---------

%Podemos realizar operaciones por conjuntos específicos por ejemplo de la siguiente manera

```
gen = dsStudents_2.gender;
race = dsStudents_2.race_ethnicity;
math = dsStudents_2.mathScore;
reading = dsStudents_2.readingScore;
writing = dsStudents_2.writingScore;
auxTable = table(gen,race,math, reading, writing)
```

auxTable = 1000x5 table

	gen	race	math	reading	writing
1	female	group D	59	70	78
2	male	group D	96	93	87
3	female	group D	57	76	77
4	male	group B	70	70	63
5	female	group D	83	85	86
6	male	group C	68	57	54
7	female	group E	82	83	80
8	female	group B	46	61	58
9	male	group C	80	75	73
10	female	group C	57	69	77
11	male	group B	74	69	69
12	male	group B	53	50	49
13	male	group B	76	74	76
14	male	group A	70	73	70
15	male	group C	55	54	52
16	male	group E	56	46	43
17	female	group C	35	47	41
18	female	group C	87	92	81
19	female	group E	80	82	85
20	female	group D	65	71	74
21	male	group C	66	66	62
22	female	group D	67	71	76
23	female	group B	70	71	71
24	male	group E	89	88	86
25	male	group D	99	85	88
26	male	group B	74	83	72
27	male	group D	58	52	51
28	male	group D	70	66	59
29	female	group E	80	79	71
30	male	group D	90	87	86
31	female	group B	80	81	85
32	female	group D	68	76	79
33	female	group B	69	78	75
34	female	group D	32	35	37
35	male	group D	82	82	82
36	female	group A	57	53	54
37	female	group E	69	74	75
38	male	group D	68	66	72
39	male	group C	74	85	87
40	male	group E	89	85	78
41	male	group C	46	46	48
42	male	group C	76	82	77
43	male	group B	86	82	72
44	male	group D	69	73	67
45	female	group B	52	56	54
46	male	group C	63	71	65
47	male	group A	96	82	90
48	male	group C	80	76	68

	gen	race	math	reading	writing
49	female	group E	59	52	56
50	male	group D	80	77	80
51	female	group E	65	77	74
52	female	group E	74	83	84
53	male	group D	90	93	84
54	female	group B	69	72	72
55	male	group C	69	67	63
56	female	group C	62	64	61
57	female	group D	67	75	80
58	female	group E	89	93	93
59	female	group C	79	86	78
60	male	group C	67	66	66
61	male	group D	82	74	75
62	male	group C	63	69	63
63	female	group D	71	83	80
64	female	group C	55	68	73
65	female	group B	61	74	71
66	female	group B	35	34	36
67	male	group C	75	77	66
68	female	group B	73	91	88
69	female	group C	56	62	57
70	male	group D	80	70	73
71	male	group C	83	81	78
72	female	group D	64	82	80
73	female	group C	23	33	33
74	female	group D	41	58	59
75	male	group E	61	49	52
76	male	group B	63	46	46
77	male	group B	84	91	89
78	male	group C	55	61	59
79	male	group A	85	75	74
80	male	group B	65	61	57
81	male	group C	88	80	81
82	male	group D	91	93	95
83	female	group A	51	46	42
84	male	group C	73	77	76
85	female	group D	73	89	89
86	male	group D	100	97	91
87	female	group D	48	68	68
88	male	group E	98	79	85
89	male	group B	68	65	60
90	male	group C	64	62	58
91	male	group C	72	67	61
92	female	group C	63	74	75
93	male	group C	43	51	38
94	male	group D	80	75	74
95	female	group C	71	88	83
96	female	group C	91	96	97
97	female	group D	68	84	87
98	female	group B	73	80	78
99	female	group B	75	90	95
100	male	group C	83	62	64

:

```
gender_raceMean = groupsummary(auxTable, [ "gen", "race"], "mean")
```

```
gender_raceMean = 10×6 table
```

	gen	race	GroupCount	mean_math	mean_reading	mean_writing
--	-----	------	------------	-----------	--------------	--------------

1	female	group A	GroupCount	41	mean_math	72.9512	mean_reading	72.683	mean_writing	77.073
2	female	group B		112	62.2679		72.3839		71.8929	
3	female	group C		151	63.0199		71.9934		71.4768	
4	female	group D		118	63.8390		72.8729		74.1610	
5	female	group E		70	75.2143		80.1286		79.9429	
6	male	group A		38	68.6579		65.8947		63.6842	
7	male	group B		86	66.4186		63.5116		59.9767	
8	male	group C		172	67.6977		65.6395		62.7035	
9	male	group D		139	73.1583		69.2806		68.4245	
10	male	group E		73	79.5479		73.2466		70.3288	

*** Ejercicio ***

Realiza las siguientes operaciones

- Cual es la media del grupo genero vs nivel de estudios de los padres
- Cual es el promedio de la raza etnica vs si se preparó o no para la prueba
- Cual es el promedio del genero vs si se preparó o no para la prueba

1) Cual es la media del grupo genero vs nivel de estudios de los padres

```
% gen = dsStudents_2.gender;
levelOfEducation = dsStudents_2.parentalLevelOfEducation;
% math = dsStudents_2.mathScore;
% reading = dsStudents_2.readingScore;
% writing = dsStudents_2.writingScore;
auxTable1 = table(gen,levelOfEducation,math, reading, writing);
genderMeanvsLevelOfEducation = groupsummary(auxTable1,[ "gen","levelOfEducation"], "median")
```

genderMeanvsLevelOfEducation = 12x6 table

	gen	levelOfEducation	GroupCount	median_math	median_reading	median_writing
1	female	associate's degree	101	69	76	77
2	female	bachelor's degree	52	66	78	78
3	female	high school	116	61	71	69
4	female	master's degree	33	68	75	77
5	female	some college	106	66	74	75
6	female	some high school	84	62	72.5000	72
7	male	associate's degree	103	74	70	67
8	male	bachelor's degree	53	71	67	67
9	male	high school	99	72	68	65
10	male	master's degree	42	75	70	71.5000
11	male	some college	118	72.5000	68	67
12	male	some high school	93	67	62	59

2) Cual es el promedio de la raza etnica vs si se preparó o no para la prueba

```
preparationCourse= dsStudents_2.testPreparationCourse;

auxTable2 = table(race,preparationCourse,math, reading, writing);
genderMeanvsLevelOfEducation = groupsummary(auxTable2,[ "race","preparationCourse"], "mean")
```

genderMeanvsLevelOfEducation = 10x6 table

	race	preparationCourse	GroupCount	mean_math	mean_reading	mean_writing
1	group A	completed	31	70.0645	72.4516	73.2258
2	group A	none	48	62.8750	67.1042	64.3750
3	group B	completed	66	66.8636	73.1818	73.8939
4	group B	none	132	62.6742	66.2045	63.1288
5	group C	completed	101	66.7822	72.6139	73.4257
6	group C	none	222	64.9324	66.7883	63.7928
7	group D	completed	97	71.1753	75.5258	77.4124
8	group D	none	160	67.4875	68.1437	67.2062
9	group E	completed	49	80.8367	81.0204	81.7551
10	group E	none	94	75.6489	74.3191	71.5319

3) Cual es el promedio del genero vs si se preparó o no para la prueba

```
auxTable3 = table(gen,preparationCourse,math, reading, writing);
```

```
genderMeanvsLevelOfEducation = groupsummary(auxTable3, ["gen","preparationCourse"], "mean")
```

genderMeanvsLevelOfEducation = 4x6 table

	gen	preparationCourse	GroupCount	mean_math	mean_reading	mean_writing
1	female	completed	177	67.5650	77.7288	79.8644
2	female	none	315	63.2063	71.0825	69.8286
3	male	completed	167	73.2695	71.5449	71.5090
4	male	none	341	69.5161	65.3519	61.7771

Modificación de las columnas en modelos de datos

Podemos realizar ligeras modificaciones a las columnas (características de nuestro modelo de datos) realizando las siguientes operaciones

- Crear tablas a partir de datos específicos
- Agregar una columna de datos
- Mover columnas de datos
- Remover columnas de datos

```
%Creación de tablas a partir de datos originales
```

```
auxTable_1 = table(gen, math, reading);  
head(auxTable_1)
```

gen	math	reading
female	59	70
male	96	93
female	57	76
male	70	70
female	83	85
male	68	57
female	82	83
female	46	61

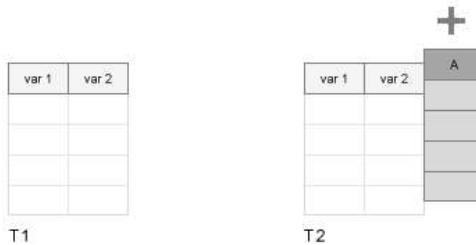


Figura 5.- Agregado de características de un modelo de datos

```
%Agregado de valores a tablas ya existentes  
auxTable_2 = addvars(auxTable_1,writing);  
head(auxTable_2)
```

gen	math	reading	writing
female	59	70	78
male	96	93	87
female	57	76	77
male	70	70	63
female	83	85	86
male	68	57	54
female	82	83	80
female	46	61	58

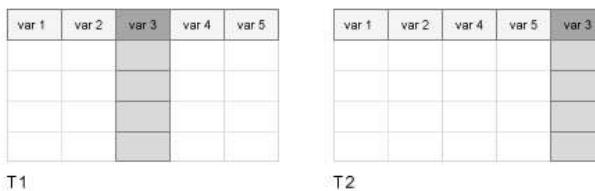


Figura 6.- Reordenamiento de caracteristicas de un modelo de datos

```
%Reordenamiento de caracteristicas de la tabla
auxTable_3 = movevars(auxTable_2,'reading','after','writing');
head(auxTable_3)
```

gen	math	writing	reading
female	59	78	70
male	96	87	93
female	57	77	76
male	70	63	70
female	83	86	85
male	68	54	57
female	82	80	83
female	46	58	61

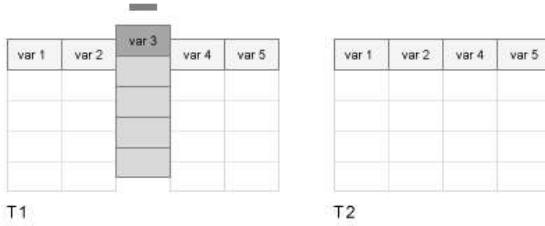


Figura 7.- Eliminacion de caracteristicas de un modelo de datos

```
%Eliminacion de caracteristicas de la tabla
auxTable_4 = removevars(auxTable_2,['reading']);
head(auxTable_4)
```

gen	math	writing
female	59	78
male	96	87
female	57	77
male	70	63
female	83	86
male	68	54
female	82	80
female	46	58

###Filtrado de informacion basado en reglas de valores

```
%Podemos tambien filtrar los datos por grupos de informacion
%filtrado de la informacion de la tabla auxTable_2 para resultados de datos
%de prueba Matematicas mayores a 60 pero menores a 80
auxTable = table(gen,math,reading,writing);
head(auxTable)
```

gen	math	reading	writing
female	59	70	78
male	96	93	87
female	57	76	77
male	70	70	63
female	83	85	86
male	68	57	54
female	82	83	80
female	46	61	58

```
mathBetween_60_80 = groupfilter(auxTable,"math",@(x) min(x) >= 60 && max(x) <= 80, "math");
head(mathBetween_60_80)
```

gen	math	reading	writing
male	70	70	63
male	68	57	54
male	80	75	73
male	74	69	69
male	76	74	76

```

male    70    73    70
female  80    82    85
female  65    71    74

```

```
auxMean_Table = groupsummary(mathBetween_60_80, "gen", "max")
```

```
auxMean_Table = 2x5 table
```

	gen	GroupCount	max_math	max_reading	max_writing
1	female	243	80	97	99
2	male	245	80	88	87

*** Ejercicio ***

Realiza las siguientes operaciones

- Crea una tabla que contenga las siguientes categorías en este orden: Preparacion de la prueba, genero, Resultado de matemáticas, Resultado de lectura, Resultado de escritura
- Calcula el promedio de los tres valores de las pruebas y coloca el valor al final de las columnas
- Agrupa los resultados por promedio y clasificalos para valores mayores o iguales a 60 pero menores o iguales a 80. ¿Cuantos datos quedaron en total?
- Cual es el promedio general del género masculino que si se preparo para la prueba
- Cual es el promedio general del género femenino que no se preparo para la prueba
- De acuerdo con los datos, ¿Existe una relación directa entre prepararse para la prueba y no prepararse para la prueba?

1) Crea una tabla que contenga las siguientes categorías en este orden: Preparacion de la prueba, genero, Resultado de matemáticas, Resultado de lectura, Resultado de escritura.

```
Ej3auxTable1 = table(preparationCourse,gen,math,reading,writing);
head(Ej3auxTable1)
```

preparationCourse	gen	math	reading	writing
completed	female	59	70	78
none	male	96	93	87
none	female	57	76	77
none	male	70	70	63
none	female	83	85	86
none	male	68	57	54
none	female	82	83	80
none	female	46	61	58

2) Calcula el promedio de los tres valores de las pruebas y coloca el valor al final de las columnas.

```
Ej3auxTable2 = table(preparationCourse,gen,math,reading,writing);
head(Ej3auxTable2);
```

preparationCourse	gen	math	reading	writing
completed	female	59	70	78
none	male	96	93	87
none	female	57	76	77
none	male	70	70	63
none	female	83	85	86
none	male	68	57	54
none	female	82	83	80
none	female	46	61	58

```
promedio = mean([Ej3auxTable2.math, Ej3auxTable2.reading, Ej3auxTable2.writing], 2)
```

```

promedio = 1000×1
69.0000
92.0000
70.0000
67.6667
84.6667
59.6667
81.6667
55.0000
76.0000
67.6667

```

%Agregado de valores a tablas ya existentes

```
Ej3auxTable2_1 = addvars(Ej3auxTable2,promedio);
head(Ej3auxTable2_1)
```

preparationCourse	gen	math	reading	writing	promedio
completed	female	59	70	78	69
none	male	96	93	87	92
none	female	57	76	77	70
none	male	70	70	63	67.667
none	female	83	85	86	84.667
none	male	68	57	54	59.667
none	female	82	83	80	81.667
none	female	46	61	58	55

3) Agrupa los resultados por promedio y clasificalos para valores mayores o iguales a 60 pero menores o iguales a 80. ¿Cuantos datos quedaron en total?

%Reordenamiento de columna promedio de la tabla

```
Ej3auxTable3_1 = movevars(Ej3auxTable2_1,'promedio','Before','preparationCourse');
head(Ej3auxTable3_1)
```

promedio	preparationCourse	gen	math	reading	writing
69	completed	female	59	70	78
92	none	male	96	93	87
70	none	female	57	76	77
67.667	none	male	70	70	63
84.667	none	female	83	85	86
59.667	none	male	68	57	54
81.667	none	female	82	83	80
55	none	female	46	61	58

%Agrupar por promedio

```
avgProm = groupsummary(Ej3auxTable3_1,[ "promedio","preparationCourse","gen"], "mean");
head(avgProm);
```

promedio	preparationCourse	gen	GroupCount	mean_math	mean_reading	mean_writing
20	none	male	1	20	25	15
25.667	none	female	1	21	30	26
25.667	none	male	1	31	27	19
26	none	male	1	28	27	23
29.333	none	female	1	28	32	28
29.667	none	female	1	23	33	33
30	none	female	1	25	33	32
31.333	none	female	1	15	39	40

%Valores mayores o iguales a 60 pero menores o iguales a 80

```
mathBetween_60_80 = groupfilter(avgProm,"promedio",@(x) min(x) >= 60 && max(x) <= 80, "promedio");
head(mathBetween_60_80);
```

promedio	preparationCourse	gen	GroupCount	mean_math	mean_reading	mean_writing
60	none	male	3	65	59	56
60.333	completed	female	1	44	69	68
60.333	completed	male	1	64	62	55
60.333	none	female	1	52	66	63
60.333	none	male	3	68.333	56	56.667
60.667	completed	female	2	53.5	64	64.5
60.667	completed	male	1	65	60	57
60.667	none	female	3	53.667	62.333	66

```
suma = sum([mathBetween_60_80.GroupCount], 1);
disp(['En total quedaron: ', num2str(suma), ' datos']);
```

En total quedaron: 511 datos

4) Cual es el promedio general del género masculino que si se preparo para la prueba.

%Reordenamiento de columna promedio de la tabla para el ej 4

```
Ej3auxTable3_4 = movevars(Ej3auxTable2_1,'promedio','after','gen');
head(Ej3auxTable3_4);
```

preparationCourse	gen	promedio	math	reading	writing
completed	female	69	59	70	78
none	male	92	96	93	87
none	female	70	57	76	77
none	male	67.667	70	70	63

none	female	84.667	83	85	86
none	male	59.667	68	57	54
none	female	81.667	82	83	80
none	female	55	46	61	58

%Agrupar por Curso de Preparación (principalmente)

```
avgProm4 = groupsummary(Ej3auxTable3_4, ["preparationCourse", "gen", "promedio"], "mean");
```

%Filtro //Basado en una respuesta generada por ChatGPT.

```
filtro = (avgProm4.gen == 'male') & (avgProm4.preparationCourse == 'completed');
```

```
tabla_filtrada = avgProm4(filtro, :);
```

%Promedio de los estudiantes masculinos que se prepararon para la prueba

```
promedio_general_masculino_preparados = mean(tabla_filtrada.promedio);
```

```
disp(['Promedio general de estudiantes masculinos que se prepararon: ', num2str(promedio_general_masculino_preparados)]);
```

Promedio general de estudiantes masculinos que se prepararon: 72.2145

5) Cual es el promedio general del género femenino que no se preparo para la prueba.

%Reordenamiento de columna promedio de la tabla para el ej 5

```
Ej3auxTable3_5 = movevars(Ej3auxTable2_1, 'promedio', 'after', 'gen');
```

```
head(Ej3auxTable3_5);
```

preparationCourse	gen	promedio	math	reading	writing
completed	female	69	59	70	78
none	male	92	96	93	87
none	female	70	57	76	77
none	male	67.667	70	70	63
none	female	84.667	83	85	86
none	male	59.667	68	57	54
none	female	81.667	82	83	80
none	female	55	46	61	58

%Agrupar por Curso de Preparación (principalmente)

```
avgProm4 = groupsummary(Ej3auxTable3_5, ["preparationCourse", "gen", "promedio"], "mean");
```

%Filtro //Basado en una respuesta generada por ChatGPT.

```
filtro = (avgProm4.gen == 'female') & (avgProm4.preparationCourse == 'none');
```

```
tabla_filtrada = avgProm4(filtro, :);
```

%Promedio de los estudiantes femeninos que no se prepararon

```
promedio_general_masculino_preparados = mean(tabla_filtrada.promedio);
```

```
disp(['Promedio general de estudiantes femeninos que no se prepararon: ', num2str(promedio_general_masculino_preparados)]);
```

Promedio general de estudiantes femeninos que no se prepararon: 65.8833

6) De acuerdo con los datos, ¿Existe una relación directa entre prepararse para la prueba y no prepararse para la prueba?

Si existe una relación directamente proporcional. Los primeros 5 estudiantes con la mayor calificación tomaron el curso de preparación para la prueba y en los promedio más bajos estan aquellos que no tomaron el curso de preparación. En los resultados entre 60 y 80 de promedio, hay una diferencia aproximada de 6 puntos entre aquellos que no tomaron el curso contra aquellos que si lo tomaron y salieron un poco más por encima.

Discretizacion de datos en categorias

%Vamos a establecer los siguientes valores de calificaciones en el formato

%americano de calificaciones donde

%Valor | calificacion

%0 59	F
%60 69	D
%70 79	C
%80 89	B
%90 99	A
%100	A+

%Primero estableceremos nuestras nuevas categorias de datos

%etiquetas

```
calAme = {'F', 'D', 'C', 'B', 'A', 'A+'};
```

%Limites de calificaciones

```
limit = [0 59 69 79 89 99 100];
```

auxVal = dsStudents_3.mathScore; %Solo para guardar el valor

```
dsStudents_3.mathScore = discretize(dsStudents_3.mathScore, limit, 'categorical', calAme);
```

```
summary(dsStudents_3.mathScore)
```

F	256
D	256
C	219
B	186
A	70
A+	13

```
head(dsStudents_3)
```

gender	mathScore	readingScore	writingScore
female	D	70	78
male	A	93	87
female	F	76	77
male	C	70	63
female	B	85	86
male	D	57	54
female	B	83	80
female	F	61	58

```
dsStudents_3.mathScore = auxVal; %Regreso el valor original para no entorpecer la BD  
head(dsStudents_3)
```

gender	mathScore	readingScore	writingScore
female	59	70	78
male	96	93	87
female	57	76	77
male	70	70	63
female	83	85	86
male	68	57	54
female	82	83	80
female	46	61	58