## Lab-1

District Him was

(finish sported ) to sky

Many Sail V expire

steller palin It. smil

MEDO ANTH

Write python code for fall considering Ilm Emmily of ) Line

- OTo load . cer file its datafrane import pardas as pd filerane = "/content/housing. car" of = pd. read - esv (filerane)
- To display if of all columns print ("Dataset Info: ") print (df. info ())

## Dataset Info:

#	Column No	Null court	Dtype	ent parin 1)	In grand
0	longitude	20640	fleat 64		hoted - betat
t	latitude	20640	_ n		
2	housing_median-age	20140			
3	total-soons	20640			- itestimes
- h	total_sedrams	20640	datastal. Was	od in the	dos Darly C.
5	population				all allow
6 7	households				300,000
8	median-income median-house-nolue	- · · ·	1/2016	Her Sules pr	<b>以</b>
a		(júš	object		
7.4	ocean-proximity			were hinethy	
		T-4019	The Children and	+ + + + + + + + + + + + + + + + + + + +	4 7 1

3 To dieplay statistical info of all numerical columns print ("In Statistical sunnary of Numerical column : ") print (df. describe (1)

the file colors - s was been a solver to pool of

- (4) To display court of wique labels for "ocea proximity" column print (" \n Unique Value counts for Ocean Proximity": ") BANDA ENTAL O pent (of ['ocean\_proximity "], ralue\_courts())
- (5) To display which columns is a dataset have missing values court greater than yere

```
miseing- natures = df. ismill (). sum ()
missing-columns = missing_natures [missing-natures >0]
                                          Appl who work how
peint ("In Columns with missing values: ")
                            annighted the talkforms
print (missing-columns)
                               "footest Thousand cin
off:
  Unique Value Courts fore Ocean Rescinity
  ocean- prominity
  CIH OCEAN
                9136
                                and of all relians
   INLAND
               6551
                                       Crops but
  NEAR OCEAN
               2658
               2290
  NEAR BAY
                 5
  ISLAND
Columns with Missing Values: may to when the world
                         2064 o Bleath
 total-bedramy: 207
              Chief of the lattered
auestions:
i) which columns in the datasets had mixing natures? How did you
hardle then?
Ars: Mining rathe columns:
   Adult income dataset - Age, Salary
    Diabetes dataset - Glucose, BMI
 Hardling approach the Dallamer Ung of los testors
 Adult inome dataset propose age pused median since its less
         Lo for solony - wes mean as salaries typically
    follow rosmal distribution
 Diabetes dataset - Columose - used median since glucose lend
now have outliers
               - BMI - pued near assuming normal
                            distribution.
```

2) which categorical columns did you identify in the dataset?
How did you encode them?

## Any - Adult Ireane dotaget:

categorical valuery: Gendes -> 0 riginal encoding to

## Diabetes dotaset:

cotegorical columns: Gender - s desgrial enoding

3) What is the difference bln Min-Man realing and standardization? When yould you use one-over another?

Ans. Min Max Scaling:

- X = x - Xmin Xnex - Xmin

- scales ralues bln 0 &1

- senitive to authors

Standardization;

- X = X - M

mean = 0 and ratione -1

Tand In Thomas

When date is not normally distributed and has known bounds, min-more scaling is used.

When data follows a normal distribution, standardization is used

Jelo B 3/25