

## DSBDA Lab Assignment-02- Pata wangling-II

| PICT, PUNE |  |
|------------|--|
|            | Title: Data wrongling  |
| 4414       | AMAZADA MANTANANANANANANANANANANANANANANANANANAN   |
|            | Problem Statement:   |
|            | Create as anademic performance datasets of students +  |
|            | perform the following operations using python:   |
|            | & scan all variable for missing values & inconsistencies. It   |
|            | there are missing values and @ inconsistences  |
|            | use any of the suitable techniques to deal with  |
| 24/        | themse the state of the state o |
| 1 700      | 2) San all numeric variable for outliers. If there are   |
|            | we any of the suitable techniques to deal with   |
|            | 31 And data transformations or at locate once of the   |
|            | 3> Apply data transformations on at least one of the variables. The purpose of this transformation   |
| 2 N Ustus  | should be one of the following reasons; to   |
|            | chage the scale for botter understanding of the  |
|            | variable to convert a non-linear relation into a   |
| Ten        | dinear one of to decrease the skewness tronvert  |
|            | the dishibution into a normal dishibution.   |
|            | Reason & downent your approach properly  |
|            | 1) Discounting volues  |
|            | Objective:   |
|            | 17 And null values & handle them.  |
| Own        | y Scar data for outliers of handle them.   |
|            | in Apply data transformation on data.  |
|            | Outrome: Mal Dellar paralle and A.   |
| parken     | Student will be able to handle null value, outliers  |
|            | & baneform data  |
| balance    | Ain sublance white the second and a  |
|            |  |



S/w & Ww requirements.

operating system: windows-10-Home, 8GBRAM, 64-bit

programming language fythom 3.8

Programming Tool: Jupyter Notebook & Pandas.

Theory.

Pata Wengling: I data muging:
The autivity of taking input data frame its
original state to a format where we can
perform meaningful analysis on it is called
data wangling.

Missing Values:

In order to check whether ene dataset contains mussing values, we use function wince!

Syntax: df. isnull()

will return if cell of dataset is Nan 60 not.

Technique to handle missing values:

17 Prop missing values

-dropnacj function med

-syntax: df. dropna (axis=0) - delete row.

df. dropna (axis=1) - delete volumn

A good stratergy when dealing with missing values involver their replacement with another values

- un for numerical value replace with mean media



| PICT, PUNE |   |
|------------|---|
|            | for categorical values replace with most frequent   |
|            | -fynction used Knduaci  |
| and a last | 3) keep missing value as if is  |
|            | Tep mystry vacate as 15 G   |
|            | entliers:   |
|            | 17 An ontles us a data point that differs   |
|            | Significantly from other observations.  |
|            | 2> Visualizing outliers:  |
|            | -A way to visually e se the outlier is the  |
|            | - Observations shown outside the whisters are   |
|            | outliers.   |
|            |   |
|            | Handle Outliers:  |
|            | y Interquatile large (IOK) method   |
|            | 2> 2-score method   |
|            |   |
|            | IOR: Method:  |
|            | - Data points that falls outside of 1.5 times of an interquartile range above the 3rd quartile (93) & below the 1st quartile (91) are outliers. |
|            | interquaitile raige above the 319 quartile (93)   |
|            | Leapon the 1st quartile (31) are originets.   |
|            | Transformation:   |
|            | Standardization:  |
|            | - If doesn't have any fixed min. I was value.   |
|            | Here, the values of all the columns are sealed  |
|            | in such a way that they all have wear =0 &  |
|            | In such a way that they all have mean =0 & standard demation st. This scaling technique   |
|            |   |



| PICT, PUNE |  |
|------------|--|
| 1000       | is preferred it outliers present in the dataset.   |
|            | Conduston:   |
|            | Through this assignment ne performed data wrongling such as hardling musting value outlier & transform data by Standardization.  |
|            | anothernolds would plan throughout   |
|            | medicine medicine  |
|            | A could be set an applying of your A-  |
|            | - Applications Smiles and the Markets of   |
|            |  |
|            | Line Alexander Countries and C |
|            | boulen (227) gont alternations   |
|            | boultan arox (2)   |
|            |  |
|            | the state of the s |
|            | Hour be at and and althoughts  |
|            | village of vival although the are could a  |
|            |  |
|            | Landon of the state of the stat |
|            | The second of th |
|            | where have been break him have the   |
|            | aca rampoles at the de parties att and   |
|            | and the wall that their a ward or  |
|            | Mint parint will in a street hands   |