{

"nbformat": 4,

"nbformat\_minor": 0,

"metadata": {

"colab": {

"provenance": []

},

"kernelspec": {

"name": "python3",

"display\_name": "Python 3"

},

"language\_info": {

"name": "python"

}

},

"cells": [

{

"cell\_type": "markdown",

"source": [

"1.Handling Missing Values"

],

"metadata": {

"id": "ySDvuk4g986j"

}

},

{

"cell\_type": "code",

"source": [

"import pandas as pd\n",

"import numpy as np"

],

"metadata": {

"id": "0yERbUSk-E8\_"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"import sklearn import preprocessing\n",

"import matplotlib.pyplot as plt\n"

],

"metadata": {

"id": "zxjpt9pz-Nl2"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"from sklearn import preprocessing\n",

"from sklearn import model\_selection\n",

"from sklearn import metrics\n",

"from sklearn import Linear\_model\n",

"from sklearn import ensemble\n",

"from sklearn import tree\n",

"from sklearn import svm\n",

"import xgboost"

],

"metadata": {

"id": "weJjLbzL-uic"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data = pd.read\_csv(\"E:\\IBM\_project\\weatherAUS.csv\")"

],

"metadata": {

"id": "6YDz6BRK\_cLr"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data.head()"

],

"metadata": {

"id": "dhQg9Bmz\_zI1"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data.describe()"

],

"metadata": {

"id": "hHJ68lDG\_9Dq"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data.info()"

],

"metadata": {

"id": "qgt96fzmAFPn"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data.shape#gives the dimension of the data"

],

"metadata": {

"id": "zQ-CZtaPAKYi"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data.isnull().sum()"

],

"metadata": {

"id": "PMDoVrE-AUW7"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"import missingno as msno\n",

"msno.matrix(data,color=(0,55,0,255,0,255),fontsize=16)"

],

"metadata": {

"id": "ohCVqzyIAjwc"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"import pandas as pd\n",

"import numpy as np"

],

"metadata": {

"id": "pUfQSLRDBFBu"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"import seaborn as sns\n",

"import matplotlib.pyplot as plt"

],

"metadata": {

"id": "T\_1AGWmgBM2V"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"from sklearn import preprocessing\n",

"from sklearn import model\_selection\n",

"from sklearn import metrics\n",

"from sklearn import Linear\_model\n",

"from sklearn import ensemble\n",

"from sklearn import tree\n",

"from sklearn import svm\n",

"import xgboost"

],

"metadata": {

"id": "bH-eIusGBZZw"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data = pd.read\_csv(\"E:\\IBM\_project\\weatherAUS.CSV\")"

],

"metadata": {

"id": "fX0Kw3lhCH2x"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data\_cat = data [['RainToday','WindGustDir','WindDir9am','WindDir3pm']]\n",

"data.drop(columns=['Evaporation','Sunshine','cloud9am,'Cloud3pm'],axis=1,inplace=True)\n",

"data.drop(column=['RainToday','WindGustDir','windDir9am','windDir3pm'],axis=1,inplace=True)"

],

"metadata": {

"id": "XVwg\_SlxCeNK"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data['MinTemp']fillna(data['MinTemp'].mean(),inplace=True)\n",

"data['maxTemp']fillna(data['maxTemp'].mean(),inplace=True\n",

"data['Rainfall']fillna(data['Rainfall'].mean(),inplace=True\n",

"data['windGustSpeed']fillna(data['windGustSpeed'].mean(),inplace=True\n",

"data['windSpeed9am']fillna(data['windSpeed9am'].mean(),inplace=True\n",

"data['windSpeed3pm']fillna(data['windSpeed3pm'].mean(),inplace=True\n",

"data['Humidity9am']fillna(data['Humidity9am'].mean(),inplace=True\n",

"data['Humidity3pm']fillna(data['Humidity3pm'].mean(),inplace=True\n",

"data['pressure9am']fllna(data['pressure9am'].mean(),inplace=True\n",

"data['pressure3pm']fillna(data['pressure3pm'].mean(),inplace=True \n",

"data['Teamp9am']fillna(data['Temp9am'].mean(),inplace=True\n",

"data['Temp3pm']fillna(data['Teamp3pm'].mean(),inplace=True\n",

" \n",

" "

],

"metadata": {

"id": "nTogXbRoGbD0"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"#filling the missing data of numeric variables with mean\n",

"cat\_names=data\_cat.columns"

],

"metadata": {

"id": "MZ5Zf4D1LPHi"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"import numpy as np\n",

"from sklearn.impute import simpleImputer\n",

"imp\_mode=SimpleImputer(missing\_values=np.nan,strategy='most\_frequent')"

],

"metadata": {

"id": "LjHknxGNLfPC"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data\_cat=imp\_mode.fit\_transform(data\_cat)"

],

"metadata": {

"id": "zAFELsuJL-zR"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data\_cat=pd.DataFrame(data\_cat,columns=cat\_names)"

],

"metadata": {

"id": "d696G4dJMKq-"

},

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"data=pd.concat([data,data\_cat],axis=1)"

],

"metadata": {

"id": "Jj56c6kzMbcW"

},

"execution\_count": null,

"outputs": []

}

]

}