{

"nbformat": 4,

"nbformat\_minor": 0,

"metadata": {

"colab": {

"provenance": []

},

"kernelspec": {

"name": "python3",

"display\_name": "Python 3"

},

"language\_info": {

"name": "python"

}

},

"cells": [

{

"cell\_type": "markdown",

"source": [

"# \*\*Data\*\* \*\*Visualization\*\* \*\*and\*\* \*\*Pre\*\*-\*\*processing\*\*"

],

"metadata": {

"id": "0JlJ3PLHWPuv"

}

},

{

"cell\_type": "markdown",

"source": [

"1. \*\*Download\*\* \*\*the\*\* \*\*dataset\*\*: \*\*Dataset\*\* \n",

"2. \*\*Load\*\* \*\*the\*\* \*\*dataset\*\*"

],

"metadata": {

"id": "whIAaLT2OVhO"

}

},

{

"cell\_type": "code",

"source": [

"import pandas as pd\n",

"import numpy as np\n",

"import matplotlib.pyplot as plt\n",

"import seaborn as sns\n",

"data=pd.read\_csv('/content/chrun\_modelling.csv')"

],

"metadata": {

"id": "Xj6M4THRoM\_M"

},

"execution\_count": 42,

"outputs": []

},

{

"cell\_type": "markdown",

"source": [

" 3.\*\*Perform\*\* \*\*Below\*\* \*\*Visualizations\*\*"

],

"metadata": {

"id": "dGv24pfYOURI"

}

},

{

"cell\_type": "code",

"source": [

"sns.kdeplot(data['CreditScore'], shade=True)\n",

"sns.kdeplot(data['Age'], shade=True)\n",

"sns.kdeplot(data['Balance'], shade=True)\n",

"sns.kdeplot(data['EstimatedSalary'], shade=True)\n",

"sns.kdeplot(data['Tenure'], shade=True)"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 296

},

"id": "3n0g\_mAX-9vg",

"outputId": "29c8c4a9-176e-4653-aa5c-c984f7b117e5"

},

"execution\_count": 11,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"<matplotlib.axes.\_subplots.AxesSubplot at 0x7fc73ba72d90>"

]

},

"metadata": {},

"execution\_count": 11

},

{

"output\_type": "display\_data",

"data": {

"text/plain": [

"<Figure size 432x288 with 1 Axes>"

],

"image/png": "\n"

},

"metadata": {

"needs\_background": "light"

}

}

]

},

{

"cell\_type": "code",

"source": [

"from sklearn.preprocessing import StandardScaler\n",

"stand= StandardScaler()\n",

"for column in ['CreditScore','Age','Balance','EstimatedSalary','Tenure']:\n",

" data[column] = stand.fit\_transform(data[column].values.reshape(-1,1))"

],

"metadata": {

"id": "nZmJapEfGKNU"

},

"execution\_count": 15,

"outputs": []

},

{

"cell\_type": "code",

"source": [

"sns.kdeplot(data['CreditScore'], shade=True)\n",

"sns.kdeplot(data['Age'], shade=True)\n",

"sns.kdeplot(data['Balance'], shade=True)\n",

"sns.kdeplot(data['EstimatedSalary'], shade=True)\n",

"sns.kdeplot(data['Tenure'], shade=True)"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 296

},

"id": "5xURsBQOF1f0",

"outputId": "3d807d3b-2444-4da8-a68e-6cb564ed9bd1"

},

"execution\_count": 16,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"<matplotlib.axes.\_subplots.AxesSubplot at 0x7fc73b4a6fd0>"

]

},

"metadata": {},

"execution\_count": 16

},

{

"output\_type": "display\_data",

"data": {

"text/plain": [

"<Figure size 432x288 with 1 Axes>"

],

"image/png": "\n"

},

"metadata": {

"needs\_background": "light"

}

}

]

},

{

"cell\_type": "code",

"source": [

"sns.countplot(data['Exited'])"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 351

},

"id": "im6vO-RuGuNF",

"outputId": "7e54e46a-e35b-482a-9d83-afa0678809a0"

},

"execution\_count": 18,

"outputs": [

{

"output\_type": "stream",

"name": "stderr",

"text": [

"/usr/local/lib/python3.7/dist-packages/seaborn/\_decorators.py:43: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.\n",

" FutureWarning\n"

]

},

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"<matplotlib.axes.\_subplots.AxesSubplot at 0x7fc73ac92950>"

]

},

"metadata": {},

"execution\_count": 18

},

{

"output\_type": "display\_data",

"data": {

"text/plain": [

"<Figure size 432x288 with 1 Axes>"

],

"image/png": "\n"

},

"metadata": {

"needs\_background": "light"

}

}

]

},

{

"cell\_type": "markdown",

"source": [

"4. \*\*Perform\*\* \*\*descriptive\*\* \*\*statistics\*\* \*\*on\*\* \*\*the\*\* \*\*dataset\*\*"

],

"metadata": {

"id": "BylDUOplO8hh"

}

},

{

"cell\_type": "code",

"source": [

"data.describe()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 364

},

"id": "qRnre7j5PJHb",

"outputId": "331b9410-913d-46cd-8bcf-3b0a26eb80d3"

},

"execution\_count": 36,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" RowNumber CustomerId CreditScore Age Tenure \\\n",

"count 10000.00000 1.000000e+04 1.000000e+04 1.000000e+04 1.000000e+04 \n",

"mean 5000.50000 1.569094e+07 -4.824585e-16 2.318146e-16 -1.078249e-16 \n",

"std 2886.89568 7.193619e+04 1.000050e+00 1.000050e+00 1.000050e+00 \n",

"min 1.00000 1.556570e+07 -3.109504e+00 -1.994969e+00 -1.733315e+00 \n",

"25% 2500.75000 1.562853e+07 -6.883586e-01 -6.600185e-01 -6.959818e-01 \n",

"50% 5000.50000 1.569074e+07 1.522218e-02 -1.832505e-01 -4.425957e-03 \n",

"75% 7500.25000 1.575323e+07 6.981094e-01 4.842246e-01 6.871299e-01 \n",

"max 10000.00000 1.581569e+07 2.063884e+00 5.061197e+00 1.724464e+00 \n",

"\n",

" Balance NumOfProducts HasCrCard IsActiveMember \\\n",

"count 1.000000e+04 10000.000000 10000.00000 10000.000000 \n",

"mean -6.252776e-17 1.530200 0.70550 0.515100 \n",

"std 1.000050e+00 0.581654 0.45584 0.499797 \n",

"min -1.225848e+00 1.000000 0.00000 0.000000 \n",

"25% -1.225848e+00 1.000000 0.00000 0.000000 \n",

"50% 3.319639e-01 1.000000 1.00000 1.000000 \n",

"75% 8.199205e-01 2.000000 1.00000 1.000000 \n",

"max 2.795323e+00 4.000000 1.00000 1.000000 \n",

"\n",

" EstimatedSalary Exited \n",

"count 1.000000e+04 10000.000000 \n",

"mean -2.877698e-17 0.203700 \n",

"std 1.000050e+00 0.402769 \n",

"min -1.740268e+00 0.000000 \n",

"25% -8.535935e-01 0.000000 \n",

"50% 1.802807e-03 0.000000 \n",

"75% 8.572431e-01 0.000000 \n",

"max 1.737200e+00 1.000000 "

],

"text/html": [

"\n",

" <div id=\"df-ad5a430e-e8d7-40bd-9eb3-d626a89d4e1e\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>RowNumber</th>\n",

" <th>CustomerId</th>\n",

" <th>CreditScore</th>\n",

" <th>Age</th>\n",

" <th>Tenure</th>\n",

" <th>Balance</th>\n",

" <th>NumOfProducts</th>\n",

" <th>HasCrCard</th>\n",

" <th>IsActiveMember</th>\n",

" <th>EstimatedSalary</th>\n",

" <th>Exited</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>count</th>\n",

" <td>10000.00000</td>\n",

" <td>1.000000e+04</td>\n",

" <td>1.000000e+04</td>\n",

" <td>1.000000e+04</td>\n",

" <td>1.000000e+04</td>\n",

" <td>1.000000e+04</td>\n",

" <td>10000.000000</td>\n",

" <td>10000.00000</td>\n",

" <td>10000.000000</td>\n",

" <td>1.000000e+04</td>\n",

" <td>10000.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>mean</th>\n",

" <td>5000.50000</td>\n",

" <td>1.569094e+07</td>\n",

" <td>-4.824585e-16</td>\n",

" <td>2.318146e-16</td>\n",

" <td>-1.078249e-16</td>\n",

" <td>-6.252776e-17</td>\n",

" <td>1.530200</td>\n",

" <td>0.70550</td>\n",

" <td>0.515100</td>\n",

" <td>-2.877698e-17</td>\n",

" <td>0.203700</td>\n",

" </tr>\n",

" <tr>\n",

" <th>std</th>\n",

" <td>2886.89568</td>\n",

" <td>7.193619e+04</td>\n",

" <td>1.000050e+00</td>\n",

" <td>1.000050e+00</td>\n",

" <td>1.000050e+00</td>\n",

" <td>1.000050e+00</td>\n",

" <td>0.581654</td>\n",

" <td>0.45584</td>\n",

" <td>0.499797</td>\n",

" <td>1.000050e+00</td>\n",

" <td>0.402769</td>\n",

" </tr>\n",

" <tr>\n",

" <th>min</th>\n",

" <td>1.00000</td>\n",

" <td>1.556570e+07</td>\n",

" <td>-3.109504e+00</td>\n",

" <td>-1.994969e+00</td>\n",

" <td>-1.733315e+00</td>\n",

" <td>-1.225848e+00</td>\n",

" <td>1.000000</td>\n",

" <td>0.00000</td>\n",

" <td>0.000000</td>\n",

" <td>-1.740268e+00</td>\n",

" <td>0.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>25%</th>\n",

" <td>2500.75000</td>\n",

" <td>1.562853e+07</td>\n",

" <td>-6.883586e-01</td>\n",

" <td>-6.600185e-01</td>\n",

" <td>-6.959818e-01</td>\n",

" <td>-1.225848e+00</td>\n",

" <td>1.000000</td>\n",

" <td>0.00000</td>\n",

" <td>0.000000</td>\n",

" <td>-8.535935e-01</td>\n",

" <td>0.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>50%</th>\n",

" <td>5000.50000</td>\n",

" <td>1.569074e+07</td>\n",

" <td>1.522218e-02</td>\n",

" <td>-1.832505e-01</td>\n",

" <td>-4.425957e-03</td>\n",

" <td>3.319639e-01</td>\n",

" <td>1.000000</td>\n",

" <td>1.00000</td>\n",

" <td>1.000000</td>\n",

" <td>1.802807e-03</td>\n",

" <td>0.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>75%</th>\n",

" <td>7500.25000</td>\n",

" <td>1.575323e+07</td>\n",

" <td>6.981094e-01</td>\n",

" <td>4.842246e-01</td>\n",

" <td>6.871299e-01</td>\n",

" <td>8.199205e-01</td>\n",

" <td>2.000000</td>\n",

" <td>1.00000</td>\n",

" <td>1.000000</td>\n",

" <td>8.572431e-01</td>\n",

" <td>0.000000</td>\n",

" </tr>\n",

" <tr>\n",

" <th>max</th>\n",

" <td>10000.00000</td>\n",

" <td>1.581569e+07</td>\n",

" <td>2.063884e+00</td>\n",

" <td>5.061197e+00</td>\n",

" <td>1.724464e+00</td>\n",

" <td>2.795323e+00</td>\n",

" <td>4.000000</td>\n",

" <td>1.00000</td>\n",

" <td>1.000000</td>\n",

" <td>1.737200e+00</td>\n",

" <td>1.000000</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-ad5a430e-e8d7-40bd-9eb3-d626a89d4e1e')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-ad5a430e-e8d7-40bd-9eb3-d626a89d4e1e button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-ad5a430e-e8d7-40bd-9eb3-d626a89d4e1e');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 36

}

]

},

{

"cell\_type": "markdown",

"source": [

"5 .\*\*Handle the Missing values\*\*"

],

"metadata": {

"id": "DxKRUwnMrn0F"

}

},

{

"cell\_type": "code",

"source": [

"data.isnull().sum()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "gKvzyUUUrmoV",

"outputId": "f81ee9a7-83ca-480d-d830-5a823df46f5f"

},

"execution\_count": 8,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"RowNumber 0\n",

"CustomerId 0\n",

"Surname 0\n",

"CreditScore 0\n",

"Geography 0\n",

"Gender 0\n",

"Age 0\n",

"Tenure 0\n",

"Balance 0\n",

"NumOfProducts 0\n",

"HasCrCard 0\n",

"IsActiveMember 0\n",

"EstimatedSalary 0\n",

"Exited 0\n",

"dtype: int64"

]

},

"metadata": {},

"execution\_count": 8

}

]

},

{

"cell\_type": "markdown",

"source": [

" 6.\*\*Find\*\* \*\*the\*\* \*\*outliers\*\* \*\*and\*\* \*\*replace\*\* \*\*the\*\* \*\*outliers\*\*"

],

"metadata": {

"id": "zYvdriOMSYTM"

}

},

{

"cell\_type": "code",

"source": [

"lowerlimit=data['Age'].quantile(0.05)\n",

"lowerlimit\n",

"data[data['Age']<lowerlimit]\n",

"upperlimit=data['Age'].quantile(0.95)\n",

"upperlimit\n",

"data[data['Age']<upperlimit]\n",

"data=data[(data['Age']>lowerlimit)&(data['Age']<upperlimit)]\n",

"data"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 468

},

"id": "v1Dugi1jSY-m",

"outputId": "f56cde91-70ce-4e03-ebf3-fe024231f70d"

},

"execution\_count": 41,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" RowNumber CustomerId Surname CreditScore Geography Gender Age \\\n",

"0 1 15634602 Hargrave 619 France Female 42 \n",

"1 2 15647311 Hill 608 Spain Female 41 \n",

"2 3 15619304 Onio 502 France Female 42 \n",

"3 4 15701354 Boni 699 France Female 39 \n",

"4 5 15737888 Mitchell 850 Spain Female 43 \n",

"... ... ... ... ... ... ... ... \n",

"9995 9996 15606229 Obijiaku 771 France Male 39 \n",

"9996 9997 15569892 Johnstone 516 France Male 35 \n",

"9997 9998 15584532 Liu 709 France Female 36 \n",

"9998 9999 15682355 Sabbatini 772 Germany Male 42 \n",

"9999 10000 15628319 Walker 792 France Female 28 \n",

"\n",

" Tenure Balance NumOfProducts HasCrCard IsActiveMember \\\n",

"0 2 0.00 1 1 1 \n",

"1 1 83807.86 1 0 1 \n",

"2 8 159660.80 3 1 0 \n",

"3 1 0.00 2 0 0 \n",

"4 2 125510.82 1 1 1 \n",

"... ... ... ... ... ... \n",

"9995 5 0.00 2 1 0 \n",

"9996 10 57369.61 1 1 1 \n",

"9997 7 0.00 1 0 1 \n",

"9998 3 75075.31 2 1 0 \n",

"9999 4 130142.79 1 1 0 \n",

"\n",

" EstimatedSalary Exited \n",

"0 101348.88 1 \n",

"1 112542.58 0 \n",

"2 113931.57 1 \n",

"3 93826.63 0 \n",

"4 79084.10 0 \n",

"... ... ... \n",

"9995 96270.64 0 \n",

"9996 101699.77 0 \n",

"9997 42085.58 1 \n",

"9998 92888.52 1 \n",

"9999 38190.78 0 \n",

"\n",

"[8863 rows x 14 columns]"

],

"text/html": [

"\n",

" <div id=\"df-02cb1053-963d-40bf-8b42-859e54403a77\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>RowNumber</th>\n",

" <th>CustomerId</th>\n",

" <th>Surname</th>\n",

" <th>CreditScore</th>\n",

" <th>Geography</th>\n",

" <th>Gender</th>\n",

" <th>Age</th>\n",

" <th>Tenure</th>\n",

" <th>Balance</th>\n",

" <th>NumOfProducts</th>\n",

" <th>HasCrCard</th>\n",

" <th>IsActiveMember</th>\n",

" <th>EstimatedSalary</th>\n",

" <th>Exited</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>1</td>\n",

" <td>15634602</td>\n",

" <td>Hargrave</td>\n",

" <td>619</td>\n",

" <td>France</td>\n",

" <td>Female</td>\n",

" <td>42</td>\n",

" <td>2</td>\n",

" <td>0.00</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>101348.88</td>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>2</td>\n",

" <td>15647311</td>\n",

" <td>Hill</td>\n",

" <td>608</td>\n",

" <td>Spain</td>\n",

" <td>Female</td>\n",

" <td>41</td>\n",

" <td>1</td>\n",

" <td>83807.86</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>112542.58</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>3</td>\n",

" <td>15619304</td>\n",

" <td>Onio</td>\n",

" <td>502</td>\n",

" <td>France</td>\n",

" <td>Female</td>\n",

" <td>42</td>\n",

" <td>8</td>\n",

" <td>159660.80</td>\n",

" <td>3</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>113931.57</td>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>4</td>\n",

" <td>15701354</td>\n",

" <td>Boni</td>\n",

" <td>699</td>\n",

" <td>France</td>\n",

" <td>Female</td>\n",

" <td>39</td>\n",

" <td>1</td>\n",

" <td>0.00</td>\n",

" <td>2</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>93826.63</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>5</td>\n",

" <td>15737888</td>\n",

" <td>Mitchell</td>\n",

" <td>850</td>\n",

" <td>Spain</td>\n",

" <td>Female</td>\n",

" <td>43</td>\n",

" <td>2</td>\n",

" <td>125510.82</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>79084.10</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>...</th>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" <td>...</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9995</th>\n",

" <td>9996</td>\n",

" <td>15606229</td>\n",

" <td>Obijiaku</td>\n",

" <td>771</td>\n",

" <td>France</td>\n",

" <td>Male</td>\n",

" <td>39</td>\n",

" <td>5</td>\n",

" <td>0.00</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>96270.64</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9996</th>\n",

" <td>9997</td>\n",

" <td>15569892</td>\n",

" <td>Johnstone</td>\n",

" <td>516</td>\n",

" <td>France</td>\n",

" <td>Male</td>\n",

" <td>35</td>\n",

" <td>10</td>\n",

" <td>57369.61</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>101699.77</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9997</th>\n",

" <td>9998</td>\n",

" <td>15584532</td>\n",

" <td>Liu</td>\n",

" <td>709</td>\n",

" <td>France</td>\n",

" <td>Female</td>\n",

" <td>36</td>\n",

" <td>7</td>\n",

" <td>0.00</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>42085.58</td>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9998</th>\n",

" <td>9999</td>\n",

" <td>15682355</td>\n",

" <td>Sabbatini</td>\n",

" <td>772</td>\n",

" <td>Germany</td>\n",

" <td>Male</td>\n",

" <td>42</td>\n",

" <td>3</td>\n",

" <td>75075.31</td>\n",

" <td>2</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>92888.52</td>\n",

" <td>1</td>\n",

" </tr>\n",

" <tr>\n",

" <th>9999</th>\n",

" <td>10000</td>\n",

" <td>15628319</td>\n",

" <td>Walker</td>\n",

" <td>792</td>\n",

" <td>France</td>\n",

" <td>Female</td>\n",

" <td>28</td>\n",

" <td>4</td>\n",

" <td>130142.79</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>38190.78</td>\n",

" <td>0</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>8863 rows × 14 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-02cb1053-963d-40bf-8b42-859e54403a77')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-02cb1053-963d-40bf-8b42-859e54403a77 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-02cb1053-963d-40bf-8b42-859e54403a77');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 41

}

]

},

{

"cell\_type": "markdown",

"source": [

" 7.\*\*Check\*\* \*\*for\*\* \*\*Categorical\*\* \*\*columns\*\* \*\*and\*\* \*\*perform\*\* \*\*encoding\*\*"

],

"metadata": {

"id": "ykvJzJbtMMwA"

}

},

{

"cell\_type": "code",

"source": [

"x = pd.get\_dummies(x)\n",

"\n",

"x.head()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "9InjESSpMI2T",

"outputId": "118c99d1-1270-49af-8711-7fc0bd9001b5"

},

"execution\_count": 26,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" RowNumber CustomerId CreditScore Age Tenure Balance \\\n",

"0 1 15634602 -0.326221 0.293517 -1.041760 -1.225848 \n",

"1 2 15647311 -0.440036 0.198164 -1.387538 0.117350 \n",

"2 3 15619304 -1.536794 0.293517 1.032908 1.333053 \n",

"3 4 15701354 0.501521 0.007457 -1.387538 -1.225848 \n",

"4 5 15737888 2.063884 0.388871 -1.041760 0.785728 \n",

"\n",

" NumOfProducts Surname\_Abazu Surname\_Abbie Surname\_Abbott ... \\\n",

"0 1 0 0 0 ... \n",

"1 1 0 0 0 ... \n",

"2 3 0 0 0 ... \n",

"3 2 0 0 0 ... \n",

"4 1 0 0 0 ... \n",

"\n",

" Surname\_Zubarev Surname\_Zubareva Surname\_Zuev Surname\_Zuyev \\\n",

"0 0 0 0 0 \n",

"1 0 0 0 0 \n",

"2 0 0 0 0 \n",

"3 0 0 0 0 \n",

"4 0 0 0 0 \n",

"\n",

" Surname\_Zuyeva Geography\_France Geography\_Germany Geography\_Spain \\\n",

"0 0 1 0 0 \n",

"1 0 0 0 1 \n",

"2 0 1 0 0 \n",

"3 0 1 0 0 \n",

"4 0 0 0 1 \n",

"\n",

" Gender\_Female Gender\_Male \n",

"0 1 0 \n",

"1 1 0 \n",

"2 1 0 \n",

"3 1 0 \n",

"4 1 0 \n",

"\n",

"[5 rows x 2944 columns]"

],

"text/html": [

"\n",

" <div id=\"df-94821227-0c99-441d-b485-1b8a5726fcd8\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>RowNumber</th>\n",

" <th>CustomerId</th>\n",

" <th>CreditScore</th>\n",

" <th>Age</th>\n",

" <th>Tenure</th>\n",

" <th>Balance</th>\n",

" <th>NumOfProducts</th>\n",

" <th>Surname\_Abazu</th>\n",

" <th>Surname\_Abbie</th>\n",

" <th>Surname\_Abbott</th>\n",

" <th>...</th>\n",

" <th>Surname\_Zubarev</th>\n",

" <th>Surname\_Zubareva</th>\n",

" <th>Surname\_Zuev</th>\n",

" <th>Surname\_Zuyev</th>\n",

" <th>Surname\_Zuyeva</th>\n",

" <th>Geography\_France</th>\n",

" <th>Geography\_Germany</th>\n",

" <th>Geography\_Spain</th>\n",

" <th>Gender\_Female</th>\n",

" <th>Gender\_Male</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>1</td>\n",

" <td>15634602</td>\n",

" <td>-0.326221</td>\n",

" <td>0.293517</td>\n",

" <td>-1.041760</td>\n",

" <td>-1.225848</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>2</td>\n",

" <td>15647311</td>\n",

" <td>-0.440036</td>\n",

" <td>0.198164</td>\n",

" <td>-1.387538</td>\n",

" <td>0.117350</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>3</td>\n",

" <td>15619304</td>\n",

" <td>-1.536794</td>\n",

" <td>0.293517</td>\n",

" <td>1.032908</td>\n",

" <td>1.333053</td>\n",

" <td>3</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>4</td>\n",

" <td>15701354</td>\n",

" <td>0.501521</td>\n",

" <td>0.007457</td>\n",

" <td>-1.387538</td>\n",

" <td>-1.225848</td>\n",

" <td>2</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>5</td>\n",

" <td>15737888</td>\n",

" <td>2.063884</td>\n",

" <td>0.388871</td>\n",

" <td>-1.041760</td>\n",

" <td>0.785728</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>...</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>0</td>\n",

" <td>1</td>\n",

" <td>1</td>\n",

" <td>0</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 2944 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-94821227-0c99-441d-b485-1b8a5726fcd8')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-94821227-0c99-441d-b485-1b8a5726fcd8 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-94821227-0c99-441d-b485-1b8a5726fcd8');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 26

}

]

},

{

"cell\_type": "markdown",

"source": [

"8.\*\*Split\*\* \*\*the\*\* \*\*data\*\* \*\*into\*\* \*\*dependent\*\* \*\*and\*\* \*\*independent\*\* \*\*variables\*\*"

],

"metadata": {

"id": "iziIg9B4JPg\_"

}

},

{

"cell\_type": "code",

"source": [

"x = data.iloc[:,0:10]\n",

"y = data.iloc[:,10]\n",

"\n",

"print(x.shape)\n",

"print(y.shape)\n"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "nRHHpBIbIgws",

"outputId": "8061b090-614f-4c89-ce9a-ba3ac9f1c632"

},

"execution\_count": 22,

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"(10000, 10)\n",

"(10000,)\n"

]

}

]

},

{

"cell\_type": "markdown",

"source": [

"9.\*\*Scale\*\* \*\*the\*\* \*\*independent\*\* \*\*variables\*\* "

],

"metadata": {

"id": "oqIJkD-zMzAJ"

}

},

{

"cell\_type": "code",

"source": [

"from sklearn.preprocessing import StandardScaler\n",

"from sklearn.model\_selection import train\_test\_split\n",

"x\_train, x\_test, y\_train, y\_test= train\_test\_split(x,y,test\_size=0.25,random\_state=0)\n",

"sc = StandardScaler()\n",

"x\_train=sc.fit\_transform(x\_train)\n",

"x\_test = sc.fit\_transform(x\_test)\n",

"\n",

"x\_train = pd.DataFrame(x\_train)\n",

"x\_train.head()"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 300

},

"id": "aaDRzlNmM-\_b",

"outputId": "3f2e22e4-f743-4999-9200-609aaaf88f8b"

},

"execution\_count": 32,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

" 0 1 2 3 4 5 6 \\\n",

"0 -0.702176 -1.343330 -0.735507 0.015266 0.008860 0.673160 2.535034 \n",

"1 -1.485722 1.558330 1.024427 -0.652609 0.008860 -1.207724 0.804242 \n",

"2 -0.524522 -0.655156 0.808295 -0.461788 1.393293 -0.356937 0.804242 \n",

"3 -1.167396 1.200594 0.396614 -0.080145 0.008860 -0.009356 -0.926551 \n",

"4 -1.451159 0.778798 -0.467915 1.255605 0.701077 -1.207724 0.804242 \n",

"\n",

" 7 8 9 ... 2934 2935 2936 2937 2938 \\\n",

"0 -0.016332 0.0 -0.0231 ... -0.011548 0.0 -0.011548 -0.011548 -0.016332 \n",

"1 -0.016332 0.0 -0.0231 ... -0.011548 0.0 -0.011548 -0.011548 -0.016332 \n",

"2 -0.016332 0.0 -0.0231 ... -0.011548 0.0 -0.011548 -0.011548 -0.016332 \n",

"3 -0.016332 0.0 -0.0231 ... -0.011548 0.0 -0.011548 -0.011548 -0.016332 \n",

"4 -0.016332 0.0 -0.0231 ... -0.011548 0.0 -0.011548 -0.011548 -0.016332 \n",

"\n",

" 2939 2940 2941 2942 2943 \n",

"0 -1.015588 1.760216 -0.574682 1.087261 -1.087261 \n",

"1 0.984651 -0.568112 -0.574682 1.087261 -1.087261 \n",

"2 -1.015588 -0.568112 1.740094 1.087261 -1.087261 \n",

"3 -1.015588 -0.568112 1.740094 -0.919743 0.919743 \n",

"4 0.984651 -0.568112 -0.574682 -0.919743 0.919743 \n",

"\n",

"[5 rows x 2944 columns]"

],

"text/html": [

"\n",

" <div id=\"df-95dc407f-35d5-4c21-949b-962392757830\">\n",

" <div class=\"colab-df-container\">\n",

" <div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>0</th>\n",

" <th>1</th>\n",

" <th>2</th>\n",

" <th>3</th>\n",

" <th>4</th>\n",

" <th>5</th>\n",

" <th>6</th>\n",

" <th>7</th>\n",

" <th>8</th>\n",

" <th>9</th>\n",

" <th>...</th>\n",

" <th>2934</th>\n",

" <th>2935</th>\n",

" <th>2936</th>\n",

" <th>2937</th>\n",

" <th>2938</th>\n",

" <th>2939</th>\n",

" <th>2940</th>\n",

" <th>2941</th>\n",

" <th>2942</th>\n",

" <th>2943</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>-0.702176</td>\n",

" <td>-1.343330</td>\n",

" <td>-0.735507</td>\n",

" <td>0.015266</td>\n",

" <td>0.008860</td>\n",

" <td>0.673160</td>\n",

" <td>2.535034</td>\n",

" <td>-0.016332</td>\n",

" <td>0.0</td>\n",

" <td>-0.0231</td>\n",

" <td>...</td>\n",

" <td>-0.011548</td>\n",

" <td>0.0</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.016332</td>\n",

" <td>-1.015588</td>\n",

" <td>1.760216</td>\n",

" <td>-0.574682</td>\n",

" <td>1.087261</td>\n",

" <td>-1.087261</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>-1.485722</td>\n",

" <td>1.558330</td>\n",

" <td>1.024427</td>\n",

" <td>-0.652609</td>\n",

" <td>0.008860</td>\n",

" <td>-1.207724</td>\n",

" <td>0.804242</td>\n",

" <td>-0.016332</td>\n",

" <td>0.0</td>\n",

" <td>-0.0231</td>\n",

" <td>...</td>\n",

" <td>-0.011548</td>\n",

" <td>0.0</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.016332</td>\n",

" <td>0.984651</td>\n",

" <td>-0.568112</td>\n",

" <td>-0.574682</td>\n",

" <td>1.087261</td>\n",

" <td>-1.087261</td>\n",

" </tr>\n",

" <tr>\n",

" <th>2</th>\n",

" <td>-0.524522</td>\n",

" <td>-0.655156</td>\n",

" <td>0.808295</td>\n",

" <td>-0.461788</td>\n",

" <td>1.393293</td>\n",

" <td>-0.356937</td>\n",

" <td>0.804242</td>\n",

" <td>-0.016332</td>\n",

" <td>0.0</td>\n",

" <td>-0.0231</td>\n",

" <td>...</td>\n",

" <td>-0.011548</td>\n",

" <td>0.0</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.016332</td>\n",

" <td>-1.015588</td>\n",

" <td>-0.568112</td>\n",

" <td>1.740094</td>\n",

" <td>1.087261</td>\n",

" <td>-1.087261</td>\n",

" </tr>\n",

" <tr>\n",

" <th>3</th>\n",

" <td>-1.167396</td>\n",

" <td>1.200594</td>\n",

" <td>0.396614</td>\n",

" <td>-0.080145</td>\n",

" <td>0.008860</td>\n",

" <td>-0.009356</td>\n",

" <td>-0.926551</td>\n",

" <td>-0.016332</td>\n",

" <td>0.0</td>\n",

" <td>-0.0231</td>\n",

" <td>...</td>\n",

" <td>-0.011548</td>\n",

" <td>0.0</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.016332</td>\n",

" <td>-1.015588</td>\n",

" <td>-0.568112</td>\n",

" <td>1.740094</td>\n",

" <td>-0.919743</td>\n",

" <td>0.919743</td>\n",

" </tr>\n",

" <tr>\n",

" <th>4</th>\n",

" <td>-1.451159</td>\n",

" <td>0.778798</td>\n",

" <td>-0.467915</td>\n",

" <td>1.255605</td>\n",

" <td>0.701077</td>\n",

" <td>-1.207724</td>\n",

" <td>0.804242</td>\n",

" <td>-0.016332</td>\n",

" <td>0.0</td>\n",

" <td>-0.0231</td>\n",

" <td>...</td>\n",

" <td>-0.011548</td>\n",

" <td>0.0</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.011548</td>\n",

" <td>-0.016332</td>\n",

" <td>0.984651</td>\n",

" <td>-0.568112</td>\n",

" <td>-0.574682</td>\n",

" <td>-0.919743</td>\n",

" <td>0.919743</td>\n",

" </tr>\n",

" </tbody>\n",

"</table>\n",

"<p>5 rows × 2944 columns</p>\n",

"</div>\n",

" <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-95dc407f-35d5-4c21-949b-962392757830')\"\n",

" title=\"Convert this dataframe to an interactive table.\"\n",

" style=\"display:none;\">\n",

" \n",

" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",

" width=\"24px\">\n",

" <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",

" <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",

" </svg>\n",

" </button>\n",

" \n",

" <style>\n",

" .colab-df-container {\n",

" display:flex;\n",

" flex-wrap:wrap;\n",

" gap: 12px;\n",

" }\n",

"\n",

" .colab-df-convert {\n",

" background-color: #E8F0FE;\n",

" border: none;\n",

" border-radius: 50%;\n",

" cursor: pointer;\n",

" display: none;\n",

" fill: #1967D2;\n",

" height: 32px;\n",

" padding: 0 0 0 0;\n",

" width: 32px;\n",

" }\n",

"\n",

" .colab-df-convert:hover {\n",

" background-color: #E2EBFA;\n",

" box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",

" fill: #174EA6;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert {\n",

" background-color: #3B4455;\n",

" fill: #D2E3FC;\n",

" }\n",

"\n",

" [theme=dark] .colab-df-convert:hover {\n",

" background-color: #434B5C;\n",

" box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",

" filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",

" fill: #FFFFFF;\n",

" }\n",

" </style>\n",

"\n",

" <script>\n",

" const buttonEl =\n",

" document.querySelector('#df-95dc407f-35d5-4c21-949b-962392757830 button.colab-df-convert');\n",

" buttonEl.style.display =\n",

" google.colab.kernel.accessAllowed ? 'block' : 'none';\n",

"\n",

" async function convertToInteractive(key) {\n",

" const element = document.querySelector('#df-95dc407f-35d5-4c21-949b-962392757830');\n",

" const dataTable =\n",

" await google.colab.kernel.invokeFunction('convertToInteractive',\n",

" [key], {});\n",

" if (!dataTable) return;\n",

"\n",

" const docLinkHtml = 'Like what you see? Visit the ' +\n",

" '<a target=\"\_blank\" href=https://colab.research.google.com/notebooks/data\_table.ipynb>data table notebook</a>'\n",

" + ' to learn more about interactive tables.';\n",

" element.innerHTML = '';\n",

" dataTable['output\_type'] = 'display\_data';\n",

" await google.colab.output.renderOutput(dataTable, element);\n",

" const docLink = document.createElement('div');\n",

" docLink.innerHTML = docLinkHtml;\n",

" element.appendChild(docLink);\n",

" }\n",

" </script>\n",

" </div>\n",

" </div>\n",

" "

]

},

"metadata": {},

"execution\_count": 32

}

]

},

{

"cell\_type": "markdown",

"source": [

" 10.\*\*Split\*\* \*\*the\*\* \*\*data\*\* \*\*into\*\* \*\*training\*\* \*\*and\*\* \*\*testing\*\*"

],

"metadata": {

"id": "8qZovkTIJbTj"

}

},

{

"cell\_type": "code",

"source": [

"from sklearn.model\_selection import train\_test\_split\n",

"x\_train, x\_test, y\_train, y\_test= train\_test\_split(x,y,test\_size=0.25,random\_state=0)\n",

"print(' x\_train.shape : ',x\_train.shape)\n",

"print(' y\_train.shape : ',y\_train.shape)\n",

"print(' x\_test.shape : ',x\_test.shape)\n",

"print(' y\_test.shape : ',y\_test.shape)"

],

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "CChYIcJHtHod",

"outputId": "fccadf82-15f4-47d6-8628-0a3aeb09d898"

},

"execution\_count": 23,

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

" x\_train.shape : (7500, 10)\n",

" y\_train.shape : (7500,)\n",

" x\_test.shape : (2500, 10)\n",

" y\_test.shape : (2500,)\n"

]

}

]

}

]

}