|  |
| --- |
| { |
|  | "cells": [ |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "import pandas as pd\n", |
|  | "import numpy as np" |
|  | ], |
|  | "metadata": { |
|  | "id": "SJ6tAkjj18NL" |
|  | }, |
|  | "execution\_count": 1, |
|  | "outputs": [] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "file=pd.read\_csv(\"/Churn\_Modelling.csv\")\n", |
|  | "pro=pd.DataFrame(file)\n", |
|  | "pro.head()" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 270 |
|  | }, |
|  | "id": "v0zOpikV2AJU", |
|  | "outputId": "87b38785-c693-45e1-c40b-f6a23dfa4659" |
|  | }, |
|  | "execution\_count": 12, |
|  | "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", |
|  | " 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", |
|  | " 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 " |
|  | ], |
|  | "text/html": [ |
|  | "\n", |
|  | " <div id=\"df-02feae72-c761-4186-b470-318fcd07dce7\">\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", |
|  | " </tbody>\n", |
|  | "</table>\n", |
|  | "</div>\n", |
|  | " <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-02feae72-c761-4186-b470-318fcd07dce7')\"\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-02feae72-c761-4186-b470-318fcd07dce7 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-02feae72-c761-4186-b470-318fcd07dce7');\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": 12 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "pro['HasCrCard'] = pro['HasCrCard'].astype('category')" |
|  | ], |
|  | "metadata": { |
|  | "id": "EdSjkjzx2gaG" |
|  | }, |
|  | "execution\_count": 13, |
|  | "outputs": [] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "pro['IsActiveMember'] = pro['IsActiveMember'].astype('category')\n", |
|  | "pro['Exited'] = pro['Exited'].astype('category')" |
|  | ], |
|  | "metadata": { |
|  | "id": "-JTPRKQy2jNE" |
|  | }, |
|  | "execution\_count": 14, |
|  | "outputs": [] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "\n", |
|  | "pro = pro.drop(columns=['RowNumber', 'CustomerId', 'Surname'])" |
|  | ], |
|  | "metadata": { |
|  | "id": "-5brBCAa2lmZ" |
|  | }, |
|  | "execution\_count": 15, |
|  | "outputs": [] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "pro.head()" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 270 |
|  | }, |
|  | "id": "yo7Z3pwb2n\_5", |
|  | "outputId": "15f9b900-27fc-46bd-808b-19b5c7cc0ee3" |
|  | }, |
|  | "execution\_count": 16, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | " CreditScore Geography Gender Age Tenure Balance NumOfProducts \\\n", |
|  | "0 619 France Female 42 2 0.00 1 \n", |
|  | "1 608 Spain Female 41 1 83807.86 1 \n", |
|  | "2 502 France Female 42 8 159660.80 3 \n", |
|  | "3 699 France Female 39 1 0.00 2 \n", |
|  | "4 850 Spain Female 43 2 125510.82 1 \n", |
|  | "\n", |
|  | " HasCrCard IsActiveMember EstimatedSalary Exited \n", |
|  | "0 1 1 101348.88 1 \n", |
|  | "1 0 1 112542.58 0 \n", |
|  | "2 1 0 113931.57 1 \n", |
|  | "3 0 0 93826.63 0 \n", |
|  | "4 1 1 79084.10 0 " |
|  | ], |
|  | "text/html": [ |
|  | "\n", |
|  | " <div id=\"df-4f226628-a85d-41b8-80da-b9420a48fb14\">\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>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>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>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>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>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>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", |
|  | " </tbody>\n", |
|  | "</table>\n", |
|  | "</div>\n", |
|  | " <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-4f226628-a85d-41b8-80da-b9420a48fb14')\"\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-4f226628-a85d-41b8-80da-b9420a48fb14 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-4f226628-a85d-41b8-80da-b9420a48fb14');\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": 16 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "import seaborn as sns\n", |
|  | "density = pro['Exited'].value\_counts(normalize=True).reset\_index()\n", |
|  | "sns.barplot(data=density, x='index', y='Exited', );\n", |
|  | "density" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 374 |
|  | }, |
|  | "id": "y0D8vZlX2qpU", |
|  | "outputId": "c870051e-cf28-4377-d3f9-2e48bd9d0806" |
|  | }, |
|  | "execution\_count": 17, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | " index Exited\n", |
|  | "0 0 0.7963\n", |
|  | "1 1 0.2037" |
|  | ], |
|  | "text/html": [ |
|  | "\n", |
|  | " <div id=\"df-80209c26-0c42-4011-8023-8f80bc381df8\">\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>index</th>\n", |
|  | " <th>Exited</th>\n", |
|  | " </tr>\n", |
|  | " </thead>\n", |
|  | " <tbody>\n", |
|  | " <tr>\n", |
|  | " <th>0</th>\n", |
|  | " <td>0</td>\n", |
|  | " <td>0.7963</td>\n", |
|  | " </tr>\n", |
|  | " <tr>\n", |
|  | " <th>1</th>\n", |
|  | " <td>1</td>\n", |
|  | " <td>0.2037</td>\n", |
|  | " </tr>\n", |
|  | " </tbody>\n", |
|  | "</table>\n", |
|  | "</div>\n", |
|  | " <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-80209c26-0c42-4011-8023-8f80bc381df8')\"\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-80209c26-0c42-4011-8023-8f80bc381df8 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-80209c26-0c42-4011-8023-8f80bc381df8');\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": 17 |
|  | }, |
|  | { |
|  | "output\_type": "display\_data", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<Figure size 432x288 with 1 Axes>" |
|  | ], |
|  | "image/png": "\n" |
|  | }, |
|  | "metadata": { |
|  | "needs\_background": "light" |
|  | } |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "import matplotlib.pyplot as plt" |
|  | ], |
|  | "metadata": { |
|  | "id": "ukLd8rcG2s\_o" |
|  | }, |
|  | "execution\_count": 18, |
|  | "outputs": [] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "categorical = pro.drop(columns=['CreditScore', 'Age', 'Tenure', 'Balance', 'EstimatedSalary'])\n", |
|  | "rows = int(np.ceil(categorical.shape[1] / 2)) - 1\n", |
|  | "\n", |
|  | "# create sub-plots anf title them\n", |
|  | "fig, axes = plt.subplots(nrows=rows, ncols=2, figsize=(10,6))\n", |
|  | "axes = axes.flatten()\n", |
|  | "\n", |
|  | "for row in range(rows):\n", |
|  | " cols = min(2, categorical.shape[1] - row\*2)\n", |
|  | " for col in range(cols):\n", |
|  | " col\_name = categorical.columns[2 \* row + col]\n", |
|  | " ax = axes[row\*2 + col] \n", |
|  | "\n", |
|  | " sns.countplot(data=categorical, x=col\_name, hue=\"Exited\", ax=ax);\n", |
|  | " \n", |
|  | "plt.tight\_layout()" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 441 |
|  | }, |
|  | "id": "pjZgJrab2vd\_", |
|  | "outputId": "e04a5991-d246-42c2-ec67-235fff0b664e" |
|  | }, |
|  | "execution\_count": 22, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "display\_data", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<Figure size 720x432 with 4 Axes>" |
|  | ], |
|  | "image/png": "\n" |
|  | }, |
|  | "metadata": { |
|  | "needs\_background": "light" |
|  | } |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "pro.info" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/" |
|  | }, |
|  | "id": "215KdTlT2xbD", |
|  | "outputId": "6b907ba3-bd99-488a-c175-d424babfc174" |
|  | }, |
|  | "execution\_count": 23, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<bound method DataFrame.info of CreditScore Geography Gender Age Tenure Balance NumOfProducts \\\n", |
|  | "0 619 France Female 42 2 0.00 1 \n", |
|  | "1 608 Spain Female 41 1 83807.86 1 \n", |
|  | "2 502 France Female 42 8 159660.80 3 \n", |
|  | "3 699 France Female 39 1 0.00 2 \n", |
|  | "4 850 Spain Female 43 2 125510.82 1 \n", |
|  | "... ... ... ... ... ... ... ... \n", |
|  | "9995 771 France Male 39 5 0.00 2 \n", |
|  | "9996 516 France Male 35 10 57369.61 1 \n", |
|  | "9997 709 France Female 36 7 0.00 1 \n", |
|  | "9998 772 Germany Male 42 3 75075.31 2 \n", |
|  | "9999 792 France Female 28 4 130142.79 1 \n", |
|  | "\n", |
|  | " HasCrCard IsActiveMember EstimatedSalary Exited \n", |
|  | "0 1 1 101348.88 1 \n", |
|  | "1 0 1 112542.58 0 \n", |
|  | "2 1 0 113931.57 1 \n", |
|  | "3 0 0 93826.63 0 \n", |
|  | "4 1 1 79084.10 0 \n", |
|  | "... ... ... ... ... \n", |
|  | "9995 1 0 96270.64 0 \n", |
|  | "9996 1 1 101699.77 0 \n", |
|  | "9997 0 1 42085.58 1 \n", |
|  | "9998 1 0 92888.52 1 \n", |
|  | "9999 1 0 38190.78 0 \n", |
|  | "\n", |
|  | "[10000 rows x 11 columns]>" |
|  | ] |
|  | }, |
|  | "metadata": {}, |
|  | "execution\_count": 23 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "pro.isna().sum()" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/" |
|  | }, |
|  | "id": "g9oYJ3Ol3nOm", |
|  | "outputId": "48d62ed5-01eb-4448-c688-4ce52dfc82c0" |
|  | }, |
|  | "execution\_count": 25, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | "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": 25 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "for i in pro:\n", |
|  | " if pro[i].dtype=='object' or pro[i].dtype=='category':\n", |
|  | " print(\"unique of \"+i+\" is \"+str(len(set(df[i])))+\" they are \"+str(set(df[i])))" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/" |
|  | }, |
|  | "id": "l9WWTxRb3tCt", |
|  | "outputId": "132fd37c-1aaa-4c09-f3fa-8b64803f97cb" |
|  | }, |
|  | "execution\_count": 26, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "stream", |
|  | "name": "stdout", |
|  | "text": [ |
|  | "unique of Geography is 3 they are {'Germany', 'France', 'Spain'}\n", |
|  | "unique of Gender is 2 they are {'Male', 'Female'}\n", |
|  | "unique of HasCrCard is 2 they are {0, 1}\n", |
|  | "unique of IsActiveMember is 2 they are {0, 1}\n", |
|  | "unique of Exited is 2 they are {0, 1}\n" |
|  | ] |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "def box\_scatter(data, x, y): \n", |
|  | " fig, (ax1, ax2) = plt.subplots(nrows=2, ncols=1, figsize=(16,6))\n", |
|  | " sns.boxplot(data=data, x=x, ax=ax1)\n", |
|  | " sns.scatterplot(data=data, x=x,y=y,ax=ax2)" |
|  | ], |
|  | "metadata": { |
|  | "id": "xdDOmq0L3xdJ" |
|  | }, |
|  | "execution\_count": 27, |
|  | "outputs": [] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "box\_scatter(pro,'CreditScore','Exited');\n", |
|  | "plt.tight\_layout()\n", |
|  | "print(f\"# of Bivariate Outliers: {len(pro.loc[pro['CreditScore'] < 400])}\")" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 375 |
|  | }, |
|  | "id": "bG8cCm1D32ij", |
|  | "outputId": "e2fc30c4-eea6-4350-a226-7780971cb304" |
|  | }, |
|  | "execution\_count": 28, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "stream", |
|  | "name": "stdout", |
|  | "text": [ |
|  | "# of Bivariate Outliers: 19\n" |
|  | ] |
|  | }, |
|  | { |
|  | "output\_type": "display\_data", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<Figure size 1152x432 with 2 Axes>" |
|  | ], |
|  | "image/png": "\n" |
|  | }, |
|  | "metadata": { |
|  | "needs\_background": "light" |
|  | } |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "box\_scatter(pro,'Age','Exited');\n", |
|  | "plt.tight\_layout()\n", |
|  | "print(f\"# of Bivariate Outliers: {len(pro.loc[pro['Age'] > 87])}\")" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 375 |
|  | }, |
|  | "id": "prJ9bPI338Qg", |
|  | "outputId": "f93eb05f-575e-4511-b09c-873a130eab33" |
|  | }, |
|  | "execution\_count": 29, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "stream", |
|  | "name": "stdout", |
|  | "text": [ |
|  | "# of Bivariate Outliers: 3\n" |
|  | ] |
|  | }, |
|  | { |
|  | "output\_type": "display\_data", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<Figure size 1152x432 with 2 Axes>" |
|  | ], |
|  | "image/png": "\n" |
|  | }, |
|  | "metadata": { |
|  | "needs\_background": "light" |
|  | } |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "box\_scatter(pro,'Balance','Exited');\n", |
|  | "plt.tight\_layout()\n", |
|  | "print(f\"# of Bivariate Outliers: {len(pro.loc[pro['Balance'] > 220000])}\")" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 375 |
|  | }, |
|  | "id": "TGMOsMF84Biu", |
|  | "outputId": "548c050b-616e-41a3-c977-60b0a1112943" |
|  | }, |
|  | "execution\_count": 30, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "stream", |
|  | "name": "stdout", |
|  | "text": [ |
|  | "# of Bivariate Outliers: 4\n" |
|  | ] |
|  | }, |
|  | { |
|  | "output\_type": "display\_data", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<Figure size 1152x432 with 2 Axes>" |
|  | ], |
|  | "image/png": "\n" |
|  | }, |
|  | "metadata": { |
|  | "needs\_background": "light" |
|  | } |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "box\_scatter(pro,'EstimatedSalary','Exited');\n", |
|  | "plt.tight\_layout()" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 357 |
|  | }, |
|  | "id": "Pfob2xOV4G3o", |
|  | "outputId": "3ebc8545-40de-4107-bd6d-d488754af9d9" |
|  | }, |
|  | "execution\_count": 31, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "display\_data", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<Figure size 1152x432 with 2 Axes>" |
|  | ], |
|  | "image/png": "\n" |
|  | }, |
|  | "metadata": { |
|  | "needs\_background": "light" |
|  | } |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "for i in pro:\n", |
|  | " if pro[i].dtype=='int64' or pro[i].dtypes=='float64':\n", |
|  | " q1=pro[i].quantile(0.25)\n", |
|  | " q3=pro[i].quantile(0.75)\n", |
|  | " iqr=q3-q1\n", |
|  | " upper=q3+1.5\*iqr\n", |
|  | " lower=q1-1.5\*iqr\n", |
|  | " pro[i]=np.where(pro[i] >upper, upper, pro[i])\n", |
|  | " pro[i]=np.where(pro[i] <lower, lower, pro[i])\n", |
|  | " " |
|  | ], |
|  | "metadata": { |
|  | "id": "FLLla9bX4KuP" |
|  | }, |
|  | "execution\_count": 32, |
|  | "outputs": [] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "box\_scatter(pro,'CreditScore','Exited');\n", |
|  | "plt.tight\_layout()\n", |
|  | "print(f\"# of Bivariate Outliers: {len(pro.loc[pro['CreditScore'] < 400])}\")" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 375 |
|  | }, |
|  | "id": "poNE\_CUo4U2m", |
|  | "outputId": "586932d9-e3c9-40cb-cd28-4cd6f3682315" |
|  | }, |
|  | "execution\_count": 33, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "stream", |
|  | "name": "stdout", |
|  | "text": [ |
|  | "# of Bivariate Outliers: 19\n" |
|  | ] |
|  | }, |
|  | { |
|  | "output\_type": "display\_data", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<Figure size 1152x432 with 2 Axes>" |
|  | ], |
|  | "image/png": "\n" |
|  | }, |
|  | "metadata": { |
|  | "needs\_background": "light" |
|  | } |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "box\_scatter(pro,'Age','Exited');\n", |
|  | "plt.tight\_layout()\n", |
|  | "print(f\"# of Bivariate Outliers: {len(pro.loc[pro['Age'] > 87])}\")" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 375 |
|  | }, |
|  | "id": "OJA8dEwz4aKx", |
|  | "outputId": "1b69d2f7-f974-425a-b8f9-2fe6bec22f1a" |
|  | }, |
|  | "execution\_count": 34, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "stream", |
|  | "name": "stdout", |
|  | "text": [ |
|  | "# of Bivariate Outliers: 0\n" |
|  | ] |
|  | }, |
|  | { |
|  | "output\_type": "display\_data", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<Figure size 1152x432 with 2 Axes>" |
|  | ], |
|  | "image/png": "\n" |
|  | }, |
|  | "metadata": { |
|  | "needs\_background": "light" |
|  | } |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "box\_scatter(pro,'Balance','Exited');\n", |
|  | "plt.tight\_layout()\n", |
|  | "print(f\"# of Bivariate Outliers: {len(pro.loc[pro['Balance'] > 220000])}\")" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 375 |
|  | }, |
|  | "id": "3rOST9\_j4fy5", |
|  | "outputId": "e7a45806-7784-4de4-bd3d-2ff4fc6acc8d" |
|  | }, |
|  | "execution\_count": 35, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "stream", |
|  | "name": "stdout", |
|  | "text": [ |
|  | "# of Bivariate Outliers: 4\n" |
|  | ] |
|  | }, |
|  | { |
|  | "output\_type": "display\_data", |
|  | "data": { |
|  | "text/plain": [ |
|  | "<Figure size 1152x432 with 2 Axes>" |
|  | ], |
|  | "image/png": "\n" |
|  | }, |
|  | "metadata": { |
|  | "needs\_background": "light" |
|  | } |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "from sklearn.preprocessing import LabelEncoder\n", |
|  | "encoder=LabelEncoder()\n", |
|  | "for i in pro:\n", |
|  | " if pro[i].dtype=='object' or pro[i].dtype=='category':\n", |
|  | " pro[i]=encoder.fit\_transform(pro[i])" |
|  | ], |
|  | "metadata": { |
|  | "id": "jg4CVChY4l9s" |
|  | }, |
|  | "execution\_count": 36, |
|  | "outputs": [] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "x=pro.iloc[:,:-1]\n", |
|  | "x.head()" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/", |
|  | "height": 270 |
|  | }, |
|  | "id": "ciUAjg9L4sL1", |
|  | "outputId": "706fca47-aba7-4a8e-b7c5-ddf13ea36f89" |
|  | }, |
|  | "execution\_count": 37, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | " CreditScore Geography Gender Age Tenure Balance NumOfProducts \\\n", |
|  | "0 619.0 0 0 42.0 2.0 0.00 1.0 \n", |
|  | "1 608.0 2 0 41.0 1.0 83807.86 1.0 \n", |
|  | "2 502.0 0 0 42.0 8.0 159660.80 3.0 \n", |
|  | "3 699.0 0 0 39.0 1.0 0.00 2.0 \n", |
|  | "4 850.0 2 0 43.0 2.0 125510.82 1.0 \n", |
|  | "\n", |
|  | " HasCrCard IsActiveMember EstimatedSalary \n", |
|  | "0 1 1 101348.88 \n", |
|  | "1 0 1 112542.58 \n", |
|  | "2 1 0 113931.57 \n", |
|  | "3 0 0 93826.63 \n", |
|  | "4 1 1 79084.10 " |
|  | ], |
|  | "text/html": [ |
|  | "\n", |
|  | " <div id=\"df-71330b13-2873-41c2-9dc0-b1458361a20d\">\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>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", |
|  | " </tr>\n", |
|  | " </thead>\n", |
|  | " <tbody>\n", |
|  | " <tr>\n", |
|  | " <th>0</th>\n", |
|  | " <td>619.0</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>42.0</td>\n", |
|  | " <td>2.0</td>\n", |
|  | " <td>0.00</td>\n", |
|  | " <td>1.0</td>\n", |
|  | " <td>1</td>\n", |
|  | " <td>1</td>\n", |
|  | " <td>101348.88</td>\n", |
|  | " </tr>\n", |
|  | " <tr>\n", |
|  | " <th>1</th>\n", |
|  | " <td>608.0</td>\n", |
|  | " <td>2</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>41.0</td>\n", |
|  | " <td>1.0</td>\n", |
|  | " <td>83807.86</td>\n", |
|  | " <td>1.0</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>1</td>\n", |
|  | " <td>112542.58</td>\n", |
|  | " </tr>\n", |
|  | " <tr>\n", |
|  | " <th>2</th>\n", |
|  | " <td>502.0</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>42.0</td>\n", |
|  | " <td>8.0</td>\n", |
|  | " <td>159660.80</td>\n", |
|  | " <td>3.0</td>\n", |
|  | " <td>1</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>113931.57</td>\n", |
|  | " </tr>\n", |
|  | " <tr>\n", |
|  | " <th>3</th>\n", |
|  | " <td>699.0</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>39.0</td>\n", |
|  | " <td>1.0</td>\n", |
|  | " <td>0.00</td>\n", |
|  | " <td>2.0</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>93826.63</td>\n", |
|  | " </tr>\n", |
|  | " <tr>\n", |
|  | " <th>4</th>\n", |
|  | " <td>850.0</td>\n", |
|  | " <td>2</td>\n", |
|  | " <td>0</td>\n", |
|  | " <td>43.0</td>\n", |
|  | " <td>2.0</td>\n", |
|  | " <td>125510.82</td>\n", |
|  | " <td>1.0</td>\n", |
|  | " <td>1</td>\n", |
|  | " <td>1</td>\n", |
|  | " <td>79084.10</td>\n", |
|  | " </tr>\n", |
|  | " </tbody>\n", |
|  | "</table>\n", |
|  | "</div>\n", |
|  | " <button class=\"colab-df-convert\" onclick=\"convertToInteractive('df-71330b13-2873-41c2-9dc0-b1458361a20d')\"\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-71330b13-2873-41c2-9dc0-b1458361a20d 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-71330b13-2873-41c2-9dc0-b1458361a20d');\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": 37 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "y=pro.iloc[:,-1]\n", |
|  | "y.head()" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/" |
|  | }, |
|  | "id": "rkOlvrHU4vLo", |
|  | "outputId": "5193081f-f0d5-417a-9b6e-6d3951706315" |
|  | }, |
|  | "execution\_count": 38, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | "0 1\n", |
|  | "1 0\n", |
|  | "2 1\n", |
|  | "3 0\n", |
|  | "4 0\n", |
|  | "Name: Exited, dtype: int64" |
|  | ] |
|  | }, |
|  | "metadata": {}, |
|  | "execution\_count": 38 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "from sklearn.preprocessing import StandardScaler\n", |
|  | "scaler=StandardScaler()\n", |
|  | "x=scaler.fit\_transform(x)" |
|  | ], |
|  | "metadata": { |
|  | "id": "3aVuJwYg4xnR" |
|  | }, |
|  | "execution\_count": 39, |
|  | "outputs": [] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "x" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/" |
|  | }, |
|  | "id": "FUABWKY640Oa", |
|  | "outputId": "1830203b-397b-49b9-d5a2-f996f61debef" |
|  | }, |
|  | "execution\_count": 40, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | "array([[-0.32687761, -0.90188624, -1.09598752, ..., 0.64609167,\n", |
|  | " 0.97024255, 0.02188649],\n", |
|  | " [-0.44080365, 1.51506738, -1.09598752, ..., -1.54776799,\n", |
|  | " 0.97024255, 0.21653375],\n", |
|  | " [-1.53863634, -0.90188624, -1.09598752, ..., 0.64609167,\n", |
|  | " -1.03067011, 0.2406869 ],\n", |
|  | " ...,\n", |
|  | " [ 0.60524449, -0.90188624, -1.09598752, ..., -1.54776799,\n", |
|  | " 0.97024255, -1.00864308],\n", |
|  | " [ 1.25772996, 0.30659057, 0.91241915, ..., 0.64609167,\n", |
|  | " -1.03067011, -0.12523071],\n", |
|  | " [ 1.4648682 , -0.90188624, -1.09598752, ..., 0.64609167,\n", |
|  | " -1.03067011, -1.07636976]])" |
|  | ] |
|  | }, |
|  | "metadata": {}, |
|  | "execution\_count": 40 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "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=0.33)\n", |
|  | "x\_train.shape" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/" |
|  | }, |
|  | "id": "W5RG9ibx43hT", |
|  | "outputId": "f7fea99c-04f2-4b96-eac8-586082d0c5be" |
|  | }, |
|  | "execution\_count": 41, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | "(6700, 10)" |
|  | ] |
|  | }, |
|  | "metadata": {}, |
|  | "execution\_count": 41 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "x\_test.shape" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/" |
|  | }, |
|  | "id": "eVLr7zta47gq", |
|  | "outputId": "1b65fe69-ad5f-4af0-e7bb-2a91aa5374ea" |
|  | }, |
|  | "execution\_count": 42, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | "(3300, 10)" |
|  | ] |
|  | }, |
|  | "metadata": {}, |
|  | "execution\_count": 42 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "y\_train.shape" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/" |
|  | }, |
|  | "id": "JqH9-c4d47l-", |
|  | "outputId": "05d84cb3-1157-4821-c958-9fa0a3d0ad28" |
|  | }, |
|  | "execution\_count": 43, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | "(6700,)" |
|  | ] |
|  | }, |
|  | "metadata": {}, |
|  | "execution\_count": 43 |
|  | } |
|  | ] |
|  | }, |
|  | { |
|  | "cell\_type": "code", |
|  | "source": [ |
|  | "y\_test.shape" |
|  | ], |
|  | "metadata": { |
|  | "colab": { |
|  | "base\_uri": "https://localhost:8080/" |
|  | }, |
|  | "id": "mEZHXROH4\_lg", |
|  | "outputId": "3cd8ec97-8ad3-49b0-aee1-3b42e8b370eb" |
|  | }, |
|  | "execution\_count": 44, |
|  | "outputs": [ |
|  | { |
|  | "output\_type": "execute\_result", |
|  | "data": { |
|  | "text/plain": [ |
|  | "(3300,)" |
|  | ] |
|  | }, |
|  | "metadata": {}, |
|  | "execution\_count": 44 |
|  | } |
|  | ] |
|  | } |
|  | ], |
|  | "metadata": { |
|  | "accelerator": "GPU", |
|  | "colab": { |
|  | "collapsed\_sections": [], |
|  | "provenance": [] |
|  | }, |
|  | "kernelspec": { |
|  | "display\_name": "Python 3", |
|  | "name": "python3" |
|  | } |
|  | }, |
|  | "nbformat": 4, |
|  | "nbformat\_minor": 0 |
|  | } |