{

"cells": [

  {

   "cell\_type": "code",

   "execution\_count": 1,

   "id": "33bfe219",

   "metadata": {},

   "outputs": [],

   "source": [

    "## Importing the necessary libraries\n",

    "\n",

    "import numpy as np\n",

    "import pandas as pd\n",

    "import matplotlib.pyplot as plt\n",

    "import seaborn as sns"

   ]

  },

  {

   "cell\_type": "code",

   "execution\_count": 2,

   "id": "c3169ac4",

   "metadata": {},

   "outputs": [],

   "source": [

    "# Loading the data set\n",

    "\n",

    "df=pd.read\_csv(\"Walmart DataSet.csv\")"

   ]

  },

  {

   "cell\_type": "code",

   "execution\_count": 3,

   "id": "e2f35174",

   "metadata": {},

   "outputs": [

    {

     "data": {

      "text/html": [

       "<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>Store</th>\n",

       "      <th>Date</th>\n",

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

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

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

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

       "      <th>CPI</th>\n",

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

       "    </tr>\n",

       "  </thead>\n",

       "  <tbody>\n",

       "    <tr>\n",

       "      <th>0</th>\n",

       "      <td>1</td>\n",

       "      <td>05-02-2010</td>\n",

       "      <td>1643690.90</td>\n",

       "      <td>0</td>\n",

       "      <td>42.31</td>\n",

       "      <td>2.572</td>\n",

       "      <td>211.096358</td>\n",

       "      <td>8.106</td>\n",

       "    </tr>\n",

       "    <tr>\n",

       "      <th>1</th>\n",

       "      <td>1</td>\n",

       "      <td>12-02-2010</td>\n",

       "      <td>1641957.44</td>\n",

       "      <td>1</td>\n",

       "      <td>38.51</td>\n",

       "      <td>2.548</td>\n",

       "      <td>211.242170</td>\n",

       "      <td>8.106</td>\n",

       "    </tr>\n",

       "    <tr>\n",

       "      <th>2</th>\n",

       "      <td>1</td>\n",

       "      <td>19-02-2010</td>\n",

       "      <td>1611968.17</td>\n",

       "      <td>0</td>\n",

       "      <td>39.93</td>\n",

       "      <td>2.514</td>\n",

       "      <td>211.289143</td>\n",

       "      <td>8.106</td>\n",

       "    </tr>\n",

       "    <tr>\n",

       "      <th>3</th>\n",

       "      <td>1</td>\n",

       "      <td>26-02-2010</td>\n",

       "      <td>1409727.59</td>\n",

       "      <td>0</td>\n",

       "      <td>46.63</td>\n",

       "      <td>2.561</td>\n",

       "      <td>211.319643</td>\n",

       "      <td>8.106</td>\n",

       "    </tr>\n",

       "    <tr>\n",

       "      <th>4</th>\n",

       "      <td>1</td>\n",

       "      <td>05-03-2010</td>\n",

       "      <td>1554806.68</td>\n",

       "      <td>0</td>\n",

       "      <td>46.50</td>\n",

       "      <td>2.625</td>\n",

       "      <td>211.350143</td>\n",

       "      <td>8.106</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",

       "    </tr>\n",

       "    <tr>\n",

       "      <th>6430</th>\n",

       "      <td>45</td>\n",

       "      <td>28-09-2012</td>\n",

       "      <td>713173.95</td>\n",

       "      <td>0</td>\n",

       "      <td>64.88</td>\n",

       "      <td>3.997</td>\n",

       "      <td>192.013558</td>\n",

       "      <td>8.684</td>\n",

       "    </tr>\n",

       "    <tr>\n",

       "      <th>6431</th>\n",

       "      <td>45</td>\n",

       "      <td>05-10-2012</td>\n",

       "      <td>733455.07</td>\n",

       "      <td>0</td>\n",

       "      <td>64.89</td>\n",

       "      <td>3.985</td>\n",

       "      <td>192.170412</td>\n",

       "      <td>8.667</td>\n",

       "    </tr>\n",

       "    <tr>\n",

       "      <th>6432</th>\n",

       "      <td>45</td>\n",

       "      <td>12-10-2012</td>\n",

       "      <td>734464.36</td>\n",

       "      <td>0</td>\n",

       "      <td>54.47</td>\n",

       "      <td>4.000</td>\n",

       "      <td>192.327265</td>\n",

       "      <td>8.667</td>\n",

       "    </tr>\n",

       "    <tr>\n",

       "      <th>6433</th>\n",

       "      <td>45</td>\n",

       "      <td>19-10-2012</td>\n",

       "      <td>718125.53</td>\n",

       "      <td>0</td>\n",

       "      <td>56.47</td>\n",

       "      <td>3.969</td>\n",

       "      <td>192.330854</td>\n",

       "      <td>8.667</td>\n",

       "    </tr>\n",

       "    <tr>\n",

       "      <th>6434</th>\n",

       "      <td>45</td>\n",

       "      <td>26-10-2012</td>\n",

       "      <td>760281.43</td>\n",

       "      <td>0</td>\n",

       "      <td>58.85</td>\n",

       "      <td>3.882</td>\n",

       "      <td>192.308899</td>\n",

       "      <td>8.667</td>\n",

       "    </tr>\n",

       "  </tbody>\n",

       "</table>\n",

       "<p>6435 rows × 8 columns</p>\n",

       "</div>"

      ],

      "text/plain": [

       "      Store        Date  Weekly\_Sales  Holiday\_Flag  Temperature  Fuel\_Price  \\\n",

       "0         1  05-02-2010    1643690.90             0        42.31       2.572   \n",

       "1         1  12-02-2010    1641957.44             1        38.51       2.548   \n",

       "2         1  19-02-2010    1611968.17             0        39.93       2.514   \n",

       "3         1  26-02-2010    1409727.59             0        46.63       2.561   \n",

       "4         1  05-03-2010    1554806.68             0        46.50       2.625   \n",

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

       "6430     45  28-09-2012     713173.95             0        64.88       3.997   \n",

       "6431     45  05-10-2012     733455.07             0        64.89       3.985   \n",

       "6432     45  12-10-2012     734464.36             0        54.47       4.000   \n",

       "6433     45  19-10-2012     718125.53             0        56.47       3.969   \n",

       "6434     45  26-10-2012     760281.43             0        58.85       3.882   \n",

       "\n",

       "             CPI  Unemployment  \n",

       "0     211.096358         8.106  \n",

       "1     211.242170         8.106  \n",

       "2     211.289143         8.106  \n",

       "3     211.319643         8.106  \n",

       "4     211.350143         8.106  \n",

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

       "6430  192.013558         8.684  \n",

       "6431  192.170412         8.667  \n",

       "6432  192.327265         8.667  \n",

       "6433  192.330854         8.667  \n",

       "6434  192.308899         8.667  \n",

       "\n",

       "[6435 rows x 8 columns]"

      ]

     },

     "execution\_count": 3,

     "metadata": {},

     "output\_type": "execute\_result"

    }

   ],

   "source": [

    "df"

   ]

  },

  {

   "cell\_type": "code",

   "execution\_count": 4,

   "id": "d8de5f3b",

   "metadata": {},

   "outputs": [

    {

     "name": "stdout",

     "output\_type": "stream",

     "text": [

      "<class 'pandas.core.frame.DataFrame'>\n",

      "RangeIndex: 6435 entries, 0 to 6434\n",

      "Data columns (total 8 columns):\n",

      " #   Column        Non-Null Count  Dtype  \n",

      "---  ------        --------------  -----  \n",

      " 0   Store         6435 non-null   int64  \n",

      " 1   Date          6435 non-null   object \n",

      " 2   Weekly\_Sales  6435 non-null   float64\n",

      " 3   Holiday\_Flag  6435 non-null   int64  \n",

      " 4   Temperature   6435 non-null   float64\n",

      " 5   Fuel\_Price    6435 non-null   float64\n",

      " 6   CPI           6435 non-null   float64\n",

      " 7   Unemployment  6435 non-null   float64\n",

      "dtypes: float64(5), int64(2), object(1)\n",

      "memory usage: 402.3+ KB\n"

     ]

    }

   ],

   "source": [

    "# Checking the info of dataframe\n",

    "\n",

"[df.info](http://df.info)()"

   ]

  },

  {

   "cell\_type": "code",

   "execution\_count": 5,

   "id": "972e550d",

   "metadata": {},

   "outputs": [

    {

     "data": {

      "text/plain": [

       "(6435, 8)"

      ]

     },

     "execution\_count": 5,

     "metadata": {},

     "output\_type": "execute\_result"

    }

   ],

   "source": [

    "df.shape"

   ]

  },

  {

   "cell\_type": "code",

   "execution\_count": 6,

   "id": "87f48628",

   "metadata": {},

   "outputs": [

    {

     "data": {

      "text/plain": [

       "Store           0\n",

       "Date            0\n",

       "Weekly\_Sales    0\n",

       "Holiday\_Flag    0\n",

       "Temperature     0\n",

       "Fuel\_Price      0\n",

       "CPI             0\n",

       "Unemployment    0\n",

       "dtype: int64"

      ]

     },

     "execution\_count": 6,

     "metadata": {},

     "output\_type": "execute\_result"

    }

   ],

   "source": [

    "# Checking for null values\n",

    "\n",

    "df.isnull().sum()"

   ]

  },

  {

   "cell\_type": "code",

   "execution\_count": 7,

   "id": "186fed43",

   "metadata": {},

   "outputs": [

    {

     "data": {

      "text/plain": [

       "0"

      ]

     },

     "execution\_count": 7,

     "metadata": {},

     "output\_type": "execute\_result"

    }

   ],

   "source": [

    "# Checking for duplicate values\n",

    "\n",

    "df.duplicated().sum()"

   ]

  },

  {

   "cell\_type": "markdown",

   "id": "73d4c48b",

   "metadata": {},

   "source": [

    "# QUESTION 1 : You are provided with the weekly sales data for their various outlets. Use statistical analysis, EDA, outlier analysis, and handle the missing values to come up with various insights that can give them a clear perspective on the following:\n"

   ]

  },

  {

   "cell\_type": "markdown",

   "id": "1397aa30",

   "metadata": {},

   "source": [

    "# Question 'a': If the weekly sales are affected by the unemployment rate, if yes - which stores are suffering the most?"

   ]

  },

  {

   "cell\_type": "code",

   "execution\_count": 8,

   "id": "2d253732",

   "metadata": {},

   "outputs": [

    {

     "data": {

      "image/png": "