Assignment No.13

```
import pandas as pd
import tensorflow as tf
import numpy as np
#Creating dataframe
df = pd.read_csv('sales_data_sample 2.csv', encoding = "latin")


dfcsv = pd.DataFrame(df)
#Printing first 5 rows of dataset
dfcsv.head(5)
```

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	...	ADDRESSLINE1
0	10107	30	95.70	2	2871.00	2/24/2003 0:00	Shipped	1	2	2003	...	897 Long Airport Avenue
1	10121	34	81.35	5	2765.90	5/7/2003 0:00	Shipped	2	5	2003	...	59 rue de l'Abbaye
2	10134	41	94.74	2	3884.34	7/1/2003 0:00	Shipped	3	7	2003	...	27 rue du Colonel Pierre Avia
3	10145	45	83.26	6	3746.70	8/25/2003 0:00	Shipped	3	8	2003	...	78934 Hillside Dr.
4	10159	49	100.00	14	5205.27	10/10/2003 0:00	Shipped	4	10	2003	...	7734 Strong St.

5 rows × 25 columns

```
dfcsv.shape
```

```
(2823, 25)
```

```
dfcsv.isna().sum()
```

```
ORDERNUMBER          0
QUANTITYORDERED      0
PRICEEACH            0
ORDERLINENUMBER      0
SALES                0
ORDERDATE            0
STATUS               0
QTR_ID               0
MONTH_ID             0
YEAR_ID              0
PRODUCTLINE          0
MSRP                 0
PRODUCTCODE          0
CUSTOMERNAME         0
PHONE                0
ADDRESSLINE1         0
ADDRESSLINE2      2521
CITY                 0
STATE             1486
POSTALCODE          76
COUNTRY              0
TERRITORY         1074
CONTACTLASTNAME      0
CONTACTFIRSTNAME     0
DEALSIZE             0
dtype: int64
```

```
dfcsv.describe
```

```
<bound method NDFrame.describe of       ORDERNUMBER  QUANTITYORDERED  PRICEEACH  ORDERLINENUMBER    SALES  \
0           10107               30      95.70                2  2871.00
1           10121               34      81.35                5  2765.90
2           10134               41      94.74                2  3884.34
3           10145               45      83.26                6  3746.70
4           10159               49     100.00               14  5205.27
...           ...              ...        ...              ...      ...
2818        10350               20     100.00               15  2244.40
2819        10373               29     100.00                1  3978.51
2820        10386               43     100.00                4  5417.57
2821        10397               34      62.24                1  2116.16
2822        10414               47      65.52                9  3079.44

            ORDERDATE    STATUS  QTR_ID  MONTH_ID  YEAR_ID  ...  \
0      2/24/2003 0:00   Shipped       1         2     2003  ...
1       5/7/2003 0:00   Shipped       2         5     2003  ...
2       7/1/2003 0:00   Shipped       3         7     2003  ...
3      8/25/2003 0:00   Shipped       3         8     2003  ...
4     10/10/2003 0:00   Shipped       4        10     2003  ...
...               ...       ...     ...       ...      ...  ...
2818   12/2/2004 0:00   Shipped       4        12     2004  ...
2819   1/31/2005 0:00   Shipped       1         1     2005  ...
2820    3/1/2005 0:00  Resolved       1         3     2005  ...
2821   3/28/2005 0:00   Shipped       1         3     2005  ...
2822    5/6/2005 0:00   On Hold       2         5     2005  ...

                     ADDRESSLINE1  ADDRESSLINE2           CITY STATE  \
0           897 Long Airport Avenue           NaN            NYC    NY
1               59 rue de l'Abbaye           NaN          Reims   NaN
2     27 rue du Colonel Pierre Avia           NaN          Paris   NaN
3               78934 Hillside Dr.           NaN       Pasadena    CA
4                  7734 Strong St.           NaN  San Francisco    CA
...                            ...           ...            ...   ...
2818            C/ Moralzarzal, 86           NaN         Madrid   NaN
2819                 Torikatu 38           NaN           Oulu   NaN
2820            C/ Moralzarzal, 86           NaN         Madrid   NaN
2821         1 rue Alsace-Lorraine           NaN       Toulouse   NaN
2822            8616 Spinnaker Dr.           NaN         Boston    MA

     POSTALCODE  COUNTRY TERRITORY CONTACTLASTNAME CONTACTFIRSTNAME DEALSIZE
0         10022      USA       NaN             Yu             Kwai    Small
1         51100   France      EMEA        Henriot             Paul    Small
2         75508   France      EMEA       Da Cunha           Daniel   Medium
3         90003      USA       NaN          Young            Julie   Medium
4           NaN      USA       NaN          Brown            Julie   Medium
...         ...      ...       ...            ...              ...      ...
2818      28034    Spain      EMEA         Freyre            Diego    Small
2819      90110  Finland      EMEA      Koskitalo           Pirkko   Medium
2820      28034    Spain      EMEA         Freyre            Diego   Medium
2821      31000   France      EMEA         Roulet          Annette    Small
2822      51003      USA       NaN        Yoshido             Juri   Medium

[2823 rows x 25 columns]>
```

```
dfcsv = dfcsv.drop(['ADDRESSLINE1','ADDRESSLINE2','CITY','STATE','TERRITORY'],axis = 1)
```

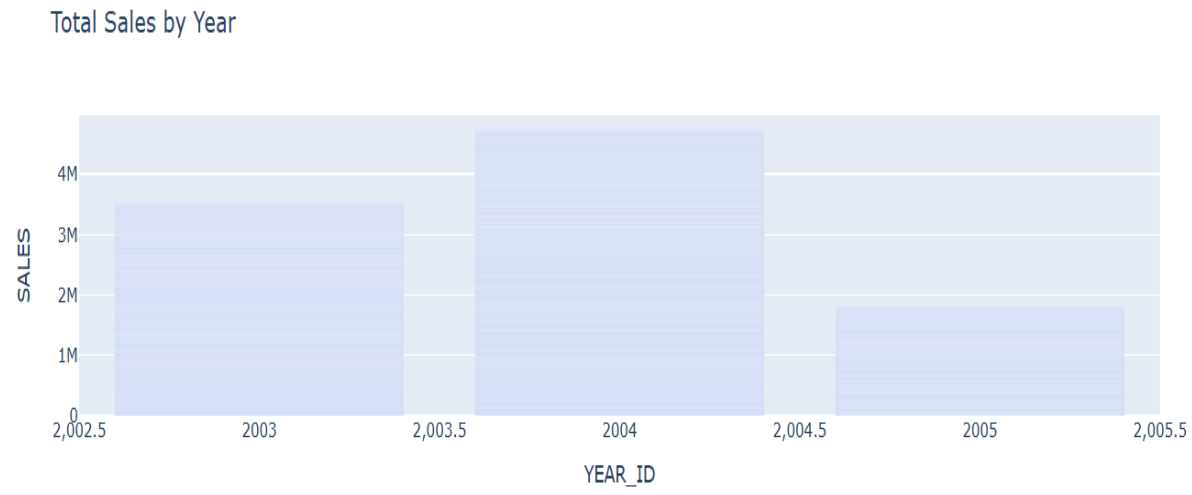
```
dfcsv.isna().sum()
```

```
ORDERNUMBER          0
QUANTITYORDERED      0
PRICEEACH            0
ORDERLINENUMBER      0
SALES                0
ORDERDATE            0
STATUS               0
QTR_ID               0
MONTH_ID             0
YEAR_ID              0
PRODUCTLINE          0
MSRP                 0
PRODUCTCODE          0
CUSTOMERNAME         0
PHONE                0
POSTALCODE          76
COUNTRY              0
CONTACTLASTNAME      0
CONTACTFIRSTNAME     0
DEALSIZE             0
dtype: int64
```

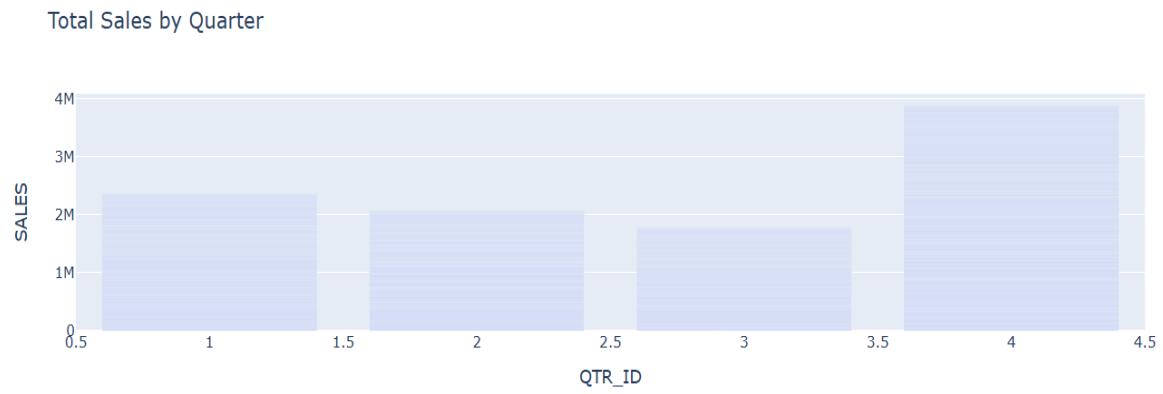
```
dfcsv = dfcsv['POSTALCODE'].fillna(dfcsv.POSTALCODE.mode(), inplace=True)
columns_to_drop = ['ADDRESSLINE1', 'ADDRESSLINE2', 'CITY', 'STATE', 'TERRITORY']
```

```
import plotly.express as px

fig = px.bar(df, x='YEAR_ID', y='SALES', title='Total Sales by Year')
fig.show()
```



```
fig = px.bar(df, x='QTR_ID', y='SALES', title='Total Sales by Quarter')
fig.show()
```




```
fig = px.line(df, x='ORDERDATE', y='SALES', title='Sales Over Time')
fig.show()
```