

In [2]:
`import pandas as pd
import numpy as np
import seaborn as sns`

In [7]:
`df=pd.read_csv('C:/Users/ambalkt/Downloads/Power BI Practice/Projects/Sales/Car_Sales.csv',sep=",", encoding='Latin-1')
df.head(5)`

Out[7]:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	...	ADDRESSLINE1	ADDRESSLINE2	CI
0	10107	30	95.70	2	2871.00	2/24/2003 0:00	Shipped	1	2	2003	...	897 Long Airport Avenue	NaN	N
1	10121	34	81.35	5	2765.90	5/7/2003 0:00	Shipped	2	5	2003	...	59 rue de l'Abbaye	NaN	Rei
2	10134	41	94.74	2	3884.34	7/1/2003 0:00	Shipped	3	7	2003	...	27 rue du Colonel Pierre Avia	NaN	Pe
3	10145	45	83.26	6	3746.70	8/25/2003 0:00	Shipped	3	8	2003	...	78934 Hillside Dr.	NaN	Pasade
4	10159	49	100.00	14	5205.27	10/10/2003 0:00	Shipped	4	10	2003	...	7734 Strong St.	NaN	S Francis

5 rows × 25 columns

In [8]:
`df.isnull().sum()`

Out[8]:

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

In [9]:
`mapping = (df.query('POSTALCODE != "NaN"')
 .drop_duplicates('CITY')
 .set_index('CITY')
 .POSTALCODE)

mapping`

Out[9]:

CITY	
NYC	10022
Reims	51100
Paris	75508
Pasadena	90003
San Francisco	NaN
...	
Oslo	N 0106
Bergamo	24100
Glen Waverly	3150
Munich	80686
Charleroi	B-6000

Name: POSTALCODE, Length: 73, dtype: object

In []:
`df['POSTALCODE'] = df['CITY'].map(mapping).fillna('NaN')

df`

In [27]:
`df['POSTALCODE'][df.CITY == "San Francisco"] = 94016
df['POSTALCODE'].head(5)`

C:\Users\ambalkt\AppData\Local\Temp\1\ipykernel_15148\1981405433.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
`df['POSTALCODE'][df.CITY == "San Francisco"] = 94016`

Out[27]:

0	10022
1	51100
2	75508
3	90003
4	94016

Name: POSTALCODE, dtype: object

In [26]:
`df['POSTALCODE'].head(5)`

Out[26]:

0	10022
1	51100
2	75508
3	90003
4	400028

Name: POSTALCODE, dtype: object

In []: