What is EDA

Exploratory Data Analysis refers to the critical process of performing initial investigation on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations

importing libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
```

Reading the data

In [2]:
```
df=pd.read_csv(r'C:\Users\Anand Mishra\Downloads\Indian_railway1.csv')
df
```

Out[2]:

	Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Stat
0	107	SWV-MAO-VLNK	1	SWV	SAWANTWADIR	00:00:00	10:25:00	0	SWV	SAW

	Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Stat
1	107	SWV-MAO-VLNK	2	THVM	THIVIM	11:06:00	11:08:00	32	SWV	SAW
2	107	SWV-MAO-VLNK	3	KRMI	KARMALI	11:28:00	11:30:00	49	SWV	SAW
3	107	SWV-MAO-VLNK	4	MAO	MADGOAN JN.	12:10:00	00:00:00	78	SWV	SAW
4	108	VLNK-MAO-SWV	1	MAO	MADGOAN JN.	00:00:00	20:30:00	0	MAO	M
...	...	...	...	...	...	...	...	...	...	
186119	99908	EMU	8	AKRD	AKURDI	23:30:00	23:31:00	19	PUNE	
186120	99908	EMU	9	DEHR	DEHU ROAD	23:35:00	23:36:00	24	PUNE	
186121	99908	EMU	10	BGWI	BEGDAEWAI	23:39:00	23:40:00	28	PUNE	
186122	99908	EMU	11	GRWD	GHORAWADI	23:41:00	23:42:00	31	PUNE	
186123	99908	EMU	12	TGN	TALEGAON	23:50:00	00:00:00	34	PUNE	

186124 rows × 12 columns

In [3]:
```
df=pd.read_csv(r'C:\Users\Anand Mishra\Downloads\Indian_railway1.csv',
index_col=0)
df
```

Out[3]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
107	SWV-MAO-	1	SWV	SAWANTWADI R	00:00:00	10:25:00	0	SWV	SAWANTWADI ROAD

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
107	SWV-MAO-VLNK	2	THVM	THIVIM	11:06:00	11:08:00	32	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	3	KRMI	KARMALI	11:28:00	11:30:00	49	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	4	MAO	MADGOAN JN.	12:10:00	00:00:00	78	SWV	SAWANTWADI ROAD
108	VLNK-MAO-SWV	1	MAO	MADGOAN JN.	00:00:00	20:30:00	0	MAO	MADGOAN JN.
...	...	...	...	...	...	...	...	...	...
99908	EMU	8	AKRD	AKURDI	23:30:00	23:31:00	19	PUNE	PUNE JN.
99908	EMU	9	DEHR	DEHU ROAD	23:35:00	23:36:00	24	PUNE	PUNE JN.
99908	EMU	10	BGWI	BEGDAEWAI	23:39:00	23:40:00	28	PUNE	PUNE JN.
99908	EMU	11	GRWD	GHORAWADI	23:41:00	23:42:00	31	PUNE	PUNE JN.
99908	EMU	12	TGN	TALEGAON	23:50:00	00:00:00	34	PUNE	PUNE JN.

186124 rows × 11 columns

To view all data

In [5]:
```
df
```

Out[5]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
107	SWV-MAO-VLNK	1	SWV	SAWANTWADI R	00:00:00	10:25:00	0	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	2	THVM	THIVIM	11:06:00	11:08:00	32	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	3	KRMI	KARMALI	11:28:00	11:30:00	49	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	4	MAO	MADGOAN JN.	12:10:00	00:00:00	78	SWV	SAWANTWADI ROAD
108	VLNK-MAO-SWV	1	MAO	MADGOAN JN.	00:00:00	20:30:00	0	MAO	MADGOAN JN.
...	...	...	...	...	...	...	...	...	...
99908	EMU	8	AKRD	AKURDI	23:30:00	23:31:00	19	PUNE	PUNE JN.
99908	EMU	9	DEHR	DEHU ROAD	23:35:00	23:36:00	24	PUNE	PUNE JN.
99908	EMU	10	BGWI	BEGDAEWAI	23:39:00	23:40:00	28	PUNE	PUNE JN.
99908	EMU	11	GRWD	GHORAWADI	23:41:00	23:42:00	31	PUNE	PUNE JN.
99908	EMU	12	TGN	TALEGAON	23:50:00	00:00:00	34	PUNE	PUNE JN.

186124 rows × 11 columns

Describe

This function returns the count, mean, standard deviation, minimum and maximum values and the quantiles of the data.

In [6]:
```
df.describe()
```

Out[6]:

	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name	Dest
count	186124	186124	186124	186124	186119	186119	186119	186114	186114	
unique	7585	235	8151	8100	1443	1444	3829	926	921	
top	MSB-TBM EMU	2	CSMT	CST-MUMBAI	00:00:00	00:00:00	0	HWH	HOWRAH JN.	
freq	2436	7186	1027	1027	2002	1969	7191	7977	7977	

print all statistical information

In [7]:
```
df.describe(include='all')
```

Out[7]:

	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name	Dest
count	186124	186124	186124	186124	186119	186119	186119	186114	186114	
unique	7585	235	8151	8100	1443	1444	3829	926	921	
top	MSB-TBM EMU	2	CSMT	CST-MUMBAI	00:00:00	00:00:00	0	HWH	HOWRAH JN.	
freq	2436	7186	1027	1027	2002	1969	7191	7977	7977	

info()

print the information of the dataset

In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 186124 entries, 107 to 99908
Data columns (total 11 columns):
Train Name                  186124 non-null object
SEQ                         186124 non-null object
Station Code                186124 non-null object
Station Name                186124 non-null object
Arrival time                186119 non-null object
Departure Time              186119 non-null object
Distance                    186119 non-null object
Source Station              186114 non-null object
Source Station Name         186114 non-null object
Destination Station         186114 non-null object
Destination Station Name    186114 non-null object
dtypes: object(11)
memory usage: 17.0+ MB
```

head()

To print top 5 rows

In [9]:
```
df.head()
```

Out[9]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
107	SWV-MAO-VLNK	1	SWV	SAWANTWADIR	00:00:00	10:25:00	0	SWV	SAWANTWADI ROAD

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
107	SWV-MAO-VLNK	2	THVM	THIVIM	11:06:00	11:08:00	32	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	3	KRMI	KARMALI	11:28:00	11:30:00	49	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	4	MAO	MADGOAN JN.	12:10:00	00:00:00	78	SWV	SAWANTWADI ROAD
108	VLNK-MAO-SWV	1	MAO	MADGOAN JN.	00:00:00	20:30:00	0	MAO	MADGOAN JN.

tail()

To print bottom 5 rows

In [10]:
```
df.tail()
```

Out[10]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name	Destin St
99908	EMU	8	AKRD	AKURDI	23:30:00	23:31:00	19	PUNE	PUNE JN.	
99908	EMU	9	DEHR	DEHU ROAD	23:35:00	23:36:00	24	PUNE	PUNE JN.	

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name	Destin St
99908	EMU	10	BGWI	BEGDAEWAI	23:39:00	23:40:00	28	PUNE	PUNE JN.	
99908	EMU	11	GRWD	GHORAWADI	23:41:00	23:42:00	31	PUNE	PUNE JN.	
99908	EMU	12	TGN	TALEGAON	23:50:00	00:00:00	34	PUNE	PUNE JN.	

shape

To ckeck how many row and colunm

In [11]:
```
df.shape
```

Out[11]:
```
(186124, 11)
```

dtypes

To check data type of attribute

In [12]:
```
df.dtypes
```

Out[12]:
```
Train Name                object
SEQ                       object
Station Code              object
Station Name              object
Arrival time              object
Departure Time            object
Distance                  object
```

```
Source Station              object
Source Station Name         object
Destination Station         object
Destination Station Name    object
dtype: object
```

index

To print the index

In [13]:
```
df.index
```

Out[13]:
```
Index([  '107',   '107',   '107',   '107',   '108',   '108',   '108',
'108',
         '128',   '128',
        ...
       '99908', '99908', '99908', '99908', '99908', '99908', '99908',
'99908',
       '99908', '99908'],
      dtype='object', name='Train No', length=186124)
```

To check is ther any missing value

In [14]:
```
df.isna()
```

Out[14]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name	Destination Station
107	False	False	False	False	False	False	False	False	False	False
107	False	False	False	False	False	False	False	False	False	False
107	False	False	False	False	False	False	False	False	False	False

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name	Destination Station
107	False	False	False	False	False	False	False	False	False	False
108	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
99908	False	False	False	False	False	False	False	False	False	False
99908	False	False	False	False	False	False	False	False	False	False
99908	False	False	False	False	False	False	False	False	False	False
99908	False	False	False	False	False	False	False	False	False	False
99908	False	False	False	False	False	False	False	False	False	False

186124 rows × 11 columns

Sum of missing value

In [15]:
```
df.isna().sum()
```

Out[15]:
```
Train Name                  0
SEQ                         0
Station Code                0
Station Name                0
Arrival time                5
Departure Time              5
Distance                    5
Source Station             10
Source Station Name        10
Destination Station        10
Destination Station Name   10
dtype: int64
```

Sum of null value

In [16]:
```
df.isnull().sum()
```

Out[16]:
```
Train Name                  0
SEQ                         0
Station Code                0
Station Name                0
Arrival time                5
Departure Time              5
Distance                    5
Source Station             10
Source Station Name        10
Destination Station        10
Destination Station Name   10
dtype: int64
```

To check the duplicated value

In [17]:
```
df.duplicated()
```

Out[17]:
```
Train No
107      False
107      False
107      False
107      False
108      False
         ...
99908    False
99908    False
99908    False
99908    False
99908    False
Length: 186124, dtype: bool
```

Number of unique values in data set

In [18]:
```
df.nunique()
```

Out[18]:
```
Train Name                  7585
SEQ                          235
Station Code                8151
Station Name                8100
Arrival time                1443
Departure Time              1444
Distance                    3829
Source Station               926
Source Station Name          921
Destination Station          928
Destination Station Name     923
dtype: int64
```

To find unique station name

In [39]:
```
df['Station Name'].unique()
```

Out[39]:
```
array(['SAWANTWADI R', 'THIVIM', 'KARMALI', ..., 'MANKHURD', 'CHEMBUR',
       'NERUL'], dtype=object)
```

To find Destination Station Name

In [38]:
```
df['Destination Station Name'].unique()
```

Out[38]:
```
array(['MADGOAN JN.', 'SAWANTWADI ROAD',
       'CHHATRAPATI SHAHU MAHARAJ TERMINUS', 'DELHI-SAFDAR JANG',
       'VARANASI JN.', 'SHRI MATA VAISHNO DEVI KATRA', 'LUCKNOW JN.',
       'SIRSA', 'AMBALA CANTT JN', 'BATHINDA JN', 'PATIALA', 'PANJIPAR
A',
```

```
                'AMRITSAR JN.', 'NAGPUR JN.(CR)', 'LOKMANYA TILAK TERMINUS',
                'KARMALI', 'AJNI', 'CST-MUMBAI', 'MANGALORE JN', 'PUNE JN.',
                'KHADKI', 'JABALPUR', 'BANDRA TERMINUS', 'ATARI JN',
                'SANTRAGACHI JN.', 'BHOPAL', 'REWA', 'SILCHAR', 'POLLACHI',
                'PALANI', 'JAIPUR JN.', 'HYDERABAD DECCAN', 'RAJKOT',
                'CHENNAI CENTRAL', 'HATIA', 'PURI', 'KOLKATA', 'DHANBAD JN.',
                'ASANSOL MAIN', 'KUSUNDA', 'PHULWARTANR', 'SONARDIH',
                'ANAND VIHAR TERMINAL', 'JASIDIH JN.', 'DELHI JN.',
                'HAZRAT NIZAMUDDIN JN', 'HARIDWAR JN', 'BIKANER JN.',
                'SHRI GANGANAGAR', 'HANUMANGARH JN.', 'SADULPUR JN',
                'SURATGARH JN.', 'MAKRANA JN.', 'PARVATSAR CITY', 'JODHPUR JN.',
                'RATANAGARH JN', 'MERTA ROAD JN.', 'CHURU', 'SAHARANPUR JN.',
                'JAMMU TAWI', 'PATHANKOT JN.', 'BARHNI', 'GORAKHPUR JN.',
                'CHHAPRA JN.', 'BASTI', 'FARRUKHABAD JN', 'KANPUR ANWARGANJ',
                'LALKUA JN.', 'BAREILLY CITY', 'SAHARSA JN.', 'JAMALPUR JN.',
                'KOTA', 'JHALAWAR CITY', 'MURKEONG SELEK', 'NORTH LAKHIMPUR',
                'GUWAHATI', 'KISHANGANJ', 'NEW JALPAIGURI JN', 'SILIGURI JN.',
                'KATIHAR JN.', 'RANGAPARA NORTH', 'RANGIYA JN.', 'TIRUNELVELI J
N',
                'CHENNAI EGMORE', 'PONDICHERRY', 'QUILON', 'KOCHUVELI',
                'MADURAI JN', 'CHENGALPATTU', 'COIMBATORE  JN.', 'PAYYANUR',
                'BANASWADI', 'MARIKUPPAM', 'PANDHARPUR', 'YESVANTPUR JN.',
                'ERNAKULAM. JN', 'VISAKHAPATNAM', 'BELGAUM', 'HOSUR',
                'BAIYYAPPANAHALLI', 'WHITE FIELD', 'DEVANHALLI', 'BILASPUR JN.',
                'KORBA', 'PALGHAT TOWN', 'NAGORE', 'KARAIKAL', 'BIJAPUR JN', na
n,
                'HUBLI JN.', 'RAXAUL JN.', 'DARBHANGA JN.', 'SECUNDERABAD JN.',
                'VIJAYWADA JN.', 'KAKINADA PORT', 'KAKINADA TOWN', 'KACHEGUDA',
                'NARASAPUR', 'AURANGABAD', 'NAGARSOL', 'TIRUPATI', 'TATANAGAR J
N',
                'NEW BONGAIGAON  JN.', 'KRISHNA RAJA PURAM', 'NANDED', 'NIZAMABA
D',
                'RENIGUNTA JN.', 'JHARGRAM', 'KHARAGPUR', 'BALASORE', 'SHALIMA
R',
                'BANKURA', 'BISHNUPUR', 'CHAKARADHARPUR', 'CHAIBASA',
                'SAMBALPUR JN.', 'KIRANDUL', 'CHANDAFORT', 'BALHARSHAH', 'RANCH
I',
                'HOWRAH JN.', 'GONDIA JN.', 'NEW DELHI', 'MUMBAI CENTRAL',
                'GANDHIDHAM JN.', 'AJMER JN.', 'UDAIPUR CITY', 'SAI NAGAR SHIRD
I',
```

```
        'DADAR', 'BHUBANESWAR', 'BHUSAVAL JN.', 'SOLAPUR', 'MYSORE JN.',
        'MIRAJ JN.', 'AHMEDABAD', 'AZAMGARH', 'SALEM JN.', 'FAIZABAD J
N.',
        'BIDAR', 'KAZIPET JN.', 'VERAVAL', 'BHAGAT-KI-KOTHI', 'NEW BHU
J',
        'GWALIOR JN', 'JHANSI JN', 'BALRAMPUR,', 'BARAUNI JN.',
        'INDORE BG', 'AMBIKAPUR', 'ITARSI',
        'KRANTIVIRA SANGOLLI RAYANNA (BENGALURU STATION)', 'GULBARGA',
        'AMRAVATI', 'SOMNATH', 'SINGRAULI', 'MHOW', 'ETAWAH',
        'KURUKSHETRA JN.', 'MATHURA JN.', 'HABIBGANJ', 'KALKA',
        'DEHRA DUN', 'BARBIL', 'PATNA JN.', 'KANPUR CENTRAL JN.',
        'AGRA  FORT', 'LUDHIANA JN.', 'KATHGODAM', 'MOGA', 'CHANDIGARH',
        'FIROZPUR CANTT JN.', 'AGRA CANTT', 'UNA HIMACHAL', 'JORHAT TOW
N',
        'RAIGARH', 'JALNA', 'TRIVANDRUM CENTRAL', 'CALICUT', 'KANNUR',
        'MAYILADUTURAI JN.', 'DIBRUGARH', 'NAHARLAGUN', 'MANMAD JN.',
        'PATLIPUTRA', 'SULTANPUR JN.', 'DANAPUR', 'PARTAPGARH JN.',
        'MUZAFFARPUR JN.', 'DELHI-SARAI ROHILLA', 'BHAGALPUR', 'SEALDA
H',
        'ALLAHABAD JN.', 'RAJENDRANAGAR TERMINAL', 'NANGAL DAM',
        'ALLAHABAD CITY', 'BOLPUR', 'RAMPUR HAT', 'HALDIBARI', 'JAISALME
R',
        'NEW ALIPURDUAR', 'GAYA JN.', 'RAJGIR', 'ISLAMPUR',
        'DIBRUGARH TOWN', 'HALDIA', 'MANIKPUR JN.', 'HAPA', 'JAMNAGAR',
        'JOGBANI', 'KAMAKHYA JN', 'BANGALORE CANTT.', 'RAIPUR JN.',
        'MANDUADIH', 'DURG', 'HISAR JN', 'JAYNAGAR', 'MANGALORE CENTRA
L',
        'KARAIKKUDI', 'KANNIYAKUMARI', 'TIRUCHIRAPPALLI JN.',
        'NAGERCOIL JN.', 'SENGOTTAI JN', 'METTUPALAIYAM JN.',
        'SRI SATYA SAI PRASANTI NILAYAM', 'TUTICORIN', 'GUNTUR JN.',
        'GUDUR JN.', 'DHARWAD', 'VASCO-DA-GAMA', 'VIKRABAD JN.',
        'MACHILIPATNAM', 'SIRPURKAGHAZ NAGAR', 'KARIM NAGAR', 'CHITTOO
R',
        'PURULIA JN.', 'BOKARO STEEL CITY', 'DIGHA FS', 'TITLA GARH J
N.',
        'BHOJUDIH JN.', 'BANGRIPOSI', 'BALANGIR', 'PORBANDAR', 'VALSAD',
        'SURAT', 'VADODARA JN.', 'DAHOD', 'BHAVNAGAR TERMINUS',
        'MALDA TOWN', 'BARDDHAMAN JN.', 'AZIMGANJ JN.', 'SIURI',
        'BALURGHAT', 'LALGOLA', 'BALLIA', 'GEDE', 'GHAZIPUR CITY',
        'PETRAPOLE', 'RADHIKAPUR', 'NEW COOCH BEHAR', 'ALIPUR DUAR JN.',
```

```
        'SITAMARHI', 'SAHIBGANJ', 'BANKA', 'BHABUA ROAD', 'DEOGHAR',
        'BARKAKANA', 'ALLEPPEY', 'NABADWIP DHAM', 'GONDA JN.',
        'HOSHIARPUR', 'AGARTALA', 'SIKAR JN.', 'KOTDWARA', 'TILAK BRIDG
E',
        'PRAYAG JN.', 'CHITRAKUTDHAM KARWI', 'ALIGARH JN.', 'MANKAPUR J
N.',
        'MEERUT CITY JN.', 'RAE BARELI JN.', 'JAUNPUR JN.', 'BAREILLY J
N.',
        'MUGHAL SARAI JN.', 'UJJAIN JN', 'ROHTAK JN.', 'CHOPAN',
        'BARWADIH JN', 'FARUKHNAGAR', 'AMB  ANDAURA', 'RISHIKESH',
        'CHHINWARA JN', 'BARMER', 'JALANDHAR CITY JN.', 'LALGARH JN.',
        'BHIWANI JN.', 'FAZILKA JN.', 'PANIPAT JN.', 'ALWAR JN.',
        'MUNABAO', 'VARANASI CITY', 'GOMTINAGAR F', 'NAUTANWA', 'MAU J
N.',
        'RAMNAGAR', 'KASGANJ JN.', 'OKHA', 'PANVEL', 'BADSHAHNAGAR',
        'RAMESWARAM', 'NARKATIAGANJ JN.', 'SILGHAT TOWN', 'UDHNA JN.',
        'LEDO', 'FURKATING JN.', 'TAMBARAM', 'DIMAPUR', 'MARIANI JN.',
        'DHUBRI', 'LUMDING JN.', 'DEKARGAON', 'NEW TINSUKIA JN', 'JHAJH
A',
        'JOLARPETTAI JN', 'ARAKKONAM JN', 'TIRUCHENDUR', 'GURUVAYUR',
        'VANCHI MANIYACHCHI JN', 'MANNARGUDI', 'THANJAVUR JN.',
        'VELANKANNI', 'NAGAPPATTINAM', 'TALGUPPA', 'SHORANUR JN.',
        'NILAMBUR ROAD', 'KAWR', 'BANGARPET JN.', 'HARIHAR JN.',
        'SHIMOGA TOWN', 'ERODE JN.', 'PALAKKAD JN', 'PUNALUR',
        'GUNTAKAL JN.', 'KURNOOL TOWN', 'MANUGURU', 'DHARMAVARAM JN.',
        'NIDADAVOLU JN.', 'LONDA JN.', 'BAGALKOT', 'GUDIVADA JN.',
        'ADILABAD', 'PURNA  JN.', 'REPALLE', 'AKOLA JN.', 'NARKHER',
        'DHARMABAD', 'JAGADALPUR', 'BHANJPUR', 'KORAPUT JN.', 'ROURKEL
A',
        'GUNUPUR', 'RAYAGADA JN.', 'MURI', 'PARADEEP', 'JUNAGARH ROAD',
        'DUMKA', 'PURNEA COURT', 'UDHAMPUR', 'PALANPUR JN',
        'GANDHINAGAR CAPITAL', 'RATLAM JN', 'NATHDWARA', 'MADAR JN',
        'KHAJAURAHO', 'CHITTAURGARH JN', 'MAHOBA', 'BHIND', 'DAMOH',
        'VILLUPURAM JN.', 'HASSAN JN.', 'ADRA', 'BHILAD', 'PALITANA',
        'MAHUVA', 'RANINAGAR JALPAIGURI', 'SHAKTINAGAR', 'GARWA ROAD J
N.',
        'NAJIBABAD JN.', 'MORADABAD JN.', 'SAGAULI JN.', 'MANDSOR',
        'BARRACKPORE', 'MAJERHAT', 'BENOY BADAL DINESH BAG', 'NAIHATI J
N.',
        'BALLYGUNGE JN.', 'NEW ALIPUR (CALCUTTA)', 'RANAGHAT JN.',
```

```
        'KRISHNA NAGAR CITY JN.', 'DUM DUM JN.', 'BARASAT JN.',
        'DATTAPUKUR', 'HASANABAD', 'PRINCEP GHAT', 'HABRA', 'MADHYAM GRA
M',
        'GHUTIARI SHARIF', 'BARUIPUR JN.', 'LAKSHMIKANTAPUR', 'KATWA J
N.',
        'KALYANI SIMANTA', 'SHANTIPUR JN.', 'DANKUNI JN.', 'BARUI PARA',
        'GOBARDANGA', 'BANGAON JN.', 'DUMDUM CANTONMENT', 'THAKURNAGAR',
        'BUDGE BUDGE', 'CANNING', 'SONARPUR JN.', 'KWAKDWIP', 'NAMKHAN
A',
        'DIAMOND HARBOUR', 'MAGRAHAT', 'BALGONA', 'SHRIPAT SHRIKHANDA',
        'CHANDANPUR', 'GURAP', 'MASAGRAM', 'SHRIRAMPUR', 'SEORAPHULI J
N.',
        'BELURMATH', 'BANDEL JN.', 'SINGUR', 'HARI PAL', 'TARAKESHWAR',
        'ARAMBAG', 'BALLY', 'PUNDOOAH', 'MEMARI', 'PANSKURA', 'ULUBARI
A',
        'MECHEDA', 'BALICHAK', 'KOLAGHAT', 'MIDNAPORE', 'AMTA',
        'CHENNAI BEACH', 'PALLAVARAM', 'ST. THOMAS MOUNT', 'TIRUMALPUR',
        'KANCHIPURAM', 'VELACHEERY', 'GUMMIDIPUNDI', 'MOOR MARKET',
        'PONNERI', 'SULLURUPETTA', 'ENNORE', 'AVADI', 'PATTABIRAM E DEPO
T',
        'TIRUVALLUR', 'TIRUTTANI', 'KADAMBATTUR', 'LINGAMPALLI',
        'FALAKNUMA', 'RATNAGIRI', 'DIVA JN', 'DAUND JN.', 'DHULE',
        'CHALISGAON JN.', 'BADNERA JN.', 'NEW AMRAVATI', 'DEVLALI',
        'KATNI', 'CHHEOKI', 'WARDHA JN.', 'BETUL', 'AMLA JN.', 'BORDHA
I',
        'KARJAT', 'BARAMATI', 'CASTLE ROCK', 'BELLARY JN.', 'IGATPURI',
        'PARLI VAIJNATH', 'SANGLI', 'SATARA', 'KURDUVADI JN.',
        'KATNI MURWARA', 'BINA', 'CHIRMIRI', 'GUNA', 'SATNA', 'BIR',
        'KHANDWA JN. (CR)', 'NAINPUR JN.', 'MADANMAHAL', 'ANUPPUR JN.',
        'BANDA JN', 'KONCH', 'AIT JN', 'SHAMLI', 'MAVLI JN.', 'MARWAR J
N.',
        'SHEOPUR KALAN', 'GWALIOR (NG)', 'SABALGARH', 'SIRMUTTRA',
        'DHOLPUR (NG)', 'TANTPUR', 'BARI', 'MAILANI JN.', 'PILIBHIT J
N.',
        'SHAHJAHANPUR JN', 'TIKUNIA', 'BAHRAICH', 'NEPALGANJ ROAD',
        'SHIMLA', 'JAWALAMUKHI ROAD', 'BAIJNATH PAPROLA', 'JOGINDER NAGA
R',
        'DARJEELING', 'KURSEONG', 'AHMEDABAD JN.', 'RANUJ',
        'DHASA JN (MG)', 'VERAVAL MG', 'JETALSAR JN (MG)', 'GANDHIGRAM',
        'BOTAD JN (MG)', 'DELVADA', 'JUNAGADH JN', 'KODINAR',
```

```
            'KHIJADIYA JN', 'SANAWAD', 'BARHARWA JN.', 'MOKAMA JN.', 'FATUH
        A',
            'SASARAM', 'BAKHTIYARPUR JN.', 'TILAIYA', 'SINDRI TOWN',
            'NSC BOSE J GOMO', 'DEHRI-ON-SON', 'CHUNAR JN.', 'KODERMA',
            'KAWAR', 'KIUL JN.', 'HANSDIHA', 'RAJMAHAL', 'TINPAHAR JN.',
            'GIRDIH', 'MADHUPUR JN.', 'DILDARNAGAR JN', 'TARIGHAT', 'JIND J
        N',
            'NARWANA JN.', 'JAKHAL JN.', 'SITAPUR CITY JN.', 'REWARI JN.',
            'FATEHPUR', 'RAHIMABAD', 'RAGHURAJ SINGH', 'UNCHAHAR JN.',
            'AYODHYA', 'ARA', 'CHANDAUSI JN.', 'SITAPUR CANTT', 'BALAMAU J
        N.',
            'BURHWAL JN.', 'ROZA JN', 'GAJRAULA JN.', 'SAMBAHAL HATIM SARA
        I',
            'KHURJA JN', 'MEERUT CANT', 'BANDIKUI JN.', 'TUNDLA JN', 'ETAH',
            'DHURI JN.', 'LOHIAN KHAS JN.', 'JAIJON DOABA', 'ABOHAR',
            'SAWAI MADHOPUR JN', 'BILARA', 'BHATNI JN.', 'SIWAN JN.',
            'SAMASTIPUR JN.', 'NAKAHA JUNGLE', 'BETTIAH', 'DALIGANJ JN',
            'KAPTANGANJ JN.', 'BARHAJ BAZAR', 'SALEMPUR JN.', 'THAWE JN.',
            'SHAHGANJ JN.', 'AUNRIHAR JN.', 'DURAUNDHA JN', 'MAHARAJGANJ',
            'CHHAPRA KACHEHRI', 'KASHIPUR JN', 'SHIKOHABAD JN.',
            'ACHHNERA JN.', 'BHARATPUR JN', 'MAJHOLA PAKARIYA', 'BAMANHAT',
            'HAJIPUR JN.', 'BATHUA BAZAR HALT', 'PURNEA JN.',
            'DAURAM MADHEPURA', 'BIRAUL', 'MAIRABARI', 'MAISHASHAN',
            'VANGAICHUNGPAO', 'BHAIRABI', 'DHARMANAGAR', 'UDAIPUR',
            'DULLABCHERRA', 'BADARPUR JN.', 'MALDA COURT', 'OLD MALDA',
            'SINGHABAD', 'BHALUKPONG', 'BUNIYADPUR', 'MENDIPATHAR',
            'TINSUKIA JN.', 'SIMALUGURI JN.', 'KATPADI JN', 'VELLORE CANTT',
            'CUDDAPAH', 'TRICHUR', 'METTUR DAM', 'KARUR JN.', 'UDHAGAMANDALA
        M',
            'COONOOR', 'CHAMRAJANAGAR', 'NANJANGUD TOWN', 'TUMKUR',
            'ARSIKERE JN.', 'CHIKMAGALURU', 'KOTTAYAM', 'EDAMANN',
            'KAYANKULAM JN', 'CHITRADURG', 'CHIKJAJUR JN.', 'KABAKA PUTTUR',
            'SUBRAHMANYA ROAD', 'CHERVATTUR', 'BYNDOOR', 'KASARAGOD',
            'DINDIGUL JN', 'BAGAVATHIPURAM', 'KUMBAKONAM', 'THIRUVARUR JN.',
            'HOSPET JN.', 'BIRUR JN.', 'KULEM', 'SIRPUR TOWN',
            'BHADRACHALAM ROAD', 'BOLARUM', 'RAICHUR', 'DORNAKAL JN.',
            'BITRAGUNTA', 'BHIMAVARAM JN.', 'MEDCHAL', 'MACHERLA',
            'NADIKUDE JN.', 'DHONE JN', 'MAHBUBNAGAR', 'MIRZAPALI', 'BODHA
        N',
            'RAYA DRUG', 'TANDUR', 'PARBHANI JN.', 'GADAG JN.', 'WADI JN.',
```

```
            'BHADRAKH', 'KHURDA ROAD JN.', 'JAJPUR KEONJHAR ROAD', 'CHAKULI
A',
            'GUA', 'ITWARI JN.', 'JHARSUGUDA JN.', 'BARSUAN', 'BIRMITRAPUR',
            'GARBETA', 'GEVRA-ROAD', 'CHANDIA ROAD', 'MANENDRAGARH',
            'LANJI GARH ROAD', 'CUTTACK JN.', 'BERHAMPUR', 'TALCHER', 'PALAS
A',
            'ANGUL', 'KENDUJHAR', 'NAYAGARHTOWN', 'LOHARDAGA  BS', 'TORI',
            'SHAHDOL', 'DHAMTARI', 'KENDRI', 'RAJIM', 'ABHANPUR JN.', 'WADS
A',
            'RAMTEK', 'TIRODI', 'TUMSAR RD', 'NAGBHIR JN.', 'VIRAR', 'VAPI',
            'VIRAMGAM JN', 'KATHANA', 'CHHOTA UDEPUR', 'PRATAP NAGAR',
            'MODASA', 'NADIAD JN.', 'ANAND JN.', 'BHARUCH JN.',
            'VADTAL SWAMINARAYAN', 'RAJPIPLA', 'ANKLESHWAR JN.',
            'SURENDRANAGAR', 'KANALUS', 'BHANVAD', 'BOTAD', 'DHRANGADHRA',
            'DHOLA JN', 'RAJULA CITY', 'NAGDA JN', 'YAMUNA BRIDGE', 'BAYAN
A',
            'MAKSI JN', 'PATAN', 'MAHESANA JN', 'PUSHKAR TERMINUS', 'ANUPGAR
H',
            'PHULERA JN.', 'FATEHPUR-SHEKHAWATI', 'NIMACH', 'BAIDYANATHDHA
M',
            'RAGHUNATHPUR', 'BUXAR', 'SONPUR JN.', 'DURGAPUR', 'ANDAL JN.',
            'SAINTHIA JN.', 'HAZARIBAGH ROAD', 'SONIPAT', 'SAHIBABAD',
            'SHAKUR BASTI', 'PALWAL', 'GHAZIABAD JN.', 'KOSI KALAN', 'DANKAU
R',
            'PHAPHUND', 'PANKI', 'BARA BANKI JN.', 'KALIANPUR', 'BULANDSHAH
R',
            'HATHRAS QILAH', 'HATHRAS JN.', 'BALLABGARH', 'NELLORE',
            'MELMARUVATHUR', 'HINDUPUR', 'KUPPAM', 'RAMANAGARAM',
            'RAJAHMUNDRY', 'TENALI JN.', 'ONGOLE', 'WARANGAL', 'BHONGIR',
            'JANGAON', 'VIZIANAGARAM JN.', 'SRIKAKULAM ROAD', 'RAVIKAMPADU',
            'TUNI', 'GHATSILA', 'JALESWAR', 'BELDA', 'BARABHUM', 'KHANODIH',
            'BHAGA JN', 'CHANDRAPURA JN.', 'HIJLI', 'DONGAR GARH',
            'PENDRA ROAD', 'GODHRA JN.', 'DAKOR', 'SANJAN', 'UMBERGAON ROA
D',
            'DAHANU ROAD', 'VASAI ROAD', 'NANDURBAR', 'BORIVLI', 'DAHEJ',
            'PERNEM', 'BOISAR', 'ROHA', 'IDGAH', 'VRINDAVAN', 'DOHRIGHAT',
            'INDARA JN.', 'NALHATI JN.', 'NIMTITA', 'JANGIPUR ROAD',
            'PATNA GHAT', 'DIGHA GHAT', 'R BLOCK HALT', 'TILRATH',
            'KHAGARIA JN.', 'DHANAURI', 'SULTANGANJ', 'BARAUT',
            'GARHI HARSARU JN.', 'TARN TARAN', 'BEAS JN', 'BUDGAM', 'BANIHA
```

```
L',
       'BARAMULLA', 'MANANWALA', 'DERABABA NANAK', 'VERKA JN.',
       'KHEM KARAN', 'BHAGTANWALA', 'QADIAN', 'MERTA CITY', 'BHILDI J
N',
       'SARDARSAHAR', 'NAKODAR JN.', 'KOT KAPURA JN', 'PHEPHNA JN.',
       'DINHATA', 'HAIBARGAON', 'TEZNARAYANPUR', 'BARSOI JN.', 'TELTA',
       'DANGARI', 'KOLAR', 'CHANNA PATNA', 'DHARMAPURI', 'LALGUDI',
       'MANAMADURAI', 'VIRUDUNAGAR JN', 'CUDDALORE PORT JN.',
       'VRIDHA CHALAM JN.', 'BHIMAVARAM TOWN', 'MARKAPUR ROAD',
       'MAHBUBABAD', 'PEDDAPALLI', 'JAGITYAL', 'KOTIPALLI', 'NANDYAL',
       'PENDLIMARRI', 'MANOHARABAD', 'JAKLAIR', 'UMDANAGAR', 'HUMNABA
D',
       'MIRYALAGUDA', 'PIDUGURALLA', 'GADWAL', 'RUPSA JN.', 'BARIPADA P
H',
       'BADAMPAHAR', 'GRAM MASAGRAM', 'SONAMUKHI PH', 'SALUR',
       'BOBBILI JN.', 'KATANGI', 'WARASEONI', 'GUDUM', 'DALLI-RAJHARA',
       'BHILWARA', 'KHAMBHAT', 'ABU ROAD', 'MORVI', 'WANKANER',
       'MALIYA MIYANA JN.', 'ADRAJ MOTI', 'VIJAPUR', 'AMBLIYASAN JN.',
       'CHURCHGATE', 'RAVLI JN', 'ANDHERI', 'PALASDARI', 'KALYAN JN',
       'BADLAPUR', 'AMBERNATH', 'KASARA', 'ASANGAON', 'TITVALA',
       'DOMBIVLI', 'THANE', 'KURLA', 'GHAT KOPAR', 'MUMBAI VADALA ROA
D',
       'VASHI', 'BELAPUR C.B.D', 'MANKHURD', 'CHEMBUR', 'NERUL',
       'LONAVLA', 'SHIVAJINAGAR', 'TALEGAON'], dtype=object)
```

Count the SEQ

In [40]:
```
df.SEQ.value_counts()
```

Out[40]:
```
2      7186
1      7186
3      6174
4      6072
5      5885
       ...
117       1
116       1
```

```
115      1
114      1
118      1
Name: SEQ, Length: 235, dtype: int64
```

To view the record of LUCKNOW JN.

In [41]:
```
d=df[df['Station Name']=='LUCKNOW JN.']
d
```

Out[41]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Stati Nar
401	BSB BHARATDA	7	LKO	LUCKNOW JN.	23:50:00	23:55:00	1295	AWB	AURANGABA
421	LKO-SVDK FTR	1	LKO	LUCKNOW JN.	23:00:00	23:00:00	0	LKO	LUCKNOW J
422	SVDK-LKO FTR	5	LKO	LUCKNOW JN.	14:30:00	14:30:00	1277	SVDK	SHRI MA VAISHNO DE KATR
5065	CPR-LJN	15	LJN	LUCKNOW JN.	08:15:00	08:15:00	503	CPR	CHHAPRA J
5066	LJN-CPR-EXP	1	LJN	LUCKNOW JN.	20:25:00	20:25:00	0	LJN	LUCKNOW J
...	...	...	...	...	...	...	...	...	
64275	BBK-LJN MEMU	10	LJN	LUCKNOW JN.	21:45:00	21:45:00	36	BBK	BARA BAN J
64281	SLN-LKO MEMU	18	LKO	LUCKNOW JN.	11:00:00	11:00:00	139	SLN	SULTANPU J
64282	LKO-SLN MEMU	1	LKO	LUCKNOW JN.	16:45:00	16:45:00	0	LKO	LUCKNOW J

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Stati Nar
74201	PBH-LKO DMU	25	LKO	LUCKNOW JN.	10:45:00	10:45:00	172	PBH	PARTAPGAR J
74202	LKO/PBH DMU	1	LKO	LUCKNOW JN.	16:30:00	16:30:00	0	LKO	LUCKNOW J

385 rows × 11 columns

To count number of Destination Station Name

In [42]:
```
d['Destination Station Name'].value_counts()
```

Out[42]:
```
LUCKNOW JN.              66
GORAKHPUR JN.            23
VARANASI JN.             18
ANAND VIHAR TERMINAL     15
NEW DELHI                14
                         ..
DANAPUR                   1
GHAZIPUR CITY             1
UDAIPUR CITY              1
NAHARLAGUN                1
TRIVANDRUM CENTRAL        1
Name: Destination Station Name, Length: 103, dtype: int64
```

To check data is aligned in a tabular fashion in rows and columns.

In [43]:
```
dt=pd.DataFrame(df)
```

```
dt
```

Out[43]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
107	SWV-MAO-VLNK	1	SWV	SAWANTWADI R	00:00:00	10:25:00	0	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	2	THVM	THIVIM	11:06:00	11:08:00	32	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	3	KRMI	KARMALI	11:28:00	11:30:00	49	SWV	SAWANTWADI ROAD
107	SWV-MAO-VLNK	4	MAO	MADGOAN JN.	12:10:00	00:00:00	78	SWV	SAWANTWADI ROAD
108	VLNK-MAO-SWV	1	MAO	MADGOAN JN.	00:00:00	20:30:00	0	MAO	MADGOAN JN.
...	...	...	...	...	...	...	...	...	...
99908	EMU	8	AKRD	AKURDI	23:30:00	23:31:00	19	PUNE	PUNE JN.
99908	EMU	9	DEHR	DEHU ROAD	23:35:00	23:36:00	24	PUNE	PUNE JN.
99908	EMU	10	BGWI	BEGDAEWAI	23:39:00	23:40:00	28	PUNE	PUNE JN.
99908	EMU	11	GRWD	GHORAWADI	23:41:00	23:42:00	31	PUNE	PUNE JN.
99908	EMU	12	TGN	TALEGAON	23:50:00	00:00:00	34	PUNE	PUNE JN.

186124 rows × 11 columns

To sort the value of Destination Station Name

TO TOP 15

In [44]:
```
dn=d.sort_values('Destination Station Name',ascending=False).head(15)
dn
```

Out[44]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Sour Station Nan
15015	GKP- YPR EXP	4	LJN	LUCKNOW JN.	11:30:00	11:50:00	277	GKP	GORAKHPU J
12540	LKO-YPR EXPR	1	LKO	LUCKNOW JN.	20:00:00	20:00:00	0	LKO	LUCKNO J
12591	GKP-SBC EXP	8	LJN	LUCKNOW JN.	11:30:00	11:50:00	277	GKP	GORAKHPU J
22684	LKO-YPR WEEK	1	LKO	LUCKNOW JN.	00:00:00	18:30:00	0	LKO	LUCKNO J
15023	GKP-YPR-EXP	5	LKO	LUCKNOW JN.	15:55:00	16:20:00	298	GKP	GORAKHPU J
14266	DDN BSB EXPR	32	LKO	LUCKNOW JN.	07:45:00	07:55:00	544	DDN	DEHRA DU
19167	SABARMATI EX	45	LKO	LUCKNOW JN.	00:50:00	01:05:00	1277	ADI	AHMEDABA
14866	MARUDHAR EXP	26	LKO	LUCKNOW JN.	02:50:00	03:25:00	870	JU	JODHPU J
14864	MARUDHAR EXP	27	LKO	LUCKNOW JN.	02:50:00	03:25:00	870	JU	JODHPU J
14854	MARUDHAR EXP	27	LKO	LUCKNOW JN.	02:50:00	03:25:00	870	JU	JODHPU J
12238	JAT-BSB EXP.	9	LKO	LUCKNOW JN.	07:10:00	07:20:00	977	JAT	JAMMU TAV

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Sour Station Nan
22408	ANVT-BSB GAR	4	LKO	LUCKNOW JN.	02:25:00	02:35:00	479	ANVT	ANAN VIHA TERMINA
14258	KASHI V EXPR	18	LKO	LUCKNOW JN.	21:35:00	21:50:00	492	NDLS	NEW DEL
14236	BERELLY VARA	30	LKO	LUCKNOW JN.	23:15:00	23:25:00	235	BE	BAREILLY J
14228	VARUNA EXP	1	LKO	LUCKNOW JN.	17:50:00	17:50:00	0	LKO	LUCKNO J

Extract data

Extract data of the Destination Station Name NEW DELHI to check how many train NEW DELHI

In [45]:
```
nd=dt.loc[(dt['Destination Station Name']=='NEW DELHI')]
nd
```

Out[45]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
9005	BCT NDLS BI	1	BCT	MUMBAI CENTR	16:00:00	16:00:00	0	BCT	MUMBAI CENTRAL
9005	BCT NDLS BI	2	BRC	VADODARA JN.	20:27:00	20:37:00	391	BCT	MUMBAI CENTRAL

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
9005	BCT NDLS BI	3	KOTA	KOTA	02:35:00	02:40:00	918	BCT	MUMBAI CENTRAL
9005	BCT NDLS BI	4	NDLS	NEW DELHI	07:55:00	07:55:00	1373	BCT	MUMBAI CENTRAL
12001	BPL - NDLS S	1	HBJ	HABIBGANJ	15:00:00	15:00:00	0	HBJ	HABIBGANJ
...	...	...	...	...	...	...	...	...	...
64912	ROK-NDLS MEM	13	SSB	SHAKUR BASTI	08:38:00	08:39:00	59	ROK	ROHTAK JN.
64912	ROK-NDLS MEM	14	DBSI	DAYA BASTI	08:47:00	08:48:00	64	ROK	ROHTAK JN.
64912	ROK-NDLS MEM	15	VVKP	VIVEKANANDPU	08:51:00	08:52:00	65	ROK	ROHTAK JN.
64912	ROK-NDLS MEM	16	DKZ	DELHI-KISHAN	08:56:00	08:57:00	66	ROK	ROHTAK JN.
64912	ROK-NDLS MEM	17	NDLS	NEW DELHI	09:15:00	09:15:00	71	ROK	ROHTAK JN.

1408 rows × 11 columns

Extract data

Extract data of the Source Station Name

In [46]:
```
nd=dt.loc[(dt['Source Station Name']=='MUMBAI CENTRAL')]
nd
```

Out[46]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Sour Stati Nar
9005	BCT NDLS BI	1	BCT	MUMBAI CENTR	16:00:00	16:00:00	0	BCT	MUMB CENTRA
9005	BCT NDLS BI	2	BRC	VADODARA JN.	20:27:00	20:37:00	391	BCT	MUMB CENTRA
9005	BCT NDLS BI	3	KOTA	KOTA	02:35:00	02:40:00	918	BCT	MUMB CENTRA
9005	BCT NDLS BI	4	NDLS	NEW DELHI	07:55:00	07:55:00	1373	BCT	MUMB CENTRA
12009	SHATABDI EXP	1	BCT	MUMBAI CENTR	06:25:00	06:25:00	0	BCT	MUMB CENTRA
...	...	...	...	...	...	...	...	...	
59441	AHMEDABAD PA	70	BJD	BAREJADI NAN	13:43:00	13:44:00	474	BCT	MUMB CENTRA
59441	AHMEDABAD PA	71	GER	GERATPUR	14:00:00	14:01:00	477	BCT	MUMB CENTRA
59441	AHMEDABAD PA	72	VTA	VATVA	14:06:00	14:08:00	482	BCT	MUMB CENTRA
59441	AHMEDABAD PA	73	MAN	MANINAGAR	14:13:00	14:15:00	487	BCT	MUMB CENTRA
59441	AHMEDABAD PA	74	ADI	AHMEDABAD	14:55:00	14:55:00	490	BCT	MUMB CENTRA

623 rows × 11 columns

Extract data

Train schedule for MUMBAI CENTRAL to NEW DELHI

In [47]:
```
nd=dt.loc[(dt['Source Station Name']=='MUMBAI CENTRAL') & (dt['Destination Station Name']=='NEW DELHI')]
nd
```

Out[47]:

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
9005	BCT NDLS BI	1	BCT	MUMBAI CENTR	16:00:00	16:00:00	0	BCT	MUMBAI CENTRAL
9005	BCT NDLS BI	2	BRC	VADODARA JN.	20:27:00	20:37:00	391	BCT	MUMBAI CENTRAL
9005	BCT NDLS BI	3	KOTA	KOTA	02:35:00	02:40:00	918	BCT	MUMBAI CENTRAL
9005	BCT NDLS BI	4	NDLS	NEW DELHI	07:55:00	07:55:00	1373	BCT	MUMBAI CENTRAL
12951	MUMBAI RAJDH	1	BCT	MUMBAI CENTR	17:00:00	17:00:00	0	BCT	MUMBAI CENTRAL
12951	MUMBAI RAJDH	2	BVI	BORIVLI	17:30:00	17:32:00	29	BCT	MUMBAI CENTRAL
12951	MUMBAI RAJDH	3	ST	SURAT	19:53:00	19:58:00	262	BCT	MUMBAI CENTRAL
12951	MUMBAI RAJDH	4	BRC	VADODARA JN.	21:18:00	21:28:00	391	BCT	MUMBAI CENTRAL
12951	MUMBAI RAJDH	5	RTM	RATLAM JN	00:37:00	00:40:00	652	BCT	MUMBAI CENTRAL

Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name
12951	MUMBAI RAJDH	6	NAD	NAGDA JN	01:18:00	01:20:00	697	BCT	MUMBAI CENTRAL
12951	MUMBAI RAJDH	7	KOTA	KOTA	03:20:00	03:25:00	918	BCT	MUMBAI CENTRAL
12951	MUMBAI RAJDH	8	NDLS	NEW DELHI	08:35:00	08:35:00	1373	BCT	MUMBAI CENTRAL
22209	NDLS DURONTO	1	BCT	MUMBAI CENTR	23:15:00	23:15:00	0	BCT	MUMBAI CENTRAL
22209	NDLS DURONTO	2	BRC	VADODARA JN.	03:52:00	04:05:00	391	BCT	MUMBAI CENTRAL
22209	NDLS DURONTO	3	RTM	RATLAM JN	07:35:00	07:40:00	652	BCT	MUMBAI CENTRAL
22209	NDLS DURONTO	4	KOTA	KOTA	10:25:00	10:30:00	918	BCT	MUMBAI CENTRAL
22209	NDLS DURONTO	5	NDLS	NEW DELHI	16:30:00	16:30:00	1373	BCT	MUMBAI CENTRAL

unique()

Use to find station name from Station Name row

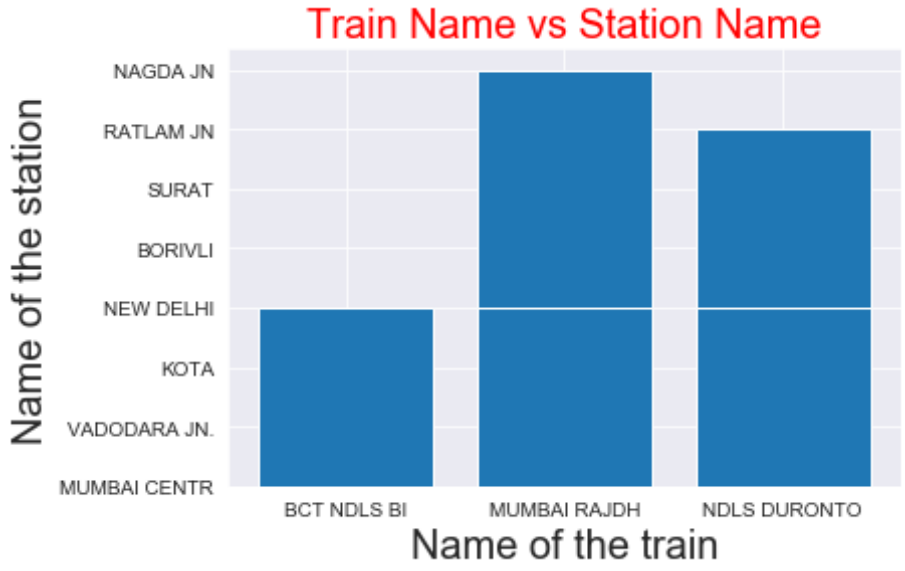
In [48]:
```
nd['Station Name'].unique()
```

Out[48]:
```
array(['MUMBAI CENTR', 'VADODARA JN.', 'KOTA', 'NEW DELHI', 'BORIVLI',
       'SURAT', 'RATLAM JN', 'NAGDA JN'], dtype=object)
```

Data Visualization

Plot a graph of Indian train dataset

In [49]:
```
plt.bar(nd['Train Name'],nd['Station Name'])
plt.title('Train Name vs Station Name',fontsize=20,color='red')
plt.xlabel('Name of the train',fontsize=20)
plt.ylabel('Name of the station',fontsize=20)
plt.grid(True)
plt.show()
```



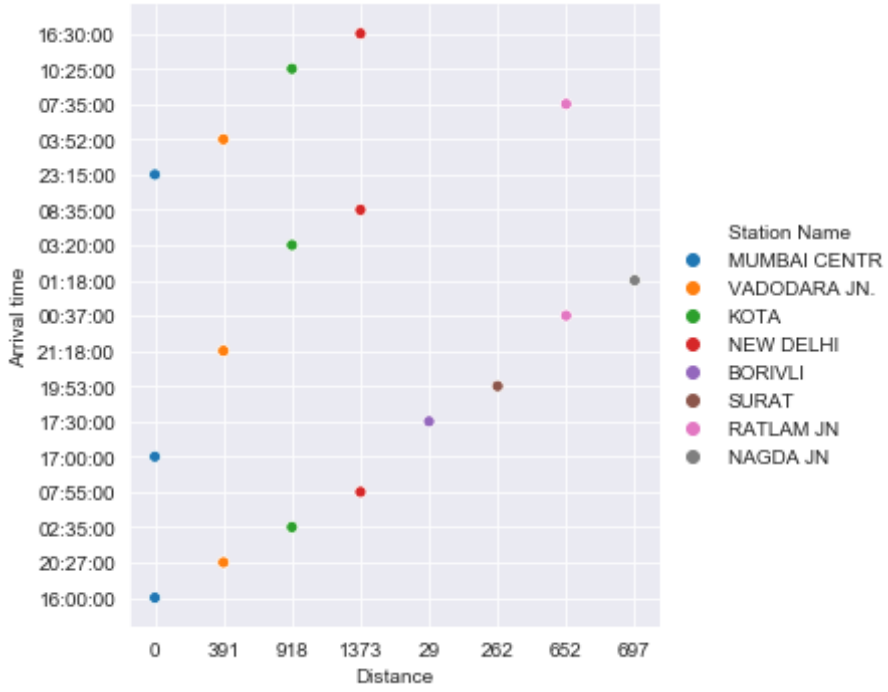
relplot()

This is a figure-level function for visualizing statistical relationships using two common approaches

In [50]:
```
plt.figure(figsize=(20,22))
sns.set_style('darkgrid')
```

```
sns.relplot(x='Distance',y='Arrival time',hue='Station Name',data=nd)
plt.show()
```

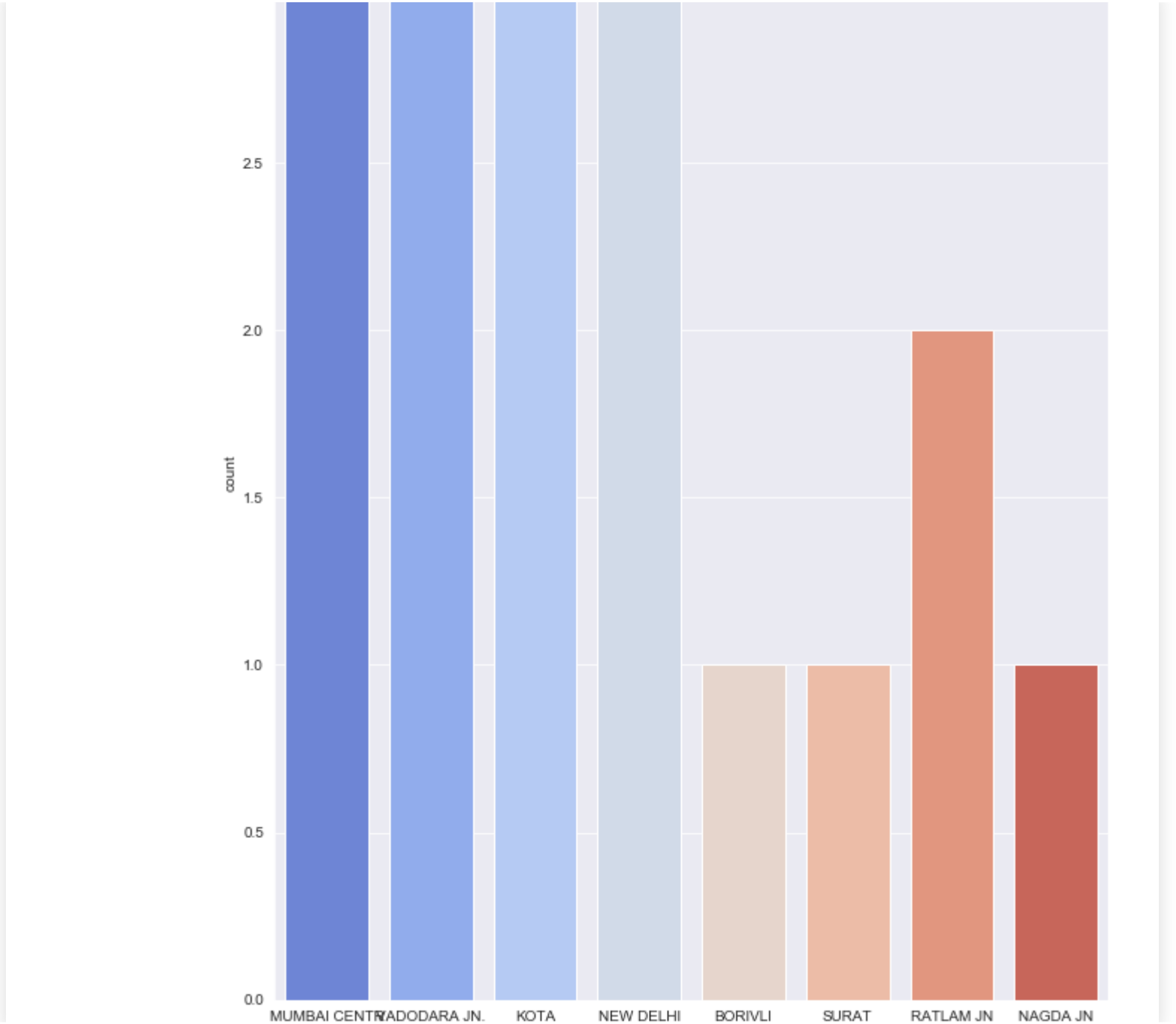
```
<Figure size 1440x1584 with 0 Axes>
```



Countplot()

A countplot is kind of likea histogram or a bar graph for some categorical area. It simply shows the number of occurrences of an item based on a certain type of category.

In [51]:
```
plt.figure(figsize=(10,13))
sns.countplot(x='Station Name',data=nd,palette='coolwarm')
plt.show()
```

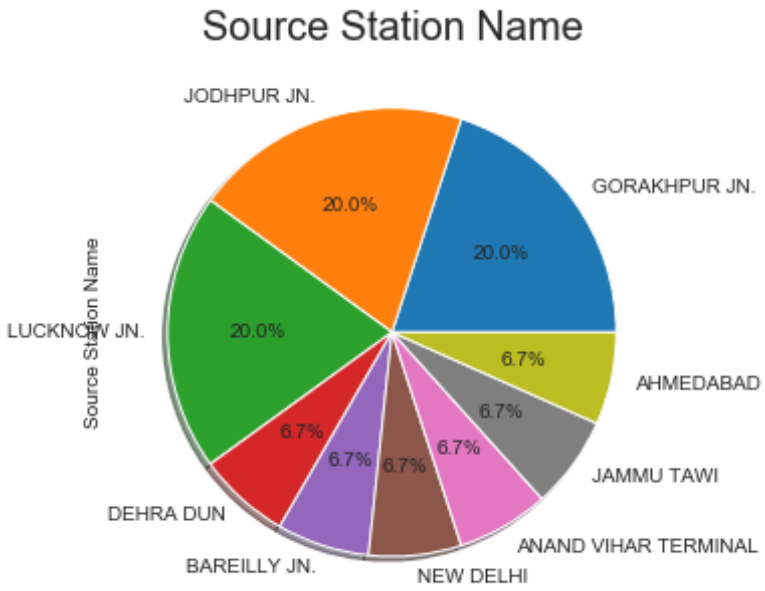


Station Name

Pie plot()

Pie charts show the size of items (called wedge) in one data series,

In [52]:
```
plt.figure(figsize=(5,9))
plt.title('Source Station Name',fontsize=20)
dn['Source Station Name'].value_counts().plot.pie(autopct='%1.1f%%',sha
dow=True)
plt.show()
```

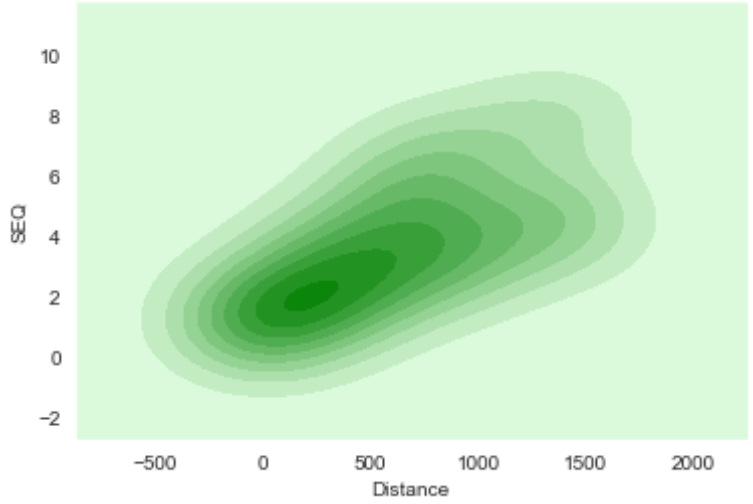


kdeplot()

A kernel density estimate (KDE) plot is a method for visualizing the distribution of observations in a dataset, analagous to a histogram. KDE represents the data using a continuous probability

density curve in one or more dimensions.

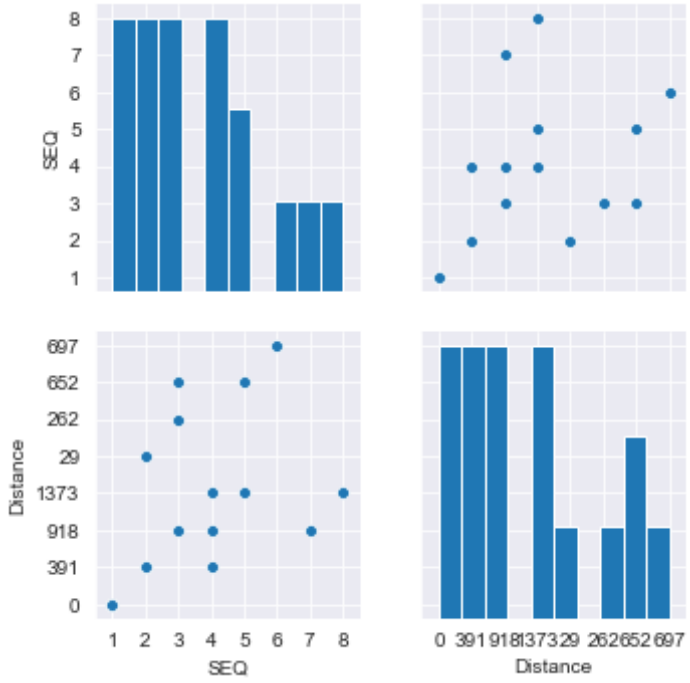
In [53]:
```
sns.kdeplot(nd['Distance'],nd['SEQ'],color='green',shade=True)
plt.show()
```



pairplot()

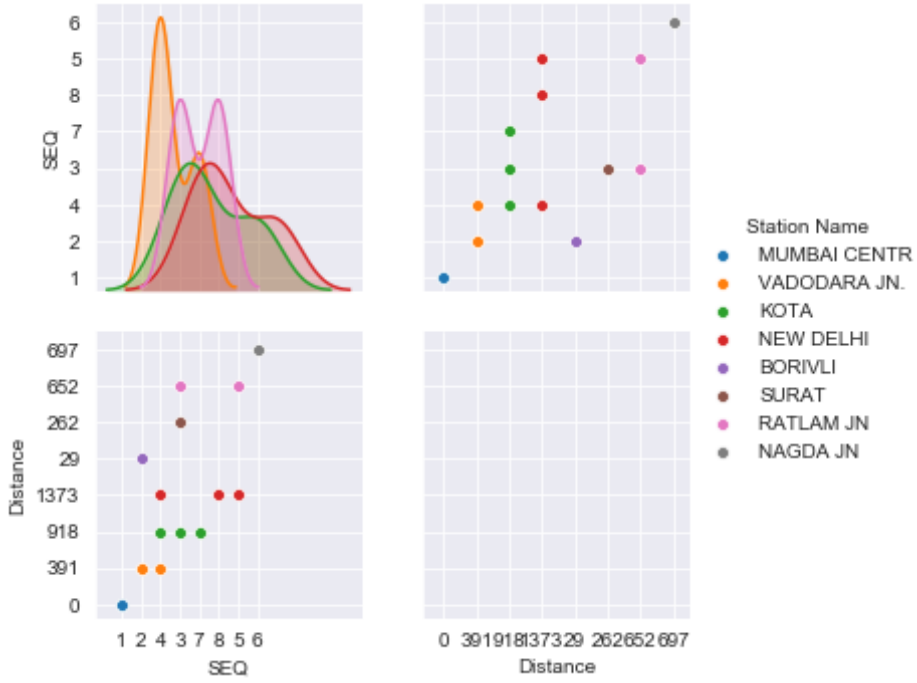
To plot multiple pairwise bivariate distributions in a dataset, you can use the pairplot() function. This shows the relationship for (n, 2) combination of variable in a DataFrame as a matrix of plots and the diagonal plots are the univariate plots.

In [54]:
```
sns.pairplot(nd)
plt.show()
```



Pairplot()

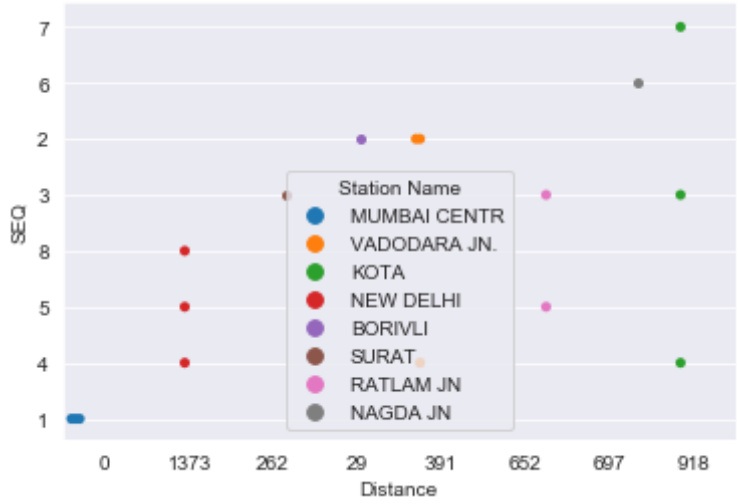
In [55]:
```
sns.pairplot(nd,diag_kind='kde',hue='Station Name')
plt.show()
```



swarm plot()

is very similar to a strip plot, yet the locations of points are adjusted automatically to avoid overlap even if the jitter value is not applied.

In [57]:
```
sns.swarmplot(x='Distance',y='SEQ',hue='Station Name',dodge=True,data=n
d)
plt.show()
```



In [ ]:

CONCLUSION

DATA SET: Indian Railway

Analysis:

My Analysis of Indain Railways with the Data Set. It Clearly shows that we can find train as per our need on the time and there are rows with Train Name and Station Name from where anyone can select Train from Stantion start there journey from Source Station Name to Destination Station Name which clrearly shows the name of the station.

In the Data set clearly mentioned Train No, Train Name, SEQ, Station Code, Arrival time, Departure Time, Distance, Source Station, Source Station Name, Destionation Station, Destination Station Name with all rows we can find Train. There are some part of visualization with that help we can plot data set in nice visual.