

In [1]:

```
import numpy as np
import pandas as pd

import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
#Loading the data from csv files.
df=pd.read_csv('Riders_List2020-22')
df.head()
```

Out[2]:

	LeadID	State	City	RiderName	LL EMP Code	RiderNumber	DOJ	Client
0	31	Karnataka	Bengaluru	SRIKANTH P	LLBBC001	6363376901	4/4/2019	Bigbasket Bangalore
1	32	Karnataka	Bengaluru	SUPREEM .	LLBBC003	7483505921	4/16/2019	Bigbasket Bangalore
2	33	Tamil Nadu	Chennai	Silambarasan V	LLBBC004	8695084040	3/1/2019	Bigbasket Chennai C
3	35	Karnataka	Bengaluru	VEERESH U	LLBBC011	9110444430	6/23/2019	Bigbasket Bangalore
4	37	Karnataka	Bengaluru	Harish M	LLBBC027	9008816586	6/17/2019	Bigbasket Bangalore

In [3]:

```
df.shape
```

Out[3]:

(7349, 15)

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7349 entries, 0 to 7348
Data columns (total 15 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   LeadID                7349 non-null   int64
 1   State                 7299 non-null   object
 2   City                  7299 non-null   object
 3   RiderName             7349 non-null   object
 4   LL EMP Code           7344 non-null   object
 5   RiderNumber           7349 non-null   object
 6   DOJ                   7257 non-null   object
 7   Client                7299 non-null   object
 8   Hub                   7299 non-null   object
 9   Previous_Status       7295 non-null   object
10   Current_Status        7294 non-null   object
11   Status_Changed        7294 non-null   object
12   Modified_by           7294 non-null   object
13   ReasonForLeaving      5737 non-null   float64
14   ReleivingDate_Entered 5769 non-null   object
dtypes: float64(1), int64(1), object(13)
memory usage: 861.3+ KB
```

In [5]:

```
df.isnull().sum()
```

Out[5]:

```
LeadID                0
State                 50
City                  50
RiderName             0
LL EMP Code           5
RiderNumber           0
DOJ                   92
Client                50
Hub                   50
Previous_Status       54
Current_Status        55
Status_Changed        55
Modified_by           55
ReasonForLeaving      1612
ReleivingDate_Entered 1580
dtype: int64
```

In [6]:

```
df1=df.copy()
```

In [7]:

df1

Out[7]:

	LeadID	State	City	RiderName	LL EMP Code	RiderNumber	DOJ	Clie
0	31	Karnataka	Bengaluru	SRIKANTH P	LLBBC001	6363376901	4/4/2019	Bigbask Bangalo
1	32	Karnataka	Bengaluru	SUPREEM .	LLBBC003	7483505921	4/16/2019	Bigbask Bangalo
2	33	Tamil Nadu	Chennai	Silambarasan V	LLBBC004	8695084040	3/1/2019	Bigbask Chenn
3	35	Karnataka	Bengaluru	VEERESH U	LLBBC011	9110444430	6/23/2019	Bigbask Bangalo
4	37	Karnataka	Bengaluru	Harish M	LLBBC027	9008816586	6/17/2019	Bigbask Bangalo
...	...	...	...	...	...	...	...	...
7344	27758	Karnataka	Bengaluru	Kranthi Kumara	13944	8861590554	6/14/2022	1ka Bangalo
7345	27761	Karnataka	Bengaluru	Darshan P	13948	7337715815	6/14/2022	1ka Bangalo
7346	27763	Telangana	Hyderabad	Madhu Kumar	13950	9180746571	6/14/2022	Bigbask Hyderabad
7347	27764	Karnataka	Bengaluru	Ravi S	13951	7975174381	6/15/2022	1ka Bangalo
7348	27782	Karnataka	Bengaluru	Aravinda reddy K	13960	7013137501	6/15/2022	1ka Bangalo

7349 rows × 15 columns

In [8]:

df1=df1.drop\_duplicates(subset=['RiderName','RiderNumber']).copy()

In [9]:

df1.shape

Out[9]:

(6685, 15)

In [10]:

```
df1
```

Out[10]:

	LeadID	State	City	RiderName	LL EMP Code	RiderNumber	DOJ	Clie
0	31	Karnataka	Bengaluru	SRIKANTH P	LLBBC001	6363376901	4/4/2019	Bigbask Bangalo
1	32	Karnataka	Bengaluru	SUPREEM .	LLBBC003	7483505921	4/16/2019	Bigbask Bangalo
2	33	Tamil Nadu	Chennai	Silambarasan V	LLBBC004	8695084040	3/1/2019	Bigbask Chenn
3	35	Karnataka	Bengaluru	VEERESH U	LLBBC011	9110444430	6/23/2019	Bigbask Bangalo
4	37	Karnataka	Bengaluru	Harish M	LLBBC027	9008816586	6/17/2019	Bigbask Bangalo
...	...	...	...	...	...	...	...	...
7344	27758	Karnataka	Bengaluru	Kranthi Kumara	13944	8861590554	6/14/2022	1ka Bangalo
7345	27761	Karnataka	Bengaluru	Darshan P	13948	7337715815	6/14/2022	1ka Bangalo
7346	27763	Telangana	Hyderabad	Madhu Kumar	13950	9180746571	6/14/2022	Bigbask Hyderabad
7347	27764	Karnataka	Bengaluru	Ravi S	13951	7975174381	6/15/2022	1ka Bangalo
7348	27782	Karnataka	Bengaluru	Aravinda reddy K	13960	7013137501	6/15/2022	1ka Bangalo

6685 rows × 15 columns

664 rows are remove due to duplicate values

In [11]:

```
# df1.sample(10)
```

In [12]:

```
df1.shape
```

Out[12]:

(6685, 15)

In [13]:

```
df1[df1['DOJ'] == '0000-00-00']
```

Out[13]:

	LeadID	State	City	RiderName	LL EMP Code	RiderNumber	DOJ	Client	
42	94	Tamil Nadu	Chennai	Karthick Subbaih	LLC0072	8220322678	0000-00-00	Bigbasket Chennai	Bigbasket Van
56	123	Telangana	Hyderabad	Nohan Kumar	LLH0001	8431291109	0000-00-00	Bigbasket Hyderabad	Bigbasket Ma
124	278	Tamil Nadu	Chennai	Vino Th .	LLC0162	9677059440	0000-00-00	Bigbasket Chennai	Bigbasket Chitla
125	284	Karnataka	Bengaluru	Manjunath Dhage	LLB0257	7795166002	0000-00-00	Bigbasket Bangalore	Bigbasket Se
127	287	Tamil Nadu	Chennai	Senthilkumar. V	LLC0379	9790914122	0000-00-00	Bigbasket Chennai	Bigbasket Van
...	...	...	...	...	...	...	...	...	...

In [14]:

```
df1[df1['DOJ'] == '0000-00-00'].shape
```

Out[14]:

(1093, 15)

In [15]:

```
df1['DOJ'] = pd.to_datetime(df1['DOJ'], errors='coerce')
```

In [16]:

```
df1['RelevingDate_Entered'] = pd.to_datetime(df1['RelevingDate_Entered'], errors='coerce')
```

In [17]:

```
df1['Status_Changed'] = pd.to_datetime(df1['Status_Changed'], errors='coerce')
```

In [18]:

```
df1['RelevingDate_Entered'] = df1.RelevingDate_Entered.fillna(df.Status_Changed, inplace=Fa
```

In [19]:

df1.head()

Out[19]:

	LeadID	State	City	RiderName	LL EMP Code	RiderNumber	DOJ	Client	
0	31	Karnataka	Bengaluru	SRIKANTH P	LLBBC001	6363376901	2019-04-04	Bigbasket Bangalore	Bi H
1	32	Karnataka	Bengaluru	SUPREEM .	LLBBC003	7483505921	2019-04-16	Bigbasket Bangalore	Bi H
2	33	Tamil Nadu	Chennai	Silambarasan V	LLBBC004	8695084040	2019-03-01	Bigbasket Chennai	Bi Chitla
3	35	Karnataka	Bengaluru	VEERESH U	LLBBC011	9110444430	2019-06-23	Bigbasket Bangalore	Bi H
4	37	Karnataka	Bengaluru	Harish M	LLBBC027	9008816586	2019-06-17	Bigbasket Bangalore	Bi H

In [20]:

df1.isnull().sum()

Out[20]:

```

LeadID          0
State           49
City            49
RiderName       0
LL EMP Code     5
RiderNumber     0
DOJ             1187
Client          49
Hub             49
Previous_Status 54
Current_Status  55
Status_Changed  55
Modified_by     55
ReasonForLeaving 1569
RelevingDate_Entered 14
dtype: int64

```

In [21]:

df1.shape

Out[21]:

(6685, 15)

In [22]:

```
pd.isnull(df1["RelevingDate_Entered"]).sum()
```

Out[22]:

14

In [23]:

```
df1=df1.dropna(subset=['RelevingDate_Entered'])
```

In [24]:

```
pd.isnull(df1["DOJ"]).sum()
```

Out[24]:

1187

In [25]:

```
df1=df1.dropna(subset=['DOJ'])
```

In [26]:

```
df1.shape
```

Out[26]:

(5484, 15)

In [27]:

```
df1=df1.drop([6995,1349,1428,4400,5748],axis=0)
```

In [28]:

```
df1.shape
```

Out[28]:

(5479, 15)

In [29]:

```
# df1.isnull().sum()
```

## working days

In [30]:

```
df1['working_days']=(df1['RelevingDate_Entered']-df1['DOJ']).dt.days
```

In [31]:

```
df1.head()
```

Out[31]:

	LeadID	State	City	RiderName	LL EMP Code	RiderNumber	DOJ	Client
0	31	Karnataka	Bengaluru	SRIKANTH P	LLBBC001	6363376901	2019-04-04	Bigbasket Bangalore
1	32	Karnataka	Bengaluru	SUPREEM .	LLBBC003	7483505921	2019-04-16	Bigbasket Bangalore
2	33	Tamil Nadu	Chennai	Silambarasan V	LLBBC004	8695084040	2019-03-01	Bigbasket Chennai
3	35	Karnataka	Bengaluru	VEERESH U	LLBBC011	9110444430	2019-06-23	Bigbasket Bangalore
5	38	Karnataka	Bengaluru	ANANDH .	LLBBC028	9095272783	2019-02-26	TOW Bangalore

In [32]:

```
df1[df1['working_days']<= 0]
```

Out[32]:

	LeadID	State	City	RiderName	LL EMP Code	RiderNumber	DOJ	Client
16	58	Karnataka	Bengaluru	Rudresh R	LLBBA066	8660622218	2021-10-20	Grofers Bangalore
53	115	Tamil Nadu	Chennai	Maria Victas .	LLC0115	9940274285	2021-07-22	Bigbasket Chennai
60	132	Telangana	Hyderabad	Nitin s	LLB0040	7483388888	2022-03-18	Bigbasket Hyderabad
69	147	Tamil Nadu	Chennai	VIGNESH D	LLC0221	9087902184	2021-01-18	Bigbasket Chennai
72	157	Tamil Nadu	Chennai	Dineshkumar R	LLC0294	7339697155	2022-04-03	Bigbasket Chennai
...	...	...	...	...	...	...	...	...
7343	27757	Karnataka	Bengaluru	Srinivas K	13945	9513162914	2022-06-14	1kart Bangalore
7344	27758	Karnataka	Bengaluru	Kranthi Kumara	13944	8861590554	2022-06-14	1kart Bangalore
7345	27761	Karnataka	Bengaluru	Darshan P	13948	7337715815	2022-06-14	1kart Bangalore
7346	27763	Telangana	Hyderabad	Madhu Kumar	13950	9180746571	2022-06-14	Bigbasket Hyderabad
7348	27782	Karnataka	Bengaluru	Aravinda reddy K	13960	7013137501	2022-06-15	1kart Bangalore

1027 rows × 16 columns



In [33]:

```
df1= df1[df1['working_days'] >0]
```

In [34]:

```
df1.shape
```

Out[34]:

```
(4452, 16)
```

Earlier data is 7349

In [35]:

```
sum(df1['working_days'])
```

Out[35]:

```
291656
```

In [36]:

```
max(df1['working_days'])
```

Out[36]:

```
1319
```

In [37]:

```
min(df1['working_days'])
```

Out[37]:

```
1
```

In [38]:

```
df1['working_days'].agg(['mean', 'median', lambda x :x.mode()[0]])
```

Out[38]:

```
mean      65.511231
median    27.000000
<lambda>   1.000000
Name: working_days, dtype: float64
```

In [39]:

```
data_outlier=df1.copy()
```

In [40]:

```
data_outlier.describe()
```

Out[40]:

	LeadID	ReasonForLeaving	working_days
count	4452.000000	3.532000e+03	4452.000000
mean	19078.661500	2.375055e+03	65.511231
std	6643.358099	1.411496e+05	106.498987
min	31.000000	0.000000e+00	1.000000
25%	15086.750000	0.000000e+00	7.000000
50%	20962.000000	0.000000e+00	27.000000
75%	24271.250000	0.000000e+00	76.000000
max	27764.000000	8.388607e+06	1319.000000

In [41]:

```
from scipy import stats
```

In [42]:

```
# q1, q3= np.percentile(data_outlier["working_days"],[25,75])  
# # q1, q3
```

In [43]:

```
# data_outlier[data_outlier["working_days"]<= q1].shape
```

In [44]:

```
# data_outlier[data_outlier["working_days"]>= q3].shape
```

In [45]:

```
# data_outlier[(data_outlier["working_days"]<q3) & (data_outlier["working_days"] >q1)]
```

In [46]:

```
# data_outlier=data_outlier[(data_outlier["working_days"]<q3) & (data_outlier["working_days"]
```

In [47]:

```
# data_outlier.shape
```

In [48]:

```
# data_outlier["working_days"].mean()
```

In [49]:

```
# data_outlier["working_days"].sum()
```

In [ ]:

In [50]:

```
Q1 = data_outlier["working_days"].quantile(.01)  
Q1
```

Out[50]:

1.0

In [51]:

```
data_outlier[data_outlier["working_days"]<= Q1].shape
```

Out[51]:

(338, 16)

In [52]:

```
Q3 = data_outlier["working_days"].quantile(.99)  
Q3
```

Out[52]:

529.48999999999998

In [53]:

```
data_outlier[data_outlier["working_days"] >= Q3].shape
```

Out[53]:

(45, 16)

In [54]:

```
data_outlier[(data_outlier["working_days"]<Q3) & (data_outlier["working_days"] >Q1)]
```

5	38	Karnataka	Bengaluru	ANANDH .	LLBBC028	9095272783	2019-02-26	TOW Bangalore	TOW HSRLayout
6	39	Karnataka	Bengaluru	Deepak Badiya	LLBBC029	9382654318	2022-05-10	BB_Now Bangalore	BBNow BLR-Bellandur
..	...	...	...	...	...	...	...	...	...
7	27671	Karnataka	Bengaluru	Nagendra Kumar	13886	8861611848	2022-06-11	1kart Bangalore	1kart BEL
0	27678	Telangana	Hyderabad	Mohammed Junaid Ahmed	13975	7799313122	2022-06-11	Amazon Hyderabad	Amazon Aramghar HYDK
2	27687	Telangana	Secunderabad	Vamsi Krishna	13909	8374079192	2022-06-14	Flipkart Hyderabad	Flipkart_ECIL
3	27688	Karnataka	Bengaluru	Pavan Kumar T	13897	7204052940	2022-03-16	1kart Bangalore	1kart Peenya
5	27691	Karnataka	Bengaluru	Ranganatha C	13901	7676822192	2022-06-13	1kart Bangalore	1kart Chandapura

In [55]:

```
df_outlier = data_outlier[(data_outlier["working_days"]< Q3) & (data_outlier["working_days"]
```

In [56]:

```
df_outlier.head()
```

Out[56]:

	LeadID	State	City	RiderName	LL EMP Code	RiderNumber	DOJ	Client	
0	31	Karnataka	Bengaluru	SRIKANTH P	LLBBC001	6363376901	2019-04-04	Bigbasket Bangalore	Bigba Hulir
1	32	Karnataka	Bengaluru	SUPREEM .	LLBBC003	7483505921	2019-04-16	Bigbasket Bangalore	Bigba Hulir
3	35	Karnataka	Bengaluru	VEERESH U	LLBBC011	9110444430	2019-06-23	Bigbasket Bangalore	Bigba Hulir
5	38	Karnataka	Bengaluru	ANANDH .	LLBBC028	9095272783	2019-02-26	TOW Bangalore	T HSRLa
6	39	Karnataka	Bengaluru	Deepak Badiya	LLBBC029	9382654318	2022-05-10	BB_Now Bangalore	BBI E Bellai

In [57]:

```
df_outlier.shape
```

Out[57]:

```
(4069, 16)
```

In [58]:

```
df_outlier.working_days.sum()
```

Out[58]:

260888

In [59]:

```
df_outlier.working_days.mean()
```

Out[59]:

64.11599901695749

In [60]:

```
df2=df_outlier.drop(['LeadID', 'State', 'LL EMP Code', 'Previous_Status', 'Current_Status',  
                    'RiderNumber', 'Modified_by', 'ReasonForLeaving'], axis=1).copy()
```

In [61]:

```
df2.shape
```

Out[61]:

(4069, 8)

In [62]:

```
df2.isnull().sum()
```

Out[62]:

City	9
RiderName	0
DOJ	0
Client	9
Hub	9
Status_Changed	17
RelevingDate_Entered	0
working_days	0
dtype: int64	

In [63]:

```
df2['City'].unique()
```

Out[63]:

```
array(['Bengaluru', nan, 'Chennai', 'Adyar', 'Hyderabad', 'Dharwad',  
      'Mysore', 'Kushalnagar', 'Nanjangud', 'Shimoga', 'Puttur',  
      'Hospet', 'Jamkhandi', 'Mangalore', 'Madurai', 'New Delhi',  
      'Delhi', 'Secunderabad', 'Hubli', 'Thrissur', 'Kayamkulam',  
      'Thiruvananthapuram', 'Kozhikode', 'Tiruchi', 'Pathanamthitta',  
      'Alappuzha', 'Noida', 'Salem', 'Erode', 'Coimbatore', 'Kottayam',  
      'Malappuram', 'Ernakulam', 'Dindigul', 'Aluva', 'Sivaganga',  
      'Mumbai', 'Malur', 'Gurgaon', 'Chengalpattu', 'Gulbarga', 'Kannur',  
      'Nilambur', 'Bhandup', 'Ghaziabad', 'Manjeri', 'Idukki',  
      'Tellicherry', 'Palakkad', 'Tirupur', 'Ahmedabad', 'Bijapur',  
      'Kolar', 'Kanhagad', 'Bidadi'], dtype=object)
```

In [64]:

```
df2['Hub'].unique()
```

Out[64]:

```
array(['Bigbasket Hulimavu', 'TOW HSRLayout', 'BBNow BLR-Bellandur',  
      'Bigbasket Siddapura', 'Bigbasket Sarjapura', nan,  
      'Bigbasket JP Nagar', 'Delhivery Wilson Garden',  
      'Delhivery TC Palya', 'Delhivery JP Nagar', 'Bigbasket Mylapore',  
      'Bigbasket Ittmadugu', '1kart_HulimavuHub_BLR',  
      'Bigbasket Jayanagar', 'TOW Whitefield', 'Bigbasket Vanagaram',  
      'Bigbasket HRBR', 'Bigbasket Thanisandra', '1kart Peenya',  
      '1kart BEL', '1Kart RajajiNagar', '1Kart Mysore Road',  
      '1Kart Ganga Nagar', '1Kart RT Nagar Bangalore',  
      'Bigbasket Mahadevpura', '1Kart Vijayanagar',  
      'SFX E-Com CHN_Velachery', '1Kart Srinagar', '1Kart Sunkuthkatte',  
      '1kart kanakapura', '1kart Bilekahalli', 'Bigbasket Thoraipakkam',  
      'SFX E-Com CHN_Pallikaranai', '1kart 8th mile',  
      'Grofers Jayanagar', 'BigBasket Kazhipattur', 'Naamdhari RT Nagar',  
      'Naamdhari IndraNagar', 'Naamdhari Koramangala', 'Flipkart Minjur',  
      '1kart_KRPuramHub_BLR', '1kart Bangalore Kammanahalli',  
      'Naamdhari Domlur', 'BigBasket Kudlu 5K', 'Grofers Velachery',  
      'Delhivery BTM', 'Mvnta HSR Layout', 'Bigbasket Chitlanakkam'])
```

In [65]:

```
df2['Client'].unique()
```

Out[65]:

```
array(['Bigbasket Bangalore', 'TOW Bangalore', 'BB_Now Bangalore', nan,
      'Delhivery Bangalore', 'Bigbasket Chennai', '1kart Bangalore',
      'Shadowfax E-Commerce ( Chennai )', 'Grofers Bangalore',
      'NAAMDHARI Bangalore', 'Flipkart Chennai', 'Grofers Chennai',
      'Myntra Bangalore', 'BlueDart Bangalore',
      'Big Basket Bangalore InActive',
      'PORTER BLR- SMARTSHIFT LOGISTICS SOLUTIONS P LTD',
      'OTP Express B2B', '1kart ROK', 'Sowkea Chennai',
      'Shadowfax Food Bangalore', 'Bigbasket Hyderabad',
      'Porter 2 W Chennai', 'Amazon Hyderabad', 'Porter 2W Bangalore',
      'Shadowfax Food - Chennai', 'Ninjacart-Chennai',
      'Ninjacart Bangalore', 'shadowfax Bangalore FLKRT QUICK',
      'Blue Dart-Chennai', 'Shadowfax E-Commerce Bangalore',
      'shadowfax Hyderabad FLKRT QUICK', 'Flipkart Delhi',
      'Flipkart Hyderabad', 'Porter_Dwarka Sec-14', 'Porter DELHI',
      'Porter 3 W Chennai', 'Flipkart Kerala', 'Grofers Delhi',
      'Flipkart Noida', 'Gourmet Garden', 'Village Milk',
      'Bigbasket Coimbatore', 'ShadowFax E-Comm Kerala'])
```

In [66]:

```
df2.groupby('City')['Client'].agg(['count']).head()
```

Out[66]:

	count
City	
Adyar	1
Ahmedabad	3
Alappuzha	4
Aluva	1
Bengaluru	2124

In [67]:

```
df_city=df2.groupby('City')['working_days'].agg(['count','mean','median',lambda x :x.mode()])
df_city.columns=['no_of_rider','average','median','mode']
```

In [68]:

```
df_city.head()
```

Out[68]:

	no_of_rider	average	median	mode
City				
Adyar	1	2.000000	2.0	2
Ahmedabad	3	11.000000	11.0	10
Alappuzha	4	97.500000	100.5	38
Aluva	1	64.000000	64.0	64
Bengaluru	2124	69.579567	34.0	2

In [69]:

```
df2['Riders_Joined_by_year'] = pd.DatetimeIndex(df2['DOJ']).year
df2['Riders_Quitted_by_year'] = pd.DatetimeIndex(df2['RelevingDate_Entered']).year
```

In [70]:

```
df2.head()
```

Out[70]:

	City	RiderName	DOJ	Client	Hub	Status_Changed	RelevingDate_Entered
0	Bengaluru	SRIKANTH P	2019-04-04	Bigbasket Bangalore	Bigbasket Hulimavu	2020-05-19	2020-04-02
1	Bengaluru	SUPREEM .	2019-04-16	Bigbasket Bangalore	Bigbasket Hulimavu	2020-05-19	2020-02-20
3	Bengaluru	VEERESH U	2019-06-23	Bigbasket Bangalore	Bigbasket Hulimavu	2020-11-10	2020-11-10
5	Bengaluru	ANANDH .	2019-02-26	TOW Bangalore	TOW HSRLayout	2020-07-20	2020-07-20
6	Bengaluru	Deepak Badiya	2022-05-10	BB_Now Bangalore	BBNow BLR-Bellandur	2022-05-14	2022-05-14

In [71]:

```
df2[df2['Riders_Joined_by_year']==2019].shape
```

Out[71]:

(40, 10)

In [72]:

```
df2[df2['Riders_Joined_by_year']==2020].shape
```

Out[72]:

(301, 10)



In [73]:

```
df2.groupby('City')['Riders_Quitted_by_year'].agg(['count']).head()
```

Out[73]:

	count
City	
Adyar	1
Ahmedabad	3
Alappuzha	4
Aluva	1
Bengaluru	2124

In [74]:

```
df2.groupby('City')['Riders_Joined_by_year'].agg(['count']).head()
```

Out[74]:

	count
City	
Adyar	1
Ahmedabad	3
Alappuzha	4
Aluva	1
Bengaluru	2124

In [75]:

```
df2['Riders_Joined_by_year'].count()
```

Out[75]:

4069

In [76]:

```
df2.groupby('Hub')['Riders_Joined_by_year'].agg(['count']).head()
```

Out[76]:

	count
Hub	
1Kart Bangalore Anekal	2
1Kart Bangalore Jakkasandra	1
1Kart Bangalore Sarjapur	3
1Kart Banshankari	20
1Kart E_City	28

In [77]:

```
df2.groupby('Client')['Riders_Joined_by_year'].agg(['count']).head()
```

Out[77]:

	count
Client	
1kart Bangalore	1006
1kart ROK	133
Amazon Fresh	3
Amazon Hyderabad	50
BB_Now Bangalore	57

In [78]:

```
df2['Client'].count()
```

Out[78]:

4060

In [79]:

```
df2.Hub.count()
```

Out[79]:

4060

In [80]:

```
bool_series = pd.isnull(df2["Hub"])
df2[bool_series]
```

Out[80]:

	City	RiderName	DOJ	Client	Hub	Status_Changed	RelevingDate_Entered	working_da
23	NaN	DINESH R	2019-06-24	NaN	NaN	2020-09-29	2020-09-29	4
574	NaN	Ashok Kumar	2020-09-19	NaN	NaN	2020-12-29	2020-12-29	1
596	NaN	Syed Khalid	2020-09-22	NaN	NaN	2021-01-16	2021-01-16	1
678	NaN	Syed Ilyas	2020-10-14	NaN	NaN	2021-01-19	2021-01-19	1
697	NaN	Mohan Raman	2020-10-09	NaN	NaN	2021-01-22	2021-01-22	1
714	NaN	Mohammed Nayeem	2022-01-10	NaN	NaN	2022-02-21	2022-02-21	1
855	NaN	Dharmendar G	2020-12-22	NaN	NaN	2021-01-22	2021-01-22	1
1973	NaN	Syed Yesdan	2021-07-12	NaN	NaN	2021-08-14	2021-08-14	1
2171	NaN	Leeladhar HP	2021-09-02	NaN	NaN	2021-09-15	2021-09-15	1

In [81]:

```
df2=df2.dropna(subset=['Client'])
```

In [82]:

```
df2.shape
```

Out[82]:

(4060, 10)

In [83]:

```
df2.sample(5)
```

Out[83]:

	City	RiderName	DOJ	Client	Hub	Status_Changed	RelevingDate_
4305	Bengaluru	Saifulla Athaulla Syed	2022- 01-13	Shadowfax E- Commerce Bangalore	Shadowfax E-Commerce Bangalore	2022-01-17	20
4210	Bengaluru	Prabhakar S	2022- 01-10	1kart Bangalore	1Kart Mysore Road	2022-02-05	20
4203	Secunderabad	SANGEM ANIL	2022- 01-22	Flipkart Hyderabad	Flipkart_ECIL	2022-02-10	20
3226	Chennai	Shyam Kumar	2021- 11-24	Bigbasket Chennai	Bigbasket Mylapore	2022-02-10	20
6811	Bengaluru	syed usman	2022- 05-21	1kart Bangalore	1Kart RT Nagar Bangalore	2022-05-25	20

In [84]:

```
df2.working_days.mean()
```

Out[84]:

64.0115763546798

In [85]:

```
df2['working_days'].max()
```

Out[85]:

529

In [ ]:

In [ ]: