S4Labour

August 18, 2021

1 S4labour Data Exercise

1.1 Marketing Data Analyst

1.1.1 Questions

Produce some descriptive statistics for the data.

What can you see that looks odd? What would you recommend doing about it?

Report the like for like sales by organisation and site by month for March 2019 and March 2021.

Which organisations have most sites?ű

In [1]:
```
# import libraries
import numpy as np
import pandas as pd
#import matplotlib.pyplot as plt
#%matplotlib inline
```

In [2]:
```
# Import csv to python

# March 2019
df2019 = pd.read_csv('March_2019_CSV.csv',index_col=0)
# March 2021
df2021 = pd.read_csv('March_2021_CSV.csv',index_col=0)
#df2019.head()
#df2021
#df1.info()
```

In [3]:
```
# Combine dataframes for 2019 and 2021 into dataframe labelled  dfAll

dfAll=pd.concat([df2019,df2021])
dfAll.head()
```

Out[3]:
```
                                    OrganisationID PostCode  ActualSalesId  \
tbl_PhysicalSites.PhysicalSiteID
```

```
1783                                          281  NG1 1NN       25136608
1783                                          281  NG1 1NN       25136609
1783                                          281  NG1 1NN       25136610
465                                            89  GU8 5HJ       25136638
465                                            89  GU8 5HJ       25136639

                                    tbl_ActualSales.PhysicalSiteID   WorkDate  \
tbl_PhysicalSites.PhysicalSiteID
1783                                                          1783  01-Mar-19
1783                                                          1783  01-Mar-19
1783                                                          1783  01-Mar-19
465                                                            465  01-Mar-19
465                                                            465  01-Mar-19

                                           Type  ActualSalesValue
tbl_PhysicalSites.PhysicalSiteID
1783                                       Food            651.75
1783                                      Drink           1432.40
1783                              Accommodation              0.00
465                                        Food           2747.20
465                                       Drink           1784.24
```

```
In [4]: round(dfAll['ActualSalesValue'].mean(),0)
```

```
Out[4]: 734.0
```

```
In [5]: round(dfAll['ActualSalesValue'].max(),0)
```

```
Out[5]: 93309.0
```

```
In [6]: sum(dfAll[dfAll['OrganisationID']==281]['ActualSalesValue'])
```

```
Out[6]: 401222.6199999999
```

```
In [7]: # Top 5 Sales
        sales_desc=dfAll.sort_values('ActualSalesValue',axis=0, ascending=False).head(5)
        sales_desc
```

```
Out[7]:                                 OrganisationID  PostCode  ActualSalesId  \
        tbl_PhysicalSites.PhysicalSiteID
        1467                                       228  CF10 1BS       25481700
        1477                                       228  GL50 3PA       25441227
        1448                                       228    M2 5QR       25602514
        1448                                       228    M2 5QR       25167677
        1448                                       228    M2 5QR       25740799

                                    tbl_ActualSales.PhysicalSiteID   WorkDate  \
        tbl_PhysicalSites.PhysicalSiteID
        1467                                                  1467  16-Mar-19
```

```
1477                                                    1477  15-Mar-19
1448                                                    1448  23-Mar-19
1448                                                    1448  02-Mar-19
1448                                                    1448  30-Mar-19

                                     Type  ActualSalesValue
tbl_PhysicalSites.PhysicalSiteID
1467                                 Drink          93309.03
1477                                 Drink          63024.79
1448                                 Drink          59454.04
1448                                 Drink          57604.22
1448                                 Drink          56861.30
```

2 Analysis by OrganisationID

```
In [8]: print('Organisation responsible for highest sale: ')
        toporg=dfAll[dfAll['ActualSalesValue']==dfAll['ActualSalesValue'].max()]
        toporg
```

```
Organisation responsible for highest sale:
```

```
Out[8]:                              OrganisationID  PostCode  ActualSalesId  \
        tbl_PhysicalSites.PhysicalSiteID
        1467                                    228  CF10 1BS       25481700

                                     tbl_ActualSales.PhysicalSiteID   WorkDate  \
        tbl_PhysicalSites.PhysicalSiteID
        1467                                                 1467  16-Mar-19

                                     Type  ActualSalesValue
        tbl_PhysicalSites.PhysicalSiteID
        1467                         Drink          93309.03
```

```
In [9]: print('Organisation responsible for lowest sale: ')
        bottomorg=dfAll[dfAll['ActualSalesValue']==dfAll['ActualSalesValue'].min()]
        bottomorg
```

```
Organisation responsible for lowest sale:
```

```
Out[9]:                              OrganisationID  PostCode  ActualSalesId  \
        tbl_PhysicalSites.PhysicalSiteID
        1547                                    242  SN13 0HB       36783172

                                     tbl_ActualSales.PhysicalSiteID   WorkDate  \
        tbl_PhysicalSites.PhysicalSiteID
        1547                                                 1547  12-Mar-21
```

```
                                                    Type  ActualSalesValue
tbl_PhysicalSites.PhysicalSiteID
1547                                Accommodation          -3004.46
```

```
In [10]: print('Maximum SalesValue for each OrganisationID')
         m=dfAll.groupby('OrganisationID')['ActualSalesValue','tbl_ActualSales.PhysicalSiteID'
         m.head()
```

```
Maximum SalesValue for each OrganisationID
```

Out[10]:

OrganisationID	ActualSalesValue	tbl_ActualSales.PhysicalSiteID	Type
2	2500.00	11	Food
4	7942.88	1093	Food
17	5329.81	547	Food
20	28386.39	1887	Food
57	19415.65	754	Food

```
In [11]: round(m.describe(),0)
```

Out[11]:

	ActualSalesValue	tbl_ActualSales.PhysicalSiteID
count	137.0	137.0
mean	6558.0	1889.0
std	9726.0	603.0
min	0.0	11.0
25%	1446.0	1525.0
50%	4514.0	1969.0
75%	8106.0	2415.0
max	93309.0	2662.0

3 Analysis by Physical site ID

```
In [12]: print('Site responsible for highest sale: ')
         topsite=dfAll[dfAll['ActualSalesValue']==dfAll['ActualSalesValue'].max()]
         topsite
```

```
Site responsible for highest sale:
```

Out[12]:

tbl_PhysicalSites.PhysicalSiteID	OrganisationID	PostCode	ActualSalesId	\
1467	228	CF10 1BS	25481700	

tbl_PhysicalSites.PhysicalSiteID	tbl_ActualSales.PhysicalSiteID	WorkDate	\
1467	1467	16-Mar-19	

```
                                        Type  ActualSalesValue
tbl_PhysicalSites.PhysicalSiteID
1467                                   Drink          93309.03
```

```
In [13]: print('Site responsible for lowest sale: ')
         bottomsite=dfAll[dfAll['ActualSalesValue']==dfAll['ActualSalesValue'].min()]
         bottomsite
```

```
Site responsible for lowest sale:
```

```
Out[13]:                                   OrganisationID  PostCode  ActualSalesId  \
         tbl_PhysicalSites.PhysicalSiteID
         1547                                         242  SN13 0HB       36783172

                                           tbl_ActualSales.PhysicalSiteID   WorkDate  \
         tbl_PhysicalSites.PhysicalSiteID
         1547                                                       1547  12-Mar-21

                                                    Type  ActualSalesValue
         tbl_PhysicalSites.PhysicalSiteID
         1547                                Accommodation          -3004.46
```

```
In [14]: print('Maximum SalesValue for each PhysicalSiteID')
         n=dfAll.groupby('tbl_ActualSales.PhysicalSiteID')['ActualSalesValue','Type','Organisat
         n.max().head()
```

```
Maximum SalesValue for each PhysicalSiteID
```

```
Out[14]:                                  ActualSalesValue  Type  OrganisationID
         tbl_ActualSales.PhysicalSiteID
         11                                        2500.00  Food               2
         12                                        4118.58  Food               4
         36                                        4743.63  Food              20
         37                                        3003.21  Food              20
         38                                        2466.32  Food              20
```

```
In [15]: top_sales=dfAll.sort_values('ActualSalesValue',axis=0, ascending=False).head()
```

```
In [16]: round(n.describe(),0).head()
```

```
Out[16]:                                 ActualSalesValue                          \
                                                   count    mean     std  min   25%
         tbl_ActualSales.PhysicalSiteID
         11                                         30.0   917.0   839.0  0.0   0.0
         12                                        117.0   702.0   770.0  0.0  83.0
         36                                         93.0  1178.0  1236.0  0.0   0.0
```

```
37                                    93.0   643.0   679.0  0.0   0.0
38                                    93.0   736.0   693.0  0.0   0.0

                                                        OrganisationID       \
                                      50%     75%     max          count  mean
tbl_ActualSales.PhysicalSiteID
11                                  825.0  1688.0  2500.0           30.0   2.0
12                                  546.0  1053.0  4119.0          117.0   4.0
36                                  942.0  1646.0  4744.0           93.0  20.0
37                                  520.0   894.0  3003.0           93.0  20.0
38                                  683.0  1143.0  2466.0           93.0  20.0


                                     std   min   25%   50%   75%   max
tbl_ActualSales.PhysicalSiteID
11                                   0.0   2.0   2.0   2.0   2.0   2.0
12                                   0.0   4.0   4.0   4.0   4.0   4.0
36                                   0.0  20.0  20.0  20.0  20.0  20.0
37                                   0.0  20.0  20.0  20.0  20.0  20.0
38                                   0.0  20.0  20.0  20.0  20.0  20.0
```

In [17]:
```
sales_desc
```

Out[17]:
```
                                    OrganisationID  PostCode  ActualSalesId  \
tbl_PhysicalSites.PhysicalSiteID
1467                                           228  CF10 1BS       25481700
1477                                           228  GL50 3PA       25441227
1448                                           228    M2 5QR       25602514
1448                                           228    M2 5QR       25167677
1448                                           228    M2 5QR       25740799

                                    tbl_ActualSales.PhysicalSiteID   WorkDate  \
tbl_PhysicalSites.PhysicalSiteID
1467                                                          1467  16-Mar-19
1477                                                          1477  15-Mar-19
1448                                                          1448  23-Mar-19
1448                                                          1448  02-Mar-19
1448                                                          1448  30-Mar-19

                                     Type  ActualSalesValue
tbl_PhysicalSites.PhysicalSiteID
1467                                Drink          93309.03
1477                                Drink          63024.79
1448                                Drink          59454.04
1448                                Drink          57604.22
1448                                Drink          56861.30
```

In [18]:
```
dfAll.head()
```

Out[18]:

tbl_PhysicalSites.PhysicalSiteID	OrganisationID	PostCode	ActualSalesId	tbl_ActualSales.PhysicalSiteID	WorkDate	Type	ActualSalesValue
1783	281	NG1 1NN	25136608	1783	01-Mar-19	Food	651.75
1783	281	NG1 1NN	25136609	1783	01-Mar-19	Drink	1432.40
1783	281	NG1 1NN	25136610	1783	01-Mar-19	Accommodation	0.00
465	89	GU8 5HJ	25136638	465	01-Mar-19	Food	2747.20
465	89	GU8 5HJ	25136639	465	01-Mar-19	Drink	1784.24

In [ ]:

In [ ]:

In [19]:
```
id_phys=dfAll.groupby('OrganisationID')['tbl_ActualSales.PhysicalSiteID']
unique_ID=id_phys.nunique()
unique_ID.head()
```

Out[19]:
```
OrganisationID
2      1
4      3
17     1
20    79
57     2
Name: tbl_ActualSales.PhysicalSiteID, dtype: int64
```

In [ ]:

3.1 Which organisations have most sites?

In [20]:
```
id_phys.nunique().max()
print('The Organisation_Id with the most physical sites is "{}2"'.format(id_phys.nunic
```

```
The Organisation_Id with the most physical sites is "882"
```

```
In [21]: q=(dfAll[dfAll['OrganisationID']==89]['tbl_ActualSales.PhysicalSiteID']).unique()
         print('The list of PhysicalSiteID for all sites under the Organisation_ID {} is:\n \n
```

```
The list of PhysicalSiteID for all sites under the Organisation_ID 88 is:

 [ 465  492  493  494  495  496  497  498  499  500  501  502  503  504
  505  506  507  508  509  510  511  512  513  514  515  516  517  518
  521  522  523  524  525  527  528  529  531  532  533  534  535  536
  537  539  550  572  610  612  649  828  829  923  975 1058 1080 1159
 1200 1414 1574 1714 1715 1716 1718 1759 1774 1808 1826  491 1642 1854
 1938 1939 1940 1941 1942 1943 1944 1946 1947  438 1945  530 1937 2336
 2001 2489 2064 2095]
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```