In [1]:

```
# Importing necessary Libraries <br>
# **1. Understanding and Analyzing the Dataset**
```

In [2]:

```
import pandas as pd
import io
print("Import success")
```

```
Import success
```

In [3]:

```
df = pd.read_csv("https://raw.githubusercontent.com/Pooja123667/Smart_Tendering_ML/main/file_finale%20(2).csv", low_memory=False)
# https://raw.githubusercontent.com/Pooja123667/Smart_Tendering_ML/main/FY19_BID_Trends_Report_Data%20(3).csv
# https://raw.githubusercontent.com/Pooja123667/Smart_Tendering_ML/main/file_name%20(2).csv
# https://raw.githubusercontent.com/Pooja123667/Smart_Tendering_ML/main/updates.csv
testing = pd.read_csv("https://raw.githubusercontent.com/Pooja123667/Smart_Tendering_ML/main/finalTesting.csv", low_memory=False)
```

```
In [4]: df.head(5)
```

```
Out[4]:
```

	index	company name	Floor Size	Full-time staff	Sanitation staff employed	Safety Inspector	Part-time staff	Current clients	Bid provides supplemental sanitation services	Types c dutie assigne t sanitatio worker
0	1	General Consulting Research	5220.0	1.0	31.0	9.0	9.0	145.0	Yes	Stree Sweepin an Baggin Snow an Ic Remo.
1	2	Analysis Analysis	1740.0	1.0	77.0	36.0	14.0	168.0	Yes	Stree Sweepin an Baggin Powe Washin Sn.
2	3	Federated Consulting Analysis	8150.0	1.0	10.0	6.0	6.0	42.0	No	Na
3	4	Atlantic Max North	41110.0	62.0	18.0	8.0	10.0	NaN	Yes	Stree Sweepin an Baggin Powe Washin Sn.
4	5	Star Consulting	3460.0	1.0	23.0	2.0	19.0	181.0	Yes	Stree Sweepin an Baggin Powe Washin Sn.

5 rows × 47 columns

In [5]:
```
df['Social media followers']
```

Out[5]:
```
0        11716.0
1          875.0
2        11231.0
3        75052.0
4         9005.0
          ...
99995    81904.0
99996     7632.0
99997     7215.0
99998    11747.0
99999     7628.0
Name: Social media followers, Length: 100000, dtype: float64
```

In [6]:
```
df.shape
```

Out[6]:
```
(100000, 47)
```

In [7]:
```
df.describe()
```

Out[7]:

	index	Floor Size	Full-time staff	Sanitation staff employed	Safety Inspector	Part-time staff	
count	100000.000000	97329.000000	97329.000000	97329.000000	97329.000000	97329.000000	8
mean	50000.500000	19928.001932	5.546291	48.539870	27.753352	12.480586	
std	28867.657797	19202.251227	11.769358	27.478816	21.614128	8.990987	
min	1.000000	1740.000000	0.000000	-14.000000	-18.000000	-9.000000	
25%	25000.750000	8070.000000	1.000000	28.000000	11.000000	6.000000	
50%	50000.500000	13140.000000	2.000000	47.000000	26.000000	12.000000	
75%	75000.250000	25460.000000	3.000000	68.000000	43.000000	19.000000	
max	100000.000000	121820.000000	62.000000	115.000000	80.000000	35.000000	

8 rows × 36 columns

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 47 columns):
 #   Column                                                             Non-N
ull Count   Dtype  
---  ------                                                             -----
---------   -----  
 0   index                                                              10000
0 non-null  int64  
 1   company name                                                       10000
0 non-null  object 
 2   Floor Size                                                         97329
non-null   float64
 3   Full-time staff                                                    97329
non-null   float64
 4   Sanitation staff employed                                          97329
non-null   float64
 5   Safety Inspector                                                   97329
non-null   float64
 6   Part-time staff                                                    97329
non-null   float64
 7   Current clients                                                    83270
non-null   float64
 8   Bid provides supplemental sanitation services                      97329
non-null   object 
 9   Types of duties assigned to sanitation workers                     94585
non-null   object 
 10  Days per week of sanitation services                               10000
0 non-null  int64  
 11  Hours logged by sanitation workers                                 97329
non-null   float64
 12  Incidents of graffiti removed                                      10000
0 non-null  int64  
 13  Trash bags collected                                               97329
non-null   float64
 14  Trash and recycling receptacles serviced                           97329
non-null   float64
 15  Bid provides supplemental public safety services                   97329
non-null   object 
 16  Duties assigned to public safety personnel                         44990
non-null   object 
 17  Hours logged by public safety officers                             67622
non-null   float64
 18  Interactions with public safety officers                           66200
non-null   float64
 19  Bid provides supplemental streetscape and beautification services  97329
non-null   object 
 20  Planters and hanging baskets maintained                            97329
non-null   float64
 21  Tree pits maintained                                               97329
non-null   float64
 22  Banners maintained                                                 97329
non-null   float64
 23  Public art installations sponsored                                 97329
non-null   float64
 24  Street furniture elements maintained                               10000
0 non-null  int64  
```

```
 25  Wayfinding elements maintained                                     84631
non-null   object
 26  Lighting elements maintained                                       83226
non-null   object
 27  Other infrastructure elements maintained                           86113
non-null   object
 28  Public spaces maintained                                           84631
non-null   float64
 29  Bid has holiday lighting program                                   97329
non-null   object
 30  Communication channels used                                        85965
non-null   object
 31  Social media followers                                             97329
non-null   float64
 32  Marketing materials distributed                                    97329
non-null   float64
 33  Public events coordinated                                          97329
non-null   float64
 34  Estimated attendees to public events coordinated                   97329
non-null   float64
 35  Special event charges                                              99990
non-null   float64
 36  Miscellaneous charges                                              99990
non-null   float64
 37  Sanitation expenses                                                99990
non-null   float64
 38  Public safety expenses                                             99990
non-null   float64
 39  Marketing, holiday lighting, and special event expenses            0 non
-null       float64
 40  Streetscape & beautification expenses                              99990
non-null   float64
 41  Salaries                                                           99990
non-null   float64
 42  Outside contractor expenses                                        99990
non-null   float64
 43  Insurance costs                                                    99990
non-null   float64
 44  Rent and utilities                                                 0 non
-null       float64
 45  Supplies and equipment costs                                       99990
non-null   float64
 46  Other G&A expenses                                                 0 non
-null       float64
dtypes: float64(32), int64(4), object(11)
memory usage: 35.9+ MB
```

In [9]:
```
df['Duties assigned to public safety personnel'][3000]
```

Out[9]:
```
'Crime prevention workshops; Coordination with NYPD'
```

In [10]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 47 columns):
 #   Column                                                             Non-N
ull Count   Dtype  
---  ------                                                             -----
---------   -----  
 0   index                                                              10000
0 non-null  int64  
 1   company name                                                       10000
0 non-null  object 
 2   Floor Size                                                         97329
non-null   float64
 3   Full-time staff                                                    97329
non-null   float64
 4   Sanitation staff employed                                          97329
non-null   float64
 5   Safety Inspector                                                   97329
non-null   float64
 6   Part-time staff                                                    97329
non-null   float64
 7   Current clients                                                    83270
non-null   float64
 8   Bid provides supplemental sanitation services                      97329
non-null   object 
 9   Types of duties assigned to sanitation workers                     94585
non-null   object 
 10  Days per week of sanitation services                               10000
0 non-null  int64  
 11  Hours logged by sanitation workers                                 97329
non-null   float64
 12  Incidents of graffiti removed                                      10000
0 non-null  int64  
 13  Trash bags collected                                               97329
non-null   float64
 14  Trash and recycling receptacles serviced                           97329
non-null   float64
 15  Bid provides supplemental public safety services                   97329
non-null   object 
 16  Duties assigned to public safety personnel                         44990
non-null   object 
 17  Hours logged by public safety officers                             67622
non-null   float64
 18  Interactions with public safety officers                           66200
non-null   float64
 19  Bid provides supplemental streetscape and beautification services  97329
non-null   object 
 20  Planters and hanging baskets maintained                            97329
non-null   float64
 21  Tree pits maintained                                               97329
non-null   float64
 22  Banners maintained                                                 97329
non-null   float64
 23  Public art installations sponsored                                 97329
non-null   float64
 24  Street furniture elements maintained                               10000
0 non-null  int64  
```

```
 25  Wayfinding elements maintained                                  84631
non-null   object 
 26  Lighting elements maintained                                    83226
non-null   object 
 27  Other infrastructure elements maintained                        86113
non-null   object 
 28  Public spaces maintained                                        84631
non-null   float64
 29  Bid has holiday lighting program                                97329
non-null   object 
 30  Communication channels used                                     85965
non-null   object 
 31  Social media followers                                          97329
non-null   float64
 32  Marketing materials distributed                                 97329
non-null   float64
 33  Public events coordinated                                       97329
non-null   float64
 34  Estimated attendees to public events coordinated                97329
non-null   float64
 35  Special event charges                                           99990
non-null   float64
 36  Miscellaneous charges                                           99990
non-null   float64
 37  Sanitation expenses                                             99990
non-null   float64
 38  Public safety expenses                                          99990
non-null   float64
 39  Marketing, holiday lighting, and special event expenses         0 non
-null       float64
 40  Streetscape & beautification expenses                           99990
non-null   float64
 41  Salaries                                                        99990
non-null   float64
 42  Outside contractor expenses                                     99990
non-null   float64
 43  Insurance costs                                                 99990
non-null   float64
 44  Rent and utilities                                              0 non
-null       float64
 45  Supplies and equipment costs                                    99990
non-null   float64
 46  Other G&A expenses                                              0 non
-null       float64
dtypes: float64(32), int64(4), object(11)
memory usage: 35.9+ MB
```

In [11]:
```
#Dropping unnecessary columns
df.drop(['Marketing, holiday lighting, and special event expenses','Rent and u
tilities','Other G&A expenses'], axis=1, inplace=True)
```

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

Out[12]:
```
(100000, 44)
```

Now we shall replace all "$" signs in the columns to a null value

In [13]:

```
#Dropping unnecessary columns
df.head(1)
```

Out[13]:

	index	company name	Floor Size	Full-time staff	Sanitation staff employed	Safety Inspector	Part-time staff	Current clients	Bid provides supplemental sanitation services	Types of duties assigned to sanitation workers
0	1	General Consulting Research	5220.0	1.0	31.0	9.0	9.0	145.0	Yes	Street Sweeping and Bagging; Snow and Ice Remo...

1 rows × 44 columns

In [14]:

```
df.shape
```

Out[14]:

```
(100000, 44)
```

In [15]:

```
df.head(1)
```

Out[15]:

	index	company name	Floor Size	Full-time staff	Sanitation staff employed	Safety Inspector	Part-time staff	Current clients	Bid provides supplemental sanitation services	Types of duties assigned to sanitation workers
0	1	General Consulting Research	5220.0	1.0	31.0	9.0	9.0	145.0	Yes	Street Sweeping and Bagging; Snow and Ice Remo...

1 rows × 44 columns

Replacing all "," by "" and all NaN values by 0
Would be required for adding the total sum later

In [16]:

```
df['Miscellaneous charges'] = df['Miscellaneous charges'].fillna("0")
df['Sanitation expenses'] = df['Sanitation expenses'].fillna("0")
df['Public safety expenses'] = df['Public safety expenses'].fillna("0")
df['Streetscape & beautification expenses'] = df['Streetscape & beautification
expenses'].fillna("0")
df['Salaries'] = df['Salaries'].fillna("0")
df['Outside contractor expenses'] = df['Outside contractor expenses'].fillna(
"0")
df['Insurance costs'] = df['Insurance costs'].fillna("0")
df['Supplies and equipment costs'] = df['Supplies and equipment costs'].fillna
("0")
```

In [17]:

```
df['Miscellaneous charges']
```

Out[17]:

```
0         16591
1         68277
2         71153
3         14927
4         61766
          ...
99995     36677
99996    114142
99997      7268
99998     16627
99999    114120
Name: Miscellaneous charges, Length: 100000, dtype: object
```

In [18]:

```
df['Public safety expenses']
```

Out[18]:

```
0         40756
1        169071
2          9179
3        293670
4        320711
          ...
99995    337186
99996     87074
99997     13808
99998     40779
99999     87125
Name: Public safety expenses, Length: 100000, dtype: object
```

In [19]:
```
df['Total Quotation'] = df['Miscellaneous charges'].astype("int") + df['Public safety expenses'].astype("int") +  df['Sanitation expenses'].astype("int") + df['Streetscape & beautification expenses'].astype("int")
+ df['Salaries'].astype("int") + df['Outside contractor expenses'].astype("int") + df['Insurance costs'].astype("int") +  df['Supplies and equipment costs'].astype("int")
df['Total Quotation']
#Streetscape & beautification expenses   Outside contractor expenses
```

Out[19]:
```
0         172826
1         499849
2         207742
3        1070827
4         779014
          ...
99995     626047
99996     927299
99997     700027
99998     172909
99999     927340
Name: Total Quotation, Length: 100000, dtype: int32
```

In [20]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 45 columns):
 #   Column                                                             Non-N
ull Count   Dtype  
---  ------                                                             -----
---------   -----  
 0   index                                                              10000
0 non-null  int64  
 1   company name                                                       10000
0 non-null  object 
 2   Floor Size                                                         97329
non-null   float64
 3   Full-time staff                                                    97329
non-null   float64
 4   Sanitation staff employed                                          97329
non-null   float64
 5   Safety Inspector                                                   97329
non-null   float64
 6   Part-time staff                                                    97329
non-null   float64
 7   Current clients                                                    83270
non-null   float64
 8   Bid provides supplemental sanitation services                      97329
non-null   object 
 9   Types of duties assigned to sanitation workers                     94585
non-null   object 
 10  Days per week of sanitation services                               10000
0 non-null  int64  
 11  Hours logged by sanitation workers                                 97329
non-null   float64
 12  Incidents of graffiti removed                                      10000
0 non-null  int64  
 13  Trash bags collected                                               97329
non-null   float64
 14  Trash and recycling receptacles serviced                           97329
non-null   float64
 15  Bid provides supplemental public safety services                   97329
non-null   object 
 16  Duties assigned to public safety personnel                         44990
non-null   object 
 17  Hours logged by public safety officers                             67622
non-null   float64
 18  Interactions with public safety officers                           66200
non-null   float64
 19  Bid provides supplemental streetscape and beautification services  97329
non-null   object 
 20  Planters and hanging baskets maintained                            97329
non-null   float64
 21  Tree pits maintained                                               97329
non-null   float64
 22  Banners maintained                                                 97329
non-null   float64
 23  Public art installations sponsored                                 97329
non-null   float64
 24  Street furniture elements maintained                               10000
0 non-null  int64  
```

```
 25  Wayfinding elements maintained                                     84631
non-null   object 
 26  Lighting elements maintained                                       83226
non-null   object 
 27  Other infrastructure elements maintained                           86113
non-null   object 
 28  Public spaces maintained                                           84631
non-null   float64
 29  Bid has holiday lighting program                                   97329
non-null   object 
 30  Communication channels used                                        85965
non-null   object 
 31  Social media followers                                             97329
non-null   float64
 32  Marketing materials distributed                                    97329
non-null   float64
 33  Public events coordinated                                          97329
non-null   float64
 34  Estimated attendees to public events coordinated                   97329
non-null   float64
 35  Special event charges                                              99990
non-null   float64
 36  Miscellaneous charges                                              10000
0 non-null  object 
 37  Sanitation expenses                                                10000
0 non-null  object 
 38  Public safety expenses                                             10000
0 non-null  object 
 39  Streetscape & beautification expenses                              10000
0 non-null  object 
 40  Salaries                                                           10000
0 non-null  object 
 41  Outside contractor expenses                                        10000
0 non-null  object 
 42  Insurance costs                                                    10000
0 non-null  object 
 43  Supplies and equipment costs                                       10000
0 non-null  object 
 44  Total Quotation                                                    10000
0 non-null  int32  
dtypes: float64(21), int32(1), int64(4), object(19)
memory usage: 34.0+ MB
```

Removing unnecessary attributes and merging some attributes

In [21]:

```
df.drop(['Types of duties assigned to sanitation workers','Duties assigned to
 public safety personnel'],axis=1,inplace=True)
```

In [22]:

```
df['Incidents of graffiti removed'] = df['Incidents of graffiti removed'].fill
na("0")
```

In [23]:

```
df['Number_Of_Sanitation_Activities'] = df['Incidents of graffiti removed'].as
type("float") + df['Trash bags collected'] + df['Trash and recycling receptacl
es serviced']
```

In [24]:

```
df.drop(['Incidents of graffiti removed','Trash bags collected','Trash and recycling receptacles serviced'],axis=1,inplace=True)
```

In [25]:

```
df['Street furniture elements maintained'] = df['Street furniture elements maintained'].fillna("0")
df['Wayfinding elements maintained'] = df['Wayfinding elements maintained'].fillna("0")
df['Lighting elements maintained'] = df['Lighting elements maintained'].fillna("0")
df['Other infrastructure elements maintained'] = df['Other infrastructure elements maintained'].fillna("0")
```

In [26]:
```
df.dtypes
```

Out[26]:
```
index                                                                  int64
company name                                                          object
Floor Size                                                           float64
Full-time staff                                                      float64
Sanitation staff employed                                            float64
Safety Inspector                                                     float64
Part-time staff                                                      float64
Current clients                                                      float64
Bid provides supplemental sanitation services                         object
Days per week of sanitation services                                   int64
Hours logged by sanitation workers                                   float64
Bid provides supplemental public safety services                      object
Hours logged by public safety officers                               float64
Interactions with public safety officers                             float64
Bid provides supplemental streetscape and beautification services     object
Planters and hanging baskets maintained                              float64
Tree pits maintained                                                 float64
Banners maintained                                                   float64
Public art installations sponsored                                   float64
Street furniture elements maintained                                   int64
Wayfinding elements maintained                                        object
Lighting elements maintained                                          object
Other infrastructure elements maintained                              object
Public spaces maintained                                             float64
Bid has holiday lighting program                                      object
Communication channels used                                           object
Social media followers                                               float64
Marketing materials distributed                                      float64
Public events coordinated                                            float64
Estimated attendees to public events coordinated                     float64
Special event charges                                                float64
Miscellaneous charges                                                 object
Sanitation expenses                                                   object
Public safety expenses                                                object
Streetscape & beautification expenses                                 object
Salaries                                                              object
Outside contractor expenses                                           object
Insurance costs                                                       object
Supplies and equipment costs                                          object
Total Quotation                                                        int32
Number_Of_Sanitation_Activities                                      float64
dtype: object
```

In [27]:

```
df['Beautification_Activities'] = df['Planters and hanging baskets maintained'] + df['Tree pits maintained'] + df['Banners maintained'] + df['Public art installations sponsored']
+ df['Street furniture elements maintained']
+ df['Wayfinding elements maintained']
+ df['Lighting elements maintained']
+ df['Public spaces maintained']

df['Beautification_Activities']
```

Out[27]:

```
0         573.0
1         558.0
2         985.0
3         617.0
4         904.0
          ...  
99995     368.0
99996     953.0
99997     635.0
99998     516.0
99999    1071.0
Name: Beautification_Activities, Length: 100000, dtype: float64
```

In [28]:

```
df.drop(['Planters and hanging baskets maintained','Tree pits maintained','Banners maintained','Public art installations sponsored','Street furniture elements maintained','Wayfinding elements maintained','Lighting elements maintained','Other infrastructure elements maintained','Public spaces maintained'],axis=1,inplace=True)
```

In [29]:

```
df.shape
```

Out[29]:

```
(100000, 33)
```

In [30]:

```
df['Beautification_Activities'].isnull()
```

Out[30]:

```
0        False
1        False
2        False
3        False
4        False
         ...  
99995    False
99996    False
99997    False
99998    False
99999    False
Name: Beautification_Activities, Length: 100000, dtype: bool
```

In [31]:

```
df['Media_Reach'] = df['Social media followers'] + df['Marketing materials distributed'] + (df['Public events coordinated']*df['Estimated attendees to public events coordinated'])
```

In [32]:
```
df.drop(['Social media followers','Marketing materials distributed','Public events coordinated','Estimated attendees to public events coordinated'],axis=1,inplace=True)
```

In [33]:
```
df.drop(['index'],inplace=True, axis=1)
```

In [34]:
```
df.dtypes
```

Out[34]:
```
company name                                                          object
Floor Size                                                           float64
Full-time staff                                                      float64
Sanitation staff employed                                            float64
Safety Inspector                                                     float64
Part-time staff                                                      float64
Current clients                                                      float64
Bid provides supplemental sanitation services                         object
Days per week of sanitation services                                   int64
Hours logged by sanitation workers                                   float64
Bid provides supplemental public safety services                      object
Hours logged by public safety officers                               float64
Interactions with public safety officers                             float64
Bid provides supplemental streetscape and beautification services     object
Bid has holiday lighting program                                      object
Communication channels used                                           object
Special event charges                                                float64
Miscellaneous charges                                                 object
Sanitation expenses                                                   object
Public safety expenses                                                object
Streetscape & beautification expenses                                 object
Salaries                                                              object
Outside contractor expenses                                           object
Insurance costs                                                       object
Supplies and equipment costs                                          object
Total Quotation                                                        int32
Number_Of_Sanitation_Activities                                      float64
Beautification_Activities                                            float64
Media_Reach                                                          float64
dtype: object
```

In [35]:
```
df.drop(['Full-time staff'], inplace=True, axis=1)
```

In [36]:
```
df['Total_Staff'] = df['Sanitation staff employed'] + df['Safety Inspector'] + df['Part-time staff']
```

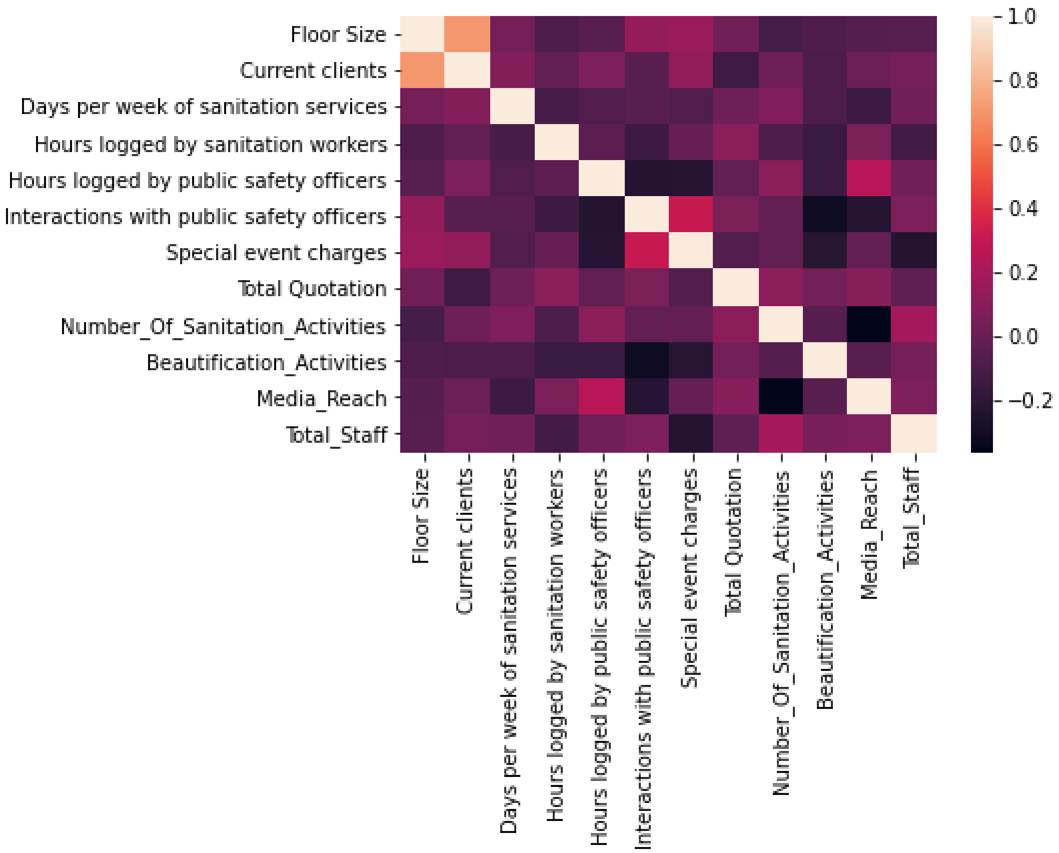
In [37]:
```
df.drop(['Sanitation staff employed','Safety Inspector','Part-time staff'], axis=1, inplace=True)
```

In [38]:
```
df.shape
```

Out[38]:
```
(100000, 26)
```

```
In [39]: import seaborn as sns
         a = df.corr()
         sns.heatmap(a)

Out[39]: <matplotlib.axes._subplots.AxesSubplot at 0x21549d946a0>
```



2. Modelling

```
In [40]: df.shape

Out[40]: (100000, 26)
```

In [41]:
```
df.dtypes
```

Out[41]:
```
company name                                                     object
Floor Size                                                      float64
Current clients                                                 float64
Bid provides supplemental sanitation services                    object
Days per week of sanitation services                              int64
Hours logged by sanitation workers                              float64
Bid provides supplemental public safety services                 object
Hours logged by public safety officers                          float64
Interactions with public safety officers                        float64
Bid provides supplemental streetscape and beautification services    object
Bid has holiday lighting program                                 object
Communication channels used                                      object
Special event charges                                           float64
Miscellaneous charges                                            object
Sanitation expenses                                              object
Public safety expenses                                           object
Streetscape & beautification expenses                            object
Salaries                                                         object
Outside contractor expenses                                      object
Insurance costs                                                  object
Supplies and equipment costs                                     object
Total Quotation                                                   int32
Number_Of_Sanitation_Activities                                 float64
Beautification_Activities                                       float64
Media_Reach                                                     float64
Total_Staff                                                     float64
dtype: object
```

In [42]:
```
df.head(4)
```

Out[42]:

	company name	Floor Size	Current clients	Bid provides supplemental sanitation services	Days per week of sanitation services	Hours logged by sanitation workers	Bid provides supplemental public safety services	Hours logged by public safety officers	Inter wit
0	General Consulting Research	5220.0	145.0	Yes	7	61051.0	No	24732.0	
1	Analysis Analysis	1740.0	168.0	Yes	7	22166.0	Yes	68063.0	
2	Federated Consulting Analysis	8150.0	42.0	No	7	139371.0	Yes	86605.0	
3	Atlantic Max North	41110.0	NaN	Yes	7	24475.0	Yes	31890.0	1

4 rows × 26 columns

In [43]:
```
pd.to_numeric(df['Hours logged by public safety officers'])
```

Out[43]:
```
0        24732.0
1        68063.0
2        86605.0
3        31890.0
4        73549.0
          ...   
99995        NaN
99996        NaN
99997        NaN
99998    37113.0
99999        NaN
Name: Hours logged by public safety officers, Length: 100000, dtype: float64
```

In [44]:
```
# Bid provides supplemental sanitation services --> Sanitation_services_provided
#Bid provides supplemental streetscape and beautification services   ---> beautification_services_provided
#Bid has holiday lighting program   ---> Holiday_program

df['Bid provides supplemental sanitation services'] = df.rename(columns={'Bid provides supplemental sanitation services': 'Sanitation_services_provided'}, inplace=True)
df['Bid provides supplemental streetscape and beautification services'] = df.rename(columns={'Bid provides supplemental streetscape and beautification services': 'Beautification_services_provided'}, inplace=True)
df['Bid has holiday lighting program'] = df.rename(columns={'Bid has holiday lighting program':'Holiday_program'}, inplace=True)
```

In [45]:
```
df.drop(['Bid provides supplemental sanitation services','Bid provides supplemental streetscape and beautification services','Bid has holiday lighting program'], inplace=True, axis=1)
```

In [46]:
```
df.head(1)
```

Out[46]:

	company name	Floor Size	Current clients	Sanitation_services_provided	Days per week of sanitation services	Hours logged by sanitation workers	Bid provides supplemental public safety services
0	General Consulting Research	5220.0	145.0	Yes	7	61051.0	No

1 rows × 26 columns

In [47]:
```
df.shape
```

Out[47]:
```
(100000, 26)
```

Converting object to numeric datatype

In [48]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 26 columns):
 #   Column                                        Non-Null Count   Dtype
---  ------                                        --------------   -----
 0   company name                                  100000 non-null  objec
t
 1   Floor Size                                    97329 non-null   float
64
 2   Current clients                               83270 non-null   float
64
 3   Sanitation_services_provided                  97329 non-null   objec
t
 4   Days per week of sanitation services          100000 non-null  int64
 5   Hours logged by sanitation workers            97329 non-null   float
64
 6   Bid provides supplemental public safety services  97329 non-null   objec
t
 7   Hours logged by public safety officers        67622 non-null   float
64
 8   Interactions with public safety officers      66200 non-null   float
64
 9   Beautification_services_provided              97329 non-null   objec
t
 10  Holiday_program                               97329 non-null   objec
t
 11  Communication channels used                   85965 non-null   objec
t
 12  Special event charges                         99990 non-null   float
64
 13  Miscellaneous charges                         100000 non-null  objec
t
 14  Sanitation expenses                           100000 non-null  objec
t
 15  Public safety expenses                        100000 non-null  objec
t
 16  Streetscape & beautification expenses         100000 non-null  objec
t
 17  Salaries                                      100000 non-null  objec
t
 18  Outside contractor expenses                   100000 non-null  objec
t
 19  Insurance costs                               100000 non-null  objec
t
 20  Supplies and equipment costs                  100000 non-null  objec
t
 21  Total Quotation                               100000 non-null  int32
 22  Number_Of_Sanitation_Activities               97329 non-null   float
64
 23  Beautification_Activities                     97329 non-null   float
64
 24  Media_Reach                                   97329 non-null   float
64
 25  Total_Staff                                   97329 non-null   float
64
dtypes: float64(10), int32(1), int64(1), object(14)
memory usage: 19.5+ MB
```

In [49]:
```
#Holiday_program, Interactions with public safety officers, Bid provides supplemental public safety services
df.drop(['Holiday_program', 'Interactions with public safety officers','Bid provides supplemental public safety services'], axis=1, inplace=True)
```

In [50]:
```
df.drop(['Floor Size'], inplace=True, axis = 1)
```

In [51]:
```
df['Special event charges'] = df['Special event charges'].replace("-","0")
```

In [52]:
```
df['Special event charges'] = df['Special event charges'].fillna("0").astype(int)
```

In [53]:
```
df['Current clients'].isnull().sum()
```

Out[53]:
```
16730
```

In [54]:
```
df.head(5)
```

Out[54]:

	company name	Current clients	Sanitation_services_provided	Days per week of sanitation services	Hours logged by sanitation workers	Hours logged by public safety officers	Beautification
0	General Consulting Research	145.0	Yes	7	61051.0	24732.0	
1	Analysis Analysis	168.0	Yes	7	22166.0	68063.0	
2	Federated Consulting Analysis	42.0	No	7	139371.0	86605.0	
3	Atlantic Max North	NaN	Yes	7	24475.0	31890.0	
4	Star Consulting	181.0	Yes	7	84739.0	73549.0	

5 rows × 22 columns

In [55]:
```
df['Miscellaneous charges'] = df['Miscellaneous charges'].astype(int)
```

In [56]:
```
df['Sanitation expenses'] = df['Sanitation expenses'].astype(int)
```

In [57]:

```
df['Public safety expenses'] = df['Public safety expenses'].astype(int)
```

In [58]:

```
df['Streetscape & beautification expenses'] = df['Streetscape & beautification expenses'].astype(int)
```

In [59]:

```
df['Salaries'] = df['Salaries'].astype(int)
```

In [60]:

```
df['Outside contractor expenses'] = df['Outside contractor expenses'].astype(int)
```

In [61]:

```
df['Insurance costs'] = df['Insurance costs'].astype(int)
```

In [62]:

```
df['Supplies and equipment costs'] = df['Supplies and equipment costs'].astype(int)
```

In [63]:

```
import numpy as np
df['Current clients'] = pd.to_numeric(df['Current clients'], errors='coerce')
df['Current clients']
```

Out[63]:

```
0          145.0
1          168.0
2           42.0
3            NaN
4          181.0
           ...  
99995      567.0
99996      518.0
99997     1423.0
99998      231.0
99999      981.0
Name: Current clients, Length: 100000, dtype: float64
```

In [64]:

```
df['Current clients'] = df['Current clients'].replace(r'^\s*$', np.nan, regex=True).fillna(method ='pad')
df['Days per week of sanitation services'] = df['Days per week of sanitation services'].replace(r'^\s*$', np.nan, regex=True).fillna(method ='pad')
df = df.replace(r'^\s*$', np.nan, regex=True).fillna(method ='pad')
df[pd.to_numeric(df['Current clients'], errors='coerce').notnull()]
```

Out[64]:

	company name	Current clients	Sanitation_services_provided	Days per week of sanitation services	Hours logged by sanitation workers	Hours logged by public safety officers	Beaut
0	General Consulting Research	145.0	Yes	7	61051.0	24732.0	
1	Analysis Analysis	168.0	Yes	7	22166.0	68063.0	
2	Federated Consulting Analysis	42.0	No	7	139371.0	86605.0	
3	Atlantic Max North	42.0	Yes	7	24475.0	31890.0	
4	Star Consulting	181.0	Yes	7	84739.0	73549.0	
...	...	...	...	...	...	...	
99995	Vision Innovation Analysis	567.0	Yes	7	119454.0	101584.0	
99996	Architecture Provider Industries	518.0	Yes	7	174290.0	101584.0	
99997	Construction Omega Vision	1423.0	No	7	92340.0	101584.0	
99998	General Virtual Innovation	231.0	Yes	7	63951.0	37113.0	
99999	Federated Systems People	981.0	Yes	7	171611.0	37113.0	

100000 rows × 22 columns

In [65]:

```
df["Current clients"]
df['Days per week of sanitation services']
```

Out[65]:

```
0        7
1        7
2        7
3        7
4        7
        ..
99995    7
99996    7
99997    7
99998    7
99999    7
Name: Days per week of sanitation services, Length: 100000, dtype: int64
```

In [66]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 22 columns):
 #   Column                                 Non-Null Count   Dtype  
---  ------                                 --------------   -----  
 0   company name                           100000 non-null  object 
 1   Current clients                        100000 non-null  float64
 2   Sanitation_services_provided           100000 non-null  object 
 3   Days per week of sanitation services   100000 non-null  int64  
 4   Hours logged by sanitation workers     100000 non-null  float64
 5   Hours logged by public safety officers 100000 non-null  float64
 6   Beautification_services_provided       100000 non-null  object 
 7   Communication channels used            100000 non-null  object 
 8   Special event charges                  100000 non-null  int32  
 9   Miscellaneous charges                  100000 non-null  int32  
 10  Sanitation expenses                    100000 non-null  int32  
 11  Public safety expenses                 100000 non-null  int32  
 12  Streetscape & beautification expenses  100000 non-null  int32  
 13  Salaries                               100000 non-null  int32  
 14  Outside contractor expenses            100000 non-null  int32  
 15  Insurance costs                        100000 non-null  int32  
 16  Supplies and equipment costs           100000 non-null  int32  
 17  Total Quotation                        100000 non-null  int32  
 18  Number_Of_Sanitation_Activities        100000 non-null  float64
 19  Beautification_Activities              100000 non-null  float64
 20  Media_Reach                            100000 non-null  float64
 21  Total_Staff                            100000 non-null  float64
dtypes: float64(7), int32(10), int64(1), object(4)
memory usage: 13.0+ MB
```

In [67]:

```
df.head()
```

Out[67]:

	company name	Current clients	Sanitation_services_provided	Days per week of sanitation services	Hours logged by sanitation workers	Hours logged by public safety officers	Beautificatio
0	General Consulting Research	145.0	Yes	7	61051.0	24732.0	
1	Analysis Analysis	168.0	Yes	7	22166.0	68063.0	
2	Federated Consulting Analysis	42.0	No	7	139371.0	86605.0	
3	Atlantic Max North	42.0	Yes	7	24475.0	31890.0	
4	Star Consulting	181.0	Yes	7	84739.0	73549.0	

5 rows × 22 columns

In [68]:

```
cat_feats = ['Sanitation_services_provided','Beautification_services_provided'
]
```

In [69]:

```
final_df = pd.get_dummies(df, columns=cat_feats,drop_first=True)
```

In [70]:

```
final_df.describe()
```

Out[70]:

	Current clients	Days per week of sanitation services	Hours logged by sanitation workers	Hours logged by public safety officers	Special event charges	Miscellaneou charg
count	100000.000000	100000.000000	100000.000000	100000.000000	100000.000000	100000.00000
mean	903.964740	6.565800	95767.975590	65579.373620	252070.033880	61996.0492
std	829.924865	1.245426	44363.815934	36575.526928	142211.597494	36506.4761
min	9.000000	1.000000	6558.000000	1065.000000	0.000000	0.0000
25%	463.000000	7.000000	61208.500000	32661.500000	167592.000000	27387.0000
50%	721.000000	7.000000	89714.500000	65600.500000	281710.000000	61759.0000
75%	981.000000	7.000000	134327.000000	98972.000000	402157.000000	99140.0000
max	5889.000000	7.000000	191164.000000	134353.000000	454427.000000	122191.0000

In [71]:
```
final_df.drop(['Communication channels used'], inplace = True, axis = 1)
```

In [72]:
```
final_df.describe()
```

Out[72]:

	Current clients	Days per week of sanitation services	Hours logged by sanitation workers	Hours logged by public safety officers	Special event charges	Miscellaneou charg
count	100000.000000	100000.000000	100000.000000	100000.000000	100000.000000	100000.0000
mean	903.964740	6.565800	95767.975590	65579.373620	252070.033880	61996.0492
std	829.924865	1.245426	44363.815934	36575.526928	142211.597494	36506.4761
min	9.000000	1.000000	6558.000000	1065.000000	0.000000	0.0000
25%	463.000000	7.000000	61208.500000	32661.500000	167592.000000	27387.0000
50%	721.000000	7.000000	89714.500000	65600.500000	281710.000000	61759.0000
75%	981.000000	7.000000	134327.000000	98972.000000	402157.000000	99140.0000
max	5889.000000	7.000000	191164.000000	134353.000000	454427.000000	122191.0000

In [73]:
```
final_df.head()
```

Out[73]:

	company name	Current clients	Days per week of sanitation services	Hours logged by sanitation workers	Hours logged by public safety officers	Special event charges	Miscellaneous charges	Sanitation expenses	Pul sat expen
0	General Consulting Research	145.0	7	61051.0	24732.0	360146	16591	76277	40
1	Analysis Analysis	168.0	7	22166.0	68063.0	360146	68277	82436	169
2	Federated Consulting Analysis	42.0	7	139371.0	86605.0	360146	71153	82436	9
3	Atlantic Max North	42.0	7	24475.0	31890.0	360146	14927	574602	293
4	Star Consulting	181.0	7	84739.0	73549.0	206317	61766	340025	320

5 rows × 21 columns

In [74]:
```
#Separating the Attribute and target attribute
y = final_df.iloc[0:1000,0] #Dependent
x = final_df.iloc[0:1000,1:21] #Independent attributes
```

In [75]:
```
x
```

Out[75]:

	Current clients	Days per week of sanitation services	Hours logged by sanitation workers	Hours logged by public safety officers	Special event charges	Miscellaneous charges	Sanitation expenses	Public safety expenses	S bea
0	145.0	7	61051.0	24732.0	360146	16591	76277	40756	
1	168.0	7	22166.0	68063.0	360146	68277	82436	169071	
2	42.0	7	139371.0	86605.0	360146	71153	82436	9179	
3	42.0	7	24475.0	31890.0	360146	14927	574602	293670	
4	181.0	7	84739.0	73549.0	206317	61766	340025	320711	
...	...	...	...	...	...	...	...	...	
995	394.0	7	118128.0	104236.0	19576	59436	177311	236721	
996	315.0	3	79633.0	104236.0	196511	49243	322515	204908	
997	627.0	7	85799.0	26315.0	281710	64092	316413	133253	
998	3145.0	7	66240.0	121560.0	281710	27345	169937	86676	
999	3379.0	7	65821.0	124580.0	281710	27398	169967	86681	

1000 rows × 20 columns

In [76]:
```
x.shape
```

Out[76]:
```
(1000, 20)
```

In [77]:
```
y
```

Out[77]:
```
0         General Consulting Research
1                   Analysis Analysis
2       Federated Consulting Analysis
3                    Atlantic Max North
4                       Star Consulting
                     ...
995            Source Atlantic Signal
996                   Provider Direct
997         Power Internet Construction
998            Vision Analysis Galaxy
999                     Power Net East
Name: company name, Length: 1000, dtype: object
```

In [78]:
```
x.shape
```

Out[78]:
```
(1000, 20)
```

In [79]:
```
y.shape
```

Out[79]:
```
(1000,)
```

In [90]:

```
TestValGiven = testing.iloc[0:20000]
TestValGiven
```

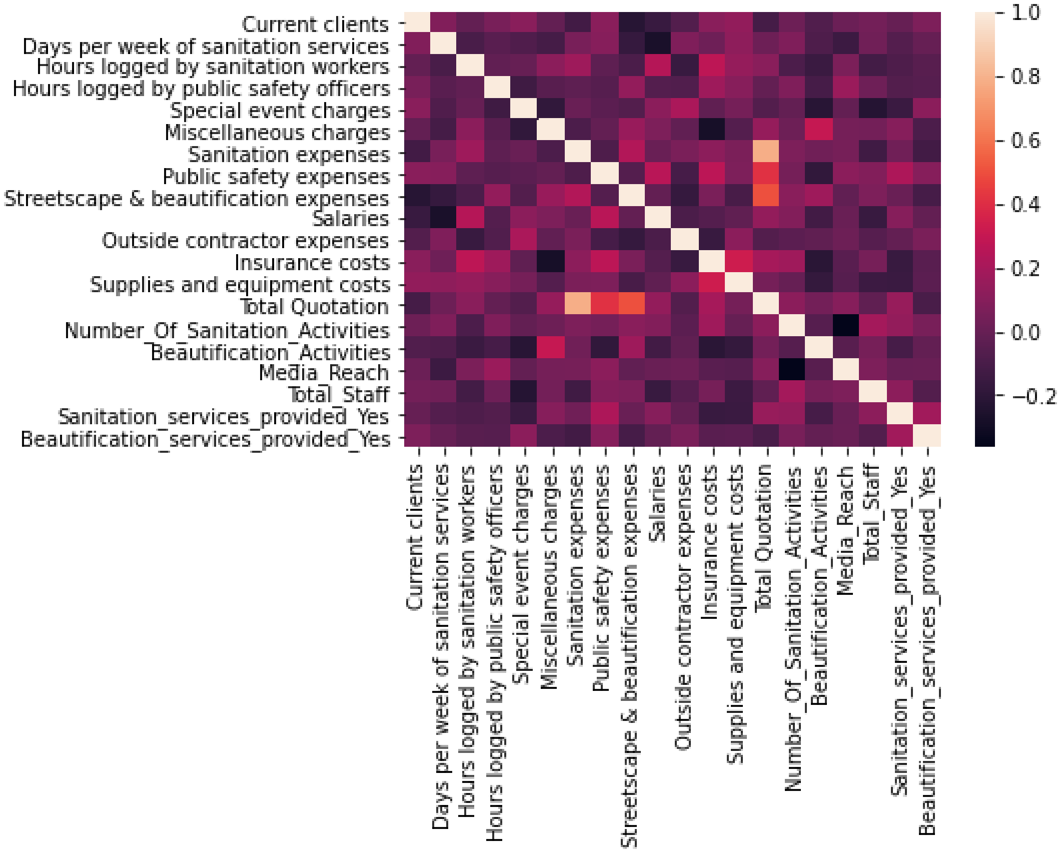
Out[90]:

	company name	Current clients	Days per week of sanitation services	Hours logged by sanitation workers	Hours logged by public safety officers	Special event charges	Miscellaneous charges	Sanitation expenses
0	Venture Universal Solutions	902	7	78588	30644	360146	16620	76314
1	Software Galaxy People	214	7	94859	84370	206317	61751	340053
2	Direct Resource Venture	214	7	94859	84370	0	0	0
3	Provider Omega Electronics	214	7	177449	84370	423722	82043	327027
4	Hill North Future	406	6	154966	48742	402157	22276	502772
...	...	...	...	...	...	...	...	...
19995	Contract Software Telecom	407	7	126790	87631	403576	54452	231242
19996	Systems Network Technology	684	7	114248	101584	281710	47112	449038
19997	Vision Innovation Analysis	567	7	119454	101584	403576	36677	61664
19998	Architecture Provider Industries	518	7	174290	101584	225721	114142	568108
19999	Construction Omega Vision	1423	7	92340	101584	403576	7268	494811

20000 rows × 21 columns

In [81]:
```
import seaborn as sns
a = final_df.corr()
sns.heatmap(a)
```

Out[81]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x2154a598910>
```



In [82]:
```
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics #Import scikit-learn metrics module for accuracy c
alculation
```

In [83]:
```
# Split dataset into training set and test set
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.5, rando
m_state=1) # 70% training and 30% test
```

In [84]:
```
4# Create Decision Tree classifer object
clf = DecisionTreeClassifier()
```

In [85]:
```
# Train Decision Tree Classifer
clf = clf.fit(X_train,y_train)
```

In [113]:

```
def Accurator():
    result = []
    Accuracies = []
    for i in range(len(TestValGiven)):
        helloArr = []
        y_pred = clf.predict([TestValGiven.iloc[i,1:21]]) #Predicting on sent company information requirements by test clients
        valuesTest = TestValGiven.iloc[i,1:21].values #Extracting company information on sent requirements
        abc = final_df[final_df["company name"]==y_pred[0]].index.values #Extracting company information based on index of predicted company
        helloArr = final_df.iloc[abc[0],1:21].values #Extracting company information based on index of predicted company
        Accuracy = 0 #Blank variable for accuracy of individual request, based on fulfilled parameters
        for j in range(0,20): #Awarding accuracy of individual request, based on fulfilled parameters with 20% margin allowed.
            if(helloArr[j]>=valuesTest[j]):
                if(helloArr[j] - valuesTest[j] <= valuesTest[j]*0.2):  #Value is within 20% higher than or equal to  required range
                    Accuracy += 5 #20 features are considered, each awards 5% points.
            elif(helloArr[j]<valuesTest[j]):
                if(valuesTest[j] - helloArr[j] <= valuesTest[j]*0.2): #Value is within 20% lower than required range
                    Accuracy += 5 #20 features are considered, each awards 5% points.
        Accuracies.append([i,Accuracy])
        result.append([i,y_pred[0],helloArr])
    sum1=0
    for i in range(0,len(TestValGiven)):
        sum1 += (Accuracies[i][1])
    print("Accuracy:",(sum1/len(TestValGiven))) #averaging fulfillment accuracy
    for i in range(5):  #company info
        print(i)
        print("Predicted Company: ",y_pred[0])
        print("Predicted Company Fulfillment %: ",Accuracies[i][1])
        print("\n")
```

In [114]:

```
Accurator()
```

```
Accuracy: 83.65375
0
Predicted Company:  Resource Universal Technology
Predicted Company Fulfillment %:  80


1
Predicted Company:  Resource Universal Technology
Predicted Company Fulfillment %:  95


2
Predicted Company:  Resource Universal Technology
Predicted Company Fulfillment %:  15


3
Predicted Company:  Resource Universal Technology
Predicted Company Fulfillment %:  95


4
Predicted Company:  Resource Universal Technology
Predicted Company Fulfillment %:  85
```

In [ ]:

In [ ]: