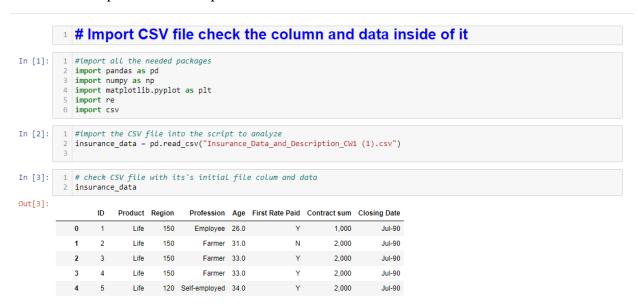
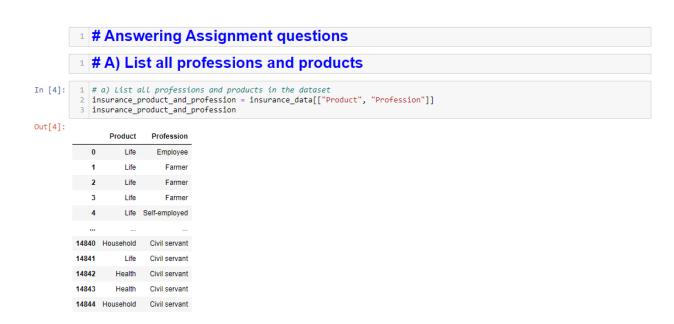
Assignment 01

Questions 01 and Answers

1. List all professions and products in the dataset.





Import all the libraries on the top and develop the script to get output for product and profession.

2. Show any problem with the data etc.

The main problem in this data set is Wrong data types and N/A, nan values – which mean Null values. The solutions are provided, then transformed into to correct data type also removed Null values from the table.

```
1 # B) Show problems & Handling problems
           # Show any problem with the data etc.
# 1. check whether the csv file have any null values
missing_value = ["N/a", "na", np.nan]
insurance_data = pd.read_csv("Insurance_Data_and_Description_CW1 (1).csv", na_values = missing_value)
In [5]:
            5 insurance_data.isnull().sum()
            6 insurance_data = insurance_data.dropna()
            7 insurance_data.isnull().sum()
Out[5]: ID
          Product
          Region
                                 0
          Profession
          First Rate Paid
          Contract sum
                                 0
          Closing Date
          dtype: int64
```

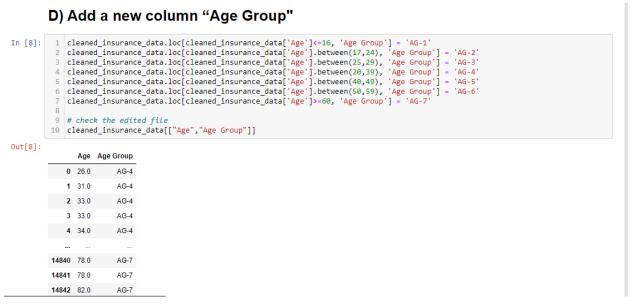
3. Clean and transform the data into the correct data types.

The abovementioned problem was sorted here.

```
In [8]: 1
In [6]: 1 #As this result there are so many Null values in both of the columns
2 #In order to clean these problem removing all the null values from dataset
3
4 insurance_data['Contract sum'] = insurance_data['Contract sum'].str.replace(',','0')
5 insurance_data['Contract sum'] = insurance_data['Contract sum'].str.replace('K','0')
6 insurance_data['Contract sum'] = insurance_data['Contract sum'].str.replace('k','0')
7
8
9 insurance_data['Contract sum'] = insurance_data['Contract sum'].replace(',',"0")
10 insurance_data['Contract sum'] = insurance_data['Contract sum'].astype(int)
11
12 cleaned_insurance_data = insurance_data.dropna()
13
14 cleaned_insurance_data
```

4. Add a new column "Age Group" and fill up its data according to below criteria.

here I am using between and comparison operators to check the age and created the colum for Age group.



5. Rank region by total insurance product.

using the count function to for region and group by the to identify the total insurance product for each Region.

```
1 #E) Rank region by total insurance product
In [9]: 1 Rank = cleaned_insurance_data.groupby('Region',sort=False).count()
2 Rank['Product']
Out[9]: Region
150     12281
120     16
160     2265
999     195
130     36
Name: Product, dtype: int64
```

6. Find the mean, median, standard deviation of Age and Contract Sum.

Finding the summary of Age group and Contract sum in the dataset, which Is showing up mean, median etc.

```
4 F) Find the mean, median, standard deviation of Age and Contract Sum
In [10]:
          cleaned_insurance_data["Age"].describe()
Out[10]: count
               14793.00000
                   44.93578
        mean
                    1.00000
         min
         25%
                   34.00000
         50%
                   45.00000
         75%
                   57.00000
         max
                   95.00000
        Name: Age, dtype: float64
In [11]: 1 # 2. For Contract Sum
          2 insurance_data['Contract sum'].astype(str).astype(int)
          3 insurance_data["Contract sum"].describe()
Out[11]: count
                1.479300e+04
                2.892438e+05
         mean
                2.584105e+05
         min
                4 9999999+91
         25%
                1.500000e+05
                2.000000e+05
         75%
                3.300000e+05
         max
                4.250000e+06
        Name: Contract sum, dtype: float64
```

7. Filter the rows of Age < 25 and Contract Sum between 50k and 100k and insurance type of Health or Life insurance.

Filtering the data with above condition, I used single line comparison to done the filtering.



8. Find total insurance product by profession.

Finding product over the profession, by grouping profession and calculate product in the professions.

9. Find correlations between product, profession, and contract sum. What are your conclusions?

The correlation is can only calculate with numerical values, so other variable could not do for the correlation.

10. Calculate and Visualize products on age group. Describe your findings