**Assignment -2**

Python Programming

| | |
|---|---|
| Assignment Date | 26 sep 2022 |
| Student Name | Ms.D.Janani |
| Student Roll Number | 510119106004 |
| Maximum Marks | 2 Marks |

## Python programming

```
import pandas as pd
import numpy as np
```

```
data=pd.read_csv("/content/drive/MyDrive/Dataset.csv")
```

```
data.head()
```

```
   RowNumber      CustomerId   Surname  CreditScore Geography  Gender  Age  \
0          1        15634602  Hargrave          619    France  Female   42
1          2        15647311      Hill          608     Spain  Female   41
2          3        15619304      Onio          502    France  Female   42
3          4        15701354      Boni          699    France  Female   39
4          5        15737888  Mitchell          850     Spain  Female   43

   Tenure  Balance  NumOfProducts  HasCrCard  IsActiveMember  \
0       2     0.00              1          1               1
1       1    83807.86           1          0               1
2       8   159660.80           3          1               0
3       1     0.00              2          0               0
4       2   125510.82           1          1               1

   EstimatedSalary  Exited0
0        101348.88        1
1        112542.58        0
2        113931.57        1
3         93826.63        0
4         79084.10        0
```

```
data.tail()
```

```
      RowNumber  CustomerId    Surname  CreditScore Geography  Gender  Age  \
9995       9996    15606229   Obijiaku          771    France    Male   39
```

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 9996 | 9997 | 15569892 | Johnstone | 516 | France | Male | 35 |
| 9997 | 9998 | 15584532 | Liu | 709 | France | Female | 36 |
| 9998 | 9999 | 15682355 | Sabbatini | 772 | Germany | Male | 42 |

```
9999       10000    15628319     Walker          792    France  Female   28

      Tenure    Balance  NumOfProducts  HasCrCard  IsActiveMember  \
9995       5       0.00              2          1               0
9996      10   57369.61              1          1               1
9997       7       0.00              1          0               1
9998       3   75075.31              2          1               0
9999       4  130142.79              1          1               0

      EstimatedSalary  Exited
9995         96270.64       0
9996        101699.77       0
9997         42085.58       1
9998         92888.52       1
9999         38190.78       0
```

```
data.shape
```

```
(10000, 14)
```

```
data.describe()
```

```
         RowNumber    CustomerId   CreditScore           Age        Tenure  \
count  10000.00000  1.000000e+04  10000.000000  10000.000000  10000.000000
mean    5000.50000  1.569094e+07    650.528800     38.921800      5.012800
std     2886.89568  7.193619e+04     96.653299     10.487806      2.892174
min        1.00000  1.556570e+07    350.000000     18.000000      0.000000
25%     2500.75000  1.562853e+07    584.000000     32.000000      3.000000
50%     5000.50000  1.569074e+07    652.000000     37.000000      5.000000
75%     7500.25000  1.575323e+07    718.000000     44.000000      7.000000
max    10000.00000  1.581569e+07    850.000000     92.000000     10.000000

             Balance  NumOfProducts    HasCrCard  IsActiveMember  \
count   10000.000000   10000.000000  10000.00000    10000.000000
mean    76485.889288       1.530200      0.70550        0.515100
std     62397.405202       0.581654      0.45584        0.499797
min         0.000000       1.000000      0.00000        0.000000
25%         0.000000       1.000000      0.00000        0.000000
50%     97198.540000       1.000000      1.00000        1.000000
75%    127644.240000       2.000000      1.00000        1.000000
max    250898.090000       4.000000      1.00000        1.000000

       EstimatedSalary        Exited
count     10000.000000  10000.000000
mean     100090.239881      0.203700
std       57510.492818      0.402769
min          11.580000      0.000000
25%       51002.110000      0.000000
50%      100193.915000      0.000000
```

```
75%      149388.247500      0.000000
max      199992.480000      1.000000
```

```
data.mean()
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError.  Select only valid columns before calling the reduction.
  """Entry point for launching an IPython kernel.
```

```
RowNumber          5.000500e+03
CustomerId         1.569094e+07
CreditScore        6.505288e+02
Age                3.892180e+01
Tenure             5.012800e+00
Balance            7.648589e+04
NumOfProducts      1.530200e+00
HasCrCard          7.055000e-01
IsActiveMember     5.151000e-01
EstimatedSalary    1.000902e+05
Exited             2.037000e-01
dtype: float64
```

```
dir(data)
```

```
['Age',
 'Balance',
 'CreditScore',
 'CustomerId',
 'EstimatedSalary',
 'Exited',
 'Gender',
 'Geography',
 'HasCrCard',
 'IsActiveMember',
 'NumOfProducts',
 'RowNumber',
 'Surname',
 'T',
 'Tenure',
 '_AXIS_LEN',
 '_AXIS_ORDERS',
 '_AXIS_REVERSED',
 '_AXIS_TO_AXIS_NUMBER',
 '_HANDLED_TYPES',
 '__abs__',
 '__add__',
 '__and__',
```

```
'__annotations__',
'__array__',
'__array_priority__',
'__array_ufunc__',
'__array_wrap__',
'__bool__',
'__class__',
'__contains__',
'__copy__',
'__deepcopy__',
'__delattr__',
'__delitem__',
'__dict__',
'__dir__',
'__divmod__',
'__doc__',
'__eq__',
'__finalize__',
'__floordiv__',
'__format__',
'__ge__',
'__getattr__',
'__getattribute__',
'__getitem__',
'__getstate__',
'__gt__',
'__hash__',
'__iadd__',
'__iand__',
'__ifloordiv__',
'__imod__',
'__imul__',
'__init__',
'__init_subclass__',
'__invert__',
'__ior__',
'__ipow__',
'__isub__',
'__iter__',
'__itruediv__',
'__ixor__',
'__le__',
'__len__',
'__lt__',
'__matmul__',
'__mod__',
'__module__',
'__mul__',
'__ne__',
'__neg__',
'__new__',
```

```
'__nonzero__',
'__or__',
'__pos__',
'__pow__',
'__radd__',
'__rand__',
'__rdivmod__',
'__reduce__',
'__reduce_ex__',
'__repr__',
'__rfloordiv__',
'__rmatmul__',
'__rmod__',
'__rmul__',
'__ror__',
'__round__',
'__rpow__',
'__rsub__',
'__rtruediv__',
'__rxor__',
'__setattr__',
'__setitem__',
'__setstate__',
'__sizeof__',
'__str__',
'__sub__',
'__subclasshook__',
'__truediv__',
'__weakref__',
'_xor_',
'_accessors',
'_accum_func',
'_add_numeric_operations',
'_agg_by_level',
'_agg_examples_doc',
'_agg_summary_and_see_also_doc',
'_align_frame',
'_align_series',
'_arith_method',
'_as_manager',
'_attrs',
'_box_col_values',
'_can_fast_transpose',
'_check_inplace_and_allows_duplicate_labels',
'_check_inplace_setting',
'_check_is_chained_assignment_possible',
'_check_label_or_level_ambiguity',
'_check_setitem_copy',
'_clear_item_cache',
'_clip_with_one_bound',
'_clip_with_scalar',
```

```
'_cmp_method',
'_combine_frame',
'_consolidate',
'_consolidate_inplace',
'_construct_axes_dict',
'_construct_axes_from_arguments',
'_construct_result',
'_constructor',
'_constructor_sliced',
'_convert',
'_count_level',
'_data',
'_dir_additions',
'_dir_deletions',
'_dispatch_frame_op',
'_drop_axis',
'_drop_labels_or_levels',
'_ensure_valid_index',
'_find_valid_index',
'_flags',
'_from_arrays',
'_from_mgr',
'_get_agg_axis',
'_get_axis',
'_get_axis_name',
'_get_axis_number',
'_get_axis_resolvers',
'_get_block_manager_axis',
'_get_bool_data',
'_get_cleaned_column_resolvers',
'_get_column_array',
'_get_index_resolvers',
'_get_item_cache',
'_get_label_or_level_values',
'_get_numeric_data',
'_get_value',
'_getitem_bool_array',
'_getitem_multilevel',
'_gotitem',
'_hidden_attrs',
'_indexed_same',
'_info_axis',
'_info_axis_name',
'_info_axis_number',
'_info_repr',
'_init_mgr',
'_inplace_method',
'_internal_names',
'_internal_names_set',
'_is_copy',
'_is_homogeneous_type',
```

```
'_is_label_or_level_reference',
'_is_label_reference',
'_is_level_reference',
'_is_mixed_type',
'_is_view',
'_iset_item',
'_iset_item_mgr',
'_iset_not_inplace',
'_item_cache',
'_iter_column_arrays',
'_ixs',
'_join_compat',
'_logical_func',
'_logical_method',
'_maybe_cache_changed',
'_maybe_update_cacher',
'_metadata',
'_mgr',
'_min_count_stat_function',
'_needs_reindex_multi',
'_protect_consolidate',
'_reduce',
'_reindex_axes',
'_reindex_columns',
'_reindex_index',
'_reindex_multi',
'_reindex_with_indexers',
'_replace_columnwise',
'_repr_data_resource_',
'_repr_fits_horizontal_',
'_repr_fits_vertical_',
'_repr_html_',
'_repr_latex_',
'_reset_cache',
'_reset_cacher',
'_sanitize_column',
'_series',
'_set_axis',
'_set_axis_name',
'_set_axis_nocheck',
'_set_is_copy',
'_set_item',
'_set_item_frame_value',
'_set_item_mgr',
'_set_value',
'_setitem_array',
'_setitem_frame',
'_setitem_slice',
'_slice',
'_stat_axis',
'_stat_axis_name',
```

```
'_stat_axis_number',
'_stat_function',
'_stat_function_ddof',
'_take_with_is_copy',
'_to_dict_of_blocks',
'_typ',
'_update_inplace',
'_validate_dtype',
'_values',
'_where',
'abs',
'add',
'add_prefix',
'add_suffix',
'agg',
'aggregate',
'align',
'all',
'any',
'append',
'apply',
'applymap',
'asfreq',
'asof',
'assign',
'astype',
'at',
'at_time',
'attrs',
'axes',
'backfill',
'between_time',
'bfill',
'bool',
'boxplot',
'clip',
'columns',
'combine',
'combine_first',
'compare',
'convert_dtypes',
'copy',
'corr',
'corrwith',
'count',
'cov',
'cummax',
'cummin',
'cumprod',
'cumsum',
'describe',
```

```
'diff',
'div',
'divide',
'dot',
'drop',
'drop_duplicates',
'droplevel',
'dropna',
'dtypes',
'duplicated',
'empty',
'eq',
'equals',
'eval',
'ewm',
'expanding',
'explode',
'ffill',
'fillna',
'filter',
'first',
'first_valid_index',
'flags',
'floordiv',
'from_dict',
'from_records',
'ge',
'get',
'groupby',
'gt',
'head',
'hist',
'iat',
'idxmax',
'idxmin',
'iloc',
'index',
'infer_objects',
'info',
'insert',
'interpolate',
'isin',
'isna',
'isnull',
'items',
'iteritems',
'iterrows',
'itertuples',
'join',
'keys',
'kurt',
```

```
'kurtosis',
'last',
'last_valid_index',
'le',
'loc',
'lookup',
'lt',
'mad',
'mask',
'max',
'mean',
'median',
'melt',
'memory_usage',
'merge',
'min',
'mod',
'mode',
'mul',
'multiply',
'ndim',
'ne',
'nlargest',
'notna',
'notnull',
'nsmallest',
'nunique',
'pad',
'pct_change',
'pipe',
'pivot',
'pivot_table',
'plot',
'pop',
'pow',
'prod',
'product',
'quantile',
'query',
'radd',
'rank',
'rdiv',
'reindex',
'reindex_like',
'rename',
'rename_axis',
'reorder_levels',
'replace',
'resample',
'reset_index',
'rfloordiv',
```

```
'rmod',
'rmul',
'rolling',
'round',
'rpow',
'rsub',
'rtruediv',
'sample',
'select_dtypes',
'sem',
'set_axis',
'set_flags',
'set_index',
'shape',
'shift',
'size',
'skew',
'slice_shift',
'sort_index',
'sort_values',
'squeeze',
'stack',
'std',
'style',
'sub',
'subtract',
'sum',
'swapaxes',
'swaplevel',
'tail',
'take',
'to_clipboard',
'to_csv',
'to_dict',
'to_excel',
'to_feather',
'to_gbq',
'to_hdf',
'to_html',
'to_json',
'to_latex',
'to_markdown',
'to_numpy',
'to_parquet',
'to_period',
'to_pickle',
'to_records',
'to_sql',
'to_stata',
'to_string',
'to_timestamp',
```

```
 'to_xarray',
 'to_xml',
 'transform',
 'transpose',
 'truediv',
 'truncate',
 'tz_convert',
 'tz_localize',
 'unstack',
 'update',
 'value_counts',
 'values',
 'var',
 'where',
 'xs']
```

```
data.iloc[1,:]
```

```
RowNumber                2
CustomerId            15647311
Surname                   Hill
CreditScore                608
Geography                Spain
Gender                  Female
Age                         41
Tenure                       1
Balance                83807.86
NumOfProducts                1
HasCrCard                    0
IsActiveMember               1
EstimatedSalary     112542.58
Exited                       0
Name: 1, dtype: object
```

```
data.loc[1]
```

```
RowNumber                2
CustomerId            15647311
Surname                   Hill
CreditScore                608
Geography                Spain
Gender                  Female
Age                         41
Tenure                       1
Balance                83807.86
NumOfProducts                1
HasCrCard                    0
IsActiveMember               1
EstimatedSalary     112542.58
```

```
Exited                        0
Name: 1, dtype: object
```

```
#dealing with missing data
data.isnull().sum()
```

```
RowNumber           0
CustomerId          0
Surname             0
CreditScore         0
Geography           0
Gender              0
Age                 0
Tenure              0
Balance             0
NumOfProducts       0
HasCrCard           0
IsActiveMember      0
EstimatedSalary     0
Exited              0
dtype: int64
```

```
data['Age'].fillna(data['Age'].mean())
```

```
0       42
1       41
2       42
3       39
4       43
        ..
9995    39
9996    35
9997    36
9998    42
9999    28
Name: Age, Length: 10000, dtype: int64
```

```
data.isnull().sum()
```

```
RowNumber           0
CustomerId          0
Surname             0
CreditScore         0
Geography           0
Gender              0
Age                 0
Tenure              0
Balance             0
```

```
NumOfProducts       0
HasCrCard           0
IsActiveMember      0
EstimatedSalary     0
Exited              0
dtype: int64
```

```
data['Age'].fillna(data['Age'].mean(),inplace=True)
```

```
data.isnull().sum()
```

```
RowNumber           0
CustomerId          0
Surname             0
CreditScore         0
Geography           0
Gender              0
Age                 0
Tenure              0
Balance             0
NumOfProducts       0
HasCrCard           0
IsActiveMember      0
EstimatedSalary     0
Exited              0
dtype: int64
```

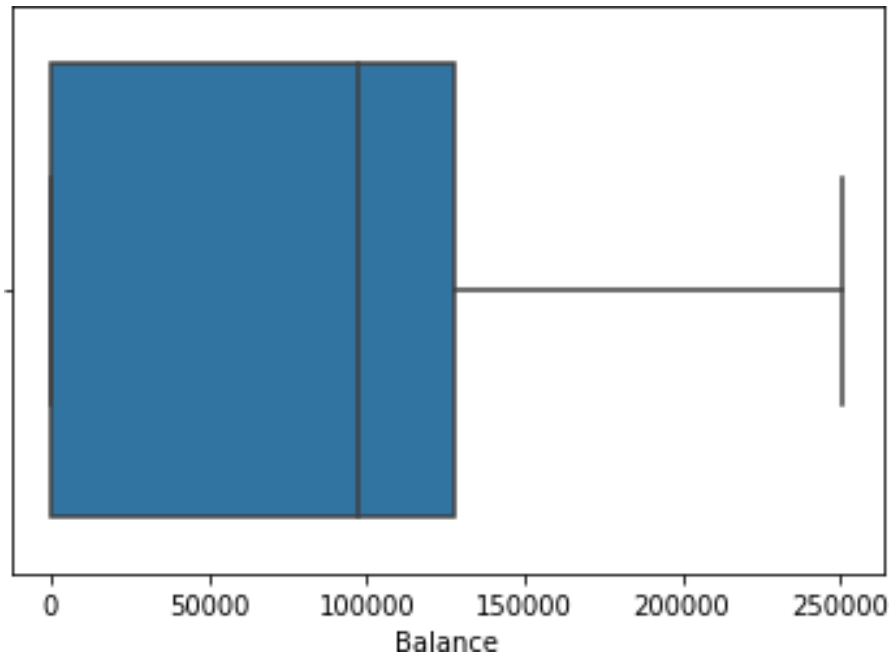
```
data.mode()
```

```
      RowNumber  CustomerId Surname  CreditScore Geography Gender   Age  \
0             1    15565701   Smith        850.0    France   Male  37.0
1             2    15565706     NaN          NaN       NaN    NaN   NaN
2             3    15565714     NaN          NaN       NaN    NaN   NaN
3             4    15565779     NaN          NaN       NaN    NaN   NaN
4             5    15565796     NaN          NaN       NaN    NaN   NaN
...         ...         ...     ...          ...       ...    ...   ...
9995       9996    15815628     NaN          NaN       NaN    NaN   NaN
9996       9997    15815645     NaN          NaN       NaN    NaN   NaN
9997       9998    15815656     NaN          NaN       NaN    NaN   NaN
9998       9999    15815660     NaN          NaN       NaN    NaN   NaN
9999      10000    15815690     NaN          NaN       NaN    NaN   NaN

      Tenure  Balance  NumOfProducts  HasCrCard  IsActiveMember  \
0        2.0      0.0            1.0        1.0             1.0
1        NaN      NaN            NaN        NaN             NaN
2        NaN      NaN            NaN        NaN             NaN
3        NaN      NaN            NaN        NaN             NaN
4        NaN      NaN            NaN        NaN             NaN
...      ...      ...            ...        ...             ...
```

```
9995     NaN      NaN            NaN        NaN             NaN
9996     NaN      NaN            NaN        NaN             NaN
9997     NaN      NaN            NaN        NaN             NaN
9998     NaN      NaN            NaN        NaN             NaN
9999     NaN      NaN            NaN        NaN             NaN

      EstimatedSalary  Exited
0            24924.92     0.0
1                 NaN     NaN
2                 NaN     NaN
3                 NaN     NaN
4                 NaN     NaN
...               ...     ...
9995              NaN     NaN
9996              NaN     NaN
9997              NaN     NaN
9998              NaN     NaN
9999              NaN     NaN

[10000 rows x 14 columns]
```

```
#dealing with outliers
import seaborn as sns

sns.boxplot(data['Balance'])
```

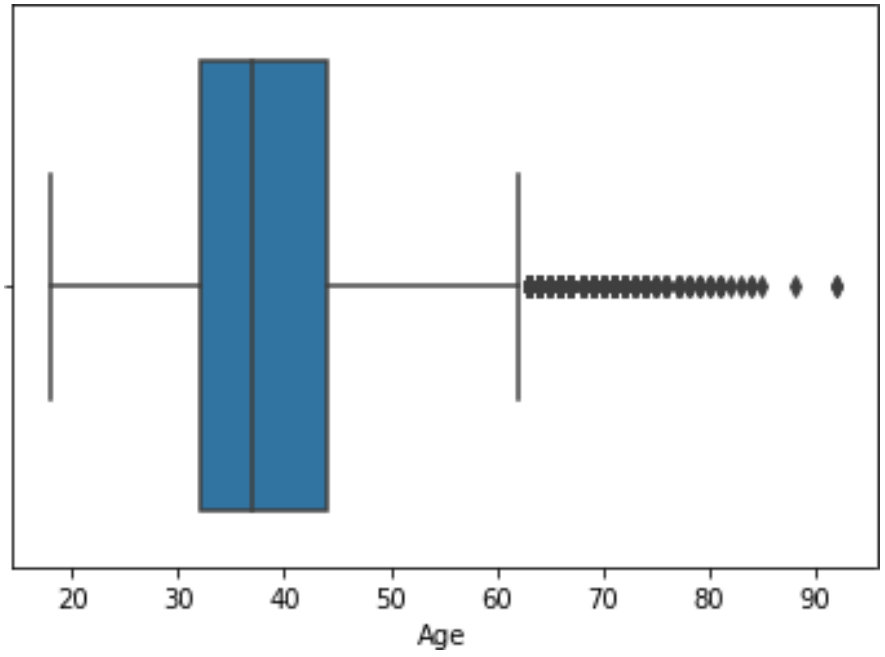
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12,
the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  FutureWarning

<matplotlib.axes._subplots.AxesSubplot at 0x7f085224e7d0>
```




```
sns.boxplot(data['Age'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12,
the only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  FutureWarning
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0852291d50>
```



```
#upper extreme=q3=1.5*IQR
#lower extreme=q1-1.5*IQR
#IQR=q3-q1
qnt=data.quantile(q=[0.25,0.75])
```

```
qnt
```

```
      RowNumber   CustomerId  CreditScore   Age  Tenure    Balance

                                                                  \
0.25    2500.75  15628528.25        584.0  32.0     3.0       0.00
0.75    7500.25  15753233.75        718.0  44.0     7.0  127644.24

      NumOfProducts  HasCrCard  IsActiveMember  EstimatedSalary  Exited
0.25            1.0        0.0             0.0       51002.1100     0.0
0.75            2.0        1.0             1.0      149388.2475     0.0
```

```
IQR=qnt.loc[0.75]-qnt.loc[0.25]
```

```
upper_extreme=qnt.loc[0.75]+1.5*IQR
```

```
upper_extreme
```

```
RowNumber          1.499950e+04
CustomerId         1.594029e+07
CreditScore        9.190000e+02
Age                6.200000e+01
Tenure             1.300000e+01
Balance            3.191106e+05
NumOfProducts      3.500000e+00
HasCrCard          2.500000e+00
IsActiveMember     2.500000e+00
EstimatedSalary    2.969675e+05
Exited             0.000000e+00
dtype: float64
```

```
lower_extreme=qnt.loc[0.25]-1.5*IQR
```

```
lower_extreme
```

```
RowNumber         -4.998500e+03
CustomerId         1.544147e+07
CreditScore        3.830000e+02
Age                1.400000e+01
Tenure            -3.000000e+00
Balance           -1.914664e+05
NumOfProducts     -5.000000e-01
HasCrCard         -1.500000e+00
IsActiveMember    -1.500000e+00
EstimatedSalary   -9.657710e+04
Exited             0.000000e+00
dtype: float64
```

```
data[data['Age']>54.25]
```

```
      RowNumber  CustomerId   Surname  CreditScore Geography  Gender  Age  \
16           17    15737452     Romeo          653   Germany    Male   58
42           43    15687946   Osborne          556    France  Female   61
44           45    15684171   Bianchi          660     Spain  Female   61
58           59    15623944     T'ien          511     Spain  Female   66
63           64    15751208   Pirozzi          684     Spain    Male   56
...         ...         ...       ...          ...       ...     ...  ...
9909       9910    15773338  Endrizzi          739    France    Male   58
9910       9911    15784042        L?          624    France    Male   55
9936       9937    15653037     Parks          609    France    Male   77
9939       9940    15808971    Lajoie          693     Spain  Female   57
9979       9980    15692664    Diribe          677    France  Female   58

      Tenure    Balance  NumOfProducts  HasCrCard  IsActiveMember  \
16         1  132602.88              1          1               0
42         2  117419.35              1          1               1
44         5  155931.11              1          1               1
58         4       0.00              1          1               0
63         8   78707.16              1          1               1
...      ...        ...            ...        ...             ...
9909       2  101579.28              1          1               1
9910       7  118793.60              1          1               1
9936       1       0.00              1          0               1
9939       9       0.00              2          1               1
9979       1   90022.85              1          0               1

      EstimatedSalary  Exited
16            5097.67       1
42           94153.83       0
44          158338.39       0
58            1643.11       1
63           99398.36       0
...               ...     ...
9909         72168.53       0
9910         95022.02       1
9936         18708.76       0
9939        135502.77       0
9979          2988.28       0

[882 rows x 14 columns]
```

```
data[data['Balance']>99000]
```

```
      RowNumber  CustomerId   Surname  CreditScore Geography  Gender  Age  \
2            3    15619304      Onio          502    France  Female   42
4            5    15737888  Mitchell          850     Spain  Female   43
5            6    15574012       Chu          645     Spain    Male   44
```

```
7            8    15656148    Obinna          376   Germany  Female   29
8            9    15792365        He          501    France    Male   44
...        ...         ...       ...          ...       ...     ...  ...
9985      9986    15586914    Nepean          659    France    Male   36
9986      9987    15581736  Bartlett          673   Germany    Male   47
9987      9988    15588839   Mancini          606     Spain    Male   30
9993      9994    15569266    Rahman          644    France    Male   28
9999     10000    15628319    Walker          792    France  Female   28

      Tenure    Balance  NumOfProducts  HasCrCard  IsActiveMember  \
2          8  159660.80              3          1               0
4          2  125510.82              1          1               1
5          8  113755.78              2          1               0
7          4  115046.74              4          1               0
8          4  142051.07              2          0               1
...      ...        ...            ...        ...             ...
9985       6  123841.49              2          1               0
9986       1  183579.54              2          0               1
9987       8  180307.73              2          1               1
9993       7  155060.41              1          1               0
9999       4  130142.79              1          1               0

      EstimatedSalary  Exited
2           113931.57       1
4            79084.10       0
5           149756.71       1
7           119346.88       1
8            74940.50       0
...               ...     ...
9985         96833.00       0
9986         34047.54       0
9987          1914.41       0
9993         29179.52       0
9999         38190.78       0

[4871 rows x 14 columns]
```

```
data[data['Age']<24]
```

```
      RowNumber  CustomerId          Surname  CreditScore Geography  Gender  \
57           58    15647091         Endrizzi          725   Germany    Male
69           70    15755648           Pisano          675    France  Female
75           76    15780961         Cavenagh          735    France  Female
86           87    15762418             Gant          750     Spain    Male
98           99    15604348           Allard          710     Spain    Male
...         ...         ...              ...          ...       ...     ...
9796       9797    15794236           Thorpe          642   Germany    Male
9888       9889    15697606          Sturdee          637    France  Female
9932       9933    15813451  Fleetwood-Smith          677     Spain    Male
9940       9941    15791972      Bergamaschi          748    France  Female
```

```
9943       9944    15659495               Fu          784     Spain    Male

      Age  Tenure    Balance  NumOfProducts  HasCrCard  IsActiveMember  \
57     19       0   75888.20              1          0               0
69     21       8   98373.26              1          1               0
75     21       1  178718.19              2          1               0
86     22       3  121681.82              1          1               0
98     22       8       0.00              2          0               0
...   ...     ...        ...            ...        ...             ...
9796   22      10  111812.52              2          1               1
9888   21      10  125712.20              1          0               0
9932   18       8  134796.87              2          1               1
9940   20       7       0.00              2          0               0
9943   23       2       0.00              1          1               1

      EstimatedSalary  Exited
57           45613.75       0
69           18203.00       0
75           22388.00       0
86          128643.35       1
98           99645.04       0
...               ...     ...
9796        183045.46       0
9888        175072.47       0
9932        114858.90       0
9940         10792.42       0
9943          6847.73       0

[325 rows x 14 columns]


data[data['Balance']<75000]

      RowNumber  CustomerId      Surname  CreditScore Geography  Gender  Age
\
0             1    15634602     Hargrave          619    France  Female   42
3             4    15701354         Boni          699    France  Female   39
6             7    15592531     Bartlett          822    France    Male   50
11           12    15737173      Andrews          497     Spain    Male   24
12           13    15632264          Kay          476    France  Female   34
...         ...         ...          ...          ...       ...     ...  ...
9992       9993    15657105  Chukwualuka          726     Spain    Male   36
9994       9995    15719294         Wood          800    France  Female   29
9995       9996    15606229     Obijiaku          771    France    Male   39
9996       9997    15569892    Johnstone          516    France    Male   35
9997       9998    15584532          Liu          709    France  Female   36

      Tenure   Balance  NumOfProducts  HasCrCard  IsActiveMember  \
0          2      0.00              1          1               1
3          1      0.00              2          0               0
6          7      0.00              2          1               1
```

```
11          3      0.00              2          1               0
12         10      0.00              2          1               0
...       ...       ...            ...        ...             ...
9992        2      0.00              1          1               0
9994        2      0.00              2          0               0
9995        5      0.00              2          1               0
9996       10  57369.61              1          1               1
9997        7      0.00              1          0               1

      EstimatedSalary  Exited
0           101348.88       1
3            93826.63       0
6            10062.80       0
11           76390.01       0
12           26260.98       0
...               ...     ...
9992        195192.40       0
9994        167773.55       0
9995         96270.64       0
9996        101699.77       0
9997         42085.58       1

[4041 rows x 14 columns]
```

```
#replacing outlier with mean
data['Balance']=np.where(data['Balance']>75000,data['Balance'].mean(),data['B
alance'])
```

```
data[data['Balance']>75000]
```

```
      RowNumber  CustomerId      Surname  CreditScore Geography  Gender  Age
\
1             2    15647311         Hill          608     Spain  Female   41
2             3    15619304         Onio          502    France  Female   42
4             5    15737888     Mitchell          850     Spain  Female   43
5             6    15574012          Chu          645     Spain    Male   44
7             8    15656148       Obinna          376   Germany  Female   29
...         ...         ...          ...          ...       ...     ...  ...
9987       9988    15588839      Mancini          606     Spain    Male   30
9991       9992    15769959  Ajuluchukwu          597    France  Female   53
9993       9994    15569266       Rahman          644    France    Male   28
9998       9999    15682355    Sabbatini          772   Germany    Male   42
9999      10000    15628319       Walker          792    France  Female   28

      Tenure       Balance  NumOfProducts  HasCrCard  IsActiveMember  \
1          1  76485.889288              1          0               1
2          8  76485.889288              3          1               0
4          2  76485.889288              1          1               1
5          8  76485.889288              2          1               0
7          4  76485.889288              4          1               0
```

```
...       ...           ...             ...         ...              ...
9987        8  76485.889288               2           1                1
9991        4  76485.889288               1           1                0
9993        7  76485.889288               1           1                0
9998        3  76485.889288               2           1                0
9999        4  76485.889288               1           1                0

      EstimatedSalary  Exited
1           112542.58       0
2           113931.57       1
4            79084.10       0
5           149756.71       1
7           119346.88       1
...               ...     ...
9987          1914.41       0
9991         69384.71       1
9993         29179.52       0
9998         92888.52       1
9999         38190.78       0

[5959 rows x 14 columns]
```

```
#Encoding
data.head()
```

```
   RowNumber       CustomerId   Surname  CreditScore Geography  Gender  Age  \
0          1         15634602  Hargrave          619    France  Female
                           42
1          2         15647311      Hill          608     Spain  Female
                           41
2          3         15619304      Onio          502    France  Female
                           42
3          4         15701354      Boni          699    France  Female
                           39
4          5         15737888  Mitchell          850     Spain  Female
                           43

   Tenure        Balance   NumOfProducts  HasCrCard  IsActiveMember\
0        2      0.000000              1          1  1
1        1  76485.889288              1          0               1
2        8  76485.889288              3          1               0
3        1      0.000000              2          0               0
4        2  76485.889288              1          1               1

   EstimatedSalary
          Exited0
          101348.88       1
1         112542.58       0
2         113931.57       1
3          93826.63       0
4          79084.10       0
```

```
#method-dummies
pd.get_dummies(data,columns=['Geography'])
```

```
      RowNumber  CustomerId    Surname  CreditScore  Gender  Age  Tenure  \
0             1    15634602   Hargrave          619  Female   42       2
1             2    15647311       Hill          608  Female   41       1
2             3    15619304       Onio          502  Female   42       8
3             4    15701354       Boni          699  Female   39       1
4             5    15737888   Mitchell          850  Female   43       2
...         ...         ...        ...          ...     ...  ...     ...
9995       9996    15606229   Obijiaku          771    Male   39       5
9996       9997    15569892  Johnstone          516    Male   35      10
9997       9998    15584532        Liu          709  Female   36       7
9998       9999    15682355  Sabbatini          772    Male   42       3
9999      10000    15628319     Walker          792  Female   28       4

           Balance  NumOfProducts  HasCrCard  IsActiveMember  EstimatedSalary
\
0         0.000000              1          1               1        101348.88
1     76485.889288              1          0               1        112542.58
2     76485.889288              3          1               0        113931.57
3         0.000000              2          0               0         93826.63
4     76485.889288              1          1               1         79084.10
...            ...            ...        ...             ...              ...
9995      0.000000              2          1               0         96270.64
9996  57369.610000              1          1               1        101699.77
9997      0.000000              1          0               1         42085.58
9998  76485.889288              2          1               0         92888.52
9999  76485.889288              1          1               0         38190.78

      Exited  Geography_France  Geography_Germany  Geography_Spain
0          1                 1                  0                0
1          0                 0                  0                1
2          1                 1                  0                0
3          0                 1                  0                0
4          0                 0                  0                1
...      ...               ...                ...              ...
9995       0                 1                  0                0
9996       0                 1                  0                0
9997       1                 1                  0                0
9998       1                 0                  1                0
9999       0                 1                  0                0

[10000 rows x 16 columns]
```

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data['Geography']=le.fit_transform(data['Geography'])
data.head()
```

```
   RowNumber      CustomerId   Surname  CreditScore  Geography  Gender  Age  \
```

```
0           1    15634602  Hargrave          619           0  Female   42
1           2    15647311      Hill          608           2  Female   41
2           3    15619304      Onio          502           0  Female   42
3           4    15701354      Boni          699           0  Female   39
4           5    15737888  Mitchell          850           2  Female   43

   Tenure       Balance  NumOfProducts  HasCrCard  IsActiveMember  \
0       2      0.000000              1          1               1
1       1  76485.889288              1          0               1
2       8  76485.889288              3          1               0
3       1      0.000000              2          0               0
4       2  76485.889288              1          1               1

   EstimatedSalary
          Exited0
          101348.88  1
1         112542.58       0
2         113931.57       1
3          93826.63       0
4          79084.10       0
```

```
#separating dependent and independent columns
x=data.iloc[:,1:6]
y=data.iloc[:,6]
```

```
data.iloc[1:3]
```

```
   RowNumber  CustomerId Surname  CreditScore  Geography  Gender  Age  Tenure
\
1          2    15647311    Hill          608          2  Female   41       1
2          3    15619304    Onio          502          0  Female   42       8

        Balance  NumOfProducts  HasCrCard  IsActiveMember  EstimatedSalary  \
1  76485.889288              1          0               1        112542.58
2  76485.889288              3          1               0        113931.57

   Exited
1       0
2       1
```

```
#splitting the dtata into train and test
from sklearn.model_selection import train_test_split

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

```
x_train.shape
```

```
(8000, 5)
```

```
x_test.shape
```

```
(2000, 5)
```