IMPORTING LIBRARIES

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import scipy as sp
import plotly.express as px
```

```
data=pd.read_csv('/content/HRDataset_v14.csv')
data
```

	Employee_Name	EmpID	MarriedID	...	LastPerfo
0	Adinolfi, Wilson K	10026	0	...	
1	Ait Sidi, Karthikeyan	10084	1	...	
2	Akinkuolie, Sarah	10196	1	...	
3	Alagbe,Trina	10088	1	...	
4	Anderson, Carol	10069	0	...	
...	...	...	...	...	
306	Woodson, Jason	10135	0	...	
307	Ybarra, Catherine	10301	0	...	
308	Zamora, Jennifer	10010	0	...	
309	Zhou, Julia	10043	0	...	
310	Zima, Colleen	10271	0	...	

311 rows × 36 columns

UNDERSTANDING DATA

1. BASIC CHECKING

DATA INFORMATION

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 311 entries, 0 to 310
Data columns (total 36 columns):
 #   Column                      Non-Null Count  Dty
---  ------                      --------------  ---
 0   Employee_Name               311 non-null    obj
 1   EmpID                       311 non-null    int
 2   MarriedID                   311 non-null    int
 3   MaritalStatusID             311 non-null    int
 4   GenderID                    311 non-null    int
 5   EmpStatusID                 311 non-null    int
 6   DeptID                      311 non-null    int
 7   PerfScoreID                 311 non-null    int
 8   FromDiversityJobFairID      311 non-null    int
 9   Salary                      311 non-null    int
 10  Termd                       311 non-null    int
 11  PositionID                  311 non-null    int
 12  Position                    311 non-null    obj
 13  State                       311 non-null    obj
 14  Zip                         311 non-null    int
 15  DOB                         311 non-null    obj
 16  Sex                         311 non-null    obj
 17  MaritalDesc                 311 non-null    obj
 18  CitizenDesc                 311 non-null    obj
 19  HispanicLatino              311 non-null    obj
 20  RaceDesc                    311 non-null    obj
 21  DateofHire                  311 non-null    obj
 22  DateofTermination           104 non-null    obj
 23  TermReason                  311 non-null    obj
 24  EmploymentStatus            311 non-null    obj
 25  Department                  311 non-null    obj
 26  ManagerName                 311 non-null    obj
 27  ManagerID                   303 non-null    flo
 28  RecruitmentSource           311 non-null    obj
 29  PerformanceScore            311 non-null    obj
 30  EngagementSurvey            311 non-null    flo
 31  EmpSatisfaction             311 non-null    int
 32  SpecialProjectsCount        311 non-null    int
 33  LastPerformanceReview_Date  311 non-null    obj
 34  DaysLateLast30              311 non-null    int
```

```
 35  Absences                    311 non-null    int
dtypes: float64(2), int64(16), object(18)
memory usage: 87.6+ KB
```

DATA SHAPE

```
data.shape
```

```
(311, 36)
```

READ FIRST 5 ROWS

```
data.head(5)
```

	Employee_Name	EmpID	MarriedID	...	LastPerforma
0	Adinolfi, Wilson K	10026	0	...	
1	Ait Sidi, Karthikeyan	10084	1	...	
2	Akinkuolie, Sarah	10196	1	...	
3	Alagbe,Trina	10088	1	...	
4	Anderson, Carol	10069	0	...	

5 rows × 36 columns

```
Warning: Total number of columns (36) exceeds max co
```

READ LAST 5 ROWS

```
data.tail(10)
```

	Employee_Name	EmpID	MarriedID	...	LastPerfor
301	Wilber, Barry	10048	1	...	
302	Wilkes, Annie	10204	0	...	
303	Williams, Jacquelyn	10264	0	...	
304	Winthrop, Jordan	10033	0	...	
305	Wolk, Hang T	10174	0	...	
306	Woodson, Jason	10135	0	...	
307	Ybarra, Catherine	10301	0	...	
308	Zamora, Jennifer	10010	0	...	
309	Zhou, Julia	10043	0	...	
310	Zima, Colleen	10271	0	...	

10 rows × 36 columns

DATA TYPES

```
data.dtypes
```

```
Employee_Name                object
EmpID                         int64
MarriedID                     int64
MaritalStatusID               int64
GenderID                      int64
EmpStatusID                   int64
DeptID                        int64
PerfScoreID                   int64
FromDiversityJobFairID        int64
Salary                        int64
Termd                         int64
PositionID                    int64
Position                     object
State                        object
Zip                           int64
DOB                          object
Sex                          object
MaritalDesc                  object
CitizenDesc                  object
HispanicLatino               object
RaceDesc                     object
DateofHire                   object
DateofTermination            object
TermReason                   object
EmploymentStatus             object
Department                   object
ManagerName                  object
ManagerID                   float64
RecruitmentSource            object
PerformanceScore             object
EngagementSurvey            float64
EmpSatisfaction               int64
SpecialProjectsCount          int64
LastPerformanceReview_Date   object
DaysLateLast30                int64
Absences                      int64
dtype: object
```

SEE NUMERIC COLUMNS

```
pd.set_option('display.max_columns',None)
data.head()
```

	Employee_Name	EmpID	MarriedID	MaritalStatusID
0	Adinolfi, Wilson K	10026	0	0
1	Ait Sidi, Karthikeyan	10084	1	1
2	Akinkuolie, Sarah	10196	1	1
3	Alagbe,Trina	10088	1	1

SEE NULL VALUES

```
data.isnull().sum().sort_values(ascending=False)
```

```
DateofTermination             207
ManagerID                       8
EmpID                           0
RaceDesc                        0
DateofHire                      0
TermReason                      0
EmploymentStatus                0
Department                      0
ManagerName                     0
Employee_Name                   0
RecruitmentSource               0
PerformanceScore                0
EngagementSurvey                0
EmpSatisfaction                 0
SpecialProjectsCount            0
LastPerformanceReview_Date      0
DaysLateLast30                  0
HispanicLatino                  0
CitizenDesc                     0
MaritalDesc                     0
FromDiversityJobFairID          0
MarriedID                       0
MaritalStatusID                 0
GenderID                        0
EmpStatusID                     0
DeptID                          0
PerfScoreID                     0
Salary                          0
Sex                             0
Termd                           0
PositionID                      0
Position                        0
State                           0
Zip                             0
DOB                             0
Absences                        0
dtype: int64
```

SEE DUPLICATED DATA

```
data[data.duplicated(keep=False)]
```

Employee_Name	EmpID	MarriedID	MaritalStatusID	

SEE DUPLICATE VALUES COUNT IN NUMBER

```
data.duplicated().sum()
```

```
0
```

SEE ALL COLUMN LABELS OF DATASET

```
data.columns
```

```
Index(['Employee_Name', 'EmpID', 'MarriedID',
'MaritalStatusID', 'GenderID',
       'EmpStatusID', 'DeptID', 'PerfScoreID',
'FromDiversityJobFairID',
       'Salary', 'Termd', 'PositionID', 'Position',
'State', 'Zip', 'DOB',
       'Sex', 'MaritalDesc', 'CitizenDesc',
'HispanicLatino', 'RaceDesc',
       'DateofHire', 'DateofTermination',
'TermReason', 'EmploymentStatus',
       'Department', 'ManagerName', 'ManagerID',
'RecruitmentSource',
       'PerformanceScore', 'EngagementSurvey',
'EmpSatisfaction',
       'SpecialProjectsCount',
'LastPerformanceReview_Date', 'DaysLateLast30',
       'Absences'],
      dtype='object')
```

DESCRIPTIVE STATISTICS SUMMARY

```
data.describe()
```

	EmpID	MarriedID	MaritalStatusID	Gend
count	311.000000	311.000000	311.000000	311.00
mean	10156.000000	0.398714	0.810289	0.43
std	89.922189	0.490423	0.943239	0.49
min	10001.000000	0.000000	0.000000	0.00
25%	10078.500000	0.000000	0.000000	0.00
50%	10156.000000	0.000000	1.000000	0.00
75%	10233.500000	1.000000	1.000000	1.00

GET THE COUNT OF UNIQUE VALUES IN DESCENDING ORDER

```
print(data['MaritalStatusID'].value_counts())
```

```
0    137
1    124
2     30
3     12
4      8
Name: MaritalStatusID, dtype: int64
```

```
print(data['GenderID'].value_counts())
```

```
0    176
1    135
Name: GenderID, dtype: int64
```

```
print(data['DeptID'].value_counts())
```

```
5    208
3     50
6     32
4     10
1     10
2      1
Name: DeptID, dtype: int64
```