

```
In [14]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

In [21]: datapd.read_csv("D:\VWO3-Pooja\Sharma\All G-drive files\Day15\IBM Attrition Data.csv")

In [31]: data.head()
```

	Age	Attrition	Department	DistanceFromHome	Education	EducationField	EnvironmentSatisfaction	JobSatisfaction	MaritalStatus	MonthlyIncome	NumCompaniesWorked	WorkLifeBalance	YearsAtCompany
0	41	Yes	Sales	1	2	Life Sciences	2	4	Single	5993	8	1	6
1	49	No	Research & Development	8	1	Life Sciences	3	2	Married	5130	1	3	10
2	37	No	Research & Development	2	2	Other	4	3	Single	2090	6	3	0
3	33	No	Research & Development	3	4	Life Sciences	4	3	Married	2909	1	3	8
4	27	No	Research & Development	2	1	Medical	1	2	Married	3468	9	3	2

In [41]: data.shape

Out[41]: (1470, 13)

In [51]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 13 columns):
 0   Age          Non-null Count  Dtype
----  -
 1   Attrition      1470 non-null  object
 2   Department     1470 non-null  object
 3   DistanceFromHome 1470 non-null  int64
 4   Education       1470 non-null  int64
 5   EducationField  1470 non-null  object
 6   EnvironmentSatisfaction 1470 non-null  int64
 7   JobSatisfaction 1470 non-null  int64
 8   MaritalStatus   1470 non-null  object
 9   MonthlyIncome   1470 non-null  int64
10   NumCompaniesWorked 1470 non-null  int64
11   WorkLifeBalance 1470 non-null  int64
12   YearsAtCompany  1470 non-null  int64
dtypes: int64(9), object(4)
memory usage: 145.4+ KB
```

In [61]: data.isnull().sum() #sum of all null values must be zero

Out[61]:

Age	0
Attrition	0
Department	0
DistanceFromHome	0
Education	0
EducationField	0
EnvironmentSatisfaction	0
JobSatisfaction	0
MaritalStatus	0
MonthlyIncome	0
NumCompaniesWorked	0
WorkLifeBalance	0
YearsAtCompany	0
dtype:	int64

In [121]: data["Age"].plot.hist(figsize=(10,6),xlabel="Age",ylabel="Count",title="Age distribution",alpha=0.5,bins=50)

• Most Employees are between the age 30 to 40

In [81]: data["Attrition"].value_counts()

Out[81]:

In [91]: data.groupby(["Age", "Attrition"]).size().unstack().head()

Out[91]:

	Attrition	No	Yes
Age			
18	40	40	
19	30	60	
20	50	60	
21	70	60	
22	110	50	

In [181]: data.groupby(["Age", "Attrition"]).size().unstack().plot(figsize=(13,6),alpha=0.7)

In [111]: x=data.groupby(["Age", "Attrition"]).size().unstack().reset_index()

Out[111]:

In [121]: x.head()

Attrition	Age	No	Yes	% Attrition
0	18	40	40	50.000000
1	19	30	60	33.333333
2	20	50	60	45.454545
3	21	70	60	33.846154
4	22	110	50	68.750000

In [131]: print("Most % of employees attritioned belong to age :",x["Age"][x["Attrition"].idxmax()], "% Percentage of leaving :",x["Attrition"].max())

Out[131]:

In [141]: x[x["Attrition"]==1].size().unstack().head()

Out[141]:

In [151]: x[x["Attrition"]==1].size().unstack().head()

Out[151]:

In [161]: data[data["Attrition"]==1].size().unstack().head()

Out[161]:

In [171]: data[data["Attrition"]==1].size().unstack().head()

Out[171]:

In [181]: data[data["Attrition"]==1].size().unstack().head()

Out[181]:

In [191]: data[data["Attrition"]==1].size().unstack().head()

Out[191]:

In [201]: data[data["Attrition"]==1].size().unstack().head()

Out[201]:

In [211]: data[data["Attrition"]==1].size().unstack().head()

Out[211]:

In [221]: data[data["Attrition"]==1].size().unstack().head()

Out[221]:

In [231]: data[data["Attrition"]==1].size().unstack().head()

Out[231]:

In [241]: data[data["Attrition"]==1].size().unstack().head()

Out[241]:

In [251]: data[data["Attrition"]==1].size().unstack().head()

Out[251]:

In [261]: data[data["Attrition"]==1].size().unstack().head()

Out[261]:

In [271]: data[data["Attrition"]==1].size().unstack().head()

Out[271]:

In [281]: data[data["Attrition"]==1].size().unstack().head()

Out[281]:

In [291]: data[data["Attrition"]==1].size().unstack().head()

Out[291]:

In [301]: data[data["Attrition"]==1].size().unstack().head()

Out[301]:

In [311]: data[data["Attrition"]==1].size().unstack().head()

Out[311]:

In [321]: data[data["Attrition"]==1].size().unstack().head()

Out[321]:

In [331]: data[data["Attrition"]==1].size().unstack().head()

Out[331]:

In [341]: data[data["Attrition"]==1].size().unstack().head()

Out[341]:

In [351]: data[data["Attrition"]==1].size().unstack().head()

Out[351]:

In [361]: data[data["Attrition"]==1].size().unstack().head()

Out[361]:

In [371]: data[data["Attrition"]==1].size().unstack().head()

Out[371]:

In [381]: data[data["Attrition"]==1].size().unstack().head()

Out[381]:

In [391]: data[data["Attrition"]==1].size().unstack().head()

Out[391]:

In [401]: data[data["Attrition"]==1].size().unstack().head()

Out[401]:

In [411]: data[data["Attrition"]==1].size().unstack().head()

Out[411]:

In [421]: data[data["Attrition"]==1].size().unstack().head()

Out[421]:

In [431]: data[data["Attrition"]==1].size().unstack().head()

Out[431]:

In [441]: data[data["Attrition"]==1].size().unstack().head()

Out[441]:

In [451]: data[data["Attrition"]==1].size().unstack().head()

Out[451]:

In [461]: data[data["Attrition"]==1].size().unstack().head()

Out[461]:

In [471]: data[data["Attrition"]==1].size().unstack().head()

Out[471]:

In [481]: data[data["Attrition"]==1].size().unstack().head()

Out[481]:

In [491]: data[data["Attrition"]==1].size().unstack().head()

Out[491]:

In [501]: data[data["Attrition"]==1].size().unstack().head()

Out[501]:

In [511]: data[data["Attrition"]==1].size().unstack().head()

Out[511]:

In [521]: data[data["Attrition"]==1].size().unstack().head()

Out[521]:

In [531]: data[data["Attrition"]==1].size().unstack().head()

Out[531]:

In [541]: data[data["Attrition"]==1].size().unstack().head()

Out[541]:

In [551]: data[data["Attrition"]==1].size().unstack().head()

Out[551]:

In [561]: data[data["Attrition"]==1].size().unstack().head()

Out[561]:

In [571]: data[data["Attrition"]==1].size().unstack().head()

Out[571]:

In [581]: data[data["Attrition"]==1].size().unstack().head()

Out[581]:

In [591]: data[data["Attrition"]==1].size().unstack().head()

Out[591]:

In [601]: data[data["Attrition"]==1].size().unstack().head()

Out[601]:

In [611]: data[data["Attrition"]==1].size().unstack().head()

Out[611]:

In [621]: data[data["Attrition"]==1].size().unstack().head()

Out[621]:

In [631]: data[data["Attrition"]==1].size().unstack().head()

Out[631]:

In [641]: data[data["Attrition"]==1].size().unstack().head()

Out[641]:

In [651]: data[data["Attrition"]==1].size().unstack().head()

Out[651]:

In [661]: data[data["Attrition"]==1].size().unstack().head()

Out[661]:

In [671]: data[data["Attrition"]==1].size().unstack().head()

Out[671]:

In [681]: data[data["Attrition"]==1].size().unstack().head()

Out[681]:

In [691]: data[data["Attrition"]==1].size().unstack().head()

Out[691]:

In [701]: data[data["Attrition"]==1].size().unstack().head()

Out[701]:

In [711]: data[data["Attrition"]==1].size().unstack().head()

Out[711]:

In [721]: data[data["Attrition"]==1].size().unstack().head()

Out[721]:

In [731]: data[data["Attrition"]==1].size().unstack().head()

Out[731]:

In [741]: data[data["Attrition"]==1].size().unstack().head()

Out[741]:

In [751]: data[data["Attrition"]==1].size().unstack().head()

Out[751]:

In [761]: data[data["Attrition"]==1].size().unstack().head()

Out[761]:

In [771]: data[data["Attrition"]==1].size().unstack().head()

Out[771]:

In [781]: data[data["Attrition"]==1].size().unstack().head()

Out[781]:

In [791]: data[data["Attrition"]==1].size().unstack().head()

Out[791]:

In [801]: data[data["Attrition"]==1].size().unstack().head()

Out[801]:

In [811]: data[data["Attrition"]==1].size().unstack().head()

Out[811]:

In [821]: data[data["Attrition"]==1].size().unstack().head()

Out[821]:

In [831]: data[data["Attrition"]==1].size().unstack().head()

Out[831]:

In [841]: data[data["Attrition"]==1].size().unstack().head()

Out[841]:

In [851]: data[data["Attrition"]==1].size().unstack().head()

Out[851]:

In [861]: data[data["Attrition"]==1].size().unstack().head()

Out[861]:

In [871]: data[data["Attrition"]==1].size().unstack().head()

Out[871]:

In [881]: data[data["Attrition"]==1].size().unstack().head()

Out[881]:

In [891]: data[data["Attrition"]==1].size().unstack().head()

Out[891]:

In [901]: data[data["Attrition"]==1].size().unstack().head()

Out[901]:

In [911]: data[data["Attrition"]==1].size().unstack().head()

Out[911]:

In [921]: data[data["Attrition"]==1].size().unstack().head()

Out[921]:

In [931]: data[data["Attrition"]==1].size().unstack().head()

Out[931]:

In [941]: data[data["Attrition"]==1].size().unstack().head()

Out[941]:

In [951]: data[data["Attrition"]==1].size().unstack().head()

Out[951]:

In [961]: data[data["Attrition"]==1].size().unstack().head()

Out[961]:

In [971]: data[data["Attrition"]==1].size().unstack().head()

Out[971]:

In [981]: data[data["Attrition"]==1].size().unstack().head()

Out[981]:

In [991]: data[data["Attrition"]==1].size().unstack().head()

Out[991]: