Automated Exploratory Data Analysis

In [19]:

```python
import pandas as pd
import numpy as np
import seaborn as sns
from pandas_profiling import ProfileReport
%matplotlib inline
```

Reading the CSV File

In [2]:

```python
df=pd.read_csv(r'C:\Users\Prasath Kamalakannan\Desktop\Prasath\Ineuron_pythonProject\FSDS_B
```

In [3]:

```python
df.head()
```

Out[3]:

	case_id	continent	education_of_employee	has_job_experience	requires_job_training	no_of_
0	EZYV01	Asia	High School	N	N	
1	EZYV02	Asia	Master's	Y	N	
2	EZYV03	Asia	Bachelor's	N	Y	
3	EZYV04	Asia	Bachelor's	N	N	
4	EZYV05	Africa	Master's	Y	N	

Pandas Profiling

In [8]:

```python
profile = ProfileReport(df, title="Pandas Profiling Report")
```

In [9]:

```
profile.to_widgets()
```

Generate report structure: 1/1 [00:02<00:00,

100%   2.62s/it]

Overview	Variables	Interactions	Correlations	Missing value	Sample

Overview	Alerts (3)	Reproduction

Number of variables	12	Categorical	6
		Boolean	3
Number of observations	25480	Numeric	3
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	2.3 MiB		
Average record size in memory	96.0 B		

Report generated by YData (https://ydata.ai/?

utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report).

In [11]:

```
profile.to_notebook_iframe()
```

Download   config.json (data:text/plain;charset=utf-
configuration   8,%7B%22title%22%3A%20%22Pandas%20Profiling%20Report%22

Variables

case_id
Categorical

HIGH

CARDINALITY
UNIFORM
UNIQUE

Distinct	25480
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Memory size	199.2 KiB

Length

Max length	9
Median length	9
Mean length	8.564481947
Min length	6

Characters and Unicode

In [13]:

```
profile.to_file("pandas_profiling.html")
```

Export report to file: 1/1 [00:00<00:00,

100%   47.94it/s]

Sweetviz

In [17]:

```
import sweetviz as sv

my_report = sv.analyze(df)
my_report.show_html()
```

Done! Use 'show' commands to display/save.   [100%] 00:00 -> (00:00 left)

Report SWEETVIZ_REPORT.html was generated! NOTEBOOK/COLAB USERS: the web bro
wser MAY not pop up, regardless, the report IS saved in your notebook/colab
files.

Dtale

In [20]:

```python
import dtale

import dtale.app as dtale_app
dtale_app.USE_COLAB = True
dtale.show(df)
```

	_training	no_of_employees	yr_of_estab	region_of_employment	prevaili
20	N	880	2008	West	
21	N	1706	2013	Midwest	1:
22	N	2878	1968	West	
23	N	1517	1884	Midwest	
24	N	241	1981	Midwest	
25	N	1100	1997	Midwest	
26	N	756	1995	South	
27	Y	2889	2005	South	1
28	N	18	2004	Northeast	
29	N	758	2009	Northeast	1
30	N	1408	1998	Midwest	
31	N	995	1880	South	1
32	N	2415	1963	Midwest	
33	N	2395	2000	South	

Autoviz

In [24]:

```python
from autoviz.AutoViz_Class import AutoViz_Class
AV = AutoViz_Class()
dft = AV.AutoViz(r'C:\Users\Prasath Kamalakannan\Desktop\Prasath\Ineuron_pythonProject\FSDS
%matplotlib inline
```

```
Shape of your Data Set loaded: (25480, 12)
##############################################################################
#############
######################### C L A S S I F Y I N G   V A R I A B L E S   #######
#############
##############################################################################
#############
Classifying variables in data set...
Data cleaning improvement suggestions. Complete them before proceeding to
ML modeling.
```

	Nuniques	dtype	Nulls	Nullpercent	NuniquePercent	Value counts Min	Data cleaning improvement suggestions
case_id	25480	object	0	0.000000	100.000000	1	combine rare categories, possible ID column: drop