MINI PROJECT

EXPLORATORY DATA ANALYSIS [EDA]

DECEMBER 19, 2022

CHELLAMALLA NITHIN

Branch: CSE

College: Sri Indu College of Engineering & Technology, Hyderabad

Submitted to:

Ameen Manna Sir

Problem Statement:

Take any dataset of your choice and apply Exploratory Data Analysis and make sure, you get 5 different conclusions/facts/analysis.

SOLUTION:

I took dataset from www.Kaggle.com, the data is about Programming Language that Includes S.no, Year, Chief Developer, etc.

I have to Apply Exploratory Data Analysis on this Data Set Then I performed so many things On that data, They are:

- Import pandas to my data
- Dataset shape
- Dataset size
- Dataset info
- Dataset sum
- Dataset Unique
- Import Seaborn to my data
- Count plot to Chief Developer
- Count Plot to Predecessors
- Count value to Chief Developer
- Count Value to Predecessors
- Moreover, I done group by for Chief Developer

I Performed above Furnished Analysis on my Dataset Successfully Without Errors……

➔ Import pandas to my data:

```
import pandas as pd
df=pd.read_csv('/content/programminglanguage.csv')
df
```

	Year	Name	Chief developer, company	Predecessor(s)
0	1804	Jacquard machine	Joseph Marie Jacquard	none (unique language)
1	1879	Begriffsschrift	Gottlob Frege	none (unique language)
2	1944	ENIAC coding system	John von Neumann, John Mauchly, J. Presper Eck...	none (unique language)
3	1948	Plankalkül	Konrad Zuse	none (unique language)
4	1948	ARC/Birkbeck Assembler	Kathleen Booth	ENIAC Short Code
...	...	...	...	...
389	2018	Fortran 2018	ISO/IEC JTC1/SC22/WG5 N2150:2018	Fortran 2008
390	2019	Bosque	Mark Marron, Microsoft	JavaScript, TypeScript, ML
391	2020	C++20	C++ ISO/IEC 14882:2020	C++, Standard C, C
392	2021	Microsoft Power Fx	Vijay Mital, Robin Abraham, Shon Katzenberger,...	Excel formulas
393	2022	Carbon	Google	C++, Rust

394 rows × 4 columns

➔ Dataset Shape:

```
df.shape
```

```
(394, 4)
```

➔ Dataset Size:

```
df.size
```

```
1576
```

➔ Dataset info:

```
df.info
```

```
<bound method DataFrame.info of     Year                 Name  \
0    1804        Jacquard machine
1    1879         Begriffsschrift
2    1944     ENIAC coding system
3    1948             Plankalkül
4    1948  ARC/Birkbeck Assembler
..    ...                     ...
389  2018           Fortran 2018
390  2019                 Bosque
391  2020                  C++20
392  2021      Microsoft Power Fx
393  2022                 Carbon

                          Chief developer, company  \
0                                  Joseph Marie Jacquard
1                                        Gottlob Frege
2    John von Neumann, John Mauchly, J. Presper Eck...
3                                          Konrad Zuse
4                                      Kathleen Booth
..                                                ...
389                ISO/IEC JTC1/SC22/WG5 N2150:2018
390                          Mark Marron, Microsoft
391                        C++ ISO/IEC 14882:2020
392  Vijay Mital, Robin Abraham, Shon Katzenberger,...
393                                          Google

              Predecessor(s)
0       none (unique language)
1       none (unique language)
2       none (unique language)
3       none (unique language)
4            ENIAC Short Code
..                        ...
389              Fortran 2008
390  JavaScript, TypeScript, ML
391          C++, Standard C, C
392              Excel formulas
393                     C++, Rust

[394 rows x 4 columns]>
```

➔ Dataset Sum:

```
[ ] df.isnull().sum()
```

```
Year                          0
Name                          0
Chief developer, company     11
Predecessor(s)               37
dtype: int64
```

➔ Dataset Unique:

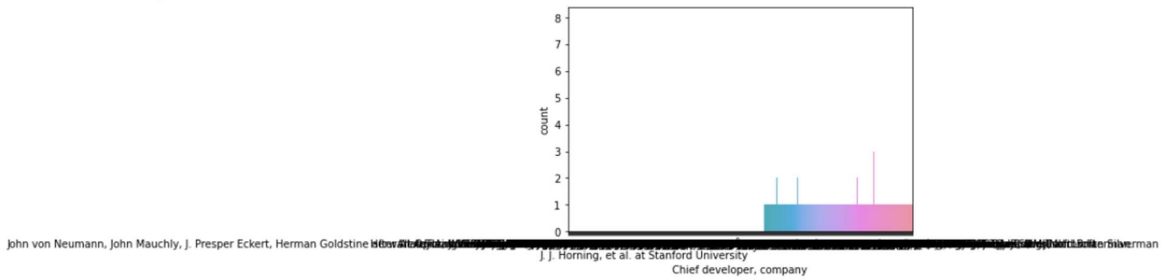
```
df.nunique()
```

```
Year                          78
Name                         393
Chief developer, company     335
Predecessor(s)               254
dtype: int64
```

➔ Count plot to Chief Developer:

```
import seaborn as sns
sns.countplot(x='Chief developer, company',data=df)
```

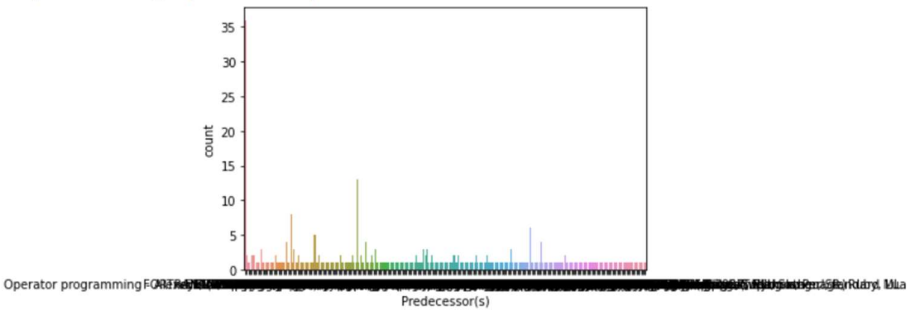
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f37ba1c62e0>
```



➔ Count plot to Predecessors:

```
import seaborn as sns
sns.countplot(x='Predecessor(s)',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f37b95f51f0>
```



➔ Count value to Chief Developer:

```
[ ] df['Chief developer, company'].value_counts()
```

```
IBM                                                                  8
Microsoft                                                            8
Apple Computer                                                       5
Google                                                               3
Allen Newell, Cliff Shaw, Herbert A. Simon                           3
                                                                    ..
Brian Kernighan                                                      1
Allan Ballard, Paul Whaley at the University of British Columbia     1
Arvind and Gostelow, University of California, Irvine                1
James S. Miller, Benjamin M. Brosgol et al. at Intermetrics          1
Vijay Mital, Robin Abraham, Shon Katzenberger, Darryl Rubin, Microsoft  1
Name: Chief developer, company, Length: 335, dtype: int64
```

➔ Count value to Predecessors:

```
df['Predecessor(s)'].value_counts()
```

```
none (unique language)    36
BASIC                     13
ALGOL 60                   8
C++, Standard C, C         6
LISP                       5
                          ..
InterPress                 1
Euclid                     1
Speakeasy-3                1
Pascal, C, ALGOL 68        1
C++, Rust                  1
Name: Predecessor(s), Length: 254, dtype: int64
```

➔ Group by for Chief Developer :

```
df.groupby('Chief developer, company').size()
```

```
Chief developer, company
ABB                                        1
ACM/GAMM                                   1
ANSI/ISO Standard C++                      1
ANSI/MIL-STD-1815A unchanged               1
ARA and Ada Europe (ISO/IEC 8652:2012)     1
                                          ..
Yukihiro Matsumoto                         1
Yves Caseau                                1
Zoltan Somogyi at University of Melbourne  1
designed by Intermetrics for NASA          1
many people at Apple Computer              1
Length: 335, dtype: int64
```