Assignment 1

Visualisation

“Programming Languages by Users, Popularity and Years”

- I have selected this data set from Kaggle. I have then loaded the csv file via pandas and displayed them.

```
************************* Dataframe *************************

            Date  Abap   Ada  C/C++    C#  Cobol  Dart  Delphi/Pascal   Go
0       July 2004  0.34  0.36  10.08  4.71   0.43   0.0           2.82  0.0
1     August 2004  0.36  0.36   9.81  4.99   0.46   0.0           2.67  0.0
2  September 2004  0.41  0.41   9.63  5.06   0.51   0.0           2.65  0.0
3    October 2004  0.40  0.38   9.50  5.31   0.53   0.0           2.77  0.0

   Groovy  Haskell   Java  JavaScript  Julia  Kotlin   Lua  Matlab  \
0    0.03     0.22  30.37        8.65    0.0     0.0  0.16    2.11
1    0.07     0.20  29.99        8.78    0.0     0.0  0.15    2.05
2    0.08     0.21  29.71        8.70    0.0     0.0  0.19    2.11
3    0.09     0.20  29.12        8.47    0.0     0.0  0.22    2.24

   Objective-C  Perl    PHP  Python     R  Ruby  Rust  Scala  Swift  \
0         0.19  7.38  18.75    2.53  0.39  0.33  0.08   0.03    0.0
1         0.18  7.11  19.26    2.64  0.41  0.40  0.09   0.03    0.0
2         0.19  7.03  19.49    2.72  0.40  0.41  0.10   0.03    0.0
3         0.20  7.17  19.34    2.92  0.42  0.46  0.11   0.04    0.0

   TypeScript   VBA  Visual Basic
0         0.0  1.44          8.56
1         0.0  1.46          8.57
2         0.0  1.55          8.41
3         0.0  1.61          8.49
```

- I have then read and look through the printed data frame and called the .info() function and get some information about the data frame.

```
****************** Information about Dataframe ******************

Data set shape =  (211, 29)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 211 entries, 0 to 210
Data columns (total 29 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Date           211 non-null    object
 1   Abap           211 non-null    float64
 2   Ada            211 non-null    float64
 3   C/C++          211 non-null    float64
 4   C#             211 non-null    float64
 5   Cobol          211 non-null    float64
 6   Dart           211 non-null    float64
 7   Delphi/Pascal  211 non-null    float64
 8   Go             211 non-null    float64
 9   Groovy         211 non-null    float64
 10  Haskell        211 non-null    float64
 11  Java           211 non-null    float64
 12  JavaScript     211 non-null    float64
 13  Julia          211 non-null    float64
 14  Kotlin         211 non-null    float64
 15  Lua            211 non-null    float64
 16  Matlab         211 non-null    float64
 17  Objective-C    211 non-null    float64
 18  Perl           211 non-null    float64
 19  PHP            211 non-null    float64
 20  Python         211 non-null    float64
 21  R              211 non-null    float64
 22  Ruby           211 non-null    float64
 23  Rust           211 non-null    float64
 24  Scala          211 non-null    float64
 25  Swift          211 non-null    float64
 26  TypeScript     211 non-null    float64
 27  VBA            211 non-null    float64
 28  Visual Basic   211 non-null    float64
dtypes: float64(28), object(1)
memory usage: 47.9+ KB
None
```

- I then had a look at the data frame for NaN by the isnull() function and could not find any null value.

```
******************** Total NaN in Dataframe ********************

0

******************** NaN values by columns *********************

Date            0
Abap            0
Ada             0
C/C++           0
C#              0
Cobol           0
Dart            0
Delphi/Pascal   0
Go              0
Groovy          0
Haskell         0
Java            0
JavaScript      0
Julia           0
Kotlin          0
Lua             0
Matlab          0
Objective-C     0
Perl            0
PHP             0
Python          0
R               0
Ruby            0
Rust            0
Scala           0
Swift           0
TypeScript      0
VBA             0
Visual Basic    0
dtype: int64
```

- Then I removed some languages and only took the most common ones used and dropped their respective columns and get a smaller data frame. I also converted the values of Date column which were string object to Date-time object to be able to plot the data against the actual period of years.

	Date	C/C++	JavaScript	PHP	Python	Ruby	Visual Basic
0	2004-07-01	10.08	8.65	18.75	2.53	0.33	8.56
1	2004-08-01	9.81	8.78	19.26	2.64	0.40	8.57
2	2004-09-01	9.63	8.70	19.49	2.72	0.41	8.41

- Then I started plotting the graph of Programming languages with respect to the years, users and their popularity among users.

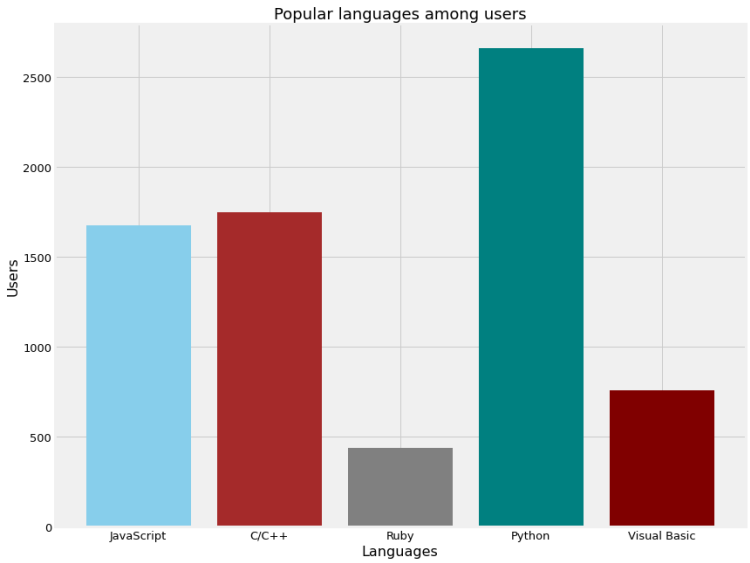

Figure: Bar Plot

- This Bar Chart shows the five most Common used programming languages with respect to their users over the span of 18 years from 2004 to 2022. We can easily see clearly that Python is the overall most popular or the most used among users over this span of time. I used this plot because it's a simple way of show casing the rank of Language in terms of usage.

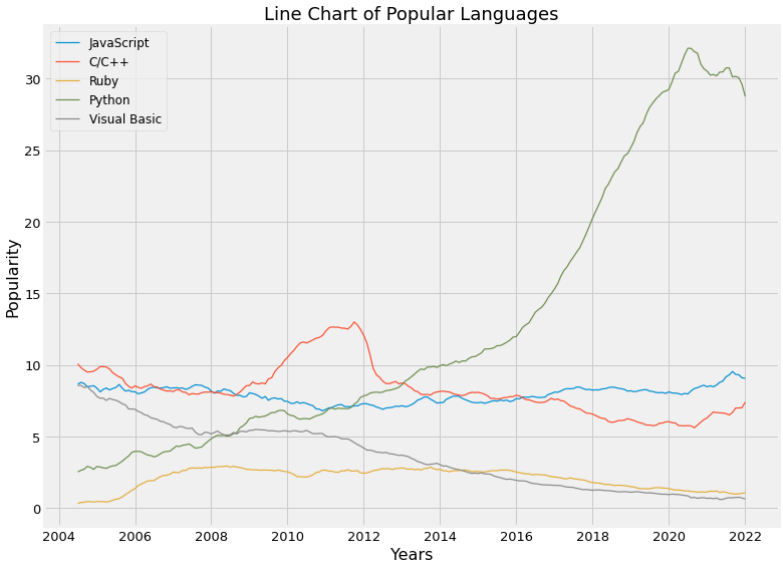

Figure: Line Chart

- This Line Chart shows some detailed variations across the years' time how these languages performed with respect to popularity in percentage. Here we can see that JavaScript almost remained constant throughout decades yet Visual basic went down and is less looked at in the new times whereas C and C++ had a significant rise from 2009 to 2012 and the remained constant. Out of all Python gets the high rise from 2013 and still emerging with its application and usage in Data science and Artificial Intelligence fields.

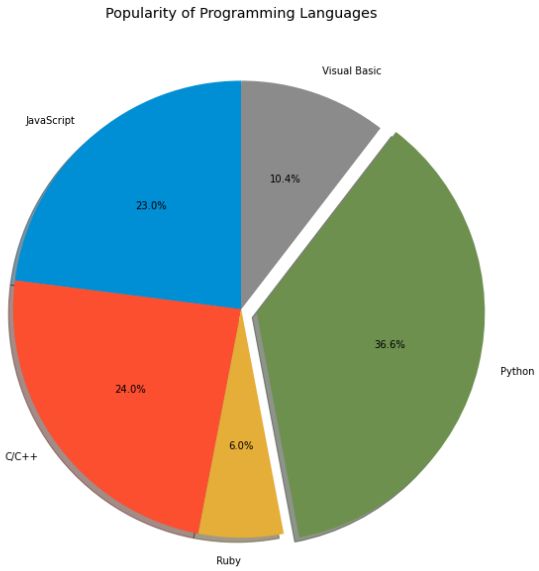

Figure: Pie Chart

- Pie Plot shows the visual portions of how much a language is used by Users, So Python being the most popular overall have the largest audience and hence the percentage of people using it, after comes C++/C and then JavaScript whereas Visual Basic and Ruby are the least used among the programmers and developers overall. I chose this chard because it's a clear visualization with we can see just by looking how much portion is popular amongst the users.

Data Source: https://www.kaggle.com/datasets/muhammadkhalid/most-popular-programming-languages-since-2004