

List Comprehension

List comprehension offers a shorter syntax when you want to create a new list based on the values of an existing list.

Example:

Based on a list of fruits, you want a new list, containing only the fruits with the letter "a" in the name.

Without list comprehension you will have to write a for statement with a conditional test inside:

```
fruits = ["ap ple ", " ban ana ", " che rry ", " kiw i", " man go"]
newlist = []
for x in fruits:
    if " a" in x:
        newlis t.a ppe nd(x)
print( new list)
```

With list comprehension you can do all that with only one line of code:

```
fruits = ["ap ple ", " ban ana ", " che rry ", " kiw i", " man go"]
newlist = [x for x in fruits if " a" in x]
print( new list)
```

Imputation

In statistics, imputation is the process of replacing missing data with substituted values.

When substituting for a data point, it is known as "unit imputation";

when substituting for a component of a data point, it is known as "item imputation".

[Pandas Imputation Article](#)

Aggregate Functions

sum()	Sums each value of an object
count()	Returns total Count
median()	Returns mathematical median
quantile([0.25, 0.75])	Quantiles of an object
min()	Lowest value in an object
max()	Highest Value in an Object
mean()	Returns mathematical mean
var()	Returns mathematical variance
std()	Returns standard deviation
df.groupby(by="col")	Groups data by value of specified column (Similar to SQL)
pd.merge(adf, bdf, how='left', on'col')	Merges to Datasheets into one based on a common column

Aggregate Functions are a way of summarizing or reshaping data

Shape of a Dataframe

Return a tuple representing the dimensionality of the DataFrame.

```
>>> df = pd.Dat aFr ame ({' col1': [1, 2], 'col2': [3, 4]})
>>> df.shape
(2, 2)
```



By CodingJinxx

Published 9th December, 2021.
Last updated 10th December, 2021.
Page 1 of 2.

Sponsored by [Readable.com](#)
Measure your website readability!
<https://readable.com>

Mean

Return the mean of the values over the requested axis.

```
DataFrame.mean(axis=None, skipna=None, level=None, numeric_only=None)
```

Median

Sorts all values in dataframe and returns the middle value

```
DataFrame.median(axis=None, skipna=None, level=None, numeric_only=None)
```

Creating a Dataframe from Scratch

```
# Import pandas library
import pandas as pd

# initialize list of lists
data = [['tom', 10], ['nick', 15], ['juli', 14]]

# Create the pandas DataFrame
df = DataFrame(data, columns = ['Name', 'Age'])
```

From Scratch means creating the Data by hand

Categorical Variable

Is data that is limited to set or range of values

They are best visualised using bar plots or balloon plot

[Example Article](#)

Quartiles vs Quantiles

Quartiles 25th percentiles of Data

Where as Quantiles can be custom percentiles

Correlation

Correlation describes the relationship between data.

Example:

If the square footage in an apartment increases, the price of the apartment increases aswell

Scatterplot

A Scatterplot plots data on an x-y grid

Histogram

A histogram plots data on a axis with the count being represented in height



By CodingJinxx

Published 9th December, 2021.

Last updated 10th December, 2021.

Page 2 of 2.

Sponsored by [Readable.com](#)

Measure your website readability!

<https://readable.com>