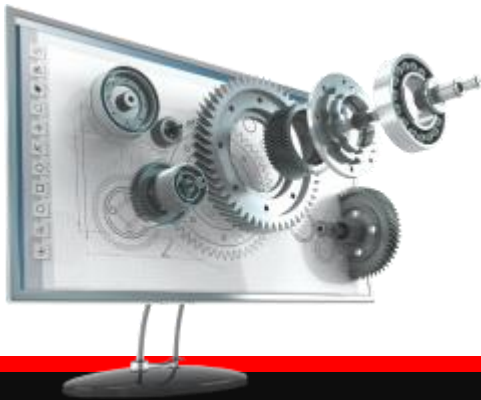




Python for Beginners

Archer Infotech , PUNE





Python - Pandas

Uploading Dataset to Google Colab



- Using Code

```
from google.colab import files  
files.upload()
```



Reading CSV file in Pandas



- Using Code

```
import pandas as pd
```

```
pd = pd.read_csv ( filepath )
```



Pandas Basic Info Functions



- **DataFrame.head($n=5$)** :This function returns the first n rows for the object based on position.
- **DataFrame.tail($n=5$)** :This function returns last n rows from the object based on position.
- **DataFrame.info()** : This method prints information about a DataFrame including the index dtype and columns, non-null values and memory usage.
- **DataFrame.describe()** : Descriptive statistics include those that summarize the central tendency, dispersion and shape of a dataset's distribution, excluding NaN values.



Pandas Analyze Data Frames



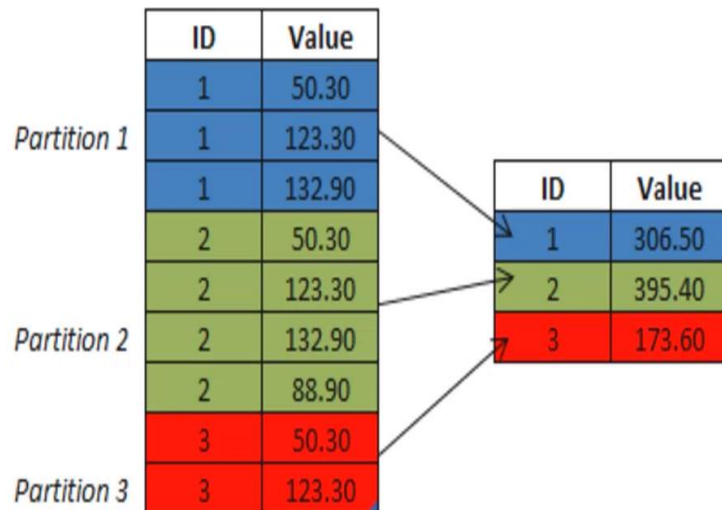
- **.head()**
- **.shape**
- **.index**
- **.columns**
- **.dtypes**
- **.unique()**
- **.nunique()**
- **.count**
- **.value_counts**
- **.info**



Pandas Groupby Function



Groupby allows you to group together rows based off of a column and perform an aggregate function on them.

A diagram illustrating the pandas groupby function. On the left, a large table is divided into three horizontal sections labeled 'Partition 1', 'Partition 2', and 'Partition 3'. Each section contains a table with 'ID' and 'Value' columns. Partition 1 has three rows with ID 1 and values 50.30, 123.30, and 132.90. Partition 2 has four rows with ID 2 and values 50.30, 123.30, 132.90, and 88.90. Partition 3 has two rows with ID 3 and values 50.30 and 123.30. Arrows point from the 'ID' column of each partition to a smaller table on the right, which shows the aggregated results for each ID: ID 1 with a sum of 306.50, ID 2 with a sum of 395.40, and ID 3 with a sum of 173.60.

	ID	Value
Partition 1	1	50.30
	1	123.30
	1	132.90
Partition 2	2	50.30
	2	123.30
	2	132.90
	2	88.90
Partition 3	3	50.30
	3	123.30

ID	Value
1	306.50
2	395.40
3	173.60

count() – Number of non-null observations

sum() – Sum of values

mean() – Mean of values

median() – Arithmetic median of values

min() – Minimum

max() – Maximum

mode() – Mode

std() – Standard deviation

var() – Variance



Pandas value_counts & nunique Function



The `value_counts()` method returns a **Series containing the counts of unique values**. This means, for any column in a dataframe, this method returns the count of unique entries in that column.

`nunique()` function **return number of unique elements in the object**. It returns a scalar value which is the count of all the unique values in the Index. By default the NaN values are not included in the count



Change Null Values by mean of Column



- Check Null Values Using `isnull()` Function / `isnull().sum()`
- **Fill Null values using `fillna` function**

To Change null values by mean :

```
df [ column_name ] .fillna ( df[column_name].mean() , inplace = True )
```



Pandas.apply() Function



- Pandas.apply allow the users to pass a function and apply it on every single value of the Pandas series.

```
s.apply(func, convert_dtype=True, args=())
```

func: .apply takes a function and applies it to all values of pandas series.

convert_dtype: Convert dtype as per the function's operation.

args=(): Additional arguments to pass to function instead of series.

Return Type: Pandas Series after applied function/operation.



Pandas isin function



- **DataFrame.isin(*values*)**

Whether each element in the DataFrame is contained in values



Pandas Sorting Dataset by Column



- `pandas.DataFrame.sort_values`
`df.sort_values`
(
 `by = ['Name', 'Rank'],`
 `axis=0,`
 `ascending=[False, True],`
 `inplace=False,`
 `kind='quicksort',`
 `na_position='first',`
 `ignore_index=True,`
 `key=None`
)



Ex 1: Cars Dataset



1. Find all Null Values in the Dataset . If there is any null value in any column then fill it with the mean of column
2. What are different types of Make are there in Dataset and what is the count (occurrence) of each make in data
3. Show all Records where origin is Asia or Europe
4. Remove All Records where weight is above 4000
5. Increase all the values of 'MPG_City' column by 3



Ex 2: Covid Dataset



1. Show the number of Confirmed , Deaths and Recovered cases in each region [use groupby function]
2. Remove All the records where Confirmed cases is less than < 10 [use \sim]
3. In which Region , maximum number of Confirmed cases were recorded?
4. In Which Region , minimum no of Deaths cases were Recorded ?
5. How many Confirmed , Deaths , Recovered cases were reported from India ?
6. Sort the entire Data wrt No of Confirmed Cases in Ascending Order
7. Sort the entire data wrt No of Recovered cases in Descending Order



Ex 3: Weather Dataset



1. Find all the unique “Wind Speed” values in the Data
2. Find the number of times when the “Weather is exactly Clear”
3. Find the number of Times when the “Wind Speed was Exactly 4 km/h”
4. Find out all Null Values in data
5. Rename the column name “Weather” of the DataFrame to “Weather Condition “
6. What is mean “Visibility”
7. What is Std Deviation of “Pressure” in data
8. What is the Variance of “Relative Humidity” in this Data
9. Find all instances when “Snow” was Recorded
10. Find All Instances when “Wind Speed is above 24” and “Visibility is 25”



Ex 4: Census Dataset



1. Hide Indexes of DataFrame [`df.style.hide_index()`]
2. How can we set the Caption / heading of DataFrame [`df.style.set_caption`]
3. Show records related with Districts – New Delhi , Lucknow , Jaipur
4. Calculate State wise :
Total Number of Population , Total No of Population with Different Religions
5. How Many Male Workers were there in Maharashtra State ?
6. How to Set a Column as a Index of dataframe [`df.set_index`]
- 7 . Add Suffix / Prefix to Column Names [`df.add_suffix`]



Ex 5: Udemy Dataset



1. What are different Subjects for which Udemy is Offering Courses ?
2. Which Subject has the Maximum No of Courses?
3. Show All the Courses which are Free of Cost ?
4. Show All the Courses which are Paid ?
5. Which are Least Selling Courses?
6. Which are Top Selling Courses?
7. Show All Courses of Graphic Design Where the Price is below 100
8. List all Courses related to Python
9. What are Courses Published in Year 2015
10. What are Max No of Subscribers for Each level of Courses ?





THANK YOU !!!

Amol Patil - 9822291613

