Pandas Part 2

```
#read data with pandas
#perform more interesting data science operations with pandas
#extract, one , multiple
#add new column
#adjust column
#delete column

#Index
#set_index
#perform extraction through index
#reset_index

#sort by, two, multiple indexes
#extracting rows with iloc

#subsetting

#lambda functions


import pandas as pd
import numpy as np


attendance = pd.read_csv("church_attendance.csv")
attendance
```

	First Name	Last Name	Worker_Status	House Holde Number	Service attended
0	Daniel	Johnson	worker	5	15
1	Blessing	Madu	worker	3	20
2	Makanjuola	Adekunmi	worker	2	14
3	John	James	Non worker	6	21
4	Jeremiah	Akindele	worker	7	25

```
#clean column names
columns = attendance.columns.to_list()
columns [0] = 'First_Name'
columns [1] = 'Last_Name'
columns [3] = 'House_Hold_Number'
columns [4] = 'Service_attended'
attendance.columns = columns
attendance
```

	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
0	Daniel	Johnson	worker	5	15
1	Blessing	Madu	worker	3	20
2	Makanjuola	Adekunmi	worker	2	14
3	John	James	Non worker	6	21
4	Jeremiah	Akindele	worker	7	25

```
#Extracting columns
attendance['First_Name']
```

```
0         Daniel
1      Blessing
2    Makanjuola
3          John
4       Jeremiah
Name: First_Name, dtype: object
```

```
attendance[['First_Name', 'House_Hold_Number']]
```

	First_Name	House_Hold_Number
0	Daniel	5
1	Blessing	3
2	Makanjuola	2
3	John	6
4	Jeremiah	7

```
gender = ["Male", "Female", "Male", "Male", "Male"]
```

```
attendance
```

	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
0	Daniel	Johnson	worker	5	15
1	Blessing	Madu	worker	3	20
2	Makanjuola	Adekunmi	worker	2	14
3	John	James	Non worker	6	21
4	Jeremiah	Akindele	worker	7	25

```
attendance['Gender'] = gender
attendance
```

	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended	Gender
0	Daniel	Johnson	worker	5	15	Male
1	Blessing	Madu	worker	3	20	Female
2	Makanjuola	Adekunmi	worker	2	14	Male
3	John	James	Non worker	6	21	Male
4	Jeremiah	Akindele	worker	7	25	Male

```
#Edit a column
#Add 4 to service attended
attendance['Service_attended'] = attendance['Service_attended'] + 4
attendance
```

	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended	Gender
0	Daniel	Johnson	worker	5	19	Male
1	Blessing	Madu	worker	3	24	Female
2	Makanjuola	Adekunmi	worker	2	18	Male
3	John	James	Non worker	6	25	Male
4	Jeremiah	Akindele	worker	7	29	Male

```
#delete a column

attendance.pop("Gender")
```

```
0      Male
1    Female
2      Male
3      Male
4      Male
Name: Gender, dtype: object
```

```
attendance
```

	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
0	Daniel	Johnson	worker	5	19

```
#create a new dataframe of worker status and age
worker_status = attendance['Worker_Status']
age = [18,20,22,24,26]
new_df = pd.DataFrame({"worker_status": worker_status,  "Age": age})
new_df
```

	worker_status	Age
0	worker	18
1	worker	20
2	worker	22
3	Non worker	24
4	worker	26

Indexing

```
#index
#set_index
#perform extraction through index
#reset_index
```

```
attendance
```

	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
0	Daniel	Johnson	worker	5	19
1	Blessing	Madu	worker	3	24
2	Makanjuola	Adekunmi	worker	2	18
3	John	James	Non worker	6	25
4	Jeremiah	Akindele	worker	7	29

```
attendance["ID"] = [1001, 1024, 1035, 1078, 1986]
attendance
```

	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended	ID
0	Daniel	Johnson	worker	5	19	1001
1	Blessing	Madu	worker	3	24	1024
2	Makanjuola	Adekunmi	worker	2	18	1035

```
attendance.set_index("ID")
```

ID	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
1001	Daniel	Johnson	worker	5	19
1024	Blessing	Madu	worker	3	24
1035	Makanjuola	Adekunmi	worker	2	18
1078	John	James	Non worker	6	25
1986	Jeremiah	Akindele	worker	7	29

```
attendance
```

	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended	ID
0	Daniel	Johnson	worker	5	19	1001
1	Blessing	Madu	worker	3	24	1024
2	Makanjuola	Adekunmi	worker	2	18	1035
3	John	James	Non worker	6	25	1078
4	Jeremiah	Akindele	worker	7	29	1986

```
#to make the change persistent

attendance.set_index("ID", inplace = True)

attendance
```

ID	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended

attendance

ID	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
1001	Daniel	Johnson	worker	5	19
1024	Blessing	Madu	worker	3	24
1035	Makanjuola	Adekunmi	worker	2	18
1078	John	James	Non worker	6	25
1986	Jeremiah	Akindele	worker	7	29

```
#Extracting a row if the index are not sequential
attendance.loc[1001]
```

```
First_Name            Daniel
Last_Name            Johnson
Worker_Status         worker
House_Hold_Number          5
Service_attended          19
Name: 1001, dtype: object
```

```
attendance.loc[1001]['First_Name']
```

```
'Daniel'
```

```
attendance.loc[[1001, 1024]]
```

ID	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
1001	Daniel	Johnson	worker	5	19
1024	Blessing	Madu	worker	3	24

```
#Going back to the original dataframe after making the change persistent using inplace

attendance.reset_index()
```

	ID	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
0	1001	Daniel	Johnson	worker	5	19
1	1024	Blessing	Madu	worker	3	24
2	1035	Makanjuola	Adekunmi	worker	2	18
3	1078	John	James	Non worker	6	25

```
#Extracting a row if the index are sequential

attendance.iloc[0:2]
```

ID	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
1001	Daniel	Johnson	worker	5	19
1024	Blessing	Madu	worker	3	24

```
attendance.reset_index()
```

	ID	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
0	1001	Daniel	Johnson	worker	5	19
1	1024	Blessing	Madu	worker	3	24
2	1035	Makanjuola	Adekunmi	worker	2	18
3	1078	John	James	Non worker	6	25
4	1986	Jeremiah	Akindele	worker	7	29

```
attendance = attendance.reset_index()


#sort by, two, multiple indexes.
#extracting rows with iloc

#sort by 1 column
attendance.sort_values("First_Name")
```

	ID	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
0	1001	Daniel	Johnson	worker	5	19

```
attendance.sort_values(["Worker_Status", "House_Hold_Number"])
```

	ID	First_Name	Last_Name	Worker_Status	House_Hold_Number	Service_attended
3	1078	John	James	Non worker	6	25
2	1035	Makanjuola	Adekunmi	worker	2	18
1	1024	Blessing	Madu	worker	3	24
0	1001	Daniel	Johnson	worker	5	19
4	1986	Jeremiah	Akindele	worker	7	29

```
attendance[['ID',
 'First_Name',
 'Worker_Status',
 'Last_Name',
 'Service_attended',
 'House_Hold_Number']]
```

	ID	First_Name	Worker_Status	Last_Name	Service_attended	House_Hold_Number
0	1001	Daniel	worker	Johnson	19	5
1	1024	Blessing	worker	Madu	24	3
2	1035	Makanjuola	worker	Adekunmi	18	2
3	1078	John	Non worker	James	25	6
4	1986	Jeremiah	worker	Akindele	29	7

```
attendance.columns.to_list()
```

```
['ID',
 'First_Name',
 'Last_Name',
 'Worker_Status',
 'House_Hold_Number',
 'Service_attended']
```

Colab paid products - Cancel contracts here

✓ 0s completed at 3:54 AM

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.