

MBBS//BDS Extraction Manipulation and Closing Rank finder Script

Description about Script:

The MBBS/BDS Extraction, Manipulation, and Closing Rank Finder Script is a Python-based tool designed to streamline the process of extracting and analyzing data from PDF files. Initially, the script extracts relevant information, such as student ranks and college details, and converts this data into a structured CSV file. This allows for easy handling and further processing of the data. By leveraging Python libraries, the script ensures that the extraction process is accurate, capturing all necessary details without omitting any critical information.

Once the data is extracted, the script performs various data manipulation tasks to correct any inconsistencies. For example, if certain values are mixed up or misplaced during extraction, the script reorganizes and sorts the data neatly. This ensures that the final dataset is clean and ready for analysis. After organizing the data, the script calculates and identifies the maximum and minimum ranks for each category, providing valuable insights into the closing ranks for MBBS and BDS admissions. This automated approach saves time and ensures precision in handling large volumes of data.

Scripts with line by line explanation:

1.Script for extraction and Manipulation

```
# pip install pdfplumber openpyxl: Installs the necessary Python packages: pdfplumber for extracting text from PDFs, and openpyxl for working with Excel files.
# import re: Imports the re module for working with regular expressions.
# import pdfplumber: Imports the pdfplumber module to extract text from PDF files.
# import pandas as pd: Imports the pandas library, commonly used for data manipulation and analysis.
# from collections import namedtuple: Imports namedtuple from the collections module to create a simple class-like structure for storing data.
# from openpyxl import Workbook: Imports Workbook from the openpyxl module to create and manipulate Excel files.
```

```
import re
import pdfplumber
import pandas as pd
from collections import namedtuple
from openpyxl import Workbook

# Adjusted named tuple to combine 'code', 'college', and everything else into 'code_college'
Line = namedtuple('Line', 'sr_no air neet_roll_no cet_form_no first_name middle_name last_name gender category quota code_college')

# Adjust the regular expressions based on your PDF structure
line_re =
re.compile(r'\d+\s+\d+\s+\d+\s+\d+\s+\d+\s+.\s+\s+[MF]\s+.\s+\s+\s+.\s+\s+\s+\s+.\s+')

file = 'example.pdf'
lines = []

with pdfplumber.open(file) as pdf:
```

MBBS//BDS Extraction Manipulation and Closing Rank finder Script

```
pages = pdf.pages
for page in pages:
    text = page.extract_text()
    for line in text.split('\n'):
        if line_re.match(line):
            items = line.split()
            # Print the items and their length to inspect
            print(f"Length: {len(items)}, Items: {items}")

            # Identify the gender index to correctly capture the name
            gender_index = 5
            for i in range(4, len(items)):
                if items[i] in ('M', 'F'):
                    gender_index = i
                    break

            sr_no = items[0]
            air = items[1]
            neet_roll_no = items[2]
            cet_form_no = items[3]

            name_parts = items[4:gender_index]
            if len(name_parts) == 2:
                first_name = name_parts[0]
                last_name = name_parts[1]
                middle_name = ''
            elif len(name_parts) == 3:
                first_name = name_parts[0]
                middle_name = name_parts[1]
                last_name = name_parts[2]
            else:
                first_name = name_parts[0]
                last_name = ' '.join(name_parts[1:])
                middle_name = ''

            gender = items[gender_index]
            category = items[gender_index + 1]
            quota = items[gender_index + 2]
            code_college = ' '.join(items[gender_index + 3:])

            lines.append(Line(sr_no, air, neet_roll_no, cet_form_no,
                             first_name, middle_name, last_name, gender, category, quota, code_college))

# Create a pandas DataFrame
df = pd.DataFrame(lines)

# Save the DataFrame to an Excel file
df.to_csv('output.csv', index=False)
```

MBBS//BDS Extraction Manipulation and Closing Rank finder Script

Above script is designed to extract specific data from a PDF file, organize it into a structured format, and then save that data as a CSV file. Here's a breakdown of its use:

1. Data Extraction from PDF:

- The script uses `pdfplumber` to open and read the text from a PDF file named `'example.pdf'`.
- It processes each line of the PDF, applying a regular expression to match lines that contain the relevant data.

2. Data Parsing:

- The script splits each matched line into individual data elements (like serial number, rank, roll number, etc.).
- It identifies key elements like the student's name, gender, category, and college code by examining the position of these items in the line.

3. Data Structuring:

- The script uses a named tuple (`Line`) to organize these elements into a structured format, with fields for serial number, rank, roll number, name components, gender, category, quota, and college code.
- Each line of data from the PDF is converted into an instance of the `Line` named tuple and added to a list (`lines`).

4. Data Conversion to DataFrame:

- After processing all pages of the PDF, the list of `Line` named tuples is converted into a pandas DataFrame (`df`).
- This DataFrame provides a tabular structure that is easy to manipulate and analyze.

5. Saving Data as CSV:

- Finally, the DataFrame is saved as a CSV file (`output.csv`), which can be used for further analysis, reporting, or integration into other systems.

Use Case:

- This script is particularly useful in scenarios where you need to extract structured data from a large number of PDF documents, such as processing student rank lists, admission data, or any other tabular data stored in PDFs.
- It automates the extraction and organization of data, saving time and reducing the potential for manual errors.

MBBS//BDS Extraction Manipulation and Closing Rank finder Script

Input SS:



GOVERNMENT OF MAHARASHTRA
State Common Entrance Test Cell, Maharashtra, Mumbai
Admissions to Health Science Courses, NEET(UG)-2024-2025
PROVISIONAL SELECTION LIST ONLY FOR MBBS/BDS COURSES-CAP ROUND 1

- Note: 1. Last Date of joining the respective college: 05/09/2024 upto 5.30 pm(Including Holiday).
2. Last Date to fill the Status Retention Form at College: 05/09/2024 upto 5.30 pm(Including Holiday).
3. Admitting Institute will verify the original documents and ascertain eligibility of the candidates as per the NEET(UG)-2024 brochure.
4. Candidates should submit all the original documents and pay requisite fees in the selected college within the stipulated time; failing which this selection stands cancelled.
5. This seat allotment is conducted on the seat(s) that are approved/permitted by MCI/NMC/DCI and affiliated to MUHS,Nashik Only.

Printed On :31/08/2024 Pg: 1

Sr. No.	AIR	NEET Roll No.	CET Form. No.	Name	G	Cat.	←-----Current Selection Details-----→	Code College
Quota								
1	109	3108020911	245022961	POTDAR YASH DEEPAK	M	OBC	OPEN (EMD)	1103:GSMC MUMBAI
2	121	3103110206	245026959	VRAJ SACHIN KALE	M	OBC	OPEN	1103:GSMC MUMBAI
3	314	3109170301	245017579	PATIL SARTH PRASHANT	M		OPEN	1103:GSMC MUMBAI
4	543	3110130236	245002432	AASTHA VIRAL SHAH	F		OPEN (W)	1103:GSMC MUMBAI
5	657	3112040419	245015391	SAIRAM KARLE	M	NTC	OPEN (EMD)	1103:GSMC MUMBAI
6	689	3110010764	245028306	ARYA PARAG JADHAV	F	OBC	OPEN (W) (EMD)	1103:GSMC MUMBAI
7	714	3107050049	245049061	VEDANT SANJAY BAWISKAR	M	OBC	OPEN	1103:GSMC MUMBAI
8	720	3115010053	245027936	THANGE SOHAM TABAJI	M	OBC	OPEN (EMD)	1103:GSMC MUMBAI
9	721	3112270021	245049548	SHREYAS SONUNE	M		OPEN	1103:GSMC MUMBAI
10	782	3104050120	245040737	ADITYA RAJESH UMBARKAR	M	SC	OPEN (EMD)	1327:GMC CH. SAMBHAJINAGAR
11	993	3115220149	245023340	NIDHISH SHIVAJI JADHAV	M		OPEN	1110:BJMC PUNE
12	1053	3108020545	245033169	SATYAJIT SUDHIR JAGTAP	M	SEBC	OPEN (EMD)	1103:GSMC MUMBAI
13	1055	3108030503	245019667	THANEKAR ROHINI NAMDEV	F	SEBCHA	HOPEN (EMD)	1110:BJMC PUNE
14	1074	3905090341	245014576	RUSHIL WAGH	M	OBC	OPEN (EMD)	1103:GSMC MUMBAI
15	1089	3112010404	245027645	SUKHADA SURESHRAO JADHAV	F		OPEN (W)	1103:GSMC MUMBAI
16	1105	3110230136	245003371	NISHAD GIRISH PATIL	M		OPEN	1103:GSMC MUMBAI
17	1146	3112310282	245009182	PANDALWAD PAWAN	M	NTB	OPEN (EMD)	1103:GSMC MUMBAI
18	1173	2713020481	245010503	HAJWANE MOHAMMED IBRAHIM ALTAF	M		OPEN	1103:GSMC MUMBAI
19	1187	3111180221	245010378	RABAB SADIQ ALI	F		OPEN (W)	1221:GMC NAGPUR
20	1221	3112400211	245009201	PRAJAKTA HIRAMAN BHAGAT	F	SC	OPEN (W) (EMD)	1103:GSMC MUMBAI
21	1229	3109400269	245046226	WAGALGAVE ADITYA TANAJI	M	NTC	OPEN	1103:GSMC MUMBAI
22	1261	3115290364	245013924	ARIHAM NILESH KANHARIYA	M		OPEN	1110:BJMC PUNE
23	1315	3108050056	245044430	KARTIK JAYKUMAR SATPUTE	M	NTC	OPEN (EMD)	1103:GSMC MUMBAI
24	1324	3108100110	245006387	OM SUDHIR YESANE	M	HA	HOPEN	1103:GSMC MUMBAI
25	1367	3110280145	245047162	SANKARSHAN NANASAHEB MEMANE	M		OPEN	1103:GSMC MUMBAI
26	1394	3108020289	245002380	RAMTIRTHAKAR PRATIK GAJANAN	M	EWS	OPEN	1103:GSMC MUMBAI
27	1403	3114040017	245021317	NETRA KALPESH MODI	F		OPEN (W)	1103:GSMC MUMBAI
28	1433	3112030108	245036698	HARSHVARDHAN SURALKAR	M		OPEN	1103:GSMC MUMBAI
29	1465	3108060132	245007159	SAJAGANE SANKET SUNIL	M	NTC	OPEN (EMD)	1103:GSMC MUMBAI
30	1470	3105080373	245057924	PAWAR SHREYA VIJAY	F	OBC	OPEN (W) (EMD)	1103:GSMC MUMBAI
31	1507	3110250313	245003776	DIYA SHYAMAL LANGALIA	F		OPEN (W)	1103:GSMC MUMBAI
32	1513	3110230009	245003307	VEDANT MURAL	M		OPEN	1103:GSMC MUMBAI
33	1547	3110460002	245025803	KRISH DHIRAJ JAIN	M		OPEN	1103:GSMC MUMBAI
34	1567	3110210424	245036098	ZEESHAN SAMIR SHEIKH	M		OPEN	1103:GSMC MUMBAI
35	1610	3103040338	245048653	ANSHUL MANOHAR ANDE	M		OPEN	1103:GSMC MUMBAI
36	1659	3110060085	245046991	MAITREYI RAMESH DABHILKAR	F	SC	OPEN (W)	1103:GSMC MUMBAI
37	1672	3118070160	245029472	SUBBULAKSHMI SHANKARANARAYANAN	F		OPEN (W)	1103:GSMC MUMBAI

Legends : AIR: All India Rank Cat: Category HA: Hilly Area D1: Def1 D2: Def2 D3: Def3 VJA:VJ(A),NTB:NT(B)/NT1, NTC:NT(C)/NT2, NTD:NT(D)/NT3, I.Q.:Institute Quota, Min: Minority Candidate, ORP-A/ORP-B: Orphan (Institutional-IWC/IOG), ORP-C: Orphan (Outside Institution-NNN), G:Gender M:Male F:Female W:Women Quota EMD: EarMarking Donor, EMR: EarMarking Receiver

Output SS:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
sr_no	air	neet_roll_no	cet_form_no	first_name	middle_name	last_name	gender	category	quota	code_college					
1	1	109	3108020911	245022961	POTDAR	YASH	DEEPAK	M	OBC	OPEN	(EMD) 1103:GSMC MUMBAI				
2	2	121	3103110206	245026959	VRAJ	SACHIN	KALE	M	OBC	OPEN	1103:GSMC MUMBAI				
3	4	543	3110130236	245002432	AASTHA	VIRAL	SHAH	F	OPEN	(W)	1103:GSMC MUMBAI				
4	5	657	3112040419	245015391	SAIRAM	KARLE	M	NTC	OPEN	(EMD) 1103:GSMC MUMBAI					
5	6	689	3110010764	245028306	ARYA	PARAG	JADHAV	F	OBC	OPEN	(W) (EMD) 1103:GSMC MUMBAI				
6	7	714	3107050049	245049061	VEDANT	SANJAY	BAWISKAR	M	OBC	OPEN	1103:GSMC MUMBAI				
7	8	720	3115010053	245027936	THANGE	SOHAM	TABAJI	M	OBC	OPEN	(EMD) 1103:GSMC MUMBAI				
8	10	782	3104050120	245040737	ADITYA	RAJESH	UMBARKAR	M	SC	OPEN	(EMD) 1327:GMC CH. SAMBHAJINAGAR				
9	12	1053	3108020545	245033169	SATYAJIT	SUDHIR	JAGTAP	M	SEBC	OPEN	(EMD) 1103:GSMC MUMBAI				
10	13	1055	3108030503	245019667	THANEKAR	ROHINI	NAMDEV	F	SEBCHA	HOPEN	(EMD) 1110:BJMC PUNE				
11	14	1074	3905090341	245014576	RUSHIL	WAGH	M	OBC	OPEN	(EMD) 1103:GSMC MUMBAI					
12	15	1089	3112010404	245027645	SUKHADA	SURESHRAO	JADHAV	F	OPEN	(W)	1103:GSMC MUMBAI				
13	17	1146	3112310282	245009182	PANDALWAD	PAWAN	M	NTB	OPEN	(EMD) 1103:GSMC MUMBAI					
14	19	1187	3111180221	245010378	RABAB	SADIQ	ALI	F	OPEN	(W)	1221:GMC NAGPUR				
15	20	1221	3112400211	245009201	PRAJAKTA	HIRAMAN	BHAGAT	F	SC	OPEN	(W) (EMD) 1103:GSMC MUMBAI				
16	21	1229	3109400269	245046226	WAGALGAVE	ADITYA	TANAJI	M	NTC	OPEN	1103:GSMC MUMBAI				
17	23	1315	3108050056	245044430	KARTIK	JAYKUMAR	SATPUTE	M	NTC	OPEN	(EMD) 1103:GSMC MUMBAI				
18	24	1324	3108100110	245006387	OM	SUDHIR	YESANE	M	HA	HOPEN	1103:GSMC MUMBAI				
19	26	1394	3108020289	245002380	RAMTIRTHAKAR	PRATIK	GAJANAN	M	EWS	OPEN	1103:GSMC MUMBAI				
20	27	1403	3114040017	245021317	NETRA	KALPESH	MODI	F	OPEN	(W)	1103:GSMC MUMBAI				
21	29	1465	3108060132	245007159	SAJAGANE	SANKET	SUNIL	M	NTC	OPEN	(EMD) 1103:GSMC MUMBAI				
22	30	1470	3105080373	245057924	PAWAR	SHREYA	VIJAY	F	OBC	OPEN	(W) (EMD) 1103:GSMC MUMBAI				
23	31	1507	3110250313	245003776	DIYA	SHYAMAL	LANGALIA	F	OPEN	(W)	1103:GSMC MUMBAI				
24	36	1659	3110060085	245046991	MAITREYI	RAMESH	DABHILKAR	F	SC	OPEN	(W) 1103:GSMC MUMBAI				
25	37	1672	3118070160	245029472	SUBBULAKSHMI	SHANKARANARAYAN	F		OPEN	(W)	1103:GSMC MUMBAI				
26	38	1688	3115020920	245006726	ADITI	RAHUL	SAWANT	F	OPEN	(W)	1103:GSMC MUMBAI				
27	39	1696	3114050356	245012826	EVA	MISHRA	F	OPEN	(W)	1103:GSMC MUMBAI					
28	40	1701	3112080257	245038609	ANKITA	ANANDRAO	HAMBARDE	F	OPEN	(W)	1103:GSMC MUMBAI				
29	41	1710	3110170333	245003009	ADITI	SANJAY	GUPTA	F	OPEN	(W)	1103:GSMC MUMBAI				
30	42	1734	3109210453	245007113	SYED	ABDUL RAHMAN	IRF M	EWS	OPEN	1103:GSMC MUMBAI					
31	43	1801	3115130211	245017175	PALLERLA	PRANAV	GANESH	M	SOBC	OPEN	(EMD) 1103:GSMC MUMBAI				
32	45	1847	3111090409	245000581	PAULOMI	SAMIR	PALEWAR	F	OPEN	(W)	1103:GSMC MUMBAI				
33	46	1853	3110060590	245013407	ISHITA	HRIIDYANATH	MHATRE	F	OPEN	(W)	1103:GSMC MUMBAI				
34	47	1885	3112270229	245016471	RIYA	LALIT	PATIL	F	OBC	OPEN	(W) 1140:HBT & COOPER MC JUHU, MUMBAI				
35	48	1887	3118060336	245003119	PATIL	SUSHRUTA	PRAVIN	F	OPEN	(W)	1103:GSMC MUMBAI				
36	49	1928	3110310242	245012536	NISCAA	CHANDRASEKHARAN	NAIR	F	OPEN	(W)	1103:GSMC MUMBAI				
37	51	1969	3118010857	245016763	HUBLI	ANUSHKA	SHIVAKUMAR	F	OPEN	(W)	1103:GSMC MUMBAI				

MBBS//BDS Extraction Manipulation and Closing Rank finder Script

2.Script for Sorting data neatly:

```
import pandas as pd
import re

# Load the CSV file
df = pd.read_csv('output.csv')

# Function to extract text before numbers
def extract_text_before_numbers(value):
    match = re.match(r"([^\d]+)", value)
    return match.group(0).strip() if match else ''

# Apply the function to the 'cod_college' column
df['new_column'] = df['code_college'].apply(extract_text_before_numbers)

# Function to remove text before numbers in 'cod_college'
def remove_text_before_numbers(value):
    match = re.search(r"\d.*", value) # Find the first occurrence of a digit
    # and everything after it
    return match.group(0).strip() if match else value # Return the matched
    # part or the original value if no match

# Apply the function to the 'cod_college' column
df['code_college'] = df['code_college'].apply(remove_text_before_numbers)

# Merge 'Quota' and 'new_column' into a single column
df['Quota'] = df['category'].fillna('') + ' ' + df['quota'].fillna('') +
df['new_column'].fillna('')
# Save the updated DataFrame to a new CSV file

print(df)

# Save the updated DataFrame to a new CSV file
df.to_csv('updated_file.csv', index=False)
```

Above script is used to clean and manipulate data in a CSV file, specifically focusing on the `code_college` column. Here's the purpose and use of each part of the script:

1. Import necessary libraries:

- o `pandas`: For data manipulation and analysis.

MBBS//BDS Extraction Manipulation and Closing Rank finder Script

- `re`: For working with regular expressions to handle text patterns.
- 2. **Load the CSV file:**
 - `df = pd.read_csv('output.csv')`: Reads the data from the `output.csv` file into a pandas DataFrame for further processing.
- 3. **Function to extract text before numbers:**
 - `extract_text_before_numbers(value)`: This function takes a string and uses a regular expression to extract and return the text that appears before any digits. If no match is found, it returns an empty string.
 - `df['new_column'] = df['code_college'].apply(extract_text_before_numbers)`: Applies the function to the `code_college` column in the DataFrame and stores the result in a new column called `new_column`. This column will contain the text part of the `code_college` values (i.e., everything before the numbers).
- 4. **Function to remove text before numbers in `code_college`:**
 - `remove_text_before_numbers(value)`: This function removes the text before the first number in the `code_college` values and returns the remaining string (i.e., the numeric part along with any following text). If no numbers are found, it returns the original value.
 - `df['code_college'] = df['code_college'].apply(remove_text_before_numbers)`: Applies the function to the `code_college` column, updating it to only contain the numeric part (and any text after the number).
- 5. **Merge Quota and new_column into a single column:**
 - `df['Quota'] = df['category'].fillna('') + ' ' + df['quota'].fillna('') + df['new_column'].fillna('')`: Combines the values of the `category`, `quota`, and `new_column` columns into a new `Quota` column. The `fillna('')` ensures that any missing values are replaced with an empty string before concatenation.
- 6. **Save the updated DataFrame:**
 - `print(df)`: Prints the updated DataFrame to the console for verification.
 - `df.to_csv('updated_file.csv', index=False)`: Saves the modified DataFrame to a new CSV file named `updated_file.csv`, without including the index column.

Overall

Purpose:

This script processes the `code_college` column by splitting and cleaning its content, then merges the cleaned data with other columns to create a more organized and useful structure. The result is saved in a new CSV file for further analysis or use.

MBBS//BDS Extraction Manipulation and Closing Rank finder Script

Input SS:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	sr_no	air	neet_rol_no	cet_form_no	first_name	middle_name	last_name	gender	category	quota	code_college				
1	1	109	3108020911	245022961	POTDAR	YASH	DEEPAK	M	OBC	OPEN	(EMD) 1103:GSMC MUMBAI				
2	2	121	3103110206	245026959	VRAJ	SACHIN	KALE	M	OBC	OPEN	1103:GSMC MUMBAI				
3	4	543	3110130236	245002432	AASTHA	VIRAL	SHAH	F	OPEN	(W)	1103:GSMC MUMBAI				
4	5	657	3112040419	245015391	SAIRAM		KARLE	M	NTC	OPEN	(EMD) 1103:GSMC MUMBAI				
5	6	689	3110010764	245028306	ARYA	PARAG	JADHAV	F	OBC	OPEN	(W) (EMD) 1103:GSMC MUMBAI				
6	7	714	3107050049	245049061	VEDANT	SANJAY	BAWISKAR	M	OBC	OPEN	1103:GSMC MUMBAI				
7	8	720	3115010053	245027936	THANGE	SOHAM	TABAJI	M	OBC	OPEN	(EMD) 1103:GSMC MUMBAI				
8	10	782	3104050120	245040737	ADITYA	RAJESH	UMBARKAR	M	SC	OPEN	(EMD) 1327:GMC CH. SAMBHAIINAGAR				
9	12	1053	3108020545	245033169	SATYAJIT	SUDHIR	JAGTAP	M	SEBC	OPEN	(EMD) 1103:GSMC MUMBAI				
10	13	1055	3108030503	245019667	THANEKAR	ROHINI	NAMDEV	F	SEBCHA	HOPEN	(EMD) 1110:BJMC PUNE				
11	14	1074	3905090341	245014576	RUSHIL		WAGH	M	OBC	OPEN	(EMD) 1103:GSMC MUMBAI				
12	15	1089	3112010404	245027645	SUKHADA	SURESHRAO	JADHAV	F	OPEN	(W)	1103:GSMC MUMBAI				
13	17	1146	3112310282	245009182	PANDALWAD		PAWAN	M	NTB	OPEN	(EMD) 1103:GSMC MUMBAI				
14	19	1187	3111180221	245010378	RABAB	SADIQ	ALI	F	OPEN	(W)	1221:GMC NAGPUR				
15	20	1221	3112400211	245009201	PRAJAKTA	HIRAMAN	BHAGAT	F	SC	OPEN	(W) (EMD) 1103:GSMC MUMBAI				
16	21	1229	3109400269	245046226	WAGALGAVE	ADITYA	TANAJI	M	NTC	OPEN	1103:GSMC MUMBAI				
17	23	1315	3108050056	245044430	KARTIK	JAYKUMAR	SATPUTE	M	NTC	OPEN	(EMD) 1103:GSMC MUMBAI				
18	24	1324	3108100110	245006387	OM	SUDHIR	YESANE	M	HA	HOPEN	1103:GSMC MUMBAI				
19	26	1394	3108020289	245002380	RAMTIRTHAKAR	PRATIK	GAJANAN	M	EWS	OPEN	1103:GSMC MUMBAI				
20	27	1403	3114040017	245021317	NETRA	KALPESH	MODI	F	OPEN	(W)	1103:GSMC MUMBAI				
21	29	1465	3108060132	245007159	SAJAGANE	SANKET	SUNIL	M	NTC	OPEN	(EMD) 1103:GSMC MUMBAI				
22	30	1470	3105080373	245057924	PAWAR	SHREYA	VIJAY	F	OBC	OPEN	(W) (EMD) 1103:GSMC MUMBAI				
23	31	1507	3110250313	245003776	DIYA	SHYAMAL	LANGALIA	F	OPEN	(W)	1103:GSMC MUMBAI				
24	36	1659	3110060085	245046991	MAITREYI	RAMESH	DABHILKAR	F	SC	OPEN	(W) 1103:GSMC MUMBAI				
25	37	1672	3118070160	245029472	SUBBULAKSHMI		SHANKARANARAYAN	F	OPEN	(W)	1103:GSMC MUMBAI				
26	38	1688	3115020920	245006726	ADITI	RAHUL	SAWANT	F	OPEN	(W)	1103:GSMC MUMBAI				
27	39	1696	3114050356	245012826	EVA		MISHRA	F	OPEN	(W)	1103:GSMC MUMBAI				
28	40	1701	3112080257	245038609	ANKITA	ANANDRAO	HAMBARDE	F	OPEN	(W)	1103:GSMC MUMBAI				
29	41	1710	3110170333	245003009	ADITI	SANJAY	GUPTA	F	OPEN	(W)	1103:GSMC MUMBAI				
30	42	1734	3109210453	245007113	SYED		ABDUL RAHMAN IRF	M	EWS	OPEN	1103:GSMC MUMBAI				
31	43	1801	3115130211	245017175	PALLERLA	PRANAV	GANESH	M	SOBC	OPEN	(EMD) 1103:GSMC MUMBAI				
32	45	1847	3111090409	245000581	PAULOMI	SAMIR	PALTEWAR	F	OPEN	(W)	1103:GSMC MUMBAI				
33	46	1853	3110060590	245013407	ISHITA	HRIADYNATH	MHATRE	F	OPEN	(W)	1103:GSMC MUMBAI				
34	47	1885	3112270229	245016471	RIYA	LALIT	PATIL	F	OBC	OPEN	(W) 1140:HBT & COOPER MC JUHU, MUMBAI				
35	48	1887	3118060336	245003119	PATIL	SUSHRUTA	PRAYIN	F	OPEN	(W)	1103:GSMC MUMBAI				
36	49	1928	3110310242	245012536	NISCAA	CHANDRASEKHARAI	NAIR	F	OPEN	(W)	1103:GSMC MUMBAI				
37	51	1969	3118010857	245016763	HUBLI	ANUSHKA	SHIVAKUMAR	F	OPEN	(W)	1103:GSMC MUMBAI				

Output SS:

	Unnamed: sr_no	air	neet_rol_no	cet_form_no	first_name	middle_name	last_name	gender	category	quota	code_college	Quota		
1	1	1	109	3108020911	245022961	POTDAR	YASH	DEEPAK	M	OBC	OPEN	1103:GSMC MUMBAI	OBC OPEN(EMD)	
2	2	2	121	3103110206	245026959	VRAJ	SACHIN	KALE	M	OBC	OPEN	1103:GSMC MUMBAI	OBC OPEN	
3	3	4	543	3110130236	245002432	AASTHA	VIRAL	SHAH	F	OPEN	(W)	1103:GSMC MUMBAI	OPEN (W)	
4	4	5	657	3112040419	245015391	SAIRAM		KARLE	M	NTC	OPEN	1103:GSMC MUMBAI	NTC OPEN(EMD)	
5	5	6	689	3110010764	245028306	ARYA	PARAG	JADHAV	F	OBC	OPEN	1103:GSMC MUMBAI	OBC OPEN(W) (EMD)	
6	6	7	714	3107050049	245049061	VEDANT	SANJAY	BAWISKAR	M	OBC	OPEN	1103:GSMC MUMBAI	OBC OPEN	
7	7	8	720	3115010053	245027936	THANGE	SOHAM	TABAJI	M	OBC	OPEN	1103:GSMC MUMBAI	OBC OPEN(EMD)	
8	8	10	782	3104050120	245040737	ADITYA	RAJESH	UMBARKAR	M	SC	OPEN	1327:GMC CH. SAMBHAIINAGAR	SC OPEN(EMD)	
9	9	12	1053	3108020545	245033169	SATYAJIT	SUDHIR	JAGTAP	M	SEBC	OPEN	1103:GSMC MUMBAI	SEBC OPEN(EMD)	
10	10	13	1055	3108030503	245019667	THANEKAR	ROHINI	NAMDEV	F	SEBCHA	HOPEN	1110:BJMC PUNE	SEBCHA HOPEN(EMD)	
11	11	14	1074	3905090341	245014576	RUSHIL		WAGH	M	OBC	OPEN	1103:GSMC MUMBAI	OBC OPEN(EMD)	
12	12	15	1089	3112010404	245027645	SUKHADA	SURESHRAO	JADHAV	F	OPEN	(W)	1103:GSMC MUMBAI	OPEN (W)	
13	13	17	1146	3112310282	245009182	PANDALWAD		PAWAN	M	NTB	OPEN	1103:GSMC MUMBAI	NTB OPEN(EMD)	
14	14	19	1187	3111180221	245010378	RABAB	SADIQ	ALI	F	OPEN	(W)	1221:GMC NAGPUR	OPEN (W)	
15	15	20	1221	3112400211	245009201	PRAJAKTA	HIRAMAN	BHAGAT	F	SC	OPEN	1103:GSMC MUMBAI	SC OPEN(W) (EMD)	
16	16	21	1229	3109400269	245046226	WAGALGA	ADITYA	TANAJI	M	NTC	OPEN	1103:GSMC MUMBAI	NTC OPEN	
17	17	23	1315	3108050056	245044430	KARTIK	JAYKUMAR	SATPUTE	M	NTC	OPEN	1103:GSMC MUMBAI	NTC OPEN(EMD)	
18	18	24	1324	3108100110	245006387	OM	SUDHIR	YESANE	M	HA	HOPEN	1103:GSMC MUMBAI	HA HOPEN	
19	19	26	1394	3108020289	245002380	RAMTIRTH	PRATIK	GAJANAN	M	EWS	OPEN	1103:GSMC MUMBAI	EWS OPEN	
20	20	27	1403	3114040017	245021317	NETRA	KALPESH	MODI	F	OPEN	(W)	1103:GSMC MUMBAI	OPEN (W)	
21	21	29	1465	3108060132	245007159	SAJAGANE	SANKET	SUNIL	M	NTC	OPEN	1103:GSMC MUMBAI	NTC OPEN(EMD)	
22	22	30	1470	3105080373	245057924	PAWAR	SHREYA	VIJAY	F	OBC	OPEN	1103:GSMC MUMBAI	OBC OPEN(W) (EMD)	
23	23	31	1507	3110250313	245003776	DIYA	SHYAMAL	LANGALIA	F	OPEN	(W)	1103:GSMC MUMBAI	OPEN (W)	
24	24	36	1659	3110060085	245046991	MAITREYI	RAMESH	DABHILKA	F	SC	OPEN	1103:GSMC MUMBAI	SC OPEN(W)	
25	25	37	1672	3118070160	245029472	SUBBULAKSHMI		SHANKAR	F	OPEN	(W)	1103:GSMC MUMBAI	OPEN (W)	
26	26	38	1688	3115020920	245006726	ADITI	RAHUL	SAWANT	F	OPEN	(W)	1103:GSMC MUMBAI	OPEN (W)	

MBBS//BDS Extraction Manipulation and Closing Rank finder Script

3.Script for finding Max AIR (Closing Rank) for Extracted sorted CSV file

```
import pandas as pd

# Load the CSV file
df = pd.read_csv('updated_file.csv')

# Group by college, category, and gender to find max AIR
max_air_df = df.loc[df.groupby(['code_college', 'category', 'gender',
'Quota'])['air'].idxmax()]

# Rename columns to indicate max
max_air_df = max_air_df.rename(columns={'air': 'max_air', 'sr_no':
'max_sr_no', 'neet_roll_no': 'max_neet_roll_no', 'cet_form_no':
'max_cet_form_no', 'Quota': 'max_quota'})

# Select relevant columns for the final output
columns_to_keep = [
    'code_college', 'category', 'gender',
    'max_air', 'max_quota',
]

# Display the result
final_df = max_air_df[columns_to_keep]

# Convert to CSV
final_df.to_csv('output_Max.csv', index=False)
print(final_df)
```

Here's a simple explanation of what each part of the script does:

1. **Group by College, Category, and Gender to Find Maximum AIR:**, and quota. It then finds the row with the highest All India Rank (AIR) within each group and creates a new DataFrame (`max_air_df`) with these maximum AIR values.
2. **Rename Columns to Indicate Maximum Values:**This renames the columns in `max_air_df` to clearly indicate that they contain the maximum values (e.g., 'max_air' instead of 'air').
3. **Merge Minimum and Maximum DataFrames:**This merges `min_air_df` (which presumably contains minimum AIR values) with `max_air_df` on the columns `code_college`, `category`, `gender`, and `Quota`. The result is a DataFrame (`result_df`) that contains both minimum and maximum AIR values for each group.
4. **Select Relevant Columns for Final Output:**This selects only the columns of interest from `result_df` (e.g., college code, category, gender, maximum AIR, and maximum quota) and creates a new DataFrame (`final_df`) with these selected columns.

MBBS//BDS Extraction Manipulation and Closing Rank finder Script

5. **Convert Final DataFrame to CSV:** This saves the `final_df` DataFrame to a CSV file named `output_Max.csv` without including row indices.

In summary, the script processes data to find the maximum AIR for each college, category, gender, and quota combination, merges this information with minimum AIR data, selects relevant columns, and then saves the final results to a CSV file.

Input SS:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
sr_no	air	neet_rol_no	cet_form_no	first_name	middle_name	last_name	gender	category	quota	code_college						
1	1	109	3108020911	245022961	POTDAR	YASH	DEEPAK	M	OBC	OPEN	(EMD) 1103-GSMC MUMBAI					
2	2	121	3103110206	245026959	VRAJ	SACHIN	KALE	M	OBC	OPEN	1103-GSMC MUMBAI					
3	4	543	3110130236	245002432	AASTHA	VIRAL	SHAH	F	OPEN	(W)	1103-GSMC MUMBAI					
4	5	657	3112040419	245015391	SAIRAM	KARLE	JADHAV	M	NTC	OPEN	(EMD) 1103-GSMC MUMBAI					
5	6	689	3110010764	245028306	ARIYA	PARAG	BAWISKAR	F	OBC	OPEN	(W) (EMD) 1103-GSMC MUMBAI					
6	7	714	3107050049	245049061	VEDANT	SANJAY	SOHAM	M	OBC	OPEN	1103-GSMC MUMBAI					
7	8	720	3115010053	245027936	THANGE	SOHAM	TABAI	M	OBC	OPEN	(EMD) 1103-GSMC MUMBAI					
8	10	782	3104050120	245040737	ADITYA	RAJESH	UMBARKAR	M	SC	OPEN	(EMD) 1327-GMC CH. SAMBAJI NAGAR					
9	12	1053	3108020545	245033169	SATYAJIT	SUDHIR	JAGTAP	M	SEBC	OPEN	(EMD) 1103-GSMC MUMBAI					
10	13	1055	3108030503	245019667	THANEKAR	ROHINI	NAMDEV	F	SEBCHA	HOPEN	(EMD) 1110-BJMC PUNE					
11	14	1074	3905090341	245014576	RUSHIL	WAGH	M	OBC	OPEN	(EMD) 1103-GSMC MUMBAI						
12	15	1089	3112010404	245027645	SUKHADA	SURESHRAO	JADHAV	F	OPEN	(W)	1103-GSMC MUMBAI					
13	17	1146	3112310282	245009182	PANDALWAD	PAWAN	M	NTB	OPEN	(EMD) 1103-GSMC MUMBAI						
14	19	1187	3111180221	245010378	RASAB	SADIQ	ALI	F	OPEN	(W)	1221-GMC NAGPUR					
15	20	1221	3112400211	245009201	PROAKTA	HIRAMAN	BHAGAT	F	SC	OPEN	(W) (EMD) 1103-GSMC MUMBAI					
16	21	1229	3109400269	245046226	WAGALGAVE	ADITYA	TANAI	M	NTC	OPEN	1103-GSMC MUMBAI					
17	23	1315	3108050056	245044430	KARTIK	JAYKUMAR	SATPUTE	M	NTC	OPEN	(EMD) 1103-GSMC MUMBAI					
18	24	1324	3108100110	245006387	OM	SUDHIR	YESANE	M	HA	HOPEN	1103-GSMC MUMBAI					
19	26	1394	3108020289	245002380	RAMTIRTHAKAR	PRATIK	GAJANAN	M	EWS	OPEN	1103-GSMC MUMBAI					
20	27	1403	3114040017	245021317	NETRA	KALPESH	MODI	F	OPEN	(W)	1103-GSMC MUMBAI					
21	29	1465	3108060132	245007159	SAJAGANE	SANKET	SUNIL	M	NTC	OPEN	(EMD) 1103-GSMC MUMBAI					
22	30	1470	3105080373	245057924	PAWAR	SHREYA	VIJAY	F	OBC	OPEN	(W) (EMD) 1103-GSMC MUMBAI					
23	31	1507	3110250313	245003776	DIVA	SHYAMAL	LANGALIA	F	OPEN	(W)	1103-GSMC MUMBAI					
24	36	1659	3110060085	245046991	MAITREYI	RAMESH	DABHILKAR	F	SC	OPEN	(W) 1103-GSMC MUMBAI					
25	37	1672	3118070160	245029472	SUBBULAKSHMI	SHANKARANARAYAN	F	OPEN	(W)	1103-GSMC MUMBAI						
26	38	1688	3115020920	245006726	ADITI	RAHUL	SAWANT	F	OPEN	(W)	1103-GSMC MUMBAI					
27	39	1696	3114050356	245012826	EVA	MISHRA	F	OPEN	(W)	1103-GSMC MUMBAI						
28	40	1701	3112080257	245038609	ANKITA	ANANDRAO	HAMBARDE	F	OPEN	(W)	1103-GSMC MUMBAI					
29	41	1710	3110170333	245003009	ADITI	SANJAY	GUPTA	F	OPEN	(W)	1103-GSMC MUMBAI					
30	42	1734	3109210453	245007113	SYED	ABDUL RAHMAN IRF M	M	EWS	OPEN	1103-GSMC MUMBAI						
31	43	1801	3115130211	245017175	PALLERLA	PRANAV	GANESH	M	SOBC	OPEN	(EMD) 1103-GSMC MUMBAI					
32	45	1847	3111090409	245000581	PAULOMI	SAMIR	PALTEWAR	F	OPEN	(W)	1103-GSMC MUMBAI					
33	46	1853	3110060590	245013407	ISHITA	HRIDAYNATH	MHATRE	F	OPEN	(W)	1103-GSMC MUMBAI					
34	47	1885	3112270229	245016471	RIYA	LALIT	PATIL	F	OBC	OPEN	(W) 1140-HBT & COOPER MC JUHU, MUMBAI					
35	48	1887	3118060336	245003119	PATIL	SUSHRUTA	PRAVIN	F	OPEN	(W)	1103-GSMC MUMBAI					
36	49	1928	3110310242	245012536	NISCAA	CHANDRASEKHARAN	NAIR	F	OPEN	(W)	1103-GSMC MUMBAI					
37	51	1969	3118010857	245016763	HUBLI	ANUSHKA	SHIVAKUMAR	F	OPEN	(W)	1103-GSMC MUMBAI					

Output SS:

	code_college	category	gender	max_air	max_quota	
1	1 1101:GMC MUMBAI	NTD	M	9301	NTD D1DEF	
2	1 1101:GMC MUMBAI	OBC	M	18853	OBC D1DEF	
3	1 1102:Tilak MC MUMBAI	NTC	M	11589	NTC D1DEF	
4	1 1102:Tilak MC MUMBAI	OBC	M	11953	OBC D1DEF	
5	1 1103:GSMC MUMBAI	OBC	M	4147	OBC D1DEF	
6	1 1104:Nair MC MUMBAI	OBC	M	19010	OBC D1DEF	
7	1 1109:RGMC KALVA THANE	OBC	M	38880	OBC D1DEF	
8	1 1110:BJMC PUNE	OBC	M	9508	OBC D1DEF	
9	1 1114:V M MC SOLAPUR	OBC	F	45977	OBC D1DEF	
10	1 1114:V M MC SOLAPUR	OBC	M	45101	OBC D1DEF	
11	1 1119:B HIRE MC DHULE	OBC	F	50950	OBC D1DEF	
12	1 1119:B HIRE MC DHULE	OBC	M	50306	OBC D1DEF	
13	1 1132:GMC KOLHAPUR	OBC	M	47060	OBC D1DEF	
14	1 1149:GMC JALGAON	NTD	F	61851	NTD D1DEF	
15	1 1149:GMC JALGAON	OBC	M	61806	OBC D1DEF	
16	1 1150:GMC BARAMATI	NTC	M	33483	NTC D1DEF	
17	1 1150:GMC BARAMATI	OBC	M	39971	OBC D1DEF	
18	1 1153:GMC ALIBAUG	NTC	M	59225	NTC D1DEF	
19	1 1154:GMC SINDHUDURG	OBC	M	66748	OBC D1HADEF	
20	1 1155:GMC SATARA	OBC	M	48063	OBC D1DEF	
21	1 1159:GMC RATNAGIRI	OBC	M	64232	OBC D1DEF	
22	1 1164:GMC NASHIK	VJA	M	48754	VJA D1DEF	
23	1 1222:IGMC NAGPUR	EWS	M	43376	EWS D1DEF	
24	1 1222:IGMC NAGPUR	OBC	M	37050	OBC D1DEF	
25	1 1226:VNMC YEOTMAL	NTD	M	60370	NTD D1DEF	