

Data Visualization

RAHUL BISWAS

M.SC. COMPUTER SCIENCE (SEM - 4)

UNIVERSITY OF CALCUTTA

DATA SET DESCRIPTION

- Name: NEET Rank & Score Analysis (2015-2024)
- <u>About</u>: This dataset contains NEET (National Eligibility cum Entrance Test) rank and score information for students from 2015 to 2024. The data includes AIR (AII India Rank), state ranks, category-wise cutoffs, percentile scores, and total marks obtained.
- Source & Author: https://www.kaggle.com/datasets/tanveersinghbedi/neet-rank-and-score

-analysis-2015-2024/; Tanveer Singh

DATA SET DESCRIPTION (Continued)

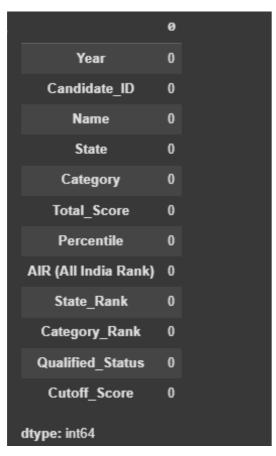
- Number of Rows: 50000
- Column details:
- **1. Year** Exam year
- 2. Candidate_ID Unique candidate identifier
- 3. Name Candidate's name
- 4. State State of domicile
- **5. Category** General, OBC, SC, ST, EWS
- **6. Total_Score** Total marks obtained out of 720

• Number of Column: 12

- 7. Percentile Percentile score
- 8. AIR(All India Rank) Overall rank in India
- 9. State_Rank Rank within the state
- **10.** Category_Rank Rank within the category
- **11. Qualified_Status** Yes/No
- **12. Cutoff_Score** minimum qualifying marks

DATA CLEANING

Checking for Null Value:



• Checking for duplicate values:

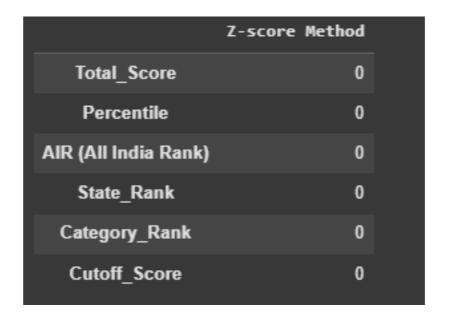
```
1 has_duplicates = df.duplicated().any()
2
3 print("Does the dataset have duplicates?", has_duplicates)

Does the dataset have duplicates? False
```

* The dataset does not have any derived attributed as well.

DATA CLEANING(Continued)

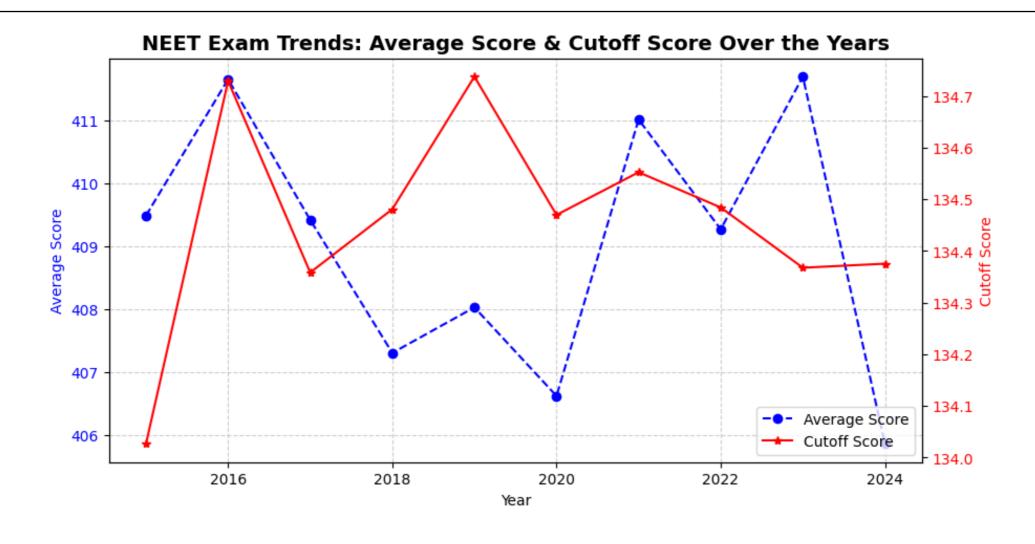
• Find outliers (Z Score method):



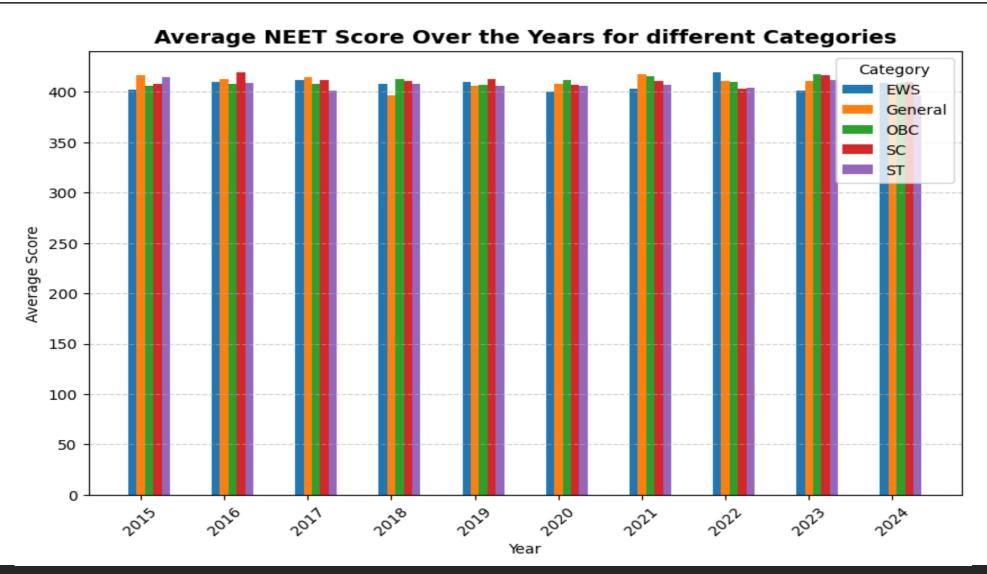
^{*} The data set is cleaned.

RESULTS

Average Score & Cutoff Score over the years



Average Score for Different Categories



Number of Candidates Qualified vs Not Qualified Over Years



