

Sanner Said This Would be on the Midterm

MIE223
Winter 2025

1 Feature Analysis and Visualization

This will be on the midterm

Univariate has discrete or numeric/continuous columns.

Discrete: bar plot of frequency

Numeric/Continuous: histogram, box plot, violin plot

Discrete heatmap of MI, PMI can produce correlation between D and d

Scatter plot, heatmap is good for C and C

Box, violin, overlapping histogram plots for D and C

2 Basic NLP Text Processing

- Match dates from 2000 to 2025 (Reg-Ex)
 - Only compound statements allowed
 - Do not just list out all the options
 - 20[0-1][0-9] — 202[0-5]