Class 5 Group 5

MUSIC GENRE CLASSIFICATION

KAGGLE COMPETITION - SHAI 2024

21 August, 2024

AGENDA OVERVIEW

HELLO!

In this Kaggle competition, we're building models to classify songs into genres using audio features like Popularity, danceability, energy, etc.

Class 5 - Group 5

HOW DO we work?

Our team embraced a collaborative workflow, with each member independently experimenting in their own notebooks. Throughout the days, findings were shared, facilitating an exchange of insights. At the task's end, we selected the best-performing model based on the lowest RMSE for submission on Kaggle. To consolidate our efforts, all notebooks were merged, creating a unified record of our progress.

Class 5 - Group 5

BIG PICTURE

We want to understand each feature and its effect on the accuracy.

- 14 futures
- 14396 records
- target is Class column -> 0 to 10

PREPARING THE DATA
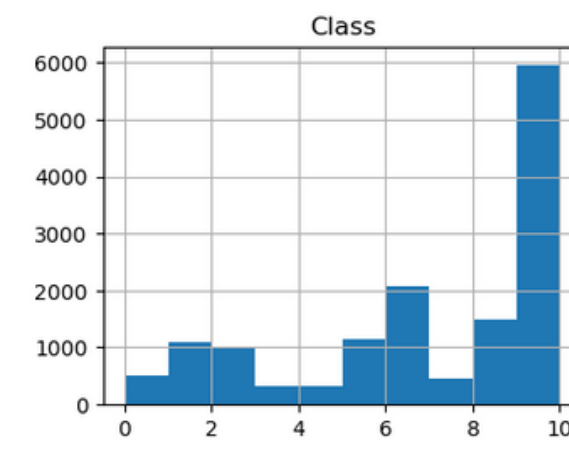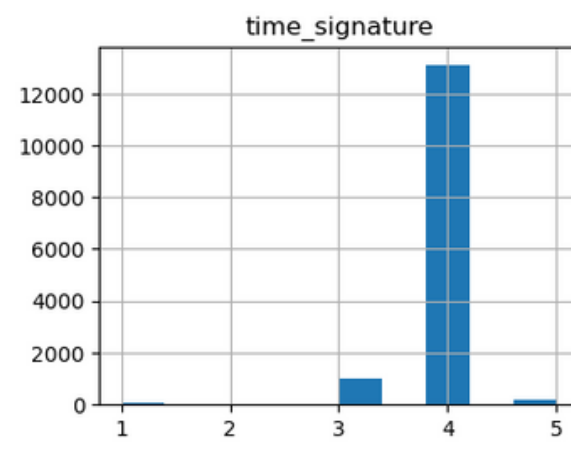
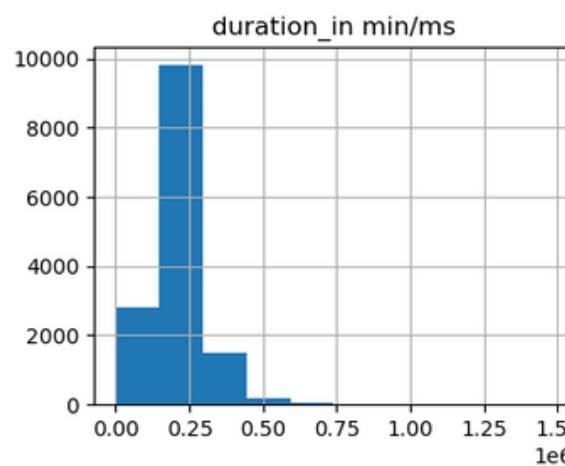## 01
HANDLE NULL VALUES

02
DEALING WITH COLUMNS

03
FEATURE SELECTION

01 HANDLE NULL VALUES

- Filling them with
  - ☐ Mean, Median, Mode
- Building models to predict them
- Drop null value
  - ☐ low accuracy

- Drop duplicated rows

VISUALISE DATA

Assist in selecting the most effective method for filling missing data.
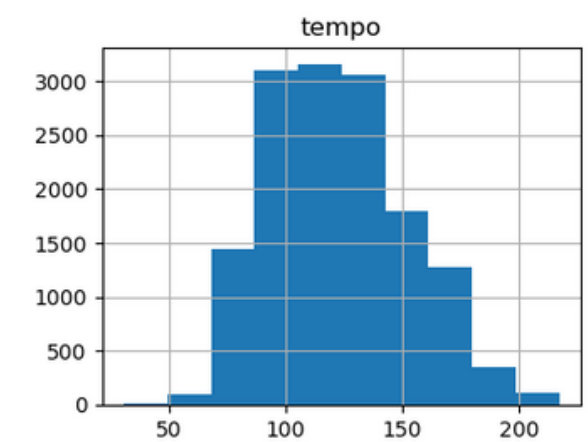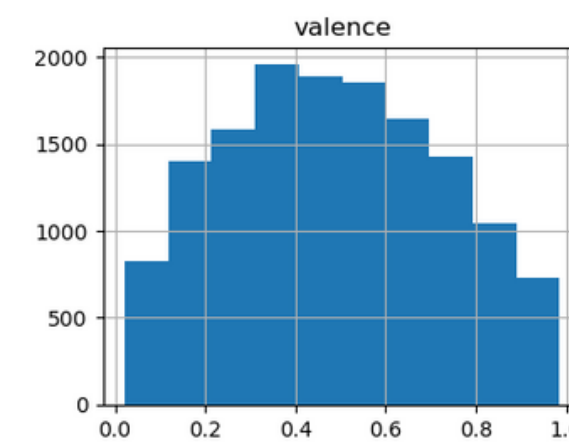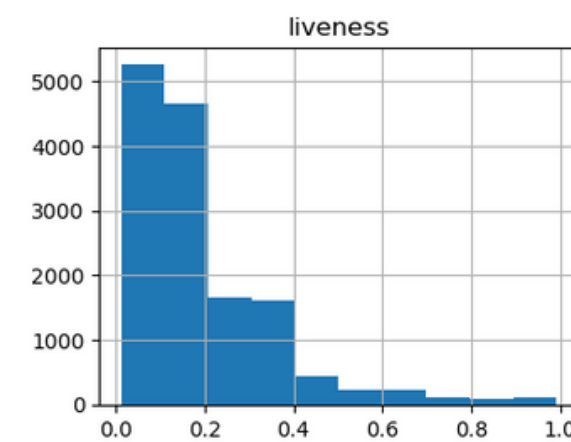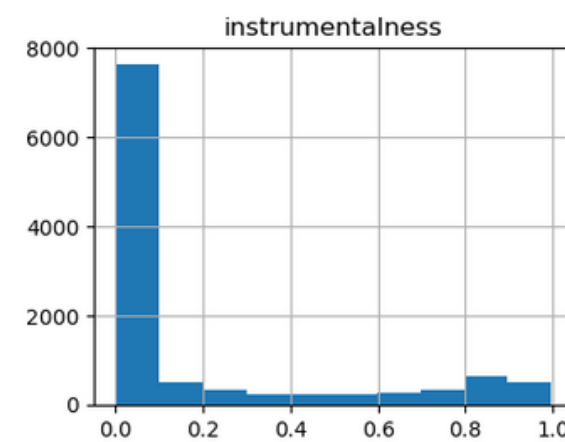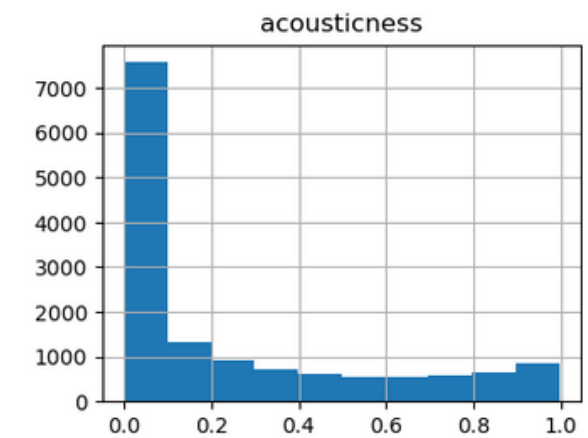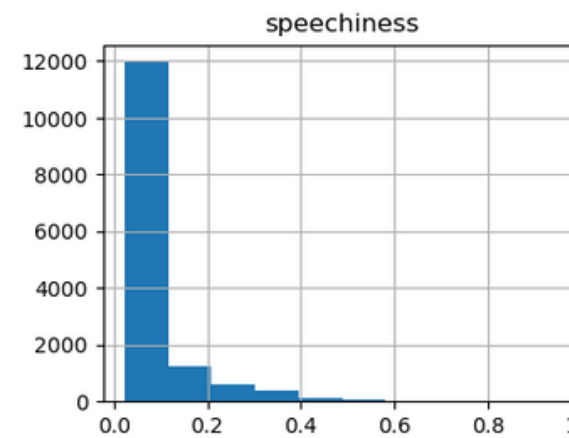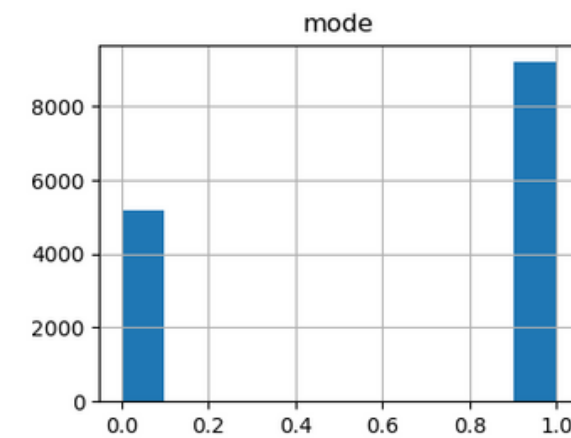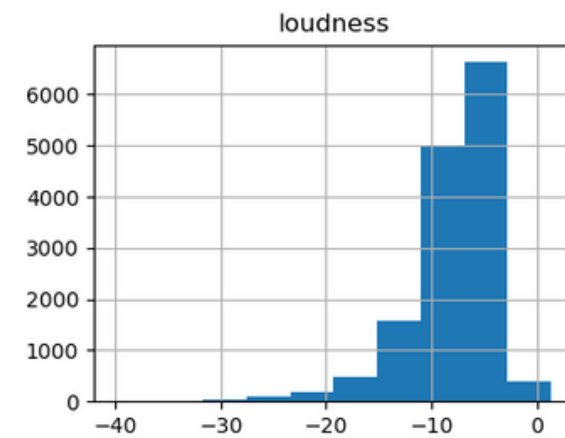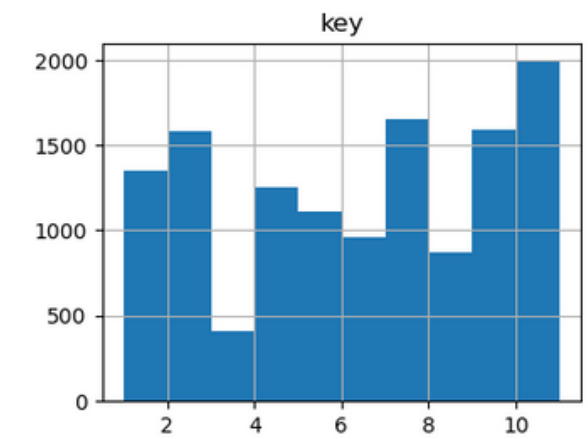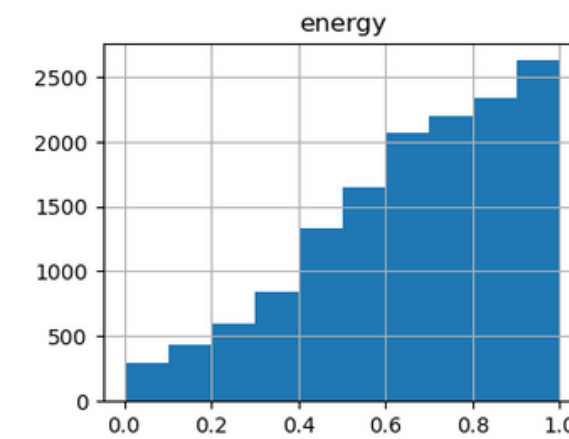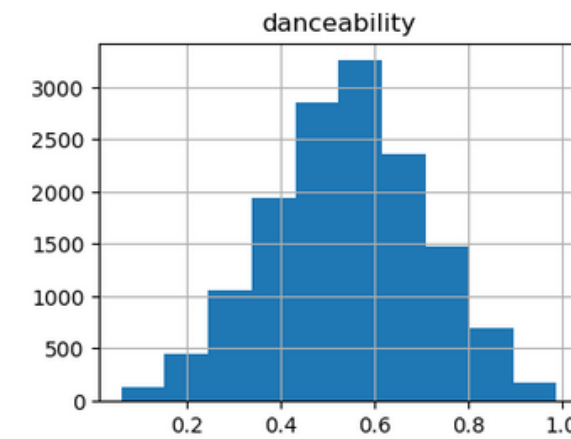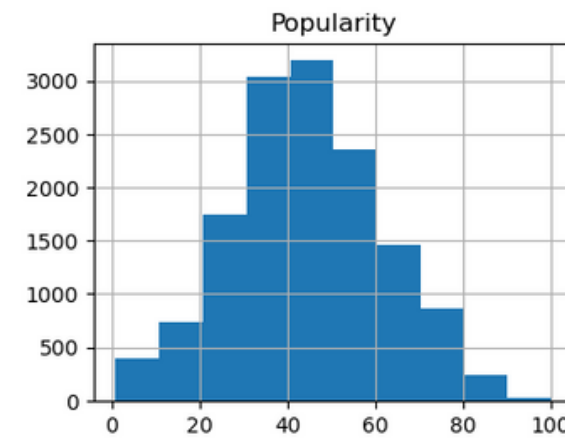
A– Categorical data

- dropping some columns like
    - Artist Name, Track Name, etc
        - increase accuracy
- Using OneHotEncoding to handle categorical data
- TargetEncoding
    - Same accuracy

B- Numerical data

- dropping some columns like
  - ID
- Standard scaling
- convert "duration_in min/ms" Column to minutes

VISUALISE DATA

Distribution of duration before and after transform to minutes
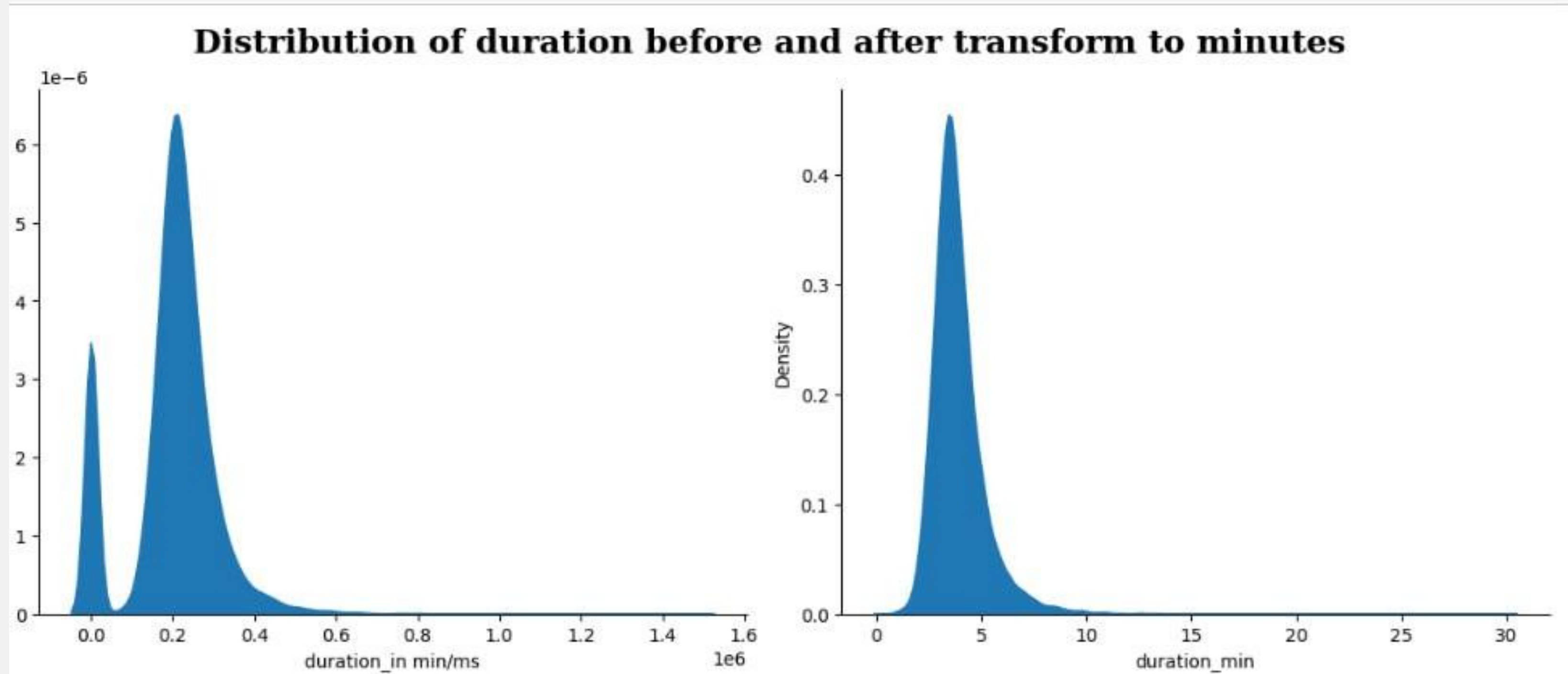
03 FEATURE SELECTION

- trying to get new columns to increase accuracy
    - □ same accuracy

- remove outliers values
    - □ using IQR

TRANSFORMATIONS

Skewness in features before apply some transformations:

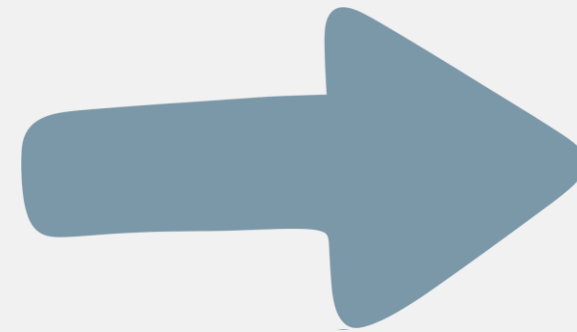|	0
Popularity	0.08
danceability	-0.08
energy	-0.66
key	-0.05
loudness	-1.75
mode	-0.58
speechiness	3.11
acousticness	1.11
instrumentalness	1.53
liveness	2.18
valence	0.09
tempo	0.38
duration_in min/ms	0.84
time_signature	-4.11

Skewness in features after transformations:

	0
Popularity	0.08
danceability	-0.08
energy	-0.66
key	-0.05
loudness	-0.18
mode	-0.58
speechiness	-0.09
acousticness	0.20
instrumentalness	-0.20
liveness	0.48
valence	0.09
tempo	0.38
duration_in min/ms	0.84
time_signature	-4.11

F1-SCORE

CATBOOST
- 0.70106

RANDOM FOREST
- 0.6037

XGBOOST
- 0.6160

LIGHTGBM
- 0.5701
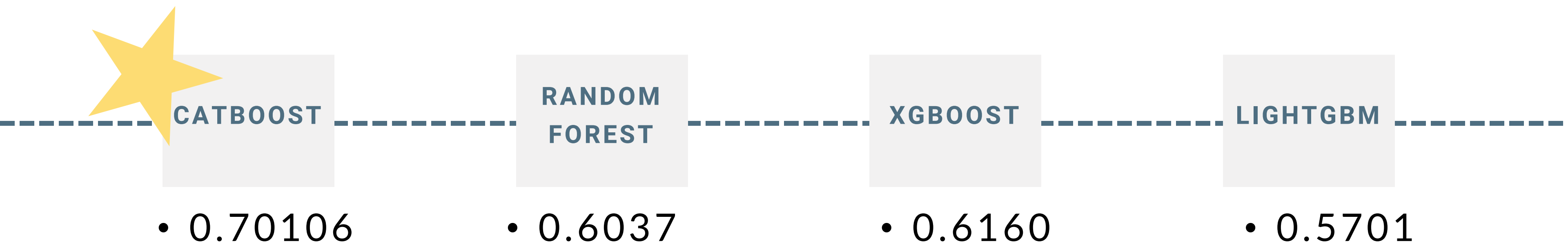
RESULT
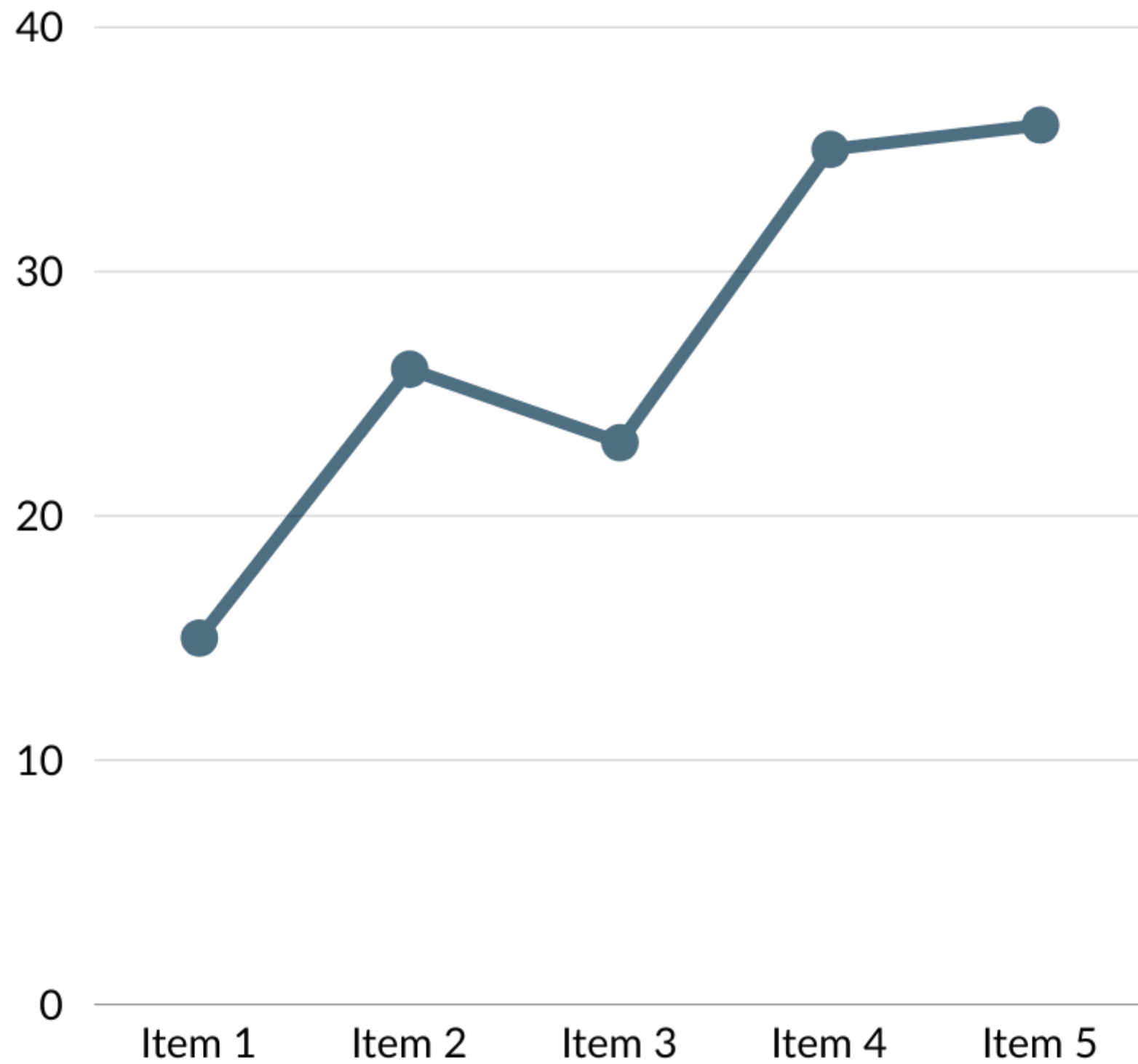
Public dashboard

7

Private dashboard

6

Class 5 - Group 5

THANK YOU

21 August, 2024