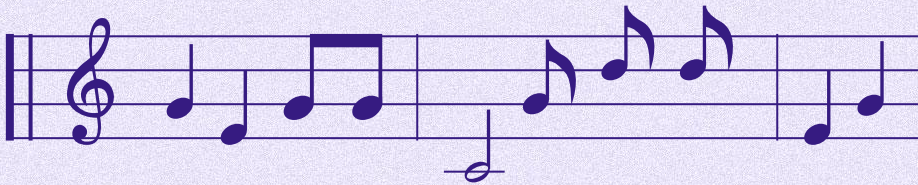
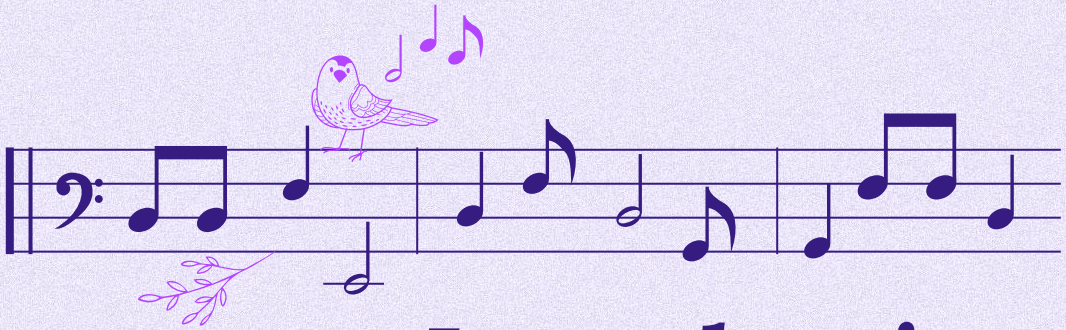
Beatmap Analysis

by Hannah Garcia





Introduction

01

Beatmaps

6

Easy

17

Hard

30

Expert

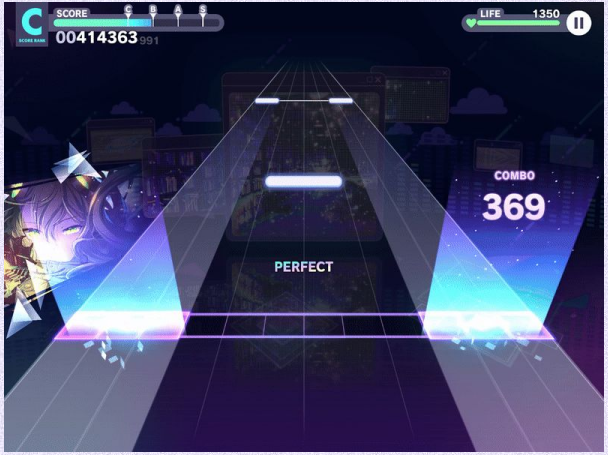
Beatmaps

30

Expert

31

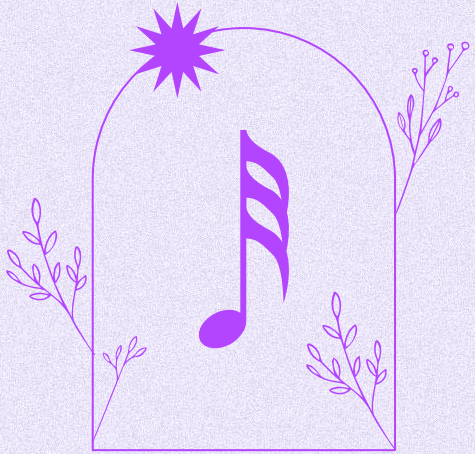
Expert



32

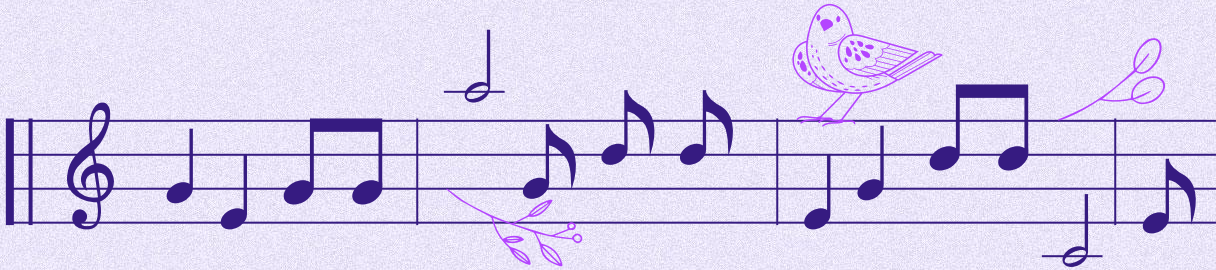
Master

Motivation



- Difficulty levels for rhythm-based video games has not been studied in detail.
- Relevant research mainly applies to mainstream rhythm-based games such as *osu!*, *Stepmania*, and *Dance Dance Revolution*).
- Mobile rhythm games in particular lack much research.

02



Data Set

id_bpm_playback.csv

songID	bpm	playback
1	150	2:03
2	150	1:47
3	185	1:56

606 songs

song_metadata.csv

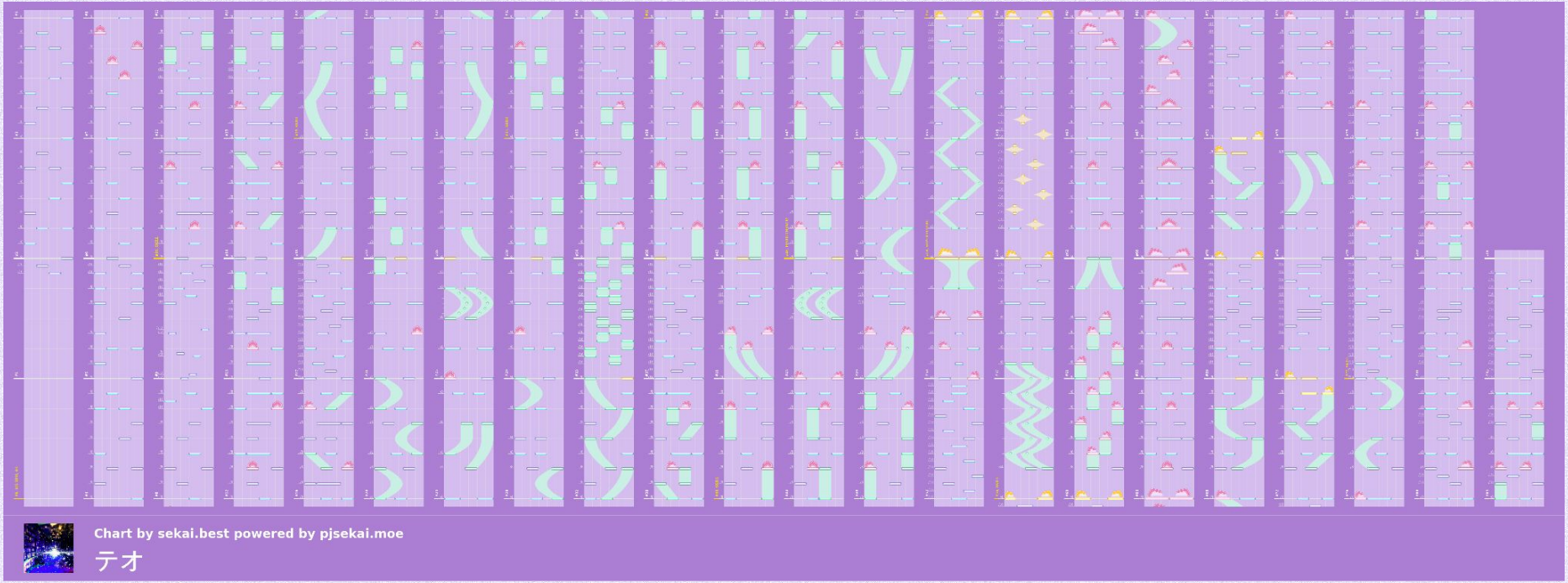
3161 charts

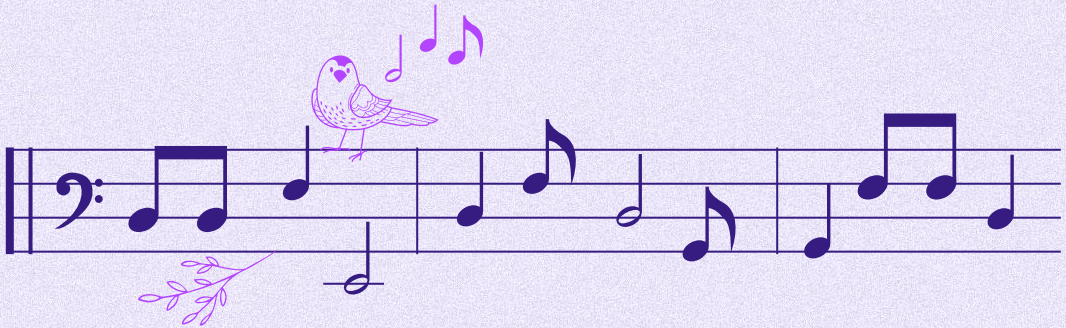
song_id	difficulty	difficulty_level	note_count	title_jpn	title_eng	bpm	playback_time_minutes	playback_time_seconds
1	easy	5	220	Tell Your World	Tell Your World	150	00:02:03	123.0
2	hard	17	635	ロキ	ROKI	150	00:01:47	107.9
3	master	32	1221	テオ	Teo	185	00:01:56	116.0

- 3_easy.svg
- 3_expert.svg
- 3_hard.svg
- 3_master.svg
- 3_normal.svg

songID_difficulty.svg

3119 total files





Data Preprocessing & Feature Extraction

03

Song Features

What are they?

Features that that differ according to each song

Tempo (t)

BPM-derived tempo characteristics

ID	Jacket art	Song name	Producer(s)	Unit(s)	BPM	Duration (game-size)
1		Tell Your World	kz (livetune)	VIRTUAL SINGER	150	2:03
2		Roki	Mikito-P	Leo/need	150	1:47
3		Teo	Omoi Sakurai (lyricist)	Leo/need	185	1:56
6		HIBANA -Reloaded-	DECO*27 Rockwell (arranger) Mes (lyricist)	Leo/need	200	1:35
8		Time Machine	1640mP	Leo/need	126	2:04
10		Happy Synthesizer	EasyPop	MORE MORE JUMP!	127	2:13
11		Viva Happy	Mitchie M	MORE MORE JUMP!	148	1:34
13		Nostalogic	yuukiss	MORE MORE JUMP!	140	2:00

Score Features

What are they?

Features that differ according to each beatmap

Note Density (n)

Number of notes per second

Note Ratio for Beat Layer (l)

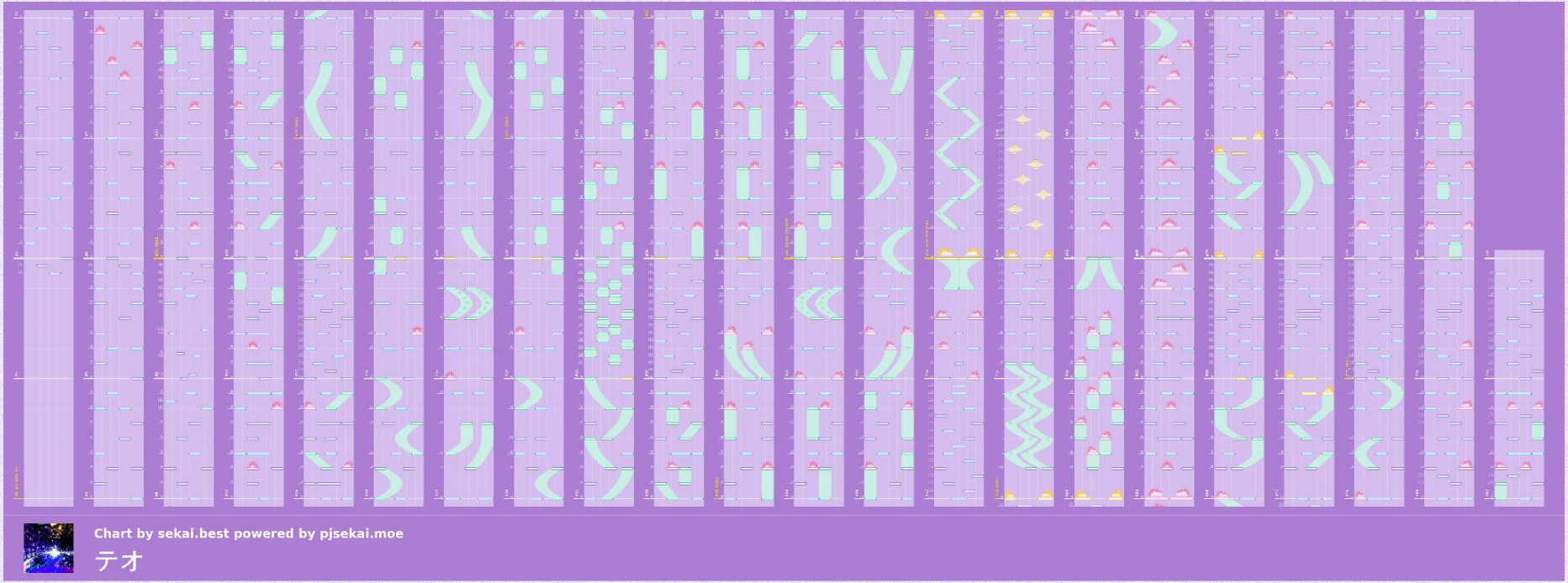
Corresponds to the complexity of a rhythm based on timing division such as ¼ or ⅛ notes

Size (s)

Total chart magnitude (number of total notes)

Algorithm

Step 1: Parsing



Class	Definition
bar-line	For bar counts in chart. About one per musical measure
beat-line	Lines for timings within each bar. Quarter notes in a musical measure
slide, slide-critical	Green slide notes, yellow slide notes
tick-line	Tick lines for note timing. Variable x1 is either 16 or 32

Step 1: Parsing

Hierarchy

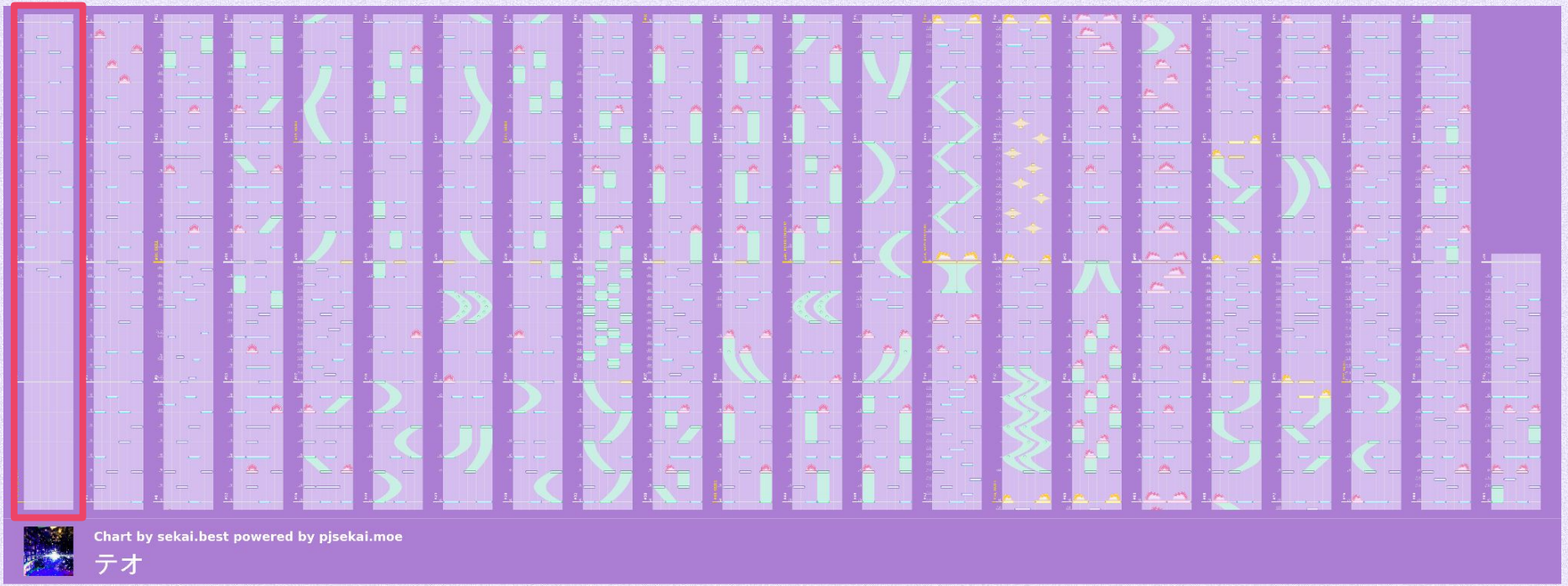
Lanes consist of several **bars** separated by a **#** and **bar lines**. Each **bar** contains **beat lines** and **tick-lines** corresponding to the note and tick timings of a chart.

Important Notes

- Doesn't contain classes for types of notes / playable notes.
- Only one type of playable note parsed: slides
- Bars are stacked on top of each other (different y-values, but the same x-values despite SVG's initial appearance).

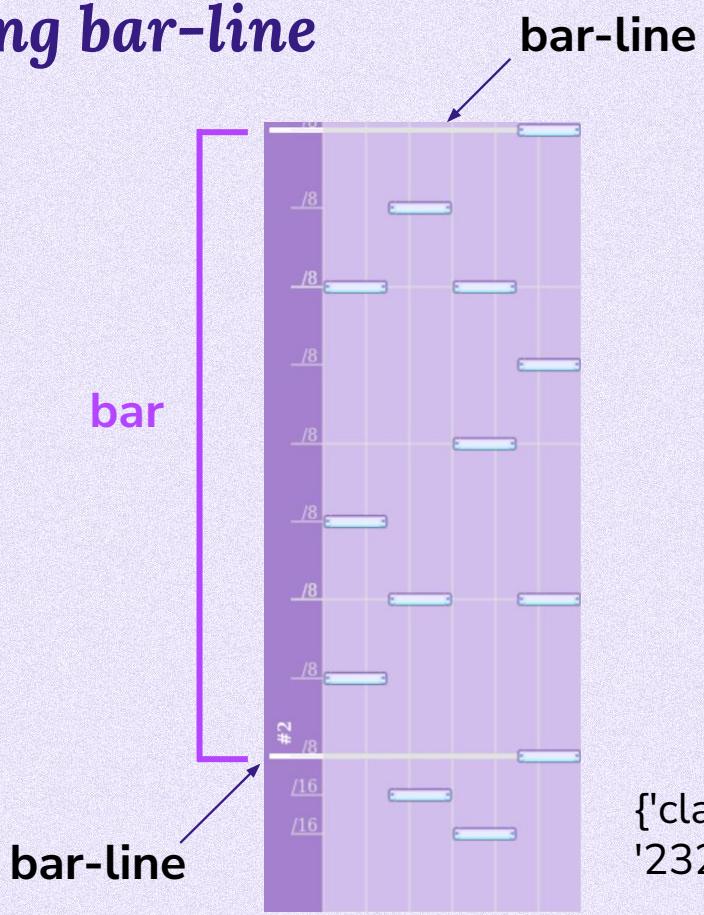
Step 1: Parsing

lane



Step 1a: Handling bar-line

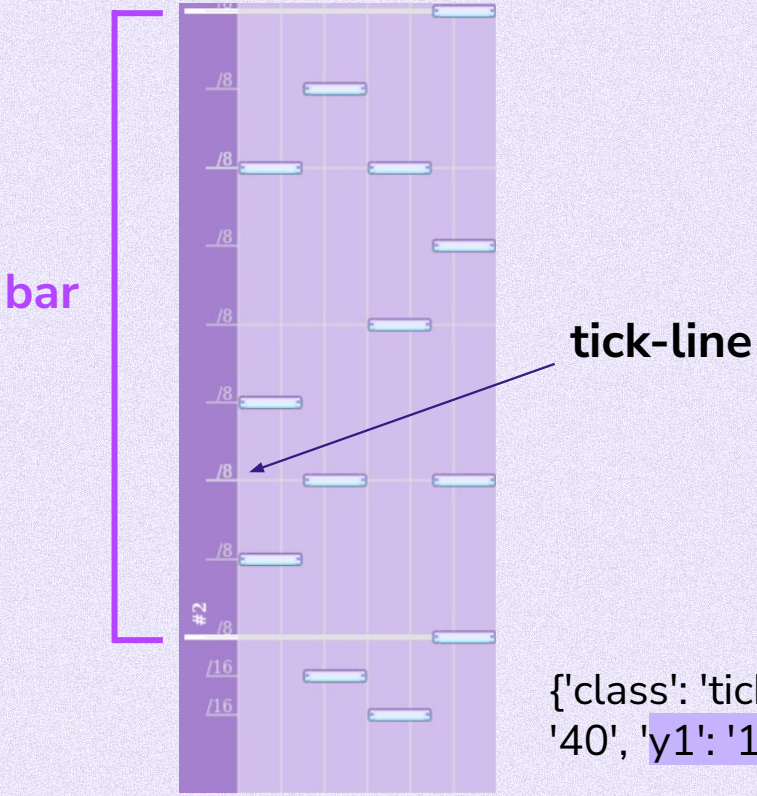
- Remove duplicate bar lines
- Sort and pair components as (top, bottom) bar boundaries
- Each pair represents a bar



{'class': 'beat-line', 'x1': '40', 'x2': '232', 'y1': '2505', 'y2': '2505'}

Step 1b: Handling tick-line

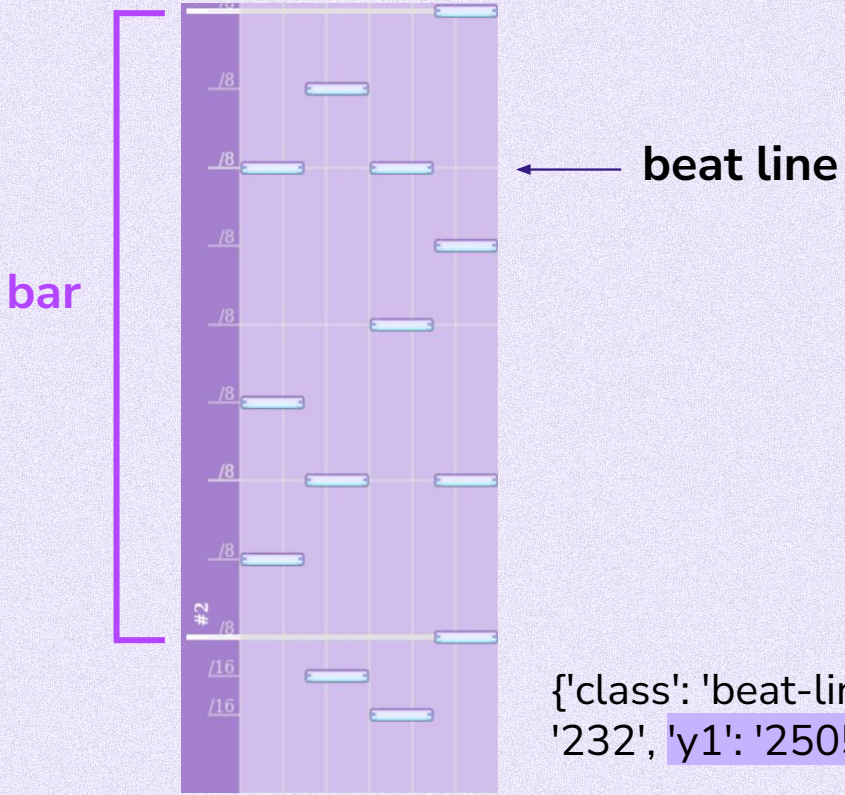
- Assign each note to a bar through the bar index
- For each note we look at its Y position (y1 or y2) and find the bar whose Y-range contains the note



{'class': 'tick-line', 'x1': '16', 'x2': '40', 'y1': '1024', 'y2': '1024'}

Step 1c: Handling beat-line

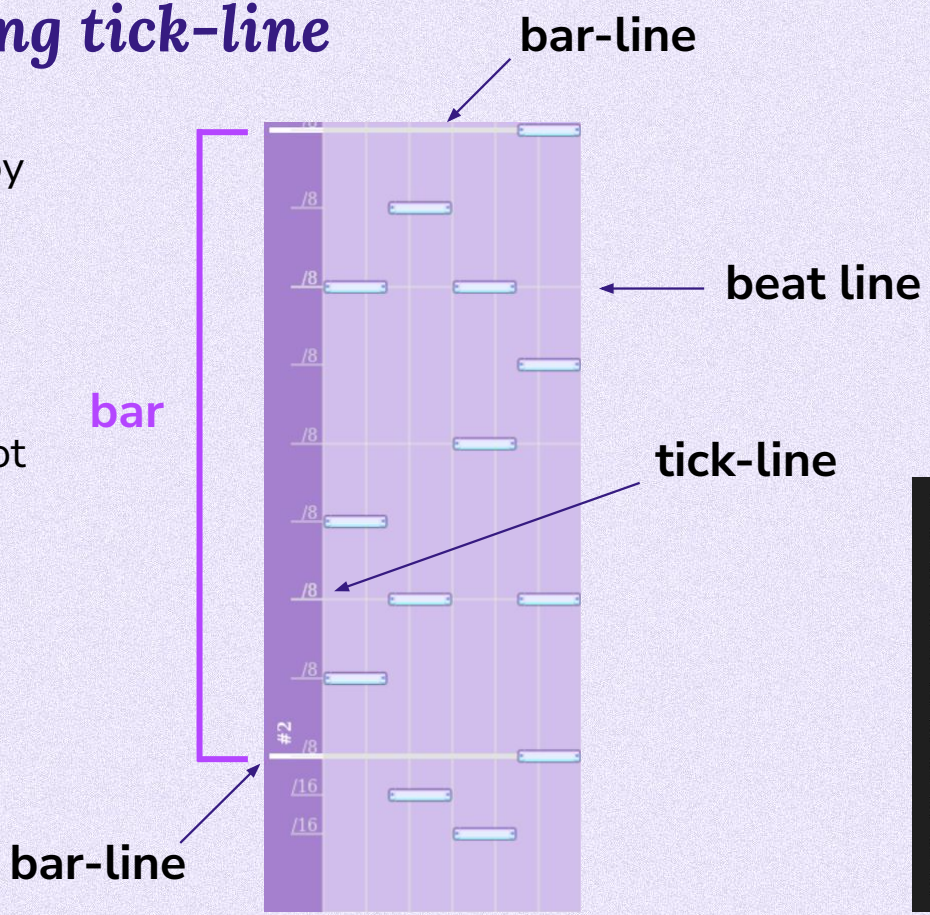
- We map beat lines to bars using bar midpoints



{'class': 'beat-line', 'x1': '40', 'x2': '232', 'y1': '2505', 'y2': '2505'}

Step 1d: Handling tick-line

- Assign each tick by calculating its fractional, relative positions within a beat segment.
- If beat lines are not available, we fallback to linear mapping.




```
songID,bpm,playback
1,150,2:03
2,150,1:47
3,185,1:56
6,200,1:35
8,126,2:04
10,127,2:13
11,148,1:34
13,140,2:00
15,130,1:46
18,185,2:08
```

Step 2: Feature Extraction

Create a synthetic beat grid (normalized timeline)

Initialization

Compute tempo features using BPM

Tempo (t)

Keep track of total notes in a chart

Total Notes/Scale (s)

Beat Layers (l)

Compute q-layer note ratios using the synthetic beat grid and timing divisions

Note Density (n)

Compute notes (ticks) per second using note times, song duration, etc.

Step 3: PCA and Clustering

Standardize Features

First, we standardize feature groups to reduce noise, scaling appropriately.

Apply PCA

We apply PCA per feature group. EX: PCA for note density $n_l, n_m, n_{mu}, n_{var}$ gives PC1: 87.6%, PC2: 9.9%, etc.

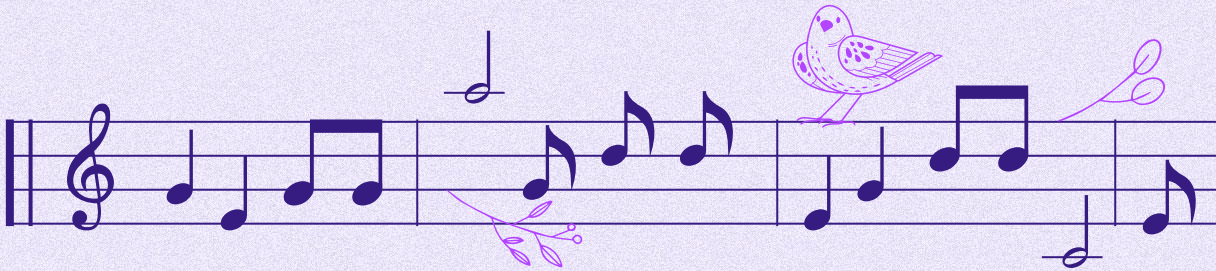
Combine PCs

We combine our selected PCs into a feature matrix for clustering.

Clustering

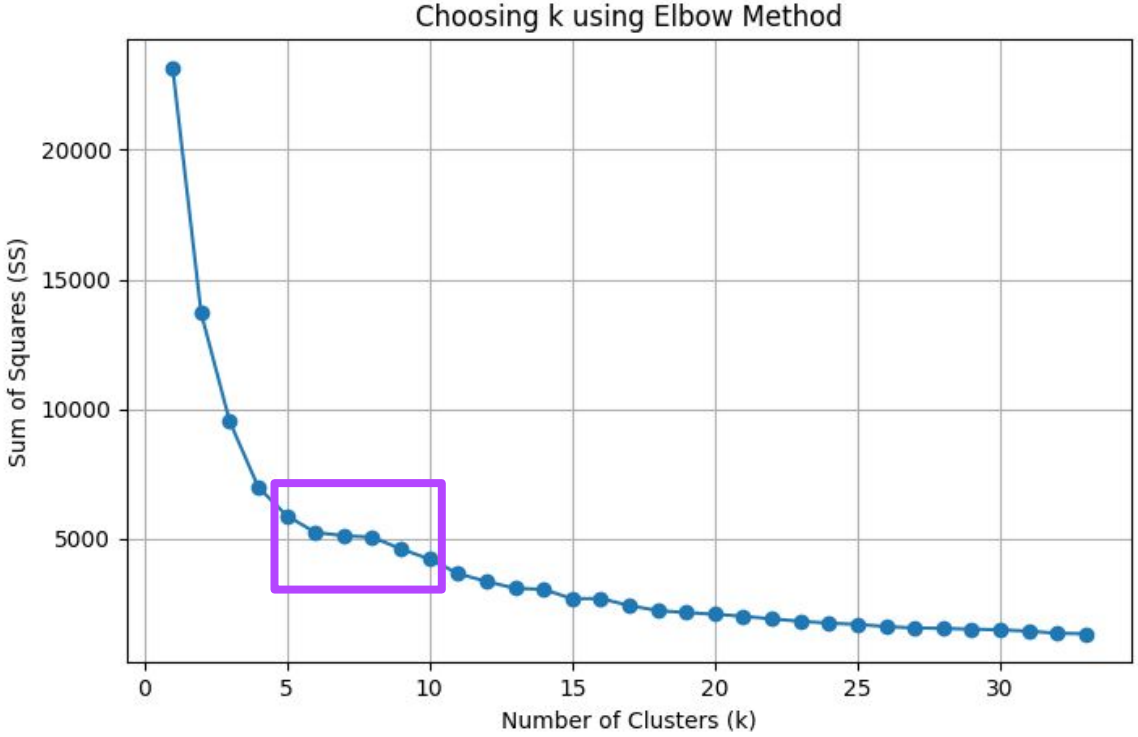
Lastly, we run KMeans clustering.

04

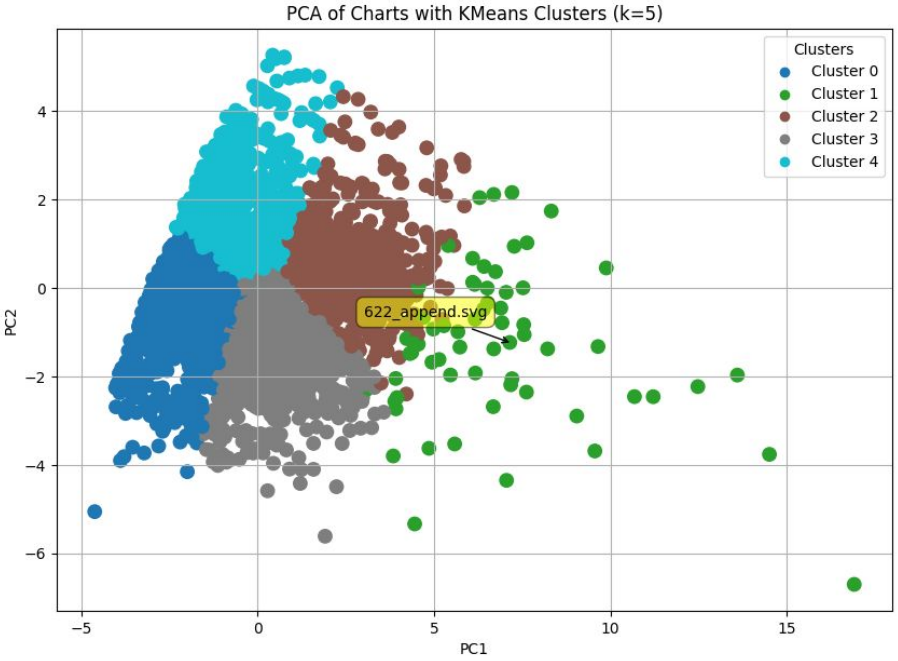


Data Analysis

Elbow Plot


Choosing k using Elbow Method
Sum of Squares (SS)
20000
15000
10000
5000
0
5
10
15
20
25
30
Number of Clusters (k)

$k = 5$

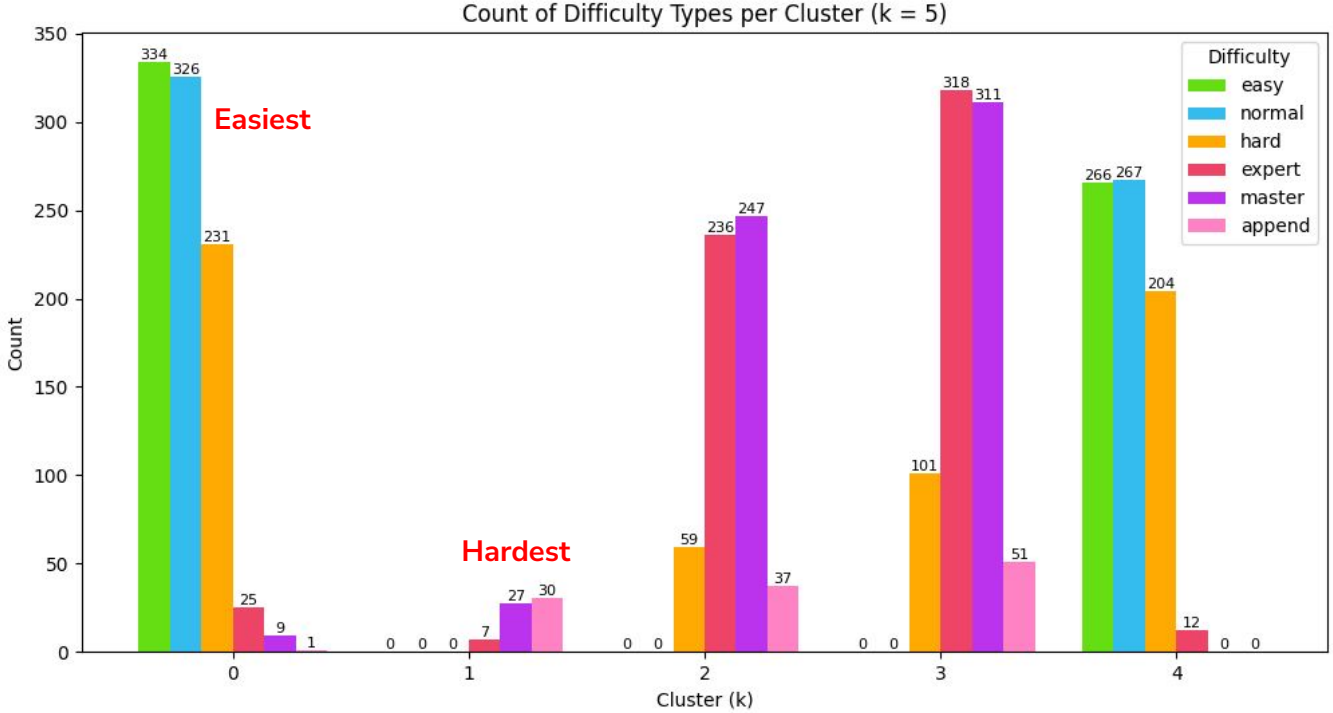

PCA of Charts with KMeans Clusters (k=5)
Clusters
Cluster 0
Cluster 1
Cluster 2
Cluster 3
Cluster 4
622_append.svg
PC1
PC2
−5
0
5
10
15
4
2
0
−2
−4
−6

KMeans Clustering with k=5

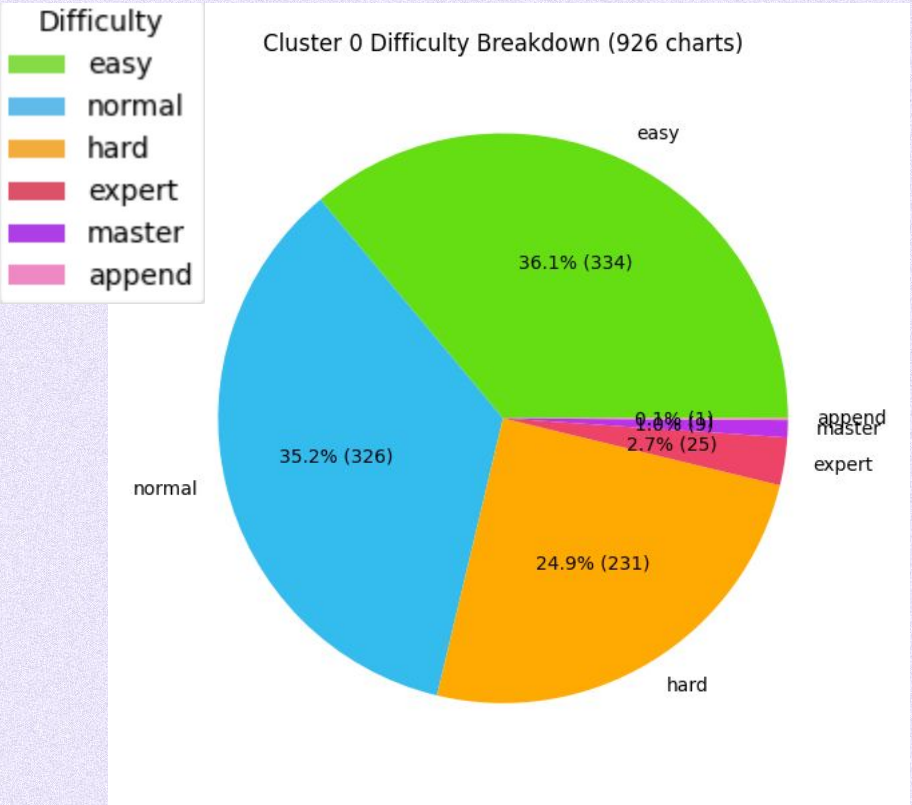
Total Chart Count:
3099

Difficulty Breakdown:

- **Easy:** 600 (19.4%)
- **Normal:** 593 (19.1%)
- **Hard:** 595 (19.2%)
- **Expert:** 598 (19.3%)
- **Master:** 594 (19.2%)
- **Append:** 119 (3.8%)



Cluster 0



Cluster 0

Easy
334 charts (36.1% of cluster)

Normal
326 charts (35.2% cluster)

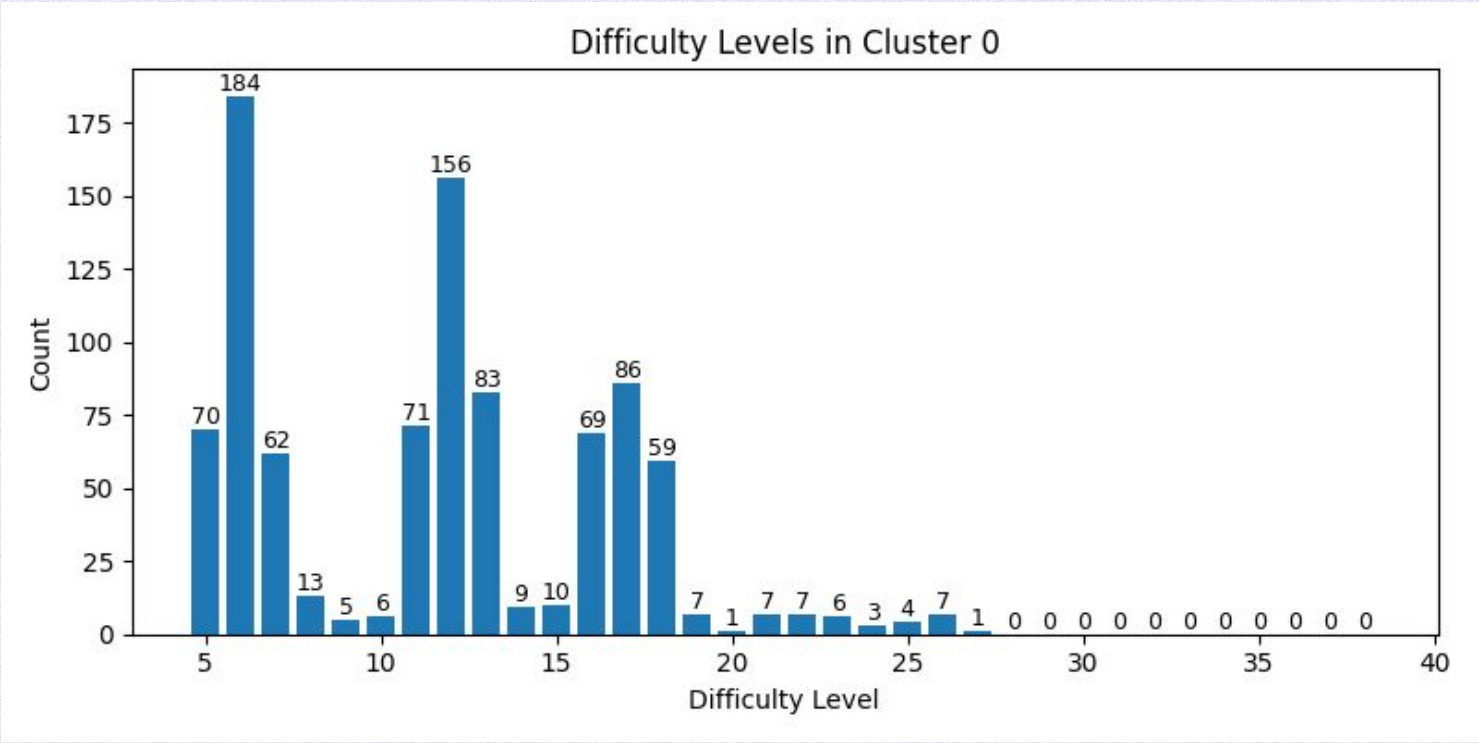
Hard
231 charts (24.9% of cluster)

Expert
25 charts (2.7% of cluster)

Master
9 charts (1.0% of cluster)

Append
1 charts (0.1% of cluster)

Cluster 0



Top 5 Difficulty Lvls

Level 6
184 charts (5.9%)

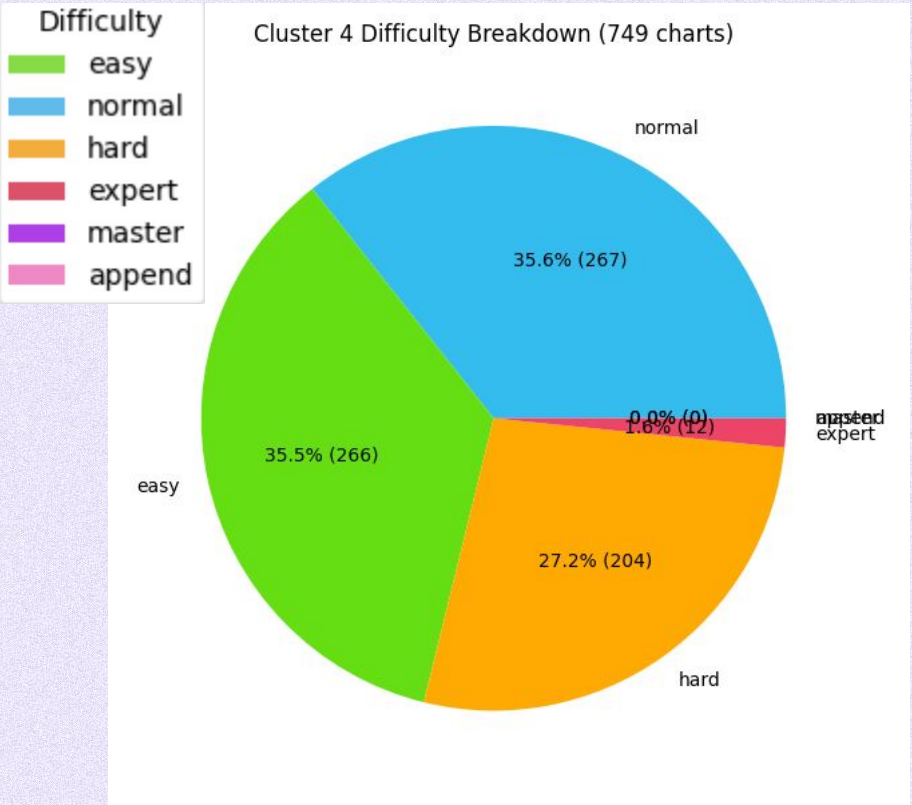
Level 12
156 charts (5.0%)

Level 17
86 charts (2.8%)

Level 13
83 charts (2.7%)

Level 11
71 charts (2.3%)

Cluster 4



Cluster 4

Easy
266 charts (35.5% of cluster)

Normal
267 charts (35.6% of cluster)

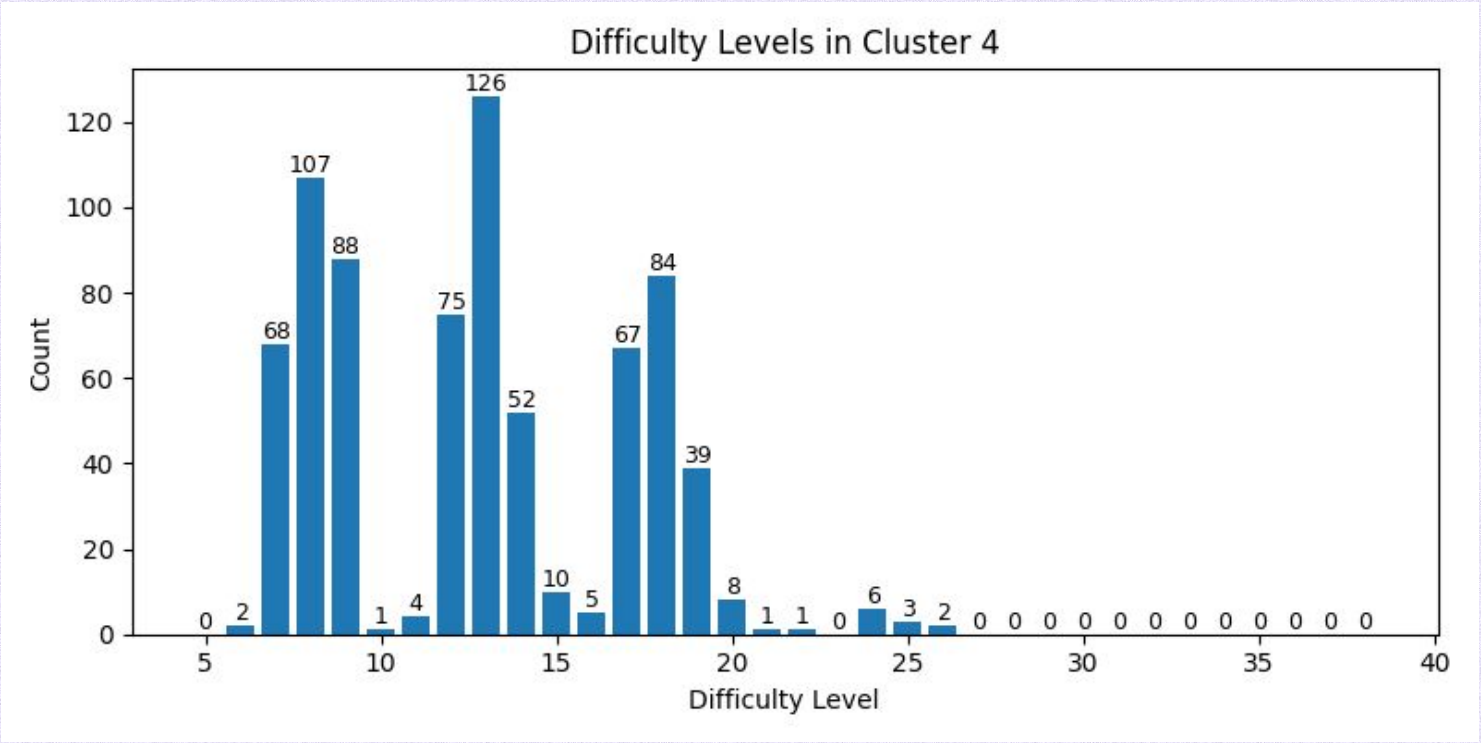
Hard
204 charts (27.2% of cluster)

Expert
12 charts (1.6% of cluster)

Master
0 charts (0.0% of cluster)

Append
0 charts (0.0% of cluster)

Cluster 4



Top 5 Difficulty Lvls

Level 13
126 charts (4.1%)

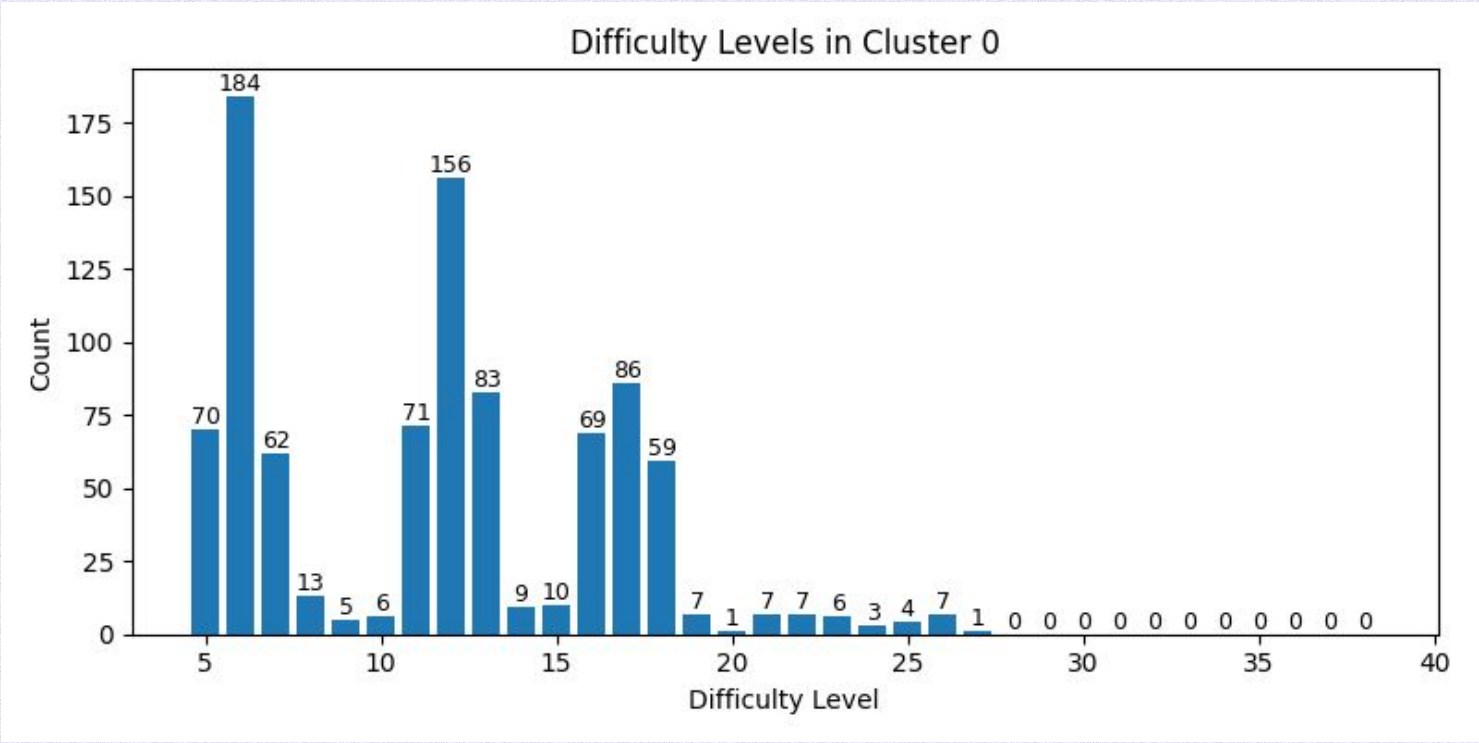
Level 8
107 charts (3.5%)

Level 9
88 charts (2.8%)

Level 18
84 charts (2.7%)

Level 12
75 charts (2.4%)

Cluster 0



Top 5 Difficulty Lvls

Level 6
184 charts (5.9%)

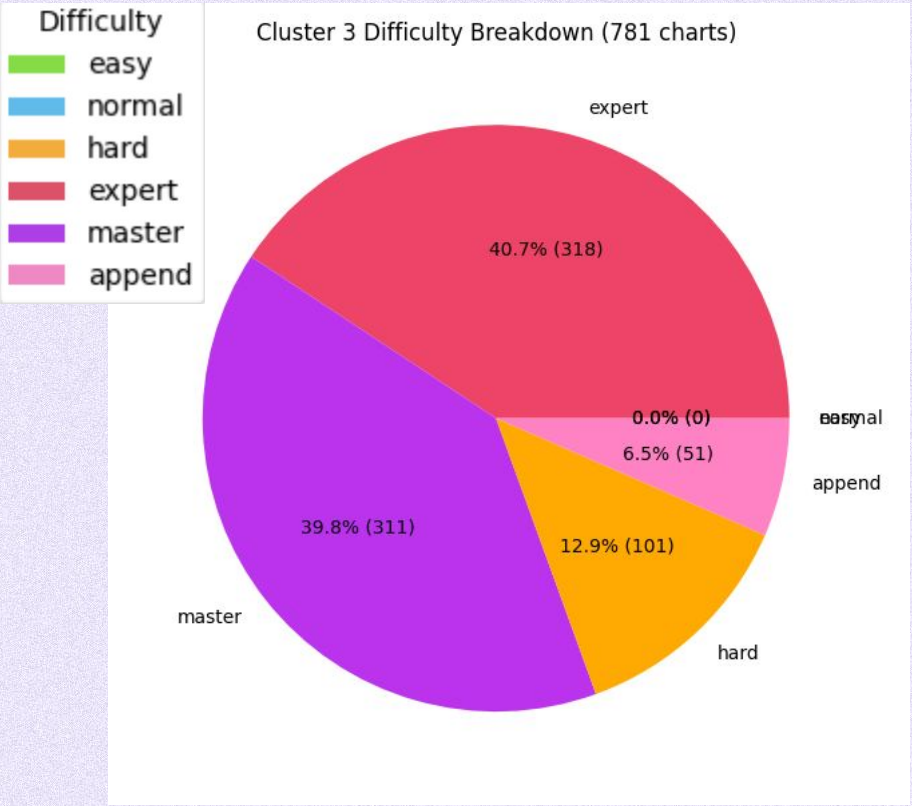
Level 12
156 charts (5.0%)

Level 17
86 charts (2.8%)

Level 13
83 charts (2.7%)

Level 11
71 charts (2.3%)

Cluster 3



Cluster 3

Expert
318 charts (40.7% of cluster)

Master
311 charts (39.8% of cluster)

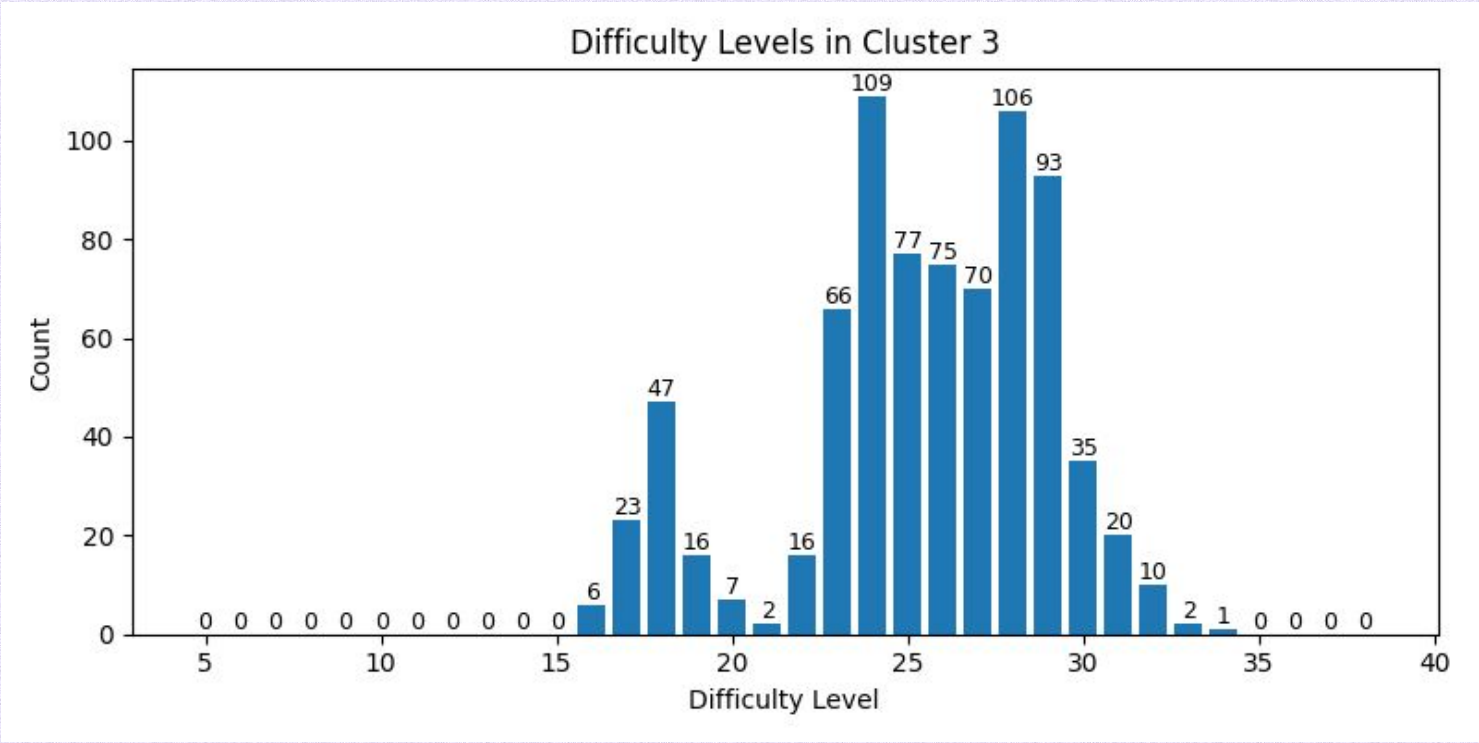
Hard
101 charts (12.9% of cluster)

Append
51 charts (6.5% of cluster)

Easy
0 charts (0.0% of cluster)

Normal
0 charts (0.0% of cluster)

Cluster 3



Top 5 Difficulty Lvls

Level 24
109 charts (3.5%)

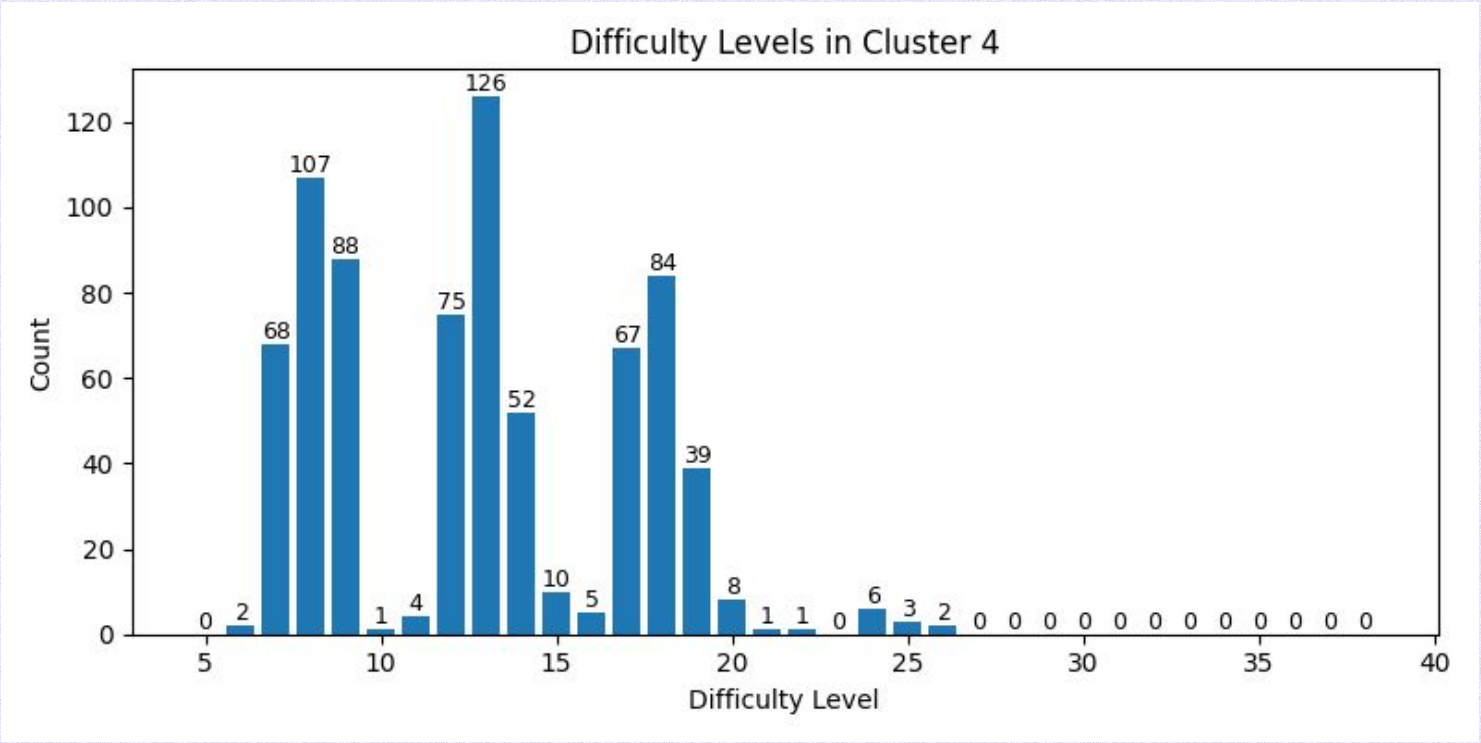
Level 28
106 charts (3.4%)

Level 29
93 charts (3.0%)

Level 25
77 charts (2.5%)

Level 26
75 charts (2.4%)

Cluster 4



Top 5 Difficulty Lvls

Level 13
126 charts (4.1%)

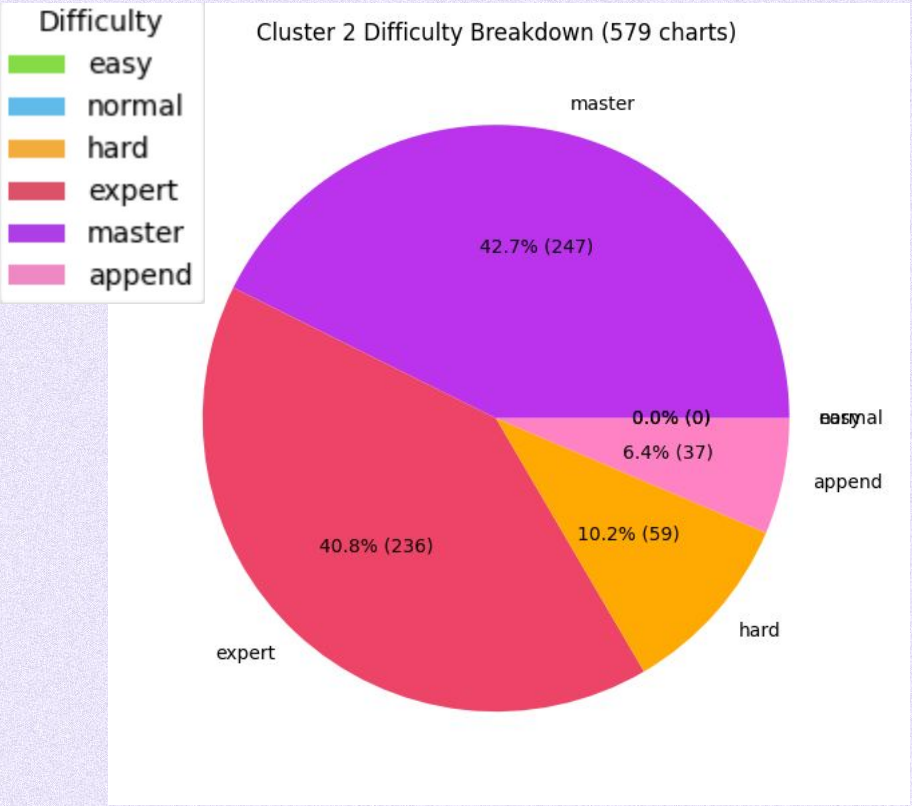
Level 8
107 charts (3.5%)

Level 9
88 charts (2.8%)

Level 18
84 charts (2.7%)

Level 12
75 charts (2.4%)

Cluster 2



Cluster 2

Master
247 charts (42.7% of cluster)

Expert
236 charts (40.8% of cluster)

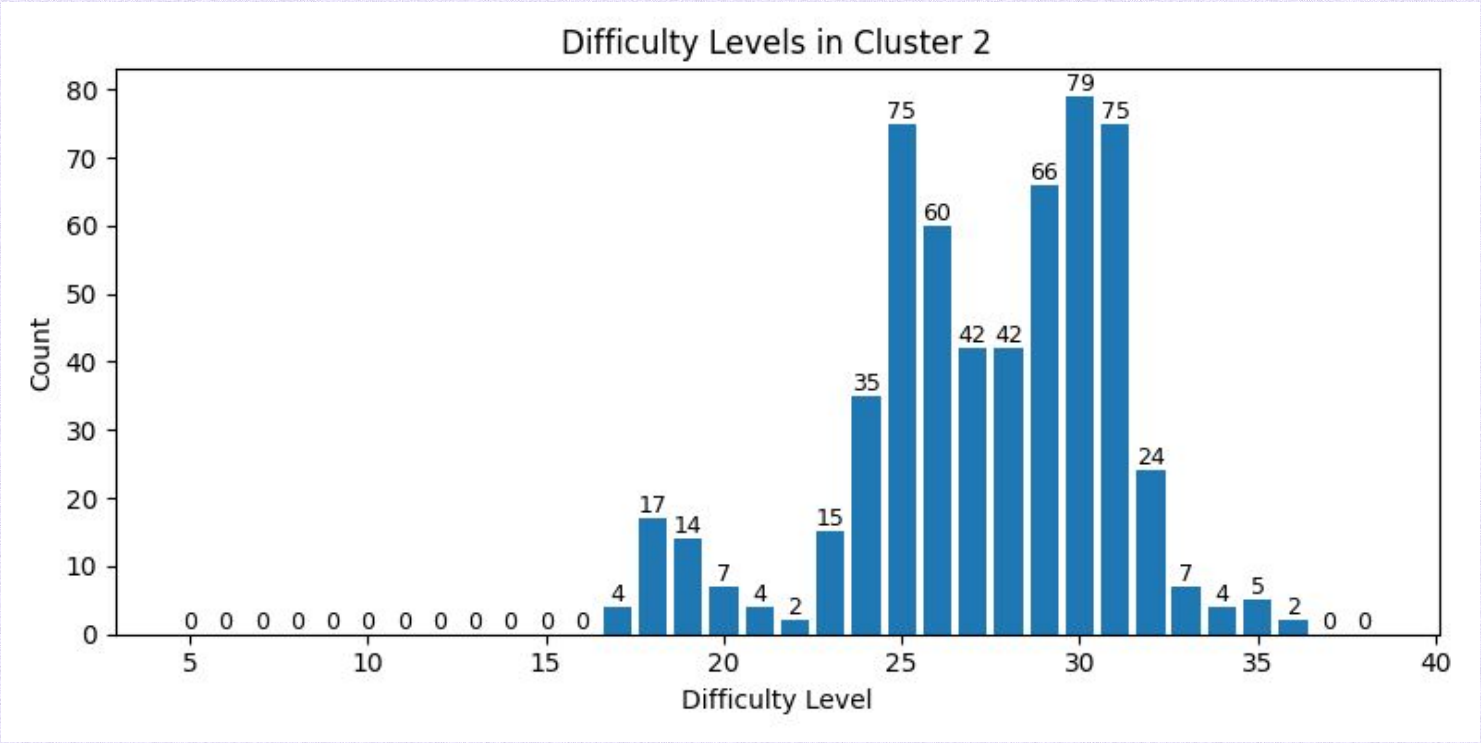
Hard
59 charts (10.2% of cluster)

Append
37 charts (6.4% of cluster)

Easy
0 charts (0.0% of cluster)

Normal
0 charts (0.0% of cluster)

Cluster 2



Top 5 Difficulty Lvls

Level 30
79 charts (2.5%)

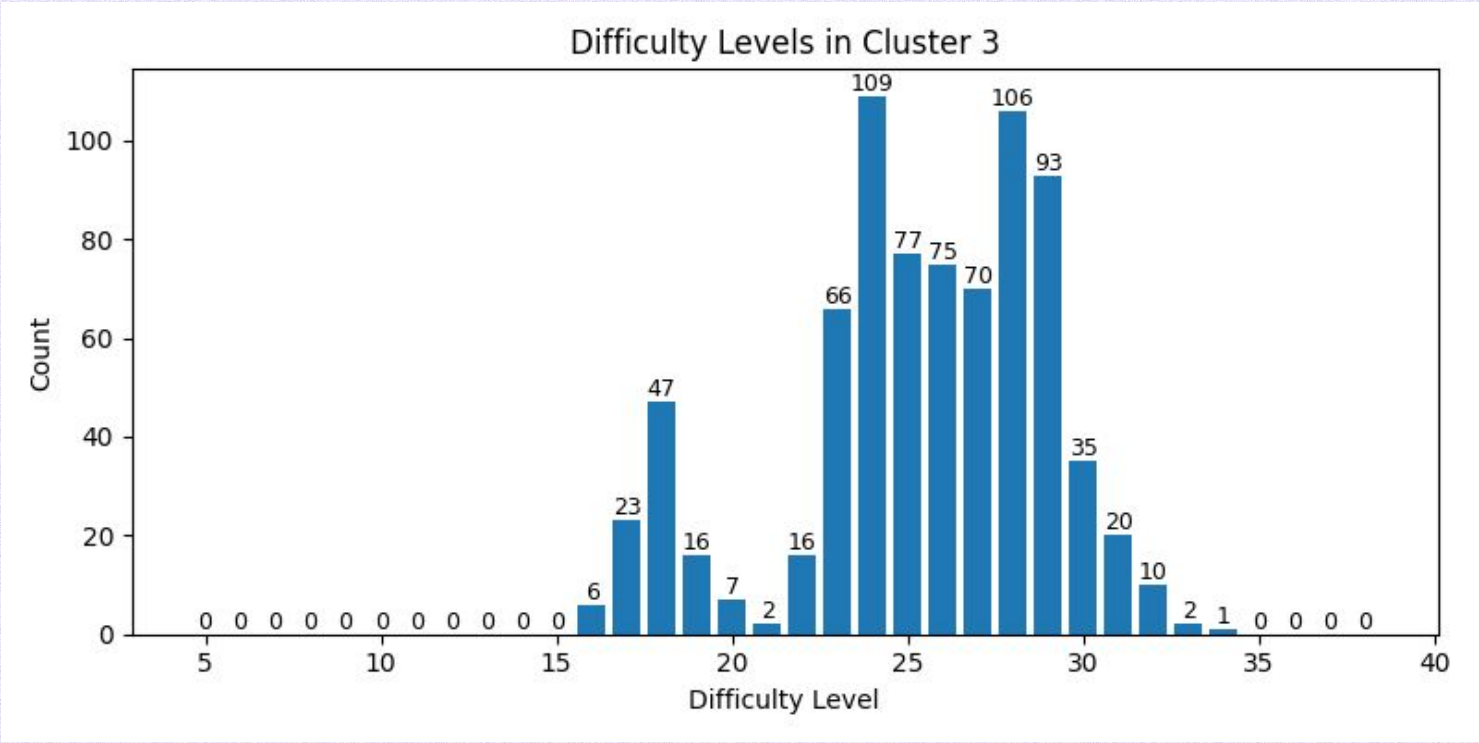
Level 25
75 charts (2.4%)

Level 31
75 charts (2.4%)

Level 29
66 charts (2.1%)

Level 26
60 charts (1.9%)

Cluster 3



Top 5 Difficulty Lvls

Level 24
109 charts (3.5%)

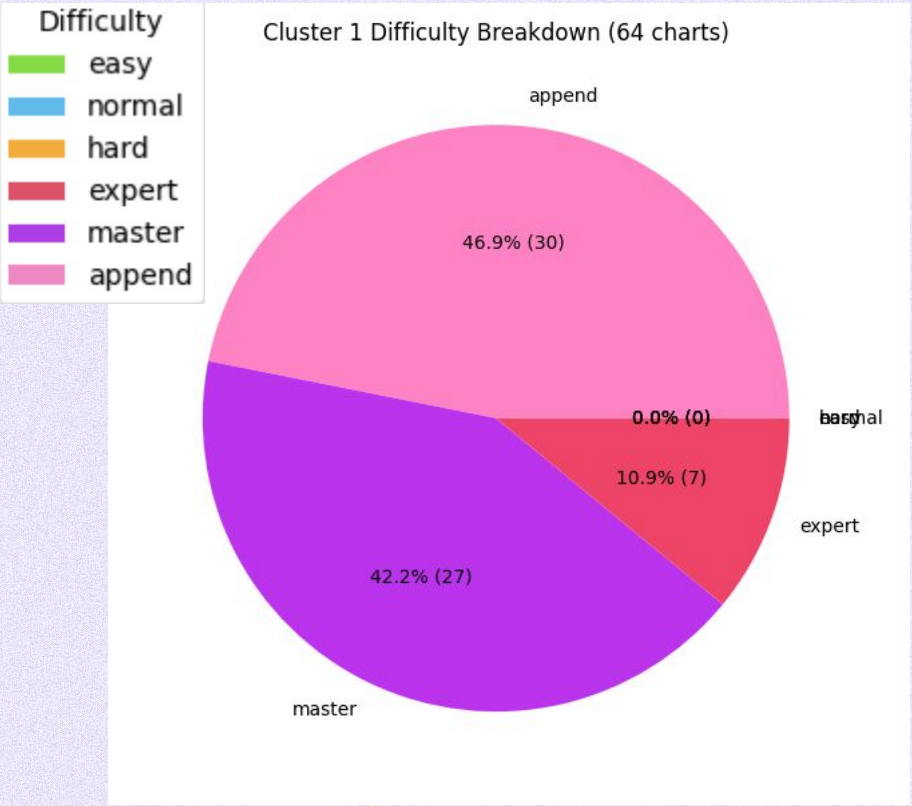
Level 28
106 charts (3.4%)

Level 29
93 charts (3.0%)

Level 25
77 charts (2.5%)

Level 26
75 charts (2.4%)

Cluster 1



Level 4 / Cluster 1

Append
30 charts (46.9% of cluster)

Master
27 charts (42.2% of cluster)

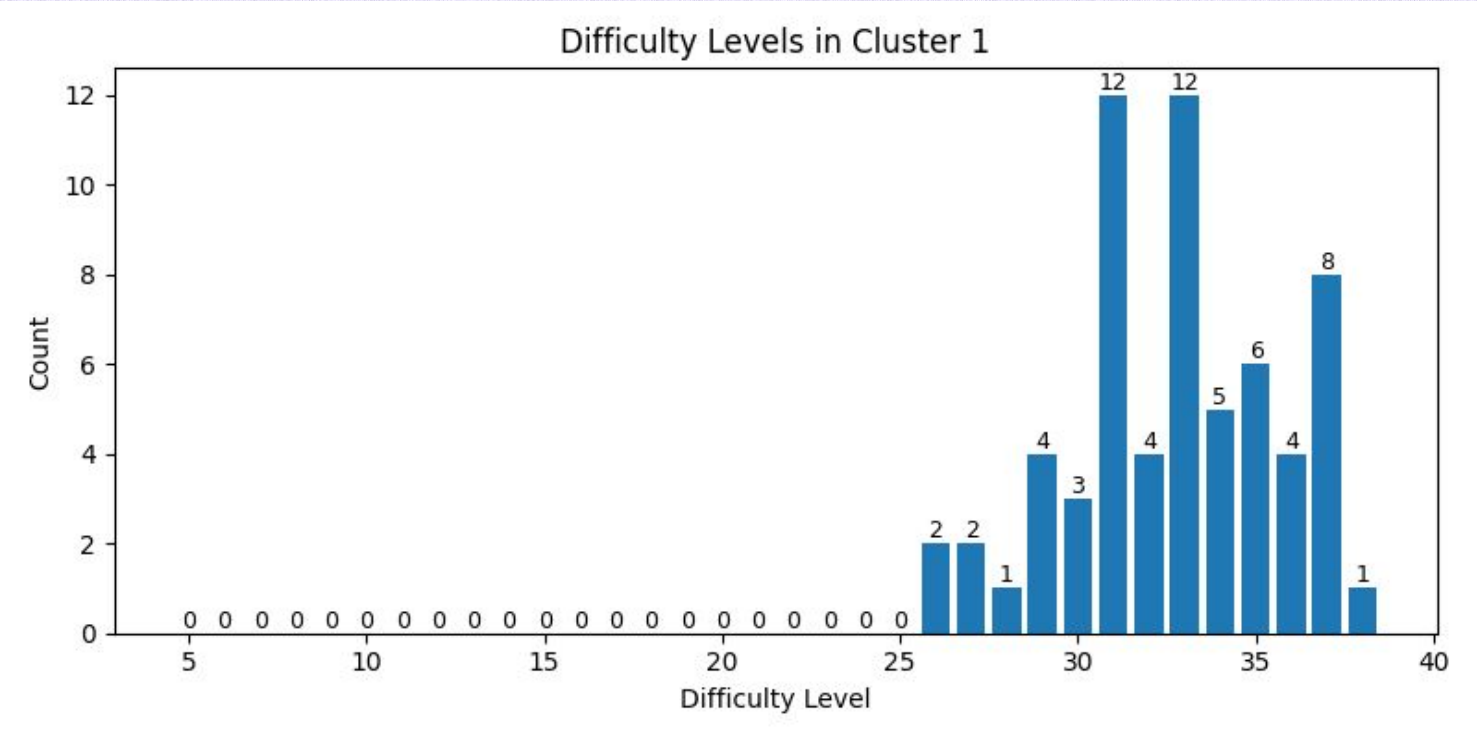
Expert
7 charts (10.9% of cluster)

Easy
0 charts (0.0% of cluster)

Normal
0 charts (0.0% of cluster)

Hard
0 charts (0.0% of cluster)

Cluster 1



Top 5 Difficulty Lvls

Level 33
12 charts (18.6%)

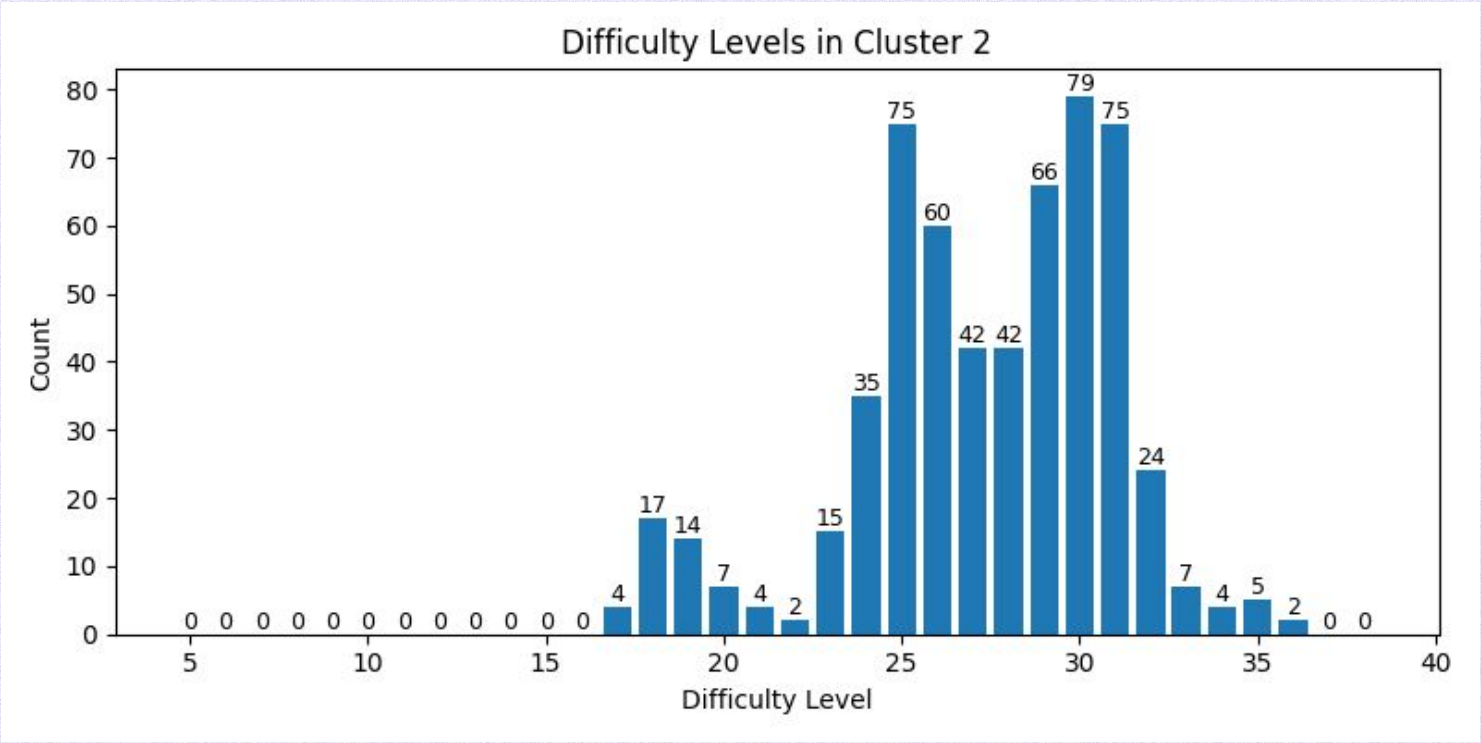
Level 31
12 charts (18.6%)

Level 37
8 charts (12.5%)

Level 35
6 charts (9.4%)

Level 34
5 charts (7.8%)

Cluster 2



Top 5 Difficulty Lvls

Level 30
79 charts (2.5%)

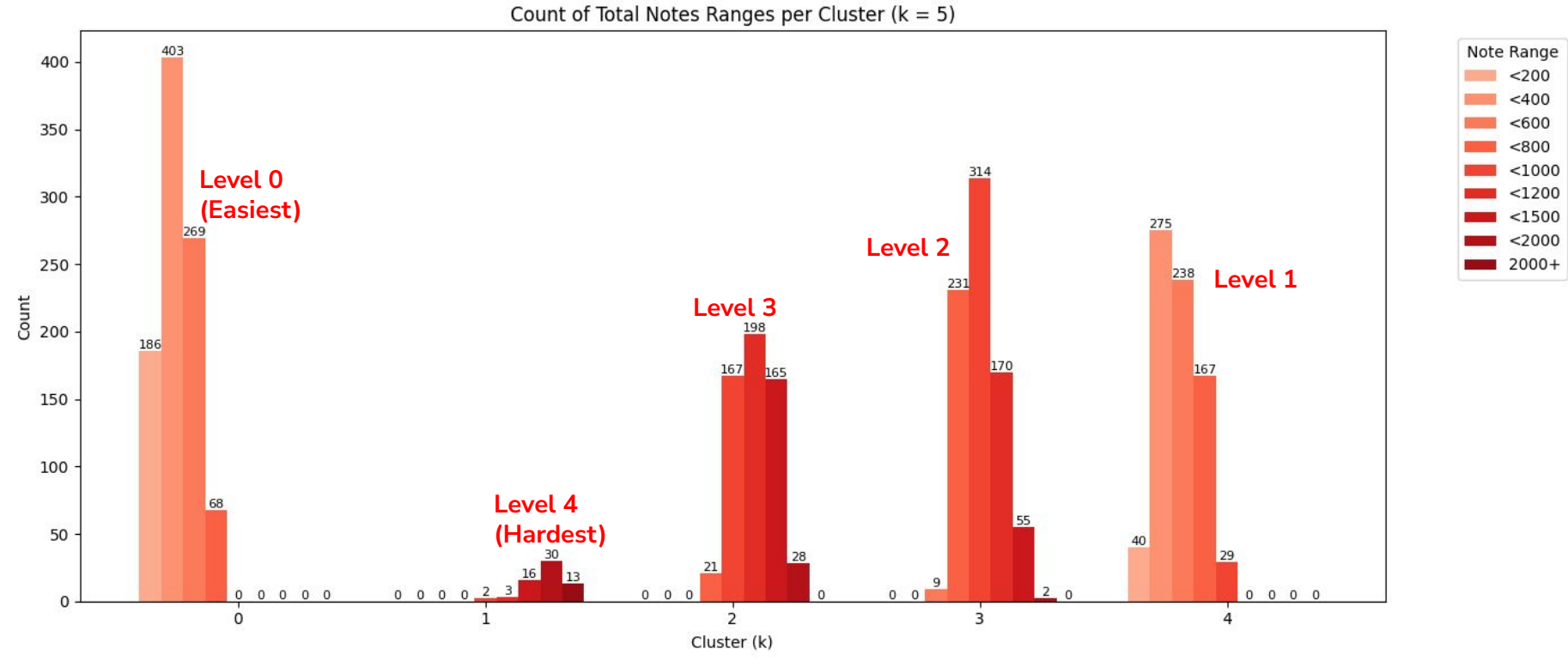
Level 25
75 charts (2.4%)

Level 31
75 charts (2.4%)

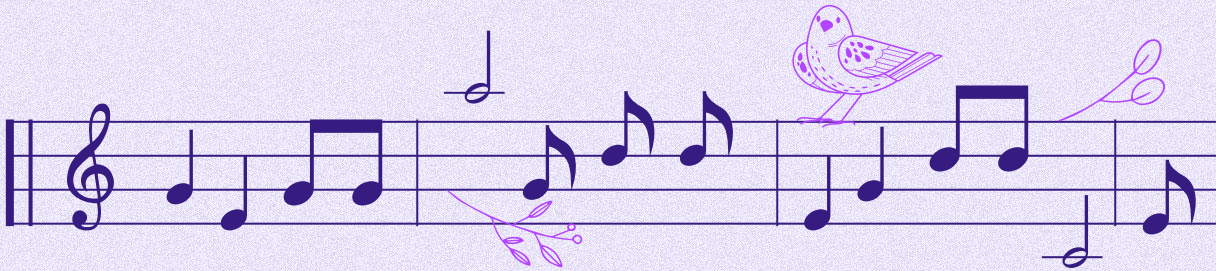
Level 29
66 charts (2.1%)

Level 26
60 charts (1.9%)

Note Count

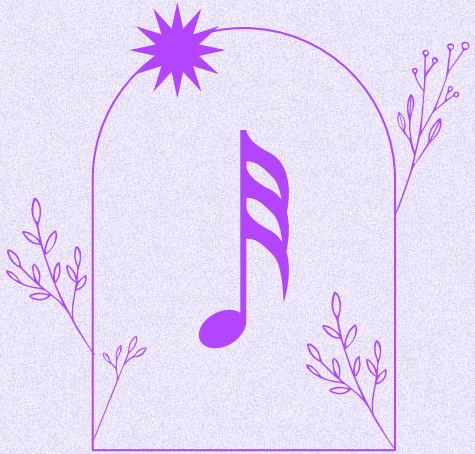

Count of Total Notes Ranges per Cluster (k = 5)
Count
Cluster (k)
Note Range
<200
<400
<600
<800
<1000
<1200
<1500
<2000
2000+
Level 0
(Easiest)
Level 1
Level 2
Level 3
Level 4
(Hardest)
186
403
269
68
40
275
238
167
29
231
314
170
55
2
9
167
198
165
28
21
2
3
16
30
13

05

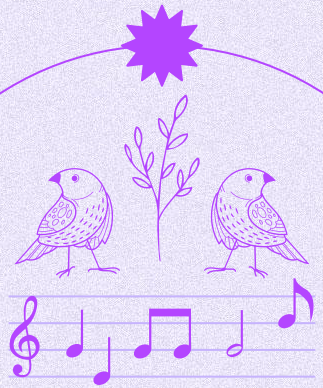


Conclusion

Conclusion



- Data analysis confirms that the designated features used for feature selection differed through the difficulty levels.
- High difficulty charts are mainly characterized by a large number of notes and a high note density, how fast a song is, and its beat layer ratios.
- Comparing our resulting results against the actual difficulty levels support our results and analysis.



Thank you!

Questions?

References: Tsujino, Y., Yamanishi, R., & Yamashita, Y. (2019). Characteristics study of dance-charts on rhythm-based video games. In *Proceedings of the 2019 IEEE Conference on Computational Intelligence and Games (CIG)* (pp. 1–8). IEEE.