

Projektbericht: Spotify Analyse

1st Eric Kaufmann
Jena, Germany
eric.kaufmann@uni-jena.de

2nd Maria Gogolev
Jena, Germany
maria.gogolev@uni-jena.de

Abstract—This document is a model and instructions for \LaTeX . This and the `IEEEtran.cls` file define the components of your paper [title, text, heads, etc.]. ***CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

This document is a model and instructions for \LaTeX . Please observe the conference page limits.

II. DATENSATZ

Der Datensatz *Spotify 1.2M+ Songs* [2] enthält Audio-Eigenschaften von über 1.2 Millionen Liedern, die mit der Spotify API abgerufen wurden.

| Attribut | Beschreibung |
|------------------|--|
| id | Eindeutige Identifikation des Lieds. |
| name | Der Name des Lieds. |
| album | Der Name des Albums, zu dem das Lied gehört. |
| album_id | Eindeutige Identifikation des Albums. |
| artists | Der Name des Künstlers, der das Lied erstellt hat. |
| artist_ids | Eindeutige Identifikation des Künstlers. |
| track_number | Die Liednummer auf dem Album. |
| disc_number | Die Disc-Nummer des Albums, auf dem das Lied ist. |
| explicit | Gibt an, ob das Lied expliziten Inhalt enthält. |
| danceability | Eignung eines Liedes zum Tanzen. |
| energy | Eine Messung der Intensität und Aktivität des Lieds. |
| key | Der musikalische Schlüssel des Lieds. |
| loudness | Der Gesamtlautstärke des Lieds in Dezibel (dB). |
| mode | Lied Dur- oder Moll-Tonart geschrieben |
| speechiness | Der Sprachanteils eines Lieds. |
| acousticness | Eine Schätzung, wie akustisch das Lied ist. |
| instrumentalness | Der Anteil des Lieds der instrumentale Musik enthält. |
| liveness | Wahrscheinlichkeit, dass ein Track live ist. |
| valence | Empfundene Positivität/Negativität. |
| tempo | Das Tempo des Lieds, gemessen in BPM. |
| duration_ms | Die Dauer des Lieds in Millisekunden. |
| time_signature | Angabe wie viele Schläge in jedem Takt enthalten sind. |
| year | Das Erscheinungsjahr des Lieds. |
| release_date | Das Datum, an dem das Lied veröffentlicht wurde. |

TABLE I
ATTRIBUTBESCHREIBUNGEN DES MUSIK-DATENSATZES

Der zweite Datensatz, *Spotify Playlists* [1] enthält Informationen zu User-erstellten Playlists. Jeder Playlist ist eine eindeutige User-ID zugeordnet. Jede Spalte enthält einen Song, der durch dessen Namen und zugeordneten Künstler identifiziert wird, und die jeweilige Playlist, die den Song enthält.

Allgemeine Fakten zu den Datensätzen: Struktur/Schema, Erklärung der musikalischen Attribute

III. VORVERARBEITUNG

Hauptsächlich Merging von Playlist-Daten mit Song-Daten

IV. ANALYSE

A. Attributes over Time

Aus den vorliegenden Daten geht hervor, dass sich die Musik im Laufe der Jahre in Bezug auf verschiedene Attribute deutlich verändert hat, und gewisse Trends ersichtlich sind. Die Analyse bezieht sich auf Abbildung 1. Im Folgenden werden nicht alle Attribute aufgegriffen, da bei ihnen Statistiken wie Standardabweichung und Durchschnittswert weniger aussagekräftig sind (bspw. "key" also der Notenschlüssel).

- **Danceability** Beginnend mit der Danceability sehen wir, dass der Mindestwert und der Maximalwert über die Jahre leicht auseinander gehen. Nach einem Absinken der Danceability in den 1940er Jahren zeigt der Durchschnittswert wieder einen leichten Aufwärtstrend. Die Standardabweichung bleibt relativ stabil über die Zeit.
- **Energy** Auch bei der Energie ist ein ähnlicher Trend zu beobachten: Der Minimalwert bleibt relativ konstant, der Maximalwert steigt stetig an, der Durchschnittswert zeigt einen Aufwärtstrend und die Standardabweichung nimmt im Laufe der Zeit zu. Dies deutet darauf hin, dass die Musik im Laufe der Jahre energiegeladener geworden ist, es gleichzeitig aber mehr Diversität in diesem Attribut gibt.
- **Loudness** Bei der Lautstärke ist ein sinkender Minimalwert und eine starke Zunahme des Maximalwerts zu verzeichnen, während der Durchschnitt relativ konstant bei ca. 4 Minuten und 6 Sekunden mit nur geringer Standardabweichung bleibt.
- **Mode** Der Durchschnittswert des Modus bleibt über die gesamte Zeit innerhalb von 0,5, was aussagt, dass stets mehr Musik in Dur produziert wird als in Moll, wobei ein leichter Abstieg des Durchschnittswertes in den letzten Jahren zu verzeichnen ist. Es wurde also in den letzten Jahren mehr in Moll produziert. Die Standardabweichung bleibt konstant hoch, der Modus nur zwei Werte annehmen kann.
- **Speechiness** Das Minimum des Sprachanteils bleibt nahe bei 0, da Instrumentalmusik früher wie auch heute relevant ist. Der anfangs geringe Maximalwert entwickelt sich hingegen nach oben. Durchschnittswert und Standardabweichung bleiben gering, was darauf hindeutet, dass der Großteil der Musik nach wie vor instrumental ist. Die Sprachlichkeit zeigt einen konstanten Rückgang der

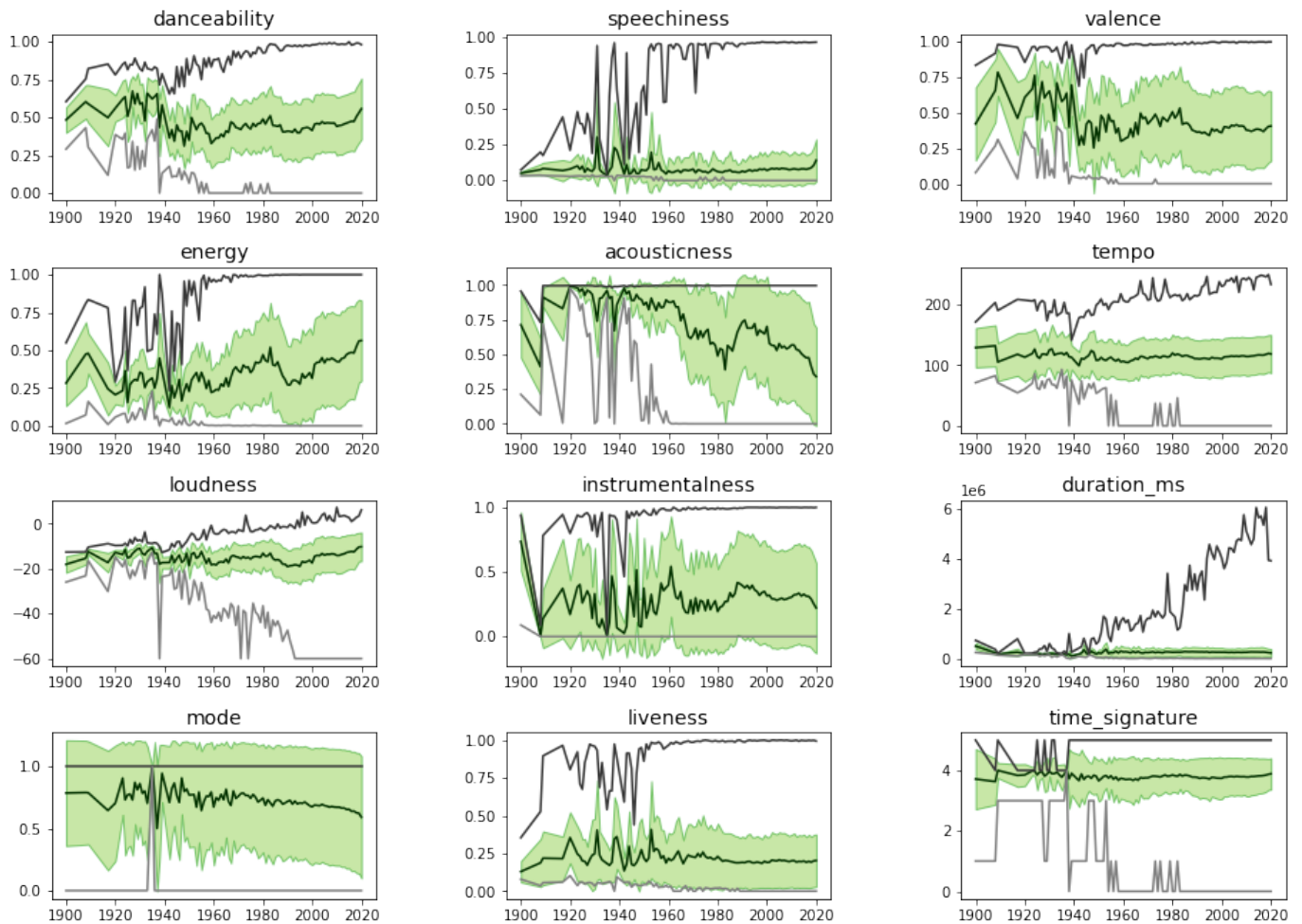


Fig. 1. Entwicklung der Musik-Attribute von 1900 bis 2020. Die in der Mitte verlaufende Linie ist der Durchschnittswert, die obere Linie das Maximum, die untere Linie das Minimum und der grüne Bereich ist die Standardabweichung im gegebenen Jahr.

Minimal-, Maximal- und Durchschnittswerte, was darauf hindeutet, dass die Musik im Laufe der Jahre weniger sprachorientiert geworden ist.

- **Acousticness** Die Akustizität zeigt einen konstanten Rückgang der Durchschnittswerte, was darauf hindeutet, dass die Musik im Laufe der Jahre elektronischer geworden ist. Die Standardabweichung für das Attribut "Akustizität" nimmt im Laufe der Zeit zu, was darauf hindeutet, dass die Verteilung der Akustizitätswerte im Laufe der Zeit immer vielfältiger und unterschiedlicher wird. Elektronische Musik ersetzt akustische Musik also nicht, sondern kommt zusätzlich hinzu. Es könnte auch darauf hindeuten, dass die Definition oder Interpretation dessen, was als "akustische" Musik gilt, nuancierter und vielfältiger wird.
- **instrumentalness** Bei der Instrumentalität ist ein auf und ab zu sehen, ohne klaren Trend hinsichtlich wachsendem oder fallendem Instrumentalanteil. Die Standardabweichung bleibt konsistent hoch, was bedeutet, dass nach wie vor viele Lieder mit hohem als auch niedrigem Instrumentalanteil produziert werden.
- **liveness** Auch "Liveness" zeigt keinen eindeutigen Trend. Der Durchschnittswert ist relativ niedrig mit einer geringen Standardabweichung, was darauf hindeutet, dass der Großteil der Musik nicht "live" aufgenommen wurde und immernoch wird.
- **valence** Die Valenz zeigt keinen konstanten Abstieg der Statistiken, aber eine Reduktion des Durchschnittswerts ab den 1940ern ähnlich zur danceability, was darauf hindeutet, dass die Musik seit dieser Zeit weniger positiv ist.
- **Tempo** Das Maximaltempo wird über die Jahre höher und das Minimaltempo wird niedriger, während der Durchschnittswert relativ konstant im mittleren Bereich der Attributskala bleibt. Die Standardabweichung ist niedrig und konstant. Die meisten Lieder sind also weder langsam noch schnell, aber es wird über die Jahre immer wieder mit neuen Tempo-Rekorden experimentiert.
- **Duration (Dauer):** Die Dauer zeigt einen starken Anstieg des Maximalwertes. Minimal- und Durchschnittswerte bleiben relativ konsistent und gering, genauso wie die

Standardabweichung, das heißt die Musik ist in Bezug auf die Dauer stabil.

- Time signature Der Anteil der Songs mit 4/4-Takt ist im Laufe der Zeit relativ stabil geblieben, die meisten Songs sind im 4/4-Takt, die Standardabweichung der Taktart ist gering und im Laufe der Zeit stabil, was darauf hindeutet, dass die meisten Songs hauptsächlich im 4/4-Takt sind.

Insgesamt hat es den Anschein, dass die Musik im Laufe der Jahre energiegeladener und etwas tanzbarer und lauter geworden ist, wobei die durchschnittliche Tanzbarkeit und das Energieniveau allmählich gestiegen sind. Die durchschnittliche Akustik und Instrumentalität schwanken. Die Musik scheint aber insgesamt zu mehr elektronischen oder synthetischen Klängen zu tendieren, da sich der Anteil der akustischen Musik über die Jahre eindeutig verringert. Für die Attribute deren Durchschnittswert einem klaren Trend folgt, also der relative Rückgang akustischer Elemente und der Anstieg der Energie, lässt sich auch ein Anstieg in der Standardabweichung, also der Diversität in der Ausprägung des Attributs feststellen.

V. EASE OF USE

A. Maintaining the Integrity of the Specifications

The IEEEtran class file is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

VI. PREPARE YOUR PAPER BEFORE STYLING

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections VI-A–VI-E below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not number text heads— \LaTeX will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, ac, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.

- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter”, not “webers/m²”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm³”, not “cc”.)

C. Equations

Number equations consecutively. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \quad (1)$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

D. \LaTeX -Specific Advice

Please use “soft” (e.g., `\eqref{Eq}`) cross references instead of “hard” references (e.g., (1)). That will make it possible to combine sections, add equations, or change the order of figures or citations without having to go through the file line by line.

Please don’t use the `{eqnarray}` equation environment. Use `{align}` or `{IEEEeqnarray}` instead. The `{eqnarray}` environment leaves unsightly spaces around relation symbols.

Please note that the `{subequations}` environment in \LaTeX will increment the main equation counter even when there are no equation numbers displayed. If you forget that, you might write an article in which the equation numbers skip from (17) to (20), causing the copy editors to wonder if you’ve discovered a new method of counting.

\BIBTeX does not work by magic. It doesn’t get the bibliographic data from thin air but from .bib files. If you use \BIBTeX to produce a bibliography you must send the .bib files.

\LaTeX can’t read your mind. If you assign the same label to a subsection and a table, you might find that Table I has been cross referenced as Table IV-B3.

\LaTeX does not have precognitive abilities. If you put a `\label` command before the command that updates the counter it’s supposed to be using, the label will pick up the last counter to be cross referenced instead. In particular, a `\label` command should not go before the caption of a figure or a table.

Do not use `\nonumber` inside the `{array}` environment. It will not stop equation numbers inside `{array}` (there

won't be any anyway) and it might stop a wanted equation number in the surrounding equation.

E. Some Common Mistakes

- The word “data” is plural, not singular.
- The subscript for the permeability of vacuum μ_0 , and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
- Do not use the word “essentially” to mean “approximately” or “effectively”.
- In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
- Do not confuse “imply” and “infer”.
- The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the “et” in the Latin abbreviation “et al.”.
- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [9].

F. Authors and Affiliations

The class file is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

G. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for

these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced.

H. Figures and Tables

a) *Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 2”, even at the beginning of a sentence.

TABLE II
TABLE TYPE STYLES

| Table Head | Table Column Head | | |
|------------|------------------------------|---------|---------|
| | Table column subhead | Subhead | Subhead |
| copy | More table copy ^a | | |

^aSample of a Table footnote.



Fig. 2. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [3]. The sentence punctuation follows the bracket [4]. Refer simply to the reference number, as in [5]—do not use “Ref. [5]” or “reference [5]” except at the beginning of a sentence: “Reference [5] was the first . . .”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [6]. Papers that have been accepted for publication should be cited as “in press” [7]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [8].

REFERENCES

- [1] Pichl, Martin; Zangerle, Eva; Specht, Gnther: ”Towards a Context-Aware Music Recommendation Approach: What is Hidden in the Playlist Name?” in 15th IEEE International Conference on Data Mining Workshops (ICDM 2015), pp. 1360-1365, IEEE, Atlantic City, 2015.
- [2] Figueroa, Rodolfo. (2020). Spotify 12M Songs [Dataset]. Kaggle. <https://www.kaggle.com/datasets/rodolfofigueroa/spotify-12m-songs>
- [3] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.
- [4] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [5] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [6] K. Elissa, “Title of paper if known,” unpublished.
- [7] R. Nicole, “Title of paper with only first word capitalized,” *J. Name Stand. Abbrev.*, in press.
- [8] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [9] M. Young, *The Technical Writer’s Handbook*. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.