

Analysing Financial Firms' Terms & Conditions

How to Make T&Cs Easier
for Retail Customers to
Understand

Project Lead: Sally Lai

Project Developers: Dylan Butcher, Nicole Xiao, Rei Yi Lai, Katie Lee



Look around ... Terms & Conditions are everywhere

The image displays a grid of nine screenshots from different financial institutions' websites, all featuring a page titled 'Terms and Conditions' or 'Terms of Use'. The institutions shown are:

- HSBC: Who we are, Insight, Investors, News and media, Careers
- Blackstone
- LLOYDS BANK: PERSONAL, BUSINESS, PRIVATE ETC, LLOYDS BANK logo
- Barclays: Digital Banking, Bank accounts, Borrowing, Credit, Back to: Important information, Terms and Conditions for Customers
- Morgan Stanley
- Goldman Sachs
- LAZARD ASSET MANAGEMENT: About, Investments
- Santander: Current accounts, Credit cards, Insurance, Loans, Mortgages, Savings and investments, Support, Personal > Support > Ways to bank > Online Banking Service Terms & Conditions
- Online Banking Service Terms & Conditions

The 'Terms and Conditions' pages are highlighted in each screenshot, demonstrating their prevalence across the industry.

In July 2022 ...

Financial Conduct Authority (FCA) released a new Consumer Duty:

- 2A.5.7 G In supporting the understanding of *retail customers* through its communications, a *firm* should:
- (1) explain or present information in a logical manner;
 - (2) use plain and intelligible language and, where use of jargon or technical terms is unavoidable, explain the meaning of any jargon or technical terms as simply as possible;
 - (3) make key information prominent and easy to identify, including by means of headings and layout, display and font attributes of text, and by use of design devices such as tables, bullet points, graphs, graphics, audio-visuals and interactive media;
 - (4) avoid unnecessary disclaimers; and
 - (5) provide relevant information with an appropriate level of detail, to avoid providing too much information such that it may prevent *retail customers* from making effective decisions.



Project Motivation

Key Takeaways from FCA Consumer Duty

- Best interests
 - Requiring financial firms to support retail customers
- Readability
 - Using plain language to avoid complex/ technical terms
- Legibility
 - Making key information easy to identify



Why data science?

- **More efficient and robust**
 - Web-scraping: access and collect large amounts of information (T&C) from the internet
 - Use NLP to determine text complexity - both readability and legibility



Project Goals

1

Construct a measure of T&C interpretability based on readability and legibility

2

Obtain measure for a set of companies for cross-industry comparison

3

Evaluate effectiveness of measure and suggest directions for further research

Project Plan

- 1 Defining scope of companies**
- 2 Getting URLs to T&Cs of all companies**
- 3 Extracting features from T&Cs through web-scraping**
 - Text
 - Formatting features
- 4 Constructing Readability Measures**
- 5 Constructing Legibility Measures**
- 6 Results & Overall Metrics**

1. Defining Scope of Companies

- Institutions monitored by the FCA
- Selected sectors with the most complete data on company names
- Removed defunct companies

Sector	Data Source	# of Companies
Banks	Bank of England	322
Building Societies		43
Insurers		1046
Credit Institutions	Financial Conduct Authority	382
Investment Firms		3075

*Unbalanced data
across sectors - *Not a concern* because we
are not running models

Data as of 25th October 2022

2. Getting URLs to T&Cs of Companies



Focused on website T&Cs instead of financial products T&Cs

- Performance on general website T&Cs as a proxy for overall compliance
- More difficult to find financial products T&Cs (different names, PDF formats, etc.)

Used BeautifulSoup to perform web scraping

- Research Paper: *Automatic Detection of Terms and Conditions in German and English Online Shops*
 - Grabbed the highest ranking link according to Google
 - Recursed through all sub-pages whose URLs contained items of interest ('terms-and-conditions', 't&c', 'privacy', 'cookies' etc.)

This was difficult as...

- Not always accurate - Picked up some firms' social pages
 - Deleting links: 5'000 to 4'500
- Blocked by Google reCAPTCHA

3. Feature Extraction: Text

Used a package that specifically extracts terms and conditions

- Package: *Structured extraction of content from German and English Terms and Conditions* ([GitHub](#))
 - Gets main content of the webpage
 - Uses visual and numerical information to build a hierarchical tree of content
 - Better extracts core T&C information without also extracting too much noise



Cleaned the resulting JSON output by

- Extracting text sections
- Using a recursive function to crawl through the output and extract clean sentences

3. Feature Extraction: Style

HTML

Used BeautifulSoup & **HTML** (Header Text, All text etc.)

```
<main role="main" tabindex="-1" aria-hidden="false">
  <section class="sublayout sublayout--s1" data-index>
    <div class="container-fluid">...</div>
  </section>
  <div class="sublayout article-sublayout ">
    <div class="container-fluid">
      ::before
      <div class="row"> (flex)
        <div class="col-sm-8 col-sm-offset-2 article-sublayout__content">
          <p>
            "Before using this Site, you should read these Terms carefully, as
            <a href="/cookie-notice" class="rte-internal-link no-icon">Cookie P
            ". By accessing this Site, you agree to be bound by the Terms."
          </p>
        </p>
        <p>
          "HSBC Holdings plc has a Companies House number of 14259 and a prin
          is authorised by the Prudential Regulation Authority and regulated
          Regulation Authority." == $0
        </p>
      </div>
    </div>
  </div>
</main>
```

css

Used BeautifulSoup & CSS (Line Height & Font Size)

4. Constructing Readability Measures: Starting point



What is Readability?

→ “The degree to which a given class of people find certain reading matter compelling and comprehensible” (DuBay, 2004)

Banks	14.25
Building Societies	13.15
Credit Institutions	13.73
Insurers	14.63
Investment Firms	14.28

Flesch–Kincaid Readability Tests:

- Based on two components
- FK is neither perfect or unique
 - Standard for assessing readability for gov. & corporations

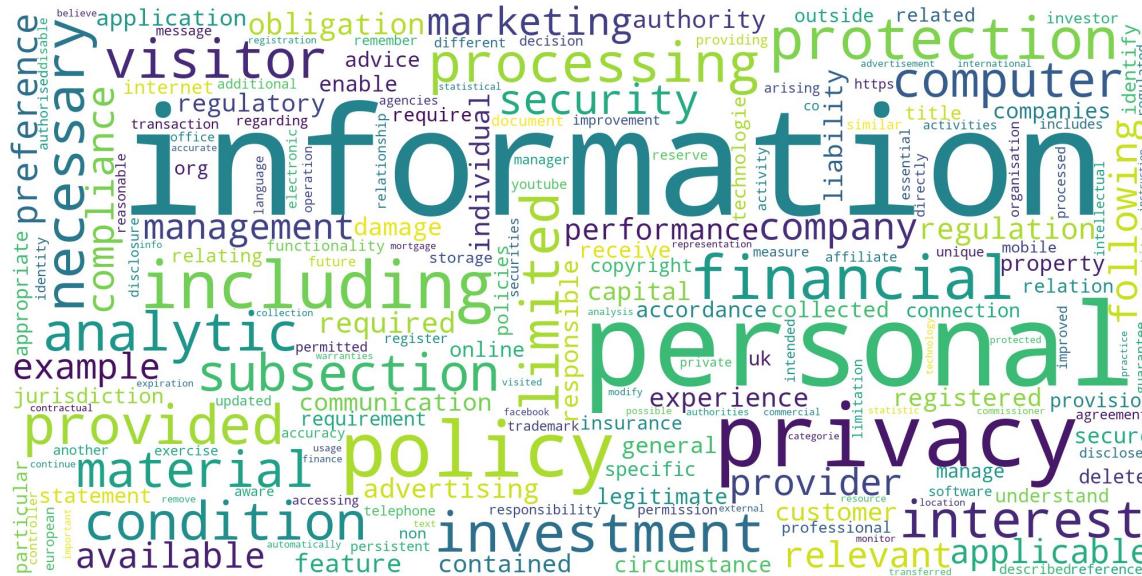
FK: 8 (schooling age 13-14);
FK>14: college-level reading material

4. Constructing Readability Measures: Complex Words

The most frequent “complex words” ...

(words with more than two syllables)

$$0.39 \left(\frac{\text{total words}}{\text{total sentences}} \right) + 11.8 \left(\frac{\text{total syllables}}{\text{total words}} \right) - 15.59$$



4. Constructing Readability Measures:

Traditional measure

- By nature, T&C for financial firms has high % of multi-syllable words to describe operations
 - Yet they are presumably easy fo average reader to understand
- Word complexity adds measurement error for the readability metrics

More reasonable measures of readability

1. Average sentence length

- longer sentence = harder to read

2. Log (Number of Characters)

- longer document = harder to read

3. Common Words

- more common words = easier to read

5. Constructing Legibility Measures:

Legibility Guideline	Metrics
Use bolding to highlight in order to emphasize keywords and concepts	Ratio of bold text to all text
Use bullets or numbers rather than continuous prose when possible	Ratio of text in numbered/bullet-point lists to all text
Use section headings to organize the content	Ratio of header text to all text
Use font sizes of at least 12 pt	Median font size of body text
Keep the between-line spacing of 1.5 point	Average ratio of text line height to font size

5. Constructing Legibility Measures:

1. Bolded Text

No. of words in Bold Text / All Body Text

- ``

4. Font Size

Median font size of all body text

- `font-size:(\d+(?:[.\,]\d{1,3}))?`

2. Bullet / Numbered Points

No. of words in Bullet or Numbered Points /
All Body Text

- ` / `

5. Line Height

Median line height of all body text

- `line-height:(\d+(?:[.\,]\d{1,3}))?`

3. Header Text

No. of words in Header Text / All Body Text

- `<h1>, <h2>, ... , <h7>`

6. Results & Overall Metrics

Combining Metrics by Standardisation

Step 1: Standardise all metrics (*via sklearn package*)

$$\text{Standardised Metric} = (\text{Original Metric} - \text{Mean})/\text{Std Deviation}$$

Step 2: For each link, get the mean of the all standardised metrics

Choosing Standardisation over Normalisation

- Our data approximately follows a normal distribution, but with many outliers
- Standardisation helps to address the outliers, but also standardised the scale for each metric

6. Results & Overall Metrics: Standardisation Formulas

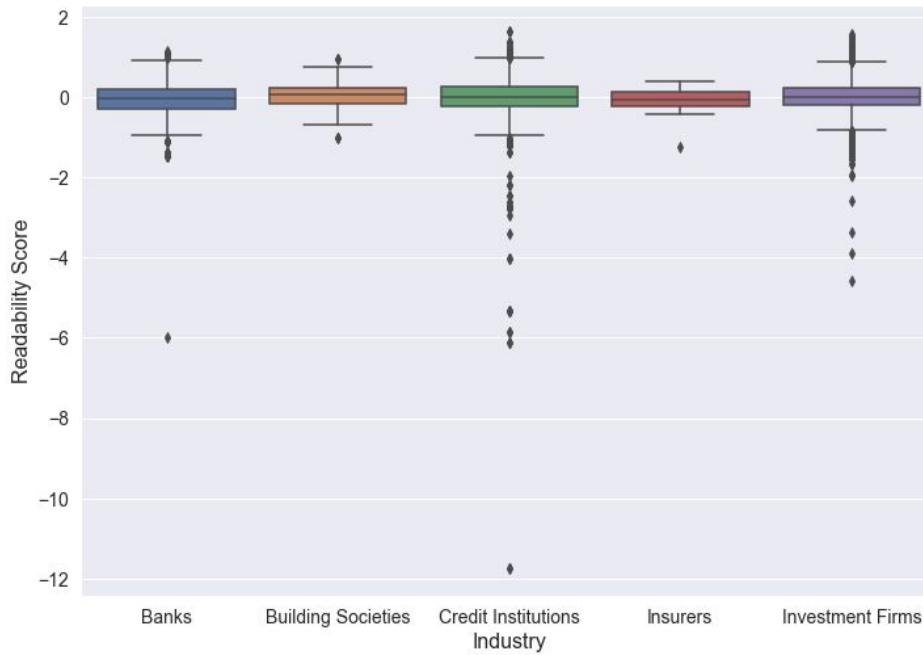
Combined = Readability + Legibility

$$\begin{aligned} \text{Readability} = & \frac{1}{3} [-\text{Standardised}(\text{Avg Sentence Length}) \\ & + \text{Standardised}(\% \text{ of Common Words}) \\ & - \text{Standardised}(\log(\text{Number of Character}))] \end{aligned}$$

$$\begin{aligned} \text{Legibility} = & \frac{1}{4} [\text{Standardised}(\text{Bolded Text Ratio}) \\ & + \text{Standardised}(\text{Bullet point List Ratio}) \\ & + \text{Standardised}(\text{Header Text Ratio}) \\ & + \text{Standardised}(\text{Line Height Font Size Ratio})] \end{aligned}$$

6. Results & Overall Metrics: Readability

Readability By Sector



Median Readability Score by Sector:

Banks	-0.007919
Building Societies	0.066019
Credit Institutions	0.026923
Insurers	-0.051291
Investment Firms	0.010421

6. Results & Overall Metrics: Readability

Excerpt From Most Readable T&C

Dropbox Sign Privacy Policy

Posted: October 25, 2022

Effective as of: October 25, 2022

INTRODUCTION

Your privacy is important, so Dropbox has created the following Privacy Policy to let you know what information we collect when you visit or use the Dropbox Sign, Dropbox Forms, and Dropbox Fax websites and services (collectively "Dropbox Sign Services"), why we collect it, and how it is used. This Privacy Policy explains the data collection and use practices of Dropbox in connection with the Dropbox Sign Services. The terms "you," "your," "yours" and "Customer" refer to the customer/purchaser utilizing our Site. The terms "we," "us," and "our" refer to Dropbox.

This Privacy Policy is governed by the Dropbox Sign Terms of Service ("Terms"). The terms "Site", "Website", "Application", "Product," and "Service" refer to the Dropbox Sign Services. This Privacy Policy does not apply to any third-party websites and apps that you may use, including any that are linked to in the Dropbox Sign Services. You should review the terms and policies for third-party websites and apps before clicking on any links.

Excerpt From Least Readable T&C

سياسة الخصوصية وشروط الاستخدام www.riyadbank.com -

MEMBERSHIP GENERAL TERMS AND CONDITIONS

متى يتم استخدام موقعي بنك
متى يتم استخدام المواقع علىكم

معلومات فقط

وأنت تناولت على

في الشبكة و ما ينفع

على جهود الشروط و

و ممارسات أو عروض

أيضاً يتفق أن توافق

الصادقة في التي تعتبر

< Back to DOCUMENTATION

Market Data >

Market Rules >

Membership General Terms and Conditions

Member Portal >

Regulatory Structure >

Connectivity >

Regulatory Market Notices >

Participant Complaints >

Here you can find the general terms and conditions regulating the relationship between MTS and the market participants to be read in conjunction with the Request for Services and all the relevant schedules.

Markets managed by MTS SpA

MTS S.p.A. Terms and Conditions effective as of 01/04/2021 (i.e. BondVision Europe MTF, MTS Italy Cash and Repo, MTS Depo, MTS Cash Domestic Markets and EBM)

MTS S.p.A. Markets (BondVision Europe MTF MTS Italy Cash and Repo, MTS Cash Domestic Markets and EBM) Terms and Conditions (English version)

MTS S.p.A. Markets (BondVision Europe MTF, MTS Italy Cash and Repo, MTS Cash Domestic Markets and EBM) Terms and Conditions (Italian version)

Schedule GUI Licence Terms and Conditions

MTS S.p.A. Membership and Trading Fees

BondVision Europe MTF (Cash & Repo Segments) - Dealer & Quote Provider - Fee schedule effective 01 June 2022

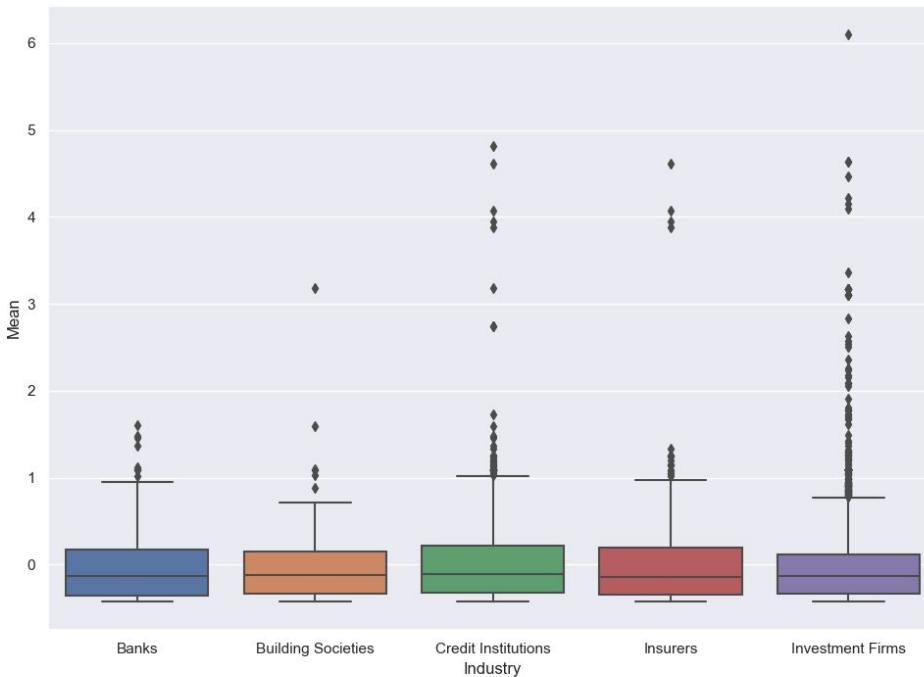
BondVision Europe MTF (Cash & Repo Segments) - Dealer & Quote Provider - Fee schedule effective 01 January 2023

BondVision Europe MTF (Cash & Repo Segments) - End User & Quote Requester - Fee schedule effective 01 June 2022

BondVision Europe MTF (Cash & Repo Segments) - End User & Quote Requester - Fee schedule effective 01 January 2023

6. Results & Overall Metrics: Legibility

Legibility By Sector



Median Legibility Score by Sector:

Banks	-0.133337
Building Societies	-0.123201
Credit Institutions	-0.110909
Insurers	-0.145995
Investment Firms	-0.131316

6. Results & Overall Metrics: Legibility

Excerpt From Most Legible T&C

6. IF YOU ARE LOCATED IN THE UNITED KINGDOM OR EUROPE

- Important information
- Lawful basis
- Sharing personal information with third parties
- Transfers of personal information outside of UK and Europe

7. IF YOU ARE LOCATED IN THE DIFC

8. IF YOU ARE A CALIFORNIA RESIDENT

9. IF YOU ARE LOCATED IN CHINA, HONG KONG, JAPAN, OR SINGAPORE

10. YOUR RIGHTS IN RELATION TO YOUR PERSONAL INFORMATION

11. HOW TO CONTACT US

1. IMPORTANT INFORMATION

Evercore respects your privacy and is committed to protecting your personal information.

This privacy notice describes how we collect and use personal information about you during and after your relationship with us, in accordance with applicable laws.

It applies to Evercore's business contacts, suppliers and professional advisors, parties with whom Evercore conducts business (including clients, counterparties

basis that Evercore relies on when using your personal information.

► Please click here for information about how Evercore shares your personal information with third parties.

► Please click here for information about how Evercore transfers your personal information outside of the United Kingdom or Europe.

7. IF YOU ARE LOCATED IN THE DIFC

Evercore is providing the following information for individuals whose personal information is collected or held by Evercore Advisory (Middle East) Limited, which is the controller of your personal information and so responsible for this privacy notice.

Where you are such an individual:

- Evercore Advisory (Middle East) Limited is registered in the Dubai International Finance Centre (DIFC) with a DIFC registered number 2478 and with its registered office at Unit 202, 203A & 204A, Level 2, Gate Village Building 10, DIFC, Dubai, 507204, United Arab Emirates.

- The processing of your personal data by Evercore Advisory (Middle East) Limited is governed by the DIFC Data Protection Law, No.5

Excerpt From Least Legible T&C

Terms and Conditions

Before accessing and using the Eriswell Capital Management LLP's website and its contents please read these terms and conditions as they constitute a Legal Notice and contain important legal information. This Legal Notice is governed by English Law and the English courts shall have exclusive jurisdiction over any matter arising out of this Legal Notice or from your accessing of the website. It is your responsibility to be aware of and to observe all applicable laws and regulations of any relevant jurisdiction.

By accessing and using the Eriswell Capital Management LLP's website you acknowledge that you have reviewed this Legal Notice and understand and agree to the terms and conditions contained herein. If you do not agree to the terms and conditions below, do not access or use this website in any way.

Eriswell Capital Management LLP's website contains certain information about its approach to providing investment advice and is presented for informational purposes only. It does not represent that the services described on the site are suitable for any specific investor. You are advised not to rely on any information contained in this site in the process of making a fully informed investment decision. Instead, you are urged to base investment decisions upon a thorough investigation and to obtain all necessary professional advice.

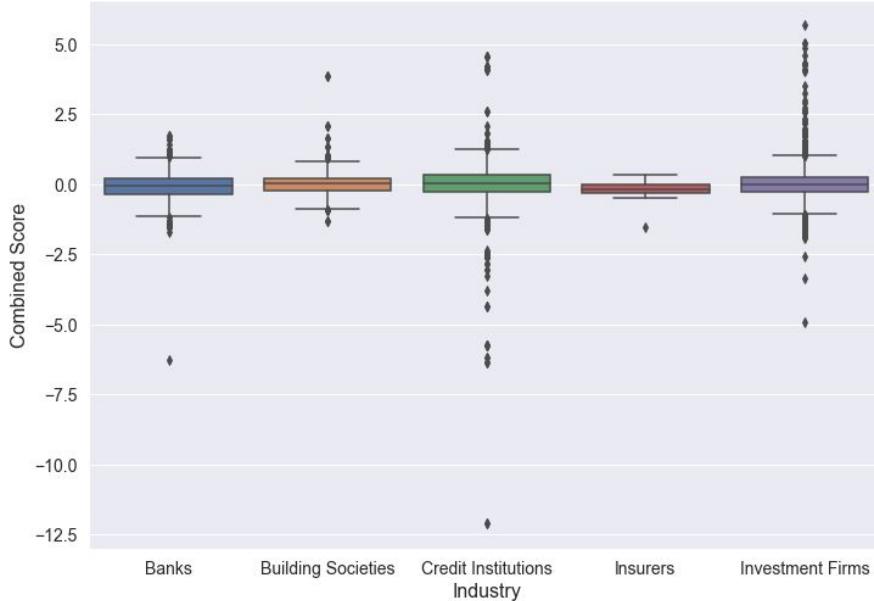
The information provided on this website is not intended for distribution to, or use by, any person or entity in any jurisdiction or country where such distribution, publication or use would be contrary to local law or regulation or in which Eriswell Capital Management LLP does not hold any necessary registration or license. Individuals or legal entities in respect of whom such prohibitions apply, whether on grounds of their nationality, their place of residence or on other grounds, must not access or use this website. In particular Eriswell Capital Management LLP is not registered as an investment advisor with the Securities and Exchange Commission (SEC) and therefore this website is neither directed at nor intended for investors in the United States of America.

To the extent that material on this Site is issued in the United Kingdom ("UK"), it is issued for the purposes of the UK Financial Services and Markets Act 2000 by Eriswell Capital Management LLP which is authorised and regulated in the UK by the Financial Conduct Authority.

Past performance data contained on this website is no indication of future performance and nothing on this website should be interpreted to state or imply otherwise. The value of investments may fall as well as rise and investors may not get back the full amount invested. In addition, the information and materials herein shall not constitute an offer or solicitation, or an offer to sell, shares of any of the funds or any advisory or management service in any jurisdiction.

Additionally, the information on this website is provided "AS IS" and "AS AVAILABLE". Eriswell Capital Management LLP is under no obligation to update the information to reflect changes after the publication date. It is presented without warranty of any kind, either express or implied, including without limitation of any warranties concerning the availability, reliability, accuracy, completeness, timeliness or sequencing of the site or the content, products or services available on or via the website. Also, the information offered

6. Results & Overall Metrics: Combined Score



Banks	-0.097586
Building Societies	0.053637
Credit Institutions	-0.019384
Insurers	-0.215751
Investment Firms	0.013479

Evercore Global Privacy Notice

Date effective and last updated: November 3rd, 2021

Evercore respects your privacy and is committed to protecting your personal information. This privacy notice will inform you as to how we collect, share, use and look after your personal information, both online and offline, and tells you about your privacy rights and how the law protects you.

This privacy notice is provided in a layered format so you can click the arrows to reveal further detail. Please also use the following table contents to click through to the specific areas set out below. Alternatively, you can download a pdf version of the full privacy notice here.

1. IMPORTANT INFORMATION
2. PERSONAL INFORMATION WE HOLD ABOUT YOU, AND THE PURPOSES FOR WHICH WE USE IT
 - Visitors to this website
 - Visitors to Evercore's offices
 - Our business contacts
 - Our suppliers and professional advisers
 - Parties with whom we conduct business (including clients, counterparties and potential clients and counterparties)
 - Our investors
3. DATA SHARING
4. DATA SECURITY
5. DATA RETENTION

Conclusion: Evaluation

Data Collection

- Privacy/Cookies not a good proxy
- T&Cs in HTML form can introduce biases
- Some websites not exactly T&Cs / not in English



Feature Extraction

- Technical issues (Dynamic / Static) with extracting text and features



Evaluation of Metrics

- Measures may not fully and accurately capture readability or legibility
- If we design T&Cs with the sole purpose of optimising our metrics, may end up with something that is undesirable



Data Analysis

- Comparison across more dimensions (e.g. firm size, firm incorporation year)
- Potential issues with observed outliers - difficult to look into

Conclusion: Implications for Society

Through our study, we have shown how it is possible to distill down T&Cs' readability and legibility to a select key features:

- **Science and technology** — Shows promise for developments in legal tech applications to improve the comprehensibility of complex terms of agreements between parties.
- **Fairness** — Shifts the onus of responsibility to large corporations to make their T&Cs readable, leading to improved individualised products and services
- **Equality** — Accounts for disadvantaged, less educated groups who may be less able to comprehend the terms of their agreements
- **Ethical & Economic** — Helps to address asymmetric information problem: While making T&Cs more readable may not change consumer's attitudes and behaviour, consumers have the right to know what their contracts say

Conclusion: Extensions & Further Research

If we had more time, we would...

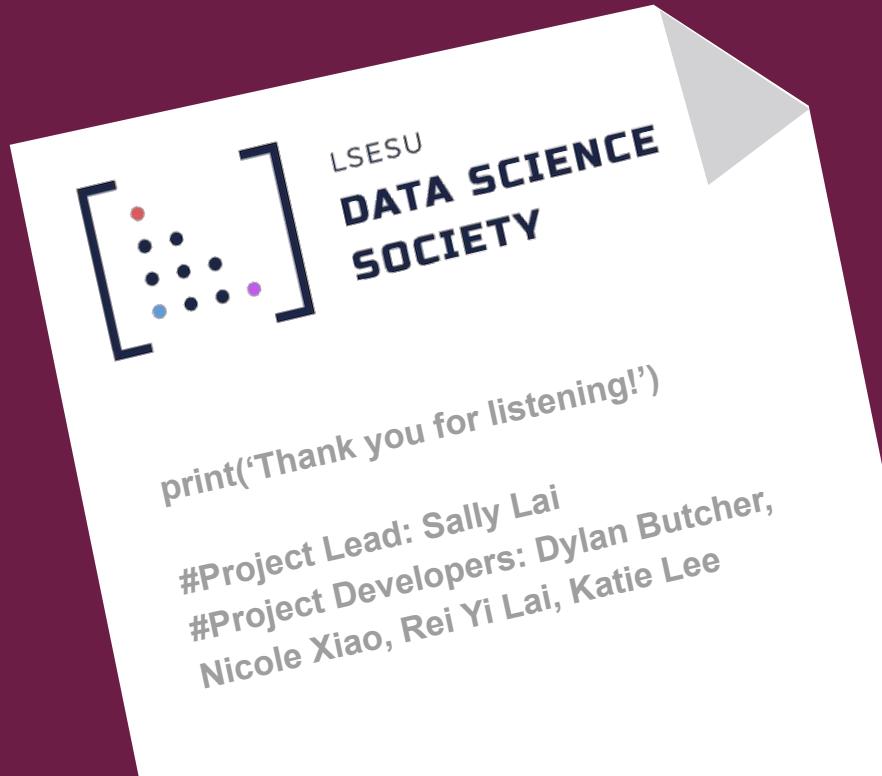
- Count the no. of hyperlinks on each T&C page
- Analyse T&C readability and legibility by firm size and value
- Create measures to take account of non-English pages

Further research

- Use readability and legibility metrics on other sectors
- Explore alternative features/measures of legibility and readability using ML methods - build classifier
- Further look into measuring obligation count (a good measure for interpretability?)
 - Obligations can be found by looking at sentences which contains an entity and an obligation keyword (“must”, “will”, “provide” etc.)

Q&A

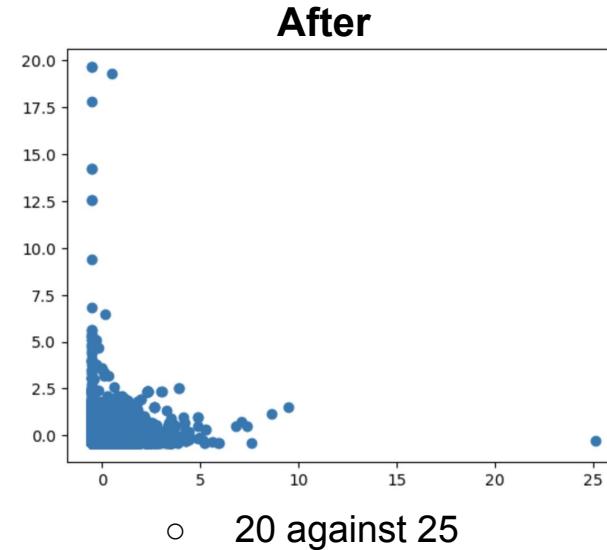
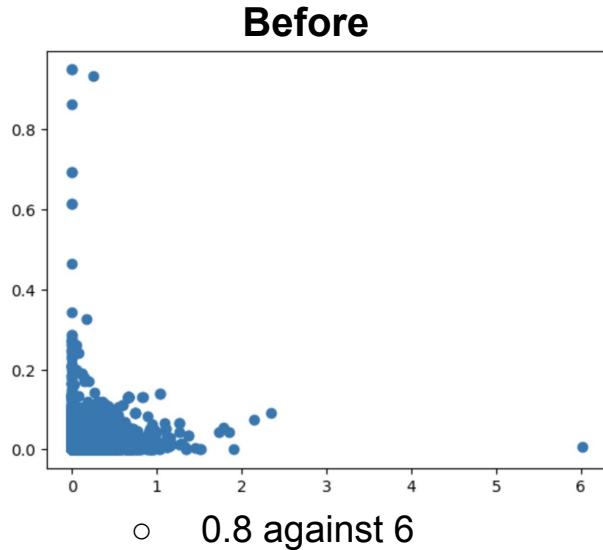
LSE Data Science Society
MT Project Showcase 2022



Appendix A

Why standardisation over normalisation

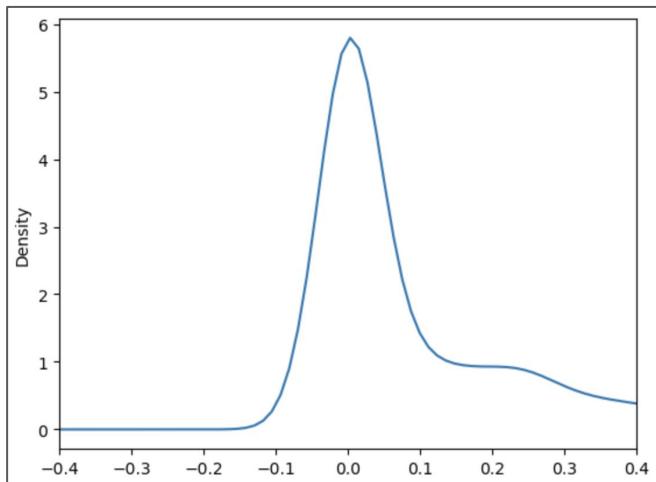
- 1) Helps to standardise the scale for each metric
 - Such that the larger range of one metric doesn't contribute more than the other metric



Appendix A

Why standardisation over normalisation

- 2) Our data approximately follows a normal distribution, but with many outliers
 - Standardisation is more robust to outliers



Appendix B

CHI Conference on Human Factors in Computing Systems: Guideline-Based Evaluation of Web Readability (May 2019)

Table 1. Readability guidelines tested in the study.

ID	Guideline Text
G1	<i>Use left-justified text with the right edge being ragged, non-justified.</i>
G2	<i>Use an off-white color for your background, like light gray or tan; use dark gray for text instead of pure black.</i>
G3	<i>Use a plain, evenly spaced sans serif font such as Arial and Comic Sans.</i>
G4	<i>Avoid using italics in the main body of the text.</i>
G5	<i>Use bolding to highlight in order to emphasize keywords and concepts.</i>
G6	<i>Avoid underlining large blocks of text as it makes reading harder.</i>
G7	<i>Use font sizes larger than 12pt.</i>
G8	<i>Avoid capital letters, apart from the beginning of sentences, abbreviations, and where it is grammatically correct.</i>
G9	<i>If appropriate, use bullets or numbers rather than continuous prose.</i>
G10	<i>Use short, simple sentences in a direct style.</i>
G11	<i>Use active rather than passive voice.</i>
G12	<i>Avoid complex language and jargon.</i>
G13	<i>Consider using short paragraphs.</i>
G14	<i>Embed in Webpage texts the hyperlinks to the pages with the text-related concepts.</i>
G15	<i>Avoid images that are 'busy', cluttered, and include too much extra detail.</i>
G16	<i>Avoid placing images above text or text around images.</i>
G17	<i>Place the main point at the very top of page.</i>
G18	<i>Place important content in a single main column and avoid two-dimensional layouts.</i>
G19	<i>Ensure navigation menus use a text size that allows for comfortable reading.</i>
G20	<i>Avoid starting a new sentence at the end of a line.</i>
G21	<i>Keep the between-line spacing of 1.5 point.</i>

Appendix C: Legibility (w/ URL Count)

★ URL Count

Number of URLs of a webpage

- <a href>

Appendix C: Legibility (w/ URL Count)

Excerpt From Most Legible T&C

- Financial status including CCJs and declarations of bankruptcy
- Previous Claims

2. Why we use the information about you

We collect your personal information for a number of reasons: so we can provide you with a quote, administer your policies, handle claims, protect us and you from fraud, to improve the products and services we offer, and finally to contact you about your policy where necessary.

To process your information, we rely mainly on the following legal bases:

Performance of a contract: The use of your information is necessary to perform the contract that you, or a person covered on the policy, has with us.

Legitimate interests: We may use your information for our legitimate interests, such as to provide you with the most suitable policy, website experience, emails or newsletters, or to improve and promote our products and services, along with administrative, fraud prevention and detection and legal purposes.

Consent: We may rely on your consent to use your personal information for certain products or to assist in a claim. You may withdraw your consent at any time by contacting us. Please contact us using the details end of this Policy

3. How will we use the information about you?

We use the personal information we collect for a number of reasons, these include:

- To service your policy and any claims you make
- Where necessary, we'll use your information to verify your identity, or those transacting on the policy
- For fraud and financial crime prevention and detection, and for legal purposes where we may need to use your information to investigate or respond to legal disputes, regulatory investigations or for compliance purposes
- To identify your additional needs (e.g. A change in health, disability adjustments or certain significant life events) which allow us to adapt our services to your needs
- We may also use your personal information to ensure we comply with legal and regulatory requirements. This includes internal audits, monitoring and assurance activities to examine and evaluate the systems and controls within Avon Insurance
- When you call us, we record calls for quality control and training purposes. How long we keep recordings depends on our retention policy (see Section 9), unless Avon Insurance has a legitimate interest to keep recordings longer (as reasonably necessary), including for fraud detection and financial crime investigation purposes. Additionally, we record calls to:
 - Ensure that we have correctly followed your instructions
 - Resolve any queries, issues or disputes
 - Help us understand how we could improve customer service
 - Respond to regulator queries

Excerpt From Least Legible T&C

[Manage Cookie Settings](#)

Changing your browser settings

You can also change the settings on your web browser to accept or reject all, or certain types of cookies. You will still be able to use most of our websites with no loss of functionality if cookies are disabled from the web browser. However, if you reject necessary cookies you may not be able to use or access certain banking service on your device.

Information on how to manage and disable cookies can generally be found in the privacy or help section of your browser's website. If you want to disable cookies on other devices or browsers, you will need to disable the cookies on each device and on each browser separately.

Find out how to manage and disable cookies on popular browsers:

- [Google Chrome](#)
- [Microsoft Edge](#)
- [Mozilla Firefox](#)
- [Microsoft Internet Explorer](#)
- [Opera](#)
- [Apple Safari](#)

Visit www.aboutcookies.org for info about managing or opting out of receiving cookies.

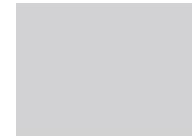
Visit www.youronlinechoices.com for info about opting out of advertising cookies.

FCA Color HEX Codes:

Wine red: #6c1d45



Sea Blue: #00749e



Mint Blue: #00bfb3

Pink red: #ff7c80

Light grey: #d2d2d4

https://www.flaticon.com/search?author_id=1&style_id=133&type=standard&word=coding

Got the color scheme from
the FCA website

