Analysing Email Reference Transactions in a Legal Library

Nicholas Lobraico

Presentation Outline

1. **Introduction**

2. **Project Overview & Outline**

3. **Major Challenges**

Introduction

1. **Who am I?**

- Recent graduate of MISt programme at McGill's School of Information Studies
- Interest in academic & research libraries, text analysis, data mining
- Background in humanities, philosophy and political science

2. **What are we talking about?**

- Practicum project (Jan. – April, 2023)
- Davies Ward Phillips & Vineberg (Montreal)
- Email transactions of legal librarians

3. **What will you takeaway?**

- General understanding of the project, how it worked, who might be suitable for it, what to expect
- A sense of the major challenges faced
- A template for you to implement this sort of project in your own library

Overview of Project

- Davies Ward Phillips & Vineberg :
  - Large legal firm with over 200 lawyers
  - Offices in Toronto, Montreal and new York
  - Montreal office across the street from McGill University
  - Montreal library consists of 4 team members, two of whom focus on research

- *"The aim of this project is to provide an assessment of the reference services provided by the library at the Montreal office of Davies Ward Phillips & Vineberg"*

Project Outline

1 Determine Scope

2 Export emails

3 Clean Dataset

4 Develop & Apply Transaction Categories

5 Deliver Final Analysis

1 Determine Scope

- **How many email transactions can we expect to analyse?**
  - 100 hours
  - All emails from Jan. 2022 – present (roughly 1 year)
  - Roughly 4,000 emails
- **Which emails are most interesting / important?**
  - Prioritize research questions over known item and other more routine transactions (eg., press reviews, access issues)
- **What is the purpose of the analysis?**
  - Making an argument for the need of a email ticketing system similar to those used by IT departments

2 Export emails

- 4 mail boxes of interest (1 general, 3 personal)
- Microsoft environment : Outlook, Mimecast, Excel
  - Export directly from Outlook into Excel

Challenges:
- Email body often cut short in Excel
- Organizing inboxes in separate tabs
- Determining un-needed fields (ex. HasAttachments)

Folder | Sent | From | Date Sent | In Copy | Sent To | Subject | Preview | ID | Category | Notes

3 Clean Dataset

- 3957 > 1212 records
- Removing irrelevant transactions (duplicates, press review, etc.)

Challenges:
- Rich-textual data
  - multi-lingual
  - processing time (eg. find & replace)
- Duplicates (within and between mailboxes)
- How to delineate the reference transaction?
  - Cascading questions
  - Complex email chains
  - Unclear questions
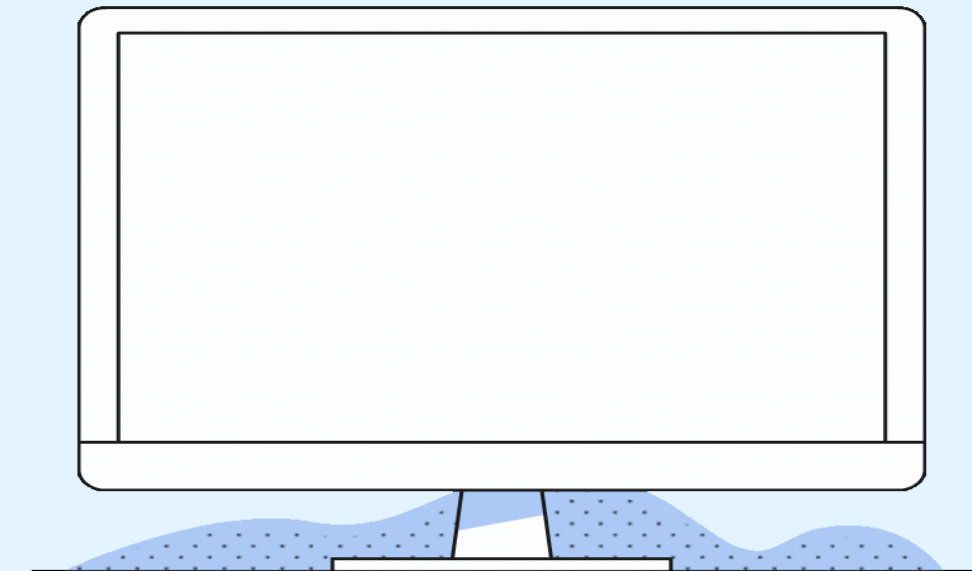
# **4** Develop & Apply Transaction Categories

- How to define transaction categories?
  - Origin of question (practice group)
  - Team member responding
  - Question topic (annotation, doctrine, noting up)
  - Question type (access, research, known item)
  - READ Scale

Challenges:
- Transactions involving multiple persons / practice groups
- More granularity leads to less insight (sub-categories)

5 Deliver Analysis

Count of Transactions by Category

Count of Transactions by Practice Group
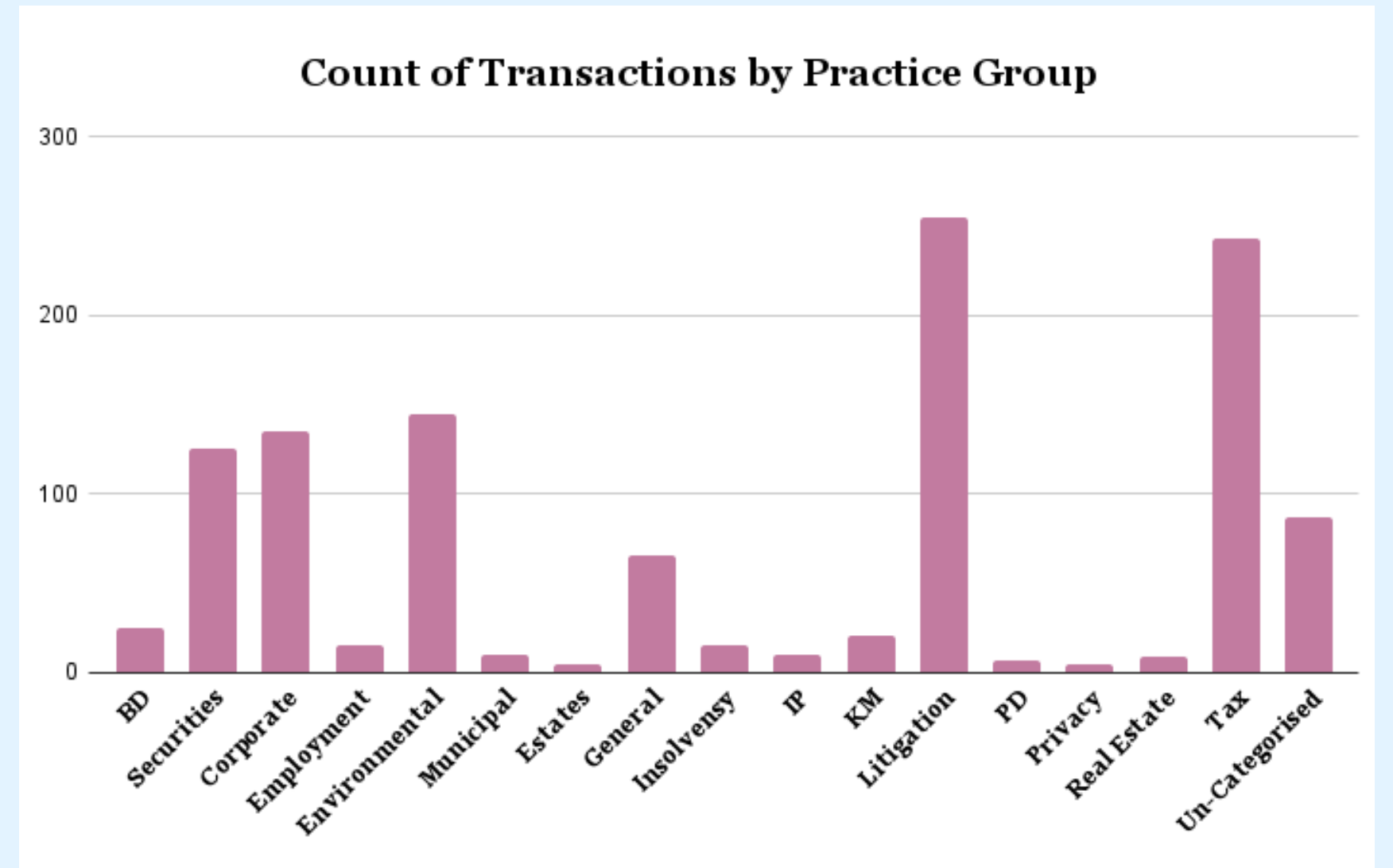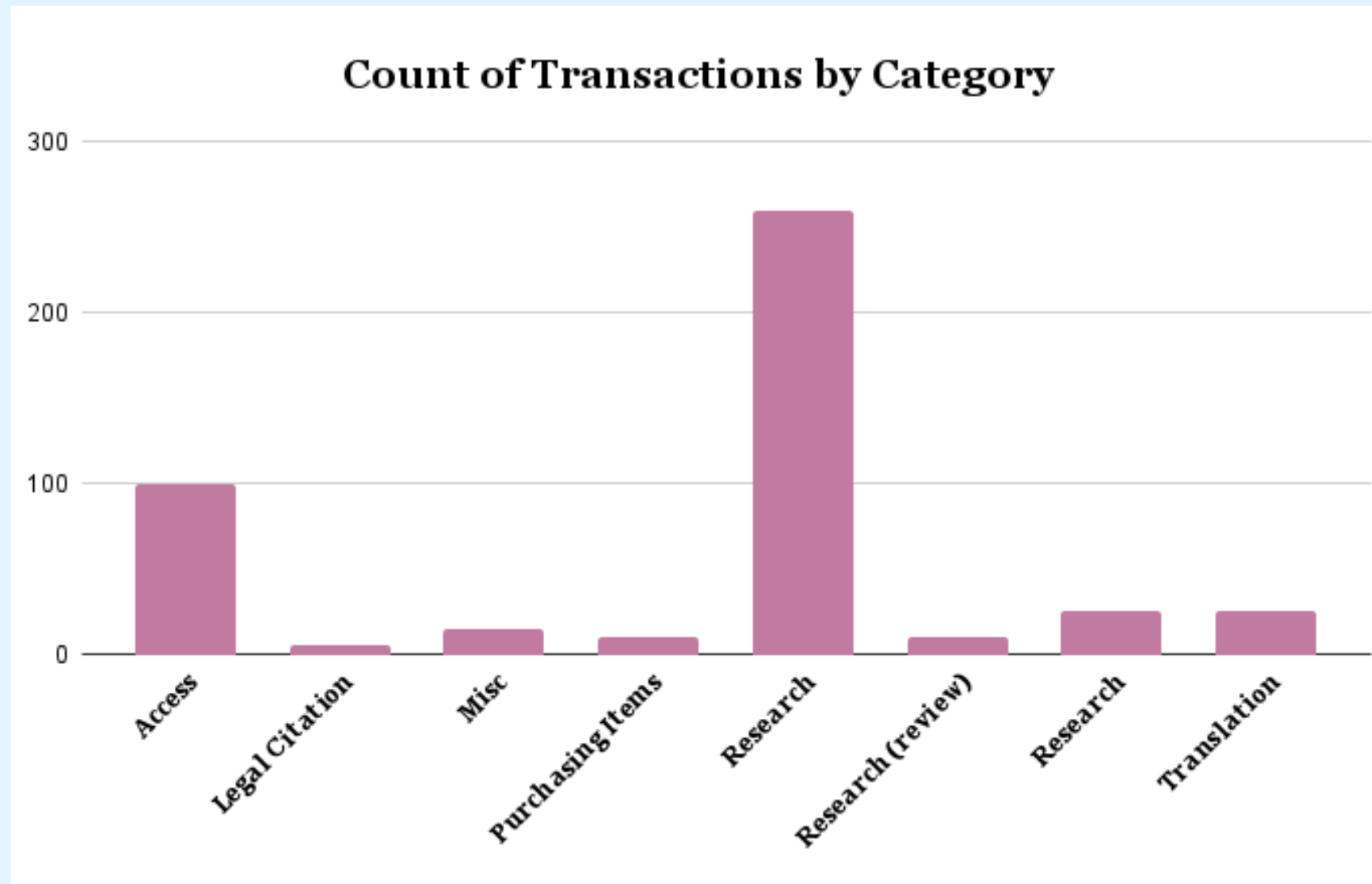
Disclaimer : transactions can belong to multiple categories & practice groups, some transactions are represented more than once

Review of Major Challenges

- Email body often cut short in Excel
  - Means that some transactions had to be read through on Outlook

- Lots of doubles and triples (3957 > 1212 records)

- Rich-textual data poses problems for Excel
  - Slow processing speed, routine crashing
  - Combined with multi-lingual text made parsing tricky

- Delineate the reference transaction is not straight-forward, transactions are more complex than we can represent in Excel

Lessons Learned
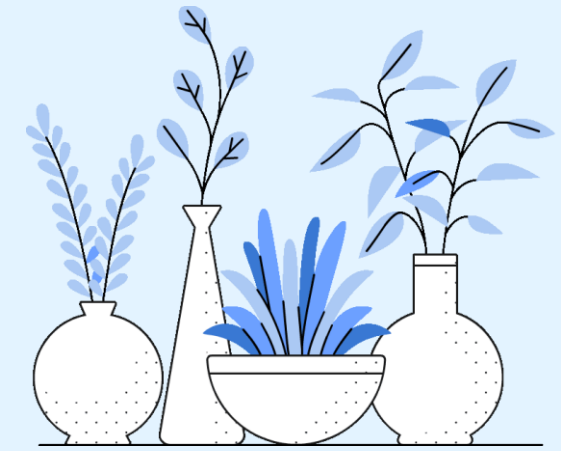
- More granularity can lead to less insight
  - The more categories we develop the fewer trends we identify
- Exporting from Outlook to Excel is messy
- Rich textual data is difficult to work with in Excel

- **This type of project is time consuming!**
  - 100 hours to clean 4000 transaction records and categorise & analyse 1212…

Thank You!