Procurement Dynamics: Analyzing Federal Contract Awards to Microsoft*

A Comprehensive Study of Microsoft's Role in Canadian Government Procurement (NEED a better SUBTITLE being more specific)

Deyi Kong

November 26, 2024

This paper analyzes the federal procurement contracts awarded to Microsoft, focusing on trends in award values, the categories of the contract, and the relationships between Microsoft and government buyers. Using IJF procurement dataset, we found that Microsoft engages significant role in IT infrastructure and modernization efforts. The analysis reveals that Microsoft dominates high-value contracts and demonstrates consistent partnerships with key government departments. By examining these procurement dynamics, this paper providing insights into the broader implications of tech-industry participation in government procurement.

Table of contents

*Code and data are available at: https://github.com/eeeee-cmd/Procurement/.

1 Introduction

The main estimand is the dollar amount of federal contracts.

The structure of the paper is as follows: Section 2 outlines the data sources and variables considered, followed by the model setup in Section 3.1 and justification in Section 3.2. The results in Section 4 presents the key findings of the analysis, with a discussion on the implications. Section 5 then discusses potential limitations and suggestions for future research. Section 6 provides additional detailed information about the data, model and methodology.

2 Data

2.1 Overview

The data used in this analysis comes from a combination of publicly available procurement data (Investigative Journalism Foundation 2024). The analysis uses the statistical programming language R (R Core Team 2023) and several libraries, including `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2023), `knitr` (Xie 2024), `dplyr` (Wickham et al. 2023), `arrow` (Richardson et al. 2024), `purrr` (Wickham and Henry 2023), `sf` (Pebesma et al. 2024), and `here` (Müller 2020) for data manipulation. `ggplot2` (Wickham 2016), `ggcorrplot` (Kassambara and Patil 2023) and `kableExtra` (Zhu 2024) for visualization. The dataset covers various polls conducted across multiple states, capturing the support for each major candidate—Donald Trump and Kamala Harris—along with detailed attributes of the polls.

2.2 Measurement

The dataset: .

The measurement process begins with: .

2.3 Data Cleaning

This paper first select and rename key variables from raw data to focus on relevant information. The key variables of interest in our analysis include Contract, StartDate, AwardDate, EndDate, Buyer, and Amount, with PreparatoryPhase and ContractTime being variables created.

More information on the data cleaning process can be found in Section 6.

2.4 Outcome Variables

The main outcome variable of interest is: .

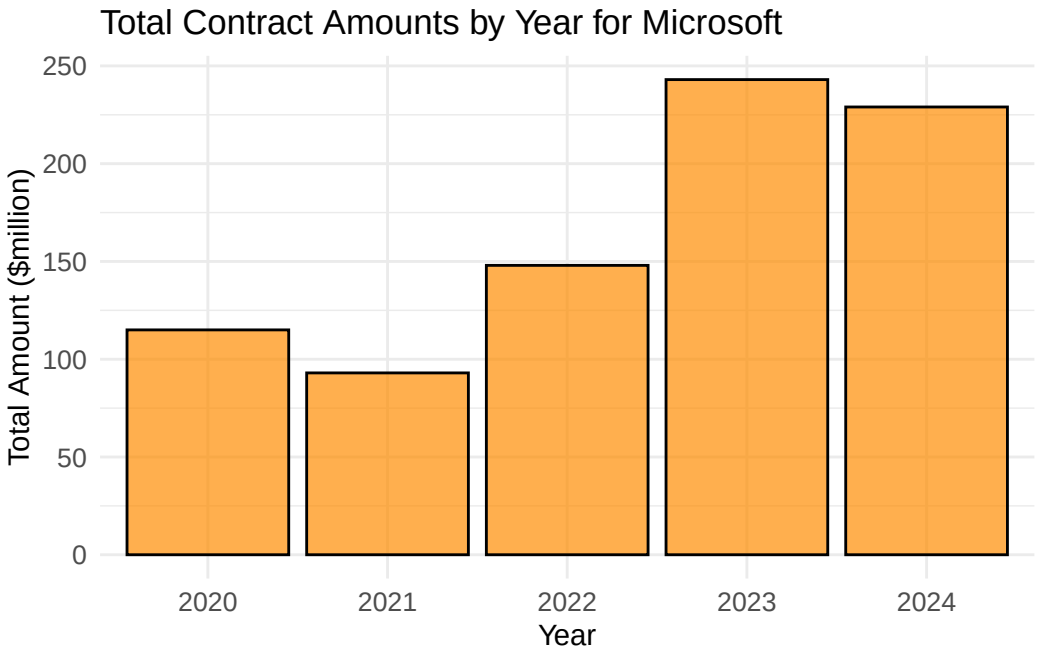

Figure 1: The histogram shows the distribution of the total amount($million) of each year's federal contract (since 2020). [An increasing trend may reflect the growing importance of technology in government operations, including software, cloud solutions, and IT infrastructure.]

2.5 Predictor Variables

In this analysis, several key predictors were identified to evaluate proocurement data effectively.

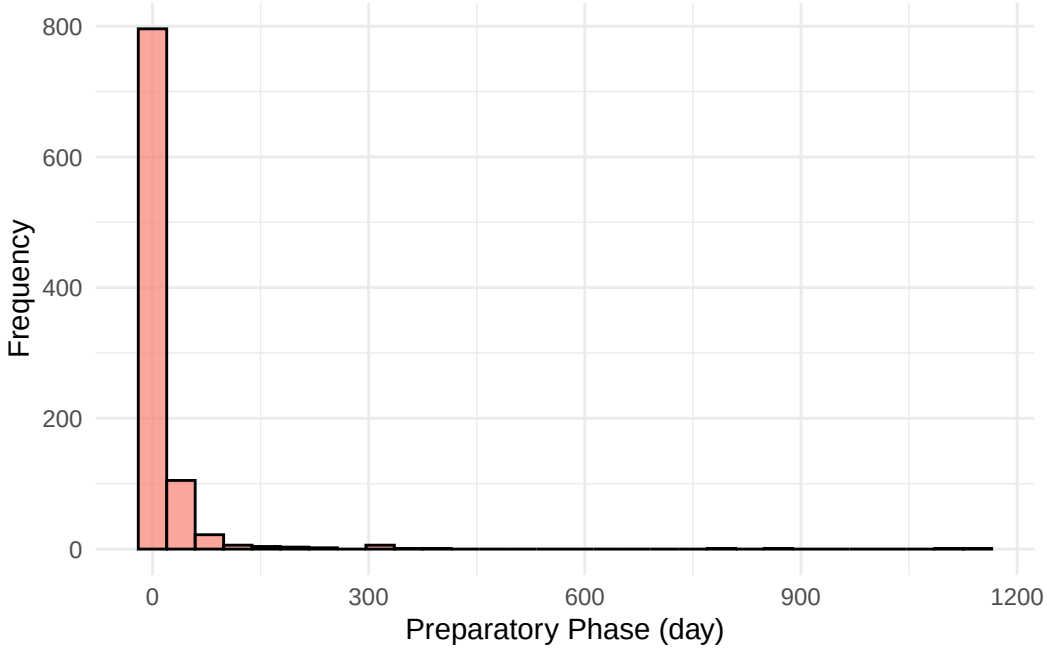

Figure 2: The histogram shows the distribution of preparatory period, which demonstrates the days between start and awarded date. A high concentration around 0 indicates rapid procurement processes or some emergency software services. There are a few outlying contracts with a mobilization period of over one year. These long mobilization periods suggest issues such as delays in the procurement process, a long negotiation period, or administrative approvals.

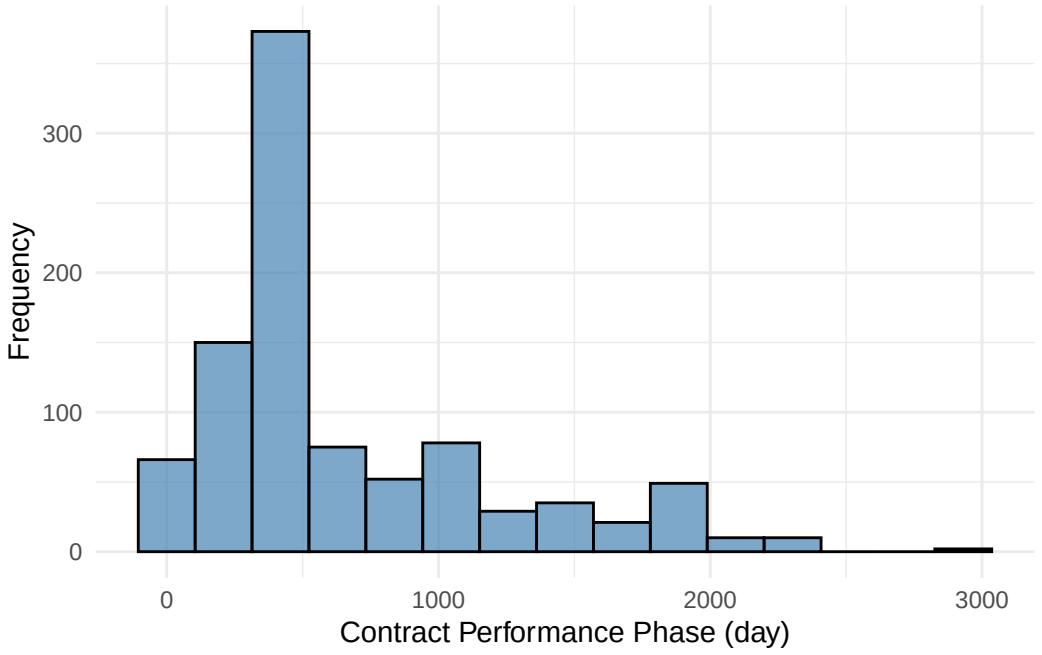

Figure 3: The histogram shows the distribution of performance period, which demonstrates the days between start and end date. The right skewed distribution indicates that the majority of contract duration is within a year. An over 5 years performance period implies a massive project that requires extensive preparation.

3 Model

3.1 Model Set-Up

3.2 Model Justification

3.3 Assumptions and Limitations

3.4 Model Validation

4 Results

5 Discussion

6 Appendix

6.1 Data Cleaning Notes

6.2 Additional Tables & Figures

6.3 Methodology

6.4 Idealized Methodology

References

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Investigative Journalism Foundation. 2024. "Procurement – Microsoft." Toronto, Canada. https://theijf.org/procurement/supplier/microsoft?region=Federal.

Kassambara, Alboukadel, and Indrajeet Patil. 2023. *Ggcorrplot: Visualization of a Correlation Matrix Using 'Ggplot2'*. https://doi.org/10.32614/CRAN.package.ggcorrplot.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Pebesma, Edzer, Roger Bivand, Etienne Racine, Michael Sumner, Ian Cook, Tim Keitt, Robin Lovelace, et al. 2024. *Sf: Simple Features for r*. https://doi.org/10.32614/CRAN.package.sf.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Lionel Henry. 2023. *Purrr: Functional Programming Tools*. https://CRAN.R-project.org/package=purrr.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.