# Data Science for Business - Replication Project

Tony Osei, Christian F. R. Casas

2025-06-06

---

### About This Project

This is part of a research exercise for replicating graphs and tables from (Calvano et al. 2020) using **R** and **Quarto**.

We focus on:

- Summary statistics
- Replication code using R
- Data cleaning and transformation

---

## Research Replication

We aim to replicate tables and charts taken from the paper ***The Effect of Gender on Credit Access*** (Calvano et al. (2020)), that was Published in *American Economic Review: Microeconomics*

Please follow the next link to find the replication package and further information:
**DOI:** 10.1257/mic.20210158

---

## Tools Used

- **R** (data analysis & visualization)
- **Quarto** (reproducible publishing)
- **GitHub Pages** (to share everything online)
- **Libraries**:

-**haven:**: this library created by Wickham, Miller, and Smith (2025) allows R to read and write data files from statistical software like Stata, SPSS, and SAS.

-**dplyr:** provides a grammar of data manipulation, making it easy to filter, select, summarize, and transform data frames. (Wickham et al. (2025))

-**ggplot:** As stated by Wickham (2016) powerful and flexible system for creating static data visualizations using the grammar of graphics.

- **janitor:** Offers simple functions for cleaning and formatting data, especially column names and tabular summaries.

---

## Repository

View the full repository here:
github.com/christianfcasas/ProjectData

---

## View the Presentation

Click on the presentation you want to view:
Christian
Tony

---

## Downloadable Version ('pdf' and 'html')

**pdf:**
Index PDF Version
Presentation PDF Version

**html:**
Index PDF Version
Presentation PDF Version

---

## Next Steps

- Expand replication to additional tables
- Align replication figures tables to those in the paper
- Fix replication issues
- Create an academic-style appendix

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. 2020. "Artificial Intelligence, Algorithmic Pricing, and Collusion." *American Economic Review* 110 (10): 3267–97. https://doi.org/10.1257/mic.20210158.

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

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2025. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, Evan Miller, and Danny Smith. 2025. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://haven.tidyverse.org.