OPTIMIZING THE COMMODITY FLOW SURVEY, v2

Economic Reimbursables Division / Commodity Flow Branch

James Hinckley — Branch Chief
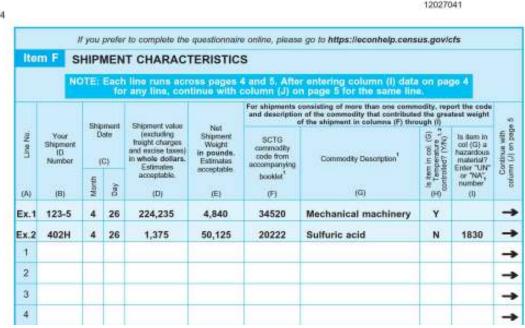
CHRISTIAN MOSCARDI

NYU CUSP

MS Urban Informatics

CIVIC DIGITAL FELLOWSHIP

United States Census Bureau

OVERVIEW

- Commodity Flow Survey

- Commissioned by BTS

- Conducted every 5 years (2017, 2022)

- Respondents provide sampling of shipments from each quarter

OVERVIEW
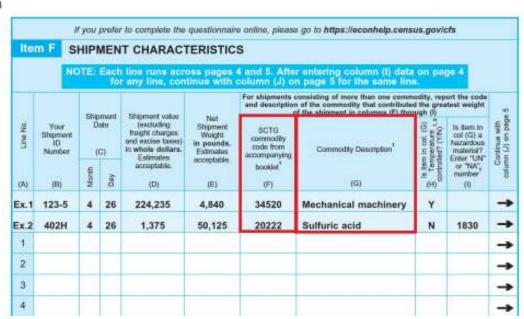
- Commodity Flow Survey

- Commissioned by BTS

- Conducted every 5 years (2017, 2022)

- Respondents provide sampling of shipments from each quarter

OVERVIEW

For shipments consisting of more than one commodity, report the code and description of the commodity that contributed the greatest weight of the shipment in columns (F) through (I)

SCTG commodity code from accompanying booklet[1] (F)	Commodity Description[1] (G)	Is item in col. (G) Temperature controlled? (Y/N)[1,2] (H)	Is item in col (G) a hazardous material? Enter "UN" or "NA", number[1] (I)
34520	Mechanical machinery	Y	
20222	Sulfuric acid	N	1830

OVERVIEW

ITEM G - Other Clarifying Information

"Pulling this information was a huge spend of time and resources."

"Just glad this is over!!"

OBJECTIVE

Using Machine Learning, can we automate the assignment of SCTG codes to shipments?

OBJECTIVE

(spoiler alert)

OBJECTIVE

(spoiler alert)

YES!

METHOD

- Training Data: **6.4 million** labelled shipment records from 2017 CFS

- Bag-of-Words, Logistic Regression (logit) model

- Use shipment description and NAICS Code
  - "fasteners"

- Clean records, de-duplicate, disambiguate: ~**400,000 unique training records**

- Initial results: **50% accuracy** (on 40,000 unseen test set records)

METHOD

- Investigate poorly performing codes

- **40994**: Sewing and knitting needles (includes for machines) crochet hooks, hook and eye **fasteners**, safety pins, straight pins, buttons, buckles and clasps, tubular and bifurcated rivets, **snap-fasteners**, zippers, and similar notions.

METHOD

- **33310**: Nails, screws, bolts, nuts, washers, staples except in strips, and similar **fastening** articles

METHOD

- Manually validating, about **50%** of items labelled 40994 by respondents were **miscoded.**

- **However, the model was getting it right!**

Commodity Code Search (Commodity Codes List)

To help find your commodity code and its description, enter SCTG code or keyword below.

Search by SCTG code or keyword: fastener   Search

Results found: 2 for 'fastener'

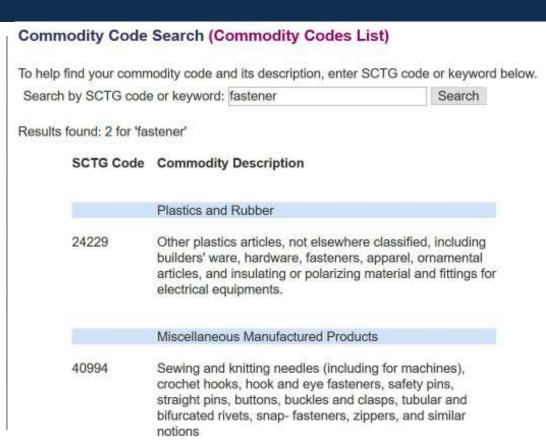| SCTG Code	Commodity Description
Plastics and Rubber	
24229	Other plastics articles, not elsewhere classified, including builders' ware, hardware, fasteners, apparel, ornamental articles, and insulating or polarizing material and fittings for electrical equipments.
Miscellaneous Manufactured Products	
40994	Sewing and knitting needles (including for machines), crochet hooks, hook and eye fasteners, safety pins, straight pins, buttons, buckles and clasps, tubular and bifurcated rivets, snap- fasteners, zippers, and similar notions

CIVIC
DIGITAL
FELLOWSHIP

RESULTS

- **Proof-of-concept**: ran model on 170,000 unlabelled records

- 70,000 with confidence score above predefined threshold

- Manually validate a sample of 350 unique records

- **89% accurate!**

- Batch-edits have saved ~**500 hours** of manual editing time, **$35,000**

RESULTS

- 2022 CFS: save respondents **50,000+ hours**, **$2.1 million** in respondent lookup costs

- This **does not include** Census' editing costs, cleaning up invalid / messy data

- And, we get more, more accurate, and **real-time** shipment data!

- Cleaner data (via **Amazon's Mechanical Turk**) will improve these savings
  - **Better model == more savings**

ACKNOWLEDGMENTS

- **James Hinckley**, CFS Branch Chief

- **Jessica Young/Berin Linfors**, CFS Section Chiefs

- **Ja Lee**, **Ben Schultz**, CFS Analysts

- **Julie Parker**, **Joy Sharp**, **Mehdi Hashemipour**, BTS

- **Cavan Capps**, Census Big Data Lead

- **Andrew Cukierwar**, 2017 Civic Digital Fellow

THANK YOU!

Twitter: @c_moscardi
Github: cmoscardi
Email: christian.moscardi@gmail.com