

Luce County Timber Fund

Financial Data Pipeline & Analysis

2026 Innovate Michigan! Summit • July 9, 2026

Michigan State University • CMSE 495 — Community-Engaged Data Science

Team: Karen Sandy • Christine Yoo • Tianjian Xing • Corei Laury

Student Technical Assistant: **Bibin Varghese**, MS Data Science | varghe64@msu.edu | **Community Partner:** Tammy Lasley-Henry, Luce County EDC



Background & Motivation

Why Luce County needed this project



Luce County manages a multi-thousand acre of state-owned timber land. Revenue from timber harvesting funds a Revolving Loan Fund (RLF) for local economic development. As the county prepares for lease renewal with the State of Michigan, it needs a documented 40-year financial history.

Lease Renewal

The county must demonstrate responsible fund management to renew its state timber lease.

40 Years of Paper

Records from 1983–2022 exist only as scanned PDFs: checks, invoices, and journal entries.

Two Disconnected Systems

Modern records (2014–2025) live in QuickBooks. Older records are scanned paper. No unified history.

Automation Is Necessary

Manual transcription of 40 years is impractical. A reusable pipeline lets the county process documents independently.

Two Data Sources

Together covering the full 40-year financial history



Scanned PDFs

1983 – 2022

Format: Scanned images of paper documents

Volume: Hundreds of pages across 5- year folders

Types: Checks, journal entries, transfer vouchers, ledger registers, bank statements

Challenge: Handwriting, red pen annotations, mixed layouts per PDF

Status: Pipeline built & tested on 1990 document



QuickBooks Export

2014 – 2025

Format: Structured CSV/Excel with 13 columns per row

Volume: 351 transactions, fully cleaned and categorized

Types: Deposits, checks, transfers, journal entries

Strength: Complete, structured, verifiable — most reliable source

Status: Fully analyzed with interactive dashboard

How the OCR + AI Pipeline Works

Three core stages: Convert → Read → Extract



1

PDF → Images

pdf2image

Each PDF page converted to high-res image so the computer can "see" the document.

2

OCR

Tesseract

Computer reads text from the image — like squinting at a photocopy. Imperfect for handwriting.

3

Send to AI

Ollama + Phi-3

Full page text sent to local AI model. No internet required — runs entirely on laptop.

4

AI Extracts

Phi-3 prompt

AI finds date, amount, payee, direction (IN/OUT), category, and rates its own confidence.

5

Export

pandas + CSV

Results saved to a spreadsheet. Low-confidence rows are flagged for human review.



NDA-Compliant: Phi-3 runs 100% on your laptop via Ollama — no financial data ever leaves the machine.

Pipeline Test: 1990 Document

1990_002651.pdf • 20 pages • 3.3 MB • Mixed document types



95%

Amounts
extracted

61%

High confidence
Classifications
**Self-reported, not verified*

41

Total transactions
found

2

Pages completely
failed

Document Types Found in One PDF

Physical Check

*[Timber Company] → Luce County, \$XX,XXX
— Jan 1990



*Cursive signature, handwritten amounts,
red pen overlay*

Journal Entry

*Transfer Voucher — [Date]
Debit/Credit table with fund codes



*Printed table but with handwritten
annotations*

Payment Voucher

*Page 18: Stumpage transfer
\$XX,XXX — account codes XXX.000 / XXX.000



*Account codes misread as dollar amounts by
OCR*

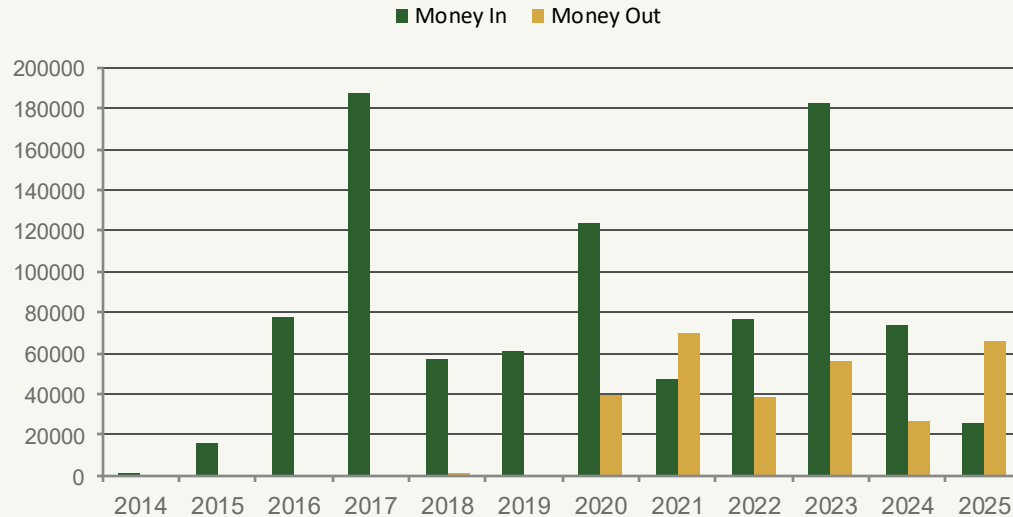
**Values anonymized for NDA compliance.*

QuickBooks Analysis: 2014-2025

351 transactions cleaned, categorized, and visualized



Money In vs. Money Out by Year (\$)



Transaction Categories

Money In

~146 rows (42%)

Interest, timber sales

Money Out

~31 rows (9%)

Tax, misc expenses

Transfer

~38 rows (11%)

Between accounts

RLF Payback ✓

132 rows (38%)

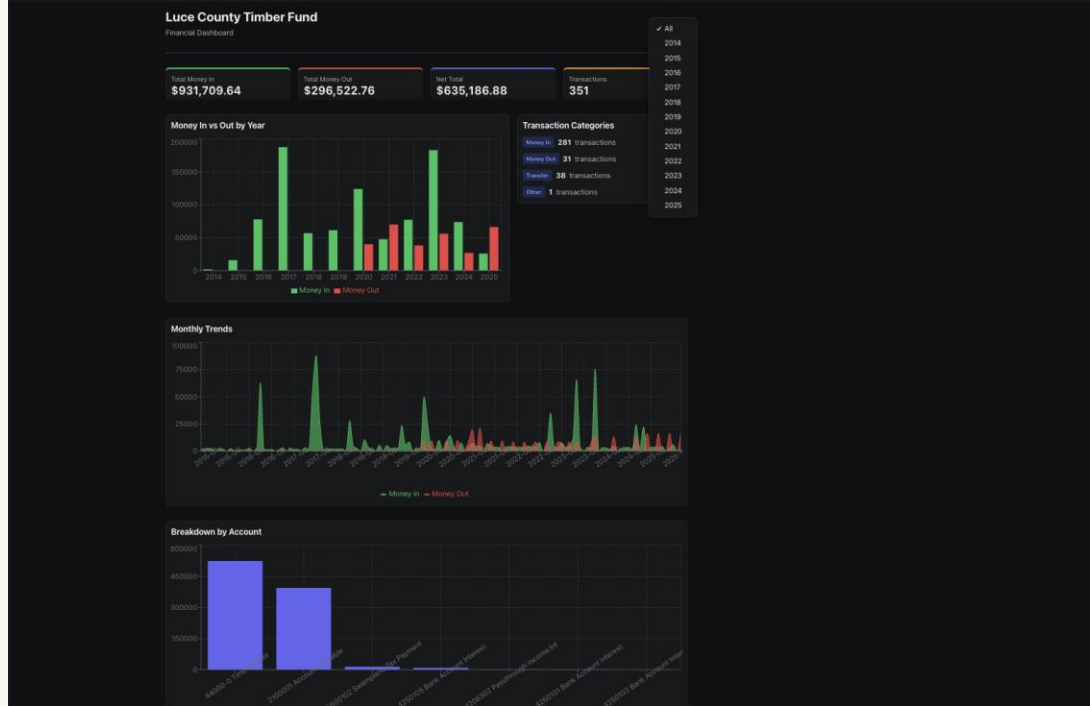
Confirmed: loan repayments



Confirmed: 132 "RLF Payback" transactions are loan repayments back into the timber fund. Classified as Money In per Tammy Lasley-Henry. 8 additional A/P deposits with no memo remain under review.

Interactive Dashboard

Built with Reflex — runs at localhost:3000 — fully NDA-compliant



Dashboard Features

KPI Cards

Money In, Money Out, Net Total, Transaction Count at a glance

Yearly Bar Chart

Side-by-side Money In vs. Out (2014–2025)

Monthly Trend

Area chart showing seasonal cash flow patterns

Account Breakdown

Which accounts drive most inflows/outflows

Year Filter

Isolate any single year across all charts instantly

No internet connection required. No data leaves the machine. Works on any laptop.

Limitations & Honest Assessment

What the pipeline can and cannot reliably do



Area	What Works	What Needs Review
QuickBooks 2014–2025	Near-perfect extraction & categorization, RLF Payback (38%) confirmed as Money In	8 unlabeled A/P deposits under review
Printed journal entries	Amounts, dates, directions reliably extracted	Account codes sometimes confused with dollar amounts
Physical checks (scanned)	Large amounts usually found	IN/OUT direction sometimes reversed when both parties named
Handwritten records	Clear printed numbers extracted	Cursive, faded ink, red pen annotations cause OCR failure
Date extraction	Standard MM/DD/YYYY formats work well	Short formats (2/6/90) and cover-page dates missed
Overall reliability	Speeds up transcription significantly vs. manual entry	Not an automated auditor — every row should be spot-checked before use in official reporting



Hardest year range: 1983–1990 — mixed document types + handwriting + no digital cross-check = most human review needed

What Luce County Receives

A reusable system the county operates independently



All tools run locally. No subscriptions, no cloud accounts, no internet required for processing. The county owns and controls everything.



Interactive Dashboard

localhost:3000 — Reflex

View 2014–2025 trends: Money In vs. Out by year and month. Year filter. No coding required.



OCR Extraction Pipeline

smart_extraction.py

Run on any new scanned PDF. Output is a CSV the treasurer opens in Excel.



Cleaned Transaction Data

cleaned_data.xlsx — 351 rows

All QuickBooks transactions categorized. Shared as Excel file under NDA.



Analysis Notebooks

Jupyter + Python

Step-by-step documented analysis. Reproducible. Runs in the conda environment.



No Dependencies

Python + Ollama + Tesseract

All free tools. No subscriptions. Runs offline. Works on any Mac or Windows laptop.



NDA-Compliant by Design

100% local processing

No county data ever transmitted externally. Private GitLab repo.

 *First-pass assistant, not an automated auditor. Low-confidence rows and pre-1990 records need human review before official reporting.*

Next Steps

- 1 Complete batch processing of 2001–2010, 2011–2014, and 2015–2022 scanned PDF folders
- 2 Merge both pipelines into a unified yearly summary table covering 1983–2025
- 3 Add image preprocessing (contrast, deskew) to improve OCR on pre-1990 handwritten records
- 4 Explore vision-capable local AI (e.g., LLaVA via Ollama) to replace OCR step — directly reads scanned pages for better handwriting accuracy
- 5 Package dashboard + pipeline with installation guide for independent county use

Thank You

Questions & Discussion

Community Partner:

Tammy Lasley-Henry
Luce County EDC

Advisor:

Dr. Justin Gambrell
MSU CMSE 495

Special thanks to:

MSU Center for Community and Economic Development (CCED)
MSU REI (Regional Economic Innovation)

Data Sources:

QuickBooks (2015–2025)
Scanned PDFs (1983–2022)



Southwest Lansing

Data & Investment Tool

Community Investment Summit • July 9, 2026



Michigan State University • Center for Community & Economic Development

Capstone Team: Mason Lee • Kyle Landolt • Vibu Darshan • William McNeilly

Student Technical Assistant: **Bibin Varghese**, MS Data Science | varghe64@msu.edu | Contact: Kevin Brooks, University Outreach & Engagement



What Is This Project?

Southwest Lansing has ~18,000 residents across 7 census tracts. Community partners needed a free, open-source tool to visualize their data clearly — in one place, without a data scientist.

Web Application (Capstone)

Interactive map built by CMSE 495. Filters for Food, Income, Race, Household, Health. Live on Render.com with ArcGIS data hosting.

Data Pipeline (Student TA)

ACS Census API pipeline: 7 tracts • 11 map layers • Poverty & digital equity analysis • Correlation charts.

Community Partners

Built for CCED, SWAG (SW Action Group), LEAP, and Ingham County to make data-driven investment decisions.

Anchor Institution Opportunity

MSU is positioned to direct CDFI and anchor institution funds toward measurable, place-based outcomes in Southwest Lansing. This tool provides the evidentiary foundation to make that case to funders and policymakers at this Summit.



Where the Data Comes From

U.S. Census Bureau • ACS 5-Year Estimates (2019–2023)

All data was pulled directly from the U.S. Census Bureau's official API (api.census.gov) for 7 tracts in Ingham County (FIPS 26065). 5-year estimates provide more reliable small-area statistics than single-year estimates.

Tract Code	Label	Coverage	Block Groups
001703	Tract 17.03	Full	1, 2, 3
003601	Tract 36.01	Full	1, 2, 3
003602	Tract 36.02	Full	1, 2, 3
003700	Tract 37	Partial *	3, 4, 5 only
005100	Tract 51	Full	1, 2, 3
005201	Tract 52.01	Full	1, 2, 3, 4, 5
007000	Tract 70	Partial *	4, 5 only

ACS Variables Used

B19013	Median Household Income
B01003	Total Population
B17001	Poverty Status
B25003	Homeownership Status
B25071	Rent Burden (% income to rent)
B25077	Median Home Value
B15003	Educational Attainment
B23025	Employment Status
B28001/B28002	Internet Devices & Subscriptions
B28004	Internet Access by Income

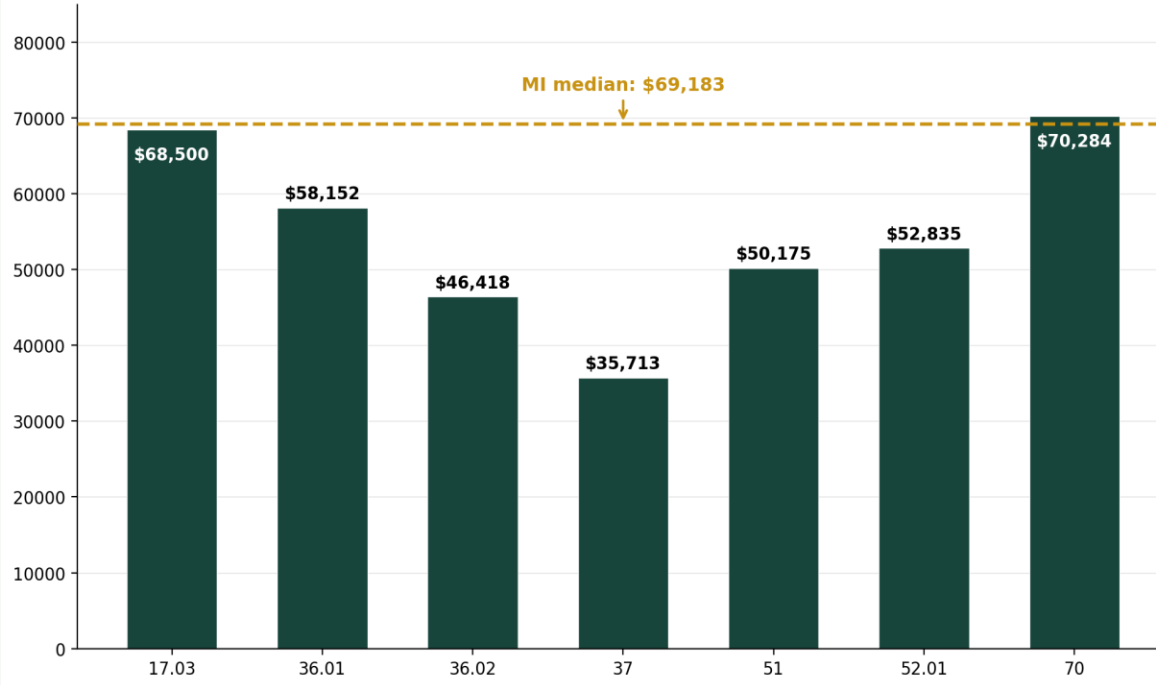
* Partial tracts: ACS data covers the full tract boundary. Figures may not precisely represent the project area. Block group-level refinement is a recommended next step.



Income & Poverty Across 7 Tracts

Michigan benchmark: median \$69,183 • poverty rate 13.5%

Median Household Income (\$)



Tract	Income	Poverty	Priority
17.03	\$68,500	9.3%	Stable
36.01	\$58,152	30.7%	HIGH
36.02	\$46,418	28.9%	HIGH
37	\$35,713	24.0%	HIGH
51	\$50,175	14.4%	Moderate
52.01	\$52,835	20.4%	HIGH
70	\$70,284	8.7%	Stable

All tracts except 70 fall below Michigan median income. Poverty rates in tracts 36.01 and 36.02 are more than double the state average (13.5%); tract 37 is also significantly elevated at 24.0%



The Digital Divide: Not Infrastructure — Affordability

Source: ACS B28001, B28002, B28004 (2019–2023)

81–93%

Broadband coverage
across all 7 tracts

29.8%

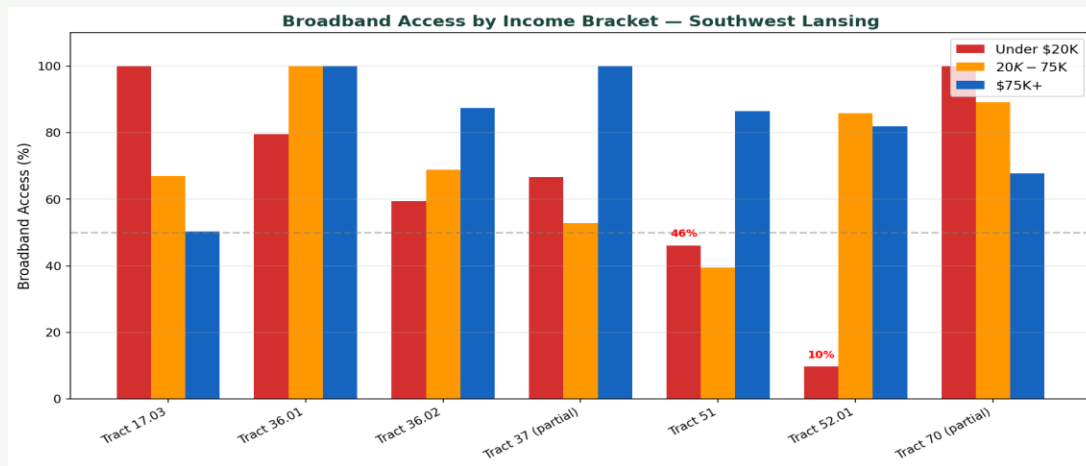
Tract 36.02
smartphone-only

90.3%

Tract 52.01 low-income
households offline

24.6%

Tract 36.01
cellular-only internet



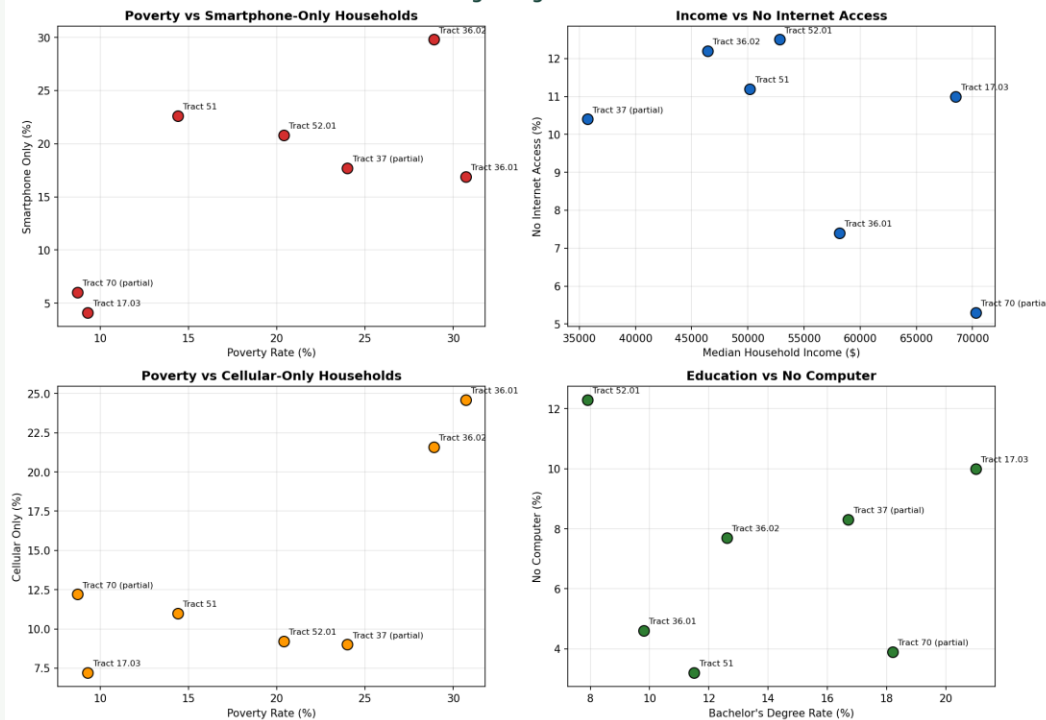
Key finding: Broadband infrastructure reaches 81–93% of all homes. The barrier is cost — Tract 52.01 (10% broadband for households under \$20K) represents a near-total digital blackout among the lowest-income residents.



Correlation Analysis — Digital Divide Drivers

Poverty, income, education vs. digital access across 7 tracts

Southwest Lansing — Digital Divide Correlations



Charts from ACS 5-Year Estimates via Census API. Each data point = one census tract.

Poverty → Smartphone-only

Tract 36.02 has the highest smartphone-only rate (29.8%) with 28.9% poverty. Tract 36.01 has the highest poverty rate (30.7%) but a lower smartphone-only rate (16.9%).

Poverty → Cellular-only

High-poverty tracts (36.01, 36.02) rely on expensive mobile data instead of home broadband.

Income anomaly (17.03)

High income but 11% no-internet — likely older homeowners opting out, not affordability.

Education alone ≠ access

Tract 17.03 has the highest education (21%) but still 10% no-computer — no independent effect.



Priority Zones for Investment

HIGH PRIORITY

Tract 52.01

90% low-income offline
Ownership 45.6%
Values \$87,700

Tract 36.01

Poverty 30.7% (highest)
24.6% cellular-only
Rent burden 45.2%

Tract 36.02

29.8% smartphone-only
Poverty 28.9%
40.5% under-\$20K offline

Tract 37

Lowest income \$35,713
33.3% low-income offline
Poverty 24.0%

MODERATE

Tract 51

19.9% unemployment
53.9% low-income offline
22.6% smartphone-only

RELATIVELY STABLE

Tracts 17.03 & 70

Income near/above state median
Low poverty & digital disconnection
Broad coverage across income brackets

Recommended Investment Actions

Device Distribution

Computers & laptops for Tracts 36.02, 51, 52.01

Broadband Subsidies

Affordable home internet for Tracts 36.01, 36.02, 52.01

Comprehensive Package

Device + broadband + digital literacy for Tract 52.01

Housing Investment

Homeownership programs for Tracts 36.02, 37, 52.01



The Web Application

Live: msu-southwest-lansing-project.onrender.com

A free, open-source tool. Any community partner can explore Southwest Lansing data without a data scientist or specialized software.

Interactive Map

MapLibre GL JS + ArcGIS. Block-group and block-level selection across 22 census block groups in 7 tracts.

Multi-Layer Filters

Population, Income, Race, Household Size, Food Access, and Health resources — all filterable.

Flexible Area Selection

Block groups, block levels, named neighborhoods (Lansing Neighborhood Org.), or custom radius tool.

Open Source & Documented

Full GitHub repo: [landoltk/MSU-Southwest-Lansing-Project](https://github.com/landoltk/MSU-Southwest-Lansing-Project). Documented for future teams to expand.

Hosted on Render.com

Free-tier hosting. Redeployable from GitHub. Data hosted on ArcGIS under MSU account.

Data Catalog

Excel + Word catalog of all explored data sources — descriptions, links, usage notes for future teams.



What We Delivered

Web Application

Live interactive map on Render.com — 6 filterable data categories

4 Clean CSV Datasets

Income/poverty, housing/education, internet access, internet-by-income

Correlation Charts

4 scatter plots + broadband-by-income bar chart (.png) from ACS data

Combined Final Report

11-section Word document with full methodology, tables, and conclusion

Jupyter Notebook

Full ACS data pipeline with Census API calls, cleaning, and export

11-Layer Interactive Map

Toggle-able choropleth (.html) with clickable tract popups showing all metrics

Data Catalog

All explored data sources documented — links, descriptions, notes for next teams

This Presentation

11-slide Summit deck — July 9, 2026 — MSU/CCED branded with real logos



Conclusion & Call to Action

What Southwest Lansing should prioritize for investment

Census data was drawn from the U.S. Census Bureau ACS 5-Year Estimates (2019–2023) via the official Census API, covering 7 tracts in Ingham County. The data consistently points to the same investment priorities.

1 Device Access First

Broadband exists (81–93%). The gap is devices, not pipes. Distribute computers & laptops to Tracts 36.02, 51, and 52.01.

2 Broadband Affordability

Subsidize home broadband in Tracts 36.01, 36.02, and 52.01 — where cellular-only rates show residents can't pay for home internet.

3 Comprehensive: Tract 52.01

90% of low-income residents offline. Needs device + broadband + digital literacy combined — the most urgent intervention.



MSU as Anchor Institution

MSU is positioned to direct CDFI resources and program investment toward measurable, place-based outcomes. This tool provides the evidentiary foundation.

Next Steps

1. Integrate ACS data layers into the capstone FastAPI application
2. Pull block group-level data for partial tracts 37 & 70
3. Add food access (USDA FARA) and health insurance coverage layers
4. Cross-reference FCC Broadband data with ACS adoption rates by tract
5. Transfer ArcGIS hosting from MSU account to CCED-managed email

Thank You

Questions & Discussion

CCED Partners:

Kevin Brooks & John Melcher

Data Source:

ACS 5-Year Estimates 2019–2023
api.census.gov

GitHub:

github.com/landoltk/MSU-Southwest-Lansing-Project

Special thanks:

Grace Densham (food/garden data)
Amanda Tickner (ArcGIS access)

Michigan State University • Center for Community & Economic Development

Capstone: Mason Lee • Kyle Landolt • Vibu Darshan • William McNeilly | Student TA: Bibin Varghese

