
PROJECT PLAN: KANSAS COST OF LIVING

Joseph Mwagura

Joseph.Mwagura@ks.gov

William Duncan

williamduncan@ku.edu

ABSTRACT

This project seeks to provide insight into the cost of living in all 105 counties in Kansas. While national statistics exist to estimate cost of living changes over time, these estimates do not adequately reflect changes in rural areas and cannot capture the variation in cost of living changes happening across Kansas. This project seeks to develop a web platform and mobile app for Kansans to upload receipts for the costs they experience in their lives. The information from these receipts is transformed into a consumer price index (CPI)-like estimate. These granular insights inform how prices have changed for Kansans across a range of products that typically go into cost-of-living calculations.

1 INTRODUCTION

In order to build a cost of living estimate for all 105 counties in Kansas, this project will undertake an effort to build a web platform and mobile app where granular data can be collected by crowdsourcing individual purchases from citizens across the state. These data collection efforts will examine prices of goods and services that Kansans frequently purchase.

Working with the Kansas Data Science Consortium (KDSC), the project will build a web platform and an app where the data can be collected through submissions of electronic files such as screenshots of physical receipts and e-receipts, process the electronic files using an Object Character Recognition (OCR) algorithm, to convert the text data to structured data using a Large Language Model (LLM), and then use the structured data to create a cost-of-living index for each county.

After the web platform and app are built, the project will pilot the data collection concept with students in classes at every community college, liberal arts college, and public four-year institution in Kansas. These students will upload receipts from the costs they face in their daily lives and participate in a statewide data project to analyze the price data from their purchases.

2 BASELINE OR INITIAL ANALYSIS

To begin this project, the Kansas Data Science Consortium (KDSC) will build a web platform and an app for the Kansas Department of Commerce where Kansans can upload receipts of their purchases.

The KDSC will use an OCR algorithm to turn images of receipts uploaded to the app and web platform into text data. Then, using an LLM, background code will categorize items listed on the receipt by product type and record the price paid for each item. This will turn unstructured text data into a structured dataset (i.e. sales tax and store name into columns).

In the initial analysis, the receipts used to test the algorithm will come from students at partner institutions with the KDSC. This will enable the KDSC and the Kansas Department of Commerce to work through glitches in the web platform and the app, the OCR algorithm, the LLM used to classify products.

3 FINAL ANALYSIS

Once the web platform and the app is built, the OCR and LLM algorithms are validated, the

KDSC will work to feed the structured data into a cost of living formula that produces an estimate for each county in Kansas as well as a statewide estimate. This index will follow best practices in weighting baskets of goods that are used in other indices such as the Consumer Price Index. The index will be made available on the web platform and the app and will update in real time as Kansas upload their receipts.

4 FINAL GOALS & EVALUATION

The deliverable for this project is an app where Kansans can upload receipts from purchases they make in stores and online. The app will provide Kansans the ability to look at data for their county over time, look at product categories over time, and make comparisons between sectors and areas of the state.

Additionally, the app will compare the Kansas-specific metric with federal metrics about cost of living for urban areas, for rural areas, and for different baskets of goods so that Kansans can see how their region compares with national estimates.

5 RELATED WORK

6 DATA & TECHNICAL REQUIREMENTS

This project will require two sets of data:

- The first dataset is publicly available from Federal sources that track Consumer Price Indices and cost of living estimates across the nation.
- The second dataset is crowdsourced receipts from Kansans who share the receipts from their purchases as images and electronic files on the app.

This project will require various software:

- The project team will decide how to build and maintain an application that Kansans can download.
- The project team will decide what OCR algorithm to deploy to process receipts that Kansans share on the app.
- The project team will determine what LLM algorithm to deploy to categorize product and expenditure data from the processed receipts
- The project team will decide what statistical software package to use to analyze the structured price data.