# Project Scope

This project aims to detect and reduce e-commerce delivery fraud at Walmart by analyzing delivery data from Central Florida. The goal is to identify patterns indicating items marked as delivered but not received by customers. Findings will help create a model to minimize theft and fraud across all U.S. regions.

# Contextualization

Walmart is the largest retail chain in the United States, generating an average of US$ 1.6 billion in daily revenue. In terms of in-store sales, Walmart averages US$ 17,000 per second, US$ 1.1 million per minute, and US$ 68 million per hour.

Its online platform and retail sales contribute a large portion of its profits. Walmart is also the largest food retailer in the U.S., with over US$ 264 billion in food sales last year.

According to a recent Lending Tree survey, self-checkout theft is a real issue, and individuals who engage in it are likely to repeat the behavior. Walmart has faced growing retail theft losses — estimated at US$ 3 billion in 2021, US$ 6.1 billion in 2022, and US$ 6.5 billion in 2023 — showing an increase of US$ 400 million in theft-related losses over the last year.

As a data scientist working in Walmart’s e-commerce division, you’ve identified that the largest proportional increase in theft occurred in online purchases where customers report missing items from their orders. Of the total theft growth from 2022 to 2023, 53% came from e-commerce transactions.

You were assigned to lead a project to reduce fraud and theft in Walmart’s e-commerce deliveries. It was identified that these thefts primarily occur during the delivery process.

The main goal of the project is to identify possible frauds in deliveries made by Walmart in Central Florida, which will serve as a model for other U.S. regions if successful. The focus is on analyzing delivery data to detect patterns and anomalies that could indicate that items reported as delivered were not actually received by customers. Walmart has received numerous consumer complaints about incomplete deliveries, and through data analysis, aims to determine whether the issue stems from the delivery drivers or other causes.

# Project Guideline

* **First Week (Week 43, 2025):** During the first week, it will be done an overall analysis for the project. We’ll define goals, analyse the databases to detect possible anomalies, brainstorm possible hypotheses and structuring the project.
* **Second Week (Week 44, 2025):** The second week will be dedicated to *data preparation*, where we will be analysing our databases and doing important data manipulations using *SQL*.
* **Third and Fourth Week (Week 45 - 46, 2025):** We will dedicate 2 weeks for the *EDA (Exploratory Data Analysis)*, where we will export our data to *Jupyter Notebook* and perform data cleaning, initial data analysis, data segmentation and data visualization to be displayed in the final printed report.
* **Fifth and Sixth Week (Week 47 - 48, 2025):** Those weeks will be