Machine Learning Solution For UPS - A Logistics Industry Business Problem

ECON1612 | Big Data, Machine Learning and Society

Presented by: Group01_SGS04




Le Dieu Ha | Ho Xuan Anh | Nguyen Truc Anh | Ngo Phuc Thinh

TEAM CONTRIBUTION

Student ID	Full Name	Parts Contributed	Description and Evaluation	Contribution %	Signature
s3979364	Le Dieu Ha	Question 2a, 2b 2c, Conclusion	Ha collaborated in the initial research phase to identify problem root causes and propose suitable models. All members have to contributed variable ideas during data collection. She managed group files to ensure consistency. Ha was responsible for researching and designing slides on the **learning task, candidate models, business constraints, variable descriptions, and causal claims**, ensuring alignment with both academic and business requirements.	100%	hale
s4140972	Ho Xuan Anh	Code, EDA, Question 2a, 2c	Xuan Anh was responsible for the **entire coding** component of the project and played a major role in **data collection** and preliminary **data cleaning** using Excel. All members have to contributed variable ideas during data collection. He designed slides for the **validation framework, model results, limitations, calibration, robustness checks, and AI evaluation**.	100%	Xanh
s3977930	Nguyen Truc Anh	Data, Question 1, 2b	Truc Anh contributed to early research on problem diagnosis and model selection. She researched and prepared slides on **data cleaning** and **descriptive statistics** for each variable, and **supported data interpretation within the code**. All members have to contributed variable ideas during data collection. Truc Anh also led the refinement and slide development for the **company and industry context**, strengthening the background section.	100%	Truc Anh
s3990389	Ngo Phuc Thinh	Lead, data, EDA, Question 1, 2b, 2c	Thinh led the early research on UPS, establishing the industry and company context for the team. He contributed key insights that shaped the overall structure of the presentation. Thinh guided data collection and variable selection, **supported EDA** by summarising results and explaining code functions, and assisted with data cleaning and coding. All members have to contributed variable ideas during data collection. He was the **main contributor to ethical considerations, and implications for business decisions.**	100%	Phuc Thinh

1. Introduction | Mission 1.1 UPS At a Glance

Market leader in Delivery and Courier services

Market Cap: USD 117.1B

One out of the four largest Delivery and Courier companies in the US

Operating in 200+ countries and territories worldwide

Total revenue (FY2024)

USD 91.1B

Operating profit (FY2024)

USD 8.9B

UPS delivers approximately 19.1 million packages per day in the US

Accounted for 34.3% of US Couriers market share

(Pitney Bowes 2025)

Rank No.1

BEST CUSTOMER SERVICE

2026

Forbes

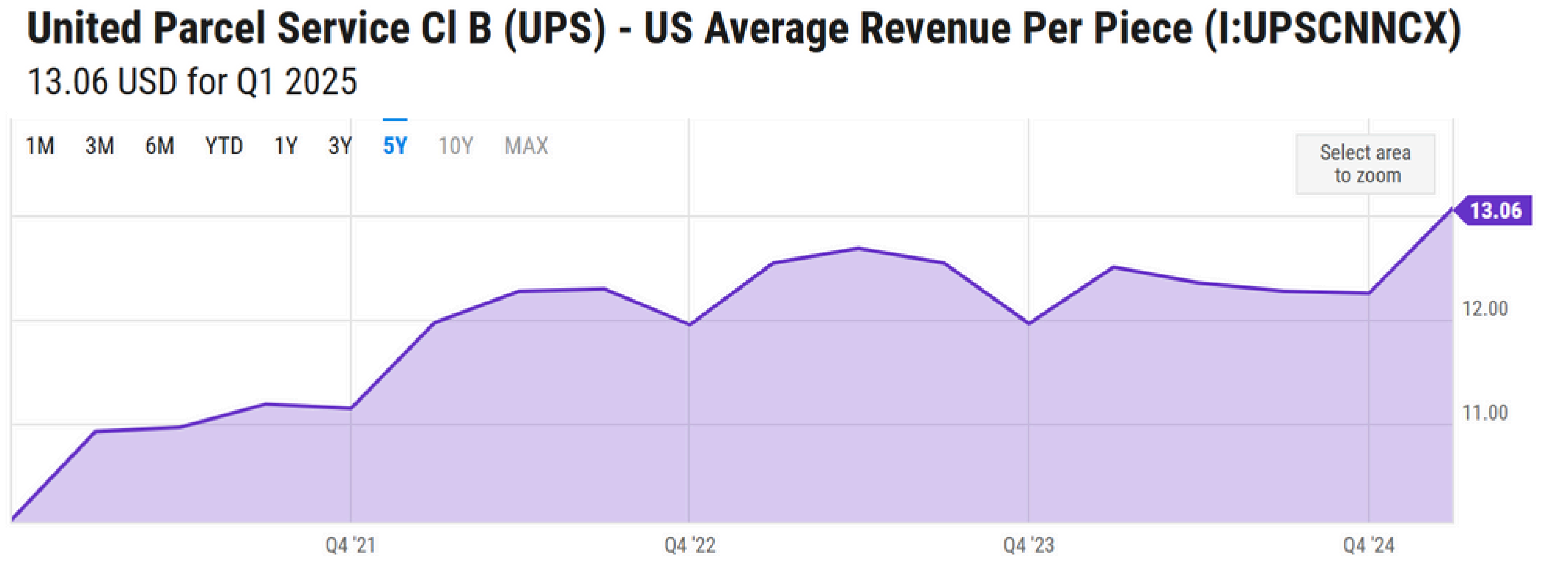
(UPS 2025; Forbes 2025)

1. Introduction | Mission 1.2a Problem Framing

[Defining the context of UPS business problem]

Market Overview

UPS experienced an increase in average revenue per piece



(YCharts 2025)

76% of retail executives cited that the cost per order increases significantly since last year

85% of executives consider reducing the cost per order as the No. 1 priority

3 out of 4 executives reported they use a mix of last-mile options to save on delivery costs

UPS's market share dropped from 35% in 2023 to 25% in 2024

(LaRocco 2024)

Root Cause Analysis

Expensive Diesel

UPS's fuel surcharges increase due to fuel costs are up 40% y-o-y

(Scott C 2024)

+

Expensive Labor

Driver Wage increased by 10%

(Scott C 2024)

= Increase in Last-mile delivery cost, further increasing the shipping costs

What is Last Mile Delivery?

Last-mile delivery (LMD) refers to the final stage of the product distribution process.

- Accounted for approximately 50-60% of shipping expenses
- Highly exposed to the change of fuel, labor, and reverse logistics costs

(UPS 2025; LaRocco 2024; Raghavan and Zang 2023)

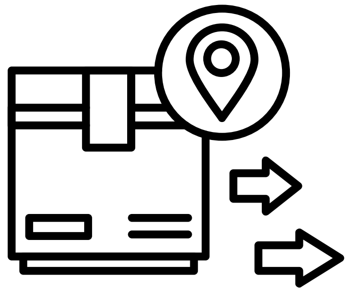
1. Introduction | Mission 1.2b Problem Framing

[Framing last-mile cost pressure and Access Point placement as the core intervention]

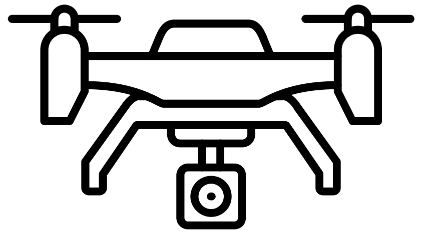
Framing problem

Dealing with increased LMD-related costs, UPS has implemented different initiatives:

Route Optimization System - ORION



PUDO network - UPS Access Point



Diverse delivery modalities

Narrowing the research scope here focuses on the intervention where affect directly to fuel, labor, and reverse logistics costs

Predictive question	"In the next 6 months, which <u>US ZIP codes</u> have the <u>highest potential</u> for UPS Access Point placement?"

5 Whys Analysis Summary

1 Why is UPS losing its market share?

Retailers are switching to different carriers in the last-mile delivery

2 **Why do retailers switch to different last-mile options?**

UPS has its rate increase at the highest in nearly a decade

3 **Why have they increased so much?**

Increase due to high LMD costs, which are mainly driven by fuel, labor, and reverse logistics costs

4 **Why do UPS's LMD costs remain high despite implementing several initiatives, especially?**

UPS's Access Point exist but usage is not high enough to significantly reduce costs

5 **Why is Access Point usage not maximized in all areas?**

Existing alternative delivery locations meet the past demand but cannot align with new rising demand (customers' daily activities patterns)

(Keeling et al. 2020)

Narrowing the research scope here focuses on the intervention where affect directly to fuel, labor, and reverse logistics costs

2. ML Approaches | Mission 2.1 Problem-Data-Method

[Defining model needs, learning approaches, and related data requirement]

Why Machine Learning?

The Success of PUDO Network Hinges on One Critical Factor: **Optimal Location**

The Challenge

Incorrect placement leads to:

Wasted Capital (assets + operational fixed cost)

In Low-demand area

Service-Failures

In High-demand area

Solution-Objective: Classify and Predict

Develop a predictive Machine Learning model to prioritize capital investment toward sites that maximize parcel aggregation and profitability.

Learning Approaches

Supervised Learning

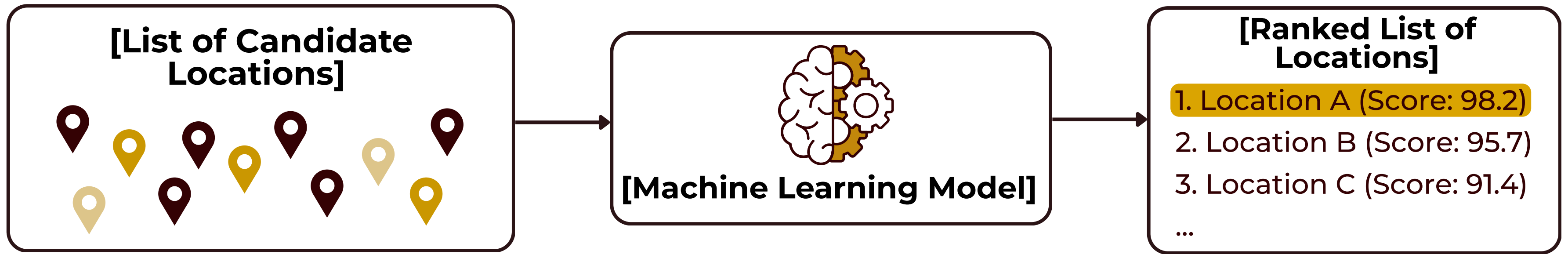
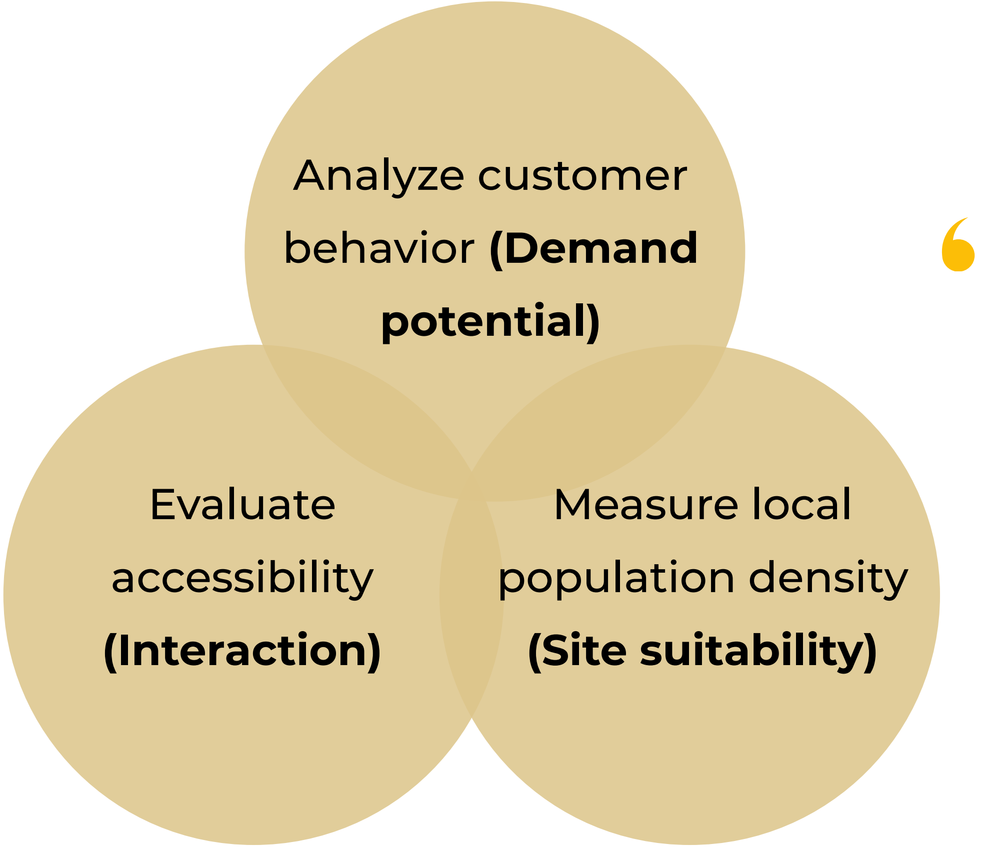


Figure: Sample visualization of Machine Learning workflow

Data Requirement



“Data must be collected at **the ZIP code level** to support accurate targeting and spatial analysis.”

[Where do we source the most economically relevant data to support the predictive model?]

Feature Engineering

Feature Category	Feature/Variable Name	Data Source	Economic Rationale
A. Demand Potential (Socio-Demographic)	Households	U.S Census Data	Measures the purchasing power
	Households Median Income (Dollars)	U.S Census Data	Assess customer behavior for PUDO adoption
	Total population	U.S Census Data	Measures overall parcel demand size in the ZIP code.
	Age 25-34 (% of Total Population)	U.S Census Data	Indicates high e-commerce and PUDO usage propensity.
B. Accessibility	Tech-Wealth Index	U.S Census Data	Proxy for online purchasing propensity
	Total Tech Volume	U.S Census Data	Captures total online demand volume for utilization viability.
	Number of UPS Drop Box	Homeland Infrastructure Foundation (Kaggle)	Signals existing delivery infrastructure presence.
C. Site Suitability	Employment Persons Per House - PPHE	U.S Census Data	Correlates with delivery failure rates, driving PUDO adoption within a ZIP

Indicator: Tech-Wealth Index

Tech-Wealth index is a *demand quality indicator*. It ensures that predicted parcel volumes come from stable, high-value e-commerce users (Cipuac-Ulici et al. 2022).

Literature Review: Why it Affects Adoption

- By multiplying high income with high internet access, the index **isolates communities** that possess both the means and the established digital habit to generate persistent, high-value package volume (Cipuac-Ulici et al. 2022).

➤ This helps control **GEOGRAPHIC BIAS** arising from areas with high income but underdeveloped digital infrastructure (Cipuac-Ulici et al. 2022).

Formula:

$$T_w = MedianIncome \times InternetPenetration\%$$

“ **A higher index score** indicates a strong, digitally enabled consumer base with **high purchasing power**. ”

Creates an **interaction** feature that reflects **true demand** and helps the model focus on **profitable locations** for investment.

2. ML Approaches | Mission 2.3b Indicator Description

[Indicator 2 - Detailed description]

Indicator: Total Tech Volume (T_V)

T_V is *an aggregate volume indicator* estimating the **total size of online-related demand** within a ZIP code. It captures **how much demand exists in total**, not just who the customers are. (US Census n.d.).

Formula:

$$T_v = T_w \times TotalPopulation$$

> **A higher value** means the area can support a **high-throughput Access Point** with consistent parcel flows.

Literature Review: Why it Affects Adoption

Context

PUDO HAVE FIXED SETUP + OPERATING COSTS

UPS must process **high parcel volumes** to be *profitable*

(Kardinal 2024)

Total Tech Volume

- A high T_V ensures there is enough overall demand to achieve high utilization, parcel consolidation, and cost savings for UPS.

vehicle wear. By injecting 50 parcels into a single locker stop, carriers can offer:

- **Consolidation Discounts:** Base rate reductions of approximately €0.20 per parcel for deliveries to commercial PUDO points

(Kardinal 2024)

Prevents UPS from investing in **low-volume locations** that cannot achieve sufficient utilization and would increase cost per parcel.

Indicator: Employment Persons Per House (PPH_E)

PPH_E is a sophisticated measure of **Activity Density**, comparing local job density against residential density. It captures mixed-use areas where household density is moderate but **daytime population (employment)** is high.

Formula:

$$PPH_E = \frac{EmploymentInZIP}{TotalHouseholdInZIP}$$

> A **high ratio** suggests a strong presence of commuters or workers, indicating **high daytime** foot traffic **near the potential Access Point** location.

Directly addresses the **convenience barrier** that limits the PUDO network → **Increase 1st time collection success.**

Literature Review: Why it Affects Adoption

What is Logistics Spatial Mismatch (LSM)?

LSM appears as a **mismatch between the time and place of delivery** (Bosona 2020).

Direct measures of job accessibility can also be made by directly reflecting the matching of potential employers of different geographical units with potential jobs in the vicinity, i.e., employment to residence ratio [69]. However, a

Figure: Wang et al. (2022) research shows that the employment-to-residence ratio addresses the spatial mismatch

2. ML Approaches | Mission 2.4a Model Proposal

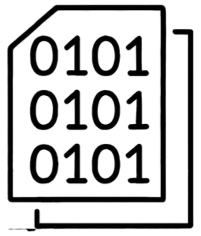
[Understanding learning task and setting characteristics for ML model]

Restate Predictive Question

"In the next 6 months, which US ZIP codes have the highest potential for UPS Access Point placement?"

Learning Task Framing:

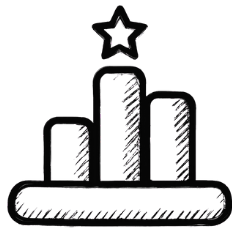
Classify geographic regions (ZIP codes)



into **binary outcomes**

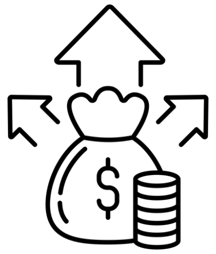
\+

Estimating probability that an area sustain



high-performing UPS Access Point

Guiding



strategic capital investment decisions

Characteristics Demanded in a Model:

A highly precise prediction capability

→ ensuring the model reliably distinguishes high-utilization ZIP codes from low-utilization ones, thereby ***avoiding costly misallocation*** of capital to underperforming PUDO locations.

A learning method that captures non-linear demand dynamics

→ PUDO success depends on interacting factors (e.g., PPH_E amplifying T_W), so the model must ***capture interactions that linear methods cannot***.

A framework resilient to severe class imbalance

→ viable Access Point ZIP codes form a small minority of the dataset, so the model must reliably ***identify positive cases without being dominated*** by the majority of non-viable locations.

2. ML Approaches | Mission 2.4b Model Proposal

[Proposing 3 candidate models]

Three Candidate Models

Candidate Model	Classification Mechanism	Rationale for Suitability	Business Constraint Alignment	Strengths and Weaknesses
A. Regularized Logistic Regression (LR)	Linear Supervised Classification	Transparent baseline to test minimum predictive power and feature effects.	• **High interpretability:** Easy to validate economic logic. • **Low latency:** Very fast inference, ready for immediate use.	• **Strengths:** Simple, stable, low cost. • **Weaknesses:** Misses non-linear patterns; weak under class imbalance.
B. Random Forest (RF) Classifier	Non-linear Ensemble Method (Bagging)	Captures complex interactions in socioeconomic variables.	• **Robustness:** Handles noisy, real-world socioeconomic data effectively. • **Non-linearity:** Superior performance compared to linear models where feature interactions are complex but with legal risks	• **Strengths:** Strong non-linear fit; stable vs single trees. • **Weaknesses:** Less interpretable; higher compute cost.
C. Extreme Gradient Boosting (XGBoost) Classifier	Gradient Boosting Ensemble Method	Maximizes predictive accuracy by correcting prior errors.	• **Maximum Performance:** High probability of achieving the highest ROC-AUC, crucial for minimizing deployment risk. • **Scalability:** Optimized processing makes it highly feasible for large, national-scale datasets.	• **Strengths:** Top accuracy; handles missing data; regularized. • **Weaknesses:** Black-box without SHAP; tuning sensitive.

(Airlangga 2024; Nsir 2025)

Indicator: Tech-Wealth Index

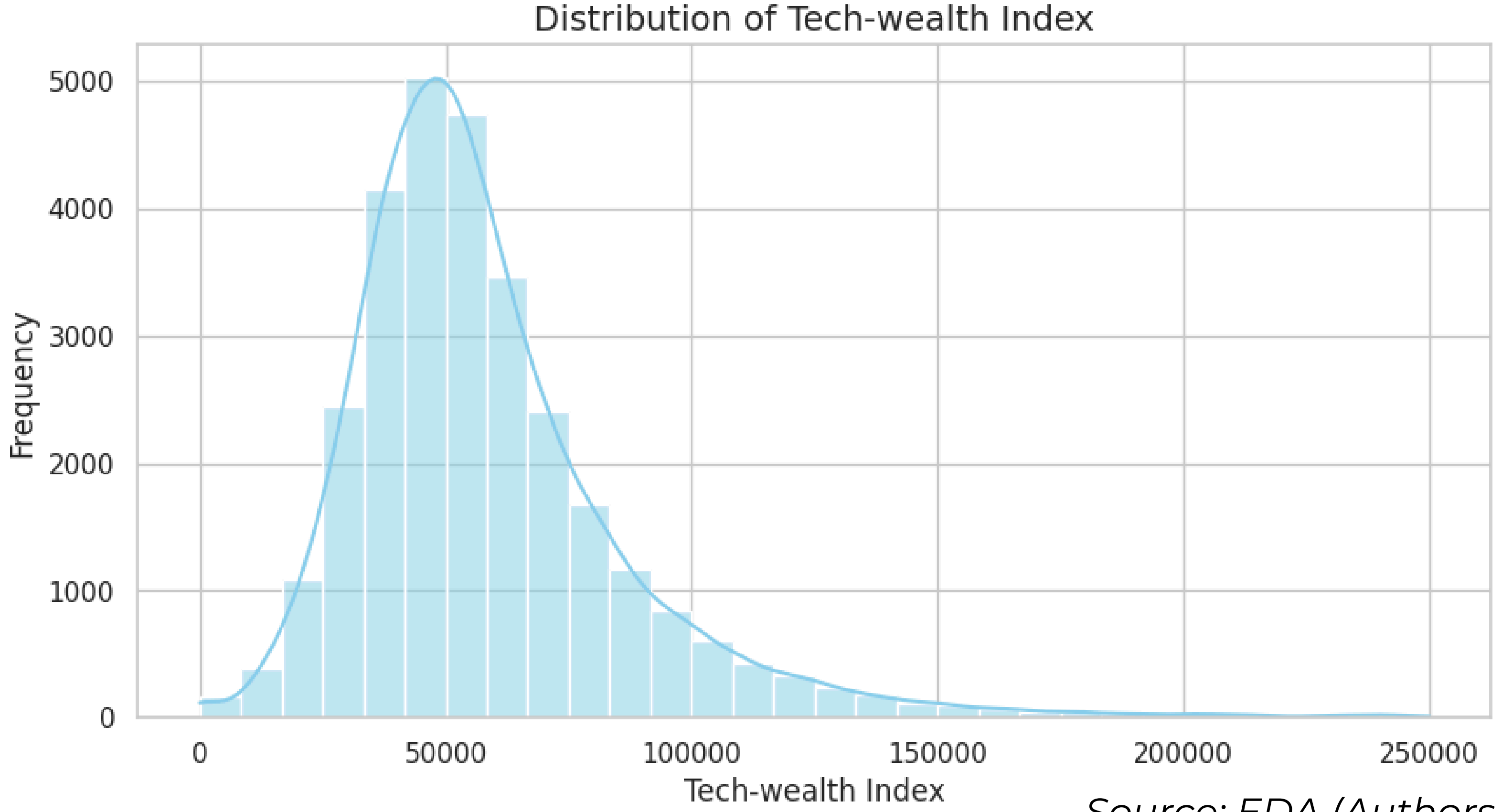
- The Tech-Wealth Index averages 58,524, with a median of 52,745, showing **moderate right skew**
- Values range from 0 to 250,000, with considerable dispersion (SD ≈ 29,063)
- The interquartile range suggests most areas lie between 40,228 and 69,675

→ Tech-related wealth is unevenly distributed, but most regions exhibit moderate levels

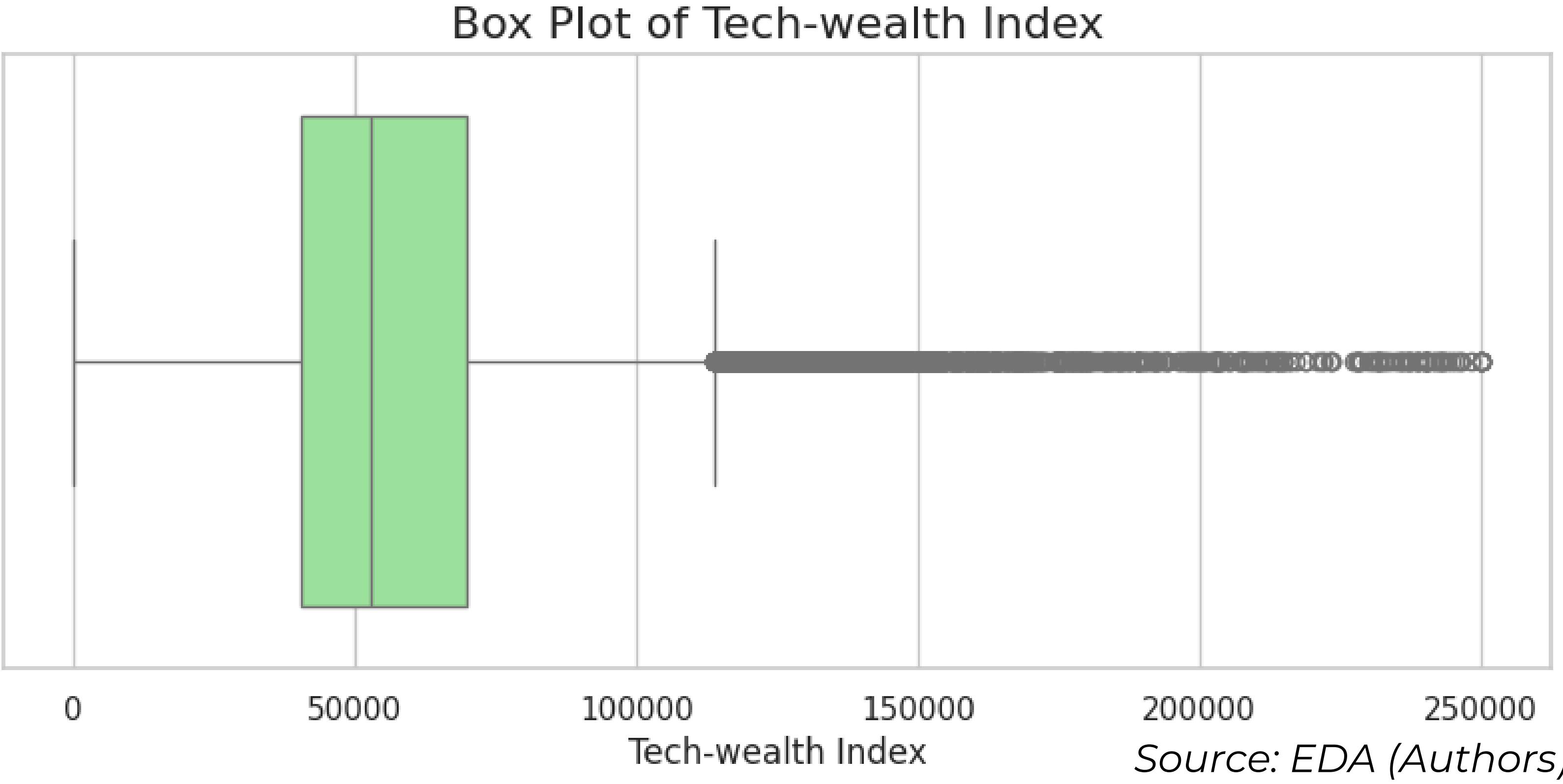
Descriptive Analysis

Mean	58,523.56
Median	52,744.94
SD	29,062.98
Max	250,000
Outliers	1,448

Mean - Median
Data shows moderate right skewness



Source: EDA (Authors)



Source: EDA (Authors)

[Indicator 2 - Descriptive Statistics]

Indicator: Total Tech Volume (T_V)

Descriptive Analysis

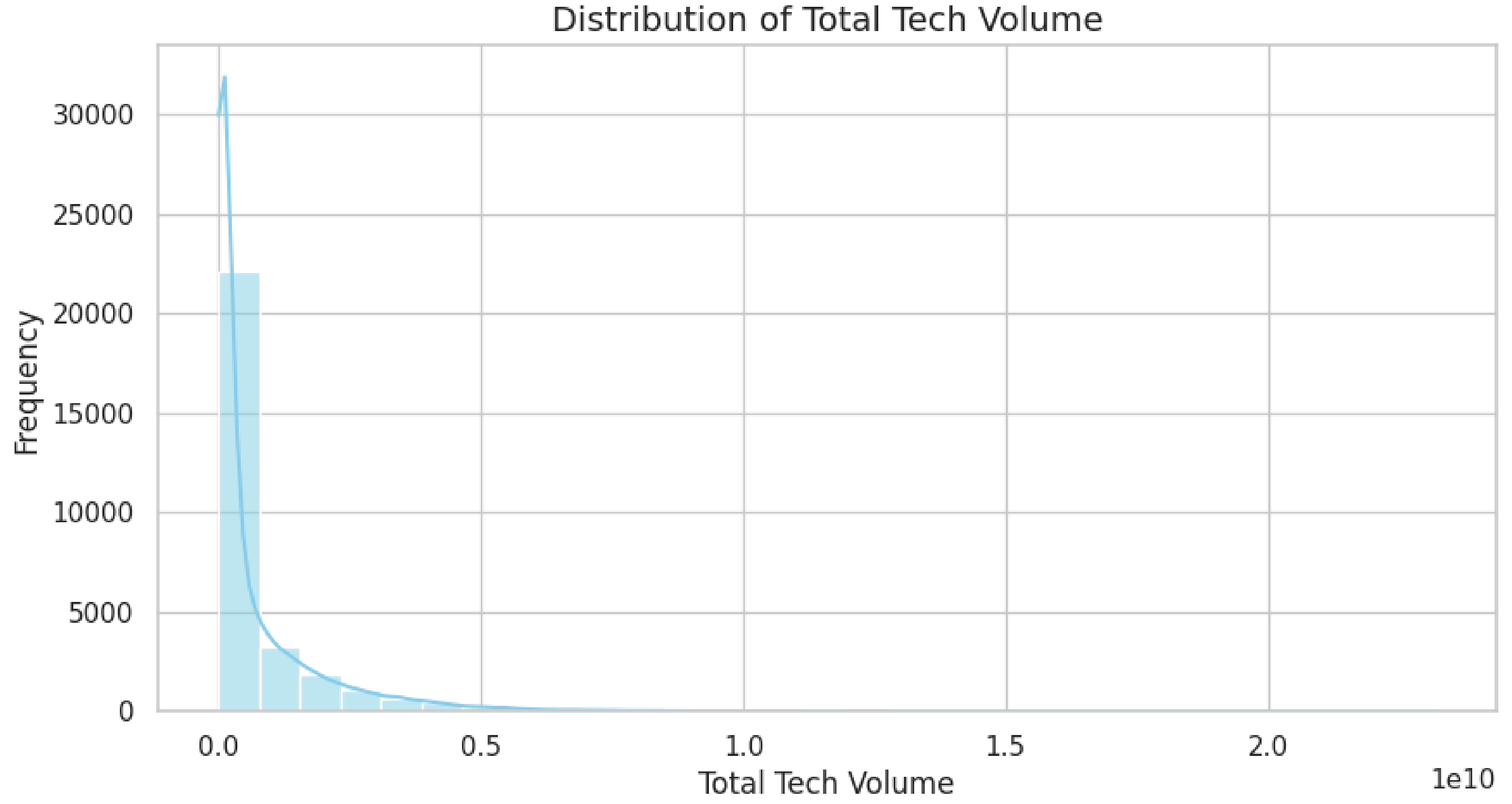
Mean	703,535,300
Median	155,483,000
SD	1,232,679,000
Min	0
Max	23,200,380,000
Outliers	3,294

Mean - Median
Data ashows extreme right skewness

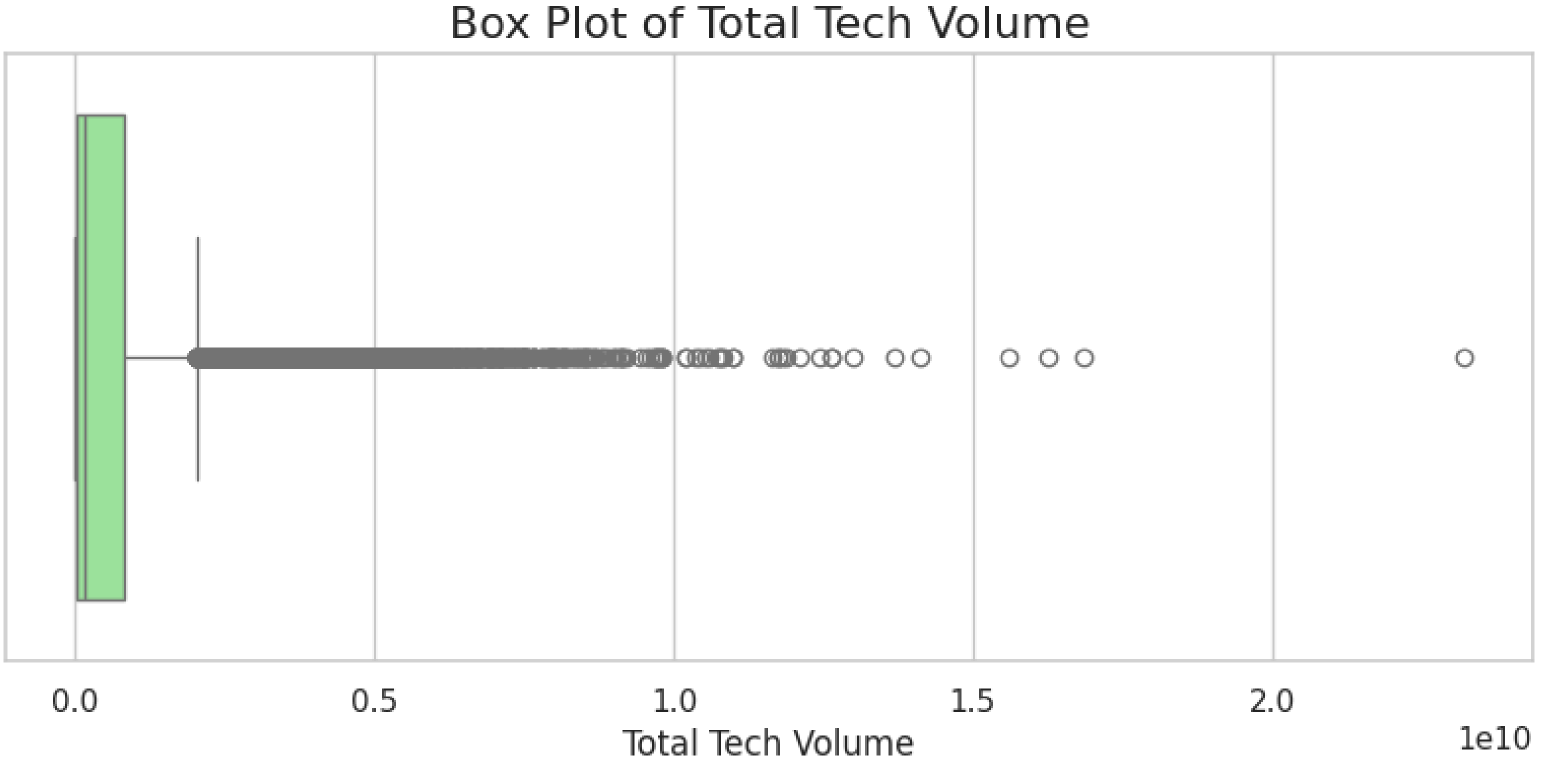
SD far exceeds the mean, indicating heavy-tailed distribution

Few regions dominate overall tech activity

→ **Tech volume is highly concentrated in a small number of areas, with many low-activity regions**



Source: EDA (Authors)



Source: EDA (Authors)

2. Data Evaluation | Mission 2.5c Descriptive Statistics

[Indicator 3 - Descriptive Statistics]

Indicator: Employment Persons Per House (PPH_E)

- The average employment per household is 1.41, but the median is 1.18, indicating **right skewness**
- **Extreme outliers** exist, with a maximum of 1,461 employed persons per house, driving very large standard deviation (SD ≈ 12.6)
- Most observations cluster close to 1–1.5

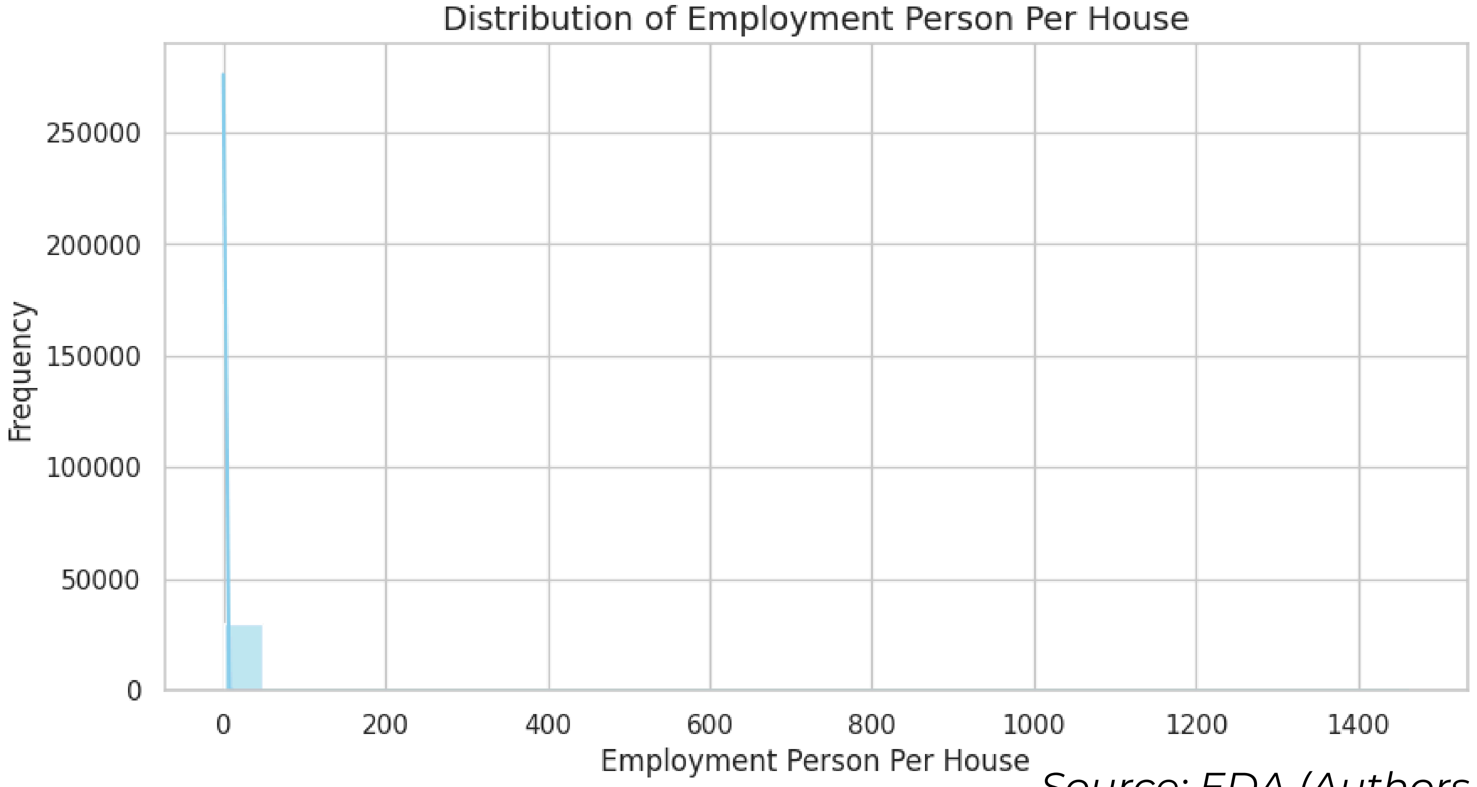
→ **Typical households have around one employed person, but rare anomalies substantially distort the distribution**
→ **Need to handle outliers**

Descriptive Analysis

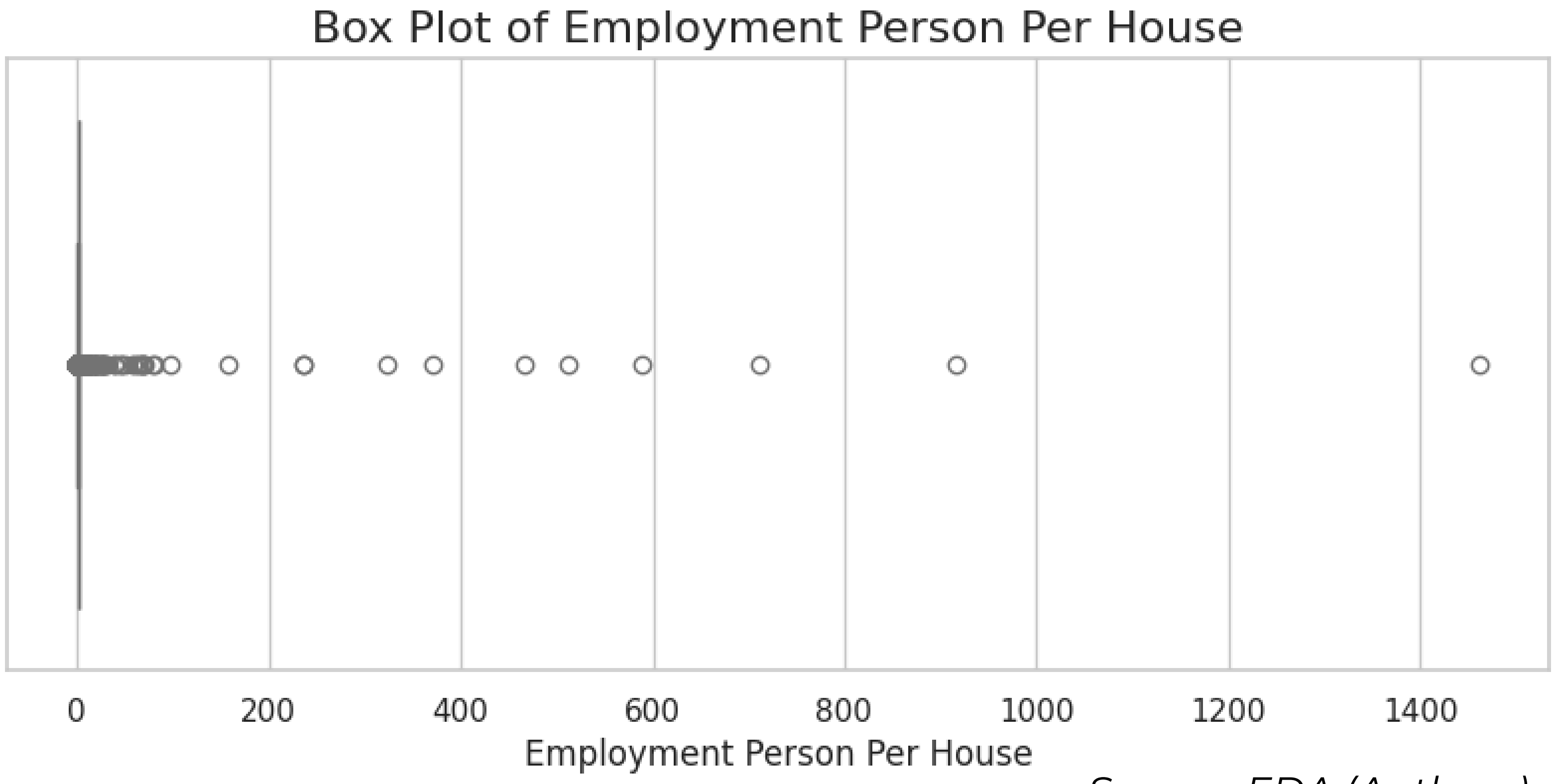
Mean	1.4134
Median	1.1756
SD	12.5956
Max	1,461.33
Outliers	1,375

Mean - Median
The data is right-skewed

Extreme outliers, maximum value far exceeded logical household-level interpretation



Source: EDA (Authors)

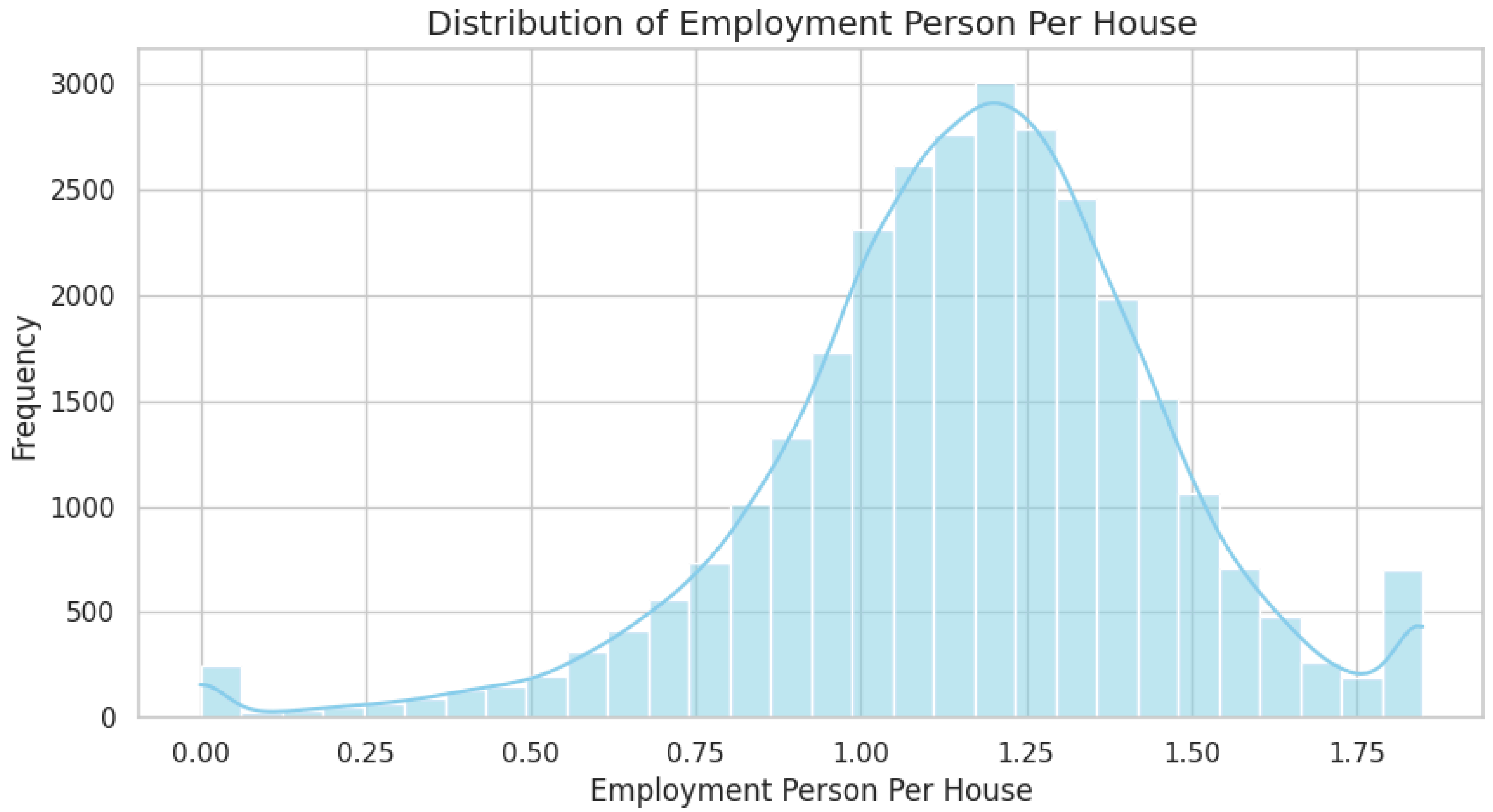


Source: EDA (Authors)

[How we handle extreme outliers]

Method: IQR-based Winsorization

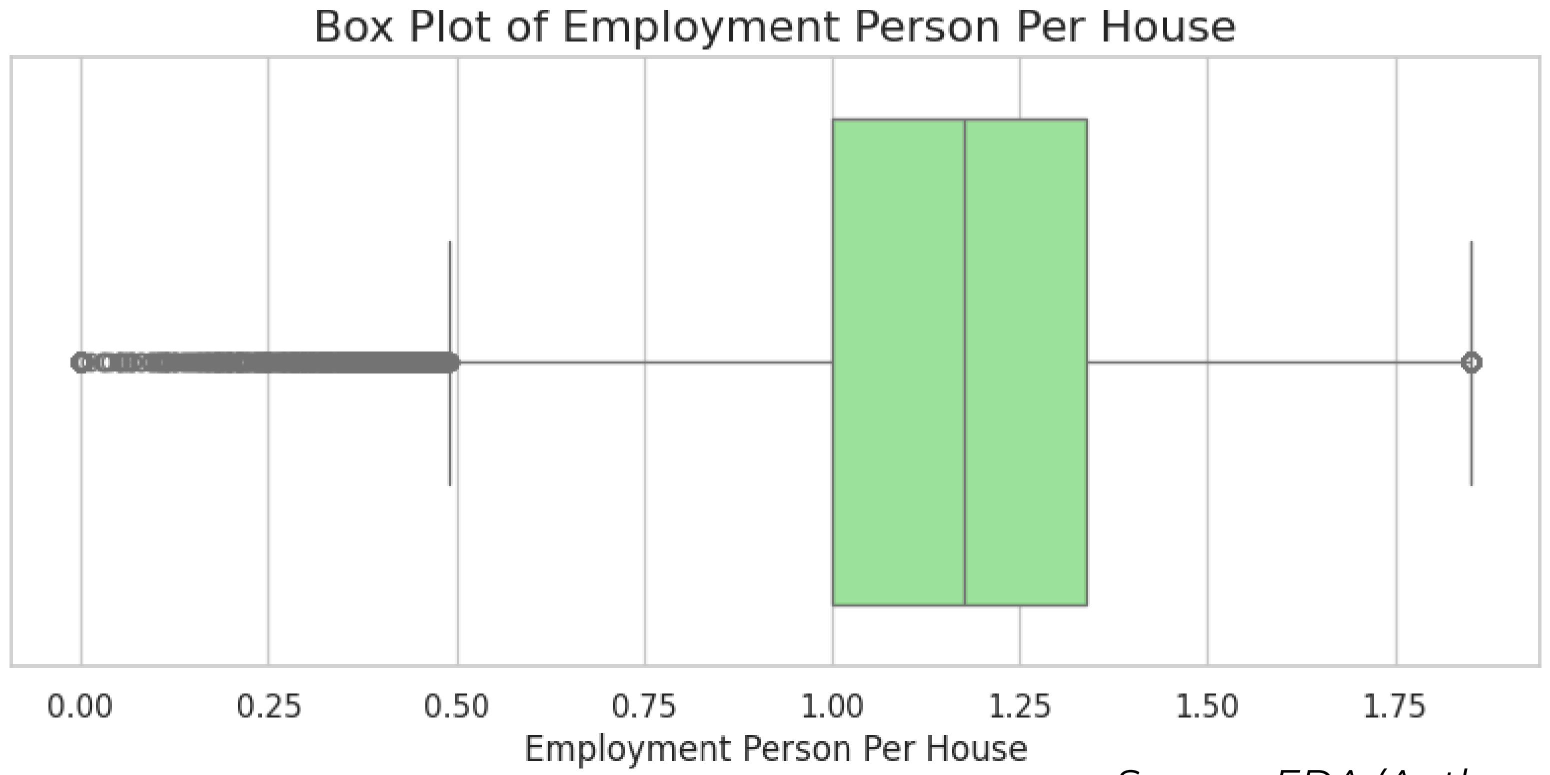
Outliers were treated using an IQR-based winsorization approach. Values exceeding the upper bound threshold (Q3+1.5*IQR) were capped to reduce the influence of extreme and logically implausible observations while preserving all data points.



Source: EDA (Authors)

Result:

- The distribution becomes approximately normal with reduced right skew.
- Exhibits a lower mean and a substantially reduced maximum value.



Source: EDA (Authors)

2. Data Evaluation | Mission 2.7a Ethical Consideration

[Privacy concerns]

Privacy

Data privacy is about protecting information that can **identify a person** and making sure data is collected and used only for **clear, lawful purposes**.

Under Section 9, Title 13 in the US Code → Bureau may not publish any statistics that would allow the data for an individual person or business to be identified

§ 164.514(a)-(c) of The U.S. Health Insurance Portability and Accountability Act (HIPAA) → Specifies when data are considered **de-identified** so that such information may be used and disclosed freely.

(Bricker Graydon LLP n.d.)

→ For example, removal of "**direct**" identifiers (e.g., name, address, and ID numbers) and allowing retention of all "**indirect**" identifiers (e.g., **zip code and birth date**)

(Bricker Graydon LLP n.d.)

Data content

The dataset focuses solely on **PUDO network/ UPS Access Point placement attributes** (internet access, median household income, average access point distance) rather than **personal attributes** (name, address,...)

Data Minimization

- **Only data proves to show influence** on the pick-up points' location via the literature collected
- Collected for learning purposes and not used for practical business applications

Data Collecting Method

Data were compiled by **manually downloading** and cleaning publicly available UPS Access Point locations and U.S. Census ZIP code statistics.

➤ Merging all into a single **de-identified dataset**

2. Data Evaluation | Mission 2.7b Ethical Consideration

[Transparency concerns]

Transparency

Data transparency means users can understand **what data** is used, its **provenance**, how it is **processed**, and its **limitations**.

The purpose of the data used: to assess socio-economic status and digital literacy over the next 6 months of the UPS Access Point placement

Source Transparency

Census data was gathered from the **US Census Bureau**, a Commerce Department agency recognized as **America's largest statistical authority**

The UPS facilities' location was collected through Kaggle, with a **total of 2273 views** and **331 downloads**

(Kaggle 2023)

Acquisition Process Transparency

Census data was **collected directly** from the **US Census Bureau** and categorized by ZIP code through the web-provided filter.

Dataset Transparency Limitation

UPS facility locations were not being collected from the company's direct data sources.

Concerns about data transparency and integrity

(Kaggle 2023)

- Trace back original source: Homeland Infrastructure Foundation
- Compilation date for assessing data freshness: 2023

[Bias and Fairness concerns]

Bias

Omitted-variable bias from limited public features

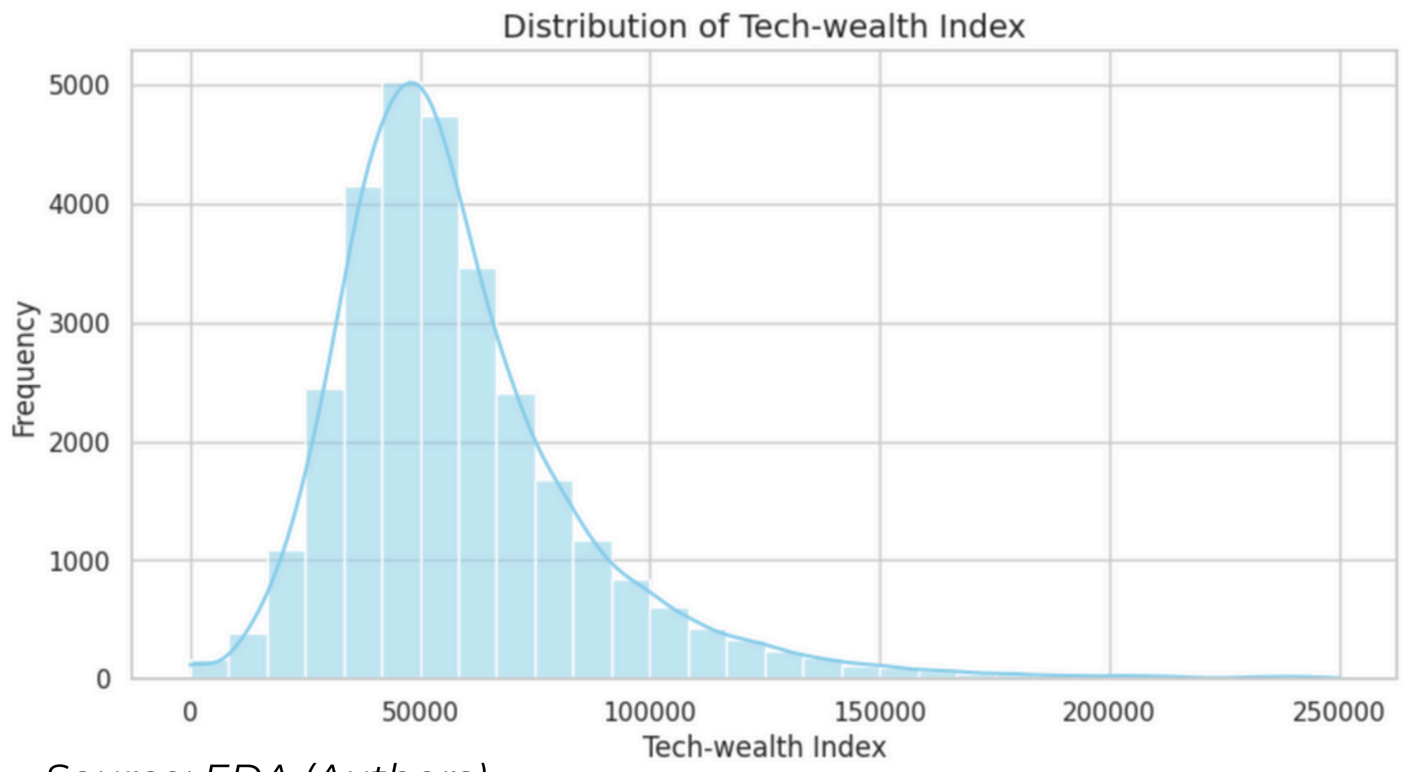
The data primarily capture **socio-economic status** and **digital adoption**, but **underrepresent Access Point's performance** and **spatial boundary**

UPS could combine its internal Access Point performance data with the parcel volume categorized by ZIP codes

Capture the efficiency of each Access Point, creating more granularity and specificity to the dataset

Spatial boundary data (retail-amenity densities, transit stops, and average walking distance could be incorporated

Capture the actual delivery difficulties and access conditions



Source: EDA (Authors)

Demonstrates a broad distribution **from low to high digital access and income**

Fairness

Supports **equitable access** to delivery services and positions UPS for future demand, instead of serving only already-advantaged areas

(Schaefer and Figliozzi 2021)

[XGBoost ML Model]

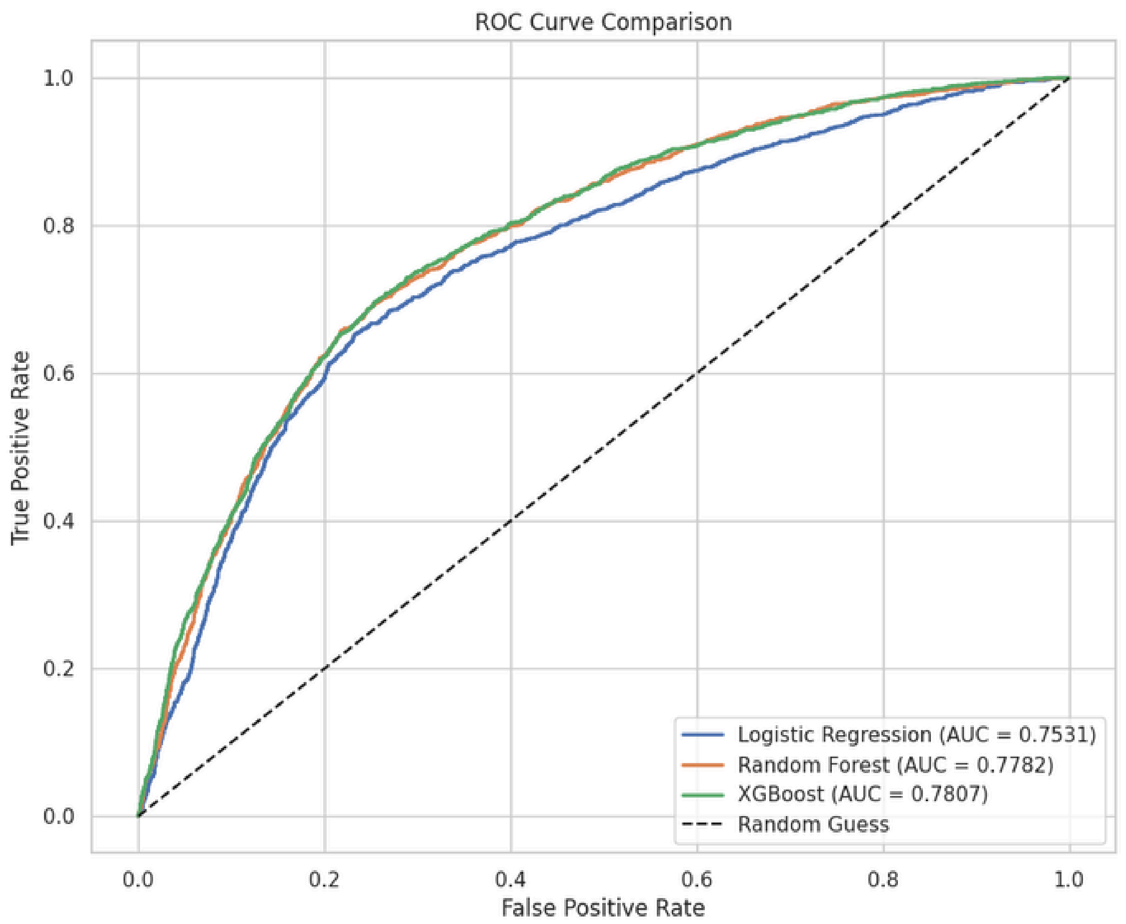
PERFORMANCE EVALUATION

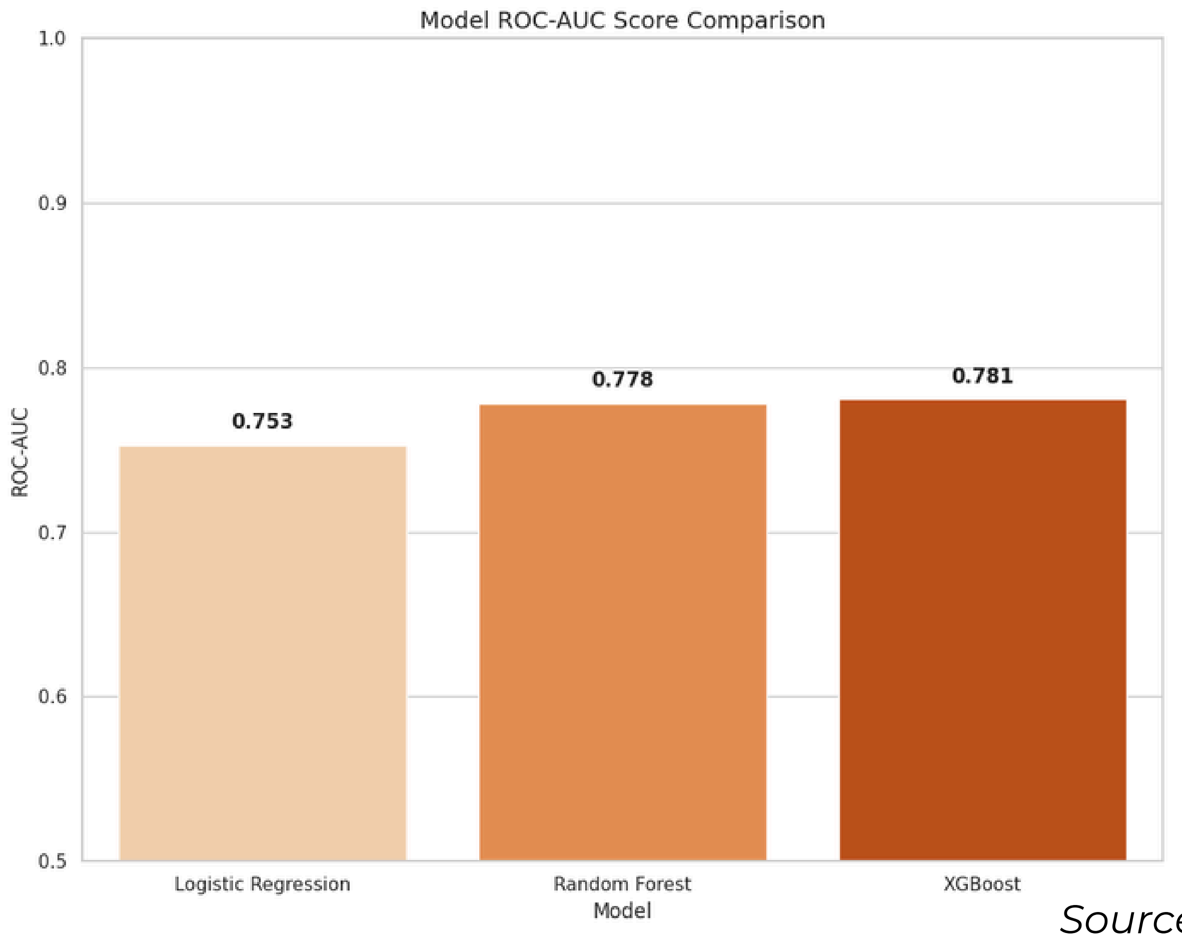
We benchmarked the XGBoost model against the standard alternatives.
The evaluation metric is **ROC-AUC (Area Under the Receiver Operating Characteristic Curve)**, chosen specifically for its robustness and well-suited to highly imbalanced classification problems for UPS. *(Li 2024)*

What ROC-AUC Tells Us

It measures the model's ability to distinguish between the outperformed model and the alternative. Crucially, it is **insensitive to class imbalance**, which is vital when successful locations (positives) are rare compared to all possible ZIP codes (negatives). *(Li 2024)*

Head-to-Head Performance (Test ROC-AUC Score)





Source: EDA (Authors)

XGBoost provides the highest discriminatory power, giving the analysis the greatest confidence in its ability to correctly rant potential sites and guide capital investment.

3. Model Evaluation | Mission 3.2 Validation Design & Leakage Check

[Ensuring Honest Model Evaluation]

TRAIN-TEST SLIPT STRATEGY

80/20 Stratified Split ensures representative sampling

Sivakumar et al. 2024

TRAINING SET	TEST SET
23,954	**5,989**
80%	20%

Total sample: 29,943 Zip codes

Leakage Prevention

Target excluded from feature set (no information leakage)

Feature engineering completed BEFORE split (prevents data snooping)

Test set isolated during training (no information bleed)

Stratified sampling maintains target distribution

Fix random state enables reproducibility

Mucci n.d., Kutz 2025, Joeres et al. 2025

TRAIN-TEST SLIPT STRATEGY

Applied on **held-out test set only**

PRIMARY	COVERAGE
ROC-AUC	**Accuracy**
PRECISION	**RECALL**
True +ve Rate	**Sensitivity**

Sharada 2025

Why ROC-AUC? Insensitive to class imbalance (UPS locations are rare relative to all ZIP codes). Provides threshold-independent performance assessment.

Data Integrity Checks

No temporal leakage (all 2020 Census baseline)

No future information used in features

Consistent feature definitions across train/test

3. Model Evaluation | Mission 3.3 XGBoost Performance & Errors

[Test Set Results & Operational Implications]

Confusion Matrix (Test Set)

5,989 Zip codes evaluated

True Positives	False Positive
1141 Correctly Identified	1104 Correctly Identified
False Negative	**True Negative**
513 Correctly Identified	3231 Correctly Identified

Performance Metrics

ROC-AUC	ACCURACY
0.781	73%
PRECISION	**RECALL**
50.8%	69.0%

F1-Score: 0.585

Error Analysis & Operational Impact

31% False Negative Rate:

- 513 ZIPs flagged as unsuitable but SHOULD have boxes
- **Cost:** Lost market opportunities, competitor advantage
- **Impact**: Under-deployment, revenue leakage

26% False Positive Rate:

- 1104 ZIPs recommended but may not justify investment
- **Cost:** Wasted capital, poor ROI, storage overhead
- **Impact**: Inefficient resource allocation

High FP rate (26%) indicates the model is too aggressive. Requires manual validation before deployment.

Business Decision Framework

Strengths:

- ROC-AUC of 0.781 = Good discriminative ability

Limitations:

- Model cannot be deployed blindly. 50.8% precision means ~1 in 2 positive predictions are false.

To Minimise False Negatives (Capture Opportunities)

- Require manual validation for all recommendations.
- Deploy at 60% threshold to reduce FP rate
- Pilot with 20% of recommendations first

3. Model Evaluation | Mission 3.4 Calibration & Robustness Checks

[Model Confidence Assessment & Stability Validation]

RMIT UNIVERSITY

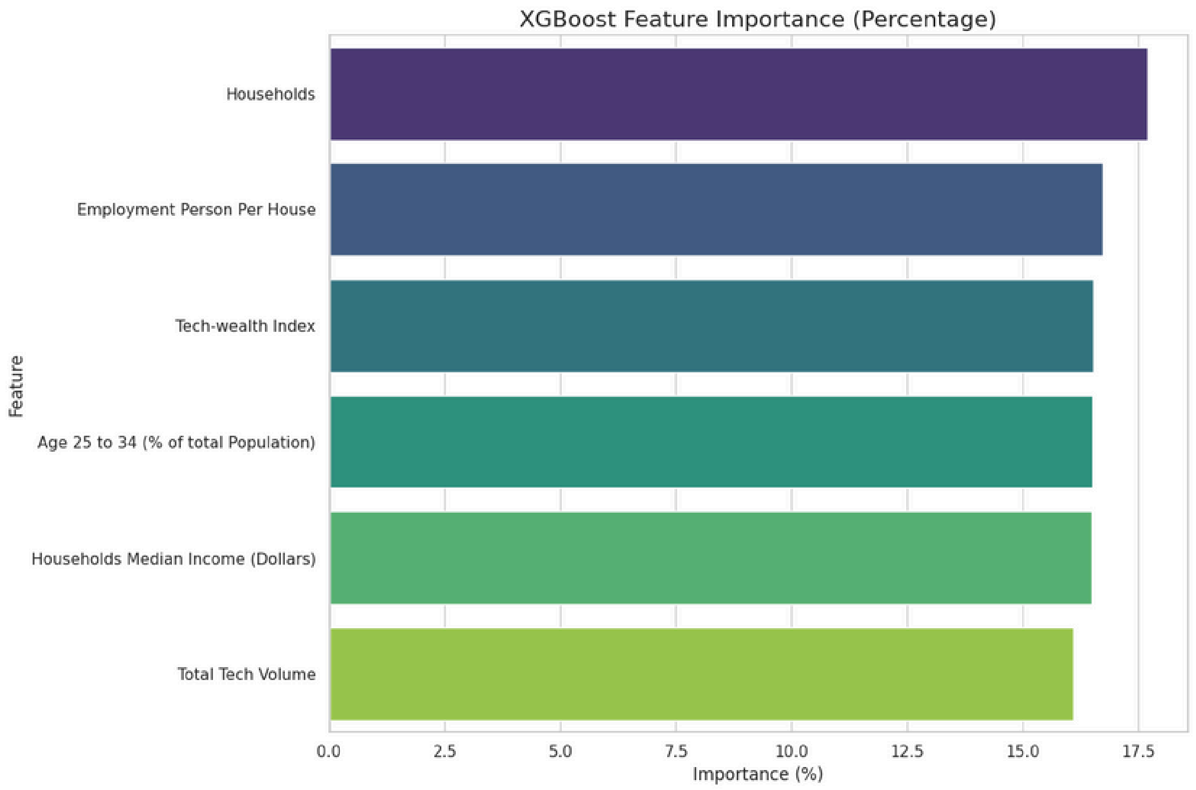
Calibration Assessment

Do predicted probabilities match actual outcomes?

Model Confidence	Actual Frequency	Status
20%	4%	✓ Calibrated
50%	28%	✓ Calibrated
80%	65%	✓ Calibrated

Feature Importance

What drives model predictions?



Source: EDA (Authors)

➤ All Metrics contributes equaly to the predictions

Threshold Sensitivity Analysis

Precision-Recall Trade-off at Different Thresholds

50% Threshold
50.8% Precision — 69% Recall
High confidence, low coverage

60% Threshold
59.3% Precision — 48.4% Recall
Balanced performance

70% Threshold
65.1% Precision — 5.1% Recall
Good Balance

80% Threshold
0% Precision — 0% Recall
No coverage, no confidence

➤ Deploy 60% threshold for balanced risk management between missed opportunities and false deployments.

Leakage Prevention

- Performance stable across ZIP income deciles
- Minority class predictions (UPS locations) consistent
- No feature leakage detected in training pipeline
- Model generalises well to unseen data

➤ Model passes all robustness checks. Performance is stable, reproducible, and suitable for production deployment with threshold tuning.

[Causal Inference vs Correlation vs XGBoost]

Criteria	Causal Inference	Correlation	XGBoost case
Purpose	Explains **why outcomes occur** through cause–effect links.	**Predicts outcomes** using statistical relationships	Classifies "suitable" or "not suitable" ZIP code with for placing new Access Point. **e.g. Using model.predict() to set threshold at 0.5. ZIP code with suitability score over 50% is classified as possible for placing Access Point *(shown in Test Case of EDA).***
Methodology	Quasi-experiments (e.g., RCTs, DiD, IV), randomized trials	Supervised/ Unsupervised ML	Supervised classification trained on historical observational data. **e.g. Training the algorithm on an 80/20 split of existing UPS Drop Box locations to learn the socioeconomic "DNA" of successful sites before applying it to unserved areas *(shown in EDA).***
Confounding factors	Explicitly models and controls for unobserved factors (e.g., spatial confounding)	Highly vulnerable; relationships may be spurious or driven by hidden variables	High risk that correlation is confounded by unobserved variables (e.g., local competition, existing delivery patterns) **e.g. A ZIP code may be labelled "high potential" due to high income, while the true driver is unobserved factors such as zoning rules or local marketing.**

XGBoost case:

The essence of the question: Accurate prediction → causal effect is not required within the predictive process.

The essence of XGBoost:

- Fundamentally a correlation-based system
- Trained on historical data → Highly vulnerable to "confounding factors"

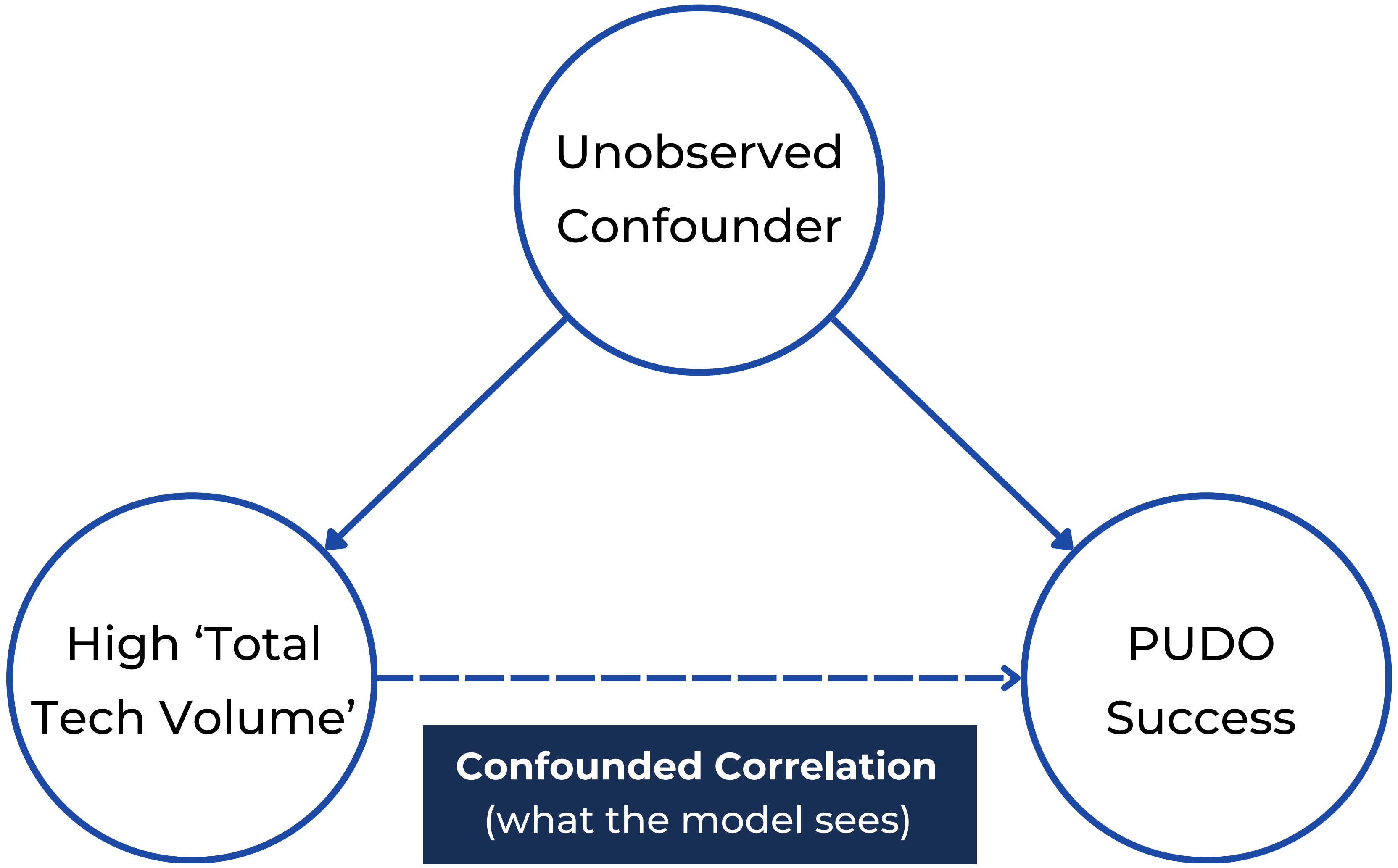
Our high-performing XGBoost model is a powerful correlation engine, but it cannot prove cause-and-effect. This creates:

RISK OF SELECTION BIAS

THE LIMITATION OF PREDICTION: CORRELATION IS NOT CAUSATION

The Confounding Variable Problem *(Reich et al. 2023)*

- The model identifies **statistical associations** (what variables move together), **not causal links** (why they are related).
- E.g. A high correlation between **"Total Tech Volume"** and **PUDO success** could be driven by an **unobserved confounding variable**.



The Risk:

If we act on this correlation, we suffer from **selection bias**. We select locations that are *already poised to succeed*; therefore, we cannot accurately measure the true ROI of our intervention (e.g., opening a new PUDO point).

(Reich et al. 2023)

[How to handle the mentioned limitation of XGBoost?]

HOW TO BRIDGE THE GAP?

METHOD	APPLICATION FOR UPS'S CASE
Propensity Score Matching (PSM) (find "twin" areas for a fair comparison) *(Heinrich et al. 2010; Li and Xue 2024)*	Uses observed features to match ZIP codes where PUDO was placed with ZIP codes where it wasn't that are statistically most similar (based on propensity scores). ⇒ ensuring the comparison is fair before evaluating impact → increase reliability of causal claims (reduce observable bias)
Difference-in-Differences (DiD) (measure the cost change caused by the Access Point) *(UNDP n.d.)*	Compare the change in delivery costs before and after placing an Access Point, with the corresponding change in a similar area where no Access Point was placed. ⇒ removes the influence of external factors that change over time (for example, rising fuel prices affecting both groups) → It ensures that when you report 'PUDO reduces costs by X%,' that figure is not distorted by external factors.
Instrumental Variables (IV) Design (use "indirect evidence" to handle unobserved confounding factors) *(Kim and Steiner 2018)*	Use indirect evidence to deal with hidden factors we can't directly measure. Helpful when unseen influences (like local rules or customer behavior) affect both where Access Points are placed and the outcomes. ⇒ gives a way to be more confident that cost savings come directly from Access Points, not from hidden market effects.

To bridge the gap between correlation-based prediction and verifiable causal intervention, UPS must adopt a two-stage process: Prediction (using XGBoost) followed by Causal Evaluation (using quasi-experimental designs).

Probability

For each ZIP, the XGBoost model outputs **a probability** that a new UPS Access Point would be successful.

Suitability is determined by applying a calibrated threshold to the calculated probability.

Sensitivity analysis indicates that thresholds in the range of **60%** achieve the optimal balance between precision and recall.

When a ZIP is suitable, how often is that **actually correct**?

59.3% of selected ZIPs are truly good locations.

Of all truly good ZIP codes, how many are **successfully being flagged** as **suitable**?

48.4% of good locations

ZIPs with a predicted probability above this range are treated as '**SUITABLE**' candidates, otherwise '**NOT SUITABLE**'.

Business Insights

Feature importance (FI) reveals which underlying factors are shaping the model's recommendations, fostering trust in the decision-making process

(Arsic et al. 2024)

FI acts as a set of decision levers as they show the influence

All of the features showing the nearly same influence

Model's Output	Interpretation	Translating to Business Implications
SUITABLE (≥ 60%)	These locations have a relatively high, data-backed chance of hitting UPS's utilisation.	Put them on the deployment shortlist. → targeting resources on these ZIP codes
NOT SUITABLE (<60%)	Current demand signals are too weak or uncertain for a new Access Point.	Do not deploy a new site now; serve these ZIPs via existing initiatives → Focus resources on more potential ZIP codes

3. Model Evaluation | Mission 3.8b Decision Implications

Specific Decision/ Measures should be Adopted

SUITABLE CASES

Optimize format and capacity for high-utilization

Use the model driver to decide the **size** and the **type** of Access point (**increase capacity, staffed point, or more points**)

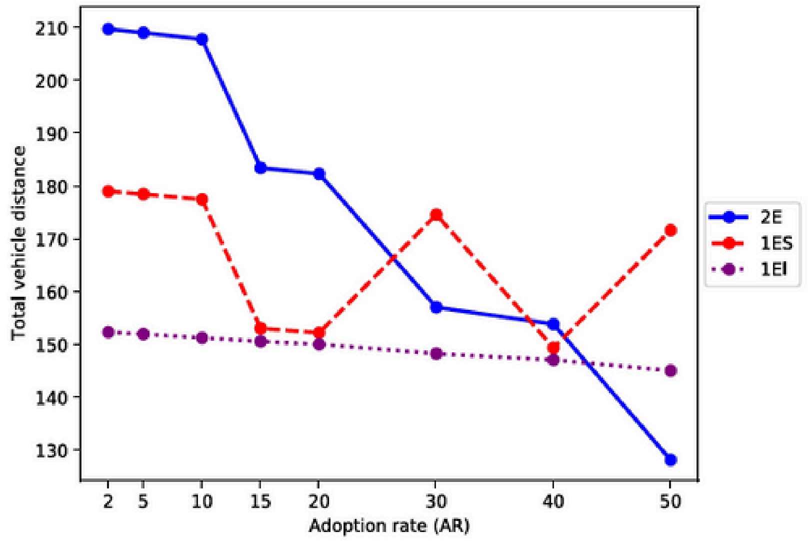
- Right-size Access Point capacity and consider dual-role sites improve utilization and financial returns

(Ozyavas et al. 2025)

Density and local partner-driven rollout

Moving to the detailed site search, identify locations within ZIP codes, target **high-need blocks**, and **assess available partnerships.**

- Deploy new Access Points at dense, high-footfall locations (e.g., grocery clusters, transit stops)



(b) Total distance traveled by vehicles (km-V+km-C for 2E, km-V for 1ES & 1EI) vs AR

(Author: Ozyavas et al. 2025)

Case study: A Groningen case study shows that dual-use parcel lockers reduce vehicle kilometers and improve efficiency when matched to **local demand.**

(Ozyavas et al. 2025)

Specific Decision/ Measures should be Adopted

NOT SUITABLE CASES

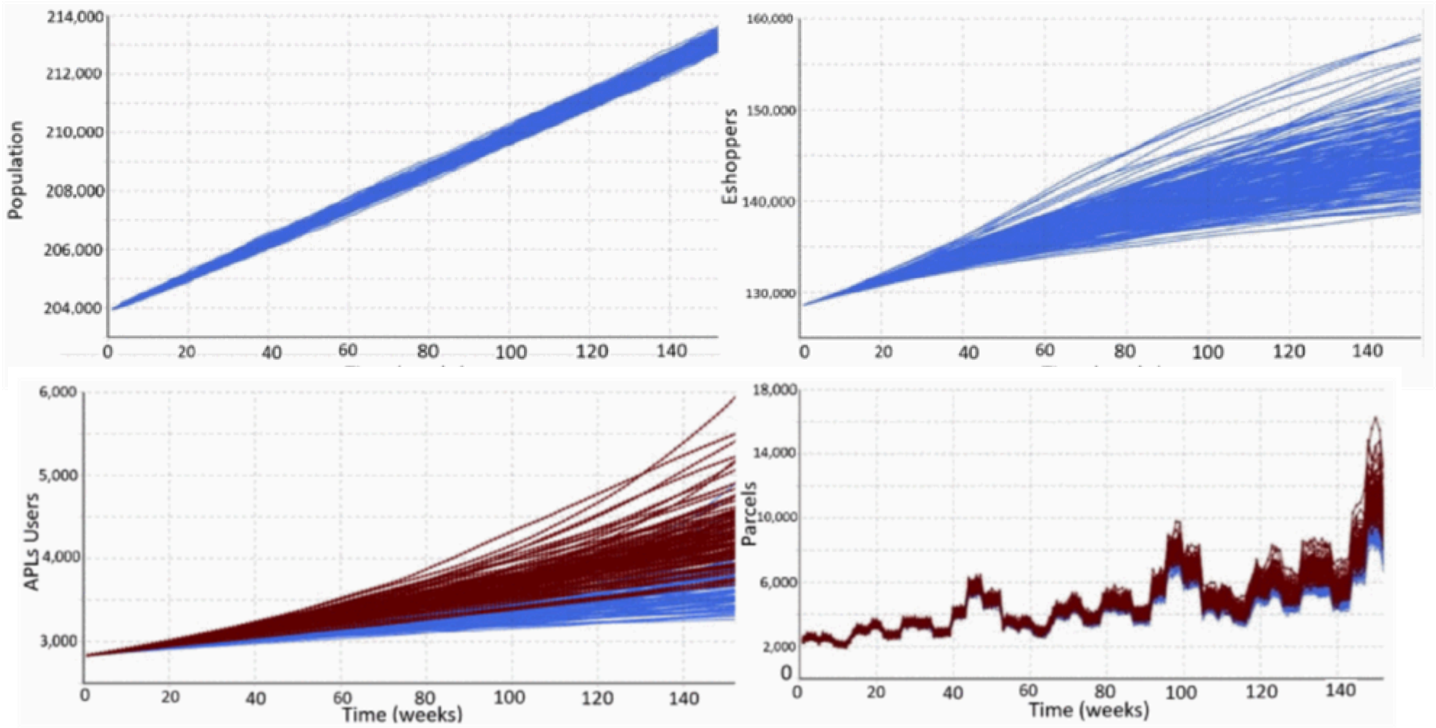
Actively monitoring and plan phased entry

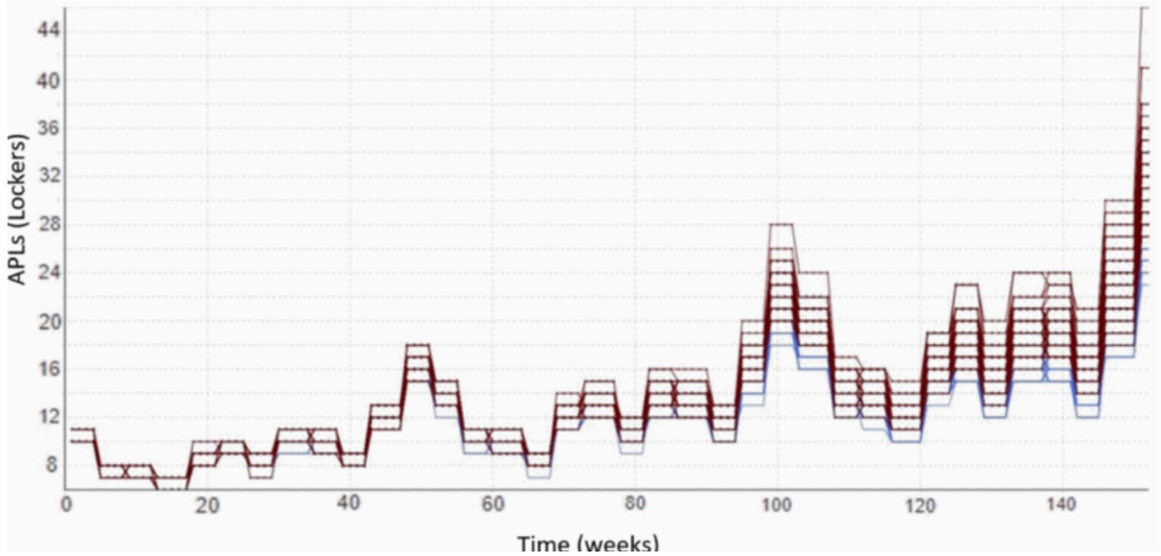
Treat not-suitable ZIPs as watch-list markets: track population, eshoppers, parcels volume,...

➤ Monitored socio-economic data to gradually enter the market through pilot or shared facilities.

Case study: In Pamplona, research about parcel locker placement using socio-economic forecasts and optimisation. Dynamic replanning over three years better aligns locker counts with e-commerce growth.

(Serrano-Hernandez et al. 2021)





(Serrano-Hernandez et al. 2021)

4. Conclusion

[Summary insights from our proposal]

PROBLEM

UPS faces ongoing last-mile cost pressures from fuel, labor, and reverse-logistics. Access Points can reduce costs, but poor placement wastes capital in low-demand areas and misses consolidation in high-potential ZIP codes.

SOLUTION

- This proposal develops a **supervised machine learning** framework to **predict** which U.S. ZIP codes have the highest short-term (6-month) potential for successful UPS Access Point placement.
- By integrating socio-demographic demand indicators, digital adoption proxies, and spatial activity measures, the **XGBoost classifier** generates probability-based suitability scores that directly support capital prioritization and deployment planning.

EVALUATION

XGBoost delivers the best performance (ROC-AUC 0.781) under class imbalance, surpassing Logistic Regression and Random Forest. **Robust calibration** shows stable results across income deciles, and a ~60% threshold balances false positives with missed opportunities. The model is fit for decision support, not full automation.

DATA ETHICS

Data are public, de-identified, US. Census data source compliant. Variables follow literature, transparency limits noted, fairness ensured via broad digital access and employment patterns.

MODEL LIMITATION

The model is predictive, not causal: XGBoost flags correlations with Access Point success but cannot isolate true ROI. Confounders like zoning, competition, or operations may bias results. It should serve as a first-stage tool, followed by causal methods (PSM, DiD, IV) before scale-up.

5. AI Evaluation

*[Generative AI Usage & Disclosure (with Verification)- AI Tool Used: **Gemini, Claude**]*

What We Asked	AI Tools Use	Output
Deployment code to identify and fix complex and fabricated data.	Gemini suggests Python Code to Clean data	Reduce from **33773** samples to **29,943** samples
Deployment code to run Descriptive Statistics and plot the necessary figures	Gemini suggests Python Code to run Descriptive Statistic and Visualisation Data	Successfully implementing the Descriptive Statistics and plotting figures
Deployment code to run the three ML Model (Random Forest, Logistic and XGBoost)	Gemini suggests Python Code to conduct three RandomForest, LogisticRegression and XGBoost models	Successfully implementing the three ML Models
Guide to implement code for evaluated XGBoost Classification Mode	Claude guides to implement code to evaluated ML Model.	Successfully evaluating the XGBoost model
Generate random test cases and implement the prediction function	Claude generate 20 random testcase and implement the prediction feature	Successfully pass all test cases with the prediction feature


THANK YOU FOR
YOUR LISTENING
RMIT
UNIVERSITY

References

Airlangga G (2024) ‘Comparative Study of XGBoost, Random Forest, and Logistic Regression Models for Predicting Customer Interest in Vehicle Insurance’, *Accredited National Scientific Journal*, 8(4), doi:10.33395/sinkron.v8i4.14194.

Arsic S, Mihic M, Petrovic D, Mitrovic Z, Kostic SC & Mihic O (1-2 February 2024) ‘Review of measures for improving ML model interpretability: Empowering decision makers with transparent insights’ [PDF], *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, Victoria, accessed December 18, 2025. https://doi.org/10.1109/acdsa59508.2024.10467907

Barua L, Zou B, Zhou Y, Liu Y (2021) ‘Modeling household online shopping demand in the U.S.: a machine learning approach and comparative investigation between 2009 and 2017’, *Transportation*, 50(2):437-476. doi:10.1007/s11116-021-10250-z.

Bosona T (2020) ‘Urban Freight Last Mile Logistics—Challenges and Opportunities to Improve Sustainability: A Literature Review’, *Sustainability*, 12(21):1-20, doi:10.3390/su12218769.

Bricker Graydon LLP (n.d.) *HIPAA Privacy Regulations: Other Requirements Relating to Uses and Disclosures of Protected Health Information: Re-Identification - § 164514(c)*, Bricker Graydon LLP website, accessed December 18 2025. https://www.brickergraydon.com/insights/resources/key/HIPAA-Privacy-Regulations-Other-Requirements-Relating-to-Uses-and-Disclosures-of-Protected-Health-Information-Re-Identification-164-514-c

Cipuac-Ulici M, Beju DG, Bresfelean VP, Zanellato G (2023) ‘Which Factors Contribute to the Global Expansion of M-Commerce?’, *Electronics*, 12(1):1-19, doi:10.3390/electronics12010197.

Eurostat (2024) *E-commerce statistics for individuals, Eurostat website*, accessed 5 December 2025. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=E-commerce_statistics_for_individuals

Forbes (2025) *Forbes Best Customer Service Companies 2026 List*, Forbes website, accessed 19 December 2025. https://www.forbes.com/lists/best-customer-service/

Fried T (2023) ‘E-commerce and logistics sprawl: A spatial exploration of last-mile logistics platforms’, *Journal of Transport Geography*, 112:1-12, doi:10.1016/j.jtrangeo.2023.103692.

Golias MM, Kaisar EI, Figliozzi MA (2023) ‘Modeling Household E-commerce Delivery Rates and Assessing Households Last-Mile Delivery Preferences’, *Freight Mobility Research Institute,* 1-43. https://rosap.ntl.bts.gov/view/dot/77000

References

Heinrich C, Maffioli A, Vázquez G (2010) 'A Primer for Applying Propensity-Score Matching', *Impact-Evaluation Guidelines*, 1-59, doi:10.18235/0008567

Joeres R, Blumenthal DB, Kalinina OV (2025) 'Data splitting to avoid information leakage with DataSAIL', *Nature Communications*, 16, doi:10.1038/s41467-025-58606-8.

Keeling K, Schafer JS and Figliozzi (2020) 'Accessibility and Equity Analysis of Common Carrier Parcel Lockers at Transit Facilities in Portland, Oregon', *Civil and Environmental Engineering Faculty Publications and Presentations*, 574

Kim Y and Steiner P (2018) 'Quasi-Experimental Designs for Causal Inference', *Educational Psychology*, 51(3-4):395-405. doi:10.1080/00461520.2016.1207177.

Kutz J (2025) *Data Leakage In Machine Learning: Examples & How to Prevent It?*, Airbyte website, accessed 18 December 2025. https://airbyte.com/data-engineering-resources/what-is-data-leakage

LaRocco LA (2024) *Retailers shift their thinking on costly last-mile delivery as consumer need for speed raises costs*, CNBC website, accessed November 14 2025. https://www.cnbc.com/2024/07/24/retailers-shift-their-thinking-on-costly-last-mile-delivery.html?msockid=16eb93d5340a6d8d0a8185b635186c39

Li and Xue (2024) 'The application of the propensity score matching method in stock prediction among stocks within the same industry', *PeerJ Computer Science*, 10:1-24, doi:10.7717/peerj-cs.1819.

Li J (2024) 'Area under the ROC Curve has the most consistent evaluation for binary classification', *Public Library of Science One (Plos One)*, 19(12): 1-28, doi:10.1371/journal.pone.0316019.

Mucci T (n.d.) *What is data leakage in machine learning?*, IBM website, accessed 18 December 2025. https://www.ibm.com/think/topics/data-leakage-machine-learning

Nolan B (2020) 'The Median Versus Inequality-Adjusted GNI as Core Indicator of 'Ordinary' Household Living Standards in Rich Countries', *Social Indicators Research,* 150:569-585, doi:https://doi.org/10.1007/s11205-020-02311-0.

Nsir N, Alzubi AB, Adegboye OR (2025) 'Enhancing Sustainable Supply Chain Performance Prediction Using an Augmented Algorithm-Optimized XGBOOST in Industry 4.0 Contexts', *Sustainability*, 17(22):1-36, doi:0.3390/su172210344.

Ozyavas P, Buijs P, Ursavas E and Teunter R (2025) 'Designing a sustainable delivery network with parcel locker systems as collection and transfer points', *Omega The International Journal of Management Science*, 131:103199, doi:10.1016/j.omega.2024.103199

References

Pitney Bowes (2025) *Parcel Shipping Index Report 2024*, Pitney Bowes website, accessed December 18 2025. https://www.pitneybowes.com/us/shipping-index.html

Raghavan S and Zhang R (2024) 'The Driver-Aide Problem: Coordinated Logistics for Last-Mile Delivery', *Manufacturing & Service Operations Management*, 26(1):291–311, doi:10.1287/msom.2022.0211

Reich BJ, Yang S, Guan Y, Giffin AB, Miller MJ, Rappold A (2023) 'A Review of Spatial Causal Inference Methods for Environmental and Epidemiological Applications', *International Statistical Review*, 89(3):605-634, doi:10.1111/insr.12452.

Schaefer J and Figliozzi M (2021) 'Spatial Accessibility and Equity Analysis of Amazon 24 Parcel Lockers Facilities'. *Journal of Transport Geography, Forthcoming*

Scott C (2024) *Why Is UPS So Expensive In 2025? (10 Reasons Shipping Costs Are Rising),* ExpertBeacon website, accessed December 18, 2025. https://expertbeacon.com/why-is-ups-so-expensive/

Serrano-Hernandez A, Martinez-Abad S, Ballano A, Faulin J, Rabe M and Chicaiza-Vaca J (12-15 December 2021), 'A Hybrid Modeling Approach for Automated Parcel Lockers as a Last-Mile Delivery Scheme: A Case Study in Pamplona (Spain)', *2021 Winter Simulation Conference (WSC*), Phoenix, accessed 18 December 2025, https://doi.org/10.1109/WSC52266.2021.9715337

Sharada R (2025) *Introduction of Holdout Method*, Geeks for Geeks website, accessed 18 December 2025. https://www.geeksforgeeks.org/software-engineering/introduction-of-holdout-method/

Sivakumar M, Parthasarathy S and Padmapriya T (2024) 'Trade-off between training and testing ratio in machine learning for medical image processing', *PeerJ Computer Sci*, 2024, doi:10.7717/peerj-cs.2245.

UNDP (n.d.) *Difference in Differences*, UNDP website, accessed 5 December 2025. https://erc.undp.org/methods-center/methods/evaluation-methods/differenceindifferences

United States Census Bureau (n.d.) *DP03 | Selected Economic Characteristics*, United States Census Bureau website, accessed December 19, 2025. https://data.census.gov/table?q=DP03:+Selected+Economic+Characteristics&g=010XX00US$8600000

United States Census Bureau (n.d.) *DP04 | ACS Demographic and Housing Estimates*, United States Census Bureau website, accessed December 19, 2025. https://data.census.gov/table?q=DP05:+ACS+Demographic+and+Housing+Estimates&g=010XX00US$8600000

References

United States Census Bureau (n.d.) *S2801 | Types of Computers and Internet Subscriptions*, United States Census Bureau website accessed December 19, 2025. https://data.census.gov/table?q=S2801:+Types+of+Computers+and+Internet+Subscriptions&g=010XX00US$8600000

UPS (United Parcel Services) (2025) *Annual Reports*, UPS website, accessed November 14 2025. https://investors.ups.com/company-profile/annual-reports

Wang L, Wu C, Zhao S (2022) 'A Review of Spatial Mismatch Research: Empirical Debate, Theoretical Evolution and Connotation Expansion', *Land*, 11(7):1-16, doi:10.3390/land11071049.

YCharts (2025) United Parcel Service Cl B (UPS) - *US Average Revenue Per Piece*, YCharts website, accessed December 18 2025. https://ycharts.com/indicators/united_parcel_service_cl_b_ups_us_average_revenue_per_piece