Methodology Justification

My approach to revenue adjustment follows a structured pipeline to ensure accuracy and robustness. First, I performed **record linkage** using address parsing and similarity matching techniques to associate financial transactions with their respective business locations. Then, I applied a **multiplier-based revenue adjustment** method, assuming that the available raw revenue represents only card transactions in someway of 80% of total revenue. To estimate total revenue, I considered the logic of different business sizes could have different payment ways and adjusted for business size (square footage) to refine my estimations.

The methodology was validated using **statistical distributions**, ensuring that adjustments align with expected revenue proportions. If actual sales data were available, I would further validate using **hypothesis testing** and **error metrics** like MAE, MSE, and R^2 to compare different adjustment models.

Key Assumptions

1. **Card Transactions Coverage** – I assume that the raw revenue represents approximately 80% of total revenue, based on industry standards and sources on the internet.
2. **Payment Distribution Consistency** – I assume that the proportion of transactions made via different payment methods remains relatively stable over time and across similar business types.
3. **Store Size Correlation** – Larger stores tend to have higher revenues, leading to weight the revenue adjustment based on store square footage.
4. **Data Matching Accuracy** – My record linkage process successfully aligns financial transactions with business locations without significant mismatches.

Solution Limitations

- **Lack of Ground Truth** – Without actual verified revenue data for validation, my method relies on external sources and news on the internet, which may introduce biases.
- **State-Level Generalization** – Payment method distributions are applied at a state level, potentially overlooking local variations in cash usage and digital payment adoption.
- **Multiplier Sensitivity** – The revenue adjustment heavily depends on the assumed multiplier, which may not capture real-world variations across different business sectors.
- **Potential Data Matching Errors** – Address parsing and record linkage matching may introduce occasional mismatches, affecting the accuracy of financial attributions, due for it, maybe some business couldn't be matched in between tables.

Despite these limitations, my approach provides a **structured, scalable**, and **statistically sound** method for estimating total revenues from partial transaction data.