Research Question: How do variations in pharmacy reimbursement rates influence the closure rates of pharmacies in the US?

Link to the Spreadsheet:
https://docs.google.com/spreadsheets/d/1z6unTUrCVujxDsQOuH12EpJyn3VDGBUURzrwC_8T-uE/edit?gid=1589256610#gid=1589256610

Datasets To Explore

New Datasets (As of 12/03)

Data on Reimbursement Rates

Average Sales Price (ASP) Methodology: For medications administered in clinical settings (e.g., chemotherapy), Medicare Part B reimburses providers at 106% of the drug's Average Sales Price (ASP). The ASP is calculated quarterly by manufacturers, reflecting the average price after discounts and rebates.

1. **AACP - Pharmacy Closures in the US**

   Summary: The AACP provides comprehensive data on pharmacy closures across the United States, highlighting trends over time and distinguishing between independent and chain pharmacies. This dataset offers insights into regional patterns and the factors contributing to these closures.

2. **Health Affairs - More US Pharmacies Closed Than Opened In 2018–21**

   Summary: A study published in *Health Affairs* reveals that nearly 30% of U.S. drugstores operational in the past decade had closed by 2021. The closures disproportionately affected Black and Latino neighborhoods, exacerbating existing healthcare access issues in these communities.

Pharmacy Reimbursement Rates Data:

1. **CMS Medicaid Spending by Drug**

   Summary: The Centers for Medicare & Medicaid Services (CMS) provides

detailed data on Medicaid spending for prescription drugs. This dataset offers insights into reimbursement rates over time, which is crucial for analyzing financial pressures on pharmacies.

2. **National Average Drug Acquisition Cost (NADAC)**

   Summary: NADAC reflects the average invoice price pharmacies pay for medications in the U.S., offering a benchmark for drug acquisition costs. This data is essential for understanding the cost dynamics that pharmacies face.

Retail Pharmacy Economic Impact Reports:

1. **NCPA - Economic Impact Reports**

   Summary: The NCPA publishes reports on the economic health of independent pharmacies, including data on profit margins, reimbursement challenges, and closure statistics. These reports provide valuable context on the financial pressures facing community pharmacies.

San Francisco-Specific Data:

1. **California State Board of Pharmacy - Licensee Lists**

   Summary: This resource provides information on licensed pharmacies in California, including openings and closures. The data can be filtered to focus on San Francisco, offering localized insights into pharmacy operations.d

2. **San Francisco Department of Public Health - Health Indicators**

   Summary: This dataset includes health service availability metrics in San Francisco, potentially offering insights into pharmacy distributions and closures. It can help assess the impact of pharmacy availability on public health outcomes.

   Pharmacy Closures

Notes on Statistics

1. **Independent Variable (x):**

    - **Pharmacy Reimbursement Rates:** This variable represents the amount of money pharmacies receive from insurers or government programs. It is the factor you expect to have an influence on the outcome.

2. **Dependent Variable (y):**

    - **Pharmacy Closure Rates:** This is the outcome variable, indicating the frequency or probability of pharmacy closures over a given period. The research aims to see how changes in reimbursement rates (x) affect this rate (y).

Modeling the Relationship

A common way to quantify the relationship between two variables is by using a simple linear regression model, which can be expressed as:

```
y = mx + b
```

- **m (Slope):** Represents the expected change in the pharmacy closure rate for each unit change in the reimbursement rate. For example, if m is positive, an increase in reimbursement might be associated with an increase in closure rates (or vice versa, depending on the context and data), although the hypothesis might be that higher reimbursement rates reduce closures.

- **b (Intercept):** Represents the predicted pharmacy closure rate when the reimbursement rate is zero. While a zero reimbursement rate might be an extreme or unrealistic scenario, b helps anchor the model.

In this setup, our goal is to determine the values of m and b that best describe the relationship observed in the data.

Statistical Measures for Model Evaluation

After fitting the model to your data, you would use several statistical metrics to assess its performance and reliability:

1. **Sum of Squares:**

    - **Total Sum of Squares (SST):** Measures the total variability in the observed pharmacy closure rates.
    - **Residual Sum of Squares (RSS):** Measures the variability in closure rates that remains unexplained by your model. A lower RSS indicates that the model's predictions are closer to the actual data.
2. **Coefficient of Determination (R^2):**

    - R^2 quantifies the proportion of the variance in the dependent variable that is explained by the independent variable in your model. An R^2 close to 1 indicates that the model explains most of the variability in the data, whereas an R^2 near 0 indicates a weak relationship.
3. **Root Mean Squared Error (RMSE):**

    - RMSE is the square root of the average of the squared differences between the observed and predicted values. It provides a measure of the typical error size in your predictions. Lower RMSE values suggest that your model's predictions are more accurate.

Steps for Pharmacy Project

1) **Load Data**
2) **Change Data into usable form in Python**
    a) **Using a dictionary**
    b) **Cleaning gaps in the data**
3) **Visualize Data**
    a) **Print out rows and columns**
    b) **Plot the data**
4) **Come up with an equation or possible line of Normal Equation**
5) **Plot the line on the data**
6) **Measure the fit (Using Root Mean Squared Error (RMSE), which measures the average magnitude of prediction errors.)**
7) **Plot line versus the test data**
8) **Measure the fit (Using R^2, RSS, RSE)**
    a) **R-squared (R^2): Indicates the proportion of variance in the dependent variable that is predictable from the independent variable(s).**

b) **Residual Sum of Squares (RSS): Represents the sum of the squared residuals (differences between observed and predicted values).**
c) **Residual Standard Error (RSE): Provides an estimate of the standard deviation of the residuals.**

Measure the same fit for other datasets including crime rate, rents, and the pharmacy closures from before 2013
