

JOINT STUDY: **WHO BUYS THE BUD?**

Jasmine Le & Ashley Yang

RESEARCH QUESTION

Does median income, racial makeup, type of city, crime rate, and median age play a role in marijuana sales in CA cities?

SIGNIFICANCE AND BACKGROUND

Post War on Drugs: Who is really economically benefiting
from the increase in cannabis sales?

OVERVIEW



Data Collection

Finding the relevant data and explanatory variables

Data Exploration

Creating plots to better understand our data

Model Fitting

First order model and interaction models

Model Diagnostics

Checking Model Adequacy

DATA

1

Cannabis Sales

CA Dept of Tax and Fee
Administration - 2024

2

Median Income, Age

Data Commons (Google) - 2023

3

Total Crime

US FBI- 2019

4

Population

Data Commons (Google) - 2023

5

Racial Makeup

US Census Bureau

6

Type of City

Niche

DATA SUMMARY

	Overall (N=53)
other	
Mean (SD)	18.7 (9.36)
Median [Min, Max]	17.2 [6.60, 46.0]
asian	
Mean (SD)	9.53 (8.14)
Median [Min, Max]	6.40 [0.100, 33.5]
black	
Mean (SD)	6.16 (6.21)
Median [Min, Max]	3.20 [0.600, 28.1]
native	
Mean (SD)	1.78 (1.82)
Median [Min, Max]	1.10 [0.400, 11.5]
hispanic	
Mean (SD)	40.1 (21.2)
Median [Min, Max]	37.3 [9.30, 98.1]
White	
Mean (SD)	63.8 (14.0)
Median [Min, Max]	64.8 [36.2, 86.7]

Med.Age	
Mean (SD)	37.0 (5.73)
Median [Min, Max]	36.8 [27.2, 57.8]
Med.Income	
Mean (SD)	40300 (10700)
Median [Min, Max]	37000 [21600, 69300]
sales.per.capita	
Mean (SD)	421 (495)
Median [Min, Max]	292 [92.3, 3480]
crime.per.capita	
Mean (SD)	0.0710 (0.0315)
Median [Min, Max]	0.0666 [0.0282, 0.172]
Urban.rural.class	
Rural	9 (17.0%)
Suburban	25 (47.2%)
Urban	19 (35.8%)

HEATMAP





1

SIMPLE ADDITIVE MODEL

Hypothesized and Predicted Model

$$\begin{aligned}\text{sales.per.capita}_i = & \beta_0 + \beta_1 \cdot \text{Med.Age}_i + \beta_2 \cdot \text{Med.Income}_i + \beta_3 \cdot \text{Urban.rural.class.Suburban}_i \\ & + \beta_4 \cdot \text{Urban.rural.class.Urban}_i + \beta_5 \cdot \text{crime.per.capita}_i + \beta_6 \cdot \text{other}_i + \beta_7 \cdot \text{asian}_i \\ & + \beta_7 \cdot \text{asian}_i + \beta_8 \cdot \text{black}_i + \beta_9 \cdot \text{native}_i + \beta_{10} \cdot \text{hispanic}_i + \beta_{11} \cdot \text{White}_i + \epsilon_i\end{aligned}$$

Where $\epsilon \sim N(0, \sigma)$

*Where *Urban.rural.class.Rural_i* is our baseline

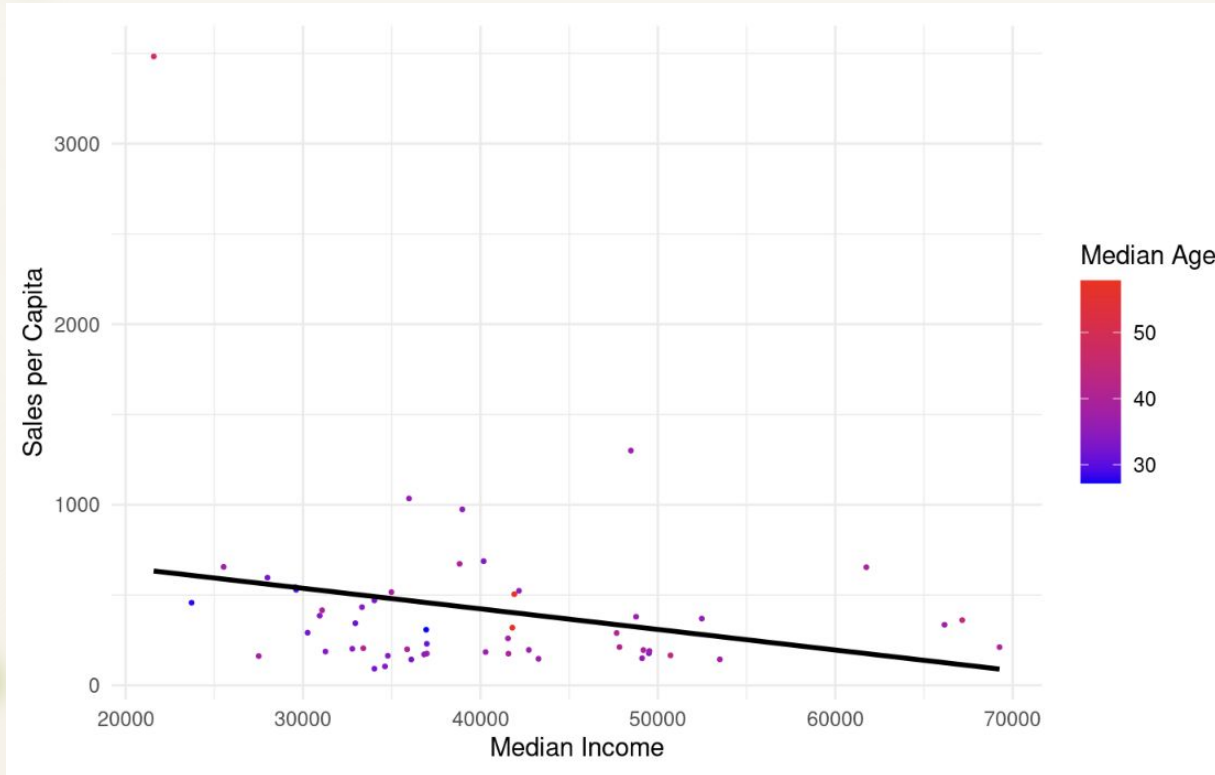
$$\begin{aligned}\text{Sales.per.capita}_i = & .000014 + 6.147 \text{Med.Age}_i + .004 \text{Med.Income}_i + 331.9 \text{Urban.rural.classSuburban}_i \\ & + 169.7 \text{Urban.rural.classUrban}_i - 562.5 \text{crime.per.capita}_i - 1399 \text{Other}_i - 1419 \text{asian}_i \\ & - 1415 \text{black}_i - 1182 \text{native}_i - 3.434 \text{hispanic}_i - 1411 \text{White}_i + \epsilon_i\end{aligned}$$

Adj. R Squared: 0.5374

2

INTERACTION MODEL 1

INTERACTION PLOT



Hypothesized and Predicted Model

$$\begin{aligned} \text{sales.per.capita}_i = & \beta_0 + \beta_1 \cdot \text{Med.Age}_i + \beta_2 \cdot \text{Med.Income}_i + \beta_3 \cdot \text{Urban.rural.classSuburban}_i + \\ & \beta_4 \cdot \text{Urban.rural.classUrban}_i + \beta_5 \cdot \text{crime.per.capita}_i + \beta_6 \cdot \text{other}_i + \beta_7 \cdot \text{asian}_i \cdot \\ & + \beta_8 \cdot \text{black}_i + \beta_9 \cdot \text{native}_i + \beta_{10} \cdot \text{hispanic}_i + \beta_{11} \cdot \text{White}_i + \beta_{12} \cdot \text{Med.Age} * \text{Med.Income}_i + \epsilon_i \end{aligned}$$

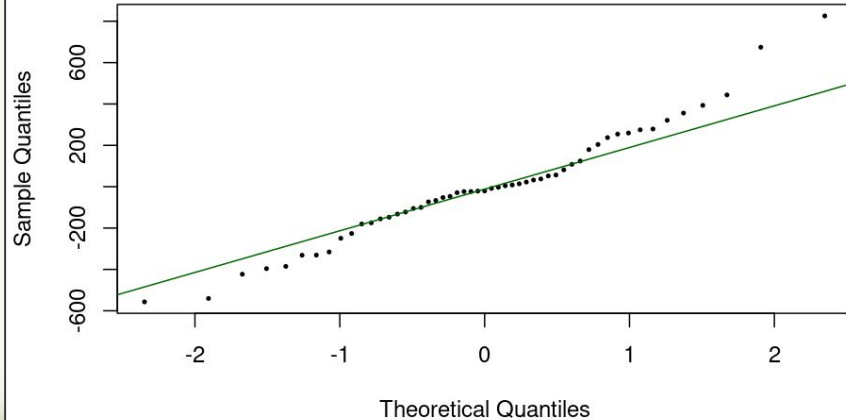
Where $\epsilon \sim N(0, \sigma)$

*Where *Urban.rural.class.Rural_i* is our baseline

$$\begin{aligned} \text{sales.per.capita}_i = & .00001 + 118.9\text{Med.Age}_i + .1170\text{Med.Income}_i + 217.8\text{Urban.rural.classSuburban}_i + \\ & .5662\text{Urban.rural.classUrban}_i - 474.6\text{crime.per.capita}_i - 1391\text{other}_i - 1412\text{asian}_i \\ & - 1408\text{black}_i - 1225\text{native}_i - 3.790\text{hispanic}_i - 1405\text{White}_i - .002\text{Med.Age} * \text{Med.Income}_i \end{aligned}$$

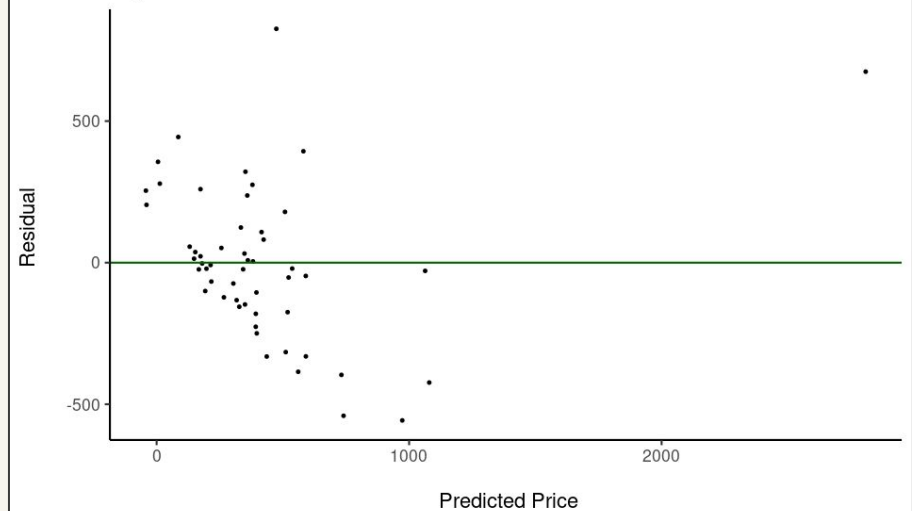
Model Assumptions

Figure 1. Normal Q-Q Plot



Normality

Figure 2. Residual Plot



**Linearity & Equal
Variance**

SUMMARY

Variable	Coefficient (p-value)
Med.Age	118.9 (0.006) ***
Med.Income	0.117 (0.006) ***
Med.Age * Med.Income	-0.003 (0.007) ***

Adj. R Squared: 0.6047

3

Interaction Model 2

Hypothesized and Predicted Model

$$\begin{aligned} \text{sales.per.capita}_i = & \beta_0 + \beta_1 \cdot \text{Med.Age}_i + \beta_2 \cdot \text{Med.Income}_i + \beta_3 \cdot \text{Urban.rural.classSuburban}_i + \beta_4 \cdot \text{Urban.rural.classUrban}_i + \beta_5 \cdot \text{crime.per.capita}_i \\ & + \beta_6 \cdot \text{other}_i + \beta_7 \cdot \text{asian}_i + \beta_8 \cdot \text{black}_i + \beta_9 \cdot \text{native}_i + \beta_{10} \cdot \text{hispanic}_i + \beta_{11} \cdot \text{White}_i + \beta_{12} \cdot \text{Med.Age} \cdot \text{Med.Income}_i \\ & + \beta_{13} \cdot \text{Suburban} \cdot \text{Crime}_i + \beta_{14} \cdot \text{Urban} \cdot \text{Crime}_i + \epsilon_i \end{aligned} \quad \text{Where } \epsilon \sim N(0, \sigma)$$

*Where *Urban.rural.class.Rural_i* is our baseline

$$\begin{aligned} \text{sales.per.capita}_i = & 180800 + 117.9 \text{Med.Age}_i + .1113 \text{Med.Income}_i - .9443 \text{Urban.rural.classSuburban}_i - 682.1 \text{Urban.rural.classUrban}_i \\ & - 8938 \text{crime.per.capita}_i - 1827 \text{other}_i - 1858 \text{asian}_i - 1859 \text{black}_i - 1644 \text{native}_i - 7.580 \text{hispanic}_i - \\ & 1848 \text{White}_i - .002 (\text{Med.Age}_i \times \text{Med.Income}_i) + 6061 (\text{Urban.rural.classSuburban}_i \times \text{crime.per.capita}_i) \\ & + 11760 (\text{Urban.rural.classUrban}_i \times \text{crime.per.capita}_i) \end{aligned}$$

INTERACTION PLOT



SUMMARY

Variable	Coefficient (p-value)
Med.Age	117..9 (0.005) ***
Med.Income	0.111 (0.007) ***
Med.Age * Med.Income	-0.003 (0.007) ***

Adj. R Squared: 0.627

RESULTS

Model Fit (Adjusted R Squared): 60.47%

Significant Predictors:

- Median Income (Individual) (+)
- Median Age (+)
- Median Income * Median Age (-)

DISCUSSION

Limitations:

- Limited dataset (not enough diverse cities in the dataset) – Overfitting
- Outdated Data (2019 – Crime)
- Omitted Variable Bias – Education, (Local) Business Density
- Equal Variance Assumption (?)

Recommendations & Future Work:

- PCA, LASSO, stepwise regression (reduce model complexity)
- Possible remediation of Equal Variance Assumption (Box-Cox)
- Data from different cities

REFERENCES

- "Open Data Portal." *Wwww.cdtfa.ca.gov*,
www.cdtfa.ca.gov/DataPortal/dataset.htm?url=CannabisSalesByCounty.
- "Download Tool - Data Commons." *Datacommons.org*, 2025,
datacommons.org/tools/download#pt=City&place=geold%2F06&sv=Median_Age_Person__Median_Income_Person&dtType=LATEST&facets=%7B%7D. Accessed 27 Apr. 2025.
- Federal Bureau of Investigation. "California." *FBI*, 2019,
ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/tables/table-8/table-8-state-cuts/california.xls.
- "Niche." *Niche*, 2025, www.niche.com/places-to-live. Accessed 27 Apr. 2025.
- US Census Bureau. "Census Datasets." *Census.gov*, 3 Jan. 2024, www.census.gov/data/datasets.html.