Project 4:
GMV Signal Analysis (Statistical Analysis)

In my GMV Signal Analysis project at Wonders, I employed Pearson's Chi-squared test and linear regression to identify how median household income and restaurant location impact GMV.

This analysis confirmed significant correlations, demonstrating that restaurants in lower-income zip codes generate higher GMV.

These insights are crucial for refining our marketing strategies, allowing us to effectively target and serve our clients based on robust data-driven evidence.

GMV Signal Analysis - Statistical Analysis

GMV Market Signal Analysis | Code for Calculation

The analysis employs two statistical methods, **Pearson Chi-Square** and **Linear Regression**, to assess the statistical significance of our preliminary findings from the **GMV Market Signal Analysis**.

→ **Pearson Chi-Square Test**: Utilized to discern differences among two or more categories of data.
i.e. school zones (categorical variable) v.s. GMV tier (categorical variable).

→ **Linear Regression**: Employed to identify correlations between continuous numeric variables and other numeric or categorical variables.
i.e. average monthly GMV (continuous **numeric** variable) v.s. Region (categorical variable).

The rationale for selecting these two statistical methods over merely evaluating the confidence level is that the confidence level primarily indicates how likely it is that a sample falls within a particular range, without elucidating the relationships between the groups or variables under examination. This distinction is critical as our objective extends beyond understanding the range of data points to uncovering the underlying dynamics between distinct groups or variables.

Rule of Thumb:
- p-value < 0.001 - Very Strong Evidence
- p-value < 0.01 - Strong Evidence
- p-value < 0.05 - Some Evidence
- p-value < 0.1 - Very weak evidence
- p-value ≥ 0.1 - No evidence

Terminology & Examples:	
P-value:	Assesses evidence against the null hypothesis in hypothesis testing, used to decide if study results are meaningful.
Confidence Level:	Represents the degree of certainty that the parameter lies within the specified interval, indicating how sure results fall within a range.
Statistical Significance:	Indicates whether the result of an analysis reflects a true effect rather than random variation.

*p-value ≤ 0.05 indicates statistical significance at the 95% confidence level.

Test Group - Association w/ GMV	Initial Conclusion	Statistcal Singnificance	P-Value*	Statistiscal Evidence	Testing Variables/Features	Methods
1. Income Level	The lower the median household income of the restaurant zipcode, the higher the GMV from our clients.	☑	0.0185	strong	Median Household Income x GMV Tier (categorical x categorical)	Pearson's Chi-squared
2. Population Density	No correlation between population density and the GMV levels of our voice platform clients.	☑	0.0474	less strong	Population Density Band x GMV Tier (categorical x categorical)	Pearson's Chi-squared
3. Region	Our client base (active + churned), average GMV per client, and ICP SAM all trend higher from the east towards the west.	☑	0.0112	strong	Region x Average Monthly GMV (categorical x numerics)	Linear Regression
4. School Zone Rating	The lower the client's zipcode's local school zone rating, the higher the average monthly GMV.	☑	0.0025	strong	School Zone Rating x GMV Tier (categorical x categorical)	Pearson's Chi-squared

Hypothesis:	The lower the median household income of the restaurant zipcode, the higher the GMV from our clients.		
Method:	Pearson's Chi-squared test	with simulated p-value (based on 2000 replicates)	*≤0.05 - statistical significace
		X-squared	26.397
		df	NA
		p-value	0.01849
Interpretation:	*The p-value is below 0.05, which means we can reject the null hypothesis that there is no association between Median Household Income and GMV Tiers. This suggests that there is a statistically significant relationship between the two variables.* *The significant result implies that the GMV Tiers vary across different income brackets. This could mean that customer spending behavior on our platform is influenced by their income level.*		

	GMVTier				
IncomeRange	< $10,000	$10,000 to $29,999	$30,000 to $49,999	$50,000 to $99,999	>$100,000
$1 - $60,000 (Low)	197	483	178	52	4
$100,001 - $150,000 (Mid)	122	254	60	13	1
$150,001+ (High)	20	23	5	2	0
$60,000 - $100,000 (Low-mid)	357	782	247	59	3