1) median housing income for national, what story doe this tell?

2) median housing income for state, what story doe this tell?

3) expenditure by state  
financial health index

I'm building an end-to-end **ETL data pipeline project** titled **"Malaysia Household Financial Health & Cost of Living Index"**, leveraging **public datasets** from trusted sources such as:  
  
Department of Statistics Malaysia (DOSM)  
  
📌 Project Scope & Goals capable of demonstrating: 1. Data engineering proficiency 2. Solving real world issues 3. Applied machine learning 4. Dashboard and visualization skills  
  
Critical Requirement 1.   
Help me outline the project outcome detail,   
1) how can I use the columns and table data below to display what types of charts on the dashboard   
2) what ML model can I train?   
3) what are the KPI cards i should show?   
4. what types of questions can this dashboard help answer?

**Table 1: National-Level Indicators**

| **Column** | **Explanation** |
| --- | --- |
| **year** | Calendar year of the data (e.g., 2022). |
| **population\_national** | Total resident population of Malaysia in that year. |
| **inflation (%)** | Annual change in the Consumer Price Index (CPI), showing percentage increase in average prices. |
| **gdp\_national (RM)** | Gross Domestic Product at current market prices, in Malaysian Ringgit (total economic output). |
| **gdp\_national\_per\_capita (RM)** | GDP divided by total population, showing average economic output per person (in RM). |
| **labour\_force\_national** | Number of residents aged 15+ who are either employed or actively seeking work. |
| **employed\_persons\_national** | Count of labour-force participants who have worked for pay or profit in the reference week. |
| **unemployed\_persons\_national** | Count of labour-force participants without work but actively seeking and available for work. |
| **outside\_labour\_force\_national** | Residents aged 15+ who are neither employed nor actively seeking (e.g., students, retirees, homemakers). |
| **unemployment\_rate\_national (%)** | (Unemployed ÷ Labour force) × 100; percentage of the labour force without work. |
| **participation\_rate\_national (%)** | (Labour force ÷ Working‑age population) × 100; share of 15+ population economically active. |
| **income\_mean\_national (RM)** | Arithmetic average of all annual household incomes (in RM). |
| **income\_median\_national (RM)** | The middle value of the annual household‑income distribution (in RM), splitting households into two equal halves. |
| **poverty\_relative\_national\_mean\_income (%)** | Percentage of households with income below a defined poverty line, often set at a fraction (e.g., 50%) of the national mean household income. |

**Table 2: State-Level Indicators**

| **Column** | **Explanation** |
| --- | --- |
| **year** | Calendar year of the data (e.g., 2022). |
| **state** | Name of the Malaysian state or federal territory (e.g., Selangor, Johor, Kuala Lumpur). |
| **population\_state** | Total resident population of that state in the given year. |
| **gdp\_per\_state (RM)** | Gross Domestic Product of the state at current market prices, in Malaysian Ringgit. |
| **labour\_force\_state** | Number of state residents aged 15+ who are employed or actively seeking work. |
| **employed\_persons\_state** | Count of state labour‑force participants who are employed. |
| **unemployed\_persons\_state** | Count of state labour‑force participants who are actively seeking work but currently jobless. |
| **outside\_labour\_force\_state** | State residents aged 15+ not in the labour force (e.g., students, retirees). |
| **unemployed\_rate\_state (%)** | (Unemployed\_State ÷ Labour\_Force\_State) × 100; percentage of the state’s labour force without work. |
| **income\_mean\_state (RM)** | Arithmetic average of annual household incomes within the state (in RM). |
| **income\_median\_state (RM)** | Middle value of the state’s household‑income distribution (in RM). |
| **expenditure\_mean\_state (RM)** | Arithmetic average of annual household expenditures (in RM), indicating typical spending levels. |
| **expenditure\_median\_state (RM)** | Middle value of the state’s household‑expenditure distribution (in RM). |
| **poverty\_relative\_state\_mean\_income (%)** | Percentage of state households with income below a relative poverty line (commonly set as a fraction of the state’s mean household income, e.g., 50%). |

Title: **Malaysia Household Financial Health & Cost‑of‑Living Index**

**Key Business Questions Your Dashboard Answers**

1. **How has household financial health evolved over time?**
2. **Which regions are most/least “financially healthy”?**
3. **Is income growth keeping pace with cost‑of‑living increases?**
4. **Where are the highest concentrations of poverty or debt risk?**
5. **What is the correlation between unemployment and financial health?**
6. **Projected trends: Will any region’s financial health fall below a critical threshold?**

**Time‑Series Forecasting**

* **Cost of Living with linear regression**

**Clustering**

* **Goal: Segment regions (or household clusters) into “High‑Risk”, “Moderate”, “Healthy”**
  1. **Survival threshold in each state:** survival\_income\_state ≈ expenditure\_mean\_state × (1 + inflation\_rate) × buffer\_factor
  2. **Predict mean income -** Multivariate Linear Regression to forecast median household income in each state for the next 5 years
  3. **Predict Future Cost of Living**
  4. **Compare Income vs. Cost of Living –** 
     1. financial\_stress\_index = income\_mean\_state / expenditure\_mean\_state
     2. 1.0 → Income is higher than average expenses → **comfortable state**
     3. <1.0 → Income is lower than cost of living → **financial stress**
     4. ≈1.0 → Breakeven
  5. Classify **states into financial health clusters - (e.g., "High Risk", "Stable", "Affluent")**

National

* Median household income vs inflation rate
* Household income inequality
* Malaysia household economic index