**MARKET ANALYSIS REPORT FOR NATIONAL CLOTHING CHAIN**

With the Aim of this project is to help create a targeted marketing campaign by advertising specific products (Shirts: $25, Sweater: $100, and Leather Bag: $1,000) to specific customers in specific locations.  
  
Using linear Regression to support and mainly as the foundation of my Marketing Analysis which I will be using suggested X and Y variable and choosing another X and Y variables to analyse further and to support the suggested variables. So, in this Project the below variable will be the ones I will be using in this project;

Suggested X and Y variables  
  
1. Predicting Average Income from Average Purchase   
X = Average Sales/Purchase (Per State)  
Y = Average Income (Per State)  
  
NB: The above variable are group or base on state level.  
  
Analysis Before Correlation Coefficient and Linear Regression Analysis / State Level  
  
**Histogram Distribution: Income distribution on state level**  
  
Shape of Distribution  
The highest frequencies are in the middle (81,338 – 87,007 and 87,007 – 92,676 which make the distribution look a bit like normal distribution (roughly bell-shaped). As shown in the Power Bi Report Page 2.

**Heat Map: House income by Location**

The like of Alaska, Hawaii, Washington, California and Colorado are top state with the Highest average income for customers per sate

**Scatter plot with Trendline:**  
 Correlation between average household income by state and average six-month sales by state indicate a strong Correlation R^2 = 0.78 which mean Purchase strongly depend on Income and Can go ahead to use this support/archive the aim of this project.

* R^2 = 0.78 with positive relationship (which when Purchase/Sales increase as Income across state increase.

**Analysis Questions**

1. What is the Correlation (R2 Value) between sales and income?

* Using Average Purchases on state level as (X) and Average Income on state level as (Y), gives R^2 = 0.78 (strong positive R^2 that is good enough to predict Y)

2. What is the Correlation (R2) between Customer Ratings and Product return rate?

* With Customer Rating as (X) and Return Rates as (Y), the correlation indicates R^2 = 0.67 (A moderate correlation) which stipulate a negative relationship stating that when X increase, the Y decreases.

3. What are the linear regression formulas to predict customer income from customer Sales

* Since I set up Sales/Purchases as x and using y = mx + b predicts Customer Income, the linear regression formula;  
    
  n = 51, m = 72.43, b = 72638.21

**y = 72.43\*(x) + 72638.21**

* Example  
  Where x = 345 in sales
* **y = 72.43\*(x) + 72638.21  
  y = 97,630.00**

4. Which product will be advertised the most

I am going to use two histogram visual diagram and decomposition Tree visual to illustrate this, However the $25 Shirt will be advertised the most and reason being that is because.

* As I have created an 8bins range categorising the predicted customer Income, the histogram a right skewed position meaning smaller income customers are pulling data to the right
* Which make me create a conditional recommended 3bins range category, condition a group of lower earner customers to be recommended to $25 Shirt, moderate earner be recommended to $100 sweater and higher earner customers to be recommended to $1,000 Leather bag product and in this case, the low earner pull the data to right
  + 679 customers fall under the bracket Lower Earner = $25 Shirt product
  + 288 Customers fall under the bracket of Moderate Earner = $100 Sweater Product.
  + Lastly, only 33 Customers fall under the income bracket of high earner = $1,000 Leather Bag product.
* The decomposition tree also supported this

CONCLUSION   
There is a strong correlation between Average Income and Average Purchase and using linear regression formular to predict Average income (y) from Average Purchases/Sales help to know what products to sell to what customers.

**FURTHER SUPPORTING ANALYSIS**

Other Choosing Variables   
  
2. Predicting Total Purchases from Total Population - Location Targeting X = Total Population (On State Level)  
Y = Total Purchase (On State Level)  
  
NB: The above X and Y variables are group or base on state level as well  
- Aim is to see how we can use X and Y variable specific state each of the specific product can be advertise at  
- larger population = Higher total purchases + high Avg Income - Indicate Strong Market Area to target.  
  
Analysis Before Correlation Coefficient and Linear Regression Analysis / State Level  
  
**Histogram Distribution: Predicted Purchase distribution on state level**  
  
Shape of Distribution  
The highest frequencies are in the middle (683.83 – 6345.63 and 6345.63 – 120007.43 which make the distribution to be right skewed meaning the smaller purchase is dragging the data to the right). As shown in the Power Bi Report Page 2.

**Scatter plot with Trendline:**  
 Correlation between population by state and Total purchases by state also indicate a strong Correlation R^2 = 0.74 which mean Total Purchase strongly depend on population and can go ahead to use this support/supplement the aim of this project.

* R^2 = 0.74 with positive relationship (which mean when Population increases the total purchases across state also increase).

**Analysis Questions**

1. What is the Correlation (R2 Value) between Population and Purchases?

* Using Population on state level as (X) and Total purchase on state level as (Y), gives R^2 = 0.74 (strong positive R^2 that is good enough to predict Y)

2. What are the linear regression formulas to predict customer income from customer Sales

* Since I set up Sales/Purchases as x and using y = mx + b predicts Customer Income, the linear regression formula;  
    
  n = 51, m = 0.00241, b = -1896.58

**y = 0.00241\*x + (-1896.58)**

* Example  
  Where population (X) = 19,078,101.00

**y = 0.00241\*19,078,101.00 + (-1896.58)  
Predicted Purchases (y) = 45878.22**

4. Which product will be advertised the most

From the regression and segmentation analysis, the **$25 Shirt will be advertised the most**, because the majority of states have **larger populations but only moderate to low-income levels**. This aligns with affordability and ensures higher conversion rates, while **sweaters and bags** are reserved for smaller, more affluent niche markets.

5. What product should be advertised to each state and why??

To achieve this, I created segmentation bins (High/Medium/Low) for **Population, Income, Purchases**.

**$25 Shirt** is recommended if meet the below scenarios:

* Moderate Purchase + Medium Population + Low Income = 5 State
* Moderate Purchase + Low Population + Low Income = 3 State
* Low Purchase + Low Population + Moderate Income = 11 State
* Low Purchase + Low Population + Low Income = 11 States

**$100 Sweater** is recommended if meet the below scenarios:

* Moderate Purchase + Medium Population + High Income = 6 State
* Moderate Purchase + Medium Population + Moderate Income = 4 Sate
* High Purchase + High Population + Moderate Income = 3 State
* High Purchase + Medium Population + Moderate Income = 2 State
* Low Purchase + Low Population + High Income = 5 State

**$1,000 Leather Bag** is recommended if meet the below scenarios:

High Purchase + High Population + High Income = 1 State

CONCLUSION

This further analysis support and contributed to previous regression analysis which corelation between Population (X) and Total Purchases indicate a strong R^2 and segmentation analysis by creating bins (High/Medium/Low) for **Population, Income, Purchases**. I was able to Observed that **most combinations across the dataset** leading back to the $25 Shirt, same as suggested regression analysis