Project Conclusion and Insights

Campaign Acceptance:

- **The effects of adjusted columns on decision tree algorithm**
  After running the decision tree algorithm on the data, in order to create a model that predicts campaign acceptance, we can see that in most campaigns the model accuracy has either improved or stayed the same but resulted in a simpler decision tree with less decision parameters.
  The only exception was on campaign 1 decision tree were the adjusted columns lowered the algorithm accuracy. The reason is because the decisive parameter for all the decision trees was mainly **Income** while in campaign 1 **Teenhome** was also important.

- **Decision tree predictions**
  Upon inspecting the decision tee accuracy on the test data, using sklearn CunfusionMatrix, we can observe that the decision tree model is very accurate in predicting costumers that will not accept the campaign (accuracy ranging around 95% with all campaigns), but mostly inaccurate in predicting campaign acceptance (even low as 0% accuracy). This though is not a problem nor surprising. Since campaign acceptance in general in the population is around 7% to all campaigns, there is very little data for the model to became precise on predicting campaign acceptance. But because **Income** is the major predictive parameter, and the accuracy in predicting non acceptance is so high, the model still provides a very good prediction tool for campaign acceptance in general.

- **Campaign acceptance**
  Over all we can conclude the **Income** is the most decisive parameter to predict campaign acceptance. The results of the decision tree prediction, together with the **'Campaign Acceptance by Income'** graph on the **Campaign_Exploration** tab on the excel file and the campaign correlation heat map, all confirm that campaigns 1 & 5 were more successful with high income population, campaigns 2 & 4 with middle rage income population and campaign 1 with low income population.

Other Conclusions and Insights:

- **Product purchases**
  By visually analyzing the data in the **Products_Exploration** tab on the **marketing_data.xlsx** file, and comparing it to the population distribution across most parameter from the **Customer_Explroation** tab, we can clearly see that every part of the population is buying the share of products relative to its size in the population. This is even more obvious if one examines the pie charts comparison in the **Purchases_Vs_Precntages** tab. So by analyzing product purchases by either **Education, Marital Status, Kidshome** & **Teenshome**, we do not find any meaningful insights.

One meaningful insight here is the composite of the shopper's population which is mostly in a relationship (either married or together) with graduation level education, 0-1 children at home and between the ages of 40-70.
One thing is important to notice on the distribution of purchases by income. We can observe on the **'Average Purchases by Income'** graph on the **Products_Exploration** tab that the higher the income the more on average the customers spend, with the 79,000 - 91,000 income group spending the most on average. But by looking on **the 'Sum of Purchases By Income'** graph we observe the most earnings are from the 63,000 - 79,000 income group.
If we relate this conclusion to the conclusions from the campaign acceptance research, we will see that campaigns 1, 4 & 5, that were most successful for costumers with income in the range of 59,000-87,000, are the most likely to produce significant incomes to the company.

- **Venue purchases**

  We can observe from the **Venue_Exploration** tab, on the **'Toatal Deals'** Graph, that most of the purchases are made through the store with 12969 purchases. Followed by web sales with 9147 purchases. Deals and catalog sales are around 5500 purchases each.

  Again, as in the product purchases analysis, we can see in the **Venue_Exploration** tab that across most population parameters the purchases by venue distribution are identical to the general parameter distribution in the population.
  Again, the meaningful insights are obtained by looking and the **'Num Deals by Income'** graph. It is clearly seen the number of deal purchases is greater in the low to mid income population, web and store purchases is greater in the mid to high income population and catalog purchases are greater in the top high-income population.

Summary of main conclusions:

- the main population parameter to predict campaign acceptance is by far **Income**.
- Campaigns 1 & 5 were successful among the high-income population (in the range of 71,000 – 91,000 income).
- Campaigns 2 & 4 were successful among the middle-income population (in the range of 47,000 – 75,000 income).
- Campaign 1 was successful among the low-income population (in the range of 31,000 – 47,000 income).
- Most sales profits are from the 63,000 - 79,000 income group.
- Most sales are from the store or web site purchases.
- Most shoppers are in a relationship, with graduation level education with 0-1 children at home and between the ages of 40-70.
- **We there for conclude that campaigns directed towards high-income population (similar to campaigns 1,4,5) and with exposure to store and website shoppers have the highest potential to produce significant profits.**