1. Describe the Data

- **Dataset:** I used the shoes_sales_dataset.csv which contains transaction-level data for footwear sales.
- **Outcome Variable:** Sales_Channel (Categorical: "Online", "Retail Store", or "Mall").
- **Features:** The primary features used for prediction include Revenue_USD, Price_USD, Shoe_Type, Brand, and Country.
- **Insights:** This analysis helps a business understand where specific products are most likely to sell. By linking price points and shoe types (like Sneakers or Formal shoes) to the sales channel, companies can optimize their inventory—for instance, deciding whether to stock a high-priced formal shoe in a physical retail store or focus on online marketing.

2. Best Value for Minimum Split

- **Best Value: 50**
- Based on the loop performed in the Jupyter Notebook, the min_samples_split of **50** yielded the highest accuracy of **0.3533**. While the overall accuracy is low due to the complexity of the dataset, a split of 50 helps prevent the tree from becoming too "noisy" or overfit. It ensures that a node must have at least 50 samples before it attempts to split further, which leads to more reliable, generalized patterns.

3. Node with the Highest and Lowest Probability

- **Highest Probability Node:** Based on the decision tree rules, the node where **Revenue_USD > 4426.21** has the highest probability for the **"Retail Store"** class.
    - *Insight:* High-revenue transactions are very strongly associated with physical retail stores, suggesting that "big-ticket" bulk buys or high-end luxury purchases in this dataset happen in person rather than online.
- **Lowest Probability Node:** The branch for **Revenue_USD <= 137.45** mixed with various **Dates** shows the lowest probability (highest uncertainty).
    - *Insight:* For low-revenue items, the sales channel is highly fragmented and fluctuates based on the date and country. This indicates that for cheaper items, consumer behavior is less predictable and doesn't lean strongly toward one specific channel.

4. Examples of Data Points and Predictions

Using the logic from the generated tree, here are four predicted outcomes:

Example	Feature Values	Predicted Outcome
1	Revenue > $4,426	**Retail Store**
2	Revenue < $362 AND Shoe Type: Sneakers	**Online**
3	Revenue between $137 and $302 AND Shoe Type: Running	**Retail Store**
4	Revenue < $137 AND Country: Saudi Arabia	**Mall**