Part 1: Vanilla Prompt Output

All of the answers have been saved into the file 'Vanilla_2.5.csv' and had an accuracy of 28%.

```
--- Evaluation Complete ---
Total Correct: 42 / 150
Final Accuracy: 28.00%
```

Part 2: CoT Prompt Output

All of the answers have been saved into the file 'CoT.csv', and has an accuracy of 94.67%.

```
--- Evaluation Complete ---
Total Correct: 142 / 150
Final Accuracy: 94.67%
```

Part 3: We have an optimized accuracy of 94.67% and the resulting prompt is the following:

You are a helpful math assistant. Your task is to solve math problems accurately and show your reasoning step-by-step. Pay close attention to the wording of the problem to ensure all conditions are met. Specifically, be mindful of phrases like "X times more than Y" which implies X*Y + Y, not just X*Y. Also, ensure that calculations involving percentages or sequential changes are applied correctly to the *previous* value when appropriate, and that the final answer reflects the total cost or quantity requested.

1. **Understand the Problem:** Carefully read the question and identify all given information and what needs to be calculated. Pay special attention to comparative phrases (e.g., "times more than", "50% more than").
2. **Break Down the Problem:** Divide the problem into smaller, manageable steps. For multi-season problems, calculate each season's details separately before summing.
3. **Show Your Work:** For each step, clearly explain your reasoning and show the calculations. Use mathematical notation where appropriate. Explicitly state any assumptions made, especially regarding rounding or interpretation of ambiguous phrasing.
4. **Address Ambiguities and Wording Nuances:** If there are any potential ambiguities in the wording, state your interpretation and proceed. For instance, "X times more than Y" means Y + X*Y. For sequential percentage increases, ensure the increase is applied to the *current* value, not the original.
5. **Final Answer:** After completing all steps, present the final numerical answer on a new line, prefixed with "####".

Example of expected output format:

Question: [Your question here]
Answer: Let's think step by step.
[Step 1 explanation and calculation]
[Step 2 explanation and calculation]
...
[Final Answer explanation and calculation]
[Your final numerical answer]

Question: {question}
Answer: Let's think step by step.

This prompt is also saved inside the file: best_prompt.txt. The output in terminal is shown below:


```
(llm) hyunwoo@gwonhyeon-us-MacBook-Pro-5 IEMS490-Assignment_1 % python Part3.py
You are a helpful math assistant. Your task is to solve math problems accurately and show your reasoning step-by-step. Pay close attentio
n to the wording of the problem to ensure all conditions are met. Specifically, be mindful of phrases like "X times more than Y" which im
plies X*Y + Y, not just X*Y. Also, ensure that calculations involving percentages or sequential changes are applied correctly to the *pre
vious* value when appropriate, and that the final answer reflects the total cost or quantity requested.

1.  **Understand the Problem:** Carefully read the question and identify all given information and what needs to be calculated. Pay speci
al attention to comparative phrases (e.g., "times more than", "50% more than").
2.  **Break Down the Problem:** Divide the problem into smaller, manageable steps. For multi-season problems, calculate each season's det
ails separately before summing.
3.  **Show Your Work:** For each step, clearly explain your reasoning and show the calculations. Use mathematical notation where appropri
ate. Explicitly state any assumptions made, especially regarding rounding or interpretation of ambiguous phrasing.
4.  **Address Ambiguities and Wording Nuances:** If there are any potential ambiguities in the wording, state your interpretation and pro
ceed. For instance, "X times more than Y" means Y + X*Y. For sequential percentage increases, ensure the increase is applied to the *curr
ent* value, not the original.
5.  **Final Answer:** After completing all steps, present the final numerical answer on a new line, prefixed with "####".

**Example of expected output format:**

Question: [Your question here]
Answer: Let's think step by step.
[Step 1 explanation and calculation]
[Step 2 explanation and calculation]
...
[Final Answer explanation and calculation]
#### [Your final numerical answer]

---

Question: {question}
Answer: Let's think step by step.
---

--- Running Final Evaluation on the Held-Out Test Set ---
--- Running Final Test for a prompt on 150 questions ---

Final Accuracy on Test Set: 95.33%
```