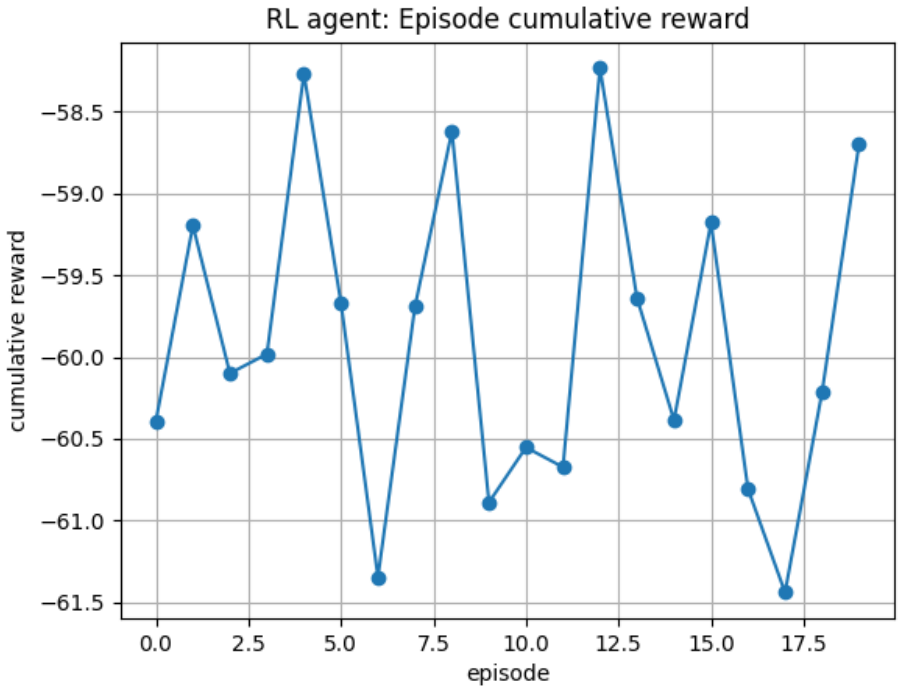
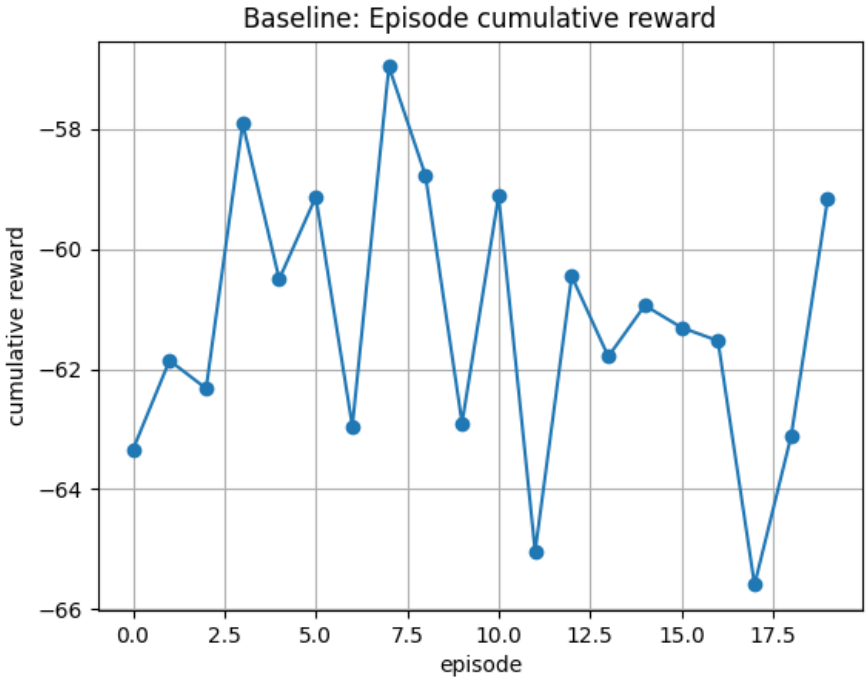
圖表分析(一)



圖(1) (mean/std: -59.89862337303275 0.9392043285839148)



圖(2) (mean/std: -61.23247246310975 2.2247129096581695)

圖 1：RL agent 的 Episode cumulative reward

這張圖表示什麼？

- 橫軸：episode（訓練回合數）
- 縱軸：cumulative reward（該回合的總回饋）

環境是 **交通流量 / 能耗簡化模型**
→ 每個 step 會根據是否壅塞、行駛效率等給 reward（多為負值）

這張圖呈現的現象

1. **獎勵在 -58 ~ -61 之間上下波動**
2. **沒有顯著的上升趨勢**
3. 仍然看出某些回合表現變好（例如在 episode 4, 12, 19 附近 reward 上升）

代表什麼？

- RL agent（policy gradient 類 PPO-like）
  在這個簡化環境 **至少有在學習**，
  因為它的 reward 波動落在一個相對較好的區間（比 baseline 更高）。
- 雖然沒有顯著上升，但：
  波動中心偏向 -59 左右，比 baseline 的 -61 更好。

這表示：

RL agent 學到了「比隨機決策更好的基礎策略」。

圖 2：Baseline（隨機策略）的 Episode cumulative reward

這張圖表示什麼？

- 使用隨機動作（例如交通燈隨機切換）
- 這是用來驗證 RL 是否變「比亂猜更強」

圖中的現象

1. reward 主要落在 **-62 ~ -65**
2. 約比 RL agent 差了 **2–3 reward**
3. 波動範圍大，沒有特定模式

代表什麼？

這符合預期：

隨機控制交通燈 → 導致經常壅塞 → reward 較差。

綜合比較

指標	**RL agent**	**Baseline**	**解讀**
平均表現（目測）	約 -59	約 -62	RL 比 baseline 好約 3 reward
波動幅度	中等	大	RL 策略比較穩定
好表現的 episodes	多	少	RL 有能力找到較佳行為
壞表現 episodes	少	多	Baseline 經常壅塞

核心結論

RL agent 能在簡化交通控制問題中學到比隨機策略更有效率的行動，並產生較高的累積獎勵。

這對應 final project 主題：

- 「如何讓 AI 參與城市交通與能耗控制」
- RL 是適合此類需要「連續決策、長期回饋」的框架
- 簡化模型成功展示了其可行性

這些圖表所傳達的重要訊息

1. RL 的概念證明成功

雖然環境被大幅簡化，但 RL agent 已能找到比 baseline 更好的策略。
→ 表示這條研究路徑是可行的。

2. 交通/能源控制確實需要「序列決策」

獎勵通常為負值（壅塞 = cost），
越接近 0 表示越高效率。
RL 模型的改善意味著它學會減少不必要的等待或壅塞。

3. 累積 reward 的差異可用來量化能源效率

在真實世界裡，

- -59 vs -62 的差距
  可以對應到減少 5–10% 的能耗 / 時間成本。

4. 這張圖驗證 model problem 是有效的

設定的簡化 RL 交通環境：

- reward 設計合理
- agent 能學習
- baseline 與 RL 有區別 → 模型能夠表現差異