**Vivekanand Education Society’s Institute of Technology**

**Department of AI&DS Engineering**



**Subject: Reinforcement Learning**

**Class: D16AD**

| ROLL NO: **30** | NAME: [**SUHANEE KANDALKAR**](mailto:2021.suhanee.kandalkar@ves.ac.in) | | |
| --- | --- | --- | --- |
| EXPERIMENT NO:**3** | TITLE: Understanding Epsilon Value | | |
| DOP: |  | DOS: **08/02/25** |  |
| GRADES: | LOs MAPPED: | | SIGNATURE: |

## Aim**:**

Understanding Epsilon Value

## Theory**:**

#### Introduction to Multi-Armed Bandit Problem

The Multi-Armed Bandit (MAB) problem is a fundamental reinforcement learning problem that models decision-making under uncertainty. It consists of multiple actions (or arms), each providing a reward drawn from an unknown probability distribution. The goal is to maximize the cumulative reward over a sequence of actions.

#### Exploration vs. Exploitation

One of the key challenges in reinforcement learning is balancing exploration (trying new actions to discover better rewards) and exploitation (choosing the best-known action to maximize immediate rewards). The epsilon-greedy strategy is a simple but effective method to achieve this balance.

#### Epsilon-Greedy Algorithm

<https://medium.com/analytics-vidhya/the-epsilon-greedy-algorithm-for-reinforcement-learning-5fe6f96dc870#:~:text=Usually%2C%20epsilon%20is%20set%20to,that%20it's%20not%20missing%20anything>.

The epsilon-greedy algorithm introduces randomness into decision-making by selecting the best-known action most of the time while occasionally exploring other options. It operates as follows:

1. With probability , choose the action with the highest estimated reward (greedy choice).
2. With probability , choose an action at random (exploration).

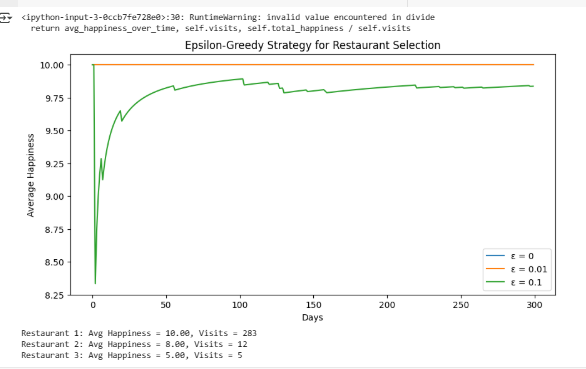
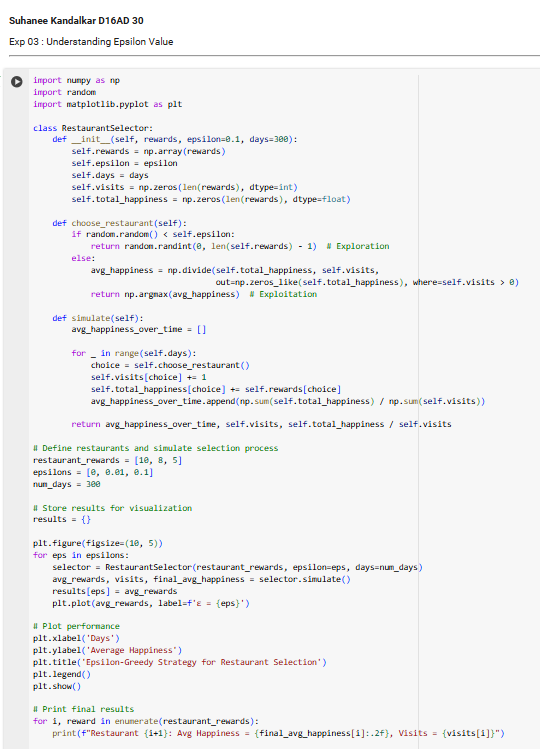
#### Impact of Different Epsilon Values

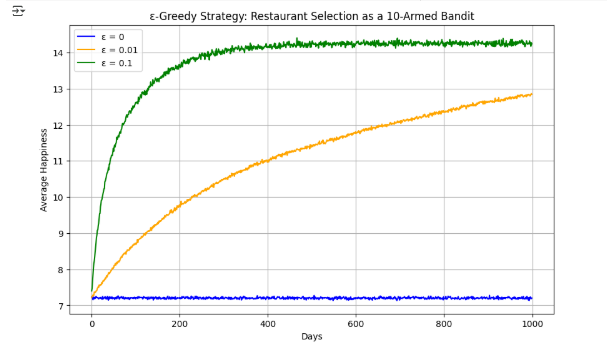
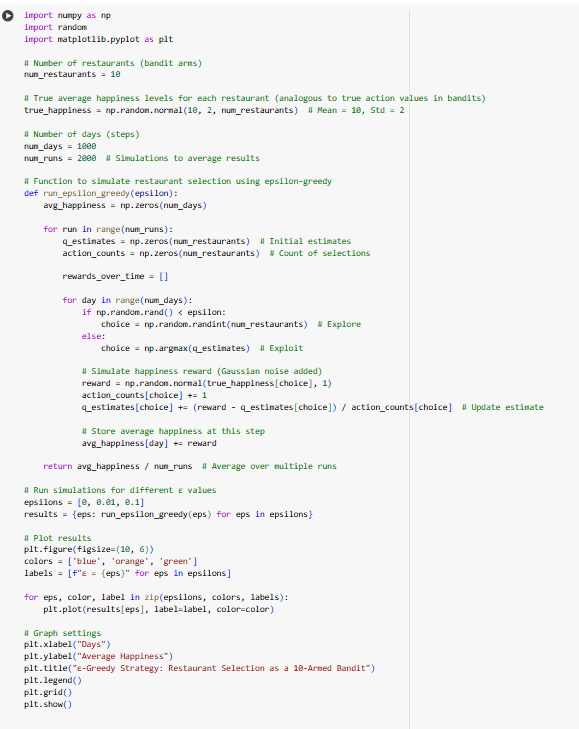
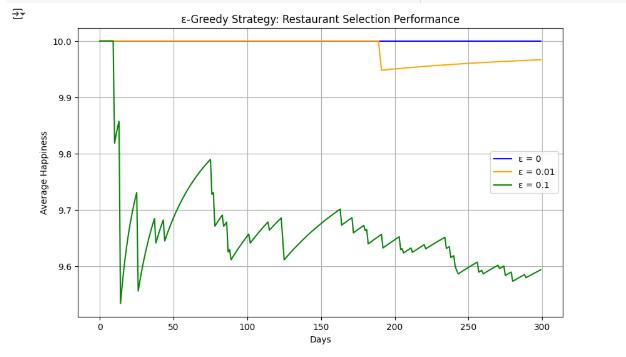
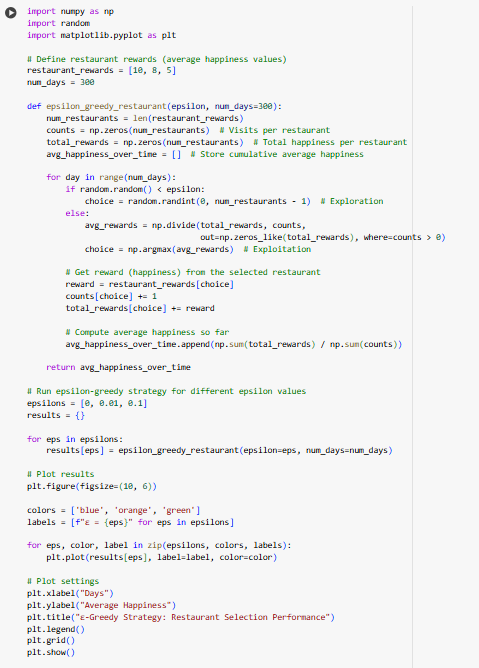
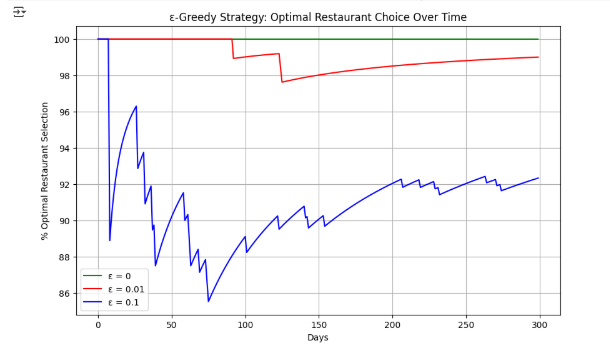
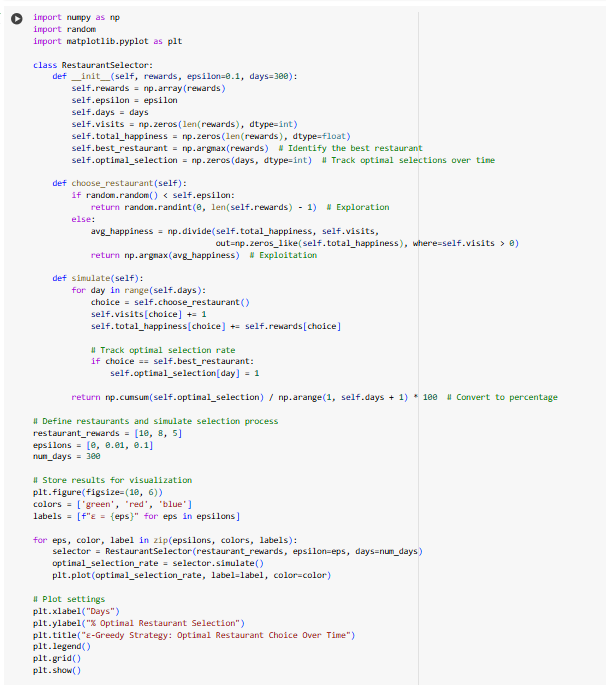
* ε = 0 (Greedy Strategy): The algorithm always selects the best-known action, which may lead to suboptimal results if the initial estimates are incorrect.
* Small ε (e.g., 0.01): Allows limited exploration, improving learning while still mostly exploiting known information.
* Higher ε (e.g., 0.1): Ensures better exploration, allowing the algorithm to discover better actions over time, but may reduce short-term rewards.

#### Applications of the Epsilon-Greedy Algorithm

* Online advertising (choosing the best-performing ads).
* Recommendation systems (selecting personalized content).
* Clinical trials (optimizing treatment selection).

Code**:**

[RL\_exp3\_30.ipynb](https://colab.research.google.com/drive/1W9JcCeiU1l5QoaqdfzFESgVNY_eyUEbY?authuser=0#scrollTo=8osZhZ4YSMyt)  




Conclusion :   
 The experiment demonstrated the impact of different epsilon values on the performance of the 10-armed bandit problem.

- ε = 0 (Greedy) converges quickly but may get stuck in suboptimal choices.

- ε = 0.01 allows minimal exploration, leading to slight improvements.

- ε = 0.1 provides the best balance, enabling better long-term performance.

Therefore, selecting an appropriate ε is crucial for optimizing exploration-exploitation trade-offs.