1 Question 2

a)

Here's an example with $t = 2$ and $W_2 = \dfrac{3}{4}W^*$. Consider $E = \{a, b, c, d\}$ and $S_1 = \{a, b\}, S_2 = \{c, d\}, S_3 = \{a, c\}$. Suppose that the weight of all elements in E is 1. Then we see in the first step, the algorithm may pick any of the three sets as the total weight of uncovered elements in the three sets are equal at 2. Suppose that it picks S_3. In the second step, the total weight of uncovered elements in both S_1 and S_2 are 1. Suppose the algorithm picks S_1. Thus, the total weight of our cover is 3. However, we see that the optimal cover is $S_1 \cup S_2$, with a weight of 4. Thus, we see that $W_2 = \dfrac{3}{4}W^*$.

Let's prove that for $t = 2$, $W_2 \leq \dfrac{1}{2}W^*$ cannot happen. Let $w(S_i|C)$ denote the weight of uncovered elements in a set S_i given some cover C - we call this the effective weight of a set. The greedy algorithm will first pick $T_1 = \arg\max_{S_i} w(S_i|\emptyset)$. For its second choice, the greedy algorithm would then pick $T_2 = \arg\max_{S_i} w(S_i|T_1)$. We know that since the max weight set is picked first, $w(T_1|\emptyset) \geq w(T_2|T_1) \geq 1$.

Let's consider the optimal sets picked: O_1 and O_2. We know that $w(O_1|\emptyset) + w(O_2|O_1) \geq w(T_1|\emptyset)$ because otherwise, the solution would not be optimal (i.e. pick $O_1 = T_1$ would be a better solution). We also know that $w(O_1|\emptyset) + w(O_2|O_1) \leq 2w(T_1|\emptyset)$ since $w(T_1|\emptyset)$ is the largest effective weight we can ever pick (as it is the largest weight set on first pick). Since effective weight of sets goes down after every pick (elements deleted), both the first and second pick are at most $w(T_1|\emptyset)$ effective weight each.

At minimum $w(T_2|T_1) = 1$, so $W_2 = w(T_1|\emptyset) + w(T_2|T_1) > \dfrac{1}{2}2w(T_1|\emptyset) \geq \dfrac{1}{2}(w(O_1|\emptyset) + w(O_2|O_1))$, and so $W_2 > \dfrac{1}{2}W^*$.

b)

Prove that $W_1 \geq \dfrac{W^*}{t}$.

Observe that $W_1 = \max_{S_i \in E} w(S_i|\emptyset)$ as the set picked first by the greedy algorithm has the largest effective weight possible (as elements are deleted for subsequent picks, which decreases the effective weight). This bound holds for each of the t sets picked in the optimal solution: $W^* \leq t \times W_1$. Rearrange the previous equation and we get $W_1 \geq \dfrac{W^*}{t}$.

Prove that $W_i - W_{i-1} \geq \dfrac{W^* - W_{i-1}}{t}$.

To prove the above claim, we will prove the equivalent statement $t(W_i - W_{i-1}) \geq$

$W^* - W_{i-1}$. Observe that in the ith iteration of the greedy algorithm, a set with effective weight $W_i - W_{i-1}$ is picked. Note that the set the greedy algorithm picks is the one with the largest effective weight (i.e. after "deleting" all the elements in the cover).

Let's now formulate an adjusted version of the problem. Suppose we remove all the elements covered by the first $i-1$ sets chosen by the greedy algorithm and try to solve the max weight set cover problem again. Let W'^* be the optimal solution to this problem. If we run the greedy algorithm from the very start, it is clear that the greedy algorithm would first pick a set with effective weight $W_i - W_{i-1}$ since it is the largest effective weight possible after deleting the covered elements (i.e. same situation mentioned above, but as first pick instead). In the optimal solution, we know that each of the t picks must have an effective weight less than or equal to $W_i - W_{i-1}$ as it is the largest possible. Then, it is clear that $t(W_i - W_{i-1}) \geq W'^*$.

Now, consider the optimal solution for the original problem, with total weight W^*. If we remove all the elements that were "deleted" from the optimal solution, then we get t sets that solve the adjusted version of the problem, which may no longer be optimal. In the best case, none of the deleted elements were in the optimal cover anyway, and $W'^* = W^*$, but in the worst case all the deleted elements were in the cover such that $W'^* \geq W^* - W_{i-1}$ (note that W'^* is the maximal solution to the adjusted version of the problem and W^* with all deleted elements removed is just *a solution* to the adjusted version of the problem). Thus, we arrive at the following: $t(W_i - W_{i-1}) \geq W^* - W_{i-1}$.

c)

We will prove by induction that $W_i \geq (1 - (1 - \frac{1}{t})^i)W^*$.

Let's prove the base case where $i = 1$, and so $W_1 \geq \frac{1}{t}W^*$. This is proven in part b).

Now, let's assume that for $i = k$, $W_k \geq (1 - (1 - \frac{1}{t})^k)W^*$. Prove for $i = k + 1$ that $W_{k+1} \geq (1 - (1 - \frac{1}{t})^{k+1})W^*$.

$$W_{k+1} - W_k \geq \frac{W^* - W_k}{t} \qquad\qquad \text{by our result in part b)}$$

$$W_{k+1} \geq \frac{W^* - W_k}{t} + \frac{tW_k}{t}$$

$$W_{k+1} \geq \frac{W^* - W_k + tW_k}{t}$$

$$W_{k+1} \geq \frac{W^*}{t} + \frac{(t-1)W_k}{t}$$

$$W_{k+1} \geq \frac{W^*}{t} + (1 - \frac{1}{t})W_k$$

$$W_{k+1} \geq \frac{W^*}{t} + (1 - \frac{1}{t})(1 - (1 - \frac{1}{t})^k)W^* \qquad \text{by our inductive hypothesis}$$

$$W_{k+1} \geq \frac{W^*}{t} + (-\frac{1}{t} + 1 - (1 - \frac{1}{t})^{k+1})W^*$$

$$W_{k+1} \geq \frac{W^*}{t} - \frac{W^*}{t} + (1 - (1 - \frac{1}{t})^{k+1})W^*$$

$$W_{k+1} \geq (1 - (1 - \frac{1}{t})^{k+1})W^*$$

Thus, we see that if $W_k \geq (1 - (1 - \frac{1}{t})^k)W^*$ then $W_{k+1} \geq (1 - (1 - \frac{1}{t})^{k+1})W^*$. And so, by our base case where $i = 1$ and the validlity of the inductive step, $W_i \geq (1 - (1 - \frac{1}{t})^i)W^*$ is true for all i.

Q.E.D.