Week 3:

New Topics:

- Creating calculated columns
- Visual EDA: bar plots, histograms, box plots, swarm plots, scatterplots

Coding Task:

1. Using the `tn_income` DataFrame, create a bar plot showing the total number of returns per income bucket for the state of Tennessee. You can get the total numbers for the state by filtering to rows where the `county` variable is "Tennessee".
2. Create a calculated column in the `income_county_agg` DataFrame which gives the approximate average household income for each county. Calculate this as 1000*`total_inc_amt` / `return_count`.
3. Create a histogram showing the distribution of average incomes across all counties in Tennessee. Be sure to remove the Tennessee row prior to creating this histogram.
4. For both `tn_cancer_costs` and `tn_ha_costs`, create boxplots and swarmplots comparing the distribution of analysis_value for urban counties vs. rural counties. What do you notice?