Data munging and analysis exercise

Background:
You have two files from a USAID country. One contains quarterly data for all of USAIDs programs going back to FY17Q1[i] through FY18Q4. The other is daily or weekly off-cycle data reported to USAID/HQ for a subset of indicators for October and November 2018. Both datasets have been de-identified and the country/district combinations replaced with names of planets and their moons.

1. **Quarterly** filename: MER_Structured_TRAINING_Dataset_PSNU_IM_FY17-18_20181221_v2_
2. **Daily/weekly** filename: Daily Programme Monitoring_4.5.2019

Both files are the records of PEPFAR indicators. The two files have a common geographic hierarchy and common indicators although the quarterly file has much more detail than the daily/weekly file. The daily dataset is at the facility level, the quarterly dataset is at the district level. District = PSNU[1]. In the organization/geographic hierarchy, facilities sit under or live inside of districts.

USAID/HQ receives the quarterly file 50 days after the end of the quarter, however we receive a version of the daily file every week.

Goal & Objectives:
We want to be able to look at the weekly file alongside the data from the quarterly file.
PEPFAR routine health data is disaggregated by age and sex (ie, not just the number of tests performed (HTS_TST) but the number of test performed on women 20-24 for example). In the quarterly file, each observation is a disaggrate of the indicator, along with the geographic hierarchy. For this exercise, only use the observations from the quarterly file that have "Total Numerator"[2] in the "standardizedDisaggregate" column.

A key of the column headers in both files:

Daily	quarterly
Indicator	indicator
PSNU	PSNU
Siyenza Sites	NA
TIER	NA
Facility	NA

Indicator definitions
HTS_TST_POS = Number of HIV tests where the result was positive
TX_NEW = Number of new clients initiated on Antiretroviral treatment

---

[1] PSNU = Priority Sub National Unit. In this instance, PSNU = District
[2] The Total Numerator is typically created in DA

Objective 1: Munging

We want to align the daily file with the quarterly file. The geographic hierarchy differs between the two, therefore the daily data will have to be aggregated up to their districts in order to be compared to the quarterly file, by indicator.

Objective 2: Analysis

We would like to look at trends over time by PSNU of the two indicators common to both files (note that not all the data needs to be displayed or reported back). We would also like to look at FY18 performance (defined as FY18apr / FY18 Targets) by PSNU alongside their results for the first two months of FY19. The visualization should be clean, readable, and crafted in such a way to be accessible to both data saavy and non-data savvy users.

---

[i] USAID uses the Federal Fiscal Year cycle.