St. Paul city salary data exercise

Use the Excel file called “stpaulsalaries.txt”

The data contains one salary record for each person paid by the city of St. Paul in the 2009 calendar year. It shows their total wages (“total\_wages”), plus fields that break that figure down into overtime wages (“OT\_pay”) and other pay (“Other\_Pay”-- which could include things like bonuses, vacation-sick payout, etc). In addition to the dollar amount paid out in overtime, it also shows the number of OT hours (“OT\_hours”) the person worked (which can be two very different because some people have higher or lower pay rates that would influence the dollar amount yielded from an hour of OT). **On separate sheet, see record layout and codes.**

**Step 1:** The data file is a tab-delimited text file. So our first step is to import the data into Excel.

**Step 2:** Let’s start by making a summary table to see what patterns we see at the department level.

Make a PivotTable that has “Dept” as the row and the following fields in the Data section: TotWages (sum), Empid (count), OT\_Pay (sum), OT\_Hours (sum). Make sure to double-click on the items in the data section to change them to “sum” or “count” as identified here.

Then copy and paste just the contents of the PivotTable (don’t take the headers or the Grand total) and paste it into a new worksheet. You should have a new table that shows one record for each department and the total wages paid, the number of employees paid, the total overtime paid out and the total number of overtime hours.

**Step 3: Using the new table…**

1. Sort the table to find out which department paid out the most in total wages?
2. In a new column, calculate the average salary for each department (divide total wages by number of employees). Which department had the highest average salary and what was the dollar amount?
3. In a new column, calculate the percentage that overtime pay makes up of the total wages for each department. What department has the highest rate of OT pay and what was percentage?
4. In a new column, calculate the average overtime hours per person (hint: you have a field with the number of employees in each department). Which department had the highest rate and what was the rate?
5. In a new row below your table, tally up the total number of OT hours clocked by city employees. What’s that number (rounded)?

**Extra credit:**

* Now that you know the total number of OT hours clocked by all city employees, how many overtime hours does that amount to on a per-week basis? (hint: there are 52 weeks in a year)
* Use the RANK function, to rank the departments on total wages. Which department paid out the most? (see below for directions on how to use RANK)
* Then RANK the departments on Overtime Per Person. Which department had the highest rate?

**Step 4: Go back to the original data table that we started with:**

1. Who was the highest paid person in 2009 and what was the dollar amount?
2. In a new column, calculate what percentage OT was of each person’s total wages. (hint: you’re doing a “percent of total” calculation for each record). Then sort the table ….Who has the highest percentage? Do you see anything else that’s interesting and/or newsworthy here? What questions does this raise?
3. Who got paid the most overtime in total and how much? (hint: you don’t need to do any calculations)

**More extra credit:**

* For those with OT hours… in a new column, calculate how many OT hours they clocked, on average, each week (hint: 52 weeks per year). Who had the most and how much?
* Which bargaining unit has the highest average pay in 2009 and how much? (hint: this requires you to summarize your data)
* Can we figure out which job titles have more overtime than others? How would you do that?

**Rank:**

Instead of simply sorting your results to put them in “order”, this is a more sophisticated way to rank your records and to account for ties.

=RANK(This Number, $Start Range$:$End Range$, Order)

This Number should be the cell where your data starts.

Start Range should be the cell where your data starts. Anchor with dollar signs. End Range should be the last cell of your data. Anchor with dollar signs.

Order is either a 1 (smallest value will get assigned #1) or a 0 (largest value will get assigned #1).

Example: =RANK(B2,$B$2:$B$100,1)

Created by:   
@MaryJoWebster  
Updated: March 2014