

ACL Guru



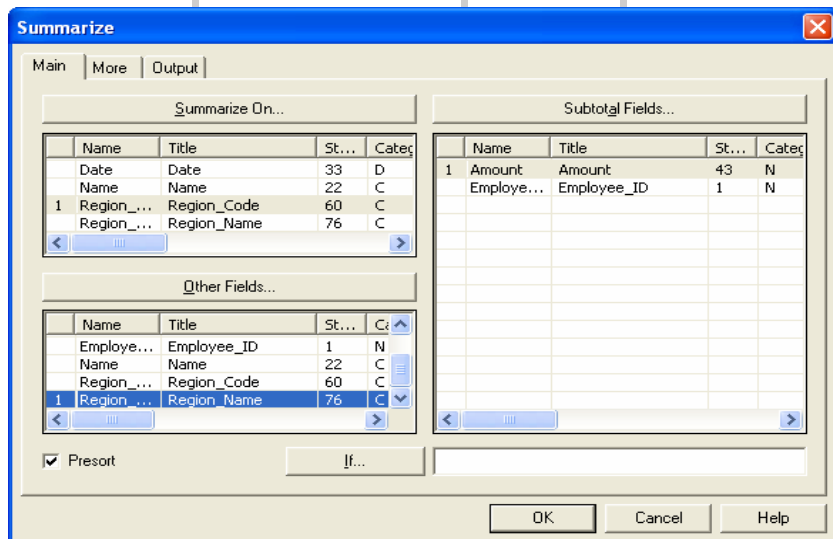
Advanced Use of Summarize and Other Fields

Quick tip this time around!! Summarize provides for us two very basic database abilities. Summarize allows us to group data together with the option of totaling by category. In addition, we receive, by default, a distinct listing of the summarized field. Those two simple concepts, group and distinct are crucial to data analytics. In ACL, we have the option of selecting other fields. What does this option do?

The 'other fields' option allows us to pull through the first instance of another field when summarizing. For example, let's look at the following transactions:

	Employee ID	Name	Date	Amount	Region Code	Region Name
1	1000001	Dave Brand	03/22/2008	100.00	01	North
2	1000002	Steve Davis	03/14/2008	125.00	03	East
3	1000003	Sally Green	03/22/2008	90.00	01	North
4	1000004	Alice Troop	03/14/2008	210.00	04	West
5	1000006	John Lucas	03/22/2008	175.00	02	South
6	1000007	Ann Stanzi	03/14/2008	215.00	02	South
7	1000009	Abdul Gharb	03/02/2008	160.00	03	East
8	1000010	Tom Anzac	03/02/2008	155.00	04	West

If I wanted to know how much was spent by each region, I can summarize by region code and subtotal the count. If I add the region as an 'other field', then I will get the region name as well.

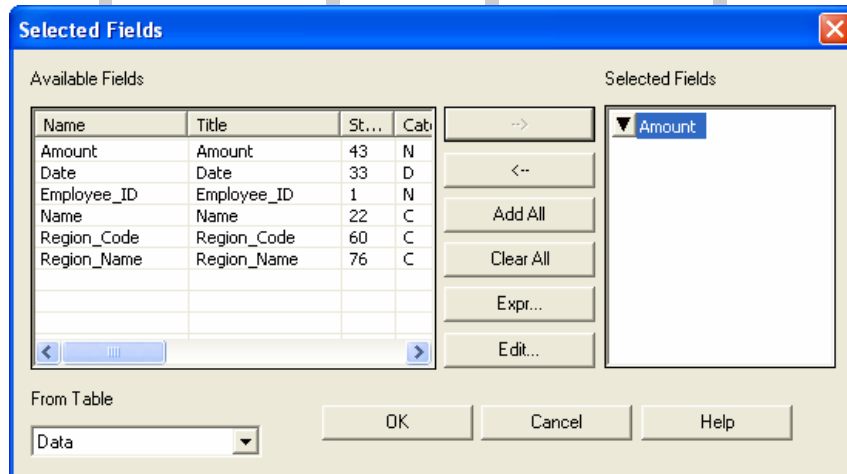


The output is the following table. Notice that the region name has been pulled through with the region code.

Region Code	Amount	COUNT	Region Name
01	190.00	2	North
02	390.00	2	South
03	285.00	2	East
04	365.00	2	West

<< End of File >>

Fairly simple but let's think about what it allows me to actually do. The region name was pulled from the first instance of its presence in the table. So, if I reorder the table, the other field could pull from a different position. This can be very useful. Let's say I want to pull the highest transaction from each day. First, I sort the table by amount in descending order to create a table in a different physical order as follows.

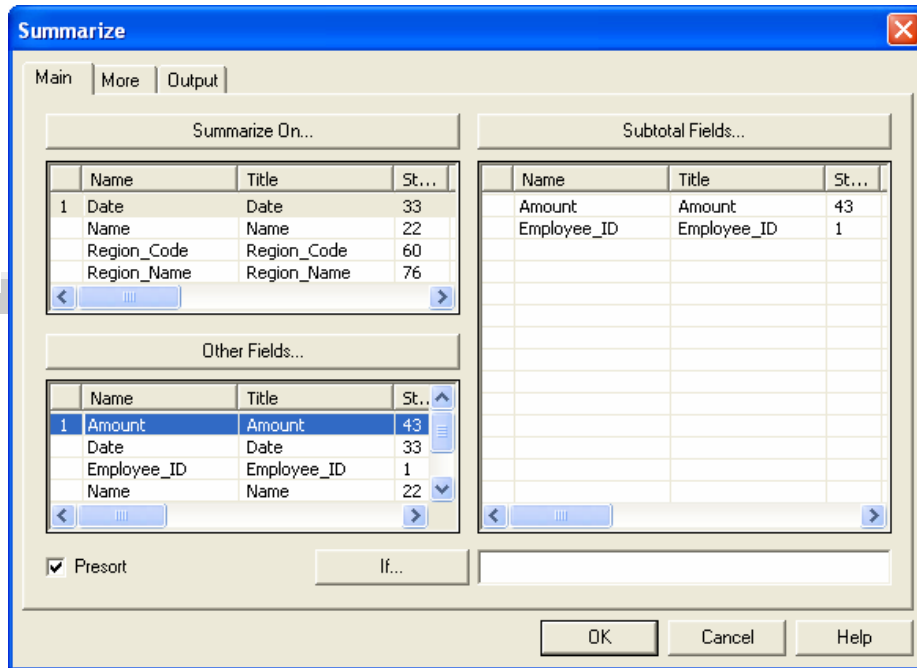


Yielding the following table:

Employee ID	Name	Date	Amount	Region Code	Region Name
1000007	Ann Stanzi	03/14/2008	215.00	02	South
1000004	Alice Troop	03/14/2008	210.00	04	West
1000006	John Lucas	03/22/2008	175.00	02	South
1000009	Abdul Gharb	03/02/2008	160.00	03	East
1000010	Tom Anzac	03/02/2008	155.00	04	West
1000002	Steve Davis	03/14/2008	125.00	03	East
1000001	Dave Brand	03/22/2008	100.00	01	North
1000003	Sally Green	03/22/2008	90.00	01	North

<< End of File >>

Next, I summarize the table by Date and pull the Amount as the other field with no subtotal. This will group the data by Date and pull the first occurrence of the other field, in this case the Amount, which will be the highest Amount because the table is in descending order. The result is the highest Amount by Date.



And the data table:

Date	COUNT	Amount
03/02/2008	2	160.00
03/14/2008	3	215.00
03/22/2008	3	175.00

<< End of File >>

As you can see, the highest transaction for each day was isolated.

So, we're done! The concept here is to use the physical order of the data in conjunction with the mechanical operation of the summarize command to automate our analysis. Data analytics at its finest!!

Randall Shane is the president of Base XVI, LLC and a Certified ACL Trainer. Base XVI performs data analytics including data mining, database development, computer forensics and electronic discovery for corporations, government agencies and law firms. Dr. Shane possesses a PhD in Computer Science with both civil and criminal testimonial experience. Base XVI and its staff are located in North Florida and serve clients worldwide. And no, Dr. Shane does not have a beard!!