Assignment:

1- Downloaded and extract csv files and ReadMe file

2- Study ReadMe document for explanation and abbreviations and correlate data with supplied csv file.

The below method is suitable for small number of data records. For large data records (~>1M), I will be using the Microsoft Excel Data Model for data analysis and visualization.

3- Refer to supplied READNE on Storm Data document to identify the headers for wanted criteria such for County/Zone (CZ_NAME, EVENT_TYPE, STATE, YEAR, MONTH_NAME, BEGIN_DATE_TIME, END_DATE_TIME)

4- Start New Excel workbook with 3 worksheets: Filter_Criteria, StormEvents_2000, StormEvents_2015

5- For initial filtering, create filter-criteria ranges with the same header title (CZ_NAME) and specify the filtering criteria:

e.g.: Does Not Begin With A, B, C (using wildcards), in same-row format (AND operator)

CZ_NAME

<>A* <>B* <>C*

You may also add additional filters to exclude some unused columns (e.g. FIPS, LAT, LON,...) to reduce resource demands.

6- Open the wanted downloaded csv file for StormData 2015.

7- Navigate back to activate the corresponding new blank sheet (e.g. StormEvents_2015)

8- Within the active blank sheet, apply Advanced Filter (with copy to another location) using the Data List from the original csv sheet (using ctrl+A to select range) and the defined criteria to exclude County/Perish, Zone or Marine Name that begins with the letters A, B and C. Place in first cell of the new sheet.

The original csv file can be closed after this operation to free system resources.

9- Convert copied and filtered data range in worksheet into Table.

10- Create PivotTable from the data range in the filtered worksheet.

Q1: Find the month in 2015 where the State of Washington had the largest number of storm events. How many days of storm-free weather occurred in that month?

Answer:

* Adjust Pivot Table Fields as below:

    Rows: MONTH_NAME

    Columns: BEGIN_DAY

    Filters: EVENT_TYPE

    Slicer: STATE

    Values: Function of Count EVENT_TYPE

* Create number of storm-free weather days using additional column adjacent to the pivot table by subtracting the count of days with events from total number of days for each month:

=DAY(EOMONTH(DATE(year,month,1),0))-COUNTA(day_range)

Where 'day_range' are the range of cells containing the count of EVENT_TYPE for each day of the month

The above equation can be adjusted to make the year and month automatically obtained for versatility:

For example:

month = ROW()-ROW(1st_month) +1, where 1st_month is a constant cell containing the first month "January"

* Slice data by Washington STATE using Slicer. Apply any other filters (e.g. EVENT_TYPE contains "storm" only)

For this example, I'm including two results:

a) All events including different storms and other weather phenomena: December (140 events, 12 event-free days)

b) Events contains the word "storm": August (8 storms, 27 storm-free days)

Q2. How many storms impacting trees happened between 8PM EST and 8AM EST in 2000?

1- Open the related downloaded csv file (StormData 2000)

2- Navigate back to activate the corresponding new blank sheet (e.g. StormEvents_2000)

3- Within the active blank sheet, apply Advanced Filter (with copy to another location) using the Data List from the original csv sheet (using ctrl+A to select range) and the defined criteria to exclude County/Perish, Zone or Marine Name that begins with the letters A, B and C. Place in first cell of the new sheet.

The original csv file can be closed after this operation to free system resources.

4- Convert copied and filtered data range in worksheet into Table.

5- Use "Find" to search for the keyword " tree" (with preceding space). Results show that details on storms impacting trees can be found under the EPISODE_NARRATIVE and EVENT_NARRATIVE fields.

Create new column to indicate if any of the two fields include the word " tree":

Field Name: Tree_Related

=NOT(OR(ISERROR(SEARCH(" tree", [@[EPISODE_NARRATIVE]])), ISERROR(SEARCH(" tree", [@[EVENT_NARRATIVE]]))))

6- Assuming that no time-zone shifting is required, and areas within the EST time zone is needed for this analysis:

Create new column to abstract the time from the BEGIN_DATE_TIME and END_DATE_TIME fields:

Field Name: Time_8

=OR(TIME(HOUR([@[BEGIN_DATE_TIME]]), MINUTE([@[BEGIN_DATE_TIME]]), SECOND([@[BEGIN_DATE_TIME]])) >= TIMEVALUE("8:00 PM"), TIME(HOUR([@[BEGIN_DATE_TIME]]), MINUTE([@[BEGIN_DATE_TIME]]), SECOND([@[BEGIN_DATE_TIME]])) < TIMEVALUE("8:00 AM"))

7- Create PivotTable from the data range in the filtered worksheet.

Adjust Pivot Table Fields as below:

Filters: Tree_Related, Time_8, CZ_TIMEZONE

Values: Count of EVENTS_TYPES

Answer: 31 storms/events impacting trees happened between 8PM EST and 8AM EST in 2000? (EST locations only. No zone-shift)

Q3. In which year (2000 or 2015) did storms have a higher monetary impact within the boundaries of the 13 original colonies?

Add another column for total monetary damage for each table

Field: Money_Damage

=IF(ISERROR(SEARCH("K", [@[DAMAGE_PROPERTY]])), VALUE([@[DAMAGE_PROPERTY]]), VALUE(LEFT([@[DAMAGE_PROPERTY]],SEARCH("K", [@[DAMAGE_PROPERTY]])-1)) * 1000) + IF(ISERROR(SEARCH("K", [@[DAMAGE_CROPS]])), VALUE([@[DAMAGE_CROPS]]), VALUE(LEFT([@[DAMAGE_CROPS]],SEARCH("K", [@[DAMAGE_CROPS]])-1)) * 1000)

Add new Pivot Tables or adjust/refresh current tables to the following Fields:

Slicer: STATE

Values: Money_Damage

Answer:

Monetary impact for year 2000: $31,662

Monetary impact for year 2015: $8,408

Year 2000 has higher monetary impact with the boundaries of the 13 original colonies and DC (excluding Atlantic N, Atlantic S, Lake Erie and Lake Ontario regions)