refocus

Home assignment for the lesson "Data Cleaning and Manipulation"

Mentor-Review

This task needs to be submitted for review.

Watch the video lessons to complete this task. You may ask your mentor's help via the chat if you have any questions.

To ensure a thorough understanding of the topic, we suggest you repeat everything after the speaker before moving on to the next tasks.

Download this dataset and proceed with the Data Cleaning tasks below.

Learning outcomes

You will be able to:

- import datasets into Power BI
- clean provided data

What you need to do

We will cover Lesson One (Importing data in Power BI) and Lesson Two (Data Cleaning Process) for this home assignment.

Task 1. Data Importing

- Import data from a single folder.
- Update file tab names to *All Users*.

Task 2. Data Cleaning

- Delete the *source.name* column.
- Split columns by the delimiter.
- Promote the first rows as the column headers.
  Note: Do not remove the change type after this step.
- Select All Columns (*CTRL+A*) and remove rows with errors.
  Reminder: You should have no errors in your data at this moment. If you apply the changes and see errors in the queries, you miss selecting all columns. To fix this, go to the query workspace and make sure you select all columns. Then apply the changes again.
- Remove the following columns:

account_type (small letters)	day_10
bannertag_1	gid
year	global_id
quarter	not found affiliates
month	onetimers w28
day	playerid_11
close_reason_2	year_12
column	quarter_13
cross brand type (demo)	month_14
year_3	day_15
quarter_4	reigstration_ip2
month_5	rewardplaintype
day_6	status_id
year_7	top country
quarter_8	unique
month_9	vip gm

- Update the data type of the following columns:
  bannertag as text data type
  close reason as text data type
  receive_email as text data type
- Replace the values of the following columns:

Account type
- Replace *Acq Cross with Acquisition*
- Replace *Crm Cross with CRM*
- Replace *Crm .Fsub with CRM*
- Replace *Existing Ftd with Existing*
- Replace *New Ftd with Ftd*

Close_reason
- Replace *Null Values with Not Applicable*
- Replace *Blank Values with Not Applicable*
- Replace *0 with Quit*
- Replace *1 with Not Happy With The Site*
- Replace *2 with Poor Customer Service*
- Replace *3 with Bad Campaigns*
- Replace *4 with No Activity For More Than 60 Days*
- Replace *5 with Game Options Limited*
- Replace *6 with Poor Payment Processing*
- Replace *7 with No Activity For More Than 90 Days*
- Replace *8 with Fraudulent Account*
- Replace *9 with Bonus Abusers*
- Replace *10 with Multiple IPs*

Country
- Replace *Blank Values with No Data Provided*

Crosstags
- Replace *Blank Values with Organic*

Deals
- Replace *Blank Values with Normal Deal*

Fsub Traffic
- Replace *506 with 506 – CRM*

Last Login Country
- Replace *Blank Values with No Data Available*

Link
- Replace *Blank Values with No Link Available*

Market
- Replace *Other with Others*

Receive_email
- Replace *0 with No*
- Replace *1 with Yes*
- Replace *Null Values with Pending*

Referer
- Replace *Blank Values with Organic*

Registration_IP
- Replace *Blank Values with No Data Available*
- Replace *Null Values with No Data Available*

Registration_time
- Replace *Null Values with 01/01/2010 12:00 Am*

Self_close_account
- Replace *Null Values with Not Applicable*
- Replace *Blank Values with Not Applicable*
- Replace *1 with Yes*

Site_ID
- Replace *Null Values with 161229585*

Status
- Replace *Blank Values with Pending*
- Replace *Null Values with Pending*

Status_Comment
- Replace *Blank Values with No Data Provided*
- Replace *Null Values with No Data Provided*

Test Player
- Replace *Blank Values with Test*
- Replace *Null Values with Test*

Traffic Source Type
- Replace *Blank Values with 198.162.234.166*
- Replace *Null Values with 198.162.234.166*
- Replace *Internal Tags with 41.146.210.186*
- Replace *Paid Tags with 37.43.250.13*

Unique_Global_ID
- Replace *Blank Values with No Data Available*
- Replace *Null Values with No Data Available*

Unsubscribed
- Replace *Blank Values with No Data Available*
- Replace *Null Values with No Data Available*

Wheel of Luck Players
- Replace *Blank Values with No Data Available*
- Replace *Null Values with No Data Available*
- Rename *Column Name with Wheel Of Luck Players*

- Rearrange Columns
  - *Registration Time* should be the first column.
  - *Registration IP* should be after *Registration Time*.
  - *Playerid* should be after the *Registration IP*.
  - *Username* should be after *Playerid*.
  - *Status* should be after *Account Type*.
  - *Status comment* should be after *Status*.
  - *Close Reason* should be after *Status comment*.
  - *Self Close Account* should be after *Close Reason*.
  - *Test Player* should be after *Self Close Account*.
  - *Receive Email* should be after *Test Player*.
  - *Unsubscribed* should be after *Receive Email*.
  - *Last Login Country* should be after *Unsubscribed*.
  - *Country* should be after *Last Login Country*.
  - *City* should be after the *Country*.
  - *Conversion RND* should be the last column.
  - *New Jackpot Players* should be after *FSUB Traffic*.
  - *Old Jackpot Players* should be after *New Jackpot Players*.
  - *Wheel of Luck Players* should be after *Jackpot Winners*.
  - *Link* should be after *Referer*.
  - *Site Name* should be after *Wheel of Luck Players*.
  - *Unique Global ID* should be after *Market*.
- Replace the values of the following columns:

Site Name
- If *Site_ID* is 62 then the result is *All Spin Win*.
- If *Site_ID* is 73 then the result is *Go Wild*.
- If *Site_ID* is 92451887 then the result is *Big Bonanza*.
- If *Site_ID* is 159136232 then the result is *Spin Million*.
- If *Site_ID* is 160570919 then the result is *Pokie Spins*.
- If *Site_ID* is 161229585 then the result is *Europoker*.

Test Player
- If *Test Player* is OK then the result is *Test Player*.
- Else *Non-Test Player*.

- Created Group By

By Account Type
- Update the *Table name* to *Account Type*.
- Name the *Column Name* to *Number of Players*.

By Close Reason
- Update the *Table name* to *Close Reason*.
- Name the *Column Name* with *Number of Players*.
- Remove *Not Applicable Row*.

By Country
- Update the *Table name* to *Country*.
- Name the *Column Name* with *Number of Players*.

By Market
- Update the *Table name* to *Market*.
- Name the *Column Name* with *Number of Players*.

By New Jackpot Players
- Update the *Table name* to *New Jackpot Players*.
- Name the *Column Name* with *Number of Players*.

By Old Jackpot Players
- Update the *Table name* to *Old Jackpot Players*.
- Name the *Column Name* with *Number of Players*.

By Status
- Update the *Table name* to *Status*.
- Name the *Column Name* with *Number of Players*.

By Wheel Of Luck Players
- Update the *Table name* to *Wheel of Luck Players*.
- Name the *Column Name* with *Number of Players*.
- Remove *No Data Available Row*.

By Site Name
- Update the *Table name* to *Site Name*.
- Name the *Column Name* with *Number of Players*.

By Test Players
- Update the *Table name* to *Test Players*.
- Name the *Column Name* with *Number of Players*.

Tips and recommendations

- Ensure your tab names match each file when importing multiple files to avoid errors.
- Update each step name to know which step you have applied.
- Always save your report by clicking the save icon on the topmost part in the quick options or by simply pressing *Ctrl+S*.
- If you're unsure of how to import and clean data, watch the demonstration again, and follow the instructions.

Completion Criteria

You have one Users Table and Ten Sub Tables.

This task needs to be submitted for review.

Watch the video lessons to complete this task. You may ask your mentor's help via chat if you have any questions.