**Turning event logs into business metrics**

Congratulations, you’ve been hired as a junior analyst at an e-commerce company!!! After searching for months, the company has finally found the right candidate to analyze their raw transaction logs and they want you to get started immediately.

Open the Business Analytics Project Google spreadsheet at the top of this lesson and explore the data in the “raw\_user\_activity” tab. Each row represents an activity, or event, by a user on the company’s website. Each time a user views a product page, opens their shopping cart, or completes a purchase, the event is captured in the activity logs.

The dataset has the following columns:

* user\_id: unique customer IDs
* event\_type: the type of activity by the user
* category\_code: category of the product being viewed or purchased
* brand: company that makes the product
* price: price of the product, in USD
* event\_date: date of the user activity, in YYYY-MM-DD format

**Part 1: Build a conversion funnel**

The executive team is interested in understanding how well the website is converting product page views into purchases. Your first job is to create a conversion funnel to better understand how users interact with the website.

* Using data from the “raw\_user\_activity” sheet, create the funnel in a pivot table as a new sheet called “conversion\_funnel”
* Make sure you are counting the unique users for each stage of the funnel (Hint: the funnel should have 3 stages)
* Use formulas to create two new columns in your pivot table: total conversion rates and conversion rates to the next step (Hint: reference the Overview of Funnels lesson)

**Part 2: Prepare data for cohort analysis**

The company wants you to build acquisition cohorts based on the month of a user’s first purchase, and they want you to track cohort metrics month by month.

The largest part of this project is data preparation for cohort analysis. Get ready to practice some advanced spreadsheet techniques!

Filter purchases

The “raw\_user\_activity” sheet contains events for product page views and the opening of shopping carts, but your cohort analysis should focus only on purchases.

* Create a new blank sheet tab called “purchase\_activity”
* Using the filters in the “raw\_user\_activity” sheet, select only event types that are purchases
* After applying the filter, copy the entire sheet and paste the data into your “purchase\_activity” sheet
* Double check that you only have data for purchase events in the new sheet (Hint: there should be 4,845 rows in the new sheet, including column headers)
* Don’t forget to reset the filter in the “raw\_user\_activity” sheet after you’re done copying the purchase data

Calculate first purchase dates

Now that purchase activity has been isolated in its own table, you can calculate the first purchase date for each user who made a purchase. These dates will eventually be used to assign users to cohorts.

* Using the “purchase\_activity” sheet data, insert a pivot table as a new sheet called “first\_purchase”
* Configure the settings of the pivot table to calculate the minimum event\_date for each user

Next, you need transfer the first purchase dates to a new column in the purchase activity data.

* Enter first\_purchase\_date as a column header in cell G1 of the “purchase\_activity” sheet
* In cell G2, write a formula that uses the VLOOKUP() function to find the date from the “first\_purchase” sheet that corresponds to the user ID in cell A2
* Make sure the ranges of your formula in G2 are configured correctly and then copy the formula to every row in column G; you can browse the result to check that none of the first\_purchase\_date values are greater than the purchase\_date values

Set up monthly data to build and track cohorts

Although you have both the event\_date and the first\_purchase\_date in the “purchase\_activity” tab, you need to group the users and transactions by month for the cohort analysis. Create three new columns in the “purchase\_activity” sheet to help build the cohorts: event\_month, first\_purchase\_month, and cohort\_age.

* Use the TEXT() function to create event\_month in column H and first\_purchase\_month in column I (Hint: the dataset contains more than one year, so use the YYYY-MM format for your months)
* Use the DATEDIF() function to create cohort\_age in column J as the number of months between the first purchase and the event (Hint: the cohort ages should range from 0 to 4 months)

**Part 3: Calculate retention rates**

The last steps of the analysis are to aggregate the purchase data into cohorts and then calculate retention rates for each cohort month by month.

Group data into cohorts

* Using the data from the “purchase\_activity” sheet, insert another pivot table as a new sheet called “cohort\_analysis”
* Configure your pivot table so that each represents one cohort, which are based on the month in which customers made their first purchase (Hint: there should be 6 cohorts)
* The pivot table should also have the count of unique users for each cohort\_age in the columns of the pivot table

Calculate overall retention rates

* Create a new blank sheet called “retention\_rates”; you can use this sheet to layout a table that looks similar to the “cohort\_analysis” sheet
* Add row labels in cells A3 to A7 for each cohort in chronological order (these should match the row labels in the “cohort\_analysis” sheet)
* Add column labels in cells B2 to E2 that represent the cohort ages from 1 to 4 months (you can’t calculate retention rates for a cohort’s first month, so you don’t need to include age 0)
* In cell B3, write a formula that calculates the retention rate for each cohort at each cohort age in the table you created, based on the starting cohort sizes (Hint: if you utilize a fixed column reference in your formula, you can copy the same formula in every cell of your table)

**Part 4 - Organize and document your spreadsheet**

You’ve completed the calculations for your analysis, but you want to make sure your first project looks polished and professional before submitting it to the executive team. Apply some best practices that you learned in the Advanced Spreadsheets course.

* Fill in the results synopsis and analysis descriptions in the “Executive Summary” sheet:
  + Results — Briefly summarize the results and conclusions from your analysis:
    - The “conversion\_funnel” sheet
    - The “retention\_rates” sheet
  + Analysis — Explain important properties about the data and any assumptions you made in your analysis:
    - Describe the raw data (e.g. timespan, event types, how it was collected, what parts of it you actually used, etc.)
    - Explain what data you used to calculate the conversion rates in your funnel (e.g. was it user counts, number of transactions, or something else?)
    - Explain the decisions you made in your cohort analysis of retention rates (e.g. how did you form cohorts, over what time were you tracking them, and how did you calculate retention rates?)
* Reorder the sheets tabs so that the “Table of Contents” and “Executive Summary” sheets come first, followed by the sheets with you analytical results, then the sheets used for calculations, and finally the raw data sheets last. We’ve created a legend to help you organize the sheets. Fill in the “Table of Contents” table with the organized sheet order and a brief one-sentence description of each section.
* Format your spreadsheets for readability (e.g., format numbers and dates, add borders to tables, utilize bold fonts for column headers, freeze rows, highlight cells with calculations)
* Use your best judgment to make a spreadsheet that you’d like to receive if you were an executive.