yipitDATA

SpendHound Data Transformation Specialist

Take-Home Exercise

Please be sure to read the instructions carefully.

SpendHound is quickly growing, and we are scaling our internal operations to support new customer growth and onboarding. We are looking for a Data Transformation Specialist to join our SpendHound team and manage a large portion of the data cleaning and quality assurance processes that power the data and insights we deliver to our customers.

The purpose of this assignment is to gauge your ability to clean and analyze datasets and express your findings clearly and concisely.

Background

A core part of our SpendHound product is ingesting financial data from our customers, cleaning and analyzing their data, and sharing a dashboard that shows an accurate view of all of the software applications their company uses.

Each time we sign a new customer, we run a comprehensive data cleaning and analysis process for their financial transactions data to (1) correctly identify all their software application transactions and (2) extract out key values like annual contract values (ACVs), renewal dates, and license counts that we use to populate their SpendHound dashboard.

While we have developed internal matching systems to do an initial matching of software applications and vendors, sometimes our systems will match applications and vendors incorrectly.

In this assignment, you will go through part of the quality assurance process for a sample set of data. This work is representative of the type of work you will have the opportunity to do in the Data Transformation Specialist role. You will be graded on accuracy, communication, and logic.

Goal

Your goal is to perform quality assurance and data extraction on the sample data in the attached Excel file.

You will work on two main pieces of our quality assurance and data extraction processes in this assignment by:

- (1) Reviewing all matched software applications and vendors that were matched by our system, determining if they are correct, and making suggestions if they are not correct
- (2) Extracting out three key pieces of information from the software transaction data that we use to populate the SpendHound dashboard

yipitDATA

This assignment should take approximately 2 hours to complete. Please reach out to Julia (jyue@yipitdata.com) with any questions.

Instructions

Please perform and submit your work in the attached Excel file according to the instructions below.

There are two tabs in the Excel file:

- **1. Matched Transactions**
  - Includes financial transactions data that our systems have matched to software companies for you to review
- **2. Data Extraction**
  - Includes cells highlighted in yellow to input your responses into

Part One

Instructions

1) Review the software transactions on the "*1. Matched Transactions*" tab to determine if the software applications are correctly matched for each transaction.
    a) As you go through the transactions, follow the below steps to determine if the transaction is correctly matched:
        i) Review the full transaction line item including all fields and determine if you think the transaction is correctly matched.
            (1) Fields like "*account", "memo", "line_item_description"* can all be helpful when making this determination.
        ii) You can research software applications online to determine if they look correct based on the transaction values:
            (1) When researching software companies, first search on G2
            (2) If you can't find it on G2, search using Google for the company's website or LinkedIn
2) If you believe the transaction is correctly matched to the correct software application, leave yellow highlighted cells in the row blank.
3) If you believe the transaction is incorrectly matched:
    a) Add your suggestions correct values and notes to the "*input_*" columns in the yellow highlighted cells:
        i) Add in your suggested correct product name, vendor name, and G2 link or website URL to the "*input_product_name*", "*input_vendor_name*", and "*input_company_link*" columns.
        ii) In the *"input_mistmatch_notes"* field, write a sentence explaining why you think the transaction is mismatched and why you suggested the new product instead.

yipitDATA

Notes

- This tab contains financial transaction data with transactions that were matched by our internal systems as software transactions for the period 7/1/23 - 12/31/23.
- Transaction data can be messy, and sometimes it can be difficult to determine the correct application. **You should use your best judgment. If you're not sure about the application or transaction, feel free to include notes in the "*input_other_notes*" field.**

Part Two

Instructions

1) For each software product and vendor that you believe is correctly identified on the "*1. Matched Transactions*" tab, copy and paste a unique list of the "*matched_product*" and "*matched_vendor*" names into the "*product*" and "*vendor*" columns.
    a) For example, if you think all of the software applications are correctly matched, you should have 40 rows of products and vendors in the "*2. Data Extraction*" tab.
2) For each product on the "*2. Data Extraction*" tab, review the financial transaction data and input the key values for ACV, renewal date, and license count when possible.
    a) As you go through the transactions, follow the below steps to determine the key values to extract:
        i) First review the transaction data "*memo"* and "*line_item_description"*
            (1) If it looks like the subscription is no longer active, put "Retired" in the renewal_date column and leave the other columns blank.
            (2) If it looks like the subscription is still active, and there's information about the current subscription, fill in the "*renewal_date", "acv",* and "*num_licenses"* fields if available.
        ii) If you can't collect all the information from the "*memo"* or "*line_item_description"* fields:
            (1) You may be able to use the "*line_item_amount"* field to calculate the ACV
            (2) You may be able to use the "*date"* field to estimate the "*renewal_date"*
3) For each application, write a sentence in the "*explanation*" column about how you determined each of the values.

Notes

- Not all values will be available. For any you can't find, you can leave the fields blank.
- **You should use your best judgment. If you're not sure about the application or transaction, feel free to include notes in the "*other_notes"* field.**

yipitDATA

Dataset Descriptions and Data Dictionaries

1. Matched Transactions tab

- This dataset includes the below columns:

date	• Date of the transaction
vendor	• Name of the company/vendor of the transaction pulled from the financial transaction data source • This field may be available or may be null
matched_product	• The matched product/application name based on our internal system's matching logic • Examples: "Atlassian Cloud", "Evernote"
matched_vendor	• The matched vendor/company name associated with the product/application based on our internal system's matching logic • Examples: "Atlassian", "Evernote"
line_item_amount	• Total amount of the transaction
account	• The account where the transaction occurred
memo	• The memo associated with the transaction
line_item_description	• The description associated with the transaction

The below columns are where you will input your suggestions for the product, vendor, and other fields when there is mismatched data. These columns will all be highlighted in yellow in the data file:

input_product_name	• Suggested product name for the transaction • Examples: "Atlassian Cloud", "Evernote"
input_vendor_name	• Suggested vendor/company name for the application and transaction • Examples: "Atlassian", "Evernote"
input_company_link	• G2 or company website URL to be used for reference/review
input_mismatch_notes	• Sentence explaining why you think the transaction is mismatched and why you suggested a different product instead.
input_other_notes	• Any other notes you'd like to include

2. Data Extraction tab

This tab includes the below columns, which is where you will input your data. These columns will all be highlighted in yellow in the data file:

product	• The value corresponding to "*matched_product*" on the "*1. Matched Transactions*" tab • Copy over every unique "*matched_product*" value on the "*1. Matched Transactions*" tab for each correct software vendor
vendor	• The value corresponding to "*matched_vendor*" on the "*1. Matched Transactions*" tab • Copy over the corresponding "*matched_vendor*" value for every unique "*matched_product*" value on the "*1. Matched Transactions*" tab that you copied over

renewal_date	• If there is a contract / subscription:
	  ○ This should be one day after the end date of the contract / subscription.
	  ■ For example, if the subscription is "1/31/23 - 1/30/24", you should fill in the "*renewal_date*" with 1/31/24
	  ○ Sometimes, you may see a subscription date like "1/31/23 - 1/31/24".
	  ■ In this case, the subscription date is actually "1/31/23 - 1/30/24", and you should fill in the "*renewal_date*" with "1/31/24"
	  ○ The "*renewal_date*" should always be a date in the future
	• If there is a monthly recurring subscription:
	  ○ This should be "Monthly"
	• If it looks like there is no longer an active subscription:
	  ○ This should be "Retired"

acv	• The annual value of the subscription or contract each year (also called ACV)
	• If it is a multi-year subscription, it should be the total subscription or contract value divided by the number of years in the contract
	  ○ *For example, if a contract is $300,000 over 3 years, the "acv" is $100,000*
	• If it is a monthly or shorter than annual subscription, you should scale up the cost of the monthly or shorter time period value to the annual cost
	  ○ *For example, if a subscription is $100/month, the "acv" is $1200*

	• If “*renewal_date*” is “*Retired*”, this field should be blank
num_licenses	• The number of licenses included in the subscription • If “*renewal_date*” is “*Retired*”, this field should be blank
summary	• Explanation of how you determined each of the values
other_notes	• Any other notes, comments, or questions about the data