**Mini Project 1**

Besides technical skills and knowledge in data analytics, it is essential that one should develop and hone their problem solving and critical thinking skills to address business and organizational issues. The purpose of mini projects in this class is to provide an opportunity for students to practice applying their technical knowledge to support organizational decisions. An important aspect of these projects is that simulated real-life scenarios and realistic (though fictitious) data are used in the projects.

**The Company:**

*eLinks* is an enterprise networking company that provides a platform for organizations to communicate and collaborate. Many companies in diverse areas of business including, for example, manufacturers, producers, suppliers, retailers, transportation, and others use *eLinks*.

Basic membership of *eLinks* is free though the company provides many other value-added services at modest cost upfront or through annual subscription. *eLinks* has a Data Analytics (DA) department, whose primary responsibility is to support better product and business decisions using data. The DA teams conduct studies, carry out projects to address specific business problems, and perform ad-hoc analyses to support business decisions.

**The Problem at Hand:**

The management of *eLinks* has noticed that user engagement with the company’s platform appears to have dropped in the most recent days. The management is unsure whether this is actually the case, and if so, what possible reasons for drop in user activity may be. The DA has been asked to look into this issue and advise the management team.

The Data

|  |  |
| --- | --- |
|  | **USERS Table** |
| user\_id: | A unique ID per user. Can be joined to user\_id in either of the other tables. |
| created\_at: | The time the user was created (first signed up) |
| state: | The state of the user (active or pending) |
| activated\_at: | The time the user was activated, if they are active |
| company\_id: | The ID of the user's company |
| language: | The chosen language of the user |
|  | EVENTS Table |
| user\_id: | The ID of the user logging the event. Can be joined to user\\_id in either of the other tables. |
| occurred\_at: | The time the event occurred. |
| event\_type: | The general event type. There are two values in this dataset: "signup\_flow", which refers to anything occuring during the process of a user's authentication, and "engagement", which refers to general product usage after the user has signed up for the first time. |
| event\_name: | The specific action the user took. Possible values include: create\_user: User is added to Yammer's database during signup process enter\_email: User begins the signup process by entering her email address enter\_info: User enters her name and personal information during signup process complete\_signup: User completes the entire signup/authentication process home\_page: User loads the home page like\_message: User likes another user's message login: User logs into Yammer search\_autocomplete: User selects a search result from the autocomplete list search\_run: User runs a search query and is taken to the search results page search\_click\_result\_X: User clicks search result X on the results page, where X is a number from 1 through 10. send\_message: User posts a message view\_inbox: User views messages in her inbox |
| location: | The country from which the event was logged (collected through IP address). |
| device: | The type of device used to log the event. |
|  | Emails Table |
| user\_id: | The ID of the user to whom the event relates. Can be joined to user\_id in either of the other tables. |
| occurred\_at: | The time the event occurred. |
| action: | The name of the event that occurred. "sent\_weekly\_digest" means that the user was delivered a digest email showing relevant conversations from the previous day. "email\_open" means that the user opened the email. "email\_clickthrough" means that the user clicked a link in the email. |

**Understanding the Problem**:

*eLinks* defines user activity as an engagement with its online portal, i.e., the customers (users) having made some type of server call by interacting with the company’s website/web server. Such events are listed as “engagement” in the event\_type column of the EVENTS table.

**Your Task:**

Please do necessary analyses using SQL to address the following:

(1) Has actually user activity or engagement dropped recently and if so, how serious or significant is it?  
(2) Think about possible reasons (at least three) for drop in activity, i.e., develop some hypotheses that you can later test if/as necessary in a future analysis. Investigate each of these potential reasons by conducting analysis using the relevant data, writing SQL queries, and generating related visualizations.

**Your Recommendations:**

What are your findings regarding whether drop in user activity is significant or no. What seems like the most likely cause of the drop in engagement?

Additional (optional) questions some of which you might want to include in your report:

If there are questions that you can't answer using data alone, how would you go about answering them (hypothetically, assuming you actually worked at this company)? What, if anything, should the company do in response?

Do the answers to any of your original hypotheses lead you to further questions? If so, what are they and how will you test them?

**Deliverables:**

It is recommended that you should please use the Databricks platform where you should create a Python notebook. In the notebook, you should use code cells for the SQL queries, and markdown cells to describe your findings, interpretations, and recommendations. You can also necessary charts within the notebook as well. Here’s a link to a cheat sheet for markdown: <https://www.markdownguide.org/cheat-sheet/>

Another link: <https://duckduckgo.com/?t=ffnt&q=Markdown+cheat+sheet&ia=answer>

**Please publish your notebook in Databricks and submit the link to the notebook. (To make sure that the link is correct and it works, use that link after closing logging off your Databricks account and closing the browser.)**