**INTRODUCTION**

This is project from DataTribe SQL Workshop Challenge.

**SCENARIO**

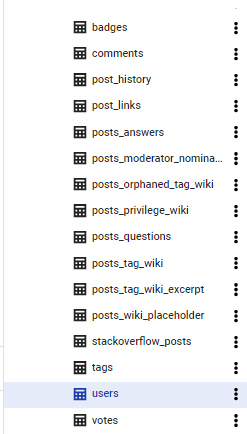
You’re currently employed as an junior analyst for stackoverflow.com. The business stakeholders found out that currently. The monthly total number of comments have declining trend since 2016

**POINT TO ANALYZE**

1. Statistic activities by user reputation (2017 – 2022)
2. Question with answer ratio (2017 – 2022).
3. Day of week with the less question answer (within an hour) in the last 3 years

**DATA SOURCE**

Data of Stackoverflow.com available on google cloud and can be access by BigQuery. The data has 16 tables.



**PROCESS**

First, I make measure plan by looking to the poin thaw want to analyze and which data from dataset that needed. Because Stack Overflow (SO) is a community-based question answering service that targets developers and software engineers, I consider to use table users, post\_questions, and post\_answers.

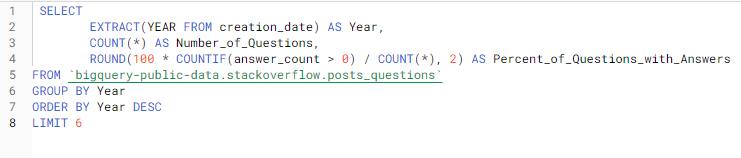
**DATA TO ANALYZE**

The next step is collecting data to analyze based on point to analyze. In order to get the data, I use some query to get the data.

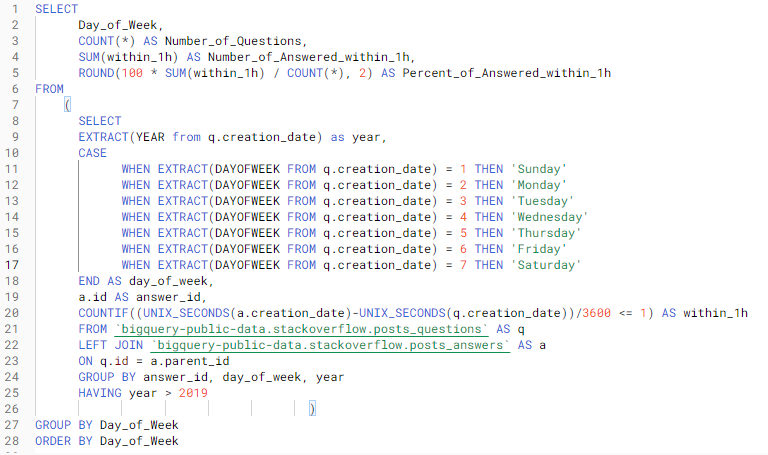
Point 1 - Statistic activities by user reputation (2017 – 2022).



Point 2 – Question with Answer ratio (2017 – 2022)



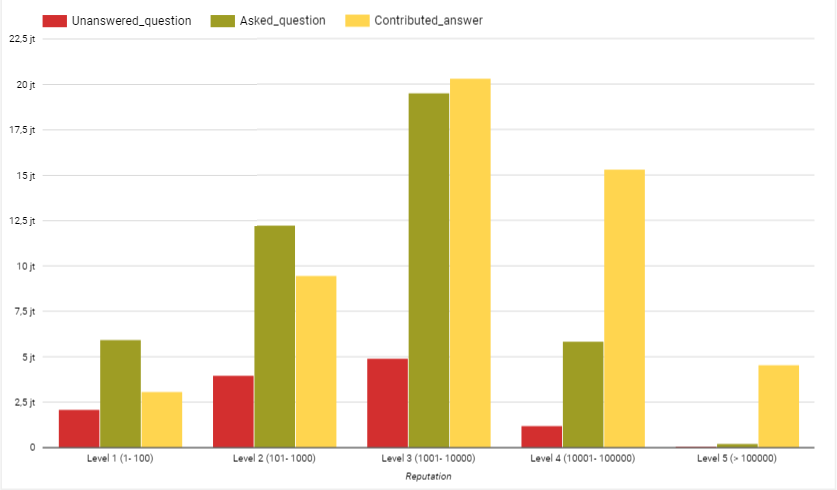
Point 3 - Day of week with the less question answer (within an hour) in the last 3 years



After performing that query and get the data, now time for analyze and make some presentation about insight we get from data. We explore the data result from that quey using google data studio and make some chart and insight.

**ANALYSIS AND INSIGHT**

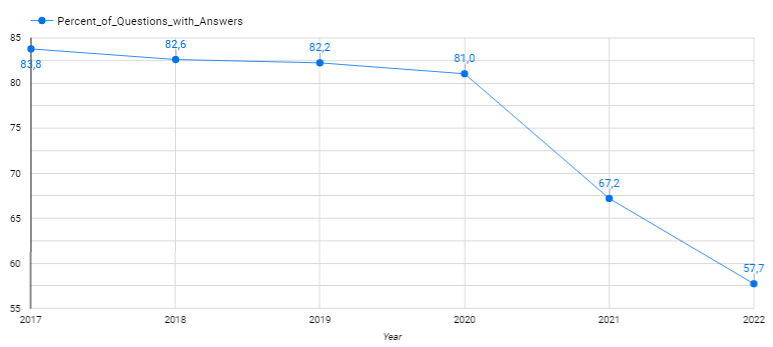
First, we look for user activities by their reputation. There is 3 activities to highlight, user with question to asked, user with question without get any answered, and user that contribute to answer some question. For the reputation, I divide to 5 groups level 1 (1-100 reputations), level 2 (101-1000 reputations), level 3 (1001-10000 reputations), level 4 (10001-100000 reputations), and level 5 (>100001 reputations).

****

From the data and chart, we can see that:

* User level 3 very active based on their three activities, based on the chart we can see that in the user level 3 have ‘asked question’ and ‘contribute answer’ almost equal.
* By the ratio of ‘asked question’ and ‘contribute answer’, we can assume that user with higher ‘asked question’ than ‘contribute answer’ is beginner (level 1 and level 2), and user with higher ‘contibute answer’ than ‘asked question’ is professional (level 4 and level 5).

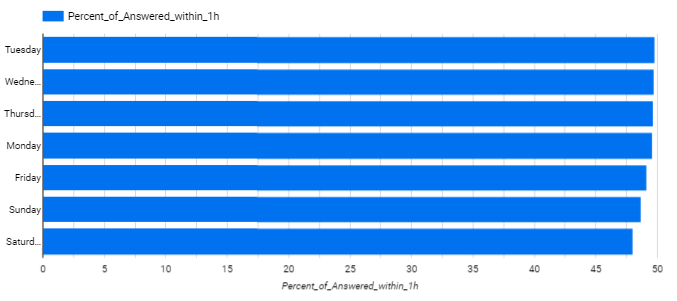
Second, we look for the percentage ratio question that have answer in the past 5 years (2017 – 2021).



From the data, we can see that:

* For the first 3 year (2017 – 2020) the trend is quite stable but on the 2021 there is massive declining trend.

And the last, we look for day of week percentage question that have answer within 1 hour. We use this data to see commonly active user to respon for question by day.



From the data and chart, there is no big difference everyday either is weekday or weekend. The ratio respon question within 1 hour everyday is above 45% but still less than 50%.

**CONCLUSION**

According to the data, Stack Overflow (SO) is a community-based question answering service that targets developers and software engineers. This is very depending on user activities. But in the 2021 there is a massive declining trend for question that have answer. My suggestion as junior data analyst to solve the problem by data are:

* We can reach out the user level 5 reputation (professional) to be more active.

People usually come we spesifict problem to solve and need it ASAP, and its sometime can be solved by professional user.

* We can create event or challenge ‘bounty hunter’ for user level 1,2 and 3 to compete for solved every problem to keep engangement and increase activation user for the website.