**Part 1.1 - Understanding APIs (5 points)**

• **1.1.1 (2)** How many API calls were required to collect the submissions?

The process of a client application submitting a request to an API and the API getting the desired data from an external server is known as an API call.

Only, 60 request can be made in a minute and per request we can fetch 100 items.

In the given question, we had to fetch 2991 items so total requests that we required to make was 30.

• **1.1.2 (1)** Why did we set the submission limit at 1000?

To understand the rules which we must adhere to while making the API calls

For machine learning project we need more data to train the algorithm so we tried to fetch more than 1000 post at a time but we couldn’t as it’s the limit set by reddit API. The limit is set to such a number because a normal user wouldn’t look more than 1000 top posts in a subreddit. Therefore, post over 1000 wouldn’t impact and will over fit our model

so obtain more than 1000 we have to use push API, which is third party API.

• **1.1.3 (2)** How long, in minutes, would it take you to collect 1000 posts

from 25 different subreddits? What about from 500 different subreddits?

*Hint: You’ll have to consider how many API requests you are allowed to*

*Make*

*As par rules for API, we can make at max 60 request and each request can fetch up to 100 post so in a minute we can fetch at max 6000 items so in 4 minute we can fetch 24000 post and rest thousand will require 0.003 sec. So it takes almost 4 min to fetch 25000 items from praw.*

*And similarly for 500 subreddit it will take around 83 minutes to fetch.*

**Part 1.2 Thinking about your sample (3 points)**

• **1.2.1 (1)** Do you think these posts are representative of **all** the posts on

that subreddit?

Yes, as most normal users wouldn’t look at post over 1000 so this data represents all the top viewed post on a particular subreddit.

• **1.2.2 (2)** Why or why not? That is, if you think so, why do you think

there’s not much sampling bias here? If not, what do you think might be

different about these top posts than other posts?

As most users will be looking at top 1000 post most of the time so post over 1000 is not much related and wont be affecting more model in any ways and in 1000 top post are post were most people reacted to it which it was most relevant to them or were useful for them.