1. Data summary
   a. How many **unique** encounters are there in the data?
   b. How many **unique** users are there in the data?

Answer:

There are three tables provided. Based on the typical flow of a patient's visit, the order of the three tables should be: 1. Id_verification 2. Doctor_enters 3. First_patient_response. Using the "SELECT COUNT(DISTINCT encounter_id) AS unique_encounter", the unique encounters in each table are determined:

| | Unique encounters | Unique users |
|---|---|---|
| Id_verification | 650 | 639 |
| Doctor_enters | 461 | - |
| First_patient_response | 448 | - |

The values of unique encounters in the three tables are decreasing, which can be explained that some patients left before they could talk with a doctor.

In the "id_verification" table, the number of unique users is less than that of unique encounters. Using the following query, it was found 11 users have two encounters:

SELECT DISTINCT user_id AS unique_user,
        COUNT(DISTINCT encounter_id) AS unique_encounter
FROM `curai-patientidverify.aa.id_verification`
GROUP BY user_id
ORDER BY unique_encounter DESC

The results are listed in the table below:

| unique_user | unique_encounter | Reason |
|---|---|---|
| User:475599 | 2 | Patient did not confirm selfie at 1st encounter_id |
| User:474965 | 2 | |
| User:477044 | 2 | |
| User:478327 | 2 | |
| User:476766 | 2 | |
| User:479123 | 2 | Patient did not start the ID verification at 1st encounter_id |
| User:479602 | 2 | |
| User:474392 | 2 | |
| User:476887 | 2 | Patient did not start the ID verification at both encounter_ids |
| User:474942 | 2 | |
| User:479491 | 2 | Patient did not confirm selfie at 1st encounter_id; Patient did not start the ID verification at 2nd encounter_id |

When looking into the details of these 11 users, it was found that these patients had troubles in starting ID verification or confirming selfie. Considering ID information are sensitive, it might be the following reasons that caused this:

1. these patients were cautious about sharing their personal information
2. these patients did not have their ID card in hand
3. these patients did not trust Curai platform.

Curai can do the following things to keep more patients:

1. do patient survey, understand why these patients did not go to next step
2. make both steps (ID verification and selfie confirmation) simple
3. improve the platform design or platform reputation to earn more patients' confidence.

Since some patients' ID verification was not successful, they could not wait for a doctor to enter. When using the following query to filter these patients out,

```
SELECT COUNT(DISTINCT encounter_id) AS unique_encounter,
       COUNT(DISTINCT user_id) AS unique_user
FROM `curai-patientidverify.aa.id_verification`
WHERE verification_confirmed_timestamp IS NOT NULL
```

we can calculate the number of unique encounters and unique users whose ID verification was successful, which is 493. The potential patient loss is caused by the ID verification and confirming selfie. This is further demonstrated by the activity of verification confirmed patients. For example, User 479721 (Encounter 152649) saw the ID verification request twice (10/10/2020 17:51 and 10/10/2020 17:55), and started the ID verification twice (17:53 and 17:55), but confirmed selfie once at 17:59. It's possible that this patient did not confirm selfie successfully in the first time and then tried the second time. If this was true, Curai might need to think about designing a more user-friendly platform to save patient/user's time for selfie confirmation.

2.   What is the drop-off rate between each step from saw_verification to patient_response on a **per encounter basis?** (Further breakdown by platform.)

Answer:

To calculate the drop-off rate at each step on a per encounter basis, it's essential to simplify the table "id_verification" since some rows are replicated. The following query was used to find the necessary information:

```
SELECT DISTINCT encounter_id, user_id, min(starts_at) as starts_at,
        max(ends_at) as ends_at, platform,
        min(saw_verification_timestamp) as saw_verification_timestamp_min,
        max(start_veritifcation_timestamp) as start_veritifcation_timestamp_max,
        max(confirmed_selfie_timestamp) as confirmed_selfie_timestamp_max,
        max(verification_confirmed_timestamp) as
verification_confirmed_timestamp_max
FROM `curai-patientidverify.aa.id_verification`
GROUP BY encounter_id, user_id, platform
ORDER BY encounter_id
```

The results showed there were 654 encounter_ids, which is different from the number of unique encounter_id (650) calculated in Question 1. The reason is that 4 encounter_ids used two different platform, as showed in the following table:

| Encounter_id | Platform | Activity details (based on original dataset "id_verification") |
|---|---|---|
| Encounter:144711 | ios | User saw verification at 3:37 on ios.  It's possible that the user did not start verification successfully on web, then ios was tried. |
| | web | User saw verification at 3:33 on Web. |
| Encounter:145059 | android | The user saw verification at 17:42 on android, and then started verification at 17:43 |
| | web | User behavior on web is the same as that on android. No idea why this user used two different platforms. |
| Encounter:150532 | ios | "start_verification_timestamp" (0:19) is older than "saw_verification_timestamp" (0:25), which does not make sense. |
| | web | start_verification_timestamp" (0:19) is as same as "saw_verification_timestamp" (0:19) |
| Encounter:150831 | ios | "start_verification_timestamp" (3:13) is older than "saw_verification_timestamp" (3:16, 3:20), which does not make sense. |
| | web | The user saw verification at 3:11 on web, and then started verification at 3:13 |

Based on the above table, it seems that web activity is more representative. Considering the following step requires further breakdown by platform, the above 4 encounter_ids were deleted for simplification. That is to say, there are 646 unique encounter_ids remaining.

The three tables were joined to calculate the drop-off rate at each step:

| dfrate_StartVeri_SawVeri | dfrate_ConfirmSelfie_StartVeri | dfrate_VeriConfirm_ConfirmSelfie | dfrate_DrEnters_VeriConfirm | dfrate_PatientResp_DrEnters |
|---|---|---|---|---|
| 0.17 | 0.06 | 0.02 | 0.07 | 0.03 |

The above table shows the step from saw_verification to start_verification has the highest drop off rate.

When breaking down by platform, the results are:

| platform | dfrate_StartVeri_SawVeri | dfrate_ConfirmSelfie_StartVeri | dfrate_VeriConfirm_ConfirmSelfie | dfrate_DrEnters_VeriConfirm | dfrate_PatientResp_DrEnters |
|---|---|---|---|---|---|
| web | 0.2 | 0.05 | 0.03 | 0.07 | 0.04 |
| ios | 0.11 | 0.08 | 0.01 | 0.05 | 0.01 |
| android | 0.12 | 0.2 | 0 | 0 | 0 |

The above table shows that for all three platforms, the step from saw_verification to start_verification has the highest drop off rate. Moreover, for Android platform, the drop off rate from start_ verification to confirmed_selfie is as high as 0.2. Is this data reliable? When looking into the number of users by platform, it was found around 70% of users are using web platform, while only 3% users are using android platform. Therefore, the high drop off rate of android platform is not representative due to the low number of android users.

| platform | Number of Users | Percentage |
|---|---|---|
| web | 453 | 70% |
| ios | 176 | 27% |
| android | 17 | 3% |

Notes:
- dfrate_StartVeri_SawVeri: drop off rate from saw_verification to start_verification
- dfrate_ConfirmSelfie_StartVeri: drop off rate from start_verification to confirmed_selfie
- dfrate_VeriConfirm_ConfirmSelfie: drop off rate from confimed_selfie to verification_confirmed
- dfrate_DrEnters_VeriConfirm: drop off rate from verification_confirmed to dr_enters
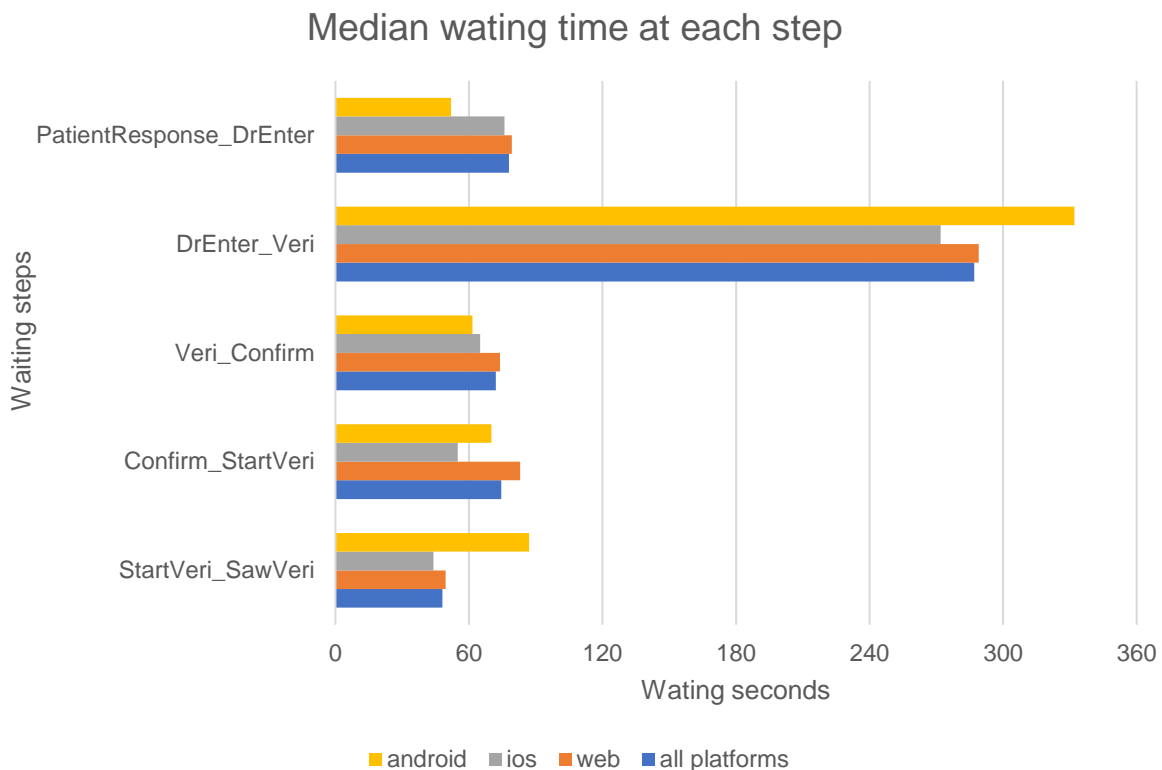- dfrate_PatientResp_DrEnters: drop off rate from dr_enters to patient_response

3. What is the median time between each step from saw_verification to patient_response on a **per user basis?** (Further breakdown by platform.)

Answer:

The median time from saw_verification to patient_response is 762 seconds, around 13 mins. When further breaking down by platform, the median time is shown in the table below, which indicates that the ios platform has the shortest median time (667 seconds, 11 mins) while the android platform has the longest median time (904 seconds, 15 mins).

| platform | Median time (saw_verification to patient_response) | Median time in minutes |
|---|---|---|
| web | 793 | 13 mins |
| ios | 667 | 11 mins |
| android | 904 | 15 mins |

The following figure shows the median waiting time between each step. Overall, it took a patient around 5 mins to wait for a doctor to enter after the verification was confirmed, while android users experience the longest waiting time, 5.5 mins. Also, for all waiting steps, the median waiting time on ios platform is shorter than that on the web platform.

## Median wating time at each step



Waiting steps / Wating seconds

Legend: android, ios, web, all platforms

4. Are there any other additional valuable insights from the data?

Answer:

From the data in the three tables, it was found that some patients left earlier before they could talk with a doctor, which is mainly caused by the long waiting time. To keep more patients, Curai may focus on the following areas:

- shorten waiting time
  - o The step from "verification confirmation" to "doctor enters" has the longest median waiting time, around 5 mins. What caused this? Is it because there are not enough doctors? Or there are enough doctors, but doctors did not respond in time?
  - o The time period from "doctor enters" to "patient response" is longer than 1 min. Is it because patients did not know doctor enter in time? If yes, how can patients get notified as soon as possible?
- make the ID verification step simple
  - o The drop off rate from "saw verification" to "start verification" is the highest among all steps, meaning some patients are cautious about sharing their personal information. Is it OK to ask patients to input the least personal information? Or is it possible to connect with other medical platforms to get patients' information directly so that patients do not need to input their information?
  - o Sometimes it took a patient multiple times to get ID verified successfully, which resulted in potential patient loss. Is it possible to design a more user-friendly platform so that ID can be verified successfully when a patient only tries once?
- improve algorithms to make sure the waiting time of web platform is the shortest
  - o Around 70% of current patients are from web platform. However, the median waiting time of web platform is longer than that of the ios platform. Why does ios platform has shorter waiting time? Is there anything the web platform can learn from?
- penetrate to more android users
  - o Currently only 3% users are using android platform. Is it because Curai did not penetrate to enough android users, or is it because android users prefer to use web platform due to the long waiting time on android platform?
  - o Curai may need to redesign the app on android platform to shorten patients' waiting time.
- earn more patients' confidence
  - o Some patients left before they could confirm their selfie. Is it because they do not trust Curai platform? If yes, Curai may need to strengthen a positive business reputation, either by marketing campaigns or by earning existing patients' referrals.