**Análisis desde los logs/registros en txt:**

* log\_carga\_de\_procesador\_con\_preguntas\_cerradas.txt
* log\_preguntas\_efectivamente\_pesistidas\_1.txt
* log\_preguntas\_efectivamente\_pesistidas\_2.txt
* log\_preguntas\_efectivamente\_pesistidas\_3.txt

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| JSON | Cantidad de Registros Iniciales (sin aplicar filtros) | Cantidad de Registros no vacíos | Cantidad de Registros solo gif o emoticón | Cantidad de Registros signos raros (ejemplo: +1) | Cantidad de Registros solo símbolos | Cantidad de Registros para procesar |
| 1 | 793 | 775 | 0 | 0 | 1 | 774 |
| 2 | 601 | 586 | 2 | 0 | 0 | 584 |
| 3 | 622 | 604 | 1 | 0 | 1 | 602 |
| Total | 2016 | 1965 | 3 | 0 | 2 | 1960 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| JSON | Mensajes Sueltos (que no son pregunta ni respuesta) | Mensajes Concatenados  (que se unieron a una pregunta) | Mensajes Cierre de Alumnos (Por ejemplo : gracias! ) | Mensajes considerados Preguntas | Mensajes considerados Respuestas | Total mensajes analizados |
| 1 | 1 | 18 | 79 | 189 | 487 | 774 |
| 2 | 0 | 20 | 33 | 153 | 378 | 584 |
| 3 | 1 | 28 | 32 | 137 | 404 | 602 |
|  | 2 | 66 | 144 | 479 | 1269 | 1960 |

|  |  |  |
| --- | --- | --- |
| JSON | Cantidad de mensajes detectados como preguntas | Cantidad de mensajes asignados como respuestas |
| 1 | 189 | 775 |
| 2 | 153 | 619 |
| 3 | 137 | 626 |
| Total | 479 | 2020 |

**Nota:** La suma de la cantidad de preguntas y respuestas detectadas (2499) no coincide con la cantidad de registros analizados (1960), dado a que algunas respuestas se han asignado a varias preguntas (se duplicaban), priorizando no perder la información.

**Análisis desde la Base de Datos:**

* **Tabla Mensajes**
  + Cantidad de Mensajes: 2499
  + Cantidad de Mensajes que son preguntas (con mensajes repetidos): 479
  + Cantidad de Mensajes que son respuestas (con mensajes repetidos): 2020
  + Cantidad de Mensajes distintos: 1729
  + Cantidad de Mensajes que son preguntas (sin mensajes repetidos): 479
  + Cantidad de Mensajes que son respuestas (sin mensajes repetidos): 1250

La cantidad de mensajes distintos: 1729, no coincide con la cantidad de registros analizados 1960, dado a que pudo haber respuestas idénticas almacenadas varias veces dado a que se correspondían a preguntas distintas.

* **Tabla Preguntas**
  + Cantidad de Preguntas: 479
  + Cantidad de Preguntas Distintas: 479
  + Cantidad de preguntas marcadas como sin\_contexto (que tienen 6 palabras):
    - Considerando repetidas: 8
    - Sin repetidas: 8
  + Cantidad de preguntas administrativas:
    - Considerando repetidas: 116
    - Sin repetidas: 116
  + Cantidad de preguntas que son administrativas y sin contexto a la vez: 2
  + Cantidad de preguntas que son administrativas o sin contexto : 122
  + Cantidad de preguntas que van a embeddings (las que no son administrativas y no son sin contexto): 479 – 122 = 357
* **Tabla Respuestas**
  + Cantidad de Respuestas: 2020
  + Cantidad de Respuestas Distintas: 1250
  + Cantidad de respuestas marcadas como cortas (hasta14 caracteres):
    - Considerando repetidas: 125
    - No repetidas: 72
* Cantidad de respuestas validadas:
  + Con repetidos: 1182
  + Sin repetidos: 704
* Cantidad de preguntas sin respuestas validadas: 0
* Cantidad de preguntas sin respuestas: 0
* Cantidad de respuestas sin preguntas: 0
* **Tabla autores**
  + Cantidad de registros: 153
  + Cantidad de registros distintos: 153