

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



El futuro es de todos

DNP
Departamento
Nacional de Planeación

Buzón de Ingreso Solidario

Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Social

Sector

Salud y protección social

Lenguaje

Python

Fuente de datos

Buzón de correo electrónico
Ingreso Solidario

Presentación

En el marco de la actualidad que está viviendo el mundo, a causa de la pandemia de enfermedad por coronavirus o también llamada COVID 19, en Colombia el gobierno nacional ha creado un esquema mediante el cual se entrega una transferencia monetaria en favor de los hogares en situación de pobreza y vulnerabilidad que no sean beneficiarios de programas sociales del Estado, esto ha despertado inquietudes en cientos de ciudadanos que han requerido múltiples plataformas para que los entes públicos den respuesta a estas, ante estos requerimientos el DNP habilitó un buzón de correo electrónico dispuesto a recibir las inquietudes de los ciudadanos. El uso de este medio por parte de los colombianos fue masivo excediendo la capacidad de respuesta, lo cual trajo consigo el represamiento de al menos 160.000 correos. Desde la unidad de ciencia de datos se propone el desarrollo de un script que permita clasificar los correos electrónicos en 10 etiquetas generadas por la Dirección de Desarrollo Social presentes en el buzón de entrada y posteriormente automatizar la respuesta a cada uno de estos. El proceso de clasificación se generó a través de reglas de asociación.

In the current framework that the world is experiencing, due to the coronavirus disease pandemic or also called COVID 19, in Colombia the national government has created a scheme through which a monetary transfer is delivered in favor of households in situation of poverty and vulnerability who are not beneficiaries of State social programs, this has aroused concerns in hundreds of citizens who have required multiple platforms for public entities to respond to these, in response to these requirements the DNP enabled an email mailbox available to receive the concerns of citizens. The use of this medium by Colombians was massive, exceeding the response capacity, which brought with it the damming of at least 160,000 emails. The data science unit proposes the development of a script that allows to classify the emails into 10 labels generated by the Social Development Department present in the inbox and subsequently automate the response to each of these. The classification process was generated through association rules

Objetivo general

Clasificar la mayor cantidad posible de correos alojados en el buzón de ingreso solidario, mediante el uso de reglas de asociación, generando el insumo requerido para dar respuesta automática a dichos correos a través del uso de paquetes del software libre Python.

Objetivos específicos

1. Descargar los correos electrónicos alojados en el buzón de ingreso solidario a través de Outlook.
2. Implementar un script para la clasificación automática de los correos electrónicos en 10 etiquetas definidas por la Dirección de Desarrollo Social.
3. Automatizar el proceso de envío o respuesta de correos electrónicos a través del software libre Python.
4. Realizar envío de correos electrónicos, teniendo en cuenta las restricciones de Outlook.



Metodología

La necesidad de este proyecto se fundamenta en los siguientes factores:

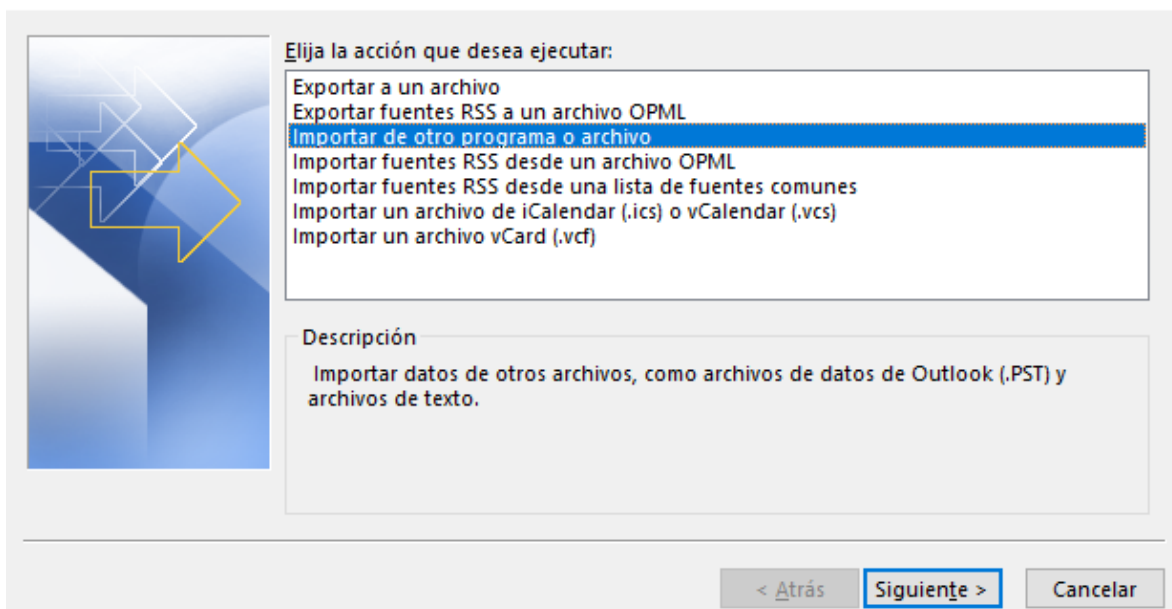
- Correos electrónicos entrantes superan las capacidades operacionales de los encargados de dar respuesta a estos.
- Necesidad de estandarizar el proceso de respuesta de correos electrónico

Descarga de correos electrónicos

La descarga de correos electrónicos se realizó a través de la aplicación de escritorio de Outlook. El proceso realizado fue abrir la aplicación, luego clic en archivo en esta sección clic en abrir y exportar, posteriormente en la opción importar y exportar. La aplicación abre una ventana emergente con el título de “Asistente para importar y exportar” (Figura 1), en esta se seleccionó la primera acción “Exportar a un archivo”, luego se seleccionó el tipo de archivo “Valores separados por comas”, luego se selecciona la carpeta que se quiere descargar (En este punto se establecieron cuatro descargas definidas por fechas), se puso la ruta deseada y se procedió a la selección de variables de los correos para realizar la clasificación y el envío, estas variables fueron Asunto, Cuerpo, De: (nombre), De: (dirección) y De: (Tipo).

Figura 1: Asistente para importar y exportar

Asistente para importar y exportar



Fuente: elaboración propia

Como se mencionó anteriormente el tipo de archivo que se descarga de la aplicación Outlook es de tipo valores separado por comas, para realizar su tratamiento se desarrolló un script¹ en Python que permitiera convertir el archivo .csv a un archivo .xlsx.

¹ Archivo en el cual se guarda el código escrito en un lenguaje de programación.



Clasificación de correos electrónicos:

Posterior a la conversión del archivo se realiza un procesamiento a la columna de asunto y cuerpo de los correos electrónicos, este procesamiento consistió en la remoción de acentos y la conversión a textos a minúscula, tras este proceso y bajo el insumo de palabras comunes suministrado por los colaboradores que estaban realizando la clasificación de manera manual, se lograron determinar reglas de asociación o reglas “fuertes” variadas por el uso de grillas en Python.

Las etiquetas fueron nombradas por una letra mayúscula y un número, estas fueron suministradas por la Dirección de Desarrollo Social. Las etiquetas fueron:

- A1: Aún no he recibido un SMS
- B1: ¿Cómo puedo acceder al programa?
- A6: ¿Cómo reclamar el dinero?
- A12: He tenido problemas con el banco
- A4: Mi número de celular es nuevo
- A10: Envío datos para actualizarlos
- A5: No tengo cuenta bancaria
- A2: Aún no he recibido el dinero
- D3: ¿Estoy dentro de los beneficiarios?
- A13: Ahora no aparezco como beneficiario

Inicialmente la clasificación se estableció como variables dummie donde la columna hace referencia a la etiqueta, por lo tanto, si un correo tiene un 1 en la columna “A1” significa que según las palabras que contiene el correo en el asunto o en el cuerpo este pertenece a esa etiqueta, sin embargo, el correo podría implicar más de una etiqueta. Para solucionar este problema se desarrolló un script de etiqueta que tenía en cuenta la longitud del correo por caracteres (menos de 1500) y la jerarquización de las etiquetas, es decir, que si un correo se asignaba con más de una etiqueta se seleccionaría aquella etiqueta más relevante, el orden de esta clasificación tiene el orden asignado en la lista de etiquetas descrita anteriormente.

Cruce entre el proceso manual y el proceso automático

Si bien es cierto que el proceso se empezó a desarrollar automáticamente, fue necesaria la clasificación manual de correos que no se fueron clasificados por las reglas de asignación o porque su contenido era extenso y requería una revisión más detallada. En el transcurso del proyecto se tuvieron una gran cantidad de datos por parte de los colaboradores que estaban desarrollando la clasificación manual, a partir de esta información se realizó un cruce entre las bases de datos.

Automatizar el proceso de envío de correos electrónicos

Luego de tener una base de datos consolidada se generó la necesidad de responder masivamente los correos electrónicos, para esto se desarrolló un script en Python que lo hiciera automáticamente. Se usó la librería “smtplib” y “MIME” para el envío masivo de correos y el lenguaje “HTML” para dar formato a los textos presentes en el correo electrónico.

Como se mencionó anteriormente la DDS desarrolló casi 30 etiquetas, por lo cual la automatización del envío de correos electrónicos debería tener en cuenta cada una de estas, a partir de la dirección se establecieron respuestas tipos por grupos de etiquetas, se terminaron construyendo 11 mensajes diferentes para dar respuesta a las 30 etiquetas, estos se alojan en funciones dentro del script.



Las librerías usadas para el envío masivo de correos fueron aplicadas en una función y posteriormente con el uso de un ciclo iterativo se desarrollo un condicional que llamaba a la función de envío y a la función del tipo de mensaje que enviaría.

Proceso de envío de correos electrónicos

Por restricciones de la aplicación Outlook indicadas por la OTSI se definió que el número máximo de correos que se podía responder diariamente es de 5.000 correos electrónicos, sin embargo, el número de correos electrónicos represado llego a alcanzar más de 90.000, debido a esta situación y teniendo en cuanto al flujo entrante y continuo de correos se estableció una estrategia de envíos masivos diariamente con el objetivo de responder a cada uno de los ciudadanos que requerían una respuesta.

Resultados

Descarga de correos electrónicos

Se descargaron un total de 179.895 correos electrónicos desde el buzón de entrada, también se descargaron algunos correos de carpetas clasificados manualmente y de correos no deseados, alrededor de 4000 correos adicionales. Se descargó con éxito el 100% de los correos electrónicos.

Clasificación de correos electrónicos:

El proceso de clasificación automática de correos electrónicos alcanzo un rendimiento del 45%, por lo tanto, se logró clasificar automáticamente alrededor de 80.952 en las etiquetas dispuestas en la metodología. Esto significo la disminución de carga de trabajo para los colaboradores que estaban encargados de la clasificación manual.

Cruce entre el proceso manual y el proceso automático

El cruce entre el proceso manual y el proceso automático permitió obtener la clasificación del 100% de los correos descargados. Se debe tener en cuenta que todo el proyecto se desarrolló en un trabajo continuo, por lo tanto, a medida que se iban teniendo avance por parte de los colaboradores del DNP se realizaba un cruce con las bases de datos con el objetivo de enviar correos constantemente. En total se realizaron más de 12 cruces entre las bases de datos enviadas manualmente con la base de datos destinada a enviar los correos electrónicos.

Automatizar el proceso de envío de correos electrónicos

La automatización permitió que se realizaran envíos masivos de 5000 correos diariamente, además de imprimir un reporte para cada uno de los envíos. El reporte de los envíos es guardado en un archivo de formato Excel, que contiene el correo electrónico y el estado de dicho correo. Los estados que pueden obtener los correos son:

- *email sent*: correos enviados de manera exitosa.
- *error*: correos no enviados, ya sea porque el correo no se logra identificar o porque el correo ha sido rechazado.
- *no enviado*: a pesar de que no es directamente adicionado por el script, este hace referencia a correos que no fueron etiquetados para pasar por el script.

Proceso de envío de correos electrónicos

El proceso de envío de correos electrónicos tuvo como resultados los siguientes reportes:



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación

Tabla 1. Reporte de envíos

Descarga	Correos Totales	Correos Enviados	Correos No Enviados	Correos Error
Primera descarga	93.830	91.451	9	2.370
Segunda descarga	44.201	41.284	31	2.886
Tercera descarga	11.008	10.886	0	122
Cuarta descarga	30.856	30.568	0	288

Fuente: elaboración propia

Respecto a los correos que tuvieron error se coordinó un proceso de recuperación manual del correo electrónico, en este se identificaron el correo de 1635 correos que habían tenido error en los procesos anteriores.

Conclusiones y recomendaciones

1. El proceso de clasificación y envío de correos término satisfactoriamente, se respondió el 100% de los correos electrónicos entrantes antes de entregar el proceso al Departamento de Prosperidad Social.
2. Se recomienda al DNP en general que cuando se realicen aperturas al público de una herramienta de comunicación masiva, se planeen los procesos a seguir posterior a la apertura, con eso podrían evitarse tener procesos sobre la marcha de un proyecto.

Socialización

Los resultados del presente proyecto se socializaron con la Dirección de Desarrollo Social, la Unidad de Científicos de Datos y el Director de Desarrollo Digital del DNP.