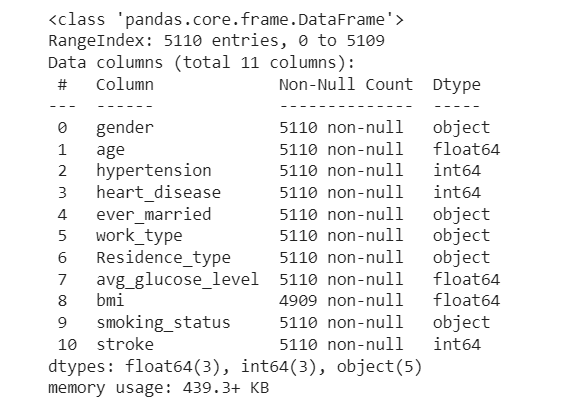
**Informe acerca de las tendencias en ataques cerebrovasculares**

El siguiente informe constituye una aproximación a las tendencias detrás de los ataques cerebrovasculares en la población. Se considera importante que las Autoridades conozcan estas causas debido a que, según la Organización Mundial de la Salud, los ataques cerebrovasculares son responsables del 11% de muertes a nivel global.

En primer lugar, se realizó una aproximación preliminar del comportamiento de los datos mediante el lenguaje de Python, con uso de las bibliotecas Numpy, Pandas, Matplotlib y Scikit-learn. Un A continuación, se aprecian las características generales del dataset en el que se basa el presente trabajo, donde destaca la existencia de 5110 observaciones, con 11 características principales.



Según los datos recopilados, el promedio de edad de las personas entrevistadas es de 44,20 años. Tal como se puede visualizar en el siguiente gráfico:

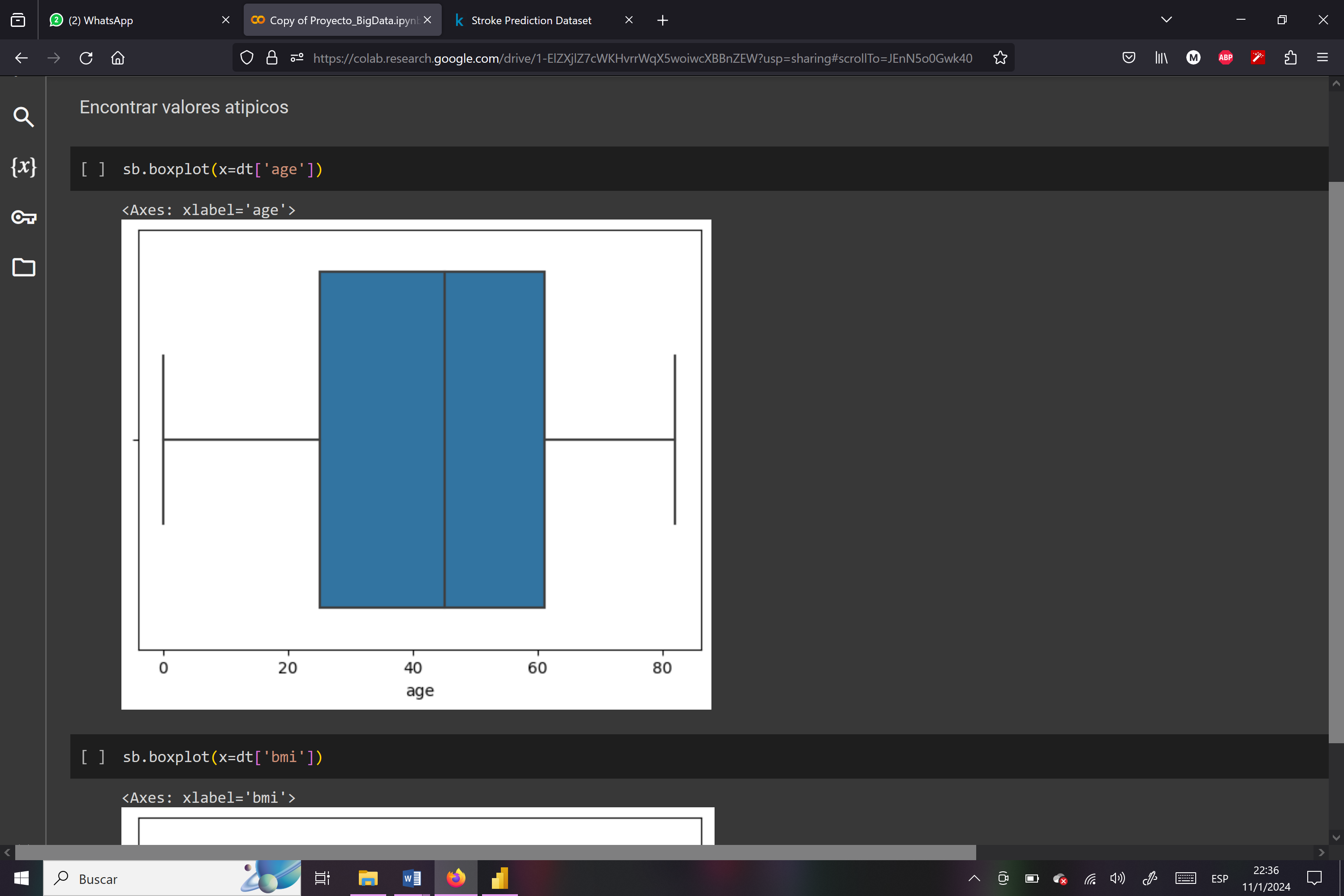


Gráfico 1. Promedio de edad.

Asimismo, el promedio de BMI (Body Max Index) de las personas entrevistadas es de 30,21 tal como se ve en el siguiente gráfico:

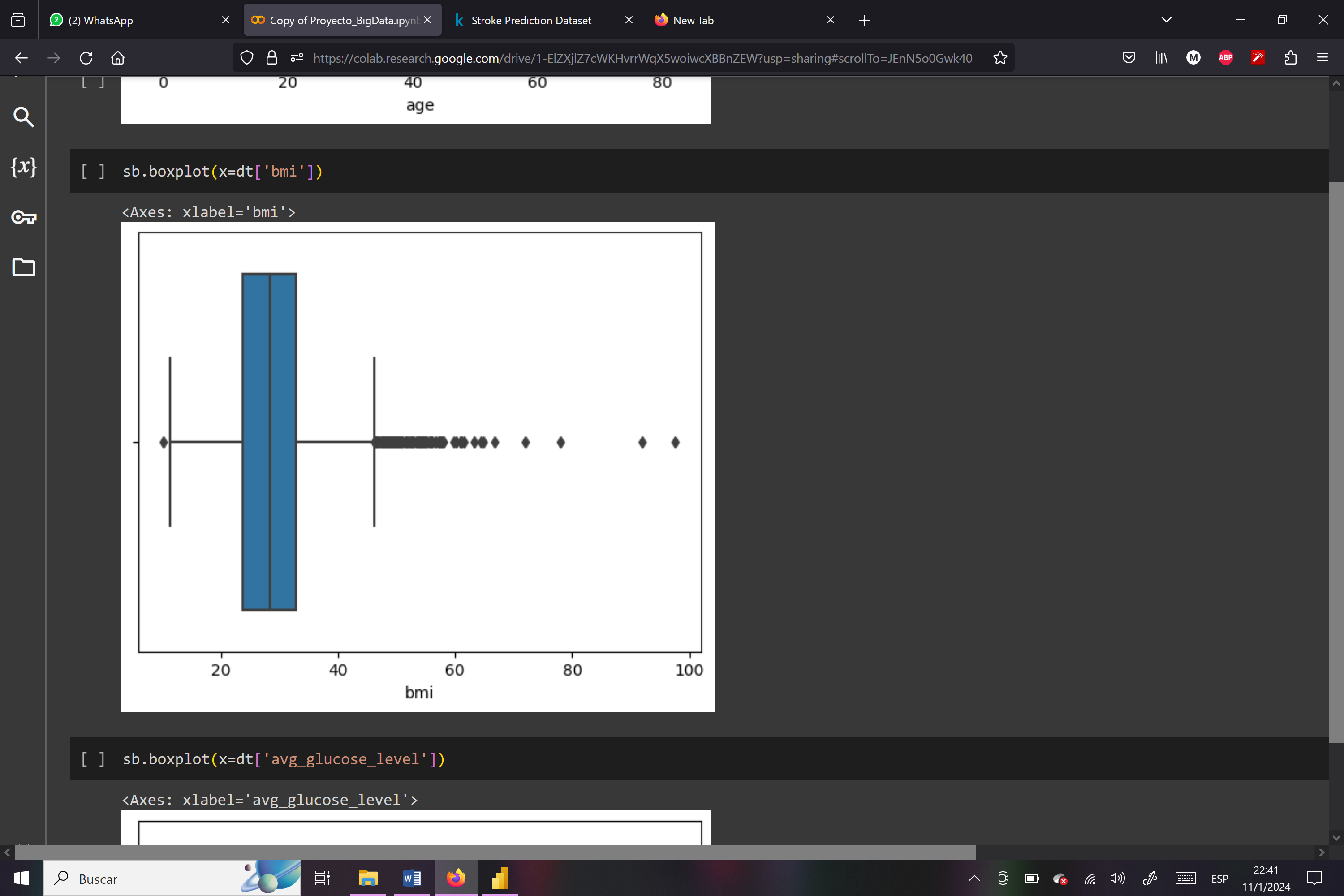
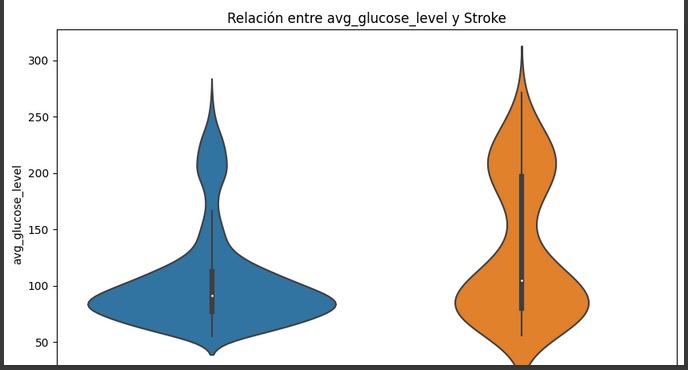


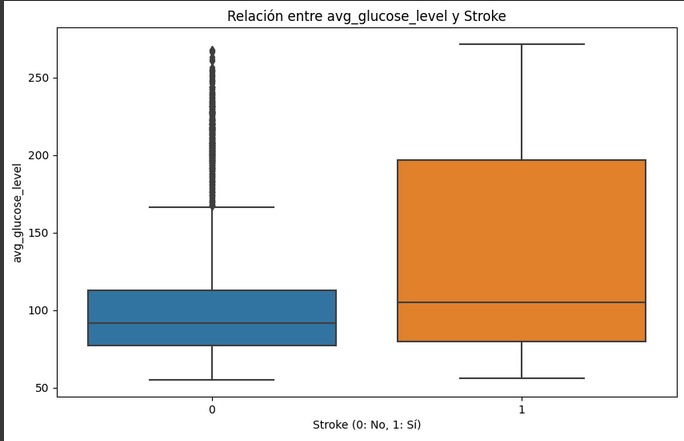
Gráfico 2. Promedio de BMI.

**Preprocesamiento de los datos para la construcción de un modelo de predicción**

Se realizaron también otros gráficos que muestran la relación de otras condiciones que inciden en los ataques cerebrovasculares, como por ejemplo los siguientes gráficos muestra la relación entre los niveles de glucosa y los ataques cerebrovasculares.

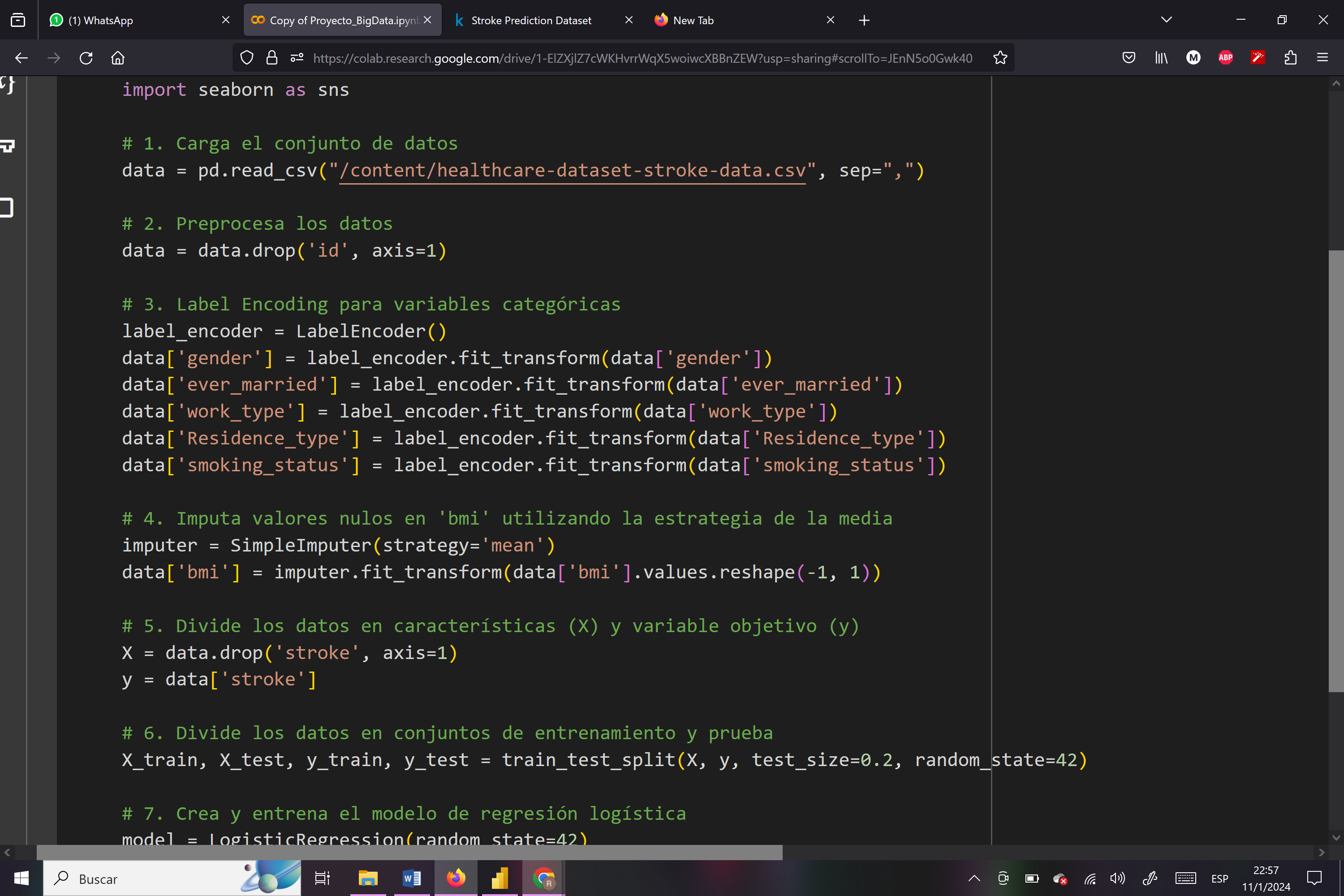
En este caso se usaron dos gráficos, uno tipo caja y bigotes y otro tipo violín para ejemplificarlo. También se debe tener en cuenta que se usaron dos grupos de muestra.





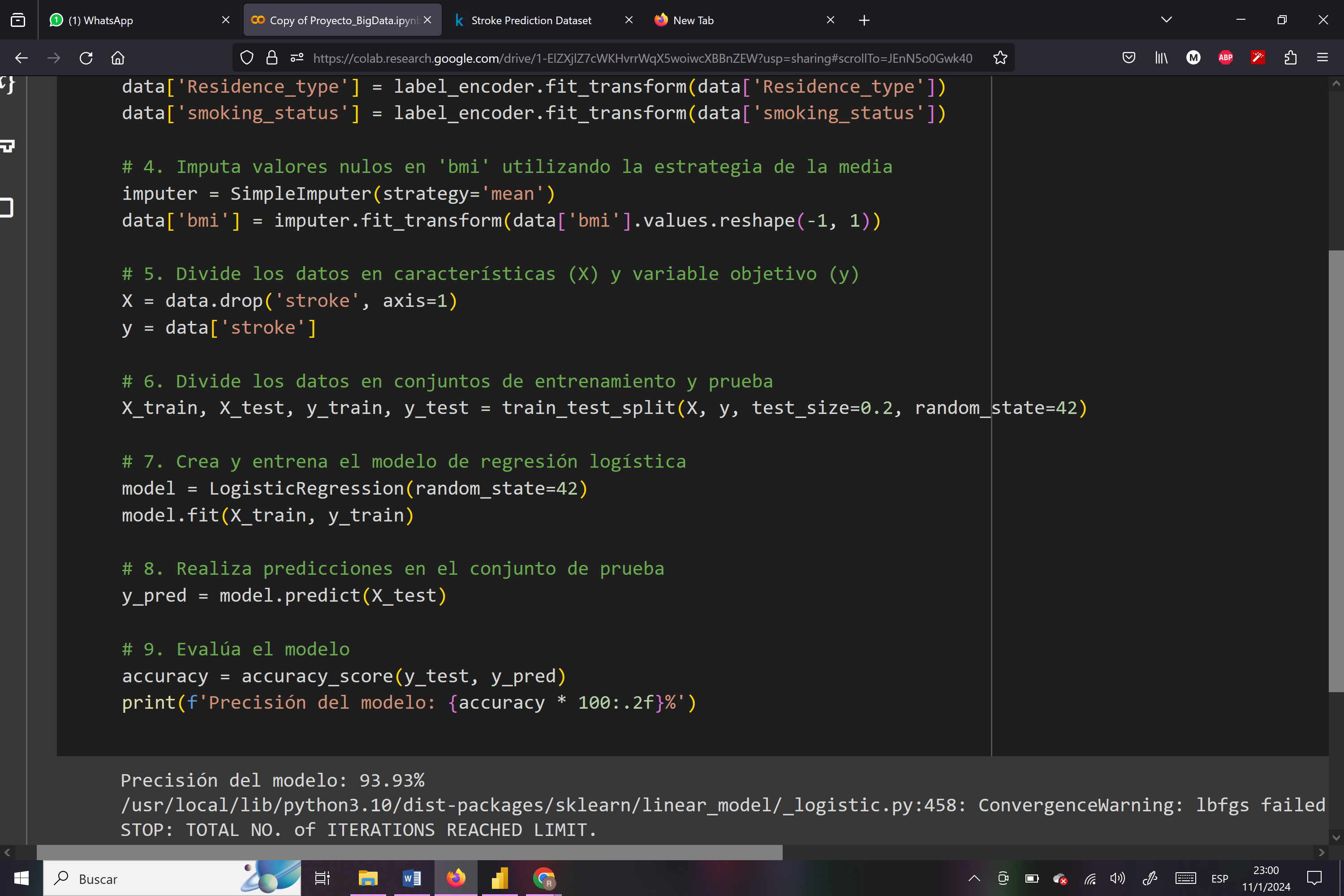
Se puede observar que la correlación si bien no es alta, existe y se podría decir que es significativa.

Ahora bien, acerca del modelo empleado para poder ver la predicción sobre los ataques cerebrovasculares, se utilizó un modelo de regresión lineal, pero para poder usarlo primero se utilizó el método de Label Encoding para que las variables categóricas fueran más fáciles de usar por el modelo.

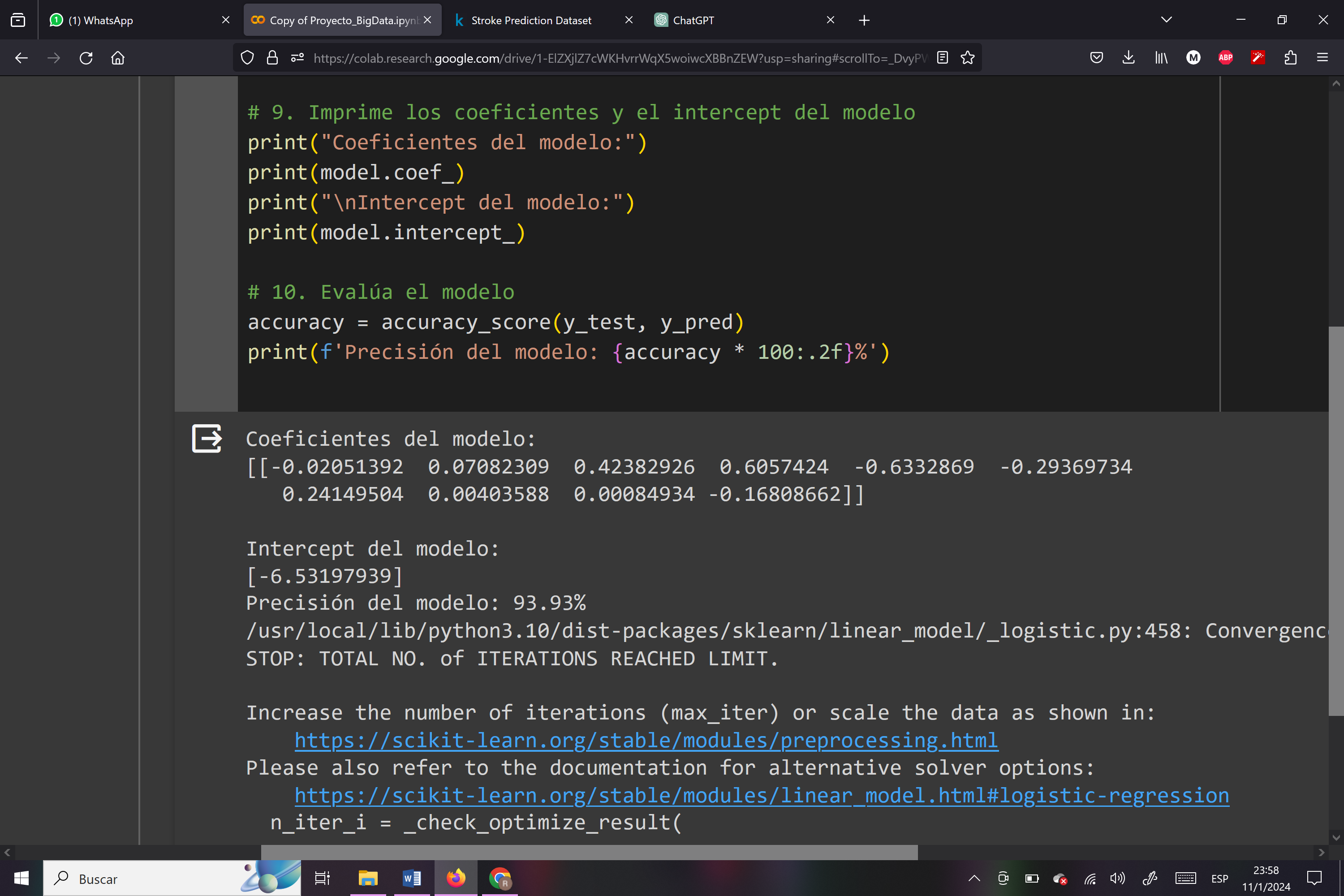


**Construcción del modelo de predicción**

Se dividieron los datos en conjuntos de entrenamiento y prueba y se crea el modelo de predicción.



Se decidió utilizar los coeficientes para poder observar la relación entre las diferentes causas y los ataques cerebrovasculares, tal como se puede observar:

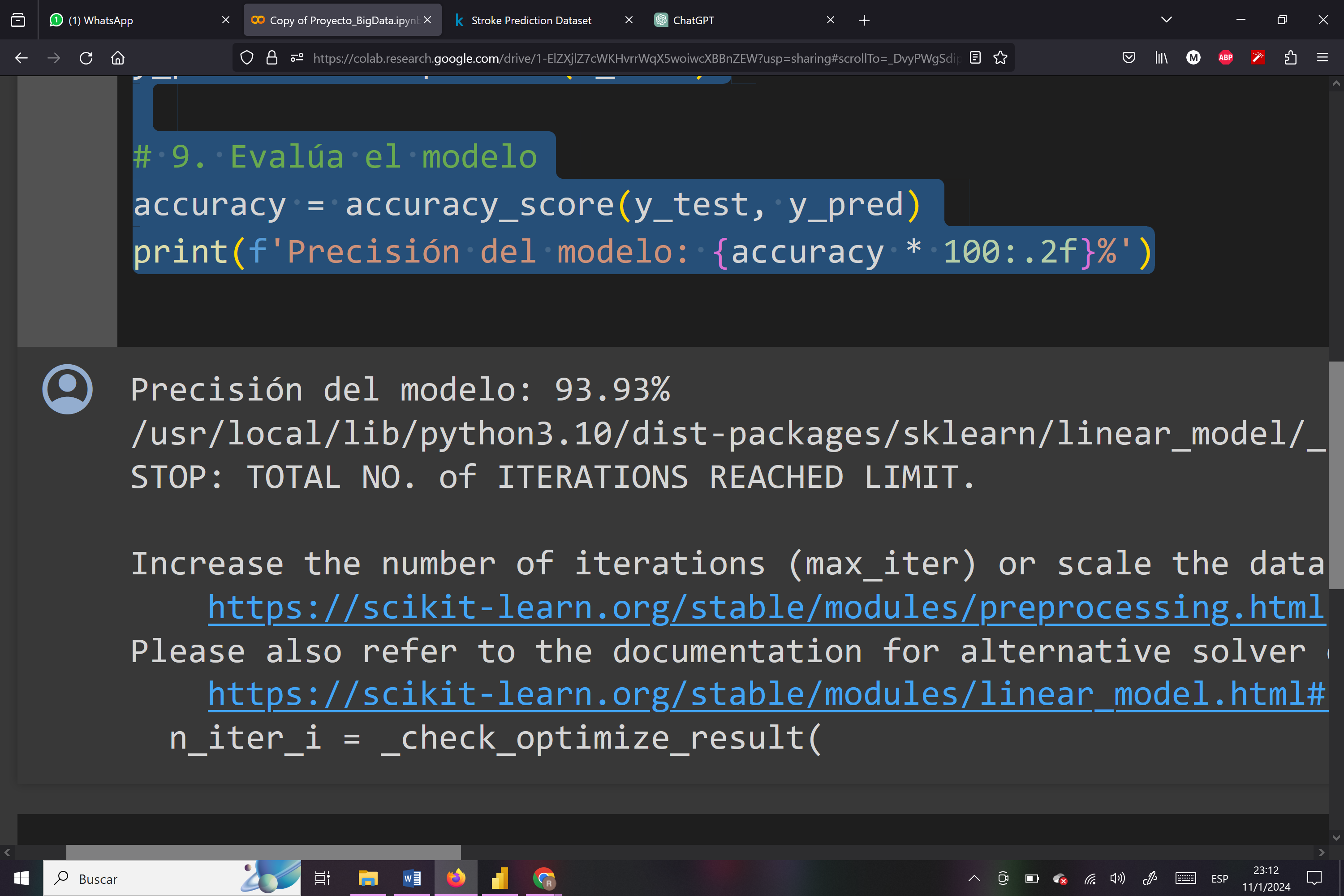


En este caso, cada número está relacionado a la característica o columna. Por lo que va de esta manera:

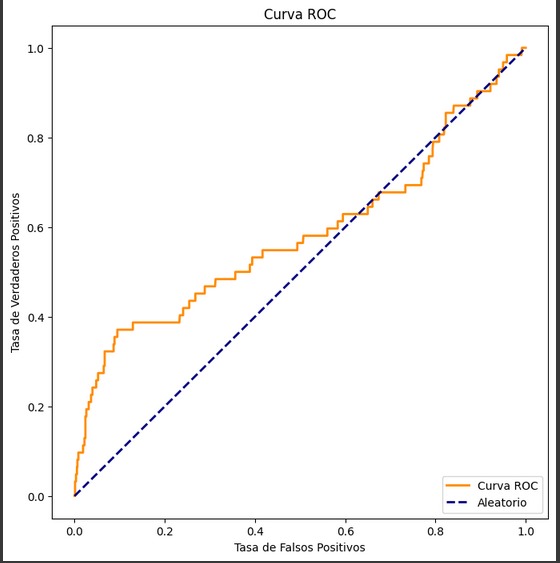
* Género: -0.02051392.
* Edad: 0.07082309.
* Hipertensión: 0.42382926.
* Enfermedades del corazón: 0.6057424.
* Casado: -0.6332869.
* Tipo de trabajo: -0.29369734.
* Tipo de residencia: 0.24149504.
* Promedio de los niveles de glucosa: 0.00403588.
* BMI: 0.00084934.
* Estatus de fumador: -0.16808662.

Se debe de tener en cuenta que estos coeficientes al ser de un modelo de regresión lineal por lo que si el coeficiente está cerca de 0, ya sea positivo o negativo, significa que esa variable tiene un impacto menor a otras que estén más alejadas del 0.

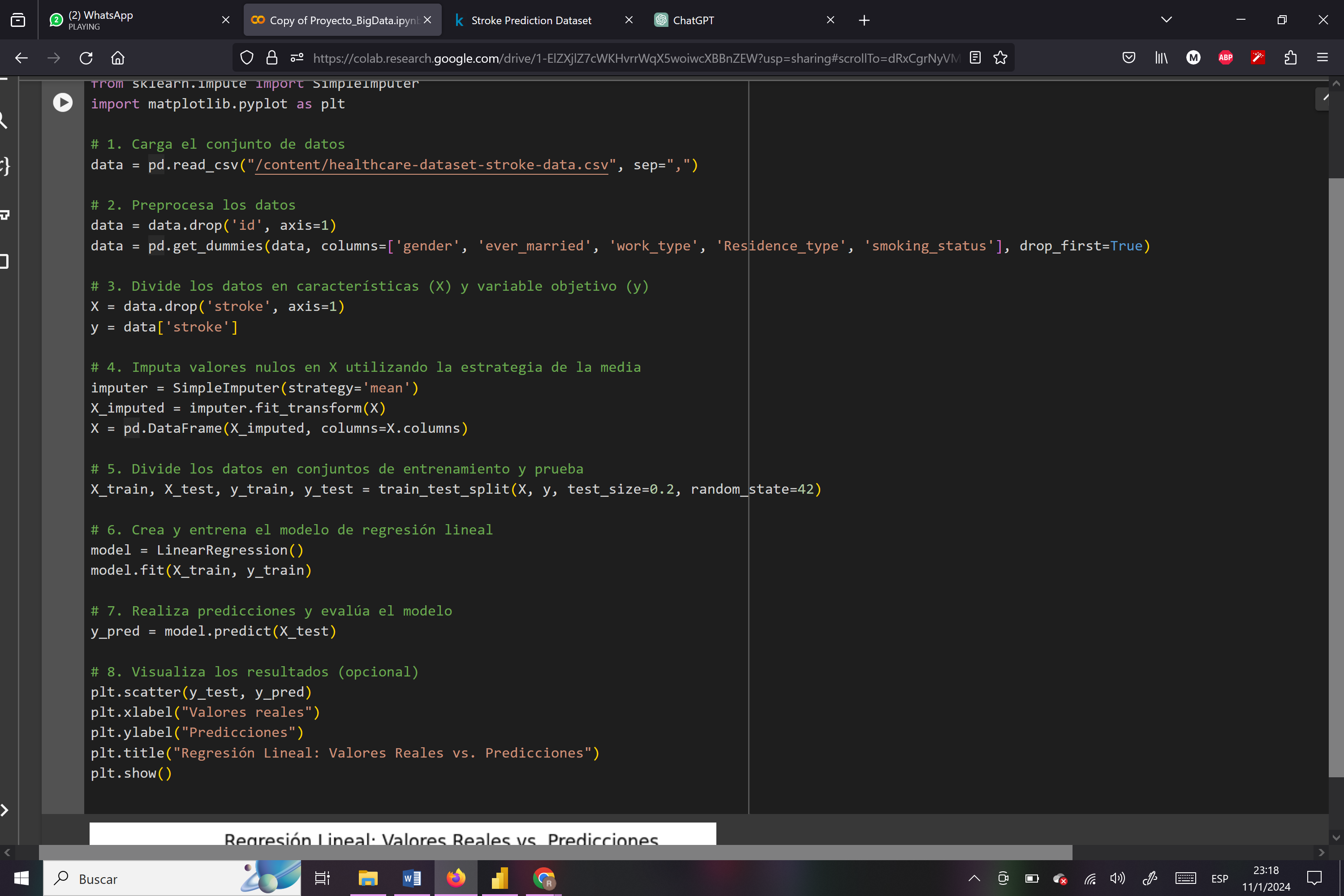
Sobre el modelo, se obtuvo un 93,93% de precisión sobre el modelo, dando a entender que el modelo fue exitoso.



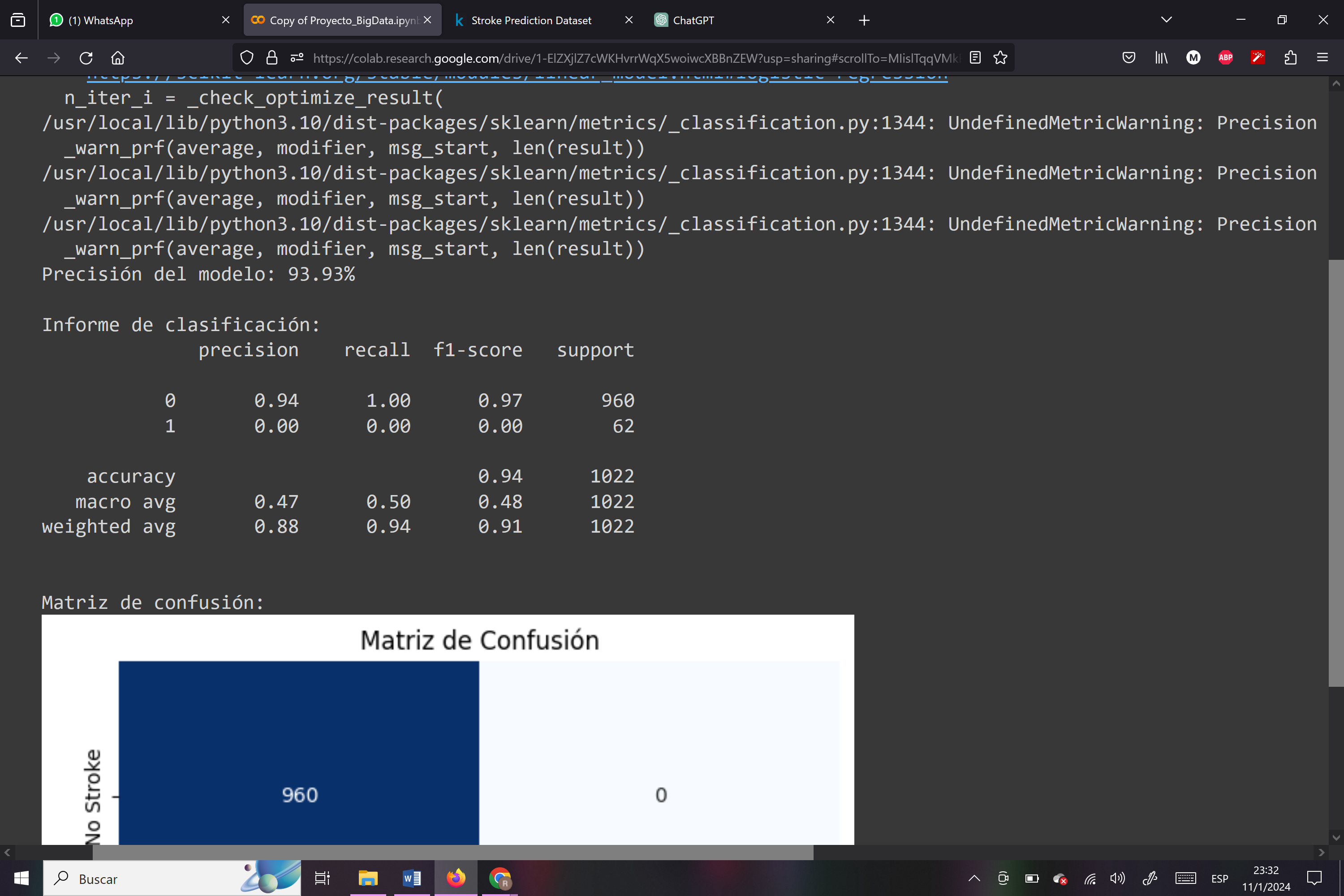
Asimismo, se utilizó un gráfico de la curva ROC (curva de eficacia diagnóstica) ya que este se utiliza para determinar la exactitud diagnóstica del test con la finalidad de determinar el punto de corte más adecuado y evalúa la capacidad discriminativa del test.



Se realizó también otro modelo de regresión lineal no obstante los resultados obtenidos no fueron satisfactorios, tal como se ve en la siguiente imagen:

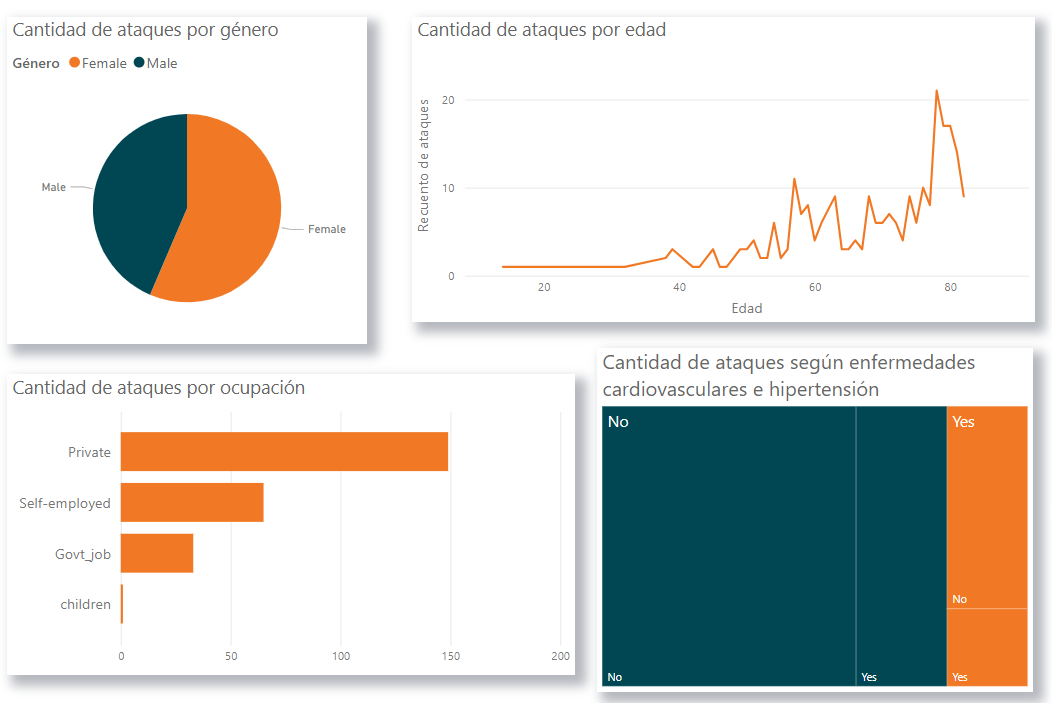


Finalmente, también se creó una matriz de confusión con el fin de evaluar el rendimiento del modelo. En este caso, se obtuvo la precisión del modelo para la clase 0 de 94%, pero por otro lado la clase 1 tiene una precisión de 0. El caso con el recall es el mismo, mientras que para la clase 0 la precisión fue perfecta, para la clase 1 esta fue errónea. El F-1 para la clase 0 indica buena precisión y recall, mientras que para la clase 1 fue todo lo contrario.

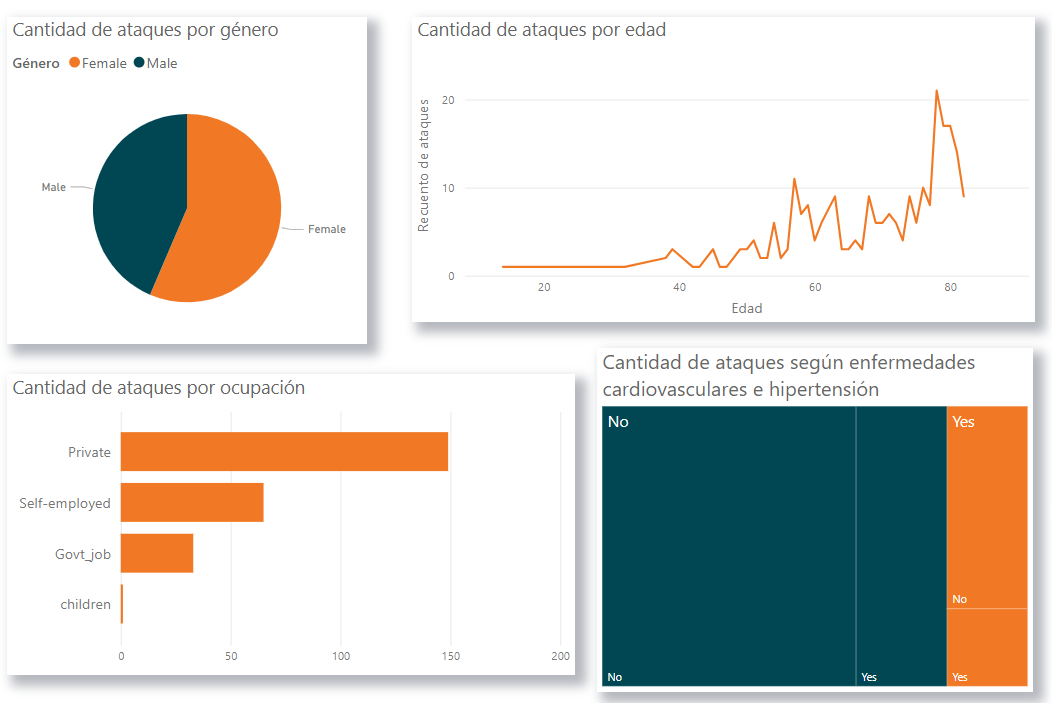


**Visualización de resultados**

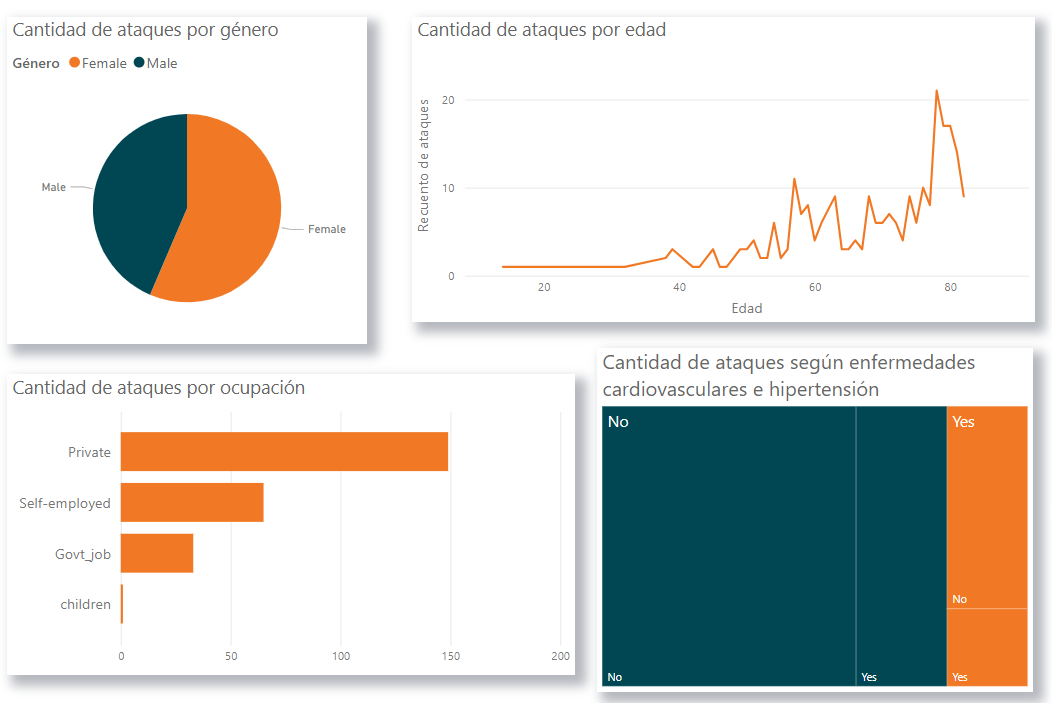
Se acudió a la herramienta de Power BI para la representación gráfica de los resultados obtenidos y la visualización más clara de las tendencias en cuanto a accidentes cerebrovasculares, tomando únicamente la muestra de los datos correspondientes a aquellas personas entrevistadas que sí han sufrido un ataque cerebrovascular. Una primera observación en cuanto a la distribución por género muestra que este tipo de accidentes son más recurrentes dentro de la población femenina.

****

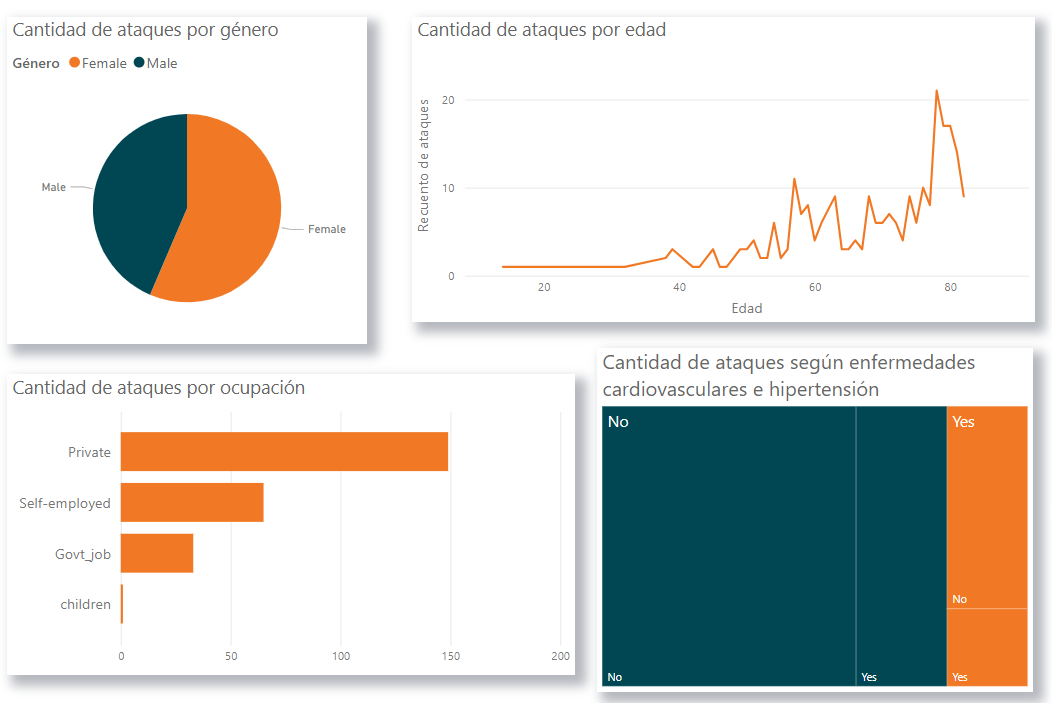
Otro aspecto de interés recayó en observar el comportamiento según un criterio de edad, donde se obtuvo como resultado que a partir de los 40 años hay una tendencia creciente en cuanto a la posibilidad de sufrir un accidente cerebral, con un pico importante en la edad de 78 años.

****

A su vez, se observó la tendencia según la ocupación de los entrevistados, con una clara superioridad de casos en aquellas personas dedicadas al sector privado; lo cual, para estudios posteriores podría ser interesante profundizar en cuanto a una correlación con el fenómeno del ¨burnout¨, el cual es más común dentro de la empresa privada.

****

Por último, se consideró importante observar la correlación entre la tenencia de antecedentes de enfermedades cardiovasculares e hipertensión con la posibilidad de sufrir un ataque cerebrovascular, a partir de lo cual se observó que la proporción prevaleciente de accidentes cerebrovasculares corresponde a personas que no han padecido ninguna de las enfermedades en cuestión, seguida de aquellos que padecen únicamente hipertensión pero no enfermedades cardiovasculares.

****

**Conclusiones**

A partir del procesamiento de los datos realizado mediante los pasos expuestos a lo largo de este informe, se obtuvieron las siguientes conclusiones principales del análisis del comportamiento de los datos:

* Del total de personas entrevistadas, únicamente un 4,96% ha sufrido algún tipo de accidente cerebrovascular en su vida.
* Los accidentes cerebrovasculares tienen una predominancia dentro de la población femenina.
* Después de los 40 años aumenta considerablemente el riesgo de sufrir un accidente cerebrovascular, con un pico alrededor de los 78 años de edad.
* La tendencia de ataques cerebrovasculares por ocupación es mayoritaria en los casos de personas que se dedican al sector privado; mientras que, aquellos que se dedican a labores de cuido familiares (específicamente niños) constituyen el grupo sobre el cual se observa una menor recurrencia de ataques.
* La presencia de antecedentes de enfermedades cardiovasculares e hipertensión no son representativos en cuanto a la posibilidad de sufrir un accidente cerebrovascular, pues de la muestra tomada, la mayoría de quienes sufrieron un ataque no poseían ninguna de estas enfermedades en su historial.