**Mentoría - Diplomatura en Ciencia de Datos y Aprendizaje Automático**

Modelo de recomendación de profesionales médicos para una mejor experiencia de usuarios

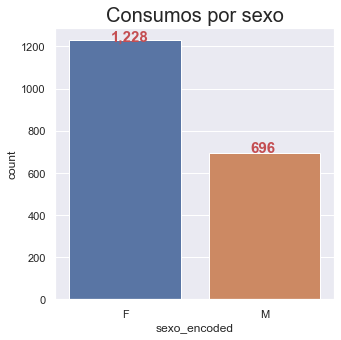
**Análisis de demanda**

1. Características de los pacientes y relaciones

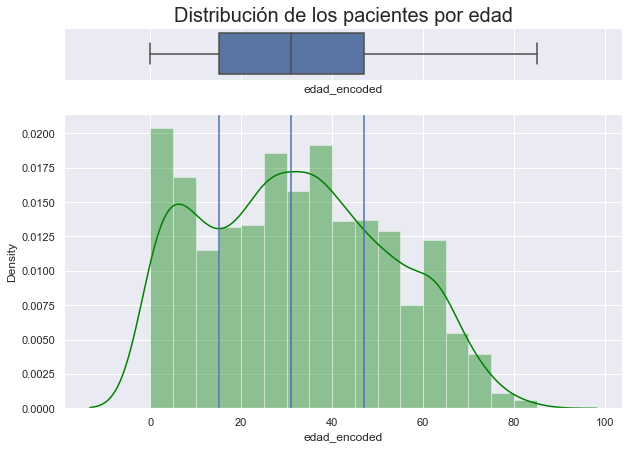
Se han analizado el grupo de variables que se relacionan con las características propias de los consumidores y las generadas a través de su vínculo con el prestador de salud.

Cada registro del dataset original se corresponde con un paciente individual y no existen duplicados.

El dataset cuenta con 1924 registros de los cuales 1228 corresponden a individuos de sexo femenino, por lo cual el dataset se encuentra desbalanceado en relación a esta variable.



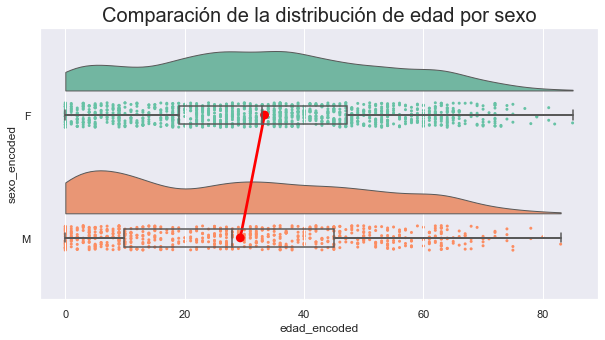
En cuanto a la edad de los socios, esta variable va desde 0 a 85 años, presentando claramente una asimetría sesgada hacia la derecha. La forma multimodal de la distribución nos permite apreciar rangos de grupos de edad que nos podrían servir en el futuro para categorizar esta variable.



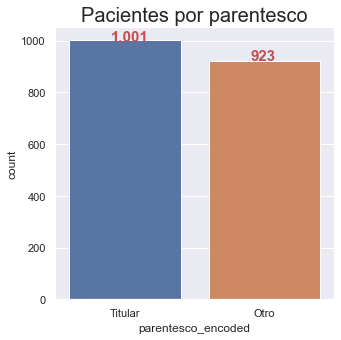
El gráfico anterior nos muestra la distribución teórica de la distribución y nos permite detectar una moda y otras dos modas relativas.

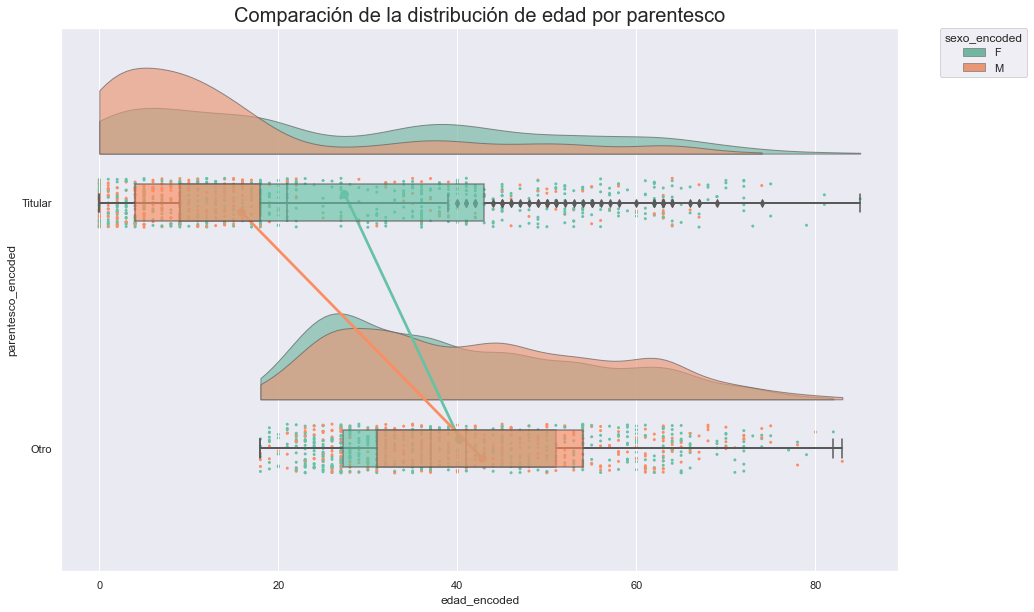
Claramente existen una gran cantidad de asociados recién nacidos y pediátricos, motivados por la necesidad de asistencia médica de calidad en los primeros años de vida. Esta cantidad disminuye para individuos entre los 10 y 25 años para aumentar luego en la mediana edad, probablemente por el deterioro en el estado de salud en general.

Combinando estas dos variables, podemos calcular el intervalo de confianza para la diferencia de medias y como dicho intervalo no incluye el cero, podemos asegurar con una probabilidad del 95% que las medias son distintas.

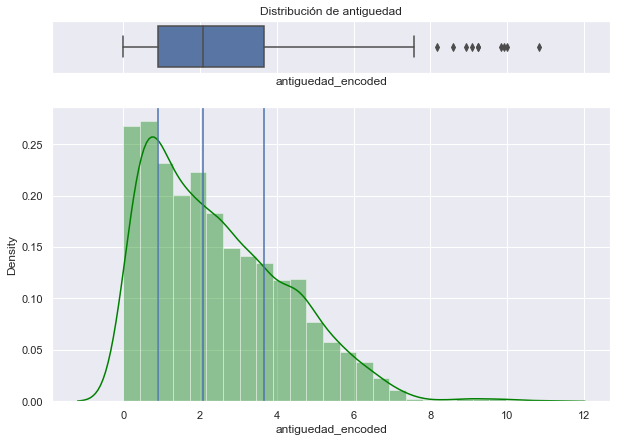


En cuanto a la variable Parentesco, se puede observar que no existe una gran diferencia entre la cantidad de socios titulares y otros. Relacionándola con la edad, se puede observar que la cantidad de afiliados del parentesco “otro” incluye a individuos a partir de los 17 años aproximadamente, mientras que es posible apreciar que el grupo “titular” comienza a disminuir a medida que aumenta la edad. Existen más pacientes de sexo masculino de temprana edad que de sexo femenino.

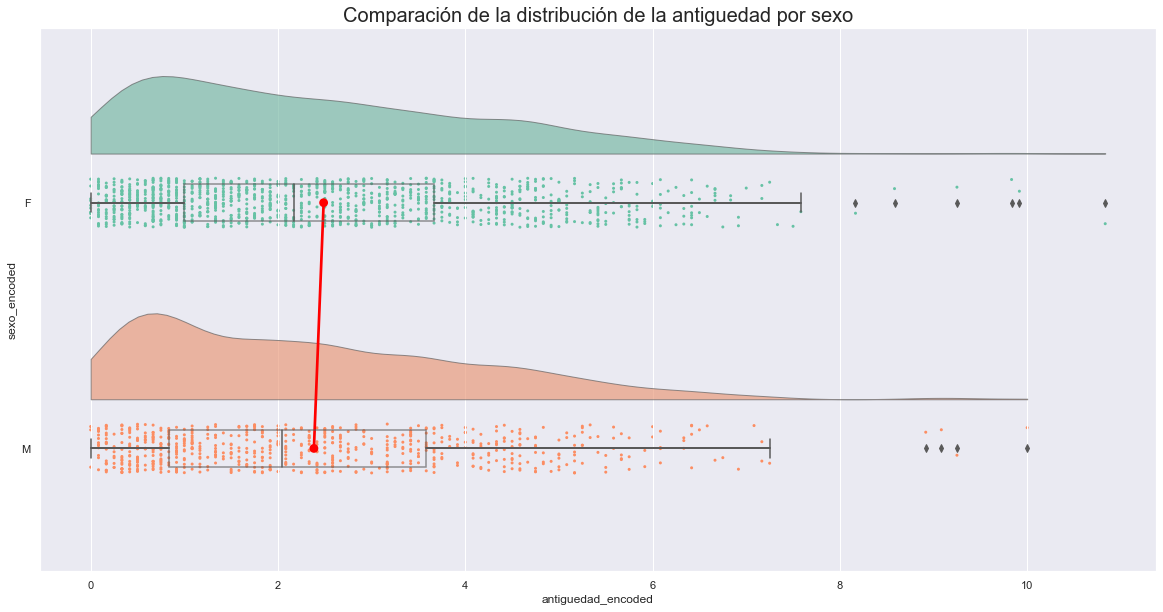




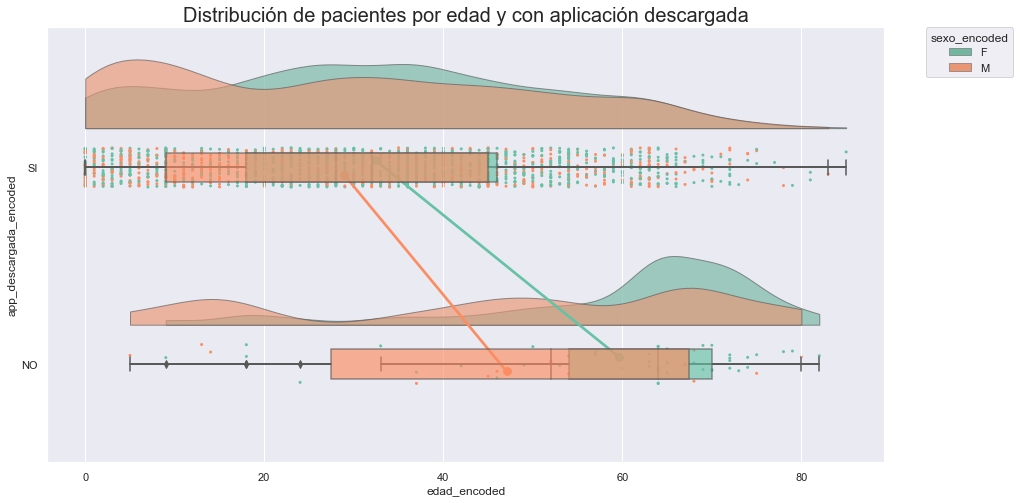
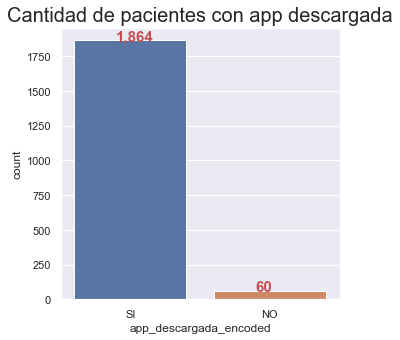
La variable antigüedad nos dice que la mayoría de los pacientes del dataset se afiliaron dentro del último año. Como es esperable, esta cantidad va disminuyendo casi de forma lineal a medida que aumenta la cantidad de años de afiliación.



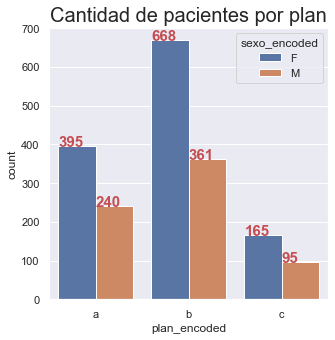
Relacionándola con la variable sexo, no se encuentran diferencias significativas en la antigüedad por sexo, siendo las medias de ambas distribuciones a simple vista casi iguales.

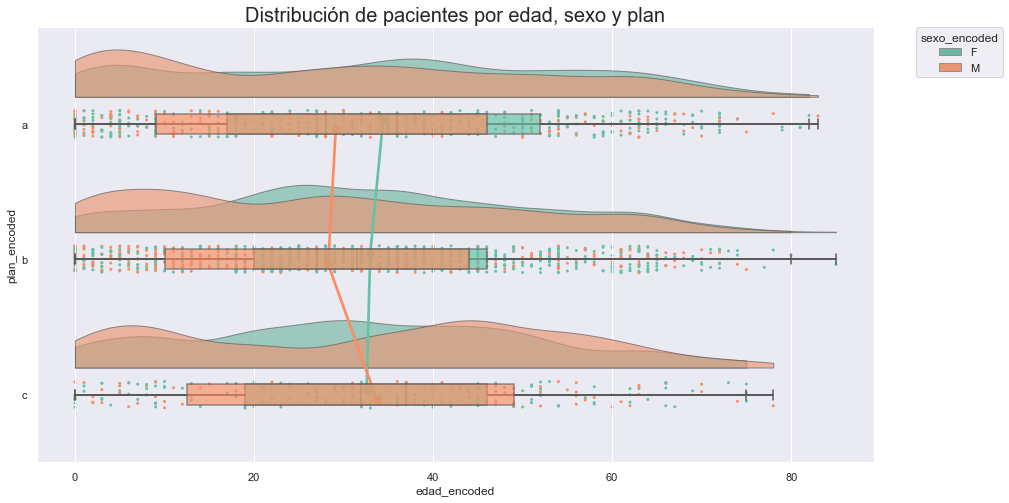


Se observa que casi la totalidad de los pacientes tiene la aplicación descargada, habiendo solamente una cantidad mínima (60 socios) que no cuentan con ella. Este grupo corresponde en mayor medida a pacientes de elevada edad, de sexo femenino.



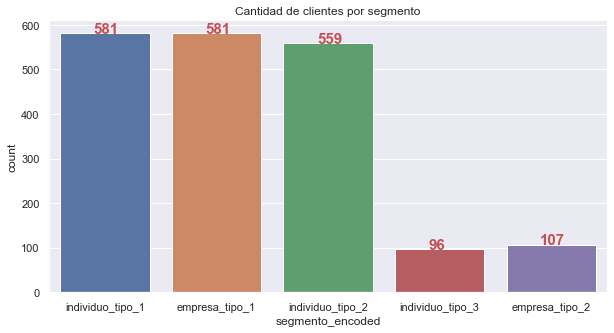
En el dataset se indica que los asociados pueden pertenecer a tres planes (a, b o c) siendo el b el que más afiliados tiene. En los grupos se respeta la proporción por sexo observada en el análisis inicial. El plan c es donde menos individuos se contabilizan, tal vez por ser el plan más costoso.

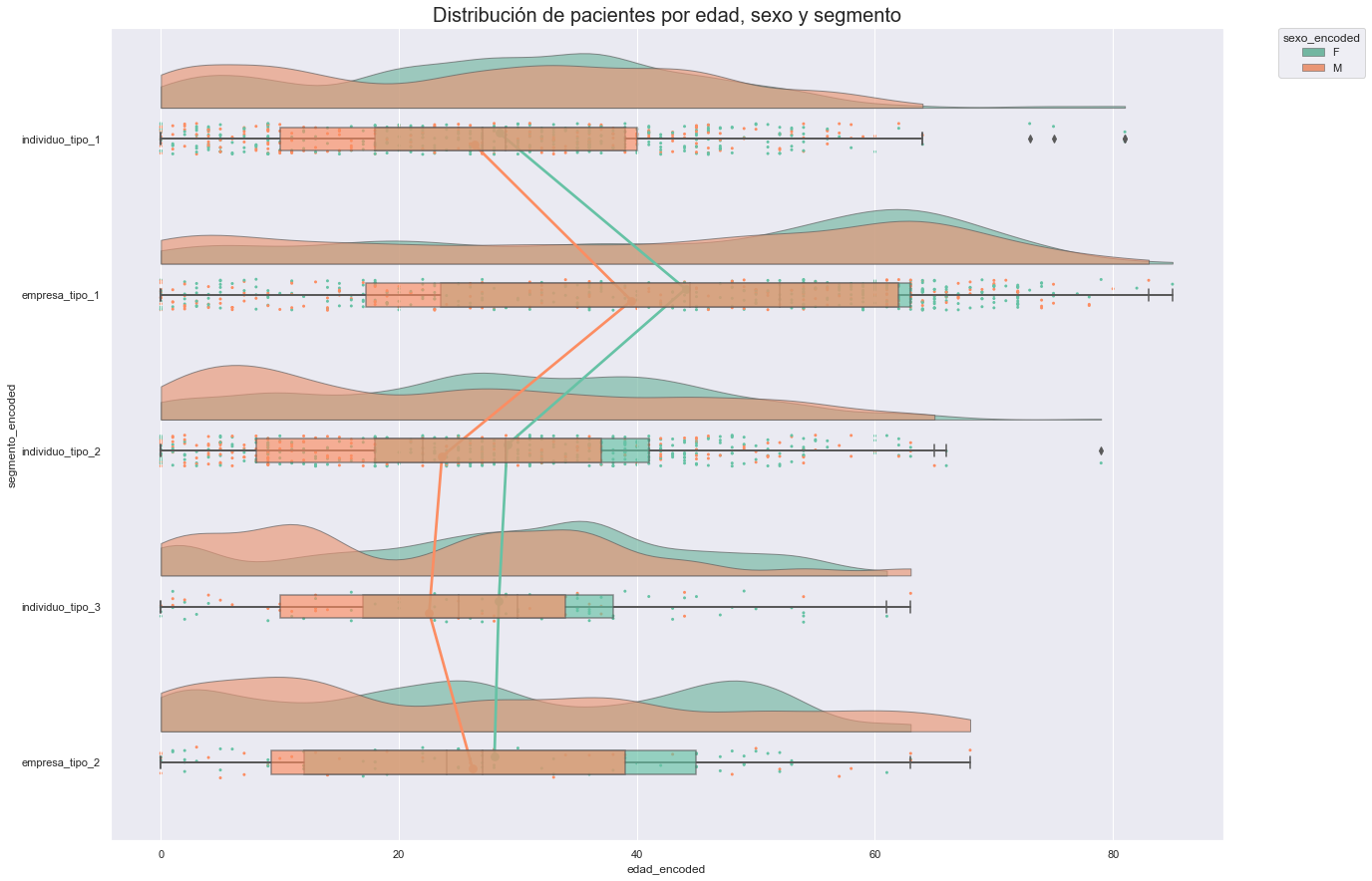




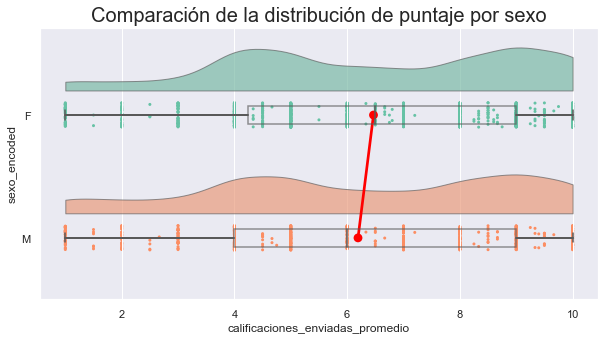
No se observan diferencias significativas en la distribución de estos tres grupos por edad y por sexo. Se realizar algunas observaciones menores como:

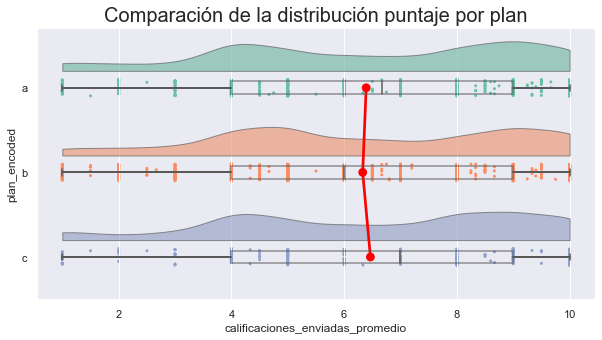
* El plan b es elegido en los primeros años de vida en mayor medida por individuos de sexo masculino.
* La disminución en la cantidad de afiliados entre los 15 y 25 años es más pronunciada en los hombres que en las mujeres.





En cuanto al análisis por segmentos, se puede observar claramente que en todos los casos los afiliados de sexo femenino tienen mayor edad. El segmento empresa\_tipo\_1 es el mayoritariamente elegido por los pacientes de elevada edad, ya que se puede apreciar cómo la cantidad de pacientes en el resto de segmentos disminuye a medida que aumenta edad\_encoded, mientras que en el segmento de interés ocurre lo contrario.





Se observa que la distribución de la puntuación por sexo en ambos casos es prácticamente la misma, siendo los afiliados de sexo femenino los que mejor devolución del servicio han dado.

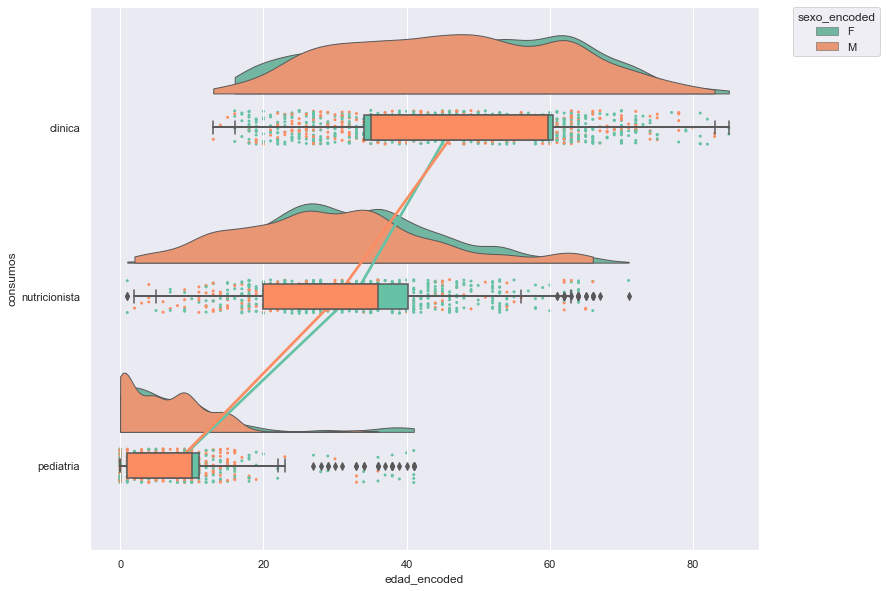
Resultaría interesante poder categorizar esta variable en malo – regular- bueno o muy bueno.

Aparentemente, a pesar de ser el grupo minoritario, el plan c es el que mejor puntuación tiene. Sin embargo, es importante aclarar que la diferencia en la media de calificaciones por plan no varía de manera considerable.

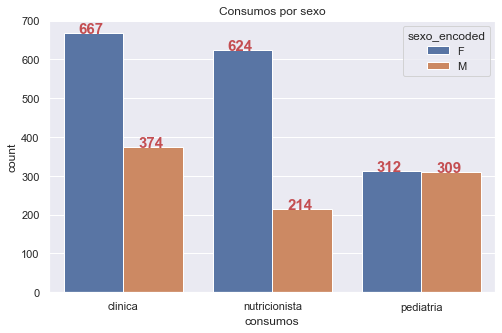
El puntaje es una de las variables sobre las cuales se debe centrar nuestro modelo de recomendación y nos permitirá enfocarnos en la mejora de un determinado grupo.

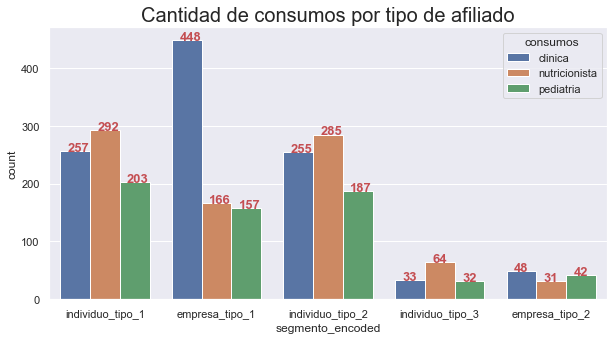
1. Consumos

En este caso se han evaluado los consumos de los afiliados para los servicios Pediatría, Nutrición y Clínica Médica.

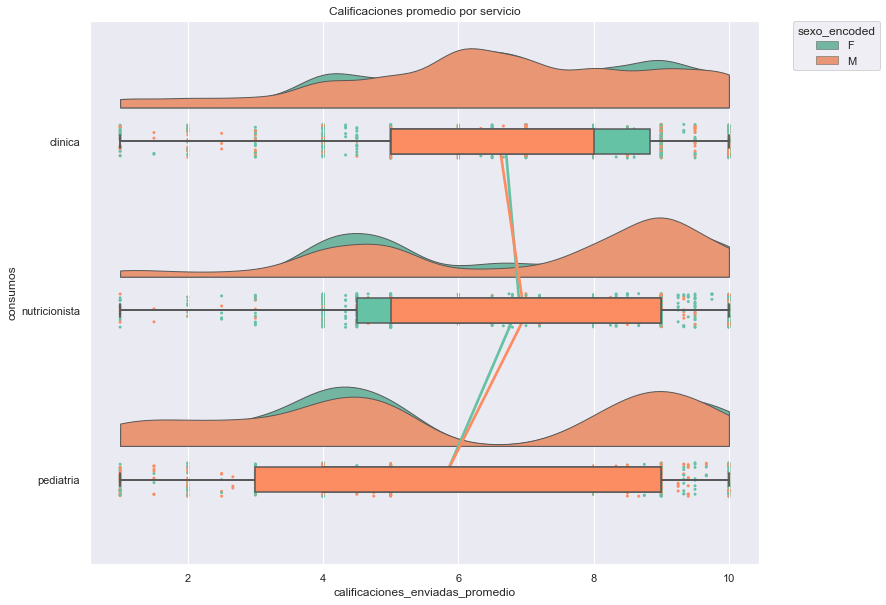


Es interesante observar cómo el servicio de nutricionista es utilizado en casi todo el rango de edad, mientras que los consumos en clínica médica empiezan desde los 13 años y se extiende durante el rango restante. En cuanto al servicio de pediatría, existen algunos consumos de pacientes de edad en torno a los 40 años, que hay que ver si se los trata como outliers.



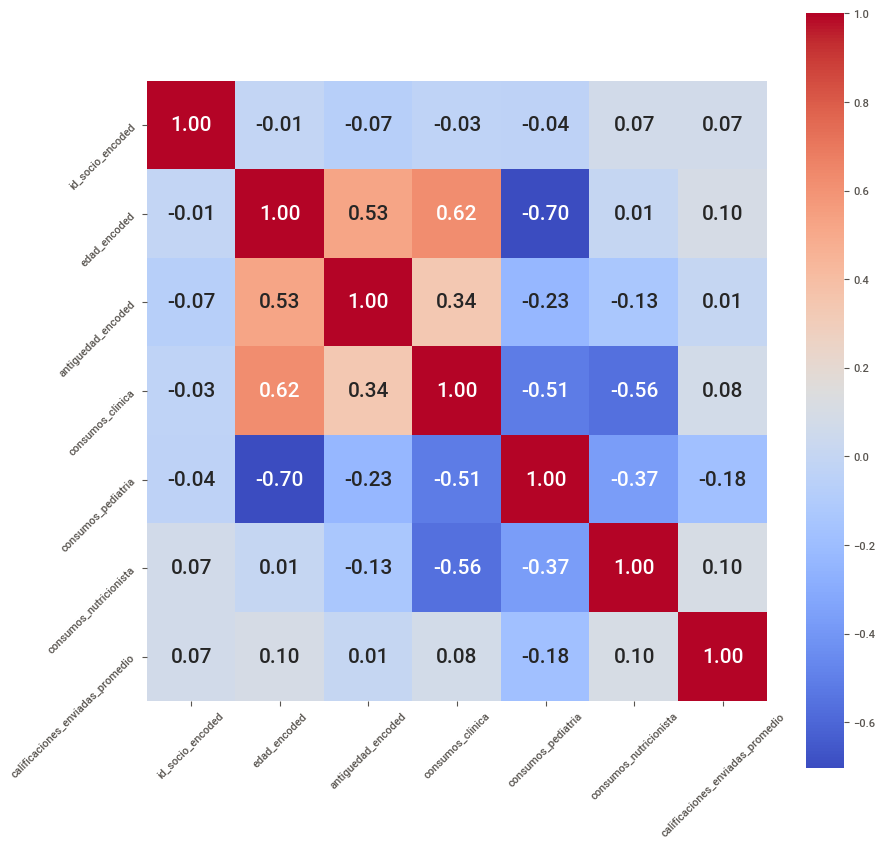


El segmento empresa\_tipo\_1 es el que más consumos de clínica médica tiene. Anteriormente se había observado que este tipo de afiliación era el que más caracteriza a los asociados de edad más avanzada.



El servicio de nutrición es donde observa que el promedio del puntaje otorgado es mejor. Se observan una mayor cantidad de calificaciones negativas para el servicio de pediatría.

**Gráfico de correlaciones**



En el gráfico de correlaciones es posible observar las relaciones positivas y negativas entre cada una de las variables. A simple vista:

* Edad tiene una correlación positiva con antigüedad y con consumos en clínica médica. Esto pudo ser comprobado en los análisis de la distribución de consumos por edad.
* Edad tiene una correlación negativa con consumos\_pediatría. Esto también ha sido observado al analizar las distribuciones.
* Edad y consumos\_nutricionista no presenta correlación. Esto también había sido analizado debido a que era el servicio que prevalecía en casi todo el rango de edades.