



Informe Ejecutivo Prueba Senior Data Scientist

GLOBANT

Unidad de Ciencia de Datos

Daniel Jiménez M.

SEPTIEMBRE - 2020

COLOMBIA

Índice

	Página
Introducción	3
Feature Engineering e Importancia de las variables.	4

Introducción

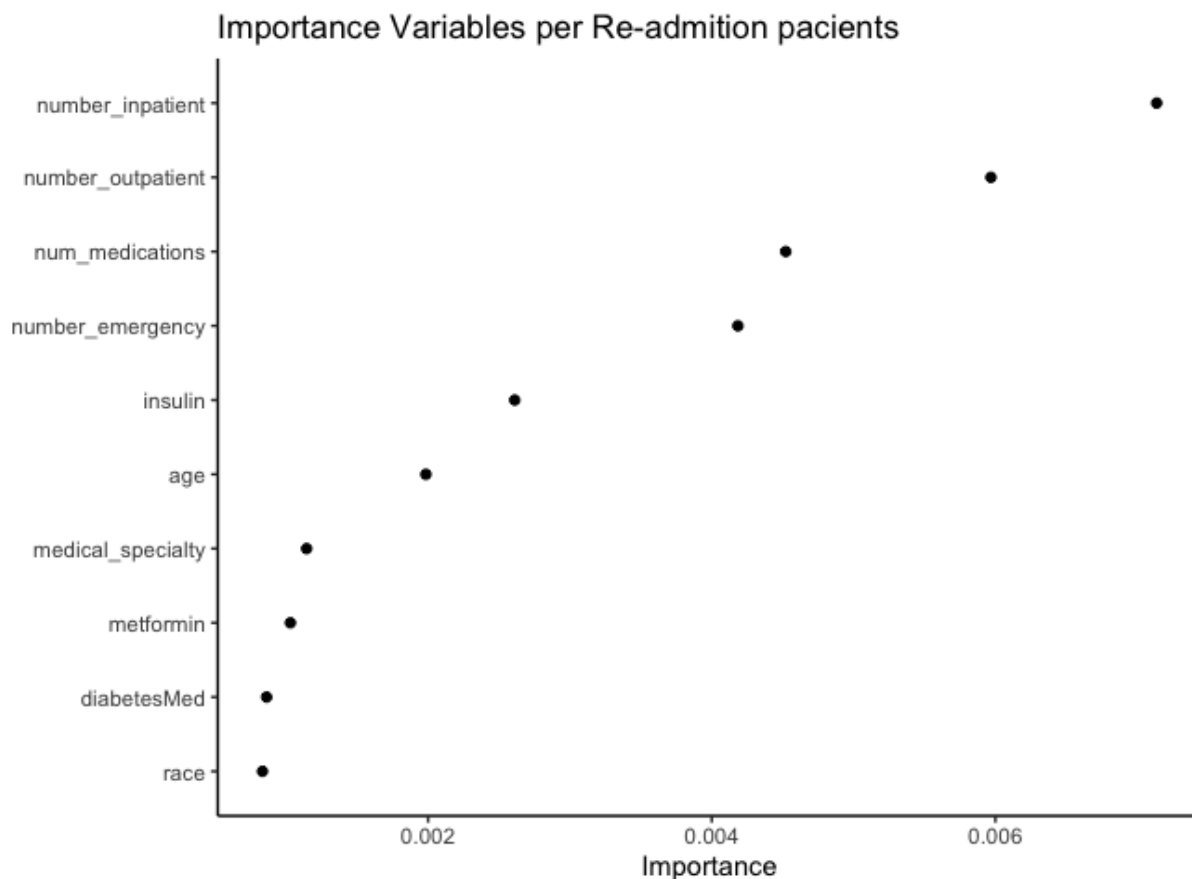
El presente documento hace parte de la ficha **ejecutiva** de la prueba de selección para la vacante de **Senior Data Scientist**, si desea ver el desarrollo Técnico, por favor haga click **aquí**¹, acá encontrará una ficha tipo git document en Inglés, con los códigos para que puedan evaluar y observar el desarrollo que implemente, por lo tanto en la medida de lo posible no se hará uso de un lenguaje **Devops**, pero en esta introducción si haré mención de lo siguiente: todo lo presentado a nivel de Exploración y Desarrollo de Insigths, esta sustentado bajo herramientas como : estadística y machine learning aplicado, modelamiento matemático, y un framework muy usado en ciencias sociales *-updown-*, donde a través de comparaciones es que se desarrollan las preguntas más oportunas de un tema, del cual se desconoce, como es el caso de este set de datos.

Dicho lo anterior, partiré al revés en el desarrollo de Insigths, con un modelo de Machine Learning para identificar las variables que explican los resultados de pacientes re-hospitalizados, con y sin diabetes.

¹El documento presentado, es desarrollado con Markdown, por lo cual los hipervinculos estaran documentados dentro de las palabras.

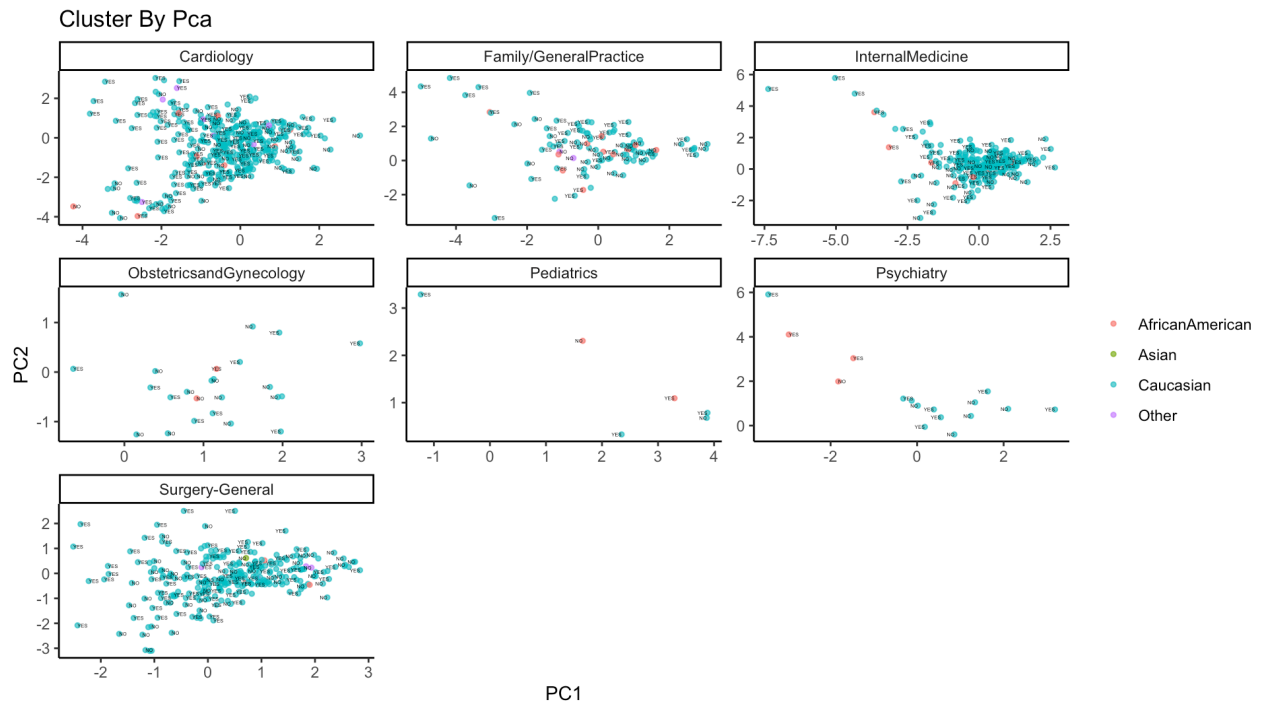
Feature Engineering e Importancia de las variables.

La importancia de las variables, se basa en un modelo en el entendimiento de cuales son las variables que generan información en este caso la reincidencia de pacientes hospitalizados bajo la premisa que muchos de ellos son pacientes de diabetes, por lo tanto en el siguiente gráfico se basa en esta premisa.



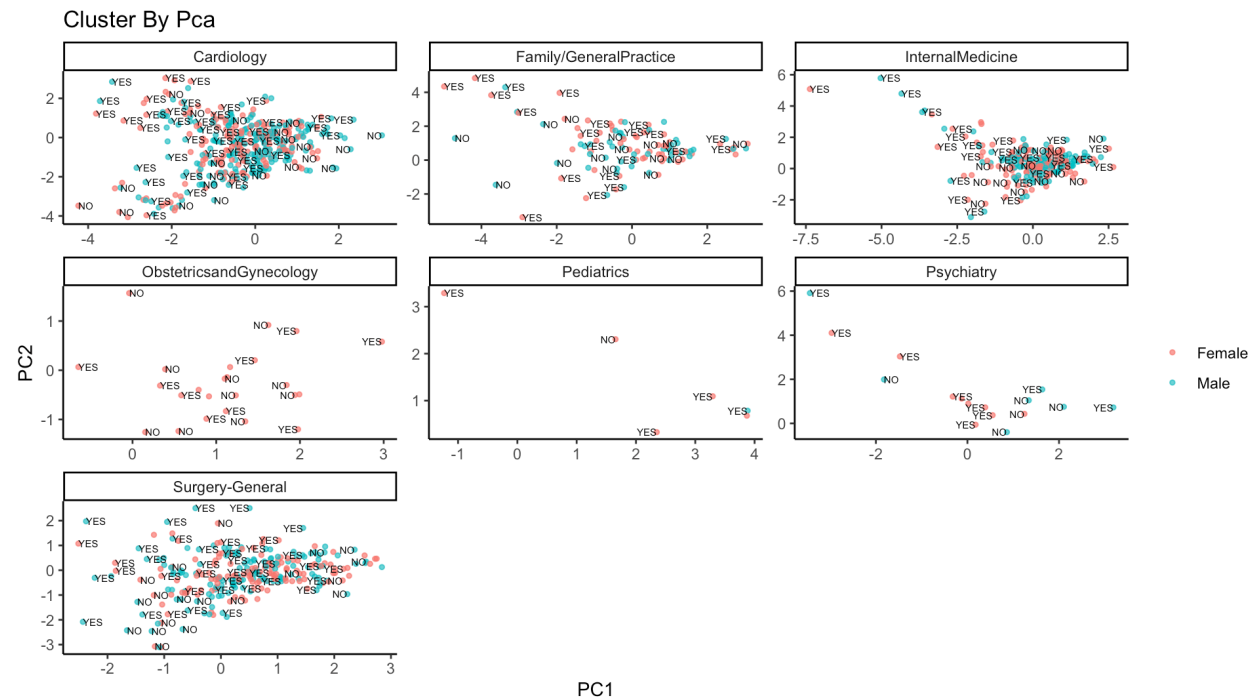
La anterior gráfico muestra que la re-incidencia de los pacientes tiene poco que ver con el tema de la diabetes (más este es el estudio de caso), pero, independiente a ellos, la raza, la edad y el genero, generan poca información sobre las re-adminsiones de pacientes, pero son factores que combinados con las variables que desde ahora se les llamara :*importantes*, si dan una vista del comportamiento clusterizado de dicho problema, como se verá a continuación.

Observación especial: Los cluster aquí presentados son desarrollados a través de de PCA's, la cual es una técnica que captura la camyor cantidad de información a través de heurísticas estadísticas.



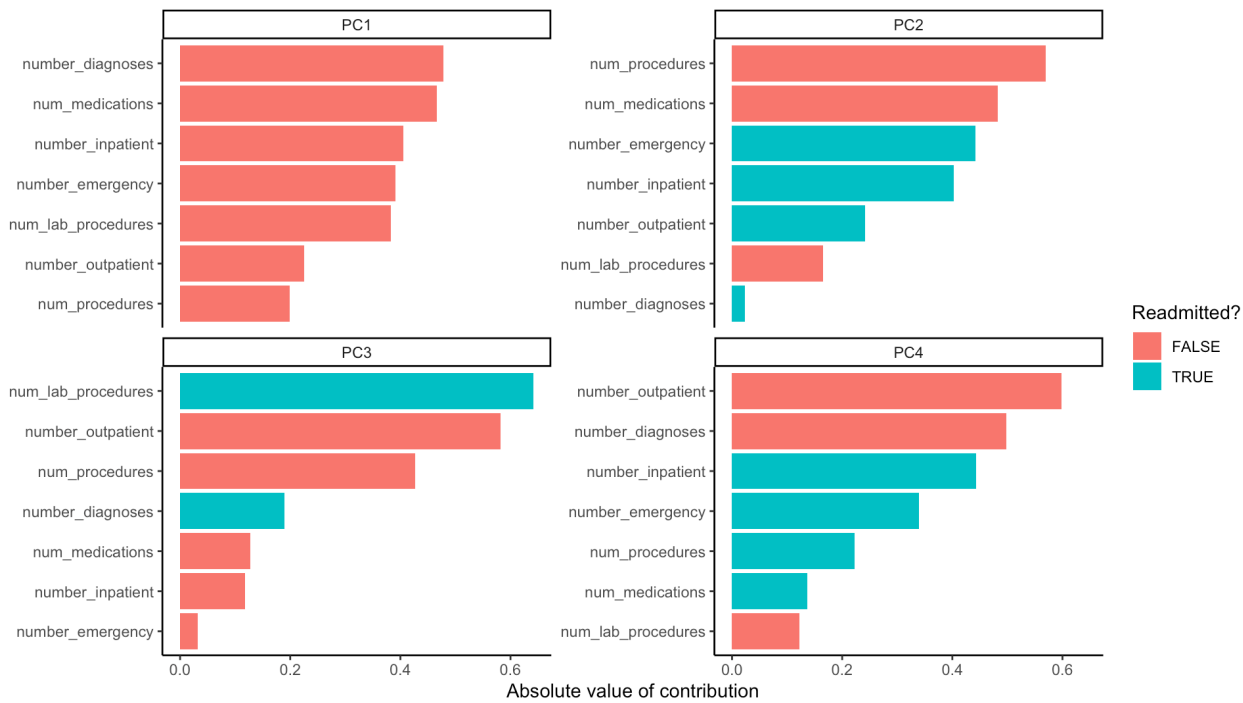
Este gráfico representa a través del cluster Raza - Tipo de Medicina, en que casos se presentan casos reincidentes de hospitalización.

Mientras que en el caso del genero parece ser que la re-admisión es más frecuente en las mujeres (independiente a su edad), en los casos de visitas al medico especializadas y el número de medicinas o tratamientos sometidas.



Para finalizar esta parte un insight a destacar es el desarrollo de los cluster a través de unas variables fijas,

las cuales denotaron en que casos se presentan re-admisiones dado un número de características, notesé que el número de pacientes hospitalizados es un feature que se mantiene, dado que la correlación del mismo con las otras variables es indudablemente robusto.



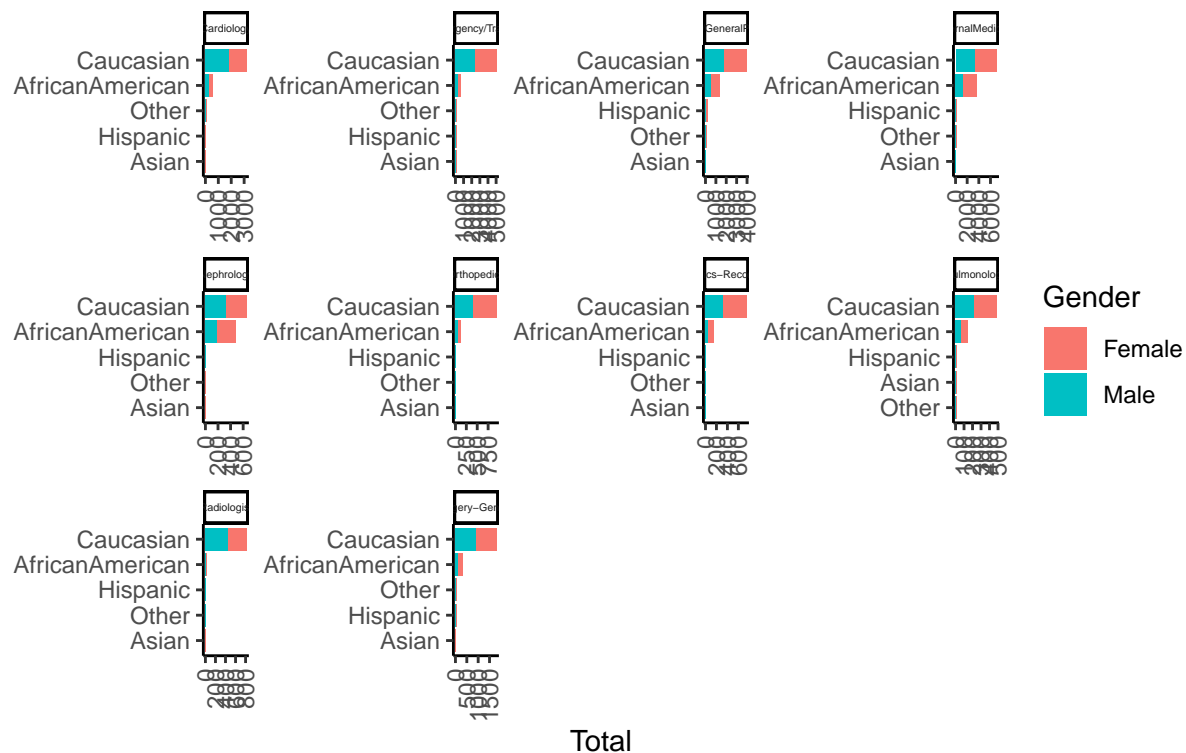
Insights basado en hallazgos

En esta sección desarrollo más la idea de los hallazgos de relevancia desarrollados en el punto anterior.

Una de las inquietudes importante a responder es ¿A través de la raza y el genero que tipo de enfermedades son causantes de reincidencia en terminos de hospitalización?

Relation Between Medical Specialty per Gender – Race

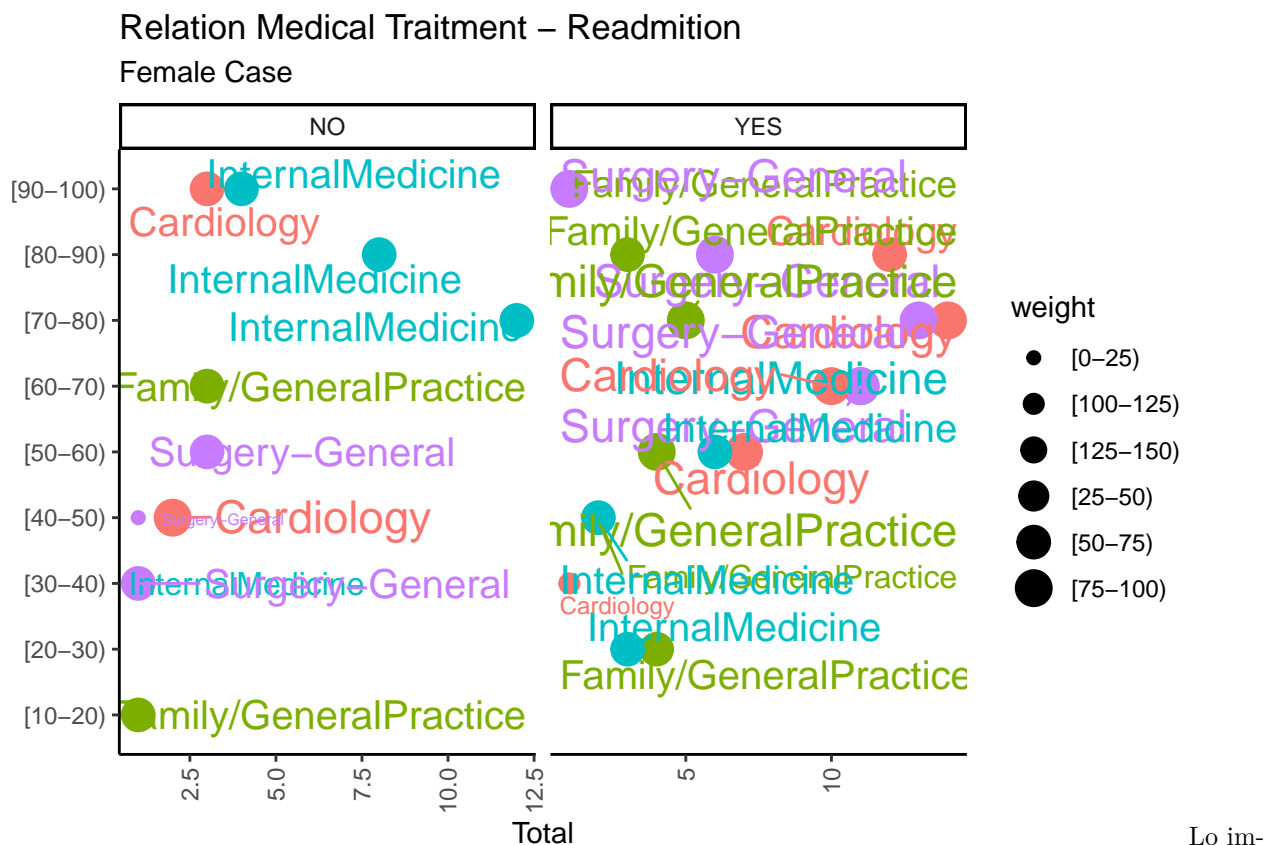
Top 10 in Medical Specialty



Unos puntos importantes a mencionar :

- En todos los casos, la raza Caucásica es la más re- incidente dada cualquier tipo de especialidad médica;
- Los Asiaticos son los menos re-incidentes;
- Se evidencia que las mujeres son las que más tienden a sufrir re-admisiones cuando la diabetes es positiva.
- Los hombres caucásicos son más re incidentes en el caso de tener atenciones con :
 - Cardiología;
 - Nefrología;
 - Radiología.

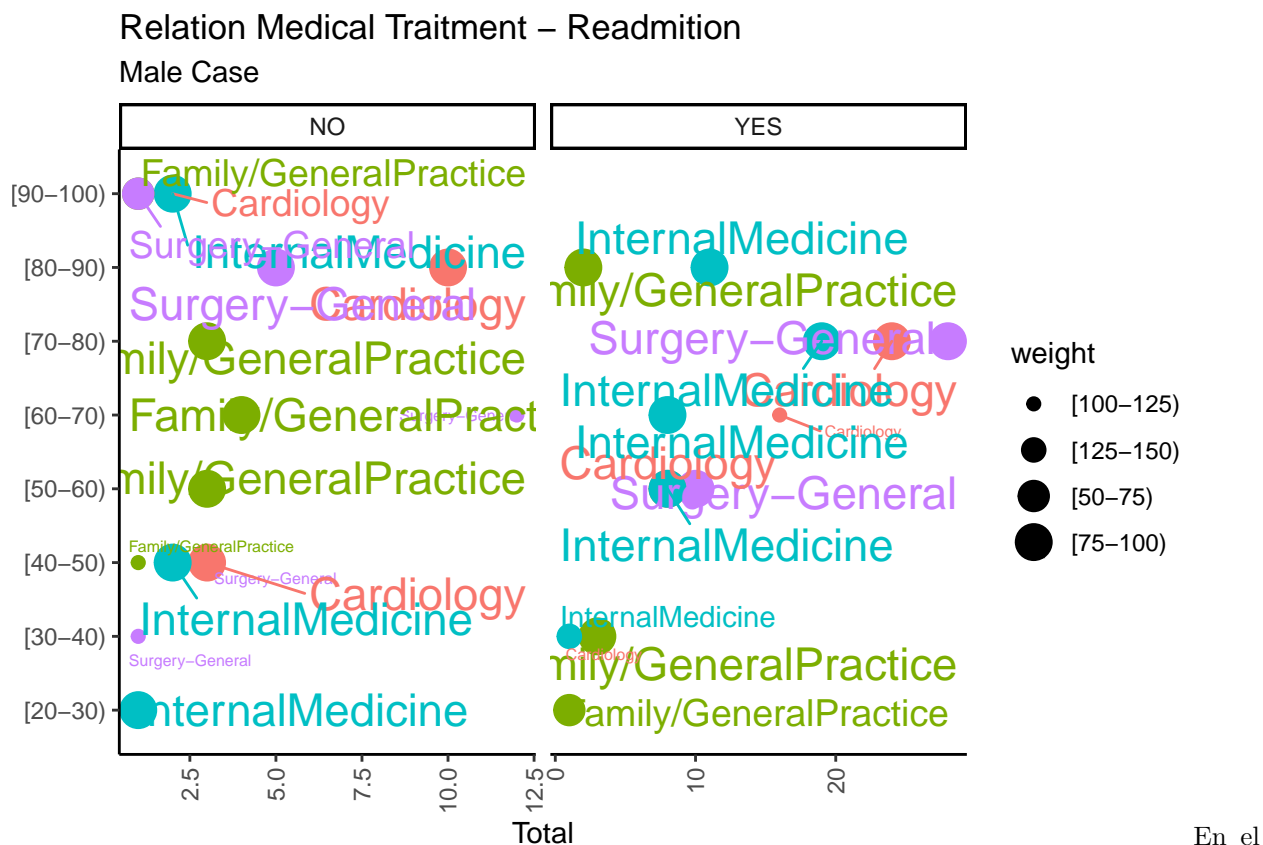
Ahora exploro a través de un gráfico de importancia, que tipo de enfermedades son aquellas de generan mayor probabilidad de ser re-incidente o no en el caso de presentar diabetes y teniendo presente el rango de edad. Para el caso de las mujeres los hallazgos fueron los siguientes:



portante a destacar es lo siguiente:

- A menor edad, los efectos de medicina general (tratamientos) generan poco o nulo proceso de re-admisión;
- Los factores cardiológicos son bajos detonadores de re-incidencias;
- A mayor edad y en especial a partir de los 30 años (rango), la medicina interna, general, y cirugías, todas las mujeres son propensas a re-incidir en hospitalizaciones dado el factor de diabetes.

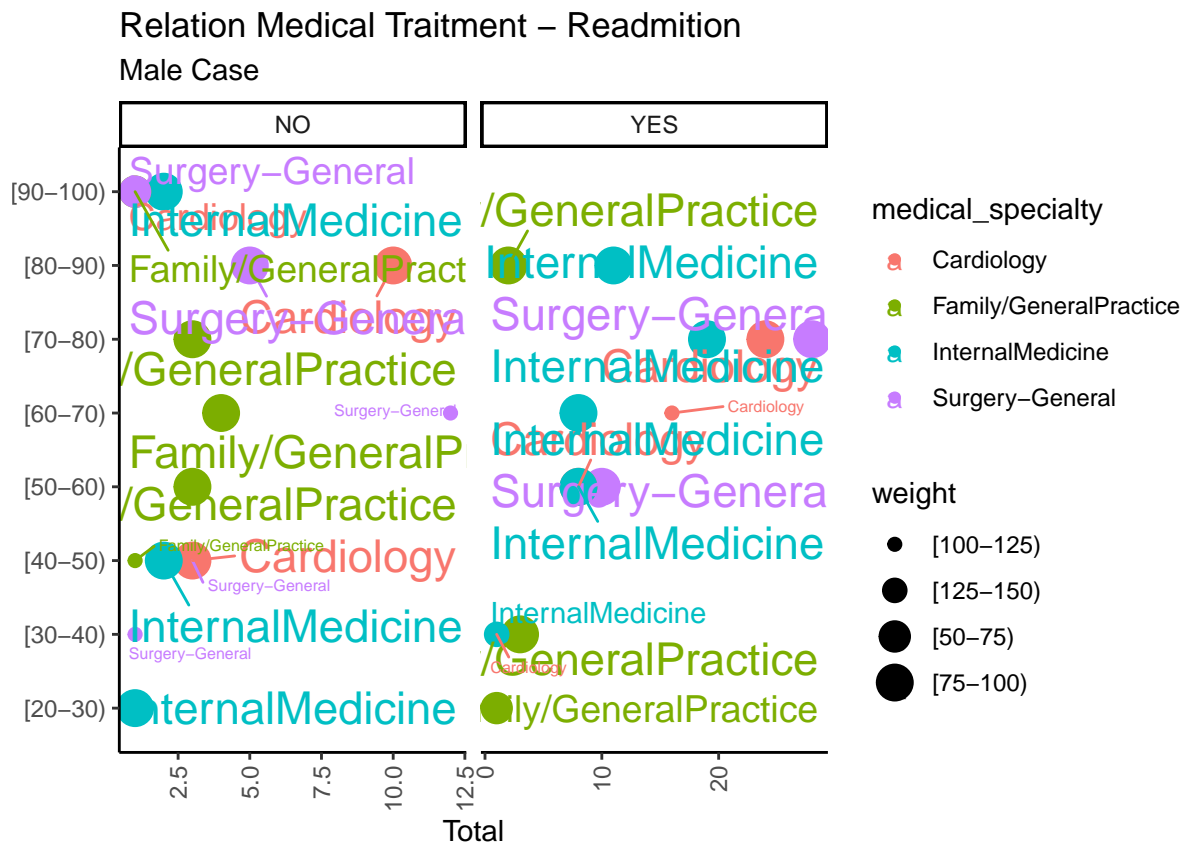
En el caso de los hombres el resultado es el siguiente:



caso de los hombres hay poco por decir :

- Entre los 40 y los 80 años, las reincidencias de hospitalización son más frecuentes cuando los pacientes tienen antecedentes de medicina general, cardiología, medicina interna y cirugías.
- Mientras que entre los 30 y 60 años, y con antecedentes de diabetes, es baja la re-incidencia dado los tratamientos de medicina interna y cardiología.

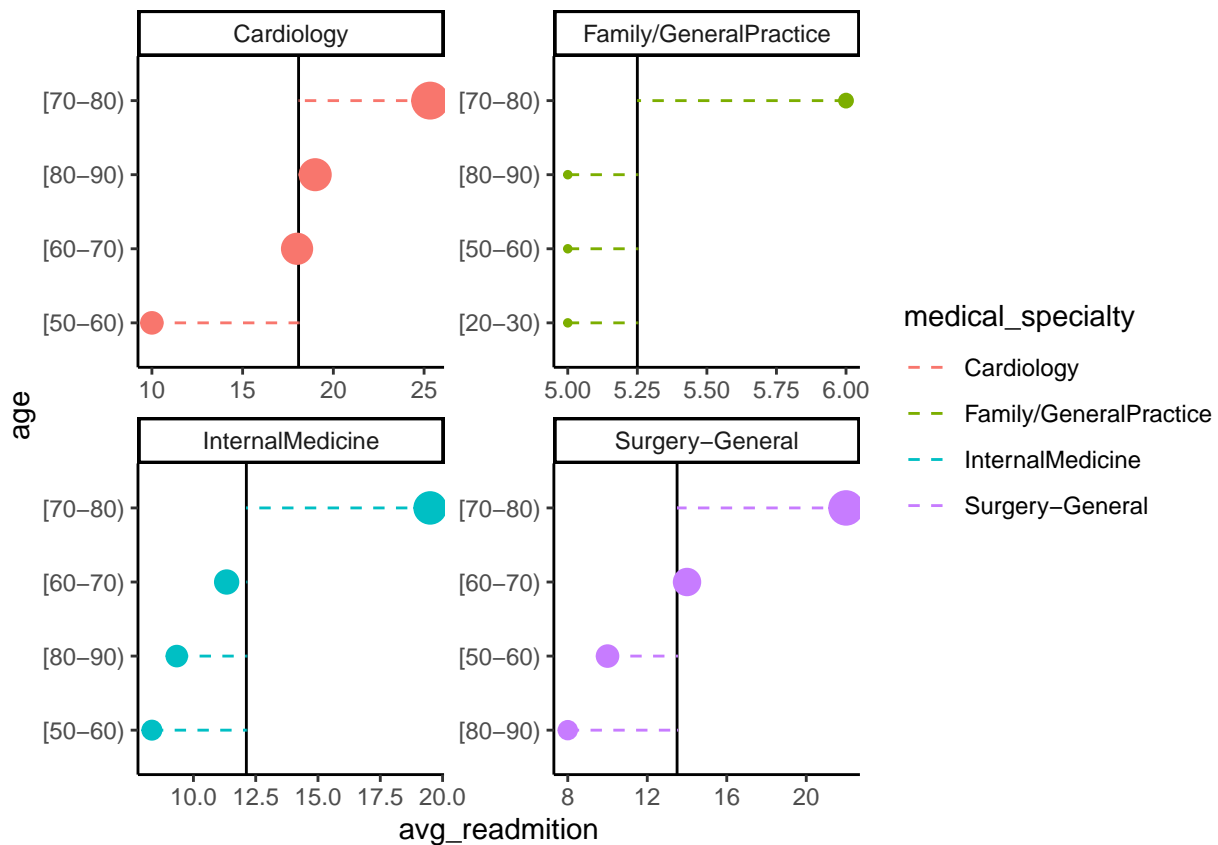
Si lo anterior se mira a través del peso el resultado es el siguiente:



Conclusiones practicas:

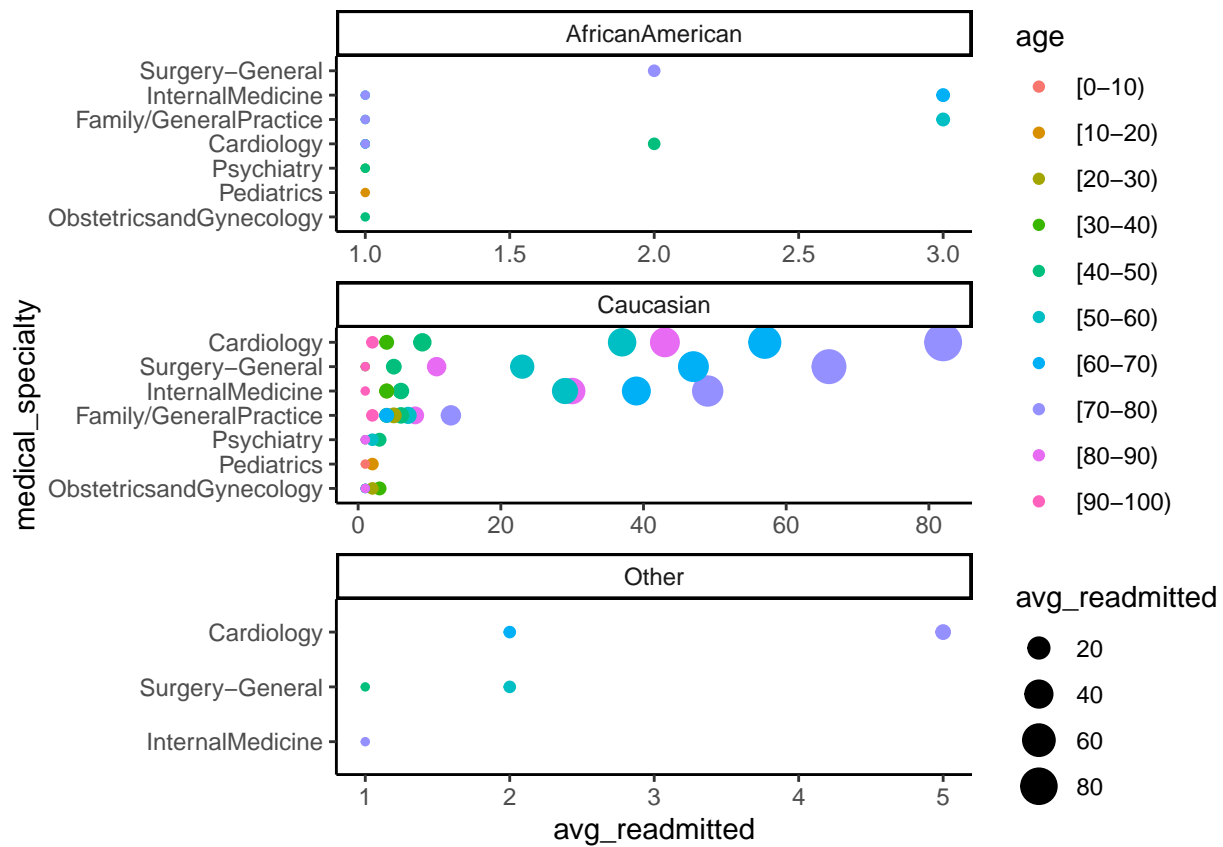
- Entre los 20 y los 80, las personas que se encuentren entre 75 y 100 Kilos, son mas propensas a sufrir de reincidencia dada la diabetes,
- Las personas entre 20 y 30 años con este rango de peso (75-100 kilos), son poco propensos a ser readmitidos a pesar de la diabetes.

Para finalizar este segmento voy hacer una síntesis de lo antes expuesto a través de un gráfico de paletas.

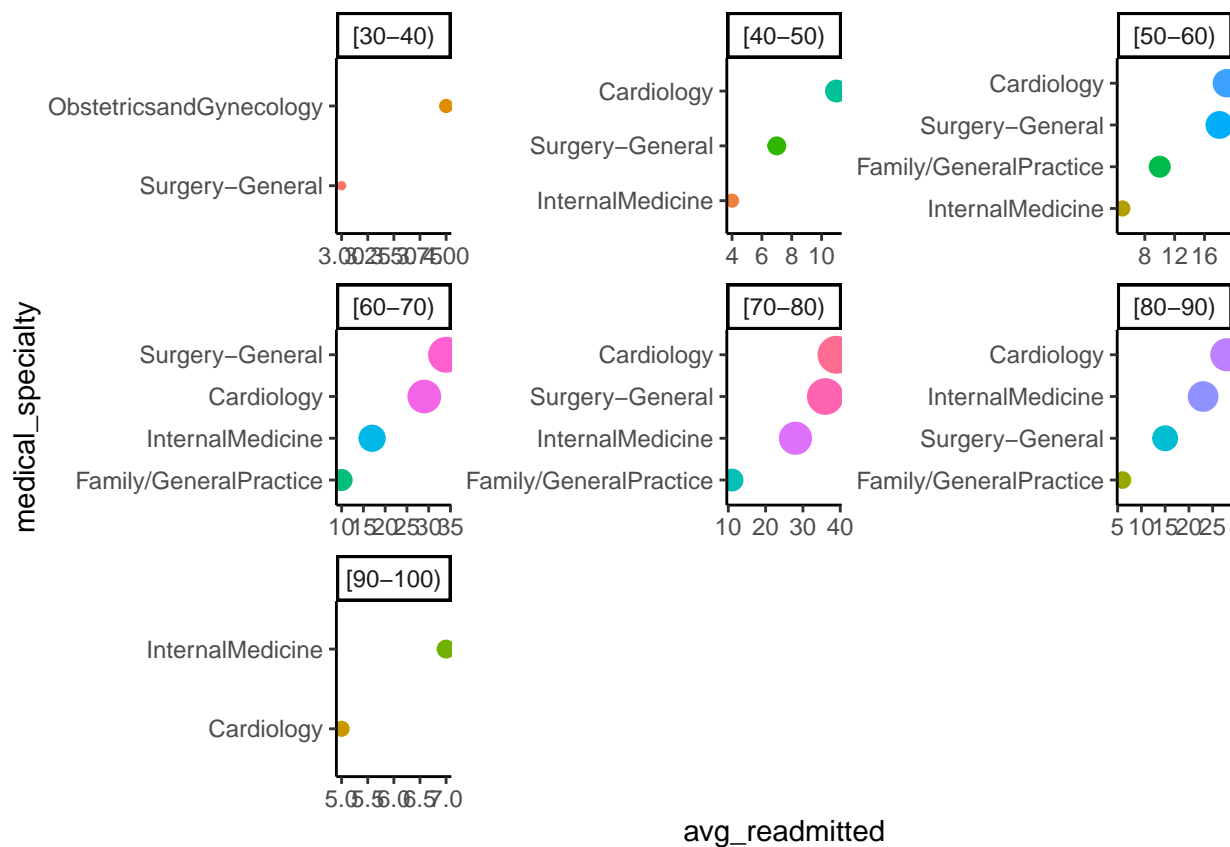


- Los efectos del peso y la edad son detonadores más probables para re-admisión dada la diabetes en el caso de cardiología, medicina interna y cirugías;
- El efecto de la edad en y el peso en medicina general es poco o nulo;
- La tasa de re-admisión en el caso de cardiología es baja hasta los 60 años, de ahí en adelante explota.

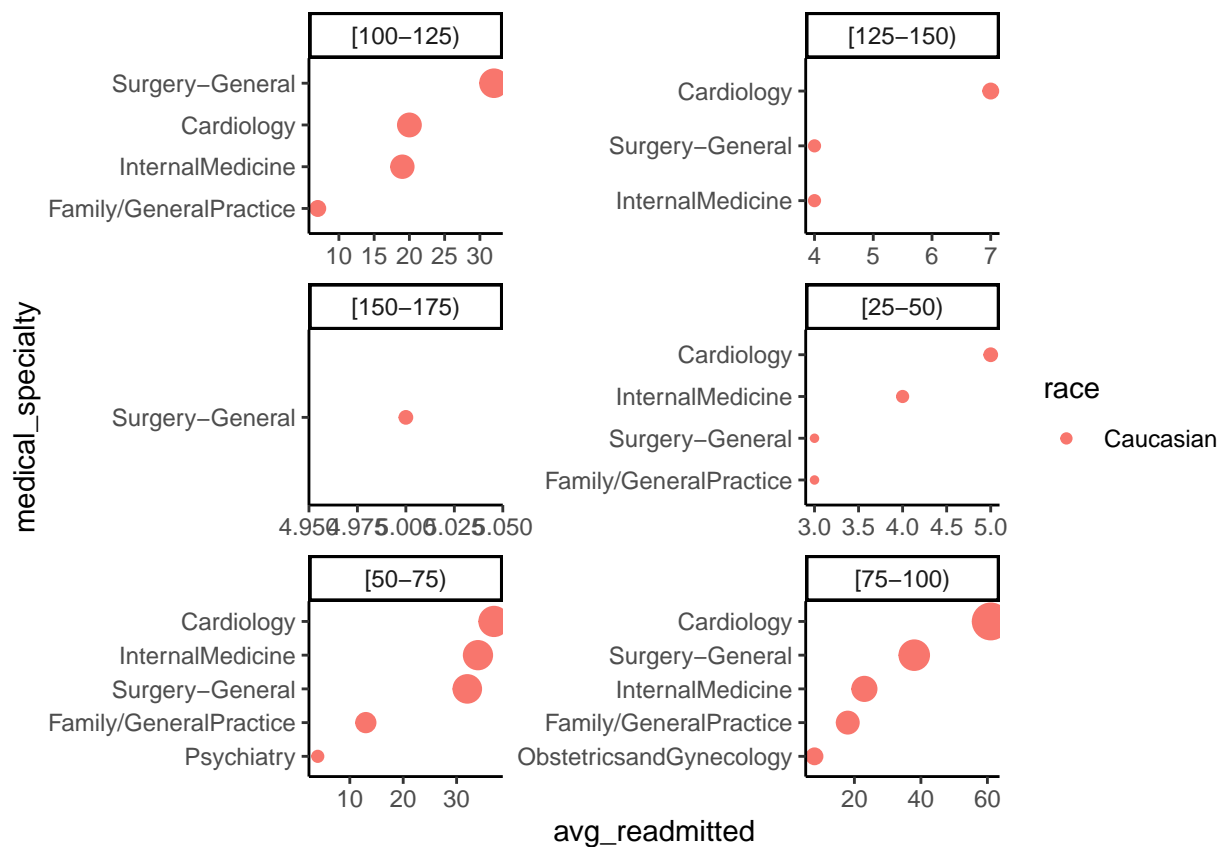
Una forma de validar lo anterior, agregando la raza, es la siguiente:



Notesé que los Caucásicos, independiente a la edad son los más propensos a sufrir de re-admisiones. Si se evalua este caso por edad, los resultados detallados son los siguientes



Lo anterior indica que los caucásicos entre los 40 y los 60 años son los más propensos a sufrir de re-admisión dada la diabetes.



el gráfico anterior, representan que después de los 75 kilos y hasta los 125, la propensión de re-admisión es extremadamente alta.

Para finalizar este documento, en la ficha técnica se crearon modelos que ayudan a evaluar en que casos las re-admisiones son altamente probables y cuando no, y con base a ello, fue que se presento estos hallazgos.