**Olariaga Sandra, Pesce Cristian**

**Documentación del Práctico de Análisis Exploratorio y Curación de Datos**

**Chequeos puntuales por campo**

-Estado del turno: se unificaron los campos “Atendido", "Asignado" y "Receptado" por “Atendido”. Luego de esto los campos quedaron en solo dos categorías: “Atendido” y “Ausente”.

-Edad: para el campo edad hemos decidido eliminar los campos que contenían el número 120, ya que asumimos que se trataba de un error en la carga del mismo. Luego de eliminarlos, vemos que el número máximo es 102, el cual nos parece correcto ya que se trata de una misma persona (pudimos verlo por el número de historia clínica).

-Sexo: cualquier categoría diferente de “M” o “F” se la considera un error de carga y se lo considera un valor nulo.

-Dirección (localidad/provincia): observamos que estos campos son muy sucios y contienen gran cantidad de errores en la carga. Se procedió a hacer una limpieza de estos campos a fin de unificar las diferentes formas en las que estaba escrita la palabra “Córdoba”. Decidimos dividir el campo en dos categorías: Córdoba y otro, tanto para localidad como para provincia.

-Nombre del canal: unificamos las categorías “Teléfono” y “Call center” como “Teléfono”.

-Prestador: para este campo hemos decidido unificar a las distintas obras sociales y prepagas en la categoría “Otros”, resultando en dos categorías finales: “Otros” y “Consumidor final”, siendo consumidor final a las personas sin obra social o prepaga.

-Fecha: en este campo se comprobó que las fechas se encuentren entre el 2018 y 2019. Para esto utilizamos las funciones max y min.

-Fecha otorgado: se comprobó que la fecha de otorgamiento del turno sea inferior a la fecha efectiva del turno. Se detectó que existen fechas de otorgamientos posteriores a la fecha del turno llegando a la conclusión de que se trata de prácticas del servicio diagnóstico por imágenes y con sobreturno.

-Prestación asignada: en este campo decidimos dividirlo en dos categorías: “Consulta Médica” que es la más frecuente y “Otra” agrupando al resto de prestaciones.

-Servicio: para esta campo eliminamos los campos de “Guardia”, “Chequeo” y los que comenzaban con la palabra “Demanda”. Al eliminar esto solo vemos las distintas especialidades.

-Es sobre turno: revisamos que los valores posibles en este campo sean “SÍ” y “NO”.

**Creación de nuevos campos**

Al data frame le incorporamos 4 columnas:

-Difdias: aquí comparamos la diferencia de días entre la fecha donde se otorgó el turno y la fecha efectiva del turno.

-DiaTurno: en esta columna identificamos el día del turno desde el número 0, correspondiente al día lunes, hasta el día domingo identificado con el número 6.

-MesTurno: en esta columna identificamos el mes de turno, siendo el 1 el mes de enero y el 12 el mes de diciembre.

-DiaTurnoNombre: aquí le asignamos el nombre del día de la semana a cada número de día (columna DiaTurno) para poder identificar de manera más simple la cantidad de turnos, ausencias, etc. según el día de la semana.

**Incorporación de información climática**

Aquí tomamos la decisión de incorporar una nueva columna a nuestro data frame, utilizando la columna de precipitación del data frame del clima. Esto se basa en querer observar si las ausencias tienen alguna relación con la caída de lluvia.

**Chek List usado**

1) Rotulos de las variables, controlar que no posean caracteres especiales. Controlado y arreglado.

2) Valores NaN y valores nulos, controlado y arreglado

3) Codificación de variables categoricas: No se aplica.

4) Testear la consistencia de las variables, controlado y arreglado

5) Identificar y eliminar outliers, controlado y arreglado