**NOTE BOOK DATASET ICBF**

**Problema**

¿Cómo mejorar la calidad de vida de los beneficiarios? -- Que es calidad de vida? ¿Cómo disminuir la muerte por desnutrición en el país? --Que es disminuir? ¿Cómo invertir mejor los recursos del programa primera infancia? --implicaría restructuraciones del programa o manteniendo el programa original. Objetivo del análisis: Definir un objetivo smart a partir de las preguntas suministradas por los interesados y determinar si los datos suministrados son suficientes para dar respuesta o cumplir con el objetivo smart planteado y las fuentes de los demás que sean necesarios para lograr tal fin.

**Objetivo del análisis**

Realizar un análisis preliminar del conjunto de datos aplicando estadística descriptiva.

**Preguntas Descriptivas**

1. ¿Cuál es la clasificación de los servicios de acuerdo a las 8 atenciones?
2. ¿Cómo es la distribución del gasto por departamento?
3. \*\*¿Cómo es la distribución del presupuesto de acuerdo a las 8 atenciones?
4. \*\*Cómo ha sido la ejecución de los presupuestos en los últimos 8 años, se ha ejecutado al 100%?
5. ¿Cuál es el periodo de ejecución de los valores programados?
6. ¿Cuál o cuáles fueron los factores para la elección de la distribución del presupuesto?
7. ¿Cuáles son las características de los servicios?
8. ¿Cuál ha sido el presupuesto destinado al ICBF por el gobierno precedente?
9. ¿Qué características tenían los programas del gobierno precedente?
10. ¿Cuáles fueron los resultados de esa ejecución del presupuesto del gobierno precedente?
11. ¿Cuál es el significado del código del servicio?
12. ¿Cuáles son los factores para determinar el número de cupos de cada programa o servicio?
13. ¿Cuál es el número de contratistas por región?
14. ¿Cuál es el número de municipios atendidos por contratista?
15. ¿Cuál es el número de niños atendidos por servicio de 8 atenciones?
16. ¿Cuáles son los parámetros de evaluación de los contratistas?
17. ¿Cuál es el presupuesto asignado a cada contratista?
18. ¿Cuál es el número de niños atendidos por contratista?
19. ¿Cuál es el número de niños atendidos por el programa y por municipio?
20. ¿Cuál es el número de niñas atendidos por el programa y por municipio?
21. ¿Cuál es el número de niños foráneos atendidos por el programa y por municipio?
22. ¿Cuál es el número de niñas foráneas atendidos por el programa y por municipio?
23. ¿Cuál es la nacionalidad de los niños foráneos atendidos por el programa?
24. ¿Cuál es el número de niños atendidos por etnia?
25. ¿Cuál es el número de niños atendidos con algún tipo de discapacidad por municipio?
26. ¿Cuál es el número de niños atendidos sin identificación?
27. ¿Cuál es el peso y talla promedio por grupo de edades?
28. ¿Cuál es el peso y talla adecuado por grupo de edades?
29. ¿Cuál es el número de niños con problemas de nutrición?
30. ¿Cuáles son los municipios de los niños con problemas de nutrición?
31. ¿Cuáles son los contratistas que atienden a los niños con problemas de nutrición?
32. ¿Cuál es la diferencia entre número de cupos y cupos UDS?

**Preguntas Diagnosticas**

1. ¿Por qué (aproximado) una cuarta parte del presupuesto está asignado a Bogotá D.C. Sede Nacional? cuando no se indican cupos programados o unidades programadas?
2. ¿Por qué no existe una correlación directa entre el valor programado y los cupos programados?
3. ¿Por qué no existe una correlación directa entre las unidades programadas y los cupos programados?
4. ¿Por qué existe un valor programado para un servicio que no tiene unidades ni cupos programados?
5. ¿Por qué la mayor cantidad de unidades programadas esta asignada a Antioquia cuando la mayor cantidad de cupos está asignada a Bogotá?
6. ¿Por qué Cali es la ciudad con más servicios asignados?
7. ¿Por qué el año de vigencia es 2016 para el archivo base\_prim\_inf y el presupuesto esta referenciado a febrero de 2018?
8. Por qué en el presupuesto para Arauca , municipio Saravena aparecen 5 servicios y en el cuadro base\_prim\_inf aparecen 8 servicios?
9. ¿Por qué para Arauca en el servicio HCB TRADICIONAL- COMUNITARIO (T) tienen 360 cupos programados y en el cuadro base\_prim\_inf aparecen para ese servicio 716 beneficiarios?
10. ¿Por qué se existen niños sin vacunación, pero si cuentan con afiliación a salud?
11. Cuál es la fórmula para el cálculo de los score?
12. ¿Cuál es el score ideal?
13. ¿Cuál es la fórmula para tomar los diferentes rangos de talla peso edad?
14. ¿Por qué existen adultos en la hoja de nutrición catalogados como niño o niña entre 6 meses y 5 años y 11 meses?
15. ¿Por qué existen 998 registros con fecha de medición 1900-01-01?
16. ¿Por qué no existe una relación en cuanto a las regiones de los datos en los diferentes archivos?
17. ¿Porque la totalización por región de los valores programados difiere de la del aporte del ICBF en el cuadro de contratos?

**Preguntas predictivas**

1. ¿Por qué (aproximado) una cuarta parte del presupuesto está asignado a Bogotá D.C. Sede Nacional? cuando no se indican cupos programados o unidades programadas?
2. ¿Por qué no existe una correlación directa entre el valor programado y los cupos programados?
3. ¿Por qué no existe una correlación directa entre las unidades programadas y los cupos programados?
4. ¿Por qué existe un valor programado para un servicio que no tiene unidades ni cupos programados?
5. ¿Por qué la mayor cantidad de unidades programadas esta asignada a Antioquia cuando la mayor cantidad de cupos está asignada a Bogotá?
6. ¿Por qué Cali es la ciudad con más servicios asignados?
7. ¿Por qué el año de vigencia es 2016 para el archivo base\_prim\_inf y el presupuesto esta referenciado a febrero de 2018?
8. Por qué en el presupuesto para Arauca , municipio Saravena aparecen 5 servicios y en el cuadro base\_prim\_inf aparecen 8 servicios?
9. ¿Por qué para Arauca en el servicio HCB TRADICIONAL- COMUNITARIO (T) tienen 360 cupos programados y en el cuadro base\_prim\_inf aparecen para ese servicio 716 beneficiarios?
10. ¿Por qué se existen niños sin vacunación, pero si cuentan con afiliación a salud?
11. Cuál es la fórmula para el cálculo del score?
12. ¿Cuál es el score ideal?
13. ¿Cuál es la fórmula para tomar los diferentes rangos de talla peso edad?
14. ¿Por qué existen adultos en la hoja de nutrición catalogados como niño o niña entre 6 meses y 5 años y 11 meses?
15. ¿Por qué existen 998 registros con fecha de medición 1900-01-01?
16. ¿Por qué no existe una relación en cuanto a las regiones de los datos en los diferentes archivos?
17. ¿Porque la totalización por región de los valores programados difiere de la del aporte del ICBF en el cuadro de contratos?

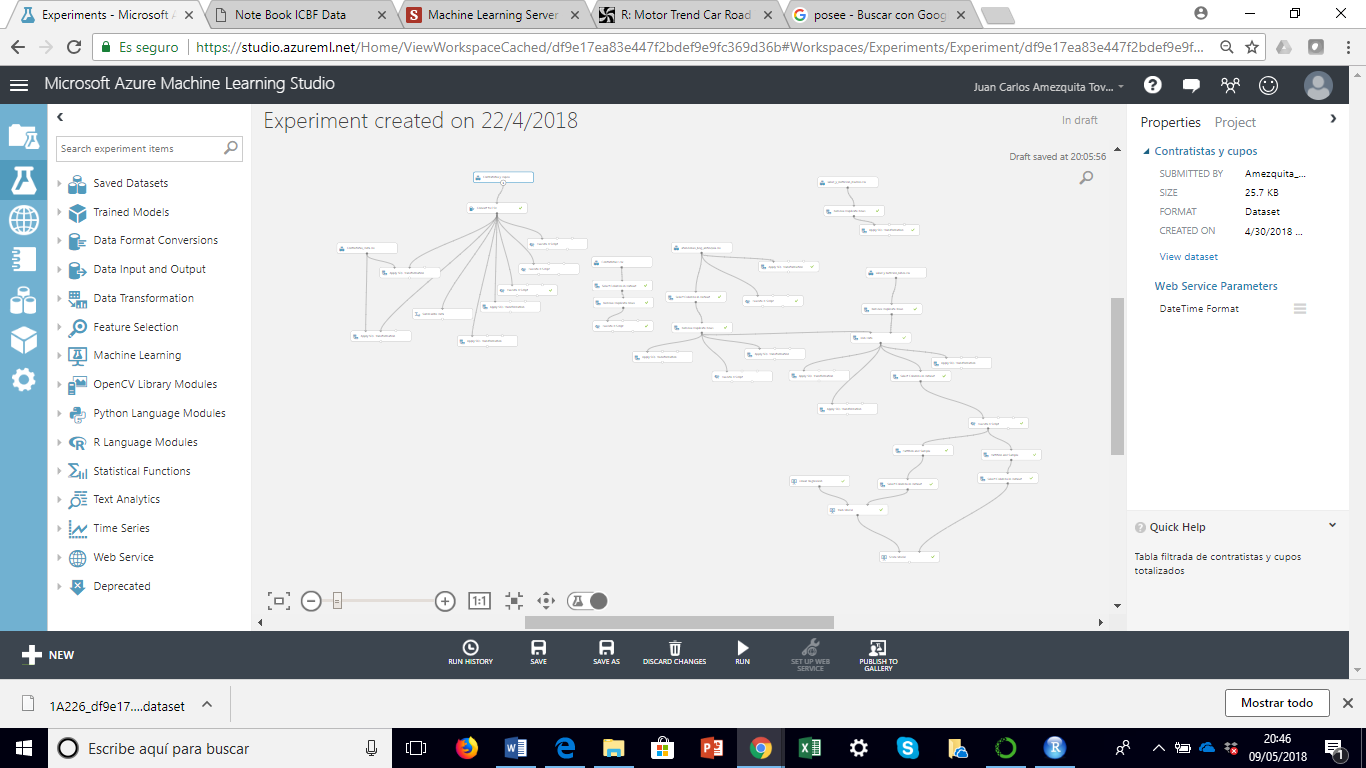
**Limpieza de archivos**

Se procede a realizar un limpieza preliminar de los archivos suministrados por el interesado antes de proceder a gargarlos en el programa R.

Para el archivo 8\_atenciones\_Bogota.xls se eliminan las siguintes columnas:

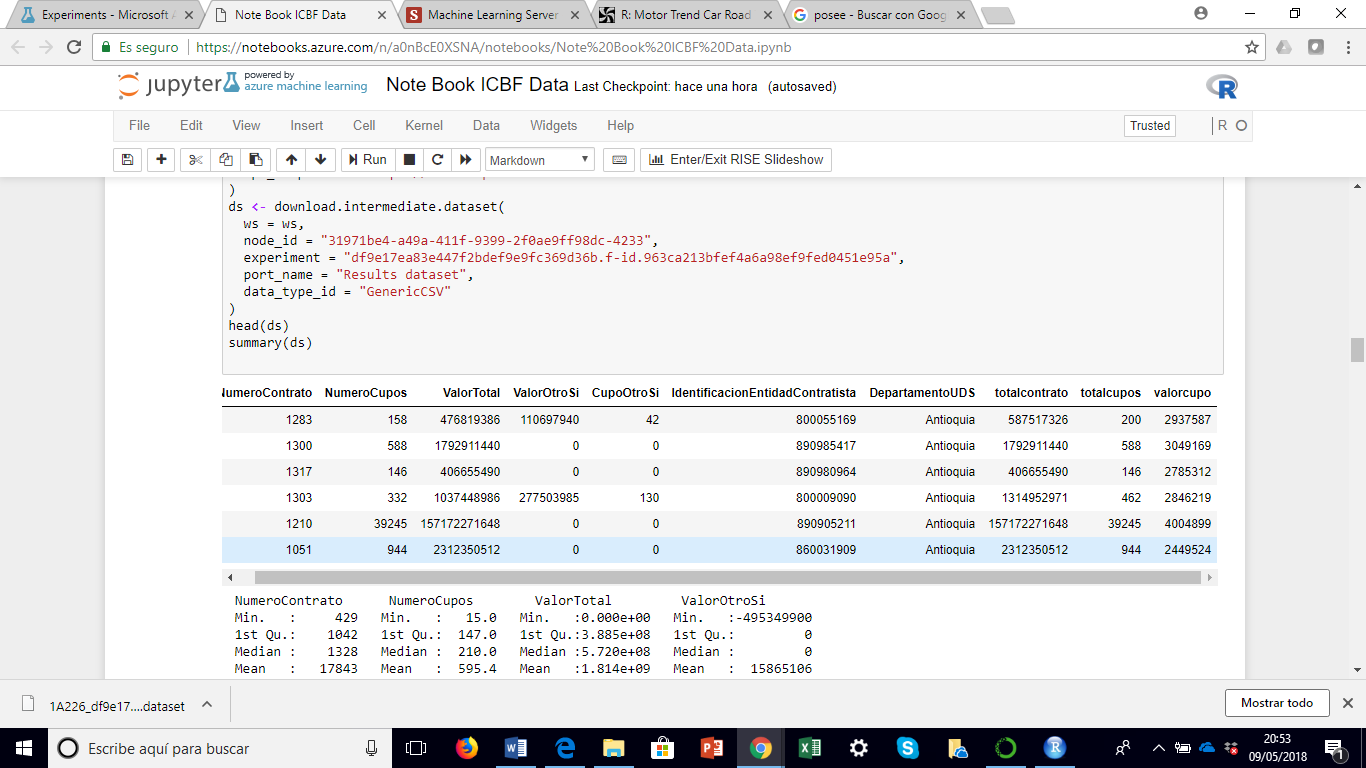
* RegionalContrato:Dato innecesario debido a que se sabe que los datos pertenecen a la ciudad de Bogotá.
* IdContrato: Se considera información redundante ya que se tienen los números de los contratos. Para el archivo beneficiarios\_Dic\_217\_bta\_antioquia.xls se eliminan las siguientes columnas:
  + Región Contrato: Dato repetido.
  + FechaInicioOtrosi: Dato no relevante para el análisis.
  + FechaFinOtrosi: Dato no relevante para el análisis.
  + CodigoRegionalContrato: Dato duplicado.
  + NombreTipoOtrosi: Dato no relevante para el análisis.
  + CodDocumento: Todos los registros tienen NIT en ese campo.
  + CodigoRegionaUDS: Dato duplicado.
  + RegionalIUD: Dato duplicado. Las siguietnes columnas se elimina de esta tabla y se utiliza en un nuevo archivo en el cual solo se tiene los datos de los contratistas
  + AñoVigencia
  + FechaInicio
  + FechaFin
  + NumeroCupos
  + ValorTotal
  + ValorOtrosi
  + CupoOtrosi
  + NombreEntidadContratista
  + IdentificacionEntidadContratista
  + TipodeEntidadContratista
  + NombreUnidadServicio
  + Zona de Ubicación
  + DirecciónUDS
  + CuposUDS
  + LatitudUDS
  + LongitudUDS
  + FechaCapturaUDS
  + HoraCapturaUDS
  + FechaVinculacionUDScontrato Se procede a cargar el data set de atenciones Bogotá Antioquia.

**DESARROLLO DEL ANÁLISIS DESCRIPTIVO.**

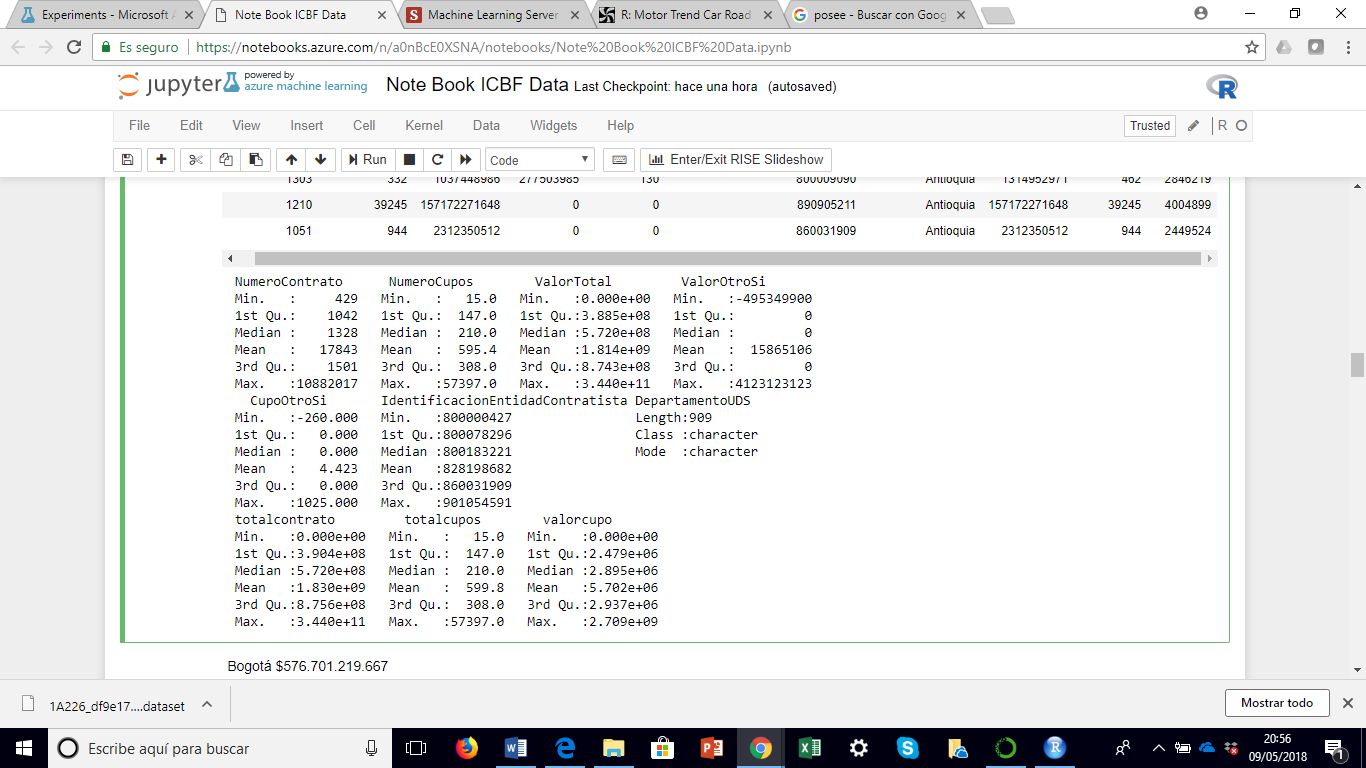
Para el análisis exploratorio de los datos se utilizo la herramienta de azur estudio machine learning. El resultado del experimento se muestra a continuación. 

**DATASET CONTRATISTAS\_UDS**

* 1. Se realizó ajuste al dataset totalizando los atributos valor del contrato y cupos generándose las columnas totalcontrato, totalcupos, valorcupo.



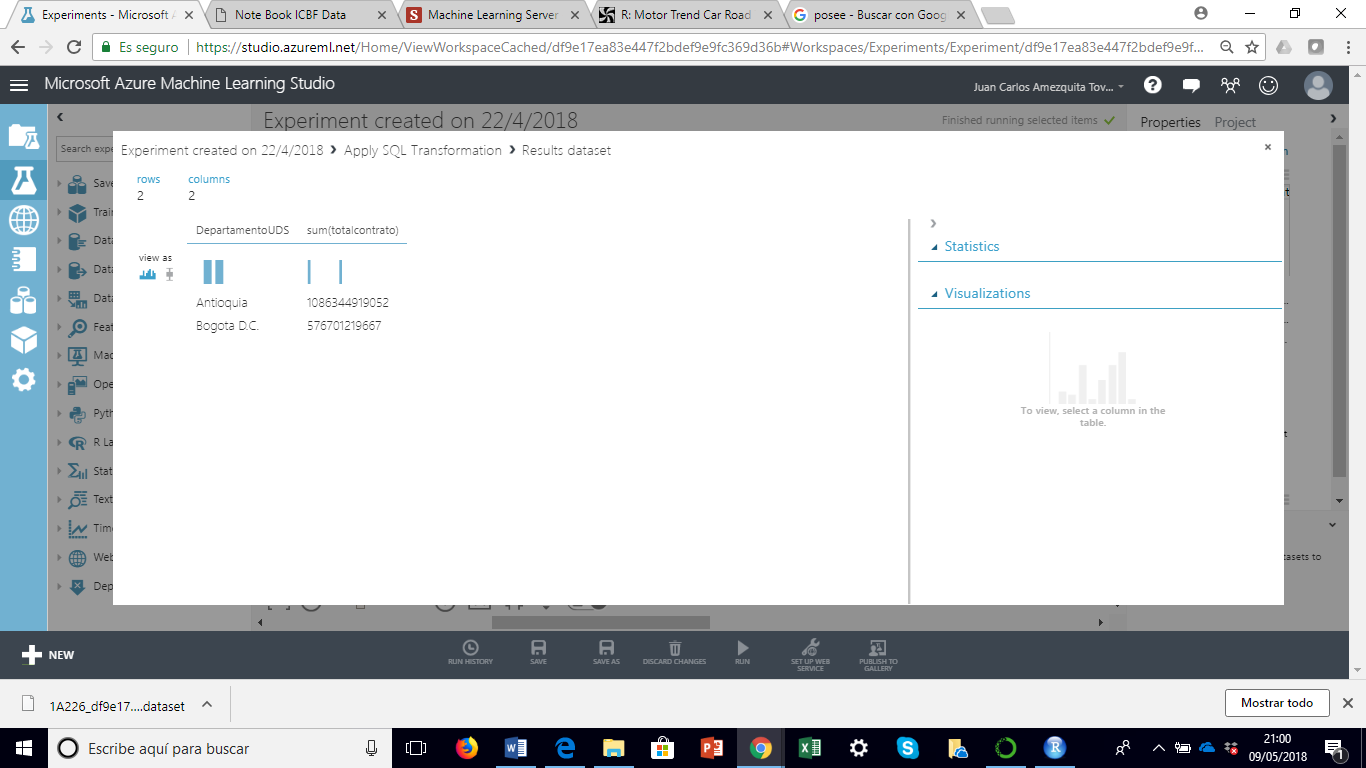
* 1. Se aplica la función de summary(ds) en R.



* 1. ¿Cómo es la distribución del gasto por departamento?

-Bogotá $576.701.219.667

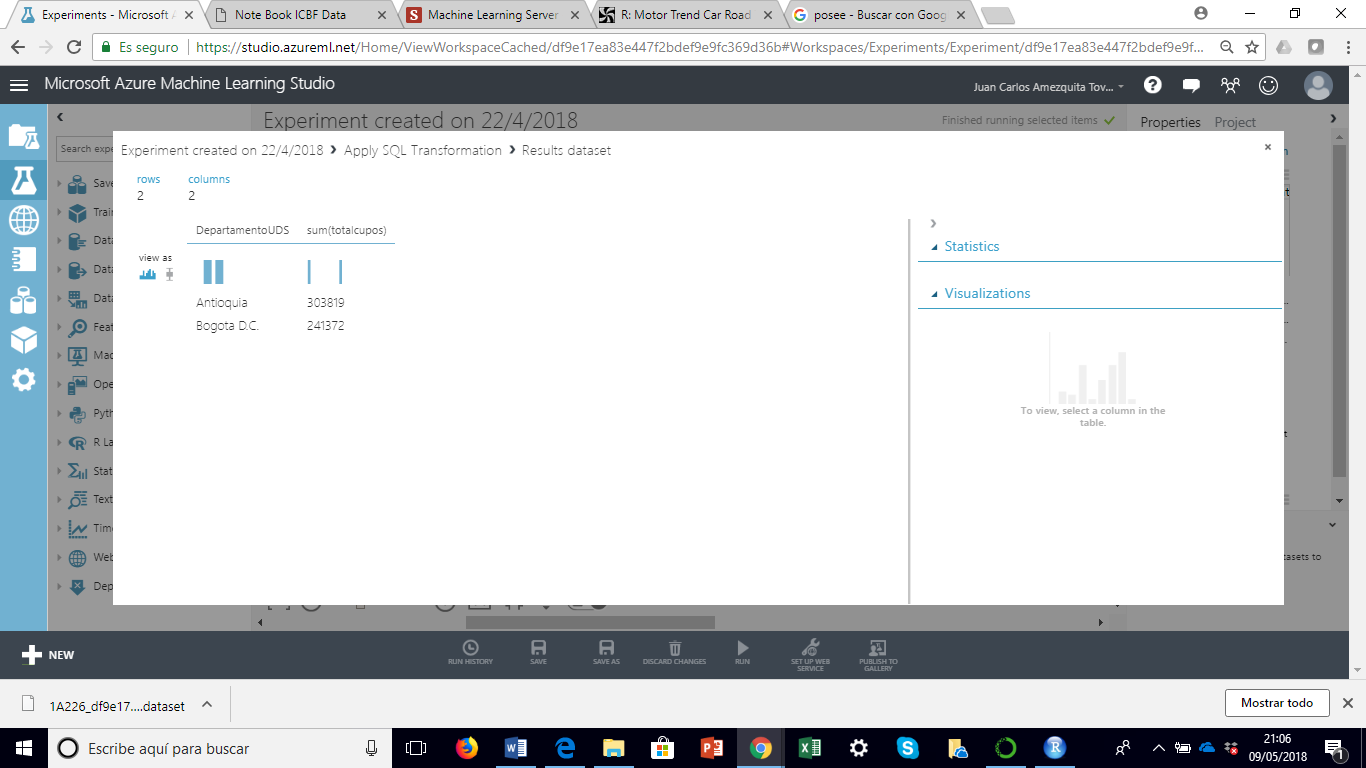
-Antioquia $1.086.344.919.052



* 1. ¿Cuál es el número de cupos ejecutados por departamento?

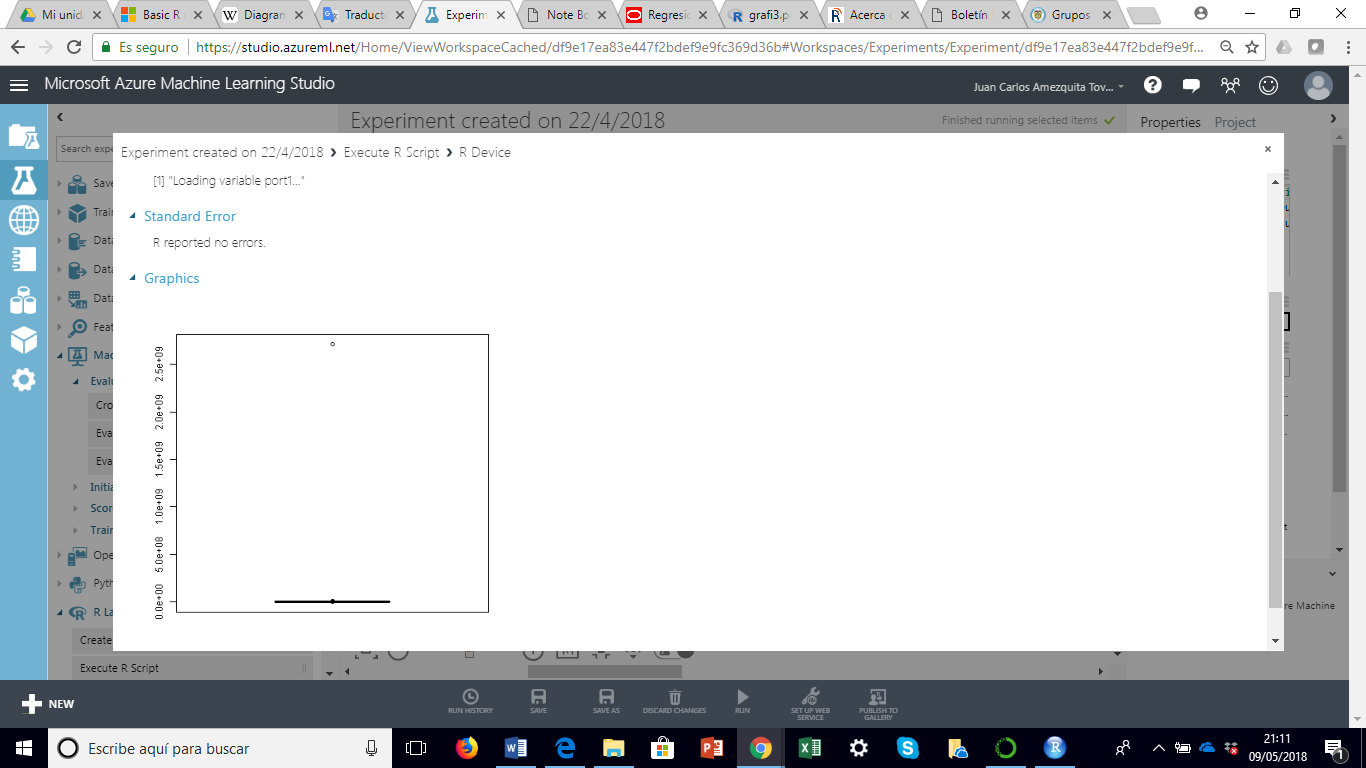
Bogotá 241.372

Antioquia 303.819



* 1. ¿Cual es el valor máximo y mínimo de un cupo por contrato?

Máximo 2.709.017.484 "COOPERATIVA MULTIACTIVA DE SAN ANTONIO DE PRADO COOMULSAP" Min 0



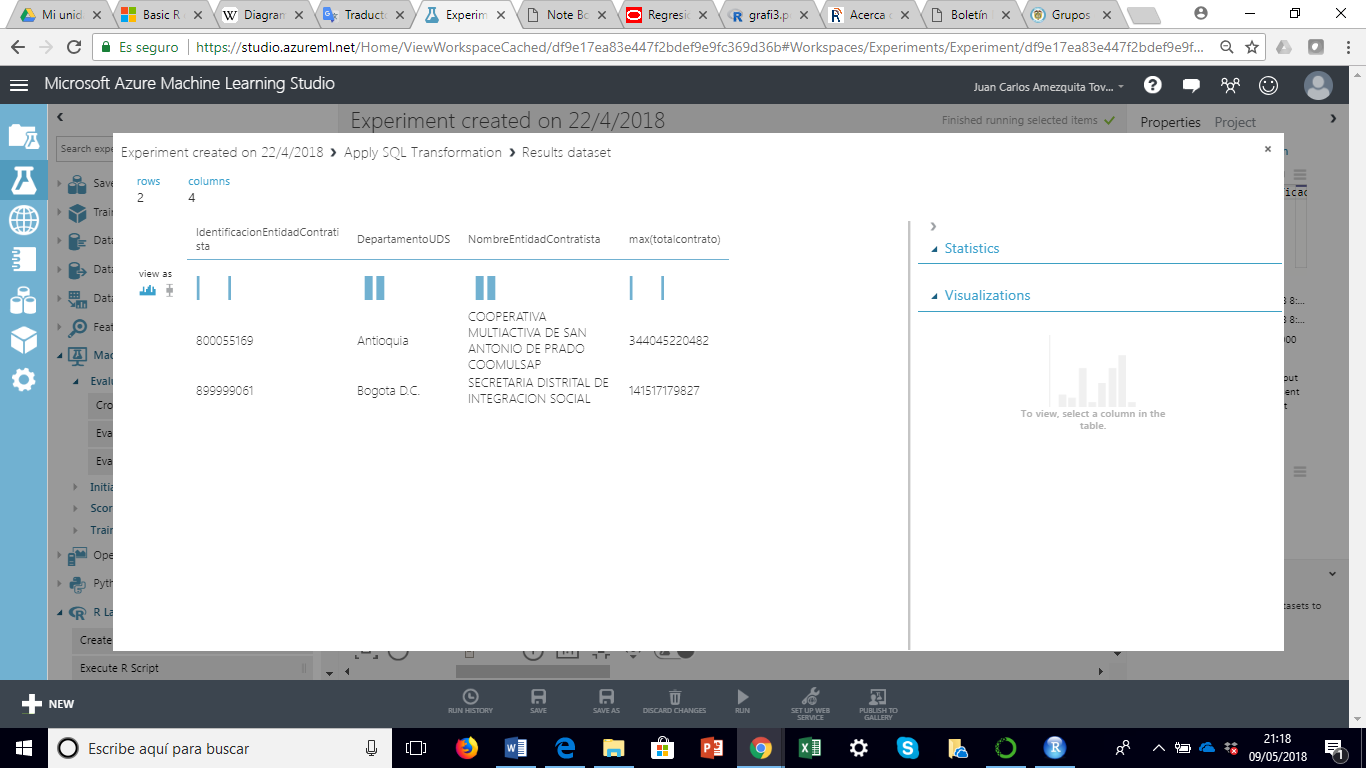
* 1. ¿Cuál es el número de cupos máximo y mínimo por contrato?

Min 15 y Máximo 57.397.

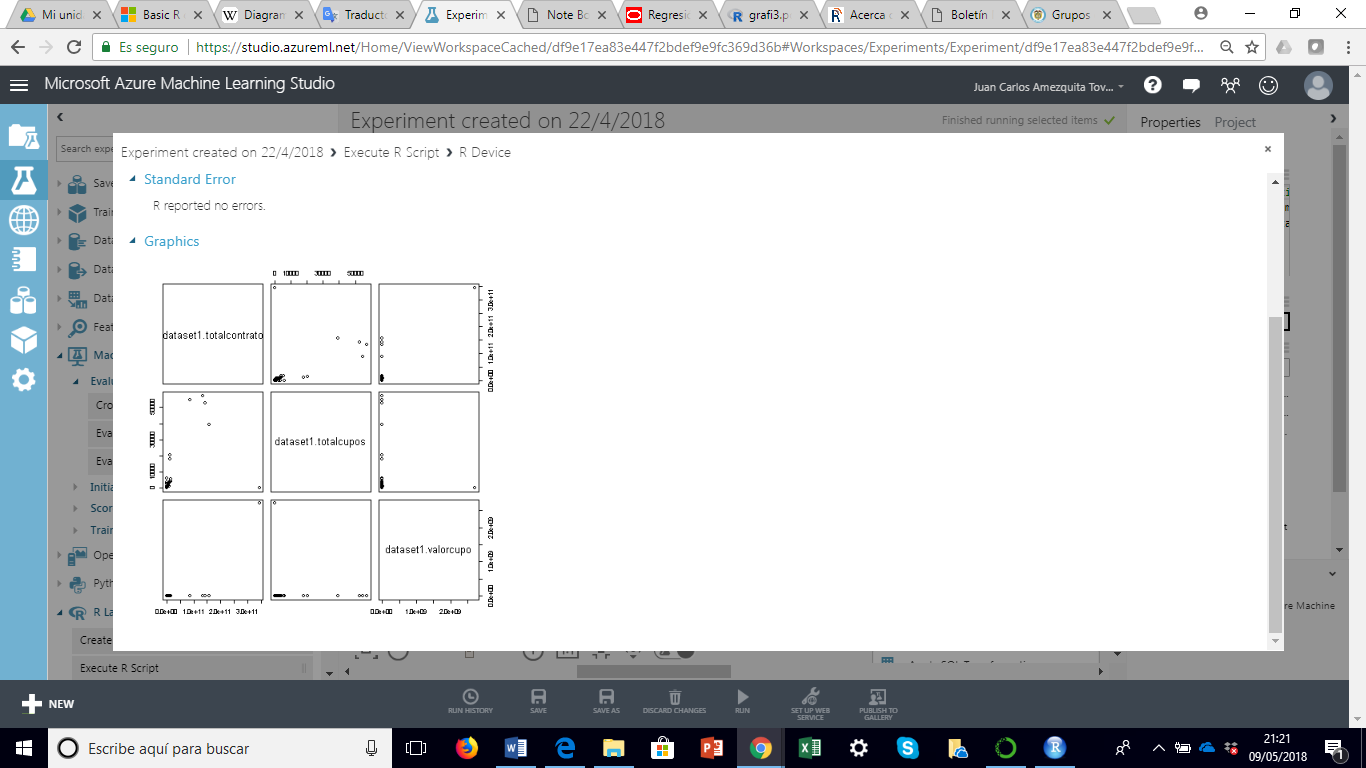
G) ¿Cuál es el contratista con mayor presupuesto asignado por departamento y por contratista?

Bogotá SECRETARIA DISTRITAL DE INTEGRACION SOCIAL $141.517.179.827

Antioquia COOPERATIVA MULTIACTIVA DE SAN ANTONIO DE PRADO COOMULSAP $344.045.220.482

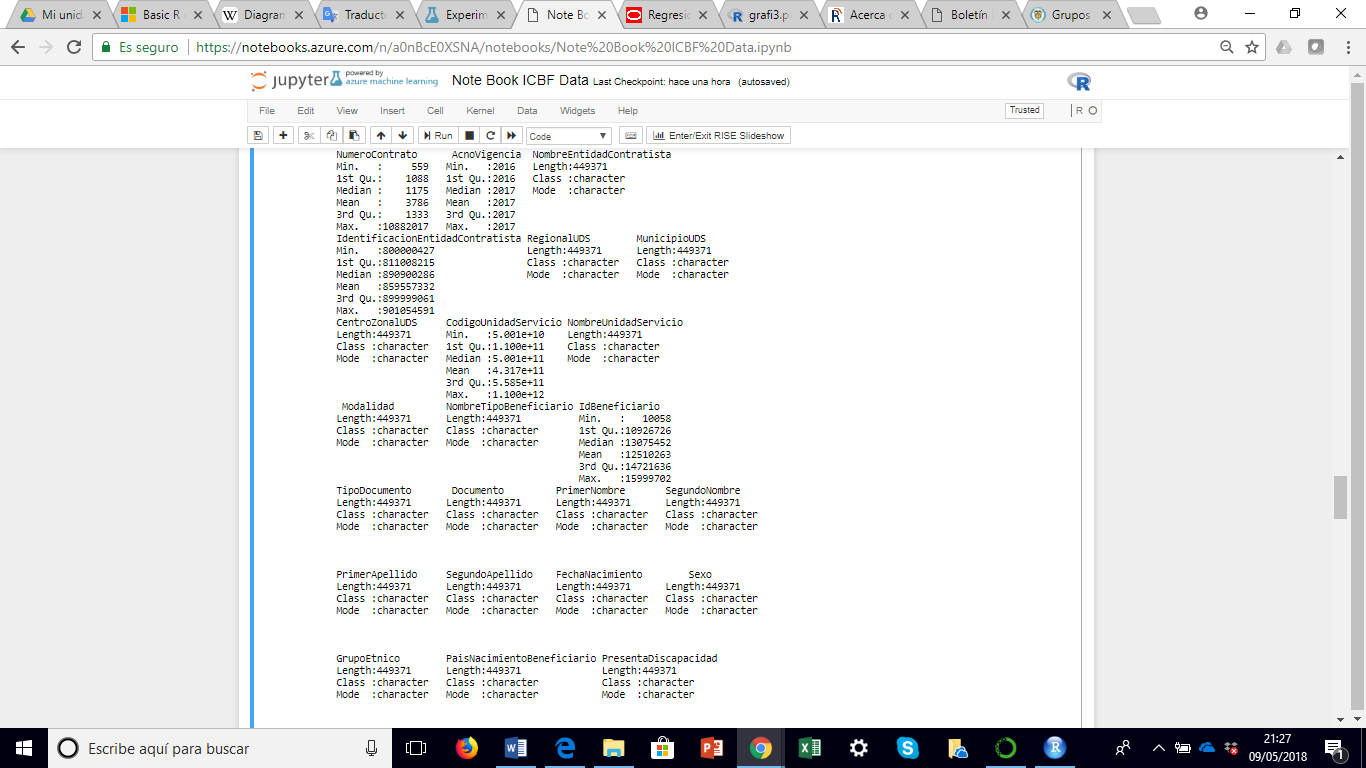
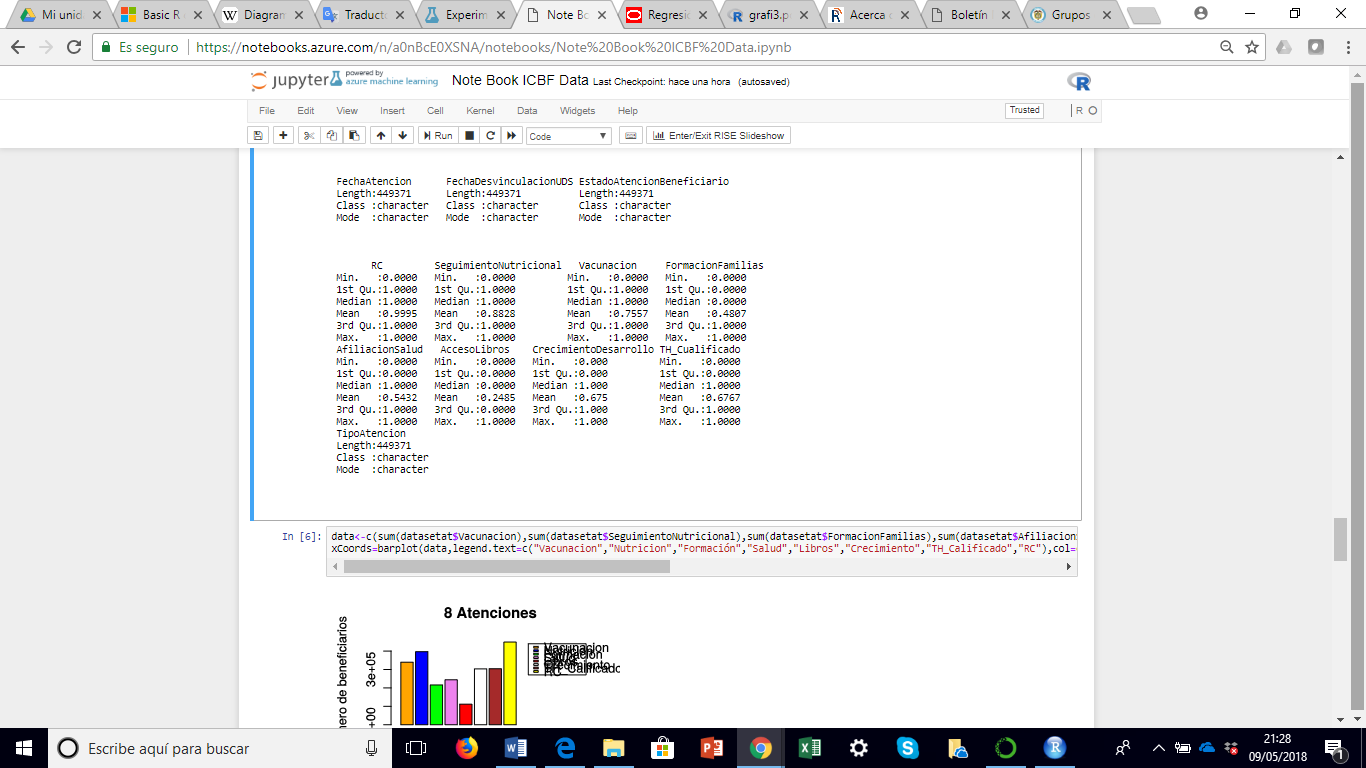


* 1. ¿Cuál es la relación entre el número total de cupos el valor de los contratos y el valor del valor del cada cupo?

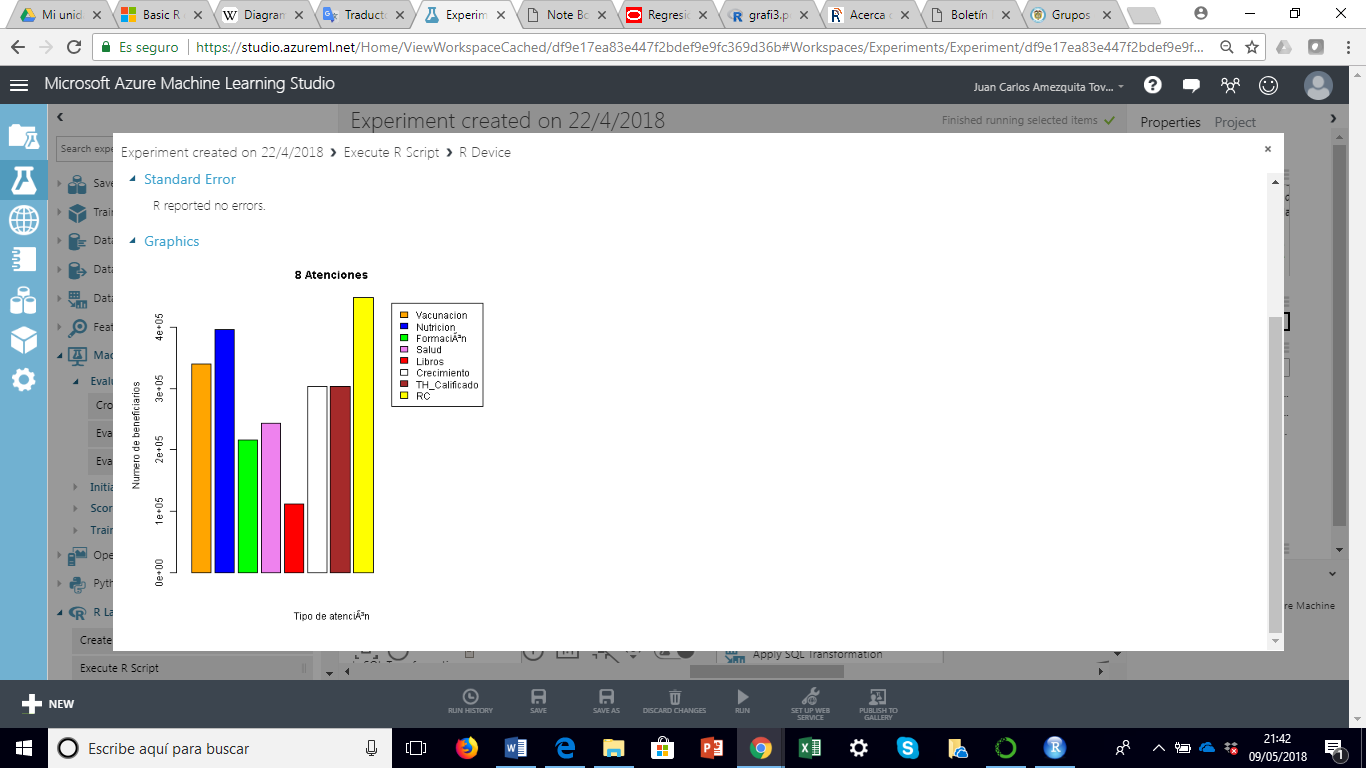


**DATASET ATENCIONES BOGOTÁ ANTIOQUIA**

1. Función summary en R.

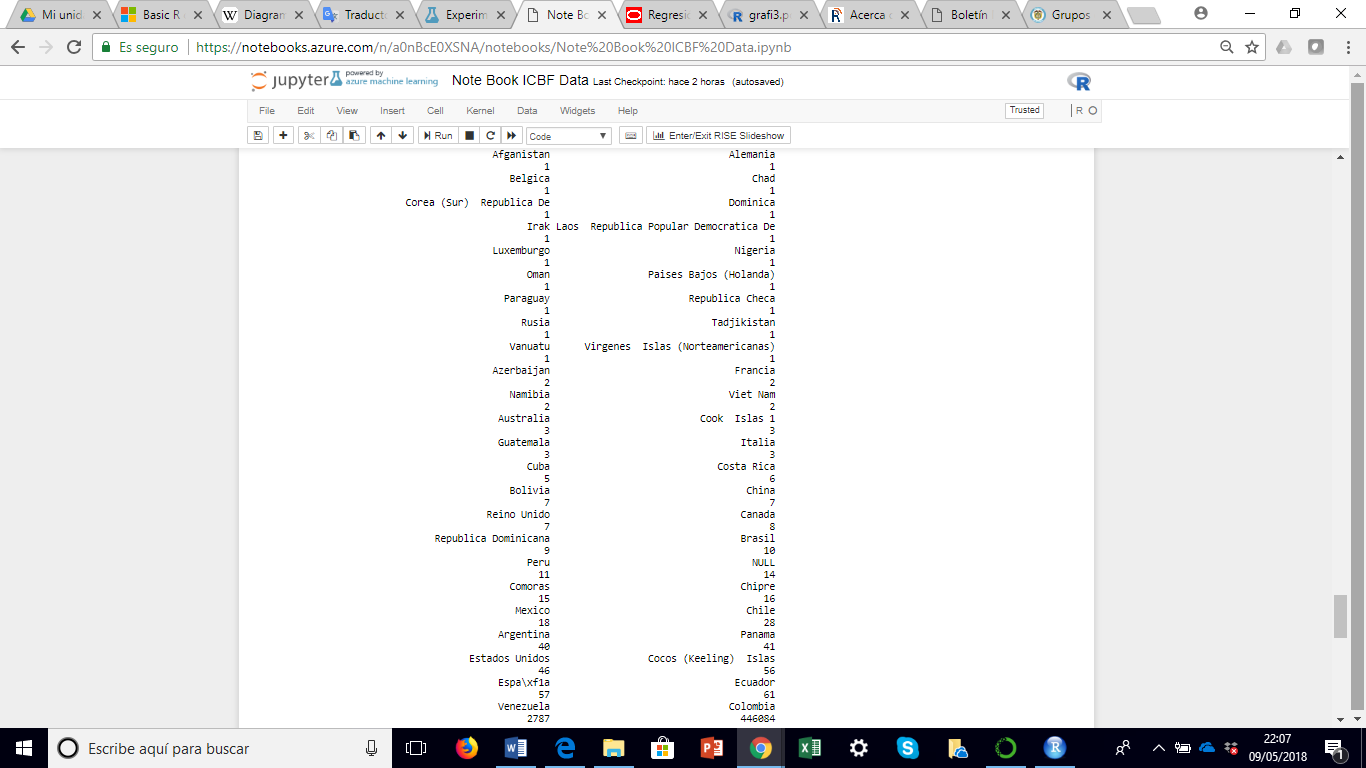
1. ¿Cuál es el porcentaje de cobertura para cada atención?



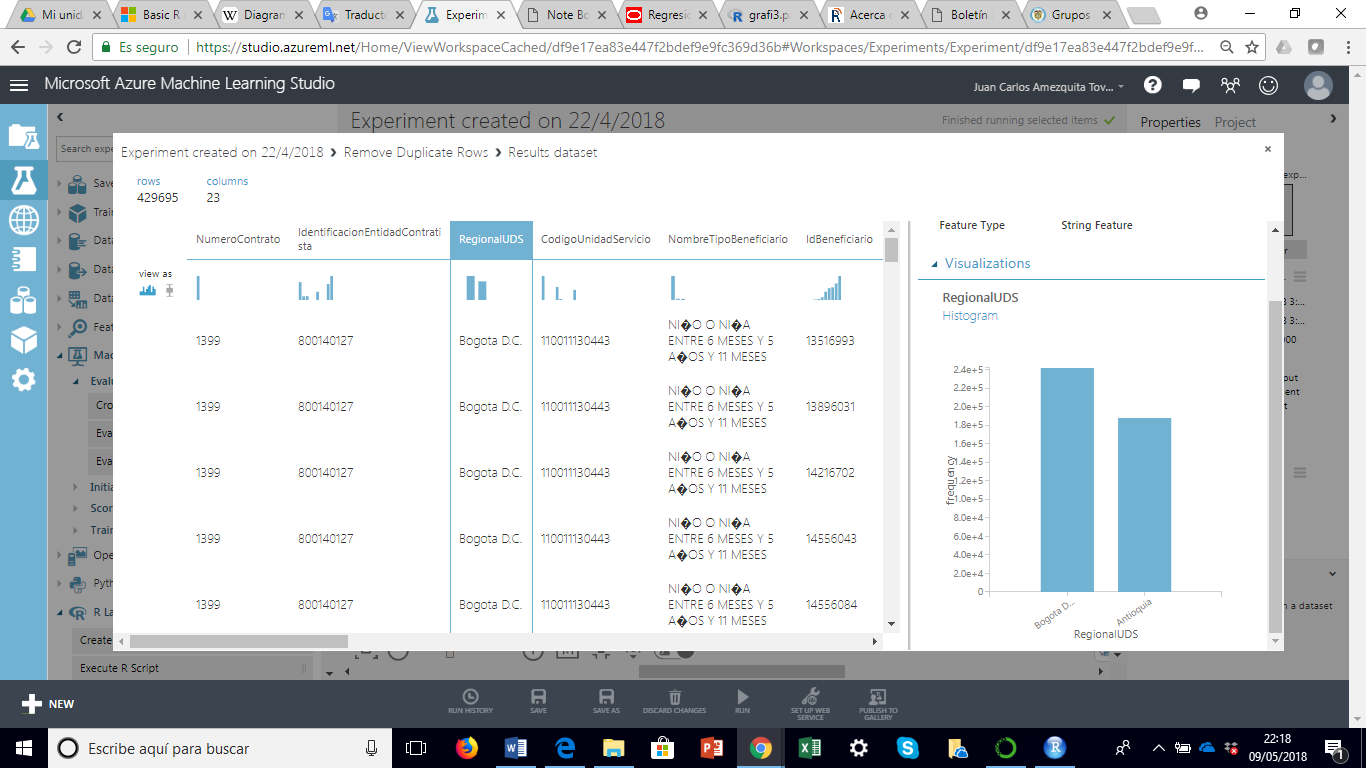
1. ¿Cuál es la distribución de tipo de atención?



1. ¿Cómo es la distribución de usuarios por país de nacimiento?



1. ¿Cuál es el número de usuarios?



Bogotá 241.852 (56%) Antioquia 187.843 (44%).

1. ¿Porque existen usuarios atendidos por más de un contratista?

El experto del dominio explico que es a causa de que algunos niños cuentan con servicios complementarios.

1. ¿Cuál es el porcentaje de niños extranjeros atendidos por el programa?

Se establece que el número de niños extranjeros es 3.123 que representa el 1% del total de la muestra.

1. ¿Cual es la distribución de niños con discapacidad?

Se establece: NULL 5941, R 428, S 311, SI 307, X 228.

# CONCLUSIONES

1. Se debe fortalecer la atención libros para lo cual se puede diseñar una estrategia en conjunto con Mintic para dotar a los centros de atención con computadores o tablet y de esta manera ambas entidades estarian fortaleciendo sus programas en cuanto a coberturas de cero a siembre y Colombia vive Digital.
2. Es fundamental hacer una revisión de la forma en la que el valor de los contratos es asignado, ya que en el análisis se evidencio que existen contratos cuyo valor no es proporcional al número de cupos ofrecidos en relación con los otros contratos ( COOPERATIVA MULTIACTIVA DE SAN ANTONIO DE PRADO COOMULSAP).
3. Es importante determinar las causas por las cuales Antioquia con un valor de asignación de recursos del casi el doble a Bogotá tiene 303.819 cupos disponeibles sin embargo el número de usuarios es de 187.343.
4. Segun el censo del 2005 para solo Bogotá existian al rededor de 27.009 niños con discapacidad pero de acuerdo a los datos suministrados en la tabla de atenciones existen 618 niños, por lo cual se debe evaluar si en realidad es un problema en la cobertura que esta ofreciendo el ICBF a este tipo de población o es una falla en el proceso de registro de los mismos ya que aparecen 5941 con ese atributo en NULL.
5. En cuanto a grupos étnicos también se aprecia un valor de cobertura muy pequeño 6.684 usuarios (1.56%) de la muestra por lo cual también se recomienda evaluar las causas para que este valor ya que de acuerdo al censo del 2005 el 14.4% de las personas pertenece a esta población.