 **ESCUELA POLITÉCNICA NACIONAL**

ESCUELA DE FORMACIÓN DE TECNÓLOGOS

**BASES DE DATOS MULTIDIMENSIONALES**

|  |  |
| --- | --- |
|  | |
| PERÍODO ACADÉMICO: | 2018-B |
|  | |

**Examen supletorio**

|  |  |
| --- | --- |
|  | |
| TÍTULO:  **ELK Y Batch** | |
| **ESTUDIANTE** | |
| Daniel Flores | |
|  | |
|  | |
|  | |
|  | |
| FECHA DE ENTREGA: | 12 / 02 / 2019 |
| CALIFICACIÓN OBTENIDA: |  |
|  | |

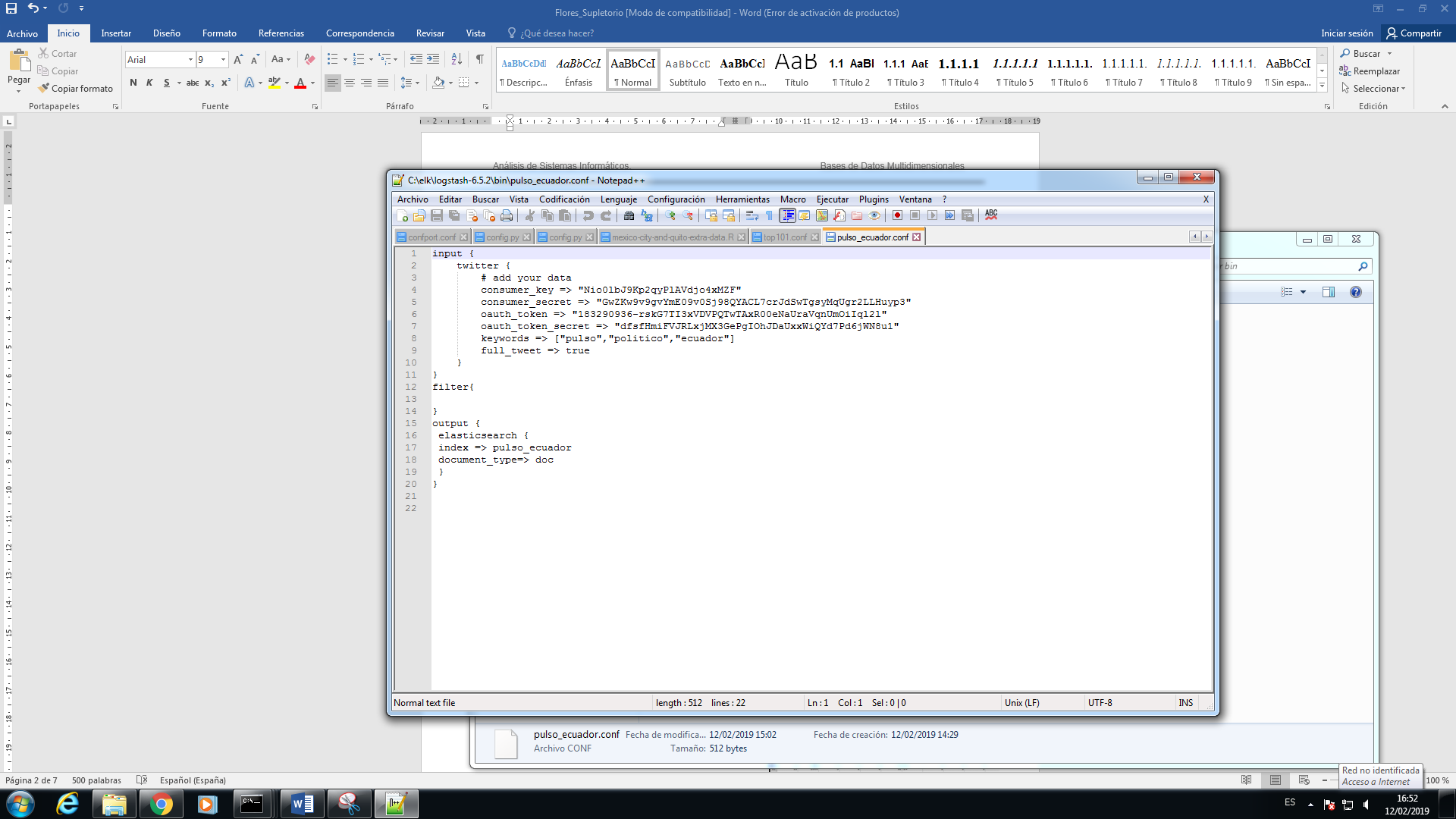
Recolección, almacenamiento y visualización de datos por medio de ELK y bath.

Flores Daniel

e-mail: [danielflorestamayo@gmail.com](mailto:danielflorestamayo@gmail.com)

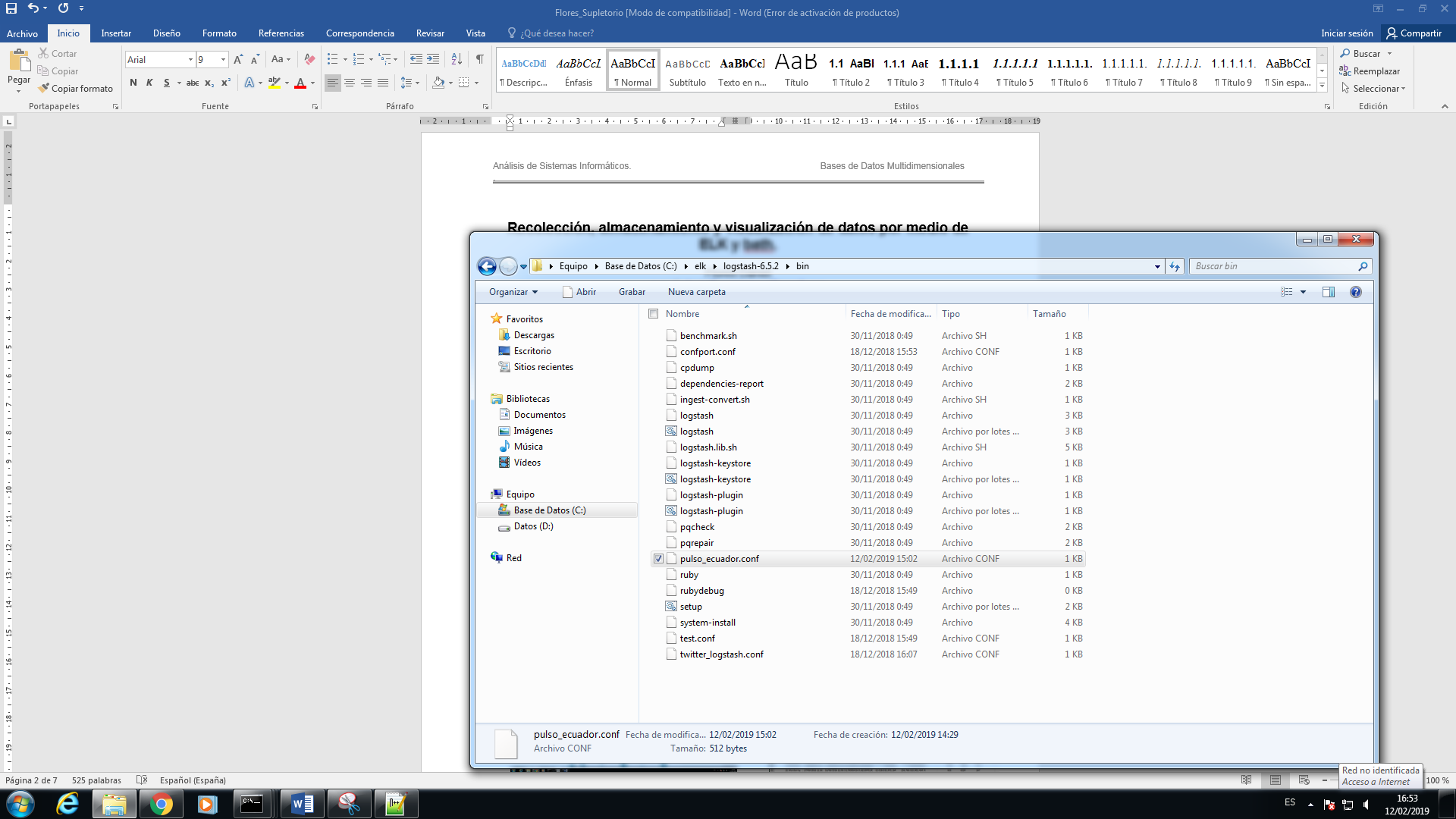
# DESARROLLO

Tiempo real:

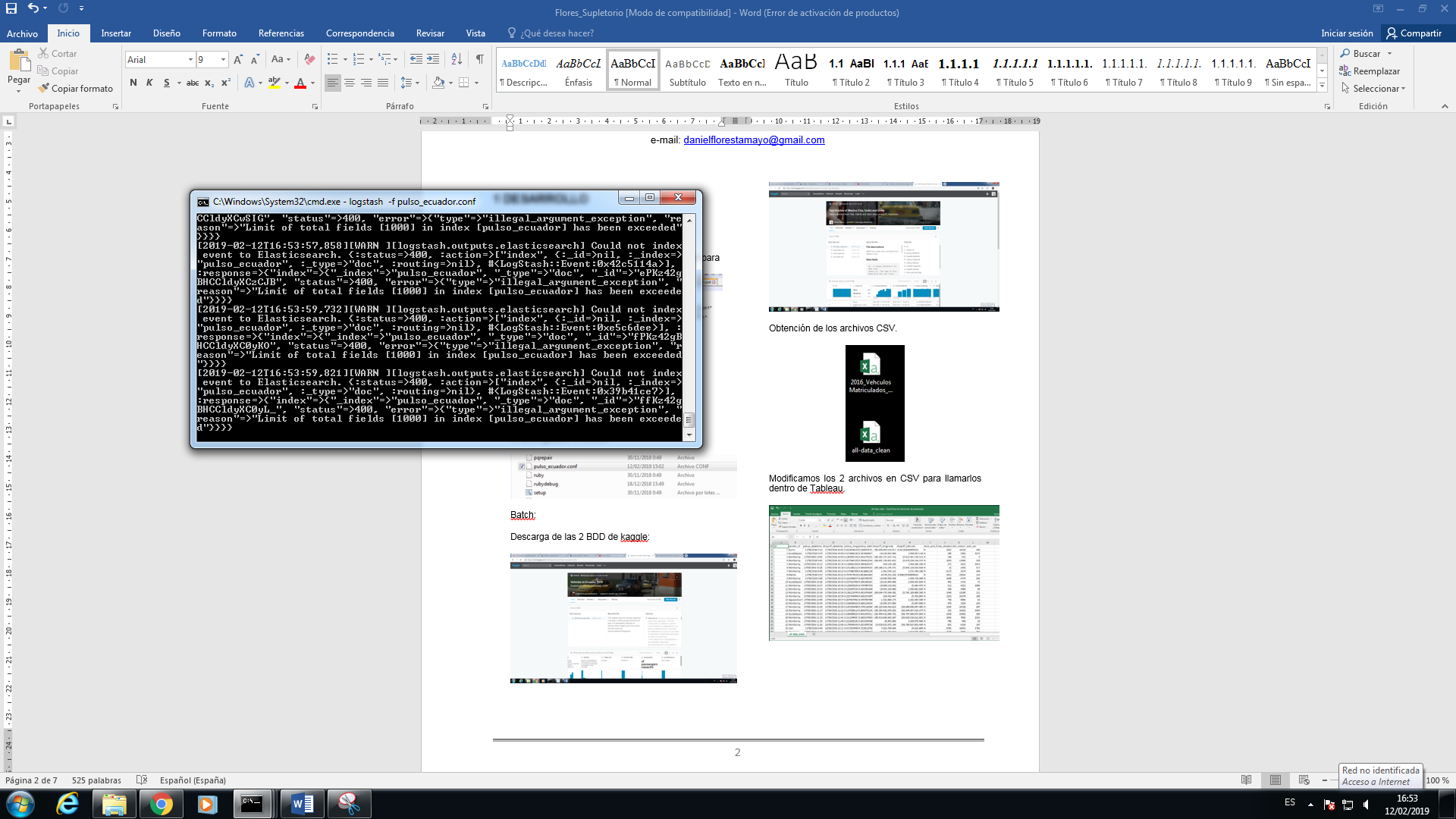
Para la extracción de los tweets en tiempo real para el pulso político se usa el siguiente script. 

Lo copiamos dentro e este directorio:

C:\elk\logstash-6.5.2\bin



Lo ejecutamos dentro del CMD correspondiente y vemos como los tweets comienzan a recogerse.

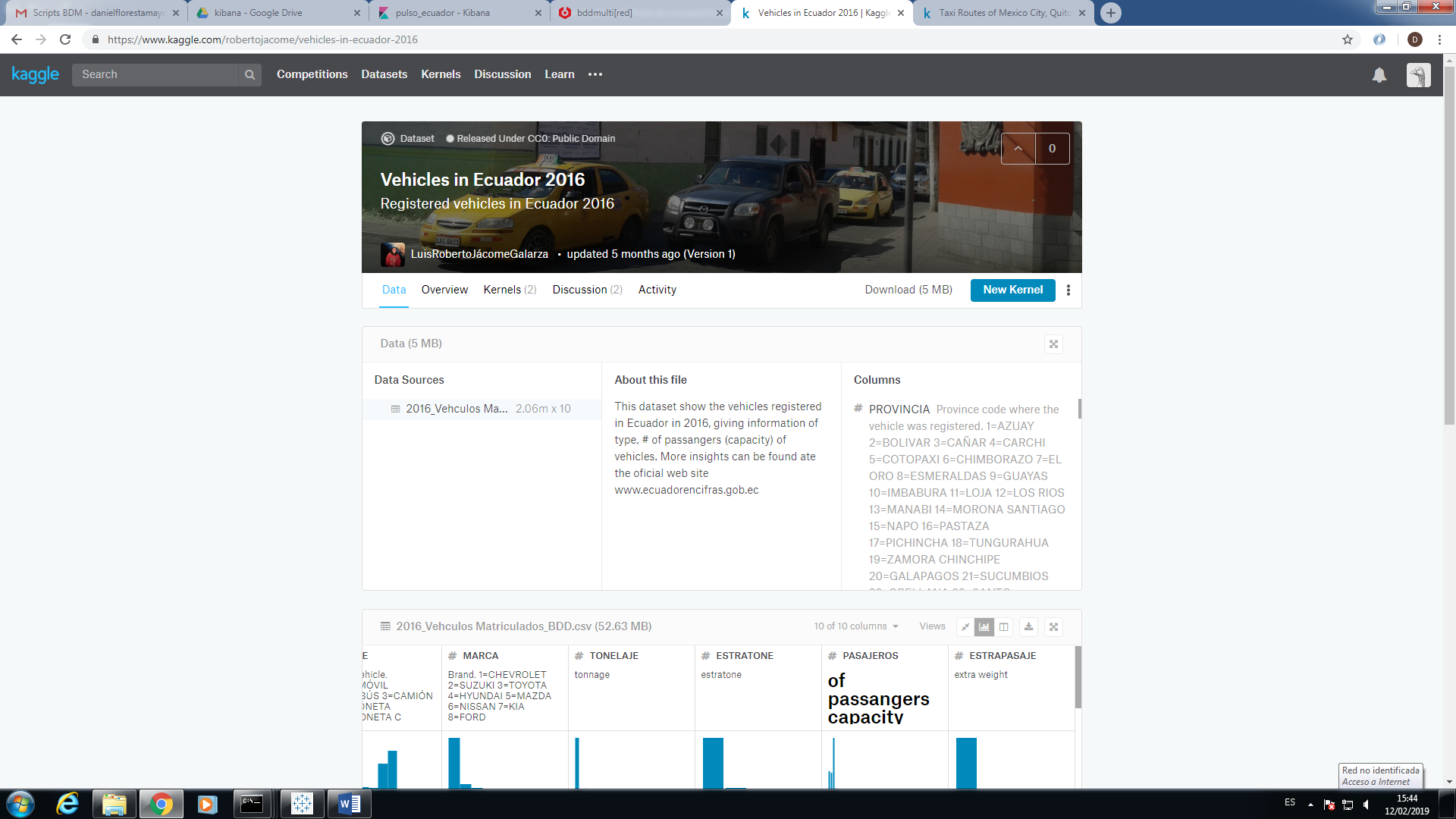


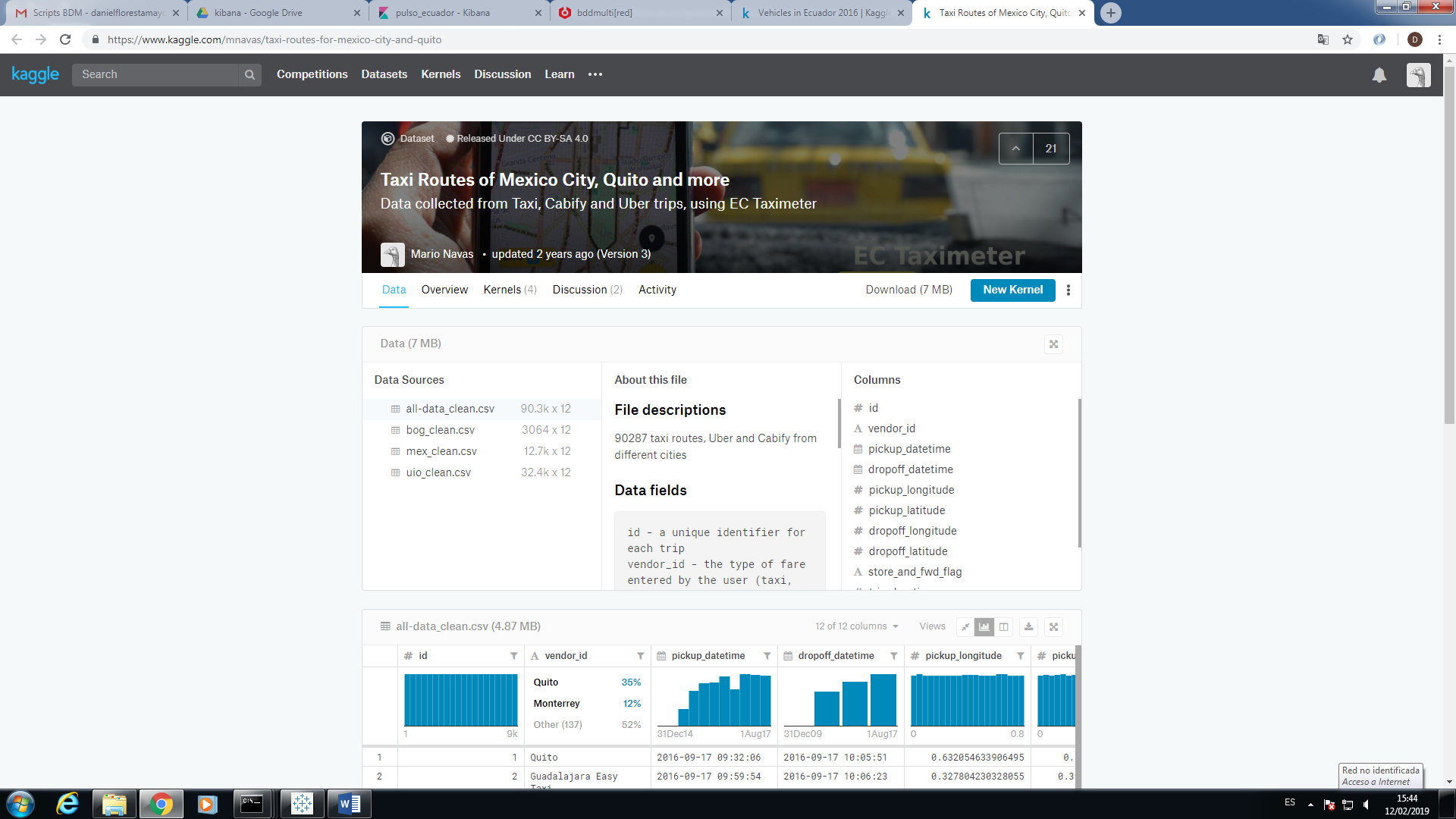
Dentro del Cerebro se ve como se crea y se cosechan los registro conforme el tiempo avanza.



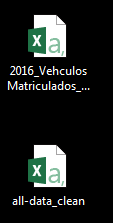
Batch:

Descarga de las 2 BDD de kaggle:

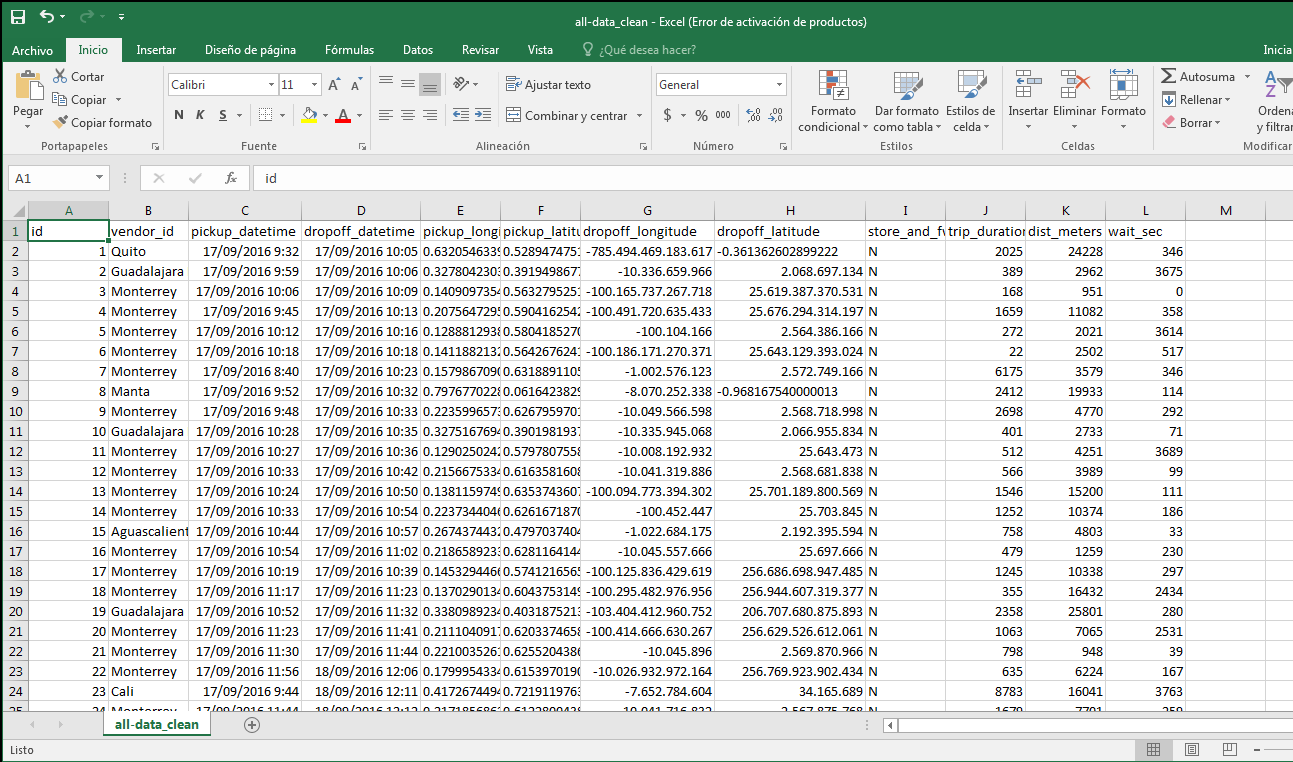


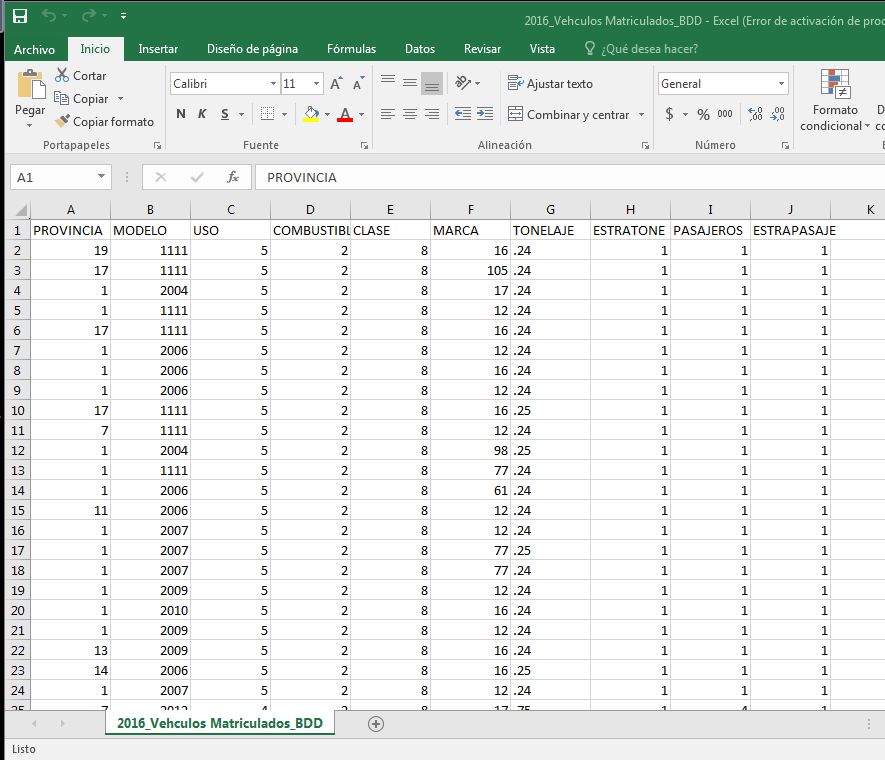


Obtención de los archivos CSV.

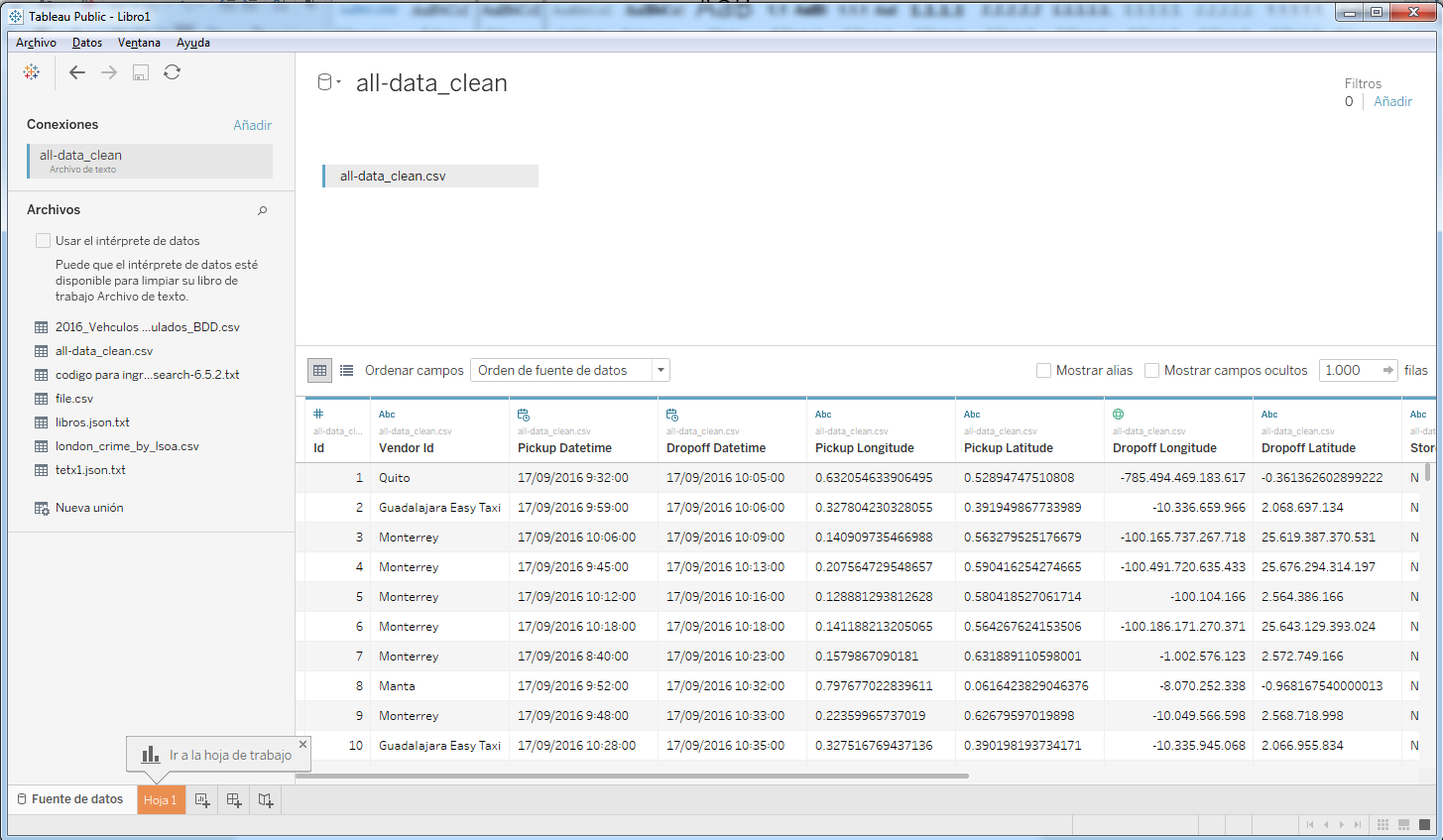


Modificamos los 2 archivos en CSV para llamarlos dentro de Tableau.

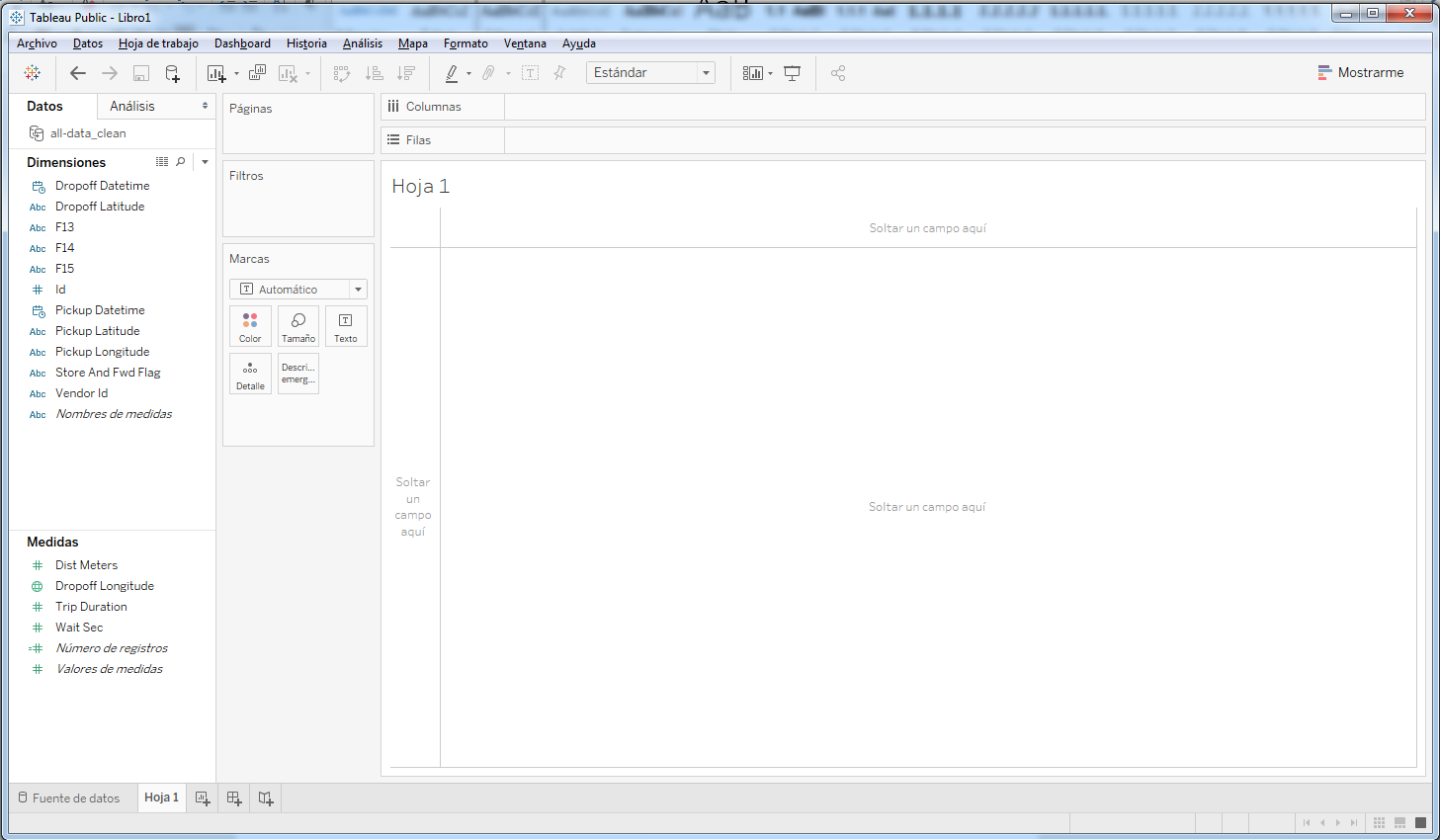




Abrimos Tableau y cargamos a nuestro primer archivo: all-data\_clean

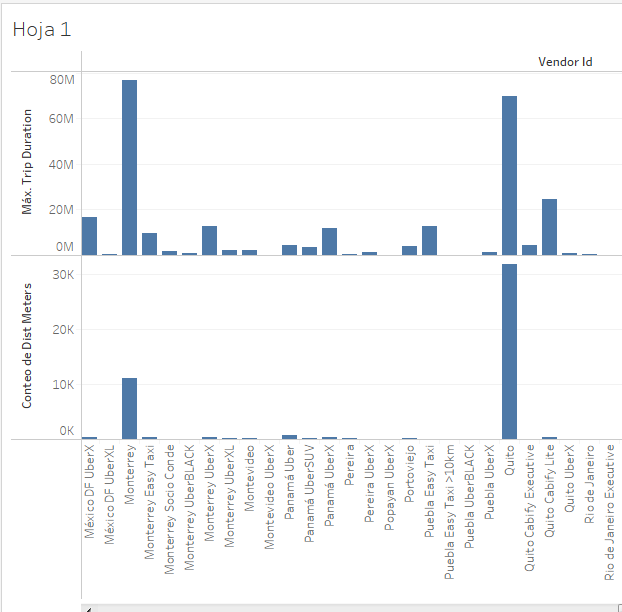


Esperamos a que carguen los atributos de nuestra BDD.

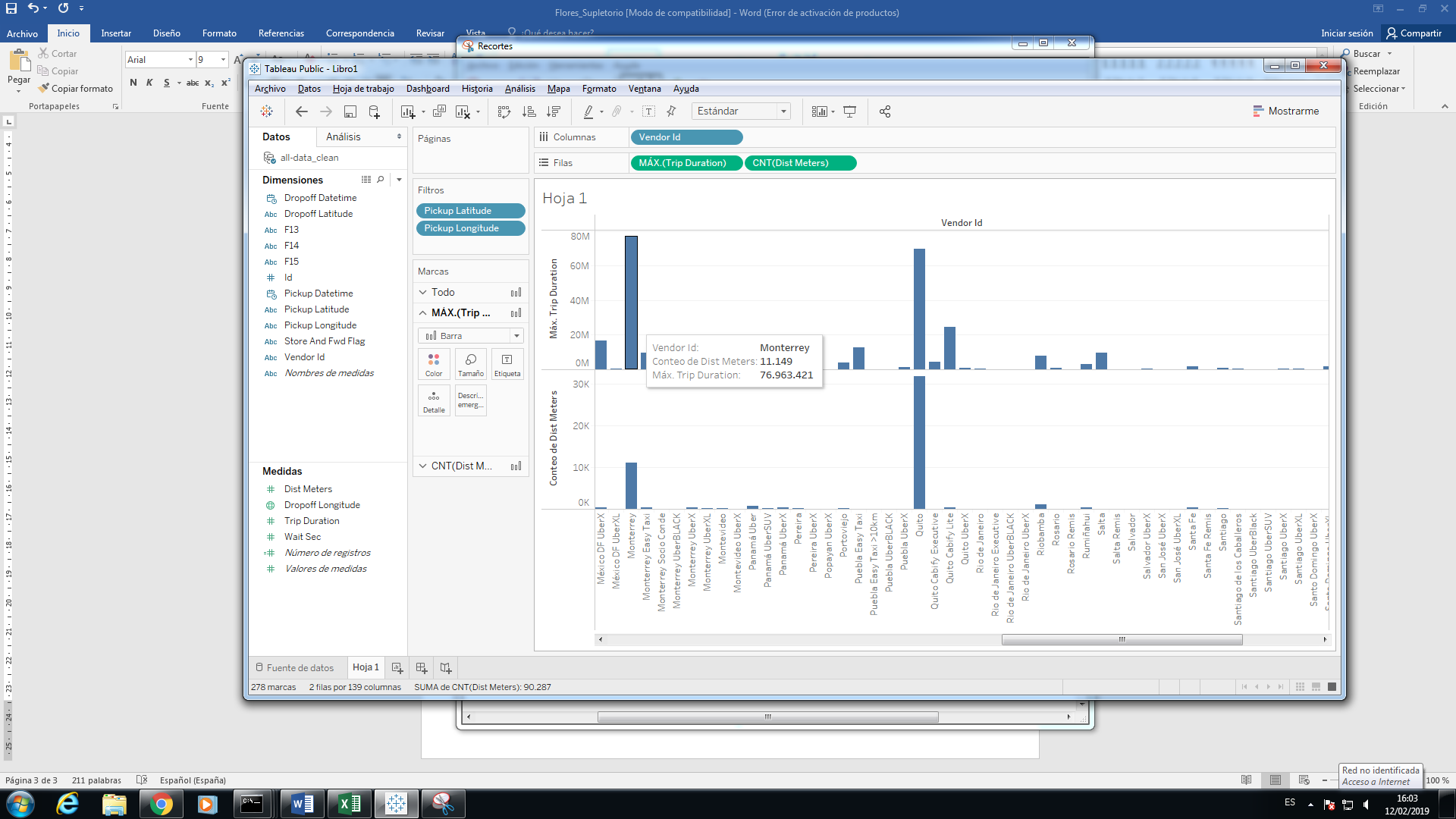


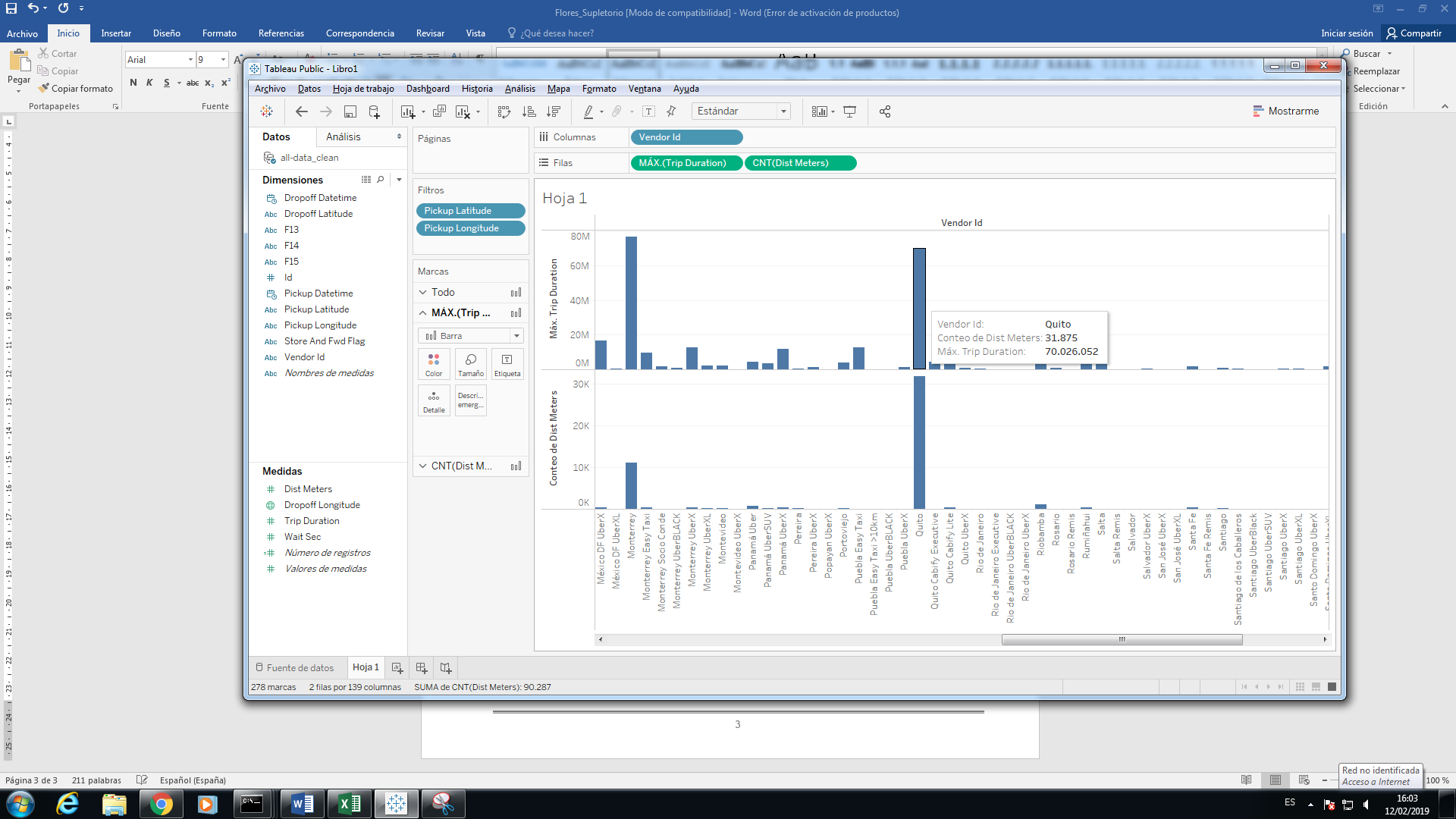
**Grafico\_1**

Para el primer caso de estudio se escogieron los atributos de: vendor\_Id, dist meters, Trip duration. Esto con la finalidad de la duración del viaje debido a los metros totales que existe entre cada viaje realizado.



Se hace una comparación entre los datos del viaje en Quito y Monterrey: su duración y su tiempo de duración.

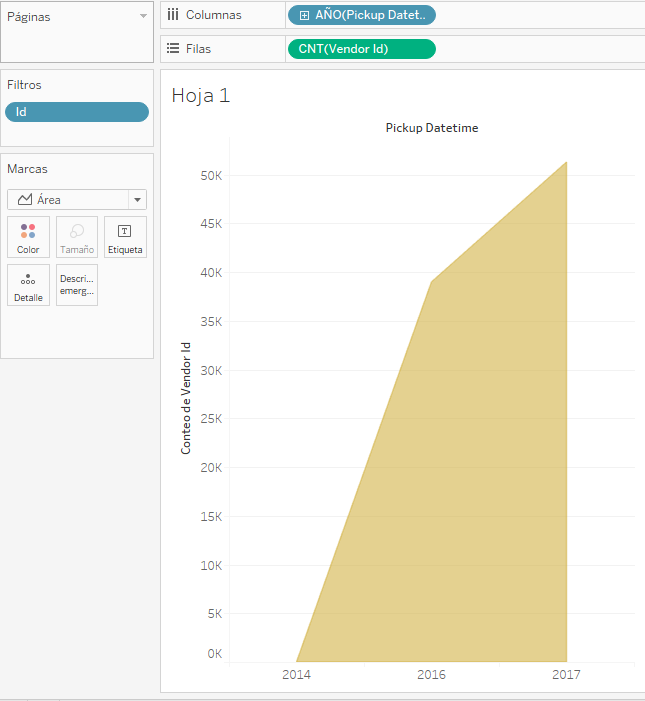


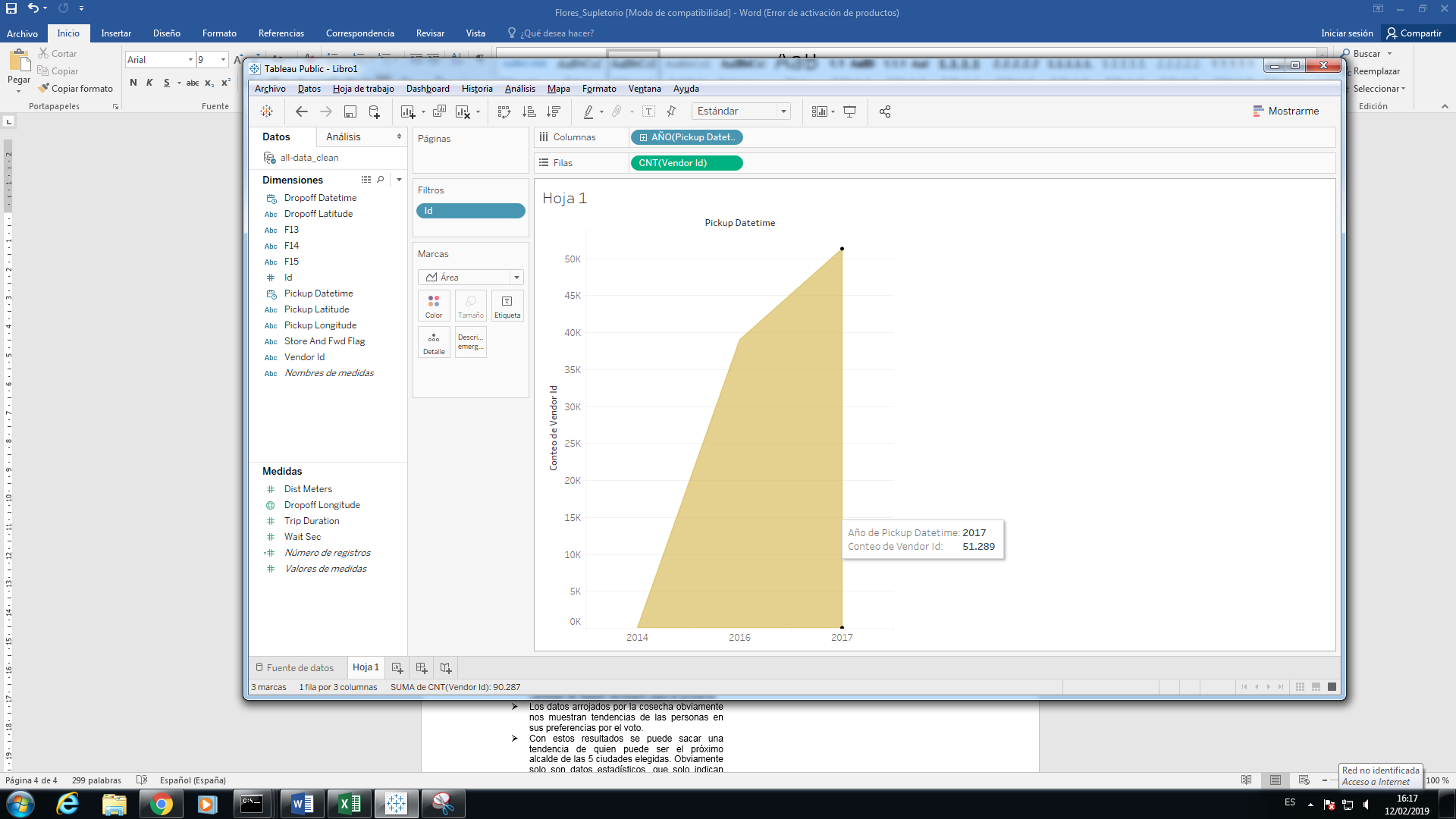


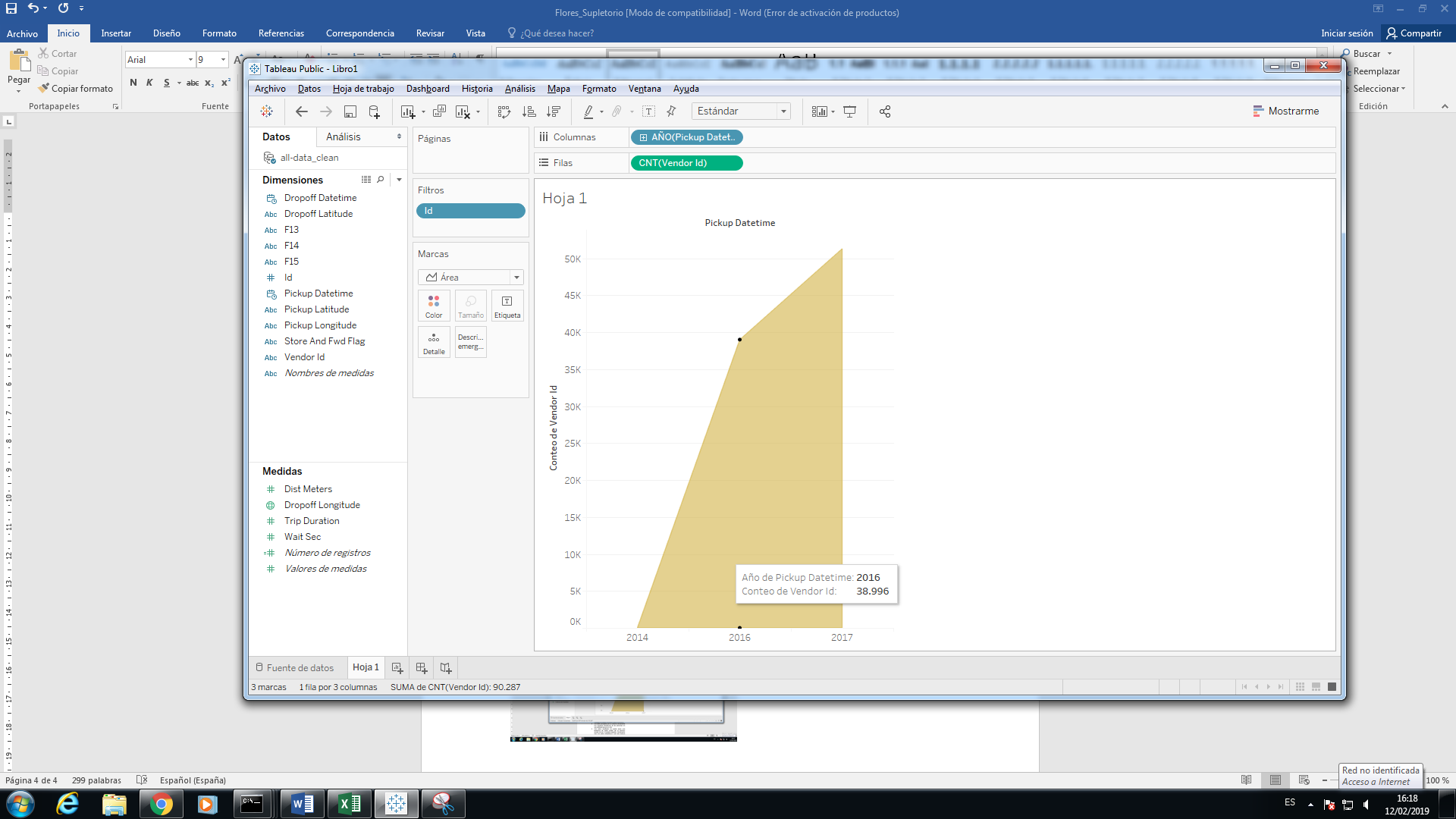
**Conclusión:** Vemos que a pesar de que la distancia del viaje de Quito es casi el triple de la de Monterrey. La duración del viaje de Monterrey y Quito es casi similar. Existe una gran variación entre la distancia. Esto debido a la latitud y la longitud que existe entre los viajes realizados.

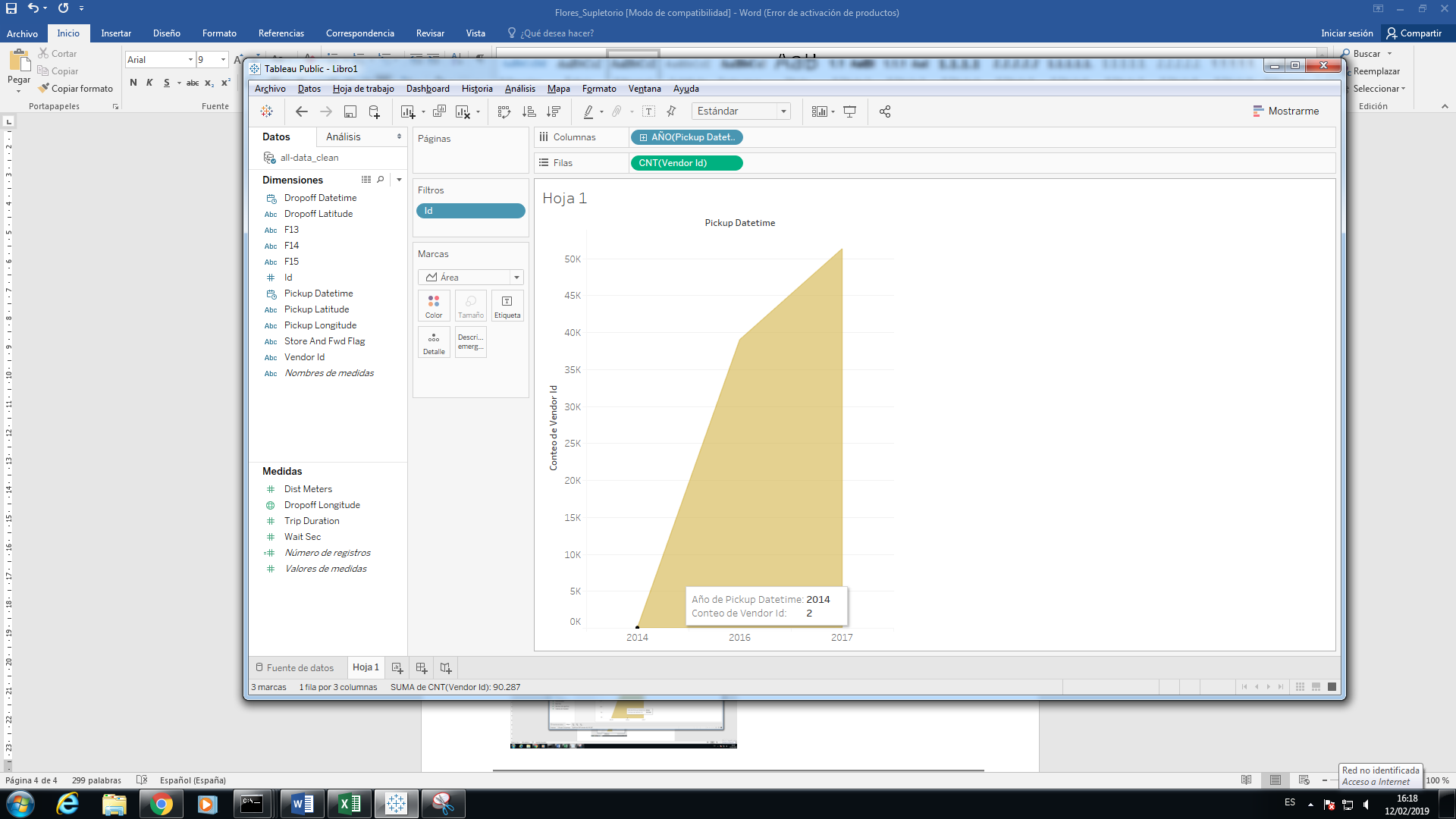
**Gráfico\_2**

Para el segundo gráfico se escogen los atributos de: Pickup Datetime, vendor Id.





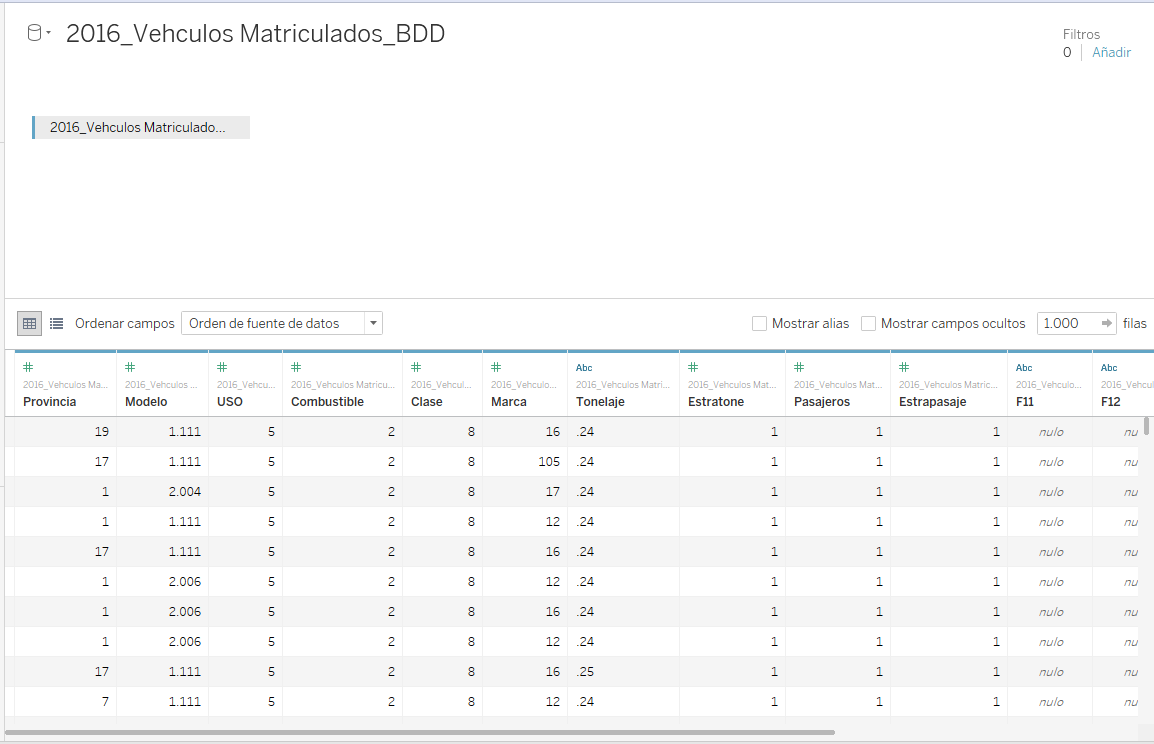




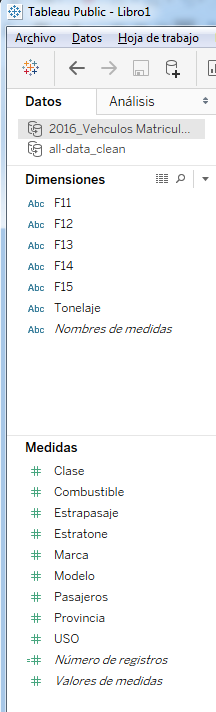
Conclusión: Como consecuencia de escoger estos 2 atributos se nos muestra un incremento conforme los años a transcurren del 2014 al 2017. En el año 2014 existe un total de 2 registros. En el 2016 existe un aumento pronunciado de 38 996 registro. En 2017 existe un total de 51 289 registros. Lo que se entiende es que en un futuro para los años 2018 a 2022, existirá un aumento igual de pronunciado en los viajes (rutas) realizados en lo que va del rango de los años.

BDD Vehiculos-in-ecuador-2016

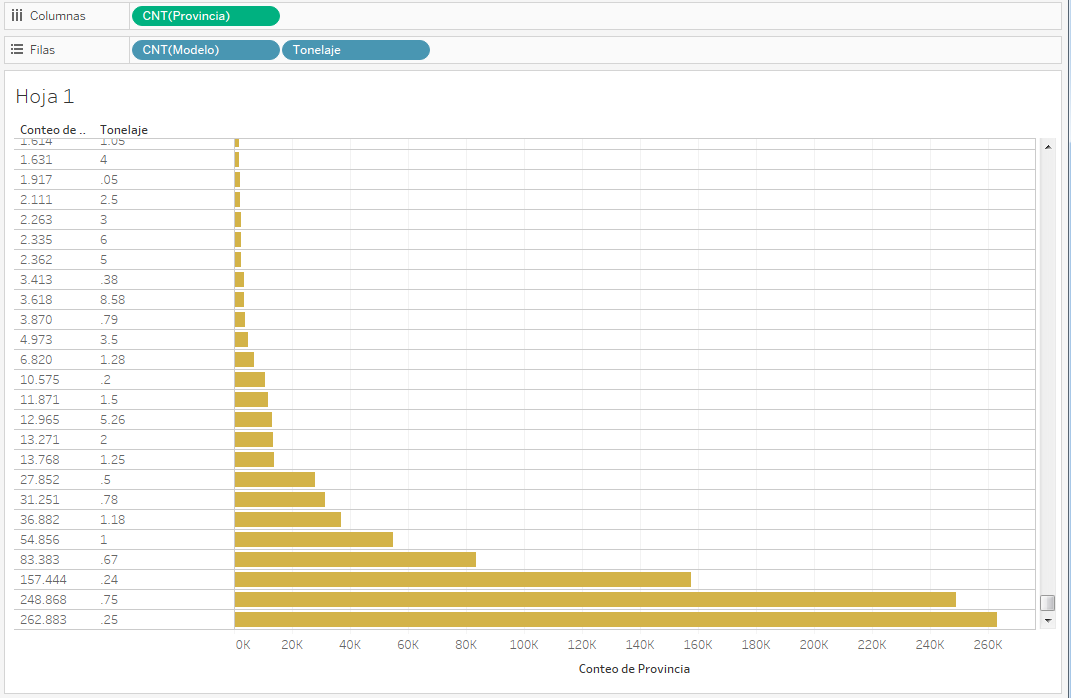
Cargamos el archivo CSV, dentro de tableau.

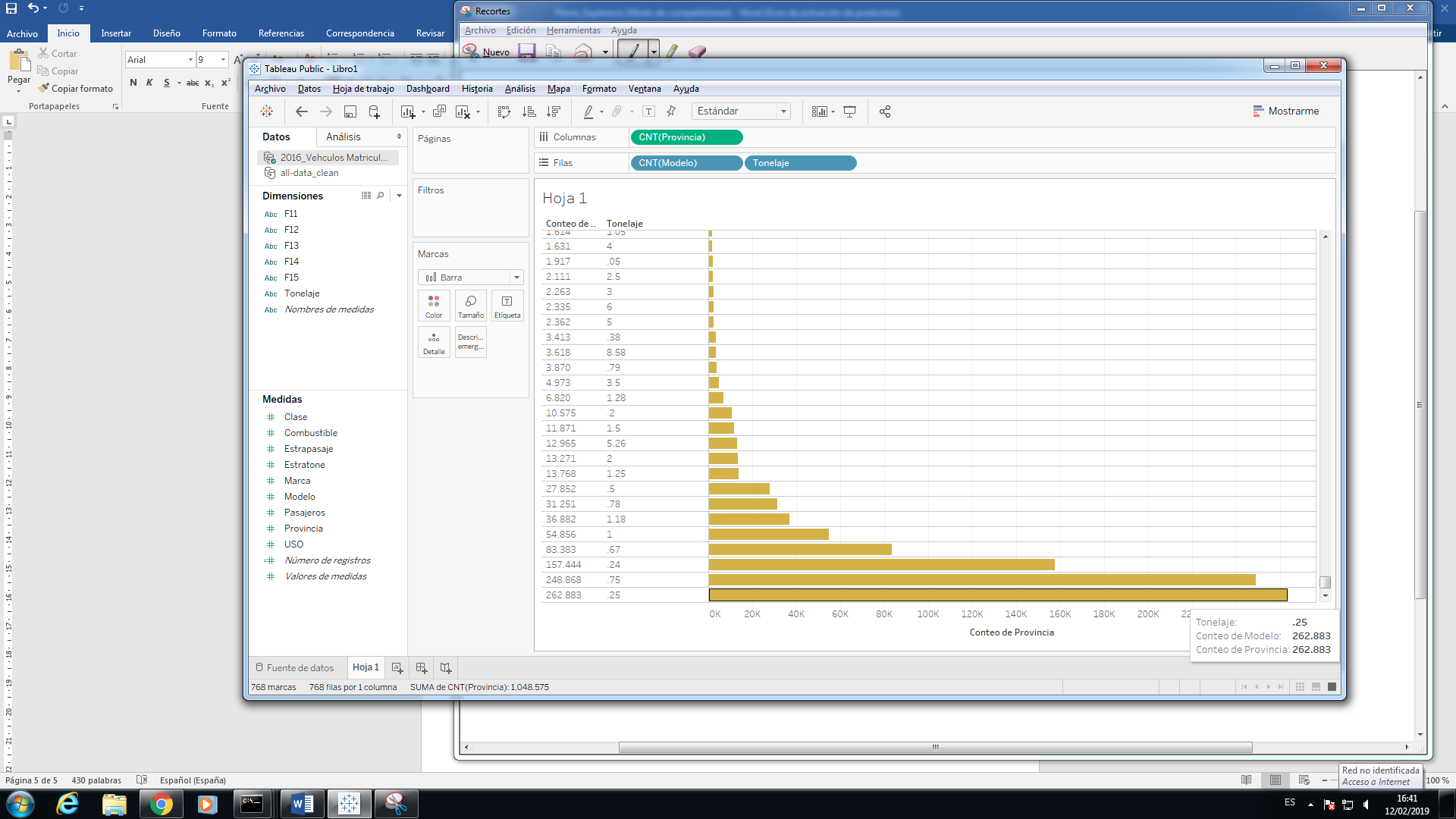


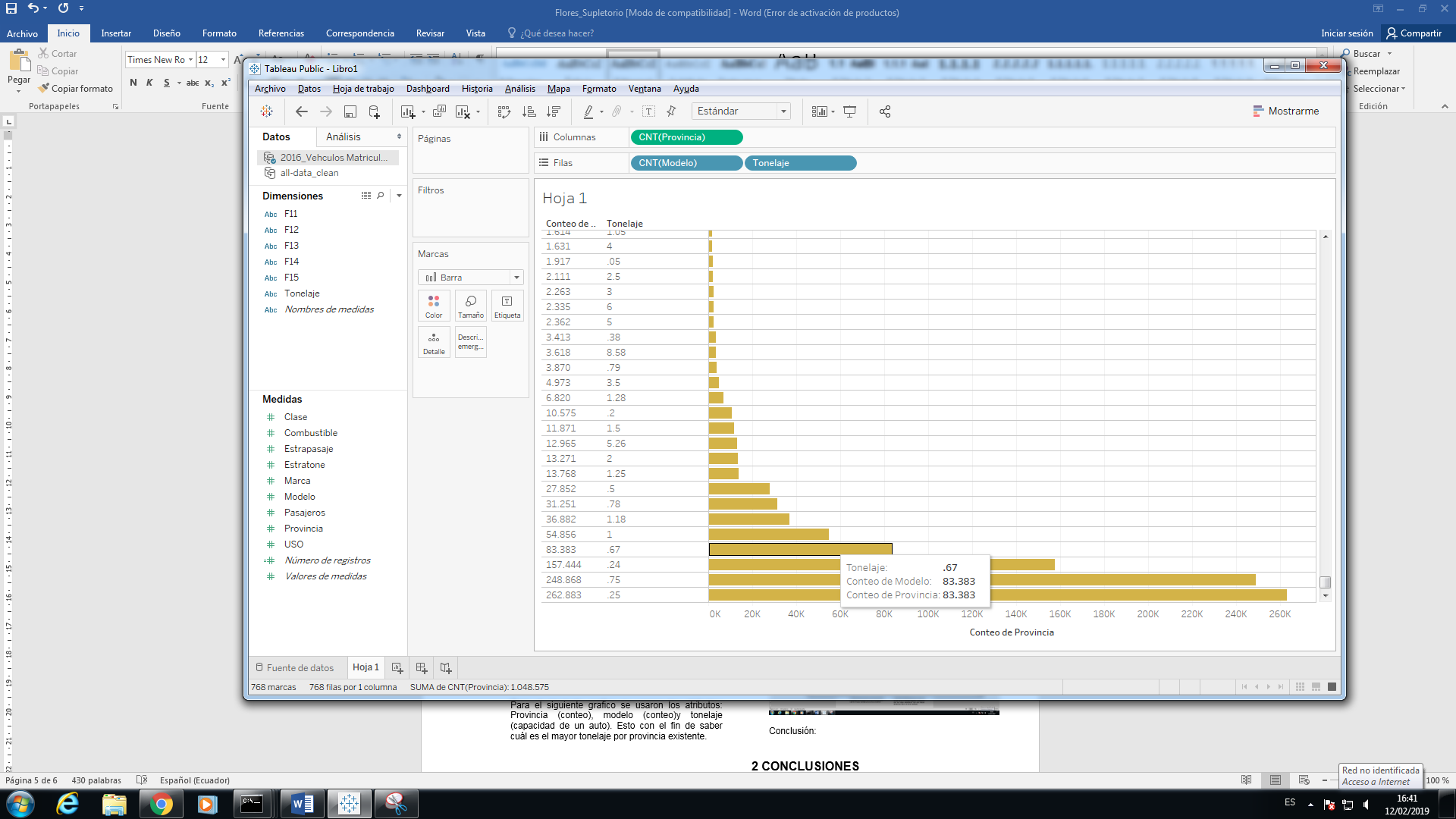
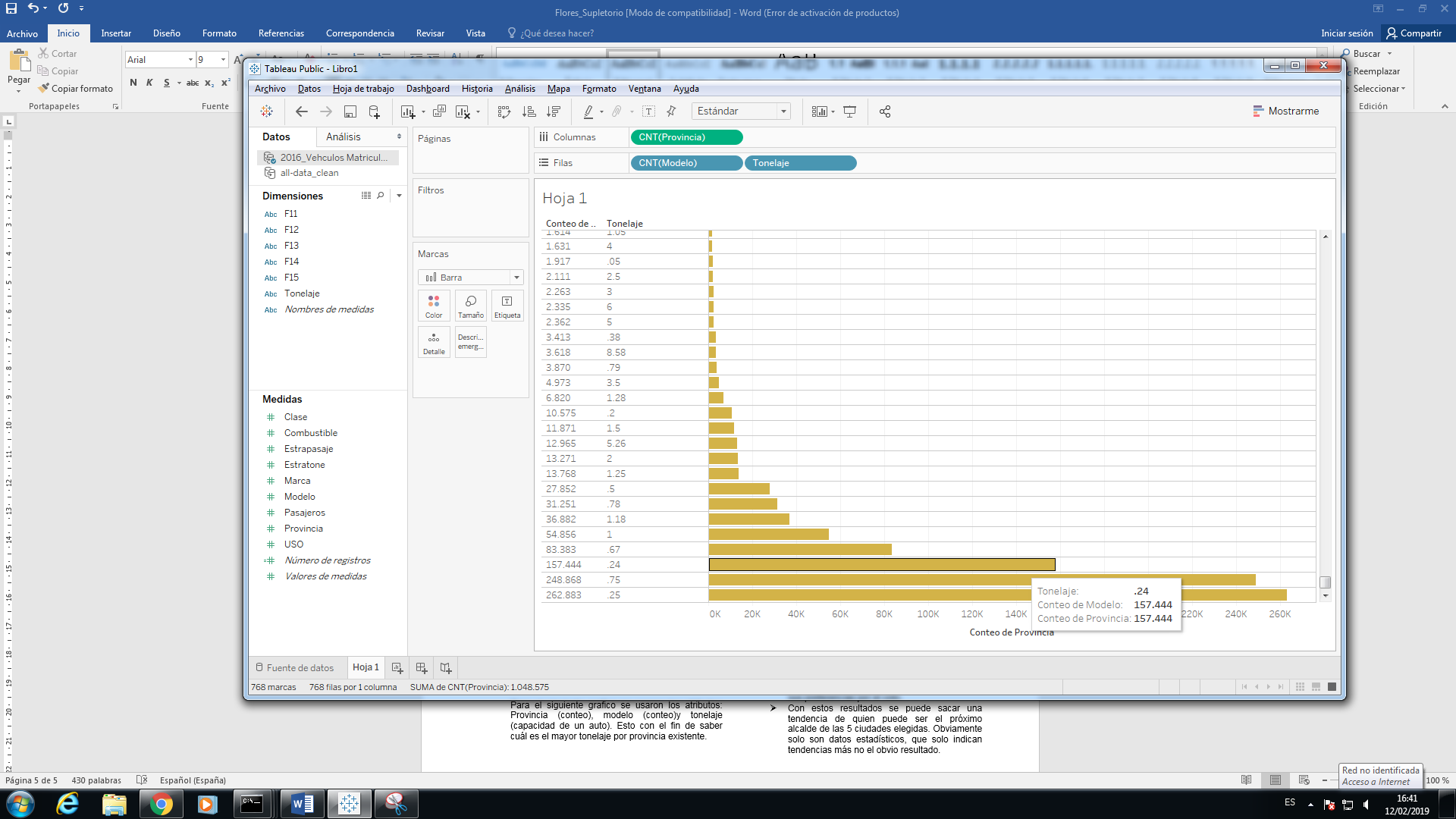
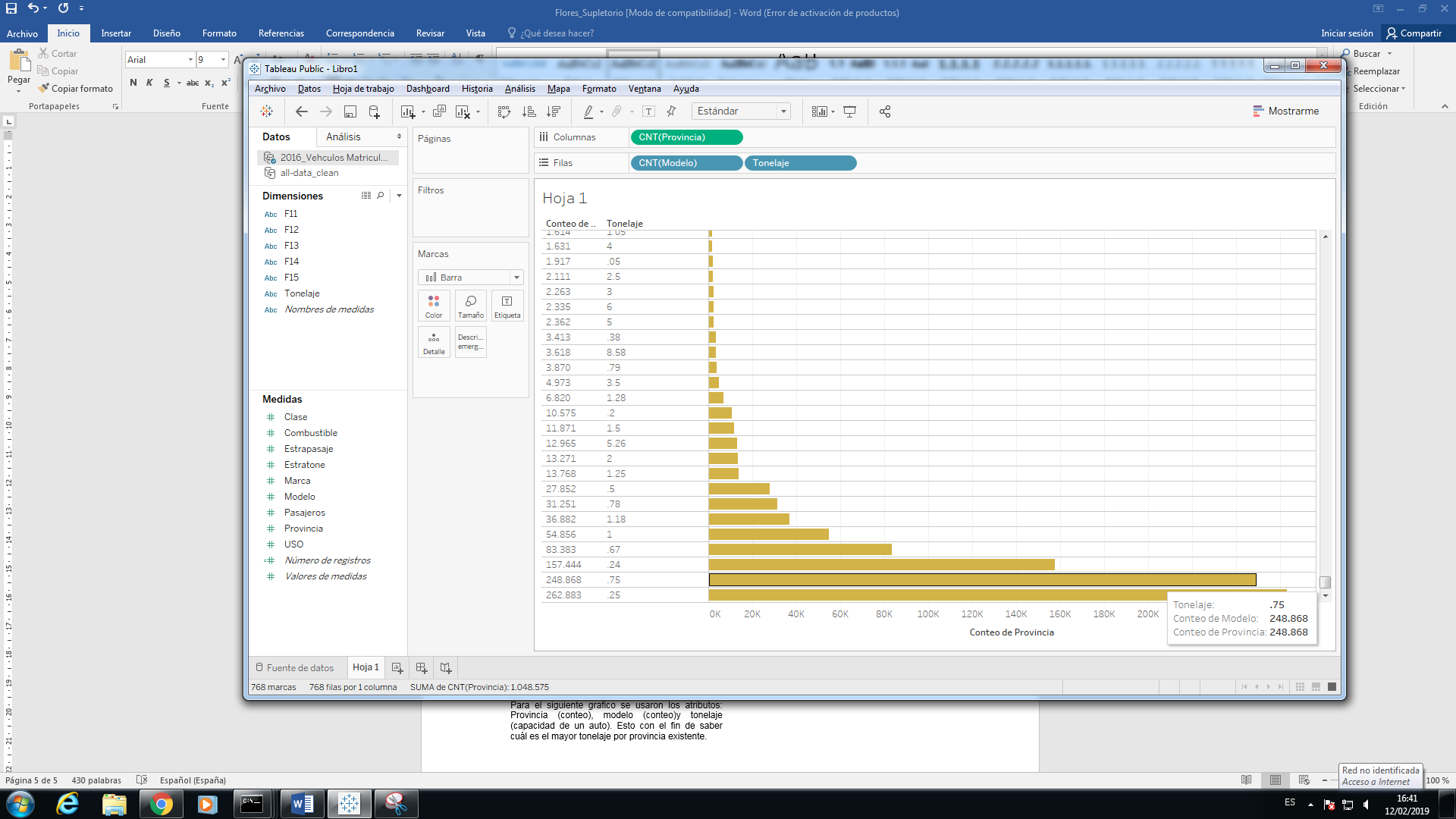
Tenemos los siguientes atributos.



Para el siguiente grafico se usaron los atributos: Provincia (conteo), modelo (conteo)y tonelaje (capacidad de un auto). Esto con el fin de saber cuál es el mayor tonelaje por provincia existente.



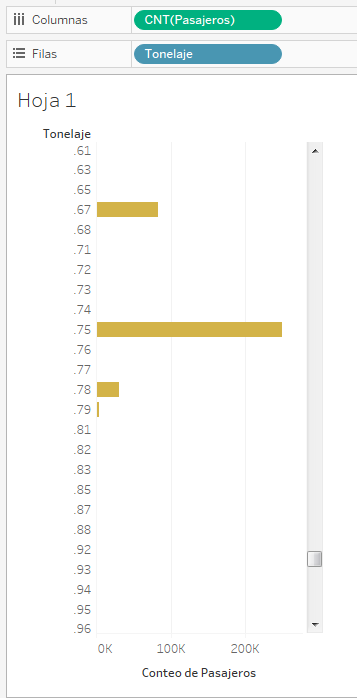


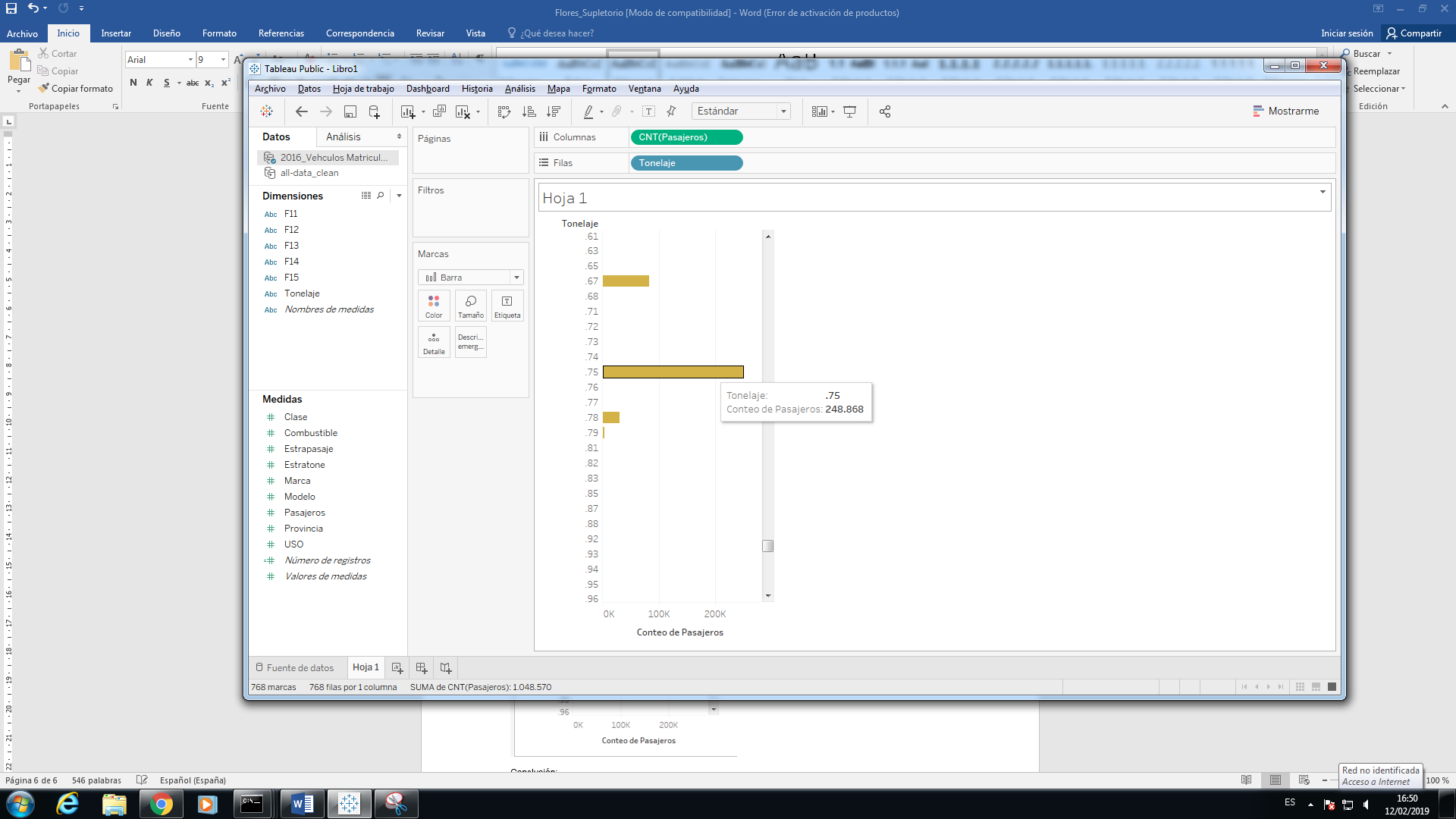


Conclusión: Tenemos un Top de los 4 tonelajes más contados en las provincias y modelos. En el tonelaje de .67 tiene menor número de modelos en circulación con un total de 83 383 modelos. Le sigue el tonelaje de .24 con un total de 157 444 modelos. En segundo puesto esta el tonelaje de .75 con un total de 248 868 modelos. En primer lugar esta el tonelaje de .25 con un total de 262 883 modelos contados dentro de las provincias.

Grafico\_2

Para este gráfico se tiene como parámetros el conteo de pasajeros vs el tonelaje que estos representan en los vehículos.





Conclusión:

Se tiene un mayor numero de pasajeros en el tonelaje de tipo .75 con un total de 248 868 pasajeros.

# CONCLUSIONES

* Para la recolección de los tweets se necesita de mucho tiempo, para ello medimos la cantidad de tweets necesario para el proyecto.