**PRÁCTICA FINAL**

**EXPLORACIÓN Y VISUALIZACIÓN DE DATOS**

**-. TABLEAU .-**

# **Índice de Contenidos**

1. Índice de Contenidos · - ·· - · - ·· - · - 1
2. Datos de la Práctica · - ·· - · - ·· - · - 2
3. Presentación y Objetivo · - ·· - · - ·· - · - 2
4. Origen de los Datos y Tratamiento Inicial · - ·· - · - ·· - · - 3
5. Carga del Dataset en Tableau · - ·· - · - ·· - · - 4
6. Diseño del Dashboard · - ·· - · - ·· - · - 5
7. Marco 1 · Evolución Anual del Promedio

del Precio Total Día por Distrito · - ·· - · - ·· - · - 5

* 1. Campos Calculados - ·· - · - ·· - · - ·· - · 5
  2. Elaboración de la Gráfica - ·· - · - ·· - · - ·· - · 6

1. Marco 2 · Evolución Anual del Promedio

del Precio Total Día por Tipo de Inmueble

y Alquiler · - ·· - · - ·· - · - 6

* 1. Elaboración de la Gráfica - ·· - · - ·· - · - ·· - · 7

1. Marco 3 · Volúmenes de Inmuebles y Reviews

por Distrito · - ·· - · - ·· - · - 7

* 1. Elaboración de la Gráfica - ·· - · - ·· - · - ·· - · 8

1. Marco 4 · Geoposicionamiento de Distritos · - ·· - · - ·· - · - 9
   1. Elaboración de la Gráfica - ·· - · - ·· - · - ·· - · 9
2. Marco 5 · Volumen de Reviews por Tipo

de Inmueble y Alquiler & Score Promedio · - ·· - · - ·· - · - 10

* 1. Elaboración de la Gráfica - ·· - · - ·· - · - ·· - · 10

1. Composición Final del Dashboard · - ·· - · - ·· - · - 11
   1. Elaboración de la Gráfica - ·· - · - ·· - · - ·· - · 10
2. Agradecimientos y Comentarios Finales · - ·· - · - ·· - · - 12

# **Datos de la Práctica**

|  |  |
| --- | --- |
| **Bootcamp IV Big Data & Machine Learning - KeepCoding** | |
| **Asignatura** | Exploración y Visualización de Datos con Tableau |
| **Profesor** | José Antonio Gascón |
| **Alumno** | Francisco Javier Gonzálvez Chico |
| **Título** | Analítica de Precios y Reviews -- Inmuebles AIRBNB Madrid |

# **Presentación y Objetivo**

El objetivo de la siguiente práctica será la construcción y descripción de un Dashboard, apoyado sobre la aplicación de Tableau, en el que mostrar una visión analítica de precios y reseñas sobre los inmuebles registrados en Madrid dentro del dataset de AirBNB.

A grandes rasgos, tomaremos los datos de un dataset descargado de la web de AirBNB como origen, realizaremos sobre ellos un tratamiento de limpieza de “ruido” eliminando columnas que no utilizaremos la elaboración de nuestra presentación y descartando las filas con datos que consideramos atípicos o “outlayers”.

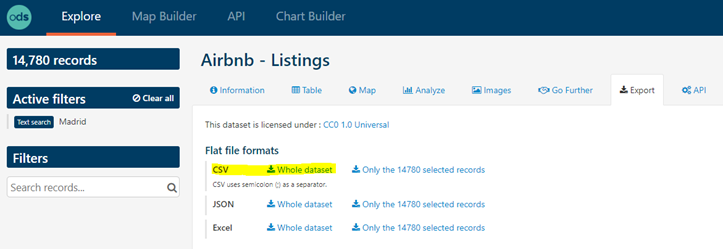
He tomado la licencia de, a la hora de confeccionar esta práctica final, con el deseo y la intención de que su presentación fuera lo más adecuada posible, utilizar ***Microsoft*** ***Word*** como herramienta para documentar y detallar mi labor al profesor, además de elegir una orientación horizontal con el objeto de que las gráficas se aprecien de la mejor forma posible.

**4. Origen de los Datos y Tratamiento Inicial**

Utilizaremos un fichero procedente de AIRBNB, con datos de los inmuebles disponibles, que descargaremos manualmente de este enlace:

**https://public.opendatasoft.com/explore/dataset/airbnb-listings/export/?disjunctive.host\_verifications&disjunctive.amenities&disjunctive.features&q=Madrid&dataChart=%3D&location=16,41.38377,2.15774&basemap=jawg.streets**

La descarga a elegir será la del enlace resaltado en el recorte de la página.



Lo que haremos será filtrar y limpiar el archivo descargado que contiene 89 columnas con 494.954 (casi medio millón de) filas.

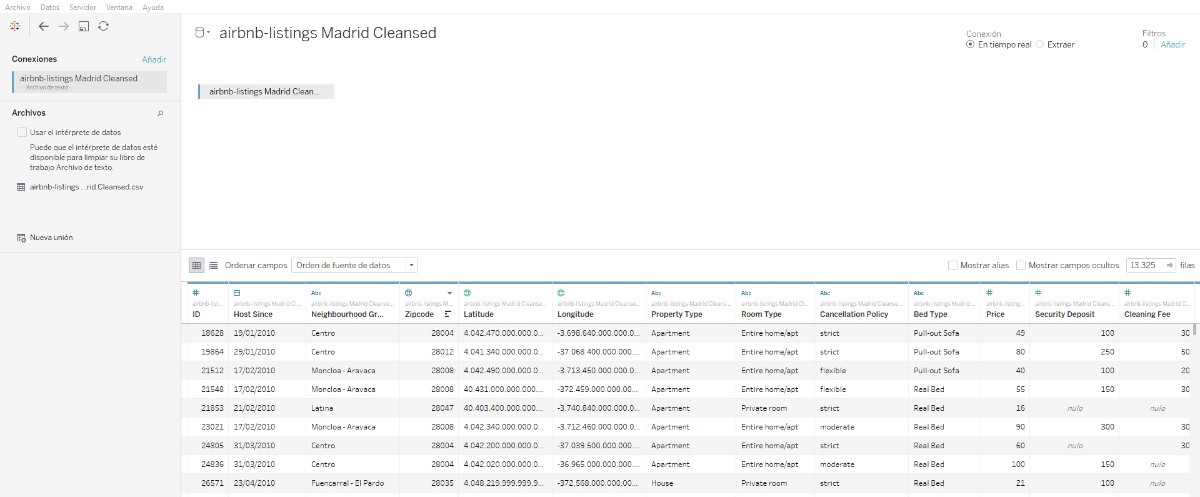
Filtraremos este gran dataset para quedarnos únicamente con los registros de inmuebles ubicados en la Comunidad de Madrid (España), obteniendo un subset de 13.215 filas.

Para nuestro trabajo nos quedaremos sólo con las siguientes 42 columnas. Aun sabiendo que no las utilizaremos todas, nos parecen lo suficientemente significativas:

* **ID** – Identificador único del inmueble en el dataset
* **Last Scraped** – Fecha de última actualización del inmueble
* **Host Since** – Fecha de alta del inmueble
* **Host Response Time** – Tiempo de respuesta del propietario
* **Host Response Rate** – Tasa de respuesta del propietario
* **Host Listings Count** – Veces que ha sido visitado el inmueble
* **Neighbourhood Group Cleansed** – Distrito en el que está ubicado el inmueble
* **City** – Ciudad en la que está ubicado el inmueble
* **State** – Comunidad en la que está ubicado el inmueble
* **Zipcode** – Código postal en el que está ubicado el inmueble
* **Country** – País en el que está ubicado el inmueble
* **Latitude** – Coordenada de latitud geoposicional del inmueble
* **Longitude** – Coordenada de longitud geoposicional del inmueble
* **Property Type** – Tipo de propiedad
* **Room Type** – Tipo de alquiler
* **Cancellation Policy** – Política de Cancelación
* **Bathrooms** – Número de baños
* **Bedrooms** – Número de habitaciones
* **Beds** – Número de camas
* **Bed Type** – Tipo de camas
* **Square Feet** – Superficie del inmueble en pies cuadrados
* **Price** – Precio de alquiler por día completo
* **Weekly Price** – Precio de alquiler por semana completa
* **Monthly Price** – Precio de alquiler por mes completo
* **Security Deposit** – Fianza
* **Cleaning Fee** – Tarifa por limpieza
* **Accommodates** – Capacidad en número de huéspedes
* **Guests Included** – Número de Invitados permitido
* **Extra People** – Precio por persona extra
* **Minimum Nights** – Número mínimo de noches para alquilar
* **Maximum Nights** – Número máximo de noches para alquilar
* **Number of Reviews** – Número de reseñas
* **First Review** – Fecha de la primera reseña
* **Last Review** – Fecha de la última reseña
* **Review Scores Rating** – Porcentaje total de puntuación de las reseñas
* **Reviews per Month** – Porcentaje de reseñas por mes
* **Review Scores Accuracy** – Puntuación de las reseñas por veracidad
* **Review Scores Cleanliness** – Puntuación de las reseñas por limpieza
* **Review Scores Checkin** – Puntuación de las reseñas por el recibimiento
* **Review Scores Communication** – Puntuación de las reseñas por nivel de comunicación
* **Review Scores Location** – Puntuación de las reseñas por la localización
* **Review Scores Value** – Puntuación total de las reseñas

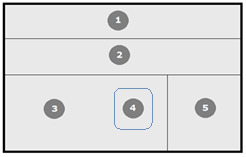
# **Carga del Dataset en Tableau**

Para cargar el dataset en la herramienta Tableau pulsamos sobre la opción “Añadir” y elegimos la opción de “Archivo de texto”. Elegiremos nuestro archivo “limpio de ruido” [***airbnb-listings Madrid Cleansed.csv***] y aparecerá en la pestaña de “Fuente de datos” de la herramienta.

Optaremos por ocultar varias columnas (degradadas en color gris, en la relación del párrafo anterior), para optimizar el rendimiento y renderizado de las consultas que realizaremos en la elaboración del Dashboard.

# **Diseño del Dashboard**

Planteamos para nuestro Dashboard un diseño de cinco marcos, repartiendo el área de la hoja horizontalmente en tres secciones y dividiendo la sección horizontal inferior en 2 espacios más uno flotante.



Como no tenemos predeterminada la utilidad del Dashboard, parametrizamos la elección de su tamaño en “Automático”.



# **Marco 1 · Evolución Anual del Promedio del Precio Total Día por Distrito**

Pretendemos visualizar en este primer marco una evolución del promedio del precio total por día en los inmuebles, clasificados por distrito, durante los cinco últimos años registrados en el dataset (de 2013 a 2017).

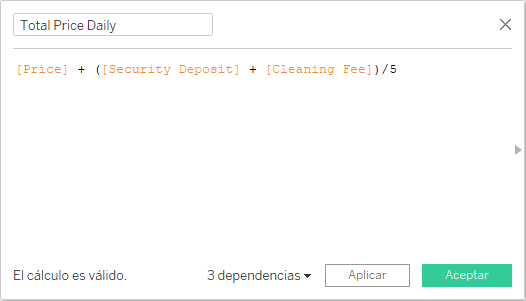
## **Campos Calculados**

Lo primero que haremos será crear dos campos calculados que vamos a necesitar para nuestras representaciones:

* **Rango\_13-17**
* **Total Price Daily**

El primero de ellos será una acotación temporal de años, entre 2013 y 2017, que será la que deseamos representar en nuestro análisis.

El segundo será una suma de importes que consolidará el precio real total diario de alquiler del inmueble.

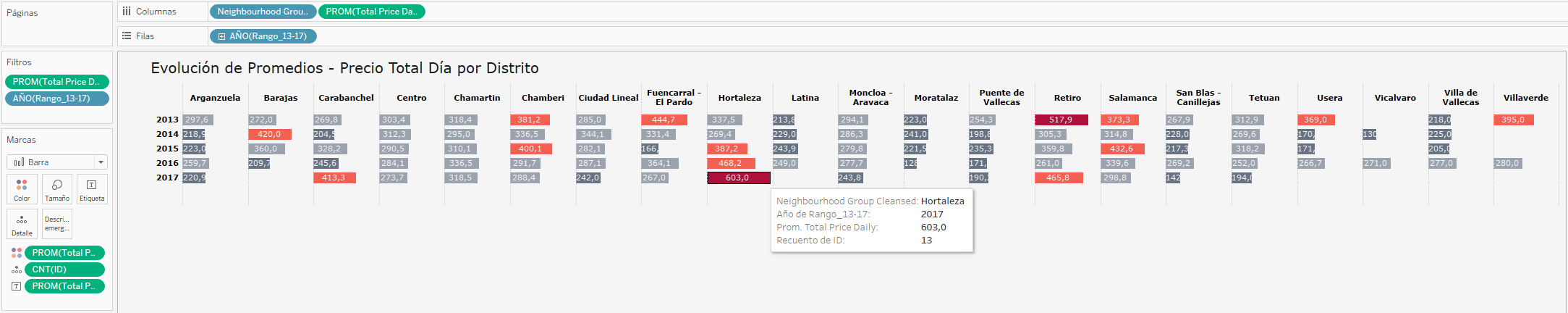
Entendemos que tanto la fianza como la tarifa de limpieza sólo se pagan una vez por alquiler, pero también que son cuotas obligatorias al formalizar el arrendamiento, no incluidas en el precio por día, por tanto establecemos un quinto de su suma como importe a añadir a ese precio, decidiendo que el periodo de alquiler medio de los inmuebles es de 5 días.

## **Elaboración de la Gráfica**

Pintaremos por columnas los distritos [**Neighboorhoud Group Cleansed**] y su promedio de precio total diario [**PROM(Total Price Daily)**], mientras en las filas fijaremos el rango de los años [**AÑO(Rango\_13-17)**], el cual al ser una dimensión temporal, será desplegable.

Será una gráfica de barras horizontales sin líneas de ejes, salvo unos tenues separadores para los distritos, en la que las barras representativas del promedio de precios contendrán su valor e irán coloreadas en función del mismo.

Ambos campos calculados se incluirán en el filtro para eliminar valores nulos. Y en la descripción emergente de cada elemento aparecerá, además de lo representado en los ejes, el recuento de inmuebles en ese distrito/año [**CNT(ID)**].



# **Marco 2 · Evolución Anual del Promedio del Precio Total Día por Tipo de Inmueble y Alquiler**

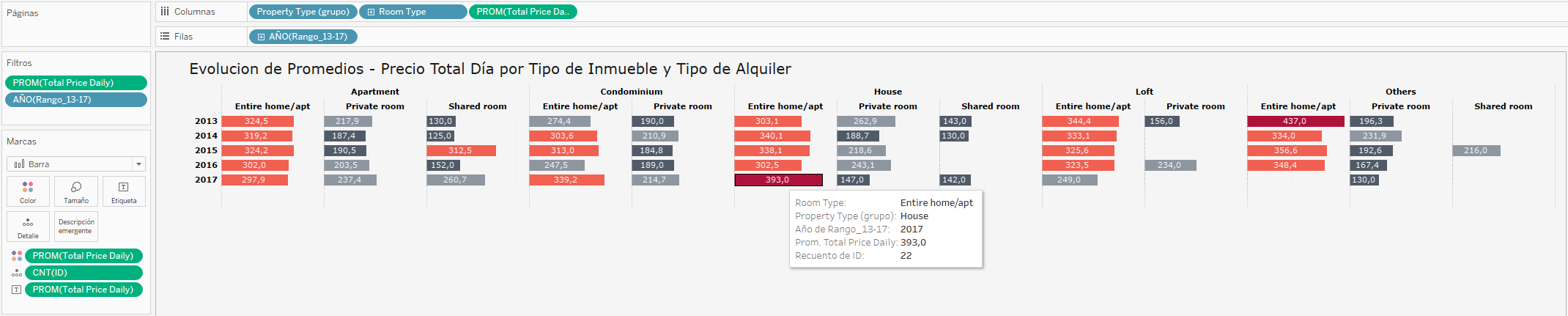
En el segundo marco visualizaremos la evolución del promedio del precio total por día en los inmuebles, clasificados por tipo de propiedad y tipo de alquiler, durante los cinco últimos años registrados en el dataset (de 2013 a 2017), aprovechando los campos que ya tenemos calculados.

## **Elaboración de la Gráfica**

Pintaremos por columnas los tipos de inmueble [**Property Type (grupo)**] desglosados por tipo de alquiler [**Room Type**], que se podrá desplegar al tipo de cama [**Bed Type**], y su promedio de precio total diario [**PROM(Total Price Daily)**], en las filas fijaremos de nuevo el rango de los años [**AÑO(Rango\_13-17)**], que será desplegable.

Será también una gráfica de barras horizontales sin líneas de ejes, salvo unos tenues separadores para los tipos de inmueble y de alquiler, en la que las barras representativas del promedio de precios, análogamente a la gráfica anterior, contendrán su valor e Irán coloreadas en función del mismo.

Para que la gráfica sea suficientemente representativa y no contenga espacios sin valores, elegimos los tipos de propiedad de “Apartment”, “Condominium”, “House” y “Loft”, agrupando el resto bajo el epígrafe “Others” (de ahí que el campo **Property Type** aparezca en modo grupo y no sea desglosable). Nótese también que se ha definido la jerarquía de **Property Type** / **Room Type** / **Bed Type** para hacer posibles desgloses.

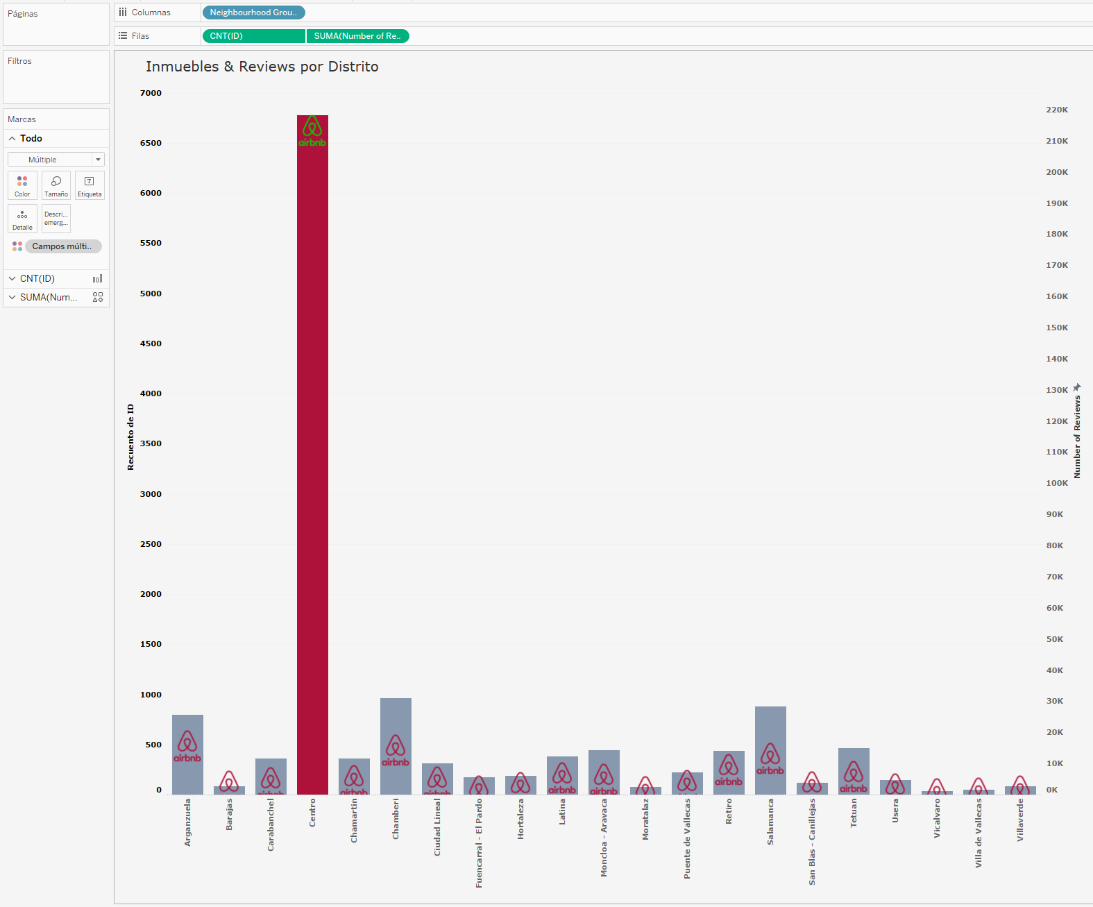
Ambos campos calculados se incluirán en el filtro para eliminar valores nulos. Y en la descripción emergente de cada elemento aparecerá, además de lo representado en los ejes, el recuento de inmuebles en ese distrito/año [**CNT(ID)**].

# **Marco 3 · Volúmenes de Inmuebles y Reviews por Distrito**

Nuestro tercer marco presentará un doble eje en el que mediremos los volúmenes de inmuebles por distrito y su número de reseñas registradas.

## **Elaboración de la Gráfica**

Pintaremos por columnas los distritos [**Neighboorhoud Group Cleansed**] y con un doble eje mediremos, a la izquierda el volumen de inmuebles [**CNT(ID)**] y a la derecha el número total de reseñas [**SUMA(Number of Reviews)**].

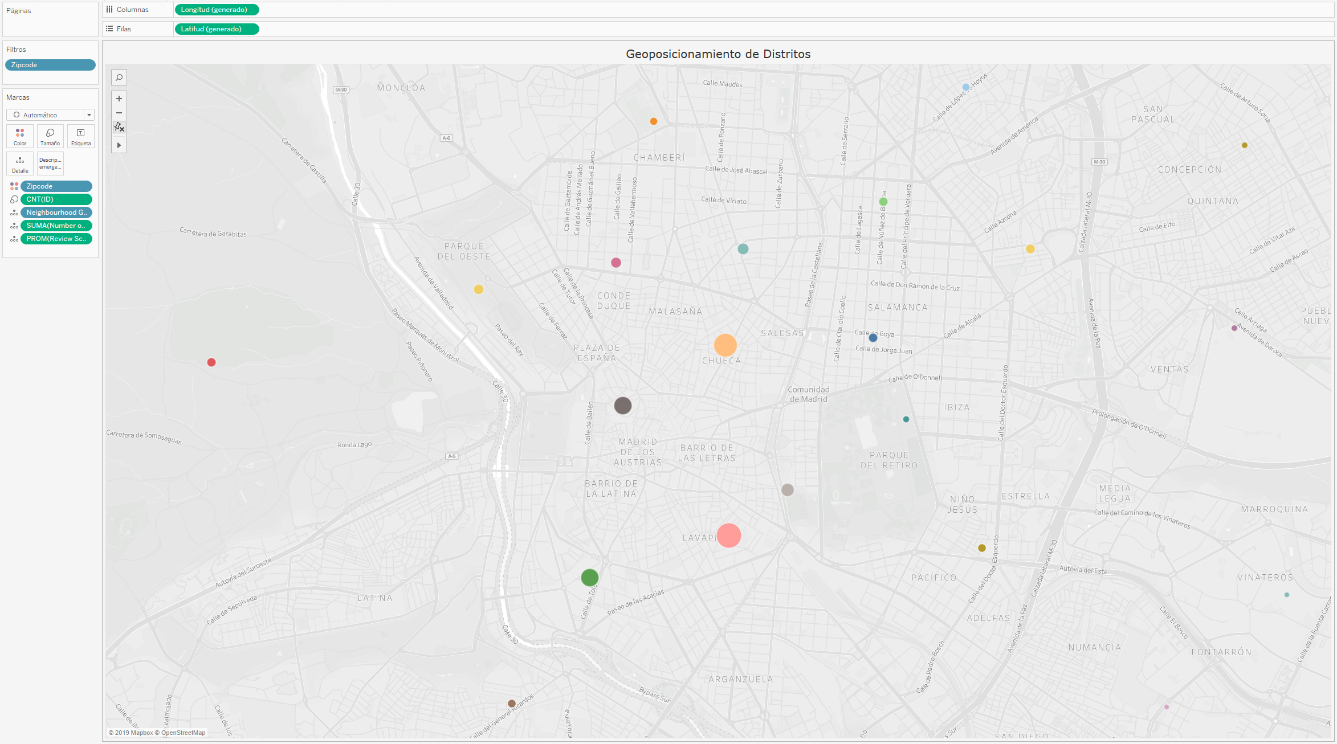
Montamos una gráfica de barras verticales para el volumen de inmuebles por distrito que tomaran color dependiendo de su valor y una ponderación de formas (con el logo de AirBNB) para el numero de reseñas que también tomaran color dependiendo de su altura (valor). Se mitigan las líneas de ejes, salvo una tenue línea de cero, y se etiquetan los distritos en la parte inferior en vertical.

# **Marco 4 · Geoposicionamiento de Distritos**

El cuarto marco viene a romper un poco el diseño del dashboard, pero sin agredir sus líneas maestras. Es una extensión del marco anterior, pero situando los distritos geolocalizados en un mapa, de hecho lo diseñamos en forma flotante para poder superponerlo sobre él.

## **Elaboración de la Gráfica**

Sobre un mapa, posicionado con zoom sobre la comunidad de Madrid, se detallan calles y distritos y se pinta una forma circular coloreada por código postal cuyo tamaño crece en función del volumen de inmuebles que contienen. Filtramos los códigos postales que nos parecen “outlayers.

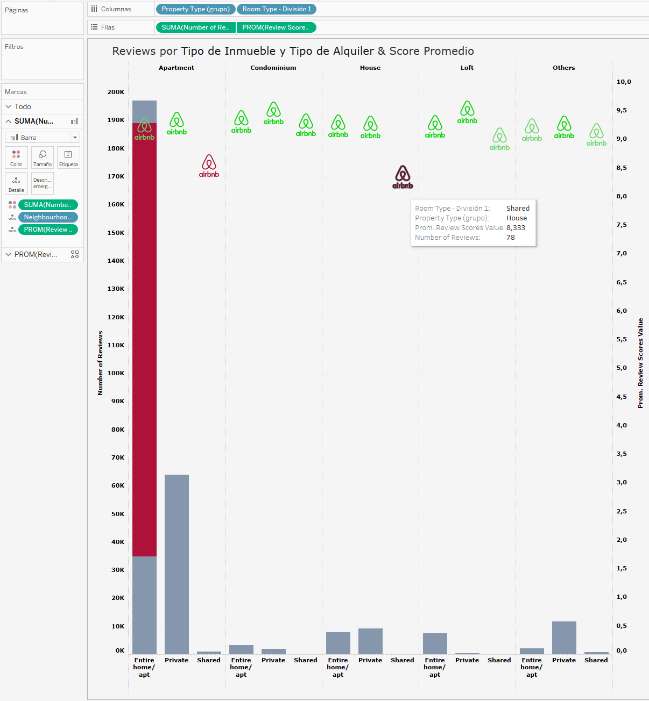
En la descripción emergente de cada elemento aparecerá su distrito [**Neighboorhoud Group Cleansed**], código postal [**Zipcode**] su volumen de inmuebles [**CNT(ID)**] su volumen de reseñas [**SUMA(Number of Reviews)**] y el promedio de sus valoraciones [**PROM(Review Scores Value)**].

# **Marco 5 · Volumen de Reviews por Tipo de Inmueble y Alquiler & Score Promedio**

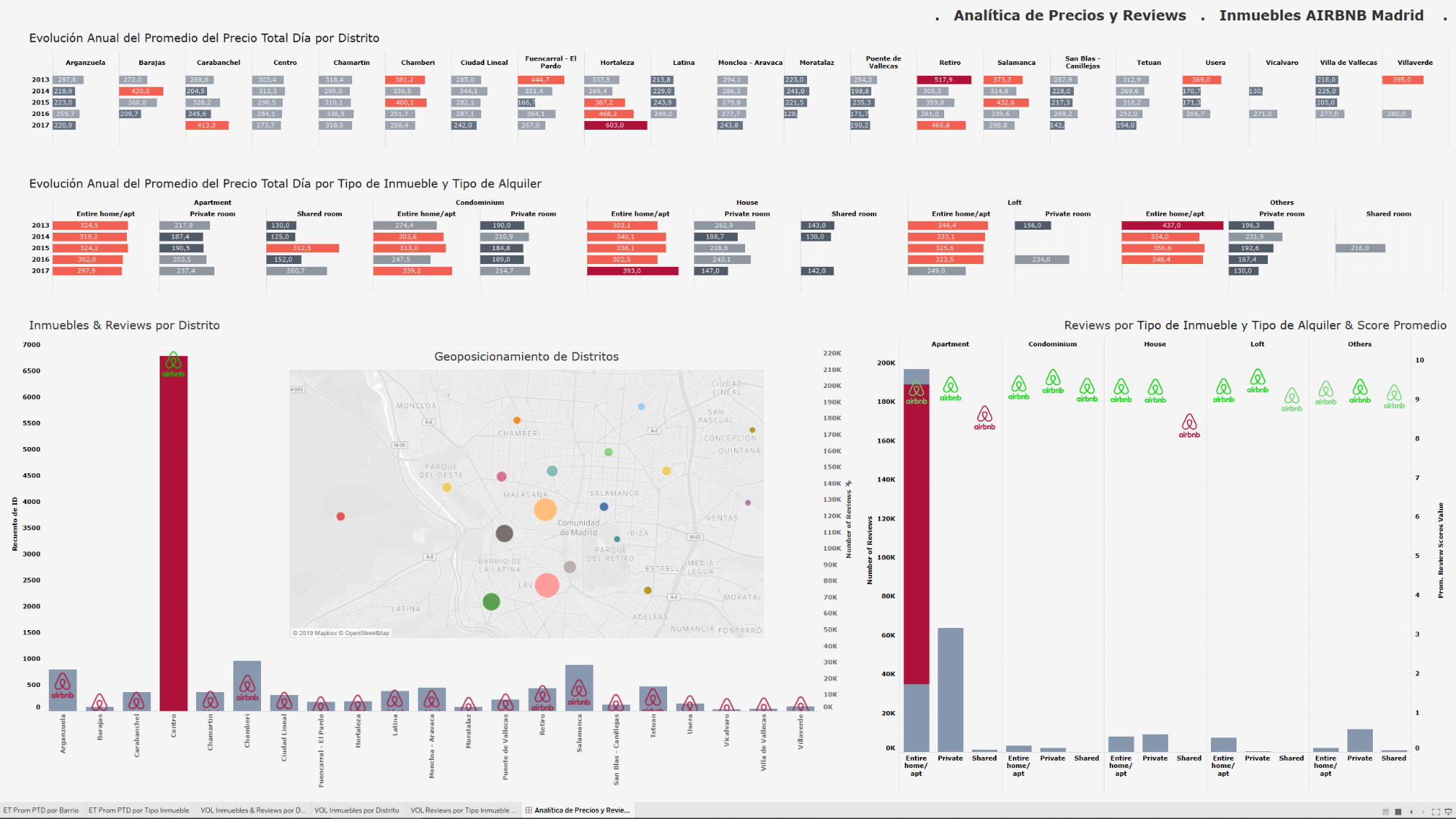
El último marco, el quinto, representará un doble eje en el que mediremos por tipo de inmueble y tipo de alquiler su número de reseñas registradas y el promedio de su puntuación general en dichas reseñas.

## **Elaboración de la Gráfica**

Pintaremos por columnas los tipos de inmueble [**Property Type (grupo)**] eligiendo como representativos los mismos que los del tercer marco (y agrupando el resto bajo “Others”) y los tipos de alquiler [**Room Type**] como su desglose. Con un doble eje mediremos, a la izquierda el volumen de reseñas [**SUMA(Number of Reviews)**] y a la derecha el promedio de la puntuación general en las reseñas [**PROM(Review Scores Value)**].

****Creamos una gráfica de barras verticales para el volumen de reseñas, subdivididas por distrito [**Neighboorhoud Group Cleansed**] el cual tomará color dependiendo de su volumen de reseñas y una ponderación de formas (con el logo de AirBNB) para el promedio de puntuación general que también tomaran color dependiendo de su altura (valor). Se mitigan las líneas de ejes, salvo una tenue línea de cero e igual de tenues líneas separadoras para los tipos de inmueble.

# **Composición Final del Dashboard**

Una vez colocadas hojas en sus respectivos marcos, el dashboard aparecerá de la siguiente forma.

Los dos primeros marcos y el marco flotante (mapa) actúan como filtros sobre el resto de marcos incluidos ellos.

Hemos elegido un fondo claro, pero no blanco puro para que la vista no sufra en demasía y se adapte mejor a la lectura.

A continuación aportamos el archivo twbx con el trabajo realizado.



# **Agradecimientos y Comentarios Finales**

Como nota final de este trabajo me gustaría dar las gracias al profesor, **José Antonio Gascón**, por su trabajo y su atención a la hora de preparar la asignatura de forma tan clara y amena.

Tableau es una herramienta gráfica muy agradable y divertida de usar. Tras aprender su uso estoy convencido de que es la manera de encontrar la cara amable de los datos.

Comentar únicamente que el dashboard que he diseñado no ha tenido como objetivo el virtuosismo o la estética sobredimensionada. En este trabajo he querido plasmar los conceptos que he interiorizado realizando los ejercicios durante la clase junto al profesor He intentado que el diseño no fuera excesivamente recargado, que los colores no resultarán agresivos, dar una visión analítica con 4 representaciones y utilizar para ello gráficas combinadas, ejes dobles, desgloses por jerarquía, elementos resaltados por valor, campos calculados con fórmulas, agrupaciones y filtros.

Sin más, mi agradecimiento por las cosas que he aprendido, que espero se vean reflejadas en esta práctica.

**F. Javier Gonzálvez**