

Proyecto Final

KeepCoding

Bootcamp

- Mujeres en Tech -

Glovo - Big Data

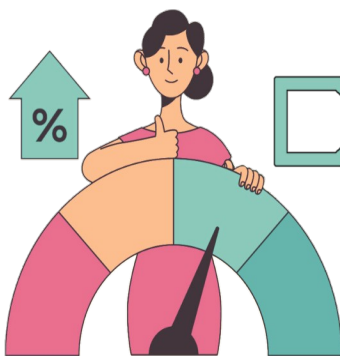
María Lourdes Ramos

Laura Medina

Olivia Moreno

Ana Alonso

Carolina Martínez



DataDonas

ÍNDICE

1. Introducción
2. Tareas del proyecto
 - 2.1. Primeras preguntas
 - 2.2. Definir Data Set
 - 2.3. Validación de los datos
 - 2.3.1. Muestreo y exploración inicial de los datos
 - 2.4. Análisis Exploratorio
 - 2.4.1. Revisión de la calidad de los datos
 - 2.4.2. Detección de outliers (boxplots)
 - 2.4.3. Gráficos (histogramas, etc.)
 - 2.4.4. Normalización de los valores
 - 2.5. Arquitectura de los datos
 - 2.5.1. Definir e implementar el Datawarehouse
 - 2.6. Visualización de las métricas
 - 2.7. Pre-procesamiento y Modelado
 - 2.8. Presentación
 - 2.9. Conclusiones

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

1. Introducción

El presente documento trata sobre la realización del proyecto final para el Bootcamp Mujeres en Tech (Glovo) especialización Big Data de la escuela Keep Coding, basado en el análisis de un dataset (ambientado en 2017) con datos de la empresa de Airbnb proporcionado por la propia escuela.

Este proyecto final es la demostración de las habilidades así como, conocimientos adquiridos a lo largo del Bootcamp en diferentes módulos (Introducción a Estadística, Data Mining & Business Intelligence, Visualización de datos, R, Python, SQL...).

El objetivo principal del proyecto final es la creación de un dashboard (Tableau) a partir del análisis del dataset. El algoritmo utilizado será el de regresión lineal que prediga el precio de un inmueble en función de las características que elijamos. Asimismo del análisis resultante, se creará un informe en el que se simulará la presentación de resultados en un entorno real de empresa. Incluyendo los apartados de suposiciones iniciales, las métricas seleccionadas, en cuanto lo aprendido y sus conclusiones.

Al ser un proyecto realizado en grupo, la definición de tareas y objetivos ha sido crucial para el desarrollo correcto del proyecto utilizando la metodología Agile. Trello es una de aplicaciones basadas en esta metodología (ágil-Kanban), dónde se crean tableros (tipo una pizarra digital) y listas (que abarcan la información correspondiente a los desarrollos del proyecto).

Las tareas predefinidas para poder realizar el análisis han sido: Definir el dataset, Arquitectura y validación de los datos, Análisis Exploratorio de los datos validados, Visualización de las métricas, Pre-procesamiento y Modelado, Informe final, así como las conclusiones.

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

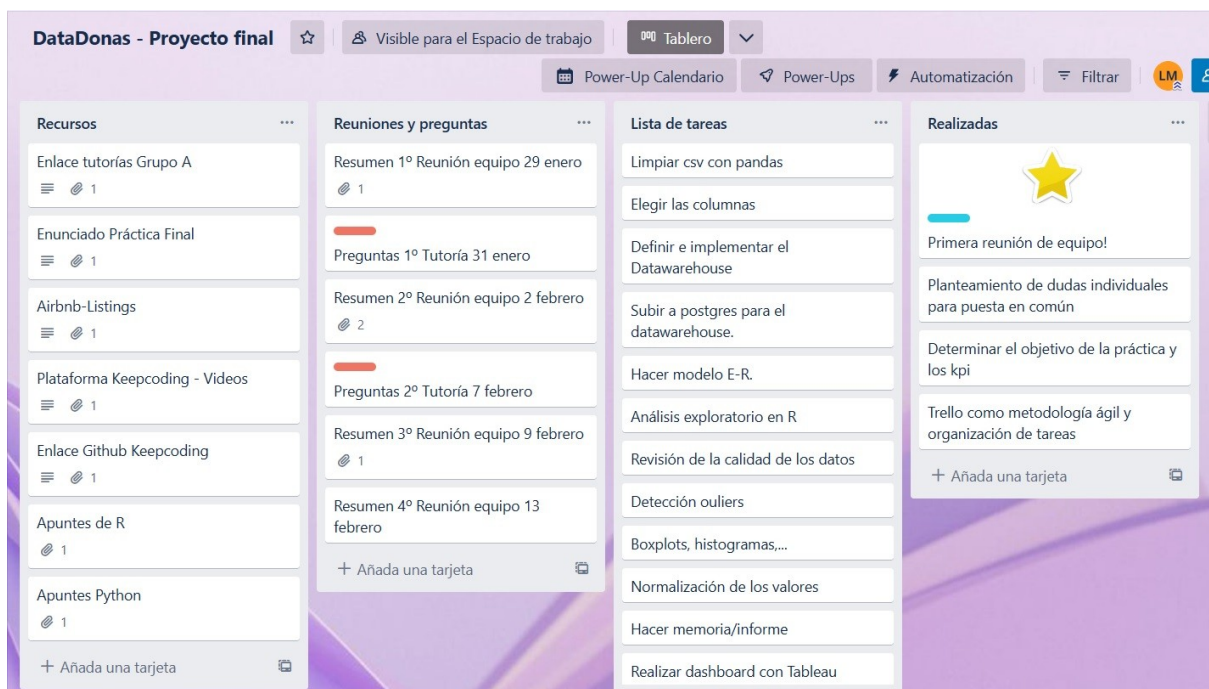
2. Tareas del proyecto

2.1. Primeras preguntas

El grupo se reúne para enfocar el proyecto con una perspectiva profesional ya que se parte de un caso ficticio pero con datos proporcionados de una empresa. Se hacen las primeras preguntas:

- ¿Qué datos se disponen?
- ¿Qué objetivos se marcan?
- ¿Cuáles son las tecnologías que se van a utilizar?

Primeramente, se decide usar Trello como metodología ágil para organizar las tareas y reuniones. Por lo que se crean diferentes apartados para la organización del proyecto: un apartado con los recursos que se necesitan (fuentes externas y apuntes de clase), un apartado con las reuniones y conclusiones de las mismas, otro con las tareas a realizar y un último con las completadas.

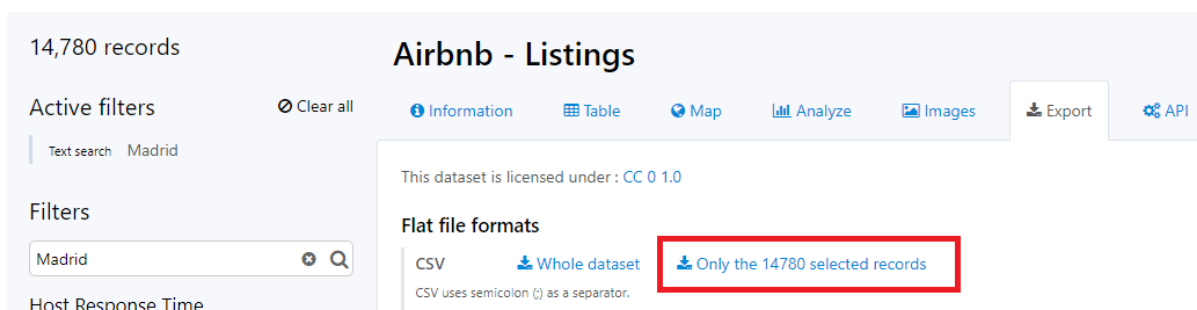


Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

Para el análisis y la limpieza de datos se plantea la utilización de los lenguajes R y Python. Así como, la creación de dashboard en la aplicación Tableau.

2.2. Definir el dataset

Los datos utilizados en este proyecto son a partir de este [dataset](#), scrapeado de Airbnb. Se descarga el fichero airbnb-listings.csv, se elige hacerlo en formato csv ya que es lo que se ha utilizado anteriormente. En las primeras reuniones, se familiariza con el dataset y que tipos de datos y columnas se encuentran.



Se abre una muestra de los campos existentes en un spreadsheets (Excel), para una primera inspección, se seleccionan cuáles serían adecuados utilizar según el objetivo que se empieza a plantear.

ID	Scraped ID	Last Scraped	Experiences Offered	Neighborhood Overview	Transit	Access	Interaction	House Rules	Host ID	Host Location	Host Res
14712301	2017030406572	2017-03-04	none	Quiet residential street	The house is loc	Guests will have	We will be away but will make our	19959228	GB		
12884849	2017030406572	2017-03-05	none	Central Soho, Oxford Street, Berwick Street	Whole Apartment Private Space	- Only 2 People		51343611	London, England, United K		

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

OBJETIVO:

Se parte del supuesto que el grupo forma parte de una consultoría con un cliente que quiere invertir en el mercado inmobiliario de Madrid. Nuestro objetivo es brindar herramientas y guía que faciliten y contribuyan en la toma de decisiones de la selección del inmueble a adquirir, a partir de nuestro análisis de datos, creación del dashboard, del datawarehouse y del modelo predictivo.

Por lo tanto, se decide que el precio/m2 responderá cual es el mejor barrio, las características del piso a comprar y por último, una predicción de cuál sería el precio de alquiler.

Preguntas que ponderamos hacer a los datos a partir de nuestro objetivo.

Preguntas
¿Qué amenities generan mejores valoraciones (media de review score)?
¿Qué barrios están mejor valorados (top 3)?
¿Qué cantidad de habitaciones están mejor valoradas?
¿Qué cantidad de baños están mejor valorados?
¿Cuál es la relación entre el número de habitaciones y el precio (sólo teniendo en cuenta propiedades con más de 4 reviews)?
¿El tipo de cama influye en la valoración de la vivienda?

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

¿En qué barrio hay más reviews per month?

¿Cómo influye la política de cancelación en la valoración del piso?

Features: ¿Cuántas hay? ¿Influyen en el número de reviews per month?

**Las resaltadas en rosa, son las que finalmente se respondieron y hacia dónde se apuntó el análisis junto con la variable de tipo de propiedad.*

Fuente de datos externa:

Se decide enriquecer la base de datos con precios por m2 de viviendas en Madrid del año 2017, que se extrae de la siguiente fuente: “PRECIO DE LOS PISOS EN MADRID – GUÍA POR DISTRITOS [2017]”: [Precios de la vivienda](#)

2.3. Validación de los datos

2.3.1. Muestreo y exploración inicial de los datos

Se procede al análisis exploratorio y primera limpieza del dataset desde Python con la librería Pandas, se realiza en Jupyter Notebook.

Limpiamos las columnas que no utilizaremos

```
#CAMPOS QUE NO APORTAN, LOS ELIMINAMOS
col_drop = ['Listing Url', 'Scrape ID', 'Last Scraped', 'Name', 'Summary', 'Space', 'Description', 'Experiences Offered', 'Neighborhood Overview',
'Notes', 'Transit', 'Access', 'Interaction', 'House Rules', 'Thumbnail Url', 'Medium Url', 'Picture Url', 'XL Picture Url', 'Host ID', 'Host URL',
'Host Name', 'Host Location', 'Host About', 'Host Response Time', 'Host Response Rate', 'Host Acceptance Rate', 'Host Thumbnail Url',
'Host Picture Url', 'Host Neighbourhood', 'Host Listings Count', 'Host Total Listings Count', 'Host Verifications', 'Street',
'Neighbourhood', 'Neighbourhood Cleansed', 'State', 'Zipcode', 'Market', 'Smart Location', 'Country Code', 'Country',
'Weekly Price', 'Monthly Price', 'Guests Included', 'Extra People', 'Minimum Nights', 'Maximum Nights', 'Calendar Updated',
'Has Availability', 'Availability 30', 'Availability 60', 'Availability 90', 'Availability 365', 'Calendar last Scraped', 'First Review', 'Last Review',
'Review Scores Accuracy', 'Review Scores Cleanliness', 'Review Scores Checkin', 'Review Scores Communication', 'Review Scores Location',
'Review Scores Value', 'License', 'Jurisdiction Names', 'Calculated host listings count', 'Geolocation']

airbnb.drop(col_drop, axis=1, inplace=True)

# COLUMNAS QUE QUEDAN
list(airbnb.columns)

airbnb.info()

#ANÁLISIS DE MISSING
airbnb.isnull().sum().sort_values(ascending=False)
airbnb.isna().sum().sort_values(ascending=False)

#GUARDAMOS EL DATAFRAME MODIFICADO EN UN CSV
airbnb.to_csv('airbnb_clean.csv', index=False, sep=',')
```

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

Exploración con Python (Pandas)

Primeramente se eliminan los campos que no aportan, por ejemplo columnas dónde los datos no son relevantes para los objetivos marcados ('License', 'Jurisdiction Names'...) Por lo tanto, seguidamente, se lista las columnas que pueden ayudar en el objetivo. Se revisa de esas columnas cuáles contienen datos NULL y en qué proporción. Finalmente, se guarda el dataframe modificado en un csv.

dataframe modificado

2.4. Análisis exploratorio

2.4.1. Revisión de la calidad de los datos

En los valores se encuentran muchos datos NULL por lo que el enriquecimiento del dataset anteriormente mencionado ayuda a que el filtro propuesto (solo Zona Madrid) no quede afligida por pocos datos. Se decide utilizar R en este caso para el análisis del dataframe modificado. De manera que, se procede a la imputación de los NULL con el cambio de NULL/NA por la media de cada columna.

Acto seguido, se cambia los tipos de datos de las columnas para no causar futuros errores:

- Latitud y Longitud float
- Bathrooms, Bedrooms, Beds a int
- price, security deposit, cleaning fee a float
- review.Scores.Rating a int
- Reviews.per.month a float

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

```
ELEJIMOS EL FILE airbnb_clean.csv EN EL DIRECTORIO NUESTRO
'''{r}
file.choose()
'''

Copiamos el resultado con comillas incluidas y lo pongo en ruta_csv
'''{r}
ruta_csv <- "/Users/marialourdesramos/Downloads/airbnb_clean.csv"
airbnb1<-read.csv2(ruta_csv)
head(airbnb1)
'''
'''{r}
typeof(airbnb1)
'''

lo convertimos en dataframe
'''{r}
airbnb <- as.data.frame(airbnb1)
typeof(airbnb)
'''
'''{r}
summary(airbnb)
'''
'''{r}
glimpse(airbnb)
'''
'''{r}
colSums(is.na(airbnb))
```

[airbnb_cleaning_data.qmd](#)

Asimismo, se filtran algunas columnas y así como, se eliminan cuando el precio total ya está filtrado. Luego, se procede a la separación de las tablas “amenities” y “features” en distintas filas en concordancia con el “ID”. Además, se realiza la normalización de valores (quitamos tildes, dobles espacios, etc.). Por último, se guardan los datasets resultantes.

NORMALIZAMOS DATOS, ELIMINANDO DOBLE CARACTERES, TILDES, ETC.

```
'''{r}
airbnb <- as.data.frame (sapply(airbnb1, function(x) iconv(x,from = "UTF-8", to = "ASCII//TRANSLIT")))
airbnb
'''
```

[datasets resultantes](#)

2.4.2. Detección outliers (boxplots)

Tras realizar varios boxplots para la detección de outliers y comprobar que en su mayoría los valores atípicos podrían deberse a errores en la recolección de datos válidos hemos decidido conservarlos. No obstante, mostramos un caso en el que

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

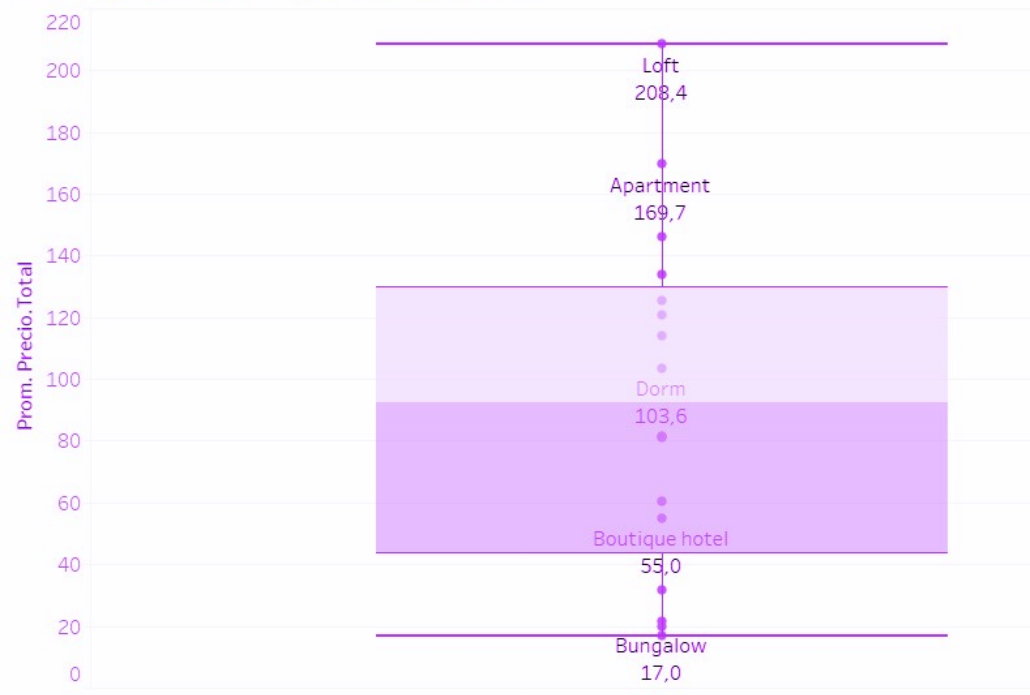
consideramos necesaria su eliminación no solo por el desvío en la variable que se estudia sino por el objetivo de nuestro análisis.



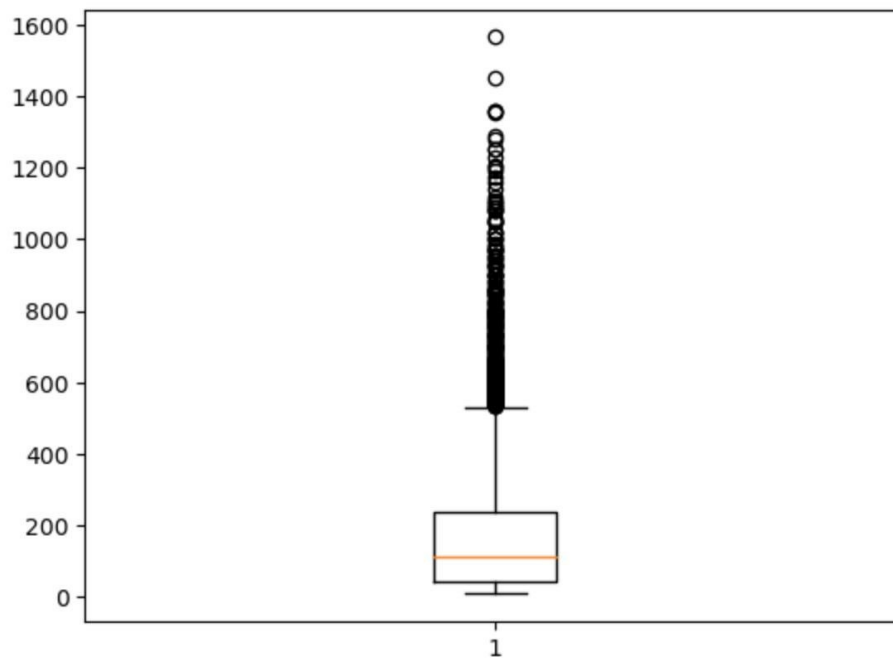
Eliminamos aquellos tipos de viviendas que no tendría sentido incorporar en nuestra valoración por las características que busca nuestro cliente.

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

Precio promedio-Tipo de propiedad

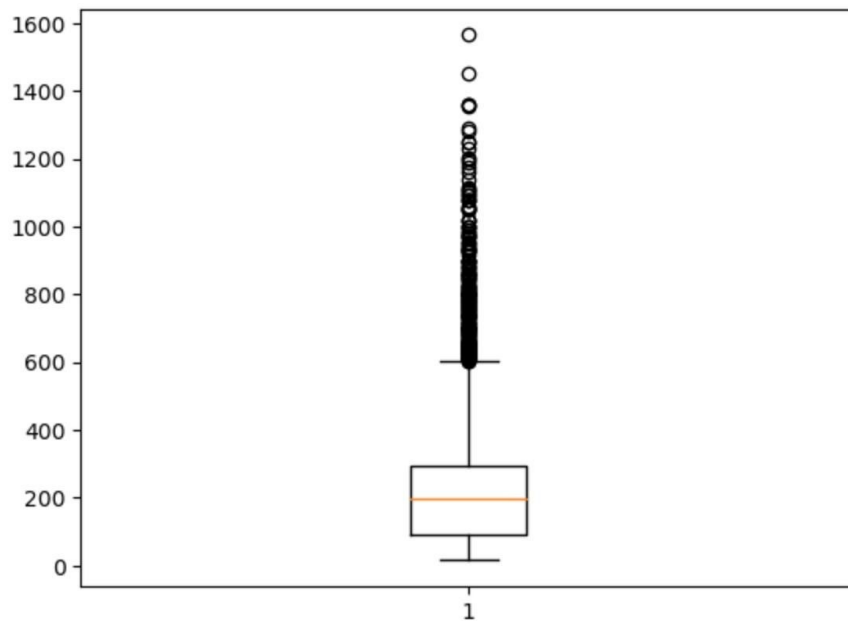


A continuación, se encuentra el promedio por tipo de vivienda.



Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

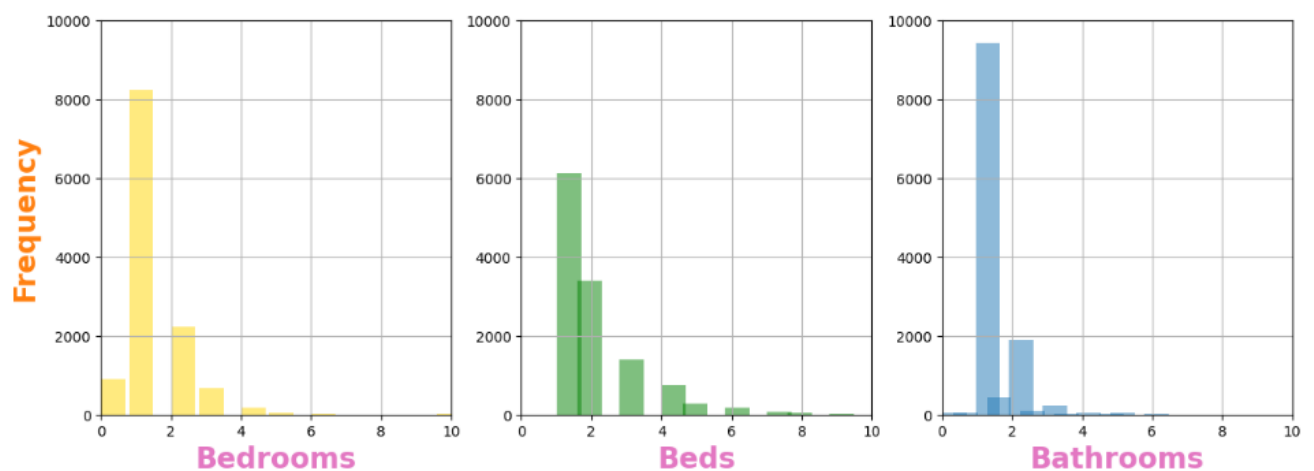
Luego, el promedio del precio final de “Entire home/apt”. Se evidencia que no hay mucha diferencia.



2.4.3. Gráficos (histogramas, etc.)

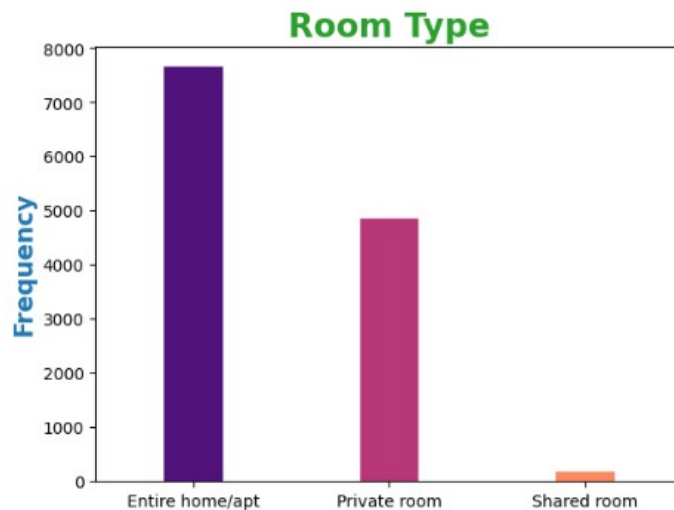
Dentro de los gráficos utilizados para el análisis de datos, se emplean las librerías matplotlib y seaborn para su creación.

Primeramente, histogramas para ver la frecuencia con la que se repite el número de habitaciones, camas y baños en los registros encontrados.



Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

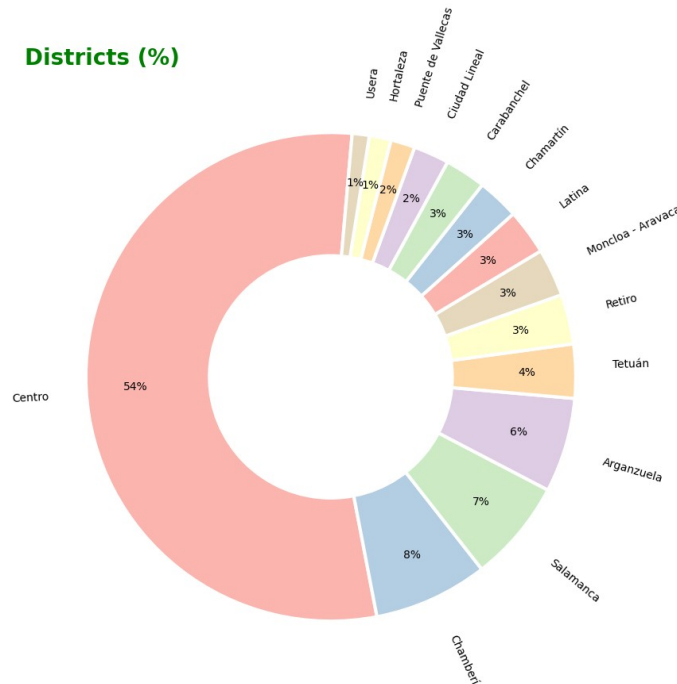
Seguidamente, para observar la frecuencia con la que se repite cada tipo de habitación. Claramente, aquí se muestra que la opción *Entire Home/apt* es la opción con más registros.



Otro gráfico utilizado es pie chart (con efecto de dona), para observar las cantidades de muestra de cada distrito de Madrid. Visiblemente, el distrito Centro es el que obtiene más registros de opciones en el dataset de Airbnb.

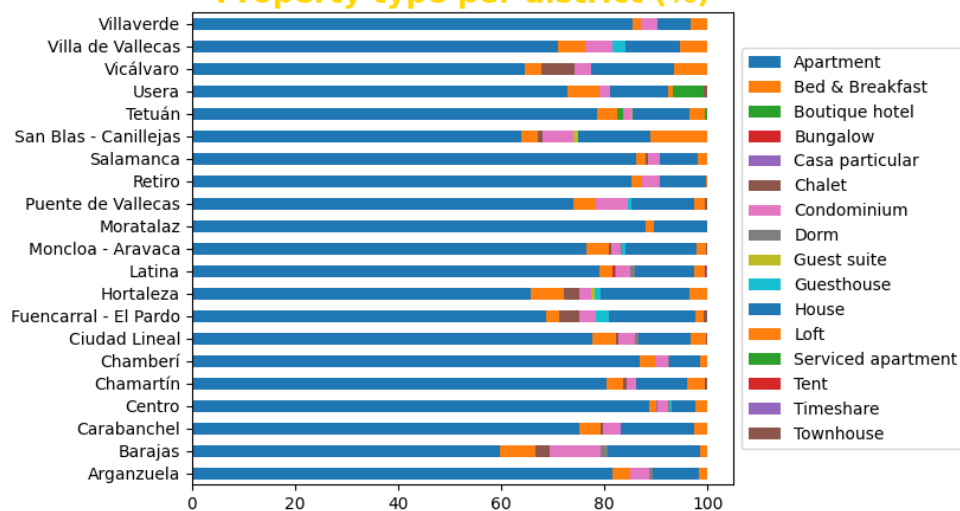
Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

Districts (%)



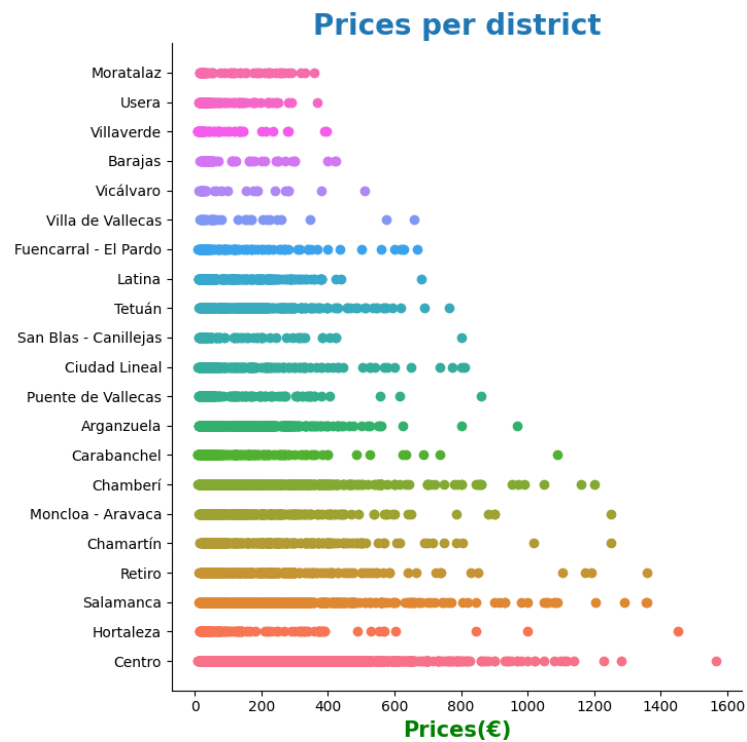
A continuación, se agrupa el data frame por los distritos y se cuenta la frecuencia de cada tipo de propiedad, de tal modo que se presenta en porcentaje. Como resultado, la opción *apartment* es la que obtiene el mayor porcentaje (mínimo 60%) en cada uno de los distritos.

Property type per district (%)

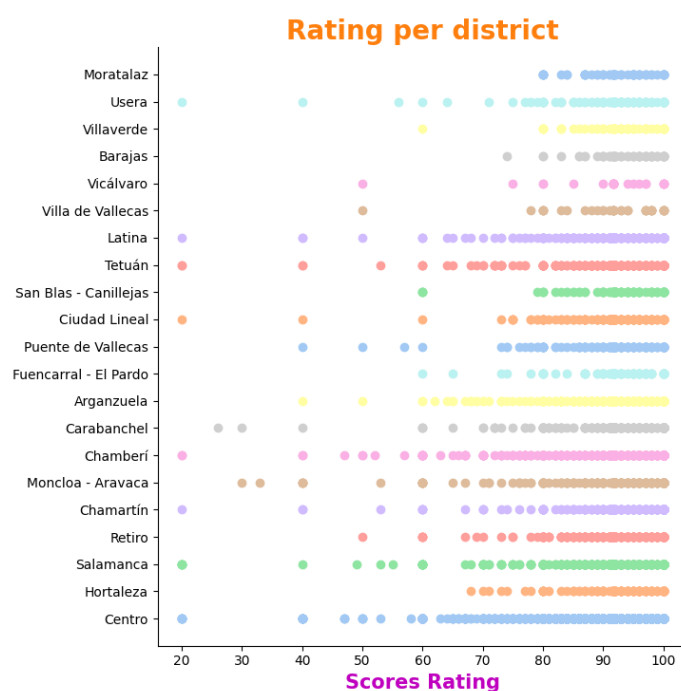


Con respecto a la correlación entre los distritos y el precio total, se utiliza un scatter plot ya que facilita ver los patrones de las dos variables.

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data



Además, se utiliza este tipo de gráfico para la observación de la correlación entre los distritos y su calificaciones obtenidas (*scores ratings*) por los usuarios de la plataforma (huéspedes).



2.4.4. Normalización de los valores

Suma de todos los pasos realizados para el análisis del dataset original, así como de su limpieza para que los datos obtenidos fomenten una respuesta acorde a la propuesta-objetivo inicial del cliente.

Se revisa en los 3 datasets definitivos creados a partir de diferentes análisis con R y Python. Por ejemplo, con Python se normaliza la lista de “Amenities”, que es una de las variables más importantes, ya que con su análisis se responde a una de las cuestiones principales de este estudio.

```
AMENITIES = ["Wireless Internet", "Kitchen", "Heating", "Essentials", "Washer", "TV", "Elevator in building",  
            "Family/kid friendly", "Iron", "Hair dryer", "Air conditioning", "Laptop friendly workspace",  
            "Smoke detector", "Wheelchair accessible", "Lock on bedroom door", "Free parking on premises"]  
  
def ingest_csv(file):  
    with open(file) as f, open('/Users/omoreno/Downloads/mioutput.csv', 'w') as outf:  
        reader = csv.DictReader(f, delimiter=';')  
        writer = csv.DictWriter(outf, fieldnames=OUTF_FIELDS)  
        writer.writeheader()  
        for row in reader:  
            new_row = row  
            for amenity in AMENITIES:  
                print(amenity)  
                if row["Amenities"].find(amenity) != -1:  
                    new_row[amenity] = "TRUE"  
                else:  
                    new_row[amenity] = "FALSE"  
            writer.writerow(new_row)
```

También, se normaliza la columna de “room type” para tener el tipo de habitación bien ajustada. Se procede a quitar el tipo “shared room” ya que no es algo que se acepte dentro de la política de Airbnb y por tanto no será de interés para nuestro cliente.

Además, se comprueba, con un conteo de valores, que el tipo de cama (“bed type”) no es de interés, por su mínima variación a lo largo del dataset.

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

```
#Parece que Tipo de habitación y tipo de cama no nos interesan mucho, por si acaso comprobamos porcentajes:

print(f"Del total de 12685 líneas hay la siguiente distribución por tipo de habitación y de cama")

print(df.groupby('Room.Type')['ID'].count().sort_values(ascending=False))
print("")
print(df.groupby('Bed.Type')['ID'].count().sort_values(ascending=False))
```

0] ✓ 0.1s

Del total de 12685 líneas hay la siguiente distribución por tipo de habitación y de cama

Room.Type

Entire home/apt	7654
Private room	4851
Shared room	180

Name: ID, dtype: int64

Bed.Type

Real Bed	12403
Pull-out Sofa	228
Futon	34
Couch	15
Airbed	5

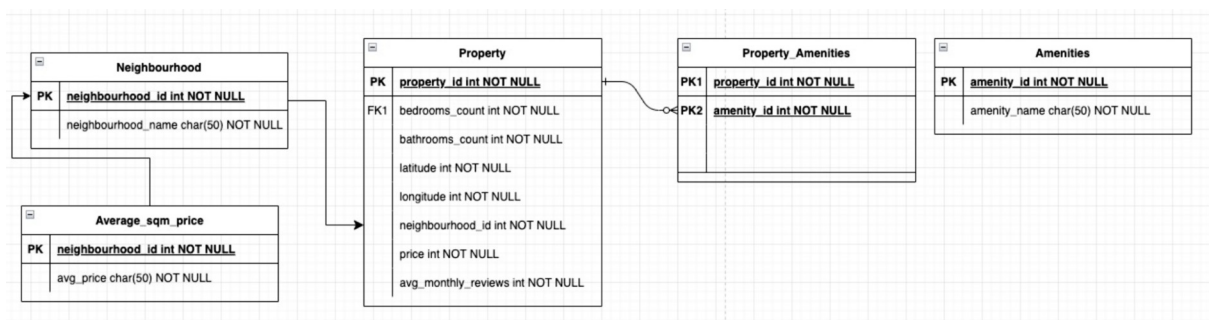
Name: ID, dtype: int64

Después de revisar los datos extraídos de los análisis, se comprueba que el precio promedio del dataset del 2017 no se ha ingestado correctamente, ya que en varias ocasiones, el nombre del barrio en la tabla de precios promedio no estaba completo (por ejemplo: Moncloa en lugar de Moncloa - Aravaca). Como solución se crea un pequeño script como hotfix.

2.5. Arquitectura de los datos

2.5.1. Definir e implementar el Datawarehouse

En la definición e implementación del Datawarehouse, primeramente, se diseña un diagrama Entidad-Relación para mostrar un modelo lógico y sencillo de la base de datos.



Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

Luego se produce el script con las sentencias DDL en lenguaje SQL.

```
--Creamos el esquema

create schema practica_final authorization npmjcswr;

--Creamos la tabla amenities
create table practica_final.amenities(
    amenity_id integer not null, -- PK
    amenity_name varchar(10) not null,,
    constraint amenity_PK primary key (amenity_id)
);

--Creamos la tabla neighbourhoods

create table practica_final.neighbourhoods(
    neighbourhood_id integer not null, -- PK
    neighbourhood_name varchar(10) not null,,
    constraint neighbourhood_PK primary key (neighbourhood_id)
);
```

[practica_final_schema.sql](#)

2.6. Visualización de las métricas

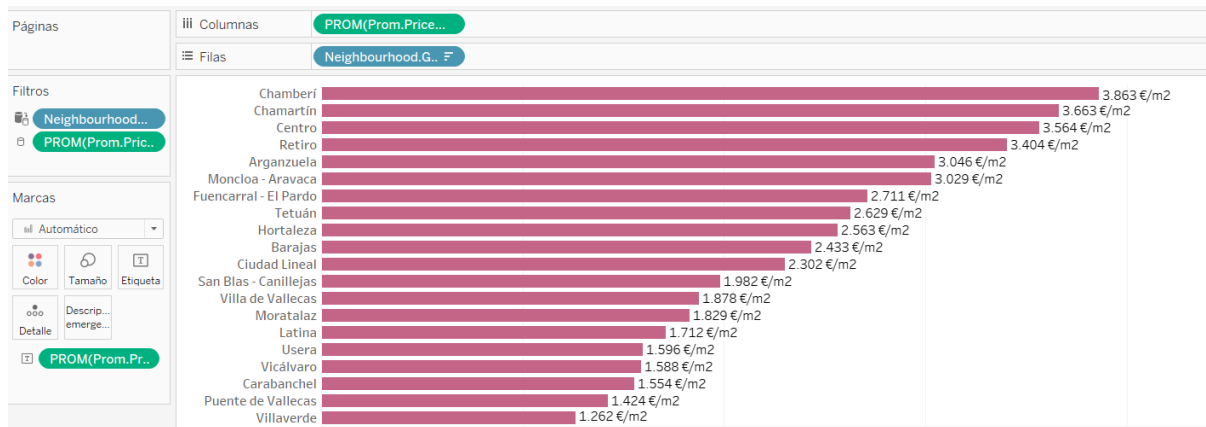
Con motivo de mejorar el desempeño y alcanzar las metas deseadas, se utiliza la aplicación Tableau que permite presentar datos de forma visual para facilitar la comprensión y/o el análisis de dicha información. Le permitirá al cliente utilizar la información visual que puede conducir a una toma de decisiones más eficaz.

Los gráficos se muestran tal y como se han dejado para incorporarse al dashboard.

Consideramos que un buen inicio para mostrar los datos al cliente es partir de la base de cuánto dinero está dispuesto a invertir por m2. En relación al rango que filtre se le mostrarán unas características u otras. Dicho rango lo puede modificar e incluso dejar mínimo y máximo por defecto si prefiere comenzar filtrando por otras características o jugando con los filtros según sus preferencias en el momento de la presentación.

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

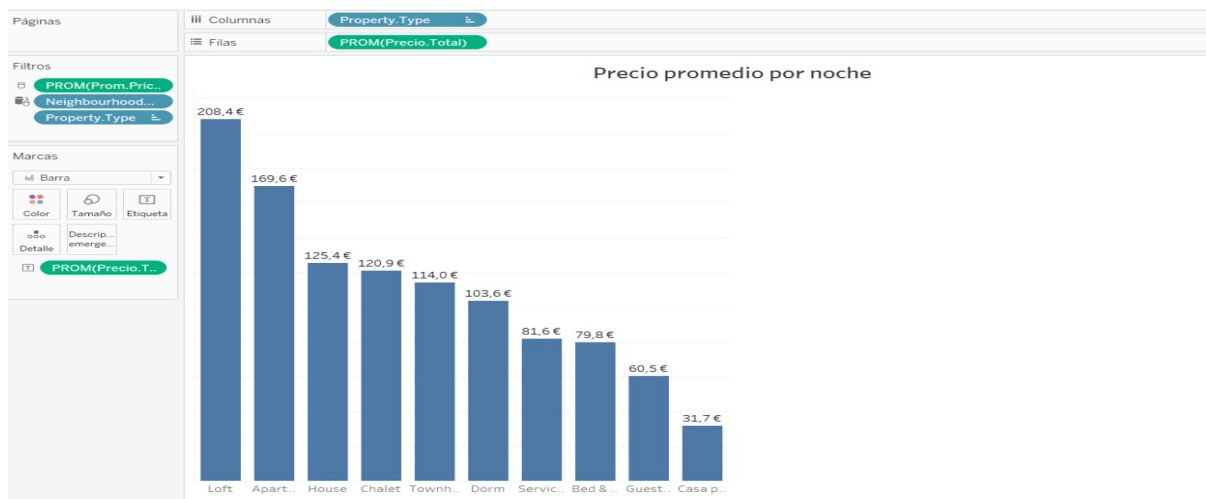
En el primer gráfico mostramos por orden descendente el promedio del precio por m2 en los distritos 21 de Madrid.



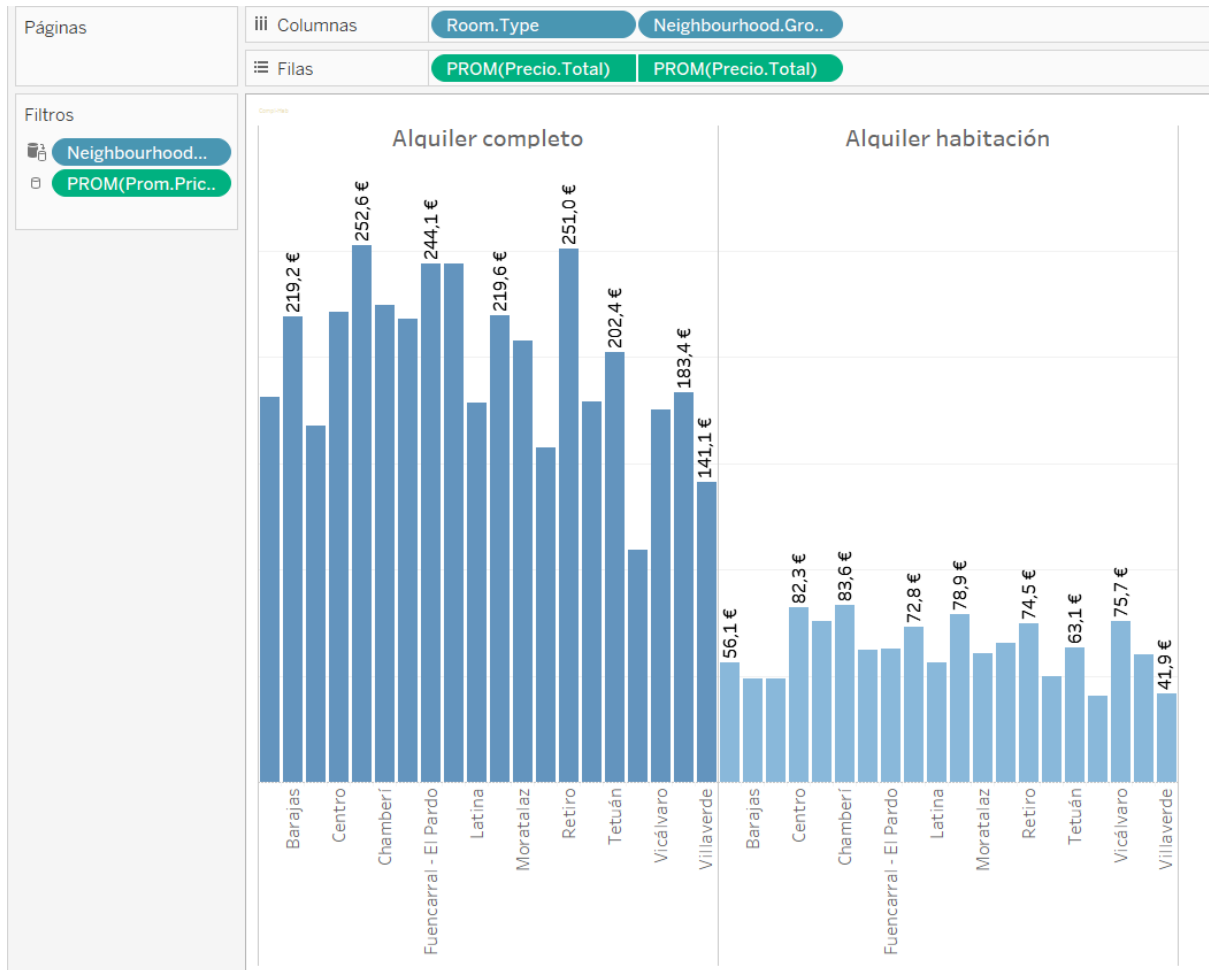
Añadimos un filtro para seleccionar los n distritos con mayor puntuación por parte de los usuarios de Airbnb.

Tanto el filtro de precio por m2 como el de los distritos con mayor puntuación modificarán los valores mostrados en todos los gráficos.

Realizamos un gráfico para mostrar el precio de alquiler promedio por tipo de inmueble y por tipo de alquiler (inmueble completo o solo habitación)

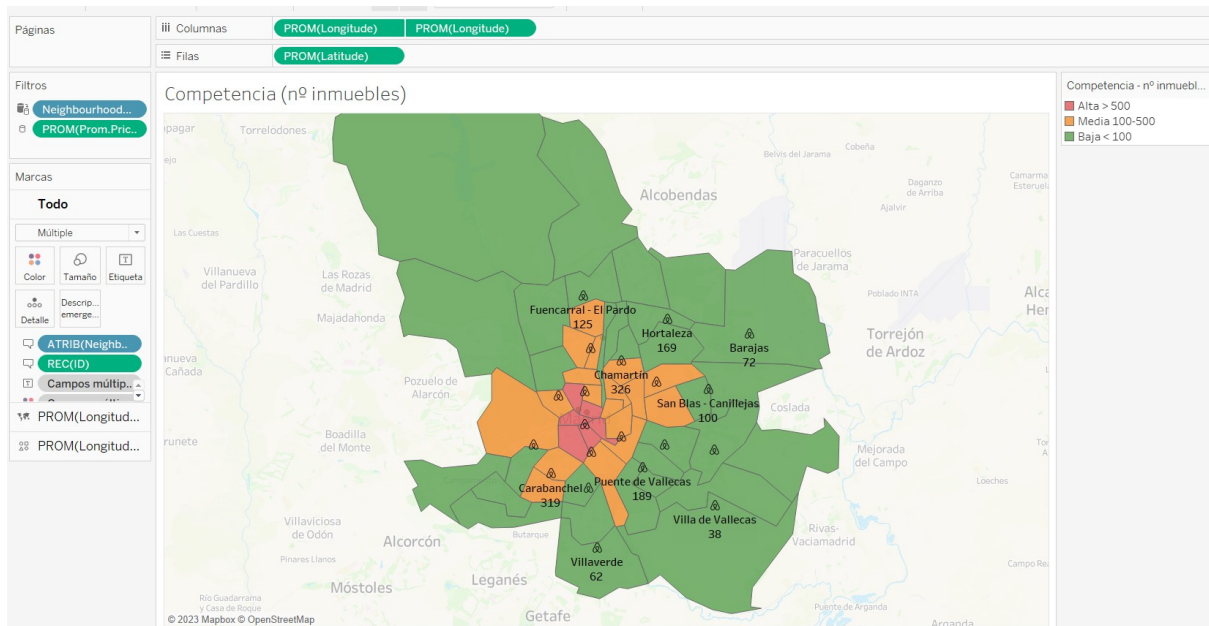


Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

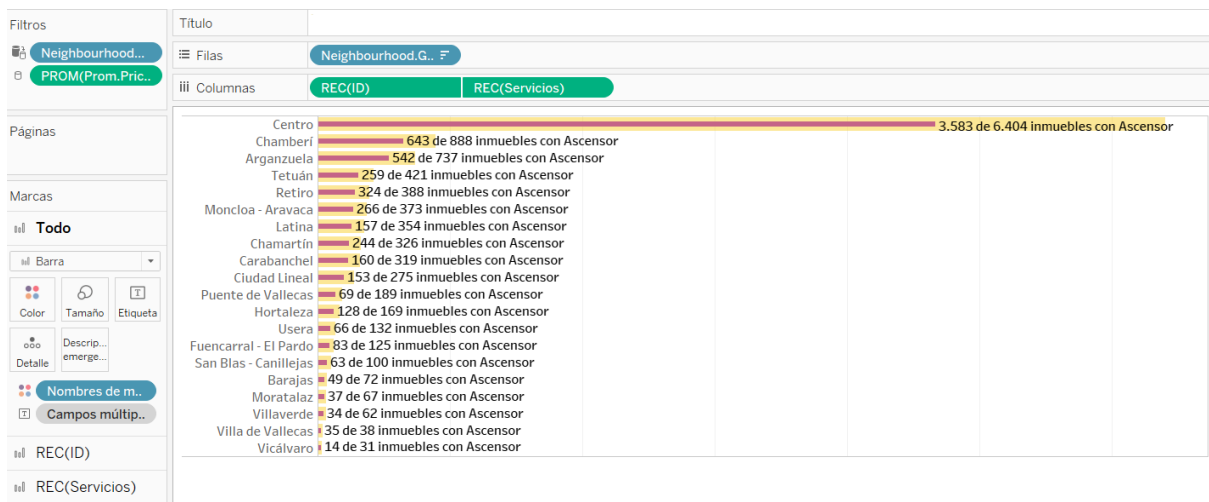


A nivel de impacto visual diseñamos un mapa que muestra por colores cuales son las zonas por código postal con mayor número de inmuebles en alquiler, es decir, la competencia. Creímos mas conveniente la separación por zipcode en lugar de por distritos (información que también se puede visualizar en el mapa) ya que distritos extensos como Fuencarral El Pardo poseen gran diferencia de número de inmuebles entre sus códigos postales, de esta forma los resultados expuestos son más exactos.

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data



Por último, realizamos un filtrado por los servicios más destacados en Airbnb. Dicho filtro mostrará los distritos resultantes de la manipulación del dashboard como consecuencia de las decisiones tomadas por nuestro cliente. Este gráfico es una herramienta que podrá usar el cliente para incorporar aquellos servicios que menos tenga la competencia como elemento diferenciador del resto.



2.7. Pre-procesamiento y Modelado

Planteamos un primer modelo, y evaluamos las variables más representativas en todo el dataset, para ajustar el modelo, de manera tal que sea más sencillo y eficiente.

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

Se utiliza el modelo de regresión lineal ya que es una técnica de análisis de datos que modela matemáticamente la variable desconocida o dependiente y la variable conocida o independiente como una ecuación lineal. En otras palabras, con la información proporcionada por el dataset de Airbnb del año 2017 sobre el precio promedio final de los tipos de propiedad en los barrios de Madrid. La técnica de regresión lineal analiza esos datos y determina cuales son los tipos de propiedad mejor valorados con diferentes variables. Luego calcula cuál será el precio promedio futuro final en el tipo de propiedad.

Residuals:					
	Min	1Q	Median	3Q	Max
	-545.0664637	-78.2223409	-17.7608279	52.1443078	1097.4803992

Residual standard error: 125.113684 on 11500 degrees of freedom	
Multiple R-squared: 0.380211901,	Adjusted R-squared: 0.378002222
F-statistic: 172.066553 on 41 and 11500 DF, p-value: < 2.220446e-16	

Luego, replanteamos el modelo.

Call:	
lm(formula = precio.total ~ Zipcode + Property.Type + Room.Type + Accommodates + Bathrooms + Bedrooms + Number.of.Reviews + Review.Scores.Rating + Reviews.per.Month + precio_m2_2017 + Kitchen + TV + Elevator.in.building + Air.conditioning + Laptop.friendly.workspace + Smoke.detector + Wheelchair.accessible + Free.parking.on.premises, data = airbnb_)	
Residuals:	
	Min 1Q Median 3Q Max
	-535.3688901 -78.2414400 -17.8633571 52.0084385 1095.9581489

--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	
Residual standard error: 125.159915 on 11509 degrees of freedom	
Multiple R-squared: 0.379268372,	Adjusted R-squared: 0.377542469
F-statistic: 219.750749 on 32 and 11509 DF, p-value: < 2.220446e-16	

2.8. Presentación

Como si fuéramos la empresa y el resultado que le mostramos al cliente final.

[Final_project_ppt](#)

2.9. Conclusiones

- **Suposiciones iniciales: ¿Cuáles han demostrado ser válidas y cuáles no? ¿Por qué?**

Las suposiciones iniciales que se han demostrado ser válidas es que el tipo de habitación más valorado y dónde se encuentran más registros es "Entire home/apt".

El precio promedio de los barrios del 2017 no se pudo ingestar correctamente, seguramente los datos no fueron recolectados correctamente. Además, se procedió a la imputación de los NULL con el cambio de NULL/NA por la media de cada columna en algunos casos.

- **Métricas seleccionadas: ¿han sido las correctas o no? ¿por qué?**

Para dar respuesta al objetivo principal del cliente, se seleccionaron diferentes métricas. Buscando con el análisis, la respuesta de diferentes preguntas: se decide que el precio/m2 responderá cual es el mejor barrio, las características del piso a comprar y por último, una predicción de cuál sería el precio de alquiler.

Se decide ocultar algunos parámetros ya que las variables disparaban los datos y no se podía obtener unos resultados concluyentes.

Proyecto Final - Bootcamp Mujeres en Tech (Glovo) - Especialización Data

- **Teniendo en cuenta lo aprendido: ¿Qué cosas se harían igual y cuáles se harían de otra forma? ¿Por qué?**

Los tiempos de desarrollo del proyecto han sido demasiado justos, ya que cada componente del equipo ha intentado hacerse cargo de un bloque del proyecto. Sin embargo, si los primeros bloques no se terminan a tiempo ralentizan demasiado el progreso de los bloques finales.

- **¿Qué les ha aportado el desarrollar este proyecto?**

Este proyecto ha ayudado a conectar diferentes conocimientos obtenidos de los diferentes módulos cursados en el bootcamp durante la especialización de Data y el pre-bootcamp. A fin de cuentas, el tener un objetivo con un cliente ficticio y poder desarrollar el problema, la búsqueda de métricas que podrían ser reales y poder colaborar en equipo ayuda a que el desarrollo del proyecto sea más fácil.