Rental Dataset: Take Home Assessment

Subhra Mukherjee

Part 1

1. **three** most important KPIs for Client
    - **Conversion Rate** = (Total Customers renting on the platform)/(Total Customers visiting the Platform). The CR is an indicator how attractive the platform is and how convenient users find HA to be
    - **Average Transaction Margin** = (Total Transaction Margin for HA)/(Total number of Transaction). The Transaction margin would give us a sense of profitability per transaction on the HA platform
    - **Total Visits on the Platform**

Total Transaction Margin (Profit) = **Total Visits on the Platform * Conversion Rate * Average Transaction Margin.** Any reduction in profitability could in principle be traced back to any of the three metrics stated here and then further taken a deeper dive into

Part 1

1. Convert

2. Assuming **Conversion Rate** goes down:

    - Investigate the causes that may impact Conversion Rate. e.g., the customer mix of new repeat visitors on the HA platform may have changed due to a recent influx of new visitors because of some recently run campaign
    - It's possible that the channel mix of the visitors may have changed with some of the Paid Search channels bringing in more visitors because of increased spends on them. Typically such visitors may be less prone to convert as they typically are looking for the best deals or compare prices
    - It's also possible that the HA platform has inducted some new feature on the platform (like an Automated Chatbot) which is turning away users
    - **Conversion Rate** broken down into segments (like Conversion Rate by Channel, or by new repeat users) would be helpful to understand the aggregated Conversion Rate and may help pinpoint the source of the problem

Part 1

1. Convert
2. three examples
3. **Three example of analyses**
    - Estimation of Fair prices of Listing based on the Features of the listing property. This will help regulate over inflated prices that listings are likely to put up causing dissatisfaction on part of the renters who then are likely to turn away from HA
    - An Life Time Value analysis identifying prospective renters that are likely to keep returning on the platform and turn into loyal customers over time.
    - A churn propensity model that will indicate how likely is a current renter likely to stop using the platform which could be used to reach out to users who are highly likely to churn
4. An **Lifetime Value Analysis** would need the visit information per prospect as they use the platform. This model for LTV prediction could be used as soon as a new user lands on the platform. If the prospect is recognized as a High potential LTV user then HA could provide some personalized discount on the first booking to ensure that they convert on the platform and become regulars on it. Standard Regression techniques could be used to generate such predictive models based on historical data

Part 2

2. **Task 1: Cleaning the Dataset**

- Firstly i ingest the 10000 row CSV file in bigquery and create a table out it in Google Bigquery (since i have access to it)
- A lot of the columns beginning with 'Furnished' seems to be having blank quotes "", Text Nulls , and actual values enclosed in "" as a string
- Total_size is also not a numeric column but has numbers enclosed in " " and blank quotes
- I have written out a **Data Cleaning.sql** script which should address all of these issues and transform like the following
    - "Yes" ⟶ Yes
    - Null (String) ⟶ *Null* (Data type null)
    - "" ⟶ *Null* (Data type null)
    - "34" ⟶ 34 (Numeric)

- The SQL file is to be found in the SQL folder attached herewith (Please note, the SQL dialect is for Bigquery and may slightly differ from standard SQL dialects like MySQL or Snowflake Sql)

Part 2

2. **Task 2: Rolling average of median monthly rental price development over time**

   - SQL attached herewith (**3 Month Rolling average of Median Price.sql**)
   - The peaking the rolling average of the median price is because of a spike in the median price in November 2018. Almost a 220% increase in monthly median price

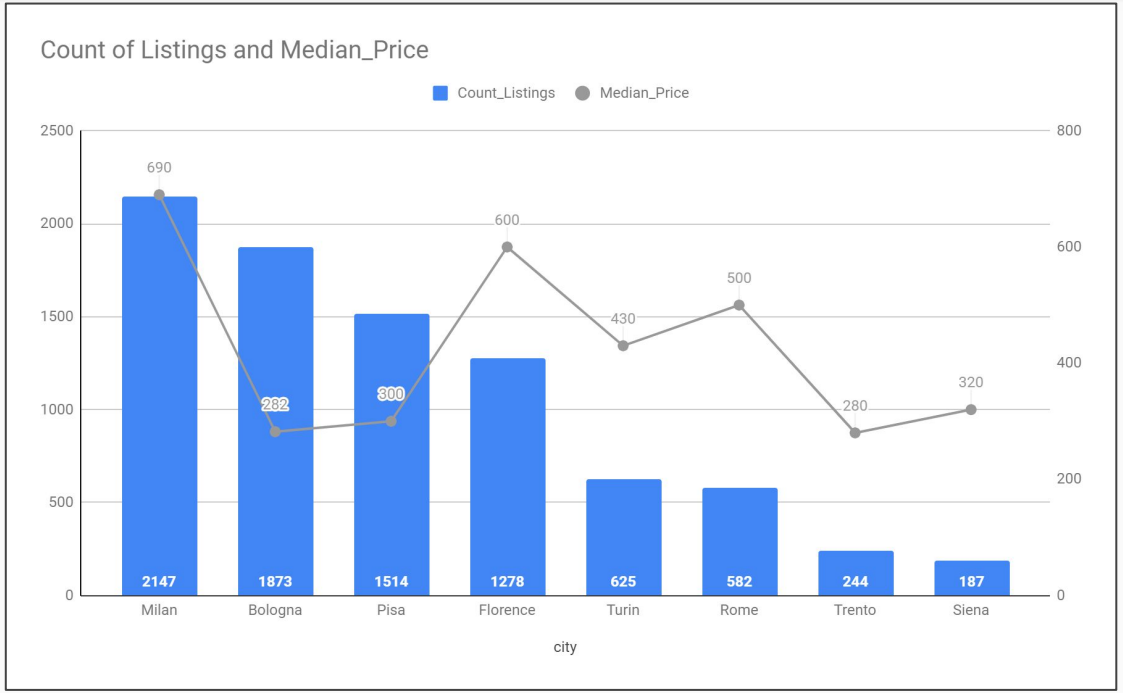
3. **Task 3 :** The Most important cities being Milan, Bologna , Pisa, Florence, Turin in terms of cities with the highest number of listings

   - SQL attached (**Important Cities.sql**)

Part 2

3. **Task 3 :** The Most important cities being Milan, Bologna , Pisa, Florence, Turin in terms of cities with the highest number of listings

    - SQL attached (**Important Cities.sql**)



Part 2: Insights

Looking at which of the Listing Features impacts the price that is set
(based on a Linear Model which explains price as a function of these features)

(R code of the linear model shared as HA_Linear_Price.R)

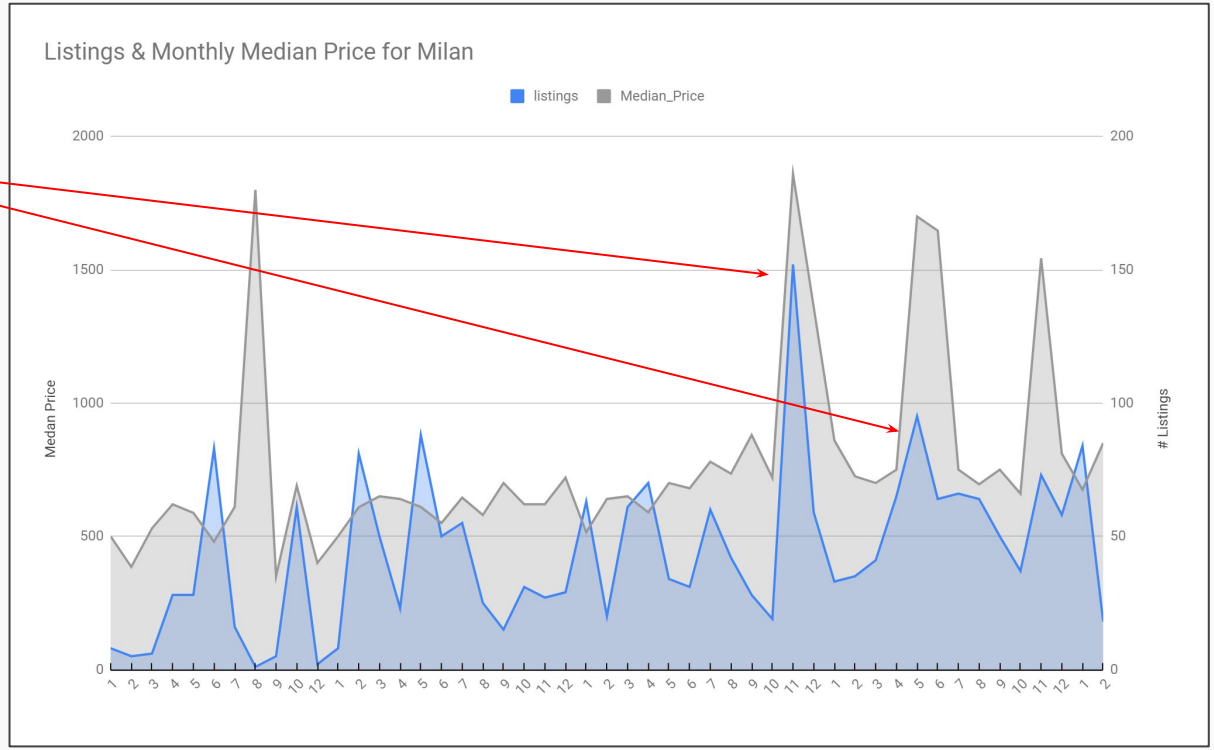
1. Size of the Apartment is directly correlated to price, larger size implies higher price
2. Listing category is important in determining price. Apartments and Studios is expected to be priced ~ €900 and €400 more respectively compared to a shared room
3. A furnished room is likely to be priced ~€70 more than a unfurnished room
4. Having a TV means that the room is ~€100 more expensive than an room without a TV
5. Balcony and Terrace doesn't seem to be that important in determining the price to be set
6. A private garden means that the room is ~€200 more expensive but a shared garden means that the room is ~€125 cheaper than a room without garden
7. Expectedly, as total size increases the expected price increases too. Every square meter increase increases the price by €0.05

Part 2: Task 3

Focusing on the City of Milan: Spikes in Listings is accompanied with a spike in the Median Prices

Hypothesis: Seasonally, a lot of listings will be coming up at certain times of the year and a spike in listings is most likely to be accompanied with a jump in the list prices.

Recommendation: Client could try to predict such spikes and pre-acquire some of the inventory for cheap in the months prior and re-rent it out during months when list prices are going to be high (if legally allowed)



Thanks!

-Subhra