Mix-shift Analysis for Demand Change in Budapest Real Estate Market During COVID

Capstone Project Summary

By Li Deborah Jia

Submitted to

Department of Economics and Business at Central European University

In partial fulfilment of the requirements for the degree of Masters of Science

Budapest, Hungary

2021

Table of Contents

About the Client

Ingatlan operates a marketplace platform (Ingatlan.com) for real estate selling and rental services in most Hungarian cities. The client is already a dominant player in this space since its establishment 20 years before.

After COVID erupted, the client's business was mostly affected. Thus, the client would like to updates their business recognition of their clients' demand regarding various listings. For this project, we worked directly under the supervision of a member of the data insights team of Ingatlan.

About the Project

COVID made a significant impact on the economy of Hungry. The real estate market from both the demand and supply ends has also changed in this period.

Ingatlan is the leading real estate platform in Hungary. Its core business is to connect sellers and buyers in its platform; meanwhile, this operation generates valuable data related to the Hungarian real estate market.

In this project, we would like to analyze users' demand change before and after COVID via data visualization, summarize the conclusion from these visual analytics, and transform these statistics into business suggestions for our client. This advice will help our client know its consumers better and make pertinent marketing policies for their business development.

We retrieve data from the client, clean the data and add extra variables. We first plot some general charts using R studio to gain a basic understanding of our data set. Then we create a Tableau dashboard and study the data based on property types and districts. Finally, we sum up all conclusions and give suggestions to the client.

Process the data

In this section, we check and delete duplicates, flag extreme values, winsorize outliers, drop rows with missing values. We also add extra columns explaining other variables and providing details of the data set.

We start with duplicates check, where we define "duplicates" as observations with all the same variables. We use the "duplicated" function to examine all observations.

Then we decide to check and process extreme values. We first replace some unique numbers that were input by the machine as default values. Then we draw a histogram of all the numeric variables. It turns out there are a considerable amount of outliers in the data set. We check the descriptive and decide on multiple benchmarks to winsorize the extreme values.

We also examine the missing values. As these missing values result from user behaviour rather than measurement error, simply dropping these observations won't help our study. We decide to drop observations only if they have missing values in the essential variables that we want to study.

In this step, I complete an R script that can iterate the data cleaning and processing; the result is a clean data set stored in a CSV file.

Data visualization in R and Tableau

We attempt to use the R language to plot some charts on a general level. Then, we create a Tableau dashboard to examine some crucial measurements based on two filters. These two filters are based on two categorical variables in our data set. A combination of filters will give us hundreds of charts (treemaps, histograms, doughnut charts, box plots, etc.) to explore. Considering the size of data and the number of plots we can create, we decided to interpret all general plots in R but only a few interesting figures in Tableau.

Project Outcomes

We start the project from scratch: learn about the client's demand, gather, process, visualize data, interpret the charts, and convert interpretations to business suggestions. We use SQL and R language to retrieve and clean data; we exploit R studio and Tableau to achieve data visualization.

We finally share a complete Tableau dashboard with the client, an HTML file based on R plotting, a technical discussion file, and a presentation video attached with PowerPoint slides. Also, we elaborate on the policy advice to our client, including the dominant property types after COVID, famous districts, change of people's expected price for a particular type of listing, etc.

Our first audience is the senior management of the data platform. Hopefully, through this mix-shift analysis, the leadership can first understand the core changes of customer demand and then transform their core business accordingly. Decisions to be driven include the marketing and staffing section. The senior leadership will consider adjusting more advertisement and human resources to consumer-preferred districts and listing types.

Lessons Learned

Doing a project means more than completing it once for all, nor does it tell that we only learnt relevant knowledge and skills from the surface. Through this project, I went beyond theories and case studies on textbooks, learnt more practical data manipulation and visualization skills while honing my old techniques. I came to understand two industries (real estate, online advertisement) better. Moreover, this project experience will help familiarize me with the industry operation and structure to guide my research work in these two industries in the future. In a word, everything I learnt from this project will be a massive asset to the establishment and refinement of my career path and academic interests.