Analytics Project – Predictive Modeling

M.Sc. Data Analytics and Decision Sciences

Univ.-Prof. Dr. rer. pol. habil. Sven Müller
Chair of Business Analytics
European University Viadrina

Summer Term 2020

1. **Pricing of real estate apartments (macro factors)**
   Assignment: NN, NN
   Description
   ALLMYHOMES is the platform for data-based development and
   marketing of condominiums located in Berlin, Germany
   (https://www.allmyhomes.com/category/uber-uns-en/)
   The project represents actual business problems with strategic relevance
   for ALLMYHOMES. The findings will support management decisions. We
   expect (at least): description of chosen data set, including source
   Explicit formulation of hypotheses (assumptions)
   Visualization of results (e.g. variable dependencies and trends). Due to the
   nature of our business model (real estate projects with long transaction
   cycles), we consciously decided not to supply (limited) internal data but
   rather let students research and gather relevant and sufficient external
   data (publicly available on online real estate portals like
   immobilienscout24.de or neubaukompass.de). By "new development" real
   estate, we refer to newly constructed projects in Germany built by large
   property development companies (Bauträger) with at least 10 apartment
   units (excluding e.g. smaller multi-family homes). ALLMYHOMES are
   available for questions at any time. Contact: Gaurav Sing (Director
   Business Development) & Paul Koeper (Pre-Sales Analyst).

   ● How are German new development real estate apartments priced,
   considering (in relation to) "macro" factors (e.g. environmental attributes)
   such as
   ○ location (e.g. ZIP code, expensive neighbourhood)
   ○ demographics (e.g. local population structure, income levels)
   ○ infrastructure (e.g. proximity to public transportation, schools,
   hospitals,
   playgrounds, restaurants)?
   ● Which of these macro factors (i.e. variables) correlate with each other,
   and which (can be isolated to) have the highest impact on price?
   ● Bonus: Based on historical data, is there any trend recognizable that
   could predict future (optimal) pricing patterns?

2. **Pricing of real estate apartments (micro factors)**
   Assignment: NN, NN
   Description
   ALLMYHOMES is the platform for data-based development and
   marketing of condominiums located in Berlin, Germany
   (https://www.allmyhomes.com/category/uber-uns-en/)

The project represents actual business problems with strategic relevance for ALLMYHOMES. The findings will support management decisions. We expect (at least): description of chosen data set, including source Explicit formulation of hypotheses (assumptions) Visualization of results (e.g. variable dependencies and trends). Due to the nature of our business model (real estate projects with long transaction cycles), we consciously decided not to supply (limited) internal data but rather let students research and gather relevant and sufficient external data (publicly available on online real estate portals like immobilienscout24.de or neubaukompass.de). By "new development" real estate, we refer to newly constructed projects in Germany built by large property development companies (Bauträger) with at least 10 apartment units (excluding e.g. smaller multi-family homes). ALLMYHOMES are available for questions at any time. Contact: : Gaurav Sing (Director Business Development) & Paul Koeper (Pre-Sales Analyst).

How are German new development real estate apartments priced, considering (in relation to) "micro" factors (i.e. attributes amongst apartments within the same building) factors such as

- Size of the apartment (in sqm)
- Number of rooms
- Floor level (ground floor, ..., penthouse)
- Structural apartment features (e.g. balcony, ...)
- Optional: Orientation (e.g. south-west, more light)
- Which of these micro factors (i.e. variables) correlate with each other, and which (can be isolated to) have the highest impact on price?
- Bonus: Based on historical data, is there any trend recognizable that could predict future (optimal) pricing patterns?

3. **Relationship of rent vs purchase prices (multiplier)**
   Assignment: NN, NN
   Description
   ALLMYHOMES is the platform for data-based development and marketing of condominiums located in Berlin, Germany (https://www.allmyhomes.com/category/uber-uns-en/)
   The project represents actual business problems with strategic relevance for ALLMYHOMES. The findings will support management decisions. We expect (at least): description of chosen data set, including source Explicit formulation of hypotheses (assumptions) Visualization of results (e.g. variable dependencies and trends). Due to the nature of our business model (real estate projects with long transaction cycles), we consciously decided not to supply (limited) internal data but rather let students research and gather relevant and sufficient external data (publicly available on online real estate portals like immobilienscout24.de or neubaukompass.de). By "new development" real

estate, we refer to newly constructed projects in Germany built by large property development companies (Bauträger) with at least 10 apartment units (excluding e.g. smaller multi-family homes). ALLMYHOMES are available for questions at any time. Contact: : Gaurav Sing (Director Business Development) & Paul Koeper (Pre-Sales Analyst).

- Context:
- In recent years the percental increase of German real estate purchasing prices
has been considerably higher than the percental increase of rents
- The Gross Rent Multiplier (Property Price / Gross Annual Rent) is one measurement to describe this relationship
- Not only has this multiplier increased in the past few year, but it also varies
significantly
- How has the multiplier developed in the past?
- Which factors determine the multiplier (i.e. the difference of rent vs purchase prices)?
E.g.
- Building (apartment) Location (ZIP code)
- Size of apartment (in sqm)

- Which of these micro factors (i.e. variables) correlate with each other, and which (can be isolated to) have the highest impact on price? Potentially discuss/exchange with Topic groups 1 and 2 (see projects above) to investigate major drivers (i.e. independent variables driving macro and micro pricing)
- Bonus: Based on historical data, is there any trend recognizable that could predict future multiplicator patterns?

4. **Latent variables mode choice model**
    Assignment: NN, NN
    Description
    Besides classical variables like distance and weather recent research discovers that so-called latent variables seem to strongly impact mode choice behavior. For example, not only the actual outside temperature impacts whether individuals choose to bike to work but also the perceived temperature. Based on a large mode choice data set from Rotterdam, NL the task is first to specify and estimate a simple MNL of mode choice considering standard variables. In a second step then we introduce latent variables build on individuals' evaluation of the weather into the MNL yielding a hybrid choice model. Implementation is done in R (Apollo package). Optionally, in a third step, we test for IIA and estimate a series of nested logit models. The report should contain a short review of the methods and the recent literature.

5. **Predicting electric vehicle purchase choices in Beijing, CN**
    Assignment: NN, NN
    Description
    In 2018 we carried out a stated choice experiment on car purchase choices in Beijing, China. The local government is interested in how the adoption of EV vehicles can be increased by providing a given density of charging stations. Our data set includes -- besides information on charging stations -- plenty further variables as controls. The task of this project is to (i) specify and estimate an interpretable MNL and (ii) build on this to come up with a nested logit specification trying to identify correlation between EV and hybrid cars. The models will be implemented in R (mlogit). The report should contain a short review of the methods and the recent literature.

6. **Predicting park-and-ride choice in Karlsruhe, GER**
    Assignment: NN, NN
    Description
    We carried out a stated choice experiment on mode choice in Karlsruhe in 2018. The alternatives are car, park-and-ride, and public transport alone. The task is to specify first an interpretable MNL. Based on this we investigate (i) the correlation structure between the alternatives, i.e., we specify various nesting patterns. (ii), we account for the panel structure of the data by an adequate specification of the error terms. The models will be implemented in R (mlogit / Apollo). The report should contain a short review of the methods and the recent literature.

7. **Predicting boarding time in public transportation vehicles.**
    Assignment: NN, NN
    Description
    In 2018 we carried out a pedestrian experiment on boarding and egressing passengers in busses / trams. The experiment settings accounted for

varying accessing and egressing passengers. The task is to specify a model that is capable of predictions of total boarding times of a coach given the number of egressing and accessing passengers at given doors. The models will be implemented in R. The report should contain a short review of the methods and the recent literature.