Dokumentation zur Visualisierung der Open Government Data (Visualisierungstechniken ILV S2021)

# Hinweise zur Ausführung des Dashboards

Das Dashboard ist online verfügbar unter: <https://biegt.shinyapps.io/real_estate_dashboard/>

Um das Dashboard lokal auszuführen, die Datei *04\_dashboard/dashboard.R* in RStudio oder einer anderen DIE öffnen und ausführen („*Run app*“). Die Installation zusätzlicher Packages ist gegebenenfalls notwendig.

# Dokumentation zum Entstehungsprozess und inhaltlichen Entscheidungen bei der Lösung der Aufgabe

Die größte Herausforderung bei der Lösung der Aufgabe war die **geringe Qualität des vorliegenden Datensatzes**. Durch diese geringe Qualität ist es schwierig sinnvolle, valide Interpretationen auf Basis der Daten zu finden, die gleichzeitig Relevanz im Sinne einer „Story“ besitzen. Symptome dieser geringen Datenqualität sind unter anderem:

* Der Datensatz beinhaltet ca. 22% fehlenden Werte
* Die Kodierungen zur Widmung sind uneinheitlich, so dass sie zur Analyse kaum bis gar nicht herangezogen werden können
* Oft fehlt der Preis pro Fläche in Bezug auf die jeweiligen Transaktionen
* Die Kodierungen für Erwerber und Veräußerer sind nur ungenau aufgeschlüsselt
* Es ist unklar, inwieweit die angegebenen Preise inflationskorrigiert sind
* Es ist unklar, inwieweit der Datensatz repräsentativ ist, d.h. die Transaktionen in den jeweiligen Jahren erschöpfend erfasst

Zunächst wurden die einzelnen Variablen im Datensatz genauer betrachtet. Insbesondere die Widmung sowie die Zuordnung von Wohnzonen und Schutzzonen wäre für die Analyse interessant gewesen. Allerdings sind in Bezug auf Wohnzonen und Schutzzonen die Fallzahlen sehr gering. Darüber hinaus ist die Kodierung der Widmung wie bereits erwähnt inhomogen, so dass eine tiefergehende Analyse dieser Variablen nicht möglich beziehungsweise sinnvoll war. Hinsichtlich der **Aggregationsebene** erschien jedenfalls die **Betrachtung über die Zeit** sinnvoll, da so die im Datensatz enthaltenen Fälle in vollem Umfang in der Analyse berücksichtigt werden können (allerdings wurden nur Daten zwischen 1987 und 2019 gewählt wegen viele fehlenden Werten bzw. unvollständigen Daten in diesen Jahren). Da die Fallzahlen in den einzelnen Katastralgemeinden relativ gering sind, wurde entschieden die **Analyse pro Gemeindebezirk** durchzuführen. Aufgrund der Unklarheit hinsichtlich der Inflationskorrektur wurde nur die Anzahl der Transaktionen als wesentliche Variable innerhalb der „Story“ gewählt. Die Preise (relativ und gesamt) wurde aufgrund von Ausreißern als Median aggregiert und als sekundäre Variable zur Exploration in das Dashboard aufgenommen. Um die Story aus Sicht eines Unternehmens relevanter zu gestalten, wurde eine Perspektive auf **verschiedene Kundengruppen** und deren „Kaufverhalten“ gewählt (Privatpersonen und private Unternehmen als die zwei Erwerbergruppen mit den höchsten Fallzahlen im Datensatz). Die auf dieser Basis gewählte Fragestellung lautet entsprechend:

**Wie entwickelt sich das Kaufverhalten bei Immobilien von Privatpersonen und privaten Unternehmen über die Zeit und nach Gemeindebezirk?**

Zur Präsentation dieser Fragestellung wurde ein **Szenario** gewählt, in dem das fiktive Unternehmen „Müller Immobilien“ in einem internen Meeting auf Basis der Daten Investitionsentscheidungen in Bezug auf Kundensegmente und Bezirke treffen möchte. Es wurde exploriert, für welche Erwerber und in welchen Bezirken sich ansteigende **Trends hinsichtlich steigender Transaktionen** erkennen lassen. Um die Argumentation zu untermauern, wurden für die explorativ identifizierten Bezirke **Zeitreihen-Modelle** angepasst, die dabei unterstützen zukünftige Entwicklungen besser abschätzen zu können. Dabei ist anzumerken, dass diese Zeitreihenmodelle aufgrund der jährlichen Aggregationen, vergleichsweise kurzen Zeitreihen und fehlender Berücksichtigung äußerer Faktoren mit Vorsicht zu interpretieren ist. Anhand der Analysen wurden also Bezirke identifiziert, für die vermehrte Käufe durch private Unternehmen zu erwarten sind. Der Analyst von Müller Immobilien würde entsprechend den Standpunkt nahelegen, in diesen Bezirken in Grundstücke zu investieren, die an der Nutzung durch private Unternehmen ausgerichtet sind.

Als Tool zur Umsetzung der Visualisierung habe ich Shiny gewählt. Tableau ist ein intuitives, mächtiges Instrument zur Erstellung von interaktiven Visualisierungen (z.B. ist die Erstellung von sehr spezifischen Layouts und reaktiven Verknüpfungen mit nur wenigen Klicks möglich). Die Umsetzung mit Shiny ist im Vergleich dazu anspruchsvoller. Ich habe mich zur Umsetzung der Aufgabe für Shiny entschieden, da es sich hier um ein Open Source-Tool handelt, was auf der einen Seite in der Umsetzung komplizierter erscheint, aber auf der anderen Seite in vielerlei Hinsicht noch mächtiger und vielseitiger als Tableau ist. Aufgrund dieser Vielseitigkeit, Offenheit und Flexibilität denke ich, dass die vertiefenden Kenntnisse, die ich durch die Arbeit mit Shiny gewinne, sehr wertvoll für meine weitere berufliche Tätigkeit sind.