

# پروژه R درس آمار و کاربردها

فاز دوم

در این فاز از دیتاست `lyon_housing` که شامل اطلاعات مربوط به قیمت معاملات ملکی در ناحیه لیون و ویلوگتن (Villeurbanne) فرانسه بین سالهای ۲۰۱۷ تا ۲۰۲۱ است استفاده می‌کنیم که از اطلاعات عمومی منتشرشده توسط دولت فرانسه استخراج شده‌اند. ستون‌های این دیتاست به شرح زیر است:

<code>date_transaction</code>	Date of the sale (yyyy-mm-dd format)
<code>type_purchase</code>	Type of transaction ("ancien" = old, "VEFA" = sale before completion transaction <sup>1</sup> )
<code>type_property</code>	Type of property ("maison" = house, "appartement" = apartment)
<code>rooms_count</code>	Number of rooms
<code>surface_housing</code>	Surface of the housing in m <sup>2</sup> rounded to the lowest integer
<code>surface_effective_usable</code>	Surface of the housing in m <sup>2</sup> according to the Carrez law, surface with height under the ceiling inferior to 1.80 meters is omitted
<code>surface_terrain</code>	Surface of terrain in m <sup>2</sup>
<code>parkings_count</code>	Number of parking spots
<code>price</code>	The selling price of the property in €. This includes VAT and estate agency fees, but not the notarial costs
<code>address</code>	Address of the property
<code>district</code>	Name of the district/city the property is located in
<code>latitude</code>	Latitude of the property
<code>longitude</code>	Longitude of the property
<code>date_construction</code>	Inferred construction date of the property (yyyy/mm/dd format). Dates too old are set to 1990-01-01

---

<sup>1</sup> پیش‌فروش

در کنار این دیتاست، یک فایل `json` شامل مختصات ایستگاه‌های مترو و تراموای لیون قرار داده شده است که می‌توانید از آن استفاده کنید.

در مواجهه با یک دیتاست این‌طور نیست که همواره پرسش‌ها مشخص باشند، طرح پرسش‌های مناسب که منجر به استنباط‌های آماری ارزشمند شود یک وظیفه حیاتی آمردان است؛ لذا برخلاف فاز قبلی که سؤال‌ها و جواب‌ها مشخص بودند در این فاز آزادی عمل بسیار بیشتری دارید و خلاقیت شما در استنباط از داده‌ها و به دست آوردن نتایج نابیهی، در کسب نمره کامل از این فاز راهگشا خواهد بود.

همانند فاز پیشین، پاسخ خود را در یک فایل R Markdown آماده کنید به‌طوری‌که کدها، توضیحات و خروجی آن‌ها نمایان باشد. به‌طور مناسب از کامنت‌گذاری بر روی کد خود استفاده کنید. درنهایت فایل را `knit` کنید و فایل `.html` و `.rmd` را زیپ کرده در CW بارگذاری کنید.

## بخش اول - EDA

مانند هر دیتاست دیگری در ابتدا باید به تحلیل اکتشافی داده‌ها بپردازید. هدف از این کار شناخت بهتر داده‌ها و ویژگی‌هایشان جهت استنباط‌های آماری است. روش کاملاً مشخصی برای این کار وجود ندارد اما به‌طورکلی باید در این مرحله به بررسی توزیع مقادیر ستون‌های مختلف، ارتباط ستون‌ها با یکدیگر، `missing-value`ها و مقادیر معیوب، داده‌های پرت، الگوهای قابل‌مشاهده و ستون‌های جدید و مفیدی که می‌توانند از داده‌ها مشتق شوند بپردازید.

## بخش دوم - Inference & Visualization

حال سعی کنید با استفاده از ابزارهایی که در درس یاد گرفته‌اید چند استنباط آماری قابل‌توجه (دست‌کم ۴) و چند نمودار معنادار (دست‌کم ۴) را گزارش کنید. توضیح دهید که گزاره‌های استنباط شده چرا از نظر شما ارزشمند هستند همچنین دقت کنید که گزاره‌های آماری شما باید از نظر آماری قابل‌اعتنا باشند و نمودارهایی که برای نمایش [بخشی از] داده‌ها از آن استفاده کردید برای آن منظور مناسب باشند.

## بخش سوم - Estimation

فرض کنید شما دانشجوی دانشگاه لیون فرانسه هستید که در مختصات جغرافیایی `45.780234113880425, 4.865561717882041` قرار گرفته و قصد خرید منزلی را دارید. بهترین خانه را در کجا می‌توانید پیدا کنید؟ اگر لازم است ترجیحات خود در مورد منزلتان را که در محاسبه استفاده کردید ذکر کنید.

## Dataset License

*Copyright 2021 Benoit Favier*

*Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:*

*The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.*

*THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.*