**Data Analytics Challenge**

***Let’s analyze the Luxembourg Real Estate Market***

For this part of the PwC Student challenge, we want you to consider the hypothetical case of a client of PwC Luxembourg who is interested in real estate asset management in Luxembourg. They are looking into a variety of metrics and considering how various stakeholders will be impacted by these, as well as likely responses.

We have datasets including the price information for different properties in Luxembourg over time. The data also includes other features such as the size of the property, the status (existing or future construction), and the different prices for each region.

We provide you with the datasets and a set of instructions to help you exploit your data analytics skills and explore the real estate dataset. What we are interested in is to hear your side of the story after analyzing the data!

Don't hesitate to come back to us if you have any questions about how to analyze the data!

## 

## **Data Analytics Challenge - Key information**

### **What do you need?**

You need to choose a data analytics tool and download the dataset.

### **What do we expect?**

At the end of this challenge, we expect to receive your answers in the format described below.

### **How much time do you have?**

In order to provide a fair assessment, we are limiting the time of this challenge to 2 weeks.

### **Is the challenge meant for individuals or teams?**

This challenge is intended to be completed **on your own**, not in teams.

### **How to prepare your answers?**

We ask you to send us the following documents in a .zip file:

* You are free to use your choice of data analytics tools that are convenient for you, be it Excel, Python, Alteryx, Power BI, Tableau or something else. What we want is to be able to follow the workflow, or the source code, to understand exactly what you did (and why you did it).
* A short report including the plots you have produced and your description.
* In case you have produced multiple files, please also make a readme file explaining where to look first.

### **What happens next?**

The shortlisted students will get an invitation to join us at PwC Luxembourg for the next stage of the challenge! More details to follow.

Please format your .zip file as YOUR NAME\_StudentChallenge\_2022.zip.

You can then send your file to the following email address: lu\_student\_challenge@pwc.com

### **What are the assessment criteria?**

As mentioned, we will not score you on the tool you have utilized nor the quality of your source file. However, the important factors to assess your answers are as follows:

* We would like to see the steps have been done on your side.
* That the plots that you provide in your answers are valuable.
* The insights and the story you come up with after analyzing the data has a great impact on our assessment as well.
* If you can provide us with any further analysis on top of the questions that are asked here, extra marks will be considered for you.
* Readability counts! The easier it is to follow what you have done, the easier someone can verify and benefit from your insights. Bonus points are awarded for clarity.

### Evaluation System (100/100)

* Data Cleaning (15/100)
* Descriptive analytics (15/100)
  + Data set description
  + Rankings
* Main Analytics (30/100)
  + Comparisons
  + Advanced analytics
* Predictive Analytics (10/100)
* Real Estate Market Insights (30/100)
  + Median Multiple
  + Business Insights
* Report (readability, figures, Readme file …) *(10 bonus points)*

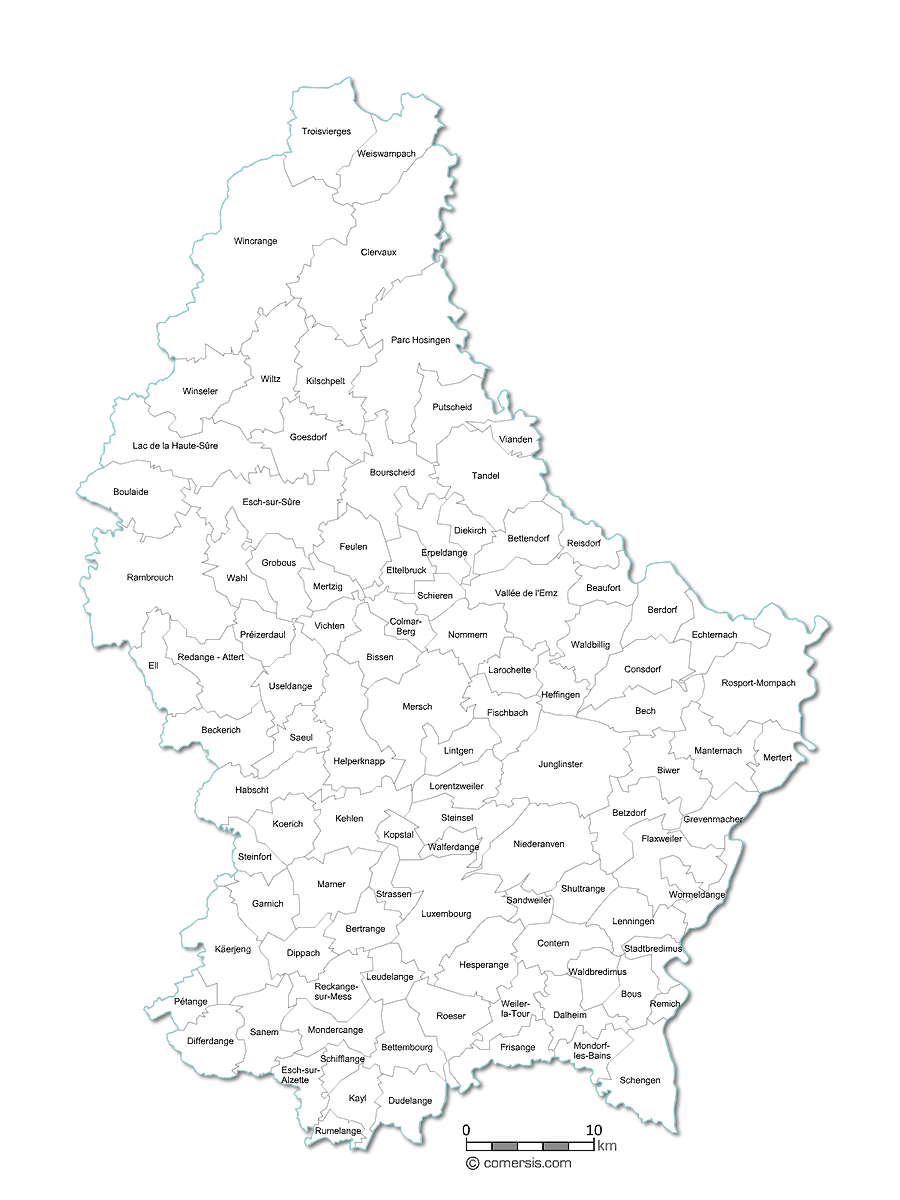
## 

## **About the Data**

The data set for this exercise concerns real estate prices in Luxembourg (the country and the city). The data is real, collected and anonymized by the [official Luxembourgish data platform](https://data.public.lu/en/), and publicly available.

The data comes in the form of Excel files showing the average price (whether to buy or rent) for a property in the specific area in a given year. All prices are expressed in Euro.

Not all prices are on the same level of aggregation. The two levels you will encounter are [Communes](http://www.citypopulation.de/en/luxembourg/cities/) and [Quarters](https://en.wikipedia.org/wiki/Quarters_of_Luxembourg_City). Communes are divisions on the level of the country; warning as Luxembourg City is also a commune. Quarters are divisions on the level of a city; in this case, the quarters are all part of Luxembourg City.

Fig. 1. Map of **Commune** divisions in Luxembourg

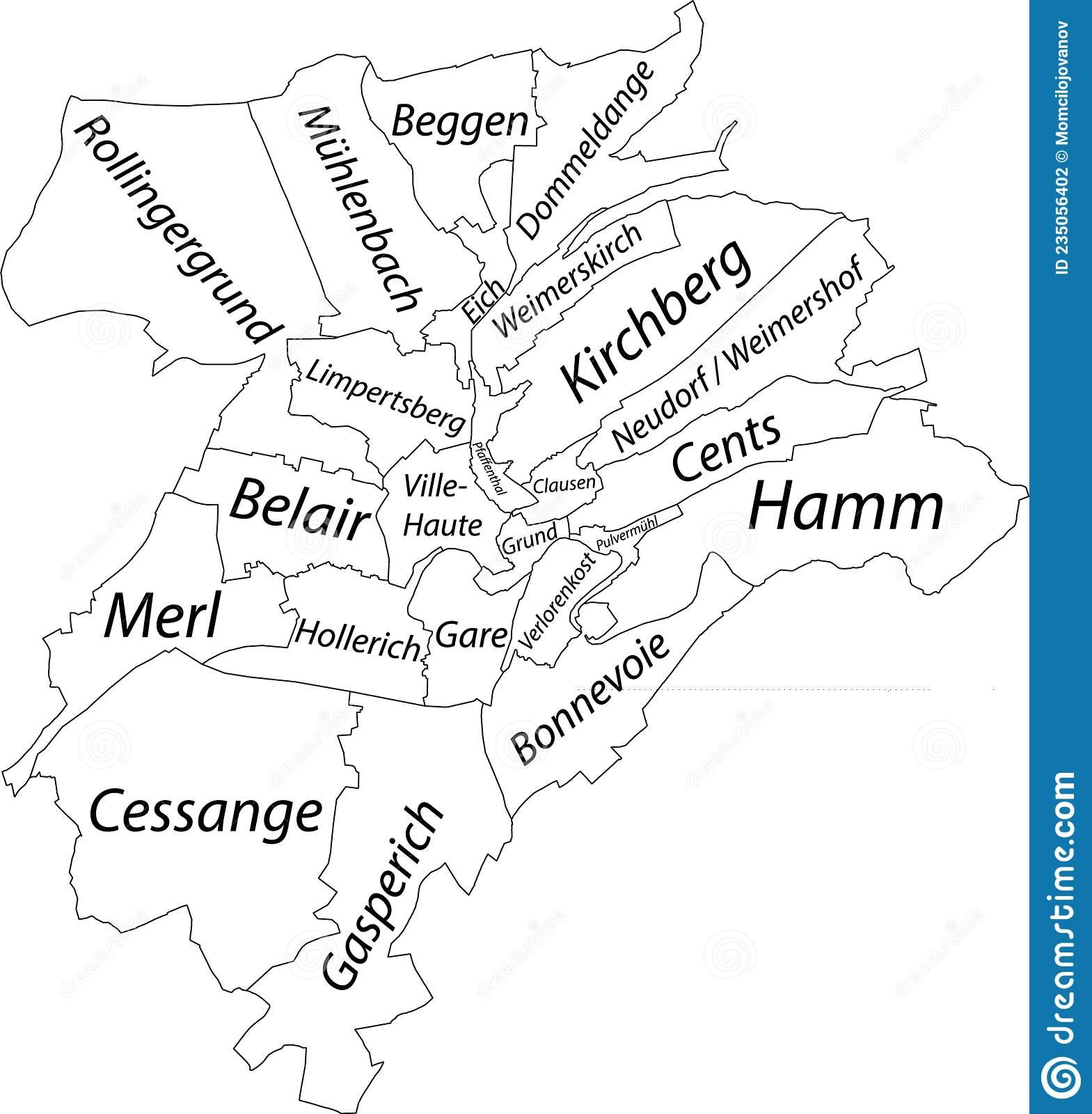


Fig. 2. Map of the **Quarters** of Luxembourg City

## 

## **Real estate data analysis**

### Let's learn about the data!

Download the data from the attached dataset and load it into your data analysis tool. Now that the data is imported, let's look at the fields and clean the data where necessary.   
Here is a checklist you can go through to clean your data.

* Check the data types and be sure that the values match the type. If necessary, adjust the type of the field.
* Look for the `null` values and decide how to treat them. Share with us your decision and let us know why you chose that option.
* Make sure that none of the entries are unexpectedly different from the others.
* Understand which fields connect different tables, and whether you expect to see the same entry in multiple places (why/ why not?).
* Pay attention to different levels of aggregation, and type differences and see if you understand what they mean.

### Descriptive analytics

The first thing to do, when you receive data, is to appraise it. You need to get an idea for what kind of information is in there, as well as some idea for the big picture view of the data. So let’s start by taking a look at it, and seeing what we notice. This type of analysis is known as *descriptive analytics*. We have prepared a set of questions to lead you through descriptive analytics in this dataset.

#### Data set description

1. What is the size of the data and the number of **observations** you find in the dataset for each field? Are there any unique or unexpected values?
2. What are the min, max, and mean of the **values** of each field?
3. How does data from one table **connect** to another?

#### Rankings

1. Which 5 **communes** have the highest/lowest prices?
2. Which property type (existing apartments, apartments still under construction, or houses) receives the highest number of offers? Are they directly **comparable**?

### Main Analytics

Our clients are interested in understanding the real estate market and would like to find the properties that retailers are investing in. Thus, we would like to present them with different scenarios by answering the following questions.

For clarity, terms are explained here:

The **sale** price refers to the price of buying a property outright, while the **rent** refers to the monthly price for renting a property.

The already **constructed** apartments are those who are already finished at the time of sale while the **future** apartments are those that are still under construction at the time of sale.

The **announced** price is the price at which the property has been put on offer. The **registered** price is the price for which the property has been sold.

#### Comparisons

1. How do **renting** prices compare to **buying (sale)** prices for apartments?
2. How do the **constructed** apartments compare to the **future** (VEFA) apartments?
3. How do the **announced** prices compare to the **registered** ones?

In the last few years, in Luxembourg city, are people going more for apartments or houses? Is that reflected in the prices? Do apartments and houses tell a similar story?

Have you found any interesting points when comparing these prices and how do you explain your findings?

#### Advanced analytics

1. Show the **growth evolution** for prices per commune. Does this differ significantly from the overall trend? Are there any outliers?
2. Calculate the year on year **growth per quarter** in Luxembourg City. What are the places with the highest (percentage) growth over the whole period?

### 

### Predictive Analytics

1. According to Fig. 3, how do you predict the future of apartment prices in Luxembourg? Will they stay the same, rise, fall?

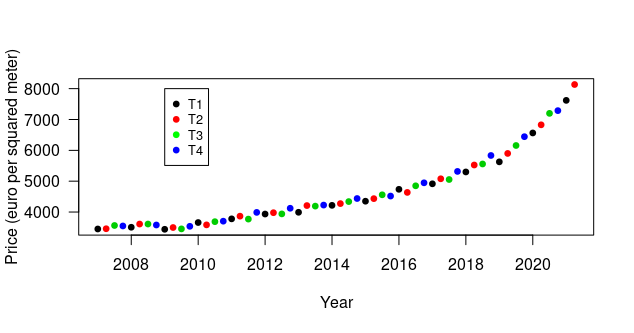
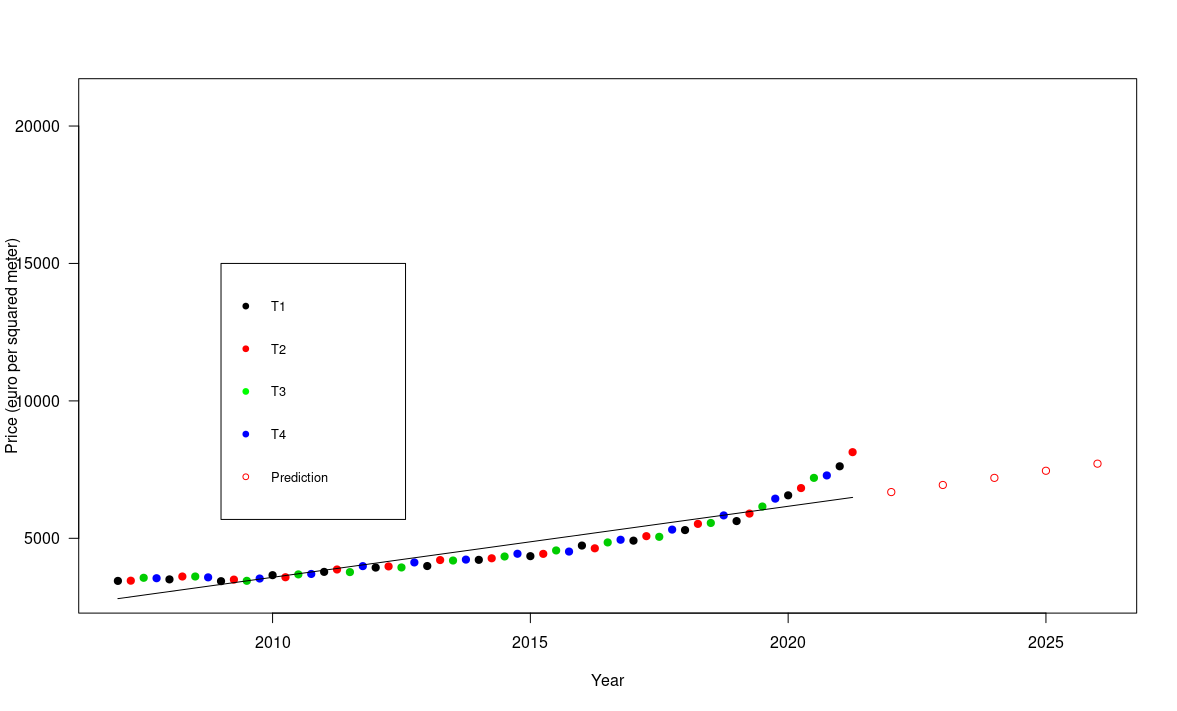
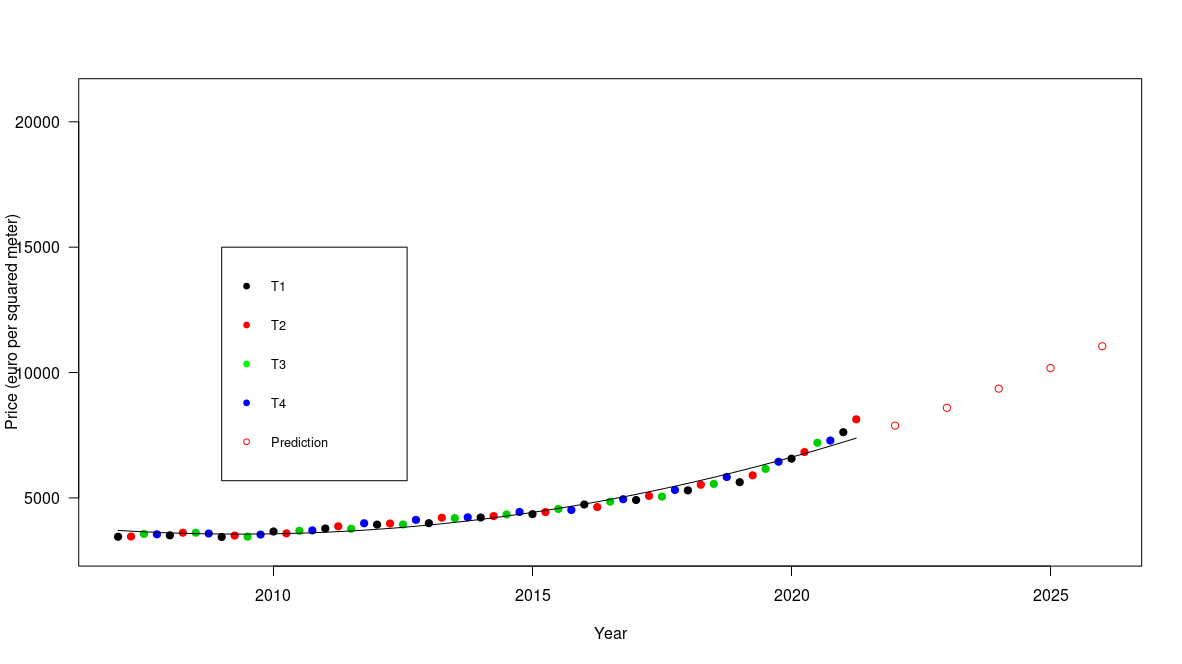


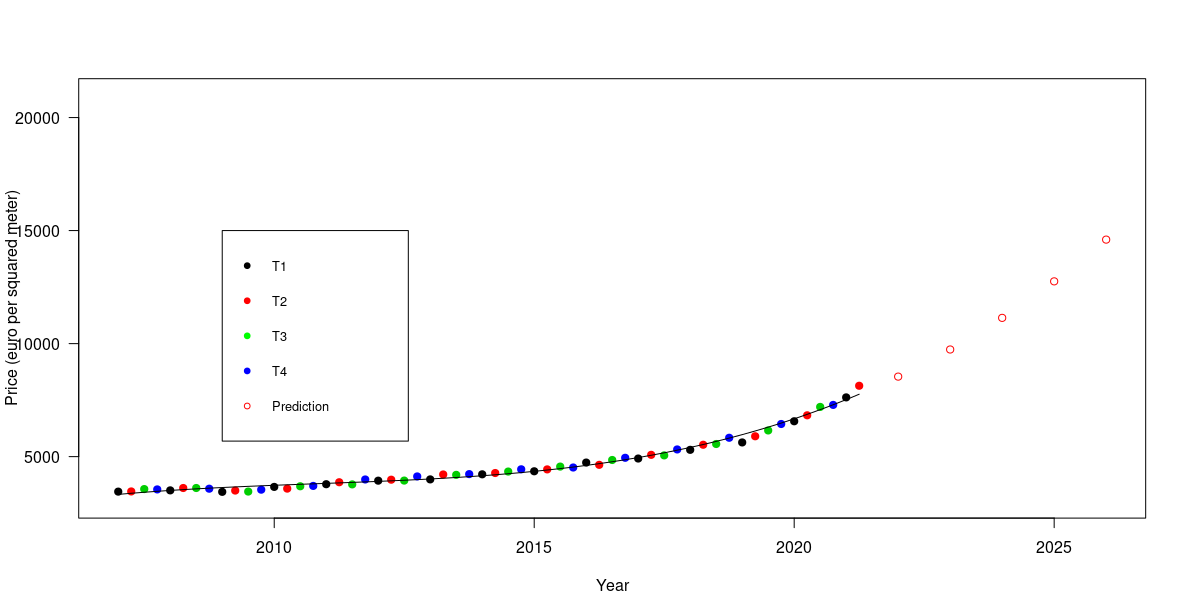
Fig. 3. The evolution of the apartment prices (per square meter) in Luxembourg (country).

The colors represent the trimester.

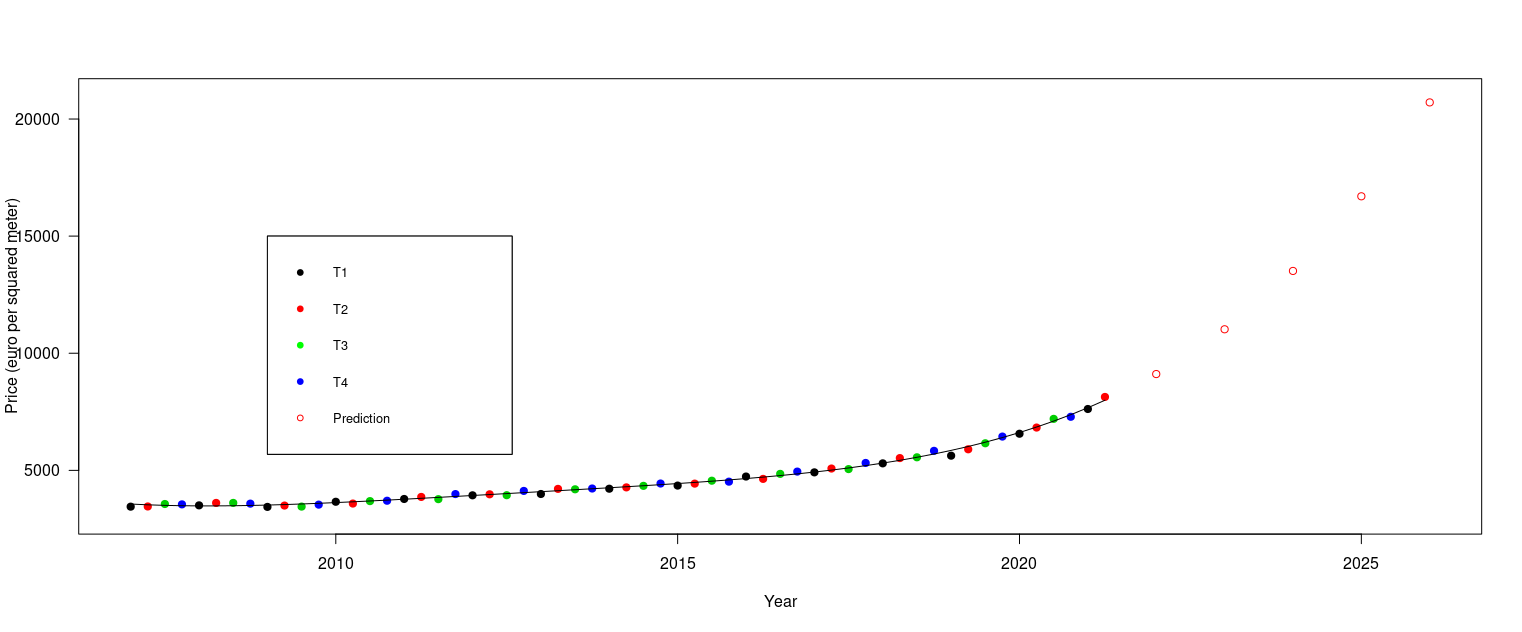
Out of the following predictions, which would you consider the most likely, and why?

a)

b)

c)

d)



### 

### Real Estate Market Insights

Now that you have had a look into the data, what are the key takeaways for you? You should always be keeping the objective in mind: coming up with actionable insights that generate value for the company that hired you. To conclude, here are some questions to allow you to express your insights.

#### Median Multiple

The **median multiple** is used to indicate the affordability of housing in any given community. It is defined as the ratio of the median house price by the median annual household (disposable) income.

Based on the [STATEC report](https://statistiques.public.lu/dam-assets/catalogue-publications/luxembourg-en-chiffres/2021/luxembourg-chiffres-21-en.pdf), the median disposable income per household for Luxembourg in 2019 was 5,454 €/month.

We would like you to calculate the median multiple for Luxembourg for 2019 using the house price shown in the following plot.

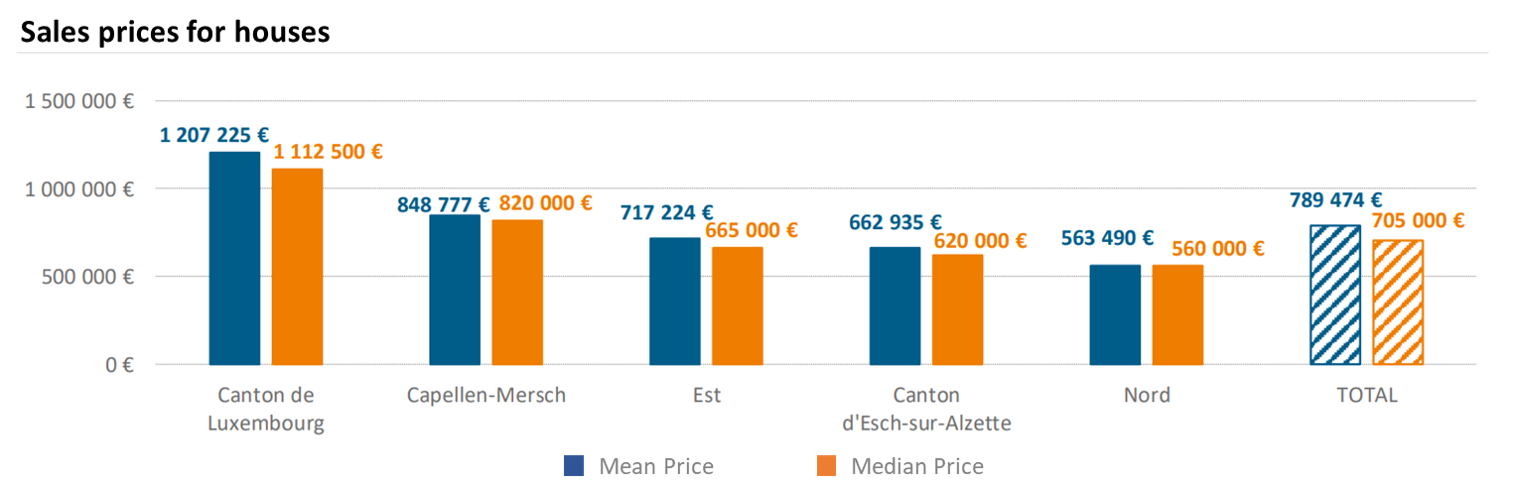


Fig. 4. Sales prices for houses in Luxembourg (country)

Based on the results, you can then discuss the housing affordability rate of Luxembourg according to the housing affordability ratings detailed in the Demographia International Housing Affordability Survey 2020.

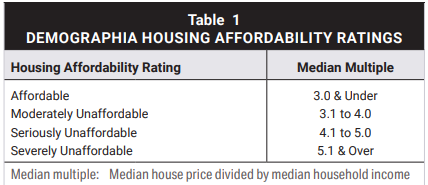


Fig 5. Housing affordability ratings

Also in the same report, the median multiples of some big cities are calculated as follows: Hong Kong, China (20.8), Sydney, Australia (11.0), Vancouver, Canada (11.9), Los Angeles, U.S. (9.0), Auckland, N.Z (8.6), London, UK (8.2). How would you compare Luxembourg's median multiple rate to these cities?

#### Business Insights

Now that you have a first perspective on Luxembourg’s real estate market, consider the following set of open discussion questions that the client might ask to get an overview over the data and inform their next decisions.

How would you answer or approach these questions?

1. What are the big trends (to buy/rent, invest in future/existing apartments) you have identified and what do you think is driving them?
2. What are likely government responses to what you have seen here? Is that what you would recommend, if you were the government?
3. How will the likely response(s) affect the other stakeholders?
4. What does the future look like in Luxembourg? Is this a good time to buy an apartment, or what would you wait for?

**Thank you and good luck for the challenge!**