Master 2 Econometrics and

Empirical Economics, TSE

Database Project, 2022-2023

**DATABASE PROJECT**  
**An Information System for Monitoring and Improving the Social Well-Being of Elderly People in a Cooperative Neighborhood**

**1. Description of the project**

In this project we are interested in neighborhoods such as Abricoop in Toulouse, which is “an intergenerational eco-district created and self-managed by its inhabitants”. This type of housing based on solidarity between inhabitants, the sharing of common places and resources (gardens, DIY workshops, laundry, guest rooms, etc.) allows the elderly to stay at home. Elderly people (seniors) will thus live surrounded by and interact with younger people and families with children. Nurses also regularly visit elderly people to monitor their states of health.

In addition to this solidarity, the cooperative has connected part of the district with IOT devices with the aim of improving the comfort of the inhabitants, making homes safer but also reducing the ecological footprint (such as controlling energy consumption). As far as the elderly are concerned, it calls on you to exploit this equipment and offer IT solutions to ensure their well-being (health, collective activities, social life) and to assist them in case of problems, and especially:  
  
1) *Improving situation awareness and social life*. The idea is to build **a database** that will provide meaningful information and services to : a) *inhabitants*: to keep them informed of collective activities and interact through a safe and homemade socio-media; b) *families:* to keep them informed of their parents’ health; c) *medical staff:* to supervise elderly people’s health.   
  
2) *Learn their habits* (normal behavior) during their day and night activities, and detect worrying changes (confinement, lack of social activities, malnutrition, etc.). These anomalies may be a sign of cognitive impairment or a loss of autonomy. From a trace of elderly people’s activities over several days (waking up, bathing, having lunch, going out, tinkering, etc.), the idea here is to discover and represent their behavior using processes, and to detect possible deviations from “normal” (expected) behavior. Alerts could be sent to their families and/or nurses if risky situations are detected. *This activity falls under the discipline of* ***process mining.***3) *Recommend daily activities.* Based on the habits of all elderly people, the idea is to suggest them actions or activities according to their situation (at home, outside, morning, afternoon, etc.) in order to improve their well-being. *This activity falls under the area of* ***Data Mining*** *in particular recommender systems.*

Une image contenant herbe, ciel, extérieur

Description générée automatiquement

**2. Description of the universe to be modelled in the database**

This database should give access to the following information. It should first record information about each *inhabitant*: last name, first name, category (senior, adult, young), address, phone number, gender and for elderly people (seniors) the references (name and phone number) of *a parent* (son, daughter…).

To engage inhabitants in***collective activities,*** *the cooperative* organizes different activities (painting, cooking, pottery, dancing, aerobics...). For each participant, we will record the list of activity sessions in which he or she participates. Each activity is also linked to a physical difficulty (0 to 5) and a price (often 0) to be paid by session and participant. It is assumed that, for a given activity, the price is the same whatever the session. For each activity session, we record also its date, starting and ending times, as well as a rating given individually by each participant. Each session is led by an animator (inhabitant) competent in the corresponding activity.

We also wish to set **up *a social network*** for the members (inhabitants) supported by this database. In this database, the members characterized by a name, a first name, a pseudo, an email, a login, a password and their skills. Each skill (painting, playing piano, cooking, dancing…) is characterized by a name. A member can indicate his/her skills and the level (beginner, intermediate, expert...). A member can also be recommended on a given skill by other members. Via the social network, the network members can interact: they are given the possibility to leave short comments (140 characters max). The date and the content of these comments must be recorded. These comments are only visible to other members who have requested to receive this member’s comments. Members should then be able to react to a comment in two ways: by simply indicating they appreciate the comment, or by writing a comment (140 characters maximum as above) on the previously one. These additional comments will be visible by all members who requested to see the comments produced by the author of the original comment.

For following **elderly people’s health state**, we also record medical information. A nurse is known by his/her surname, name, professional address, phone number and the seniors she/he follows and visits. A visit takes place on a given date, is described by a start time, a duration, reason of the visit (alarm, prescription…), a textual remark (changing behavior, health issues, falls…) and blood pressure and sugar measurements. We also record the prescribed *medication with* its starting time and its duration in number of days. Medicines are characterized by a name and a unique active substance (the one that will have a therapeutic effect). For the same active substance, there may be several medicines. For example, for the active substance called Paracetamol, there are several medicines (Doliprane, Dafalgan, Perfalgan, Efferalgan, etc.).

**Your mission, should you accept it, will be to:**

1. Draw an entity relationship model of this universe with the software Looping. Explain your choices; give your assumptions, data structures and integrity constraints.
2. Derive the previous model into a relational model.
3. Generate a physical database with ACCESS, MySQL, SQLite or other DBMS of your choice. Fill in the DB with relevant data. Create views or forms to ease your application use.
4. Give answers to the questions of Package 1 to 3 (see appendix 1) in SQL. Represent some answers with charts (bar graphs, histograms, pie charts…) using Excel, Tableau, R or Python.
5. Discover seniors’ behavior, detect worrying changes and check compliance (see section 3 for logfile structure and precise instructions).
6. *Research Question:* Implement an activity recommendation algorithm (see section 4, for instructions and file to be used).

**Expected deliverables**

* A document composed of an introduction and for each question, its answer (SQL or others) and a screenshot of the results produced. Also, include your assumptions, explanations and arguments when answering questions. The document should end by a conclusion.
* The sources of your software productions.
* The slides of your presentation.

**Deadline:**

* November 20: Document including Step 1) and 2) to be upload in moodle
* November 21: Solution for step 1) and 2) ;
* December 15th: pdf document and software developed to be put on moodle;
* January 9th: slides of your presentation to be put on moodle;
* January 10th: oral presentation and demonstration (1/2 hour by group)

**Conditions:** project to be done by pair.

**3. Discovering behavior and detecting worrying changes** **(*Process mining task*)**

We assume here that we have a trace over four days of the daily events (activities) of 2 seniors (Solange and Mario). These traces are given in the csv file: Solange\_And\_Mario.csv. This file represents normal behaviors of these two seniors (see also an extract of this file in appendix 2).

*Log file structure*. Each line corresponds to an activity carried out by a person in the context of a day. Each day relating to a person corresponds to a case. Thus, cases 1 to 4 relate to four days spent by Solange, cases 5 to 8 relate to 4 days spent by Mario.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** | 15/06/2022 16:30 | Collective Dancing | Solange | Adrian |

Figure 1. A logfile line

The line in figure 1 means that “In the context of case **1**, **Solange** started the **Collective Dancing** activity with the assistance of **Adrian** on **15/06/2022** at **16:30.**

Solange's behavior on the first day (case 1) can be summarized by the following sequence of activities: *WakeUp. Going to toilet. Taking a shower. Dressing. Taking a breakfast. Walking in the garden. Collective Game. Having Lunch. Watch TV. Walking in the garden. Phoning. Collective Dancing. Having Dinner. Reading. Sleeping.*

From day to day the behavior varies.

**Your mission**

From the log file, you will discover, represent and analyze the behavior of each person over several days. You will provide high-level models, visuals and indicators.

1. **Calculate of indicators to better understand the log file (SQL, Excel, R, BI, tool, etc.) and represent them with graphics.**
2. Create a new case for Mario (different from the previous one, with a difficulty or an improvement: more interactions with others, more activities…);
3. Time spent by each activity for each person?
4. Average time per activity for each senior (Solange and Mario).
5. Comparative evolution of the time devoted to collective activities
6. Percentage of time spent outdoors for each person.
7. What are the tasks carried out by Solange that Mario does not do and vice versa?
8. What are their common tasks?
9. Can we deduce a profile of each of them?
10. **Mining Solange’s behavior**

i. Discover the process synthesizing cases 1,2,3 and 4 in the form of a Petri net.

j. Deduce also Solange’s social network (handover relation, consider the helper attribute).  
k. Play the Petri Net with the Tina software. Be sure we can retrieve Solange’s behavior.

**3. Analysis of conformity and deviations in the behavior of Solange (SQL and/or programming language)**

l. A new day (day 5) is observed for Solange following this sequence:

*Wake Up (A). Going to toilet (B). Sleeping (C). Dressing room (D). Going to toilet (B). Sleeping (C). Having Lunch (E). Watching TV(F). Going to toilet (B). Having Dinner (G). Going to toilet B). Sleeping (C).*

Should we be worried? How to measure this problem?

m. Measure the rate of essential activities carried out by Solange considering the 5 days? Draw the evolution of this rate.

n. Draw the evolution of the percentage of outdoor activities?

o. conformance checking

If we represent a trace by a word, for example this one above by: ABCDBCEFBGBC. We can measure the difference between two behaviors by the distance between these two words: that is to say the number of modifications to make the two words identical. We can use the *Levenshtein distance* for which we can find an algorithm [here](https://fr.wikipedia.org/wiki/Distance_de_Levenshtein) (to be implemented in Python). By comparing this new trace to those that constitute Solange's normal behavior, what conclusions can you draw? Discuss the relevance of Levenstein measure for this problem?

Dorsaf Zekri, [Thierry Delot](https://dblp.org/pid/d/ThierryDelot.html), [Marie Thilliez](https://dblp.org/pid/t/MarieThilliez.html), [Sylvain Lecomte](https://dblp.org/pid/04/523.html), [Mikael Desertot](https://dblp.org/pid/56/3168.html):  
A Framework for Detecting and Analyzing Behavior Changes of Elderly People over Time Using Learning Techniques. [Sensors 20(24)](https://dblp.org/db/journals/sensors/sensors20.html#ZekriDTLD20): 7112 (2020).

1. **Recommending activities to Solange (Research Question)**

The file activity-rating.csv (to be downloaded on the course page) contains the rating  
of users (including Solange and Mario) on an online platform for the seniors:  
● There are 21 possible activities : “Morning gym” is Id 21 and “Painting” is Id 18  
● A large number of users (how many?) : Solange is Id 66 and Mario is Id 69  
● Each line of the file is of the form (USER ID, ACTIVITY ID, RATING)  
Answer the following questions using existing tools and/or by programming them with your favorite programming language:

1) Calculate the similarity (Euclidean distance) between Solange and Mario.  
2) According to a recommendation algorithm of your choice, should we recommend “Morning Gym” to Mario? What are the top three activities that are recommended to Solange?

**Appendix 1: Queries to implement (mainly with SQL)**

**Package 1: Management of activities**

F1. Create a form or a view that displays for each activity its sessions and animators.   
R1. Activities planned between two dates.   
R2. Free activities having a difficulty less than a given value.   
R3. List of activities mixing seniors and young people.  
R4. Number of sessions per activity with their average rate and maximum duration (in minutes). Limited to activities having at least two sessions and display them in the descending order of the average rate (most appreciated first!)  
R5. Percentage of activities animated by elderly people.

**Package 2: Monitoring elderly people’s health state**

F2. Create a form or a view that displays for each senior all the information about his/her nursery’s visits.

R6. Number and average time of nursery’s visits for each senior? To be displayed in the descending order of the number of visits.

R7. Seniors having had bad measurements (about blood pressure or blood sugar) for the past 7 days. (Checker avec blood pressure)

R8. Draw the evolution of the time spent per day in collective activities by a given senior for the past 7 days.

R9. Senior who did not interact through the socio-media for the past 7 days?

R10. Draw a word cloud with the most significant words appearing in nurses’ remarks.

**F2.**

SELECT I.FirstName & " " & I.LastName As NameSenior,V.\*

FROM Visit as V

INNER JOIN Inhabitant as I

ON V.InHabitantId = I.InHabitantId

WHERE I.Category = "Senior"

No where statement bc INNER JOIN, ptetre ajouter d’autres tables.

**R6.**

SELECT I.FirstName & " " & I.LastName As NameSenior,

FROM Visit as V

INNER JOIN Inhabitant as I

ON V.InHabitantId = I.InHabitantId

WHERE I.Category = "Senior"

GROUP BY I.FirstName, I.LastName

ORDER BY COUNT(V.VisitId) DESC

**R7.**

SELECT I.FirstName & " " & I.LastName As NameSenior

FROM Visit as V

INNER JOIN Inhabitant as I

ON V.InHabitantId = I.InHabitantId

WHERE V.DateVisit >= DATE() - 7

AND ( V.BloodPressureSystolic > 2

OR V.BloodPressureDiastolic > 2

OR V.SugarLevel > 2)

**R8.**

SELECT

**R9.**

SELECT

**R10.**

SELECT V.TextualRemark

FROM Visit as V;

**Package 3: Situation awareness**

F3. Create a form or a view that displays for each senior his/her social network   
R12. Number and percentage of seniors participating to activities.  
R13. Percentage of young and adult participating to activities.  
R14. List of inhabitants in the descending order of the number of followers. Give their name, category, and the number of followers.

R15. Draw the social network of the “follower relation” (explore Gephi tool or do it with R or Python). Analyze it.  
R16. Draw a word cloud with the most significant words appearing in the comments.

**F3.**

SELECT

FROM

WHERE

**R12.**

SELECT COUNT(T.SeniorInActivity) As CountOfSeniorInActivity,

COUNT(T.SeniorInActivity) \* 100 / COUNT(I.InHabitantId) as Pct\_Of\_Senior\_In\_Activity

FROM (SELECT DISTINCT P.InHabitantId as SeniorInActivity

FROM Inhabitant AS I

INNER JOIN Participation AS P

ON I.InHabitantId = P.InHabitantId

WHERE I.Category = "senior") AS T

RIGHT JOIN Inhabitant AS I

ON I.InHabitantId = T.SeniorInActivity

WHERE I.Category = "senior"

**R13.**

SELECT COUNT(T.YoungAdultInActivityID) As CountOfYoungAdultInActivity,

COUNT(T.YoungAdultInActivityID) \* 100 / COUNT(I.InHabitantId) & " %" as PctOfYoungAdultInActivity

FROM (SELECT DISTINCT P.InHabitantId as YoungAdultInActivityID

FROM Inhabitant AS I

INNER JOIN Participation AS P

ON I.InHabitantId = P.InHabitantId

WHERE I.Category = "young" or I.Category = "adult") AS T

RIGHT JOIN Inhabitant AS I

ON I.InHabitantId = T.YoungAdultInActivityID

WHERE (I.Category = "young" OR I.Category = "adult")

**R16.**

SELECT C.Content

FROM Comments as C;

**Appendix 2**

***Extract of the logfile:***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| \_CASE\_KEY | EVENTTIME | ACTIVITY\_EN | Senior | Helper |
| 1 | 15/06/2022 09:00 | WakeUp | Solange | Self |
| 1 | 15/06/2022 09:03 | Going to toilet | Solange | Self |
| 1 | 15/06/2022 09:10 | Taking a shower | Solange | Self |
| 1 | 15/06/2022 09:25 | Dressing | Solange | Self |
| 1 | 15/06/2022 09:34 | Taking a breakfast | Solange | Self |
| 1 | 15/06/2022 10:00 | Walking in the garden | Solange | Self |
| 1 | 15/06/2022 10:30 | Collective Game | Solange | Elsa |
| 1 | 15/06/2022 12:05 | Having Lunch | Solange | Self |
| 1 | 15/06/2022 13:00 | Watching TV | Solange | Self |
| 1 | 15/06/2022 15:00 | Walking in the garden | Solange | Self |
| 1 | 15/06/2022 15:30 | Phoning | Solange | Self |
| 1 | 15/06/2022 16:30 | Collective Dancing | Solange | Adrian |
| 1 | 15/06/2022 19:00 | Having Dinner | Solange | Self |
| 1 | 15/06/2022 20:30 | Reading | Solange | Self |
| 1 | 15/06/2022 22:30 | Sleeping | Solange | Self |