Managing the safety-index of cities

in frame of the service management

Laszlo Pitlik (sen), Laszlo Pitlik (jun), Matyas Pitlik, Marcell Pitlik (MY-X team)

Abstract: The paper tries to demonstrate in a rel. reproducible way how easy/simple models can be derived for rational decision-making constellations based on real data assets and based on similarity analyses. The focused data are the crime statistics for a lot of US cities (2017). Through the safety-index as such, it is possible to show how can be interpreted an unknown index? How can be explored the impact of new variables? It means: whether a higher population size of the cities makes the life in these cities more or less safe? What kind of crime-type should be decreased at first in order to achieve the highest safety increasing without involving information about cost-levels of the development actions or even based on information about the resource allocations of each improvement action? The managing of the safety of cities can be interpreted as a kind of service management challenges being relevant for e.g. the tourism…

Keywords: chained similarity analyses, risks of modelling, consistence of modelling

# Introduction

This paper is the newest part of the series about experiences of the QuILT-based education processes. Previous articles and their annexes can be downloaded here:

* <https://miau.my-x.hu/miau/quilt/Definitions_of_knowledge.docx> + annexes like:
  + <https://miau.my-x.hu/miau/quilt/demo_questions_to_important_messages.docx>
  + <https://miau.my-x.hu/mediawiki/index.php/QuILT-IK045-Diary>
  + <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK045-Diary>
  + <https://miau.my-x.hu/mediawiki/index.php/QuILT-IK059-Diary>
  + <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK059-Diary>
* <https://miau.my-x.hu/miau/quilt/reality_driven_education.docx> + annexes like:
  + <https://miau.my-x.hu/miau/quilt/chained-translations-legal-slang.docx>
  + <https://miau.my-x.hu/miau/quilt/demo_chained_translations.docx>
  + <https://miau.my-x.hu/miau/quilt/demos_chained_translations.docx>
  + <https://miau.my-x.hu/miau/quilt/forum_details.docx>
  + <https://miau.my-x.hu/mediawiki/index.php/QuILT-IK057-Diary>
  + <https://miau.my-x.hu/mediawiki/index.php/Vita:QuILT-IK057-Diary>
* <https://miau.my-x.hu/miau/quilt/Exercises_for_critical_thinking_and_doing.docx>

* <https://miau.my-x.hu/miau/quilt/st1_all.docx>
* <https://miau.my-x.hu/miau/quilt/20Q.docx>
* <https://miau.my-x.hu/miau/quilt/GDP_final_en.doc>
* <https://miau.my-x.hu/miau/quilt/st2_all.docx>
* <https://miau.my-x.hu/miau/quilt/harmony.docx>

Parallel, there are a lot of spreadsheets supporting the needs for details: <https://miau.my-x.hu/miau/quilt/?C=M;O=D>

The last paper about the harmony-index is a kind of adaptive cloning of the GDP-oriented thinking experiments. This paper demonstrates an analytical process (based on the KNUTH’s principle) where the first impulse was just a single phenomenon (a keyword for searching): “**crime statistic of cities**”. Each detail of the whole modelling process will be found in the background document: <https://miau.my-x.hu/miau/quilt/crime_v1.xlsx>

Based on the given searching keyword, online data assets can mostly be explored within minutes: e.g. <https://www.google.com/search?ei=v-ahXLjmDdH6qwGbjZGgBQ&q=crime+index+by+city+usa> where the first 2 items could be used at once (see <https://miau.my-x.hu/miau/quilt/crime_v1.xlsx>).

After the potential data assets could be identified, it means: we have an impression about the amount and type of objects (like US cities) and their attributes (like crime-types and/or aggregated index-values), the analyst should derive the real problems needing a solution – as far as possible a robot solution (c.f. KNUTH’s principle).

The Figure Nr1 is a part of the background spreadsheet: <https://miau.my-x.hu/miau/quilt/crime_v1.xlsx>

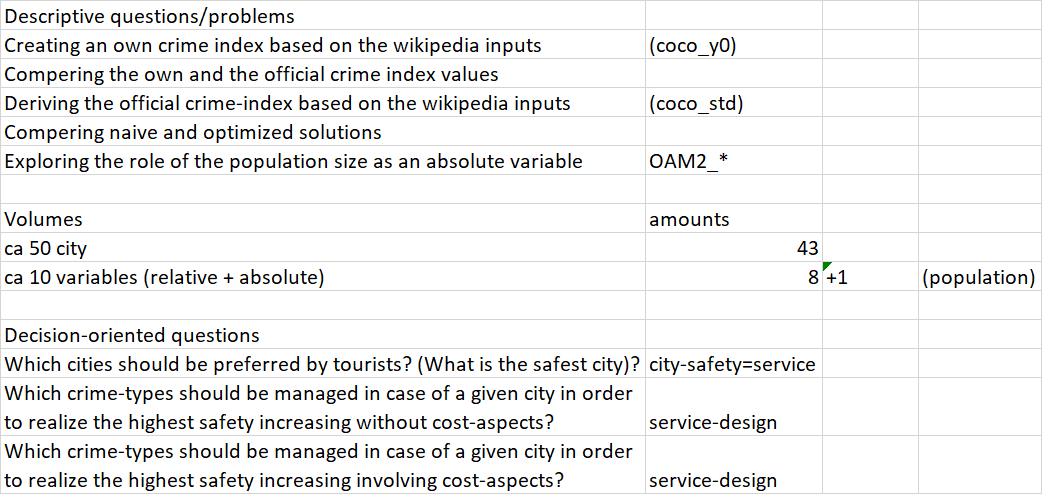


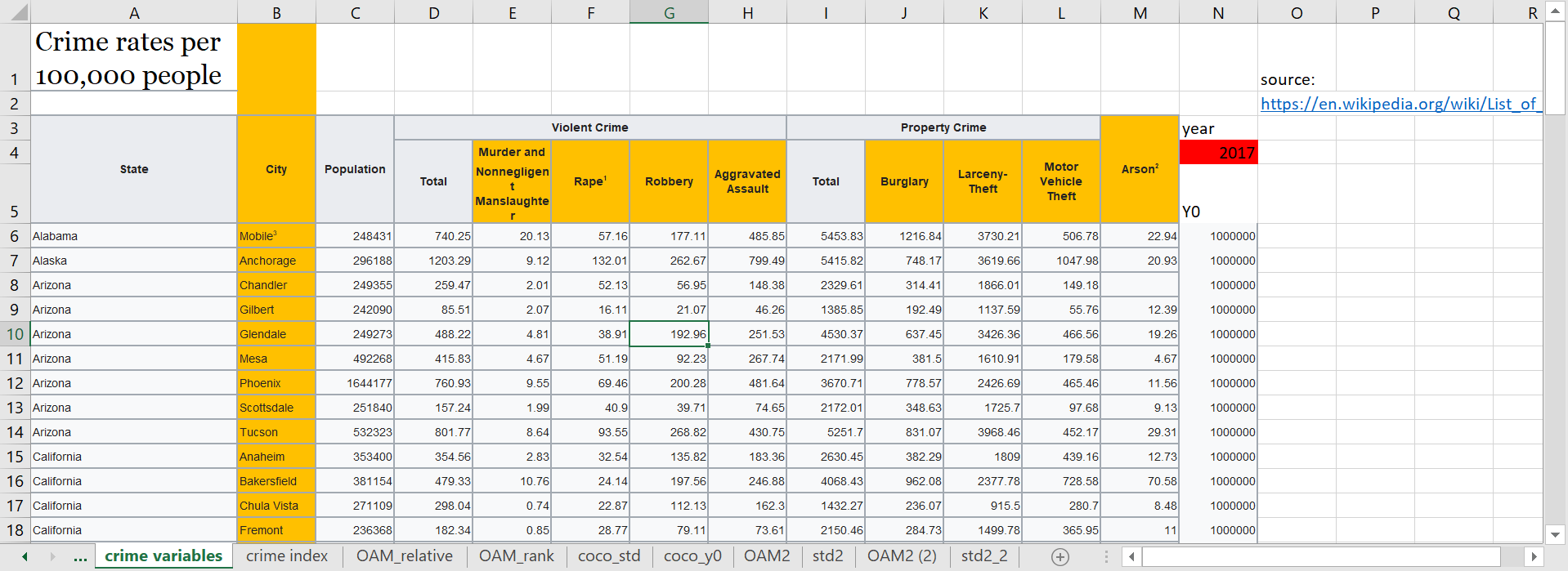
Figure Nr1: Descriptive and decision-oriented problems for a set of available data (source: own presentation)

If somebody (a team) has the data and the problems, then it is possible to start with the detailed working:

# Steps of the data processing

The raw data has mostly not the expected structures – even the raw data should not have any structure (c.f. PDF-pages without any possibility to use visual given structures). In this particular case, the data has the necessary structure and quality:

* objects (cities) can be identified and searched (c.f. VLOOKUP)
* attributes of the objects can be identified pro object
* the different sources (c.f. basic crime-types and index-values) can be joined based on the objects (it means – cities are available in both sources)
* the description of the objects seems to be consolidated (it means – cities can be identified without any misunderstandings in the both sources)
* lack of data can be interpreted, and it is possible to handle them (it means: the not available data about arson-events can be substituted through average values
* the crime-types are relative data it means we do not need any conversion for comparing them
* the population-data are absolute data where it seems to be relevant to derive whether the more (is the population) the more (is the safety) or the less the more principle seems to be more valid
* crime-index + safety index = 100 in case of each city
* the calculation scheme for the crime-index of the literature is already not known



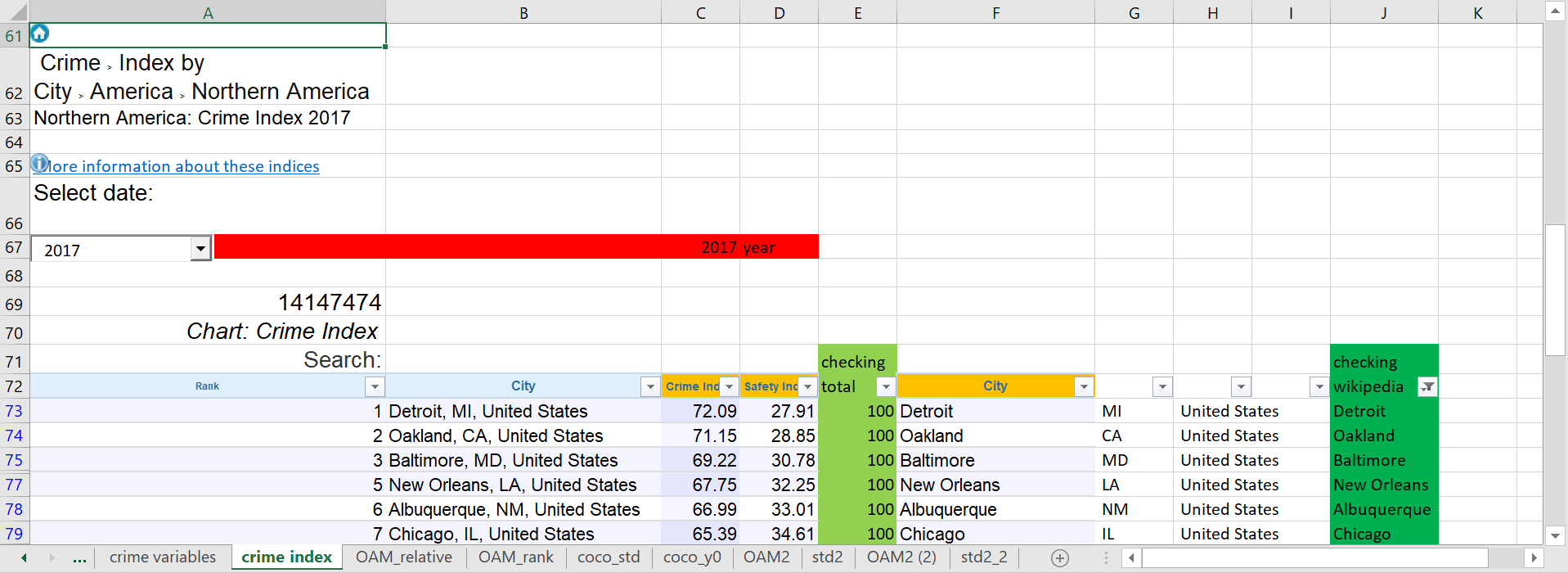


Figure Nr2: Raw data (source: own presentation)

The cities (in case of the crime-index) are available integrated in a cell with data about the (US) states (see column “B” in lower part of the figure). The delimiter is the “comma”. Therefore, it is possible to create three columns from the original one (see columns B=F+G+H).

The crime-index and the safety-index can be checked (see: checking total). Based on a VLOOKUP-function, it is possible to identify cities being available in both datasets. The filter-effect (see column “J” ensure the cities both with index values and crime-types.

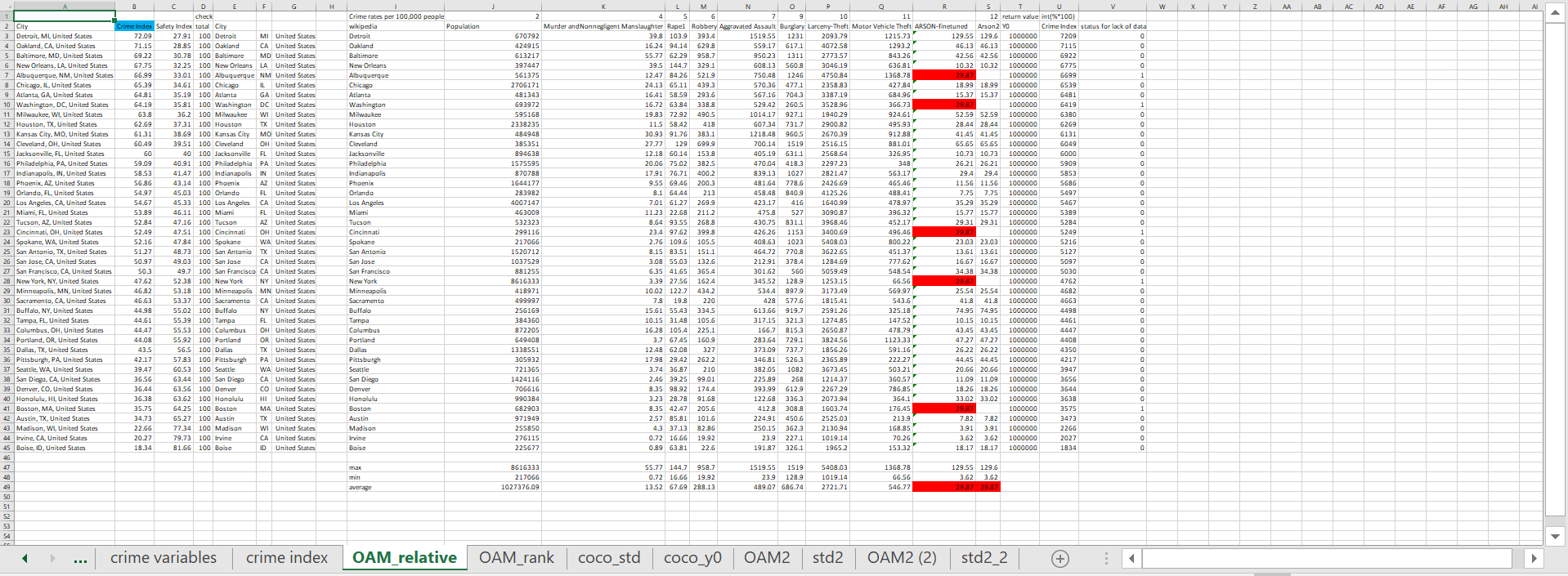
Figure Nr3: The OAM (source: own presentation)

Figure Nr3 presents the cities (objects) and their attributes based on a new VLOOKUP-transformation. The lack of data concerning the arson-statistics can be identified in case of 5 cities from the selected 43 (see Figure Nr2). A binary status variable got created in order to describe these cities (see column “V”. The blank cells were filled with the average number derived from the available positions. It should be highlighted, that the maximum and minimum values of the already relativized variables are very different.

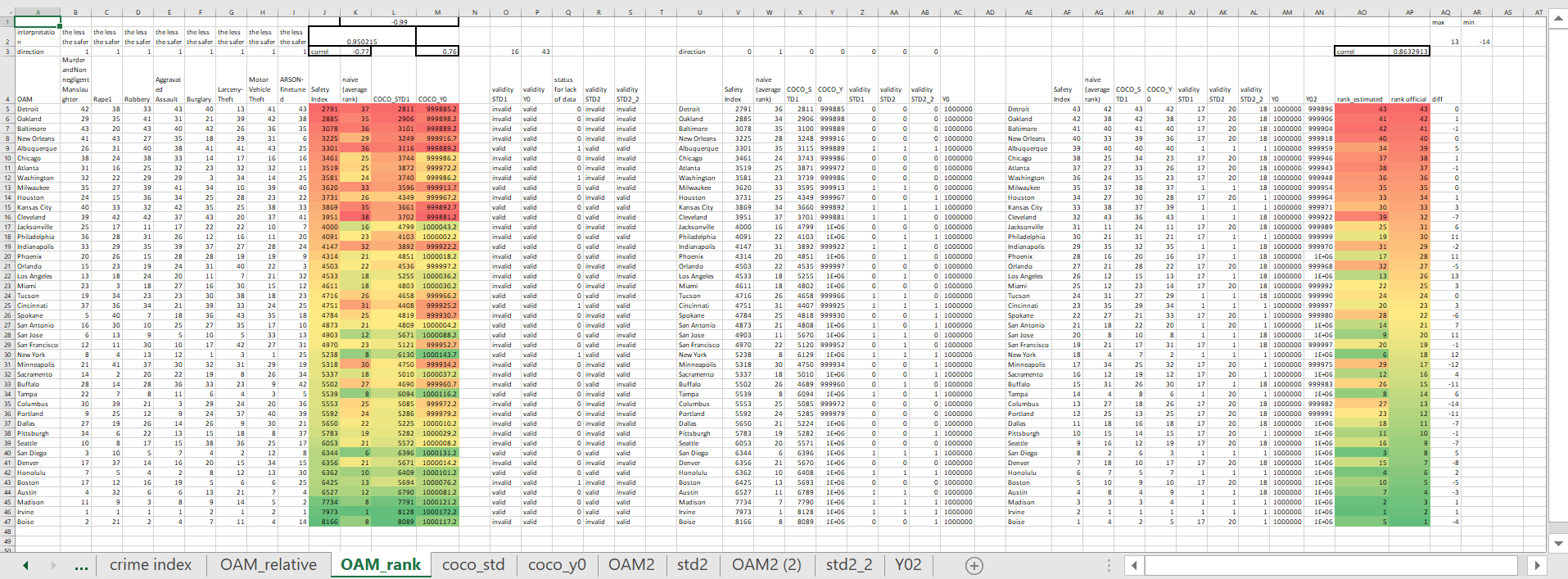


Figure Nr4: The ranking view of the OAM with the final Y0-model (source: own presentation)

The ranking view (see Figure Nr4) includes the crime-types as independent variables and the safety index (as Y-variable). The coloured cells (in 4 columns) demonstrate

* the original safety-index (on the left side – column “J”),
* the naïve safety-index (derived from the ranking values as the average value of them – column “K”),
* the first “learning” model (STD1 – column “L”),
* and the anti-discriminative model (Y0 – column “M”).

The first model (STD1) is a kind of production function where the online robot engine (responsible for similarity based learning - <https://miau.my-x.hu/myx-free/coco/index.html>) tries to reproduce the original safety-index-values derived from the ranking values of the crime-types.

The anti-discriminative model (Y0) tries to derive an anti-discriminative evaluation value for each city based on the ranking values of the crime-types without involving the information about the “official” safety-index.

Both modelling approaches (STD and Y0) work similarity-based. It means they can be evaluated based on function-symmetries. The validity values are for the anti-discriminative model excellent. But the production function produced a lot of invalid signs. This high-level invalidity can be interpreted so, that the safety index will be created from other/further crime-types and/or the calculation mechanism of the safety index is very complex.

Figure Nr4 delivers a lot of correlation values:

* the correlation between the naïve and the official safety index is: -0.77 (it means: the naïve approach uses ranking numbers – what leads to the minus sign in a logical way – and the value 0.77 is rel. high but not unlimited high what can be seen as a further sign to the already discussed validity signs
* the correlation between the official safety index and the anti-discriminative model is: 0.76 (it means: the naïve model could not be improved – and the calculation scheme of the original safety index could not be approximated better)
* the correlation between the naïve and the anti-discriminative values is: -0.99 (it means: the crime-types do not have identifiable weighting)
* the correlation between the original safety index values and the values of the first model is: 0.95 (it means: the identified crime-types could be deep involved into the calculation scheme of the official safety index or the used variables should have high-level correlations to other variables used for calculations of the original safety index
* the coloured cells seem to have a massive parallelism concerning the red and/or green highlighted values where green is for the higher safety levels and red for the lower levels

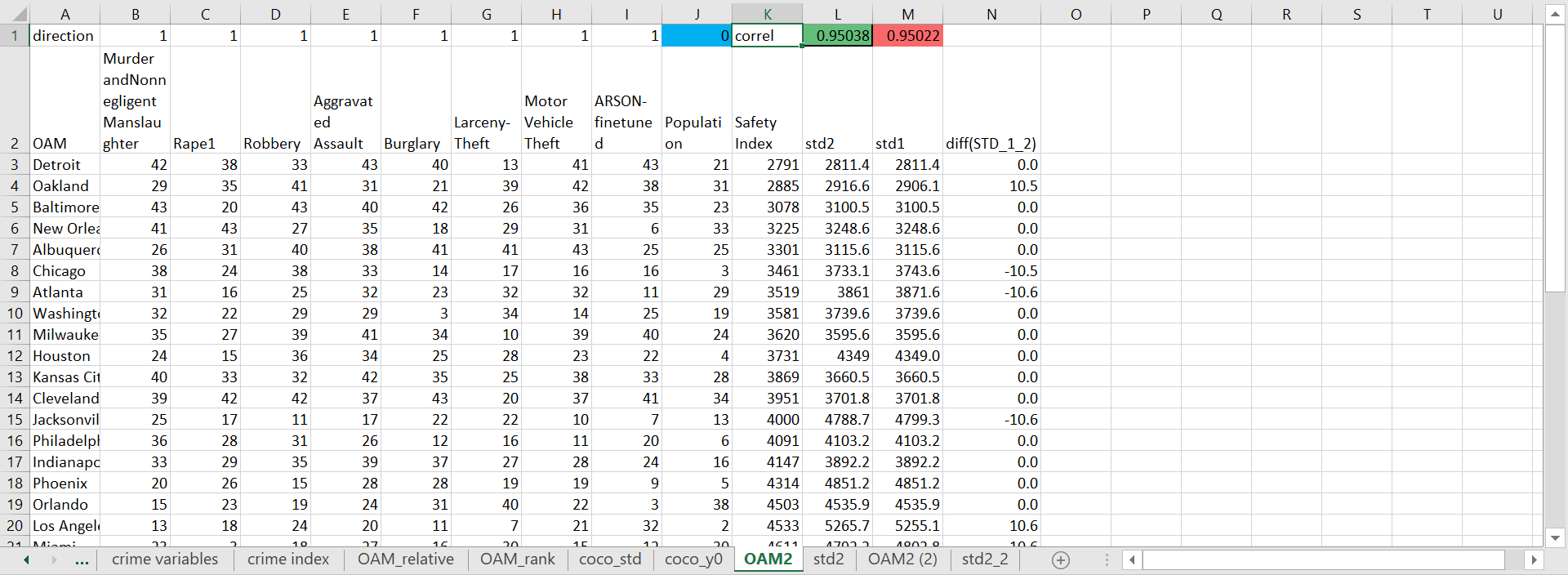


Figure Nr5: The advanced model including population size (source: own presentation)

Figure Nr5 shows a new OAM where the population size as absolute variable is integrated based on the direction: the higher is the population the higher is the safety (see blue-highlighted cell). As it can be seen: the new column (“J”) delivered hardly information. The correlation values are quasi the same. The more is information (STD2) the more is the correlation (0.95038 > 0.95022). The differences between STD1 and STD2 are mostly 0.

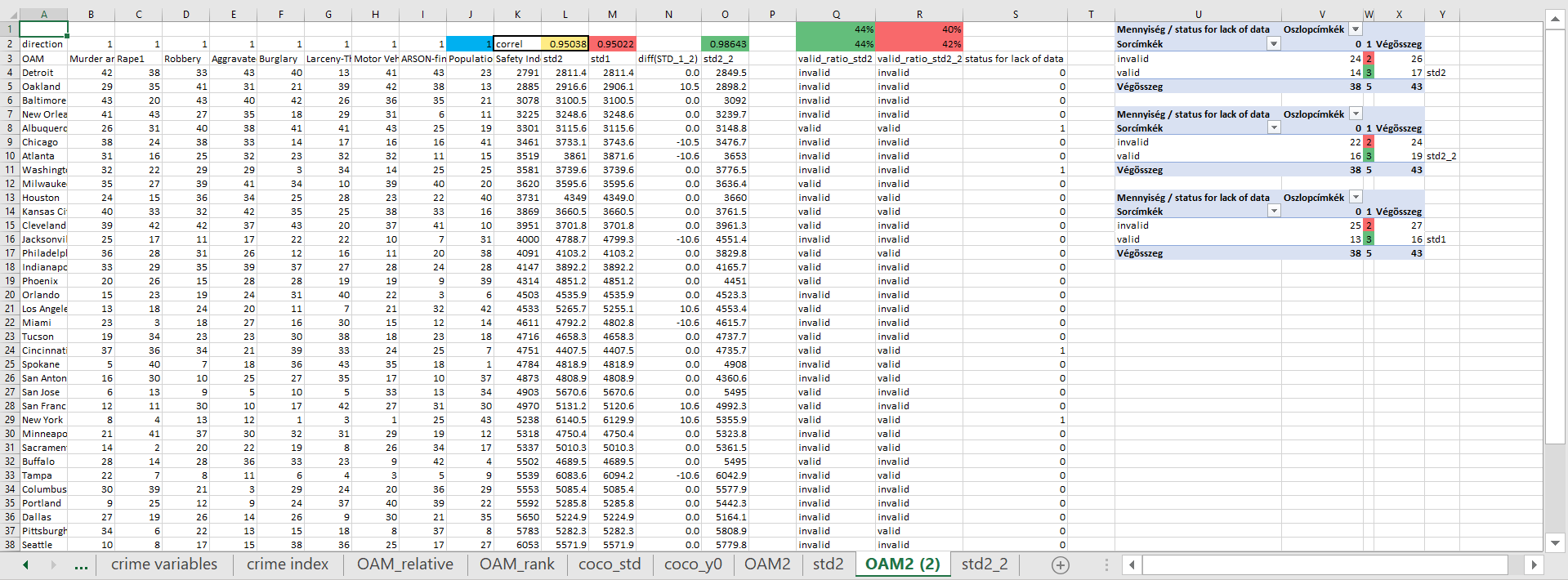


Figure Nr6: The new model with inverse direction (source: own presentation)

The population size can also be involved based on the inverse direction: the less is the population the higher is the safety. From system-theoretical point of view, this interpretation (direction) seems to be more robust because the higher systems (more people) might cause more problems through the more complex/diverse interactions. Figure Nr6 demonstrates the correlation values between the original safety index and the estimation of the model STD2-2. This correlation is higher than before for the model STD2 with the other direction for the impact mechanism concerning the population size and safety. Unfortunately, the validity values would lead to the inverse conclusion where the validity levels about trivial and/or fine-tuned model-risk show more advantage for the model STD2 than for the model STD2\_2. This is a typical inconsistence effect which should be minimized. The columns “U-V-W-X-Y” can not bring any new impulses for the solving of the above-mentioned inconsistence problem. The validity ratios are the same for each model (STD1, STD2, STD3 in case of the cities having a substitution value for the ARSON-positions).

# Decision-oriented interpretations

The question “Which is the safest city?” could have been answered quasi at once

* based on the safety index values – if somebody has these information units and believe them
* in those cases, if somebody has just the data about the crime-types, then it is possible to create a naïve and/or anti-discriminative approximation of the safety index
* Figure Nr4 demonstrates that the original/official safety values are not robust enough compared to the naïve and/or anti-discriminative evaluations – therefore, it was necessary to create a final anti-discriminative model (see Figure Nr4) where
  + each evaluation information unit was integrated into a final model
    - Safety Index
    - naive (average rank)
    - COCO\_STD1
    - COCO\_Y0
    - validity STD1
    - validity STD2
    - validity STD2\_2
  + the antidiscriminative model technique evaluated the more or less robust but not consistent basic information units for each city
  + the final estimation value set has a correlation compared to the official safety index value set of 0.86 based on the ranking values of the cities
  + the winner is not the winner of the official statistics (see Irvine > Boise)
* Figure Nr4 highlights that the official ranks can have some risks because the estimated ranks could have a position difference between -14 and +13 ranking units compared to an object-set with 43 objects.

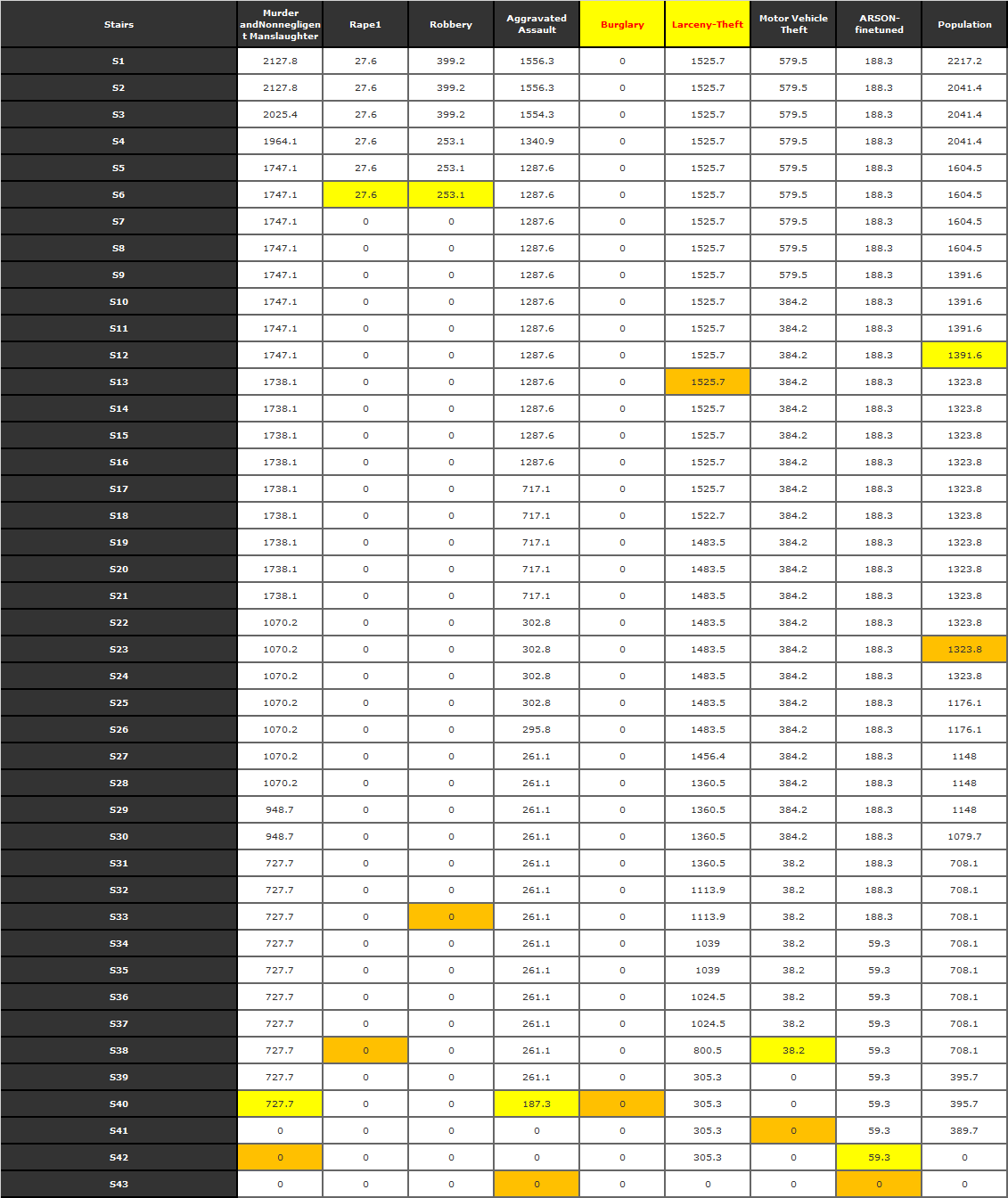
The above interpreted ranking information could be used as a kind of input for geographical decisions for private persons (see: where to study? where to organize a sightseeing?) and/or for enterprises (see: where to work?)

Of course, it is necessary to involve each further variable. From these analyses, two information units (variables) could be selected:

* the higher is the estimated safety index the better is the place for activities
* the lower is the difference between the “official” and estimated safety ranks the better is the place for activities

# The case of Detroit without cost-aspects





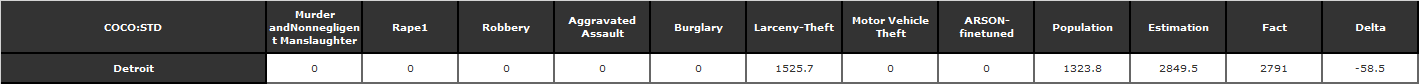


Figure Nr7: Knowledge map for Detroit (based on the model STD2\_2) – (source: own presentation)

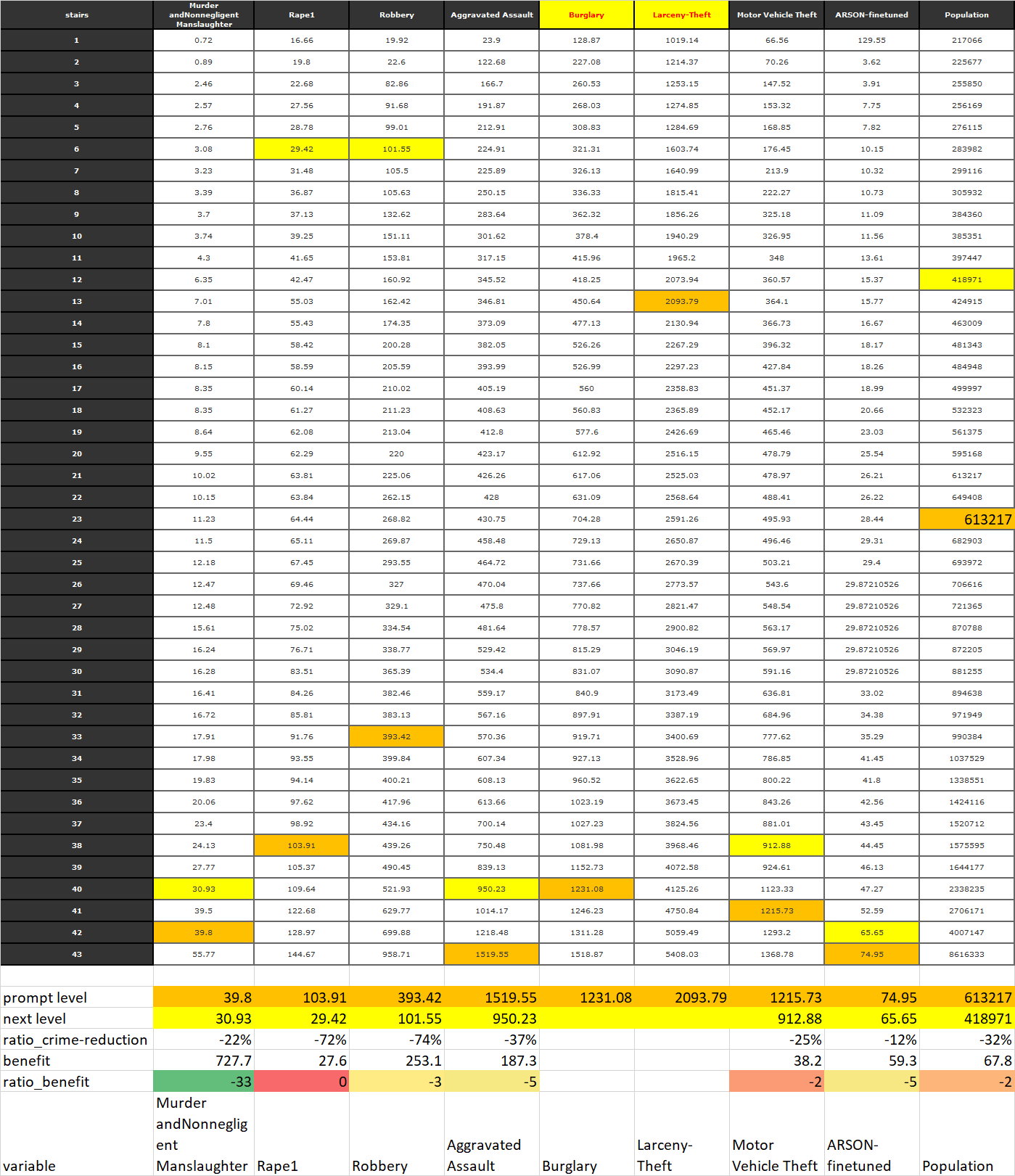


Figure Nr8: Impact-estimation for crime-types (source: own presentation)

Figure Nr7 shows the real levels of Detroit (orange) and the next ranking level having the less safety-increasing (yellow). The cells are the safety-increasing parameters as such. The sum of the safety increasing parameters for a city should be close to the real safety index based on the similarities STD2\_2).

Figure Nr8 shows the same coloured pattern but in case of the ordered raw input values for each variable.

* The prompt level means therefore: Detroit has a value for 100.000 people of 39.8 cases concerning “Murder and Nonnegligent Manslaughter”, etc.
* The next level means: Detroit could reduce the above derived value from 39.8 to 30.93
* Ratio of crime-reduction = -22 % where the 100% is the prompt level of Detroit
* Benefit means the difference between the yellow and the orange cells of the Figure Nr7 (727.7 safety scores where Detroit has a safety index of 27.91 (2791)
* Ratio-benefit means: what kind of benefit can be expected in case of 1% ratio-crime-reduction.

Figure Nr8 has some special interpretations like

* in cases of “burglary” and “larceny theft” the reduction of the crime-level might not bring any safety-increasing based on the simulation model (STD2\_2) and the OAM as such
* Detroit can realize the highest relative benefit-increasing concerning the safety index if the variable “Murder and Nonnegligent Manslaughter” will be reduced (-22%) – see benefit ratio and ist green coloured background

# The case Detroit with cost-aspects

The starting point for this simulation should be the statement, that the city of Detroit is capable of creating an action plan just for one single crime-type. In this case, the action plans (for each variable see before) are the objects for a new OAM. The attributes (Xi) should be the cost positions like

* expenditure for experts deriving the action plan,
* marketing cost for publishing information towards the populations
* cost of new equipment
* cost of maintaining
* cost of licences
* etc.

The last attribute (Y) should be the ratio-benefit value-set where the ratio-benefit is higher then 0. The new OAM should be interpreted by the online robot engine of COCO-Y0.

It could be permitted to create more than just one single action plan for a variable (crime-type). The amount of the attributes (Xi) can be quasi unlimited.

The best action can be derived based on the antidiscriminative calculation. The best action should have the highest estimation value (as far as possible valid after checking function-symmetries).

# Conclusions

The paper presented simple modelling steps after identifying necessary/useful data and appropriate questions/tasks/problems. The modelling steps are similarity-based units like antidiscriminative modelling and/or production function-oriented modelling. The analysts do not need any complex and/or specific models just OAMs. The conclusions (model-interpretations) can be derives in an automated way too. The models/model-chains can be seen as kinds of robot experts being capable of deriving decisions like a human expert or even in a more consistent way…