Understanding Family Dynamics in France: An FCA-Based Study

Omar AbedelKader, Maria Alejandra Cely Latorre
Institut des sciences du Digital, Management Cognition
University of Lorraine
Nancy , France
{omar.abedelkader9, maria-alejandra.cely-latorre6}@etu.univ-lorraine.fr

Prof. Miguel Couceiro
LORIA (CNRS – Inria Nancy Grand Est)
University of Lorraine
Nancy , France
miguel.couceiro@loria.fr

Egor Dudyrev
PhD student in LORIA
University of Lorraine
Nancy , France

I. INTRODUCTION

This report relates to the topic of family dynamics in France. By employing the Formal Concept Analysis (FCA) approach, it is intended to analyze conceptual structures, extract interesting patterns and make connections regarding family structure and composition across all five country zones. Among the dimensions we want to discuss there is cohabitation of members, single-parenthood, willingness to have children, naturalization through marriage and the underlying reasons for these to happen.

This paper is composed by five main sections: Section II. corresponds to the state of the art which discusses the current approaches that other researchers have utilised in the field of FCA. Section III. describes the methodology of our project including details about the problem definition, technical approach and theoretical approach. Section IV. refers to the result analysis per zone and in France in general. Section V. we dedicate this to speak our minds about the project implementation and the results.

II. STATE OF THE ART

A. Formal Concept Analysis

Formal Concept Analysis (FCA) offers a framework for analyzing rules in terms of frequency distribution and confidence, revealing relationships and patterns in datasets. The resulting visual representation, known as the concept lattice, helps identify concepts formed by sets of objects and attributes sharing common characteristics.[2] A concept lattice is conformed by the following elements: intents, which are closed sets of attributes that represent concepts in the lattice; pseudo-intents, sets of attributes that are close to being closed; proper premises, that represent incomplete or partial concepts in the lattice; keys, they are the smallest sets of attributes that can generate a closed set and its number indicates the complexity of the dataset; finally, passkeys, which are the smallest sets of attributes that can generate a closed set and cannot be further reduced. What these elements do is to highlight closed sets, incomplete concepts, and minimal generators and provide insights into dataset complexity and structure. [2] Furthermore, a foundational element in FCA for generating formal concepts is the formal context. It finds extensive applications in representing and understanding complex domains. It organizes objects and attributes in lattices, providing insights into domain structure, exploring all common dependencies and defining various formal concepts.[8]

In our project we can identify the steps to create a formal context as follows:

1) Define the set of objects: the items we want to analyse are a representative sample of the French population, 3 million individuals per region.
2) Define the set of attributes: the characteristics of the objects that we want to analyse are the twelve features we chose for understanding the family dynamics.
3) Define the incidence relation: this specifies which objects have which attributes. For example, that we associate the attribute DIPL_15: highest diploma to each of the ones polled who are above 14 years old.
4) Create a matrix representation: with all the elements above, we can represent the formal context. The rows of the matrix represent the objects, the columns represent the attributes, and the entries indicate whether an object has a particular attribute.

5) Apply Formal Concept Analysis: identify the concepts and relationships within the context, which we are going to describe in below sections.

B. Scaling

Regarding techniques for analysing and visualizing the relationships between objects and attributes within a formal context; Wang et al. [8] remark the status of Conceptual Scaling and Attribute Scaling in the context of FCA. These techniques play a crucial role in understanding the structure and organization of complex data sets. They offer valuable insights into the underlying patterns and associations given that interest measures aid in data exploration, knowledge discovery, and decision-making. On the one hand, Conceptual Scaling organizes and prioritize concepts according to their importance within the given context based on criteria like frequency, generality, association, uniqueness, and discrimination ability. The transformation of many-valued context into a formal context with conceptual scaling is crucial for FCA because it impacts the resulting concept lattice's size, interpretation, and computation. For instance, the choice of a scale should align with data and goals, highlighting its significance in FCA processes.[2] On the other hand, Attribute Scaling focuses on representing various attribute types within the concept lattice to understand the relevance of attributes in relation to each other. These can be of different types: nominal, ordinal or numeric. Nominal attributes have no inherent order or ranking. The are treated as binary attributes, where each attribute is either present or absent for a given object they describe. Other are ordinal attributes which are scaled to preserve order capturing the relationships between attributes and objects. Finally, numeric attributes are quantitative variables that represent measurable quantities. These can be scaled and discretized to fit within the concept lattice for instance by categorizing the numeric values into intervals. This representation allows FCA to uncover underlying relationships and dependencies in datasets, providing a versatile framework for understanding complex data structures and extracting meaningful insights. [2]

Other indicators that provide insights in assessing the complexity of a dataset and that can be used to identify hidden patterns and relationships within the lattice to make more informed decisions are introduced by Buzmakov et al. [2]. In their work they used two alternative indicators called the Linearity Index and Nonlinear Distributivity Index. The first one measures how closely the concept lattice resembles a chain by calculating the probability of comparability between random concepts. The second one measures the probability that the union of two intents forms another intent, indicating the lattice's proximity to a distributive lattice.

C. Encoding

In FCA, concepts can be represented through their intents, extents, or a combination of both, using two types of encoding full or narrow. Full encoding involves using all attributes or objects to represent a concept, whereas narrow encoding selectively includes only those attributes or objects present in the concept's intent or extent, excluding those in lower or higher concepts according to the partial order $\leq$ and therefore such attributes or objects occur only once in the lattice. Marquer's[10] thesis used two numerical representations: a matrix based on full encoding, where entries indicate the presence of attributes or objects in the intent or extent of a concept, and a vector-based on narrow encoding, where each entry corresponds to an attribute or object, and its value represents the index in the ordered list where the attribute or object appears.

D. Formal Concept Analysis Algorithms

The 2002 study conducted by Kuznetsov and Obiedkov, shows a comprehensive comparative analysis of ten concept-generating algorithms, including Bordat, NextClosure, Close by One, Lindig, Chein, Nourine, Norris, Godin, Dowling, and Titanic. These are categorized into two main types based on their approach to concept generation: Incremental Algorithms, exemplified by Godin et al. (1995)[7], generate the concept set or graph for the first i objects at each *ith* step. Conversely, Batch Algorithms build these for the entire context from scratch. They employ either a top-down approach (from the maximal extent to the minimal one), or a bottom-up approach (from the minimal extent to the maximal one) considering attributes instead of objects and vice versa. [9] One latent challenge of these algorithms is the repetition of identical concepts. Different strategies are used to tackle this, including organizing concepts into disjoint sets (e.g., Chein algorithm), using hash functions for distribution and search reduction (e.g., Godin algorithm with intent cardinality as hash). Algorithms like NextClosure and ClosebyOne adopt the lexicographical order approach, ensuring recognised concept generation without the need for checking duplicates or relying on previously generated concepts.[9]

E. Formal Concept Analsis Usage and Applications

Hazarika & Sinha [9] performed a review work aiming at categorizing the research, usage and trends of FCA. They primarily pointed out two fields in which FCA has a wide scope: Software Engineering and Data Mining. In the first realm, concepts are formed from software source code artifacts which are organized in a concept lattice, facilitating navigation, visualization, and the generation of additional concepts. FCA particularly operates for the maintenance of large amounts of documentation and formal methods during initial software development phases, enhancing the readability of formal specifications. In the second, FCA has mainly been applied as an analytical tool in Knowledge Discovery in Databases (KDD). Given the flow in online data and database utilization there is the need to extract meaningful knowledge like frequent patterns, particularly in association rule mining from vast datasets and web discovery tools. Further applications of this, cover areas like linguistics and ontology engineering in which FCA contributes to disambiguating natural language, organizing lexical information, capturing generalizations, building multilingual databases, and facilitating knowledge reuse in Expert Systems.

F. Implementation Tools

Uta Priss Github page [11] contains a collection of open-source software applications, demos and other tools for FCA analysis and visualization. From this list, we want to highlight that packages and libraries allow programmatic control over the analysis process. This can be particularly beneficial if you have specific requirements or want to integrate FCA into a larger workflow. Packages often allow for more customization, and their specific needs or requirements are better managed with code rather than most software. To exemplify, there are the packages called GALACTIC [1] and fcaR [3], guiding their approach to analyse and extract meaningful patterns from complex and diverse datasets. GALACTIC introduces plugins, providing a modular and extensible architecture. In contrast, fcaR utilizes classes and methods to achieve similar organizational goals. Notably, both address additional complexities, for instance, GALACTIC broadens FCA through the incorporation of the "Next Priority Concept" algorithm that facilitates versatile analyses. On the other hand, fcaR extends FCA to handle ambiguous datasets, demonstrating adaptability and incorporating a simplification logic. Additionally, GALACTIC is implemented in Python and introduces a modular system featuring plugins for characteristics, descriptions, strategies, and meta-strategies, whereas fcaR is an R package that proposes specific classes such as "Set," "Concept," "ConceptLattice," and "ImplicationSet" to represent formal contexts, concepts, concept lattices, and implications.

III. METHODOLGY

A. Problem Definition

In this project we want to examine the various dimensions in which households in France are constructed. It needs to be explored per zone, who cohabit to identify patterns in which French families can be framed and the underlying reasons for it to happen. For instance, single-parent household: a parent and a child, a traditional household: a couple with children, couples with no children, non-traditional households that have presence of non-family members in the place of residence, other households: if a blended or extended family structure prevails and why there is a mix of nationalities in the household, referring to French citizens married foreigners.

We want to focus on the following aspects:

- Family structure: Identify the different types of family structures with the kind and number of members that cohabit.
- Single parenthood: Identify the characteristics of single-parent households.
- Marriages with(out) children: Identify the reasons for having or not having children in some couples.
- Naturalization: Identify the rate and possible reasons for naturalization in France. People who have acquired French nationality because of marriage to a French national.

In order to make the correct assumptions regarding how households in France are constructed, we took the following variables into consideration:

1) 'STAT_CONJ': Marital Status: either married or not married.
2) 'APAF': Belonging to a household in the form of male/female adult from principal or secondary household, child from main or secondary household or a person not belonging to a household.
3) 'LPRM': Relationship to household reference. It could be the same person, his/her partner, his/her own or couple's child, an infant, another parent, friend, ancestor, subtenant.
4) 'INFAM': Number of families in the household: no family, one or two families.
5) 'COUPLE': Life as a couple, whether someone lives as a couple or not.
6) 'NPERR': Number of people in the household: one to six people or more.
7) 'NAT49': Current nationality detailed in 49 positions. For example, born in France, French by acquisition, or the most representative nationalities at a national level.
8) 'NATN12': Nationality at birth detailed in 12 positions. For example, born in France, born in the UE, other European nationalities, foreigners, other nationalities.
9) 'IMMI': Immigrant status: immigrant, non-immigrant.
10) 'INAT': Nationality indicator: French by birth, French by acquisition, foreigner.
11) 'DIPL_15': Highest diploma: no diploma, high school diploma, professional certificate, higher education.
12) 'STAT': Professional status: employed, independent, employer or family helper.

Remark: we excluded common variables like age and gender, since we did not consider them as relevant for our initial hypotheses.

B. Technical Approach

We conducted our experiments on the *PNY NVIDIA A100*, on a private server. We used FCApy [5] Python package to work with FCA [5], and two other data mining Python packages using FCA framework developed by Egor Dudyrev called [6] (Pattern-Structures-pailleur) which mines dependencies in complex data and [4](Characteristc-Attribute-Sets-pailleur) which mines dependencies in binary data. Given our limited RAM resources, we were not be able to create a lattice with the full dataset. Therefore, we randomly "random_state = 42" extracted 3 million rows per zone from the original dataset to create our new one in which all the information is based on.

1) Dataset: The original data called "Housing, individuals, activity, educational and occupational mobility, residential migration in 2016" is a population census obtained from The National Institute of Statistics and Economic Studies (INSEE) which is in charge of collecting, analysing and disseminating information on the French economy and society. The data describe the characteristics of each person enumerated, those of his or her household and principal residence. This set contains detailed modalities, including restricted variables like nationality, country of birth, length of time in France. An administrative division of France is shown in Fig. 1.

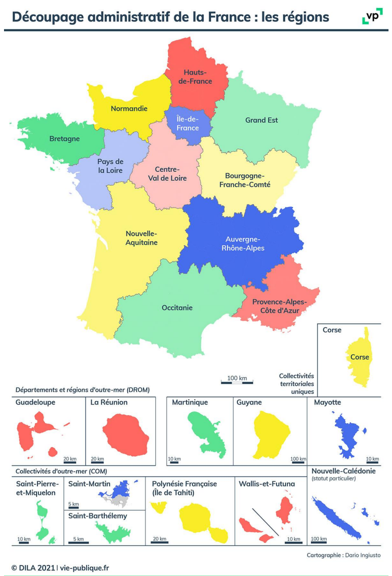

Fig. 1. Administrative division of France: the regions]Administrative division of France: the regions[1]

French regions are divided in the following zones:

- Zone A : Region Île-de-France (région 11)
- Zone B : Regions Centre-Val de Loire (région 24), Bourgogne-Franche-Comté (région 27), Normandie (région 28) and Hauts-de-France (région 32)
- Zone C : Regions Grand Est (région 44), Pays de la Loire (région 52) and Bretagne (région 53)
- Zone D : Regions Nouvelle-Aquitaine (région 75) and Occitanie (région 76)
- Zone E : Regions Auvergne-Rhônes-Alpes (région 84), Provence-Alpes-Côte d'Azur (région 93), Corse (région 94), Guadeloupe (région 01), Martinique (région 02), Guyane (région 03) and La Réunion (région 04).

The data can be found in dBase or csv formats. Each zone contains 97 columns corresponding to the attributes and different numbers of rows or objects. In total there are 19 426 531 rows.

- Zone A: 4 315 657 rows
- Zone B: 3 898 453 rows
- Zone C: 3 341 347 rows
- Zone D: 3 116 443 rows
- Zone E: 4 754 631 rows

Feature	Value	Encoding
STAT_CONJ	A	0
	B	1
APAF	Z	999
LRPM	Z	999
INFAM	Z	999
NPERR	Z	999
NAT49	AFR	101
	AME	102
	AOC	103
	CAR	104
	EUR	105
	HUE	106
NATN12	ZZ	999
DIPL_15	A	1
	B	2
	C	3
	D	4
	Z	999
STAT	ZZ	999

TABLE I
FEATURES ENCODING

2) Features Encoding: This step is crucial to handle different data types appropriately and to analyze the mining patterns. First, we need to transform those values into intervals and group infrequent categories. Then, the preprocessing steps require feature encoding to ensure the data is in the correct format. Table I Encoded values of our chosen variables.

C. Theoretical Approach

We conducted an experiment that is composed of 3 main parts: creating the pattern structures, visualising data and mining patterns, and analyzing and visualising stable patterns.

1) Creating Pattern Structure: Pattern Structures are the extension of the classical binary FCA that allow it to work with many complex data .

We can define our formal context as:

$$(G, M, I) \tag{1}$$

where the G represents the objects, M presents the binary attributes and the I connection between G & M. In the many-valued context we can define the following formula of the formal context as:

$$(G, (D, \sqcap), \delta) \tag{2}$$

where G represents the set of objects, D represents the space of descriptions and $\sqcap : DxD \rightarrow D$ represent the operation that outputs the maximal description smaller than the two given descriptions, and $\delta : G \rightarrow D$ maps every object to its description.

We started by creating the pattern structure that determines the datatype of each column and assign appropriate pattern structures ('SuperSetPS' for categorical and 'IntervalPS' for numerical data), then combining these into a single Cartesian pattern structure. Finally converting numeric values to intervals.

2) Visualizing data and Mining Patterns: The binary attributes are generated and iterated over to show how to handle a large number of attributes. After having a normalized data sample, the frequency of values in the categorical columns is analyzed. Finally, using the caspailleur package it is iterated over binary attributes to find the stable extents. The algorithm we used for extracting top best patterns for itemset mining is called Sofia Algorithm (Searching for Optimal Formal Intents Algorithm).

3) Analyzing and Visualizing Stable Patterns : Our table intents are calculated from the extents that are created in the previous part and computed the statistical measures. First, we used Support, a measure attached to an association rule, which shows how often one can find the premise in the dataset. Here, an association rule between an itemset X and an itemset Y is denoted by $X \rightarrow Y$, where X is called the premise and Y is called the conclusion of the rule [2]. Moreover, we used the associated measure of concept relevancy called Delta-measure which is an estimate of stability that can be computed in polynomial time, thus being a criteria for pattern selection in a large dataset [2]. We chose this also because it can be used in combination with Sofia Algorithm.

Therefore, Interestingness can be considered “the product of support and delta measure" then the top patterns are identified. Actually, Interestingness is a qualitative measure of how unusual or meaningful a pattern or concept is within a dataset. [2]

Afterwards, a partially ordered set “Poset” of these patterns is created and visualized as our lattice to show the relationships and hierarchy among the different stable concepts.

IV. RESULT ANALYSIS

A. Zone A

This zone consists only of the Île-de-France region.

1) Family structure and cohabitation: The predominant family structure is typically composed by two adults who are likely to be married This is a normalized behaviour by the culture and society, which establish that family formation starts with marriage. Moreover, there were mostly identified households with only one family that contains from two to five members. For instance, a house where a married couple live and consist of 5 members, we could infer that this represents parents and at most 3 children. Finally, there seems no to be significant differences in family composition based on educational level or professional status.

2) Single-parent households: The implications associated to a household reference person indicates that single-parents live obviously with their children and rare but possibly they may also cohabit with other relatives or friends. We can justify the presence of people out of the main family as a support for the parent or a need for shared resources. Finally, no special relationship was found regarding a person’s origin, highest diploma or professional status, indicating that single-parent households can occur in various demographic groups.

3) Having children or not: First of all, the population that is not married points to a higher value than the ones who are married. Therefore, an initial implication for not having children is mainly being single. For obvious reasons, married couples are more likely to have children than unmarried couples, this suggests that marriage plays a role in this decision. Another influential rule is that couples with at least one person working as a salaried employee and with higher diploma education are more liable to have children since they could have a more stable financial position and have the resources to upbring a child.

4) Naturalization in France: The largest number of people who acquire french nationality through marriage is found in this Zone. We discovered that Portuguese people are prone to acquire French nationality by marrying a French national. This may be due to factors such as shared cultural backgrounds or the ease of integration within EU countries. However, we also observed that certain non-European nationalities such as Moroccans, Turks, and Algerians are also naturalized by marriage. This may be linked to historical and cultural ties between France and these countries, or due to the presence of large communities within the region. Lastly, common professional or educational backgrounds among those who naturalize through marriage indicate that they may have more opportunities for social and economic mobility within France.

B. Zone B

This zone comprises the regions of Centre-Val de Loire, Bourgogne-Franche-Comté, Normandie and Hauts-de-France.

1) Family structure and cohabitation: The family structure in these regions is highly varied. The predominant pattern is one person living in the household. There are also other rules that relate to the typical married parents living with their children. Less traditionally, this tendency repeats even if the parents are not married. We noticed that there could be up to four members living together in the household, inferring that this corresponds to parents and at most two children. Regarding the cohabitation with members belonging to a secondary family, this seems not to be predominant in these regions, meaning that it is not common to have two generations of a family or blends with other families living under the same roof.

2) Single-parent households: The implication rules revealed that the characteristics of single-parent households correspond to individuals who have French nationality by birth and possess at least a higher education diploma. In this home, there are no other family members that belong to an extended or blended family.

3) Having children or not: Once again, there seems to be a clear tendency of constructing a nuclear family towards the notion of marriage and it could increase the likelihood of having children. Even though the desire for having or not having children is not explicitly linked to levels of reached education and employment status, we observed that salaried employees are more prone to procreate than self-employed people.

4) Naturalization in France: The lowest number of people who acquired french nationality through marriage is found in this Zone. A pattern to highlight is that most of the naturalized population from this region are salaried employees and own at least a professional certificate, which may suggest financial stability and a desire for a more permanent status. Furthermore, this can occur in a diverse range of nationalities, with North African countries featuring prominently, possibly indicating the desire for cultural integration and assimilation into French society.

C. Zone C

This zone comprises the regions Grand Est, Pays de la Loire and Bretagne.

1) Family structure and cohabitation: In these regions, family units can range from solitary individuals to larger family constructions, ranging from one to four people. In this span, we can find the typical family structure composed by a married adult male member of the main family with his spouse, and their children. However, a very different and new pattern was also found, in which a single individual is living with other non-family members as a non-couple household, this could indicate shared accommodations.

2) Single-parent households: In this area, the pattern of a single person who has children and lives with more than two people is evidence that relatives or even non-relatives might be living with the single parent for reasons that could correlate with factors such as financial support and childcare assistance.

3) Having children or not: The rules we found indicate that married couples, living as a couple, just one family living in the household consisting of 2 to 4 people with French nationality. They can have wither a degree or not but still are likely to have children. This reveals there are no specific relationships between education and the willingness of having children.

4) Naturalization in France: Acquiring French nationality not by birth but through marriage to a French citizen is quite common among the respondents of these regions. However, no particular reason seems to justify this finding. On the one hand, it was observed that people from all sorts of employment statuses and educational background are represented, suggesting that these aspects may not be significant determining factors for naturalization through marriage.

D. Zone D

This zone comprises the regions Nouvelle-Aquitaine and Occitanie.

1) Family structure and cohabitation: From the association rules we obtained from the lattice we can identify that a family structure in these regions of France is diverse. There are people who live alone, there are couples with children and possibly multi-generational households. The household range between one up to five members, suggesting that there are extended families living together.

2) Single-parent households: Overall, single-parent households are headed by a female adult member of the main family, with either one or two children and they do not share the household with any other member. Reasons for single-parent households could include divorce, separation, or being a single parent by choice. There are varied professional status and educational level, meaning that being a single parent is not directly linked to these.

3) Having children or not: The professional status and educational background influences couples' decisions about children. For instance, people reaching higher levels of education may delay starting a family to focus on their career and personal goals. Conversely, salaried employees and self-employed individuals could consider having children perhaps due to their income and ability to provide for a family, they might want to expand their family and provide siblings for their existing children.

4) Naturalization in France: It seems that there is a varied range of educational backgrounds and professional statuses among those naturalizing. This suggests that education and employment status are not decisive factors in naturalization through marriage. However, we can assume that immigrants who are married to a French national may also be more likely to seek naturalization to stay in France permanently and avoid the uncertainty of a temporary visa. On the other hand, people with higher levels of education may be more likely to seek naturalization as they may have a better understanding of the benefits and requirements of French citizenship.

E. Zone E

This zone comprises the regions Auvergne-Rhônes-Alpes, Provence-Alpes-Côte d'Azur, Corse, Guadeloupe, Martinique, Guyane, La Réunion.

1) Family structure and cohabitation: From the association rules we obtained from the lattice we can identify that a family structure in these regions of France is again diverse. There are people living alone, as couples with children and possibly extended families with ascendants and other relatives. The household size consists of one up to four members.

2) Single-parent households: The patterns found suggest that single-parent households of these regions are conformed by the parent and the child and they cohabit with other members of the family. Since this parent is likely to be employed, he/she can profit of the company of external member of the main family to take care of the children.

3) Having children or not: For obvious reasons, not being married reduces the chances of having children and this factor is quite latent in this region. Nevertheless, there seem to be no specific connection between educational background and type of employment with the inclination of having children.

4) Naturalization in France: This rule reveals that naturalization through marriage is common in specific nationalities like Algerian, Moroccan, Tunisian, Portuguese and Italian. A possible explanation for naturalization is the geographical proximity of those countries. It seems that the variety of educational backgrounds and professional statuses among those naturalizing is not a decisive factor for this choice.

F. Overview of families in France

1) Family structure and cohabitation: In analyzing the family structure and cohabitation patterns across France, we found that all zones match a family structure consisting of two adults, often married, and with a child, reflecting cultural norms that associate family formation with marriage. One-family households raging from two to five members are also identified, without significant variations based on educational or professional status mainly in zones D and E. Besides, zones C, D and E both exhibit diverse family structures, with individuals living alone, and potentially multi-generational households. Notably, a novel pattern emerges with single individuals living with non-family members, suggesting shared accommodations in zone C.

2) Single-parent households: Since there are varied professional statuses and educational levels present in France, it was observed that being a single parent is not directly linked to these factors in all zones except for Zone B, where the characteristics of single-parent households correspond to mostly individuals who have French nationality by birth and possess at least a higher education diploma. Conversely, patterns in Zones C, D, and E suggest that this type of household might cohabit with other members of the family. Among the reasons for this, they serve as support systems, whether for financial assistance, childcare, or shared resources.

3) Having children or not: In general, there is a clear tendency of constructing a nuclear family towards the notion of marriage that increases the likelihood of having children. However, only 33.7 percent of the people in our sample have this marital status. Conversely, zone B suggests that couples with at least one person working as a salaried employee and with higher diploma education are more likely to have children due to their stable financial position, whereas in zone C concerning the employment status, we observed that salaried employees are more prone to procreate than self-employed people.

4) Naturalization in France: Around 5.4 percent of the sampled population across zones are considered French by acquisition. Among the common reasons for this is geographical proximity, historical links, cultural connections, or the presence of such communities in France. Some representative nationalities are Portugal and Italian. Similarly, non-European nationalities, such as Moroccans, Turks, and Algerians. Moreover, for regions A and C professional and educational backgrounds appear to influence them whereas Zones B, D, E reveal a varied range of educational backgrounds and professional statuses, indicating that these factors are not decisive in naturalization through marriage.

G. Support and Delta Measurement

Appendix **??** shows scatter plots for each zone, these plots show that there is a higher density of data points towards the lower end of the "support" range and the lower end of the "delta_measure" range, as "support" increases, the number of data points and the " delta_measure" values seem to decrease. Moreover, a negative correlation is observed meaning a general trend where the points are more concentrated at the lower right and upper left corners of the plot. The scatter plot shows a dense clustering of concept towards the lower end of the support spectrum, with a high variance in the "delta_measure" meaning the concepts with a lower support tend to have a wider range of "delta_measure" values indicating that some infrequent patterns are considered to be quite interesting as they have a high delta_measure values despite their infrequency. The vertical line at "support = 1000" suggests a cutoff point. The concepts to be right are considered interesting based on their support level being above this threshold. However, it's important to mention that some concepts with a lower support still have high delta_measure values indicating that they might be considered interesting for reasons other than their frequency of occurrence.

V. Discussion and Limitations

Based on the experience carrying out this project we want to discuss some technical and methodological limitations and experiences encountered during the elaboration process.

First, the suggested implementation for extracting mining patterns using FCA exhibits several limitations. The code is in its preliminary stages and contains several rudimentary elements that need further development, refinement, optimization and streamlining. These issues impact the performance, accuracy, usability of the code. We would suggest exhibits substantial resource consumption, with operational zone requiring approximately 100GB of RAM. Due to the unavailability of slots on the GRID5K platform, the code was executed on a private server. This level of hardware usage poses a significant challenge, especially in environments where such resources are limited or where efficiency is a critical factor. Furthermore, the allocation and utilization of system resources, including CPU and GPU, indicate a need for optimization. Addressing these aspects is crucial for enhancing the performance and scalability of the code. Future iterations would benefit from a focused effort on optimizing resource partitioning and management.

On the other hand, while analyzing the association rules we were expecting to obtain more varied, precise and revealing information that could lead to more interesting results. Therefore, we think the dissatisfaction with the results is due to different possible reasons. First, some initial hypothesis that we might have set are deviated from the scope of the original dataset. This means mean that there is no sufficient information suitable for our expectations. For instance, in the case of marital status, it is hard to determine other marital status other than married and not married excluding values such as divorced, widowed etc that might help to explain reasons for single-parenthood. Secondly, the algorithm implemented does not extract the precise patterns that aligns to our initial hypotheses.

VI. CONCLUSIONS

The FCA analysis of family dynamics in France seems to extract varied relations among the chosen variables that directly influences the reasons of cohabitation patterns and naturalization trends. For example, the influence of nationality, education, and employment status on family dynamics differs between zones, highlighting the diversity of experiences within the country. There is a high variance in delta measure values for concepts with lower support meaning that some rare patterns are considered interesting due to their large impact.

References

[1] Karell Bertet and Christophe Demko. GALACTIC : une nouvelle approche d'analyse de données complexes et hétérogènes, June 2023. Communication orale aux Journées du GT Explicon en collaboration avec le séminaire "Interdisciplinary Seminar Algorithms and Society" (LAMSADE-IRISSO).

[2] Alexey Buzmakov, Sergei O Kuznetsov, Tatiana Makhalova, Amedeo Napoli, and Egor Dudyrev. Data complexity: An FCA-based approach. working paper or preprint, January 2023.

[3] Pablo Cordero, Manuel Enciso, Domingo López-Rodríguez, and Ángel Mora. fcar, formal concept analysis with r. *R J.*, 14:341–361, 2022.

[4] Egor Dudyrev. Caspailleur. Caspailleur.

[5] Egor Dudyrev. Fcapy. FCApy.

[6] Egor Dudyrev. Paspailleur. Paspailleur.

[7] Robert Godin, Guy W. Mineau, Rokia Missaoui, Marc St-Germain, and Najib Faraj. Applying concept formation methods to software reuse. *Int. J. Softw. Eng. Knowl. Eng.*, 5:119–142, 1995.

[8] Longchun Wang Lankun Guo and Qingguo Li. Continuous domains in formal concept analysis. *arXiv preprint arXiv:1912.04496*, 2019.

[9] Shyamanta Hazarika and Smriti Sinha. Formal concept analysis: current trends and directions. *Artificial Intelligence Review*, 44:1–40, 06 2013.

[10] Esteban Marquer. LatticeNN – Deep Learning and Formal Concept Analysis. Master's thesis, Université de Lorraine, September 2020.

[11] Uta Priss. Fca software. FCA Software.

APPENDIX - I - SUPPORT AND DELTA MEASURMENT

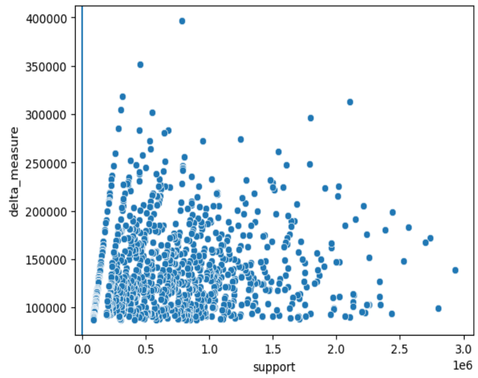

Fig. 2. Zone A: Support & Delta Measurement

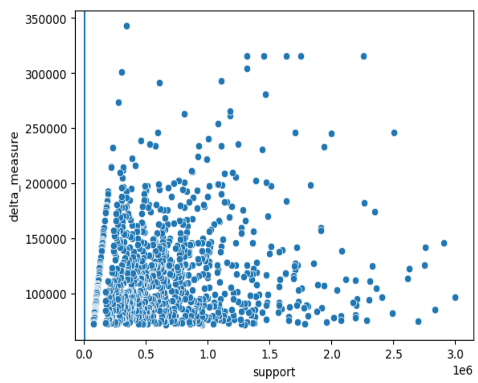

Fig. 3. Zone B: Support & Delta Measurement

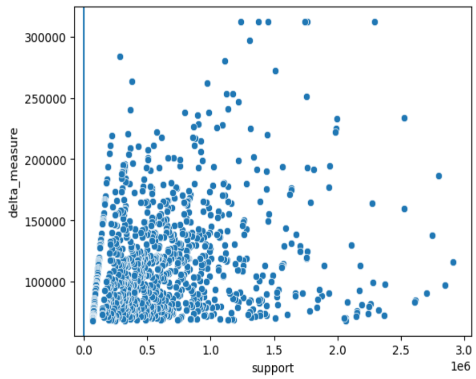

Fig. 4. Zone C: Support & Delta Measurement

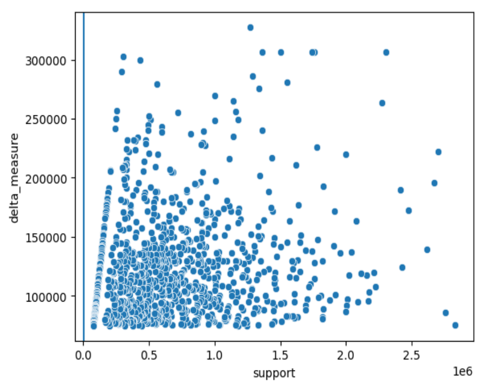

Fig. 5. Zone D: Support & Delta Measurement

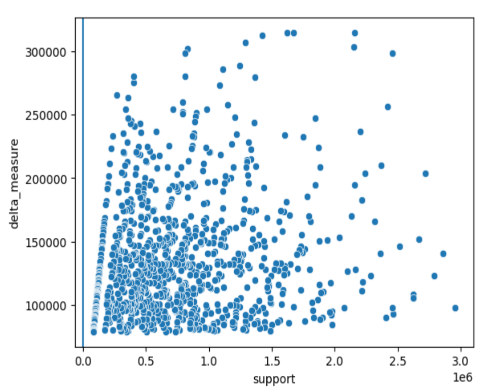

Fig. 6. Zone E: Support & Delta Measurement

APPENDIX - II- LATTICE PRESENTATION

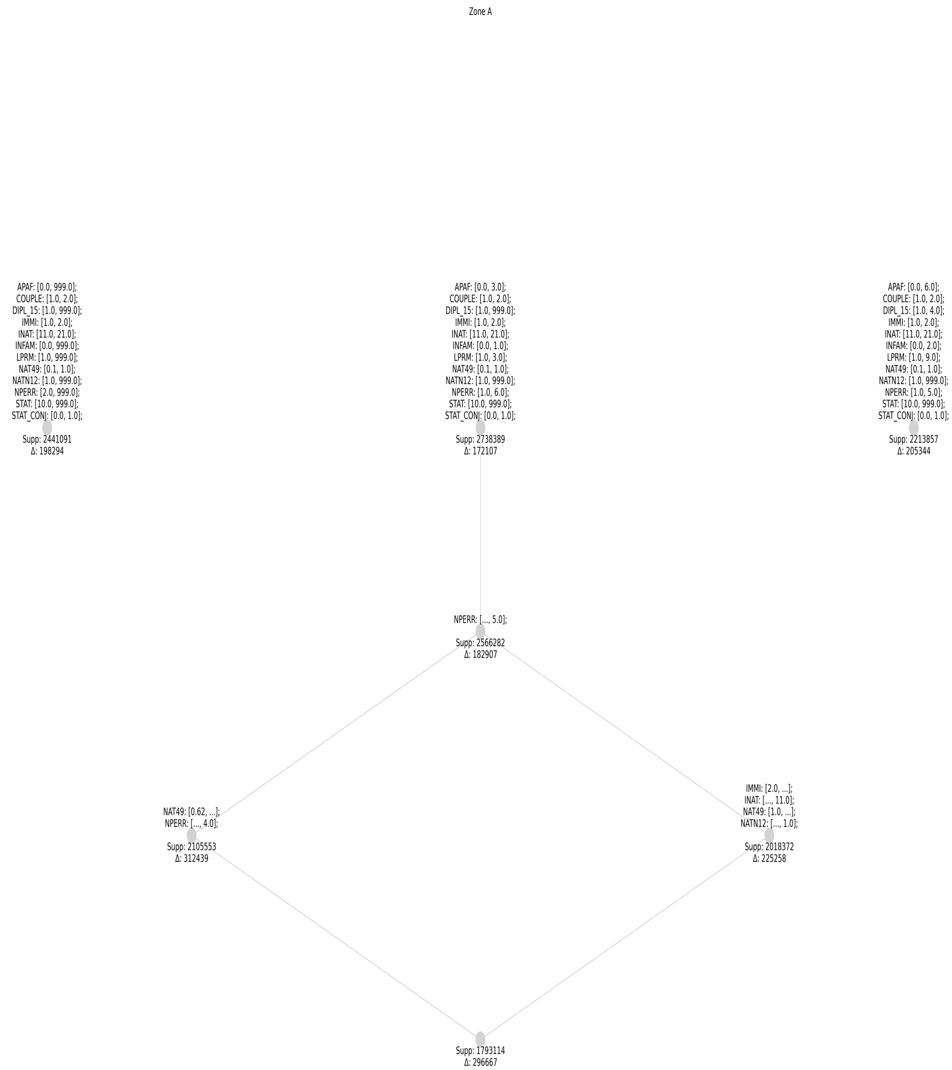

Fig. 7. Zone A

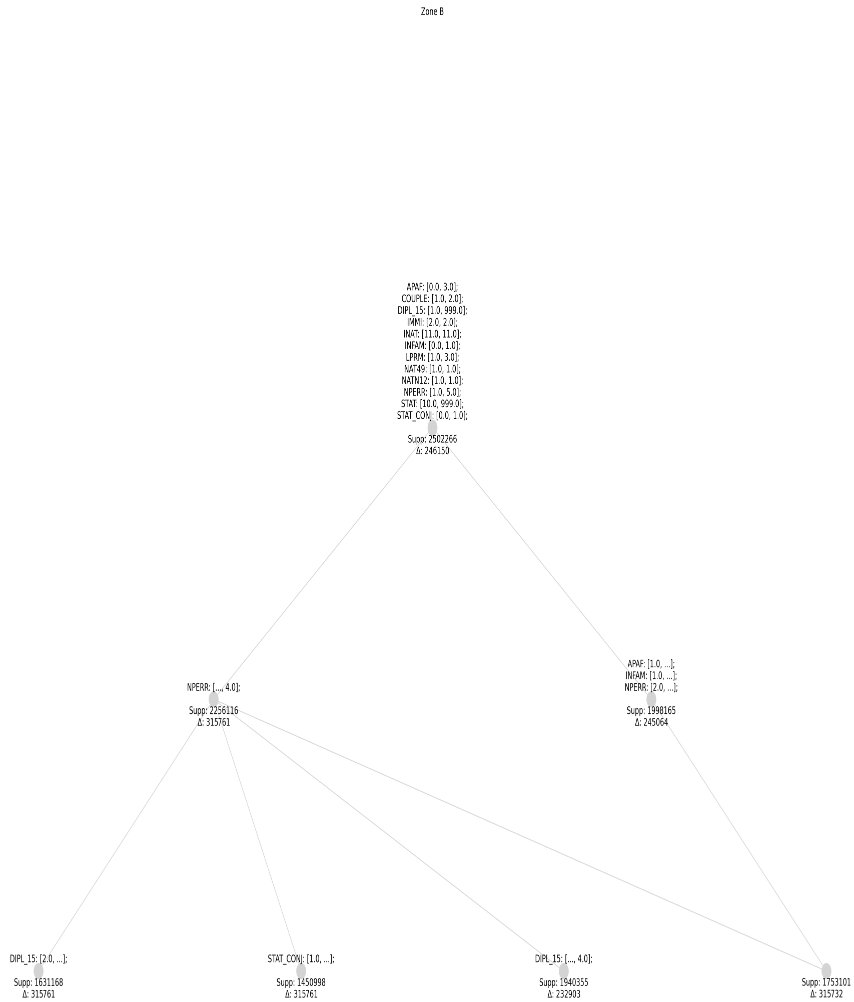
Zone B
APAF: [0.0, 3.0];
COUPLE: [1.0, 2.0];
DIPL_15: [1.0, 999.0];
IMMI: [2.0, 2.0];
INAT: [11.0, 11.0];
INFAM: [0.0, 1.0];
LPRM: [1.0, 3.0];
NAT49: [1.0, 1.0];
NATN12: [1.0, 1.0];
NPERR: [1.0, 5.0];
STAT: [10.0, 999.0];
STAT_CONJ: [0.0, 1.0];
Supp: 2502266
Δ: 246150
NPERR: [..., 4.0];
Supp: 2256116
Δ: 315761
APAF: [1.0, ...];
INFAM: [1.0, ...];
NPERR: [2.0, ...];
Supp: 1998165
Δ: 245064
DIPL_15: [2.0, ...];
Supp: 1631168
Δ: 315761
STAT_CONJ: [1.0, ...];
Supp: 1450998
Δ: 315761
DIPL_15: [..., 4.0];
Supp: 1940355
Δ: 232903
Supp: 1753101
Δ: 315732

Fig. 8. Zone B

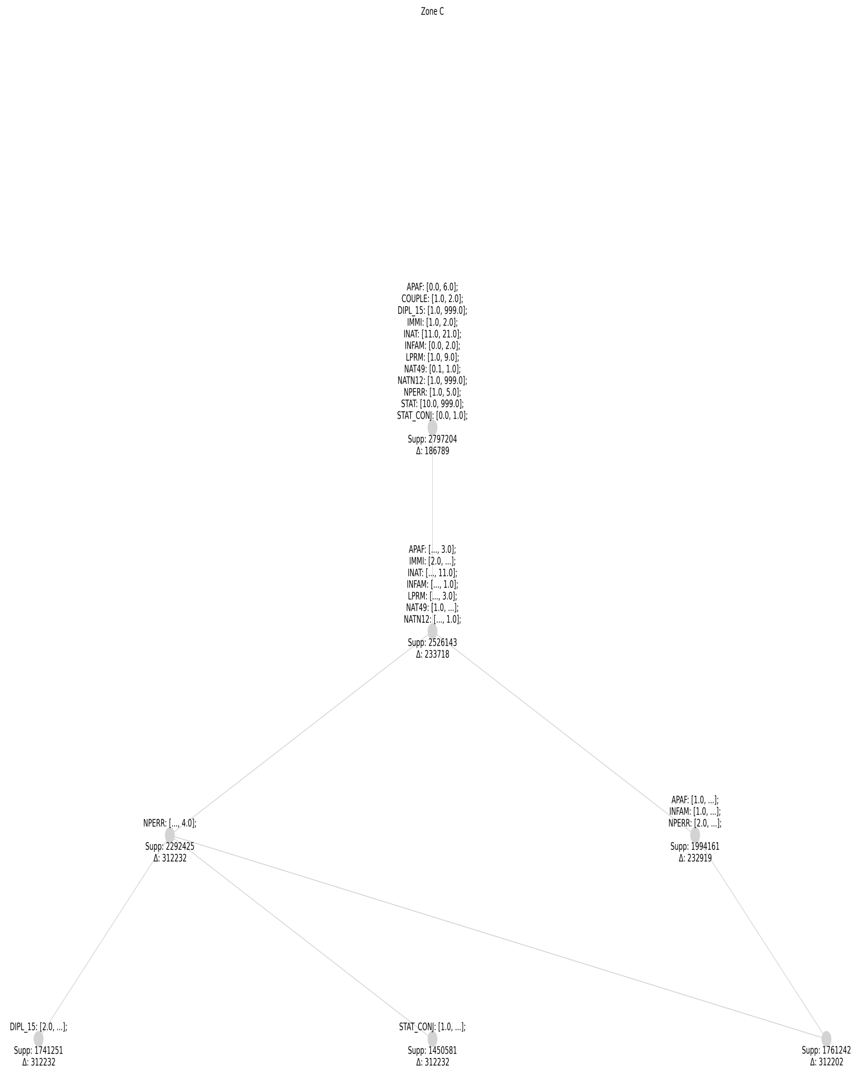
Zone C
APAF: [0.0, 6.0];
COUPLE: [1.0, 2.0];
DIPL_15: [1.0, 999.0];
IMMI: [1.0, 2.0];
INAT: [11.0, 21.0];
INFAM: [0.0, 2.0];
LPRM: [1.0, 9.0];
NAT49: [0.1, 1.0];
NATN12: [1.0, 999.0];
NPERR: [1.0, 5.0];
STAT: [10.0, 999.0];
STAT_CONJ: [0.0, 1.0];
Supp: 2797204
Δ: 186789
APAF: [..., 3.0];
IMMI: [2.0, ...];
INAT: [..., 11.0];
INFAM: [..., 1.0];
LPRM: [..., 3.0];
NAT49: [1.0, ...];
NATN12: [..., 1.0];
Supp: 2526143
Δ: 233718
NPERR: [..., 4.0];
Supp: 2292425
Δ: 312232
APAF: [1.0, ...];
INFAM: [1.0, ...];
NPERR: [2.0, ...];
Supp: 1994161
Δ: 232919
DIPL_15: [2.0, ...];
Supp: 1741251
Δ: 312232
STAT_CONJ: [1.0, ...];
Supp: 1450581
Δ: 312232
Supp: 1761242
Δ: 312202

Fig. 9. Zone C

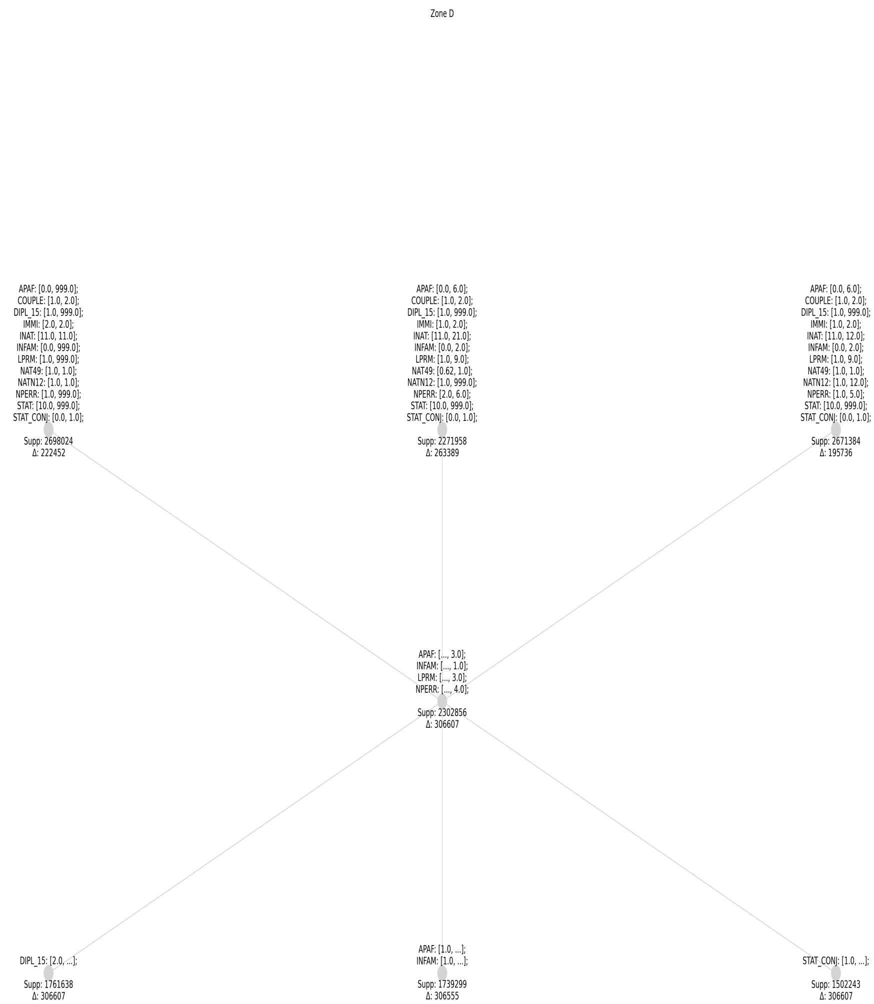
Zone D
APAF: [0.0, 999.0];
COUPLE: [1.0, 2.0];
DIPL_15: [1.0, 999.0];
IMMI: [2.0, 2.0];
INAT: [11.0, 11.0];
INFAM: [0.0, 999.0];
LPRM: [1.0, 999.0];
NAT49: [1.0, 1.0];
NATN12: [1.0, 1.0];
NPERR: [1.0, 999.0];
STAT: [10.0, 999.0];
STAT_CONJ: [0.0, 1.0];
Supp: 2698024
Δ: 222452
APAF: [0.0, 6.0];
COUPLE: [1.0, 2.0];
DIPL_15: [1.0, 999.0];
IMMI: [1.0, 2.0];
INAT: [11.0, 21.0];
INFAM: [0.0, 2.0];
LPRM: [1.0, 9.0];
NAT49: [0.62, 1.0];
NATN12: [1.0, 999.0];
NPERR: [2.0, 6.0];
STAT: [10.0, 999.0];
STAT_CONJ: [0.0, 1.0];
Supp: 2271958
Δ: 263389
APAF: [0.0, 6.0];
COUPLE: [1.0, 2.0];
DIPL_15: [1.0, 999.0];
IMMI: [1.0, 2.0];
INAT: [11.0, 12.0];
INFAM: [0.0, 2.0];
LPRM: [1.0, 9.0];
NAT49: [1.0, 1.0];
NATN12: [1.0, 12.0];
NPERR: [1.0, 5.0];
STAT: [10.0, 999.0];
STAT_CONJ: [0.0, 1.0];
Supp: 2671384
Δ: 195736
APAF: [..., 3.0];
INFAM: [..., 1.0];
LPRM: [..., 3.0];
NPERR: [..., 4.0];
Supp: 2302856
Δ: 306607
DIPL_15: [2.0, ...];
Supp: 1761638
Δ: 306607
APAF: [1.0, ...];
INFAM: [1.0, ...];
Supp: 1739299
Δ: 306555
STAT_CONJ: [1.0, ...];
Supp: 1502243
Δ: 306607

Fig. 10. Zone D

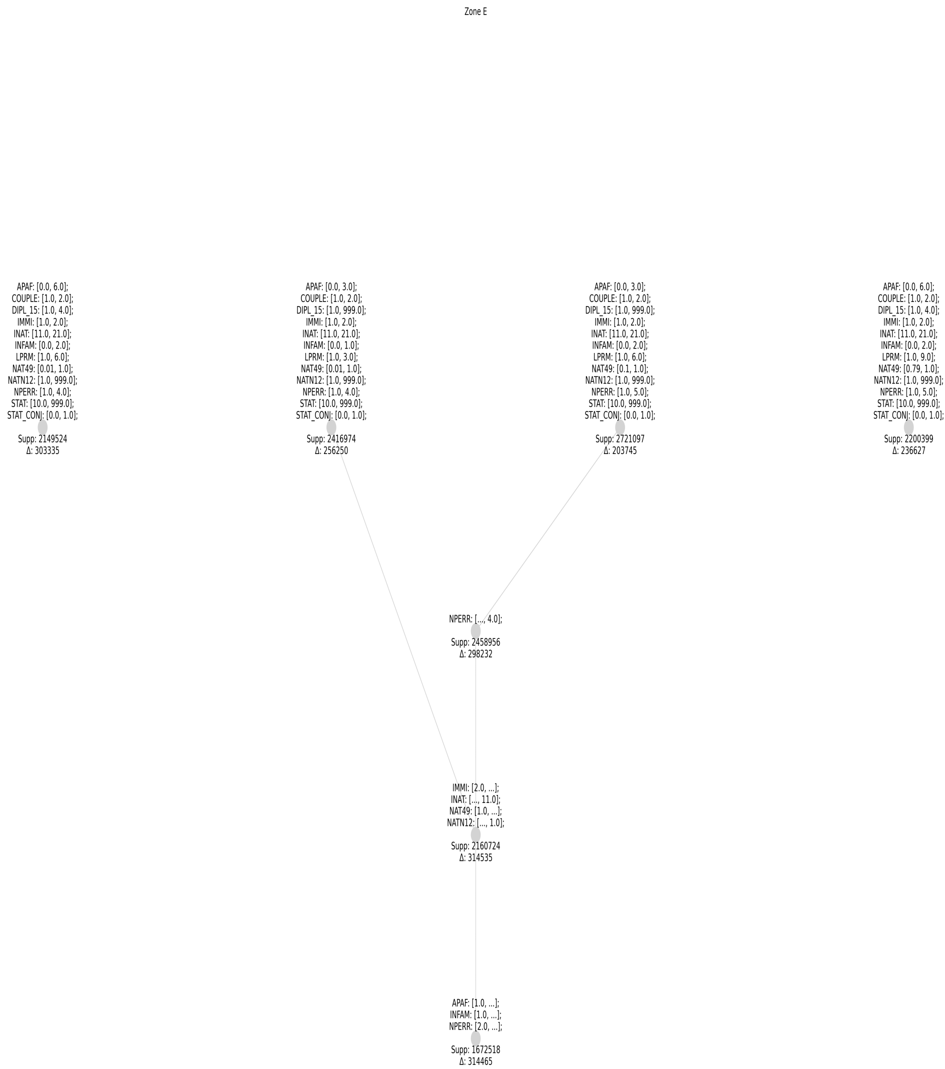
Zone E
APAF: [0.0, 6.0];
COUPLE: [1.0, 2.0];
DIPL_15: [1.0, 4.0];
IMMI: [1.0, 2.0];
INAT: [11.0, 21.0];
INFAM: [0.0, 2.0];
LPRM: [1.0, 6.0];
NAT49: [0.01, 1.0];
NATN12: [1.0, 999.0];
NPERR: [1.0, 4.0];
STAT: [10.0, 999.0];
STAT_CONJ: [0.0, 1.0];
Supp: 2149524
Δ: 303335
APAF: [0.0, 3.0];
COUPLE: [1.0, 2.0];
DIPL_15: [1.0, 999.0];
IMMI: [1.0, 2.0];
INAT: [11.0, 21.0];
INFAM: [0.0, 1.0];
LPRM: [1.0, 3.0];
NAT49: [0.01, 1.0];
NATN12: [1.0, 999.0];
NPERR: [1.0, 4.0];
STAT: [10.0, 999.0];
STAT_CONJ: [0.0, 1.0];
Supp: 2416974
Δ: 256250
APAF: [0.0, 3.0];
COUPLE: [1.0, 2.0];
DIPL_15: [1.0, 999.0];
IMMI: [1.0, 2.0];
INAT: [11.0, 21.0];
INFAM: [0.0, 2.0];
LPRM: [1.0, 6.0];
NAT49: [0.1, 1.0];
NATN12: [1.0, 999.0];
NPERR: [1.0, 5.0];
STAT: [10.0, 999.0];
STAT_CONJ: [0.0, 1.0];
Supp: 2721097
Δ: 203745
APAF: [0.0, 6.0];
COUPLE: [1.0, 2.0];
DIPL_15: [1.0, 4.0];
IMMI: [1.0, 2.0];
INAT: [11.0, 21.0];
INFAM: [0.0, 2.0];
LPRM: [1.0, 9.0];
NAT49: [0.79, 1.0];
NATN12: [1.0, 999.0];
NPERR: [1.0, 5.0];
STAT: [10.0, 999.0];
STAT_CONJ: [0.0, 1.0];
Supp: 2200399
Δ: 236627
NPERR: [..., 4.0];
Supp: 2458956
Δ: 298232
IMMI: [2.0, ...];
INAT: [..., 11.0];
NAT49: [1.0, ...];
NATN12: [..., 1.0];
Supp: 2160724
Δ: 314535
APAF: [1.0, ...];
INFAM: [1.0, ...];
NPERR: [2.0, ...];
Supp: 1672518
Δ: 314465

Fig. 11. Zone E

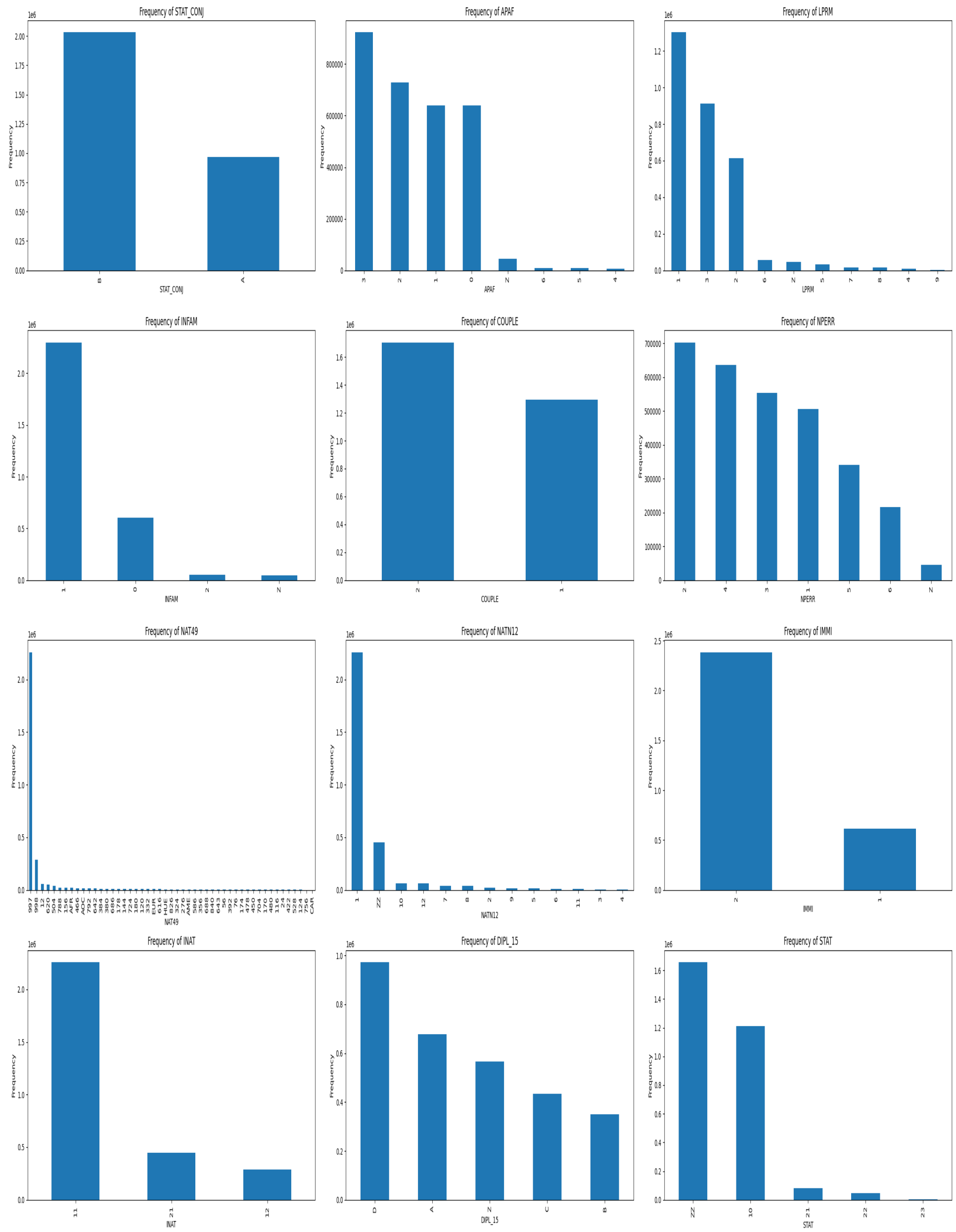

Fig. 12. Zone A: Exploratory Data Analysis

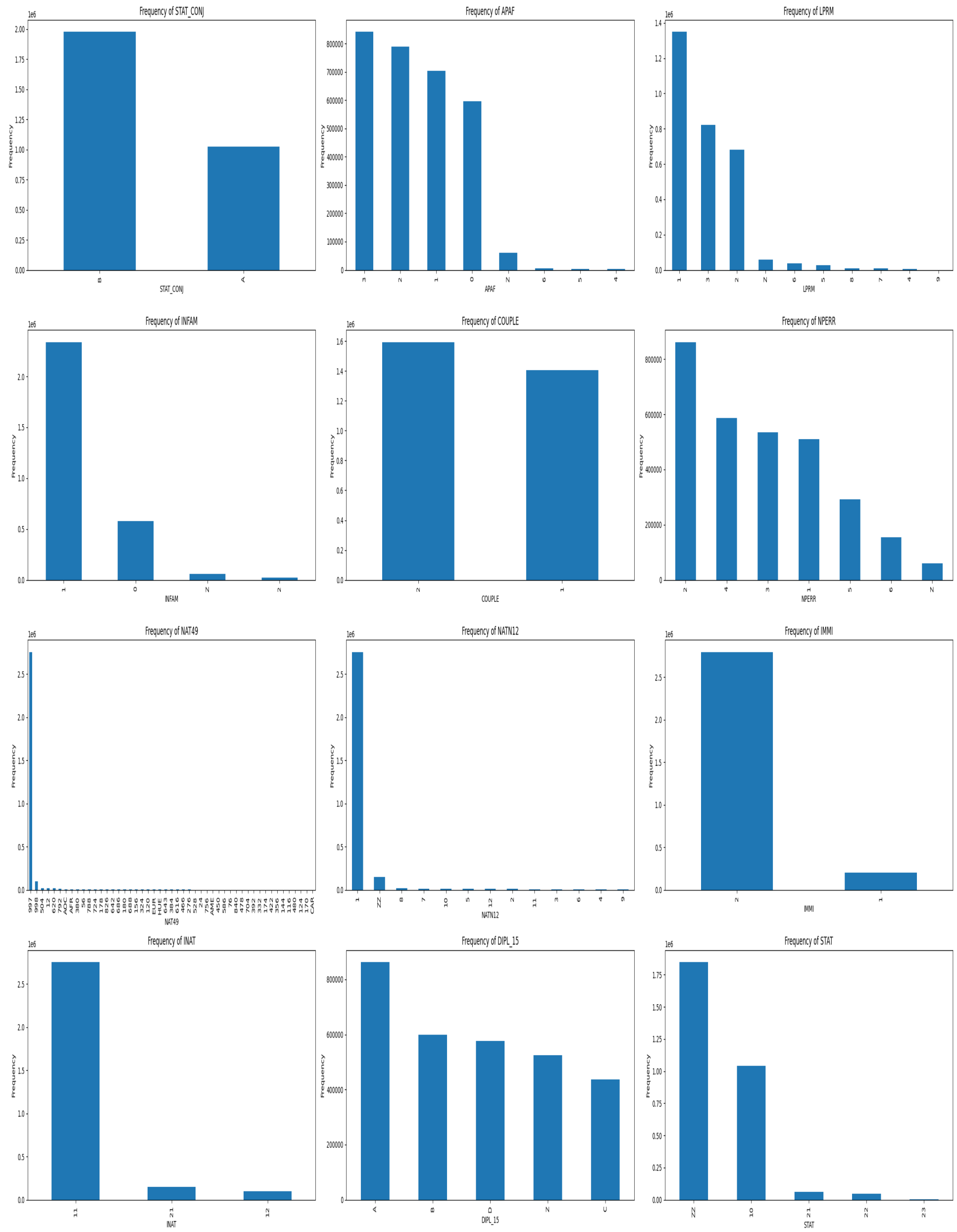

Fig. 13. Zone B: Exploratory Data Analysis

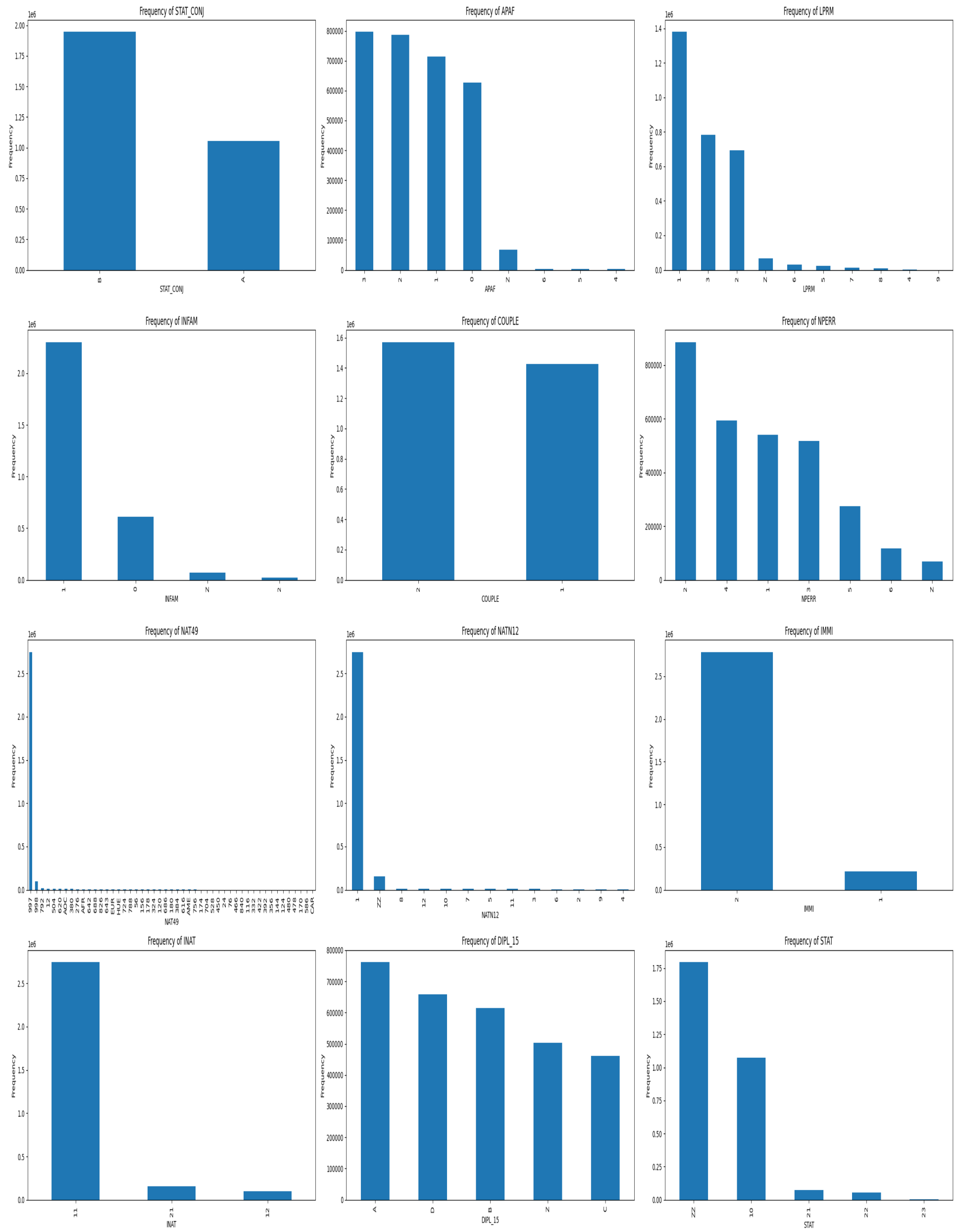

Fig. 14. Zone C: Exploratory Data Analysis

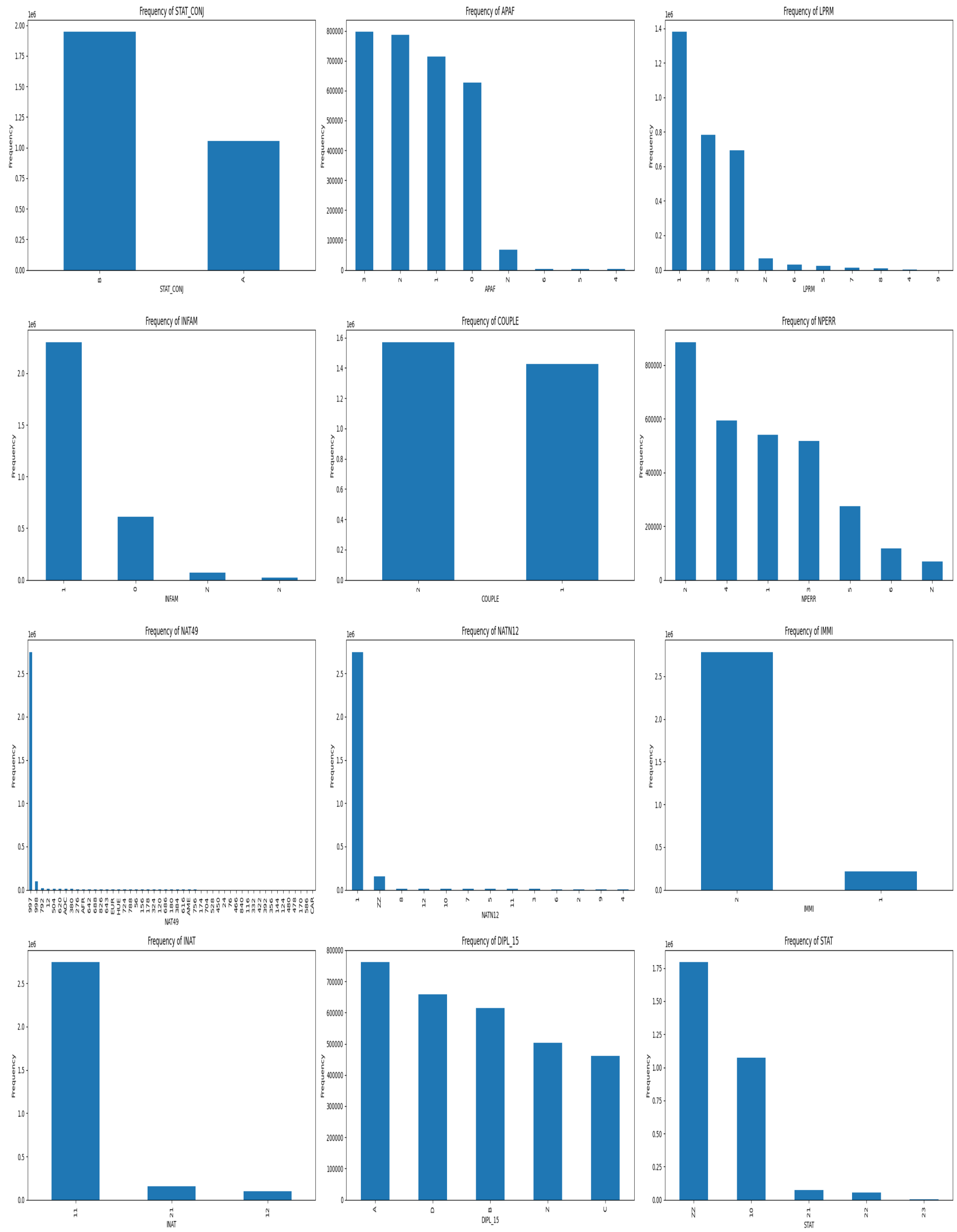

Fig. 15. Zone D: Exploratory Data Analysis

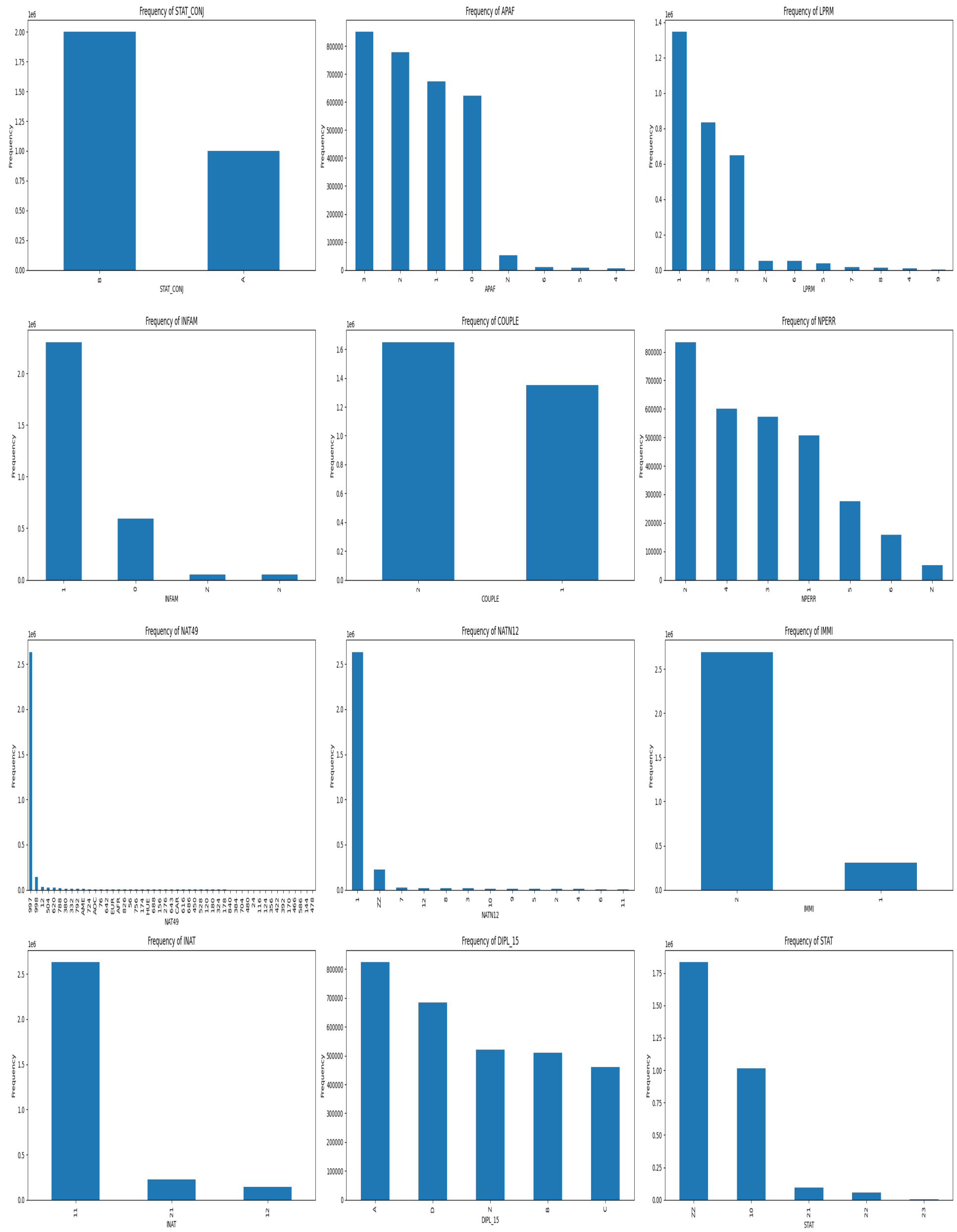

Fig. 16. Zone E: Exploratory Data Analysis