

Child Development Project Report: Predictive Analysis of Early Childhood Development Deficits

NGO Applied Project - BREIT

Angel Choquehuanca Luis Torpoco Alan Fraquita Edson

2025-10-10

Table of contents

Executive Summary	3
1. Introduction and Context	3
1.1. Background on TANI	3
1.2. Child Development Deficit Context	4
1.3. Project Objectives	5
Primary Objective	5
Secondary Objectives	6
1.4. Scope and Limitations	6
Scope	6
Limitations	6
1.5. Key Stakeholders	6
Primary Beneficiaries	6
Secondary Stakeholders	7
2. Data Reception and Consolidation	7
2.1. Data Sources	7
Internal Data	7
2.2. Data Challenges and Quality Issues	7
Structural Issues	8
Data Quality Issues	8
2.3. Consolidation Process	8
Phase 1: Data Profiling	8
Phase 2: Data Cleaning Functions	12

Phase 3: Data Integration	12
2.4. Consolidated Dataset Characteristics and created features	12
2.4.1 Consolidated dataset overview	12
2.4.2 Features created for exploratory analysis	14
3. Exploratory Data Analysis (EDA)	15
3.1. Defining population	15
3.2. Descriptive Statistics	18
3.3. Outlier Detection and Treatment	18
3.4. Correlation Analysis	19
3.5. Preliminary Insights	19
4. Key Patterns and Insights	19
5. Predictive Modeling	19
6. Recommendations for TANI	19
7. Conclusions and Next Steps	20
References	20

Executive Summary

This report presents a comprehensive predictive analysis of early childhood development deficits for Asociación Taller de los Niños (TANI), a Peruvian NGO with over 45 years of experience serving vulnerable children and families. Through data science methodologies, we analyze pre- and post-pandemic health records to develop actionable predictive models that identify children at highest risk of developmental delays.

Methodology Overview:

- Multi-phase data consolidation from pre/post-pandemic periods
- Strategic population filtering to address class imbalance
- Feature engineering

Key Findings:

- Identification of critical risk factors for targeted intervention

Impact: This work provides TANI with a decision-support framework to optimize early intervention strategies, improving outcomes for vulnerable children through data-driven prioritization.

1. Introduction and Context

1.1. Background on TANI

Asociación Taller de los Niños (TANI) is a non-governmental organization based in Peru, dedicated to improving the quality of life for children and families in vulnerable situations. With over 45 years of operational experience, TANI focuses particularly on the critical early years of life (0-5 years), implementing a comprehensive approach that integrates:

- **Health and Nutrition:** Growth monitoring, nutritional assessment, and intervention programs
- **Community Strengthening:** Family engagement and capacity building
- **Preventive Care:** Systematic health check-ups and early detection of developmental issues

1.2. Child Development Deficit Context

Fundamental Concepts of Growth and Development:

Child growth and development constitute two interrelated sets of indicators of critical utility for determining health status. Growth refers to the physical increase in weight and height/length measurements while development encompasses the progressive improvement of functional capacity and skills across multiple domains (motor, cognitive, language, and socio-emotional). Both processes are fundamentally dependent on the interaction of genetic, nutritional, and environmental factors. Therefore, systematic monitoring of normal developmental trajectories is essential to identify risk factors that may interfere with these processes, ensuring that each child achieves their full developmental potential.

The Global Burden of Early Childhood Development Deficits:

Early childhood development (ECD) deficits represent a critical public health challenge in low- and middle-income countries. According to WHO estimates, approximately 250 million children under five years in these settings are at risk of not reaching their developmental potential due to poverty and stunting (Black et al., 2017). These deficits manifest across multiple, often overlapping dimensions:

- **Stunting (Chronic Malnutrition):** Linear growth failure resulting from prolonged nutritional deficiencies, affecting approximately 22% of children globally (UNICEF, 2021)
- **Wasting (Acute Malnutrition):** Low weight-for-height indicating recent or current severe nutritional deficit
- **Developmental Delays:** Failure to achieve age-appropriate milestones in gross motor, fine motor, cognitive, language, or social-emotional domains
- **Micronutrient Deficiencies:** Particularly iron-deficiency anemia, affecting cognitive development and immune function

The Peruvian Context:

In Peru, despite significant economic growth over recent decades, substantial disparities persist in child health outcomes:

- **Chronic Malnutrition:** Affects 12.1% of children under 5 nationally, with rates exceeding 25% in rural Andean regions (INEI-ENDES, 2023)
- **Anemia Prevalence:** Reaches 43.1% in children aged 6-35 months, a critical window for brain development (INEI-ENDES, 2023)
- **Developmental Delays:** Studies indicate that children exhibit delays in one or more developmental domains, with higher rates among socioeconomically disadvantaged populations (Diaz et al., 2017)
- **Geographic Inequities:** Urban-rural and coastal-highland disparities create vastly different developmental trajectories for children

The Critical Window: First 1,000 Days:

The period from conception to age 3 (approximately 1,000 days) represents a critical window for brain architecture development. During this time, neural connections form at a rate exceeding 1 million per second, establishing the foundational circuitry for all future learning, behavior, and health (Shonkoff & Phillips, 2000). Nutritional and environmental insults during this window can result in:

- **Irreversible Stunting:** Chronic undernutrition affecting linear growth and cognitive capacity
- **Neurodevelopmental Impairment:** Compromised executive function, memory, and academic readiness
- **Immunological Vulnerability:** Increased susceptibility to infectious diseases
- **Intergenerational Transmission:** Girls who experience stunting are more likely to have low-birth-weight infants, perpetuating cycles of disadvantage

Multidimensional Determinants:

Following Bronfenbrenner's ecological systems framework (1979), child development results from interactions across multiple levels:

- **Individual Factors:** Genetic endowment, birth characteristics (gestational age, birth weight), sex
- **Family/Microsystem:** Maternal nutrition and health, breastfeeding practices, caregiver-child interactions, household food security
- **Community/Exosystem:** Access to healthcare services, availability of nutritious foods, water and sanitation infrastructure
- **Societal/Macrosystem:** Economic policies, health system organization, cultural beliefs about child-rearing

This multidimensional causality necessitates comprehensive assessment approaches that capture not only anthropometric measurements but also functional developmental capacity and contextual risk factors.

1.3. Project Objectives

This project aims to leverage TANI's data to develop actionable predictive models. Specific objectives include:

Primary Objective

Develop and validate machine learning models to predict early childhood development deficits, enabling proactive identification of at-risk children for targeted interventions.

Secondary Objectives

1. **Exploratory Analysis:** Conduct comprehensive exploratory data analysis (EDA) to identify patterns, risk factors, and data quality issues
2. **Feature Engineering:** Transform raw clinical measurements into actionable predictive features
3. **Model Development:** Compare multiple predictive algorithms and select the optimal approach for deployment
4. **Actionable Insights:** Provide evidence-based recommendations for TANI's operational and strategic decisions
5. **Data Infrastructure:** Document data quality issues and recommend improvements for future data collection

1.4. Scope and Limitations

Scope

- **Temporal Coverage:** Analysis covers post pandemic observations.
- **Population:** Children aged 0-5 years receiving services at TANI health centers
- **Predictive Modeling:** Focus on classification models for binary outcomes (deficit vs. normal)

Limitations

- **Data Quality:** As with most real-world clinical data, missing values and measurement inconsistencies are present
- **Selection Bias:** Population consists of families actively seeking TANI services, which may not represent the broader vulnerable population
- **Data Bias:** Some variables like deficit may be bias due to the nurse in turn who diagnosed the children. TANI have each age group assigned to certain nurse.
- **Causality:** While we identify predictive associations, establishing causal relationships requires additional study designs
- **External Validity:** Models will be optimized for TANI's specific population and may require recalibration for other contexts

1.5. Key Stakeholders

Primary Beneficiaries

- **Children (0-5 years):** Direct beneficiaries through improved early detection and intervention

- **Families:** Receive targeted support and guidance based on predictive risk assessments
- **TANI Clinical Staff (nurses):** Gain decision support tools for prioritizing cases and planning interventions

Secondary Stakeholders

- **TANI Leadership:** Use insights for strategic planning, resource allocation, and fundraising
 - **Public Health Authorities:** Evidence base for policy recommendations and scaling interventions
 - **Research Community:** Methodological contributions to applied machine learning in global health
-

2. Data Reception and Consolidation

2.1. Data Sources

The analysis utilizes three primary data sources provided by TANI:

Internal Data

1. **MALNUTRITION Sheet:** Longitudinal records of children's growth monitoring and nutritional assessments
2. **DEVELOPMENT Sheet:** Developmental screening results across five domains
3. **INITIAL DIAGNOSIS Sheet:** First medical diagnosis when taking the program
4. **MEDICAL ADVICE Sheet:** Relevant medical recommendations given by the nurses

2.2. Data Challenges and Quality Issues

Initial data exploration revealed several challenges common to clinical registries:

Structural Issues

- **Inconsistent Formatting:** Age represented in mixed formats (days, months, years: “4d”, “6m”, “1a8m”)
- **Variable Types:** Anthropometric measurements stored as text instead of numeric values
- **Categorical Encoding:** Multiple encoding schemes for the same concept (e.g., “SI”/“Normal” for development)

Data Quality Issues

- **Missing Values:** Inconsistent handling of “not applicable” vs. “not measured” vs. “refused”
- **Measurement Errors:** Outliers in weight, height, and head circumference requiring validation
- **Temporal Inconsistencies:** Some records show impossible sequences (e.g., height decreasing over time)
- **Duplicate Records:** Need to establish deduplication rules for repeated visits

2.3. Consolidation Process

To address these challenges, first we implemented a systematic data consolidation pipeline which consist in merging two main datasets excluding duplicates validated by TANI:

Phase 1: Data Profiling

```
=====
DATASET: MALNUTRITION
=====
```

Shape: 257,178 rows × 27 columns

Data Types Distribution:
object 23
float64 3
datetime64[ns] 1
Name: count, dtype: int64

Top 10 Variables with Missing Values:
Missing_Count Missing_Percentage

Tam_graha	253123	98.42
Tam_para	252117	98.03
Recuperado	247952	96.41
Tam_hb	243562	94.71
Razón	238809	92.86
Mantiene_Diag_Fav/Desf	188524	73.30
Lactancia	186632	72.57
CN-CA	167056	64.96
ACA	112621	43.79
P/E	47695	18.55

=====

DATASET: DEVELOPMENT

=====

Shape: 257,178 rows × 18 columns

Data Types Distribution:

object	15
float64	2
datetime64[ns]	1
Name: count, dtype: int64	

Top 10 Variables with Missing Values:

	Missing_Count	Missing_Percentage
(M) - FF	130726	50.83
CabPC	42520	16.53
(S) - Soc	12915	5.02
(L) - Len	12902	5.02
(C) - Cog	12905	5.02
(M) - FG	12902	5.02
Talla	18	0.01
Nº_HC	9	0.00
Peso	7	0.00
Nº_Control	1	0.00

=====

DATASET: MALNUTRITION (PRE PANDEMIC)

=====

Shape: 197,832 rows × 28 columns

Data Types Distribution:

```

object          22
float64         4
datetime64[ns]   1
int64           1
Name: count, dtype: int64

```

Top 10 Variables with Missing Values:

	Missing_Count	Missing_Percentage
Tam_hb	197832	100.00
Tam_graha	197832	100.00
Tam_para	197832	100.00
Recuperado	195054	98.60
Razón	187570	94.81
CabPC	133848	67.66
Lactancia	129627	65.52
ACA	68300	34.52
Mantiene_Diag_Fav/Desf	39477	19.95
CN-CA	38391	19.41

```
=====
DATASET: DEVELOPMENT (PRE PANDEMIC)
=====
```

Shape: 197,832 rows × 19 columns

Data Types Distribution:

```

object          16
datetime64[ns]   1
float64         1
int64           1
Name: count, dtype: int64

```

Top 10 Variables with Missing Values:

	Missing_Count	Missing_Percentage
(M) - FF	197831	100.00
CabPC	133848	67.66
(S) - Soc	395	0.20
(L) - Len	390	0.20
(C) - Cog	390	0.20
(M) - FG	392	0.20
Nº_HC	95	0.05
Talla	5	0.00
Peso	2	0.00

Enfermera 6 0.00

Note: The following datasets of ‘Initial diagnosis’ & ‘Medical advice’ were given after first exploratory analysis which have the objective of narrow down the population. Therefore these dataset considered just recent years.

=====

DATASET: INITIAL DIAGNOSIS

=====

Shape: 35,705 rows × 45 columns

Data Types Distribution:

object	31
float64	7
bool	5
datetime64[ns]	2
Name: count , dtype: int64	

Top 10 Variables with Missing Values:

	Missing_Count	Missing_Percentage
Motivo_baja	35704	100.00
Otros_Motivos	35705	100.00
Motivo_Relactancia	35705	100.00
Diag_Nutricional	35672	99.91
Talla_actual	35659	99.87
PC_actual	35659	99.87
Condicion_mama	35643	99.83
Dx 2	35642	99.82
Dx 1	35081	98.25
complicacion_parto_bebe	35069	98.22

=====

DATASET: MEDICAL ADVICE

=====

Shape: 78,918 rows × 14 columns

Data Types Distribution:

bool	7
object	3

```

int64      2
datetime64[ns]    1
float64     1
Name: count, dtype: int64

Top 10 Variables with Missing Values:
      Missing_Count  Missing_Percentage
Nº_Control           2                  0.0

```

Phase 2: Data Cleaning Functions

Functions for parsing variables like Age, Date, Development area, and others were implemented.

Phase 3: Data Integration

```

Data Cleaning Results:
Original records (Malnutrition): 257,178
Original records (Development): 257,178
Original records (Malnutrition Pre-pandemic): 197,832
Original records (Development Pre-pandemic): 197,832
After cleaning and consolidation (pre and post pandemic datasets): 454,901

```

2.4. Consolidated Dataset Characteristics and created features

2.4.1 Consolidated dataset overview

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 454901 entries, 0 to 454900
Data columns (total 32 columns):
 #   Column          Non-Null Count  Dtype  
--- 
 0   Fecha           454901 non-null   object 
 1   N_HC            454901 non-null   float64
 2   Tipo_Paciente  454901 non-null   object 
 3   Categoría      454901 non-null   object 
 4   Sexo             454901 non-null   object 
 5   Edad            454901 non-null   object 
 6   N_Control       454900 non-null   float64
 7   Peso            454892 non-null   object 

```

```

8   Talla           454878 non-null  object
9   CabPC          278593 non-null  object
10  Diag_Nacimiento 454901 non-null  object
11  P/T            407210 non-null  object
12  T/E            407216 non-null  object
13  P/E            407210 non-null  object
14  Ganancia_Peso_Talla 412822 non-null  object
15  Dx_Nutricional 454894 non-null  object
16  CN-CA          249503 non-null  object
17  Mantiene_Diag_Fav/Desf 226959 non-null  object
18  Recuperado     12003 non-null   object
19  Lactancia      138719 non-null  object
20  Razón          28618 non-null   object
21  ACA            274016 non-null  object
22  Tam_para       5061 non-null   object
23  Tam_graha      4055 non-null   object
24  Tam_hb          13614 non-null  float64
25  Acompaña_control 454901 non-null  object
26  Enfermera      454890 non-null  object
27  (M) - FG        441618 non-null  object
28  (M) - FF        126435 non-null  object
29  (C) - Cog       441617 non-null  object
30  (L) - Len        441620 non-null  object
31  (S) - Soc        441602 non-null  object
dtypes: float64(3), object(29)
memory usage: 111.1+ MB
None

```

Consolidated Dataset Summary Statistics:

Numeric variables:

	count	mean	std	min	25%	50%	\
N_HC	454901.0	66360.885111	17133.564981	0.0	52871.0	67695.0	
N_Control	454900.0	4.246393	2.783653	0.0	2.0	4.0	
Tam_hb	13614.0	11.411892	1.295860	6.6	10.8	11.4	
	75%	max					
N_HC	81026.0	95801.0					
N_Control	6.0	14.0					
Tam_hb	12.0	107.0					

2.4.2 Features created for exploratory analysis

ORIGINAL FEATURES:

- **Fecha:** Medical care date for the child
- **N_HC:** Medical record number assigned to the patient
- **Sexo:** Child's gender. F: Female; M: Male
- **Peso:** Recorded weight of the child in kilograms
- **Talla:** Recorded height/stature of the child in centimeters
- **CabPC:** Head circumference of the child in centimeters
- **Diag_Nacimiento:** Birth diagnosis. Values: Normal, Pretérmino (Preterm), BPN (Low Birth Weight), Macrosómico (Macrosomic)
- **Ganancia_Peso_Talla:** Weight and height gain (adequate or inadequate). Combinations: GIP/GIT, GAP/GIT, GIP/GAT, GAP/GAT
- **Dx_Nutricional:** Nutritional diagnosis. N: Normal, O: Obesity, R: Risk, S: Overweight, DC: Chronic M., DA: Acute M., DG: Global M.
- **CN-CA:** Continuity of Nutritional Dx. CN: New Case, CA: Old Case
- **Mantiene_Diag_Fav/Desf:** Nutritional diagnosis follow-up. Fav: Favorable, Desf: Unfavorable
- **Recuperado:** Recovery in nutritional diagnosis (e.g.: Global M. to Normal, Risk to Normal, etc.)
- **Lactancia:** Current breastfeeding type (up to 6 months). LME: Exclusive Breastfeeding, LMX: Mixed Breastfeeding, LA: Artificial Feeding
- **Razón:** Reason for LMX or Artificial. 1: Others, 2: Own Will, 3: Low Production, 4: Study/Work, 5: Medical Indication
- **ACA:** Active complementary feeding. YES/NO
- **Tam_para:** Parasite test result. Positive/Negative
- **Tam_graha:** Graham test result. Positive/Negative
- **Tam_hb:** Hemoglobin dosage result
- **T/E_cat:** Height/Age percentile category. N: Normal, R: Risk, DC: Chronic M.
- **P/E_cat:** Weight/Age percentile category. N: Normal, R: Risk, DA: Acute M.
- **P/T_cat:** Weight/Height percentile category. N: Normal, O: Obesity, S: Overweight, DA: Acute M.

TRANSFORMATIONS:

- **edad_meses:** Child's age expressed in months
- **flg_cognitivo:** Binary flag for cognitive development area (0: Normal, 1: Deficit)
- **flg_lenguaje:** Binary flag for language development area (0: Normal, 1: Deficit)
- **flg_motora_fina:** Binary flag for fine motor development area (0: Normal, 1: Deficit)
- **flg_motora_gruesa:** Binary flag for gross motor development area (0: Normal, 1: Deficit)
- **flg_social:** Binary flag for social development area (0: Normal, 1: Deficit)

- **flg_alguna:** Flag indicating if there is any deficit in any development area (0: No, 1: Yes)
 - **flg_total:** Flag indicating if there is deficit in all development areas
 - **flg_social_lenguaje:** Binary flag for language or social development area (0: Normal, 1: Deficit)
 - **control Esperado:** Expected check-up number according to child's age
 - **primer_alguna:** Check-up number where the first deficit was detected (**flg_alguna=1**)
 - **primer_control Esperado:** First expected check-up number according to child's age
 - **ultimo_control:** Last recorded check-up number
 - **cant_controles_primer_alguna:** Number of check-ups until first deficit detection
 - **cantidad_controles:** Total number of check-ups performed
 - **num_controles_previos_deficit:** Number of previous check-ups before deficit detection
 - **nro_control_recreado:** Recreated or recalculated check-up number
 - **num_controles_posteriores_deficit:** Number of check-ups after detected deficit
-

3. Exploratory Data Analysis (EDA)

3.1. Defining population

For the definition of the population to work with TANI recommended to focus on social and language areas. This due to the rare cases in motor development deficits. On the other hand, the election of records since 2023 were defined because of imbalanced target in previous years. When validating with TANI they confirmed that there had been changed when defining "deficits" and that is why there is more cases in recent years.

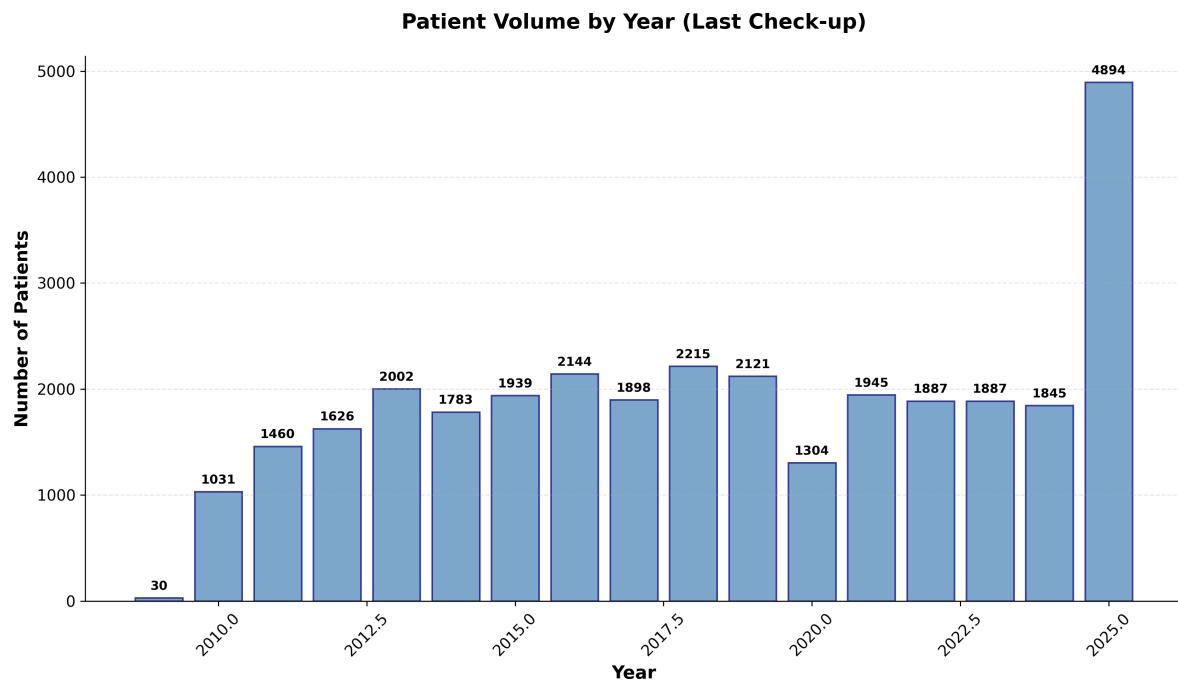
```
=====
EXPLORATORY ANALYSIS: POPULATION DEFINITION (BY PATIENT N_HC)
=====
```

Total records in dataset: 454,901
 Total unique patients (N_HC): 32,011

DEFICIT PREVALENCE BY PATIENT

Any deficit	:	100 (0.31%)
Language deficit	:	91 (0.28%)
Social deficit	:	15 (0.05%)
Social OR Language deficit	:	86 (0.27%)

Chart saved as: patient_volume_by_year.png



PREVALENCE BY YEAR (LAST CHECK-UP OF EACH PATIENT)

year	Patients	flg_alguna	%_alguna	flg_lenguaje	%_lenguaje	flg_social	\
2009	30	0	0.00	0	0.00	0	
2010	1031	0	0.00	0	0.00	0	
2011	1460	0	0.00	0	0.00	0	
2012	1626	0	0.00	0	0.00	0	
2013	2002	0	0.00	0	0.00	0	
2014	1783	0	0.00	0	0.00	0	
2015	1939	0	0.00	0	0.00	0	
2016	2144	0	0.00	0	0.00	0	
2017	1898	0	0.00	0	0.00	0	
2018	2215	0	0.00	0	0.00	0	
2019	2121	7	0.33	6	0.28	0	
2020	1304	21	1.61	22	1.69	1	
2021	1945	5	0.26	4	0.21	0	
2022	1887	2	0.11	1	0.05	1	

2023	1887	6	0.32	5	0.26	1
2024	1845	8	0.43	8	0.43	1
2025	4894	51	1.04	45	0.92	11

year	%_social	flg_social_lenguaje	%_soc_leng
2009	0.00	0	0.00
2010	0.00	0	0.00
2011	0.00	0	0.00
2012	0.00	0	0.00
2013	0.00	0	0.00
2014	0.00	0	0.00
2015	0.00	0	0.00
2016	0.00	0	0.00
2017	0.00	0	0.00
2018	0.00	0	0.00
2019	0.00	6	0.28
2020	0.08	17	1.30
2021	0.00	4	0.21
2022	0.05	1	0.05
2023	0.05	5	0.26
2024	0.05	8	0.43
2025	0.22	45	0.92

The final population meets the following filters:

3. FILTER CRITERIA ANALYSIS (YEAR >= 2023)

Total patients with last check-up 2023+: 8,626

Patients with primer_alguna_en [1,2,3]: 8,059
 Patients with cant_controles_primer_alguna >= 6: 7,527
 Patients with ultimo_control >= 19: 4,218
 Patients with ALL criteria: 3,721

FILTERED POPULATION:

Total: 3,721 patients
 With deficit (flg_social_lenguaje=1): 47
 Without deficit (flg_social_lenguaje=0): 3,674
 Positive class balance: 1.26%

3.2. Descriptive Statistics

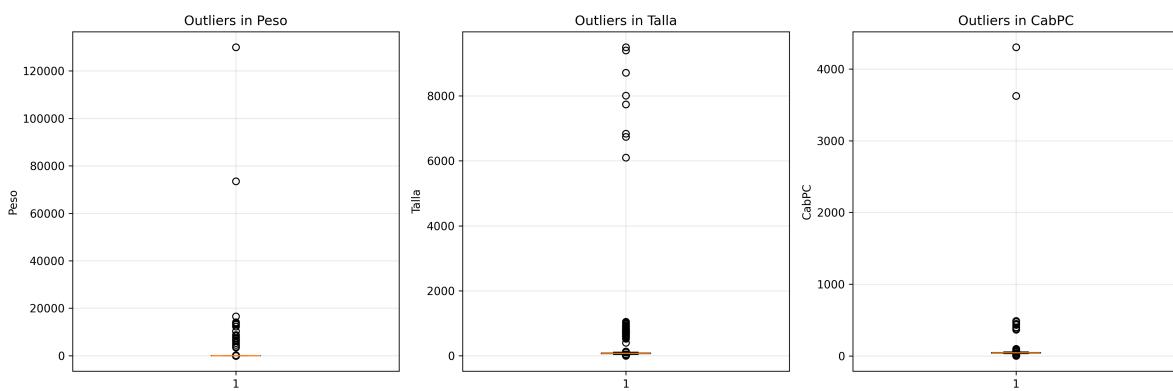
3.3. Outlier Detection and Treatment

Some numeric variables were treated with the 99th percentile after validating that most of the outliers were human error when transcribing to their databases from physical records.

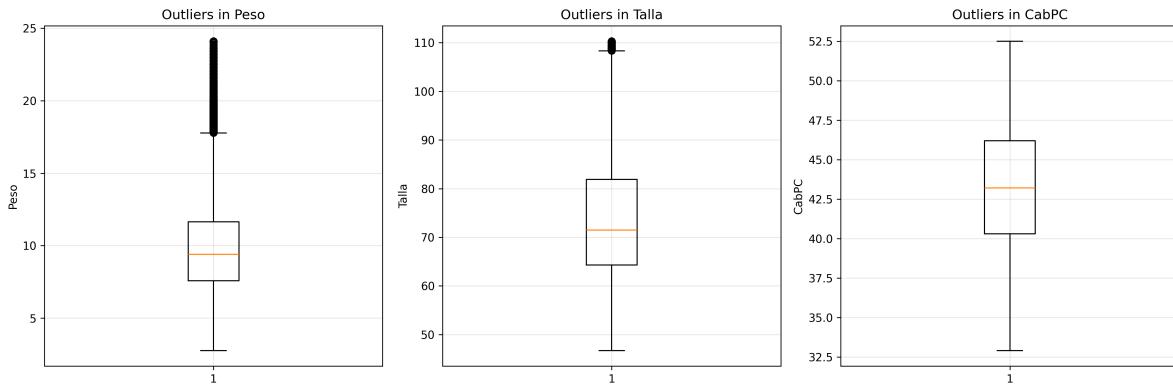
HOJA DE CONTROL DE DESARROLLO INFANTIL TEMPRANO (DIT)												FICHA DIT:																			
NOMBRE		Perez Abemayor Angel Halen			HISTORIA CLINICA:			84.892			Niños			Fam.			Varuna			Ex.His.			Parasito			Pto.			Sintomas		
FECHA	TALLER	EDAD	PESO	TALLA	PERCENTIL	Ganar	EXAMEN FISICO	Nº C	DACTYLIA ALIMENTACION	ACA	Dx Nutri	D.Soc	D.Fis	D.Len	D.Cog	DF	Fam.	Vacuna	Ex.His	Parasito	Pto.	Sintomas									
18 MAY 2023	4	10m 4m	12000	79	>12 -1	SIP	Sano	3	2(3) T(1)	ACA N	L	EP	L	L	VC																
21 JUL 2023	4	10m 6m	81	>12 -1	SIP	Sano	4	1(1) 2(1)	ACA N	L	EP	L	L	VC																	
20 SET 2023	16	10m 8m	81	>12 -1	SIP	Sano	NO ASISTIO A SU CITA CRED COLECTIVO	5	1(1) 2(1)	ACA N	L	EP	L	L	VC																
22 SET 2023	16	10m 8m	85	>12 -1	SIP	Sano	6	1(1) 2(1)	ACA N	L	EP	L	L	VC																	
20 NOV 2023	14	10m 8m	84.4	S	N	N	GT S	5.1	1(1) 2(1)	ACA N	L	EP	L	L	VC																
23 ENR 2024	20	14m 8m	85.6	S	N	N	GT S	5.1	1(1) 2(1)	ACA N	L	EP	L	L	VC																
19 ABR. 2024	20	14m 8m	85.6	S	N	N	NO ASISTIO A SU CITA CRED COLECTIVO	7	1(1) 2(1)	ACA N	L	EP	L	L	VC																
24 ABR. 2024	20	14m 8m	86.9	S	N	N	GT S	5.1	1(1) 2(1)	ACA N	L	EP	L	L	VC																
19 JUL. 2024	20	14m 8m	86.9	S	N	N	NO ASISTIO A SU CITA CRED COLECTIVO	8	1(1) 2(1)	ACA N	L	EP	L	L	VC																
21 AGO. 2024	20	16-200 90 APR	85	S	NN	NN	GT S	5.1	1(1) 2(1)	ACA N	L	EP	L	L	VC																
25 ENE. 2025	30	16-200 92.3	85	S	NN	NN	GT S	5.1	1(1) 2(1)	ACA N	L	EP	L	L	VC																
28 ENE. 2025							ASISTIO CONSEJERIA NUTRICIONAL N°1																								
26 ABR. 2025	30	16-200 94.2	85	S	NN	NN	GT S	5.1	1(1) 2(1)	ACA N	L	EP	L	L	VC																

Figure 1: Original records TANI

Before treatment:



After imputation:



3.4. Correlation Analysis

3.5. Preliminary Insights

4. Key Patterns and Insights

[To be developed based on EDA findings]

5. Predictive Modeling

[To be developed in next phase]

6. Recommendations for TANI

[To be developed based on model results]

7. Conclusions and Next Steps

[Final synthesis]

References

- Black, M. M., et al. (2017). Early childhood development coming of age: science through the life course. *The Lancet*, 389(10064), 77-90.
 - World Health Organization. (2006). WHO Child Growth Standards.
 - Instituto Nacional de Estadística e Informática (INEI). (2023). Encuesta Demográfica y de Salud Familiar.
 - Díaz AA, Gallesteay JB, Vargas-Machuca R, Velarde RA. Desarrollo infantil en zonas pobres de Perú [Child development in poor areas of Peru]. Rev Panam Salud Publica. 2017 Jun 8;41:e71. Spanish. doi: 10.26633/RPSP.2017.71. PMID: 28614480; PMCID: PMC6660845.
 - National Research Council (US) and Institute of Medicine (US) Committee on Integrating the Science of Early Childhood Development; Shonkoff JP, Phillips DA, editors. From Neurons to Neighborhoods: The Science of Early Childhood Development. Washington (DC): National Academies Press (US); 2000. 8, The Developing Brain. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK225562/>
 - Bronfenbrenner, U. (1979). The ecology of human development. Harvard University Press.
-