Supplementary Table 3. Results Based on Data Understanding in CRISP-DM (N = 125)

Index	Author (Year)	Software	Data Sources and Types	EDA Method	Subjects	Sample Size
A1	Marketta Hiissa et al. (2006)	SPSS	EMRs	NI	Nursing narratives	43 patients (1,363 narrative segments)
A2	Manabu Nii et al. (2007)	MeCab	EMRs	NI	Nursing-care texts	370 nursing-care text samples
A3	Laurence G. Moseley et al. (2008)	SPSS	Student records	Frequency, Percentage	Nursing students	528 nursing students (3,978 records)
A4	Alexander Zlotnik et al. (2015)	Stata, Weka	EMRs	Frequency	Nursing staff	553,000
A5	Manabu Nii et al. (2016)	MeCab	EMRs	NI	Nursing-care texts	8,313 nursing care texts
A6	Robert Sherwin et al. (2017)	NI	EHRs	Frequency, Percentage	Emergency department patients	912 sepsis patients, 975 non-sepsis patients.
A7	Shinichiroh Yokota et al. (2017)	R	EMRs	Frequency, Percentage	Hospital inpatients	45,257 patients (1,223,687 patient-days)
A8	Jeungok Choi et al. (2018)	MATLAB	Survey data	Frequency, Percentage	Community-dwelling older adults	3,377
A9	Stephen I. Gallant et al. (2018)	NI	EHRs	Frequency, Percentage	Hospital inpatients	203,000 adult inpatient admissions
A10	Eliezer Bose et al. (2019)	R	EMRs	NI	Home visiting clients who received public health nursing services	756
A11	Gerald C. Gannod et al. (2019)	NI	Survey data	NI	Nursing home residents	255
A12	Steven G. Johnson et al. (2018)	Python	EHRs	NI	Flowsheet data types in electronic health records	126,957 flowsheet rows
A13	Zfania Tom Korach et al. (2019)	Python	EHRs	Frequency	Hospitalized patients	45,299 patients
A14	Jae Yung Kwon et al. (2019)	R	EHRs	Frequency, Percentage	Diabetes patients	101,766
A15	Suzanne S. Sullivan et al. (2019)	SPSS	EHRs	Frequency, Percentage	Community-dwelling older adults	635,590
A16	Maxim Topaz et al. (2019)	R	EHRs	Frequency, Percentage	Home health care patients	1,149,586 notes from 89,459 patients
A17	Heather Brom et al. (2020)	JMP Pro, SAS, STATA	EHRs	Frequency, Percentage	Patients	2,165 clinical encounters
A18	Roschelle L. Fritz et al. (2020)	NI	IoT, Digital based data	Frequency, Percentage	Independent older adults reporting pain	Approx 720,000(11 participants, with 27 pain-related events analyzed)

(Continued on the next page)

Supplementary Table 3. Continued

Index	Author (Year)	Software	Data Sources and Types	EDA Method	Subjects	Sample Size
A19	Christopher M. Horvat et al. (2020)	R	EHRs	Frequency, Percentage	Pediatric patients	11,826
A20	Renjie Hu et al. (2020)	NI	Survey data	Frequency	Nurses	328
A21	Mireia Ladios-Martin et al. (2020)	R	EMRs	Frequency, Percentage, Mean, Standard deviation	ICU patients at risk of developing pressure injuries	6,694
A22	Soo-Kyoung Lee et al. (2020)	R	Survey data	Frequency, Percentage, Mean, Standard deviation, Minimum, Maximum	Nursing home residents	60 nursing homes
A23	Chen Liang et al. (2020)	NI	EHRs	NI	Patient safety event reports	3,333
A24	David S. Lindberg et al. (2020)	R	EHRs	Frequency, Percentage	Inpatients	814
A25	Jung In Park et al. (2020)	Python, SQL, Weka	EHRs	NI	Patients	11,226
A26	Maxim Topaz et al. (2020)	KNIME	EHRs	NI	Home health care patients	112,237 home health care episodes (89,459 unique patients)
A27	Dana M. Womack et al. (2020)	Python	IoT, Digital based data	Frequency, Mean, Standard deviation	Nurses	366 work shifts with data from 144 nurses
A28	Ran An et al. (2021)	Python, R, SPSS	EMRs	Frequency, Percentage, Mean, Standard deviation	ICU patients	300
A29	Linyan Chen et al. (2021)	Python	Image or voice or video records	Frequency, Percentage, Mean, Standard deviation	Cancer patients	232
A30	Aaron Conway et al. (2021)	R	IoT, Digital based data	Frequency, Percentage, Median, IQR	Patients	384
A31	Alberto Garcés-Jiménez et al. (2021)	WEKA, Apache Spark	IoT, Digital based data	Frequency, Percentage, Mean, Standard deviation	Nursing home residents	60 Nursing homes (6,297 data)
A32	Li Hannaford et al. (2021)	R	Student records	Frequency, Percentage, Missing value detection	Nursing students	773
A33	Farinaz Havaei et al. (2021)	R	Survey data	Frequency, Mean, Standard deviation, Minimum, Maximum	Nurses	4,029
A34	Elizabeth P. Howard et al. (2021)	C++	EHRs	Frequency, Percentage, Mean, Standard Error	Patients	3,592,995
A35	Mingyue Hu al. (2021)	R	Survey data	NI	Community-based older adults	6,718
A36	Oleksandr Ivanov et al. (2021)	Java, Python	EHRs	Frequency, Percentage	Patients	166,175

(Continued on the next page)

Supplementary Table 3. Continued

Index	Author (Year)	Software	Data Sources and Types	EDA Method	Subjects	Sample Size
A37	Liuqi Jin et al. (2021)	NI	EHRs	NI	Patients	1,483
A38	Jisu Kim et al. (2021)	SPSS, R	EMRs	Frequency, Percentage, Mean, Standard deviation	Elderly patients with mechanical valve replacements	598
A39	Soo-Kyoung Lee et al. (2021)	SPSS, R	Survey data	Frequency, Percentage, Mean, Standard deviation	Nursing home residents	60 nursing homes
A40	Chia-Hui Liu et al. (2021)	SPSS, WEKA	EMRs	NI	Inpatients	108,940
A41	Tamara G. R. Macieira et al. (2021)	R	EHRs	Percentage	Patients	4,354
A42	Takuro Nagata et al. (2021)	Python	Image data	Frequency, Mean, Standard deviation	Patients	31
A43	Gojiro Nakagami et al. (2021)	Python	EHRs	Frequency, Percentage, Mean, Standard deviation	Inpatients	75,353
A44	Wenyu Song et al. (2021)	R, Python	EHRs	Mean, Standard deviation	Hospitalized patients	15,313
A45	Rumei Yang et al. (2021)	Python	Survey data	NI	Community-dwelling older adults	214,777
A46	Huaqiong Zhou et al. (2021)	R	EMRs	Frequency, Percentage, Mean, Standard deviation	Pediatric patients	940
A47	Yanhong Dong et al. (2022)	Python	Survey data	Frequency, Percentage, Mean, Median, IQR	Healthcare workers	1,122
A48	Farinaz Havaei et al. (2022)	R	Survey data	NI	Nurses	4,029
A49	Tingting Hu et al. (2022)	MATLAB, SPSS	EHRs	Frequency, Percentage, Mean, Standard deviation	Patients	907
A50	Shuai Jin et al. (2022)	R	EMRs	Frequency, Percentage	Cancer patients	1,035
A51	Mireia Ladios-Martin et al. (2022)	Azure ML, R	EMRs	Frequency, Percentage, Mean, Standard deviation	Patient	22,515
A52	Young Ji Lee et al. (2022)	Python	Social media	Frequency	Patient	853
A53	Anup Kumar Mishra et al. (2022)	Python	IoT, Digital based data	Frequency, Mean, Standard deviation	Older adults	92
A54	Kyoung Ja Moon et al. (2022)	NI	Survey data	Frequency, Percentage, Mean, Standard deviation	Patients	206
A55	Nikhil Padhye et al. (2022)	JMP Pro, Python, R	IoT, Digital based data	Frequency, Percentage, Median, IQR	Patients	26
A56	Dongni Qian et al. (2022)	SPSS	EMRs	Frequency, Percentage, Mean, Standard deviation	Patients	80

(Continued on the next page)

Supplementary Table 3. Continued

Index	Author (Year)	Software	Data Sources and Types	EDA Method	Subjects	Sample Size
A57	Javier Rojo et al. (2022)	NI	Survey data	NI	Older adults	50 social and healthcare centers (exact number of individuals not mentioned)
A58	Jiyoun Song et al. (2022)	R, Weka	EHRs	Frequency, Percentage, Mean, Standard deviation	Home health care patients	86,823 episodes from 66,317 unique patients
A59	Tobias R. Spiller et al. (2022)	R	EHRs	Frequency, Percentage, Mean, Standard deviation	Patients	324,050
A60	Katie Walker et al. (2022)	Python	EMRs	Frequency, Percentage, Median, IQR	Patients	1,930,609 patient episodes
A61	Melyana Nurul Widyawati et al. (2022)	Python	Survey data	NI	Pregnant women	5,324
A62	Jie Xu et al. (2022)	R	EHRs	Frequency, Percentage, Mean, Standard deviation, Median, IQR	Patients	618
A63	Olga Yakusheva et al. (2022)	R	EHRs	Frequency, Percentage, Mean, Standard deviation	Patients	137,778
A64	Ayla İ rem Aydın et al. (2023)	Python, SPSS	Image or voice or video records	Frequency, Percentage, Mean, Standard deviation	Children	83 children (151 pain assessments)
A65	Rui CHEN et al. (2023)	NI	EMRs	NI	Cancer patients	230
A66	Ya-Huei Chen et al. (2023)	RapidMiner	EHRs	Frequency, Percentage, Mean, Standard deviation	Inpatients	53,122
A67	Pei-Yu Dai et al. (2023)	NI	EMRs	NI	Patients	121,306 data points from 1,827 patients
A68	Odai Y. Dweekat et al. (2023)	Minitabl, Python, R	EHRs	NI	Patients	158,892
A69	Juliet Edgcomb et al. (2023)	NI	EHRs	Frequency, Percentage	Patients	1,093
A70	Ajeet Gajra et al. (2023)	CORE	EMRs	NI	Cancer patients	28,578
A71	Farinaz Havaei et al. (2023)	R	Survey data	Frequency, Percentage, Mean, Standard deviation, Minimum, Maximum	Nurses	4,066
A72	Sharon Hewner et al. (2023)	R	EHRs	Frequency, Percentage, Mean	Patients	3,438
A73	Sunho Im et al. (2023)	SAS, Python	EHRs	Frequency, Percentage, Mean, Standard deviation	Patients	5,211
A74	Junglyun Kim et al. (2023)	SPSS, R	Survey data	Frequency, Percentage, Mean, Standard deviation	Older adults	650

(Continued on the next page)

Supplementary Table 3. Continued

Index	Author (Year)	Software	Data Sources and Types	EDA Method	Subjects	Sample Size
A75	Seong-Kwang Kim et al. (2023)	Python	Others (Hospital personnel department data)	Frequency, Percentage, Mean, Standard deviation	Nurses	1,410
A76	Hyungbok Lee et al. (2023)	R	EHRs	Frequency, Percentage, Mean, Standard deviation	Patients	968
A77	Hyungbok Lee et al. (2023)	R	EHRs	Frequency, Percentage, Mean, Standard deviation	Nurses	7,259
A78	Lin-Lin Lee et al. (2023)	SAS, SigmaPlot, SPSS	Image or voice or video records	Percentage, Mean, Standard deviation	Patients	30
A79	Lingjuan Li et al. (2023)	R	EHRs	Frequency, Percentage	Patients	234
A80	Pei-Hung Liao et al. (2023)	Modeller, SPSS	Survey data	Frequency, Percentage, Mean, Standard deviation	Home-based individuals at risk for sarcopenia	120
A81	Sarah R. Martha et al. (2023)		EMRs	Frequency, Percentage	Patients	5
A82	Aruna Jothi Shanmugam et al. (2023)	NI	Survey data	NI	Students	180
A83	Araceli Rodríguez Vico et al. (2023)	NI	EHRs	NI	Patients	1,572
A84	Zeping Yan et al. (2023)	R	Survey data	Frequency, Percentage, Mean, Standard deviation	Patients	470
A85	Metin Yildiz et al. (2023)	R, SPSS	Survey data (Intercultural Sensitivity Scale, Ethnocentrism Scale, Health Tourism Awareness Scale)	Frequency, Percentage, Mean, Standard deviation, Kurtosis, Skewness	Nurses	386
A86	Ying Zhou et al. (2023)	Python, SPSS	Image or voice or video records	Frequency, Percentage, Mean, Standard deviation	Older adults	319
A87	Maryam Zolnoori et al. (2023)	Python	Image or voice or video records	Frequency, Percentage, Mean, Standard deviation, Quartiles	Patients, Nurses	46 patient-nurse encounters (23 patients with 2 encounters each), 3,494 utterances
A88	Young-Taek Park et al. (2024)	Python	Others (Health Insurance Review & Assessment Service billing data for 708 hospitals)	Frequency, Percentage, Mean, Standard deviation, Minimum, Maximum	Hospital patients (inpatients and outpatients)	708 hospitals(mean Inpatients: 16,537.8, Outpatient: 49,878.2)

(Continued on the next page)

Supplementary Table 3. Continued

Index	Author (Year)	Software	Data Sources and Types	EDA Method	Subjects	Sample Size
A89	Cynthia ABI KHALIL et al. (2024)	Python	EMRs	Frequency, Mean, Standard deviation, Handling missing value, Standard deviations	Patients	2,086
A90	Yaser ALQARRAIN et al. (2024)	NI	EHRs	NI	Nurses	266
A91	Komal Aryal et al. (2024)	R	EHRs	Frequency, Percentage	Nursing home residents	14,977
A92	Ranjana Chavan et al. (2024)	NI	Others (Historical placement data from Symbiosis College of Nursing)	NI	Students	2,253
A93	Xiaomei Chen et al. (2024)	Python	EMRs	Frequency, Percentage	Pediatric patients	438
A94	Colum Crowe et al. (2024)	Python	IoT, Digital based data	Mean, Standard deviation Median, Maximum, Minimum, Range	Patients	113
A95	Tian Dai et al. (2024)	Python, R	EHRs	NI	Patients	495
A96	Martha Duarte et al. (2024)	Python	EHRs	NI	Patients	74,636
A97	Yu-Fang Guo et al. (2024)	Python, SPSS	Survey data	Frequency, Percentage, Mean, Standard deviation, Minimum, Maximum	Nurses	1,235
A98	Rui Jin et al. (2024)	NI	EHRs	NI	Patients	5,196
A99	Arisa Kawashima et al. (2024)	Python	EHRs	Frequency, Percentage, Mean, Standard deviation	Patients	561
A100	Yeonju Kim et al. (2024)	Python	EMRs	NI	Patients	46,693
A101	Ju Hee Lee et al. (2024)	Python	EHRs	Frequency, Percentage, Mean, Standard deviation	Patients	1,529,557
A102	Pin-Chieh Lee et al. (2024)	Modeller, SPSS	EHRs	Frequency, Percentage, Mean, Standard deviation	Patients	817
A103	Renee C. B. Manworren et al. (2024)	NI	Video records	NI	Patients	1,918 frames from 120 videos
A104	Ninon Girardon da Rosa et al. (2024)	PostgreSQL, Python	EHRs	NI	Patients	43,871 patient assessments from 11,774 patients.
A105	Jihye Kim Scroggins et al. (2024)	Python	Voice records	Frequency, Percentage	Patients	23 audio recordings from 15 patients.
A106	Lu Shao et al. (2024)	R	Interview	Frequency, Percentage, Mean, Standard deviation	Patients	864

(Continued on the next page)

Supplementary Table 3. Continued

Index	Author (Year)	Software	Data Sources and Types	EDA Method	Subjects	Sample Size
A107	Madeleine Stanik et al. (2024)	Python	EMRs	NI	Patients	262,301
A108	Imam Tahyudin et al. (2024)	Python	EMRs	NI	Patients	106
A109	Metin Yıldız et al. (2024)	SPSS, R	Survey data	Frequency, Percentage, Mean, Standard deviation, Minimum, Maximum	Nurses	235
A110	Cheng Yu et al. (2024)	R	Survey data	Frequency, Percentage, Mean, Standard deviation	Nurses	1,384
A111	Wei Zhang et al. (2024)	SPSS, R	Survey data	NI	Older adults	6,997
A112	Maryam Zolnoori et al. (2024)	Python	EHRs	Frequency, Percentage, Mean, Standard deviation, Quartiles	Home health care patients	47 patients, 126 audio-recorded encounters
A113	Yunping Zhang et al. (2022)	JointPoint, R, SAS	EHRs	Frequency, Percentage, Mean, Standard deviation	Nursing home residents	107,000
A114	Goran Erfani et al. (2024)	SPSS	Survey data	Frequency, Percentage, Mean, Standard deviation	Nurses	294
A115	Tae Youn Kim et al. (2006)	Oracle, SPSS, Weka	EHRs	Frequency	Patients	2,347
A116	In Sook Cho et al. (2011)	Netica, SPSS	EMRs	Frequency, Percentage	Patients	3,348
A117	Yoko Setoguchi et al. (2016)	RapidMiner Studio	EMRs	Frequency, Percentage	Patients	8,286
A118	Mikyung Moon et al. (2017)	R	Others (National Inpatient Sample from Health Insurance Review & Assessment Service)	Frequency, Percentage, Mean, range	Patients	15,856
A119	Pacharmon Kaewprag et al. (2017)	NI	EHRs	Frequency, Percentage	Patients	7,717
A120	Xiaohong Deng et al. (2017)	SPSS	EMRs	Frequency, Percentage, Mean, Standard Deviation	patients	468
A121	Hong-Lin Chen et al. (2018)	SPSS	Others (Medical chart data and patient interviews)	Frequency, Percentage, Mean	Patients	149
A122	Hsiu-Lan Li et al. (2019)	SPSS, WEKA	EMRs	Frequency, Percentage, Mean, Standard Deviation	Patients	2,062
A123	Seul Ki Park (2020)	SPSS	EHRs	Frequency, Percentage, Mean, Standard Deviation	Patients	400

(Continued on the next page)

Supplementary Table 3. Continued

Index	Author (Year)	Software	Data Sources and Types	EDA Method	Subjects	Sample Size
A124	Sookyung Hyun et al. (2021)	NI	EHRs	Frequency, Percentage	Patients	12,654
A125	Ji-Yu CAI et al. (2021)	NI	EHRs	Frequency, Percentage, Mean, Standard Deviation	Patients	149

Azure ML: Azure Machine Learning; CORE: Care Optimization and Recommendation Enhancement; EHRs: Electronic Health Records; EMRs: Electronic Medical Records; IoT: Internet of Things; KNIME: Konstanz Information Miner; NI: No Information; SQL: Structured Query Language.