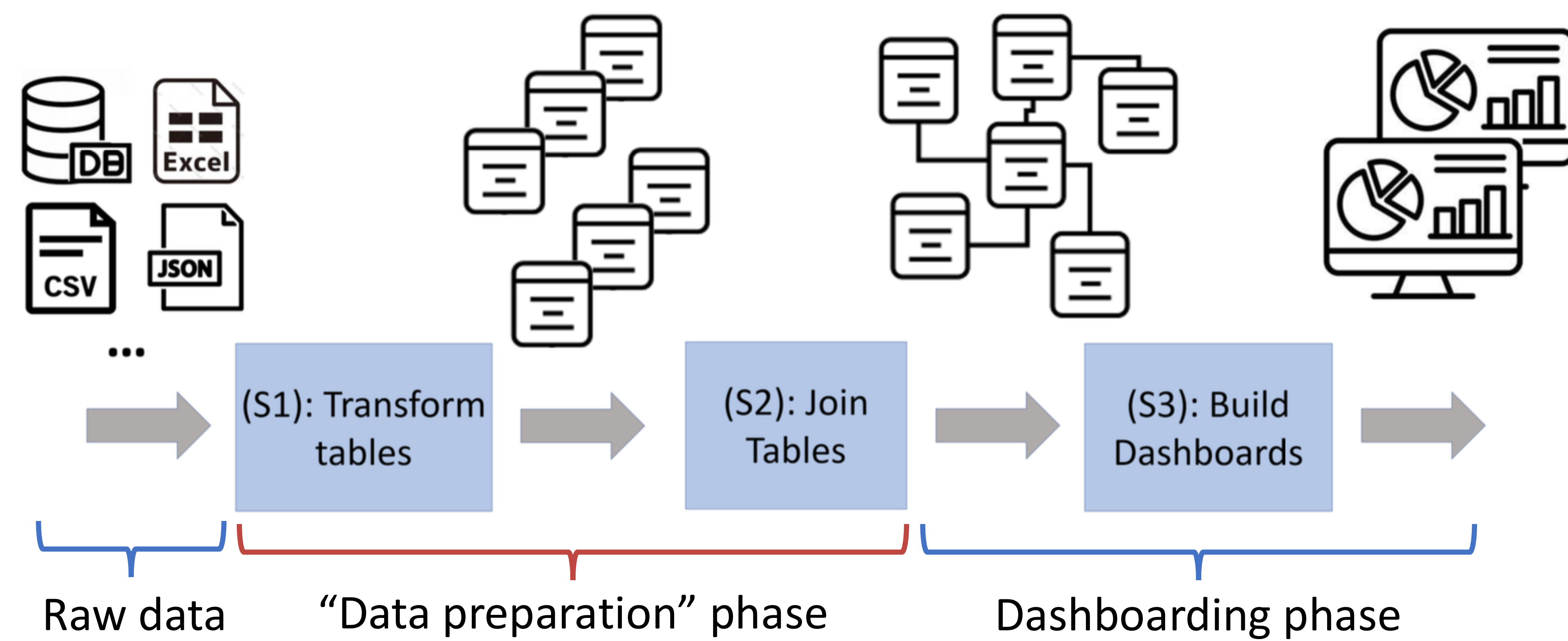


Motivation

- Business Intelligence (BI) tools like Power BI and Tableau allow enterprise users to build visualizations (or dashboards) easily.
- However, “data preparation” phase of BI continues to be a key pain point for BI users.



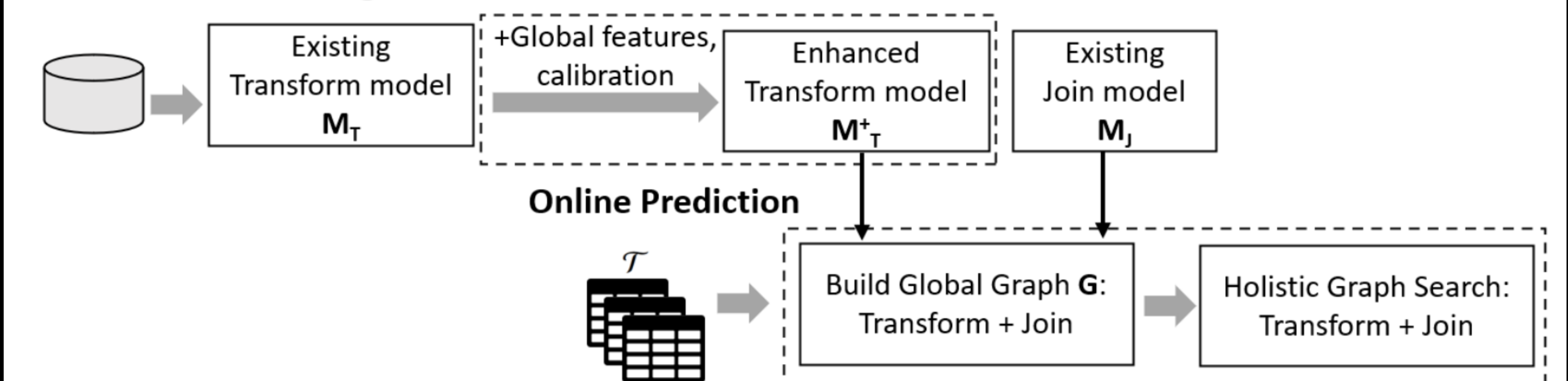
Key Insight

We systematically studied around 2K real BI projects and found:

- Transformation and join** steps are often intertwined.
- Considering the **two steps holistically** is crucial for accurate prediction.

Architecture Overview

Offline Model Training



Motivating Example

Table 1: Fertility (fact)

Country	2010	2011	2012
Poland	1.38	1.3	1.3
Chile	1.86	1.84	1.83
Morocco	2.58	2.65	2.71
Turkey	2.1	2.08	2.06

Table 2: Date (dim)

Year	IsLeap	Days
2010	No	365
2011	No	365
2012	Yes	366

Table 4: Country (dim)

Code	POL	CHL	TUR	MAR
Country	Poland	Chile	Turkey	Morocco
Continent	Europe	S. America	Europe	Africa
Developed	Yes	No	Yes	No

Table 3: Economics (fact)

Line-ID	Code	Metric	Value
CHL-CY2010	CHL	GDP	12756
CHL-CY2011	CHL	GDP	14637
CHL-CY2012	CHL	GDP	15397
CHL-CY2010	CHL	CPI	100
CHL-CY2011	CHL	CPI	103.3
CHL-CY2012	CHL	CPI	106.4
CHL-CY2010	CHL	Payroll	55.16
CHL-CY2011	CHL	Payroll	57.01
CHL-CY2012	CHL	Payroll	57.43

Table 5: Fertility (fact)

Country	Year	Fertility
Chile	2010	1.86
Chile	2011	1.84
Chile	2012	1.83
Turkey	2010	2.30
Turkey	2011	2.21
Turkey	2012	2.34

Table 7: Date (dim)

Year	IsLeap	Days
2010	No	365
2011	No	365
2012	Yes	366

Table 6: Economics (fact)

Year	Code	GDP	CPI	Payroll
2010	CHL	12756	100	55.16
2011	CHL	14637	103.3	57.01
2012	CHL	15397	106.4	57.43
2010	TUR	71022	100	93.29
2011	TUR	78102	102.3	98.10
2012	TUR	82019	104.1	99.38

Table 8: Country (dim)

Code	Country	Continent	Developed
POL	Poland	Europe	Yes
CHL	Chile	S. America	No
TUR	Turkey	Europe	Yes
MAR	Morocco	Africa	No

Experimental Evaluation

Method	AUTO-PREP		Join-only baselines		Transform + Join baselines					
	AP-P	AP-O	BI-join	GPT-j	SOP+BI-join	MOP+BI-join	GPT-t+BI-join	SOP+GPT-j	MOP+GPT-j	GPT-t+GPT-j
P_{join}	0.807	0.807	0.737	0.274	0.767	0.769	0.764	0.222	0.244	0.202
R_{join}	0.707	0.703	0.671	0.350	0.631	0.600	0.589	0.271	0.299	0.246
F_{join}	0.754	0.752	0.702	0.307	0.705	0.674	0.665	0.244	0.269	0.222

