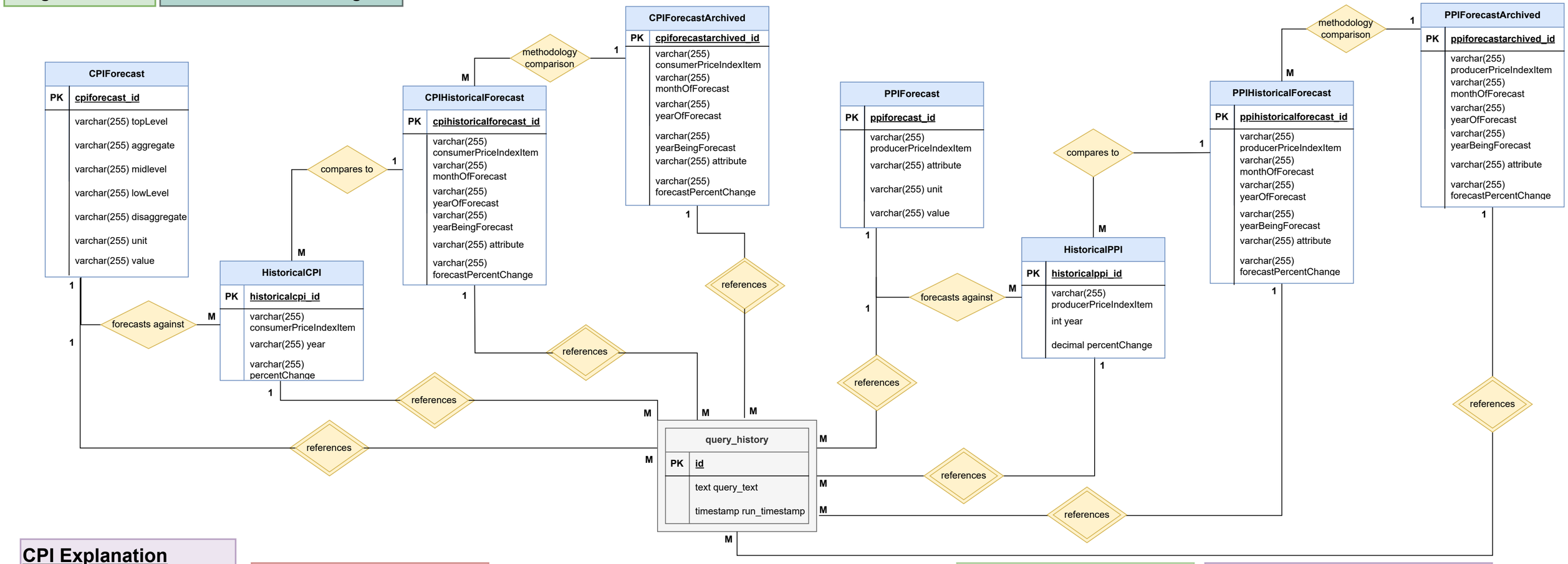
Keigen Godlaski

# Price Visualizer ER Diagram




## CPI Explanation

**CPIForecast to HistoricalCPI**

-one CPI forecast can relate to many historical CPI records

**CPIHistoricalForecast to HistoricalCPI**

-one historical forecast can be compared against many actual historical records

**CPIHistoricalForecast to CPIForecastArchived**

-many new methodology forecasts can be compared to one old methodology forecast

**query_history to ALL CPI Tables**

-many queries can be from one table and one table generates multiple queries

-all weak relationships since query_history needs the CPI/PPI tables to exist

## Note About Data Types

Practically each data type, apart from query_history, historicalppi, and all the table IDs, are stored as VARCHAR(255) in this database. At first, this seems like a strange choice since an attribute like year could be stored as an int and percentChange as a double. However, this decision was intentional. The raw data sources had inconsistent formatting (i.e. missing values, unexpected characters, and variable field types), and so storing them as strings allowed for smoother parsing and data reading. The type-specific casting is then done in the SQL queries and Java code during data processing. There's debate on which method is better, but this is the one I found to be most efficient.

## Assumptions

-forecasts are compared to historical data in one direction (ex: forecast --> actual) and not in reverse

-query_history records SQL queries but does not enforce foreign key relationships. it assumes the existence of the referenced tables without constraining them

-the raw CPI and PPI datasets did not contain a single column or combination of columns that uniquely identified each row. Due to duplicate values across fields like item, year, monthOfForecast, and attribute, no natural composite key could be reliably constructed. As a result, each table uses an artificial integer primary key (e.g., cpiforecast_id)

## PPI Explanation

**PPIForecast to HistoricalPPI**

-one PPI forecast can relate to many historical PPI records

**PPIHistoricalForecast to HistoricalPPI**

-one historical forecast can be compared against many actual historical records

**PPIHistoricalForecast to PPIForecastArchived**

-many new methodology forecasts can be compared to one old methodology forecast

**query_history to ALL PPI Tables**

-many queries can originate from one PPI table, and each PPI table can generate multiple queries

-all are weak relationships since query_history depends on the existence of the core PPI tables