Section 1: Overview of Dataset

The chosen dataset offers a multi-faceted view of the Canadian housing market, providing depth and breadth of data necessary for a comprehensive analysis. The inclusion of seasonally adjusted data ensures that the analysis accounts for seasonal variations, enabling a focus on underlying trends rather than seasonal fluctuations. The combination of price, sales, and year-over-year change data allows for a holistic view of the market, from pricing trends to sales dynamics and annual growth rates. This dataset is particularly valuable for understanding how different factors interact within the housing market, making it possible to address complex questions about market behavior and economic conditions.

The dataset not only contributes to a deeper understanding of the Canadian housing market's current state but can inform stakeholders, including policymakers, investors, and potential homeowners, making our analysis an essential tool for navigating the complexities of the housing landscape.

Data Sources and Past Uses

I am utilizing the House Price Index (HPI) dataset, sourced from the Canadian Real Estate Association (CREA). This dataset is publicly available through CREA's statistical release and the MLS® HPI tool:

- Source of Data:
    - News Release: CREA Statistics - January 15, 2024, News Release
    - MLS® HPI Tool: Try the MLS® HPI tool.
    - License Information: The data are distributed under license terms provided by CREA, which typically allow for personal, non-commercial use and require attribution when the data are used in public-facing analyses.
- Past Uses of the Data:
    - The HPI data have been instrumental in numerous studies and reports focusing on the housing market trends in Canada. Researchers and economists have used this dataset for analyzing market dynamics, forecasting trends, and examining the impact of economic policies on housing prices. For example, the Bank of Canada has utilized the HPI data in its analysis of housing market health and its implications for financial stability.

NOTE: I have not generated or simulated any part of the data. All datasets are as obtained from the official sources mentioned above.

Plans with the Dataset

I intend to use this dataset to conduct a comprehensive analysis of the Canadian housing market, focusing on long-term trends and their implications for potential homebuyers and investors. Key questions include, but are not limited to, the following:

⇒ What are the trends in housing prices for different types of properties, and how do these trends correlate with sales volumes?

⇒ Is there a correlation between sales volume and housing prices? This question seeks to understand whether higher transaction volumes lead to higher prices.

⇒ What are the patterns in year-over-year changes in housing prices and sales? This involves analyzing the data to identify trends of growth or decline.

Section 2: Description Tables

This section of our project report shows the structured of database. The database is composed of four key tables, each serving a distinct purpose in aggregating and organizing data related to housing prices, sales volumes, and market trends. These tables include SeasonallyAdjusted, PriceData, SalesData, and YearOverYearChange, collectively offering a comprehensive dataset that facilitates a nuanced examination of market dynamics over time.

Our database comprises four primary tables, each capturing different dimensions of the housing market:

- SeasonallyAdjusted: Contains seasonally adjusted home price indices for a variety of property types, crucial for analyzing market trends devoid of seasonal fluctuations.

  Attributes:

  Important Attributes:

  - Date: Unique date of record, serving as a temporal anchor for each entry.
  - Composite_HPI_SA: The overall seasonally adjusted Home Price Index, reflecting broad market movements.

  Others:

  - Single_Family_HPI_SA, One_Storey_HPI_SA, Two_Storey_HPI_SA, Townhouse_HPI_SA, Apartment_HPI_SA, Composite_Benchmark_SA, Single_Family_Benchmark_SA, One_Storey_Benchmark_SA, Two_Storey_Benchmark_SA, Townhouse_Benchmark_SA, Apartment_Benchmark_SA
  - Primary Key: Date.
  - Dimensions: 228 rows × 13 columns.

- PriceData: Holds information on average and aggregate composite prices of residential properties across Canada, providing insight into general pricing trends.
  Attributes:
  - Date, AveragePriceCanada, AggregateCompositeCanada.
  - Primary Key: Date.
  - Foreign Keys: Date references SeasonallyAdjusted (Date).
  - Dimensions: 228 rows × 3 columns.

- SalesData: Records monthly home sales data along with a ten-year monthly sales average.
  Attributes:
  - Date, MonthlyHomeSales, TenYear_Monthly_HomeSales_Average.
  - Primary Key: Date.
  - Foreign Keys: Date references PriceData (Date).
  - Dimensions: 204 rows × 3 columns.

- YearOverYearChange: Tracks the year-over-year percentage change in the Home Price Index, offering a measure of market growth or contraction.
  Attributes:
  - Date, CompositeBenchmarkSA, YearOverYearChangePercentage.
  - Primary Key: Date.
  - Foreign Keys: CompositeBenchmarkSA references PriceData(AggregateCompositeCanada).
  - Dimensions: 216 rows × 3 columns.

Analysis of Foreign Key Relationships

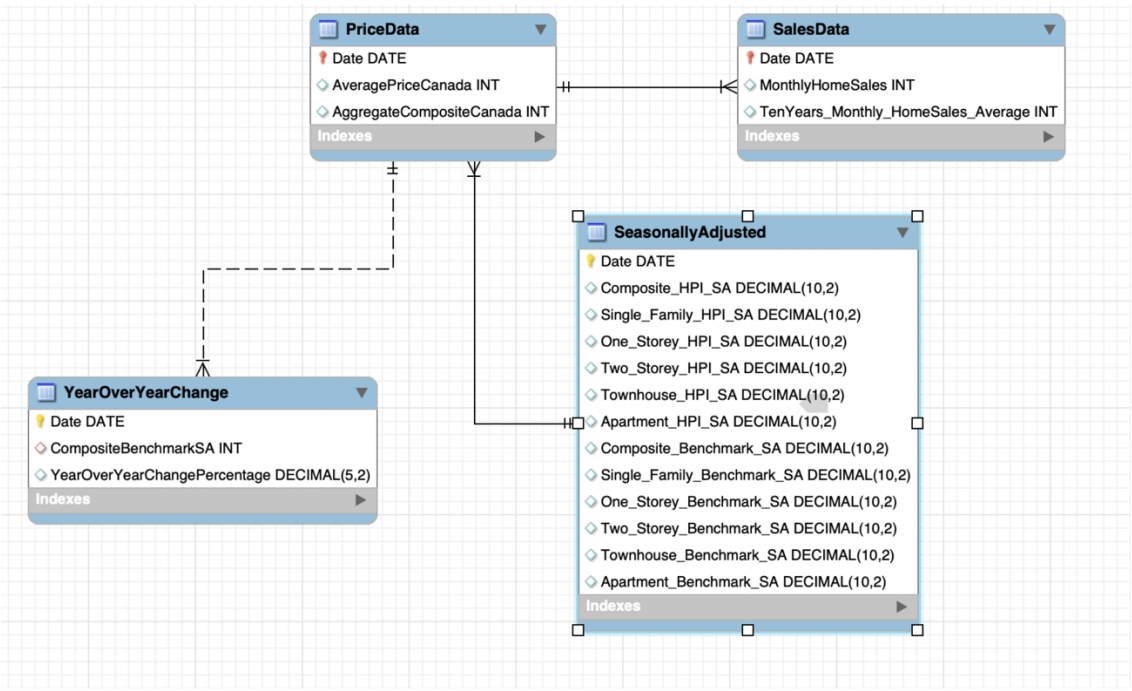
The foreign key relationships within the database ensure data integrity and facilitate complex, cross-table analyses. The **Date** attribute serves as a critical link across tables, ensuring that analyses of price adjustments, sales data, and year-over-year changes are temporally aligned. The foreign key from **YearOverYearChange** to **PriceData** ties annual market changes to aggregate pricing data.

Section 3: Internal Schema and Normalization

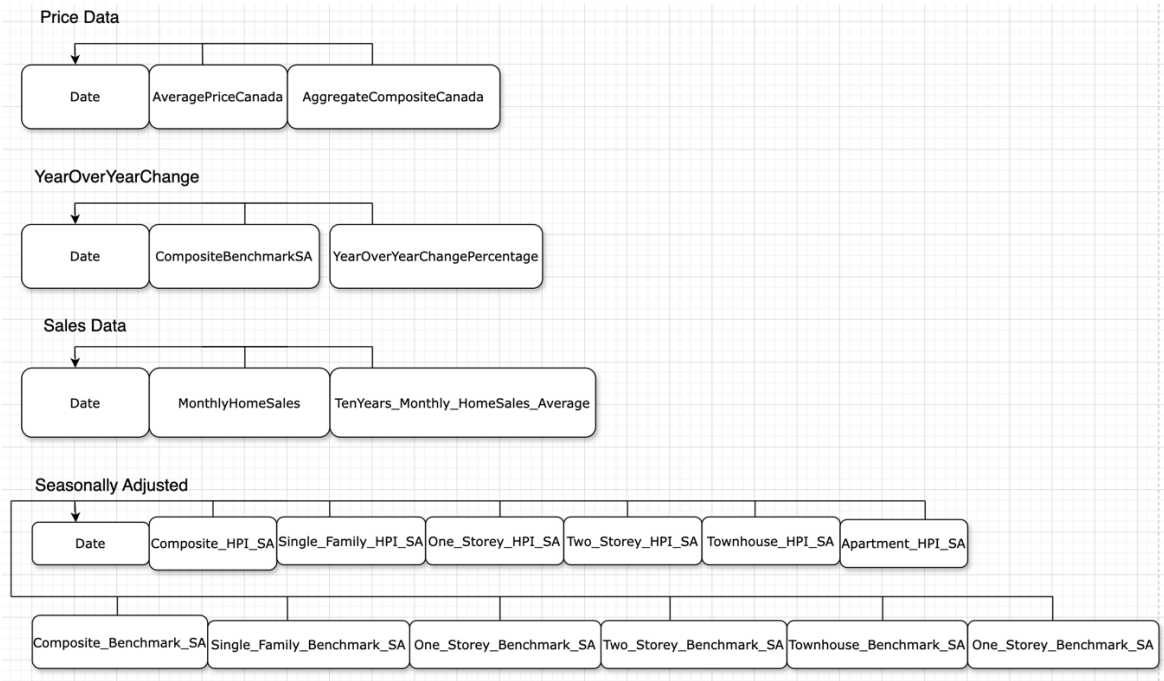
To design a database schema in 3NF, we need to ensure that:

- The schema is in second normal form (2NF).
- All attributes are non-transitively dependent on the primary key.

Internal Schema Diagram
Generated using MySQL reverse engineering tool.



Dependency diagrams



The internal schema design given above ensures each table is in 3NF:

- All are in 2NF as there are no partial dependencies of any non-primary key attribute on a part of a composite primary key (since we're using single-attribute primary keys).

- All attributes are fully functionally dependent on the primary key and only the primary key, ensuring no transitive dependencies.

Business Rules Enforced by the Database

1. Seasonal Adjustments Must Precede Price Data Entries

The foreign key constraint on PriceData (Date) referencing SeasonallyAdjusted (Date) enforces a rule that any price data entry must correspond to an existing seasonally adjusted record. This ensures chronological integrity between these datasets, implying that any adjustments for seasonal variations in property prices are accounted for before logging average and aggregate prices.

2. Unique Date Entries for Data Records

The use of PRIMARY KEY on the Date column in each table ensures that there can only be one entry per date in SalesData, SeasonallyAdjusted, PriceData, and YearOverYearChange tables. This rule enforces a clear, unambiguous timeline of data without overlaps or duplications, crucial for maintaining accurate historical records and analyses.

3. Linking Composite Benchmarks to Price Data

The foreign key constraint in YearOverYearChange linking CompositeBenchmarkSA to PriceData (AggregateCompositeCanada) ensures that any recorded change in the home price index over a year is directly associated with a corresponding aggregate composite benchmark. This rule solidifies the relationship between general market assessments and specific yearly performance metrics, allowing for more grounded and referential data analysis.

Explanation of the Queries

Query 1: Selects records from SalesData post-2023 and computes the difference between monthly sales and the ten-year average, showcasing immediate sales performance against historical averages.

Query 2: Joins SalesData and PriceData to correlate sales figures with average prices, offering insights into how sales volumes might relate to price trends on the same dates.

Query 3: Groups sales data by year and averages the sales, providing a yearly overview of sales activity and spotting trends over time.

Query 4: Uses a subquery to calculate the average difference between monthly sales and the ten-year average across all records, offering a holistic view of how current sales stand against the long-term average.

Query 5: Demonstrates the dynamic nature of views by creating a view that combines sales and price data, then updating underlying data to show how views reflect these changes immediately.

Using the Stored Procedures

1. Update Record SalesData:

Purpose: Updates or inserts sales data for a given date, allowing for the dynamic updating of sales records.
CALL(): Provide a date ('YYYY-MM-DD'), monthly sales figure, and ten-year average as parameters.
Example: CALL UpdateRecordSalesData('2023-01-01', 33466, 43061);

This procedure does not return data via OUT or INOUT variables but modifies the database directly.

2. Insert And Update SeasonalllyAdjusted And PriceData:

Purpose: Inserts a new seasonally adjusted record and updates or inserts corresponding price data, ensuring both datasets remain synchronized.
CALL(): Provide a date, composite HPI seasonally adjusted value, average price in Canada, and aggregate composite benchmark as parameters.
Example: CALL InsertAndUpdateSeasonalAndPriceData('2024-01-01', 130.75, 600000, 650000);

Like the first procedure, this one directly modifies the database without returning data via OUT or INOUT variables. Each procedure is transaction-safe, meaning changes are either fully committed or rolled back to ensure database integrity, particularly useful in maintaining consistent and accurate records across related tables.